Ph.D. thesis

February 2014

Doctorado en Física

Departamento de Física,
Universitat Autònoma de Barcelona
y
Institut de Física d'Altes Energies

Probing new physics at the LHC: searches for heavy top-like quarks with the ATLAS experiment

Doctoranda
Antonella Succurro

Director
Prof. Aurelio Juste Rozas
Instituciò Catalana de Recerca i Estudis Avancat
Institut de Física d'Altes Energies

Tutora
Prof. María Pilar Casado Lechuga
Universitat Autònoma de Barcelona
Institut de Física d'Altes Energies

@ Malpensa airport, Milan (Italy)

Contents

Introduction **1**

Part I **5**

1 Theoretical framework **7**
1.1 The Standard Model from 1963 to 2012 . . . 7
1.1.1 Building the Standard Model . . . 11
1.2 Unanswered questions and new physics quests . . . 16
1.2.1 Supersymmetry . . . 18
1.2.2 Fourth generation SM4 . . . 19
1.2.3 Composite and Little Higgs . . . 20
1.2.4 Extra-dimensions . . . 21
1.3 Going beyond the SM with vector-like quarks . . . 23
1.3.1 Production . . . 23
1.3.2 Decay . . . 24

2 The ATLAS experiment at the Large Hadron Collider **27**
2.1 The Large Hadron Collider at CERN . . . 27
2.2 The ATLAS detector . . . 31
2.2.1 Coordinate system . . . 32
2.2.2 Magnets . . . 34
2.2.3 Inner detector . . . 34
2.2.4 Calorimeters . . . 37
2.2.5 Muon spectrometer . . . 41
2.3 Forward sub-detectors . . . 42
2.4 Trigger system . . . 43
2.4.1 Level 1 trigger . . . 44
2.4.2 Level 2 trigger . . . 45
2.4.3 Event filter . . . 45

2.5 Data Quality . . . 45

3 Monte Carlo simulation **47**
3.1 Phenomenology of pp collisions . . . 47
3.1.1 Proton structure . . . 48
3.1.2 Factorization theorem . . . 49
3.2 Simulation of pp collisions . . . 50
3.2.1 Hard interaction . . . 50
3.2.2 Parton shower . . . 52
3.2.3 Hadronization . . . 55
3.2.4 Underlying event . . . 56
3.3 Generators . . . 57
3.4 ATLAS detector simulation . . . 59
3.5 Monte Carlo samples corrections . . . 59

4 Event reconstruction **61**
4.1 ID Tracks . . . 61
4.2 Primary vertices . . . 62
4.3 Energy clusters . . . 63
4.4 Electrons . . . 64
4.5 Muons . . . 66
4.6 Jets . . . 68
4.6.2 *b*-tagging . . . 70
4.7 Missing Transverse Energy . . . 72

Part II **73**

5 Searches for vector-like top partner pairs in the single lepton channel **75**
5.1 General strategy for vector-like quark pairs searches . . . 75
5.2 Data sample and common event preselection . . . 78
5.3 Background and signal modeling . . . 80
5.3.1 Monte Carlo simulated samples . . . 80
5.3.2 W+jets background normalisation . . . 82
5.3.3 Multi-jet background . . . 84
5.4 Data to Monte Carlo comparison . . . 86
5.5 Systematic uncertainties . . . 86
5.5.1 Luminosity . . . 88
5.5.2 Object definitions . . . 88
5.5.3 Theoretical cross-sections . . . 90
5.5.4 Normalizations of data-driven backgrounds and background modeling 91
5.6 Statistical analysis . . . 91

6 Search for $T\bar{T}$ decaying to $Wb + X$ **93**
6.1 Analysis strategy . . . 93
6.2 W boson reconstruction . . . 94
6.3 Event selection . . . 95
6.4 Control regions . . . 98
6.5 Final discriminant: heavy quark reconstructed mass . . . 101
6.6 Systematic uncertainties . . . 103
6.6.1 Merging of non-$t\bar{t}$ Backgrounds . . . 103
6.6.2 Jet energy scale . . . 106
6.6.3 Jet energy resolution . . . 106
6.6.4 Jet mass scale and resolution . . . 107
6.6.5 W/Z+jets Normalisation . . . 107
6.6.6 $t\bar{t}$ Modelling . . . 107
6.6.7 V+jets Modelling . . . 108
6.6.8 Overall effect of systematic uncertainties . . . 108
6.7 Results . . . 108

7 Search for $T\bar{T}$ decaying to $Ht + X$ **113**
7.1 Analysis strategy . . . 113
7.2 Event selection . . . 114
7.3 Control regions . . . 116
7.4 Final discriminant: event transverse momentum . . . 117
7.5 Systematic uncertainties . . . 118
7.5.1 Jet energy scale . . . 119
7.5.2 Normalization of backgrounds . . . 120
7.5.3 $t\bar{t}$+jets Modelling . . . 120
7.5.4 $t\bar{t}$+jets Heavy-Flavour Content . . . 121
7.5.5 Overall effect of systematic uncertainties . . . 122
7.6 Results . . . 122

8 Final results **125**
8.1 Combination of the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses . . . 125
8.2 Potential coverage of the mixing plane . . . 127

Conclusions **131**

A Multi-jet background estimation in the single muon plus jets channel **133**

B Data to Monte Carlo comparison in the preselection region **139**
B.1 Data to Monte Carlo comparison vetoing $b$jets . . . 139
B.1.1 Electron channel . . . 140
B.1.2 Muon channel . . . 141
B.1.3 Electron+Muon channel . . . 142
B.2 Data to Monte Carlo comparison requiring at least 1 $b$jet . . . 142
B.2.1 Electron channel . . . 142

B.2.2 Muon channel . . . 144
B.2.3 Electron+Muon channel . . . 145

C The Tag Rate Function method **147**
C.1 TRF method principle . . . 147
C.1.1 Validation . . . 148
C.2 TRF in the $T\bar{T} \to Ht + X$ analysis . . . 149
C.3 TRF in the $T\bar{T} \to Wb + X$ analysis . . . 150

D Search for $T\bar{T} \to WbWb$ with pp collisions at $\sqrt{s}$= 7 TeV **153**
D.1 Jet mass related systematic uncertainties . . . 153

E $T\bar{T} \to Wb + X$ analysis: SR cut-flow **155**
E.1 Event yields in the electron and muon channels . . . 156
E.2 Reconstructed mass in the SRs . . . 156

F $T\bar{T} \to Wb + X$ analysis: data to background comparison in SDRs **159**
F.1 Data to background comparison in SDR0 . . . 159
F.2 Data to background comparison in SDR1 . . . 161
F.3 Data to background comparison in SDR2 . . . 163
F.4 Data to background comparison in SDR3 . . . 165
F.5 Data to background comparison in SDR4 . . . 167
F.6 Data to background comparison in SDR5 . . . 169
F.7 Data to background comparison in SDR6 . . . 171
F.8 Data to background comparison in SDR7 . . . 173
F.9 Data to background comparison in SDR8 . . . 175
F.10 Data to background comparison in SDR9 . . . 177

G $T\bar{T} \to Wb + X$ analysis: search configuration **179**

H $T\bar{T} \to Ht + X$ analysis: data to background comparison in control regions **181**
H.1 e+jets Channel . . . 181
H.2 μ+jets Channel . . . 185
H.3 Combined e+jets and μ+jets Channels . . . 189
H.4 H_{T} tail dedicated control regions . . . 193

I $T\bar{T} \to Ht + X$ analysis: comparison of signal prediction between singlet and doublet scenarios **195**

Bibliography **197**

Introduction

July 4th, 2012, represents a milestone for high-energy physics, being the date when the ATLAS and CMS experiments at CERN announced the discovery of a new particle consistent with a Standard Model Higgs boson with mass $m_H \sim 125$ GeV. Is it going to be celebrated, 20 years from now, as the beginning of a new era of discoveries or as the end of the adventure? Is there something *more*, out there, in the outer space or 100 m underground, awaiting for being discovered? As outlined in Chapter 1 there is quite some evidence something must be there lying "beyond the Standard Model". A successful theory finally completed by the identification of the Higgs boson, the Standard Model as it is still leaves too many questions unanswered. What is "Dark Matter", this exotic form of matter different from atoms, being immune to electromagnetic interactions, but which is known to account for ∼27% of the total matter in the Universe? After the Big Bang, what caused the asymmetry in particles and antiparticles production that made matter prevail over antimatter? Why is the top quark so much heavier than the other quarks? Why is the Higgs boson so much lighter than the Planck mass?

It was to find an explanation to this puzzle that the Large Hadron Collider project was initiated 20 years ago. The ATLAS collaboration then started the design of the detector described in Chapter 2, outlining a challenging physics program in which the search for the Higgs boson was "just" the first bullet of the list. The LHC first run was on the 20th of November 2009, with the ATLAS experiment beginning to record data from these early proton-proton collisions at a center of mass energy of 900 GeV just three days later. Since then outstanding performances of both the accelerator and the detector allowed to collect a huge amount of data from proton-proton collisions at increasing center of mass energies, reaching in 2012 a total of ∼20 fb^{-1} at a center of mass energy of 8 TeV.

However, this large amount of data alone would not tell much if it were not possible to compare them to precise theoretical predictions. Chapter 3 describes the Monte Carlo techniques used to obtain simulated samples of either "Standard Model" or "new physics" events. Starting from the computation of the matrix element of a particular process cross section, Monte Carlo tools are combined to obtain the complete picture of how the event of interest develops, including as a last step the simulation of the particles interactions with

the detector material.

Being real data or Monte Carlo simulated samples, at the "raw" level events are simple digital outputs, coming respectively from the real or simulated response of the read-out electronics of the different ATLAS detector subsystems. How these outputs are assembled into physical objects is described in Chapter 4, where the reconstruction process is explained. The outcome is a dataset containing all the information needed about physics objects such as leptons, jets and energy imbalance of the event, ready to be processed by analyses.

Using these kind of datasets, the Exotics group of the ATLAS collaboration defined a search strategy for exotic heavy quarks different from the first three generations and called "vector-like". Even though these quarks are predicted in various contexts, like extra-dimensions or composite Higgs models, no details on their masses are given and their decay branching fractions are very model dependent. Searches aiming at inclusivity must therefore rely the minimum as possible on assumptions from the model, and this is the approach chosen for the two searches in the single lepton channel for pair-produced heavy vector-like top partners presented in this dissertation. The general quasi-model independent strategy common to the two analyses, performed analyzing $\sim$14 fb^{-1} of data from proton-proton collisions at the center of mass energy $\sqrt{s} = 8$ TeV recorded during the year 2012 at the ATLAS experiment, is presented in Chapter 5.

The search for pair-produced heavy vector-like top partners where at least one of them decays into a W boson and a bottom quark is detailed in Chapter 6. The key point in this analysis is the reconstruction of the W boson from its hadronic decay products which allows for the reconstruction of the heavy quark mass, a very good discriminating variable between signal and Standard Model background processes.

Chapter 7 presents the search for pair-produced heavy vector-like top partners where at least one of them decays into a Standard Model Higgs boson and a top quark. In this case the main decay of the Higgs boson into two bottom quarks is exploited resulting in a final state signature characterized by a high number of recontructed jets with a large fraction of them identified as originating from the hadronization of bottom quarks.

While the individual results from the searches are presented in the respective chapters, higher sensitivity is achieved combining the two analyses. This is described in Chapter 8, and the result of these searches is compared with other similar searches exploiting multi-lepton signatures.

Personal contributions and acknowledgement

The results presented in this dissertation represent a small sample of the achievements made possible by the combined effort of many, many people. The ATLAS collaboration itself consists of $\sim$3000 scientists, working in different subgroups. In particular, the author of this dissertation participated to calibration and performance studies of the hadronic calorimeter and to the improvement of data-driven estimation of multi-jet backgrounds for analyses with top quarks decaying in the single-muon channel. Regarding the two analyses object of this dissertation, the author has been amongst the main analyzers, implementing and running the signal selection and statistical analysis and performing the needed cross checks for data to background modeling. The results are documented in two preliminary notes [1, 2], the

work for the final analyses being still on-going at the time of the writing of this dissertation. The author also significantly participated to a previous, published analysis [3] performed on the data from lower center of mass energy proton-proton collisions which set the strategy for heavy vector-like quark searches in ATLAS.

A particular acknowledgement goes to the “IFAE-top” group, present and former members, for their fundamental contributions to the common analysis framework used for these searches.

Part I

Theoretical framework

Today I have done something which no theoretical physicist should ever do in his life: I have predicted something which shall never be detected experimentally!

Wolfang Pauli to his friend Walter Baade

The Standard Model (SM) of particle physics is the most successful, beautyful and precise theory describing the interactions between fundamental particles. Its validity has been tested by precision measurements at the Large Electron-Positron Collider (LEP) at CERN and confirmed by the observation of all the particles it predicts, including the Higgs-like boson discovered at the Large Hadron Collider (LHC) in July of 2012 which up to now behaves as expected from the SM.

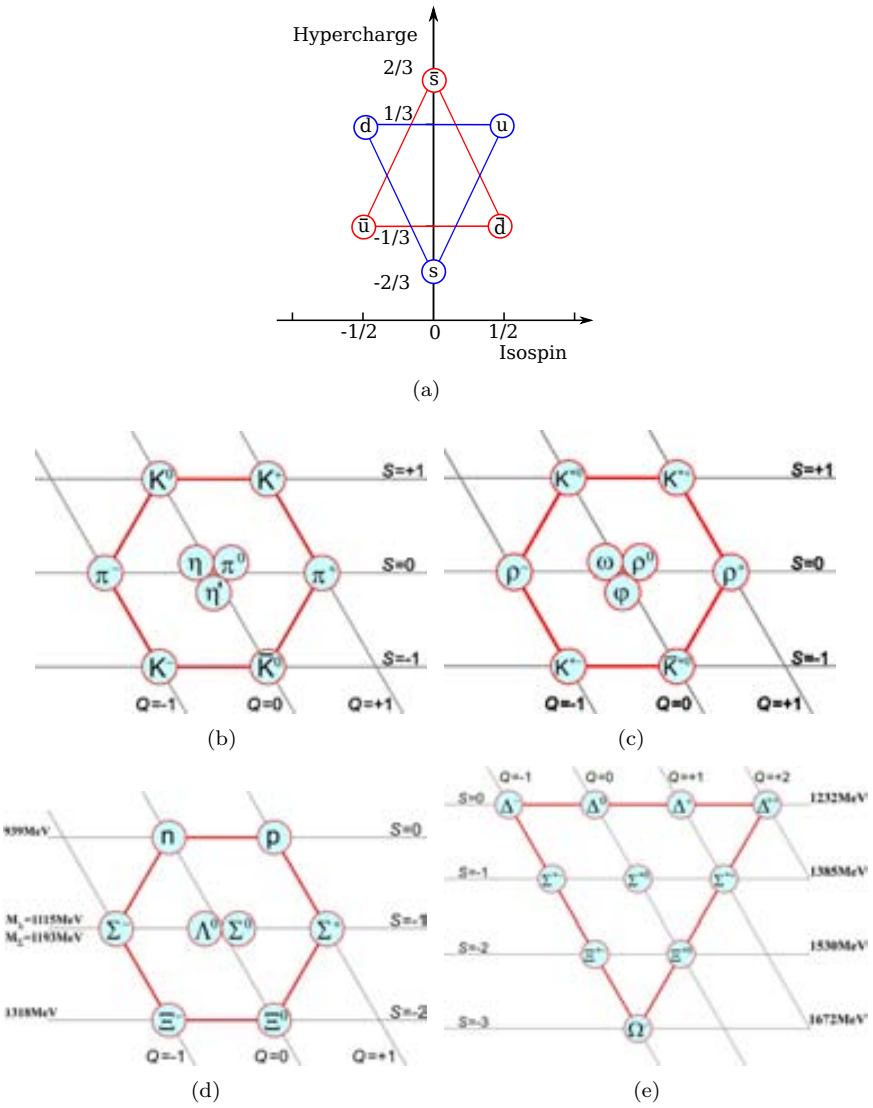
What makes the SM "only" and effective theory is the fact that unstabilities appear at high energy scales of the order of the Planck mass. In this Chapter we will show

1.1 The Standard Model from 1963 to 2012

During the first half of the 20th century, many experimental discoveries were challenging particle physicists to find a coherent model to explain the existence of new particles and forces. A "heavy electron", the muon, was observed in 1936 by C.Anderson and S.Neddermeyer in cosmic radiation, and later observed in a cloud chamber experiment by J.C. Street and E. C. Stevenson [4]. The neutrino, postulated in 1930 by W. Pauli[1] to explain the shape of the electron spectrum in beta decay, was experimentally detected in 1956 by C.L. Cowan and F. Reines [6], and few years later a second neutrino type was discovered by L.M. Lederman, M. Schwartz and J. Steinberger [7]. By 1963 a huge number of new *mesons* and *baryons* were populating what was called the "particle zoo" until Murray Gell-Mann and George Zweig independently proposed a classification for all these new particles supposing that hadrons were made by three smaller components [8, 9] dotated of a new quantum number, the hypercharge $Y = 2(Q - I_3)$, where Q is the electric charge and

[1] Pauli did not write a scientific paper of his great intuition, which is only testified by a letter he sent to the 1930 Gauverein meeting in Tübingen, famous also for his funny opening "*Dear Radioactive Ladies and Gentlemen,[...]*". A copy of the letter can be found in http://tinyurl.com/cjgoubj, and a nice history of the neutrino is in [5].

I_3 the third componend of the isospin. Figure 1.1 shows this particle classification referred to as "the eightfold way"[2].



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

Figure 1.1: Gell-Mann's *Eightfold Way* (proposed also independently by Yuval Ne'eman) identifies three fundamental components (a) and classifies through their electric charge q and strangeness s: a spin-0 meson octet (b), a spin-1 meson octet (c), a spin-1/2 baryon octet (d) and a spin-3/2 baryon decuplet (e). It is now understood that this structure is a consequence of flavour symmetry.

It was the beginning of the *quark model*, a theory that had to wait for multiple experimental evidence before being accepted, nevertheless it successfully predicted a new particle, the strangeness s=-3 particle Ω^- of the spin-3/2 baryon decuplet of Figure 1.1, discovered in 1964 at Brookhaven [10]. Even when in 1968 deep inelastic experiments at the Stanford Linear Accelerator Center (SLAC) found out evidence for a substructure in protons [11, 12],

[2]See e.g. http://lccn.loc.gov/65013009.

physicists were reluctant to accept this point-like objects to be the quarks. Richard Feynman called them *partons*, the term now used to identify quarks and antiquarks as well as gluons.

However, back then there was a tougher problem tormenting theoretical physicists. Quantum field theory was in fact apparently unsuitable for the description of the dynamics of particles interactions, since divergences appeared in the high energy domain. In 1954 Chen N. Yang and Robert Mills proposed a new gauge theory [13] based on the principle of *local gauge invariance*, i.e. the property of space-time regions of not being affected by a symmetry transformation performed locally in a different region. With the addition of a scalar field proposed by Peter Higgs, François Englert and Robert Brout [14, 15] and the implied modification of the vacuum structure, the Yang-Mills field became a very accurate description of the weak force interactions. Such unified model was consistently proposed, independently, in the 1960s by Abdus Salam, Sheldon Glashow and Steven Weinberg [16, 17], but it suffered of a problem: as it was a perturbative theory, equations had to be expanded in a power series to be calculated but only the leading order term did not show ultraviolet divergences[3].

By the first years of the 1970s Gerard't Hooft demonstrated in his PhD thesis under the supervision of Martinus Veltman the renormalization for the theory [18, 19], with the result that divergences could be cancelled and physical observables obtained with precisions higher than the leading order. The concept of *renormalization group* was introduced and Yang-Mills theories were found to have a β-function (a function typical of gauge theories) generally negative. This was the discovery of *asymptotic freedom*, a property that made Yang-Mills theory suitable also to describe strong interactions and that matched properly with the experimental effect named *Bjorken scaling*[4].

At the same time, the three-quark model by Gell-Mann and Zweig was about to be expanded. In 1963 Nicola Cabibbo proposed the mixing of up, down and strange quark [20] in order to explain the non-conservation of quark flavour in weak interactions as $\Lambda \to p^+\pi^-$ with ΔS=1 and the empirical law $\Delta S = \Delta Q$ for strangeness changing processes. In 1970 Glashow, Iliopoulos and Maiani (GIM) predicted a fourth quark [21], the charm, to account for the non-observation of Strangeness Changing Neutral Current (SCNC) processes. Thus, the quark mixing between the two families of quarks (u, d) and (c, s) could be described with a 2×2 matrix, parameterized by the Cabibbo angle θ_C and referred to as the Cabibbo-GIM matrix:

$$V_c = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \quad . \tag{1.1}$$

Furthermore, after the observation of events violating the Charge-Parity (CP) symmetry [22], Makoto Kobayashi and Toshihide Maskawa supposed the existence of two more quarks, the

[3] Divergences in computations are classified according to the energy scale at which they appear. Following the Planck relation $E = h\nu$ and the relation between the wave lenght and the frequency of radiation $\lambda\nu = c$, in natural units ($\hbar = c = 1$) energies of the order of 1 TeV or more lie in the ultraviolet (UV) range: $\lambda(\text{UV}) \sim 10^{-9}$ m. Divergencies appearing at the energy scale of 1 GeV or less are in the infrared (IR) range: $\lambda(\text{IR}) \sim 10^{-6}$ m.

[4] At SLAC it was observed during deep inelastic scattering experiments that strong interactions show a decrease of strenght at short distances (i.e. high momentum transfer) together with a scaling behaviour. A property is said to "scale" when it depends only by dimensionless kinematic quantities, such as a scattering angle or the ratio of the energy to a momentum transfer.

bottom and the top, thus increasing the number of quark flavours to six [23]. This allowed the introduction in the extended 3×3 mixing matrix (the Cabibbo Kobayashi Maskawa CKM matrix) of, besides three angles, a complex phase that is responsible of CP violation. All this conjectures found an important confirmation in November 1974, a date later known as "the November Revolution" probably because it set the beginning of a real trust in the quark theory. Almost simultaneously at SLAC and at Brookhaven the charm quark was discovered in the bound state $c\bar{c}$, called J meson by the Brookhaven team and ψ by the SLAC one, so that in the end it was named J/ψ [24, 25]. Not too much later, the bottom quark was observed in 1977 at Fermilab [26], enhancing the belief in the top quark existence and in the six flavours theory.

The discovery of the tau lepton in 1975 [27] and of the W and Z bosons in 1983 [28, 29, **?**] finally set the scene for the Standard Model (SM) of particle physics. Table 1.1 shows the fundamental particles composing the SM: three generations of fermions, each of them having a corresponding antiparticle, the three forces (three Yang-Mills fields) with their vector bosons, and the Higgs boson whose field interaction with particles results in the lagrangian mass terms. A great achievement of the SM was the unification of electromagnetic and weak theories in the Electroweak Theory by Salam, Glashow and Weinberg. In fact, since at a scale of about 100 GeV the coupling constants converge, it is possible to describe them within the same mathematical model. Quantum ChromoDynamics (QCD) instead describes the strong interaction in terms of the color threefold charge and up to now is not known if also the strong coupling constant can become equal to the others at some high energy scale. However, a unified theory is strongly desired, as will be stressed in Section 1.2.

	Leptons spin 1/2		Quarks spin 1/2	
	$q=-1$	$q=0$	$q=2/3$	$q=-1/3$
I	e^-	ν_e	u	d
II	μ^-	ν_μ	c	s
III	τ^-	ν_τ	t	b
Force	Elm	Weak		Strong
Carrier boson	γ	$W^\pm$	Z	g
spin	1	1	1	1
q	0	±1	0	0
Higgs boson H	$q=0$, spin=0			

Table 1.1: Elementary particles and forces of the SM.

Even if the SM was somehow born to be merely a stepping stone, it consolidated through years, standing all experimental tests sometimes with a precision greater than 0.1%. Experiments carried out at the Large Electron-Positron (LEP) collider at CERN, thanks to the clean signals given by e^+e^- collision events, allowed to obtain very high precision measurements of the SM parameters (see Figure 1.2 for a summary of the measurements, with their errors, of the free parameters of the SM).

In particular experiments ALEPH, DELPHI, L3 and OPAL performed the measurements of the Z boson mass with a precision of 0.0023% that made it one of the most precisely known

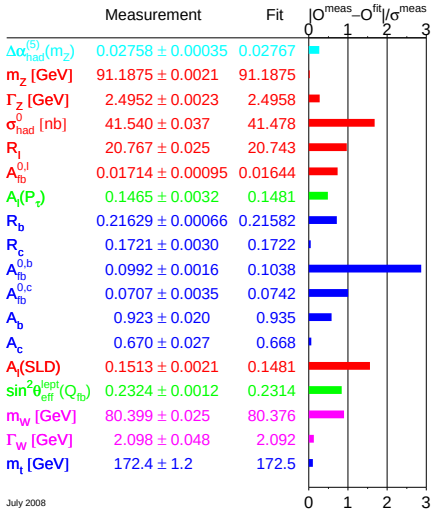


Figure 1.2: Measured free parameters of the SM with the corresponding precision and the result of the consistency fit [30].

quantities within the SM. Furthermore, measurements of its total decay width and of its partial decay widths for all processes with a visible final state (i.e. different from $\nu\bar{\nu}$) allowed to set the number of light neutrino flavours to three, confirming the three-generation SM and excluding the possibility for a fourth family of leptons with masses lower than half of m_Z.

The penultimate discovery inside the SM has been the observation of the top quark in 1995 at the Fermilab's experiments CDF and D0 [31, 32]. The mass of the top quark resulted consistent with the predicted constraints, thus confirming again the SM as an accurate framework. At that point, and for almost 20 years, the last missing piece whose absence could invalidate all the previous beautiful corroborations, was the observation of a Higgs boson with a light mass for self-consistency of the overall fit of data. When the LEP physics program was terminated, direct searches gave, at 95% CL, a lower limit of 114 GeV and an upper limit of 144 GeV on the mass of the Higgs boson [30].

Then LEP was dismantled, the LHC was built, and in 2012 the ATLAS and CMS experiments observed a new ∼125 GeV mass boson [33, 34] with the same spin-parity as the one expected from a Standard Model Higgs boson. A precise *identikit* of this new particles, necessary in order to say the final word about its nature (is it a Higgs boson? Is it the Standard Model Higgs boson? Is it a "new physics" Higgs boson?), is expected to come within the next decades of activity of the LHC and its experiments.

1.1.1 Building the Standard Model

The Standard Model (SM) is a gauge theory invariant under the symmetry transformation $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The three terms of the product of groups are the matrix representations of the fundamental symmetries acting on the forces of Nature (*quantum fields*) governing the interactions of particles: $SU(3)_C$ is the *color* unbroken symmetry act-

ing on the gluon field G_a; $SU(2)_L \otimes U(1)_Y$ is the unified *electroweak* broken symmetry, with the weak symmetry $SU(2)_L$ acting on the vector boson fields $W_{1,2,3}$ and on the scalar Higgs field ϕ, and the electromagnetic symmetry $U(1)_Y$ acting on the vector boson field B and on the scalar Higgs field ϕ.

The quantum number C, the color charge, is carried only by quarks, antiquarks and gluons in three different values labelled "red", "gree", "blue". The quantum number I, the weak isospin, differentiates between left-handed ($I = 1/2$) and right-handed ($I = 0$) fermions, with the latter not undergoing weak interactions. The quantum number Y, the hypercharge, is defined as $Y = 2(Q - I_3)$, where Q is the electric charge and I_3 the third componend of the isospin, which is $I_3 = +1/2$ for up-type quarks and negatively charged leptons, and $I_3 = -1/2$ for down-type quarks and neutrinos (and *vice-versa* for the antiparticles).

Gravity is not (yet) included in the model, and even though this is a flaw of the SM that is desirable to be fixed in a Grand Unified Theory (GUT), its action on particles is of many order of magnitudes lower than the others' and is, therefore, negligible at the fundamental components scale.

Building the electroweak lagrangian

The Lagrangian of the SM is built by *gauging* the symmetries in order to obtain invariance under those transformations. Fermions are spin-1/2 particles that can be represented as spinors. Using ψ_L and ψ_R to denote the left-handed and right-handed fermion fields respectively, the bare electroweak Lagrangian of the SM (considering for simplicity only leptons) is made of two terms:

$$\mathcal{L}_0 = \mathcal{L}_{lept} + \mathcal{L}_{gauge} \tag{1.2}$$

that are

$$\begin{cases} \mathcal{L}_{lept} &= \bar{\psi}_L i \not{D}_L \psi_L + \bar{\psi}_R i \not{D}_R \psi_R \\ D_L^\mu &= \partial^\mu + ig\dfrac{\bar{\sigma}\cdot\bar{W}^\mu}{2} + i\dfrac{g'}{2} Y_L B^\mu \\ D_R^\mu &= \partial^\mu + i\dfrac{g'}{2} Y_R B^\mu \end{cases} , \tag{1.3}$$

$$\begin{cases} \mathcal{L}_{gauge} &= -\frac{1}{4} W^l_{\mu\nu} W^{\mu\nu\, l} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ W^l_{\mu\nu} &= \partial_\mu W^l_\nu - \partial_\nu W^l_\mu - g\varepsilon^{jkl} W^j_\mu W^k_\nu \end{cases} . \tag{1.4}$$

The gauge invariance is obtained through the definition of the covariant derivatives $\not{D} = \gamma_\mu D^\mu$, which are different for the left- and right-handed components of the field. The chirality of the electroweak interactions does not find a theoretical motivation, but $SU(2)_L \otimes U(1)_Y$ transformations do make distinction between left and right helicity of the spinors. Introducing the Weyl representation of the γ matrices

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \text{and} \quad \gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} , \tag{1.5}$$

where σ_i are the Pauli matrices, the left- and right-handed spinors transform as:

$$\begin{cases} \psi_L = \frac{1}{2}(1-\gamma_5)\phi & \rightarrow \psi'_L = e^{iY\beta(x)+i\bar{\sigma}\bar{\alpha}(x)}\psi_L \\ \psi_R = \frac{1}{2}(1+\gamma_5)\phi & \rightarrow \psi'_R = e^{iY\beta(x)}\psi_R \end{cases} . \tag{1.6}$$

Equation 1.3 is the "free matter" Lagrangian describing the transformation under the symmetry $SU(2)_L$ of weak isospin with coupling constant g, three boson fields $W^l_{\mu\nu}$ and their weak generators $\bar{\sigma}$[5], and under the symmetry $U(1)_Y$ of hypercharge with coupling constant $g'/2$, the boson field $B_{\mu\nu}$ and its hypercharge generator Y. Equation 1.4 is the Lagrangian for the vector bosons dynamics with self-couplings, including trilinear and quadrilinear terms.

The Lagrangian 1.2 is invariant under group $SU(2)_L \otimes U(1)_Y$ transformation, but has the problem of leaving fermions massless. Thus a scalar Lagrangian with a quartic auto-interaction is introduced:

$$\begin{cases} \mathcal{L}_\phi &= (D^\mu_\phi \phi)^\dagger (D_{\phi\,\mu}\phi) - V(\phi^\dagger\phi) \\ V &= \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \\ D^\mu_\phi &= \partial^\mu + ig\frac{\bar{\sigma}\cdot\bar{W}^\mu}{2} + i\frac{g'}{2}Y_\phi B^\mu \end{cases} , \tag{1.7}$$

where ϕ is an Higgs doublet, accounting for four degrees of freedom needed to give mass to three vector bosons (the neutral Z boson and the two $W^\pm$ bosons) and to a scalar Higgs boson:

$$\phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} . \tag{1.8}$$

The potential $V(\phi)$ depends on two parameters, μ^2 and λ. The case $\lambda < 0$ is unphysical, leading to no stable minima. For $\lambda > 0$, two cases arise: $\mu^2 > 0$ and $\mu^2 < 0$. In the first case (see Figure 1.3(a)) there is a single solution to the minimization which corresponds to $|\phi| = 0$ and gives as vacuum expectation value $\langle 0|\,\phi\,|0\rangle = 0$. In the second case (see Figure 1.3(b)) the potential has a non-vanishing vacuum expectation value $\langle 0|\,\phi\,|0\rangle = v$ and the minimum is not unique anymore. The fundamental vacuum state is no more invariant under $SU(2)_L \otimes U(1)_Y$, meaning that these two symmetries are now broken[6]: this is the Spontaneous Symmetry Breaking (SSB) mechanism. The vacuum state is chosen as:

$$\phi_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{1.9}$$

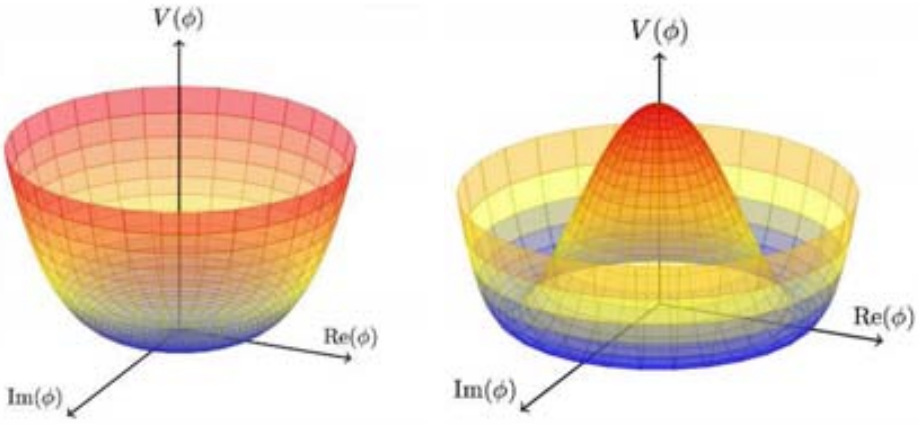
where $v = \sqrt{-\mu^2/\lambda}$ is the vacuum expectation value. Finally, since only the vector bosons got their mass up to now, the only thing left to do is to introduce the scalar-fermion interaction in order to give mass to the fermions. This is done defining:

$$\mathcal{L}_{Yukawa} = -G_{lept}\left[\bar{\psi}_R(\phi^\dagger\psi_L) + (\bar{\psi}_L\phi)\psi_R\right]. \tag{1.10}$$

[5]These are the Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

[6]Instead, the symmetry $U(1)_{elm} \subset SU(2)_L \otimes U(1)_Y$ is not broken. This means that vacuum is electrically uncharged while it has isospin and hypercharge charges.



(a) (b)

Figure 1.3: Vacuum pontential for $\lambda > 0$ and (a) $\mu^2 > 0$ or (b) $\mu^2 < 0$, with the typical shape of a mexican hat [35].

Thus the complete electroweak Lagrangian of the SM, that can be generalized also to quarks (see Table 1.2), is

$$\mathcal{L} = \mathcal{L}_{lept} + \mathcal{L}_{gauge} + \mathcal{L}_{\phi} + \mathcal{L}_{Yukawa}. \tag{1.11}$$

Weak isospin left-handed doublets ψ_L			
$\begin{pmatrix}\nu_{e_L}\\ e_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\mu_L}\\ \mu_L\end{pmatrix}$	$\begin{pmatrix}\nu_{\tau_L}\\ \tau_L\end{pmatrix}$	**Leptons**
$\begin{pmatrix}u_L\\ d'_L\end{pmatrix}$	$\begin{pmatrix}c_L\\ s'_L\end{pmatrix}$	$\begin{pmatrix}t_L\\ b'_L\end{pmatrix}$	**Quarks**
Weak isospin right-handed singlets ψ_R			
e_R	μ_R	τ_R	**Leptons**
u_R , d'_R	c_R , s'_R	t_R , b'_R	**Quarks**

Table 1.2: Weak isospin multiplets. q' refers to the flavour eigenstate that correspond to the mass eigenstate transformed with the CKM matrix.

An important consequence of the introduction of the scalar Higgs field and the consequent spontaneous symmetry breaking is the mixing of the vector bosons B_μ and $W_\mu^{1,2,3}$ to give

the photon A_μ, the two $W^\pm_\mu$ and the Z_μ bosons:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right), \tag{1.12}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}, \tag{1.13}$$

where the Weinberg angle θ_W is defined through

$$\begin{cases} \dfrac{g}{\sqrt{(g')^2+g^2}} = \cos\theta_W \\ \dfrac{g'}{\sqrt{(g')^2+g^2}} = \sin\theta_W \end{cases}. \tag{1.14}$$

Masses acquired through the Higgs mechanism are listed in Table 1.3.

Force Carriers			
γ	$W^\pm$	Z	g
0	$\sqrt{\dfrac{g^2v^2}{4}}$	$\sqrt{\dfrac{(g^2+g'^2)v^2}{4}}$	0
0	80.4	91.2	0

Leptons and Quarks									
e	μ	τ	ν	u	d	c	s	t	b
				$\dfrac{v}{\sqrt{2}}g_f$					
$0.511\cdot10^{-3}$	0.105	1.777	~ 0	0.0024	0.0048	1.27	0.104	171.2	4.2

Table 1.3: Particle Masses in GeV.

Adding the strong interaction

Adding the color transformations $SU(3)_C$, the Standard Model is finally described by the symmetry group $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$. In the same way electromagnetic interactions are described by Quantum ElectroDynamics (QED), strong interactions are described by Quantum ChromoDynamics (QCD [36]). However, while the photon does not carry electric charge, the gluon does carry color charge (in particular it exists in eight states, being a color octet) and can, therefore, self-interact.

The QCD lagrangian only involves quarks and gluons, and reads:

$$\begin{cases} \mathcal{L}_{QCD} &= \sum_{f=1}^{6} \bar{q}(i\gamma^\mu D_\mu - m_q)q - \frac{1}{4}F^{\mu\nu}_a F^a_{\mu\nu} \\ D_\mu &= \partial_\mu + ig_s G^a_\mu T^a \\ F^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f^a{}_{bc} G^b_\mu G^c_\nu \end{cases} \tag{1.15}$$

where q is the quark field and m_q is the quark mass, summed over the six types of quarks, and a runs over the eight degrees of freedom of the gluon field, T^a being the generators of the $SU(3)_C$ group. The field tensor $F^a_{\mu\nu}$ is derived from the gluon field G^a_μ and its third

term describes the gluon self-interaction, with f^a_{bc} being the structure constants. The strong coupling constant α_s is defined as $\alpha_s = g/4\pi$ and its values, large at low energies (i.e. large distances), and small at high energies (i.e. short distances), determine the peculiar property called *asymptotic freedom* which explains the confinement of the quarks inside the hadrons (see Figure 1.4). Indeed, at small distances α_s is so low that the strong interaction can be treated perturbatively and quarks and gluons described as free particles. On the other hand, a quark cannot be found isolated since at large distances the field strength will increase enough to create new quarks from the vacuum and colorless hadrons will be formed.

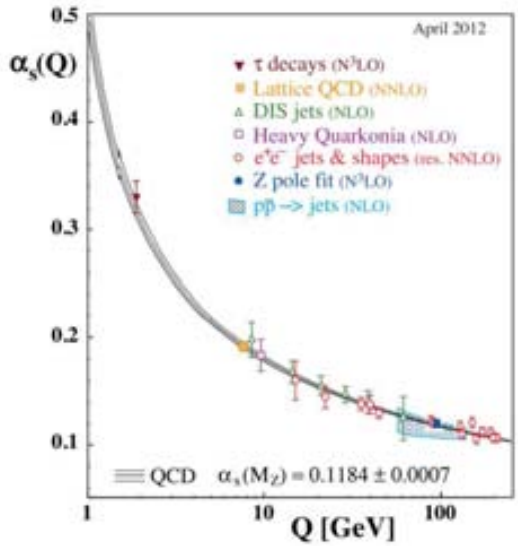


Figure 1.4: Running of the strong coupling α_s with the energy scale Q, proven from different measurements [37].

As an example of how quarks behave in hadrons, protons are composed of two *valence* up quarks and one *valence* down quark. The quarks continuosly exchange gluons, which create in turn quark-antiquark pairs, resulting in what is referred to as a "sea" of quarks and gluons. As the sum of the rest masses of the valence quarks contributes only to about 1% of the total nucleon mass, what accounts for the missing 99% is the binding energy of gluons and sea quarks. Further details on QCD will be given in Section 3.1, where the phenomenology of proton-proton collision is discussed.

1.2 Unanswered questions and new physics quests

In the case the new boson discovered in July 2012 really is a Standard Model Higgs boson, there would be no more arguments to contradict the Standard Model as an *effective theory*. Indeed many facts, both in theory and experiments, hints that the Standard Model, despite is great success in describing the interaction of fundamental particles, might be just an approximation at low energy regimes of a more complete theory.

One of the principal objections to the Standard Model as "the final theory" is the high number of arbitrary parameters of the theory. In fact, 19 parameters are needed to fit data

from experimental observations. Three of them are the couplings of the gauge groups g_3, g, g' for the strong, electromagnetica and weak interactions respectively, also written as:

$$\alpha_s = \frac{g_3^2}{4\pi}, \quad \alpha_{elm} = \frac{e^2}{4\pi} = \frac{g^2 \sin^2\theta_W}{4\pi}, \quad \sin^2\theta_W = \frac{(g')^2}{g^2 + (g')^2}. \tag{1.16}$$

Then, 13 parameters are associated with the nine charged fermion masses and the four parameters of the CKM matrix (three quark-mixing angles and one phase), two are needed to describe the Spontaneous Symmetry Breaking mechanism, i.e. the Higgs vacuum expectation value v and the quartic coupling constant λ, and the last one is the QCD θ parameter. Additionally, if neutrinos are massive (as it is almost certain from neutrino oscillation observations, see e.g. [38]) there will be even more arbitrary parameters describing their masses and their mixing. Furthermore, massive neutrinos cannot exist in the Standard Model, where only left-handed neutrinos are predicted and thus no Dirac mass term can appear[7].

The arbitrariety of parameters, and in particular of the fermion masses, introduces what goes under the name of *naturalness problem*. A "natural" theory is characterized by free parameters with values at, more or less, the same order of magnitude. This does not happen in the Standard Model, where the top quark, as an example, has a mass $\sim 10^5$ larger than the up quark. This issue further develops as follows. If the Standard Model is valid only up to an energy scale Λ (which, if it's the Planck scale, differs from the electroweak scale by $\sim 10^{17}$!), then the scalar Higgs boson mass should encounter radiative corrections from vacuum polarization diagrams (like the one in Figure 1.5) of the order of Λ giving to the mass the value:

$$M_H^2 \sim M_{H_0}^2 + \frac{\lambda}{4\pi^2}\Lambda^2 + \delta M_H^2. \tag{1.17}$$

If the mass counterterm δM_H^2 does not cancel the quadratically divergent contribution and if the cutoff scale is chosen as the Planck scale, then

$$M_H^2 \sim 10^{32}, \tag{1.18}$$

i.e. many orders of magnitude bigger than the experimentally measured value coherent with the Standard Model and with the unitarity constraint. This is the *hierarchy problem*, and could be fixed within the Standard Model by choosing a fine-tuned mass counterterm, a solution considered not really elegant also because fine tuning will be required for every order in the perturbative expansion[8].

Another disturbing feature of the Standard Model as it is is the lack of theoretical explanation for the generations of quarks and leptons to be exactly three, as suggested (under certain assumptions) by precision measurements performed at LEP at the Z-pole ($\sqrt{s} \sim 91$ GeV). From QCD comes the only constraint for quark generation to be less that nine.

Cosmology and cosmological observations also challenge the Standard Model. The reason for baryon-antibaryon asymmetry is still not understood although we know that it is

[7] A way out of this problem postulates a new type of neutrinos, namely *Majorana neutrinos*, in contrast to *Dirac neutrinos*.

[8] This problem does not arise with loop corrections to fermion masses, which are protected by chiral symmetry, nor with boson masses, which are protected by gauge invariance. It is, actually, an issue of scalar particles like the Higgs boson.

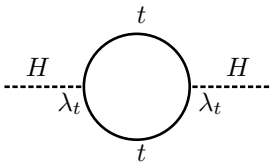


Figure 1.5: The typical vacuum polarization diagram for the Higgs is a top quark loop.

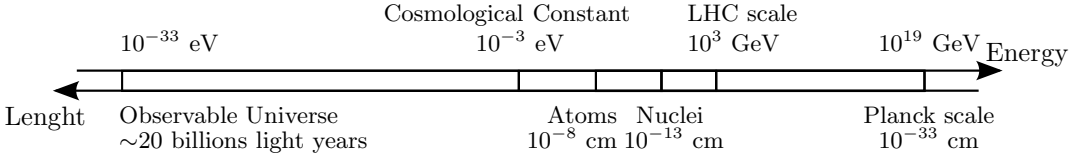


Figure 1.6: Typical length and energy scales of some of the fundamental parameters of the Universe.

connected to CP violation[9]. Besides, astronomical observations [39] tell us that the energy density of the Universe is made only for a 4-5% of ordinary baryonic matter, the other components being dark matter (20-25%) and dark energy (70-76%). Dark matter is non-baryonic matter that interacts only weakly and gravitationally and, therefore, cannot be observed with telescopes but it is revealed by its gravitational interaction with ordinary matter in space. It is now believed that Dark Matter is composed of Weakly Interacting Massive Particles (WIMPs) whose masses range from a few GeV to a few TeV and are not predicted within Standard Model. Dark energy instead is still more mysterious and maybe new physics will give some hints for its interpretation.

Another topic making the Standard Model likely to need improvements is the desire to go further in the unification of theories. Gravity is not implemented in the Standard Model, nor is available a widely accepted quantum theory of gravity. This is acceptable at the electroweak scale of few hundreds of GeV where the strength of gravity is negligible, but its effect should become relevant going up to the Planck scale $\Lambda \sim 10^{19}$ GeV. Also, electroweak and strong forces forming the Standard Model gauge group $SU(2)_L \otimes U(1)_Y \otimes SU(3)_C$ are expected to unify at high energy since their coupling constants are running constants dependent on the energy scale (Figure 1.7, $\alpha_i^{-1} = g_i^2/(4\pi)$).

In the following some "beyond-Standard Model" (BSM) theories proposed to solve most of the issues illustrated before are briefly illustrated.

1.2.1 Supersymmetry

Since Supersymmetry [40] is one of the most popular BSM scenarios, having many trustful supporters even now that after the first LHC run no trace of it has been found, a brief explanation of its concepts is included in this dissertation.

The basic idea is to postulate the invariance of the theory under a symmetry operation which trasforms fermionic fields into bosonic fields (and viceversa), called *supersymmetry*

[9]The CP violating phase introduced in the CKM mechanism cannot, however, account for the total baryon-antibaryon asymmetry measured.

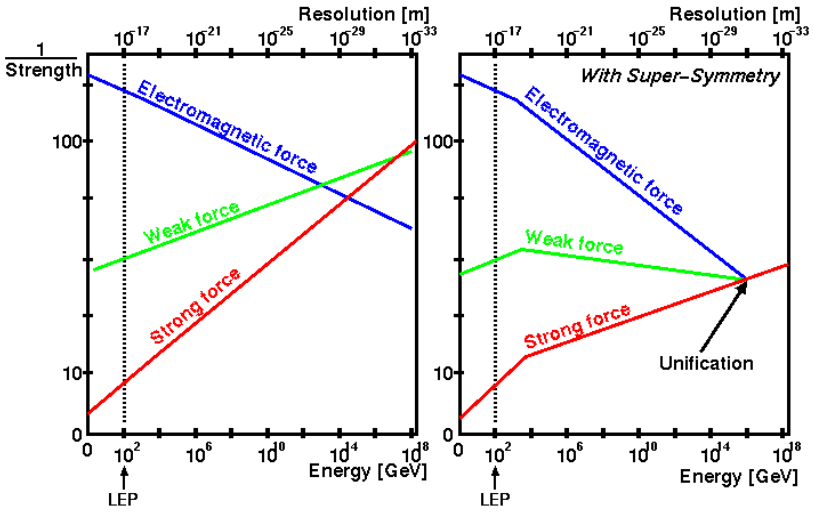


Figure 1.7: Running coupling constants in the Standard Model (left) and in a hypothetical Supersymmetric Model (right, see Section 1.2.1) as functions of the renormalization scale (picture from http://scienceblogs.com, original credits unknown). The energy scale explored at LEP is marked on the two figures and corresponds to 10^2 GeV. The LHC is able to go just one order of magnitude further.

(SUSY). In this theory to each fermionic or bosonic degree of freedom of the Standard Model is associated a "superpartner". These superpartners have all of the quantum numbers identical to the corresponding Standard Model particles, except for the spin quantum number transformed as $s' = |s - 1/2|$. Thanks to the presence of superpartners, the Fermi statistics allows to write the new version of Equation 1.17 as:

$$M_H^2 \sim M_{H_0}^2 + \frac{g_F^2}{4\pi^2}(\Lambda^2 + m_F^2) - \frac{g_S^2}{4\pi^2}(\Lambda^2 + m_S^2), \tag{1.19}$$

where the subscripts F and S indicate respectively fermionic and scalar degrees of freedom. If $g_F = g_S$ and masses are equal as in an unbroken supersymmetry scenario, the Λ^2 terms cancel and the hierarchy problem is solved. However, since no superpartners of known particles have been observed, supersymmetry must be a broken symmetry and sparticles masses have to lie in an energy range not yet accessed by experiments.

Supersymmetry also provides a natural candidate for Dark Matter, the Lightest SUSY Particle (LSP), which, in a *R-parity*[10] conserving scenario [41], would be stable, weakly interacting and neutral.

1.2.2 Fourth generation SM4

One of the simplest modifications to the Standard Model is the addition of a fourth generation of fermions, a scenario referred to as SM4 [42]. This does not contraddict the precision measurements of the Z boson decay width performed at LEP since the fourth

[10]The quantum number *R-Parity* is defined as $R \equiv (-1)^{3(B-L)+2s}$ to prevent lepton and baryon numbers from being violated. It has value $+1$ for all Standard Model particles and -1 for their superpartners.

neutrino might have a mass $m_{\nu_4} > m_Z/2$.

The new fermions are sometimes called f_4 ($f = u, d, l^-, \nu$) or, more frequently, t', b', τ', ν'_τ. In the Standard Model lagrangian additional Yukawa couplings appear and the CKM matrix is extended to a 4×4 matrix that transforms the weak eigenstates into the mass eigenstates. This introduces three complex phases allowing for more sources of CP violation[11]. Combining experimental measurements and unitarity constraints, the elements of the SM4 CKM matrix can be evaluated [43]. It can be intuitively expected that the fourth generation mixes primarily with the third generation, but this is not always the case. If it is further assumed that the mass splitting between the two new quarks t' and b' is lower than the value of the W boson mass, the decay $t' \to Wb'$ is not allowed and the fourth generation quarks will decay into a W boson and a standard model quark, according to the matrix elements $|V_{qt'}|$ ($q = u, c, t$) and $|V_{qb'}|$ ($q = d, s, b$).

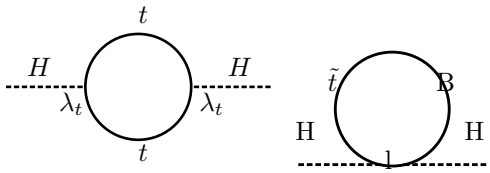


Figure 1.8: Additional contribution to the Higgs mass corrections canceling the divergent term from the top quark loop.

After the discovery of a Standard Model-like Higgs boson, the measurements performed of its cross section and decay branching ratio apparently rule out the possibility of extra quarks that receive mass through the Yukawa couplings with the Higgs doublet [44]. Indeed the Higgs production through gluon-gluon fusion in the Standard Model would be 9 times larger than the measured value due to the additional contributions, but there are still some models that allow for a chiral fourth generation [45].

1.2.3 Composite and Little Higgs

The SSB mechanism, described previously in Section 1.1.1, does not apply only to the Standard Model. It is, indeed, a feature of many models. In general, when a symmetry is spontaneously broken, *Goldstone bosons* (scalar, massless particles) arise [46], like excitations of the field. The Goldstone boson can aquire mass if the symmetry is not exact and is both spontaneously and explicitly broken. In this case the boson is called *pseudo-Goldstone bosons* (PGB). To each generator of the broken symmetry corresponds a (pseudo-)Goldstone boson. In the Higgs mechanism, where the $SU(2)_L \otimes U(1)_Y$ symmetry is gauged, three out of four Goldstone bosons are "eaten" by the Z and $W^\pm$ bosons, while the fourth is what then becomes the Higgs boson.

A nice example of a pseudo-Goldstone boson is the pion, which was considered the responsible for quark masses. In QCD the flavor chiral symmetry of the Lagrangian is broken spontaneously, generating three massless scalar bosons. The further explicit symme-

[11] The 3×3 CKM matrix has only one complex phase.

try breaking operated by the quark masses gives mass also to the pseudo-Goldstone bosons which is, however, much smaller than the other mesons' masses. The three pseudo-Goldstone bosons are the $\pi^{\pm}$ and π^{0} particles. However, the pion is not a fundamental particle as it was believed before the idea of quarks was first proposed but is, indeed, composed by quarks.

More or less in the same way one can construct a class of BSM theories that go under the name of *strong EWBS* (from ElectroWeak Symmetry Breaking). These scenarios predict either no Higgs boson at all (as in *technicolor*) or a light pseudo-Goldstone Higgs bosons. The latter is the case for *composite Higgs* models [47, 48], where the mass of the Higgs boson is now naturally low and hence there is no hierarchy problem to worry about. What happens here is that up to a certain energy scale (imposed to be "far" from the electroweak scale), the Higgs does not show its composite nature (these new strong constituents being, for now, undefined). There are then in the theory two sectors: one well described by the Standard Model gauge and particle fields (the *elementary* sector), and another "strongly" coupled containing the Higgs field and new heavy resonances (*composite* sector). In the composite sector a global symmetry is spontaneously broken and then, thanks to small mixing with the elementary sector, it is also explicitly broken, giving a pseudo-Goldstone bosons.

Many models developed from the composite Higgs scenario, such as Little Higgs models. predict exotic heavy quarks [49, 50].

1.2.4 Extra-dimensions

All the physics developed up to now assumes a four-dimensional spacetime reality described by four-vectors that specify the spatial (three-dimensional) position and the temporal (one-dimensional) position of an event. From cosmological observations it looks like we are living in an expanding universe and the spacetime, therefore, expands with it. If a particular moment in time is chosen, the universe can be described as a three-dimensional flat surface with cartesian topology.

Extra Dimensions (ED) theories propose that our universe is a four-dimensional "wall" or "three-brane" embedded in a bigger multi-dimensional space ("bulk") where the only communication comes through the *graviton*, the hypothetical vector boson of gravity. And, actually, these extra dimensions is where the gravity "dilutes" itself, explaining why in our brane it is 10^{32} times weaker than the electroweak force. The first model was proposed in the 1920's by Kaluza [51] and Klein [52] in the "grand unification" attempt to unify electromagnetism with gravity. This model generated little interest in the scientific community until 1998, when Arkani-Hamed, Dimopoulos, and Dvali [53] proposed the "ADD" model, where extra-dimensions are compactified around the three-dimensional brane (Figure 1.9). The idea is that extra dimensions are hidden to us because they are compactified on very small scales up to now impossible to probe.

Many ED models developed in the following years because of the interesting features of the basic idea that the Standard Model particles are confined to our three-dimensional brane. In general, in most of ED theories Kaluza-Klein (KK) states appear. These are infinite modes, also called "towers", existing for every particle that can propagate to the extra dimensions as resonances. In other terms, the same particle exists with larger and larger mass the further goes towards extra dimensions, and these states can be observed on

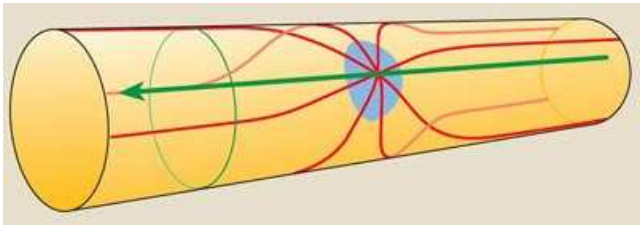
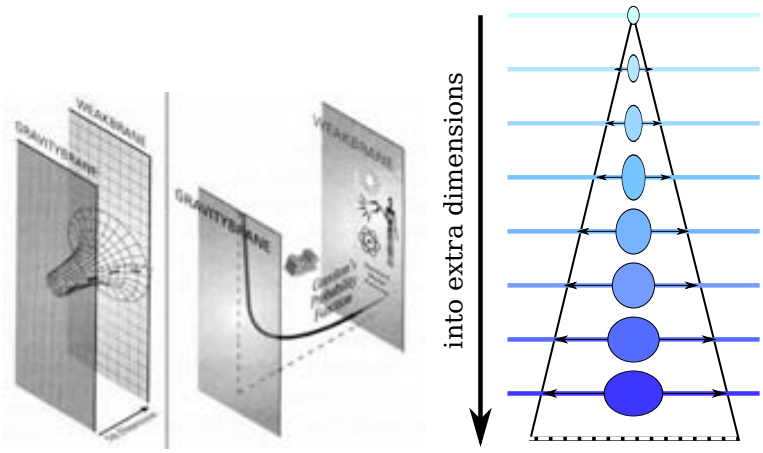


Figure 1.9: Drawing of compactified extra dimensions. The three-dimensional brane lies along the green line, while infinite extra dimensions (green circles) make up the bulk (picture from the web).

the three-dimensional brane (Figure 1.10(b)).



(a) (b)

Figure 1.10: (a): Pictorial representation of the 5th dimension (from http://universe-review.ca). (b): Pictorial representation of a KK tower, with particles aquiring more mass oscillating towards the 5th dimension.

1.3 Going beyond the SM with vector-like quarks

New physics models proposed to go BSM have been briefly presented in Section 1.2. Some of these theories predict new heavy quarks [54] with spin 1/2, transforming as triplets under $SU(3)_C$ and with the same electroweak quantum numbers for their left- and right-handed components. They will therefore transform the same under $SU(2)_L \otimes U(1)_Y$. This also means that those new quarks will not contribute to the Higgs production and decay processes, avoiding then the limitations encountered by a fourth chiral generation hypothesis.

In this section is described the general phenomenology of new heavy top and bottom partners, T and B with charges $2/3$ and $-1/3$ respectively, as well as new heavy exotic quarks, X and Y with charges $5/3$ and $-4/3$ respectively. They can appear, depending on the model considered, as $\mathrm{SU}(2)_L$ isosinglets, isodoublets or isotriplets, but here only the first two cases will be considered. A new $T_{L,R}$ isosinglet is present, e.g., in Little Higgs models [55, 56], where it aids solving the hierarchy problem, as well as (in form of a tower of isosinglets $T^{(n)}_{L,R}$) in extra-dimensional models with t_R in the bulk [57].

These vector-like quarks can couple to Standard Model quarks and in general their mixings are of order m/M [58], with m and M being the masses of Standard Model and new quarks respectively. Hence, since $m_t \gg m_{u,d}$, $m_b \gg m_{d,s}$, the new heavy quarks will mainly couple to the third generation of the Standard Model, unless some model-specific symmetries are imposed. This is encouraging also from an experimental point of view since the measured constraints on couplings involving the top quark are weaker than the ones on the lighter quarks [59, 60]. It is worth noticing that however some models allow for large mixing of the new heavy quarks with the Standard Model first and second generations [61].

1.3.1 Production

The new heavy vector-like quarks can be produced in proton-proton collisions in pairs via QCD interactions or in single mode associated with Standard Model quarks via electroweak interactions. Figure 1.11 shows the cross section for pair production ($Q\bar{Q}$) and single production in v

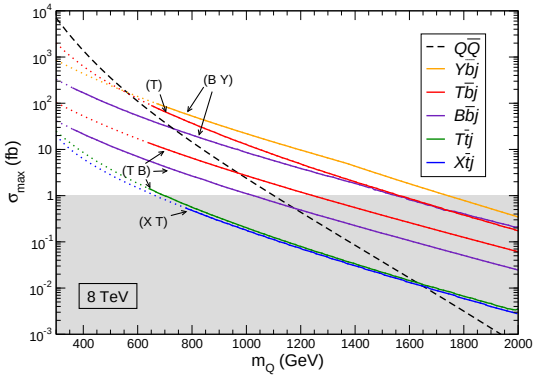
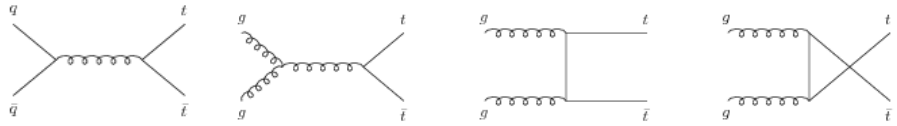


Figure 1.11: Production cross section for heavy quarks as a function of their mass, for pair production and for single production in different channels [62]

Pair production via strong interaction is analogous to pair production of Standard Model top quarks (see Figure 1.12):

$$gg, q\bar{q} \to Q\bar{Q} \qquad (Q = T, B, X, Y)\,, \tag{1.20}$$

and is dominant with respect to electroweak single production at lower m_Q masses.



(a) (b)

Figure 1.12: Pair production of top quarks from $q\bar{q}$ (a) and gg QCD interactions [63]. Pair production of heavy quarks is analogous.

1.3.2 Decay

The electroweak interactions of the new heavy quarks determine the possible decay channels of the isosinglets and isodoublets. In general, the allowed decay modes are:

$$T \to W^+ b\,, \quad T \to Zt\,, \quad T \to Ht; \tag{1.21}$$

$$B \to W^- t\,, \quad B \to Zb\,, \quad B \to Hb; \tag{1.22}$$

$$X \to W^+ t; \tag{1.23}$$

$$Y \to W^- b. \tag{1.24}$$

The flavor changing neutral currents entering in the interactions of the T and B quarks come from the breaxing of the GIM mechanism in the modified lagrangian [54].

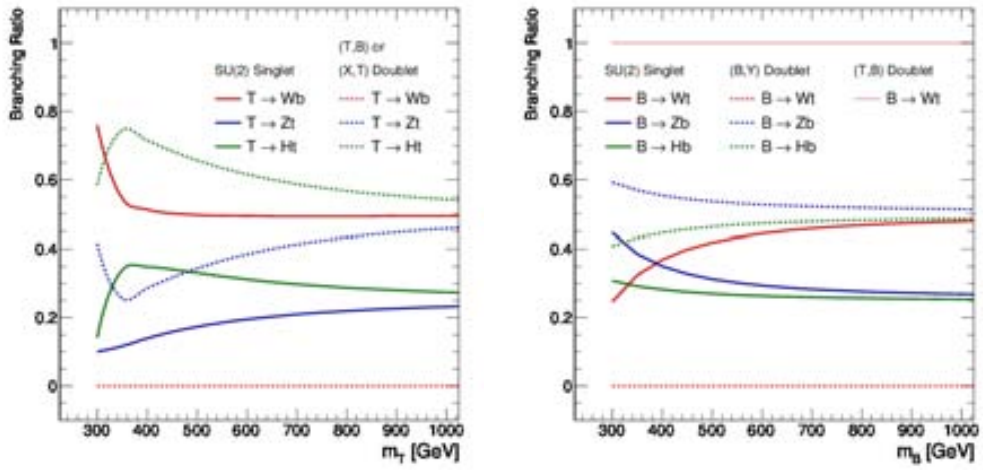
For the isospin singlets $T_{L,R}$ and $B_{L,R}$ all the three decay modes of Equations 1.21 and 1.22 respectively are possible, while for isospin doublets different options are present. Table 1.4 summarizes the allowed decay modes for vector-like isosinglets and isodoublets.

In the case of the $(T\,B)_{L,R}$ doublet the two quarks are almost degenerate in mass and the decays strongly depend on the mixing factors of the extended CKM matrix V_{Tb} and V_{tB}. If $V_{Tb} \sim V_{tB}$ then the T and B quarks have the same decays as the corresponding isosinglets but different angular distributions since only the right-handed component of $(T\,B)_{L,R}$ couples to the Standard Model quarks. In the most natural case where $V_{Tb} \ll V_{tB}$ with the Standard Model top quark, much heavier than the bottom quark, mixes more strongly with the heavy quark, the $T \to W^+ b$ decay is suppressed as well as the $B \to Hb$ and $B \to Zb$, and this is the scenario considered in Table 1.4.

The X and T quarks of the $(X\,T)_{L,R}$ doublet are also almost degenerate in mass and also in this case they only couple to Standard Model quarks with the right-handed component. Here, however, charged currents are not present in the first order lagrangian and hence the decay $T \to W^+ b$ is not present, as reported in Table 1.4. For the $(B\,Y)_{L,R}$ doublet exactly the same arguments apply.

VLQ Singlets	Decay modes	VLQ Doublets	hyper-charge	Decay modes
$T(+2/3)$	W^+b, Ht, Zt	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\frac{1}{6}$	Ht, Zt W^-t
$B(-1/3)$	W^-t, Hb, Zb			
$X(+5/3)$	W^+t	$\begin{pmatrix} T \\ X \end{pmatrix}$	$\frac{7}{6}$	Ht, Zt W^+t
$Y(-4/3)$	W^-b	$\begin{pmatrix} B \\ Y \end{pmatrix}$	$-\frac{5}{6}$	Hb, Zb W^-b

Table 1.4: Allowed decay modes for vector-like isosinglets and isodoublets (the L, R subscript is omitted).



(a) (b)

Figure 1.13: Branching ratio of vector-like top (a) and bottom (b) partners as a function of the heavy quark mass m_T and m_B respectively [64] for isosinglets and isodoublets.

It is interesting to notice that for "simple" searches, i.e. searches only aimed at revealing the presence of a new heavy quark and not sensitive, e.g., to the particle charge or angular distribution, a $Y_{L,R}$ will be indistinguishable from a $(B\,Y)_{L,R}$ or a chiral fourth generation t'. However it would be possible to understand better the vector-like scenario by performing searches in the different decay channels in a model-independent way.

Decay ratios do not only depend on the model, but also vary as a function of the heavy quark mass: Figure 1.13 shows the decay branching ratios of the vector-like top and bottom partners for isosinglets and isodoublets as a function of the heavy quark mass, for a Higgs boson with mass $m_H = 125$ GeV.

The ATLAS experiment at the Large Hadron Collider

The time when particle physics experiments could fit in one's loft is well passed, if it ever existed. The reason is simple, as the deeper we want to investigate matter, the highest the energy we need. According to the Standard Model, we now know all the particles composing ordinary matter present in nature, so if we want to see something new, we need to produce it. The way to do it is suggested by one of the fundamental principles of relativity, $E = mc^2$, according to which we can smash massive particles and observe what other kind of matter comes out of the available energy. Soon enough after the discovery of the muon, after observing all the observable from cosmic rays, physicists started to do that using particle accelerators, the last of them in history being the Large Hadron Collider.

The Large Hadron Collider, built to collide protons at a center of mass energy of 14 TeV, is the world's highest energy particle accelerator, overcoming the Tevatron proton-antiproton collider where the top quark was discovered in 1995 [32, 31]. The ATLAS experiment is one of the collaborations that take advantage of the collisions provided by the Large Hadron Collider, and has been conceived to pursuit a challenging physics program with, at the head of the list, the discovery of the Higgs boson, achieved in 2012 [33].

In the following Chapter we will briefly describe the main features of the accelerator and with some more details the ATLAS detector, both located at the CERN laboratories in Geneva, Switzerland.

2.1 The Large Hadron Collider at CERN

The LHC program was approved by CERN Council in 1994, followed by the approval of the four main experiments: ATLAS [65] and CMS [66] in 1996; ALICE [67] in 1997; LHCb [68] in 1998. Works towards the installation of the most powerful particle accelerator of the world started when the Large Electron Positron Collider (LEP) was dismantled in 2000 to give up its place in the tunnel to the LHC, which was then fully operational by 2008.

The ATLAS experiment [65] is situated at Point 1 along the Large Hadron Collider

(a) (b)

Figure 2.1: Left: View of Point 1, just above the ATLAS cavern, with a mural painting of the detector, reproduced at a scale of about 1:3 by artist Josef Kristofoletti[1]. Right: A drawing of the LHC complex.

(LHC) [69] 27 km long ring (Figure 2.1). The accelerator tunnel can reach an underground depth of 175 meters and is spread between Swiss and French territory, while the cave where ATLAS is allocated is about 100 meters underground in the CERN Swiss site of Meyrin.

The collider accelerates protons up to 4 TeV, but is designed to reach 7 TeV per beam when it will be operated at his full potential. This energy is achieved through various steps, shown in Figure 2.2. To start with, protons are extracted from Hydrogen gas and injected in the first machine, the linear accelerator LINAC2 that starts the acceleration chain. When protons reach an energy of 50 MeV they are injected into the Proton Synchrotron Booster (PSB) and accelerated up to the energy of 1.4 GeV. The second circular accelerator, the Proton Synchrotron (PS) brings the energy of the protons to 25 GeV previous to injecting them into the last machine before the LHC, the Super Proton Synchrotron (SPS). Protons of 450 GeV finally enter the LHC where they are boosted to energies of up to 4 TeV. The four main LHC experiments are shown on the collider ring.

The LHC is composed of eight arcs 2.7 km long, each of which contains 154 dipole magnets, whose function is to bend the beams along the circular trajectory, and 49 quadrupole magnets, that focus the beam. These superconducting magnets operate at a temperature of 1.9 K, maintained by means of liquid Helium vessels. Eight insertions are placed inbetween the arches. Each insertion has a specific role that characterizes its design and can be injection, beam dumping, beam cleaning, or "physics", i.e. make the beams collide within an experiment.

First proton beams were circulated on 10th September 2008 and right on the verge of getting the first collisions at a center of mass energy $\sqrt{s} = 900$ GeV nine days later, an electrical connection joining superconducting wires of a dipole and a quadrupole failed. This caused the release of liquid Helium in the insulating vacuum, resulting in an explosion that severely damaged the machine. After more than one year devoted to repair the damage and consolidate the security, on 30th November 2009 the LHC became the world's highest energy particle accelerator[2]:

[1] More info at: http://www.atlas.ch/mural/.
[2] http://press.web.cern.ch/press/PressReleases/Releases2009/PR18.09E.html

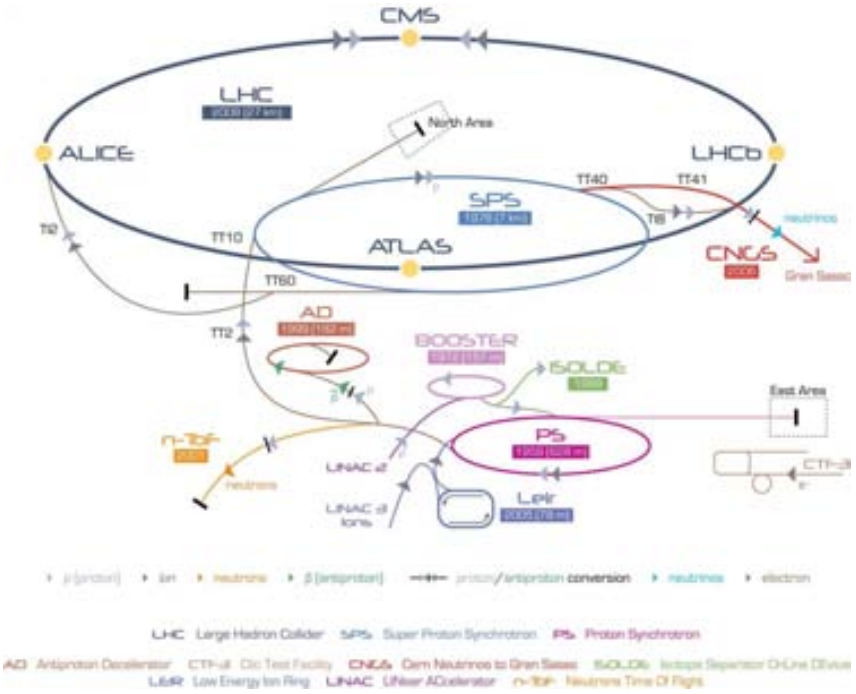


Figure 2.2: A schematic showing the accelerator complex at CERN.

> Geneva, 30 November 2009. CERN's Large Hadron Collider has today become the worlds highest energy particle accelerator, having accelerated its twin beams of protons to an energy of 1.18 TeV in the early hours of the morning. This exceeds the previous world record of 0.98 TeV, which had been held by the US Fermi National Accelerator Laboratorys Tevatron collider since 2001. It marks another important milestone on the road to first physics at the LHC in 2010.

One of the main characteristics for an accelerator is the luminosity, the instantaneous luminosity $\mathcal{L}$ being defined as

$$\mathcal{L} \times \sigma = \frac{dN}{dt} = f \times n \frac{N_1 \times N_2}{A} \times \sigma. \tag{2.1}$$

Here dN/dt is the event rate of a certain process and σ is its cross section. This rate is directly proportional to the the frequency f, the number of bunches n and the number of particles in the two bunches N_1, N_2, and inversely proportional to the beam cross-section A. The instantaneous luminosity is measured by dedicated subdetectors that are described in Section 2.3.

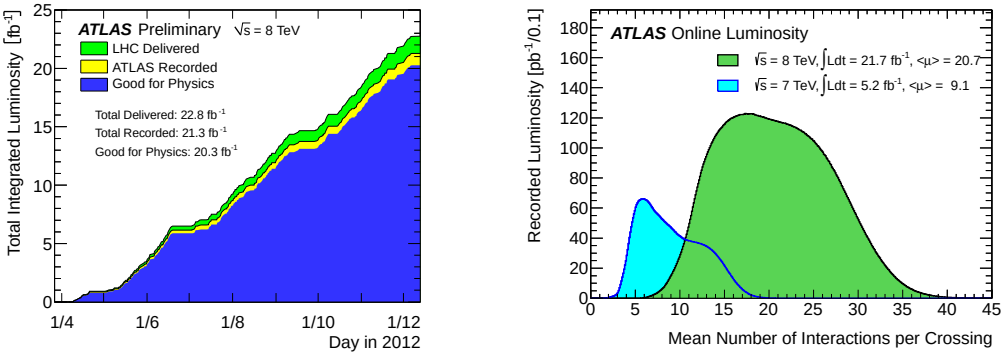
Integrating over the accelerator active time (a "fill", when stable beams are kept colliding) gives the *integrated luminosity*, relating the total number of produced events N_{tot} to the cross-section:

$$\int \mathcal{L} dt = \frac{N_{tot}}{\sigma}. \tag{2.2}$$

In 2010 ATLAS collected about 45 pb^{-1}of pp collision data at $\sqrt{s}$=7 TeV, and in 2011 reached about 5 fb^{-1}of the same data. In 2012, with $\sqrt{s}$=8 TeV collisions, LHC reached

Parameter	designed	2010	2011	2012
Beam energy (TeV/c)	7	3.5	3.5	4
Beta function $\beta*$ (m)	0.55	2.0/3.5	1.5/1.0	0.6
Max. No. bunches/beam	2808	368	1380	1380
Max. No. protons/bunch	1.15×10^{11}	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}
Bunch spacing (ns)	25	150	75/50	50
Peak luminosity ($cm^{-2}s^{-1}$)	1×10^{34}	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}
Emittance ε_n (μrad)	3.75	2.0	2.4	2.5
Max. $< \mu >$	19	4	17	37

Table 2.1: Overview of some parameters for the LHC performance comparing the design values with their time evolution during the first long run operation in 2010-2013 [70].



(a) (b)

Figure 2.3: (a) Total integrated luminosity versus time delivered by the LHC to ATLAS (in green), recorded by the experiment (in yellow) and selected as "good data" for analysis (in blue) for pp collisions at $\sqrt{s}$=8 TeV. (b) Mean number of interactions per beam crossing during 2011 and 2012 LHC runs, where $\mu = \mathcal{L} \times \sigma_{\text{inelastic}}/f$ depends on the instantaneous luminosity $\mathcal{L}$, the pp inelastic cross section $\sigma_{\text{inelastic}}$ and the revolution frequency f. [71]

a peak luminosity of 7.7×10^{33} $cm^{-2}s^{-1}$ which is more than half the design luminosity, as shown in Table 2.1 together with other parameters relevant for the accelerator performance. Over 2012, the last year of data taking before the long shutdown[3], ATLAS collected about 20 fb^{-1} of pp collision data at $\sqrt{s}$=8 TeV. Figure 2.3(a) shows the delivered luminosity from the start of stable beams until beam dump and the luminosity recorded by ATLAS during stable beam conditions, the difference with respect to the delivered luminosity being due to Data AQuisition (DAQ) inefficiencies. Of the recorded luminosity, only a part is usable for analysis, and is what is called "good data", i.e. the data that satisfy Data Quality (DQ) requirements assessed after reprocessing (see Section 2.5).

In order to increase the luminosity LHC operates with a high number of protons per bunch as well as a high number of bunches per beam and reduces the inter-bunch latency

[3]LHC terminated the pp program at the end of 2012, operated proton-heavy ion collisions for two months at the beginning of 2013 and then stopped for what is called the first long shutdown. During this two-years time the accelerator and the experiments as well will undergo substantial maintenance and upgrade works, in order to be re-operated in 2015 with higher performance at a higher center of mass energy for particle collisions.

time. This overall defines a set of challenges that physics analysis will face associated to the high luminosity. Even at the detector design stage, the high frequency of collision environment foreseen influenced the choice of radiation resistance material for the experiment sub-systems. Concerning directly the physics instead, the main problematic is *pile-up*.

Pile-up events are distinguished between *in-time* and *out-of-time* pile-up. The first ones come from the multiple inelastic scatterings of protons in the same bunch, as if we consider a cross-section of 80 mb at the nominal luminosity of 10^{34} cm^{-2}s^{-1} the number of events per second will be something like a billion. This translate, at a collision frequency of one crossing every 25 ns, to about 20 interactions per crossing that will be detected simultaneously. A useful observable to estimate in-time pile-up is the number of reconstructed primary vertices (see Section 4.2) N_{PV}. In addition, on the other hand, the inter-bunch time interval is so short that the electronics reading the detector might not keep up with the frequency of collisions, leading to the cumulation of events that happened in different beam crossings. This is the effect we refer to as out-of-time pile-up, and a good estimator for it is the average number of pp interactions per bunch crossing at the time of the event, $< \mu >$, which recalling Equation 2.1 is defined as:

$$< \mu >= \frac{LA}{nf}, \tag{2.3}$$

with L being the average instantaneous luminosity over a time period $\Delta t \gg 600$ ns. The maximum values reached by the variable $< \mu >$ during the three years of data taking are reported in Table 2.1.

Finally, ATLAS makes use of a three-level trigger system (described in Section 2.4) to identify and record only the events of interest, while the pile-up issues are dealt with at the analysis reconstruction level.

2.2 The ATLAS detector

ATLAS (A Toroidal LHC ApparatuS) [65] is a general purpose experiment aimed at exploring a vast range of physics scenarios and designed to measure the particles produced in pp collisions at the LHC at unprecedented energies and instantaneous luminosities. It is the biggest detector of its kind ever built (it's 46 m long and 25 m high) is characterized by a full coverage of the space around the pp interaction point and complete containment of the particles produced in the collision. Different subsystems are layered concentrically one after the other, each of them pursuing a specific task. Right around the interaction point (IP) where the LHC makes protons collide there is the Vertex Detector, reconstructing charged particles trajectories that are bent by the first solenoid magnet surrounding the Vertex Detector. Particles going through it then encounter the two calorimeter systems, the Electromagnetic and the Hadronic one. Muons are the only particles that will pass the calorimeters material (beyond neutrinos) and a dedicated Muon Spectrometer is the last piece of detector, embedded in a huge toroidal magnet. The detector complex is presented as a schematic in Figure 2.4, and a drawing of particle detection in the various subdetector systems is shown in Figure 2.5.

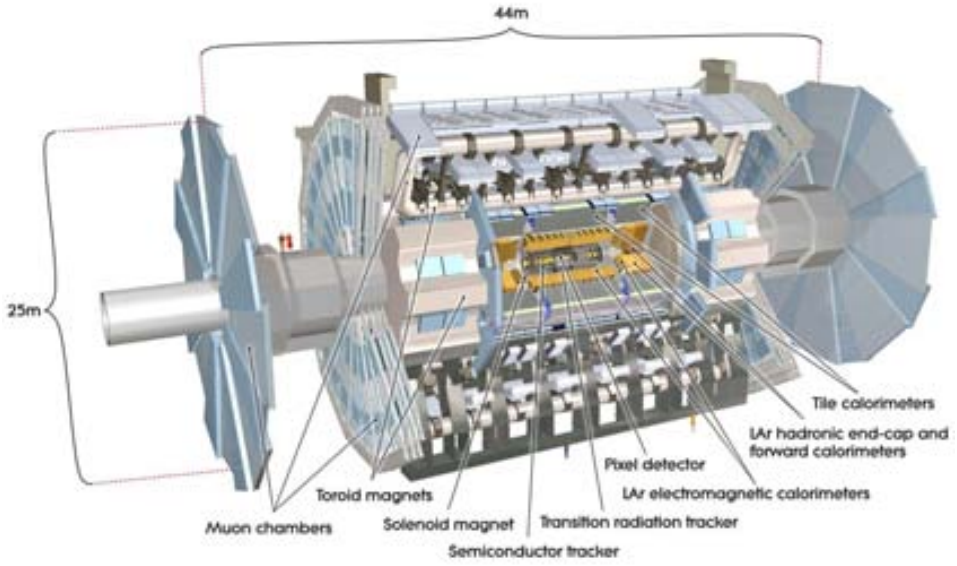


Figure 2.4: Schematic drawing of the ATLAS experiment. The detector subsystem are indicated as well as the total dimensions.

2.2.1 Coordinate system

Protons from the two circulating beams are made to collide in the center of the ATLAS detector, in the region that takes the name of Interaction Point (IP). The IP is taken as the origin of a three dimensional XYZ right-handed coordinate system. The Z axis is tangent to the trajectory of the beams while the XY plane is perpendicular to it and defines a symmetry plane for the detector, dividing it into the A and C sectors, respectively in the positive and negative Z semi-axes. Figure 2.6(a) shows a schematic of the coordinate system.

In terms of polar coordinates, the Z axis is again along the beam axis and in the transverse plane the R and ϕ coordinates are defined with ϕ ranging between $-\pi$ and $+\pi$ with respect to the X axis. In terms or spherical coordinates (see Figure 2.6(b)), the radial vector R originates from the IP, the azimuth ϕ is the same as the polar angle ϕ, and the polar angle θ is measured with respect to the Z axis and ranges between 0 and π.

Since the interaction initial energy is unknown, being dependent on the parton distribution functions for the proton energy, it is useful to define the transverse component of variables of interest[4] like the energy and the momentum, being taken as the projection on the XY plane:

$$E_{\mathrm{T}} = E \sin\theta, \qquad p_{\mathrm{T}} = p \sin\theta. \tag{2.4}$$

Another common variable used at hadron colliders to describe the polar distribution and

[4]These quantities transverse initial value will be, indeed, zero, as the protons are accelerated along the Z axis.

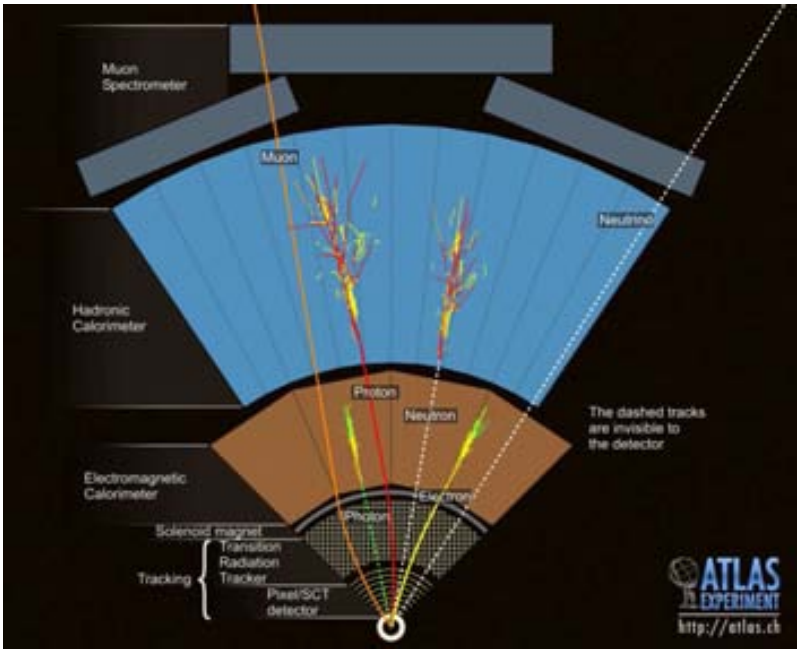


Figure 2.5: Drawing of the detection of particles going from the interaction point through the whole detector.

preferred to the simple polar angle θ is the pseudorapidity η:

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right); \tag{2.5}$$

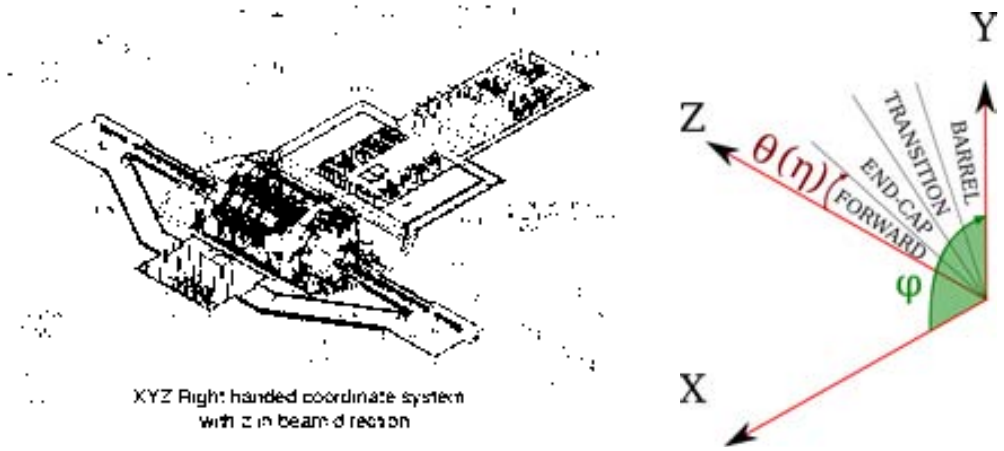
which, for relativistic regimes, is equal to the rapidity y:

$$y \equiv \frac{1}{2}\ln\left(\frac{E+p_Z}{E-p_Z}\right); \tag{2.6}$$

and Δy and $\Delta\eta$ are Lorentz invariant. The pseudorapidity is preferred to the rapidity as it does not require knowing the particle mass but only its polar position. The distance between two particles is often referred to in terms of ΔR:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.7}$$

Figure 2.6(b) shows how different pseudorapidity regions are named. Particles along the Z axis have a pseudorapidity $|\eta| = \infty$, particles along the Y axis have a pseudorapidity $|\eta| = 0$. ATLAS has an excellent hermeticity and is able to cover pseudorapity regions up to $|\eta| = 4.9$. Typically, physics analysis consider objects in the pseudorapity region $|\eta| < 2.5$. For a quick visualization of the correspondence in terms of polar angle distribution, some pseudorapidity values are reported in Table 2.2.



(a) (b)

Figure 2.6: (a) Drawing of the ATLAS experiment with the cartesian coordinate system. The positive X axis points towards the center of the LHC ring. The positive Z axis points todards the anti-clockwise circulating direction of beam 2. (b) Simple schematic showing the spherical coordinates and the region definition in terms of the absolute value of the pseudorapidity η. These regions are symmetrical with respect to the transverse XY plane.

θ	0°	5°	10°	20°	30°	45°	60°	80°	90°
η	∞	3.13	2.44	1.74	1.31	0.88	0.55	0.175	0

Table 2.2: Pseudorapidity vs polar angle values.

2.2.2 Magnets

ATLAS is provided with four superconducting magnets that allow the measurement of charged particles momenta by curving their trajectory.

A central solenoid, 5.3 m long and 2.4 m in diameter, sits around the inner detector and produces a 2 T magnetic field along the direction parallel to the beam axis. It is only 45 mm thick (equivalent to 0.66 radiation lenghts X_0) and is cooled with liquid Helium, sharing the cryostat with the electromagnetic calorimeter.

Paired to the muon spectrometer, the superconducting air-core toroid magnet (Figure 2.7) has an open structure with eight superconducting toroidal coils in the barrel part (each 25.3 m long, located at the outer diameter of 20.1 m) and two end-cap systems made of eight coils. The field strength varies strongly with ϕ: in the barrel region ($|\eta| < 1.4$) is 1.5-5.5 Tesla·m; in the end-caps ($1.6 < |\eta| < 2.7$) 1-7.5 is Tesla·m. Such configuration of the magnets gives a field orthogonal to the muons trajectory.

2.2.3 Inner detector

The Inner Detector (ID) is the subsystem closest to the IP and tracking charged particles arising from collisions allows for the measurement of their momentum and vertex reconstruction with excellent resolution. At the design choices level, radiation resistance had to be taken into account, as well as reducing the amount of material to be placed in front of

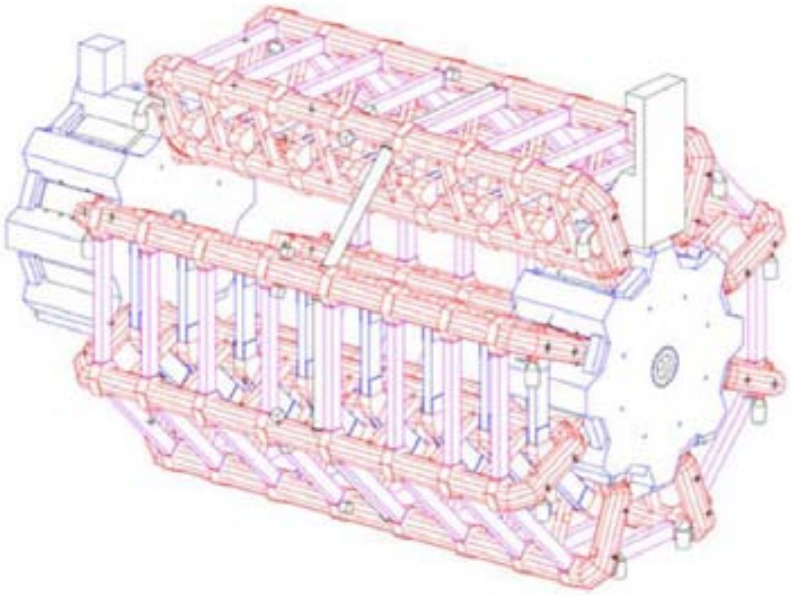


Figure 2.7: Toroidal magnet system.

the calorimeters to avoid spoiling the energy measurement. This quantity varies between 0.5 and 2.5 X_0 depending on the pseudorapidity region, most of it coming from supporting equipment. This material is responsible for photon conversions and electron bremsstrahlung.

The ID is surrounded by the central solenoid magnet (Section 2.2.2) and is composed by three subsystems, from the closest to the furthest from the IP: the pixel detector, the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

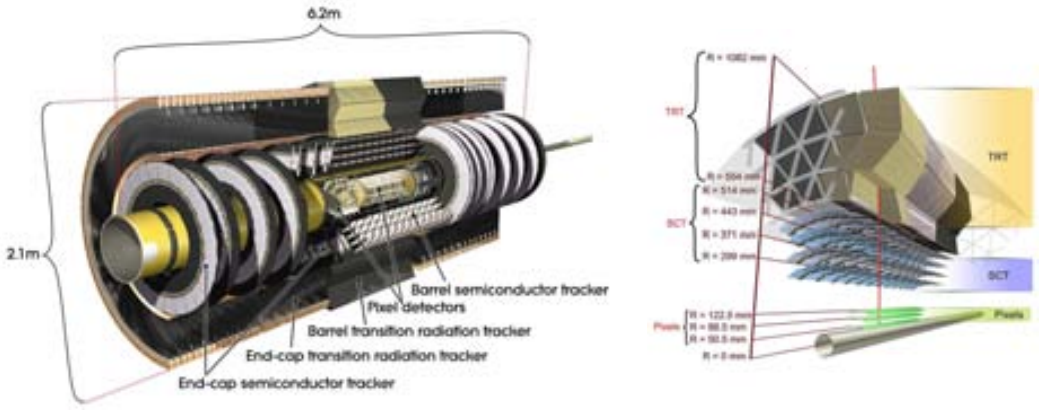
Pixel detector

The first subsystem covers the region $|\eta| < 2.5$ and is composed by three cylindrical layers in the barrel region, each of them distant from the beam by 50.5 mm, 88.5 mm and 122.5 mm respectively, and by three concentric discs in the end-cap region, each of them distant from the beam by 49.5 mm, 58.0 mm and 65.0 mm respectively. Each silicon pixel has a size of 50×400 μm^2 and is 250 μm thick, with in total $\sim$80.4 million readout channels to achieve a very fine granularity. The precision is of 10 μm in $R\phi$ and 115 μm in Z and R in the barrel and end-cap region respectively.

The very first layer is called $B-$layer as, thanks to its position really close to the IP, allows for the reconstruction of secondary vertices associated with the production of short lived particles such as $B-$hadrons. This information is very useful to identify jets originating from the fragmentation of b quarks.

Semiconductor tracker

After the three layers of pixel detectors, come four layers of silicon strip detectors. The SemiConductor Tracker (SCT) also covers the region $|\eta| < 2.5$ with a barrel and end-cap



(a) (b)

Figure 2.8: (a) Schematic of the ID system. (b) Detailed schematic of the barrel section of the ID showing the three subsystems and reporting the distance to the center of the beam pipe.

design similar to the pixel detector one, being composed by eight silicon strips (two per layer) 128 mm long and 80 μm large. It makes use of ∼6.3 millions readout channels and the resolution achieved is of 17 μm in $R\phi$ and 580 μm in Z (R) in the barrel (end-cap) region.

By allowing for four redundant position measurements[5], the SCT contributes mainly to the momentum reconstruction.

Transition Radiation Tracker

In order to reduce the amount of material in front of the calorimeters, and to reduce the construction costs as well, in the third subsystem the semiconductor technology has been substituted with straw detectors. The Transition Radiation Tracker (TRT) consists of thin proportional chambers made of straw polyimide drift tubes, 4 mm in diameter. The drift tubes are filled with a gas mixture composed of: 70% Xenon, 27% Carbon Dyoxide, 3% Oxygen. The anode collecting the electrons from the ionized gas at the passage of the charged particle is made of tungsten covered in gold.

In the barrel region the tubes are 144 cm long and placed parallel to the beam axis, while in the end-cap region they are 37 cm long and positioned radially in wheels, with layers of radiator foils alternated to layers of straws. The resolution achieved is of 130 μm in $R\phi$ and Zϕ in the two regions respectively. The covered pseudorapidity region is of $|\eta| < 2.0$ and the readout is composed by ∼351000 channels.

About 36 measurements per track are taken, and since each channel provides two independent thresholds per hit, it is possible to discriminate between electrons and pions, since the former will more likely reach the high threshold.

In the end, the combination of the three ID subsystems gives very precise $R\phi$ and Z measurements, as well as good track pattern recognition. The resolution on the transverse

[5] One of the coupled layers is rotated of 40mrad with respect to the other, which is parallel to the axis, giving a small stereo angle for a redundancy in the ϕ coordinate measurement.

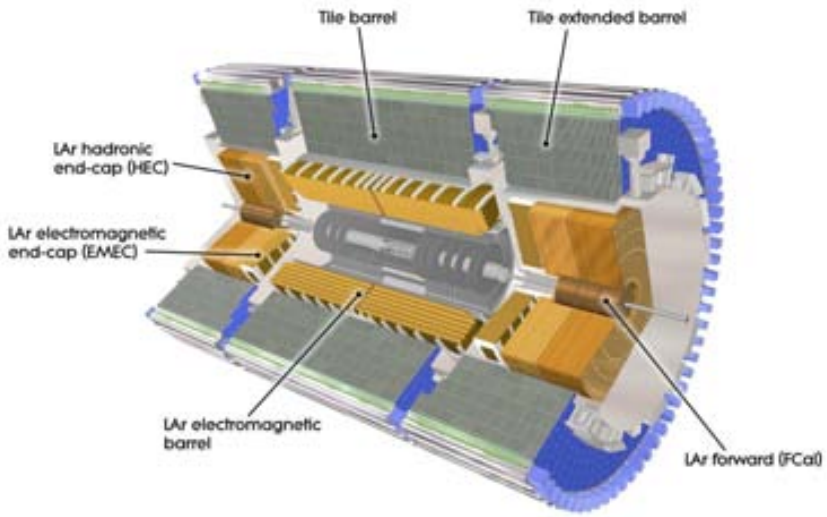


Figure 2.9: Schematic of the calorimeter complex of the ATLAS detector.

momentum, measured with cosmic muon calibration runs [72], is:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = P_1 \oplus P_2 \times p_{\mathrm{T}}, \tag{2.8}$$

where $P_1 = 1.6 \pm 0.1\%$ and $P_2 = (53 \pm 2) \times 10^{-5}$ GeV^{-1}. This means a resolution of $\sim$1.6% for tracks with p_{T} $\sim$1 GeV and $\sim$50% for tracks with p_{T} $\sim$1 TeV.

2.2.4 Calorimeters

Particles leaving the ID and surviving the crossing of the central solenoid magnet will face the calorimeter system, depicted in Figure 2.9. The full system is characterized by a coverage in pseudorapidity up to $|\eta| < 5$ and an almost full coverage in ϕ. With its 22 X_0 and 24 X_0 radiation lengths of material in the barrel and end-cap regions respectively it is also able to stop most of the non-muon particles from the interaction. Besides particles energy measurement, the calorimeters provide particle identification information, discriminating electrons, photons and jets, and the determination of the missing transverse energy.

Different technologies are used in the barrel, end-cap and forward regions for both the electromagnetic and the hadronic calorimeters. All of them are sampling calorimeters, with a dense medium acting as absorber to stop particles and start showers, and an active material to detect the signal from ionization. For the electromagnetic calorimeters and the forward hadronic calorimeter liquid argon is used as active medium, while the barrel and extended-barrel hadronic calorimeter uses scintillating tiles. The liquid argon is cooled at a temperature of about 88 K, with the use of two sets of cryostats: the barrel electromagnetic calorimeter shares the cryostat with the central solenoid; the end-cap and forward electromagnetic calorimeter and the hadronic end-cap calorimeter share a cryostat in the foward region.

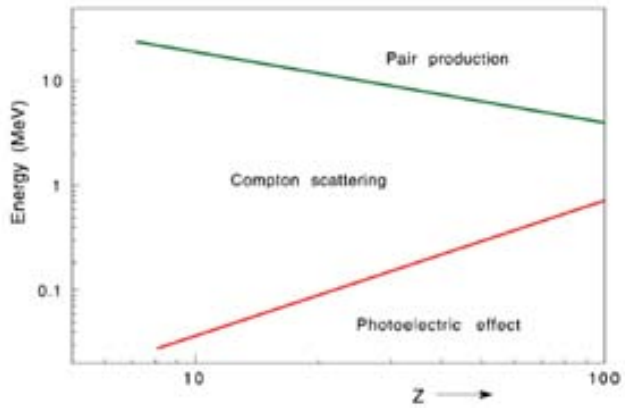


Figure 2.10: Domains in term of photon energy and Z number of the absorber material in which photoelectric eect, Compton scattering and pair production are the favorite processes for energy loss [73].

Particles interact both with the passive and active material, but only the energy released in the active samples will be detected. The processes involved in the shower formation are several and mainly electromagnetic. Photons in matter can undergo the photoelectric effect, Compton scattering and $\gamma \to e^+e^-$ pair formation. The general contribution of these processes depends both on the photon energy and on the atomic number Z of the material, and is shown in Figure 2.10. Electrons and positrons can ionize atoms and molecules, produce bremsstrahlung $e^{\pm} \to e^{\pm} + \gamma$ and emit Cerenkov radiation. Unless the calorimeter has been specifically designed for it, Cerenkov radiation does not contribute much, while ionization is the main process for energies up to $\sim$ 100 MeV, where bremsstrahlung starts to dominate.

In general, these cascade of events continues until a certain threshold is reached, and the final number of particles produced is proportional to the energy of the first particle originating the shower.

Also hadrons interact with matter, either ionizing it (if charged) or by nuclear interactions. The problem of the latter process is that this energy release is often not directly detectable, like in nuclear breakups and excitations, and is therefore called "invisible energy". Secondary hadrons will be produced, forming the hadronic part of the shower, but sooner or later something like $\pi^0 \to \gamma\gamma$ will happen and the shower will develop electromagnetically further on.

The average fraction of electromagnetic and hadronic shower components is a characteristic of the sampling calorimeter and depends on the choice of the passive and active material and on the design. Calorimeters are said to be *non-compensating* if, like the ATLAS calorimeters, the detection of hadronic showers is less efficient than the one of electromagnetic showers. Calorimeters with a similar response for the two components are called *compensating*, while calorimeters more efficient when revealing hadronic showers are *over-compensating*.

The performance for the energy resolution is parametrized by the following formula:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \tag{2.9}$$

where the terms of the sum correspond, respectively, to a "stocastic" term related to how shower develops in the sampling calorimeter; to a "noise" term including the contribution from electronic noise and pile-up energy fluctuation; to a systematic term that depends on calibration, shower containment, inactive material and on the linearity of the response as well.

The goal energy resolution for the liquid argon calorimeters is [74]:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{170 \text{ MeV}}{E} \oplus 0.7\%, \tag{2.10}$$

while for the hadronic barrel calorimeter is [75]:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 5\%. \tag{2.11}$$

Test-beam runs to measure the response of the two calorimeters to electrons and single pions respectively have shown results comparable to the goal resolutions.

Electromagnetic calorimeter

The electromagnetic calorimeter, also called LAr calorimeter (from Liquid Argon, the active material), can measure electrons and photons energies in the range from 50 MeV to 3 TeV. In the barrel region it is referred to as EMB (ElectroMagnetic Barrel), is divided into two identical semi-barrels EMBA and EMBC separated at Z=0 by a 6 mm gap and covers the pseudorapidity region $|\eta| < 1.475$. Two end-cap detectors (EMEC, ElectroMagnetic End-Cap), divided into two coaxial wheels, cover the pseudorapidity regions $1.375 < |\eta| < 3.2$. A pre-sampler, extended over $|\eta| < 1.8$, stands in front of the EMB and allows for the measurement of the energy the particles lost before reaching the EMB i.e. crossing the material of the ID, the central solenoid and the cryostat.

Three longitudinal samples in the EMB are designed for different tasks. The first sample, $4.3X_0$ long, is finely segmented in η to precisely measure the direction in pseudorapidity of the particles with thin readout strips of $\Delta\eta \times \Delta\phi = 0.0031 \times 0.098$. This helps for photon/$\pi^0$ discrimination and as well for separate close-by γs from π^0 decay. The second sample, $16X_0$ long, contains the bulk of electrons and photons energy deposit. It is divided in towers with dimension $\Delta\eta \times \Delta\phi = 0.025 \times 0.0245$ and provides the position measurement of the cluster. The 95% of the energy of the shower is deposited in a matrix of 3×7 towers $\Delta\eta \times \Delta\phi$. The third sample, $2X_0$ long, is coarsely segmentes and collects the last bit of the longitudinal development of the electromagnetic showers. Towers in this region have a dimension of $\Delta\eta \times \Delta\phi = 0.05 \times 0.0245$.

Also the EMEC is divided in three longitudinal samples (two in the region $1.375 < |\eta| < 1.5$), and besides the lead, also the thickness of the liquid argon layers are varied in the radial direction.

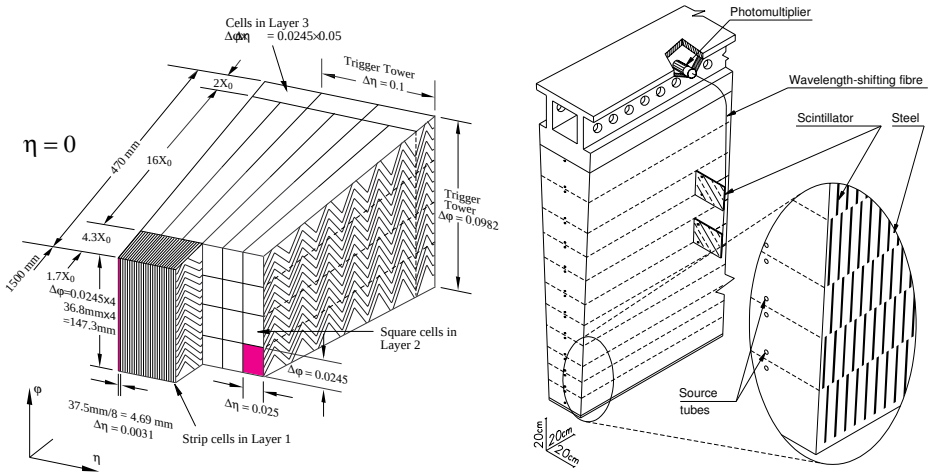


Figure 2.11: (a) Schematic drawing of a module of the Electromagnetic barrel calorimeter. (b) Schematic drawing of a module of the Hadronic barrel calorimeter.

The absorbing material is lead shaped into an accordion geometry to achieve full symmetry in ϕ, as shown in the drawing of Figure 2.11(a). Signal from the ionization produced in the liquid argon is collected by an electrode in the middle of the active material region, fixed into a honeycomb structure.

The thickness of the absorber layers depend on the pseudorapidity in order to make particles entering the system with different incident angles cross the same amount of material.

Hadronic calorimeters

Hadronic showers have typically a much longer shape than electromagnetic ones, and need therefore in general more interaction lenghts of material to be fully contained. Hadronic calorimeters are therefore designed to completely absorb high-energy hadrons, which will deposit only some (small) part of their energy in the electromagnetic calorimeter.

Hadronic barrel calorimeter

The hadronic calorimeter in the barrel and extended barrel region, going up to $|\eta| < 1.7$, is made of scintillating tiles as active material with lead as absorber and is commonly referred to with the name of TileCal. The light in the ultraviolet range that is generated in the tiles is collected through wavelenght shifting optical fibre (Figure 2.11(b)).

TileCal sits just after the electromagnetic calorimeter and measures the energy and position of jets and isolated hadrons. It is divided in depth in three layers with varying lenght (1.4, 4.1, 1.8 hadronic interaction leghts λ in the barrel and 1.5, 2.6, 3.3λ in the extended barrel) and segmentation ($\Delta\eta \times \Delta\phi = 0.1{\times}0.1$ in the first two layers, $\Delta\eta \times \Delta\phi = 0.2{\times}0.1$ in the third), and in 64 slices in ϕ, each of $\Delta\phi \sim 0.1$.

The redout channels are grouped into cells that form a pseudo-projective geometry in η.

Hadronic end-cap calorimeter

The Hadronic End-Cap calorimeters (HEC) use copper as passive material and liquid argon as active material, chosen for its radiation hardness in a region ($1.5 < |\eta| < 3.2$) exposed to a significant amount of particle flux. Each HEC is composed by two independent wheels with granularity varying with η: in $1.5 < |\eta| < 2.5$ $\Delta\eta \times \Delta\phi$ is 0.1×0.1 in the first two longitudinal layers, 0.2×0.1 in the last one; in $2.5 < |\eta| < 3.2$ $1.5 < |\eta| < 2.5$ $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ in all the three samples.

The HECs collect the energy from particles that are not completely contained in the EMECs and in particular are used to reconstruct jets and the missing transverse energy.

Forward calorimeter

The Forward Calorimeter (FCal) cover the very forward region of pseudorapidity $3.1 < |\eta| < 4.9$ making the calorimeter system achieve its good hermeticity and minimize the energy losses. It has an electromagnetic part that uses copper as absorber and two hadronic compartments with tungsten as passive material.

2.2.5 Muon spectrometer

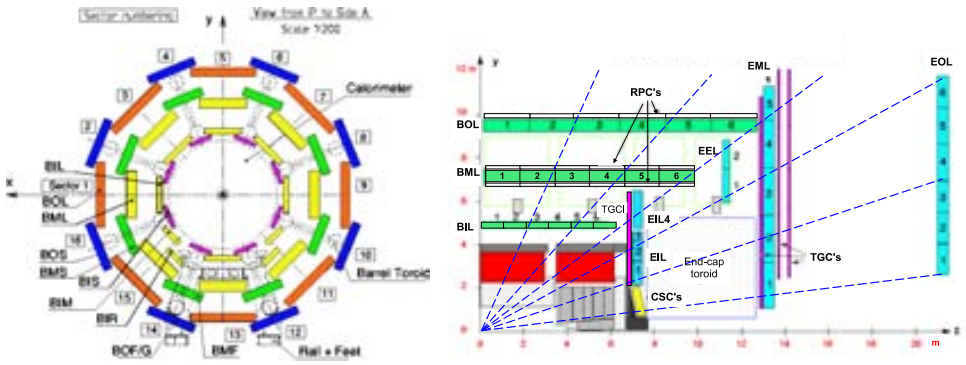
The most external detector system is the muon spectrometer, a combination of toroidal superconducting magnets (Section 2.2.2) and precision chambers providing a measurement of the momentum of muons in $|\eta| < 2.7$ in addition to the measurement from the ID. It is also equipped with an independent trigger system used for the first event triggering stage (see Section 2.4.1) active in the pseudorapidity region $|\eta| < 2.4$.

Four sub-detectors compose the muon system: Monitored Drift-Tube (MDT) chambers, Cathode Strips Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The layout changes in the barrel and end-cap regions, and is schematically shown in Figure 2.12(b): in the barrel region, chambers are arranged in three cylindrical layers around the beam axis, one layer being inside the magnet; in the end-caps these three layers are placed perpendicular to the beam axis.

Detection chambers

MDTs and CSCs are used to detect muons in the pseudorapidity regions $|\eta| < 2.0$ and $2.0 < |\eta| < 2.7$ respectively. MDTs are proportional chambers constituted by pressurised drift tubes made of aluminium with a diameter of 30 mm and lenght varying from 0.9 m to 6.2 m. The gas mixture in them is 93% argon and 7% carbon dioxyde, the anode is a 50 μm tungsten-rhenium wire producing a radial electric field. Each chamber is composed by a group of six or eight tubes placed transverse to the beam axis. This number of tubes allows for a very good track reconstruction and high reduction of the fake tracks from random associations of background hits, providing a resolution on position of 80 μm.

The CSCs are used at higher η to better cope with the higher particle flux. They are arranged in a system of two disks with eight chambers each. Each chamber contains four multiwire proportional chambers (the CSCs) with wires oriented in the radial direction, spaced by 2.5 mm and in the same gas mixture of argon and carbon dioxyde as the MDTs.



(a) (b)

Figure 2.12: (a) Cross section of the barrel muon system. (b) Lateral section of the muon system. Barrel MDTs are shown in green, end-caps MDTs in light blue, CSC in yellow, TGCs in magenta, RPCs in white.

The cathode strips are oriented one perpendicularly to the anode wires (and gives the precision coordinate) and the other parallel to the wires (and gives the transverse coordinate). The resolution provided by the interpolation between the charges induced on neighbouring cathode strips ranges between 50 and 70 μm.

Trigger chambers

For trigger purposes detectors with faster response than drift tubes are needed[6]. MDTs and CSCs are then coupled with special layers of trigger chambers: in the barrel region, the MDT's second layer is covered on both sides by RPCs, while MDT's third layer is covered by a RPC alternatively on the inner and outer side; in the endcaps, TGCs cover the inner side of MDT's first and third layers.

A RPC is a detector with a gas-gap between two resistive bakelite plates separated by 2 mm and containing a gas mixture of $C_2H_2F_4$ (94.7%), Iso-C_4H_{10} (5%) and SF_6 (0.3%). RPCs measure six points per coordinate for each particle, quickly collecting the avalanches with two orthogonal sets of pick-up strips that provides a position resolution of 1 cm in each plane and 1 ns time resolution, allowing for individual bunch crossing discrimination. Also RPCs provide the ϕ coordinate for the tracks in the final analysis, since MDTs only give the η coordinate.

TGCs are similar to CSCs, have 1.8 mm wire-to-wire separation and 1.4 mm wire-to-cathode separation. They use a highly quenching gas mixture of CO_2 55% and n-C_5H_{12} 45% and provide a spatial resolution of about 1 mm and a time resolution of 5 ns.

2.3 Forward sub-detectors

ATLAS is equipped with some detectors in the forward regions to perform additional measurements or monitoring studies. In particular, the Minimum Bias Trigger Scintillators

[6]Drift-time in tubes with a diameter of $\mathcal{O} \sim 10$ mm can be of ~ 500 ns, too long with respect to the 25 ns spacing of the bunch crossings.

Figure 2.13: Schematic of a section of TileCal barrel and extended barrel modules, with the cells division. The parts labelled with "E" are the MBTS.

(MBTS), that are somehow embedded in the structure of TileCal extended barrel modules (see Figure 2.13) and share with it the readout electronics, as they are also read by wavelenght-shifting fibers. The MBTS consist of 32 scintillator paddles assembled in two disks covering the pseudorapidity region $2.09 < |\eta| < 3.84$ and are used for trigger purposes to detect minimum bias activity during the first runs of the LHC.

MBTS are also used for relative luminosity measurements, but there are two detectors specifically built to determine the luminosity delivered to ATLAS: LUCID and ALFA. LUCID (LUminosity measurements using Cerenkov Integrating Detector) is made of 32 tubes surrounding the beam pipe 17 m far from the interaction point on both sides of ATLAS and measures the luminosity bunch by bunch. ALFA (Absolute Luminosity For ATLAS) is only activated during special runs, and consists of 8 scintillating fibers detectors placed at 240 m from the interaction point inside roman pots, above and below the beam pipe.

Another luminosity monitorer is the Zero-Degree Calorimeter, whose main purpose is to determine the centrality of heavy-ion collisions. Placed at 140 m from the interaction point on both sides of the beam axis, is made of quartz rods alternated with tungsten plates.

Finally, the Beam Condition Monitor (BCM) is made of two sets of diamond sensors located 184 cm close to the interaction point along the beam and 5.5 cm close along R. Its task is to detect beam losses, potentially harmful for ATLAS, and in that case to alert LHC in order to stop the accelerator.

2.4 Trigger system

It was already introduced at the beginning of this Chapter the issue faced by LHC experiments of dealing with a huge amounts of events at very high frequencies. We remind that considering the nominal LHC luminosity of 10^{34} cm^{-2} s^{-1} a rate of interactions of 40 MHz is expected! This poses serious technical difficulties as the maximum frequency at which data can be recorded is limited to 200 Hz considering the limited capacity for storage.

ATLAS developed a trigger system able to reduce by a factor of 10^6 the amount of data to be kept by selecting only interesting physics events. The system is divided in three levels characterized by increasing sofistication and diminishing speed. At the very first indeed we will need a really quick and simple criterium to reject uninteresting events. The reduced information can then be processed with somehow slower logic by the other two High Level Triggers (HLT). A drawing of the system is shown in Figure 2.14.

Most of the trigger chains used for physics are un-scaled in the sense that all the events passing the selection are kept, but there are also pre-scaled trigger chains that contain either too many events or events considered not physically interesting. These trigger chains are used for checks or calibration rather than physics analysis, and the prescaling value P means that of all the events that would have passed the trigger, $1/P$ were accepted.

With the term "trigger chain" we refer to the sequence of selections defining a certain

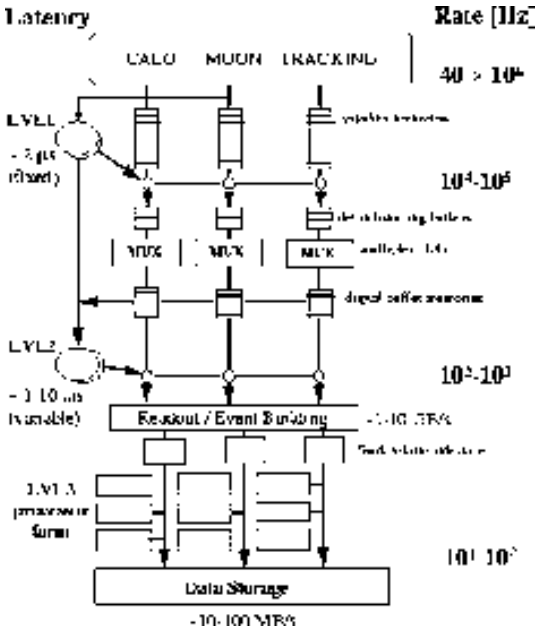


Figure 2.14: Schematic drawing of the three-level trigger system of ATLAS.

trigger object, with a naming convention like:

[LEVEL][N][TYPE(S)][THRESHOLD][ISOLATION][QUALITY],

where the components, from left to right, are: the trigger level used; the multiplicity of the type; the object candidate; the threshold applied to the transverse momentum or energy of the object candidate; the object isolation; the severity of the final algorithm requirements (this applies only to the Event Filter level).

Trigger chains define a *trigger menu*, where they are associated to their prescale value P, and which is chosen based on the physics program of the data taking period taking into account the LHC luminosity.

Defining the data taking period time unit as "Luminosity Block" (LB), typically a few minutes of data taking, information on beam conditions, detector performance and events passing any of the trigger chains of the trigger menu are stored to be then used in the analyses. All the LB occurring between the start and the end of a stable beam collision period compose a "run". Runs are finally grouped in "Data Periods", labelled with capital letters ("Period A", "Period B", *etc.*), when they pertain to the same general detector condition, machine configuration and trigger menu.

2.4.1 Level 1 trigger

The Level 1 trigger (L1) is completely based on the hardware of the detector, taking information from calorimeters, from the muon spectrometer trigger systems RPC and TGC (Section 2.2.5) and from the MBTS (Section 2.3) at 40 MHz (the frequency of the beam crossing) and reducing it to 75 kHz by choosing events with high transverse momentum or

high missing transverse energy.

Using dedicated fast front-end electronics (the typical decision time being less than 2 μs), calorimeter cells are analogically summed to build calorimetric towers which, if having an energy higher than a certain threshold, will activate a trigger chain.

These trigger chains will then be combined with the information from the muon spectrometer to form the so-called Region of Interest (RoI) that is passed to the next trigger level.

2.4.2 Level 2 trigger

Starting from the RoI, the Level 2 trigger (L2) will reduce the 75 kHz to 3.5 kHz of events with an average decision time of 40 ms. At this stage the information from the trackers is incorporated to the RoI to build candidate object (electrons, photons, muons) and better obtain its position and energy with simplified algorithms quick enough to respect the limit on the decision time.

2.4.3 Event filter

The last trigger, Level 3, is called Event Filter (EF) since at this point the physics objects are built using the same algorithms as the off-line reconstruction, with looser selections. With an execution time amounting to 4 s, the EF reduces the event rate to the goal value of 200 Hz. Events passing the EF are assigned to *streams* defined to separate the events into different datasets for different analysis interests, e.g. electron streams, muon streams, jet streams *etc.*

As an example, one of the trigger chains used in our analysis is `EF_mu24i_tight`: it selects events at the EF level with one muon with $p_T > 24$ GeV and some isolation requirement which passes the muon reconstruction algorithm cuts defined as "tight" (more on event reconstruction is reported in the dedicated Chapter 4).

2.5 Data Quality

The totality of pp collisions recorded by ATLAS, which differs from the amount delivered by the LHC because of data-taking inefficiencies, is still not 100% usable by physics analyses. Indeed, every subdetector needs to perform some routine checks on the quality of the data they recorded in order to certify that its performace was conform to the expectations. So-called "Good Runs Lists" (GRL) are compiled stating for each LB what was "OK" and what not. The single analyses will then decide which GRL to use, based on their specific needs of the individual subsystems.

Monte Carlo simulation

In science, little should be left to chance. Still, randomness over a huge number of trials leads to insights of something we could consider as "real". This is particularly useful when dealing with complex environments that can be described with a mathematical model in order to know what to expect from the actual data. Monte Carlo methods can describe hadron-hadron collisions, using pseudorandom numbers to simulate event-by-event fluctuations, and hence help us e.g. understanding the detector response and developing analysis strategies by predicting the sensitivity to the physics under study.

In this chapter, we will first go through a very brief overview of some concepts of the Quantum Chromodynamics theory useful to understand the evolution of a pp collision event (Section 3.1). Thanks to perturbation theory we can predict hard scattering cross sections, which is what is done as the first step of the Monte Carlo simulation chain, described in Section 3.2.1. Despite the theoretical ability to compute fixed order calculations for hard scattering cross sections, we still lack the possibility to describe QCD at low energy and, hence, hadron final states formation. *Shower algorithms* can associate to an hard event an arbitrary number of partons to constitute a final state with quarks and gluons (see Section 3.2.2). The *hadronization* of this yet unphysical final state is performed by means of phenomenological models of hadron formation, introduced in Section 3.2.3). The last ingredient for a complete picture is the treatment of the remnants from the incoming protons from which the hard interacting partons came from. For this we rely on the so-called "underlying event model", discussed in Section 3.2.4.

Finally, the products from the generated event are passed through a simulation of the ATLAS systems and digitized to give an output identical to the real detector one (Section 3.4). At this point objects are reconstructed in the same way for Monte Carlo and real data, as discussed in Chapter 4.

3.1 Phenomenology of pp collisions

Of the interactions making up the processes, for Monte Carlo simulating hadron colliders physics the most challenging part is related to the description of Quantum Chromodynamics

(QCD) phenomenology. Indeed, despite its theorethical framework being successful and verified, calculations are difficult and often need approximations.

3.1.1 Proton structure

The proton is a bound state of three *valence* quarks which carry each a fraction x of the proton momentum unpredicted theoretically and described by parton distribution functions (PDFs) $f_i(x)$. The probability for a parton i to carry a momentum fraction between x and $x + dx$ is $f_i(x)dx$ and the following condition holds:

$$\int_0^1 x \sum_i f_i(x)dx = 1. \tag{3.1}$$

PDFs are measured in deep inelastic scattering experiments and are universal, not depending on the particular process used. It is observed that the valence quarks only carry about half of the proton total momentum, the rest being carried by virtual gluons continuatively exchanged by the quarks. These gluons in turn produce virtual $q\bar{q}$ pairs called *sea* quarks.

Various parametrizations are available and the most widely used come from the CTEQ and MRST/MSTW collaborations[1] The 2008 NLO 68 version of PDFs of valence quarks, gluon and sea quarks from the MSTW group are shown in Figure 3.1 as a function of the momentum fraction for two values of transferred momentum Q^2 at which the proton is probed.

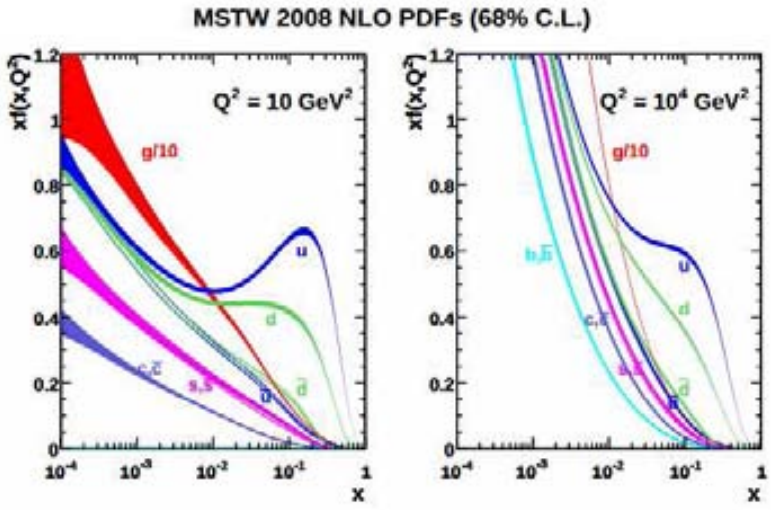


Figure 3.1: Proton PDF functions at transfer momentum Q^2=10 GeV (Q^2=10000 GeV) on the left (right) [76].

[1] See for more information the Les Houches Accord PDFs (LHAPDF): http://lhapdf.hepforge.org/ and the collaborations pages: http://www.phys.psu.edu/cteq, http://mstwpdf.hepforge.org/.

3.1.2 Factorization theorem

To treat infinities arising from divergent contributions in loop diagrams, an arbitrary renormalization scale μ_R is to be introduced. As a consequence from requiring that the physical observables be independent from the choice of μ_R, the strong coupling constant α_S does depend on the energy scale at which the coupling is observed and is, at the leading order:

$$\alpha_S(\mu^2) = \frac{\alpha_S(\mu_R^2)}{1 + (11 - \frac{2}{3}n_f)\frac{\alpha_S(\mu_R^2)}{2\pi}\ln\frac{\mu^2}{\mu_R^2}}, \tag{3.2}$$

with n_f being the number of quark flavors of the theory and μ the energy at which we observe the process.

This means that the coupling α_S decreases with increasing energy scale (small distances) as a consequence from the factor $11 - \frac{2}{3}n_f$ (the number 11 coming from the self interaction of gluons) being positive in the theory with $n_f = 6$, while it increases at lower energies (high distances). These two properties goes under the name of *asymptotic freedom* and *confinement* respectively: at low α_S quarks and gluons interact very weakly with each other and it is possible to use perturbation theory and the parton model [77], which treats partons as free and non-interacting; when α_S is large instead, partons tend to bound together into colorless hadrons, predictions from perturbative calculations become less reliable and soft QCD interactions are typically modeled using tunings from experimental data.

The *factorization theorem* [78] allows us to separate the two components to compute the cross section as a product of probability functions with the short distance cross section $\hat{\sigma}_{ab}$ computable in perturbation theory (pQCD, from perturbatve QCD) as a power expansion of the strong coupling constant $\alpha_S(\mu_R)$:

$$\begin{aligned}\sigma_{pp\to X} &= \sum_{a,b}\int_0^1 dx_a dx_b\ f_a(x_a,\mu_F) f_b(x_b,\mu_F)\hat{\sigma}_{ab}(x_a p_a, x_b p_b, \mu_R, \mu_F) \\ &= \sum_{a,b}\int_0^1 dx_a dx_b\ f_a(x_a,\mu_F) f_b(x_b,\mu_F) \times \left[\hat{\sigma}_0(\hat{s}) + \alpha_S(\mu_R^2)\hat{\sigma}_1(\hat{s},\mu_F^2) + \dots\right].\end{aligned} \tag{3.3}$$

Here, f_i $(i = a, b)$ are the standard PDFs for partons $a, b = \{g, u, \bar{u}, d, ...\}$ carrying fractions x_a, x_b of the proton longitudinal momentum, and $\sigma_{pp\to X}$ is the partonic scattering cross-section calculated in fixed-order perturbation theory. The μ_F coefficient is newly introduced *factorization scale*, μ_R is the renormalization scale for the QCD running coupling. Figure 3.2 shows a pictorial representation of the generic pp process.

Equation 3.3 refers to a sum of final state and is, therefore, inclusive. The choice for the renormalization and factorization scale values is usually to take them of the order of some typical hard scale entering the process like, e.g., the mass of the top quark for top-anti-top pair production. The cross section calculations are done in pQCD are in general either provided at Leading Order (LO) or at Next-to-LO (NLO). NLO calculations include corrections from virtual exchange or emission of a massless parton.

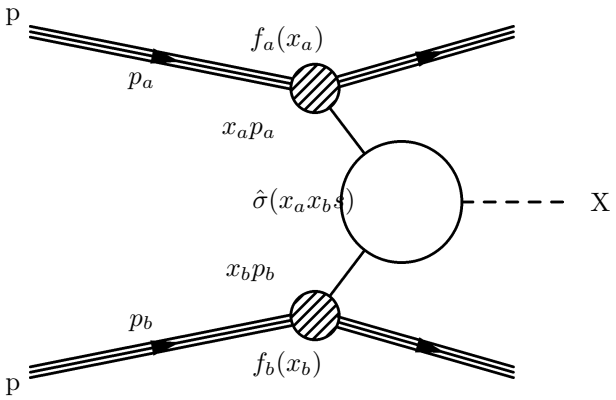


Figure 3.2: Diagram of a generic hard scattering process. The partons, extracted from the colliding pp pair, carry a momentum fraction with respect to the proton energy described by a parton distribution function. The scattering of the partons is computed perturbatively and hence the kinematic properties of the final state object X are predicted.

3.2 Simulation of pp collisions

The most interesting phenomena under study are scattering events with large momentum transfer or with production of massive particles, also called "hard scattering" events. Typical pQCD calculations can only provide an inclusive description of the process, while Monte Carlo programs give an exclusive picture of the event. A pp collision event in Monte Carlo simulation is the combination of different sub-processes, illustrated in Figure 3.3: it comes with a large library of Standard Model and Beyond Standard Model cross section from which the hard scattering process is chosen; it has a showering algorithm to generate the dominant pQCD effects, adding the emission of colored partons to the hard process enhanced with collinear and soft singularities; it implements hadronization for the high energy partons of the final state; it describes the underlying event according to some phenomenological model; it includes libraries for the weak decay of unstable hadrons.

A set of drawings in Figure 3.4 shows the sequence of the evolution of Monte Carlo event simulation, starting from the plain hard scatter and adding step by step the other components.

3.2.1 Hard interaction

Recalling what discussed in Section 3.1.2 about the computation in pQCD of the hard scattering cross section of a typical LHC $pp \to X$ event, we can rewrite Equation 3.3 as:

$$\begin{aligned}\sigma_{\mathrm{pp}\to\mathrm{X}} &= \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b \int f_a(x_a,\mu_F) f_b(x_b,\mu_F) \mathrm{d}\hat{\sigma}_{\mathrm{ab}}(x_a p_a, x_b p_b, \mu_R, \mu_F) \\ &= \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b \int \mathrm{d}\Phi_{\mathrm{X}} \; f_a(x_a,\mu_F) f_b(x_b,\mu_F) \times \frac{1}{2x_a x_b s} \left|\mathcal{M}_{\mathrm{ab}}\right|^2 (\Phi_{\mathrm{X}}, \mu_R, \mu_F),\end{aligned} \tag{3.4}$$

where we introduced the dependence on the final state phase space Φ_{X}, the parton flux $\frac{1}{2x_a x_b s}$ (with s being the center of mass energy squared) and the *matrix element* $\mathcal{M}_{\mathrm{ab}}$, which

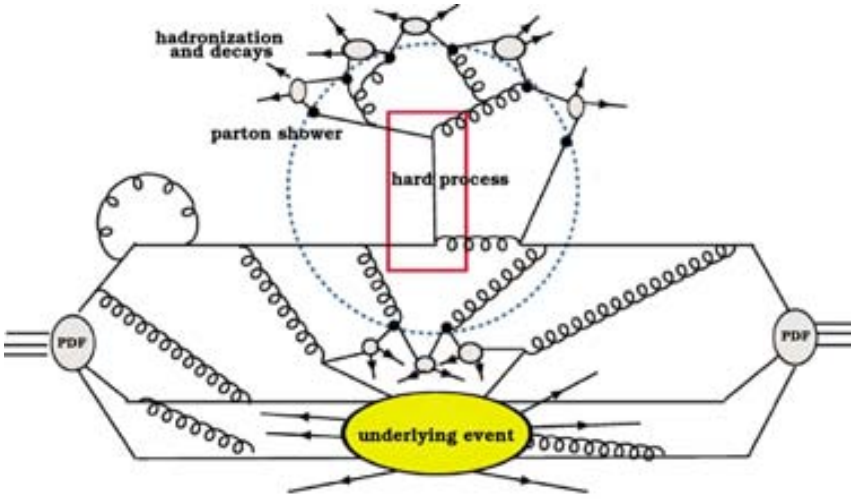


Figure 3.3: Drawing describing a hadron-hadron collision from the Monte Carlo point of view. Partons from the hadron share its energy according to the PDFs. The dotted circle separates pQCD events (hard scattering and initial and final state radiation) from non-perturbative effects (parton shower, hadronization, initial emissions included in the PDFs, and underlying event)[79].

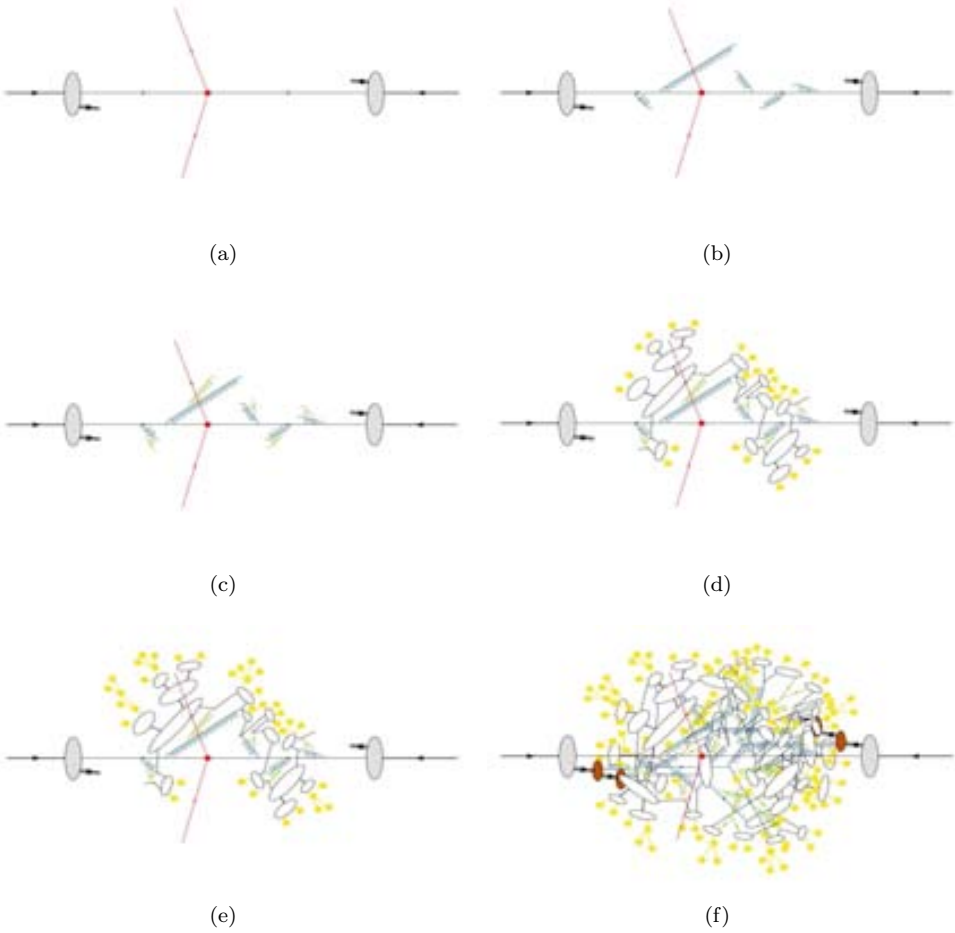


Figure 3.4: Set of frames of Monte Carlo event generation evolution: (a) hard scattering of two partons; (b) and (c) parton showering; (d) hadronization; (e) final particle decays; (f) underlying event simulation. Drawings from [80].

is also written as a sum over Feynman diagrams:

$$\mathcal{M}_{\mathrm{ab}} = \sum_i \mathcal{F}_{\mathrm{ab}}^{(i)}. \tag{3.5}$$

3.2.2 Parton shower

Parton shower adds higher order corrections to the hard scatter using an approximation scheme, since real radiative corrections to any inclusive quantity (like the hard cross section as computed at fixed order in pQCD) are divergent. The dominant contributions below a cut-off parameter, associated to collinear parton splitting or soft gluon emission, are included iteratively ordered in sequence of, typically, smaller emission angles.

There are three possible processes for QCD emission: $q \to gq$, $g \to gg$ and $g \to q\bar{q}$. The cross section then factorizes into the product of the parent parton production cross section times a splitting factor. Considering e.g. the $q \to gq$ splitting from a tree level process with $n+1$ final state particles we can graphically represent it as in Figure 3.5(a), with the quark k and the gluon l being emitted at a small angle θ. Mathematically we have:

$$\left|\mathcal{M}_{n+1}\right|^2 \mathrm{d}\Phi_{n+1} \to \left|\mathcal{M}_n\right|^2 \mathrm{d}\Phi_n \frac{\alpha_{\mathrm{S}}}{2\pi} \frac{\mathrm{d}t}{t} P_{q,qg}(z) \frac{\mathrm{d}\phi}{2\pi}, \tag{3.6}$$

where ϕ is the azimuth defined by $\vec{k}$ and$\vec{l}$ around the $k + l$ direction, z is an arbitrary parameter in general defined as a ratio between the energy of the particles emitted:

$$z = \frac{k^0}{k^0 + l^0}, \tag{3.7}$$

and t is the *hardness* parameter characterizing the divergence and the ordering of the splittings. It has the dimensions of a mass, and the preference is to take it as $t = E^2\theta^2$. The values for ϕ, z and t are generated randomly during the Monte Carlo simulation process. The function $P_{q,qg}(z)$ is the Altarelli-Parisi splitting function, and is the only term that changes in Equation 3.6 between $q \to gq$, $g \to gg$ and $g \to q\bar{q}$ splittings.

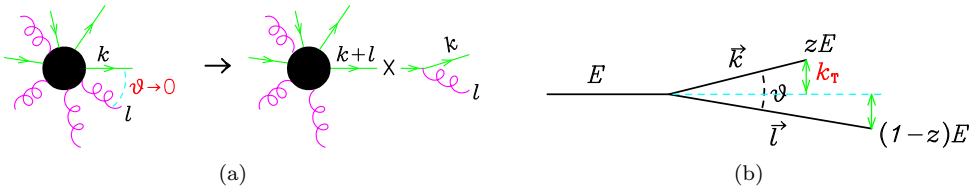


Figure 3.5: Left: graphical representation of the $q \to qg$ splitting. The black circles represent the $\mathcal{M}_{n+1}$ and $\mathcal{M}_n$ amplitudes of the tree-level processes. Right: kinematic of the splitting [81].

Factorization holds if the virtuality of the splitting parton $q^2 = (k+l)^2$ is negligible with respect to the energies entering the process, and can be applied iteratively as shown graphically in Figure 3.6. At this point, allowing for n splitting processes, the cross section can be written as:

$$\sigma_0 \frac{1}{n!} \alpha_S^n \log^n \frac{Q^2}{\lambda^2}, \tag{3.8}$$

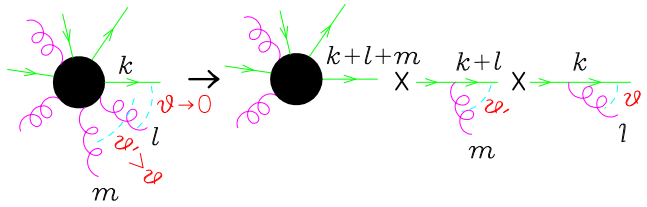


Figure 3.6: Recursive application of factorization, with the angles ordered as $\theta' \gg \theta \to 0$ [81].

where Q is the upper cut-off scale called *annihilation energy* that determines when the showering starts, and λ is the infrared cut-off. The shower ends when the virtuality q^2 reaches the *hadronization scale*, which is of the order of 1 GeV^2. From the cross section expression of Equation 3.8, this procedure is called "leading log approximation".

Once the shower is developed, the vertices and lines of the final configuration are assigned weights, which are for each vertex:

$$\theta(t - t_0) \quad \frac{\alpha_S(t)}{2\pi} \, \frac{dt}{t} \, P_{i,jl}(z) \, dz \, \frac{d\phi}{2\pi}, \tag{3.9}$$

and for each line are the so-called *Sudakov form factors*:

$$\Delta_i(t', t'') = \exp\left[-\sum_{(jl)} \int_{t''}^{t'} \frac{dt}{t} \int_0^1 dz \quad \frac{\alpha_S(t)}{2\pi} \; P_{i,jl}(z) \right] \tag{3.10}$$

where t' is the value of t at the upstream vertex, and t'' at the downstream vertex. If we reached the end of the graph, t'' is substituted by a cut-off t_0. The Sudakov form factors specify the range of the z parameter for which the splitting is resolvable and represent the probability of *not* splitting. Figure 3.7 shows the typical graph shape of a shower evolved with splittings strongly ordered in angle.

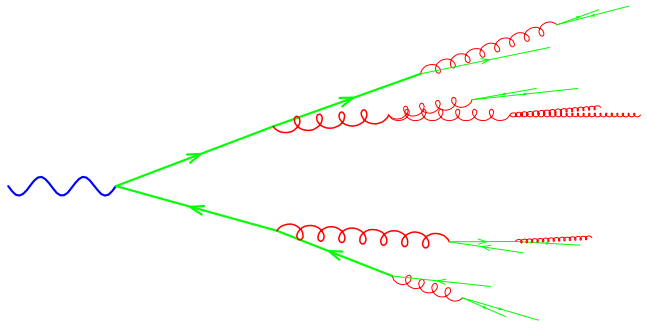


Figure 3.7: An example of a typical shower final graph [81].

Initial and final state radiation

We discussed up to now the development of parton shower arising from the emission from hadrons produced in the hard scattering, i.e. starting in general at an high annihilation scale Q^2 and progressively reaching the hadronization scale. This process goes under the name of *final state radiation* (FSR), as it is generated from outgoing partons of the hard interaction.

Parton shower can of course also happen before the hard interaction, and is called *initial state radiation* (ISR) as the incoming partons of the hard scattering originate the emission. In this case there is an important difference in the shower evolution, that is that the final energy of the showering is the hard interaction energy scale. To respect this fact, Monte Carlo simulation of ISR adopts a "backward evolution", first setting the correct parton momentum distributions for the hard scatter, and then developing the shower backward, with the intermediate partons aquiring energy at each emission. The Sudakov form factors are then slightly different from Equation 3.10, being rescaled by a factor that takes into account the PDFs of the parton at the two vertices.

Matrix element and parton shower matching

So far, we introduced two powerful ways to describe a QCD event, one relying on our capability to compute the Matrix Elements (ME) in pQCD at the LO and NLO, the other exploiting a procedure to develop Parton Showers (PS) including soft and collinear emissions. ME can also introduce soft and collinear emissions, but the ME weight is in this case divergent, while in the PS scheme divergences are eliminated through the Sudakov form factors. We need therefore to set the rules to conveniently split the phase space of the event into two regions, one characterized by hard and large angle emission to be described by ME, the other of soft and collinear emission to be described by PS. This is achieved in the so-called "ME and PS matching", where some resolution parameters are introduced with the role of soft and collinear cut-offs for ME.

The baseline idea is to compute the weight of an event using ME and then develop PS giving as inputs the ME weight as well as the event kinematics and color flow. A complication arises in this simple approach as the same final state can be obtained in multiple ways if ME and PS generated partons are swapped, as graphically shown in Figure 3.8, where the same event has three possible configurations. This issue goes under the name of "double counting", and the matching scheme aim is to possibly avoid it or, at least, reduce its impact.

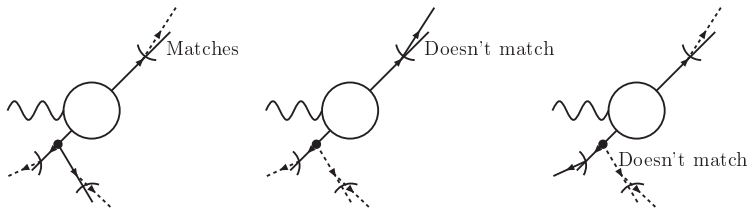


Figure 3.8: Example of double counting in hadron production. Dashed lines are PS emissions, solid lines are ME emissions. On the left, we have one hard large angle emission from ME and soft collinear emissions from PS. In the middle and on the right, one soft collinear emission from ME and both hard and soft PS emissions [81].

The main requirements on the matching scheme besides avoiding double counting are to perform a smooth transition from the ME and PS regions and to make sure the appropriate Sudakov form factors reabsorb the divergencies eventually introduced by the ME by reweighting the ME weight.

There are two main matching schemes: the Catani-Krauss-Kuhn-Webber (CKKW [82]) and the Michelangelo L. Mangano (MLM [83]) methods. They separate the phase space into the ME and PS regions by introducing resolution parameters that distinguish between *resolved* and *non-resolved* jets, to be described by means of the ME and PS respectively.

In the CKKW scheme, the parton-parton separation is measured by defining the distances between two final state partons and the distance between the parton and the beam using the $k_\perp$ jet algorithm [84]. For a parton to be resolved, and described through ME, both variables have to be greater than a resolution parameter Y_{sep}. A branching tree is developed clustering together the two closest partons and ME elements are reweighted first to the strong coupling α_S value at the ME scale and then using a combination of Sudakov form factors. Then, PS is evolved and emissions at scales greater than Y_{sep} are vetoed, thus avoiding overlap between configurations.

In the MLM scheme, partons are first clustered into samples with different multiplicity and then, like in the CKKW method, the $k_\perp$ jet algorithm is used to develop a branching tree and ME are reweighted to the proper strong coupling α_S value. At this point PS is performed and a jet finding algorithm matches partons into clusters. Jets with $p_T > p_{T_{\text{min}}}$, separation $R > R_{\text{min}}$ and pseudorapidity $\eta < \eta_{\text{max}}$ are considered for matching partons. Events are kept only if there is a one-to-one correspondence of partons to jets, else is rejected and thus double counting is avoided. As final step, the ME is reweighted with a Sudakov form factor.

The performance of the two matching methods has been extensively tested with data, and in particular it is important that the independence of the final result from the resolution parameters is verified.

3.2.3 Hadronization

When partons reach the hadronization scale energy $Q \sim 1$ GeV after showering, they recombine in bound colorless states according to the confinement principle, holding at low momentum. The so-called *parton-hadron duality* assumes that no high momentum transfer is needed in the recombination, as it happens close in phase space. This is a property of QCD experimentally observed, but there are no theoretical arguments explaining the hadronization. The solution is then to rely on phenomenological models.

The principle at the basis of hadronization models used in Monte Carlo is the *large* N_c *limit*, or *planar limit*, where N_c is the number of colors and we consider the QCD value $N_c = 3$ as just the dominant contribution. Simple rules are then defined: color and anticolor indices go from 1 to N_c; quark and antiquark lines are oriented and assigned a color and anticolor index respectively; gluon lines are oriented and assigned a pair of color-anticolor indices to achieve color neutrality. The color structure of the three splitting processes according to these rules are shown in Figure 3.9(a). While the assignment of color connections in the case of $q \to qg$ and $g \to q\bar{q}$ splittings is univocal, for $g \to gg$ there are two possible configurations (where by inverting the two final state gluons transform one into the other) and when the

color connections are reconstructed they are chosen with a 50% probability each. The final picture looks e.g. like the graph in Figure 3.9(b), where the important information is not the specific color assigned to the final state parton, but rather the *color flow*.

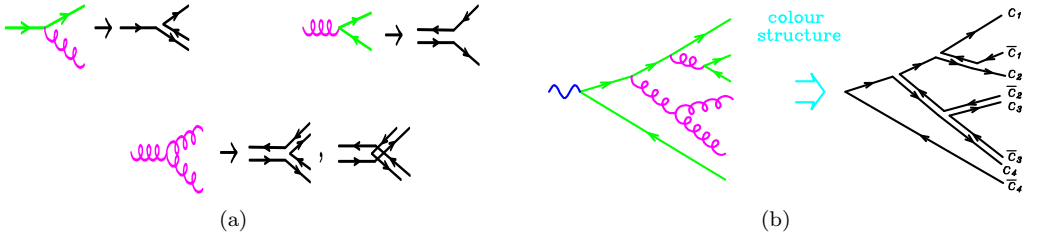


Figure 3.9: Left: rules to assign color connections in the splittings $q \to qg$ (top left), $g \to q\bar{q}$ (top right) and $g \to gg$ (bottom). Right: example of a color connected shower [81].

At this point, two possible phenomenological fragmentation models come in the game to bound partons into hadrons starting from the color connected final state. They both have been tuned using collider measurements over the last decades to properly describe the final state hadron multiplicity and need in general a large number of parameters.

The first hadronization scheme is called *Lund string model* and ties a quark with an antiquark plus a number of intermediate gluons, like stretching a string (or a "color flux tube") from the quark to its color-connected antiquark taking in the gluons that lie inbetween them. This can be see as an illustration in Figure 3.10(a) and Figure 3.10(b): the first string collects the quark with color c_1, the antiquark with color $\bar{c}_4$ and all the final state gluons along the path. uses string dynamics to describe the color flux between quarks. In other words, the string between the quark and antiquark produces a linear confinement potential.

The other hadronization scheme is the *cluster model*, where final state gluons are forced to split into a quark-antiquark pair and then color connected quark-antiquark pairs are bounded together. Figure 3.10(c) presents a graphical illustration of the concept. Because of *preconfinement* (the fact that color connected pairs with large invariant masses are Sudakov suppressed in angular ordered showers) the cluster can be associated with an hadronic two-body system.

During the last decades, different measurements at colliders have been used to tune these models to properly describe the hadron multiplicity in the final state.

3.2.4 Underlying event

With "underlying event" (UE) we refer to the secondary parton interactions at low transfered momentum that accompany the main hard process. The underlying event is flavor- and color-connected to the hard scattering and in real data is in general not separable from the event of interest. It is typically observed as jets of particles close to the direction of the beam and cannot be modeled with perturbative QCD but is instead studied from experimental data on *minimum bias* events at low momentum.

Besides a backward shower modeling starting from the beam remnants, an important contribution to the underlying event are multiple parton interactions, i.e. secondary relatively hard collisions between the incoming hadron remnants. Phenomenological models are

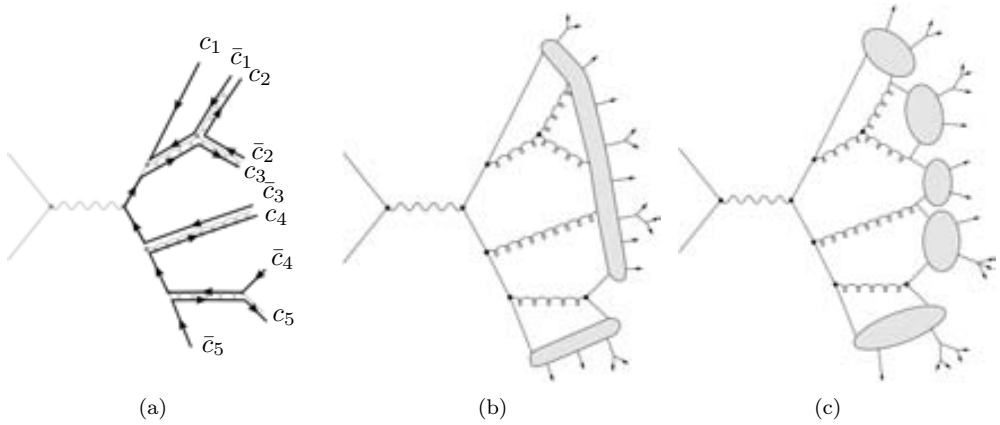


Figure 3.10: Drawing of a color connected parton shower graph (a) completed with hadronization from the Lund string model (b) and the cluster model (c) [79].

used to simulate the underlying event, and additional semi-hard interactions are generated with different primary vertices to account for the effect of pile-up.

3.3 Generators

Monte Carlo generators can either be *multi-purpose* generators, i.e. capable of performing the full simulation chain described in this chapter, or *specialized* generators, i.e. devoted to a single aspect of the simulation. As should be clear after the discussion about the modeling of hadronic collisions, there are some choices to be done that might result being better suited for some particular events or others. Therefore the Monte Carlo generators should be chosen according to their performance in modeling the event of interest.

We briefly present in the following sections the main characteristics of the generators that are used in the analyses that are the object of this dissertation.

PYTHIA

`PYTHIA` [85] is a multi-purpose Monte Carlo generator using ME computed at LO for $2 \to n$ ($n \leq 3$) events and PS with emissions ordered in p_T instead of angle. The Lund string model is used for hadronization and UE simulation is included. Minimum bias events with $p_T > \hat{p_{\mathrm{T}}}_{min}$ are used to model interactions between proton remnants.

HERWIG

`HERWIG` [86] is a multi-purpose Monte Carlo generator using ME computed at LO for $2 \to 2$ events and PS with emissions ordered in angle. The cluster model is used for hadronization and for the UE description, `HERWIG` is typically interfaced with the external package `JIMMY` [87] which simulates UE as scattering between proton remnants in $2 \to 2$ ME computed at LO.

ALPGEN

ALPGEN [88] is a Monte Carlo generator specialized for ME computation of $2 \to n$ ($n \leq 9$) events at LO, with cross sections computed using the ALPHA algorithm [89]. It is interfaced either with PYTHIA or HERWIG for PS development and ME/PS matching is done in the MLM scheme, where the resolution parameter $p_{T_{\min}}$ is called *jet matching scale*. Hadronization and UE are simulated through HERWIG and hence JIMMY.

The various parton multiplicity samples are then normalized to their LO cross section and combined into an inclusive sample, which is finally typically normalized to an inclusive cross section calculated at higher order in pQCD. For the inclusive production of Z and W bosons in pp collisions, the MCFM [90] and FEWZ [91] programs are used to predict cross sections at NLO and NNLO respectively.

MC@NLO

MC@NLO [92] is a Monte Carlo generator simulating ME at NLO, where the use of 1-loop corrections introduce the possibility of having negative weighted events. Theoretical uncertainties on the inclusive cross section is reduced thanks to the use of full NLO corrections, but higher multiplicity parton emission is simulated through PS in HERWIG which has a poorer description of hard emissions. Hadronization and UE are also simulated through HERWIG (and hence JIMMY).

SHERPA

SHERPA [93] is a multi-purpose Monte Carlo generator interfaced with PYTHIA for the parton shower. ME/PS matching is performed with the CKKW scheme. A modular design for SHERPA allows for simple implementation of other techniques, e.g. NLO corrections can be introduced in the CKKW matching scheme using a NLO Monte Carlo generator with the MENLOPS procedure [94]. Hadronization is done within PYTHIA and a multiple parton scattering model for UE simulation.

POWHEG

POWHEG [95] is a Monte Carlo generator computing ME at NLO and typically interfaced either with PYTHIA or HERWIG for the modeling of PS, hadronization and UE.

MADGRAPH

MADGRAPH [96] is a Monte Carlo generator specialized for ME computation of $2 \to n$ ($n \leq 6$) events at LO interfaced with PYTHIA for the modeling of PS, hadronization and UE.

ACERMC

ACERMC [97] is a Monte Carlo generator computing ME at LO and typically interfaced either with PYTHIA or HERWIG for the modeling of PS, hadronization and UE.

3.4 ATLAS detector simulation

Events generated with Monte Carlo simulation can be directly used at *parton level*, i.e. without any further operation on the generated partons, or be reconstructed either at *truth level*, where the particles go through object reconstruction (see Chapter 4) without interacting in the detector material and the original kinematics is maintened, or go through detector simulation [98] and after that through object reconstruction, to obtain the so-called *reconstructed level*.

The detector material, geometry and response are modeled using the `GEANT4` [99] package. The interaction of particles with ATLAS subsystems is converted into detector signals of the same sort of the real read-out and at this point the same kind of reconstruction algorithms shape the detector output into physical objects. During test-beam periods the `GEANT4` parameters have been tuned to best simulate the ATLAS configuration and the performance of detector simulation has been extensively checked also with data calibrations.

The detector simulation relies on the usage of two databases: the *geometry database* contains the information on the detector volumes like dimensions, geometry, position and material composition, while the *conditions database* is constantly updated with the information on the real detector real-time conditions as dead channels, misalignments, temperature. Since conditions vary from run to run, it is important that the detector simulation reproduces as close as possible the real status of ATLAS during a particular data period. Also for this reason, Monte Carlo samples are regularly reproduced consistently with *data reprocessings* or *data releases*. For the analyses presented in this dissertation, the Monte Carlo production tagged as `mc12` is used, performed within release 17 of the ATLAS analysis framework `ATHENA` [**?**].

3.5 Monte Carlo samples corrections

At the end of the full Monte Carlo simulation chain, after the detector simulation and event reconstruction steps, the generated samples need some corrections to better model data. The main event reweighting is to correct Monte Carlo samples to the correct theoretical cross section of the process and to the number of expected data events, which comes from the luminosity measurements. As typical in Monte Carlo techniques, an higher number of randomly produced events assures a better modeling of the system under study, and therefore usually a very high number of events are produced for each sample to maximize the confidence that all the relevant configurations have been copiously simulated. The event weight w to be applied (event by event) is defined as:

$$w = \frac{\sigma \times k}{N} \int \mathcal{L} dt, \tag{3.11}$$

where σ is the process theoretical cross section, N is the number of Monte Carlo events, $\int \mathcal{L} dt$ the integrated luminosity and k is the so-called $k-$factor, which is a correction to higher orders as, e.g., to bring the accuracy of a LO cross section computation to the NLO.

Furthermore, a weight to account for pile-up effect is to be applied, to match the expected number of interactions per bunch crossing $< \mu >$ in real data-taking conditions.

Event reconstruction

After having described the ATLAS detector in Chapter 2 and the procedure for Monte Carlo simulation of events in Chapter 3, we understand that what we deal with when we talk about "data" is raw digital signals from the detector, either the real one or the simulated one.

In the following Chapter we will explain how, starting from these outputs, objects are reconstructed to be used in physics analyses [100]. This process is what is called "offline event reconstruction" since it is not done in real time, due to the time required by the algorithms to perform their tasks.

In general we could describe the full procedure as subdivided into three main steps: a pre-reconstruction stage where the electronic signals are translated into measurements; a pattern-recognition step where the measurements are assembled into the building blocks of particles, e.g. tracks and energy clusters; a particle identification final leg where the full detector information elaborated is combined to match a candidate physics object (electrons, muons, jets and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$).

The expected signatures for the various particles in terms of interaction with the detector system are schematically shown in Figure 4.1.

4.1 ID Tracks

Particle trajectories ("tracks") are used both to reconstruct the particle itself, giving the momentum measurement, and to identify the interaction vertices. The parameters describing a track are: q/p, the charge divided by the momentum; θ, or more used η, the angle with respect to the Z axis in the RZ plane measured from the perigee[1]; ϕ_0, the angle with respect to the X axis in the XY plane measured from the perigee; d_0, the impact parameter, or perigee with respect to the Z axis in the XY plane; z_0, Z component of the perigee. These parameters are shown in the double-view drawing of Figure 4.2.

In order to reconstruct the track, the first step is to retrieve the information from the ID hits, which are converted into three-dimensional space points. Then, the *inside-out*

[1]The perigee is the point of the track closest to the origin.

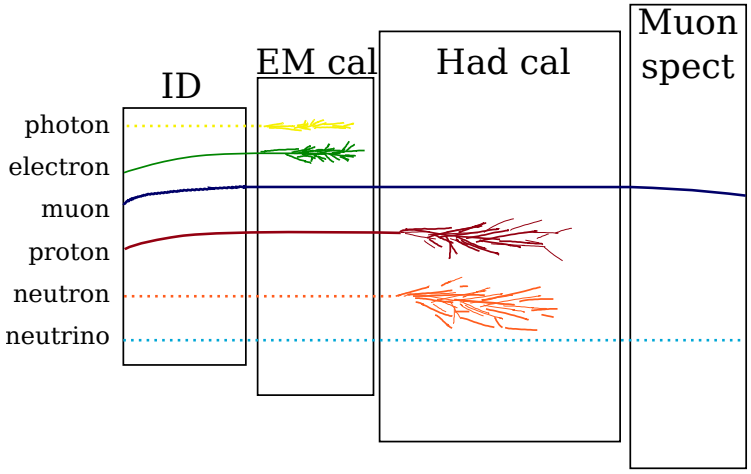


Figure 4.1: Drawing illustrating how particles are detected in the ATLAS sub-systems.

algorithm [101] is used, starting from a seed of three aligned hits in the pixel detector or in the SCT. From there, a path is formed along the seed directional information adding space points one by one. This is done by using a Kalman filter algorithm [102] which checks progressively the compatibility between the track (also progressively updated) and the new point. The five track parameters described before are also computed at this step. A cleaning procedure then rejects incomplete tracks or tracks sharing hits with others, or composed by false space points. The candidate tracks are extended into the TRT and re-fitted taking into account the effects from the interaction of the charged particle with the detector material.

A second algorithm, called *outside-in*, is applied in order to better reconstruct tracks from secondary charged particles. This algorithm does the opposite of the inside-out one, taking as seeds hits in the TRT (the ones not associated to any track candidate in by the inside-out reconstruction) and extrapolating back to the SCT and pixel detector.

4.2 Primary vertices

In general, a primary vertex (PV) is identified by the tracks associated to it. The reconstruction is performed via an iterative procedure [103] starting from a seed defined as the maximum in the distribution of the z_0 parameter of reconstructed tracks. After tracks are assigned to the PV with the aid of an iterative χ^2 fit, the ones that fall out of more than 7σ from the PV are used to seed another PV until no track is left without being assigned to a vertex (one track can be associated to more than one vertex).

A PV must have at least two associated tracks and its position must be consistent with the beam collision region in the XY plane. The hard-scatter PV is chosen as the one with the highest sum of squared transverse moments of the tracks. The other reconstructed PVs are identified with pile-up interactions. Another kind of vertices, not compatible with the requirement of coming from close to the proton collision spot, are the secondary vertices, originating from the decay of short-lived particles. These vertices are useful to identify B-hadrons and will be described in Section 4.6.2.

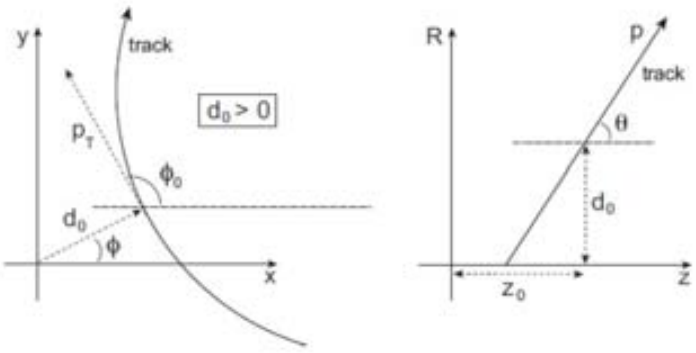


Figure 4.2: Schematic drawings of the parameters used for track reconstruction in the XY and RZ planes (left and right respectively) where the origin is the beam spot, i.e. where the protons collide and interact.

As can be expected, high pile-up environments deteriorate the performance of vertex reconstruction, as more fake tracks are introduced and nearby interaction might lead to the misreconstruction of distinct vertices as a single one [104].

4.3 Energy clusters

With the name "energy cluster" we generically refer to energy deposits in the calorimeter cells that are grouped together on the basis of some criteria [105]. In particular, we are interested in *topological clusters* and *electromagnetic towers*, used respectively for jets and electron/photon reconstruction.

Topological clusters, abbreviated as "topoclusters", are built from neighboring calorimeter cells starting from a seed deposit with a signal (S, the cell measured energy) to noise (N, the RMS of the cell noise distribution) ratio higher than a certain threshold. Cells with $S/N \geq 4$ are taken as seeds, and starting from the one with the highest S/N all the neighboring cells with $S/N \geq 2$ are added to the topocluster. Topoclusters are treated as massless and their energy at the electromagnatic scale is the sum of the constituent cells. Their position and direction parameters are obtained from a weighted sum of the constituent cells' pseudorapidity and azimuth angle based on the absolute value cell's energy. Since energy measurement can be negative (due to noise fluctuations), clusters with negative energies are rejected.

Towers are built using the *sliding window* algorithm [106] starting from single energy deposits in the EM calorimeter middle layer of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$. As schematically shown in Figure 4.3, a window of 3×5 cell units is defined, centered on the maximum of energy and finally expanded to optimize the cluster reconstruction, with a size that depends on the object (electron or photon) and the position in the detector (3×7 in regions with $|\eta| < 1.4$ and 5×5 elsewhere).

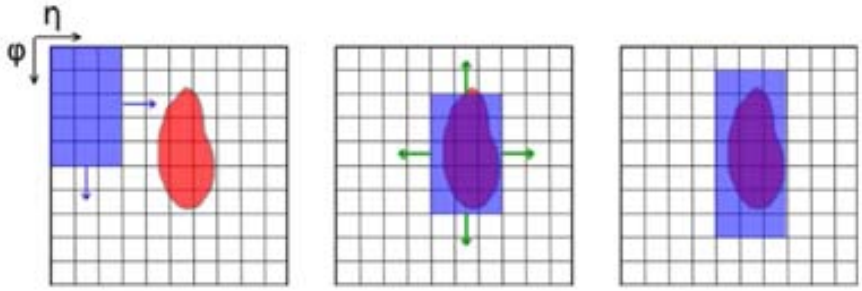


Figure 4.3: The three steps of the sliding window algorithm.

4.4 Electrons

Electrons [106] are reconstructed for pseudorapidities up to $|\eta| = 2.47$, where information from the ID is available, matching a track (see Section 4.1) with a cluster in the electromagnetic calorimeter reconstructed with the sliding window algorithm (see Section 4.3). In order to account for bremsstrahlung losses the matching is done within a region of dimension $\Delta\eta \times \Delta\phi = 0.05 \times 0.10$ and if more candidate tracks are matched, of all the ones with hits in the silicon detectors the track with the smallest ΔR with respect to the energy cluster is chosen. In addition, the track momentum has to be compatible with the cluster energy, which is calibrated to the electromagnetic scale derived from Monte Carlo based corrections (to account for dead material losses), test-beam studies and calibration from $Z \to ee$ events [107].

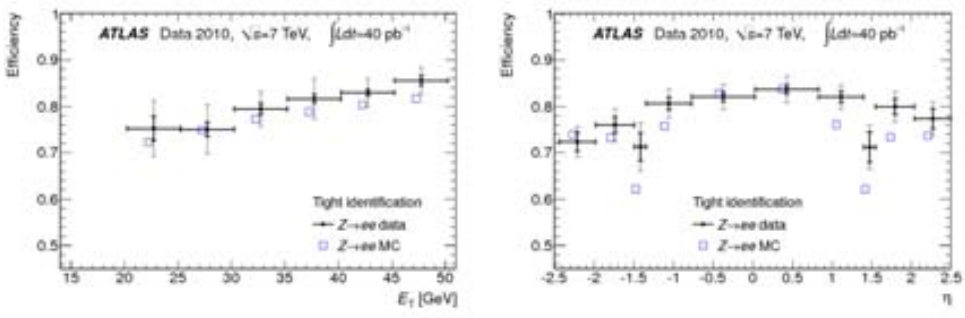
In general, electron can be distinguished from hadrons thanks to various characteristics of their shower development: electrons deposit the most of their energy in the second layer of the EM calorimeter; the width of their shower is narrower; they have smaller hadronic leakage[2]; the E/p variable (ration of cluster energy and track momentum) is higher.

Some difficulties arise when dealing with π^0 and η particles, which decay into two γs that produce two close showers reconstructed as a single one in the second layer of the EM calorimeter, and in general with jets faking electrons from, e.g. QCD processes. There are then six different electron definitions to help separate real electrons from fake ones, described in the following ordered from the looser requirements to the tightest. Performance studies on electron reconstruction and identification where done using 2010 data and Monte Carlo $Z \to ee$ and $W \to e\nu$ events [106] (see Figure 4.4).

`Loose` electrons lie in the pseudorapidity region $|\eta| < 2.47$ and have low hadronic leakage and requirements on the variables defining the shower shape. The identification efficiency is high but the jet rejection is low (about 500).

`Loose++` electrons are `loose` electrons whose track has at least one hit in the pixel detector and at least 7 hits in the combined silicon detectors and the $|\eta_{\text{firstEM}}|$ distance between the track estrapolated to the first EM layer and the matched cluster is lower than 0.015. The identification efficiency is similar to the loose one but the rejection is ten times

[2]The hadronic leakage is the ratio of the transverse energy reconstructed in the first layer of the hadronic calorimeter to the total transverse energy reconstructed in the EM calorimeter.



(a) (b)

Figure 4.4: `Tight` electron identification efficiencies measured from $Z \to ee$ events and predicted by MC as a function (left) of E_{T} (integrated over $|\eta| < 2.47$ excluding the transition region $1.37 < |\eta| < 1.52$ and (right) of η and integrated over $20 < E_{\mathrm{T}} < 50$ GeV. [106]

higher.

`Medium` electrons are `loose++` electrons where additional requirements on shower shape are made as well as on their tracks: $|d_0|$ <5 mm and $|\eta_{\mathrm{firstEM}}| < 0.01$. The efficiency drops to 88% and the rejection is higher than the previous.

`Medium++` electrons are `medium` electrons whose track has at least one hit in the first pixel detector layer, a requirement that allows to reject electrons from photon conversion. Charged hadrons contamination is reduced by discarding candidates whose track has a low fraction of high-threshold TRT hits. In addition, $|\eta_{\mathrm{firstEM}}| < 0.005$ and stricter cuts are applied to shower shaper of clusters in $|\eta| < 2.01$. The efficiency is about 85% and rejection is about $50{\times}10^3$.

`Tight` electrons are `medium++` electrons with additional requirements on the distance between the track and the matched cluster ($|\Delta\phi| < 0.02$, $|\Delta\eta| < 0.005$) and on the E/p variable. Stricter cuts are imposed on the fraction of high-threshold TRT hits and on the impact parameter ($|d_0|$ <1 mm). The efficiency drops to 75% and the rejection is higher than the previous one.

`Tight++` electrons are `tight` electrons with asymmetric $\Delta\phi$ cuts, which give both better efficiency and rejection.

4.4.1 Additional requirement and corrections for analyses

For our analyses [108], electrons in the transition region $1.37 < |\eta_{\mathrm{cluster}}| < 1.52$ with inactive material are excluded. Electrons are required to satisfy `tight++` criteria and to have $E_{\mathrm{T}} = E_{\mathrm{cluster}}/\cosh\eta_{\mathrm{track}} > 25$ GeV and $z_0 < 2$ mm. In addition, to suppress further the QCD multijet background, isolation cuts are imposed both as calorimeter (using the energy in a cone of size $\Delta R < 0.2$, `EtCone20`) and track isolation (using the scalar sum of p_Ts from tracks within a cone of $\Delta R < 0.3$, `PtCone30`). The `EtCone20` and `PtCone30` isolation cuts are chosen to give 90% efficiency. In addition, jets (see Section 4.6) within $\Delta R = 0.2$ of the selected electron are discarded, and if an additional jet with $p_T > 25$ GeV and $|JVF| > 0.5$ is found within $\Delta R = 0.4$, then the electron is rejected. The electron is

matched to the single electron trigger `EF_e24vhi_medium1` combined with a logical `OR` to the `EF_e60_medium1` trigger, which recovers some efficiency loss at $E_T > 80$ GeV.

The efficiency in selecting electrons can be factorized as:

$$\varepsilon = \varepsilon_{\text{reco}} \cdot \varepsilon_{\text{tight++}} \cdot \varepsilon_{\text{isolation}} \cdot \varepsilon_{\text{trigger}} \tag{4.1}$$

where the various components represent respectively: the efficiency in recostructing the electron in terms of track-cluster match, track quality and hadronic leakage; the efficiency for the `tight++` identification criteria; the efficiency for the isolation cuts; the efficiency from trigger selection. Scale factors are derived in bins of (η, E_T), and the trigger scale factors are separated into four data-taking periods (`A-B3`, `B4-D3` without `C1-C5`, `C1-C5` and `D4+`). The efficiency scale factors are applied as weights to Monte Carlo events.

The electron energies in data are corrected using scale factors $\alpha(\eta)$ derived from data-to-simulation comparison in $Z \to ee$ events in order to match the Z boson mass peak.

4.5 Muons

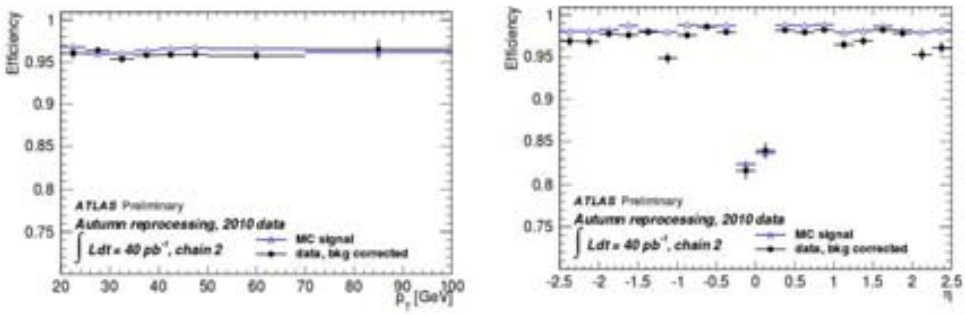
As suggested in Figure 4.1, muons interact with all of ATLAS sub-detectors, even though they act as minimum ionizing particles (mip) for the calorimeters and hence will deposit only a very small fraction of their energy in the material. Their track instead is precisely measured both in the ID and in the muon spectrometer (MS). Based on how we decide to combine the various information, we can list the following types of reconstructed muons: `standalone` muons take the MS track and extrapolate it back to the interaction point; `combined` muons match the MS track with the tracks from the ID; `segment tagged` muons extrapolate ID tracks to the spectrometer and match the result with MS segments; `calorimeter tagged` muons extrapolate ID tracks to the calorimeters and match the result with energy deposits.

We will only consider `combined` muons, recontructed using an algorithm called `Muid` [109] and whose pseudorapidity is limited to $|\eta| < 2.5$ by the ID acceptance. Starting from $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ regions where interesting activity has been triggered, track segments are searched for in the RPC and TGC and combined into a single track by means of a least-square fitting method. These track candidates are hence extrapolated back to the interaction point and their momentum corrected for the mip energy loss in the calorimeter material.

At this points a χ^2test (checking the difference between the extrapolated track coordinates weighted with combined covariance matrix) on the matching of the candidate MS track and the tracks reconstructed in the ID is performed to obtain the final muon candidate track. Only ID tracks that satisfy some quality requirements are considered for the matching: they need to have at least two pixel hits, of which at least one in the first layer; at least two pixel hits plus number of crossed dead pixel sensors; at least six SCT hits plus number of crossed dead SCT sensors; maximum two pixel or SCT holes[3]; defining the number of TRT outliers[4] and the number of TRT hits as N_{TRT_o} and N_{TRT_h} respectively, $N_{\text{TRT}_h} > 5$ and $N_{\text{TRT}_o}/N_{\text{TRT}_h} < 0.9$ for $|\eta| < 1.9$, $N_{\text{TRT}_o}/N_{\text{TRT}_h} < 0.9$ if $N_{\text{TRT}_h} > 5$ for $|\eta| \geq 1.9$. In

[3] A "hole" in the silicon detectors is a region where the module did not perform as expected even though the surrounding ones did.

[4] "Outlier" is an hit that is deviated from the track path.



(a) (b)

Figure 4.5: `Combined` muon reconstruction efficiencies using the `Muid` algorithm measured from $Z \to \mu\mu$ events and predicted by MC as a function (left) of p_{T} and (right) of η. [109]

case no matching is found, no muons are reconstructed, while if more candidates arise, the one giving the best $\chi^2$is chosen. The momentum is computed as a weighted average of ID and MS measurements.

Performance studies on muon reconstruction and identification where done using 2010 data and Monte Carlo $Z \to \mu\mu$ events [109] (see Figure 4.5).

4.5.1 Additional requirement and corrections for analyses

`Combined` muons are used in our analyses [108] with an additional cut on the longitudinal impact parameter $|z_0| < 2$ mm to ensure the track comes from the hard-scattering primary vertex. A requirement on the muon momentum of $p_T > 25$ is used to obtain 90% efficiency from the chosen single muon trigger, which is the logical `OR` combination of the triggers `EF_mu24i_tight` and `EF_mu36_tight`. The `EF_mu24i_tight` trigger includes an isolation requirement for which the p_{T} sum of the tracks in a cone of size $\Delta R = 0.2$ around the muon has to be less than the 12% of the muon transverse momentum. Muons overlapping with any jet (see Section 4.6) with $p_{\mathrm{T}} > 25$ GeV and $|JVF| > 0.5$ within a $\Delta R < 0.4$ cone are rejected.

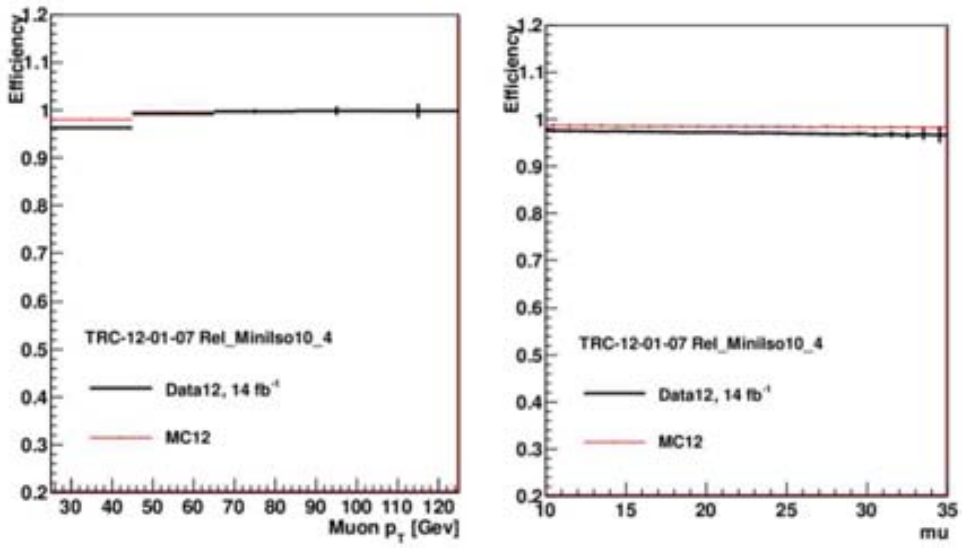
In addition to the previous isolation requirements, a "mini-isolation" is defined [108] to better deal with the high pile-up present in $\sqrt{s}$ =8 TeV collision events. The mini-isolation is defined as

$$I^l_{mini} = \sum_{tracks} p_{\mathrm{T}}^{track} / p^l_{\mathrm{T}} \tag{4.2}$$

where p^l_{T} is the lepton transverse momentum and the summation runs over all tracks found in a cone whose radius varies as a function of the muon momentum as:

$$\Delta R(l, track) = \frac{10 \text{ GeV}}{p^l_{\mathrm{T}}}. \tag{4.3}$$

The tracks also have to satisfy: $p_{\mathrm{T}}^{track} > 1$ GeV; $d_0 < 10$ mm; $z_0 \sin\theta_{track} < 10$ mm; at least four hits or dead sensors crossed in the silicon detectors. The cut on the mini-isolation variable is chosen as $I^l_{mini} < 0.05$ The performance of the mini-isolation is shown



(a) (b)

Figure 4.6: Efficiency of the mini-isolation as a function of the muon momentum (left) and of the average number of bunch crossings $< \mu >$ (right) [108].

in Figure 4.6.

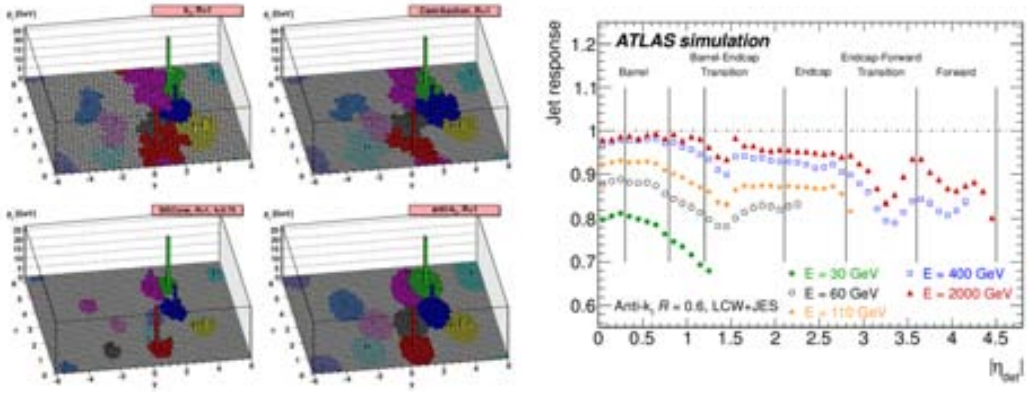
As is done for electrons, a set of corrections are applied to correct for minor discrepancies between Monte Carlo simulation and data events. The scale factors to compensate reconstruction, isolation and trigger i nefficiencies are derived from tag-and-probe measurements and applied to Monte Carlo events. In addition, the muon momentum in simulated events is smeared to obtain agreement between the momentum resolutions in Monte Carlo and data.

4.6 Jets

With the name "jet" we generically refer to the object formed as a consequence of parton hadronization from a spray (or *jet*) of particles. These particles will leave signals both as tracks in the ID and as energy deposits in the calorimeters and two type of jets can then be defined using either the former or the latter information: track jets and calorimeter jets. In the following, we will only deal with calorimeter jets.

In order to interpret the detector information, first topoclusters are formed from the calorimeter cells signals, as explained in Section 4.3. Then, different algorithms were developed to associate topoclusters into a jet. Because of the need for a stable and precise performance over the QCD processes from pp collisions, a set of requirements has been defined for the algorithms to be valid [110].

First of all, the splitting of one particle into two collinear particles must not change the result of the algorithm reconstruction, as well as the presence of additional soft emission. The importance of Infrared and Collinear (IRC) safety is evident considering e.g. that a hard parton, as part of the fragmentation process, will undergo many collinear splittings, and that QCD events always include emission of some soft particles, perturbatively or not. In addition, we want the algorithm result to be invariant under Lorentz boost along the



(a) (b)

Figure 4.7: Left: The same input produces different results in terms of jet reconstruction using various jet algorithms [110]. Right: Average jet energy response from simualted events at the LCW scale for various calibrated energies (E) as a function of pseudo-rapidity. The inverse of the response shown in each bin is equal to the average jet energy scale correction [114].

beam direction, to be as insensitive as possible to detector effects like noise or resolution, and to be light in terms of computing resourse usage.

A set of jet algorithm that satisfy these requirements are the sequential recombination algorithms [111, 112, 113], which combine topoclusters into jets using as criteria a distance parameter defined as:

$$d_{ij} = min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \tag{4.4}$$

$$d_i = p_{T_i}^{2p}, \tag{4.5}$$

where p_{Ti} is the transverse momentum of topocluster i, $\Delta R_{ij}{=}\sqrt{\Delta\eta^2 + \Delta\phi^2}$ the distance between constituents i and j, R a parameter of the algorithm that approximately controls the size of the jet, p the parameter that defines the type of algorithm as:

$$\begin{array}{lcl} p = 1 & : & k_t \text{ algorithm;} \\ p = 0 & : & \text{Cambridge/Aachenalgorithm;} \\ p = -1 & : & \text{anti}{-}k_t \text{ algorithm.} \end{array} \tag{4.6}$$

The algorithms compute d_{ij}, the distance between the two topocluster inputs i and j, and d_i, the distance between the input i and the beam axis in the momentum space. By computing the minimum of the two distances the choice made is to combine i and j into a new input if $d_{ij} < d_i$, or take i as a jet candidate and remove it from the input list if $d_i < d_{ij}$. The cluster combination is done by summing the four-momentum of each input. The distances are recalculated with the updated list of input objects and the process repeated until no further cluster is left.

The anti-k_t algorithm is chosen by most of analyses in ATLAS as it is particularly performant against pile-up, since it starts summing up constituents with higher momentum, and produces jets with a conical structure (see Figure 4.7(a)).

4.6.1 Additional requirement and corrections for analyses

For our analyses [108] jets are reconstructed using the anti-k_t algorithm with a radius parameter $R = 0.4$ (from which the algorithm is often referred to as anti-k_t4) using calorimeter energy deposits corrected for effects of non-compensation[5], dead detector material and out-of-cluster leakage. Other effects affecting jet energy are low momentum particles that are deflected by the magnetic field and energy losses in topocluster formation and jet reconstruction [114].

The initial energy is reconstructed at the electromagnetic (EM) scale as the calorimeter signals arise from electromagnetic interaction of particles with matter. The energy calibration to EM scale was derived during test-beam runs using electron beams, validated with muons from both test-beam and cosmic-rays runs and corrected using simulated $Z \to ee$ events.

Of the several energy calibration schemes derived in ATLAS, we will be using the Local Cluster Weighting (LCW) calibration [115, 116]. This scheme exploits properties of the topoclusters shapes to classify the clusters as "mainly electromagnatic" or "mainly hadronic" and then derives the calibration from Monte Carlo simulation of charged and neutral pion events. The calibration corrections are applied before the jet reconstruction algorithm is operated, and after the jet is formed a final correction is applied to ensure linearity in response.

In our analyses we consider only jets with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. Furthermore, a variable called "jet vertex fraction" (JVF) is defined as the fraction of the sum of p_{T} of tracks with $p_{\mathrm{T}} > 1$ GeV associated with the jet that comes from tracks originating from the primary vertex. By requiring JVF> 0.5 we avoid selecting jets from in-time pile-up events.

To avoid double counting of energy deposits from electrons as jets, if jets are found within ΔR of 0.2 of the selected electron, the jet closest to the lepton is removed and then electrons that lie within $\Delta R < 0.4$ of the remaining jets are discarded.

4.6.2 b-tagging

When a bottom quark is produced in an events, it hadronizes into a B hadron, which has a lifetime of the order of 10^{-12} s and hence can travel about 3 mm before decaying. The result is a displaced secondary vertex that, if correctly reconstructed, can allow for the identification of the bottom quark. Since this capacity relies on track reconstruction from the ID, its applicability is limited by the ID acceptance to the pseudorapidity region $|eta| < 2.5$.

This technique is called b-tagging [117] and is widely used in ATLAS analyses with top quarks. There are three types of algorithms, and they can be combined to obtain better performance. In general they define a weight corresponding to the probability for the jet to be tagged, and a working point is chosen as the threshold for this weight to discriminate between $b-$ and not-$b$jets by finding a good compromize between a good efficiency (the ratio between tagged b-jets and true b-jets) and a high light-jet rejection (the inverse of the number of light-jets misidentified as b-jets).

[5] The energy response to hadrons is lower than the response to electrons of the same energy due to the presence of invisible processes.

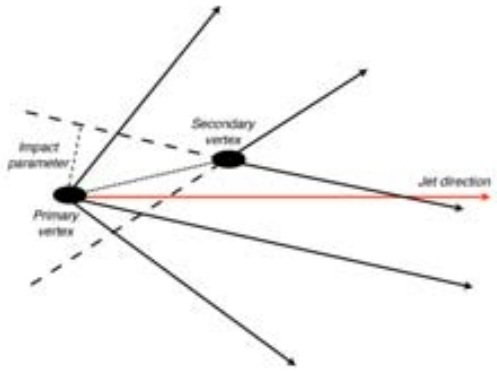


Figure 4.8: Simple schematic of the displaced secondary vertex.

Algorithms like `IP1D`, `IP2D` and `IP3D` are based on information from impact parameters of the tracks contained in the jet, z_0/σ_{z_0}, d_0/σ_{d_0} and a combination of the two respectively. A likelihood is computed to obtain the $b-$tag weights.

Other algorithms reconstruct the secondary vertex from the B hadron decay, allowing for a better discrimination between $b$jets and light jets. The `SV1` algorithm uses the number of track pairs in the secondary vertex, their total invariant mass and the ratio of the sum of the energies of tracks from the secondary vertex to the one of the total tracks of the jet to compute likelihood ratios the logarithm of which are then summed to obtain the $b-$tag weights.

Finally, the `JetFitter` algorithm uses the full decay chain reconstruction of b and c hadrons by fitting it with a Kalman filter to determine a common path between the primary vertex and the vertices from the b and c hadrons inside the jet. The likelihood is computed with the flight length significances of the vertices and the variables from the `SV1` algorithm.

The algorithm emploied in our analyses is called `MV1` and uses a neural network to combine information from the `JetFitter`, `IP3D` and `SV1` algorithms. The working point corresponding to 70% efficiency, ∼130 light-jet rejection and a charm-jet rejection of 5 is chosen (see Figure 4.9).

The tagging efficiencies in Monte Carlo are corrected for b and c flavours with the appropriate $\epsilon_{data}/\epsilon_{MC}$ scale factors, determined in bins of jet p_{T} and η.

Tag Rate Function method

When requiring ≥ 1 b-tagged jet the available Monte Carlo statistics is significantly reduced for some particular background processes, leading to large fluctuations in the resulting distributions. To overcome this problem the Tag Rate Function (TRF) method is introduced. Here, no event is rejected based on its b-tagging count, but instead all the events are kept and weighted according to the probability of the given event to contain the desired number of $b$jets. Appendix C describes the TRF method in more details.

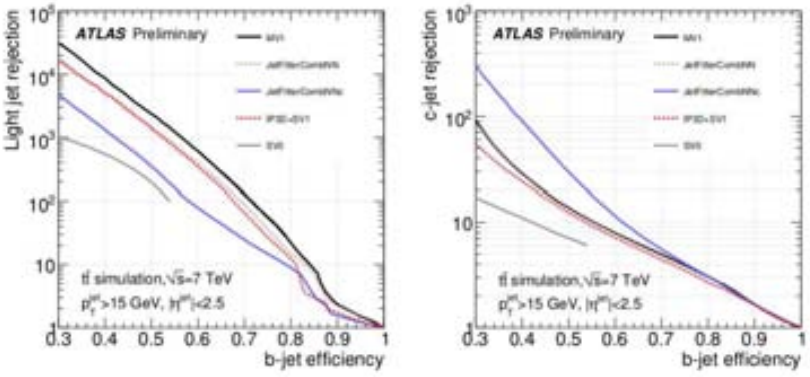


Figure 4.9: Light- (left) and $c-$jet (right) rejection as a function of the $b$jet tagging efficiency for different tagging algorithms. These values refer to jets with $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 2.5$ in simulated $t\bar{t}$ events [118].

4.7 Missing Transverse Energy

To estimate the momentum of invisible particles in the event, i.e. neutrinos and, eventually, new particles, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ [119] is defined [108] to balance the total transverse momentum of the event. Indeed, while the longitudinal energy of the interacting partons is unknown, as they carry an unpredictable fraction of the total proton momentum, its transverse component is, initially, zero. Possible sources of fake contributions to the $E_{\mathrm{T}}^{\mathrm{miss}}$ are detector coverage, dead or noisy regions and finite detector resolution.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is computed by first matching each calorimeter energy deposit is with a high-p_{T} object, in the following order: electrons, photons, jets and muons. These are respectively the RefEle, RefGamma, RefJet, RefMuon terms, wether the low-p_{T} jets are grouped into the SoftJet term. Then, the energies of these objects are corrected accordingly to the respective calibration constants. The calorimeter clusters that did not get associated with any high-p_{T} object are calibrated for energy losses in dead material regions and for the different response to the electromagnetic and hadronic components of particle showers and added as the CellOut term. Finally the $E_{\mathrm{T}}^{\mathrm{miss}}$ is computed as:

$$\begin{aligned} E_{x,y}^{\mathrm{miss}} &= E_{x,y}^{\mathrm{RefEle}} + E_{x,y}^{\mathrm{RefGamma}} + E_{x,y}^{\mathrm{RefJet}} + E_{x,y}^{\mathrm{RefMuon}} + E_{x,y}^{\mathrm{SoftJet}} + E_{x,y}^{\mathrm{CellOut}} \\ E_{T}^{\mathrm{miss}} &= \sqrt{(E_{x}^{\mathrm{miss}})^2 + (E_{y}^{\mathrm{miss}})^2} \end{aligned} \tag{4.7}$$

Part II

Searches for vector-like top partner pairs in the single lepton channel

Starting from this chapter, and continuing in Chapter 6 and Chapter 7, we are going to describe two searches for vector-like top partners $T\bar{T}$ pairs performed in the single lepton[1] channel. These analyses are optimized for different final states and are thus complementary. The analyses are performed using a partial dataset of the pp collisions at the center of mass energy of $\sqrt{s}$=8 TeV collected during 2012 at the ATLAS detector, corresponding to an integrated luminosity of 14.3 fb^{-1}. The first search focuses on vector-like top partners decay channels with high Branching Ratio (BR) to a W boson and a bottom quark, while the second search is optimized for events with high BR to a Higgs boson and a top quark.

This chapter is devoted to the presentation of the general features that are common to the two analyses and is organized as follows: first in Section 5.1 we review the strategy for vector-like quark searches adopted by the Exotics group of the ATLAS collaboration; Section 5.2 summarises the common event preselection for data and few general concepts in the analyses design; Section 5.3 describes the Monte Carlo samples used in the searches, which are in general common to both analyses with only few exceptions that are reported, and how the multi-jet background from QCD events is obtained; Section 5.5 introduces the general treatment of systematic uncertainties.

5.1 General strategy for vector-like quark pairs searches

The phenomenology for vector-like quarks was described already in Section 1.3 of this dissertation. Here we will only briefly re-introduce the concepts on which the strategy for the searches has been built. Table 5.1 collects the decay modes for vector-like quarks in the singlet and doublet models. It is evident from the richness of the final state phase space, combined with the unpredicted mass of the heavy objects that could span from few hundreds

[1] In the following, with the word "lepton" we will refer either to an electron or a muon, assumed to come from the leptonic decay of a W boson or a leptonic τ decay.

of GeVs (down to the values exluded by previous searches) up to ~ 1 TeV (since we focus on pair production of vector-like quarks, which is favoured up to this mass scale)that is impossible to cover it with a single inclusive search.

VLQ Singlets	Decay modes	VLQ Doublets	Decay modes
$T(+2/3)$	W^+b, Ht, Zt	$\begin{pmatrix} T \\ B \end{pmatrix}$	W^+b, Ht, Zt W^-t, Hb, Zb
$B(-1/3)$	W^-t, Hb, Zb		
$X(+5/3)$	W^+t	$\begin{pmatrix} T \\ X \end{pmatrix}$	Ht, Zt W^+t
$Y(-4/3)$	W^-b	$\begin{pmatrix} B \\ Y \end{pmatrix}$	Hb, Zb W^-b

Table 5.1: Allowed decay modes for vector-like singlets and doublets.

The BR of vector-like top and bottom partners to the allowed decay modes depends on the mass of the heavy quark and on the considered model (in our case, singlet or doublet scenario), as shown in Figure 5.1. Each decay mode has specific features that allow to define powerful, optimized searches. Therefore in order to exploit this opportunity and at the same time stay as model independent as possible, different searches for vector-like quarks are performed at ATLAS to be later combined, each of them sensitive to specific channels. To ensure a comprehensive coverage of the phase space, a two-dimensional plane is defined (Figure 5.2) as follows: along the Y axes is the BR of the decay modes with a Higgs boson in the final state; along the X axis is the BR of the decay modes with a W boson in the final state. The BR to the channel with a Z boson in the final state is then fixed by the unitarity requirement BR($T/B \to Zt/b$) = 1 - BR($T/B \to Ht/b$) - BR($T/B \to Wb/t$). A plane of this kind is defined for every vector-like quark mass point considered in the analysis. Each point of each plane therefore represents a uniquely defined model, and analyses are performed for every configuration to either find deviations from expectations or to set a 95% Confidence Level (CL) exclusion. The final objective of the joint strategy is to cover the full plane by combining analyses probing the different signatures.

Up to the date of the writing of this dissertation, four complementary and quasi model-independent analyses have been performed by the Exotics working group on the partial dataset of 14.3 fb^{-1} of proton-proton collision data at $\sqrt{s} = 8$ TeV. Two analyses investigated dilepton channels and two probed single lepton channels.

The search for vector-like bottom and top partners in the same-sign dilepton channel [120] investigates a channel with very small contamination from Standard Model backgrounds and is also sensitive to four-top production $pp \to t\bar{t}t\bar{t}$, either through the Standard Model process or a beyond-SM source such as pair production of scalar color-octets (sgluons) or gluinos, with subsequent decays to top quark pairs. The approach of this search is to select via restrictive cuts (that also impose a veto on a $Z \to l\bar{l}$ boson) the eventual signal and compare the final counts with the expected yields from background sources. Figure 5.3 shows how the decays of vector-like bottom and top partner pairs can contribute to the same-sign dilepton

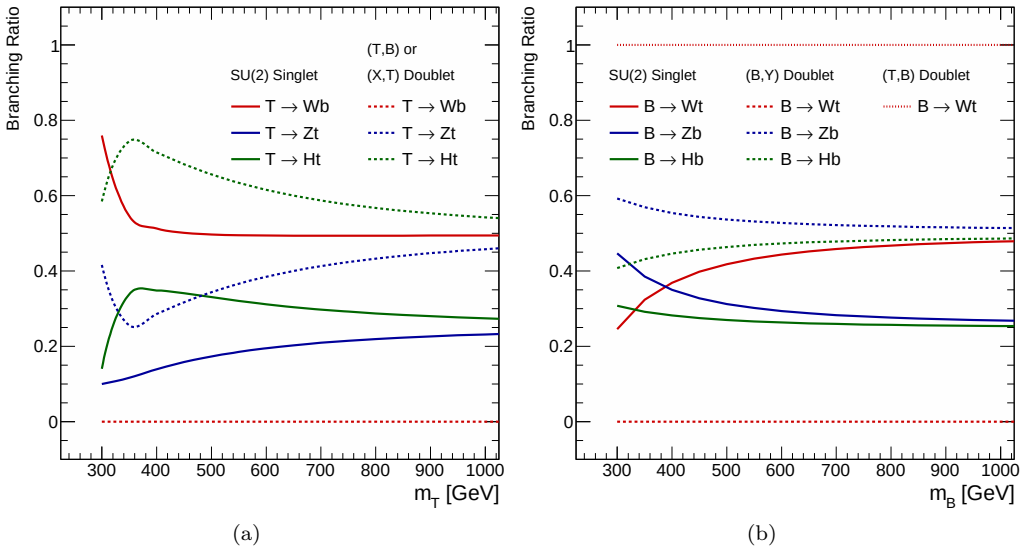


Figure 5.1: Branching ratio of vector-like top (a) and bottom (b) partners as a function of the heavy quark mass m_T and m_B respectively [64] for singlet and doublet models.

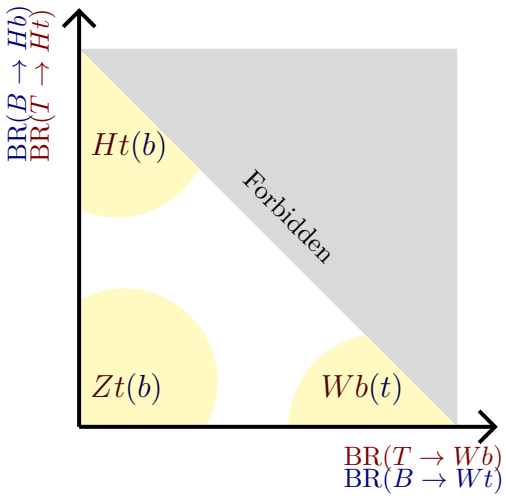


Figure 5.2: Two dimensional plane used to represend the comprehensive scan of model mixing. Searches with a Higgs boson in the final state cover the top left corner; searches with a Z boson in the final state cover the bottom left corner; searches with a W boson in the final state cover the bottom right corner. The shaded area labelled as "forbidden" is the unphysical region where $\mathrm{BR}(T/B \to Ht/b) + \mathrm{BR}(T/B \to Wb/t) + \mathrm{BR}(T/B \to Zt/b) > 1$

signature. From this it is easy to understand that this search will be mainly covering the bottom right corner of the $B\bar{B}$ two dimensional plane of Figure 5.2 and the top left corner of the same plane for $T\bar{T}$.

The search in the opposite-charge dilepton channel [64] focuses on vector-like bottom (top) partners decay channels where at least one heavy quark decays into a Z boson and

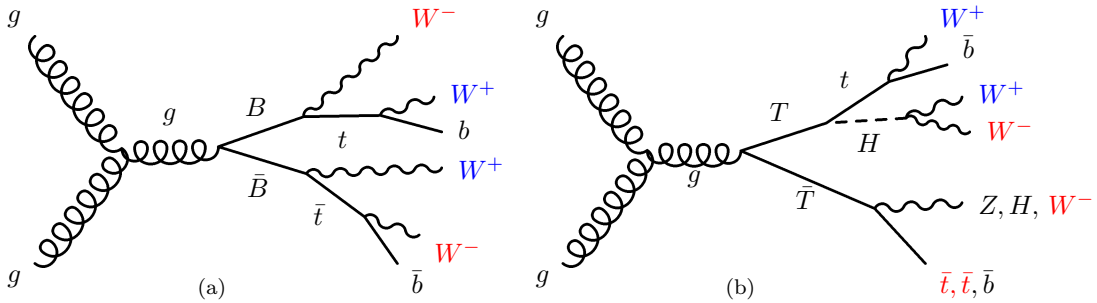


Figure 5.3: Feynman diagrams for $B\bar{B}$ (left) and $T\bar{T}$ (right) decays that can result in a final state with two same-sign leptons, in the case that the same-sign W bosons (highlighted in different colors; on the right the anti-top quarks associated to the Z or Higgs bosons are also highlighted as they will originate a W^- boson) decay into a lepton and a neutrino.

a bottom (top) quark. Here the strategy is to reconstruct the Z boson from the opposite-charge lepton pair and use the invariant mass of the Z boson candidate paired with the highest p_T b-jet as final discriminant variable to perform the statistical analysis. It is then straight-forward to expect this search to efficiently cover the bottom left corners of both the $B\bar{B}$ and $T\bar{T}$ two dimensional planes of Figure 5.2.

The searches in the single lepton channel are focused on vector-like top partner pairs decays and are optimized for two distinctive signatures. One search exploits the boosted kinematics of the W boson from vector-like top partners decays to reconstruct it from its hadronic channel final state particles [1]. The heavy quark invariant mass is then reconstructed pairing the boosted W boson with a $b$jet and this distribution is used to perform the statistical analysis. It is evident that this search is going to cover the bottom right corner of the $T\bar{T}$ two dimensional plane of Figure 5.2.

The other search in the single lepton channel considers final states with high jet and $b$jet multiplicities as a result of the decay of at least one heavy quark into a Higgs boson (assumed to decay into $b\bar{b}$) and a top quark [2] (see Figure 5.4). The distribution of the total transverse momentum of the event is then used to perform the statistical analysis. This search is mainly sensitive to the top left corner of the $T\bar{T}$ two dimensional plane of Figure 5.2.

We will in the following treat in details the two searches for vector-like top partners performed in the single lepton channels, starting with the discussion of the common features between the two analyses.

5.2 Data sample and common event preselection

The data from pp collision events recorded at the ATLAS experiment during 2012 at a center of mass energy of $\sqrt{s} = 8$ TeV are considered. Physics object definitions were discussed in Chapter 4. Events collected during stable beam periods are required to pass data quality requirements and single lepton trigger selection. In order to maximize trigger efficiency, different transverse momentum threshold triggers are combined through a logical

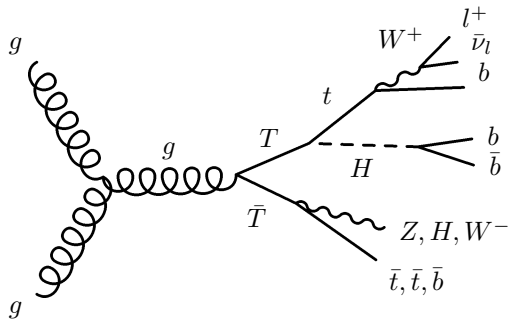


Figure 5.4: Feynman diagrams for the $T\bar{T} \to Ht + X$ decay entering the high jet and $b$jet multiplicity final states. Assuming the single lepton condition, in this picture all the bosons produced in the $\bar{T}$ decay will decay hadronically.

OR, with the lower p_{T} ones including isolation requirements that result in inefficiencies for high p_{T} lepton candidates, recovered with the use of the higher threshold triggers. The electron triggers have p_{T} thresholds of 24 and 60 GeV, the muon ones of 24 and 36 GeV (Section 4.4).

After passing trigger requirements, events with more than one lepton are discarded. In addition, the only lepton of the event has to match within $\Delta R < 0.15$ the triggered one. As basic preselection, four jets satisfying the conditions described in Section 4.6 are required, at least one of them being tagged as a $b$jet.

In order to suppress the multi-jet background from QCD processes, combined cuts on the $E_{\mathrm{T}}^{\mathrm{miss}}$ and on the tranverse mass of the leptonically decaying W boson m_T[2] are defined: $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} + m_T > 60$ GeV.

At this point, a simple consideration about the typical expected jet (and $b$jet) multiplicity based on counting the parton multiplicities and their flavor is made so as to define an orthogonality cut between the two analyses. Table 5.2 shows the number of jets ($b$jets) per decay channel combinations of $T\bar{T}$ pairs, in the case of single lepton selection with at least four jets (i.e. one W boson will always decay into lepton and neutrino, and Z boson decay to neutrinos is excluded in the $WbZt$ channel) and assuming that the Higgs boson decays to a bottom quark-antiquark pair. To avoid overlap between selected events from the two analyses, in the $T\bar{T} \to Wb + X$ analysis events with $\geq$6 jets and $\geq$3 $b$jets are rejected[3].

[2] $m_T = \sqrt{2p_{\mathrm{T}}^{\ell} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi)}$, with p_{T}^{ℓ} being the transverse momentum (energy) of the muon (electron) and $\Delta\phi$ the azimuthal angle separation between the lepton and the direction of the missing transverse momentum.

[3] As will be explained later in Section 7.2, another orthogonality cut will be applied in the low $b$jet multiplicity channel of the $T\bar{T} \to Ht + X$ analysis.

	Wb	Ht	Zt
Wb	**4 (2)**	6 (4)	**6** (**2**/4)
Ht	6 (4)	8 (6)	max: 8 (4/6) min: **6 (2)**
Zt	**6 (2/4)**	max: 8 (4/6) min: **6 (2)**	max: **8** (**2**/6) min: **6** (**2**/4)

Table 5.2: Jets ($b$jets) multiplicities in the various possible final states. Z boson decays 55% hadronically, 15% of the times into $b\bar{b}$, therefore the min/max number of $b$jets is reported. Highlighted in bold characters are the channels that after the orthogonality cut will contribute to the $T\bar{T} \to Wb + X$ analysis.

5.3 Background and signal modeling

All samples are enterely modeled using Monte Carlo simulation with the exception of QCD multi-jet events, which are derived using data-driven techniques, and background from W boson production in association with jets, which is obtained from Monte Carlo at first but then is normalized to data.

The main background for both analyses is $t\bar{t}$ production in association with jets ($t\bar{t}$+jets) and different choices for the generator are made in the analyses because of the specific needs of having well modeled regions. In the case of the $t\bar{t}$+jets background prediction for the $T\bar{T} \to Ht + X$ analysis further corrections to match the data are applied, due to a mismodeling in the heavy- and light-flavour content of the simulated sample (see Section 7.2).

W boson production in association with jets (W+jets) and QCD multi-jet events also contributes, the latter entering into the event selection via the misidentification of a jet or a photon as an electron or the presence of a non-prompt lepton from, e.g., semileptonic b- or c-hadron decay. Small background contributions originate from single top quark, Z+jets, diboson (WW, WZ, ZZ), and associated $t\bar{t}V$ ($V = W, Z$) and $t\bar{t}H$ production.

5.3.1 Monte Carlo simulated samples

With the exception of the signal samples, all simulated samples utilise `PHOTOS 2.15` [121] to model photon radiation and `TAUOLA 1.20` [122] to model τ decays.

All simulated samples include multiple pp interactions and make use of the `GEANT4` [99] detector geometry and response simulation [98] with the exception of the signal samples, for which a fast simulation of the calorimeter response is used.

All simulated samples are then processed through the same reconstruction software as the data and are reweighted to match the instantaneous luminosity profile in data. For more details on the Monte Carlo simulation chain we refer the reader to Chapter 3 and in particular to Section 3.3.

$t\bar{t}$ MC@NLO

Simulated samples of $t\bar{t}$ pair production in association with jets ($t\bar{t}$+jets or simply $t\bar{t}$ in the following) are generated with `MC@NLO` v4.01 [123, 124, 125] using the `CT10` set of

parton distribution functions (PDFs) [126], with the parton-shower and fragmentation steps being performed by HERWIG v6.520 [86]. The top quark mass is assumed to be equal to 172.5 GeV and the samples are normalized to approximate next-to-next-to-leading-order (NNLO) theoretical cross section [127]; the cross section used has been computed with HATHOR 1.2 [127] using the MSTW2008 NNLO PDF set [76] and is $\sigma_{t\bar{t}} = 238^{+22}_{-24}$ pb, where the total uncertainty results from the sum in quadrature of the scale and PDF+α_S uncertainties according to the MSTW prescription [128]. This is the $t\bar{t}$ used in the $T\bar{T} \to Wb + X$ analysis.

$t\bar{t}$ Alpgen

Simulated samples of $t\bar{t}$+jets are generated using the ALPGEN v2.13 [88] leading-order (LO) generator and the CTEQ6L1 PDF set [129], with parton shower and fragmentation modelled through HERWIG v6.520 [86].

A parton-jet matching scheme called "MLM matching" [130] is used in orderd to avoid double-counting of partonic configurations eventually generated both at the matrix-element calculation level and at the parton-shower evolution step.

Separate samples are generated for $t\bar{t}$+light jets ($t\bar{t}$+light or $t\bar{t}$+LF in the following, from "light flavour") with up to three additional light partons (u, d, s quarks or gluons), and for $t\bar{t}$+heavy-flavour jets ($t\bar{t}$+HF in the following), including $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$. An algorithm based on the angular separation between the extra heavy quarks is used to remove the overlap between $t\bar{t}q\bar{q}$ ($q = b, c$) generated from the matrix element calculation and from parton-shower evolution in the $t\bar{t}$+light samples is employed: matrix-element prediction is chosen over the parton-shower one when $\Delta R(q, \bar{q}) > 0.4$, else vice-versa.

Again a top quark mass of 172.5 GeV is assumed, and normalisation to the NNLO theoretical cross section is used (see 5.3.1)

W/Z+jets

Simulated samples of W/Z boson production in association with jets (W/Z+jets in the following) are generated with up to five additional partons using the ALPGEN v2.13 [88] LO generator and the CTEQ6L1 PDF set [129], interfaced to HERWIG v6.520 for parton showering and fragmentation. The MLM matching scheme is used also here to avoid double-counting of partonic configurations between matrix-element calculation and parton showering.

The W+jets samples are generated separately for W+light jets, $Wb\bar{b}$+jets, $Wc\bar{c}$+jets, and Wc+jets, with the relative contributions normalized using the fraction of b-tagged jets in W+1-jet and W+2-jets data control samples [131], while the Z+jets samples are generated separately for Z+light jets, $Zb\bar{b}$+jets, and $Zc\bar{c}$+jets and normalized to the inclusive NNLO theoretical cross section [132]. Overlap between $W/Zq\bar{q}$+jets ($q = b, c$) events generated from the matrix element calculation and those generated from parton-shower evolution in the W/Z+light jets samples is avoided via the same algorithm used for $t\bar{t}$ Alpgen.

Other backgrounds

Simulated samples of single top quark backgrounds corresponding to the s-channel and Wt production mechanisms are generated with MC@NLO v4.01 [123, 124, 125] using the CT10

PDF set [126]. In the case of t-channel single top quark production, the ACERMC v3.8 LO generator [97] with the MRST LO** PDF set is used.

Simulated samples of $t\bar{t}$ produced in association with a W or Z boson ($t\bar{t}V$ ($V = W, Z$) in the following) are generated with the MADGRAPH v5 LO generator [96] and the CTEQ6L1 PDF set.

Samples of $t\bar{t}$ produced in association with a Higgs boson ($t\bar{t}H$ in the following) are generated with the PYTHIA 6.425 [133] LO generator and the MRST LO** PDF set [134], assuming a Higgs boson mass of 125 GeV and considering the $H \to b\bar{b}$, $c\bar{c}$, gg, and W^+W^- decay modes.

Parton shower and fragmentation are modelled with HERWIG v6.520 [86] in the case of MC@NLO, with PYTHIA 6.421 in the case of ACERMC, and with PYTHIA 6.425 in the case of MADGRAPH. All these samples are generated assuming a top quark mass of 172.5 GeV. The single top quark samples are normalised to the approximate NNLO theoretical cross sections [135, 136] using the MSTW2008 NNLO PDF set, while the $t\bar{t}V$ samples are normalised to the NLO cross section predictions [137, 138]. The $t\bar{t}H$ sample is normalised using the NLO theoretical cross section and branching ratio predictions [139]. Finally, the diboson backgrounds are modelled using HERWIG with the MRST LO** PDF set, and are normalised to their NLO theoretical cross sections [140].

Signal samples

For vector-like T signals, samples corresponding to a singlet T quark decaying to Wb, Zt and Ht are generated with the PROTOS v2.2 LO generator [54, 141] using the MSTW2008 LO PDF set, and interfaced to PYTHIA for the parton shower and fragmentation.

For each decay channel (Wb, Zt and Ht) the branching ratio has been set to 1/3. Events are reweighted in order to reproduce any desired branching ratio configuration. The predicted branching ratios in the weak-isospin singlet and doublet scenarios as a function of m_T are given in Table 5.3.

The m_T values considered range from 350 GeV to 850 GeV in steps of 50 GeV, with the Higgs boson mass assumed to be 125 GeV. All Higgs boson decay modes are considered, with branching ratios as predicted by HDECAY [142].

Signal samples are normalized to the approximate NNLO theoretical cross sections [127] using the MSTW2008 NNLO PDF set. The cross section values used are summarized in Table 5.4.

5.3.2 W+jets background normalisation

For the W+jets background, a normalisation from data for the shapes obtained from the simulation is derived since the simulation overestimates the number of W+jets events by up to ∼20%, depending on the jet multiplicity. Data driven techniques are also used to correct the heavy flavor (HF) composition of the W+jets events, such that at the end the scale factors applied are the product of the overall normalization scale factor and the scale factors obtained for Wbb, Wcc, Wcj and Wjj components.

In protons the PDF of quarks and anti-quarks are different, hence the production of W bosons in pp collisions will have different cross sections for processes like $\sigma(u\bar{d} \to W^+)$ and

	Singlet			Doublet		
m_T (GeV)	$BR(T \to Wb)$	$BR(T \to Zt)$	$BR(T \to Ht)$	$BR(T \to Wb)$	$BR(T \to Zt)$	$BR(T \to Ht)$
350	0.545	0.116	0.338	0.000	0.255	0.745
400	0.513	0.139	0.348	0.000	0.285	0.715
450	0.502	0.158	0.341	0.000	0.316	0.684
500	0.497	0.173	0.330	0.000	0.343	0.657
550	0.495	0.185	0.321	0.000	0.365	0.635
600	0.494	0.194	0.312	0.000	0.383	0.617
650	0.494	0.202	0.304	0.000	0.399	0.601
700	0.494	0.208	0.298	0.000	0.411	0.589
750	0.494	0.214	0.292	0.000	0.422	0.578
800	0.494	0.218	0.288	0.000	0.431	0.569
850	0.494	0.222	0.284	0.000	0.439	0.561

Table 5.3: Branching ratios for T decay as a function of m_T as computed with `PROTOS` in the weak-isospin singlet and doublet scenarios. The same values are used in the graphical representation of Figure 5.1

m_T (GeV)	$\sigma(TT)$ (pb)	Scale uncertainties (pb)	PDF+α_s uncertainties (pb)	Total uncertainty (pb)
350	5.083	+0.140/-0.285	+ 0.569/-0.488	+0.586/-0.565
400	2.296	+0.066/-0.130	+ 0.269/-0.221	+0.277/-0.257
450	1.113	+0.034/-0.063	+ 0.136/-0.107	+0.140/-0.125
500	0.5702	+0.0185/-0.0327	+ 0.0723/-0.0545	+0.0746/-0.0636
550	0.30545	+0.01040/-0.01769	+ 0.04012/-0.02889	+0.0414/-0.0339
600	0.1696	+0.0060/-0.0099	+ 0.0230/-0.0161	+0.0238/-0.0189
650	0.09707	+0.00359/-0.00571	+ 0.01363/-0.00936	+0.01410/-0.01097
700	0.05694	+0.00218/-0.00338	+ 0.00828/-0.00559	+0.00856/-0.00653
750	0.03411	+0.00135/-0.00204	+ 0.00513/-0.00343	+0.00530/-0.00400
800	0.02080	+0.00085/-0.00126	+ 0.00329/-0.00216	+0.00340/-0.00250
850	0.01287	+0.00054/-0.00079	+ 0.00215/-0.00138	+0.00222/-0.00159

Table 5.4: Theoretical cross section at NNLO for TT production as a function of m_T as computed by `HATHOR`, and scale and PDF uncertainties.

$\sigma(d\bar{u} \to W^-)$. This charge asymmetry in W+jets production is predicted theoretically [143] and can be measured in data and then used to derive the correct overall normalization of the process. The total number of W+jets events in data ($N_W = N_{W^+} + N_{W^-}$) is estimated measuring the difference between the number of positively- and negatively-charged W bosons ($(N_{W^+} - N_{W^-})_{\text{meas}}$) and compared to the prediction from Monte Carlo simulation:

$$N_W = \left(\frac{N_{W^+} + N_{W^-}}{N_{W^+} - N_{W^-}}\right)_{\text{MC}} (N_{W^+} - N_{W^-})_{\text{meas}}. \tag{5.1}$$

Equation 5.1 is evaluated in the signal region for top searches of at least 4 jets without requirements on the b-tagging multiplicity. The ratio of event with at least one b-tagged jet is estimated multiplying the prediction from Equation 5.1 by the tagging fractions in the different jet bins evaluated in data by subtracting the Monte Carlo predictions of non-W + jets backgorunds.

To account for the different flavor composition, additional scale factors are derived in the 2 jets bin considering that the total number of tagged W + jets events in the ith jet bin

is given by:

$$N^{W,\text{tag}}_{i\ \text{jets}} = N^{W,\text{pretag}}_{i\ \text{jets}} \left(\sum_{\text{x flavor}} F_x P_x \right)_{i\ \text{jets}}, \tag{5.2}$$

where F_x are the flavor fractions $N^{\text{pretag}}_x / N^{\text{pretag}}$ (which add up to unity for each jet bin) and P_x are the b-tagging probabilities for each flavor type $x = bb, cc, c, light$. After evaluating Equation 5.2 in the 2 jets bin for data and Monte Carlo the scale factors are obtained as $K_x = F^{\text{data}}_x / F^{\text{MC}}_x$ and propagated into the other jet multiplicities bins by correcting for the different normalization:

$$\sum_{\text{x flavor}} K_{x,2\ \text{jets}} F^{\text{MC}}_{x,i\ \text{jets}} = A, \qquad \neq 1 \text{ if } i \neq 2. \tag{5.3}$$

5.3.3 Multi-jet background

QCD multi-jet production can pass the event selection in the electron channel as non-prompt electrons from e.g. heavy flavor quark decays, as electrons from photon conversions or as jets mis-identified as electrons because they left a high amount of energy in the electromagnetic calorimeter like in the case of π^0 and η decays into two close-by photons. For events in the muon channel the main contributions come from non-prompt leptons from semileptonic b- and c-hadron decays.

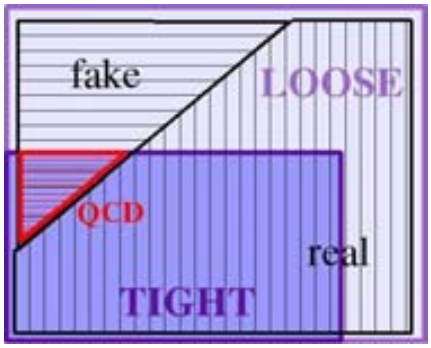
Although these kind of events rarely pass the quality cuts required at the lepton reconstruction stage, the production cross section is so high (orders of magnitude more than $t\bar{t}$ production) that the contribution to the background from QCD multi-jet events is no longer negligible. The QCD multi-jet contribution is estimated via data-driven methods, since simulation is not expected to predict this contribution with the desired level of accuracy.

The technique used is called "Matrix Method" (MM in the following) [144]. The basic principle is to divide the data sample into two categories, one of events passing the standard selection criteria ("tight" events), the other including also leptons satisfying looser requirements ("loose" events). Loose leptons can reasonably be considered as either *real* leptons or *fake* leptons, and it can be assumed that most of the real leptons will pass the tight selection. A good evaluation of multi-jet contamination is then that fraction of fake events going through the tight requirements. A pictorial representation of the sampling space is shown in Figure 5.5(a). Typically tight selection requirements are the same as the ones defined to reconstruct the lepton (Sections 4.4 and 4.5) where then some cuts are either removed or changed to obtain the loose sample. In the case of muons, loose muons are as final leptons but without the mini-isolation requirement. Loose electrons are like final electrons where the `tight++` requirement is replaced by the `medium++` one, isolation requirements are omitted and a condition to veto against conversion is added.

Defining $N^{\text{loose}}_{\text{real}}$ ($N^{\text{loose}}_{\text{fake}}$) as the number of real (fake) leptons events satisfying the loose selection requirements, and $N^{\text{tight}}_{\text{real}}$ ($N^{\text{tight}}_{\text{fake}}$) as the number of real (fake) leptons events satisfying the tight selection requirements, we can write:

$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{loose}_{\text{fake}}, \tag{5.4}$$

$$N^{\text{tight}} = \epsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \epsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}}. \tag{5.5}$$



(a)

Figure 5.5: The events passing loose selection criteria can be real or fake leptons. Tighter requirements are added to the loose selection ones to define a sample of tight event which will contain real leptons as well as multi-jet background events.

where ϵ_{real} (ϵ_{fake}) is the *efficiency* of selecting real (fake) loose leptons as tight leptons, i.e.:

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}}, \tag{5.6}$$

$$\epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}. \tag{5.7}$$

The number we are interested in to estimate the background from multi-jet events is the amount of fake leptons leaking into the tight selection region, which comes out from elaborating the previous equations and is:

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}}(N^{\text{loose}}\epsilon_{\text{real}} - N^{\text{tight}}). \tag{5.8}$$

An important condition for this method to work is that $\epsilon_{\text{real}} \gg \epsilon_{\text{fake}}$, which holds as $\epsilon_{\text{real}} \sim 1$, while ϵ_{fake} is in general well below 1. Efficiencies for fake and real leptons to pass the tight requirement are measured in dedicated control regions, enriched respectively with fake and real leptons. The ϵ_{fake} is measured in control regions with small $E_{\text{T}}^{\text{miss}}$ and $m_T(W)$ (a triangular cut on the sum of the two variables is applied), which is enriched in multi-jet event contributions. The ϵ_{real} is in general measured on the contrary in regions with high $E_{\text{T}}^{\text{miss}}$ or $m_T(W)$ to select leptons from a leptonic decay of a W boson. While ϵ_{real} shows basically no dependency on the event topology and has a stable value close to unity, ϵ_{fake} needs to be parametrized in terms of some observables towards which it shows dependency and its values vary, typically, between 0.2 and 0.5.

In Appendix A the MM used for the estimation of multi-jet background in the muon channel is described in some more details, as the author of this dissertation directly contributed to its development.

5.4 Data to Monte Carlo comparison

In order to validate the good modeling of the main backgrounds, we present here a first set of data to Monte Carlo comparisons in control regions defined at the preselection level (see Section 5.2), very far from the signal regions that will be defined for the two analyses. We are indeed interested in checking the data and backgrounds agreement in selections free from eventual signal, and therefore a *blinding cut* is defined using the H_T variable defined as the scalar sum of the lepton transvers momentum, the first four leading jets transvers momenta and the missing transverse energy. Considering the typical hardness of vector-like top partners decay events, the region with $H_T < 800$ GeV can safely be considered as signal-free. It is worth noticing here that in the $T\bar{T} \to Wb + X$ analysis this cut is going to be inverted, obtaining an efficient reduction of background contributions, while in the $T\bar{T} \to Ht + X$ analysis the H_T variable (with a slightly different definition, i.e. all jets are included in the sum and not only the first four) is going to be used to discriminate signal and background in the statistical analysis. The preselection requirement of at least one $b$jet enriches these control regions in $t\bar{t}$+jets background. A more complete set of plots is available in Appendix B, where also a selection vetoing $b$jets is considered to check the modeling of W+jets background. Yields for both selections are shown in Table 5.5 for electron and muon channels combined. Individual tables are available in the appendices.

	$N_{jets} \geq 4 N_{b-jets} = 0$	$N_{jets} \geq 4 N_{b-jets} \geq 1$
MultiJet	15134.22 ± 94.93	6263.90 ± 73.82
Single top	3949.71 ± 58.97	14375.50 ± 107.31
Diboson	2171.71 ± 21.67	548.06 ± 11.52
Z+jets	31401.11 ± 378.84	5804.27 ± 145.76
W+jets	167551.08 ± 946.56	35920.56 ± 525.37
$t\bar{t}$V	112.69 ± 1.00	680.19 ± 2.31
$t\bar{t}$H (125)	23.52 ± 0.29	220.03 ± 0.74
$t\bar{t}$ MC@NLO	34563.32 ± 131.34	202041.60 ± 284.98
$t\bar{t}$ Alpgen	35634.36 ± 113.42	219150.18 ± 254.97
Tot Bkg w/ MC@NLO	254907.36 ± 1034.26	265854.11 ± 628.95
Tot Bkg w/ Alpgen	255978.41 ± 1032.14	282962.69 ± 615.94
$T\bar{T}$ (600) chiral	3.28 ± 0.63	36.30 ± 2.13
Data	238709.00 ± 488.58	256993.00 ± 506.94

Table 5.5: Yields for data, backgrounds and signal in the two blinded control regions defined as "preselection" with a veto on b-tagged jets (enriched in $W+$jets) and with at least one b-tagged jet (enriched in $t\bar{t}$).

5.5 Systematic uncertainties

In addition to the uncertainty that comes from the stocastical nature of events we are subject to other sources that are systematical in the sense that they will bias the result towards a definite direction. They can come from detector measurements, from the way we reconstruct the objects, from the Monte Carlo modelling, and can affect either only the

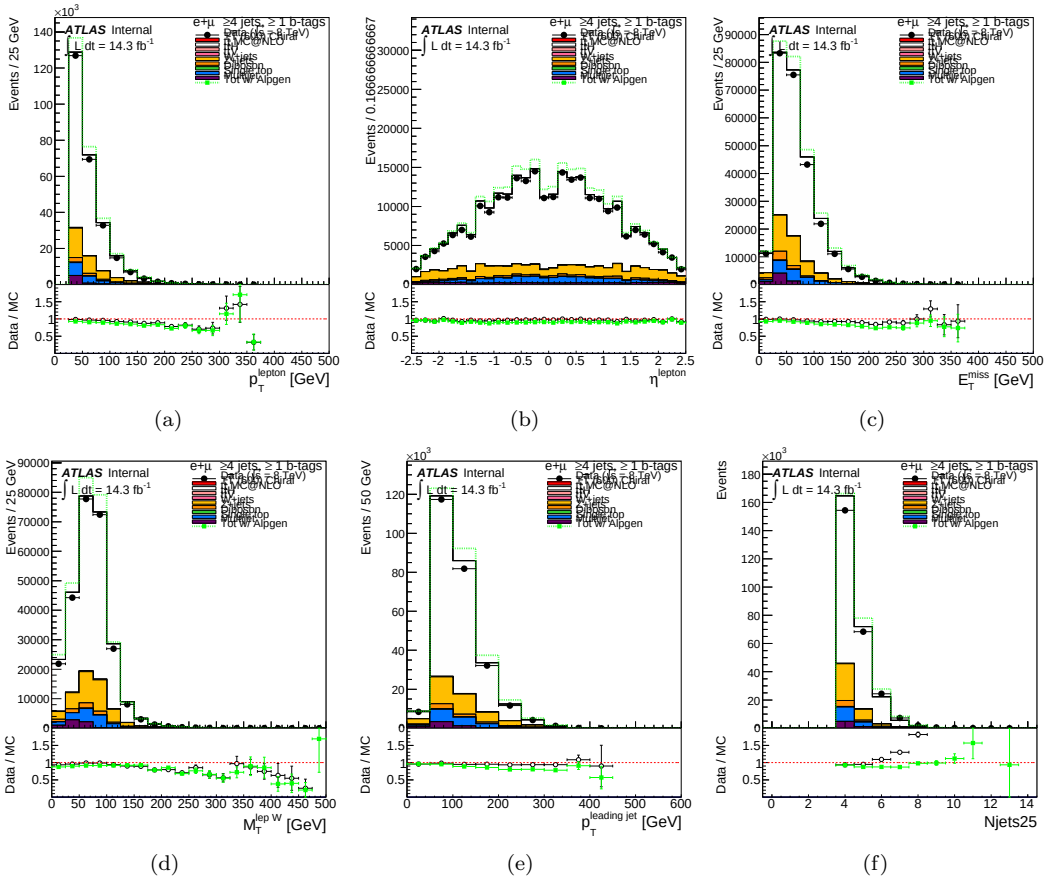


Figure 5.6: Data to Monte Carlo comparison plots for (a) lepton transverse momentum and (b) pseudorapidity, (c) missing tranverse energy, (d) transverse mass of W boson, (e) leading jet transverse momentum, (f) number of jets in the event with $p_{\mathrm{T}} > 25$ GeV.

normalization of the total event yield (and are called "normalization-only" sytematics) or also the shape of the distributions (and are called "shape and normalization" systematics).

The individual sources of systematics are treated as uncorrelated, while the correlations eventually present in the particual systematic uncertainty are kept for the various processes and channels. Most of the systematic uncertainties are common to the two analyses presented with only minor differences, e.g. in the $T\bar{T} \to Ht + X$ analysis the systematic uncertainty affecting the jet energy scale is split into 9 components, while it has a unique component in the $T\bar{T} \to Wb + X$ analysis. The full list of systematics considered is presented in Table 6.7, labelling them as "normalization-only" or "shape and normalization" systematics and indicating the number of components.

The systematic uncertainties are treated inside the `MCLimit` package [**?**] developed for ATLAS heavy quark searches based on the original `MCLimit` code developed by the CDF collaboration [145, 146, 147]. Here the histograms are interpolated between the nominal and the systematically shifted templates bin-by-bin, with a shift of $+0.5\sigma$ corresponding to half way between the nominal and the $+1\sigma$ shifted template. This interpolation method is

Systematic uncertainty	$T\bar{T} \to Wb + X$		$T\bar{T} \to Ht + X$	
	Status	Components	Status	Components
Luminosity	N	1	N	1
Lepton ID+reco+trigger	N	1	N	1
Jet vertex fraction efficiency	SN	1	SN	1
Jet energy scale	SN	1	SN	8
Jet energy resolution	SN	1	SN	1
b-tagging efficiency	SN	9	SN	9
c-tagging efficiency	SN	5	SN	5
Light jet-tagging efficiency	SN	1	SN	1
$t\bar{t}$ cross section	N	1	N	1
$t\bar{t}V$ cross section	N	1	N	1
$t\bar{t}H$ cross section	-	-	N	1
Single top cross section	N	1	N	1
Dibosons cross section	N	1	N	1
W+jets normalization	N	5	-	-
Z+jets normalization	N	1	-	-
V+jets normalization	-	-	N	1
Multijet normalization	-	-	N	1
$t\bar{t}$ modelling	SN	3	SN	3
V+jets modelling	SN	1	-	-
$t\bar{t}$+heavy-flavour fractions	-	-	N	1

Table 5.6: List of systematic uncertainties considered in the two analyses. We label as "N" ("SN") uncertainties taken as "normalization-only" (both "shape" and "normalization") for all processes and channels. Some of the systematic uncertainties are split into more components for a more accurate treatment.

called vertical morphing and uses a linear bin-by-bin interpolation, but for variations below 1σ we use quadratic interpolation to ensure a continuous derivative at zero shift. Pseudoexperiments are generated using these interpolated numbers for all systematic uncertainties.

Details on specific treatments of systematics in particular channels will be given in the dedicated sections of the corresponding analysis chapters (Section 6.6 and 7.5).

5.5.1 Luminosity

The uncertainty on the absolute integrated luminosity is estimated to be of 3.6% [71]. This systematic uncertainty is applied to all processes except the QCD multi-jet background.

5.5.2 Object definitions

The event reconstruction introduces uncertainties on the definition of leptons, jets and on the b-, c-, and light flavour-tagging. In the following the related systematic uncertainties considered are described.

Lepton reconstruction, identification and trigger scale factors

In Sections 4.4 and 4.5 the reconstruction of leptons was introduced, explaining the need of adjusting the differences between data and simulation in the efficiency for reconstruction,

identification and trigger. This is done by applying to Monte Carlo samples some scale factors derived with tag-and-probe techniques on $Z \to \ell^+\ell^-$ ($\ell = e, \mu$) data and simulated samples. For each of these three sources of systematic uncertainty, the overall systematic uncertainty is obtained as the quadratic sum of the statistical and systematic uncertainties on the corresponding scale factor.

In the e+jets channel, the systematic uncertainties corresponding to electron reconstruction, identification and trigger, are 0.3%, 1.1% and 0.2%, respectively. In the μ+jets channel, the systematic uncertainties corresponding to muon reconstruction, identification and trigger, are 0.2%, 1.1% and 1.4%, respectively. A total uncertainty on the signal and background acceptances of 2% is estimated.

Lepton momentum scale and resolution

To check the accuracy of the lepton momentum scale and resolution simulated samples of $Z \to \ell^+\ell^-$ and $J/\psi \to \ell^+\ell^-$ are used to reconstruct the particles masses (for electrons also $W \to e\nu$ events are used from E/p studies). The small discrepancies observed between data and simulation are corrected adjusting lepton energy scale and resolution in Monte Carlo samples only for muons, while for electrons energy resolution corrections are also applied only to Monte Carlo samples but energy scale corrections are applied to data in all detector regions and to simulation only in the calorimeter transition region. The systematic uncertainties on these scale factors are varied separately and the result on the total yields are are at the sub-percent level and considered therefore ngligible in the analyses.

JVF efficiency

Recalling the cut applied on the JVF variable (Section 4.6) of $|\text{JVF}| > 0.5$, the per-jet efficiency of this requirement is estimated in $Z(\to \ell^+\ell^-)$+1-jet events both in data and Monte Carlo simulation. Event enriched in hard-scatter jets are selected separately from events enriched in jets from pile-up interactions and specific efficiency and inefficiency scale factors are measured. Scale factors for pileup jets are estimated to be consistent with 1, while efficiency for hard-scatter jets goes from $\sim$1.03 for jets with $p_{\text{T}} = 25$ GeV down to $\sim$1.01 for jets with $p_{\text{T}} > 150$ GeV. An overall event weight is obtained as the product of all per-jet scale factors and is applied to the Monte Carlo samples. The systematic uncertainty from the propagation of the per-jet scale factor uncertainty gives an overall uncertainty on the signal and background acceptance of $\sim$2.5%.

Jet energy scale

The systematic uncertainty on the Jet Energy Scale (JES) has been derived combining the information from both test-beam and collision data and Monte Carlo simulation [148, 149, 150]. Pile-up activity produces an additional source of systematic uncertainty which depends on the number of primary vertices and on the average number of interactions per bunch crossing $< \mu >$. Momentum balance techniques in Z+jets, γ+jets and multi-jet events are combined to derive a small residual correction for jets in the transverse momentum range 20 GeV $< p_{\text{T}} <\sim$ 1 TeV.

The overall variation due to JES systematic uncertainty evaluated in the central detector region is ∼4% for jets with $p_{\mathrm{T}} = 25$ GeV and improves to ∼1% for jets with $p_{\mathrm{T}} = 500$ GeV [151]. The effect of this systematic uncertainty is implemented in the analyses by varying in the Monte Carlo samples the transverse momentum of all the selected jets by ±1 standard deviation. In each event the missing transverse momentum $E_{\mathrm{T}}^{\mathrm{miss}}$ is then corrected consistently to the varied p_{T} of the jets and all the variables involving jets are also recomputed.

As can be seen in Table 6.7, for the $T\bar{T} \to Ht+X$ analysis the JES systematic uncertainty is split into 8 uncorrelated components, each with a different jet p_{T} and η dependence, which are treated independently. The $T\bar{T} \to Wb + X$ instead uses the total JES uncertainty as a single uncertainty resulting from the sum in quadrature of all individual sources.

Jet energy resolution

The Jet Energy Resolution (JER) was measured with two *in-situ* techniques [148] as a function of the jet transverse momentum and pseudo-rapidity. It is consistent in data and Monte Carlo simulation and no corrections are needed. To account for the systematic uncertainty the quadratic difference between the JER in data and in simulated samples is used to smear the energy of jets in Monte Carlo simulation and a new varied sample is obtained with a different normalisation and variable distributions shapes. The final result is then symmetrised to obtain both positive and negative variations.

Heavy- and light-flavour tagging

The efficiencies in heavy flavour (b and c) jets identification with the b-tagging algorithm are measured in data and depend on the individual jet flavour [152, 153, 154]. These efficiencies are measured from data and depend on the jet flavour: in Monte Carlo events b (c) jet efficiencies are corrected with scale factors of 0.9–1.0 (1.1–1.2) depending on p_{T}, light jet efficiencies are corrected with a scale factor of ∼1.3. Every jet in the Monte Carlo simulated events is corrected depending on its flavour, p_{T} and η The uncertainty on these scale factors is between 7% and 13% for b jets, between 15% and 39% for c jets, and ∼25% for light jets.

As was reported in Table 6.7, the systematic uncertainty on b-tagging (c-tagging) efficiency is divided into nine (five) independent components that correspond to an eigenvector from the diagonalization of the matrix containing the information on the total uncertainty per p_{T} bin and the bin-to-bin correlations (see [155, Appendix P] for more details). These individual sources of systematic uncertainties are taken as uncorrelated between b, c jets, and light flavour jets. In Monte Carlo simulated events a per-jet weighting procedure [156] is applied in order to propagate the b-tagging calibration and related uncertainties.

5.5.3 Theoretical cross-sections

Normalization-only systematic uncertainties on the theoretical cross-sections are considered as follows: +10%/-11% for the inclusive $t\bar{t}$ production cross section evaluated at approximate NNLO using `HATHOR` [127]; +5%/-4% and ±5% for the theoretical cross sections of the single top [135, 136] and diboson [140] backgrounds respectively; +12%/-17%

and $\pm 30\%$ for the theoretical cross sections of the $t\bar{t}H$ [139] and $t\bar{t}V$ [137, 138] backgrounds respectively.

5.5.4 Normalizations of data-driven backgrounds and background modeling

Because of the differences between the effects of these systematic uncertainties in the $T\bar{T} \to Wb+X$ and $T\bar{T} \to Ht+X$ analyses, the reader is referred to the specific Sections 6.6 and 7.5

5.6 Statistical analysis

To test the presence or absence of signals from new physics we use the CL_{s} method [157, 158] originally developed in the context of Higgs searches at the LEP collider [159]. The fundamental principle of this technique is that, in order to exclude or verify a theory predicting some kind of signal over some other kind of background, both "background only" and "signal plus background" hypotheses have to be tested. Once a test statistic Q has been chosen analysis-wise, two confidence levels (CL) are defined for the two hypotheses:

$$CL_{\mathrm{b}} = P(Q \leq Q_{\mathrm{obs}}|\mathcal{H}_b), \tag{5.9}$$

$$CL_{\mathrm{s+b}} = P(Q \leq Q_{\mathrm{obs}}|\mathcal{H}_{s+b}), \tag{5.10}$$

with Q_{obs} being the observed value in data of the test statistic. It is good to identify well separated test statistic distributions in order to be able to clearly distinguish between events consistent with the backgroud only prediction and events showing deviations consistent with the signal+background hypothesis.

The CL_{s} is then defined as the ratio of the two:

$$CL_{\mathrm{s}} = CL_{\mathrm{s+b}}/CL_{\mathrm{b}} \tag{5.11}$$

and its meaning is given by its relation with the confidence level CL for the exclusion of the signal hypothesis $(1 - CL_{\mathrm{s}}) \leq CL$. Similarly, the confidence level for excluding the background only hypothesis is the p-value $1 - CL_{\mathrm{b}}$ and the value required to claim discovery is of $\sim 10^{-7}$. The motivation that led to the definition of CL_{s} instead of using $CL_{\mathrm{s+b}}$, which also corresponds to the confidence level in excluding the signal+background hypothesis, is that the latter might wrongly exclude scenarios to which the analysis is simply not sensitive to like is the general case for searches for rare events.

For the two analyses presented in this dissertation, the test statistic is defined as a log-likelihood ratio

$$LLR = -2 \log \frac{\mathcal{L}(\mathrm{data}|\mathcal{H}_{s+b})}{\mathcal{L}(\mathrm{data}|\mathcal{H}_b)}, \tag{5.12}$$

where the likelihoods $\mathcal{L}(\mathrm{data}|\mathcal{H}_{s+b})$ and $\mathcal{L}(\mathrm{data}|\mathcal{H}_b)$ for signal+background and background only hypothesis are built from the chosen discriminant variable distribution as the bin by bin product of Poisson probabilities to observe the data under one or the other hypothesis.

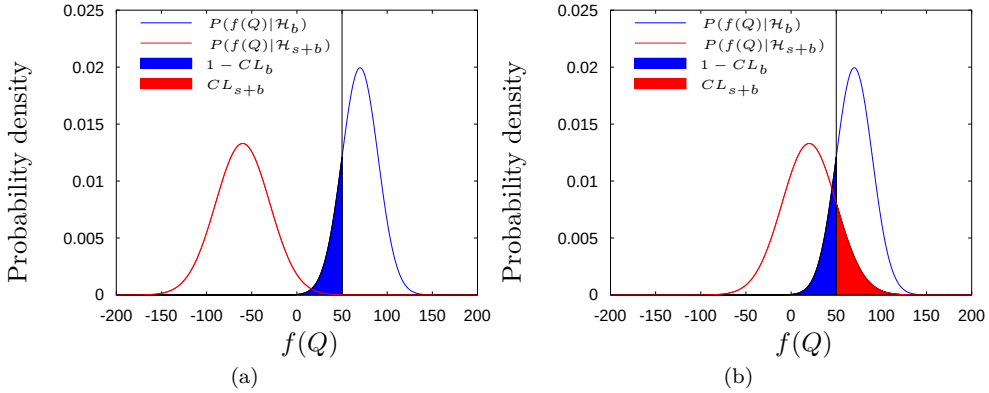


Figure 5.7: Example of (a) sensitive and (b) not sensitive analyses using the test statistics $f(Q)$, often chosen as $-2\log Q$.

Eventually more than one channel can be combined (like will be the case in the $T\bar{T} \to Ht+X$ analysis) and the complete formula for the likelihood computation reads:

$$\begin{aligned} -2\log\mathcal{L}(\text{data}|\mathcal{H}_x) &= -2\log\mathcal{L}(\vec{n}|R,\vec{s},\vec{b},\vec{\theta}) \\ &= -2\sum_{i=1}^{N_{chan}}\sum_{j=1}^{N_{bins}}(n_{ij}\log\mu_{ij}-\mu_{ij}) + \sum_{k=1}^{N_{par}}\theta_k^2. \end{aligned} \tag{5.13}$$

Here the first sum is over the number of channels combined in the analysis $N_{chan} \geq 1$ and the second sum is over the number of the discriminant variable histogram bins N_{bins}. n_{ij} (μ_{ij}) is the number of events in data (expected number of events) for channel i and histogram bin j. μ_{ij} is given by $\mu_{ij} = Rs_{ij}(\vec{\theta}) + b_{ij}(\vec{\theta})$, where s_{ij} and b_{ij} represent the expected signal and background yields, the first being equal to zero in the background only hypothesis. R is a scaling parameter applied to the signal to test the sensitivity of the search and $\vec{\theta}$ are the nuisance parameters parametrizing the effect of systematic uncertainties. The statistical uncertainty of the Monte Carlo samples is also taken into account when computing our likelihoods as an uncertainty of the templates, which in the case of the non-$t\bar{t}$ background are merged into a single template with different weights. For both hypotheses pseudoexperiments are generated to account in each bin for statistical fluctuations (Poisson-distributed) and systematic variations (Gaussian-distributed). The effect of systematic uncertainties are described by nuisance parameters taken at their nominal values and no parameter fitting is performed. In the case of the $T\bar{T} \to Ht+X$ analysis it will be explained that two additional nuisance parameters are introduced and fitted to help costraining the sensitivity degradation due to a poor Monte Carlo modeling of $t\bar{t}$ heavy flavor component.

In absence of data excess over background prediction, values of $CL_s < 0.05$ are considered to exclude a signal cross section at 95% CL.

Search for $T\bar{T}$ decaying to $Wb + X$

After an overview of the general features common to the two searches for vector-like top partner pairs in the single-lepton channel that are the object of this dissertation, we present in this chapter the search for $T\bar{T}$ production with at least one heavy quark decaying to a W boson and a bottom quark, referred to as the $T\bar{T} \to Wb + X$ analysis. This search is particularly optimized for the $T \to Wb$ channel and updates an analysis performed with the data collected by ATLAS in 2011 of pp collisions at a center of mass energy of $\sqrt{s}$=7 TeV [3], whose main features and results are reported in Appendix D. The strategy followed is to reconstruct the W boson decaying into two light jets exploiting the different kinematic characteristics of W bosons from a heavy object like the vector-like top partner and the lighter Standard Model top quark (see Section 6.1). The reconstruction of the W bosons is illustrated in Section 6.2. Section 6.3 describes the requirements to select the final signal region and, as will be explained, two signal regions are identified, one called "Loose" and the other "Tight". In Section 6.4 we define "control regions" depleted of signal that gradually approach the final signal region in order to check the good modeling of real data by the various expected backgrounds contributions. The final discriminant chosen to search for the signal is the heavy quark reconstructed mass, defined in Section 6.5. Section 6.6 summarizes the analysis-specific systematic uncertainties considered in this search, which are in addition to those described in Section 5.5. Section 6.7 is devoted to the results.

6.1 Analysis strategy

The very high center of mass energy available in the pp collisions provided by the LHC can either be used by Nature to produce massive particles like the heavy quarks we are looking for or to provide lighter particles with high momentum. In the case of $t\bar{t}$ production, the main and irreducible background of this search as both particles decay to a W boson and a bottom quark, the particles will receive a boost which is transferred to their decay products. This means that the W boson and the bottom quark from a Standard Model top quark will be produced close-by along the direction of the parent top quark. On the contrary, heavy top-like quarks will be produced almost at rest, as most of the center of mass energy

goes into their mass, and their decay products will be produced "almost" back-to-back. In this case, both the W boson and the bottom quark have the typical high momentum of the decay products fo heavy particles, resulting in a high-p_T $b$jet and close-by light jets (lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$) from the W boson decay in the hadronic (leptonic) channel. This difference, represented in the drawing of Figure 6.1, is at the basis of the $T\bar{T} \to Wb + X$ search strategy.

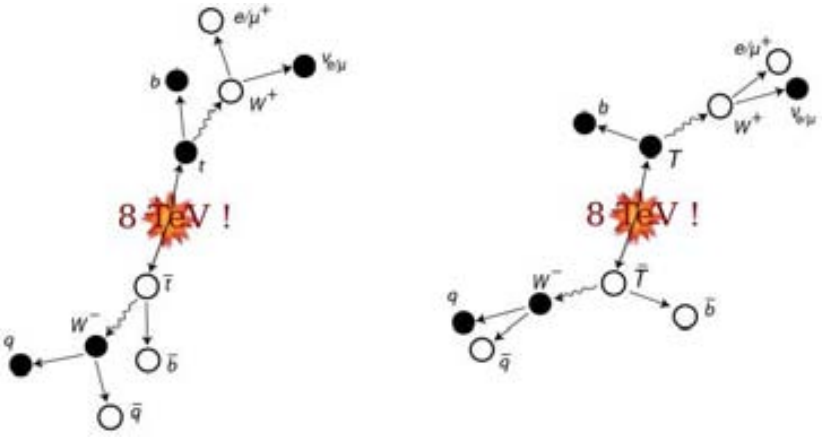


Figure 6.1: Pictorial representation of the typical kinematic configuration for top pairs decays (left), where the boosted objects are the top quarks themselves, and vector-like top pairs decays (right), where the boosted objects are the W bosons.

The strategy is therefore settled as follows: the two W bosons, one decaying leptonically and one decaying hadronically, are reconstructed as detailed in Section 6.2; cuts based on the hardness of the event, as well as on the angular separation of the objects, are defined in order to achieve a high S/B ratio; the heavy quark mass is reconstructed and used for the statistical analysis.

6.2 W boson reconstruction

In order to exploit the difference between the kinematics of signal and background, the angular separation of the final state objects is considered. Since the W boson decay in the lepton channel cannot be fully reconstructed because of the presence of a neutrino, the analysis focuses on the reconstruction of the W boson that decays in the hadronic channel (W_{had}). The two jets from the decay of the W_{had} are expected to be light-flavored jets. At the preselection level at least one b-tagged jet is required. Since two b-tagged jets are expected in the final state, and in order to avoid the loss in acceptance that would follow a strict cut like requiring at least two b-tagged jets[1], the two jets with the highest weight computed from the b-tagging algorithm are considered "$b$jets". Amongst the light jets two cases are considered: either the two light flavoured quarks were produced so close-by that after hadronization the jet reconstruction algorithm identifies a single jet, or two

[1]Considering our choice of the working point for the b-tagging `MV1` algorithm with a 70% efficiency, a selection requiring at least two b-tagged jets would result in a 50% selection efficiency.

nearby, distinct jets are reconstructed. To account for the first category of events, jets with $p_{\mathrm{T}} > 250$ GeV and a mass between 60 GeV and 120 GeV, an interval chosen to cover both the world-average W and Z boson masses values of $m_W = 80.4$ GeV and $m_Z = 91.2$ GeV and hence increasing the acceptance for $T \to Zt$ events, are classified as $W_{\mathrm{had}}^{\mathrm{type\ I}}$ candidates. If no $W_{\mathrm{had}}^{\mathrm{type\ I}}$ candidates are found, jets are paired in di-jet systems if their angular separation $\Delta R(j,j)$ is lower than 0.8 and, if $p_T(jj) > 200$ GeV and their invariant mass m_{jj} lies in the same window as for the $W_{\mathrm{had}}^{\mathrm{type\ I}}$ candidates, they are taken as a $W_{\mathrm{had}}^{\mathrm{type\ II}}$ candidate. If multiple candidates are found, the one with mass closest to the nominal W boson mass is chosen as the hadronic W boson of the event. Figures 6.2(a) and 6.2(b) show the mass distribution of the $W_{\mathrm{had}}^{\mathrm{type\ I}}$ and $W_{\mathrm{had}}^{\mathrm{type\ II}}$ candidates before the mass window cut is applied. The number of W_{had} candidates after preselection is shown in Figure 6.2(c).

For what concerns the W boson decaying in the leptonically (W_{lep}), it is reconstructed using the lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$, which is considered as the transverse momentum of the neutrino. In order to define the neutrino longitudinal momentum, it is required that the invariant mass of the two-body system composed by the lepton and the neutrino equals the nominal W boson. The neutrino 4-momentum is therefore set using the $E_{\mathrm{T}}^{\mathrm{miss}}$ X and Y components while the Z component p_{Z_ν} is computed from:

$$P_W^2 = (P_l + P_\nu)^2 = M_W^2, \tag{6.1}$$

resulting in two possible solutions for the Z component of the neutrino momentum:

$$p_{Z_\nu} = \frac{\lambda \pm \sqrt{\delta}}{2}, \tag{6.2}$$

with:

$$\lambda = 2\beta \frac{p_{Z_l}}{E_l^2 - p_{Z_l}^2}; \tag{6.3}$$

$$\delta = \lambda^2 - 4\gamma; \tag{6.4}$$

$$\gamma = -\frac{\beta^2 - E_l^2(p_{X_\nu}^2 + p_{Y_\nu}^2)}{E_l^2 - p_{Z_l}^2}; \tag{6.5}$$

$$\beta = \alpha + p_{X_\nu} p_{X_l} + p_{Y_\nu} p_{Y_l}; \tag{6.6}$$

$$\alpha = \frac{1}{2}(M_W^2 - M_l^2). \tag{6.7}$$

The chosen solution is the one giving the smallest difference between the reconstructed heavy quark mass for the leptonic and hadronic side of the decays (see Section 6.5 for the details on the invariant mass reconstruction). In the case no real solution exists, the neutrino pseudorapidity is set equal to that of the lepton, since in the kinematic regime of interest the decay products of the W boson tend to be collinear.

6.3 Event selection

After preselection, as was explained in Section 5.2, events with ≥6 jets and ≥3 b-tagged jets are rejected in order to keep the $T\bar{T} \to Wb + X$ and the $T\bar{T} \to Ht + X$ analyses

Selection	Signal Region	Requirements
Preselection		One electron or muon $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV, $E_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{T}} > 60$ GeV ≥ 4 jets, ≥ 1 b-tagged jets orthogonality cut: reject events with ≥ 6 jets and ≥ 3 b-tagged jets
Loose selection	SR0	Preselection
	SR1	+ ≥ 1 W_{had} candidates
	SR2	+ $H_{\mathrm{T}} > 800$ GeV
	SR3	+ $p_{\mathrm{T}}(b_1) > 160$ GeV
	SR4	+ $p_{\mathrm{T}}(b_2) > 80$ GeV
	SR5 (≡Loose)	+ $\Delta R(\ell, \nu) < 1.2$
Tight selection	SR5	Loose selection
	SR6	+ $\min\Delta R(\ell, b) > 1.4$
	SR7 (≡Tight)	+ $\min\Delta R(W_{\mathrm{had}}, b) > 1.4$

Table 6.1: Summary of event selection requirements.

orthogonal. Once all the objects taking part to the event are identified, the selections summarized in Table 6.1 are applied with the aim of rejecting as much Standard Model background as possible. As can be seen in the table, two final selections referred to as Loose and Tight are defined, the latter being a subset of the former with two very restrictive cuts applied on top of it. The "Signal Region" (SR) label is applied to the selections progressively reaching the Tight selection.

After reconstructing at least one W_{had} candidate (SR1), a variable called H_{T} is defined as the scalar sum of the lepton p_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of the four highest-p_{T} jets. This H_{T} distribution peaks at $\sim 2m_T$ for signal events and is therefore very useful to discriminate between the signal and the background, as clearly shown in Figure 6.2(d). A cut at $H_{\mathrm{T}} > 800$ GeV (SR2) is chosen as particularly efficient in rejecting background while keeping most signal events. Considering then the boost inflicted to the bottom quarks from the heavy vector-like top decay, cuts on the b-tagged jets are defined. Figures 6.3(a) and 6.3(b) show the transverse momentum distributions of the highest-p_{T} b-jet candidate (b_1) and the next-to-highest-p_{T} b-jet candidate (b_2). The cuts chosen are $p_{\mathrm{T}}(b_1) > 160$ GeV (SR3) and $p_{\mathrm{T}}(b_2) > 80$ GeV (SR4). To further discriminate between signal and background, the angular separation between the lepton and the reconstructed neutrino (shown in Figure 6.3(c)) is required to satisfy $\Delta R(\ell, \nu) < 1.2$, thus defining the Loose selection (SR5). In the Tight selection two further cuts on the minimum angular separation between the lepton and a b-tagged, $\min(\Delta R(\ell, b_{1,2}))$ (see Figure 6.4(a)), and the minimum angular separation between the W_{had} candidate and a b-tagged jet, $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ (see Figure 6.4(b)) are defined as $\min(\Delta R(\ell, b_{1,2})) > 1.4$ (SR6) and $\min(\Delta R(W_{\mathrm{had}}, b_{1,2})) > 1.4$ (SR7, i.e. Tight selection). In Table 6.2 the expected and observed yields in these SRs are reported for the electron and muon channels combined. In Appendix E the same numbers are reported for the electron and muon channels separately.

Tables 6.3 and 6.4 summarize the acceptance times efficiency for the Loose and Tight

	$T\bar{T}$ (600) chiral	$t\bar{t}$	non-$t\bar{t}$	Tot Bkg	Data
SR0	380 ± 7	201693 ± 282	59666 ± 349	261359 ± 449	261881 ± 512
SR1	168 ± 5	6336 ± 51	1417 ± 45	7752 ± 68	8401 ± 92
SR2	161 ± 4	1575 ± 26	564 ± 30	2138 ± 40	2359 ± 49
SR3	138 ± 4	812 ± 19	296 ± 20	1108 ± 28	1223 ± 35
SR4	106 ± 4	437 ± 14	121 ± 14	558 ± 19	598 ± 24
SR5	88 ± 3	264 ± 10	53 ± 7	317 ± 13	348 ± 19
SR6	67 ± 3	27 ± 4	22 ± 4	49 ± 5	61 ± 8
SR7	54 ± 3	10 ± 2	11 ± 3	21 ± 4	37 ± 6

Table 6.2: Number of observed events, integrated over the whole mass spectrum, compared to the Standard Model expectation for the combined e+jets and μ+jets channels in the Signal Regions (see Table 6.1 for the region definitions). The expected signal yields for a chiral fourth-generation T quark with $m_T = 600$ GeV are also shown. The quoted uncertainties are only statistical.

selections, respectively, separately for each of the allowed $T\bar{T}$decay modes that can enter the selections and as a function of m_T.

	Decay mode					
m_T (GeV)	$WbWb$	$WbZt$	$ZtZt$	$WbHt$	$ZtHt$	$HtHt$
350	0.48%	0.21%	0.11%	0.18%	0.06%	0.08%
400	0.95%	0.34%	0.13%	0.26%	0.14%	0.08%
450	1.79%	0.57%	0.24%	0.47%	0.22%	0.15%
500	2.26%	0.89%	0.36%	0.68%	0.28%	0.21%
550	3.25%	1.21%	0.52%	1.07%	0.46%	0.35%
600	3.92%	1.64%	0.60%	1.34%	0.62%	0.58%
650	4.20%	2.08%	1.08%	1.66%	0.82%	0.60%
700	5.05%	2.42%	2.10%	2.01%	1.04%	0.76%
750	5.17%	2.84%	1.46%	2.40%	1.27%	0.82%
800	5.72%	3.35%	1.67%	2.69%	1.43%	1.18%

Table 6.3: Acceptance times efficiency for different $T\bar{T}$ decay modes as a function of m_T for the Loose selection.

In Table 6.5 the background estimates for the Loose and Tight selections are presented together with the total predicted and observed yields and the expectated number of events for two signal models, the chiral and the singlet scenarios. The quoted uncertainties include both statistical and systematic contributions, which are discussed in Section 6.6. The yields predicted from the Standard Model and the observed yields are in agreement within these uncertainties. The absence of QCD multijet background from the table is due to the fact that after the $H_{\mathrm{T}} > 800$ GeV requirement, the estimate from the Matrix Method gives negative results, consistent with zero, and is therefore set to zero. It has been checked that its contribution is small ($\sim 0.2\%$ of the total) even withouth the H_{T} requirement, and is therefore considered negligible. Since in the final selection the Standard Model background contributions from W+jets, Z+jets, diboson, single-top, $t\bar{t}V$ and multi-jet events are very small, it is chosen to show these backgrounds combined into a single component called "non-

	Decay mode					
m_T (GeV)	$WbWb$	$WbZt$	$ZtZt$	$WbHt$	$ZtHt$	$HtHt$
350	0.17%	0.023%	0.011%	0.022%	0.0018%	0.0072%
400	0.46%	0.054%	0.011%	0.047%	0.023%	0.014%
450	1.11%	0.14%	0.047%	0.11%	0.027%	0.054%
500	1.38%	0.32%	0.068%	0.21%	0.065%	0.029%
550	1.92%	0.45%	0.10%	0.34%	0.12%	0.11%
600	2.45%	0.64%	0.10%	0.47%	0.18%	0.16%
650	2.46%	0.74%	0.22%	0.60%	0.21%	0.19%
700	3.21%	0.97%	0.24%	0.74%	0.27%	0.25%
750	3.16%	1.06%	0.29%	0.94%	0.34%	0.29%
800	3.57%	1.35%	0.30%	1.06%	0.38%	0.34%

Table 6.4: Acceptance times efficiency for different $T\bar{T}$decay modes as a function of m_T for the TIGHT selection.

$t\bar{t}$". In addition, in order to increase the available statistics for these non-$t\bar{t}$ backgrounds, instead of a direct cut on the b-tagging algorithm weight to select at least one heavy flavor jet, the Tag Rate Function method is used for the non-$t\bar{t}$ samples only. The method is detailed in Appendix C, together with a study to validate the consistency of this choice.

	LOOSE selection			TIGHT selection		
$t\bar{t}$	264	±	80	10	±	6
$t\bar{t}V$	5.1	±	1.8	0.5	±	0.2
W+jets	16	±	11	6	±	5
Z+jets	1.1	±	1.4	0.2	±	0.5
Single top	30	±	7	4.4	±	1.6
Dibosons	0.21	±	0.15	0.06	±	0.05
Total background	317	±	90	21	±	9
Data		348			37	
$T\bar{T}$(600 GeV)						
Chiral fourth-generation	88	±	10	54	±	7
Vector-like singlet	41	±	4	20.3	±	2.2

Table 6.5: Number of observed events, integrated over the whole mass spectrum, compared to the Standard Model expectation for the combined e+jets and μ+jets channels after the LOOSE and TIGHT selections. The expected signal yields in two different scenarios, a chiral fourth-generation T quark and a vector-like singlet T quark, assuming $m_T = 600$ GeV, are also shown. The quoted uncertainties include both statistical and systematic contributions.

6.4 Control regions

In order to check the good modeling of Standard Model background and its agreement with data a number of "signal-depleted" regions (SDR) have been studied. These regions, summarized in Table 6.6, progressively go from the preselection to the final selections re-

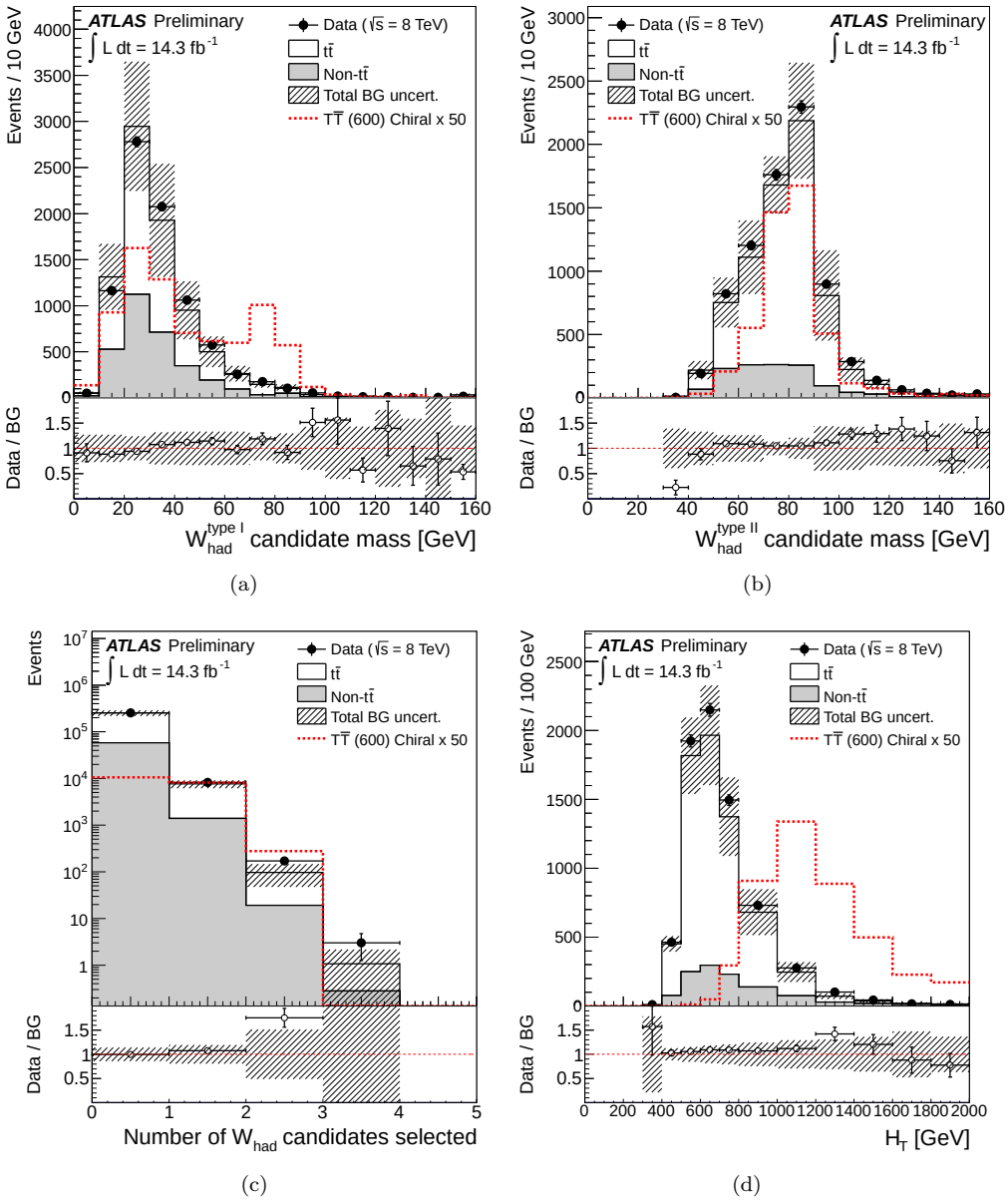


Figure 6.2: Distribution of the reconstructed mass for (a) $W_{\mathrm{had}}^{\mathrm{type\ I}}$ and (b) $W_{\mathrm{had}}^{\mathrm{type\ II}}$ candidates for the combined e+jets and μ+jets channels after preselection, prior to apply the mass window cut; distribution of (c) number of W_{had} candidates at the preselection level, and (d) H_{T} after requirement of ≥ 1 W_{had} candidate, for the combined e+jets and μ+jets channels. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV, multiplied by a factor of 50, is also shown (red dashed histogram). The lower panel shows the ratio of data to background prediction. The overflow has been added to the last bin.

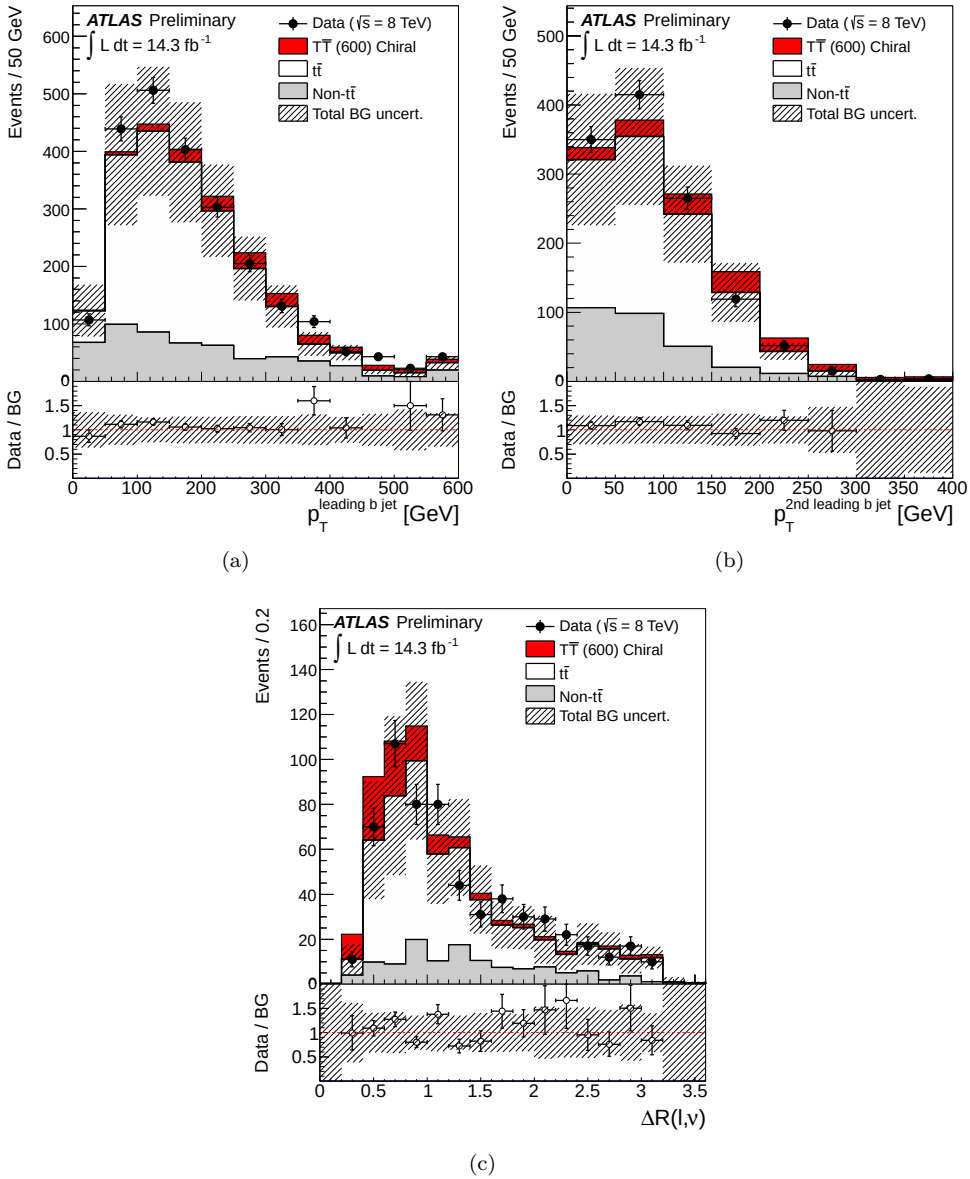
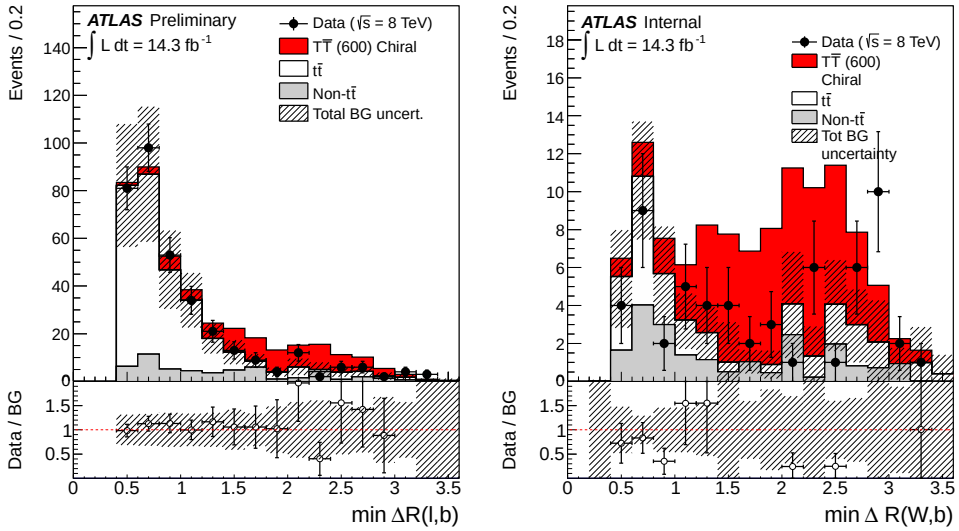


Figure 6.3: Distribution of (a) p_{T} of leading b jet candidate, $p_{\mathrm{T}}(b_1)$, and (b) p_{T} of subleading b jet candidate, $p_{\mathrm{T}}(b_2)$, for the combined e+jets and μ+jets channels after all previous selection requirements (see text for details), except for the requirements on $p_{\mathrm{T}}(b_1)$ and $p_{\mathrm{T}}(b_2)$ themselves; distribution of (c) $\Delta R(\ell, \nu)$ and (d) $\min(\Delta R(\ell, b_{1,2})) > 1.4$ for the combined e+jets and μ+jets channels after all previous selection requirements (see text for details), except for the requirements on $\Delta R(\ell, \nu)$ and $\min(\Delta R(\ell, b_{1,2})) > 1.4$ themselves. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.



(a) (b)

Figure 6.4: Distribution of (a) $\min(\Delta R(\ell, b_{1,2}))$ and (b) $\min(\Delta R(W_{\text{had}}, b_{1,2}))$ for the combined e+jets and μ+jets channels after all previous selection requirements (see text for details), except for the requirements $\min(\Delta R(\ell, b_{1,2})) > 1.4$ and $\min(\Delta R(W_{\text{had}}, b_{1,2})) > 1.4$ themselves, respectively. In the case of (a), this corresponds to the LOOSE selection. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.

jecting the signal by means of reverting one of the TIGHT selection criteria. This allows to test regions close enough to the final signal selections but still suppressing a potential signal contribution. Comparisons between data and background for selected kinematic variables in SDR4 and SDR5 are shown in Figure 6.5 and Figure 6.6 respectively. Additional plots for each of these SDRs can be found in Appendix F. In general, a reasonable agreement is found between data and Standard Model prediction, with the $t\bar{t}$ contribution being modeled by the Monte Carlo generator `MC@NLO`, which was chosen after comparison with other generators as providing the best description in the event configurations this analysis is mostly interested in.

6.5 Final discriminant: heavy quark reconstructed mass

The discriminant variable chosen to build the binned log-likelihood ratio for the statistical analysis is the reconstructed mass of the heavy vector-like top partner quark that decays into the boosted W_{had}, referred to as m_{reco}. In order to build this variable the W_{had} candidate reconstructed as described in Section 6.2 has to be paired with the associated bottom quark, to be chosen from the two b-tagged jets selected in the analysis. Another source of ambiguity comes from the computation of the longitudinal momentum of the neutrino from the W

Region	Requirements
SDR0	preselection, ≥ 1 W_{had} candidates, reversed cut: $H_{\text{T}} < 800$ GeV
SDR1	preselection cuts, ≥ 1 W_{had} candidates, reversed cut: $m_{\text{reco}} < 200$ GeV
SDR2	LOOSE selection, reversed cut: $H_{\text{T}} < 800$ GeV
SDR3	LOOSE selection, reversed cut: $p_{\text{T}} < 160$ GeV and $p_{\text{T}} < 80$ GeV
SDR4	LOOSE selection, reversed cut: $\Delta R(\ell, \nu) > 1.2$
SDR5	LOOSE selection, reversed cut: $\min(\Delta R(W_{\text{had}}, b_{1,2})) < 1.4$ and $\min(\Delta R(\ell, b_{1,2})) < 1.4$)
SDR6	LOOSE selection, reversed cut: $m_{\text{reco}} < 200$ GeV
SDR7	TIGHT selection, reversed cut: $H_{\text{T}} < 800$ GeV
SDR8	TIGHT selection, reversed cut: $p_{\text{T}} < 160$ GeV and $p_{\text{T}} < 80$ GeV
SDR9	TIGHT selection, reversed cut: $\Delta R(\ell, \nu) > 1.2$

Table 6.6: List of signal-depleted regions considered.

boson decay in the lepton channel. The various combinations for the leptonic solution and the b-tagged jets pairing are attempted and the one that in the end returns the smallest difference between m_{reco} and $m(W_{\text{lep}}, b)$ is chosen as final configuration. Figure 6.7 shows the m_{reco} distributions for chiral fourth-generation heavy top-partners with masses of $m_T = 400$ GeV, $m_T = 600$ GeV and $m_T = 800$ GeV. The distributions nicely peak around a slightly lower value of the heavy quark mass. It can also be seen from the small difference between the m_{reco} distribution in the LOOSE (Figure 6.7(a)) and TIGHT (Figure 6.7(b)) channels that the mass reconstruction is almost not affected by the two additional cuts of the TIGHT selection. The evolution of the signal peak in the m_{reco} distribution in the first four steps of the signal region selection (see Table 6.1 for the definition of the requirements) can be appreciated in Figure 6.8. In Figure 6.9 the m_{reco} distribution in the LOOSE and TIGHT final selections is shown. From Figure 6.9(a) it is clear that this variable works well in discriminating between the signal from the heavy quark and the backgrounds, with an evident low m_{reco} sideband dominated by the $t\bar{t}$ background contribution. This allows a useful check of the background modeling and, in case this channel were to be used for the statistical analysis, the sideband would help constraining the fluctuations. Figure 6.9(b) instead shows a nice and clear peak from a chiral heavy top partner with mass of 600 GeV with almost no background left after the TIGHT selection requirements have been imposed. In Appendix E the final discriminant variable is shown for data, background and signal in the selection steps through the signal region of Table 6.1. As will be explained in Section 6.7, the TIGHT selection shows the best sensitivity and therefore the distribution of Figure 6.9(b) is the one chosen to derive the final test statistics.

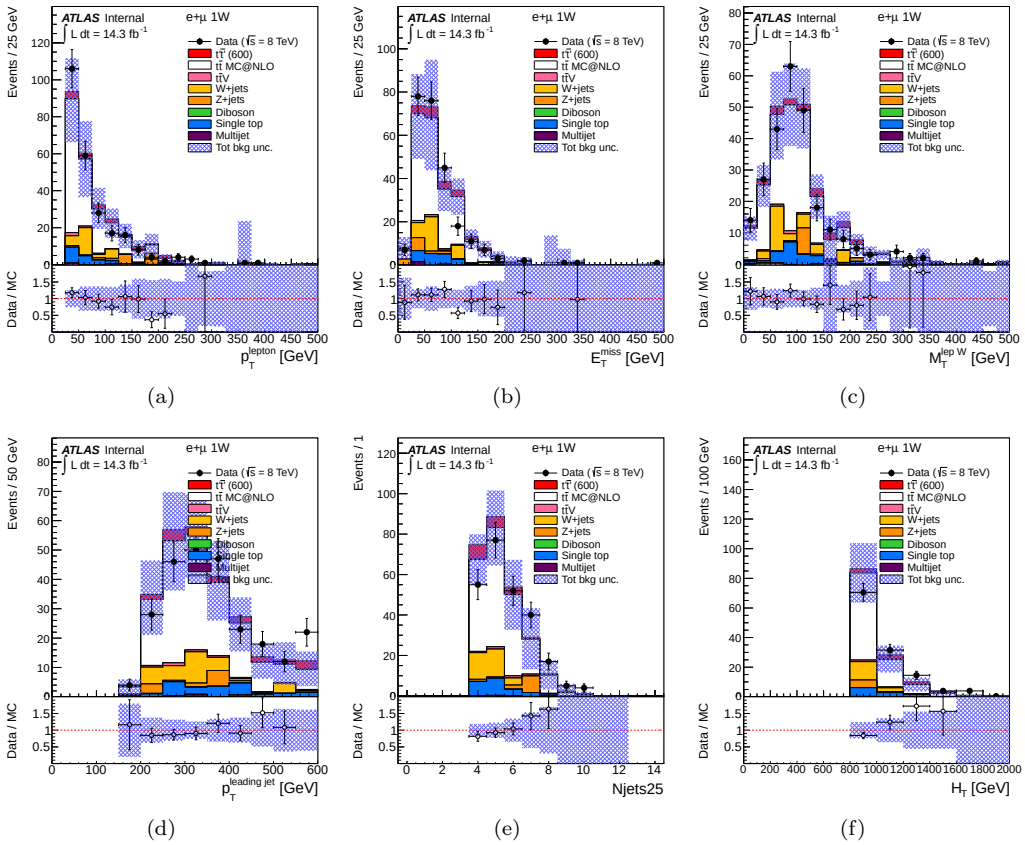


Figure 6.5: Data to Monte Carlo comparison plots for (a) lepton transverse momentum (b) missing tranverse energy, (c) transverse mass of W boson, (d) leading jet transverse momentum, (e) number of jets in the event with $p_{\mathrm{T}} > 25$ GeV, (f) H_{T} variable in SDR4.

6.6 Systematic uncertainties

The general aspects of the systematic uncertainties considered in the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses were illustrated in Section 5.5, here the traits specific to the $T\bar{T} \to Wb + X$ analysis will be described.

6.6.1 Merging of non-$t\bar{t}$ Backgrounds

The very stringent cuts defined to select signal and reject background work so well that very low statistics is left for "non-$t\bar{t}$" backgrounds (W+jets, Z+jets, dibosons, single top, $t\bar{t}V$)[2]. This can lead to problems with the treatment of some systematic uncertainties as e.g. an empty bin in the nominal case might have non-zero content in the systematically varied sample. For this reason, considering that no resonances are expected for these backgrounds, these non-$t\bar{t}$ samples are merged into a single component. Systematic uncertainties like cross-section normalization, which are specific for each background, are applied

[2]The prediction for QCD multi-jet background contribution is negative and consistent with zero and is then set to zero.

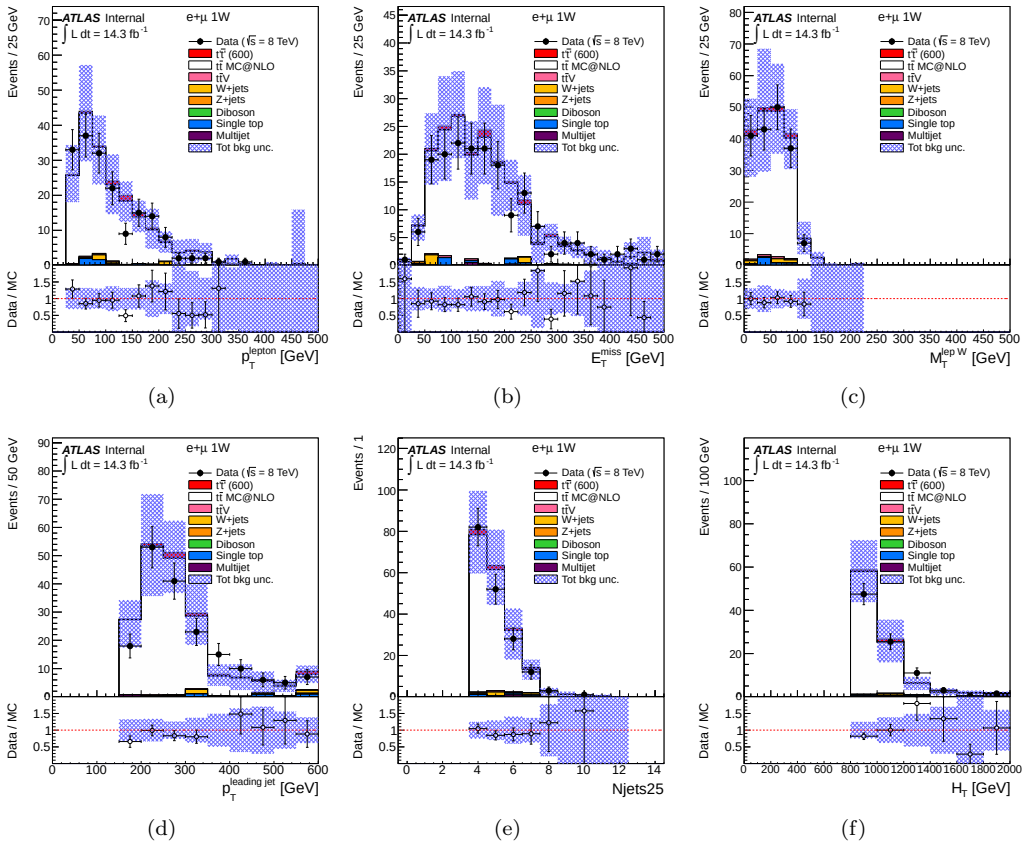


Figure 6.6: Data to Monte Carlo comparison plots for (a) lepton transverse momentum (b) missing tranverse energy, (c) transverse mass of W boson, (d) leading jet transverse momentum, (e) number of jets in the event with $p_{\mathrm{T}} > 25$ GeV, (f) H_{T} variable in SDR5.

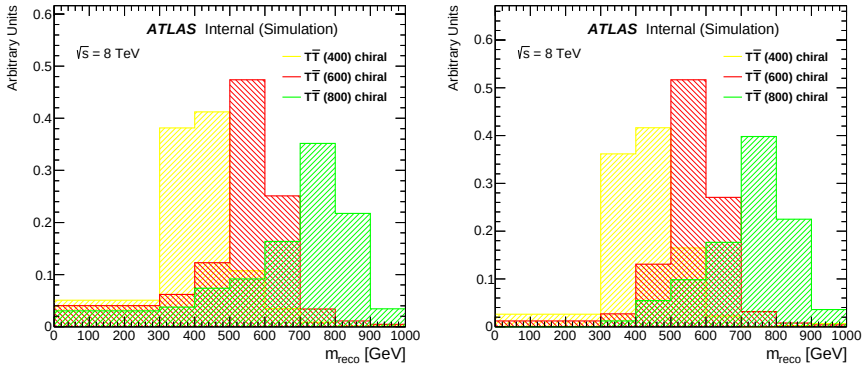


Figure 6.7: Distribution of the reconstructed mass m_{reco}, normalized to unity, for chiral fourth-generation heavy top-partners with different masses in the Loose (a) and Tight (b) channels.

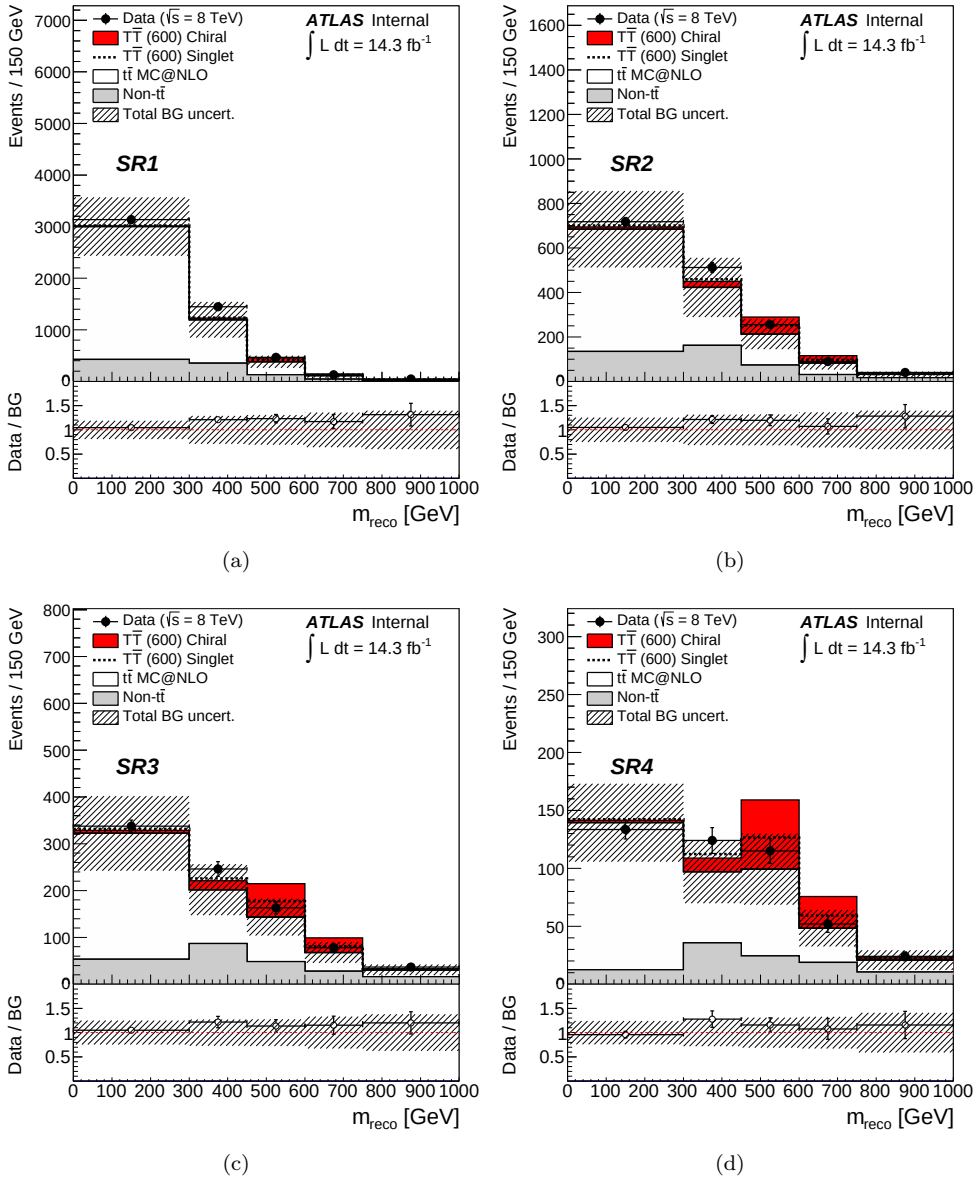
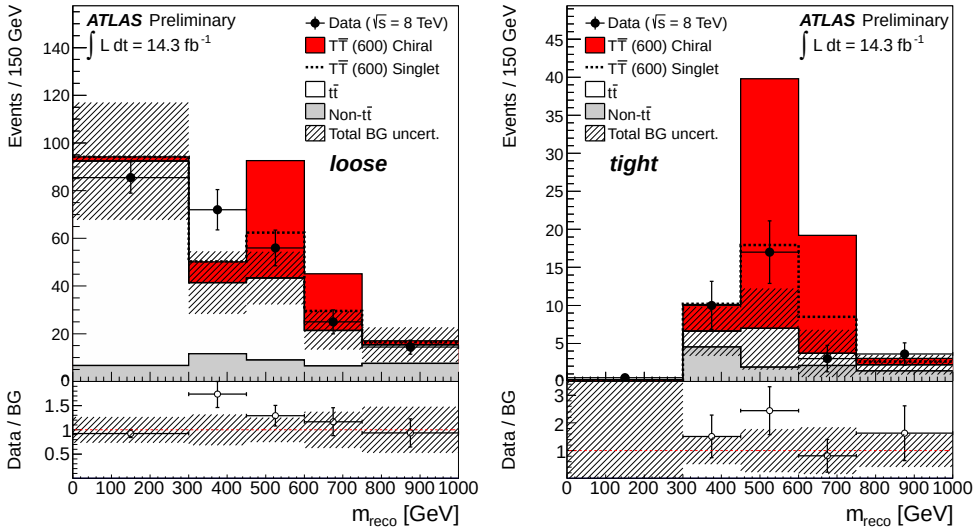


Figure 6.8: Distribution of the reconstructed mass m_{reco} in the combined electron and muon channel for the various signal regions: (a) SR1, (b) SR2, (c) SR3, (d) SR4. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.

to the single component and the nominal version of the other samples are added to obtain a varied histogram. For the non-$t\bar{t}$ background all systematic uncertainties are treated as normalization-only.



(a) (b)

Figure 6.9: Distribution of m_{reco} for the combined e+jets and μ+jets channels after the (a) Loose and (b) Tight selection. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.

6.6.2 Jet energy scale

Concerning the analysis performed in the Tight channel, we recall here (see Section 5.5.2 for details) that it was chosen to treat the systematic uncertainties from jet energy scale as a single component. Even with this treatment large unphysical bin-to-bin fluctuations are observed in the m_{reco} distribution for both $t\bar{t}$ and non-$t\bar{t}$ backgrounds in the Tight selection when the varied distributions are compared with the nominal one. In order to overcome this issue and the related noise that it would inject in the analysis, two different approaches are followed for the $t\bar{t}$ and the non-$t\bar{t}$ background: for the former, the systematic uncertainty shape is taken from the smoother (because of the higher statistical population) template of the Loose channel, while in the case of the latter, where no resonances are expected, a flat distribution is assumed and the systematic variation is taken as normalization-only.

6.6.3 Jet energy resolution

In addition to the general treatment explained in Section 5.5.2, for the $t\bar{t}$ and non-$t\bar{t}$ backgrounds the same behaviour is observed in the Tight channel as for the systematic variations from the jet energy scale uncertainties. Therefore the same treatment as previously outlined in Section 6.6.2 is applied in order to reduce the impact on the analysis from large unphysical statistical fluctuations.

6.6.4 Jet mass scale and resolution

The jet mass variable is used to identify the most energetic $W_{\text{had}}^{\text{type I}}$ candidate. Following previous studies carried out in the context of the search for $T\bar{T} \to WbWb$ analysis performed with the pp collision data at the center of mass energy of $\sqrt{s} = 7$ TeV [3], briefly summarized in Appendix D.1, given the small impact on the sensitivity of that analysis and since it was judged to be a conservative uncertainty at the time, this systematic uncertainty is neglected.

6.6.5 W/Z+jets Normalisation

The W/Z+jets cross sections as computed at the leading-order in the `Alpgen` generator framework are affected by large uncertainties. It was explained in Section 5.3.2 that the overall W+jets normalization is corrected using data-driven methods performing the estimation in each jet multiplicity bin (in the case of the $T\bar{T} \to Wb + X$ analysis, the 4 jet bin and the ≥ 5 jet bin) in order to get the best possible evaluation for the W+jets background contribution. Additional uncertainties come from the fractions of $Wb\bar{b}$, $Wc\bar{c}$ and Wc events and from their extrapolation from 2-jet events to higher jet multiplicity. The total uncertainty on the estimated W+jets normalization in the TIGHT channel, as obtained from the sum in quadrature of the above contributions, is of ∼30%.

6.6.6 $t\bar{t}$ Modelling

In order to evaluate the effect of systematic uncertainties affecting the modelling of the $t\bar{t}$ background from the `MC@NLO` generator, the following factors are considered in the $T\bar{T} \to Wb + X$ analysis: the choice of NLO event generator; the modeling of initial and final state radiation from QCD precesses; the choice of parton shower and fragmentation models.

Concerning the NLO event generator choice, the effect of systematics uncertainties on the parton-level modeling of the $t\bar{t}$ events is evaluated by comparing the distributions from two NLO Monte Carlo generators, namely `MC@NLO` (used for the nominal prediction in the $T\bar{T} \to Wb + X$ analysis) and `PowHeg`, both interfaced to `Herwig` for parton showering. This choice is based on detailed comparisons between data and Monte Carlo Standard Model backgrounds in a number of control regions, often defined starting from the LOOSE selection but with one of the cuts inverted to reject a possible signal contribution. In these studies three different $t\bar{t}$ generators, `MC@NLO`, `PowHeg` and `Alpgen` were compared. In general, it was found that data was bracketed by `MC@NLO` and `PowHeg` predictions, with `MC@NLO` providing the best description overall, which motivated its choice as the main $t\bar{t}$ generator for this analysis. In contrast, `Alpgen` was found to be the most inconsistent model with data, predicting yields above `PowHeg`, which motivated its rejection as a valid alternate $t\bar{t}$ model in the kinematic region explored by this analysis.

Differences between `MC@NLO` and `PowHeg` arise from the details on how the NLO calculation is interfaced with the parton shower, resulting in `PowHeg` predicting higher jet multiplicities than `MC@NLO`. The two samples used for the comparison have been processed through a fast simulation of the detector, and the difference in shape and normalization from these two samples of `PowHeg`+`Herwig` and `MC@NLO` is symmetrised and propagated to the `MC@NLO` fully simulated sample as systematic uncertainty. The resulting uncertainty on

the $t\bar{t}$ yield in the TIGHT selection is 48%, and is the higher contribution to the total uncertainty on the $t\bar{t}$ background. The same uncertainty in the LOOSE selection is 16%. The significant increase in the uncertainty for the TIGHT selection is mainly due to the increase in jet multiplicity in the sample simulated with `PowHeg` with respect to the one simulated with `MC@NLO`, which leads to a larger rate of misreconstructed W_{had} candidates. This makes the final cut of $\min(\Delta R(W_{\text{had}}, b_{1,2})) > 1.4$ cut less effective in discriminating signal and background than in the case of the nominal `MC@NLO`.

To assess the systematic uncertainty on the modeling of initial state (ISR) and final state radiation (FSR), dedicated $t\bar{t}$ samples are generated with `AcerMC` combined with `Pythia`, using modified `Pythia` parameters (PARP(67), PARP(64) and PARP(72)) in order to increase or reduce the amount of parton shower. The range of variation for these parameters is chosen to be consistent with existing measurements such as the gap fraction in dileptonic $t\bar{t}$ events [160] and jet shapes in QCD multijet events [161]. These samples have been processed through a fast simulation of the detector. Half the difference between these alternate samples is symmetrized and propagated to the `MC@NLO` fully-simulated sample, resulting in a systematic uncertainty on the $t\bar{t}$ yield in the TIGHT channel of 8.8%.

Parton-Shower and Fragmentation Models: Finally, systematic uncertainties on the simulation of the parton shower and fragmentation models chosen are studied by comparison of two different hadronisation models applied to the same parton level generator: `PowHeg`+`Herwig` vs `PowHeg`+`Pythia`, both samples being processed through a fast simulation of the detector. The relative difference measured between `PowHeg`+`Herwig` and `PowHeg`+`Pythia` is then symmetrised and propagated to the `MC@NLO` nominal sample, with a final result of a systematic uncertainty on the $t\bar{t}$ yield in the TIGHT selection of 25%.

6.6.7 V+jets Modelling

The effect of systematic uncertainties affecting the modeling of the V+jets bacground kinematics by the `Alpgen` generator is assessed by changing the factorization scale from the nominal choice, $Q^2 = m_{\text{W}}^2 + \sum m_{\text{T}}^2$, to $Q^2 = m_{\text{W}}^2 + p_{\text{T,W}}^2$. The resulting variation is then symmetrized.

6.6.8 Overall effect of systematic uncertainties

In Table 6.7 the final results on the effect of the different systematic uncertainties affecting the $T\bar{T} \to Wb + X$ analysis are reported. It can be seen that the main contributions come from jet energy scale and resolution, b-tagging efficiency and, in the case of the $t\bar{t}$ background, from the uncertainty on the Monte Carlo modeling of the simulated $t\bar{t}$ sample.

6.7 Results

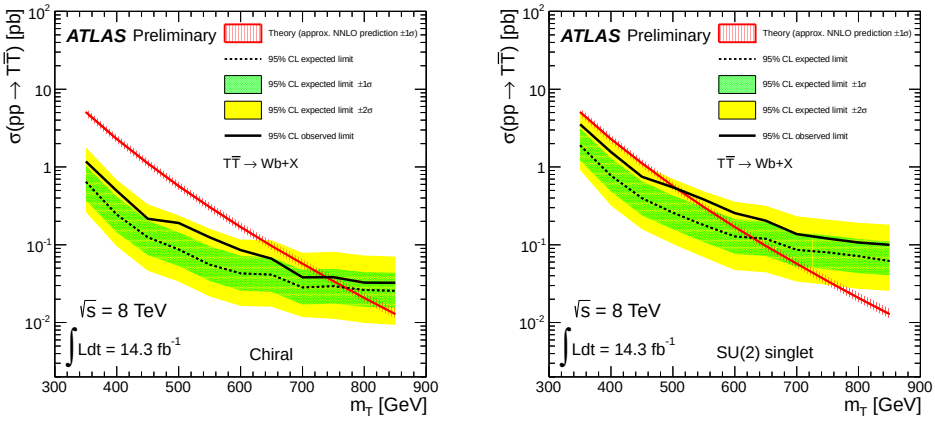
In the $T\bar{T} \to Wb+X$ analysis no significant excess of data over the expected background has been observed in the m_{reco} spectra of Figure 6.9. The observed and expected upper limits on the $T\bar{T}$ production cross section times branching fraction as a function of m_T are shown in Figure 6.10 for the two chosen benchmark scenarios, namely the chiral model with 100%

	$T\bar{T}$ (600 GeV)	$t\bar{t}$	Non-$t\bar{t}$
Uncertainties [%] affecting only the normalisation of the m_{reco} distribution:			
Luminosity	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6
Lepton trigger, reconstruction and ID efficiency	+2.0/-2.0	+2.0/-2.0	+2.0/-2.0
$t\bar{t}$ cross section	–	+10/-11	–
Uncertainties [%] affecting both normalisation and shape of the m_{reco} distribution:			
Jet energy scale	+6.6/-8.4	+15/-15	+33/-22
Jet energy resolution	+8.4/-8.4	+3.6/-3.6	+9.3/-9.3
Jet identification efficiency	+2.3/-2.7	+2.3/-2.5	+1.9/-2.6
b-quark tagging efficiency	+6.7/-7.3	+6.7/-8.9	+1.8/-2.2
c-quark tagging efficiency	+1.6/-1.6	+4.1/-4.1	+5.6/-5.6
Light-jet tagging efficiency	+0.3/-0.3	+0.7/-0.7	+2.7/-2.7
$t\bar{t}$ modelling: NLO MC generator	–	+48/-48	–
$t\bar{t}$ modelling: parton shower and fragmentation	–	+25/-25	–
$t\bar{t}$ modelling: initial and final state QCD radiation	–	+8.8/-8.8	–
W+jets normalisation	–	–	+8.9/-7.8
W+heavy-flavor fractions	–	–	+18/-19
W+jets modelling: scale variation	–	–	+11/-11
Z+jets cross section	–	–	+1.1/-1.1
Single top cross section	–	–	+1.9/-1.5
Diboson cross section	–	–	$< 0.1\%$
$t\bar{t}V$ cross section	–	–	+1.5/-1.5
Total	+14/-15	+59/-59	+42/-35

Table 6.7: List of all systematic uncertainties (in %) considered in the analysis, indicating which ones are treated as normalisation and/or shape uncertainties, with their impact on normalisation in the case of the TIGHT selection, for signal and backgrounds. The signal given here is a chiral fourth-generation T quark with mass $m_T = 600$ GeV.

BR($T \to Wb$) and the weak-isospin singlet model. These results include both statistical and systematic uncertainties, and the consistency of the data with the background prediction is assessed following the concepts presented in Section 5.6 by computing the p-value under the background-only hypothesis (1-CL_{b}) for each point of the two-dimensional plane (each point corresponding to a signal scenario) and for every heavy quark mass point considered (one two-dimensional plane is built for each m_T value). The smallest p-value found is of 0.095 for a vector-like top-partner with mass $m_T = 350$ GeV, $BR(T \to Zt) = 0.9$ and $BR(T \to Wb) = BR(T \to Ht) = 0.05$, which corresponds to a significance of 1.7 standard deviations above the background-only prediction, which is, therefore, not significant. For a chiral fourth-generation T quark, an observed (expected) 95% CL limit $m_T > 740\,(770)$ GeV is obtained for the central value of the theoretical cross section, pushing ∼100 GeV further the limit set by the previous analysis using pp collisions at the center of mass energy of 7 TeV [3]. This result can also be applied to a Y vector-like quark with electric charge of $-4/3$ and decaying into a W^- boson and a b quark. For a vector-like singlet T quark, an observed (expected) 95% CL limit $m_T > 505\,(630)$ GeV is obtained for the central value of the theoretical cross section.

Concerning the quasi-model independent strategy illustrated in Section 5.1, we recall here that a the two-dimensional plane was defined in order to perform a scan of the allowed BRs. To probe the full plane the signal samples are reweighted by the ratio of desired branching



(a) (b)

Figure 6.10: Observed (solid line) and expected (dashed line) 95% CL upper limits on the $T\bar{T}$ cross section times branching fraction for (a) a chiral fourth-generation T quark and (b) a vector-like singlet T quark as a function of the T quark mass. The surrounding shaded bands correspond to the ± 1 and ± 2 standard deviations around the expected limit. The thin red line and band show the theoretical prediction and its ± 1 standard deviation uncertainty.

ratio to the original branching ratio generated by PROTOS and the complete analysis are repeated in each point. The 95% CL exclusion limits obtained in the scan of the two-dimensional plane by varying the mixing of the three decay channel contributions for different values of m_T are shown in Figure 6.11. This plot reads as follow: taking for instance the 600 GeV vector-like top partner, a heavy quark with $BR(T \to Wb) > 0.7$ is excluded at $\geq 95\%$ CL, regardless of the value of the vector-like quark branching ratios to Ht and Zt.

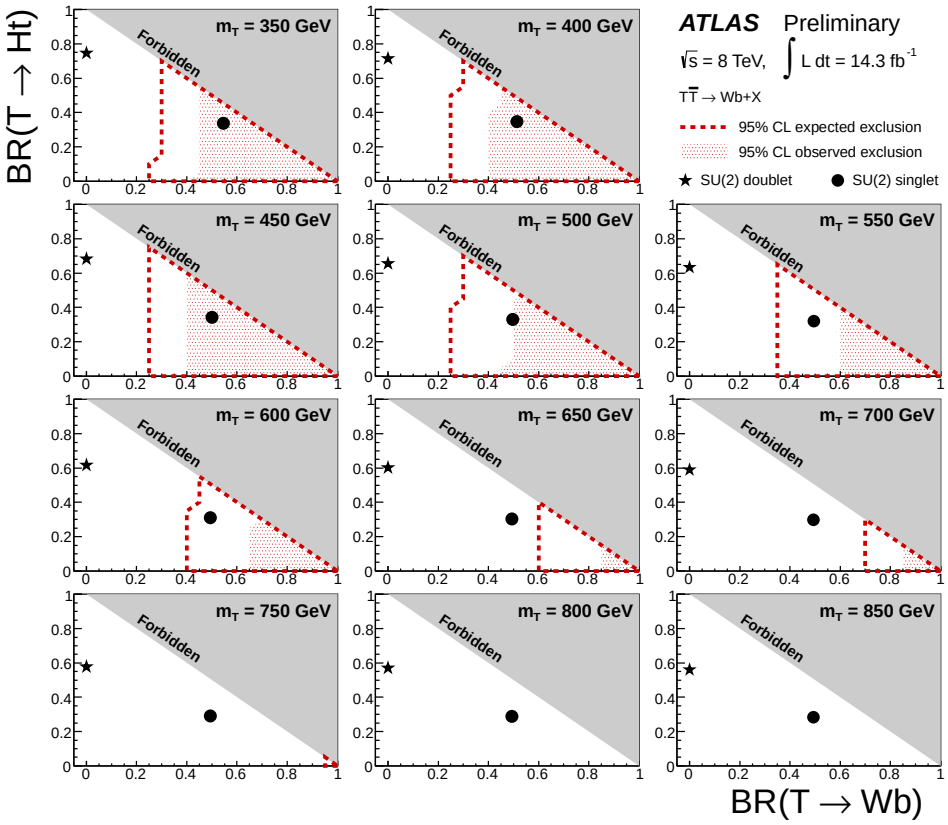


Figure 6.11: Observed (red filled area) and expected (red dashed line) 95% CL exclusion in the plane of $BR(T \to Wb)$ versus $BR(T \to Ht)$, for different values of the vector-like T quark mass. The grey (dark shaded) area corresponds to the unphysical region where the sum of branching ratios exceeds unity. The default branching ratio values from the PROTOS event generator for the weak-isospin singlet and doublet cases are shown as plain circle and star symbols, respectively. This result includes both statistical and systematic uncertainties.

Search for $T\bar{T}$ decaying to $Ht + X$

Having presented in Chapter 5 the general features common to the two searches for vector-like top partner pairs in the single lepton channel that are the object of this dissertation, we present in this chapter the search for $T\bar{T}$ pairs with at least one heavy quark decaying to one Standard Model Higgs boson with a mass $m_H = 125$ GeV and a top quark, shortly called $T\bar{T} \to Ht + X$ analysis. After outlining the chosen strategy in Section 7.1, Section 7.2 describes the event selection and the definition of the three channels used for the statistical analysis. In Section 7.3 the good modeling of Standard Model backgrounds is discussed by identifying a set of "control regions" depleted of signal contamination. Section 6.6 completes the summary of the systematic uncertainties started in Section 5.5 with the the discussion of the uncertainties treated specifically in the context of this search. The results are finally presented in Section 7.6.

7.1 Analysis strategy

Assuming a Standard Model Higgs boson with a mass of $m_H = 125$ GeV, its main decay mode is in the $H \to b\bar{b}$ channel, with a BR of about 60% (see Figure 7.1). Considering that the other competing channel is $H \to WW$ with a BR of about 20%, assuming for the sake of illustration and without loss of generality that the only lepton of the final state comes from the decay of the W from the top quark of the $T \to Ht$ channel, the signature for the heavy vector-like quark will have either high b-tagged jets multiplicity ($T \to Ht \to bbbl\nu$) or high jet multiplicity ($T \to Ht \to qqqbl\nu$). The decay of the other pair-produced heavy vector-like quark will further contribute to these multiplicities as a minimum with one b-tagged jet and two light jets if it decays in the $T \to Wb$ channel. Table 7.1 reports the number of jets and b-tagged jets in the possible final states with one heavy vector-like quark decaying to a Higgs boson and a top quark, and the other decaying in one of the three allowed channels. It can be noticed that the lowest $b$jets multiplicity is 2, and corresponds to high jet multiplicities (≥ 8). We remind here that in the $T\bar{T} \to Wb + X$ analysis an orthogonality cut is applied, rejecting events that have ≥ 6 jets and ≥ 3 b-tagged jets. This would still cause an overlap of the two analyses in the case of 2 b-tagged jets, and therefore

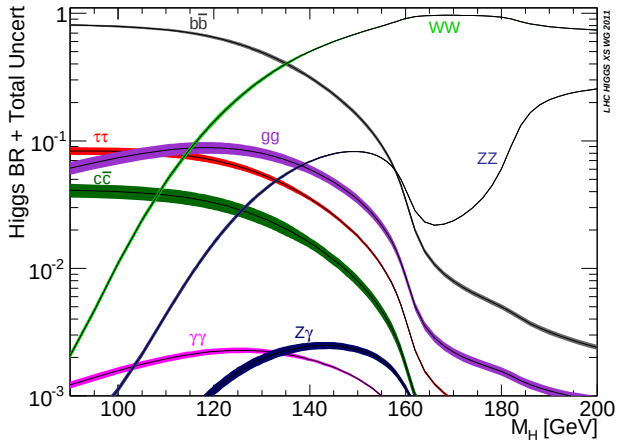


Figure 7.1: Theoretical computation of Standard Model Higgs branching ratios with uncertainties as a function of the mass of the boson, from [162].

		Wb		Ht		Zt	
		#jets	#$b$jets	#jets	#$b$jets	#jets	#$b$jets
Ht	min #$b$jets	8	2	12	2	10	2
	max #$b$jets	6	4	8	6	8	6

Table 7.1: Jets and $b$jets multiplicities in the various possible final states when at least one heavy quark decays into a Higgs boson and a top quark. The table reports the cases with the minimum and the maximum number of b-tagged jets in the final state.

in the $T\bar{T} \to Ht + X$ analysis a cut on the H_T variable opposite to the one required in the signal regions of the $T\bar{T} \to Wb + X$ analysis is applied. Since the definition of the H_T variable is slightly different in the two analyses, the cut applied results conservative (see Section 7.2 for more details). The corresponding multiplicities for the main irreducible background contribution, $t\bar{t}$ events produced in association with jets, start from a baseline of 4 jets, 2 of them b-tagged, from the $t\bar{t} \to W^+bW^-\bar{b} \to qqbbl\nu$, to which the associated jets add up. In particular, to model the $t\bar{t}$ background the `ALPGEN` Monte Carlo generator is chosen, which well describes high jet multiplicity regions and allows the separation between the heavy-flavor ("$t\bar{t}$+HF") and the light-flavor ("$t\bar{t}$+light") components of these events.

7.2 Event selection

Following the general ideas developed in Section 7.1, the natural choice to further optimize the preselection cuts presented in Section 5.2 is to require higher jets and $b$jets multiplicities. The lowest number of jets in the final state comes from the $T\bar{T} \to HtW^-\bar{b}$ channel and amounts to 6 jets. Therefore, in the final selection at least 6 jets (with the same characteristics explained in Section 7.1) are required. Then, in order to maximize the signal acceptance, considering the fact that the b-tagging efficiency is not 100%, at least 2 of these jets are required to be b-tagged. As was already hinted in Section 4.6.2, applying direct cuts

	$\geq$ 6 jets, 2 b-tags	$\geq$ 6 jets, 3 b-tags	$\geq$ 6 jets, $\geq$ 4 b-tags
$t\bar{t}$+heavy-flavour jets	1500 ± 900	900 ± 400	170 ± 70
$t\bar{t}$+light-flavour jets	9600 ± 1000	1900 ± 350	75 ± 22
W+jets	250 ± 130	50 ± 30	5 ± 3
Z+jets	50 ± 40	9 ± 6	0.5 ± 0.9
Single top	300 ± 70	75 ± 18	7 ± 3
Diboson	1.7 ± 0.6	0.3 ± 0.1	0.03 ± 0.03
$t\bar{t}V$	70 ± 20	36 ± 12	7 ± 3
$t\bar{t}H$	28 ± 4	31 ± 6	12 ± 3
Multijet	49 ± 23	1.7 ± 0.8	0.15 ± 0.06
Total background	11860 ± 260	2990 ± 210	270 ± 60
Data	11885	2922	318
$T\bar{T}$(600)	±	±	±
Vector-like doublet	4.3 ± 1.2	94 ± 7	79 ± 18
Vector-like singlet	2.3 ± 0.4	61 ± 7	36 ± 9

Table 7.2: Predicted and observed yields in the combined e+jets and μ+jets "2 b-TAGGED JETS", "3 b-TAGGED JETS" and "$\geq$4 b-TAGGED JETS" channels. The $t\bar{t}$ background prediction is after fitting to data using the full H_{T} spectrum (see text for details). Also shown is the expected $T\bar{T}$ signal in both the doublet and singlet scenarios for $m_T = 600$ GeV. The uncertainties shown are post-fit and include the effect of statistical and systematic uncertainties. The uncertainty on the total background is smaller than the sum in quadrature of the uncertainties on the individual background sources due to the anti-correlation between the $t\bar{t}$+light jets and $t\bar{t}$+heavy-flavour jets components resulting from the fit.

on the events based on the number of b-tagged jets can result in a dramatic reduction of statistic population of the Monte Carlo simulated samples. To overcome this problem the Tag Rate Function method is used for all the Monte Carlo background samples. Details on the method and on the validation checks can be found in Appendix C.

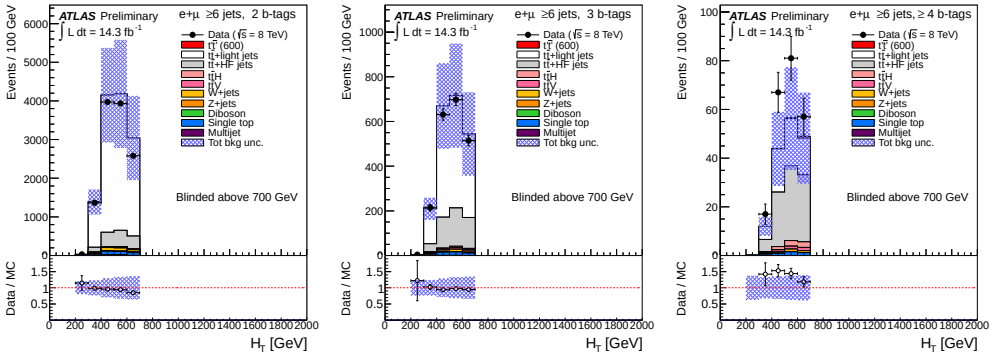
The final selection is further splitted in different $b$jet multiplicity channels in order to optimize the search sensitivity, as it can be easily understood that the higher the number of b-tagged jets are identified, the less background contamination, the better S/B ratio is obtained. Three channels are defined as follows: the "2 b-TAGGED JETS" channel, with exactly two b-tagged jets; the "3 b-TAGGED JETS" channel, with exactly three b-tagged jets; the "$\geq$4 b-TAGGED JETS" channel, with at least four b-tagged jets. In order to ensure orthogonality with the $T\bar{T} \to Wb + X$ analysis in the "2 b-TAGGED JETS" channel, the choice was done to blind this channel from signal and maintain it to exploit the high contamination from $t\bar{t}$ background contribution in the containment of systematic uncertainties. The blinding cut applied relies on the H_{T} variable, which as will be explained in Section 7.4 is the chosen final discriminant, defined as:

$$H_{\mathrm{T}} = p_{\mathrm{T}}(l) + E_{\mathrm{T}}^{\mathrm{miss}} + \sum_{j=1}^{\mathrm{Njets}} p_{\mathrm{T}}(j). \tag{7.1}$$

This definition differs from the one in the $T\bar{T} \to Wb + X$ analysis, where the scalar sum over the jets' transverse momenta run only over the four leading jets. In the $T\bar{T} \to Wb + X$ analysis signal region the events were selected with $H_{\mathrm{T}}^{4j} > 800$ GeV. In the $T\bar{T} \to Ht + X$ analysis events in the "2 b-TAGGED JETS" channel with $H_{\mathrm{T}} > 700$ GeV are rejected. This cut is somehow over-conservative, as $H_{\mathrm{T}} > H_{\mathrm{T}}^{4j}$ and the lower threshold of 700 GeV was actually chosen with in mind the previous search for fourth generation top quarks performed with ATLAS data from pp collisions at a center of mass energy of 7 TeV [3]. Table 7.2 shows the final expected and observed number of events in the three channels, after applying a rescaling procedure to the $t\bar{t}$ Monte Carlo simulated sample explained in Section 7.3.

7.3 Control regions

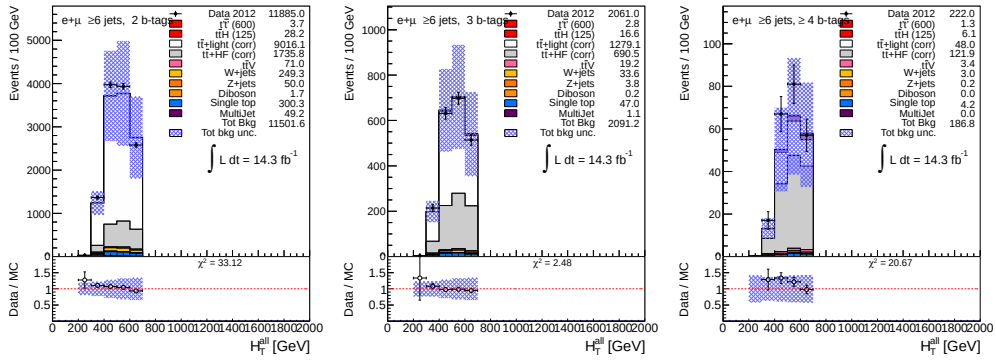
In order to assess the good modeling of the background contributions simulated with Monte Carlo generators, dedicated "control regions" depleted of signal are defined. The preselection region already provides useful control regions where the signal presence is vetoed by applying the same blinding cut used for the "2 b-TAGGED JETS" channel: $H_{\mathrm{T}} < 700$ GeV. The preselection is splitted in various control regions with different jet and $b$jets multiplicities. Appendix H reports the outcome of these checks in the electron and muon channels separately and combined. While low $b$jet multiplicity selections show rather good agreement between data and Standard Model backrounds (Figures 7.2(a) and 7.2(b)), it is observed that in the control region corresponding to the blinded "$\geq$4 b-TAGGED JETS" channel the background prediction appears systematically below the data (Figure 7.2(c)).



(a) (b) (c)

Figure 7.2: Comparison of H_{T} between data and simulation in the combined e+jets and μ+jets (a) "2 b-TAGGED JETS", (b) "3 b-TAGGED JETS" and (c) "$\geq$4 b-TAGGED JETS" channels with the requirement of $H_{\mathrm{T}} < 700$ GeV to suppress a possible signal contribution. The $t\bar{t}$+jets background is the nominal `ALPGEN` prediction before the fit to data (see text for details). Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The bottom panel displays the ratio between data and the background prediction. The shaded area represents the total background uncertainty.

In order to correct for the mismodeling of the $t\bar{t}$+jets Monte Carlo prediction from `ALPGEN` affecting in particular the heavy-flavor component, two scaling factors are introduced, one for $t\bar{t}$+light and one for $t\bar{t}$+HF, and are determined by performing a simultaneous fit to the



(a) (b) (c)

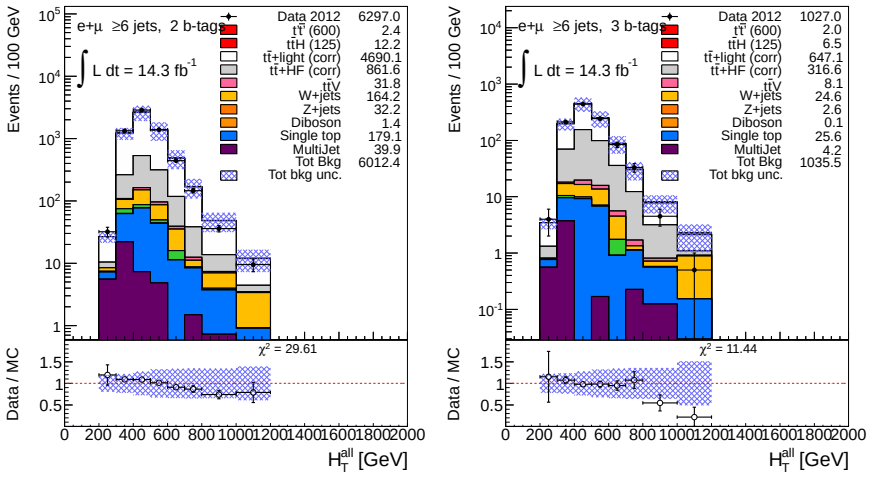
Figure 7.3: Comparison of H_{T} between data and simulation in the combined e+jets and μ+jets (a) "2 b-TAGGED JETS", (b) "3 b-TAGGED JETS" and (c) "$\geq$4 b-TAGGED JETS" channels with the requirement of $H_{\mathrm{T}} < 700$ GeV to suppress a possible signal contribution. The $t\bar{t}$+jets background is the nominal `ALPGEN` prediction after the fit to data (see text for details). Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The bottom panel displays the ratio between data and the background prediction. The shaded area represents the total background uncertainty.

data distributions of the H_{T} variable in the three analysis channels. The measured scaling factors in the blinded channels are $0.87 \pm 0.02\,(\mathrm{stat.})$ for $t\bar{t}$+light and $1.35 \pm 0.11\,(\mathrm{stat.})$ for $t\bar{t}$+HF. The H_{T} distributions in the "2 b-TAGGED JETS", "3 b-TAGGED JETS" and "$\geq$4 b-TAGGED JETS" channels corresponding obtained after applying this scaling to the `ALPGEN` prediction are shown in Figure 7.3. This rescaled prediction is taken as default from now on.

Considering that the previously defined control regions do not allow to investigate the data to Standard Model backgrounds comparison in the tails of the H_{T} distribution where the eventual signal would lay, an additional control region is defined as follows: at most two jets with $p_{\mathrm{T}} > 60$ GeV, $H_{\mathrm{T}} < 1.2$ TeV, and either 2 or 3 b-tagged jets. The ≥ 4 b-tagged jets is not considered as it still has a large signal content and too low background statistics to provide a useful cross-check. A comparison between data and simulation for the H_{T} distribution in these two additional control regions is shown in Figure 7.4. More details are given in Appendix H.4. Data are found to be in reasonable agreement with the prediction within the assigned systematic uncertainties. The last two bins of Figure 7.4(b) have too low statistics (10 and 1 data events) for their error to be properly computed.

7.4 Final discriminant: event transverse momentum

Considering the higher mass of the heavy vector-like quark, compared to its backgrounds it will transfer more momentum to its decay products. Therefore the HT variable, already introduced in Section 7.2 and defined as the scalar sum of the lepton transverse momentum, the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ and the transverse momentum of all the jets of the event, is a good choice as discriminant variable to perform the statistical analysis on. Indeed as a consequence of its construction the H_{T} distribution typically peaks at $2m_T$ and at lower



(a) (b)

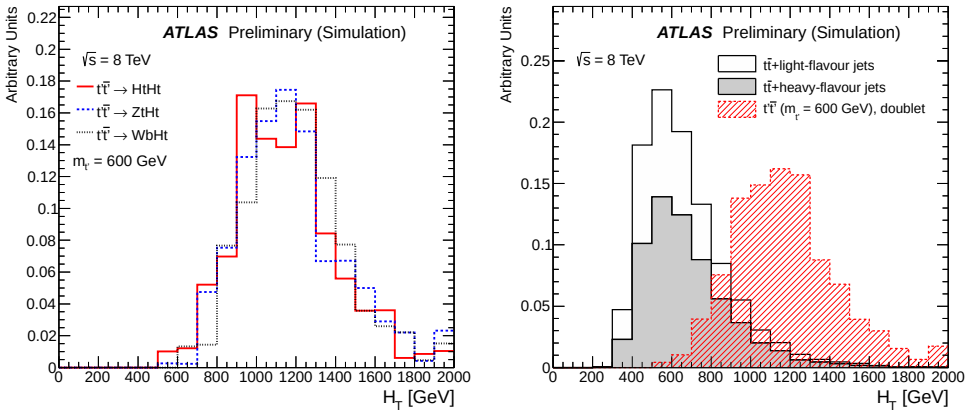
Figure 7.4: Comparison between data and simulation for the H_{T} variable in the combined e+jets and μ+jets channels with ≥ 6 jets, at most two jets with $p_{\mathrm{T}} > 60$ GeV, $H_{\mathrm{T}} < 1.2$ TeV and (a) 2 b tags, and (b) 3 b tags. The $t\bar{t}$ background prediction is after fitting to data in the $H_{\mathrm{T}} < 700$ GeV region (see text for details). Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The last bin in all figures contains the overflow. The bottom panel displays the ratio between data and background prediction. The shaded area represents the total post-fit background uncertainty.

values for the dominant $t\bar{t}$ background, as can be seen in Figure 7.5(b). Furthermore this variable is rather independent of the signal decay mode, as shown in Figure 7.5(a). Figure 7.6 instead shows a slightly harder H_{T} distribution for the $t\bar{t}$+HF background compared to the $t\bar{t}$+light one.

The three search channels, the "2 b-TAGGED JETS", "3 b-TAGGED JETS" and "$\geq$4 b-TAGGED JETS" channels, are now unblinded (except for the "2 b-TAGGED JETS", which is kept orthogonal to the $T\bar{T} \to Wb + X$ analysis signal region) and the fit to data of the two scaling factors for the $t\bar{t}$+HF and $t\bar{t}$+light background components is performed over the full range of the H_{T} variable. Consistent values are found with respect to the ones measured in the blinded regions, and are $0.88 \pm 0.02\,(\mathrm{stat.})$ and $1.21 \pm 0.08\,(\mathrm{stat.})$ for $t\bar{t}$+light jets and $t\bar{t}$+heavy-flavour jets, respectively. Figure 7.7 displays the H_{T} distribution in each of the search channels considered, showing a large S/B ratio and good discrimination in the "$\geq$4 b-TAGGED JETS" channel.

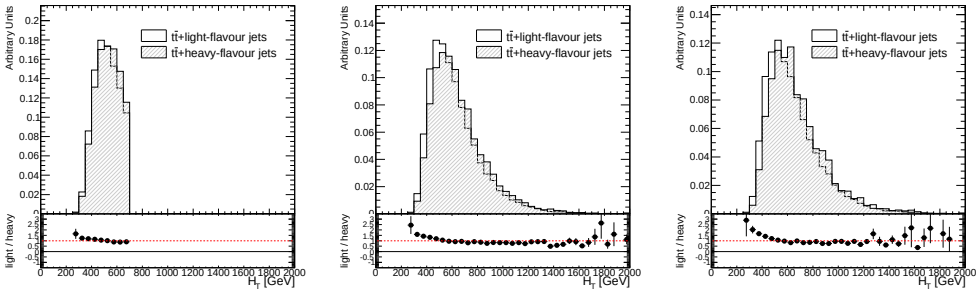
7.5 Systematic uncertainties

The general aspects of the systematic uncertainties considered in the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses were illustrated in Section 5.5, here the traits specific to the $T\bar{T} \to Ht + X$ analysis will be described.



(a) (b)

Figure 7.5: Comparison of the shape of the H_{T} distribution in simulation for (a) different $T\bar{T}$ decay modes, assuming $m_T = 600$ GeV, and (b) between $t\bar{t}$+jets background (with $t\bar{t}$+light jets and $t\bar{t}$+heavy-flavour jets shown stacked) and $T\bar{T}$ signal ($m_T = 600$ GeV) in the T doublet scenario. The selection used corresponds to the combined e+jets and μ+jets channels with ≥ 6 jets and ≥ 4 b tags. The last bin in all figures contains the overflow.

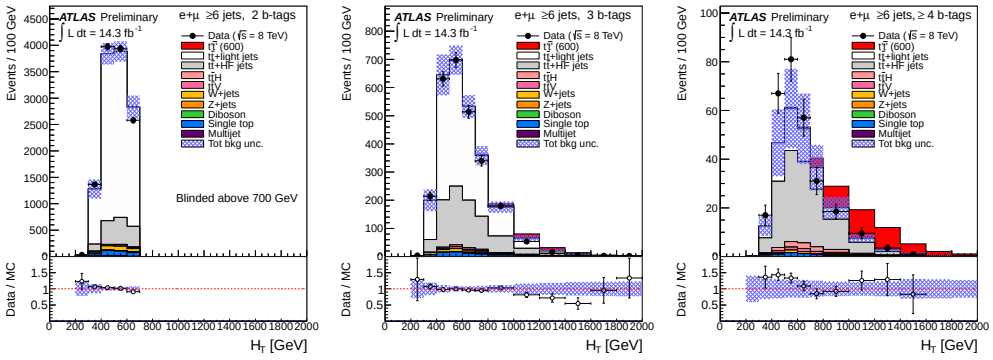


(a) (b) (c)

Figure 7.6: Comparison of the shape of the H_{T} distribution in simulation between the $t\bar{t}$+light jets and $t\bar{t}$+heavy-flavour jets backgrounds. The selection used corresponds to the combined e+jets and μ+jets channels with ≥ 6 jets and (a) 2 b tags, (b) 3 b-tags and (c) ≥ 4 b tags. The last bin in all figures contains the overflow.

7.5.1 Jet energy scale

As was seen in Table 6.7, for the $T\bar{T} \to Ht+X$ analysis the JES systematic uncertainty is split into 8 uncorrelated components, each with a different jet p_{T} and η dependence, which are treated independently. Looking at the effects of the individual sources of systematic uncertainty, it is evident that the dominant contribution comes from the first eigenvector ("BASELINE"), while the rest of the eigenvectors lead in general to very small systematic uncertainties, except for those Monte Carlo samples characterized by low statistics in the final selection channels, where unphysical fluctuations can lead to artificially large uncertainties.



(a) (b) (c)

Figure 7.7: Comparison between data and simulation for H_{T} in the combined e+jets and μ+jets (a) "2 b-TAGGED JETS" (b) "3 b-TAGGED JETS" and (c) "$\geq$4 b-TAGGED JETS" channels. The $t\bar{t}$ background prediction is after fitting to data using the full H_{T} spectrum (see text for details). Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The last bin in all figures contains the overflow. The bottom panel displays the ratio between data and background prediction. The shaded area represents the total post-fit background uncertainty.

7.5.2 Normalization of backgrounds

The W/Z+jets cross sections as computed at the leading-order in the `Alpgen` generator framework are affected by large uncertainties. It was explained in Section 5.3.2 that the overall W+jets normalization is corrected using data-driven methods performing the estimation in each jet multiplicity separately for events with exactly 4 and ≥ 5 jets in order to ensure the best possible central value for the predicted W+jets yield. An additional 24% uncertainty is assigned to the extrapolation of the data driven estimate to events with ≥ 6 jets. Additional normalization uncertainties are evaluated by varying the fractions of heavy- and light-flavor components of the W+jets background in different ways and by studying the W+heavy-flavour fractions as a function of `Alpgen` paramteters, as explained in [108]. The sum in quadrature of all the above contributions result in a total uncertainty of $\sim$50% on the estimated W+jets normalisation for events with ≥ 6 jets and ≥ 2 b tagged jets. The same uncertainty is also assigned to the Z+jets normalisation.

Systematic uncertainties on the QCD multijet background estimate via the Matrix Method receive contributions from the limited data statistics, particularly at high jet and b-tag multiplicities, as well as from the uncertainty on the method, based on the difference between estimates obtained using different control regions and from the calibration of the method using simulated QCD multijet events. The uncertainty due to the method is assessed to be 50%, which is taken as correlated across jet and b-tag multiplicity bins.

7.5.3 $t\bar{t}$+jets Modelling

A number of systematic uncertainties affecting the modelling of $t\bar{t}$+jets are considered in this analysis. Systematic uncertainties associated with the choice of factorisation and renormalisation scales in ALPGEN are considered. For the former, two different uncertainties

are taken into account.

$\mathbf{Q_{fac}}$

The factorisation scale for the hard scatter is varied by a factor of two up and down relative to the original scale, $Q^2 = \sum_{\text{partons}}(m^2 + p_{\text{T}}^2)$. Since sometimes both variations can go in the same direction, the largest of the two is taken and symmetrised.

Functional form of the factorisation scale (iqopt2)

On the other hand, the default choice for the dynamic factorisation scale, $Q^2 = \sum_{\text{partons}}(m^2 + p_{\text{T}}^2)$, is compared to an alternate choice, $Q^2 = x_1 x_2 s$. This uncertainty is significantly larger than that obtained by simply scaling the factorization scale up and down by a factor two and is symmetrised to obtain a two-sided uncertainty.

$\mathbf{k_{Tfac}}$

The renormalisation scale associated with the evaluation of α_s at each local vertex in the matrix element calculation is varied by a factor of two up and down relative to the original scale, k_{T}, between two partons. This uncertainty is only applicable for the $t\bar{t}$+light partons sample, since that is the only sample to which the MLM matching prescription [130] is applied. As a result, this uncertainty cannot be applied to the events originating from the dedicated $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ simulated samples. However, this uncertainty is applied to the subset of $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ events selected from the $t\bar{t}$+light partons MC samples after the heavy-flavour overlap removal procedure.

7.5.4 $t\bar{t}$+jets Heavy-Flavour Content

The fraction of $t\bar{t}Q\bar{Q}$ ($Q = b, c$) events relative to all $t\bar{t}jj$ events, where j denotes any parton, is one of the most important systematic uncertainties in this analysis. Currently there are no available theoretical predictions for the $t\bar{t}$+heavy-flavour fractions in pp collisions at $\sqrt{s} = 8$ TeV at NLO matched to a parton shower. In order to estimate a systematic uncertainty, the dependence of the ratio of cross sections for $t\bar{t}b\bar{b}$ over $t\bar{t}jj$ as a function of the factorisation scale choice is examined in Alpgen. These cross sections are computed requiring the extra partons to satisfy $p_{\text{T}} > 20$ GeV, $|\eta| < 2.5$ and $\Delta R(\text{j},\text{j}) > 0.4$, which are similar requirements to those used in this analysis. The ratio of cross sections is computed for the default factorisation scale choice in Alpgen, $Q^2 = \sum_{\text{partons}}(m^2 + p_{\text{T}}^2)$, which is then scaled up and down by a factor of two in a correlated way for $t\bar{t}b\bar{b}$ and $t\bar{t}jj$. The variation in the ratio of cross sections is found to be $\leq 25\%$. A similar conclusion is reached if a different dynamic scale, $Q^2 = x_1 x_2 s$, is chosen, and then scaled up and down by a factor of two. The systematic uncertainty assigned to the $t\bar{t}$+heavy-flavour fraction is 50%, conservatively doubling the variation found in the generator-level study with Alpgen.

Therefore, the fraction of $t\bar{t}Q\bar{Q}$ ($Q = b, c$) events relative to all $t\bar{t}$+jets events is varied up and down by $\pm 50\%$ (relative) with respect to the original Alpgen prediction. This uncertainty is taken to be fully correlated between the $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ fractions. The fraction of

$t\bar{t}$+light jet events is adjusted accordingly to preserve the total $t\bar{t}$ yield in each jet multiplicity bin prior to any b-tagging requirement.

7.5.5 Overall effect of systematic uncertainties

In Table 7.3 the final results on the effect of the different systematic uncertainties affecting the $T\bar{T} \to Ht + X$ analysis are reported. The overall systematic uncertainty in the "≥4 b-TAGGED JETS" channel before the two-parameter fit (*pre-fit*) on the background normalization was ∼42%, with the dominant uncertainties being from b tagging efficiency (16%), c tagging efficiency (11%), jet energy scale (11%), $t\bar{t}$ modelling (11%), $t\bar{t}$+heavy-flavour fractions (32%) and $t\bar{t}$ cross section (10%). As a result of the two-parameter fit, the total background uncertainty in this channel is reduced by about 80%. The total systematic uncertainty in the signal normalisation in the ≥ 4 b-tags channel is ∼21%, completely dominated by the uncertainty in the b tagging efficiency.

	≥ 6 jets, ≥ 4 b-tags									
	vlt	$t\bar{t}$H (125)	$t\bar{t}$-HF	$t\bar{t}$-Light	W+jets	Z+jets	Single top	Diboson	$t\bar{t}V$	Multijet
BTAGBREAK0	+0.0/-0.0	+0.2/-0.2	+0.0/-0.0	+0.1/-0.1	+0.1/-0.0	+1.0/-1.0	+0.1/-0.1	+0.3/-0.3	+0.1/-0.1	–
BTAGBREAK1	+0.7/-0.7	+0.5/-0.5	+0.5/-0.5	+0.3/-0.3	+0.1/-0.0	+0.2/-0.2	+1.1/-1.2	+2.8/-2.8	+0.4/-0.4	–
BTAGBREAK2	+0.4/-0.4	+0.2/-0.2	+0.1/-0.1	+0.1/-0.1	+0.5/-0.5	+0.8/-0.8	+0.2/-0.2	+2.4/-2.4	+0.2/-0.2	–
BTAGBREAK3	+0.9/-0.9	+0.3/-0.3	+0.2/-0.2	+0.3/-0.3	+0.5/-0.5	+0.2/-0.1	+0.1/-0.1	+1.9/-1.8	+0.1/-0.1	–
BTAGBREAK4	+1.4/-1.4	+1.8/-1.8	+1.6/-1.6	+1.5/-1.5	+0.3/-0.3	+0.1/-0.1	+1.0/-1.0	+1.8/-1.7	+1.3/-1.3	–
BTAGBREAK5	+2.7/-2.7	+1.4/-1.4	+1.0/-1.0	+0.7/-0.7	+0.4/-0.4	+2.1/-2.1	+1.7/-1.7	+0.9/-0.6	+1.2/-1.2	–
BTAGBREAK6	+1.0/-1.0	+0.7/-0.7	+0.7/-0.8	+0.5/-0.5	+1.5/-1.5	+1.4/-1.4	+0.7/-0.7	+1.3/-1.2	+0.7/-0.7	–
BTAGBREAK7	+0.1/-0.1	+4.1/-4.2	+4.1/-4.3	+3.4/-3.5	+5.5/-5.6	+1.1/-1.1	+4.4/-4.7	+0.5/-0.3	+3.4/-3.5	–
BTAGBREAK8	+20.4/-22.7	+18.7/-21.6	+15.8/-17.8	+12.2/-13.1	+13.5/-15.0	+13.0/-13.9	+15.9/-17.8	+22.0/-27.4	+16.4/-18.6	–
CTAGBREAK0	+0.3/-0.3	+0.3/-0.3	+0.9/-0.9	+1.2/-1.2	+1.5/-1.6	+0.8/-0.8	+1.1/-1.1	+0.4/-0.5	+0.6/-0.6	–
CTAGBREAK1	+0.0/-0.0	+0.2/-0.2	+0.7/-0.7	+1.0/-1.0	+0.2/-0.3	+2.7/-2.8	+0.6/-0.6	+0.0/-0.0	+0.5/-0.5	–
CTAGBREAK2	+0.1/-0.1	+0.2/-0.2	+0.5/-0.5	+0.6/-0.6	+1.7/-1.7	+2.6/-2.7	+0.5/-0.5	+0.4/-0.4	+0.5/-0.5	–
CTAGBREAK3	+1.5/-1.5	+1.9/-2.0	+5.0/-5.2	+5.3/-5.4	+6.2/-6.6	+6.1/-6.2	+3.8/-3.9	+2.7/-2.9	+4.8/-5.0	–
CTAGBREAK4	+2.4/-2.4	+3.5/-3.6	+9.4/-10.1	+10.2/-10.5	+11.8/-13.8	+16.5/-18.3	+8.3/-8.8	+3.8/-4.3	+8.5/-9.1	–
Dibosons XS	–	–	–	–	–	–	–	+5.0/-5.0	–	–
JER	+0.9/-0.9	+0.5/-0.5	+1.9/-1.9	+4.3/-4.3	+7.9/-7.9	+21.9/-21.9	+9.6/-9.6	+63.2/-63.2	+0.6/-0.6	–
JESBREAK1	+3.1/-3.1	+7.3/-7.3	+10.5/-10.5	+13.7/-13.7	+18.1/-18.1	+18.2/-18.2	+19.9/-19.9	+5.2/-5.2	+8.4/-8.4	–
JESBREAK2	+0.7/-0.7	+1.8/-1.8	+3.1/-3.1	+3.4/-3.4	+7.2/-7.2	+0.4/-0.4	+2.9/-2.9	+0.2/-0.2	+1.9/-1.9	–
JESBREAK3	+0.0/-0.0	+0.7/-0.7	+0.9/-0.9	+1.2/-1.2	+6.6/-6.6	+10.8/-10.8	+1.4/-1.4	+0.3/-0.3	+0.7/-0.7	–
JESBREAK4	+0.3/-0.3	+0.2/-0.2	+0.6/-0.6	+0.7/-0.7	+0.8/-0.8	+11.1/-11.1	+2.0/-2.0	+0.6/-0.6	+0.4/-0.4	–
JESBREAK5	+0.1/-0.1	+1.6/-1.6	+2.2/-2.2	+2.7/-2.7	+8.7/-8.7	+1.0/-1.0	+2.3/-2.3	+0.7/-0.7	+1.8/-1.8	–
JESBREAK6	+0.7/-0.7	+2.0/-2.0	+4.0/-4.0	+7.0/-7.0	+0.2/-0.2	+0.2/-0.2	+2.6/-2.6	+0.7/-0.7	+3.1/-3.1	–
JESBREAK7	+0.2/-0.2	+1.1/-1.1	+2.2/-2.2	+4.0/-4.0	+1.5/-1.5	+1.9/-1.9	+0.8/-0.8	+0.1/-0.1	+1.7/-1.7	–
JESBREAK8	+1.2/-1.2	+3.2/-3.2	+3.5/-3.5	+2.7/-2.7	+8.9/-8.9	+1.3/-1.3	+5.8/-5.8	+0.8/-0.8	+2.9/-2.9	–
JVFSF	+3.0/-3.0	+2.1/-2.8	+1.9/-2.7	+2.0/-2.9	+1.7/-2.3	+2.2/-2.7	+2.3/-3.3	+1.9/-1.8	+2.1/-2.7	–
LEPTONSYS	+2.1/-2.1	+2.1/-2.1	+2.1/-2.1	+2.1/-2.1	+2.1/-2.1	+1.5/-1.5	+2.1/-2.1	+2.1/-2.1	+2.1/-2.1	–
LTAG	+1.7/-1.8	+1.6/-1.6	+3.1/-3.2	+16.8/-17.7	+6.8/-7.3	+13.4/-14.5	+6.7/-7.0	+6.9/-7.4	+3.3/-3.3	–
Luminosity	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	+3.6/-3.6	–
QCD norm	–	–	–	–	–	–	–	–	–	+50.0/-50.0
Vjets XS jet6	–	–	–	–	+50.0/-50.0	+50.0/-50.0	–	–	–	–
ttbar iqopt2	–	–	+6.9/-6.9	+20.1/-20.1	–	–	–	–	–	–
ttbar ktfac	–	–	+7.5/-9.2	+13.8/-17.0	–	–	–	–	–	–
ttbar qfac	–	–	+0.7/-0.7	+1.6/-1.6	–	–	–	–	–	–
singleTop XS	–	–	–	–	–	–	+4.7/-3.7	–	–	–
ttH125 XS	–	+12.0/-17.0	–	–	–	–	–	–	–	–
ttbarHF	–	–	+50.0/-50.0	+13.0/-13.0	–	–	–	–	–	–
ttbarV XS	–	–	–	–	–	–	–	–	+30.0/-30.0	–
ttbar XS	–	–	+9.9/-10.7	+9.9/-10.7	–	–	–	–	–	–
Total	+21.9/-24.0	+25.2/-30.0	+57.3/-58.4	+42.0/-44.1	+60.0/-61.0	+65.2/-66.2	+31.7/-32.9	+68.2/-70.2	+37.6/-38.8	+50.0/-50.0

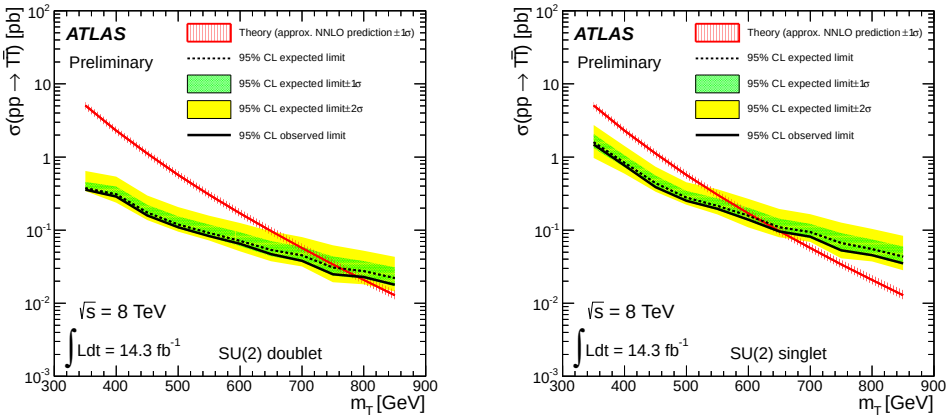
Table 7.3: List of all systematic uncertainties (in %) considered in the analysis, indicating which ones are treated as normalisation and/or shape uncertainties, with their impact on normalisation in the case of the "≥4 b-TAGGED JETS" channel, for signal and backgrounds.

7.6 Results

In the $T\bar{T} \to Ht + X$ analysis no significant excess of data over the expected background has been observed in the signal enriched "≥4 b-TAGGED JETS" channel of Figure. Since the

prediction for the $t\bar{t}$ background in the "$\geq$4 b-tagged jets" channel (the most sensitive one) is affected by large systematic uncertainties originating from $b$jet identification, jet energy calibration and physics modelling, including the fraction of $t\bar{t}$+heavy-flavour jets, two nuisance parameters are introduced. These parameters correspond to scaling factors on the overall yields of $t\bar{t}$+HF and $t\bar{t}$+light, and by allowing them to be fitted to data during the statistical analysis particularly in the "2 b-tagged jets" and "3 b-tagged jets" channels dominated by background, the degrading impact of systematic uncertainties on the sensitivity of the search is significantly reduced. The observed and expected upper limits on the $T\bar{T}$ production cross section times branching fraction as a function of m_T are shown in Figure 7.8 for the two chosen benchmark scenarios, namely the weak-isospin singlet and doublet models are chosen. These results include both statistical and systematic uncertainties, and the consistency of the data with the background prediction is assessed following the concepts presented in Section 5.6 by computing the p-value under the background-only hypothesis (1-CL_{b}) for each point of the two-dimensional plane (each point corresponding to a signal scenario) and for every heavy quark mass point considered (one two-dimensional plane is built for each m_T value). For a vector-like T quark from a weak-isospin doublet, an observed (expected) 95% CL limit $m_T > 790\,(745)$ GeV is obtained for the central value of the theoretical cross section. For a vector-like singlet T quark, an observed (expected) 95% CL limit $m_T > 640\,(615)$ GeV is obtained for the central value of the theoretical cross section.

Concerning the quasi-model independent strategy illustrated in Section 5.1, we recall here that a the two-dimensional plane was defined in order to perform a scan of the allowed BRs. To probe the full plane the signal samples are reweighted by the ratio of desired branching ratio to the original branching ratio generated by `PROTOS` and the complete analysis are repeated in each point. The 95% CL exclusion limits obtained in the scan of the two-dimensional plane by varying the mixing of the three decay channel contributions for different values of m_T are shown in Figure 6.11. This plot reads as follow: taking for instance the 600 GeV vector-like top partner, a heavy quark with $BR(T \to Ht) > 0.3$ is excluded at $\geq$ 95% CL, regardless of the value of the vector-like quark branching ratios to Wb and Zt.



(a) (b)

Figure 7.8: Observed (solid line) and expected (dashed line) 95% CL upper limits on the $t'\bar{t}'$ cross section times branching fraction for a weak-isospin (a) doublet and (b) singlet T quark as a function of the T quark mass.

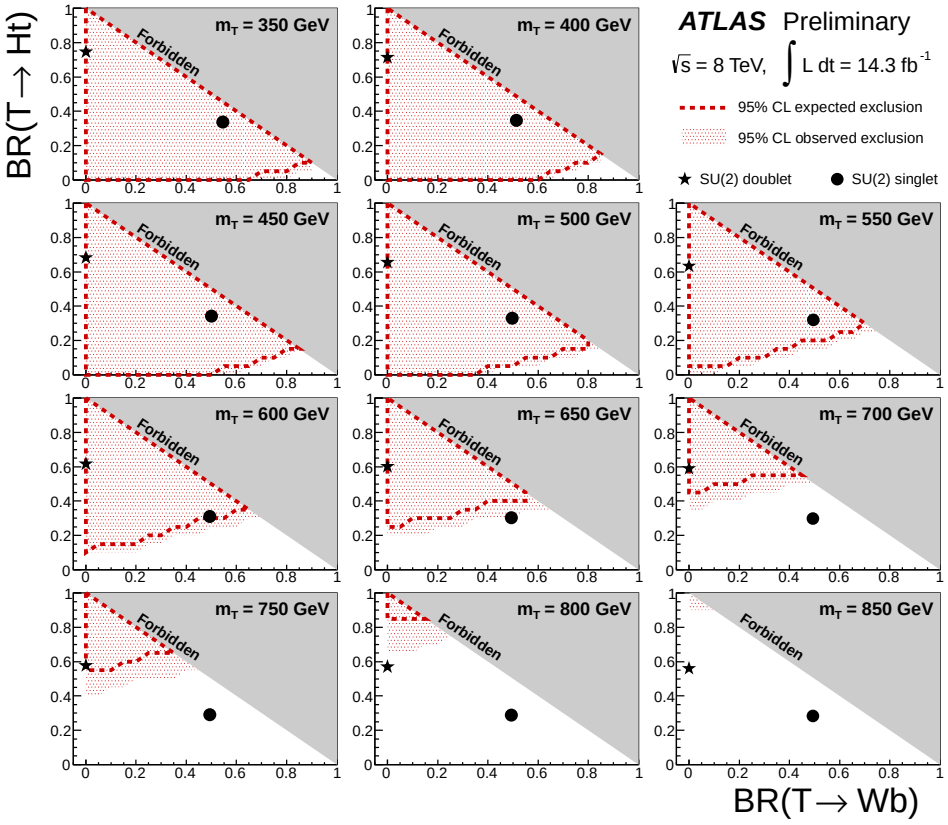


Figure 7.9: Observed (red filled area) and expected (red dashed line) 95% CL exclusion in the plane of $BR(T \to Wb)$ versus $BR(T \to Ht)$, for different values of the vector-like T quark mass.

Final results

Through Chapters 5, 6 and 7 we presented the strategy adopted for the searches of vector-like top partners in the single lepton channel and implemented into two complementary analyses: the $T\bar{T} \to Wb + X$ and the $T\bar{T} \to Ht + X$ analyses. Each of these analyses is probing a different area of the two-dimensional plane (described in Section 5.1) defined in order to perform a model-independent scan of the three possible decay channels BR mixing phase space. In this chapter we are going to illustrate how the results obtained by the individual analyses (Sections 6.7 and 7.6) perform when the search channels are combined (Section 8.1). In Section 8.2 we discuss the potential coverage of the BRs two-dimensional mixing plane by the four quasi-model independent searches for vector-like quarks performed by the Exotics group.

8.1 Combination of the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses

As the $T\bar{T} \to Wb + X$ and the $T\bar{T} \to Ht + X$ analyses do not overlap thanks to the orthogonality requirements (rejection of events with ≥ 6 jets and ≥ 3 b-tagged jets in the $T\bar{T} \to Wb + X$ analysis, rejection of events with $H_T > 700$ GeV in the "2 b-tagged jets" channel of the $T\bar{T} \to Ht + X$ analysis), it is possible to obtain a fully combined result. Figure 8.1 reports the final search channels to be used.

For the purpose of a combined statistical analysis of the four channels, choices on the systematic uncertainties treatment have to be done. Uncertainties that are common to both analyses and that are treated in the same way are considered as fully correlated and are: integrated luminosity; lepton reconstruction, identification and trigger (1 component); jet vertex fraction; jet energy resolution; b tagging (9 components); c tagging (5 components); light-jet tagging (1 component); background cross sections ($t\bar{t}$, single top, diboson, $t\bar{t}V$). For the JES uncertainty, while the $T\bar{T} \to Wb + X$ analysis considers a single component the $T\bar{T} \to Ht + X$ uses the 8 components breakdown and is then impossible to correlate the individual JES uncertainty sources one by one. Since in the $T\bar{T} \to Ht + X$ analysis the

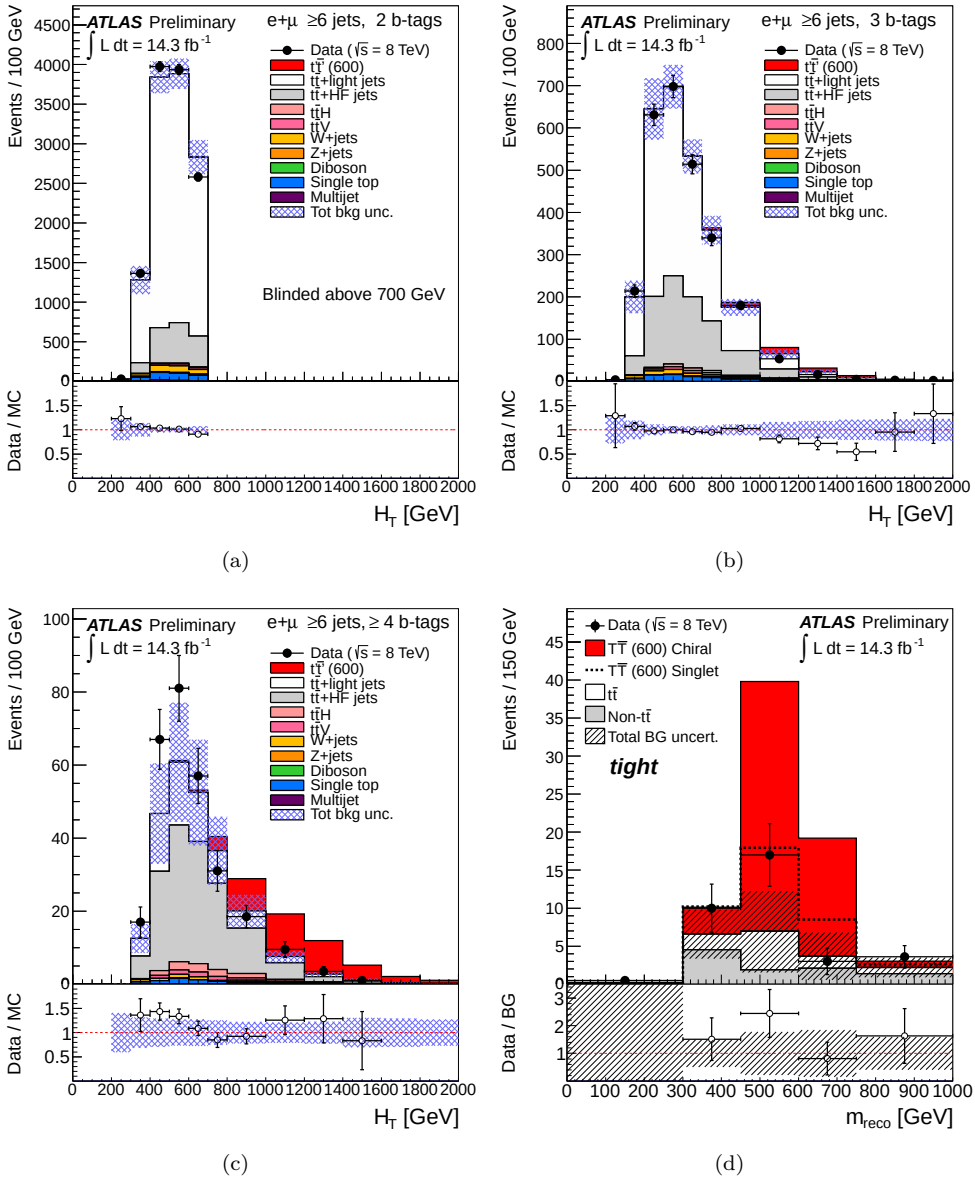


Figure 8.1: Final discriminant variables distributions in the search channels of the $T\bar{T} \to Ht + X$ analysis (H_{T} in (a) "2 b-TAGGED JETS", (b) "3 b-TAGGED JETS" and (c) "$\geq$4 b-TAGGED JETS" channels) and of the $T\bar{T} \to Wb + X$ analysis (m_{reco} in the TIGHT channel).

dominant JES uncertainty eigenvector is the BASELINE (see Section 7.5) the choice is to correlate the $T\bar{T} \to Wb + X$ analysis JES uncertainty with the BASELINE uncertainty of the $T\bar{T} \to Ht + X$ analysis. The systematic uncertainties that are not taken as correlated are: W+jets normalization, divided into 5 components in the $T\bar{T} \to Wb + X$ analysis and only 1 in the $T\bar{T} \to Ht + X$ analysis where, however, is a negligible background; $t\bar{t}$ modeling, as the two analyses use different $t\bar{t}$ Monte Carlo generators and probe very different final state kinematics region with different flavor composition.

The benchmark model chosen to show the exclusion limit as a function of the mass is the vector-like singlet T quark scenario, as it is the common benchmark model for the two analyses. In this case the $T\bar{T} \to Ht + X$ analysis performed better than the $T\bar{T} \to Wb + X$ analysis, giving an observed (expected) 95% CL limit value of $m_T > 640\,(615)$ GeV. Figure 8.2 shows the observed and expected upper limits on the $T\bar{T}$ production cross section times branching fraction as a function of m_T for a weak-isospin singlet after combination of the two analyses. The observed (expected) 95% CL limit is $m_T > 670\,(675)$ GeV for the central value of the theoretical cross section, improving by ∼30 GeV the expected sensitivity obtained by the $T\bar{T} \to Ht + X$ analysis alone.

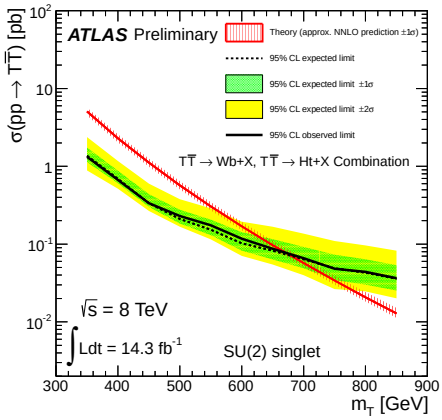


Figure 8.2: Observed (solid line) and expected (dashed line) 95% CL upper limit on the $T\bar{T}$ cross section times branching fraction for a vector-like singlet T quark as a function of the T quark mass, resulting from the combination of the $T\bar{T} \to Wb + X$ and the $T\bar{T} \to Ht + X$ analyses. The surrounding shaded bands correspond to the ±1 and ±2 standard deviations around the expected limit. The thin red line and band show the theoretical prediction and its ±1 standard deviation uncertainty.

Figure 8.3 shows the two-dimensional BR plane for different values of m_T with the resulting 95% CL exclusion limits. Comparing this result with the ones of Figures 6.11 and 7.9 it is evident the improvement resulting from the combination of the two analyses, covering a much larger area than the simple addition of two individual ones. From this picture, vector-like top partners are completely excluded no matter of the model in the mass range from 350 GeV up to 550 GeV (almost up to 600 GeV).

8.2 Potential coverage of the mixing plane

As briefly introduced in Section 5.1, two additional analyses have been performed on the same dataset as the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses to search for heavy vector-like top and bottom quarks in final states with exactly two leptons: a search in the same-sign dilepton channel [120] and a search in the opposite-charge dilepton channel [64]. It was shown that the four analyses probe different regions of the 2-dimensional mixing plane and are, hence, complementary. Even though the single lepton and the dilepton channels can be considered enough separated, it was not possible to ensure a complete orthogonality

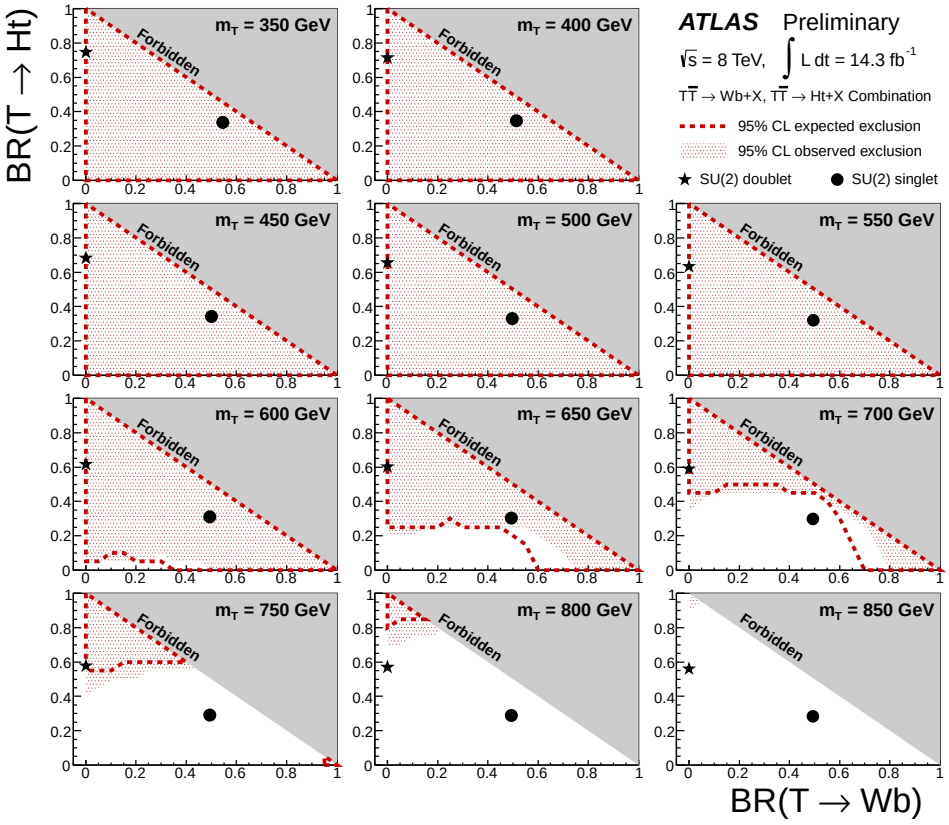


Figure 8.3: Observed (red filled area) and expected (red dashed line) 95% CL exclusion in the plane of $BR(T \to Wb)$ versus $BR(T \to Ht)$, for different values of the vector-like T quark mass. The grey (dark shaded) area corresponds to the unphysical region where the sum of branching ratios exceeds unity. The default branching ratio values from the PROTOS event generator for the weak-isospin singlet and doublet cases are shown as plain circle and star symbols, respectively. This result from the combination of the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses includes both statistical and systematic uncertainties.

between all the analyses and they have not, therefore, been combined[1]. It is however useful to visualize how the four searches contribute to the full coverage of the BRs mixing plane by observing Figure 8.4. Here the results from the four searches, each obtained independentely, are simply overlapped. This picture corresponds somehow to a "worst case scenario" of a possible real combination of the four searches, as in general the statistical analysis in the case of additional signal enriched channels would gain sensitivity. This can easily be seen comparing the combined result of the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses with the plain overlap of the separate coverages, presented in Figure 8.5. Looking e.g. at the 650 GeV mass point, it can be clearly seen how the singlet scenario, lying outside of the expected exclusion region in the case of the single $T\bar{T} \to Ht + X$ analysis, is then swallowed in the exluded area when the same analysis is combined with the $T\bar{T} \to Wb + X$ analysis, which alone does not even reach the proximities of that benchmark point.

[1]This fact is mainly due to the different timescales and frameworks at which the analyses have been developed. The $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ searches have been conceived to be combined since the early stages of their design and were performed inside the same analysis framework.

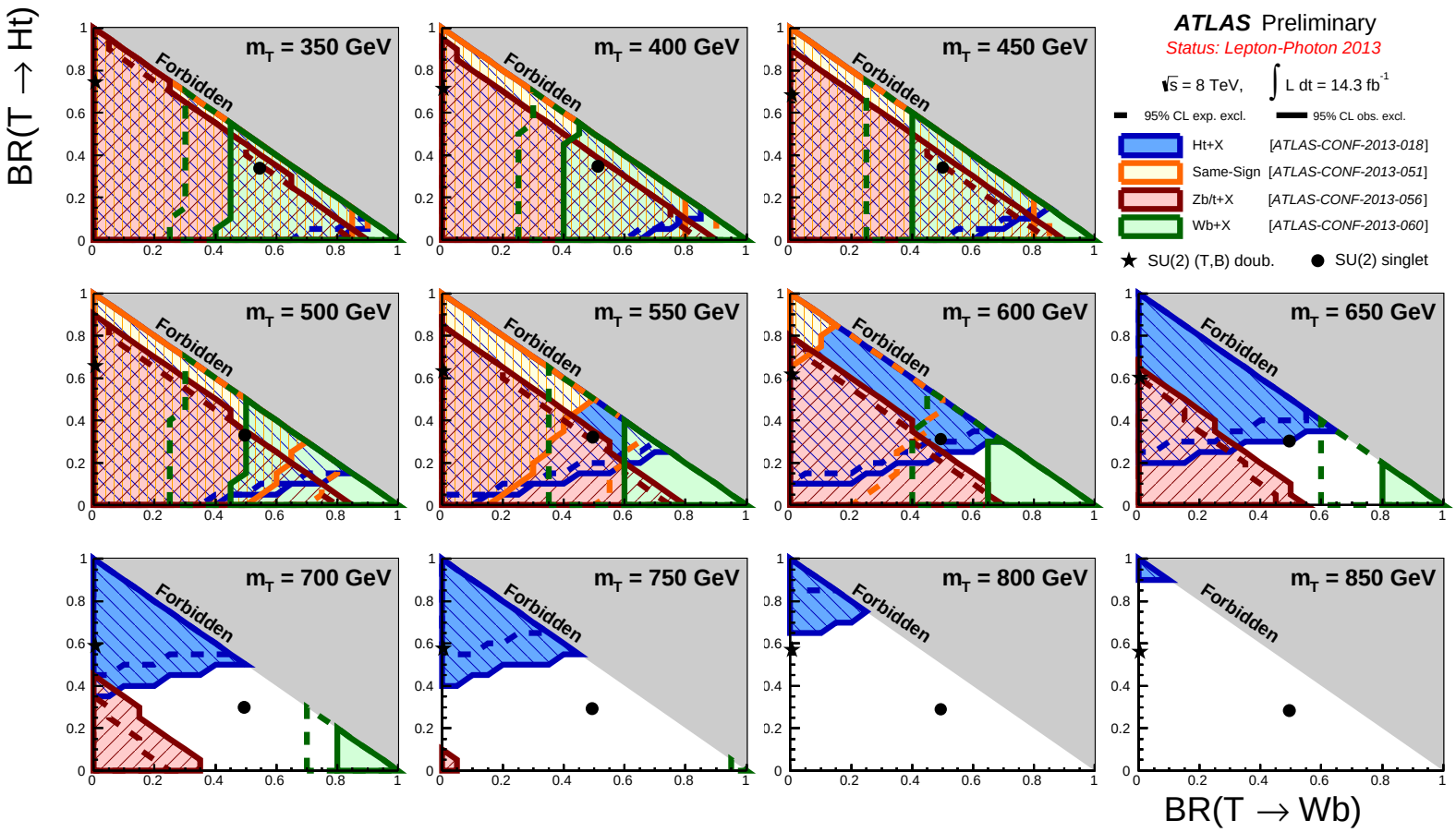


Figure 8.4: Observed (filled area) and expected (dashed line) 95% CL exclusion in the plane of $BR(T \to Wb)$ versus $BR(T \to Ht)$, for different values of the vector-like T quark mass. In blue is shown the area excluded by the single lepton $T\bar{T} \to Ht + X$ analysis; in orange is shown the area excluded by the same-sign dilepton analysis; in red is shown the area excluded by the opposite-sign dilepton $T\bar{T} \to Zt + X$ analysis; in green is shown the area excluded by the single lepton $T\bar{T} \to Wb + X$ analysis. The grey (dark shaded) area corresponds to the unphysical region where the sum of branching ratios exceeds unity. The default branching ratio values from the `PROTOS` event generator for the weak-isospin singlet and doublet cases are shown as plain circle and star symbols, respectively. This result includes both statistical and systematic uncertainties.

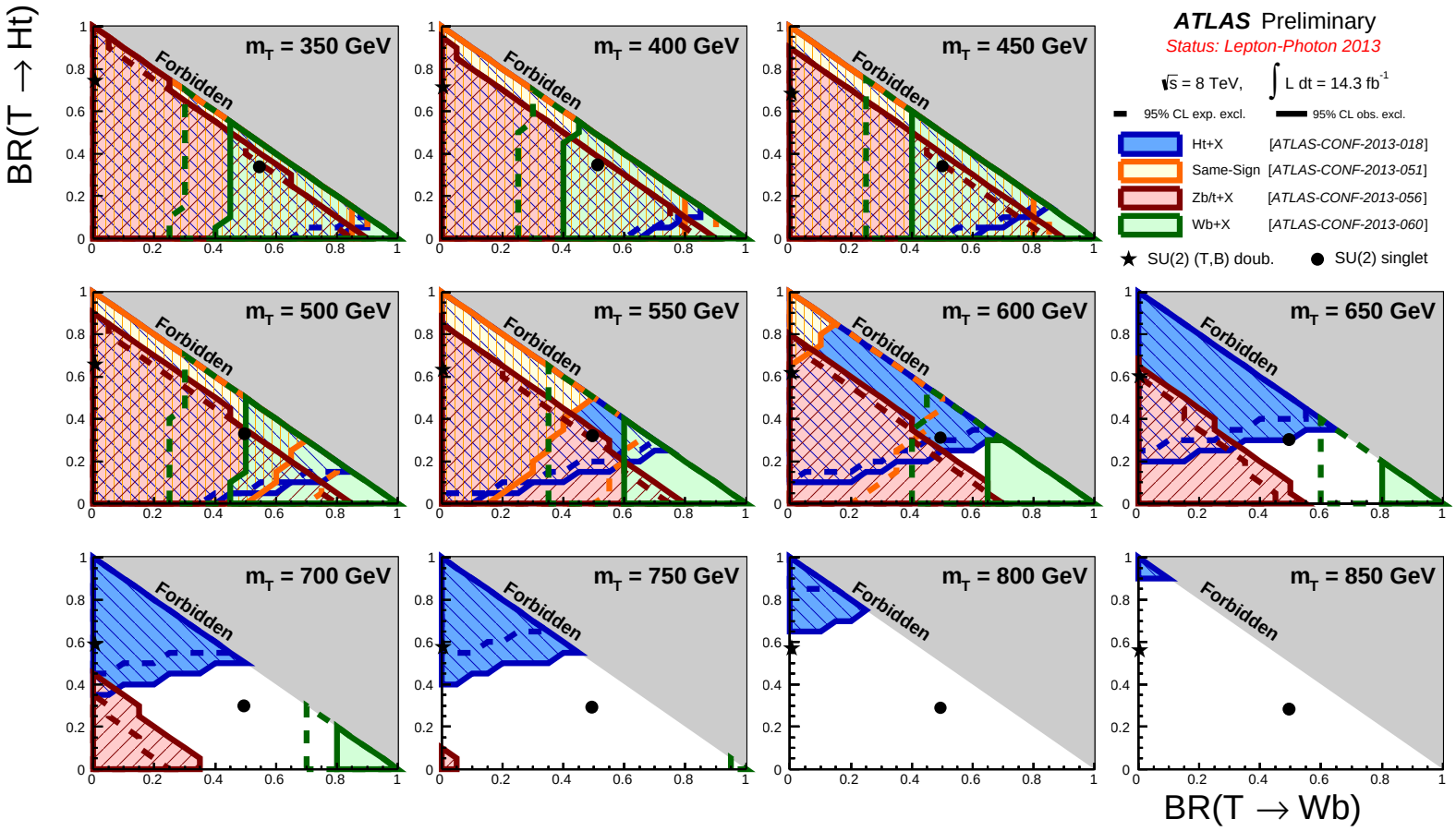


Figure 8.5: AAAAAAAAAAAAAAAA change!!! Observed (filled area) and expected (dashed line) 95% CL exclusion in the plane of $BR(T \to Wb)$ versus $BR(T \to Ht)$, for different values of the vector-like T quark mass. In blue is shown the area excluded by the single lepton $T\bar{T} \to Ht + X$ analysis alone; in green is shown the area excluded by the single lepton $T\bar{T} \to Wb + X$ analysis alone; in yellow is shown the area excluded by the combination of the $T\bar{T} \to Ht + X$ and $T\bar{T} \to Wb + X$ analyses. The grey (dark shaded) area corresponds to the unphysical region where the sum of branching ratios exceeds unity. The default branching ratio values from the `PROTOS` event generator for the weak-isospin singlet and doublet cases are shown as plain circle and star symbols, respectively. This result includes both statistical and systematic uncertainties.

Conclusions and outlook

Two quasi-model independent searches for vector-like top partners pair-produced in proton-proton collisions at a center of mass energy of 8 TeV have been presented in this dissertation. The final states considered for both analyses involve one lepton and many jets but different strategies are adopted in order to achieve sensitivities in different corners of the decay phase space. Indeed, a peculiar fact for these searches that has been stressed many times over these pages is the unpredicted nature of the heavy vector-like top partners model. As a direct consequence, the two analyses have been designed and developed to be optimized for a particular decay mode and to have orthogonal final signal rich search channels, to allow for a combined search that could exploit the specific sensitivities.

Three particular models, interesting from a theoretical point of view (but not for this more favoured than others), are considered over the two analyses: the chiral fourth-generation, with BR($T \to Wb$) = 1 for any value of the heavy quark mass; the singlet vector-like, with BR($T \to Wb$) $\sim$ 0.5 and BR($T \to Ht$) $\sim$ 0.3 for almost all values of the heavy quark mass considered in the searches; the doublet vector-like, with BR($T \to Wb$) = 0 and BR($T \to Ht$) $\in$ [0.50, 0.75] for all the values of the heavy quark mass. In the $T\bar{T} \to Wb + X$ analysis, it was possible to exclude at a 95% CL pair-produced chiral fourth-generation top partners and vector-like Y quarks with masses up to 740 GeV, and pair-produced vector-like singlet top partners with masses up to 505 GeV. In the $T\bar{T} \to Ht + X$ analysis, it was possible to exclude at a 95% CL pair-produced vector-like singlet and doublet top partners with masses up to 640 GeV and 790 GeV respectively. When the two analyses are combined, the observed exclusion limit for the only model where both analyses are sensitive, the vector-like singlet T, is pushed $\sim$30 GeV further the best result of the two, obtained by the $T\bar{T} \to Ht + X$ analysis, achieving a 95% CL exclusion of pair-produced vector-like singlet top partners with masses up to 670 GeV. While this might not look like an incredible improvement, the power of the combination of the two searches is evident looking at the coverage of the two-dimensional BR plane, where 95% CL exclusion is set for pair-produced vector-like top partners with masses up to 550 GeV independently from the model, and also the plane for the 600 GeV mass point is almost fully excluded. This strongly encourages to perform, in the future, full combination of searches for vector-like quarks.

Further improvements are also possible for the searches presented in this dissertation, without changing the core of the analysis strategies. Regarding the $T\bar{T} \to Wb + X$ analysis, it has been shown that the search channel, the TIGHT selection, suffered from poor statistical population of Monte Carlo simulated backgrounds, an issue that resulted in a degradation of the final sensitivity. The problem can easily be overcome by using larger Monte Carlo samples. Another possible improvement might come from an optimization of the cut on the total transverse momentum, the H_{T} variable, which peaks at about $H_{\mathrm{T}} \sim 2m_T$ and can therefore be tightened further for higher mass values of the heavy quark. Regarding the $T\bar{T} \to Ht + X$ analysis, one evident issue presented in this dissertation is the poor modeling of the heavy-flavor component of the main Standard Model background, the $t\bar{t}$+jets background. Since this comes from the Monte Carlo generators parameters, studies are needed to identify the corrections to this problem. Finally, the sensitivity of the search can gain much from an eventual full profiling of the nuisance parameters, thanks to the background-enriched sidebands in the three final search channels.

In any case, the masses range excluded at 95% CL up to now is getting closer and closer to the point where pair-production of vector-like quarks will start to be disfavoured with respect to single production. In this sense, while it is desirable to exploit the know-how achieved up to now with the searches for pair-produced vector-like quarks in the single lepton and dilepton channels, it is a good idea to start designing searches for single-produced vector-like quarks for LHC Phase-II.

Appendix A

Multi-jet background estimation in the single muon plus jets channel

We report in this appendix the method developed in year 2011 to better predict the contribution from multi-jet background events in analyses with a single muon and jets in the final state. For what concerns the studies presented in this appendix, data from pp collisions at center of mass energy of 7 TeV collected with the ATLAS detector in 2011 are used, with a total integrated luminosity of 689.5 pb^{-1}. Details on the status of top analyses using this dataset can be found in Reference [163]. We refer to Section 5.3.3 for the description of the general approach of the Matrix Method and will present here the improvements we made in the estimation by introducing a parametrization of the ϵ_{fake} as a function of the leption transverse momentum and of the minimum of $\Delta R(\mu, j)$. These parametrizations for the fake efficiencies are combined with the already consolidated parametrization in terms of muon pseudorapidity, which was used before.

The idea is that an increase in leading jet p_T corresponds to higher hadronic activity nearby the lepton, which results in the fact that the event will no longer satisfy the tight selection requirement of isolation $\min \Delta R(\mu, j) > 0.4$. This means a lower ϵ_{fake}. For the same reason we expect the fake efficiency to be lower for muons closer to jets. We also expect this effects to increase with the number of jets in the event, a dependence that should be entering in the ϵ_{fake} parametrization as a function of $\min \Delta R(\mu, j)$ (see Figure A.1 and A.2).

The muons are selected as "tight" if they pass the standard selection that was used in 2011 top analyses [163] (`combined` muons passing `EF_mu18` trigger and track quality and isolation cuts, summarized in Table A.1), while for the "loose" selections the requirements on calorimeter and tracker isolation (*etcone30*$<$ 4 GeV and *ptcone30*$<$ 4 GeV respectively) are dropped.

ϵ_{real} is estimated in a sample of $Z \to \mu\mu$ restricted to events with exactly 2 muons, one "tight" and one "loose" and requiring the dilepton reconstructed mass of the boson to be between 80-100 GeV. The control region to estimate ϵ_{f} is chosen as 5 GeV$< E_T^{Miss} <$ 15 GeV in order to isolate a sample enriched in QCD, then the event is required to have a "loose" muon and at least one jet. Since contamination from muons from W and Z decays is still

cut	loose	tight
track ID quality cuts	✓	✓
`combined` muon	✓	✓
`tight` muon	✓	✓
$\min\Delta R(\mu, j) > 0.4$	✓	✓
e/μ overlap removal	✓	✓
`etcone30` < 4 GeV		✓
`ptcone30` < 4 GeV		✓

Table A.1: Selection cuts for tight and loose muons. The track ID quality cuts include $p_{\mathrm{T}}(\mu) > 20$ GeV and $|\eta(\mu)| < 2.5$.

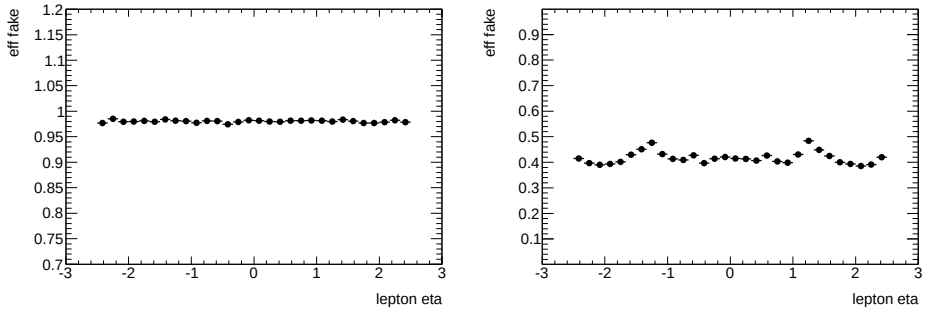


Figure A.1: Lepton η dependency of ϵ_{r} (left plot) and ϵ_{f} (right plot).

present in the low E_T^{Miss} region, we try to achieve higher purity by correcting N^{tight} and N^{loose} as

$$N_{corr}^{tight} = N^{tight} - N_{W+jets,MC}^{tight} - N_{Z+jets,MC}^{tight}) - N_{t\bar{t},MC}^{tight} \quad (A.1)$$

$$N_{corr}^{loose} = N^{loose} - N_{W+jets,MC}^{loose} - N_{Z+jets,MC}^{loose}) - N_{t\bar{t},MC}^{loose} \quad (A.2)$$

Figure A.1 and A.2 show ϵ_{r} as a function of muon η and ϵ_{f} as a function of the variables on which we parametrize, i.e. muon η, leading jet p_T and the minimum of $\Delta R(\mu, j)$. The functions used don't have a specific physical meaning and are respectively

$$f_{LJpT}(x) = p_0 + p_1/(x/100)^{p_2} \quad (A.3)$$

and

$$f_{min\Delta R}(x) = 0.5p_0(1 + TMath :: Erf((x - p_1)/(\sqrt{2}p_2))). \quad (A.4)$$

If we add the requirement of having at least one tagged jet in the event, the efficiencies change (see Table A.2). Efficiencies are computed in the same way as the untagged case but separately and requiring at least 1 b-tagged jet, where at the time the b-tagging algorithm used was `SV0` with a weight cut of 5.85. While different fake efficiencies where obtained in the parametrizations with respect to lepton η and jet p_{T}, no significant variations were observed in the dependency on the minimum $\Delta R(\mu, j)$ and the pre-tag efficiencies are used in that case to exploit the higher statistics, like it's done for ϵ_{r} where again the b-tagged events show no significant differences and more statistics for the estimation is available.

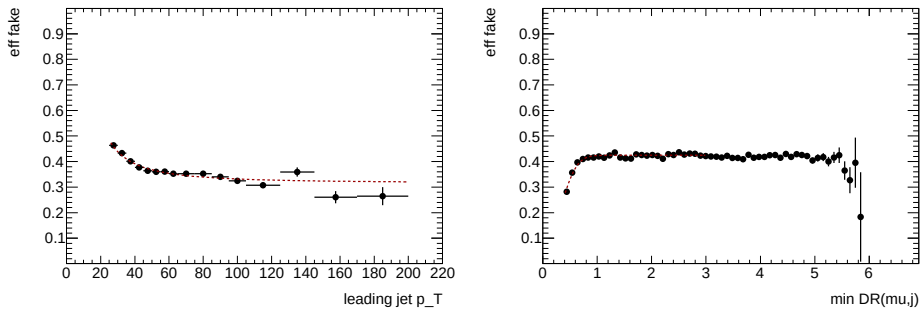


Figure A.2: Parametrization of ϵ_{f} as a function of the leading jet p_T (left plot) and of the minimum ΔR between muon and jets (right plot).

	ϵ_{f}	ϵ_{r}
untagged	0.4178 ± 0.0006	0.9805 ± 0.0003
tagged	0.353 ± 0.002	0.973 ± 0.006

Table A.2: Average values for ϵ_{f} and ϵ_{r} in untagged and tagged channels. Error is only statistical.

The two ϵ_{f} dependencies on leading jet p_T and minimum ΔR between muon and jets are then combined together to obtain a weight for the value of the fake efficiency at a given η:

$$\epsilon_{\mathrm{f}} = \epsilon_{\mathrm{f}}(\eta)\frac{f_{min\Delta R}(min\Delta R)}{<\epsilon_{\mathrm{f}}^{min\Delta R}>}\frac{f_{LJpT}(LJpT)}{<\epsilon_{\mathrm{f}}^{LJpT}>}. \tag{A.5}$$

Figure A.3 shows the agreement between data and Monte Carlo backgrounds when the QCD multi-jet backgorund estimated with this Matrix Method is considered. Here the events satisfy the full standard selection for top analyses [163] with exactly 1 jet before and after applying the triangular cut $E_T^{Miss} + m_T(W) > 60$ GeV, which basically kills QCD multi-jet contributions, and no btagging information is required. Adding the tagging selection leads to the comparison plots of Figure A.4 where the full selection (with and without the triangular cut) leave exactly 2 jets of which at least one has been tagged as a $b$jet. The variables $p_T(\mu)$, E_T^{Miss} and $m_T(W)$ are chosen to illustrate the QCD multi-jet background prediction since it is known that here the QCD multi-jet will peak at low values.

The two plots in Figure A.5 show the total amount of events for data and backgrounds in the pre-tagged and b-tagged channels in different jet multiplicity bins. The numerical values for the QCD estimate are reported in Table A.3 and Table A.4 for the pre-tagged and b-tagged case respectively.

An estimation of the systematic uncertainties on the QCD multi-jet background as derived in this Matrix Method can be evaluated considering the following sources:

1. statistical error on ϵ_{f} and ϵ_{r};
2. statistical error on the QCD estimation;
3. different control regions for the estimation of fake efficiency;
4. changes in the parametrization used.

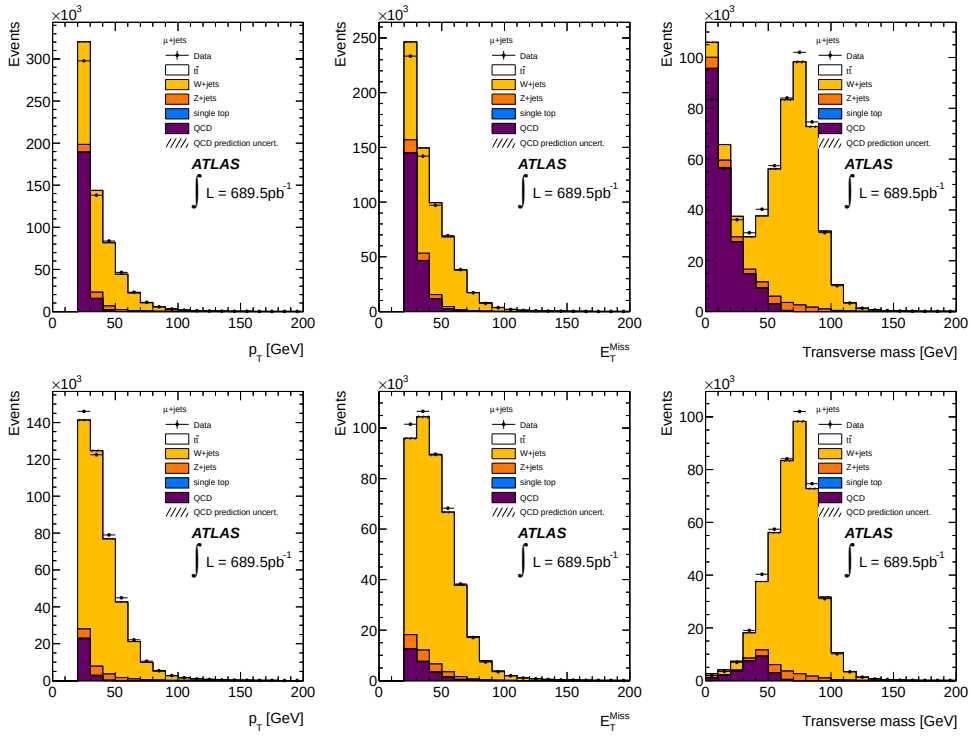


Figure A.3: Comparison plots between data and backgrounds for the muon transverse momentum (left column), missing transverse energy (central column) and the transverse mass of the W (right column). The full event selection of 1 jet exclusive with no btagging information is used without and with the triangular cut (top and bottom respectively).

	= 1 jet	= 2 jets	= 3 jets	$\geq$ 4 jets
$t\bar{t}$	304.33 ± 7.41	1320.55 ± 15.50	2709.63 ± 22.13	4702.33 ±29.49
QCD	24803.36 ± 153.57	10511.66 ± 87.94	2942.08 ± 43.72	1049.85 ±25.02
W+jets	385242.06 ± 1129.55	98826.98 ± 373.93	23614.51 ± 154.62	7419.73 ±81.07
Z+jets	17257.90 ± 63.81	5478.57 ± 35.64	1553.30 ± 18.79	592.78 ±11.28
Single top	1002.70 ± 10.77	1126.85 ± 10.55	578.51 ± 6.47	285.44 ±4.15
Total prediction	428610.34 ±1141.80	117264.61 ±386.23	31398.04 ±163.41	14050.14 ±90.62
Data	437526	112984	29135	12779

Table A.3: Yields table for the data and background samples for different jet multiplicities in the untagged full event selection.

For the points 1 and 2, the values can be taken from what already shown in Table A.2 and Tables A.3 and A.4. For what concerns point 3, we compared the results obtained in control region 5 GeV$< E_T^{Miss} <$ 15 GeV with an estimation in control region $E_T^{Miss} <$ 10 GeV, while no studies have yet been performed about point 4. Table A.5 and A.6 summarize the systematic uncertainties for different jet multiplicity in the untagged and tagged channels respectively.

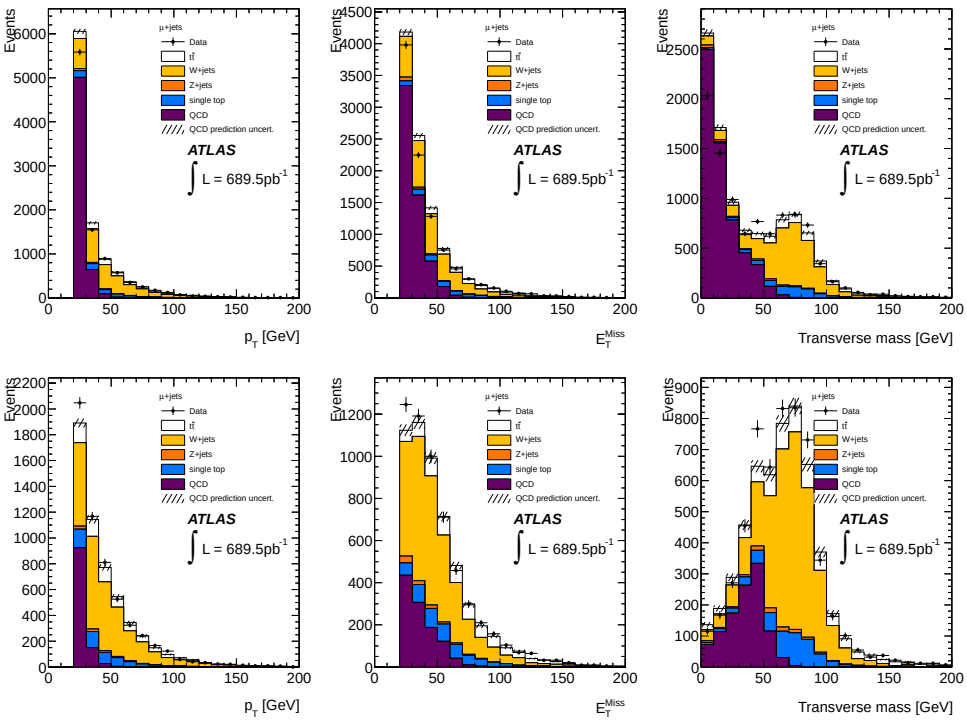


Figure A.4: Comparison plots between data and backgrounds for the muon transverse momentum (left column), missing transverse energy (central column) and the transverse mass of the W (right column). The full event selection of 2 jet exclusive with at least 1 btagged jet is used without and with the triangular cut (top and bottom respectively).

	= 1 jet	= 2 jets	= 3 jets	$\geq$ 4 jets
$t\bar{t}$	108.35 ± 4.22	689.18 ± 10.58	1659.28 ± 16.48	3185.88 ±23.19
QCD	1449.22 ± 29.45	1109.84 ± 24.17	420.08 ± 14.44	211.11 ±10.35
W+jets	4392.92 ± 77.83	3016.70 ± 56.17	1280.12 ± 38.41	612.50 ±26.04
Z+jets	104.68 ± 4.76	94.69 ± 4.56	52.10 ± 3.32	28.77 ±2.44
Single top	359.78 ± 6.16	522.23 ± 6.80	307.53 ± 4.50	159.42 ±2.96
Total prediction	6414.95 ±83.69	5432.64 ±62.60	3719.11 ±44.57	4197.68 ±36.58
Data	7243	5634	3876	4406

Table A.4: Yields table for the data and background samples for different jet multiplicities in the tagged full event selection (at least one bjet).

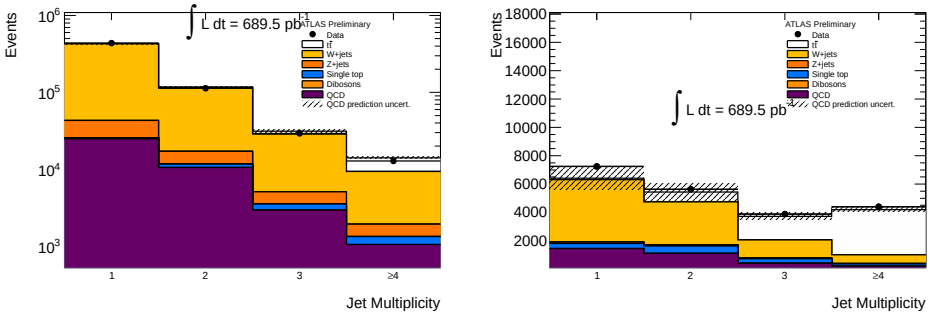


Figure A.5: Yields plots for data and backgrounds requiring full event selection (left plot) and full event selection plus at least one btagged jet (right plot) in jet multiplicity bins.

	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
ε_1			0.1%	
ε_2	0.6%	0.8%	1.5%	2.4%
ε_3	7.9%	21.2%	31.0%	41.3%

Table A.5: Systematic uncertainties on QCD estimation for different jet multiplicity in the untagged case.

	= 1 jet	= 2 jets	= 3 jets	≥ 4 jets
ε_1			0.5%	
ε_2	2.0%	2.2%	3.4%	4.9%
ε_3	6.4%	18.5%	26.5%	32.4%

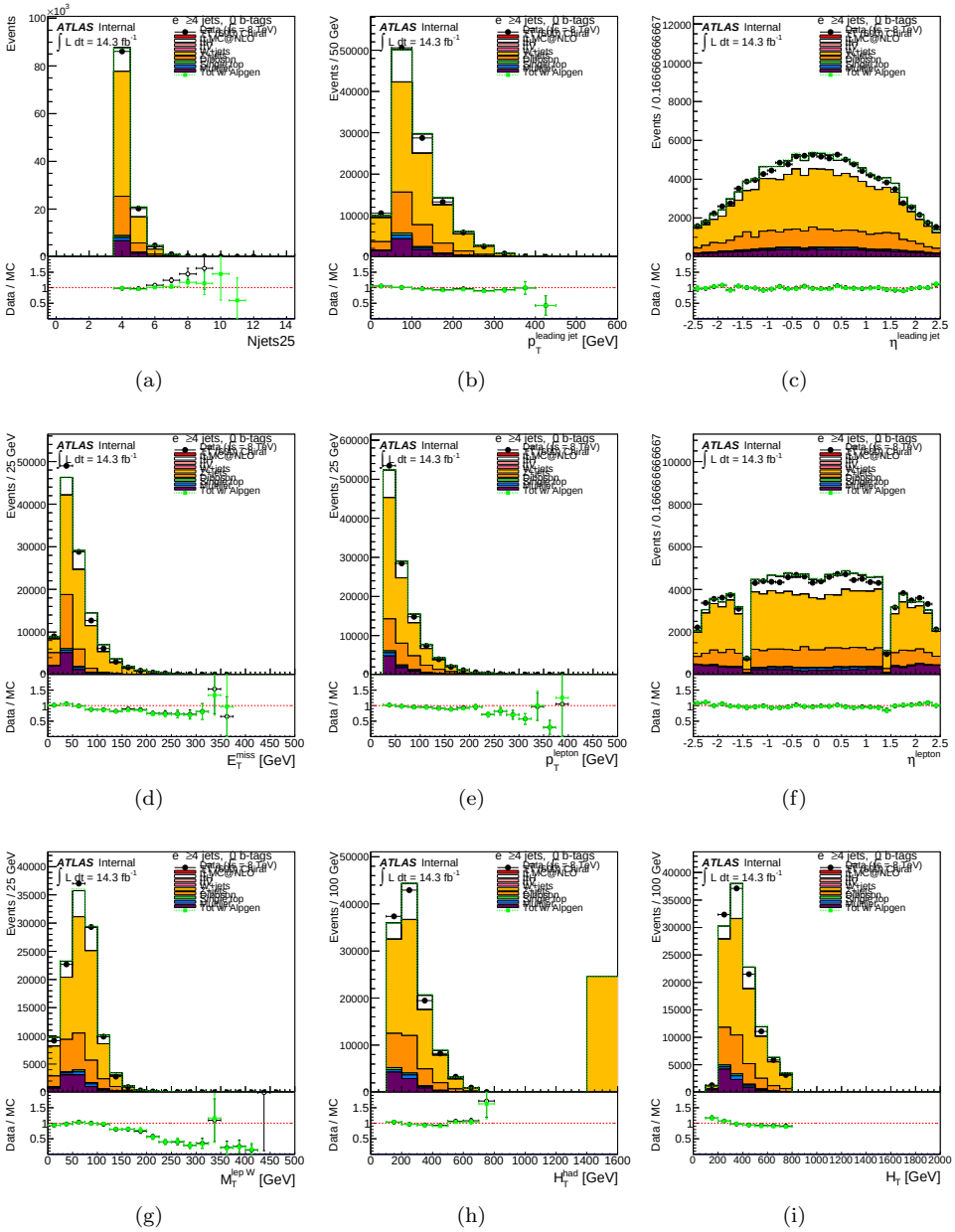
Table A.6: Systematic uncertainties on QCD estimation for different jet multiplicity in the tagged case.

Appendix B

Data to Monte Carlo comparison in the preselection region

B.1 Data to Monte Carlo comparison vetoing $b$jets

B.1.1 Electron channel



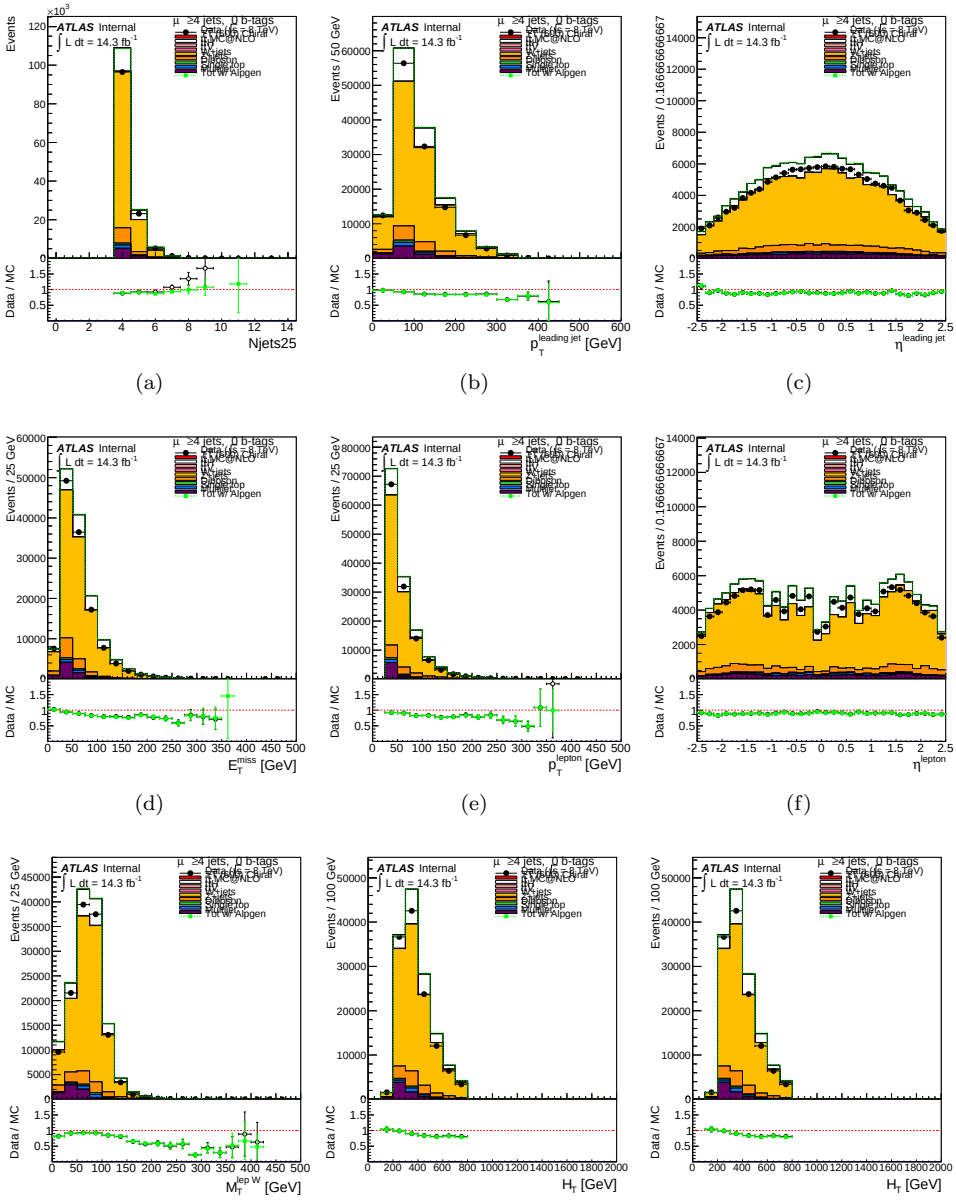
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.1

B.1.2 Muon channel



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure B.2

B.1.3 Electron+Muon channel

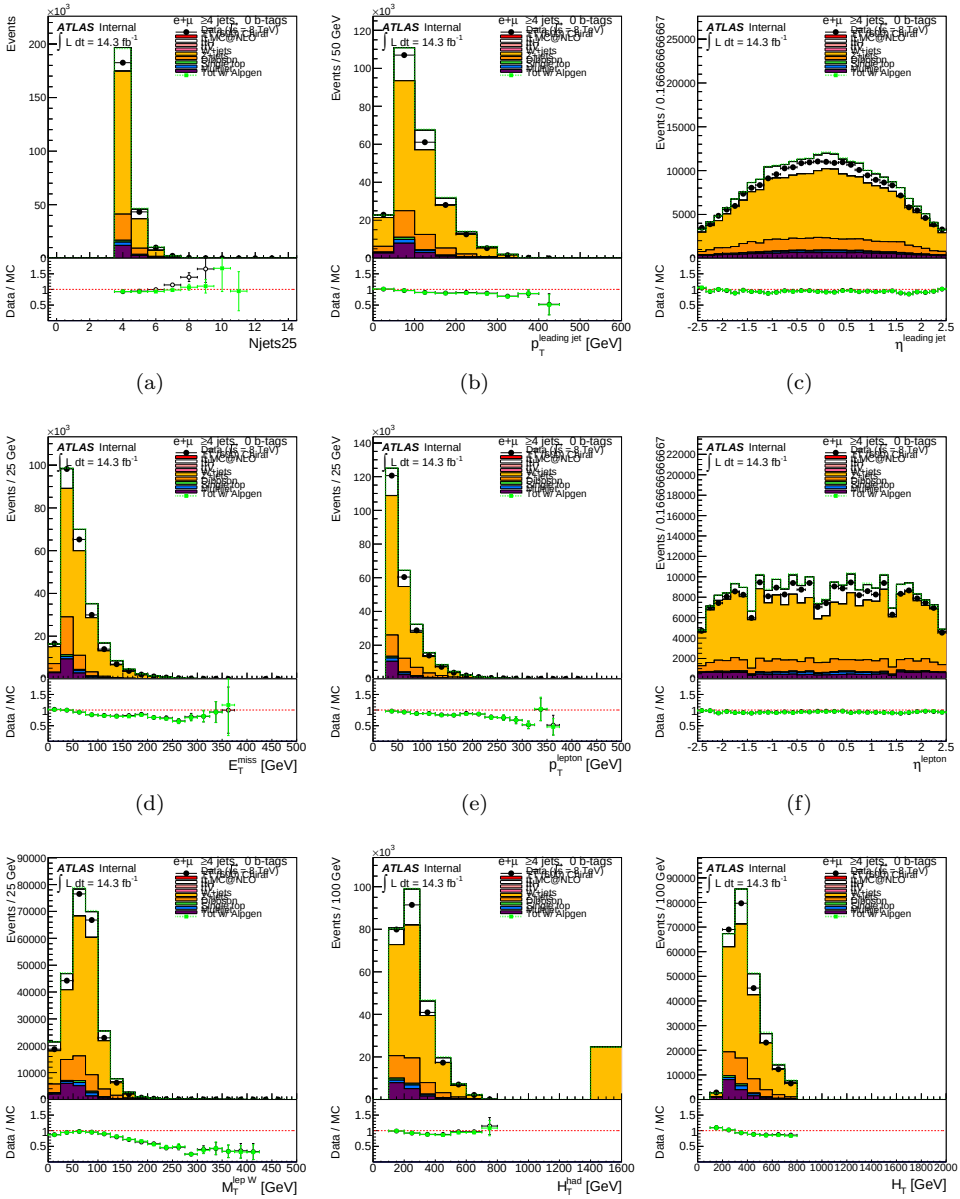


Figure B.3

B.2 Data to Monte Carlo comparison requiring at least 1 *b*jet

B.2.1 Electron channel

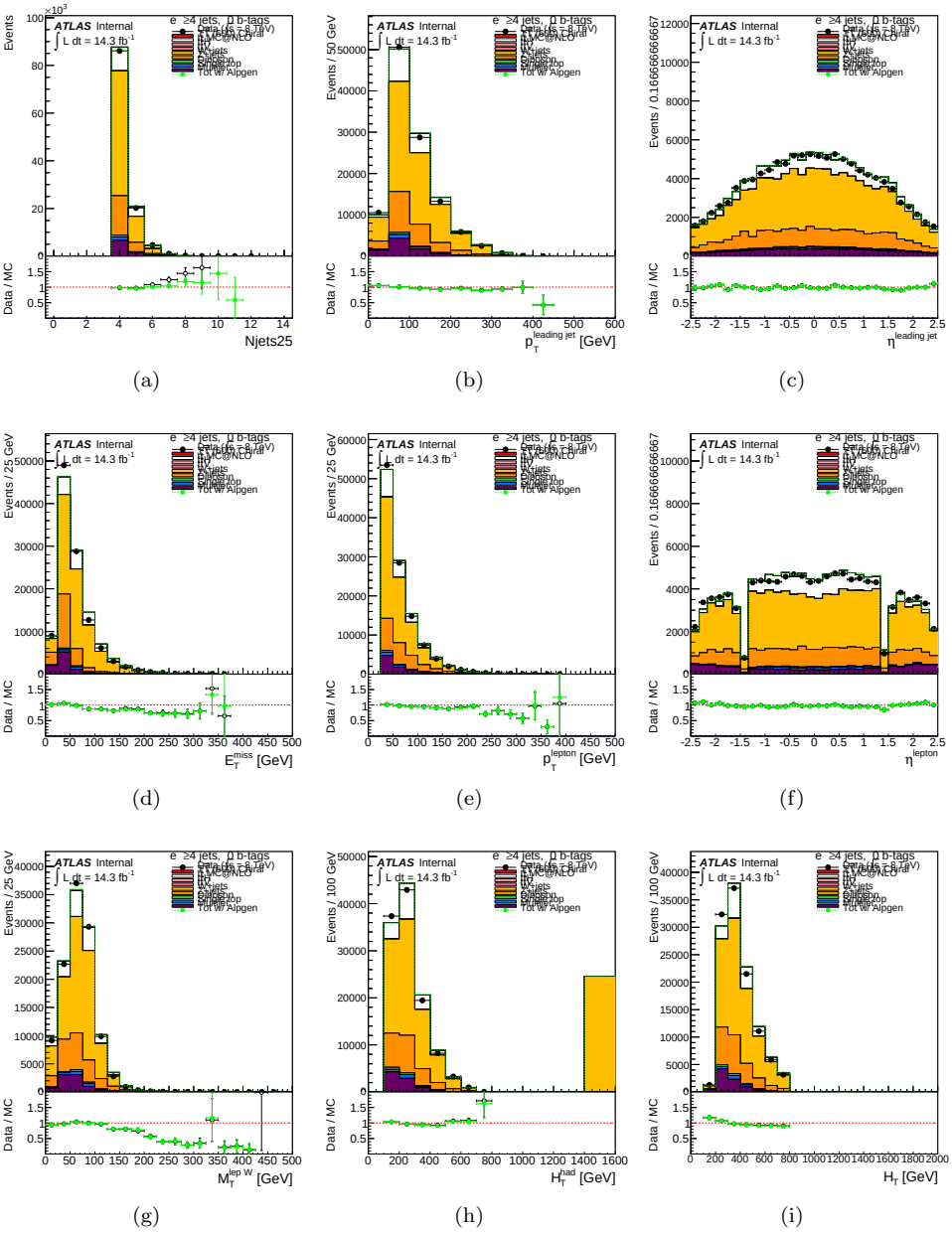

ATLAS Internal
e ≥4 jets, ≥1 b-tags
∫ L dt = 14.3 fb-1
Njets25
Data / MC
(a)
Events / 50 GeV
leading jet [GeV]
(b)
Events / 0.16666666666667
(c)
Events / 25 GeV
(d)
(e)
(f)
Events / 100 GeV
(g)
(h)
(i)

Figure B.4

B.2.2 Muon channel

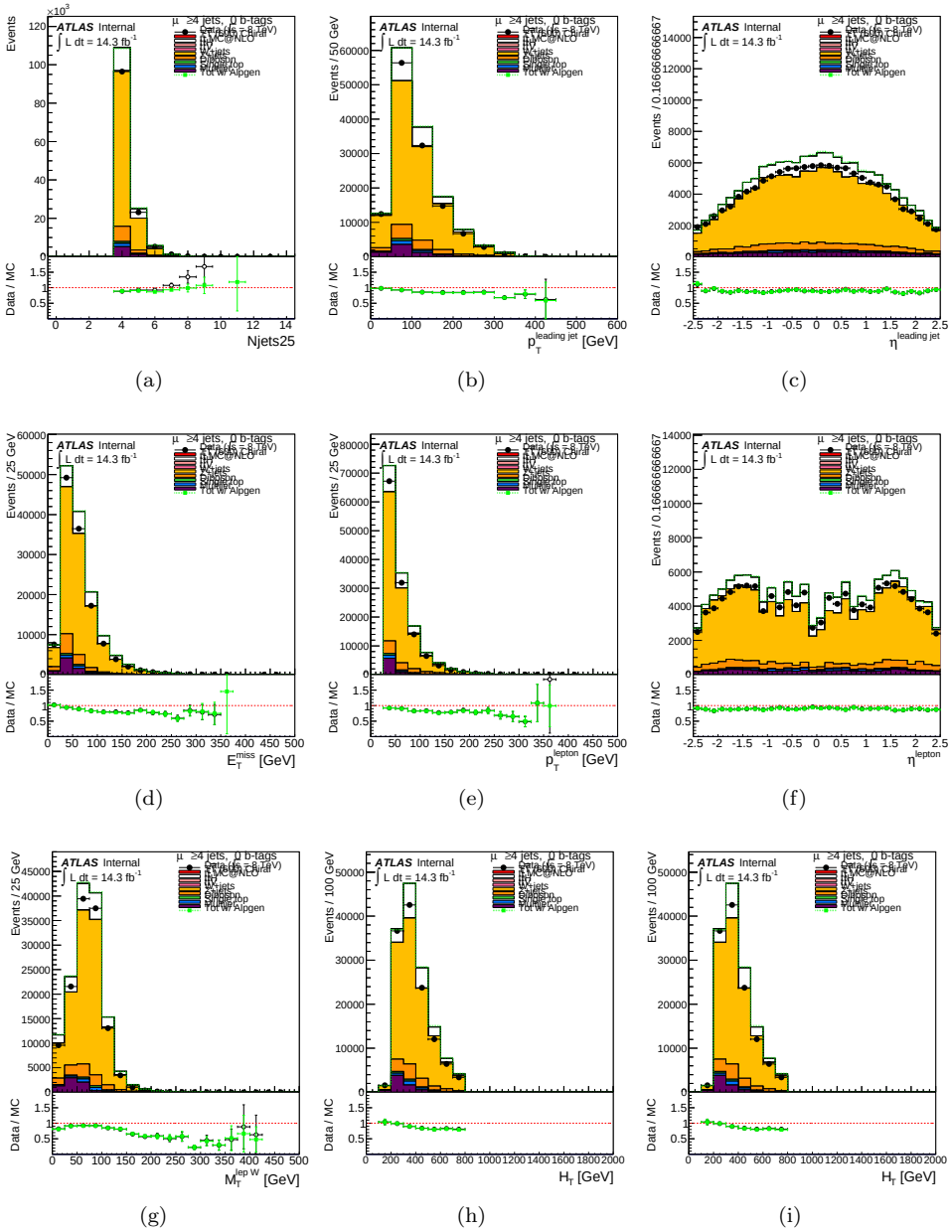


Figure B.5

B.2.3 Electron+Muon channel

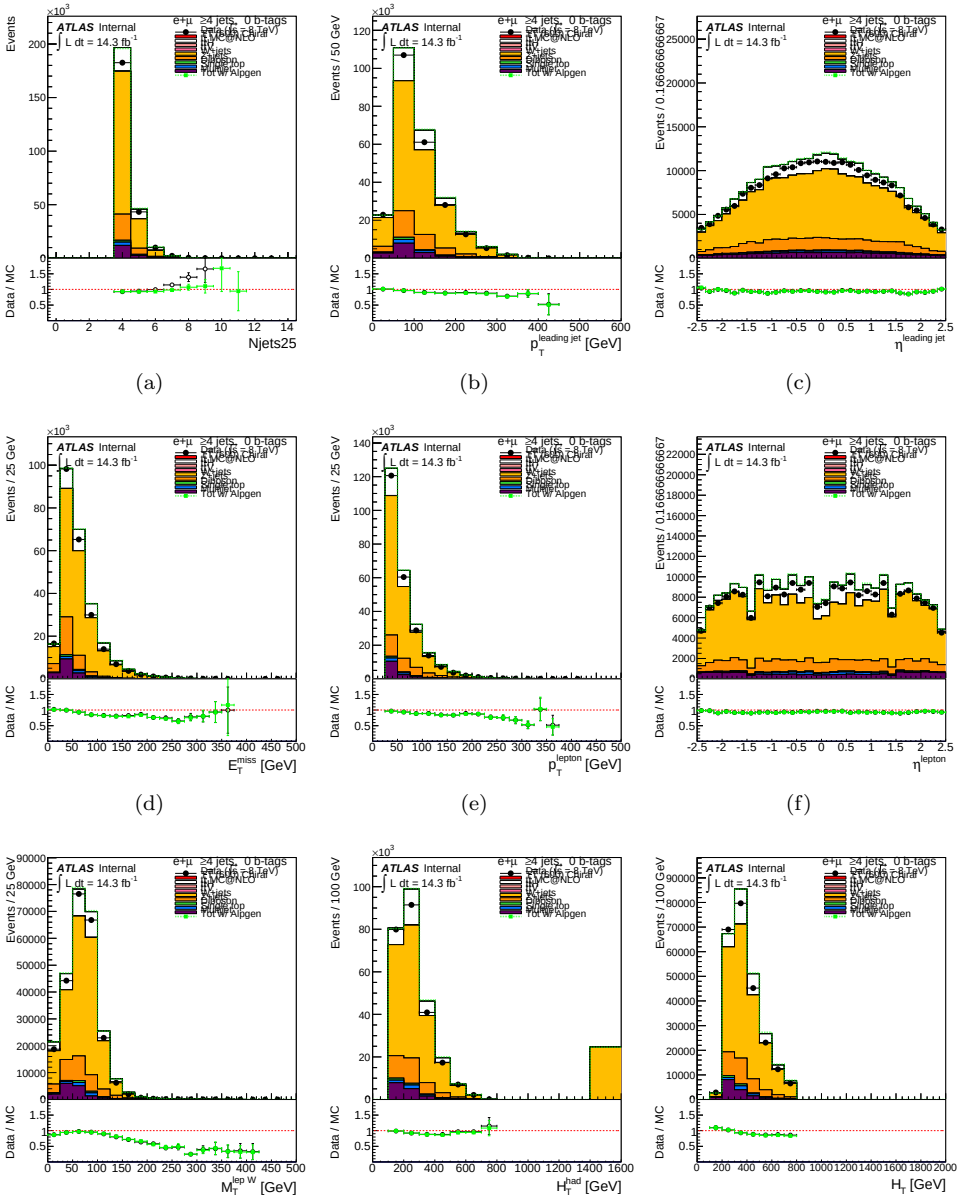


Figure B.6

The Tag Rate Function method

When requiring high b-tagged jet multiplicity in an analysis, usually the available statistics in Monte Carlo simulated samples is significantly reduced. This leads to large fluctuations for the final discriminant variable distribution and, as a direct consequence, reduced sensitivity of the search because of unphysical variations of the systematic uncertainties affected by the unreliable statistical uncertainties. Furthermore this can introduce a bias in the observed limits depending on the side of the fluctuation of the Monte Carlo templates with respect to the data in the final signal region. A Tag Rate Function (TRF) method can mitigate this problem by keeping the full Monte Carlo pre-tag statistics and deriving the shape and normalization predictions in the b-tagged channels through a reweighting procedure.

While the vector-like signal samples, obtained using a fast simulation of the detector, have enough statistics in the final channels for both the $T\bar{T} \to Wb + X$ and $T\bar{T} \to Ht + X$ analyses, Monte Carlo backgrounds are highly affected by the cuts specifically designed to reduce their presence in the final selection. Therefore, the TRF method is applied to all Monte Carlo samples in the $T\bar{T} \to Ht + X$ analysis, which requires high b-tagged jet multiplicities in its search channels, and to all non-$t\bar{t}$ Monte Carlo samples (W+jets, Z+jets, diboson, single-top, $t\bar{t}V$) in the $T\bar{T} \to Wb + X$ analysis, which requires only one b-tagged jet but drastically reduces the background contribution to the signal region through tight requirements.

In this appendix the principle of this method and the various studies performed to validate its usage in the analyses are discussed.

C.1 TRF method principle

When a direct cut on the b-tagging weight returned by the b-tagging algorithm is applied, events that do not have the requested number of jets satisfying this cut are rejected. Considering that commonly chosen working points for the b-tagging algorithms have a measured efficiency of 70%, it is easy to imagine the potential loss in acceptance when requiring more that one b-tagged jets. By using the TRF method, no event is cut based on how many b-

taged jets are counted, but instead all the pre-tag events are reweighted. The event weight is calculated using the b-tagging efficiency (which depends on the jet's η, p_{T} and true flavour) and based on the kinematics and flavour of the jets found in each event. This weight can be interpreted as the probability of the given event to contain the desired number of b-tagged jets.

Given a jet with η, p_{T} and flavour f, its tagging probability can be written as:

$$\varepsilon\left(f, |\eta|, p_{\mathrm{T}}\right). \tag{C.1}$$

For a given event with N jets, its probability of containing exactly one b-tagged jet can be computed as:

$$P_{=1} = \sum_{i=1}^{N} \left(\varepsilon_i \prod_{j \neq i} (1 - \varepsilon_j) \right). \tag{C.2}$$

This can be generalized as the probability of containing exactly M b-tagged jets by iterating over all the possible subsets of M and $N - M$ jets as:

$$P_{=M} = \sum_{i,\ldots,m=1}^{N} \left(\prod_{i=1}^{m=M} \varepsilon_i \prod_{j=m+1}^{N} (1 - \varepsilon_j) \right), \tag{C.3}$$

and in general the probability for inclusive b-tagging selections can be computed:

$$P_{=0} = \prod_{i=1}^{N} (1 - \varepsilon_j), \tag{C.4}$$

$$P_{\geq 1} = 1 - P_{=0}, \tag{C.5}$$

$$\ldots$$

$$P_{\geq m} = 1 - \sum_{i=0}^{m-1} P_{=i}. \tag{C.6}$$

C.1.1 Validation

This method relies on the correct calibration of the b-tagging efficiency in Monte Carlo samples. Closure tests performed with the official calibration files have shown that the efficiency parametrization is not as accurate as expected. Assuming a correct calibration, the average of the histogram of $1/\varepsilon$ vs η, p_{T} and true jet flavour should be flat and with mean equal to one.

In Figure C.1 the result of this test is shown, and it can be observed that for the official maps (the left colums) there are departures from closure of to up to 40% in some regions of the light flavor map, and on average of 13%. New efficiency maps have been derived using a combination of $t\bar{t}$ `MC@NLO`, $t\bar{t}$ `Alpgen`, and `Protos` $T\bar{T}$ Monte Carlo samples. The plots on the right column of Figure C.1 show reasonably good closure for the newly derived efficiency map, which will therefore be used for the probability computations in the TRF method.

As a validation check, Figure C.2 compares the spectrum of the number of b-tagged jets

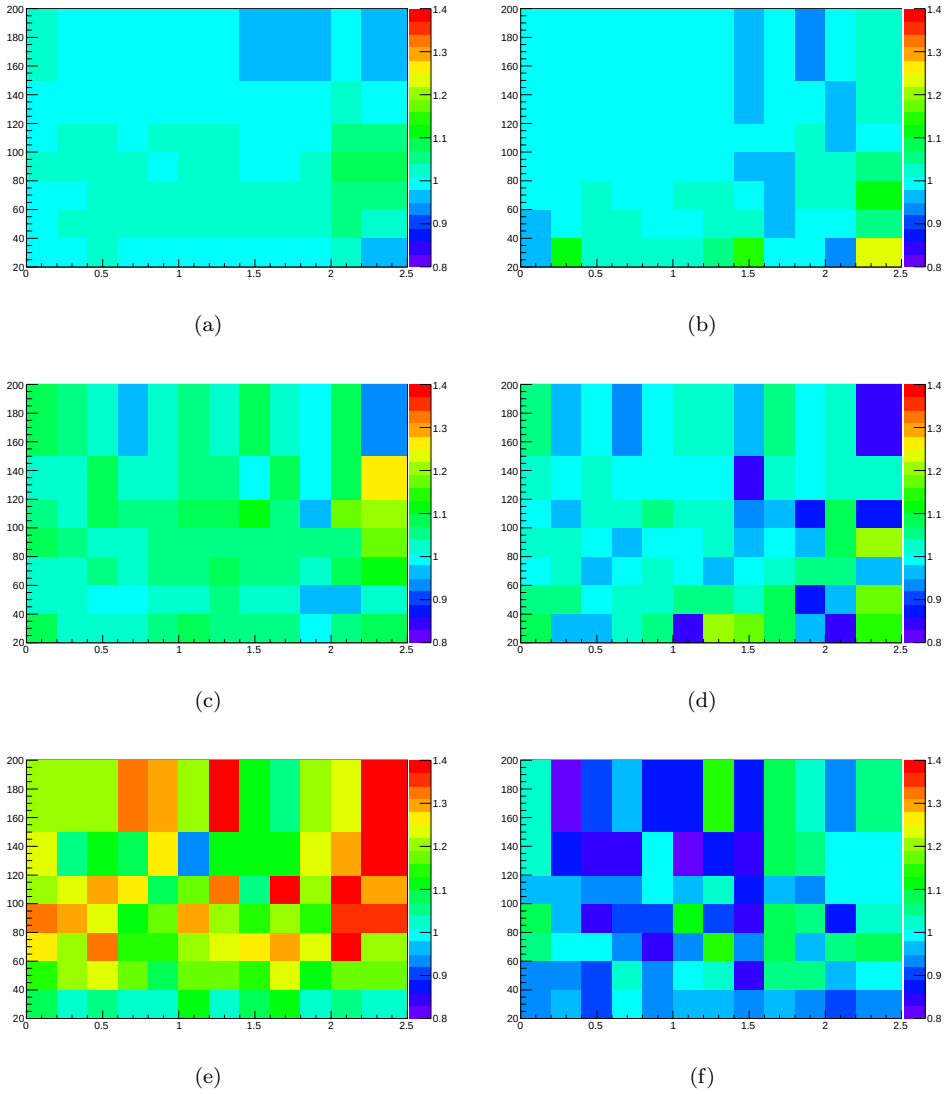
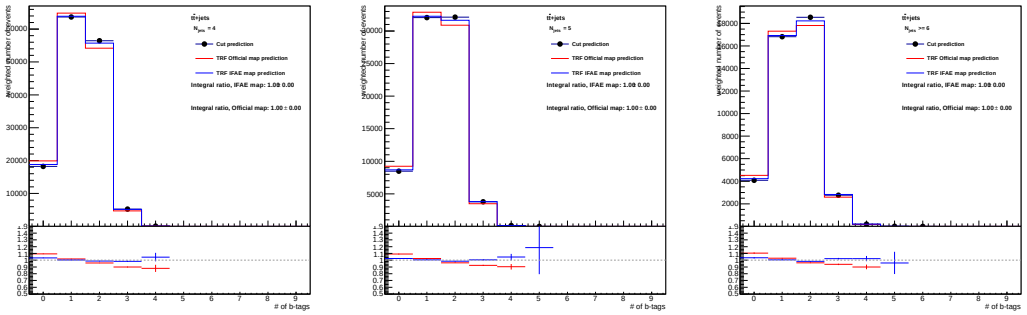


Figure C.1: Results of the closure test using efficiency from the official calibration file (left column) and the private efficiency map (right column). The test is split in the different jet flavours: b-jets (top), c-jets (middle) and light jets (bottom).

distribution in the $t\bar{t}$ Monte Carlo sample simulated with `ALPGEN` obtained using the TRF method and the direct b-tagging. The shapes are found to be compatible.

C.2 TRF in the $T\bar{T} \to Ht + X$ analysis

The TRF method is used for all the Monte Carlo backgrounds in the $T\bar{T} \to Ht + X$ analysis. The validity of the method is checked by comparing the normalization and shape of final discriminant H_T distribution for the most relevant background, the $t\bar{t}$ sample simulated with `ALPGEN`, in different jet and $b$jetmultiplicity channels, using the TRF method and direct



(a) (b) (c)

Figure C.2: Comparison of the TRF and direct b-tag cut prediction for the b-tag spectrum, in the 4 jet exclusive (a), 5 jet exclusive (b) and 6 jet inclusive (c) channels.

b-tagging. As it can be seen in these plots, the prediction obtained with the TRF method is accurate up to the statistical error.

C.3 TRF in the $T\bar{T} \to Wb + X$ analysis

As previously mentioned, the $T\bar{T} \to Wb + X$ analysis applies the TRF method only to Monte Carlo simulated samples of W/Z+jets, single top, diboson and $t\bar{t}V$. Table C.1 compares the predicted yields at the various steps of the selection obtained applying the TRF method and the direct cut on the b-tagging weight for each of these simulated processes. It can be seen that good agreement is observed for selections where the direct b-tagging still leaves sufficient Monte Carlo statistics, while going tighter into the signal regions only the TRF method returns non-zero prediction. Also to be noted is that in all cases the statistical uncertainty on the predicted yield is improved by the TRF method.

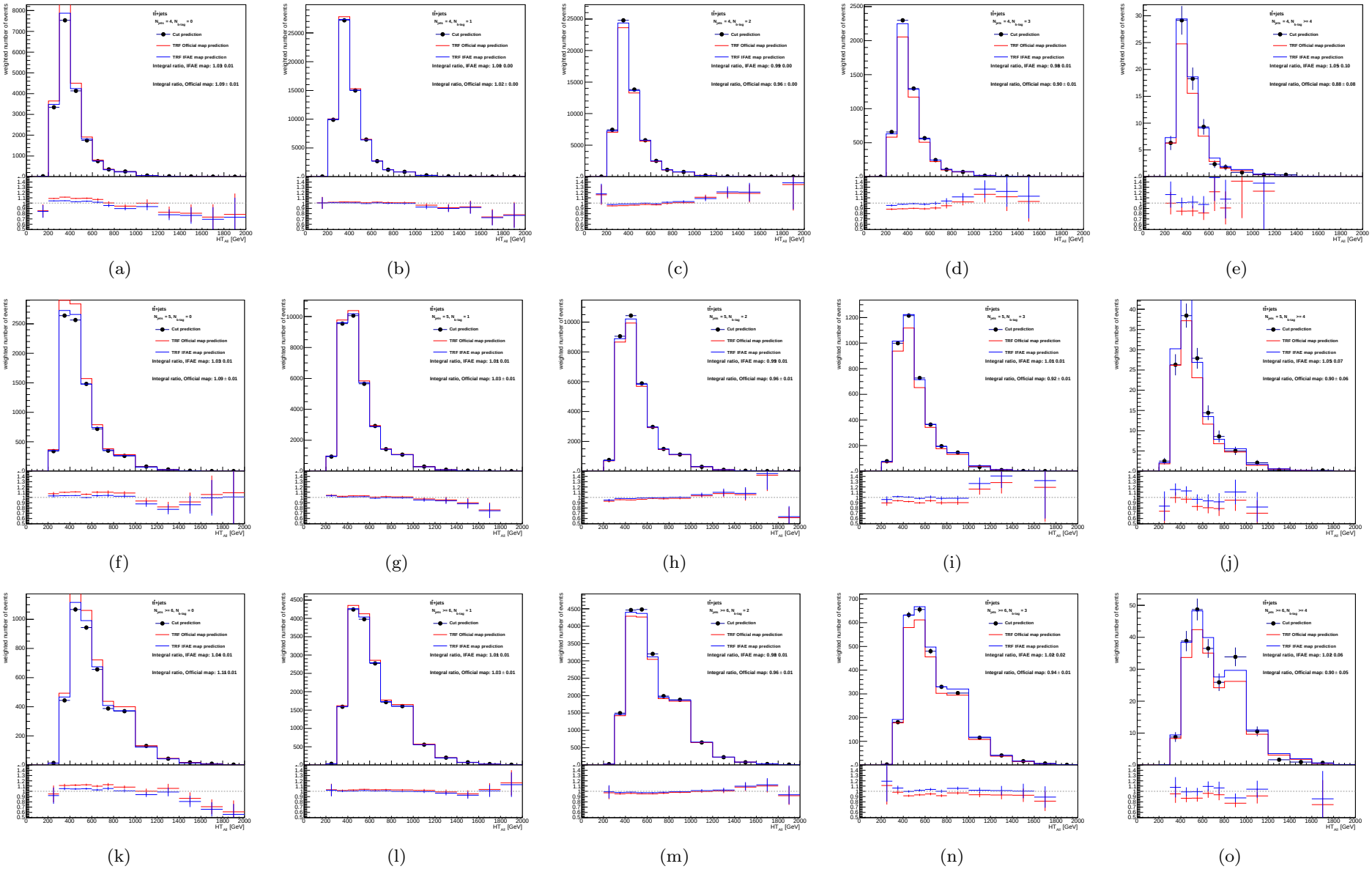


Figure C.3: Comparison of the TRF and b-tag cut prediction for the HT_{all} distribution in the (a–e) 4 jet exclusive, (f–j) 5 jet exclusive and (k–o) 6 jet inclusive channels, for different b-tagging multiplicities (from left to right, 0, 1, 2, 3 exclusive and 4 inclusive) for the $t\bar{t}$ `ALPGEN` sample.

	TRF		Direct tagging	
	Entries	Predicted yield	Entries	Predicted yield
		W+jets		
Preselection	66381	37679.2 ± 324.0	8880	37317.3 ± 527.8
≥ 1 W	1321	723.2 ± 40.3	185	713.9 ± 66.7
$H_T > 800$ GeV	520	314.0 ± 27.4	84	308.0 ± 41.7
$p_T(b_1) > 160$ GeV	262	146.9 ± 17.8	44	155.3 ± 30.5
$p_T(b_2) > 80$ GeV	63	46.9 ± 11.5	11	39.9 ± 13.8
$\Delta R(l,\nu) < 1.2$	28	16.3 ± 6.0	5	14.7 ± 7.2
$\min\Delta R(l,b) > 1.4$	15	6.3 ± 2.8	2	5.2 ± 4.0
$\min\Delta R(W,b) > 1.4$	9	5.5 ± 2.8	1	3.7 ± 3.7
		Z+jets		
Preselection	19500	6054.4 ± 84.3	4573	6015.1 ± 147.7
≥ 1 W	331	133.9 ± 14.2	87	125.9 ± 19.0
$H_T > 800$ GeV	130	47.8 ± 8.7	32	49.4 ± 12.2
$p_T(b_1) > 160$ GeV	74	25.2 ± 6.7	19	29.8 ± 10.2
$p_T(b_2) > 80$ GeV	22	12.8 ± 6.0	9	15.9 ± 7.4
$\Delta R(l,\nu) < 1.2$	5	1.1 ± 0.6	1	0.8 ± 0.8
$\min\Delta R(l,b) > 1.4$	2	0.2 ± 0.2	0	0.0 ± 0.0
$\min\Delta R(W,b) > 1.4$	2	0.2 ± 0.2	0	0.0 ± 0.0
		Dibosons		
Preselection	18629	555.5 ± 7.0	3532	552.2 ± 11.4
≥ 1 W	336	10.9 ± 1.1	50	8.6 ± 1.4
$H_T > 800$ GeV	85	2.9 ± 0.6	14	2.4 ± 0.7
$p_T(b_1) > 160$ GeV	32	0.9 ± 0.3	4	0.5 ± 0.3
$p_T(b_2) > 80$ GeV	14	0.5 ± 0.2	2	0.3 ± 0.2
$\Delta R(l,\nu) < 1.2$	9	0.2 ± 0.1	1	0.1 ± 0.1
$\min\Delta R(l,b) > 1.4$	8	0.1 ± 0.1	1	0.1 ± 0.1
$\min\Delta R(W,b) > 1.4$	4	0.1 ± 0.0	0	0.0 ± 0.0
		Single top		
Preselection	74327	14670.8 ± 97.9	59854	14722.9 ± 107.0
≥ 1 W	2799	469.9 ± 14.1	2349	468.0 ± 14.9
$H_T > 800$ GeV	986	164.7 ± 7.9	826	162.8 ± 8.4
$p_T(b_1) > 160$ GeV	624	105.2 ± 6.4	539	107.6 ± 6.9
$p_T(b_2) > 80$ GeV	292	51.6 ± 4.4	263	53.8 ± 4.7
$\Delta R(l,\nu) < 1.2$	165	30.2 ± 3.6	147	30.4 ± 3.7
$\min\Delta R(l,b) > 1.4$	61	14.0 ± 2.4	55	14.0 ± 2.4
$\min\Delta R(W,b) > 1.4$	21	4.4 ± 1.3	19	4.8 ± 1.4
		$t\bar{t}V$		
Preselection	171489	706.1 ± 2.1	142296	709.0 ± 2.3
≥ 1 W	19492	78.6 ± 0.7	15862	78.3 ± 0.8
$H_T > 800$ GeV	8516	34.2 ± 0.5	6963	34.2 ± 0.5
$p_T(b_1) > 160$ GeV	4419	17.9 ± 0.3	3657	18.0 ± 0.4
$p_T(b_2) > 80$ GeV	2267	9.3 ± 0.2	1912	9.5 ± 0.3
$\Delta R(l,\nu) < 1.2$	1227	5.1 ± 0.2	1029	5.2 ± 0.2
$\min\Delta R(l,b) > 1.4$	321	1.3 ± 0.1	265	1.4 ± 0.1
$\min\Delta R(W,b) > 1.4$	138	0.5 ± 0.1	104	0.6 ± 0.1

Table C.1: Comparison of expected yields between TRF and direct tagging as a function of cuts applied from the preselection level up to the TIGHT selection.

Search for $T\bar{T} \to WbWb$ with pp collisions at $\sqrt{s}$= 7 TeV

Using the data collected by the ATLAS experiment in 2011 from pp collisions at a center of mass energy of $\sqrt{s}$= 7 TeVthe first quasi-model independent search for heavy vector-like top quarks was performed [3]. Originally designed for searches of chiral fourth-generation top partners, this analysis is optimized for the $T\bar{T} \to WbWb$ decay channel.

D.1 Jet mass related systematic uncertainties

(see Section 8.2.7 in Ref. [**?**].

The existing jet mass scale/resolution uncertainties had been derived for larger anti-k_t jets ($R = 1.0$) and in release 16 [**?**] and were believed to be conservative when applied to smaller $R = 0.4$ jets using the refined calibration from release 17.

The jet mass scale uncertainties used in Ref. [**?**] were 4.5% for jets with $p_{\mathrm{T}} < 400$ GeV and 6% for $p_{\mathrm{T}} > 400$ GeV (from Table 18 in Ref. [**?**]). The jet mass resolution uncertainty used was 20%, independent of jet p_{T} (from Table 18 in Ref. [**?**]).

The propagation of these uncertainties in the previous analysis led to a degradation in the expected sensitivity of only a few GeV (compare yellow and light blue curves in Fig. 25(b) of Ref. [**?**]).

$T\bar{T} \to Wb + X$ analysis: SR cut-flow

In this appendix some more information about the cut-flow in the signal regions for the $T\bar{T} \to Wb + X$ analysis is given. We remind in Table E.1 the signal regions definition. In the following sections the selected number of events selected in the electron (Table E.2) and muon (Table E.3) channels are reported, and the distribution of the discriminant variable m_{reco} is shown in the various signal regions.

Selection	Signal Region	Requirements
Preselection		One electron or muon $E_{\text{T}}^{\text{miss}} > 20$ GeV, $E_{\text{T}}^{\text{miss}} + m_{\text{T}} > 60$ GeV ≥ 4 jets, ≥ 1 b-tagged jets
Loose selection	SR0	Preselection
	SR1	+ ≥ 1 W_{had} candidates
	SR2	+ $H_{\text{T}} > 800$ GeV
	SR3	+ $p_{\text{T}}(b_1) > 160$ GeV
	SR4	+ $p_{\text{T}}(b_2) > 80$ GeV
	SR5 (≡Loose)	+ $\Delta R(\ell, \nu) < 1.2$
Tight selection	SR5	Loose selection
	SR6	+ $\min\Delta R(\ell, b) > 1.4$
	SR7 (≡Tight)	+ $\min\Delta R(W_{\text{had}}, b) > 1.4$

Table E.1: Summary of event selection requirements.

E.1 Event yields in the electron and muon channels

	$T\bar{T}$ (600) chiral	$t\bar{t}$	non-$t\bar{t}$	Tot Bkg	Data
SR0	190 ± 5	92135 ± 190	29216 ± 229	121352 ± 297	117565 ± 343
SR1	89 ± 3	2918 ± 34	678 ± 32	3596 ± 46	3845 ± 62
SR2	86 ± 3	727 ± 18	274 ± 22	1001 ± 28	1109 ± 33
SR3	75 ± 3	356 ± 12	153 ± 16	509 ± 20	560 ± 24
SR4	59 ± 3	195 ± 9	62 ± 11	258 ± 14	263 ± 16
SR5	49 ± 2	119 ± 7	25 ± 6	144 ± 9	164 ± 13
SR6	38 ± 2	11 ± 2	11 ± 2	22 ± 3	27 ± 5
SR7	30 ± 2	4 ± 2	4 ± 2	9 ± 2	18 ± 4

Table E.2: Number of observed events, integrated over the whole mass spectrum, compared to the Standard Model expectation for the e+jets channel in the Signal Regions (see Table E.1 for the region definitions). The expected signal yields for a chiral fourth-generation T quark with $m_T = 600$ GeV are also shown. The quoted uncertainties are only statistical.

	$T\bar{T}$ (600) chiral	$t\bar{t}$	non-$t\bar{t}$	Tot Bkg	Data
SR0	190 ± 5	109558 ± 209	36747 ± 274	146304 ± 344	144316 ± 380
SR1	79 ± 3	3417 ± 37	757 ± 33	4174 ± 50	4556 ± 67
SR2	75 ± 3	848 ± 20	280 ± 21	1128 ± 28	1250 ± 35
SR3	64 ± 3	456 ± 14	139 ± 13	595 ± 19	663 ± 26
SR4	48 ± 2	242 ± 10	56 ± 9	297 ± 13	335 ± 18
SR5	39 ± 2	144 ± 8	25 ± 4	169 ± 9	184 ± 14
SR6	29 ± 2	16 ± 3	11 ± 3	26 ± 4	34 ± 6
SR7	23 ± 2	6 ± 2	6 ± 3	12 ± 3	19 ± 4

Table E.3: Number of observed events, integrated over the whole mass spectrum, compared to the Standard Model expectation for the μ+jets channel in the Signal Regions (see Table E.1 for the region definitions). The expected signal yields for a chiral fourth-generation T quark with $m_T = 600$ GeV are also shown. The quoted uncertainties are only statistical.

E.2 Reconstructed mass in the SRs

Figure E.1 shows the progress of signal selection and background rejection in the various signal regions that progressively approach the final TIGHT selection.

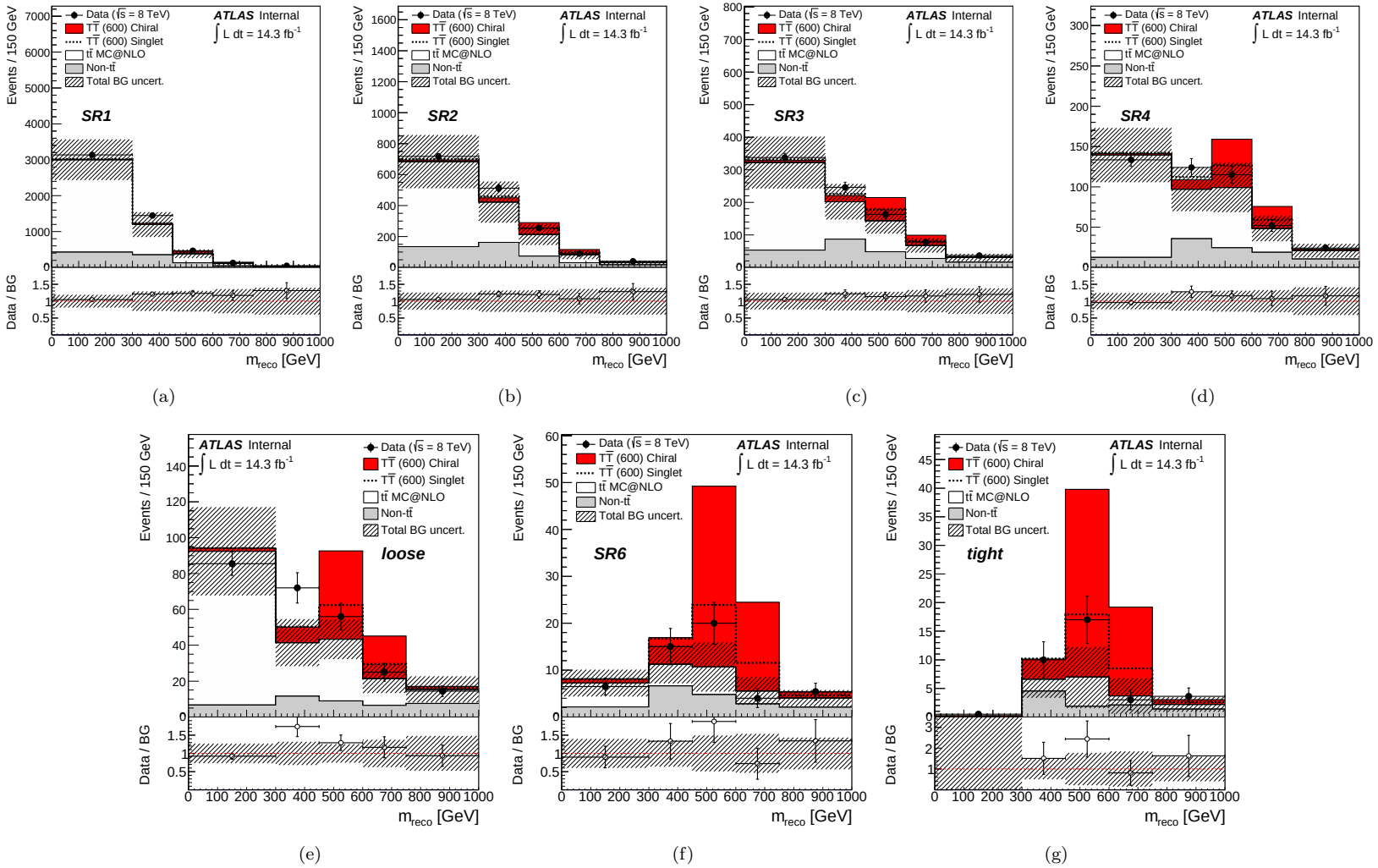


Figure E.1: Distribution of the reconstructed mass m_{reco} in the combined electron and muon channel for the various signal regions: (a) SR1, (b) SR2, (c) SR3, (d) SR4, (e) SR5 also known as LOOSE selection, (f) SR6 and (e) SR7 also known as TIGHT selection. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.

$T\bar{T} \to Wb + X$ analysis: data to background comparison in SDRs

F.1 Data to background comparison in SDR0

SDR0: preselection cuts, ≥ 1 W_{had} candidates, $H_{\text{T}} < 800$ GeV.

	ELEMUONCR0_1W
$T\bar{T}$(600 GeV) (Chiral)	$7.05 \pm 0.87\ ^{+1.01}_{-1.76}$
$t\bar{t}$	$4760.66 \pm 43.08\ ^{+805.03}_{-837.88}$
W+jets	$409.22 \pm 29.61\ ^{+169.34}_{-136.66}$
Z+jets	$86.14 \pm 11.27\ ^{+42.07}_{-45.56}$
Diboson	$7.95 \pm 0.95\ ^{+2.04}_{-2.28}$
Single top	$305.21 \pm 11.66\ ^{+43.96}_{-35.20}$
$t\bar{t}V$	$44.46 \pm 0.53\ ^{+13.57}_{-13.66}$
Multijet	$27.22 \pm 7.44 \pm 13.61$
Total bkg.	$5640.87 \pm 55.24\ ^{+946.22}_{-948.39}$
Data	6042

Table F.1: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR0 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

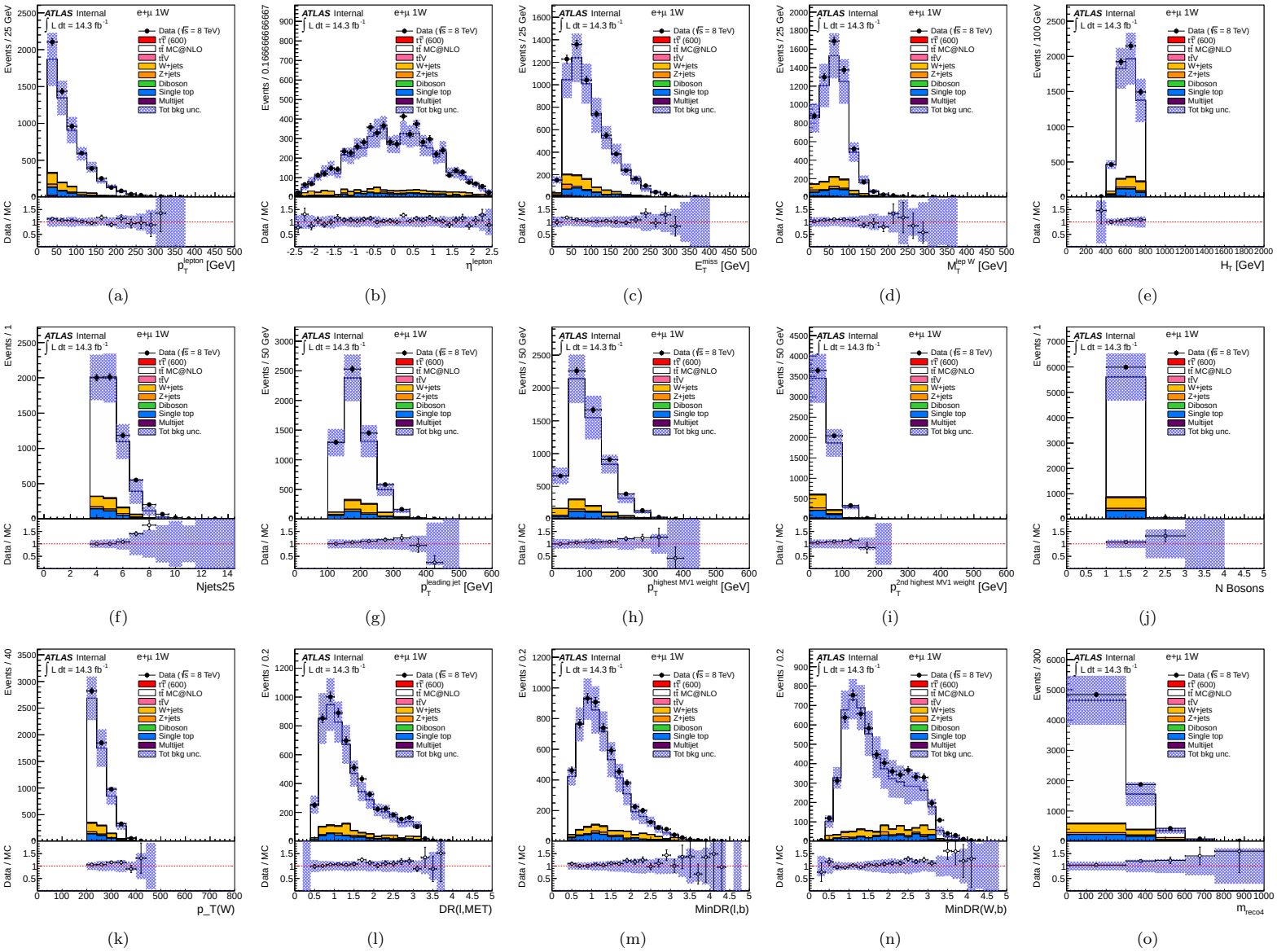


Figure F.1: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR0 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.2 Data to background comparison in SDR1

SDR1: preselection cuts, ≥ 1 W_{had} candidates, $m_{\text{reco}} < 200$ GeV. The m_{reco} variable represents the reconstructed heavy quark mass and is defined in Section 6.5.

	ELEMUONCR5_1W
$T\bar{T}$(600 GeV) (Chiral)	$7.07 \pm 0.94\ ^{+2.91}_{-2.19}$
$t\bar{t}$	$3870.01 \pm 38.16\ ^{+612.74}_{-702.94}$
W+jets	$218.14 \pm 21.80\ ^{+92.42}_{-85.97}$
Z+jets	$41.06 \pm 7.35\ ^{+20.80}_{-25.36}$
Diboson	$1.88 \pm 0.35\ ^{+0.77}_{-0.74}$
Single top	$136.68 \pm 7.11\ ^{+22.98}_{-18.41}$
$t\bar{t}V$	$30.24 \pm 0.44\ ^{+9.28}_{-9.32}$
Multijet	$-2.84 \pm 23.67 \pm 0.00$
Total bkg.	$4295.16 \pm 50.96\ ^{+679.61}_{-773.95}$
Data	4174

Table F.2: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR1 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

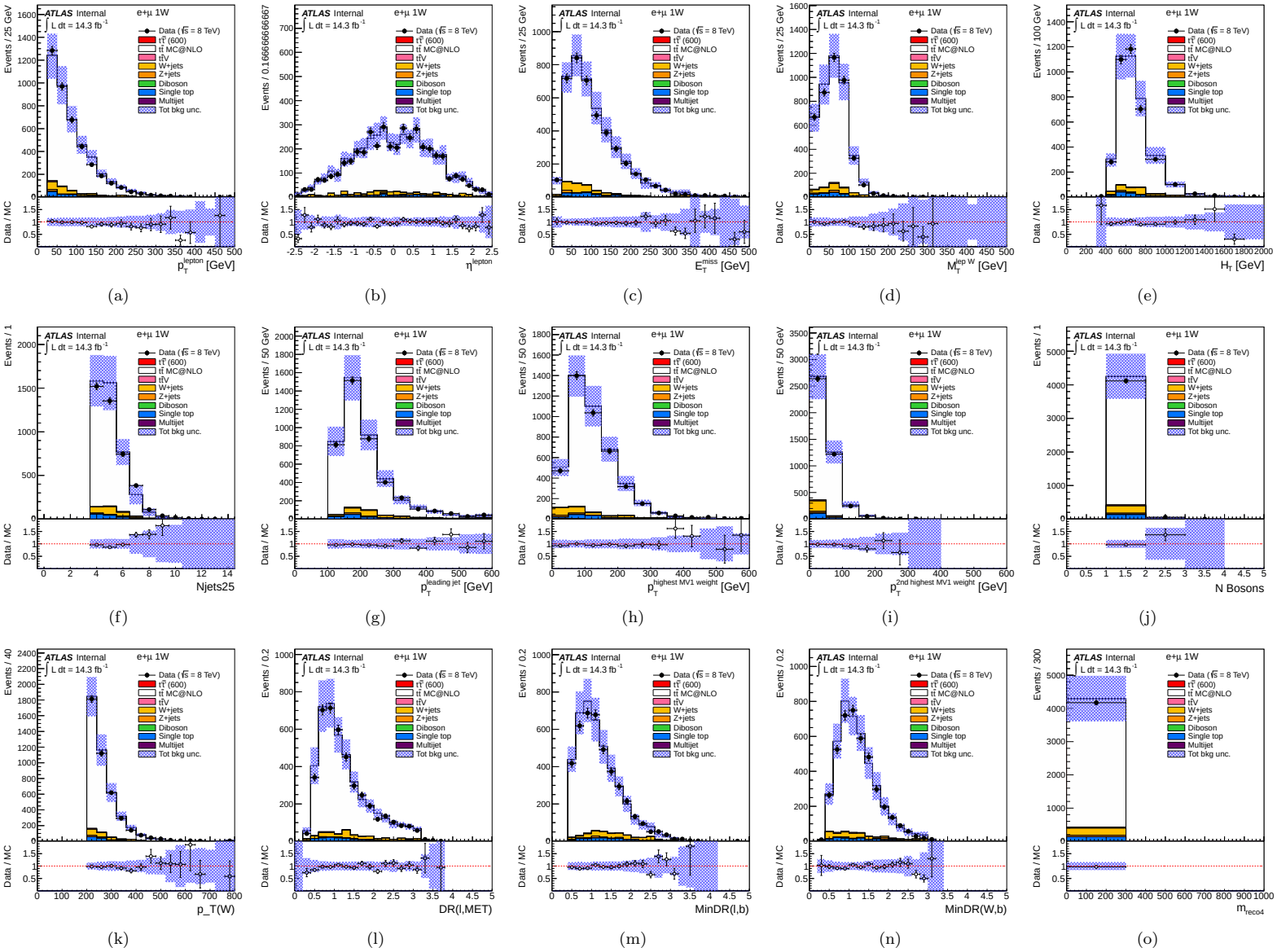


Figure F.2: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR1 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.3 Data to background comparison in SDR2

SDR2: *loose* selection with reversed H_{T} cut (i.e. $H_{\mathrm{T}} < 800$ GeV).

	ELEMUONCR1_1W
$T\bar{T}$(600 GeV) (Chiral)	$0.73 \pm 0.31\ ^{+0.40}_{-0.45}$
$t\bar{t}$	$72.59 \pm 5.46\ ^{+18.64}_{-18.93}$
W+jets	$6.46 \pm 5.61\ ^{+2.77}_{-6.57}$
Z+jets	$0.91 \pm 0.91\ ^{+1.01}_{-1.36}$
Diboson	$0.10 \pm 0.08\ ^{+0.02}_{-0.02}$
Single top	$1.13 \pm 0.57\ ^{+1.38}_{-0.42}$
$t\bar{t}V$	$0.47 \pm 0.05\ ^{+0.20}_{-0.18}$
Multijet	$0.36 \pm 0.66 \pm 0.18$
Total bkg.	$82.01 \pm 7.93\ ^{+20.71}_{-23.47}$
Data	85

Table F.3: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR2 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

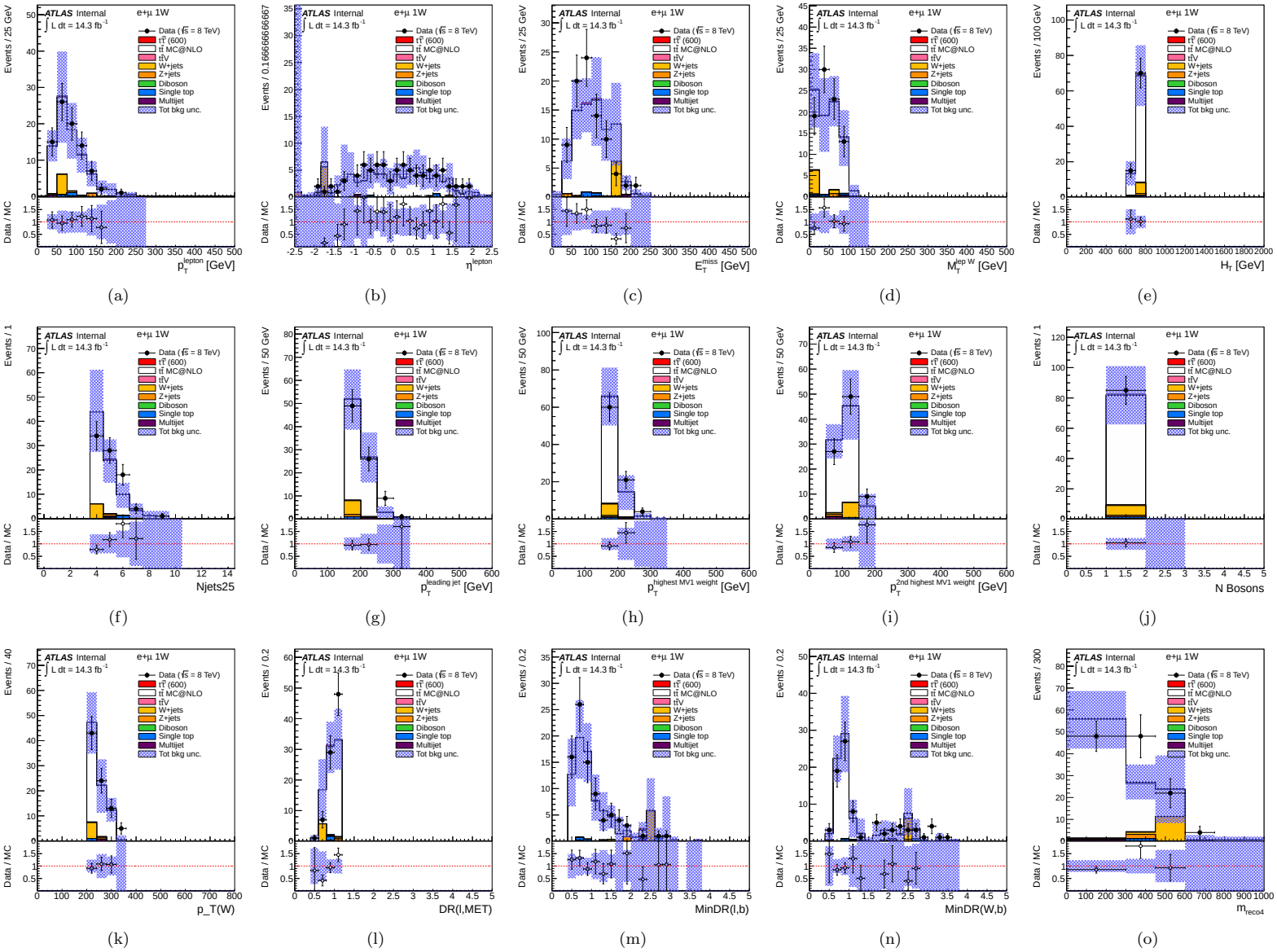


Figure F.3: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR2 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.4 Data to background comparison in SDR3

SDR3: *loose* selection with reversed b-jet p_{T} cuts (i.e. $p_{\mathrm{T}} < 160$ GeV and $p_{\mathrm{T}} < 80$ GeV).

	ELEMUONCR2_1W
$T\bar{T}$(600 GeV) (Chiral)	$12.84 \pm 1.28\ ^{+3.53}_{-5.30}$
$t\bar{t}$	$472.53 \pm 14.37\ ^{+119.04}_{-117.67}$
W+jets	$92.31 \pm 16.04\ ^{+52.50}_{-43.49}$
Z+jets	$7.52 \pm 2.77\ ^{+5.81}_{-4.95}$
Diboson	$1.81 \pm 0.45\ ^{+0.18}_{-0.45}$
Single top	$39.14 \pm 3.73\ ^{+9.44}_{-6.71}$
$t\bar{t}V$	$8.87 \pm 0.23\ ^{+3.03}_{-2.93}$
Multijet	$-3.68 \pm 2.63 \pm 0.00$
Total bkg.	$618.50 \pm 22.19\ ^{+167.53}_{-149.64}$
Data	661

Table F.4: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR3 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

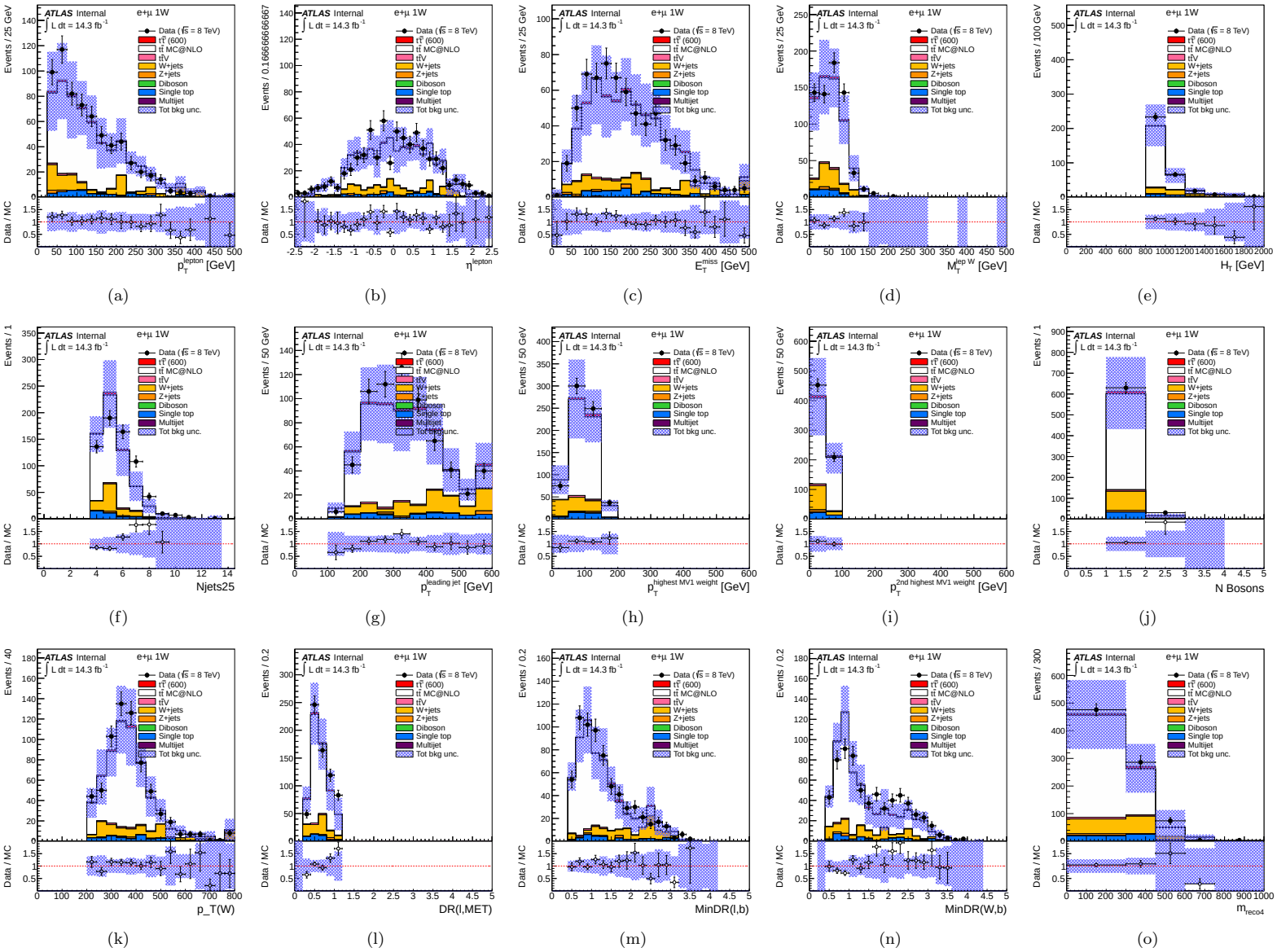


Figure F.4: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR3 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.5 Data to background comparison in SDR4

SDR4: *loose* selection with reversed $\Delta R(\ell, \nu)$ cut (i.e. $\Delta R(\ell, \nu) > 1.2$).

	ELEMUONCR3_1W
$T\bar{T}$(600 GeV) (Chiral)	$18.47 \pm 1.48\ ^{+1.09}_{-1.64}$
$t\bar{t}$	$173.13 \pm 8.82\ ^{+46.92}_{-48.59}$
W+jets	$30.64 \pm 9.78\ ^{+13.74}_{-12.43}$
Z+jets	$11.68 \pm 5.93\ ^{+5.89}_{-6.96}$
Diboson	$0.29 \pm 0.19\ ^{+0.17}_{-0.17}$
Single top	$21.46 \pm 2.54\ ^{+2.60}_{-2.54}$
$t\bar{t}V$	$4.21 \pm 0.16\ ^{+1.33}_{-1.33}$
Multijet	$0.49 \pm 0.91 \pm 0.25$
Total bkg.	$241.90 \pm 14.70\ ^{+53.57}_{-55.95}$
Data	250

Table F.5: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR4 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

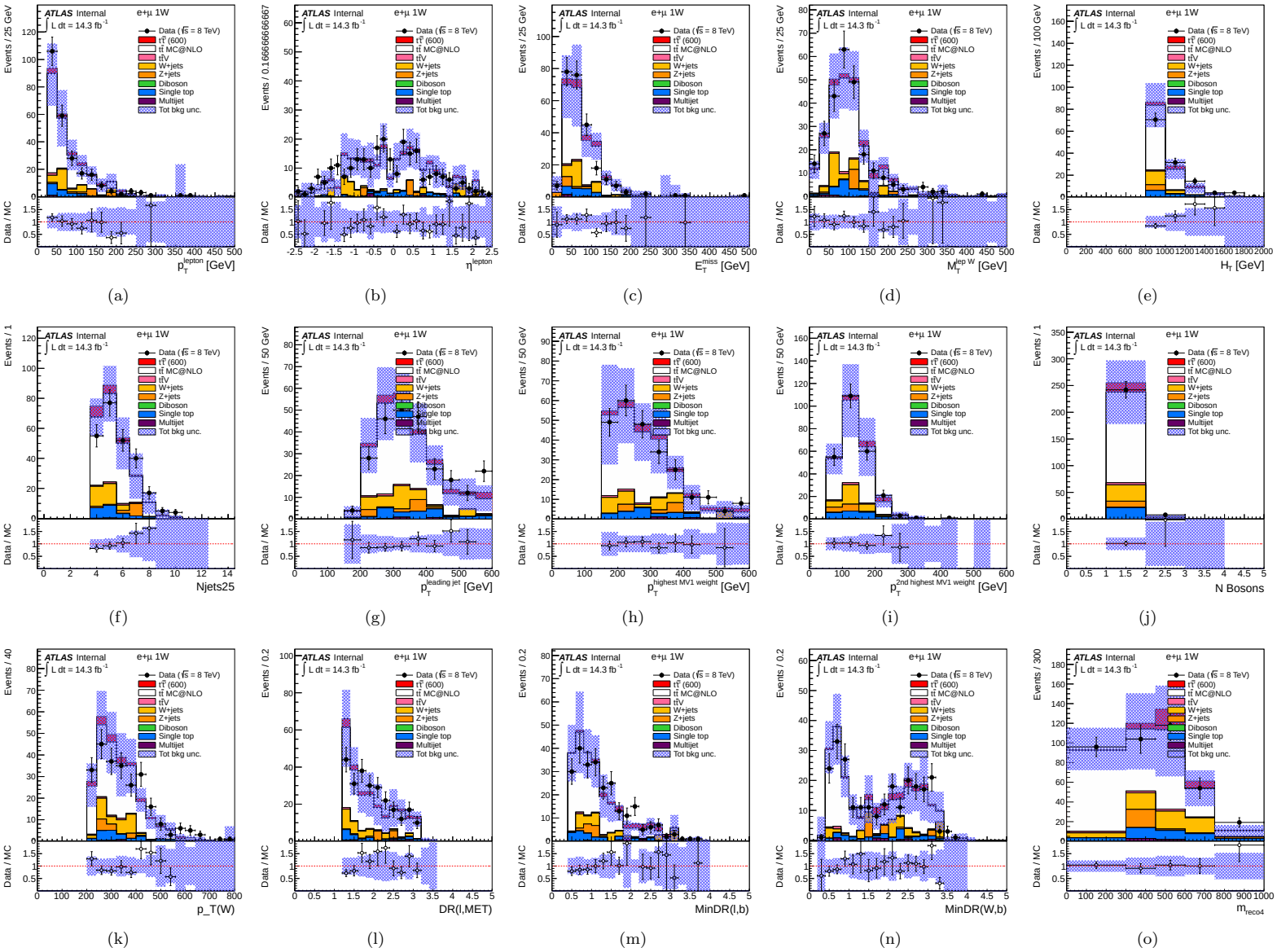


Figure F.5: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR4 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.6 Data to background comparison in SDR5

SDR5: *loose* selection with reversed $\min(\Delta R(W_{\text{had}}, b_{1,2}))$ and $\min(\Delta R(\ell, b_{1,2}))$ cuts (i.e. $\min(\Delta R(W_{\text{had}}, b_{1,2})) < 1.4$ and $\min(\Delta R(\ell, b_{1,2})) < 1.4$).

	ELEMUONCR4_1W
$T\bar{T}$(600 GeV) (Chiral)	$6.48 \pm 0.87\ ^{+2.02}_{-2.06}$
$t\bar{t}$	$180.06 \pm 8.35\ ^{+42.97}_{-52.75}$
W+jets	$3.29 \pm 1.79\ ^{+2.58}_{-2.48}$
Z+jets	$0.34 \pm 0.31\ ^{+0.21}_{-0.21}$
Diboson	$0.00 \pm 0.00 \pm 0.00$
Single top	$5.17 \pm 1.33\ ^{+1.28}_{-1.49}$
$t\bar{t}V$	$2.25 \pm 0.12\ ^{+0.74}_{-0.77}$
Multijet	$-1.04 \pm 0.83 \pm 0.00$
Total bkg.	$190.07 \pm 8.69\ ^{+43.97}_{-54.68}$
Data	178

Table F.6: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR5 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

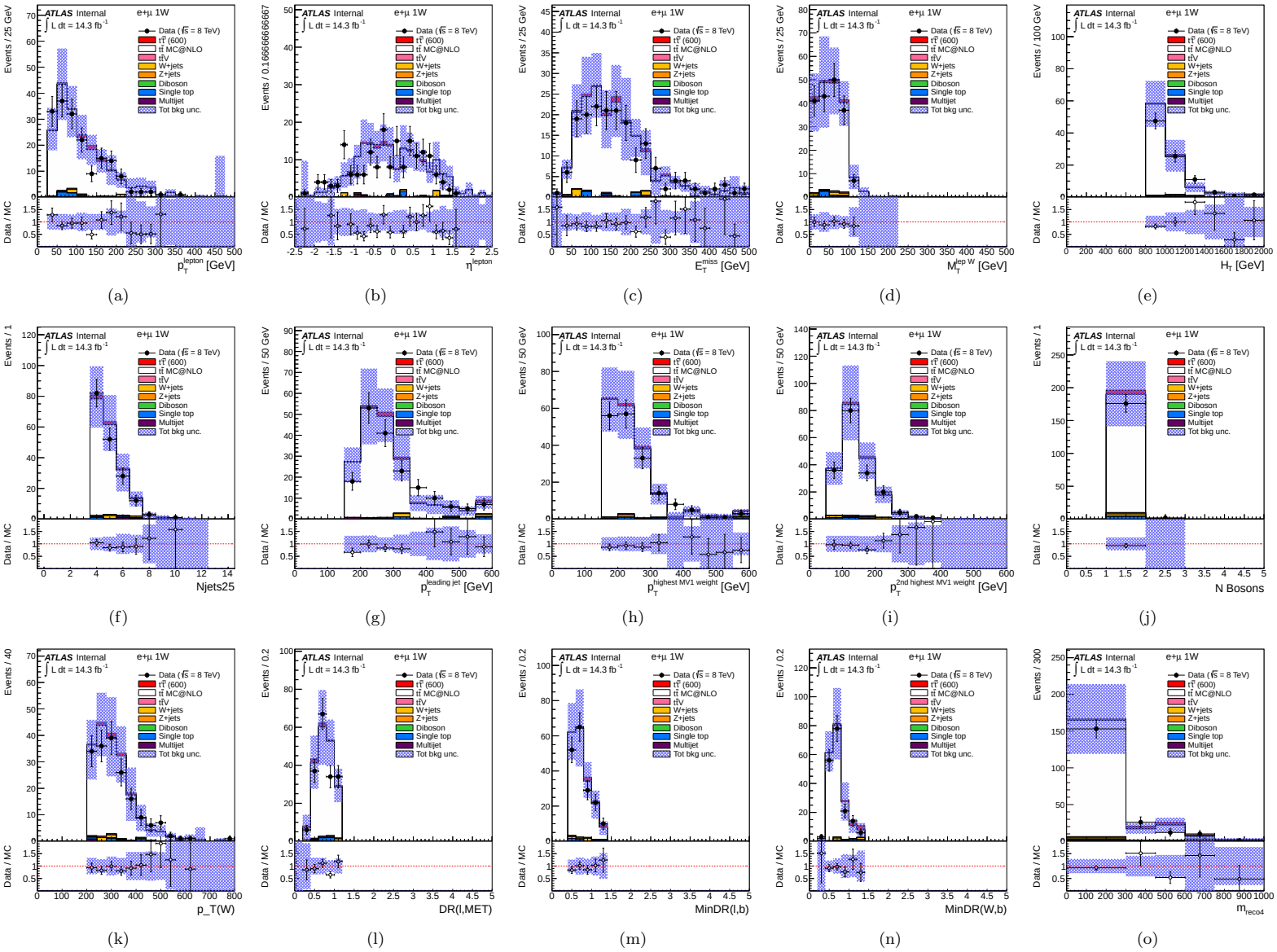


Figure F.6: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR5 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.7 Data to background comparison in SDR6

SDR6: *loose* selection, $m_{\text{reco}} < 200$ GeV.

	ELEMUONCR6_1W
$T\bar{T}$(600 GeV) (Chiral)	$0.38 \pm 0.19\ ^{+0.06}_{-0.13}$
$t\bar{t}$	$148.05 \pm 7.63\ ^{+30.06}_{-44.43}$
W+jets	$0.42 \pm 0.26\ ^{+0.16}_{-0.37}$
Z+jets	$0.00 \pm 0.00 \pm 0.00$
Diboson	$0.01 \pm 0.01\ ^{+0.01}_{-0.01}$
Single top	$3.37 \pm 1.32\ ^{+0.76}_{-1.00}$
$t\bar{t}V$	$1.59 \pm 0.10\ ^{+0.52}_{-0.54}$
Multijet	$-1.77 \pm 1.78 \pm 0.00$
Total bkg.	$151.66 \pm 7.95\ ^{+30.49}_{-45.71}$
Data	132

Table F.7: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR6 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

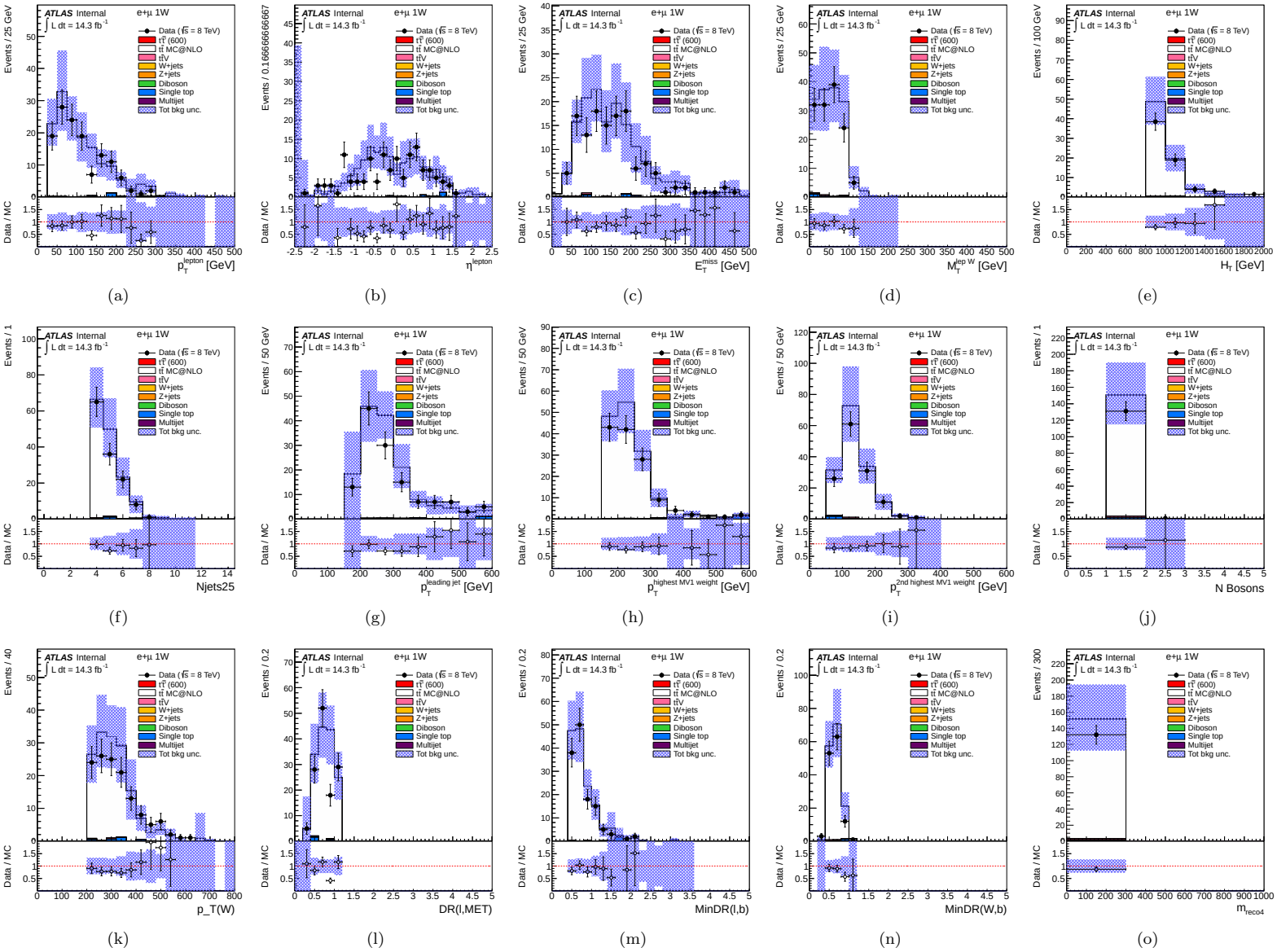


Figure F.7: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR6 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.8 Data to background comparison in SDR7

SDR7: *tight* selection with reversed H_{T} cut (i.e. $H_{\mathrm{T}} < 800$ GeV).

	ELEMUONCR7_1W
$T\bar{T}$(600 GeV) (Chiral)	$0.43 \pm 0.22\ ^{+0.06}_{-0.11}$
$t\bar{t}$	$4.86 \pm 1.51\ ^{+5.07}_{-5.83}$
W+jets	$5.92 \pm 5.60\ ^{+2.15}_{-5.98}$
Z+jets	$0.91 \pm 0.91\ ^{+1.01}_{-1.36}$
Diboson	$0.02 \pm 0.01\ ^{+0.00}_{-0.01}$
Single top	$0.62 \pm 0.55\ ^{+0.47}_{-0.12}$
$t\bar{t}V$	$0.07 \pm 0.02\ ^{+0.03}_{-0.03}$
Multijet	$0.43 \pm 0.41 \pm 0.21$
Total bkg.	$12.82 \pm 5.91\ ^{+5.66}_{-11.03}$
Data	9

Table F.8: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR7 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

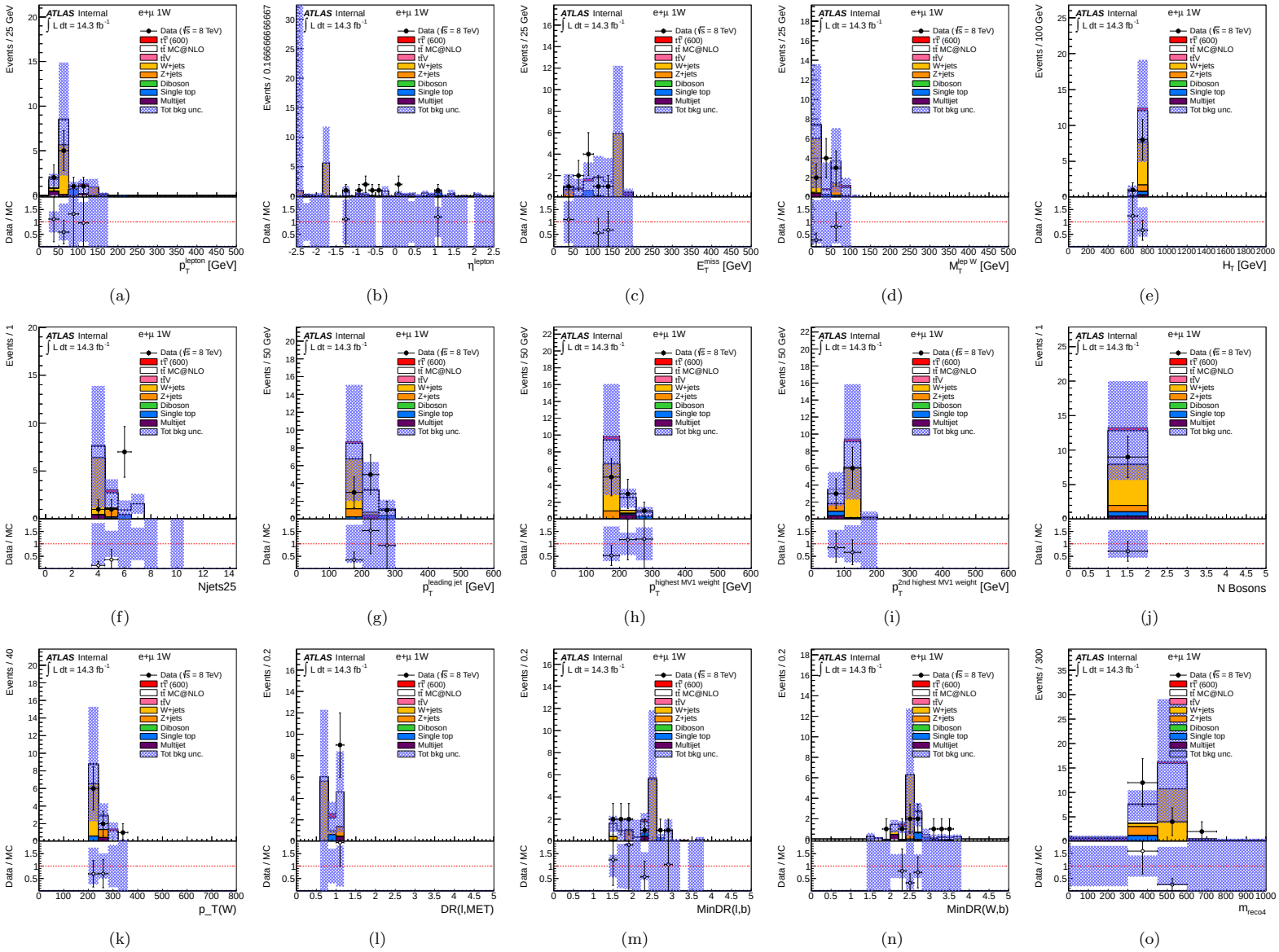


Figure F.8: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR7 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.9 Data to background comparison in SDR8

SDR8: *tight* selection with reversed b-jet p_{T} cuts (i.e. $p_{\mathrm{T}} < 160$ GeV and $p_{\mathrm{T}} < 80$ GeV).

	ELEMUONCR8_1W
$T\bar{T}$(600 GeV) (Chiral)	$5.63 \pm 0.87\ ^{+1.93}_{-2.50}$
$t\bar{t}$	$34.67 \pm 4.20\ ^{+22.00}_{-23.65}$
W+jets	$28.48 \pm 9.21\ ^{+16.58}_{-20.31}$
Z+jets	$0.85 \pm 0.53\ ^{+1.07}_{-1.05}$
Diboson	$0.98 \pm 0.37\ ^{+0.24}_{-0.38}$
Single top	$4.91 \pm 1.46\ ^{+2.01}_{-1.49}$
$t\bar{t}V$	$1.22 \pm 0.08\ ^{+0.44}_{-0.44}$
Multijet	$-0.42 \pm 0.25 \pm 0.00$
Total bkg.	$70.68 \pm 10.26\ ^{+32.96}_{-39.16}$
Data	98

Table F.9: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR8 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

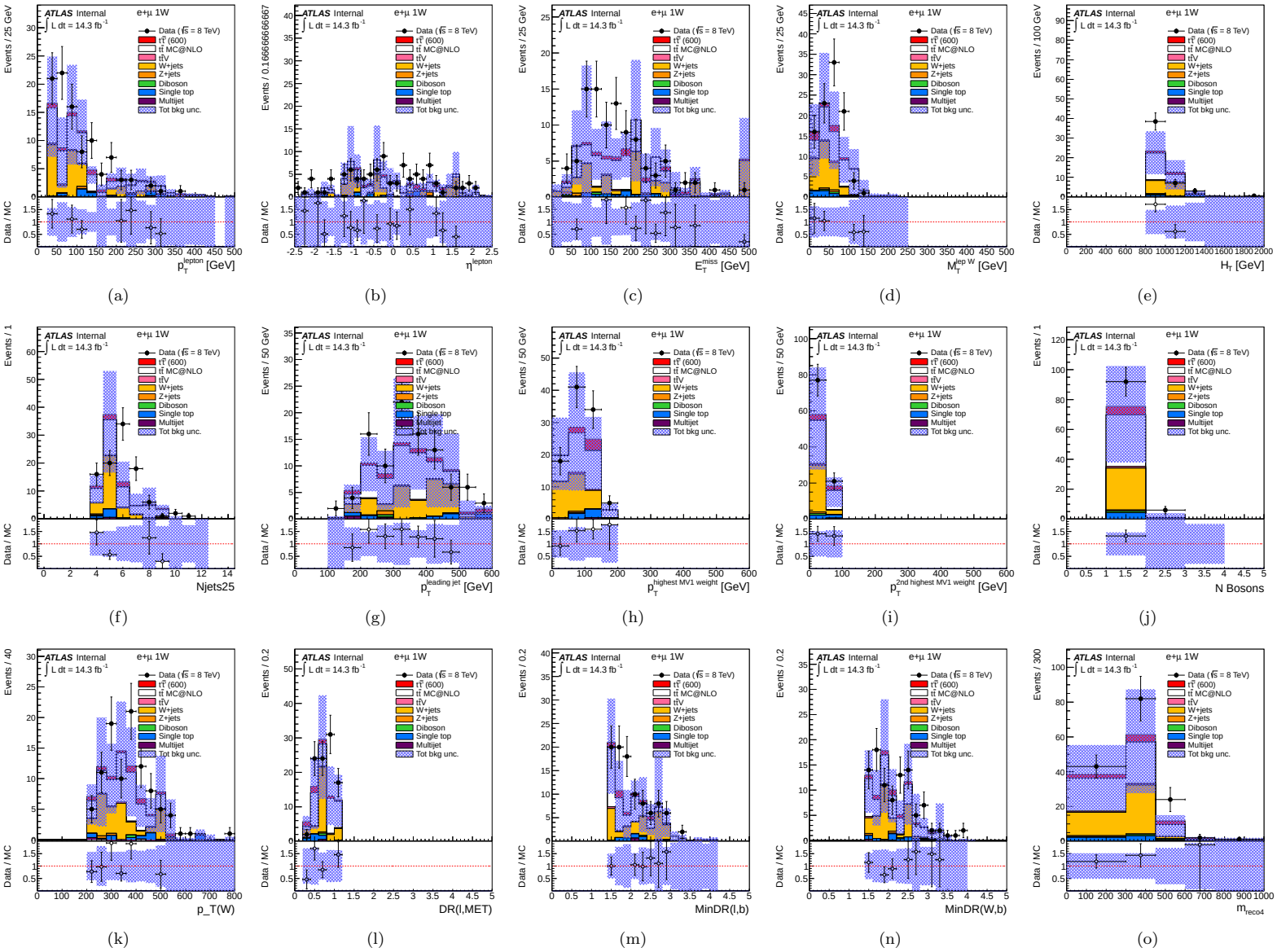


Figure F.9: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR8 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

F.10 Data to background comparison in SDR9

SDR9: *tight* selection with reversed $\Delta R(\ell,\nu)$ cut (i.e. $\Delta R(\ell,\nu) > 1.2$).

	ELEMUONCR9_1W
$T\bar{T}$(600 GeV) (Chiral)	$13.17 \pm 1.25\ ^{+0.77}_{-1.80}$
$t\bar{t}$	$22.86 \pm 3.37\ ^{+9.24}_{-8.59}$
W+jets	$10.48 \pm 5.30\ ^{+6.89}_{-6.64}$
Z+jets	$1.35 \pm 0.69\ ^{+0.76}_{-0.84}$
Diboson	$0.20 \pm 0.18\ ^{+0.17}_{-0.17}$
Single top	$3.37 \pm 1.05\ ^{+2.05}_{-1.63}$
$t\bar{t}V$	$0.82 \pm 0.07\ ^{+0.27}_{-0.27}$
Multijet	$0.50 \pm 0.33 \pm 0.25$
Total bkg.	$39.58 \pm 6.42\ ^{+15.04}_{-13.66}$
Data	64

Table F.10: Number of observed events compared to the SM expectation for the combined e+jets and μ+jets channels in SDR9 (see Section 6.4 for details) . The expected signal yield assuming $m_T = 600$ GeV for the chiral scenario is also shown. The quoted uncertainties include both statistical and systematic contributions.

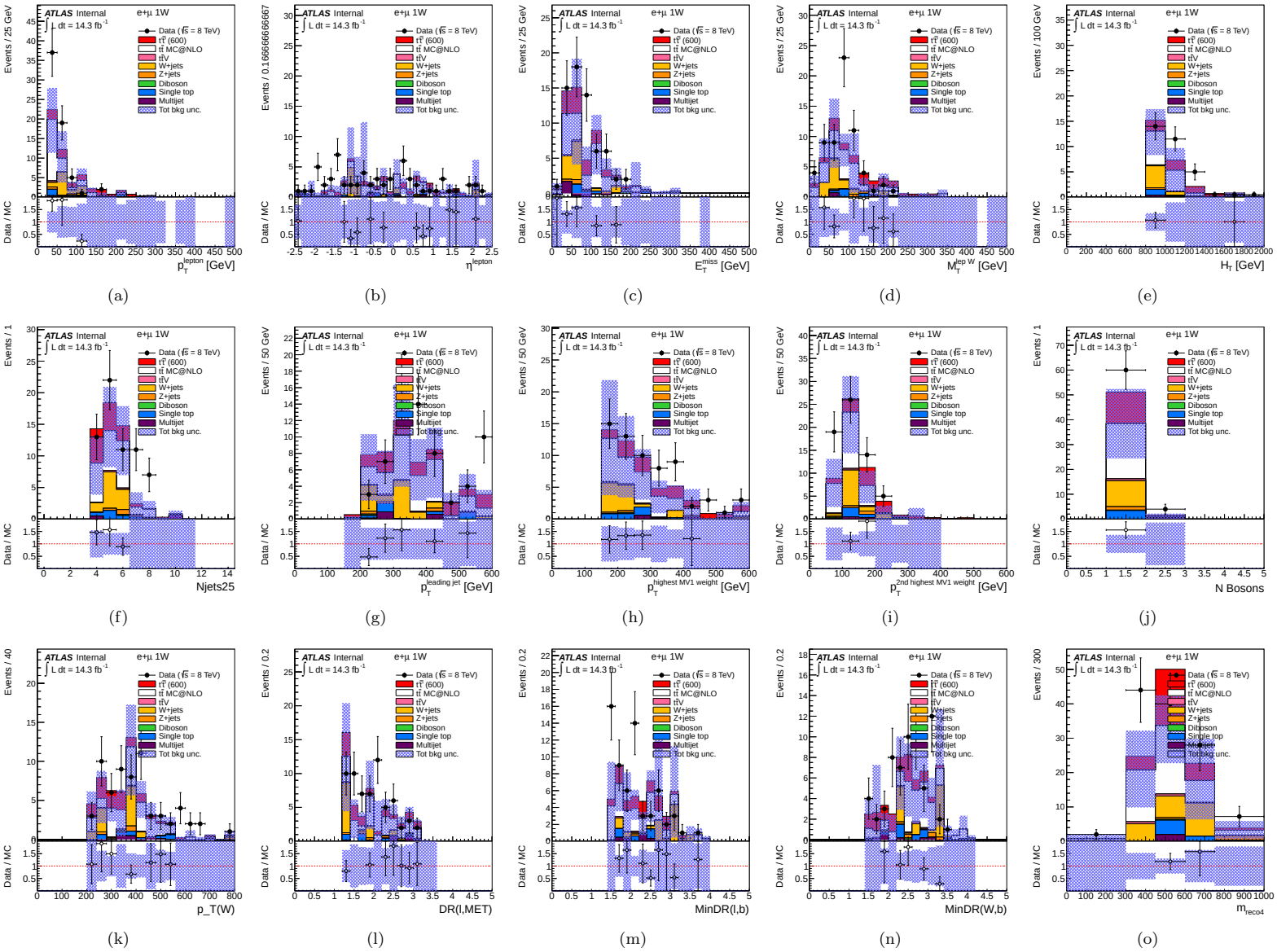
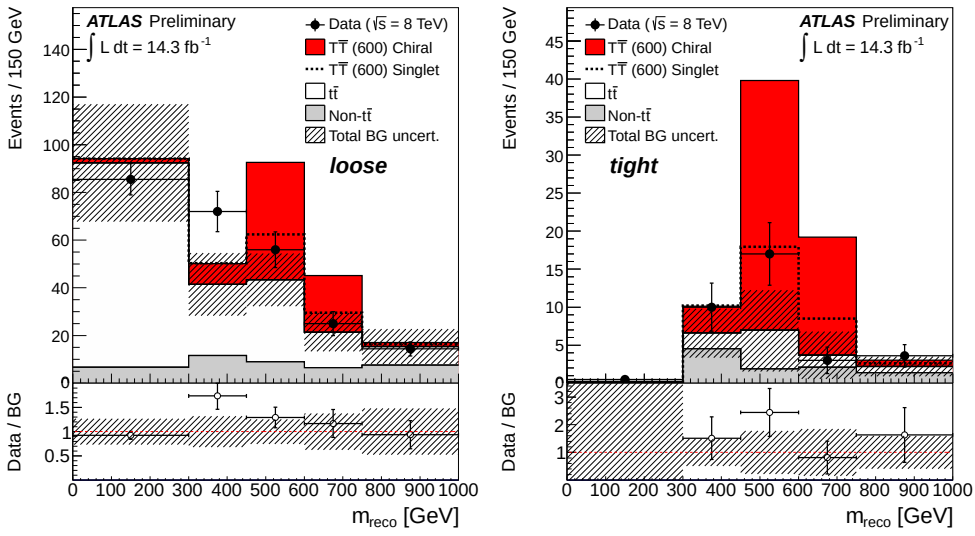


Figure F.10: Comparison between data and prediction in combined e+jets and μ+jets combined channel in SDR9 for a number of kinematic variables: (a) lepton p_{T}, (b) lepton η, (c) missing transverse energy, (d) W transverse mass, (e) H_{T} variable, (f) number of jets with $p_{\mathrm{T}} > 25$ GeV, (g) leading jet p_{T}, (h) p_{T} for leading $b$jet, (i) p_{T} for second-leading $b$jet, (j) number of W_{had} candidates, (k) p_{T} of selected W_{had} candidate, (l) $\Delta R(\ell, \nu)$, (m) $\min(\Delta R(\ell, b_{1,2}))$, (n) $\min(\Delta R(W_{\mathrm{had}}, b_{1,2}))$ and (o) m_{reco}. The shaded area represents the total background uncertainty.

$T\bar{T} \to Wb + X$ analysis: search configuration

In the $T\bar{T} \to Wb + X$ analysis, since no significant excess of data over the expected background has been observed in neither the Loose nor the Tight channels for the m_{reco} spectra of Figure G.1, three configurations have been tested to derive the final results of this search, namely: Loose selection using m_{reco} and profiling of overall $t\bar{t}$ yield ("Loose"); Tight selection using m_{reco} ("Tight"); Tight selection considering just the overall yield and not the shape of m_{reco} ("Tight cut-and-count"). The expected value of CL_{s} (see Section 5.6) as a function of m_T is used to choose the best performing strategy. As was shown in Table 6.7, the prediction for the $t\bar{t}$ background is affected by large systematic uncertainties originating from b-tagged jet identification efficiency, jet energy calibration and resolution and physics modeling in the Monte Carlo generators. In the case of the Loose selection configuration the low-mass sideband region dominated by $t\bar{t}$ can be used to exploit the available data statistics to reduce the degrading impact of systematic uncertainties on the sensitivity of the search. This is accomplished by fitting, during the statistical analysis, a single nuisance parameter representing a scaling factor on the overall $t\bar{t}$ yield. Such procedure is not possible in the case of the Tight selection, where no sidebands are present as its selection is designed to achieve a very high background rejection and high signal-to-background ratio.

As can be seen in Figure G.2, the Tight configuration using m_{reco} distribution information achieves an expected exclusion for a chiral fourth-generation T quark which is ~ 50 GeV higher than the reach of both the Loose and Tight cut-and-count analyses, whose sensitivities are comparable. Therefore, the Tight selection is chosen *a priori* to obtain the main result of this search.



(a) (b)

Figure G.1: Distribution of m_{reco} for the combined e+jets and μ+jets channels after the (a) LOOSE and (b) TIGHT selection. The data (solid black points) are compared to the background prediction from Standard Model (stacked histograms). The total uncertainty on the background estimation (see Section 6.6 for details) is shown as a black hashed band. The expected contribution from a chiral fourth-generation T quark with mass $m_T = 600$ GeV is also shown (red shaded histogram), stacked on top of the Standard Model background. The lower panel shows the ratio of data to Standard Model prediction. The overflow has been added to the last bin.

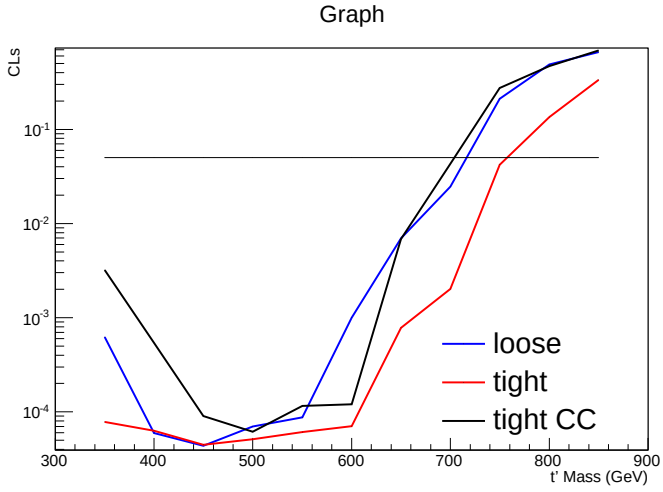


Figure G.2: Expected CL_{s} as a function of m_T taking into account systematic uncertainties. Compared are three possible configurations for the analysis (see text for details).

$T\bar{T} \to Ht + X$ analysis: data to background comparison in control regions

In this appendix the results from the comparison between data and Standard Model backgrounds are outlined, in the electron (Section H.1), muon (Section H.2), and combined channel (Section H.3). The selections applied investigate for different jet multiplicities (4, 5 and at least 6 jets) different b-tagged jet multiplicities (0, 1, 2, 3 and at least 4 b-tagged jets). Each region is blinded to possible signal contribution by means of a cut on the H_{T} variable (H_{T} being defined as the sum of the lepton p_{T}, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the p_{T} of all the jets selected in the event), namely $H_{\mathrm{T}} < 700$ GeV. In each of these sections, tables summarizing the expected and observed yields in the various selections are presented, followed by plots of the distributions for data and Standard Model backgrounds of selected variables.

An additional set of control regions is then presented in Section H.4, were a study was performed in order to identify a suitable selection to investigate the final discriminant variable H_{T} in signal depleted regions.

H.1 e+jets Channel

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	9360.57 ± 27.15	27214.10 ± 62.06	20786.34 ± 51.29	2051.40 ± 7.83	23.08 ± 0.26
$t\bar{t}$+HF	452.59 ± 5.53	1380.94 ± 13.18	1200.53 ± 11.77	255.26 ± 3.64	12.20 ± 0.44
$t\bar{t}V$	19.94 ± 0.23	49.73 ± 0.43	33.61 ± 0.33	4.70 ± 0.07	0.26 ± 0.01
$t\bar{t}H$ (125)	2.77 ± 0.05	8.05 ± 0.10	7.77 ± 0.08	3.10 ± 0.03	0.45 ± 0.01
W+jets	50073.46 ± 458.97	9138.80 ± 140.59	1067.65 ± 38.20	42.70 ± 3.73	1.13 ± 0.69
Z+jets	16011.05 ± 229.06	2289.54 ± 39.92	300.05 ± 10.59	14.14 ± 1.51	0.18 ± 0.03
Diboson	826.95 ± 12.19	173.62 ± 3.31	26.92 ± 1.22	1.48 ± 0.16	0.06 ± 0.05
Single top	1364.50 ± 18.53	3031.69 ± 35.32	1398.67 ± 22.41	113.67 ± 3.56	3.16 ± 0.52
MultiJet	6640.13 ± 51.38	1798.49 ± 27.65	142.71 ± 16.44	13.43 ± 1.46	1.66 ± 0.18
totalBKG	84751.96 ± 516.75	45084.96 ± 165.55	24964.25 ± 71.52	2499.88 ± 10.28	42.18 ± 1.02
$T\bar{T}$ (600)	0.19 ± 0.04	0.91 ± 0.16	0.85 ± 0.15	0.22 ± 0.05	0.06 ± 0.02
Data	84236.00 ± 290.23	40538.00 ± 201.34	25155.00 ± 158.60	2281.00 ± 47.76	42.00 ± 6.48

Table H.1: Event yields as a function of b-tag multiplicity in the $= 4$ jets sample for the e+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	4095.46 ± 17.01	12871.51 ± 42.27	10890.35 ± 37.79	1308.40 ± 6.49	29.97 ± 0.36
$t\bar{t}$+HF	246.49 ± 3.71	837.31 ± 9.86	861.58 ± 10.05	265.70 ± 3.95	28.29 ± 0.80
$t\bar{t}V$	15.39 ± 0.18	42.96 ± 0.39	34.05 ± 0.33	6.95 ± 0.10	0.76 ± 0.03
$t\bar{t}H$ (125)	2.65 ± 0.05	8.74 ± 0.10	9.91 ± 0.10	5.11 ± 0.05	1.31 ± 0.02
W+jets	9807.47 ± 197.64	2239.64 ± 70.08	356.49 ± 25.26	23.47 ± 2.98	1.14 ± 0.56
Z+jets	3599.10 ± 104.67	674.99 ± 24.58	119.92 ± 7.67	8.40 ± 1.06	0.30 ± 0.09
Diboson	126.19 ± 4.82	30.83 ± 1.45	6.61 ± 0.63	0.51 ± 0.09	0.02 ± 0.01
Single top	327.76 ± 8.44	811.17 ± 17.63	484.09 ± 12.92	61.17 ± 2.65	3.35 ± 0.49
MultiJet	1360.58 ± 22.69	432.55 ± 14.82	14.60 ± 10.47	3.65 ± 2.11	0.98 ± 0.57
totalBKG	19581.10 ± 225.68	17949.68 ± 89.06	12777.60 ± 50.03	1683.35 ± 8.90	66.13 ± 1.28
$T\bar{T}$ (600)	0.41 ± 0.12	1.25 ± 0.22	1.29 ± 0.23	0.69 ± 0.14	0.24 ± 0.06
Data	18928.00 ± 137.58	15310.00 ± 123.73	11871.00 ± 108.95	1500.00 ± 38.73	69.00 ± 8.31

Table H.2: Event yields as a function of b-tag multiplicity in the $= 5$ jets sample for the e+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	1654.34 ± 10.48	5456.07 ± 27.49	4917.89 ± 25.72	695.21 ± 4.90	26.16 ± 0.45
$t\bar{t}$+HF	128.12 ± 2.48	477.14 ± 7.33	555.00 ± 8.17	222.05 ± 3.87	39.91 ± 1.22
$t\bar{t}V$	10.95 ± 0.14	34.31 ± 0.34	31.74 ± 0.32	8.47 ± 0.12	1.52 ± 0.05
$t\bar{t}H$ (125)	2.77 ± 0.04	9.89 ± 0.11	12.42 ± 0.12	7.41 ± 0.07	2.75 ± 0.04
W+jets	1900.43 ± 83.87	500.76 ± 30.94	91.75 ± 11.74	9.29 ± 2.06	0.85 ± 0.45
Z+jets	745.83 ± 46.85	151.04 ± 11.97	29.12 ± 4.69	1.90 ± 0.38	0.08 ± 0.02
Diboson	18.22 ± 1.74	5.17 ± 0.59	1.01 ± 0.22	0.13 ± 0.05	0.00 ± 0.00
Single top	76.13 ± 4.30	194.78 ± 9.23	135.89 ± 7.13	20.20 ± 1.67	1.33 ± 0.41
MultiJet	329.36 ± 11.13	108.62 ± 8.61	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
totalBKG	4866.14 ± 97.41	6937.78 ± 45.49	5774.83 ± 30.64	964.66 ± 6.80	72.60 ± 1.44
$T\bar{T}$ (600)	0.49 ± 0.13	1.60 ± 0.24	1.39 ± 0.18	1.03 ± 0.17	0.68 ± 0.24
Data	4964.00 ± 70.46	6140.00 ± 78.36	5351.00 ± 73.15	886.00 ± 29.77	90.00 ± 9.49

Table H.3: Event yields as a function of b-tag multiplicity in the ≥ 6 jets sample for the e+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

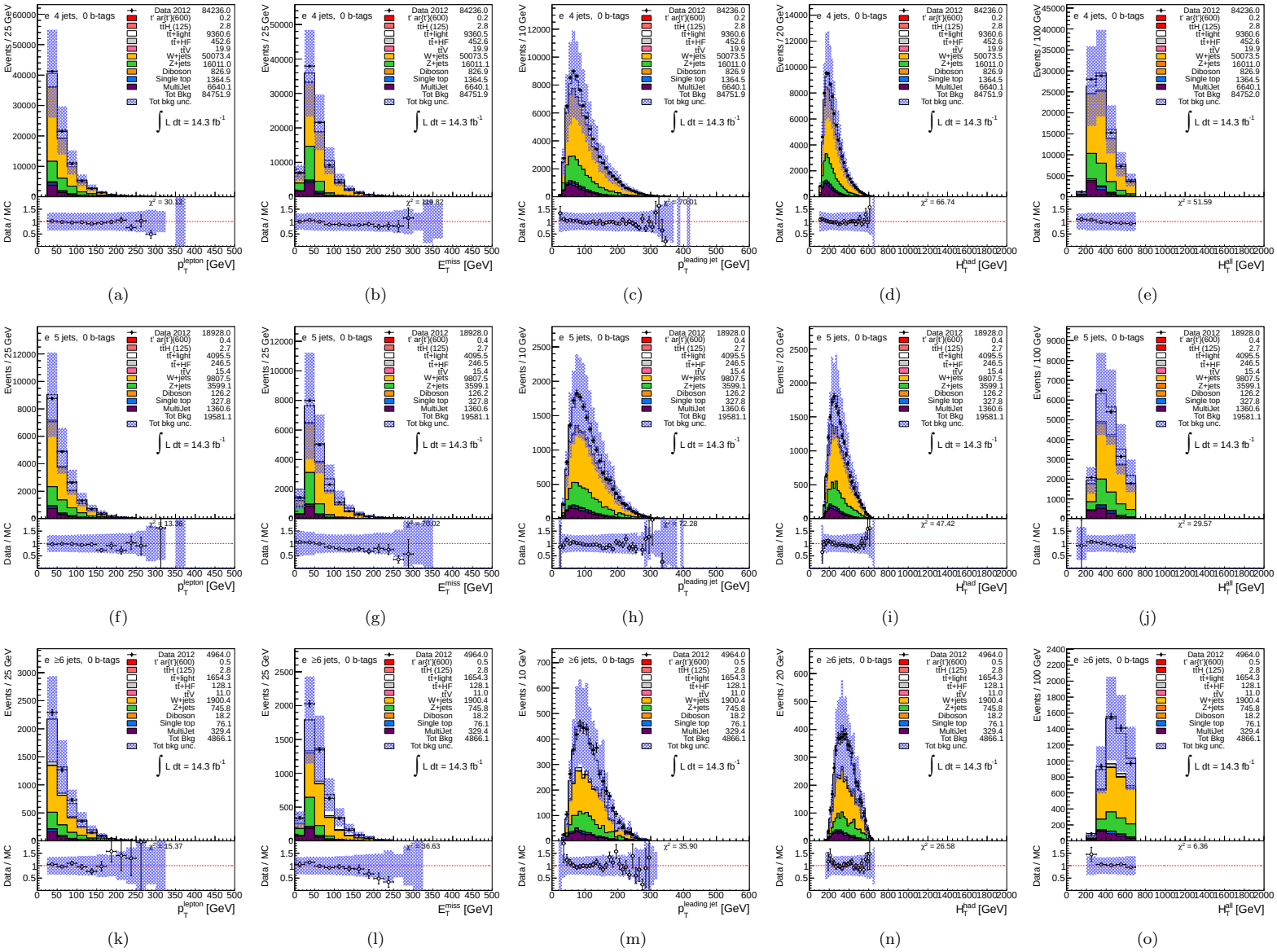


Figure H.1: Comparison between data and prediction in the e+jets channel in the control samples with $= 4$ jets (a–e), $= 5$ jets (f–j), ≥ 4 jets (k–o), and $= 0$ b-tagged jets for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

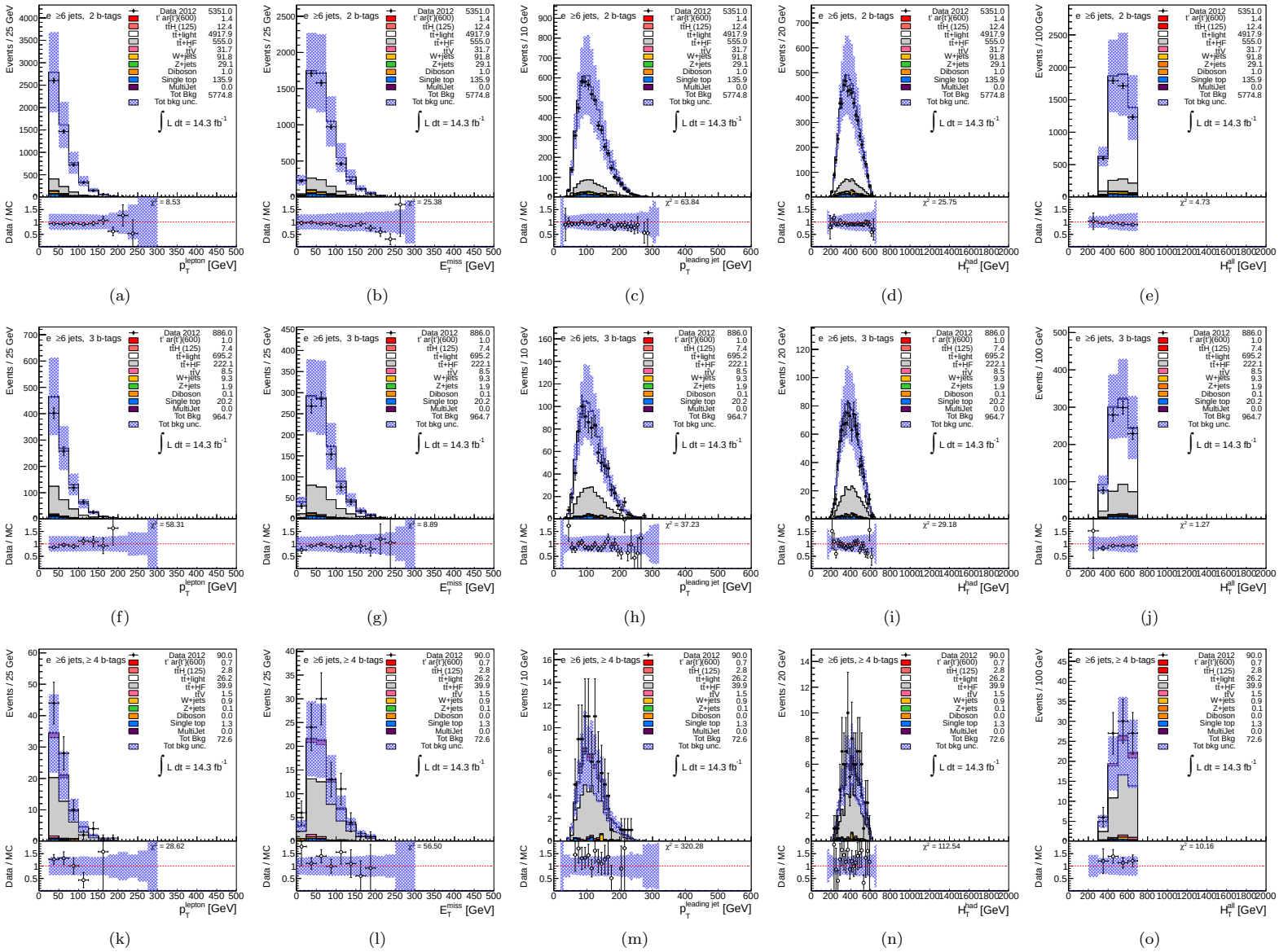


Figure H.2: Comparison between data and prediction in the e+jets channel in the blindend channels "2 b-TAGGED JETS" (a–e), "3 b-TAGGED JETS" jets (f–j), "≥4 b-TAGGED JETS" jets (k–o) for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

H.2 μ+jets Channel

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	11148.15 ± 29.92	32231.47 ± 68.03	24568.32 ± 56.31	2422.31 ± 8.61	27.54 ± 0.26
$t\bar{t}$+HF	542.04 ± 6.23	1629.36 ± 14.41	1408.93 ± 12.85	295.11 ± 3.93	15.20 ± 0.55
$t\bar{t}V$	23.83 ± 0.25	59.32 ± 0.48	39.88 ± 0.35	5.77 ± 0.08	0.32 ± 0.02
$t\bar{t}H$ (125)	3.46 ± 0.06	9.83 ± 0.11	9.47 ± 0.09	3.71 ± 0.04	0.54 ± 0.01
W+jets	76005.82 ± 615.81	13574.63 ± 170.71	1510.73 ± 48.64	78.37 ± 6.67	1.58 ± 0.80
Z+jets	8483.92 ± 161.95	1353.77 ± 32.59	191.53 ± 8.84	8.97 ± 0.81	0.21 ± 0.05
Diboson	931.55 ± 13.24	196.76 ± 3.60	31.16 ± 1.36	1.60 ± 0.16	0.02 ± 0.00
Single top	1595.64 ± 19.94	3493.76 ± 38.63	1598.82 ± 23.94	123.20 ± 3.64	3.08 ± 0.57
MultiJet	5238.75 ± 66.27	2574.56 ± 44.97	266.95 ± 25.06	7.10 ± 0.66	1.07 ± 0.10
totalBKG	103973.18 ± 641.36	55123.47 ± 196.39	29625.80 ± 83.56	2946.13 ± 12.18	49.55 ± 1.16
$T\bar{T}$ (600)	0.49 ± 0.11	1.12 ± 0.20	0.98 ± 0.18	0.52 ± 0.15	0.32 ± 0.14
Data	94459.00 ± 307.34	50232.00 ± 224.12	30427.00 ± 174.43	2757.00 ± 52.51	46.00 ± 6.78

Table H.4: Event yields as a function of b-tag multiplicity in the = 4 jets sample for the μ+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	4901.14 ± 18.80	15353.44 ± 46.69	12911.75 ± 41.52	1553.96 ± 7.16	35.92 ± 0.42
$t\bar{t}$+HF	290.54 ± 4.17	986.83 ± 10.80	1024.05 ± 11.05	320.37 ± 4.43	34.35 ± 0.92
$t\bar{t}V$	18.16 ± 0.20	51.84 ± 0.43	42.36 ± 0.37	8.86 ± 0.11	1.01 ± 0.03
$t\bar{t}H$ (125)	3.26 ± 0.05	10.75 ± 0.12	12.22 ± 0.11	6.17 ± 0.06	1.54 ± 0.02
W+jets	14933.27 ± 270.19	3378.79 ± 87.80	536.34 ± 31.98	39.26 ± 7.40	0.76 ± 0.17
Z+jets	1731.88 ± 72.73	339.37 ± 18.26	59.69 ± 5.99	4.45 ± 0.81	0.12 ± 0.03
Diboson	150.49 ± 5.39	37.37 ± 1.61	6.65 ± 0.59	0.48 ± 0.09	0.02 ± 0.01
Single top	370.68 ± 9.23	935.37 ± 19.22	565.52 ± 13.80	66.19 ± 2.91	3.60 ± 0.58
MultiJet	1140.26 ± 30.80	587.65 ± 23.35	48.24 ± 14.48	5.45 ± 1.47	1.83 ± 0.49
totalBKG	23539.67 ± 282.35	21681.39 ± 106.09	15206.83 ± 57.49	2005.19 ± 11.70	79.15 ± 1.28
$T\bar{T}$ (600)	0.32 ± 0.09	0.74 ± 0.13	0.85 ± 0.15	0.44 ± 0.10	0.08 ± 0.03
Data	21775.00 ± 147.56	19116.00 ± 138.26	14767.00 ± 121.52	1910.00 ± 43.70	91.00 ± 9.54

Table H.5: Event yields as a function of b-tag multiplicity in the = 5 jets sample for the μ+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	1997.84 ± 11.54	6575.23 ± 30.44	5892.83 ± 28.36	838.53 ± 5.46	31.44 ± 0.43
$t\bar{t}$+HF	157.15 ± 2.86	575.60 ± 8.14	667.39 ± 9.13	264.24 ± 4.28	45.91 ± 1.26
$t\bar{t}$V	13.21 ± 0.15	42.08 ± 0.38	39.21 ± 0.36	10.70 ± 0.13	1.90 ± 0.05
$t\bar{t}$H (125)	3.68 ± 0.05	12.69 ± 0.13	15.74 ± 0.13	9.22 ± 0.08	3.36 ± 0.04
W+jets	3290.28 ± 118.40	887.19 ± 45.46	157.58 ± 16.74	24.31 ± 5.57	2.16 ± 0.71
Z+jets	362.23 ± 31.67	98.73 ± 12.09	20.92 ± 3.95	1.87 ± 0.71	0.09 ± 0.04
Diboson	20.63 ± 2.10	5.88 ± 0.65	0.68 ± 0.12	0.04 ± 0.01	0.00 ± 0.00
Single top	80.38 ± 4.28	227.00 ± 9.77	164.43 ± 7.68	26.85 ± 2.01	2.87 ± 0.75
MultiJet	298.47 ± 15.30	179.90 ± 13.78	49.18 ± 10.97	1.11 ± 0.25	0.00 ± 0.00
totalBKG	6223.87 ± 124.18	8604.30 ± 59.09	7007.96 ± 36.91	1176.87 ± 9.15	87.73 ± 1.69
$T\bar{T}$ (600)	0.72 ± 0.15	1.55 ± 0.21	2.34 ± 0.27	1.78 ± 0.25	0.59 ± 0.12
Data	5579.00 ± 74.69	7639.00 ± 87.40	6534.00 ± 80.83	1175.00 ± 34.28	132.00 ± 11.49

Table H.6: Event yields as a function of b-tag multiplicity in the $\geq$ 6 jets sample for the μ+jets channel. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

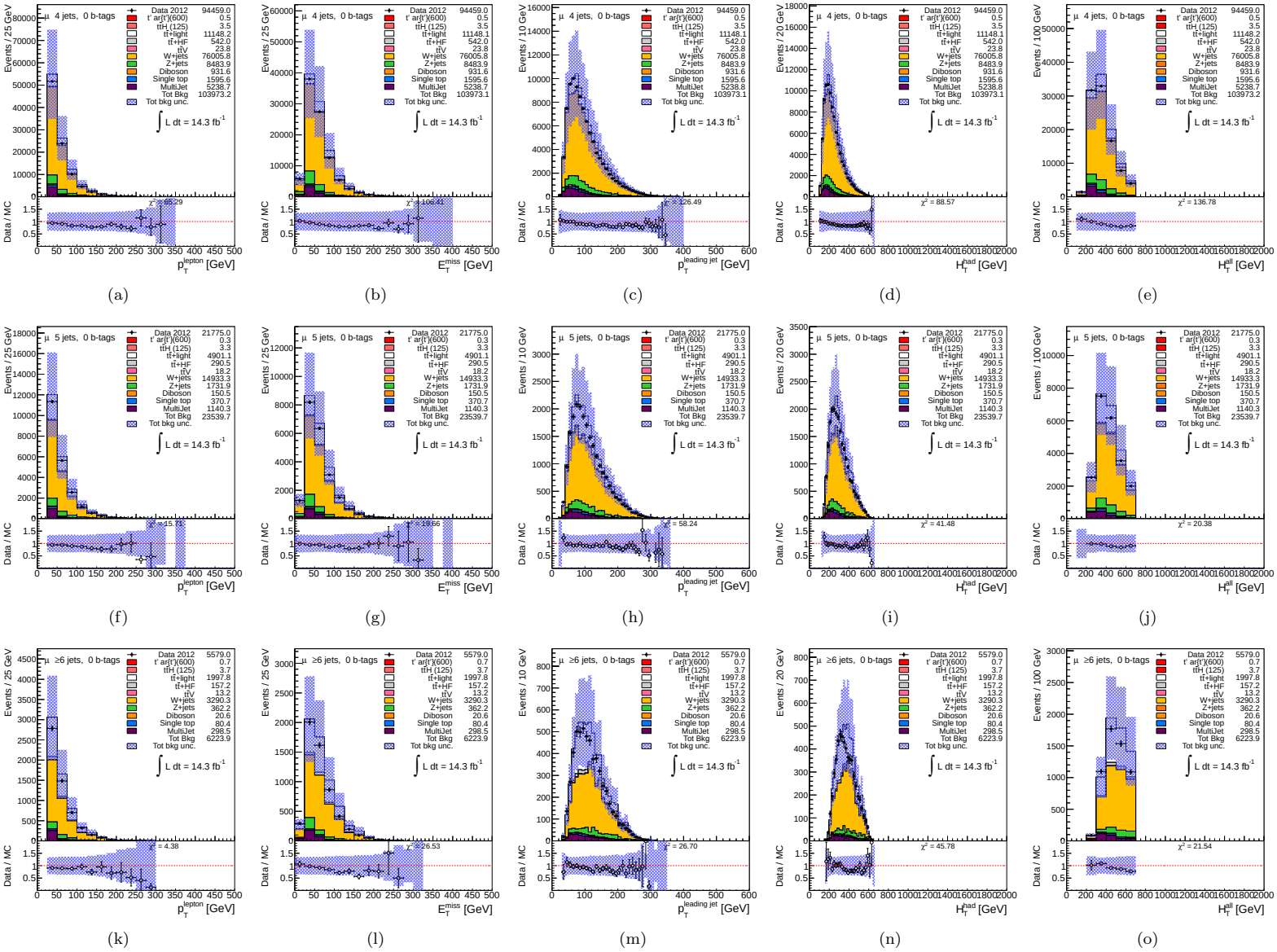


Figure H.3: Comparison between data and prediction in the e+jets channel in the control samples with = 4 jets (a–e), = 5 jets (f–j), $\geq$ 4 jets (k–o), and = 0 b-tagged jets for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

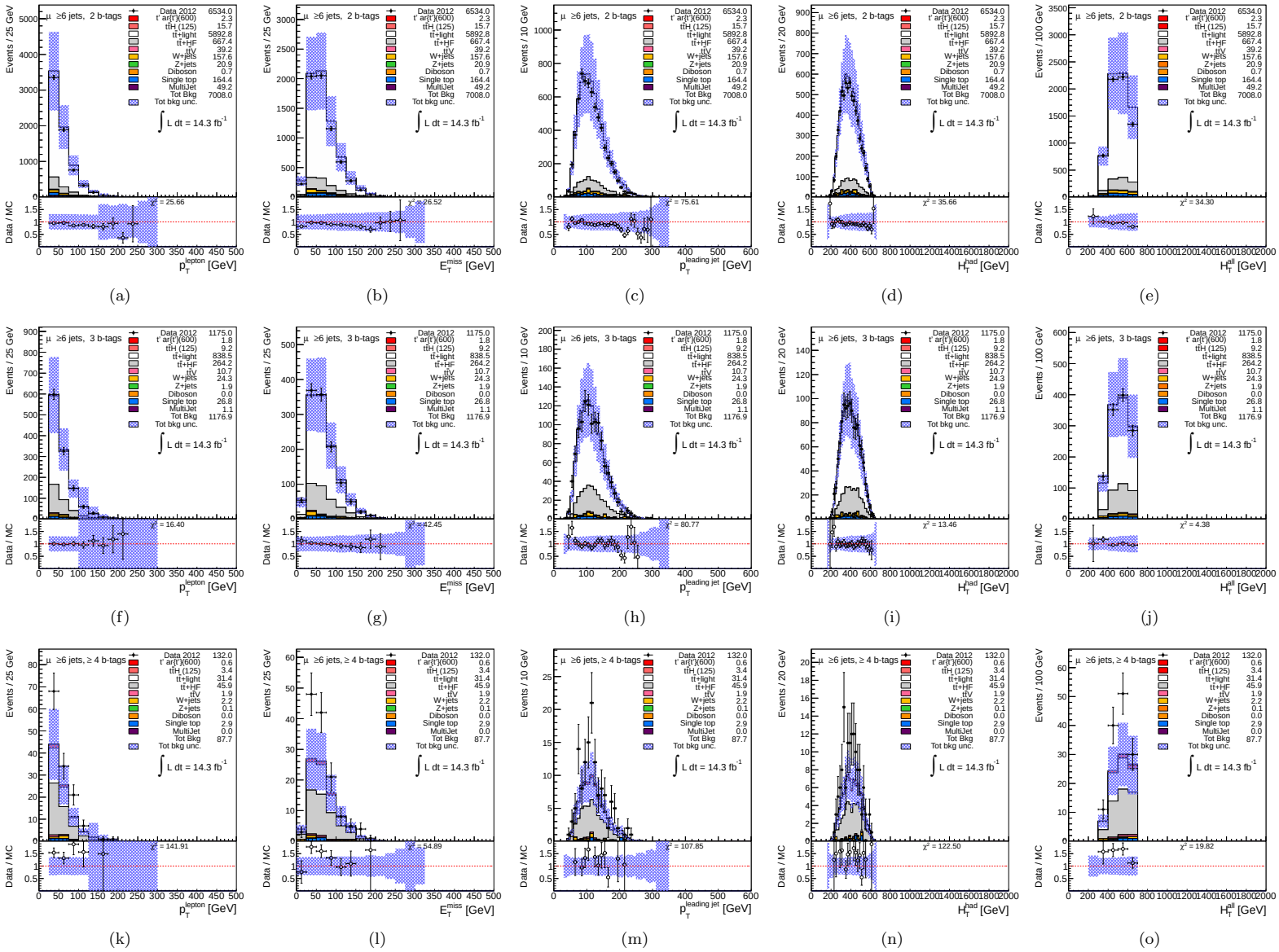


Figure H.4: Comparison between data and prediction in the e+jets channel in the blindend channels "2 b-TAGGED JETS" (a–e), "3 b-TAGGED JETS" jets (f–j), "≥4 b-TAGGED JETS" jets (k–o) for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

H.3 Combined e+jets and μ+jets Channels

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	20508.73 ± 40.40	59445.57 ± 92.09	45354.66 ± 76.16	4473.71 ± 11.63	50.62 ± 0.36
$t\bar{t}$+HF	994.63 ± 8.33	3010.30 ± 19.53	2609.46 ± 17.43	550.36 ± 5.36	27.39 ± 0.70
$t\bar{t}$V	43.77 ± 0.34	109.05 ± 0.64	73.49 ± 0.48	10.47 ± 0.11	0.59 ± 0.02
$t\bar{t}$H (125)	6.23 ± 0.08	17.88 ± 0.15	17.24 ± 0.12	6.81 ± 0.05	0.99 ± 0.01
W+jets	126079.28 ± 768.04	22713.43 ± 221.15	2578.38 ± 61.85	121.08 ± 7.64	2.70 ± 1.05
Z+jets	24494.97 ± 280.53	3643.31 ± 51.53	491.58 ± 13.79	23.11 ± 1.72	0.39 ± 0.06
Diboson	1758.50 ± 18.00	370.39 ± 4.89	58.09 ± 1.83	3.07 ± 0.23	0.08 ± 0.05
Single top	2960.14 ± 27.23	6525.45 ± 52.35	2997.48 ± 32.79	236.87 ± 5.09	6.23 ± 0.78
MultiJet	11878.88 ± 83.85	4373.04 ± 52.79	409.65 ± 29.98	20.53 ± 1.60	2.73 ± 0.21
totalBKG	188725.13 ± 823.63	100208.43 ± 256.86	54590.04 ± 109.99	5446.01 ± 15.94	91.73 ± 1.54
$T\bar{T}$ (600)	0.68 ± 0.12	2.03 ± 0.26	1.83 ± 0.23	0.74 ± 0.16	0.37 ± 0.14
Data	178695.00 ± 422.72	90770.00 ± 301.28	55582.00 ± 235.76	5038.00 ± 70.98	88.00 ± 9.38

Table H.7: Event yields as a function of b-tag multiplicity in the = 4 jets sample for the combined e+jets and μ+jets channels. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	8996.60 ± 25.35	28224.95 ± 62.98	23802.09 ± 56.14	2862.36 ± 9.66	65.89 ± 0.56
$t\bar{t}$+HF	537.03 ± 5.58	1824.13 ± 14.62	1885.64 ± 14.94	586.07 ± 5.94	62.63 ± 1.22
$t\bar{t}$V	33.55 ± 0.27	94.79 ± 0.59	76.41 ± 0.50	15.81 ± 0.15	1.77 ± 0.04
$t\bar{t}$H (125)	5.91 ± 0.07	19.49 ± 0.16	22.13 ± 0.15	11.28 ± 0.08	2.85 ± 0.03
W+jets	24740.74 ± 334.76	5618.42 ± 112.34	892.83 ± 40.75	62.73 ± 7.98	1.90 ± 0.58
Z+jets	5330.98 ± 127.46	1014.36 ± 30.61	179.61 ± 9.73	12.85 ± 1.33	0.43 ± 0.09
Diboson	276.69 ± 7.23	68.20 ± 2.17	13.26 ± 0.87	1.00 ± 0.13	0.05 ± 0.02
Single top	698.44 ± 12.51	1746.54 ± 26.08	1049.61 ± 18.90	127.36 ± 3.94	6.95 ± 0.76
MultiJet	2500.85 ± 38.25	1020.19 ± 27.66	62.84 ± 17.87	9.10 ± 2.57	2.81 ± 0.75
totalBKG	43120.78 ± 361.46	39631.07 ± 138.52	27984.43 ± 76.21	3688.54 ± 14.70	145.27 ± 1.81
$T\bar{T}$ (600)	0.73 ± 0.15	1.99 ± 0.26	2.13 ± 0.28	1.13 ± 0.17	0.32 ± 0.07
Data	40703.00 ± 201.75	34426.00 ± 185.54	26638.00 ± 163.21	3410.00 ± 58.40	160.00 ± 12.65

Table H.8: Event yields as a function of b-tag multiplicity in the = 5 jets sample for the combined e+jets and μ+jets channels. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

	$N_{b-jets} = 0$	$N_{b-jets} = 1$	$N_{b-jets} = 2$	$N_{b-jets} = 3$	$N_{b-jets} \geq 4$
$t\bar{t}$+light	3652.18 ± 15.59	12031.31 ± 41.01	10810.72 ± 38.28	1533.73 ± 7.34	57.60 ± 0.62
$t\bar{t}$+HF	285.28 ± 3.79	1052.74 ± 10.96	1222.40 ± 12.25	486.29 ± 5.77	85.82 ± 1.76
$t\bar{t}V$	24.16 ± 0.21	76.39 ± 0.52	70.95 ± 0.49	19.17 ± 0.18	3.42 ± 0.07
$t\bar{t}H$ (125)	6.46 ± 0.07	22.58 ± 0.17	28.16 ± 0.18	16.64 ± 0.11	6.11 ± 0.05
W+jets	5190.71 ± 145.10	1387.95 ± 54.99	249.33 ± 20.44	33.60 ± 5.94	3.01 ± 0.84
Z+jets	1108.05 ± 56.55	249.76 ± 17.01	50.04 ± 6.14	3.77 ± 0.80	0.16 ± 0.05
Diboson	38.84 ± 2.73	11.05 ± 0.88	1.69 ± 0.25	0.16 ± 0.05	0.00 ± 0.00
Single top	156.51 ± 6.07	421.78 ± 13.44	300.32 ± 10.48	47.05 ± 2.61	4.20 ± 0.86
MultiJet	627.83 ± 18.92	288.52 ± 16.25	49.18 ± 10.97	1.11 ± 0.25	0.00 ± 0.00
totalBKG	11090.01 ± 157.83	15542.08 ± 74.57	12782.79 ± 47.97	2141.53 ± 11.40	160.33 ± 2.22
$T\bar{T}$ (600)	1.21 ± 0.20	3.15 ± 0.32	3.73 ± 0.33	2.81 ± 0.30	1.27 ± 0.26
Data	10543.00 ± 102.68	13779.00 ± 117.38	11885.00 ± 109.02	2061.00 ± 45.40	222.00 ± 14.90

Table H.9: Event yields as a function of b-tag multiplicity in the ≥ 6 jets sample for the combined e+jets and μ+jets channels. Also shown is the expected $T\bar{T}$ signal corresponding to $m_T = 600$ GeV in the T doublet scenario. The uncertainties shown are statistical only.

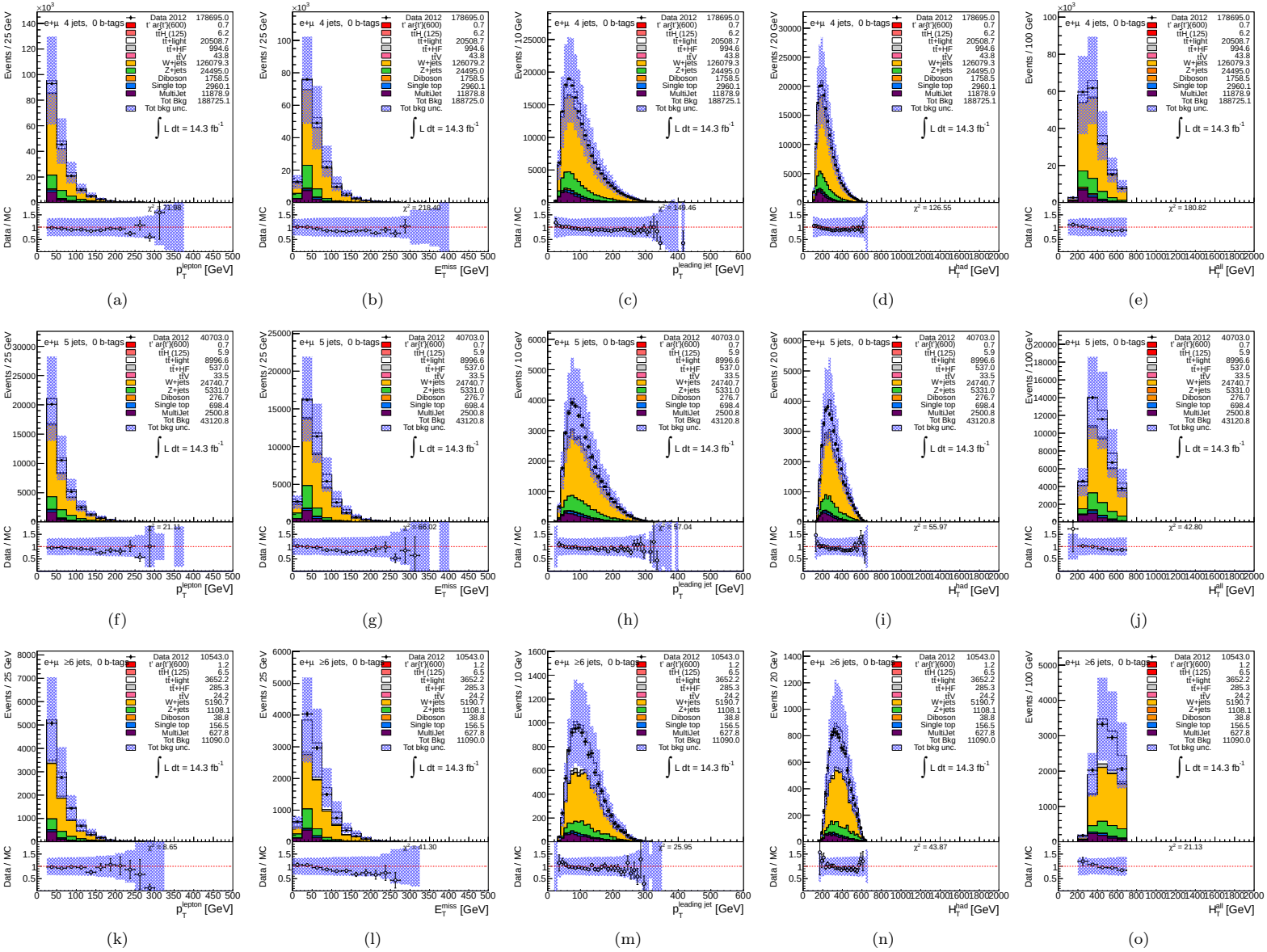


Figure H.5: Comparison between data and prediction in the e+jets channel in the control samples with = 4 jets (a–e), = 5 jets (f–j), $\geq$ 4 jets (k–o), and = 0 b-tagged jets for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

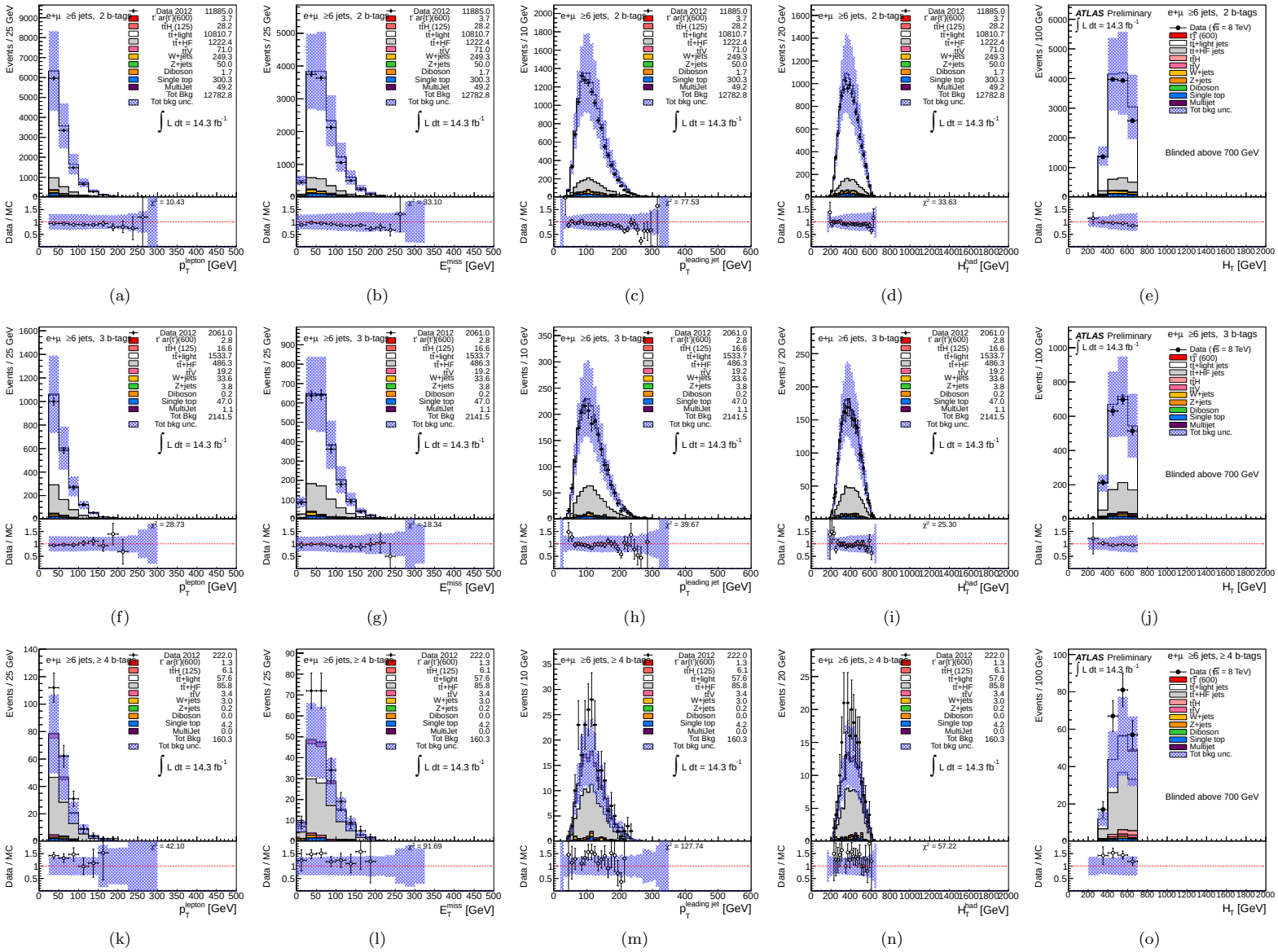


Figure H.6: Comparison between data and prediction in the e+jets channel in the blindend channels "2 b-TAGGED JETS" (a–e), "3 b-TAGGED JETS" jets (f–j), "≥4 b-TAGGED JETS" jets (k–o) for a number of kinematic variables: from left to right, lepton p_{T}, missing transverse energy, leading jet p_{T}, H_T^{had} and H_{T}. The shaded area represents the total background uncertainty.

H.4 H_{T} tail dedicated control regions

A study has been performed to identify a set of cuts that could deplete the signal contribution at high H_{T} but still leave enough background contribution to check the Standard Model bagrounds modeling for the signal region blinded in the control regions previously described. The signal veto cut studied is an upper cut on the number of jets above a given p_{T} threshold, higher than the usual $p_{\mathrm{T}} > 25$ GeV. For different thresholds the S/B ratio was estimated as a function of H_{T} for each of the three analysis channels separately ("2 b-tagged jets", "3 b-tagged jets" and "$\geq$4 b-tagged jets") and for two extreme signal scenarios (vector-like quark doublet and chiral fourth generation quark with $BR(T \to Wb) = 1$, both with $m_T = 600$ GeV), characterized by very different topologies and therefore differently affected by any additional signal veto cuts applied, and a H_{T} region with $S/B \leq 10\%$ per bin and sufficient statistical populetion was identified. It was found that the distribution of number of jets with $p_{\mathrm{T}} > 60$ GeV offers reasonable discrimination between the $t\bar{t}$ background and the two signal scenarios.For a requirement ≤ 2 jets with $p_{\mathrm{T}} > 60$ GeV about 35% of the $t\bar{t}$ background remains while the signal is suppressed by a factor ~ 30. Admittedly, this cut selects a subset of background events where the high H_{T} region will be populated by events with at most two hard jets and possibly large lepton p_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ and high multiplicity of lower-p_{T} jets. Nevertheless, it is useful to be able to probe the high H_{T} tail for at least a subset of the background It was then found that in the "2 b-tagged jets" and "3 b-tagged jets" channels the H_{T} distribution could be examined up to 1.2 TeV with an expected $S/B \leq 10\%$ per bin for either of the two signal scenarios considered. In the case of the "$\geq$4 b-tagged jets" channel, it was not possible to go significantly above the original blinding cut of 700 GeV. Being able to go up to $H_{\mathrm{T}} \sim 1.2$ TeV probes the kinematic region where a T signal with $m_T = 600$ GeV would peak and therefore constitues a useful control region for the background. Therefore, the control region to examine the high H_{T} tail are defined as: selected events with 2 b-tags or 3 b-tags, ≤ 2 jets with $p_{\mathrm{T}} > 60$ GeV, and $H_{\mathrm{T}} < 1.2$ TeV. The comparisons between data and prediction (using scaled $t\bar{t}$ Alpgen) in this control region, including the expected pre-fit systematic uncertainties, are shown in Figure H.8.

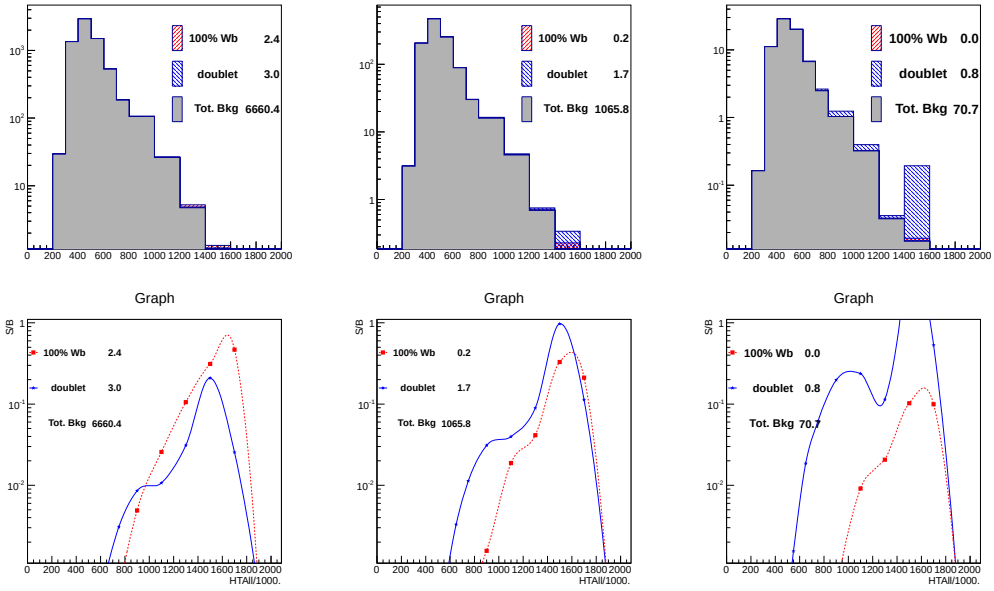
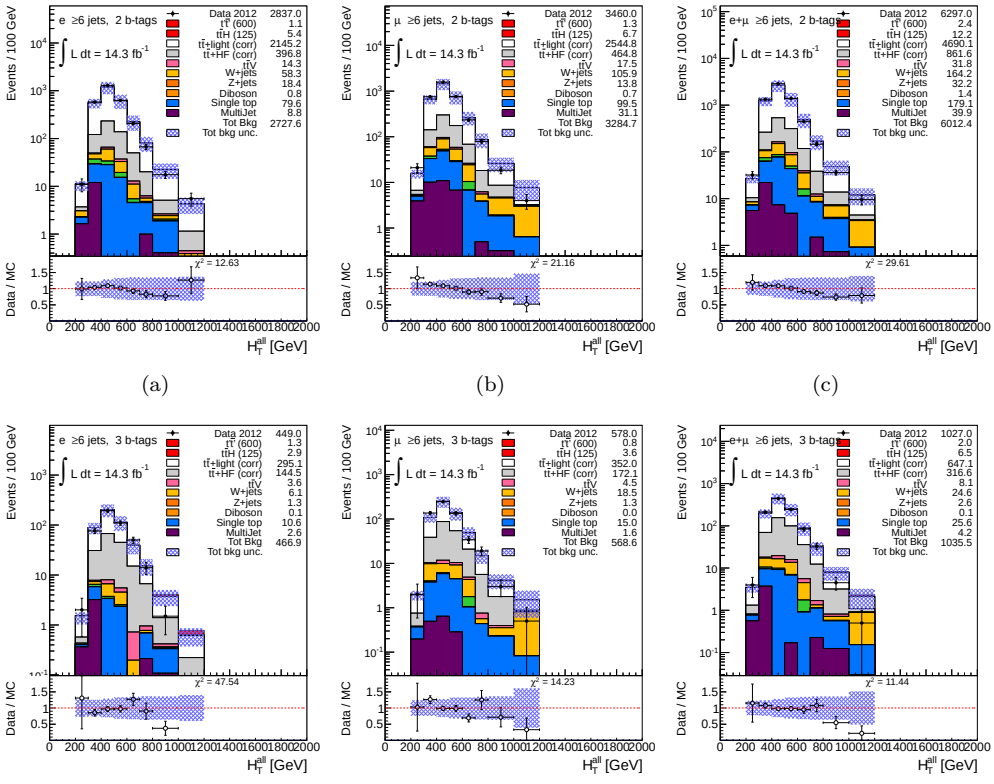


Figure H.7: Top: Expected H_{T} distribution for events after final selection and including the requirement of ≤ 2 jets with $p_{\mathrm{T}} > 60$ GeV, in the (left) 2 b-tags, (center) 3 b-tags and (right) ≥ 4 b-tags channels. Shown is the spectrum for total background background as well as the expected signal in the two scenarios considered (see text for details). Bottom: Expected S/B as a function of H_{T} after final selection and including the requirement of ≤ 2 jets with $p_{\mathrm{T}} > 60$ GeV, in the (left) 2 b-tags, (center) 3 b-tags and (right) ≥ 4 b-tags channels. The two signal scenarios considered are shown.



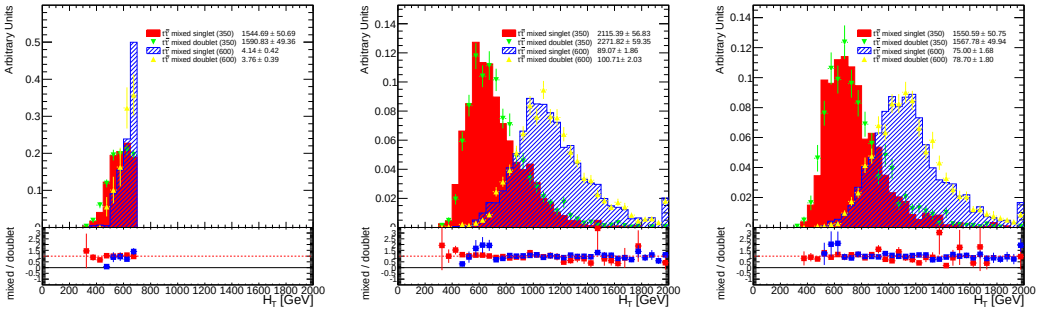
(a) (b) (c)

(d) (e) (f)

Figure H.8: Data to Standard Model comparison for the H_{T} variable in the e+jets, μ+jets and combined channels, in the control regions specified in the text requiring 2 b-tagged jets (a–c) and 3 b-tagged jets (d–f).

$T\bar{T} \to Ht + X$ analysis: comparison of signal prediction between singlet and doublet scenarios

As discussed in Sect. **??**, the signal MC samples used in this analysis were generated in the singlet scenario, for which the mixing between the T quark and SM quarks is left-handed. For simplicity, these samples are reweighted to reproduce any desired branching-ratio configuration, included that corresponding to a doublet scenario. There is a small concern that for the latter the mixing between the T quark and SM quarks is right-handed, slightly affecting the kinematics and thus the signal acceptance and shape of the H_{T} distribution. Two MC samples for the doublet scenario, corresponding to $m_T = 350$ and 600 GeV, are available, which have been used to check this effect. Figure I.1 compares, for both mass points and each of the three analysis channels considered, the yield and shape of the H_{T} distribution for the predicted signal using the singlet and doublet samples. In both cases the samples were reweighted to reproduce the branching ratios corresponding to the doublet model. As it can be appreciated, the shapes of the distributions are in reasonable agreement and discrepancies in the yields are below 5% in the highest-sensitivity channel (≥ 4 b tags).



(a) (b) (c)

Figure I.1: Comparison of the yields and shape of the H_{T} distribution in simulation for T signal using the singlet samples (symbols) and using the doublet samples (histogram). In both cases the signal has been reweighted to reproduce the branching ratios corresponding to the doublet model. The selection used corresponds to the combined e+jets and μ+jets channels with ≥ 6 jets and (a) 2 b tags, (b) 3 b-tags and (c) ≥ 4 b tags. The comparison is made for two different mass points, $m_T = 350$ and 600 GeV. The last bin in all figures contains the overflow.

Bibliography

[1] ATLAS Collaboration, *Search for pair production of heavy top-like quarks decaying to a high-pT W boson and a b quark in the lepton plus jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2013-060 (Jun, 2013) .

[2] ATLAS collaboration, *Search for heavy top-like quarks decaying to a Higgs boson and a top quark in the lepton plus jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2013-018 (Mar, 2013) .

[3] ATLAS Collaboration, *Search for pair production of heavy top-like quarks decaying to a high-pT W boson and a b quark in the lepton plus jets final state at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Lett. **B718** (2012) 1284–1302, arXiv:1210.5468 [hep-ex].

[4] J. C. Street and E. C. Stevenson, *New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron*, Ph http://link.aps.org/doi/10.1103/PhysRev.52.1003.

[5] H. Pietschmann, *Neutrino: Past, Present and Future*, ArXiv Physics e-prints (Mar., 2006) , physics/0603106.

[6] C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, *Detection of the Free Neutrino: A Confirmation*, Science **124** (July, 1956) 103–104.

[7] G. Danby, J.-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger, *Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos*, Physical

[8] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964) no. 3, 214 – 215. http://www.sciencedirect.com/science/article/pii/S0031916364920013.

[9] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*, Tech. Rep. CERN-TH-401, CERN, Geneva, Jan, 1964.

[10] V. E. Barnes, P. L. Connolly, D. J. Crennell, B. B. Culwick, W. C. Delaney, W. B. Fowler, P. E. Hagerty, E. L. Hart, N. Horwitz, P. V. C. Hough, J. E. Jensen, J. K. Kopp, K. W. Lai, J. Leitner, J. L. Lloyd, G. W. London, T. W. Morris, Y. Oren, R. B. Palmer, A. G. Prodell, D. Radojičić, D. C. Rahm, C. R. Richardson, N. P. Samios, J. R. Sanford, R. P. Shutt, J. R. Smith, D. L. Stonehill, R. C. Strand, A. M. Thorndike, M. S. Webster, W. J. Willis, and S. S. Yamamoto, *Observation of a Hyperon with Strangeness Minus Three*, Phys. Rev. Lett. **12** (Feb, 1964) 204–206. http://link.aps.org/doi/10.1103/PhysRevLett.12.204.

[11] E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor, M. Breidenbach, J. I. Friedman, G. C. Hartmann, and H. W. Kendall, *High-Energy Inelastic $e-p$ Scattering at 6° and 10°*, Phys. Rev. Lett. **23** (Oct, 1969) 930–934. http://link.aps.org/doi/10.1103/PhysRevLett.23.930.

[12] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, L. W. Mo, and R. E. Taylor, *Observed Behavior of Highly Inelastic Electron-Proton Scattering*, Phys. Rev. Lett. **23** (Oct, 1969) 935–939. http://link.aps.org/doi/10.1103/PhysRevLett.23.935.

[13] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (Oct, 1954) 191–195. http://link.aps.org/doi/10.1103/PhysRev.96.191.

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (Aug, 1964) 321–323. http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[15] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (Oct, 1964) 508–509. http://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[16] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) no. 4, 579 – 588. http://www.sciencedirect.com/science/article/pii/0029558261904692.

[17] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (Nov, 1967) 1264–1266. http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[18] G. 'tHooft, *Renormalization of massless Yang-Mills fields*, Nuclear Physics B **33** (1971) no. 1, 173 – 199. http://www.sciencedirect.com/science/article/pii/0550321371903956.

[19] G. Hooft, *Renormalizable Lagrangians for massive Yang-Mills fields*, Nuclear Physics B **35** (1971) no. 1, 167 – 188. http://www.sciencedirect.com/science/article/pii/0550321371901398.

[20] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (Jun, 1963) 531–533. http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[21] S. L. Glashow, J. Iliopoulos, and L. Maiani,
Weak Interactions with Lepton-Hadron Symmetry, Phys. Rev. D **2** (Oct, 1970) 1285–1292.
http://link.aps.org/doi/10.1103/PhysRevD.2.1285.

[22] R. G. Sachs, CP *Violation in K^0 Decays*, Phys. Rev. Lett. **13** (Aug, 1964) 286–288.
http://link.aps.org/doi/10.1103/PhysRevLett.13.286.

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

[24] J. J. Aubert, U. Becker, P. J. Biggs, J. Burger, M. Chen, G. Everhart, P. Goldhagen, J. Leong, T. McCorriston, T. G. Rhoades, M. Rohde, S. C. C. Ting, S. L. Wu, and Y. Y. Lee,
Experimental Observation of a Heavy Particle J, Phys. Rev. Lett. **33** (Dec, 1974) 1404–1406.
http://link.aps.org/doi/10.1103/PhysRevLett.33.1404.

[25] J. E. Augustin, A. M. Boyarski, M. Breidenbach, F. Bulos, J. T. Dakin, G. J. Feldman, G. E. Fischer, D. Fryberger, G. Hanson, B. Jean-Marie, R. R. Larsen, V. Lüth, H. L. Lynch, D. Lyon, C. C. Morehouse, J. M. Paterson, M. L. Perl, B. Richter, P. Rapidis, R. F. Schwitters, W. M. Tanenbaum, F. Vannucci, G. S. Abrams, D. Briggs, W. Chinowsky, C. E. Friedberg, G. Goldhaber, R. J. Hollebeek, J. A. Kadyk, B. Lulu, F. Pierre, G. H. Trilling, J. S. Whitaker, J. Wiss, and J. E. Zipse,
Discovery of a Narrow Resonance in e^+e^- Annihilation, Phys. Rev. Lett. **33** (Dec, 1974) 1406–1408.
http://link.aps.org/doi/10.1103/PhysRevLett.33.1406.

[26] S. W. Herb, D. C. Hom, L. M. Lederman, J. C. Sens, H. D. Snyder, J. K. Yoh, J. A. Appel, B. C. Brown, C. N. Brown, W. R. Innes, K. Ueno, T. Yamanouchi, A. S. Ito, H. Jöstlein, D. M. Kaplan, and R. D. Kephart,
Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions, Phys. Rev. Lett
http://link.aps.org/doi/10.1103/PhysRevLett.39.252.

[27] M. L. Perl, G. S. Abrams, A. M. Boyarski, M. Breidenbach, D. D. Briggs, F. Bulos, W. Chinowsky, J. T. Dakin, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Goldhaber, G. Hanson, F. B. Heile, B. Jean-Marie, J. A. Kadyk, R. R. Larsen, A. M. Litke, D. Lüke, B. A. Lulu, V. Lüth, D. Lyon, C. C. Morehouse, J. M. Paterson, F. M. Pierre, T. P. Pun, P. A. Rapidis, B. Richter, B. Sadoulet, R. F. Schwitters, W. Tanenbaum, G. H. Trilling, F. Vannucci, J. S. Whitaker, F. C. Winkelmann, and J. E. Wiss,
Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation, Phys. Rev. Lett. **35** (Dec, 1975) 14
http://link.aps.org/doi/10.1103/PhysRevLett.35.1489.

[28] UA1 Collaboration Collaboration, G. A. *et al.*, *Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV*, Physics Letters B **122** (1983) no. 1, 103 – 116.
http://www.sciencedirect.com/science/article/pii/0370269383911772.

[29] UA2 Collaboration Collaboration, M. B. *et al.*, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at*

the {CERN} pp collider, Physics Letters B **122** (1983) no. 56, 476 – 485. http://www.sciencedirect.com/science/article/pii/0370269383916052.

[30] P. Renton, *Global Electroweak Fits and the Higgs Boson Mass*, arXiv:0809.4566 [hep-ph].

[31] CDF Collaboration Collaboration, F. Abe and et al., *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett http://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[32] D0 Collaboration Collaboration, S. Abachi and et al., *Search for High Mass Top Quark Production in $p\overline{p}$ Collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **74** (Mar, 19 http://link.aps.org/doi/10.1103/PhysRevLett.74.2422.

[33] 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. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[34] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the {CMS} experiment at the {LHC}*, Physics Letters B **716** (2012) no. 1, 30 – 61. http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[35] M. M. H. Luk and U. Heintz, *The Search for a Heavy Top-Like Quark*. PhD thesis, Brown U., 2013.

[36] G. Ecker, *Quantum ChromoDynamics*, 2006. http://arxiv.org/abs/hep-ph/0604165v1.

[37] S. Bethke Eur. Phys. J. **C64** (2009) 689, arXiv:09081135 [hep-ex].

[38] P. Langacker, *Neutrino Physics (theory)*, Int. J. Mod. Phys. A **20** (2004) 5254. http://arxiv.org/abs/hep-ph/0411116v1.

[39] Planck Collaboration Collaboration, P. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, arXiv:1303.5076 [astro-ph.CO].

[40] S. Dawson, *SUSY and such*, NATO Adv.Study Inst.Ser.B Phys. **365** (1997) 33–80, arXiv:hep-ph/9612229 [hep-ph].

[41] S. P. Martin, *A Supersymmetry primer*, arXiv:hep-ph/9709356 [hep-ph].

[42] B. Holdom, W. Hou, T. Hurth, M. Mangano, S. Sultansoy, et al., *Four Statements about the Fourth Generation*, PMC Phys. **A3** (2009) 4, arXiv:0904.4698 [hep-ph].

[43] H. Lacker and A. Menzel, *Constraints on CKM matrix elements within a fourth generation*, Taipei, Second Workshop on Beyond 3 Generation Standard Model (2010) .

[44] A. Djouadi and A. Lenz, *Sealing the fate of a fourth generation of fermions*, Phys.Lett. **B715** (2012) 310–314, arXiv:1204.1252 [hep-ph].

[45] S. Cetin, G. Hou, V. Ozcan, A. Rozanov, and S. Sultansoy, *Status of the Fourth Generation: A Brief Summary of B3SM-III Workshop in Four Parts*, arXiv:1112.2907 (2011) , arXiv:1112.2907 [hep-ph].

[46] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (Aug, 1962) 965–970. http://link.aps.org/doi/10.1103/PhysRev.127.965.

[47] L. Vecchi, *The Natural Composite Higgs*, arXiv:1304.4579 [hep-ph].

[48] F. del Aguila, A. Carmona, and J. Santiago, *Neutrino Masses from an A4 Symmetry in Holographic Composite Higgs Models*, JHEP **1008** (2010) 127, arXiv:1001.5151 [hep-ph].

[49] R. Contino and G. Servant, *Discovering the top partners at the LHC using same-sign dilepton final states*, JHEP **0806** (2008) 026, arXiv:0801.1679 [hep-ph].

[50] J. Mrazek and A. Wulzer, *A Strong Sector at the LHC: Top Partners in Same-Sign Dileptons*, Phys.Rev. **D81** (2010) 075006, arXiv:0909.3977 [hep-ph].

[51] T. Kaluza, *Zum Unitätsproblem in der Physik*, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **966** (1921) . https://archive.org/details/sitzungsberichte1921preussi.

[52] O. Klein, *Quantentheorie und fnfdimensionale Relativittstheorie*, Zeitschrift fr Physik **37** (1926) no. 12, 895–906. http://dx.doi.org/10.1007/BF01397481.

[53] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, Phys.Lett. **B429** (1998) 263–272, arXiv:hep-ph/9803315 [hep-ph].

[54] J. Aguilar-Saavedra, *Identifying top partners at LHC*, JHEP **0911** (2009) 030, arXiv:0907.3155 [hep-ph].

[55] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, *Electroweak symmetry breaking from dimensional deconstruction*, Phys.Lett. **B513** (2001) 232–240, arXiv:hep-ph/0105239 [hep-ph].

[56] M. Perelstein, *Little Higgs models and their phenomenology*, Prog.Part.Nucl.Phys. **58** (2007) 247–291, arXiv:hep-ph/0512128 [hep-ph].

[57] C. Csaki, *TASI lectures on extra dimensions and branes*, arXiv:hep-ph/0404096 [hep-ph].

[58] F. del Aguila and M. J. Bowick, *The Possibility of New Fermions With Δ $I = 0$ Mass*, Nucl.Phys. **B224** (1983) 107.

[59] F. del Aguila, J. Aguilar-Saavedra, and R. Miquel, *Constraints on top couplings in models with exotic quarks*, Phys.Rev.Lett. **82** (1999) 1628–1631, arXiv:hep-ph/9808400 [hep-ph].

[60] J. Aguilar-Saavedra, *Effects of mixing with quark singlets*, Phys.Rev. **D67** (2003) 035003, arXiv:hep-ph/0210112 [hep-ph].

[61] A. Atre, M. Carena, T. Han, and J. Santiago, *Heavy Quarks Above the Top at the Tevatron*, Phys.Rev. **D79** (2009) 054018, arXiv:0806.3966 [hep-ph].

[62] J. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Perez-Victoria, *A handbook of vector-like quarks: mixing and single production*, arXiv:1306.0572 [hep-ph].

[63] N. Kidonakis and B. D. Pecjak, *Top-quark production and QCD*, Eur.Phys.J. **C72** (2012) 2084, arXiv:1108.6063 [hep-ph].

[64] ATLAS Collaboration, *Search for pair production of new heavy quarks that decay to a* $\mathbf{Z}$ *boson and a third generation quark in* $\mathbf{pp}$ *collisions at* $\sqrt{\mathbf{s}} = \mathbf{8}$ *TeV with the ATLAS detector*, ATLAS-CONF-2013-056 (Jun, 2013) .

[65] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[66] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[67] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[68] LHCb Collaboration, *The LCHb experiment at the CERN LHC*, JINST **3** (2008) S08005.

[69] L. Evans, P. Bryant (Eds.), *LHC Machine*, JINST **3** (2008) S08001.

[70] M. Lamont, *The First Years of LHC Operation for Luminosity Production*, in *Proceedings of 4th International Particle Accelerator Conference (IPAC 2013)* (2013) .

[71] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 7$ TeV *using the ATLAS detector at the LHC*, Eur. Phys. J. C **71** (2011) 1630, arXiv:1101.2185 [hep-ex].

[72] ATLAS Collaboration, *Studies of the performance of the ATLAS detector using cosmic-ray muons*, Eur. Phys. J. **C 71** (2011) 1593.

[73] W. R., *Calorimetry*, Scientifica Acta **2 (1)** (2008) 1855.

[74] ATLAS Collaboration, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, Eur. Phys. J. **C 70** (2010) 723–753.

[75] ATLAS Collaboration, *Readiness of the ATLAS Tile calorimeter for LHC collisions*, Eur. Phys. J. **C 70** (2010) 11931236.

[76] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[77] R. P. Feynman, *Very high-energy collisions of hadrons*, Phys.Rev.Lett. **23** (1969) 1415–1417.

[78] J. M. Campbell, J. Huston, and W. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, Rept.Prog.Phys. **70** (2007) 89, hep-ph/0611148.

[79] M. L. Mangano and T. J. Stelzer, *Tools for the simulation of hard hadronic collisions*, Annu. Rev. Nucl. Part. Sci. **55** (2005) no. CERN-PH-TH-2005-074, 555–588.

[80] S. Gieseke, *Parton Shower Monte Carlos*, . http://th-workshop2009.desy.de/e59393/e59379/infoboxContent59381/Gieseke.pdf.

[81] F. Ambroglini, R. Armillis, P. Azzi, G. Bagliesi, A. Ballestrero, et al., *Proceedings of the Workshop on Monte Carlo's, Physics and Simulations at the LHC. Part I*, arXiv:0902.0293 [hep-ph].

[82] S. Catani, F. Krauss, R. Kuhn, and B. Webber, *QCD matrix elements + parton showers*, JHEP **0111** (2001) 063, arXiv:hep-ph/0109231 [hep-ph].

[83] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, JHEP **0701** (2007) 013, arXiv:hep-ph/0611129 [hep-ph].

[84] S. Catani, Y. Dokshitzer, M. Olsson, G. Turnock, and B. Webber, *New clustering algorithm for multijet cross sections in e+e annihilation*, Physics Letters B **269** (1991) no. 34, 432 – 438. http://www.sciencedirect.com/science/article/pii/037026939190196W.

[85] S. M. T. Sjostrand and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[86] G. Corcella *et al.*, *HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010, arXiv:hep-ph/0011363.

[87] J. Butterworth, J. Forshaw, and M. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637, arXiv:9601371 [hep-ph].

[88] M. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv:hep-ph/0206293.

[89] F. Caravaglios, M.L. Mangano, M. Moretti and R. Pittau, *A New Approach to Multi-jet Calculations in Hadron Collisions*, Nucl. Phys. B **539** (1999) 215, arXiv:hep-ph/9807570.

[90] J. Campbell, R. K. Ellis, and D. Rainwater, *Next-to-leading order QCD predictions for $W + 2$ jet and $Z + 2$ jet production at the CERN LHC*, Phys. Rev. D **68** (2003) 094021.

[91] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, *FEWZ 2.0: A code for hadronic Z production at next-to- next-to-leading order*, arXiv:1011.3540 [hep-ph].

[92] S. Frixione and B. Webber, *The MC@NLO 3.2 event generator*, arXiv:hep-ph/0601192.

[93] F. K. M. S. S. S. e. a. T. Gleisberg, S. Hoeche, *Event generation with SHERPA 1.1*, JHEP **007** (2009) 0902, arXiv:0811.4622 [hep-ex].

[94] P. N. Keith Hamilton, *Improving NLO-parton shower matched simulations with higher order matrix elements*, arXiv:1004.1764 [hep-ex].

[95] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040.

[96] J. Alwall et al., *MadGraph/MadEvent v4: the new web generation*, JHEP **09** (2007) 028, arXiv:0706.2334 [hep-ph].

[97] B. P. Kersevan and E. Richter-Was, *The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5*, arXiv:0405247 [hep-ph].

[98] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur.Phys.J. **70** (2010) 823, arXiv:1005.4568 [physics.ins-det].

[99] S. Agostinelli et al., *Geant4: a simulation toolkit*, Nucl. Instr. Meth. A **506** (2003) 250.

[100] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

[101] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007.

[102] R. Frhwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research **262** (1987) no. 23, 444 – 450. http://www.sciencedirect.com/science/article/pii/0168900287908874.

[103] ATLAS Collaboration Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at* $\sqrt{s} = 7$ *TeV in the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2010-069, CERN, Geneva, Jul, 2010.

[104] ATLAS Collaboration Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, Tech. Rep. ATLAS-CONF-2012-042, CERN, Geneva, Mar, 2012.

[105] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, ATL-LARG-PUB-2008-002 (2012) . https://cdsweb.cern.ch/record/1099735.

[106] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur.Phys.J. **C72** (2012) 1909, arXiv:1110.3174 [hep-ex].

[107] E. Abat, J. Abdallah, T. Addy, P. Adragna, M. Aharrouche, et al., *Combined performance studies for electrons at the 2004 ATLAS combined test-beam*, JINST **5** (2010) P11006.

[108] B. Acharya, J. Adelman, S. Adomeit, M. Aoki, B. Alvarez, F. Balli, W. Bell, K. Becker, K. Behr, D. Benjamin, E. Bergeaas Kuutmann, C. Bernard, K. Black, S. Calvet, R. Camacho, Y. Coadou, G. Cortiana, N. Cooper-Smith, T. Cornelissen, M. Cristinziani, V. Dao, U. De Sanctis, C. Doglioni, F. Derue, K. Finelli, K. Grahn, J. Groth-Jensen, S. Head, A. Henrichs, D. Hirschbuehl, V. Kaushik, O. Kind, H. Khandanyan, A. Krasznahorkay, T. Kuhl, E. Le Menedeu, H. Lee, A. Lister, K. Loureiro, L. Miljovic, J. Morris, R. Moles Valls, O. Nackenhorst, D. Pelikan, M. Owen, M. Pinamonti, K. Rao, K. Rosbach, M. Rudolph, G. Salamanna, J. Schwindling, J. Searcy, E. Shabalina, K. Shaw, J. Sjolin, R. Soualah, S. Stamm, D. Ta, T. Theveneaux-Pelzer, E. Thompson, K. Uchida, L. Valery, M. Vreeswijk, C. Wasicki, I. Watson, K. Yau, J. Zhong, H. Zhu, and M. zur Nedden, *Object selection and calibration, background estimations and MC samples for the Winter 2013 Top Quark analyses with 2012 data*, Tech. Rep. ATL-COM-PHYS-2013-088, CERN, Geneva, Jan, 2013.

[109] ATLAS Collaboration Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-063, CERN, Geneva, Apr, 2011.

[110] G. P. Salam, *Towards Jetography*, Eur.Phys.J. **C67** (2010) 637–686, arXiv:0906.1833 [hep-ph].

[111] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189v2 [hep-ph].

[112] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57, arXiv:0512210v2 [hep-ph].

[113] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[114] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv:1112.6426 [hep-ex].

[115] C. Cojocaru et al., *Hadronic calibration of the ATLAS liquid argon end-cap calorimeter in the pseudorapidity region $1.6 < |\eta| < 1.8$ in beam tests*, Nucl. Instr. Meth. A **531** (2004) 481, arXiv:0407009 [physics].

[116] T. Barillari et al., *Local hadronic calibration*, ATL-LARG-PUB-2009-001 (2009) . https://cds.cern.ch/record/1112035.

[117] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102 (2011) . https://cds.cern.ch/record/1369219.

[118] ATLAS Collaboration, *Measurement of the b-tagging efficiency in a sample of jets containing muons with 5 fb^{-1} of data from the ATLAS detector*, ATLAS-CONF-2012-043 (2012) . https://cdsweb.cern.ch/record/1435197.

[119] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, Eur. Phys. J. C **72** (2012) 1844, arXiv:1108.5602 [hep-ex].

[120] ATLAS collaboration, *Search for anomalous production of events with same-sign dileptons and b jets in 14.3 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, ATLAS-CONF-2013-051 (May, 2013) .

[121] P. Golonka and Z. Wąs, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, arXiv:0506026 [hep-ph].

[122] S. Jadach, J. H. Kühn, and Z. Waąs, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1991) 275.

[123] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:0204244 [hep-ph].

[124] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Single-top production in MC@NLO*, JHEP **03** (2006) 092, arXiv:0512250 [hep-ph].

[125] S. Frixione, E. Laenen, P. Motylinski, C. White, and B. R. Webber, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv:0805.3067 [hep-ph].

[126] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[127] M. Aliev et al., *HATHOR: a Hadronic Top and Heavy quarks cross section calculator*, Comput. Phys. Commun. **182** (2011) 1034, arXiv:1007.1327 [hep-ph].

[128] A. D. Martin et al., *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv:0905.3531 [hep-ph].

[129] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[130] M. L. Mangano et al., *Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study*, Nucl. Phys. **B 632** (2002) 343, arXiv:0108069 [hep-ph].

[131] ATLAS Collaboration, *Measurement of the t-channel single top-quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Lett. **B717** (2012) 330, arXiv:1205.3130 [hep-ex].

[132] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$*, Phys. Rev. D **74** (2006) 114017, arXiv:0609070 [hep-ph].

[133] T. Sjostrand et al., *High-energy-physics event generation with Pythia 6.1*, Comput. Phys. Commun. **135** (2001) 238, arXiv:0010017 [hep-ph].

[134] A. Sherstnev and R. Thorne, *Parton distributions for LO generators*, Eur. Phys. J. C **55** (2008) 553, arXiv:0711.2473 [hep-ph].

[135] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, arXiv:1103.2792 [hep-ph].

[136] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.

[137] J. M. Campbell and R. K. Ellis, *$t\bar{t}W$ production and decay at NLO*, arXiv:1204.5678 [hep-ph].

[138] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trsyi, *$t\bar{t}W$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **1211** (2012) 056, arXiv:1208.2665 [hep-ph].

[139] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[140] J. Campbell and R. Ellis, *An update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006, arXiv:9905386 [hep-ph].

[141] J. A. Aguilar-Saavedra, *PROTOS, a Program for Top Simulations*, . http://jaguilar.web.cern.ch/jaguilar/protos/.

[142] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension*, Comput. Phys. Commun. **108** (1998) 56, arXiv:9704448 [hep-ph].

[143] ATLAS Collaboration, *Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Eur. Phys. J. C **72** (2012) 2039, arXiv:1203.4211 [hep-ex].

[144] ATLAS Collaboration, *Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **71** (2011) 1577, arXiv:1012.1792 [hep-ex].

[145] J. Heinrich, *Bayesian limit software: multi-channel with correlated backgrounds and efficiencies*, .

[146] T. Junk, *Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties*, .

[147] T. Junk, *Building a more general χ^2*, .

[148] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, submitted to Eur. Phys. J. , arXiv:1112.6426v1 [hep-ph].

[149] ATLAS Collaboration, *In-situ jet pseudorapidity intercalibration of the ATLAS detector using dijet events in* $\sqrt{s} = 7$ TeV *proton-proton 2011 data*, ATLAS-CONF-2012-124 (2012) . http://cds.cern.ch/record/1474490.

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

[151] ATLAS Collaboration, *JES uncertainties for 2012 data*, . https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2013JESUncertainty.

[152] 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 (2012) .

[153] ATLAS Collaboration, *b-jet tagging calibration on c-jets containing* D^* *mesons*, ATLAS-CONF-2012-039 (2012) .

[154] ATLAS Collaboration, *Measurement of the Mistag Rate of b-tagging algorithms with 5* fb^{-1} *of Data Collected by the ATLAS Detector*, ATLAS-CONF-2012-040 (2012) .

[155] F. Ahmadow et al., *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair using up to 4.7* fb^{-1} *of pp collision data at* $\sqrt{s} = 7$ TeV *with the ATLAS detector at the LHC*, ATL-COM-PHYS-2011-1648 (2011) .

[156] M. Bosman et al., *Weighting method to propagate heavy-flavor tagging calibrations and related uncertainties*, ATL-COM-PHYS-2010-331 (2010) .

[157] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instr. Meth. A **434** (1999) 435, arXiv:9902006 [hep-ex].

[158] A. L. Read, *Presentation of search results: the* CL_s *technique*, J. Phys. G **28** (2002) 2693.

[159] A. L. Read, *Modified frequentist analysis of search results (the* CL_s *method)*, .

[160] ATLAS Collaboration, *Measurement of* $t\bar{t}$ *production with a veto on additional central jet activity in pp collisions at* $\sqrt{s} = 7$ TeV *using the ATLAS detector*, Eur. Phys. J. C **72** (2012) 2043, arXiv:1203.5015 [hep-ex].

[161] *Liza Mijovic's talk at the MC Generator Group:*, . https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materia%lId=slides&confId=172994

[162] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur.Phys.J. **C71** (2011) 1753, arXiv:1107.5909 [hep-ph].

[163] B. Acharyia et al., *Object selection and calibration, background estimations and MC samples for the Summer 2012 Top Quark analyses with 2011 data*, ATL-COM-PHYS-2012-499 (2012) .