Reconstruction of the top-antitop invariant mass spectrum (Candidacy proposal)

J-R. Lessard

April 18, 2008

Abstract

This document highlights the principal steps of my Ph.D. research proposal. The main topic is the reconstruction of the $M_{t\bar{t}}$ spectrum at ATLAS. That includes probing for hypothetical resonances as well as comparing the $M_{t\bar{t}}$ spectrum with the Standard Model (SM) prediction. The proposed analysis will be described after a brief review of our current knowledge of $t\bar{t}$ events. The analysis itself will be divided in two parts, the lower $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} < 1$ TeV) and the upper $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} > 1$ TeV) reconstruction, since both contain different experimental challenges. My activities in the operation of the ATLAS detector will also be discussed. A tentative time line for the proposed research will be presented. Finally, an alternative plan will be discussed in case the LHC does not provide enough data for the intended analysis.

Contents

1 Introduction

The primary goal for my Ph.D. research is to reconstruct the $M_{t\bar{t}}$ spectrum at ATLAS with early data. Once the spectrum is reconstructed, it will be possible to set limits on Beyond Standard Model (BSM) resonances or discover them. The $M_{t\bar{t}}$ spectrum is harder to reconstruct for high $M_{t\bar{t}}$ value since the number of events starts to be suppressed ($\frac{d\sigma_{t\bar{t}}}{dM_{t\bar{t}}}$ decreases exponentially passed $M_{t\bar{t}} \approx 450$ GeV)[1] and because high p_{T} tops are harder to identify (merged decay products issue). Therefore, the proposed analysis will be separated in two parts: lower ($M_{t\bar{t}} < 1$ TeV) and upper ($M_{t\bar{t}} > 1$ TeV) $M_{t\bar{t}}$ spectrum. In an optimistic scenario, the upper limit of the reconstructed $M_{t\bar{t}}$ spectrum could go up to 5 TeV while 1 TeV is a very conservative estimate.

First, the theory behind $t\bar{t}$ events and the phenomenology surrounding $M_{t\bar{t}}$ spectrum in the ATLAS context will be examined. A brief description of the LHC accelerator ring and the ATLAS detector will be given in the *LHC and ATLAS* section.

The first part of the analysis should be robust enough to be used with early ATLAS data (integrated luminosity of about 1 fb^{-1}). Therefore, the procedure and methodology must be simple and transparent. Namely, the number of 'tuning' parameters in the analysis should be as small as possible as well as their effect on the final result. Moreover, the use of Monte Carlo (MC) simulations to extract experimental results should be kept to a minimum. Finally, the signal to background ratio (purity) should be optimized and an efficient way to evaluate this ratio in-situ should be developed. The proposed methodology and a feasibility MC study (to be completed) for this analysis will be discussed in the *Lower $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} < 1$ TeV)* section. Additional challenges with LHC data will also be discussed in that section.

To reconstruct top pair events with $M_{t\bar{t}}$ higher than 1 TeV, a refined analysis will be needed. Indeed, since there will be few top events with high $M_{t\bar{t}}$ it is important to keep as many as possible. This means compromising purity to achieve better efficiency. The distribution of the background, including the total amount, should be known after the early data phase. It is therefore expected that the background could be removed from the $M_{t\bar{t}}$ spectrum. The major issue with high $M_{t\bar{t}}$ events will then be to recover those events with merged decay products, whether they are merged jets or the overlap of a lepton and jet(s). There are many promising approaches to solve that problem. Extensive MC studies will be done to identify the best tool, or combination of tools, to reconstruct collimated decay products. This is discussed in the *Upper $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} > 1$ TeV)* section.

Finally, service tasks related to the operation of the ATLAS detector will be discussed. Those will be put in perspective with the proposed schedule for the Ph.D. completion in the *Research time line* section.

[1] Natural units ($\hbar = c = 1$) are used throughout the text.

2 Theory

The discovery of the top quark in 1995 at Fermilab was a great achievement and an important milestone for the Standard Model of particle physics (SM). It completed the third (and last) family of quarks. However, its high mass ($m_{\mathrm{top}} = 172.5 \pm 2.6$ GeV [1]) is a mystery. Its mass is the highest of all known elementary particles, about a factor of two higher than the electroweak bosons. As a consequence, no particle of the SM are allowed to decay into a top-antitop quark pair ($t\bar{t}$).

However, many new models propose to enlarge the SM with new heavy particles. They are refereed to as Beyond the Standard Model (BSM) theories. Technicolor and extra dimensions are examples of BSM in which the heavy new particles could decay into a top-antitop quark pair. Before considering details of physics BSM, the properties of the process $pp \rightarrow t\bar{t}$ in the SM will be explored.

2.1 Production of $t\bar{t}$

The processes that generate top-antitop quark pair from proton-proton collisions ($pp \rightarrow t\bar{t}$) are well known in the SM. The main Feynman diagrams that contribute to the top-antitop cross section ($\sigma_{t\bar{t}}$) are shown in Figure 1, where gluons (g) or quarks (q) come from the protons.

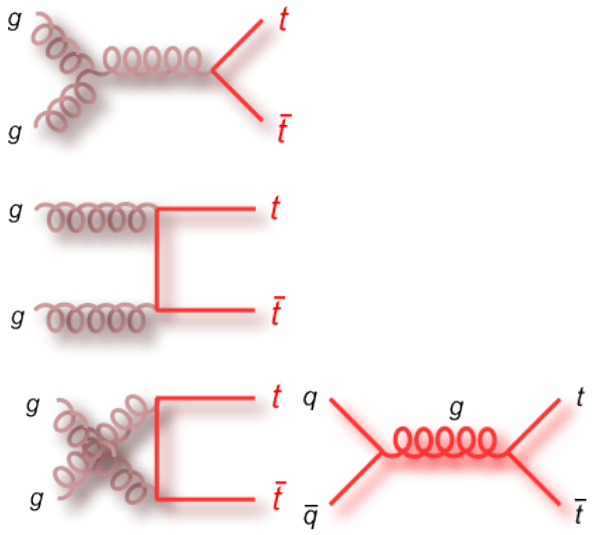


Figure 1: Leading order Feynman diagrams for the process $p\bar{p} \rightarrow t\bar{t}$.

2.2 Decay of $t\bar{t}$

Once produced, the top quark[2] is highly unstable. Namely, it will decay before it starts hadronizing (making a jet) like every other quarks do. More than 99% of the time, the decay products will be a W boson and a bottom (b) quark. Then, the W boson can either decay into a pair of light quarks (hadronically) or into a lepton-neutrino pair (leptonically). The overall decay channel for the top pair is therefore characterized by the way the W^+ from the top quark and W^- from the antitop quark decay. The fully hadronic (alljets) channel refers to both W bosons decaying hadronically. The semi-leptonic (lepton + jets) channel means that one W boson decays hadronically while the other decays leptonically. Finally, the fully leptonic (dileptons) channel is when both W bosons decay leptonically. The branching ratio for each of these channels is presented in Figure 2.

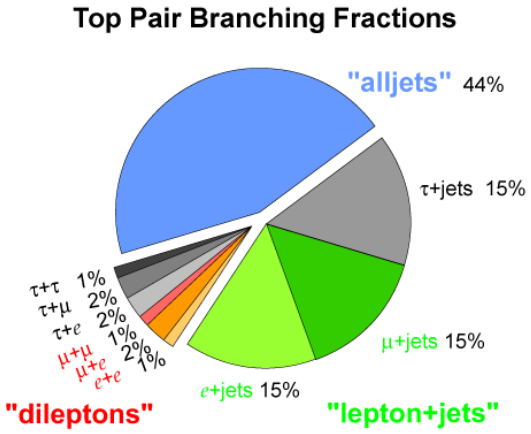


Figure 2: Branching ratio of the different decay channels of a top pair event.

2.3 Semi-Leptonic channel

Although the fully hadronic (6 jets) channel represents a large fraction of the top events, it is not often used for top analysis. From Quantum ChromoDynamics (QCD), there are many processes that generate events with multiple jets (6 or more). The cross section for these processes is many orders of magnitude higher than the one for top events. Consequently, it is difficult to separate fully hadronic top events from the QCD background.

The semi-leptonic channel leaves a more characteristic signature: 4 jets, a lepton and a neutrino. The neutrino is not directly detected, but its presence is deduced from an unbalance in the transverse energy of the event referred as $\not{E}_{\mathrm{T}}$[3]. None of the dominant

[2]Every time a top quark is mentioned, it should be understood that the equivalent is true for the antitop.

[3]Since the initial p_{T} of both protons is zero before the collision, the sum of all the produced particle p_{T} must also be

QCD backgrounds generate a lepton and some $\not{E}_T$. It is therefore possible to have a sample with high purity when considering the semi-leptonic channel. Note however that while the electrons (muons) is stable (has a life time long enough to be detected), tau leptons decay before reaching the detector. Therefore, hadronic decays of taus are often mis-identified as jets from quarks. For this reason, most of the analyses that use the semi-leptonic channel as signal, like the one presented here, consider only channels with an electron, or a muon, but not a tau. Figure 3 shows an example of a Feynman diagram for the semi-leptonic process.

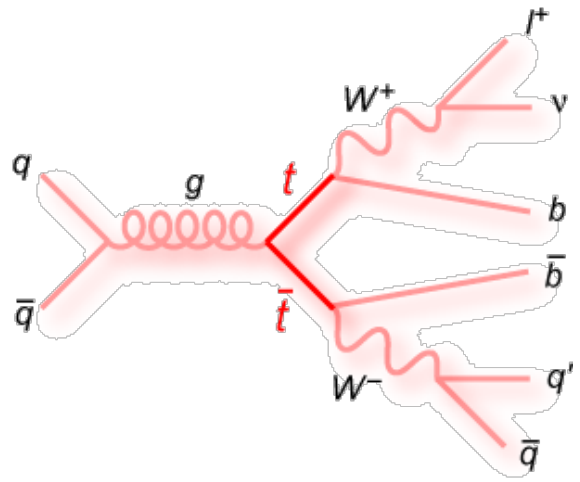


Figure 3: Example of a Feynman diagram for a semi-leptonic top pair decay, l stands for lepton (e or μ).

The fully leptonic channel cannot be used to reconstruct $M_{t\bar{t}}$ since the loss of information from having two undetected neutrinos is too important.

zero after: $\sum \vec{p}_T = 0$. The calorimetry only detects energy of particles, not the momentum. It is therefore convenient to work in the massless approximation for particles ($\vec{p}_T \approx \vec{E}_T$) and add energy vector instead such that $|\sum \vec{E}_T| = \not{E}_T$.

3 Phenomenology

3.1 $M_{t\bar{t}}$ spectrum in the SM context

Although the processes that generate $t\bar{t}$ are known, the theoretical cross section $\sigma_{t\bar{t}}$ has a large uncertainty. This is mainly due to the renormalization and factorization scale uncertainty and the parton density function (pdf) uncertainty. At the LHC these uncertainties represent $\pm 13\%$ and $\pm 3.2\%$ respectively of the total cross section $\sigma_{t\bar{t}}$ [2]. These uncertainties on $\sigma_{t\bar{t}}$ are of the same order of magnitude as the dependence of the top mass on $\sigma_{t\bar{t}}$. Therefore, it is not possible to estimate m_{top} from $\sigma_{t\bar{t}}$. However, studying the cross section as a function of the top-antitop invariant mass, the $M_{t\bar{t}}$ spectrum: $\frac{d\sigma_{t\bar{t}}}{dM_{t\bar{t}}}(M_{t\bar{t}})$, opens new possibilities. It has been shown that simple quantities like the mean of the $M_{t\bar{t}}$ spectrum are strongly correlated to the mass of the top quark:

$$\frac{\Delta m_{\text{top}}}{m_{\text{top}}} = 1.2 \frac{\Delta < M_{t\bar{t}} >}{M_{t\bar{t}}} + 0.005. \qquad (1)$$

It means that being able to evaluate the mean of the $M_{t\bar{t}}$ spectrum up to 1% uncertainty allows to estimate m_{top} with 1.7% uncertainty [2]. This conclusion assumes a full reconstruction of the $M_{t\bar{t}}$ distribution. Experimental considerations, like removing background from signal, will require the loss of some $t\bar{t}$ events. Having a partially reconstructed $M_{t\bar{t}}$ spectrum will sligthly reduce our power to probe m_{top}. This technique is probaly not the most precise to evaluate m_{top}, but it offers a cross check with more direct approches.

3.2 $M_{t\bar{t}}$ spectrum in the BSM context

For many BSM theories, the $M_{t\bar{t}}$ spectrum is the main signature. Indeed, due to the high mass of the top quark, some heavy BSM particles would primarily decay into a top-antitop quark pair. The Topcolor Z′ (leptophobic or not) [3] and KK-gluon/graviton [4]/[5] are examples of such BSM particles that would reveal themselves as resonances in the $M_{t\bar{t}}$ spectrum. Figure 4 and 5 show how the $M_{t\bar{t}}$ distribution would be affected. Note that the mass and width of these

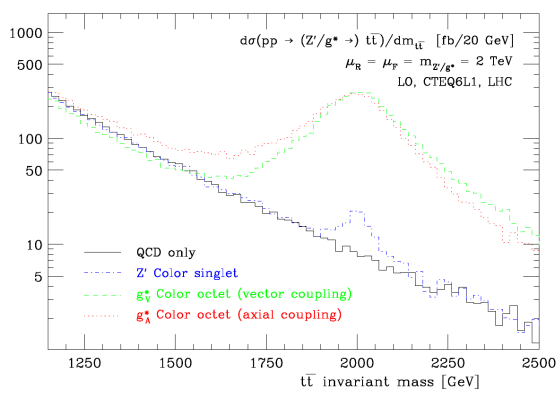


Figure 4: $M_{t\bar{t}}$ spectrum for different resonances set at 2 TeV: Z′ color singlet in blue and color octet vector (axial) coupling in green (red) [2].

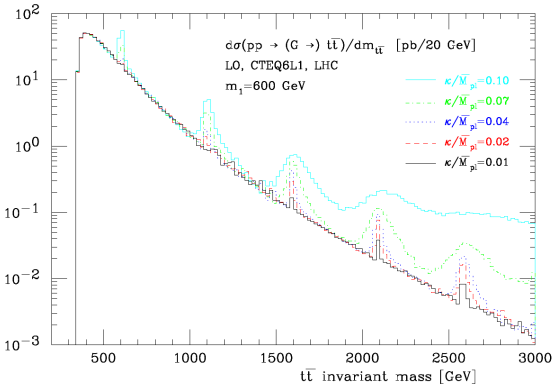


Figure 5: $M_{t\bar{t}}$ spectrum in the presence of KK-gravitons. The first mass (600 GeV) is arbitrary while the masses of the other KK-gravitons are determined by the zeros of the Bessel function $J_1(x)$ [2].

resonances are model dependent and somewhat arbitrary at this stage. For more models, see [2].

The main goal of the analysis is to reconstruct the $M_{t\bar{t}}$ spectrum without choosing any model in particular that could generate a resonance in the $M_{t\bar{t}}$ spectrum. Once a resonance is identified, if any, it will be possible to investigate it further to understand which process or particle could be responsible for the resonance.

4 Previous experimental studies

A search for resonances in the $M_{t\bar{t}}$ spectrum has already been performed at the Tevatron [6]. The reconstructed $M_{t\bar{t}}$ spectrum is shown in Figure 6. A direct search for a leptophobic Z′ with $\Gamma_{Z'} = 1.2\% M_{Z'}$ has been performed as well, see Figure 7. Resonances for $M_{Z'} < 725$ GeV have been excluded at 95% confidence level.

These studies have been performed in 2007. At that time, the Collider Detector at Fermilab (CDF) Run II had acquired an integrated luminosity of 680 pb^{-1}. The analysis only used the $t\bar{t}$ semi-leptonic (e or μ) channel. This channel corresponds to a cross section of 2.2 pb at the Tevatron ($\sqrt{s} = 1.96$ TeV), leaving only a few hundred events to perform the analysis. This is to be compared with a design luminosity of 30 fb^{-1}/year and a cross section of 362 pb for the $t\bar{t}$ semi-leptonic (e or μ) channel at the LHC ($\sqrt{s} = 14$ TeV). Therefore, at the LHC, tens of millions of top-antitop events are expected yearly. The $M_{t\bar{t}}$ spectrum extracted at the LHC will have an unprecedent reach for resonances with masses up to a few TeV.

5 LHC and ATLAS

The LHC is an underground particle accelerator ring schedule to start operation in 2008. Its large circumference of 27 km and its powerful superconducting dipole magnets of 8.4 Tesla will permit proton-proton (pp) collisions with 14 TeV center of mass energy. There are three interaction points where such pp collisions occur: the ATLAS, CMS and LHCb detectors. All detectors are designed to be able to identify visible SM particles created by the pp collision. While the CMS detector puts a larger emphasis on precise reconstruction of muons and the LHCb on

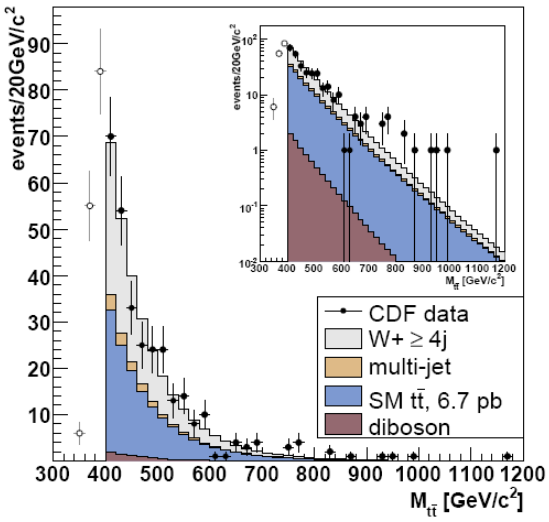


Figure 6: The reconstructed $M_{t\bar{t}}$ spectrum (data) and SM prediction in the search region above 400 GeV/c^2 [6].

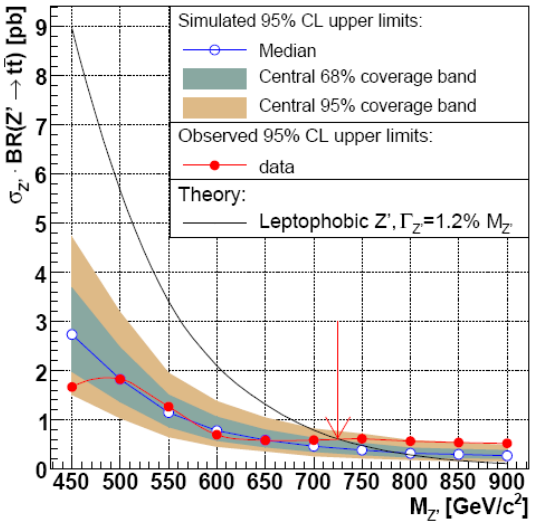


Figure 7: Predicted and experimental 95% CL upper limits on the leptophobic Z′ cross section ($\sigma_{Z'} \cdot BR(Z' \to t\bar{t})$) using data corresponding to 680 pb^{-1} of integrated luminosity. Dark and light areas define the 1σ and 2σ ranges for the expected limits. The predicted leptophobic topcolor Z′ cross section is overlaid in black. The arrow marks the Z′ mass upper limit [6].

b-physics, the ATLAS detector focuses on jet resolution. The proposed analysis will be performed using data collected by the ATLAS detector.

The ATLAS detector, Figure 8, can be divided in four main parts: the inner tracker, the calorimeter, the muon spectrometer and the magnet system. The inner tracker, composed of the Pixel Detector and SCT/TRT Trackers, enables the reconstruction of charged particles as well as identifying the primary vertex of interaction. The calorimeter is divided in two parts, the electromagnetic (EM) calorimeter and the hadronic calorimeter. They both estimate the energy deposited by the EM particles (electrons and photons) and by hadrons (protons, neutrons, etc.) respectively. The EM calorimeter is located in the Liquid Argon Calorimeter, while the hadronic calorimeter is in the end cap region of the Liquid Argon Calorimeter (HEC) and in the Tile Calorimeter. The muon spectrometer is designed to accurately reconstruct muons. Finally, the magnet system produces the magnetic field to curve charge particle trajectories allowing a measurement of their momentum. The Solenoid Magnet is used by the inner detector while the Toroid Magnets are there for the Muon detectors.

6 Lower $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} < 1$ TeV)

6.1 Methodology

6.1.1 Software and samples used

In order to test the feasibility of the proposed analysis to reconstruct the $M_{t\bar{t}}$ spectrum with early data, the sample 5200 (TopView version 1213, MuidTauRec) created by the top working group has been used. It consist of a MC simulation of $t\bar{t}$ events where fully hadronic events are removed. The top and antitop are generated according to their matrix element evaluated using the MC@NLO generator [7] and the un-

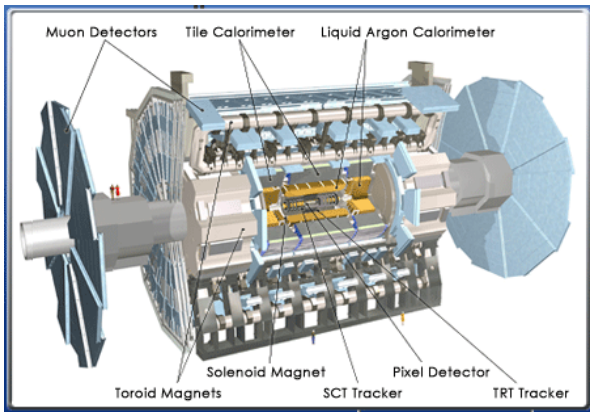


Figure 8: Overview of the ATLAS detector.

derlying event of $t\bar{t}$ is simulated using the HERWIG generator [8]. The simulated events are then pass trough a full MC simulation of the ATLAS detector. The software TopView is then used to treat the reconstructed objects in the detector and to make a data file (ntuple). The muons are reconstructed using the MuidTauRec technique. The important quantities for the sample 5200 are displayed in Table 1. Note that the MC@NLO generator is a next to leading order (NLO) simulation and can output events with negative weight -1; this is why the number of entries is not the same than the number of events.

The main background for the $t\bar{t}$ semi-leptonic signal is W+jets events (direct production of a W boson from $qq' \rightarrow W$ + jets from QCD background) where the W decays leptonically. Samples 8240-8251 (TopView version 1213, MuidTauRec) have been created by the top working group to study this source of background. Generators Alpgen [9] (matrix elements) and HERWIG (underlying event) where used to simulate these samples to leading order (LO). Details of these samples can be found in Table 1.

6.1.2 Event selection

The current analysis uses only top events decaying in the semi-leptonic channel with an electron or muon ($t\bar{t} \rightarrow b\bar{b}qq'l\nu_l$, where l = electron or muon). Therefore, 4 jets, a lepton and some $\not{E}_{\mathrm{T}}$ is required to select an event. The commissioning cuts, those proposed by the top working group to have a good purity with early data, are as follows.

The three highest p_{T} (momentum in the transverse plane) jets must have $p_{\mathrm{T}} > 40$ GeV while the fourth jet must have $p_{\mathrm{T}} > 20$ GeV. Events that have more than 4 jets, unlike those with less than four jets, are not discarded. These extra jets are simply ignored. They can come from underlying events, pile-up, mis reconstruction of jet(s) or bad signal, see *Treating the background*. The jets algorithm used is cone 0.4 (tower). Moreover, a jet needs to have $|\eta| < 2.5$ and to have no lepton within a distance $\Delta R < 0.4$ to be kept, $R = \sqrt{\eta^2 + \phi^2}$.

There must be only one high p_{T} lepton (electron or muon) in the event. This lepton must have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. In addition, electrons with $1.37 < |\eta| < 1.52$ (crack region between barrel EM calorimeter and forward EM calorimeter) are excluded of the analysis; the energy resolution for the electron is considerably reduced in that region. Muons are not affected since they are detected in the muons chamber. Moreover, there is a $E_{\mathrm{T,cone20}} < 6$ GeV cut on the lepton, meaning that there must not be more than 6 GeV of the transverse energy of the lepton found outside a cone $\Delta R = 0.2$ around the lepton (which ensures a good isolation of the lepton).

Finally, to ensure the presence of a neutrino, an overall $\not{E}_{\mathrm{T}} > 20$ GeV is required. A summary of the commissioning cuts is shown in Table 2 while the efficiency of these cuts on the signal and background samples can be seen in Table 1.

Samples:	No Cut			Commissioning Cuts		$\chi^2 < 5$ Cut	
	# Entries	# Events	Lumi (pb-1)	# Events	Efficiency	# Events	Efficiency
ttbar (5200)	592300	433578	941	53473	12.33%	10960	2.53%
Signal (evqq' or μvqq')	474506	347350	941	48695	14.02%	10694	3.08%
Fully leptonic	52340	38230	941	2098	5.49%	108	0.28%
(τν + X, X not qq')	52257	38345	941	2472	6.45%	150	0.39%
(ττν)	13197	9653	941	208	2.15%	8	0.08%
W+Jets (8240-8251)	112100	112100	N/A	2139	1.91%	152	0.14%
8240 (W->eν + 2 partons)	18750	18750	88	15	0.08%	2	0.01%
8241 (W->eν + 3 partons)	11250	11250	91	91	0.81%	6	0.05%
8242 (W->eν + 4 partons)	6000	6000	111	290	4.83%	27	0.45%
8243 (W->eν + 5 partons)	4950	4950	225	404	8.16%	27	0.55%
8245 (W->μν + 3 partons)	11300	11300	167	248	2.19%	20	0.18%
8246 (W->μν + 4 partons)	3200	3200	89	339	10.59%	26	0.81%
8247 (W->μν + 5 partons)	4500	4500	225	646	14.36%	38	0.84%
8248 (W->τν + 2 partons)	20600	20600	234	5	0.02%	1	0.00%
8249 (W->τν + 3 partons)	26000	26000	299	44	0.17%	4	0.02%
8250 (W->τν + 4 partons)	5000	5000	109	48	0.96%	1	0.02%
8251 (W->τν + 5 partons)	550	550	26	9	1.64%	0	0.00%

Table 1: The samples used (signal in red) for the analysis. The commissioning cuts and $\chi^2 < 5$ cut are explained in the text.

Event Selection [signal = semi-leptonic (e/μ only)]	
1 isolated lepton (e/μ)	3 leading jets with $p_{\mathrm{T}} > 40$ GeV
$\not\!\!E_{\mathrm{T}} > 20$ GeV	4^{th} jet with $p_{\mathrm{T}} > 20$ GeV
Good lepton (e or μ)	**Good jet**
$p_{\mathrm{T}} > 20$ GeV	Cone 0.4 (tower)
$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
$1.37 < \lvert\eta\rvert < 1.52$ (excluded for e)	No lepton within $\Delta R < 0.4$
$E_{\mathrm{T,cone20}} < 6$ GeV	

Table 2: Commissioning cuts used to select semi-leptonic $t\bar{t}$ events.

6.1.3 Determination of $M_{t\bar{t}}$

The top-antitop invariant mass is obtained by summing all the top decay product 4-vectors: $M^2_{t\bar{t}} = (p_b + p_{\bar{b}} + p_q + p_{q'} + p_l + p_\nu)^2$. However, the observables detected for a $t\bar{t}$ event are 4 jets, a lepton and some $\not{E}_{\mathrm{T}}$. Therefore, the goal is to associate these observables to the top decay product 4-vectors: ($\mathrm{jet}_{1...4}$, lepton, $\not{E}_{\mathrm{T}}$) $\rightarrow (p_b, p_{\bar{b}}, p_q, p_{q'}, p_l, p_\nu)$. There are two main problems in that assignment. First, the only information about p_ν comes from $\not{E}_{\mathrm{T}}$ which gives no information about the momentum in the longitudinal (z) direction : ($\not{E}_{\mathrm{T}} \rightarrow p_{\mathrm{T}\nu}$) such that $p_{\mathrm{z}\nu}$ is unknown. Moreover, there is a large uncertainty in the association ($p_{\mathrm{jet}} \rightarrow p_{\mathrm{quark}}$) since the energy resolution of the jets is non negligible. This is also true for the $\not{E}_{\mathrm{T}}$ resolution. One way to handle both problems is to use a Bayesian statistical approach called maximum posterior probability. This technique is used to define a χ^2 quantity (for details leading to this equation, see [10]):

$$\chi^2 = \sum_{i=1}^{4} \left(\frac{\alpha_i E_i - E_i}{\sigma_{\mathrm{jets}}(\alpha_i E_i)} \right)^2 + \left(\frac{\lambda \not{E}_{\mathrm{T}} - \not{E}_{\mathrm{T}}}{\sigma_{\not{E}_{\mathrm{T}}}(\lambda \not{E}_{\mathrm{T}})} \right)^2 + \sum_{\mathrm{type=lep,had}} \left(\frac{M_W^{\mathrm{type}} - M_W^0}{\Gamma_{M_W}} \right)^2 + \sum_{\mathrm{type=lep,had}} \left(\frac{M_{\mathrm{top}}^{\mathrm{type}} - M_{\mathrm{top}}^0}{\Gamma_{M_{\mathrm{top}}}} \right)^2 . \quad (2)$$

The idea is to chose the set of six 4-vectors $(p_b, p_{\bar{b}}, p_q, p_{q'}, p_l, p_\nu)$ that minimizes this χ^2 given the following constraints: $(p_{\mathrm{lepton}} \rightarrow p_l = p_{\mathrm{lepton}}), (\not{E}_{\mathrm{T}} \rightarrow p_{\mathrm{T}\nu} = \lambda \not{E}_{\mathrm{T}})$ and $(p_{\mathrm{jet}_i} \rightarrow p_{\mathrm{quark}} = \alpha_i p_{\mathrm{jet}_i})$, where the subscript i runs over the four different jets, the type superscript runs over leptonic (lep) and hadronic (had), the 0 superscript refers to the known values of the masses [1], σ is the resolution, Γ is the natural particle width while λ and α_i are free parameters. The longitudinal momentum of the neutrino ($p_{\mathrm{z}\nu}$) is treated as a free parameter as well, for a total of 6 free parameters to fit to get the best (smallest) χ^2: $\alpha_{1..4}, \lambda$ and $p_{\mathrm{z}\nu}$. In other words, Equation (2) is used to rescale the observables, "within their errors", to more accurately reconstruct the masses of the Ws and tops.

While this approach improves the poor resolution issue and gives $p_{\mathrm{z}\nu}$, it creates a new problem: which jets should be associated with which quarks, $(\mathrm{jet}_1, \mathrm{jet}_2, \mathrm{jet}_3, \mathrm{jet}_4) \rightarrow (p_b, p_{\bar{b}}, p_j, p_{j'})$? Since j and j' are interchangeable, it leaves 12 *a priory* legitimate hypotheses. The one giving the smallest χ^2 after the minimization of equation (2) is chosen as the "good" hypothesis. In that context, the chosen hypothesis without applying the rescaling factors ($\alpha_{1..4}, \lambda$ and $p_{\mathrm{z}\nu}$) will be referred as an event *before minimization*. The 4-momentum of the neutrino is re-evaluated using $\not{E}_{\mathrm{T}}$ and constraining the top and W masses. The event *after minimization* is the chosen hypothesis with the rescaling factors applied.

Note that this selection sometime fails to lead to the right association $(\mathrm{jet}_1, \mathrm{jet}_2, \mathrm{jet}_3, \mathrm{jet}_4) \rightarrow (p_b, p_{\bar{b}}, p_j, p_{j'})$ and therefore results in a bad reconstruction of $M_{t\bar{t}}$. Taking the wrong hypothesis is known as combinatoric background.

6.1.4 Treating the background

The background events considered for the semi-leptonic $t\bar{t}$ signal are the fully leptonic, $\tau\nu + X(X \neq qq')$ and $\tau\nu\tau\nu$ channels of $t\bar{t}$ as well as the W+jets processes. When the commissioning cuts are applied, the Signal over Background ratio (S/B) is 2.6, while the efficiency of the signal is 14%.

In addition to background events, there is background inside a signal event. There are two such sources of background called underlying event and pile-up. The underlying event is formed of all the partons that did not participate directly in the process of creating the top event, but that did interact to form other decay products. The underlying event is responsible for the fact that most top events will have more than four jets, up to about 10 jets in certain cases. These jets from the underlying event are often less energetic and this is why the four leading jets (highest p_{T} jets) should, most of the times, originate from the $t\bar{t}$ process. Sometimes it will not be the

case so that none of the 12 hypotheses will be good. Such situation will be referred as *bad signal* event. The reason why the analysis looks only at the four leading jets is to reduce combinatoric background.

Pile-up is a combination of two problems. First, the in-time collisions are caused by the crossing of two bunches of protons. It can therefore happen that more than one pair of protons interact. The event of interest can therefore be contaminated by other in-time pp collisions. Second, the events from previous bunch crossing affect the event. This is because the detector cannot instantaneously "reset" and interactions left by previous events can sometime affect the reconstruction of the in-time event. Pile-up was not simulated in the samples used; the present analysis therefore neglects that effect. This will have to be addressed in the future.

6.2 Feasibility MC study (completed part)

A way to verify if the chosen hypothesis is indeed the good one is to verify if the reconstructed top and W yield the right mass before the minimization, Figure 9. Many of the selected hypotheses have significantly wrong reconstructed masses for the hadronic W and top, even when the events are $t\bar{t}$ events (signal in red). This cannot be attributed only to the resolution of jets and $\not{E}_{\mathrm{T}}$. It rather indicates that the rate of bad signal ($t\bar{t}$ events where none of the 12 hypotheses are good) is quite high. The asymmetric shift of the distribution toward higher masses suggests that these bad signal events are caused by considering only the four leading jets of the event. These bad signal events are well removed by using a $\chi^2_{\mathrm{cut}} = 5$, meaning that only events with $\chi^2_{\mathrm{best}} < 5$ are kept. This is somehow similar to using window cuts around the W and top masses, see Equation (2). This can be observed in Figure 10. The choice of $\chi^2_{\mathrm{cut}} = 5$ as been made[4] considering Figure 11; it yields a good S/B of 8 and the efficiency of the signal is not too low, 3%.

[4]From now on, $\chi^2_{\mathrm{cut}} = 5$ is default.

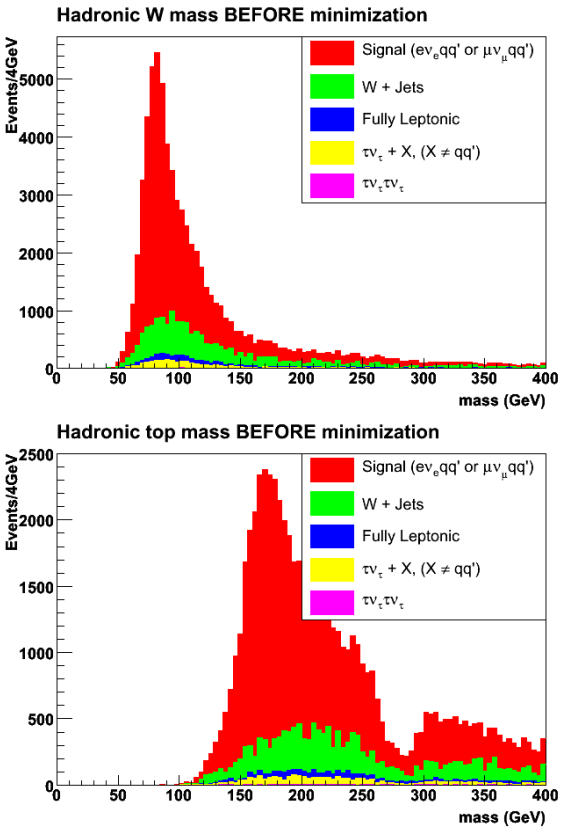
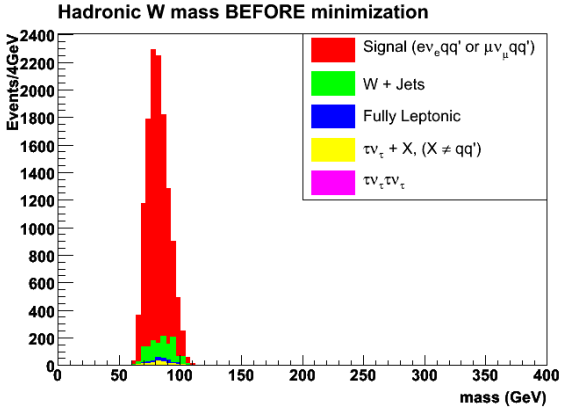


Figure 9: The reconstructed masses for the hadronic W and top before the minimization when no χ^2_{cut} is applied. The histograms are stacked.

Figure 12 shows the resolution of $M_{t\bar{t}}$ for different $M_{t\bar{t}}$ [true] bin when a $\chi^2_{\mathrm{cut}} = 5$ is used[5]. It can be observed that the resolution is narrower after the rescaling factors are applied (left plots vs right plots). However, theses distributions are quite asymmetric with large tails. No explicit studies of that effect have been done, but it is most likely caused by rescaling a wrong hypothesis (combinatoric background), since most of the bad signal has been cut by $\chi^2_{\mathrm{cut}} = 5$ as can be seen in Figure 10. In the case of before minimization events, the combinatoric background has a considerably weaker effect. In fact, the

[5]Of course, this is done only for signal events since background events do not have $M_{t\bar{t}}$ [true].



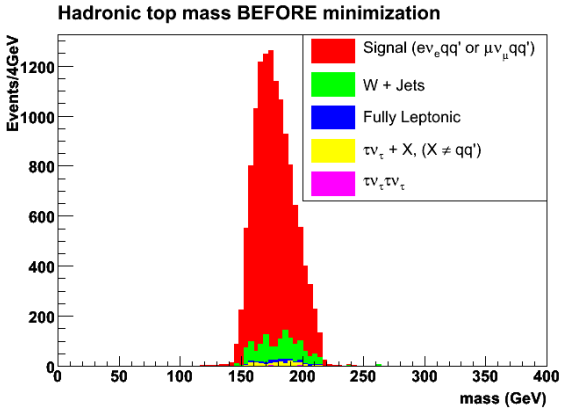


Figure 10: The reconstructed masses for the hadronic W and top before the minimization when a $\chi^2_{\text{cut}} = 5$ is applied. The histograms are stacked.

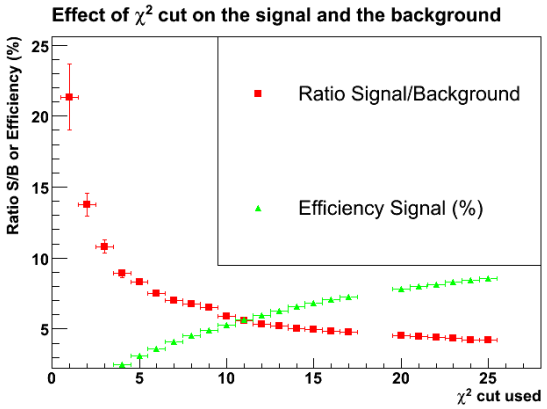


Figure 11: Signal to background ratio and total efficiency of the signal in percentage as a function of the χ^2_{cut} used. Note that one curve is a ratio while the other one is in percentage (The intersection point has therefore no particular meaning).

only change between a good or wrong hypothesis selection is which jet will be used as leptonic b-quark to find $p_{z\nu}$. But a wrong estimate of $p_{z\nu}$ has a small effect on the $M_{t\bar{t}}$ reconstruction. In comparison, for rescaled events, a wrong hypothesis leads to a wrong shift in most or all the jets energy and/or the complete 3-momentum of the neutrino (both $\not\!\!E_{\text{T}}$ and $p_{z\nu}$).

The way the resolution of the reconstructed $M_{t\bar{t}}$ behaves as function of $M_{t\bar{t}}$ [true], $R(M_{t\bar{t}}[\text{true}])$, is of prime importance since it will smear the true $M_{t\bar{t}}$ spectrum: $M_{t\bar{t}}[\text{reco}] = M_{t\bar{t}}[\text{true}] \otimes R(M_{t\bar{t}}[\text{true}])$. The effect of both resolutions, before and after minimization, on the reconstructed $M_{t\bar{t}}$ spectrum can be observed in Figure 13. The effect of having a significant systematic error in $R(M_{t\bar{t}}[\text{true}])$ after minimization leads to that unwanted behavior for the reconstructed $M_{t\bar{t}}$ compared to the true $M_{t\bar{t}}$ spectrum for $M_{t\bar{t}}$ [true] < 500 GeV. Although it is possible to "undo" this systematic by studying carefully $R(M_{t\bar{t}}[\text{true}])$ after minimization and apply a deconvolution to the reconstructed $M_{t\bar{t}}$ spectrum [10], it implies an important dependence on the MC simulation which would make the analysis less robust. For that reason, it has been chosen to do the subsequent analysis without applying the rescaling factors since robustness is favored over precision measurement in an early data context.

Nevertheless, even without rescaling, there is a systematic error in the resolution $R(M_{t\bar{t}}[\text{true}])$ as can be seen in figure 14. The reconstructed $M_{t\bar{t}}$ passes from being over estimated by 5% at $M_{t\bar{t}} = 350$ GeV to being under estimated by 5% after $M_{t\bar{t}} = 500$ GeV. This systematic effect "pushes" reconstruted events toward the 400 GeV peak of the $M_{t\bar{t}}$ spectrum. This effect counter-balances the smearing of the true $M_{t\bar{t}}$ spectrum by the width of the resolution $(\sigma_{M_{t\bar{t}}}/M_{t\bar{t}})$ of $9 \pm 1\%$ and explains why the reconstructed $M_{t\bar{t}}$

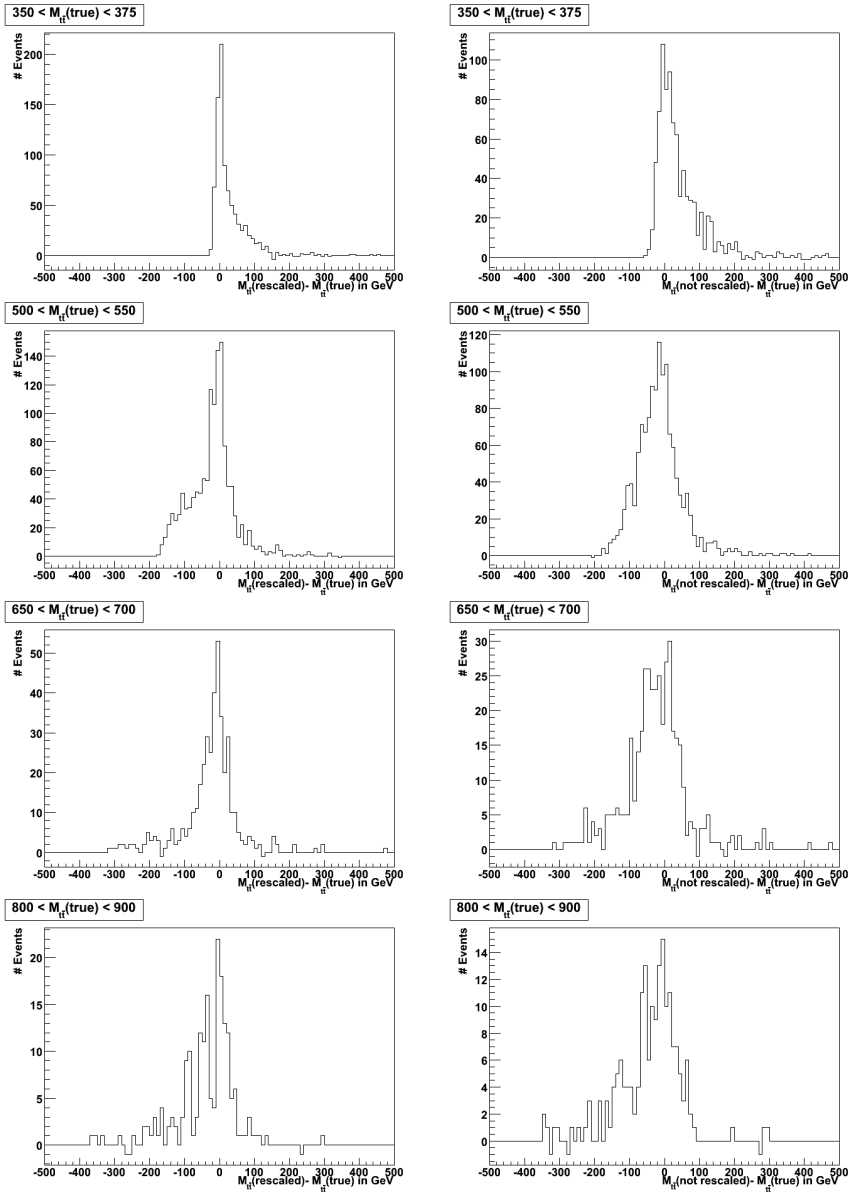


Figure 12: The left hand side plots show the resolution after minimization ($M_{t\bar{t}}$ (rescaled) - $M_{t\bar{t}}$ (true)) for different ranges of $M_{t\bar{t}}$ (true), from top to bottom: [350,375[; [500,550[; [650,700[; [800,900[. The right hand side plots show the resolution before the minimization ($M_{t\bar{t}}$ (not rescaled) - $M_{t\bar{t}}$ (true)) for the same $M_{t\bar{t}}$ (true) ranges.

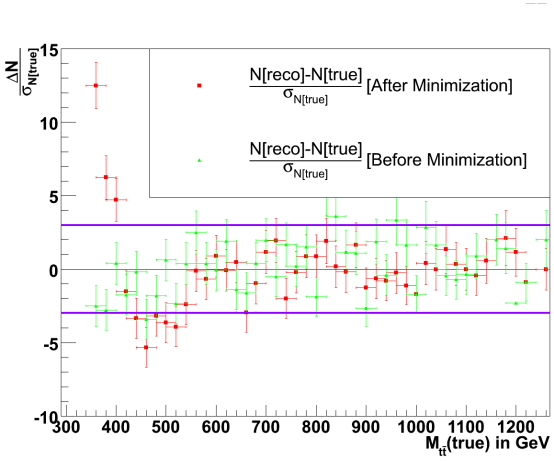


Figure 13: For each bin of 20 GeV in the $M_{t\bar{t}}$ spectrum, the difference between the number of events reconstructed, N[reco], minus the number of true events, N[true], over the error on the number of true events, $\sigma_{N[\text{true}]} = \sqrt{N[\text{true}]}$, is recorded: $\frac{N[\text{reco}]-N[\text{true}]}{\sigma_{N[\text{true}]}}$. In other words, the difference between the reconstructed $M_{t\bar{t}}$ spectrum and the true $M_{t\bar{t}}$ spectrum in unit of $\sigma_{N[\text{true}]}$ as a function of $M_{t\bar{t}}$ [true] is plotted. The before minimization $M_{t\bar{t}}$ spectrum is in green, while the after minimization one is in red. The purple lines delimit the $\pm 3\sigma$ zone.

spectrum agree so well with the true $M_{t\bar{t}}$ spectrum.

To summarize the procedure, the χ^2 is used to select the good hypothesis (removes some combinatoric background) and to cut bad signal and event backgrounds through a χ^2 cut (which improves considerably the purity). The rescaling factors are not applied since they are a source of systematic error. The $p_{z\nu}$ is then found by minimizing an equation similar to (2), but without the first two terms and the sum term only have leptonic type. The MC reconstructed $M_{t\bar{t}}$ spectrum using that procedure is shown in Figure 15.

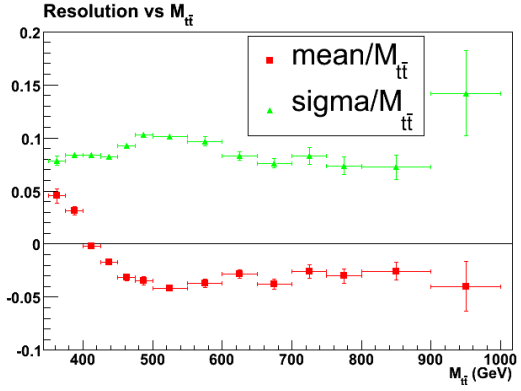


Figure 14: Fitted Gaussian mean and sigma of the resolution $R(M_{t\bar{t}}[\text{true}])$ over $M_{t\bar{t}}$ as a function of $M_{t\bar{t}}$ [true].

6.3 Feasibility MC study (to be completed)

6.3.1 Probing for resonances

One thing that has not been done yet, but should be looked at soon, is to come up with exclusion zones for a possible resonance (X) in the $M_{t\bar{t}}$ spectrum, namely reproduce a plot similar to the one that the CDF collaboration published, Figure 7. The quantity of interest is the upper limit of the cross section of $pp \to X$ multiplied by the branching ratio $BR(X \to t\bar{t})$[6] as function of the resonance mass (M_X). A possible prescription to exclude σ_X at a given confidence level (CL) is to use

$$1 - \text{CL} = \int_{\sigma_X\,(\text{up})}^{\infty} p(\sigma_X|\vec{n})\,\mathrm{d}\sigma_X, \qquad (3)$$

where $p(\sigma_X|\vec{n})$ is the probability of obtaining the measured $M_{t\bar{t}}$ spectrum histogram ($\vec{n}$) giving the cross section of the resonance (σ_X). A Bayesian approach is used to find that quantity

$$p(\sigma_X|\vec{n}) = \frac{L(\vec{n}|\sigma_X)\pi(\sigma_X)}{\int L(\vec{n}|\sigma'_X)\pi(\sigma'_X)\,\mathrm{d}\sigma'_X}, \qquad (4)$$

[6] From now on, this quantity will be referred simply as the cross section (σ_X).

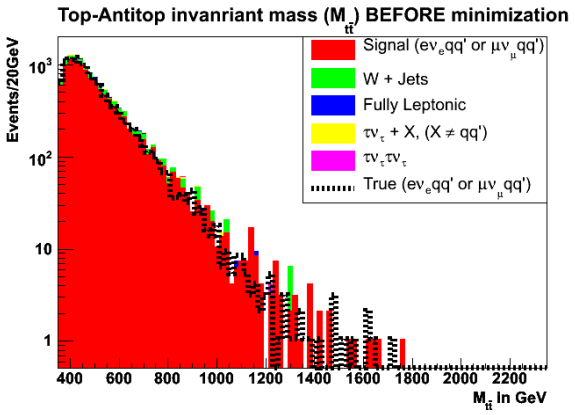


Figure 15: $M_{t\bar{t}}$ spectrum for the final proposed procedure (before minization $M_{t\bar{t}}$).

where $\pi(\sigma_X)$ is the prior probability density function for σ_X [1]. Since a model independent analysis is privileged, $\pi(\sigma_X)$ is taken to be 1. The number of measured events in bin i (n_i) follows a Poisson distribution. The likelihood $L(\vec{n}|\sigma_X)$ can therefore be written as

$$L(\vec{n}|\sigma_X) = \prod_{i\in\{\text{bins}\}} \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!}, \qquad (5)$$

where μ_i is the expected number of events in bin i for a given σ_X. So everything comes down to knowing accurately $\vec{\mu}$. Note that $\vec{\mu}$ is the expected reconstructed $M_{t\bar{t}}$ spectrum such that the resolution on $M_{t\bar{t}}$ is already accounted for in that quantity. Moreover, $\vec{\mu}$ does not only include the signal for the X resonance, but also the known QCD $t\bar{t}$ signal as well as all the backgrounds that passed the cuts. The vector $\vec{\mu}$ can be rewritten as $\vec{\mu}/\mathcal{L} = \epsilon_X\sigma_X\vec{\mu}_X + \epsilon_1\sigma_1\vec{\nu}_1 + \ldots + \epsilon_N\sigma_N\vec{\nu_N}$, where $\vec{\mu}_X$ and $\vec{\nu_k}$ are the normalized $M_{t\bar{t}}$ distribution for X and the various N background (including QCD $t\bar{t}$) respectively, ϵ is the efficiency, σ the cross section for a given signal or background and $\mathcal{L}$ is the integrated luminosity.

Hopefully all the background distributions will be sufficiently similar[7] so that it will be possible to combine all the backgrounds together: $\vec{\mu} = \epsilon_X\sigma_X\vec{\mu}_X + \epsilon_\nu\sigma_\nu\vec{\nu}$. In that case, the quantity $\epsilon_\nu\sigma_\nu$ could be trivially evaluated by comparing $\vec{n}$ to $\vec{u}$ in the low $M_{t\bar{t}}$ region of the spectrum where we know there are no resonances. If this is not the case, the knowledge of all the cross section and their efficiency, or at least their relative σ and ϵ, of all dominant backgrounds will be needed. No matter what, at that point, the use of full MC simulation seems nescessary to find $\vec{\mu}$ since it accounts for detector effects.

In addition to the resolution on $M_{t\bar{t}}$ and the accuracy of the MC, the limiting factors in the reconstruction of a resonance X are the natural width of the resonance (Γ_X) and the accumulated integrated luminosity ($\mathcal{L}$). Note that Γ_X will be a significant factor only if it is greater than the experimental resolution of $M_{t\bar{t}}$. The upper limit on σ_X should go down as the square root of the integrated luminosity since $\vec{n}$ follows a Poisson distribution.

6.4 Additional consideration for real data analysis

One thing that has not been discussed explicitly yet, is the resolution of the jets (σ_{jets}) and the missing transverse energy ($\sigma_{\not{E}_T}$). These quantities need to be well known for equation (2) to make sense. In the MC simulation, the following parametrizations are used:

$$\frac{\sigma_{E_{jets}}}{E_{jets}} = \frac{1}{\sqrt{E_{jets}}} \oplus 0.1 \text{ and } \frac{\sigma_{\not{E}_T}}{\not{E}_T} = 0.5\sqrt{\sum \not{E}_T}, \qquad (6)$$

where the energy is in GeV. However, this is a rough estimate. Moreover, it is known that σ_{jets} will be a function of other variables such as η. Therefore, it will be important to establish more accurate resolution functions once real data will be available.

It will also be important to find ways to estimate the $t\bar{t}$ background in-situ. One way to do that is

[7] Up to now, it seems to be the case, but a more detailed analysis should be done to confirm it.

to establish different selection cuts that isolate the background instead of the signal. This will require MC studies.

7 Higher $M_{t\bar{t}}$ spectrum ($M_{t\bar{t}} > 1$ TeV)

7.1 Proposed methodology

The methodology for the higher $M_{t\bar{t}}$ spectrum has not been decided yet, but it will likely be quite different from the one used for the lower $M_{t\bar{t}}$ spectrum. Due to the very few events with high $M_{t\bar{t}}$, the efficiency of the analysis will need to be optimized. This will be done by sacrifying some purity through looser cuts. Some MC studies are planned to see what, if any, should be the χ^2_{cut}. The commissioning cuts will also be revisited. In particular, the lepton isolation requirement might be dropped since events with high $M_{t\bar{t}}$ are more collimated, meaning that the leptonic b-jet will often be closer than $\Delta R < 0.4$ to the lepton. A b-tagging requirement of one or two b-jets might be added in the event selection.

It is also known that higher $M_{t\bar{t}}$ events lead to merged jets [4], [12], [13]. Events with unidentified merged jets will result in bad signal. However, when the merged jets are identified, the signal can be recovered in principle. This is because the sum of two merged jet 4-vectors $(p_{\text{jet1}}, p_{\text{jet2}})$ is equal to their merger 4-vector (p_{merger}), $p_{\text{jet1}} + p_{\text{jet2}} = p_{\text{merger}}$. While this is certainly true for particle jets (sum of all the final particle decay 4-vectors coming from a parton), it does not necessarily hold for jets reconstructed in the detector (sum of all the energy topocluster (or tower) 4-vectors[8] determined by a jet algorithm). Topoclusters are cluster of calorimeter cells [11].

[8] An energy topocluster (or tower) is a massless 4-vector defined by the energy and position of the topocluster (tower) with respect to the primary vertex of interaction.

7.2 Frequency of merged jets

A detailed study of the frequency of $t\bar{t}$ events with a jet merger has already been done [14], [15], [16]. The analysis has been done using the cone 0.4 tower jet algorithm which sums all the tower 4-vectors within a radius of $R = \sqrt{\eta^2 + \phi^2} = 0.4$. One must however keep in mind that different jet algorithms would lead to different results.

It has been demonstrated that hadronic quarks emitted within $\Delta R < 0.5$ are almost always reconstructed as a single jet. Figure 16 shows the proportion of $t\bar{t}$ events for which this is the case as a function of $M_{t\bar{t}}$. At 2 TeV, already 40% of the events would be bad signal if nothing is done to identify merged jets. Note that this number is for the events passing the commisioning cuts, but there is no χ^2_{cut} applied, meaning that these events do not have their leptonic b-quark overlapping with their lepton. The percentage would be higher if those were included.

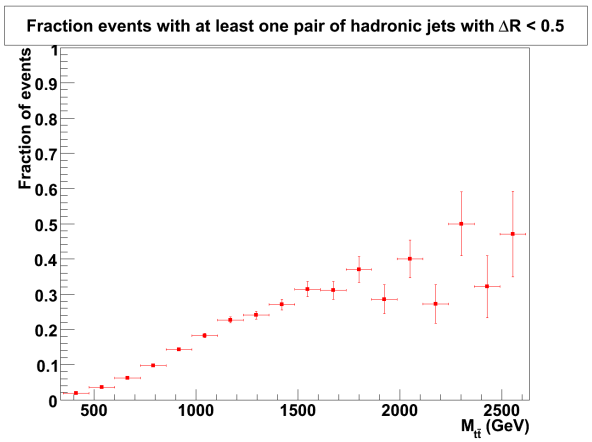


Figure 16: Fraction of events with at least one pair of hadronic quark with $\Delta R < 0.5$ as a function of $M_{t\bar{t}}$.

7.3 Identifying merged jets

There are many groups of physicist in ATLAS that already work on the jet merging problem caused by high p_{T} tops: the ATLAS Exotic Working Group created a sub-working group that focuses only on high

p_{T} bottom and top quarks. I plan to work in close collaboration with this new working group to identify the best tool(s) to find jet mergers. Some of the most promising paths to solve the problem are briefly enumerated in that subsection.

Although the top working group use tower to reconstruct their jets with satisfactory results in the low $M_{t\bar{t}}$ region, towers are quite unlikely to be useful to identify merged jets since they are less detailed (too big granularity) than topoclusters. The planned analyses are therefore with topoclusters.

An important part of the analysis will be to compare the different jet algorithms on topoclusters and see which one is the best to resolve collimated jets. The k_{T} algorithm is quite interesting in this respect since it offers a quantity called Y-scale [12]. This quantity gives an indication on how close were the two last groups of topoclusters to be merged to form the reconstructed jets. It therefore provides information on how likely the jet is a merger of two jets.

Some studies already showed [12] that, as expected, a jet merger from two jets of a W boson will have a mass close to the mass of the W. A similar result is found for a merger of the 3 jets from a top: the single reconstructed jet will yield a mass close to the one of the top. Therefore, the jet mass could be an efficient way to identify merged jets.

Another signature of a jet merger could be in the way the energy is distributed between the different topoclusters. Information about jet moments could consequently be useful quantity. Principal Component Analysis (PCA) has been shown to be also useful to identify merged jets [17].

Finally, information provided by the inner detector could be useful to resolve two jets that have been reconstructed as a single jet. For example, the number of tracks (charged particles) associated with a jet merger might be different than the number associated with a single jet for a given jet energy [13].

8 Research time line

Based on passed experience, my Ph.D. should take no more than five years to complete including the fact that I was allowed to transfer to a Ph.D. program without completing a M.Sc. thesis. The first eight months (September 2006 - May 2007) have been devoted to eight graduate courses; two are still required to meet the UVic Ph.D. standards. After the completion of these courses, most of my time has been spent on research using MC simulations. The frequency of merged jets in the high $M_{t\bar{t}}$ regime has been studied. The procedure to reconstructed the low $M_{t\bar{t}}$ spectrum has also been determined. Over the current summer (2008), most of the MC analysis related to low $M_{t\bar{t}}$ spectrum should be completed. This includes a method to probe low $M_{t\bar{t}}$ resonances as well as a way to identify the background in-situ. During fall 2008, ATLAS is expected to start collecting data. I will therefore move to CERN for an extensive period of time, about a year, to do shifts and help with the calibration of the detector using early data. In particular, I will join the jets and liquid argon calorimeter data quality monitoring since they both enter directly in the process of reconstructing $t\bar{t}$ events. It is hard to tell how long it will take before the data quality and integrated luminosity will allow applying the lower $M_{t\bar{t}}$ spectrum analysis to real data. For now, we estimate a period of about six months to two years. Meanwhile, it will be possible to do more MC study to validate the higher $M_{t\bar{t}}$ spectrum analysis. Hopefully, it will be possible to apply this analysis to real data after one or two years of data collection, leaving me enough time to write a thesis with the $M_{t\bar{t}}$ spectrum reconstructed up to few TeV.

Depending on how much integrated luminosity is achieved by 2009-2010 at ATLAS, the part concerning the reconstruction of the higher $M_{t\bar{t}}$ spectrum with real data might be abandoned. In the case that the integrated luminosity is not even sufficient to reconstruct the lower $M_{t\bar{t}}$ spectrum, other analyses can be considered. There are many interesting alternative studies that are related to $t\bar{t}$ physics that need less data than the reconstruction of the $M_{t\bar{t}}$ spectrum.

One of them is a careful determination of the W + jets cross section at ATLAS. This quantity is important since W + jets process is the main background for semi-leptonic $t\bar{t}$. Moreover, the W + jets cross section is predicted by the SM, but with a large uncertainty. It is therefore useful to get an accurate measurement of this quantity.

References

[1] W.-M. Yao et al. (Particle Data Group), J. Phys. G 33, 1 (2006) and 2007 partial update for the 2008 edition.

[2] R. Frederix and F. Maltonie, Top pair invariant mass distribution: a windows on new physics, 2007, hep-ph:0712.2355v1.

[3] R. M. Harris, C. T. Hill, S. J. Parke, Cross section for Topcolor Z′ decaying to top-antitop., 1999, hep-ph:9911288.

[4] B. Lillie, L. Randall and L.-T. Wang, The Bulk RS KK-gluon at the LHC, 2007, hep-ph:0701166v1.

[5] L. Fitzpatrick, J. Kaplan, L. Randall and L.-T. Wang, Searching for the Kaluza-Klein Graviton in Bulk RS Models, 2007, hep-ph:0701150v1.

[6] CDF Collaboration, Search for Resonant $t\bar{t}$ Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, 2007, hep-ex:0709.0705v1.

[7] S. Frixione, P. Nason and B.R. Webber, Matching NLO QCD and parton showers in heavy flavour production, JHEP 0308 (2003) 007 [hep-ph/0305252].

[8] G. Corcella *et al.*, HERWIG 6.5 Release Note, 2005, hep-ph/0210213.

[9] M.L. Mangago *et al.*, ALPGEN, a generator for hard multiparton processes in hadronic collisions, 2003, hep-ph/0206293.

[10] J-R. Lessard and M. Lefebvre, Top-antitop invariant mass spectrum, 2007, presentation to the ATLAS top working group: http://indico.cern.ch/conferenceDisplay.py?confId=25487.

[11] TileCal Group-IFIC(Valencia), Clustering of very low energy particles, ATL-CAL-INIT-2005-001, 29 July 2005.

[12] G. Brooijmans, High p_{T} Hadronic Top Quark Identification, Part 1:Jet Mass and YSplitter, ATL-COM-PHYS-2008-001.

[13] M. Vos, High p_{T} Hadronic Top Quark Identification, Part 2:the lifetime signature, ATL-PHYS-PUB-2008-000.

[14] J-R. Lessard and M. Lefebvre, Identifying Highly Energetic Tops, 2007, poster presented in the First ATLAS Physics Workshop of the Americas.

[15] J-R. Lessard and M. Lefebvre, Jets merging in top-antitop events, September 5th 2007, http://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=20527

[16] J-R. Lessard and M. Lefebvre, ATLAS top-antitop invariant mass spectrum reconstruction, presentation at the Atlas-Canada Physics Workshop, Montreal 2007, http://indico.cern.ch/conferenceOtherViews.py?view=standard&confId=20640

[17] Ben Smith, Recovering merged jets in semileptonic Zprime to ttbar decays, presentation in the High p_{T} b/top, http://indico.cern.ch/getFile.py/access?contribId=1&resId=1&materialId=slides&confId=30847, March 27th, 2008.