Fermilab FERMILAB-FN-0705 August 2001

A Study of $t\bar{t}$ + Higgs at CMS

Dan Green, Kaori Maeshima, Richard Vidal, Weimin Wu*

Fermi National Accelerator Laboratory, Batavia, Illinois, U.S.A.

Shuichi Kunori

University of Maryland, College Park, Maryland, U.S.A.

Abstract

This note describes a study of the detection of a light Higgs boson produced in association with $t\bar{t}$, using the CMS detector at the LHC. We conclude that CMS will be able to isolate a sample of Higgs events with a signal to background of order 1:1 with a clear peak visible in the $b\bar{b}$ invariant mass distribution. We find that the dominant background is $t\bar{t}$ + $b\bar{b}$, coming from massive virtual gluons splitting into two b-quarks. We conclude that it would be very worthwhile to develop a b-quark tagging algorithm with high efficiency and good rejection against mistags specifically for this signal. We also find that there is a definite angular correlation between one of the t-quarks and the $b\bar{b}$ coming from the Higgs decay. This angular correlation is not present in the $t\bar{t}\,b\bar{b}$ background.

*) send comments/questions to weimin@fnal.gov

1 Introduction

The t-quark differs significantly from the other fundamental hadrons in at least a couple ways. Not only does it have a very large mass, significantly affecting the fundamental parameters of the Standard Model, but it never hadronizes. Its lifetime is so short that it never passes beyond the effective radius of the strong interaction and thus never forms an independent hadron. Most of the time, it decays immediately into a W-boson and a b-quark. The large energy available and the simplicity of the decay (affected by QCD but free from the complications of QCD bound states) permit the possibility of finding new and previously unknown particles produced in association. This is particularly true for a Standard Model Higgs since there should be a very large Yukawa coupling between the t-quark and Higgs boson. Top quarks should radiate Higgs bosons just as they do Z-bosons and photons. With its large energy and high luminosity, the LHC is an excellent place to study associated production of new particles with the t-quark. Despite the large mass of the t-quark, it is produced copiously at the LHC, unlike lower energy colliders.

We propose studying this channel for a variety of reasons. First, this can be a possible channel of discovery for a light Higgs in the mass region of 110-150 GeV. Second, this channel involves multi-jet detection, b-quark tagging, $\not\!\!E_T$ measurement, multi-jet triggering, lepton identification, and charged particle tracking, which exercises nearly all components of the CMS detector, especially the hadronic calorimeter. It offers a good opportunity to test our detector performance and triggering system, as well as our Monte-Carlo programs and analysis software. Since CMS has an excellent electro-magnetic calorimeter, allowing us to detect the two photon decay of the Higgs boson, this channel also can provide complementary and corroborating evidence for a Higgs boson decaying into b quark pairs in this mass region. [1][2][3][4].

The cross section for $t\bar{t}$ + Higgs is not large by LHC standards, and there are significant QCD backgrounds at the LHC. The main irreducible background is QCD production of $t\bar{t}$ + $b\bar{b}$, which has the same final state and very similar kinematics to $t\bar{t}$ + Higgs (Higgs decaying to $b\bar{b}$). At the LHC, there is also a large QCD production of $t\bar{t}$ plus two light quarks or gluons. The reliability of current Monte Carlo programs has not been well established for the production and kinematics of these complex backgrounds at the LHC, making it difficult to estimate convincingly the background contributions. Also, the tagging of b-quarks in events with multiple b-quarks is not well established. Not only the tagging efficiency, but also the mis-tagging rate, are critical for this study. At present, we do not have a mature b-quark tagging algorithm in CMS. The much larger backgrounds and small mis-tagging rates tax the computer resources and inhibit the generation of large data samples in this study. Therefore, we view this study as simply the beginning of a much more sophisticated analysis that will evolve and be improved over the next few years until the actual data from the LHC becomes available. Nevertheless, even at this early stage, tangible results can be achieved.

At the LHC, a light Higgs boson is produced in association with $t\bar{t}$ when two initial-state gluons make a $t\bar{t}$ pair, either via s-channel annihilation, or via the t-channel exchange of a virtual t-quark (see Figure 1). Since the Yukawa coupling of the t-quark to the Standard Model Higgs is close to one, there is a significant probability that one of the t-quarks (virtual or real) can radiate a Higgs, which then decays into two b-quarks. The t-quark decays almost exclusively to a W-boson and a b-quark. The W-boson subsequently decays into two light quarks, with a 7/9 branching ratio, or into a charged lepton and a neutrino with a 2/9 branching ratio. So, for the $(W\bar{b})(\bar{W}b)(b\bar{b})$ final state, about 2* 2/9 * 7/9 , or roughly one third of the events, contain two light quark jets and one charged lepton. Thus, the final state we wish to investigate consists of four b-quark jets, two or more non-b jets, an isolated charged lepton, and missing transverse energy coming from the neutrino.

The background from $t\bar{t}$ plus two light quarks (or gluons) produces 4 b-tagged jets in the final state only in the case where two of the light quark or gluons jets are mis-identified as b-quark jets. Although the relative cross section is much larger for this background, by applying a reliable b-tagging to the jets, requiring at least 4 b-tagged jets, and requiring at least 5 jets per event, this background can be made negligible for this study. However, the irreducible $t\bar{t}$ + $b\bar{b}$ background nearly always contains 4 genuine b-quark jets in the final state, and we must rely on other criteria to distinguish this background from the signal. Therefore, we focus mainly on the $t\bar{t}\,b\bar{b}$ background in this study.

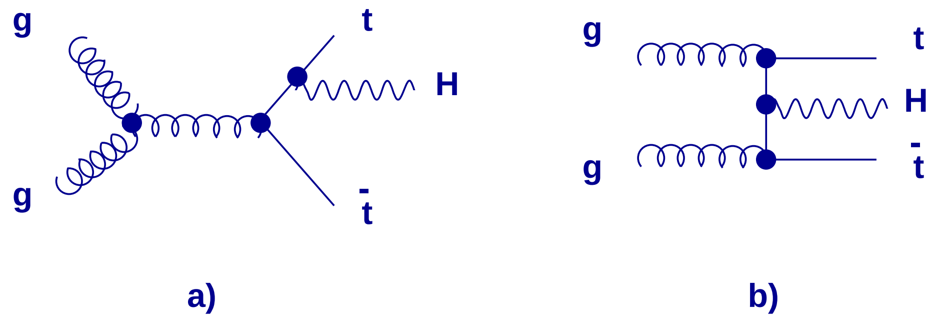

Figure 1: Representative Feynman diagrams for $t\bar{t}$ + Higgs production at the LHC: (a) s-channel gluon-gluon annihilation, (b) t-channel top quark exchange.

2 Event Generation and Simulation

All events were generated using the Monte Carlo program CompHEP V41.10 [5],[6] by V.Ilyin, and the CompHEP-PYTHIA interface program V46[6], which was developed at Moscow State University. This program interfaces CompHEP with PYTHIA 6, and we used PYTHIA 6.157. PYTHIA was used for further decay and hadronization of the CompHEP final state particles. The CTEQ41 structure functions were used in the Monte Carlo. The events from CompHEP and PYTHIA were simulated using the program CMS120, which is a GEANT-based simulation of the CMS detector at the LHC. The production cross sections calculated in CompHEP and the number of events generated for the signal and backgrounds are given in Table 1. In order to reduce the number of events to be fully simulated, we have attempted to determine which jets most likely would be selected by the final CMS b-tagging algorithm. The final algorithm will use the full potential of the 3-D tracks found by the CMS Silicon Tracking detector. Here, we apply a simplified algorithm before the events are passed to the detector simulation, and remove any events where there are less than 4 jets "pre-tagged". This "pre-tagging" algorithm is discussed below. By pre-tagging the events, we can achieve nearly two order of magnitude reduction in the number of events to be simulated, as can be seen in Table 1. The Higgs mass was chosen to be 120 GeV in this study.

process	$t\bar{t}$ + Higgs	$t\bar{t}$ + bb	$t\bar{t}$ + jj	$t\bar{t}$ + Z
Cross section (pb)	0.784	3.28	507	0.646
Number of events generated	600K	600K	1,000K	600k
Luminosity (fb^{-1})	765.3	182.9	1.97	928.7
Number of events after pre-tagging	6600	6035	125	1558

Table 1: The number of events generated for the signal and backgrounds.

From Table 1, it appears as if a large amount of LHC running time is required to accumulate enough events to establish a clear signal. An integrated luminosity of approximately 700 fb^{-1} at the LHC would take a few years of running. However, this is mainly due to the low efficiency assumed for the pre-tagging algorithm. Since the efficiency of the final CMS b-tagging algorithm is hoped to be nearer 60%[10] the equivalent data sample should require only about $90\,fb^{-1}$ when the final algorithm is used (*i.e.* about one year of running at the LHC). Thus, this study uses effectively the "worst-case scenario", where the b-tagging rate is no better than average, and the mis-tagging rate is relatively large, compared to what is expected in CMS.

In the actual "pre-tagging" algorithm [7], the charged particles and their vertices are smeared by the expected resolution of the CMS Silicon Tracking system. This is simpler and faster than using a full simulation and reconstruction of the CMS tracking, which would be prohibitive for the large number of events needed in this study. Each particle is propagated to the CMS calorimeter, and its full energy is assigned to the calorimeter tower it strikes. A energy-clustering algorithm is then executed to find all "jets" with $E_T > 20\,GeV$. After smearing, if any vertex (in a cone of radius 0.5 in $\eta - \phi$ space around each "jet" direction) is displaced by $1.5\,\sigma$ from the primary vertex, then this jet is considered "pre-tagged". We require at least 4 "pre-tagged's" for the event to proceed to full simulation and analysis. A minimum charged particle momentum of 2 GeV is required to be included in a vertex, and as few as 2 such tracks defines a vertex.

In addition to the "pre-tagging" selection, other kinematic cuts were applied to the backgrounds to reduce the data sample before simulation. For all samples, we set $\alpha_s(M_Z) = 0.132$, $M_{top} = 175\,GeV$, and $M_b = 4.62\,GeV$. For the generation of $t\bar{t}$+ Higgs events, we used $Q^2 = M_H^2$, for the $t\bar{t}$+ $b\bar{b}$ events, $Q^2 = M_t^2 + ((P_T)_t^2 + (P_T)_{\bar{t}}^2 + (P_T)_b^2 + (P_T)_{\bar{b}}^2)/4$, and for the $t\bar{t}$+ jj events, $Q^2 = M_t^2 + ((P_T)_t^2 + (P_T)_{\bar{t}}^2 + (P_T)_{j1}^2 + (P_T)_{j2}^2)/4$. For the $t\bar{t}$+ $b\bar{b}$ events, we required $P_T^b > 15\,GeV$, $P_T^{\bar{b}} > 15\,GeV$, $|\eta_b| < 3.0$, $|\eta_{\bar{b}}| < 3.0$, and $\Delta R(b,\bar{b}) > 0.3$, where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. Similar kinematic cuts were applied to the $t\bar{t}$+ jj events.

There are other alternatives to pre-tagging jets that might be used to reduce the number of events simulated. One might require at least one lepton in the event, which reduces the events by about a third, or make selective mass cuts. None of these are as effective as pre-tagging the jets, and then selecting events with multiple tags. The $t\bar{t}$+ Higgs signal (as well as the $t\bar{t}$+ $b\bar{b}$ background) has four real b-quarks in its final state, and this can be readily exploited to reduce the number of events.

The CMS calorimeter was simulated using GEANT3 within the framework of the CMS general simulation program CMSIM (version 120). The electro-magnetic calorimeter (ECAL) consists of $23\,\mathrm{cm}\ \mathrm{PbWO_4}$ crystals followed by the support structure. The hadron calorimeter (HCAL) consists of layers of copper absorber, 5.0 cm thick (7.9 cm in the endcap) separated by air gaps containing the scintillator packages. Outside the solenoid, two "tail catcher"(HO) layers are implemented in the muon system. The minimum thickness of the calorimeter inside the solenoid was 6.9λ at $\eta = 0$. The detector model follows the TDR-2 design described in the HCAL Technical Design Report.

Hadronic showers were simulated using the GHISHA software, with energy cutoff values of 1 MeV for electrons and photons, and 10 MeV for hadrons. The energy deposited in the crystals and scintillators was in an array of cells of size $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 2.262$ and 0.174×0.174 for $|\eta| > 2.262$.

Although the $t\bar{t}$+ Z background is not large, we generated 600,000 $t\bar{t}$+Z events by the same methods used to produce the $t\bar{t}$+H events in order to test our analysis. The cross section is about 0.646 pb, which is about 80 % of the cross section for $t\bar{t}$+H. The basic Feynman diagrams for $t\bar{t}$+H and $t\bar{t}$+Z are similar (see Figure 1), and the coupling constant are quite comparable. After applying pre-b-tagging, we got 1558 events. The branching ratio of Z decay into $b\bar{b}$ is much smaller (about 15 %) than that of the Higgs, so that we find fewer $b\bar{b}$'s in the $t\bar{t}$+Z events compared to $t\bar{t}$+H.

In practice, the real data sample where the Z decays into di-leptons can be used to set the cross section scale using this known process. That data sample can be used to test much of the analysis chain with the Higgs decay into b pairs replaced by Z decay into lepton pairs.

3 Event Selection and Analysis

3.1 Jet Finding and Reconstruction

Jets were reconstructed using a simple cone algorithm. The ECAL and HCAL responses were first summed into towers using the transverse segmentation given above. Then the tower with the largest E_T was used as the initial seed for a jet, and the energies within a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.5$ around the seed tower were added to the jet. The jet axis was recalculated after the addition of each new tower. Once all the towers within the cone radius were merged, those towers were removed from the list to be clustered, and the procedure was repeated until all the towers had been associated with a jet. The jet mass was calculated from transverse energy spread within the jet radius, and only jets with $E_T > 20\,GeV$ were used. Figure 2 shows the jet multiplicity for the $t\bar{t}$ + Higgs signal, and for the $t\bar{t}$ + $b\bar{b}$ background samples, before and after the removal of EM clusters from electrons.

In $t\bar{t}$ + Higgs and $t\bar{t}$ + $b\bar{b}$ events, the multiplicity of jets is larger than one might expect since jets can come not only from the decay products of the t-quarks and Higgs, but also from radiated gluons and other light quarks from the initial or final state particles. One can try to reduce the number of extraneous jets by adjusting the cone size and E_T cut. However, this often compromises the jet energy resolution (and thus the mass resolution of the Higgs). In contrast, mis-identification of the jets from the Higgs decay has a similar bad effect. Thus, the cone size and E_T cut were chosen to eliminate irrelevant jets as much as possible, without compromising the energy resolution.

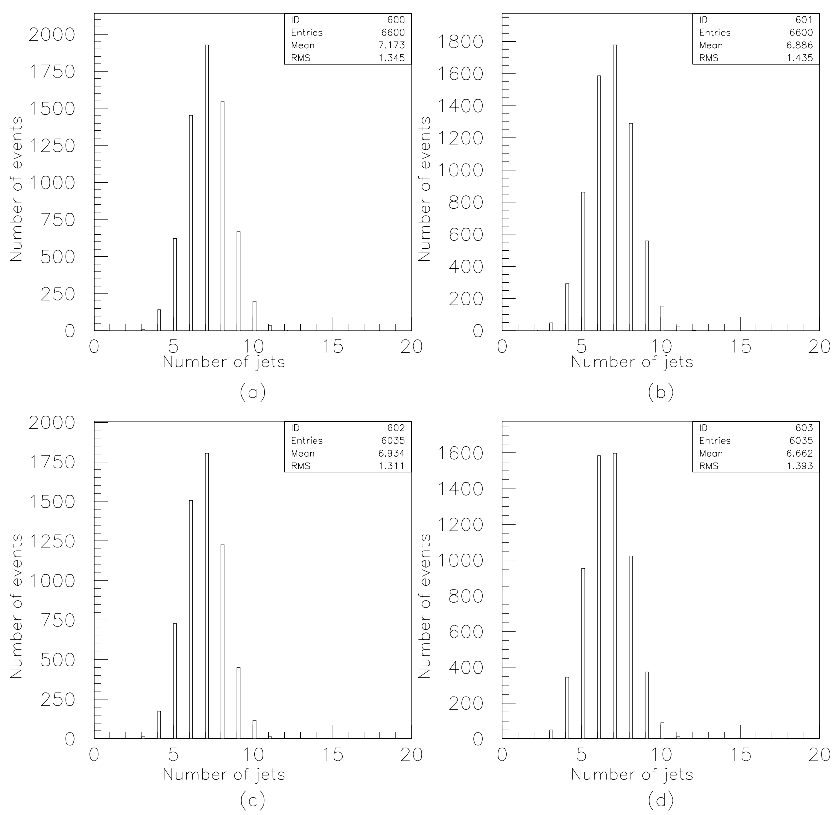

Figure 2: Jet multiplicity: (a) $t\bar{t}$+H all, (b) $t\bar{t}$+H with electron clusters removed, (c)$t\bar{t}$+$b\bar{b}$ all (d) $t\bar{t}$+$b\bar{b}$ with electron clusters removed,

We have used a jet E_T correction routine based on simulated QCD jets from CMSIM version 116,[8] which is the best available jet Et correction at this time. Note that this Et correction has over-corrected the jet E_T, mainly due to the differences between version 116 and 120 of CMSIM. (Also, the b jet should have a slightly different correction compared to QCD jets.) Since the mass resolution is improved significantly by the Et correction, and since we know that the jet Et is over-corrected, we have scaled the reconstructed mass of the top, W and $b\bar{b}$ by an overall normalization such that the Z mass is 90 GeV.

The CMS trigger and the data analysis for this sample is expected to require one isolated lepton, $\not{E}_T$, and one jet. Figure 3 shows the E_T of the leading jet for the signal and background samples. By requiring at least one jet with $E_T > 100\,GeV$, the trigger rate should be small enough to be accommodated by the CMS DAQ system and reconstruction software.

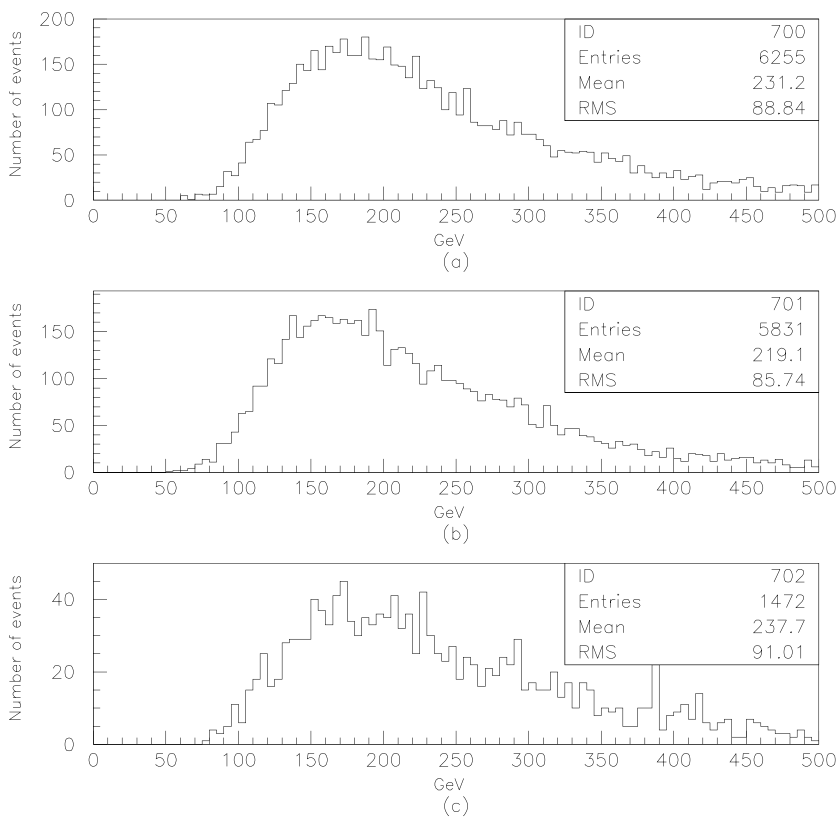

Figure 3: E_T (GeV) of leading jet for (a) $t\bar{t}$+H, (b) $t\bar{t}\,b\bar{b}$, and (c) $t\bar{t}$+Z event samples

3.2 Identification of the Lepton

In order to satisfy the expected trigger and simplify the top reconstruction, we require that each event has only one charged lepton. That is, one W decays into two jets, and the other W decays into a lepton and a neutrino. (About 35 % of the events satisfy this topology.) In this type of analysis, the charged lepton identification usually depends critically on the associated track. As we stated previously, since we need to generate and simulate a considerable number of events to study the signal and backgrounds, it is prohibitive in this first attempt at an analysis to try to use the tracking simulation and reconstruction for charged leptons (or jets). Therefore, we have chosen to forgo a formal set of criteria for charged lepton identification in favor of a minimum set, and take the momenta of charged leptons directly from the Monte Carlo generator. This is probably adequate in this case since the charged leptons from the signal and background have similar kinematics and identification probabilities. We expect that even a final, very selective set of criteria for charged lepton identification will not seriously degrade the statistical and systematic significance of the signal despite the smaller efficiency. However, a more realistic set of charged lepton criteria will require a proportionately larger luminosity to achieve the same statistical significance for the signal as found in this study.

In order to satisfy a minimum set of trigger and analysis selection criteria, we require that the charged leptons have $P_T > 20\,\mathrm{GeV/c}$ and $|\eta| < 2.5$. Furthermore, we require each muon or electron to be loosely isolated. By loosely isolated, we mean the additional transverse energy in a cone of radius 0.5 in $\eta - \phi$ space around each lepton direction be less than 50 % of the E_T of the lepton:

$$(E_T^C - E_T^l)/E_T^C < 0.5$$

We require one and only one such isolated lepton in each event. This isolation criteria is much looser than is normally used for lepton identification (usually about 10 %). In this complicated topology containing many fragmenting quarks, there are many sources of photons[9] that make lepton identification using the normal criteria inefficient (only about 70 % of the events satisfy the normal isolation criteria for electrons). With this much looser cut, about 95 % of the events are selected and less than 5 % of the leptons did not come from the W decay. Of course, with this looser criteria, the energy of the electrons as measured in the calorimeter may be less accurate, and the momentum from the tracking may be needed instead (as well as complicating any E/P criteria that may be required if these cuts are not enough). However, we believe this looser cut, coupled with fits to the W and t-quark mass, is sufficient to select the signal event with reasonable high efficiency.

3.3 Reconstructing the W-bosons and t-quarks

To reduce the combinatorics in identifying the Higgs boson via the $b\bar{b}$ mass distribution, we attempted to eliminate two of the b-tagged jets by associating them with the t-quark decays. Each event contains at least two jets that are not b-tagged and four b-tagged jets. We first try to find the two jets coming from the decay of W-boson and then the associated b-quarks coming from the t-quark decays using the invariant mass distributions.

From all the jets in the event that were not b-tagged, we associate the two jets whose invariant mass is closest to the W-boson mass with the t-quark that decayed purely hadronically. Figure 4a shows the jet-jet mass for the closest pair. Because some b-quark jets are not tagged, and some quark jets from actual W decays are mis-tagged as b-quark jets or lost, there is an undelying contribution from incorrect identification of one or both of these jets as coming from the W-boson decay. However, these mis-identified pairs are not likely to reconstruct the correct top mass when paired with a b-tagged jet.

To reconstruct the t-quark decaying semi-leptonically, we use the charged lepton P_T and $\not{E}_T$ to reconstruct the W-boson decaying leptonically. We fit the lepton and $\not{E}_T$ to the W-boson mass[10], picking the neutrino P_z that gives the smallest longitudinal momentum to the W-boson.

The two W-boson candidates are then paired with the b-tagged jets in the event, and we pick the jet combination which has the smallest deviation from the top mass for both the “hadronic” t-quark and “leptonic” t-quark combined. Figure 4b shows the reconstructed mass for the t-quark that decayed purely hadronically. Figure 4c shows the recontructed mass for the t-quark that decayed semi-leptonically. After removing the b-tagged jets used to form the t-quarks, the remaining b-tagged jets are used to form the $b\bar{b}$ invariant mass distribution, shown in Figure 4d for the $t\bar{t}$ + Higgs signal sample.

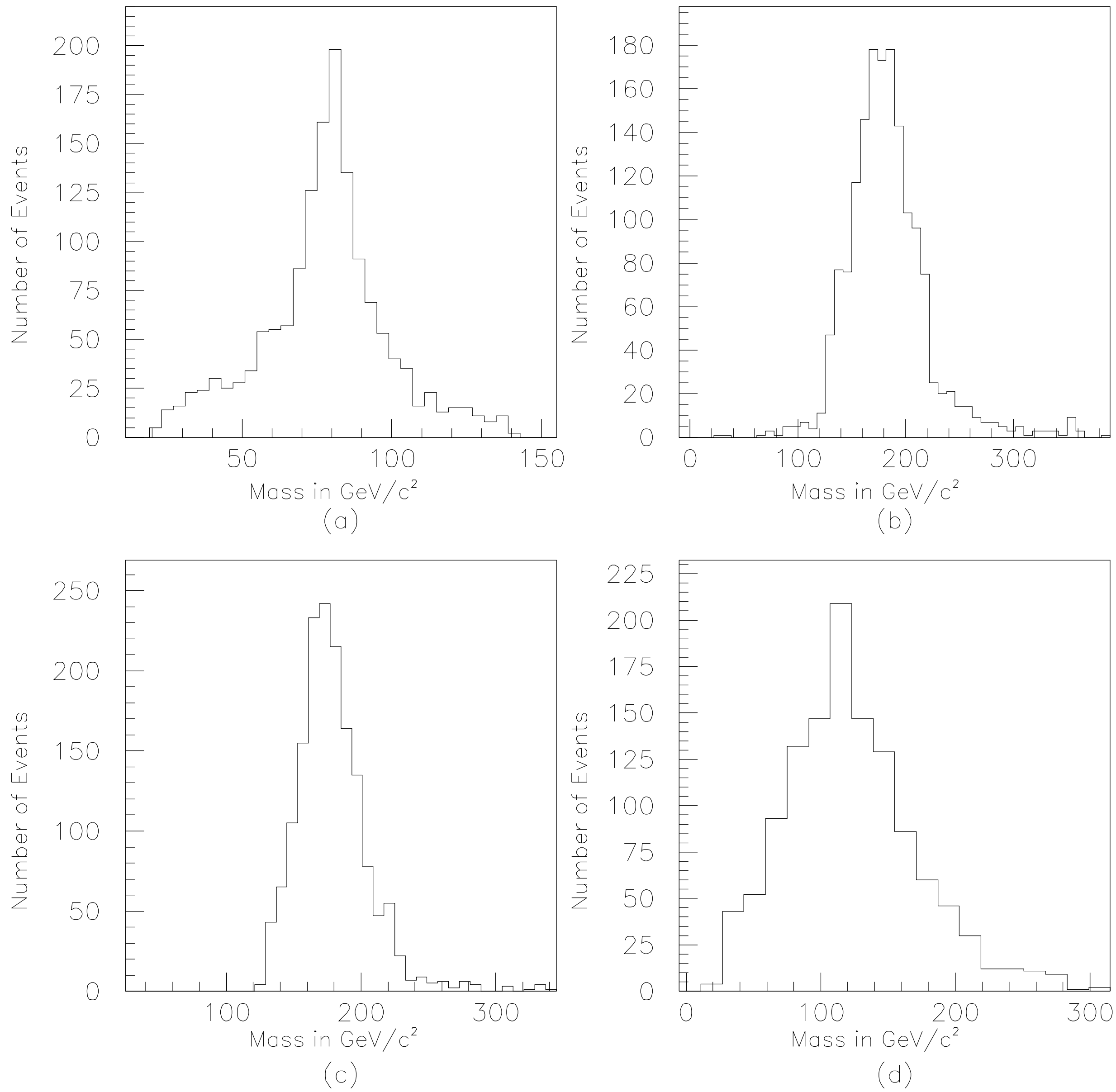

Figure 4: Reconstructed mass (a) $q\bar{q}$, (b)$q\bar{q}b$, (c)$l\nu b$ and (d) $b\bar{b}$ for the $t\bar{t}$+H signal sample

4 Further Background Suppression

Even with all the requirements applied so far (including 4 b-tagged jets, and the mass constraints), the ratio of signal to background is only 1: 3.83: 0.19 for the $t\bar{t}$ + H, $t\bar{t}$ + $b\bar{b}$ and $t\bar{t}$ + Z samples. We still need to reduce the $t\bar{t}$ + $b\bar{b}$ background further to find a clear signal for the Higgs boson. Although the $t\bar{t}$ + jj background is quite large, since we require 4 b-tagged jets, and a minimum of 5 jets, these criteria reduce the $t\bar{t}$ + jj background given in Table 1 to a negligible level. Therefore, we are left with only the $t\bar{t}$ + $b\bar{b}$ as the only significant background to consider.

One can consider the Higgs to be coming from the decay of a virtual t^* that decays into a t-quark (on mass-shell) and the Higgs. Unfortunately, this is generally only true for the s-channel production (see Figure 1), and not for the t-channel exchange. Since these diagrams are inseparable, it is not clear how much contribution comes from the s-channel production relative to the t-channel exchange. In addition, it is not always obvious to which of the two reconstructed t-quarks we should associate the Higgs to reconstuct the t^*. The $b\bar{b}$ in the background $t\bar{t}$ + $b\bar{b}$ generally comes from the decay of a virtual gluon, and its association with either of the two t-quarks is less strong since any colored intermediate particle could have radiated the virtual gluon. Since the $b\bar{b}$ association with one of the t-quarks is much stronger for the $t\bar{t}$ + Higgs signal, and since the spins of these particles are different, we might be able to exploit the angular and momentum distributions of these intermediate states to further discriminate the signal from the $t\bar{t}$ + $b\bar{b}$ background. To explore the possibility of using this information to further reduce the background without loosing too much signal, we have investigated the following parameters:

a) the invariant mass of the t^*, where we combine the $b\bar{b}$ with the reconstructed t-quark nearest in angle to make the t^*;

b) the angle between the $b\bar{b}$ and the reconstructed t-quark nearest in angle;

c) the rapidity difference Δy between the $b\bar{b}$ and the reconstructed t-quark nearest in angle;

d) the ΔR between the $b\bar{b}$ and the reconstructed t-quark nearest in angle, calculated using rapidity;

e) the ΔR between the $b\bar{b}$ and the reconstructed t-quark nearest in angle, calculated using pseudo-rapidity;

f) the angle Θ^* between the b-quark and the $b\bar{b}$ direction in the $b\bar{b}$ rest frame.

Note that the parameters in b) through e) are similar, and try to use the same information inherent in the ”bremstrallung” nature of Higgs radiation from a t-quark. The last parameter attempts to exploit the different angular distributions of the spin-0 Higgs and the spin-1 virtual gluon. However, as is evident in Figure 5f, we observe little usable difference between the two in our samples. Figures 5a-f shows the distributions for each parameter for the signal and background samples. While the t^* mass shows a distinct difference, and might be usefull in further discriminating the signal from background, it depends weakly on the mass of the Higgs boson. After carefull consideration, we prefered the Δy parameter since it reduces the $t\bar{t}$ + $b\bar{b}$ background sufficiently without eliminating too much of the signal, and has little mass dependence.

Figure 6 shows the Δy distribution including the $t\bar{t}$ + Z test sample (normalized to the same number of $t\bar{t}$ + H events). One can see that $t\bar{t}$ + H and $t\bar{t}$ + Z have same shape (both ”bremstrallung-like”) while $t\bar{t}$ + $b\bar{b}$ is more widely distributed. If we make a cut on $|\Delta y| < 1$, we can significantly reduce the $t\bar{t}$ + $b\bar{b}$ background without eliminating too much of the signal.

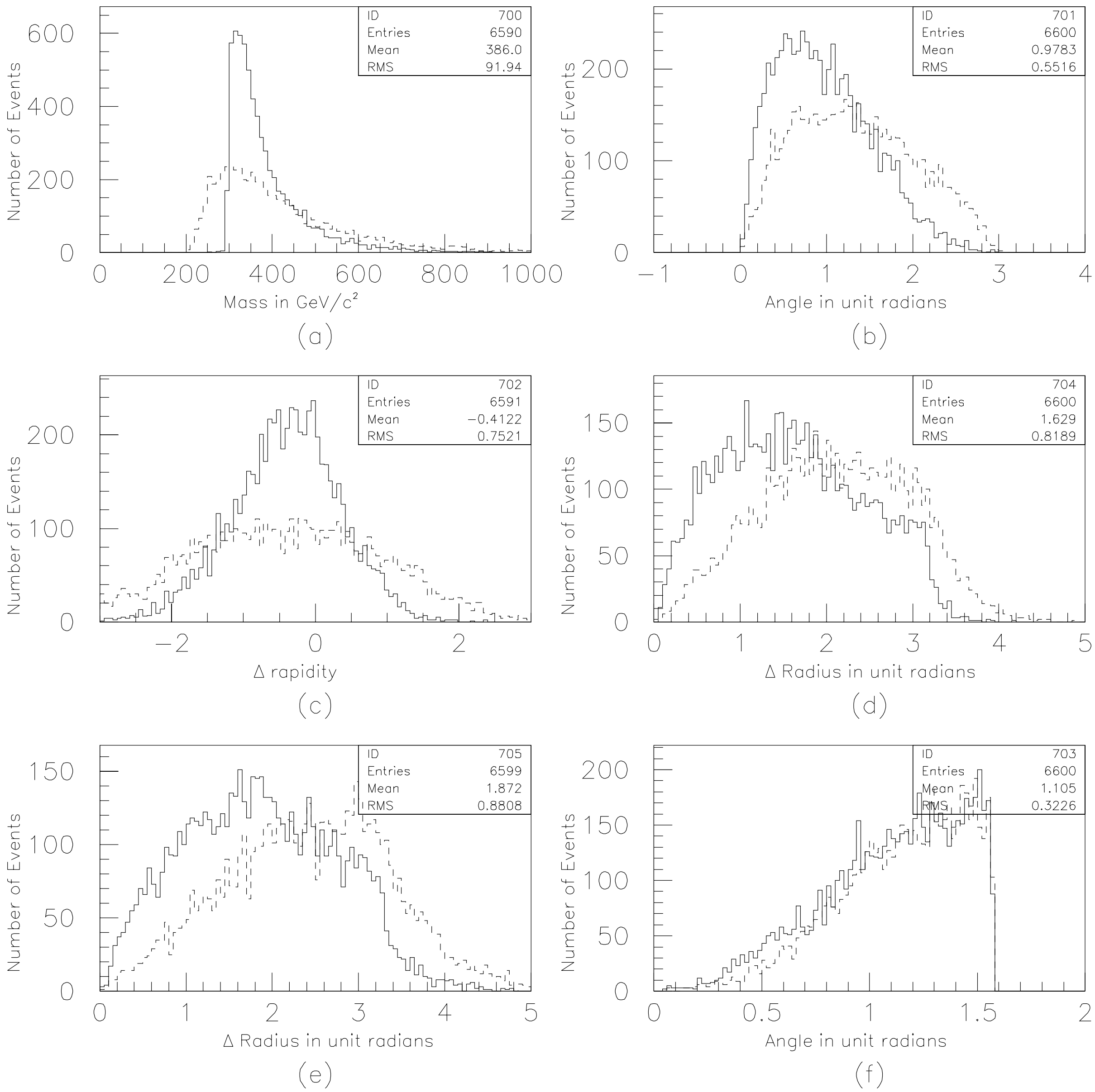

Figure 5: Distribution of the kinematics variables for $t\bar{t}$ + H (solid) and $t\bar{t} + b\bar{b}$ (dash) events (see text for the definitions of these variables).

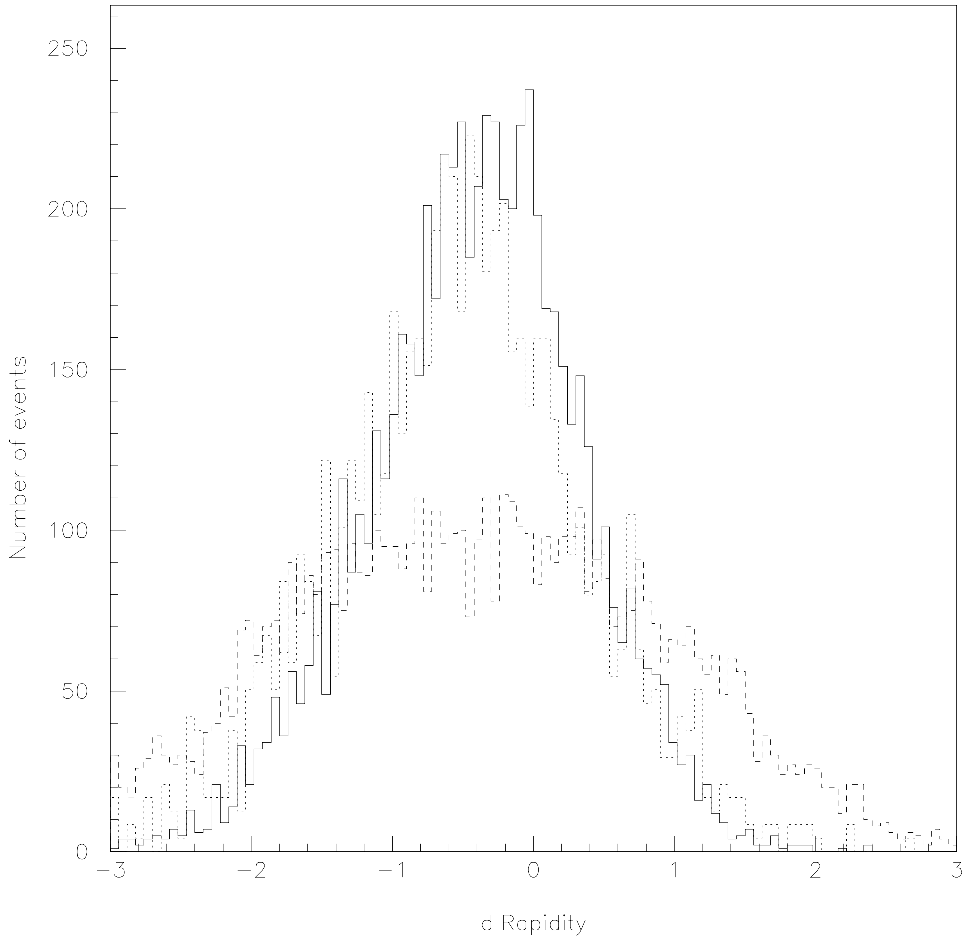

Figure 6: Δy distribution for $t\bar{t}$+H(solid), $t\bar{t}\,b\bar{b}$(dash) and $t\bar{t}$+Z(dot)

5 Results

5.1 Expectations before Simulation

We can examine what we would expect if we had a perfect CMS detector by using the energies and momenta of particles before simulation in the actual CMS detector. After extrapolation to the CMS calorimeters, jets are found by a simple clustering algorithm that assumes the total energy is deposited in the incident tower for each particle. All the previously specified analysis criteria are applied at this level. We identify which jets resulted from b-quark framentation by determining which jets align with initial b-quark direction (*i.e.* within a cone of $\Delta R = 0.5$), and consider each of these jets as b-tagged. The two b-tagged jets associated with the t-quark decays are identified by the methods specified previously (*i.e* the mass combination closes to the t-quark mass), leaving the other b-tagged jets as candidates for the Higgs decay.

Figure 7a shows the $b\bar{b}$ invariant mass for the sum of the $t\bar{t}$+ H, $t\bar{t}$+ $b\bar{b}$, and $t\bar{t}$+ Z data samples normalized to the same luminosity, where the $|\Delta y| < 1.0$ requirement has not been applied. A peak in the mass spectrum is clearly evident. If we then require $|\Delta y| < 1.0$, Figure 7b shows a significant inprovement in discriminating the Higgs mass peak relative to the background events. The resonance appears at a mass of 120 GeV as expected. This gives us confidence that the Δy requirement should actually improve the situation.

5.2 Results after Simulation

Using the CMS dectector simulation and the event selection described earlier, we plot the $b\bar{b}$ invariant mass in Figure 7c for the $t\bar{t}$ + H, $t\bar{t}$ + $b\bar{b}$, and $t\bar{t}$ + Z data samples normalized to the same luminosity, where the $|\Delta y| < 1.0$ requirement has not been applied. The number of events surviving after simulation has been reduced, but a clear peak is still evident. If we then require $|\Delta y| < 1.0$, Figure 7d shows the same improvement in discriminating the Higgs mass peak relative to the background events as was evident before simulation (if not slightly better).

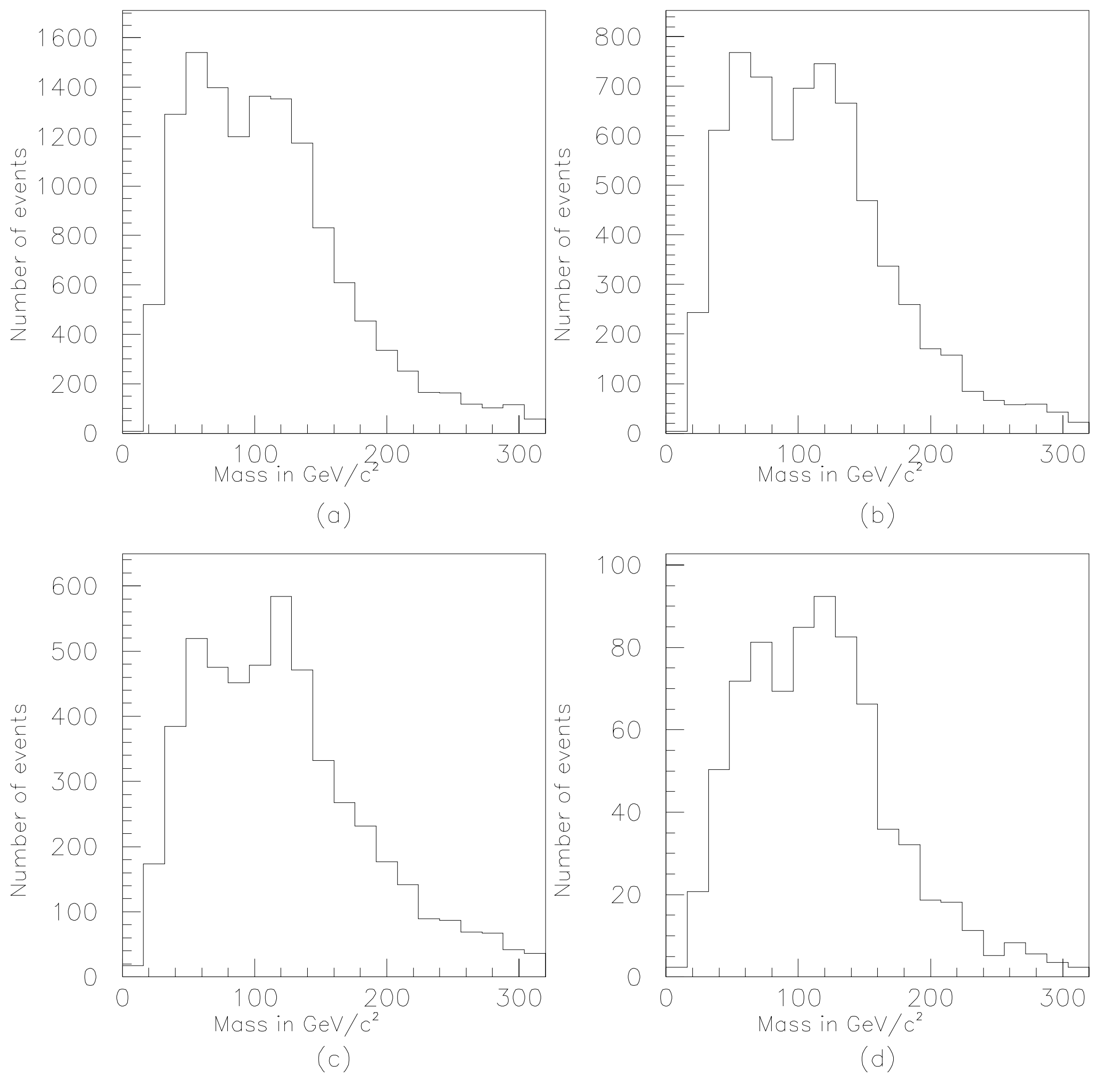

Figure 7: The $b\bar{b}$ invariant mass for the $t\bar{t}$ + H signal and the $t\bar{t}$ + $b\bar{b}$ and $t\bar{t}$ + Z backgrounds: (a) before simulation, without the Δy requirement, (b) before simulation with the Δy requirement, (c) after simulation, without the Δy requirement, and (d) after simulation, with the Δy requirement.

From Figure 7d, we can estimate the significance of the Higgs signal after the simulation and $|\Delta y|$ requirement. The number of $t\bar{t}$ + Higgs signal events in the peak divided by the square root of the $t\bar{t}$ + $b\bar{b}$ plus $t\bar{t}$ + Z background is:

$$\frac{N_S}{\sqrt{N_B}} \sim 11.8$$

with a signal to background ratio approximatly 1:1. For comparison, a previous study [11], found a signal to background ratio of 0.73 with the ratio of signal to square root of the background about 5.3.

5.3 Effect of Improved B-tagging

Figure 8a shows that the reconstructed $b\bar{b}$ mass for $t\bar{t}$ + Higgs if we had perfect b-tagging. Figure 8b shows the $b\bar{b}$ mass with the simulated b-tagging after the analysis procedure described above. Clearly, the mass resolution gets worse due to mis-identification of some jets. Figure 8c shows the same plots a) and b) overlapped for comparision. Figure 8d shows the same overlap of plots for the $t\bar{t}$ + Z sample. The results for the $t\bar{t}$ + Z sample are not as good mainly because we have optimized our analysis for the $t\bar{t}$ + Higgs signal.

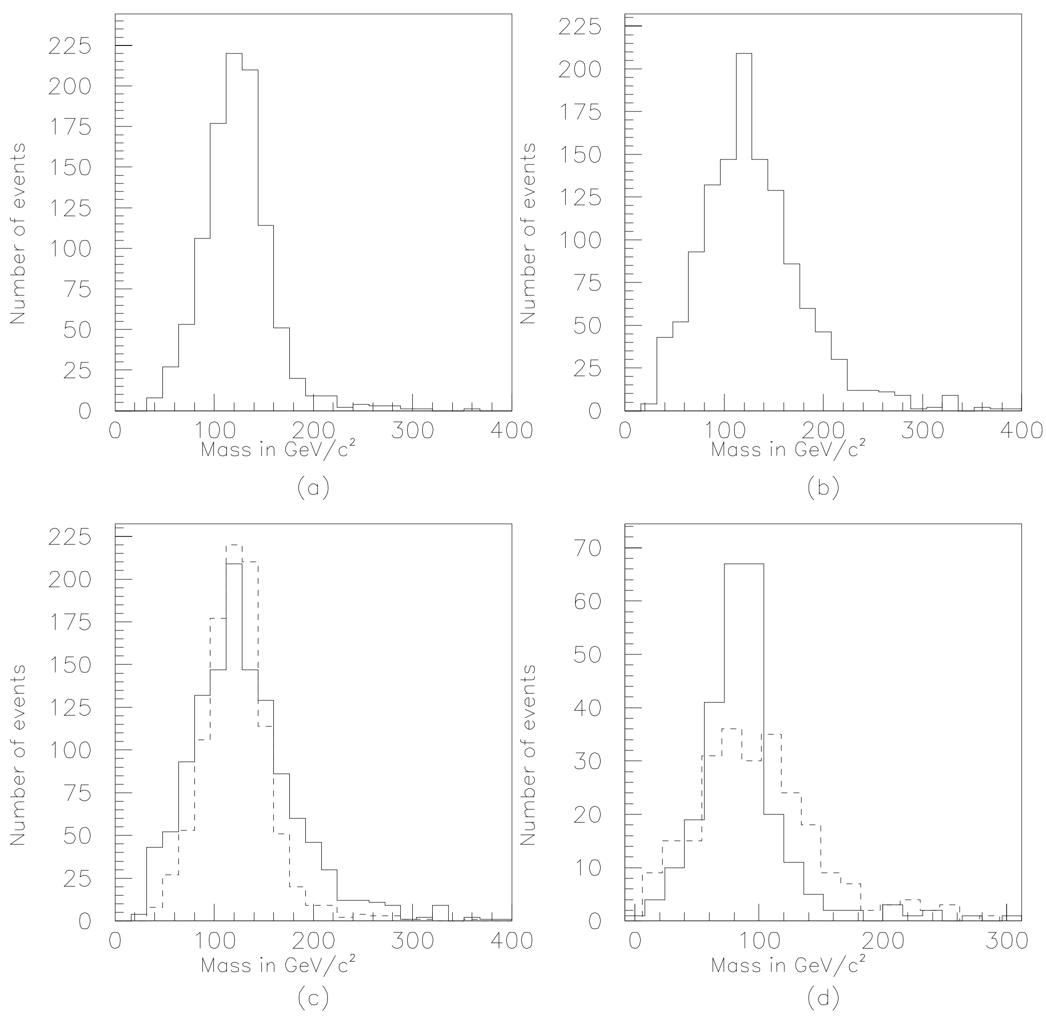

Figure 8: The $b\bar{b}$ mass for (a) the $t\bar{t}$ + H sample with perfect b-tagging, (b) the $t\bar{t}$ + H sample after analysis cuts, (c) overlap of (a) (dotted) and (b) (full), (d) overlap for $t\bar{t}$ + Z sample

6 Conclusion

If the electoweak symmetry breaking is caused by a Standard Model Higgs or a similar scaler particle decaying to $b\bar{b}$, and the Higgs has a large coupling to the t-quark, then the CMS detector at the LHC should be able to identify the Higgs via production of $t\bar{t}$ + Higgs after a reasonable amount of luminosity. Detection of the Higgs depends critically on efficient b-tagging and lepton identification, which are characteristics of the CMS detector. Reduced mis-tagging is also very important for identifying the two b-tagged jets from t-quark decays. Improvements in the calorimeter resolution and calibration also would be helpful. We believe CMS will be able to isolate a sample of Higgs events with a signal to background of order 1:1 with a clear peak in the $b\bar{b}$ invariant mass distribution.

We also find that there is a definite angular correlation between one of the t-quarks and the $b\bar{b}$ coming from the Higgs decay, which is is not present in the $t\bar{t}$ + $b\bar{b}$ background. This correlation can be an important tool in further discriminating the Higgs signal from the dominant background.

Acknowledgments

We would like to express our sincere thanks to Dr. V.A.Ilyin, Dr. A. Kryukov and the CompHEP group from Moscow State University, Russia. They generated all our events, and created library of partonic events for LHC processes as PEVLIB[6] with very close collaboration. We also would like to express our grateful thanks to David Rainwater, FNAL theory department. He has given us several talks to explain the physics, verify the cross section, and make cuts to suppress the $t\bar{t}\,b\bar{b}$ background. During our analysis, we also had a very valuable discussion with Prof.Ulrich Baur , University of Buffalo, New York.

References

[1] Associated production of Higgs and top pairs at CMS, Dan Green, Fermilab-FN-686, Nov. 1999.

[2] A discovery mode for the Higgs Boson at the Tevatron J.Goldstein et., Fermilab-Pub-00/146-T, June 27, 2000.

[3] Possibilities of $t\bar{t}$+H Event Reconstruction V.Drollinger et. CMS Note 1999/001, Oct. 1998.

[4] ATLAS detector and physics performance TDR May,1999

[5] CompHEP- a package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's Manual Preprint INP MSU 98-41/542, 1998 hep-ph/9908288

[6] CompHEP-PYTHIA interface: integrated package for the collision events generation based on exact matrix elements, hep-ph/0101232. A.S.Belyaev et al.

[7] CMS b-tag program from Volker Drollinger,CMS Note 1999/001, Oct. 1998.

[8] Energy Corrections for QCD jets, CMS Note, IN-2001/001

[9] Private communication with V.A.Ilyin and Torbjorn Sjostrand

[10] Single Top at CMS, CMS note 1999/048, Dan Green et., May 1999

[11] Finding H0—$b\bar{b}$ at the LHC V.Drollinger, hep-ex/0105017. 10th May 2001