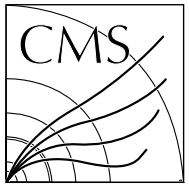
Available on CMS information server

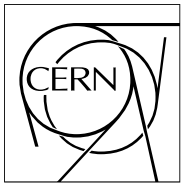
CMS NOTE 2006/119



The Compact Muon Solenoid Experiment

CMS Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



12 June 2006

Search for $H \rightarrow b\bar{b}$ in association with a $t\bar{t}$ pair at CMS

S. Cucciarelli

CERN, Geneva, Switzerland

A. Schmidt, C. Weiser

Institut für Experimentelle Kernphysik, Universität Karlsruhe (TH), Karlsruhe, Germany

C. Riccardi, P. Torre

Dipartimento di Fisica Nucleare e Teorica, Università degli Studi di Pavia and INFN, Pavia, Italy

D. Benedetti, A. Santocchia

Dipartimento di Fisica, Università degli Studi di Perugia and INFN, Perugia, Italy

C. Hill, J. Incandela, S.A. Koay

University of California, Santa Barbara

Abstract

The discovery of a low mass Higgs boson is particularly challenging because the leading decay mode is $H \rightarrow b\bar{b}$ for $m_H < 135 \mathrm{GeV}/c^2$. Direct Higgs production is thus overwhelmed by the more mundane QCD production of $b\bar{b}$ pairs. This leads one to more rare, but also more accessible modes such as $H \rightarrow \gamma\gamma$. Nevertheless, the degree to which the Higgs is manifested in a variety of channels and modes could hold important clues to its true identity, and whether it heralds from a more rich Higgs sector than that of the Standard Model. A potentially promising channel is the production of Higgs in association with a $t\bar{t}$ pair. The search for $H \rightarrow b\bar{b}$ is enhanced by properties of top quark decays that allow a dramatic reduction in possible backgrounds. These events contain four energetic b jets and fall into three salient subclasses depending upon whether neither, one, or both of the W bosons in the event decay to electrons or muons and their respective neutrinos. In this note we present results of studies of all three subclasses of $t\bar{t}H$ events in CMS with an integrated luminosity of 60 fb^{-1}. The search for a low mass Higgs in $t\bar{t}H$ relies upon every subsystem in the CMS experiment and relies particularly strongly upon the CMS Tracking System. It is very challenging.

1 Introduction

The Higgs boson decay to $b\bar{b}$ is the dominant branch for the Higgs mass range up to $m_H \sim 135$ GeV$/c^2$. Direct Higgs production is impossible to detect via this decay as a result of the combination of an overwhelming QCD cross section for $b\bar{b}$ production and the inability to reconstruct the Higgs mass very precisely. While the latter is still true in the case of Higgs production in association with a $t\bar{t}$ or $b\bar{b}$ pair, these channels hold promise because they entail substantially lower backgrounds. The Top quark decay to Wb has a branching fraction of very nearly 100%. As a result, $t\bar{t}H$ events in which the Higgs decays to bottom quarks that almost always contain four b quarks in the final state. A further separation of these events into 3 salient topologies follows as a result of the ways in which the two W bosons decay. Thus, in addition to the four b jets, roughly 49% of these events also contain four hadronic jets (the all-hadron channel), while some 28% have two hadronic jets together with an isolated electron or muon and missing E_t (the semi-leptonic channel), finally of order 5% contain two oppositely charged leptons (either can be an electron or muon) and missing E_t (the di-lepton channel). The remaining 14% or so of events correspond to those cases where one or both of the W bosons decay to a tau lepton and neutrino and are not easily distinguished as such, as a result of the rich decay repertoir of the tau meson. In fact, these events do contribute in small part to the three other classes of events in actual analyses. Additional hadronic jets appear in these events and originate from initial and final state QCD radiation (IFSR). The variety and complexity of these events, and particularly the appearance of a such a large number of high energy jets and/or leptons requires one to rely upon and fully exploit the the performance of all components of the CMS detector. The Tracking system is of particular importance, since the detection of b-jets is crucial to the identification and reconstruction these events.

In this note we present results of studies performed on Monte Carlo data samples for which the CMS detector has been fully simulated. Every attempt has been made to process the data as would be the case for real data acquired in proton-proton collisions at a center of mass energy of 14 TeV. Thus, online triggers have been simulated, and all of the standard offline reconstruction algorithms as they currently exist have beem employed in the processing of the data.This emphasis on realism has several important consequences and by-products. Firstly, the analyses are substantially more complicated than studies involving more simple approaches involving parametric detector modelling. This has the advantage of providing a more true picture of the complexity of the events under study, highlighting some of the limitations that might otherwise be overlooked. On the other hand, because these studies are undertaken at a time when no real data is available, it is also not possible to take advantage of the most important tools in the arsenal available to high energy hadron collider experimentalists: the multitude of control samples resulting from the broad-band nature of hadron collider physics. In all previous and currently operating hadron collider experiments, it is the availability of a rich array of data sets containing a variety of final state objects in useful ranges of transverse energy that allows one to obtain even very precise measurements in a very difficult environment. With this in mind, the studies chronicled in this note attempt to answer two parallel lines of questions. One line focuses on our current uncertainties in such things as, for example theoretical cross sections for signal and background, and estimates of detector performance. The other line attempts to understand the ways in which real data can and will be used to greatly diminish such uncertainties. The conclusions of this note include an attempt to combine these threads to provide the most useful forward looking information that can be provided at this moment. Namely, the smaller uncertainty estimates expected during actual data taking are used to determine specific expectations for sensitivities to a Higgs signal in $t\bar{t}H$ events, while present uncertainties are used to provide expected ranges for these sensitivities. Thus, as large control data sets become available in early operation and background estimates and detector performance are better understood, this information can be used to provide more refined estimates of sensitivity to a Higgs boson in this channel. In addition, sensitivities calculated in the manner currently prescribed by CMS for discovery studies are also presented.

The note has the following overall structure. In Section 2 the event generation and simulation of signal and background samples are described. Section 4 details how the various signature objects (leptons, jets, b jets, and missing E_T) of the final state are reconstructed, while a detailed description of the event selection for each of the main event topologies and the attempt to reconstruct the invariant Higgs mass for some of the channels is given in Section 5. All the results are for an integrated luminosity of 60 fb^{-1}.

2 Event Generation and Simulation

Because the identification of signal relies upon the presence of top quark decay products, it comes as no surprise that the most significant backgrounds are those associated with $t\bar{t}$ events themselves. The main backgrounds turn out to be $t\bar{t}jj$, $t\bar{t}b\bar{b}$ and $t\bar{t}Z$ with $Z \rightarrow b\bar{b}$.

These processes are studied in detail as presented in this note. Secondary background sources include pure QCD

multijet events in the case of the all-hadron channel, and W/Z+jets or diboson+jets events in the case of the semi-leptonic and di-lepton channels. With the exception of QCD multijets, the latter have substantially lower production cross-sections than $t\bar{t}$ events but very analogous structures. They are therefore not studied in detail, and instead their much smaller contributions are estimated and bounded by conservative back-of-envelope calculations, as discussed in Appendix B of this note.

The primary Monte Carlo data samples used in this analysis are summarised in Table 1 along with some of their generation information. For the $t\bar{t}H$ signal and the irreducible $t\bar{t}b\bar{b}$ background we used compHEP [1] plus PYTHIA. Though a leading order Monte Carlo, PYTHIA is known to do a very good job of reproducing IFSR as well as parton shower effects. For the $t\bar{t}$ plus jets backgrounds, greater care must be exercised. In particular, PYTHIA alone cannot be expected to do a realistic job since the processes that are relevant are not leading order. On the other hand, there is currently no full NLO MC for $t\bar{t}$ plus jets. As a result, one uses higher order matrix elements including additional radiated partons in conjunction with the parton showering of PYTHIA to produce the appropriate event topologies. This is not trivial because the soft QCD effects represented by the PYTHIA parton shower program are not completely distinct from the higher order perturbative diagrams. While there are jet energies for which the two are clearly distinct, they nevertheless represent two extremes in a continuum and so one is forced to artificially place a boundary betwen them. This is done by what is called matching.

The analyses presented in this note use ALPGEN + PYTHIA for the matrix elements and parton showering, respectively, for the $t\bar{t}$ plus n jets backgrounds. The matching is done in ALPGEN as discussed in Ref. [3]. In particular, all of the matrix elements for $t\bar{t}$ plus n additional hard partons are included and properly combined at each order, taking into account interference between amplitudes. These are interfaced to PYTHIA which then proceeds to generate parton showers and IFSR. The final generated event is then checked to see if the number of hard partons in the final state is in fact n and not greater than n, for exclusive n additional parton samples. Events with more than n hard partons can occur as a result of the high energy extremes of the parton shower program in PYTHIA. A carefully chosen set of criteria for selecting partons based on their origin, angular separation from other partons, and their energy, allows one to avoid overproduction of extra jets due to this overlap of the parton showering and matrix element sources of hard partons.

It should be noted that the Alpgen samples used for these studies were available only recently. As a result, much of the work that was done to optimize the selection of events was carried out with CompHEP plus PYTHIA samples. These are also described in the tables and in further detail below. These samples did not use the pruning mechanism used in the matching portion of Alpgen and so typically had more additional partons as a result of the double counting alluded to above. In other words, CompHEP substantially over-estimated the cross sections for production of top quark pairs with additional hard partons. The actual character of the events, for a given number n of additional partons, do not differ substantially from ALPGEN events with the same number of additional partons. A detailed comparison of these samples is presented in Appendix A. The only reason for mentioning the CompHEP samples at all is that the analysis cuts were tuned on these samples and then used on the final ALPGEN samples. For this reason it was important to verify that for a given number of additional partons, the events did not differ too much in character.

Table 1: Generation parameters used for producing signal and background datasets.

Channel	Generator	PDF	Detector Simulation	Digitization
Fully-hadronic $t\bar{t}H$	CompHEP + PYTHIA 6.215	CTEQ4L	CMSIM 133 (GEANT 3)	ORCA 7.6.1
Semi-leptonic $t\bar{t}H$	CompHEP + PYTHIA 6.215	CTEQ4L	CMSIM 133 (GEANT 3)	ORCA 7.6.1
Dilepton $t\bar{t}H$	PYTHIA 6.227	CTEQ5L	OSCAR 3.9.8 (GEANT 4)	ORCA 8.13.1
$t\bar{t}b\bar{b}$	CompHEP + PYTHIA 6.215	CTEQ4L	CMSIM 133 (GEANT 3)	ORCA 7.6.1
$t\bar{t}Z$	CompHEP + PYTHIA 6.215	CTEQ4L	CMSIM 133 (GEANT 3)	ORCA 7.6.1
$t\bar{t}$Nj	ALPGEN 2 + PYTHIA 6.325	CTEQ5L	OSCAR 3.9.8 (GEANT 4)	ORCA 8.13.1

QCD events were generated with PYTHIA in the $\hat{p}_t$ ranges from 120 to 170 GeV/c and greater than 170 GeV/c. For the simulation of the interaction with the detector, the CMS detector simulation program, CMSIM, was employed, which provides a full GEANT3 simulation of the CMS detector. The detector response was digitized with low luminosity Pile-Up using the CMS reconstruction software, (ORCA). The next-to-leading order (NLO) signal cross-sections for different Higgs mass hypotheses are listed in Table 2 together with the branching ratios of $H \rightarrow b\bar{b}$ [2].

A generator level filter,described below, was applied to the QCD, $t\bar{t}b\bar{b}$ and $t\bar{t}Z$ events. In order to apply requirements

Table 2: NLO signal cross-sections and H $\rightarrow$ b$\bar{\text{b}}$ branching ratios for different Higgs mass hypotheses.

m_{H}	115 GeV/c^2	120 GeV/c^2	130 GeV/c^2
σ_{NLO} (pb)	0.747	0.664	0.532
$BR(\text{H} \rightarrow \text{b}\bar{\text{b}})$	0.731	0.677	0.525

on jet kinematics, jet reconstruction was performed at generator level with the PYCELL routine in PYTHIA. The filter used for the QCD events consists of the following cuts: $\sum_{jet} E_t^{jet}$ greater than 300 GeV, E_t of the two most energetic jets in the event must be above 40 GeV and the E_t of the third and fourth leading jets must be above 30 GeV. Background events were passed on to the full simulation only if they pass the above filter or if they have an isolated lepton with p_t >10 GeV/c in the pseudorapidity range $|\eta|$ <3, and the summed p_t of particles inside a cone with $\Delta R = 0.2$ around the lepton direction is below 1 GeV/c. In order to select a b-enriched QCD sample, an additional requirement of at least three b quarks in the final state was applied in addition to the aforementioned filter. The leading order CompHEP cross-sections of background processes together with the effective cross-

Table 3: LO CompHEP cross-sections and effective cross-sections after the generator filters of the considered background processes.

	QCD $\hat{p}_t$=120-170 GeV/c	QCD $\hat{p}_t$ >170 GeV/c	t$\bar{\text{t}}$b$\bar{\text{b}}$	t$\bar{\text{t}}$Z
σ_{LO} (pb)	$3.82 \cdot 10^5$	$1.05 \cdot 10^5$	3.28	0.65
$\sigma_{LO} \times \varepsilon$ (pb)	76.4	336.0	2.82	0.565

sections after the generator filters are listed in Table 3.

Since no parton shower matching criteria are applied to these CompHEP samples they suffer from double countng of hard partons in the final state as discussed earlier, leading to a significant overestimate of the rate of extra jet production. For the t$\bar{\text{t}}$ +1j exclusive, (i.e one and only one extra jet), t$\bar{\text{t}}$ +2j exclusive, t$\bar{\text{t}}$ +3j exclusive, and t$\bar{\text{t}}$ +4j inclusive, (i.e. with at least 4 extra jets) samples that were used in this analysis, the ALPGEN generator was used. Taken together, these constitute a t$\bar{\text{t}}$jj sample with essentially no double-counting. The ALPGEN samples were generated at parton level using CTEQ5L parton distribution functions and a Top mass of 175 GeV/c^2. The fragmentation and IFSR have been performed with PYTHIA. For the interaction with the detector, the newer CMS detector simulation, OSCAR, wa used, which provides a full GEANT4 simulation of the CMS detector. The detector response was digitized with low luminosity Pile-Up using ORCA. The t$\bar{\text{t}}$jj events in which the additional jets originate from b quarks are rejected. They are expected to be uncorrelated with events from the t$\bar{\text{t}}$b$\bar{\text{b}}$ event sample, although a small amount of double counting could be present due to the gluon splitting into b$\bar{\text{b}}$. In view of the fact that the ALPGEN samples are expected to have the most accurate associated cross-sections, while the cross sections for the CompHEP samples are too high by a substantial amount, the use of the CompHEP samples for the study of this particular background is conservative. As will be seen later, this background comes nowhere near to dominating the total background contribution. Nevertheless, the irreducible t$\bar{\text{t}}$b$\bar{\text{b}}$ background should probably not be separated from the t$\bar{\text{t}}$ +n jets samples in future studies.

Table 4: LO ALPGEN cross-sections for the different jet multiplicity samples.

	exclusive t$\bar{\text{t}}$+1j	exclusive t$\bar{\text{t}}$+2j	exclusive t$\bar{\text{t}}$+3j	inclusive t$\bar{\text{t}}$+4j
σ_{LO} (pb)	170	100	40	61

The ALPGEN cross sections for the different jet multiplicity processes are listed in Table 4. A detailed comparison of ALPGEN versus CompHEP for the t$\bar{\text{t}}$jj background is presented in Appendix A. As noted earlier, the CompHEP samples in general are only relevant because they were used to optimize the event selection criteria. The comparison demonstrates that the optimisation should also be valid for ALPGEN samples. All the results that are presented in the following for the t$\bar{\text{t}}$Nj backgrounds were obtained using ALPGEN samples.

3 Level 1 and High Level Trigger Selections

A dedicated $t\bar{t}H$ trigger was not available and therefore could not be implemented in the analysis. As a result, it is assumed in what follows that the signal is recorded by the CMS Level 1 (L1) and High Level Triggers (HLT) as described in Ref. [4]. Wherever possible, the cleaner signature of at least one isolated lepton in the final state is exploited. The semi-leptonic channels thus use the single muon (stream #43) or single electron (stream #2) trigger, using a p_t threshold of 19 GeV/c or 26 GeV/c, respectively. The OR of the single muon, single electron, and single tau streams is taken for the di-lepton channel. The same trigger setups as for streams #43 and #2 were used, except that the p_T threshold was lowered to 15 GeV/c to permit selection of 20 GeV/c leptons later in the analysis. The tau trigger is the official stream (bit #91). When there is no lepton, jet triggers are used to select all-hadron events. In particular, the single-jet, 3-jet and 4-jet triggers with low luminosity thresholds[4] are combined (stream #120 or #122 or #123), using E_t thresholds of 572, 195 and 80 GeV, respectively.

Efficiencies for the various HLT and L1 triggers that were used are presented in Table 5, which requires some additional explanation to be correctly interpreted. The efficiencies quoted are determined by counting the numbers of accepted events relative to the total numbers of events in these samples. In order to streamline the various studies that were performed, the analyses used different MC samples, produced with different final state constraints. Thus, efficiencies for single muon, single electron and fully hadronic final states were defined with respect to exclusive signal samples and inclusive background samples, as described in the preceeding section of this note. The di-lepton channel efficiency on the other hand, was defined with respect to samples containing at least one leptonic top decay for the signal and inclusive samples for the backgrounds.

As a final comment, it should be noted that during data taking, more sophisticated triggers will certainly be available and could implemented for this search. They would combine such elements of the signatures of these events as the presence of leptons, missing transverse energy (MET), multiple jets, and possibly even some loose b tagging information. As such, the efficiency for collecting these events on tape will very likely be much higher than implied by the results shown here. This point is revisited in the concluding sections of this note where the implications of various scenarios affecting the estimated counts of signal and background are addressed.

Table 5: Signal and background efficiencies of the Level 1 and High Level Triggers.

	Single μ	Single e	Single e OR μ OR τ	Jets
$H \to b\bar{b}$ (%) with $m_H = 120\text{GeV}/c^2$	63.5	52.4	76.7	24.9
$t\bar{t}b\bar{b}$ (%)	19.0	16.1	83.6	18.3
$t\bar{t}$1j (%)	13.9	11.3	53.0	2.9
$t\bar{t}$2j (%)	14.0	11.1	59.8	6.2
$t\bar{t}$3j (%)	14.0	11.1	68.5	11.4
$t\bar{t}$4j (%)	13.4	11.1	78.6	31.4
$t\bar{t}Z$ (%)	20.4	18.8	84.4	25.3
QCD 120-170 GeV/c (%)	0.08	0.8	4.3	1.7
QCD >170 GeV/c (%)	0.07	2.1	4.4	10.3

4 Reconstruction

4.1 Muon Reconstruction

The process of muon reconstruction begins in the Muon Chambers and is then extended to the tracking system, as described in Ref. [6]. For the studies presented here it is important to identify muons coming from W decays. To this end, additional selection criteria are applied to distinguish these muons, which will be referred to as *signal* muons, from the muons coming from other sources such as b decays and fakes. The latter will be referred to as *background* muons, even though they arise in signal events as well as background events. The desired discrimination between *signal* and *background* muons is achieved by constructing a discriminator that is based upon Probability Density Functions (PDF) for the following observables associated with muon candidates:

- Transverse momentum, p_t
- Track isolation, $IsoTk$
- Calorimeter isolation, $IsoCalo$

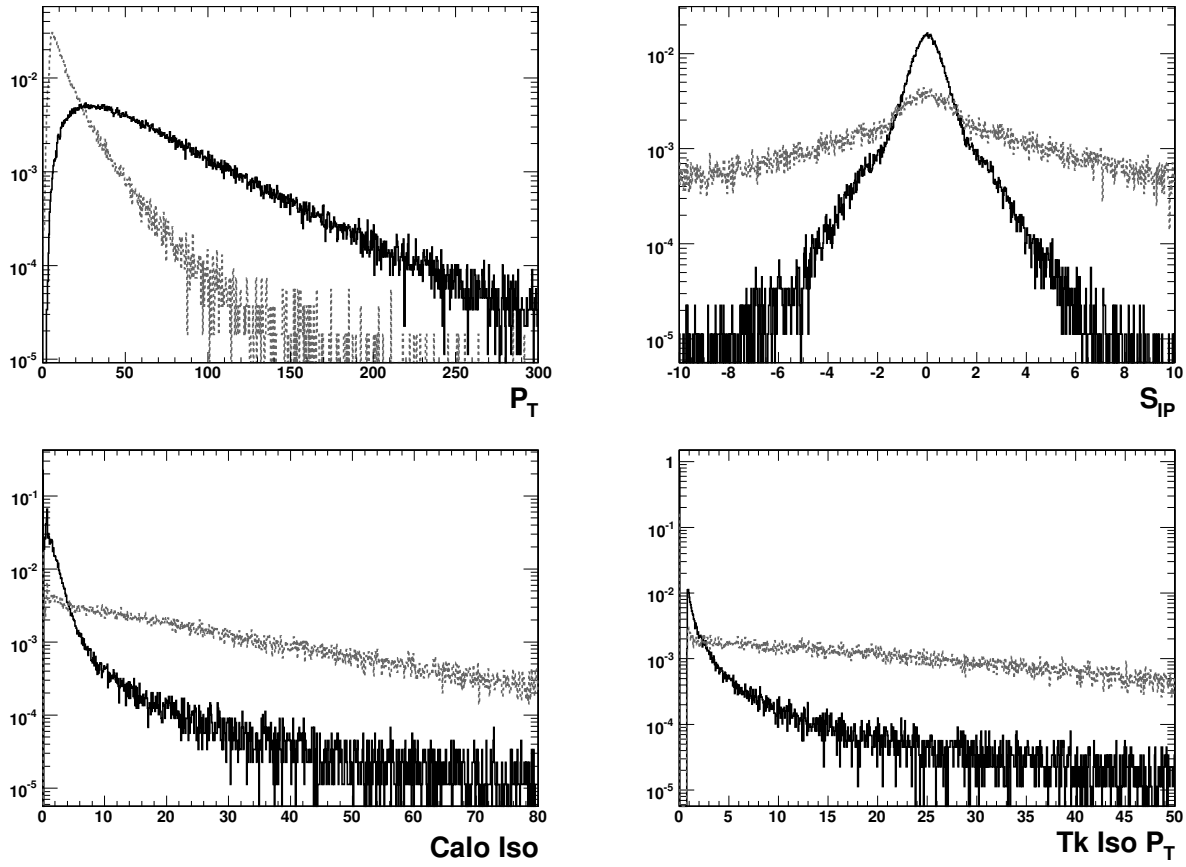


Figure 1: Probability density function used in the construction of the muon likelihood: the black line refers to *signal* muons while the grey line refers to *background* muons, as defined in the text.

- Significance of track impact parameter, $S_{ip} = d/\sigma_d$

The PDF's associated with these variables (Figure 1) for *signal* and *background* muons are obtained by matching to generated muons. A reconstructed muon is identified as originating from a W boson decay when the separation ($\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$) from the generated muon in azimuth ϕ and pseudorapidity η space is less than 0.01. Each of the selected variables provide good discriminating power between *signal* and *background* muons. The PDF's are combined into the following likelihood ratio:

$$L = \Pi_i \frac{P_i^{sig}(x_i)}{P_i^{sig}(x_i) + P_i^{bkg}(x_i)} \tag{1}$$

where P_i^{sig} and P_i^{bkg} are the PDF's of an observable x_i for signal and background muons, respectively. The likelihood distribution as well as the performance for *signal* and *background* muon discrimination are displayed in Figure 2. At a signal muon efficiency of 90%, only 1% of background muons are selected. The PDF's are constructed using a sample of $t\bar{t}H$ events with m_H =120 GeV/c^2 in which one and only one of the W bosons decays to a muon and neutrino, while the other one decays hadronically.

In Figure 3 the $t\bar{t}H$ muon selection efficiency versus the QCD ($\hat{p}_t$ >170 GeV$/c$) efficiency is shown. The performance of the likelihood selection is compared to the performance of the HLT single muon stream selection. If the likelihood selection is used in place of the HLT, Figure 3 shows that at the same QCD ($\hat{p}_t > 170$ GeV$/c$) efficiency the likelihood gives 74% of signal efficiency as compared to 63% for the HLT single muon stream. Alternatively, if the likelihood selection is used after the HLT, dramatic improvement in QCD ($\hat{p}_t$ >170GeV$/c$) rejection is possible at little or no loss in signal efficiency. For example, a small drop in signal efficiency from 63% to 60% lowers the QCD efficiency by more than a factor of 3 (i.e. from 0.07% to 0.02%).

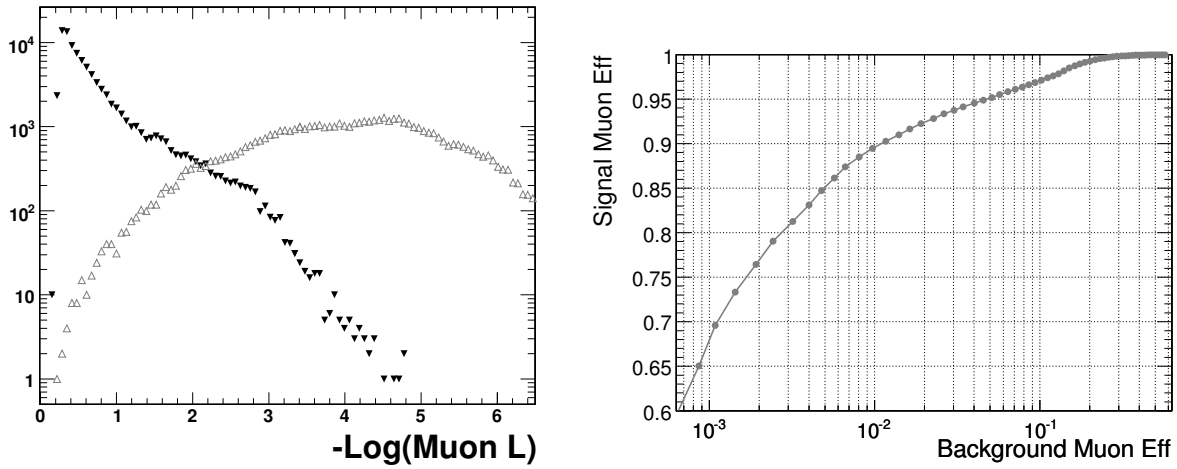


Figure 2: On the left: Distribution of $-Log$ of muon likelihoods. The black line refers to *signal* muons and the grey line to *background* muons, as defined in the text. On the right: Performance of the muon likelihood discriminator for the $t\bar{t}H$ channel.

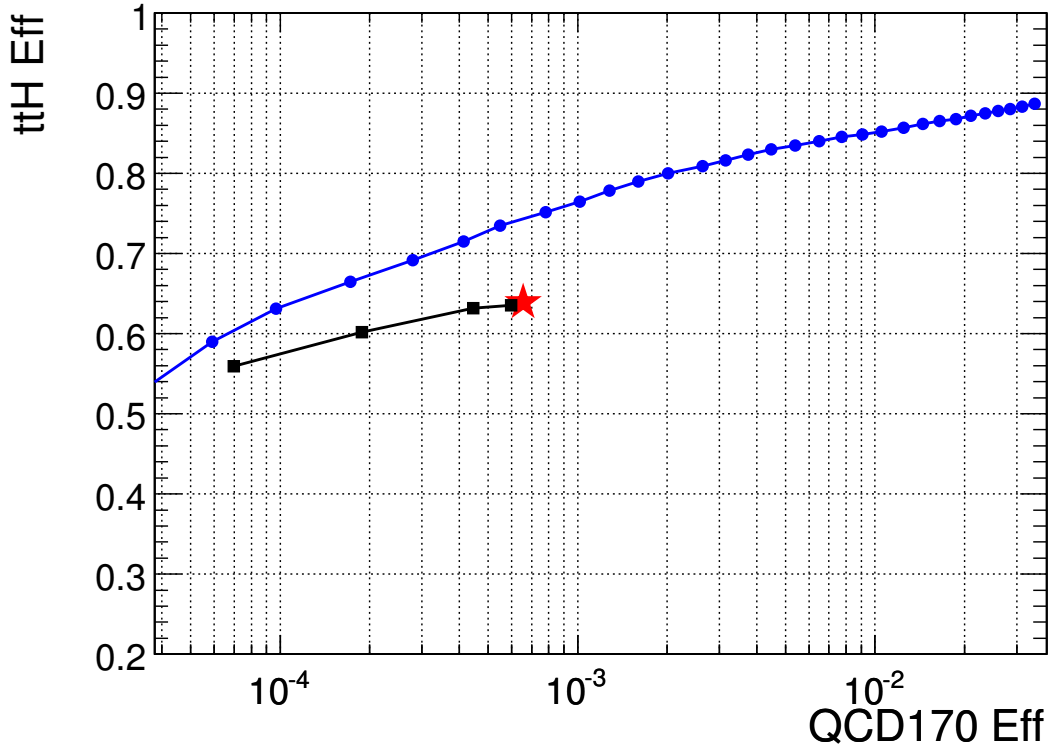


Figure 3: $t\bar{t}H$ signal efficiency versus QCD ($\hat{p}_t > 170$ GeV$/c$) efficiency. Circles: likelihood performance without HLT selection; Star: HLT selection; Squares: Likelihood performance after the HLT selection.

4.2 Electron Reconstruction

An electron candidate is a directed cluster of energy reconstructed in the electromagnetic calorimeter where the direction is calculated using the cluster centroid position and the nominal position of the primary interaction vertex. This information is then matched with the information from the Tracking system. A full description of the electron reconstruction in CMS can be found in Ref. [9]. Electrons coming from W boson decays are typically characterized by isolated high transverse energy clusters. These electrons are thus efficiently identified by means of an isolation requirement applied to the electron candidate with respect to other reconstructed tracks in the event.

Isolation is defined by means of two variables. The first is the transverse momentum (p_t) sum of tracks inside an *isolation cone* of radius $\Delta R = 0.3$ in the η-ϕ plane about the candidate electron's direction, as defined by the calorimeter. A *veto cone* of radius $\Delta R = 0.015$ around the electron direction is defined in order to exclude the electron p_t from this sum. The second variable is the distance $\Delta R = \sqrt{((\phi_{ele} - \phi_{track})^2 + (\eta_{ele} - \eta_{track})^2)}$ between the electron candidate and the closest track in the η-ϕ plane outside the *veto cone*.

To further optimize the electron selection, three additional variables are considered:

- the transverse momentum of the electron candidate, p_t
- the ratio between the cluster energy and the track momentum, E/p
- the ratio between the hadronic and electromagnetic energies of the cluster, H/E

All of these five discriminating variables are used in order to construct a likelihood hypothesis for an isolated electron.

The relevant PDF's are constructed in the form of reference histograms and combined into a likelihood in a manner completely analogous to that which was described earlier for muons in Equation 1. Also as in the muon case, the reference histograms for *signal* and *background* electrons are again constructed by matching to generator level electrons in Higgs signal events. Unit-normalized distributions for *signal* and *background* electrons are compared in Figure 4. Proceeding from Figure 4a to Figure 4e, distributions of p_t, E/P, H/E, $p_t isolation$ and ΔR are displayed, respectively. The $-Log(L_e)$ curves for *signal* (continuous line) and *background* electrons (dashed line) are superimposed in Figure 5.

An appropriate choice of likelihood cut value has been studied by comparing *signal* versus *background* electron efficiencies as shown in Figure 6.

For a $-Log(L_e)$ cut value of 1.27, *signal* electrons are selected with an efficiency of 84% and *background* electrons with an efficiency of 1.5%. This value was chosen as the working point for the analyses described in subsequent sections of this note.

To evaluate the efficiency of the likelihood cut with regard to background rejection, a study was performed with a sample of $t\bar{t}jj$ events in which there were no isolated electrons from W decays. With a likelihood cut of 1.27, 6% of these events were accepted.

In Figure 7 the signal efficiency is plotted versus QCD background electron efficiency for various choices of likelihood cut values. The star identifies the efficiency levels obtained by applying only the HLT trigger selection. Applying the $-Log(L_e)$ working point value of 1.27, the efficiency for accepting QCD background electrons is 1%. Furthermore, as was true for the muon sample, the likelihood approach can be used to augment the HLT selection efficiency. Maintaining roughly constant signal efficiency, the likelihood cut in combinaton with the HLT trigger yields an order of magnitude reduction in the QCD background selection efficiency. After applying the likelihood cut, less than 1% of these events have more than one electron. In the few cases where more than one is found, the highest likelihood electron is selected.

4.3 Jet and Missing Transverse Energy Reconstruction for Channels with Leptons in the Final State

The jet reconstruction for channels with leptons in the final state is described in this section. A separate optimization of jet reconstruction was carried out for the all-hadron channel, as described in the next section.

Jets are reconstructed using the iterative cone algorithm. A cone with $\Delta R = 0.5$ is used when at least one W boson decays into leptons, while a smaller cone size was found to be more suitable for the more dense jet environments

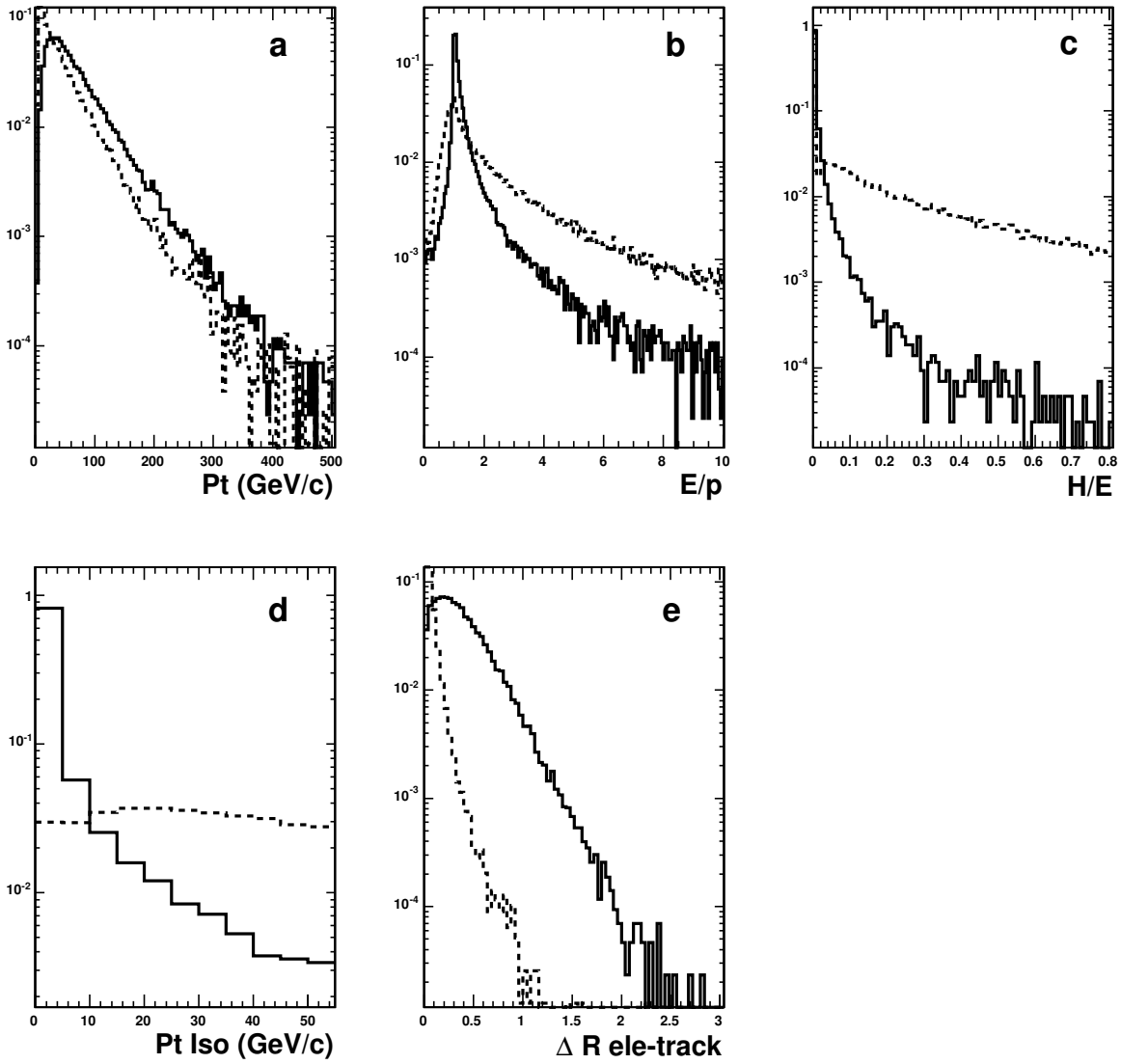


Figure 4: Unit-normalized distributions for the five electron variables used to construct the electron likelihood, as discussed in the text.

associated with the all-hadron channel, (see below). A calorimetric-tower energy threshold of 0.8 GeV and a transverse-energy threshold of 0.5 GeV are used. Calorimetric towers that exceed 1 GeV are considered as jet seeds. For the leptonic channels, the jet energy is calibrated using MC calibrations [5] provided by the JetMET group for the corresponding set of reconstruction parameters.

The single lepton analyses, as described in more detail below, make use of an event likelihood to help select and properly reconstruct events in order to reconstruct the $H \rightarrow b\bar{b}$ decay and subsequently the invariant mass of the Higgs. This is facilitated, in part, by making use of the various invariant mass constraints associated with top quark decays. The corresponding likelihoods thus rely upon the resolutions that are obtained for the invariant masses of the hadronically decaying W boson and the two top quarks. The invariant mass distributions for the hadronically decaying W bosons, the hadronically decaying top quarks and the leptonically decaying top quarks are reconstructed using jet matching to generator-level parton information. The corresponding distributions are shown in Figure 8. A reconstructed jet is considered as matched to the corresponding parton if their separation, ΔR_{j-p}, is less than 0.3.

The missing transverse energy of the event E_t^{miss} is computed as [5]:

$$E_t^{miss} = \sum_i E_t^{tower} - (\sum_j E_t^{RawJet} - \sum_k E_t^{CaliJet}) + \sum_m E_t^{Muon} \quad (2)$$

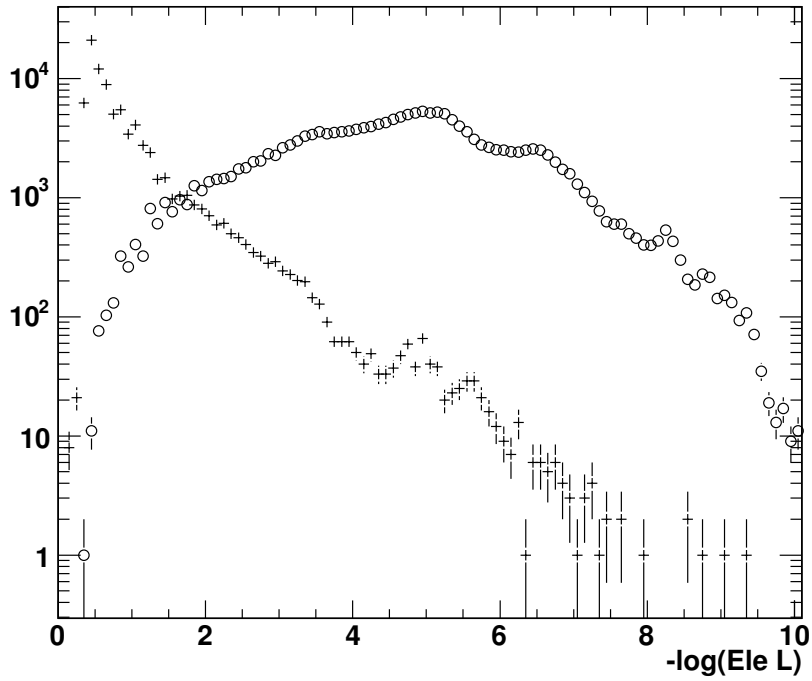


Figure 5: $-Log$ of the electron likelihood distributions for *signal* (crosses) and *background* (circles) electrons.

where the sum with index i runs over the calorimetric towers, that with j runs over raw jets, k runs over calibrated jets, and m runs over the reconstructed muons of the event. Equation 2 thus takes into account the corrections due to jet calibration and the contributions of muons that are not measured in the calorimeter system. The distribution of the reconstructed missing transverse energy of semi-leptonic $t\bar{t}H$ events and the missing transverse energy resolution are shown in Figure 9. The distribution of the reconstructed missing transverse energy of fully leptonic $t\bar{t}H$ events is shown in Figure 20.

4.4 Jet Reconstruction for the All Hadron Channel

The choice of the jet reconstruction algorithm is an important step in the event selection optimization for the all hadron $t\bar{t}H$ channel, where 8 jets in the final state are expected. For this reason, an optimization is obtained by means of a simple "proto" analysis.

In the proto analysis, the iterative cone algorithm is employed, and cone sizes ranging from 0.35 to 0.50 are considered. A simple analyis of events is then performed to extract a significance and signal to background ratio for each of various cone sizes in order to obtain a relative comparison of their Higgs discovery potential. Jets are reconstructed with the standard CMS ORCA (version 8.13.3) software and the following parameters, (whose labels appear in parentheses):

- Cone Size $\Delta R = 0.35, 0.40, 0.45, 0.50$ (ConeCut)
- Cone Seed 0 GeV (ConeSeedEtCut): no requirement on minimum E_T, any tower can seed a jet
- Energy Recombination Scheme
- Tower threshold: $E_t > 0.5$ GeV and $E > 0.8$ GeV to suppress instrumental noise contributions

A dedicated $t\bar{t}H$ calibration [11] is applied to help recover the parton's original transverse energy. Reconstructed jets with a b-tagging discriminator value higher than 0.4 are calibrated using b-jet calibration functions.

The proto analysis is applied to the signal and the 3 most dangerous background samples ($t\bar{t}2j$, $t\bar{t}b\bar{b}$ and qcd170). The top quarks are reconstructed using a χ^2 method applied to the 8 most energetic jets in the tracker acceptance ($|\eta| < 2.7$). For each possible combination of jet-parton assignments for the 8 jets, the following invariant mass

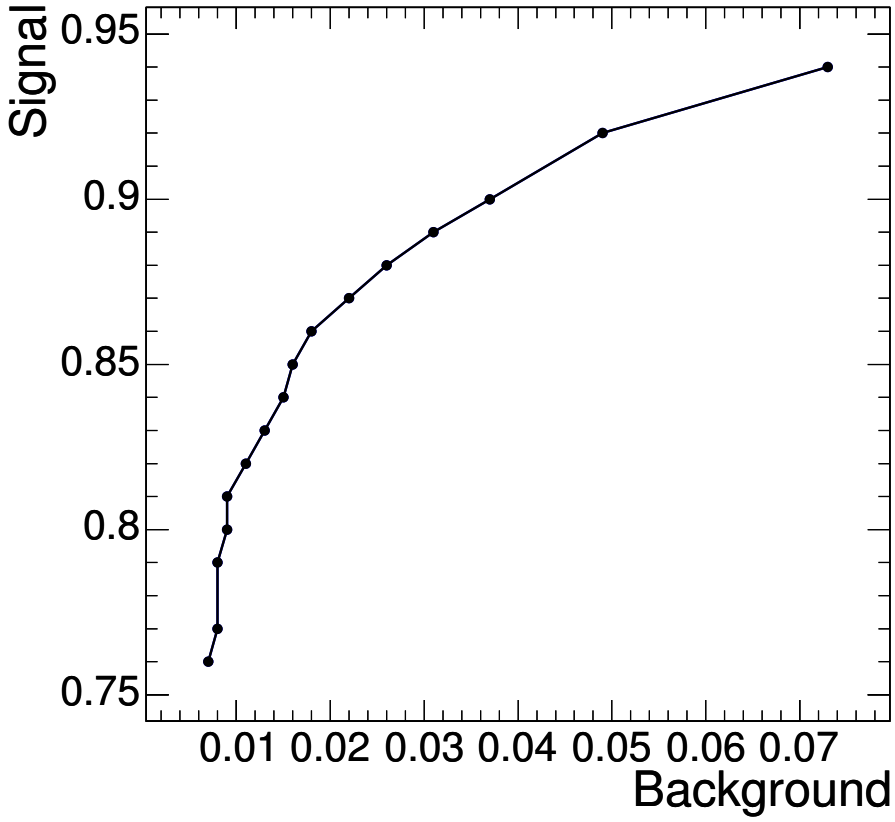


Figure 6: *Signal* versus *background* electron efficiencies for likelihood values ranging from 0.006 (the upper point) with a step size of 0.006, (i.e. approximately in the range $1.0 < -Log(L_e) < 2.0$).

χ^2 is computed:

$$\chi^2_{mass} = \left(\frac{m_{W^+} - m_{jj}}{\sigma(m_W)}\right)^2 + \left(\frac{m_{W^-} - m_{jj}}{\sigma(m_W)}\right)^2 + \left(\frac{m_t - m_{jjj}}{\sigma(m_t)}\right)^2 + \left(\frac{m_{\bar{t}} - m_{jjj}}{\sigma(m_t)}\right)^2. \quad (3)$$

The expected mass values and their σ's for top quarks and W bosons were obtained by parton-jet matching with $\Delta R_{j-p} < 0.3$. The values differ for the different cone sizes as reported in [11]. The following conditions must be satisfied for any combination of jet-parton assignments to become a candidate for the $\bar{\text{t}}$t H hypothesis:

- the mass χ^2 for two W bosons and the two t quarks are within 3 sigma of the expected value
- $E_t > 25$ GeV for the 8 selected jets
- The jets assigned to the Higgs and to the b daughters of the top quarks must be b-tagged

The jet-pairing combination with the lowest χ^2_{mass} that satisfies these requirements is chosen and the event is selected.

Figure 10 plots significance with respect to the S/N ratio for a range of b-tag discriminator values for each of the several cone sizes indicated. Lower discriminator values yield higher significance but only at the cost of low S/N while, on the contrary, higher discriminator values give lower significance but higher S/N. A good compromise is in the middle range of each of the curves where neither S/N nor significance are unreasonably low. With this in mind, the best choice for the jet cone is seen to be $\Delta R = 0.40$.

4.5 b-Tagging

The identification of jets from b-quarks is done with the *Combined Secondary Vertex* algorithm. This algorithm exploits secondary vertex and track properties to calculate a discriminator value wich separates b-jets from non b-jets.

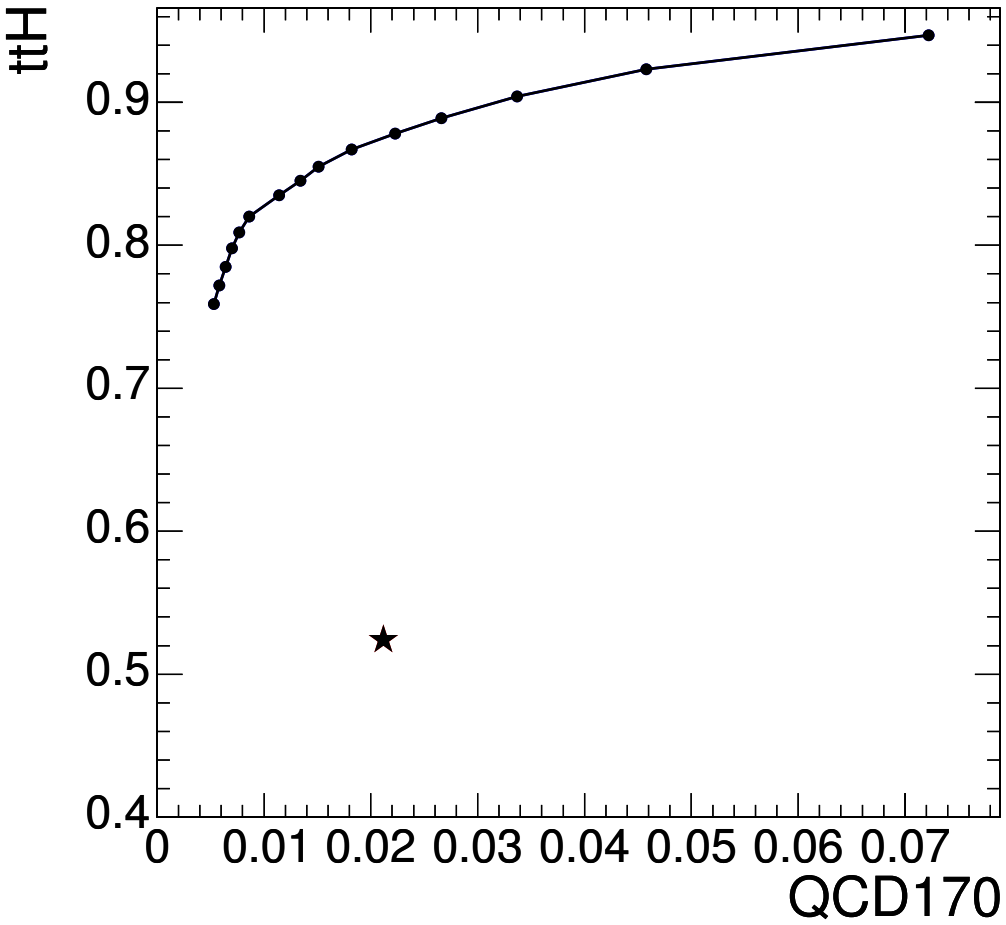


Figure 7: Signal efficiency versus QCD efficiency for likelihood cut values ranging from 0.006 (the upper point) with a step of 0.006 (i.e. approximately in the range $1.0 < -Log(L_e) < 2.0$).

A detailed description is published in Ref. [10] which also presents results of detailed studies of the performance of the b-tagging algorithm as applied to Monte Carlo $t\bar{t}$ and QCD samples.

In the $t\bar{t}H$ analyses, a fixed cut value for the b-tagging discriminator is applied, and four jets are required to pass this cut in the semileptonic and all-hadron channels, while only 3 jets are required to be tagged in the dilepton analysis. The misidentification rate of charm and light flavour jets as a function of the b-tagging efficiency is shown in Fig. 11 for the $t\bar{t}H$ and the $t\bar{t}jj$ samples, respectively. It is seen that the efficiencies are similar in these samples.

This fixed-cut b-tagging approach is quite reasonable but not necessarily optimal. Some potential improvements are possible and have been investigated, but these were not used in the current analysis.

Due to several effects, e.g. a poor impact parameter resolution due to multiple scattering for low p_t tracks and higher track multiplicity at the primary vertex for high p_t jets, the b-tagging performance peaks for a transverse momentum of about 80 GeV$/c$ and falls relatively quickly as the momentum scale falls [10]. Since the $t\bar{t}H$ channel provides a large number of lower p_t jets, better performance can be achieved by applying a dynamic (i.e. p_t and η dependent) discriminator cut. In this case, the discriminator cut necessary to obtain a given efficiency is calculated for each jet separately. It is thus not possible to achieve a fixed efficiency for b-identification and light flavour misidentification at the same time. In order to obtain a reliable rejection of the $t\bar{t}jj$ background, a fixed light flavour misidentification rate of, for example, 1% can be chosen. The variation of the cut versus p_t required for this rate to be maintained is shown in Fig. 12. It is worth noting that while there is some potential to improve b tagging performance by such means, there is also an increase in the complexity of the tagging machinery. From a practical standpoint, such a tagger could be more difficult to understand and as such may incur larger systematic uncertainties. This is indicative of many aspects of this analysis. Namely, that marginal improvements often come at a cost. In any case, this potential improvement was not used in the analyses reported in this note.

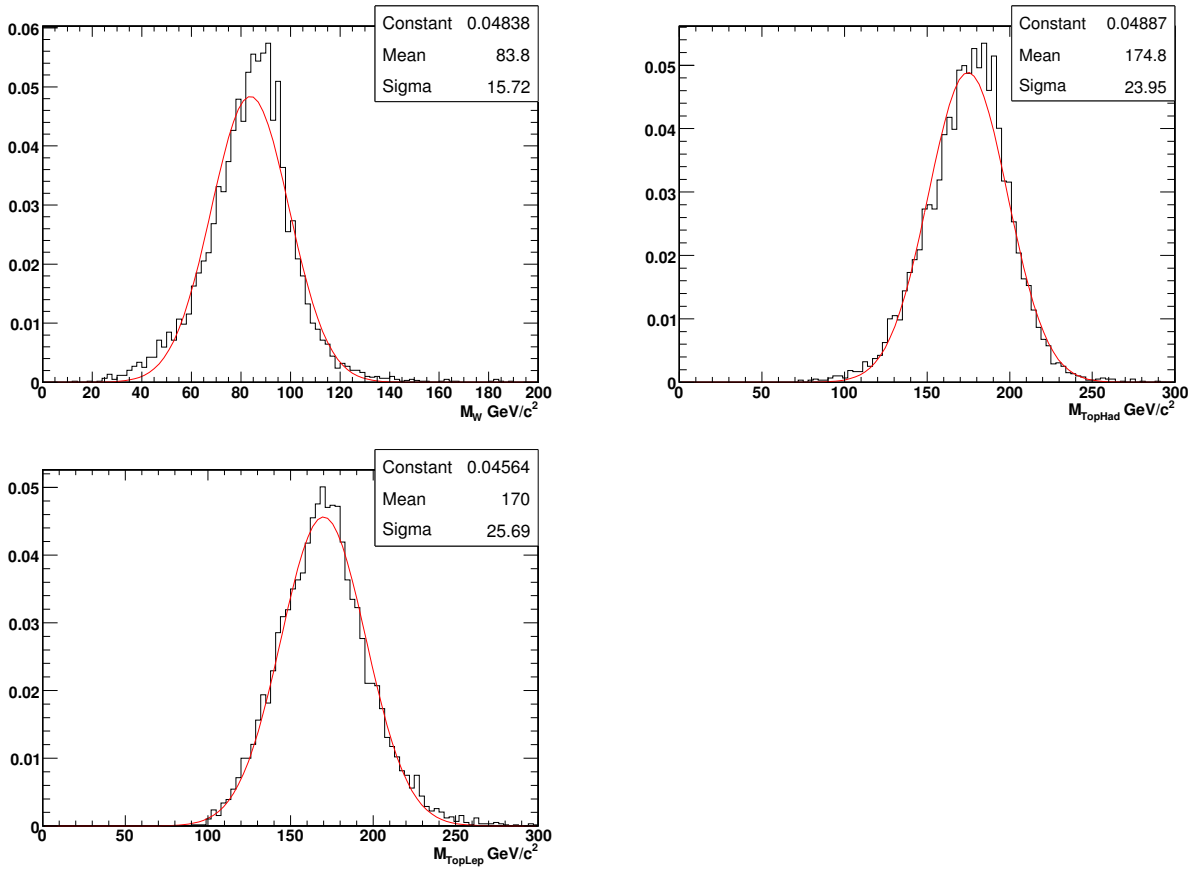


Figure 8: Invariant masses of the hadronically decaying W boson and the hadronically and leptonically decaying Top quarks using jet-parton matching with $\Delta R_{j-p} < 0.3$.

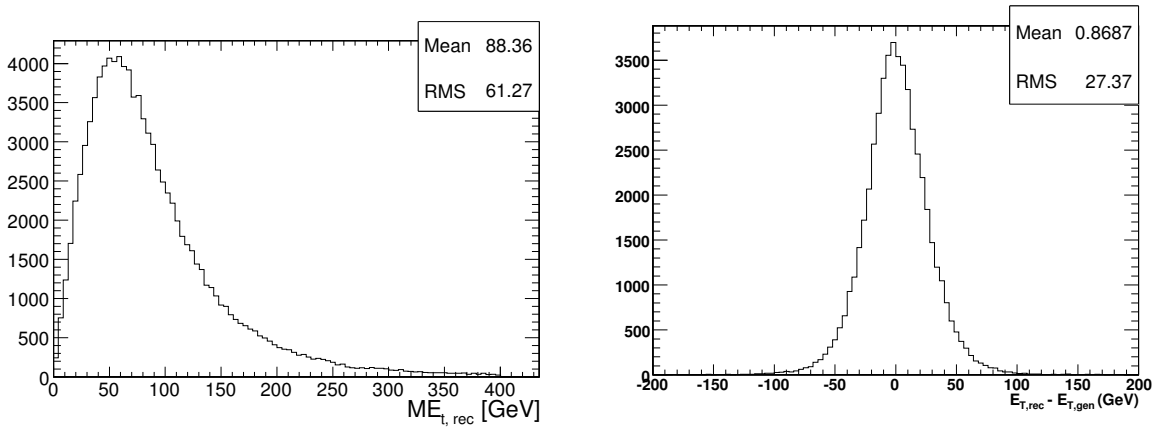


Figure 9: Left: Distribution of the reconstructed missing transverse energy. Right: Resolution of the reconstructed missing transverse energy in $t\bar{t}H$ events with semi-leptonic W decays.

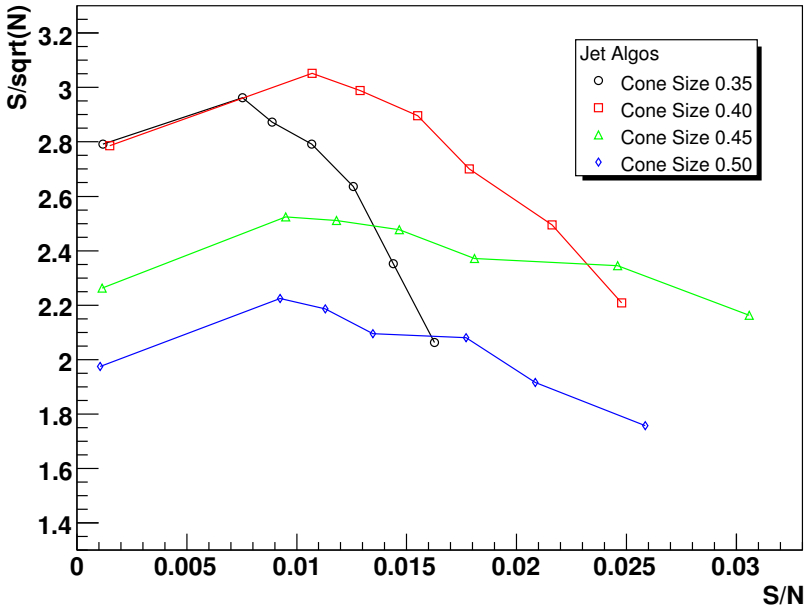


Figure 10: Change in significance and S/N resulting from variations in the bTag Discriminator Variable for the various cone sizes indicated in the legend.

The *Combined Secondary Vertex* algorithm does not include any information about leptons in b-jets. A separate algorithm, the *Soft Lepton* b-tagging algorithm [12], exploits the presence of leptons in b-jets that arise in semi-leptonic decays of B hadrons. This is achieved by means of a discriminator that is constructed for jets with non-isolated leptons. The two algorithms have some correlation but it is possible to obtain an increase in efficiency by combining them via a likelihood ratio. This has various potential benefits including a relative improvement of about 10% for the rejection of light flavour jets..

None of the aforementioned improvements in tagging were used in the analyses presented in this note. They represent an overall improvement on the order of a few percent and will be presented in a future publication.

5 Event Selection

In this section of the note the event selection for the different channels under consideration is described. Clearly there are significant differences in the selection criteria employed and each channel can arrive at a set of cuts that optimizes the final significance of the signal over background without considering what the other channels are doing. In practice however, this approach is not ideal because it is of substantial interest to be able to combine the results of all the $t\bar{t}H$ search channels, particularly in the time prior to a Higgs discovery. This is most easily facilitated by making sure that the different channels use distinct event samples. This, in turn, is most easily facilitated by coordinating how high p_t electrons and muons from W's (previously refered to as *signal* leptons) are either selected or vetoed by the different analyses.

For the analyses reported here, the different data samples were made disjoint by the following set of lepton -Log(L) selection and/or veto values:

- Single electron channel: Muon veto 1.4, Second electron veto 1.2
- Single muon channel: Electron veto 1.2, Second muon veto 1.4
- Di-lepton channel: First or second muon selection 1.4, First or second Electron selection 1.2
- All-hadron channel: Electron veto 1.2, Muon veto 1.4

These cuts were chosen largely because of where the single lepton channels optimum values appear. As it turns out, they are also not unreasonable choices for the di-lepton channel and the all-hadron channel. It is therefore

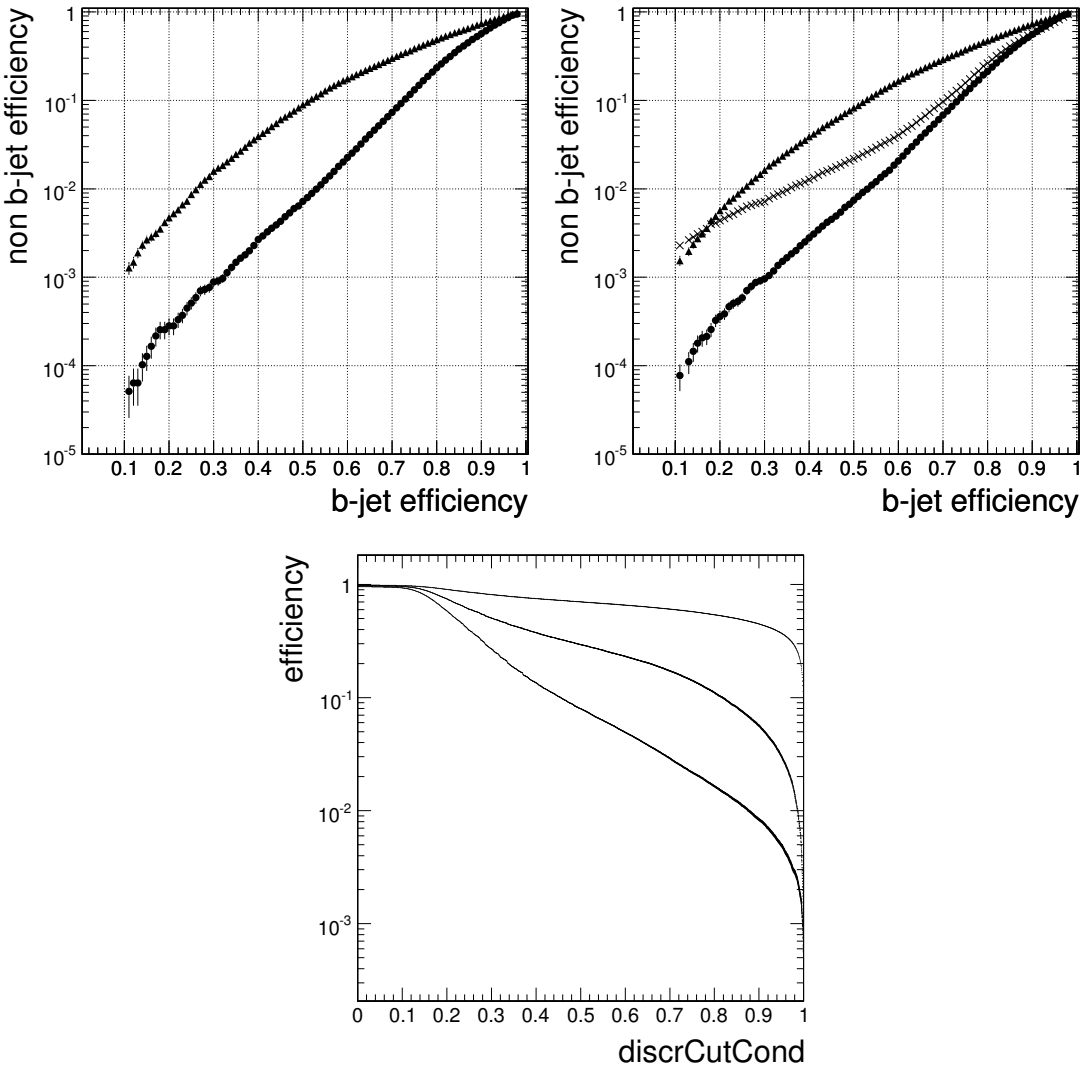


Figure 11: On the top left: Non-b jet mistagging efficiency versus b-jet tagging efficiency for c-jets (triangles), and uds-jets (stars) for the $t\bar{t}H$ sample with $m_H = 120$ GeV$/c^2$ and jets with a minimum transverse momentum of 20 GeV$/c$. For this plot the "physics definition" of the original jet flavour has been used. In this definition, jets originating from the splitting of gluons into heavy flavour pairs are labeled as gluons leading to a significantly reduced rejection power for gluon jets. There are no original gluon jets in the $t\bar{t}H$ sample. On the top right: The corresponding plot for the $t\bar{t}jj$ sample, where gluon jets are represented by crosses. On the bottom: Tagging efficiency for b-jets (topmost curve), and Mis-tagging efficiencies for c-jets (middle) and uds-jets (bottom) as a function of the discriminator cut for the $t\bar{t}jj$ sample.

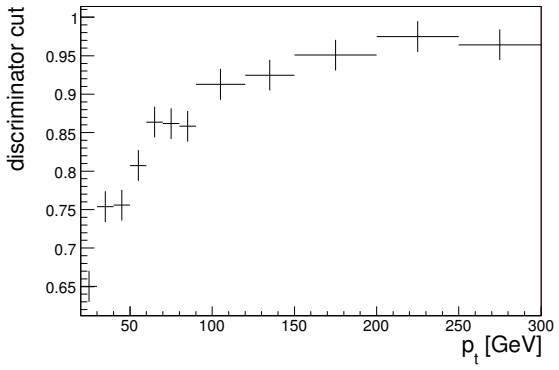


Figure 12: The b-tagging discriminator cut values required to maintain a fixed light flavour misidentification rate of 1%, as a function of jet p_t.

probable that a slightly better overall optimization could be found by varying the cuts and looking at the combined significance coming from all channels. This was not done for the work presented here.

5.1 Semi-leptonic Channel: $t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'\mu\nu_\mu$ and $b\bar{b}b\bar{b}qq'e\nu_e$

The strategy for selecting $t\bar{t}H$ events with one isolated muon or electron in the final state can be summarized in the following three steps:

- preselection
- choice of jet pairing
- selection

The preselection requires the HLT stream for a single muon or a single electron, one isolated lepton using the likelihood information as described in section 4.1 and 4.2, and 6 or 7 jets in the pseudorapidity region $|\eta| < 3.0$ with a calibrated transverse energy larger than 20 GeV. In order to recover some efficiency, jets with 10 GeV $< E_t <$ 20 GeV are also accepted if they have at least two associated tracks pointing to the signal primary vertex [1)] within a distance along the z (beam) axis of $|z_{PV} - z_{track}| < 1$ mm. The latter condition is required to reject low transverse energy fake jets, (i.e. jets that are not associated with any of the signature partons in the signal event). For the single electron channel, the misidentification of the jet with the isolated electron has been excluded by imposing a veto on the jet if the electron lies inside a jet cone radius of 0.1. The effect of this requirement is highlighted in Figure 13 where the ratio between the electromagnetic energy to the total energy is plotted for the selected jets.

At least 4 jets are required to be tagged as b-jets with a minimal discriminator value corresponding to about 70% of b-efficiency. To avoid contamination from background with both W's decaying leptonically, a double muon, double electron and muon-electron veto is applied. The second highest muon likelihood value for the exclusive semi-leptonic (μ) sample, and the inclusive $t\bar{t}2j$ background sample, after application of the cuts described above, is shown on the left of Figure 14. The electron likelihood for the same sample is shown on the right of the same Figure. To decrease the contamination from the di-lepton channel, events with the second lowest $-Log(L_\mu) < 1.4$ and events with $-Log(L_e) < 1.2$ are rejected from the analysis. In the case of the semi-leptonic electron channel the previous cuts are applied respectively to the first muon likelihood candidate and to the second electron likelihood candidate. The application of the vetoes results in a lowering of the signal efficiency by only about 2%, while the total background rejection is increased by 13%.

In order to perform a complete reconstruction of the event, the longitudinal component of the neutrino momentum has to be computed from four-momentum conservation for the W boson: $m_W^2 = (E^\mu + E^\nu)^2 - (\vec{p}^\mu + \vec{p}^\nu)^2$. This equation gives 2 real solutions for p_z^ν in 66% of the cases. In the remaining 34%, the neutrino is taken to be collinear with the lepton: $p_z^\nu = p_z^l$. The resolution of p_z^ν before and after the collinear approximation is shown in Figure 15. A small degradation in the longitudinal resolution is observed but the reconstruction efficiency of the leptonic W boson decay is increased to 100%.

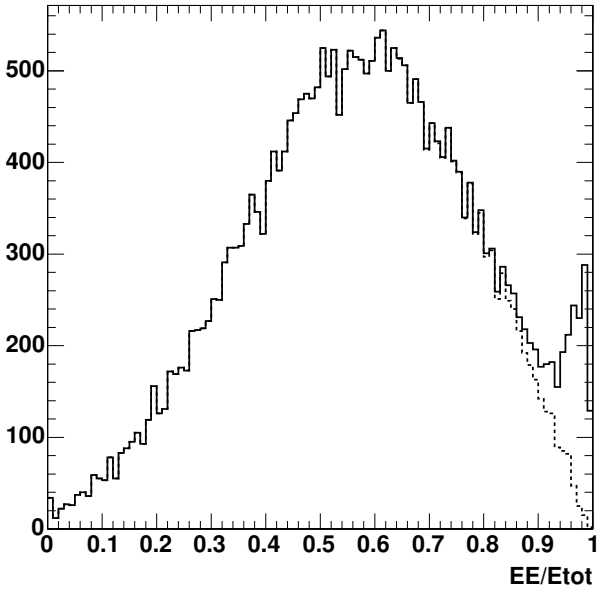


Figure 13: Ratio between electromagnetic and total energy for selected jets: The dashed line refers to jets selected after the electron veto.

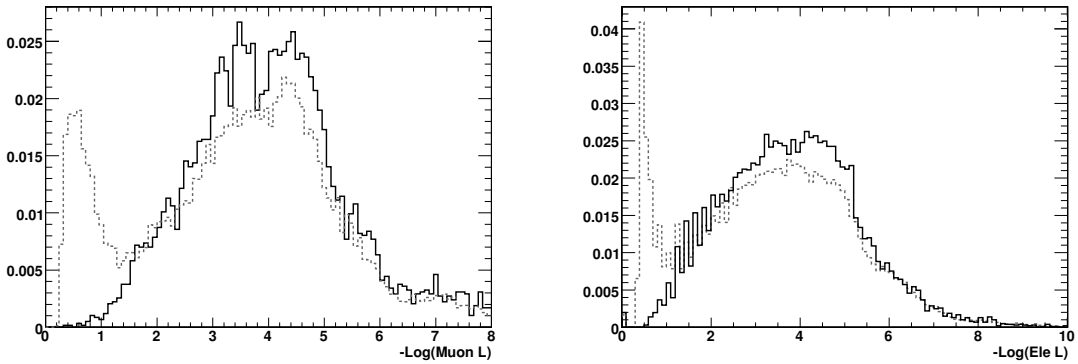


Figure 14: Avoiding the appearance of a second lepton in the single μ analysis. On the left: Muon likelihood distributions. On the right: Electron likelihood distributions for a potential second lepton candidate in these events. In both cases solid lines refer to the pure semi-leptonic (μ) sample while the dashed lines refer to the inclusive $t\bar{t}2j$ background. The peaks near to zero clearly highlight the presence of a second muon (left) or an electron (right) in the background samples which allow inclusive decays of the Ws. The distributions are normalized to Unity.

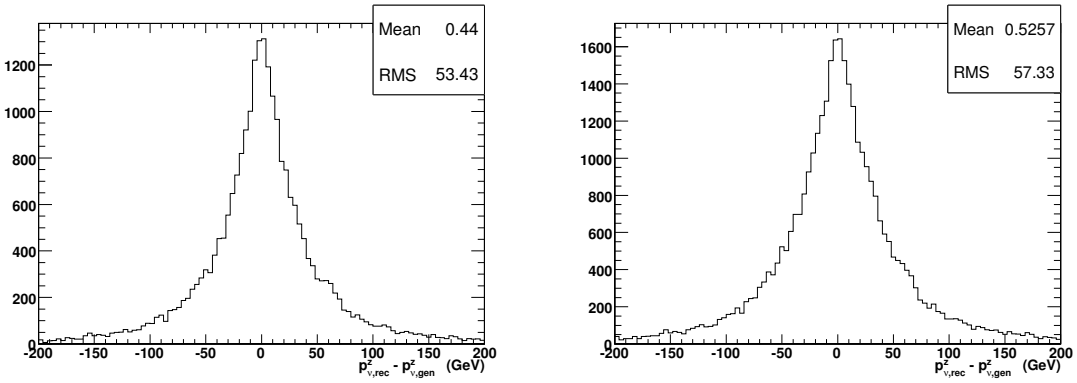


Figure 15: Longitudinal resolution of the neutrino. On the left: Only those cases in which there are real solutions obtained for the W boson's four-momentum constraint. On the right: Real plus collinear solutions as described in the text.

The strategy of the complete reconstruction of semi-leptonic $t\bar{t}H$ events is similar to that used in Reference [13]. In order to choose the jet combination that does the best job of reconstructing the two top quarks, a likelihood (L_{Event}) is defined by using masses, b-tagging and kinematic information from the whole event:

$$L_{Event} = L_{Mass} \times L_{bTag} \times L_{Kine}. \tag{4}$$

To increase the speed of the algorithm while also enhancing the rejection of events that are badly reconstructed, only those combinations with mass resonances within three sigma of expectation are considered in the jet assignments. The mean values and widths of the invariant masses used in this procedure are taken from the values obtained with jet-parton matching, as described in Section 4.3.

The mass information considered in the likelihood L_{Mass} is the probability of the kinematic fit with invariant mass constraints (top quarks and hadronically decaying W) that is described in Reference [7]. In $t\bar{t}H$ events, a mass constraint on the hadronically decaying W improves the probability to correctly identify the b-jets coming from the top quark decay and hence also improve the chance of correctly finding the b-jets originating from the Higgs boson decay. However, the relative improvement in the jet assignment is found to be only around 5% with respect to what is obtained by using simple Gaussian distributions for mass resonances and has little effect on the overall result of the analysis.

The b-tagging function L_{bTag} is defined as the product of the b-tag discriminators: $L_{bTag} = D_{Top_{Had}} \times D_{Top_{Lep}} \times D_{H_1} \times D_{H_1} \times (1-D_{W1}) \times (1-D_{W2})$; where Top_{Had} and Top_{Lep} are expected to be the two b jets from the hadronic and leptonic top, respectively, while H_1 and H_2 are expected to be the two b jets coming from Higgs and W_1 and W_2 are of course the two jets from the hadronically decaying W boson. As expected, high values of this variable correspond to combinations where b-jets are assigned to the top quarks and Higgs, and light quark jets are assigned to the W.

The kinematic function takes into account the observation that the b-jets coming from top quarks tend to be slightly more energetic than b-jets coming from the Higgs boson. It is defined as:

$$L_{Kine} = arctan\left(\frac{4(E_{Top_{Had}} + E_{Top_{Lep}} - E_{H_1} - E_{H_2})}{(E_{Top_{Had}} + E_{Top_{Lep}} + E_{b1H} + E_{b2H})}\right) \times \frac{1}{\pi} + \frac{1}{2}. \tag{5}$$

Among all possible combinations of jet-parton assignments, the one with the highest value of L_{Event} is chosen for use in the final reconstruction of the top quarks. Finally, the invariant mass of the two remaining jets with highest b-tagging discriminator values is used to reconstruct the Higgs mass.

After the jet assignment just described is complete, criteria are applied to further reject background. In particular, a stronger b-tag requirement is applied on the event variable L_{bSele} given by the combination of b-tag discriminators: $L_{bSele} = D_{Top_{Had}} \times D_{Top_{Lep}} \times D_{H_1} \times D_{H_2}$. The variable L_{bSele} is shown in Figure 16 for the semi-leptonic (μ) channel as well as for all backgrounds combined. The distributions are normalized to the number of expected events in 60 fb^{-1}. To optimize the background rejection, the signal significance as a function of the selection cut L_{bSele} has been studied. The result is shown in Figure 17. In this figure, the backgrounds that give 0 remaining events are not taken into account. The behavior of the signal significance versus different values of L_{bSele} is shown to be fairly

[1] The signal interaction is generally the one which allows the event to be triggered.

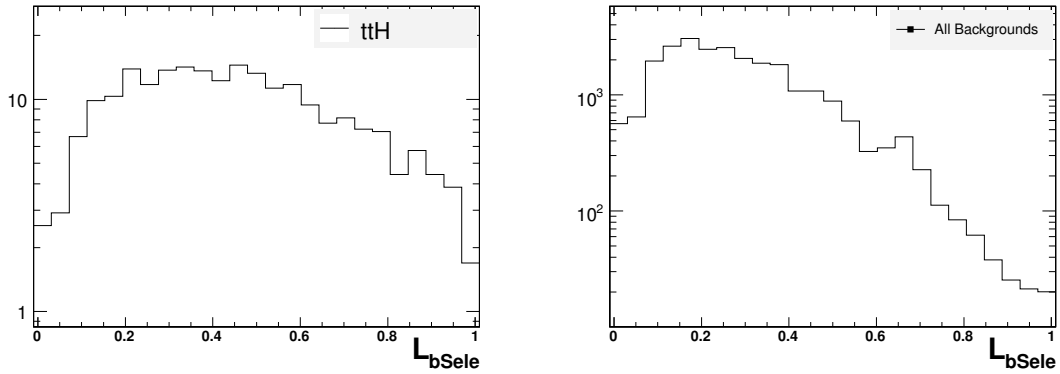


Figure 16: L_{bSele} selection cut for $t\bar{t}H$ with $W \to qq', W \to \mu\nu$ (on the left) and for background (on the right).

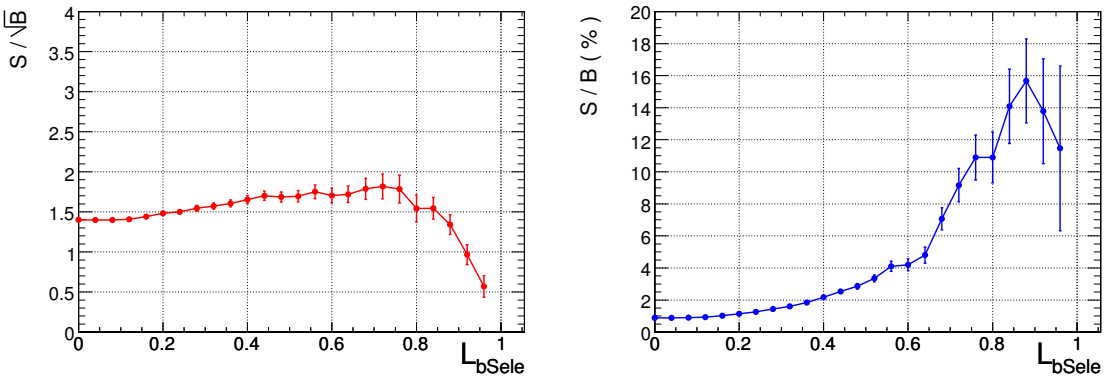


Figure 17: $t\bar{t}H$ ($W \to qq', W \to \mu\nu$): Signal Significance (left) and Signal to Background ratio (right) as function of the cut on L_{bSele}.

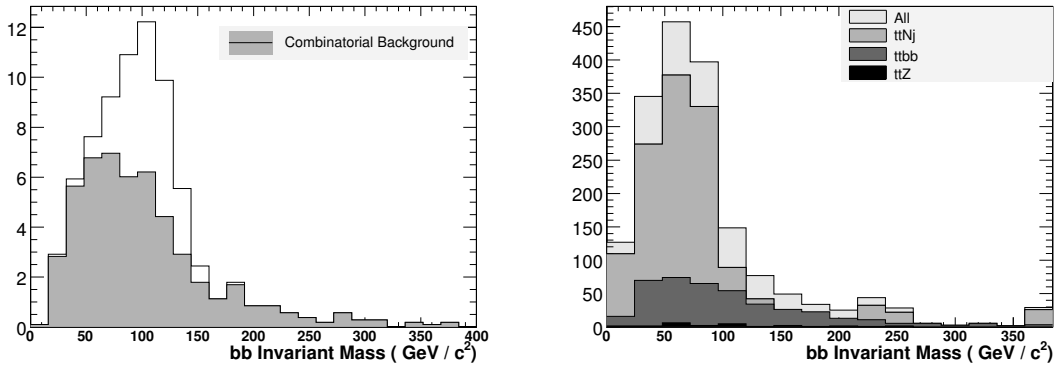


Figure 18: $t\bar{t}H$ ($W \to qq', W \to \mu\nu$). Left: Invariant $b\bar{b}$ mass for signal only (combinatorial background is shaded grey). Right: The sum of the reconstructed $m_{b\bar{b}}$ spectra for backgrounds with a value of $L_{bSele} > 0.55$. The distributions are normalized to an integrated luminosity of 60fb^{-1} .

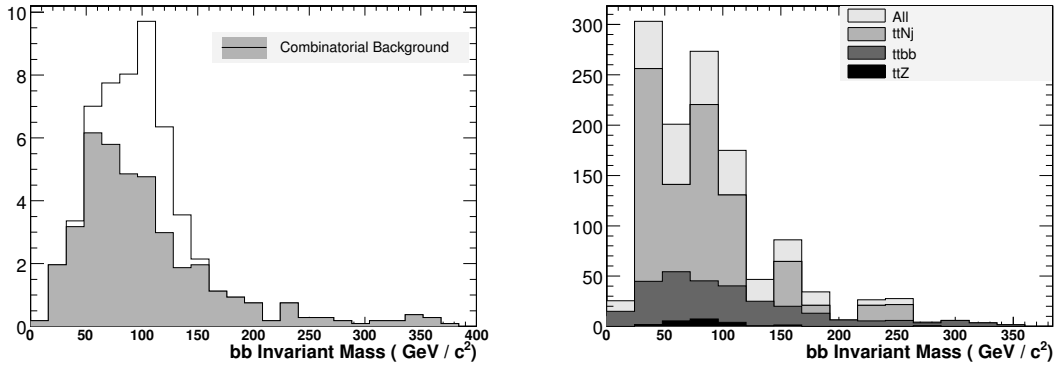


Figure 19: $t\bar{t}H$ ($W \to qq', W \to e\nu$). Left: Invariant $b\bar{b}$ mass for signal only (combinatorial background is shaded grey). Right: The sum of the reconstructed $m_{b\bar{b}}$ spectra for backgrounds with a value of $L_{bSele} > 0.55$. The distributions are normalized to an integrated luminosity of 60fb^{-1}. The large fluctuations in the $t\bar{t}$Nj background are due to a very limited number of remaining events that have to be scaled by large factors.

stable in the L_{bSele} cut range between 0.45 and 0.75. Cutting hard on L_{bSele} helps to increase S/B but is achieved at the cost of a significant reduction in signal statistics.

The distributions of reconstructed mass of the Higgs boson for final selected events are shown in Figures 18 and 19 for signal only (left) and for the combination of the different backgrounds (right) for the muon and electron channels, respectively. The grey histogram on the left distribution of Figure 18 represents the events where at least one jet has been wrongly assigned to the Higgs boson. The fraction of signal events where the two b-jets are correctly assigned to the Higgs boson (i.e. the pairing efficiency) is roughly 31% in the muon channel and about 29% for the electron channel.

5.1.1 Results

The preselection and selection efficiencies with the corresponding numbers of expected events and signal significances are reported in Table 6 for the channel with a muon or an electron in the final state. The number of expected events is computed for an integrated luminosity of 60 fb^{-1} in the Standard Model Higgs mass range from 115 to 130 GeV/c^2. In this table, the results are presented for two different working points. The tight cut for $L_{bSele} > 0.75$ gives better results for the final significance, but it suffers from a very small number of remaining background events, due to the limited Monte Carlo Statistics. For the determination of the shape of the invariant mass (Fig. 18) as well as for the systematic uncertainties (Section 6.1), a looser cut at $L_{bSele} > 0.55$ has been chosen.

Finally, it is interesting to note that in Table 6 the efficiencies for $t\bar{t}$2j to $t\bar{t}$3j to $t\bar{t}$4j are seen to be decreasing. As will be seen in the next subsections of this note, the opposite trends are seen for these samples in the di-lepton and all-hadron channels. The difference can be traced to the semi-leptonic selection requirement of 6 or 7 jets. This requirement very effectively removes all events with high jet multiplicity and so necessarily has ever increasing

impact going from $t\bar{t}2j$ to $t\bar{t}4j$. For the hadron channel this requirement is not applied and so the efficiency is seen to increase going from $t\bar{t}2j$ to $t\bar{t}4j$ because the presence of more jets in the events increases the probability of passing all of the remaining selection requirements. Note that the inclusion of this requirement in the all-hadron channel is not possible, since 8 jets are necessary to fully reconstruct the $t\bar{t}H$ event. For the di-lepton channel, 4 to 7 jets are required, since it is very likely that the signal would appear in samples with 2 leptons and only 4 jets (all b jets) prior to consideration of additional radiation in the events. A harder cut than a 7-jet maximum has been found to diminish the signal significance. However, since the 7-jet maximum corresponds to a softer cut for the channel than the same for the semi-leptonic channels (which have two additional jets from the hadronically decaying W), the efficiency going from $t\bar{t}2j$ to $t\bar{t}4j$ still increases.

Table 6: Selection efficiency for $L_{bSele} > 0.55$ ($\varepsilon_{\text{loose}}$) and for $L_{bSele} > 0.75$ ($\varepsilon_{\text{tight}}$), number of expected events and signal significance in 60 fb^{-1} for the muon and electron $t\bar{t}H$ channels. The signal datasets are labeled by the generated Higgs mass in GeV/c^2 (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the complete Higgs mass range.

muon channel					
	Analyzed Ev.	$\varepsilon_{\text{loose}}$ (%)	N^{ev}_{loose} 60fb^{-1}	$\varepsilon_{\text{tight}}$ (%)	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	27768	2.00 ± 0.08	96 ± 4	0.80 ± 0.05	38 ± 3
$t\bar{t}H$ (120)	41929	1.90 ± 0.07	75 ± 3	0.74 ± 0.04	29 ± 2
$t\bar{t}H$ (130)	19466	2.23 ± 0.11	55 ± 3	0.84 ± 0.07	21 ± 2
$t\bar{t}b\bar{b}$	372737	0.247 ± 0.008	419 ± 14	0.0877 ± 0.0048	148 ± 8
$t\bar{t}1j$	393000	0.0051 ± 0.0011	520 ± 120	0.00076 ± 0.00044	78 ± 45
$t\bar{t}2j$	568999	0.0105 ± 0.0014	633 ± 82	0.00070 ± 0.00035	42 ± 21
$t\bar{t}3j$	101000	0.0050 ± 0.0022	119 ± 53	0.	$< 27 (68\% C.L)$
$t\bar{t}4j$	86697	0.0035 ± 0.0020	126 ± 73	0.	$< 48 (68\% C.L.)$
$Zt\bar{t}$	50000	0.068 ± 0.012	23 ± 4	0.026 ± 0.007	9 ± 2
Total Background			1840		< 352
$S/\sqrt{B}$ (115)			2.2		2.0
S/B (115)			5.1%		10.8%
$S/\sqrt{B}$ (120)			1.8		1.6
S/B (120)			4.1%		8.2%
$S/\sqrt{B}$ (130)			1.3		1.1
S/B (130)			3.0%		6.0%
electron channel					
	Analyzed Ev.	$\varepsilon_{\text{loose}}$ (%)	N^{ev}_{loose} 60fb^{-1}	$\varepsilon_{\text{tight}}$(%)	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	27692	1.39 ± 0.07	66 ± 3	0.52 ± 0.04	25 ± 2
$t\bar{t}H$ (120)	42228	1.42 ± 0.06	56 ± 2	0.53 ± 0.04	21 ± 1
$t\bar{t}H$ (130)	19127	1.57 ± 0.09	39 ± 2	0.61 ± 0.06	15 ± 1
$t\bar{t}b\bar{b}$	372737	0.176 ± 0.007	297 ± 12	0.0641 ± 0.0041	109 ± 7
$t\bar{t}1j$	393000	0.0038 ± 0.0010	390 ± 100	0.00025 ± 0.00025	26 ± 26
$t\bar{t}2j$	568999	0.0067 ± 0.0011	401 ± 65	0.00123 ± 0.00046	74 ± 28
$t\bar{t}3j$	101000	0.0040 ± 0.0020	95 ± 48	0.	$< 27 (68\% C.L)$
$t\bar{t}4j$	86697	0.0023 ± 0.0016	84 ± 60	0.	$< 48 (68\% C.L.)$
$Zt\bar{t}$	50000	0.064 ± 0.011	22 ± 4	0.022 ± 0.007	7 ± 2
Total Background			1289		< 291
$S/\sqrt{B}$ (115)			1.8		1.5
S/B (115)			5.1%		8.6%
$S/\sqrt{B}$ (120)			1.6		1.2
S/B (120)			4.4%		7.2%
$S/\sqrt{B}$ (130)			1.1		0.9
S/B (130)			3.0%		5.2%

5.2 Di-lepton Channel: $t\bar{t}H \rightarrow b\bar{b}b\bar{b}l'\nu'l\nu$

Di-lepton $t\bar{t}H$ events are selected by requiring two reconstructed leptons (e,μ) accompanied by significant missing transverse energy and at least four but no more than seven jets, three or four of which have been b-tagged according to the *Combined Secondary Vertex* b-tagging algorithm. The details of these cuts follow.

Lepton identification is performed using the electron and muon likelihoods described in Section 4. In the semi-

leptonic analyses, events with more than one identified lepton are vetoed, but in the di-lepton analysis those same events are selected. The likelihood acceptance cuts used for leptons in the di-lepton channel are therefore chosen to be the same as the second-lepton veto cuts for both semi-leptonic channels. In this way, the sample of events for the di-lepton $t\bar{t}H$ analysis is by construction strictly complementary to those used in the semi-leptonic channels.

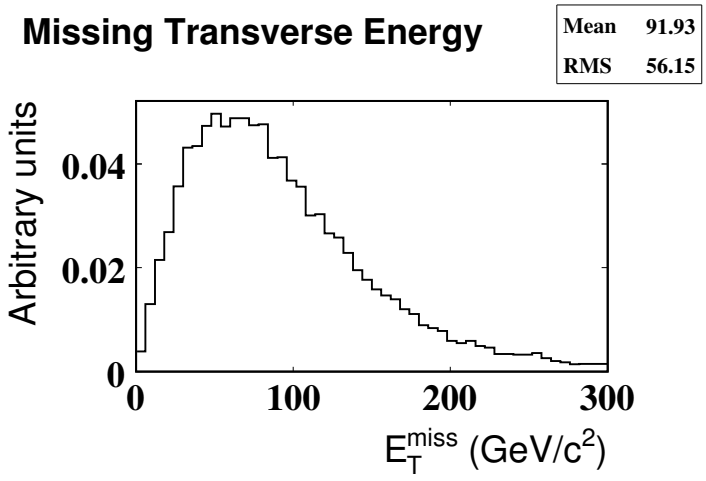


Figure 20: Missing Transverse Energy for non-exclusive $t\bar{t}H$ (m_H = 120 GeV/c^2) events.

Missing transverse energy is corrected for jet calibration and muon momenta according to Equation 2. After these corrections, the distribution shown in Figure 20 is obtained. At present, the di-lepton analysis is a counting experiment, thus no effort has been made to assign the missing transverse energy to the two neutrinos from the hard event.

Jets are reconstructed using the iterative cone algorithm, with electromagnetic plus hadronic calorimeter towers as input and a cone size of $\Delta R = 0.5$. The parameters and calibration used in the reconstruction follow the recommendations for PTDR II publications. A number of other recommended settings have been experimented with, but found to produce little variation in either the selection efficiencies or the "best case" Higgs resolution (see Figure 21 and associated caption).

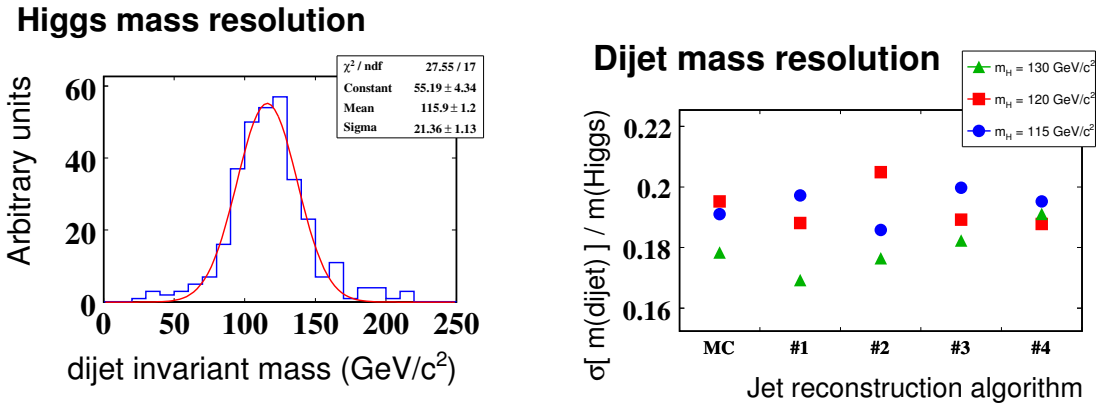


Figure 21: Left: Invariant mass of pairs of jets matched ($R < 0.3$) to the Monte Carlo b partons from Higgs. A Gaussian fit is performed to obtain the "best case" Higgs mass resolution σ. This example is for non-exclusive $t\bar{t}H$ with m_H = 120 GeV/c^2, with jets calibrated using the PTDR II recommendation 1 settings. Right: The relative "best case" Higgs mass resolution $\sigma/m(\text{Higgs})$ for various Higgs masses and jet calibrations (MonteCarlo, PTDR II recommendations 1, 2, 3, and 4). $m(\text{Higgs})$ is the mean reconstructed Higgs mass according to the Gaussian fit.

As is done for the semi-leptonic electron channel, jets that are found to be within an $R < 0.1$ cone of any selected electron are removed. This affects about 1% of all reconstructed "jets" that are in actuality mis-identified electrons.

The details of the di-lepton $t\bar{t}H$ selection are summarized below:

- 2 oppositely charged leptons (e,μ) passing id criteria ($-Log(L_\mu) < 1.4$ for muons, $-Log(L_e) < 1.2$ for

electrons)

- corrected $E_T^{\text{miss}} > 40$ GeV
- 4 to 7 jets with calibrated $E_T > 20$ GeV and $|\eta| < 2.5$
- ≥ 3 selected jets b-tagged with discriminator $D > 0.7$

The above is termed the "loose" working point because there is evidence that it is possible to increase the purity (S/B), of the selection by way of the following more stingent "tight" criteria:

- lepton selection as above
- E_T^{miss} requirement as above
- 4 to 6 jets with calibrated $E_T > 20$ GeV and $|\eta| < 2.5$
- ≥ 4 selected jets b-tagged with discriminator $D > 0.7$

to increase the purity (S/B) of the selection. Although the naive significance $S/\sqrt{(B)}$ decreases, the cleaner selection is plagued less by systematic uncertainties which dominate the more realistic significance $S/\sqrt{B+dB^2}$. However, the numbers quoted for the "tight" working point are currently insufficiently precise because of limited dataset sizes at the time of writing. One should therefore not neglect the errors accompanying them.

The efficiency of the above selections have been studied for signal with Higgs of masses m_H = 115, 120 and 130 GeV$/c^2$, as well as the dominant background processes. Signal samples were generated with the Higgs forced to decay to $b\bar{b}$. The W^- was forced to decay leptonically (e, μ, τ), but the W^+ is allowed to decay freely. Such a "non-exclusive" dataset incurs a branching ratio of $1/3$, which has been factored into the selection efficiencies reported in Table 8. This choice takes into account the contribution from semi-leptonic top decays which are manifested or mis-reconstructed as di-lepton events; the same applies to tau decays which are mis-recontructed as e, μ. Furthermore,in order to save time and effort, the large all-hadronic piece is avoided since it has a very low selection rate given that two leptons have to be "faked" for such an event to be selected.

The contributions from the background processes, $t\bar{t}b\bar{b}$, $t\bar{t}$Nj, and $t\bar{t}Z$ are estimated using the samples described in Section 2. These events have small selection efficiencies, so very large samples must be analyzed. To make these samples more manageable, a loose pre-selection requiring at least 3 b-tags with discriminator $D > 0.7$ is applied before analysis.

The software versions used in the analysis are summarized in Table 7.

Table 7: Software used in generation of di-lepton $t\bar{t}H$ signal dataset. The analysis software is the same for both signal and background.

	Software	Version
Monte-Carlo	CMKIN	4.3.1
	running: PYTHIA	6.227
Full simulation	OSCAR	3.9.8
Digitization	ORCA	8.13.1
Analysis	ORCA	8.13.3
	framework: ExRootAnlysis	customized

5.2.1 Results

The selection efficiencies for the two working points, with the corresponding number of expected events and the signal significance, are reported in Tables 8. The number of expected events is computed for an integrated luminosity of 60 fb^{-1}. In the interest of conciseness, and since it has already been factored into the eventual selection efficiencies, the background pre-selection efficiency is not shown in the table.

Since event selection is quite simple for the di-lepton channel, it is possible to formulate equations predicting the selection efficiencies. This is detailed in Appendix B where some back-of-the-envelop calculations are performed

to estimate efficiencies for signal and backgrounds. The calculations include some of the backgrounds that were not otherwise taken into account in this analysis and indicate that their yields are negligible compared to those backgrounds that were studied with Monte Carlo generated and simulated data samples.

Table 8: Selection efficiency $\varepsilon_{\text{loose}}$ (including branching fraction where applicable) and resulting number of expected events N_{loose} in 60 fb^{-1}, for the di-lepton $t\bar{t}H$ channel. For a glimpse of possible improvements, the same for a tighter set of cuts is provided ($\varepsilon_{\text{tight}}$, N_{tight}). The signal datasets are labeled by the generated Higgs mass in GeV/c^2 (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the full mass range.

	Analyzed ev.	$\varepsilon_{\text{loose}}(\%)$	N^{ev}_{loose}	$\varepsilon_{\text{tight}}(\%)$	N^{ev}_{tight}
$t\bar{t}H$ (115)	27900	0.511 ± 0.025	168 ± 8	0.088 ± 0.010	29 ± 3
$t\bar{t}H$ (120)	26141	0.490 ± 0.025	132 ± 7	0.070 ± 0.009	19 ± 3
$t\bar{t}H$ (130)	25911	0.490 ± 0.025	82 ± 4	0.072 ± 0.010	12 ± 2
$t\bar{t}b\bar{b}$	313894	0.637 ± 0.014	1080 ± 24	0.094 ± 0.007	159 ± 12
$t\bar{t}$1j	280385	0.0125 ± 0.0021	1270 ± 220	0	< 42 (68% C.L.)
$t\bar{t}$2j	276917	0.0448 ± 0.0040	2690 ± 240	0.00144 ± 0.00072	87 ± 43
$t\bar{t}$3j	90367	0.0553 ± 0.0078	1330 ± 190	0	< 31 (68% C.L.)
$t\bar{t}$4j	120042	0.0716 ± 0.0077	2620 ± 280	0.0025 ± 0.0014	92 ± 53
$t\bar{t}Z$	110156	0.304 ± 0.017 0 103 ± 6	0.0363 ± 0.0057	12 ± 2	
all backgrounds			9090		< 422
$S/\sqrt{B}$ (115)			1.8		1.4
S/B (115)			1.8 (%)		6.9 (%)
$S/\sqrt{B}$ (120)			1.4		0.9
S/B (120)			1.5 (%)		4.5 (%)
$S/\sqrt{B}$ (130)			0.9		0.6
S/B (130)			0.9 (%)		2.9 (%)

5.3 All-hadron Channel: $t\bar{t}H \rightarrow b\bar{b}b\bar{b}qq'q''q'''$

A χ^2 method has been adopted as baseline for the all-hadron channel following the recipe given in Section 4.4 to select the optimal jet cone size parameter. A number of kinematics variables, together with the b-tagging discriminator, have been studied to optimize the signal selection with respect to background rejection. Moreover, in order to combine the results from the 4 different decay sub-channels, a veto on leptons has been applied using the complementary cut developed within the semi- and di-leptonic analyses: the event is discarded if $-Log(L_\mu) < 1.4$ or $-Log(L_e) < 1.2$.

The final variables are the following:

- Jet Transverse Energies of the 8 most energetic jets in the tracker acceptance
- Combined b-Tag discriminator value for each jet
- Centrality of the event defined as $\sum_{i=0}^{8} E_T^i / E^i$
- Centrality of the Higgs defined similarly, with the sum restricted to the 2 jets paired to the Higgs

The 8 most energetic jets in the tracker acceptance ($|\eta| < 2.7$) have been used for the analysis. Fig. 22 shows the E_T distribution for the 7th and 8th most energetic jets for signal and background samples while Fig. 23 shows the b-Tag discriminator variable for the third and fourth jet (ordered by the b-Tag variable). Event and Higgs centrality distributions are shown in Fig. 24. The Higgs centrality variable has been computed after the mass χ^2 minimization which is responsible for jet to parton matching.

The jet to parton pairing is performed using the mass χ^2 defined in Eq. 3. All possible combinations in jet-parton assignments for the 8 selected jets (ordered from the lowest to the highest χ^2) are considered starting from the lowest χ^2. Two working points have been chosen: the first uses loose cuts on bTagging to get higher statistical significance (but lower S/B), while the second uses a tighter cut on bTagging to obtain a higher S/B (but lower significance).

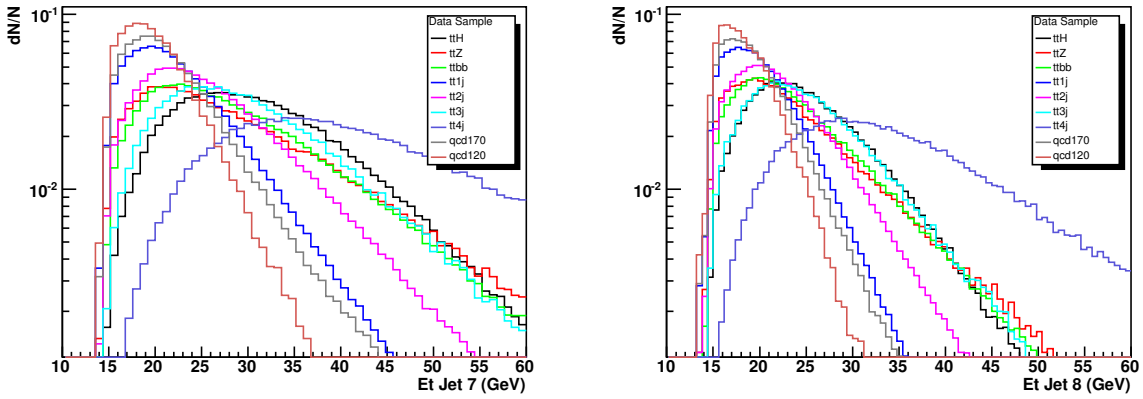


Figure 22: E_T distribution for the 7th (left) and 8th (right) (E_T ordered from highest) jet for different datasets as indicated in the legend.

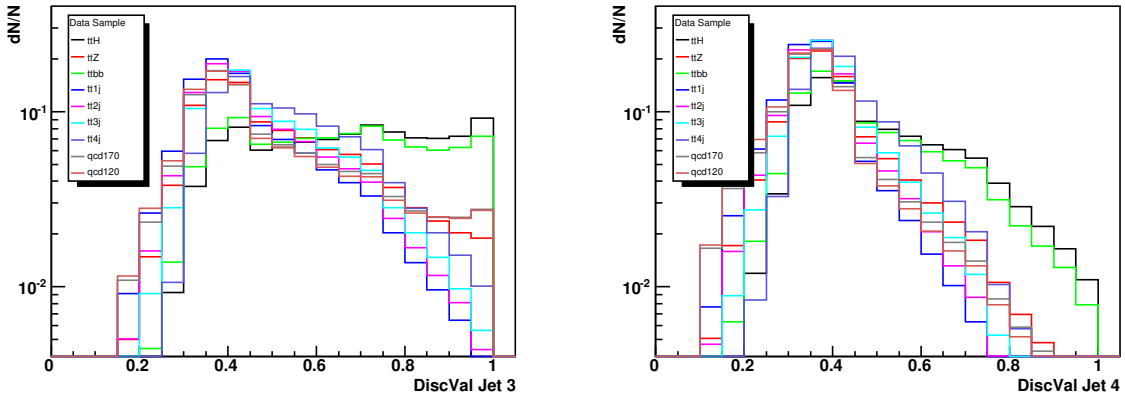


Figure 23: bTag discriminator distribution for the 3rd (left) and 4th (right) jets (ordered by the bTag variable starting from highest) for different datasets as indicated in the legend.

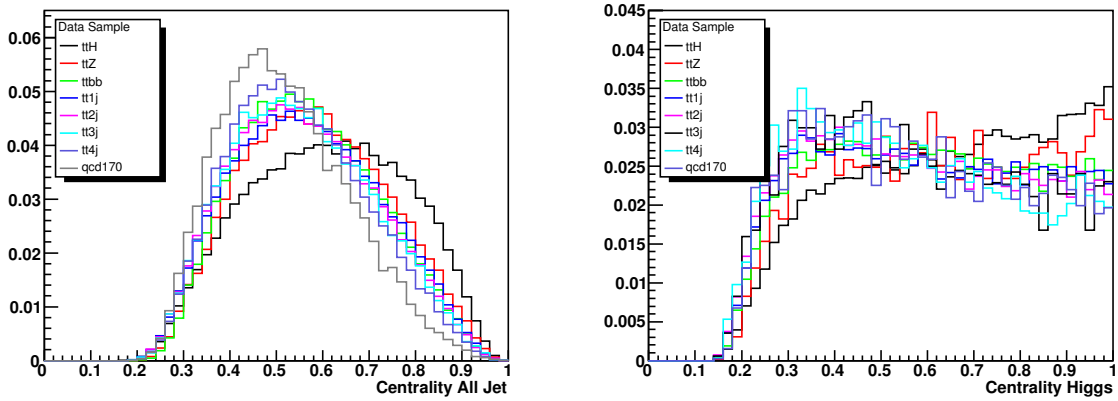


Figure 24: Event (left) and Higgs (right) centrality distributions for different datasets as indicated in the legend.

For the "loose" working point an event is selected if the following conditions are satisfied:

- $E_T^{7th} > 30$ GeV and $E_T^{8th} > 20$ GeV for the E_T ordered jets
- the χ^2 for each of the 2 W's and 2 t quarks are within 3 sigma of their expected values
- the third highest combined b-Tag discriminator for the 4 jets associated to the b-partons satisfies $D_3 > 0.80$
- Higgs centrality higher than 0.55 and no cut on Event Centrality

For the "tight" working point an event is selected if the following conditions are satisfied:

- $E_T^{7th} > 30$ GeV and $E_T^{8th} > 20$ GeV for the E_T ordered jets
- the χ^2 for each of the 2 W's and 2 t quarks are within 3 sigma of their expected values
- the 3 highest combined b-Tag discriminator for the 4 jets associated to the b-partons satisfy $D_3 > 0.85$ while the fourth satisfies $D_4 > 0.70$
- Event and Higgs centrality, respectively, in excess of 0.55 and 0.80

All the applied cuts have been optimized to obtain the highest significance while keeping the S/B ratio as high as possible. All values chosen for E_T^{7th}, E_T^{8th}, D_3, D_4, Event and Higgs centrality have been varied simultaneously, thereby mapping out the complete space of possibilities within the following limits:

- 20 GeV $< E_T^{8th} <$ 40 GeV
- $E_T^{8th} < E_T^{7th} < E_T^{8th} +$ 40 GeV
- $0.5 < D_3$ and $D_4 < 0.95$
- Event and Higgs Centrality in the range [0.50-0.95]

As an example, Figures 25 through 28 show how significance and S/B change upon varying one cut while keeping all the other cut values fixed. Variation of more than one cut has also been tested and the final, implemented set of cut values is that for which significance and S/B are optimal.

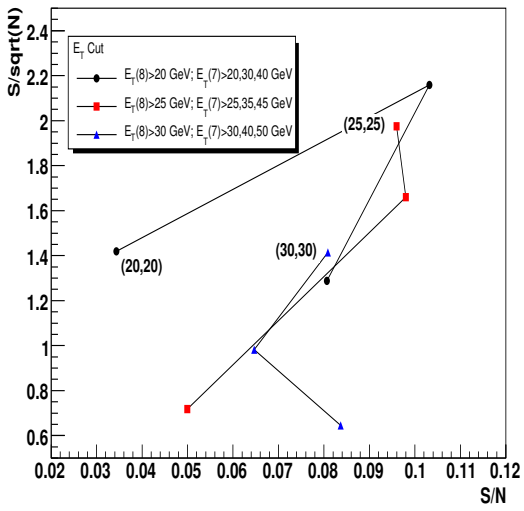


Figure 25: Different E_T cuts

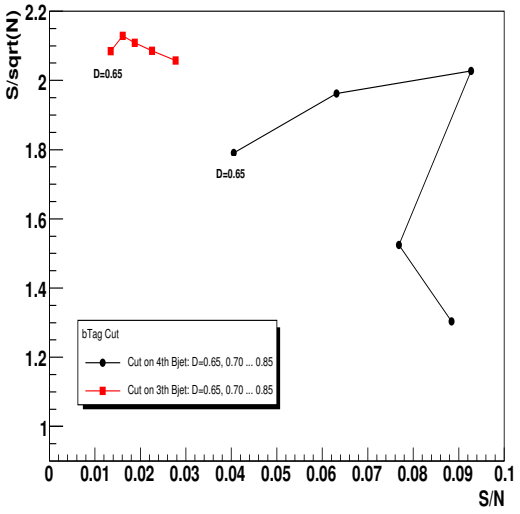


Figure 26: Different bTag discriminator cuts

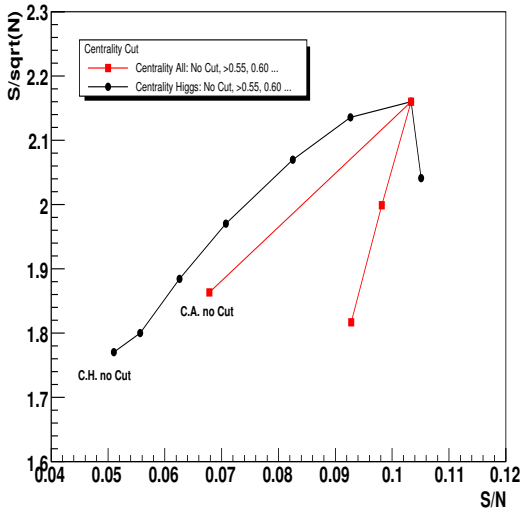


Figure 27: Different centrality cuts

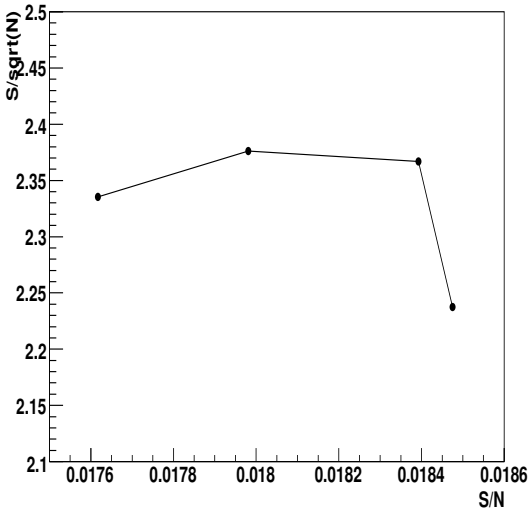


Figure 28: Different eta cuts, ranging from 2.4 to 3.0 in steps of 0.2.

5.3.1 Results

The number of analyzed events, selection efficiencies with the corresponding number of expected events and the signal significance are reported in Tables 9 for the all-hadron decay channel. The number of expected events is computed for an integrated luminosity of 60 fb^{-1}. Both working points are considered.

5.4 "Tight" vs. "loose" Working Points

Upon reading Tables 6 (semi-leptonic), 8 (di-lepton), and 9 (all-hadron), one may be persuaded that it is preferable to impose very tight selection cuts. In particular, improvements in the signal-to-background ratios by factors of greater than 2 and up to about 5 are obtainable, with a much smaller decrease in $S/\sqrt{B}$. The purity of the selection is especially important at the range of achieveable S and B for the analyses as they stand, in the presence of systematic uncertainties dB. This can be simply understood from the observation that $dB^2 \gg B$, so the significance, once systematics have been included, is $S/\sqrt{B+dB^2} \sim S/dB$. As will be seen in Section 6.1, $dB \sim 0.2B$ so that $S/dB \sim 5S/B$. Thus S/B (not $S/\sqrt{B}$) needs to be optimized for best performance.

A note of caution should, however, be injected. At conception, the analyses were planned around the "loose" working points. As such, the results for those working points are well-understood and carry with them reasonably small statistical errors. At least an order of magnitude more Monte Carlo events need to be generated in order to reduce the fluctuations seen for the "tight" selections to comparable levels. The reader is thus advised that discussion of the "tight" working points is provided in the spirit of probing what might be obtained if the analyses are stretched to these extremes.. Nevertheless, as will be seen, even for the"tight" selection, the final significance, though improved, does not qualitatively improve the observability of this channel for an integrated luminosity of 60 fb^{-1}.

6 Discussion of Systematic Uncertainties

In this section, sources of systematic uncertainty in the results presented in the previous sections are estimated.

6.1 Estimation of "Standard" CMS Systematics

The uncertainties in various quantities are estimated based on the expected knowledge of the CMS experiment when 60 fb^{-1} of data have been collected. In keeping with other CMS analyses, the following "standard" sources of systematic error are considered:

Table 9: Analyzed events, selection efficiency, number of expected events and signal significance in 60 fb^{-1} for the all-hadron $t\bar{t}H$ channel for the two working points $\varepsilon_{\text{loose}}$ and $\varepsilon_{\text{tight}}$. The signal datasets are labeled by the generated Higgs mass in GeV/c^2 (parentheses). Also quoted are binomial errors arising from the finite sizes of processed datasets. All numbers refer to the full mass range.

	Analyzed Ev.	$\varepsilon_{\text{loose}}(\%)$	N^{ev}_{loose} 60fb^{-1}	$\varepsilon_{\text{tight}}(\%)$	N^{ev}_{tight} 60fb^{-1}
$t\bar{t}H$ (115)	49636	2.32 ± 0.07	347 ± 10	0.294 ± 0.015	44 ± 4
$t\bar{t}H$ (120)	163494	2.55 ± 0.03	314 ± 5	0.366 ± 0.024	45 ± 2
$t\bar{t}H$ (130)	43254	2.80 ± 0.08	214 ± 6	0.358 ± 0.029	27 ± 2
$t\bar{t}b\bar{b}$	203135	0.702 ± 0.019	1190 ± 31	0.0645 ± 0.0056	109 ± 9
$t\bar{t}$1j	1031551	0.0084 ± 0.0009	860 ± 92	0.0005 ± 0.0002	49 ± 22
$t\bar{t}$2j	559111	0.0333 ± 0.0024	2000 ± 150	0.0009 ± 0.0004	54 ± 24
$t\bar{t}$3j	68015	0.079 ± 0.011	1910 ± 260	0.0015 ± 0.0015	35 ± 35
$t\bar{t}$4j	97334	0.182 ± 0.014	6660 ± 500	0.0021 ± 0.0015	75 ± 53
$Zt\bar{t}$	80226	0.358 ± 0.021	121 ± 7	0.0312 ± 0.0062	11 ± 2
qcd170	264310	0.0238 ± 0.0030	4810 ± 610	0.0004 ± 0.0004	76 ± 76
qcd120	55128	0.0018 ± 0.0018	83 ± 83	0 ± 0	$<$95(68%*C.L.*)
Total Background			17600		< 505
$S/\sqrt{B}$ (115)			2.6		2.0
S/B (115)			2.0%		8.7%
$S/\sqrt{B}$ (120)			2.4		2.0
S/B (120)			1.8%		8.9%
$S/\sqrt{B}$ (130)			1.6		1.2
S/B (130)			1.2%		5.4%

- Jet energy scale (JES)
- Jet resolution
- b-jet and c-jet tagging efficiencies
- uds-jet tagging efficiencies
- Luminosity

The systematics listed above are assumed to be uncorrelated. Each is varied independently according to procedures detailed below. Each variation produces a change in the selection efficiency $\Delta\varepsilon$ and the corresponding change in expected event yields ΔN_X ($X = t\bar{t}H, t\bar{t}1j, ...$) for the signal and background. In cases where the "up" and "down" variation produces an asymmetric range for $\Delta\varepsilon$, the range is symmetrized by taking the maximum of the two. The change in yields for all backgrounds are then added in quadrature to obtain the total expected change in background yields ΔN due to all systematics.

Following commonly agreed prescriptions for CMS analyses, the JES uncertainty is taken into account by shifting the calibrated energies of each jet up (down) by a relative 10% for jets with transverse momentum less than 20 GeV/c. For jets with transverse momentum between 20 GeV/c and 50 GeV/c, the relative variation falls linearly with increasing transverse momentum from 10% to a final 3%. Jets with transverse momentum greater than 50 GeV/c are assumed to have a flat 3% JES uncertainty.

The jet resolution should be smeared by an overall 10%. This is achieved by shifting the energy of each jet by a random number drawn from a Gaussian distribution of mean zero and width σ:

$$E_T^{jet} \rightarrow E_T^{jet} + Gaus(0, \sigma_{jet}) \tag{6}$$

For the semi-leptonic and di-lepton channels, σ_{jet} is taken to be 10% of the jet energy, $\sigma_{jet} = 0.1\ E_T^{jet}$. For the all-hadron channel, $\sigma_{jet} = 0.046\ \sigma(E_T^{jet}, \eta)$ where the nominal jet resolution in various regions of the detector is given in Eq. 7 ($\oplus$ indicates addition in quadrature):

$$\begin{aligned}
\sigma(E_T^{jet}, |\eta| < 1.4) &= (5.8GeV) \oplus (1.25\sqrt{E_T^{jet}}) \oplus (0.033\ E_T^{jet}) \\
\sigma(E_T^{jet}, 1.4 < |\eta| < 3.0) &= (4.8GeV) \oplus (0.89\sqrt{E_T^{jet}}) \oplus (0.043\ E_T^{jet}) \\
\sigma(E_T^{jet}, 3.0 < |\eta| < 5.0) &= (3.8GeV) \oplus (0.085\ E_T^{jet}) \qquad (7)
\end{aligned}$$

For purposes of the b-tagging systematic, all jets in the event are categorized by flavor according to Monte Carlo truth. Official CMS tools in the b-tagging subsystem of ORCA are used by the all-hadron and semi-leptonic analyses to perform jet-parton matching, but the di-lepton channel uses a simple prescription as follows: If a jet is within an $R < 0.3$ cone of some (generator) b-hadron, the jet is called a b-jet. Jets that are not b-jets are then examined to see if they lie within an $R < 0.3$ cone of some c-hadron, and if so called a c-jet. Finally, all remaining jets are called uds-jets. Given that the flavor of the jet has somehow been determined, the following relative uncertainties in the tagging efficiencies of jets of the various flavors are considered:

- 4% for b- and c-jets
- 10% for u-, d-, and s-jets[2)]

As suggested by the grouping, the tagging efficiencies are varied simultaneously for b-jets and c-jets, taking into account that any uncertainty is likely to be fully correlated between them. Similarly but in a separate run, the u-, d-, and s-jet tagging efficiencies are varied together by the stated amount. The procedure for affecting a relative uncertainty of $\delta\varepsilon$ in tagging efficiency is to throw a flat random number for each jet (in the category of interest), and untagging it a fraction $\delta\varepsilon$ of the time. Similarly, the tagging efficiency can be varied upwards by tagging untagged jets a fraction $\delta\varepsilon \times \varepsilon_{tag}/(1-\varepsilon_{tag})$ of the time, where ε_{tag} is the tagging efficiency for that category. This factor is required to properly scale the un- tagging probability so as to effect an increase that is a $\delta\varepsilon$ relative to ε_{tag}.

The above procedure assumes that any correlation between the b,c-tagging and uds- tagging systematics is small. For the di-lepton channel, the pre-selection of 3 b-tags on the background datasets makes it nonsensical to increase the tagging efficiencies, since there exist many events that will never be selected simply because they are not present in the dataset. As the observed ranges in the signal are quite symmetric, however, the downward variation is taken to be the range of the systematic.

Lastly, the systematic uncertainty in luminosity is assumed to be 3%, from which a corresponding change in yield is calculated.

Table 11 shows the effect of the various systematic uncertainties on the selection efficiencies for the di-lepton analysis. For the semi-leptonic channels, systematic effects were evaluated for the muon-channel only. Since none of the sources of uncertainty considered involve lepton identification, the results are assumed to be applicable to the electron-channel as well. Table 10 shows the effect of the various systematic uncertainties on the selection efficiencies for the muon analysis, while Table 12 shows the same for the all-hadron analysis.

Due to the finite sizes of datasets used to evaluate the reported efficiencies, there is a binomial error $\sqrt{\varepsilon(1-\varepsilon)/N}$ from having selected some number of events with efficiency ε out of N analyzed events. This "statistical" uncertainty is shown in the first column of Tables 11-12. It is not included in the quadratic sum since it is a trivially reducible contribution.

Finally, there is a potential additional source of statistical uncertainty due to the random processes required in the stated procedures for varying the systematics. Since this can also be significant, calculations are iterated. until the change in $\Delta\varepsilon$ between two successive runs is less than 5% (of $\Delta\varepsilon$).

[2)] Previous experience indicates that the material included in even the most detailed detector simulations usually underestimates that in the installed detector. As such, tracks undergo more multiple-scattering and their impact parameter resolution is worsened. Thus, for a given lifetime tagger, there will be significantly more mistagged light-quark jets than is predicted by the simulation. Since the estimation of the light flavour mistagging uncertainty is quite difficult and the impact on the analysis is of major importance, one might be interested the effect of say a 50% uds-tagging uncertainty. This has been determined to correspond roughly to a factor of 5 increase in the "uds-tagging" systematic uncertainties reported in Tables 11-12. Note that this should be interpreted as a possible shift in the central values of S and B due to a corrected tagger working point. It should not be confused with the expectation for ultimate uncertainty on uds tagging. Once data is available, measurements on control samples should allow one to know the uds-tag rate to better than 10%.

In Tables 13-16, the systematic uncertainties are propagated to the expected signal significance.

All these uncertainties have been calculated for the loose woking points of the analyses, since the tight working points do not give enough remaining background events to reasonably quantify a systematic uncertainty. Tentative numbers are provided in Appendix D to give a flavor of what they might be, but have large errors associated with them due to the limited statistics. Since the values are not vastly different from the better-understood results at the loose working points, the latter are preferred for use in estimated impacts of systematics at the tight working point.

Appendix C explores the effect of plausible eventual differences in the working parameters of the CMS experiment versus the present simulation of it.

6.2 Background Rates from Data

There are relatively large theoretical uncertainties in the cross-sections used to normalize the signal yields [15], and even larger theoretical uncertainties in those used for the $t\bar{t}$+jets backgrounds [16]. These have not been included as part of the systematic errors considered above, because when the CMS experiment reaches maturity, estimating the $t\bar{t}$+jets background directly from data ought to be possible. In this way, the uncertainty associated with Monte Carlo derived tagging rates would be avoided entirely. For example, the number of mis- tagged $t\bar{t}$+jets can be factorized as follows:

$$N_{t\bar{t}jj}^{mistag} = N_{t\bar{t}jj}^{no-tag} \times Pr(uds \rightarrow b; E_T, \eta, ...)$$

where $N_{t\bar{t}jj}^{no-tag}$ is a high purity (e.g. fully reconstructed with a mass window) top sample that has been obtained without requiring b-tagging and $Pr(uds \rightarrow b; E_T, \eta, ...)$, is a parametrized "fake matrix" that is derived from some independent dataset (e.g dijet data) which yields the probability for a light quark jet to fake a secondary vertex. It may also be possible to derive this fake matrix from the top sample itself. If a high-purity (e.g. double-tagged and fully reconstructed) semi-leptonic top sample were selected, the jets belonging to the hadronic W would provide a source of both light quark and charm jets. From these data, a measurement of the corresponding uds-tag and c-tag rates at the relevant energy could be obtained.

6.3 Combined Significance

Since the event samples for the channels studied in this note are strictly disjoint, the results can be combined by simply adding the individual signal yields (background yields) to obtain a summed S (B).

For each of the considered systematics, the resultant error in background yields are added for all four channels, since they are by definition fully correlated. The summed errors are then added by quadratures to get a combined systematic uncertainty dB. One then calculates the significance, inclusive of systematic uncertainties in the background yield, according to the formula $S/\sqrt{B+dB^2}$.

As per the discussion in Section 5.4, it is of interest to see how much better the results have the potential to be at tighter working points for the various analyses. Since the systematic uncertainties are not well quantified at these "tight" working points, because of a lack in Monte Carlo Statistics, the same uncertainties as for the "loose" working points are used to reduce spurious statistical effects. The justification for this procedure is given in Appendix D. However, since these uncertainties can only be evaluated approximately, the results for the "tight" cuts should be understood to be indicative of what could be obtained with a possibly re-optimized analysis. This exercise indicates that further studies using a larger amount of Monte Carlo statistics might be able to achieve a higher sensitivity in this channel. Even so, the overall expectation for discovery potential for an integrated luminosity of 60 fb-1 is not qualitatively improved.

Also as noted in the preceeding section, it may be possible to know the fraction of background in the event samples to relatively high precision. In the case of $t\bar{t}$Nj, this could be the case because it will be possible to isolate large, signal depleted samples of these events. For irreducible backgrounds, particularly $t\bar{t}b\bar{b}$, it is more difficult to normalize the background contributions. A data-driven estimate of this background would then rely upon universality of b fragmentation to estimate the background from a measure of $t\bar{t}$Nj production and the mistag matrix, also measured in data. Another possibility would be to loosen some selection cuts in order to accept a larger portion of the $t\bar{t}b\bar{b}$ background events in regions where there is very little signal expected. These could then be used, together with Monte Carlo, to normalize the background contribution in the signal region. Finally,

Table 10: Systematic uncertainties relative to selection efficiencies (in percent) for the semi-leptonic $t\bar{t}H$ channels. Σ is the quadrature sum of all changes in the given row, except the statistical ones (provided for comparison only). Σ also includes the 3% uncertainty of the luminosity which is the same for all the samples. The last two colums show the abolute uncertainty (in number of events) at the two different working points ε_2 and ε_3.

semi-leptonic	statistical	JES+jet res	bc-tagging	uds-tagging	Σ	# events $\varepsilon_{\text{loose}}$	# events $\varepsilon_{\text{tight}}$
$t\bar{t}H$ (115)	4.2	20.1	10.7	1.0	23.0	22	9
$t\bar{t}H$ (120)	3.5	18.9	10.9	0.8	22.0	17	6
$t\bar{t}H$ (130)	4.7	17.8	9.4	0.7	20.4	11	4
$t\bar{t}b\bar{b}$	3.3	16.6	9.5	0.8	19.4	81	29
$t\bar{t}1j$	22.3	33.9	14.3	5.0	37.3	193	29
$t\bar{t}2j$	12.9	22.5	7.2	7.0	24.8	157	10
$t\bar{t}3j$	44.9	15.4	5.0	5.0	17.2	21	5
$t\bar{t}4j$	58.2	23.2	5.3	6.6	24.9	31	12
muon channel							
all backgrounds $\varepsilon_{\text{loose}}$		23.7	9.4	4.8	25.9	476	
all backgrounds $\varepsilon_{\text{tight}}$		21.5	9.1	3.6	23.6		83
electron channel							
all backgrounds $\varepsilon_{\text{loose}}$		23.7	9.47	4.67	26.4	339	
all backgrounds $\varepsilon_{\text{tight}}$		20.2	8.01	4.08	22.6		66

Table 11: Systematic uncertainties relative to selection efficiencies (in percent) for the di-lepton $t\bar{t}H$ channel. Σ is the quadrature sum of all changes in the given row, except the statistical ones (provided for comparison only). Σ also includes the 3% uncertainty of the luminosity which is the same for all the samples. The last two colums show the abolute uncertainty (in number of events) at the two different working points $\varepsilon_{\text{loose}}$ and $\varepsilon_{\text{tight}}$.

di-lepton	statistical	JES	jet res.	bc-tagging	uds-tagging	Σ	# events $\varepsilon_{\text{loose}}$	# events $\varepsilon_{\text{tight}}$
$t\bar{t}H$ (115)	4.8	2.34	1.81	9.46	0.514	10.4	17	3
$t\bar{t}H$ (120)	5.07	3.65	1.77	9.38	0.651	10.7	14	2
$t\bar{t}H$ (130)	5.09	3.41	1.18	9.76	0.787	10.9	9	1
$t\bar{t}b\bar{b}$	2.23	5.26	0.851	8.62	0.572	10.6	114	17
$t\bar{t}1j$	16.9	25.7	3.81	10.7	9.71	29.9	380	13
$t\bar{t}2j$	8.98	12.1	1.13	11.0	5.97	17.7	477	15
$t\bar{t}3j$	14.1	6.0	7.33	10.7	7.71	16.5	220	5
$t\bar{t}4j$	10.8	16.3	5.81	6.51	7.85	20.3	532	19
$t\bar{t}Z$	5.46	6.87	3.16	8.96	3.58	12.6	13	2
all backgrounds $\varepsilon_{\text{loose}}$		13.4	3.7	9.3	6.6	18.3	1660	
all backgrounds $\varepsilon_{\text{tight}}$		11.2	2.8	9.0	4.8	15.7		66

Table 12: Systematic uncertainties relative to selection efficiencies (in percent) for the all-hadron $t\bar{t}H$ channel. Σ is the quadrature sum of all changes in the given row, except the statistical ones (provided for comparison only). The $t\bar{t}H$ signal row corresponds to m_H=120 GeV. Σ also includes the 3% uncertainty of the luminosity which is the same for all the samples. The last two colums show the abolute uncertainty (in number of events) at the two different working points ε_1 and ε_2.

all-hadron	statistical	JES	jet res.	bc-tagging	uds-tagging	Σ	# events $\varepsilon_{\text{loose}}$	# events $\varepsilon_{\text{tight}}$
$t\bar{t}H$ (120)	1.53	17.6	7.0	6.0	0.6	20.1	63.1	9.0
$t\bar{t}b\bar{b}$	2.64	14.2	5.9	6.4	0.7	16.9	201	19
$t\bar{t}$1j	10.7	43.7	9.2	2.3	4.6	45.1	388	22
$t\bar{t}$2j	7.33	23.1	10.8	5.9	5.9	27.0	539	15
$t\bar{t}$3j	13.6	18.5	3.7	3.7	3.7	19.8	377	7
$t\bar{t}$4j	7.51	5.6	0.6	6.8	4.5	10.4	689	8
$t\bar{t}Z$	5.89	17.8	8.4	7.0	1.7	21.2	26	2
qcd 170	12.6	52.4	15.9	7.9	4.8	55.6	2670	42
qcd 120	100	100	0.0	0.0	0.0	100	83	95
all backgrounds $\varepsilon_{\text{loose}}$		24.7	7.1	6.4	4.4	27	4760	
all backgrounds $\varepsilon_{\text{tight}}$		39.1	6.3	4.9	2.9	40		202

Table 13: Significance of muon channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different cuts on L_{bSele}, but assuming the same systematic uncertainties (as computed at the first working point) for both.

muon	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$L_{bSele} > 0.55$ ($\varepsilon_{\text{loose}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	96	0.052	2.2	0.20
$t\bar{t}H$ (m_H=120 GeV/c^2)	75	0.041	1.8	0.15
$t\bar{t}H$ (m_H=130 GeV/c^2)	55	0.030	1.3	0.11
$L_{bSele} > 0.75$ ($\varepsilon_{\text{tight}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	38	0.108	2.0	0.44
$t\bar{t}H$ (m_H=120 GeV/c^2)	29	0.082	1.6	0.34
$t\bar{t}H$ (m_H=130 GeV/c^2)	21	0.060	1.1	0.24

Table 14: Significance of electron channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different cuts on L_{bSele}, but assuming the same systematic uncertainties (as computed at the first working point) for both.

electron	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$L_{bSele} > 0.55$ ($\varepsilon_{\text{loose}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	66	0.051	1.8	0.20
$t\bar{t}H$ (m_H=120 GeV/c^2)	56	0.044	1.6	0.17
$t\bar{t}H$ (m_H=130 GeV/c^2)	39	0.030	1.1	0.12
$L_{bSele} > 0.75$ ($\varepsilon_{\text{tight}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	25	0.086	1.5	0.37
$t\bar{t}H$ (m_H=120 GeV/c^2)	21	0.072	1.2	0.31
$t\bar{t}H$ (m_H=130 GeV/c^2)	15	0.052	0.9	0.22

Table 15: Significance of di-lepton channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different set of cuts on number of jets and b-tagged jets, but assuming the same systematic uncertainties (as computed at the first working point) for both.

di-lepton	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
4-7 jets, 3-4 b-tagged ($\varepsilon_{\text{loose}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	170	0.018	1.8	0.10
$t\bar{t}H$ (m_H=120 GeV/c^2)	130	0.015	1.4	0.08
$t\bar{t}H$ (m_H=130 GeV/c^2)	82	0.009	0.9	0.05
4-6 jets, 4-6 b-tagged ($\varepsilon_{\text{tight}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	29	0.069	1.4	0.42
$t\bar{t}H$ (m_H=120 GeV/c^2)	19	0.045	0.9	0.27
$t\bar{t}H$ (m_H=130 GeV/c^2)	12	0.029	0.6	0.18

Table 16: Significance of all-hadron channel before and after taking into account the uncertainty dB in the total number of background events due to systematics. The result is shown for the two different working points as explained in Section 5.3.

hadron	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
Softer b-tag discriminator cut ($\varepsilon_{\text{loose}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	350	0.020	2.6	0.07
$t\bar{t}H$ (m_H=120 GeV/c^2)	310	0.018	2.4	0.07
$t\bar{t}H$ (m_H=130 GeV/c^2)	210	0.012	1.6	0.05
Harder b-tag discriminator cut, event centrality cut ($\varepsilon_{\text{tight}}$)				
$t\bar{t}H$ (m_H=115 GeV/c^2)	44	0.087	2.0	0.22
$t\bar{t}H$ (m_H=120 GeV/c^2)	45	0.089	2.0	0.22
$t\bar{t}H$ (m_H=130 GeV/c^2)	27	0.054	1.2	0.13

theoretical calculations of the $t\bar{t}b\bar{b}$ cross section at next-to-leading order, and the availability of a next-to-leading order Monte Carlo could be used to directly normalize this background.

Even with the above prescriptions, however, there will remain uncertainties in the cross-sections depending on how well the procedures can be carried out. It is difficult to predict at this time exactly what will be the level to which the backgrounds can be understood, because the tools required are not yet in existence and because this understanding requires real data. In view of this, it is interesting to consider how the combined significance of the measurements presented in this note would vary as a function of the fractional uncertainty in background cross-sections, i.e. dB_{xsec}/B. In principle these might have been included in the total systematic uncertainty dB_{sys}/B presented in Tables 11-12. However, since the uncertainties in these cross sections at the time of LHC operation are difficult to predict now, it was deemed to be more useful to factor them out and present them separately as described below.

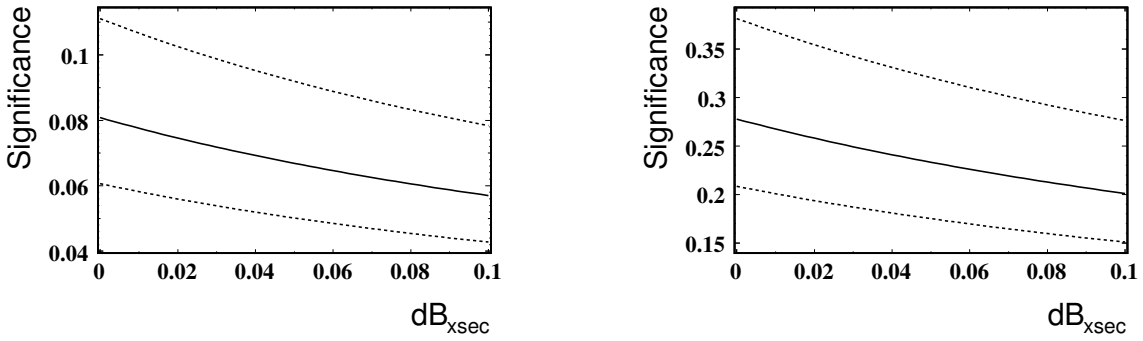


Figure 29: Expected range of combined significance (di-lepton + semi-leptonic + all-hadron, and includes the systematic uncertainties estimated in Section 6.1) versus an additional systematic uncertainty on the background cross-section as a fraction of total background. Left: Results for the "loose" working points. Right: Results for the "tight" working points.

The solid central line in Figure 29 shows how the combined significance $S/\sqrt{B+(dB_{sys}+dB_{xsec})^2}$ degrades as a function of dB_{xsec}/B. The signal and background yields for the "tight" working points are used in the right side of Figure 29, because these give the best results after inclusion of systematics.

Other than this "fundamental" cross-section uncertainty, there are also the "correctible" errors in the cross-sections used at the time of writing, which can be compensated for once data has been collected. The upper and lower dashed curves in Figure 29 show the maximum and minimum allowed excursions, should the signal and background cross-sections be off by 10% and 20% respectively. Thus the upper (lower) dashed line corresponds to the signal cross-section scaled up (down) by 10% while at the same time the background cross-section is scaled down (up) by 20%.

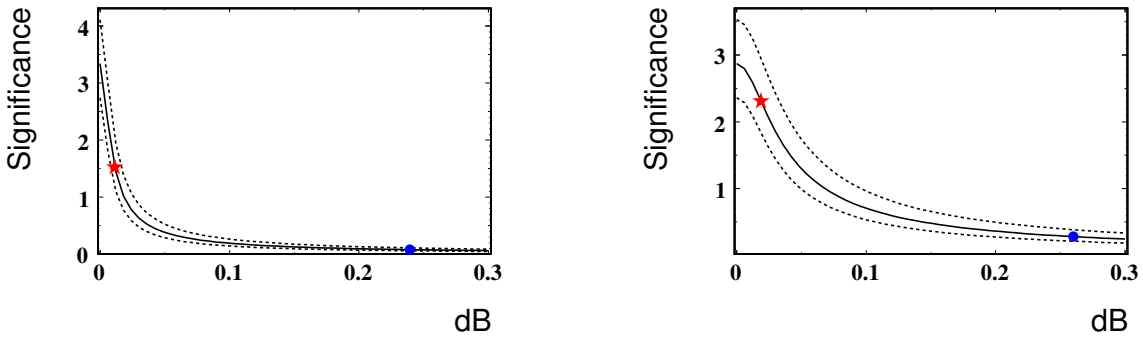


Figure 30: Expected range of combined significance (di-lepton + semi-leptonic + all-hadron) versus the total systematic uncertainty in background as a fraction of total background. Left: Results for the "loose" working points. Right: Results for the "tight" working points.

It is also of interest to see how much better the analyses could do if the total systematic uncertainty can be reduced (i.e. the region left of zero in Figure 29). Hence, Figure 30 shows the full range of obtainable significances, with the dot marking the currently estimated value with no cross-section uncertainty ($dB = dB_{sys}$). The star corresponds to what one would obtain for 1% and 4% uncertainties on the $t\bar{t}Nj$ and $t\bar{t}b\bar{b}$ backgrounds, respectively, an arbitrarily chosen reference. It is interesting to note that it does not quite yield a substantial signficance, even though background uncertainties of 1% and 4% for $t\bar{t}Nj$ and $t\bar{t}b\bar{b}$ are probably substantially better than what will be accessible in reality. This highlights the challenge that is faced in observing $t\bar{t}H$.

6.4 Prospects for Improvement

Among the most promising potential improvements that could have significant impact on this analysis is the application of an "energy flow", which combines all detector components, especially the tracker, to obtain a better jet reconstruction performance.

Also the b-tagging performance will benefit from this. There are several possible further improvements for the b-tagging as briefly mentioned in Section 4.5.

In a mature experiment, more complex triggers should have been implemented to the effect that the trigger efficiencies stated in previous sections will have been improved upon. As an example, consider lepton triggers, which appear from Table 5 to be around a rather disappointing 60%. Single lepton trigger efficiencies of 90% or better should be achieved if combined with any of the other physics objects present in $t\bar{t}H$ events (e.g. b-tags and/or missing transverse energy). If necessary, to further increase the trigger efficiency for di-lepton events, one could create an OR of multi-lepton triggers with lower p_T thresholds with single lepton triggers.

Given a 90% efficiency, datasets with purely di-lepton W^+W^- decays should have single-lepton-stream trigger efficiency close to:

$$0.9^2 + 2 \times 0.9\ (1 - 0.9) = 99\%$$

with the first term being where both leptons trigger and the second being where either leptons trigger. For fully inclusive datasets, then, the overall triggering efficiency is weighted by the relative fractions of di-lepton and semi-leptonic events, being 4/81 and 14/81 respectively (for "leptons" not including taus):

$$(4/81)[0.9^2 + 2 \times 0.9\ (1 - 0.9)] + (14/81) \times 0.9 \sim 20\%$$

From Tables 5 and 6 one can then calculate the expected yields for 90% trigger efficiencies in signal (exclusive datasets) and 20% trigger efficiencies in background (inclusive datasets). The prescription is to simply multiply the yields in Table 6 by $\varepsilon_{\text{mature}}/\varepsilon_{\text{current}}$ where $\varepsilon_{\text{mature}}$ is the mature trigger efficiencies (90% or 20%) and $\varepsilon_{\text{current}}$ the values stated in Table 5.

A 90% single-lepton trigger efficiency is thus predicted to improve the significance of the semi-leptonic channels individually by around 5%[3)].

7 Summary and Outlook

A full simulation study of all decay channels of the $t\bar{t}H$ with $H \to b\bar{b}$ channel has been performed and the discovery potential after an integrated luminosity of 60 fb^{-1} has been evaluated.

In contrast to earlier studies, the global picture for the this analysis is substantially more pessimistic. This change is due to mainly to the greater degree of realism that was made possible for this study by use of more advanced tools for event generation, detector simulation and physics reconstruction, which were not available to previous (fast-) simulation studies. Mistagging of light flavour jets in $t\bar{t}jj$ events, for example, proved to be a substantially more serious problem than had been foreseen in studies that made use of parametrized b-tagging. This effect is particularly difficult to estimate without a full detector simulation based upon a relatively detailed material description of the apparatus, coupled to an equally detailed track reconstruction program.

On the other hand, it has been noted that the availability of large contol samples of top events will enable b tagging of high energy jets to be very well understood in CMS. This will probably enable some further suppression of

[3)] For the "loose" working point and assuming no change in the systematics from Table 10.

light quark and charm jet tagging relative to b tagging. Similarly, experience with real data will likely improve jet reconstruction and energy measurements. This, in turn, will enhance the efficiency of many of the techniques described in this paper.

A Comparison of CompHEP and ALPGEN for the $t\bar{t}jj$ Background

A.1 Comparison at Generator Level

The $t\bar{t}jj$ background was originally generated with CompHEP and proved to be the most dangerous background. Since processes with extra jets in the final states are better described by ALPGEN for the reasons described in section 2, this generator has also been used, and the comparison of Alpgen and CompHEP generated events is presented in this appendix.

The main feature of ALPGEN version 2 is the matching procedure introduced in the Matrix Element (ME) to Parton Shower (PS) interfacing. The parton-shower matching criteria avoid double counting due to the fact that initial and final state radiation are added by the PS generator on top of the extra jets already described at parton level. If no matching is applied from ME to PS generator a significant overestimation of the rate of extra-jet production occurs. This major improvement in jet accounting is responsible for the reduction of the cross section of these background sources by more than a factor of two. In addition, to avoid double counting, the ALPGEN matching procedure allows the kinematics of the extra jets to be better described. Actually, a PS generator provides a more reliable description of extra jets with low transverse momentum, while a ME generator is more suitable to describe extra jets in the higher region of the p_t spectra of jets.

Even though a direct comparison of the CompHEP and ALPGEN samples is not straightforward, because of a different Q-scale and different PDFs (CTEQ4L and CTEQ5L) used, a comparison is attempted below which exploits the main difference introduced by the parton shower matching. As mentioned in Section 2, exclusive samples of $t\bar{t}$ with exactly one, two and three extra jets, respectively, and an inclusive $t\bar{t}$ sample with at least four extra jets have been generated with ALPGEN to be compared with the inclusive $t\bar{t}jj$ CompHEP sample.

The cross sections on generator level after applying similar kinematical cuts is shown in Table 17. After including

Table 17: Comparison of the cross sections of the inclusive CompHEP $t\bar{t}jj$ sample and the exclusive ALPGEN samples after application of the same kinematical cuts: $p_t > 20$ GeV, $|\eta| < 3$, $\Delta R(j,j) > 0.7$. These cuts systematically reduce the cross sections listed in Table 4.

CompHEP $t\bar{t}jj$	330pb
ALPGEN exclusive $t\bar{t}1j$	120pb
ALPGEN exclusive $t\bar{t}2j$	73pb
ALPGEN exclusive $t\bar{t}3j$	32pb
ALPGEN inclusive $t\bar{t}4j$	51pb

the two-jet and the higher jet multiplicities, the ALPGEN cross section is still a factor of two smaller than the CompHEP cross section, where the CompHEP sample represents an inclusive two-jet or higher multiplicity.

Jets have been reconstructed at generator level using the PYTHIA routine PYCELL which simulates a calorimetric grid in the η-ϕ ($|\eta| <3$) plane made of 0.1×0.1 cells and uses as input the hadronization particles. The parameters used in the MC jet reconstruction are: 2 GeV cell E_t threshold, 20 GeV for the E_t jet threshold and jet cone radius of 0.4. Variations of jet-parameter values do not significantly change the results.

Table 18 shows the event selection efficiency for a very basic choice of the selection cuts, i.e. single muon HLT, p_t cuts and b-tagging cuts.

Table 18: Comparison of the event selection efficiency after single muon HLT and cuts.

	HLT	6 jets with $p_t > 20$GeV and $\eta < 2.4$	4 b-jets with $disc > 0.7$
CompHEP $t\bar{t}jj$	18%	8.3%	0.05%
exclusive ALPGEN $t\bar{t}1j$	14%	2.2%	0.008%
exclusive ALPGEN $t\bar{t}2j$	14%	4.7%	0.019%
exclusive ALPGEN $t\bar{t}3j$	14%	8.3%	0.038%
inclusive ALPGEN $t\bar{t}4j$	13.4%	11.2%	0.13%

Figure 31 shows the distribution of the number of reconstructed jets. The CompHEP events have higher jet mul-

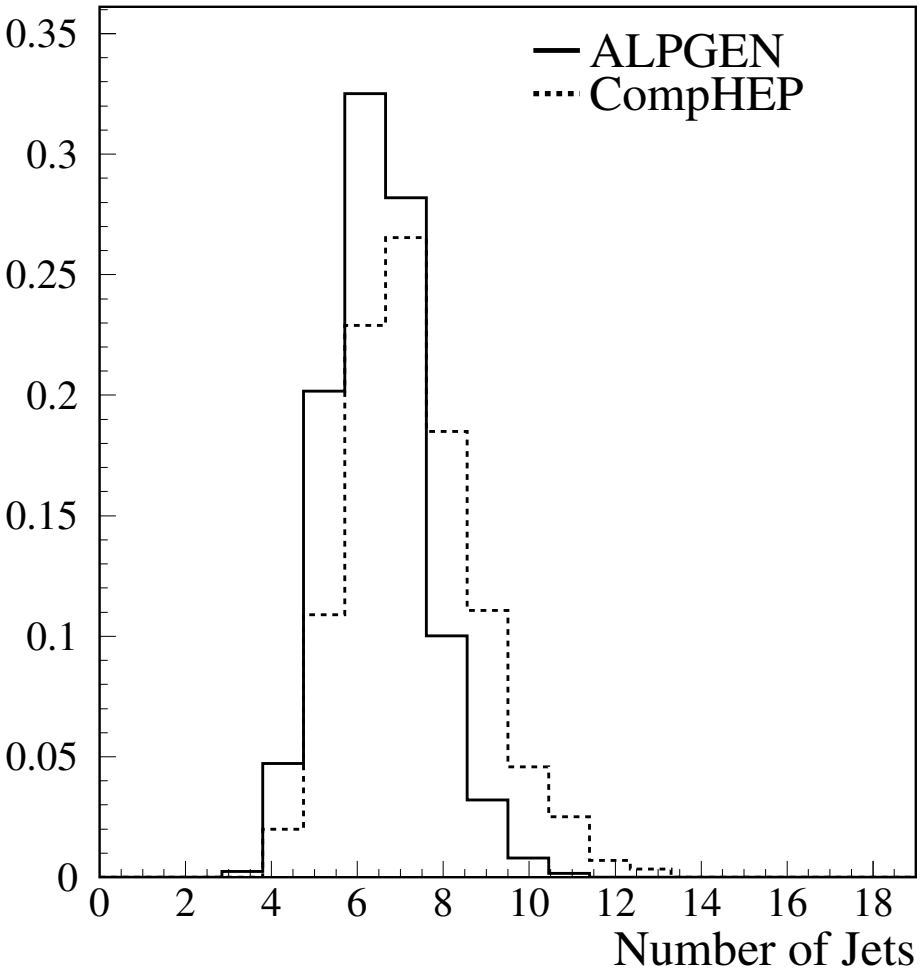


Figure 31: Comparison of the number of reconstructed jets at generator level.

tiplicity than the ALPGEN events. The distributions of the transverse momenta of the leading jets after the same cuts are shown in Fig. 32. The leading jets in CompHEP events are shown to be systematically more energetic than those in the ALPGEN events.

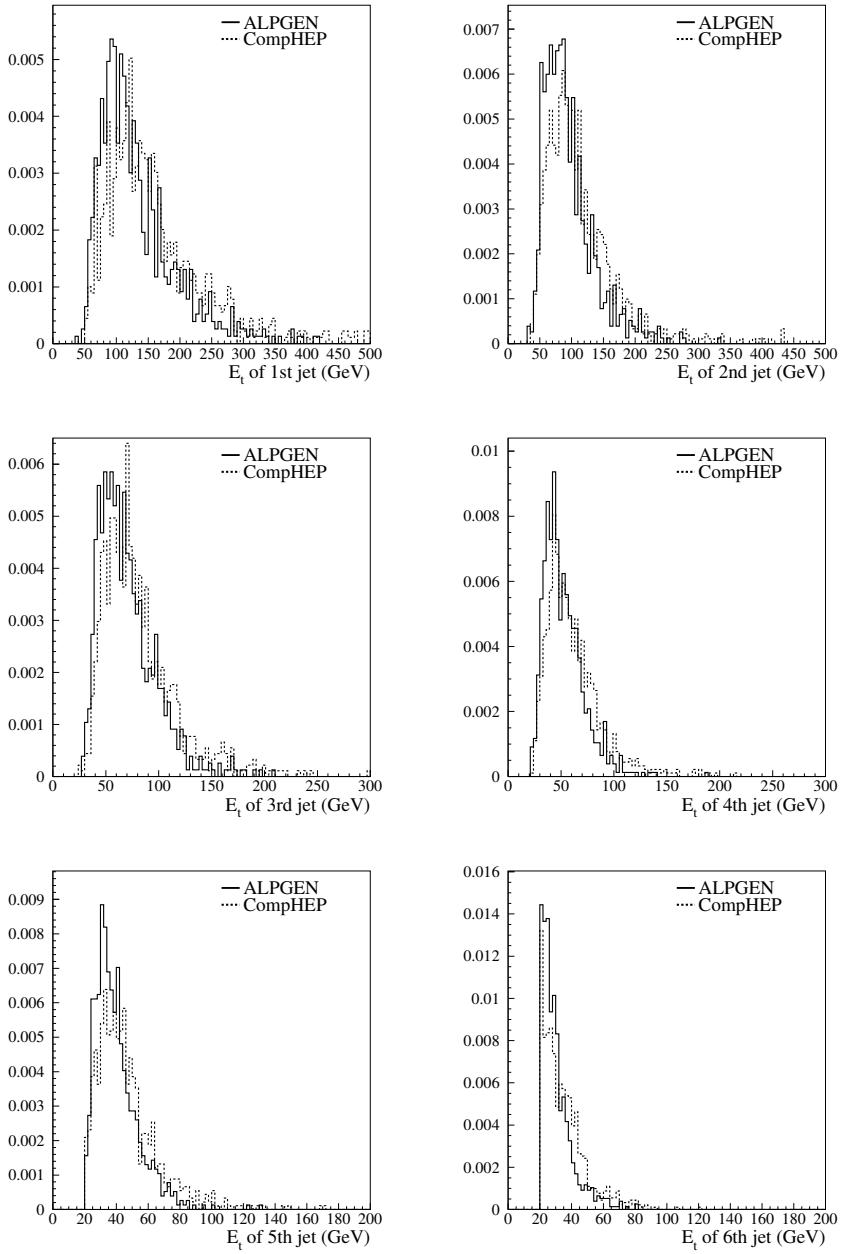


Figure 32: Comparison of the p_t spectra of the six leading jets at generator level.

A.2 Comparison of Reconstructed Information

Both the CompHEP and ALPGEN samples have been reconstructed using the setup described in Section 2. Some reconstructed values are compared in the following.

Figure 33 shows the number of reconstructed jets which pass a p_t cut of 20 GeV, while Fig. 34 and 35 show the spectra of the transverse momenta of the six leading jets.

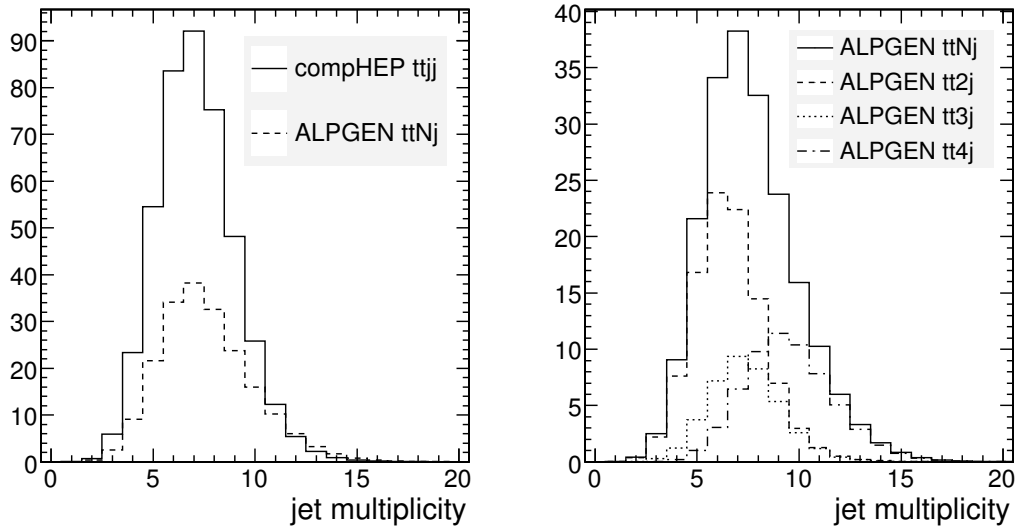


Figure 33: Comparison of the number of reconstructed jets above a p_t threshold of 20 GeV. $t\bar{t}$Nj represents the sum of the four ALPGEN multiplicities $t\bar{t}$1j, $t\bar{t}$2j, $t\bar{t}$3j and $t\bar{t}$4j. The histograms are normalized to the respective cross sections of each multiplicity.

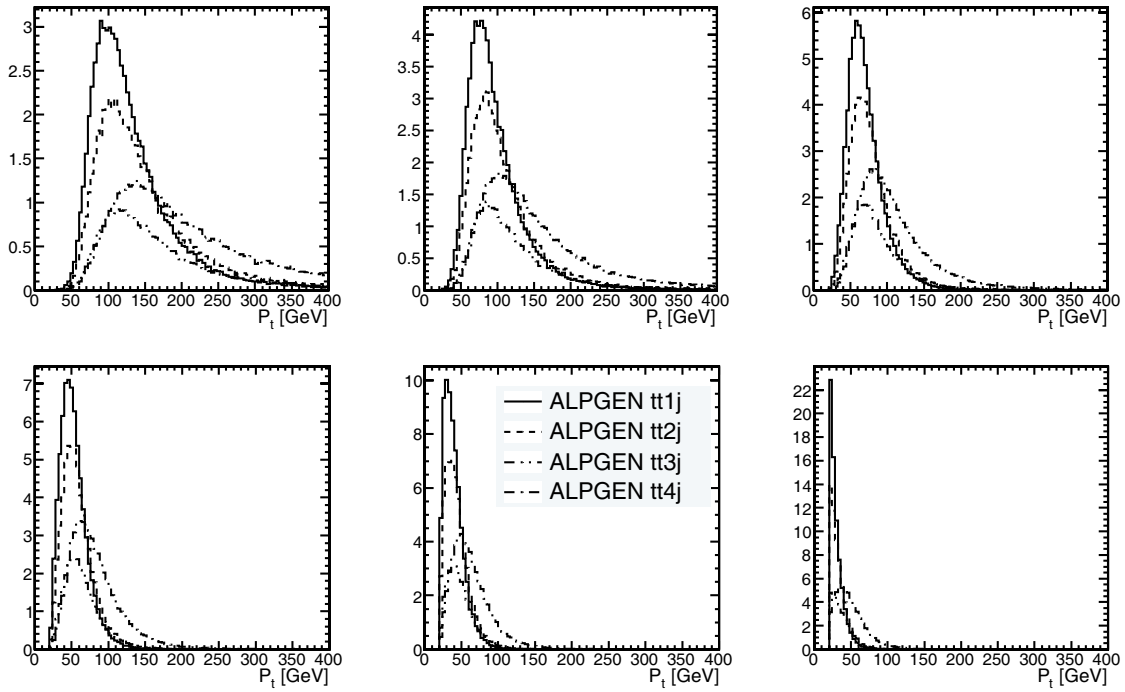


Figure 34: Comparison of the transverse momentum of the six leading jets for ALPGEN. The histograms are in order of decreasing p_t from top left to bottom right. The histograms are normalized to the respective cross sections of each multiplicity.

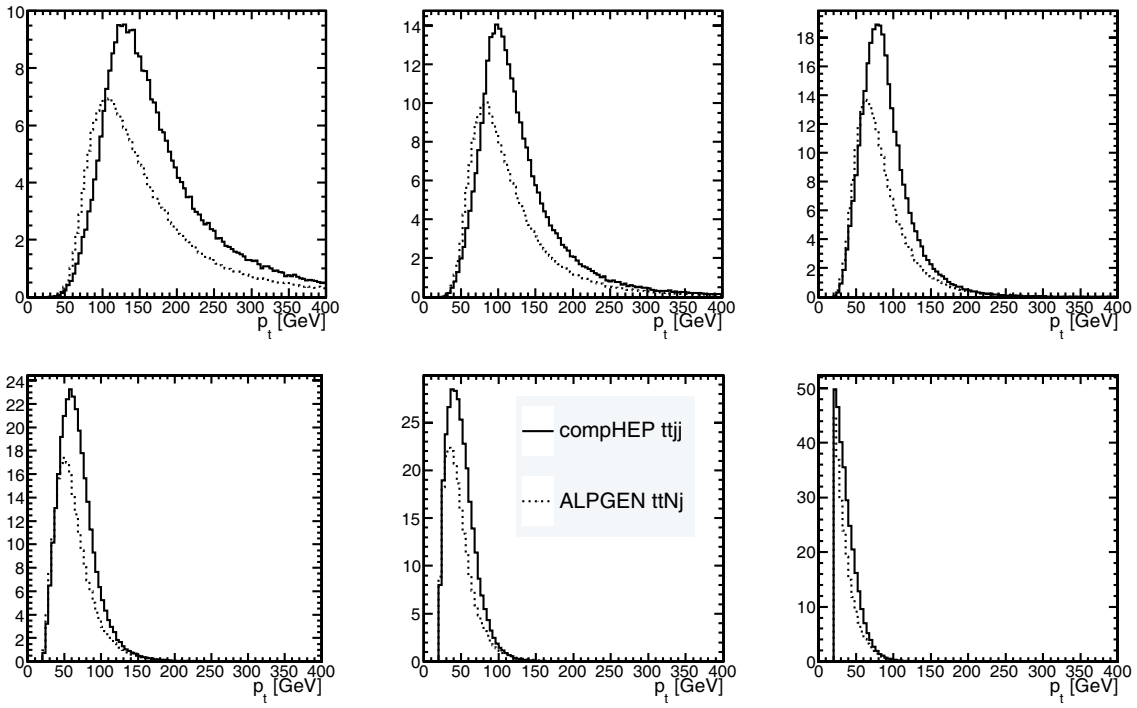


Figure 35: Comparison of the transverse momentum of the six leading jets. The histograms are in order of decreasing p_t from top left to bottom right. $t\bar{t}$Nj shows the sum of the three ALPGEN multiplicities $t\bar{t}$1j, $t\bar{t}$2j, $t\bar{t}$3j and $t\bar{t}$4j, while $t\bar{t}$jj represents the inclusive CompHEP sample. The histograms are normalized to the respective cross sections of each multiplicity.

B Estimation of the Selection Efficiencies

As mentioned in Section 5.2, it is possible to use back-of-the-envelop calculations to estimate the selection efficiencies for various channels. For simplicity, only the particles in the hard event are considered, ignoring complications such as initial- and final-state radiated jets which may additionally contribute to the selection.

Of the four event-level cuts presented for the di-lepton event selection, the number-of-b-tags requirement has the largest impact on the selection of events with four b's (signal and the irreducible $t\bar{t}b\bar{b}$ background) versus events with only two ($t\bar{t}$Nj). The two-leptons requirement is also an important ingredient, if less directly, in order to estimate the fractions of events with the various W^+W^- decay modes. These are the only two selection criteria considered in this toy calculation.

B.1 Preliminaries

In the following calculations, $\approx$ indicates a numerical approximation, whereas $\sim$ indicates an approximation in the physics. The multinomial coefficient Equation 8 gives the combinatorial factor for the number of ways i objects, then j objects, and so on, out of a total N, can be assigned without regards to order:

$$C_{i,j}^{N} \equiv \frac{N!}{i!j!\cdots(N-i-j-\cdots)!} \text{ for } 0 \leq i+j+\cdots \leq N, \text{ 0 otherwise} \tag{8}$$

The variables appearing in the calculation are defined in Table 19, and the branching ratios required to correctly mix the different samples depending on their final state leptons/partons are listed in Table 20. Henceforth, jets are categorized by their originating parton, i.e. "b", "c", "q" (for either of u, d, or s), and "g" (jets from gluon splitting).

Table 19: Definitions of variables. The values are rough estimates for the working points used in the di-lepton analysis.

	Definition	Value
ε_b	efficiency of b-tagging a b-jet	60%
ε_c	efficiency of b-tagging a c-jet	15%
ε_q	efficiency of b-tagging a u- , d-, or s-jet	2.0%
ε_g	efficiency of b-tagging a g-jet	4.0%
ε_l	efficiency of selecting a lepton	85%
ε_f	efficiency of a jet faking a lepton	3%

Table 20: Branching ratios for the W gauge boson, with numbers from [14].

	Value
$\mathrm{BR}_{W\to l\nu}$	0.32
$\mathrm{BR}_{W\to cq}$	0.34
$\mathrm{BR}_{W\to qq}$	0.34

B.2 Calculations

If exactly n b-tags is the only requirement, the efficiency for selecting an event with n_b bottom, n_c charm, n_q of up, down, or strange jets, and n_g gluon jets, is given by Equation 9. It is a composition of terms similar to $C_i^{n_b}\varepsilon_b^i(1-\varepsilon_b)^{n_b-i}$, which is the probability of b-tagging i out of n_b b-jets, with the appropriate combinatorial factor to account for all possible choices of which b-jets to tag. The composition is subject to the constraint that the total number of tags is n, and this is indeed evident since every term in the sum has total power of ε_X ($X = b,c,q,g$) equal to n.

$$\varepsilon^n(n_b,n_c,n_q,n_g) = \sum_{i=0}^{n}\sum_{j=0}^{n-i}\sum_{k=0}^{n-i-j}\left[C_i^{n_b}\varepsilon_b^i(1-\varepsilon_b)^{n_b-i}\right]\left[C_j^{n_c}\varepsilon_c^j(1-\varepsilon_c)^{n_c-j}\right]\left[C_k^{n_q}\varepsilon_q^k(1-\varepsilon_q)^{n_q-k}\right] \times\left[C_{n-i-j-k}^{n_g}\varepsilon_g^{n-i-j-k}(1-\varepsilon_g)^{n_g-(n-i-j-k)}\right] \quad (9)$$

Table 21: Factors in the selection efficiency due to the 2-lepton requirement, for various W branching modes. The values are calculated using the approximate observed efficiencies $\varepsilon_l \approx 85\%$ for selecting signal leptons and $\varepsilon_l \approx 3\%$ for selecting "fake" leptons.

W^+W^- decay channel	Variable	Expression	Value
di-lepton	ε_{ll}	ε_l^2	72.2%
semi-lepton	$\varepsilon_{lq}=\varepsilon_{ql}$	$\varepsilon_l\varepsilon_f$	2.55%
all-hadron	ε_{qq}	ε_f^2	0.0900%

The slightly more involved part of the calculation concerns counting the number of jets of each flavor and adding their contributions to the selection efficiency according to the fractions present in each sample. For samples with two tops decaying to two b's and two W's, for various W branching modes the 2-lepton requirement weights the selection efficiency by the probability of reconstructing two leptons given that there are either two, one, or zero "real" leptons (from the hard event). These factors are given in Table 21. The (potentially poor) assumption is made that in cases where there are less than two leptons in the hard event (and only those cases), a "fake" lepton can somehow be reconstructed with some fixed probability ε_f. To keep the complexity level manageable, details such as the following are omitted:

- ε_f should in principle depend on the number of jets in the event.
- In events with two "real" leptons, the analysis might still select one "real" and one "fake" lepton.

and etcetera[4)].

The above are all the ingredients required for computing the n-tags selection efficiency for $t\bar{t}b\bar{b}$, assuming that only the top and W decay products plus the two "extra" b's are relevant:

$$\begin{aligned}\varepsilon^n_{ttbb} = &\left[\varepsilon_{ll}\mathrm{BR}^2_{W\to l\nu}\varepsilon^n(4,0,0,0)\right] \\ &+\left[\varepsilon_{lq}C_1^2\mathrm{BR}_{W\to l\nu}\mathrm{BR}_{W\to cq}\varepsilon^n(4,1,1,0)+\varepsilon_{lq}C_1^2\mathrm{BR}_{W\to l\nu}\mathrm{BR}_{W\to qq}\varepsilon^n(4,0,2,0)\right] \\ &+\left[\varepsilon_{qq}\mathrm{BR}^2_{W\to cq}\varepsilon^n(4,2,2,0)+\varepsilon_{qq}C_1^2\mathrm{BR}_{W\to cq}\mathrm{BR}_{W\to qq}\varepsilon^n(4,1,3,0)+\varepsilon_{qq}\mathrm{BR}^2_{W\to qq}\varepsilon^n(4,0,4,0)\right] \quad (10)\end{aligned}$$

The first set of square brackets in Equation 10 corresponds to the selection efficiency for the subset of $t\bar{t}b\bar{b}$ with purely di-lepton, the second set to the subset with semi-leptonic, and the third set to the subset with fully hadronic W decays. The W decay products may contribute to the available charm and q jets up for selection, as reflected in the arguments to the various ε^n.

A similar if more general formula for the n-tags selection efficiency for $t\bar{t}$Nj, provided that only the top and W decay products plus the N jets are relevant, and assuming that n_c of the N jets are charmed, n_q are uds (with the remaining of course being gluon jets):

[4)] As a result of these approximations, in particular the flat ε_f fake rate, the so-called selection "efficiencies" are not properly normalized (can be greater than 1).

$$
\begin{aligned}
\varepsilon^n_{ttNj}(n_c,n_q) &= \varepsilon_{ll}\mathrm{BR}^2_{W\to lv}\varepsilon^n(2,n_c,n_q,N-n_c-n_q) \\
&+ \varepsilon_{lq}C_1^2\mathrm{BR}_{W\to lv}\mathrm{BR}_{W\to cq}\varepsilon^n(2,n_c+1,n_q+1,N-n_c-n_q) \\
&+ \varepsilon_{lq}C_1^2\mathrm{BR}_{W\to lv}\mathrm{BR}_{W\to qq}\varepsilon^n(2,n_c,n_q+2,N-n_c-n_q) \\
&+ \varepsilon_{qq}\mathrm{BR}^2_{W\to cq}\varepsilon^n(2,n_c+2,n_q+2,N-n_c-n_q) \\
&+ \varepsilon_{qq}C_1^2\mathrm{BR}_{W\to cq}\mathrm{BR}_{W\to qq}\varepsilon^n(2,n_c+1,n_q+3,N-n_c-n_q) \\
&+ \varepsilon_{qq}\mathrm{BR}^2_{W\to qq}\varepsilon^n(2,n_c,n_q+4,N-n_c-n_q) \qquad (11)
\end{aligned}
$$

The flavor composition of the N jets in the $t\bar{t}$Nj datasets are approximately (from examination of Monte Carlo truth):

$$
\begin{aligned}
f_c &\approx 2\% \\
f_{uds} &\approx 32\% \\
f_g &\approx 66\%
\end{aligned}
$$

There is in truth some variation in compositions for various N, i.e. the gluon-jet fraction decreases while the charm-jet fraction increases, but the rough values above are in the ballpark. The fraction of $t\bar{t}$Nj that contain exactly n_c charm jets and n_q uds jets are then modelled by:

$$
f(n_c,n_q;N) \sim C^N_{n_c,n_q} f_c^{n_c} f_q^{n_q} f_g^{N-n_c-n_q} \qquad (12)
$$

The above allows the relative fractions to be combined to obtain a total n-tags selection efficiency for $t\bar{t}$Nj:

$$
\varepsilon^n_{ttNj} = \sum_{n_c=0}^{N}\sum_{n_q=0}^{N-n_c} f(n_c,n_q;N)\varepsilon^n_{ttNj}(n_c,n_q) \qquad (13)
$$

B.3 Comparison to Observed Efficiencies

Allowing 3 or 4 b-tags as stated for the di-lepton selection, $(\varepsilon_{ttbb}/\varepsilon_{ttNj})_{calc} = (\varepsilon^4_{ttbb}+\varepsilon^3_{ttbb})/(\varepsilon^4_{ttNj}+\varepsilon^3_{ttNj})$ is calculated and compared to the observed ratio in the analysis.

The numbers in Table 22 show that the observed selection efficiencies are on average a factor of five smaller than the calculated ones. This is indicative of the fact that other factors in the selection (the missing transverse energy, HLT efficiency, etc.) are important for obtaining the exact efficiencies. However, the ratio of efficiencies between various channels, where one might expect such details to be "divided out", is more or less consistent with the back-of-the-envelop calculation. The larger discrepancy for $t\bar{t}$1j is likely due to the fact that Equation 13 does not (particularly for that channel) take into account other jets present in the system. Since it assumes that there are no other jets, the formula has the undesirable behavior of dropping the di-lepton contribution to $t\bar{t}$1j altogether, because at least one of the W's must undergo a hadronic decay in order to provide the fourth required jet.

B.4 Estimates of 'W'+jets and 'WW'+jets Background

It is also simple to write down the n-tags selection efficiency for W + N jets. These have no b-jets (assuming all the N jets are light-quark jets, since the gluon to $b\bar{b}$ branching ratio is negligibly small), but otherwise follow the same logic:

$$
\begin{aligned}
\varepsilon^n_{WNj} = \sum_{n_c=0}^{N}\sum_{n_q=0}^{N-n_c} f(n_c,n_q;N)\ [\ &\varepsilon_l\varepsilon_f\mathrm{BR}_{W\to lv}\varepsilon^n(0,n_c,n_q,N-n_c-n_q) \\
&+ \varepsilon_f^2\mathrm{BR}_{W\to cq}\varepsilon^n(0,n_c+1,n_q+1,N-n_c-n_q) + \varepsilon_f^2\mathrm{BR}_{W\to qq}\varepsilon^n(0,n_c,n_q+2,N-n_c-n_q)\ ] \qquad (14)
\end{aligned}
$$

Table 22: Calculated versus observed selection efficiencies. ttX refers to any one of the $t\bar{t}$ datasets as stated in the column headers.

	$t\bar{t}H$	$t\bar{t}b\bar{b}$	$t\bar{t}1j$	$t\bar{t}2j$	$t\bar{t}3j$	$t\bar{t}4j$
$(\varepsilon_{ttX})_{calc}$ (%)	4.09	4.09	0.135	0.225	0.316	0.408
$(\varepsilon_{ttX})_{obs}$ (%)	0.49	0.637	0.0125	0.0448	0.0553	0.0716
$(\varepsilon_{ttX})_{calc}/(\varepsilon_{ttX})_{obs}$ (%)	8.35	6.42	10.8	5.03	5.72	5.69
$(\varepsilon_{ttbb}/\varepsilon_{ttX})_{calc}$	1	1	30.3	18.1	12.9	10.0
$(\varepsilon_{ttbb}/\varepsilon_{ttX})_{obs}$	1.44	1	51.0	14.2	11.5	8.88
$(\varepsilon_{ttH}/\varepsilon_{ttX})_{obs}$	1	0.769	39.2	10.9	8.85	6.83

and the n-tags selection efficiency for $WW + N$ jets:

$$\begin{aligned}\varepsilon^n_{WWNj} = \sum_{n_c=0}^{N}\sum_{n_q=0}^{N-n_c} f(n_c,n_q;N)\ [\ &\varepsilon_{ll}\mathrm{BR}^2_{W\to l\nu}\varepsilon^n(0,n_c,n_q,N-n_c-n_q)\\ &+\varepsilon_{lq}C_1^2\mathrm{BR}_{W\to l\nu}\mathrm{BR}_{W\to cq}\varepsilon^n(0,n_c+1,n_q+1,N-n_c-n_q)+\varepsilon_{lq}C_1^2\mathrm{BR}_{W\to l\nu}\mathrm{BR}_{W\to qq}\varepsilon^n(0,n_c,n_q+2,N-n_c-n_q)\\ &+\varepsilon_{qq}\mathrm{BR}^2_{W\to cq}\varepsilon^n(0,n_c+2,n_q+2,N-n_c-n_q)+\varepsilon_{qq}C_1^2\mathrm{BR}_{W\to cq}\mathrm{BR}_{W\to qq}\varepsilon^n(0,n_c+1,n_q+3,N-n_c-n_q)\\ &+\varepsilon_{qq}\mathrm{BR}^2_{W\to qq}\varepsilon^n(0,n_c,n_q+4,N-n_c-n_q)\ ]\end{aligned} \tag{15}$$

Table 23: Selection efficiencies and yields in $60fb^{-1}$ for $W + N$ jets as predicted by Equation 14

	Cross-section (pb)	Efficiency (%)	Yield in $60fb^{-1}$
W + 0 jets	90000	0	0
W + 1 jet	24000	3.02×10^{-6}	43.4
W + 2 jets	7500	1.12×10^{-5}	50.3
W + 3 jets	2170	4.54×10^{-5}	59.0
W + 4 jets	522	1.25×10^{-4}	39.1
W + 5 jets	135	2.67×10^{-4}	21.6
W + $\geq$6 jets	180	4.87×10^{-4}	52.6

Table 24: Selection efficiencies and yields in $60fb^{-1}$ for $WW + N$ jets as predicted by Equation 15

	Cross-section (pb)	Efficiency (%)	Yield in $60fb^{-1}$
WW + 0 jets	33.8	1.45×10^{-5}	0.294
WW + 1 jet	17.6	8.53×10^{-5}	0.899
WW + 2 jets	8.24	2.55×10^{-4}	1.26
WW + $\geq$3 jets	11.2	7.24×10^{-3}	4.87

The predicted efficiencies and corresponding yields for the W + jets and WW + jets samples are listed in Tables 23 and 24. The cross-sections used for those calculations are those calculated using Alpgen for purely leptonic (e,μ,τ) W decays, divided by $1/3$ (W + jets) and $36/81$ (WW + jets) in order to convert them to the cross-sections for inclusive W decays. The W + $\geq$6 jets cross-section is an estimate since the information was not yet available at the time of this publication.

Evidently, W + jets is predicted to have contributions that are small compared to the other backgrounds (in particular $t\bar{t}$Nj) considered in this analysis, which have yields on the order of thousands of events. Furthermore, Table 22 argues that these simple calculations overestimate the efficiencies by a factor of five, so the efficiencies and yields in Tables 23 are expected to be upper bounds, with more realistic numbers coming out to be some tens of events. The WW + jets samples, despite the higher selection rate due to availability of leptons in the hard event, have negligible yields as shown in Table 24.

One may additionally be concerned about backgrounds such as Z + jets (ZZ + jets), but those have cross-sections closely following those of W + jets (WW + jets), which are already found to be dismissable.

C Comparison of CMS and CDF Working Points

The following is an investigation of the possible variations in results, should the systematic uncertainties and working point of the CMS experiment as a whole turn out to be different. The excercise is performed for the di-lepton channel only, but similar conclusions should apply for the other channels. Following the reasoning in Section 6.4, the HLT trigger efficiency is taken to be 100%.

To get an idea of how close the current performance of the CMS simulation is to what has already been achieved in reality, the expected signal significance for "CMS" and "CDF" working points are compared. These are taken to differ only in the secondary vertex tagging performance – one of the most significant sources of systematic uncertainty in this measurement. Table 25 lists the tagging efficiencies of b, c, and uds-jets currently obtained for di-lepton $t\bar{t}H$ signal datasets (according to the jet-categorization as described in Section 6.1), which is to be used as the "CMS" working point. The corresponding "CDF" working point in the table uses approximate average values reported by the CDF experiment.

Table 25: Tagging efficiencies at "CMS" versus "CDF" working points

	CMS	CDF
b-tagging (%)	54	45
c-tagging (%)	16	15
uds-tagging (%)	1.6	1

The following systematic uncertainties are considered:

- Flat 3% in jet energy scale (JES)
- 10% in jet resolution
- 4% in b-jet and c-jet tagging efficiencies
- 10% in uds-jet tagging efficiencies
- 3% in Luminosity

For simplicity, all uncertainties calculated for the "CMS" working point are assumed to be those of the "CDF" working point. In this study the correlation between b-, c-jet and uds-jet tagging efficiencies is also taken into account, i.e. all tagging efficiencies were varied simultaneously (in the same direction, down).

A comparison of the expected signal significance at the two working points, both with and without systematics, is provided in Table 26. From these tables it can be seen that although the "CDF" operating point is somewhat more pessimistic than the "CMS" operating point, these two are not radically different. Moreover, the total systematic uncertainties are not overly sensitive to the differences between the prescription used in here and those detailed in Section 6.1. This lends confidence that the "CMS" simulation, and the signal and background estimates derived from it, are not widely off the mark.

Details of the effect of individual systematics, as calculated at the "CMS" working point, are available in Table 27.

D Systematic Uncertainties at "tight" Working Points

This appendix contains estimates for the systematic uncertainties evaluated at the "tight" working points for the various analyses (Tables 28 to 30). These numbers are only provided for comparison to those at the better-understood "loose" working points, and the reader should keep in mind the nontrivial statistical errors associated with them due to inadequate dataset sizes. Only a very limited number of signal and background samples could be considered here, due to limited statistics. Datasets with zero number of selected events, or related problems in estimating the systematic effect(s), are left out or indicated by blank entries in the tables.

Table 26: Impact of systematic uncertainties on expected signal significance at the "CMS" and "CDF" working points. The b-, c-jet and uds-jet tagging efficiencies were assumed to be completely correlated for these calculations.

CMS	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$t\bar{t}H$ (m_H=115 GeV/c^2)	220	0.016	1.9	0.10
$t\bar{t}H$ (m_H=120 GeV/c^2)	160	0.012	1.4	0.077
$t\bar{t}H$ (m_H=130 GeV/c^2)	110	0.0081	0.93	0.051
CDF	S	S/B	$S/\sqrt{B}$	$S/\sqrt{B+dB^2}$
$t\bar{t}H$ (m_H=115 GeV/c^2)	170	0.014	1.5	0.086
$t\bar{t}H$ (m_H=120 GeV/c^2)	130	0.010	1.1	0.063
$t\bar{t}H$ (m_H=130 GeV/c^2)	81	0.0065	0.72	0.040

Table 27: Systematic uncertainties at the "CMS" working point (all in percent). All of the b, c, and uds-tagging efficiencies were varied simultaneously down. The other systematics were varied independently. The same uncertainties as shown below are taken for the "CDF" working point.

di-lepton	statistical	JES	jet res.	bcuds-tagging	Σ
$t\bar{t}H$ (m_H=115 GeV/c^2)	12.5	1.94	4.44	9.19	10.4
$t\bar{t}H$ (m_H=120 GeV/c^2)	14.3	0.73	3.0	9.31	9.81
$t\bar{t}H$ (m_H=130 GeV/c^2)	13.6	4.64	4.47	10.9	12.7
$t\bar{t}b\bar{b}$	10.1	3.4	1.76	10.5	11.1
$t\bar{t}$1j	52.3	8.0	12.0	12.6	19.1
$t\bar{t}$2j	27.5	2.86	4.64	13.6	14.6
$t\bar{t}$3j	29.2	8.7	2.4	17.0	19.2
$t\bar{t}$4j	21.2	7.89	3.51	14.5	16.9
$t\bar{t}Z$	10.3	2.04	5.15	12.1	13.3
all backgrounds ("CMS")		5.92	5.07	13.8	15.8
all backgrounds ("CDF")		6.01	5.21	13.9	16.0

From these tables, which have to be compared to Tables 11 to 12, it is seen that the propagation of the systematic errors from loose to tight working points, as done in Section 6.1, is only correct at an approximate level. In some cases the error will be underestimated, in some cases overestimated, but the orders of magnitudes of these errors are the same.

The procedure of using the "loose" cut systematics in lieu of the above can be justified by the observation that the impact of the b-tagging and uds-(mis)tagging uncertainty is smaller at the "tight" working points, and the JES uncertainty becomes dominant. Since the "tight" working points are defined by stronger b-tagging cuts, while keeping the E_T cuts constant, no major change in the relative systematic uncertainty is expected.

References

[1] M. Dubinin *et al.*, *CompHEP - a package for evaluation of Feynman diagrams and integration over multi-particle phase space*, INP-MSU **41**, 542 (1998).

[2] W. Beenhakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira and P.M. Zerwas, *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl.Phys. **B653**, 151-203 (2003).

[3] *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, JHEP 0307:001,2003.
Multijet matrix elements and shower evolution in hadronic collisions: W B Bbar + N jets as a case study, M.L. Mangano, M. Moretti, R. Pittau, Nucl.Phys. **B632**:343-362,2002.
A new approach to multijet calculations in hadron collisions, F. Caravaglios, M.L. Mangano, M. Moretti, R. Pittau, Nucl.Phys. **B539**:215-232,1999

[4] CMS Collaboration, *The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*,CERN/LHCC 2002/26.

Table 28: Systematic uncertainties relative to selection efficiencies (all in percent) for the di-lepton $t\bar{t}H$ channel at the "tight" working point. Σ is the quadrature sum of all changes in the given row, except the statistical ones (provided for comparison only). Σ also includes the 3% uncertainty of the luminosity which is the same for all the samples.

di-lepton	statistical	JES	jet res.	bc-tagging	uds-tagging	Σ
$t\bar{t}H$ (m_H=115 GeV/c^2)	11.6	14.9	4.05	15.9	1.35	22.4
$t\bar{t}H$ (m_H=120 GeV/c^2)	13.3	14.3	4.29	15.4	1.49	21.7
$t\bar{t}H$ (m_H=130 GeV/c^2)	13.3	7.14	2.14	16.1	3.21	18.3
$t\bar{t}b\bar{b}$	7.45	2.78	1.74	15.6	1.05	16.2
$t\bar{t}2j$	50.0		25.0	13.8	10.0	30.4
$t\bar{t}4j$	57.7		3.33	13.3	6.67	15.6
$t\bar{t}Z$	15.8	17.5	6.25	12.5	3.13	22.8
all backgrounds		1.88	8.08	14.4	4.81	17.6

Table 29: Systematic uncertainties relative to selection efficiencies (all in percent) for the semi-leptonic $t\bar{t}H$ channel at the "tight" b-tagging working point with 50% of b-efficiency.

muon or electron	bc-tagging	uds-tagging
$t\bar{t}H$ (m_H=120 GeV/c^2)	11	0.5
$t\bar{t}b\bar{b}$	12.5	0.7

[5] Haifeng Pi, P. Avery, D. Green, J. Rohlf, and C. Tully, *Measurement of Missing Transverse Energy With the CMS Detector at the LHC*, CMS Note 2006/035.

[6] E. James, Y. Maravin, M. Mulders and N. Neumeister, *Muon Identification in CMS*, CMS Note 2006/010.

[7] J. D'Hondt, S. Lowette, O. Buchmuller, S. Cucciarelli, F.P. Schilling, M. Spiropulu, S.P. Mehdiabadi, D. Benedetti and L. Pape, *Fitting of Event Topologies with External Kinematic Constraints in CMS*, CMS Note 2006/0023.

[8] A. Heister, O. Kodolova, V. Konoplianikov, S. Petrushanko, J. Rohlf, C. Tully and A. Ulyanov, *Measurement of Jets with the CMS Detector at the LHC*, CMS Note 2006/036.

[9] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno and Y. Sirois, *Electron reconstruction in CMS*, CMS Note 2006/040.

[10] C.Weiser, *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*, CMS Note 2006/014.

[11] A. Santocchia *Optimization of Jet Reconstruction Settings and Parton-Level Correction for the ttH Channel*, CMS Note 2006/059

[12] A. Bocci, P. Demin, R. Ranieri, S. de Visscher, *Tagging b jets with electrons and muons at CMS*, CMS Note 2006/043.

[13] V. Drollinger, Th. Muller, D. Denegri, *Searching for Higgs Bosons in Association with Top Quark Pairs in the $H \rightarrow bb$ Decay Mode*, CMS Note 2001/054.

[14] Particle Data Group, S. Eidelman, K. G. Hayes, K. A. Olive, M. Aguilar-Benitez, C. Amsler, D. Asner, K. S. Babu, R. M. Barnett, J. Beringer et al. *Review of Particle Physics*, Physics Letters B, Volume **592**, Issues 1-4, 15 July 2004, Pages 1-5.

[15] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira and P. M. Zerwas, Phys. Rev. Lett. **87**, 201805 (2001) [arXiv:hep-ph/0107081].

[16] Michelangelo Mangano (CERN),private communication.

Table 30: Systematic uncertainties relative to selection efficiencies (all in percent) for the all-hadron $t\bar{t}H$ channel at the "tight" working point.

all-hadron	JES	bc-tagging	uds-tagging
$t\bar{t}H$ (m_H=120 GeV/c^2)	15.5	8.7	1.2
$t\bar{t}b\bar{b}$	9.9	2.3	1.5