IL NUOVO CIMENTO Vol. 32 C, N. 5-6 Settembre-Dicembre 2009
DOI 10.1393/ncc/i2010-10553-9

Colloquia: LaThuile09

Search for supersymmetry in di-jet events with novel data-driven background estimation

H. U. Flaecher on behalf of the CMS Collaboration
PH Department, CERN - CH-1211 Geneva 23, Switzerland

(ricevuto il 10 Novembre 2009; pubblicato online il 25 Gennaio 2010)

Summary. — We present a search for supersymmetry (SUSY) in di-jet events with the CMS detector at the LHC. Our study is focused on a SUSY parameter space where squarks are pair produced and both directly decay to a quark and neutralino with the latter escaping the detector, thus leaving a missing energy signature. Although the background from QCD di-jet events is overwhelming, the particular kinematics of the SUSY events allow to define powerful discriminating variables which enable a clear separation of signal and QCD events. Therefore, the only important SM background left for this search is the invisible decay of the Z boson accompanied with two jets in the final state. This background can be estimated by utilizing a novel approach using gamma+jet events thus enabling a possible discovery of SUSY in the di-jet system with the early physics data.

PACS 11.30.Pb – Supersymmetry.
PACS 12.38.-t – Quantum chromodynamics.

1. – Introduction

We present a search strategy for a possible discovery of supersymmetric (SUSY) signatures at the LHC using di-jet events following a recently proposed new approach [1]. It is based on the assumption that squarks are pair-produced and subsequently decay directly to a quark and the χ_1^0, the lightest stable particle (LSP). This approach is most promising for regions in SUSY parameter space where squarks have large branching fractions to decay directly to the LSP. This configuration in turn requires the gluino to be heavier than the squarks, thus avoiding cascade decays of squarks via the gluino. Therefore the event topology under investigation consists of two high-p_{T} jets and two invisible neutralinos which lead to a missing energy signature. The main background processes for this topology are QCD di-jet events and Z + jet events where the Z decays into two neutrinos. It is however possible to define kinematic variables that can discriminate between signal and background without relying on the missing energy measurement from the calorimeters. The presented analysis [2] is an extension to the existing SUSY searches within CMS which so far have focused on missing E_{T} signatures with at least three jets and/or involving charged leptons [3].

2. – Di-jet search using the kinematic variable $\alpha_{(\mathrm{T})}$

2·1. *Event selection.* – The results presented here are carried out assuming an integrated luminosity of $1\,\mathrm{fb}^{-1}$ collected at a LHC centre-of-mass energy of 14 TeV. The Monte Carlo samples considered consist of QCD processes generated using `PYTHIA` [4], including minimum bias and high-energy jet data. Further backgrounds are $\mathrm{t\bar{t}}$ + jets, W + jets, and Z + jets events (excluding $\mathrm{Z} \to \nu\nu$), all simulated using `ALPGEN` [5] and Z + jets events with $\mathrm{Z} \to \nu\nu$, generated with `PYTHIA`. In addition, single top, γ + jets and $\mathrm{b\bar{b}}$ + jets background samples were investigated which, however, only play a negligible role in the presented search. Possible SUSY signal yields are estimated using the CMS low-mass mSuGra points LM1–LM4 [3].

We select events that pass a two-jet trigger where the Level 1 trigger conditions are either one jet with E_T greater than 150 GeV or two jets with E_T greater than 70 GeV. At the High Level Trigger this cut is raised to two jets with each E_T greater than 150 GeV. For calorimeter jet clustering, the corrected iterative cone algorithm with $R = 0.5$ is used and two jets with $p_\mathrm{T} > 50\,\mathrm{GeV}$ and the electromagnetic fraction $F_\mathrm{em} < 0.9$ are required. Based on the two leading jets two additional variables are defined: HT as the scalar sum of the two leading jet p_T's, $\mathrm{HT} = p_\mathrm{T}^{\mathrm{j1}} + p_\mathrm{T}^{\mathrm{j2}}$ and missing p_T (MHT) of the event calculated as $\mathrm{M}\vec{\mathrm{H}}\mathrm{T} = -(\vec{p}_\mathrm{T}^{\mathrm{j1}} + \vec{p}_\mathrm{T}^{\mathrm{j2}})$. In order to select clean di-jet events, any event where either an isolated electron or muon with momentum $p_\mathrm{T} > 10\,\mathrm{GeV}$ was identified is vetoed. Furthermore, events with any additional jets with $p_\mathrm{T} > 50\,\mathrm{GeV}$, which also includes jets from hadronic τ decays, are also vetoed. To protect against significant mismeasurements of jet energies, events where the missing p_T based on the two-jet system points into the same direction as one of the first three jets, are rejected by requiring $\Delta\phi(\mathrm{jet}, \mathrm{MHT}) < 0.3\,\mathrm{rad}$. This definition of missing E_T based on the two jet momenta should also be robust against fake signals and noise in the calorimeters. In addition to the selection criteria above, the leading jet must be within pseudo-rapidity $|\eta| < 2.5$.

2·2. *Event kinematics.* – As mentioned in the previous section the di-jet trigger requires already two (uncorrected) jets with $p_\mathrm{T} > 150\,\mathrm{GeV}$ each, which implies $\mathrm{HT} > 300\,\mathrm{GeV}$. For signal events, two high-p_T jets come directly from a squark decay with typical mass of the order of 500 GeV. Therefore, to make the analysis cuts more restrictive than the trigger and to further reduce background contributions it is required that HT exceeds 500 GeV. Even after requiring two high-p_T jets, sizeable background contributions from a number of processes remain, the most important of which are

- QCD di-jet events due to their (overwhelmingly) large cross-section and sizeable uncertainties in higher-oder corrections, in particular production of extra jets due to gluon emission;
- $\mathrm{Z} \to \nu\nu$ events which present an irreducible background as the invisible Z decay leads to real missing E_T;
- W + jets events, with $\mathrm{W} \to \tau\nu$ followed by a hadronic τ decay which is wrongly identified as a jet.

It is however possible to define kinematic variables to disentangle QCD events and signal-like events with real missing E_T. In well-measured QCD di-jet events, transverse momentum conservation requires the p_T of the two jets to be of equal magnitude and back-to-back in the plane transverse to the beam. In contrast, in signal-like events the two squarks decay independently of each other and therefore the resulting jet p_T's can

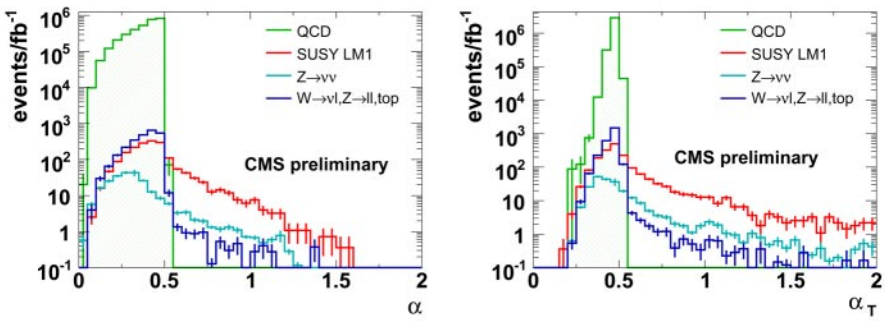

Fig. 1. – Distribution of α and α_{T} after all other selection cuts have been applied.

be of different magnitude and their ϕ values (largely) uncorrelated. In ref. [1], a new variable α was suggested which exploits the requirement of back-to-back jets of equal magnitude for QCD events:

(1)
$$\alpha = E_{\mathrm{T}}^{\mathrm{j2}}/M_{\mathrm{j1,j2}}, \text{ which for massless particles is equal to } \alpha = \frac{E_{\mathrm{T}}^{\mathrm{j2}}}{\sqrt{2E^{\mathrm{j1}}E^{\mathrm{j2}}(1-\cos\Theta)}},$$

where Θ is the angle between the two jets. As can be seen from eq. (1), α can at most have a value of 0.5 for well-measured QCD events. In addition, as the E_{T} of the second energetic jet enters in the numerator, uncertainties introduced through energy mis-measurements partly cancel out in α. (If one of the two jet energies is mis-measured by a large amount, the order of the two jets is reversed.) A modified version of this variable is also explored in which the transverse mass of the two jets is used instead of the invariant mass:

$$\alpha_{\mathrm{T}} = E_{\mathrm{T}}^{\mathrm{j2}}/M_{\mathrm{T_{j1,j2}}} = \frac{E_{\mathrm{T}}^{\mathrm{j2}}}{\sqrt{2E_{\mathrm{T}}^{\mathrm{j1}}E_{\mathrm{T}}^{\mathrm{j2}}(1-\cos\Delta\phi)}} = \frac{\sqrt{E_{\mathrm{T}}^{\mathrm{j2}}/E_{\mathrm{T}}^{\mathrm{j1}}}}{\sqrt{2(1-\cos\Delta\phi)}}, \tag{2}$$

where $\Delta\phi$ is the difference in azimuthal between the two jets. For well-measured QCD di-jet events, α_{T} is exactly 0.5. The α and α_{T} distributions are shown in fig. 1 for the different background processes and exemplary for LM1.

While the present selection is safe with respect to the effects of hard extra gluon radiation by rejecting events with extra jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$, multiple soft gluon emission might still noticeably affect the $\Delta\phi$ distribution. It is therefore safer to use α and α_{T} in the event selection as these variables also reject events where the p_{T} of the two jets is not balanced, in addition to being sensitive to the $\Delta\phi$ between the two jets. Compared with $\Delta\phi$, α and α_{T} have the additional benefit that they are more effective in rejecting $\mathrm{Z} \to \nu\nu$ events. In the following α and α_{T}, shown in fig. 1 are used in the event selection. Both variables are highly correlated to $\Delta\phi$, *i.e.* an additional cut on $\Delta\phi$, has a negligible effect. To account for finite jet energy and ϕ resolution as well as missed jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ it is required that α (α_{T}) exceeds 0.55.

2·3. *Expected event yields from simulations.* – After the selection criteria described above are applied, the event yields listed in table I are obtained for background events

TABLE I. – *Numbers of expected events after each selection cut for background samples (QCD, $t\bar{t}$, W, Z + jets, and $Z \to \nu\nu$) and LM1 signal point. The final numbers of events selected are shown after a cut on α or alternatively α_T and $\Delta\phi_{j1,j2}$.*

Selection cut	QCD	$t\bar{t}$, W, Z	$Z \to \nu\nu$	LM1
Trigger	1.1×10^8	147892	1807	25772
Preselection	3.4×10^7	9820	878	2408
$HT > 500\,\mathrm{GeV}$	3.2×10^6	2404	243	1784
$\alpha > 0.55$	0	7.2	19.7	227.6
$\alpha_T > 0.55$	0	19.9	58.2	439.6
$\Delta\phi_{j1,j2} < 2\pi/3$	0	18.7	57.2	432.4

and the LM1 signal point. All the numbers correspond to an integrated luminosity of $1\,\mathrm{fb}^{-1}$. Both α and α_T are very effective in reducing the backgrounds, particularly from QCD di-jet events but also for electroweak processes. When α_T is used instead of α, the signal yield for the LM1 point is almost doubled. The dominant background from $Z \to \nu\nu$ however rises by about a factor three while the background from $t\bar{t}$, W, and Z decays doubles as well. It is therefore proposed to study both variables with real data as the signal-to-background ratio differs in the two cases. Nevertheless, in each case signal-over-background ratios larger than five are expected.

Beside the mSuGra point LM1, the event yields for the low-mass SUSY points LM2–LM4 were studied. We find signal yields of 132 events for LM2, 138 events for LM3 and 195 events for LM4, where the selection on α_T was used. Accordingly a signal-over-background ratio in excess of 2 can be achieved for LM4 while for LM2 and LM3, the signal would still dominate over the expected background.

2˙4. *Jet-energy scale and resolution uncertainties.* – The systematic uncertainties due to miscalibration and mismeasurement of jets were estimated by applying the following systematic variations:

- a Gaussian smearing of the transverse jet momenta of 10% and a Gaussian smearing of the azimuthal angle ϕ by 0.1 rad;
- a scaling of the jet energy scale by ±5%;
- a scaling of the jet energy scale in the forward direction ($|\eta| > 1.4$) by ±3%.

It was found that the Gaussian smearing only has a small effect ($\sim 3\%$) on the selected signal and background events. The upward scaling of the transverse momenta of the jets effectively relaxes the HT cut and hence more events pass the selection. Conversely, the reduced momentum therefore leads to fewer events. The largest deviation is a 12% reduction in both, the signal and background efficiencies, leaving the signal-over-background ratio largely unchanged. The miscalibration applied for the jet energy scale in the forward regions has a negligible effect. Overall the signal-to-background ratio remains stable under varying conditions and the background from QCD remains small in all scenarios.

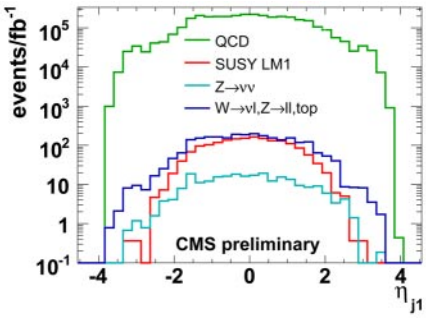

Fig. 2. – Distribution of η for QCD, $t\bar{t}$, W, Z, and SUSY LM1 events. Shown is the expected number of events for a luminosity of 1 fb^{-1}, after all selection cuts except the cut on α_T and $|\eta_{j1}|$.

3. – Data-driven background estimation

In the following we outline data-driven methods for background estimation for jets + missing energy searches. The main emphasis is on an approach where signal enhanced and depleted regions in phase space are defined and the combination of all backgrounds can be estimated simultaneously. In addition, we discuss how the dominant remaining background from $Z \to \nu\nu$ events can be estimated by using a data control sample of photon + jets events.

3˙1. *Background estimation using the η dependence of α_T via the matrix method.* – The idea of the matrix method is to find two variables, $\mathcal{V}_1$ and $\mathcal{V}_2$, which are uncorrelated for background events and for which in the 2-d plane three quadrants exists that are signal depleted and one that is signal enriched, *i.e.* each variable has a signal-enriched and a signal-depleted region. In this case it is possible to determine the amount of background events directly from the data. The two variables in question for the present analysis are the pseudo-rapidity $|\eta|$ of the leading jet and α_T. As can be seen from fig. 2, the leading jet from a SUSY event is on average more central than those from the background processes, QCD, $t\bar{t}$, W, Z + jets and $Z \to \nu\nu$. Therefore, the forward regions with $|\eta| > 2.5$ are considered as signal depleted. Similarly, the region formed by $\alpha_T > 0.55$ is signal enriched while that with $\alpha_T < 0.55$ is signal depleted.

The variable $R^i_{\alpha_T} = N_{\alpha_T>0.55}/N_{\alpha_T<0.55}$ is defined as the ratio of events with $\alpha_T > 0.55$ to those with $\alpha_T < 0.55$ for a given bin i in $|\eta|$. For the method outlined above to be applicable this ratio needs to be constant. While in real data it will not be possible to distinguish the different background processes on an event-by-event basis, Monte Carlo simulation shows that R_{α_T} is, to a good approximation, constant for all the relevant individual background contributions. It is therefore legitimate to combine all the backgrounds and to determine the sum of all backgrounds with the help of the matrix method. In fig. 3, R_{α_T} is shown for all backgrounds combined and as expected the "combined R_{α_T}" is flat as a function of $|\eta|$. In addition, R_{α_T} is shown for the case of a LM1 signal present in the data.

To estimate the number of background events in the $|\eta| < 2.5$ regions, $N_{\mathrm{pred}}(|\eta|)$, R_{α_T} needs to be multiplied with the number of events with $\alpha_T < 0.55$, $N_{\mathrm{bkgd}}(|\eta|)$, in the corresponding $|\eta|$ bin: $N_{\mathrm{pred}}(|\eta|) = R_{\alpha_T} \cdot N_{\mathrm{bkgd}}(|\eta|)$.

Figure 4 shows the numbers of background events predicted and measured in the different η regions after all selection cuts. In the absence of a signal, the background

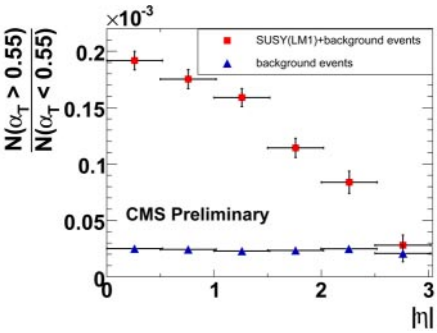

Fig. 3. – (Colour on-line) $R_{\alpha_{\mathrm{T}}}$ as a function of $|\eta|$ of the leading jet after all selection cuts except the cut on α_{T}, and $|\eta|$, shown for background events only (blue triangles) and for a mixture of background and SUSY LM1 events (red squares).

can be predicted within the simulated statistical precision. In fig. 4, the total number of signal-plus-background events is also compared to the number of predicted background events. The presence of a SUSY signal leads to a slight overestimate of the background. Despite the large statistical uncertainty on the background prediction, a clear signal is still visible. The stability of the presented matrix method was verified against the systematic variations discussed in subsect. **2**˙4.

The validity of this method can be estimated directly from data. To do so, the selection cuts are loosened until the signal contribution becomes negligible compared to the backgrounds. Then $R_{\alpha_{\mathrm{T}}}$ should be independent of $|\eta|$. Figure 5 shows $R_{\alpha_{\mathrm{T}}}$ for a mix of SUSY LM1 and background events for several different HT cuts. For relatively low requirements on HT, $R_{\alpha_{\mathrm{T}}}$ stays approximately constant while for stricter requirements on HT, $R_{\alpha_{\mathrm{T}}}$ is falling off with larger values of $|\eta|$. This study presents an elementary check that will need to be carried out once real collision data are available.

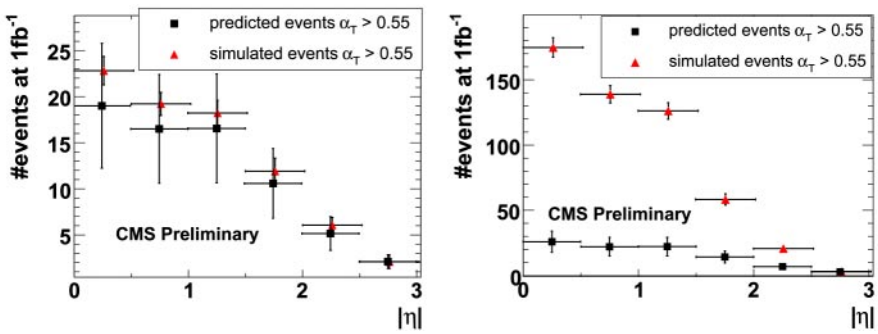

Fig. 4. – (Colour on-line) Comparison of the number of predicted and simulated (measured) events for a luminosity of $1\,\mathrm{fb}^{-1}$ with $\alpha_T > 0.55$. Left: background events only. The black squares indicate the number of predicted events, the number of simulated events is shown as red triangles. Right: background + SUSY LM1 signal. The black squares indicate the number of predicted background events, the total number of observed events is shown as red triangles.

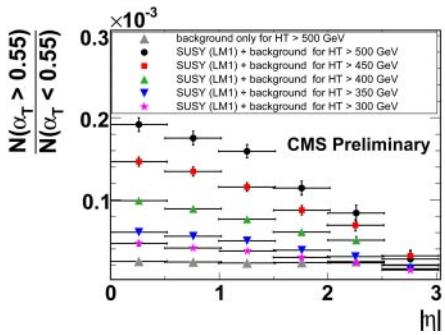

Fig. 5. – $R_{\alpha_{\mathrm{T}}}$ as a function of $|\eta|$ for different HT cuts.

4. – Estimation of $\mathrm{Z} \to \nu\nu$ missing energy spectrum from photon + jets events

In the following we present an alternative to the standard approach of using $\mathrm{Z} \to \mu\mu$ events for estimation of the background contribution from $\mathrm{Z} \to \nu\nu$ events in jets + missing energy searches. We will instead use a sample containing a high-p_{T} photon produced with high-p_{T} jets that has larger statistics. The missing energy (MET) spectrum is obtained by removing the identified photon and correcting for residual differences between these events and invisible Z events. Similarly, a sample of W + jets events could be used [6].

The differential production cross-sections for W, Z or photon plus exactly two additional partons, including all contributing subprocesses as evaluated by MadGraph [7], are shown in fig. 6. The production of W bosons is higher by a factor of three at high p_{T}, as expected, while photon production is within 20% of Z production. The γ-to-Z ratio levels out at a value that is simply predicted by the differences in the couplings of Z's *versus* photons to up-like and down-like quarks. Above $\sim 150\,\mathrm{GeV}$ boson p_{T} these ratios depend mostly on the electroweak characteristics of the events. Stated another way, the

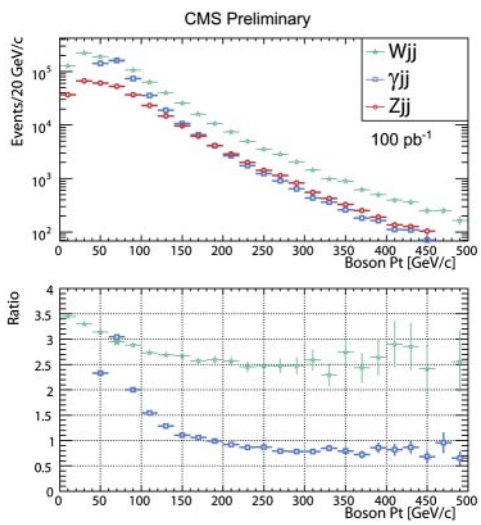

Fig. 6. – Top: differential event yield as a function of boson p_{T}, for the processes pp $\to$ boson + 2 partons (boson = W, γ, or Z) at generator level for $100\,\mathrm{pb}^{-1}$ integrated luminosity. Bottom: ratios of the yields of W relative to Z and γ relative to Z.

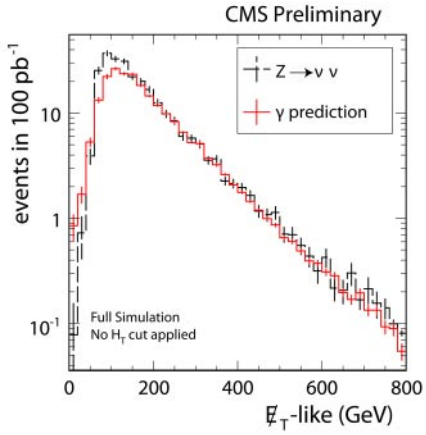

Fig. 7. – Comparison of MET from Z $\to \nu\nu$ events and the "MET-like" quantity from photon + jets events after all corrections.

hadronic parts of these events are not easily predicted, but to good approximation do not depend upon whether the boson is a Z, W, or photon. The ratios are thus relatively robust to variations in selection criteria, such as number and transverse energies of jets. In the absence of large contributions to these samples from new physics, they have the potential to be suitable for predicting the MET spectrum for invisible Z's at high p_{T}.

The Monte Carlo samples used do not take into account theoretical uncertainties such as Q^2 scale variations, and contributions from uncertainties in parton distribution functions. Initial studies indicate that the former can affect the relative normalization of photon+jets to Z+jets events at the $\sim 10\%$ level, while the latter have much smaller impact. The difference due to collinear photon production is expected to be mitigated by isolation requirements. In general there is also a difference in the η distribution of photons relative to that of Z bosons, as a result of different phase space factors for massive Z bosons *versus* massless on-shell photons, and to a lesser extent, due to the different vector and axial couplings. However, at sufficiently high p_{T} the bosons tend to be found in the central region, which significantly mitigates the difference.

Once the photon p_{T} spectrum is measured for events passing the event selection criteria (in this case three jets within $|\eta| < 3$ and with uncorrected $p_{\mathrm{T}} > 180\,\mathrm{GeV}$, $> 110\,\mathrm{GeV}$ and $> 30\,\mathrm{GeV}$, respectively, and an isolated photon with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$), the transverse component of the vector sum of photon E_{T} and event calorimeter MET is computed, and this "MET-like" quantity is corrected for the photon isolation efficiency and the Z $\to \nu\nu$ branching ratio. Taking into account residual differences associated with couplings to quarks, a final correction was calculated via the ratio of Z plus three parton to photon plus three parton generator level events obtained with ALPGEN [5]. The ratio is flat at high E_{T} as expected. The resulting spectrum is found to be in excellent agreement with that of the invisible Z events in the MET region above 200 GeV as seen in fig. 7. For this exercise, all the corrections were evaluated and applied in the barrel and endcap separately. For $100\,\mathrm{pb}^{-1}$, the contribution of Z $\to \nu\nu$ events to the MET $> 200\,\mathrm{GeV}$ region can be estimated with a statistical uncertainty of order 10% while systematic uncertainties obtained via data-driven techniques are expected to be roughly 20%. A more detailed description of how to use photon + jets but and W + jets events to determine the background contribution from Z $\to \nu\nu$ events can be found in [6].

5. – Conclusions

A prospective search for a low-mass SUSY signature with di-jet events has been carried out. In this study two new kinematic variables, α and α_{T} were explored which are very powerful in suppressing the several orders of magnitude larger background from QCD di-jet events without making explicit use of a calorimeter-based missing E_{T} measurement. With the discrimination power of α (α_{T}), several SUSY benchmark points can be discovered with a data sample smaller than $1\,\mathrm{fb}^{-1}$, for which signal-over-background ratios of up to 6 are achieved. Over the past few months the α_{T} method has been further developed and extended to multi-jet events [8]. Furthermore two independent data-driven techniques have been developed for background estimation. By defining signal-depleted and -enriched regions in the leading jet η it was shown that a matrix method can be used to predict the total number of background events in the central η region with $\alpha_{\mathrm{T}} > 0.55$. In an alternative approach that can be used as a cross-check it was demonstrated how to use photon + jets event to determine the dominant $\mathrm{Z} \to \nu\nu$ background.

REFERENCES

[1] Randall L. and Tucker-Smith D., *Phys. Rev. Lett.*, **101** (2008) 221803.
[2] The CMS Collaboration, *CMS Physics Analysis Summary*, CMS PAS SUS-08-005.
[3] Bayatian G. L. *et al.* (The CMS Collaboration), *J. Phys. G*, **34** (2007) 995.
[4] Sjostrand T. *et al.*, *Comput. Phys. Commun.*, **135** (2001) 238.
[5] Mangano M. L., Moretti M., Piccinini F., Pittau R. and Polosa A. D., *JHEP*, **07** (2003) 001.
[6] The CMS Collaboration, *CMS Physics Analysis Summary*, CMS PAS SUS-08-002.
[7] Alwall J. *et al.*, *JHEP*, **0709** (2007) 028.
[8] The CMS Collaboration, *CMS Physics Analysis Summary*, CMS PAS SUS-09-001.