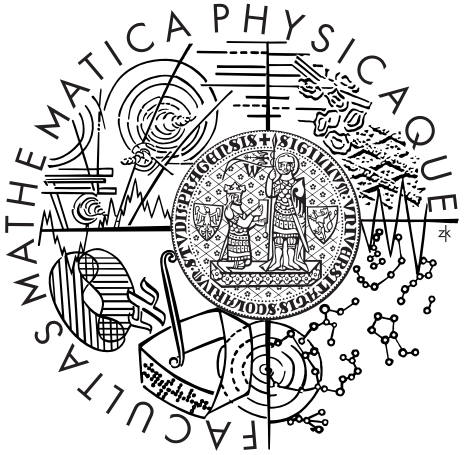
Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS



Peter Berta

Investigation of properties of the top quark with the ATLAS experiment at LHC

Institute of Particle and Nuclear Physics

Supervisor of the doctoral thesis: Prof. RNDr. Rupert Leitner, DrSc.
Study programme: Physics
Specialization: Subnuclear Physics

Prague 2016

I would like to express my gratitude to my supervisor Prof. Rupert Leitner for the continuous support of my Ph.D study. His guidance and physics insight helped me a lot in research and writing of this thesis.

I am very grateful to my advisor Jiří Kvita for his helpful discussions about physics and especially about the measurement of the top-antitop differential cross sections.

My sincere thanks also go to Martin Spousta with whom I worked on a novel pileup mitigation method for jets.

I would like to thank to David W. Miller who put forward the issue of pileup mitigation for jets at the ATLAS experiment. I gained large experience thanks to David's guidance and cooperation.

I am grateful to the ATLAS collaborators Jean-Francois Arguin, Francesco Spano, Matteo Negrini, Ana Ovcharova, Federica Fabbri, Alberto Mengarelli, Matteo Franchini, and Ofir Gabizon for their cooperation in the measurement of the top-antitop differential cross section.

I would like to thank to Fabrizio Parodi, Frank Filthaut, Tim Scanlon, and Giacinto Piacquadio for their guidance within the ATLAS Flavour Tagging Working Group.

Many thanks belong to my colleagues Tomáš Kadavý and Stanislav Štefánik for proof-reading parts of my thesis.

Special thanks to my dear parents, to my brother Milan, and to my beloved wife Ruženka for their support.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague on 27-Apr-2016 Peter Berta

Název práce: Zkoumání vlastností top kvarku pomocí experimentu ATLAS na LHC

Autor: Mgr. Peter Berta

Katedra: Ústav částicové a jaderné fyziky

Vedoucí disertační práce: Prof. RNDr. Rupert Leitner, DrSc., Ústav částicové a jaderné fyziky

Abstrakt: Tato disertační práce prezentuje měření diferenciálního účinného průřezu produkce párů kvarků top-antitop ve srážkách protonů s těžišťovou energií 8 TeV na experimentu ATLAS. Toto měření je provedeno pomocí událostí top-antitop v rozpadovém kanálu s jedním nabitým leptonem pro top kvarky s vysokou příční hybností. Měřený diferenciální účinný průřez je vyjádřen jako funkce příční hybnosti top kvarku na částicové a partonové úrovni. Změřená spektra jsou v souhlasu s předpovědí Standardního Modelu. V práci jsou popsané experimentální metody používané pro toto měření s důrazem na rekonstrukci jetů a identifikace jetů pocházejících z fragmentace b-kvarků. Rekonstrukce jetů je ovlivněna srážkami více protonů (pileup) a je studováno potlačení těchto efektů. Je navržen nový způsob na odstranění pileup efektů, který je založen na metodách v současnosti používaných na LHC experimentech. Tato nová metoda účinně opravuje jety na úrovni jejich konstituentů.

Klíčová slova: top kvark, urychlovač LHC, experiment ATLAS, pileup, b-jety

Title: Investigation of properties of the top quark with the ATLAS experiment at LHC

Author: Mgr. Peter Berta

Department: Institute of Particle and Nuclear Physics

Supervisor: Prof. RNDr. Rupert Leitner, DrSc., Institute of Particle and Nuclear Physics

Abstract: This thesis presents the measurement of the differential cross section of the top-antitop pair production in proton-proton collisions at center-of-mass energy of 8 TeV at the ATLAS experiment. The measurement is performed for top-antitop events in the single lepton decay channel in the boosted topology. The measured differential cross section is expressed as a function of the top quark transverse momentum at particle level and at parton level. The measured distributions are compatible with the theoretical predictions of the Standard Model. The experimental techniques used in this measurement are discussed with emphasis on the jet reconstruction and on the identification of jets originating from b-quark fragmentation. The jet reconstruction is influenced by simultaneous proton-proton collisions (pileup), and the mitigation of these pileup effects is studied. A novel pileup subtraction technique for jets is proposed using an extension of the methods currently being employed by the LHC experiments. The new method has a good performance in removing the pileup contributions at the level of jet constituents.

Keywords: top quark, LHC accelerator, ATLAS experiment, pileup, b-jets

Contents

Introduction

The main topic of this thesis is the measurement of the differential cross section of the top-antitop pair production at the ATLAS experiment at center-of-mass energy of 8 TeV. The measurement is done in the decay channel containing one charged lepton and the cross section is measured as a function of the top quark transverse momentum above 300 GeV. Jet substructure techniques are used to identify hadronically decaying top quarks with large transverse momentum. The measured spectra are compared to Standard Model predictions.

This thesis deals with other two topics connected with the top quarks: the identification of jets originating from b-quark fragmentation and the mitigation of effects on jets originating from additional proton-proton collisions (pileup). The identification of jets originating from b-quark is crucial for the reconstruction of top-antitop pairs and for the suppression of background events. The pileup has negative effect on the reconstruction of jets, and it lowers the ability to identify top quarks with large energy. Therefore, methods mitigating the pileup effects are important for most of the analyses performed at the LHC.

The thesis structure and the author's contributions to publications are outlined as follows. In Sec. 1, the overview of the Standard Model is summarized, focusing on the top quark. The ATLAS detector and the reconstruction of physics objects with the ATLAS detector are described in Sec. 2. The Sec. 3 describes the methods for mitigation of pileup effects. The author tested one such method on data from the ATLAS experiment (Sec. 3.3), which was published as a public ATLAS Note in [1]. The author proposed a novel method called Constituent Subtraction (Sec. 3.4) and is the main author of the publication [2] describing this method. The author contributed to the Flavor Tagging Combined Performance Group of the ATLAS collaboration by evaluating the Monte Carlo efficiencies of the jet flavor identification. This work is summarized in Sec. 4, and the author made few contributions to the refereed ATLAS publication [3] and the public ATLAS Note [4]. The author made significant contribution to the measurement of the differential cross section for top-antitop pair production at the ATLAS experiment which is presented in Sec. 5, and which was published in the refereed ATLAS publication [5].

The author presented results at three conference talks. Two conference contributions about the Constituent Subtraction pileup mitigation method were presented at the Workshop on Boosted Object Phenomenology on 21-Aug-2014, and at the Workshop on Mitigation of pileup effects at the LHC on 16-May-2014. The third conference contribution about the jet and missing transverse energy reconstruction at the ATLAS experiment was presented on behalf of the ATLAS Collaboration at the International Conference on High Energy Physics on 3-Jul-2014. The content of this talk is published as a conference proceeding in [6].

The reader is encouraged to read the App. A-C where the basic conventions, the jet clustering algorithms, and the jet shape observables are introduced, and which are commonly used at the LHC experiments.

1. Top Quark

The top quark is the heaviest elementary particle within the Standard Model (SM) of elementary particles and their interactions. The top quark is the only quark that does not form bound states due to its extremely short lifetime. The measurement of the top-antitop ($t\bar{t}$) pair production probes our understanding of the strong interactions and predictions of perturbative quantum chromodynamics, while the decay of the top quark and the production of single top quark examine the electroweak interactions. The measurements of the differential cross section of the $t\bar{t}$ pair production are therefore important to confirm the SM predictions or to search for new physics beyond the SM.

The top quark was experimentally discovered in 1995 at the Tevatron proton-antiproton ($p\bar{p}$) collider in FNAL, Chicago. The discovery was confirmed independently at two experiments, CDF and D0. The only place where the top quarks are produced and detected nowadays is the Large Hadron Collider (LHC) at CERN, Geneva. Two multipurpose detectors, ATLAS and CMS, at this proton-proton (pp) collider make measurements in top quark physics. The LHC operates since 2010, and until the end of 2012 more than $\sim 6 \cdot 10^6$ top quark events were produced per experiment that was approximately 100-times more than at the Tevatron collider.

This section summarizes the basic properties of the top quark, and gives an overview of the SM predictions and the measurements of the differential cross section of $t\bar{t}$ production.

1.1 Top Quark in the Standard Model

The SM classifies all known elementary particles. It describes the electromagnetic, weak and strong interactions and the generation of masses of gauge bosons and fermions. A detailed overview of the SM can be found in [7]. A brief summary is given in the following. The elementary particles of the SM are depicted in Fig. 1.1. The strong interactions of quarks with gluons are described by the Quantum Chromodynamics (QCD) which is a quantum field theory based on the gauge group $\mathrm{SU}(3)_C$. The QCD predicts the confinement of quarks in hadrons which makes it impossible to observe free quarks. Another property of the QCD is the asymptotic freedom which allows to assume the hadron constituents, quarks and gluons, as free particles in high-energy hadron collisions. The electromagnetic and weak interactions of quarks and leptons with bosons γ, W, and Z are described by the Electroweak Theory (EWT) which unifies the electromagnetic and weak interactions through the gauge group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$.

The masses of gauge bosons, W and Z, in the SM are generated by the spontaneous symmetry breaking of $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ which is called the Brout-Englert-Higgs mechanism. This mechanism predicts a new particle, called Higgs boson, which was experimentally discovered in 2012 [8, 9]. The masses of fermions are generated from Yukawa interaction of fermions with the Higgs boson.

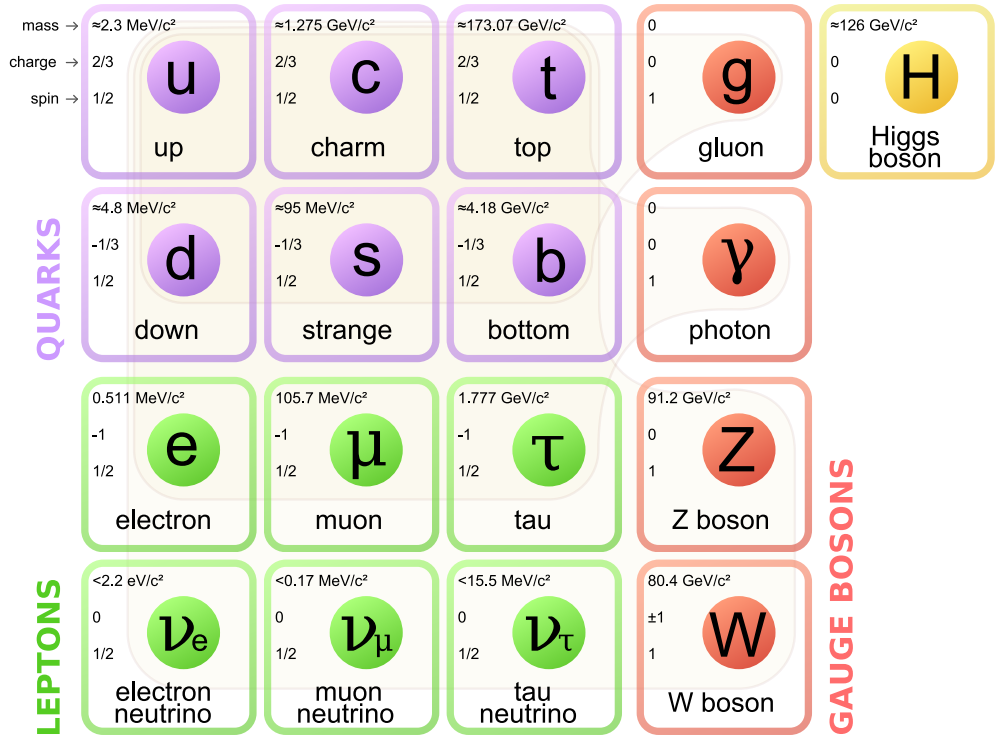


Figure 1.1: Elementary particles of the SM: the 12 fundamental fermions and 5 fundamental bosons. Brown loops indicate which bosons (red) couple to which fermions (purple and green). The electric charge is given in units of electric charge of proton. [10]

1.1.1 Mass of the Top Quark

The mass of the top quark, m^{top}, is a free parameter in the SM. The world combination of several measurements gives the value $m^{\text{top}}_{\text{meas}} = (173.34 \pm 0.76)$ GeV, [11]. It is important to define which theoretical framework is used when referring to quark masses, see [12, p. 725]. The m^{top} is renormalization scheme dependent, e.g. in the pole mass scheme, the top quark mass, $m^{\text{top}}_{\text{pole}}$, corresponds to the real part of the pole in the top quark propagator. The measured value of m^{top} from certain analysis corresponds with the top quark mass defined in the Monte Carlo (MC) generator used for calibration in that particular analysis. The MC mass definition can differ between two MC generators. It is non-trivial to match a MC mass definition with the pole mass definition. It is expected that the difference between top masses in certain MC mass definition and the pole mass definition is up to the order of 1 GeV [13].

The top quark has a special position in the SM due to its extremely large mass. It is at least 11 orders of magnitude heavier than the lightest elementary fermion and the top quark is the only fermion which is heavier than the gauge bosons W and Z and the Higgs boson. The SM cannot explain this mass hierarchy.

1.1.2 Production of Top Quarks in Hadron Collisions

In the parton model, the hadrons are composed of partons (quarks, antiquarks and gluons) which act as free particles in high-energy hadron collisions due to the asymptotic freedom in QCD. This property is used in calculations of observables

in high-energy hadron collisions using the QCD factorization theorem [14] which divides the hadron collision into two components by introducing a factorization scheme with a factorization scale, μ_F, in the calculation. These two components are long-distance phenomena and short-distance phenomena. The long-distance phenomena are universal for each process and they result in a phenomenological description of hadrons. A certain type of parton a carries fraction x_a of the 4-momentum of hadron h. The momentum fraction x_a is described by the parton distribution function $f_{a/h}(x_a, \mu_F)$ (PDF) where the quantity $f_{a/h}(x_{a,0}, \mu_F)\mathrm{d}x_a$ is the probability that the parton a carries fraction of 4-momentum of hadron within infinitesimal interval $[x_{a,0}, x_{a,0} + \mathrm{d}x_a]$. The short-distance phenomena define observables from interaction of partons, i.e. the building blocks of the SM, and therefore they are calculable in perturbative expansion. This calculation can be usually done to certain fixed order by introducing a renormalization scheme with renormalization scale, μ_R.

There are two ways to create top quarks from colliding partons, via top-antitop ($t\bar{t}$) pair production or single top production. The main contribution to the $t\bar{t}$ pair production comes from the QCD (gluon-gluon fusion $gg \to t\bar{t}$ or quark-antiquark annihilation $q\bar{q} \to t\bar{t}$). The EWT gives the main contribution to the cross section in s-channel and t-channel single top production which is represented by LO QCD partonic processes $q_1\bar{q}_2 \to t\bar{b}$ and $q_1 b \to q_2 t$, respectively. The single top production can be associated with a W boson in QCD+EWT process ($gb \to tW^-$). The $t\bar{t}$ production cross section is approximately twice higher than the single top production cross section in pp collisions at $\sqrt{s} = 8$ TeV. This thesis focuses on the $t\bar{t}$ pair production while the single top production is assumed as a possible background when selecting $t\bar{t}$ events.

The total cross section of $t\bar{t}$ production in pp collisions at center-of-mass energy, $\sqrt{s}$, can be calculated using the QCD factorization theorem:

$$\begin{aligned}&\sigma_{pp\to t\bar{t}X}\left(\sqrt{s}, \mu_R, \mu_F\right)\\ &= \sum_{p_1}\sum_{p_2}\int\int \sigma_{p_1p_2\to t\bar{t}Y}\left(\sqrt{s}, \mu_R, \mu_F, x_1, x_2\right) f_{p_1/p}\left(x_1, \mu_F\right) f_{p_2/p}\left(x_2, \mu_F\right)\mathrm{d}x_1\mathrm{d}x_2,\end{aligned} \tag{1.1}$$

where X and Y represents any additional particles, and the two sums runs over all proton constituents. In leading-order (LO) QCD, there are only two partonic cross sections which contribute: $\sigma_{gg\to t\bar{t}}$ and $\sigma_{q\bar{q}\to t\bar{t}}$, see the Feynman diagrams in Fig. 1.2. The Fig. 1.2a represents the LO QCD contribution to the quark-antiquark annihilation, and the interference of diagrams in Fig. 1.2b, 1.2c and 1.2d represents the LO QCD contribution to the gluon-gluon fusion. Contributions from quark-quark parton pair (e.g. $p_1 = u$ and $p_2 = d$) in Eq. 1.1 are also possible, although they are two orders below the gluon-gluon fusion in the perturbation expansion in QCD.

A physical quantity like the cross section of $t\bar{t}$ production must be independent on scales μ_R and μ_F. Although, the calculation in Eq. 1.1 depends on scales μ_R and μ_F which is given by the ability to compute the partonic cross section or the PDFs to certain fixed order. This dependence on scales μ_R and μ_F brings uncertainty in the theoretical prediction. If the calculation in Eq. 1.1 is done to all orders of the perturbation theory, the result would be independent of the scales μ_R and μ_F.

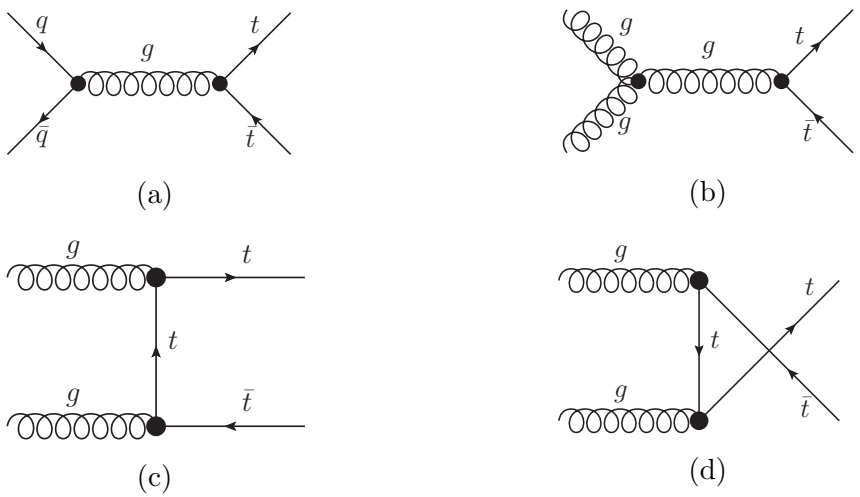


Figure 1.2: Leading order Feynman diagrams in QCD for $t\bar{t}$ production.

The Fig. 1.3 shows the theoretical predictions of total cross section of $t\bar{t}$ productions in pp and $p\bar{p}$ collisions compared to the measurement from LHC and Tevatron, respectively. The predictions are computed at next-to-next-to-leading order (NNLO) QCD complemented with soft-gluon resummation with next-to-next-to-leading logarithmic (NNLL) accuracy [15]. The NNLO+NNLL QCD prediction is in very good agreement with all measurements, including preliminary measurements at $\sqrt{s} = 13$ TeV.

The total cross section of $t\bar{t}$ production depends on the top pole mass. Using the measured world average of the top mass $m_{\text{pole}}^{\text{top}} = 173.3$ GeV, the cross section of $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8$ TeV in NNLO+NNLL QCD is $\sigma = 245.8^{+6.2+6.2}_{-8.4-6.4}$ pb [15]. The uncertainties come from the uncertainties of PDFs and the dependence on scales μ_R and μ_F. The theoretical prediction is in good agreement with the measured total cross section at the ATLAS experiment $\sigma_{\text{meas}} = (242.4 \pm 1.7 \pm 5.5 \pm 7.5 \pm 4.2)$ pb where the four uncertainties arise from data statistics, experimental and theoretical systematic effects, the knowledge of the integrated luminosity and of the LHC beam energy, [16].

1.1.3 Top Quark Decay

The top quark has three decay modes ($t \to W^+d$, $t \to W^+s$ and $t \to W^+b$) in the SM with branching ratios related to the elements of the CKM matrix V_{td}, V_{ts} and V_{tb}, see [12, p. 214]. The global fit of available measurements on the CKM matrix elements gives $V_{tb} > 0.999$ which implies that the decay channel $t \to W^+b$ is dominant, with branching ratio almost 100%. The antitop quark $\bar{t}$ decays accordingly to the charge conjugation: $\bar{t} \to W^-\bar{b}$.

The decay width of the top quark, Γ^{top}, at next-to-leading order (NLO) QCD was calculated more than 25 years ago: $\Gamma^{\text{top}}_{\text{NLO}} = 1.33$ GeV for $m^{\text{top}}_{\text{pole}} = 172.5$ GeV, [18]. From then on, a more precise prediction was obtained with NNLO QCD, NLO EWT corrections, finite bottom quark mass and W boson width effects, for details see [19]. The decay width was measured at the D0 experiment: $\Gamma^{\text{top}}_{\text{meas}} = 2.00^{+0.47}_{-0.43}$ GeV, [20], which is the most precise measurement of the top decay width by now, although this measurement is model-dependent. A model-independent

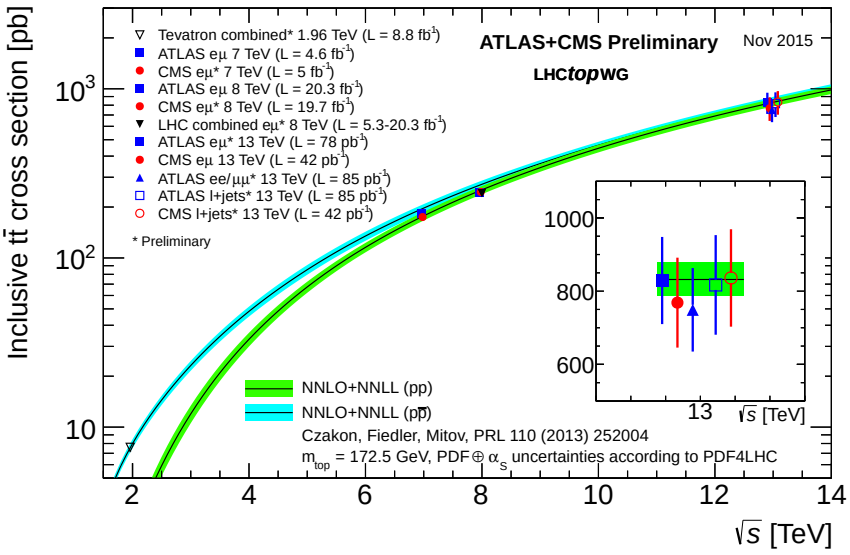


Figure 1.3: Summary of LHC and Tevatron measurements of the $t\bar{t}$ production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation (top++2.0). The theory band represents uncertainties due to scales μ_R and μ_F, PDFs and the strong coupling. The measurements and the theory calculation are quoted at $m^{\text{top}} = 172.5$ GeV. Measurements made at the same center-of-mass energy are slightly offset for clarity. [17]

measurement was performed by the CDF experiment leading to 1.10 GeV $< \Gamma^{\text{top}}_{\text{meas}} <$ 4.05 GeV at 68% confidence level for $m^{\text{top}}_{\text{pole}} = 172.5$ GeV [21]. All the measurements of the top width are in agreement with the SM prediction.

Based on the theoretical decay width $\Gamma^{\text{top}}_{\text{NLO}}$, the mean lifetime of the top quark is only ∼0.15 fm, therefore the top quark is expected to decay instead of forming hadrons containing top quark, for details see [22].

Any measurement involving identification of $t\bar{t}$ pairs must define its selection criteria, which depend on the signature of their decay products. There are more possibilities for the $t\bar{t}$ signature in a LHC detector which is overviewed in the following. The b-quark from the top quark fragments to a b-jet. The W boson has very large decay width of (2.085 ± 0.042) GeV [12, p. 562], and therefore, it almost immediately decays. The $t\bar{t}$ pair signature in a detector depends on the decay mode of the two W bosons from the decay $t\bar{t} \to W^-bW^+\bar{b}$. The W boson decay channels are listed in Tab. 1.1. The W can decay hadronically which usually leads to signature of two jets in a detector. The signature of leptonic decay of W boson depends on the lepton type. The electron is stable particle which can be detected. The muon has mean lifetime of ∼660 m which is large enough to not decay in a detector. The τ lepton has mean lifetime of ∼90 μm which means that it decays before entering a detector or in its first detection layers. The τ lepton decays leptonically or hadronically, see Tab. 1.2. In case of leptonic decay, the identification of τ lepton is practically impossible, and its signature is one charged lepton and missing transverse energy. Therefore, the signature of a W boson decaying to leptonic τ is similar to the signature of a W boson decaying to electron or muon. The identification of a hadronic τ is possible, although other objects

can be misidentified as hadronic τ, which makes this identification challenging. Overall, there are three main decay channels of $t\bar{t}$ pairs: all-hadronic, single lepton and dilepton channels. The exact definitions of these channels depend on the classification of decay chains containing τ lepton - this is a choice for particular measurement. One possible classification of the $t\bar{t}$ decay channels is presented in Tab. 1.3, in which the definition of the single lepton channel is the same as in the measurement presented in Sec. 5. The single lepton channel has relatively unique signature with one charged lepton, missing energy due to one or more neutrinos, and four jets from which at least two jets originates from b-quark, while the branching ratio is relatively high. The all-hadronic channel has even higher branching ratio, but its signature is not so unique due to large background from QCD multijet production. The dilepton channel has also relatively unique signature with two charged leptons, but its branching ratio is lower than for single lepton channel. The decay channels containing hadronic τ can be also used in measurements, although these measurements are usually less precise with respect to the measurements in other decay channels.

In the next, hadronic top quark means a top quark which decays to hadronically decaying W boson, and leptonic top quark means a top quark decaying to leptonically decaying W boson.

Table 1.1: Main decay channels of the W^+ boson with branching ratios (BR). The W^- boson decays according to the charge conjugation. [12, p. 563]

W^+ decay	BR [%]
hadrons	67.41 ± 0.27
$e^+\nu_e$	10.71 ± 0.16
$\mu^+\nu_\mu$	10.63 ± 0.15
$\tau^+\nu_\tau$	11.38 ± 0.21

Table 1.2: Main decay channels of τ^- lepton with branching ratios (BR). The branching ratio for decay to hadrons is estimated as the difference between unity and branching ratios for decay to electron and muon. The τ^+ lepton decays according to the charge conjugation. [12, p. 659]

τ^- decay	BR [%]
$e^-\bar{\nu}_e\nu_\tau$	17.83 ± 0.04
$\mu^-\bar{\nu}_\mu\nu_\tau$	17.41 ± 0.04
hadrons ν_τ	64.76

1.1.3.1 Resolved vs Boosted Top Quark

The signature of a hadronic top quark in a detector is several jets. One commonly used jet algorithm[1] is the anti-k_t $R = 0.4$. Given that a hadronic top quark results in three quarks at lowest tree level of QCD, the number of anti-k_t $R = 0.4$ jets is usually three. This number can be larger than three as predicted by higher order

[1] See App. B for more information about jet algorithms.

Table 1.3: Decay channels of the $t\bar{t}$ pair dependent on the decays of the two W bosons from the $t\bar{t}$ pair decay. Branching ratios (BR) and typical detector signatures are listed. The symbol ℓ represents electron or muon. No distinction between particles and antiparticles is done. MET is the missing transverse energy.

Channel	one W	the other W	BR [%]	$t\bar{t}$ signature
all-hadronic	hadrons	hadrons	45.7	6 jets
single lepton	hadrons	$\ell\nu_\ell$	28.8	e or μ, MET, 4 jets
	hadrons	$\tau\nu_\tau \to \ell\nu_\ell\bar{\nu_\tau}\nu_\tau$	5.4	
dilepton and tauonic channels	$\ell\nu_\ell$	$\ell\nu_\ell$	4.5	ee or $e\mu$ or $\mu\mu$, MET, 2 jets
	$\ell\nu_\ell$	$\tau\nu_\tau \to \ell\nu_\ell\bar{\nu_\tau}\nu_\tau$	1.7	
	$\tau\nu_\tau \to \ell\nu_\ell\bar{\nu_\tau}\nu_\tau$	$\tau\nu_\tau \to \ell\nu_\ell\bar{\nu_\tau}\nu_\tau$	0.2	
	hadrons	$\tau\nu_\tau \to$ hadr. $\bar{\nu_\tau}\nu_\tau$	9.8	4 jets, hadr. τ, MET
	$\ell\nu_\ell$	$\tau\nu_\tau \to$ hadr. $\bar{\nu_\tau}\nu_\tau$	3.1	2 jets, 2 hadr. τ, MET
	$\tau\nu_\tau \to$ hadr. $\bar{\nu_\tau}\nu_\tau$	$\tau\nu_\tau \to$ hadr. $\bar{\nu_\tau}\nu_\tau$	0.5	2 jets, 2 hadr. τ, MET

QCD description of the decay. On the other hand, this number can be lower than three, if the three quarks are close to each other in terms of the distance ΔR which is used in the anti-k_t algorithm. This closeness of quarks is the subject of this subsection which leads to classification of top quarks as *resolved* or *boosted*.

The closeness of the three quarks from hadronic top quark can lead to finding less than three jets or to finding three jets from which one jet contains only part of particles originating from one quark (the remaining part is contained in the other two jets). The fraction of such events increases with the top quark p_{T} by a Lorentz boost, see Fig. 1.4. A resolved top quark has signature of three well-separated jets, and each jet contains large fraction of particles originating from one of the three quarks. A boosted top quark has signature of three or less jets which may share the particles among each other from the original three quarks. There is no clear boundary in definition of resolved and boosted top quarks.

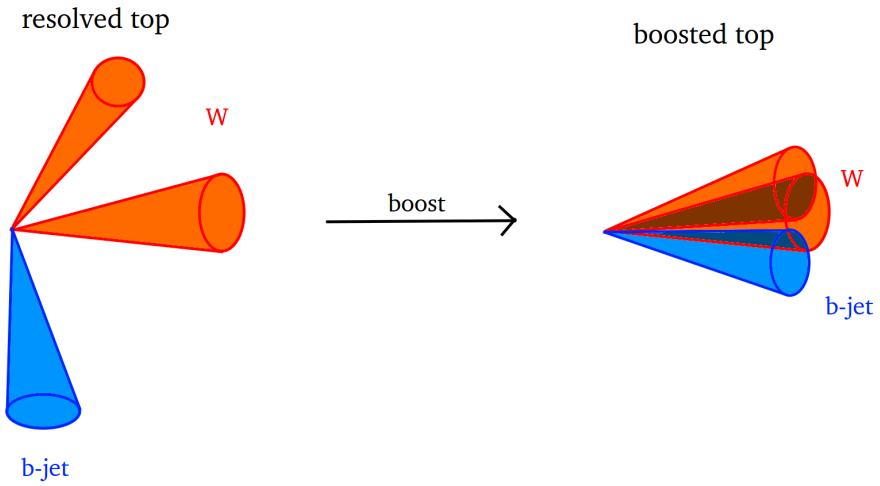


Figure 1.4: Illustration of the difference between resolved and boosted hadronic top quarks. The signature of resolved hadronic top quark is three well-separated jets. A boosted hadronic top quark has higher energy which can lead to overlap of the particles from the original three quarks in the resulting jets.

In case the top quark is resolved, the reconstruction technique of the top quark

takes into account the kinematics: invariant mass of certain two jets should be close to the W boson mass, and the invariant mass of certain three jets should be close to the top quark mass. For example, in case of $t\bar{t}$ events in lepton+jets channel, a kinematic likelihood fit can be used as is described in [23].

With increasing p_{T}, the jets start to overlap, and the efficiency to reconstruct the top quark with resolved reconstruction techniques is getting lower. The Fig. 1.5 shows the distribution of distance ΔR between the W boson and b-quarks from the top decay vs $p_{\mathrm{T}}^{\mathrm{top}}$ for a sample of top quarks. The mean value of the distance ΔR for top quarks with $p_{\mathrm{T}}^{\mathrm{top}} \sim 300$ GeV is $\sim$0.8, and it is decreasing with increasing $p_{\mathrm{T}}^{\mathrm{top}}$. This roughly means that for $p_{\mathrm{T}}^{\mathrm{top}} \gtrsim 300$ GeV, the resolved reconstruction techniques using anti-k_t $R = 0.4$ jets may be inefficient, and other (boosted) techniques are needed. The idea of these boosted techniques is to use jet clustering algorithm with large distance parameter, e.g. anti-k_t $R = 1.0$ or C/A $R = 1.2$ which ensures that large fraction of the boosted top quark decay products are clustered into one such jet. Using the jet mass and the inner structure (substructure) of this jet, one can discriminate between jets originating from a top quark or from other partons. There are several boosted top quark identification techniques, and they were tested at the ATLAS experiment as described in [24].

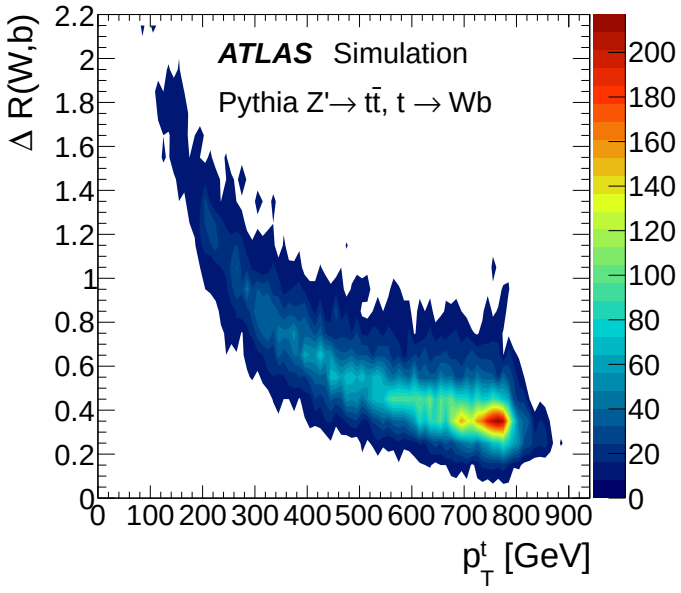


Figure 1.5: The angular separation between the W boson and the b-quark in top decays, $t \to Wb$, as a function of the top quark p_{T} in simulated Pythia $Z' \to t\bar{t}$ ($m_{Z'} = 1.6$ TeV) events. The distribution is at the generator level and does not include effects due to initial and final-state radiation, or the underlying event. [25]

1.2 Differential Cross Sections for $t\bar{t}$ Production

The measurement of the differential cross section of $t\bar{t}$ production is the main topic of this thesis (Sec. 5). The differential cross section, $\mathrm{d}\sigma/\mathrm{d}v$, can be expressed

as a function of a kinematic variable, v. From theoretical point of view, the most interesting variables are: transverse momentum (p_{T}) of the top quark[2], $\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{\mathrm{top}}$, rapidity of the top quark, $\mathrm{d}\sigma/\mathrm{d}y^{\mathrm{top}}$, transverse momentum of the $t\bar{t}$ pair, $\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{t\bar{t}}$, and invariant mass of the $t\bar{t}$ pair, $\mathrm{d}\sigma/\mathrm{d}m^{t\bar{t}}$. These differential cross sections are also called as $p_{\mathrm{T}}^{\mathrm{top}}$ distribution, y^{top} distribution, $p_{\mathrm{T}}^{t\bar{t}}$ distribution and $m^{t\bar{t}}$ distribution, respectively.

The measurements of these differential cross sections are important for many reasons. They provide test of the SM predictive power. Also, new physics theories beyond the SM can modify the predictions for these spectra. These new theories can be rejected after comparing their predictions with measurements. The deviation between prediction and measurement can occur mainly for the higher $p_{\mathrm{T}}^{\mathrm{top}}$ range or in the $m^{t\bar{t}}$ distribution. A bump in the $m^{t\bar{t}}$ distribution would indicate a new resonance decaying to $t\bar{t}$ pairs, an example candidate can be a hypothetical boson Z' [26]. The measurements can be included in the PDF fits or in the estimation of optimal MC generator parameters. Especially, the high $p_{\mathrm{T}}^{\mathrm{top}}$ region can make the gluon PDF fit more precise for high fraction of proton 4-momentum (x), since currently there is low number of available experimental data to constrain the high x region for gluon PDF.

The state-of-the-art SM predictions and measurements of the $t\bar{t}$ differential cross section are presented in the following. The terms *parton level* and *particle level* are explained as well. The focus is given to the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution since this distribution is measured in the presented thesis.

1.2.1 Standard Model Predictions

In order to make any conclusion from a measurement, one needs to have the SM prediction with its uncertainties. The basis for any SM prediction of the $t\bar{t}$ differential cross section is the QCD factorization theorem which leads to equation similar to the one for the total cross section in Eq. 1.1. To compute the differential cross section, two main parts are needed: the knowledge of PDFs and the knowledge of partonic differential cross section. The partonic differential cross section can be computed in perturbative expansion in QCD and EWT leading to a fixed order calculation. This calculation can be supplemented by approximate calculations. In this way, one can obtain numerical prediction of certain distributions.

Besides numerical predictions, there are also predictions from MC simulations. Using MC techniques, events can be generated randomly according to the PDFs and certain fixed order prediction of partonic differential cross sections. With large statistics of MC events, the obtained distributions should correspond to the ones got numerically using the same theory. To obtain more precise prediction, each MC event can go through parton showering which is an approximation of the missing radiation from higher orders of the perturbative expansion. To obtain

[2]No distinction between particle (top) and antiparticle (antitop) is done here. Although, the NLO QCD predicts that the p_{T} distribution of top quark is different from the p_{T} distribution of antitop quark. When referring to the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution, the average between the top distribution and the antitop distribution is assumed for both, theoretical predictions and measurements, throughout this thesis, unless indicated to the contrary. This is motivated by the fact that the measurements usually do not distinguish between top and antitop quarks.

predictions at particle level with stable particles, additional MC simulation models can be used in each event. The hadronization models simulate the transition of hadronic final state from partons, and further the decays of unstable particles are simulated. A detailed overview of the MC event generators for hadron collisions can be found in [27, 28], and the most relevant points concerning the $t\bar{t}$ production are summarized in this section.

Any SM prediction is affected by uncertainties. The main uncertainties are related to the PDF, to the scales μ_R and μ_F, and to the strong coupling constant. There are several groups which construct PDF sets with different techniques. Each PDF set has its own uncertainties. When making a SM prediction, one can use one particular PDF set, and use the others to estimate the PDF related uncertainty or make the SM prediction for each PDF set separately. The fixed order calculation depends on unphysical scales μ_R and μ_F which brings uncertainty to the prediction. Another uncertainty which has impact on both PDF and partonic cross section is the uncertainty on the strong coupling constant, α_{S}. There are also several smaller factors which can introduce uncertainties to the $t\bar{t}$ differential cross section such as the top quark mass, the electroweak coupling constant, α_{EW}, and the parton shower and the hadronization modeling in case of stable particle predictions.

1.2.1.1 Fixed Order Calculation

The main contribution to the SM prediction of $t\bar{t}$ production comes from QCD, although the EWT also contributes, and it is non-negligible in certain regions of the phase space. The total cross section has been known to the order NNLO QCD since 2013 but the prediction of the differential cross section was more challenging. The NLO QCD differential cross section had been computed already in year 1991 [29], and it took next 25 years to compute the differential prediction at NNLO QCD [30, 31]. Nowadays, the NNLO QCD prediction is available for the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution up to 400 GeV. The comparison of LO, NLO, NNLO QCD predictions of the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution is shown in Fig. 1.6. The NLO QCD prediction is significantly higher with respect to the LO QCD prediction, although it lies within the scale uncertainty of the LO QCD prediction. Similarly, the addition of next order to the NLO prediction (resulting to NNLO QCD) changes the normalization and the shape of the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution. For low $p_{\mathrm{T}}^{\mathrm{top}}$, the NNLO QCD prediction is $\sim$10% higher than the NLO QCD prediction - this is also observed for the total $t\bar{t}$ cross section according to [32]. For $p_{\mathrm{T}}^{\mathrm{top}} \gtrsim 300$ GeV up to the available $p_{\mathrm{T}}^{\mathrm{top}} = 400$ GeV, the NNLO QCD prediction is slightly lower than the NLO QCD prediction (the cross section belonging to this phase space region is $\sim$3% from the total cross section).

The NNLO QCD prediction involves terms with QCD coupling constant, α_{S}, of orders: α_{S}^2, α_{S}^3, and α_{S}^4. There are computations of EWT corrections to the QCD prediction which involve the electroweak coupling constant, α_{EW}, at orders: α_{EW}^2, $\alpha_{\mathrm{S}}\alpha_{\mathrm{EW}}$, and $\alpha_{\mathrm{S}}^2\alpha_{\mathrm{EW}}$. As shown in [33], the terms at order α_{EW}^2 and $\alpha_{\mathrm{S}}\alpha_{\mathrm{EW}}$ are negligible with respect to the LO QCD prediction, and the term at order $\alpha_{\mathrm{S}}^2\alpha_{\mathrm{EW}}$ contributes significantly in certain regions of the phase space. The QED and purely weak corrections can be treated separately. The weak correction is significant mainly for high $p_{\mathrm{T}}^{\mathrm{top}}$, see Fig. 1.7 in which only the prediction for Higss mass of 126 GeV is relevant. The weak correction adds negative contribution up

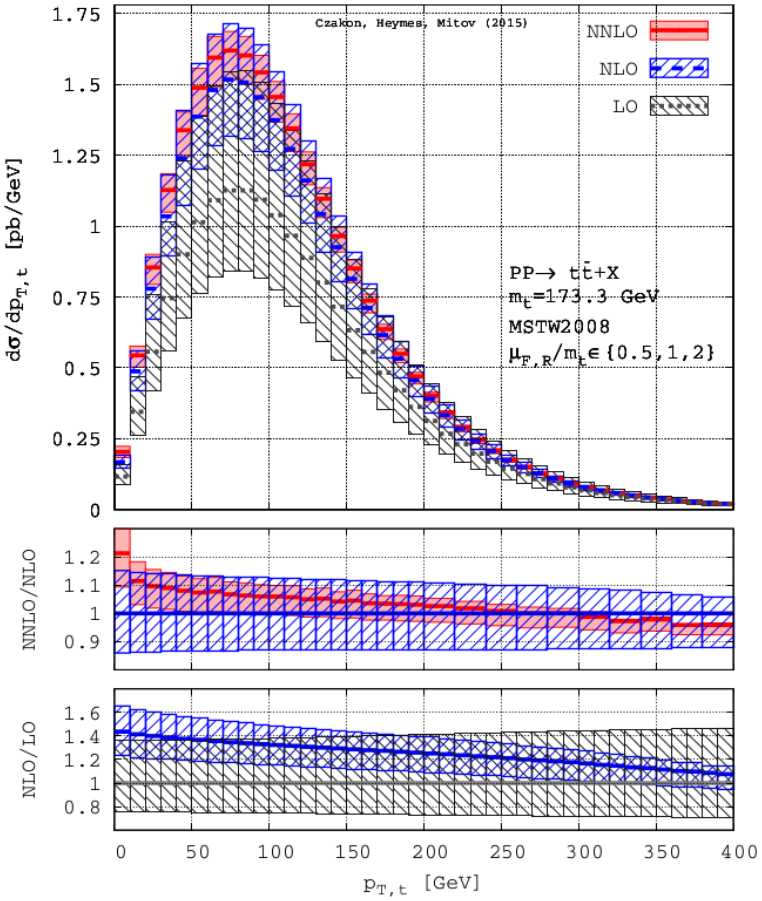


Figure 1.6: Top p_{T} distribution in LO, NLO and NNLO QCD at LHC at $\sqrt{s} = 8$ TeV. Error bands from scale variation only. [30]

to ∼15% to the LO QCD prediction in $p_{\mathrm{T}}^{\mathrm{top}}$ range $[0, 1200]$ GeV. The QED corrections contribute to three partonic processes: gluon-gluon fusion, quark-antiquark annihilation, and photon induced $t\bar{t}$ production ($g\gamma \to t\bar{t}X$) [34]. The total effect of the QED correction to gluon-gluon fusion and quark-antiquark annihilation varies from positive contribution of ∼1% to negative contribution of ∼2% to LO QCD in $p_{\mathrm{T}}^{\mathrm{top}}$ range $[0, 1200]$ GeV. The photon induced $t\bar{t}$ production can give non-negligible positive contribution which can cancel the weak corrections, although its uncertainty is large due to the large uncertainty of the photon PDF of proton [35].

1.2.1.2 Approximate Calculation

Before the full NNLO QCD prediction was available, improvements to the NLO QCD results were achieved by computing approximate NNLO corrections. One way is using soft-gluon threshold resummation methods in which the logarithmic contributions associated to the emission of soft gluons from the initial state can be added at next-to-next-to-leading logarithmic (NNLL) order, [36].

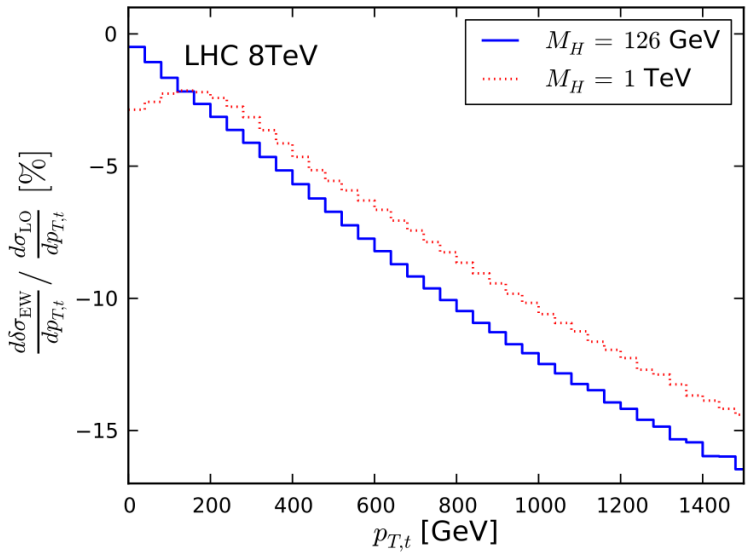


Figure 1.7: Relative weak corrections to the LO QCD for the invariant $p_{\rm T}^{\rm top}$ distribution for LHC at $\sqrt{s} = 8$ TeV for Higgs masses of 126 GeV and 1 TeV. [33]

1.2.1.3 Off-Shell and Interference Effects

All the calculations mentioned in Sec. 1.2.1.1 and 1.2.1.2 treat the top quark as a stable on-shell particle. Simulation of the top decays can be attached in the *narrow width approximation*, and the spin correlations between the produced top quark and the same decaying top quark can be included, as it is described in [37]. In this way, the $t\bar{t}$ production is treated as a separate physics process. This is the approach in most applications, such as the MC $t\bar{t}$ event generation, nowadays.

The decoupling of the top quark production and its decay is only an approximation. More correctly, the physics process involving particular set of final state particles (e.g. $pp \to e^-\bar{\nu}_e\mu^+\nu_\mu b\bar{b}X$) should be computed in which the top quark is treated as an off-shell particle, and the interference between amplitudes with and without top quark propagator is allowed. Fig. 1.8 shows selection of Feynman diagrams contributing to the partonic process $gg \to e^-\bar{\nu}_e\mu^+\nu_\mu b\bar{b}$. The diagram in Fig. 1.8a represents the $gg \to t\bar{t}$ production with top quark and W boson decays. The diagram in Fig. 1.8b can be assumed as the single top production associated with a W boson at NLO QCD. The diagram in Fig. 1.8c does not contain any top quark propagator. The interference of these diagrams with other diagrams for the process $gg \to e^-\bar{\nu}_e\mu^+\nu_\mu b\bar{b}X$ should be taken into account. The computation of the process $pp \to e^-\bar{\nu}_e\mu^+\nu_\mu b\bar{b}X$ is available at NLO order (i.e. terms $\alpha_{\rm S}^2\alpha_{\rm EW}^4$ and $\alpha_{\rm S}^3\alpha_{\rm EW}^4$) in [38].

These off-shell and interference effects are usually neglected, and the $t\bar{t}$ events with stable top quarks are treated as a separate physics process. Similarly, the single top production associated with a W boson is treated as a separate physics process. This neglecting is motivated by using measurement selection criteria, which aim to select events with on-shell top quarks, e.g. by requiring the invariant mass of certain combination of jets to be close to the top quark mass. The separation of the two processes, the $t\bar{t}$ production and the single top production

associated with a W boson, is discussed in detail in [39].

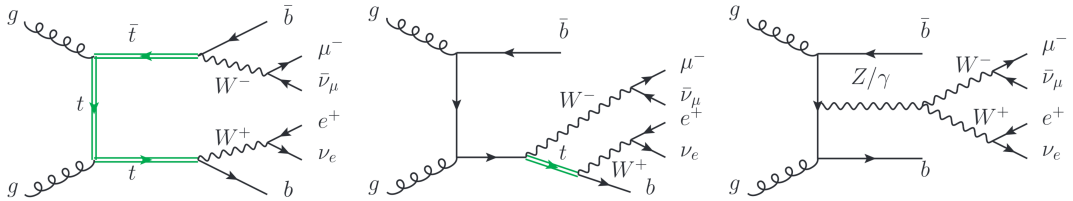


Figure 1.8: Representative set of Feynman diagrams contributing to the process $gg \to e^-\bar{\nu}_e\mu^+\nu_\mu b\bar{b}$ at leading order $\alpha_S^2\alpha_{EW}^4$. The top quark propagator is indicated with green color. [38]

1.2.1.4 Parton Shower and Hadronization

The parton shower (PS) is used in MC generators to approximately simulate the gluon radiation or gluon splitting coming from the higher order terms of the perturbation expansion which are beyond the fixed order prediction. The parton shower can be applied to the initial state partons or to the final state partons, which is called initial state radiation (ISR) and final state radiation (FSR), respectively. The PS models are built on soft and collinear approximations to the full cross sections while they cannot describe the wide-angle emissions and multi-jet final states reliably. There are several PS models, see [28].

Most MC generators generating $t\bar{t}$ events use fixed order prediction at NLO QCD which is then supplemented by PS. This matching of fixed order calculation and PS needs to be done carefully to avoid double-counting of phase-space configurations. For example, the PS applied only on the LO QCD part (Fig. 1.2) gives contribution to the next order of QCD. There are several approaches how to match PS with fixed order calculation, see [28].

The parton shower cannot give any approximate estimate on the total cross section from the missing higher order terms of the perturbation expansion. Therefore, the usual procedure of the simulation of $t\bar{t}$ events is to use MC generator at NLO QCD supplemented by PS, and normalize the distributions to the available prediction of higher order $t\bar{t}$ total cross section (nowadays NNLO+NNLL QCD).

Hadronization converts the partons into the observed hadrons. The process of hadronization is not explained by first-principles theory, and there are several phenomenological models for it. The two basic models are the string model used in the Pythia8 generator [40] and the cluster model used in the Herwig generator [41].

1.2.2 Particle vs Parton level

The predictions summarized in Sec. 1.2.1.1 are at *parton level*. It means that the prediction is done for the top quark as a stable particle. However, the top quark is not stable, and only the top quark's decay products after the hadronization process are detectable. Moreover, the parton level is defined using quarks and gluons, which may depend on the technical details of the MC generators. A measurement unfolded to parton level can be compared with a parton level

prediction. However, the parton level measurements have usually large physics modeling uncertainties.

Another approach to compare a measurement with a theoretical model is to construct prediction from final-state particles which is called as *particle level* prediction. The particle level should define observables in a theoretically safe and unambiguous way, and its definitions should be maximally independent of the technical details of MC generators. The particle level is defined from stable particles detectable by the detector. The particle level is usually defined in a *fiducial phase space* by certain event selection criteria at particle level objects. The fiducial phase space is a phase space, which can be used to compare theory with measurements or to compare two measurements. It is advantageous to define similar particle level selection criteria to the detector level selection criteria - this minimizes extrapolation uncertainties from the detector level to the particle level. For the above reasons, it makes more sense to compare certain measurement with certain theoretical model at particle level in contrast with the comparison at parton level.

In case of a certain particle level measurement, the corresponding particle level prediction from a theoretical model should be obtained from MC generators which contain PS, hadronization, and decays to stable particles. Then each event should be passed through the particle level definition of objects and selection criteria, which were used in the measurement. This step can be done using the Rivet project (Robust Independent Validation of Experiment and Theory) [42]. The Rivet project contains a convenient infrastructure to add the particle level definition used in the measurement. Then this definition can be used on a set of particles from the theoretical model, and the final comparison of the measurement and prediction at particle level is possible.

The exact definition of the particle level in certain measurement can be arbitrary. It usually aims to reduce the extrapolation uncertainties from detector level to particle level. An effort has started to unify the particle level definitions for $t\bar{t}$ measurements at the ATLAS and CMS experiments [43]. This would allow to compare the same particle level measurements between the two experiments or to combine the two measurements.

1.2.3 Overview of Measurements

The differential cross sections for $t\bar{t}$ production were measured at Tevatron at $\sqrt{s} = 1.96$ TeV and LHC colliders at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. The measurements with different colliding particles and at different energies are not comparable among each other. However, they all can be compared to the same approach of theory calculation which can give different predictions for different $\sqrt{s}$ values. And due to the QCD factorization theorem, the same approach of theory calculation can be used for both, pp and $p\bar{p}$, collisions, just the PDF sets are different.

Measurements were performed at both Tevatron experiments, CDF and D0. The D0 experiment measured the $p_{\mathrm{T}}^{\mathrm{top}}$, $m^{t\bar{t}}$, and y^{top} distributions at parton level [44]. All recorded data at the D0 experiment were used which corresponds to the integrated luminosity of 9.7 fb^{-1}. The measurements used events from the lepton+jets decay channel. The measured $p_{\mathrm{T}}^{\mathrm{top}}$ spectrum up to 400 GeV is shown

in Fig. 1.9. The CDF experiment measured the $m^{t\bar{t}}$ distribution at parton level using an integrated luminosity of 2.7 fb^{-1} [45]. All observed differential cross sections at Tevatron are consistent with the QCD predictions.

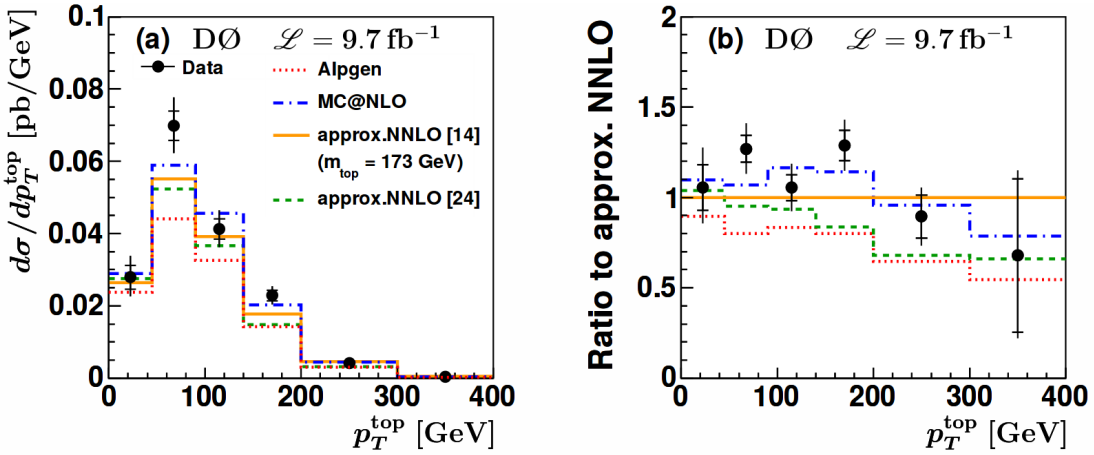


Figure 1.9: (a) Measured differential cross section as a function of $p_{\mathrm{T}}^{\mathrm{top}}$ for data compared to several QCD predictions. The inner error bars correspond to the statistical uncertainties and the outer error bars to the total uncertainties. (b) Ratio of data, ALPGEN (dashed line) and MC@NLO cross sections (dash-dotted line) to the QCD prediction at approximate NNLO [36]. MC simulations and QCD predictions use a top quark mass of 172.5 GeV unless indicated differently. Note that the correlated overall normalization uncertainty on the differential data points is about ±6.6%. [44]

Two LHC experiments, ATLAS and CMS, measured the $t\bar{t}$ differential cross sections. At $\sqrt{s} = 7$ TeV, both experiments measured only the normalized distributions at parton level for variables $p_{\mathrm{T}}^{\mathrm{top}}$, $m^{t\bar{t}}$, $p_{\mathrm{T}}^{t\bar{t}}$, and $y^{t\bar{t}}$ [46, 47]. The CMS experiment additionally measured the distribution of y^{top} and several kinematic distributions of leptons and b-jets. The Fig. 1.11 shows the measured normalized $p_{\mathrm{T}}^{\mathrm{top}}$ distribution at the ATLAS experiment compared to several predictions from MC generators. Statistical test showed that the measured distributions are in agreement with these predictions. The ATLAS experiment made a measurement at $\sqrt{s} = 7$ TeV at particle level [48] which was a novel approach to measure the $t\bar{t}$ differential cross sections leading to reduced model dependence. At $\sqrt{s} = 8$ TeV, similar set of measurements is provided from both experiments. The ATLAS experiment measured distributions of $p_{\mathrm{T}}^{\mathrm{top}}$, $p_{\mathrm{T}}^{t\bar{t}}$, $m^{t\bar{t}}$, and several other variables at particle and parton level [49]. The CMS experiment measured normalized distributions for kinematic variables connected with top quark, leptons, and b-jets [50].

All the above presented measurements used $t\bar{t}$ pair reconstruction techniques assuming resolved top quarks. This gives a limitation on the maximum measurable $p_{\mathrm{T}}^{\mathrm{top}}$. Using boosted techniques for the identification of hadronically decaying top quarks, one can explore the tail of the $p_{\mathrm{T}}^{\mathrm{top}}$ spectrum. This is performed at both LHC experiments at $\sqrt{s} = 8$ TeV. The ATLAS experiment published final measurement of the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution for boosted top quarks [5] which is presented in Sec. 5. The CMS experiment published preliminary results [51], see Fig. 1.10.

However, the measured $p_{\mathrm{T}}^{\mathrm{top}}$ distributions are consistent within uncertainties

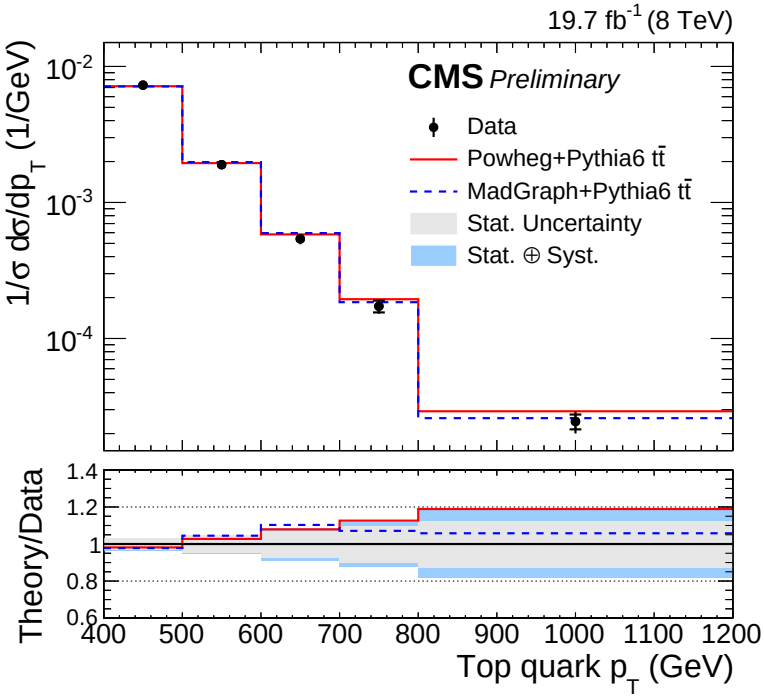


Figure 1.10: Normalized differential cross section of $p_{\mathrm{T}}^{\mathrm{top}}$ measured at the CMS experiment, including all systematic uncertainties. The experimental and theoretical uncertainties are shown separately. [51]

with NLO+PS predictions for both LHC experiments and both values of $\sqrt{s}$ in each measurement, there is a common feature. There is a slope for the ratio between central values of measurement and central values of NLO+PS prediction as a function of $p_{\mathrm{T}}^{\mathrm{top}}$ as in Fig. 1.11. This slope depends on the used MC generator or PDF set but it is positive almost in all cases. A possible explanation for this feature was given in year 2015 by evaluating the NNLO+QCD prediction as shown in Sec. 1.2.1.1. The Fig. 1.12 shows the comparison of one $p_{\mathrm{T}}^{\mathrm{top}}$ measurement from the CMS experiment with NLO QCD and NNLO QCD predictions. One can see that the NNLO QCD predicts this measurement much better than the NLO QCD prediction.

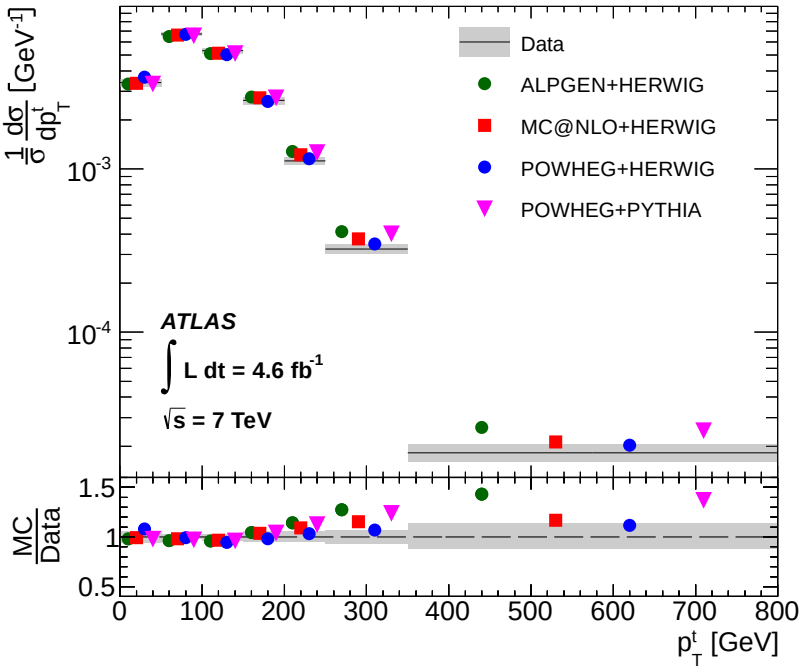


Figure 1.11: Normalized differential cross sections for the p_{T} of the hadronically decaying top quark measured at the ATLAS experiment. Generator predictions are shown as markers for ALPGEN+HERWIG (circles), MC@NLO+HERWIG (squares), POWHEG+HERWIG (triangles) and POWHEG+PYTHIA6 (inverted triangles). The markers are offset within each bin to allow for better visibility. The gray bands indicate the total uncertainty on the data in each bin. [47]

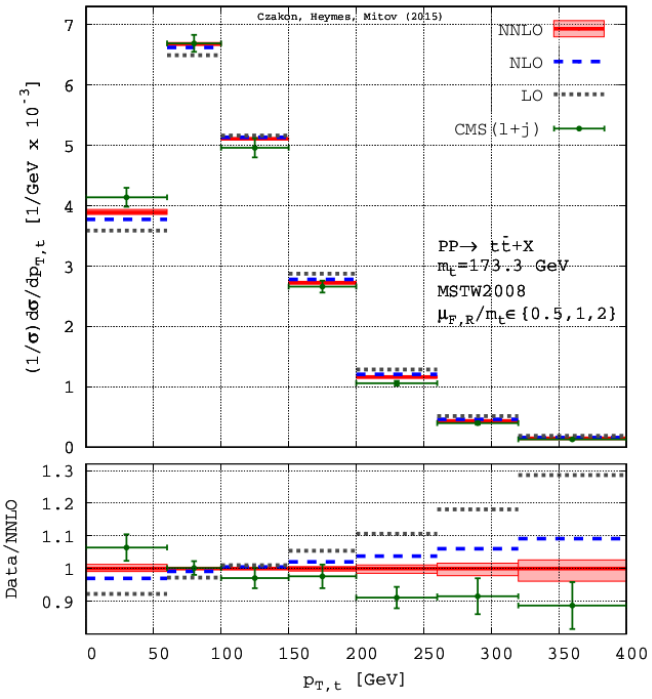


Figure 1.12: Top p_{T} distribution vs. CMS data [50]. All distributions are normalized. NNLO error band from scale variation only. [30]

2. The ATLAS Experiment

The ATLAS (A Toroidal LHC ApparatuS) experiment [52] is one of the two general purpose experiments at the LHC [53]. The LHC is the world's largest and highest-energy particle accelerator located ∼100 m underground at CERN. It is a circular collider with circumference of ∼27 km which collides protons or lead ions. The protons in colliding beams had energy of 3.5 TeV in year 2011, and energy of 4 TeV in year 2012 resulting in center-of-mass energies $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, respectively.

The ATLAS detector is about 44 meters long, more than 25 meters high, and weighs about 7000 tons. The detector has subdetectors arranged symmetrically in layers around the interaction point and the beam pipe, therefore, the detector covers almost the entire solid angle around the interaction point.

The individual subdetectors of the ATLAS detector are briefly described in this section. Further, the overview of pileup, physics object reconstruction, and Monte Carlo (MC) simulation are presented. The author of this thesis presented results about jet and missing transverse energy reconstruction at the ATLAS experiment at a conference talk on behalf of the ATLAS Collaboration. The content of the talk is published as a conference proceeding in [6], and it is summarized in Sec. 2.3.2 and 2.3.3. The jets and the missing transverse energy are important physics objects used in measurements of the $t\bar{t}$ pair production.

2.1 Subdetectors

The ATLAS detector is divided into three main parts: the inner detector, calorimeters, and muon spectrometer. These parts consist of complex detector systems which are shown in Fig. 2.1, and are outlined in the following.

The ATLAS Inner Detector [55] is the closest subdetector system to the interaction point. It consists of three different tracking technologies symmetrically distributed around the beam pipe: the Pixel Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The acceptance of the Inner Detector in pseudorapidity is $|\eta| < 2.5$ for any azimuthal angle ϕ. The Inner Detector detects the tracks of charged particles. Using these tracks, the positions of the pp interaction points (primary vertexes) can be estimated. Then for each pair of track and primary vertex, the impact parameter can be defined as a distance between the primary vertex and the point of closest approach of the track. The Inner Detector is placed in a solenoidal magnetic field of 2 T which bends the tracks of charged particles, and hence their momentum can be determined. The detector has been designed to provide a p_{T} resolution of $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} = 0.0005 \cdot p_{\mathrm{T}}[\mathrm{GeV}] \oplus 0.01$ and a transverse impact parameter resolution of 10 μm for high momentum particles in the central η region.

The next layers of ATLAS subdetectors are the calorimeters. The calorimeters measure the direction and the energy of charged and neutral particles. The calorimeters consist of two materials: absorber and active material. The absorbers are used to induce an electromagnetic or hadronic shower by the entering particle. The part of the deposited energy from these showers is measured in the active material. The electromagnetic shower is induced by a high-energetic elec-

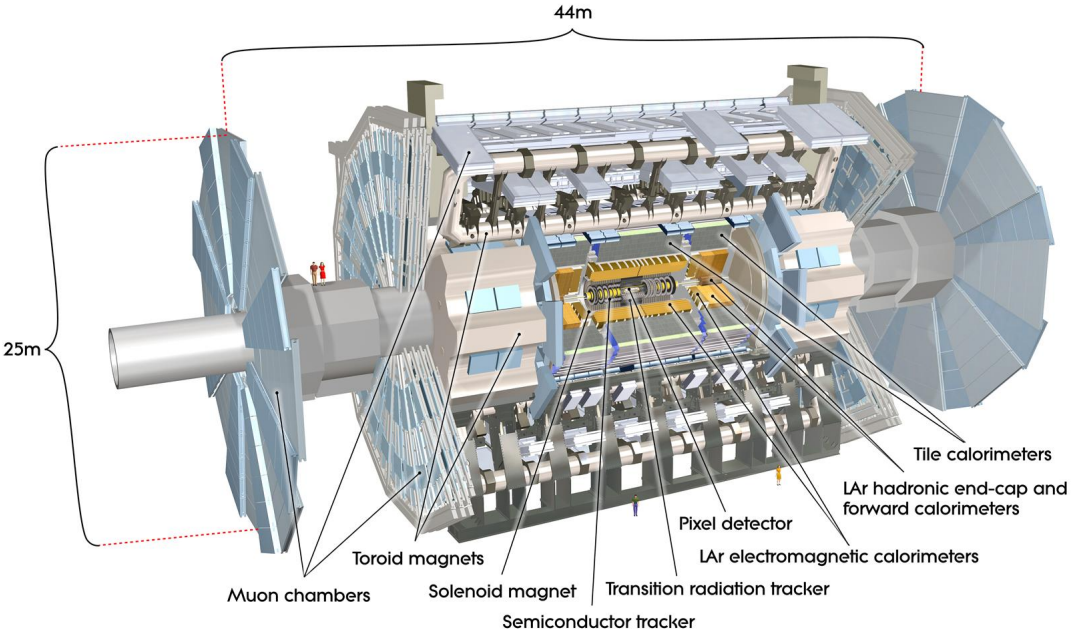


Figure 2.1: The ATLAS detector with indicated subdetectors and magnets. For scale demonstration, there are two people depicted just behind the first muon chamber on the left. [54]

tron or photon, while the hadronic shower is induced by a high-energetic hadron. The energy resolution of a calorimeter is generally parametrized by the equation

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E[\text{GeV}]}} \oplus \frac{b}{E[\text{GeV}]} \oplus c, \tag{2.1}$$

where a is the stochastic term, b is the noise term and c is the constant term. There are two main technologies used for the calorimeters in the ATLAS detector resulting in two calorimeters: the ATLAS Liquid Argon (LAr) Calorimeters [56] and the ATLAS Tile Calorimeter (TileCal) [57]. The LAr Calorimeters are composited of four calorimeters: electromagnetic barrel calorimeter (EMB), electromagnetic endcap calorimeter (EMEC), hadronic endcap calorimeter (HEC) and forward calorimeter (FCal). The basic properties of all the ATLAS calorimeters are summarized in Tab. 2.1. They detect the electromagnetic and hadronic showers up to pseudorapidity of 4.8.

The most distant subdetector system from the beam axis is the ATLAS Muon Spectrometer [58] that detects muons which are the only detectable particles passing through the calorimeters. The muon spectrometer measures the muon tracks in the magnetic field, from which the muon p_{T} is determined. The magnetic field is provided by toroidal magnets and is non-uniform. The designed p_{T} resolution $\sigma p_{\text{T}}/p_{\text{T}}$ is 4% for muons with $p_{\text{T}} \in [3, 100]$ GeV and it is increasing up to 10% for muons with $p_{\text{T}} = 1$ TeV.

Table 2.1: Calorimeters of the ATLAS detector with their properties: type (electromagnetic or hadronic), active material, absorber, $|\eta|$ coverage, and resolution terms. [56, 57]

	Type	Active material	Absorber	$\|\eta\|$ coverage	Resolution terms
EMB	elmag.	LAr	lead	$[0, 1.475]$	$a = 10\%$, $b = 30\%$ $c = 0.7\%$
EMEC	elmag.	LAr	lead	$[1.375, 3.2]$	$a = 10\%$, $b = 30\%$ $c = 0.7\%$
TileCal	hadr.	plastic scintillator	steel	$[0, 1.7]$	$a = 50\%$, $c = 3\%$
HEC	hadr.	LAr	copper	$[1.5, 3.2]$	$a = 50\%$, $c = 3\%$
FCal	elmag. and hadr.	LAr	copper and tungsten	$[3.2, 4.8]$	$a = 100\%$, $c = 10\%$

2.2 Pileup

Each hard scattering event (the event of interest) at LHC can be influenced by multiple uncorrelated *pp* interactions (pileup). There are two types of pileup: in-time pileup and out-of-time pileup. The in-time pileup are *pp* interactions occurring in the same bunch crossing while the out-of-time pileup refers to *pp* interactions from successive bunch crossings. Both types of pileup have impact on the reconstruction of the hard scattering event. The impact of the out-of-time pileup is caused by the fact that the bunch spacing can be smaller than the read-out response of many of the ATLAS subdetectors. The smaller the bunch spacing, the higher the effect of the out-of-time pileup. There are two measurable quantities which characterize the pileup in experiments: the average number of interactions per bunch crossing, μ, and the number of reconstructed primary vertexes, N_{PV}. The μ characterizes the magnitude of the out-of-time pileup on average. It depends on the instantaneous luminosity which is measured as described in [59]. The instantaneous luminosity is measured independently with a variety of detectors and using several different algorithms. One such detector is the LUCID detector, which measures the interaction rate in the pseudorapidity range $5.6 < |\eta| < 6.0$ ceaselessly during the data taking. This measurement is calibrated to the instantaneous luminosity using calibration parameters which were obtained from several van der Meer scans. The N_{PV} is a measure of the in-time pileup. The primary vertexes are reconstructed using the tracks from the Inner Detector as described in [60]. In case a certain bunch crossing contains hard scattering *pp* event, the number of pileup events $n_{\mathrm{PU}} = N_{\mathrm{PV}} - 1$ assuming perfect primary vertex reconstruction.

In year 2012, the LHC collided bunches of protons with bunch spacing 50 ns with instantaneous luminosities up to $\sim 8 \cdot 10^{33}$ cm^{-2} s^{-1} [61]. The Fig. 2.2 shows the distribution of μ in years 2011 and 2012. In year 2012, the LHC had in average 21 *pp* interactions per bunch crossing which had not been experienced in previous hadron colliders.

Pileup has negative effects in measurements involving jets. It randomly adds energy deposits to calorimeter cells, and hence it degrades the reconstruction of

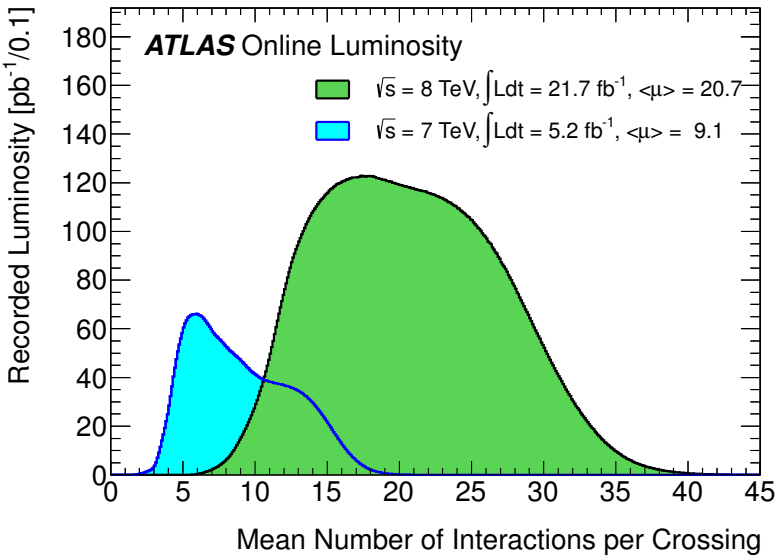


Figure 2.2: The distributions of μ in years 2011 and 2012. [61]

the jets from the hard scattering event. Each $t\bar{t}$ pair event consists of several jets, and thereby the $t\bar{t}$ pair events are highly affected by pileup. Mitigation of pileup effects is necessary to obtain better response and resolution for the jet 4-momentum using a pileup subtraction method. Moreover, the identification of boosted top quarks is based on jet substructure techniques, therefore pileup correction of the jet substructure is necessary as well. The Sec. 3 is devoted to pileup subtraction techniques for jets and jet substructure.

Another negative effect of pileup is the occurrence of additional jets (pileup jets) in the event, i.e. jets not originating from the hard scattering event, but mainly from pileup. These pileup jets can be mitigated using techniques described in Sec. 3 or using the methods developed at the ATLAS experiment as described in Sec. 2.3.2.3.

2.3 Physics Object Reconstruction

The particles from a pp collision leave typical characteristics in the complex sub-detector system of the ATLAS experiment, which are referred to as the physics objects, such as electron, photon, muon, hadronic tau and jet candidates. The neutrinos cannot be detected, although they can be partially reconstructed using the event observable called missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$. The $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed based on the momentum conservation in the transverse plane, and it represents the vector sum of p_{T} of all non-detectable particles such as neutrinos in each event.

A summary of the reconstruction of physics objects relevant for this thesis is presented in the following.

2.3.1 Electrons and Muons

An electron candidate in the pseudorapidity region $|\eta| < 2.5$ is reconstructed as a track in the Inner Detector associated with energy deposits in the electromagnetic calorimeters. The electron candidate energy is determined from the measured energy deposits in the electromagnetic calorimeters. It is corrected for the effects of lateral and longitudinal leakage and for the energy loss in the material in front of the electromagnetic calorimeters. The electron candidate spatial coordinates, η and ϕ, are taken from the spatial coordinates of the matched track. The sign of the electron charge can be determined from the curvature of the matched track. Each electron candidate must satisfy identification criteria based on the shower shape in the electromagnetic calorimeters, on track quality, and on the transition radiation observed in the TRT detector. A detailed description of the electron reconstruction, calibration and performance can be found in [62].

To find the muon candidates, the tracks are reconstructed independently in the Inner Detector and in the Muon Spectrometer. The muon candidates are found by matching tracks from the Muon Spectrometer with the tracks in the Inner Detector. The muon candidate momentum is determined through a global fit of hits from both subdetectors, and this determination takes into account the energy loss in the calorimeters. A detailed description of the muon reconstruction, calibration and performance can be found in [63].

2.3.2 Jets

Jets are key objects for many ATLAS measurements and searches. The jet reconstruction, the jet calibration, the jet energy resolution, and the suppression of pileup jets is overviewed in this section.

2.3.2.1 Reconstruction and Calibration of Jets

The highly segmented calorimeters of the ATLAS detector enable to reconstruct jets with high precision. The jet reconstruction consist of several steps which are explained in this section: topo-cluster finding, jet clustering, pileup correction, jet origin correction, jet calibration, and residual in situ calibration.

The calorimeter cells are grouped to 3-dimensional clusters of topologically connected cells called topo-clusters. The topo-cluster finding is optimized to noise and pileup suppression [64]. The first step is the identification of seeds which are cells with energy deposits $E > 4\sigma$ where σ is the noise defined as a sum in quadrature of electronic and pileup noise. The second step is the iterative adjunction of neighboring cells with $E > 2\sigma$ to the seeds. In the third step, an extra layer of cells with $E > 0$ on the perimeter of the clustered cells are added. Splitting algorithm separates the resulting topo-clusters based on local energy maxima. There are two options to calibrate topo-clusters: calibration to the electromagnetic scale (EM topo-clusters) and local calibration weighting [65] (LCW topo-clusters). In both cases, the invariant mass of the topo-clusters is set to zero. The EM topo-clusters are calibrated to the response from electrons while the LCW topo-clusters are classified as electromagnetic or hadronic and then a weighting scheme corrects for the different electron-to-pion response in

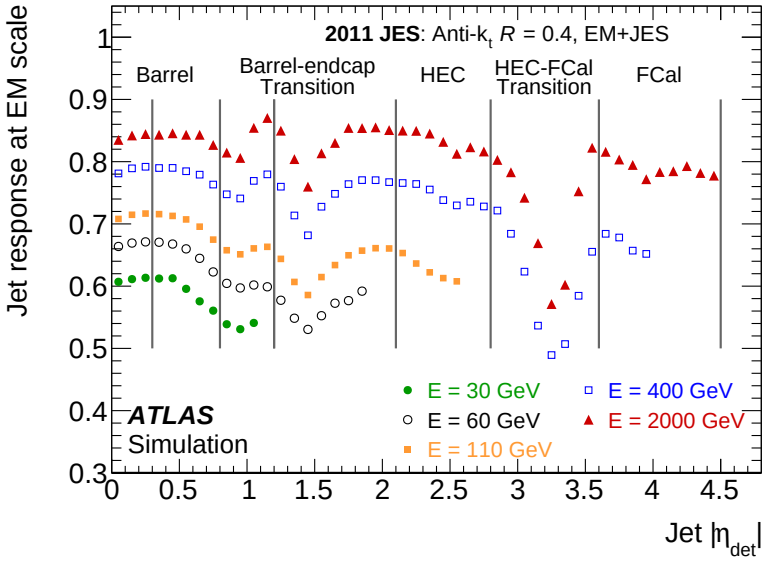


Figure 2.3: Average response of simulated EM jets as a function of jet pseudorapidity for several truth-jet energies. Also indicated are the different calorimeter regions. [67]

the calorimeters. Dead material correction and out-of-cluster correction is used for the LCW topo-clusters.

The App. B summarizes the jet finding algorithms used in this thesis. The standard jet finding algorithm at ATLAS experiment is the anti-k_t clustering algorithm with distance parameters $R = 0.4$ and $R = 0.6$. The input to the jet finding algorithm are EM and LCW topo-clusters resulting in EM and LCW jets, respectively. Additionally, clustering algorithms with large distance parameters (large-R) have been commissioned and are used, such as anti-k_t algorithm with $R = 1.0$ and C/A algorithm with $R = 1.2$.

The topo-cluster finding suppresses the effect of pileup but applying a further pileup correction on the jets is necessary. The jet area-based correction (see Sec. 3.2.1) was used followed by residual offset correction [66].

The jet origin correction [67] makes the jet pointing back to the primary event vertex instead of the nominal center of the ATLAS detector.

The jet energy and pseudorapidity are calibrated using the relation between reconstructed and truth-jets in MC simulated QCD events [67]. The jet energy calibration (Jet Energy Scale) is a multiplication by the inverse of average jet energy response. The Fig. 2.3 shows the dependence of the jet energy response on pseudorapidity. After applying the jet energy scale (JES), the EM jets and LCW jets are called EM+JES and LCW+JES jets, respectively. The jet pseudorapidity calibration corrects for a bias due to poorly instrumented regions of the calorimeter. Average difference between pseudorapidities of reconstructed and truth-jets in MC is added as a correction factor to the jet pseudorapidity.

Differences between data and MC simulation lead to miscalibration of jet energy which is removed by a residual in situ calibration applied to the data only.

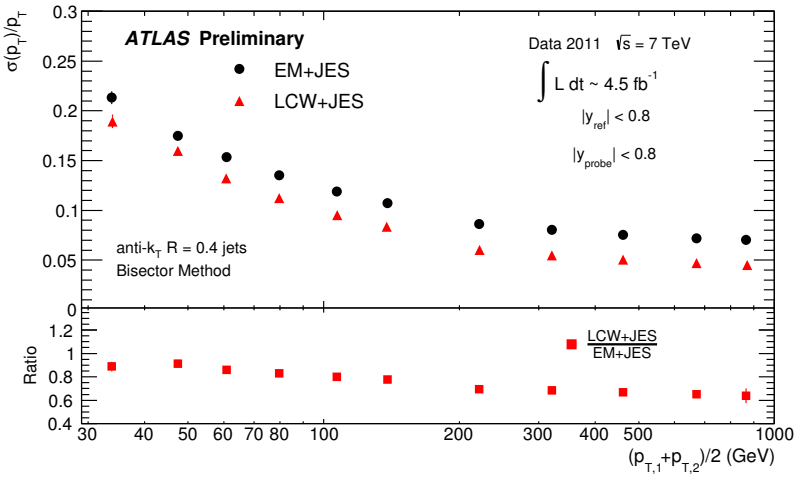


Figure 2.4: Fractional jet energy resolution for anti-$k_t R = 0.4$ jets as a function of the average jet transverse momenta measured with the bisector in-situ technique using the EM+JES calibration (circles) and the LCW+JES calibration (triangles) with 2011 data. The bottom plot shows the ratio as a function of the average jet transverse momenta. [69]

It corrects the jet p_{T} by multiplying by the response ratio of MC to data

$$\frac{\mathrm{Response}_{\mathrm{MC}}}{\mathrm{Response}_{\mathrm{Data}}} = \frac{\left\langle p_{\mathrm{T}}^{\mathrm{jet}} / p_{\mathrm{T}}^{\mathrm{ref}} \right\rangle_{\mathrm{MC}}}{\left\langle p_{\mathrm{T}}^{\mathrm{jet}} / p_{\mathrm{T}}^{\mathrm{ref}} \right\rangle_{\mathrm{Data}}}, \tag{2.2}$$

where the response is obtained from transverse momentum balance between jet and a reference object. To cover large kinematic phase space, different reference objects are used in the following methods: dijets η-intercalibration, γ+jet balance, Z+jet balance and multijet balance.

2.3.2.2 Jet Energy Resolution

The jet energy resolution (JER) is measured in data with two in situ techniques: dijet balance and bisector method, [68]. The Fig. 2.4 shows the JER obtained with the bisector method for both type of calibration of jets. The LCW+JES jets exhibit better energy resolution than EM+JES jets. The JER decreases with increasing jet p_{T}, which is expected from Eq. 2.1.

2.3.2.3 Suppression of pileup jets

To suppress pileup jets against jets from the hard scattering event, information from tracks matched to each jet is used. Several methods were commissioned at the ATLAS experiments named as Jet Vertex Fraction (JVF), corrJVF, R_{pT}, and Jet Vertex Tagger (JVT), and they are explained in detail in [70]. In each method, a discriminating variable is constructed to distinguish between pileup jets and hard scattering jets. The JVF method is used for most of analyses using 2011 and 2012 data, including the measurement presented in Sec. 5. The JVF variable is defined as the fraction of momenta of tracks matched to the jet which

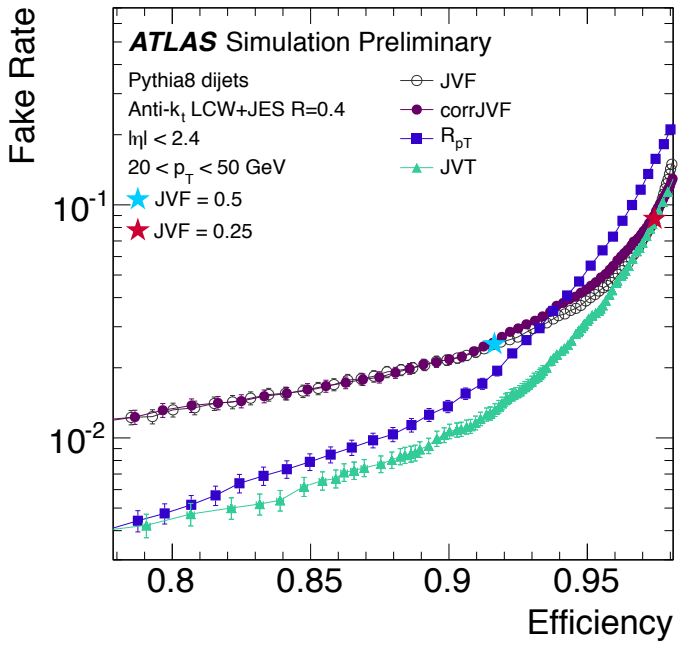


Figure 2.5: Fake rate from pileup jets versus hard scattering jets efficiency curves for JVF, corrJVF, R_{pT}, and JVT in MC simulation. The widely used JVF working points with cut values 0.25 and 0.5 are indicated with red and blue stars. [70]

are associated with the hard scattering vertex. The R_{pT} and corrJVF are other two variables with discriminating power, and they are used to construct the JVT variable which is a multivariate combination of these two variables. The Fig. 2.5 shows the comparison of the dependence of the rejection rate for pileup jets on the efficiency for hard scattering jets for the four pileup jet suppression methods. This comparison is prepared for MC events for jets with $p_{\mathrm{T}} \in [20, 50]$ GeV and $|\eta| < 2.4$. The JVT method has much better performance than the JVF method. For example, the widely used JVF working point, which selects jets with JVF > 0.5 has hard scattering jet efficiency of $\sim$92% and fake rate of $\sim$2.5%, while the JVT working point with the same hard scattering jet efficiency has much lower fake rate ($\sim$1%). Another advantage of the JVT method with respect to the JVF method is that its performance has much lower dependence on the pileup conditions as discussed in [70]. These two advantages are the reason for switching to the JVT method in analyses using 2015 data.

2.3.3 Missing Transverse Energy

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is an important signature for many physics processes such as the $t\bar{t}$ production in the single lepton channel. It is an event quantity calculated based on momentum conservation in the transverse plane [71]:

$$\begin{aligned} E_{\mathrm{T}}^{\mathrm{miss}} &= \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2}, \\ E_{x(y)}^{\mathrm{miss}} &= -\left(E_{x(y)}^{\mathrm{jets}} + E_{x(y)}^{e} + E_{x(y)}^{\gamma} + E_{x(y)}^{\tau} + E_{x(y)}^{\mu} + E_{x(y)}^{\mathrm{ST}}\right) \end{aligned} \tag{2.3}$$

where $E^{\text{jets}}_{x(y)}$, $E^{e}_{x(y)}$, $E^{\gamma}_{x(y)}$, $E^{\tau}_{x(y)}$, and $E^{\mu}_{x(y)}$ are the sum of $x(y)$-component of the momenta of all jets, electrons, photons, taus and muons in the event, respectively. All objects are corrected for the pileup and calibrated. The anti-$k_t R = 0.4$ jets calibrated with LCW+JES scheme with $p_{\text{T}} > 20$ GeV are used to calculate $E^{\text{jets}}_{x(y)}$. Suppression of pileup jets is done by rejecting jets fulfilling conditions JVF $= 0$, $p_{\text{T}} < 50$ GeV and $|\eta| < 2.4$. The $E^{\text{miss}}_{\text{T}}$ Soft Term, $E^{\text{ST}}_{x(y)}$, is defined as the sum of $x(y)$-component of the momenta of all topo-clusters and tracks not associated to the above physics objects with double counting avoided.

The pileup has significant effect on the performance of $E^{\text{miss}}_{\text{T}}$ reconstruction. There are several pileup correction methods for $E^{\text{miss}}_{\text{T}}$, see [72]. All methods correct the $E^{\text{miss}}_{\text{T}}$ Soft Term. The first method called Soft-Term Vertex-Fraction (STVF) corrects the $E^{\text{miss}}_{\text{T}}$ Soft Term by a multiplication factor constructed from all tracks in the event. This factor is the fraction of momenta of tracks matched to the primary vertex from hard scattering event. Further possibilities are to use jet-area-based methods. The basic idea in these methods is that the soft term constituents from calorimeters are clustered to jets which are corrected with jet-area-based pileup correction method. Optionally, JVF based selection can be applied.

The Fig. 2.6 shows the pileup dependence of the reconstructed average value of $E^{\text{miss}}_{\text{T}}$ for several pileup correction methods from data events. The event selection consists of criteria to select events containing two opposite sing muons from the Z boson. These events do not contain non-interacting particles, such as neutrinos, i.e. the $E^{\text{miss}}_{\text{T}}$ should be zero in case of perfect detection of all particles. The average value of $E^{\text{miss}}_{\text{T}}$ is biased by the incomplete capture of the hadronic recoil balancing the p_{T} of the Z boson. This leads to average value of $E^{\text{miss}}_{\text{T}}$ of $\sim$10 GeV for no additional pileup interactions. With increasing pileup, the average value of $E^{\text{miss}}_{\text{T}}$ increases due to increasing fluctuations in the calorimeter response. All the applied pileup correction methods enhance the reconstructed $E^{\text{miss}}_{\text{T}}$, and the STVF method has the best performance. The STVF method is used also for the $E^{\text{miss}}_{\text{T}}$ pileup correction in the analysis presented in Sec. 5.

2.4 Monte Carlo Simulation

The real pp collision events detected by the ATLAS detector are referred to as the *data* events in this thesis. These events can be simulated using MC simulations, which are extremely important for any measurement at the ATLAS experiment. The measured objects are influenced by detector response, and the corresponding observables are referred to as measured at the *detector level*. The detector level cannot be compared to any theory prediction or measurements from other experiments. On the other hand, the MC simulation of events provides three levels of information: parton, particle and detector level. The Fig. 2.7 shows the sketch of a pp collision with a jet at these three levels. The parton and particle level are discussed in Sec. 1.2.2. The detector level is defined from objects reconstructed in the detector. The exact definitions of these levels depend on the particular analysis. The availability of these three levels from MC simulations allows to extrapolate the measured observable in data from detector level to particle level or to parton level.

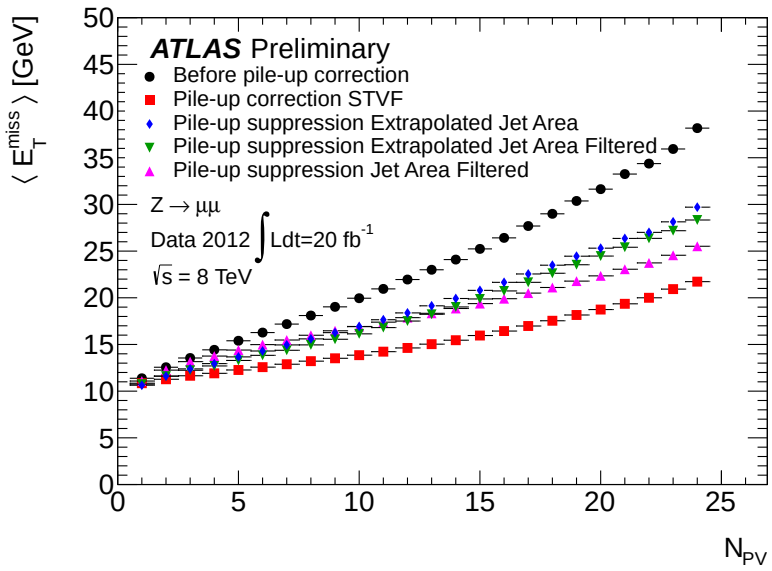


Figure 2.6: The reconstructed average $E_{\mathrm{T}}^{\mathrm{miss}}$ as a function of N_{PV} for the inclusive $Z \to \mu^+\mu^-$ data sample for several pileup correction methods. [72]

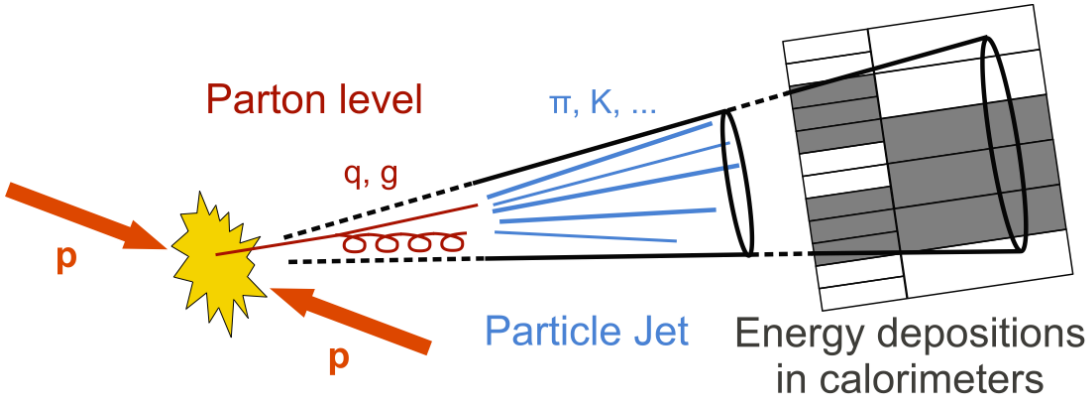


Figure 2.7: Sketch of a *pp* collision and resulting jet at parton level, particle level and detector level. [73]

The MC simulation of events at the ATLAS experiment consists of two main steps: the simulation of certain physics process and the simulation of the response of various detectors and triggers. The simulation of certain physics process involves simulation of the matrix element, parton shower, hadronization and decays of unstable particles. This give access to the parton and particle levels. Then the particles from the particle level are passed through the full GEANT4 [74] detector simulation. A detailed description of the ATLAS simulation chain can be found in [75]. After the MC events are simulated, the same trigger, event, quality, track, and physics object selection criteria used in the data analyses are applied to the MC simulation.

Each event can have a certain event weight. This event weight must be used when analyzing data, e.g. when filling histograms. In contrary to data events, which always have unit event weight, the events from a MC simulation can have non-unit event weight. There are two main reasons for using non-unit event weights for some MC simulations: to achieve better statistics in phase space regions with low cross section or to avoid double counting when matching the NLO QCD calculation with parton shower MC simulation. For example, the generation of events with MC@NLO generator uses event weights of -1 and 1 as explained in [76]. Further, additional multiplicative factors to the event weights can be applied to correct the non-perfect detector simulation as explained in Sec. 2.4.1.

2.4.1 Corrections of the Monte Carlo Simulation

The detector and physics simulation are not perfect. This leads to differences between the performance in MC simulation and data. These differences are corrected in MC using inputs from various calibration measurements. There are three main types of MC corrections: efficiency correction, 4-momentum correction, and resolution correction. The efficiency correction is done using an event scale factor which is a multiplicative factor to the total event weight. By applying such correction factor to the total event weight for each event, the reconstruction, identification and trigger efficiencies of various objects can be corrected. With an event scale factor, also the b-tagging performance can be corrected as it is in detail described in Sec. 4.3. The second type of the corrections is the 4-momentum correction. It corrects the 4-momenta of various physics objects, such as jets, muons, electrons. Also the $E_{\mathrm{T}}^{\mathrm{miss}}$ can be corrected. The third type of corrections is the resolution correction. It is applied to obtain the same p_{T} resolution of physics objects in MC as in data. Usually, the resolution in data is worse than or identical with the resolution in MC simulation. For this reason, a random smearing of the p_{T} of physics objects in MC is applied. Similarly, the angular resolution can be corrected in MC. After all the corrections, the performance in MC should be more similar to the performance in data.

Each MC correction has its associated uncertainties. According to the classification of the corrections, the corresponding uncertainties are referred to as efficiency, 4-momentum, and resolution uncertainties. These uncertainties need to be propagated to the final physics results, since they are derived using inputs from MC simulations (e.g. efficiency of selection criteria). The way how the MC correction uncertainties are propagated depends on the particular analysis.

2.4.2 Pileup Simulation

Pileup is simulated by overlaying additional soft pp onto the hard scattering event. A soft pp interaction is generated with Pythia 8.160 [77, 40] using MSTW2008LO PDF set [78] with ATLAS A2 tune [79]. The number of these soft pp interactions was generated randomly according to the expected μ distribution in future data. After the μ distribution is measured from all data recorded in year 2012, the simulated events are reweighted such that the MC distribution of μ agrees with the data. The pileup simulation takes into account the contribution from out-of-time pileup using time shifts between bunches to model the bunch structure of the LHC beam.

3. Pileup Subtraction for Jets

The pileup has large effect on the jet reconstruction as it is discussed in Sec. 2.2, therefore it is very important to mitigate the pileup effects. Only pp collisions are considered in the next, although most of the methods can be used or extended for underlying event subtraction in heavy-ion collisions, in which the jet reconstruction is challenging due to the sizable underlying event [80, 81].

There are several approaches of how to mitigate the pileup effects, and they are overviewed in this section. The main focus is given on the *pileup p_{T} density methods*, from which the shape-expansion pileup correction technique was tested by the author of this thesis on data from the ATLAS experiment. This performance work is published in the ATLAS note [1] and the author of this thesis was a co-editor of this publication. Further, a novel approach called *Constituent Subtraction* is described in this section. The author of this thesis developed this approach and is the main author of the publication [2]. Most of the figures in this section (Fig. 3.4–3.9 and Fig. 3.11–3.14) are results of the work of the author of this thesis.

3.1 Overview of Pileup Subtraction Methods

The currently available pileup subtraction methods can be categorized as follows:

- *Offset correction* - simple correction of the measured jet p_{T}. The p_{T} offset used for the subtraction is proportional to the number of observed pileup events. This method was used for analyzing $\sqrt{s} = 7$ TeV data at the ATLAS experiment [82].

- *Pileup p_{T} density methods* (also called *area-based subtraction*) - there are several methods for the correction of jet 4-momentum or jet shapes. The *area 4-vector method* [83] corrects the jet 4-momentum, and it is extended to account for hadron masses in [84]. The *shape-expansion method* [84] provides general approach to correct jet shapes. An extension of these two methods is the *Constituent Subtraction* method. Detailed description of these methods can be found in the Sec. 3.2–3.4.

- Methods using tracking information - the tracking information can identify charged particles originating from pileup interactions. These charged pileup particles can be further used to subtract pileup at the calorimeter level. One of these methods is the *charged hadron subtraction* (CHS) [85] used at the CMS experiment. The CHS method takes an advantage of the Particle-flow event reconstruction at CMS [86]. It removes the identified charged pileup particles, which corresponds roughly to the half of the pileup p_{T} offset to jets in the tracker-covered region. Another example is the *jet cleansing* method [87]. Jet cleansing attempts to correct the 4-momenta of constituents in each subjet within a jet. This correction is based on the energy levels of charged and neutral particles in each subjet.

- *Pileup per particle identification* (PUPPI) [88] - an event-level subtraction procedure which corrects the particle 4-momenta based on the collinear versus soft diffuse structure in the neighborhood of the particle.

- *Soft Killer* method [89] - an event-level subtraction procedure which removes the softest particles in an event, up to a p_{T} threshold that is determined dynamically on an event-by-event basis.

- *Grooming* techniques - methods which enhance the identification of boosted objects such as top quarks, W, Z or Higgs bosons using large-R jets. Any grooming technique yields to different jet definition than that which was used to find the jet prior to grooming. The aim of such a new jet definition is to be less susceptible to pileup, multiple parton interactions (MPI), and initial-state radiation (ISR) while incorporating as much as possible particles from the hard scatter event. The particles from pileup, MPI, and ISR are often much softer than the particles from hard-scatter event which is used in several grooming techniques. Here is a list of example grooming techniques with brief description:

  - *trimming* [90] - the jet constituents are clustered with the k_t clustering algorithm with distance parameter $R = R_{\mathrm{sub}}$. Any subjets with $p_{\mathrm{T}} < p_{\mathrm{T}}^{\mathrm{jet}} \cdot f_{\mathrm{cut}}$ are removed, where $p_{\mathrm{T}}^{\mathrm{jet}}$ is the p_{T} of the initial jet, see Fig. 3.1 for an illustration. This algorithm has two free parameters: R_{sub} and f_{cut}.
  - *pruning* [91] - the jet constituents are clustered with the k_t or Cambridge/Aachen (C/A) clustering algorithm with large enough distance parameter to cluster all jet constituents. At each recombination step, the two constituents are recombined only if $p_{\mathrm{T}}^2/p_{\mathrm{T}}^{1+2} > z_{\mathrm{cut}}$ or $\Delta R_{1,2} < 2R_{\mathrm{cut}}m^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{jet}}$ where $p_{\mathrm{T}}^2 < p_{\mathrm{T}}^1$ are the p_{T} of the two constituents, p_{T}^{1+2} is the p_{T} of the recombined constituent, and $\Delta R_{1,2}$ is the ΔR distance between the two constituents. This algorithm has two free parameters: z_{cut} and R_{cut}.
  - *splitting and filtering* [92] - grooming technique optimized to identify boosted Higgs bosons decaying to two b-quarks using large-R C/A jets.

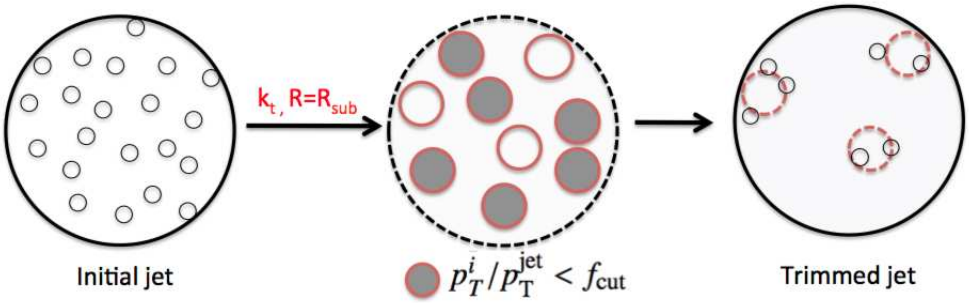


Figure 3.1: A cartoon depicting the jet trimming procedure. [25]

The above pileup subtraction methods are mainly meant to correct individual jets after jet clustering. However, some of them (Soft Killer, PUPPI) corrects

the whole event prior the jet clustering. In this way, the jet clustering is not biased by the presence of pileup particles, which is called as *back-reaction*, and it is discussed in [93]. Moreover, the whole event correction can also provide corrected event variables, such as $E_{\mathrm{T}}^{\mathrm{miss}}$.

3.2 Pileup p_T Density Methods

The concept of pileup p_{T} density was first introduced in [83]. The pileup p_{T} density, $\rho(y, \phi)$, is defined as the amount of p_{T} originating from pileup per unit area in the rapidity-azimuth $(y - \phi)$ space for a certain event. It can depend on the position (y, ϕ). This continuous approximation of pileup particles (which are point-like) can be done due to their large density in the $(y - \phi)$ space. Similarly to the pileup p_{T} density, the pileup mass density, $\rho_m(y, \phi)$, can be introduced as the density of variable m_δ

$$m_\delta = \sqrt{m^2 + p_{\mathrm{T}}^2} - p_{\mathrm{T}} \tag{3.1}$$

in the $y-\phi$ space, as done in [84]. Then the expected pileup deposition in certain direction (y, ϕ) is expressed by the 4-momentum

$$P_{\mathrm{pileup}} = [\rho \cos\phi, \rho \sin\phi, (\rho + \rho_m) \sinh y, (\rho + \rho_m) \cosh y] \cdot \Delta y \Delta\phi, \tag{3.2}$$

where $\Delta y \Delta\phi$ is a small region in the $y-\phi$ space. There are approximations to obtain the functions $\rho(y, \phi)$ and $\rho_m(y, \phi)$ in events containing both pileup and hard scattering contributions. These approximations are based on the assumption that the particles from the hard scattering event are collimated in certain directions while the particles from pileup are randomly distributed in the $y - \phi$ space, and they are much softer than the particles from the hard scattering event.

An example for the estimation of pileup densities is described in the following. This example neglects the dependence on y and ϕ. The event is divided by rectangular grid in $y - \phi$ space into *patches*. The patch area, A_{patch}, is defined as the area of rectangles in the $y - \phi$ space. The p_{T} and m_δ of each patch, p_{Tpatch} and $m_{\delta\mathrm{patch}}$, is determined by summing over all particles within that patch:

$$p_{\mathrm{Tpatch}} = \sum_{i \in \mathrm{patch}} p_{\mathrm{T}i}, \qquad m_{\delta\mathrm{patch}} = \sum_{i \in \mathrm{patch}} \left(\sqrt{m_i^2 + p_{\mathrm{T}i}^2} - p_{\mathrm{T}i} \right), \tag{3.3}$$

where $p_{\mathrm{T}i}$ and m_i are the p_{T} and mass of particle i. The final pileup p_{T} and mass densities are estimated as the median of p_{T} and m_δ densities from the set of all patches

$$\rho = \mathrm{median} \left\{ \frac{p_{\mathrm{Tpatch}}}{A_{\mathrm{patch}}} \right\}, \qquad \rho_m = \mathrm{median} \left\{ \frac{m_{\delta\mathrm{patch}}}{A_{\mathrm{patch}}} \right\}. \tag{3.4}$$

Several modifications exist to estimate the pileup densities as described in [94], e.g. apply selection criteria on patches or use position-dependent ρ and ρ_m estimations. An alternative method to obtain the patches can be used by defining the patches as jets reconstructed using the k_t algorithm with certain distance parameter, e.g. $R = 0.5$. In this case, the patch area is equal to the *jet area*. The concept of the jet area is discussed in detail in [93]. There are

several possibilities how to define the jet area. One of these definitions is called the *scalar active area* definition, which is explained in detail in Sec. 3.2.1. Using jets as patches brings ambiguity in the definition of the patch area in contrary to the example above, in which the patch area is clearly defined as the area of rectangle in the $y - \phi$ space.

The Fig. 3.2 serves to illustrate the extraction of pileup p_{T} density. It shows the densities $p_{\mathrm{Tpatch}}/A_{\mathrm{patch}}$ for an example pp dijet event overlaid with 22 pileup events. The densities $p_{\mathrm{Tpatch}}/A_{\mathrm{patch}}$ are clustered around the ρ determined by Eq. 3.4, except few patches containing hard jets. The dependence on rapidity is weak. In this example, the estimated constant ρ is a reasonable approximation of the pileup contamination.

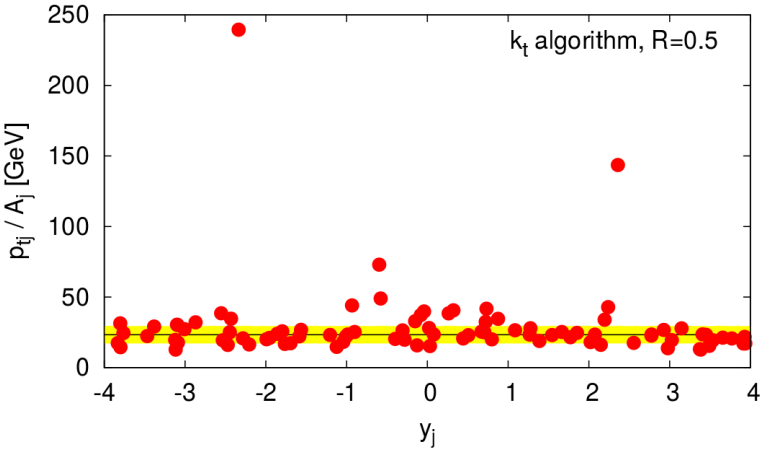


Figure 3.2: The density $p_{\mathrm{Tpatch}}/A_{\mathrm{patch}}$ as a function of the patch rapidity for an pp dijet event overlaid with 22 pileup events. The patches are obtained as jets clustered with k_t $R = 0.5$ algorithm. The black line is constant function of ρ determined by Eq. 3.4. [83]

3.2.1 Area 4-vector Method

The area 4-vector method uses the active area definition for the jet area, [93], which is explained in the following. Before the jet algorithm is used, massless particles with very low momentum (ghosts) are incorporated into the event such that they uniformly cover the $y - \phi$ space with high density. An example uniform distribution of ghosts in rapidity region $[-y^g_{\mathrm{max}}, y^g_{\mathrm{max}}]$ can be obtained by constructing 4-momenta

$$P^g = [p^g_{\mathrm{T}} \cos\phi^g, p^g_{\mathrm{T}} \sin\phi^g, p^g_{\mathrm{T}} \sinh y^g, p^g_{\mathrm{T}} \cosh y^g], \tag{3.5}$$

where the p_{T} of the ghost, p^g_{T}, is a free parameter and it is set to very low positive value, and the kinematic variables y^g and ϕ^g are acquiring values

$$y^g \in \{-y^g_{\mathrm{max}}, -y^g_{\mathrm{max}} + \delta y, -y^g_{\mathrm{max}} + 2\cdot\delta y, \ldots, y^g_{\mathrm{max}}\}, \quad \text{where} \quad \delta y = 2y^g_{\mathrm{max}}/n \tag{3.6}$$

$$\phi^g \in \{0, \Delta\phi, 2\cdot\Delta\phi, \ldots, (m-1)\cdot\Delta\phi\}, \quad \text{where} \quad \Delta\phi = 2\pi/m \tag{3.7}$$

where n and m are large integer numbers defining the density of ghosts. Each ghost covers a certain fixed area, A^g, in the $y - \phi$ space. In the above example

of ghosts distribution, the total number of ghosts is $n \cdot m$, and the ghost area for each ghost is $A^g = 4\pi \cdot y^g_{\max}/n/m$.

After adding ghosts into the event, the jet clustering algorithm is used on all real particles and ghosts in that event. The final jets contain ghosts with real particles. Only infrared safe jet algorithms can be used since the infrared safety ensures that the addition of ghosts does not change the real particle composition of the final jets. The ghosts clustered into a certain jet J are used to evaluate the area 4-vector for that jet:

$$[A_x, A_y, A_z, A_E] = \frac{A^g}{p^g_{\mathrm{T}}} \sum_{\substack{i \in \text{ghosts} \\ \text{in J}}} P^g_i \tag{3.8}$$

Assuming that the pileup particles are soft and distributed uniformly in the $y - \phi$ space similarly as ghosts, one can think of ghosts as particles which mimic the pileup particles during the running of the jet clustering algorithm. Using this assumption, the area 4-vector method corrects the jet 4-momentum $[p^x, p^y, p^z, E]$ in the following way:

$$P^{\mathrm{corr}} = [p_x - \rho A_x, p_y - \rho A_y, p_z - (\rho + \rho_m)A_z, E - (\rho + \rho_m)A_E]. \tag{3.9}$$

where the pileup densities are taken at point (y, ϕ) corresponding to the original jet.

Alternatively, one can correct only the jet p_{T} using the scalar active area, A^{jet}:

$$p^{\mathrm{corr}}_T = p_{\mathrm{T}} - \rho \cdot A^{\mathrm{jet}} \tag{3.10}$$

The scalar active area can be defined as the transverse component of the area 4-vector defined in Eq. 3.8:

$$A^{\mathrm{jet}} \equiv \sqrt{A^2_x + A^2_y} \tag{3.11}$$

There is an alternative definition for the scalar jet area:

$$A^{\mathrm{jet}} \equiv N \cdot A^g \tag{3.12}$$

where N is the number of ghosts clustered into the jet. The two definitions of the scalar active area give approximately the same result, but they are not identical since vector sum of ghosts transverse components is used in Eq. 3.8 leading to Eq. 3.11, while the scalar sum of transverse components in Eq. 3.8 would lead to Eq. 3.12. The jet area in Eq. 3.10 can be also estimated differently using the *Voronoi area definition* [93], which does not require adding ghosts.

The area 4-vector method in Eq. 3.9 leads to similarly corrected jet p_{T} as the correction in Eq. 3.10 using various jet scalar area definitions. It can be shown that the corrected p_{T} using Eq. 3.9 is greater than or equal to the corrected p_{T} using Eq. 3.10 with definition Eq. 3.11 which is greater than or equal to the corrected p_{T} using Eq. 3.10 with definition Eq. 3.12.

The p_{T} correction using Eq. 3.10 is actively used at the ATLAS experiment for $\sqrt{s} = 8$ TeV data before the calibration process as is explained in Sec. 2.3.2. The scalar jet area definition in Eq. 3.12 is used and anti-k_t $R = 0.4$ jets are used as patches in Eq. 3.4. The performance of this correction is shown in Fig. 3.3 as the

root mean square (RMS) of the $\left(p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{truth}}\right)$ distribution, where the $p_{\mathrm{T}}^{\mathrm{reco}}$ is the p_{T} of detector level jet angularly matched to a particle level jet with $p_{\mathrm{T}} = p_{\mathrm{T}}^{\mathrm{truth}}$. Dijet MC simulation is used, and particle level jets with $p_{\mathrm{T}}^{\mathrm{truth}} \in [20, 30]$ GeV are selected. The RMS is shown without any pileup correction, with the offset correction [82] used for $\sqrt{s} = 7$ TeV data, and with the p_{T} correction using Eq. 3.10 which is referred to as the area-based correction. One can see that the RMS is ∼5 GeV for low μ values. This uncertainty is mainly given by the stochastic term of the calorimeters resolution. With increasing pileup conditions, the RMS increases significantly up to ∼10 GeV for $\mu = 35$. The offset correction results in better jet p_{T} resolution, and the area-based correction has even better performance than the simple offset correction (it reduces the RMS by ∼20%). The main reason of this observation is the fact that the area-based correction acts on each jet individually depending on its area in each event, while the offset correction applies the jet p_{T} correction on jets on average in each event. Also the variable ρ gives better description of the pileup contamination in each event than the variables N_{PV} and μ used in the offset correction. The better performance of the area-based correction was the reason to switch to the area-based method as the main pileup correction method for jet reconstruction from the $\sqrt{s} = 8$ TeV data. This correction is supplemented by offset correction to remove the residual dependence on pileup for JES as it is described in [95].

3.2.2 Shape-expansion Method

The shape-expansion method is an extension of the area 4-vector method. Besides correcting the jet 4-momentum, it corrects also any jet shape. The jet shape, S, is a function of the jet components. Few example jet shapes are described in App. C. The procedure of the correction is similar to the area 4-vector method: ghosts are added to the whole event, after which a certain jet clustering algorithm is applied leading to jets containing ghosts - these ghosts represent the susceptibility to contamination from pileup, and can be used for the correction. The pileup mass density is neglected in the following in order to make clear the main idea of the shape-expansion method, which is illustrated in Fig. 3.4. One can assume the jet shape, $S(p_{\mathrm{T}}^g)$, as a function of p_{T}^g of clustered ghosts from Eq. 3.5. The value $S(p_{\mathrm{T}}^g = 0) = S_{\mathrm{reco}}$ corresponds to the reconstructed jet shape influenced by pileup. The point $p_{\mathrm{T}}^g = 0$ is basically the initial value for ghosts when added into the event. After the final jets are obtained from jet clustering, the p_{T}^g can be set to arbitrary value. By setting p_{T}^g to certain value $p_{\mathrm{T}}^{g,\mathrm{pos}} > 0$, one artificially adds more pileup with pileup p_{T} density $\rho = p_{\mathrm{T}}^{\mathrm{pos}}/A^g$. Similarly, one can set the p_{T}^g to value $p_{\mathrm{T}}^{g,\mathrm{neg}} < 0$ which should artificially remove pileup. In particular at point $p_{\mathrm{T}}^g = -\rho A^g$, the pileup contribution should be removed in case the ρ was estimated precisely. One can extrapolate to the point $p_{\mathrm{T}}^g = -\rho A^g$ by using the Taylor expansion of the function $S(p_{\mathrm{T}}^g)$ at point $p_{\mathrm{T}}^g = 0$. Then the corrected jet shape is

$$S_{\mathrm{corr}} = \sum_{k=0}^{\infty} (-\rho A^g)^k \cdot \left.\frac{\partial^k S(p_{\mathrm{T}}^g)}{\partial {p_{\mathrm{T}}^g}^k}\right|_{p_{\mathrm{T}}^g=0}. \tag{3.13}$$

In practice, the derivatives can be estimated numerically from the values of function $S(p_{\mathrm{T}}^g)$ for several low positive p_{T}^g points. This pileup correction method is implemented in [96] where only the first three terms in this expansion are used

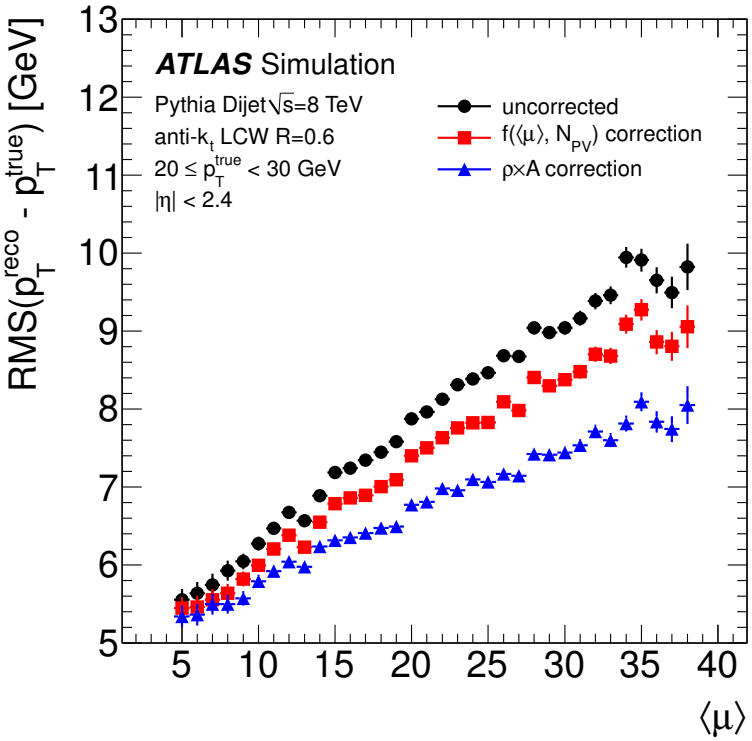


Figure 3.3: RMS of the $\left(p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{truth}}\right)$ distribution versus μ for reconstructed LCW anti-k_t $R = 0.6$ jets matched to truth-particle jets satisfying $p_{\mathrm{T}} \in [20, 30]$ GeV, in simulated dijet events. Three types of reconstructed jets are shown: without pileup correction (black circles), with the offset correction [82] (red squares), and with the area-based correction using Eq. 3.12 (blue triangles). The symbol $\langle\mu\rangle$ in the plot corresponds with the variable μ used in this thesis. [95]

for practical reasons - it was shown that the next term does not change the result significantly for most jet shapes, [84]. The p_{T}^g is identical for all ghosts in the jet when evaluating the derivatives in the original shape-expansion method. Although the method can be extended to use position $(y-\phi)$ dependence of relative magnitude of p_{T}^g for ghosts in the jet. The extension of Eq. 3.13 to account for the pileup mass density is straightforward as discussed in [84].

The Eq. 3.13 gives a generic approach to correct any jet shape. The addition of 4-momenta from Eq. 3.5 with $p_{\mathrm{T}}^g = -\rho A^g$ may be done explicitly for some jet shapes without using Eq. 3.13. This is shown for two jet shapes in the following: jet energy and energy correlation double ratio.

The jet energy can be also taken as a jet shape as function of p_{T}^g:

$$E(p_{\mathrm{T}}^g) = \sum_{i=1}^{N^{\mathrm{part}}} E_i + \sum_{j=1}^{N^{\mathrm{ghosts}}} p_{\mathrm{T}}^g \cosh y_j^g \tag{3.14}$$

where the sums run over all particles and ghosts clustered into the jet, respectively. The number of clustered ghosts is N^{ghosts}, and the number of particles is N^{part}. By substituting $p_{\mathrm{T}}^g = -\rho A^g$, the formula for the energy correction is

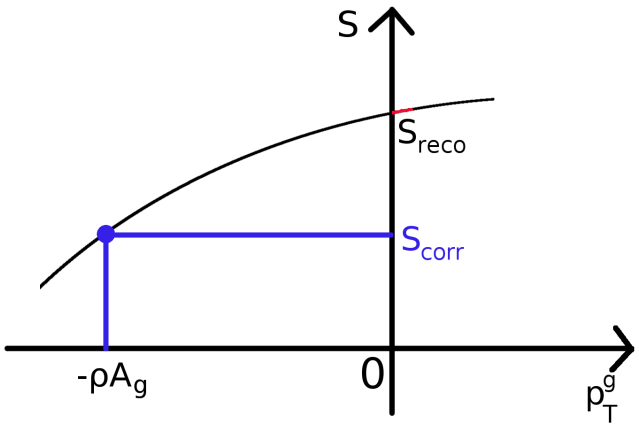


Figure 3.4: Cartoon to illustrate the shape-expansion method. The plot shows an example dependence of certain jet shape, S, on the p_{T}^g for ghosts clustered into the jet. The value S_{reco} is the value of jet shape before pileup correction. The point $p_{\mathrm{T}}^g = -\rho A^g$ for corresponds to the opposite of the expected pileup deposition, and hence to the corrected jet shape, S_{corr}. The values of S for low positive p_{T}^g (red points) can be used to numerically obtain the derivatives in Eq. 3.13.

obtained:

$$E_{\mathrm{corr}} = \sum_{i=1}^{N^{\mathrm{part}}} E_i - \sum_{j=1}^{N^{\mathrm{ghosts}}} \rho A^g \cosh y_j^g \tag{3.15}$$

which is identical with the corrected energy using the area 4-vector method in Eq. 3.9. One can also notice, that using the expansion Eq. 3.13 for jet shape in Eq. 3.14, the obtained corrected energy is identical with Eq. 3.15, and then also identical with corrected energy in the area 4-vector method. This can be shown also for other components of jet 4-momentum which then proves that the area 4-vector method corrects the jet 4-momentum and any function of it (e.g. jet mass, jet p_{T}) identically as the shape-expansion method.

The energy correlation double ratio $C_1^{(\beta)}$ (see App. C.3) can be as well expressed explicitly as a function of p_{T}^g:

$$C_1^{(\beta)}(p_{\mathrm{T}}^g) = \frac{\sum_{i=1}^{N^{\mathrm{part}}} \sum_{j=i+1}^{N^{\mathrm{part}}} p_{\mathrm{T}i} p_{\mathrm{T}j} \Delta R_{i,j}^{\beta} + \sum_{i=1}^{N^{\mathrm{part}}} \sum_{j=1}^{N^{\mathrm{ghost}}} p_{\mathrm{T}i} p_{\mathrm{T}}^g \Delta R_{i,j}^{\beta} + \sum_{i=1}^{N^{\mathrm{ghost}}} \sum_{j=i+1}^{N^{\mathrm{ghost}}} p_{\mathrm{T}}^g p_{\mathrm{T}}^g \Delta R_{i,j}^{\beta}}{\left(\sum_{k=1}^{N^{\mathrm{part}}} p_{\mathrm{T}k} + \sum_{i=1}^{N^{\mathrm{ghost}}} p_{\mathrm{T}}^g \right)^2}. \tag{3.16}$$

Consequently the formula for corrected $C_1^{(\beta)}$ is

$$C_{1,\mathrm{corr}}^{(\beta)} = \frac{\sum_{i=1}^{N^{\mathrm{part}}} \sum_{j=i+1}^{N^{\mathrm{part}}} p_{\mathrm{T}i} p_{\mathrm{T}j} \Delta R_{i,j}^{\beta} - \rho_0 A^g \sum_{i=1}^{N^{\mathrm{part}}} \sum_{j=1}^{N^{\mathrm{ghost}}} p_{\mathrm{T}i} \Delta R_{i,j}^{\beta} + \rho_0^2 A^{g2} \sum_{i=1}^{N^{\mathrm{ghost}}} \sum_{j=i+1}^{N^{\mathrm{ghost}}} \Delta R_{i,j}^{\beta}}{\left(\sum_{k=1}^{N^{\mathrm{part}}} p_{\mathrm{T}k} - \sum_{i=1}^{N^{\mathrm{ghost}}} \rho_0 A^g \right)^2}. \tag{3.17}$$

There are several disadvantages of the shape-expansion method, which can give non-perfect correction. The estimated pileup density may be imprecise at the position of the jet due to fluctuations in pileup. Also the distribution of pileup particles can be non-uniform in the $y - \phi$ space. To be able to use the Taylor expansion in Eq. 3.13, the function $S(p_{\mathrm{T}}^{g})$ must have continuous derivatives, e.g. the jet shape subjettiness can have discontinuous derivatives for certain particle configurations, although it was found that this effect can be neglected for this particular jet shape [84]. Another disadvantage of the shape-expansion method is that the corrected value for certain jet shapes can acquire non-physical values, e.g. the jet width is non-negative by definition, but the shape-expansion correction can lead to negative corrected jet width. In such case, the corrected jet shape can be set to the closest physical value.

3.3 Performance of the Shape-expansion Pileup Correction at the ATLAS Experiment

In this section, the shape-expansion method presented in Sec. 3.2.2 was tested on data recorded by the ATLAS detector, and compared to the performance in MC simulation. The results were published in the ATLAS note [1] - the author of this thesis made significant contribution to this publication, and he was the contact editor for this publication. The results from this note were included in the ATLAS publication [95].

The performance work uses the data from the LHC pp run at $\sqrt{s} = 8$ TeV in year 2012. The integrated luminosity after the data quality criteria is estimated to $L = (20.28 \pm 0.57)$ fb^{-1} using techniques similar to those described in [59]. The bunch spacing is 50 ns and the average number of pp interactions per bunch crossing is $\sim$21. The events are required to have a primary vertex reconstructed from at least two tracks with $p_{\mathrm{T}} > 400$ MeV. This requirement should reject non-collision backgrounds. Unprescaled trigger was used requiring at least one anti-k_t $R = 1.0$ jet with $p_{\mathrm{T}} > 450$ GeV and $|\eta| < 2.5$. No other event selection criteria were applied. The MC simulation of hard scattering events and pileup follows the description provided in Sec. 2.4. Two MC simulations of the inclusive jet events using LO QCD were used for comparison to data:

- Herwig++: fixed order calculation, PS, and hadronization simulated with Herwig++ 2.5.2 [97] using CTEQ6L [98] PDF set. The parameters of this MC simulation are optimized to reproduce underlying event data from the LHC experiments in the UE7-2 tune [99].

- Pythia8: fixed order calculation, PS, and hadronization simulated with Pythia 8.160 [77, 40] using CT10 [100] PDF set. The parameters of this MC simulation are optimized to reproduce underlying event data at $\sqrt{s} =$ 7 TeV from the LHC experiments in the AU2 tune [79].

The same trigger requirement is used for MC events as for events in data. The number of MC events is normalized to the number of data events. The correction was tested on anti-k_t $R = 1.0$ jets at detector level fulfilling $|\eta| < 2.5$ in several jet p_{T} ranges. The corrected detector level distributions were compared to particle

level anti-k_t $R = 1.0$ jet distributions referred to as truth-particle jets. The truth-particle jets are obtained from clustering all stable particles (with mean lifetime greater than 1 cm) from MC simulation except neutrinos and muons. The truth-particle jets contain the effect of underlying event which is removed for the corrected detector level jets on average, since the estimated ρ in Eq. 3.4 is influenced by underlying event. This makes the comparison between truth-particle and corrected detector distributions less interpretive, although the effect of underlying event is small, and it can be neglected for this comparison (the effect of underlying event roughly corresponds to the effect of three pileup events). The matching between detector and truth-particle jets is performed by requiring their distance ΔR to be less than 0.75.

The inputs to the shape-expansion pileup correction are the LCW topo-clusters (see Sec. 2.3.2). These inputs have zero mass, and therefore there is no need to use the pileup mass density term in the correction. The pileup p_{T} density was estimated using Eq. 3.4 where the patches were constructed as jets clustered with k_t $R = 0.4$ algorithm, and the jet area is defined as Voronoi area [93]. Only jets with $|y| < 2.0$ were used to estimate ρ. The estimated pileup p_{T} density is shown in Fig. 3.5 for data and the two MC generators. The two simulations use the same pileup simulation, and hence the different ρ distribution is caused by different modeling of the soft jet spectrum and underlying event. The agreement between data and MC simulation for the ρ distribution is slightly better for Herwig++ than Pythia8. Also for majority of observables the agreement is better for Herwig++ which is the reason why the Herwig++ simulation is used for comparison to data in the next.

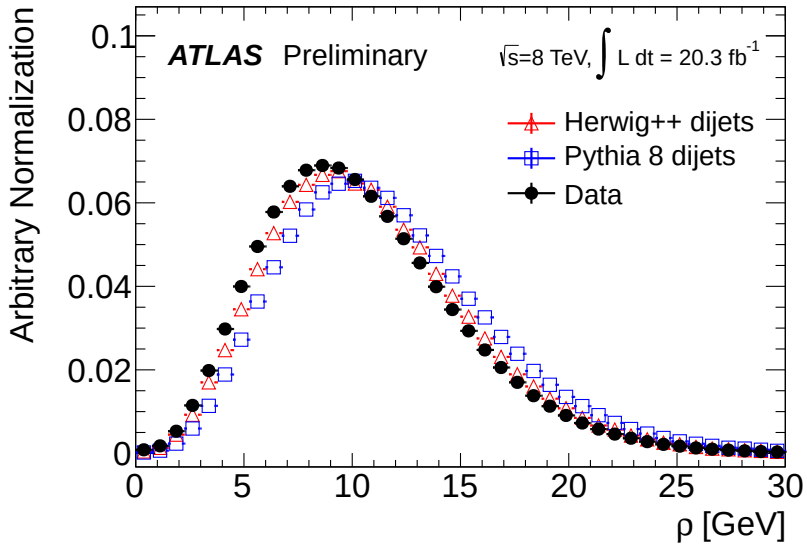


Figure 3.5: Distribution of the estimated ρ in inclusive jet events for data and two independent MC simulation samples (Herwig++ and Pythia8). Both MC generators use the same pileup simulation model. [1]

The performance of the correction on the jet shape called splitting scale, $\sqrt{d_{12}}$, (see App. C.2) for anti-k_t $R = 1.0$ jets is discussed in the following. Fig. 3.6a shows the distribution of $\sqrt{d_{12}}$ for three μ ranges without any pileup correction for data. The distribution of $\sqrt{d_{12}}$ differs significantly from the corresponding truth-

particle jet distribution, and it depends on μ. The $\sqrt{d_{12}}$ distribution is shifted to higher values with increasing μ. After the shape-expansion correction shown in Fig. 3.6b, the corrected distribution are very similar for the three μ ranges, and they are more consistent with the truth-particle jets than the uncorrected distributions.

The Fig. 3.7 shows the comparison between MC prediction and data before and after the pileup correction. There is a reasonable agreement between the MC prediction and data before the pileup correction. The small differences are caused by imperfect detector simulation and physics modeling. Similar reasonable agreement between MC prediction and data after the pileup correction can be observed. It means that the performance of shape-expansion correction is well-modeled in MC which is a key figure of merit. The corrected MC distributions of $\sqrt{d_{12}}$ more accurately reflects that expected from the truth-particle jet distributions. The small difference between truth-particle jet distribution and corrected MC distributions are caused by the detector effects, the imperfect calibration of topo-clusters, the presence of underlying event in truth-particle jets and the imperfect pileup correction.

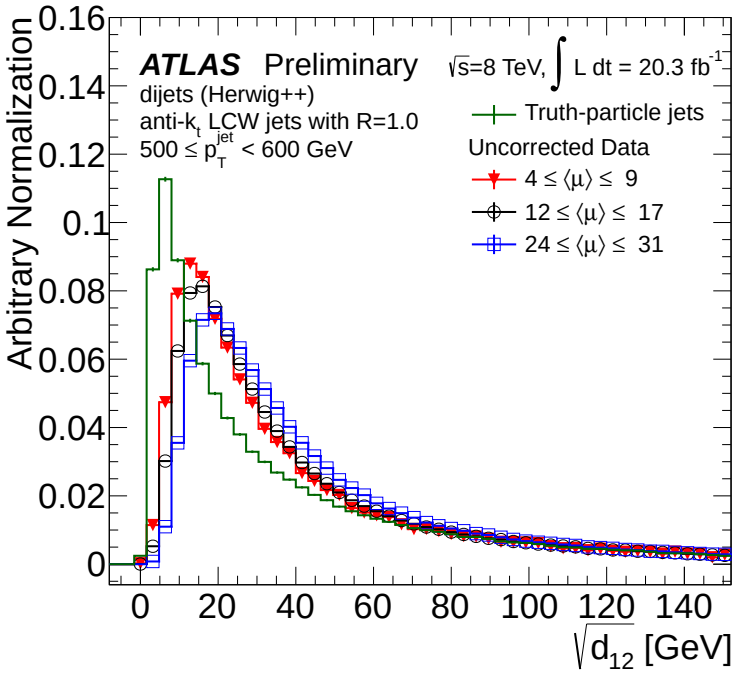
The performance of the shape-expansion method was tested also on anti-k_t $R = 1.0$ jets originating from boosted top quarks. To simulate boosted top quarks, a $Z' \to t\bar{t}$ sample was produced with PYTHIA 8.160 using MSTW2008LO PDF set [78] with AU2 tune and a Z' mass of $m_{Z'} = 1.75$ TeV. The dependence of mean $\sqrt{d_{12}}$ on μ is shown in Fig. 3.8 for jets from the $Z' \to t\bar{t}$ and dijet MC samples before and after pileup correction. The correction removes the dependence on pileup in both MC samples. The significant discrimination power between the two type of jets is preserved after the correction is applied.

3.4 Constituent Subtraction

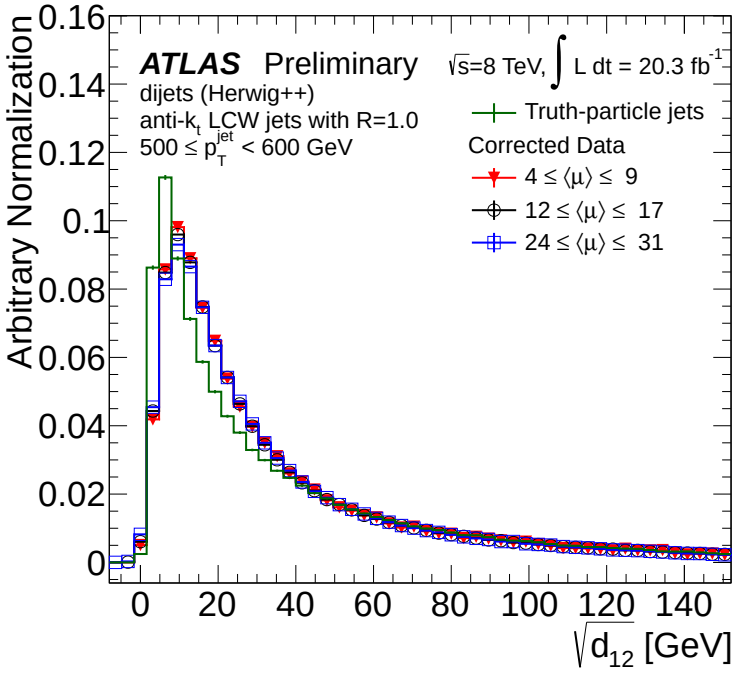
The area 4-vector method and the shape-expansion method assume that the pileup particles are uniformly distributed in $y-\phi$ space with momenta determined from the ρ estimation. Of course, fluctuations can occur in positions and p_{T} magnitude of the pileup particles. An improvement is achieved in the Constituent Subtraction method, which accounts for these fluctuations in positions. The author of this thesis developed this novel pileup subtraction method, and he made significant contribution to the publication [2].

The output of the Constituent Subtraction method is a jet with corrected constituents, therefore one can evaluate besides the corrected jet 4-momentum, also any jet shape, or perform any other operation with jet constituents. The Constituent Subtraction can be applied also on the whole event prior the jet clustering as discussed in Sec. 3.4.2. In the following, only the correction of individual jets is assumed.

The basic ingredient of the Constituent Subtraction is the pileup p_{T} density which is summarized in Sec. 3.2. Similarly as in the area 4-vector and shape-expansion methods, ghosts are uniformly added to the event before running the jet clustering algorithm. Then the jet clustering algorithm runs over all particles and ghosts delivering the same jets as in the case without the ghosts. The jets contain except the real particles also ghosts which are used to correct for pileup



(a) uncorrected



(b) corrected

Figure 3.6: Distribution of $\sqrt{d_{12}}$ for three μ ranges for anti-k_t $R = 1.0$ jets with $p_{\mathrm{T}} \in [500, 600]$ GeV for data. The distribution of $\sqrt{d_{12}}$ is shown uncorrected (left) and after the shape-expansion correction (right). The HERWIG++ prediction of $\sqrt{d_{12}}$ using stable truth particles (green) without pileup is also shown. The symbol $\langle\mu\rangle$ in the plot corresponds with the variable μ used in this thesis. [1]

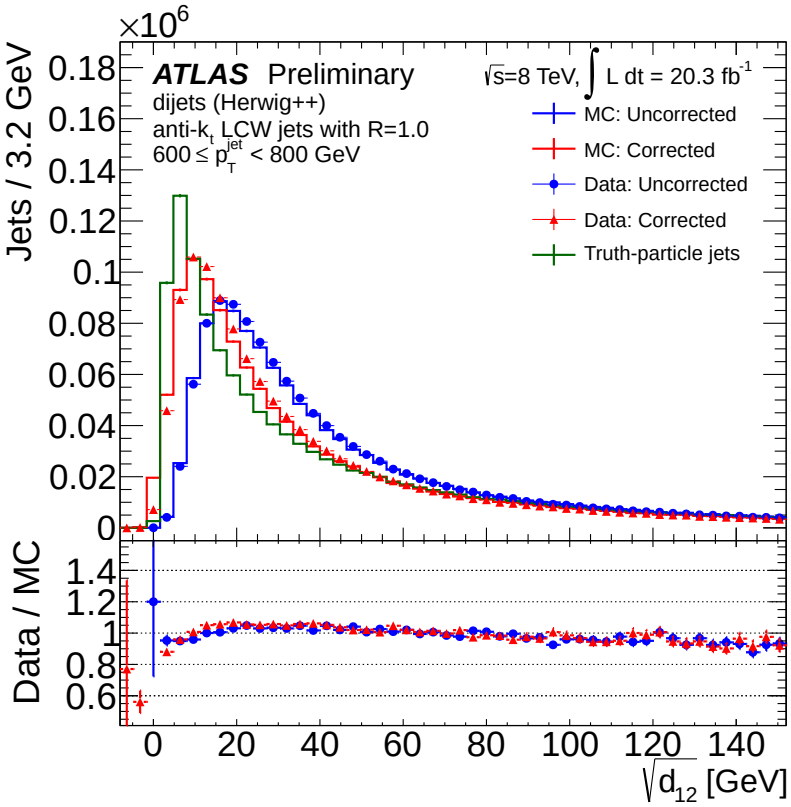


Figure 3.7: Comparison of the uncorrected (blue) and corrected (red) distributions of $\sqrt{d_{12}}$ for data (points) and MC simulation (solid histogram) for anti-k_t $R = 1.0$ jets with $p_{\mathrm{T}} \in [600, 800]$ GeV. The distribution of $\sqrt{d_{12}}$ computed using stable truth particles (green) is also shown. [1]

in the Constituent Subtraction differently as for the area 4-vector and shape-expansion methods.

Eq. 3.2 is translated into the 4-momentum of each ghost by identifying the p_{T}^g and m_δ^g with the amount of pileup within ghost area A^g:

$$\begin{aligned} p_{\mathrm{T}}^g &= A^g \cdot \rho(y, \phi), \\ m_\delta^g &= A^g \cdot \rho_m(y, \phi). \end{aligned} \tag{3.18}$$

An iterative procedure is used to define the scheme for calculating the specified amount of p_{T} and mass m_δ to subtract from each jet constituent. For each pair of particle i and ghost k, a matching scheme is implemented using the distance measure, $\Delta_{i,k}$, defined as

$$\Delta_{i,k} = p_{\mathrm{T}i}^{\alpha} \cdot \sqrt{(y_i - y_k^g)^2 + (\phi_i - \phi_k^g)^2}, \tag{3.19}$$

where α is a free parameter. The set of all distance measures, $\{\Delta_{i,k}\}$, is sorted from the lowest to the highest values. The pileup removal proceeds iteratively, starting from the particle-ghost pair with the lowest $\Delta_{i,k}$. At each step, the

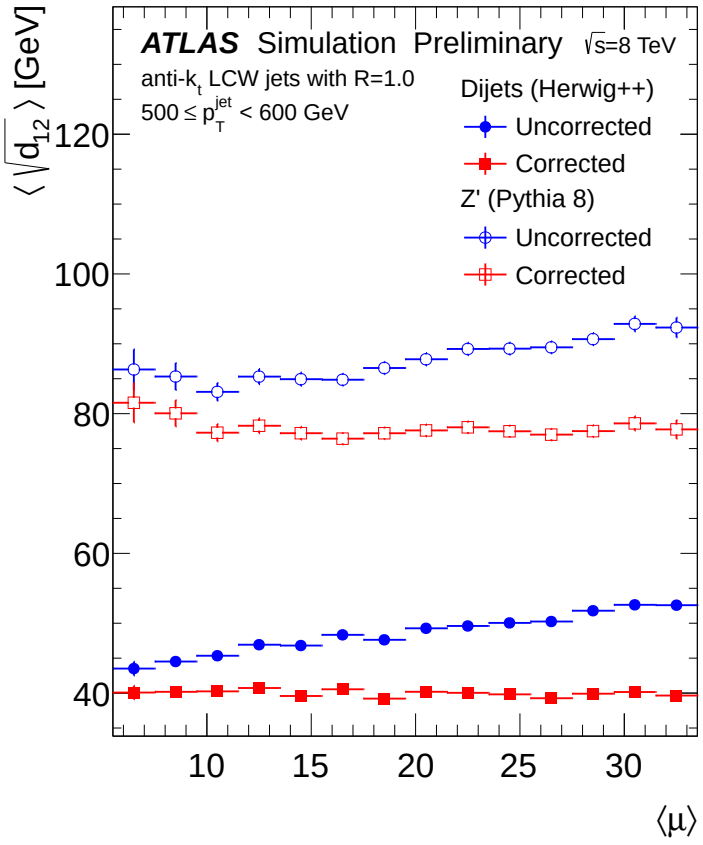


Figure 3.8: Dependence of mean $\sqrt{d_{12}}$ on μ for jets from boosted tops (open circles) and QCD jets (closed circles) before (blue) and after (red) the shape-expansion correction. The unphysical negative corrected $\sqrt{d_{12}}$ was set to zero for the computation of mean value. The symbol $\langle\mu\rangle$ in the plot corresponds with the variable μ used in this thesis. [1]

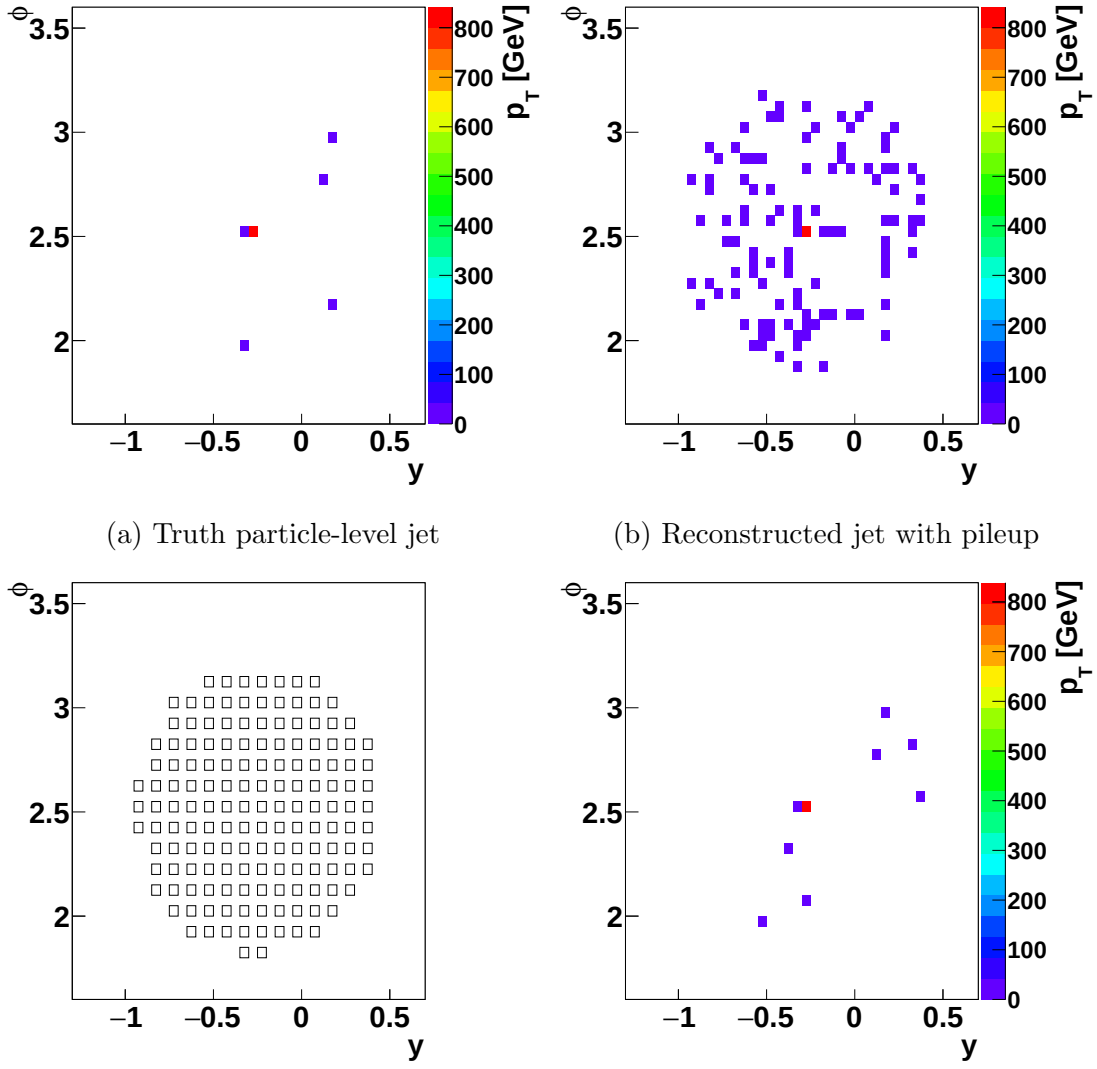
momentum p_{T} and mass m_δ of each particle i and ghost k are modified as follows.

$$\begin{aligned}
\text{If } p_{\mathrm{T}i} \geq p^g_{\mathrm{T}k}: \quad & p_{\mathrm{T}i} \longrightarrow p_{\mathrm{T}i} - p^g_{\mathrm{T}k}, \\
& p^g_{\mathrm{T}k} \longrightarrow 0; \\
\text{otherwise:} \quad & p_{\mathrm{T}i} \longrightarrow 0, \\
& p^g_{\mathrm{T}k} \longrightarrow p^g_{\mathrm{T}k} - p_{\mathrm{T}i}.
\end{aligned}
\qquad
\begin{aligned}
\text{If } m_{\delta i} \geq m^g_{\delta k}: \quad & m_{\delta i} \longrightarrow m_{\delta i} - m^g_{\delta k}, \\
& m^g_{\delta k} \longrightarrow 0; \\
\text{otherwise:} \quad & m_{\delta i} \longrightarrow 0, \\
& m^g_{\delta k} \longrightarrow m^g_{\delta k} - m_{\delta i}.
\end{aligned}
\tag{3.20}$$

The azimuth and rapidity of the particles and ghosts remain unchanged. The iterative process is terminated when the end of the sorted list is reached. A threshold $\Delta R^{\max}$ can be introduced to forbid subtraction between too distant pairs of particle and ghost where the distance is measured as in Eq. A.12. This leads to restriction: ghost-particle pairs with $\Delta R_{i,k} > \Delta R^{\max}$ are not used in the subtraction. Particles with zero p_{T} after the iterative process are discarded. It can happen that after the subtraction no real particle remains. This may be a signal that such a jet originates from pileup.

The Constituent Subtraction procedure is illustrated in Fig. 3.9 showing an example anti-k_t $R = 0.7$ jet from Pythia 8.180 dijet simulation at particle lev-

el. The event is overlaid with 25 pileup events, and ghosts are added into the event with $A^g = 0.01$. The resulting jet obtained by clustering only hard scattering particles is in Fig. 3.9a. The corresponding jet obtained when clustering hard scattering interaction and pileup particles is in Fig. 3.9b. Fig. 3.9c shows the ghosts which were clustered into that jet. The jet from Fig. 3.9b after the Constituent Subtraction is shown in Fig. 3.9d. One can observe that few pileup particles remained in the corrected jet, and also few hard scattering interaction particles are removed. However, there is significantly lower number of pileup particles in the corrected jet. More quantitative evaluation of the performance of the Constituent Subtraction is shown in Sec. 3.4.1.



(a) Truth particle-level jet

(b) Reconstructed jet with pileup

(c) Ghosts clustered into the jet

(d) Jet after Constituent Subtraction

Figure 3.9: Illustration of the Constituent Subtraction method on an example anti-k_t $R = 0.7$ jet from PYTHIA8 dijet simulation at particle level when overlaid with 25 pileup events. The p_{T} of the jet constituents in the $y - \phi$ space is shown for (a) jet without pileup, (b) jet with pileup, (c) ghosts clustered into the jet, and (d) corrected jet.

The Constituent Subtraction works equally well when applied directly to

Monte Carlo truth particles from simulation as when applied to a coarse pseudo-detector grid, over which the energy from the truth particles is distributed, to simulate calorimeter objects (the quantitative performance is done in Sec. 3.4.1). An important feature of the algorithm is that it preserves longitudinal invariance – an arbitrary jet after the correction and a subsequent boost in the direction of the colliding beams has the same constituents as the same jet which is first boosted and then corrected.

An important advantage of the Constituent Subtraction is the speed – it can be as much as twenty times faster compared to the shape-expansion method, depending on the type of the jet shape and the nominal jet radius. Furthermore, the shape-expansion correction must be determined for each jet shape in consideration, whereas the Constituent Subtraction approach provides a corrected set of constituents, from which any jet shape may be determined. Corrected constituents may also be used as inputs to jet grooming and tagging algorithms, e.g. the top-quark tagging using the shower deconstruction method [101]. In comparison to the jet cleansing method or CHS used by the CMS experiment, the Constituent Subtraction does not require any knowledge about the connection of each charged particle with the signal vertex or pileup vertexes, though such a knowledge might in principle be used to further enhance capabilities of the algorithm.

The Constituent Subtraction procedure has the following free parameters: A^g, $\Delta R^{\max}$, and α. The basic recommended settings are: $A^g = 0.01$, $\Delta R^{\max} \to \infty$, and $\alpha = 0$. The smaller the area A^g, the higher the density of the ghosts, which can lead to better performance of the Constituent Subtraction. On the other hand, too small A^g requires more CPU time, while the enhancement in performance can be negligible. The parameter $\Delta R^{\max}$ can prevent the combination of distant particle-ghost pairs. For configuration with $\alpha = 0$, the distance $\Delta_{i,k}$ is equivalent with the standard distance in $y - \phi$ space from Eq. A.12. The configuration with $\alpha > 0$ prefers to subtract the lower p_{T} constituents first, which may be beneficial since lower p_{T} constituents more often originate from pileup.

The software for the Constituent Subtraction is implemented as a part of the `FastJet Contrib` project [96].

3.4.1 Performance of the Constituent Subtraction

The performance of the Constituent Subtraction is evaluated for correction of jet p_{T}, jet mass, jet position and several jet shape variables, and it was compared to the area 4-vector method and to the shape-expansion method.

The studies presented are performed using pp events at $\sqrt{s} = 8$ TeV generated with Pythia 8.180, tune 4C [77, 40] at particle level. The effect of additional pp collisions is simulated by using soft events overlaid to the hard scattering interaction, which are also generated with Pythia 8.180. The CTEQ 5L, LO PDF set [102], configured to simulate the LHC conditions at $\sqrt{s} = 8$ TeV, are used for all event generation. Two processes are simulated without underlying event: dijet events covering the p_{T} range of $[10, 800]$ GeV and events with boosted top quarks from decay $Z' \to t\bar{t}$ of hypothetical boson Z' with mass of 1.5 TeV. The performance of the subtraction is tested using jets clustered with the anti-k_t algorithm with the distance parameter $R = 0.7$ and $R = 1.0$ and jets clustered

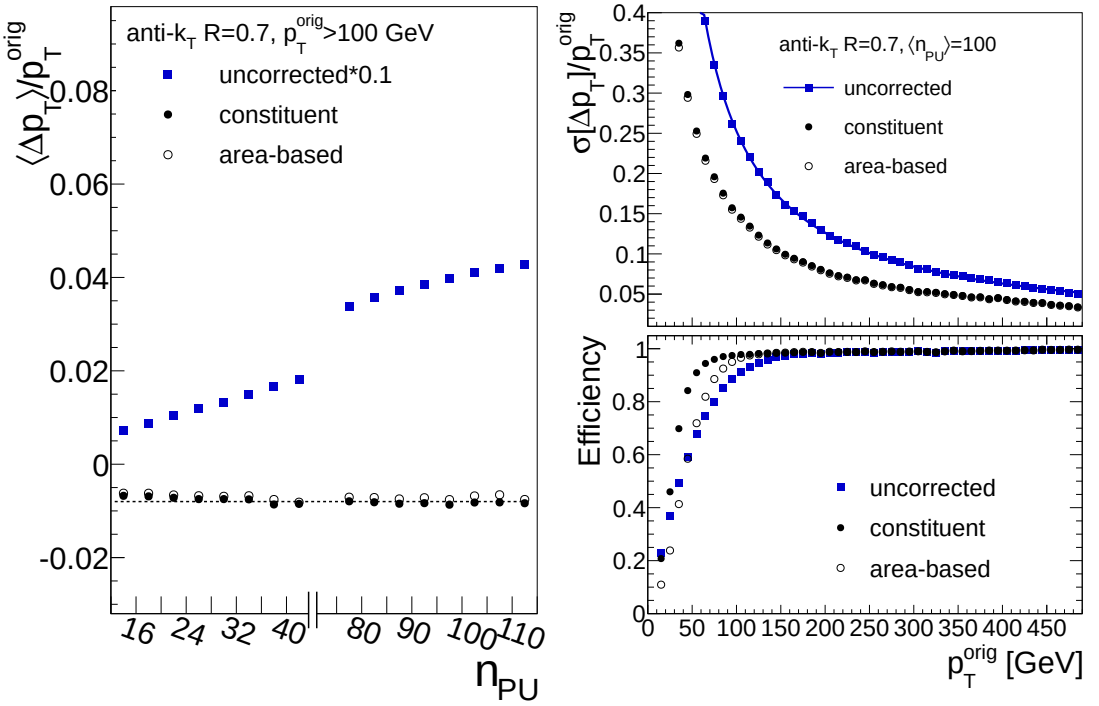
with C/A algorithm with $R = 1.2$. The number of pileup events, n_{PU}, has a Poisson distribution with a mean $\langle n_{\mathrm{PU}} \rangle$. Two pileup conditions are simulated, $\langle n_{\mathrm{PU}} \rangle = 30$ and $\langle n_{\mathrm{PU}} \rangle = 100$. These pileup configurations represent realistic conditions for the LHC Run I, and for the LHC Run II. On average, the pileup contribution to the hard scattering event can be described through mean value of the pileup p_{T} densities, $\langle \rho \rangle$, and pileup fluctuations characterized by standard deviation, $\sigma[\rho]$. For the used configuration $\langle n_{\mathrm{PU}} \rangle = 100$, these quantities are $\langle \rho \rangle \approx 75$ GeV and $\sigma[\rho] \approx 13$ GeV.

All the jet finding and background estimation is performed using `FastJet` 3.0.6 [94, 103]. The shape-expansion correction is performed using the tool called `GenericSubtractor` from the project `FastJet Contrib` 1.003 [96]. The patches in Eq. 3.4 are obtained by clustering particles with the k_t algorithm with distance parameter $R = 0.4$. The active area definition is used in Eq. 3.4. The non-negligible dependence of the pileup densities ρ and ρ_m on rapidity impacts each of the correction methods discussed below. Consequently, in order to focus the comparisons and performance evaluations, only patches with rapidity $|y| < 2.0$ are used in Eq. 3.4 and jets are required to be fairly central, with $|\eta| < 2.0$. The free parameters of the Constituent Subtraction are set to values: $A^g = 0.01$, $\Delta R^{\max} \to \infty$, and $\alpha = 0$.

The ability of the subtraction to correctly recover the kinematics of the jet can be characterized in terms of the following quantities: jet p_{T} response, jet p_{T} resolution, jet position resolution, and jet finding efficiency. These quantities are commonly used to evaluate the performance of the jet reconstruction, see e.g. [104, 105].

The jet p_{T} response can be defined as $\langle \Delta p_{\mathrm{T}} \rangle / p_T^{\mathrm{orig}} = \langle p_{\mathrm{T}}^{\mathrm{det}} - p_T^{\mathrm{orig}} \rangle / p_T^{\mathrm{orig}}$ where p_T^{orig} is the original jet p_{T} with no pileup and the $p_{\mathrm{T}}^{\mathrm{det}}$ is either the pileup corrected jet p_{T} or the p_{T} of the uncorrected jet that is jet clustered in the presence of pileup without subtraction. This quantity is also often referred to as the jet energy scale (JES). In the optimal situation, the jet p_{T} response should be zero which means that, on average, the algorithm can reconstruct the same p_{T} as with no pileup. The Fig. 3.10a shows the jet p_{T} response as a function of the number of pileup collisions, n_{PU}. The jet p_{T} response of subtracted jets differs from zero by less than 1%. The jet p_{T} response of subtracted jets is stable with respect to the pileup which is a crucial condition for the jet reconstruction. Without satisfying this condition any cut applied on the jet p_{T} would lead to a choice of different subset of jets depending on the size of the pileup. The small deviation from zero of the jet p_{T} response is resulting from ignoring the rapidity dependence of the pileup density ρ and ρ_m and from a small average bias of the pileup p_{T} density by a presence of the hard scattering event. Such small deviation can be easily corrected after the jet reconstruction by multiplying the jet p_{T} by a correction factor. The Constituent Subtraction performs equally well as the area 4-vector subtraction.

The small deviation of the jet p_{T} response from zero for corrected jets can be contrasted with the jet p_{T} response of uncorrected jets – in the case of low pileup scenario the p_{T} response is 10-20%, in the case of high pileup scenario the jet p_{T} response is around 40%. This means that in the high pileup case, there is on average approximately 40 GeV of the pileup background underneath each jet leading to a reconstruction of a typical 100 GeV jet as a 140 GeV jet if performing



(a) (b)

Figure 3.10: Jet p_{T} response (left). Jet p_{T} resolution (upper right) and jet finding efficiency (lower right). Jets prior the subtraction denoted as pileup (square markers) are compared to jets corrected by the Constituent Subtraction (full circles) and jets corrected by the area 4-vector method (open circles referred to as area-based). The dashed line in the left panel shows a constant at -0.8% to guide the eye. Curve in the upper middle panel represents a fit of the jet p_{T} resolution by c/p_{T} resulting $c \approx 25$ GeV. [2]

no pileup subtraction.

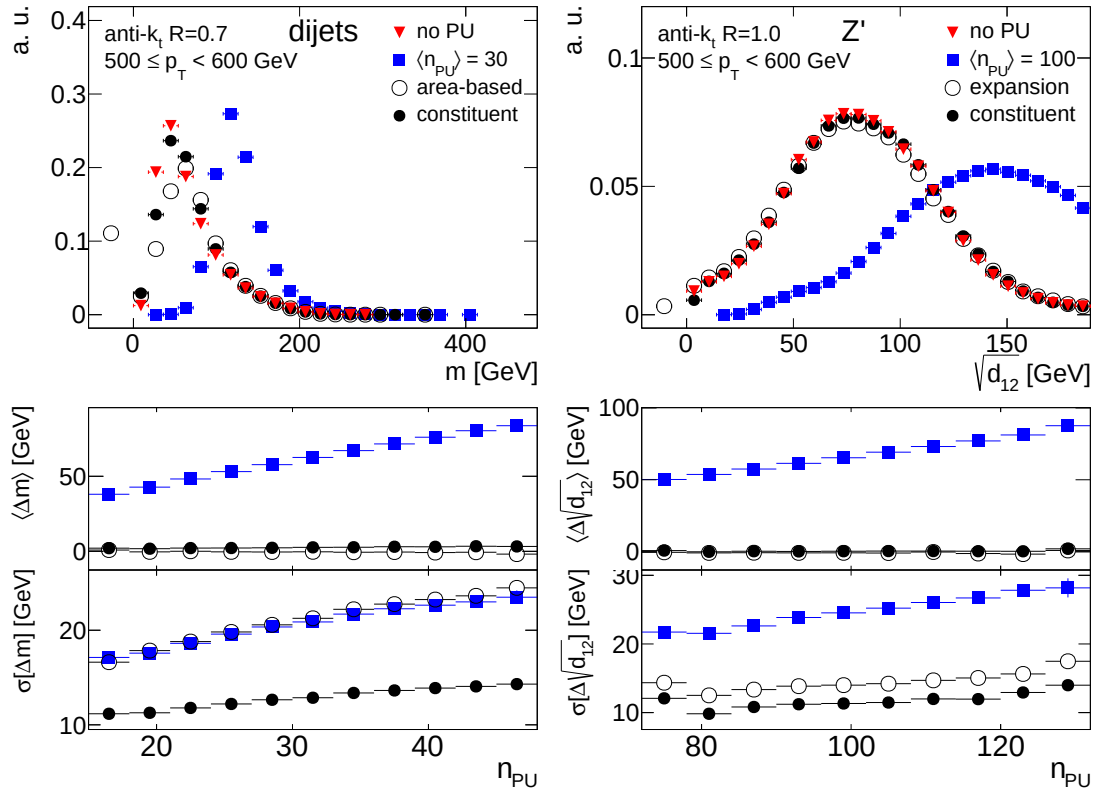
The second quantity characterizing the basic performance of the subtraction is the jet p_{T} resolution defined as $\sigma[\Delta p_{\mathrm{T}}]/p_T^{\mathrm{orig}} = \sigma[p_{\mathrm{T}}^{\mathrm{det}} - p_T^{\mathrm{orig}}]/p_T^{\mathrm{orig}}$ where $\sigma[...]$ denotes the standard deviation. The jet p_{T} resolution is dictated by the presence of fluctuations in the underlying pileup background leading to p_{T} dependence $\sigma[\Delta p_{\mathrm{T}}]/p_T^{\mathrm{orig}} = c/p_T^{\mathrm{orig}}$ where c is a constant. The jet p_{T} resolution is shown in the upper panel of Fig. 3.10b. The fit of the jet p_{T} resolution of uncorrected jets by c/p_T^{orig} leads to $c \approx 25$ GeV which results from the magnitude of $\sigma[\rho]$ and pileup fluctuations in $y - \phi$ space within each particular event. The Constituent Subtraction and area 4-vector method have similar jet p_{T} resolution while both methods significantly improve it.

The jet finding efficiency is defined as the number of original jets having a matching corrected (or uncorrected) jet divided by the number of original jets. The matching criterion is the distance in the $\eta - \phi$ plane between the original jet and corrected (or uncorrected) jet satisfying the condition $\Delta R^{\eta} < 0.2$. This quantity is plotted in the lower panel of Fig. 3.10b. It shows that in the case of high pileup events it is difficult in principle to reconstruct the jets with $p_T^{\mathrm{orig}} <$ 50 GeV due to the presence of sizable fluctuating background. The jet efficiency for jet $p_{\mathrm{T}} < 100$ GeV is better for the Constituent Subtraction than for the area

4-vector subtraction.

The jet position resolution characterizes the ability to recover the original jet axis in η or ϕ. It was found that the jets corrected with the area 4-vector method have slightly worse jet position resolution than jets corrected by the Constituent Subtraction.

The ability of the Constituent Subtraction method to recover the internal structure of jet has been tested by evaluating two jet shape variables: k_t splitting scale and N-subjettiness, see App. C for their definitions. The jet mass is also included in these tests, although it can be evaluated simply from the jet 4-momentum. All these variables are useful for identifying hadronic decays of boosted objects or quark/gluon tagging [106, 107, 108]. The Constituent Subtraction method is tested on various combinations of signal samples, pileup conditions, clustering algorithms, and jet shapes. The Constituent Subtraction can recover the original jet shape with a good accuracy in all evaluated combinations as it is shown in the following.



(a) (b)

Figure 3.11: Performance of the Constituent Subtraction for jets clustered with the anti-k_t algorithm for (a) the jet mass in dijet events and (b) the $\sqrt{d_{12}}$ variable in $Z' \to t\bar{t}$ events. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and Constituent Subtraction method, respectively. The n_{PU} dependence of mean and standard deviation are shown in the lower panel. [2]

A representative subset of performance plots for the anti-k_t algorithm is shown

in Fig. 3.11 for various jet p_{T} intervals[1] Four distributions of jet shapes are plotted in each figure: the original distribution without pileup, the distribution with pileup, the distributions corrected by the Constituent Subtraction and the shape-expansion[2] methods. To quantify precisely the performance of the correction, two quantities have been evaluated for the differences between jet shape x and its original value without pileup x^{orig}: the mean value of these differences $\langle \Delta x \rangle = \langle x - x^{orig} \rangle$ and the standard deviation of these differences $\sigma[\Delta x] = \sigma[x - x^{orig}]$ which represents the resolution. For each combination of configurations, the uncorrected distributions differ significantly from the corresponding original distribution, and have a significant dependence on n_{PU}. A substantial improvement is achieved by the Constituent Subtraction. The mean difference $\langle \Delta x \rangle$ does not exhibit the n_{PU} dependence and it is always centered near zero after the subtraction. The resolution $\sigma[\Delta x]$ is improved as well. The Constituent Subtraction method performs similarly or better when compared to the shape-expansion method.

For any of the studied jet shapes, the shape-expansion method can lead to negative corrected jet shapes that are unphysical. To better visualize the contribution of such values, the first bin with negative jet shape in plots of Figs. 3.11-3.13 is set to the fraction of negatively corrected jet shapes[3]. Unphysical values can also occur in the case of the area 4-vector correction of the jet mass when the corrected energy is smaller than the corrected momentum. Again, the negative bin represents the fraction of such jets. The fraction of unphysical jet shapes obtained from the shape-expansion method reaches up to $\sim 12\%$ depending on the p_{T} interval and the type of the jet shape.

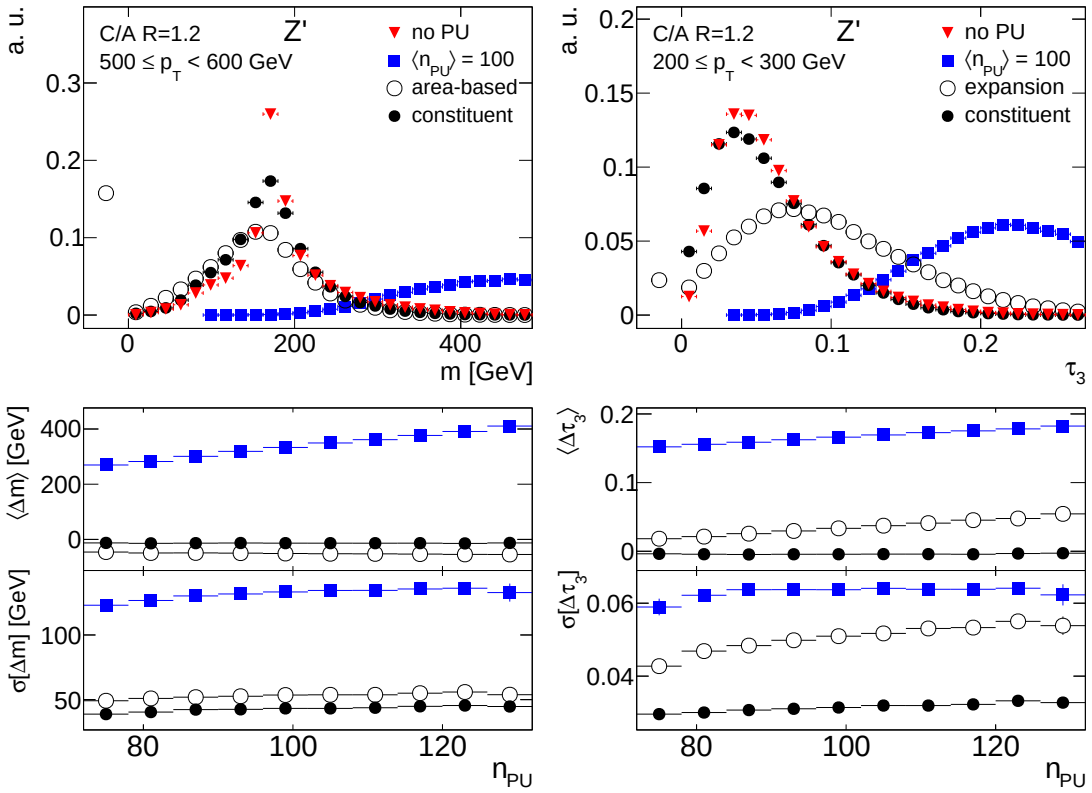
The Constituent Subtraction method has been tested also on the jets clustered with C/A algorithm which is often employed in various studies of the jet substructure and boosted objects [106]. The clustering in the C/A algorithm is based mainly on the geometry and thus it leads to jet with a different jet area compared to the anti-k_t algorithm [93]. The performance of the Constituent Subtraction method for C/A algorithm with distance parameter $R = 1.2$ is shown in Fig. 3.12. For this configuration, the impact of the pileup on jet shapes is much stronger compared to the configuration with the anti-k_t algorithm. The Constituent Subtraction can recover the original distributions and it exhibits significantly better ability to subtract the pileup compared to the shape-expansion method.

Further, the Constituent Subtraction method has been tested on jets reconstructed in events run through a simple simulation of a segmented detector. In this simulation, the $\eta - \phi$ plane is divided into cells of size 0.1×0.1. Particles pointing to the same cell are combined into one new effective particle by summing their energies. The mass of the cell is set to zero and its $\eta - \phi$ position is set

[1]The p_{T} of the jets without pileup is used to define the p_{T} intervals. The jets with pileup and the corrected jets are matched to the original jets without pileup.

[2]For the jet mass, the area 4-vector method using Eq. 3.9 is used which is identical to the shape-expansion method.

[3]For the shape-expansion correction of ratios τ_{21} or τ_{32}, the numerator and denominator are corrected individually. When at least one of these corrected variables is negative, the corrected τ_{21} or τ_{32} is counted as negative. For the calculation of the mean and resolution of τ_{21} or τ_{32}, the negative values are not used. For any other jet shapes, the negative values are set to zero which is the closest allowed physical value for all used jet shapes.



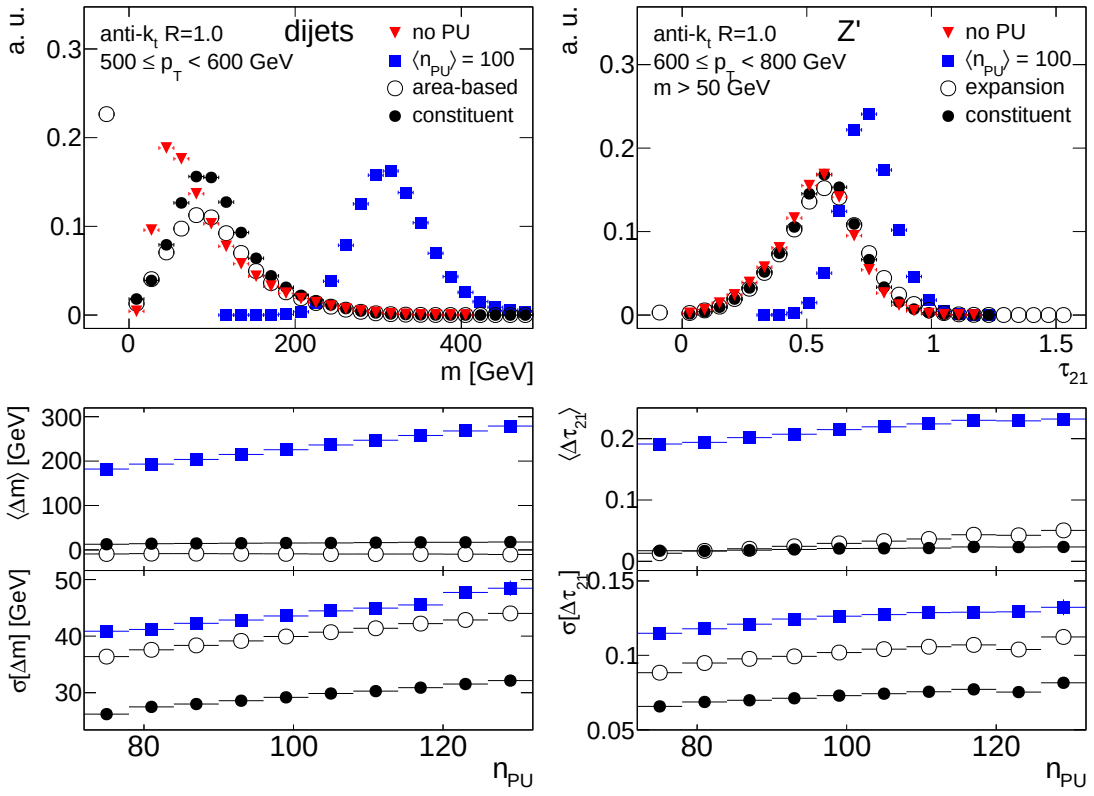
(a) (b)

Figure 3.12: Performance of the Constituent Subtraction for jets clustered with the C/A algorithm for (a) the jet mass and (b) the τ_3 variable in $Z' \to t\bar{t}$ events. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and Constituent Subtraction method, respectively. The n_{PU} dependence of mean and standard deviation are shown in the lower panel. [2]

to the center of the cell. These new effective particles have the same properties as the calorimeter clusters or towers used in real experiments and they are a combination of the pileup and signal. The jet finding algorithm runs over these events delivering jets that are corrected in the same way as for events composed of standard particles. A typical example of the performance of the subtraction methods in case of this simulation is shown in Fig. 3.13.

The Constituent Subtraction exhibits very similar performance as without the detector simulation which also applies for the shape-expansion method while the Constituent Subtraction again outperforms the shape-expansion method. Degradation of the performance of jet mass reconstruction is seen for all methods in the configuration of jet reconstruction with large distance parameter. This is a consequence of a sensitivity of jet mass to variations in the constituent transverse momenta of constituents that are at large angles with respect to the jet axis. While other variables are less sensitive to such variations, the jet mass can change significantly by adding a soft particle to the jet periphery.

An important test is to evaluate the identification of jets originating from boosted top quarks with respect to background jets. The splitting scale, $\sqrt{d_{12}}$, can be used to identify boosted top quarks from the Z' decay reconstructed with

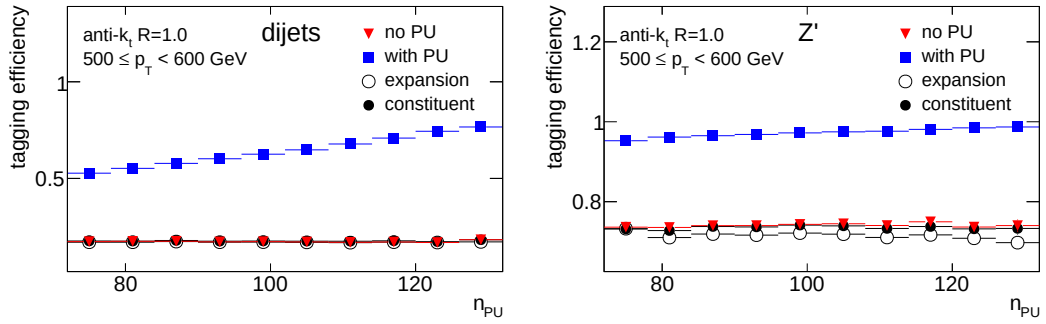


(a) (b)

Figure 3.13: Performance of the Constituent Subtraction in events simulating a segmented detector. Red triangles show distribution without pileup, blue squares show the uncorrected distribution with pileup, open and closed circles show distributions corrected by the shape-expansion and Constituent Subtraction method, respectively. The n_{PU} dependence of mean $\langle \Delta x \rangle$ and standard deviation $\sigma[\Delta x]$ are shown in the lower panel for each jet shape. [2]

the anti-k_t $R = 1.0$ algorithm, which are referred to as signal jets. The tagging cut $\sqrt{d_{12}} > 50$ GeV is tested. The tagging efficiency for signal jets is compared with the tagging efficiency for jets from the dijet sample, which are referred to as the background jets. The n_{PU} dependence of the tagging efficiencies for high n_{PU} values is shown in Fig. 3.14. Without pileup, the applied tagging cut has good discrimination between signal and background jets (the tagging efficiency for signal jets is ∼75% and the tagging efficiency for background jets is only ∼20%). The pileup shifts the $\sqrt{d_{12}}$ distribution to higher values for both, signal and background events. With pileup, the discrimination power between signal and background events is lower. By applying the Constituent Subtraction or the shape-expansion method, the tagging efficiencies are well restored. The tagging efficiency for signal jets is slightly lower when using the shape-expansion method, while the Constituent Subtraction restores this efficiency perfectly.

The performance of the Constituent Subtraction is tested also with different settings of parameters A^g and α. The above results are evaluated with settings $A^g = 0.01$ and $\alpha = 0$. It was found that the Constituent Subtraction is stable with respect to varying A^g. The variation of A^g by a factor of two does not lead to a change in any of the studied quantities that would be significant with respect



(a) (b)

Figure 3.14: Top-tagging efficiencies for (a) background and (b) signal samples using tagging cut $\sqrt{d_{12}} > 50$ GeV in events simulating a segmented detector. Red triangles show the tagging efficiency for jets without pileup. Blue squares show the tagging efficiency for jets with pileup, open and closed circles show tagging efficiency for jets corrected by the shape-expansion method and Constituent Subtraction method, respectively. [2]

to the statistical uncertainty shown on plots in this section. The performance was tested with parameter $\alpha = 1$, and it was not found any significant difference with respect to the performance with $\alpha = 0$.

3.4.2 Summary of the Constituent Subtraction

The above presented results demonstrate the stability and good performance of the Constituent Subtraction method. It has a solid ability to recover the original jet kinematics even in the presence of a sizable pileup. The Constituent Subtraction has generally similar performance as the area 4-vector method when correcting jet 4-momenta. A slightly better performance of the Constituent Subtraction in terms of jet position resolution and jet efficiency may be attributed to the correction of the jet internal structure. Further, the good ability to recover the jet internal structure is demonstrated by correcting several jet shape observables. In most cases, the performance of the Constituent Subtraction is much better than the performance of the shape-expansion method. The performance observed in the simulated events should be vetted and tested thoroughly within the LHC experiments using full detector simulation and data events.

Several improvements can enhance the Constituent Subtraction method, which is left for future work. Since the jet algorithm in the Constituent Subtraction procedure has only the role of adding expected pileup particles (ghosts), a novel application of the Constituent Subtraction approach would be to correct the whole event prior to the jet finding. In such case, the input to the Constituent Subtraction are not jet constituents with clustered ghosts, but the particles from the whole event with added ghosts. The ghosts should again mimic pileup and hence they should be distributed with the same detector acceptance criteria as the real particles. Then the ghosts can be rescaled using Eq. 3.18, and the same procedure can be applied as in case of Constituent Subtraction on individual jets. The jet finding algorithm can be applied after the subtraction on the subtracted event. Such correction have potential to avoid back-reaction or to improve the determi-

nation of event observables such as the missing transverse energy. Introducing a finite $\Delta R^{\max}$ may improve the performance and the speed of the correction when running over the whole event. The whole event subtraction can benefit from subtracting low p_{T} particles first with setting $\alpha > 0$. Another improvement can be achieved by including tracking information into the subtraction procedure. The tracks originating from charged pileup particles can serve as additional proxies besides ghosts to remove the pileup contribution in calorimeter.

4. Flavor Tagging of Jets

The flavor tagging of jets refers to the identification of jets from the fragmentation of a quark of certain flavor. In analyses at the ATLAS experiment, identification of b-jets and c-jets is possible due to the specific properties of B-hadrons and D-hadrons, respectively. The identification of b-jets is extremely useful for analyses with $t\bar{t}$ events since the $t\bar{t}$ events contain two b-quarks. Using b-tagging for $t\bar{t}$ events, the background events can be rejected, and also the reconstruction of the $t\bar{t}$ events is enhanced mainly for resolved topologies in all-hadronic and single lepton decay channels which contain many jets. The b-tagging is used in the measurement of $t\bar{t}$ production presented in Sec. 5.

The author of this thesis contributed to the Flavor Tagging Combined Performance Group of the ATLAS collaboration by evaluating the MC flavor tagging efficiencies. These MC efficiencies are used to correct the flavor tagging calibrations, and they are also used to preserve the normalization and the shape of the jet p_{T} vs. $|\eta|$ distribution after applying the flavor tagging calibrations, as described in Sec. 4.3. The author contributed to the ATLAS publication [3] and to the ATLAS Note [4]. Most of the figures in this section (Fig. 4.2, 4.3, 4.4 and 4.5) has been produced by the author.

4.1 Flavor tagging algorithms

There are algorithms for identification of b-jets and c-jets at the ATLAS experiment. At detector level in MC, the b-jets and c-jets are defined as calorimeter anti-k_t $R = 0.4$ jets which have within distance $\Delta R = 0.3$ at least one truth B-hadron and D-hadron, respectively. The light-flavor jets are the remaining calorimeter anti-k_t $R = 0.4$ jets. The identification of b-jets (b-tagging) is enhanced with respect to the identification of c-jets (c-tagging) due to more specific properties of the B-hadrons. Only b-tagging is discussed in the following, which is described in details in [3]. The details about c-tagging can be found in [109]. The b-tagging algorithm should identify as much as possible b-jets, while keeping the rate of the misidentified c-jets and light-flavor jets low. Also small fraction of jets originating from hadronic decay of a τ lepton can be misidentified by a b-tagging algorithm, which is not discussed in the following.

The b-quark fragments mostly to B^-, $\bar{B}^0$, or $\bar{B}_s^0$ meson or Λ_b baryon (the anti-b-quark fragments mostly to the corresponding antiparticles). These B-hadrons are distinct from other hadrons in the following properties: they have relatively high masses (above 5 GeV), decay multiplicity and specific mean lifetimes (in range $[0.4, 0.5]$ mm). The B-hadron mean lifetimes allow to travel specific distances which are distinguishable by the ATLAS Inner Detector. The traveled distance is a random variable and depends also on the p_{T} of the B-hadron. A typical example is given using Eq. A.15 as follows. A B^- meson with p_{T} of 50 GeV travels mean distance of $\sim$4.7 mm in the transverse plane. Such large distance between the pp primary vertex and the B-hadron decay vertex (secondary vertex) can be measured using tracks. There are several approaches of how to use this distance or the other specific properties of the B-hadrons to identify b-jets. Each of these approaches leads to one discriminating variable which expresses the

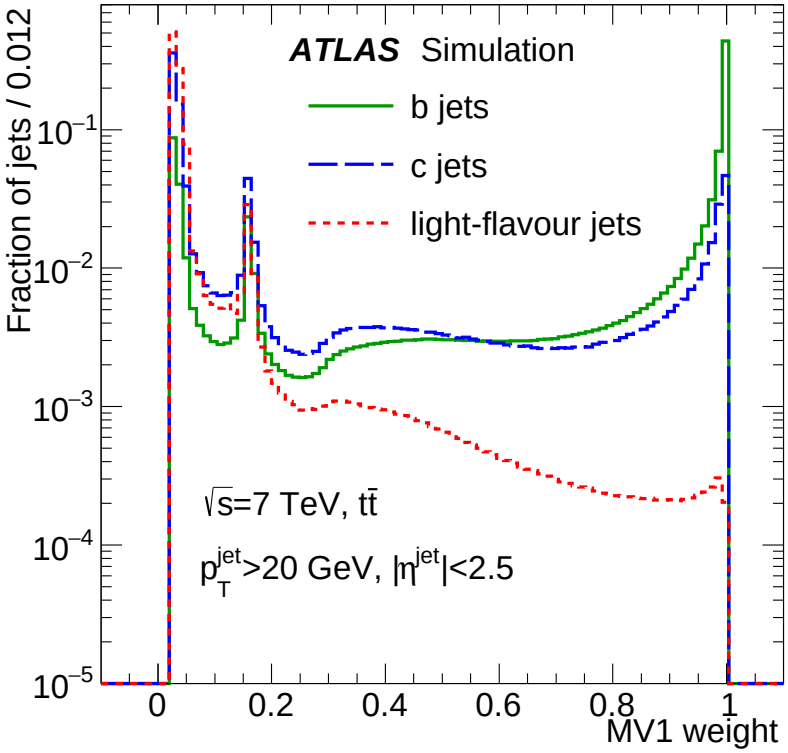


Figure 4.1: Distribution of the tagging weight w^{MV1} obtained with the MV1 algorithm, for three different flavors of jets from MC $t\bar{t}$ simulation. [3]

probability that a given jet originates from a b-quark, as outlined in the following.

The b-tagging is performed using the subset of tracks in the event that are associated with the calorimeter jet. There are two main algorithms used for the b-tagging: vertex-based and impact parameter-based algorithms. The most discriminating vertex-based method at the ATLAS experiment is the SV1 algorithm. It uses an algorithm to reconstruct the secondary vertex, and the discrimination is based on the distance between the secondary and primary vertexes. The invariant mass of all tracks associated with the secondary vertex is exploited as well in the SV1 algorithm. The most discriminating impact parameter-based algorithm at the ATLAS experiment is the IP3D algorithm, which is based on the impact parameters of the tracks with respect to the main primary vertex. The SV1 and IP3D algorithms have good discrimination power between b-jets and light-flavor jets. The discrimination between b-jets and c-jets is much weaker, since the ratio between mass and mean lifetime are similar for B-hadrons and D-hadrons (this leads to similar mean distance in Eq. A.15). To enhance the discrimination between b-jets and c-jets, another algorithm was developed at the ATLAS experiment called the JetFitter algorithm. It exploits the topological structure of B-hadron decay, and the subsequent decays of B-hadron decay products. The JetFitter algorithm uses six input nodes in an artificial neural network. Another algorithm was commissioned, called the IP3D+JetFitter, which is similar to the JetFitter algorithm. This algorithm uses the IP3D variable as an additional input node for the JetFitter algorithm.

To achieve the best performance, the variables SV1, IP3D, and IP3D+JetFitter

are combined in an artificial neural network leading to two algorithms, MV1 and MV1c. The MV1 algorithm is optimized to suppress mostly light-flavor jets, while the MV1c algorithm is optimized to suppress mostly c-jets. The MV1 algorithm is used to identify b-jets in the measurement presented in Sec. 5, and only this algorithm is discussed in the following.

The MV1 algorithm outputs one discriminating variable called the MV1 weight, w^{MV1}. The normalized distribution of w^{MV1} for different jet flavors from $t\bar{t}$ simulation is shown in Fig. 4.1. The bump around the point $w^{\mathrm{MV1}} = 0.15$ corresponds mostly to jets for which no secondary vertex was found within the SV1 algorithm. There is no clear boundary among different flavors, although the MV1 weight distributions for b-jets is shifted to higher values when compared to c-jets and light-flavor jets which gives certain discriminating power for this observable. The most common practical usage of this discriminating power is described in the following. A weight cut, w^{cut}, is chosen. Then jets with $w^{\mathrm{MV1}} > w^{\mathrm{cut}}$ are called as b-tagged jets, and the other jets are called non-b-tagged. The b-tagged jets can be used for a selection criterion in certain analyses, e.g. by requesting one or more b-tagged jets. The performance of b-tagging is expressed in b-tagging efficiencies for b-jets, c-jets and light-flavor jets (ε_b, ε_c and $\varepsilon_{\mathrm{light}}$), or rejection rates for c-jets and light-flavor jets (RR_c and RR_{light}). The ε_b is defined as the ratio between number of b-jets with $w^{\mathrm{MV1}} > w^{\mathrm{cut}}$ and the total number of b-jets. The ε_c and $\varepsilon_{\mathrm{light}}$ are defined similarly. The rejection rates RR_c and RR_{light} are defined as the inverse of the efficiency ε_c and $\varepsilon_{\mathrm{light}}$, respectively. An illustrative example is given in the following. The choice of $w^{\mathrm{cut}} = 0.7892$ gives benchmarks $\varepsilon_b \approx 70\%$, $RR_c \approx 5$, and $RR_{\mathrm{light}} \approx 140$ for anti-k_t $R = 0.4$ jets from POWHEG+PYTHIA6 simulation[1]. An analysis focusing on $t\bar{t}$ events can require one or more b-tagged jets. Since there are two b-quarks in each $t\bar{t}$ event, only $1 - (1 - \varepsilon_b)^2 \approx 9\%$ of $t\bar{t}$ events is rejected. But since only every ∼140th light-flavor jet is b-tagged, large fraction of background events can be rejected and overall, the $t\bar{t}$ purity of the selected sample of events can be enhanced.

The choice of the cut value w^{cut} depends on the strategy in particular analysis. The Fig. 4.2 shows the dependence of the rejection rate RR_{light} on the efficiency ε_b. Each point corresponds to a unique w^{cut} which can be obtained from distributions in Fig. 4.1. Going from lower w^{cut} to higher w^{cut}, there is a trade-off between large b-jet efficiencies with low rejection rates and large rejection rates with low b-jet efficiency. The analyses which require large sample of events can use selection based on smaller w^{cut} values. The analysis which require large signal purity can use selection based on higher w^{cut} values.

Another point, which needs to be taken into account in analyses, is the p_{T} dependence of the b-tagging performance, see Fig. 4.3. The discrimination power between b-jets and light-flavor jets increases with increasing jet p_{T} until the point ∼100 GeV, at which it starts to decrease. A qualitative explanation is given as follows. The impact of the jet p_{T} on the MV1 algorithm is small since the track impact parameter is approximately boost invariant. The observed jet p_{T} dependence is explained by track reconstruction effects. The impact parameter from a track with very low p_{T} has high uncertainty due to multiple scattering in the Inner Detector material. The impact parameter uncertainty due to multiple scattering decreases with increasing p_{T} leading to better discrimination between

[1]These benchmarks are only for illustration. They depend on MC generator, jet p_{T}, jet η.

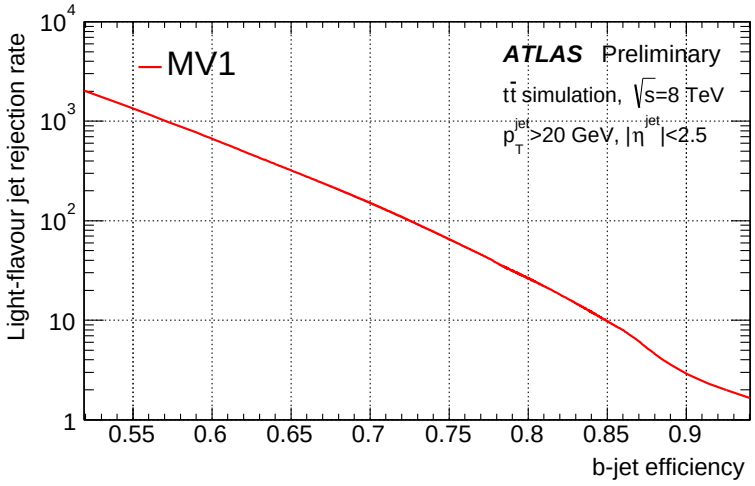


Figure 4.2: Light-flavor jet rejection rate versus b-jet efficiency using the MV1 b-tagging algorithm, as evaluated for jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$ in a sample of simulated $t\bar{t}$ events. [4]

b-jets and light-jets as it is observed. With $p_{\mathrm{T}} > 100$ GeV, another effect starts to have impact. The tracks from a high p_{T} jet are more collimated, and hence the hits in the Inner detector are more close to each other, which is referred to as the *dense environment* in the core of the jet. This causes lower performance of the track reconstruction algorithm leading to worse b-tagging performance. The track reconstruction algorithm can be improved in dense environment as it is described in [110]. It is shown that by using the improved track reconstruction algorithm, the light-jet rejection increases significantly for fixed b-jet efficiency for jets with $p_{\mathrm{T}} > 100$ GeV. There is another effect which limits the b-tagging performance for high p_{T} jets: with increasing p_{T}, larger fraction of B-hadrons decay beyond the first layer of the Pixel Detector.

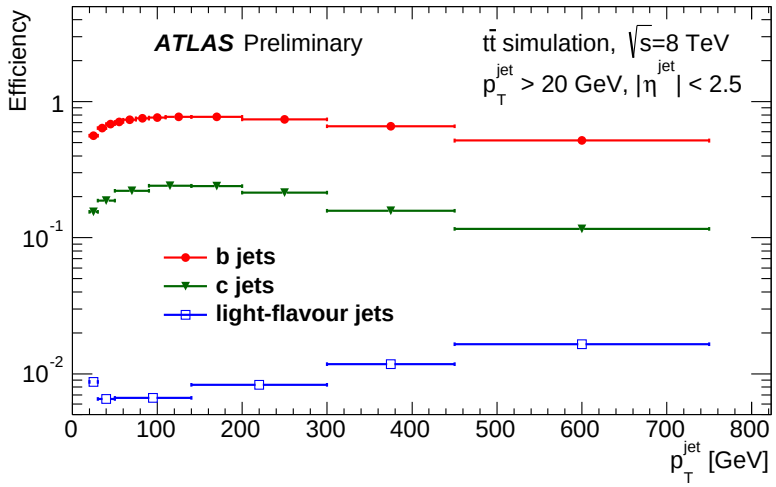


Figure 4.3: Efficiency of the MV1 tagger to select b, c, and light-flavour jets, as a function of jet p_{T}. The weight cut selection on the MV1 weight is chosen to be 70% efficient for b-jets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$, as evaluated on a sample of simulated $t\bar{t}$ events. [4]

4.2 MC Dependence of the *b*-tagging Efficiencies

There are several MC generators which uses different models to simulate the parton shower and the hadronization. Moreover, each MC generator can use different models for the decay of hadrons. These differences can lead to different *b*-tagging performance.

The Fig. 4.4 and 4.5 show the *b*-tagging efficiency for *b*-jets for several MC generators. The fixed order NLO QCD calculation was simulated using POWHEG [111, 112, 113] in all cases, and just the parton shower and hadronization is modeled with different MC generators: PYTHIA6 [77], PYTHIA8 [77, 40], HERWIG [41], and HERWIG++ [97]. The decay of the heavy-flavor hadrons, *B*-hadrons and *D*-hadrons, can be simulated using the same generator as used for hadronization (Fig. 4.4) or using EvtGen generator, [114], (Fig. 4.5) which simulates the decays of heavy-flavor hadrons more precisely in complex sequential decays. One can see that the differences of *b*-tagging efficiencies among the MC generators using purely PYTHIA6, PYTHIA8, HERWIG, and HERWIG++ for heavy-flavor decays are larger than among the MC generators using EVTGEN for heavy-flavor decays. This suggests that the modeling of the heavy-flavor decay has large impact on the *b*-tagging efficiency of *b*-jets, and the different parton shower and hadronization model has no such large impact.

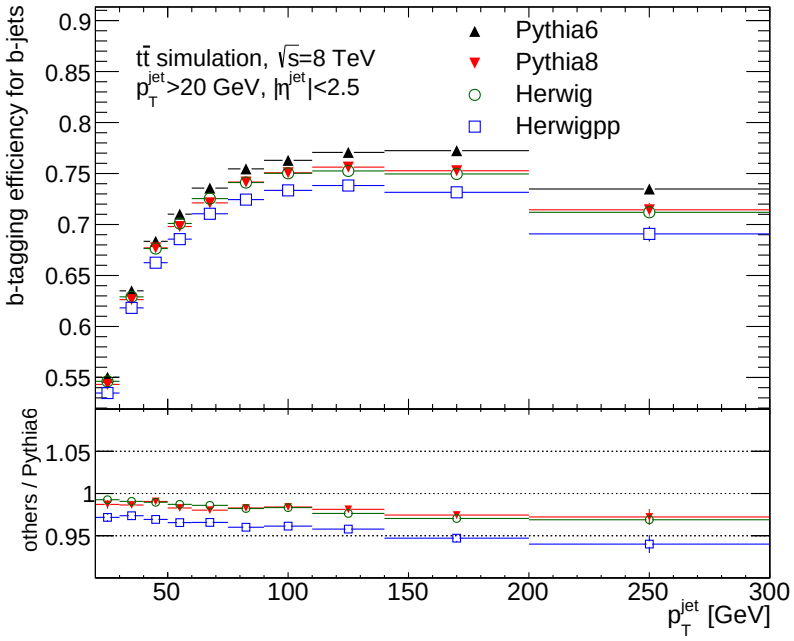


Figure 4.4: Dependence of *b*-tagging efficiency on jet p_{T} for four MC generators. The matrix element is simulated with POWHEG in all cases. The decay of the heavy-flavor hadrons are simulated using the same generator as was used for the parton shower and hadronization.

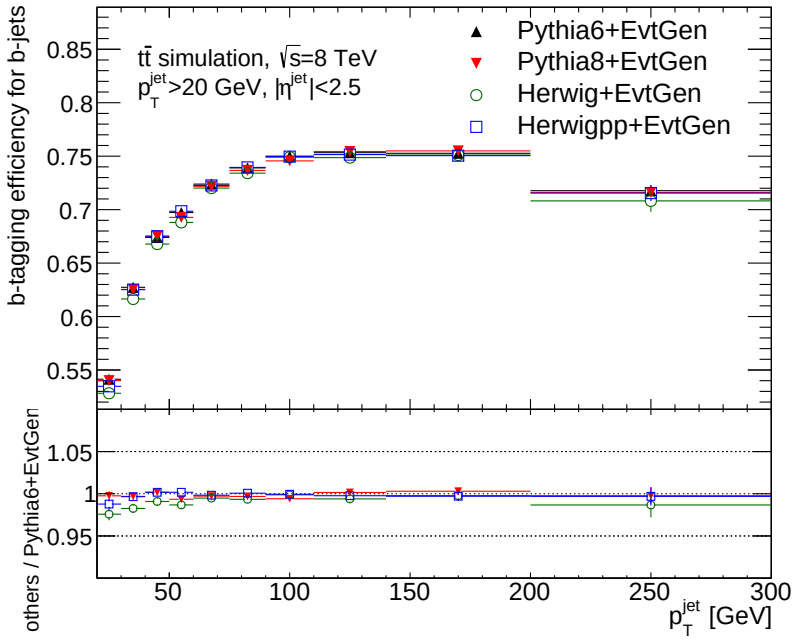


Figure 4.5: Dependence of b-tagging efficiency for b-jets on jet p_{T} for four MC generators using EvtGen for heavy flavor decays. The matrix element is simulated with Powheg in all cases.

4.3 Calibration of the Flavor Tagging Performance

The detector and physics simulation in MC are not perfect. For this reason, the b-tagging performance in MC simulated events can be different than in real data events, and hence a correction for the MC simulation is necessary which is called as flavor tagging calibration. The procedure of this correction is explained in this section.

Firstly, a calibration measurement determines the b-tagging efficiencies for certain flavor of jets in MC and in data. Several methods have been developed to measure the efficiencies ε_b, ε_c, and $\varepsilon_{\mathrm{light}}$, and they are described in detail in [3]. For example, the ε_b can be measured using the $p_{\mathrm{T}}^{\mathrm{rel}}$ method. In this method, the b-tagging efficiency is obtained from jets with a muon inside using the variable called $p_{\mathrm{T}}^{\mathrm{rel}}$ which is defined as the momentum of the muon transverse to the jet axis. Templates of $p_{\mathrm{T}}^{\mathrm{rel}}$ in simulated events are constructed for b-jets, c-jets and light-flavor jets separately, and these are fitted to the $p_{\mathrm{T}}^{\mathrm{rel}}$ distribution to obtain the fraction of b-jets before and after requiring b-tagging. This fit is done separately in MC and in data leading to the efficiencies shown in Fig. 4.6 as a function of jet p_{T}. One can see that the b-tagging efficiencies for b-jets are larger in MC than in data in this calibration measurement. Similarly, there are calibration measurements for ε_c and $\varepsilon_{\mathrm{light}}$ in which these efficiencies are again different between MC and data. Therefore, a correction of the MC is needed to remove this difference in each analysis which uses b-tagging.

The flavor tagging calibration is done by multiplying the event weight for each MC event by a scale factor, $\mathrm{SF}_{\mathrm{event}}$, as was explained in Sec. 2.4. This should ensure that the overall b-tagging performance will be the same in MC as in data.

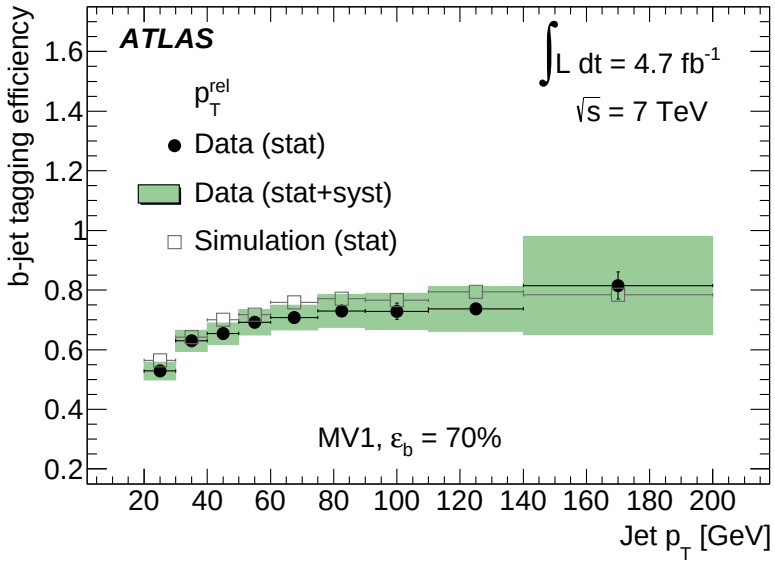


Figure 4.6: The b-jet tagging efficiency in data and simulation for the MV1 tagging algorithm at 70% efficiency obtained with the $p_{\mathrm{T}}^{\mathrm{rel}}$ method. [3]

Efficiency scale factor is defined from the calibration measurement as

$$\mathrm{SF}_{\mathrm{ef}}^{\mathrm{cal}} = \frac{\varepsilon_{\mathrm{Data}}^{\mathrm{cal}}}{\varepsilon_{\mathrm{MC,cal}}^{\mathrm{cal}}} \tag{4.1}$$

where $\varepsilon_{\mathrm{Data}}^{\mathrm{cal}}$ and $\varepsilon_{\mathrm{MC,cal}}^{\mathrm{cal}}$ are the b-tagging efficiencies for certain jet flavor obtained from the calibration measurement for data and MC, respectively. An example of these efficiencies is shown in Fig. 4.6, from which the efficiency scale factor as a function of jet p_{T} is obtained. Each jet flavor has its specific calibration measurement, and hence its own scale factors. The MC generator used in a certain analysis can be different than the MC generator which was used in the calibration measurement. Since the b-tagging performance depends on the MC generator, the efficiency scale factor is corrected for this effect by introducing the corrected jet scale factor

$$\mathrm{SF}_{\mathrm{ef}} = \mathrm{SF}_{\mathrm{ef}}^{\mathrm{cal}} \cdot \mathrm{SF}_{\mathrm{MC/MC}}, \tag{4.2}$$

where the MC-to-MC scale factor, $\mathrm{SF}_{\mathrm{MC/MC}}$, is defined as

$$\mathrm{SF}_{\mathrm{MC/MC}} = \frac{\varepsilon_{\mathrm{MC,cal}}}{\varepsilon_{\mathrm{MC,ana}}} \tag{4.3}$$

where $\varepsilon_{\mathrm{MC,cal}}$ is the b-tagging efficiency in the MC generator used for the calibration measurement and $\varepsilon_{\mathrm{MC,ana}}$ is the b-tagging efficiency in the MC generator used in the analysis. In Eq. 4.2, the efficiencies $\varepsilon_{\mathrm{MC,cal}}$ and $\varepsilon_{\mathrm{MC,ana}}$ and the $\mathrm{SF}_{\mathrm{ef}}^{\mathrm{cal}}$ are derived for the same flavor. The efficiencies $\varepsilon_{\mathrm{MC,cal}}$ and $\varepsilon_{\mathrm{MC,cal}}^{\mathrm{cal}}$ may be different because any calibration measurement uses a specific event selection while the efficiency $\varepsilon_{\mathrm{MC,cal}}$ used for $\mathrm{SF}_{\mathrm{MC/MC}}$ is derived without any event selection. In the calibration process, another scale factor is used which is the inefficiency scale factor defined as

$$\mathrm{SF}_{\mathrm{inef}} = \frac{1 - \varepsilon_{\mathrm{MC,ana}} \cdot \mathrm{SF}_{\mathrm{ef}}}{1 - \varepsilon_{\mathrm{MC,ana}}}. \tag{4.4}$$

The final scale factor used as multiplicative factor for the event weight is defined as

$$\mathrm{SF}_{\text{event}} = \mathrm{SF}_{b\text{-jets}} \cdot \mathrm{SF}_{c\text{-jets}} \cdot \mathrm{SF}_{\text{light-jets}} \tag{4.5}$$

where the scale factors $\mathrm{SF}_{b\text{-jets}}$, $\mathrm{SF}_{c\text{-jets}}$ and $\mathrm{SF}_{\text{light-jets}}$ are the scale factors constructed from all b-jets, c-jets and light-jets, respectively. The $\mathrm{SF}_{b\text{-jets}}$ is defined as

$$\mathrm{SF}_{b\text{-jets}} = \prod_{\substack{i \in b\text{-tag} \\ b\text{-jets}}} \mathrm{SF}^{i}_{\text{ef}} \cdot \prod_{\substack{j \in \text{non-}b\text{-tag} \\ b\text{-jets}}} \mathrm{SF}^{j}_{\text{inef}}. \tag{4.6}$$

where the $\mathrm{SF}^{i}_{\text{ef}}$ is the efficiency scale factor corresponding to i-th b-tagged b-jet and the $\mathrm{SF}^{j}_{\text{inef}}$ is the inefficiency scale factor corresponding to j-th non-b-tagged b-jets. Similarly are defined the scale factors $\mathrm{SF}_{c\text{-jets}}$ and $\mathrm{SF}_{\text{light-jets}}$ for which the two products runs over c-jets and light-jets, respectively.

By using the b-tagging event scale factors $\mathrm{SF}_{\text{event}}$ in certain analysis, the b-tagging performance in MC will by much more similar to the performance in data. The presence of the inefficiency scale factors $\mathrm{SF}_{\text{inef}}$ ensures that the overall normalization of the sample is conserved, since the fraction of b-tagged b-jets changes from $\varepsilon_{\text{MC,ana}}$ to $\varepsilon_{\text{MC,ana}} \cdot \mathrm{SF}_{\text{ef}}$, and the fraction of non-b-tagged b-jets changes from $(1 - \varepsilon_{\text{MC,ana}})$ to $(1 - \varepsilon_{\text{MC,ana}}) \cdot \mathrm{SF}_{\text{inef}}$ after applying the flavor tagging calibration. The sum of these fractions before calibration is unity, and the same holds for the sum of fractions after the calibration, i.e. the normalization for b-jets is preserved. Similarly, the normalization is preserved for c-jets and light-jets. Moreover, the analyses can use $\varepsilon_{\text{MC,ana}}$ as function of jet p_{T} and jet η in Eq. 4.4 which also ensures that the shape of the jet p_{T} vs jet η distribution remains the same after the flavor tagging calibration.

The scale factors $\mathrm{SF}_{\text{MC/MC}}$ from Eq. 4.3 can be deduced from Fig. 4.4. The Pythia6 MC generator was used for the calibration measurement for b-jets at $\sqrt{s} = 8$ TeV. In case, a user uses in his analysis the Herwig++ MC generator, the $\mathrm{SF}_{\text{MC/MC}}$ changes between $\sim$1.03 and $\sim$1.05 depending on the jet p_{T}. One can also observe that in case of using the EvtGen generator, the corresponding scale factors $\mathrm{SF}_{\text{MC/MC}}$ are closer to unity, see Fig. 4.5. This is the reason for using the EvtGen generator for the MC samples in the $\sqrt{s} = 13$ TeV analyses.

The scale factor in Eq. 4.1 has associated uncertainties resulting from the calibration measurement. These uncertainties are propagated to the final event weight in Eq. 4.5, and then also to a certain measurement. The uncertainties of the individual factors for the b-jets scale factor in Eq. 4.6 are fully correlated between each other. Similarly for the other flavors. Each flavor has its own calibration measurement, and therefore the uncertainties of the individual terms $\mathrm{SF}_{b\text{-jets}}$, $\mathrm{SF}_{c\text{-jets}}$, and $\mathrm{SF}_{\text{light-jets}}$ are not correlated.

5. Measurement of the Differential Cross Section of $t\bar{t}$ pair production

This section summarizes the measurement of the $p_{\mathrm{T}}^{\mathrm{top}}$ distribution of $t\bar{t}$ pair production at the ATLAS experiment at $\sqrt{s} = 8$ TeV. The measurement focuses on events from single lepton decay channel of the $t\bar{t}$ pair in the boosted topology. The single lepton channel is defined in Tab. 1.3. The $p_{\mathrm{T}}^{\mathrm{top}}$ distribution from hadronic top is measured. The results are unfolded to particle and parton level. Systematic uncertainties from all kind of sources are evaluated, and they are used to derive the covariance matrix to express the bin-to-bin correlations of statistical and systematic uncertainties. The measurement is compared to SM predictions using a statistical test.

The author of this thesis significantly contributed to this measurement which is published in [5]. The author worked in an analysis team of 20 members, and he disposed with full analysis chain and performed various studies. His main contributions are the evaluation of the PDF uncertainty, the evaluation of the bin-to-bin correlations of the statistical and all the systematic uncertainties, and the derivation of the NLO QCD prediction with all the theoretical uncertainties using the MCFM generator. The Fig. 5.1, 5.5 and 5.8 and the Tab. 5.4-5.10 has been produced by the author.

5.1 Signal Modeling and Background Estimation

MC simulation is used to model the $t\bar{t}$ events in the single lepton channel, which are referred to as $t\bar{t}$ single lepton events or signal events. The background estimation relies on several physics processes. These processes are estimated using MC simulation or using data-driven methods. The MC signal sample is used to characterize the detector response and efficiency of the event selection. The background estimate is subtracted from the measured data after event selection. The signal and background predictions are also needed to estimate the systematic uncertainties of the measurement. For each MC simulation, the top quark mass is set to $m^{\mathrm{top}} = 172.5$ GeV.

Several MC signal samples are used in this measurement and they are referred to as:

- Powheg+Pythia6 - sample generated with the NLO QCD fixed order generator Powheg-hvq patch4 [111, 112, 113] interfaced with parton shower and hadronization generator Pythia 6.425 using the Perugia 2011 tune [115] and the CT10 PDF set [100]. The parameter h_{damp}, which controls the emission with highest p_{T} in Powheg, is set to value $h_{\mathrm{damp}} = m^{\mathrm{top}}$ (more details about h_{damp} parameter can be found in [116]).

- POWHEG+PYTHIA6 $h_{\text{damp}} = \infty$ - same as the first POWHEG+PYTHIA6 sample except different h_{damp} parameter, which is set to $h_{\text{damp}} = \infty$.

- POWHEG+PYTHIA6+EWT - same as the POWHEG+PYTHIA6 sample with added electroweak corrections which are applied as a multiplicative factor to the event weight. These EWT corrections were extracted with cross section calculator HATHOR 2.1 [117], implementing the theoretical calculations from [33].

- POWHEG+HERWIG - sample generated with the NLO QCD fixed order generator POWHEG-hvq patch4 interfaced with the parton shower and hadronization generator HERWIG 6.520 [41] using the CT10 PDF set. The POWHEG h_{damp} parameter is set to $h_{\text{damp}} = \infty$. The generator JIMMY 4.31 [118] is used for the modeling of MPI.

- MC@NLO+HERWIG - sample generated with the NLO QCD fixed order generator MC@NLO 4.01 [76, 119] interfaced with HERWIG 6.520 with JIMMY 4.31, using the CT10 PDF set.

- ALPGEN+HERWIG - sample generated with the LO QCD fixed order generator ALPGEN 2.13 [120] with up to four additional partons. The ALPGEN generator is interfaced with HERWIG, for which the MLM matching scheme [121] is employed. The CTEQ6L1 [98] PDF set is used.

- ACERMC+PYTHIA6 - two samples generated with the LO QCD generator ACERMC 3.8 [122] interfaced with PYTHIA6. Two variations of the PYTHIA6 parton shower parameters are used to generate these two samples: one with increased and the other with decreased ISR and FSR.

All the signal samples are normalized to the NNLO+NNLL total cross section prediction (see Sec. 1.1.2), which for $m^{\text{top}} = 172.5$ GeV corresponds to value of 253 pb. The POWHEG+PYTHIA6+EWT signal sample is used for the event selection correction, unfolding correction and the data/MC comparison. This sample is referred to as the reference signal sample. The other signal samples are used to derive the signal modeling uncertainties or are used as predictions to compare with the final measured data.

There are several background processes, which can have similar signature as the $t\bar{t}$ events in the single lepton channel. They are referred to as:

- $t\bar{t}$ Dilepton - $t\bar{t}$ events from other decay channels than all-hadronic and single lepton according to the classification Tab. 1.3, i.e. the decay channel with two charged leptons, e or μ, and the decay channels containing one or two τ leptons. The $t\bar{t}$ Dilepton background is simulated using the same generators, EWT correction, and normalization as reference POWHEG+PYTHIA6+EWT signal sample.

- Single top - the t-channel is simulated using the NLO QCD generator ACERMC 3.8, the s-channel and the production of a top quark in association with a W boson are modeled with POWHEG [123, 124]. Both generators are interfaced with PYTHIA6 using the CTEQ6L1 PDF set and the Perugia 2011 tune. The three channels are normalized to the NLO+NNLL QCD total cross section predictions [125, 126, 127].

- W+jets and Z+jets - production of leptonically decaying vector boson, W or Z, in association with several jets. These backgrounds are modeled using generator ALPGEN 2.13, with which LO QCD and EWT vector boson production were simulated with up to five additional partons. The Z+jets sample contains the full Drell-Yan contribution also from gamma except the Z boson, and the interference between Z and gamma is taken into account. The generator PYTHIA6 is used for PS and hadronization. The normalization and heavy-flavor fraction of the W+jets background is determined from data by exploiting the expected charge asymmetry in the production of W^+ and W^- bosons at a pp collider, for more details see [128].

- Diboson - processes $pp \to WWX$, $pp \to WZX$, and $pp \to ZZX$, where X represents any other particles. The diboson background is modeled using generator Sherpa [129] with the CT10 PDF set. It is normalized to the NLO QCD total cross section prediction [130].

- Multijet - processes different from the processes listed above. This includes mainly multijet production and $t\bar{t}$ production in the all-hadronic decay channel, which is estimated using the data-driven matrix method described in [131].

All MC samples are scaled to the integrated luminosity, L, corresponding to the data used in this measurement. The integrated luminosity after the data quality criteria is estimated to $L = (20.28 \pm 0.57)$ fb^{-1} using techniques similar to those described in [59].

5.2 Event Selection

Thorough event selection is necessary to select the events of interest from all the pp collisions. The event selection exploits the $t\bar{t}$ single lepton channel signature in the boosted topology. Moreover, the events selection must take into account the detector acceptance. The object definition and event selection follow closely the ones used in the search for $t\bar{t}$ resonances in [132]. There are two sets of selection requirements: one focusing on $t\bar{t}$ events with one electron and the second focusing on $t\bar{t}$ events with one muon. These selection requirements are called $e+$jets and $\mu+$jets selections, respectively. The selection is applied at two stages: during data taking (on-line) and at the analysis stage (off-line).

The on-line event selection uses sophisticated trigger requirements to record as much as possible signal events since it is technically not possible to detect and record each pp collision. The $e+$jets selection contains trigger requiring any electron with $p_{\mathrm{T}} > 60$ GeV or an isolated electron with $p_{\mathrm{T}} > 24$ GeV [133]. The $\mu+$jets selection contains trigger requiring any muon with $p_{\mathrm{T}} > 36$ GeV or an isolated muon with $p_{\mathrm{T}} > 24$ GeV [134].

The off-line event selection contains requirements relying on the non-zero $E_{\mathrm{T}}^{\mathrm{miss}}$, the presence of one lepton, one boosted hadronic top quark and b-jets. The definitions of lepton candidates, jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$ are discussed in Sec. 2.3. To identify the hadronic top quark decay products, the LCW+JES anti-k_t $R = 1.0$ jets are used. The hadronic top is identified using jet substructure techniques, such as the trimming algorithm (see Sec. 3.1) and the jet shape variable splitting

scale, $\sqrt{d_{12}}$. The trimming algorithm uses parameters $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 0.05$. The LCW+JES anti-k_t $R = 0.4$ jets are used for finding the jet on the leptonic side. Suppression of pileup jets is done using the requirement JVT > 0.5 for anti-k_t $R = 0.4$ jets with $p_{\text{T}} < 50$ GeV and $|\eta| < 2.4$. A b-tagged jet is an LCW+JES anti-k_t $R = 0.4$ jet for which the MV1 weight is larger than the cut value $w^{\text{cut}} = 0.7892$ (for more details see Sec. 4.1). These are the detailed event requirements for $e +$ jets and $\mu +$ jets selections:

1. At least one primary vertex with five or more associated tracks with $p_{\text{T}} > 0.4$ GeV. The primary vertex with the highest $\sum p_{\text{T}}^2$ of the associated tracks is called as the signal vertex.

2. Exactly one lepton candidate. The $\mu +$ jets selection requires exactly one muon candidate with $p_{\text{T}} > 25$ GeV and $|\eta| < 2.5$. The $e +$ jets selection requires exactly one electron candidate with $p_{\text{T}} > 25$ GeV, $|\eta| < 2.47$, and $1.37 < |\eta| < 1.52$. The associated track to the lepton candidate must be close to the signal vertex, which is imposed by requiring $|z_0| < 2$ mm, where $|z_0|$ is the longitudinal impact parameter of the associated track (z-coordinate of the perigee point). The lepton candidates are required to be isolated to suppress background leptons originating from jets as described in [132]. To avoid double counting of leptons and jets, electron-overlap and muon-overlap removal is used as described in [132].

3. Missing transverse energy $E_{\text{T}}^{\text{miss}} > 20$ GeV and requirement for the W boson transverse mass $m_{\text{T}}^{W} > \left(60 \text{ GeV} - E_{\text{T}}^{\text{miss}}\right)$. The W boson transverse mass is defined as $m_{\text{T}}^{W} = \sqrt{2p_{\text{T}}^{\text{lepton}} E_{\text{T}}^{\text{miss}}(1 - \cos\Delta\phi)}$, where the $p_{\text{T}}^{\text{lepton}}$ is the p_{T} of the lepton, and $\Delta\phi$ is the azimuthal angle between the lepton and $E_{\text{T}}^{\text{miss}}$.

4. At least one anti-k_t $R = 0.4$ jet with $p_{\text{T}} > 25$ GeV within distance $\Delta R < 1.5$ from the lepton. The highest p_{T} jet fulfilling these conditions is referred to as the *leptonic jet*. The separation condition $\Delta R < 1.5$ is based on the fact that the decay products from a leptonic top are more collimated.

5. At least one anti-k_t $R = 1.0$ trimmed jet with $p_{\text{T}} > 300$ GeV, $|\eta| < 2$, $m > 100$ GeV, and $\sqrt{d_{12}} > 40$ GeV. The highest p_{T} jet fulfilling these conditions is referred to as the *hadronic top-jet candidate*.

6. Separation between hadronic top-jet candidate and the decay products from the leptonic top in the boosted regime. This is imposed by requiring distance $\Delta R > 1.5$ between the hadronic top-jet candidate and the leptonic jet and by requiring distance $\Delta\phi > 2.3$ between the lepton and hadronic top-jet candidate.

7. At least one b-tagged jet with $p_{\text{T}} > 25$ GeV and $|\eta| < 2.5$ fulfilling one of the two following conditions. It must be the same as the leptonic jet from Point 4 or another jet which is within distance $\Delta R < 1$ from the hadronic top-jet candidate.

5.3 Detector Level MC/Data Comparison

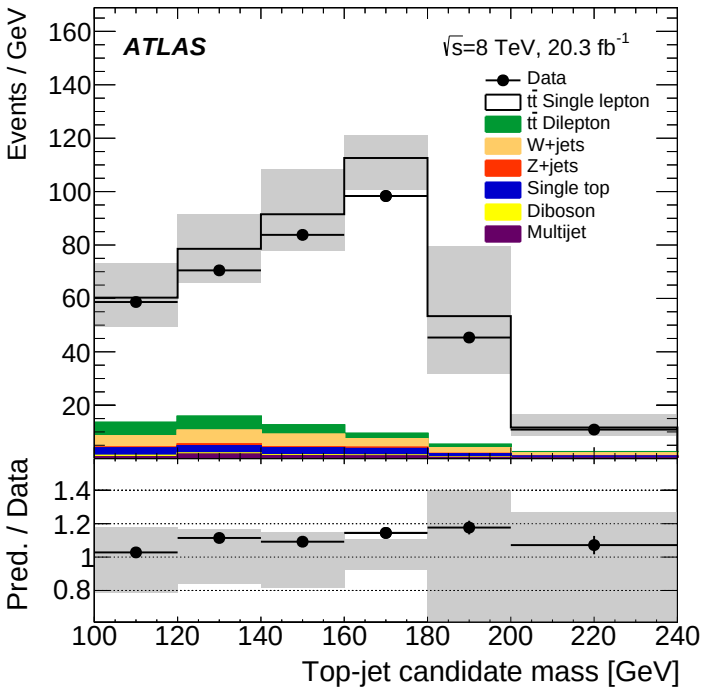
The data and the prediction are compared at the detector level after the event selection. The Tab. 5.1 gives the number of events for each background process and signal in both selection channels. Their sum gives the prediction, which is compared to the observed number of events in data. The uncertainties include the uncertainties on the background estimates, objects' energy scales and reconstruction efficiencies, and MC statistics. The observed yields in data are compatible with the prediction within uncertainties.

The $e+$ jets and $\mu+$ jets selections are combined into a common sample of events. The final unfolded results were obtained from this combination of events. As a cross-check, the extraction of final unfolded results were done individually in each selection channel, $e+$ jets and $\mu+$ jets. The final results are found to be consistent between $e+$ jets and $\mu+$ jets selections, and also with the final result obtained using the combination of $e+$jets and $\mu+$jets channels. In the following, only the combination of the two selection channels is used. The combination of the two selections gives ∼7700 observed data events, from which ∼85% are signal events according to the prediction.

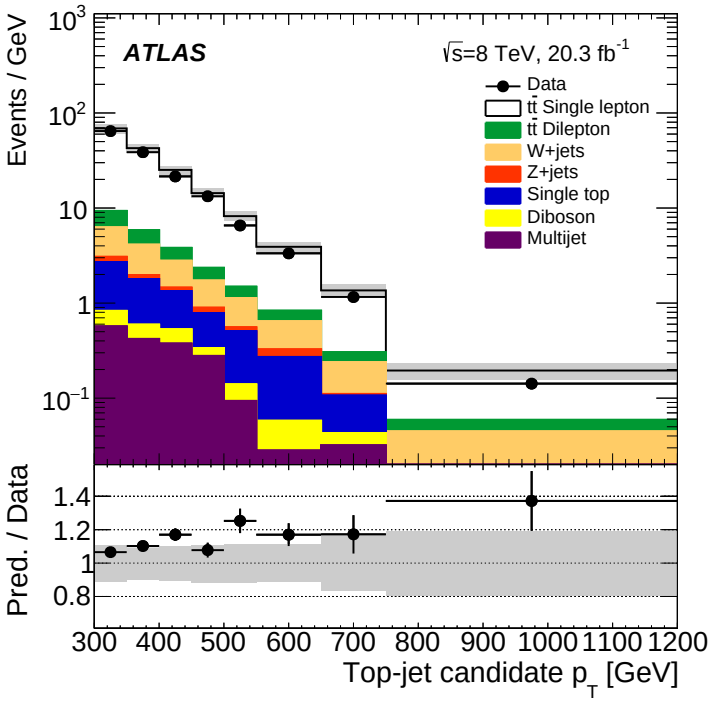
	e+jets	μ+jets
$t\bar{t}$ Single lepton	3880 ± 430	3420 ± 380
$t\bar{t}$ Dilepton	199 ± 27	169 ± 24
W+jets	235 ± 54	226 ± 50
Single top	133 ± 22	134 ± 29
Multijet	91 ± 17	3 ± 1
Z+jets	34 ± 18	14 ± 8
Dibosons	22 ± 12	18 ± 10
Prediction	4600 ± 470	3980 ± 410
Data	4145	3603

Table 5.1: Observed and expected number of events for the $e+$ jets and $\mu+$ jets selections. The systematic uncertainties include the background estimation techniques, objects' energy scales and reconstruction efficiencies, and MC statistics. [5]

The Fig. 5.1 shows the comparison of observed distributions of hadronic top-jet candidate mass and p_{T} with the prediction. The systematic uncertainties on the objects' energy scales and reconstruction efficiencies, on the background estimates, luminosity and MC statistics are shown for the prediction. The statistical uncertainty is shown for the observed distributions for data. The mass distribution is peaked around of the top quark mass as expected. The observed distributions with data are consistent with the predictions within uncertainties, which are summarized in Sec. 5.7. The distribution of the hadronic top-jet candidate p_{T} is the input to the unfolding procedure.



(a)



(b)

Figure 5.1: Distributions of (a) mass and (b) p_{T} of the hadronic top-jet candidate. The $t\bar{t}$ prediction is obtained using the nominal POWHEG+PYTHIA6 sample. The ratio of the MC prediction to the data is shown in the insets below the histograms. The hashed area includes all the object-related uncertainties (on the jet, lepton, and $E_{\mathrm{T}}^{\mathrm{miss}}$), and the uncertainties from the background estimation, luminosity, and MC statistics. The vertical lines indicate the data statistical uncertainty. [5]

5.4 Particle Level Definition of the $t\bar{t}$ Single Lepton Events

The usefulness of measurements at particle level is discussed in Sec. 1.2.2. The definition of the particle level used for this measurement is provided in the Rivet Routine in [135], and it is outlined in this section. It consists of the definition of physics objects at particle level and a fiducial phase space. The definition of physics objects at particle level is based on stable particles detectable by the detector. In this measurement, the fiducial region is defined by very similar particle level selection criteria as were used for the detector level selection. Such a choice may be advantageous to minimize the uncertainties of the efficiencies of detector and particle level selections, which are used for the unfolding procedure from the detector level to the particle level.

The definition of physics objects at particle level is outlined in the following. Electron and muon at particle level are defined as an electron and muon originating not from hadron decays. The electrons and muons are *dressed*, which means that the 4-momenta of photons within a cone of $\Delta R = 0.1$ around the electron or muon direction are added to those of the leptons. The anti-k_t $R = 0.4$ and anti-k_t $R = 1.0$ jet algorithms are used to cluster jets followed by the same trimming procedure as was applied at detector level. The jets are clustered from all stable particles with mean lifetime greater than 1 cm except neutrinos and dressed electrons and muons. Also all B-hadrons are added to the set of input particles to the clustering algorithm. The p_{T} of these B-hadrons is changed to be infinitesimally small, i.e. the presence of the B-hadrons does not change the clustering algorithm. This addition of infinitesimally small B-hadrons is useful for the identification of b-jets. An anti-k_t $R = 0.4$ jet is considered as b-tagged at particle level if there is at least one B-hadron with $p_{\mathrm{T}} > 5$ GeV clustered in this jet. The $E_{\mathrm{T}}^{\mathrm{miss}}$ at particle level is defined as the magnitude of the vector sum of the p_{T} of neutrinos not resulting from hadron decays.

The fiducial region for particle level is defined using physics objects at the particle level as follows. The requirement for leptons at particle level is simplified with respect to the detector level. Exactly one lepton, electron or muon, with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ is required. Further, exactly the same event selection criteria as for the detector level are used as it is summarized in Points 3–7 in Sec. 5.2.

5.5 Result at Particle Level

The Fig. 5.1b shows the top-jet candidate p_{T} at detector level for data after detector level selection criteria expressed as the number of observed events, D_i, in bin i. This distribution is corrected for background contribution, unfolded to the particle level defined in Sec. 5.4, and corrected for the event selection criteria. The same binning is used for the top-jet candidate p_{T} at particle level, $p_{\mathrm{T}}^{\mathrm{ptcl}}$, as for the detector level. It contains eight non-equidistant bins from 300 GeV up to 1200 GeV.

The background is subtracted from the data in two steps. In the first step, all background except $t\bar{t}$ dilepton are subtracted. The sum of these backgrounds,

B, is defined as the predicted Single top, W+jets, Z+jets, Diboson, and Multijet backgrounds from Sec. 5.1. The resulting distribution should contain only the $t\bar{t}$ events (in single lepton and dilepton channels). In the second step, the $t\bar{t}$ dilepton background is subtracted using a multiplication factor f^{Dilepton} defined as the ratio between two distributions: the $t\bar{t}$ single lepton prediction and the sum of the $t\bar{t}$ dilepton and $t\bar{t}$ single lepton predictions.

After the background subtraction, the detector effects are removed and a correction is performed due to differences between the detector and particle event selection criteria. The detector effects are removed by unfolding from detector level to particle level using migration matrix in Fig. 5.2. The migration matrix is constructed from MC reference signal sample using events which passed both, detector and particle, level selection criteria. To be able to perform the unfolding on data after background subtraction, they must be corrected for the missing particle level selection criteria using a multiplication factor $\varepsilon(\text{ptcl}|\text{det})$. It is defined as the efficiency of particle level selection on the sample of events after detector level selection[1]. This correction factor is obtained from MC reference signal sample and it is shown in Fig. 5.3a. The unfolding is performed using the RooUnfold package [136], from which the singular value decomposition (SVD) [137] method is used. The unfolding regularization parameter of the SVD method is chosen to be $k = 6$. After unfolding to particle level, the distribution is still influenced by the detector level selection criteria and it is necessary to correct for this effect. This is done by dividing the unfolded distribution by factor $\varepsilon_i(\text{det}|\text{ptcl})$, which is the efficiency of the detector level selection criteria in the sample of events after particle level selection criteria. After this step, the obtained distribution is only after particle level selection criteria as a function of $p_{\text{T}}^{\text{ptcl}}$. It is divided by the integrated luminosity and the bin width, $\Delta p_{\text{T}}^{\text{ptcl}}$, in order to obtain the differential cross section.

Overall, the extraction of the final result at particle level from the detector level distribution can be expressed by the equation

$$\mathrm{d}\sigma/\mathrm{d}p_{\text{T}}^{\text{ptcl}}\left(p_{\text{T}i}^{\text{ptcl}}\right) = \frac{D_i^{\text{ptcl}}}{\Delta p_{\text{T}i}^{\text{ptcl}} \cdot L} = \frac{\left(\mathcal{U}\left[I^{\text{ptcl}}\right]\right)_i}{\Delta p_{\text{T}i}^{\text{ptcl}} \cdot L \cdot \varepsilon_i(\text{det}|\text{ptcl})} \tag{5.1}$$

where the symbol $\mathcal{U}\left[I^{\text{ptcl}}\right]$ assigns the unfolded distribution obtained from the distribution

$$I_i^{\text{ptcl}} = \varepsilon_i(\text{ptcl}|\text{det}) \cdot f_i^{\text{Dilepton}} \cdot (D_i - B_i)\,, \tag{5.2}$$

which is corrected in bin-by-bin correction for background contribution and for the missing particle level selection criteria. The index i represents the bin number of a given histogram in both equations.

The final particle level results for the $t\bar{t}$ production in the single lepton decay channel are summarized in Tab. 5.2. The uncertainty evaluation is explained in Sec. 5.7.

[1] Another particle level selection criterion is added for the evaluation of this efficiency: top-jet candidate $p_{\text{T}} < 1200$ GeV. This is important since the migration matrix does not go above 1200 GeV.

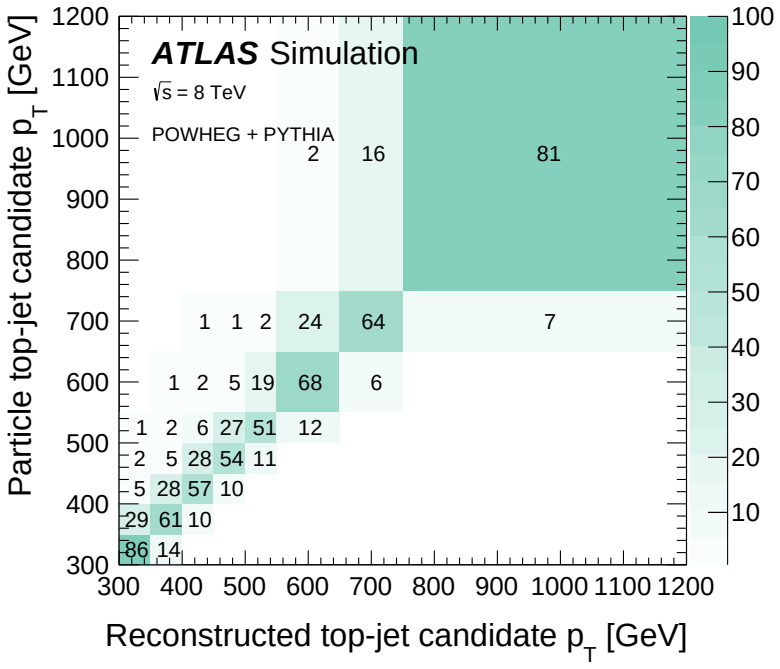


Figure 5.2: Migration matrix between the particle and detector level top-jet candidate p_{T}. The unit of the matrix elements is the probability (expressed in percentage) for an event generated at a given particle level value to be reconstructed at certain detector level value (each row adds up to 100%). [5]

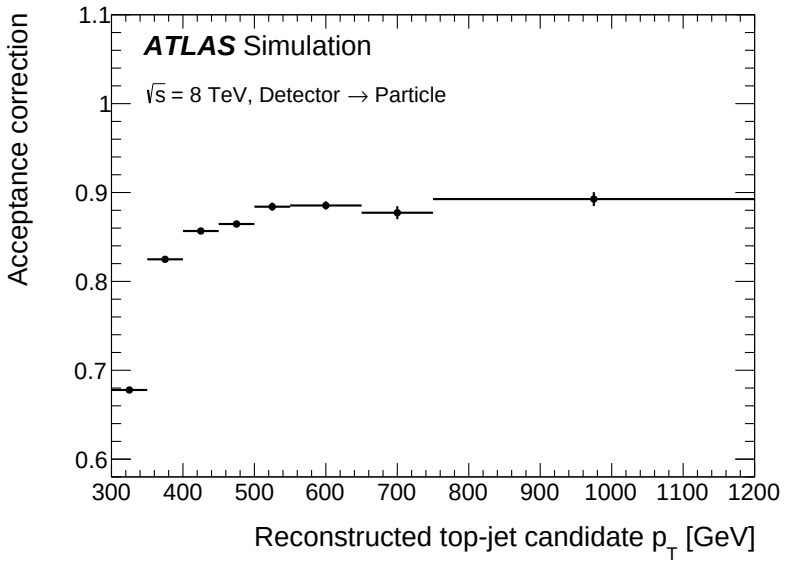
$p_{\mathrm{T}}^{\mathrm{ptcl}}$ [GeV]	$\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{\mathrm{ptcl}}$ [fb/GeV]	Stat. [%]	Syst. [%]
300 – 350	4.97	±2.7	±15
350 – 400	3.09	±3.5	±13
400 – 450	1.73	±4.2	±13
450 – 500	1.08	±4.4	±14
500 – 550	0.56	±6.1	±14
550 – 650	0.27	±6.0	±16
650 – 750	0.097	±8.1	±20
750 – 1200	0.012	±15	±24

Table 5.2: Particle level $p_{\mathrm{T}}^{\mathrm{ptcl}}$ distribution with statistical and systematic uncertainties. [5]

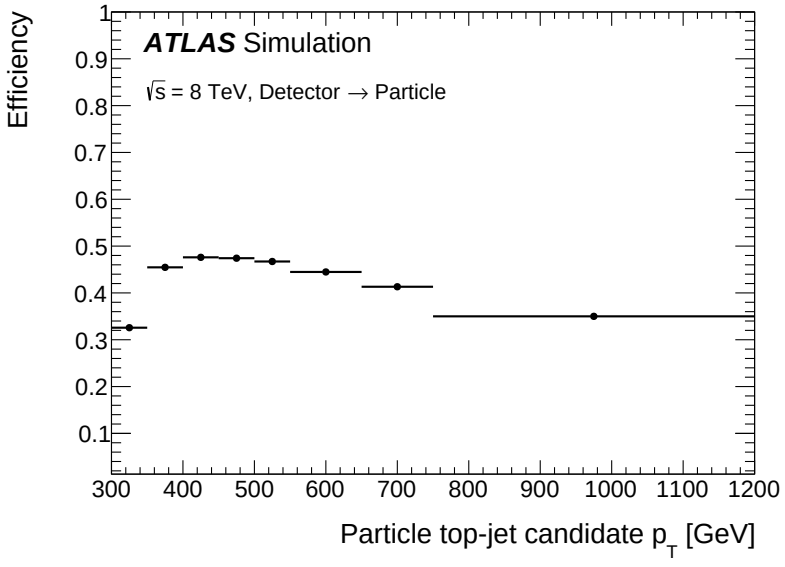
5.6 Result at Parton Level

Besides the particle level result, the measurement is unfolded also to the parton level for variable p_{T} of the hadronically decaying top quark, $p_{\mathrm{T}}^{\mathrm{parton}}$. The top quark at parton level is defined as the top quark right before its decay to W and b-quark. This observable can be obtained theoretically in a fixed-order calculation or using MC simulation of fixed-order prediction interfaced with parton shower. The distribution at parton level has the same binning as the detector level distribution, which means eight non-equidistant bins from 300 GeV up to 1200 GeV.

There are two possibilities how to obtain the result at parton level. The



(a) ε(ptcl|det)



(b) ε(det|ptcl)

Figure 5.3: (a) Correction factor ε(ptcl|det) as a function of top-jet candidate p_{T} at detector level. It represents the ratio of the number of events after detector and particle level selection requirements to the number of events after detector level selection requirements. (b) Correction factor ε(det|ptcl) as a function of top-jet candidate p_{T} at particle level. It represents the ratio of the number of events after detector and particle level selection requirements to the number of events after particle level selection requirements. [5]

particle level result of Sec. 5.5 can be unfolded to the parton level or the detector level distribution can be directly unfolded to the parton level. Both approaches were performed in this measurement, and they are consistent. Only the unfolding from particle to parton level is considered in the following. The parton level result

is obtained with similar equations as the particle level result

$$\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{\mathrm{parton}}\left(p_{\mathrm{T}i}^{\mathrm{parton}}\right) = \frac{\left(\mathcal{U}\left[I^{\mathrm{parton}}\right]\right)_i}{B \cdot \Delta p_{\mathrm{T}i}^{\mathrm{parton}} \cdot L \cdot \varepsilon_i(\mathrm{ptcl}|\mathrm{parton})}, \tag{5.3}$$

$$I_i^{\mathrm{parton}} = \varepsilon_i(\mathrm{parton}|\mathrm{ptcl}) \cdot D_i^{\mathrm{ptcl}}. \tag{5.4}$$

The index i represents the bin number of a given histogram in both equations. The result extraction starts with the particle level spectrum, D^{ptcl}, defined in Eq. 5.1. The spectrum is corrected with factor $\varepsilon(\mathrm{parton}|\mathrm{ptcl})$ which is defined as the efficiency of parton selection criteria $p_{\mathrm{T}}^{\mathrm{parton}} > 300$ GeV and $p_{\mathrm{T}}^{\mathrm{parton}} < 1200$ GeV in the sample of events after particle level selection criteria. In this way, the obtained distribution can be unfolded using migration matrix in the range $p_{\mathrm{T}}^{\mathrm{parton}} \in [300, 1200]$ GeV shown in Fig. 5.4. The same unfolding method is used as for the unfolding from detector to particle level: SVD method with regularization parameter $k = 6$. The unfolded spectrum denoted as $\mathcal{U}\left[I^{\mathrm{parton}}\right]$ is corrected for the particle level selection criteria using correction factor $\varepsilon(\mathrm{ptcl}|\mathrm{parton})$ defined as the efficiency of particle level selection criteria in the sample of events without any selection criteria. The migration matrix and the efficiencies $\varepsilon(\mathrm{parton}|\mathrm{ptcl})$ and $\varepsilon(\mathrm{ptcl}|\mathrm{parton})$ are derived using the MC reference signal sample defined in Sec. 5.1. The result is divided by the luminosity and bin width $\Delta p_{\mathrm{T}}^{\mathrm{parton}}$ to get the differential cross section that is further corrected by the branching ratio, B, of the $t\bar{t}$ single lepton channel. In this way, the obtained result for the $t\bar{t}$ production is independent of the W boson decays from the $t\bar{t}$ pair.

The final parton level results for the $t\bar{t}$ production are summarized in Tab. 5.3. The uncertainty evaluation is explained in Sec. 5.7.

$p_{\mathrm{T}}^{\mathrm{parton}}$ [GeV]	$\mathrm{d}\sigma/\mathrm{d}p_{\mathrm{T}}^{\mathrm{parton}}$ [fb/GeV]	Stat. [%]	Syst. [%]
300 – 350	60.1	±3.2	±16
350 – 400	26.2	±3.4	±15
400 – 450	11.8	±4.2	±20
450 – 500	6.27	±4.5	±21
500 – 550	3.06	±6.1	±27
550 – 650	1.21	±6.3	±26
650 – 750	0.375	±9.6	±31
750 – 1200	0.043	±17	±38

Table 5.3: Parton level $p_{\mathrm{T}}^{\mathrm{parton}}$ distribution with statistical and systematic uncertainties. [5]

5.7 Uncertainties

The measurement is influenced by both statistical and systematic uncertainties. All the uncertainties are propagated from the detector level through the unfolding procedure to the final results at particle and parton levels. Moreover, the ingredients used to extract the final results (background estimate, migration matrix, efficiencies) suffer from additional systematic uncertainties which should be taken into account as well.

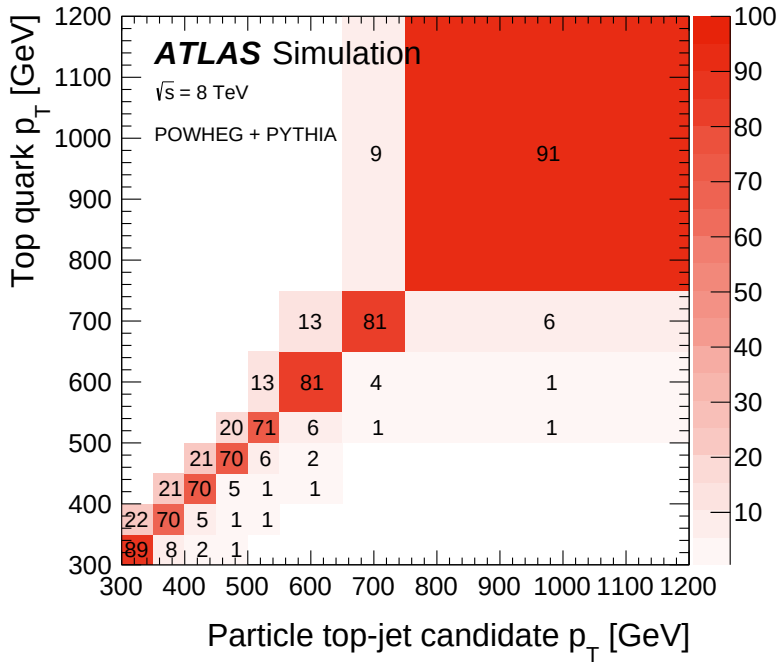


Figure 5.4: Migration matrix between the parton level top quark p_{T} and particle level top-jet candidate p_{T}. The unit of the matrix elements is the probability (expressed in percentage) for an event generated at a given parton level value to be reconstructed at certain particle level value (each row adds up to 100%). [5]

5.7.1 Statistical Uncertainty

The statistical uncertainty of the measurement originates from the statistical uncertainty of the D distribution from Eq. 5.2. This distribution contains 3227 events in the first bin and lower number of events in the next bins down to 64 events in the last bin, which corresponds to relative statistical uncertainties from ∼1.8% to ∼12.5%. The uncertainty of this distribution is propagated through Eq. 5.1–5.4 to the final results at particle and parton level.

To obtain the final statistical uncertainty, an ensemble of $2 \cdot 10^5$ *pseudo-experiments* is prepared. Each pseudo-experiment is constructed from the D distribution as follows. A random number from the Poisson distribution with mean D_i is assigned to the bin i of a pseudo-experiment. The Poisson distribution in a certain bin is independent from the Poisson distribution in another bin. The extraction of the result is performed on each pseudo-experiment individually using the same procedure as for the main result. The statistical uncertainty of the particle level measurement is the RMS of the obtained particle level results for the ensemble of pseudo-experiment evaluated bin-by-bin. The same holds for the parton level. The relative statistical uncertainties are shown in Tab. 5.2-5.3. These statistical uncertainties are greater than the statistical uncertainties of the input distribution at detector level. The reason for this is the background subtraction using distribution B and the unfolding procedure. Moreover, the unfolding introduces correlation of the statistical uncertainties between bins. The corresponding correlation matrix is evaluated using the sample of pseudo-experiments. The obtained correlation matrices at particle and parton level are shown in Tab. 5.4-5.5.

bin	1	2	3	4	5	6	7	8
1	1.00	-0.38	-0.11	0.08	0.02	-0.01	-0.01	0.00
2	-0.38	1.00	-0.09	-0.23	0.00	0.04	0.01	-0.00
3	-0.11	-0.09	1.00	0.09	-0.28	-0.05	0.04	0.02
4	0.08	-0.23	0.09	1.00	0.33	-0.20	-0.10	0.01
5	0.02	0.00	-0.28	0.33	1.00	0.31	-0.16	-0.14
6	-0.01	0.04	-0.05	-0.20	0.31	1.00	0.39	-0.04
7	-0.01	0.01	0.04	-0.10	-0.16	0.39	1.00	0.68
8	0.00	-0.00	0.02	0.01	-0.14	-0.04	0.68	1.00

Table 5.4: Correlation matrix between the bins of the particle level differential cross section as a function of $p_{\mathrm{T}}^{\mathrm{ptcl}}$ assuming only statistical uncertainties. [5]

bin	1	2	3	4	5	6	7	8
1	1.00	-0.44	-0.27	0.09	0.07	-0.01	-0.02	-0.01
2	-0.44	1.00	0.08	-0.34	-0.10	0.05	0.03	0.00
3	-0.27	0.08	1.00	0.23	-0.33	-0.19	0.02	0.04
4	0.09	-0.34	0.23	1.00	0.47	-0.17	-0.20	-0.09
5	0.07	-0.10	-0.33	0.47	1.00	0.47	-0.17	-0.27
6	-0.01	0.05	-0.19	-0.17	0.47	1.00	0.49	0.12
7	-0.02	0.03	0.02	-0.20	-0.17	0.49	1.00	0.89
8	-0.01	0.00	0.04	-0.09	-0.27	0.12	0.89	1.00

Table 5.5: Correlation matrix between the bins of the parton level differential cross section as a function of $p_{\mathrm{T}}^{\mathrm{parton}}$ assuming only statistical uncertainties. [5]

The uncertainty in the first two bins are anti-correlated for both results, while the correlation factor increases for the uncertainties between next neighboring bins. The correlations between distant bin pairs are negligible.

5.7.2 Systematic Uncertainties

The total systematic uncertainty of the measurement consists of various systematic uncertainty components which are thoroughly estimated. These uncertainties originate from imperfect detector modeling, physics modeling, luminosity measurement, or limited statistics used to derive various corrections. They affect directly the input data distribution D at detector level or the individual ingredients used to extract the particle and parton level results in Eq. 5.1–5.4. The individual components of systematic uncertainty and the evaluation of their effect on the final result are described in this section.

5.7.2.1 Object Modeling Uncertainties

The MC modeling of physics objects is associated with uncertainties for various MC correction factors as described in Sec. 2.4.1. In this measurement, uncertainties from detection of several objects contribute, such as electrons, muons, anti-k_t $R = 0.4$ jets, anti-k_t $R = 1.0$ jets, and $E_{\mathrm{T}}^{\mathrm{miss}}$.

The detection of electrons and muons are influenced each by five uncertainty components: identification efficiency, reconstruction efficiency, trigger efficiency,

energy scale, and energy resolution, which are in detail described in [62, 138, 63].

The anti-k_t $R = 0.4$ jets uncertainties have in total 25 components: jet reconstruction efficiency, JER, and 23 JES components such as the *in situ* calibration uncertainties, pileup uncertainty, flavor composition and flavor response uncertainty. Detailed overview of the anti-k_t $R = 0.4$ uncertainties evaluation can be found in [67], and the total systematic uncertainty for $\sqrt{s} = 8$ TeV can be found in [139].

The b-tagging of anti-k_t $R = 0.4$ jets is associated with three uncertainty components: b-jets, c-jets, and light-jets as it is explained in Sec. 4.3. Additional extrapolation uncertainty is added for b-jets in range $[300, 1200]$ GeV, which ranges from 10% to 30%.

The anti-k_t $R = 1.0$ jets uncertainties have in total 20 components: JER, jet mass scale (JMS), jet mass resolution (JMR), $\sqrt{d_{12}}$ scale, and 16 JES components from which the three main components are called JES topology, JES Data/MC, and JES statistics. The JES topology uncertainty reflects the fact that the JES may differ between jets originating from a quark or a gluon and jets originating from a vector boson or a top quark. The next two significant uncertainties originate from the *in situ* calibration uncertainties using p_{T} balance in gamma+jet events. The components JES Data/MC and JES statistics are, respectively, the systematic and statistical uncertainties associated with this gamma+jet calibration. The JMS, JMR, and $\sqrt{d_{12}}$ scale uncertainties are evaluated using comparison of calorimeter jets and track jets between data and MC simulation as described in [140]. The total systematic uncertainty for $\sqrt{s} = 8$ TeV can be found in [141]. It varies between 2% and 4%.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ has uncertainties associated with energy scale of physics objects used to calculate the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the $E_{\mathrm{T}}^{\mathrm{miss}}$ Soft Term as it is described in [71].

Propagation to the Final Result

The above uncertainties can be classified in three categories as it is discussed in Sec. 2.4.1: efficiency, 4-momentum, and resolution uncertainties. The uncertainties for lepton reconstruction, identification and trigger efficiencies, jet reconstruction efficiency and b-tagging belong to the efficiency uncertainties category. The uncertainty components JES, JMS, electron, muon and $E_{\mathrm{T}}^{\mathrm{miss}}$ energy scales belongs to the 4-momentum uncertainties. The resolution uncertainties are JER, JMR, muon and electron energy resolution uncertainties. The propagation of the resolution uncertainties to the final result is done in slightly different way than the propagation of the efficiency and 4-momentum uncertainties. First, the propagation of the efficiency and 4-momentum uncertainties is outlined.

At the beginning, it is found how the individual uncertainty components alter the hadronic top-jet candidate prediction at detector level from Fig. 5.1b, which is derived using the central values of the MC correction factors. This distribution is referred to as the nominal distribution, P^{nom}. For the efficiency and 4-momentum uncertainties, the lower and upper bounds of the uncertainties associated with the MC correction factors are derived in calibration measurements. These bounds corresponds to one standard deviation from the central value. For each uncertainty component k, the corresponding correction factor is shifted up and down by one standard deviation, after which the up and down predicted detector level top-jet candidate p_{T} is obtained, $P^{k,\mathrm{up}}$ and $P^{k,\mathrm{down}}$, respectively.

The shifted predictions are used to derive the relative up and down shifts, $s^{k,\mathrm{up}}$ and $s^{k,\mathrm{down}}$, of the uncertainty component k. They are defined in bin i as

$$s_i^{k,\mathrm{up}} = \frac{P_i^{k,\mathrm{up}} - P_i^{\mathrm{nom}}}{P_i^{\mathrm{nom}}}, \tag{5.5}$$

$$s_i^{k,\mathrm{down}} = \frac{P_i^{k,\mathrm{down}} - P_i^{\mathrm{nom}}}{P_i^{\mathrm{nom}}}. \tag{5.6}$$

These shifts are used to derive the lower and upper bounds, $\epsilon P^{k,+}$ and $\epsilon P^{k,+}$, of the relative systematic uncertainty component k of the top-jet candidate p_{T} distribution at detector level before background subtraction. They are defined as

$$\epsilon P_i^{k,\pm} = \max\left(0, \pm s_i^{k,\mathrm{up}}, \pm s_i^{k,\mathrm{down}}\right). \tag{5.7}$$

The relative uncertainties at particle and parton level are estimated in similar way after the individual uncertainty components are propagated further to the final result using the predictions P^{nom}, $P^{k,\mathrm{up}}$ and $P^{k,\mathrm{down}}$ instead of D in Eq. 5.1–5.4 leading to final results $\sigma\left(P^{\mathrm{nom}}\right)$, $\sigma\left(P^{k,\mathrm{up}}\right)$, and $\sigma\left(P^{k,\mathrm{down}}\right)$, respectively (not differentiating between particle and parton level). These results are used to derive the lower and upper bounds, $\epsilon\sigma^{k,+}$ and $\epsilon\sigma^{k,-}$, of the relative uncertainties of the final result

$$\epsilon\sigma_i^{k,\pm} = \max\left(0, \pm\frac{\sigma_i\left(P^{k,\mathrm{up}}\right) - \sigma_i\left(P^{\mathrm{nom}}\right)}{\sigma_i\left(P^{\mathrm{nom}}\right)}, \pm\frac{\sigma_i\left(P^{k,\mathrm{down}}\right) - \sigma_i\left(P^{\mathrm{nom}}\right)}{\sigma_i\left(P^{\mathrm{nom}}\right)}\right). \tag{5.8}$$

As mentioned in Sec. 5.1, data-driven approaches are used to estimate the normalization and heavy-flavor content for $W+$jets background prediction. These approaches are also associated with object modeling uncertainties. The derivation of $P^{k,\mathrm{up}}$ and $P^{k,\mathrm{down}}$ is correlated with this effect for the most dominant uncertainty components.

The propagation of the third category of uncertainties (resolution uncertainties) is performed differently due to the different application of a resolution correction. For example, the JES or b-tagging correction applies a multiplicative factor to scale the p_{T} or the event weight, while the muon resolution correction is performed by randomly smearing the energy of muons in MC to get in total the same muon energy resolution in MC as in data. For the propagation of the resolution uncertainty k, the up shift $P^{k,\mathrm{up}}$ is obtained by applying random smearing of the corresponding variable, which leads to degradation of the total resolution by one standard deviation. The opposite process (resolution improvement) is not possible in practice. Therefore, the down shift is not available, and to account for possible over-smearing, symmetrization is done as explained in the following. The lower and upper bounds for uncertainties are derived using Eq. 5.7 and 5.8 without the down shift term. After this procedure, one of the two bounds is equal to zero, and the other is non-zero. The bound, which is equal to zero, is changed to the opposite value of the non-zero bound.

Besides the approach described above, there are other possibilities how to propagate the object modeling uncertainties to the final results. These uncertainties directly affect the ingredients used to extract the final results in Eq. 5.1–5.4 (background estimates, selection efficiencies, migration matrices), and therefore

one can shift the MC correction factors by one standard deviation up or down when evaluating these ingredients. The difference between the final result obtained using the non-shifted and shifted ingredients can be used to define the uncertainties of the final result. In this approach, there are two options for the input detector level distribution: the data or the prediction. Both options should in principle result in similar relative uncertainties, although using data as input may be more correct since data are used for the extraction of the final results. This approach is different from the approach used in this measurement. However, both approaches should lead to similar uncertainties, which is given by the fact that the uncertainties on the MC objects can be transfered to the data objects in case the lower and upper bounds are symmetric (which is the case for most of the object uncertainty components). For example, the lower bound of the JES uncertainty in MC can be eliminated and assigned as the upper bound of the JES uncertainty in data. This is not possible to perform for all the object uncertainties in practice, e.g. the b-tagging uncertainties cannot be transferred from MC to data since they are evaluated based on the truth jet flavor in MC. And therefore, it is not possible to perform the shifts in data, and MC is used instead of data as it is outlined in this section. The cross-check of the equivalence of the two approaches was not performed for this measurement.

5.7.2.2 Background Modeling Uncertainties

The estimate for each background process is associated with various uncertainties. They can affect only the overall normalization or also the shape of the top-jet candidate p_{T} distribution.

The shape of the W+jets background is estimated from varying the renormalization and factorization scales and the matching parameter between the fixed-order calculation and the parton shower.

The uncertainty of the multijet background originates from the inputs to the data-driven estimation. It is found by varying the definition of loose leptons, changing the selection used to form the control region, propagating the statistical uncertainty of the efficiencies, and propagating object modeling uncertainties like JES. The resulting uncertainties on the normalization are 19.4% in the $e+$jets selection and 18.9% in the $\mu+$jets selection.

The normalization of the three single top background samples is associated with uncertainty from theory calculations [125, 126, 127]. Two additional uncertainties are assumed for the shape of the single top associated production with a W boson. The first uncertainty is associated with the used MC generator and the second uncertainty with the interference with $t\bar{t}$ process as explained in [142].

The diboson and $Z+$jets backgrounds are less dominant, and no detailed studies are performed to obtain their uncertainties. Therefore, a conservative uncertainty of 50% is applied to the normalization of these backgrounds.

All the MC background estimates are influenced by the luminosity uncertainty, which is discussed in Sec. 5.7.2.4. All the MC background estimates are influenced by limited number of generated MC events, which is treated as another systematic component called MC background statistical uncertainty.

Propagation to the Final Result

The background modeling uncertainty is associated with the B distribution in Eq. 5.2. One option, how to propagate a certain background uncertainty component, is to directly shift this distribution and derive the uncertainty of the final results from the difference of the non-shifted and shifted background estimates. Different approach is used similarly as for the object modeling uncertainties in Sec. 5.7.2.1.

5.7.2.3 MC Signal Modeling Uncertainties

The MC signal modeling affects the migration matrices and efficiencies in Eq. 5.1–5.4, and therefore uncertainties are associated to it. There are several components of the MC signal modeling uncertainties: PDF, $t\bar{t}$ generator, Parton Shower (PS) and Hadronization, ISR and FSR.

PDF

The PDF related uncertainty of this measurement is derived using the PDF4LHC recommendations [143]. According to them, PDF fits from three PDF groups are used: CT10, MSTW2008NLO [78], and NNPDF23_nlo [144]. Each PDF fit contains one central PDF set and several uncertainty PDF sets, which are used to obtain the intra-PDF uncertainty by following the PDF group's prescription. Using the intra-PDF uncertainties and the central PDF values, the PDF uncertainty is evaluated as an envelope. The PDF uncertainty is obtained using PDF sets derived for the strong coupling constant $\alpha_{\mathrm{S}} = 0.118$ for renormalization scale equal to the mass of the Z boson.

The MC@NLO+Herwig sample is used which is generated using the central CT10 PDF set. The samples for other PDF sets are obtained from this sample using an event reweighting scheme. The migration matrices and the efficiency corrections are obtained for each central and uncertainty PDF set. These varied unfolding inputs are used to derive the final results with Eq. 5.1–5.4 using the detector level distribution from the central CT10 PDF set as the input instead of the D distribution, and not applying background subtraction.

The PDF uncertainty of the particle level result ranges from 0.2% to 1.2% increasing with $p_{\mathrm{T}}^{\mathrm{ptcl}}$. The PDF uncertainty of the parton level result ranges from 0.3% to 2.2%. The upper bounds of the intra-PDF uncertainties and the PDF uncertainty of the parton level result are shown in Fig. 5.5. The lower bounds are identical or slightly different with the upper bounds (only the MSTW intra-PDF prescription can lead to non-symmetric uncertainties). The intra-PDF uncertainty for CT10 PDF fit is the highest among the three PDF fits. Using an envelope on the central values with uncertainties, the final PDF uncertainty is identical with the CT10 intra-PDF uncertainty.

$t\bar{t}$ Generator

The uncertainty related to the $t\bar{t}$ generator is estimated using the two signal samples Powheg+Herwig and MC@NLO+Herwig. The migration matrices and efficiency corrections are used from the sample Powheg+Herwig to obtain the unfolded final results from the input detector level distribution from

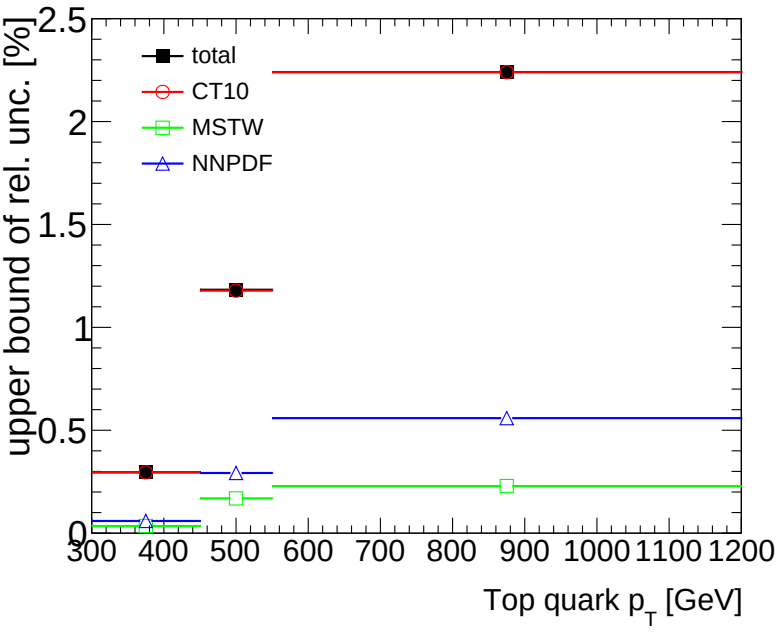


Figure 5.5: Upper bound for the intra-PDF uncertainties of the parton level result for the three PDF fits: CT10, MSTW2008NLO, and NNPDF23_nlo. The total PDF uncertainty is identical with the CT10 intra-PDF uncertainty.

MC@NLO+Herwig (no background subtraction is applied). These final results at particle and parton levels are compared to the truth MC@NLO+Herwig distributions. The ratio between them gives the relative $t\bar{t}$ generator uncertainty, which is symmetrized.

Besides the above approach, it is possible to use the inverted approach, in which the migration matrices and efficiency corrections are obtained from the MC@NLO+Herwig sample, and the truth and the detector level distributions are obtained from Powheg+Herwig. This approach is tested and it is found that the two approaches lead to similar $t\bar{t}$ generator uncertainties (their absolute difference is $< 1\%$).

Parton Shower and Hadronization

The PS and Hadronization uncertainty is derived with the same approach as the $t\bar{t}$ generator uncertainty using the signal samples Powheg+Herwig and Powheg+Pythia6. The migration matrices and efficiency corrections are obtained from the Powheg+Pythia6 sample, and the truth and the detector level distributions are obtained from the Powheg+Herwig sample.

ISR and FSR

The ISR and FSR uncertainty is derived with similar approach as the $t\bar{t}$ generator uncertainty. The migration matrices and efficiency corrections are obtained from the Powheg+Pythia6 sample. The truth and the detector level distributions are obtained from the AcerMC+Pythia6 samples with less and more PS settings. The half of the difference between uncertainties derived for less and more PS is taken as the ISR and FSR uncertainty.

MC Statistical Uncertainty

The limited number of generated MC reference signal events has impact on the final results, since the migration matrices and the efficiency correction are affected by the corresponding statistical uncertainties. This effect leads to another signal modeling systematic uncertainty called MC signal statistical uncertainty. It is evaluated similarly as the data statistical uncertainty. A set of $2 \cdot 10^5$ replicas of migration matrices and efficiency corrections are constructed by performing random smearing. The correlation between events in migration matrix and efficiency corrections is taken into account. The final results are obtained using the P^{nom} distribution as an input in Eq. 5.1–5.4 for each replica of migration matrices and efficiency corrections. The resulting set of final results is used to obtain the RMS and mean value in each bin. Their ratio is assigned as the relative MC signal statistical uncertainty of the measurement.

5.7.2.4 Luminosity Uncertainty

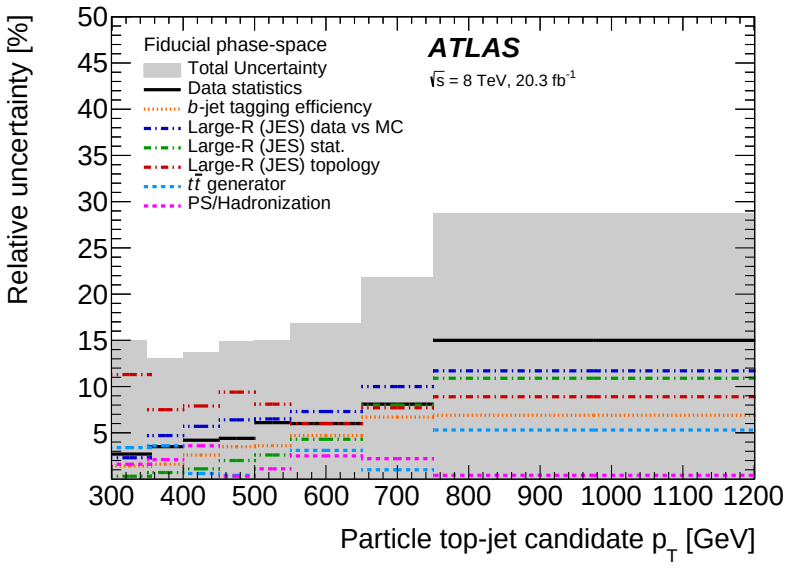
The luminosity uncertainty is 2.8%. The luminosity uncertainty affects the measurement in two ways: the background estimate from MC samples within distribution B in Eq. 5.2 depends on the luminosity, and the unfolded result is divided by the luminosity in Eq. 5.1 and 5.3. These two effects are correlated when the luminosity uncertainty for the final result is evaluated.

5.7.2.5 Total Systematic Uncertainty

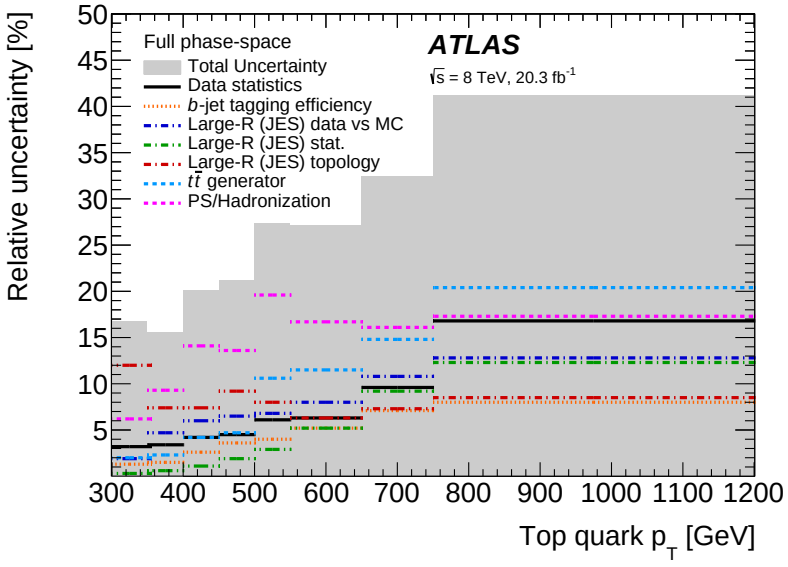
The total detector level systematic uncertainty is included in Fig. 5.1b. It is evaluated as the squared sum of luminosity, MC signal statistical uncertainty, object and background modeling uncertainties computed with Eq. 5.7. It ranges from 10% to 20%.

The total systematic uncertainties for the particle and parton level results are computed as the squared sum of the individual uncertainty components presented in Sec. 5.7.2. They are shown in Tab. 5.2-5.3. The total systematic uncertainty ranges from 13% to 24% for the particle level result, and from 15% to 38% for the parton level result. The Fig. 5.6 shows the dominant uncertainty components for particle and parton level results.

The major systematic uncertainty components for the particle level result are the anti-k_t $R = 1.0$ JES uncertainties. The JES topology component is dominant at low $p_{\text{T}}^{\text{ptcl}}$ and the gamma+jet calibration JES uncertainty components are dominant for high $p_{\text{T}}^{\text{ptcl}}$. The major systematic uncertainty components for the parton level result are the anti-k_t $R = 1.0$ JES uncertainties and the signal modeling uncertainties. For $p_{\text{T}}^{\text{parton}} < 350$ GeV, the JES topology uncertainty component dominates, and for the rest of the distribution the $t\bar{t}$ generator and PS and hadronization uncertainties are dominant and they go up to 20%. The signal modeling uncertainties for the particle level result are much lower than for the parton level result, which demonstrates the benefit of a particle level measurement in a fiducial region similar to the detector level selection criteria.



(a) Particle level



(b) Parton level

Figure 5.6: Relative uncertainties on (a) the particle level differential cross section and (b) the parton level differential cross section. The total uncertainty (grey band) is shown along with the effect of the dominant uncertainties. The uncertainties "Large-R (JES) topology", "Large-R (JES) stat." and "Large-R (JES) data vs MC" are the major components of the JES uncertainty for anti-k_t $R = 1.0$ jets. [5]

5.8 Uncertainty Correlation

Each uncertainty component presented in Sec. 5.7.2 leads to uncertainties which are correlated between bins of the final results, and therefore the total uncertainties can be correlated between bins. It is important to evaluate the corresponding covariance matrix and use it when performing any quantitative comparison between this measurement and certain prediction.

Two separate covariance matrices are evaluated which are added at the end. The first covariance matrix is associated with all the object modeling, the background modeling and MC statistical uncertainties. The second covariance matrix is associated with PDF, $t\bar{t}$ generator, PS and hadronization, ISR and FSR uncertainties. The two covariance matrices are evaluated using different approaches.

The procedure to obtain the first covariance matrix is similar to the procedure used to derive the statistical uncertainty correlations in Sec. 5.7.1, and it is outlined in the following. An ensemble of $2.5 \cdot 10^5$ pseudo-experiments is prepared performing smearing of the D distribution. First, the smearing due to the statistical uncertainty is applied, and then the smearing due to individual systematic uncertainty components is applied using the relative shifts $s^{k,\mathrm{up}}$ and $s^{k,\mathrm{down}}$ from Eq. 5.5 and 5.6. These shifts corresponds to the change of a certain correction factor by one standard deviation, and it is assumed that they are proportional to the size of this change. Each pair of up shifts in two bins from Eq. 5.5 is fully correlated in case the pair of shifts has identical sign or fully anti-correlated in case the pair of shifts has opposite sign. The same holds for the pair of down shifts from Eq. 5.6 or for the symmetrized uncertainties in case of resolution uncertainty components. The above considerations leads to the equation for the smeared number of events for the j-th pseudo-experiment, D^j, in bin i:

$$D_i^j = \mathrm{Pois}^j\left(D_i\right) \cdot \left(1 + \sum_{k \in \mathrm{sys}} \left|\mathrm{Nor}^{k,j}\right| \cdot \left(H\left(\mathrm{Nor}^{k,j}\right) \cdot s_i^{k,\mathrm{up}} + H\left(-\mathrm{Nor}^{k,j}\right) \cdot s_i^{k,\mathrm{down}}\right)\right), \tag{5.9}$$

where the sum runs over all the object and background systematic uncertainty components. The function $H\left(x\right)$ is the Heaviside step function defined as $H\left(x\right) = 0$ for $x < 0$, and $H\left(x\right) = 1$ for $x \geq 0$, the set of random numbers $\mathrm{Pois}^j\left(D_i\right)$ for $j \in \{1, 2, \ldots\}$ is generated from Poisson distribution with mean D_i for each bin i, and the set of random numbers $\mathrm{Nor}^{k,j}$ for $j \in \{1, 2, \ldots\}$ is generated from standard normal distribution (for each component k independently). The final results are obtained using each pseudo-experiment D^j as the input to Eq. 5.1–5.4 with unchanged background estimates, and with smeared migration matrices and efficiency corrections as described in Sec. 5.7.2.3 to account for the MC signal statistical uncertainty. The corresponding ensemble of final results is used to obtain the first type of covariance matrices of the particle and parton level results. It is found that the mean value final results from the ensemble of pseudo-experiments in individual bins is consistent with the final result obtained with non-smeared data within $< 2\%$.

The second covariance matrix is obtained as a sum of four covariance matrices for the individual signal modeling uncertainties: PDF, $t\bar{t}$ generator, PS and hadronization, ISR and FSR. The diagonal elements corresponds to the square of the standard deviation, which is evaluated as the product of the relative uncertainty obtained in Sec. 5.7.2.3 and the final cross section in the corresponding bin. The bin-to-bin correlations are set to a fixed value for each bin pair when evaluating the non-diagonal elements of the covariance matrices. There are no first principles for the determination of these correlation factors. They are set to unity for the $t\bar{t}$ generator, PS and hadronization, ISR and FSR uncertainties, and the correlation for PDF uncertainty is set to 0.5.

The final covariance matrix for the particle and parton level results are shown

in Tab. 5.6 and 5.8 and the corresponding correlation matrices are shown in Tab. 5.7 and 5.9. The correlations between neighboring bins are strong, and there are significant correlation between more distant pairs of bins. The total uncertainties in the individual bins can be extracted as well from the diagonal elements of these covariance matrices. They are in good agreement with the total uncertainties obtained with the first method which was summarized in Sec. 5.7.2.5, in which each systematic uncertainty component is evaluated separately. The relative differences in the individual bins between the two approaches are below 7%. These differences are caused by the different procedure for the total uncertainty evaluation: in the first approach, the individual systematic components are evaluated separately and they are summed after unfolding, and in the second approach, the total effect of all systematic components are applied before unfolding. There is another point which is different between the two approaches: in the first approach the P^{nom} distribution is smeared while in the second approach, the D distribution is smeared.

bin	1	2	3	4	5	6	7	8
1	$5.1 \cdot 10^5$	$2.2 \cdot 10^5$	$1.4 \cdot 10^5$	$9.0 \cdot 10^4$	$4.4 \cdot 10^4$	$2.1 \cdot 10^4$	$9.3 \cdot 10^3$	$1.4 \cdot 10^3$
2	$2.2 \cdot 10^5$	$1.3 \cdot 10^5$	$7.4 \cdot 10^4$	$4.6 \cdot 10^4$	$2.4 \cdot 10^4$	$1.2 \cdot 10^4$	$5.4 \cdot 10^3$	$8.2 \cdot 10^2$
3	$1.4 \cdot 10^5$	$7.4 \cdot 10^4$	$6.0 \cdot 10^4$	$3.4 \cdot 10^4$	$1.7 \cdot 10^4$	$8.8 \cdot 10^3$	$4.2 \cdot 10^3$	$5.9 \cdot 10^2$
4	$9.0 \cdot 10^4$	$4.6 \cdot 10^4$	$3.4 \cdot 10^4$	$2.5 \cdot 10^4$	$1.2 \cdot 10^4$	$5.6 \cdot 10^3$	$2.7 \cdot 10^3$	$4.0 \cdot 10^2$
5	$4.4 \cdot 10^4$	$2.4 \cdot 10^4$	$1.7 \cdot 10^4$	$1.2 \cdot 10^4$	$7.4 \cdot 10^3$	$3.3 \cdot 10^3$	$1.4 \cdot 10^3$	$2.0 \cdot 10^2$
6	$2.1 \cdot 10^4$	$1.2 \cdot 10^4$	$8.8 \cdot 10^3$	$5.6 \cdot 10^3$	$3.3 \cdot 10^3$	$2.1 \cdot 10^3$	$9.1 \cdot 10^2$	$1.2 \cdot 10^2$
7	$9.3 \cdot 10^3$	$5.4 \cdot 10^3$	$4.2 \cdot 10^3$	$2.7 \cdot 10^3$	$1.4 \cdot 10^3$	$9.1 \cdot 10^2$	$5.0 \cdot 10^2$	$7.4 \cdot 10^1$
8	$1.4 \cdot 10^3$	$8.2 \cdot 10^2$	$5.9 \cdot 10^2$	$4.0 \cdot 10^2$	$2.0 \cdot 10^2$	$1.2 \cdot 10^2$	$7.4 \cdot 10^1$	$1.4 \cdot 10^1$

Table 5.6: Covariance matrix for the particle level differential cross section as a function of $p_{\text{T}}^{\text{ptcl}}$. The elements of the covariance matrix are in units of ab^2/GeV^2. [5]

bin	1	2	3	4	5	6	7	8
1	1.00	0.83	0.79	0.79	0.72	0.63	0.58	0.51
2	0.83	1.00	0.83	0.80	0.76	0.74	0.67	0.60
3	0.79	0.83	1.00	0.87	0.79	0.78	0.77	0.63
4	0.79	0.80	0.87	1.00	0.89	0.76	0.77	0.66
5	0.72	0.76	0.79	0.89	1.00	0.84	0.75	0.62
6	0.63	0.74	0.78	0.76	0.84	1.00	0.89	0.71
7	0.58	0.67	0.77	0.77	0.75	0.89	1.00	0.87
8	0.51	0.60	0.63	0.66	0.62	0.71	0.87	1.00

Table 5.7: Correlation matrix of the particle level differential cross section as a function of $p_{\text{T}}^{\text{ptcl}}$. [5]

bin	1	2	3	4	5	6	7	8
1	$9.1 \cdot 10^7$	$3.0 \cdot 10^7$	$1.8 \cdot 10^7$	$1.0 \cdot 10^7$	$5.7 \cdot 10^6$	$2.1 \cdot 10^6$	$7.2 \cdot 10^5$	$9.5 \cdot 10^4$
2	$3.0 \cdot 10^7$	$1.4 \cdot 10^7$	$8.6 \cdot 10^6$	$4.6 \cdot 10^6$	$2.8 \cdot 10^6$	$1.0 \cdot 10^6$	$3.6 \cdot 10^5$	$4.7 \cdot 10^4$
3	$1.8 \cdot 10^7$	$8.6 \cdot 10^6$	$5.8 \cdot 10^6$	$3.1 \cdot 10^6$	$1.8 \cdot 10^6$	$7.0 \cdot 10^5$	$2.5 \cdot 10^5$	$3.3 \cdot 10^4$
4	$1.0 \cdot 10^7$	$4.6 \cdot 10^6$	$3.1 \cdot 10^6$	$1.8 \cdot 10^6$	$1.1 \cdot 10^6$	$3.9 \cdot 10^5$	$1.4 \cdot 10^5$	$1.8 \cdot 10^4$
5	$5.7 \cdot 10^6$	$2.8 \cdot 10^6$	$1.8 \cdot 10^6$	$1.1 \cdot 10^6$	$7.1 \cdot 10^5$	$2.7 \cdot 10^5$	$8.9 \cdot 10^4$	$1.2 \cdot 10^4$
6	$2.1 \cdot 10^6$	$1.0 \cdot 10^6$	$7.0 \cdot 10^5$	$3.9 \cdot 10^5$	$2.7 \cdot 10^5$	$1.1 \cdot 10^5$	$4.0 \cdot 10^4$	$5.3 \cdot 10^3$
7	$7.2 \cdot 10^5$	$3.6 \cdot 10^5$	$2.5 \cdot 10^5$	$1.4 \cdot 10^5$	$8.9 \cdot 10^4$	$4.0 \cdot 10^4$	$1.6 \cdot 10^4$	$2.3 \cdot 10^3$
8	$9.5 \cdot 10^4$	$4.7 \cdot 10^4$	$3.3 \cdot 10^4$	$1.8 \cdot 10^4$	$1.2 \cdot 10^4$	$5.3 \cdot 10^3$	$2.3 \cdot 10^3$	$3.5 \cdot 10^2$

Table 5.8: Covariance matrix for the parton level differential cross section as a function of $p_{\mathrm{T}}^{\mathrm{parton}}$. The elements of the covariance matrix are in units of $\mathrm{ab}^2/\mathrm{GeV}^2$. [5]

bin	1	2	3	4	5	6	7	8
1	1.00	0.82	0.77	0.80	0.71	0.64	0.59	0.53
2	0.82	1.00	0.93	0.91	0.87	0.82	0.75	0.66
3	0.77	0.93	1.00	0.96	0.90	0.87	0.81	0.73
4	0.80	0.91	0.96	1.00	0.94	0.87	0.81	0.73
5	0.71	0.87	0.90	0.94	1.00	0.94	0.84	0.75
6	0.64	0.82	0.87	0.87	0.94	1.00	0.94	0.84
7	0.59	0.75	0.81	0.81	0.84	0.94	1.00	0.97
8	0.53	0.66	0.73	0.73	0.75	0.84	0.97	1.00

Table 5.9: Correlation matrix of the parton level differential cross section as a function of $p_{\mathrm{T}}^{\mathrm{parton}}$. [5]

5.9 Comparison of the Measurement with Standard Model Predictions

The measurement is compared to several SM predictions based on QCD. The Fig. 5.7 shows the measured particle and parton level results compared to four MC predictions: POWHEG+PYTHIA6, ALPGEN+HERWIG, MC@NLO+HERWIG, and POWHEG+HERWIG, which are described in Sec. 5.1. All the samples are normalized to the same total cross section resulting from the NNLO+NNLL prediction.

At particle level (Fig. 5.7a), the MC@NLO+HERWIG and POWHEG+HERWIG predictions agree the best with the measured cross sections, although there are one standard deviation differences in several bins. The ALPGEN+HERWIG prediction is more distinct from the measurement since there are in some bins three standard deviation differences. The POWHEG+PYTHIA6 prediction is also discrepant in comparison with the measurement within two standard deviations. For all the predictions, there is a positive slope as a function of $p_{\mathrm{T}}^{\mathrm{ptcl}}$ for the ratio of central values of prediction and measurement. The overall normalization of the measured distribution is lower than the normalization of each MC prediction. The source of this non-normalization and slope can originate from imprecise theory predictions or systematic effects of the measurement.

At parton level (Fig. 5.7b), the MC@NLO+HERWIG and POWHEG+HERWIG predictions are consistent with the measurement. The ALPGEN+HERWIG and

POWHEG+PYTHIA6 predictions are less consistent with the measurement. Similar slope is observed between predictions and measurement as it is observed at particle level.

To decide if the above observed discrepancies between measurement and predictions are significant, the correlations of the uncertainties between bins need to be taken into account. A quantitative comparison is performed by employing the Pearson's χ^2 test. The χ^2 values are obtained from equation

$$\chi^2 = V^T \cdot \mathrm{Cov}^{-1} \cdot V \tag{5.10}$$

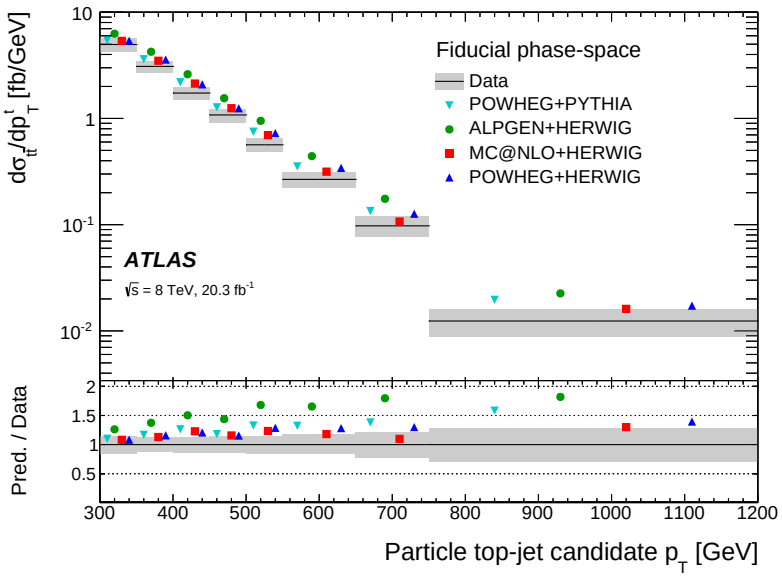
where V is the vector of differences between the central values of the measurement and the prediction, and Cov^{-1} is the inverse of the covariance matrix obtained in Sec. 5.8. The number of degrees of freedom corresponds with the number of bins, which is 8. The corresponding p-values are evaluated as well. The p-value represents the probability that under the assumption a given prediction is true, the measurement results in the observed distribution or worse distribution (in the sense with higher χ^2).

The χ^2 and p-values at particle level are shown in Tab. 5.10 for the MC predictions POWHEG+PYTHIA6+EWT and POWHEG+PYTHIA6 $h_{\mathrm{damp}} = \infty$ and for the predictions from Fig. 5.7. The measured particle level distribution is consistent with all the MC predictions except the ALPGEN+HERWIG prediction, for which the Pearson's χ^2 test leads to p-value of $5.9 \cdot 10^{-5}$. The other MC predictions are consistent with the particle level measurement according to this statistical test in spite of the fact that there are visible discrepancies in several bins in Fig. 5.7a. This consistency is obtained due to the high correlation of the uncertainties between bins. These correlations originate mainly from the dominant systematic uncertainty components (JES for anti-k_t $R = 1.0$ jets). The measurement is slightly more consistent with the POWHEG+PYTHIA6+EWT prediction than with the POWHEG+PYTHIA6 prediction, which implies that the EWT corrections may be important for high $p_{\mathrm{T}}^{\mathrm{top}}$ values. The p-values are evaluated at the parton level as well, and it is found that the measurement is consistent with all the above tested MC predictions.

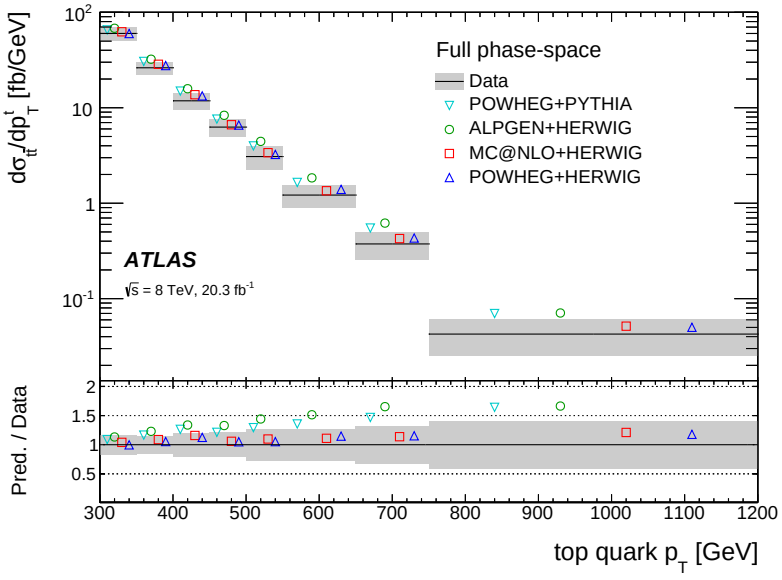
MC prediction	χ^2	p-value
POWHEG+PYTHIA6+EWT	9.8	0.28
POWHEG+PYTHIA6	13.0	0.11
POWHEG+PYTHIA6 $h_{\mathrm{damp}} = \infty$	15.6	0.05
POWHEG+HERWIG	8.2	0.41
MC@NLO+HERWIG	12.3	0.14
ALPGEN+HERWIG	33.1	$5.9 \cdot 10^{-5}$

Table 5.10: Values of χ^2 and p-value, computed for 8 degrees of freedom, obtained from the covariance matrix of the measured cross-section for various predictions. Electroweak corrections are applied only to the first prediction. The number of degrees of freedom is 8. [5]

In summary, the performed measurement is consistent within uncertainties with most of the MC predictions at particle and parton level. However, two comments are in place concerning the shape and the normalization of the compared MC predictions.



(a)



(b)

Figure 5.7: (a) Particle level differential cross-section as a function of the hadronic top-jet candidate p_{T} and (b) parton-level differential cross-section as a function of the hadronically decaying top quark p_{T}, compared to several MC generator predictions. The lower part of the figure shows the ratio of the MC prediction to the data. The shaded area includes the total statistical plus systematic uncertainties. The points of the various predictions are spaced along the horizontal axis for presentation only; they correspond to the same p_{T} range. [5]

First, a positive slope is observed as a function of $p_{\mathrm{T}}^{\mathrm{ptcl}}$ and $p_{\mathrm{T}}^{\mathrm{top}}$ for the ratio of central values between the measurement and each MC prediction based on NLO QCD. Similar slope is observed also for the low $p_{\mathrm{T}}^{\mathrm{top}}$ range in the ATLAS measurements at $\sqrt{s} = 7$ TeV and in the CMS measurements in both, resolved and boosted, regimes when compared to MC predictions based on NLO QCD (as discussed in Sec. 1.2.3). This slope is smaller in case the recent NNLO QCD

prediction from 2015 is taken into account, see Fig. 1.6.

The second comment is that the compared MC predictions are normalized to the NNLO+NNLL QCD total cross section prediction (this roughly increases the normalization by 10% with respect to the NLO QCD normalization). However, this normalization may be overestimated in the phase space, in which the measurement is performed. The Fig. 1.6 shows that for $p_{\mathrm{T}}^{\mathrm{top}} \in [300, 400]$ GeV, the NNLO QCD prediction of the cross section is lower than the NLO QCD prediction.

These two comments lead to the conclusion that the consistency between the measurement in this thesis and the SM prediction can be even better when taking into account the NNLO QCD prediction. A non-negligible role can play also the EWT contribution which also lowers the normalization and the slope between the measurement and the predictions. The NNLO QCD prediction for the high $p_{\mathrm{T}}^{\mathrm{top}}$ range for $\sqrt{s} = 8$ TeV pp collisions is not available nowadays, and hence a quantitative comparison with the measurement cannot be made.

5.9.1 Evaluation of the NLO QCD prediction and comparison with the measurement

The MC predictions in Fig. 5.7 are obtained for certain fixed settings in the MC generators and using certain PDF sets, and they do not account for any theoretical uncertainties. In this section, the measurement is compared to a SM prediction containing the main theoretical uncertainties.

A parton level NLO QCD prediction is evaluated. The MCFM generator [145] is used to produce this prediction with the uncertainties associated with the strong coupling constant α_{S}, the PDF, and the scales μ_F and μ_R. The prediction is evaluated for four PDF sets: CT10, MSTW, NNPDF, and HERAPDF [146]. The scales μ_R and μ_F are varied dynamically as $\mu_R = \mu_F = \sqrt{m^{\mathrm{top2}} + \hat{p_{\mathrm{T}}}^2}$ where $\hat{p_{\mathrm{T}}}$ is the average between top and antitop p_{T} in the event. The mass of the top quark is set to the value of 172.5 GeV. The strong coupling constant is set to the value of 0.118 for μ_R equal to the mass of the Z boson. The overall normalization is kept at the NLO QCD (it is not normalized to the NNLO+NNLL total cross section as it is done for the previous predictions).

The scale uncertainty is obtained from the envelope of four variations: $2\mu_R$, $2\mu_F$; $2\mu_R$, $0.5\mu_F$; $0.5\mu_R$, $2\mu_F$; $0.5\mu_R$, $0.5\mu_F$. The usual interpretation of the scale uncertainty is that it encompasses the missing terms in the perturbation expansion. The α_{S} uncertainty is evaluated by varying the α_{S} by 2%. The PDF uncertainty is evaluated for each PDF set using the prescription from the corresponding PDF group. The APPLgrid tool [147] is used to perform faster evaluation of the predictions for various uncertainty PDF sets. The total uncertainty is evaluated as the squared sum of the three uncertainties.

The comparison of the above NLO QCD prediction with the measurement at parton level is shown in Fig. 5.8. The measurement is in good agreement with the predictions for each PDF set within the quoted uncertainties. A statistical test is performed as well, and the measurement at parton level is consistent with the NLO QCD predictions for all the tested PDF sets.

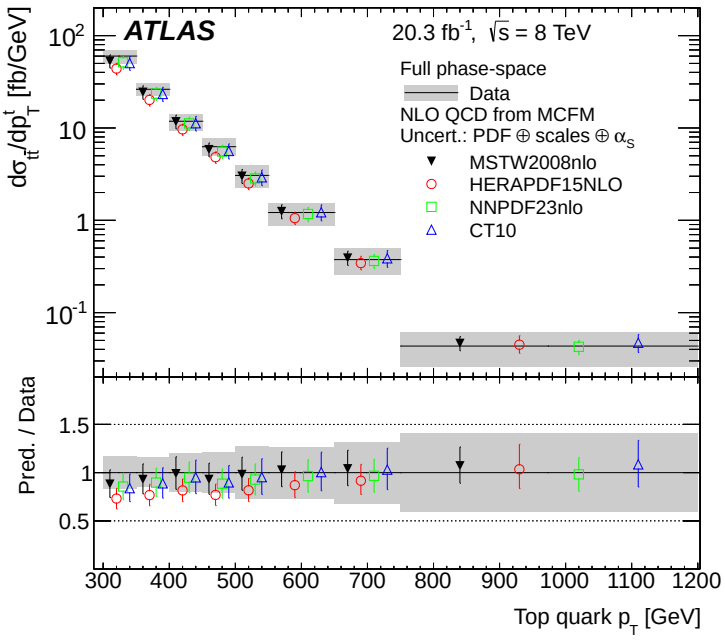


Figure 5.8: Measured parton level differential cross-section as a function of the hadronically decaying top quark p_{T}. The MCFM predictions with various PDF sets are also shown. The lower part of the figure shows the ratio of the MCFM predictions to the data. The shaded area includes the total statistical plus systematic uncertainties of the measurement. The uncertainty on the predictions include the PDF, α_{S}, μ_F, and μ_R uncertainties. [5]

Summary

This thesis presents the measurement of the $t\bar{t}$ differential cross section for high p_{T} top quarks at $\sqrt{s} = 8$ TeV at the ATLAS experiment. The experimental techniques used in this measurement are discussed with emphasis on the jet reconstruction and on the identification of jets originating from b-quark fragmentation.

The measurement of the $t\bar{t}$ differential cross section is performed in the single lepton decay channel, and it is expressed as a function of the top quark p_{T} for top quarks with $p_{\mathrm{T}} > 300$ GeV. Jet substructure techniques are used to help identify the hadronically decaying top quarks within anti-k_t $R = 1.0$ jets. The measurement is extracted at particle level and parton level. All possible uncertainty sources are taken into account. At the particle level, the largest contribution to the total uncertainty comes from the jet energy scale of the anti-k_t $R = 1.0$ jets. At the parton level, the same uncertainty is dominant together with the physics modeling of the $t\bar{t}$ production. The bin-by-bin correlations of all statistical and systematic uncertainties are evaluated. The consistency of the measurement and several MC predictions based on the SM is performed using a statistical test. The measurement is consistent with the majority of tested MC predictions. The measurement can constrain models beyond the SM, and it can be used to tune MC generator parameters. The top quark p_{T} above 300 GeV is sensitive to the gluon PDF for high proton momentum fraction, and therefore the groups performing PDF fits can benefit from this measurement as well.

The jet reconstruction is influenced by pileup, and the mitigation of the pileup effects on jets is studied in this thesis. The state-of-the-art pileup correction techniques are presented. The performance of one such technique, the shape-expansion method, is tested on the jets from $\sqrt{s} = 8$ TeV pp collisions at the ATLAS experiment. A new pileup subtraction method, Constituent Subtraction, is presented, which represents an extension and a simplification of the current methods. Its performance is evaluated using MC simulated events at particle level. The Constituent Subtraction method operates at the level of the jet constituents and provides a performance improvement compared to the existing methods: the precision of the reconstruction of the jet 4-momenta and jet shapes is improved. This performance improvement can directly impact the experimental sensitivity to new physics. The performance observed in the MC events should be vetted and tested thoroughly within the LHC experiments using full detector simulation and data.

The identification of jets originating from b-quark fragmentation at the ATLAS experiment is summarized in this thesis. The b-jet identification is essential to suppress the background events with respect to the $t\bar{t}$ events. The ideas of the b-jet identification algorithms are summarized together with the way of how the MC identification performance is calibrated to be consistent with the performance in data. The identification in simulated events is dependent on the MC generators, and this effect is included in the calibration process.

A. Conventions and Kinematic Variables

In this Appendix, the basic conventions and kinematic variables which are relevant for this thesis are introduced. The natural system of units [148] is used in which the reduced Planck constant and the speed of light, c, are equal to 1.

The 4-momentum of a particle[1] with energy E and momentum $\overrightarrow{p} = [p_x, p_y, p_z]$ is

$$P = [p_x, p_y, p_z, E]. \tag{A.1}$$

All variables are expressed in the laboratory reference frame of the ATLAS experiment. The right-handed coordinate system is used with its origin at the interaction point in the center of the detector. The z-axis points along the colliding beams. The x-axis points from the interaction point to the center of the LHC ring, and the y-axis points upward to the surface of the earth.

The magnitude of the momentum is defined as

$$p = \sqrt{p_x^2 + p_y^2 + p_z^2}. \tag{A.2}$$

The mass, m, is defined as

$$m = \sqrt{E^2 - p^2}, \tag{A.3}$$

which is a Lorentz invariant quantity. The transverse momentum, p_{T}, is defined as

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}. \tag{A.4}$$

The azimuth, ϕ, fulfills the equation

$$\cos\phi = \frac{p_x}{p_{\mathrm{T}}} \tag{A.5}$$

and the two conditions: $\phi \in [0, \pi]$ for $p_y \geq 0$ and $\phi \in (\pi, 2\pi)$ for $p_y < 0$. The polar angle, θ, fulfills

$$\cos\theta = \frac{p_z}{p} \tag{A.6}$$

and $\theta \in [-\pi/2, \pi/2]$. The pseudorapidity is defined as

$$\eta = -\ln\tan\theta/2, \tag{A.7}$$

and the rapidity is defined as

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}. \tag{A.8}$$

It holds

$$\lim_{m/E \to 0} y = \eta \tag{A.9}$$

[1] By particle, any object is meant which was created as a sum of 4-momenta, i.e. particle can be also a jet in this context.

therefore for massless particles, y is identical with η. Another useful kinematic variable is the difference between the transverse mass and the transverse momentum, m_δ:

$$m_\delta = \sqrt{m^2 + p_{\mathrm{T}}^2} - p_{\mathrm{T}} \tag{A.10}$$

The 4-momentum in Eq. A.1 can be parametrized using different four variables p_{T}, m_δ, ϕ, and y like this:

$$p^\mu = [p_{\mathrm{T}} \cos\phi, p_{\mathrm{T}} \sin\phi, (p_{\mathrm{T}} + m_\delta) \sinh y, (p_{\mathrm{T}} + m_\delta) \cosh y]. \tag{A.11}$$

One of the distance measures between two particles i and k often used in pp collisions is the ΔR distance defined as

$$\Delta R_{i,k} = \sqrt{(y_i - y_k)^2 + (\phi_i - \phi_k)^2} \tag{A.12}$$

where y_i and y_k are the rapidities of particles i and k, respectively, and ϕ_i and ϕ_k are the azimuths of particles i and k, respectively. Another possibility to define the distance between two particles is to change the rapidity difference to pseudorapidity difference, and get the distance ΔR^η:

$$\Delta R^\eta_{i,k} = \sqrt{(\eta_i - \eta_k)^2 + (\phi_i - \phi_k)^2} \tag{A.13}$$

where η_i and η_k are the pseudorapidities of particles i and k, respectively. The symbol ΔR without the η superscript always refers to the distance in Eq. A.12, and the symbol ΔR^η always refers to the distance in Eq. A.13 in this thesis. Another distance often used is the distance in the transverse plane, $\Delta\phi$, defined as

$$\Delta\phi_{i,k} = |\phi_i - \phi_k| \tag{A.14}$$

An important feature in pp collisions is the property called longitudinal invariance. An observable which is longitudinally invariant is invariant with respect to a Lorentz boost in the direction of the z-axis. Examples of the longitudinally invariant observables are the p_{T}, the difference of two rapidities, the difference of two azimuths, and hence also the distance ΔR. In pp collisions effectively, one parton from one colliding proton interacts with another parton from the second colliding proton. Each parton carries certain fraction of the proton's momentum. The longitudinal component of the momentum of the parton-parton system is a random variable, and it cannot be estimated experimentally. For this reason, it is useful to use longitudinally invariant quantities. The distance ΔR^η is not longitudinally invariant, but it has larger geometrical meaning. The ΔR^η depends only on the directions of the momenta of the two particles while ΔR depends also on the mass of the particles. Therefore it can happen that the two particles are geometrically distant, and they have low ΔR in certain configurations or vice versa.

For practical purposes, the mean lifetime, τ, of a certain particle is expressed in units of distance after multiplying the lifetime by the speed of light ($3{\cdot}10^8$ m$\cdot$s^{-1}). Then the mean traveled distance, $\langle d \rangle$, is simply given by equation

$$\langle d \rangle = \frac{p}{m}\tau. \tag{A.15}$$

B. Jets and Jet Clustering Algorithms

Jets are collimated sprays of hadrons that result from the fragmentation of a high-energy quark or gluon. In the final state of a collision event, there may be several quarks and gluons, and one needs to define rules that project a set of particles into a set of jets. These rules are referred to as a jet clustering algorithm. It can be applied to the measured objects in calorimeter, or to the particle or partonic final state of MC predictions. The jet clustering algorithm must be infrared and collinear safe in order to be able to use it in perturbative QCD calculations [149].

There are two main types of jet clustering algorithms: cone algorithms and sequential recombination algorithms, [94]. Only sequential recombination algorithms described in the following are used in this thesis. One set of such algorithms used in hadron-hadron collisions are the longitudinally invariant k_t algorithms. The generalized form of the longitudinally invariant k_t algorithms is formulated as follows:

1. For each pair of particles i and k evaluate the distances:

$$d_{ik} = \min\left(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}k}^{2p}\right)\Delta R_{i,k}^2/R^2 \tag{B.1}$$

$$d_{Bi} = p_{\mathrm{T}i}^{2p} \tag{B.2}$$

   where $p_{\mathrm{T}i}$ and $p_{\mathrm{T}k}$ are the transverse momenta of particles i and k, respectively. The parameters p and R are free parameters.

2. Find the minimum among all distances d_{ik} and d_{iB}. If the minimal distance is in the set of distances d_{ik}, then merge the two particles i and k into one by summing their 4-momenta[1]. If the minimal distance is in the set of distances d_{iB}, then remove the particle i from the list and declare it as a final jet.

3. Repeat from Step 1 until no particles are left.

The above algorithm is infrared and collinear safe. In case $p = 1$, this algorithm is referred to as the longitudinally invariant k_t jet algorithm [150, 151], or simply the k_t algorithm. In case $p = 0$, the above algorithm is referred to as the Cambridge/Aachen (C/A) jet algorithm [152, 153, 154]. In case $p = -1$, the above algorithm is referred as the anti-k_t jet algorithm [155]. All these three jet algorithms have one free parameter: the distance parameter R. The most common jet definitions at the ATLAS experiment are the anti-k_t jet algorithm with distance parameters $R = 0.4$, $R = 0.6$ and $R = 1.0$, and the C/A jet algorithm with distance parameter $R = 1.2$. They are referred to as anti-k_t $R = 0.4$, anti-k_t $R = 0.6$, anti-k_t $R = 1.0$, and C/A $R = 1.2$ algorithms in this thesis. The jet algorithms anti-k_t $R = 1.0$ and C/A $R = 1.2$ are also called large-R jet algorithms leading to large-R jets.

[1] How to combine the two particles into one defines the recombination schemes. There are several recombination schemes. The summing of the two 4-momenta defines the E-scheme recombination, and only this recombination scheme is used in this thesis.

There is another jet algorithm used in this thesis called the exclusive k_t algorithm which has one more parameter, N. It is identical with the above mentioned k_t algorithm, but the clustering stops when the number of jets is N. The distance parameter R is usually set to a larger value, e.g. $R = 1.0$, for this jet algorithm.

The particles projected into a common jet are called *jet constituents*.

C. Jet Shape Definitions

Jet shape is an arbitrary function of the 4-momenta of the jet constituents, $\{P_i\}$. There are several jet shapes used in this thesis, and they presented in the following.

C.1 N-subjettiness

The N-subjettiness, τ_N, [156] is a jet shape with one integer parameter N. Is is usually evaluated for anti-k_t or C/A large-R jets to help identify boosted objects. The constituents of the jet are re-clustered with exclusive k_t algorithm with large-enough distance parameter to cluster all constituents into one jet. The parameter N of the exclusive k_t algorithm is identical with the parameter N of τ_N. The N jets obtained from the exclusive k_t algorithm are called subjets, and they are used to evaluate τ_N:

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}k} \cdot \min(\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}) \ , \quad \text{with} \quad d_0 \equiv \sum_k p_{\mathrm{T}k} \cdot R_{\mathrm{orig}} \quad \text{(C.1)}$$

where $p_{\mathrm{T}k}$ is the transverse momentum of constituent k, $\Delta R_{i,k}$ is the distance between the subjet i and the constituent k, and R_{orig} is the distance parameter of the clustering algorithm used for the original jet. There is an alternative definition of subjets in [157]. In this thesis, the above definition using exclusive k_t algorithm is used. The N-subjettiness gives a measure how much the given jet consist of N subjets – the smaller the N-subjettiness for certain jet, the larger the probability that this jet consist of N subjets. Beside the N-subjettiness also the subjettiness ratio, $\tau_{MN} = \tau_M/\tau_N$, can be used to characterize the jet substructure. Typically, the three-to-two ratio, $\tau_{32} = \tau_3/\tau_2$, is used which provides a good discrimination between standard QCD jets and jets formed differently e.g. by boosted top quarks. The subjettiness τ_1 is a measure of the width of the jet.

C.2 k_t Splitting Scale

To evaluate the k_t splitting scale, $\sqrt{d_{12}}$, [158] the jet constituents are re-clustered with the k_t algorithm with large-enough distance parameter to cluster all constituents into one jet. Two subjets are found by going back one step in the clustering history. The $\sqrt{d_{12}}$ is defined as

$$\sqrt{d_{12}} = \min(p_{\mathrm{T}}^1, p_{\mathrm{T}}^2) \cdot \Delta R_{12}, \quad \text{(C.2)}$$

where p_{T}^1 and p_{T}^2 are the transverse momenta of the two subjets and ΔR_{12} is the distance between these two subjets. The variable $\sqrt{d_{12}}$ can be used to distinguish boosted heavy-particle decays, which tend to be reasonably symmetric, from the largely asymmetric splittings that originate from QCD radiation in light-quark or gluon jets.

C.3 Energy Correlation Functions

The energy correlation functions, [159], yield information about the internal structure of a jet without the need to explicitly reconstruct subjets. From the energy correlation functions, one can construct another variable called energy correlation double ratio, $C_1^{(\beta)}$, defined as

$$C_1^{(\beta)} = \frac{\sum\limits_{i<j} p_{\mathrm{T}i} p_{\mathrm{T}j} \Delta R_{i,j}^{\beta}}{\left(\sum\limits_{k} p_{\mathrm{T}k}\right)^2}, \tag{C.3}$$

where β represents the pairwise angular exponent. This observable have been shown to be useful for discriminating between quark- and gluon-initiated jets.

Bibliography

[1] ATLAS Collaboration, *Performance of pile-up subtraction for jet shapes* (2013), ATLAS-CONF-2013-085
URL https://cds.cern.ch/record/1572979

[2] Berta P., Spousta M., Miller D. W., Leitner R., *Particle-level pileup subtraction for jets and jet shapes*, JHEP, **1406**, p. 092 (2014)
URL http://link.springer.com/article/10.1007%2FJHEP06%282014%29092

[3] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST, **11** (2016)
URL http://iopscience.iop.org/article/10.1088/1748-0221/11/04/P04008

[4] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data* (2014), ATLAS-CONF-2014-046
URL http://cds.cern.ch/record/1741020

[5] ATLAS Collaboration, *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in sqrt(s) = 8 TeV proton-proton collisions using the ATLAS detector*, Phys. Rev. D, **93**, p. 032009 (2016)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.93.032009

[6] Berta P., *ATLAS jet and missing-ET reconstruction, calibration, and performance* (2014), ATL-PHYS-PROC-2014-158
URL http://cds.cern.ch/record/1951815

[7] Ho-Kim Q. and Pham X., *Elementary Particles and Their Interactions: Concepts and Phenomena* (Springer, 2004), 1 edition, ISBN 3540636676

[8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B, **716**, pp. 1 (2012)
URL http://www.sciencedirect.com/science/article/pii/S037026931200857X

[9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B, **716**, p. 30 (2012)
URL http://www.sciencedirect.com/science/article/pii/S0370269312008581

[10] MissMJ, *Standard Model of Elementary Particles*, This file is licensed under the Creative Commons Attribution 3.0 Unported license, via Wikimedia Commons (downloaded on 27.1.2016)
URL https://commons.wikimedia.org/wiki/File%3AStandard_Model_of_Elementary_Particles.svg

[11] The ATLAS, CDF, CMS, D0 Collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass* (2014)
URL http://arxiv.org/abs/1403.4427

[12] K.A. Olive et al. (Particle Data Group), *Review of Particle Physics*, Chin. Phys. C, **38**, p. 090001 (2014)
URL http://pdg.lbl.gov/2015/download/rpp2014-Chin.Phys.C.38.090001.pdf

[13] Juste A. et al., *Determination of the top quark mass circa 2013: methods, subtleties, perspectives*, The Eur. Phys. Jour. C, **74**, p. 3119 (2014)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-3119-5

[14] Collins J.C. et al., *Heavy particle production in high-energy hadron collisions*, Nuclear Physics B, **263**, 1, pp. 37 (January 1986)
URL http://www.sciencedirect.com/science/article/pii/055032138690026X

[15] Czakon M. et al., *The total top quark pair production cross-section at hadron colliders through O(alpha˙S^4)*, Phys. Rev. Lett., **110**, p. 252004 (2013)
URL http://arxiv.org/abs/1303.6254

[16] ATLAS Collaboration, *Measurement of the ttbar production cross-section using e/mu events with b-tagged jets in pp collisions at sqrt(s)=7 and 8 TeV with the ATLAS detector*, Eur. Phys. J. C, **74**, p. 3109 (2014)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-3109-7

[17] ATLAS Collaboration, *Summary plots from the ATLAS Top physics group* (2015), web page, cited 15.3.2016
URL https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/TOP/

[18] Jeżabek M. and Kühn J., *QCD corrections to semileptonic decays of heavy quarks*, Nuclear Physics B, **314**, 1, pp. 1 (February 1989)
URL http://www.sciencedirect.com/science/article/pii/0550321389901089

[19] Gao J. et al., *Top-Quark Decay at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett., **110**, p. 042001 (2013)
URL http://arxiv.org/abs/1210.2808

[20] D0 Collaboration, *An improved determination of the width of the top quark*, arXiv:1201.4156 (January 2012)
URL http://arxiv.org/abs/1201.4156

[21] CDF Collaboration, *A Direct Measurement of the Total Decay Width of the Top Quark*, Phys. Rev. Lett., **111**, p. 202001 (2013)

URL http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.111.202001

[22] Bigi I. et al., *Production and decay properties of ultra-heavy quarks*, Physics Letters B, **181**, 1-2, pp. 157 (November 1986)
URL http://www.sciencedirect.com/science/article/pii/037026938691275X

[23] Erdmann J. et al., *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nuclear Instruments and Methods in Physics Research Section A, **748**, pp. 18 (2014)
URL http://arxiv.org/abs/1312.5595

[24] ATLAS Collaboration, *Performance of boosted top quark identification in 2012 ATLAS data* (2013), ATLAS-CONF-2013-084
URL http://cds.cern.ch/record/1571040

[25] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at sqrt(s)=7 TeV using the ATLAS detector*, JHEP, **09**, p. 076 (2013)
URL http://link.springer.com/article/10.1007/JHEP09%282013%29076

[26] Rizzo T.G., *Z' Phenomenology and the LHC* (2006), sLAC-PUB-12129
URL http://arxiv.org/abs/hep-ph/0610104

[27] Skands P., *Introduction to QCD* (2012)
URL http://arxiv.org/abs/1207.2389

[28] Buckley A. et al., *General-purpose event generators for LHC physics*, Physics Reports, **504**, p. 145–233 (2011)
URL http://www.sciencedirect.com/science/article/pii/S0370157311000846

[29] Beenakker W. et al., *QCD corrections to heavy quark production in hadron-hadron collisions*, Nucl. Phys. B, **351**, pp. 507 (1991)
URL http://cds.cern.ch/record/208913

[30] Czakon M. et al., *High-precision differential predictions for top-quark pairs at the LHC* (2015)
URL http://arxiv.org/abs/1511.00549

[31] Czakon M. et al., *NNLO QCD predictions for fully-differential top-quark pair production at the Tevatron* (2016)
URL http://arxiv.org/abs/1601.05375

[32] Czakon M. and Mitov A., *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, Computer Physics Communications, **185**, p. 2930 (2014)
URL http://arxiv.org/abs/1112.5675

[33] Kühn J.H. et al., *Weak Interactions in Top-Quark Pair Production at Hadron Colliders: An Update*, Phys. Rev. D, **91**, p. 014020 (2015)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.91.014020

[34] Hollik W. and Kollar M., *NLO QED contributions to top-pair production at hadron collider*, Phys. Rev. D, **77**, p. 014008 (2008)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.77.014008

[35] Martin A. et al., *Parton distributions incorporating QED contributions*, Eur.Phys.J. C, **39**, pp. 155 (2005)
URL http://arxiv.org/abs/hep-ph/0411040

[36] Kidonakis N., *Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution*, Phys. Rev. D, **82**, p. 114030 (2010)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.82.114030

[37] Biswas S. et al., *Next-to-leading order QCD effects and the top quark mass measurements at the LHC*, JHEP, **1008**, p. 048 (2010)
URL http://xxx.lanl.gov/abs/1006.0910

[38] Bevilacqua G. et al., *Complete off-shell effects in top quark pair hadroproduction with leptonic decay at next-to-leading order*, JHEP, **1102**, p. 083 (2011)
URL http://link.springer.com/article/10.1007%2FJHEP02%282011%29083

[39] Frixione S. et al., *Single-top hadroproduction in association with a W boson*, JHEP, **07**, p. 029 (2008)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2008/07/029/meta

[40] Sjöstrand T. et al., *A Brief Introduction to PYTHIA 8.1*, Comp. Phys. Com., **178**, pp. 852 (2008)
URL http://www.sciencedirect.com/science/article/pii/S0010465508000441

[41] Corcella G. et al., *HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons*, JHEP, **0101**, p. 010 (2001)
URL http://arxiv.org/abs/hep-ph/0011363

[42] Buckley A. et al., *Rivet*, web page, last revision 7.10.2015, cited 11.4.2016
URL https://rivet.hepforge.org/

[43] Giammanco A., *Particle level objects and pseudo-top-quark definitions*, web page, last revision 6.4.2016, cited 11.4.2016
URL https://twiki.cern.ch/twiki/bin/view/LHCPhysics/ParticleLevelTopDefinitions

[44] D0 Collaboration, *Measurement of differential ttbar production cross sections in ppbar collisions*, Phys. Rev. D, **90**, p. 092006 (2014)
URL http://arxiv.org/abs/1401.5785

[45] CDF Collaboration, *First Measurement of the ttbar Differential Cross Section in ppbar Collisions at sqrt(s)=1.96 TeV*, Phys. Rev. Lett., **102**, p. 222003 (2009)
URL http://arxiv.org/abs/0903.2850

[46] CMS Collaboration, *Measurement of differential top-quark pair production cross sections in pp collisions at sqrt(s) = 7 TeV*, Eur. Phys. J. C, **73**, p. 2339 (2013)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-013-2339-4

[47] ATLAS Collaboration, *Measurements of normalized differential cross-sections for ttbar production in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector*, Phys. Rev. D, **90**, p. 072004 (2014)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.90.072004

[48] ATLAS Collaboration, *Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at sqrt(s)=7 TeV in the ATLAS detector*, JHEP, **06**, p. 100 (2015)
URL http://link.springer.com/article/10.1007%2FJHEP06%282015%29100

[49] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at sqrt(s)=8 TeV using the ATLAS detector* (2015)
URL http://arxiv.org/abs/1511.04716

[50] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at sqrt(s) = 8 TeV*, Eur. Phys. J. C, **75**, p. 542 (2015)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-015-3709-x

[51] CMS Collaboration, *Measurement of the differential ttbar production cross section for high-pt top quarks in e/mu+jets final states at 8 TeV*, CMS-PAS-TOP-14-012 (2015)
URL https://cds.cern.ch/record/2045404

[52] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST, **3** (August 2008)
URL http://iopscience.iop.org/1748-0221/3/08/S08003

[53] Brüning O.S. et al., *LHC Design Report* (CERN (Geneva), 2004), ISBN 9789290832249
URL http://cdsweb.cern.ch/record/782076

[54] Pequenao J., *Computer generated image of the whole ATLAS detector* (2008)
URL http://cdsweb.cern.ch/record/1095924

[55] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, arXiv:1004.5293 (April 2010), Eur.Phys.J.C70:787-821,2010
URL http://arxiv.org/abs/1004.5293

[56] The ATLAS Liquid Argon Calorimeter Group, *Commissioning and Performance of the ATLAS Liquid Argon Calorimeters*, arXiv:0809.2672 (2008)
URL http://arxiv.org/abs/0809.2672

[57] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, arXiv:1007.5423 (July 2010), Eur.Phys.J.C70:1193-1236,2010
URL http://arxiv.org/abs/1007.5423

[58] ATLAS Collaboration, *Calibration and Performance of the ATLAS Muon Spectrometer*, arXiv:1109.6933 (September 2011)
URL http://arxiv.org/abs/1109.6933

[59] ATLAS Collaboration, *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C, **73**, p. 2518 (2013)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-013-2518-3

[60] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at sqrts=7 TeV in the ATLAS experiment* (2010), aTLAS-CONF-2010-069
URL http://cds.cern.ch/record/1281344

[61] ATLAS Collaboration, *Luminosity Public Results* (2013), web page, last revision 20.7.2015, cited 15.3.2016
URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults

[62] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. C, **74**, p. 2941 (2014)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-2941-0

[63] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur. Phys. J. C, **74**, p. 3130 (2014)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-3130-x

[64] Lampl W. et al., *Calorimeter clustering algorithms: description and performance* (2008), ATL-LARG-PUB-2008-002
URL http://cdsweb.cern.ch/record/1099735

[65] Barillari T. et al., *Local hadronic calibration* (2009), ATL-LARG-PUB-2009-001
URL http://cdsweb.cern.ch/record/1112035

[66] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS* (2013), ATLAS-CONF-2013-083
URL http://cdsweb.cern.ch/record/1570994

[67] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. C, **75**, p. 17 (2015)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-3190-y

[68] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. Jour. C, **73**, p. 2306 (2013)

[69] ATLAS Collaboration, *ATLAS Public Results - JetEtmissApproved2013Jer2011* (2013), web page, last revision 10.9.2013, cited 15.3.2016
URL https://twiki.cern.ch/twiki/bin/view/\AtlasPublic/JetEtmissApproved2013Jer2011

[70] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector* (2014)
URL http://cdsweb.cern.ch/record/1700870

[71] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV* (2013), ATLAS-CONF-2013-082
URL http://cdsweb.cern.ch/record/1570993

[72] ATLAS Collaboration, *Pile-up Suppression in Missing Transverse Momentum Reconstruction in the ATLAS Experiment in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV* (2014), ATLAS-CONF-2014-019
URL http://cdsweb.cern.ch/record/1702055

[73] Kirschenmann H., *Jets at CMS and the determination of their energy scale*, web page, last revision 7.9.2012, cited 22.2.2016.
URL http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale

[74] Agostinelli S. et al., *Geant4: A simulation toolkit*, Nuclear Instruments and Methods A, **506**, 3, pp. 250 (July 2003)
URL http://www.sciencedirect.com/science/article/pii/S0168900203013688

[75] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C, **70**, pp. 823 (2010)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-010-1429-9

[76] Frixione S. and Webber B., *Matching NLO QCD computations and parton shower simulations*, JHEP, **0206**, p. 029 (2002)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2002/06/029/meta

[77] Sjöstrand T. et al., *PYTHIA 6.4 Physics and Manual*, JHEP, **05**, p. 026 (2006)
URL http://arxiv.org/abs/hep-ph/0603175

[78] Martin A. et al., *Parton distributions for the LHC*, Eur. Phys. J. C, **63**, pp. 189 (2009)
URL http://arxiv.org/abs/0901.0002

[79] ATLAS collaboration, *ATLAS Run 1 Pythia8 tunes* (2014)
URL http://cds.cern.ch/record/1966419

[80] Apolinário L. et al., *An analysis of the influence of background subtraction and quenching on jet observables in heavy-ion collisions*, JHEP, **1312**, p. 022 (2013)
URL http://link.springer.com/article/10.1007%2FJHEP02%282013%29022

[81] Spousta M., *Jet Quenching at LHC*, Mod. Phys. Lett., **A28**, p. 1330017 (2013)
URL http://arxiv.org/abs/1305.6400

[82] ATLAS Collaboration, *Pile-up corrections for jets from proton-proton collisions at* $\sqrt{s} = 7$ TeV *in ATLAS in 2011* (2012), ATLAS-CONF-2012-064
URL http://cdsweb.cern.ch/record/1459529

[83] Cacciari M. and Salam G.P., *Pileup subtraction using jet areas*, Phys.Lett.B, **659**, pp. 119 (2008)
URL http://arxiv.org/abs/0707.1378

[84] Soyez G. et al., *Pileup subtraction for jet shapes*, Phys. Rev. Let., **110**, p. 162001 (2013)
URL http://arxiv.org/abs/1211.2811

[85] Kirschenmann H. (CMS Collaboration), *Jet performance in CMS*, PoS EPS-HEP2013, p. 433 (2013)
URL https://cds.cern.ch/record/1627818

[86] CMS Collaboration, *Jet performance in CMS* (2010)
URL https://cds.cern.ch/record/1247373

[87] Krohn D. et al., *Jet Cleansing: Pileup Removal at High Luminosity*, Phys. Rev. D, **90**, p. 065020 (2014)
URL http://arxiv.org/abs/1309.4777

[88] Bertolini D. et al., *Pileup Per Particle Identification*, JHEP, **1410**, p. 59 (2014)
URL http://link.springer.com/article/10.1007%2FJHEP10%282014%29059

[89] Cacciari M. et al., *SoftKiller, a particle-level pileup removal method*, The European Physical Journal C (2015)
URL http://link.springer.com/article/10.1140/epjc/s10052-015-3267-2

[90] Krohn D. et al., *Jet Trimming*, JHEP, **1002**, p. 084 (2010)
URL http://link.springer.com/article/10.1007%2FJHEP02%282010%29084

[91] Ellis S.D. et al., *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, Phys. Rev., **D81**, p. 094023 (2010)
URL http://arxiv.org/abs/0912.0033

[92] Butterworth J.M. et al., *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett., **100**, p. 242001 (2008)
URL http://arxiv.org/abs/0802.2470

[93] Cacciari M. et al., *The Catchment Area of Jets*, JHEP, **04**, p. 005 (2008)
URL http://arxiv.org/abs/0802.1188

[94] Cacciari M. et al., *FastJet User Manual*, Eur. Phys. J. C, **72**, p. 1896 (2012)
URL http://arxiv.org/abs/1111.6097

[95] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at sqrt(s)=8 TeV using the ATLAS detector* (2015), CERN-PH-EP-2015-206
URL http://arxiv.org/abs/1510.03823

[96] *FastJet Contrib Project*
URL http://fastjet.hepforge.org/contrib/

[97] Bahr M. et al., *Herwig++ Physics and Manual*, Eur.Phys.J.C, **58**, pp. 639 (2008)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-008-0798-9

[98] Pumplin J. et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP, **0207**, p. 012 (2002)
URL http://arxiv.org/abs/hep-ph/0201195

[99] Gieseke S. et al., *Colour reconnections in Herwig++*, Eur. Phys. J. C, **72**, p. 2225 (2012)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-012-2225-5

[100] Lai H.L. et al., *New parton distributions for collider physics*, Phys. Rev. D, **82**, p. 074024 (2010)
URL http://arxiv.org/abs/1007.2241

[101] Soper D.E. and Spannowsky M., *Finding top quarks with shower deconstruction*, Phys. Rev. D, **87**, p. 054012 (2013)
URL http://arxiv.org/abs/1211.3140

[102] Lai H.L. et al., *Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions*, Eur. Phys. J. C, **12**, pp. 375 (2000)
URL http://arxiv.org/abs/hep-ph/9903282

[103] Cacciari M. and Salam G.P., *Dispelling the N^3 myth for the Kt jet-finder*, Phys. Lett. B, **641**, pp. 57 (2006)
URL http://inspirehep.net/record/700668

[104] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at sqrt(s) = 7 TeV*, Eur. Phys. J. C, **73**, p. 2304 (2013)
URL http://arxiv.org/abs/1112.6426

[105] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, Journal of Instrumentation, **6** (2011)
URL http://iopscience.iop.org/article/10.1088/1748-0221/6/11/P11002/meta

[106] Altheimer A. et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, J. Phys. G, **39**, p. 063001 (2012)
URL http://arxiv.org/abs/1201.0008

[107] Abdesselam A. et al., *Boosted objects: a probe of beyond the Standard Model physics*, Eur. Phys. J. C, **71**, p. 1661 (2011)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-011-1661-y

[108] Gallicchio J. and Schwartz M.D., *Quark and Gluon Tagging at the LHC*, Phys. Rev. Lett., **107**, p. 172001 (2011)
URL http://arxiv.org/abs/1106.3076

[109] ATLAS Collaboration, *Performance and Calibration of the JetFitterCharm Algorithm for c-Jet Identification* (2015), aTL-PHYS-PUB-2015-001
URL https://cds.cern.ch/record/1980463

[110] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments* (2015), aTL-PHYS-PUB-2015-006
URL https://cds.cern.ch/record/2002609

[111] Nason P., *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP, **0411**, p. 040 (2004)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2004/11/040/meta

[112] Frixione S. et al., *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP, **0711**, p. 070 (2007)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2007/11/070/meta

[113] Alioli S. et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP, **1006**, p. 043 (2010)
URL http://link.springer.com/article/10.1007%2FJHEP06%282010%29043

[114] Lange D.J., *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research A, **462**, p. 152–155 (2001)
URL http://www.slac.stanford.edu/~lange/EvtGen/evtgen.pdf

[115] Skands P.Z., *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D, **82**, p. 074018 (2010)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.82.074018

[116] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for gap fraction and jet multiplicity observables in top-antitop events* (2014), aTL-PHYS-PUB-2014-005
URL https://cds.cern.ch/record/1703034

[117] Alieva M. et al., *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*, Computer Physics Communications, **182**, p. 1034–1046 (2011)
URL http://www.sciencedirect.com/science/article/pii/S0010465510005333

[118] Butterworth J. et al., *Jimmy Generator*, web page, last revision 31.10.2005, cited 19.3.2016
URL http://jimmy.hepforge.org/news

[119] Frixione S. et al., *Matching NLO QCD and parton showers in heavy flavour production*, JHEP, **0308**, p. 007 (2003)
URL http://arxiv.org/abs/hep-ph/0305252

[120] Mangano M.L. et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP (2002)
URL http://arxiv.org/abs/hep-ph/0206293

[121] Mangano M.L. et al., *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, JHEP, **0701**, p. 013 (2007)
URL http://arxiv.org/abs/hep-ph/0611129

[122] Kersevan B.P. and Richter-Was E., *The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5*, CPC, **184**, p. 919–985 (2013)
URL http://arxiv.org/abs/hep-ph/0405247

[123] Alioli S. et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP, **0909**, p. 111 (2009)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2009/09/111/meta

[124] Re E., *Single-top Wt-channel production matched with parton showers using the POWHEG method*, EPJC, **71**, p. 1547 (2011)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-011-1547-z

[125] Kidonakis N., *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D, **83**, p. 091503 (2011)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.83.091503

[126] Kidonakis N., *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys. Rev. D, **82**, p. 054018 (2010)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.82.054018

[127] Kidonakis N., *NNLL resummation for s-channel single top quark production*, Phys. Rev. D, **81**, p. 054028 (2010)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.81.054028

[128] ATLAS Collaboration, *Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at sqrt(s) = 7 TeV*, EPJC, **73**, p. 2261 (2013)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-012-2261-1

[129] Gleisberg T. et al., *Event generation with SHERPA 1.1*, JHEP, **0902**, p. 007 (2009)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2009/02/007/meta

[130] Campbell J.M. et al., *Vector boson pair production at the LHC*, JHEP, **1107**, p. 018 (2011)
URL http://link.springer.com/article/10.1007%2FJHEP07%282011%29018

[131] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at sqrts=8 TeV with the ATLAS detector* (2014), aTLAS-CONF-2014-058
URL https://cdsweb.cern.ch/record/1951336

[132] ATLAS Collaboration, *A search for ttbar resonances using lepton-plus-jets events in proton-proton collisions at sqrt(s)=8 TeV with the ATLAS detector*, JHEP, **08**, p. 148 (2015)
URL http://link.springer.com/article/10.1007%2FJHEP08%282015%29148

[133] ATLAS Collaboration, *Performance of the Electron and Photon Trigger in p-p Collisions at sqrt(s) = 7 TeV* (2011), aTLAS-CONF-2011-114
URL https://cds.cern.ch/record/1375551

[134] ATLAS Collaboration, *Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV*, EPJC, **75**, p. 120 (2015)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-015-3325-9

[135] Buckley A., *Rivet Routine ATLAS_2015_I1397637*, web page, last revision 16.4.2016, cited 26.4.2016
URL https://rivet.hepforge.org/trac/browser/src/Analyses/ATLAS_2015_I1397637.cc

[136] Adye T., *PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*, p. 313–318 (2011), cERN-2011-006
URL https://cdsweb.cern.ch/record/1306523

[137] Hoecker A. and Kartvelishvili V., *SVD Approach to Data Unfolding*, Nucl. Instrum. Meth. A, **372**, pp. 469 (1996)
URL http://arxiv.org/abs/hep-ph/9509307

[138] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. C, **74**, p. 3071 (2014)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-014-3071-4

[139] ATLAS Collaboration, *ATLAS Public Results - JetEtmissApproved2013JESUncertainty* (2013), web page, last revision 10.5.2013, cited 15.3.2016
URL https://twiki.cern.ch/twiki/bin/view/\AtlasPublic/JetEtmissApproved2013JESUncertainty

[140] ATLAS Collaboration, *Jet mass and substructure of inclusive jets in sqrt(s) = 7 TeV pp collisions with the ATLAS experiment*, JHEP, **1205**, p. 128 (2012)
URL http://link.springer.com/article/10.1007%2FJHEP05%282012%29128

[141] ATLAS Collaboration, *Large-R (anti-kt R=1.0) JES and JMS for 2012 data* (2014), web page, last revision 17.8.2014, cited 15.3.2016
URL https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2014LargeRJES

[142] White C.D. et al., *Isolating Wt production at the LHC*, JHEP, **0911**, p. 074 (2009)
URL http://iopscience.iop.org/article/10.1088/1126-6708/2009/11/074/meta

[143] Botje M. et al., *The PDF4LHC Working Group Interim Recommendations* (2011)
URL http://arxiv.org/abs/1101.0538

[144] Ball R.D. et al., *Parton distributions with LHC data*, Nuclear Physics B, **867**, p. 244–289 (2012)
URL http://arxiv.org/abs/1207.1303

[145] Campbell J.M. and Ellis R.K., *MCFM for the Tevatron and the LHC*, Nucl. Phys. Proc. Suppl., **205-206**, pp. 10 (2010)
URL http://arxiv.org/abs/1007.3492

[146] H1 and Collaborations Z., *Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA*, JHEP, **1001**, p. 109 (2010)
URL http://link.springer.com/article/10.1007%2FJHEP01%282010%29109

[147] Carli T. et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: the APPLGRID project*, EPJC, **66**, pp. 503 (2010)
URL http://link.springer.com/article/10.1140%2Fepjc%2Fs10052-010-1255-0

[148] Hsu L. and Hsu J.P., *The physical basis of natural units and truly fundamental constants*, Eur. Phys. J. Plus, **127**, p. 11 (2012)
URL http://link.springer.com/article/10.1140%2Fepjp%2Fi2012-12011-5

[149] Salam G.P., *Towards Jetography*, Eur. Phys. J. C, **67**, pp. 637 (2010)
URL http://arxiv.org/abs/0906.1833

[150] Catani S. et al., *Longitudinally-invariant kT-clustering algorithms for hadron-hadron collisions*, Nuclear Physics B, **406**, pp. 87 (1993)
URL http://www.sciencedirect.com/science/article/pii/055032139390166M

[151] Ellis S.D. and Soper D.E., *Successive Combination Jet Algorithm For Hadron Collisions*, Phys. Rev. D, **48**, pp. 3160 (1993)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.48.3160

[152] Dokshitzer Y. et al., *Better Jet Clustering Algorithms*, JHEP, **9708**, p. 001 (1997)
URL http://arxiv.org/abs/hep-ph/9707323

[153] Wobisch M. and Wengler T., *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering* (1999)
URL http://arxiv.org/abs/hep-ph/9907280

[154] Wobisch M., *Measurement and QCD analysis of jet cross-sections in deep inelastic positron proton collisions at s**(1/2) = 300-GeV* (2000)
URL http://inspirehep.net/record/538471

[155] Cacciari M. et al., *The anti-k_t jet clustering algorithm*, arXiv:0802.1189 (February 2008), JHEP 0804:063,2008
URL http://arxiv.org/abs/0802.1189

[156] Thaler J. and Tilburg K.V., *Identifying Boosted Objects with N-subjettiness*, JHEP, **1103**, p. 015 (2011)
URL http://arxiv.org/abs/1011.2268

[157] Thaler J. and Tilburg K.V., *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, JHEP, **02** (2012)
URL http://arxiv.org/abs/1108.2701

[158] Butterworth J.M. et al., *WW scattering at the CERN LHC*, Phys. Rev. D, **65**, p. 096014 (2002)
URL http://journals.aps.org/prd/abstract/10.1103/PhysRevD.65.096014

[159] Larkoski A.J. et al., *Energy Correlation Functions for Jet Substructure*, JHEP, **06**, p. 108 (2013)
URL http://arxiv.org/abs/1305.0007

List of Tables

List of Figures

List of Abbreviations

ϕ	azimuth, see Eq. A.5
C/A	Cambridge/Aachen (jet algorithm)
$\sqrt{s}$	center-of-mass energy
$\Delta\phi$	distance in the transverse plane, see Eq. A.14
ΔR	standard distance in the $y-\phi$ space, see Eq. A.12
$\sqrt{d_{12}}$	k_t splitting scale, see App. C.2
$E_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy, see Sec. 2.3.3
L	integrated luminosity
m	mass
m_δ	difference between transverse mass and transverse momentum, see Eq. A.10
μ_F	factorization scale
μ	average number of pp interactions per bunch crossing
μ_R	renormalization scale
pp	proton-proton
$p\bar{p}$	proton-antiproton
p_{T}	transverse momentum with respect to the beam axis of colliding hadrons
y	rapidity, see Eq. A.8
CHS	Charged hadron subtraction
EWT	Electroweak Theory
ISR	Initial state radiation
JER	Jet energy resolution, see Sec. 2.3.2
JES	Jet energy scale, see Sec. 2.3.2
LCW	local calibration weighting, see Sec. 2.3.2
LHC	Large Hadron Collider
LO	leading order (in perturbation expansion)
MC	Monte Carlo

ME	Matrix Element
MPI	Multiple parton interactions
NLO	next-to-leading order (in perturbation expansion)
NNLL	next-to-next-to-leading logarithm
NNLO	next-to-next-to-leading order (in perturbation expansion)
PDF	Parton Distribution Function
PS	Parton Shower
QCD	Quantum Chromodynamics
RMS	root mean square
SM	Standard Model
TRT	Transition Radiation Tracker
UE	Underlying Event