Fermilab FERMILAB-FN-0714 February 2002

Study of low mass Higgs using $pp \to qqH$ at CMS

N. Akchurin[a], D. Green[b], S. Kunori[c], R. Vidal[b], W. Wu[b], M. T. Zeyrek[d,1)]

[a] *Texas Tech University, Lubbock, Texas, U.S.A.*
[b] *Fermi National Acceleration Laboratory, Batavia, Illinois, U.S.A.*
[c] *University of Maryland, College Park, Maryland, U.S.A.*
[d] *Middle East Technical University, Ankara, Turkey*

Abstract

We have studied the Standard Model Higgs Boson ($m_H = 120\,\mathrm{GeV}$) decaying to a WW pair in the vector boson fusion channel, with the subsequent decay of the W's to $l^+l^-\nu\bar{\nu}$. The vector boson fusion channel is characterized by two final state jets at large rapidity. The importance of the forward jet tagging to extract the signal from top quark backgrounds is emphasized. This study uses the full CMS detector simulation including the ORCA package, and the CompHEP and Pythia Monte Carlo generators.

1) **Now at Texas Tech University, Lubbock, Texas, U.S.A., contact person. e-mail: mehmet.zeyrek@ttu.edu**

1 Introduction

One of the main objectives of the CMS experiment is to discover or to rule out the Higgs particle. The LEP II has set a lower bound at around 113 GeV. Since a low Higgs mass is preferred by supersymmetry, the Higgs search up to several hundred GeV mass is presently the focus of intensive theoretical and experimental studies at the LHC. The Weak Boson Fusion (WBF) process has emerged as potentially the channel where a light or medium mass Higgs particle might be discovered first. In addition this production mechanism and subsequent decay depends only on the HWW coupling. [1], [2].

In this study, we investigated Higgs decay to a pair of W's, and the subsequent decay of these W's to $l^+l^-\nu\bar{\nu}$. We considered a light Higgs ($m_H = 120$ GeV) in low luminosity LHC running conditions (L $\sim$ 2 x $10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$). The reaction is:

$$pp \to jjH \quad (qq \to q\prime q\prime H), \quad H \to W^{(*)}W^{(*)} \to l^+l^-\nu\bar{\nu} \tag{1}$$

where leptons, $l^{\pm}$, are either electrons or muons. For a light Higgs, approximately 90% of the decays to two W's have one W-boson on the mass-shell and one off the mass-shell. Thus, at least one of the charged leptons nearly always has a small momentum which makes triggering on two leptons somewhat inefficient.

The distinct feature of this process is that the forward and backward jets tend to preserve the initial parton direction due to the absence of color exchange in the t-channel. This results in energetic forward jets with significant transverse momentum ($p_T \approx m_W/2$) and suppressed hadronic activity in the central region. We refer to these jets as the forward "tagging jets".

Nearly half of all these tagging jets are detected by the CMS HF forward calorimeters, which cover the pseudorapidity range of $3 < |\eta| < 5$, and the other half, by the CMS HE calorimeters which cover $1.5 < |\eta| < 3$. The two charged leptons, $l^{\pm}$, generally are produced centrally and are detected by the CMS electromagnetic calorimeters and muon systems which cover $|\eta| < 2.5$,

Although the production cross section for a low mass Higgs is largest in the gluon fusion process, the WBF cross section at the LHC is sizeable for all relevant Higgs masses. The WBF cross section is roughly 1/3 of the gluon fusion cross section. The gluon fusion cross section depends significantly on the Yukawa coupling of the Higgs to the top quark, while the WBF depends only on the weak boson couplings to the quarks and to the Higgs boson.

The largest background for the WBF channel is $t\bar{t}$ production [3]. The sample of $t\bar{t} + jets$, where each top decays to $l\nu b$, should be considered in three distinct phase space regions that could possibly mimic the WBF Higgs production with final state $l^+l^-\nu\bar{\nu} + jj$. They are given in order of ascending probability.

- The two b's from $t\bar{t}$ decay produce two forward jets.
- One b is identified as one of the forward jets; the other b, remaining central. The second forward jet originates from either QCD radiation or the scattered initial quarks.
- Both b's from the top decay remain central, and the two forward jets are created from either QCD radiation or the scattered initial quarks.

The suppression of the $t\bar{t} + jets$ background is achieved mainly by forward jet selection and an extra jet veto in the central region. The analysis by Kauer et al [3] showed that the $t\bar{t} + jets$ background can be reduced substantially with these two requirements. By requiring two forward jets with a large rapidity difference, the background from $t\bar{t}$ events with no extra jets is reduced to a negligible level. Similarly, the extra jet veto reduces the $t\bar{t}$ plus multijet background greatly. However, $t\bar{t}j$ still remains a significant background after all analysis cuts. Therefore, in this study, based on the analysis in [3], we concentrate only on the $t\bar{t}j$ background.

In addition to $t\bar{t}j$, there are several other backgrounds with complicated topologies that require close attention, $e.g.$ $b\bar{b}jj$, QCD production of $WWjj$, Electro-Weak production of $WWjj$, QCD production of $\tau\tau jj$, and Electro-Weak production of $\tau\tau jj$. We leave reduction of these backgrounds and the $t\bar{t}$ plus multijet production to future studies.

In this study, the signal is compared with the main $t\bar{t}$ background channel $t\bar{t}j$ where one b (or $\bar{b}$) is identified as one of the two forward jets. Requiring two forward jets and vetoing events with a tagged b or $\bar{b}$ jets brings significant suppression in the background. By rejecting events with QCD radiation (extra-jets) in the central region, the $t\bar{t}j$ background can be reduced further. Of course, since we need to reject events with an extra jet in the central region for this and other backgrounds, rejecting events with a tagged b-jet may seem superfulous. However, CMS has an excellent silicon tracking system and expects good b-tagging efficiency, and a small mis-tagging rate, so we choose to use all the experimental information which is available to understand the $t\bar{t} + jets$ background.

2 Event Generation and Reconstruction

The signal and background events are generated using CompHEP [4] parton level matrix elements. The complete matrix element calculation method is used in the generation of signal and background events because this method represents the three body final state in the process $qq \to qqH$ correctly.

CompHEP produces cross sections with the proper phase space weighting. This information is stored in a special data base, called PEVLIB. CompHEP generated events are then interfaced to PYTHIA to produce detectable final states through hadronization and decay, with Initial and Final State Radiation (ISR and FSR) turned on. Thus, we expect that radiation from the external lines of the Feynman diagrams of CompHEP are treated properly.

Table 1 shows the number of events and cross sections before the analysis cuts. The value for $\sigma \cdot Br$ represents the charged lepton ee, $\mu\mu$ and $e\mu$ final states from the W decays. For the signal, the branching ratio of $H \to WW$ is also included. The CompHEP cross sections include an initial set of jet cuts at the generation level: $p_T > 15$ GeV, $|\eta| < 5$ and $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.5$ separation between jets. These cuts represent weak detection criteria for jets within the CMS angular acceptance.

Table 1: Cross sections and the event generation

	Signal	$t\bar{t}j$
No. of events	136926	280846
σ	2104 fb	788000 fb
$\sigma \cdot Br$	12.7 fb	36800 fb

The CompHEP/PYTHIA Monte Carlo events are reconstructed by ORCA (Version 4.54) where the CMS detector is simulated. We have used the ntuple files created by ntuple-maker (Version 2.05) package where the CMS ECAL & HCAL hits, jets reconstructed by ORCA, and the generation information are available.

Because of limited computer resources and time, simulations of the CMS tracking systems and the CMS b-tagger were not used in this analysis. In addition, the charged leptons are identified only from the generator information. Only for a very low momentum charged lepton should there be any significant loss of efficiency.

3 Analysis

The basic event selection criteria came from the cuts previously proposed in [3]. Our motivation was to see if the conclusions drawn in [3] were sustained after a complete detector simulation. Events were selected by requiring two forward tagging jets and two charged leptons, each lepton with an η between the forward jets. Jets are reconstructed by the ORCA software using a cone of ΔR = 0.5. The jet multiplicity distribution for signal and background events ($p_{Tj} > 20$ GeV) is shown in Fig. 1. The peak at 2 final states for the signal (2 tag jets) is evident.

The selection criteria for the forward tagged jets, j_1 and j_2, is given in (2) and (3). We require each jet have a minimum transverse momentum of 20 GeV and a pseudo-rapidity well within the CMS calorimeters. In addition, we require a minimum ΔR between any two jets, a large separation in pseudorapidity between the two tagged jets, and that the two jets be in opposite hemispheres.

$$p_{Tj} \geq 20\,\mathrm{GeV}, \quad |\eta_j| \leq 4.5, \quad \Delta R_{jj} \geq 0.6 \tag{2}$$

$$\Delta\eta_{jets} = |\eta_{j1} - \eta_{j2}| \geq 4.2, \quad \eta_{j1}.\eta_{j2} < 0 \tag{3}$$

Fig. 2 shows the distribution of tagged jet pair candidates per event, as a fraction of the total number of events. Note that 55% of the signal events have just one combination satisfying the selection criteria (one forward and one backward jet), while only 15% of the signal events have multiple combinations. The background events show completely different characteristics. Here just 6% of the background events have a single combination satisfying the selection criteria, while about 4% have multiple combinations.

For events with multiple combinations of forward tagged jets, we use the combination with the largest rapidity difference to select the two tagged jets. For the signal events, 74% of the events have at least one combination of forward tagged jets, while for the $t\bar{t}j$ background only 10% survive. The p_T and η distributions for the designated forward tagged jets are shown in Fig. 3 and Fig. 4. Note that signal jets have a harder p_T than the background jets, on average. After this selection, many of the $t\bar{t}j$ background events still have a large amount of jet activity in the central and forward regions. Events with extra jet activity are removed by the b and extra-jet veto in the later stages of the analysis.

Leptons are identified from the generator level information rather than the simulated tracking data. Leptons are selected by a p_T cut that is implemented in a staggered fashion. For 90% of the events, one of the W's is off-shell, so a softer p_T cut on one of the charged leptons is more efficient in identifying the lepton from this virtual W. In addition, since both leptons are produced centrally, they are required to have an η between the selected tagging jets. Lepton isolation is maintained both by distance (ΔR_{jl}) between the jets and the charged leptons, and by asking for a wide separation in pseudorapidity ($\Delta\eta_{jl} > 0.6$). The lepton cuts are given in (4) and (5):

$$p_{Tl} \geq 20, 10\,\mathrm{GeV}, \quad |\eta_l| \leq 2.5, \quad \Delta R_{jl} \geq 0.7, \tag{4}$$

$$\eta_{j,min} + 0.6 < \eta_l < \eta_{j,max} - 0.6 \tag{5}$$

The maximum transverse momentum, p_T^{max} and η distributions for the charged leptons are shown in Fig. 5 and Fig. 6. The final state W's are produced almost at rest in the decay of a 120 GeV Higgs. Hence the two charged leptons and two neutrinos are emitted back-to-back in the rest frame of the Higgs. The invariant mass of the two leptons and neutrinos are almost equal and kinematically can not exceed M_H/2. The azimuthal lab angle between the two charged leptons prefers to be small due to decay dynamics. We applied the following cuts to the invariant mass and the azimuthal angle of the charged leptons :

$$m_{ll} < 60\,\mathrm{GeV}, \quad \phi_{ll} < 140^0 \tag{6}$$

The invariant mass distribution of the charged leptons for the signal and for the background is shown in Fig. 7. The charged leptons in the Higgs signal clearly prefer lower masses and a small separation in ϕ, as expected. Conversely, the background events have larger dilepton masses, on average, (see Fig. 7) and are uniformly distributed in azimuth (see Fig. 8).

The invariant mass distribution of the two forward jets is shown in Fig. 9. The jets in the signal tend to peak at larger values compared to $t\bar{t}j$ events. This feature is then used to suppress the background by imposing the following cut on the dijet mass:

$$m_{jj} > 600\,\mathrm{GeV} \tag{7}$$

Because the longitudinal momenta of the two neutrinos from the W decays cannot be determined, the invariant mass of the Higgs boson cannot be reconstructed uniquely. However, since the Higgs decay is nearly at threshold where $m_{ll} \sim m_{\nu\nu}$, a transverse mass of the decay product WW can be reconstructed using:

$$m_T(WW) = \sqrt{(E_T^{miss} + E_{T,ll})^2 - (\vec{p}_T^{\,miss} + \vec{p}_{T,ll})^2}. \tag{8}$$

The parameters $\vec{p}_T^{miss}$ and $\vec{p}_{T,ll}$ are the transverse momentum vectors of the missing momentum and the two charged leptons, respectively. For the $t\bar{t}j$ background, after the forward jet and lepton selection (2,3,4 and 5), this transverse mass reconstruction gives a broad distribution that peaks around the top quark mass, distinctively different from the signal.

After the kinematical cuts (particularly after the cut on the invariant mass of the leptons in (6)), the transverse mass of the background has a shape very similar to the signal distribution. Fig. 10 shows the invariant transverse mass distributions after the forward jet and lepton selection only, compared with the distribution after all the analysis cuts, for both signal and background. Although the signal and the background transverse mass distributions are distinctively different in shape after only jet and lepton selection cuts (specifically, before the m_{ll} and ϕ_{ll} cuts), we get a the better signal to background ratio (S/B) by including these cuts in the transverse mass reconstruction. To calculate the remaining signal and backgrounds, we use a window in the tranverse mass distribution defined by:

$$50\,\mathrm{GeV} < m_T(WW) < 140\,\mathrm{GeV} \tag{9}$$

Forward jet tagging ((2) & (3)) and lepton isolation ((4) & (5)), together with cuts on invariant mass and the azimuthal angle of charged leptons (6), and the cut on invariant mass of the two tag jets (7), goes along way toward removing the $t\bar{t}j$ background. Nevertheless, since many background events that survive these cuts have considerable extra jet activity in the central region, additional reduction of background can be achieved. In most of these background events, one of the forward jets originated from light quarks or gluons and the other came from a b-quark from t-decay. Usually, the other b-quark from the other t-decay ends up in the central rapidity region. For the signal, any additional jet activity from QCD is usually along the forward jets. Therefore, we can remove some of the background using this difference.

Since we did not simulate the CMS tracking systems or CMS b-tagger in this study because of limited computer resources and time, we identified central b-quarks in the $t\bar{t}j$ background events by matching the generated b-quarks in the central region ($p_T > 20$ GeV) with the jets found in the simulation, excluding the two forward tagged jets. The ΔR between those jets and the generated b-quarks is shown in Fig. 11. A cut of $\Delta R > 0.4$ is applied to identify these central b-quarks, and events are removed that have a central b jet. Since the efficiency of the CMS b-tagger for b-quark jets with $p_T > 20$ GeV is expected to be quite good, this should be sufficient for this analysis, although the rejection is better than we could obtain with a realistic CMS b-tagger.

Removing central b jets is very powerful in suppressing the $t\bar{t}j$ background. About 85% of the background events are found to contain the central b-jets and are rejected. The effect of this cut in shown in Table 2, just after the forward jet tagging cuts. Since we intended to eliminate these events eventually, this cut was applied at the beginning of the analysis to reduce the number of events processed.

For the signal events, this source of b-jets does not exist, but there is still some central jet activity from QCD radiation. Due to color coherence between initial and final state quarks in the signal, we expect most of the gluon radiation that does occur to be in the forward directions. In contrast, for $t\bar{t}j$ background events we expect most of the gluon radiation to be in the central region. The characteristics of the extra-jets in the signal and background events with respect to tagged jets is shown in Fig. 12. We define a pseudorapidity of an extra jet with respect to the perpendicular to the two tagged jets:

$$\eta_0 = \eta_{extra-jets} - (\eta_f + \eta_b)/2 \tag{10}$$

In Fig. 12, the extra jets in the signal and background are shown for $p_T > 20$ GeV. The extra jets for the signal are more likely to be along the direction of the forward tagged jets than the background events, as expected. Note that the majority of signal (87%) and many of the background (76%) events have no central ($|\eta_0| < 2$) jet activity. Clearly we keep twice as many (13% vs 24%) signal events compared to the $t\bar{t}j$ background. In the case of the background events, the rejected events are those that still contain QCD central radiation after the removal of central b jets.

Raising the p_T jet cut to 30 GeV from 20 GeV, reduces the number of central jets dramatically in the signal, indicating that many of the extra jets are soft. We kept the $p_T > 20$ GeV cut because a harder cut does not improve the signal-to-background ratio.

4 Results

In Table 2, we show the final cross sections after each cut. The cumulative efficiency of the selection at each step is also given as the percentage of the original sample. For the signal events, the selection efficiency after all cuts is about 18%, while it is only 0.02$\%$ for the background events. If we consider $t\bar{t}j$ to be the only background, the S/B = 33$\%$. The cross sections include the branching ratios for the ee, $\mu\mu$ and $e\mu$ final states.

Table 2: Cross sections and selection efficiencies

	Signal (fb)	$t\bar{t}j$ (fb)
$\sigma.Br$	12.7	36800
Forward Jet Tagging (Eq. 2,3)	9.40 (74.0%)	3620 (9.84%)
b jet (from top decay) veto	9.40 (74.0%)	534 (1.45%)
Lepton Isolation (Eq. 4,5)	5.25 (41.3%)	199 (0.54%)
ϕ_{ll} and m_{ll} cuts (Eq. 6)	4.17 (32.8%)	41 (0.11%)
m_{jj} (Eq. 7)	3.26 (25.7%)	13 (0.035%)
extra-jet veto $p_T >$ 20 GeV	2.84 (22.0%)	9.86 (0.027%)
$M_T(WW)$ cut (Eq. 9)	2.24 (17.6%)	6.83 (0.019%)

We have also considered the $t\bar{t}$ background separately. In this background channel, the two b's from $t\bar{t}$ decay could result in two forward tagged jets, although the probability is very low. We generated 200K $t\bar{t}$ events using CompHEP, and ran them through the same simulation, reconstruction and analysis. The probability of finding a pair of tagged jets in this sample is very low ($\sim 0.5\%$). After all the analysis cuts, we found a cross section of 0.28 fb before the extra jet veto, which is almost a factor of 46 less than $t\bar{t}j$. After the extra-jet veto, no events were left from the original sample. Although the statictics in the $t\bar{t}$ sample is limited, our analysis shows that $t\bar{t}$ is not a significant background.

There are other backgrounds for this channel besides $t\bar{t} + jets$. The most serious ones among them are $bbjj$ and $\tau\tau jj$. The cuts developed in [3] to suppress $bbjj$ and $\tau\tau jj$ signals affect the signal-to-background ratio in our analysis at the level of 10%. Therefore we have not applied them in this work.

5 Conclusions

With forward tagged jets, Higgs production through the WBF process giving a very clear signal topology. In addition, exploiting the unique kinematics of the isolated leptons between the jets further enhances the experimental signature of this channel. Further improvement in the signal to background can be achieved by the suppression of events with additional central jets.

Although the signal to background is less than one, the chosen mass, m_H = 120 GeV is the worst case since it is the lowest mass which is not yet experimentally excluded.

After the first year of operation of the LHC, assuming low luminosity running conditions, an accumulated luminosity of 60 fb^{-1} at CMS would result in 134 signal and 410 $t\bar{t}j$ events or S/$\sqrt{\text{B}} = 6.62$. The five standard deviation luminosity is 34 fb^{-1}.

As described in the analysis in [3], $t\bar{t}j$ events are about 50% of the total background to this Higgs signal. Therefore the total background cross section should be considered to be larger by about a factor of 2 than given in this analysis. All else being equal the five standard deviation luminosity is then 68 fb^{-1}.

As we mentioned earlier, the suppression of $t\bar{t}j$ background is enhanced greatly with efficient b jet vetoing. The method described here to reject the central b-jets that come from t-decay is estimated from the correlation of extra jets with b-quarks instead of tagging b's. Good tracking and good lepton identification efficiency are also essential. Until realistic lepton and b-tagging efficiencies are determined, the results of this study remain somewhat unproven. Nevertheless, this WBF channel, with its distinct characteristics is one of the most promising channels for detecting a low mass Standard Model Higgs particle with CMS at the LHC.

6 Acknowledgements

We would like to thank V. A. Ilyin for his CompHEP event generation. The efforts of the FNAL software group for the ORCA production are much appreciated. We also thank D. Rainwater for his valuable comments during this study.

References

[1] CDF and D0, Higgs Search at the Tevatron. FERMILAB-Conf-99/053-E

[2] Esther Ferrer Ribas, Standard Model Higgs at LEP. LAL 00-50, Sep. 2000

[3] N. Kauer, T. Plehn, D. Rainwater, D. Zeppenfeld, Physics Letters B 503 (2001) 113-120.

[4] A. Pukhov et. al., Preprint INP MSU 98-41/541, hep-ph/9908288.

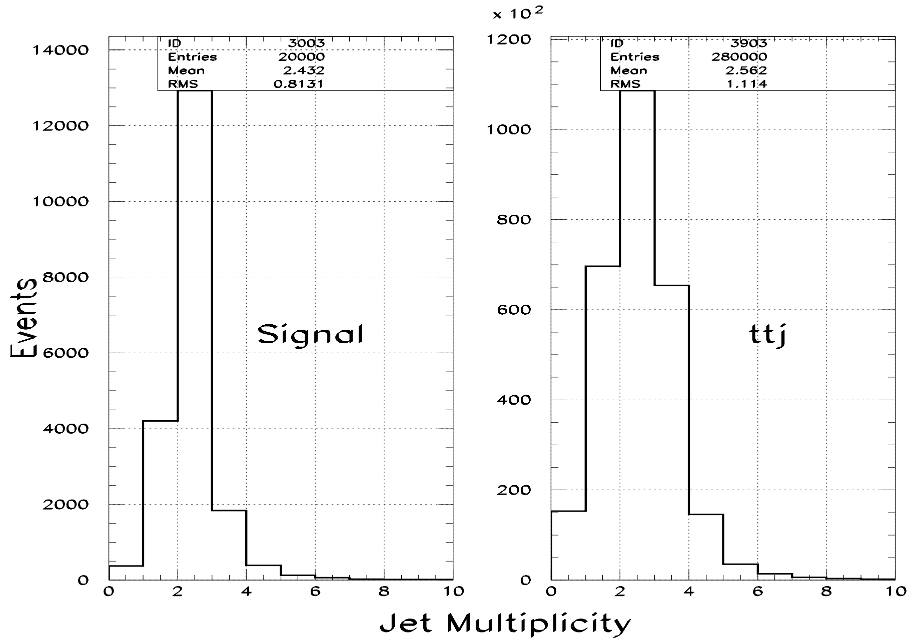

Figure 1: Jet multiplicities are shown for the signal (left) and $t\bar{t}j$ (right) background event for all jets with $p_t^j > 20$ GeV.

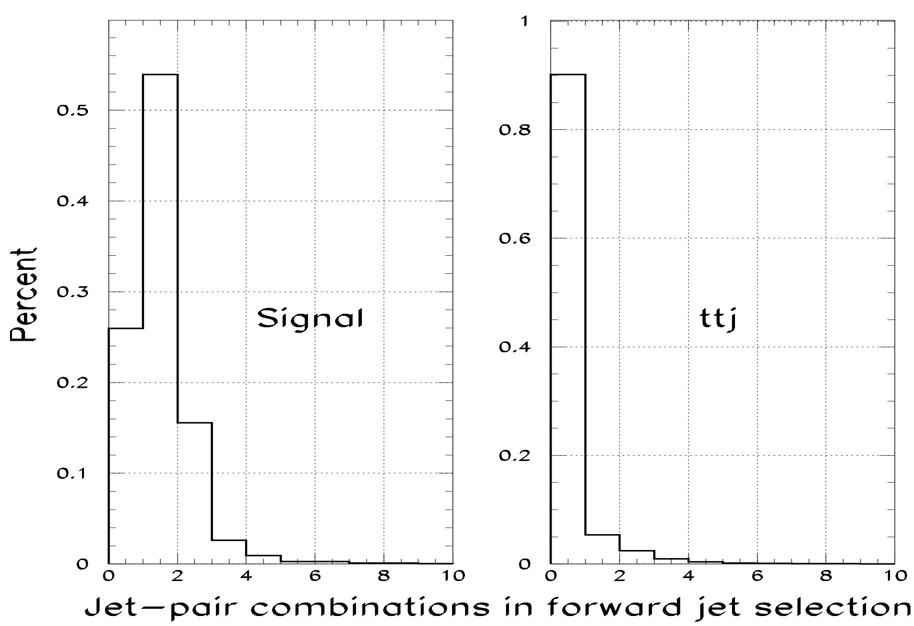

Figure 2: Numbers of found jet pairs after the tagging jet selection algorithm are shown for the signal and $t\bar{t}j$ background. More than half the time, signal events produce a single pair of tag jets. The background events do not produce a pair of tag jets 90% of the time.

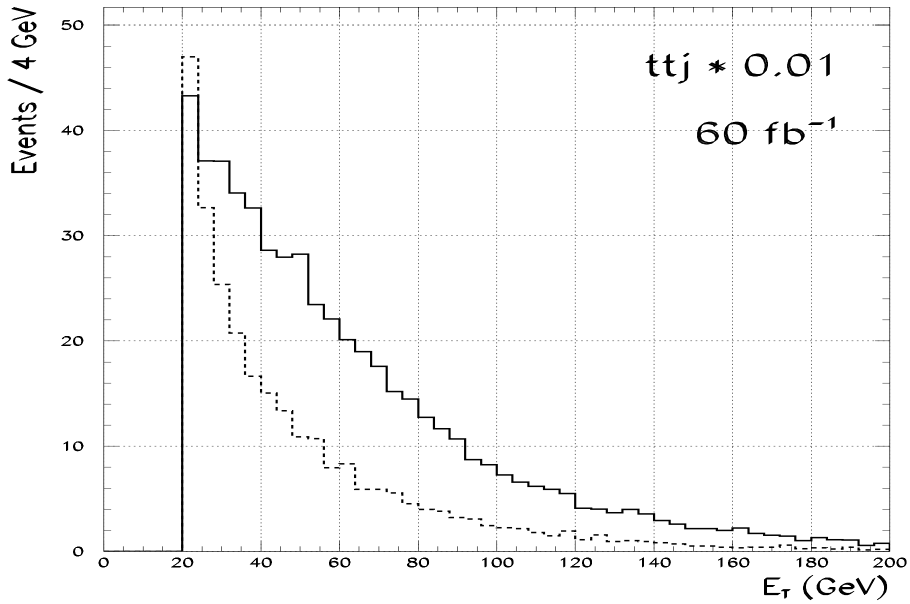

Figure 3: E_t distribution of tagged jets after jet selection cuts (Eqs. 2, 3) for the signal (solid) and the background (dotted) events. b-veto is imposed in the case of background.

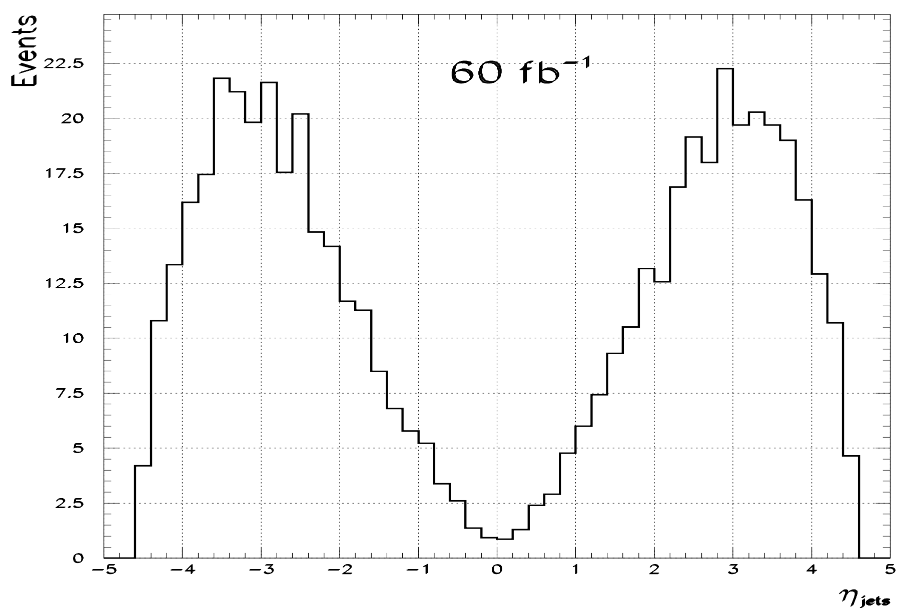

Figure 4: The pseudorapidity, η_j, distribution of tagged jets for signal events after the jet selection cuts (Eqs. 2,3) show a distinctive feature where the jets tend to go in the forward/backward direction and peak at $|\eta| \approx 3$.

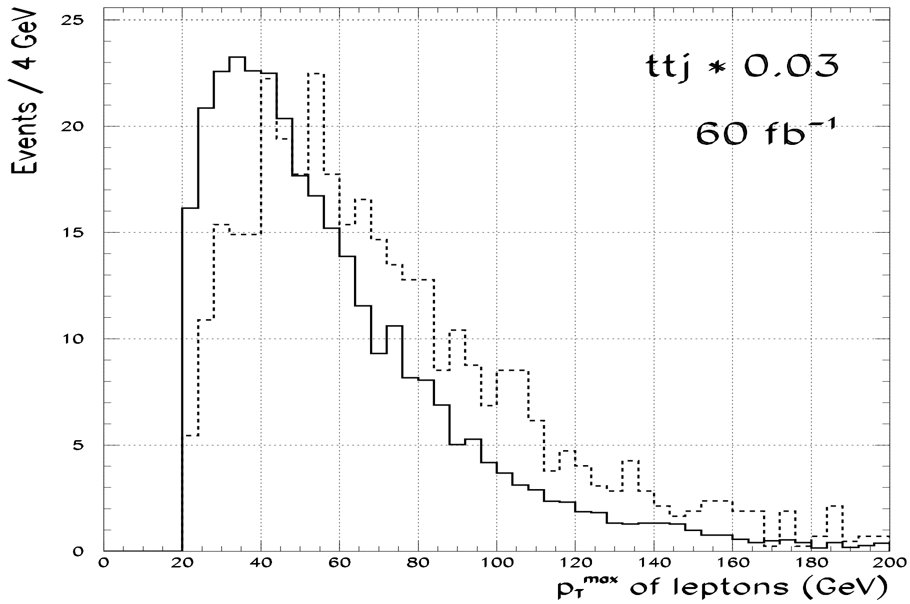

Figure 5: Maximum p_t distribution of charged leptons after jet and lepton selection cuts (Eqs. 2-5) for the signal (solid) and the background (dotted). b-veto is imposed in the case of background.

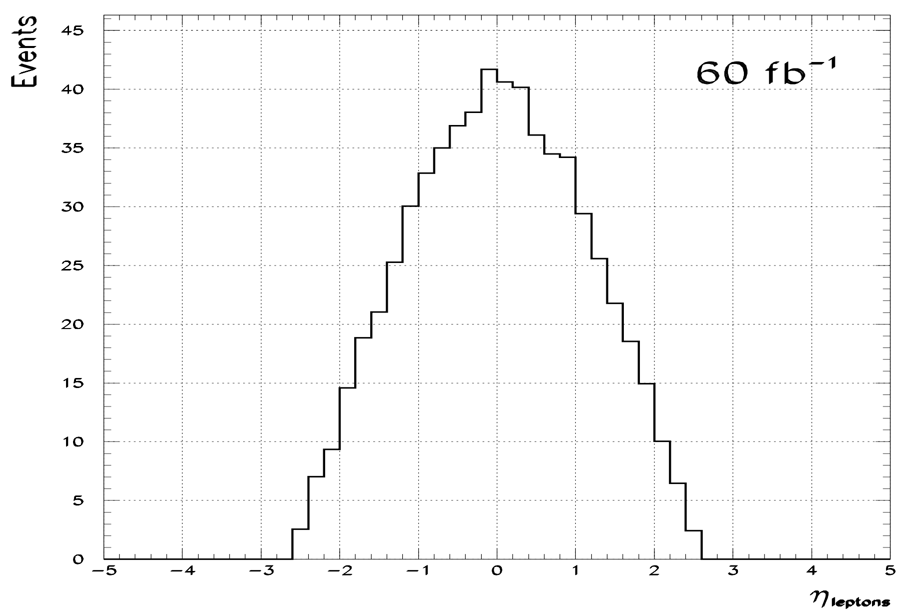

Figure 6: The pseudorapidity, η_l, distribution of charged leptons after jet and lepton selection cuts (Eqs. 2-5) for signal events.

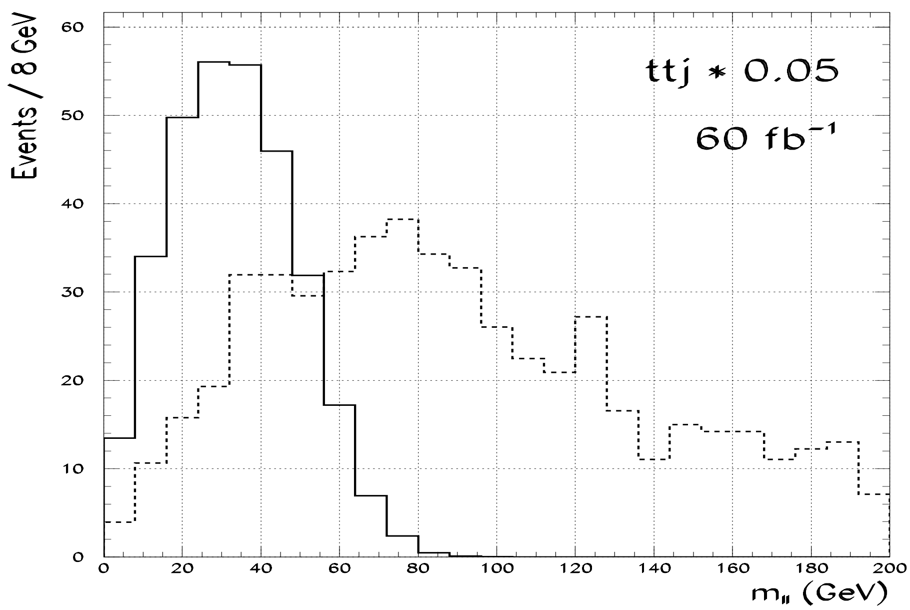

Figure 7: Invariant mass distribution of charged leptons after jet and lepton cuts (Eqs. 2-5) for the signal (solid) and the background (dotted). b-veto is imposed in the case of background. The dilepton mass cut, $m_{ll} < 60$ GeV, suppresses the background by $\approx 80\%$. (See Table 2.)

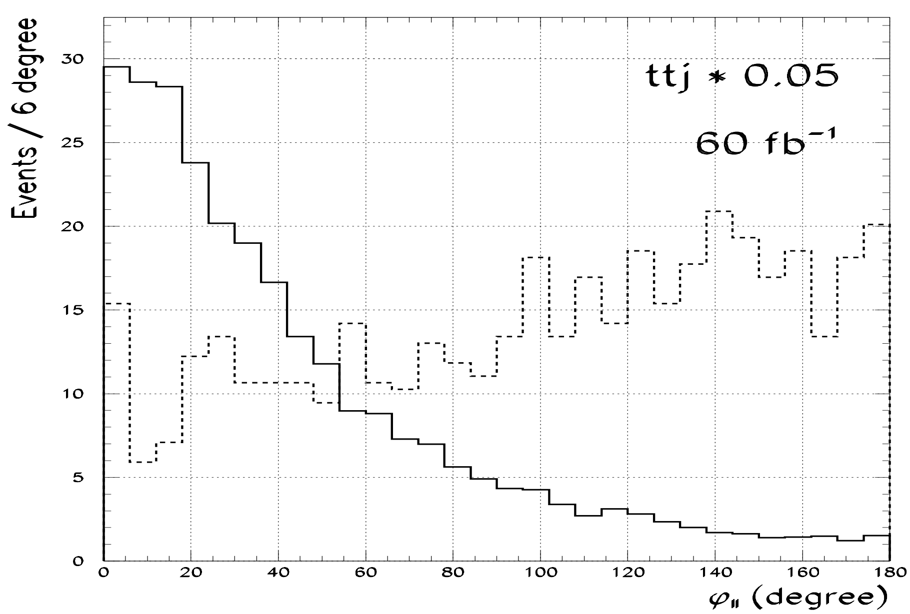

Figure 8: Azimuthal angle distribution after the same cuts as in Fig. 7 between the charged leptons. The azimuthal angular difference is small in the signal events (solid) compared to the $t\bar{t}j$ background. The two charged leptons tend to move in the same direction.

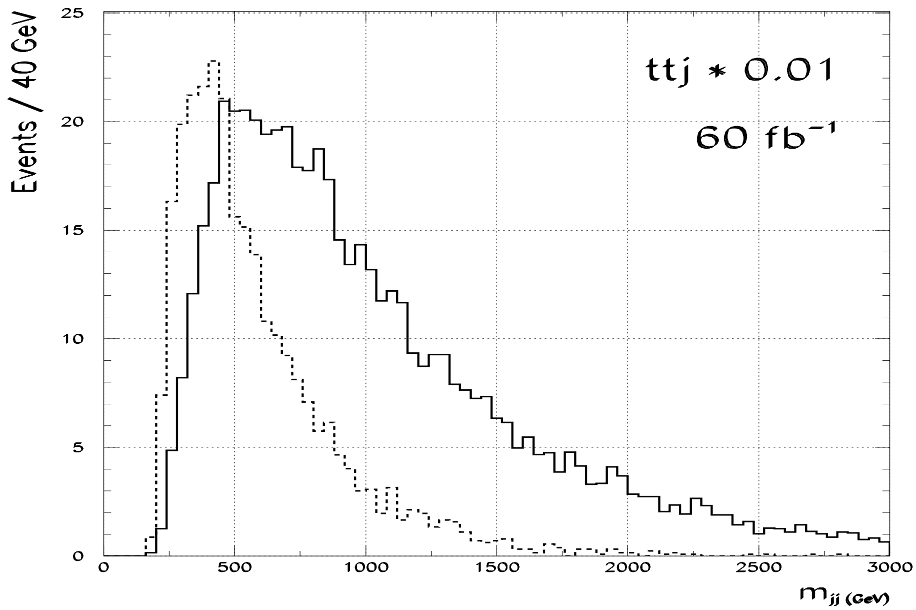

Figure 9: Invariant mass of the forward jet system after jet and lepton selection cuts cuts (Eqs. 2-5) for the signal (solid) and the background (dotted). b-veto is imposed in the case of background. m_{jj} cut at 600 GeV suppresses background by 70%.

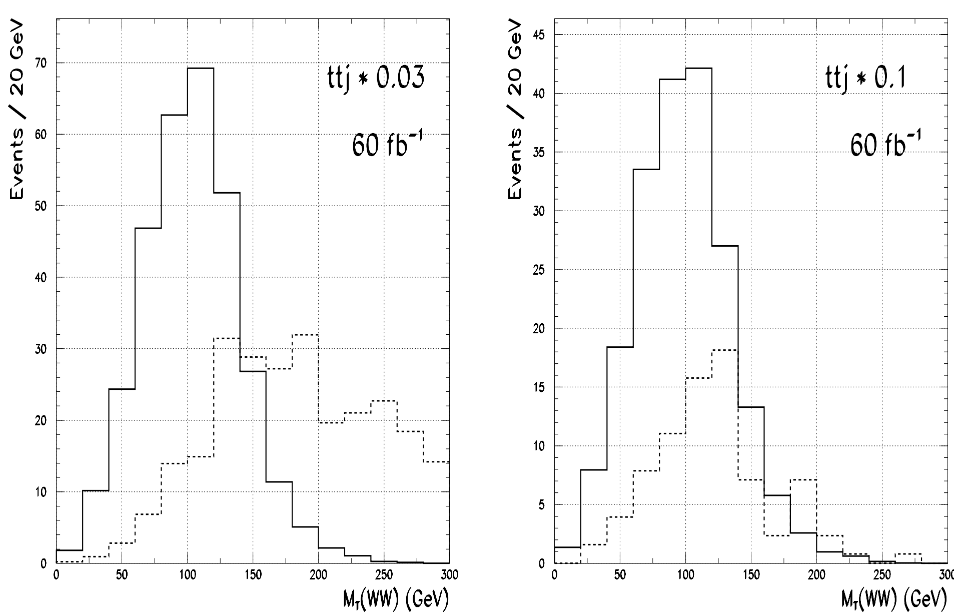

Figure 10: Transverse mass of the WW system for the signal (solid) and the background (dotted) events are shown for the jet and lepton selection cuts only in the left figure, whereas the right figure depicts the result of all cuts. (See Table 2).

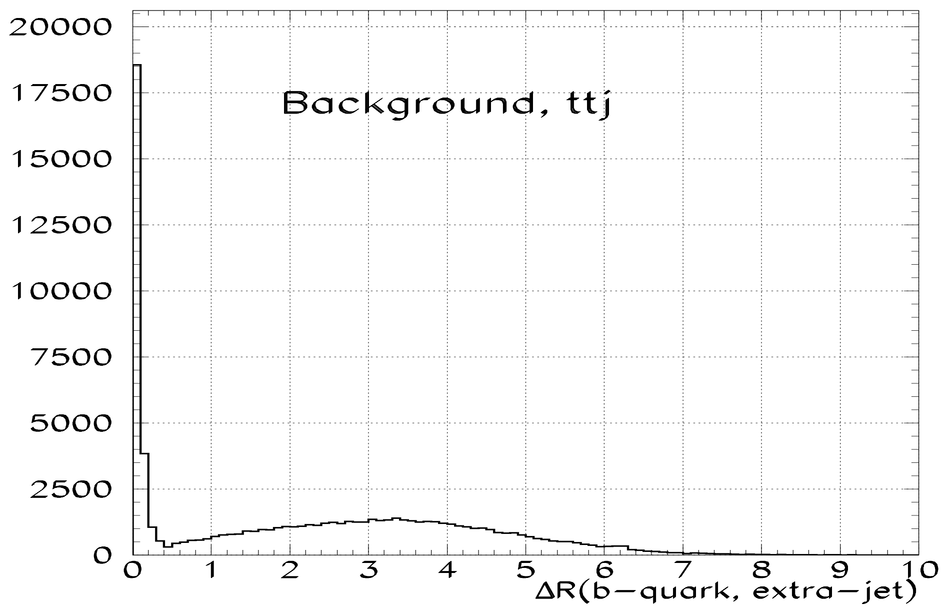

Figure 11: $\Delta R = \sqrt{(\eta_b - \eta_{extra-jet})^2 + (\phi_b - \phi_{extra-jet})^2}$ between the b-quarks in the central region ($|\eta_b| < 2.5$) and the extra-jets for the $t\bar{t}j$ background events results in two classes of events. Events with $\Delta R \geq 0.4$ are when b or $\bar{b}$ can be one of the tag jets in forward/backward directions. Events with $\Delta R < 0.4$ are rejected because neither of those can be forward/backward jets.

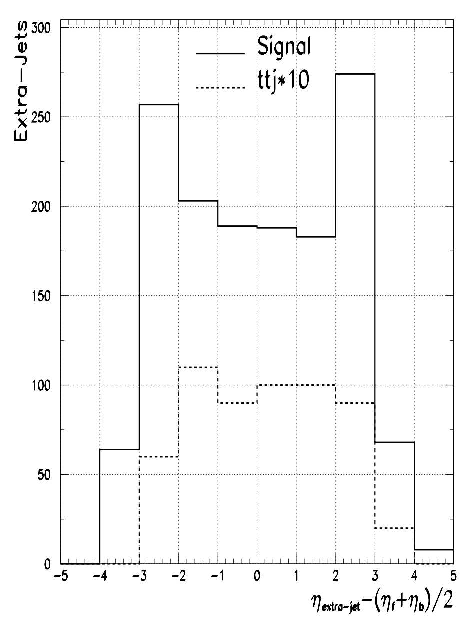

Figure 12: The pseudorapidity distributions of the extra-jets for $p_T > 20$ GeV with respect to the center of the forward/backward tagged jets, $(\eta_f + \eta_b)/2$ for the signal (solid) and the background (dashed) after all cuts.