Validation of the MadAnalysis 5 implementation of CMS-SUS-14-001: stop search in all hadronic final states using top-tagged jets

Prachi Atmasiddha (IISER Pune); Samuel Bein (Florida State U.); Seema Sharma (IISER Pune)

emails: prachi.atmariddha@students.iiserpune.ac.in, samuel.bein@gmail.com, seema@iiserpune.ac.in

December 10, 2015

1 Introduction

We present the MadAnalysis 5 (MA5) implementation of the CMS search for the production of third generation squarks in all hadronic final states using top-tagged jets [1]. Proton-proton collision data were collected at a centre of mass energy of 8 TeV with an integrated luminosity of 19.4 fb^{-1}. The analysis is the first of the three searches presented in [1], which are the multijet top-tagged search, a dijet b-tagged search, and a monojet search (also implemented in MA5 [5]). This implementation supplements the public analysis database presented in [2].

In this search, a pair of hadronically decaying top quarks is required to be reconstructed in final states with large missing transverse momentum p_T^{miss}. This final state is expected to occur in excess if the mass difference between the top squark (stop) and the stable LSP (lightest supersymmetric particle) is larger than that of top quark ($m_{\tilde{t}} - m_{\tilde{\chi}_1^0} \geq m_t$), which is assumed. The top tagging is an essential feature because it effectively rejects backgrounds from multijet events while preserving the signal acceptance.

The CMS collaboration provided LHE files for the validation of the MA5 implementation. In the generation of these events, a 100% branching ratio for $\tilde{t} \to t\tilde{\chi}_1^0$ was assumed. We then showered and hadronized these events using PYTHIA 6 [4], and performed a simulation of the CMS detector using the built-in MA5-tuned implementation of Delphes [3]. After analyzing the events at the level of the reconstruction, event cut flow and signal region event counts are compared with the results obtained by CMS. The simplified model shown in Figure 1) has been used as the benchmark signal scenario, and a range of values of the masses of the stop and neutralino have been considered.

In the following, we summarize the baseline and signal region selection of the analysis, highlighting the differences between what has been implemented in MA5 and CMS, and follow with

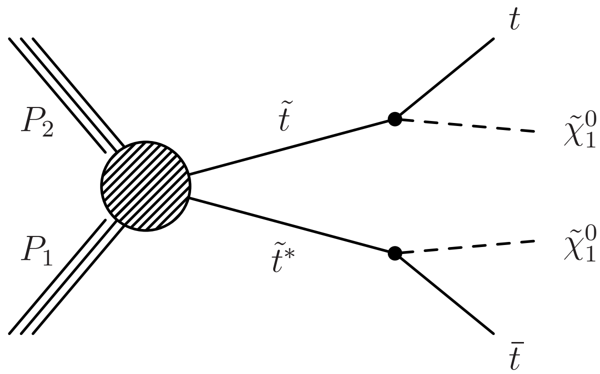


Figure 1: Diagram for the T2tt SMS topology. Several mass combinations of the stop and LSP are used in the tables below as benchmark comparison scenarios.

the summary of the agreement between the implemented and original analyses.

2 Baseline event and object selection

These are the selection are applied to events and objects.

-**event cleaning**. The cleaning described in the paper has not been performed because it requires a detailed knowledge of inhomogeneities in the CMS detector, but we apply a flat efficiency when computing the cut flow tables. The losses due to this cleaning are small and the efficiency is always greater than 95%.

- **lepton veto:**

- muons with $p_T > 5$ GeV, $|\eta| < 2.1$, isolation< 9.5.
- electrons with $p_T > 5$ GeV, $|\eta| < 1.44$ or $1.56 < |\eta| < 2.5$, isolation< 6.2.

lepton isolation: The paper describes a "directional isolation", but we perform a standard isolation, defined as

$$\text{isolation} = \frac{1}{(p_T)_{\text{lep}}} \sum_i (p_T)_i \tag{1}$$

where the sum runs over all reconstructed particles with an η-ϕ separation from the lepton less than 0.5. The large (loose) isolation thresholds select leptons with a high efficiency in order to reject standard model events with non-isolated leptons.

- **Multiplicity of reconstructed jets:** ($p_T > 30$ GeV, $\eta < 2.4$)

- (Number of jets of $p_T > 70 GeV$) ≥ 2

- (Number of jets of $p_T > 50GeV$) ≥ 4
- (Number of jets of $p_T > 30GeV$) ≥ 5

- **Azimuthal angle between three highest p_T jets and the p_T^{miss}:**

- $\Delta\phi(p_{T1}, p_T^{miss}) > 0.5$,
- $\Delta\phi(p_{T2}, p_T^{miss}) > 0.5$,
- $\Delta\phi(p_{T3}, p_T^{miss}) > 0.3$,

where $p_{T1} > p_{T2} > p_{T3}$, three highest p_T jets selected.

- **Nbjets** ≥ 1
The b-tagging efficiency stated in the paper differs from the standard MA5 prescription. A simple a 67% b-tagging efficiency and 1.4% misidentification rate is applied to match what is stated in the paper. To achieve this, we altered the detector card that specifies the b-tagging efficiency. The card, called delphesMA5tune_card_CMS.tcl, is provided in the documentation.

- **Top Reconstruction:** One fully reconstructed top quark system is required in the final state. This top candidate is the four-vector sum of three jets in the event, and the prescription given in the paper is followed. Additionally, a partially reconstructed top quark system, built from the remaining jets in the event, is required. For the partially reconstructed top quark, we consider this implementation to be approximate. The prescription was based on what is given in the paper, and the analyzers provided some further clarification. For a full description of the remnant system reconstruction, we refer the reader to the analysis code provided.

-$\mathbf{p_T^{miss}}$ > 200 GeV

- $\mathbf{M_{T2}} \geq 300$ GeV
M_{T2} is computed using the built-in MA5 function with input vectors set as the fully reconstructed and partially reconstructed top quarks, and a test mass of 0.

- $\mathbf{M_T}$: [0.5*M_T(full top)+M_T(Rsys)] ≥ 500 GeV
Here, the definition of the transverse mass M_T is taken from the paper and not from the usual definition as in the TLorentzVector class.

Signal Regions Selection: Signal regions are defined by applying tighter selection on p_T^{miss} and Nbjets.
- $p_T^{miss} \epsilon$ [200,350] and Nbjets = 1
- $p_T^{miss} > 350$ and Nbjets = 1
- $p_T^{miss} \epsilon$ [200,350] and Nbjets ≥ 2
- $p_T^{miss} > 350$ and Nbjets ≥ 2

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	98.13 (xxx)	98.13 (xxx)
No Mu	72.16 (73%)	72.22 (74%)
No Ele	55.41 (76%)	55.50 (77%)
Njet70>1	49.55 (89%)	50.11 (90%)
Njet50>3	31.16 (62%)	32.29 (64%)
Njet30>4	26.25 (84%)	27.15 (84%)
Min $\Delta(\phi)$	22.46 (85%)	23.23 (86%)
Nbjets>0	19.63 (87%)	19.85 (85%)
MET>200	12.21 (62%)	13.02 (66%)
Top Reco	– (–)	5.82 (45%)
MTsum>500	4.87 (39%)	4.95 (85%)

Table 1: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-125 and the MA5 results are given in column 3. .

3 Comparison of MA5 and CMS results

Tables 1–9 and Figures 2–4 show the results of the comparison of counts and distributions between the MA5 implementation and the CMS result. In some cases the cut flow tables give the number of events normalized to 100%, and in other cases the tables are normalized to the cross section times the integrated luminosity. In each case, the normalization convention used by CMS was followed. Dashes hold the place of values that were not provided in the CMS cut flow tables.

Agreement to within 15% is typical, but there are a few instances in which the disagreement is larger. In these cases, it is not believed that the discrepancy will significantly impact the mass limits, but this remains to be tested. The most significant difference can be seen as the discrepancy between the shapes of the M_T(Rsys) distributions. The CMS and MA5 distributions appear to be shifted by about 20% with respect to each other.

We observe differences in counts on the order of 15% in association with variables derived from the remnant system between the MA5 and CMS implementations. The effects are not visibly translated into the cut flow tables because the bins of the analysis integrate over any shape differences. It is suspected that these differences arise out of a difference in the criteria used to select the b-jet seed for the remnant system, and the fact that the MTRsys used for the last M_T cut is different from the $M_T(Rsys)$ used for M_{T2} calculation). For one of the benchmark points, (T2tt-350-0), a 50% discrepancy is seen in one of the signal regions, which we attribute to a combination of jet response mis-modeling as well as mis-modeling of the $M_T(Rsys)$ variable in the presence of highly boosted top-tagged jets. The discrepancy is approximately on the order of the typical uncertainty on the signal cross section, and so is not expected to alter any conclusions that may be derived.

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	1.19	1.22
MET>350, Nbjets=1	0.93	1.10
MET200-350, Nbjets>1	1.64	1.42
MET>350, Nbjets>1	1.11	1.21

Table 2: The signal region (SR) counts in CMS CMS-SUS-14-001 for the working point T2tt-500-125 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table 1.

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	97.44 (xxx)	97.44 (xxx)
No Mu	72.5 (74%)	71.74 (74%)
No Ele	55.55 (76%)	54.91 (77%)
Njet70>1	52.72 (94%)	51.82 (94%)
Njet50>3	34.55 (65%)	34.66 (67%)
Njet30>4	28.49 (82%)	28.88 (83%)
Min $\Delta(\phi)$	24.98 (87%)	25.24 (87%)
Nbjets>0	21.81 (87%)	21.68 (86%)
MET>200	17.6 (80%)	17.64 (81%)
Top Reco	– (–)	9.20 (52%)
MTsum>500	8.37 (47%)	8.62 (94%)

Table 3: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-25 and the MA5 results are given in column 3. .

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	1.06	0.93
MET>350, Nbjets=1	2.49	2.95
MET200-350, Nbjets>1	1.34	1.21
MET>350, Nbjets>1	3.48	3.54

Table 4: The signal region (SR) counts in CMS CMS-SUS-14-001 for the working point T2tt-650-25 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table 3.

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	15662.0 (xxx)	15662.0 (xxx)
No Mu	– (–)	11568.97 (73%)
No Ele	8802.0 (56%)	8927.82 (77%)
Njet70>1	– (–)	7380.74 (82%)
Njet50>3	– (–)	4350.13 (58%)
Njet30>4	3113.0 (35%)	3653.80 (83%)
Min $\Delta(\phi)$	2205.0 (70%)	2972.08 (81%)
Nbjets>0	2200.0 (99%)	2539.64 (85%)
MET>200	– (–)	1010.90 (39%)
Top Reco	– (–)	314.46 (31%)
MTsum>500	182.9 (8%)	213.18 (67%)

Table 5: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-350-0 and the MA5 results are given in column 3. .

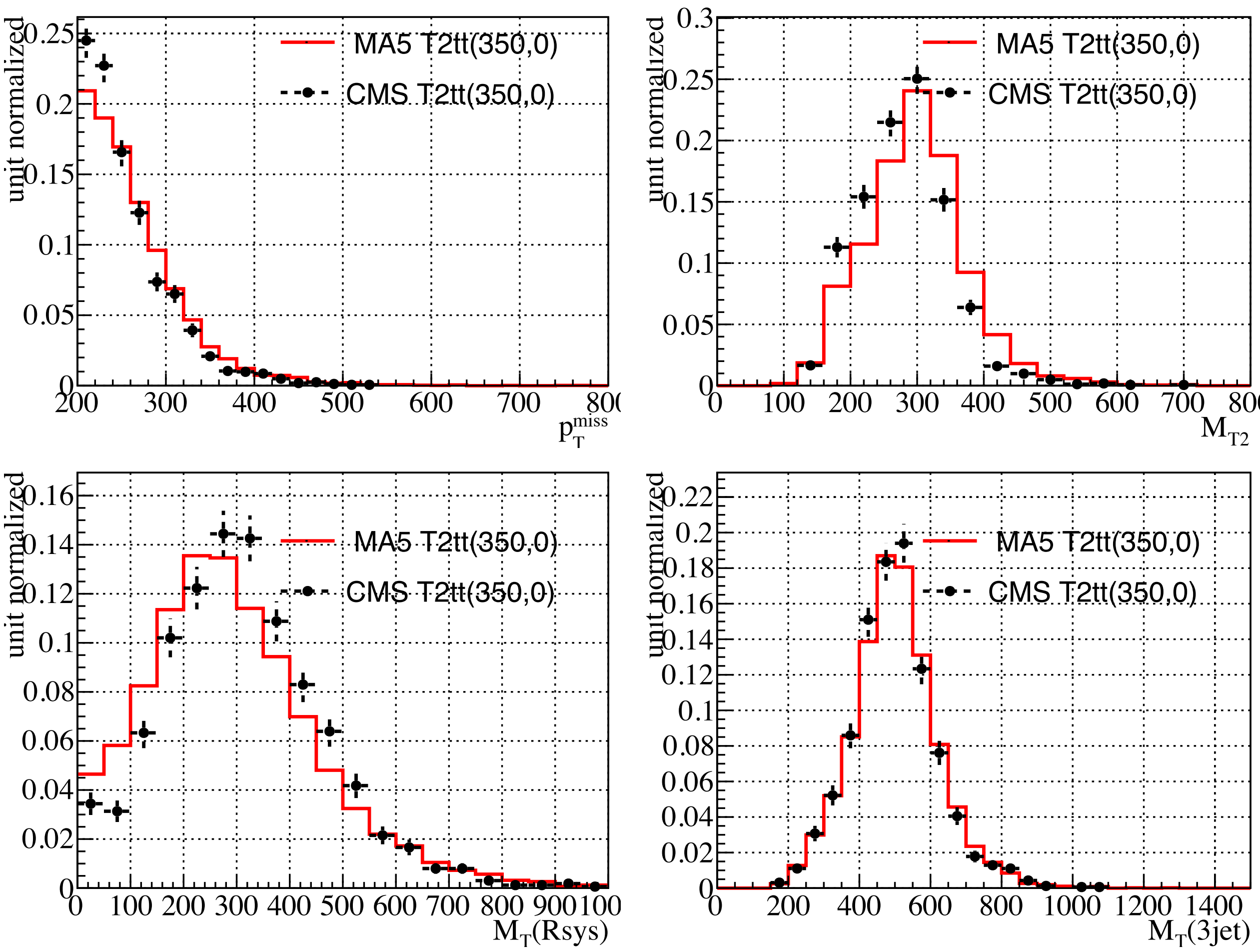
Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	–	95.08
MET>350, Nbjets=1	–	13.66
MET200-350, Nbjets>1	–	90.88
MET>350, Nbjets>1	7.5	13.55

Table 6: The signal region (SR) counts in CMS CMS-SUS-14-001 for the working point T2tt-350-0 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table 5.

Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	1660.0 (xxx)	1660.0 (xxx)
No Mu	– (–)	1229.14 (74%)
No Ele	927.0 (55%)	942.80 (76%)
Njet70>1	– (–)	856.80 (90%)
Njet50>3	– (–)	551.71 (64%)
Njet30>4	419.0 (45%)	468.27 (84%)
Min $\Delta(\phi)$	360.0 (85%)	400.29 (85%)
Nbjets>0	314.0 (87%)	341.67 (85%)
MET>200	– (–)	229.78 (67%)
Top Reco	– (–)	105.02 (45%)
MTsum>500	85.9 (27%)	90.82 (86%)

Table 7: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-500-100 and the MA5 results are given in column 3. .

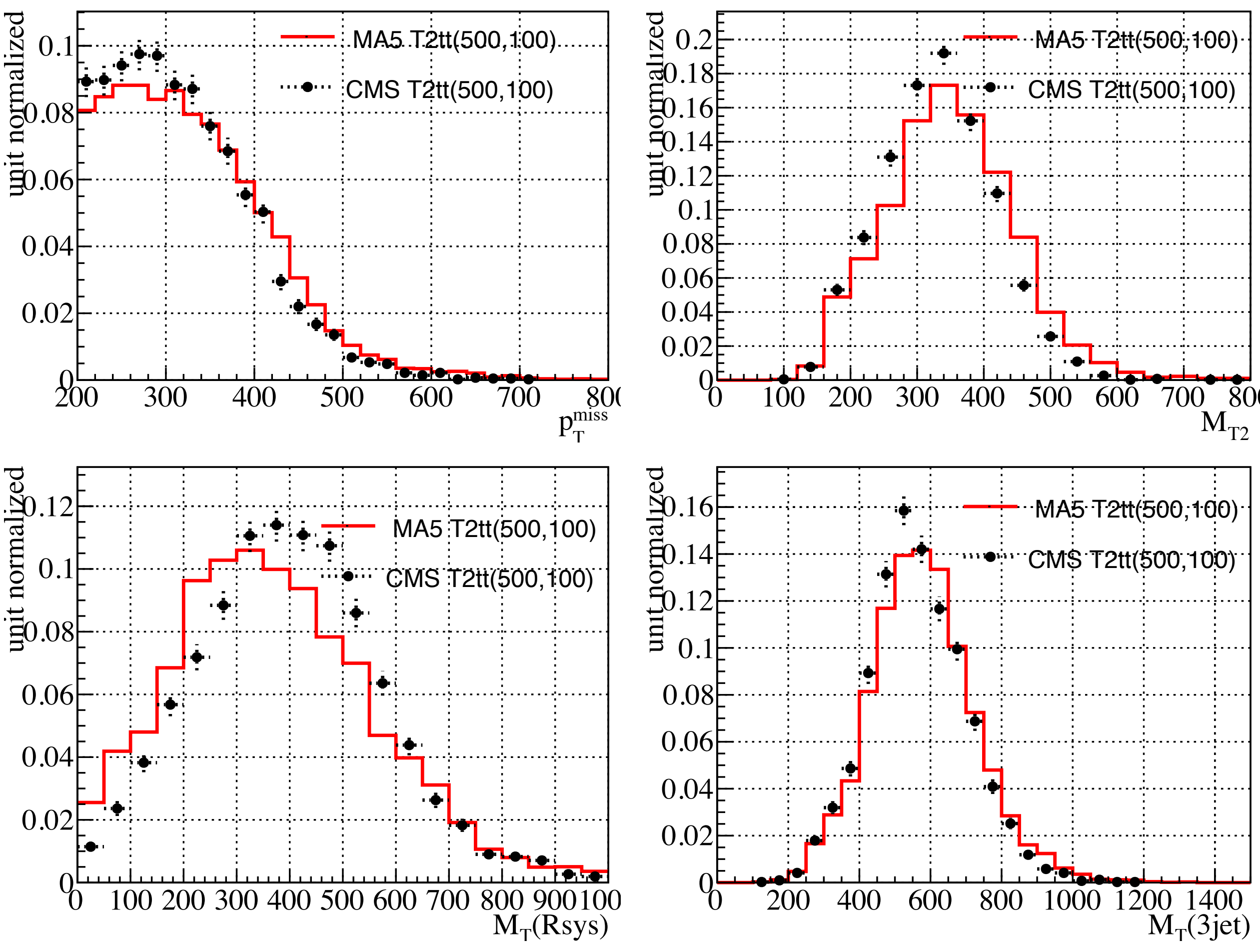
Figure 2: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (350,0).



Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	–	21.48
MET>350, Nbjets=1	–	20.52
MET200-350, Nbjets>1	–	25.31
MET>350, Nbjets>1	19.8	23.50

Table 8: The signal region (SR) counts in CMS CMS-SUS-14-001 for the working point T2tt-500-100 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table 7.

Figure 3: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (500,100).



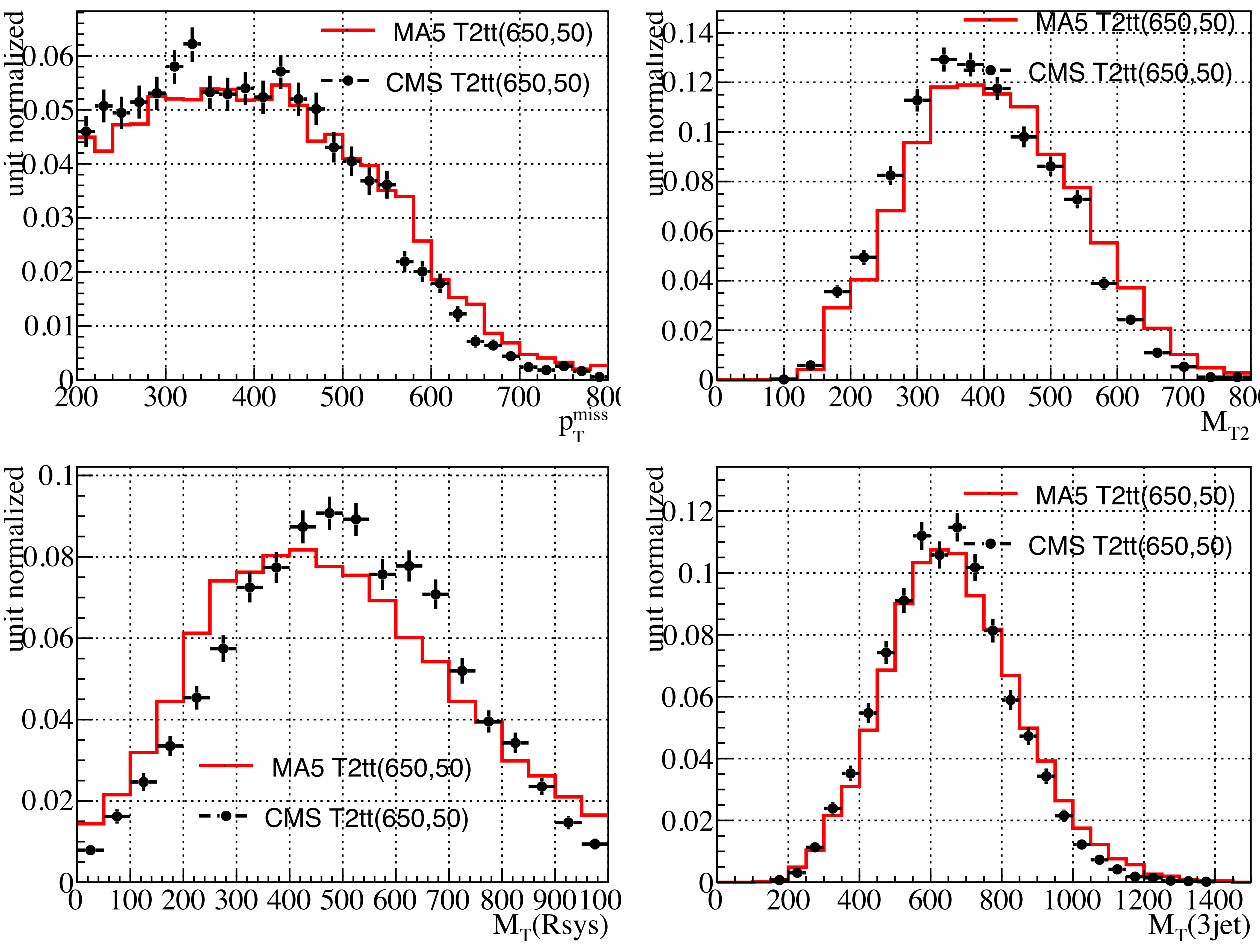
Cut Name	CMS Count(Eff)	MA5 Count(Eff)
Event Cleaning	270.8 (xxx)	270.8 (xxx)
No Mu	– (–)	199.25 (73%)
No Ele	152.0 (56%)	152.26 (76%)
Njet70>1	– (–)	143.65 (94%)
Njet50>3	– (–)	95.87 (66%)
Njet30>4	75.0 (49%)	79.95 (83%)
Min $\Delta(\phi)$	66.0 (88%)	69.84 (87%)
Nbjets>0	58.0 (87%)	59.99 (85%)
MET>200	– (–)	48.58 (80%)
Top Reco	– (–)	25.36 (52%)
MTsum>500	22.7 (39%)	23.71 (93%)

Table 9: The acceptance cut flow for the baseline selection in CMS SUS-14-001 for model point T2tt-650-50 and the MA5 results are given in column 3. .

Signal Region Name	CMS	MA5
MET200-350, Nbjets=1	–	2.77
MET>350, Nbjets=1	–	8.08
MET200-350, Nbjets>1	–	3.35
MET>350, Nbjets>1	9.3	9.49

Table 10: The signal region (SR) counts in CMS CMS-SUS-14-001 for the working point T2tt-650-50 after all selection has been applied. Column 2 is the CMS account, and our own results displayed in column 3. These counts were determined by applying the SR selection to the end of the cut flow featured in table 9.

Figure 4: MA5 and CMS unit-normalized kinematic distributions after the baseline selection for the T2tt working point (650,50).



References

[1] V. Khachatryan *et al.* [CMS Collaboration], "Searches for third-generation squark production in fully hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV," JHEP **1506**, 116 (2015) doi:10.1007/JHEP06(2015)116 [arXiv:1503.08037 [hep-ex]].

[2] B. Dumont *et al.*, "Toward a public analysis database for LHC new physics searches using MADANALYSIS 5," Eur. Phys. J. C **75**, no. 2, 56 (2015) doi:10.1140/epjc/s10052-014-3242-3 [arXiv:1407.3278 [hep-ph]].

[3] J. de Favereau *et al.* [DELPHES 3 Collaboration], "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," JHEP **1402**, 057 (2014) doi:10.1007/JHEP02(2014)057 [arXiv:1307.6346 [hep-ex]].

[4] T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **0605**, 026 (2006) doi:10.1088/1126-6708/2006/05/026 [hep-ph/0603175].

[5] S. Sharma and S. Pandey, "MadAnalysis 5 implementation of CMS-SUS-14-001 (monojet analysis)," doi:10.7484/INSPIREHEP.DATA.QGBP.K237

Acknowledgements:

We would like to thank Sabine Kraml, Dipan Sengupta, Hongxuan Liu and Shubham Pandey, and and the whole MA5 team. S.B. thanks to the LPSC Grenoble for hospitality during a research stay where this work was started, and the LabEx Enigmass for financial support.