CMS PH-CMG-PS

Summer student programme 2014

Optimization of a search for squark with M_{T2} hadronic

Author:
Tommaso Radicioni

Supervisors:
Mariarosaria D'Alfonso
Filip Moortgat

Contents

Chapter 1

Analysis

1.1 Introduction

In the following analysis, CMS data simulation at 8 TeV are used to optimize the SUSY analysis in a simplified model in which the squark decay $\tilde{q} \longrightarrow$q+$\chi^0$ is considered. Hereinafter, SUSY and backgrounds events are studied by using M_{T2} variable, namely a peculiar transverse mass defined for SUSY particles, calculated in the hadronic channel. In a multijet event, jets are clustered into two distinct pseudojets used as input for M_{T2} calculation. Furthermore, I focussed my attention on the improvement of pseudojets reconstruction exploiting the flavour of jets and the different algorithms for combining jets.

1.2 Supersimmetry scenario

A broad class of extensions of the standard model (SM) predicts the existence of heavy particles that decay to hadronic final states accompanied by large missing transverse energy (E_T^{Miss}). The most popular scenario is supersymmetry (SUSY), a theory which could simply solve different unsolved questions in SM, namely:

- **Higgs mass corrections**: the corrections due to the SUSY particles subtracted to the corrections due to the SM particles make finite the Higgs mass value;

- **Unification between coupling costants**: at high energy, SUSY theory predicts an unification of all the coupling costants (electromagnetic, weak and strong);

- **dark matter candidate**: lightest particle in the SUSY model (called neutralino LSP or χ^0) is stable and, after a decays chain, all SUSY particle decays in the LSP.

Unfortunately this theory has several disadvantages: the value of free parameters in the theory are unknown and different specific models have been built to simplify SUSY theory by reducing the huge number of parameters.

1.3 Analysis description

This section is organized as follows: after a brief introduction to M_{T2} in Section 1.3.1 and to SUSY signal and background events in Section 1.3.2 and in Section 1.3.3, I will describe the conditions I applied to discriminate background from signal events in Section 1.3.4 and I will examine more deeply M_{T2} behaviour for compressed spectrum case, namely when mass difference between the mother particle ($\tilde{q}$) and one of the daughter particle (χ^0) is smallest.

1.3.1 M_{T2} definition

Since SUSY particles production is always accompanied by E_T^{Miss} it is possible to construct a variable from this quantity, called M_{T2} or stranverse mass, introduced to measure the mass of primary pair-produced particles in a situation where both ultimately decay into undetected particles (e.g. neutralino LSPs).[1] In particular, when two sparticles give rise to identical types of decay chains with two visible systems it is possible to define M_{T2} by jets transverse momenta $p_T^{\tilde{q}(i)}$, transverse energies $E_T^{\tilde{q}(i)}$ and masses $m^{\tilde{q}(i)}$ and by the unknown neutralino LSP transverse momenta $p_T^{\chi^{0(i)}}$. Two transverse masses (i=1,2) are so defined as:

$$(M_T^{(i)})^2 = (m^{vis(i)})^2 + m_\chi^2 + 2(E_T^{\tilde{q}(i)} E_T^{\chi^{0(i)}} - p_T^{\tilde{q}(i)} p_T^{\chi^{0(i)}}) \tag{1.1}$$

These have the property that for the true LSP mass their distribution cannot exceed the mass of the parent particle of the decay and they present an endpoint at the value of the parent mass. The momenta p_T of the unseen particles are not experimentally accessible individually and only their sum, the vector of the missing transverse momentum p_T, is known. M_{T2} is so defined as:

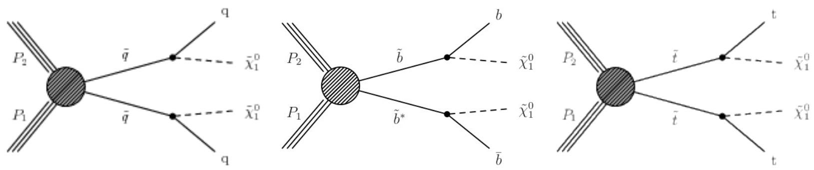
$$M_{T2}(m_\chi) = min_{p_T^{Miss}}[max(M_T^{(1)}, M_T^{(2)})] \tag{1.2}$$

where the LSP mass m_{χ^0} remains as a free parameter. As neither $M_T^{(1)}$ nor $M_T^{(2)}$ can exceed the parent mass if the true momenta are used, the larger of the two can be chosen. To make sure that M_{T2} does not exceed the parent mass, a minimization is performed on trial LSP momenta fulfilling the E_T^{miss} constraint. The distribution of M_{T2} for the correct value of m_{χ^0} then has an endpoint at the value of the primary particle mass.

1.3.2 SUSY benchmark models for fully hadronic search

In the following analysis, I'm going to examine one of the simplified SUSY events, namely the two-body decays represented in Figure 1.1:

[1] M_{T2} definition has been deduced from the paper CMS PAS SUS-13-019



(a) (b) (c)

Figure 1.1: Simplified diagram of a particular SUSY events with $\tilde{q} \longrightarrow$q+$\chi^0$ in Figure (1.1a), $\tilde{b} \longrightarrow$b+$\chi^0$ in Figure (1.1b) and $\tilde{t} \longrightarrow$t+$\chi^0$ in Figure (1.1c) where squarks and neutralino masses are unknown values

1.3.3 Main SUSY backgrounds for fully hadronic search

M_{T2} is one of the most important search variable since SUSY events with large expected E_T^{Miss} will be concentrated in the large M_{T2} region. In contrast, QCD dijet events, in which the two jets are back-to-back, populate the region of small M_{T2}, as shown in Figure 1.2:

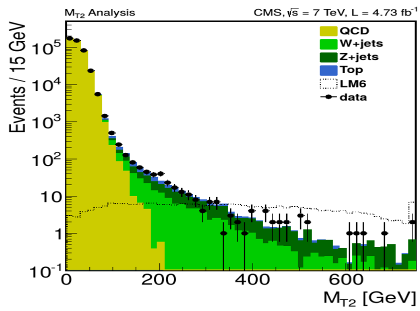


Figure 1.2: The M_{T2} distribution for the main SM backgrounds from simulation are stacked on top of each other. All distributions from simulation are normalized to the integrated luminosity of the data.

Other main SM backgrounds are:

- QCD multijet events with large E_T^{Miss} and M_{T2} values;
- Z+jets events with Z$\longrightarrow \nu\nu$;
- W+jets with the W decaying to lν, when the charged lepton is missed or absorbed into a jet;
- t$\bar{t}$ production, with W$\longrightarrow l\nu$, when the charged lepton is missed.

1.3.4 Preselection and further selection

In order to improve the M_{T2} efficiency in my analysis, first of all I imposed a first set of conditions to reduce as much as possible the background with respect to signal events. In this first selection, called preselection, I required that all the events must satisfy the following conditions:

1. **Jets**: all the jets, clustered using the anti-k_T jet clustering algorithm, are required to satisfy the condition p_T>20 GeV and at least two of them must have a momentum p_T>40 GeV;

2. **PU Jets**: all the Pile Up jets (PU), namely multiple pp collisions within a beam crossing, are neglected by imposing the FastJet pileup subtraction procedure;

3. **Missing energy**: only events such that E_T^{miss}>30 GeV are considered;

4. every event has to satisfy other conditions in order to suppress the contributions from W+jets, Z+jets and top-quark backgrounds, namely:

   - **No isolated leptons**: no e/μ with the following features:
     - Δz and Δxy less than 0.1;
     - p_T>10 and $\frac{p_T^{All}}{p_T^{Lep}}$<0.2;
     - $|\eta| \leq$2.4;
   - **No *good*** τ: no τ with $p_T > 20$ GeV;
   - **No isotrack**: no isotrack with momentum p_T>20 GeV (while for electrons and muons p_T>5 GeV) and such that $\frac{p_T^{All}}{p_T^{Iso}}$<0.2 (additional condition not in the previous analysis).

After that, I applied a further selection to improve particular aspects in the analysis, not necessarily linked to main backgrounds substraction described before:

1. H_T>450 GeV, namely the scalar sum of transverse momenta of reconstructed and energy-corrected calorimeter jets, to improve the trigger efficiency;

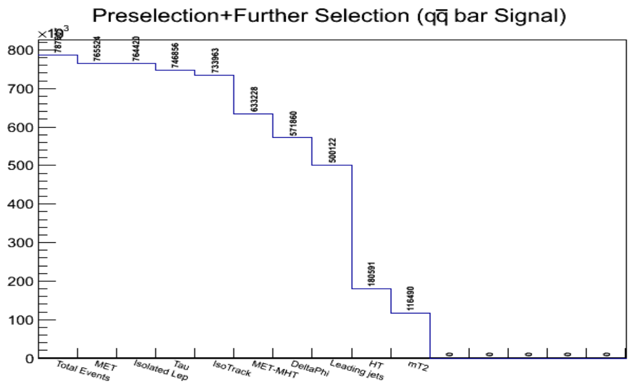
2. $|E_T^{Miss} - H_T^{Miss}|$<70 GeV to reject events where a significant fraction of the momentum imbalance arises from forward or soft jets;

3. $\Delta\Phi$(jets,E_T^{miss})>0.3 to reduce the background from QCD multijet events with large E_T^{miss}, arising from mismeasurements or leptonic heavy flavor decays;

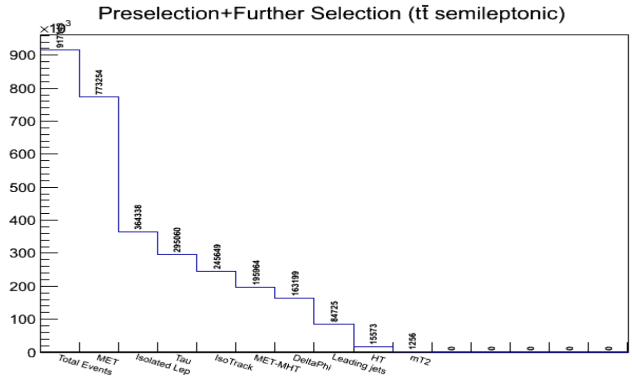
4. two leading jets with a momentum greater than 100 GeV;

5. M_{T2}>125 GeV to strongly suppress the QCD multijet background.

In the figures shown below, it is possible to observe the effect of these two selections for q$\bar{q}$ signal and for semileptonic t$\bar{t}$ background events. In particular, we notice that the percentage of signal events left after preselection is equal to ≈93% with respect to the total number of events while for the background is equal to ≈25%:



(a)



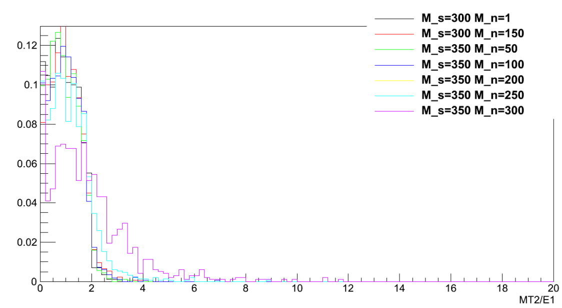
(b)

Figure 1.3: Preselection and further selection of the events for q$\bar{q}$ signal events in compressed spectrum ($m_{\tilde{q}}$,m_{χ^0})=(350,300) in Figure (1.3a) and for t$\bar{t}$ background events in Figure (1.3b)

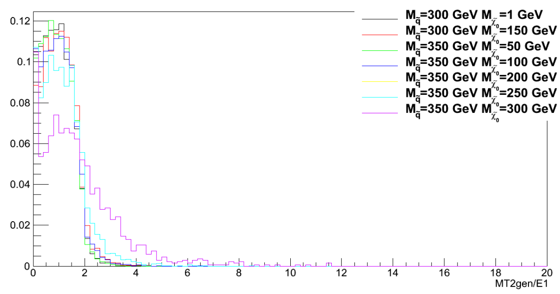
Moreover, M_{T2} has the same efficiency both after preselection and after further selection in the background events.

1.3.5 Compressed spectra deviations

In order to verify the correlation between M_{T2} calculation and the physical quantities in the SUSY scenario, I defined E_j as the jet energy which, considering the semplified case of two-body decay $\tilde{q} \longrightarrow$q+$\chi^0$, is equal to $E_j = \frac{m^2_{\tilde{q}} - m^2_{\chi^0}}{2m_{\tilde{q}}}$. This quantity is related to M_{T2} since it is defined starting from the same kinematic quantities and it is possible to observe this correlation in the figures shown below for different SUSY mass point:



(a)



(b)

Figure 1.4: $\frac{M_{T2}}{E_j}$ behaviour in q$\bar{q}$ signal events in Figure (1.4a) and $\frac{M_{T2(gen)}}{E_j}$ behaviour in q$\bar{q}$ signal events in Figure (1.4b)

The M_{T2} endpoint is related to the mass difference $m_{\tilde{q}}$-m_{χ_0} and it is possible to observe a larger tail in the compressed spectrum, namely in the line with the smallest mass difference. There isn't substantial differences with $M_{T2}(Gen)$, that is, M_{T2} calculated with E_T^{Miss} and PseudoJet at generator level. I studied more deeply the events with an high $\frac{M_{T2}}{E_j}$ in compressed spectrum and I concluded that they are linked to the ISR presence: the extra-jets spoiled the decay chain reconstruction and, for this reason, the correlation between M_{T2} and E_j is different from the other cases.

1.4 Analysis improvement

I focussed my attention on the improvement of pseudojets reconstruction exploiting the flavour of jets and the different algorithms for combining jets. In Section 1.4.1, I investigated different algorithm combinatorics. Furthermore, regarding the input to the pseudojets, the following additional points are examined:

1. specific flavour of final jets, for istance by exploiting b-jets (discussed more in Section 1.4.2);

2. presence of particular substructures in pseudojets (discussed more in Section 1.4.3):

    - Pseudojet mass;

    - Jet multiplicity inside pseudojets.

1.4.1 M_{T2} hemispheres algorithm

In order to improve this first analysis and to find other paths to discriminate background events, it is now necessary to deepen the understanding of M_{T2} calculation by introducing the concept of pseudojets. In a multijet event, jets are clustered into two distinct pseudojets and M_{T2} is defined starting with those. A method to subdivide an event into two pseudojets is the reconstruction of "event hemispheres"which works as follows:

- two initial axes (called seeds) are chosen as the directions of the two massless jets that have the largest dijet invariant mass;

- the other jets are associated to one of these axes according to a clustering criterion (hemisphere association method);

- after all jets are associated to one or the other axis, the axes are recalculated as the sum of the momenta of all jets connected to a hemisphere;

- the association is iterated using these new axes until no jet switches from one group to the other.

By modifying some aspects of the previous reconstruction method, different algorithm can be implemented. In the next table it's described the main differences between several algorithms used in the next analysis, called "Lund"(default algorithm), "MinMass", "k_T"and "Anti-k_T"(test algorithms):

	Lund	MinMass	Anti-k_T and k_T
Seeding	Seeds are chosen in order to maximize dijets invariant mass	Seeds are chosen in order to maximize dijets transverse invariant mass	seeds aren't chosen in any way
Association	minimal Lund distance	minimal mass square sum of the hemispheres	$d_{ij} = min(k_i^{2p}, k_j^{2p})\frac{\Delta_{ij}^2}{R^2}$ with p=±1

Table 1.1: Working principle of algorithms used in the analysis

The correlation between algorithms defined in Table 1.1 through M_{T2} correlation is represented in the figures shown below. We can notice that there aren't particular deviations between MinMass and Lund algorithms in the M_{T2} calculation. Instead, k_T and Anti-k_T are more uncorrelated with Lund algorithm:

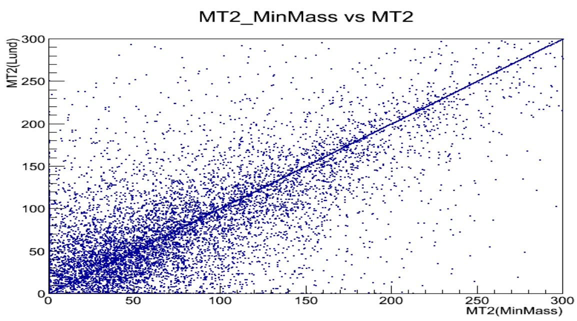
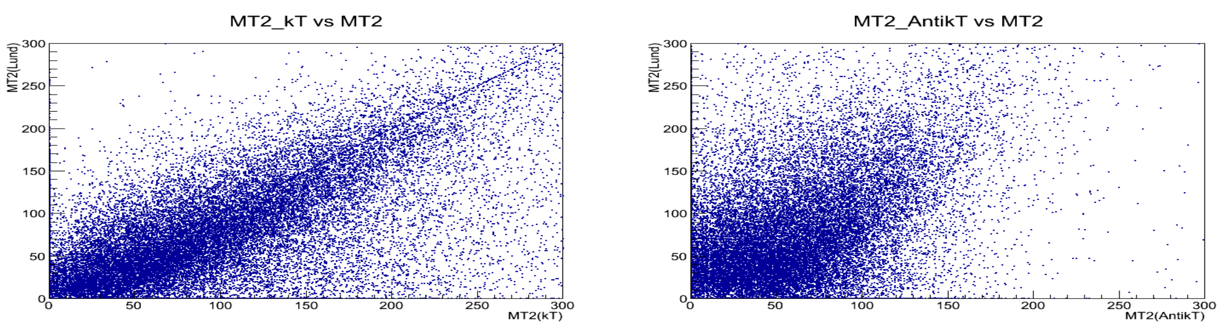


Figure 1.5: Correlation between M_{T2}(MinMass) and M_{T2}(Lund) for compressed spectrum ($m_{\tilde{t}}$,m_{χ^0})=(350,200) in t$\bar{t}$ signal events



(a) (b)

Figure 1.6: Correlation between $M_{T2}(k_T)$ and M_{T2}(Lund) in Figure (1.6a) and between M_{T2}(Anti-k_T) and M_{T2}(Lund) in Figure (1.6b) for compressed spectrum ($m_{\tilde{t}}$,m_{χ^0})=(350,200) in t$\bar{t}$ signal events

1.4.2 M_{T2}^{bb} analysis

I defined another variable starting from M_{T2} by considering, instead of all the jets, only the b jets and I called this variable M_{T2}^{bb}. In particular, I considered 2 b jets which pass b-tagging medium condition, analyzed trough jetbtagCSV value, or 1 b jet which pass this condition and the other one with the second greater value in jetbtagCSV. The goal of this part of the analysis is to use M_{T2}^{bb} distribution to improve the selection events. The conditions for the events analysis on M_{T2}^{bb}, applied to strongly suppress the QCD multijet background, are similar to M_{T2} condition, namely M_{T2}^{bb}>125 GeV. The advantages deriving from using M_{T2}^{bb} are represented hereunder for b$\bar{b}$ signal events. First of all, M_{T2}^{bb} solves the deviation of the endpoint in compressed spectrum: in the $\frac{M_{T2}^{bb}}{E_j}$ plot, the endpoint is indeed where we expected it should be. Moreover, I imposed M_{T2} and M_{T2}^{bb} cuts for signal and background varying the number of b jets and of the total jets and finally I concluded that M_{T2}^{bb} is more efficient than M_{T2} on equal terms:

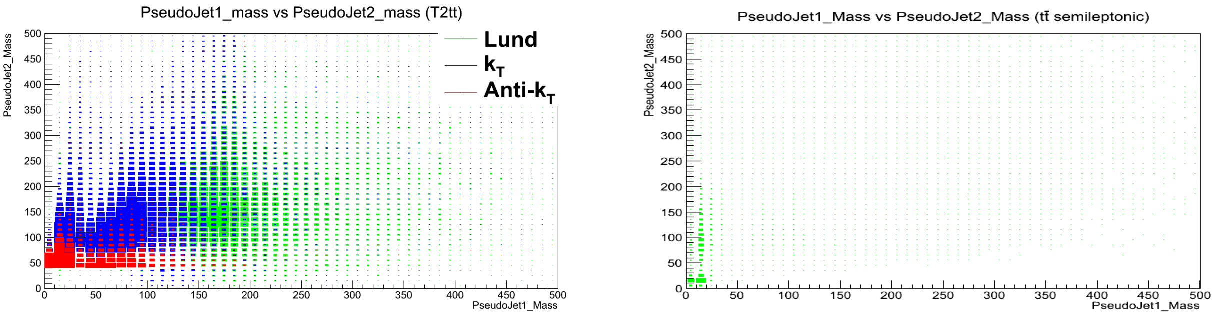
	#jets with 2 b jets			
Semileptonic t$\bar{t}$	2	3	4	5
$M_{T2} > 125$ GeV	2.87%	3.15%	4.49%	4.84%
$M_{T2}^{bb} > 125$ GeV	2.87%	1.93%	1.79%	2.19%
t$\bar{t}$ Signal	2	3	4	5
$M_{T2} > 125$ GeV	18.24%	15.84%	16.86%	18.5%
$M_{T2}^{bb} > 125$ GeV	18.24%	13.91%	12.54%	12.21%
b$\bar{b}$ Signal	2	3	4	5
$M_{T2} > 125$ GeV	44.55%	44.14%	46.24%	48.67%
$M_{T2}^{bb} > 125$ GeV	44.55%	43.09%	42.62%	41.17%

Table 1.2: Percentage (related to the number of events after preselection) of remained events after M_{T2} and M_{T2}^{bb} cuts applied after preselection for SUSY mass point equal to ($m_{\tilde{b}}$,m_{χ^0})=(375,300) in b$\bar{b}$ signal events and to ($m_{\tilde{t}}$,m_{χ^0})=(350,250) in t$\bar{t}$ signal events

In particular, in the events with 4 jets M_{T2}^{bb} has an efficiency, with respect to M_{T2} cut efficiency, equal to ≈90% in b$\bar{b}$ and ≈75% in t$\bar{t}$ signal events while it is equal to ≈40% in semileptonic t$\bar{t}$ background events.

1.4.3 Pseudojet substructures

In order to find particular internal substructures, I finally investigated the correlation between pseudojets masses for different algorithm as shown in the figures below for t$\bar{t}$ signal and t$\bar{t}$ semileptonic background events:



(a) (b)

Figure 1.7: Correlation between pseudojets masses for Lund, k_T and Anti-k_T algorithms in compressed spectrum ($m_{\tilde{t}}$,m_{χ^0})=(350,200) for t$\bar{t}$ signal events in Figure (1.7a) and for Lund algorithm in t$\bar{t}$ semileptonic events in Figure (1.7b)

In Figure 1.7a it is possible to observe that different algorithms behave in very different way: Lund algorithm recognizes two top in t$\bar{t}$ signal events while Anti-k_T and k_T behaviours have to still to be understood. In Figure 1.7b, substructures observed in t$\bar{t}$ signal events are different from those ones observed in t$\bar{t}$ semileptonic events when Lund algorithm is used. The difference between substructures recognition is due to the fact that, in signal events, two hadronic top give rise to a symmetric decay chain whereas the hadronic top in semileptonic t$\bar{t}$ background events gives rise to an asymmetric decay chain.

Chapter 2

Conclusion

From this analysis, it is quite evident that working on M_{T2}^{bb} analysis is useful to improve M_{T2} cuts in the studying of SUSY events and in particular to enhance the ratio $\frac{Signal}{Background}$ in events with more than 4 jets and that, even though I didn't find specific pseudojet substructures by analyzing jets multiplicity and the correlation between pseudojets mass and p_T, substructures analysis could help to recognize the differences between signal and background events. Finally, it could be useful looking at other combination scheme in order to prove that a different pseudojet reconstruction don't necessarily imply a different M_{T2} calculation.