Probing top quark and Higgs boson production in multi-jet events at the LHC with the ATLAS detector

Claudia Bertella

CPPM-T-2013-04

AIX-MARSEILLE UNIVERSITÉ

Faculté des sciences

Thèse de doctorat

Ecole Doctorale 352

Physique et sciences de la matière

Mention: Physique des particules et astroparticules

Présentée par

Claudia Bertella

en vue d'obtenir le grade de docteur d'Aix-Marseille Université

Probing top quark and Higgs boson production in multi-jet events at the LHC with the ATLAS detector

---

Soutenue le 30 Septembre 2013 devant le jury composé de

Prof.	J. Donini	Rapporteur
Dr.	L. Feligioni	Co-directeur de thèse
Dr.	E. Kajfasz	Président du jury
Dr.	F. Parodi	Examinateur
Prof.	M. Talby	Directeur de thèse
Prof.	R. Tenchini	Rapporteur

To my family

Résumé

Le cadre théorique traité dans la présente thèse est le Modèle Standard (SM) de la physique des particules. Ce modèle décrit les constituants élémentaires de la matière et leurs interactions, où trois des quatre interactions fondamentales, les interactions électromagnétique, faible et forte, sont décites dans le cadre de la théorie de jauge ayant comme groupe de symétrie, le groupe $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$. Les nombreuses mesuress expérimentales, réalisées dans les différentes expériences au cours des cinquantes dernières années, ont permis de tester avec succès le SM avec avec une haute précision (au niveau des ses corrections quantiques d'ordres supérieurs). Malgré son succès, le SM reste incomplet et présente quelques faiblesses et des lacunes. Par exemple le SM ne décrit pas l'interaction gravitationnelle au niveau des particules élémentaires, il n'explique pas leur hiérarchie en masse il ne propose aucune piste pour la matière et l'énergie dites noires, il considère les neurtinos comme des particules sans masse, etc. Pour répondre à ces interrogations et à fin de découvrir la théorie fondamentale des constituants ultimes de notre univers, le LHC a été construit au CERN.

Ce manuscrit est composé d'une introduction, de cinq chapitres et d'une conclusion.
Le premier chapitre du manuscrit est dédié á la description du SM. Les particules fondamentales sont ainsi divisées, selon cette théorie, en deux catégories principales : les fermions de spin 1/2 (particules qui composent la matière) et les bosons de spin 1 (médiateurs des interactions). A chaque interaction est associé un boson vecteur. L'interaction électromagnétique, qui s'exerce entre deux particules chargées, a comme médiateur le photon. L'interaction faible, qui s'exerce entre deux particules possédant un nombre quantique d'isospin faible est véhiculée par les bosons vecteurs $W^{\pm}$ et Z^0. L'interaction forte, active entre deux particules possédant le nombre quantique d'hyper-charge ou charge de couleur a comme médiateur les gluons.
Les fermions sont divisés en six leptons et six quarks, groupés en trois familles. Chaque famille de leptons est constituée d'un lepton chargé électriquement sensible á l'interaction électromagnétique et à l'interaction faible, et d'un lepton de charge électrique nulle, le neutrino, sensible seulement á l'interaction faible. Les quarks á la différence des leptons possèdent, comme indiqué plus haut, en plus d'une charge électrique fractionnaire et d'un nombre quantique d'isospin faible, un autre nombre quantique, la charge de couleur. Les six quarks observés sont: (up, down), (charm, strange), (top, bottom). Les quarks sont, à l'exception du quark top, des particules confinées dans des états

hadroniques. Ils sont confinées soit dans des mesons (états liés quark-antiquark) ou alors dans des baryons (états composés de trois quarks).
Le quark top est produit dans les collisioneurs hadroniques en deux modes : en quark célibataire ou en une paire top anti-top. Il se désintègre presque á 100% des cas en boson W et en quark b. Selon les modes de désintégrations du boson W, trois canaux de désintégration de la paire $t\bar{t}$ peuvent être identifiés: "semi-leptionic", ou un des bosons W se désintègre en lepton et neutrino, tandis que l'autre va en paire de quarks; "leptonique" ou les deux bosons W se désintègrent en leptons et neutrinos ; "hadronique" ou les deux bosons W se désintègrent en paires de quarks.
Les principales caractéristiques et propriétés du quark top ainsi que ses modes de production et de désintégration sont décrits dans ce chapitre.
Cette thèse porte également sur le boson de Higgs. Cette particule est á la base du mécanisme du même nom, proposé par P. Higgs, F. Englert et R. Brout, en 1964. Ce mécanisme a été introduit comme une possible solution au problème de la masse des bosons vecteurs. Ce mécanisme postule l'existence d'un nouveau champ scalaire conplexe qui conduit á la brisure spontanée de la symétrie électrofaible. Salam et Weinberg ont introduit le potentiel de Higgs dans la Lagrangien du SM, ainsi qu'un terme d'interaction du nouveau champ avec les fermions pour générer leurs masses. Le mode dominant de production du boson de Higgs au LHC est la fusion de gluons, les modes sous-dominants sont: la production associée á un boson vecteur, la fusion de bosons vecteurs et la production associée á une paire de quarks top. Selon la masse du boson de Higgs il est possible d'étudier différents modes de désintégrations. Dans la région de basses masses, le mode dominant est la désintégration en paire de quarks beaux.

Le 4 Juillet 2012, les collaborations ATLAS et CMS ont annoncés la découverte d'une nouvelle particule avec une masse de 125 GeV. Après une étude détaillée de la nouvelle résonance, les deux expériences ont confirmés la nature de la nouvelle particule, l'identifiant ainsi á un boson de Higgs compatible avec celui prédit par le SM. Malgré cette observation, différentes questions sur cette particule demeurent encore sans réponse, comme par exemple la nature de cette nouvelle particule, est-ce le boson de Higgs du SM ou un des nombreux bosons prédits par des théorie au-delá du SM. Davantage d'informations sur les caractéristiques de ce nouveau boson, notamment ses couplages avec les autres particules, sont nécessaires pour répondre á la question sur sa nature.

La découverte du boson de Higgs a été rendue possible grâce au LHC (Large Hadronic Collider) ; un collisionneur proton-proton circulaire de 27 Km de circonférence avec plusieurs points de croisements de faisceaux de protons auprès desquels sont placés différentes expériences, dont les expériences généralistes: ATLAS et CMS. Le LHC est conçu pour produire des collisions avec une énergie nominale au centre de masse de l'ordre de $\sqrt{s} = 14$ TeV et une luminosité de $10^{-34} cm^{-2} s^{-1}$. Il donne accès á des processus á l'échelle du TeV. En 2010 et 2011 le LHC a fournit des collisions proton-proton avec une énergie de $\sqrt{s} = 7$ TeV donnant la possibilité d'enregistrer plus de 5 fb^{-1} de données. En 2012 le LHC a fournit des collisions avec une énergie de $\sqrt{s} = 8$ TeV. La prochaine prise de données est prévu pour le début du 2015 avec une énergie au centre de masse de $\sqrt{s} = 13/14$

TeV. A coté des collisions proton-proton, le LHC fourni qussi des collisions d'ions lourds pour, en particulier, l'étude du plasma de quarks et gluons.
Le deuxième chapitre décrit le cadre expérimental de la thèse. Il passe en revu le collisionneur LHC et le détecteur ATLAS (A Torroidal LHC ApparatuS). Ce dernier a été conçu et optimisé pour détecter, reconstruire et identifier les particules produites lors des collisions proton-proton. Le programme scientifique de cette expérience couvre une large gamme de processus physiques du SM, et au-delá.

Un système de déclenchement efficace est nécessaire pour sélectionner les événements physiques produits par les collisions proton-proton. Le but du système d'acquisition de ATLAS est de filtrer et de sélectionner les événements de physique intéressants parmi le bruit du fond des interactions. Le système de déclenchement de ATLAS a fonctionné de façons efficace et fiable pendant les dernières prises des données. Il a une structure divisée en trois niveaux qui permet de reduire le flux de données á enregistrer de 1 GHz jusqu'á presque 500 Hz. Les événements de physique sont sélectionnés suite á l'identification des signatures de muons, électrons, photons, leptons tau, jets, b-jes et des méson B candidats.
Le troisième chapitre passe en revue le système de déclenchement ou "Trigger", en donnant une description détaillée des différentes parties ainsi que l'identification des principales signatures. Une partie de ce chapitre est dédiée á la description de l'algorithme d'identification des jets issus des quarks b (b-tagging). La séparation des jets légers des jets lourds représente un ingrédient important pour de nombreuses analyses de physique, en particulier dans le secteur de la physique du quark top, dans la recherche du boson de Higgs ou la recherche de nouvelle physique. L'implémentation de l'algorithme de b-tagging au niveau du système de déclenchement peut améliorer notablement l'indentification et la sélection des événements contenant des jets issus de quarks b (b-jets), parmi les événements multi-jets, comme par exemple les évenements top-anti top dans le canal complètement hadronique. Deux catégories de b-jets sont utilisées dans le système de déclenchement de l'expérience ATLAS: une qui utilise les b-riétés des hadrons b et l'autre qui cherche la présence d'un muon dans les jet issues de la désintégration de hadrons b. Le premier déclenchement est utilisé dans les analyses de physique, le deuxième est utilisé pour les études de calibration des algorithmes de b-tagging. Les deux déclenchements sont contrôlés au cours de la prise de données pour s'assurer que les algorithmes du système de déclenchement ont été bien configurés. Le contrôle du système du déclenchement se fait en deux étapes: "online" et "offline". Au cours de ma thèse, j'ai travaillé sur le contrôle du b-jet trigger et développé le système de contrôle des différents algorithmes utilisés. Grâce au système de contrôle, de mauvaises configurations d'algorithme de débranchement sont isolées et puis rapidement résolues en donnant la possibilité d'enregistrer les événements avec le déclenchement de jet b.

La première analyse effectuée au cours de ma thèse traite de la mesure de la section efficace de production des paires de quarks top anti-top dans le mode de désintégration complètement hadronique utilisant les données enregistrées par l'expérience ATLAS en 2011 avec une énergie

dans le centre de masse de $\sqrt{s} = 7$ TeV correspondant á une luminosité intégrée de 4.7 fb^{-1}. Le canal complètement hadronique a l'avantage d'être caractérisé par un rapport d'embranchement de 46%, et une absence de leptons et d'énergie transverse manquante. Il souffre par contre d'un bruit de fond multi-jet QCD élevé. L'état final de ce canal est constitué par la présence de six jets, parmi lesquels deux b-jets issu de la désintégration des quarks top. Les événements intéressants sont sélectionnés en utilisant un système de déclenchement qui demande la présence d'au moins cinq jets avec une grande impulsion transverse. Différentes coupures sont ensuite appliquées sur les objets reconstruits pour sélectionner les évènements qui ont au moins six jets dans l'état final avec une grande impulsion transverse dans la région central du détecteur. Deux des six jets doivent aussi être identifiés comme des jets issus d'un quark b utilisant un algorithm d'étiquetage, b-tagging. La reconstruction de la topologie top-anti top est effectuée avec un algorithm de maximum de vraisemblance, qui utilise les informations cinématiques des objets sélectionnés pour déterminer la meilleure combinaison de six jets (la combinaison avec la plus haute vraisemblance est retenue). La distribution de la masse du top fournie par cet algorithm est choisie comme variable discriminante pour la mesure de la section efficace.
Le principal bruit de fond pour la production des paires des quark top-antitop est dû aux processus QCD avec au mois six de quarks/gluons dans l'état final. Ce bruit de fond n'est malheureusement pas bien connu et donc difficile á reproduire par simulation Monté Carlo. Il est estimé á partir des données. La méthode utilisée est basée sur l'observation de la distribution de la masse du quark top fourni par la méthode de reconstruction indépendamment de la présence des jets étiquetés b. Donc il est possible d'estimer la forme de la distribution de la masse du quark top dans une région enrichie de bruit de fond et de l'extrapoler après dans la région où la contribution du signal est grande. L'extraction de la section efficace est réalisé grâceà un ajustement de maximum de vraisemblance appliqué sur masse du quark top en utilisant comme modèle pour le signal, la prédiction obtenue par la simulation Monte-Carlo (MC@NLO). Pour le bruit de fond multi-jet l'estimation est obtenue avec les données avant la sélection des jets b dans les événements. Le résultat est comparé avec la prédiction du SM et aussi au mesures obtenues par ATLAS dans les autres canaux des désintégrations.

La deuxième analyse décrite dans cette thèse est la recherche du boson du Higgs produit en association avec une paire de quark top, où le boson de Higgs se désintègre en paires de quarks b, et le système top-anti-top se désintégrant dans le canal complétant hadronique. L'analyse est effectuée sur les mêmes données utilisées par la mesure de la section efficace top-anti top. La signature recherchée est caractérisée par la présence d'au moins huit jets dans l'état final, avec quatre jets étiquetés comme b-jet, deux sont produits par la désintégration des quarks top et deux issus de la désintégration du boson de Higgs. Á cause de la grande multiplicité des jets dans l'état final, cette analyse souffre d'une grande contribution de bruit de fond QCD. Elle utilise la même sélection d'objets utilisée pour la mesure de la section efficace. La reconstruction de la topologie de l'événement signal recherché est effectuée aussi avec la méthode d'ajustement de maximum de

vraisemblance. Cette méthode est appliquée seulement au système de top-anti top. La reconstruction des candidats Higgs est effectuée après la reconstruction du système top-anti top: les deux jets avec la plus grande impulsion transverse n'ayant pas contribué à l'ajustement de maximum de vraisemblance sont considérés comme candidats des produits de désintégration du boson de Higgs. La variable discriminante choisie pour cette analyse est la masse invariante des pairs de jets retenues comme possibles candidats.
Une analyse multivariée MVA est utilisée pour sélectionner avec une grande pureté les évènements du signal. Un arbre de décision boosté (BDT) est ainsi entraîné utilisant de nombreuses variables cinématiques du système top-anti top, mais aucune information sur boson de Higgs n'est introduite afin d'éviter tout biais dans la reconstruction de la masse invariante de la paire de jets candidats. Deux bruits de fond principaux peuvent être distingués: la production des paires de quarks top avec la présence des jets supplémentaires et la production QCD. Le premier bruit de fond est estimé en utilisant le modèle Monte-Carlo (ALPGEN) différenciant la production de jets légers des jets lourds. Le deuxième bruit de fond est estimé avec les données utilisant une méthode appelée ABCD. L'idée est de trouver deux variables dé-corrélée permettant de séparer les données en différentes catégories ou régions, celles ressemblant au bruit de fond et celles ressemblant au signal, et d'utiliser les régions pauvres en signal pour estimer la contribution du bruit de fond dans la région du signal. Dans le cas spécifique de cette analyse, les variables sélectionnées sont; la multiplicité des jet b et la variable de sortie du MVA. Quatre bins en multiplicité des jets b ont été sélectionnés en donnant la possibilité d'identifier huit régions, avec différentes concentrations en signal et en bruit de fond. L'une de ces régions est identifiée comme région de validation de la méthode et deux régions comme régions de signal. Le Chapitre 5 décrit en détail toutes les étapes suivies dans l'analyse, passant en revue l'analyse multivariée, la méthode pour l'estimation du bruit du fond QCD ainsi que sa validation en utilisant les comparaisons données-Monte Carlo.

Contents

Introduction

The Standard Model (SM) of particle physics is the quantum field theory that describes the properties and the interactions of the fundamental particles. It unifies in a single theoretical framework three of the four fundamental interactions: electromagnetic, weak and strong interactions. The SM was developed in the second half of the twentieth century and has so far been tested thoroughly and up to its quantum corrections by many experiments at various accelerator centers, such as CERN, Fermilab, DESY, SLAC, KEK. Over the years all the particles predicted by the SM were discovered by different experiments, such as the W and Z weak bosons in 1983 by UA1 and UA2 experiments at CERN, and the top quark in 1995 by CDF and DØ at Fermilab. The missing piece of the SM up to last year was the Higgs boson. With the advent of the Large Hadron Collider (LHC) at CERN and after three years of operation, the discovery, beyond any reasonable doubts, of a (the) Higgs boson was announced in July 2012 by ATLAS and CMS collaborations. However there are still some inconsistencies between the SM and the experiments, such as neutrino masses. Moreover the nature of the dark matter and other unanswered fundamental questions as well as few conceptual weaknesses point toward the existence of new physics beyond the SM. Many effort are thus necessary to investigate and finally complete the SM of particle physics and discover new physics beyond the present standard picture. An overview of the theoretical framework of this thesis is presented in Chapter 1.

The Chapter 2 is devoted to the description of the experimental framework: the LHC accelerator and the ATLAS detector. The LHC is a proton-proton collider located in Geneva, Switzerland at the European Organization for Nuclear Research (CERN). After a first period of commissioning, it began to deliver proton-proton collisions in 2010 at a center-of-mass energy of $\sqrt{s} = 900$ GeV. During 2011 and the 2012 the LHC delivered pp collisions at $\sqrt{s} = 7$ and 8 TeV respectively. Since Mars 2013 a shutdown period of two years has started in order to make the necessary accelerator's upgrades to increase the center-of-mass energy of the pp collisions up to 13 and then to 14 TeV, as well as the instantaneous luminosity up to 10^{34} cm^{-2} s^{-1}.

ATLAS (A Toroidal LHC ApparatuS) is one of the two large general-purpose experiments installed along the LHC ring. The detector design reflects the challenging physics program which ATLAS aims to access in order to improve the knowledge of high energy physics.

Several tasks, performed during my PhD, are treated in this dissertation. The first one consists

in the development and maintenance of the b-jet and μ-jet triggers. The monitoring of the trigger algorithms during and after the data-taking is crucial for physics analysis in order to insure good quality of the detector and data acquisition system behaviours. An overview of the complex trigger system used in the ATLAS detector is provided in Chapter 3, as well as the different steps in data quality validation through the online and offline monitoring.

The particle physics area investigated in this thesis work focuses on probing top quark and Higgs boson production in the multi-jet events at the LHC. The top quark is the heaviest particle discovered so far. Due to its mass and other connected proprieties, it plays a special role in particle physics. Several theoretical predictions, developed in the last century, indicate that new physics could appear in both the production and the decay processes of the top quark. The LHC provides a very extraordinary environment for the investigation of the top quark physics. With more than 25 fb^{-1} pp collisions delivered since 2011, the era of the precision measurement in the top quark sector has begun. Based on the $\sqrt{s} = 7$ TeV pp collisions, ATLAS experiment has measured the $t\bar{t}$ cross section in almost all the decay modes. The result of the combination of the single-lepton, di-lepton and all hadronic channels is $\sigma_{t\bar{t}} = 177 \pm 3\ (stat.) \pm 8\ (syst.) \pm 7\ (lumi.)$ reaching a precision of 5%.

This thesis reports in Chapter 4 on the measurement of the top quark pair production cross section in the fully hadronic final state performed with an integrated luminosity of 4.7 fb^{-1} recorded by the ATLAS detector in 2011 at $\sqrt{s} = 7$ TeV. This measurement is quite challenging due to the busy final state characterized by a nominal six jets multiplicity. This particular topology suffers from a significant background originating from the QCD multi-jet production, which represents the key ingredient for the cross section extraction in the fully hadronic top decay mode.

The second part of my physics analysis work focuses on the search for the SM Higgs boson produced in association with a top quark pair, in which the Higgs boson decays into bottom anti-bottom pair and the $t\bar{t}$ pair in the fully hadronic decay mode. As the $t\bar{t}$ cross section measurement, the main background is the QCD multi-jets production. Accurate studies were made on the modeling of this overwhelming background as well as on the reconstruction of the final state topology. A first preliminary analysis performed using an integrated luminosity of 4.7 TeV recorded by the ATLAS detector at $\sqrt{s} = 7$ TeV is presented in Chapter 5.

Chapter 1

Theoretical framework

1.1 The Standard Model of particle physics

The modern description of fundamental particles and their interaction is based on the gauge invariant theory known as the Standard Model (SM) [1]. The model was built in the second half of the last century to describe elementary particles, their interactions and their production mechanisms.

1.1.1 Elementary particles

The SM is the theory of particle physics which successfully describes in a single theoretical framework the electromagnetic, weak and strong interactions among elementary particles. The fundamental particles are divided into spin 1/2 fermions, matter elementary particles, and spin 1 bosons, which play the role of force carriers [2]. There are three fundamental interactions carried out by three different vector bosons:

- electromagnetic interaction: acting between electrically charged particles. The carrier of this interaction is the massless spin 1 neutral photon (γ). The theory of electromagnetic interactions (QED) [3] is based on the local $U(1)$ gauge symmetry group,
- weak interaction: acting between particles carrying weak isospin quantum number. The carriers of this interaction are three heavy spin 1 vector bosons W^+, W^-, Z^0. The theory of weak interactions is based on the local $SU(2)_L$ gauge of symmetry group. This symmetry is mathematically similar to the one used in quantum mechanics to describe the spin of particles, this explain why the word isospin is used. In particular the three interaction carriers constitute an isospin-1 triplet,
- strong interaction: acting between particles carrying strong hyper-charge or color. The theory describing the strong interaction (quantum chromodynamics or QCD [4]) is based

Interaction	Boson	Charge (e)	Mass (GeV/c^2)
Electromagnetic	γ	$< 1 \cdot 10^{-35}$	$< 1 \cdot 10^{-27}$
Weak	$W^{\pm}$	± 1	80.385 ± 0.015
	Z^0	0	91.1876 ± 0.0021
Strong	8 gluons	0	0

TABLE 1.1 Gauge bosons in the Standard Model. c is the speed of the light, in this dissertation it will be fixed to $c = 1$ [7].

on the local $SU(3)_C$ gauge symmetry group and the carriers of this interaction are eight spin 1 massless bosons called gluons. They carry color charges themselves, and are thus self-interacting. Only colorless bound states are invariant under transformation of $SU(3)_C$; thus quark colorless bound states can occur as $q\bar{q}$ (meson) or $qqq/\bar{q}\bar{q}\bar{q}$ (baryon). A quark, which has a color charge, cannot be observed as a free state. This physics phenomenon is known as the *confinement of color* [5] which comes from the fact that the QCD coupling α_S[1] is large at low energy (high distance), leading to the confinement of quarks inside color-neutral hadrons. At very high energy (very low distance) α_S is so small that quarks behave as free particles. This behavious is known as asymptotic freedom [6].

Table 1.1 summarizes the characteristics of the gauge bosons for each interaction.
All the particles are characterized by the charge quantum numbers they carry. Fermions carrying a color charge are called quarks, while those with zero color charge are called leptons. All fermions (quarks and leptons) are divided into isospin 1/2 left-handed doublets and isospin 0 right-handed singlets (see Table 1.2). Three quark left-handed doublets as well as three lepton left-handed doublets have been observed, while there are six right-handed quark singlets and three right-handed lepton singlets[2].

Each interaction is described by a local gauge theory, i.e. a theory that requires invariance under some set of local transformations.

1.1.2 The Standard Model

The SM is a gauge invariant (i.e. invariant under the space-time symmetry) and renormalizable theory, based on the local gauge symmetry group product $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, which is able to describe the strong and electroweak interactions [9].

[1] The QCD coupling constant g_S is related to α_S through the following formula $\alpha_S = \frac{g_S^2}{4\pi}$. α_S is a function that varies with energy scale Q. Below an energy threshold $Q = \Lambda_{QCD}$, the QCD cannot be considered as a perturbative theory anymore ($\Lambda_{QCD} \approx 200$ MeV):

$$\alpha_S(Q) = \frac{12\pi}{(33 - 2n_f) log \frac{Q^2}{\Lambda_{QCD}}} \tag{1.1}$$

where n_f is the number of quarks. This formula expresses the fact that $\alpha_S(Q)$ becomes small for large Q. The scale Λ_{QCD} is the scale at which α_S becomes infinitely large.

[2] Neutrinos are considered in the SM as massless particles and thus have only left-handed component.

	Left handed doublets	Right handed singlets
Leptons	$\binom{\nu_e}{e}_L \binom{\nu_\mu}{\mu}_L \binom{\nu_\tau}{\tau}_L$ $T = \frac{1}{2}, Y = -1$	$e_R\ \mu_R\ \tau_R$ $T = 0, Y = -2$
Quarks	$\binom{u}{d}_L \binom{c}{s}_L \binom{t}{b}_L$ $T = \frac{1}{2}, Y = \frac{1}{3}$	$u_R\ d_R\ c_R\ s_R\ b_R\ t_R$ $T = 0, Y(u_r, c_r, t_r) = \frac{4}{3}, Y(d_r, s_r, b_r) = -\frac{2}{3}$

TABLE 1.2 Classification of fermions predicted by the SM divided into leptons and quarks. The weak isospin T and weak hyper-charge Y are given for each group. The third component of weak isospin and the hyper-charge are related to the electric charge through the Gell-Mann-Nishigima formula: $Q = T_3 + \frac{Y}{2}$ [8].

1.1.2.1 Quantum Electrodynamic theory

The electromagnetic interaction between two electrically charged fermions is a good example of how the interaction between particles can be described using the gauge field framework [10]. The simple scenario of a free fermion is described by the Dirac free Lagrangian density for a massive fermion:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \tag{1.2}$$

where $\psi(x)$ is the quantum field of the fermion and m its mass. The gauge theory requires the Lagrangian density $\mathcal{L}$ to be invariant under the action of a local gauge transformation of the fermion field $\psi(x)$.

$$\psi \to \psi' = e^{iQ\theta(x)}\psi, \qquad \bar{\psi} \to \bar{\psi}' = e^{-iQ\theta(x)}\bar{\psi} \tag{1.3}$$

where $\theta(x)$ is the parameter of the transformation which depends on the space-time point $x = (t, \overrightarrow{x})$ and Q is the generator of the transformation which in the case of the electromagnetic interaction corresponds to the electric charge of the fermion in unit of the charge of an electron. The above transformation defines the local $U(1)$ gauge symmetry. The local gauge invariance of $\mathcal{L}$ requires the introduction of a new field describing a massless, spin-1 boson: the photon. In the case where the parameter θ is not space-time dependent the gauge transformation becomes a global transformation and the Dirac free Lagrangian density $\mathcal{L}$ is invariant under a global $U(1)$ transformation. When we apply a local transformation to the Lagrangian field density (1.2), the derivate term $\partial_\mu\psi$ is not covariant:

$$\partial_\mu\psi \to e^{iQ\theta(x)}(\partial_\mu + iQ\partial_\mu\theta(x))\psi \tag{1.4}$$

To eliminate the $\partial_\mu\theta(x)$ term, we introduce the covariant derivative:

$$\partial_\mu \rightarrow \mathcal{D}_\mu \equiv \partial_\mu + iQ\mathcal{A}_\mu \tag{1.5}$$

where $\mathcal{A}_\mu$ is a spin-one field representing the spin-one boson of the electromagnetic quadri-potential. The transformation property of the A_μ four-vector field is:

$$\mathcal{A}_\mu \rightarrow \mathcal{A}_\mu - \frac{1}{e}\partial_\mu\theta(x) \tag{1.6}$$

The Lagrangian density, invariant under local $U(1)$ transformation, becomes:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\mathcal{D}_\mu - m)\psi = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \mathcal{A}_\mu eQ(\bar{\psi}\gamma^\mu\psi) - \frac{1}{4}\mathcal{F}_{\mu\nu}\mathcal{F}^{\mu\nu} \tag{1.7}$$

where the first term is the free fermion Lagrangian (Equation 1.2), the second, introduced by the covariant derivate, represents the interaction between a fermion of electric charge Qe and a photon with a vertex factor $-ieQ\gamma^\mu$ and the third is the photon kinematic energy where $F_{\mu\nu}$ is the electromagnetic tensor defined as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

1.1.2.2 Electroweak theory

The electromagnetic and weak interactions are unified in a single electroweak (EW) theory which describes the interactions among quarks and leptons mediated by four vector bosons. In the EW theory $(SU(2)_L \times U(1)_Y)$ each particle is identified by the quantum numbers (T, T_3) for $SU(2)_L$ group and Y for the $U(1)_Y$ group. T is called the weak isospin and Y the hyper-charge [4]. The EW theory is formulated using a Lagrangian invariant under the local $SU(2)_L \otimes U(1)_Y$ gauge group symmetry transformation. The elementary fields described by the EW model are:

- 4 gauge boson fields, $W^i_\mu(x)$ $(i = 1, 2, 3)$ one for each generator of the $SU(2)_L$ gauge group symmetry and the boson fields $B_\mu(x)$ associated to the $U(1)_Y$ generator,
- left-handed quark fields $Q^f_L(x)$ which are doublets of the $SU(2)_L$ group, $f = 1, 2, 3$ is the flavour index,
- right-handed quark fields $U^f_R(x)$ and $D^f_R(x)$, which are singlets of the $SU(2)_L$ group,
- left-handed lepton fields $L^f_L(x)$ ($SU(2)_L$ doublets), and right-handed $L^f_R(x)$ lepton fields ($SU(2)_L$ singlets).

The EW Lagrangian density of quark fields for instance is written as:

$$\mathcal{L} = i\bar{Q}_L^f \not{D}' Q_L^f + i\bar{U}_R^f \not{D} U_R^f + i\bar{D}_R^f \not{D} D_R^f + \mathcal{L}_{gauge} \tag{1.8}$$

The covariant derivate $\not{D}$ and $\not{D}'$ are defined as:

$$\not{D} = \gamma^\mu D_\mu = \gamma^\mu(\partial_\mu + \frac{ig}{2} W_\mu^i \cdot \tau^i + ig' \frac{Y}{2} B_\mu) \tag{1.9}$$

$$\not{D}' = \gamma^\mu D'_\mu = \gamma^\mu(\partial_\mu + ig' \frac{Y}{2} B_\mu) \tag{1.10}$$

Using this definition the three first terms of the equation (1.8) can be written as:

$$\begin{aligned} \mathcal{L}(quark, gauge) = \bar{Q}_L^f \gamma^\mu (i\partial_\mu - g\frac{W_\mu^i \cdot \tau^i}{2} - \frac{1}{6} g' B_\mu) Q_L^f + \\ \bar{U}_R^f \gamma^\mu (i\partial_\mu - \frac{2}{3} g' B_\mu) U_R^f + \bar{D}_R^f \gamma^\mu (i\partial_\mu + \frac{1}{3} g' B_\mu) D_R^f \end{aligned} \tag{1.11}$$

where τ^i $(i = 1,\ 2,\ 3)$ are the three Pauli matrices, g and g' are fermion-boson coupling constants associated respectively to W_μ^i and B_μ. The three terms of Equation 1.11 describe the interactions of vector bosons with left-handed and right-handed quarks.
The $\mathcal{L}$(gauge), term of Equation 1.8 describes the kinematic term of the free gauge W_μ^i and B_μ fields and is written as:

$$\mathcal{L}(\text{gauge}) = -\frac{1}{4} W_{\mu\nu} \cdot W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} \cdot B^{\mu\nu} \tag{1.12}$$

where

$$W_{\mu\nu}^i = \partial_\nu W_\mu^i - \partial_\mu W_\nu^i + g(W_\mu \times W_\nu)^i, \qquad B_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu \tag{1.13}$$

The Lagrangian (1.8) describes four massless vector bosons and massless quarks. Similar electroweak Lagrangian density can be written for the leptons with the exception that there are no right-handed component for neutrinos. The fermion mass terms are not allowed in the Lagrangian because these are forbidden under the EW symmetry transformation. Therefore the model doesn't describe the real observed particles because in nature there is only one massless vector boson and the fermions (quarks and charged leptons) are massive particles[3].

1.1.3 Quantum Chromodinamic theory

Quantum Chromodynamics (QCD) is the gauge field theoretical framework that describe strong interaction among quarks and gluons, the subconstituants of hadrons. It was constructed in analogy

[3]Neutrinos have also a non-zero mass. This is not accounted for in the SM but have been measured experimentally through neutrinos observation experiments [11].

with QED but with a more complicated gauge group. QCD is a non-abelian[4] gauge theory of colour charged fermions, invariant under $SU(3)_c$ colour symmetry group. Colour charge, carried by quarks and gluons, mediators of strong interaction, is an addional quantum number (degree of freedom) introduced by Han and Nambu [13] in order to reconcile the baryon spectrum of the quark model [14–16] with the spin-statistic theorem[5]. The QCD Lagrangian is invariant under transformation of the non-Abelian $SU(3)_C$ group. The quanta of the $SU(3)_C$ gauge field are the gluons. The $\mathcal{L}$(QCD) Lagrangian of the QCD theory is defined as:

$$\mathcal{L}(\text{QCD}) = \bar{\psi}_i(i\gamma^\mu D_\mu - m\delta_{ij})\psi_j - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{1.14}$$

where ψ_i is the quark field and D_μ is the covariant derivate equal to:

$$D_\mu = \partial_\mu + \frac{i}{2}g_S G^a_\mu \lambda^a \tag{1.15}$$

where g_S is the strong coupling constant, G^a_μ, $a = 1, ...8$, are the eight gluons gauge fields, mediators of the strong interaction. λ^a, $a = 1, ...8$, are the eight generators of the $SU(3)_C$ gauge symmetry group. They are represented by eight traceless matrices called Gell-Mann matrices.
$G^a_{\mu\nu}$ is the gauge invariant gluonic field strength tensor, equal to:

$$G_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_S f^{abc} G^b_\mu G^c_\nu \tag{1.16}$$

where f^{abc} is the structure constant of the $SU(3)_C$. The Lagrangian $\mathcal{L}$(QCD) is invariant under the combined transformations:

$$\psi \rightarrow e^{-\frac{i}{2}\vec{\theta}(x)\vec{\lambda}}\psi \tag{1.17}$$

$$G^a_\mu \rightarrow G^a_\mu - \frac{1}{g_S}\partial_\mu \theta^a(x) - f^{abc}\theta^b(x)G^c_\mu \tag{1.18}$$

where $\theta^a(x)$, $a = 1, ...8$, are the parameters of the $SU(3)_C$ gauge group symmetry.

1.1.4 The Higgs mechanism

A possible solution to the problem of the gauge boson mass was proposed by P. Higgs [17–19], F. Englert and R. Brout [20] in 1964. They proposed a spontaneous symmetry breaking mechanism, the so called Higgs mechanism, through the introduction of a complex scalar field Φ, as a possible way to give mass to vector bosons. Salam and Weinberg introduced the Higgs potential in the SM Lagrangian, as well as the interaction terms of the fermions with the remaining scalar field after the spontaneous symmetry breaking (Higgs field) in order to generate the fermionic masses. The

[4] The non-abelian groups is a groups which elements do not commute[12]

[5] The lightest excited state of a nucleon is a spin 3/2 particle with charge +2, Δ^{++}. This particle is interpreted as a *uuu* bound state with zero orbital angular momentum and all the three spin quark projections parallel. This state, with the three fermions in the same state, violates the Pauli exclusion principal.

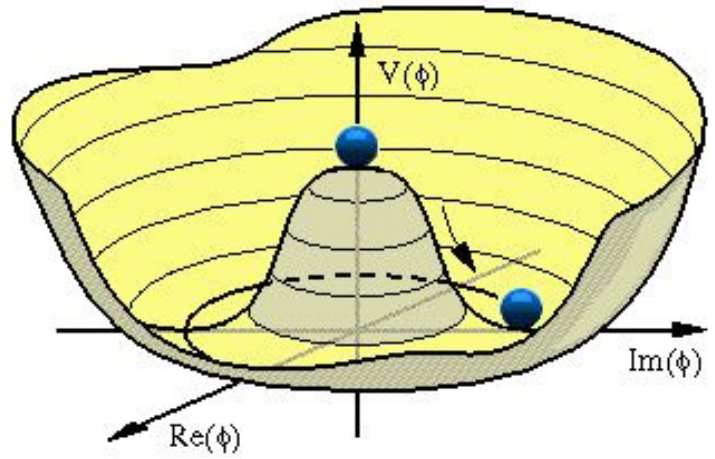


FIGURE 1.1 Shape of the scalar potential for $\mu^2 > 0$ (blue bullet at $\phi = 0$, single trivial minimum) and $\mu^2 < 0$ (blue bullet at $\phi \neq 0$, continuum minimum). In the second case there is a continuous set of degenerate vacua, corresponding to different phases θ.

Lagrangian density incorporating the new field (Φ) is

$$\mathcal{L}(Higgs) = (D^{\mu}\Phi^{+})D_{\mu}\Phi + V(\Phi) + y_f(\Phi L_L L_R + h.c.) \tag{1.19}$$

where the potential $V(\Phi)$ is defined as

$$V(\Phi) = \mu^2\phi^{+}\phi + \lambda(\phi^{+}\phi)^2 \tag{1.20}$$

The parameter μ represent the mass term and the parameter λ corresponds to the coupling constant. Under $SU(2)_L$, Φ is a doublet of complex scalar fields:

$$\Phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^{+}(x) + i\phi_2^{+}(x) \\ \phi^{0}(x) + i\phi^{0}(x) \end{pmatrix} \tag{1.21}$$

For the quadratic term of the potential $V(\Phi)$ there are two possibilities:

- $\mu^2 > 0$: the potential has only the trivial minimum $\Phi = 0$. It describes a massive scalar particle with mass μ and quartic coupling λ, Figure 1.1,
- $\mu^2 < 0$: the potential has a non-trivial minimum equal to:

$$|\Phi| \equiv \frac{v}{\sqrt{2}} = \sqrt{-\mu^2/2\lambda} \tag{1.22}$$

where $v/\sqrt{2}$ is the vacuum expectation value ($v.e.v$), Figure 1.1. In this case there are an infinite number of fundamental states and no one is preferred. With the choice of one of these infinite fundamental states, the symmetry gets spontaneously broken leading to the appearance of as many massless scalar particles as broken generators of the symmetry of the Lagrangian, as stated by the Goldstone theorem [6] [21, 22].

The latter solution leads to a non empty vacuum, since the field Φ has a $v.e.v$ which is different from zero[7]. It is possible to choose one arbitrary state among the infinity solution as the *ground state* and parametrize the scalar field Φ around its minimum:

$$\Phi = e^{i\theta^a(x)\sigma_a}\begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix} \tag{1.23}$$

To preserve the invariance under translation, the $v.e.v$ of the scalar field should not depend on space-time transformations. Performing a local gauge transformation (unitary gauge transformation) of Φ allows to eliminate three Goldstone bosons and obtain a simple scalar field:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.24}$$

The last term in the Equation 1.19 is a mass term for the fermions, $y_f L_R \phi L_L$. The Higgs coupling to the fermions y_f is proportional to the fermion mass, while the matrix mass for the vector bosons is not diagonal [23]

$$M^2 = \frac{v^2}{2}\begin{pmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & -gg' \\ 0 & 0 & -gg' & g^2 \end{pmatrix} \tag{1.25}$$

A rotation of an angle θ_W, Weinberg-angle, is necessary for the diagonalization of the mass matrix. The Weinberg-angle θ_W is defined as a function of the coupling constants g and g' through the following relation:

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}, \tag{1.26}$$

[6] Goldstone theorem: if a Lagrangian is invariant under a continuous symmetry group G, but the vacuum is only invariant under a subgroup $H \subset G$, then there must exist as many massless spin-0 particles (Nambu-Goldstone bosons) as broken generators (i.e., generators of G which do not belong to H).

[7] In the SM v is equal to: $v = (\sqrt{2}G_F)^{1/2} \sim 246.2$ GeV where G_F is the Fermi coupling constant.

One can show also that the electric charge e is proportional to the coupling constant g and g' and to the Weinberg-angle:

$$e = g \sin\theta_W = g' \cos\theta_W \tag{1.27}$$

A way to measure experimentally the Weinberg-angle is from Z-pole and neutral current processes [24]. The latest result of $sin\theta_W$ from LEP and SLD experiments is:

$$\sin\theta_W = 0.23153 \pm 0.00016^{8} \tag{1.28}$$

Using this Weinberg-angle definition, the gauge bosons Z_μ and A_μ, mediators respectively of the neutral weak and electromagnetic interactions, can be obtained by a linear combination of the gauge fields W^3_μ and B_μ

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W, \qquad A_\mu = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W \tag{1.29}$$

The charged weak bosons are defined as:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \tag{1.30}$$

The mass eigenvalues of the three weak vector bosons $W^\pm$ and Z^0 and the electromagnetic vector boson γ are:

- $W^\pm$: $M_W = v\frac{g}{2}$,
- Z^0 : $M_Z = v\frac{\sqrt{g^2+g'^2}}{2}$,
- A : zero mass.

Three degrees of freedom of the Higgs doublet were used to give mass to the weak vector bosons, while the fourth one shows up as a new particle with zero electric charge, 0-spin and a mass related to the Higgs potential parameters by the expression:

$$M_H = \sqrt{-2\mu^2} = v\sqrt{2\lambda}v \tag{1.31}$$

[8] Result obtained by the combination of all asymmetries (relative difference between the measurement of the differential cross section as a function of the scattering angle of the out-going fermion with respect to the direction of the incoming particle), which can be expressed in terms of the effective weak mixing angle, at LEP and SLD.

1.1.4.1 Higgs boson production

At the LHC the Higgs boson is searched for mainly in four exclusive production processes [25] [26]: gluon-gluon fusion $gg \to H$, vector-boson-fusion (VBF) $qq' \to qq'H$, associated production with a vector boson $q\bar{q} \to HW/Z$ and associated production with a top quark pair $q\bar{q}/gg \to t\bar{t}H$[9]. The Feynman diagrams of these four SM Higgs production processes are shown in Figure 1.2.

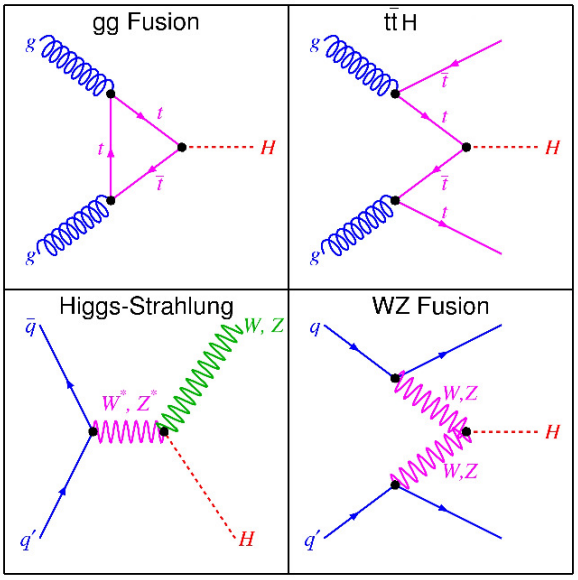


FIGURE 1.2 Leading order Feynman diagrams of the four SM Higgs production processes: gluon-gluon fusion, VBF, VH, $t\bar{t}H$.

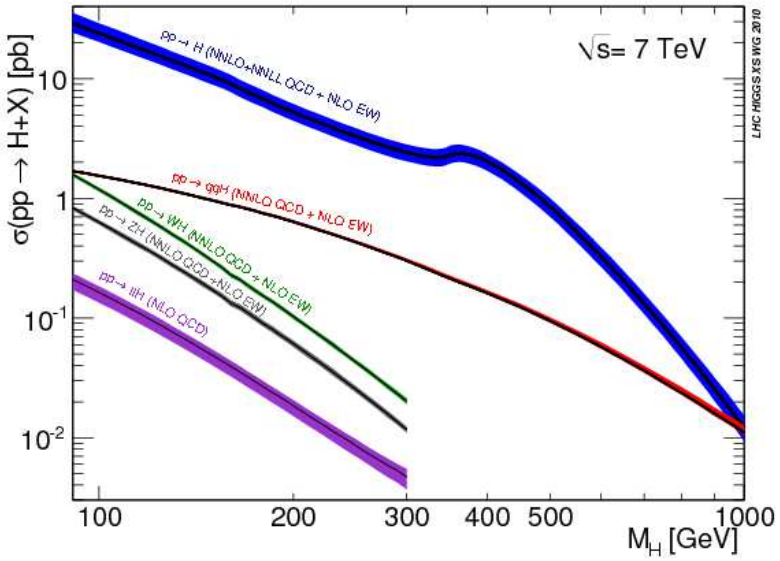


FIGURE 1.3 The SM Higgs production cross-section at $\sqrt{s} = 7$ TeV as a function of the Higgs mass M_H [27].

[9] Associated production with the other quark flavours are also possible but, due to the lower quark masses compared to the top quark mass, they are negligible and usually swamped by other physics background.

Figure 1.3 shows the Higgs boson production cross sections in pb at $\sqrt{s} = 7$ TeV as a function of its mass in the different production channels. The cross section values of the different production processes at $\sqrt{s} = 7$ TeV and 8 TeV, for a Higgs boson mass hypothesis of 125 GeV are presented in Table 1.3.

The main production mechanism of the Standard Model Higgs boson at a hadron collider is the gluon-gluon fusion via a heavy quark loop. The gluon-gluon fusion is a strong interaction mechanism. The leading order (LO) contribution to the gluon-gluon fusion cross section is proportional to the square of the QCD coupling constant α_S^2. The main contribution to the quark loop arises from the top quark, due to its large Yukawa coupling[10] to the Higgs boson.

The VBF Higgs production channel in which the Higgs boson is produced in association with two hard jets in the forward and backward regions of the detector, play as well an important role at the LHC energy. Precision studies of this channel could help in the determination of the Higgs boson coupling with the weak gauge bosons.

Other interesting production modes are the associated production with the weak gauge bosons, WH and ZH, known as Higgs-strahlung processes. They could provide further information on the Higgs-weak gauge bosons coupling, and as well as the Higgs-b quark coupling by exploiting the Higgs decay mode to a bottom quark pair, $VH \to Vb\bar{b}$.

The last process is the Higgs boson production in association with a top quark pair $t\bar{t}H$. This production process will be discussed in more detail in Section 1.2.5.

Channel	$\sqrt{s} = 7$ TeV [pb]		$\sqrt{s} = 8$ TeV [pb]	
	σ [pb]	Total Err.[%]	σ [pb]	Total Err.[%]
$gg \to H$	15.31	+19.5 -15.1	19.52	+14.7 -14.7
VBF	1.211	+2.7 -2.4	1.578	+2.8 -3.0
$qq \to WH$	0.5729	+3.7 -4.3	0.6966	+3.7 -4.1
$qq \to ZH$	0.3158	+4.9 -5.	0.3943	+5.1 -5.0
$qq \to t\bar{t}H$	0.08634	+11.8 -17.8	0.1302	+11.6 -17.1

TABLE 1.3 SM Higgs-boson production cross sections at $\sqrt{s} = 7$ TeV at NNLL QCD and NLO EW for a Higgs boson mass of 125 GeV [27, 28] .

1.1.4.2 Higgs boson Couplings

The coupling between the Higgs boson and the fermions (Figure 1.4) is proportional to the fermion masses. Considering that the Higgs field $v.e.v$ can be written as a function of the W boson mass, the fermion coupling is:

$$y_f = \sqrt{2}\frac{m_f}{v} = \frac{g m_f}{2M_W} \tag{1.32}$$

[10] The top quark Yukawa coupling $y_t$to the Higgs boson is equal to $y_t = \sqrt{2}\frac{m_t}{v} \sim 1$.

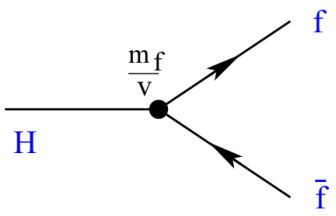


FIGURE 1.4 Fermionic coupling of the Higgs boson.

The Higgs couples also to the weak gauge fields via a trilinear and quadrilinear coupling modes. The trilinear terms are of the the form:

$$L_{hWW} = \frac{g^2 v}{2} h W_\mu^- W^{+\mu} = g M_W h W_\mu^- W^{+\mu}, \qquad L_{hZZ} = \frac{g^2 v}{4 cos^2\theta_W} h Z_\mu Z^\mu = \frac{g}{2 cos\theta_W} M_Z h Z_\mu Z^\mu \tag{1.33}$$

Where the field h is the scalar field in (1.23). Another trilinear term, that appears in the Lagrangian, is the self coupling of the Higgs boson:

$$\mathcal{L}_{hhh} = -g \frac{M_H^2}{4 M_W} hhh \tag{1.34}$$

The terms for the quadrilinear coupling are:

$$L_{hhWW} = \frac{g^2}{4} hh W_\mu^- W^{+\mu}, \qquad L_{hhZZ} = \frac{g^2}{8 cos^2\theta_W} hh Z_\mu Z^\mu, \qquad L_{hhhh} = \frac{-g^2 M_H^2}{32 M_W^2} hhh \tag{1.35}$$

The vertices corresponding to the Higgs boson couplings to the vector weak bosons are shown in Figure 1.5.

1.1.4.3 Higgs boson decay

The partial width of the Higgs boson into a pair of fermions is at the tree level equal to:

$$\Gamma(H \to ff) = \frac{N_c g^2 m_f^2}{32 \pi M_W^2} \beta^3 M_H \tag{1.36}$$

where N_c is the color charge factor (1 for leptons, 3 for quarks) and $\beta = (1 - \frac{4m_f^2}{M_H^2})$

The partial width of the Higgs boson decay into on-shell vector bosons W and Z are at the tree level:

$$\Gamma(H \to WW) = \frac{g^2 M_H^3}{64 \pi M_W^2} \sqrt{1 - x_W} (1 - x_W + \frac{3}{4} x_W^2) \tag{1.37}$$

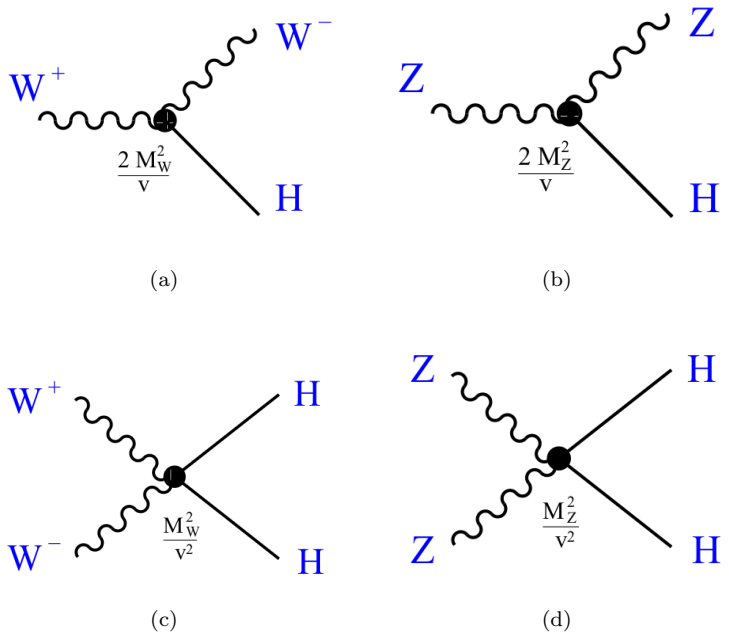


FIGURE 1.5 Higgs couplings to the weak gauge bosons. 1.5(a) shows the HWW coupling and 1.5(b) HZZ coupling. 1.5(c) and 1.5(d) are the quadrilinear terms respectively with the W and Z bosons.

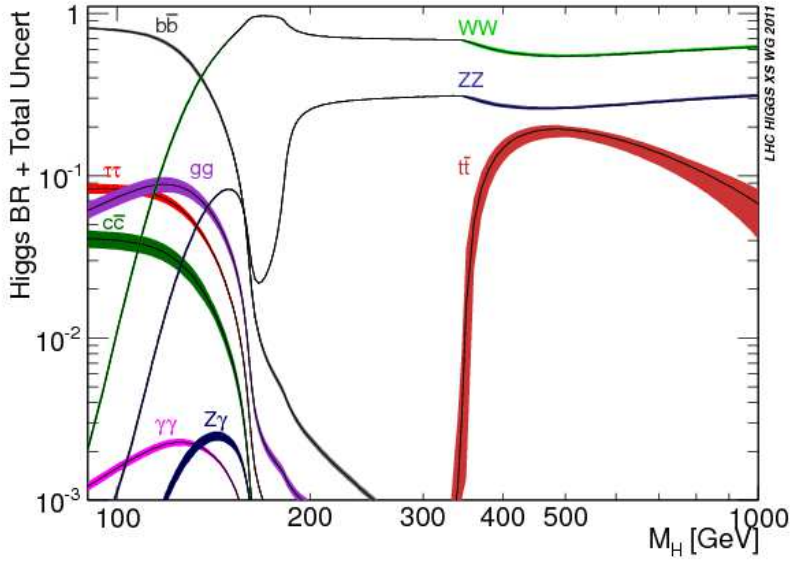


FIGURE 1.6 Branching ratios of the Standard Model Higgs boson into different decay channels as a function of its mass [27].

$$\Gamma(H \to ZZ) = \frac{g^2 M_H^3}{128\pi M_Z^2}\sqrt{1-x_Z}(1 - x_Z + \frac{3}{4}x_Z^2) \tag{1.38}$$

where $x_i = 4M_i^2/M_H^2$ $(i = W, Z)$. In the case where $M_H >> M_i$ all the terms involving x_i go to unity, so they can be dropped. Thus the Higgs width into di-boson decay increases with the third

power of the Higgs mass.

Higgs decays into gluon and photon pairs are mediated by fermionic loops. In the latter case it is also mediated by weak bosons loop. Since the Higgs-fermion coupling is proportional to the fermion mass (Equation 1.32), the main contribution in the fermionic loop is from the top quark. Figure 1.6 shows the variation of the Higgs boson decay branching ratios in different final states as a function of its mass.

1.1.5 Discovery of the Higgs boson

In July 4, 2012, the ATLAS and CMS Collaborations announced the discovery of a new particle with a mass of about 125 GeV in the search for the SM Higgs boson. Both collaborations performed Higgs searches analysis on the total data-set collected in 2011 at $\sqrt{s} = 7$ TeV and part of the data collected in 2012 at $\sqrt{s} = 8$ TeV, which corresponds to a total integrated luminosity of $\int Ldt = 4.8$ (@ 7 TeV) $+$ 5.9 (@ 8 TeV). The observed combined significance in three of the Higgs decay modes ($H \to \gamma\gamma$, $H \to ZZ^*$, and $H \to WW^*$) is 5.9 σ in ATLAS [29] and 5.0 σ in CMS [30], compatible with their respective sensitivities. Table 1.4 summarizes the observed and expected sensitivities of the new discovered resonance in the three Higgs decay channels mentioned above.

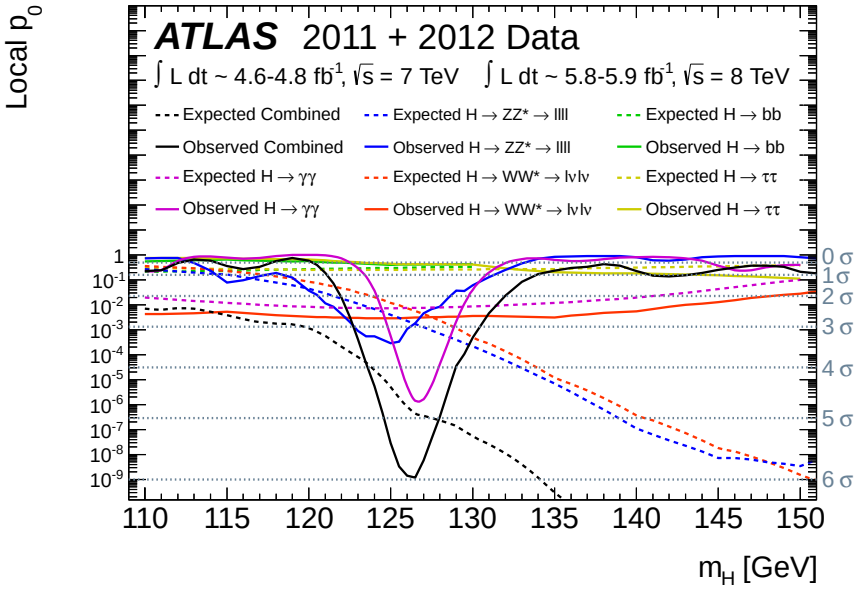


FIGURE 1.7 The local probability p_0 for a background-only experiment to be more signal-like than the observation, for individual channels and the combination. The solid curves give the observed individual and combined p_0. The dashed curves show the median expected value under the hypothesis of a SM Higgs boson signal at that mass. The horizontal dashed lines indicate the p_0 corresponding to significances of 0 σ to 6 σ [29].

The local probability for a background-only hypothesis is shown in Figure 1.7. The decay channels which contributed to the discovery are the $H \to ZZ^* \to l^+l^-l^+l^-$, $H \to \gamma\gamma$ and $H \to W^+W^- \to$

Channel	Observed sensitivity		Expected sensitivity	
	ATLAS	CMS	ATLAS	CMS
$H \to \gamma\gamma$	4.5σ	4.1σ	2.5σ	2.8σ
$H \to ZZ^* \to l^+l^-l^+l^-$	3.6σ	3.2σ	2.7σ	3.8σ
$H \to WW^* \to l^+\nu l^-\bar{\nu}$	2.8σ	1.6σ	2.3σ	2.4σ

TABLE 1.4 Summary of the observed and expected sensitivities of the new discovered resonance in the three main decay modes in ATLAS and CMS experiments at the Higgs boson mass hypothesis of about 125 GeV.

$l^+\nu_l l^-\bar{\nu_l}$. Both ATLAS and CMS Collaborations reported also a compatible measurement of the mass of the observed resonance:

$$M_H^{ATLAS} = 126.0 \ \pm \ 0.4 \text{ (stat.)} \ \pm \ 0.4 \text{ (syst.) GeV}, \tag{1.39}$$

$$M_H^{CMS} = 125.3 \ \pm \ 0.4 \text{ (stat.)} \ \pm \ 0.5 \text{ (syst.) GeV} \tag{1.40}$$

A mass measurement alone is not enough to confirm that the observed particle is the Higgs boson predicted by the SM. Specific studies on the properties of the observed particle are necessary to validate the model. The SM, in fact, predicts the Higgs boson with spin zero and positive parity.

At the Rencontres de Moriond conference in March 2013, an update on the new discovered particle was shown by both Collaborations based on the total 2011 and 2012 integrated luminosities, up to 25 fb^{-1}. A summary of the signal strength of the new particle is shown in Figure 1.8. The signal strength μ is the ratio of the measured cross section and the cross section predicted by the SM. Moreover other interesting results were presented about the properties of the observed particle, such as the spin and the parity. ATLAS results in ZZ channel exclude the spin-parity hypothesis of $J^P = 0^-$ and $J^P = 1^+$ at 97% confidence level (C.L.) with respect to the $J^P = 0^+$ one [32]. The CMS ZZ results exclude the $J^P = 2^+$ hypothesis with minimal couplings to vector bosons at greater than 98% CL, and the $J^P = 0^-$, $J^P = 1^+$, and $J^P = 1^-$ hypothesis at greater than 99.8% CL with respect to the $J^P = 0^+$ hypothesis [33]. In the WW [34] and di-photon [35] channels, the ATLAS Collaboration excluded the $J^P = 2^+$ hypothesis at 95% C.L. The CMS WW results exclude an additional standard model Higgs-like bosons in the mass range 128-600 GeV at 95% confidence level. The $J^P = 0^+$ hypothesis of the standard model Higgs boson for quantum numbers and couplings is tested against the hypothesis of a narrow spin-2 resonance produced through the gluon fusion mechanism and with minimal couplings to the WW pair [36]. From these results the two collaborations confirmed that the new particle is very likely to be a Higgs boson. However many open questions are still to be considered, such as whether the observed particle is the SM

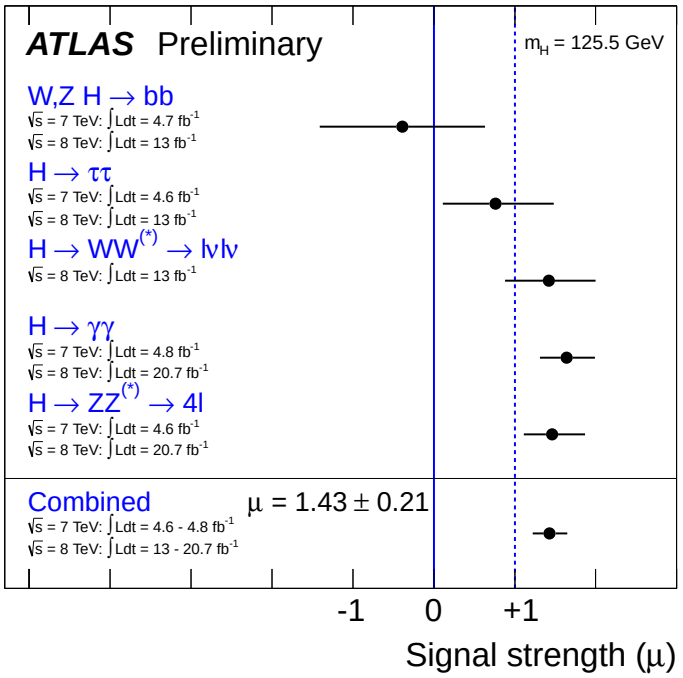


FIGURE 1.8 Measurements of the signal strength parameter μ for $M_H = 125.5$ GeV/c^2 for the individual channels and for their combination [31]. The signal strength μ is the ratio between the measured cross section and the cross section predicted by the SM. It is defined such that $\mu = 0$ correspond to the background only hypothesis and $\mu = 1$ correspond to the SM Higgs boson signal in addition to the background.

Higgs boson, or just one of many Higgs bosons predicted by some beyond the SM theories. To answer these questions, more studies on the characteristics of the Higgs boson are necessary. A key information is the Higgs boson coupling to the particles, especially to fermions.

The experimental results are also used by the theorists to determinate the individual Higgs coupling and test the overall compatibility of the SM with the collected data. In the SM, as described in Section 1.1.4.2, the Higgs boson coupling to the weak bosons, fermions and its self coupling are predicted for a specific value of the Higgs boson mass and are found to depend on the particle's masses. The presence of new physics can alter the couplings strength. All tree-level Higgs couplings and their ratios are parameterized as[37]:

$$g_{xxH} = g_x = (1 + \Delta_x) g_x^{SM}, \qquad \frac{g_{xxH}}{g_{yyH}} = \frac{g_x}{g_y} = (1 + \Delta_{x/y})(\frac{g_x}{g_y})^{SM} \tag{1.41}$$

The Δ_x term contains two components, one that takes into account the measured direct coupling to all SM particles and a second one that parameterizes additional contributions, due to the presence of new particles beyond SM, to the effective vertex.
Figure 1.9 shows the observed central values and the error bars of the coefficients which parametrizes the shift with respect to the SM expectations. The Figure shows the central coupling values for

the W and Z bosons as well as the third-generations fermions. The two massive gauge boson couplings are extracted with a good precision, while the top and bottom Yukawa couplings, which are measured indirectly, don't agree well with the SM expectations. Moreover the τ Yukawa coupling cannot yet be extracted. The studies performed on the 2011 and 2012 data give some indications concerning the next steps that are needed to complete the picture. The results suffer still from the statistical limitation; hence by enhancing the data sample, the extraction of the couplings strength measurements could be improved. A direct determination of the top quark Yukawa coupling is also necessary to test the induced Higgs-photon coupling. Another fundamental measurement is the direct measurement of the bottom Yukawa coupling which allow to probe its contribution to the Higgs boson width. An improvement in the $H \to b\bar{b}$, $H \to \tau\bar{\tau}$ and Higgs associated production analyses could help in the SM validation or in a new physics search.

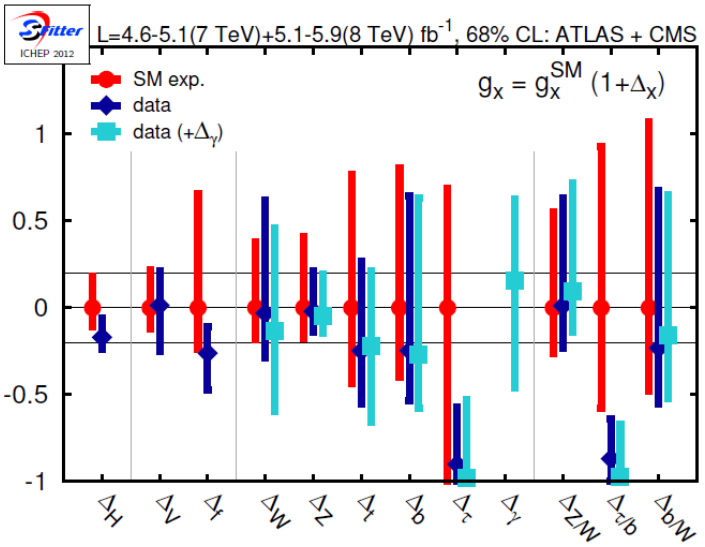


FIGURE 1.9 Observed $g_x = g_x^{SM}(1 + \Delta_x)$ for different Higgs boson couplings and their ratios. $\Delta_{i/j}$ $(i = Z, \tau, b,\ j = W, b)$ are the ratio of the single vairation.The results are based on 2011 and 2012 data, for the SM signal expectation for $M_H = 126$ GeV.The shift with respect to the SM prediction for the Higgs boson Δ_H, as well as the universal fermion and boson shift coefficient $\Delta_{V,f}$ are shown. The band indicates a $\pm 20\%$ variation [37].

1.2 Top quark physics

The top quark was discovered in 1995 at the Fermilab $p\bar{p}$ Tevatron collider by the two experiments CDF [38] and DØ [39]. It belongs to the third quark generation of the SM and it has an electric charge of $2/3$ e, where e is the elementary electric charge. The top quark is the heaviest observed elementary particle with a mass of $173.20 \pm 0.31 \pm 0.71$ GeV [40–42]. Figure 1.10 shows a table of leptons and quarks masses, charge and spin.

The top quark has a fundamental role for probing the strong and electroweak physics and for the new physics discoveries. Moreover, considering that it represents the main background of several

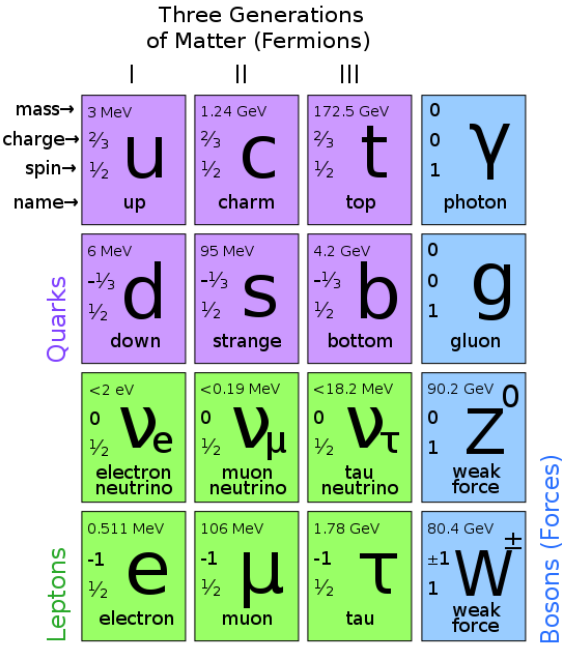


FIGURE 1.10 Table of leptons and quarks masses, charge and spin. The mass values do not correspond to the word values, they give only an order of magnitude of the mass of the particles.

new physics searches, it is necessary to measure its total and differential production cross section accurately.

Precise top mass measurement provides a constraint to the mass of the Higgs boson. It is the only fermion with a Yukawa coupling to the Higgs boson close to unity $y_t \approx 1$ and could point to new dynamics beyond the SM. Due to its high mass, the top quark has a very short lifetime, about an order of magnitude smaller than the hadronization timescale.

$$\tau_t = O(10^{-24})\ s < \tau_{hadronization} \sim O(10^{-23})\ s \tag{1.42}$$

This means that the top quark produced in a collision, decays before the hadronization process, so it cannot form a bound state allowing to test the properties of a bare quark.

Moreover the top quark properties are predicted with a high precision by the SM.Their precise measurement at the LHC may probe its validity and is also a sensitive window to discover new physics beyond it.

1.2.1 Top quark production

In hadronic colliders top quarks are expected to be produced in pairs via strong interaction processes and singly via electroweak interaction processes. The production cross sections for top quarks, both in pairs and as a single quark, depend strongly on the collision energy provided by the accelerator, as shown in Figure 1.11. Two distinct QCD processes contribute to the top quark pair production:

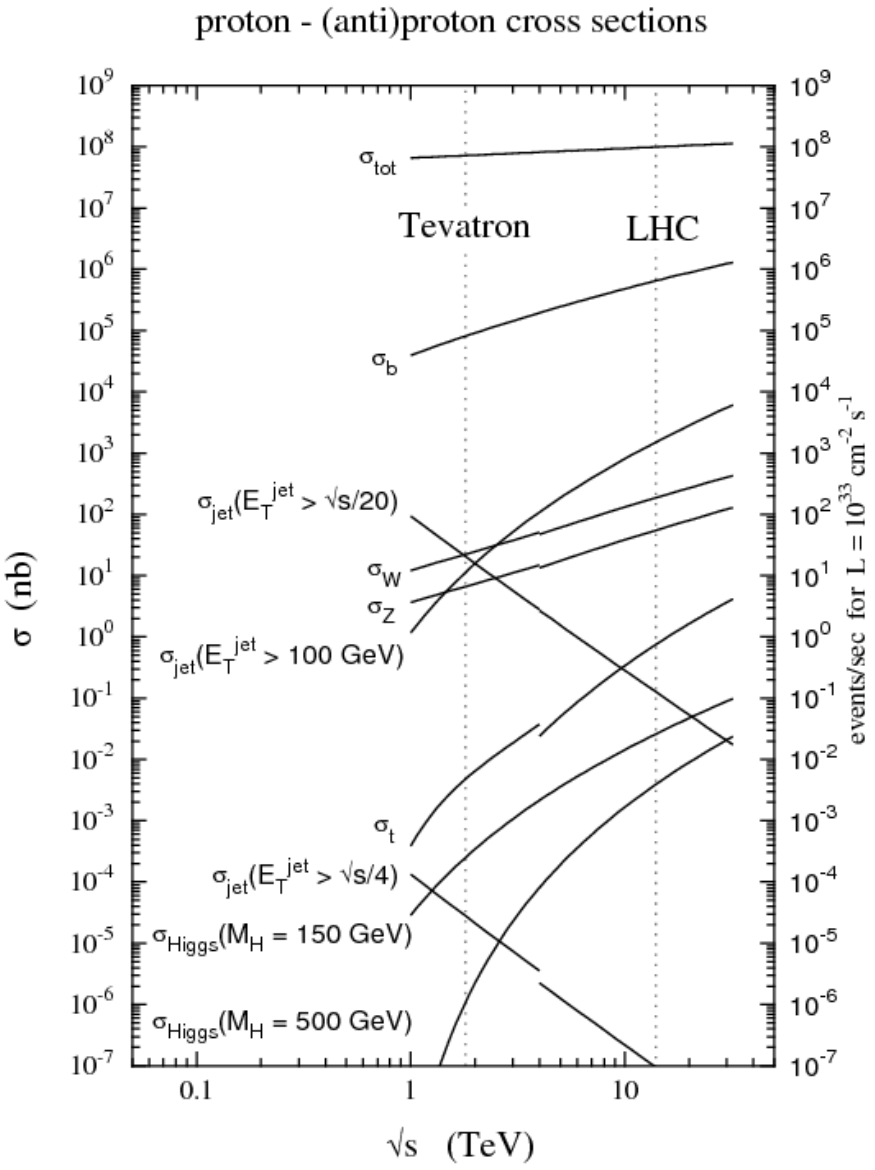


FIGURE 1.11 QCD prediction for hard scattering cross section at the TEVRATON and the LHC. The steps in the curves at $\sqrt{s} = 3$ TeV mark the transition from $p\bar{p}$ at the TEVRATON and pp at the LHC[43].

$q\bar{q}$ annihilation and gluon-gluon fusion. The Feynman diagrams of the two production processes of the top-anti top pairs are shown in Figure 1.13. At the LHC proton-proton collider there are no valence anti-quarks. The antiparticles only exist as sea quark-anti quark in the protons. Also the fraction of gluons in the protons increases with the energy scale of an event, therefore at high proton energies the gluon fraction within the protons increases, see Figure 1.12.

As a consequence about the 87% ($\sqrt{s} = 7$ TeV) of the top-antitop events at the LHC are produced via gluon-gluon fusion and 13% ($\sqrt{s} = 7$ TeV) via quark-antiquark annihilation. The total cross section for the production of heavy quarks, at a pp collider is given by the convolution of the partonic cross section with the parton distribution function (PDF) $f(x, \mu)$, where μ is the factorization scale.

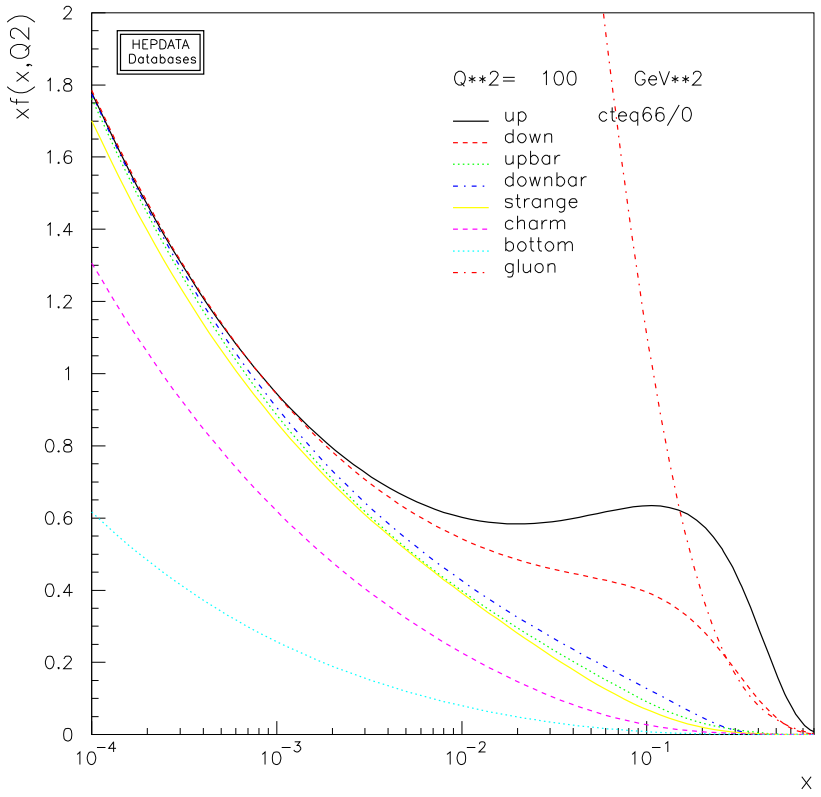


FIGURE 1.12 Parton distribution function for protons in the CTEQ66 PDF set [44] at $Q^2 = 100$ GeV [45].

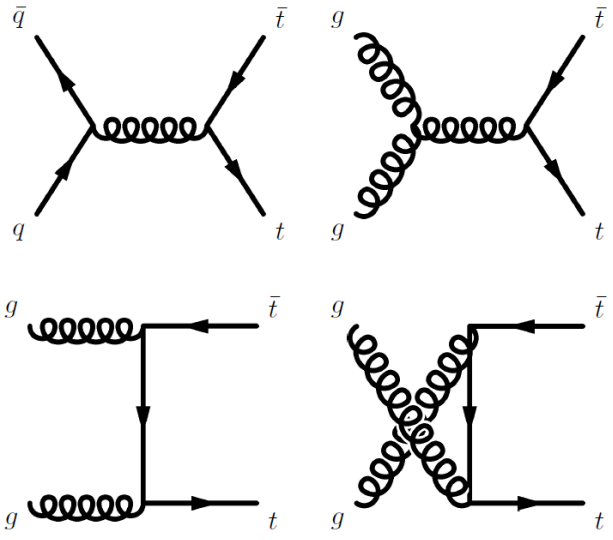


FIGURE 1.13 The lowest order Feynman diagrams of the $t\bar{t}$ production at the LHC.

The total cross section can be then written as:

$$\sigma(pp \to t\bar{t}) = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu^2) f_j(x_j, \mu^2) \sigma_{ij}(\hat{s}, \alpha_S, \mu^2, m_{top}) \tag{1.43}$$

The sum over the indices i and j in the Equation 1.43 is over all the partons (gluons and quarks) and σ_{ij} is the perturative cross section for collisions of partons i and j. The f_i (f_j) term is PDF which express the probability for a parton i (j) to carry a momentum fraction x_i (x_j) of its parent proton. The center-of-mass energy $\hat{s}$ of the $i-j$ parton system is related to the pp center-of-mass energy s by $\hat{s} = x_i \cdot x_j \cdot s$. The parameter α_S is the strong coupling constant and m_{top} the quark top mass value. The partonic cross section is independent of the factorization scale μ at leading order in the perturbative QCD, but depend logarithmically on μ at next-to-leading order and higher. Including all the orders in perturbative QCD the hadronic cross section is independent from μ, but at any finite order the cross section depends on the normalization scale. In order to obtain a reliable cross section prediction it is necessary to calculate high-order correction until the factorization scale dependence is reduced.
The current theoretical prediction at the Next-to-Next Leading Order (NNLO) and Next-to-Next Leading Logarithmic Order (NNLL) of the $t\bar{t}$ total production cross section at the LHC at $\sqrt{s} =$ 7 TeV and for a top quark mass of 173.3 GeV is $\sigma_{t\bar{t}} = 172.0^{+4.4+4.7}_{-5.8-4.8}$ pb [46].

1.2.2 Single top quark production

The top quark can also be produced singly via electroweak interaction processes. Three different modes contribute to the single top quark production, which differ in the virtually of the participating W boson:

- t-channel (Figure 1.14(a), Figure 1.14(b)): in the process $pp \to tqb + X$ the W boson is space like ($-Q^2 = q^2 = t < 0$). The predicted production cross section at approximate NNLO at LHC for $\sqrt{s} = 7$ TeV and a top quark mass of 173 GeV is $\sigma_t = 65.9^{+2.1+1.5}_{-0.7-1.7}$ pb [47];
- s-channel (Figure 1.14(e)): in the process $pp \to tqb + X$ the W boson is time like ($-Q^2 = q^2 = s \geq (m_t + m_b)^2 > 0$). The predicted production cross section at approximate NNLO at LHC for $\sqrt{s} = 7$ TeV and a top quark mass of 173 GeV is $\sigma_t = 4.56 \pm 0.07^{+0.18}_{-0.17}$ pb [47];
- W-channel (Figure 1.14(c), Figure 1.14(d)): in the process $pp \to tW$, an on-shell W is produced in association with a top quark ($Q^2 = m_W$). The predicted production cross section at approximate NNLO at LHC for $\sqrt{s} = 7$ TeV and a top quark mass of 173 GeV is $\sigma_t = 15.6 \pm 0.4 \pm 1.1$ pb [47].

The single top quark production cross section is proportional to the matrix element $|V_{tb}|^2$ of the Cabibbo-Kobayashi-Maskawa matrix (CKM) [48, 49]. It is directly sensitive to the transition width

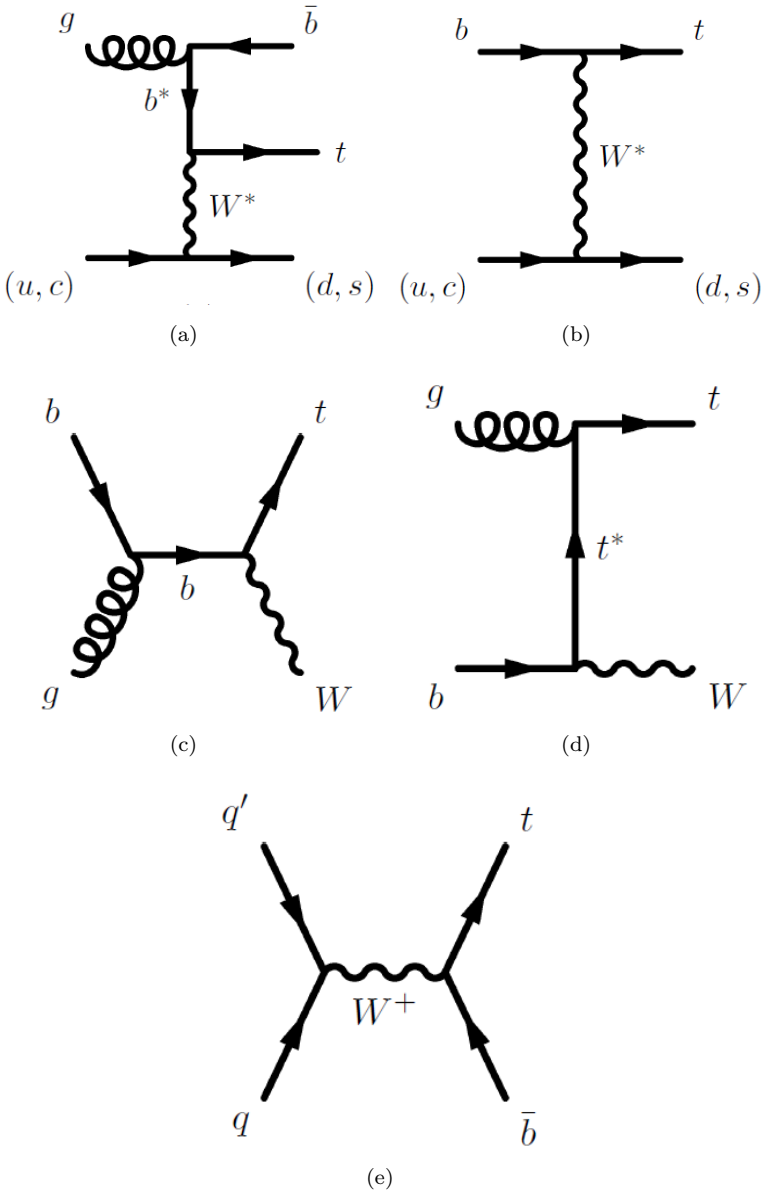


FIGURE 1.14 Leading Order Feynman diagrams for electroweak production of single top quark: (1.14(a), 1.14(b)) t-channel; (1.14(c), 1.14(d)) Wt-channel; (1.14(e)) s-channel.

of $t \to Wb$ and as a consequence to $|V_{tb}|$. Thanks to this electroweak production mode it is possible to measure the width of the top quark high dominant decay to W and b-quark $\Gamma(t \to Wb)$ and hence its lifetime. The studies of the single top quark are mainly important to directly determine the vertex coupling strength $t-W-b$. It is also important to have a precise experimental measurement of the three production modes. Moreover the various channels are sensitive in different ways to new physics and hence can be used to distinguish between several models.

1.2.3 Top quark decays

In the SM the decay probability of $t \to Wb$ is almost equal to 100% (99.8%) [7]. The $t \to Ws$ and $t \to Wd$ are allowed but suppressed by a factor of 10^{-3} to 10^{-4} by the square of the CKM matrix elements $|V_{ts}|$ and $|V_{td}|$. Assuming the unitary of the three generation CKM matrix, the values of these matrix elements are estimated to be less than 0.042 and 0.014 respectively. The top quark decay width predicted by the SM is:

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{M_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{M_W^2}{m_t^2}\right)^2 \times \left[1 - \frac{2\alpha_S}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \times |V_{tb}|^2 \tag{1.44}$$

The G_F Fermi coupling constant in the Equation 1.44 contains the largest part of the one-loop electroweak radiative corrections. The decay width depends on the top quark mass to the third power and for a top quark mass of 175 GeV/c^2 and $|V_{tb}| = 1$ the decay width is equal to:

$$\Gamma_t \approx 1.55 \text{ GeV} \to \tau_t = \frac{1}{\Gamma_t} \approx 4 \cdot 10^{-25}\text{s} \tag{1.45}$$

As already mentioned above the top quark mean lifetime is smaller than the characteristic hadronization time of the QCD. The top quark is expected to decay before top-flavoured hadrons or $t\bar{t}$ quarkonium bound states can form.
Events from top quark pair production consist of two W bosons and two b-quarks: $t\bar{t} \to W^+bW^-\bar{b}$. The W boson can decay into lepton-neutrino or quark-antiquark pair of different flavours. According to the W boson decay the $t\bar{t}$ final states can be divided in three classes:

- di-leptonic channel: $t\bar{t} \to \bar{l}\nu_l b l'\bar{\nu_{l'}}\bar{b}$,
- semi-leptonic channel: $t\bar{t} \to \bar{l}\nu_l b q\bar{q'}\bar{b} + q\bar{q'}bl\bar{\nu_l}\bar{b}$,
- hadronic channel: $t\bar{t} \to qq'bq''q'''\bar{b}$.

In the lowest order the W boson decay 1/3 of the time into lepton-neutrino pair, and 2/3 of the time into quark-antiquark pair of different flavours, see Table 1.5

Decay mode	Branching Ratio (%)
$W^+ \to e^+\nu_e$	10.8
$W^+ \to \mu^+\nu_\mu$	10.8
$W^+ \to \tau^+\nu_\tau$	10.6
$W^+ \to u\bar{d}, c\bar{s}$	69.6

TABLE 1.5 Born level theoretical branching ratios of the W^+ boson decay, assuming lepton universality. Identical values are obtained for the W^- [7].

The resulting branching ratios for the top quark pair are shown in Figure 1.15.

The di-leptonic final state is characterized by two well isolated opposite charged leptons, two b-jets from $t\bar{t}$ decays and missing energy due to the two neutrinos. This channel has the smallest

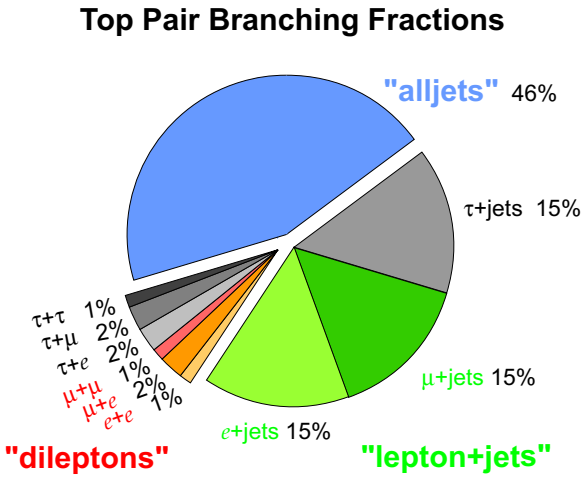


FIGURE 1.15 Decay modes of the $t\bar{t}$ events and their frequency occurrence.

background whilst the smallest branching ratio ($\sim 10\%$). In the semileptonic channel the events have one isolated lepton, two b-jets, missing energy and at least two jets originating from the hadronization of the two quarks from the W boson decay. This channel is characterized by a branching ratio of about 45%. The last channel is the all hadronic final state, characterized by at least six jets among which two are b-jets, no missing energy and no isolated leptons. Its branching ratio is about 46%, but it suffers from a huge background from QCD multi-jets production.

1.2.4 Top quark properties

1.2.4.1 Top quark mass

The top quark mass is one of the fundamental ingredients for the calculation of the radiative corrections, which connect electroweak processes. These corrections depend on the masses of the Higgs boson and top quark via the loop quantum corrections. At one loop the ρ parameter which relates the W and Z boson masses and the weak angle θ_W:

$$\rho = \frac{M_W^2}{M_Z^2}(1 - sin^2\theta_W) \equiv 1 + \Delta r \tag{1.46}$$

gets a radiative correction which is quadratic in top quark mass.

$$\Delta r = \frac{3G_F}{8\pi^2\sqrt{2}} m_{top}^2 + \frac{\sqrt{2}G_F}{16\pi^2} M_W^2 \left[\frac{11}{3} ln\left(\frac{M_H^2}{M_W^2}\right) + ...\right] + ... \tag{1.47}$$

The dominant term in the corrections of the electroweak processes is top quark mass. Precise measurement of the top quark mass can help in the test of the consistency of the SM and in the prediction of some of the unknown parameters, such as the Higgs boson mass. The most

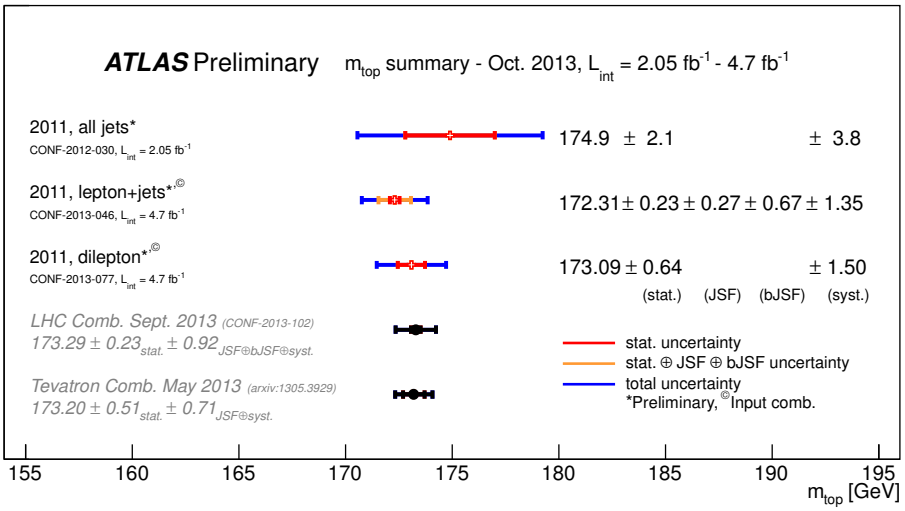


FIGURE 1.16 Summary of the latest ATLAS direct top quark mass measurements. The results are compared to the 2013 Tevatron and LHC mtop combinations. For each measurement, the statistical uncertainty, the jet scale factor (JSF) and b-jet scale factor (bJSF) contributions (when applicable) as well as the sum of the remaining uncertainties are reported separately. The JSF, bJSF contributions are statistical in nature and apply to analyses performing in-situ (top quark pair base) jet energy calibration procedures [41, 50–53].

precise measurement, $m_{top} = 173.2 \pm 0.6$ (stat) $\pm$ 0.8 (syst) GeV, is obtain by the CDF and D0 collaborations using the full available integrated luminosity collected during Run I and Run II, up to 5.8 fb^{-1} [40]. The lastest top quark mass result published by the ATLAS collaboration, using the 4.7 fb^{-1} data recorded in 2011 at $\sqrt{s} = 7$ TeV, in the lepton+jets channel is $m_{top} = 172.31 \pm 0.75$ (stat + JSF + bJSF) $\pm$ 1.35 (syst) GeV[11] [41]. The CMS top quark mass result in the lepton+ jets channel performed on the 2011 data at $\sqrt{s} = 7$ TeV is $m_{top} = 173.49 \pm 0.43$ (stat. + JES) $\pm$ 0.98(syst.) GeV[12] [42].

Figure 1.16 summarizes the results of ATLAS Collaboration on the top quark mass [41, 50, 51]. The results of the Tevatron [52] and the LHC [53] combination are also reported.

1.2.4.2 Electric charge

The measurement of the top pair production cross section $t\bar{t} \to b\bar{b}W^+W^-$ does not forbid alternative hypothesis on the top quark electric charge with respect to SM expected value $\frac{2}{3}e$. A possible interpretation could be an exotic heavy quark with electric charge of $-\frac{4}{3}e$ decaying via $Q_4 \to W^-b$. However the TEVATRON [54] and the LHC [55, 56] excluded, at more the 5 σ by the ATLAS and CMS Collaborations, the possibility of exotic top quark, validating the SM prediction. Various techniques are possible to perform an electric charge measurement in a hadron collider, such as

[11] JSF means Jet Scale Factor, bJSF stands for *b*-Jet Scale Factor
[12] *JES* means Jet Energy Scale

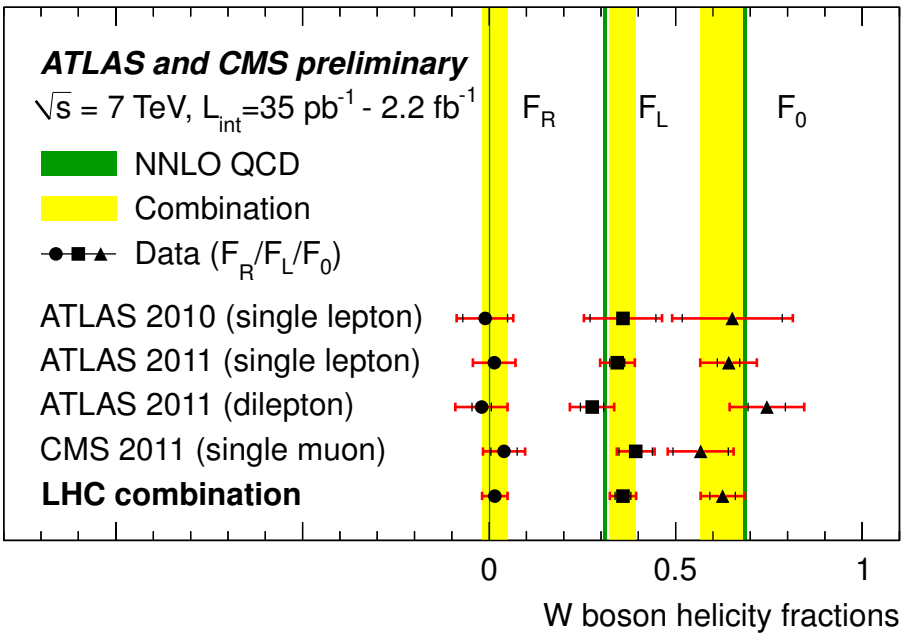


FIGURE 1.17 Overview of the four measurements of W boson helicity fractions included in the combination as well as the results of the combination. The inner and outer error bars correspond to the statistical and the total uncertainty, respectively. The green solid line indicates the predictions of NNLO QCD calculations [59].

measuring the charges of the decay products, in particular the b-jet or investigating the photon radiation in $t\bar{t}$ events through the ratio $\frac{\sigma(t\bar{t}\gamma)}{\sigma(t\bar{t})}$.

1.2.4.3 Helicity of W boson

The SM predicts that the top quark, as the others fermions, has a V-A (Vector-Axial) structure of the weak decay [57]. It implies that the W boson produced in the top quark decay cannot be right-handed (positive helicity). Considering a massless b-quark, for the V-A current the b-quark in the top decay should be left-handed. If the W boson is right-handed, the component of the total angular momentum along the decay axis would be $+3/2$. Since the initial top quark has a spin $\pm 1/2$, for the conservation of the angular momentum a right-handed boson is forbidden. The W boson can be either left-handed or longitudinally polarised. In the SM, the W boson from the top quark decay is 70% longitudinally polarized and 30% left-handed. The CMS results are obtained using an integrated luminosity of 5.0 fb^{-1} at $\sqrt{s} = 7$ TeV reporting $F_0 = 0.698 \pm 0.057$ (stat.) $\pm$ 0.063 (syst.), $F_L = 0288 \pm 0.035$ (stat.) $\pm$ 0.040 (syst.) and $F_R = 0.014 \pm 0.027$ (stat.) $\pm$ 0.042 (syst.), consistent with the standard model predictions [58]. Figure 1.17 summarizes the ATLAS and CMS measurements and reports also the LHC Combination [59].

1.2.4.4 Spin correlation

On average the top quark decays before there is time for its spin to be depolarized by the strong interaction. It means that the top quark polarization is directly observable through the angular distribution of its decay products. Thus it is possible to measure observables that are related to the top quark spin. The degree of correlation depends on the production and decay processes and as well on the reference axis used to define the top quark spin states.
In hadron collisions the $t\bar{t}$ pairs are produced unpolarised and the spins between the top and the anti-top quark are correlated in the top pair production. The spin correlation measurement[13] is useful to probe the $t\bar{t}$ production mechanism, and as well to study the weak decay property of the top quark. The measured degree of correlation, obtained by ATLAS, corresponds to $A_{helicity} = 0.40^{+0.09}_{-0.08}$, in agreement with the next-to-leading-order SM prediction. The hypothesis of zero spin correlation is excluded at 5.1 standard deviations [60]. The spin correlation in $t\bar{t}$ events, reported by the CMS Collaboration using an integrated luminosity of 5 fb^{-1}, is extracted from a fit to the angular distribution between the two selected leptons. In the helicity basis[14], the correlation coefficient is found to be 0.24 $\pm$ 0.02 (stat.) $\pm$ 0.08 (syst.) [61].

1.2.4.5 Yukawa coupling

In the SM the Yukawa coupling, see Section 1.1.4.2, of the top quark to the Higgs boson ($y_t = \sqrt{2}m_t/v$) is close to unity. This value leads to various speculations on the new physics that could be probed thought top quark coupling. The top quark coupling can be measured directly or indirectly. For a Higgs boson with a mass of 125 GeV the decay to a top pair is forbidden, so the only way to have a direct measurement of the Higgs coupling is via the associated production $t\bar{t}H$, see Figure 1.6. An indirect measurement of the top-Higgs coupling is provided by the di-photon decay mode process mediated, besides weak vector bosons, by a top quark loop. A key ingredient, to obtain indirect constrain on the top-Higgs coupling from electroweak precision observables, is a high precision measurement of the top mass.

1.2.5 $t\bar{t}$ Higgs associated production

The $t\bar{t}H$ search can be performed in different channels from the combination of the top quark and Higgs boson decays. One of most investigated final state is with the Higgs boson decaying into $b\bar{b}$

[13] The degree of correlation, A, is defined as the fractional difference between the number of the events where the top and antitop quark spin orientations are aligned and those where the top quark spins have opposite alignment. The $A_{helicity}$ represents the degree of correlation in the helicity basis, using the direction of flight of the top quark in the center-of-mass frame of the $t\bar{t}$ system.
[14] The helicity is the projection of the spin $\vec{s}$ onto the direction of the momentum $\vec{p}$.

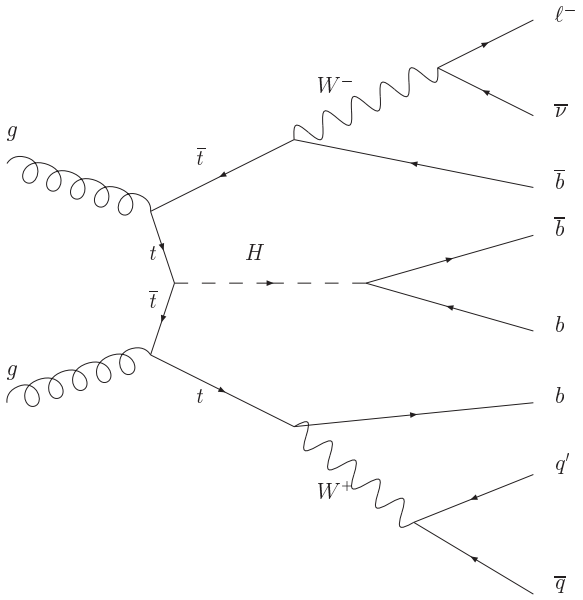


FIGURE 1.18 Feynman diagram for $t\bar{t}H$ production in the semi-leptonic final state

pair, $t\bar{t}b\bar{b}$. The signal cross section can be decomposed in the following way for all decay channels:

$$\sigma_{q\bar{q}/gg} \cdot BR_{t\bar{t}H \to t\bar{t}f\bar{f}} = \frac{\sigma_{ii} \cdot \Gamma_{ff}}{\Gamma_H} \tag{1.48}$$

where $\sigma_{q\bar{q}/gg}$ is the $t\bar{t}H$ production cross section through the initial state ii, $\Gamma_{f\bar{f}}$ the partial decay width into the final state $f\bar{f}$ and Γ_H the total width of the Higgs boson. The predicted cross section times branching ratio for the Higgs radiation off the top quark is low, $\sigma_{ttH}(m_H = 125 \text{ GeV}) = 0.0863^{+11.8}_{-17.8}$ pb and $BR_{H\to b\bar{b}} = 0.577^{+3.2}_{-3.3}$ at $\sqrt{s} = 7$ TeV [27], making a discovery of the SM Higgs boson in this channel alone not feasible with the available integrated luminosity. However the combination of the different Higgs boson decay channels, as well as the three $t\bar{t}$ decay modes, could lead to the discovery of the Higgs boson in this associated production channel. Furthermore, due to characteristic final state of $t\bar{t}H \to t\bar{t}b\bar{b}$ with at least 4 jets coming from a bottom quark (see Figure 1.18), it is interesting to search for any deviations from the SM. For example in Supersymmetric Two-Higgs-Doublet Models (2HDM) or in the Minimal Supersymmetric Standard Model (MSSM) this channel could be enhanced at low value of the ratio of the vacuum expectation values of the two Higgs doublets [62]. Anomalous contribution to the top quark Yukawa coupling could also exist [63] and the presence of a new quark singlet could enhance the $t\bar{t}H$ production [64]. The major source of background of the $t\bar{t}H$ production are the $t\bar{t}$ production with additional light or heavy jets, W+jets and multijets production.

Three different $t\bar{t}$ decay modes can be studied. The one that can provide a good discrimination between signal and background is the semi-leptonically final state. The other channels are also investigated, in particular the fully hadronic final state $t\bar{t}H \to t\bar{t}b\bar{b} \to bq_1q_2\bar{b}q_3q_4b\bar{b}$ which is studied in this thesis.

Chapter 2

Accelerator and Detector

The discovery of new physics and precise measurement of the SM theory requires a high energy and luminosity collider and experiments capable of very high performance. The Large Hadron Collider (LHC) machine, operating at CERN since November 2009, will provide proton-proton collisions at $\sqrt{s} = 13/14$ TeV with an instantaneous luminosity of the order of $\mathscr{L} = 10^{34} cm^{-2} s^{-1}$. This chapter gives an overview of the LHC and the ATLAS (A Toroidal LHC Apparatus) detector, which is one of the two general purpose detectors operating at the LHC ring.

2.1 LHC accelerator

The Large Hadron Collider (LHC) [65] is a superconducting proton machine installed at CERN inside the 27 km tunnel used in the past years for the LEP program. The LHC is designed for colliding two counter-rotating beams of protons. Each beam is injected in the LHC at 450 GeV and is then accelerated up to the nominal collision energy. The beams structured in proton bunches move around the LHC ring inside a continuous vacuum. They are guided by different varieties and sizes magnets. These include 1232 dipole magnets 15 metres in length which bend the beams, and 392 quadrupole magnets, each 57 metres long, which focus the beams.

The existing machines at CERN provide the first stages of acceleration (Figure 2.1): first, the protons are accelerated up to 50 MeV in the proton LINAC, then the Proton Synchrotron Booster boosts them to 1.8 GeV. The Proton Synchrotron accelerates them up to 25 GeV. Finally, the Super Proton Synchrotron is used accelerate protons up to 450 GeV and to inject them into the LHC. LHC dipole magnets have two different magnetic channels in one single twin bore magnet with the same yoke and cryostat. The magnets provide a magnetic field up to 8.36 T, which allows the colliding protons beam to reach the design energy of 7 TeV. The design instantaneous luminosity ($\mathscr{L} = 10^{34}$ cm^{-2} s^{-1}) for the pp collisions can be reached with 2835 bunches crossing at 25 ns intervals, corresponding to a spatial separation between bunches of 7.5 m. The number of

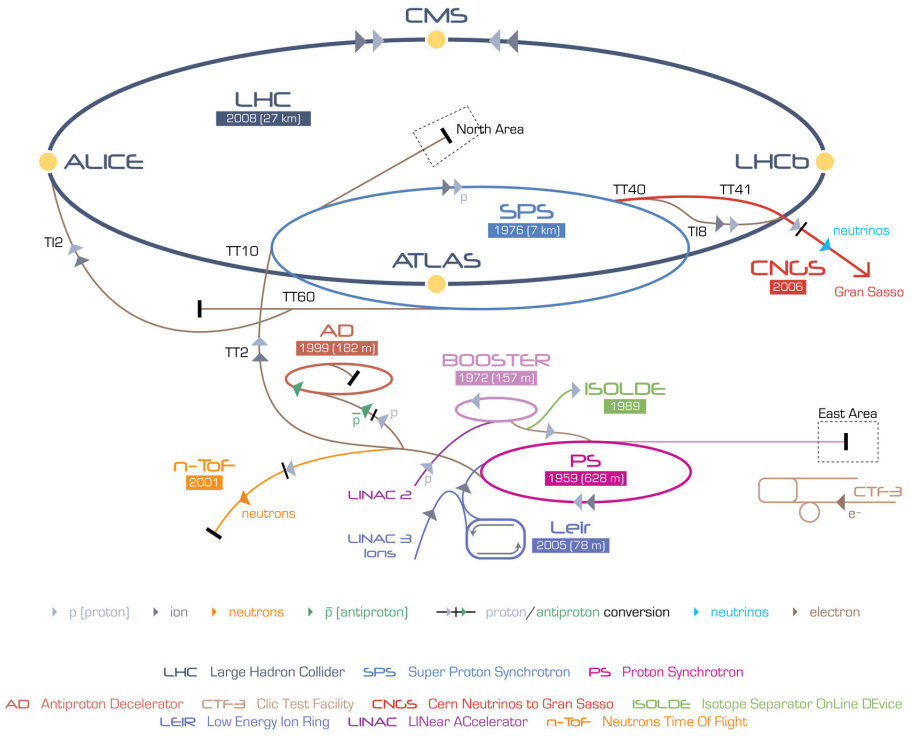


FIGURE 2.1 Schematic overview of the LHC and adjoint injection complex.

protons per bunch is of order of 10^{11}. At the LHC nominal center of mass energy, the total inelastic non-diffractive pp cross section is about 70 mb. Since the interesting processes have cross sections that are several orders of magnitude lower[1], a very selective trigger system is required. The most important parameters of an accelerator for the physics analysis are the center of mass energy and the instantaneous luminosity. In the 2010 and 2011 the pp collisions were recorded at $\sqrt{s} = 7$ TeV, whereas in 2012 at $\sqrt{s} = 8$ TeV. The luminosity is strictly connected to the accelerator parameters. From the instantaneous luminosity $\mathscr{L}$ and the cross section σ of a particular physics process, the number N of events produced per second is $N = \mathscr{L} \cdot \sigma$. The instantaneous luminosity, assuming a Gaussian distribution of the beam, is calculated by the following formula:

$$\mathscr{L} = \frac{n^2 \cdot B \cdot f_{rev}}{4\pi \cdot \sigma_x^* \cdot \sigma_y^*} \cdot F \tag{2.1}$$

where:

- n is the number of protons in a bunch;
- B is the number of bunches in the beam;
- f_{rev} is the bunch revolution frequency;
- σ_x^*, σ_y^* are the width of the Gaussian distribution of the beams in the transversal plane;

[1] At $\sqrt{s} = 7$ TeV the production cross section of $t\bar{t}$ pairs is $\sigma_{t\bar{t}} = 800$ pb and the inclusive Higgs production is well below 1 nb for any Higgs mass, see Figure 1.11

- F is a geometrical factor due to the crossing angle between the beams at the interaction point.

The main parameters of the LHC for pp collisions during 2011 data taking and at construction design are shown in Tab. 2.1

Parameters	2011	nominal
Center of mass energy (TeV) $\sqrt{s}$	7	14
Particles per bunch (B)	1.2 10^{11}	1.2 10^{11}
Number of bunches (n)	1800	2808
Bunch revolution frequency [kHz] f_{rev}	11	11
σ_x^*, σ_y^* [μm]	$\sim$ 60	$\sim$15
Bunch spacing [ns]	50	5
Instantaneous luminosity [$cm^{-2}s^{-1}$] $\mathscr{L}$	3.6510^{33}	10^{34}
Integrated luminosity $\mathscr{L}_{int}/year$	5fb^{-1}	100fb^{-1}

TABLE 2.1 Main LHC parameters at design luminosity and during 2011 data-taking.

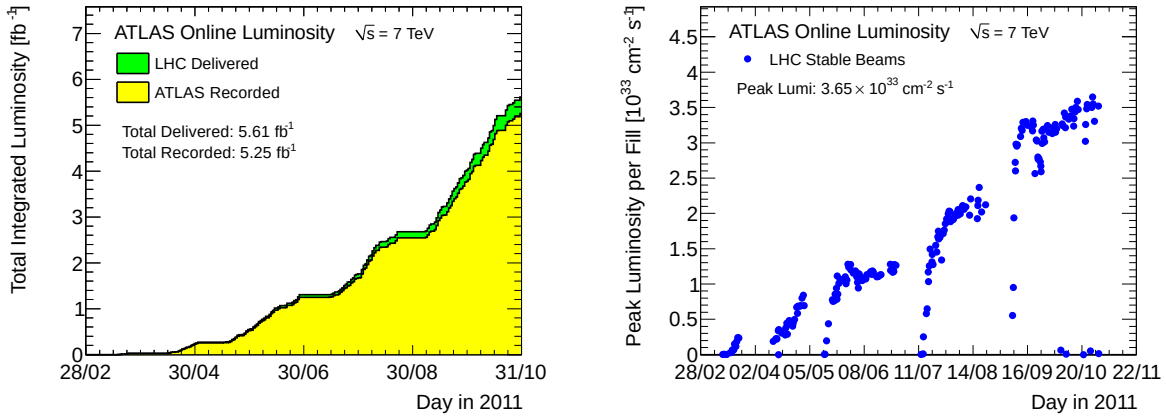


FIGURE 2.2 Cumulative luminosity versus day delivered to by the LHC (green), and recorded by ATLAS (yellow) during stable beams and for pp collisions at 7 TeV center-of-mass energy in 2011 in the left side. At right side the maximum instantaneous luminosity versus day delivered to ATLAS in the 2011 [66].

The integrated luminosity $\mathscr{L}_{int}$ is defined the integral over the LHC operation time period in a year of the instantaneous luminosity $\mathscr{L}_{int} = \int \mathscr{L}(t) \cdot dt$. At the nominal design instantaneous luminosity of 10^{34} $cm^{-2}s^{-1}$ we expect an integrated luminosity of 100 fb^{-1} in one year. During 2011 the instantaneous luminosity peak reached a peak value of 3.6 10^{33} $cm^{-2}s^{-1}$ and a total integrated luminosity ($\sim$ 120 days of collisions) of 5 fb^{-1} in ATLAS. Figure 2.2 show the cumulative luminosity per day delivered by the LHC at $\sqrt{s} = 7$ TeV in 2011 and recorded by the ATLAS detector (left), and the peak instantaneous luminosity per day delivered by the LHC and recorded by ATLAS during the same period (right).

Along the LHC tunnel, four experiments are installed: two general-purpose experiments (ATLAS and CMS [67]), one experiment dedicated to the study of heavy ion collisions (ALICE [68]), and LHCb [69], dedicated to the study of b hadrons physics.

2.2 ATLAS detector

2.2.1 Overview

The high interaction rates, radiation doses, particle multiplicities and energies, as well as the requirements for precision measurements lead to the design of the ATLAS detector. Due to these experimental conditions, the detector subcomponents require fast and radiation-hard electronics and high granularity to handle the particle fluxes and to reduce the influence of overlapping events. The general requirements for the subdetector are:

- large acceptance in pseudorapidity with almost full azimuthal angle coverage,
- good charged-particle momentum resolution and reconstruction efficiency in the inner tracker,
- very good electromagnetic (EM) calorimeter for electron and photon identification and measurements of their energy, complemented by full-coverage hadronic calorimeter for accurate jet and missing transverse energy measurements,
- good muon identification and momentum resolution over a wide range of momenta, and ability to determine unambiguously the electric charge of high p_T muons are fundamental requirements,
- highly efficient triggering on low transverse-momentum objects with sufficient background rejection.

The overall ATLAS detector layout is shown in Figure 2.3. The detector is cylindrical in shape with a total length of 44 m and a radius of 12 m and it is divided into a barrel section and two end-caps. The total weight is approximately $7 \cdot 10^3$ tons. The ATLAS detector is nominally forward-backward symmetric with respect to the interaction point.

2.2.2 Magnetic System

Two different magnetic fields are generated in the ATLAS detector volume: a central magnetic field, provided by a solenoid surrounding the inner detector, and an outer one, produced by a set of toroids surrounding the muon spectrometer [70]. The central superconducting solenoid provides a central magnetic field of 2 T, while the peak value (at the superconductor face) is 2.6 T. In order to obtain the desired calorimetric performances, in particular for photon and electron energy measurements, a careful design to minimize the amount of dead material in front of the calorimeters has been performed: the solenoid is placed inside the vacuum vessel of the LAr calorimeter. The amount of dead material due to the solenoid and the cryostat wall is of about one radiation length. The magnetic field in the barrel region of the muon spectrometer is provided by a system of 8 coils assembled radially with eight fold symmetry. The magnetic field in the forward region is delivered

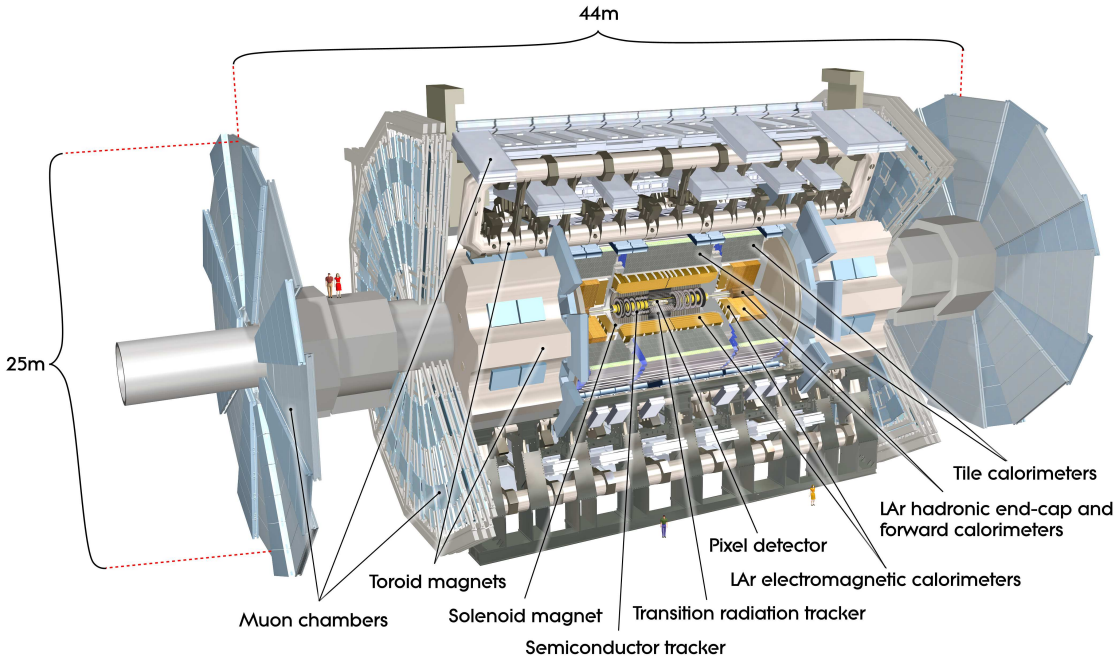


FIGURE 2.3 Cut-away view of the ATLAS detector. The dimensions of the detector are 24 m in height and 44 m in length. The overall weight of the detector is approximately 7000 t.

by the end-cap coils system, which is rotated by 22.5^o with respect to the barrel coils to provide radial overlap and to optimize the bending power in the interface regions of the two coil systems. The peak magnetic field obtainable in the barrel region is about 4 T. The coils of the barrel are 25 m long and their height is 4.5 m and there is one cryostat per coil. In the endcap region there is only one cryostat within which the coils (5 m long and 4.5 m tall) are housed.

2.2.3 Inner detector

At design energy and luminosity approximately 1000 particles emerge from the collision point every 25 ns within $|\eta| < 2.5$, leading to a very large track density in the detector. To achieve the momentum and vertex resolution requirements imposed by the benchmark physics processes, high-precision measurements must be made with fine detector granularity. The strategy used for the ATLAS tracker [71, 72] is to combine few high precision measurements close to the interaction point with a large number of lower precision measurements in the outer radius. The inner detector is immersed in a 2 T solenoidal field and is made from three different technologies at different distances from the interaction point. The combination of three inner layers of pixels and four layers of silicon micro-strips allows hermetic and robust pattern recognition and good secondary vertex identification for charged tracks above a given p_T threshold (nominally 0.5 GeV) and within a pseudorapidity range of $|\eta| < 2.5$. It allows also good momentum measurements over $|\eta| < 2.0$ and a wide range of energies (between 0.5 GeV and 150 GeV). The tracking is completed by continuous straw tube detectors with transition radiation detection capability in the outer part. The structure

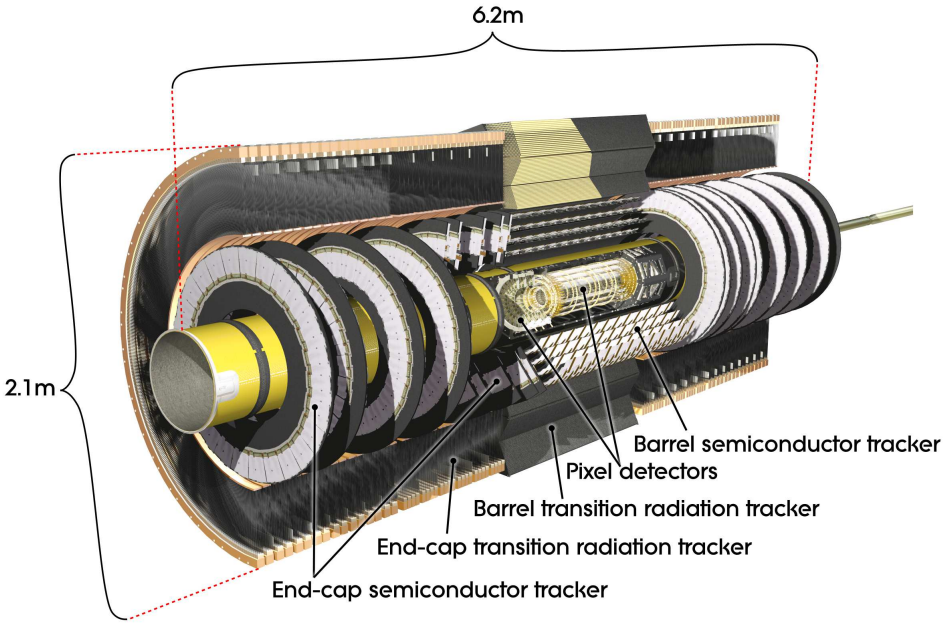


FIGURE 2.4 Cut-away view of the ATLAS Inner Detector [71, 72].

of the inner detector is shown in Figure 2.4. Table 2.2 summarizes the main parameters of the tracking system.

		Radial extension (mm)	Length (mm)
Overall ID envelope		$0 < R < 1150$	$0 < \|z\| < 3512$
Beam-pipe		$29 < R < 36$	
Pixel	overall envelope	$45.5 < R < 242$	$0 < \|z\| < 3092$
3 cylindrical layers	sensitive barrel	$50.5 < R < 122.5$	$0 < \|z\| < 400.5$
2×3 disks	sensitive end-cap	$88.8 < R < 149.6$	$495 < \|z\| < 650$
SCT	overall envelope (barrel)	$255 < R < 549$	$0 < \|z\| < 805$
	overall envelope (end-cap)	$251 < R < 610$	$10 < \|z\| < 2797$
4 cylindrical layers	sensitive barrel	$299 < R < 514$	$0 < \|z\| < 749$
2×9 disks	sensitive end-cap	$275 < R < 560$	$839 < \|z\| < 2735$
TRT	overall envelope (barrel)	$554 < R < 1082$	$0 < \|z\| < 780$
	overall envelope (end-cap)	$617 < R < 1106$	$827 < \|z\| < 2744$
73 straw planes	sensitive barrel	$563 < R < 1066$	$0 < \|z\| < 712$
160 straw planes	sensitive end-cap	$644 < R < 1004$	$848 < \|z\| < 2710$

TABLE 2.2 Main parameters of the ATLAS tracking systems. Radial extensions and lengths are in mm [71, 72].

Pixel Detector

Close to the beam line is the Pixel Detector [73] with approximately 80 million silicon pixel structured in 1744 sensors of size 50×400 μm^2 arranged in three layers in the barrel region and three end-cap disks at large pseudorapidity. At least three space points are measured by the Pixel Detector, leading to the reconstruction of track segments independently from the outer detectors. The

intrinsic hit resolutions in the barrel are 10 μm in $R-\phi$ and 115 μm in z while in the disks are 10 μm in $R-\phi$ and 115 μm in z. The pixel detector covers a pseudorapidity region of $|\eta| < 2.5$ and ensures a high granularity in the area around the proton-proton collisions, where the density of charged tracks is very high. The main purpose of the Pixel Detector is to efficiently reconstruct tracks and vertices at each beam crossing. The innermost layer is commonly referenced as the *b*-layer because it is an essential ingredient to improve the tracks impact parameter which is crucial for the selection and identification of jet originating from *b*-quarks (*b*-jets).

SemiConductor Tracker

The SemiConductor Tracker (SCT) is formed by silicon micro-strips arranged in four nested cylindrical layers located in the barrel and nine disks in each end-cap. In the barrel region, SCT detector uses two different micro-strips with one set of strips in each layer parallel to the beam direction and a relative angle of 40 mrad. This stereoscopic geometry provides the capability to perform three-dimensional position measurements. In the end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad. The modules cover a surface of 63 m^2 of silicon and provide almost hermetic coverage with at least four precision space-point measurements over the fiducial coverage of the inner detector with intrinsic hit resolutions per module of 17 μm in $R-\phi$ and 580 μm in z in the barrel region and of 17 μm in $R-\phi$ and 580 μm in z in the disk. The total number of SCT read-out channels is approximately equal to 6.3 million.

Transition Radiation Tracker

The last tracking subsystem is the Transition Radiation Tracker (TRT), which is both a straw drift-tube tracker and a transition radiation detector. The TRT consists of 2 mm radius straw tubes, arranged in two barrel sections with straws parallel to the beam-axis and in two end-caps with straws arranged radially. In the barrel region, the straws are parallel to the beam axis and are 144 cm long. They are divided into two halves, approximately at $\eta = 0$. In the end-cap region, the 37 cm long straws are arranged radially in wheels. The total number of TRT read-out channels is approximately equal to 350 thousand. The straw tubes are filled with a gas mixture (70% Xe, 27% CO_2, 3% O_2) with inside a tungsten wire. When charged particles cross a straw, they leave a trail of electron-ion pair in their wake. The electrons, drift towards the anode wire, gain energy and create other electron-ion pairs, generating an avalanche process in which a cascade of electron-ion pairs is created. The TRT can provide only $R-\phi$ information, for which it has an intrinsic accuracy of 130 mm per straw, but the combination of precision silicon-based trackers at small radii with the TRT gives very robust pattern recognition and high precision hit measurements in both $R-\phi$ and z coordinates. The straw hits contribute significantly to the momentum measurement, since

the lower precision per point compared to silicon detectors is compensated by the large number of measurements, typically $\sim$ 36 per crossing track, and longer measured track length.

2.2.4 Calorimeters

The basic idea of the calorimeter system is to detect, within layers of active material, the particles created in the shower initiated by an incoming particle which pass through layers of a dense material (absorber). The calorimeter aims to have a precise measurement of the energy and position of the incoming particles, as well as a good estimation of the missing transverse energy in an event. ATLAS calorimetric systems differ in technology and materials depending on the pseudorapidity region. Liquid Argon (LAr) technology is used as active material for the electromagnetic (EM) calorimeters in all pseudorapidity ranges and for the hadronic calorimeter in the end-cap regions (HEC). In the end-cap regions the HEC and the EM calorimeter are hosted in the same cryostat. Different absorbers are used in the different regions: lead for the LAr in the barrel up to $|\eta| < 1.7$ and in the end-caps ($1.5 < |\eta| < 3.2$), copper for the HEC. An homogeneous LAr presampler detector is placed between the cryostat wall and the EM calorimeter in the region up to $|\eta| = 1.8$. In the barrel region ($|\eta| < 1.7$) the hadronic calorimeter is composed of an iron-scintillating tiles calorimeter (TileCal) subdivided into three parts: the central barrel covers up to $|\eta| \simeq 1$, while the two extended barrels cover up to $|\eta| < 1.7$. In the very forward region, up to $\eta \simeq 5$, the system is completed by a very dense LAr calorimeter consisting of rod-shaped electrodes in a tungsten matrix. An overall view of the ATLAS calorimetric system is shown in Figure 2.5 while Table 2.3 shows the details of the segmentation of the calorimeters [74].

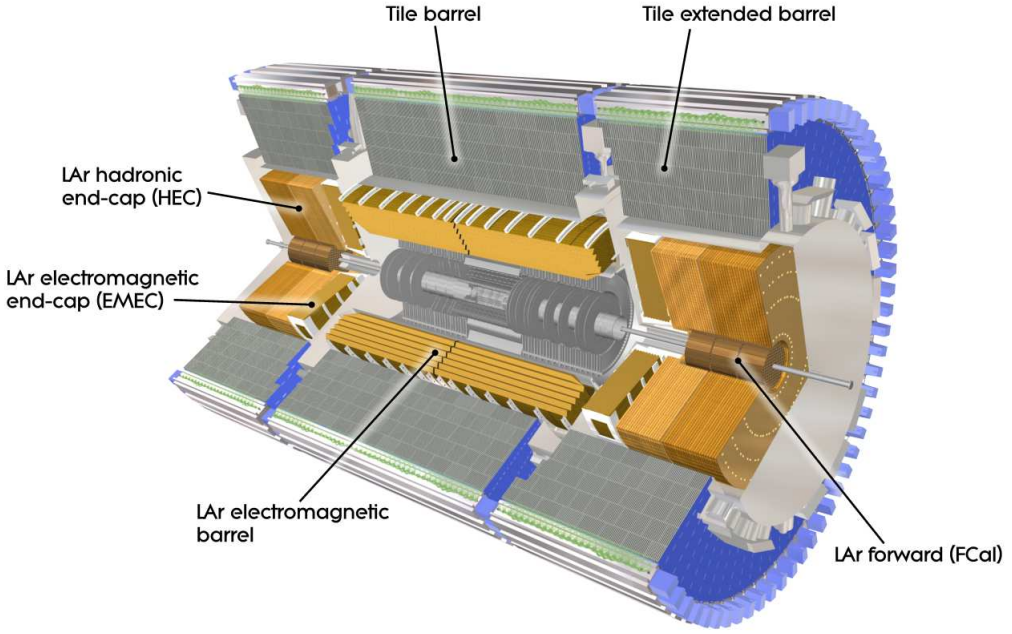


FIGURE 2.5 Cut-away view of the ATLAS calorimeter systems [74].

EM CALORIMETER	Barrel	End-cap	
Coverage	$\|\eta\| < 1.475$	$1.375 < \|\eta\| < 3.2$	
Long. segmentation	3 sampling	3 sampling	$1.5 < \|\eta\| < 2.5$
		2 sampling	$2.5 < \|\eta\| < 3.2$
Granularity($\Delta\eta \times \Delta\phi$)			
Sampling 1	0.003×0.1	0.025×0.1	$1.375 < \|\eta\| < 1.5$
		0.003×0.1	$1.5 < \|\eta\| < 1.8$
		0.004×0.1	$1.8 < \|\eta\| < 2.0$
		0.006×0.1	$2.0 < \|\eta\| < 2.5$
Sampling 2	0.025×0.025	0.025×0.025	$1.375 < \|\eta\| < 2.5$
		0.1×0.1	$2.5 < \|\eta\| < 3.2$
Sampling 3	0.05×0.025	0.05×0.025	$1.5 < \|\eta\| < 2.5$
PRESAMPLER	Barrel	End-cap	
Coverage	$\|\eta\| < 1.52$	$1.5 < \|\eta\| < 1.8$	
Long. segmentation	3 sampling	3 sampling	
Granularity($\Delta\eta \times \Delta\phi$)	0.025×0.1	0.025×0.1	
HADRONIC TILE	Barrel	End-cap	
Coverage	$\|\eta\| < 1.0$	$0.8 < \|\eta\| < 1.7$	
Long. segmentation	3 sampling	3 sampling	
Granularity($\Delta\eta \times \Delta\phi$)			
Sampling 1 and 2	0.1×0.1	0.1×0.1	
Sampling 3	0.2×0.1	0.2×0.1	
HADRONIC LAr		End-cap	
Coverage		$1.5 < \|\eta\| < 3.2$	
Long. segmentation		3 sampling	
Granularity($\Delta\eta \times \Delta\phi$)		0.1×0.1	$1.5 < \|\eta\| < 2.5$
		0.2×0.2	$1.5 < \|\eta\| < 3.2$
FCAL CALORIMETER		Forward	
Coverage		$3.1 < \|\eta\| < 4.9$	
Long. segmentation		3 sampling	
Granularity($\Delta\eta \times \Delta\phi$)		0.2×0.2	

TABLE 2.3 Design parameters of the ATLAS calorimeter systems [74].

2.2.4.1 The electromagnetic calorimeter

As mentioned above the EM calorimeter is a sampling calorimeter, using lead as absorber and liquid argon as active material. It is segmented in three parts of different granularity. The first part close to the tracking system is a fine granularity in pseudorapidity and azimuthal angle, to provide a precision η measurement and to improve the γ/π^0 and e/π^0 separation. The thickness of the EM calorimeter is more than 24 radiation lengths (X_0) in the barrel and 26 X_0 in the end-caps. The energy resolution is given as a function of the energy E of incoming particle (in GeV) by the formula (2.2) in which the first term takes into account the statistical fluctuation due to the development of the shower, and the second one is a constant term that takes into account several systematic errors, like the inhomogeneity in the calorimeter response.

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \tag{2.2}$$

2.2.4.2 The hadronic calorimeter

The EM calorimeter is surrounded by the hadronic one. The barrel part uses the iron as absorber and scintillators as active materials. The end-cap hadronic calorimeter receives a much higher radiation dose and therefore uses the intrinsically radiation-hard LAr technology. The thickness of the hadronic calorimeters is more than 10 hadron interaction length. The equation (2.3) gives the design energy resolution for the barrel hadronic calorimeter, whereas the equation (2.4) is for the end-cap part.

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad , \tag{2.3}$$

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \tag{2.4}$$

2.2.4.3 The forward calorimeter

The forward calorimeter covers the pseudorapidity region $3.1 < |\eta| < 4.9$ and uses LAr technology with copper and tungsten as absorber. It consists of an electromagnetic part and two hadronic parts along the longitudinal direction. To avoid back-scattered neutrons in the Inner Detector system, the forward calorimeter is placed 1.2 meter further away from the interaction point, compared to the electromagnetic end-cap calorimeter. The design energy resolution of the forward calorimeter is:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \tag{2.5}$$

2.2.5 Muon Spectrometer

The calorimeter system is surrounded by the muon spectrometer. It consists of an air-core toroid system, with a long barrel, in the central region, and two inserted end-cap magnets, for the coverage at small angles. They generate strong magnetic field in a large volume with a relatively light structure. Multiple-scattering effects are therefore minimized, allowing an excellent muon momentum resolution with three layers of high precision tracking chambers. The muon spectrometer defines the overall dimensions of the ATLAS detector.

The main features of the muon spectrometer [75] is the possibility of a precise standalone measurement of the muon momentum. The magnetic field provided by the superconducting air-core

toroid magnets deflects the muon trajectories that are measured by high precision tracking chambers. The magnetic field in the $|\eta| < 1.0$ range is provided by the barrel toroids, while the region $1.4 < |\eta| < 2.7$ is covered by the end-caps. In the so called transition region ($1.0 < |\eta| < 1.4$) the combined contributions of both the barrel and end-caps provide the magnetic field coverage. In the barrel region, the muon chambers are arranged in three cylindrical layers (stations), while in the end-cap they form three vertical walls. The transition region is instrumented with one extra station. Figure 2.6 offers a three dimensional view of the spectrometer. The azimuthal layout follows the magnet structure with 16 sectors. The so-called Large Sectors lie between the coils, and they overlap with the Small Sectors, placed in correspondence with the coils themselves. The choice of the different chamber technologies follows the particle flux expectation in the different regions of the detector. Criteria of rate capability, granularity, aging properties and radiation hardness have been considered. Table 2.4 summarizes the chamber technologies used in the various pseudorapidity regions. The measurement of the track non-bending coordinate (ϕ) is provided in most of the η region by the Monitored Drift Tubes (MDT), while at large pseudorapidity, the higher granularity Cathode Strip Chambers (CSC) are used.

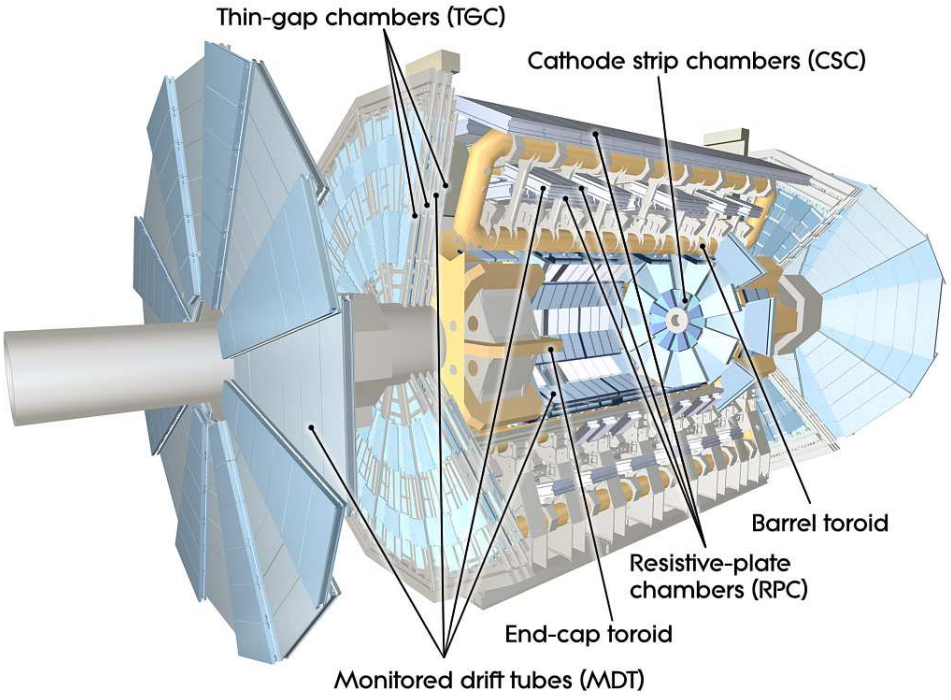


FIGURE 2.6 Cut-away view of the ATLAS muon system [75].

Region		station I	station E	station M	station O
Barrel	$\|\eta\| < 1.$	MDT		MDT RPC	MDT RPC
End-caps	$1. < \|\eta\| < 1.4$	MDT TGC	MDT		
	$1.4 < \|\eta\| < 2.$	MDT TGC		MDT TGC	
	$2. < \|\eta\| < 2.4$	CSC			MDT

TABLE 2.4 Design parameters of the Muon spectrometer [75].

To reach the transverse on the momentum resolution of $\Delta p_T/p_T \simeq 10\%$ at 1 TeV requires an accuracy of the relative positioning of chambers traversed by a muon track that matches the intrinsic resolution and the mechanical tolerances of the precision chambers. The knowledge of the chamber positioning with an accuracy of 30 μm is required within a projective tower. The accuracy required for the relative positioning of different towers to obtain adequate mass resolutions for multi-muon final states is in the millimeter range. This accuracy can be achieved by the initial positioning and survey of the chambers at the installation time. The relative alignment of muon spectrometer, calorimeters and ID relies on the measurement of the high-momentum muon trajectories. The MDT chambers are equipped with an in-plane alignment system aiming at a measurement of the tube position displacements, with respect to their nominal positions at the assembly phase, with a precision of better than 10 μm. To achieve this, the spectrometer is equipped with a laser, mounted at one side of a chamber which project a pattern to a CCD camera positioned at the other end of the chamber. From the displacement of the pattern-figure with respect to what is expected, corrections for chambers deformation can be computed. The chambers for the first level (LVL1) muon trigger system covers the region $|\eta| < 2.4$. Resistive Plate Chambers (RPC) are used in the barrel region, while the Thin Gap Chambers (TGC) are used in the end-cap. Their first task is to identify without any ambiguity the bunch crossing of the triggered event. This requires a time resolution of better than 25 ns. Next, they have to provide a well defined p_T cutoff for the LVL1 choice. This is obtained considering a window of a size defined by the LVL1 p_T threshold considered on the second RPC (or TGC) station once a super-hit has been obtained in the first station. Finally, the trigger chambers measure the bending coordinate (ϕ), in a plane orthogonal to that measured by the precision chambers, with a typical precision of 5-10 mm.

2.3 Trigger system

The ATLAS trigger system is designed to reject the largest rate of the background and at same time to select with a satisfactory efficiency the potential interesting events. The ATLAS trigger and data-acquisition system is based on three on-line event selection. More detail on the trigger system, as well as the data-taking and monitoring are presented and discussed in the Chapter 3.

2.4 Computing

The complexity of the ATLAS experiment imposes the use of new paradigms concerning the data processing once they are made available on mass storage. The event rate of few hundred Hz (see Chapter 3), the size of the events ($\sim$ 1.6 MB per event) and the number of physicists involved in the analysis require that the data distribution, processing and analysis is carried out according to a multi-tier schema that is well suited to distribute the computing and storage loads among the

different participating institutes. Similar strategies have been used in the past for other experiments, but it is the first time that this kind of distributed analysis is performed on an LHC-size scale requiring the development of completely new software tools [76]. At the output of the data acquisition system, the raw data are transferred to the CERN computing center, called Tier-0, the first layer of the ATLAS analysis system. In Tier-0 is stored a complete copy of the raw data and a first-pass reconstruction is made producing ESD (Event Summary Data) and AOD (Analysis Object Data). The ESD data-format contains the reconstructed quantities measured by the detector (energy in the calorimeter cells, clusters information, tracks, vertices) as well as reconstructed physics objects (electrons, photon, jets, taus, muons). The event size foreseen for the ESD format is about 0.8 MB/evt. The small-sized data in AOD format (0.15 MB per event) are well suited for data distribution. Here only the physics objects are recorded. Tier-0 has also the responsibility to run calibration and alignment algorithms that will be refined in future steps. The distribution of the data to the ATLAS community is performed by copying raw data, ESD, AOD to the Tier-1s. Tier-1s are big regional or national computer centers spread around the world. Tier-1s have also the responsibility to reprocess raw data performing more accurate reconstructions. Updated version of ESD, AOD are therefore constantly produced and spread among the different computer centers. Most of the physics analysis is performed at the Tier-2 centers. The Tier-2 are allowed to connect to different Tier-1s and Tier-2s from a different cloud. They have the responsibility for the official Monte Carlo production (the simulated data are stored in the Tier-1s) and physics analysis. The development and refinement of calibration and reconstruction algorithms are also performed at the Tier-2 centers. The physics analysis are performed on AOD data sets or on even more compact Derived Physics Data (DPD). DPD format is a subsample reduced data set with stricter event selection, reducing in size the information per object and dropping unwanted data objects. The multi-tier paradigm is deployed using GRID technology [77].

Chapter 3

ATLAS Trigger system

An efficient trigger system is required to select the events produced by the proton-proton collision in the LHC challenging environment. The aim of data acquisition systems of the LHC experiments consists in filtering and selecting the relevant physics events from the background of soft interactions. Section 3.1 provides an overview of the ATLAS trigger structure including a general description of its three trigger levels. In particular, a description of the jet and muon trigger as well as the tracking algorithms used in ATLAS are provided in Section 3.2. Section 3.3 gives a detailed description of the b-jet trigger. Section 3.4 is devoted to the data quality online monitoring of the b- and μ-jet trigger algorithms.

3.1 ATLAS trigger infrastructure

The trigger of the ATLAS experiment [78] is designed as a three level system that reduces the event rate from 40 MHz to about 500 Hz at which events[1] can be written to mass storage. Figure 3.1 gives an overview of the ATLAS trigger and data-taking system showing its three levels structure and describing the different steps of the trigger system, starting from the input signals received from the calorimeter and the muon spectrometer up to the storing of the data events. The input rate of the proton-proton collision and the consequent reduction performed by the trigger system are also presented. Each step refines the previous decision by using a larger fraction of the data and more advanced and time-demanding algorithms. The difficult task at each trigger consists on reaching a decision quickly enough to handle the output rate of the previous level.

[1] The average size per event is of about 1.6 MB

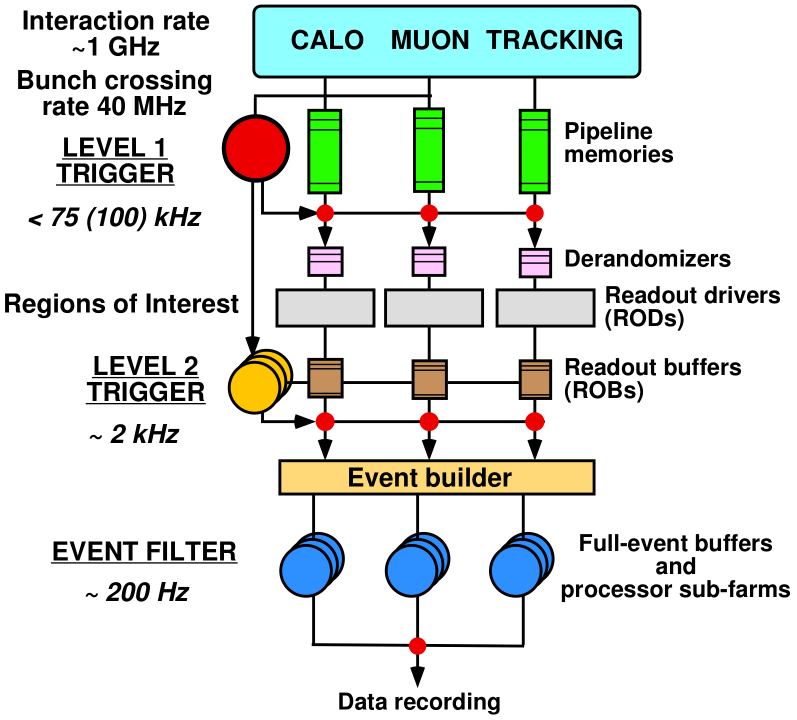


FIGURE 3.1 Schematics of the ATLAS trigger and data-taking system. The rates shown are derived for the instantaneous luminosity used in the early data-taking period.

3.1.1 First Level Trigger (LVL1)

The LVL1 trigger is a hardware-based system that receives signals from the calorimeter and muon detectors of ATLAS. Its task is to reduce the event rate from 40 MHz to 75 kHz within a latency of 2.5 μs. During that time the data from all detectors are stored in pipeline memories. The LVL1 objects information, recorded in the so called 'Regions-of-Interest' (RoI) (spatially limited areas in the detector with candidates for phenomena to be triggered) are sent to the Central Trigger Processor (CTP), which implements the different trigger combinations, trigger menu, as well as the possible configuration pre-scale factors[2]. It provides information to the High Level Trigger for selected events indicating which signatures were fulfilled. The total number of allowed L1 configurations (also called L1 items) that can be deployed at any time is 256. The data-taking run is subdivided into time ranges of about one minute, called luminosity blocks. The luminosity blocks represent the smallest size at which the data will be monitored and available for the physics analysis. Within a lumonosity blocks the trigger menu, configuration and pre-scale values, remain unchanged.

[2]A trigger pre-scale allow the optimization of the bandwidth usage for different luminosity and background conditions by recording only part of the data triggered: the portion of the recorded data is governed by the pre-scale factor.

3.1.2 Second Level Trigger (LVL2)

The LVL2 trigger is based on software selection algorithms running in processor farms. LVL2 can access data from all sub-detectors of ATLAS in the RoIs that were identified by the LVL1 system. A seed is constructed for each trigger accepted by LVL1 that consists of a p_T threshold and an $\eta - \phi$ position. The LVL2 algorithms use this seed to construct an RoI window around the seed position.
LVL2 is the first stage of the ATLAS trigger system that has access to data from the tracking sub-detectors. Hence specific algorithms to select events containing jets originating from b-quarks can be implemented at this stage. The processing time available for LVL2 algorithms is 40 ms in average, during this time the trigger algorithm should be able to perform a its events rejection. The LVL2 system must provide a reduction of the LVL1 input rate from 75 kHz to 2 kHz at nominal operations.

3.1.3 Event Filter (EF)

The EF is also based on software selection algorithms. With respect to LVL2, it runs after the event building, thus the complete event information is available to the EF algorithms. Each accepted LVL2 trigger can be used to seed a sequence of EF algorithms which provide a more refined and complete analysis. The input EF rate is equal to 2 kHz during nominal operations and must provide the additional necessary rejection to reach the output rate of $\sim$ 500 Hz, corresponding to a data rate size of 600 MB/s. In the EF, a thorough event selection and classification process is performed within a time budget of about 4 seconds. The EF algorithms are foreseen to use offline reconstruction code with the full calibration and alignment informations. Events accepted by the EF are written to mass storage and the output rate from the Event Filter is limited by the offline computing budget and storage capacity.

3.2 Trigger objects

Several objects are used in the definition of the trigger chains: electron, muon, tau, jet, track and b-jet. In the next sections jet, muon and b-jets which are used either in the physics analysis described in this dissertation or in the b-jet trigger data-quality monitoring.

3.2.1 Jet trigger objects

The calorimeter trigger receives as inputs 7200 analogue signals from a dedicated trigger-tower electronics that locally combines information from calorimeter cells in the various ATLAS calorimeters. A trigger tower has a typical granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The LVL1 jet trigger constructs jet

elements using 2×2 towers in the electromagnetic and hadronic calorimeters, leading to a granularity of $\Delta\eta\times\Delta\phi = 0.2\times0.2$ for jet trigger in the central pseudorapidity region and $\Delta\eta\times\Delta\phi = 0.4\times0.4$ in the forward region (Figure 3.2). The LVL1 jet trigger identifies objects within a pseudorapity range of $|\eta| < 3.2$. The task of the calorimeter trigger is to look for localized energy depositions that

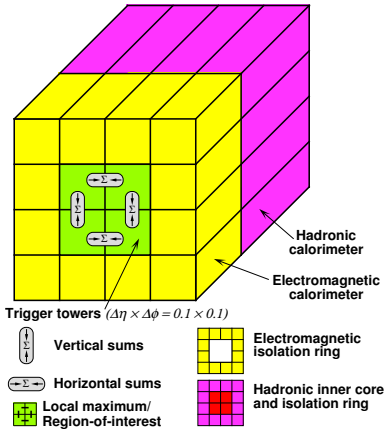


FIGURE 3.2 Schematic diagram of the Level-1 jet algorithm showing a window of 4×4 jet elements spanning the electromagnetic and hadronic calorimeters in depth, and a local maximum transverse energy cluster of 2×2 jet elements [79].

are the signatures for high transverse energy electrons/photons, τ particles, hadrons or jets. The most common LVL1 trigger level inputs are the multiplicities and the transverse energy thresholds of these objects. The energy depositions are compared to a set of programmed transverse-energy thresholds and the multiplicity of objects passing each of the thresholds is counted. Other observables can be used to create a trigger configuration, such as the sum of scalar energies in the event or the missing transverse energy. The type, position and threshold information about each candidate object are collected in the RoIs. For selected events, these RoIs are sent to the HLT[3] via the Region-of-Interest Builder (RoIB) in order to seed the LVL2 selection.

The LVL2 jet trigger algorithm is seeded by the LVL1 RoI and has data access from a limited region of the detector centered around the LVL1 seed. The calorimeter data are transfered from the detector readout drivers (RODs). The byte streams are converted into more refined objects, such as calorimeter cells, which are processed by the jet algorithm. A simple cone-like jet algorithm with a radius of $R = 0.4$ determines the energy-weighted center of the RoI. The output of the LVL2 algorithm is a reconstructed jet with a given energy and position in η and ϕ. Due to the non-compensative[4] nature of the ATLAS calorimeters, the electromagnetic scale, used to compute the reconstructed jet energy, underestimates the hadronic energy correction, so a reweighing of the jet energy is necessary to take into account this effect. In particular two weights are applied to calibrate the reconstructed jet energy [80]: one for the total energy deposited in the EM calorimeter and one for the total energy deposed in the hadronic calorimeter.

[3]The HLT trigger encompasses a LVL2 followed by the EF trigger.

[4]The average ratio between signals from electromagnetic and hadronic particles of the same incident energy is calorimeter- and energy-dependent, and for non-compensating calorimeters there is a higher response for electromagnetic particles, $e/h \neq 1$.

Finally the EF runs any offline jet algorithms, either RoI based or full event reconstruction. The EF default configuration runs in the full scan mode. The EF jet reconstruction uses as input any E_T energy ordered list of calorimeter objects. At first it reconstructs three dimensional calorimeter clusters. These clusters are used as input to the jet reconstruction which is performed using an anti-k_t algorithm with a radius of $R = 0.4$ or $R = 1.0$.

3.2.2 L1.5 Jet Trigger

Studies on the RoI approach show that it suffers for multi-jet trigger from a decreasing in the performance caused by the low efficiency in the identification of the close-by-jets[a]. In order to remove this problem, a new strategy, so called L1.5, of the LVL2 trigger algorithm has been implemented [81]. Instead of the basic LVL2 algorithm which reconstructs jets from calorimeter cell information exclusively in RoIs seeded by LVL1 algorithm, the L1.5 trigger uses data produced by the LVL1 calorimeter to reconstruct jets across the entire detector, giving the possibility to access either the trigger granularity for the electron/photon (0.1×0.1) trigger or for the jets (0.2×0.2), as shown in Figure 3.3. Several improvements can be provided using the L1.5 algorithm:

- study the entire detector at LVL2,
- run different, more modern, jet algorithms,
- enhance LVL2 input rate,
- apply jet specific calibrations to L1 calorimeter based jets.

This algorithm was in place during the 2012 data-taking period. It was not available for the 2011 data used for the physics analyses presented in this dissertation.

[a] Close-by-jets represents the non isolated jets.

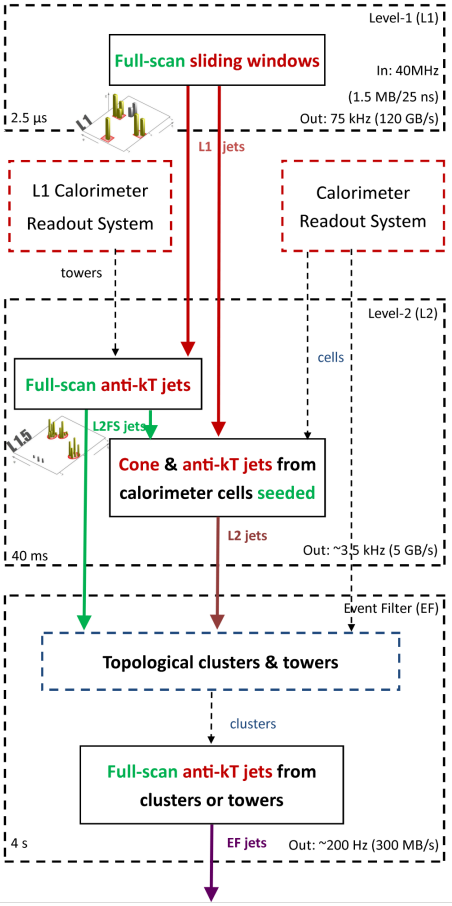


FIGURE 3.3 Architeture of ATLAS Jet Trigger System. The topological cluster label identifies the clusters derived from calorimeter cells by adding the energy in neighbouring cells if they have an energy larger than a predefined threshold.

3.2.3 Jet trigger performance

The efficiency of a jet trigger item, for example for a jet reconstructed with an anti-k_T cone algorithm ($R = 0.4$) (see Section 4.4.1), is defined as the fraction of reconstructed jets which are

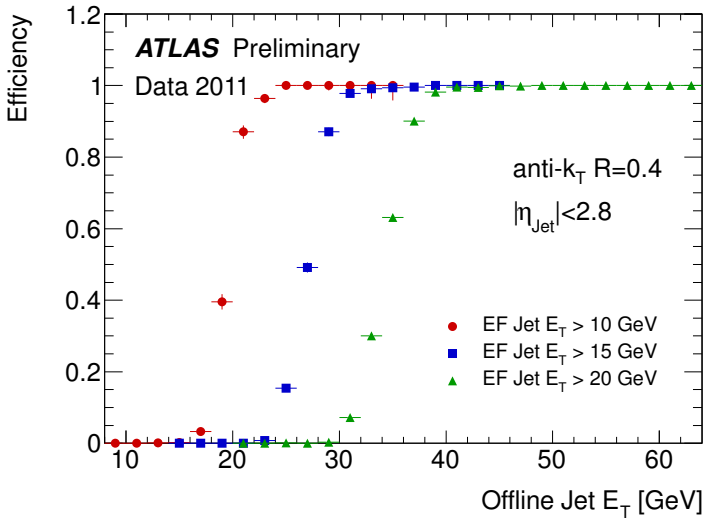


FIGURE 3.4 The efficiency for anti-kt jets with $R = 0.4$ to satisfy the Event Filter inclusive jet trigger for three choices of threshold. The efficiency is plotted as a function of the offline calibrated jet E_T for jets with central rapidities $|\eta| < 2.8$ [82].

matched to a trigger jet within a distance of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.4$ and passed the trigger threshold. An example of efficiency as function of the offline calibrated jet E_T for reconstructed jets in the central pseudorapidity region $|\eta| < 2.8$ is shown in Figure 3.4. The different curves correspond to different Event Filter trigger thresholds. It can be noticed that an EF jet trigger with a 10 GeV threshold reaches full efficiency with respect to offline jets only at $E_T \sim 25$ GeV.

3.2.4 Tracking objects

During the 2011 two LVL2 silicon tracking algorithms were available in ATLAS, namely SiTrack and IDScan [79]. They perform pattern recognition steps, starting from space points reconstructed from clusters of hits in the pixel layers and from the couples of clusters from two SCT stereo layers. Both algorithms exploit a different track-finding approach and share common tools for track fitting and extrapolation to TRT detector. The main steps for both track algorithms are:

- SiTrack is based on a combinatorial approach. It looks for pairs of space points in the inner layers consistent with the beam-line constraints, then combines the pairs with space points in outer layers to form triplets and merge the triplets to define the track candidates. It is the default algorithm for the *b*-jet trigger,
- IDScan is based on a projective approach. It reconstructs the position of the primary vertex along the beam-line, then it identifies clusters of hits in $\eta - \phi$ plane and performs a combinatorial tracking reconstruction on the groups of hits pointing back to the selected region where the *pp* collision occurred. The groups of hits space points are considered as track candidates.

For 2012 a new strategy of the track reconstruction algorithm is developed and optimized in ATLAS. The baseline of the new algorithm is the unification of tracking algorithms used previous at $\sqrt{s} = 7$ GeV; the framework is known as L2Star [83] and provides a common utility for comparing the candidate tracking algorithms, such as IDScan and SiTrack, as well as an adapted version of the offline tracking, It starts by retrieving the RoI and the associated space points and storing them into structures for the pattern recognitions. Several algorithms are implemented as tools that are used by different pattern recognition strategies to perform optimal tracking for a given trigger object (like electron, muon, tau, RoIs, b-jet). L2Star can be configured with different strategies which can use any of the different pattern recognition tools within this framework. Standard tools can be combined in a modular way within a new strategy, without any code duplication. New strategies can easily be added to support developments planed for detector upgrades.

The tracking reconstruction at the EF trigger level is performed using the ATLAS offline tracking software [84]. The tracking includes two sequences, the inside-out track reconstruction, as in the LVL2 approach, and the outside-in tracking, which starts from the TRT segments and search for the matching hits in the inner silicon detectors. The latter method aims at reconstructing late decays of neutral particles, like photon conversions to electron-positron pairs. The EF track reconstruction has a similar algorithm sequence compared to the standard inside-out approach.

3.2.5 Muon trigger objects

The LVL1 muon trigger is a custom hardware based that processes input data using fast muon detectors. The LVL1 RoI are selected using the RPC in the barrel for $|\eta| < 1.05$ and the TGC in the $1.05 < |\eta| < 2.4$ region. The trigger looks for a hit coincidence in η and ϕ within different detector layers from which the transverse momentum is defined. The LVL1 provides the η and ϕ coordinates of the selected RoI as well as the p_T. The transverse momentum is divided in six programmable thresholds. The RoI informations are used to seed the HLT trigger which consists of a fast stage L2 and EF. The LVL1 muon candidates are refined at L2 including the precision data from the MDT's. The LVL2 algorithm can access to data information either in the LVL1 RoI or in the full detector. The muon candidate information are refined by a fitting algorithm which is performed using the MDT drift time (muFast). The transverse momentum is assigned by Look-up-Tables (LUT) [79]. The LVL2 combined algorithm (muComb) uses the informations of the inner detector tracks to define a combined muon candidate. The reconstructed tracks in the inner detector are combined with the muon candidate found by the muFast. The muon spectrometer and inner detector combination allows to reject muons from the cosmic radiation. A third algorithm is accessible at LVL2 in which the information of the muon spectrometer, inner detector and calorimeter to find an isolated muon are combined. It is seeded by the muComb candidate and estimates the electromagnetic and hadronic energy in the cone around the muon direction.

At the EF level the full data are accessible. The starting point is the RoI defined by the LVL1 and LVL2 and reconstructs the tracks using the trigger and the precision chambers. The track is extrapolated to the interaction point to define the muon candidate. As in the LVL2, the muon candidate is combined with the inner detector tracks to define the combined candidate. This strategy, outside-in, is completed by a second algorithm which starts from the inner detector tracks and extrapolates to the muon spectrometer leading to an inside-out muon candidate.

3.3 On-line *b*-tagging algorithm

Due to the huge QCD multi-jet events rate produced in pp collisions at the LHC, the selection of rare signal events should start at trigger level. For example, in fully hadronic final state analyses, such as the $t\bar{t}$ in the fully hadronic decay channel, the requirement of large amount of multi-jets background events is usually obtained by increasing the inclusive jet p_T thresholds at trigger level. To avoid the consequent loss of signal events a possible solution consists on applying in the trigger selection strategy an algorithm which separates b-jets from light-jets and gluon jets. The b-tagging consists on the identification of jets stemming from the fragmentation and hadronization of b-quarks. By lowering the jet transverse momentum thresholds at LVL1 and applying the b-tagging selection in a second step, the acceptance of events with jets originating from b-quarks increases, whereas the background rate reduces: the rejection power of the b-tagging requirement at the HLT compensates for less rejection due to the lower LVL1 and HLT jet p_T thresholds.

3.3.1 The role of the *b*-tagging in physics analyses

The b-tagging relies on several physical propeties which characterize the presence of b-hadrons and allows for their discrimination from the light-quarks, these propeties are:

- hard fragmentation: b-hadrons retain about 70% of the original b-quark momentum,
- large mass, (above 5 GeV/c^2): b-hadron decay products are characterized typically by large transverse momentum with respect to the jet axis and a large angle,
- long lifetime ($\tau \sim 1.6$ ps [7]): significant flight path length: considering a b-hadron in a jet with $p_T = 50$ GeV the flight mean path length before decaying is $\langle l \rangle = \beta\gamma c\tau \approx 3$ mm.

The described proprieties of a b-hadron allow the identification of the b-jets thanks to the:

- presence of a secondary vertex which represents the decay position of the long lived particles, like b-hadron, displaced with respect to the primary vertex which signs the hard proton-proton collision. The distance between these two vertices represents the flight path length of the b-hadron,

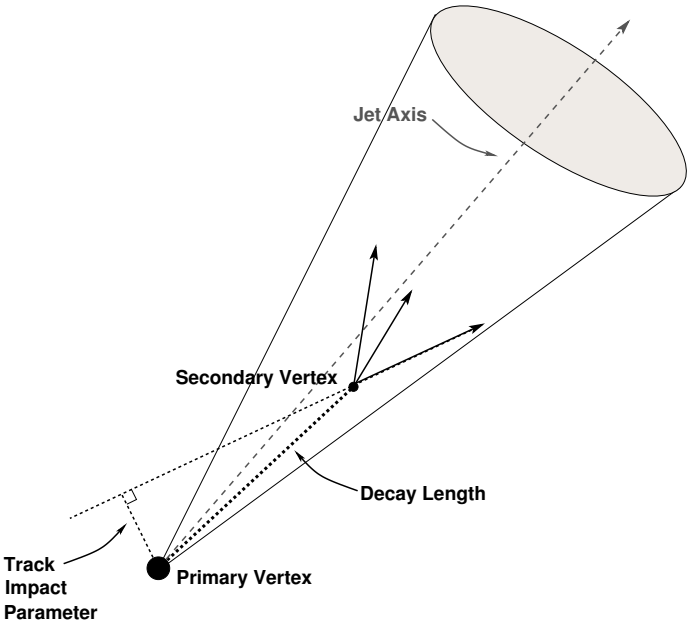


FIGURE 3.5 A secondary vertex with a significant decay length indicates the presence of a long-lived particle in a jet. The secondary vertex is reconstructed from tracks with a large impact parameter significance with respect to the primary vertex.

- large transverse impact parameter of the tracks originating from the secondary vertex. The impact parameter of these tracks represents their closest approach to the primary vertex,
- larger tracks multiplicity with respect to the light-jet.

Figure 3.5 gives a schematic view of the primary and the secondary vertices and the transverse impact parameter of a track originating from the secondary vertex.
The b-tagging algorithms exploiting the long lifetime of the b-hadron are usually divided in two categories: algorithms relying on tracks with impact parameter values not pointing to the primary vertex position and algorithms relying on the presence of a displaced secondary vertex.

3.3.2 *b*-jet trigger implementation

The b-jet selection at the trigger level is based on the tracks information, which are evaluated at the second level of the trigger. The track reconstruction at HLT starts from the RoIs, identified at LVL1. Several track algorithms are available [79]; the default one for the b-tagging is SiTrack. During the 2011 data-taking period, the b-tagging algorithm ran on the jet RoI with $\Delta\eta$ and $\Delta\phi$ dimension equal to 0.8. In order to reduce data access and processing time the LVL2 algorithms were performing in half the size in η and ϕ. In the 2012 the LVL2 and the EF jet triggers can access the full detector scan, the b-tagging algorithm uses the direction of the L2 or L1.5 or EF RoI.
The b-tagging algorithm follows three main steps: tracking reconstruction of charged particles, estimation of primary vertex interaction, definition of the discriminant variable. This chain runs

at LVL2 and consequently at EF level, if the event has passed the second level trigger.

3.3.2.1 Primary vertex reconstruction

The longitudinal impact parameter of each track with respect to the primary vertex is used to compute one of the most important b-tagging discriminant variable. The interaction point of the two beam is not known a priori, hence it must be reconstructed using the track information in the RoI. In the transverse plane the primary vertex is not calculated in the MC simulation. It is set to the origin of the coordinate system and in data sample the beam-spot position information are used to compute the transverse impact parameter. The algorithm used to reconstruct the primary vertex position along the beam-line is identical at LVL2 and EF with some difference in the performance: at LVL2 (EF) the efficiency is about 98% (99%) and the resolution on the z-position is of about 120 μm (100 μm).

3.3.2.2 *b*-tagging discriminant variables

The most important quantities used in the implementation of the b-jet associated track selection are the transverse and longitudinal impact parameter of the reconstructed tracks.

A common way to built a b-tagger discriminant variable is using the transverse impact parameter d_0 (the transverse component of the track impact parameter is in Figure 3.5).
The transverse impact parameter d_0 is not used directly to build a discriminant variable, what it is used is its significance defined as $S(d_0) = d_0/\sigma(d_0)$ where $\sigma(d_0)$ represents the error on the impact parameter. $S(d_0)$ provides the information on the track distance from the primary vertex position.

A sign is derived by considering the dot product of the jet axis direction and the line connecting the primary vertex position to the point of the closest approach of the track to the beam-line. If the track crosses the jet axis in front of the primary vertex it is positive, otherwise it is negative. Usually the tracks coming from a b-hadron decay have a positive sign whereas the experimental resolution generates a random sign for the tracks originating from the primary vertex. Since the direction of the jet axis enters into the calculation of the sign of the transverse impact parameter, a high angular resolution is necessary to achieve a good b/light-quark jet discrimination.
Figure 3.6 shows data (2011 pp collisions at $\sqrt{s} = 7$ TeV) to MC simulation comparison for the signed transverse impact parameter significance of reconstructed tracks at the EF level; the MC simulation sample is subdivided into flavour jets (light-, c- and b-jet) composition underlying the difference between the different components.

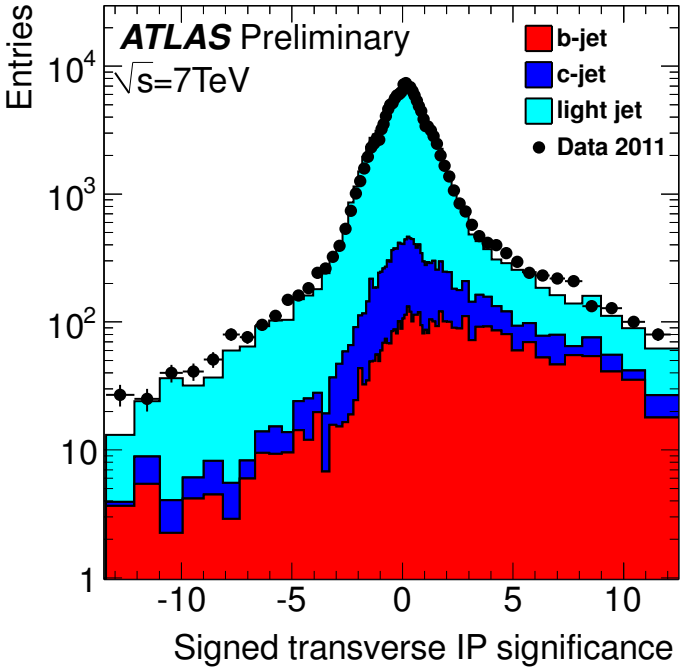


FIGURE 3.6 Signed transverse impact parameter significance of reconstructed tracks at the Event Filter level. Tracks are reconstructed starting from a low p_T jet identified by the Level 1 and are required to fulfill online b-tagging criteria [85]. The width of the signed transverse impact parameter distribution is connected to the tracking detector resolution and multiple-scattering effects.

The longitudinal transverse parameter z_0 can also be used as a discriminant variable but it has a lower discrimination power due to limited detector resolution along the z-axis.

3.3.2.3 JetProb method

During the 2010 and 2011 data-taking period the default b-jet tagger was based on the transverse impact parameter. The so called JetProb technique was developed by the ALEPH collaboration at LEP and was used by Tevatron experiments [86–88].

The signed transverse impact parameter significance of tracks within a jet is used by the JetProb algorithm to compute the probability of a track to originate from the primary vertex. The probability for a track i to originate from the primary vertex is defined by the comparison between the signed transverse impact parameter significance of each track in the RoI and a resolution function R for prompt tracks:

$$P_i = \int_{-\infty}^{-|d_0^i/\sigma_{d_0}^i|} R(x)dx \tag{3.1}$$

The JetProb probability is measured considering the probabilities of all tracks with positive transverse impact parameter in a jet:

$$P_{jet} \equiv P_0 \sum_{i=0}^{N_{tracks}-1} \frac{(-lnP_0)^i}{i!} \tag{3.2}$$

where $P_0 \equiv \sum_1^{N_{tracks}} P_i$. P_{jet} represents the probability that a set of tracks does not have decay products coming from long-lived particles. P_{jet} has a distribution between 0 and 1: tracks originating from a b-hadron have $P_{jet} \sim 0$, while tracks coming from the primary vertex are more compatible with $P_{jet} \sim 1$.

In 2011 three different thresholds, called working points, at LVL2 and at EF were defined; they correspond to 70% (loose), 55% (medium) and 40% (high) b-tagging efficiency. They were derived from MC sample of $t\bar{t}$ simulated events. Figure 3.7 shows the data (black dots) to Monte Carlo simulation (yellow histogram) comparison for the offline JetProb distribution corresponding to P_{jet} value. In the same figure the JetProb distribution after the application of the three b-tagging criteria at both trigger levels is also shown. As one can notice there is a bias in the offline JetProb weight distribution of data collected with a b-jet trigger requirements due to the trigger level selection on JetProb.

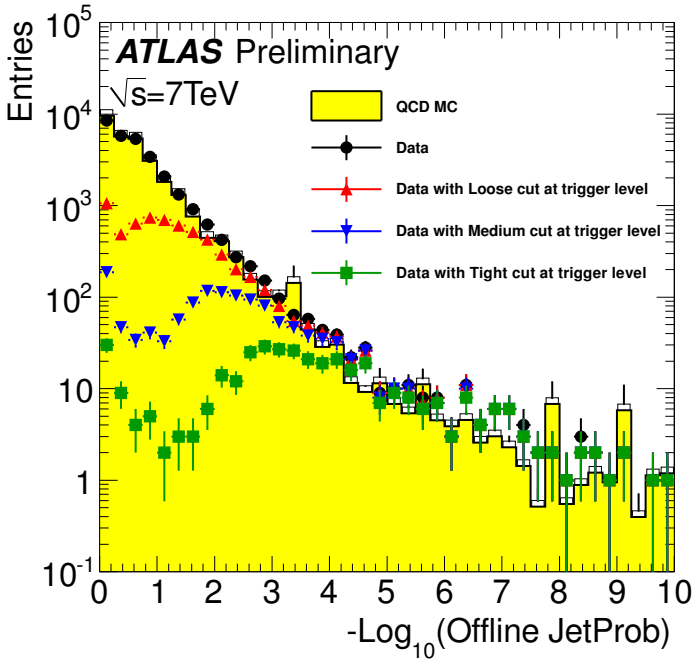


FIGURE 3.7 Offline JetProb distribution in data and simulation. The same distributions in data are shown when a b-jet requirement is added at the trigger level (both Level 2 and Event Filter) [85]. The Loose label means that the working point selected at trigger level corresponds to 70% b-tagging efficiency.

3.3.2.4 The likelihood-ratio method

A second discriminant variable is built by exploiting a *likelihood-ratio* method applied to the impact parameter distributions obtained for signal and background samples. The likelihood-ratio is able to separate two or more event categories using characteristic observables of the event. This method was not implemented directly in the early data-taking since the probability density function of the likelihood method are extracted from MC simulation and need to be carefully validated on real data. The ratio between the probability distributions for two alternative hypotheses of n individual

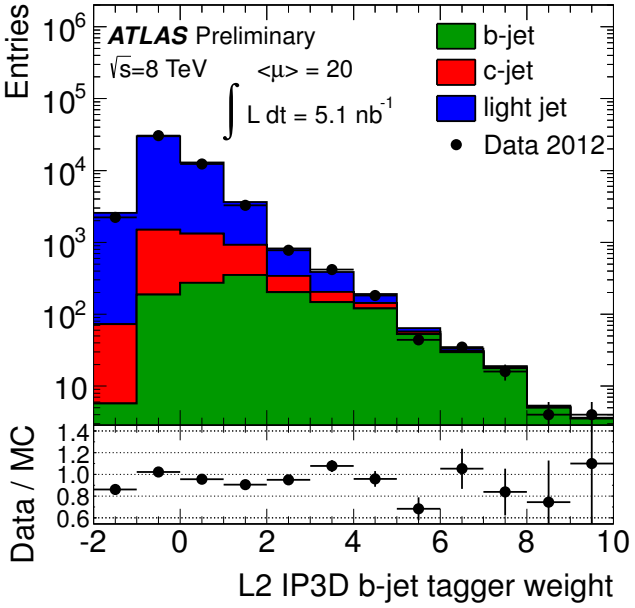


FIGURE 3.8 Jet weight distribution for the likelihood-ratio tagger based on the longitudinal and transverse impact parameter significance of prescaled Level 2 tracks in Level 2 jets with $p_T > 50$ GeV and $|\eta| < 2.5$ [89].

tracks defines the likelihood-ratio variable W:

$$W = \prod_{i=1}^{n} \frac{s(par_i)}{b(par_i)} \tag{3.3}$$

where $s(par_i)$ and $b(par_i)$ are the probability density function for the signal (b-jets) and the background (light-jets), both are parametrized as a function of track information, such as the transverse impact parameter. The $s(par_i)$ function can be based on different variables:

- the longitudinal impact parameter significance, namely IP1D,
- the transverse impact parameter significance, namely IP2D,
- the two-dimensional combination of longitudinal and transverse impact parameters, namely IP3D.

W can take any value between 0 (background-like) and $+\infty$ (signal-like). Furthermore to be sure to deal with a variable defined on a finite range, W is replaced by:

$$X = \frac{W}{W+1} \tag{3.4}$$

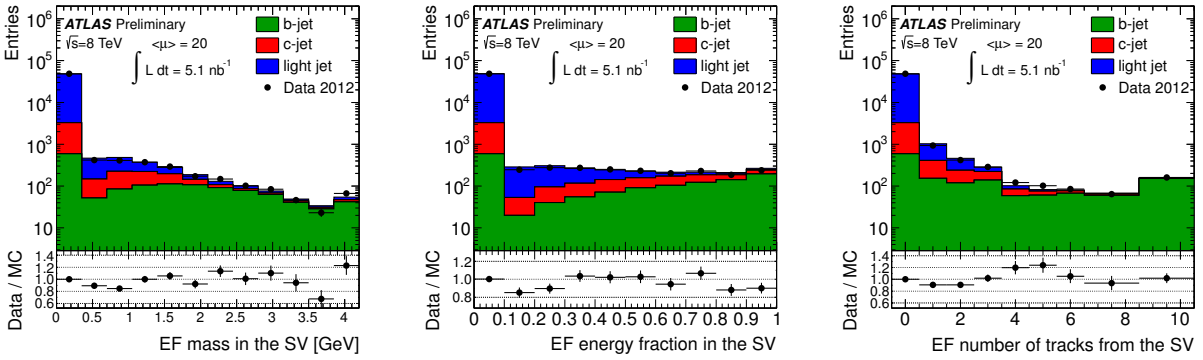
where X ranges are between 0 (background-like) and 1 (signal-like).
The IP3D algorithm, which combines more information gives the best performance. The data to MC simulation comparison, performed on pp collision recorded in 2012 at $\sqrt{s} = 8$ TeV, for the IP3D variable is shown in Figure 3.8. The three algorithms (IP3D, IP2D and IP3D) are implemented and available for online b-tagging at LVL2 and EF.

3.3.2.5 The SV1 algorithm

Using together three of the most discriminant secondary vertex properties: the invariant mass of all quality tracks associated to the secondary vertex, Figure 3.9(a), the ratio of the sum of the energies of the quality tracks in the secondary vertex to the sum of the energies of all tracks in the jet, Figure 3.9(b), and the number of tracks from secondary vertex, Figure 3.9(c), a new discriminant tagger is built, called SV1. To identify the quality tracks several requirement are applied:

- η and ϕ matching between the RoI and the tracks: $\Delta\phi < 0.2$ and $\Delta\eta < 0.2$ at LVL2 and EF,
- transverse momentum of the tracks above 1 GeV at LVL2 and EF,
- sufficient number of hits in the inner detector:
  - at least 1 hit in the b-layer at LVL2 and EF,
  - at least 4(2) hits in SCT at LVL2 (EF),
  - at least 7 hits in pixel detector at EF,
- transverse impact parameter $d_0 < 1$ mm at LVL2 and EF,
- longitudinal impact parameter $z_0 * \sin\theta < 2$ mm at LVL2 and EF,
- check on the track fitting: $\chi^2 \geq 0.001$ at LVL2.

The variables built from the tracks are combined using likelihood-ratio technique. SV1 relies on a 2D-distribution of the two first variables and a 1D-distribution of the number of two-track vertices. Additionally the distance ΔR between the jet axis and the line joining the primary vertex to the secondary one is used.



(a) (b) (c)

FIGURE 3.9 Data to MC simulation comparison for the three variables used in the SV1 definition: 3.9(a) invariant mass of quality tracks associated with the Event Filter jets' secondary vertex, 3.9(b) ratio between the energy sum of quality tracks associated with the Event Filter jets' secondary vertex and the energy sum of all quality tracks in the jet, 3.9(c) number of tracks from the secondary vertex. All the quantities are built for Event Filter jets with $p_T > 55$ GeV and $|\eta| < 2.5$. Only statistical errors are shown [89]. $< \mu >= 20$ gives the average of the number of interactions per bunch crossing.

3.3.2.6 The combined algorithm

Thanks to the likelihood ratio method used for IP3D and SV1, the algorithms can be easily combined: the weights of the individual tagging algorithms are the simply summed up. The combination of IP3D and SV1 is known as XComb tagger, shown in Figure 3.10. This tagger was chosen as default tagger during the 2012 ATLAS data-taking campaign.

Three working points (loose, medium and tight) have been implemented for the XComb tagger. They correspond to a b-jet efficiency (light-jet rejection at EF) computed on $t\bar{t}$ Monte Carlo sample of 40% (51), 50% (145) and 60% (350) respectively.

3.3.2.7 *b*-jet trigger rate

Since 2011 ATLAS has put in place various b-jet triggers, where the multi-jet requirements are combined with a b-tagging criteria. These trigger were used to collect data-sets for several channels with final states containing one or more b-jets, providing an improvement in the event efficiency with respect to the multi-jet triggers.

Figure 3.11 shows the rejection that can be obtained on the trigger rate thanks to the implementation of the b-tagging criteria, in particular this topology requires at least four jets at all trigger levels with transverse momentum larger than 10, 25 and 30 GeV respectively at LVL1, LVL2 and EF and the presence of at least two jets identified as a b-jets for a cut on the b-tagging weight such that the b-tagging efficiency is of 70% (estimated on $t\bar{t}$ Monte Carlo simulation). The trigger rate is shown for a run with a luminosity peak of $2 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$. In particular this trigger was the

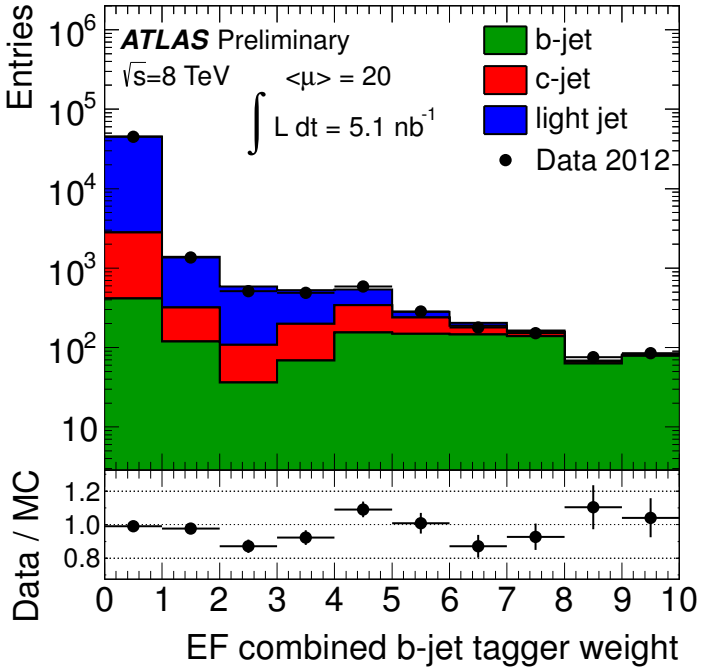


FIGURE 3.10 Jet weight distribution for the tagger based on the combination of the impact parameter significance and the secondary vertex likelihood-based taggers, derived from Event Filter tracks in Event Filter jets with $p_T > 55$ GeV and $|\eta| < 2.5$. Only statistical errors are shown [89]. $< \mu >= 20$ gives the average of the number of interactions per bunch crossing.

design trigger for the $t\bar{t}$ fully hadronic cross section measurement and was used for the top pair cross section analysis in the tau+jet channel that has been performed on an integrated luminosity of $L = 1.67$ fb^{-1} [90].

3.3.3 μ-in-jet trigger

A particular class of trigger chains are implement in the ATLAS trigger menu, these are a combination of two objects: muons and jets. The *mu*-in-jet trigger aim is to collect muons close to jets. This class of triggers is used to select a sample for online and offline b-tagging algorithm calibration. The motivation of performing the b-tagging calibration on a sample collected by a μ-in-jet trigger is that it is possible to select a sample enriched in b-jets. These, in fact, can be identified thanks to the presence of a muon coming from the semi-leptonic decay of b-hadrons. The production mechanisms of muons in b-, c-, and light-jets are different. Muons in b-jets are produced through b-hadron semi-leptonic decay ($B(b \to \mu\nu X) \approx 11\%$) [7] and the sequential semi-leptonic decay ($B(b \to c \to \mu\nu X) \approx 10\%$). On the other hand, muons in c-jets are produced via $c \to \mu\nu X$ with a branching ratio $\approx 10\%$ [7]. The sources of light-flavor jets with muons are several: muons from the decay of light hadrons (mainly pions $\pi^{\pm}$), hadrons reaching the muon spectrometer (punch-through), and mis-identified tracks (fake tracks) in the muon spectrometer.
The common technique used for the b-tagging calibration is based on p_T^{rel} [92]. Muon p_T^{rel} is defined

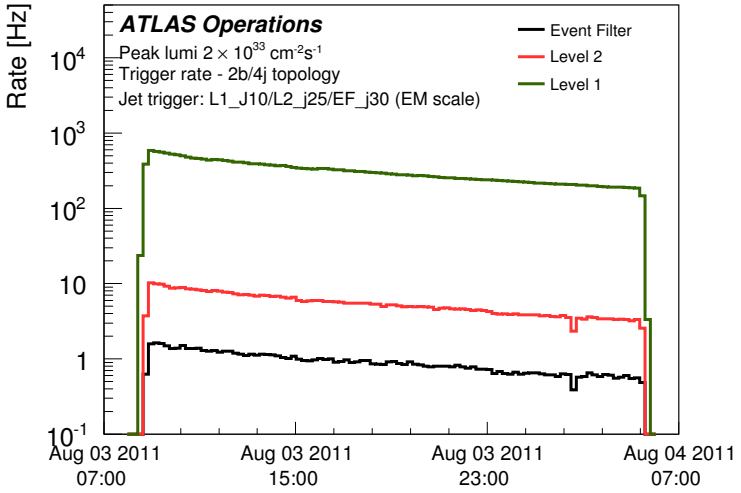


FIGURE 3.11 Trigger rate for $2b/4$j topology. LVL1, LVL2 and EF rate of a b-jet trigger requiring at least four jets in the event and at least two b-tagged jet. The jet thresholds correspond to 10, 25 and 30 GeV at LVL1, LVL2 and EF respectively with energies measured at the electromagnetic scale. The b-jet requirement is applied at LVL2 and EF and is tuned to give 70% efficiency on a b-tagged jet sample using $t\bar{t}$ MC simulation [91].

as the momentum of the muon transverse to the combined muon plus jet axis. Due to the high mass of b-hadrons, muons from direct b-quark decays are more boosted in the transverse plane of the jet+μ axis than those originating from other sources. They are characterized by harder p_T^{rel} spectrum than muons in c- and light-jets. This property can be used to separate b-jets from c- and light flavour jets. In order to perform the b-tagging calibration in different jet p_T bins, it is necessary to have a good statistic in all the p_T spectrum. So in order to obtain a flat distribution of the offline jet p_T, several μ-in-jet chains with several p_T threshold are implemented in the trigger menu. Table 3.1 and Table 3.2 present a list of the trigger used during the 2011 and 2012 data-taking period respectively. At LVL1 the muon and the jet objects are identified by the corresponding algorithm. The HLT chains are seeded by a LVL1 which ask for the presence of a muon with a low p_T and different p_T thresholds for the jets. Then a geometrical matching between a muon and a jet candidates is applied $\Delta R(\text{jet}, \mu) < 0.4$ at LVL2 and EF in order to build the μ-in-jet trigger (matched). In the 2012 the matching between the muon and jet is performed also in the z $\Delta z(\text{jet}, \mu) < 2$ mm (matchedZ).

The jet trigger item name consists of severals parts. The common structure is (level)_(n)(jet type)(theshold)_(algorithm+radius)(input object)(calibration), where:

- level: L1, L2 or EF,
- n: jet multiplicity, only for multi-jets chains,
- jet type: j stands for the central and fj forward jet
- threshold: p_T cut in GeV,

- jet algorithm name: a stands for anti-k_T [93], c for cone [94] and ca for cambridge-aachen [95],
- radius used in the jet algorithm: 4 stands for $R = 0.4$,
- algorithm input: cc stands for cell cluster[5], tc for topo cluster[6], tt for topo tower, TT for Trigger Tower[7], JE for Jet Element [8], hi for Heavy Ion baseline subracted tower,
- jet calibration: had stand for hadronic scale, em for electromagnetic scale.

Then there are other labels that can added to the nominal name scheme in order to describe more complex triggers, also present in μ-in-jet chains. There for both object is indicated the type and the p_T threshold as well as the $\Delta R(\text{jet}, \mu)$ requirements (matched/matchedZ).

EF chain name	L2 chain name	L1 seed	Rate at EF [Hz]
EF_mu4_j10_a4_EFFS	L2_mu4_j10_a4_EFFS	L1_MU0	1.0
EF_mu4_j10_a4tc_EFFS	L2_mu4_j10_a4tc_EFFS	L1_MU0	1.0
EF_mu4_j10_a4tc_EFFS_matched	L2_mu4_j10_a4tc_EFFS_matched	L1_MU0	1.0
EF_mu4_L1J10_matched	L2_mu4_L1J10_matched	L1_MU0_J10	1.0
EF_mu4_L1J15_matched	L2_mu4_L1J15_matched	L1_MU0_J15	1.0
EF_mu4_L1J20_matched	L2_mu4_L1J20_matched	L1_MU0_J20	1.0
EF_mu4_L1J30_matched	L2_mu4_L1J30_matched	L1_MU0_J30	1.0
EF_mu4_L1J50_matched	L2_mu4_L1J50_matched	L1_MU0_J50	1.0
EF_mu4_L1J75_matched	L2_mu4_L1J75_matched	L1_MU0_J75	1.0
EF_mu4_j135_a4tc_EFFS_L1matched	L2_mu4_j95_L1matched	L1_MU0_J75	1.0
EF_mu4_j180_a4tc_EFFS_L1matched	L2_mu4_j95_L1matched	L1_MU0_J75	1.0

TABLE 3.1 List of μ-in-jet triggers used in the 2011 data-taking period. The three trigger level requirements are shown for each trigger item as well as the corresponding final rate at EF. The pre-scale factors were changed several times during the 2011 data-taking due to the increasing of the instantaneous luminosity. Some trigger items were activated only for short periods in the trigger menu. The EFFS label stays for Event Filter Full Scan, meaning that the third level of trigger can access to all the detector.

3.4 Trigger Monitoring

The ATLAS trigger and data acquisition systems [96] require reliable and efficient monitoring systems to ensure proper online operation, good performance, rapid recognition and localization of potential problems. Through the monitoring system it is possible to check during and after the proton-proton collisions whether the trigger algorithm configuration runs without problems and

[5]The Cell Cluster is based on summing cells within a fixed-size rectangular window.

[6]The topological algorithm starts with a seed cell and iteratively adds to the cluster the neighbor of a cell already in the cluster.

[7]The Trigger Tower are formed by analog-summing over electromagnetic/hadronic calorimetric cells with a granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$.

[8]The Jet Element are formed by by summing 2×2 Trigger Towers in $\eta \times \phi$.

EF chain name	L2 chain name	L1 seed	Rate at EF [Hz]
EF_mu4T_j15_a4tchad_matched	L2_mu4T	L1_MU4	1.0
EF_mu4T_j15_a4tchad_matchedZ	L2_mu4T	L1_MU4	1.0
EF_mu4T_j25_a4tchad_matched	L2_mu4T	L1_MU4	1.0
EF_mu4T_j25_a4tchad_matchedZ	L2_mu4T	L1_MU4	1.0
EF_mu4T_j35_a4tchad_matched	L2_mu4T	L1_MU4	1.0
EF_mu4T_j35_a4tchad_matchedZ	L2_mu4T	L1_MU4	1.0
EF_mu4T_j55_a4tchad_L2FS_matched	L2_mu4T_j15_a4TTem	L1_MU4_J15	1.0
EF_mu4T_j55_a4tchad_L2FS_matchedZ	L2_mu4T_j15_a4TTem	L1_MU4_J15	1.0
EF_mu4T_j55_a4tchad_matched	L2_mu4T_j50_c4cchad	L1_MU4_J15	1.0
EF_mu4T_j55_a4tchad_matchedZ	L2_mu4T_j50_c4cchad	L1_MU4_J15	1.0
EF_mu4T_j80_a4tchad_L2FS_matched	L2_mu4T_j30_a4TTem	L1_MU4_J30	1.0
EF_mu4T_j80_a4tchad_matched	L2_mu4T_j75_c4cchad	L1_MU4_J30	1.0
EF_mu4T_j110_a4tchad_L2FS_matched	L2_mu4T_j50_a4TTem	L1_MU4_J50	0.25
EF_mu4T_j110_a4tchad_matched	L2_mu4T_j105_c4cchad	L1_MU4_J50	0.25
EF_mu4T_j145_a4tchad_L2FS_matched	L2_mu4T_j75_a4TTem	L1_MU4_J75	0.25
EF_mu4T_j145_a4tchad_matched	L2_mu4T_j140_c4cchad	L1_MU4_J75	0.25
EF_mu4T_j180_a4tchad_L2FS_matched	L2_mu4T_j75_a4TTem	L1_MU4_J75	0.25
EF_mu4T_j180_a4tchad_matched	L2_mu4T_j165_c4cchad	L1_MU4_J75	0.25
EF_mu4T_j220_a4tchad_L2FS_matched	L2_mu4T_j75_a4TTem	L1_MU4_J75	0.25
EF_mu4T_j220_a4tchad_matched	L2_mu4T_j165_c4cchad	L1_MU4_J75	0.25

TABLE 3.2 List of *mu*-in-jet triggers used in the 2012 data-taking period. The three trigger level requirements are shown for each trigger item as well as the corresponding final rate at EF. The label matched is used for indicate the $\Delta R(\text{jet}, \mu)$ maching, while the label matchedZ the $\Delta z(\text{jet}, \mu)$ one. L2FS indicated that the LVL2 make use of the full detector. The L2FS label stays for Level 2 Full Scan, meaning that the second level of trigger is running in the the full scan mode.

to find the presence of any processing failures. The ATLAS trigger monitoring system sits within the framework of the global ATLAS Data Quality Monitoring (DQM) system. It was developed, implemented and successfully tested by a group of ATLAS institutes (among which the CPPM). The Trigger Steering code checks the trigger decision through more than 2500 histograms which monitor detailed aspects of the trigger computation. The trigger response is checked on data, and the quality of the trigger configuration algorithm is provided by different flags which reflect the behavior of each trigger item controlled from the online and offline monitoring information. Five possible Data Quality (DQ) flags can be assigned to each sub-detector and trigger item. In the DQM display, the flags are indicated by five colors: good (green), flawed (yellow), bad (red), undefined (gray) and off (black). The online monitoring is assisted by an offline one which runs after the event reconstruction in the Tier-0. The offline monitoring has a very important role in the validation of the online decision. It usually provides further information on the selected data and it can help clarifying the undefined flags in the online monitoring, check and confirm with high accuracy the online decision. The DQ assessment are based on analysis of specific histograms created and filled during data taking in the HLT trigger algorithms (online) or within the standard raw data reconstruction (offline). The online DQM Framework (DQMF) displays histograms of trigger objects in real time within a dedicated system to the ATLAS shift crew. The offline DQMF

framework shows the histograms on a web page for a visual inspection after running over the express stream[9] and is complemented by further monitoring performed during the bulk data processing approximately 48 hours late.

3.4.1 Data Quality Monitoring online

The data quality monitoring of the ATLAS trigger, part of the general ATLAS DQM, is based on a common software infrastructure. Through the trigger DQM, it is possible to check the properties of the trigger objects which are related to physics candidates and utilized in the events selection by the trigger items, such as muons, electrons, photons, taus, jets, b-jets candidates and missing energy. A specific trigger signature group is associated to trigger chains based on different objects for cosmic muons, minimum bias, jet/τ/b/E_T^{miss}, B-physics. The trigger signature groups are characterized by a data quality flag which is stored into a database. Automated evaluations of selected histograms based on pre-defined tests run via the DQMF software. The implemented tests can compare several distribution parameters with respect to reference histograms, such as minimum statistic, width and mean of the distributions or the Kolmogorov-Smirnov discriminant of these parameters [97]. A highly configurable Data Quality Monitoring Display (DQMD) shows the DQ results obtained with DQMF, focusing the attention of the shifter on the histograms failing the automatic checks. A hierarchically structure of the information which starts from the various ATLAS subdetectors and goes down to the checking level on each histogram, is implemented to provide a complete overview of the data-taking process. Every single histogram check is performed with its own threshold and reference computing a result in terms of "good", "uncertain" or "bad". These DQ results are successively grouped into DQ regions yielding the final DQ result in the corresponding DQ flag.

3.4.1.1 DQMD for the b-jet signature

Since the b-jet signature depends on the inner detector tracking system, so some variables checked in the b-jet signature are related entirely to the HLT tracking. The b-jet signature need to control variables in the following areas:

- track and vertex related variables, such as the number of tracks per RoI, the track parameters and the primary vertex along z direction;
- variables used by b-tagging algorithms, such as the transverse and the longitudinal impact parameter;

[9] The 10% of the data collected during a run of the data-taking are immediately processed and stored in the so called express stream sample. This stream is used in the DQ to check the data and validate the reprocessing which correspond to reconstruction procedure applied on the raw data.

- weight used for event selection, such as the JetProb, the IP3D and the combined tagger (XComb).

The quantities under scrutiny by the DQMD are produced by different algorithms. The *b*-jet DQMD monitoring is composed by two HLT main folders for the two trigger level within those the histograms are grouped according to the algorithm which produces them. The interface for the evaluation of the DQ flags is shown in Figure 3.12.

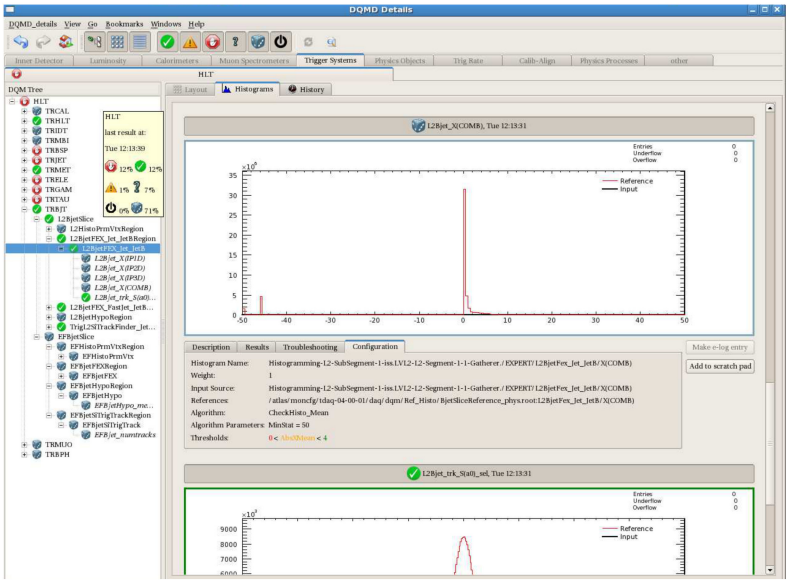


FIGURE 3.12 Display of the online data quality monitoring framework (DQMD) in charge of surveying the trigger objects quality. Individual signature groups can be selected in the panel on the left where these are also displayed using a colour legend showing the results of the check test applied on each signature. The trigger DQ flags described in the text are listed on the left panel. This example shows the the distribution of the XComb weight tagger at LVL2. The distribution of the selected data run (black line) is compared with the reference histogram (red line). In the bottom of the figure it is possible to access at more information on the pre-defined check test applied on the histogram. In this case two test are done: the check on the minimum statistics and the check on the mean.

3.4.2 Online Histogram Presenter

The Online Histogram Presenter (OHP) is a diagnostic tools used by the shift crew to survey the trigger behavior in the ATLAS control room during data taking. This application uses Qt [98] and ROOT framework and is configured via a XML [99] file. The basic concept of this monitoring tool is to provide one single panel per signature group allowing a fast check. In fact a small number of representative histograms (less than ten) per trigger signature group are presented in a tree structure or in configurable pre-defined windows. In OHP, reference histograms, made with previous data, are displayed and superimposed to the analyzed histograms in question. The OHP works with a mixed pull/push mode that interacts with the Online Histogram Service (OHS)

servers: notifications are transfered from the OHS server to OHP every time an histogram is updated. Only when an histogram is actually displayed, the histogram object is retrieved from the OHS server and updated in OHP. The OHP monitoring for the *b*-jet signature has three main folders, one for each trigger level: LVL2, L1.5 and EF. In LVL2 and EF folders four histograms are available: the number of reconstructed tracks in the jet RoI, the distribution of the transverse impact parameter of the selected tracks, the default *b*-tagger weight implemented in the trigger algorithm, a count of accepted and rejected events by the trigger criteria. In the L1.5 folder two distributions are shown: the transverse impact parameter of the selected tracks and the default *b*-tagger weight implemented in the trigger algorithm. An example of the setup provided for the ATLAS control room is given in Figure 3.13.

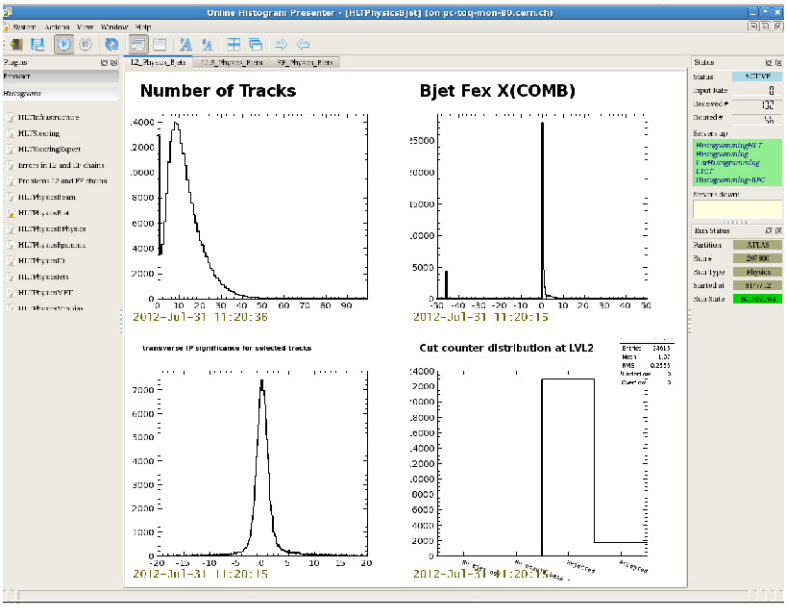


FIGURE 3.13 Display of the Online Histogram Presenter for the survey of the *b*-jet trigger in the ATLAS control room by the shift crew. Individual signature groups can be selected in the panel on the left. The corresponding pre-defined histograms are displayed in the central region. The example shows the *b*-jet signature histograms at LVL2. At the top of the central region it is possible to switch folders to see the histograms for the L1.5 level and Event Filter.

3.4.3 Offline DQMF framework

In order to understand and validate the trigger DQ flags, an offline cross-check of the basic variables used for their online determination is necessary. With a maximum delay of 36 hours after collection, the data are reconstructed and basic monitoring information is produced at Tier-0. This offline monitoring (called Tier-0 monitoring) is based on histograms containing physics distributions for the trigger objects. The offline DQ looks at the same trigger objects as the online one allowing to confirm or resolve problems thanks to the further information available after the Tier-0 reconstruction. The trigger monitoring information are accessible in the ATLAS control room but

also outside on a web-based tool, the Web Monitoring Interface (WMI) in which are shown all relevant trigger information after a run has been taken.

In the offline monitoring of the b-jet signature, two classes of primary triggers are studied and checked: b-jet and μ-jet triggers.The two trigger chains used in the monitoring for the b-jet trigger are:

- EF_b55_NoCut_j50_a4tchad: single jet trigger which requires at least one jet at L1 with $p_T > 15$ GeV, $p_T > 50$ GeV at LVL2 and $p_T > 55$ GeV at EF. The NoCut label indicates that the tracking and b-tagging algorithms are running but no decision is taken from them. The a4tchad denotes that the jet reconstruction was done with a full-scan from topological clusters using the anti-k_T reconstruction algorithm with a radius $R = 0.4$ at electromagnetic and jet energy scale, see Section 4.4.1, (EM+JES) scale,

- L2_b50_NoCut_j50_c4cchad: single jet trigger which requires at least one jet at L1 with $p_T > 15$ GeV, $p_T > 50$ GeV at LVL2. The b-tagging algorithm run at LVL2 but it does not take a decision on the event (NoCut). The c4cchad stands for the LVL2 triggers denotes jets, which are reconstructed from RoI-based calorimeter cells using a cone algorithm with a radius of $R = 0.4$ at electromagnetic and jet energy scale, see Section 4.4.1, (EM+JES) scale.

For both trigger chains all the implemented tagger weights, such as JetProb and XComb, are shown at LVL2 and EF.
Figure 3.14 shows the JetProb distribution at LVL2 for a selected data run. The peak at ~ 0 is an artifact due to RoIs with no selected tracks, while the peak at 1 reflects the displaced tracks.

The default tagger used in the 2011 data taking is shown in Figure 3.15. The nominal distribution for the XComb weight (combination of IP3D and SV1, see section 3.3.2.6) accounts for several pathological cases that may occur during data taking:

- no reconstructed tracks, hence no secondary vertex found in the event,
- tracks have been found, but no secondary vertex is associated to them,
- negative input to a logarithmic weight,
- wrong initialization value.

In order to improve the quality of the monitoring and focus the attention on the critical region of the tagger weight, it was decided to decouple the pathological cases from the core distribution and to have separate histograms. Figure 3.15(a) shows the distribution of the XComb tagger weight in the core region. In the counter histogram, Figure 3.15(b), if no reconstructed tracks and no secondary vertex have been found an entry is added at 0, if a track and associated secondary vertex have been found an entry is added at 1. In all the other cases the bin at -1 is filled.

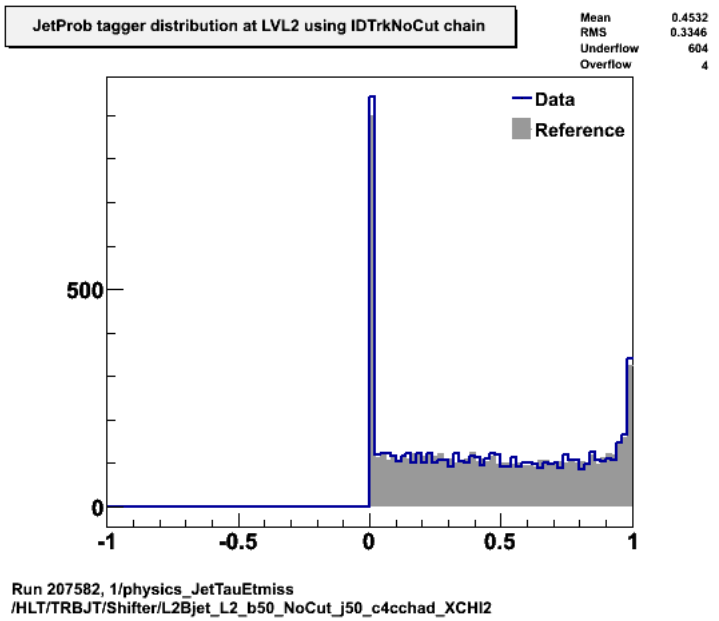
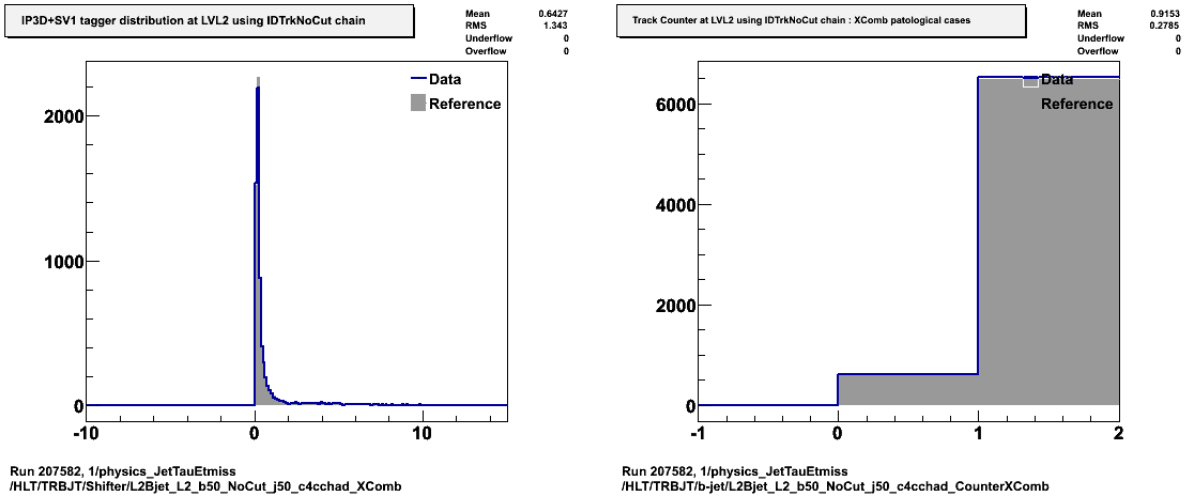


FIGURE 3.14 Histograms taken from the web display of Tier-0 offline monitoring. JetProb tagger probability distribution for events passing LVL2 trigger taken from the web display of Tier-0 offline monitoring. The blue line represents the selected data run to be check, whereas the fill histogram represents the reference, made by a previous data run. The IDTrackNoCutlabel means that the *b*-tagging algorithm runs in the trigger chain but It does not take a decision.



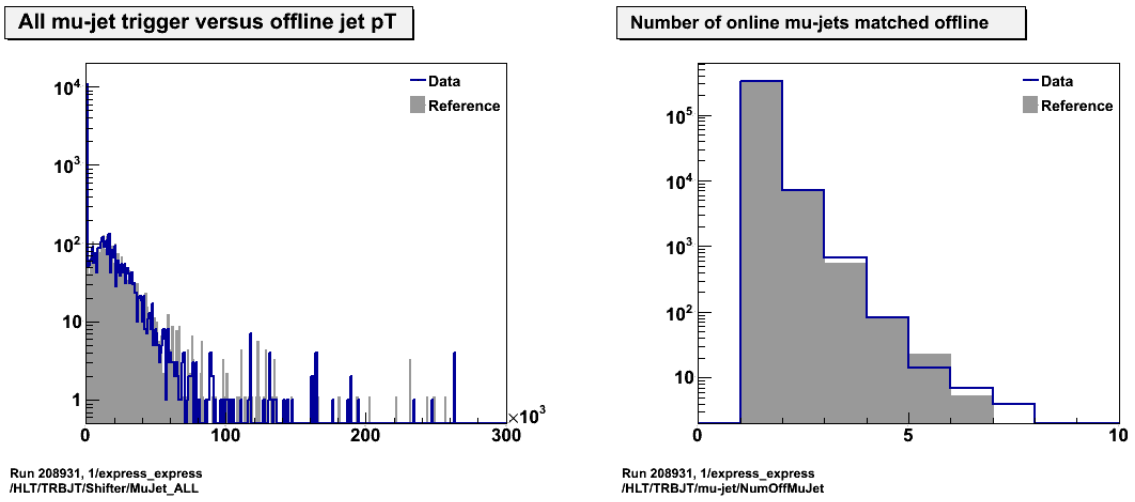
(a) (b)

FIGURE 3.15 Histograms taken from the web display of Tier-0 offline monitoring. 3.15(a) shows the core distribution of the XComb tagger weight for all the events passing the LVL2 trigger criteria. 3.15(b) shows the counter histogram for the pathological cases. The blue line represents the selected data run to be check, whereas the filled histogram represents the reference.

3.4.3.1 μ-jet trigger

Due to the fact that the data stream where the monitoring of *b*-jet signature is run, is dominated by light flavour jets, distributions more sensitive to *b*-tagging efficiency have been analyzed in events passing μ-jet trigger chain(s). In the offline monitoring some basic distributions, like the offline

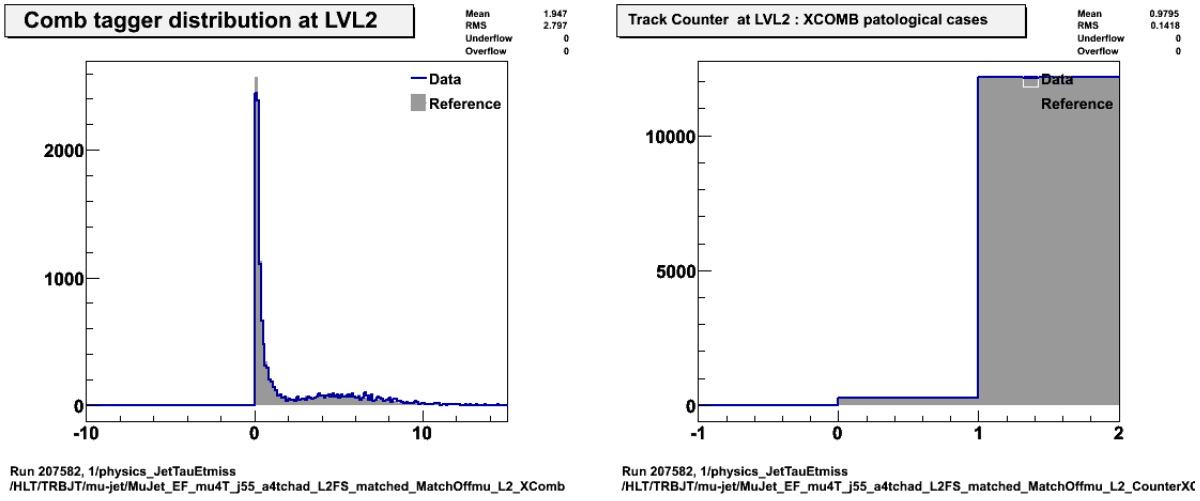
jet p_T spectra (Figure 3.16(a)), μ-jet multiplicity (Figure 3.16(b)) are displayed for different μ-jet trigger chains (see Table 3.2). Additionally, the main b-tagger weights are shown at LVL2 and



(a) (b)

FIGURE 3.16 Histograms taken from the web display of Tier-0 offline monitoring. 3.16(a) shows offline jet transverse momentum in MeV where one of the μ-jet trigger is fired. 3.16(b) represents the number of muons that are matched to an offline jet. The blue line represents the selected data run to be checked, whereas the filled histogram represents the reference

EF, for a full scan trigger and for a RoI based trigger: EF_mu4T_j55_a4tchad_L2FS_matched and EF_mu4T_j55_a4tchad_matched. Figure 3.17(a) and Figure 3.17(b) show the core distribution of XComb tagger and the counter for the pathological cases for events passing a trigger request of at least one jet with $p_T > 55$ GeV matched with a μ-jet, respectively. In order to further enhance the



(a) (b)

FIGURE 3.17 Histograms taken from the web display of Tier-0 offline monitoring. Figure 3.17(a) is core distribution of the XComb tagger weight for all the events passing the μ-jet trigger criteria. Figure 3.17(b) represents ounter histogram for the pathological cases. The blue line represents the selected data run to be checked, whereas the filled histogram represents the reference.

content of the b-jets in the selected sample, a study on p_T^{rel} of the muon jet was made following the selection used in the b-tagging calibration [92] (see Section 3.3.3). This study was performed using the following events selection:

- events passing a μ-jet trigger,
- selection of offline jets with $p_T > 15$ GeV and $|\eta| < 2.5$,
- selection of offline jets matched to a muon within an angle $\Delta R < 0.4$,
- basic quality cuts on the muons, like cut on transverse momentum, pseodorapidity, transverse impact parameter (for more details see [92]).

The p_T^{rel} spectra obtained after the events selection is shown in Figure 3.18. The choice of the p_T^{rel}

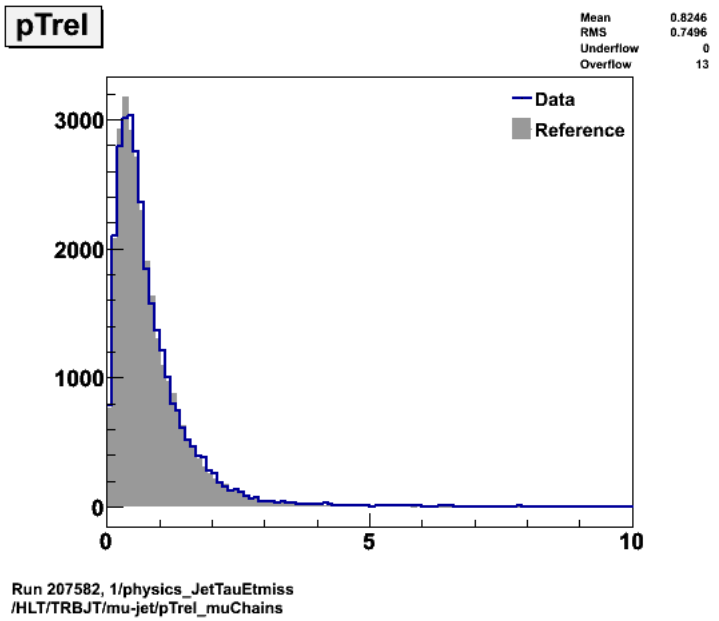


FIGURE 3.18 Histograms taken from the web display of Tier-0 offline monitoring. p_T^{rel} spectrum for events passing μ-jet trigger chain. The measurement unit of the x-axis is GeV. The blue line represents the selected data run to be check, whereas the filled histogram line represents the reference.

cut was made taking into account the fraction of the b-jets and also the statistics left after requiring specific p_T^{rel} threshold. The aim was to achieve a good balance between these two contributions. Figure 3.19 show the XComb tagger weight for events passing a μ-jet trigger and a cut on p_T^{rel} larger than 2 GeV. The impact of the μ-jet trigger as well as of the p_T^{rel} requirement on the data sample composition is clear in the XComb distribution. By comparing Figure 3.15(a), Figure 3.17(a) and Figure 3.19, we can notice that the mean value of the XComb distribution increases proportionally to the heavy flavour composition, when adding a muon to be in the jet radius and then when cutting on the p_T^{rel}.

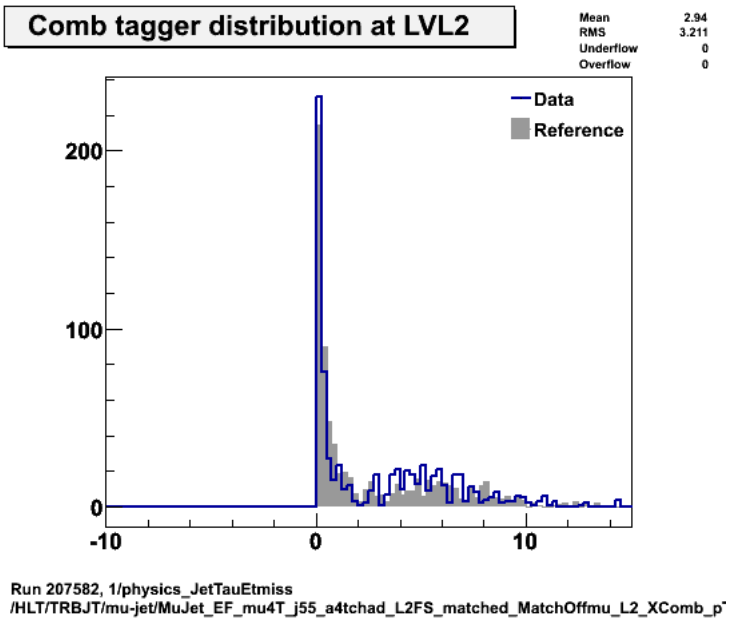


FIGURE 3.19 Histograms taken from the web display of Tier-0 offline monitoring. Core distribution of the XComb tagger weight for all the events passing a μ-jet trigger criteria and a cut on $p_T^{rel} > 2$ GeV. The blue line represents the selected data run to be check, whereas the filled histogram represents the reference.

3.5 Conclusion

The trigger system during the last two years demonstrated to function well, satisfying operational requirements and evolving to answer the demands of increasing LHC luminosity. The complex trigger system selects interesting physics events by identifying signatures of muon, electron, photon, tau lepton, jet, b-jet and B meson candidates and missing transverse energy. An overview of the ATLAS trigger system has been presented on this chapter, pointing out the implementation of the b-tagging at trigger level. The ability to separate heavy flavour jets from light-quark is an important ingredient for many physics analyses, such as measurements in the top-quark sector and searches for Higgs bosons or other new physics signatures and gives access to signals that would be hidden by multi-jet background, such as $t\bar{t}$ in the fully hadronic decaying mode. Two different categories are available in ATLAS: the first that exploits the b-hadron proprieties, b-jet trigger, and the second that looks for the presence of a muon in the b-hadron decays. Both trigger are monitoring during the data-taking period in oder to survey the trigger algorithm behavior. I collaborated in the development and maintenance of the online and offline monitoring of the b-jet trigger signature. Thanks to the monitoring system relevant problems on the trigger algorithm configuration were figured out, providing the change to quickly fix the configuration and collecting data with b-jet trigger. For the future data-taking campaigns, the trigger monitoring will evolve, like for example with the introduction of efficiency and fake rate distributions, to provide more accurate checks on the algorithms.

Chapter 4

Measurement of the $t\bar{t}$ production cross section in the fully hadronic final state

This Chapter presentes the $t\bar{t}$ production cross section measurement in the fully hadronic channel. The analysis is performed on a dataset consisting of an integrated luminosity of 4.70 fb^{-1} of proton-proton collisions produced in 2011 at the LHC with a centre of mass energy of $\sqrt{s} = 7$ TeV and recorded with the ATLAS detector [100].

4.1 The $t\bar{t}$ fully hadronic channel topology

The fully hadronic $t\bar{t}$ events, represented by the Feynman diagram shown in Figure 4.1, are about 44% of the total $t\bar{t}$ production. The topology of this channel consists of six or more jets with relatively high transverse jet-momenta. Two of these jets originate from b-quarks, while the remaining four jets originate from the decay of the two W bosons. The initial and final state radiation, as well as, the overlapping of additional events from soft interactions (pile-up) can lead to additional jet activity in the event. On the other hand, due to the detector acceptance some jets can be missed or mis-identified leading to a lower jet multiplicity in the final state. In general the purity of the signal signature may be improved by the identification of jets originating from b-quarks (b-jet identification see Section 4.4.3). In general the $t\bar{t}$ signal in the fully hadronic mode suffers from a huge background due mainly to the QCD multi-jet background events with the same experimental signature. A very strict event selection can help in decreasing this important background contribution, however it is not sufficient to eliminate it completely. Therefore its modeling is studied very carefully using data-driven methods. Also $t\bar{t}$ signal events are carefully studied with Monte Carlo

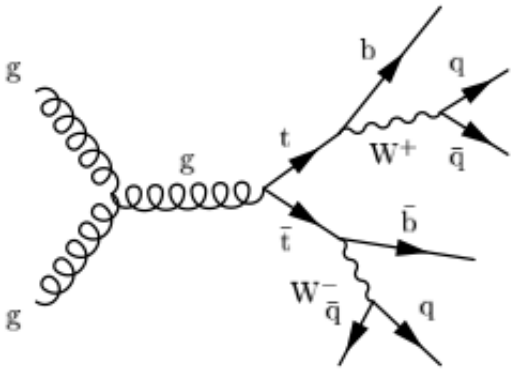


FIGURE 4.1 Feynman diagram for the fully hadronic $t\bar{t}$ decay.

simulation in order to better understand its kinematics and derive signal template to discriminate $t\bar{t}$ events with respect to the background.

4.2 Data and Monte Carlo datasets

The data used for this analysis was recorded by ATLAS during the 2011 data taking period with a stable beam and solenoidal field in nominal configuration. A data quality requirement was applied to ensure that all the subdetectors were fully operational. The dataset used corresponds to an integrated luminosity of 4.7 fb^{-1}.

The modelling of $t\bar{t}$ signal events is derived using Monte Carlo generated events. For our studies of the $t\bar{t}$ signal, the MC@NLO v3.41 [101] generator with PDF set CT10 [102, 103] was used to tune the selection criteria and to build a signal template, for a top quark mass of 172.5 GeV. The generated events were processed through the full ATLAS detector simulation based on GEANT4 [104] and subsequently reconstructed as real data events through the ATLAS event reconstruction chain. At $\sqrt{s} = 7$ TeV the cross section of $t\bar{t}$ production in the fully hadronic decay mode for the MC@NLO+Herwig[1] generators is $\sigma_{t\bar{t}}^{had} = 76.18$ pb. Due to the large uncertainty in the QCD multi-jet cross section prediction [106], the estimation of this important background is derived using a data-driven method. For the background modeling validation, MC QCD multi-jet samples generated by ALPGEN [107] were used, in particular the samples with multi-jet production and exclusive $b\bar{b}$ production. ALPGEN is a generator that is capable of calculating the matrix-element at leading order for a given hard process plus N additional partons. Events are simulated separately for every multiplicity of N partons. Samples with $N = 0$ to 4 additional partons are exclusively generated, the last bin with $N \geq 5$ is inclusive. The additional partons are generated for p_T value above a given threshold. Below the p_T threshold the partons stem from the parton shower simulated by Herwig [105], which is also used to simulate the showering of the partons generated by the matrix-element.

[1] Herwig [105] is used for the modeling the hadronization

4.3 Data and MC treatment at pre-analysis level

Before going through the specific event selection required for the $t\bar{t}$ production cross section measurement, it is important to explain the selection applied on the data sample in order to eliminate the non-collision background events and events collected during detector problems.
Considering that the objects present in signal events are reconstructed in all ATLAS subdetectors, only events recorded when the full detector was fully operational are considered in the analysis. A set of data-quality checks, based on the quality and performance of different subdetectors, is applied on the collected data. The results of the data-quality decision are stored in flags used to fill a list of good run and luminosity blocks, GoodRunLists (GRLs). The rejection of the events originating from cosmic rays or other sources of non-collision background is requested by applying a cut on the number of tracks related to the reconstructed primary vertex. If the primary vertex is reconstructed with less then four tracks the event is not considered and is rejected from the physical data-set.

4.3.1 Pile-up

Any hard proton-proton collision recorded by the ATLAS detector contains a superposition of particles coming from several soft proton-proton collisions. These non interesting soft collisions are called pile-up events. Their rate depend on the magnitude of the instantaneous luminosity at operation.
The number of interactions that occur during the beam crossing follows a Poisson distribution with an expected mean value $< \mu >$ of 23 interactions at the design luminosity of the LHC. However, due to the long tail of the Poisson distribution, an important fraction of the collisions will have a higher number of interactions. MC simulation events are generated with constant beam condition. This is not the case for real data as the beam condition vary with the time. As consequence the amount of pile-up events is not constant. Therefore a re-weighting of the simulated events, based on the exact configuration of extra pile up events in the run periods, is implemented in order to account for the different level of pile-up.

4.4 Object identification and selection

In the $t\bar{t}$ final state various objects can be identified, depending on the decay mode under investigation, such as light jets, b-jets, leptons (electron and muon) and neutrinos (presence of missing transverse energy E_T^{miss}). While jets and b-jets are used to identify the fully hadronic channel, leptons and E_T^{miss} are vetoed to suppress contamination from other $t\bar{t}$ final states and from other SM backgrounds. A good object identification in the final state has a crucial role in the selection

of signal events. In the following Sections the main characteristics of ATLAS objects identification are described in details.

4.4.1 Jets

The jets are the manifestation of scattered sub-nuclear partons. A jet is reconstructed [93, 108] from the energy deposition in the calorimeter cells. Due to the high jet multiplicity in the $t\bar{t}$ all hadronic final state signature, the study and the understanding of jet reconstruction and kinematics is important. According to QCD, the scattered partons loose energy pulling out from the vacuum more partons (fragmentation process). Those then have to rearrange themselves in color singlet states (hadronization process) producing several hadrons set around the initial hard parton direction. Most of these hadrons are very short lived particles. They decay inside the detector and the decay products are absorbed and measured by the detector, in particular by the calorimeter system. Figure 4.2 shows a schematic view of the different phases of jets production from proton-proton scattering.

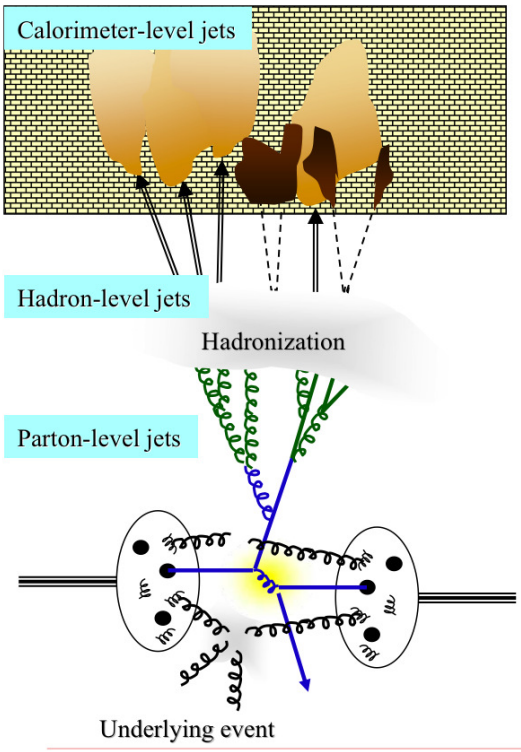


FIGURE 4.2 *pp* scattering in different phases. The initial pair of partons forms jets. The image shows the phase just before the hadronization (parton jets), after the hadronization (particle jets) and the calorimeter jets reconstructed from the particles energy deposition in the calorimeter.

Hence a jet is measured as a set of calorimeter clusters, which are a collections of geometrically closed calorimetric cells. To allocate the particles and energy depositions in the calorimeter to different jets a reconstruction algorithm is used. The jet clustering algorithm designed to be both infrared and collinear safe is the anti-k_T algorithm. This means that the jet reconstruction does

not depend on additional soft gluon radiation between jets and collinear splitting of the initial parton. The inputs to the jet reconstruction are topological clusters calibrated at the EM scale [80], which corresponds to the baseline calorimeter energy scale. The anti-k_T algorithm performs jet reconstruction through the following steps:

- defines for each entity i the quantity $d_i = P_{Ti}^2$,
- computes the distances d_{ij} between entities i and j:

$$d_{ij} = min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2}$$

  where $\Delta_{ij}^2 = \Delta(\eta_i - \eta_j)^2 + \Delta(\phi_i - \phi_j)^2$ is the standard setting for the reconstruction of cone jets. It is a parameter of the jet algorithm. The variables k_{ti}, η_i and ϕ_i are respectively the transverse ij momentum, pseudorapidity and azimuth angle of particle i,
- computes the distance of an entity i and the beam (B) $d_{iB} = k_{ti}^{-2}$,
- finds the minimum (d_{min}) among all the d_{iB} and d_{ij},
- if d_{min} is d_{ij}, recombine entities i and j in a new particle k with $p_k^\mu = p_i^\mu + p_j^\mu$. The new particle is put back into the list while removing i and j objects,
- if d_{min} is d_{iB}, the particle i is considered as a jet and it is removed from the list of entities,
- recalculates the distances and repeat the procedure until no entities are left.

Pre-clustering techniques are applied to reduce the number of input jet components, because the anti-k_T algorithm can be very expensive in term of execution time. The choice of the ATLAS collaboration is to consider a seeded cone algorithm with a transverse energy threshold of ${E_T}^{th} = 2$ GeV , and a radius of $R = 0.6$ or $R = 0.4$. A merging procedure of two overlapping jets is applied if they share more than 50% of the energy of the least energetic one. Once the jet reconstruction is performed, the jets are calibrated from the EM to the hadronic scale to take into account the energy losses due to the detector acceptance and non-active parts. The corrections are derived as a function of the pseudorapidity and transverse momentum from MC simulation. Further corrections such as those due to additional proton-proton interactions and those originating from displaced vertices are applied as well. The residual correction functions are derived using in situ data measurements, and MC simulation. The jet energy response at EM scale is shown in Figure 4.3 as a function of pseudorapidity, the different marker colors are used to distinguish the jet energy response behavior at various energy thresholds. The grey vertical lines indicate the ATLAS calorimeter zones. In this analysis only jets with $p_T > 20$ GeV and $|\eta| < 2.5$ are considered. Moreover the events in which at least one calibrated jet with positive energy and transverse momentum larger than 20 GeV identified as a "LooseBad" (see Section 4.5) by the data quality group are discarded. Three main sources of bad jet reconstruction can be found: jets stemming from cosmic rays or other non

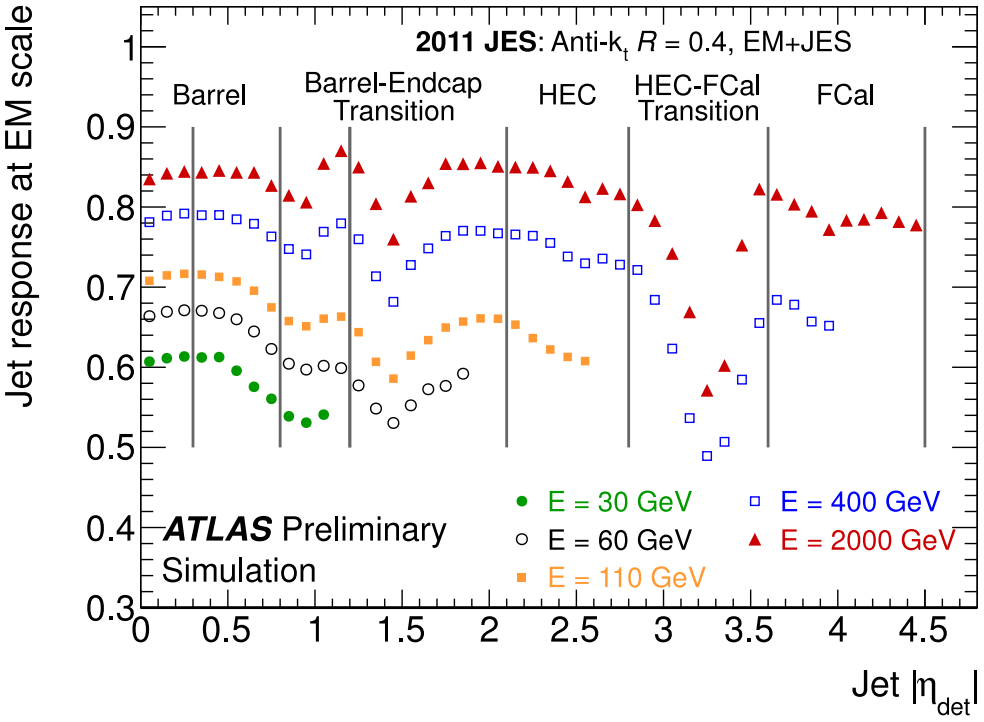


FIGURE 4.3 Average energy of jets formed from topological clusters calibrated at the EM scale with respect to the truth jet energy in Monte Carlo simulation (Ejet(EM)/Ejet(truth)) as a function of the jet pseudorapidity before applying the correction for the event vertex. Also indicated are the different calorimeter regions. The inverse of the response shown in each bin is equal to the average jet energy scale correction. This result is based on PYTHIA inclusive jet samples [80].

collision backgrounds, coherent noise in the electromagnetic calorimeter and energy spikes in the hadronic calorimeter.

4.4.1.1 Jet vertex fraction

To identify jets originating from the hard-scatter interaction a discriminating variable, called jet vertex fraction (JVF), is built by combining the information of the tracks and their primary vertex with the calorimeter jets. The JVF measures the probability that a jet comes from a particular vertex. To built this discriminant variable, the jets are matched to the tracks with a $\Delta R(jet, track) < 0.4$ where the track parameters are computed at the origin because the aim is to associate tracks to jets produced in the same primary vertex. The JVF is defined as the fraction of each jet's constituent transverse momentum in each vertex, in other words the sum of p_T of all matched tracks to a given vertex divided by the total jet matched track p_T for all vertices. Formally for a single jet jet_i the JVF with respect to the vertex vtx_j in the event is:

$$\mathrm{JVF}(jet_i, vtx_j) = \frac{\sum_k p_T(trk_k^{jet_i}, vtx_j)}{\sum_n \sum_l p_T(trk_l^{jet_i}, vtx_n)} \tag{4.1}$$

An ATLAS study performed on different MC simulation samples investigated the effect of different selections on the JVF value. When jets are required to have $|\text{JVF}| < 0.75$, the jet multiplicity distribution becomes flat with respect to the number of primary vertices. This results highlights the importance of the jet-vertex fraction algorithm in physics analyses: the JVF is applied in the fully hadronic analysis in order to be insensitive to the contributions of simultaneous uncorrelated soft collisions from the pile-up events.

4.4.2 Lepton and missing transverse momentum

The fully hadronic $t\bar{t}$ final state is characterized by at least six jets and does not contain any real missing transverse energy or isolated leptons. To reduce the presence of background due to events containing W boson decaying leptonically and the overlap with the other $t\bar{t}$ cross section measurement a veto against high-p_T isolated leptons is applied.

4.4.2.1 Electrons

The electron algorithm is seeded by energy clusters in the electromagnetic calorimeter. A sliding window algorithm is implemented in the calorimeter to form the cluster. The calorimeter position yielding the maximum energy deposition in a $\Delta\eta \times \Delta\phi$ cone is chosen as the cluster that will seed the electron reconstruction algorithm. The information on the track in the inner detector are used to discriminate between electrons and muons. The TRT information, instead, is important to distinguish electrons from hadrons. A set of cuts on different electron characteristics helps in the identification of a real electron with respect to a mis-identified one. The electrons candidates should have a transverse energy larger than 25 GeV and should be within pseudorapidity range of $|\eta| < 2.47$, excluding the pseudorapidity region between 1.37 and 1.52 which correspond to the barrel-endcap calorimeter transition region, crack region. These are required to pass an identification selection cuts [109], so called "tight++" which include cuts on variables related to the first calorimeter layer to reject the electrons originating from π^0 decay and on track quality. An isolation criteria is applied in order to suppress background candidates from hadrons identified as electron, electron from heavy flavour decay and photon conversions. The electron candidate should have a small jet activity in the space around its direction. The energy deposited in a cone of $\Delta R = 0.2$ has to be less than 3.5 GeV. Electron candidates are also likely to be reconstructed as jets. Electron and jet candidates close to each other are likely to have the same magnitude of transverse momentum. To remove this electron-jet overlap a jet is excluded if it is closer than $\Delta R < 0.2$ to the selected electron.

4.4.2.2 Muons

In ATLAS the muons are reconstructed in the muon spectrometer. They are also reconstructed as inner muon tracks in the inner detector and in the calorimeter as well. Hence the information from the different subsystems are combined in order to identify a muon candidate: reconstruction of the track direction and determination of its momentum. Two algorithms are available within the ATLAS Collaboration to perform muon identification: MuID and Staco [79]. The MuID algorithm forms a track in the muon spectrometer and looks for the associated track in the inner detector. The final muon track is provided by a global fit of tracks from the inner detector and muon system. The Staco algorithm looks for a track in the muon system and extrapolates it back to the inner detector toward the interaction region.
The muon reconstruction used in this dissertation is the latter one. A very strict definition is used for the muon considering that in top events a muon coming from the decay of a W boson is expected to be well identified. Selected muon candidates [109] are required to have a transverse momentum $p_T > 20$ GeV and a pseudorapidity $|\eta| < 2.5$. Additional criteria are on the quality of the association of an inner detector track with a muon candidate in the spectrometer. A muon originating from hard proton-proton scattering is expected to leave a track along all the inner detector systems; for this reason a hit in the innermost layer of the pixel detector is required, as well as at least one hit in the pixel detector, at least five hits in the SCT, and hits in the TRT. Moreover an isolation criteria is applied in order to separate a muon produced by hard interaction, so called prompt-muon, from a muon present inside a jet as product of hadron decays. A prompt muon is characterized by a clear signature in the detector, instead a non prompt one presents a significant amount of energy deposition around its trajectory. The amount of energy deposition is computed in a cone of a fixed radius around the muon track and it is measured in the calorimeter or in the inner detector. Two isolation criteria are required: calorimetric and track isolation. The calorimeter isolation requires that the energy deposition in the calorimeter within a cone of radius $R = 0.2$ is less than 4 GeV, excluding the energy deposition along the muon. The track isolation requires that the sum of the transverse momentum of the tracks in a cone of radius $R = 0.3$ is less then 2.5 GeV, excluding the transverse momentum of the muon tracks. The isolation criteria is required to select only muons with transverse energy in a cone of radium 0.3 around the muon direction less than 2.5 GeV and the transverse energy in a cone of $R = 0.2$ less of 4 GeV. Also the muons should be separated from the reconstructed jet with $\Delta R > 0.4$ to decrease the multi-jet background contribution.

4.4.2.3 Missing Transverse Energy

The presence of missing transverse energy is expected in the top pair final state in which at least one of two W bosons decay into a lepton and neutrino. It represents the energy associated with the neutrino which, due to its lower interaction with matter, escape from the detector without

leaving any signal. The missing transverse momentum E_T^{miss} is calculated from the topological clusters calibrated at EM scale with an additional object-dependent calibration applied to the cluster associated to higher level objects like electron or jets. The calculation of the missing energy include the clusters associated to all the physics objects [110].

4.4.3 Offline identification of *b*-jets

A main key in the top pair cross section measurement is the ability of identifying b-quarks coming from the $t\bar{t}$ quarks decay. The ATLAS collaboration developed, implemented, and calibrated various algorithms making use of offline tracking reconstruction with different performances. Most of the algorithms are based on the presence of a secondary vertex in the event. The existence of a displaced vertex from the primary vertex signs the presence of a long lived particles, like b-hadron, in the event. Also for the offline algorithm the main variables used are the secondary vertex position and the impact parameter of the tracks. All these informations can also be used in a multivariate analysis to improve the algorithm performance.
In this analysis the identification of jets originating from b-quarks is performed using a discriminant built from the combination of three offline b-tagging algorithms: JetFitter [2], IP3D (Section 3.3.2.4) and SV1 [112] (Section 3.3.2.5). As done online, these algorithms are combined using a likelihood-ratio technique to build a final tagging discriminant, so called MV1, used to perform b-tagging decisions.
The efficiency of the b-jet identification and as well the rejection of the light-jet and c-jets are measured in data and compared with MC simulation. Several working points, corresponding to a precise b-jet efficiency and light-jet rejection of the tagger, are chosen and then calibrated. The b-jet efficiency as a function of the light-jet rejection is show in Figure 4.4 for different offline tagging algorithms. The b-tagging efficiency studies are performed on a $t\bar{t}$ MC sample, considering jets with $p_T > 15$ GeV and $|\eta| < 2.5$ [92].

The cut used in this analysis on the MV1 weight accepts b-jets with approximately 60% efficiency and corresponds to a light-jet rejection factor of about 500 on simulated $t\bar{t}$ events [113]. Two methods are used to calibrate the b-tagging algorithms: $system8$ [114] and p_T^{rel} [92]. The efficiency of the algorithms are provided as a function of the jet transverse momentum and pseudorapidity. The results of both methods are in good agreement with the data. In order to improve the performance the two methods are combined. Scale factors which compensate the data to MC simulation disagreement are then calculated as:

$$SF_{WP}(flavor, p_T, \eta) = \frac{\mathcal{P}_{WP}^{data}(flavor, p_T, \eta)}{\mathcal{P}_{WP}^{MC}(flavor, p_T, \eta)} \tag{4.2}$$

[2] JetFitter [111] exploits the topology of weak b-and c-hadron decays inside the jet. A Kalman filter is used to find a common line on which the primary vertex and the b- and c-vertices lie, as well as their position on this line, giving an approximated flight path for the b-hadron.

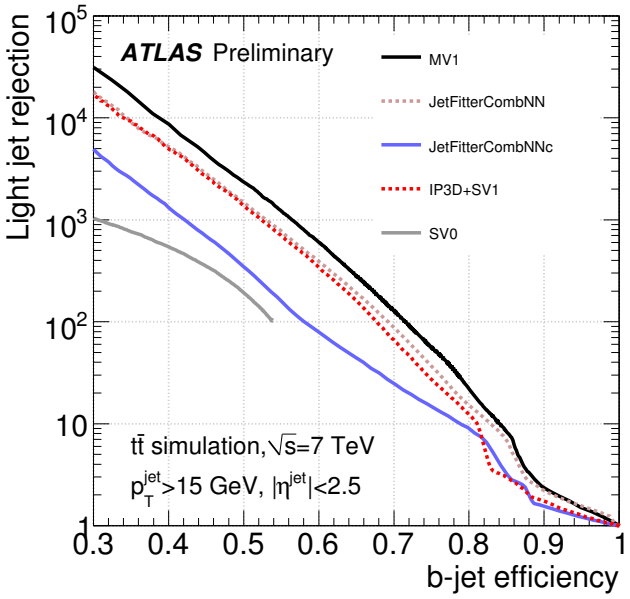


FIGURE 4.4 Light-jet rejection as a function of the b-tag efficiency for several offline b-tagging algorithms. The results are based on simulated $t\bar{t}$ events [92].

where P_{WP}^{data} and P_{WP}^{MC} are the probabilities as a function of the jet p_T, η and flavour for a particular working point in data and MC simulation respectively.

The scale factor for the MV1 working point used in this thesis is shown in Figure 4.5

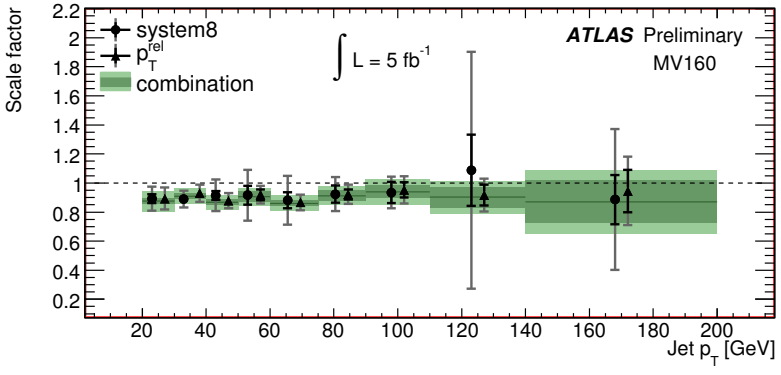


FIGURE 4.5 The data-to-simulation scale factor for the MV1 tagging algorithm at 60% efficiency as a function of the jet p_T, obtained by combining the p_T^{rel} and $system8$ results. The dark green band represents the statistical uncertainty of the combined scale factor while the light green band shows the total uncertainty.

After assigning to each jet a scale factor, a global weight is applied to the event defined as the product over all the scale factors of the jets.

4.5 Event Selection

Several studies were carried out [100] to perform an ad-hoc selection of the $t\bar{t}$ events in order to reduce the fraction of the background events and to obtain a pure $t\bar{t}$ sample.
After the data and MC treatment presented in Section 4.3, a veto on an isolated lepton in the event, defined in Section 4.4.2, is applied. In order to remove hardware problems in the calorimeter, fake jets or cosmic muons, a jet-cleaning criteria is applied by asking the jets to be non identified as "LooseBad". This label identifies the jets that are not associated to in-time real energy deposition in the calorimeter. The events in which at least one jet is identify as "LooseBad" are removed from the data sample used in this analysis.
The data analyzed has been collected by non prescaled multi-jet trigger which request the presence of at least five jets in a pseudorapidity range of $|\eta| < 3.2$ with a transverse energy larger than 10 GeV at LVL1, 25 GeV at LVL2 and 30 GeV at the Event Filter level. After applying a selection of events, such as the trigger one, the corresponding selection efficiency must be measured. This selection efficiency represents a conditional probability that one event passes the selection given the initial conditions. For example considering an initial sample A, the trigger requirement T and a variable x, the selection efficiency is given by $\epsilon_T(x, A) = P(T|A, x)$. A possible estimation of this probability is the measured success efficiency defined as the ratio between the number of events passing the trigger selection T in each bin i of variable x ($k_A(i)$) and the number of the events in the initial sample A ($n_A(i)$): $\epsilon_T(x, A) = P(T|A, x) \equiv \frac{k_A(i)}{n_A(i)}$. This ratio converges for a large initial sample to the selection efficiency thanks to the law of large numbers (Bernoulli's theorem). This method used to measure the selection efficiency is called Tag&Probe where the initial sample, the Tag, is used to test the selection, the Probe. When the number of the initial sample is not enough to compute the selection efficiency, as in the case of a multi-jet trigger, a different technique, called *bootstrapping*, is implemented. The *bootstrapping*[3] technique overcomes the problem of insufficient statistics in samples collected with unbiased triggers by using a biased event sample where the efficiency is known. In this technique the trigger under study is labelled Tag and the trigger used to collect the event sample is called Probe. In the case of a Tag N-jet trigger, the efficiency is

[3]The bootstrapping technique is based on a Bayes's theorem. Considering the context of the measurement of a trigger efficiency where the trigger under study is labelled as T, the Probe, and the trigger of reference is B, the Tag, (the efficiency of the Tag trigger is known) the trigger efficiency is given by :

$$P(T) = \frac{P(T|B) \cdot P(B)}{P(B|T)} \tag{4.3}$$

where $P(T)$ and $P(B)$ are the probabilities that an event passes the trigger requirement T and B respectively, the $P(T|B)$ represents the probability that the Probe trigger fires an event already fired by Tag one and $P(B|T)$ is the probability that an event passes the Tag trigger requirement after firing the Probe trigger. Generally the trigger under study and the reference one are chosen in order to give $P(B|T) = 1$. As a consequence it is possible to estimate the trigger efficiency $P(T)$ using the efficiency of the Tag trigger $P(B)$ and the biased efficiency of the Probe trigger T on events collected by the Tag trigger B:

$$P(T) = P(T|B) \cdot P(B). \tag{4.4}$$

estimated with respect to a Probe (N-1)-jet trigger as a function of the Nth reconstructed jet. In this analysis the Probe requires the presence of four jets with $p_T > 30$ GeV at EF level and the Tag is a five jets trigger. The reference trigger used is fully efficient with respect to the preselection ($P(4j30) = 1$). In the $t\bar{t}$ cross section analysis the trigger efficiency is measured as a function of the transverse momentum of the fifth leading jet in the event and is is computed using the following formula: $\epsilon_{5j30}(p_T^{5th}, 4j30) = \frac{n(i_{p_T^{5th}})^{5j30}}{n(i_{p_T^{5th}})^{4j30}}$.

Figure 4.6 shows the efficiency for the Tag trigger as a function of the fifth reconstructed jet

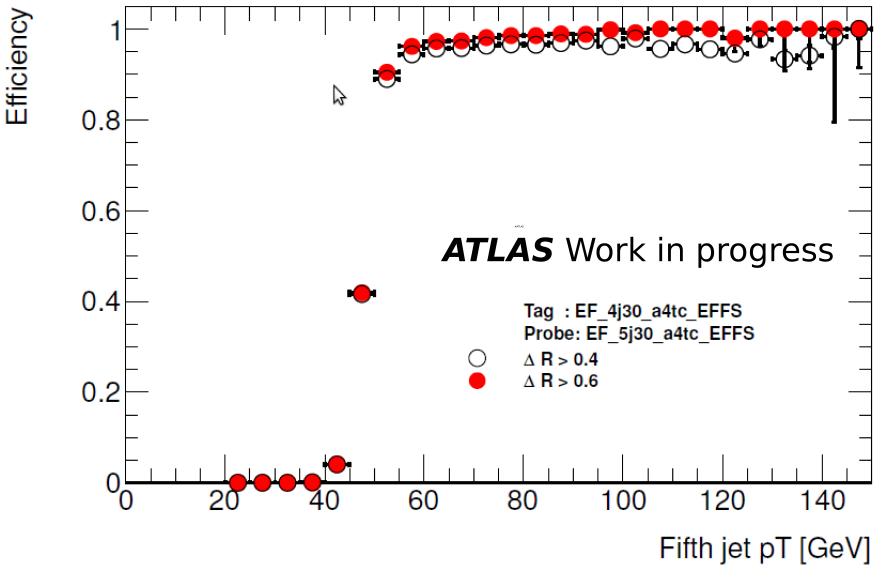


FIGURE 4.6 Trigger efficiency for the 5-jet trigger as a function of the fifth jet p_T with respect to the 4-jet trigger.

p_T. The turn-on curve plateau is achieved for the fifth jet p_T above 55 GeV. The open dots present the trigger efficiency for non isolated jet $\Delta R(\text{jet}, \text{jet}) > 0.4$. Using these non isolated jets, the trigger turn-on curve is not completely efficient. Whereas selecting only the isolated jets ($\Delta R(\text{jet}, \text{jet}) > 0.6$) in each event (red dots) the trigger turn-on curve becomes fully efficient at the plateau. Furthermore by comparing of the trigger efficiency in data and in $t\bar{t}$ MC simulation, Figure 4.7, it is obvious that in data the number of isolated jets is lower than in MC simulation. A better data to MC agreement in the plateau of the trigger turn-on curve is reached adding an isolation jet criteria. In order to have a full trigger efficiency a cut on the minimum distance between any two jets is added in the event selection asking for a $\Delta R > 0.6$.

The all hadronic $t\bar{t}$ channel has two b-quarks, originating from top quark decay, in the final state; thus at least two of the selected jets should be identified as b-jets by the b-tagging algorithm. As already mentioned in Section 4.4.3, the b-tagging criteria used in this analysis is MV1. The transverse momentum of the b-tagged jets should be larger than 55 GeV and within $|\eta| < 2.5$.
An important background is connected to the $b\bar{b}$ production arising from the gluon splitting. It

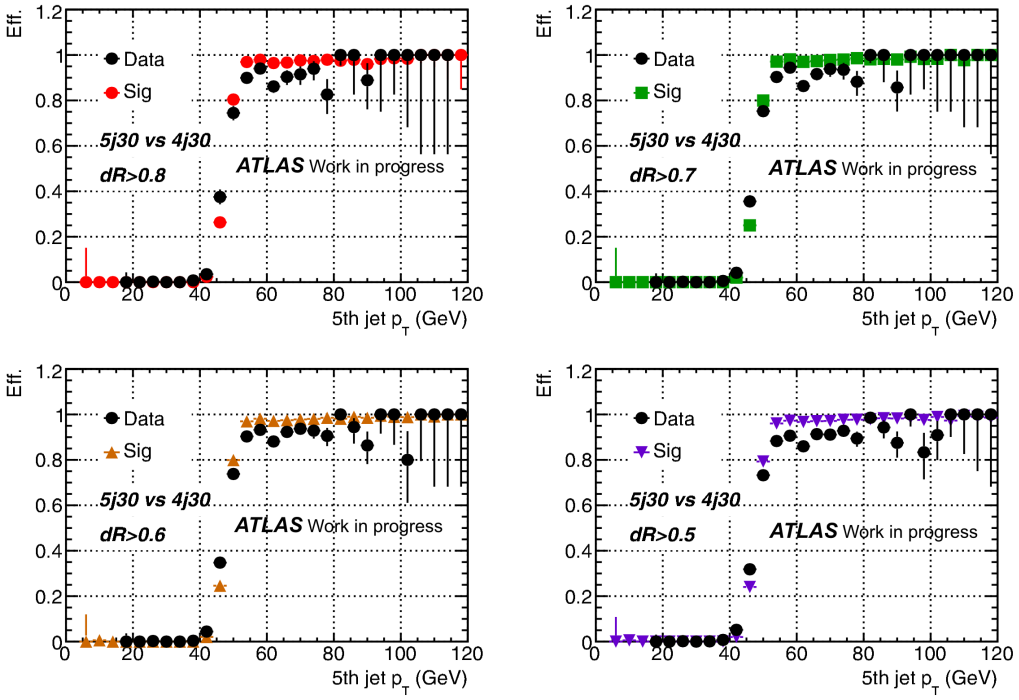


FIGURE 4.7 Turn on curve on the fifth leading jet of five jets trigger ($p_T > 30$ GeV) with the respect to four jets trigger ($p_T > 30$ GeV). The turn on curves are shown for four cuts on the minimum distance ΔR between any two jets in the event, in particular from the top left to bottom right $\Delta R > 0.8$, $\Delta R > 0.7$, $\Delta R > 0.6$ and $\Delta R > 0.5$.

contributes to the QCD multi-jet background. Figure 4.8 shows the distance $\Delta R(b-jet_1, b-jet_2)$ between the two b-tagged jets before the event selection. In the events with more than two b-tagged jets the $\Delta R(b-jet_1, b-jet_2)$ is built using the two b-jets with the high p_T. The b-jets originating from the $t\bar{t}$ events are mostly produced back-to-back (black line), hence the distance $\Delta R(b-jet_1, b-jet_2)$ is peaking at π, while the QCD multi-jet is characterized by a low distance between the two b-jets (green line). As a consequence this observable is used to reduce the $b\bar{b}$ background requiring that the distance between the two b-jets is above $\Delta R(b-jet_1, b-jet_2) > 1.2$.

Selection cut	$t\bar{t}$	Data
Initial number of events	489993	65215848
Trigger EF_5j30	71235 ± 246	12756145
at least 5 jets $p_T > 55$ GeV, $\|\eta\| < 2.5$	27899 ± 162	3701029
non isolated jets $\Delta R > 0.6$	15309 ± 121	922026
at least two b-tagged jets	7447 ± 85	31766
$\Delta R(b-jet_1, b-jet_2)$	4589 ± 66	22969

TABLE 4.1 Events selection cut flow in data and in fully hadronic $t\bar{t}$ MC simulation normalized to the the data luminosity using the theoretical cross section. Uncertainties shown are statistical only.

Table 4.1 presents the events left after the events selection cuts in Data and in MC.

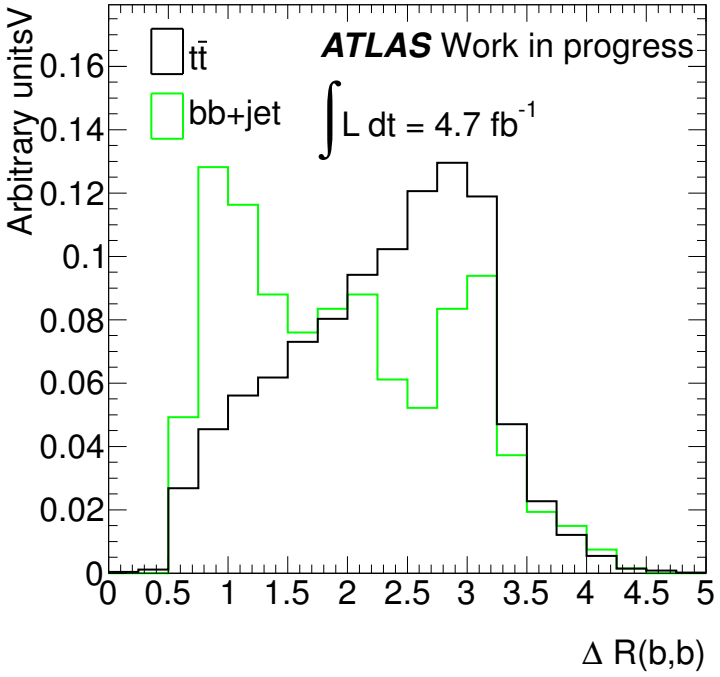


FIGURE 4.8 Minimum distance distribution between the two b-tagged jets for $t\bar{t}$ and $b\bar{b}$+jets MC.

4.6 Characterization of the background sources to the $t\bar{t}$ fully hadronic events

Fully hadronic $t\bar{t}$ events are hidden among background events. The most challenging task when performing the $t\bar{t}$ cross section measurement in the all hadronic final state is the estimation of the dominant sources of background. This section describes the main backgrounds, their impact on measurement of the $t\bar{t}$ fully hadronic cross section, as well as the background modeling performed in the analysis.

4.6.1 W and Z boson production

A background to this analysis is the W and Z boson production in association with additional jets in the final state, leading to a topology which could be similar to the $t\bar{t}$ fully hadronic one.

Table 4.1(b) and Table 4.1(a) presents the branching ratio of the W and Z bosons respectively. For both, the dominant decay mode is the hadronic channel. The events in which the EW bosons do not decay hadronically, are characterized by a presence of isolated leptons and missing transverse energy that can be easily separated from the fully hadronic $t\bar{t}$ signal events. The exclusion of events that present an isolated lepton in the final state reduces the amount of these background events. Therefore due to lepton veto required (Section 4.4.2) it can be considered negligible in this analysis. The W/Z decaying in the hadronic mode cannot be eliminated completely. The main

Decay mode	Branching ratio (%)
$Z \to qq$	69.9
$Z \to \nu\nu$	20
$Z \to e^+e^-$	3.4
$Z \to \mu^+\mu^-$	3.4
$Z \to \tau^+\tau^-$	3.4

Decay mode	Branching Ratio (%)
$W^+ \to u\bar{d}, c\bar{s}$	69.6
$W^+ \to e^+\nu_e$	10.8
$W^+ \to \mu^+\nu_\mu$	10.8
$W^+ \to \tau^+\nu_\tau$	10.6

TABLE 4.2 Born level theoretical branching ratios of Z boson decay (left) and the W^+ boson decay (right), assuming lepton universality. Identical values are obtained for the W^- [7].

background in this case is when the W/Z are produced in association with at least four jets, since this signature can mimic the $t\bar{t}$ final state. The cross section of fully hadronic W boson produced in association with at least four jets at $\sqrt{s} = 7$ TeV is about 190 pb [115], while for the hadronic decaying Z boson is about 13.5 pb [116]. While their contribution was not considered explicitly an estimation of the fully hadronic W/Z after preselection comes from the semi-leptonic analysis [117] and is of order of 4% and 2%, respectively. This contamination is then furtherly reduced by requirements applied at analysis level when $t\bar{t}$ reconstruction algorithms are applied (Sections 4.9, 4.10).

4.6.2 $t\bar{t}$ non hadronic background

The other $t\bar{t}$ decay modes, semi-leptonic and di-lepton channels, can also contribute to the background in the fully hadronic $t\bar{t}$ events. Therefore the amount of these background was investigated. Thanks to the optimized preselection and in particular to the veto on the isolated leptons in the final state, these contributions are small ($\sim 4\%$ after the event selection). They are not considered in the fully hadronic $t\bar{t}$ cross section measurement.

4.6.3 Multi-jet QCD background

A QCD multi-jet background events stem from QCD processes with a pair of light-quarks or gluon in the final state, instead of a pair of top quarks. These particles emit gluons which then hadronize. Thus, jets in the background events originate mainly from gluon radiation whereas jets in $t\bar{t}$ events are predominantly coming from the hadronization of quarks. The QCD processes with two, three and four jets can easily be rejected from the fully hadronic $t\bar{t}$ event topology by requiring a lower limit on the number of reconstructed jets.

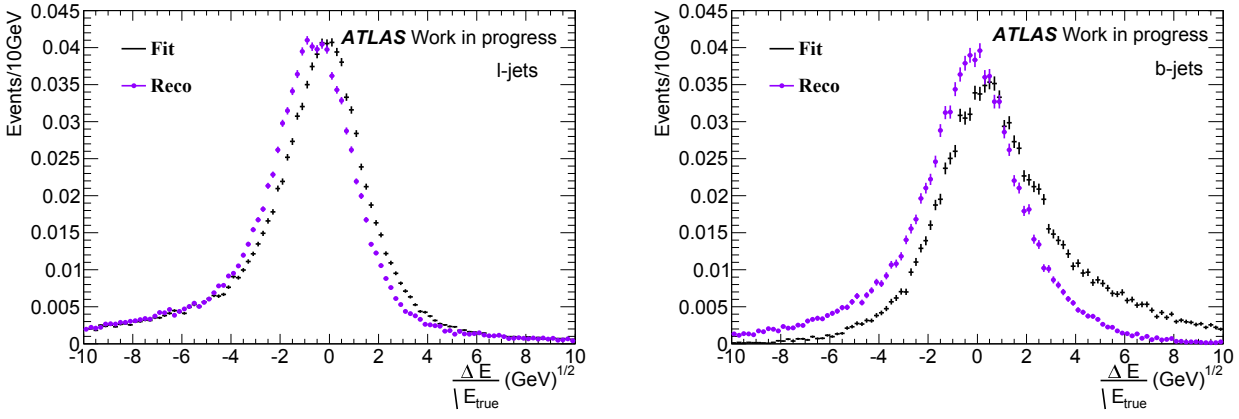
The challenging background processes are the ones that lead to five and six jets per event for which the corresponding cross section production at $\sqrt{s} = 7$ TeV is about 10^3 pb [118]. Their event topology is similar to the $t\bar{t}$ fully hadronic one, so the discrimination between these processes is very hard.

The QCD multi-jet estimation is based on a data-driven technique consisting on the estimation of

the background in an untagged data sample, which means that the whole event selection is applied without the requirement of the presence of b-tagged jets in the events. More detail on the QCD background modeling are provided in Section 4.8.

4.7 Kinematic Fit Likelihood

Once the event selection is performed, it is necessary to reconstruct the $t\bar{t}$ event topology. This is performed by using a kinematic likelihood fitter (KLF) which assigns the observed objects to the parton level predictions from $t\bar{t}$ Monte Carlo simulation in a leading-order picture. The KLF is done using a likelihood approach. The likelihood describes the probability of obtaining some measured quantities given a model. In this analysis the model corresponds to $t\bar{t}$ event in the fully hadronic final decay mode. The quantities used in the likelihood building are the energies and the direction of the six quarks, four light-quarks and two b-quarks.
Given the experimental energy and angular resolution, the quarks information can be only known with an associated uncertainty, thus they are parametrized by transfer functions (TFs), $\mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet_i}} \mid \mathrm{E}_{\mathrm{q_j}}\right)$. The TFs map the energy of an object, $\hat{E}_{jet_i}$, to the energy of the final state particle, E_{q_i}. These functions are derived from $t\bar{t}$ simulation MC@NLO signal samples using reconstructed objects which are geometrical matched in $\eta - \phi$ space to their parent partons. The matching criteria requires that the two objects are inside a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$. The TFs are derived from two dimensional binned likelihood fit: the energy of the reconstruction object and the relative difference between the measured and the true energy. The TFs are parametrized by double Gaussians and are derived for light and b-quarks separately in bins of p_T in four bins of pseudorapidity regions up to $|\eta| < 2.5$. The impact of applying the transfer functions on the measured jet energies is presented in Figure 4.9(a) and Figure 4.9(b) for the light-jets and b-jets respectively.



(a) (b)

FIGURE 4.9 Energy resolution of light jets 4.9(a) and b-tagged jets 4.9(b), before (black dots) and after applying the transfer functions used in the KLFitter (magenta circles).

The mass of the hadronically decaying W bosons are distributed according to a Breit-Wigher distribution around the pole mass of $m_W = 80.4$ GeV.
The kinematic likelihood is defined as:

$$\begin{aligned} \mathrm{L_{kin}} \;=\; & \mathrm{BW}\left(\mathrm{m_{q_1q_2}} \mid \mathrm{m_W}, \Gamma_\mathrm{W}\right) \cdot \mathrm{BW}\left(\mathrm{m_{q_3q_4}} \mid \mathrm{m_W}, \Gamma_\mathrm{W}\right) \cdot \\ & \mathrm{BW}\left(\mathrm{m_{q_1q_2b_1}} \mid \mathrm{m^{reco}_{top}}, \Gamma_\mathrm{top}\right) \cdot \mathrm{BW}\left(\mathrm{m_{q_3q_4b_2}} \mid \mathrm{m^{reco}_{top}}, \Gamma_\mathrm{top}\right) \cdot \\ & \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_1} \mid \mathrm{E_{b_1}}\right) \cdot \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_2} \mid \mathrm{E_{b_2}}\right) \cdot \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_3} \mid \mathrm{E_{q_1}}\right) \cdot \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_4} \mid \mathrm{E_{q_2}}\right) \cdot \\ & \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_5} \mid \mathrm{E_{q_3}}\right) \cdot \mathrm{W}\left(\hat{\mathrm{E}}_{\mathrm{jet}_6} \mid \mathrm{E_{q_4}}\right) \end{aligned} \tag{4.5}$$

where q_i with $i = 1, ..., 4$ are the light quarks from the decay of the two W bosons, b_i with $i = 1, 2$ are the b-quarks from the decay of the two top quarks, j_i represent the calorimeter jets that are assigned to one of the final state partons.

Breit-Wigner functions $\mathrm{BW}(\mathrm{m_{q_1q_2}}|\mathrm{m_W}, \Gamma_\mathrm{W})$ and $\mathrm{BW}(\mathrm{m_{q_1q_2b_2}}|\mathrm{m^{reco}_{top}}, \Gamma_\mathrm{top})$ are used to constrain the di-jet m_{ji} and the triplet m_{ijk} masses to the W boson and top quark masses, respectively. In the BW for W boson both mass (m_W) and width (Γ_W) are kept constant at the known values, whereas for top quark only the width Γ_{top} is considered as constant. The top pole mass ($\mathrm{m^{reco}_{top}}$) is treated as an additional parameter of the fit and is required to be identical for the top and anti-top candidates. The likelihood is characterized by seven parameters: the energy of the six jets $\hat{E}_{jet_i}$ and the top quark pole mass $\mathrm{m^{reco}_{top}}$. The parameter ranges are set of each event. The energy of the partons have to be inside a given range around the measured values, $min(0, \hat{E}-7\cdot\sqrt{\hat{E}}) < E < max(0, \hat{E}+7\cdot\sqrt{\hat{E}})$. The top pole mass is constrained to be between 100 GeV and 400 GeV.
Since a prior association of the jets with the quarks is not possible, all the possible combinations of six jets in each event are made. The fit is then performed on all distinguishable permutations. It is, in fact, possible to find several specific "invariant laws" of the $t\bar{t}$ fully hadronic topology which help to drop from the permutation list redundant combinations. Two invariance laws are considered: invariance under swap of the two triplets and two objects in the triplet. For example we consider six objects (j_1, j_2, j_3, j_4, j_5, j_6), among which we must form two triplets of three objects. The combinations (j_1, j_2, j_3)+(j_4, j_5, j_6) and (j_4, j_5, j_6)+(j_1, j_2, j_3) are not distinguished in the fit procedure, hence it is enough to reconstruct only one and drop the second. Furthermore, there is no way to distinguish among the combinations where an exchange of the two objects, which make the W boson, is made in a triplet, like $(j_1, j_2^W, j_3^W) \equiv (j_1, j_3^W, j_2^W)$. Thus the fit is performed just on one of these jets combinations.
In the fit, the function $-\ln(\mathrm{L_{kin}})$ defined in Equation 4.5 is minimized for each possible combination with respect to the seven parameters of the fit. The fit returns the best fit parameters, the corresponding value of the likelihood and a relative weight for each jet permutation *event probability*. The combination which has the lowest $-\ln(\mathrm{L_{kin}})$ is chosen as the best one.

In the KLF baseline procedure, the b-tagging requirement is ignored. However, the b-tagging could improve the reconstruction efficiency in KLF. Therefore two different methods are available to deal with the b-tagging:

1. the simple veto method,
2. the weight method using a particular working point.

The first method consists in vetoing jet permutations in which the b-tagged jet is placed in the position of a parton coming from a decay of the hadronic W boson. The latter method is a more sophisticated way of using the b-tagging information, and it gives the possibility of choosing a particular working point with a given efficiency for b-tagging and a given light-jets rejection R.
In the analysis presented in this dissertation, the weight method is implemented. A term Δp is introduced in order to take into account whether the jet assigned to a parton has been b-tagged or not. This is defined in the following way:

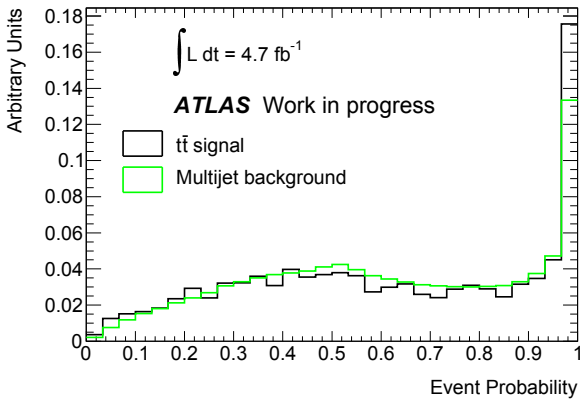


FIGURE 4.10 Event probability distribution for $t\bar{t}$ MC simulation and background samples. Histograms are normalized to the unit area.

$$\Delta \mathrm{p} = \prod_{\mathrm{i}=1}^{2} \left\{ \begin{array}{c} \epsilon; \mathrm{b_i\ b-tagged} \\ (1-\epsilon); \mathrm{b_i\ not\ b-tagged} \end{array} \right\} \cdot \prod_{\mathrm{i}=1}^{4} \left\{ \begin{array}{c} \frac{1}{R}; \mathrm{q_i\ b-tagged} \\ (1-\frac{1}{R}); \mathrm{q_i\ not\ b-tagged} \end{array} \right\}, \tag{4.6}$$

This terms modifies the likelihood to select the permutations in which the b-tagged jet is placed in the b-quark position of the likelihood. The multiplication of the minus the logarithm of the likelihood (Equation 4.5) and the Δp term in Equation 4.6 gives the *event probability*. It can take a value between 0 and 1, see Figure 4.10. The combination which resembles more to a $t\bar{t}$ event gets an *event probability* close to 1, on the contrary the events which represent more a background topology spread up over all the allowed values.
The generation and the fitting procedure of all the combinations of six jets considered by the KLF for each event requires a very important computing time. Thus in order to reduce the CPU time needed by the $t\bar{t}$ system reconstruction procedure an additional cut on the jet multiplicity is applied which requires $6 \leq N_{jet} \leq 10$ with $p_T > 20$ GeV and $|\eta| < 2.5$.

4.7.1 Top mass distribution

The four-vectors of the top and anti-top candidates obtained after the kinematic fit are used to compute the top and anti-top mass m_t. Figure 4.11 shows the top mass distribution reconstructed using the fitted four-vector for $t\bar{t}$ signal events . Comparing the m_t distribution obtained using the reconstructed jet energies of best combination (magenta triangles) with the kinematically fitted ones (in black dots), shows that the latter leads to a better top quark mass resolution.

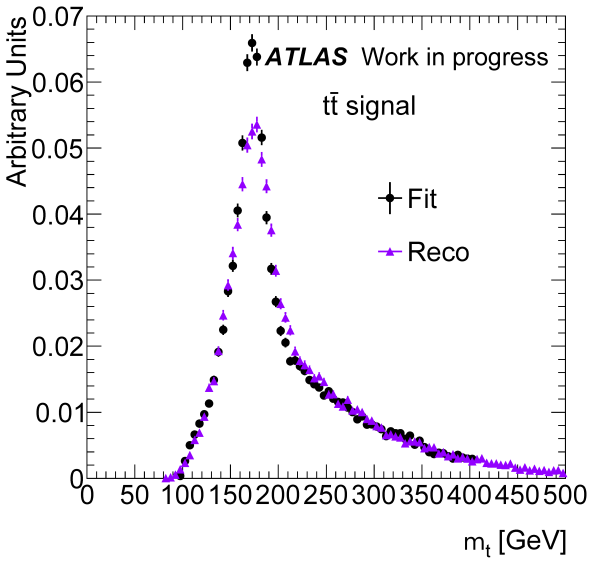


FIGURE 4.11 Results of the kinematic likelihood fit on signal events passing the event selection. The distributions are normalized to unity. The distribution with filled circles is the output of the kinematic fit, whereas the distribution with magenta filled triangles is the mass distribution obtained using the reconstructed jet energies of the best combination.

The distribution shows an asymmetric shape with a clear peak at $\approx$ 170 GeV which is a value compatible with the top quark mass. After matching[4] the truth partons in $t\bar{t}$ decays at the generator level with the reconstructed jets, the top quark mass distribution can be divided in two components:

1. combinations in which at least one jet is wrongly assigned to the true parton,
2. combinations where the $t\bar{t}$ topology is completely matched at truth level.

The top mass distribution for the first category is shown in Figure 4.12 and for the second one in Figure 4.13.

The validation of the KLF tool with the fully hadronic $t\bar{t}$ channel is extensively investigated. The main aspect studied concerns the purity of signal events used in the KLF reconstruction. After event selection the fraction of the events where the reconstructed jets are matched to the

[4]The matching between an offline jet and a truth parton in the MC is made by calculating the angular distance $\Delta R(jet, parton)$ between the two objects. In order to assign a truth parton to a reconstructed jet the corresponding distance should be less then $\Delta R(jet, parton) < 0.3$.

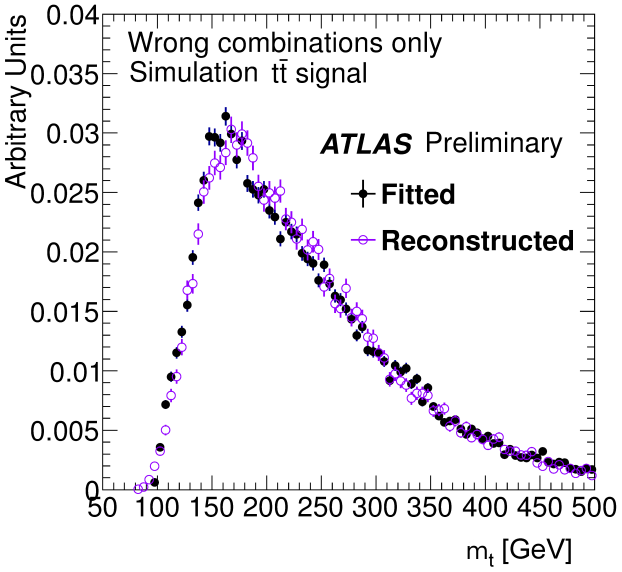


FIGURE 4.12 Mass distributions obtained from the KLFitter for signal events passing the event selection. The distributions are normalized to unity. The distributions are obtained from permutations with at least one jet that is not correctly matched to the $t\bar{t}$. The distribution with filled circles is the output of the kinematic fit, whereas the distribution with magenta filled triangles is the mass distribution obtained using the reconstructed jet energies of the best combination.

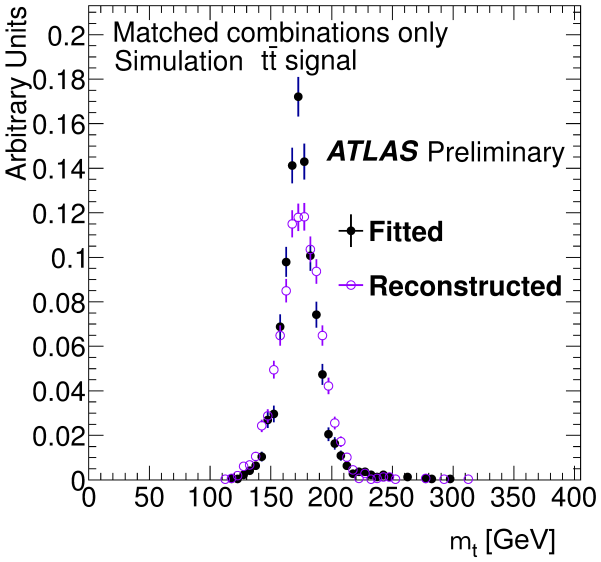


FIGURE 4.13 Mass distributions obtained from the KLFitter for signal events passing the event selection. The distributions are normalized to unity. The distributions are for permutations that are fully matched to the $t\bar{t}$ system. The distribution with filled circles is the output of the kinematic fit, whereas the distribution with magenta filled triangles is the mass distribution obtained using the reconstructed jet energies of the best combination.

truth partons from top decays is 29%. The figure of merit for the KLF performance is the purity defined as the ratio between the number of events in which the KLF assigns the correct jets to the reconstructed objects and the total number of events where all partons are matched to jets. The

fraction of events where the KLF uses both W boson decay products (all the jets which are used to reconstruct the W boson are matched with the decay products of a truth W boson) is 48%. Then, the fraction of events in which at least one of the top quark candidate is matched to the correct partons is 71%. The fraction of the events in which the KLF is able to entirely reconstruct the topology is 40%. Table 4.3 summarizes the results of the study.

Configuration	Purity
at least 1 b_{KLF} matched	0.95
2 b_{KLF} matched	0.58
at least 1 W_{KLF} matched	0.84
2 W_{KLF} matched	0.48
at least 1 t_{KLF} matched	0.71
$t\bar{t}_{KLF}$ matched	0.40

TABLE 4.3 Summary of the study of the KLF purity with "at least 1 b_{KLF} matched" is indicated the fraction of events in which the KLF uses at least one b-jet which is matched to a truth b-quark, with "2 b_{KLF}" the fraction of events in which the KLF finds the good assignment for both b-quarks. With "at least 1 W_{KLF} matched (2 W_{KLF} matched)" label is indicated the purity for the good assignment of at least one (two) W boson(s) and "at least 1 t_{KLF} matched" label includes the events where at least one top reconstructed by KLF is built using only jets matched to the truth level. Finally, "$t\bar{t}_{KLF}$ corresponds to the matched" is the fraction of the events where the KLF finds the true jet combination for the $t\bar{t}$ system. All fractions are computed with respect to the number of events where the $t\bar{t}$ system is fully reconstructed (29% of all MC $t\bar{t}$ events after pre-selection).

4.8 Multi-jet QCD background modeling

The strategy followed in the analysis consists of a data-driven procedure to estimate the QCD multi-jet background. The idea implemented is quite simple and is based on the assumption that the top quark mass m_t reconstructed by the KLF does not depend strongly on the flavour composition of multi-jet processes. Using this hypothesis it is possible to study and model the background properties in an un-tagged data sample where the contamination of the $t\bar{t}$ signal events is very low. The un-tagged sample is defined as events passing all the analysis cuts except the b-tagging requirement. In order to validate this strategy, several studies were performed on ALPGEN MC multi-jet simulation, mainly to understand the impact of flavour composition on m_t. The effect of the b-tagging requirement consists in sculpting the jet transverse momentum spectrum and in the deformation of the event kinematic variables. The size of this effect is evaluated by studying separately the general multi-jet production and exclusive $b\bar{b}$+jets events. Figure 4.14, Figure 4.15 and Figure 4.16 summarize the results of the studies. Figure 4.14 and Figure 4.15 show the comparison of the fitted top mass distribution before (red triangles) and after (black dots) applying the b-tagging requirement on the inclusive multi-jet production and in the $b\bar{b}$ exclusive one, respectively. The low pads show the ratio of the distributions fitted with a first order polynomial $p_0 + p_1 \cdot x$, in order the quantitatively estimate the b-flavour effects.

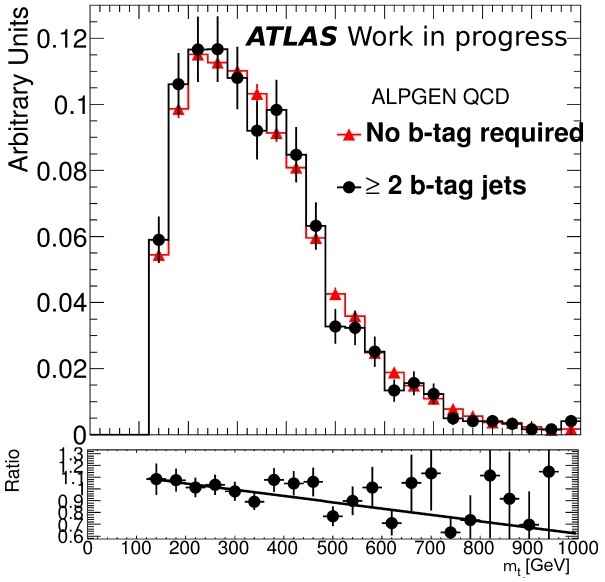


FIGURE 4.14 m_t distribution obtained after the kinematic fit on MC multi-jet events generated with ALPGEN. The distribution is shown before requesting b-tagged jets (red triangles) and after requesting at least two b-tagged jets (black dots).The function used for the fit of the ratio is a first order polynomial : $p_0 + p_1 \cdot x$.

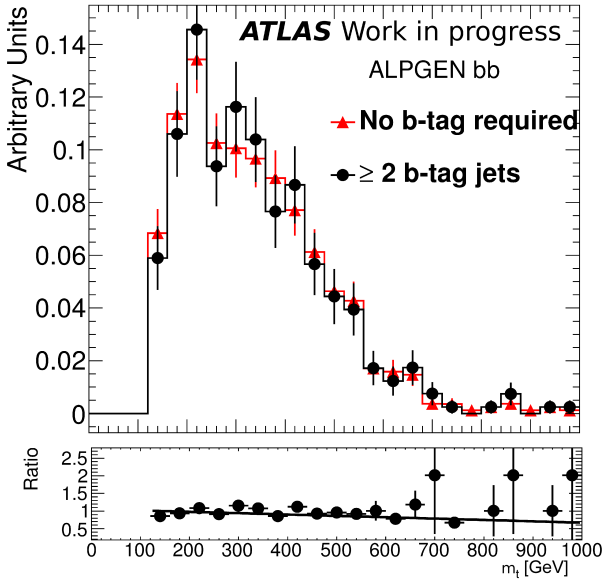


FIGURE 4.15 m_t distribution obtained after the kinematic fit on MC $b\bar{b}$+jets events generated with ALPGEN. The distribution is shown before requesting b-tagged jets (red triangles) and after requesting at least two b-tagged jets (black dots). The funtion used for the fit of the ratio is a first order polynomial : $p_0 + p_1 \cdot x$.

The results of the fit are presented in Table 4.4. For any particular sample the effect of having or not two b-tagged jets in the events is of order of $15 - 20\%$. Other effects can be due to the difference between the kinematics of light and b-jet production. Figure 4.16 shows the latter effect comparing the fitted top mass extracted in the untagged sample for the inclusive light-jets (red) and exclusive (black) flavour multi-jets production. The difference in the m_t shape between an

Fit Results	p_0	p_1	χ^2/ndf
ALPGEN QCD	1.15 ± 0.06	-0.0005 ± 0.0001	0.7
ALPGEN bb	1.06 ± 0.06	-0.0004 ± 0.0001	0.9
ALPGEN QCD vs bb	1.23 ± 0.06	-0.0008 ± 0.0001	1.6

TABLE 4.4 Fit results for the three comparison: m_t in tagged and untagged sample for ALPGEN QCD, m_t in tagged sample for ALPGEN $b\bar{b}$ and ALPGEN QCD. The function used for the fit of the ratio is a polynomial : $p_0 + p_1 \cdot x$.

inclusive light-jet sample and an exclusive $b\bar{b}$+jets one is quantified to be of the order of 20% . To take into account the effects due to the above assumption, an associated systematic uncertainty is introduced which is discussed in Section 4.13.

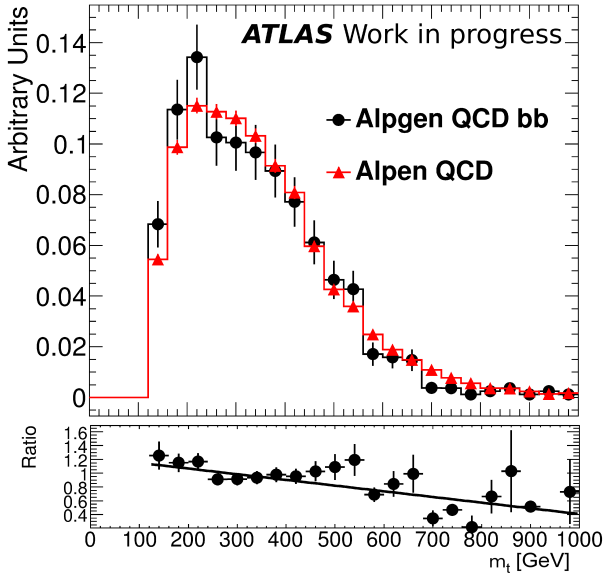


FIGURE 4.16 m_t distribution obtained after the kinematic fit on MC events generated with ALPGEN. The distribution is shown before requesting b-tagged jets for the inclusive light-jets (red triangles) and exclusive (black dots) flavour multi-jets production. The function used for the fit of the ratio is a first order polynomial : $p_0 + p_1 \cdot x$.

4.9 Minimum χ^2 discriminant

A χ^2-based discriminant observable is used to further test the compatibility of the selected events with the $t\bar{t}$ hypothesis. After assigning jets to the different decay products, it aims to distinguish the $t\bar{t}$ signal from the multi-jet background by looking at the consistency of the kinematics with the expected top quark and W boson masses. The χ^2 is computed for each of the different $t\bar{t}$ hypothesis, which are the the not redundant jets combination. The correct assignment in each event corresponds to the jet combination which minimizes the Equation 4.7.

$$\chi^2 = \frac{(m_{j_1,j_2} - m_W)^2}{\sigma_W^2} + \frac{(m_{j_1,j_2,b_1} - m_t)^2}{\sigma_t^2} + \frac{(m_{j_3,j_4} - m_W)^2}{\sigma_W^2} + \frac{(m_{j_3,j_4,b_2} - m_t)^2}{\sigma_t^2}, \qquad (4.7)$$

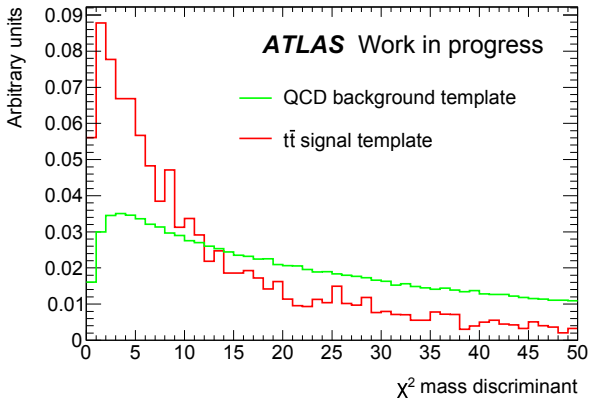


FIGURE 4.17 Mass chi-square discriminant probability density functions for the $t\bar{t}$ signal (red) and the QCD background (green) derived from an untagged data sample.

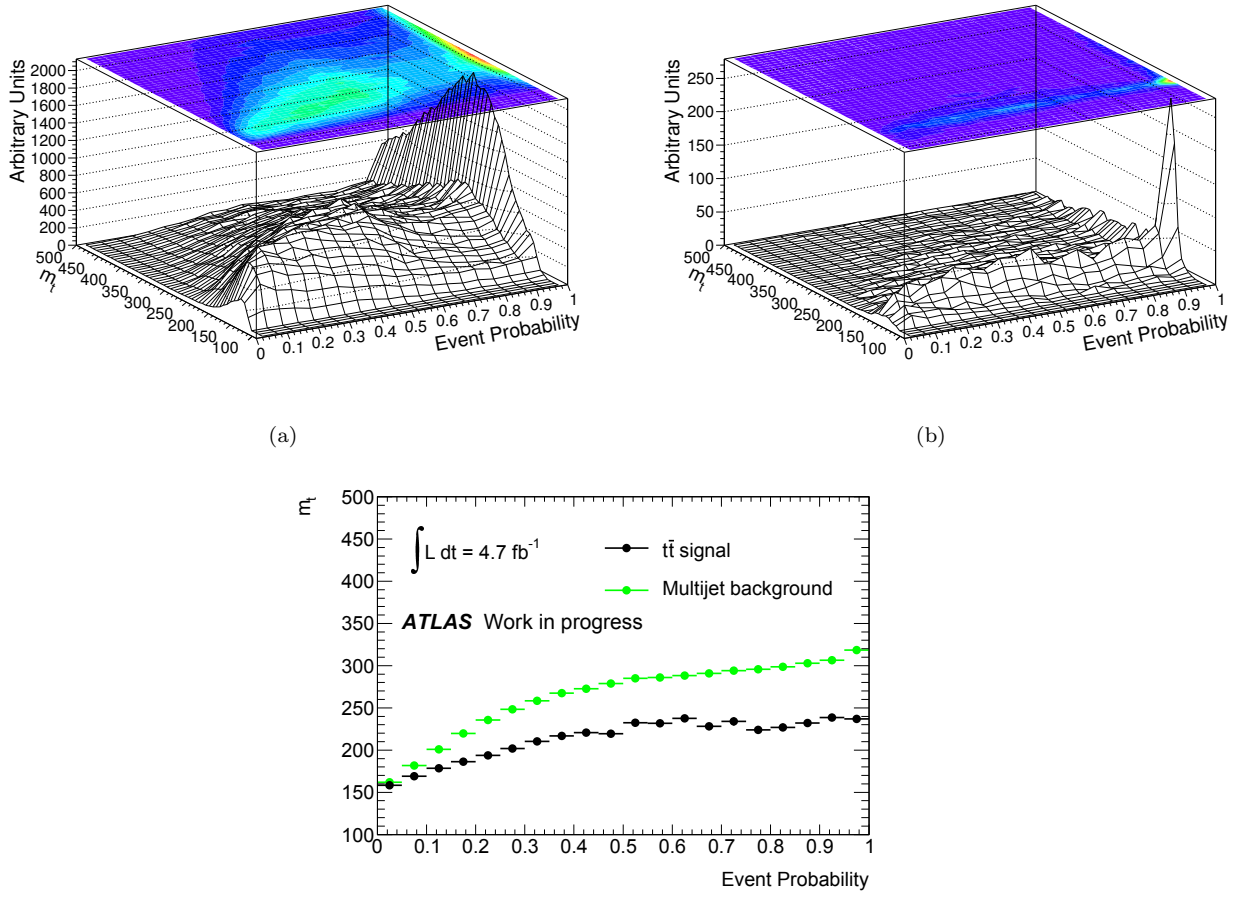
where m_t, m_W, σ_t and σ_W are the reconstructed top and W masses and the associated resolutions respectively. Their values are extracted by matching the reconstructed jets to the truth jets from the original $t\bar{t}$ partonic process and then by fitting the reconstructed invariant masses with a gaussian function and are found to be $m_W = 85.7$ GeV, $\sigma_W = 10.2$ GeV, $m_t = 177.9$ GeV and $\sigma_t = 17.4$ GeV.

If more than two b-tagged jets are found in the events, the χ^2 method uses just two of them in the minimization procedure, dropping the other jets from the $t\bar{t}$ system reconstruction, keeping the two giving a lower χ^2 value. The same procedure is followed also if the events have more than six jets: only six jets, including two b-tagged jets, are used to reproduce the $t\bar{t}$ final state. Figure 4.17 shows the χ^2 mass distribution for the $t\bar{t}$ signal generated with MC@NLO and for the background modelled with a data-driven template.

To reduce the contribution of the background process, a cut on the χ^2-discriminant variable is applied. As one can see, the QCD multi-jet background is characterized by a more spread distribution with a long asymmetric tail, whereas the $t\bar{t}$ events are more accumulated at low value of the χ^2 mass discriminant. The loose requirement added to the event selection requires a χ^2 mass discriminant value to be less than 30.

4.10 Event Probability

By comparing the behaviour of the event probability as a function of the fitted top quark mass reconstructed by the KLF in the background and $t\bar{t}$ signal samples, it is possible to find an additional characteristic of signal events. From the two dimensional histograms, the event probability versus fitted top quark mass are shown in Figure 4.18(a) and Figure 4.18(b) for the untagged sample and the $t\bar{t}$ MC simulation respectively. One can see that the mean of the fitted top quark mass in the background events increases with the event probability values, whereas in the $t\bar{t}$ events it shows an almost constant mean value. Figure 4.18(c) compares the profile of the two dimensional histogram,



(a) (b)

(c)

FIGURE 4.18 4.18(a) Event probability versus m_t distribution obtained after the kinematic fit for the untagged sample, background sample. 4.18(b) Event probability versus m_t distribution obtained after the kinematic fit for the $t\bar{t}$ MC simulation sample. 4.18(c) comparison of the profile of the two dimensional histogram in background and in $t\bar{t}$ samples.

making clearer the separation between the two production processes.

The shift of the fitted top mass in the background is very useful for the cross section measurement, because it helps in the separation of the contributions at the peak mass value. Therefore the last criteria to define a signal region is that the event probability is chosen to be larger than 0.8 which is the cut that provides a good discrimination between the $t\bar{t}$ signal and the background.

4.11 Cross-section measurement

Additional event requirements are added to the selection outlined in Section 4.5, such as the event probability requirement, the cut on the mass χ^2 and lower limit to the top mass $m_t > 125$ GeV. Note that with these additional selection requirement, the contamination of the EW bosons production backgrounds is further reduced. The $t\bar{t}$ production cross section in the fully hadronic

final state is extracted by the means of an extended unbinned maximum likelihood fit [119] of the signal and background templates to the data sample.
The probability density function, PDF, for signal and background events are derived from the MC simulation and the data untagged sample, respectively. The regular likelihood function is defined as the product over the observed events of the total PDF (signal plus background). The extended likelihood can be then formed by multiplying the regular likelihood with a Poisson term

$$\mathcal{L}(N,\alpha) = \frac{e^{-N}N^{N_{obs}}}{N_{obs}} \times \prod_{i=i}^{N_{obs}} PDF(x,\alpha) \tag{4.8}$$

where N and N_{obs} are the number of the expected and observed events, respectively; x is the observable used and α is a parameter to be estimated. In this case the observable is the fitted top mass reconstructed by the kinematic fit and the α parameter represents the fraction of the signal event yields. The best fit parameter is achieved by a minimization procedure of the $-\log\mathcal{L}$. The

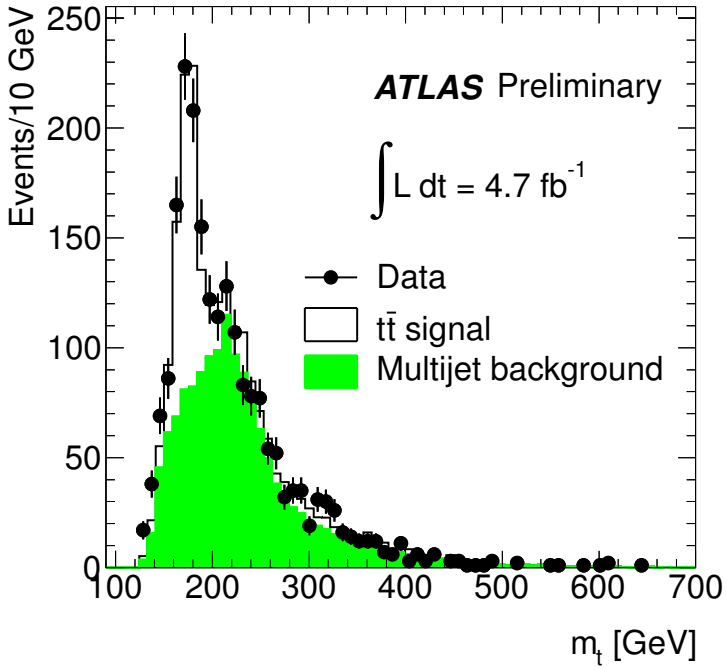


FIGURE 4.19 Fit of the m_t distribution with an unbinned likelihood function to the selected data sample (dots). The errors bars associated to the data are statistical only.

result of the extended unbinned likelihood is shown in Figure 4.19. The $t\bar{t}$ production cross section in the fully hadronic final state is given by the following formula:

$$\sigma_{t\bar{t}} = \frac{N^{data} \times C_{fit}}{L_{int} \times \epsilon_{t\bar{t}} \times BR} \tag{4.9}$$

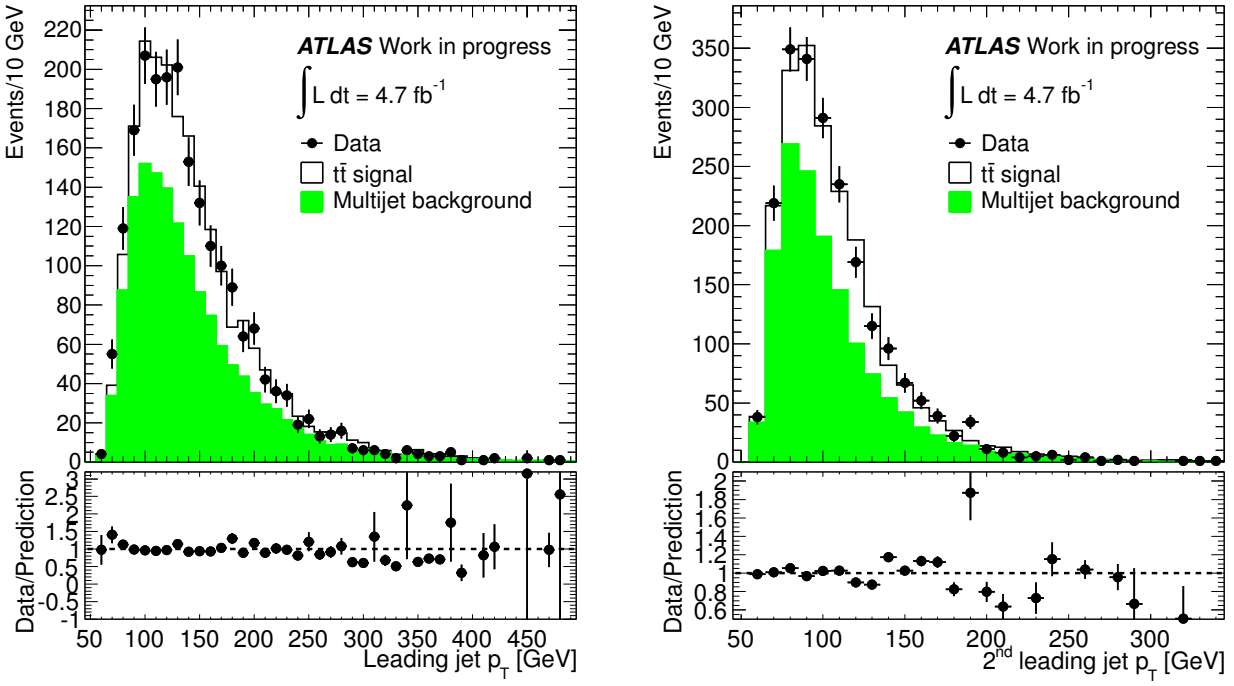
where:

- N^{data} is the number of events observed in the data sample,
- C_{fit} is the fraction of signal events in the observed data, extracted from the fit,
- L_{int} is the integrated luminosity,
- $\epsilon_{t\bar{t}}$ is the $t\bar{t}$ signal efficiency, defined as the ratio between the number of events passing the selection and the total number of events in the generated sample,
- BR is the branching ratio of the fully hadronic $t\bar{t}$ channel 44%.

The unbinned likelihood is performed on a dataset of 2118 events and the result of the fit predicts 666 $\pm$ 48 (stat.) signal event and 1452 $\pm$ 104 (stat.) background events. Considering that the efficiency of simulated $t\bar{t}$ events is 0.187% the computed value for the total $t\bar{t}$ cross section is from the Equation 4.9, equal to: $\sigma_{t\bar{t}} = 168 \pm 12$ ($stat.$) pb.

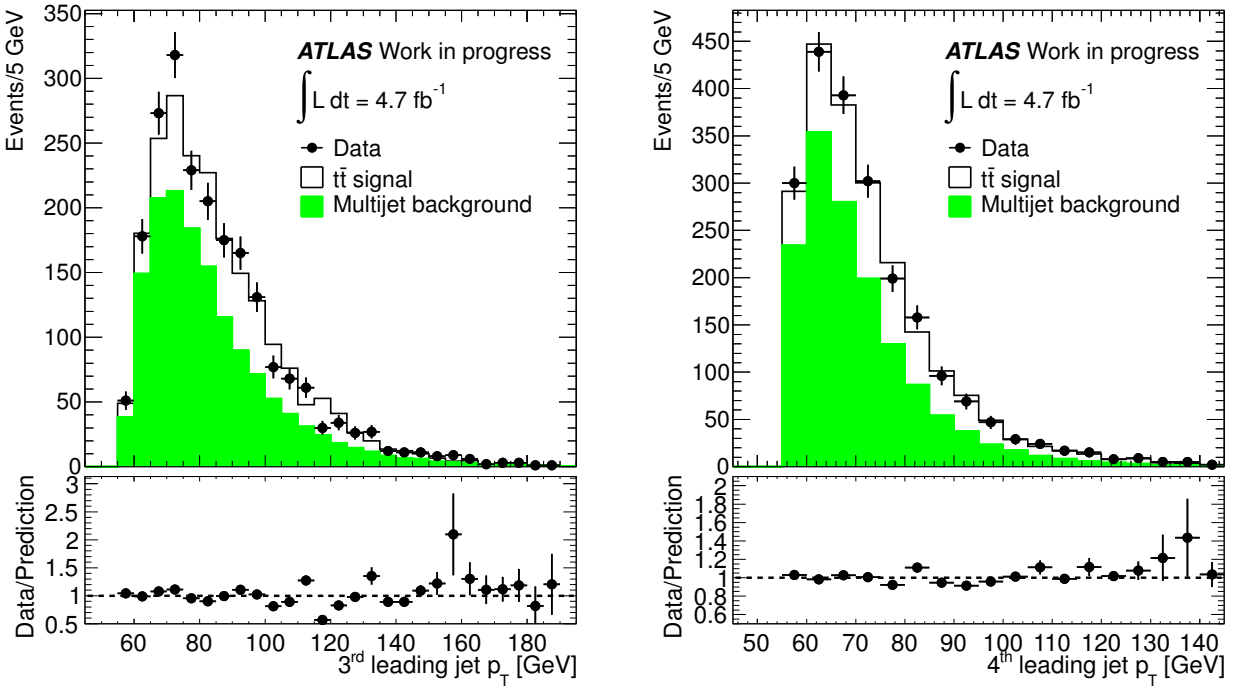
4.12 Control Plots for the main kinematic variables

In order to assess the data to prediction agreement achieved in the analysis using the signal and background templates and the normalization extracted by the maximum likelihood fit procedure, we look at distributions of the main kinematical variables. A set of these distributions are presented in this section. In particular the transverse momentum spectrum, the pseudorapidity and the azimutal angular distributions for the six leading jets in the events are shown in Figures 4.20, 4.21, 4.22, 4.23, 4.24, 4.25. The scalar sum of the jet transverse energy in each event is also presented in Figure 4.26(d). The last set of plots (Figures 4.26(a), 4.26(b), 4.26(c)) show angular distributions between different reconstructed objects.



(a)

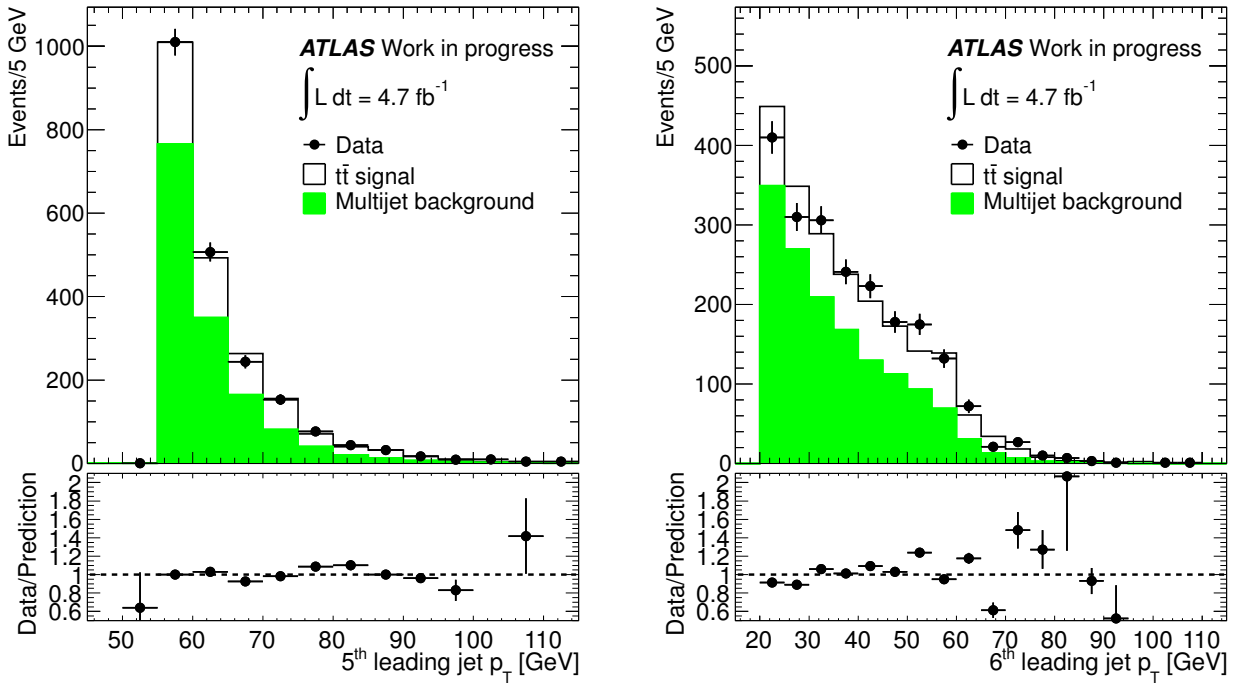
(b)



(c)

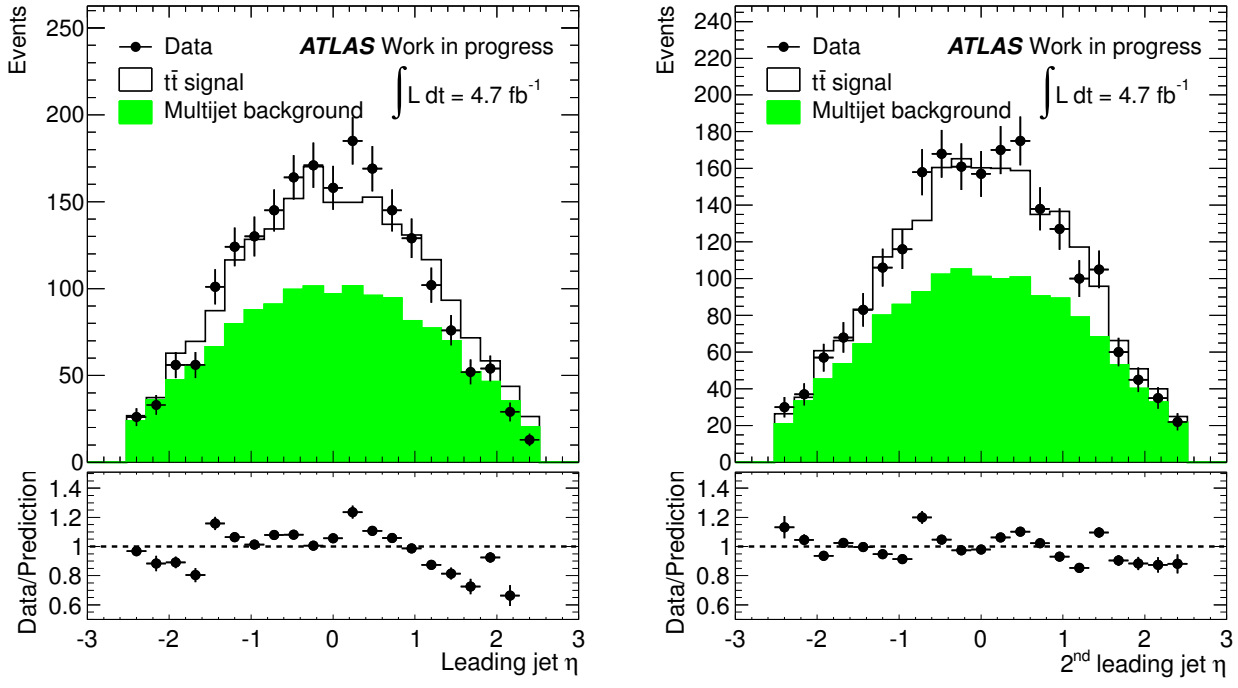
(d)

FIGURE 4.20 Data versus prediction comparison of the p_T distribution of the four leading jets.



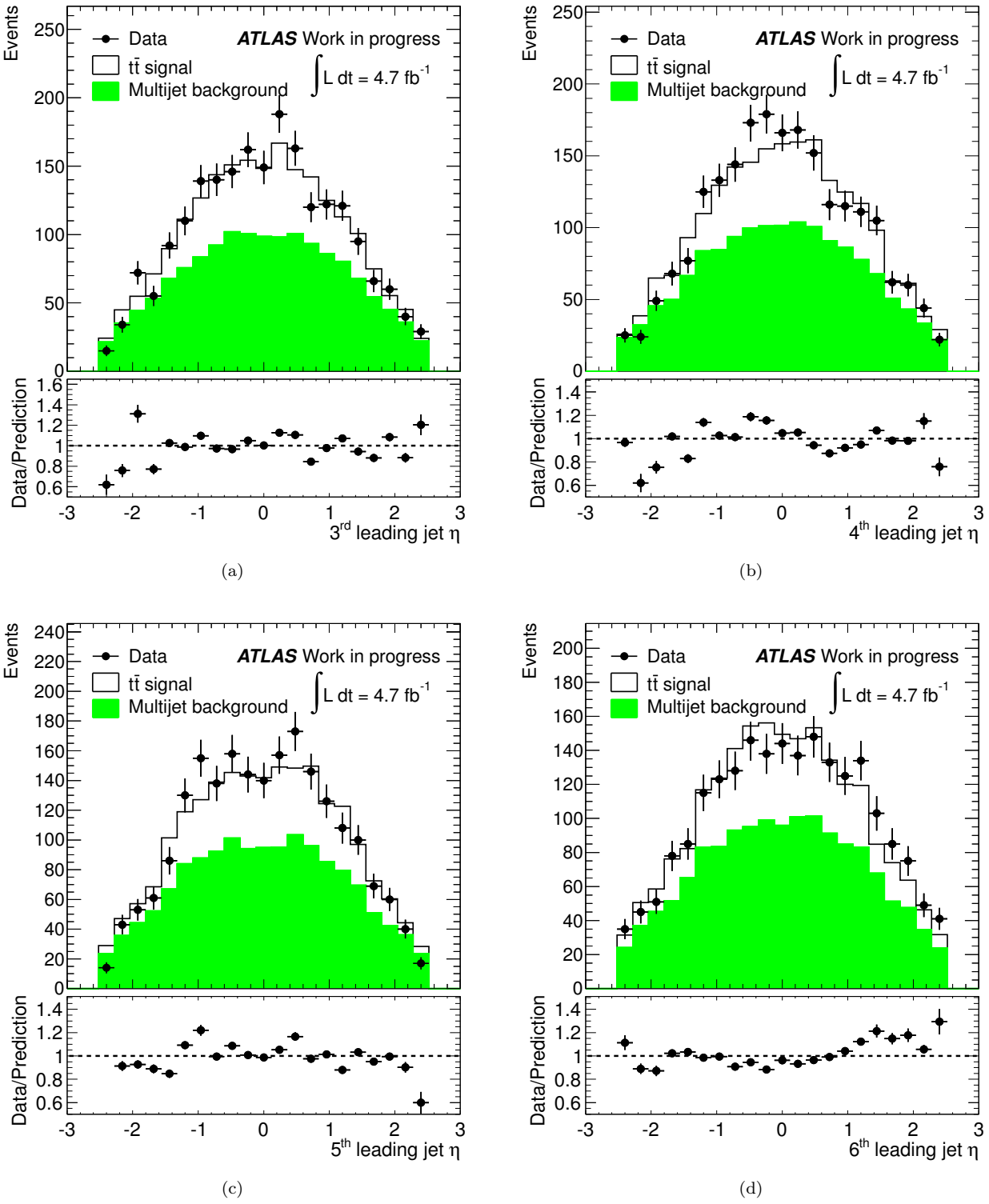
(a) (b)

FIGURE 4.21 Data versus prediction comparison of the p_T distribution of the fifth and sixth leading jets.



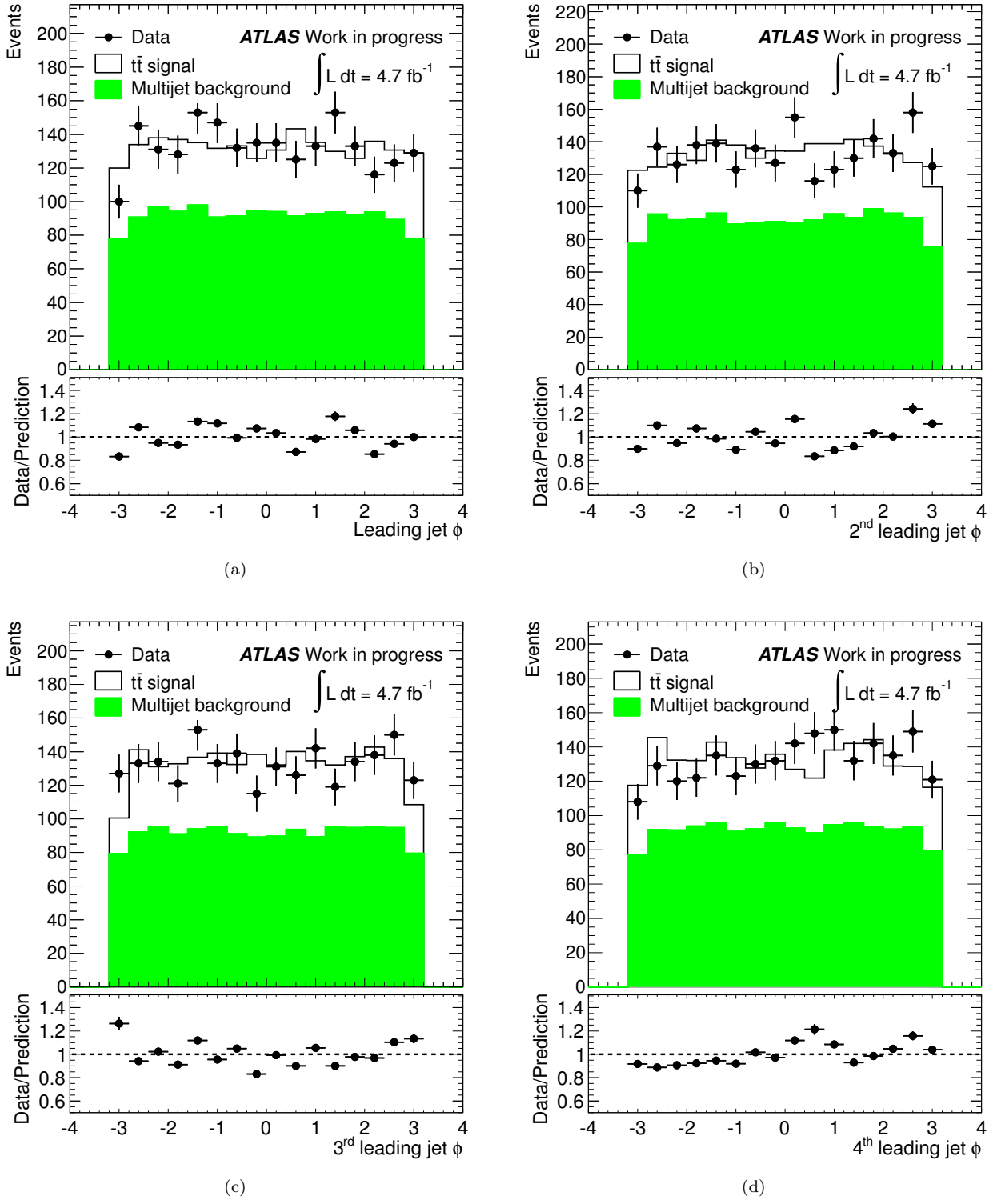
(a) (b)

FIGURE 4.22 Data versus prediction comparison of the η distribution of the first and second leading jets.



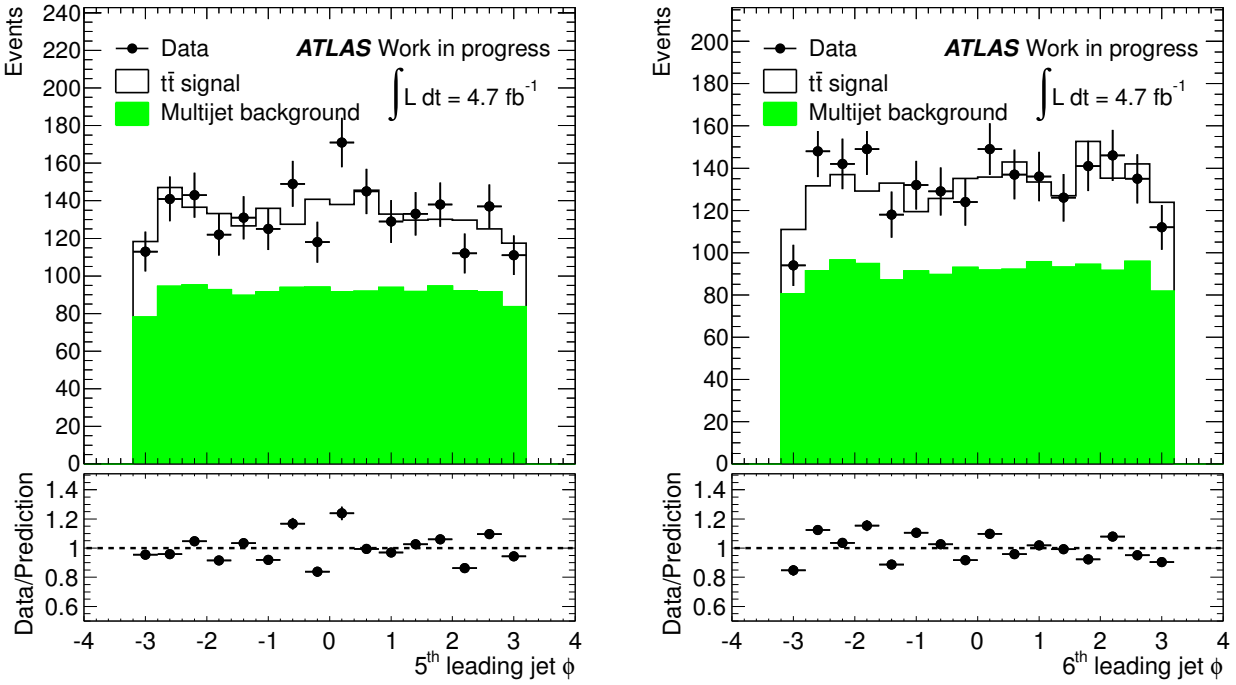
(a) (b) (c) (d)

FIGURE 4.23 Data versus prediction comparison for η distribution of the third, fourth, fifth and sixth leading jets.



(a) (b) (c) (d)

FIGURE 4.24 Data versus prediction comparison of the ϕ distribution of the four leading jets.



(a) (b)

FIGURE 4.25 Data versus prediction comparison of the ϕ distribution of the fifth and sixth leading jets.

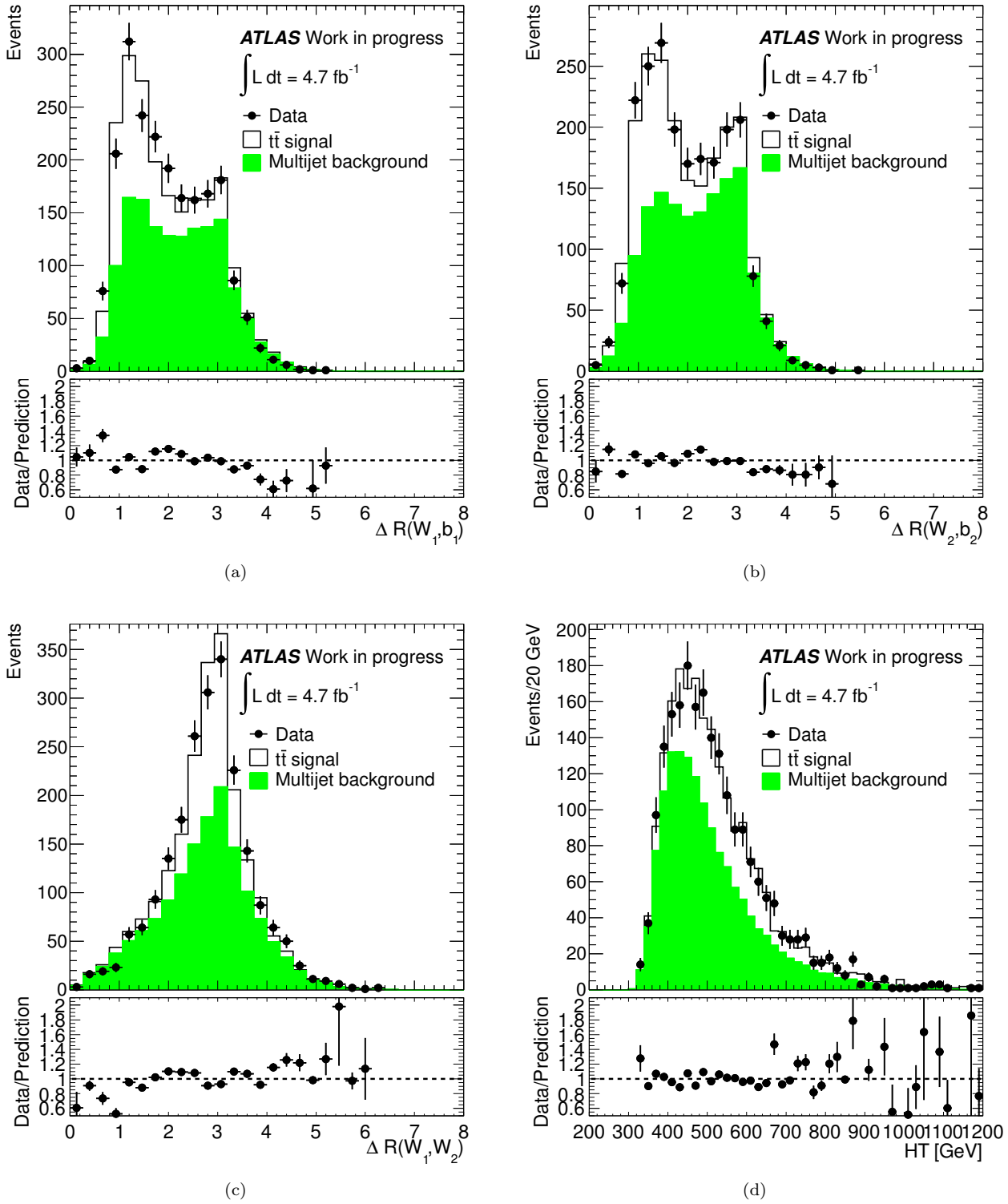


FIGURE 4.26 Data versus prediction comparison of the distance between the W boson and b-jet originating from the top quark decay 4.26(a) (second top quark in the event 4.26(b)), the angular distance 4.26(c) between the two W bosons in the event and the scalar sum of the jets energy 4.26(d).

4.13 Systematic uncertainties

Systematic uncertainties represent the imperfect knowledge of various parameters that could affect the measurement. Most of the systematic uncertainties in this analysis are connected to the signal modelling predicted using the MC simulation. Precise studies are dedicated to their evaluation and understanding with the aim to reduce their influence on the final result. The systematic uncertainties are defined as $\pm 1\ \sigma$ variations of a parameter with respect to the normal settings, or as comparison between different models. The only systematic uncertainty related to the background is the one related to shape modelling. Concerning the signal $t\bar{t}$ MC sample the systematic uncertainties can be divided into two categories, one related to the shape effect and the other to the acceptance effect. The different systematic sources are described in the following sub-sections.

4.13.1 Jet energy scale (JES) and the associated uncertainty

The hight jet multiplicity that characterizes the $t\bar{t}$ system in the fully hadronic channel lead to a high sensitivity to the jet energy scale uncertainty. The jet energy scale and its uncertainty come from the combination of the information recorded during the test-beam, LHC collision data and MC simulation. The JES uncertainty has various components originating from the calibration method, the calorimeter response, the detector simulation and the specific choice of the parameter in the physics model employed in the MC event generator. The JES uncertainty is parametrized as a function of the transverse momentum p_T and pseudorapidity η. It varies in the central region between 1.5% and 6% depending on the jet transverse momentum, see Figure 4.27. These values include the uncertainties in the flavour composition of the sample and different response of the detector depending on the jet flavour, pileup, presence of not isolated jets (close by jet), but does not include the b-jet energy scale [80].

4.13.1.1 Flavour composition and response

The JES uncertainty due to flavour effects covers the possible shifts in the JES caused by:

- uncertainties in the flavour composition of the sample;
- different flavour composition in data and MC;
- uncertainty on the flavour response itself;
- different calorimeter response depending on the jet flavour.

This uncertainty can be expressed by the following formula:

$$\Delta\mathcal{R}_S = \Delta f_g \times (\mathcal{R}_q - \mathcal{R}_g) \oplus f_g \times \Delta\mathcal{R}_g \tag{4.10}$$

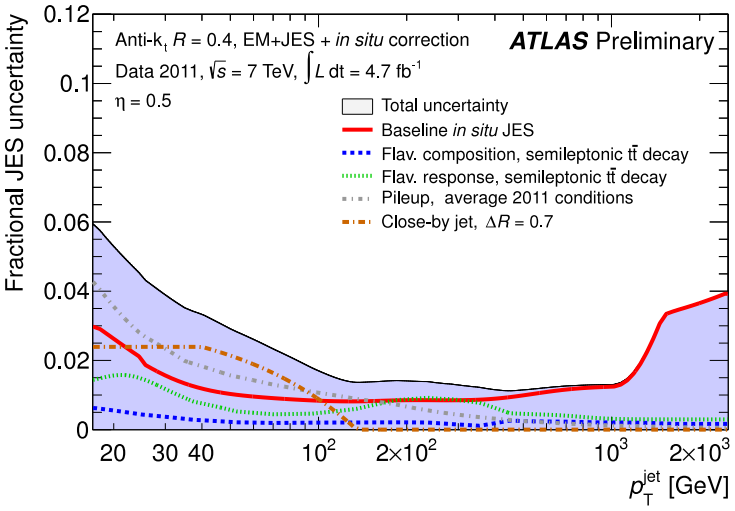


FIGURE 4.27 Sample-dependent fractional jet energy scale systematic uncertainty as a function of jet p_T and jet pseudorapidity η for anti-kt jets with distance parameter of R = 0.4 calibrated using the EM+JES calibration scheme. The uncertainty shown applies to semileptonic top decays with average 2011 pile-up conditions, and does not include the uncertainty on the jet energy scale of b-jets [80].

where f_g and Δf_g are the gluon jet fraction and its uncertainty respectively. The $\mathcal{R}_q$ and $\mathcal{R}_g$ are the jet responses for quark and gluon initiated jet respectively.

4.13.1.2 *b*-jet energy scale (*b*-JES)

The b-jet energy scale takes in to account the remaining differences among the jets originating from light quarks and from a b-quarks after the global JES calibration. An uncertainty, ranging from 1.5% to 3% and depending on jet transverse momentum p_T and pseudorapidity η, is assigned to the jets that have been b-tagged in order to take in account the difference between the light jets, gluon jets, and jets containing b-hadrons.

4.13.1.3 Close by jets

The JES calibration is extracted from isolated jets, however the jet calibration can be affected by the presence of close-by jets. Hence an additional component should be added to the JES uncertainty in order to include the response for non-isolated jets. The variation of the jet energy scale and its uncertainty due to the close-by jets is studied on anti-k_T jets built from the inner detector tracks and comparing the data to MC simulation. The jet p_T is compared to the track p_T matched in the (ϕ, η)-plane when a second jet with transverse momentum at EM-scale larger than 7 GeV is near a high-p_T jet. The p_T ratio of the calorimetric jet and the track jet is examined as a function of the minimum distance (R_{min}) between the two calorimetric jets. Then a Data to Monte Carlo simulation comparison is performed on the ratio of the calorimeter to the track jet

p_T between isolated ($R_{min} > 1.0$) and non-isolated ($R_{min} < 1.0$)

$$A_{close-by} = \frac{[r^{calo/trackjet}_{non-iso/iso}]_{Data}}{[r^{calo/trackjet}_{non-iso/iso}]_{MC}}, \qquad r^{calo/trackjet}_{non-iso/iso} = \frac{r^{calo/trackjet}_{non-iso}}{r^{calo/trackjet}_{iso}}, \qquad r^{calo/trackjet} = \frac{p_T^{calojet}}{p_T^{trackjet}} \tag{4.11}$$

The $A_{close-by}$ represents the uncertainty due to the close by jets effect. It is equal to about $1-2\%$ for a jet $p_T < 100$ GeV and a $R_{min} = 0.7$.

4.13.1.4 Pile-up

The pile-up correction is derived from the Monte Carlo simulation samples. It is estimated by studying the difference between the reconstructed jet p_T and the truth jet p_T as a function of the number of reconstructed primary vertices N_{PV} [5] and the expected average number μ of the interactions per bunch crossing [6] . The reconstructed jet p_T increase approximatively by 370 MeV per reconstructed primary vertex and by 60 MeV per expected average additional interaction. The resulting systematic uncertainty for a jet with $p_T > 40$ GeV is at most 3% in the highest pile-up condition reached during 2011 data-taking. A significant reduction is achieved at high jet p_T $(0.1-0.2\%)$.

4.13.2 Jet reconstruction efficiency (JRE)

The calorimeter jet reconstruction efficiency, related to jets reconstructed from the charged tracks in the inner detector, is determined using a tag and probe technique. This efficiency is defined as the fraction of the probe-track matched to the calorimeter jet. The MC simulation and data are found to be in agreement on the jet reconstruction efficiency at the level of 2% . The uncertainty on the Data to MC agreement is applied randomly to eliminate a fraction of jets in the simulation sample. The variation obtained on the cross section is taken as a systematic uncertainty.

4.13.3 Jet energy resolution (JER)

In order to take into account the jet energy resolution [120], the energy of each reconstructed jet in the MC simulation is smeared by a Gaussian function such that the width of the resulting Gaussian distribution includes the uncertainty on the JER. The jet resolution is measured with two different methods: the di-jet balance and the bi-sector techniques. Due to the good agreement found between data and MC simulated events for the jet energy resolution, the jet smearing is applied only to extract the corresponding systematic uncertainty.

[5] N_{PV} is a good measure of the number of proton collisions in a event.
[6] μ is a variable sensitive to the out-of-time pile-up activity.

4.13.4 Trigger efficiency

The five jets trigger efficiency is derived from the ratio between the number of events passing the trigger selection and the cut on the fifth-jet transverve momentum $p_T > 55$ GeV with respect to the events collected with the same trigger. The associated efficiency ranges from 90% to 100%, so a conservative 10% systematic uncertainty is assigned to the trigger turn-on curve.

4.13.5 *b*-tagging efficiency and mistag rate

The *b*-tagging algorithm used to identify jets originating from *b*-quarks can have different performance on data and simulation; to take in account this difference in the efficiency in the two samples, a set of scale factors parametrized as a function of transverse momentum p_T and pseudorapidity η were applied to *b*-, *c*- and light-jets. To each jet an efficiency or an inefficiency scale factor is associated depending on whether it was *b*-tagged or not. These scale factors were varied individually within their maximal associated uncertainty and propagated to the analysis. The variation on the cross section gives the systematic uncertainty.

4.13.6 Theoretical Uncertainty

Theoretical uncertainties also limit our ability to estimate the signal efficiency.

4.13.6.1 Initial and Final State Radiation (ISR and FSR)

The effect of the variation of the amount of the initial and final state radiation (ISR/FSR) were studied using the leading order ACERMC [121] generator interfaced to PYTHIA by varying the parameter controlling ISFR and FSR in the range consistent with experimental data. Two ACERMC samples were generated with less or more parton shower activity (ps). The systematic uncertainty is defined as the half of the difference between the less and more ps samples.

4.13.6.2 Parton Distribution Function (PDF)

The signal samples are generated usign the CTEQ10 (Coordinated Theoretical-Experimental Project on QCD) [103] proton parton distribution functions, PDFs. These PDFs, obtained from experimental data, have an uncertainty that is represented in 22 pairs of additional PDF sets provided by the CTEQ group. To evaluate the impact of the PDF uncertainty on the signal templates, the events are reweighted with the corresponding ratio of PDFs, and 22 pairs of additional signal templates are constructed. The uncertainty is calculated as half of the quadratic sum of differences

of 22 pairs. Two other PDF sets were considered as well, MWST2008nlo68cl [122] and NNPDF20 [123]. The maximum uncertainty is then considered as PDF uncertainty on the cross section value.

4.13.6.3 Parton shower and generator uncertainties

The uncertainty related to the default MC simulation used to generate $t\bar{t}$ events is estimated by the comparison with another MC generator. The uncertainties on the signal modeling arise from the choice of the event generator used to simulate events. They are evaluated by comparing the MC@NLO prediction of $t\bar{t}$ kinematics and acceptance to the one from POWHEG [124]. The uncertainty from the parton shower simulation is determined by comparing POWHEG interfaced to HERWIG [121] or PYTHIA [125].

4.13.7 Luminosity

The uncertainty on the integrated luminosity has an effect only on signal acceptance, it is due to the beam intensity uncertainty (n^2 in the Eq. 2.1) [126]. It propagates linearly to the cross section measurement, leading to a systematic uncertainty of 3.9% [127].

4.13.8 Background modelling

The background modelling for the top quark mass is derived, as described in Section 4.8, from the shape of the muti-jet events in an untagged data sample. The assumption made was validated using a multi-jet sample generated by the ALPGEN MC program. After evaluating the order of the difference between the tagged and the untagged top quark mass distribution on the simulated event, and quantifying the difference coming from the generic multi-jet QCD production and the exclusive $b\bar{b} + jets$ production, a correction factor is computed to take into account these two effects on the fitted top quark mass in the background region. Figure 4.28 shows the top quark mass distribution with and without the correction derived on ALPGEN MC simulation. The background uncertainty due to the background shape modelling is defined as the maximum variation between the nominal cross section and the corrected one. It is estimated to be of order of 4.1%.

4.14 Summary of systematic uncertainties

The break down of the systematic uncertainties obtained for the fully hadronic $t\bar{t}$ analysis, presented in this Section, are summarized in Table 4.5.
The main contributions to the total systematic uncertainty come from the jet energy scale, the b-tagging and the ISR/FSR uncertainties.

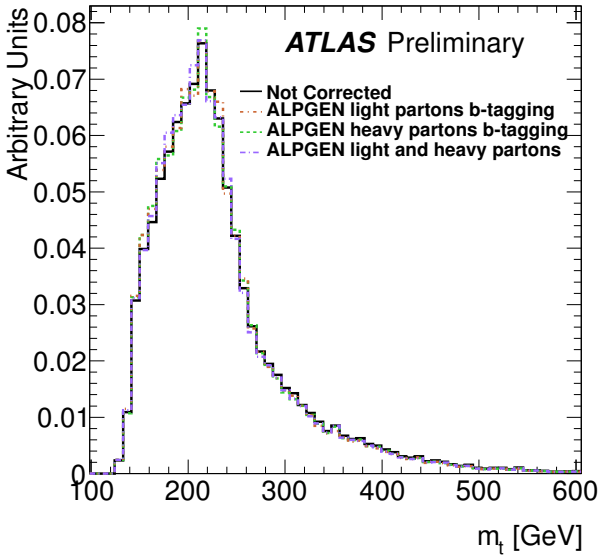


FIGURE 4.28 The distributions after applying different corrections derived from Monte Carlo, as explained in Section 4.8, are overlaid.

Source of uncertainty	$\Delta\sigma/\sigma$ (%)
Jet energy scale (JES)	$^{+20}_{-11}$
b-tagging	$\pm$ 17
ISR, FSR	$\pm$ 17
Parton Shower	$\pm$ 13
Multi-jet trigger	$\pm$ 10.0
Generator	$\pm$ 7
PDF	$^{+7}_{-4}$
Pile-up	$^{+5}_{-7}$
Background model	$\pm$ 4.
Luminosity	$\pm$ 4
Jet energy resolution	$\pm$ 3
Jet reconstruction efficiency	< 1
Total	$^{+36}_{-34}$

TABLE 4.5 Summary of the different systematic uncertainties associated with the kinematic fit analysis using the selected data events, $t\bar{t}$ MC signal and data-driven background estimated events. Uncertainties are given in %. The asymmetric uncertainties are derived using [128].

4.15 Final results for $t\bar{t}$ cross section measurement in the fully hadronic final state

The $t\bar{t}$ production cross section in the fully hadronic decay channel at the LHC has been measured with the ATLAS detector at a centre of mass energy of $\sqrt{s} = 7$ TeV and with an integrated luminosity of 4.7 fb^{-1} recorded in 2011. The cross section is extracted using a kinematic fit technique. The measured value is:

$$\sigma(pp \to t\bar{t}) = 168 \pm 12 \text{ (stat.)} \, ^{+60}_{-57} \text{ (syst.)} \, \pm 7 \text{ (lum.) pb} \tag{4.12}$$

The measurement is compatible with the Standard Model expectation of $\sigma^{t\bar{t}}_{SM} = 167^{+17}_{-18}$ pb. The result obtained is limited by the systematic uncertainty. The main contributions to the total systematic uncertainty come from the jet energy scale, the b-tagging and the modeling of the initial and final state radiation. Several studies have been carried out to improve the jet energy scale calibration and the ISR/FSR modeling, as well as the analysis framework in order to decrease the impact of the systematic uncertainties in the full hadronic $t\bar{t}$ cross section result.
An improvement of the measurement could be achieved by requiring a different event selection where the b-jet identification is implemented already at the trigger level. The b-jet trigger represents in fact a crucial ingredient especially for the $t\bar{t}$ cross section in the fully hadronic mode measurement because it helps in enhancement of the signal efficiency and could reduce the impact of the QCD background contribution. The trigger has bee implemented successfully for the proton proton collision collected during the 2012 data taking period at $\sqrt{s} = 8$ TeV.

Chapter 5

Search for associated Higgs boson production together with $t\bar{t}$ pairs

This Chapter focuses on the discussion of the search for the Higgs boson associated production channel, $t\bar{t}H$, subsequently decaying in $b\bar{b}$ pairs. This analysis is presented as an extension of the fully hadronic $t\bar{t}$ cross section analysis presented in Chapter 4. Both the pre-selection strategy and the event topology reconstruction are taken from the previous analysis. A study of the purity of $t\bar{t}$ and $t\bar{t}H$ system reconstruction is also presented. For this analysis, a boosted decision tree is used in order to discriminate between the $t\bar{t}$ and QCD multi-jet production, details on its implementation and performance are discussed thoroughly. The estimation of the QCD multi-jet background is performed using a data driven technique, so called "ABCD" method, the basic idea of the method as well as the application in the $t\bar{t}H$ analysis are described also in this Chapter.

5.1 Motivation

The dominant Higgs boson production mechanism is the gluon-gluon fusion and the Higgs boson decay mode with the highest branching ratio in the low mass region is $H \to b\bar{b}$ (see Section 1.1.4.1 and Section 1.1.4.3), unfortunately the $gg \to H \to b\bar{b}$ process is very difficult to detect due to the huge QCD background it suffers from. For this reason the Higgs to $b\bar{b}$ decay channel is normally investigated in the VBF Higgs production channel, $qq \to qqH \to qqb\bar{b}$, in the VH associated production, $qq \to VH \to Vb\bar{b}$, where V is one of the two vector bosons $W^{\pm}$ and Z, and with a top quark pair, $qq \to t\bar{t}H \to t\bar{t}b\bar{b}$. In this Chapter the $t\bar{t}H$ ($H \to b\bar{b}$) channel is investigated in one particular final state where the two W bosons from the top quark decay, disintegrate exclusively into two quarks. The $t\bar{t}H$ analysis is also aimed at understanding the electroweak sector of the SM, and gives the possibility to perform a direct measurement of the top and bottom quark Yukawa couplings to the Higgs boson (see Section 1.1.4.2).

5.2 Previous results on the search for $t\bar{t}H$ ($H \to b\bar{b}$)

The ATLAS collaboration has already performed a search for the $t\bar{t}H$ production with a $H \to b\bar{b}$ using an integrated luminosity of 4.7 fb^{-1} collected in 2011 at $\sqrt{s} = 7$ TeV [117]. This search is performed in the $t\bar{t}$ semi-leptonic final state. The typical signature of the semi-leptonic $t\bar{t}H$ production was characterized by the presence of one high transverse momentum isolated lepton (e/μ), high transverse missing energy and six jets, four of which originating from b-quarks. Nine exclusive topologies are considered. Four topologies are referred as signal regions characterized by high jet and b-jet multiplicity. Five topologies, dominated by background contributions are used in order to constraint the systematic uncertainties in the fit procedure. No significant events excess with respect to the background expectation was observed and 95% confidence level (C.L.) upper limit on the production cross section $\sigma(t\bar{t}H) \times BR(H \to b\bar{b})$ for a Higgs boson mass range are provided, in particular for $M_H = 125$ GeV. The observed and expected 95% C.L. upper limits obtained are 13.1 and 10.5 times the SM cross section respectively.
The most recent result is presented by the CMS collaboration which uses an integrated luminosity of 19.5 fb^{-1} at $\sqrt{s} = 8$ TeV [129]. Two $t\bar{t}$ decay mode are considered in the analysis: semileptonic and dileptonic. The search for the $t\bar{t}H$ has been performed in two Higgs boson decay modes: $H \to b\bar{b}$ and $H \to \tau\bar{\tau}$. Multivariate techniques are used in order to discriminate between the background and signal events. The combination of the three channels gives an observed (expected) 95% C.L. limit on the $\sigma(t\bar{t}H)$ of 5.2 (4.1) times the SM expectation for a Higgs boson mass hypothesis of $M_H = 125$ GeV. The best fit for the signal strength is $\mu = 0.85^{+2.47}_{-2.41}$ (68% C.L.).
Until now no results on the fully hadronic decay mode of the $t\bar{t}$ system has been shown by the two LHC collaborations. This thesis presents a first analysis performed with the ATLAS detector on the search for Higgs boson produced in association with a top-anti top pair in the fully hadronic decay mode.

5.3 Background contributions

The background sources can be divided into two categories: irreducible and reducible backgrounds. The former includes all the processes which share the same topology with the signal, so it is not possible to eliminate these contributions; instead the latter includes all the processes which are characterized by different object contributions in the final state with respect to the signal. In the $t\bar{t}H$ ($H \to b\bar{b}$) analysis the irreducible background includes all the processes where there are at least eight jets in the final state among which four b-jets. Such topology can be found in the $t\bar{t}b\bar{b}$ production as well as in the $t\bar{t}Z$ ($Z \to b\bar{b}$) process. While the $t\bar{t}Z$ background can be ignored due to it very small production cross section (see Table 5.3), the $t\bar{t}b\bar{b}$ with a predicted production cross section of about 1.6 pb, represents one of the most important irreducible backgrounds. On the other hand, the reducible backgrounds are for example the top pair produced in association with

light jets: a final state characterized by only two b-jets.
In summary the main background for the $t\bar{t}H$ process comes from the $t\bar{t}$ production with at least two extra jets. If the two extra jets are b-jets the background signature is exactly the same as the signal. The $t\bar{t}b\bar{b}$ final states is produced via the EW and QCD processes. Some three level Feynman diagrams contributing to these two processes are shown in Figure 5.1(a) and Figure 5.1(b). The

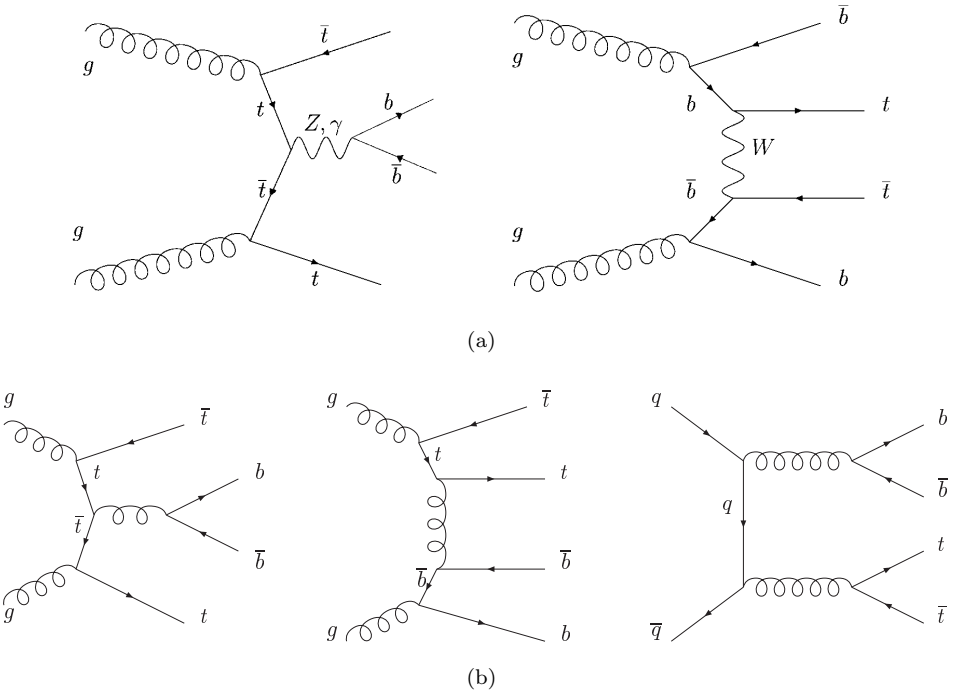


FIGURE 5.1 Some of the tree level Feynman diagrams for $t\bar{t}b\bar{b}$ production via EW 5.1(a) and QCD 5.1(b).

contribution of the processes in which the extra jets are c-jets or light-jets could be reduced using the b-tagging. The shape template as well as the normalization of the $t\bar{t} + jets$ sample relies on MC simulation.

5.4 Data and MC simulation samples

The analysis is performed on the data recorded by the ATLAS detector during the 2011 data-taking period, corresponding to an integrated luminosity of 4.7 fb^{-1} (see Section 4.2). The $t\bar{t}H$ signal sample is generated with PYTHIA 6.425 leading order generator [125], the MRST LO**PDF [130] and AUET2B LO** underlying event tune for a Higgs boson mass $M_H = 125$ GeV. The sample is generated assuming a top quark mass of 172.5 GeV. The $t\bar{t}H$ sample is normalized using the NLO theoretical cross section and a Higgs boson decay branching ratio to $b\bar{b}$ of $BR(H \to b\bar{b}) = 0.57$ [131]. The cross section and the branching ratios for a Higgs mass $M_H = 125$ GeV and a top mass

$m_t = 172.5$ GeV are summarized in Table 5.1.

The $t\bar{t}$+jets background events are generated using the ALPGEN v2.13 [107] LO generator and

Channel	Cross Section (pb)	BR($H \to b\bar{b}$)	BR($t\bar{t} \to b\bar{b}q_1q_2q_3q_4$) [%]
$t\bar{t}H$	0.106	0.57	0.44

TABLE 5.1 $t\bar{t}H$ cross section at $\sqrt{s} = 7$ TeV for a Higgs mass $M_H = 125$ GeV and a top mass $m_t = 172.5$ GeV, branching ratio of Higgs decaying in a bottom-anti bottom pair and all hadronic decay of $t\bar{t}$ pair branching ratio are also indicated [131].

the CTEQ6L1 PDF set [102]. Parton shower and fragmentation are modeled with Herwig v6.520 [105]. A parton-jet matching scheme ("MLM matching") [132] is used to avoid double-counting of partonic configurations generated by both matrix-element calculation and the parton-shower evolution. The $t\bar{t}$+jets background samples are generated separately for light-jets (u, d, s quarks and gluons), b-jets and c-jets. A selection criteria is applied on the heavy flavour composition in the $t\bar{t}$+jets samples to check the exclusion production of each sample. In the ALPGEN samples there is no flavour matching, which means that the $b\bar{b}$ ($c\bar{c}$) can arise either from the parton shower or from the matrix-element calculation leading to a double counting of the events. The overlap between the heavy flavour (b and c) events generated from the matrix-element calculation and those generated from parton-shower evolution in the $t\bar{t}$+jets samples is avoided via an algorithm, heavy flavour overlap removal, based on the angular separation between the extra heavy quarks: if $\Delta R(q, \bar{q}) > 0.4$, the matrix-element prediction is used, otherwise the parton-shower prediction is used. These samples are generated assuming a top quark mass of 172.5 GeV and are normalized to the approximate next-to-next-leading order (NNLO) theoretical cross section [133] using the MSTW2008 NNLO PDF set [122]. The cross sections for these productions are presented in Table 5.2. Other background processes which have similar signature in the final state as $t\bar{t}H$ are

Physics process	$\sigma \times BR_{t\bar{t}\to allhad}$ [pb]
$t\bar{t}b\bar{b}$	1.59
$t\bar{t}c\bar{c}$	3.06
$t\bar{t}$+ light jets	76.19

TABLE 5.2 Cross sections at $\sqrt{s} = 7$ TeV for $t\bar{t}$+jets (light and heavy flavoured jets) productions [133].

the $t\bar{t} + V$ ($V = Z, W$), where the W decays into c- and $\bar{s}$-quarks and Z goes to $b\bar{b}$ or $c\bar{c}$. However the expected contributions of these productions are negligible due to their low cross sections, see Table 5.3.

Physics process	$\sigma \cdot BR_{t\bar{t}\to allhad}$ [pb]
$t\bar{t}W$	0.1244
$t\bar{t}W + jets$	0.0835
$t\bar{t}Z$	0.0956
$t\bar{t}Z + jets$	0.0816

TABLE 5.3 Cross sections at $\sqrt{s} = 7$ TeV for $t\bar{t}V(+jets)$ ($V = Z, W$) productions [133].

For the same reason also the $W/Z + jets$ production are negligible and will not be taken into account. All event generators using Herwig are also interfaced to Jimmy v4.31 [134] to simulate the underlying events. All simulated samples use Photos 2.15 [135] to simulate photon radiation and Tauola 1.20 [136] to simulate τ decays. Finally, all simulated samples include multiple pp interactions and are processed through a simulation of the detector geometry and response using GEANT4 [137]. For the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) signal, two samples are produced with different detector simulations:

- full simulation (FullSim): this is the official setup used in the collaboration where the pp interaction are processed with a very detailed simulation of the detector geometry and response with GEANT4,
- fast simulation (AFII): the pp interactions are processed considering a parametric cell response of the calorimeter and a simplified geometry model which keeps the exact description of the sensitive elements. This fast simulation requires less CPU and computing times giving the possibility to generate samples with high statistics.

The analysis is performed mainly on the full simulation but due to the low statistics available in the FullSim sample (30 M events in FullSim instead of 600 M events in the AFII), AFII sample is used in the training steps of the multivariate analysis (Section 5.9).
All simulated samples are processed through the same reconstruction software as the data. Simulated events are corrected so that the object identification efficiencies, energy scales and energy resolutions match those determined in data control samples.

5.5 The fully hadronic $t\bar{t}H(H \to b\bar{b})$ event topology

Three possible final states can be investigated with the $t\bar{t}H$ production channel with the Higgs decaying into $b\bar{b}$ pair. They correspond to the three different decay modes of the $t\bar{t}$ system: dileptonic, semileptonic and fully hadronic. Considering the work and time invested to perform the $t\bar{t}$ cross section measurement in the fully hadronic mode and to utilize the knowledge acquired on this particular signature, the $t\bar{t}H$ in the all-hadronic final state $t\bar{t}H \to q\bar{q}'q\bar{q}'b\bar{b}$ is investigated. This process can be visualized in the Feynman diagram depicted in Figure 5.2. The fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) final state is characterized by at least eight jets among which four are b-jets, two b-jets coming from the Higgs boson decay and two from the top quark via $t \to Wb$ decay. The analysis of this multi-jet final state channel is very challenging, since, like in the $t\bar{t}$ cross section analysis shown in Chapter 4, the multi-jet background contribution is overwhelming. For this reason all the events selection and reconstruction technique must be improved in order to reduce as much as possible the fraction of this background. A careful study of the multi-jet background is necessary to obtain a good modelling for it.

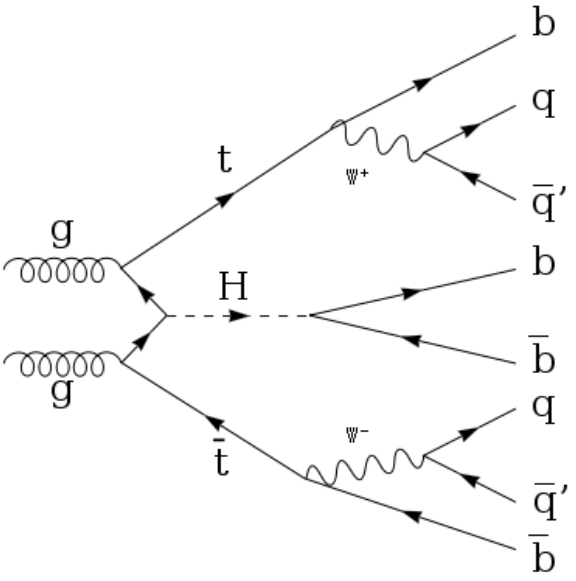


FIGURE 5.2 One of the most important $t\bar{t}H(H \to b\bar{b})$ tree level Feynman diagrams for the all-hadronic final state.

5.6 Event Selection

The dataset used for this analysis is the same as the one used for the $t\bar{t}$ cross section measurement (see Section 4.3), together with the object definition and the veto on the isolated leptons (Section 4.4.2). The events are collected with a multi-jet trigger asking for the presence of at least five jets with transverse momentum above 30 GeV at EF trigger level. The jets considered in the analysis must have a transverse momentum above 25 GeV and be within a pseudorapidity range of $|\eta| < 2.5$. With respect to the $t\bar{t}$ cross section analysis a lower p_T threshold is applied on the fifth leading reconstructed jets, 45 GeV instead of 55 GeV. Furthermore no requirement on the jets isolation[1] is added. By comparing the events yield with and without the higher jet p_T and jet isolation requirements in Table 5.4 and Table 5.5 respectively, one can notice the large increase in the efficiency for the signal events (a factor of 3.25) by removing the highest jet selection. It should be reminded that the high p_T cut on the fifth leading jets as well as the $\Delta R > 0.6$ cut were used in the cross section measurement in order to reach a good data to MC simulation agreement in the multi-jet trigger turn on curve (see Figures 4.6 and 4.7). An ongoing ATLAS study on the trigger turn-on curve shows that the correction factors needed to absorb the data to MC simulation difference are of the order of unity. Thus their impact on the analysis is considered as negligible. Considering the presence of four b-quarks in the final state, two of them originating from the top quarks decay and the other two from the Higgs boson decay, at least four of the selected jets should be identified as b-jets by the b-tagging algorithm MV1, which is detailed in Section 4.4.3. The b-tagging working point used in the analysis provides a 60% efficiency on the b-jets and a light

[1] This requirement is applied in the $t\bar{t}$ cross section measurement.

jets rejection of about 500.

The signal region in this analysis requires the presence of at least eight jets in the event and includes two exclusive b-tagging multiplicity subcategories identified for exactly three b-tagged jets and at least four b-tagged jets. A validation region, dominated by the background contribution, is also considered by asking for the presence of exactly two b-tagged jets. It could also happen that the signal topology is not completely reconstructed and therefore not all the eight jets are present in the final state. For this reason the low jet multiplicity bins, in particular six and seven jets, must be taken into account in order to have complete overview of the signal topology and to gain in the final sensitivity of the analysis. The following sections go through the inclusive eight jets measurement and will not touch the low jets multiplicity analysis strategy which is currently not fully developed yet.

Selection	$t\bar{t}H$	$t\bar{t}qq$	$t\bar{t}b\bar{b}$	$t\bar{t}c\bar{c}$	data
Total Events	131	359146	7512	14425	65215848
Preselection & Trigger	72 ± 6	73773 ± 240	1322 ± 30	2125 ± 40	12132491
Heavy flavour removal	72 ± 6	71906 ± 240	1110 ± 31	1750 ± 39	12132491
$\Delta R(j,j) > 0.6$	38 ± 5	48848 ± 205	721 ± 30	1183 ± 30	8078776
≥ 5 jets $p_T > 55$ GeV	20 ± 4	17654 ± 130	259 ± 20	414 ± 20	1420963
≥ 8jets $p_T > 25$ GeV	7.2 ± 3	2817 ± 53	56 ± 7	77 ± 9	66492

TABLE 5.4 Event yield in data and MC simulation. The latter is normalized to the luminosity in data (4.7 fb^{-1}) using the respective cross sections. The jet isolation criteria is applied to the event selection. Only the events with at least five jets with $p_T > 55$ GeV in the final state are considered. Uncertainties are statistical only.

Cut	ttH	ttqq	ttbb	ttcc	data
Total Events	131	359146	7512	14425	65215848
Preselection & Trigger	72 ± 6	73773 ± 242	1322 ± 35	12125 ± 40	12132491
Heavy flavour	72 ± 6	71907 ± 240	1096 ± 30	1753 ± 40	12132491
≥ 5 jets $p_T > 45$ GeV	65 ± 6	56442 ± 220	863 ± 30	1357 ± 35	1046147
≥ 8jets $p_T > 25$ GeV	29 ± 4	11384 ± 100	226 ± 15	308 ± 20	447431

TABLE 5.5 Event yield in data and MC simulation. The latter is normalized to the luminosity in data (4.7 fb^{-1}) using the respective cross sections. No requirement on the jet isolation is considered. Lower p_T (45 GeV) threshold is applied on the fifth leading jets. Uncertainties are statistical only.

5.7 $t\bar{t}H$ topology reconstruction

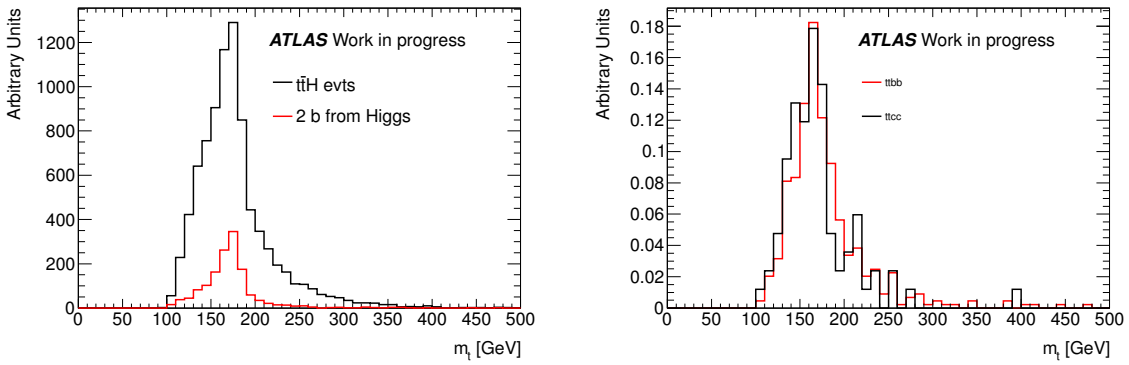
After the event selection, it is necessary to reconstruct the $t\bar{t}H$ final state topology. In order to not introduce a bias on the modeling of the invariant mass of the two jets considered as Higgs decay products candidates, which represents the discriminant variable that is ultimately used to estimate the $t\bar{t}H$ ($H \to b\bar{b}$) sensitivity, the reconstruction procedure tries to find the correct association

of the jets with the final state partons only for the $t\bar{t}$ system without using any information on the Higgs decay products. Then, after the reconstruction of the $t\bar{t}$ system, the jets left out from the top pair reconstruction are considered as candidates for the Higgs boson decays products. In particular in this analysis the two highest p_T jets among the jets not used to reconstruct the $t\bar{t}$ pair, are used to compute the invariant mass of the Higgs candidates.
The technique used for this purpose is a kinematic likelihood fit. As discussed previously in Section 4.7, the kinematic fit method is based on a likelihood approach (KLF). The KLF tests all the possible combinations of six jets among the jets in the event taking into account the invariances which characterize the $t\bar{t}$ hadronic system in order to avoid double combination counting and to optimize the CPU time. On each combination a fit is performed. The combination which minimizes the likelihood function is the one used for the $t\bar{t}$ system reconstruction and is considered as the "best" combination.
Several options are available to take into account the b-tagging information. Only two of them are considered in this analysis: vetoing of the combination in which a b-tagged jet is placed in the light parton position in the KLF structure, and the WorkingPoint option which consider the b-jet efficiency and the light-jet rejection corresponding to the b-tagging working point chosen. Some constraints[2] especially on the W boson are imposed in the likelihood. For the top quark two different implementations are considered:

- Floating top quark mass: the top quark pole mass, m_{top}^{reco}, is treated in the likelihood (Equation 4.5) as a free parameter of the fit. The top quark width[3] is constant,
- Fixed top quark mass: both the top quark pole mass $m_{top}^{reco} = 172.5$ GeV and its width are constants in the fit.

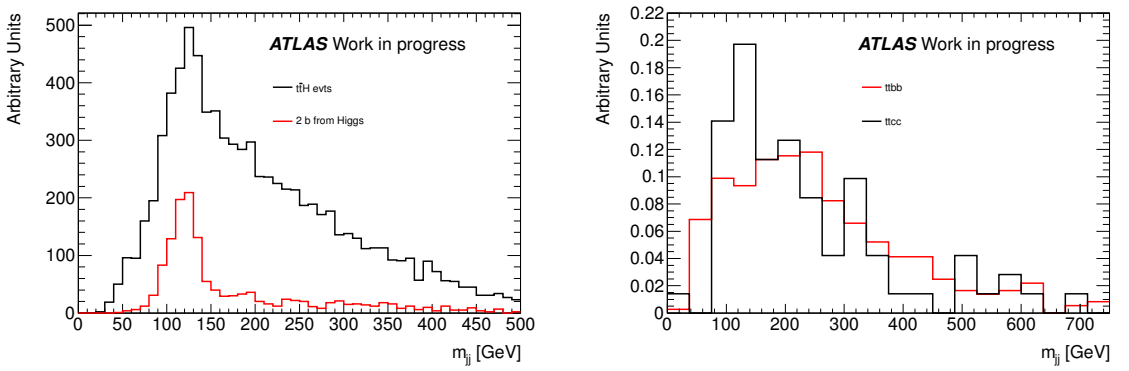
To choose which KLF setup, for different b-tagging and the top quark pole mass configuration, provides the best performance, the purity of the reconstruction is investigated. The purity is defined as the ratio between the number of events in which the jets used to reconstruct the $t\bar{t}$ and Higgs are matched to the truth level and either the total number of events or the number of events where all the jets are matched to $t\bar{t}H$ partons. The performance studies are conducted on the MC simulation samples; the $t\bar{t}H$ dataset as well as the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ are used. Only the events left after the selection presented in Section 5.6 are considered. The first step consists in the estimation of the number of events which have at least one or exactly two jets matched with b-quark(s) originating from the Higgs boson decay. Table 5.6 shows that the best purity is reached with the fixed top quark pole mass leading to about 76% (23%) for at least one (exactly two) jet(s) matched to the quark originating from the Higgs boson disintegration. Figure 5.3(a) compares the top quark mass reconstructed by the KLF in the $t\bar{t}H$ signal sample with (red curve) and without (black curve) the matching of exactly two jets with the Higgs boson decay. Figure **??** shows the top quark mass distribution for $t\bar{t}b\bar{b}$ (red histogram) and $t\bar{t}c\bar{c}$ background events. The invariant mass distribution

[2]The W boson mass and width are kept constant at the known values, 80.4 GeV and 2.1 GeV, respectively.
[3]The top quark width is fixed at 1.5 GeV.



(a) (b)

FIGURE 5.3 Distribution of the top quark mass reconstructed by the KLF where the top quark pole mass m_{top}^{reco} is treated as a free parameter in the fit procedure. Figure 5.3(a) shows the comparison in the fully hadronic $t\bar{t}H(H \to b\bar{b})$ MC sample of the events with at least eight jets and four b-tagged jets and the events in which exactly two jets are matched with the Higgs decays products at parton level. Figure **??** shows the top quark mass distribution in the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ MC samples for events with at least eight jets and four b-tagged jets.

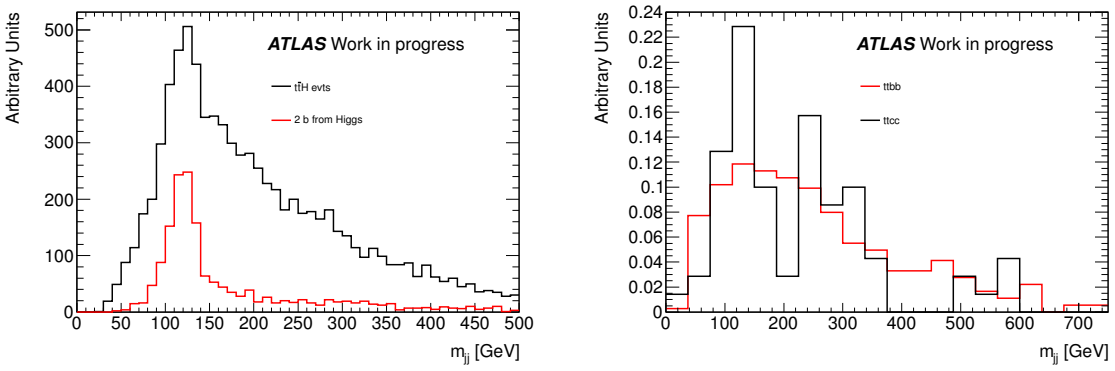


(a) (b)

FIGURE 5.4 Distribution of the invariant mass of the two jets not used in the KLF topology reconstruction where the top quark pole mass m_{top}^{reco} is treated as a free parameter in the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$ MC sample (Figure 5.4(a)) and in the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ MC background samples (Figure 5.4(b)).

of the two jets not considered in the KLF is shown in Figure 5.4(a) and Figure 5.5(a) where the top quark mass pole is treated as a free parameter and constant in the KLF, respectively. The same distribution is presented also for the $t\bar{t}b\bar{b}$ (red histogram) and $t\bar{t}c\bar{c}$ background sample, Figures 5.4(b) and 5.5(b).

A more restricted dataset which contains events where the eight jets in the final state are matched to the parton level is also considered (8 $jets_{match}$). Then the matching of at least (exactly) one (two) jet(s) to the b-quark(s) originating from the Higgs boson at parton level is performed. In this sample the highest purity for events with exactly two jets matched to the Higgs decay products is equal to 45% and is achieved by the KLF working point setup using a fixed top quark pole mass (Table 5.7).



(a) (b)

FIGURE 5.5 Distribution of the invariant mass of the two jets not used in the KLF topology reconstruction where the top quark pole mass m_{top}^{reco} is treated as a constant in the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) MC sample (Figure 5.5(a)) and in the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ MC background samples (Figure 5.5(b)).

To validate the KLF performance on the background processes, where no Higgs boson is present

	WorkingPoint Fix Mass (%)	WorkingPoint Floating Mass (%)	Veto Fix Mass (%)	Veto Floating Mass (%)
$\geq 1 b_{Higgs}$	71 ± 0.5	70 ± 0.5	76 ± 0.5	74 ± 0.5
$= 2 b_{Higgs}$	20 ± 0.4	17± 0.4	23 ± 0.4	19 ± 0.4

TABLE 5.6 Reconstruction purity studies on the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) MC sample. The study is performed on the events that passed the selections described in Section 5.6. The "$\geq 1\, b_{Higgs}$" ("$= 2\, b_{Higgs}$") label represents the events where at least one (exactly two) jets are matched at the parton level to a quark originating from the Higgs boson. Uncertainties are statistical only. The purity is defined as the ratio between the number of events in which the jets used to reconstruct the $t\bar{t}$ and Higgs are matched to the truth level and either the total number of events or the number of events where all the jets are matched to $t\bar{t}H$ partons.

	WorkingPoint Fix Mass (%)	WorkingPoint Floating Mass (%)	Veto Fix Mass (%)	Veto Floating Mass (%)
$\geq 1 b_{Higgs}$	91 ± 0.7	86 ± 0.8	85 ± 0.8	81 ± 1
$= 2 b_{Higgs}$	45 ± 1	37 ± 1	41 ± 1	32 ± 1

TABLE 5.7 Reconstruction purity studies on the fully hadronic $t\bar{t}H(H \to b\bar{b})$ MC sample. The study is performed on the events that passed the selections described in Section 5.6. The "$\geq 1\, b''_{Higgs}$ ("$= 2\ b_{Higgs}$") label represents the events where at least one (exactly two) b-jets are matched at the parton level to a quark originating from the Higgs boson. The purity fractions are with respect to the number of events were the $t\bar{t}H$ system is fully reconstructed (20 ± 0.4% of all MC $t\bar{t}H$ events after the pre-selection. Uncertainties are statistical only.

and therefore can not be used to quantify the purity, a subsample of events where all six jets are matched to the top quarks partons, "6 $jets_{MatchToTop}$", is considered. To quantify the purity of the algorithm this number is compared to the number of events where the six jets selected by the KLF are fully matched to the top quark partons, "6 $jets^{KLF}_{MatchToTop}$". In this way the purity of the reconstruction of the $t\bar{t}$ system topology is compared in the signal sample (Table 5.8) and in the

	WorkingPoint Fix Mass (%)	WorkingPoint Floating Mass (%)	Veto Fix Mass (%)	Veto Floating Mass (%)
6 $jets_{MatchToTop}$	27 ± 0.5	27 ± 0.5	27± 0.5	27 ± 0.5
6 $jets^{KLF}_{MatchToTop}$	33 ± 1	26 ± 1	23 ± 1	18 ± 1

TABLE 5.8 Reconstruction purity studies on the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) sample. The study is performed on the events that completely passed the selections explained in the Section 5.6."6 $jets_{MatchToTop}$" correspond to events in which six jets are matched at the parton level to the six quarks originating from the top quark decay and is computed with respect to all the events left after the selection. Instead "6 $jets^{KLF}_{MatchToTop}$" corresponds to events where the jets used in the kinematic fit are matched to the six quarks of the top decay products and is computed with respect to the number of events where the $t\bar{t}$ system is fully reconstructed.

background one (Table 5.9). Figure 5.6(a) and Figure 5.6(b) show the top quark mass distributions for $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ background samples respectively for the events where the $t\bar{t}$ system is completely resolved by the KLF (six jets used in the topology definition are matched to truth level) and for the events where the topology is not resolved (at least one jet is not matched to a quark originating from the top decay). The top quark mass distribution in "6 $jets^{KLF}_{MatchToTop}$" configuration shows a more narrower peak around the top quark mass value.

To summarize, the purity study asserts that the b-tagging setting with the best performance for the KLF is the WorkingPoint. Concerning the top quark pole mass parameter m^{reco}_{top}, both setups are used in the analysis for different purposes:

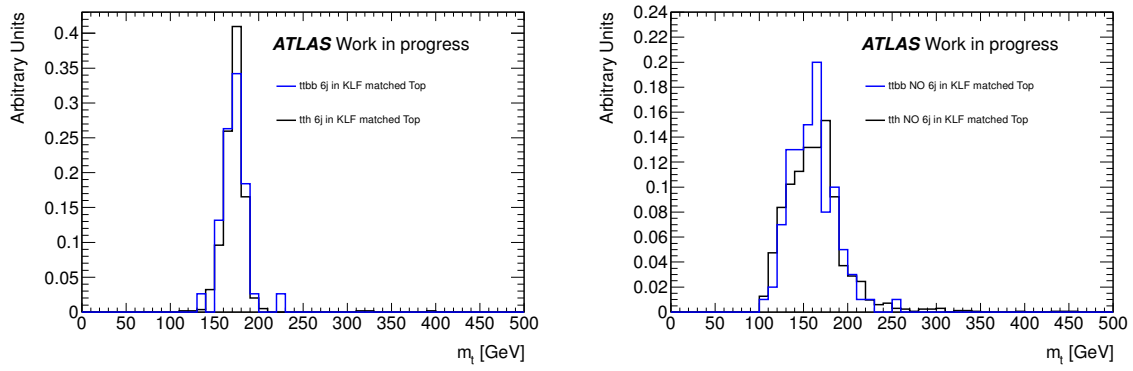
- KLF with the m^{reco}_{top} as a free parameter is used to discriminate the top quark process against the multi-jet production,
- KLF with the m^{reco}_{top} constant parameter is used to completely solve the $t\bar{t}$ system and to identify the two Higgs boson jets candidates.

	WorkingPoint Fix Mass (%)	WorkingPoint Floating Mass (%)	Veto Fix Mass (%)	Veto Floating Mass (%)
		$t\bar{t}b\bar{b}$		
6 $jets^{KLF}_{MatchToTop}$	31 ± 4	28 ± 4	17± 3	14± 3
		$t\bar{t}c\bar{c}$		
6 $jets^{KLF}_{MatchToTop}$	26 ± 7	29 ± 8	13 ± 6	13± 6

TABLE 5.9 Purity studies for reconstructing top pairs in $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ samples. The "6 $jets^{KLF}_{MatchToTop}$" label is the percentage of those events where the jets used in the kinematic fit are matched to the six quark of the top decay products and is computed to the number of events where the $t\bar{t}$ system is fully reconstructed which corresponds to 31% (37%) of all MC $t\bar{t}b\bar{b}$ ($t\bar{t}c\bar{c}$) events after the pre-selection. Uncertainties are statistical only.

5.8 QCD multi-jets background estimation: "ABCD" method

An important physics background that has to be considered in the analysis is the QCD multi-jet background. The MC simulation is not considered able to predict it with the necessary accuracy,



(a) (b)

FIGURE 5.6 Distribution of the top quark mass reconstructed by the KLF where the top quark pole mass m_{top}^{reco} is treated as a free parameter in the fit procedure. Figure 5.6(a) shows the comparison between the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ MC background samples for events where the six jets matched at parton level to the top quark decay products used by the output of the kinematic fit. Figure 5.6(b) shows the distribution in the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ MC background samples for the events where the six jets are matched at the parton level are not used in the KLF. Uncertainties are statistical only.

for this reason a data-driven technique is the only sensible choice to estimate it. The method we used to predict the shape and the normalization of the QCD multi-jet background is the so called "ABCD" method. The basic idea of the "ABCD" method is to choose two uncorrelated variables in order to divide data events into four regions: control and signal regions. The control regions are used to estimate the contribution of the background in the signal one.

For example, calling the two observables *Obs*1 and *Obs*2, it is possible to define four regions: A, B, C and D as shown in Figure 5.7.

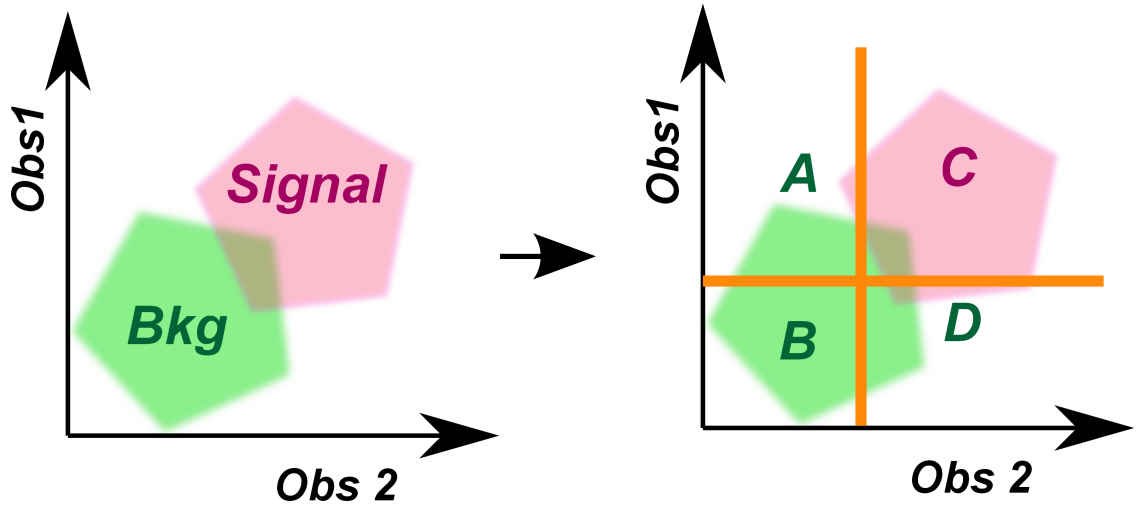


FIGURE 5.7 "ABCD" method. *Obs*1 and *Obs*2 are two uncorrelated observables used to subdivide the sample in signal and background regions.

Assuming that the signal is characterized by large values of *Obs*1 and *Obs*2, the region C identifies the signal region, instead the regions A, B and D are background-like regions, characterized by

different fraction of background and signal contibutions. With the "ABCD" method, it is possible to estimate the background contribution in the signal region C. The number of expected QCD background events in the background dominated regions is defined as the number of events in the data minus the number of events, $N_i^{other-bkg}$, from other background processes estimated from MC simulation (i.e. $t\bar{t}$+ jets).

$$N_i^{QCD-bkg} = N_i^{data} - N_i^{other-bkg}, \qquad i = A, B, D. \tag{5.1}$$

The value $N_i^{other-bkg}$ is estimated from the MC simulation using the theoretical cross section corresponding to the considered physics background. The template for QCD multi-jets background distribution x is constructed in the same way by subtracting the other background distributions obtained by the MC simulation from the data distribution. For each bin j of the kinematic variable x, the QCD multi-jet background events can be expressed via Equation 5.2:

$$\frac{dN_{i,j}^{QCD-bkg}}{dx_j^i} = \frac{dN_{i,j}^{data}}{dx_j^i} - \frac{dN_{i,j}^{other-bkg}}{dx_j^i}, \qquad i = A, B, D. \tag{5.2}$$

Considering the relation that connects the background fractions in the four regions $\frac{N_C^{QCD-bkg}}{N_D^{QCD-bkg}} = \frac{N_A^{QCD-bkg}}{N_B^{QCD-bkg}}$, it is possible to obtain the expected background event for each bin of a given variable x in the signal region C:

$$\frac{dN_{j,C}^{QCD-bkg}}{dx_j^C} = \frac{dN_{j,A}^{QCD-bkg}}{dx_j^A} \frac{\sum_j N_{j,D}^{QCD-bkg}}{\sum_j N_{j,B}^{QCDbkg}} \tag{5.3}$$

The only region which contributes to the modeling of QCD multi-jet kinematic variables is the region A, while the regions B and D provide the normalization of the QCD background in the signal region. In this analysis one of the two uncorrelated variables used for the "ABCD" method is the number of the b-tagged jets. The second variable can be either a kinematical variable, such as the top quark mass reconstructed by the KLF, or the output variable of a multivariate analysis, where different kinematic variables can be used in order to improve the discrimination power. The choice made in the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis is to put in place a multivariate analysis, see Section 5.9. The aim of the multivariate analysis is therefore two fold: in one hand it allows to discriminate $t\bar{t}$-like events from the QCD-like ones, on the other it allows to predict the QCD contribution in the signal region. So calling for the moment the second variable still $Obs2$, the data sample can be subdivided in four bins in the b-jet multiplicity and in two bins in the $Obs2$. The four bins considered in the b-tagged jet multiplicity are either exactly one, two, three or inclusive four, while an optimization is performed to choose the best value (α) to be used when cutting on $Obs2$. Table 5.10 summarizes the eight regions. In this analysis there are two signal regions: E and G, whereas the region C is the validation region. In the region C the shape of the QCD background

	$Obs2 > \alpha$	$Obs2 < \alpha$
$n_{b-jets} = 1$	A	B
$n_{b-jets} = 2$	C	D
$n_{b-jets} = 3$	E	F
$n_{b-jets} \geq 4$	G	H

TABLE 5.10 Setup for the "ABCD" method in the fully hadronic $t\bar{t}H(H \to b\bar{b})$ analysis. The definition of the eight exclusive regions uses two uncorrelated variables the b-jet multiplicity and the $Obs2$ (defined latter in this chapter).

is taken from the region A where there is the presence of exactly one b-tagged jet and $Obs2 > \alpha$, while for the normalization are used the regions B and D with exactly one and two b-tagged jets respectively and $Obs2 < \alpha$. The signal regions E and G ask for the presence of exactly three and at least four b-tagged jets respectively and $Obs2 > \alpha$. The QCD shape used in both signal regions are the same and are taken from the regions with exactly one and two b-tagged jets and $Obs2 > \alpha$ (A and C). The normalization in the case of the signal region E with exactly three b-tagged jets is estimated using the regions with exactly one, two and three b-tagged jets and $Obs2 < \alpha$ (B, D, F), instead in the case of the signal region G with at least four b-tagged jets, is estimated using the regions with exactly one, two and at least four b-tagged jets and $Obs2 < \alpha$ (B, D, H). The expected multi-jets background in the three signal regions (C, E, G) can be summarize using the Equation 5.4, Equation 5.5 and Equation 5.6 respectively.

$$\frac{dN_C^{bkg}}{dx_j^C} = \frac{dN_A^{bkg}}{dx_j^A}\frac{N_D^{bkg}}{N_B^{bkg}} \tag{5.4}$$

$$\frac{dN_E^{bkg}}{dx_j^E} = \left(\frac{dN_A^{bkg}}{dx_j^A} \cdot \frac{1}{N_B^{bkg}} + \frac{dN_C^{bkg}}{dx_j^C} \cdot \frac{1}{N_D^{bkg}}\right)\frac{N_F^{bkg}}{2} \tag{5.5}$$

$$\frac{dN_G^{bkg}}{dx_j^G} = \left(\frac{dN_A^{bkg}}{dx_j^A} \cdot \frac{1}{N_B^{bkg}} + \frac{dN_C^{bkg}}{dx_j^C} \cdot \frac{1}{N_D^{bkg}}\right)\frac{N_H^{bkg}}{2} \tag{5.6}$$

The estimation in the region C follows the prescription explained in the first example. The multi-jet background extraction in signal regions E and G is more complicated because the two regions are merged in order to predict the background in the signal region. A simple approach is then followed for the combination of the two regions, which consists in the relative average of the two contributions. A discussion of the systematic uncertainty connected to this assumption will be described in Section 5.11.6.

From the regions introduced in the "ABCD" method, the one where the contamination of the $t\bar{t}H(H \to b\bar{b})$ signal is the lowest is the exclusive one b-tagged jet bin. This region is used to model the multi-jet background. The model of the background in the control region is extracted by subtracting the simulated $t\bar{t}$+jets contribution from the data. To achieve this a weighting procedure is implemented in the data as well as in MC simulation consisting in simply weighting the data events by one and the $t\bar{t}$+jets by minus one of the expected number of events according

to the integrated luminosity ($weight_{tt+\text{jets}} = \sigma_{tt+\text{jets}} \cdot BR_{t\bar{t}\to allhad} \cdot L$). By doing so one can hope to mimic the whole kinematic of the QCD events. This assumption is validated in the strictly two b-tagged jets region.

5.9 Multivariate (MVA) technique

A multi-variate technique is used to discriminate potential signal events from background ones. This method is very useful in the case where single variables do not exhibit a clear separation power among signal and background. The MVA implemented in this analysis is the Boosted Decision Tree (BDT) [138]. A decision tree (DT) is a decision support tool that uses a tree-like model of decisions. Decision Tree is a machine learning algorithm based on the recursive growth of a tree-like structure in which an internal node represents a test on an attribute, each branch represents an outcome of the test and each leaf node represents a class label (decision taken after computing all attributes). A path from root to leaf represents classification rules. A typical MVA analysis consists of two independent phases: the training phase, where the multivariate methods are trained, tested and evaluated, and an application phase, where the chosen methods are applied to the data samples. For enhancing the classification and the performance, and for increasing the stability of statistical fluctuations, it is possible to reweight (boost) the decision tree. The Boosted procedure consists in the increasing of the weight of a misclassified event, i.e. signal event that lands on a background leaf or a background event lands on a signal leaf. The original weight obtained by the training of the first tree is then modified by a multiplicative factor, *boost weight* α, in the case of the misclassified events. The boost weight is derived by the misclassification rate, err, of the previous tree $\alpha = \frac{1-err}{err}$. This is the most popular boosting procedure, so called *Adaboost* [139]. Three different boosting setups are investigated in the analysis:

- *AdaBoost*: default boosting setup,
- *AdaBoost.R2*: similar to the *AdaBoost* with a redefined loss per event to account for the deviation of the estimated value from the true one,
- *Gradient*: uses a binomial log-likelihood loss, with respect to the exponential loss implemented in the *AdaBoost.R2*. The gradient of the loss function is calculated and the tree grows with the leaf values adjusted to match the mean value of the gradient in each region defined by the tree structure.

The BDT is implemented using the Toolkit for Multivariate Analysis (TMVA) [139] which provides a ROOT-integrated environment for the processing, parallel evaluation and application of multivariate classification technique. The software package consists of object-oriented implementation in C++/ROOT for each of these multivariate analysis (MVA) techniques.
The kinematic variables used as input for the training phase are:

- top quark mass reconstructed by the KLF using the m_{top}^{pole} as a free parameter,
- minimum invariant mass of jet pairs,
- average of the sum of all jets transverse energy, except the two leading jets, multiplied by $\sin\theta^*$, where θ^* is the angle between the jet and the beam axis in the jet rest frame,
- transverse momentum of the eighth leading jet,
- minimum ΔR distance between all the jets,
- aplanarity, defined as 3/2 times the smallest eigenvalue of the momentum tensor $M_{ij} = \sum_{k=1}^{N_{objects}} p_{ik}p_{jk} / \sum_{k=1}^{N_{objects}} p_k^2$, where p_{ik} is the i-th momentum component of the k-th jet and p_k is the modulus of its momentum,
- distance between the two jets placed in the b-parton positions by the KLF $\Delta R(b_1, b_2)$,
- global transverse thrust, defined as $T_{T,g} = max_{\overrightarrow{n_T}} \frac{\sum_i |\overrightarrow{p_{T,i}} \cdot \overrightarrow{n_T}|}{\sum_i p_{T,i}}$, where the sum runs over all particles p_i in the final state, $p_{T,i}$ represents the two momentum components transverse to the beam, and n_T is the transverse vector that maximises the sum. The observable which is resummed is then $\tau_{T,g} = 1 - T_{T,g}$ [140],
- maximum ΔR distance between all the jets,
- transverse momentum of the softest jet in the event,
- $\cos\theta^*$ assuming no boost in x and y.

The training is made on AFII $t\bar{t}H$ sample, requiring the presence of at least one b-tagged jet, and on background sample, defined as described in Section 5.8 (subtracting the $t\bar{t} + jets$ events from the data ones in exclusively one b-jet multiplicity bin).
Figure 5.8 and Figure 5.9 show the distributions of the input variables used for the BDT training in the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) events and the background ones. From these distributions it is possible to see the discrimination power of each variables.

In order to understand whether the AFII $t\bar{t}H$ sample can be used in the training step and the corresponding output can be applied on the full detector simulation of $t\bar{t}H$ ($H \to b\bar{b}$), it is necessary to investigate the compatibility between these two samples. A comparison between fast and full simulation on the twelve inputs variables is then made. Figures 5.10, 5.11, 5.12 show the agreement between the two samples on the variables of interest. It can be argued then that the two samples are in a satisfactory agreement; thus we are allowed to perform the BDT training on AFII simulation, taking the advantage of more statistics.
The boosting method that provides the best discrimination power is the Gradient Boost which is then kept and used in the next steps of the analysis. The shape of the output of the BDT is shown in Figure 5.13 for the fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) (signal) and for the multi-jet background.

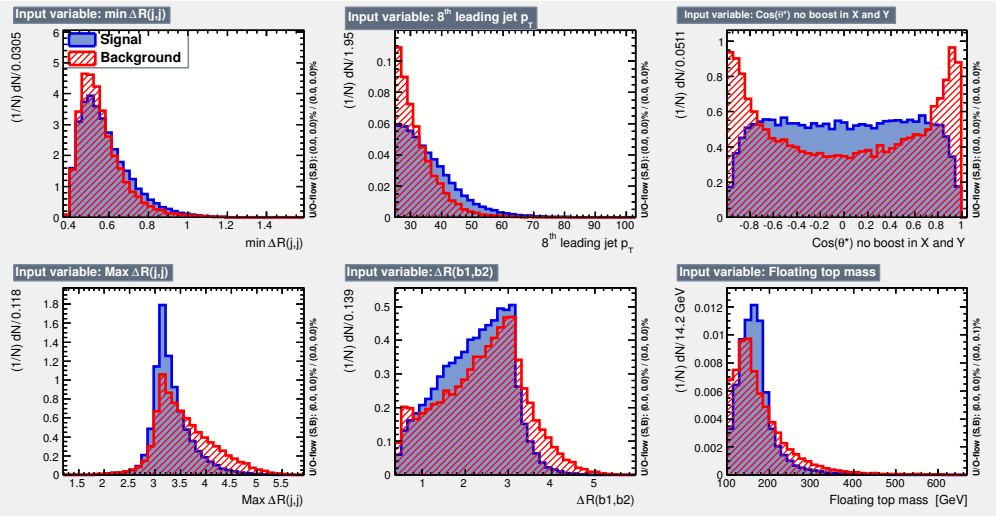


FIGURE 5.8 Variables used as inputs for the BDT classifier. The red histograms represent the $t\bar{t}H$ signal and the blue ones represent the multi-jet background.

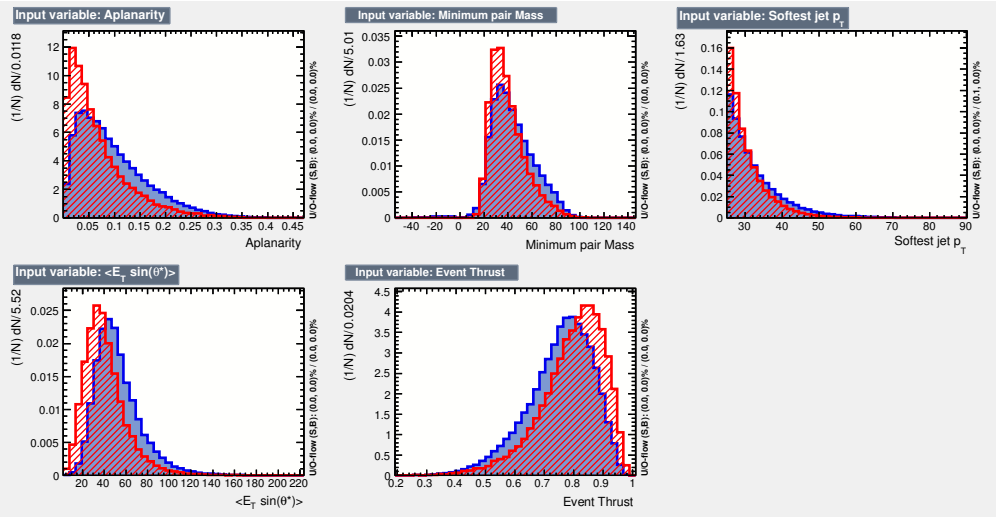


FIGURE 5.9 Variables used as inputs for the BDT classifier. The red histograms represent the $t\bar{t}H$ signal and the blue ones represent the multi-jet background.

Figure 5.13 shows also the comparison between the training and the test samples.

The TMVA framework provides also the correlation coefficient of each pair of input variables. The BDT expects the input variables to be uncorrelated. A non-vanishing correlation would be lead to incorrect efficiency estimation and hence to a lower discriminant power of the output variable. So if two variables show a high level of correlation one of the two variables must be removed from the input variable list. Figure 5.14(a) and Figure 5.14(b) show the correlation matrix obtained after the training for background and signal respectively. The highest correlation coefficient is between the transverse momentum of the eighth leading jet and the softest jet ($\sim 50\%$). A useful check consists in the elimination of the observable with the less discriminant power from the training and investigate the effect on the BDT performance with the respect to the previous setup.

Once the events classification is ready a ranking list of the variables is provided. The ranking value

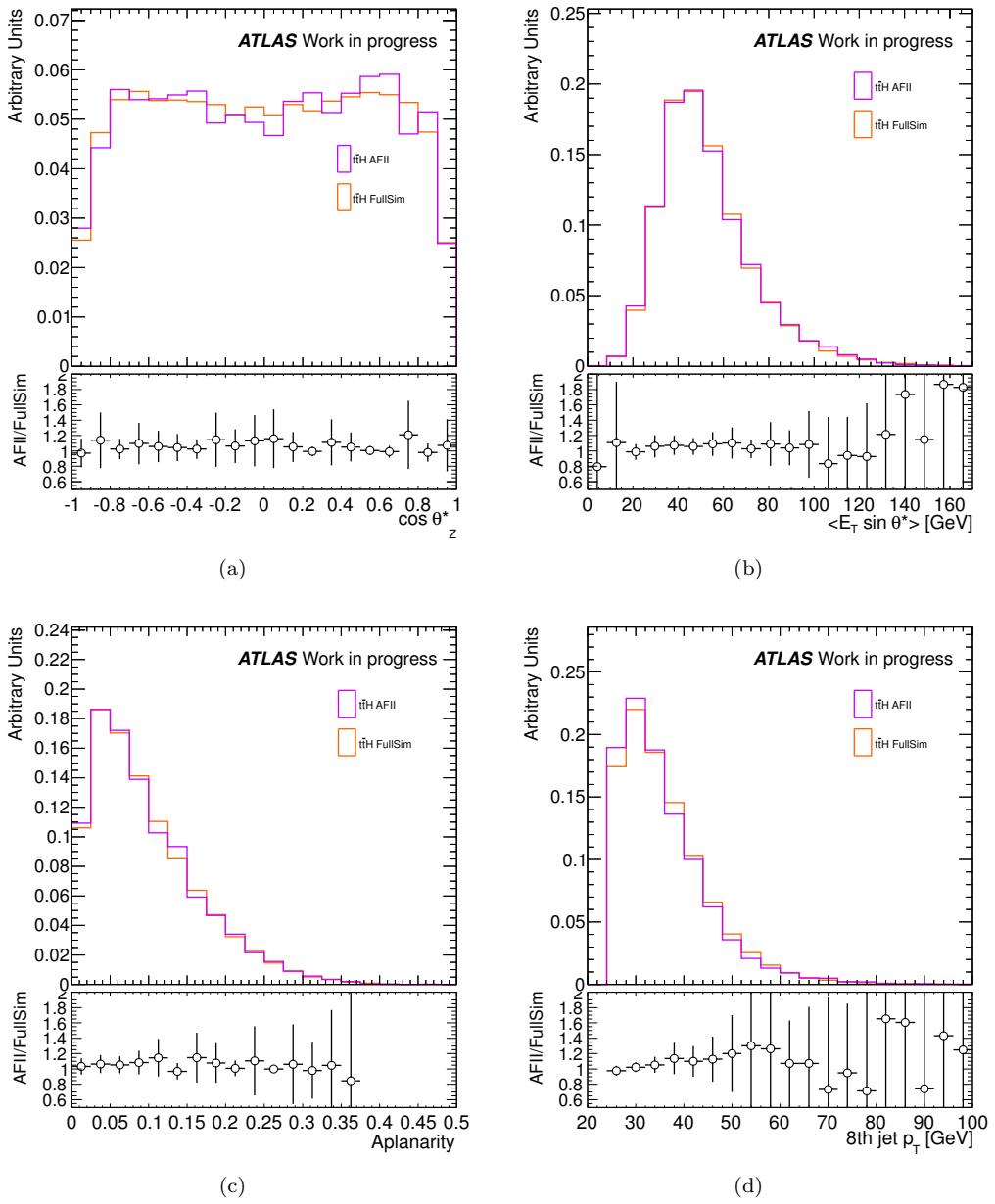
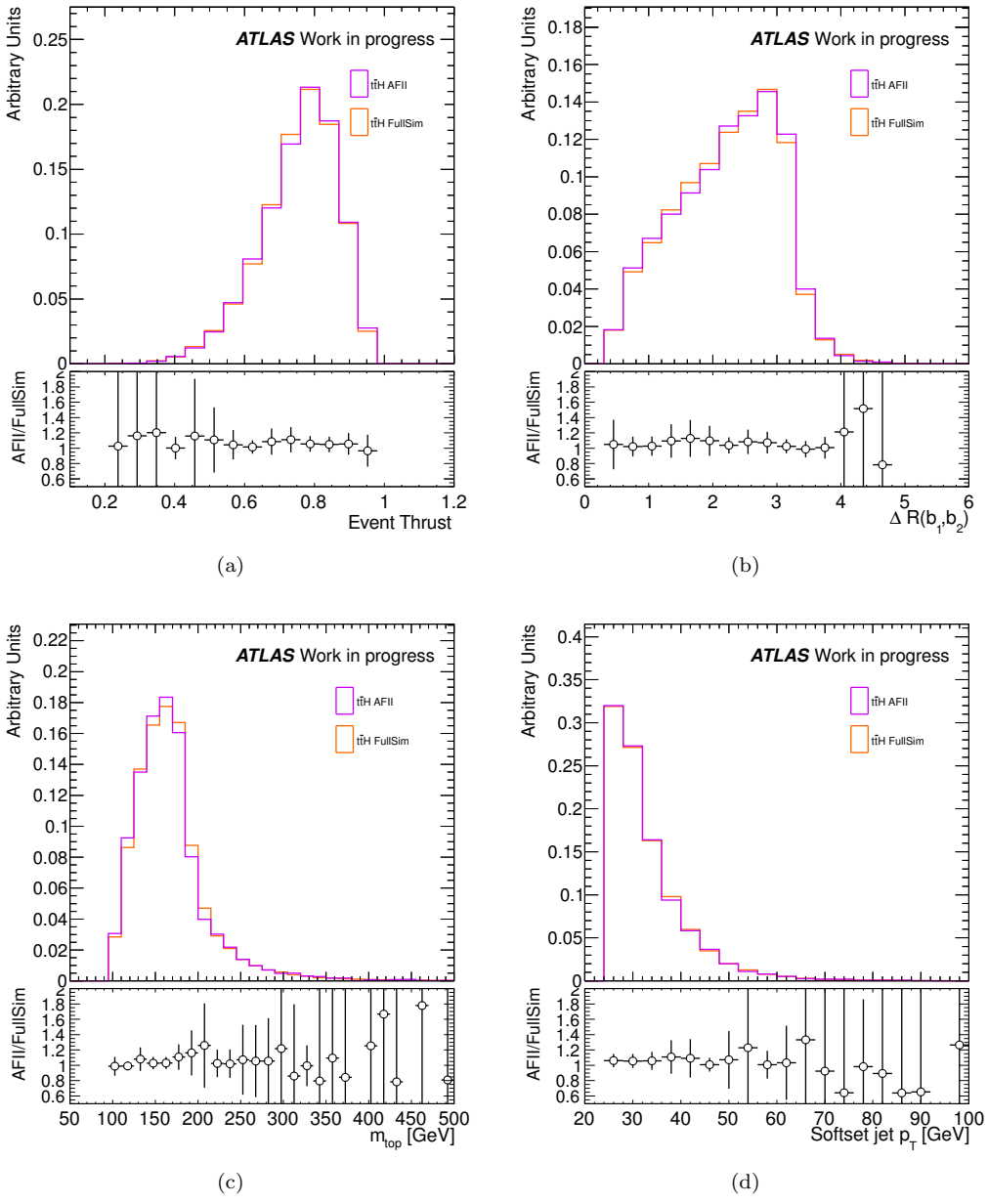


FIGURE 5.10 Comparison of AFII and FullSim for the fully hadronic $t\bar{t}H(H \to b\bar{b})$ sample. Four BDT input variables are presented: $\cos\theta^*$ (5.10(a)), average of the sum of all jets transverse energy multiplied by $\sin\theta^*$ (5.10(b)), aplanarity (5.10(c)) and eighth leading jet p_T (5.10(d)). The purple histograms represent the AFII simulation, the orange ones represent FullSim. The uncertainties on the ratio are statistical only.

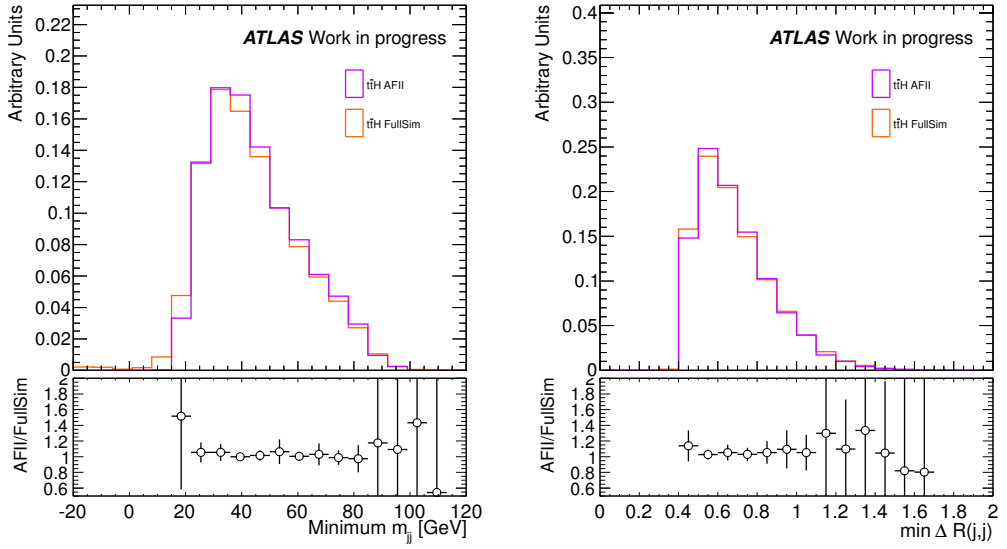
estimates the discriminant power of the observables. The ranking of the BDT input variables is derived by counting how often the variables are used to split decision tree nodes and by weighting each splited occurrence by the separation gain-squared it has achieved and by the number of the events in the node. Table 5.11 shows the ranking lists of the chosen BDT. The most powerful variable is the top quark mass reconstructed by KLF.



(a) (b)

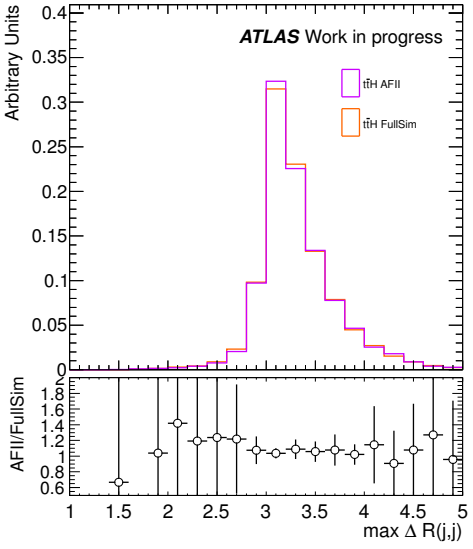
(c) (d)

FIGURE 5.11 Comparison of AFII and FullSim for the fully hadronic $t\bar{t}H(H \to b\bar{b})$ sample. Four BDT input variables are presented: for event thrust (5.11(a)), $\Delta R(b_1, b_2)$ (5.11(b)), top floating mass (5.11(c)) and softest jet p_T (5.11(d)). The purple histograms represent AFII simulation, the orange ones represent FullSim. The uncertainties on the ratio are statistical only.



(a)

(b)



(c)

FIGURE 5.12 Comparison of AFII and FullSim for the fully hadronic $t\bar{t}H(H \to b\bar{b})$ sample. Three BDT input variables are presented: minimum invariant mass of jet pairs (5.12(a)), maximum distance between two jets (5.12(b)) and minimum distance between two jets (5.12(c)). The purple histograms represent AFII simulation, the orange ones represent FullSim. The uncertainties on the ratio are statistical only.

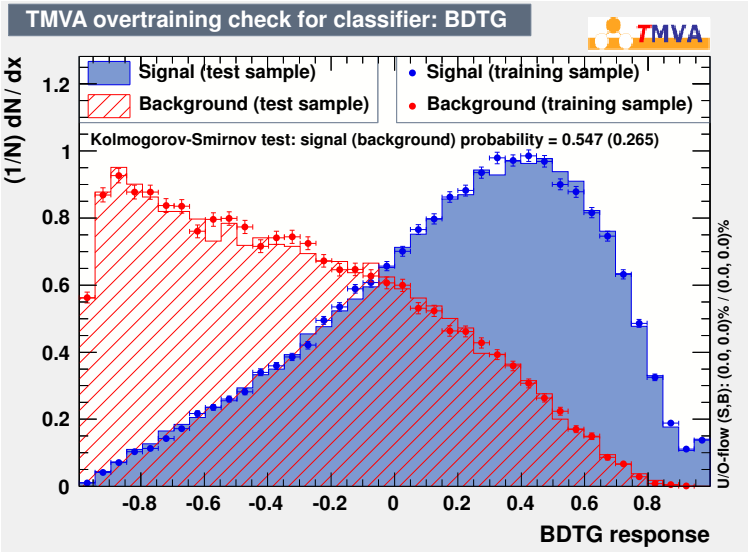
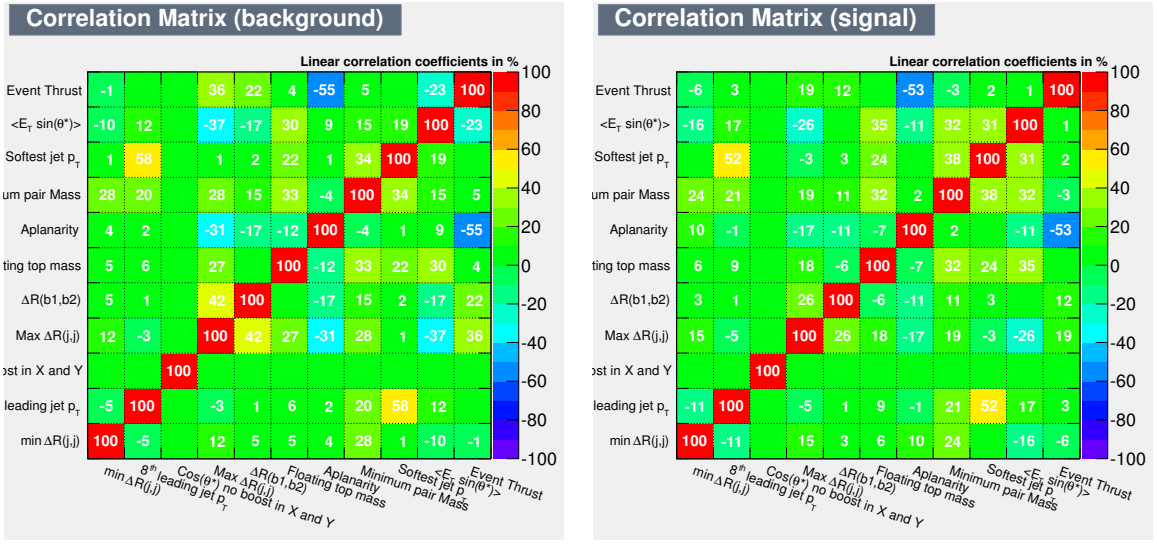


FIGURE 5.13 BDTG (G means Gradient boosting setup) output for the $t\bar{t}H(H \to b\bar{b}$ signal (blue) and for the multi-jet background (red). The sample used for the training (filled area) are compared with the test ones (solid marker).



(a) (b)

FIGURE 5.14 Correlation matrix of the BDT input variables for the multi-jet background (5.14(a)) and for the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$ signal sample (5.14(b)).

Variable	Ranking
Top quark mass	0.110
Min(m_{jj})	0.109
$< E_T \sin\theta^* >$	0.105
8th jet p_T	0.098
Min($\Delta R(j,j)$)	0.091
Aplanarity	0.090
$\Delta R(b1b2)$	0.082
Event Thrust	0.079
Max($\Delta R(j,j)$)	0.078
Softest jet p_T	0.078
$\cos\theta^*$ JustZ	0.077

TABLE 5.11 Ranking list of the input variables used in the training of BDT on $t\bar{t}H$ and background samples.

5.10 Application of "ABCD" method to the $t\bar{t}H$ ($H \to b\bar{b}$) analysis

For this first preliminary study the cut applied on the BDT output is not optimized and is chosen to be equal to 0.1. In the future a scan of expected sensitivity as a function of the BDT output variables will be performed. Once the choice of the BDT cut is done, it is possible to define eight regions, as shown in Table 5.12. The signal regions are then E and G, and the validation region C.

	$BDT_{output} > 0.1$	$BDT_{output} < 0.1$
$n_{b-jets} = 1$	A	B
$n_{b-jets} = 2$	C	D
$n_{b-jets} = 3$	E	F
$n_{b-jets} \geq 4$	G	H

TABLE 5.12 "ABCD" method: definition of the eight exclusive regions. The region C is the validation region, while the regions E and G are the signal ones.

The multi-jets background contribution are computed using the Equation 5.4, Equation 5.5 and Equation 5.6.

Table 5.13 shows the predicted events for the QDC multi-jet background, the expected events for $t\bar{t}$+jets, fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) as well as the data in the signal and validation regions. From comparison of the number of events in data and in the total background, which includes QCD multi-jet and $t\bar{t}$ productions, it is possible to argue that the "ABCD" method is able to predict the QCD yields with a quite good precision. Once the QCD multi-jet process is modeled in each region, it is possible to check its shape for several kinematical variables distributions in order to estimate the agreement between data and prediction. The control plots are made in exclusive bins of the b-tagged jet multiplicity: exactly two (Figures 5.15, 5.16 and 5.17), exactly three (Figures 5.18, 5.19 and 5.20) and at least four (Figures 5.21, 5.22 and 5.23) b-tagged jets in the events. The list of systematic uncertainties is described in Section 5.11. There is no multi-jet systematic

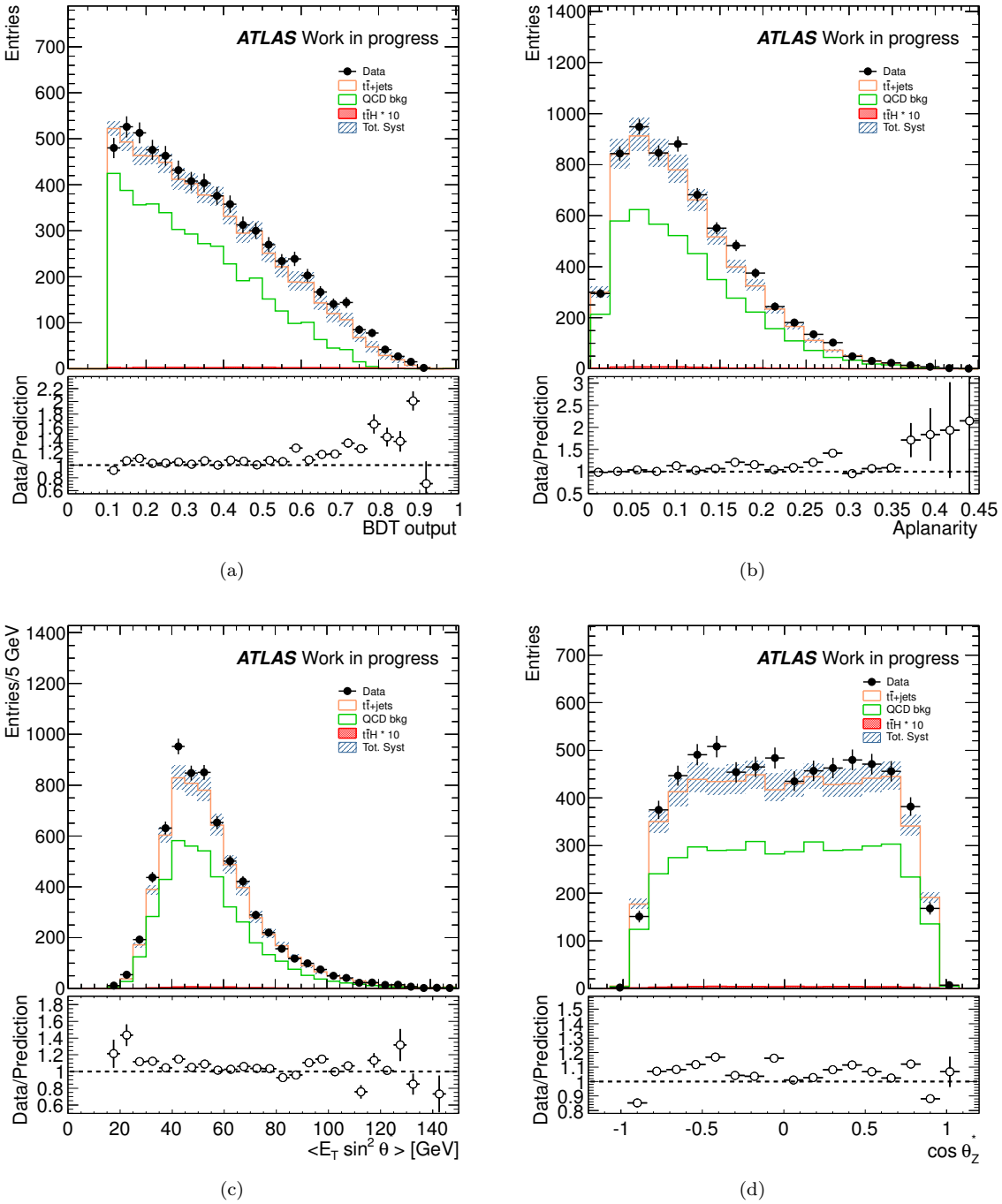
	Region C	Region E	Region G
Data	6679	1070	111
$t\bar{t}H$	5.7 ± 2.3	4.7 ± 2.1	2 ± 1
multi-jet	4265 ± 55	557 ± 23	79 ± 7
$t\bar{t} + jets$	2014 ± 45	357 ± 19	30 ± 5
Total BKG	6279	914	109

TABLE 5.13 Data, signal and background event yield in regions C, E and G.

uncertainties included at this stage though preliminary study indicates it can be of order of 20%. This is supported by the fact that data versus prediction agreement in the validation region C is of about 10%.

5.10.1 $t\bar{t}H(H \to b\bar{b})$ validation and discriminant variable distributions

The Higgs boson candidates are searched for after the reconstruction of the $t\bar{t}$ system by the kinematic fit. In particular the KLF setup used at this step is the so called "fixed top quark mass". Once the KLF finds the combination of six jets which maximizes the likelihood, the remaining other jets are considered as possible Higgs decay product candidates. The two leading jets not used in the KLF are used to built the invariant mass of the Higgs candidate. The invariant mass distributions for the three signal regions are shown in Figure 5.24.



(a) (b) (c) (d)

FIGURE 5.15 Data and background prediction in the exclusive two b-tagged jets region for the BDT output variable (5.15(a)) as well as for the BDT input variables: aplanarity (5.15(b)), average of the transverse energy multiplied by $\sin\theta^*$ (5.15(c)) and $\cos\theta^*$ not boosted in the x and y axes (5.15(d)). The shaded area represents the total systematic uncertainty (sec Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.

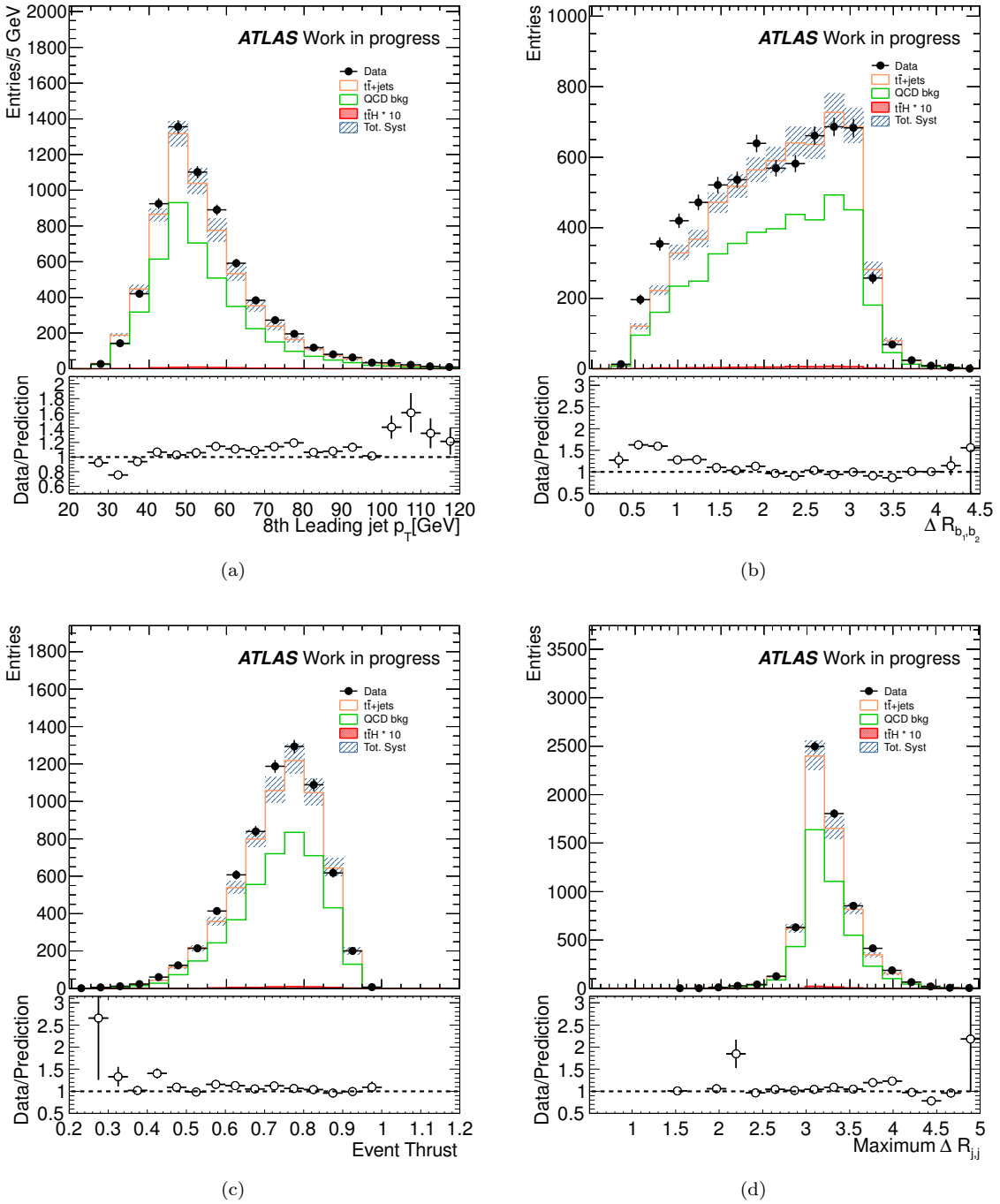
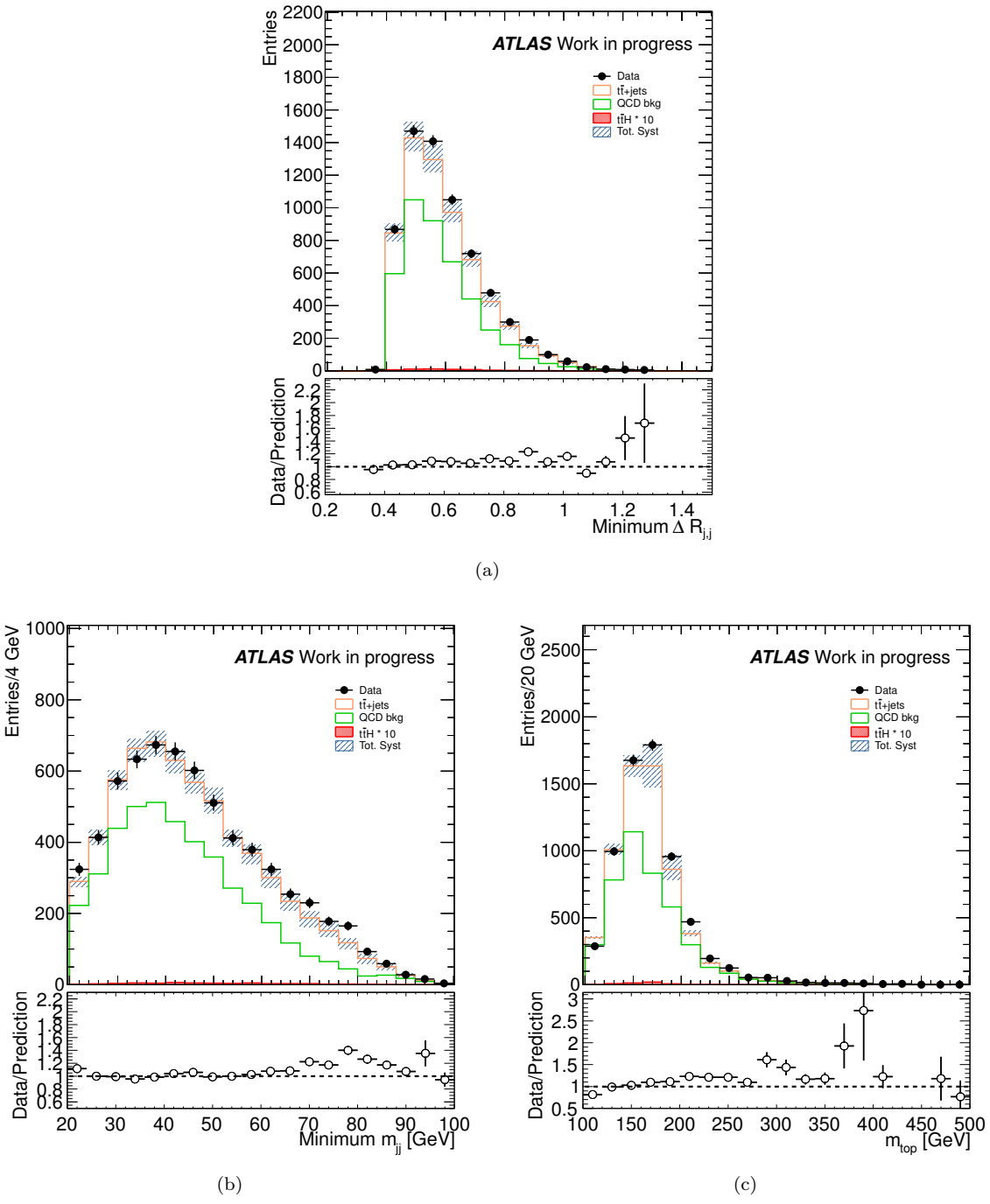


FIGURE 5.16 Data and background prediction in the exclusive two b-tagged jets region for the BDT input variables: transverse momentum of the eight leading jet (5.16(a)), distance betwen two b-jets (5.16(b)), event thrust (5.16(c)) and maximum distance between each jets pair (5.16(d)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.

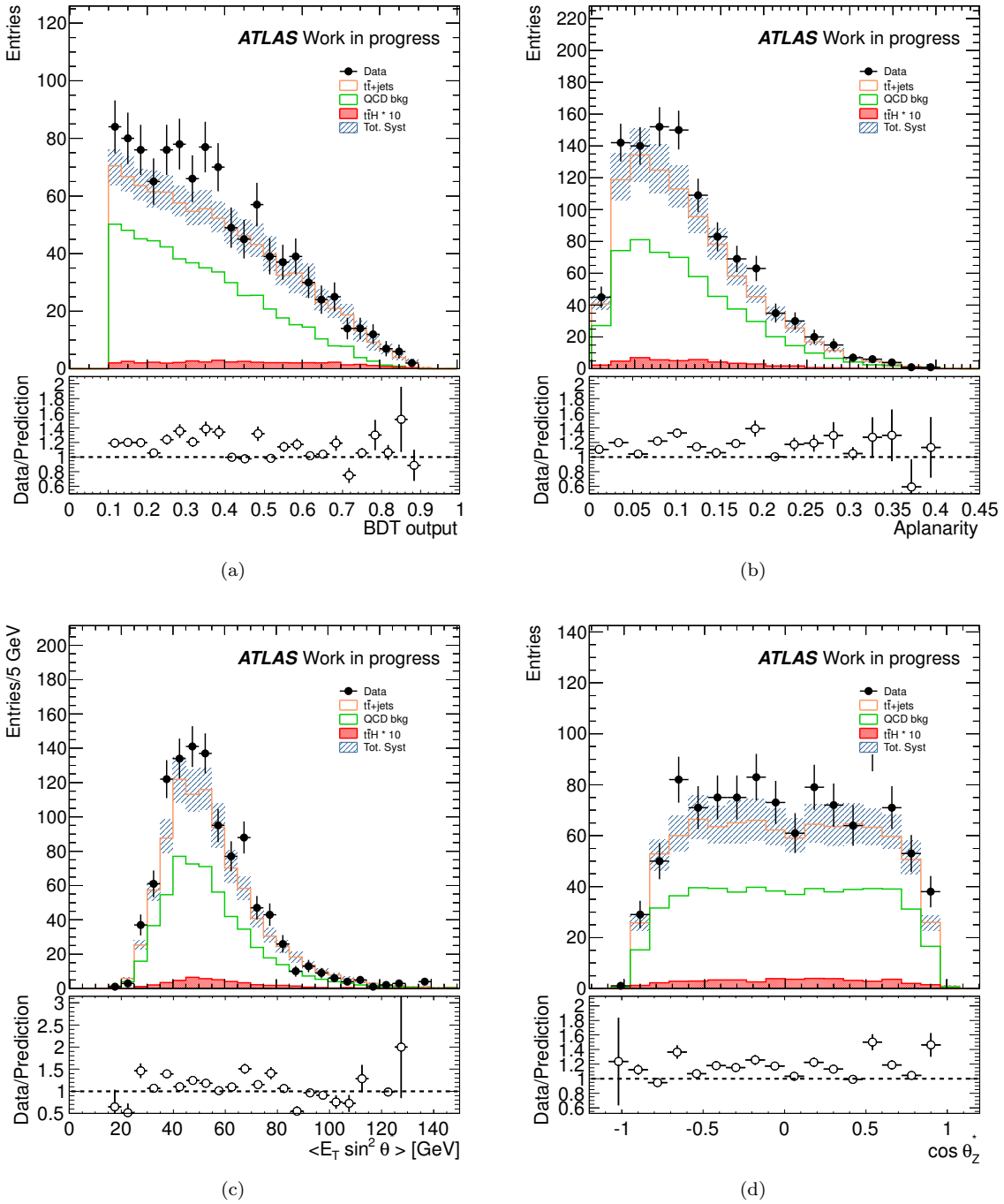


(a)

(b)

(c)

FIGURE 5.17 Data and background prediction in the exclusive two b-tagged jets region for the BDT input variables: minimum distance between each jets pair (5.17(a)), minimum invariant mass of each jets pair (5.17(b)) and reconstructed top quark mass (5.17(c)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.



(a) (b) (c) (d)

FIGURE 5.18 Data and background prediction in the exclusive three b-tagged jets region for the BDT output variable 5.18(a) as well as for the BDT input variables: aplanarity 5.18(b), average of the transverse energy multiply by $\sin\theta^*$ 5.18(c) and $\cos\theta^*$ not boosted in the x and y axes 5.18(d). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.

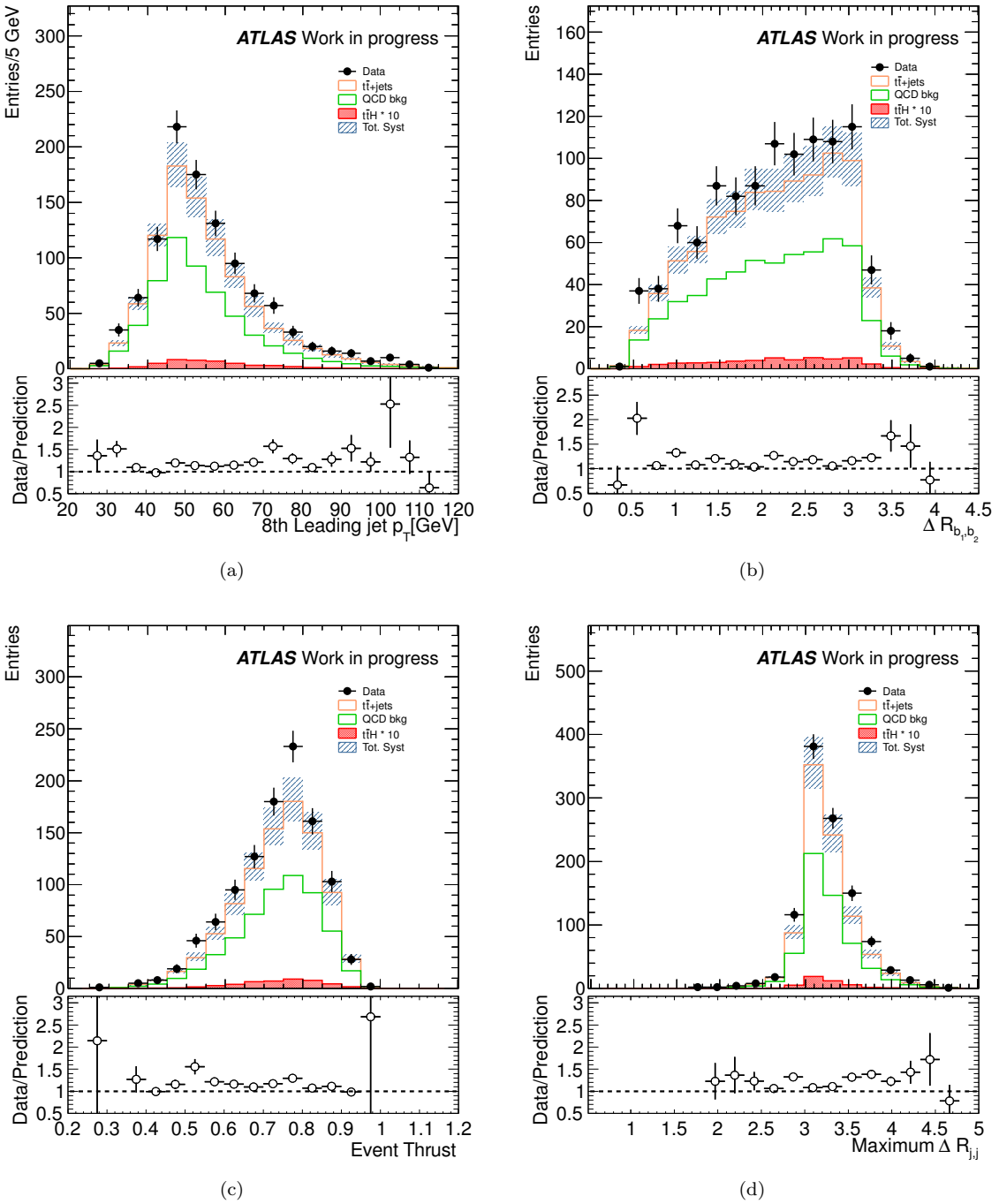
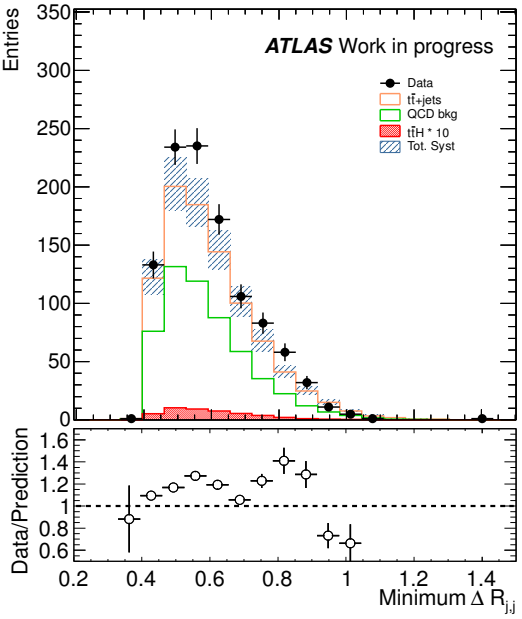
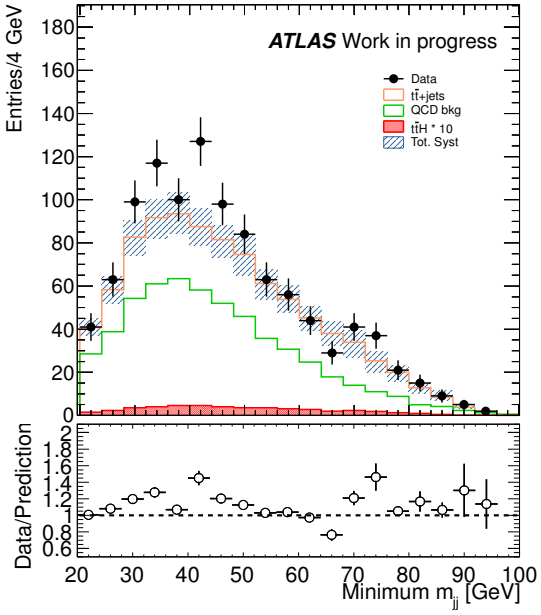


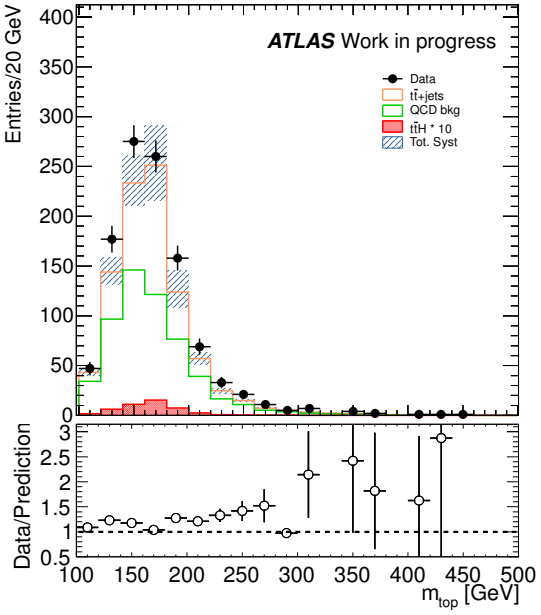
FIGURE 5.19 Data and background prediction in the exclusive three b-tagged jets region for the BDT input variables: transverse momentum of the eight leading jet 5.19(a), distance betwen two b-jets (5.19(b)), event thrust (5.19(c)) and maximum distance between each jets pair (5.19(d)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visuaze it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.



(a)

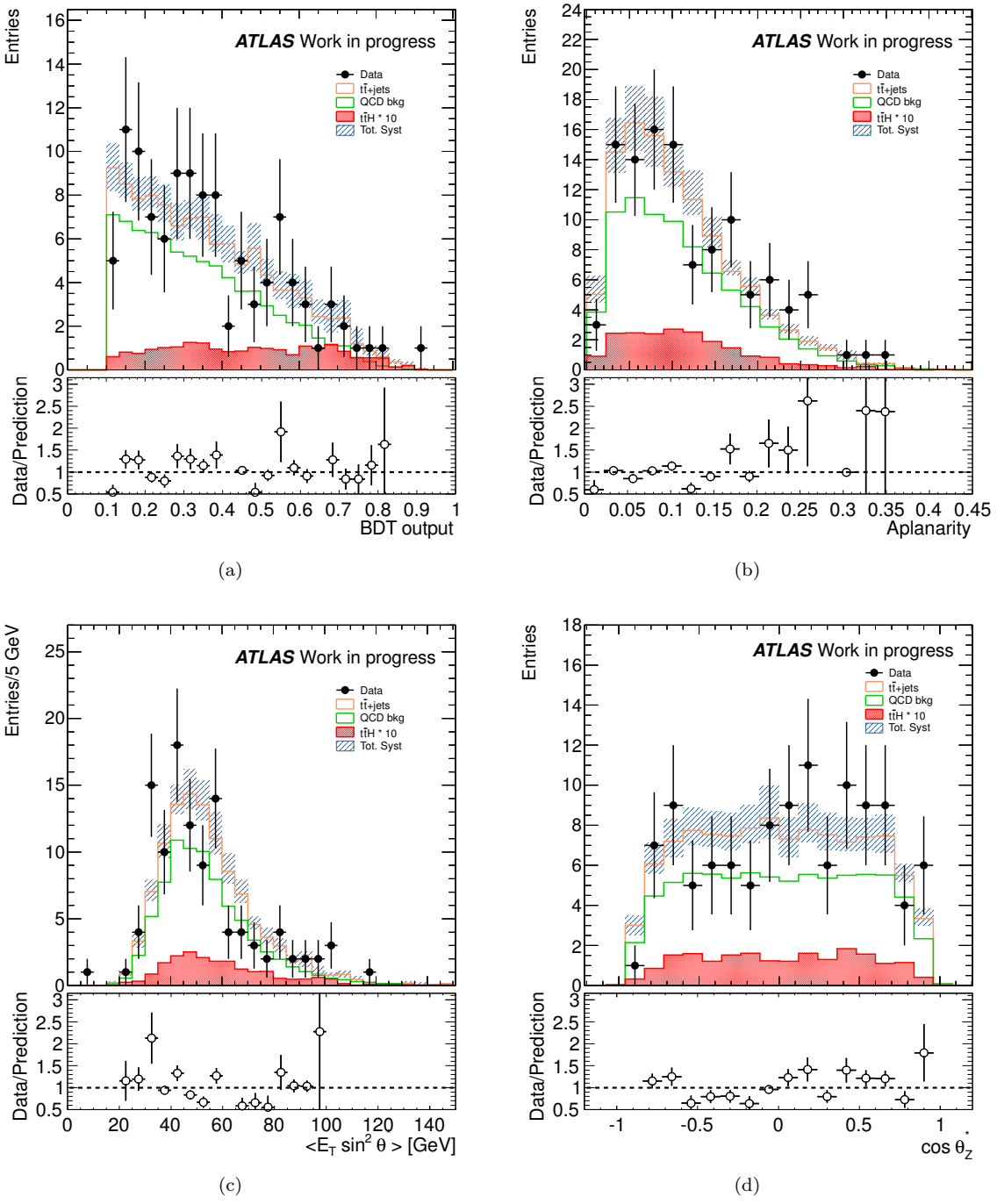


(b)



(c)

FIGURE 5.20 Data and background prediction in the exclusive three b-tagged jets region for the BDT input variables: distance betwen two b-jets (5.19(b)), event thrust (5.19(c)), maximum distance between each jets pair (5.19(d)), minimum distance between each jets pair (5.20(a)), minimum invariant mass of each jets pair (5.20(b)), reconstructed top quark mass (5.20(c)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.



(a) (b) (c) (d)

FIGURE 5.21 Data and background prediction in the inclusive four b-tagged jets region for the BDT output variable (5.21(a)) as well as for the BDT input variables: aplanarity (5.21(b)), average of the transverse energy multiplied by $\sin\theta^*$ (5.21(c)), $\cos\theta^*$ not boosted in the x and y axes (5.21(d)), transverse momentum of the eight leading jet (5.22(a)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.

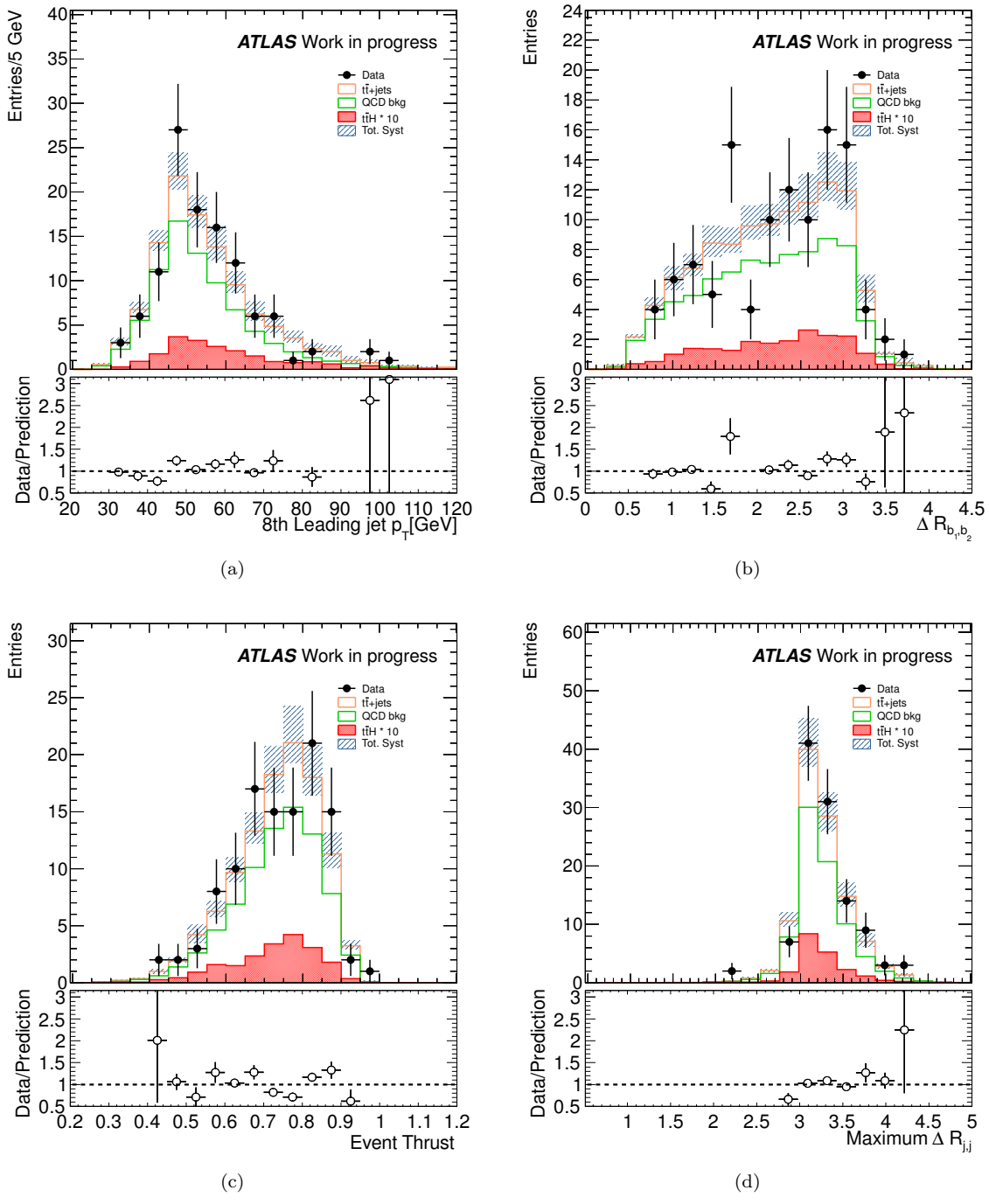
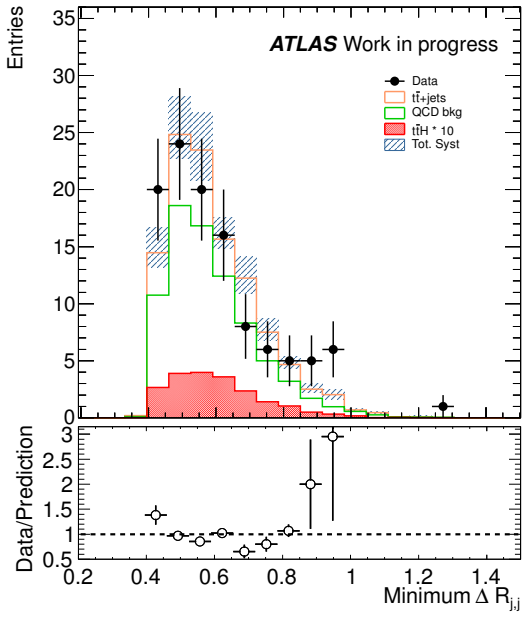
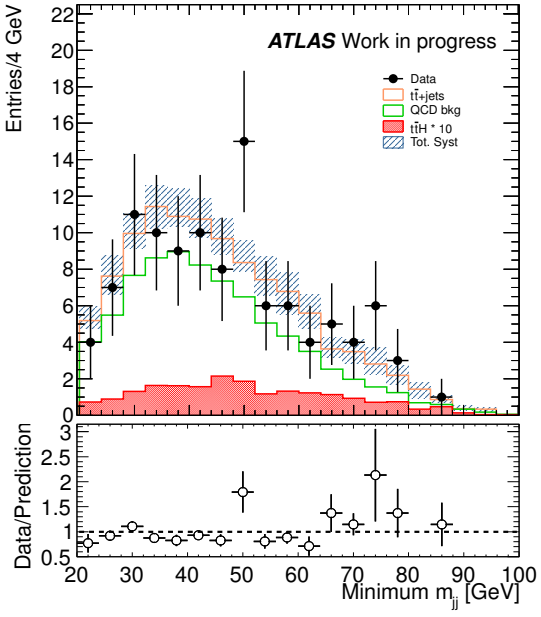


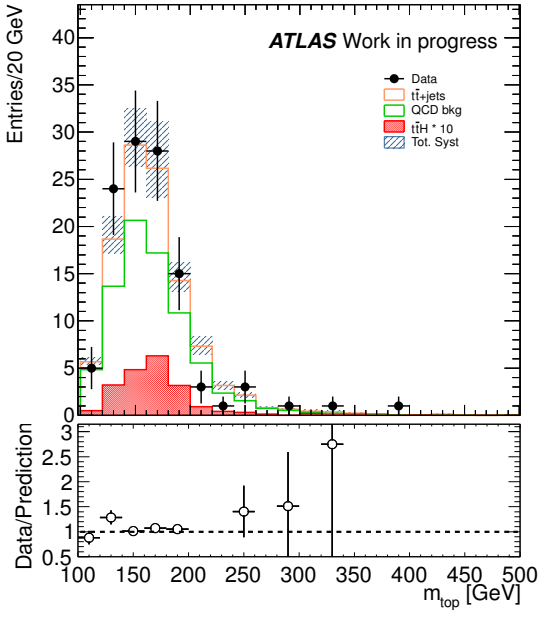
FIGURE 5.22 Data and background prediction in the inclusive four b-tagged jets region for the BDT input variables: distance betwen two b-jets (5.22(b)), event thrust (5.22(c)) and maximum distance between each jets pair (5.22(d)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.



(a)

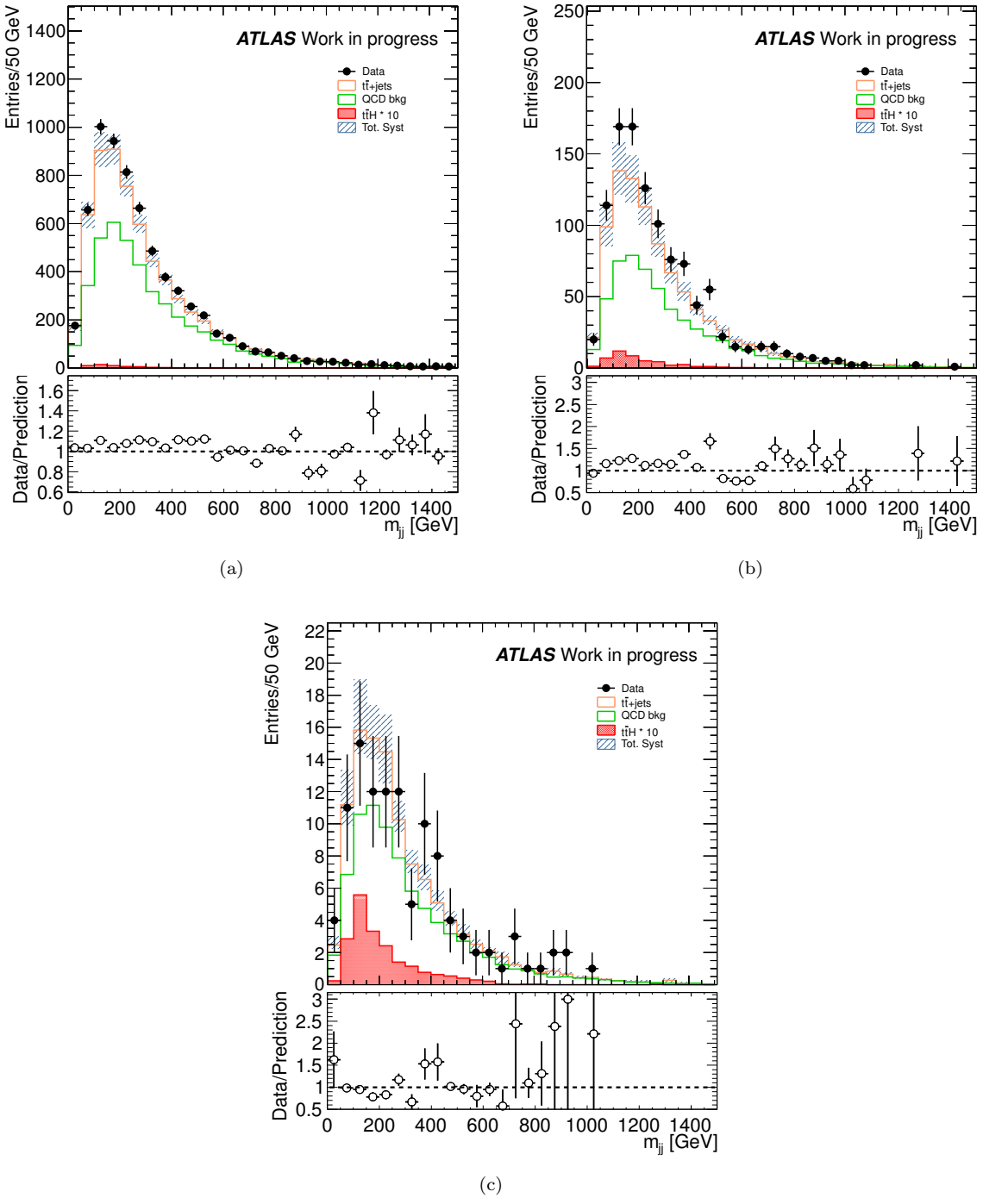


(b)



(c)

FIGURE 5.23 Data and background prediction in the inclusive four b-tagged jets region for the BDT input variables: minimum distance between each jets pair (5.23(a)), minimum invariant mass of each jets pair (5.23(b)) and reconstructed top quark mass (5.23(c)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.



(a)

(b)

(c)

FIGURE 5.24 Invariant mass of the two jets chosen to be from the Higgs boson decay by the Kinematic Likelihood Fit in region C (5.24(a)), region E (5.24(b)) and region G (5.24(c)). The shaded area represents the total systematic uncertainty (see Section 5.11 for the systematic uncertainties estimation). Data are represented with black markers. The orange and green histograms represent $t\bar{t}$+jets and QCD respectively. The red histogram represents the $t\bar{t}H(H \to b\bar{b})$. To better visualize it the signal contribution is multiplied by a factor of 10. The bottom panel displays the ratio between data and background prediction.

5.11 Systematic uncertainties

The sources of systematic uncertainties considered in this analysis affect both the normalization of the signal and the background, and the shape of the final discriminant distributions. The details of the systematics uncertainties affecting the signal and the backgrounds in each of the three considered signal regions are summarized in Table 5.15, Table 5.16 and Table 5.17, respectively for region C, D and E. The following section describes how these uncertainties are implemented in the analysis and gives an estimation of their magnitudes.

5.11.1 Luminosity

The systematic uncertainty related to the collected luminosity has been measured by ATLAS in dedicated runs and is estimated to be of order of 1.8% [141] for the 2011 data-taking period. This uncertainty affects only the normalization. After the fully hadronic $t\bar{t}$ cross section analysis an improvement on the estimation of the luminosity uncertainties was obtained by ATLAS, leading to a lower luminosity uncertainty contribution.

5.11.2 Jet energy scale

The JES and its uncertainty have been derived from the combination of the test-beam data, LHC collision data and MC simulation [80]. More details on the JES can be found in Section 4.13.1.

5.11.3 Jet energy resolution and Jet reconstruction efficiency

The procedure followed for the estimation of the jet energy resolution and the jet energy efficiency is the same as the one used for the $t\bar{t}$ cross section analysis, see Section 4.13.3 and Section 4.13.2

5.11.4 Heavy flavour tagging

The effects of the systematic uncertainties on the efficiency of the heavy flavour identification algorithm is evaluated in the analysis. The working point chosen corresponds to a b-tagging efficiency of 60% for b-jets. The efficiency obtained in the simulation is corrected by a p_T/η depended scale factor different for the three quark flavours ($b/c/light$). To each scale factor is assigned a systematics uncertainty. The scale factors and their uncertainties are applied to each jet in the simulation samples depending on the quark flavour, transverse momentum and pseudorapidity.

5.11.5 $t\bar{t}H$ modelling

In order to evaluate the impact of the parton shower modelling, the analysis is performed on specific samples. These samples are produced varying the amount of the initial and final state QCD radiation through a variation of the relevant parameter in PYTHIA generator in a range that is consistent with the corresponding experimental data. In particular two samples are used: one with high amount of initial and final state radiations and the second with a lower amount. The difference between the two samples is symmetrized and taken as a systematic $\pm\sigma$ variation. The contribution of the $t\bar{t}H$ modelling changes in the three signal regions; it is estimated to be within $4 - 19\%$.

5.11.6 Discussion on QCD multi-jet background systematic

The systematic uncertainty on the QCD background estimation can be divided in two categories: one connected to the normalization estimation and the other to the shape modeling.
In order to assign a systematic uncertainty for the QCD background normalization estimated using the "ABDC" method, a study of the MC samples used in the analysis is performed. The aim is to evaluate the variation of the number of fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) and $t\bar{t}$+jets events in the signal region ($BDT > 0.1$) and in background region ($BDT < 0.1$) as a function of the b-tagged jets multiplicity. The fraction α, defined as the number of events in signal region N_{sig} over the number of events in the background region N_{bkg}, $\alpha = \frac{N_{sig}}{N_{bkg}}$, in four bins of b-tagged jet multiplicity is summarized in Table 5.14.

b-tagged jet multiplicity	$\alpha = \frac{N_{sig}}{N_{bkg}}$ (%)	
	fully had. $t\bar{t}H$ ($H \to b\bar{b}$)	$t\bar{t}$+ jets
$n_{b-jets} = 1$	1.29	0.96
$n_{b-jets} = 2$	1.35	1.10
$n_{b-jets} = 3$	1.62	1.06
$n_{b-jets} \geq 4$	1.82	1.11

TABLE 5.14 Fraction of the number of MC simulation events with $BDT > 0.1$ (N_{sig}) over the number of MC simulation events with $BDT < 0.1$ (N_{bkg}) α in four b-tagged jets multiplicity bins.

The relative variation between the maximum and the minimum of the α fraction $\frac{\alpha_{max}-\alpha_{min}}{\alpha_{min}}$ is taken an indication of the uncorrelation assumption and corresponds to a variation of a $\pm 20/40\%$ for the $t\bar{t}$+jets and $t\bar{t}H$. The correlation between the b-jet multiplicity and the BDT output variable can also be noticed in Figure 5.25 where the distribution of the BDT output variable is shown as a function of the b-jet multiplicity in the $t\bar{t}$+jets MC sample, after the application of all the analysis selection.

The estimation of the QCD background shape is explained in detail in Section 5.10. What follows is a first assessment of the uncertainties related to the shape modeling. Considering just the signal

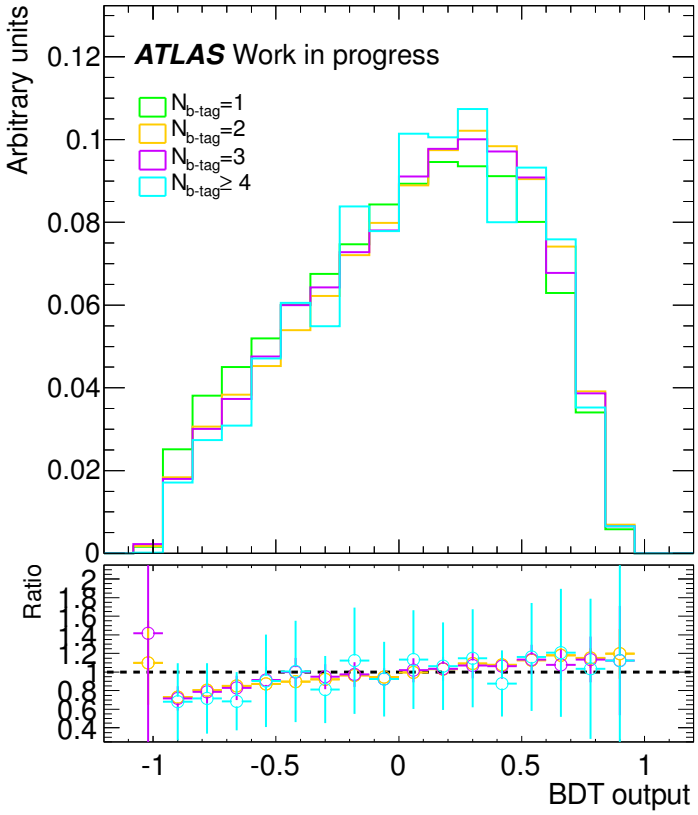


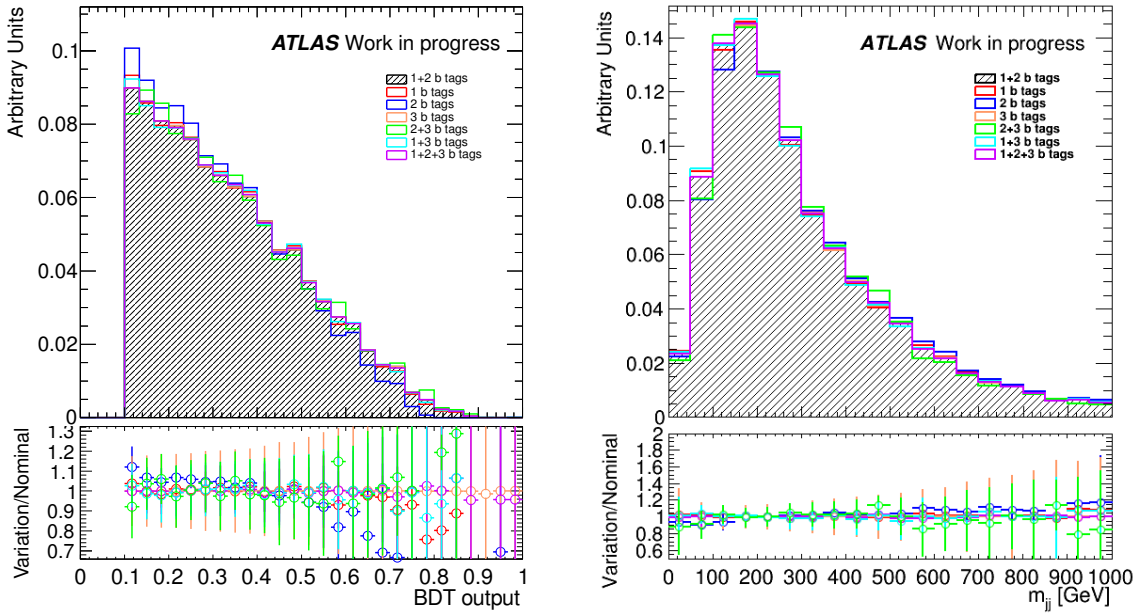
FIGURE 5.25 BDT output varibale as a function of the b-jet multiplicity in $t\bar{t}$+jets MC sample, after applying all the analysis selection.

region E in which there are at least four b-tagged jets and the BDT output variable is above 0.1, it is interesting to compare the default background template used in this region with other alternative combinations. As a reminder, the default configuration is constructed by the combination of A and C regions, exactly 1 b-tagged and 2 b-tagged jets respectively Equation 5.6. Other background templates can be made using single or combination of all the control regions:

- 1 b-tag: region A ,
- 2 b-tags: region C ,
- 1 and 3 b-tags: regions A and E,
- 1, 2 b-tags: regions C and E,
- 1, 2 and 3 b-tags: regions A, C and E.

Figure 5.26(a) and Figure 5.26(b) show the comparison between the different templates for the BDT output variable and for the invariant mass of the two leading jets not considered in the $t\bar{t}$ system reconstruction by the KLF. In the below pads it is shown the ratio among the alternative shape modelings and the default one.

The maximum variation among the different templates gives a qualitative assessment of the



(a) (b)

FIGURE 5.26 Comparison between default background template used in region E with other alternative combinations for the BDT output variable and for the invariant mass of the two jets chosen to be the Higgs bosons decay products. In the below pads it is shown the ratio among the alternative shape models and the default one.

uncertainty on the QCD background modeling performed with the "ABCD" method.

Source of Unc,	$t\bar{t}+jets$	$t\bar{t}H$	Total
JES	15.8/-16.6	12.5/-13.9	
b-tagging	11.6/-12.5	9.0/-5.3	
c-tagging	2.5/-2.5	0.7/-0.7	
l-tagging	0.54/-0.54	0.2/-0.2	
JER	$<<1$	$<<1$	
JRE	$\pm$ 0.17	0.1	
ISR/FSR		18.71	
Luminosity			1.8
Total	19.9/-20.7	23.9/-24	

TABLE 5.15 Systematic uncertainties contribution for the $t\bar{t}+jets$ and for the $t\bar{t}H$ samples in region C.

5.12 Conclusion

In this chapter the first analysis strategy of the search for the Higgs boson in associate production with a top quark pair, with the Higgs boson decaying in the bottom anti bottom pair and the top quark in the fully hadronic decay mode, has been presented. The preliminary studies on the topology reconstruction and associated performance have been discussed, as well as the modeling

Source of Unc,	$t\bar{t}+jets$	$t\bar{t}H$	Total
JES	16.2/-17.3	13.2/-12	
b-tagging	14.8/-15.8	11.8/-9.4	
c-tagging	19.6/-19.6	1.9/-1.8	
l-tagging	3.3/–3.2	0.1/-0.1	
JER	$<<1$	$<<1$	
JRE	$<<1$	$<<1$	
ISR/FSR		3.69	
Luminosity			1.8
Total	29.8/-30.4	17.7/-16.3	

TABLE 5.16 Systematic uncertainties contribution for the $t\bar{t}+jets$ and for the $t\bar{t}H$ samples in region E.

Source of Unc,	$t\bar{t}+jets$	$t\bar{t}H$	Total
JES	16.1/-13.4	12.8/-9.5	
b-tagging	19.2/-22.0	24.9/-30.4	
c-tagging	29.1/-34.9	9.9/-10.3	
l-tagging	8.2/-8.6	2.1/-2.1	
JER	$<<1$	$<<1$	
JRE	$\pm$ 0.4	$<<1$	
ISR/FSR		10	
Luminosity			
Total	39.7/-43.4	31.3/-34.8	1.8

TABLE 5.17 Systematic uncertainties contribution for the $t\bar{t}+jets$ and for the $t\bar{t}H$ samples in region G.

of the background sources, especially the QCD multi-jet production which represents the most challenging background in this analysis. A multivariate analysis is used in order to discriminate between the background and the signal. The data and background comparison in the signal regions are presented showing a quite good agreement. The contribution of the systematic uncertainties has been quantified in the signal and validation regions. Unfortunately due to the lack of time, the limit has not been set on the signal strength. This first attempt to measure the $t\bar{t}H(H \to b\bar{b})$ production at the $\sqrt{s} = 7$ TeV and with integrated luminosity of 4.7 fb^{-1} shows that one can reach a good potential in this measurement with more integrated luminosity, improved analysis tools and better handle of main components of the systematics uncertainties. The signal over square root of background in the two signal regions, 0.16 in the exclusive three b-tagged jet bin and 0.19 in the inclusive four b-tagged jet bin, gives an idea of the future performance that could be achieved by the fully hadronic $t\bar{t}H(H \to b\bar{b})$. Also the analysis shows a good discrimination power between the $t\bar{t}H(H \to b\bar{b})$ signal and the $t\bar{t}$+jets background, of about 1.3% and 6.7% in the exclusive three b-tagged jet bin and the inclusive four b-tagged jet bin respectively. Considering this first preliminary study, the fully hadronic $t\bar{t}H(H \to b\bar{b})$ seems a quite promising analysis which could be integrated in the future combination where all the top anti-top final states, as well as Higgs boson decays, are merged. There are still some crucial points to be investigated and to understand

in order to achieve a competitive analysis, such as the QCD multi-jet background modeling. An important improvement of the fully hadronic $t\bar{t}H(H \rightarrow b\bar{b})$ can be obtained by extending the analysis to the low jet multiplicity bins where the system topology is not completely reconstructed and by performing the analysis on the proton proton collision collected by the ATLAS detector during the 2012 data taking at $\sqrt{s} = 8$ TeV.

Conclusions

Over the last century the theorists and experimental physicists have joined the efforts to explore the fundamental structure of the matter. The best picture which describes how the fundamental particles and the fundamental forces are related is provided by the Standard Model (SM) of particle physics. The SM is a very remarkable theory which has been deeply and thoroughly tested and validated by a large number of experiments over the last five decades. This allowed a high precision description of the properties of the fundamental particles and their interactions. Nevertheless this successful theoretical model presents some conceptual weaknesses and it is not able to account for or answer some of the important and pending fundamental questions such as the nature of the dark matter and dark energy, gravity, matter-antimatter asymmetry in the universe, the mass hierarchy of elementary particles, etc. With the LHC[4] accelerator, the particle physics community hopes to bring answers to part of these fundamental questions and shade light in the others. The announcement in the fourth of July 2012 by the ATLAS and CMS collaborations of the discovery of a Higgs-like particle is a first step towards other potential discoveries at the LHC in the next decade.

The LHC started in 2010, with the collection of the first pp collisions at $\sqrt{s} = 900$ GeV. It was followed by two long data-taking periods at center-of-mass energy of $\sqrt{s} = 7$ TeV in 2011 and at $\sqrt{s} = 8$ TeV in 2012, with a total integrated luminosity of 25 fb^{-1}. The LHC machine has shown excellent performances together with the detectors operating along its ring, proving the possibility to investigate a wide spectrum of particle physics sectors in a very short time. After few months of data-taking, the main Standard Model measurements were validated and first limits were set on several Beyond Standard Model scenarios. Shortly after in July 2012, the particle physics community has lived an impotant and historical moment in the current century with the discovery of the Higgs-like scalar boson.

All these results were made possible thanks to the hard work of few thousands of physicists through a continuous control of the detectors status and performances during the proton-proton collisions. Furthermore an efficient trigger and data-taking system were key elements in this success and allowed to select the interesting collisions originating from the hard scattering out of the overwhelming soft ones. A relevant part of this thesis was devoted to the development of the online

[4]The LHC was designed and constructed during the last twenty years by a world wide collaboration of physicists and engineers.

and offline monitoring of two specific trigger chains: b-jet and μ-jet triggers. The b-jet trigger represents a crucial ingredient for many physics analyses which search for a rare signal in a final state characterized by high jet multiplicity with the presence of b-jets, such as the $t\bar{t}$ and in $t\bar{t}H$ in the fully hadronic decay mode. Thanks to the monitoring system wrong configurations of the trigger algorithms were pointed out during the data-taking and then quickly fixed. Though an optimization work is still necessary for the future data-taking campaigns in order to improve the trigger monitoring algorithms through the introduction of the efficiency and fake rate distributions for the b- and μ-jet triggers.

The discovery of the Higgs-like boson does not represent the end point of the LHC scientific program. It rather marks the beginning of a new era in the field of particle physics. A first important point is to validate the nature of this new particle in order to understand whether it is the SM Higgs boson or one of the Higgs bosons predicted by beyond SM theories. An answer to this question can be provided by the study of the Higgs coupling with the SM particles, in particular with the third generation of the fermions, top and bottom quarks. In order to explore one of the most excitant areas of particle physics, my thesis work focused on probing top quark and Higgs boson physics, starting from a detailed study of the top pair production and finishing with the search for the Higgs boson produced in association with a top quark pair.

In this thesis the measurement of the $t\bar{t}$ cross section in the fully hadronic decay mode performed using an integrated luminosity of 4.7 fb^{-1} collected by the ATLAS detector at $\sqrt{s} = 7$ TeV is presented. The result reported is $\sigma(pp \to t\bar{t}) = 168 \pm 12$ (stat.) $^{+60}_{-57}$ (syst.) ± 7 (lum.) pb. This measurement, limited by the systematic uncertainty, is compatible with the measurements performed in the other $t\bar{t}$ final state decays and with the SM prediction. In order to reduce the impact of the systematic uncertainty, very accurate performance studies on the jet energy scale and on the b-tagging are being performed, as well as an improvement in the MC simulation modelling for example of the initial and final radiations. For this reason the ATLAS collaboration investigates the jet multiplicity in $t\bar{t}$ events with the aim to constrain the ISR/FSR model [142], one of the most important uncertainties in the fully hadronic top quark measurement.

Furthermore additional performing techniques could be implemented in the analysis, such as profiling likelihood methods [143]. In the profiling likelihood technique the systematic uncertainties are treated by the introduction of nuisance parameters aiming at the reduction of the single and total systematic contributions.

The $t\bar{t}$ cross section measurement, presented in this dissertation, can been compared with the single lepton, di-lepton final state results, as well as with the previous fully hadronic result, considered in the combination. The combined cross section reported by ATLAS is $\sigma_{t\bar{t}} = 177 \pm 3\ (stat.)^{+8}_{-7}\ (syst.) \pm 7\ (lumi.)$ pb which is in good agreement with the SM expectation. Figure 5.27 summarizes the central value, the statistical and systematic uncertainties obtained in each channel as well as the combined cross section.

An important topic addressed in this thesis is the search for the Higgs boson produced in association with top quarks pair, which then decays into bottom-anti bottom final state. The first

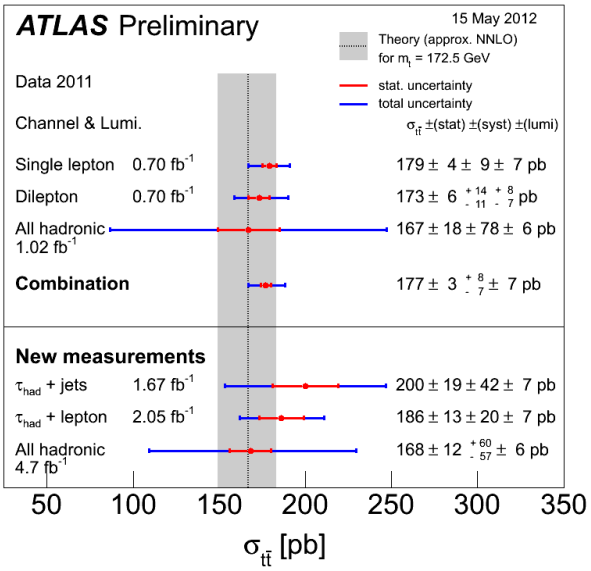


FIGURE 5.27 The measured value of $\sigma_{t\bar{t}}$ by the ATLAS experiment in various channels at $\sqrt{s} = 7$ TeV and the resulting combination. Error bars include statistical and systematic uncertainties. The approximate NNLO theoretical prediction together with its uncertainty is also shown.

analysis in ATLAS of the $t\bar{t}H$ in the multi-jet final state, described in this dissertation, makes use of the same dataset as the one used in the fully hadronic $t\bar{t}$ analysis. The aim of this work is to provide a first estimation of the potential that can be achieved by this analysis. It makes use of a kinematic likelihood fit for the $t\bar{t}$ system reconstruction and of a Boosted Decision Tree technique in order to discriminate between the main background source, QCD multi-jet production, and the $t\bar{t}$. The estimation of the QCD multi jet background is based on a data driven technique, "ABCD" method. The modeling of the other main physics background, $t\bar{t}$+jets, relies on the MC simulation. The missing part of this analysis is the estimation of its sensitivity and the limit setting. Unfortunately due to the lack of time it was not possible to provide in this dissertation a limit on the signal strength for the $t\bar{t}H$ in the fully hadronic final state. The analysis power can however be estimated by looking at the signal-over-square root of background achieved in the signal regions and making a comparison with the public results presented by the ATLAS collaboration for the semileptonic $t\bar{t}H(H \rightarrow b\bar{b})$. Table 5.18 shows the signal-over-square root of background in the validation and signal regions for the fully hadronic $t\bar{t}H(H \rightarrow b\bar{b})$, as well as the corresponding regions in the semileptonic analysis (the presence of at least eight jets in the fully hadronic final state corresponds to at least six jets in the semileptonic one). Moreover the performance of the fully hadronic analysis can also be quantified by estimating the fraction of the $t\bar{t}H$ with respect to the $t\bar{t}$+jets events yield. In the signal regions, $= 3$ b-tagged jets and ≥ 4 b-tagged jets, this fraction is about 1.3% and 6.7% respectively, instead the signal fractions in the semileptonic analysis are 0.7% and 4.1% in at least six jets with $= 3$ b-tagged jets and ≥ 4 b-tagged jets.

More studies are necessary in order to improve the modeling of the QCD multi-jet background and hence the data versus background prediction comparison. A disagreement of about 10% was

Signal region	$S/\sqrt{B}$	
	Hadronic	Semileptonic
$t\bar{t}$+jets and 2 b-tagged jets	0.07	0.32
$t\bar{t}$+jets and 3 b-tagged jets	0.16	0.16
$t\bar{t}$+jets and at least 4 b-tagged jets	0.19	0.28

TABLE 5.18 $S/\sqrt{B}$ for the fully hadronic and semileptonic $t\bar{t}H(H \to b\bar{b})$ for $\sqrt{s} = 7$ TeV analysis using 2011 dataset.

pointed out between the expected data and the total background events yield in the validation region. Moreover some important kinematic variables, such as the distance between the two jets used by the KLF as b-jets of the top quark decay, show a clear mismodeling. For this reason different BDT and "ABCD" method configuration should be investigated.

After the optimization of the background modeling, the next step for this analysis is to explore the low jet multiplicity bins in order to improve the sensitivity. The last optimization of the analysis will be the profiling likelihood of the systematic uncertainties in order to reduce the large systematics which affect the measurement.

The proton-proton collisions collected during the 2012 data-taking period at $\sqrt{s} = 8$ TeV should also be added to search for the fully hadronic $t\bar{t}H(H \to b\bar{b})$ decay channel. The main ingredient of the 2012 analysis is the b-jets trigger. As discussed in this dissertation, the good performance of the b-jet identification at the HLT level may lead to a significant improvement of the $t\bar{t}H(H \to b\bar{b})$ signal efficiency and as a consequence of the sensitivity of the analysis to this challenging decay channel.

Appendix A

ATLAS nomenclature

The coordinate system and nomenclature used to describe the ATLAS detector and the particles emerging from the proton-proton collisions are briefly summarised in this section. The coordinate system is based on the cartesian reference system where the nominal interaction point is defined as the origin of the coordinate system, while the beam direction defines the z-axis and the $x-y$ plane is transverse to the beam direction. The positive x-axis is defined as pointing from the interaction point to the centre of the LHC ring and the positive y-axis is defined as pointing upwards. The detector is divided in two sides, the side-A of the detector is defined as the one with positive z (toward the point 8) and side-C is with negative z (toward the point 2). The azimuthal angle ϕ is measured as usual around the beam axis between the x-axis and the selected point; the polar angle θ is the angle from the beam axis measured among the z-axis and the selected point (see Figure A.1.

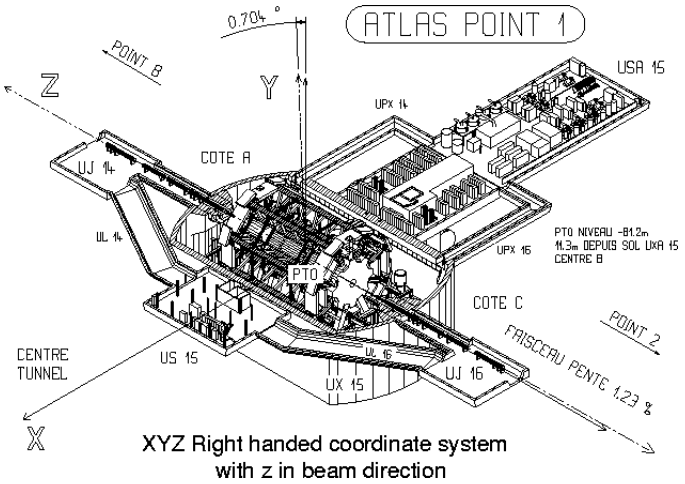


FIGURE A.1 The ATLAS coordinate system

The variable more used with respect the polar angle is the pseudorapidity. The pseudorapidity is a dimensionless quantity, defined as:

$$\eta = \ln tan(\frac{\theta}{2}) \tag{A.1}$$

It takes the zero value along the y-axis, and tends to $+\infty$ or $-\infty$ along the positive or negative z-axis respectively. The angular covarege of the detector usually is given in pseudorapidity units. For example the inner detector covers a pseudorapidity region $|\eta| < 2.5$ which corresponds to 9.4^o from the beam, the calorimeter covers a region of $|\eta| < 4.9$ which corresponds to 0.9^o from the beam. The pseudorapidity is the relativistic limit of the rapidity. In the limit where the particle is traveling close to the speed of light or in the approximation that the mass of the particle is nearly zero, the equation A is equivalent to rapidity formula: $y = \frac{1}{2} ln[\frac{(E+p_z)}{(E-p_z)}]$.

The transverse momentum p_T , the transverse energy E_T are defined in the $x - y$ plane.

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

$$E_T = \sqrt{E^2 - p_z^2} \tag{A.3}$$

where E is the total particle energy and p_x , p_y and p_z are the cartesian components of the particle three-momentum vector. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

Natural units $\hbar = c = 1$ are the defaults as usual in elementary particle physics.

Bibliography

[1] Glashow, S. L., "*Partial Symmetries of Weak Interactions*," *Nucl. Phys.* **22** (1961) 579–588.

[2] Weinberg, S., "*A Model of Leptons*," *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[3] Aitchison, I. J. R. and Hey, A. J. G., *Gauge Theories in Particle Physics, Vol.1: From Relativistic Quantum Mechanics to QED*. Taylor & Francis Group, 2003.

[4] Aitchison, I. J. R. and Hey, A. J. G., *Gauge Theories in Particle Physics, Vol.2: QCD and electoweak theory*. Taylor & Francis Group, 2003.

[5] Wilson, K. G., "*Confinement of quarks*," *Phys. Rev. D* **10** (Oct, 1974) 2445–2459.

[6] Gross, D. J. and Wilczek, F., "*Asymptotically Free Gauge Theories. I*," *Phys. Rev. D* **8** (Nov, 1973) 3633–3652.

[7] **Particle Data Group** Collaboration, Beringer, J. and others, "*Review of Particle Physics*," *Phys. Rev. D* **86** (Jul, 2012) 010001.

[8] Tadao, N. and Kazuhiko, N., "*Charge Independence for V-particles*," *Progress of Theoretical Physics* **10** no. 5, (1953) 581–582.

[9] Salam, A., "*Weak and Electromagnetic Interactions*," *Conf. Proc.* **C680519** (1968) 367 – 377.

[10] Feynman, R. P., "*Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*," *Phys. Rev.* **80** (Nov, 1950) 440–457.

[11] Weinheimer, C. and Zuber, K., "*Neutrinos Masses*," *Annalen der Physik* **525** (2013) 565 – 575.

[12] Dummit, D. S.and Foote, R. M., *Abstract Algebra*. John Wiley & Sons, 2004.

[13] Han, M. Y. and Nambu, Y., "*Three Triplet Model with Double* SU(3) *Symmetry*," *Phys. Rev.* **139** (Aug, 1965) B1006–B1010.

[14] Gell-Mann, M., "*A Schematic Model of Baryons and Mesons*," *Phys. Lett.* **8** (1964) 214–215.

[15] Zweig, G., "*An SU(3) model for strong interaction symmetry and its breaking. Part I*,". CERN-TH-401.

[16] Zweig, G., "*An SU(3) model for strong interaction symmetry and its breaking; Part II*,". CERN-TH-412.

[17] Higgs, P. W., "*Broken symmetries, massless particles and gauge fields*," *Phys. Lett.* **12** (1964) 132–133.

[18] Higgs, P. W., "*Broken symmetries and the masses of gauge bosons*," *Phys. Rev. Lett.* **13** (1964) 508–509.

[19] Higgs, P. W., "*Spontaneous symmetry breakdown without massless bosons*," *Phys. Rev.* **145** (1966) 1156–1163.

[20] Englert, F. and Brout, R., "*Broken Symmetry and the Mass of Gauge Vector Mesons*," *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[21] Goldstone, J., "*Field theories with Superconductor solutions*," *Il Nuovo Cimento* **19** no. 1, (1961) 154–164.

[22] Goldstone, J. and Salam, A. and Weinberg, S., "*Broken Symmetries*," *Phys. Rev.* **127** (Aug, 1962) 965–970.

[23] Weinberg, S., The Quantum Theory of Fields, Vol.1: Foundations. Cambridge University Press, New York, 1995.

[24] **ALEPH and CDF and D0 and DELPHI and L3 and OPAL and SLD Collaboration** Collaboration, Alcaraz, J., "Precision Electroweak Measurements and Constraints on the Standard Model." CERN-PH-EP-2009-023, FERMILAB-TM-2446-E, LEPEWWG-2009-01, TEVEWWG-2009-01, ALEPH-2009-001-PHYSICS-2009-001, CDF-NOTE-9979, D0-NOTE-6005, DELPHI-DELPHI-2009-PHYS-951, L3-NOTE-2836, OPAL-PR430, SLAC-PUB-13830, 2009.

[25] Djouadi, A. and Spira, M. and Zerwas, P. M., "*Production of Higgs bosons in proton colliders: QCD corrections*," *Phys. Lett.* **B264** (1991) 440 – 446.

[26] Spira, M. and Djouadi, A. and Graudenz, D. and Zerwas, P. M., "*Higgs boson production at the LHC*," *Nucl. Phys.* **B453** (1995) 17–82.

[27] Dittmaier, S. and Mariotti, C. and Passarino, G. and Tanaka, R., "*Report of the LHC Higgs Cross Section Working Group*," *CERN-2011-002* (2011) .

[28] de Florian, D. and Grazzini, M., "*Higgs production at the LHC: updated cross sections at* $\sqrt{s} = 8$ *TeV*," arXiv:1206.4133 [hep-ex].

[29] 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** no. 1, (2012) 1 – 29.

[30] CMS Collaboration, "*Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*," *Phys. Lett. B* **716** no. 1, (2012) 30 – 61.

[31] ATLAS Collaboration, "*Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data*,". ATLAS-CONF-2013-014.

[32] ATLAS Collaboration, "*Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25* fb^{1} *of proton-proton collision data*,". ATLAS-CONF-2013-013.

[33] CMS Collaboration, "*Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at $\sqrt{s} = 7$ and 8 TeV*,". CMS-PAS-HIG-13-002.

[34] ATLAS Collaboration, "*Study of the spin properties of the Higgs-like particle in the $H \to WW^{(*)} \to e\nu\mu\nu$ channel with 21 fb^{-1} of $\sqrt{s} = 8$ TeV data collected with the ATLAS detector*,". ATLAS-CONF-2013-031.

[35] ATLAS Collaboration, "*Study of the spin of the Higgs-like boson in the two photon decay channel using 20.7* fb^{-1} *of pp collisions collected at $\sqrt{s} = 8$ TeV with the ATLAS detector*,". ATLAS-CONF-2013-029.

[36] CMS Collaboration, "*Evidence for a particle decaying to W+W- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at the LHC*,". CMS-PAS-HIG-13-003.

[37] Plehn, T. and Rauch, M., "*Higgs Couplings after the Discovery*," *Eur. Phys. Lett.* **100** (2012) 11002, arXiv:1207.6108 [hep-ph].

[38] CDF Collaboration, "*Observation of top quark production in $\bar{p}p$ collisions*," *Phys. Rev. Lett.* **74** (1995) 2626–2631, arXiv:hep-ex/9503002 [hep-ex].

[39] D0 Collaboration, "*Observation of the top quark*," *Phys. Rev. Lett.* **74** (1995) 2632–2637, arXiv:hep-ex/9503003 [hep-ex].

[40] CDF and D0 Collaborations, "*Combination of CDF and D0 results on the mass of the top quark using up to* 5.8 fb^{-1} *of data*," arXiv:1107.5255 [hep-ex]. FERMILAB-TM-2504-E, CDF-NOTE-10549, D0-NOTE-6222.

[41] ATLAS Collaboration, "*Measurement of the Top Quark Mass from $\sqrt{s} = 7$ TeV ATLAS Data using a 3-dimensional Template Fit*,". ATLAS-CONF-2013-046.

[42] CMS Collaboration, "*Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s}$=7 TeV*," *JHEP* **12** (Sep, 2012) 105. 33 p, arXiv:1209.2319. CMS-TOP-11-015. CERN-PH-EP-2012-250.

[43] Plehn, T., "*Lectures on LHC Physics*," *Lect. Notes Phys.* **844** (2012) 1–193, arXiv:0910.4182 [hep-ph].

[44] Nadolsky, Pavel M. and Lai, Hung-Liang and Cao, Qing-Hong and Huston, Joey and Pumplin, Jon and Stump, Daniel and Tung, Wu-Ki and Yuan, C.-P., "*Implications of CTEQ global analysis for collider observables*," *Phys. Rev. D* **78** (Jul, 2008) 013004.

[45] The Durham HepData Project, "http://hepdata.cedar.ac.uk/pdf/pdf3.html.".

[46] Czakon, M. and Fiedler, P. and Mitov, A., "*The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*." CERN-PH-TH-2013-056, TTK-13-08, 2013.

[47] Kidonakis, N., "*NNLL threshold resummation for top-pair and single-top production*," arXiv:1210.7813 [hep-ph].

[48] Cabibbo, N., "*Unitary Symmetry and Leptonic Decays*," *Phys.Rev.Lett.* **10** (1963) 531–533.

[49] Kobayashi, M. and Maskawa, T., "*CP Violation in the Renormalizable Theory of Weak Interaction*," *Prog. Theor. Phys.* **49** (1973) 652–657.

[50] ATLAS Collaboration, "*Determination of the Top Quark Mass with a Template Method in the All Hadronic Decay Channel using 2.04 fb^{-1} of ATLAS Data*,". ATLAS-CONF-2012-030.

[51] ATLAS Collaboration, "*Measurement of the Top Quark Mass in Dileptonic Top Quark Pair Decays with $\sqrt{s} = 7$ TeV ATLAS Data*,". ATLAS-CONF-2013-077.

[52] CDF and DO Collaborations, "*Combination of CDF and DO results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron*," arXiv:1305.3929 [hep-ph]. ATLAS-CONF-2012-030.

[53] ATLAS and CMS Collaborations, "*Combination of ATLAS and CMS results on the mass of the top-quark using up to 4.9 fb^1 of $\sqrt{s} = 7$ TeV LHC data*,". ATLAS-CONF-2013-046.

[54] D0 Collaboration, "*Experimental discrimination between charge 2e/3 top quark and charge 4e/3 exotic quark production scenarios*," *Phys.Rev.Lett.* **98** (2007) 041801, arXiv:hep-ex/0608044 [hep-ex].

[55] ATLAS Collaboration, "*Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*,". ATLAS-CONF-2011-141.

[56] CMS Collaboration, "*Constraints on the Top-Quark Charge from Top-Pair Events*,". CMS-PAS-TOP-11-031.

[57] Marshak, R. E., "*Chirality Invariance and the Universal V-A Theory of Weak Interactions*," *World Scientific Publishing Company* (1991) 18 – 42.

[58] CMS Collaboration, "*Measurement of W-Polarization in Di-Leptonic $t\bar{t}$ Events in pp Collisions with $\sqrt(s) = 7$ TeV*,". CMS-PAS-TOP-12-015.

[59] ATLAS and CMS Collaborations, "*Combination of the ATLAS and CMS measurements of the W-boson polarization in top-quark decays*,". ATLAS-CONF-2013-033.

[60] ATLAS Collaboration, "*Observation of Spin Correlation in $t\bar{t}$ Events from pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector*," *Phys. Rev. Lett.* **108** (May, 2012) 212001.

[61] CMS Collaboration, "*Measurement of Spin Correlations in ttbar production*,". CMS-PAS-TOP-12-004.

[62] Stange, A. and Willenbrock, S., "*Yukawa correction to top-quark production at the Fermilab Tevatron*," *Phys. Rev. D* **48** (Sep, 1993) 2054–2061.

[63] Feng, Tai-Fu and Li, Xue-Qian and Maalampi, Jukka, "*Anomalous Higgs top-quark couplings in the MSSM*," *Phys. Rev. D* **69** (Jun, 2004) 115007.

[64] Aguilar - Saavedra, J.A., "*Light Higgs boson discovery from fermion mixing*," *JHEP* **0612** (2006) 033, arXiv:hep-ph/0603200 [hep-ph].

[65] Brüning, O. S. and Collier, P. and Lebrun, P. and Myers, S. and Ostojic, R. and Poole, J. and Proudlock, P., *LHC Design Report*. CERN, Geneva, 2004.

[66] ATLAS Collaboration, "https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults#2011_pp_Collisions," 2011.

[67] CMS Collaboration, *Technical proposal*. LHC Tech. Proposal. CERN, Geneva, 1994. Cover title : CMS, the Compact Muon Solenoid : technical proposal.

[68] ALICE Collaboration, *ALICE Technical Design Report*. Tech. Proposal. CERN, Geneva, 2001.

[69] LHCb Collaboration, *LHCb: Technical Proposal*. Tech. Proposal. CERN, Geneva, 1998.

[70] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.

[71] ATLAS Collaboration, *ATLAS inner detector: Techincal Design Report,1*. Technical Design Report ATLAS. CERN, Geneva, 1997.

[72] ATLAS Collaboration, *ATLAS inner detector: Techincal Design Report,2*. Technical Design Report ATLAS. CERN, Geneva, 1997.

[73] ATLAS Collaboration, *ATLAS pixel detector: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1998.

[74] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.

[75] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. distribution.

[76] ATLAS Collaboration, *ATLAS Computing: technical design report*. CERN, Geneva, 2005.

[77] Aderholz, A. and others, "*Models Of Networked Analysis At Regional Centres For LHC Experiments (monarc), Phase 2 Report*,". CERN-LCB-2000-001. KEK-2000-8.

[78] ATLAS Collaboration, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003.

[79] ATLAS Collaboration, Expected performance of the ATLAS experiment: detector, trigger and physics. CERN, Geneva, 2009.

[80] ATLAS Collaboration, "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV," *Eur. Phys. J.* **C73** (2013) 2304, arXiv:1112.6426 [hep-ex].

[81] Tamsett, M. and Agustoni, M. and Begel, M. and Conde Muino, P. and Damazio, D. and Rubbo, F. and Sandoval, C. and Santamarina Rios, C. and Sutton, M. and Subramaniam, R., "*The L1.5 Jet Trigger*,". ATL-COM-DAQ-2012-015.

[82] Begel, M. and Campanelli, M. and Chislett, R and Kasieczka, G. and Machado Miguens, J. and Vaz Gil Lopes, L., "*Jet Trigger Performance with 2011 Data*,". ATL-COM-DAQ-2011-063.

[83] Bernat, P., "*Architecture and Performance of the Inner Detector Trigger of the ATLAS detector*,". ATL-DAQ-PROC-2012-032.

[84] Cornelissen, T. and Elsing, M. and Fleischmann, S. and Liebig, W. and Moyse, E. and Salzburger, A., "*Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*,". ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002.

[85] Coccaro, A. and Hansson, P. and Ferreira de Lima, D. E., "*b-jet trigger plots for EPS 2011*,". ATL-COM-DAQ-2011-052.

[86] Buskulic, D. and others, "*A precise measurement of* $\Gamma(Z \to bb/\Gamma(Z \to hadrons)$," *Physics Letters B* **313** no. 34, (1993) 535 – 548.

[87] D0 Collaboration, "*Search for $Wb\bar{b}$ and WH Production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*," *Phys. Rev. Lett.* **94** (Mar, 2005) 091802.

[88] CDF Collaboration, "*Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using lepton + jets events with jet probability b-tagging*," *Phys. Rev. D* **74** (Oct, 2006) 072006.

[89] Ferreira de Lima, D. E., "*b-jet tagger data to simulation comparison using 8 TeV data*,". ATL-COM-DAQ-2013-018.

[90] ATLAS Collaboration, "*Measurement of the $t\bar{t}$ production cross section in the tau + jets channel using the ATLAS detector*," *Eur. Phys. J. C* **73** no. 3, (2013) .

[91] Coccaro, A. and Hansson, P., "*b-jet trigger rate plots*,". ATL-COM-DAQ-2011-067.

[92] ATLAS Collaboration, "*Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with 5 fb^1 of Data from the ATLAS Detector*,". ATLAS-CONF-2012-043.

[93] Cacciari, M. and Salam, G. P., "*Dispelling the myth for the jet-finder*," *Phys. Lett. B* **641** no. 1, (2006) 57 – 61.

[94] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[95] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms," *JHEP* **9708** (1997) 001, arXiv:hep-ph/9707323 [hep-ph].

[96] ATLAS Collaboration, "*ATLAS high-level trigger, data acquisition and controls: Technical design report*,". CERN-LHCC-2003-022, ATLAS-TRD-016.

[97] Lopes, R. H. C. and Hobson, P. R. and Reid, I. D., "*Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test*," *J. Phys. Conf. Ser.* **119** (2008) 042019.

[98] "The Qt-Cross Platform Application." http://doc.trolltech.com/.

[99] "The Extensible Markup Language." http://www.w3.org/XML/.

[100] ATLAS Collaboration, "*Measurement of the $t\bar{t}$ production cross section in the all-hadronic channel in ATLAS with $\sqrt{s} = 7$ TeV data*,". ATLAS-CONF-2012-031.

[101] Frixione, S. and Webber, B. R., "*Matching NLO QCD computations and parton shower simulations*," *JHEP* **0206** (2002) 029, arXiv:hep-ph/0204244 [hep-ph].

[102] Nadolsky, P. M. and Lai, H. L. and Cao, Q. H. and Huston, J. and Pumplin, J. and others, "*Implications of CTEQ global analysis for collider observables*," *Phys. Rev.* **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[103] Lai, H. L. and Guzzi, M. and Huston, J. and Li,Z. and Nadolsky, P. M. and others, "*New parton distributions for collider physics*," *Phys.Rev.* **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[104] Agostinelli, S. and others, "*Geant4: a simulation toolkit*," *Nucl. Instr. Meth. Phys. Res. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** no. 3, (2003) 250 – 303.

[105] Corcella, C. and others, "*HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*," *JHEP* **01** (2001) 010.

[106] S. Badger, B. Biedermann, P. Uwer, and V. Yundin, "NLO QCD corrections to multi-jet production at the LHC with a centre-of-mass energy of $\sqrt{s} = 8$ TeV," *Phys. Lett.* **B718** (2013) 965–978, arXiv:1209.0098 [hep-ph].

[107] Mangano, M. L. and Moretti, M. and Piccinini, F. and Pittau, R. and Polosa, A. D., "*ALPGEN, a generator for hard multiparton processes in hadronic collisions*," *JHEP* **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph].

[108] Cacciari, M. and Salam, A. and Gavin P. and Soyez, G., "*The Anti-k_T jet clustering algorithm*," *JHEP* **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[109] ATLAS Collaboration, "*Measurement of the cross section for top-quark pair production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector using final states with two high-pt leptons*," *JHEP* **1205** (2012) 059, arXiv:1202.4892 [hep-ex].

[110] ATLAS Collaboration, "Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS," *Eur. Phys. J.* **C72** (2012) 1844, arXiv:1108.5602 [hep-ex].

[111] Piacquadio, G. and Weiser, C., "*A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*," *J. Phys. Conf. Ser.* **119** no. 3, (2008) 032032.

[112] ATLAS Collaboration., "*Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*,". ATLAS-CONF-2011-102.

[113] ATLAS Collaboration, "*Calibrating the b-Tag Efficiency and Mistag Rate in* 35 pb^{-1} *of Data with the ATLAS Detector*,". ATLAS-CONF-2011-089.

[114] D0 Collaboration, "*b-Jet Identification in the D0 Experiment*," *Nucl. Instrum. Meth.* **A620** (2010) 490–517, arXiv:1002.4224 [hep-ex].

[115] ATLAS Collaboration, "*Measurement of the production cross section for W-bosons in association with jets in pp collisions at with the ATLAS detector*," *Phys. Let. B* **698** no. 5, (2011) 325 – 345.

[116] ATLAS Collaboration, "*Measurement of the production cross section of jets in association with a Z boson in pp collisions at* $\sqrt{(s)} = 7$ *TeV with the ATLAS detector*," *JHEP* **07** (2013) 032, arXiv:1304.7098 [hep-ex].

[117] ATLAS Collaboration, "*Search for the Standard Model Higgs boson produced in association with top quarks in proton-proton collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*,". ATLAS-CONF-2012-135.

[118] ATLAS Collaboration, "*Measurement of multi-jet cross sections in proton-proton collisions at a 7 TeV center-of-mass energy*," *Eur. Phys. J.* **C71** (2011) 1763, arXiv:1107.2092 [hep-ex].

[119] Barlow, R., "*Extended maximum likelihood*," *Nucl. Instrum. Meth. Phys. Res. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **297** no. 3, (1990) 496 – 506.

[120] ATLAS Collaboration, "*Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy* $\sqrt{s} = 7$ *TeV*,". ATLAS-CONF-2010-054.

[121] Kersevan, B. P. and Richter-Was, E., "*The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5*." TPJU-6-2004, 2004.

[122] Martin, A. D. and Stirling, W. J. and Thorne, R.S. and Watt, G., "*Parton distributions for the LHC*," *Eur. Phys. J.* **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[123] Martin, A.D. and Stirling, W.J. and Thorne, R.S. and Watt, G., "*Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections*," *Eur. Phys. J.* **C64** (2009) 653–680, arXiv:0905.3531 [hep-ph].

[124] Frixione, S. and Nason, P. and Oleari, C., "*Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*," *JHEP* **0711** (2007) 070, arXiv:0709.2092 [hep-ph].

[125] Sjostrand, T. and Mrenna, S. and Skands, P. Z., "*PYTHIA 6.4 Physics and Manual*," *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[126] ATLAS Collaboration, "*Luminosity Determination in pp Collisions at* $\sqrt{(s)} = 7$ *TeV Using the ATLAS Detector at the LHC*," *Eur. Phys. J. C: Particles and Fields* **C71** (2011) . Article Number 1630.

[127] 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** no. 8, (2013) 1 – 39.

[128] Barlow, R., "*Asymmetric Errors*," arXiv:physics/0406120 [hep-ph].

[129] CMS Collaboration, "*Search for Higgs Boson Production in Association with a Top-Quark Pair and Decaying to Bottom Quarks or Tau Leptons*,". CMS-PAS-HIG-13-019.

[130] Sherstnev, A. and Thorne, R. S., "*Parton Distributions for LO Generators*," *Eur. Phys. J.* **C55** (2008) 553–575, arXiv:0711.2473 [hep-ph].

[131] **LHC Higgs Cross Section Working Group** Collaboration, Dittmaier, S. and others, "*Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*," arXiv:1101.0593 [hep-ph].

[132] Mangano, M. L. and Moretti, M. and Pittau, M., "*Multijet matrix elements and shower evolution in hadronic collisions: $Wbb+n$-jets as a case study*," *Nucl. Phys. B* **632** no. 1 - 3, (2002) 343 – 362.

[133] Aliev, M. and Lacker, H. and Langenfeld, U. and Moch, S. and Uwer, P. and others, "*HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR*," *Comput.Phys.Commun.* **182** (2011) 1034–1046, arXiv:1007.1327 [hep-ph].

[134] Butterworth, J. M. and Forshaw, Jeffrey R. and Seymour, M. H., "*Multiparton interactions in photoproduction at HERA*," *Z. Phys.* **C72** (1996) 637–646, arXiv:hep-ph/9601371 [hep-ph].

[135] Golonka, P. and Was, Z., "*PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*," *Eur. Phys. J.* **C45** (2006) 97–107, arXiv:hep-ph/0506026 [hep-ph].

[136] Jadach, S. and Kuhn, J. H. and Was, Z., "*TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons*," *Comput. Phys. Commun.* **64** (1990) 275–299.

[137] ATLAS Collaboration, "*The ATLAS Simulation Infrastructure*," *Eur. Phys. J. C* **70** no. 3, (2010) 823–874.

[138] Byron, P. R. and others, "*Boosted deciosion trees as an alternative to artificial neural networks for particles identification*," *Nuclear Instrumentation and Methods in Physics Research Section A: Accelerator, Spectrometers, Detectors and Associated Equipment* **543** no. 2-3, (2005) 577–584.

[139] Hoecker, A. and others, "*TMVA - Toolkit for Multivariate Data Analysis*," arXiv:physics/0703039.

[140] Banfi, A. and Gavin P. and Zanderighi, G., "*Phenomenology of event shapes at hadron colliders*," *JHEP* **1006** (2010) 038, arXiv:1001.4082 [hep-ph].

[141] ATLAS Collaboration, "*Improved Luminosity Determination in* **pp** *Collisions at* $\sqrt{\mathbf{s}} = \mathbf{7}$ *TeV using the ATLAS Detector at the LHC*,". ATLAS-CONF-2012-080.

[142] ATLAS Collaboration, "*Measurement of the jet multiplicity in top anti-top final states produced in 7 TeV proton-proton collisions with the ATLAS detector*,". ATLAS-CONF-2012-155.

[143] Conway, J. S., "*Incorporating Nuisance Parameters in Likelihoods dor Multisource Spectra*," arXiv:hep-ex/1103.0354 [hep-ex].

Abstract

Probing top quark and Higgs boson production in multi-jet events at the LHC with the ATLAS detector

The ATLAS detector is a general purpose detector located at the Large Hadron Collider (LHC). It aims at the discovery of new physics phenomena and improving our knowledge of the Standard Model (SM). The LHC is an hadron collider designed to provide proton proton collisions at 14 TeV center-of-mass energy and $10^3 4\text{cm}^2\text{s}^1$ peak luminosity. The ATLAS experiment collected 4.7 fb^1 pp collisions delivered by the LHC at a centre-of-mass energy of 7 TeV in 2011 and about 20 fb^{-1} pp collisions in 2012 at a centre-of-mass energy of 8 TeV.

The thesis reports the measurement of top-anti top cross section performed using the 4.7 fb^{-1} data recorded by the ATLAS detector during the 2011 data-taking campaign with a center-of-mass energy of 7 TeV. The top-anti top pair decay mode investigated is the all-hadronic, where both the W bosons, produced in the top quark disintegration, decay in a quarks pair. The all hadronic decay mode has the advantage of a high branching fraction, about 46 %, but on the other hand it suffers from an irreducible QCD multi-jet background. The main background is coming from the QCD processes with at least six quaks or gluons in the final state.This processes is estimated using dedicated data-driven techniques. The measured total cross section is 168 ± 12 (stat.)$^{+60}_{-57}$ (syst.) $\pm$ 7 (lum.) pb.

The thesis presents also the first search for the Higgs boson produced in the association with a top quark pair, where the Higgs boson decays in a bottom quark pair and the top-anti top pair in the fully hadronic mode. The analysis is performed on the same data sample used in the top-anti top cross section measurement. The main background of the $t\bar{t}H$ signal coming from the $t\bar{t} + jets$ production and from the QCD multi-jet process. The former is estimated using the Monte-Carlo simulation, instead the latter using a dedicated data-driven technique. The aim of this study it is to show the power of fully hadronic channel looking at the signal-over-square root of background achieved in the signal regions and making a comparison with the public results presented by the ATLAS in the semileptonic channel.

Key-words: LHC, ATLAS, top quarks, Higgs boson

Résumé

Sonder la production du quark top et du boson de Higgs dans les événements multi-jet avec expérience ATLAS auprès du LHC

Le détecteur ATLAS est un expérience généraliste placé auprès du collisionneur proton-proton circulaire de 27 Km de circonférence, LHC. Le LHC est conçu pour produire des collisions avec une énergie nominale au centre de masse de l'ordre de $\sqrt{s} = 14$ TeV et une luminosité instantané de $10^{-34} cm^{-2} s^{-1}$. Il donne accès à des processus á l'échelle du TeV. En 2010 et 2011 le LHC a fournit des collisions proton-proton avec une énergie dans le centre de mass de $\sqrt{s} = 7$ TeV donnant la possibilité d'enregistrer plus de 5 fb^{-1} de données. En 2012 le LHC a ensuite fournit des collisions avec une énergie dans le centre de mass de $\sqrt{s} = 8$ TeV.

La présente thèse traite de la mesure de la section efficace de production des paires de quarks top anti-top dans le mode de désintégration complètement hadronique utilisant les données enregistrées par l'expérience ATLAS en 2011 avec une énergie dans le centre de masse de $\sqrt{s} = 7$ TeV correspondant á une luminosité intégrée de 4.7 fb^{-1}. Le canal complètement hadronique a l'avantage d'être caractérisé par un rapport d'embranchement de 46%. Il souffre par contre d'un bruit de fond multi-jet QCD élevé. Le principal bruit de fond pour la production des paires des quark top-antitop est dû aux processus QCD avec au mois six de quarks/gluons dans l'état final. Ce bruit de fond n'est malheureusement pas bien connu et donc difficile á reproduire par simulation Monté Carlo. Il est estimé á partir des données. La section efficace ainsi mesuré est 168 ± 12 (stat.)$^{+60}_{-57}$ (syst.) ± 7 (lum.) pb.

La deuxième analyse presentée dans la cette thèse se foncalise sur la recherche du boson du Higgs produit en association avec une paire de quark top, où le boson de Higgs se désintègre en paires de quarks b, et le système top-anti-top se désintégrant dans le canal complétant hadronique. L'analyse est effectuée sur les mêmes données utilisées par la mesure de la section efficace top-anti top. Deux bruits de fond principaux peuvent être distingués: la production des paires de quarks top avec la présence des jets supplémentaires et la production QCD. Le premier bruit de fond est estimé en utilisant le modèle Monte-Carlo différenciant la production de jets légers des jets lourds. Le deuxième bruit de fond est estimé avec les données. Le but de cette étude est de montrer le potential du canal complétant hadronique en regardant le signal sur la racine du bruit du fond dans les

régions de signal et en faisant une comparaison avec les résultats publics présentés par l'ATLAS dans le canal semileptonic.

Mots-clefs : LHC, ATLAS, top quarks, Higgs boson