Baseline analysis for the observability study of the $t\bar{t}H(H \to b\bar{b})$ process in the lepton plus jets channel

Georges Aad, Lorenzo Feligioni, Alexandre Rozanov, Laurent Vacavant*
Centre de Physique des Particules de Marseille

November 16, 2007

Abstract

This note is intended to be the first of a series of studies performed in the framework of the HG5 CSC note, aimed at the investigation of the observability of a Higgs boson in the $t\bar{t}H(H \to b\bar{b})$ semi-leptonic channel. After the description of the Monte Carlo samples needed for this analysis, we discuss the pre-selection of the one lepton and 4 b-jet final state and two analysis techniques to discriminate the signal from the combinatorial backgrounds. The aim of the note is to establish and document several corrections and techniques that could define a baseline analysis for the channel. This analysis leads to a statistical significance of $S/\sqrt{B} = 2.3$ for 30 fb^{-1} and for $m_{\mathrm{H}} = 120$ GeV/c^2.

1 Introduction

If the Standard Model Higgs boson exists with a mass in the vicinity of its experimental upper limit of 114.4 GeV [1], as indicated by the electroweak precision fits, its decay products consist of a $b\bar{b}$ pair about 70% of the time. Unfortunately this signature is very difficult to detect when the Higgs boson is produced directly in the 14 TeV center-of-mass pp interactions of the LHC. The main difficulties are represented by the presence of an overwhelming QCD $b\bar{b}$ background and to a lesser extent by the capability to trigger on these events. The same arguments can be applied for the vector boson fusion production mechanism which has the second highest cross-section, about one tenth of the direct production.

For these reasons only the associated production of an Higgs boson with a gauge boson (W and Z) or a $t\bar{t}$ pair is usually considered, with a preference for the latter since, despite the smaller cross-section, its physics background rate is lower. The $t\bar{t}H(H \to b\bar{b})$ channel is expected to be the only mode where the decay $H \to b\bar{b}$ is observable at the LHC. In addition, this is the only channel allowing to measure the top-Higgs Yukawa coupling for a low-mass Higgs boson.

*Partly supported by the European Commission under Marie Curie International Reintegration Grant MIRG-CT-2005-21700

1.1 Signal

At LHC the $t\bar{t}H$ signal is dominantly (90%) produced by gluon-gluon interactions, illustrated in Fig. 1. The remaining 10% arises from quark-antiquark interactions. For a Higgs boson mass of 115 GeV/c^2 and 130 GeV/c^2 the production cross-section times branching ratio to $b\bar{b}$ varies roughly between 0.4 and 0.2 pb at leading order. The top quarks decay almost exclusively to bW, and therefore the various final states can be classified according to the decays of the W bosons.

With a value of 43% the all-hadronic channel is the one with the highest branching fraction. Unfortunately the large QCD multijet cross-section does not allow easily to trigger on jets. Only tight requirements on the jet p_T and on the jet multiplicity could lead to reasonable rates in the first level of the trigger, but these requirements come at the expense of the signal efficiency. Another option currently being studied for this final state is to use the b-tagging at the second level of the trigger to reduce the p_T threshold on jets.

Despite presenting a simpler signature to trigger on, given the presence of two isolated leptons, the all-leptonic final state (electrons or muons) analysis is probably the less feasible. The branching fraction (5%) is low and the two neutrinos prevent the reconstruction of the top quarks.

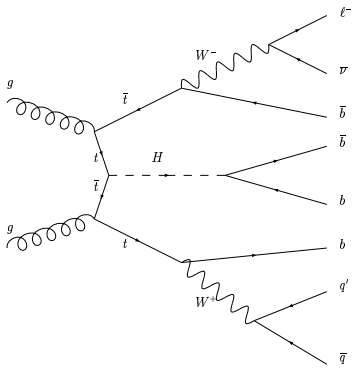

Figure 1: *A Feynman diagram for $t\bar{t}H$ in the semi-leptonic final state.*

The semi-leptonic final state, the choice for this analysis, is a good compromise with a branching fraction of about 28% (excluding taus). The trigger relies on the presence of a high-p_T lepton, as shown in Fig. 1, which can be identified with reasonable efficiency. The same lepton can be used for the reconstruction of one of the top quarks. The non-b-tagged jets can be used for reconstructing the hadronic W boson, decreasing the combinatorial background with respect to the all-hadronic final state.

1.2 Physics backgrounds

The $t\bar{t}H$ semi-leptonic final state consists of one energetic lepton and a high jet multiplicity. This signature alone is not sufficient to discriminate against the large $W(\rightarrow \ell\nu)$ + jets and QCD multijet production where the presence of an isolated lepton could come from the decay of an heavy flavored meson or, in case of the electron, even from the contamination of hadronic jets faking the electron signature. Even though requesting 4 b-jets in the event, as explained in Section 4.3.2, could still lead to a contamination of QCD $b\bar{b}b\bar{b}$ production (cross-section is a few hundreds of nb), the reconstruction of the $t\bar{t}$ system allows a certain degree of safety against non-top background.

The production of $t\bar{t}$ events, with a cross-section of about 800 pb, is the main background for the $t\bar{t}H$ process. Given the high jet multiplicity in the signal process ($\geq$ 6 jets), only $t\bar{t}$ events produced together with at least two extra jets contribute to the pre-selected data sample. Since most of these extra jets comes from the hadronization of light quarks, this contribution is greatly reduced by asking four jets to be identified as b-jets.

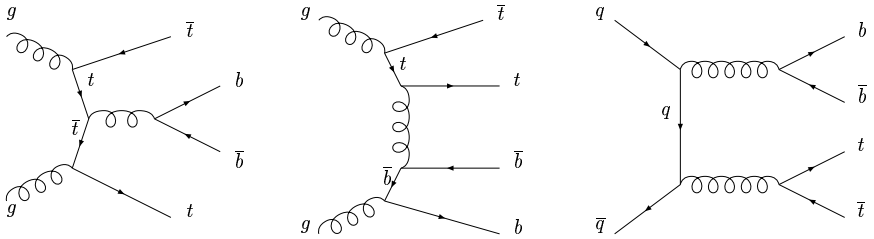

Figure 2: *Some Feynman diagrams for the $t\bar{t}b\bar{b}$ QCD production.*

The irreductible background comes from $t\bar{t}b\bar{b}$ production. The production can proceed via QCD or via electroweak interactions with a total cross-section of the order of 9 pb. Some of the Feynman diagrams involved in the two production mechanisms are shown in Fig. 2 and Fig. 3. While the QCD production is by far the largest, with a cross-section 10 times higher, the EW production is also important: the two b-jets not coming from the $t\bar{t}$ decay have hard momenta and the total invariant mass of this $b\bar{b}$ system is close to the Z boson mass and can therefore contaminate the signal region.

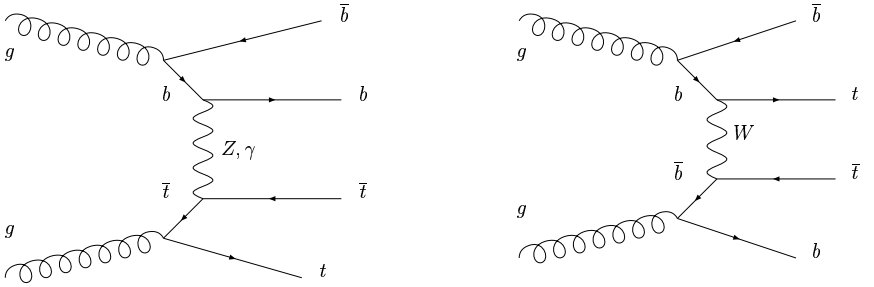

Figure 3: *Some Feynman diagrams for the $t\bar{t}b\bar{b}$ EW production.*

With a cross-section roughly 40% higher than $t\bar{t}b\bar{b}$ the $t\bar{t}c\bar{c}$ background could also play an important role. To estimate the need for a dedicated $t\bar{t}c\bar{c}$ sample, two AlpGen [2] samples of $t\bar{t}c\bar{c}$ and $t\bar{t}b\bar{b}$ were generated and compared using ATLAS fast simulation Atlfast [3]. The Atlfast default c-jet rejection was decreased to follow the latest estimates from full simulation. Even a modest c-jet rejection of 8, for a b-tagging efficiency of 60%, leads to a $t\bar{t}b\bar{b}$ over $t\bar{t}c\bar{c}$ ratio of 20 after performing a simple cut-based analysis, as explained in Section 7. Therefore no dedicated sample was generated for the full-simulation study. Some $t\bar{t}c\bar{c}$ events are present in the inclusive $t\bar{t}$+jets sample used.

Several other backgrounds, such as W+jets and single-top+W, could also have a non-negligible impact on the analysis. Even though the W+jets cross-sections is huge (about 15000 pb per lepton flavor), it has been shown [5] that the contribution to the analysis can be reduced to a negligible level if the four b-tags requirement is applied. This is true as well for the less abundant single-top background (10 pb). None of those samples were studied in what follows.

1.3 Previous studies

The first ATLAS sensitivity study for the $t\bar{t}H$ channel has been presented in the ATLAS physics TDR [4]. In that study, based on fast simulation, a cut-based analysis aimed to reconstruct the entire final state was performed. For a Higgs boson mass of 120 GeV/c^2 the analysis result for 30 fb^{-1} of data was a significance of 3.6.

This study was repeated by Cammin and Schumacher [5] using mostly fast simulated events. In this later analysis a direct comparison with the old TDR study gave an updated significance of 2.0. The reasons for such a change were studied in detail and found to be related to the differences in the Monte Carlo generators and to the updated cross-sections used. Pythia was used in order to simulate the $t\bar{t}b\bar{b}$ background in the TDR while this study used a matrix element generator, AcerMC, which naturally led to a better selection efficiency for those events. Updating PDF's which were used and changing the default Q^2 scale resulted in a decrease of 20% for all cross sections. As a step forward, this work also proposed an improved analysis based on a likelihood technique using kinematical distributions to reconstruct the two top quarks. With this procedure it was possible to recover some of the significance lost with respect to the TDR analysis, resulting in an estimate for $S/\sqrt{B}$ of 2.7.

The latest ATLAS thesis work was carried out by Corréard [6]. It made use of the likelihood techniques developped in [5] performing the analysis on "mixed simulated" events (fast simulated calorimeter and muon system and fully simulated inner detector) in order to exploit the implicit correlations between b-tagging algorithms and kinematical observables. Promising results were found, with a potential increase of the significance up to a value around 4.0. Unfortunately the lack of statistics for the simulated $t\bar{t}$+jets process did not allow to make any strong statement.

Finally, it is worth mentionning that CMS reported in its Physics TDR [7] a significance of 1.8 for the electron channel and 1.6 for the muon channel, in both cases for 60 fb^{-1}.

1.4 Scope of this study

This work is intended to be a back-up document for the CSC note devoted to the observability study of the Higgs boson in the $t\bar{t}H(H \to b\bar{b})$ semi-leptonic channel (note HG5). In Section 2 we describe in details how we defined the Monte Carlo samples used in this analysis, also used in the CSC note. Notice that all samples are fully simulated, an important step forward with respect to the previous $t\bar{t}H(H \to b\bar{b})$ analyses. Compared to the previous studies, the main difference regarding generators is the use of MC@NLO [8] for the simulation of the $t\bar{t}$ background. MC@NLO provides a better description of the leading extra jet for the $t\bar{t}$ process. This was previously simulated using the showering in Pythia which could result in a softer jet p_T spectrum and gives a lower jet multiplicity. More importantly, MC@NLO was found [9] to provide a reasonable description of the sub-leading additional jet. Thanks to the implementation a dedicated jet filter, the study presented in this note makes use of the largest fully simulated $t\bar{t}$+jets sample to date. In Section 2.4 we present a discussion of the treatment of the double-counting in $t\bar{t}b\bar{b}$ events.

After giving an overview of the analysis in Section 3, the pre-selection of one lepton and 4 b-jets is explained in Section 4. Identification cuts for electrons and muons, as derived from the AOD objects in release 12, are presented together with the corresponding efficiencies and fake rates estimation. Some specific corrections are shown in details: the electron/jet overlap removal, the soft muon jet correction and the jet calibration.

The discussion of the reconstruction of the two W bosons, Sections 5 and 6, is followed by the presentation of how a cut-based analysis (à la TDR), Section 7, and a likelihood analysis (à la Cammin), Section 8, are performed to discriminate the signal from the combinatorial backgrounds.

The aim of this note is to establish and document the corrections and techniques, mentionned above, which could define a baseline analysis for the channel. Several points will be discussed in foreseen future notes. This is notably the case for the important issues of the background measurement and the signal extraction in data, as well as for the trigger aspects. Even though our cuts on the reconstructed leptons from the W boson imply a single lepton trigger, neither a detailed study of this trigger nor a discussion on potential improvements on the trigger strategy are presented here.

2 Monte Carlo samples and cross-sections

Unless stated otherwise, all the Monte Carlo samples were generated, fully-simulated and reconstructed with the release 12.0.6. All samples were simulated with the so-called 30 μm G4 cut. No pile-up is included.

The signal and the $t\bar{t}b\bar{b}$ events have been generated with a lepton filter requiring at least one electron or one muon with a pseudorapidity $|\eta| < 2.7$ and a transverse momentum above 10 GeV/c.

An accidental event duplication during production of the signal and $t\bar{t}b\bar{b}$ (QCD) files was treated properly and duplicated events removed. In addition, some events with the same kinematics for the top quarks and the Higgs boson were found in the two different signal samples and were removed. Both problems are now understood and fixed for future production.

2.1 Signal

Two signal samples were generated for a Higgs boson mass of $m_{\mathrm{H}} = 120$ GeV/c^2 with Pythia [10] 6.403 (within Athena release 12.0.6.1). Samples 5870 and 5871 correspond respectively to the positively and negatively charged lepton from the leptonic W. The exact generated process is $pp \rightarrow t\bar{t}HX \rightarrow \ell\nu bq\bar{q}'b\bar{b}X$, with ℓ = e or μ (ISUB=121,122). The scale choices for the generation correspond to the Pythia defaults. In particular, the factorisation and maximal ISR scales are identical, their value being $Q^2 = m_{\mathrm{t}}^2 + \max(p_{T\mathrm{t}}^2, p_{T\bar{\mathrm{t}}}^2)$ with $m_{\mathrm{t}} = 175$ GeV/c^2. The new showering of Pythia is used.

The cross-section of this process is known at next-to-leading order with an uncertainty of roughly 20%. Values from different computations are summarized in Table 1. A K-factor of about 1.2 is obtained for $Q^2 = (m_{\mathrm{H}}/2 + m_{\mathrm{t}})^2$, but we used the LO cross-section in this study.

Source	Pythia	HQQ [11]	LO [12]	NLO [12]	NLO [12]
Structure functions	CTEQ6L1	CTEQ6L	MRST	MRST	CTEQ6M
$\sigma(t\bar{t}H)$ (fb)	520	536	577	702	665

Table 1: *Different estimations of the $t\bar{t}H$ cross-section for $m_H = 120$ GeV/c^2. The factorisation and renormalisation scales are set to $Q^2 = (m_H/2 + m_t)^2$ except in Pythia (see text).*

The cross-section of the generated process, the number of events and the equivalent luminosity are summarized in Table 2. The cross-section includes the H $\rightarrow$ b$\overline{\text{b}}$ (68.4% at 120 GeV/c^2 [13]) and W$\rightarrow \ell\nu$ (10.66% [1]) branching ratios, as well as the lepton filter efficieny of $\epsilon = 0.953$.

2.2 t$\overline{\text{t}}$b$\overline{\text{b}}$ QCD and EW processes

Both processes can be initiated by a gluon-gluon or a q$\overline{\text{q}}$ pair, but only the dominant (95%) gluon fusion has been simulated. They have been generated with the AcerMC [14] program interfaced to Pythia. The exact generated process is gg$\rightarrow$ t$\overline{\text{t}}$b$\overline{\text{b}}$X$\rightarrow \ell\nu$bq$\,\overline{\text{q}}'$bb$\overline{\text{b}}$X, with ℓ = e or μ and the Q-scale used in AcerMC is $Q = m_{\text{H}}/2 + m_{\text{t}} = 235$ GeV with $m_{\text{H}} = 120$ GeV/c^2. It is important to note that, depending on the choice of this scale, the total cross-section can be as much as twice smaller or twice bigger than with our current choice [14]. In both cases the CTEQ6L1 structure functions have been used and the release 12.0.31 was used for simulation.

The QCD process (sample 6886) corresponds to all the diagrams with electroweak gauge bosons switched off. The versions of the generators used are AcerMC 3.4 and Pythia 6.403 (within Athena release 12.0.7.1). The leading order cross-section is $\sigma(\text{pp} \rightarrow \text{t}\overline{\text{t}}\text{b}\overline{\text{b}}) = 8.18(\text{gg})(+0.52(\text{q}\overline{\text{q}}))$ pb and the lepton filter efficiency is $\epsilon = 0.946$. The Q-scale which in Pythia fixes the maximum virtuality in the initial state parton shower is $Q = 235$ GeV. For this process, another sample (5855) was previously generated with AcerMC 2.4 and Pythia 6.323 (within Athena release 11.0.41.3). For this latter sample, the Q-scale for ISR in Pythia was set to $Q = 235^2$ GeV. A priori the differences between the two samples are useful to assess the systematics associated with the choice of scale. However, while some small differences were seen in the selection, the impact at the end of the analysis and for instance on the significance is not visible given the statistics, and this is not further discussed in this note.

The electroweak process (sample 5214) corresponds to diagrams with at least one electroweak gauge boson. The generator versions are AcerMC 3.3 and Pythia 6.403 (within Athena release 12.0.31.4). The leading order cross-section is $\sigma(\text{pp} \rightarrow \text{t}\overline{\text{t}}\text{b}\overline{\text{b}}) = 0.90(\text{gg})(+0.04(\text{q}\overline{\text{q}}))$ pb and the lepton filter efficiency is $\epsilon = 0.943$. The Q-scale used in Pythia is $Q = 235$ GeV.

2.3 t$\overline{\text{t}}$ process

The reducible t$\overline{\text{t}}$ background events have been generated with the MC@NLO program, interfaced to Herwig and Jimmy for the underyling event. The common background sample (sample 5200) has been used. The events in this sample correspond to the processes pp $\rightarrow$t$\overline{\text{t}} \rightarrow (\ell\nu, \text{q}\,\overline{\text{q}}')\text{b}\ell\nu\text{b}$ with $\ell = \text{e}, \mu, \tau$.

The generator versions used are MC@NLO 3.1 and Herwig 6.510. The Q-scale corresponds to $Q^2 = m_{\text{t}}^2 + \frac{1}{2}(p_{T\text{t}}^2, p_{T\overline{\text{t}}}^2)$. The inclusive t$\overline{\text{t}}$ cross-section from MC@NLO is $\sigma(\text{pp} \rightarrow \text{t}\overline{\text{t}}) = 760$ pb. The fraction of negative weight events in this sample is $f \sim 0.14$. The CTEQ6M structure functions have been used.

To increase the Monte Carlo statistics, we completed the common sample with a similar sample biased toward a high jet multiplicity thanks to a filter on true jets (sample 5212). Jets of cone size 0.4 are built from the truth with the JetRec algorithm. The overlap between electrons and jets is treated as follows: isolated electrons from W's or from taus are removed from the jet list if $\Delta R(j, e) < 0.4$ and $|p_{Tj} - p_{Te}|/p_{Tj} < 0.1$. The complete filter then corresponds to the following requirements :

[1] In Pythia, Br(W$\rightarrow \ell\nu$) = 10.82%.

- only the process pp $\rightarrow t\bar{t}X \rightarrow \ell\nu bq\bar{q}'b$ with $\ell = e, \mu$ has been generated
- at least six jets with $p_T > 14$ GeV/c and $|\eta| < 5.2$
- at least four jets with $p_T > 14$ GeV/c and $|\eta| < 2.7$
- at least one electron or one muon with $p_T > 14$ GeV/c and $|\eta| < 2.7$

The efficiency of this filter on inclusive $t\bar{t}$ events is 0.146 and the fraction of negative weight events in this sample is $f \sim 0.135$.

2.4 Double-counting treatment for $t\bar{t}b\bar{b}$ events

There exists an overlap between the $t\bar{t}X$ samples (5200 and 5212) and the $t\bar{t}b\bar{b}$ QCD sample (6886), since parton shower can generate $b\bar{b}$ pairs in addition to a given hard process. In the analysis presented here, the events with additional $b\bar{b}$ pairs in the MC@NLO samples have been rejected, since the Matrix Element $t\bar{t}b\bar{b}$ generation is expected to better describe such events in the region of the phase space selected by this analysis (especially for relatively large opening angle between the two quarks of the $b\bar{b}$ pair). The corresponding number of events was also removed from 5212 for normalization purposes: in principle this might be a bit optimistic but the effect is small since only 10% of events are removed.

2.5 Summary

Table 2 summarizes the cross-sections, calculated using the Monte Carlo generators, of the different processes considered for this analysis together with the corresponding numbers of generated events and the equivalent integrated luminosity. All branching fractions and filter efficiencies are included. To account for the $q\bar{q}$ production which is not simulated in the $t\bar{t}b\bar{b}$ samples, the cross-sections were increased by 6% and 4% for respectively the QCD and EW productions. For the sample 5212, about 10% of events were removed in the number of events and from the cross-section number, following the $t\bar{t}b\bar{b}$ overlap treatment explained in the previous section. In addition for this sample, the fraction of events with a negative weight (106173 out of 766591 events) was taken into account in the equivalent integrated luminosity.

Process	Sample	Cross-section (fb)	Events	L (fb^{-1})
$t\bar{t}H$ (LO)	5870+5871	98	42748	436
$t\bar{t}b\bar{b}$ QCD (LO)	6886	2371	14250	6.0
$t\bar{t}b\bar{b}$ EW (LO)	5214	255	10000	39
$t\bar{t}$ filtered (NLO)	5212	100515	766591	5.5

Table 2: *Summary of the different samples used for the analysis. The cross-sections are taken from the generators and include all branching fractions and filter efficiencies. The last column shows the equivalent integrated luminosity, taking into account all corrections (see text).*

3 Analysis overview

The analysis consists of two main parts. A first set of pre-selection requirements is applied to the events to ensure that the fundamental physics objects associated with the production of $t\bar{t}H$ are reconstructed. The second step involves the reconstruction of the top quark pairs and the Higgs boson through the identification of their decay products.

Several techniques are used to associate jets and leptons and to reconstruct the top quark pairs. This part of the analysis is directly related to the quality of the reconstructed Higgs signal. It mainly suffers from the presence of extra light jets from ISR/FSR faking the hadronic W boson decay products and from the misassociation of the four b-tagged jets to the original partons. For this reason multivariate algorithms are used to improve the performances in terms of rejection of this combinatorial background.

4 Pre-selection

At the pre-selection level we require one isolated high-p_T lepton (muon or electron) coming from the decay of one of the W bosons. Vetoing the presence of a second isolated lepton is intended to remove additional sources of background. We require at least six calorimeter jets, and at least four b-tagged jets from the decay of the top quarks and the Higgs boson.

4.1 Trigger lepton

The presence of one high-p_T lepton, together with missing tranverse momentum, is a distinct signature of W boson production. These leptons can generally be used to trigger on W production with high efficiency. In this Section we explain the selection criteria used for reconstructed muons and electrons produced in the semi-leptonic decay of the $t\bar{t}$ system and the efficiencies for such selections.

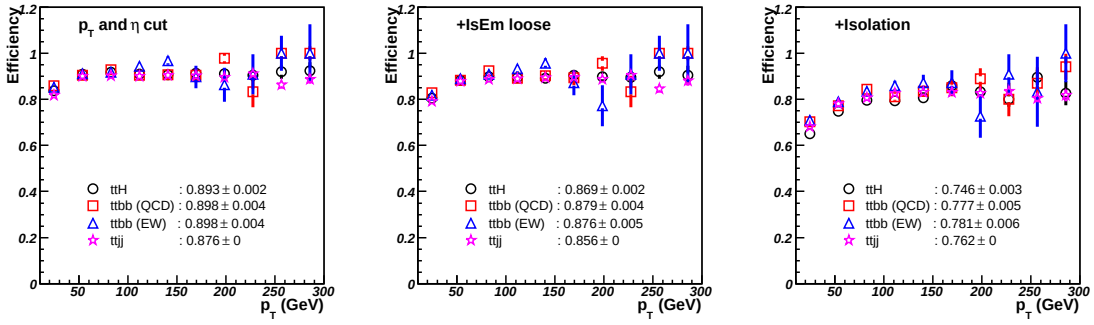

Figure 4: *Efficiency for electron reconstruction and identification as a function of generated electron p_T, for all the samples considered. From left to right, selection cuts are cumulated.*

4.1.1 Electron selection

Electron candidates are taken from the AOD ElectronCollection. Only candidates reconstructed by the high-p_T algorithm (author=1) or by both the high-p_T and the soft-electron algorithms

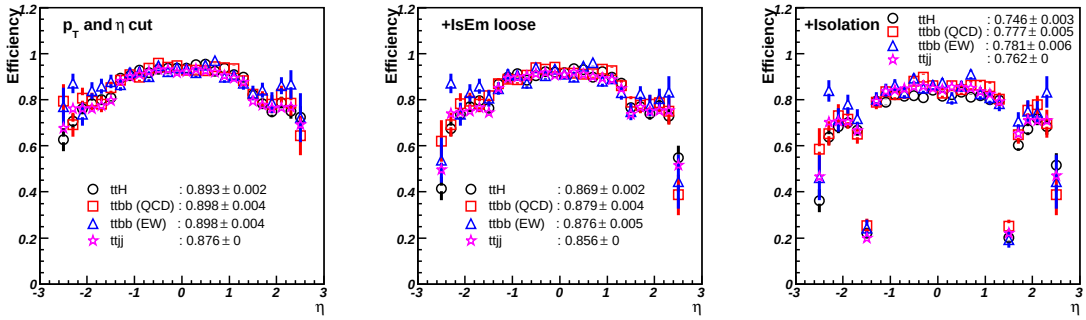

Figure 5: *Efficiency for electron reconstruction and identification as a function of generated electron η, for all the samples considered. From left to right, selection cuts are cumulated.*

(author=3) are considered, provided they fulfil the acceptance cuts: $p_T > 25$ GeV/c and $|\eta| < 2.5$. The electron-id cuts are usually combined in the general isEM flag. Since for this release the matching criteria with tracks were too tight, only the cuts based on calorimetric information are used: candidates are required to satisfy the requirement (isEM & 0x7)=0.A loose matching cut is applied anyway beforehand at the reconstruction level: in each sampling of the EM calorimeter the distance between the cluster and the extrapolated track is required to fulfil $|\Delta\varphi| < 0.05$ and $|\Delta\eta| < 0.025$. Finally, an isolation cut is applied in the form of an upper limit of 0.15 on the ratio between the transverse energy deposited inside a cone of size 0.3 in ΔR around the electron track (the *etcone* isolation variable) and the electron p_T. The efficiency, as a function of electron p_T and η for all different physics samples, are shown in Fig. 4 and 5. Since there is no requirements applied to the clusters in the crack region, it is visible only after the isolation cut. It should be noted that in a wide region around the crack ($1.35 < |\eta| < 1.65$), the crack scintillator cells are not removed at all from the *etcone* computation because of a problem in release 12.0.6, and therefore the inefficiency after the isolation cut in this region is overestimated. Different techniques, like a track-based isolation, will be investigated in the future to correct this problem.

The fake high-p_T electrons after final cuts, shown on Fig. 6, are actual electrons coming from B- and D-mesons in about 30% of the times. For the remaining fakes, about a fourth of them were traced back as originating from photon conversions where the photon was emitted by a W.

4.1.2 Muon selection

Muon candidates are taken from the AOD StacoMuonCollection. Candidates must be reconstructed with the high-p_T algorithm (author=1) combining the Inner Detector and the Muon Spectrometer and pass the acceptance cut: $p_T > 20$ GeV/c and $|\eta| < 2.5$. In order to remove poorly reconstructed muons an upper cut of 30 on the reduced χ^2 of the muon combined fit is applied. Further cuts on the muon track transverse impact parameter ($|d_0| < 50$ μm) and on the isolation are used to discriminate against muons generated by the decay of long-lived mesons. The isolation consists of a cut of 15 GeV on the transverse energy reconstructed in the calorimeter within a cone of size 0.2 in ΔR around the muon track. The efficiency and fake rate, as a function of muon p_T and η for all different physics samples are shown in Figures 7, 8 and 9.

The fake high-p_T muons, after final cuts, are still in 90% of the cases actual muons coming from B- and D-meson decays.

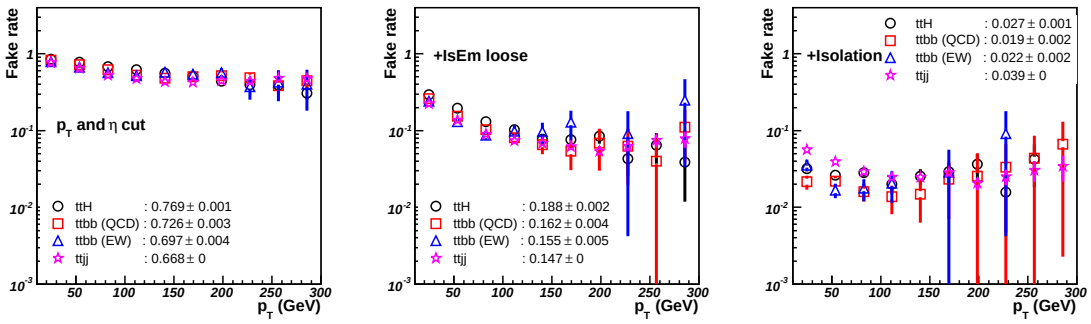

Figure 6: *Fake rate for reconstructed high-p_T electrons as a function of p_T, for all the samples considered. From left to right, selection cuts are cumulated. Soft electrons from B and D decays are counted as fakes.*

4.2 Treatment of soft muons

About 20% of the times a B-meson decay cascade gives rise to a muon. With a four b-jet signature in this channel, these muons, also called *soft muons* since in general have relatively low momentum, are present in almost every event. In order to have a better estimate of the momentum of the original b-parton, these muons must be used to correct the jet measured in the calorimeters. These muons are reconstructed by two different algorithms. All the candidates from the high-p_T algorithm (author=1) which are different from the trigger lepton and from the low-p_T algorithm (author=2) are considered for addition, provided they fulfil $p_T > 4$ GeV/c (and $p_T < 100$ GeV/c), $|\eta| < 2.5$ and $\chi^2/n < 30$ for the combined fit. In addition a loose anti-isolation cut is applied, requiring that the energy reconstructed in the calorimeter within a cone of size 0.3 in ΔR around the muon track is higher than 5 GeV.

The effect of adding a muon to at least one b-jet for reconstructing the top and Higgs boson masses is illustrated on Fig. 10: both the mass peak location and the resolution are improved.

4.3 Jets

To reconstruct the energy of the partons produced in the original collision, calorimeter jets are reconstructed using a seeded fixed-cone algorithm with a cone size of ΔR = 0.4. These jets are found in the AOD collection Cone4TowerParticleJets. Cuts on $p_T > 20$ GeV/c and $|\eta| < 5.0$ are applied. Only events with at least 6 jets are kept for the analysis. The jet multiplicity and the p_T of the leading jets are shown in Figures 11 and 12.

4.3.1 Overlaps with electrons

Since all electrons are reconstructed by the jet algorithm it becomes necessary to identify them in the jet collection in order to avoid double counting. The criteria for the jet-electron overlap removal is the following: each jet matching a well-reconstructed electron (i.e. fulfilling the cuts defined in Section 4.1.1) within a ΔR of 0.2, and for which the ratio of the electron to the jet transverse momenta is greater than 0.75, is discarded from the jet collection. About 4% of the jets are removed by this selection, 99% of them being actual true electrons.

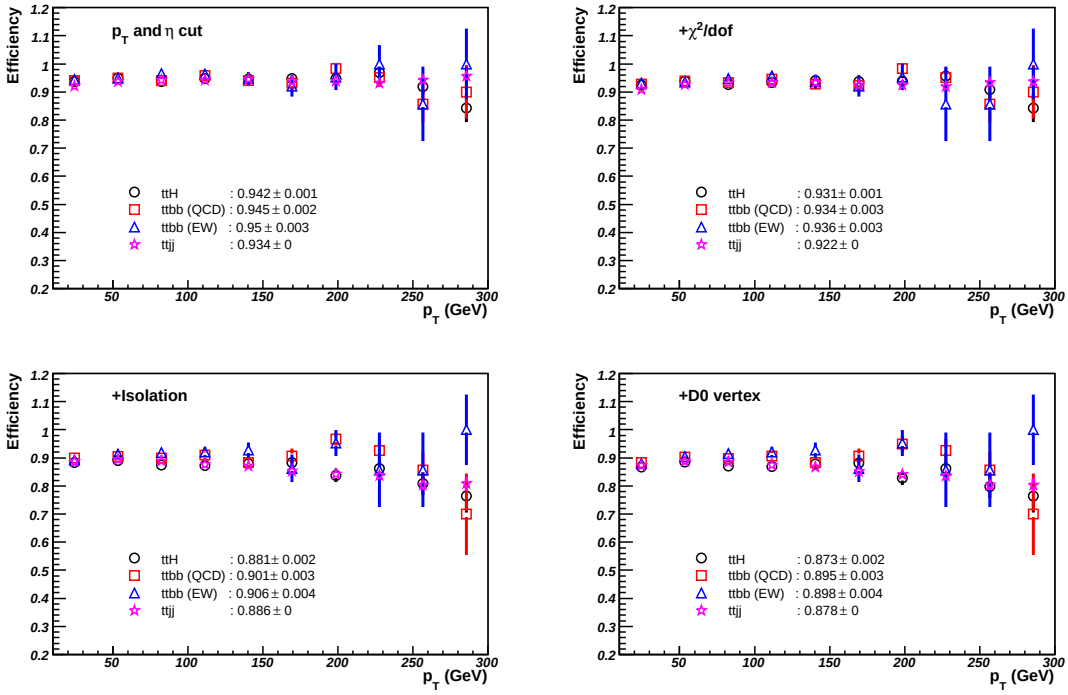

Figure 7: *Efficiency for muon reconstruction and identification as a function of generated muon p_T, for all the samples considered. From top left to right bottom plots, selection cuts are cumulated.*

4.3.2 b-tagging

B-jets are identified using the IP3D+SV1 tagger [15], which exploits both the impact parameter of tracks and the properties of an inclusive secondary vertex, using a likelihood approach which leads to a single discriminating variable: the b-tagging weight. The b-tag weight spectrum for b-, c-, and light jets in the signal sample is shown in Fig. 13. A cut on the weight defines which jets will be eventually identified as b-jets in the analysis. The rejection of c- and light jets versus the b-jet efficiency, obtained by varying the weight cut, is also shown in Figure 14. The different samples exhibit a very similar behaviour. In the same plot (bottom plots) the rejection for purified jets is shown, *i.e.* only jets away ($\Delta R > 0.8$) from any other heavy flavor (b, c, τ) are considered. For this analysis a b-tagged jet is defined as jet having a b-tag weight ≥ 6, which implies a b-tagging efficiency of about 62% and a rejection of light (c-) jets of about 100 (8).

At the event selection level we require at least four b-tagged jets, otherwise the event is discarded. Therefore it is important to work at the highest possible b-tagging efficiency while keeping a large enough rejection of light jets. This requirement of 4 b-jets is the last cut of the preselection. If more than four b-tagged jets are present in the event, only the jets with the highest b-tagging weight are considered as b-jets.

4.3.3 Calibration

A Monte Carlo based jet correction has been derived to take into account residual calibrations, *i.e.* out-of-cone effects. The jet 4-momentum is corrected by a rescaling factor: $\mathbf{p} \leftarrow \alpha(p_T) \times \mathbf{p}$.

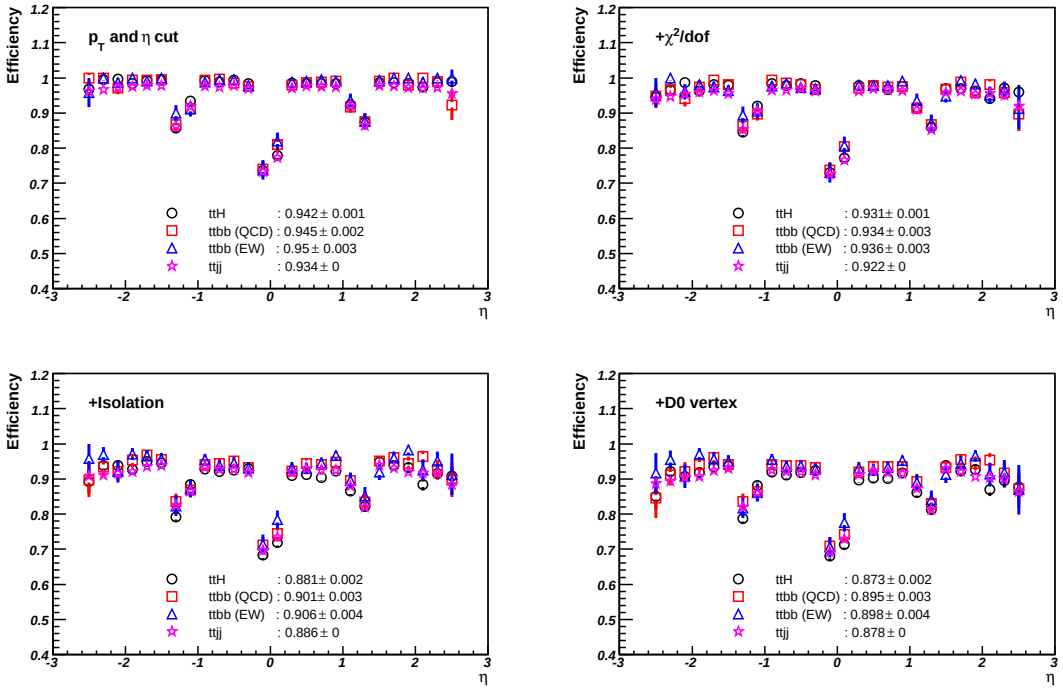

Figure 8: *Efficiency for muon reconstruction and identification as a function of generated muon η, for all the samples considered. From top left to right bottom plots, selection cuts are cumulated.*

An attempt to use a 2D-calibration, as a function of the jet p_T and η, did not bring any significant improvement. For this analysis we use a parametrization which was derived for Atlfast from full simulation and is proved to still calibrate properly the jets. The rescaling factor $\alpha(p_T)$ is shown in Tables 3 and 4 for b- and light jets. The impact of this calibration is shown on Fig. 15.

jet p_T	$\alpha_b(p_T)$ with p_T in GeV
$p_T <$ 55 GeV	$1.27+0.12\ p_T - 0.01\ {p_T}^2 + 0.33\cdot10^{-3}{p_T}^3$ $-0.47\cdot10^{-5}{p_T}^4 + 0.25\cdot10^{-5}{p_T}^5$
$55 \leq p_T <$ 200 GeV	$1.18-0.17\cdot10^{-2}p_T+0.44\cdot10^{-5}{p_T}^2$
$p_T \geq$ 200 GeV	$1.18-0.17\cdot10^{-2}\cdot200+0.44\cdot10^{-5}\cdot200^2$

Table 3: *Momentum-rescaling factor $\alpha(p_T)$ for b-jets.*

4.4 Results of pre-selection on signal and background

The effect of the event pre-selection on signal and background samples is illustrated in Table 5. The efficiency for the four b-tag cut is different in the signal and in the irreducible background because the two additional b-jets have different p_T spectra.

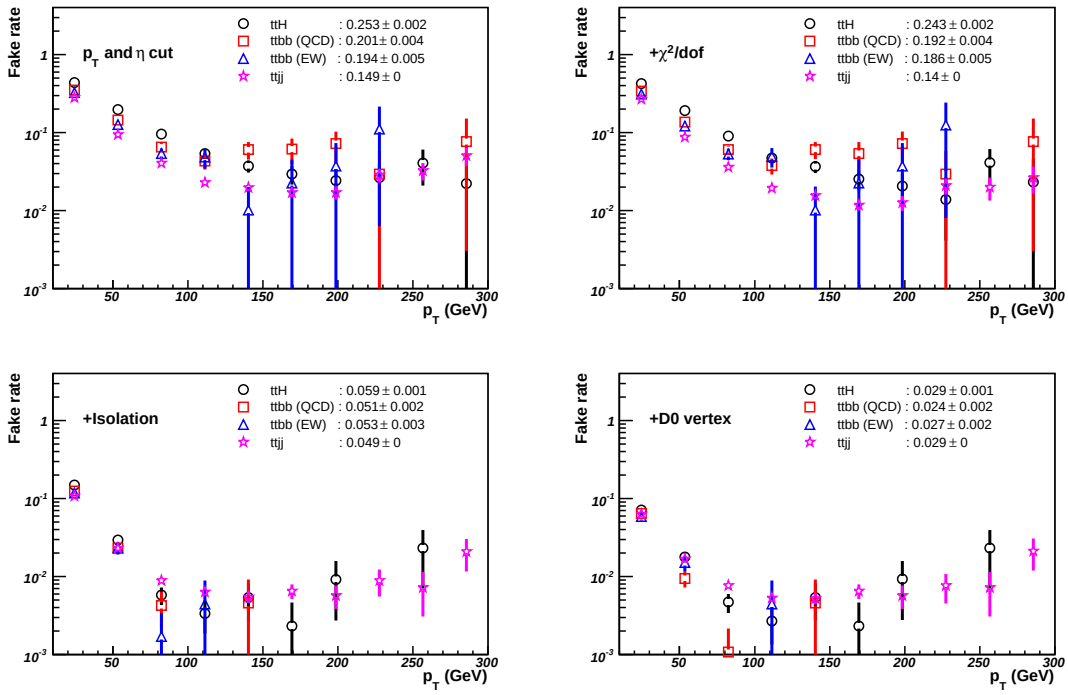

Figure 9: *Fake rate for reconstructed high-p_T muons as a function of p_T, for all the samples considered. From top left to right bottom plots, selection cuts are cumulated. Soft muons from B and D decays are counted as fakes.*

jet p_T	$\alpha_l(p_T)$ with p_T in GeV
$p_T <$ 45 GeV	$1.51+0.31\cdot10^{-1}p_T-0.37\cdot10^{-2}{p_T}^2+0.11\cdot10^{-3}{p_T}^3$ $-0.14\cdot10^{-5}{p_T}^4+0.61\cdot10^{-8}{p_T}^5$
$45 \leq p_T <$ 200 GeV	$1.18-0.17\cdot10^{-2}p_T+0.44\cdot10^{-5}{p_T}^2$
$p_T \geq$ 200 GeV	$1.18-0.17\cdot10^{-2}\cdot200+0.44\cdot10^{-5}\cdot200^2$

Table 4: *Momentum-rescaling factor $\alpha(p_T)$ for light jets.*

5 Reconstruction of the leptonic W boson

Reconstructing the leptonic W boson 4-momentum is necessary to reconstruct the top quarks. While the lepton 4-momentum is completely measured in the detector, in the case of the neutrino, only the transverse momentum can be inferred by measuring the imbalance of the measured transverse energy in the event. This measured quantity is referred to as missing transverse energy.

5.1 Neutrino p_z estimation

Once the transverse missing energy is identified with p_{T_ν}, the invariant mass of the sum of the lepton and neutrino 4-momentums can be constrined to the W boson mass:

$$M_W^2 = 2(E_\ell E_\nu - \mathbf{P}_\ell \cdot \mathbf{P}_\nu). \quad (1)$$

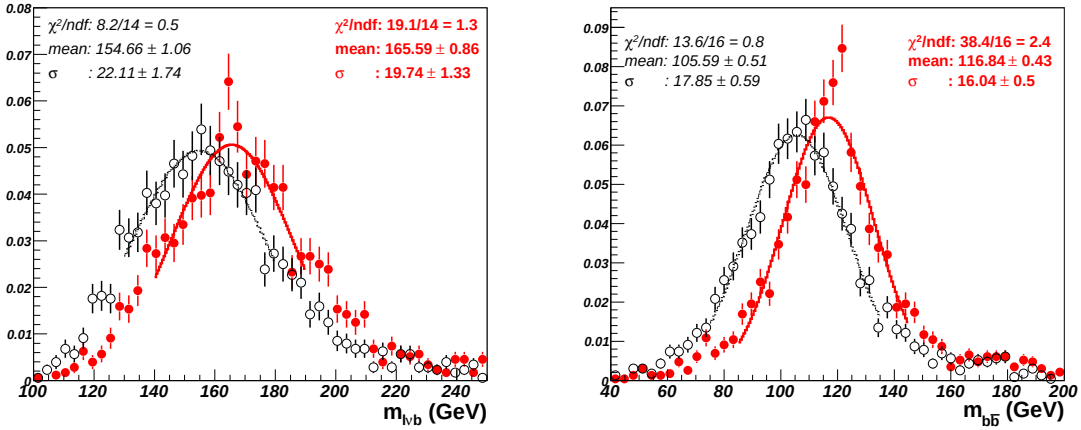

Figure 10: *Effect of adding a reconstructed soft muon to at least one of the b-jets, shown for the leptonic top mass and the Higgs boson mass. Red solid (black open) markers and solid (dotted) line show the mass distribution and fitted values for jets after (before) soft muon correction. The effect of the accidental wrong muon matches can be seen on the shape distorsion. All distributions are normalized to unity.*

preselection cut	$t\bar{t}H$	$t\bar{t}b\bar{b}$(EW)	$t\bar{t}b\bar{b}$(QCD)	$t\bar{t}X$
lepton	1938 ± 12	5144 ± 63	48443 ± 492	2098150 ± 3971
$+ \geq 6$ jets	1221 ± 9	2774 ± 46	23363 ± 341	853263 ± 2582
$+ \geq 4$ b-tags	108 ± 3	113 ± 9	888 ± 67	796 ± 80

Table 5: *Number of events after each preselection cut for signal and background. All number are normalized for 30 fb^{-1} of integrated luminosity. In the last column the contribution of $t\bar{t}b\bar{b}$ has been removed. The errors are statistical only.*

This leads to a quadratic equation with one unknown, p_{z_ν}. If we define

$$\beta = M_W^2 + 2p_{x_\ell}p_{x_\nu} + 2p_{y_\ell}p_{y_\nu} \tag{2}$$

and

$$\Delta = E_\ell^2 \left(\beta^2 + (2p_{z_\ell}p_{T_\nu})^2 - (2E_\ell p_{T_\nu})^2\right) \tag{3}$$

then the two solutions for p_{z_ν} can be written as:

$$p_{z_\nu}^{\pm} = \frac{1}{2}\frac{p_{z_\ell}\beta \pm \sqrt{\Delta}}{E_\ell^2 - p_{z_\ell}^2} \tag{4}$$

The two solutions are chosen later on in the analysis, either by the χ^2 or by the likelihood.

5.2 $\Delta = 0$ approximation

Because of the limited measurement resolution on the transverse missing energy, for a significant fraction of events Equation 4 does not have a real solution.

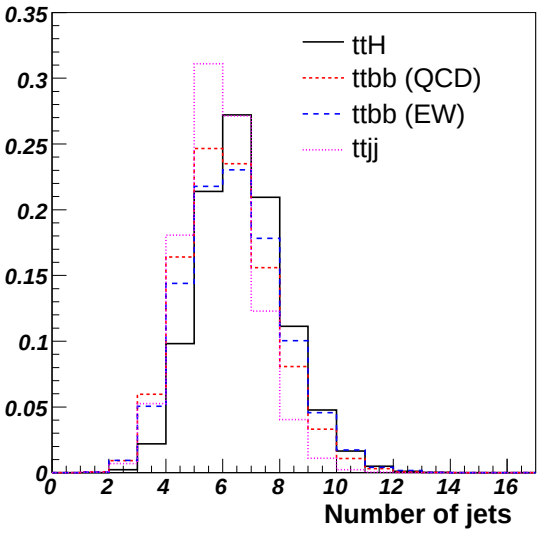

Figure 11: *Multiplicity of jets, inside p_T and η acceptance. All distributions are normalized to unity.*

In this case an approximation can be made by considering events with negative Δ (Eq. 3) as events where a small value of Δ is measured negative because of resolution effects. Therefore Δ can be forced to be equal to zero in order to have exactly one solution ("$\Delta = 0$ approximation").

Another method starts from considering that the large mass of the top quark produced a very boosted W boson. The decay products of the W boson must then be produced preferentially along the direction of the boost. For this reason one can consider $p_{z_\ell} = p_{z_\nu}$ ("collinear approximation").

Considering $t\bar{t}H$ events where one lepton is reconstructed, 32% of the times p_{z_ν} does not have solutions. The actual Δ distributions can be seen in Figure 17. For these events both approximations have been tried, as shown in Figure 17 the direction and the mass of the W boson is better represented by the $\Delta = 0$ approximation.

Since the mass constraint is lost when using the $\Delta = 0$ approximation (same would be for the collinear approximation) there is an actual cut for the reconstructed W boson mass. Only W candidates having a mass less than 140 GeV/c^2 are considered.

6 Reconstruction of the hadronic W boson

Not all the pairs of jets are tried to reconstruct the hadronic W. If exactly 4 b-tagged jets are present, all the other jets are paired to form W candidates. With these criteria, Figure 18 shows the mass distribution of all W candidates found on 30 fb $^{-1}$ of data. Only candidates within 25 GeV from the true W mass are kept. Even with these cuts the hadronic W candidates multiplicity is still very high. All jets used to form W candidates are calibrated with the light-jet calibration.

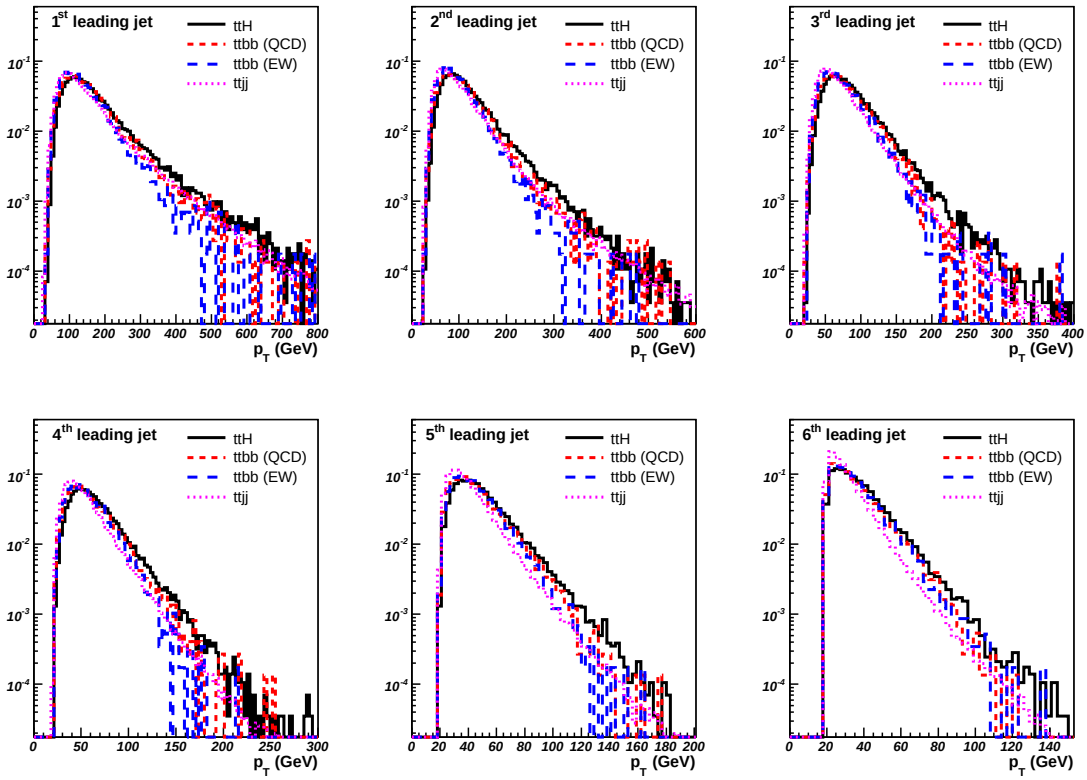

Figure 12: *p_T spectra for the first 6 leading jet in events with at least 6 jets. All distributions are normalized to unity.*

7 Cut-based analysis

This Section describes the algorithms used to reconstruct the $t\bar{t}$ system and consequently the Higgs boson. Starting from the list of leptonic and hadronic W boson candidates, b-jets are associated in order to build the list of top candidates. The combination of b-jets resulting in the best reconstruction for the top candidates is taken as the final choice. Events where there is no combination giving a satisfactory top mass reconstruction are discarded. The two remaining b-jets are used to form the Higgs candidate.

7.1 Top-antitop system and combinatorial background

In each event, top quarks are reconstructed by pairing the two b-jets with the W's in the way which minimizes the χ^2 expressed as:

$$\chi^2 = \left(\frac{m_{jjb} - m_{top}}{\sigma_{m_{jjb}}}\right)^2 + \left(\frac{m_{l\nu b} - m_{top}}{\sigma_{m_{l\nu b}}}\right)^2, \tag{5}$$

where $\sigma_{m_{jjb}}$ and $\sigma_{m_{l\nu b}}$ are the reconstructed mass resolutions estimated in simulated signal events, as shown in Figure 19. The mass distributions for all top selected candidates in the signal sample is shown in Figure 20. In order to reduce background from events where no $t\bar{t}$ pairs were produced,

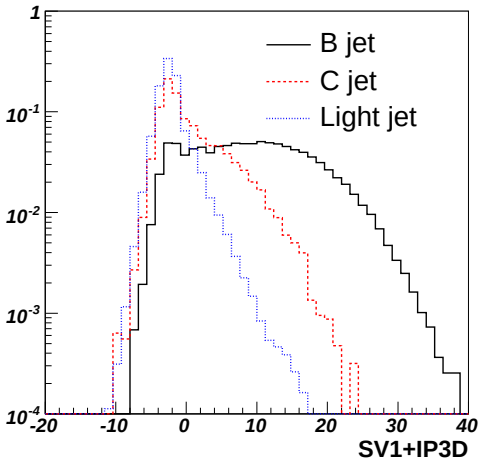

Figure 13: *Distribution of the b-tagging weight for b-, c- and light jets in $t\bar{t}H$ events, using the IP3D+SV1 tagger. All distributions are normalized to unity.*

only events where both top quarks are reconstructed with a mass not more than 25 GeV away from the true top mass are kept.

The two remaining b-tagged jets are used to form the Higgs candidates. The mass distributions for all Higgs candidates in the signal sample is shown in Figure 21, the same for the signal and physics backgrounds normalized to 30 fb $^{-1}$ is shown in Figure 22. Here the difficulty of the analysis is clearly shown, requiring dedicated studies to measure the background normalization and probably its shape in data.

As a final cut, to discriminate against $t\bar{t}$ events where no Higgs is produced, only events in a mass window of 30 GeV from the nominal Higgs mass are used for the final estimation of the cut-based analysis significance.

The effect of the final selection for the cut-based analysis on signal and background samples is shown in Table 6.

mass window cuts	$t\bar{t}H$	$t\bar{t}b\bar{b}$(EW)	$t\bar{t}b\bar{b}$(QCD)	$t\bar{t}X$
$W_{had} + W_{lep}$	71 ± 2	75 ± 8	494 ± 50	289^{+54}_{-51}
$+ t\bar{t}$	58 ± 2	59 ± 7	379 ± 44	191^{+43}_{-40}
+ Higgs	26.4 ± 1.4	$15.4^{+4.2}_{-3.4}$	130 ± 25	60^{+32}_{-23}

Table 6: *Number of events after each final selection cuts for signal and background. All number are normalized for 30 fb^{-1} of integrated luminosity. In the last column the contribution of $t\bar{t}b\bar{b}$ has been removed.*

8 Pairing likelihood analysis

In the previous Section we used a cut-based approach to isolate the top quark decay products. A straightforward improvement to such approach is to use all discriminating topological distributions combined together in order to build a pairing likelihood. As a first step the analysis

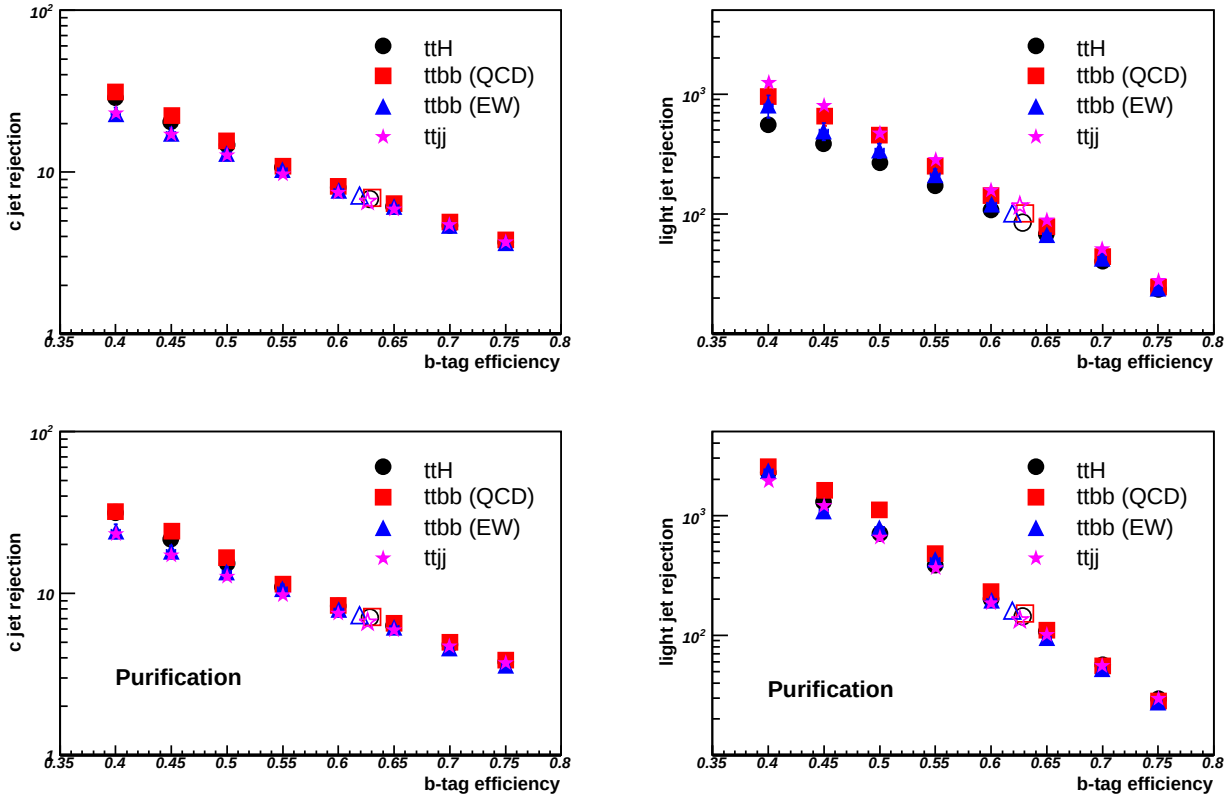

Figure 14: *Rejection of light and c-jets versus b-tagging efficiency. With open red markers is indicated the working point used in this analysis.*

considers only top quark properties as likelihood templates. Even though Higgs boson properties could help associating b-jets, if used, those could lead to bias the background distributions. The variables used, shown in Figure 23 are:

- m_{jj}: invariant mass of the light jets from the hadronic W decay,
- m_{jjb}: invariant mass of the hadronic top decay products,
- $m_{l\nu b}$: invariant mass of the leptonic top decay products,
- $\hat{jj}$: angle between the light jets from the hadronic W decay,
- $\Delta R(jj, b)$: distance in R between the light jets from the hadronic W decay,
- $\Delta R(l, b)$: distance in R between the reconstructed hadronic W and the b-jets from the hadronic top decay.

The output of the pairing likelihood for the correct and wrong b-jets combinations is shown in Figure 24. As shown in this plot, even though the correct distributions are peaked at 1, the wrong combinations have still a large probability to be selected. The only combination used is the one which maximizes the likelihood output. In order to avoid the presence of a large combinatorial contribution only well-reconstructed events are chosen by a cut on the likelihood

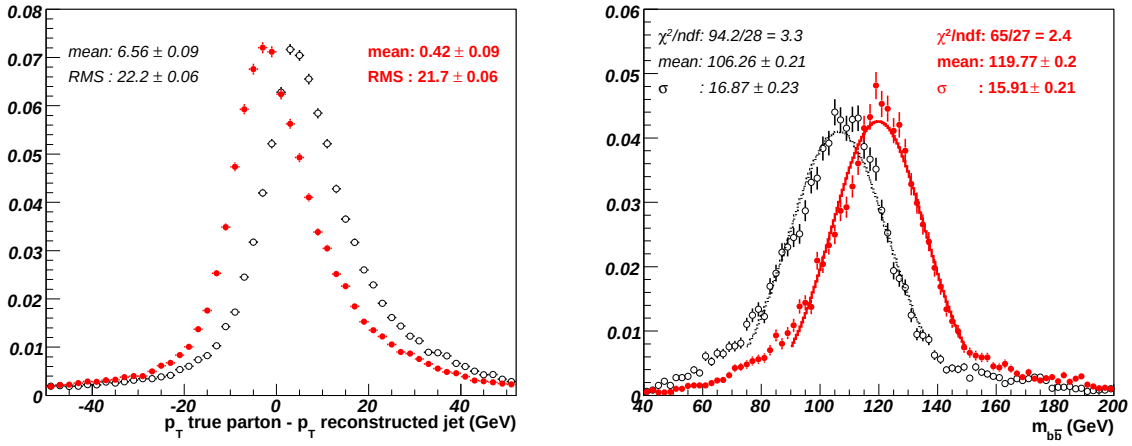

Figure 15: *Impact of jet calibration. The left plot shows the difference between the jet p_T and the actual parton p_T, without (black open circles) and with (red solid circles) the recalibration, for b-jets. Right plot: impact on the Higgs boson mass reconstruction.*

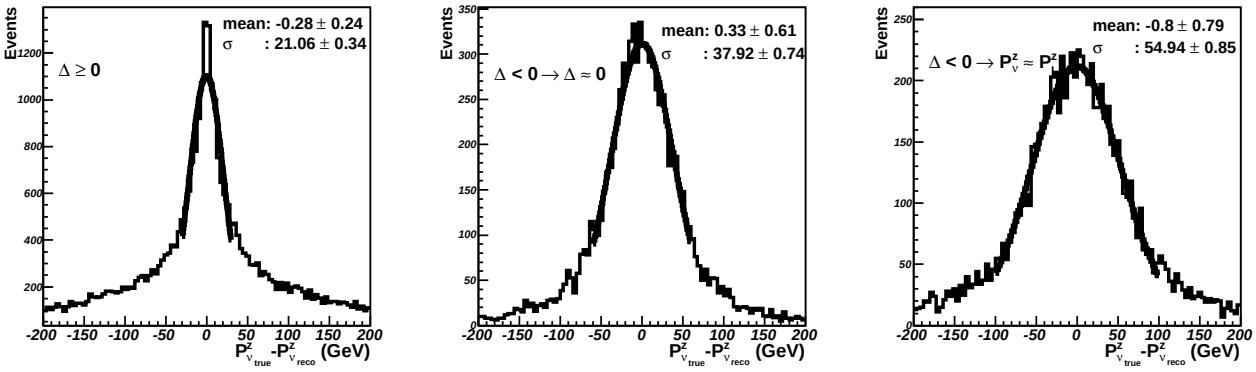

Figure 16: *Neutrino p_z resolution when there are solutions to Eq. 1 (left plot), or when there is no solution and two different approximations are used (center: $\Delta = 0$ approximation, right plot: collinear approximation).*

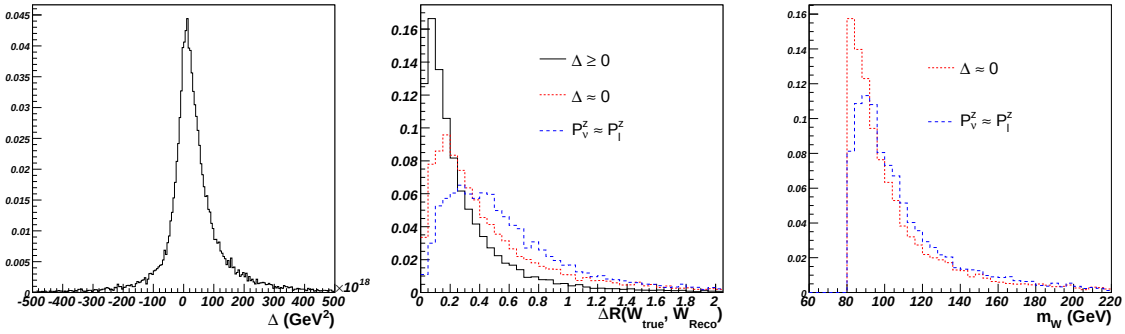

Figure 17: *Distributions of Δ (left hand side), of the ΔR between the true and the reconstructed W boson (center) and of the reconstructed leptonic W mass (right hand side) for events where a solution for p_z is found (black line) and events where an approximation is used (red for $\Delta = 0$, blue for the collinear approximation). All distributions are normalized to unity.*

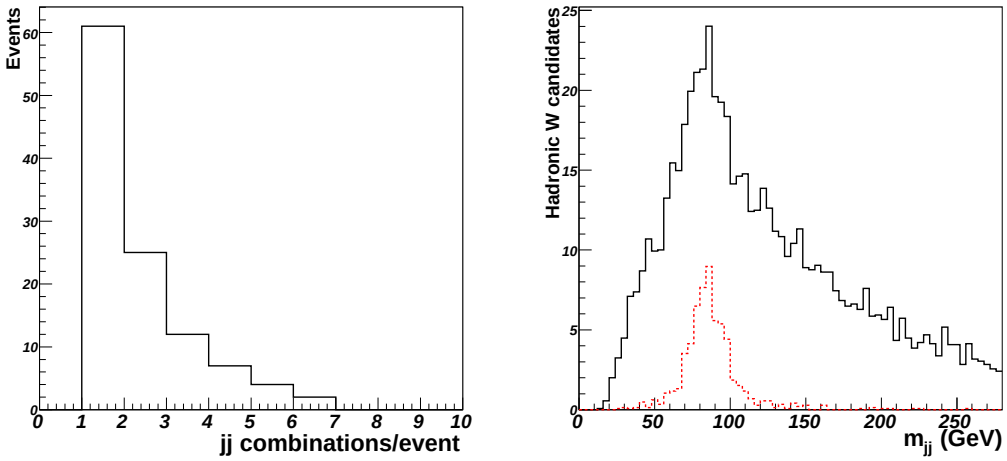

Figure 18: *Mass reconstruction of the hadronic W. Left plot: number of combinations per event. Right plot: invariant mass spectrum for hadronic W candidates. The dotted line are the correct combinations. All distributions are normalized to 30 fb $^{-1}$.*

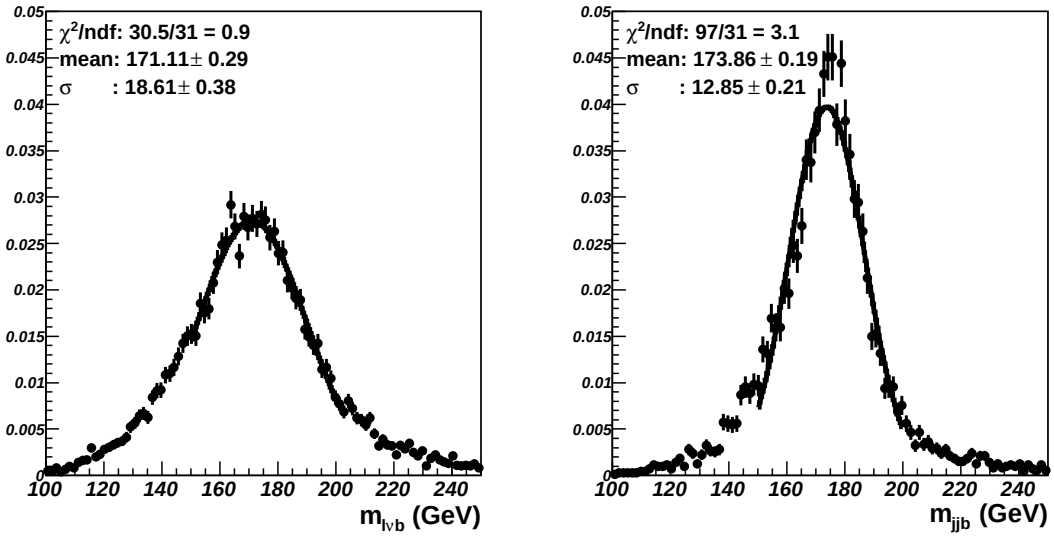

Figure 19: *Mass resolution for reconstructed top quarks in the signal sample, using only correct assignments from the Monte Carlo truth. The results of the fit are used for computing the* χ^2 *to choose the right combination.*

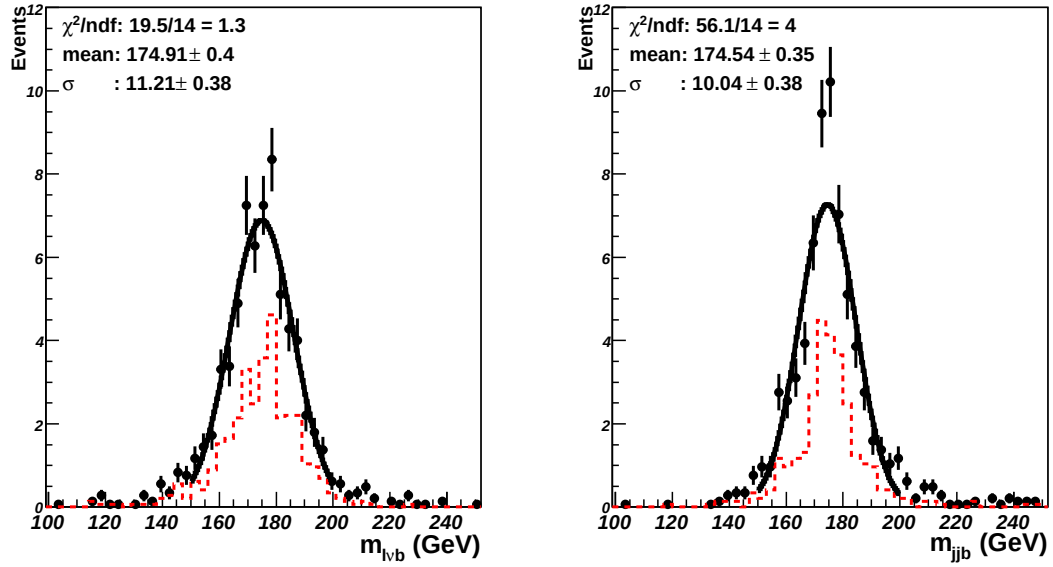

Figure 20: *Reconstructed invariant mass spectrum for top quark selected candidates in the signal sample. The dotted red line indicates the candidates formed by assigning the correct b-jets. All distributions are normalized to 30 fb* $^{-1}$.

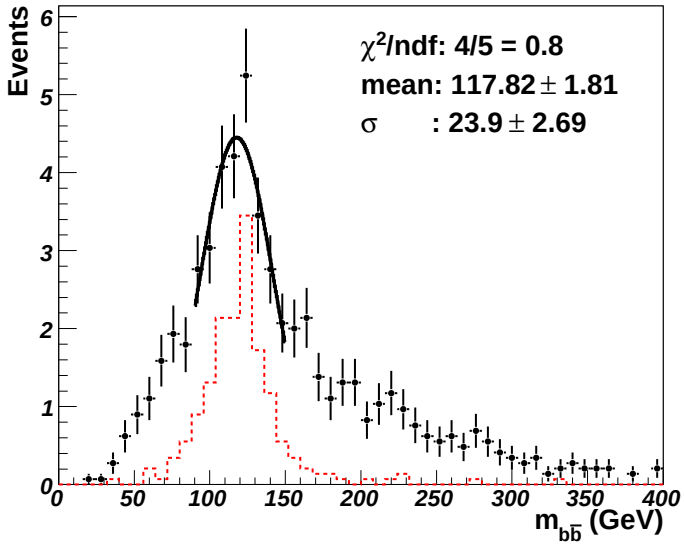

Figure 21: *Reconstructed invariant mass spectrum for Higgs candidates in the signal sample. The dotted red line indicates the candidates formed by assigning the correct b-jets. All distributions are normalized to 30 fb* $^{-1}$.

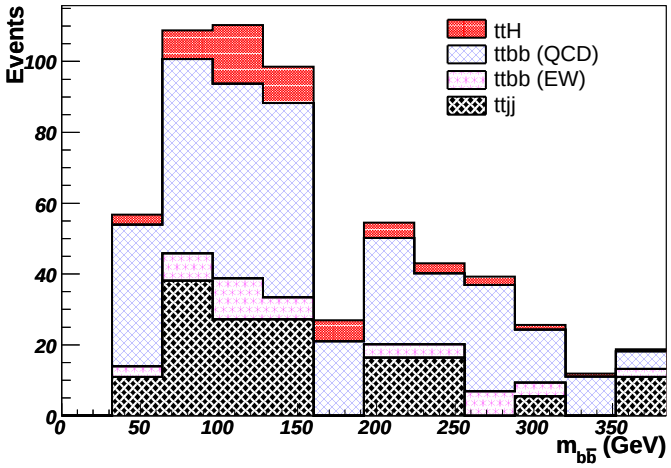

Figure 22: *Reconstructed invariant mass spectrum for Higgs candidates for signal and backgrounds after the cut-based selection. Distributions are normalized to 30 fb* $^{-1}$.

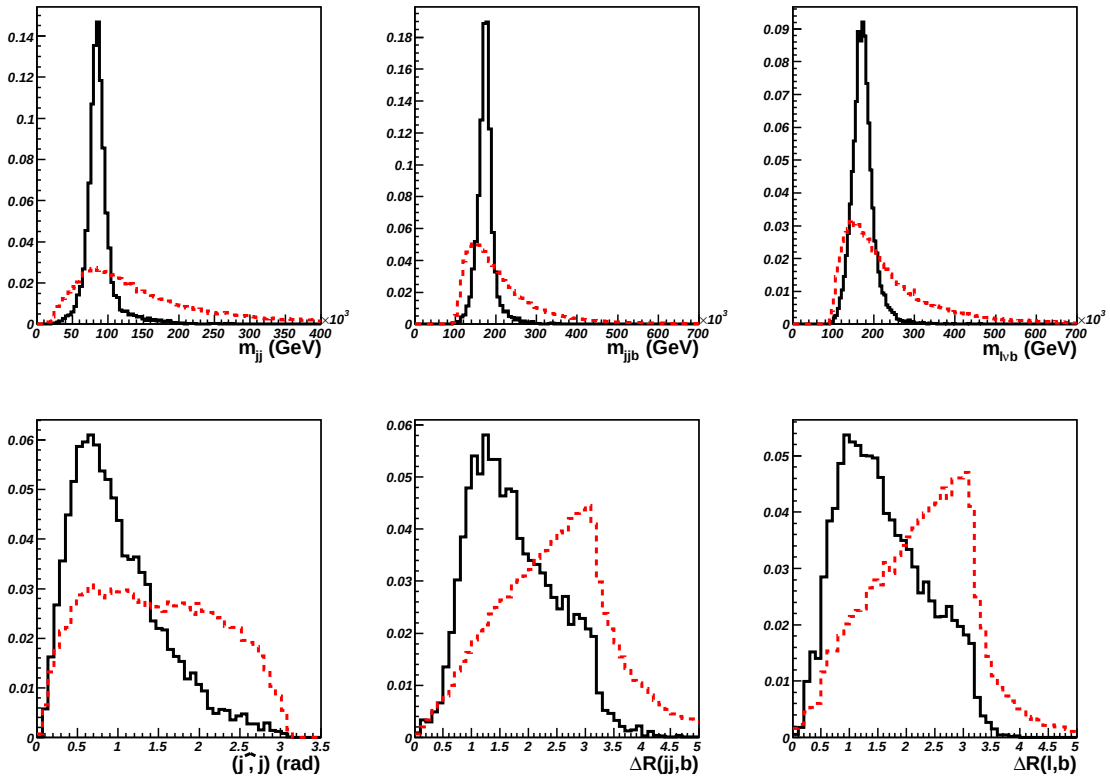

Figure 23: *Pairing likelihood templates for top quark topological distributions, derived from the $t\bar{t}H$ signal sample. Solid lines represent the correct combination while the dotted lines show the combinatorial background in the signal itself. See text for a description of the variables.*

output of 0.9. After this cut, b-jets are associated to reconstruct the Higgs boson, as shown in Figure 24, and the two top quarks, shown in Figure 25. The invariant mass for the selected Higgs candidates for the signal and physics backgrounds normalized to 30 fb $^{-1}$ is shown in Figure 26.

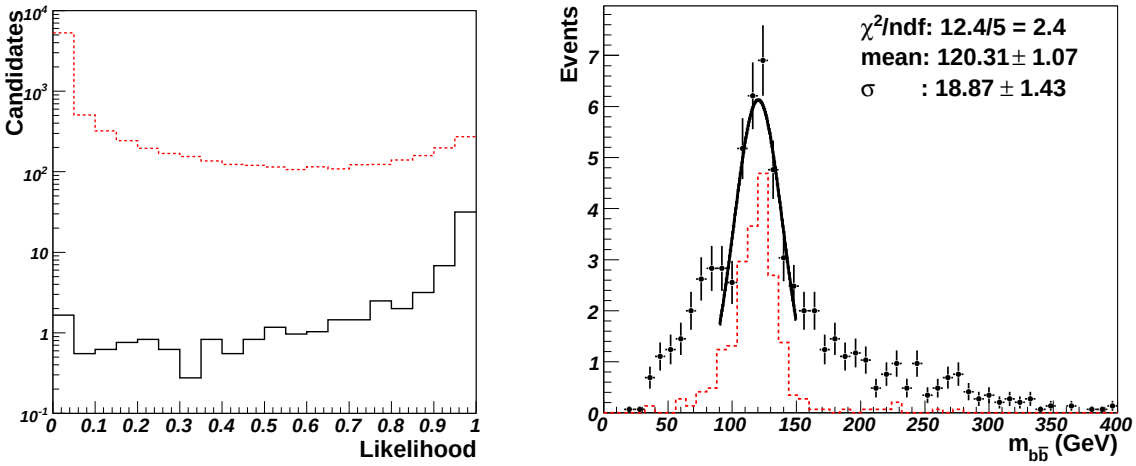

Figure 24: *Left hand side: combinatorial likelihood output for $t\bar{t}H$ events. Black solid (red dotted) histogram indicates the correct (wrong) combinations. Right hand side: invariant mass for the Higgs candidates reconstructed using the maximum likelihood configuration, after applying a cut on the likelihood. Dotted histogram indicates the correct combinations. Distributions are normalized to 30 fb $^{-1}$.*

A final cut on the reconstructed Higgs mass, requiring it to be within 30 GeV of the Higgs boson nominal mass is applied. The event yield for the whole analysis using pairing likelihood is shown in Table 7.

applied cuts	$t\bar{t}H$	$t\bar{t}b\bar{b}$(EW)	$t\bar{t}b\bar{b}$(QCD)	$t\bar{t}X$
Leptonic W	103 ± 3	110 ± 9	793 ± 63	752 ± 76
+ Best likelihood > 0.9	66 ± 2	69 ± 7	454 ± 48	251^{+49}_{-46}
+ Higgs mass window	32.5 ± 1.5	17.7 ± 3.7	125 ± 25	60^{+36}_{-27}

Table 7: *Number of events after each final selection cuts for signal and backgrounds for the pairing likelihood analysis. All number are normalized for 30 fb^{-1} of integrated luminosity. In the last column the contribution of $t\bar{t}b\bar{b}$ has been removed.*

9 Comparison between the pairing likelihood and the cut-based analysis

The performance of the two approaches in terms of purity versus selection efficiency at the end of the analysis can be seen in Figure 27. In the case of the likelihood-based analysis, the different working points are obtained by varying the final cut on the likelihood discriminant. In the case of the cut-based analysis, the same variation is achieved by applying different cuts on the hadronic

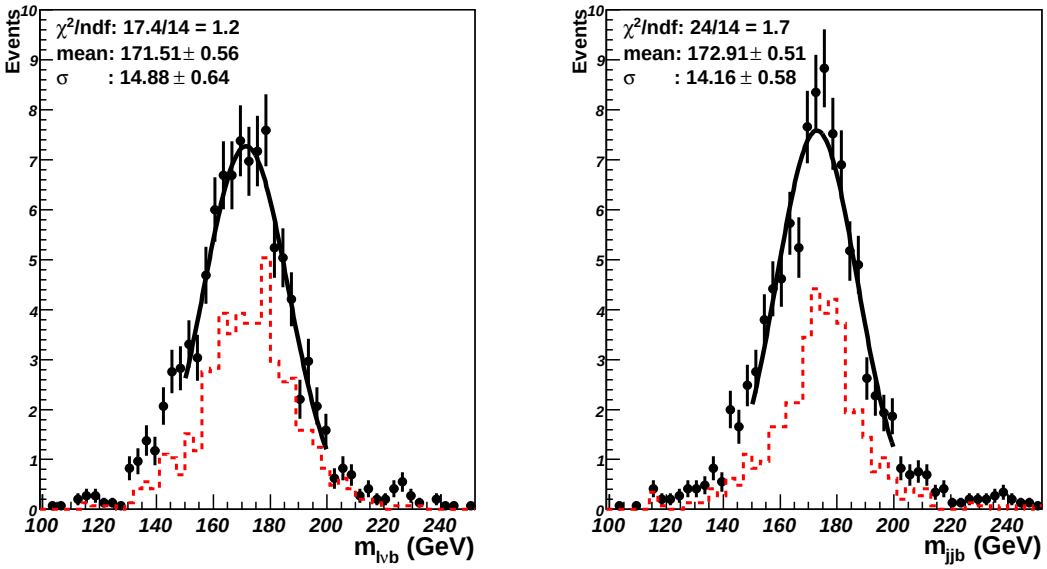

Figure 25: *In the left (right) hand side is shown the leptonic (hadronic) top quark candidates reconstructed invariant mass using the maximum likelihood configuration, after applying a cut on the likelihood output. The dotted histogram indicates the correct combinations. Distributions are normalized to 30 fb* $^{-1}$.

W and top quark reconstructed invariant masses. For this study the efficiency is defined as the selection efficiency relative to the events after the preselection and the leptonic W reconstruction. The purity is defined in terms of the correctness of the assignment of b-jets used to reconstruct the final objects. For instance, one has a *pure* hadronic top quark when the b-jet matches the true b parton from the top quark decay, regardless of whether the same happens for the hadronic W boson decay products.

Figure 27 clearly shows the increase of performance when using more information (likelihood) than just the mass of the reconstructed particles (cut-based).

The chosen working points are indicated with solid markers on Fig. 27. Those points have not been optimized in terms of statistical significance, because of a lack of statistics for the $t\bar{t}X$ background. However, the significance does not change much with the choice of the cut on the likelihood output since the pairing likelihood is not designed to discriminate signal from physics background events.

10 Significance estimates

The number of remaining events in the Higgs mass window have been used to compute a crude estimate of the statistical significance for this channel with 30 fb^{-1}. For such a channel in which the signal and backgrounds are very alike, this naive estimate is not the most relevant figure of merit, but it is still useful to compare analysis. For the cut-based analysis, a significance of 1.8 is achieved. With the likelihood approach the significance is 2.3.

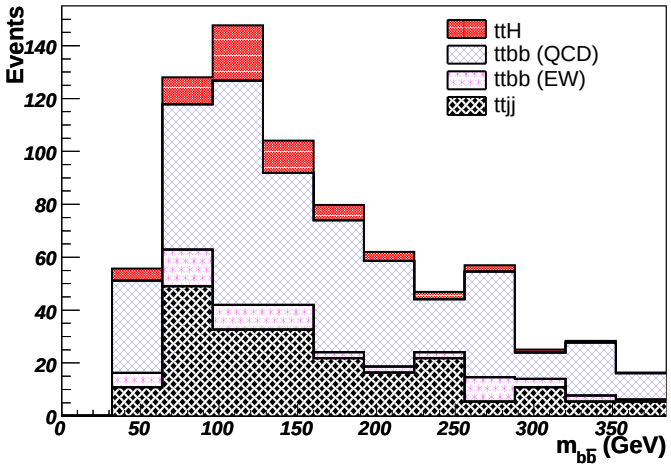

Figure 26: *Reconstructed invariant mass spectrum for Higgs candidates for signal and backgrounds after likelihood selection. Distributions are normalized to 30 fb* $^{-1}$.

11 Nature of the wrong assignments in the signal

As shown in Fig. 24 an incorrect assignment of b-jets when reconstructing the $t\bar{t}$ or the Higgs boson systems has a large effect on the final Higgs mass spectrum. In this section we will discuss the origin of these wrong combinations in the signal sample, in the case of the likelihood analysis.

The first issue we investigated is the composition of the selected objects at the end of the likelihood analysis. In order to understand this, all selected leptons and jets are matched to particles from the Monte Carlo truth. For the case of the jets, Fig. 28 shows the composition of each reconstructed particle in the analysis: b-jets used for the leptonic top, b-jets used for the hadronic top, b-jets used for the Higgs mass, jets used for the hadronic W. First of all, it can be noticed that only in 50% of the cases the reconstruction of the hadronic W uses a jet matched to an actual parton from the W. In about 40% of the times a jet used for the W comes from jets not matched to the $t\bar{t}H$ system, *i.e.* mostly jets from radiations. The remaining source for incorrect jets is due to b-jets mis-tagged as light jets. Therefore right at the beginning of the analysis procedure a large fraction of events have already a wrong combination for the hadronic W. Obviously the effect of the combinatorial background worsen when looking at the two top quarks reconstruction. In about a third of the times, the b-jet used for the top mass reconstruction is matched to a jet coming from the Higgs boson decay. For the Higgs mass reconstruction, the two jets are actually coming from the Higgs decay only in 27% of the times.

To isolate the effect of the likelihood pairing the same exercise is repeated in events where all jets used to reconstruct the $t\bar{t}H$ system have been correctly matched. In addition, we use only events where all the jets matched to the b partons are actually b-tagged. The final object composition is then shown in Fig. 29. At this point there is no more exchange possible between light and b-jets: this is the case of a perfect b-tagging and jet pre-selection. As espected, the hadronic W reconstruction is much better, since the mis-tagging contribution was already small. This points to a pre-selection problem for the jets forming the W boson: unreconstructed jets or acceptance problem. The impact on the hadronic top mass reconstruction is also noticeable bringing to the conclusion that improving the b-tagging and perhaps lowering the threshold for some of the jets could lead to an overall amelioration of the analysis. However, the exchange

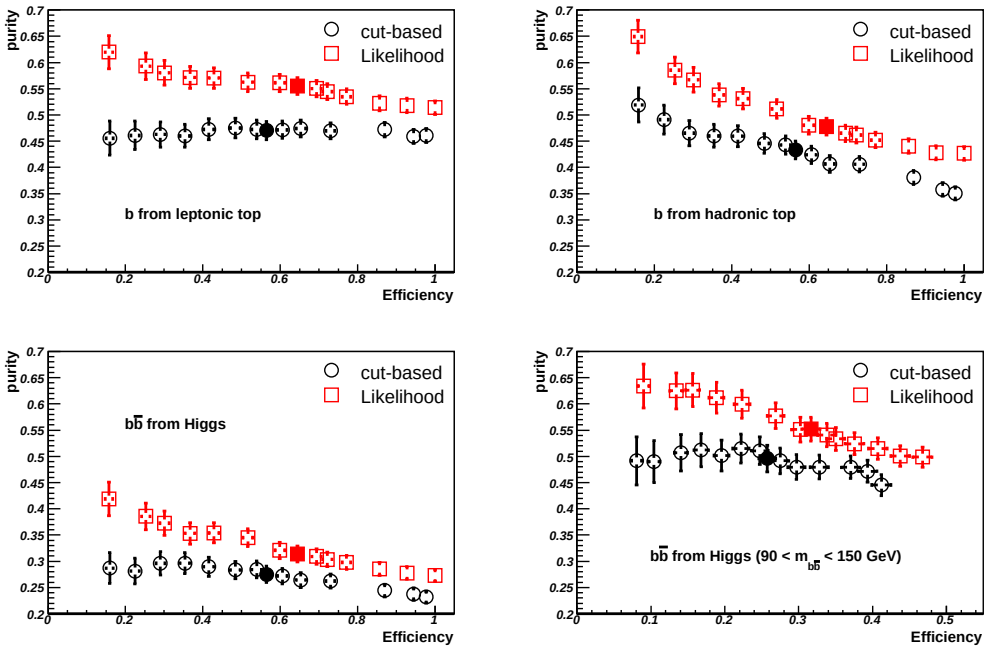

Figure 27: *Comparison of the purity of reconstructed objects versus selection efficiency for the cut-based and the pairing likelihood analysis. Top left (right) side: percentage of b-jets matched to b-parton from leptonic (hadronic) top decay. Bottom left (right) side: purity of the reconstructed bb invariant mass before (after) the final mass window cut. The chosen working points are shown with solid markers.*

of b-jets between top and Higgs decays still remains: more powerful multivariate analysis could help in resolving this ambiguity.

12 Conclusion

We performed a baseline sensitivity study for the detection of a SM Higgs when produced together with a $t\bar{t}$ pair. After the definition of a common pre-selection, two different techniques are used, both aimed at the reconstruction of the $t\bar{t}$ system. The first one is based on the reconstruction of the top and W candidate masses (cut-based analysis). The second one uses a more complete description of the kinematic properties of the $t\bar{t}$ system to build a likelihood discriminant and isolate the jets coming from the Higgs boson decay. We see an increase of the overall significance of about 25% using the latter approach. In the last part of this work we investigated the sources of the combinatorial background with some conclusions on the possible ameliorations.

From this study emerges the necessity of a strong b-tagging algorithm which is important not only to suppress the $t\bar{t}$+jets physics background but also to help reducing the combinatorial background by improving the hadronic W reconstruction. It is also clear that the combinatorial background, responsible for the dilution of the Higgs mass peak, needs to be further reduced using possibly multivariate techniques, in order to improve the statistical significance of the channel.

The next studies need to cover the fundamental aspects of the measurement of the background

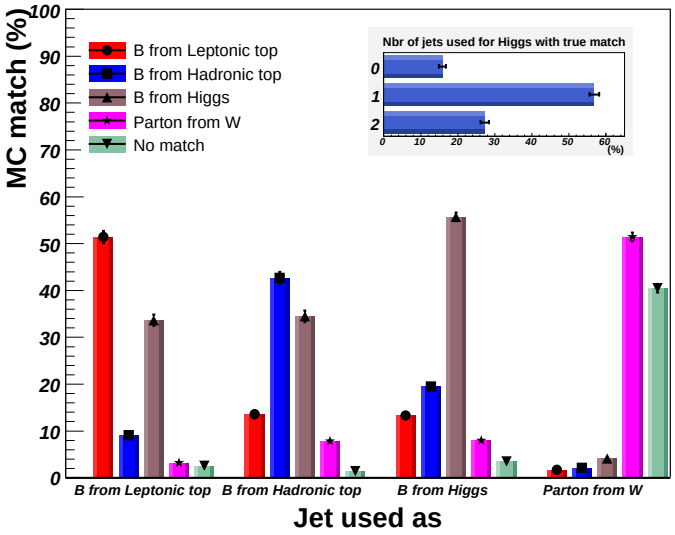

Figure 28: *Origin of jets used for the reconstruction of the various objects in the analysis, from left to right: b-jet used for the leptonic top, b-jet used for the hadronic top, b-jets used for the Higgs mass, jets used for the hadronic W. The colors show the actual origin from the Monte Carlo truth. The inset box shows the fraction of events for which none, one or the two jets used for the Higgs mass are matched to the actual daughters of the Higgs boson in the Monte-Carlo truth.*

rates and possibly shapes in data, and of the extraction of the signal in the presence of a *quasi-signal-like* background, as shown in the invariant mass plots at the end of both analysis.

Acknowledgments

We thank the European Commission partly responsible for the funding of this work through the Marie Curie International Reintegration Grant MIRG-CT-2005-21700. Our special thanks go to Jean-Baptiste de Vivie for his numerous useful contributions to this work, for his help in defining and validating most of the Monte Carlo samples, for his availability on discussing the issues related to the generation of the physics background and finally for a careful reading of this paper. We also thank the production team for producing the Monte Carlo samples, Wouter Verkerke for providing us with the MC@NLO input files for the sample 5212 and Karim Bernardet for the production of this sample and for his technical help throughout the whole process. Finally, we thank the members of the CSC $t\bar{t}H(H \to b\bar{b})$ sub-group and the Higgs working group conveners for useful discussions, and Chris Collins-Tooth for his help with the production of the $t\bar{t}H$ ntuples.

A Details on datasets and analysis software

This analysis used the officially produced AOD files mentioned in Table 8. From those AOD files, ntuples were built using release 12.0.6 and the package HiggsAssocWithTopToBBbar which

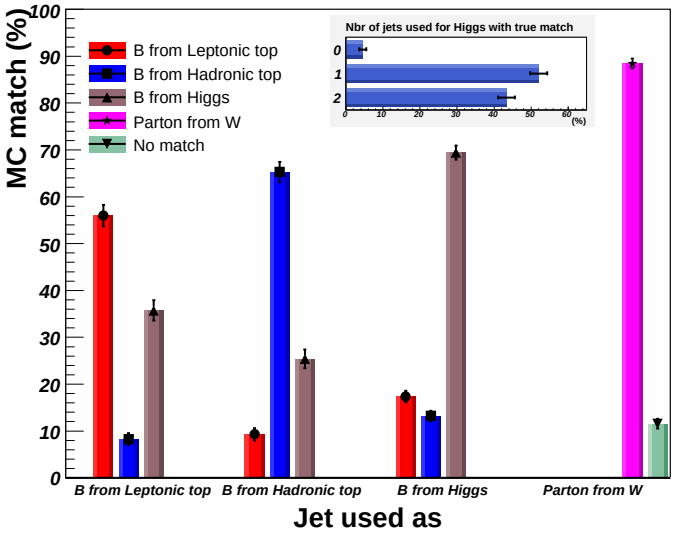

Figure 29: *Same as Fig. 28. However only events having correct matches for all the jets in the $t\bar{t}H$ system, and for which all the jets matched to the b partons are b-tagged, are used.*

is the standard interface to Event View [16] for the $t\bar{t}H(H \to b\bar{b})$ working group.

Process	Dataset
$t\bar{t}H$	trig1_misal1_mc12_V1.005870.ttH_poslepnu_jj_bb.recon.AOD.v12000601_tid008657 trig1_misal1_mc12_V1.005870.ttH_poslepnu_jj_bb.recon.AOD.v12000601_tid010770 trig1_misal1_mc12_V1.005871.ttH_neglepnu_jj_bb.recon.AOD.v12000601_tid008658 trig1_misal1_mc12_V1.005871.ttH_neglepnu_jj_bb.recon.AOD.v12000601_tid010773
$t\bar{t}b\bar{b}$ (QCD)	trig1_misal1_mc12.006886.AcerMCttbb.recon.AOD.v12000605_tid012137
$t\bar{t}b\bar{b}$ (EW)	trig1_misal1_mc12_V1.005214.AcerMCttbb_EW.recon.AOD.v12000601_tid007853
$t\bar{t}$+jets	trig1_misal1_mc12.005212.McAtNlo_ttbar_plus_jets.recon.AOD.v12000605_tid010093

Table 8: *AOD datasets used for this study.*

References

[1] ALEPH, DELPHI, L3 and OPAL Collaboration, R. Barate et al., *Phys. Lett.* B565 (2003) 61

[2] M.L. Mangano et al., *JHEP* 0307 (2003) 001 (and updates)

[3] E. Richter-Was, D. Froidevaux and L. Poggioli, ATL-PHYS-98-131 (and updates)

[4] ATLAS Collaboration, ATLAS detector and physics performances. Technical design Report. Vol 2, CERN-LHCC-99-15 (1999)

[5] J. Cammin and M. Schumacher, ATL-PHYS-2003-024

[6] S. Corréard, Ph.D. Thesis. CPPM, 2005.

[7] CMS Collaboration, CMS Physics TDR, CERN-LHCC-2006-21 (2006)

[8] S. Frixione and B.R. Webber, *JHEP* 0206 (2002) 029 (and updates)

[9] Y.Q. Fang et al., ATL-PHYS-2004-035

[10] T. Sjostrand, S. Mrenna and P. Skands, *JHEP* 0605 (2006) 26

[11] M. Spira, http://people.web.psi.ch/spira/proglist.html

[12] W. Beenakker et al., *Nucl. Phys.* **B653** (2003) 151

[13] A. Djouadi, J. Kalinowski and M. Spira, *Comp. Phys. Comm.*108 (98) 56

[14] B. Kersevan and E. Richter-Was, *Comp. Phys. Comm.*149 (2003) 142 (and updates)

[15] S. Corréard et al., ATL-PHYS-2004-006

[16] https://twiki.cern.ch/twiki/bin/view/Atlas/EventView