Supersymmetry Searches in multijet events with CMS

Raffaele Tito D’Agnolo for the CMS Collaboration
Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa, Italy
raffaele.dagnolo@sns.it

Abstract. We describe the searches for Supersymmetry performed by the CMS collaboration in events with jets and significant missing transverse energy, characteristic of the decays of heavy, pair-produced squarks and gluinos. The results presented here have been obtained with the first 36 pb−1 collected by the CMS detector during the 2010 data taking period.

1. Introduction
One of the many challenges, that confront us at the LHC, is the lack of a precise intuition of how physics beyond the Standard Model will manifest itself. Even if we consider only a specific class of models, such as supersymmetric theories, many possibilities are left open in terms of experimental signatures. Furthermore these signatures might depend on the fine structure of the new particles spectrum. Therefore the CMS collaboration has designed inclusive analyses based on event topology. In addition to that a great effort has been made to construct data-driven methods to estimate the Standard Model (SM) backgrounds. This has led to the development of new techniques and the production of reliable limits on new physics observables.

In these early stages of the LHC running the most promising channels for the discovery of supersymmetry involve the strong production of squarks and gluinos. If we assume R-parity conservation, their decays will be characterized by the presence of jets and high missing momentum. There are three CMS analyses designed to be sensitive to these final states. The so called $\alpha_T$ [1] and $MH_T$ [2] analyses exploit mainly the high $MH_T$ in the signal events, that we define as the magnitude of the negative vectorial sum of the $p_T$s of the jets in the event. The third analysis [3], that we will call the Razor, adopts a different approach studying less traditional observables that take advantage of the kinematical properties of the pair production of the supersymmetric partners. The main SM backgrounds to these searches are: $Z(\rightarrow \nu\nu)$+jets, $W$+jets, $t\bar{t}$+jets and multijet events with large missing momentum from leptonic decays of heavy flavor hadrons inside jets, jet energy mismeasurement or instrumental noise and dead components.

After a brief description of the CMS detector in section [2] we examine the strategies adopted by the three all hadronic analyses to distinguish the production of supersymmetric particles from these backgrounds. In section [3] we review some of the data-driven techniques used to estimate the residual SM contribution in the signal regions and in the last section we show the results obtained with the first 36 pb−1 collected by the CMS detector during the 2010 running.
2. The CMS detector

The Compact Muon Solenoid has been described in detail in [4]. At its core is placed a 3.8 Tesla superconducting solenoid in which silicon strips and pixels, that constitute the main tracking apparatus, are contained. The charged particles momenta can be measured with a full azimuthal coverage and up to $|\eta| < 2.5$, where $\eta$ is the pseudorapidity, defined as $\eta = -\ln(\tan(\theta/2))$ and $\theta$ is the polar angle of the particle trajectory with respect to the counterclockwise proton beam direction. Outside of the solenoid the steel return yoke is instrumented with gas detectors used to identify muons.

A lead-tungstate crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL) surround the tracking volume and cover the region $|\eta| < 3$. An iron/quartz-fiber hadronic calorimeter in the forward region ($3 < |\eta| < 5$) contributes to the quasi-hermeticity of the detector to allow energy-balance measurements in the plane normal to the beam direction.

3. Analysis Strategies

The three CMS hadronic searches for supersymmetry do not rely on any assumption on the spectrum of the supersymmetric particles. The $\alpha_T$ and $M_{HT}$ analyses explore the tails of the $M_{HT}$ with two very different approaches towards jets mismeasurement and detector effects, while the Razor takes advantage of the properties of pair production of heavy particles. In this section we would like to describe the ideas behind the different strategies without entering into the details of each analysis, for which we refer to CMS papers and public physics analysis summaries in the bibliography.

3.1. $\alpha_T$

In this search the variable $\alpha_T$, inspired by [6] and first introduced in [7], was used as the main discriminator between events with real and fake missing transverse energy. Events with at least two calorimetric jets with a cone parameter of 0.5, clustered using the anti-$k_T$ algorithm [8] and with $E_T > 100$ GeV, are selected, requiring the sub-leading jets to have at least an $E_T$ of 50 GeV. Additional cuts on the $H_T$ and the consistency of the jet based estimate of the $M_{ET}$ are applied and a QCD-killer cut, $\alpha_T > 0.55$, is performed. This last cut is particularly effective in reducing the background, since for perfectly measured dijet events $\alpha_T=0.5$, while an imbalance of the $E_T$’s can only make $\alpha_T$ smaller than 0.5. Advantages and drawbacks of this approach
emerge clearly from figure 1(a) where it is easy to see how the cut on $\alpha_T$ is suitable for a fast discovery. On one hand it allows to ignore to a large extent the difficulties associated to cleaning the ME$_T$ tails and is very effective in reducing the backgrounds, on the other it is not very efficient for typical mSUGRA [9–12] signals.

3.2. MH$_T$
This analysis focuses on final states with at least three jets and high missing momentum. The baseline selection that defines this class of events consists in asking for

- $\geq 3$ Jets with $|\eta| < 2.5$ and $p_T > 50$ GeV
- $H_T > 300$ GeV
- $MH_T > 150$ GeV
- $\Delta\phi$(Jet$_{1,2}$, MH$_T$) > 0.5
- $\Delta\phi$(Jet$_3$, MH$_T$) > 0.3
- Plus event cleaning cuts

where we have indicated with Jet$_{1,2}$ the two leading jets and with Jet$_3$ the additional jet required. In order to suppress the high MH$_T$ tails of QCD multijets events, all the jets with $|\eta| < 5$ and $p_T > 30$ GeV are used in its measurement.

After the baseline selection two search regions are defined

(i) High $H_T$, where $H_T > 500$ GeV
(ii) High MH$_T$, where MH$_T > 250$ GeV

All the observables mentioned above, from jets to the MH$_T$ in figure 1(b) have been reconstructed using the Particle Flow algorithm [5]. The jets have a cone parameter of 0.5 and have been clustered using the anti-$k_T$ algorithm. It would be worth spending several pages on the Particle Flow technique, however for our purposes it is enough to mention that the jet energy resolution improves considerably with respect to calorimetric jets for $p_T < 100$ GeV and that the MH$_T$ tails are less prone to fakes coming from jet mismeasurements. This has allowed to design the analysis retaining a good efficiency for models where the masses of the new particles are low enough to be produced with sizable yield at limited integrated luminosities.

The very high efficiency on the signal is one of the main strengths of this approach, as will emerge in section 5.

3.3. The Razor
The pair production of two heavy particles, each decaying to an unseen LSP plus jets, is studied using the idea of event hemispheres following [13]. All the reconstructed objects in each hemisphere are combined into a single mega-jet. Therefore all events are forced into a dijet-like topology. In this scenario it is natural to introduce the $R$-frame, which is the longitudinally boosted frame that equalizes the magnitude of the two mega-jets 3-momenta. If the two heavy particles are produced approximately at rest, the $R$-frame is simply the center of mass frame of the collision. In this case, the magnitude of one of the two 3-momenta is proportional to an event by event estimate of the scale $M_\Delta$

$$|\vec{p}_{hem_{1,2}}^R| = \frac{M_\Delta}{2} = \frac{M_{heavy}^2 - M_{LSP}^2}{M_{heavy}},$$

so it is natural to define the new variable

$$M_R \equiv 2|\vec{p}_{hem_{1,2}}^R|.$$
Figure 2. [2(a)] Simulated SUSY events with parameter set LM1, in the \((R,M_R)\) plane. [2(b)] Simulated QCD multijet events in the \((R,M_R)\) plane. [2(c)] Simulated \(W\)+jets events in the \((R,M_R)\) plane. [2(d)] Simulated \(t\bar{t}\)+jets events in the \((R,M_R)\) plane.

It is clear that \(M_R\) peaks around a characteristic scale. For example a QCD dijet event has \(M_R \sim \sqrt{\hat{s}}\), whereas the production of a \(W\) boson, recoiling against a multijet system, lives in the vicinity of \(M_R \sim M_W\).

The definition of \(M_R\) will indeed prove very useful, but it relies only on a subset of the full kinematical information in the event. If we use also the unseen LSPs as a source of background rejection we can define another kinematical observable as

\[
R \equiv \frac{M_R^T}{M_R},
\]

where \(M_R^T\) is the transverse variable:

\[
M_R^T \equiv \sqrt{|M_E T| (p_T^{hem1} + p_T^{hem2}) - M_E T \cdot (\vec{p}_T^{hem1} + \vec{p}_T^{hem2})}. \tag{4}
\]

Since \(M_R^T\) is a transverse version of \(M_R\), it can be proven that \(R\) peaks at 0.5 and has its endpoint at \(M_\Delta\). In figure 2 we show the distribution of SM and SUSY events in the \((R,M_R)\) plane. Our intuitions about \(R\) and \(M_R\) are confirmed and it is clear that a pair of cuts in that plane would be enough to obtain a very pure signal region, without further requirements on
other observables. We will review the chosen values for the cuts while presenting the results. In the next section we will see how these two variables allow a fully data-driven determination of the surviving background.

4. Background determination

All the analyses devoted to supersymmetric searches at CMS have designed independent data-driven methods to estimate the backgrounds in the signal region. These techniques will be more and more performing as the statistics increases. In this way our future sensitivity will grow much faster than $\sqrt{\int L dt}$. In addition to that, these data-driven methods are intrinsically easy to adapt to different searches, as we will see in the following.

In the next subsection we will review the main strategies used for the three analyses, without aiming to completeness, but rather trying to give a general picture.

4.1. $\alpha_T$

The total background is estimated from two control regions of low $H_T$. The method is based on the ratio $R_{\alpha_T}$ of events which pass and fail the full selection requirements. This ratio can be extrapolated from the two low $H_T$ regions to the signal region, by virtue of its independence from $H_T$,

$$\frac{R_{\alpha_T}(HT300)}{R_{\alpha_T}(HT250)} = \frac{R_{\alpha_T}(HT350)}{R_{\alpha_T}(HT300)}$$

for events with real $M_{H_T}$. This can be seen in figure 3 where $R_{\alpha_T}$ is shown to be constant in $H_T$ for the electroweak backgrounds, while it decreases slowly for QCD events, where the missing energy is mostly fake.

In addition to this procedure, the $W+\text{jets}$ and $t\bar{t}+\text{jets}$ backgrounds can be determined independently from a muon sample, while the $Z(\rightarrow \nu\nu)+\text{jets}$ background can be estimated from a $\gamma+\text{jets}$ sample. The two different methods, the inclusive one based on $R_{\alpha_T}$ and the exclusive ones for the electroweak backgrounds, give consistent results as will be shown in section 5.

4.2. $M_{H_T}$

A number of novel techniques have been developed to measure the background for this search in a fully data-driven way. In this limited space we will only sketch briefly the Rebalance+Smear method intended to assess the QCD multijet background. The R+S method makes a prediction by starting from a sample of events consisting of seed jets. The events are produced using an
inclusive multijet data sample as input. At first all of the jet momenta are adjusted to return the event back into approximate transverse momentum balance. Then, in the Smear step, a random value of the jet response is drawn from the jet resolution distribution for each seed jet, and its true momentum is scaled by this factor. As the input to the smearing step is representative of a pure QCD multijet sample, the resulting events can be used to apply search cuts.

As mentioned above, one of the advantages of data-driven techniques is that they are often “analysis-independent”. For instance, in this case, the electroweak backgrounds have been estimated in analogy to what we have already seen for the $\alpha_T$ analysis. The details of the techniques are different, but the key idea of exploiting the approximate equivalence of electroweak bosons at high energies remains. So the $Z(\rightarrow\nu\nu)+$jets background can be modeled using either a $\gamma+\text{jets}$ sample or a $W+\text{jets}$ sample where the muon is treated as a neutrino.

It is not unreasonable to object to the choice of purely data-driven methods. For example in the R+S technique the jet resolution distribution must me modeled using data and nothing assures us a priori that the total error would be smaller than that coming from simply taking it from the Montecarlo. However a very thorough work has been conducted, the details of which can be found in [2] and in an upcoming paper, assuring a good sensitivity to the analysis as we will see in section 5.

4.3. The Razor
As noticed in the previous section $M_R$ and $R$ are strongly correlated and their correlation can be used to estimate the residual background. The $M_R$ distribution for all the relevant backgrounds was found to fall exponentially after a turn on determined by the scale of the process under study. In general the exponential tail has a single component whose slope has a simple functional dependence on the cut on $R$

$$S = a + b(R_{cut})^2.$$ 

(6)

This allows to extrapolate the number of background events from regions of low $M_R$, where the SM is known to dominate, to the signal region. Therefore to estimate the background it is
sufficient to select exclusive boxes where the single components of the SM dominate and then extrapolate from low $M_R$. These boxes are an hadronic one for QCD and a muon and an electron box for the electroweak backgrounds. Their complete definition can be found in [3]. In figure 4 the behaviour of $M_R$ is shown for data, together with the results of the fit to the exponential slope.

5. Results
The three analyses have not found any significant excess with respect to the expected yield from SM processes. Therefore we can only set limits in the SUSY parameter space. As an example, we show the number of events found by the $\alpha_T$ analysis with the first 36 pb$^{-1}$ collected by the CMS detector.

- In the data we are left with 13 events.
- For the backgrounds there is the inclusive prediction from $R_{\alpha_T}$: $9.4^{+4.0}_{-4.8}$ (stat)$\pm1.0$ (syst).
- The exclusive predictions for the electroweak backgrounds. $W$+jets and $t\bar{t}$+jets, from a muon+jets sample: $6.1^{+2.8}_{-1.9}$ (stat)$\pm1.8$ (syst). $Z(\rightarrow \nu\nu)$+jets, from a $\gamma$+jets sample: $4.4^{+2.3}_{-1.6}$ (stat)$\pm1.8$ (syst).

We can easily see that the sum of the two electroweak components is consistent with the inclusive prediction, which is in turn compatible with the number of events from the data and with a Montecarlo prediction of $9.3\pm0.9$ (stat). The Razor and MH$_T$ analyses have found the same degree of compatibility between data and SM predictions, but with different yields. The Razor with $R > 0.5$ and $M_R > 500$ GeV sees 7 events in the hadronic box and, in the muon box, for slightly relaxed cuts on the two variables, just 3.

In figure 5 we show a comparison between limits obtained by different CMS analyses. It is easy to see that the MH$_T$ setup gives a limit similar to the Razor, whereas the cut on $\alpha_T$ affects the sensitivity of the analysis. The limit set by the MH$_T$ strategy is dominated by the high MH$_T$ search region defined in section 3 whereas the high $H_T$ region is more prone to background contamination giving a smaller $S/B$.

We have shown that the CMS SUSY hadronic analyses are performing well and that their sensitivity will increase non trivially with 2011 data. In addition to that several novel background estimation methods have been designed, that might be relevant also for future searches and make us confident in the robustness of our results. In one sentence: we are ready for a supersymmetric year.

References
[1] Khachatryan V et al. (CMS Collaboration) 2011 Phys.Lett. B698 196–218 * Temporary entry * (Preprint 1101.1628)
[2] Khachatryan V et al. (CMS Collaboration) 2011 CMS-PAS-SUS-10-005
[3] Khachatryan V et al. (CMS Collaboration) 2011 CMS-PAS-SUS-10-009
[4] Adolphi R et al. (CMS Collaboration) 2008 JINST 3 S08004
[5] Khachatryan V et al. (CMS Collaboration) 2009 CMS-PAS-PFT-09-001
[6] Randall L and Tucker-Smith D 2008 Phys.Rev.Lett. 101 221803 (Preprint 0806.1049)
[7] Khachatryan V et al. (CMS Collaboration) 2009 CMS-PAS-SUS-09-001
[8] Cacciari M, Salam G P and Soyez G 2008 JHEP 0804 063 (Preprint 0802.1189)
[9] Chamseddine A H, Arkani H, R L and Nath P 1982 Phys.Rev.Lett. 49 970
[10] Barbieri R, Ferrara S and Savoy C A 1982 Phys.Lett. B119 343
[11] Hall L J, Lykken J D and Weinberg S 1983 Phys.Rev. D27 2359–2378
Figure 5. Observed limits from 2010 CMS supersymmetry searches in the mSUGRA ($m_0, m_{1/2}$) plane.

[12] Nath P, Arnowitt R L and Chamseddine A H 1983 *Nucl. Phys.* B227 121

[13] Rogan C 2010 (*Preprint* 1006.2727)