Table of Contents

SUS-15-009: Search for natural GMSB in events with top quark pairs and photons (8 TeV)

Contact

- Brian Francis brian.patrick.francis@cernNOSPAMPLEASE.ch
- Brad Cox brad.cox@cernNOSPAMPLEASE.ch

Links

CADI⌐
HN⌐

Timeline

- Full status update⌐: June 15, 2015
- Pre-approval⌐: August 24, 2015 --> Conditionally pre-approved
- Questions and responses from pre-approval
- Pre-approval announcement: October 9, 2015
- ARC-authors first meeting⌐: November 17, 2015

Analysis Summary

The analysis searches for an excess of high MET events in the lepton (e/mu), jets, and photon final state. The search targets the direct production of light stops in a GMSB scenario with a very bino-like neutralino NLSP.

The dominant background is Standard Model top-antitop pair production with associated photons or additional jets that may be mis-identified as photons. Additional backgrounds include typical backgrounds to ttbar searches: W/Z + gamma, diboson, W/Z + jets, and single top production. All backgrounds are simulated in Monte Carlo and several scale factors are determined to best fit the data in several control regions. An additional control region using poorly isolated photons ("fakes") is examined to characterize the MET shape of the MC.

The results of the search are interpreted as a shape comparison between data and expected backgrounds. To eliminate dependance on the SM ttbar+gamma+gamma cross section and the jet-->photon fake rate, the total background normalization is allowed to float freely to be an entirely shape-based comparison. Upper limits are calculated against a private MC sample of very bino-like NLSP GMSB models with the stop being much lighter than all other squarks or gluinos.

Signal MC

The search uses a privately generated FastSim set of samples. For information about these, see:

- David Mason's presentation⌐ to the SUSY Trigger/MC/xPAG meeting on July 7, 2014
- Twiki page describing the grid

ARC review:

Manfred Paulini on AN v3:

♦ Sec. 2.2: What checks were performed to gain some confidence that the privately produced GMSB signal samples can be trusted as if they were centrally produced?

These samples were created similarly to those for the di-photon inclusive searches (SUS-12-001 and SUS-12-018), in which the production was not stops but first/second generation squarks and gluinos (scalar udsg). Thus the checks performed for the "stop-bino" samples were mainly against these well-known older samples, which were also private FastSim.
We found that for stops and binos in our sample, the kinematics agreed favorably to that of squarks and binos with the same masses produced in these older samples. For an executive summary see Dave Mason's xPAG presentation◻ and the twiki for the scan.
If he'd like, perhaps Dave Mason could comment on this since he oversaw their creation firsthand.

♦ Sec. 2.5: why was the tt= sample re-weighted by the weights squared and not by a variation of no re-weight and 2x the weight (instead of weight squared)?

Weighting by the weight squared is the TOP PAG recommendation for estimating the upwards systematic fluctuation of this effect: see their twiki on the matter.

♦ Sec. 3.3 & 3.4: what bothesr me a bit is the fact that the eta regions for e (2.5) and mu (2.1) are different for tight but for loose you use 2.5 for mu, too. Can this difference in choice cause any effect on the CR estimates?

The requirement |eta|<2.1 for tight muons is due to the SingleMuon trigger requiring it, and is not necessary for other/additional muons in the event. The loose lepton veto is kept constant between signal and control regions, so this should not affect control regions. Where it could affect the analysis is if the object kinematics or MET differed greatly between ttbar-->(e+e, e+mu, mu+mu), in which case the different efficiencies for each combination would be important; however this is not the case.
The |eta|<2.1 cut in the trigger does make the tight muon requirement tighter than for the tight electron, which is one cause of the difference between electron and muon event counts. Beyond all this, these vetoes are what is recommended by the TOP PAG for semi-leptonic selections.

♦ Tab. 8: why is there a lower cut of 0.001 on sigma_IetaIeta? Is this standard photon ID? I don't recall ...

This cut, and the one on sigma_IphiIphi, are for ECAL spike removal and general anomaly protection. They are not required by EGamma but are fairly common; for example the 8 TeV inclusive search used these as well. Concerning its effect, zero otherwise selected photons in the TTGamma MC sample fail these cuts.

♦ Fig. 9: the fits seem okay around the Z region but are less from optimal away from the Z. Is this anything to worry about? Was is treated in a systematic uncertainty?

While not considered important for the signal regions, what you are seeing is the lack of Drell Yan for 20 GeV < M(lep lep) < 50 GeV, which in Figure 9 is exaggerated compared to signal regions due to the di-lepton selection here. You can see in Figure 9 that when requiring b-jets (the top two plots) this is not an issue. What can be done to study the effect of this is to re-do the template fit excluding this low-mass region, and see that the scale factor doesn't change much (it should be dominated by on-mass Z-->dilepton). Furthermore, since these events are more accurately Z/gamma* Drell Yan, the fit range can be extended to higher masses to observe how much the scale factors change. Keep in mind here that the non-btagged muon channel (bottom right of Fig. 9) is not used in the analysis: the non-btagged electron sample is only useful as an input to the electron mis-id rate measurement. When varying the fit range of Figure 9, the scale factors for this are:

Z(gamma) SF in channel	Normal (0 - 180)	50 - 180	50 - 600

ele_jjj	1.24	1.26	1.25
ele_bjj	1.38	1.39	1.39
muon_bjj	1.60	1.62	1.62

These are within the fit uncertainties summarized in Table 14 of the AN. So in short this is not seen as a cause for concern as the Z peak dominates the fit result, and was not given its own systematic. When other systematics are fluctuated, these fits are re-performed and so there is a reflection of these fits in the final results beyond just the fit/stat uncertainties. Plots of the results for this are shown below for the ele_jjj channel:
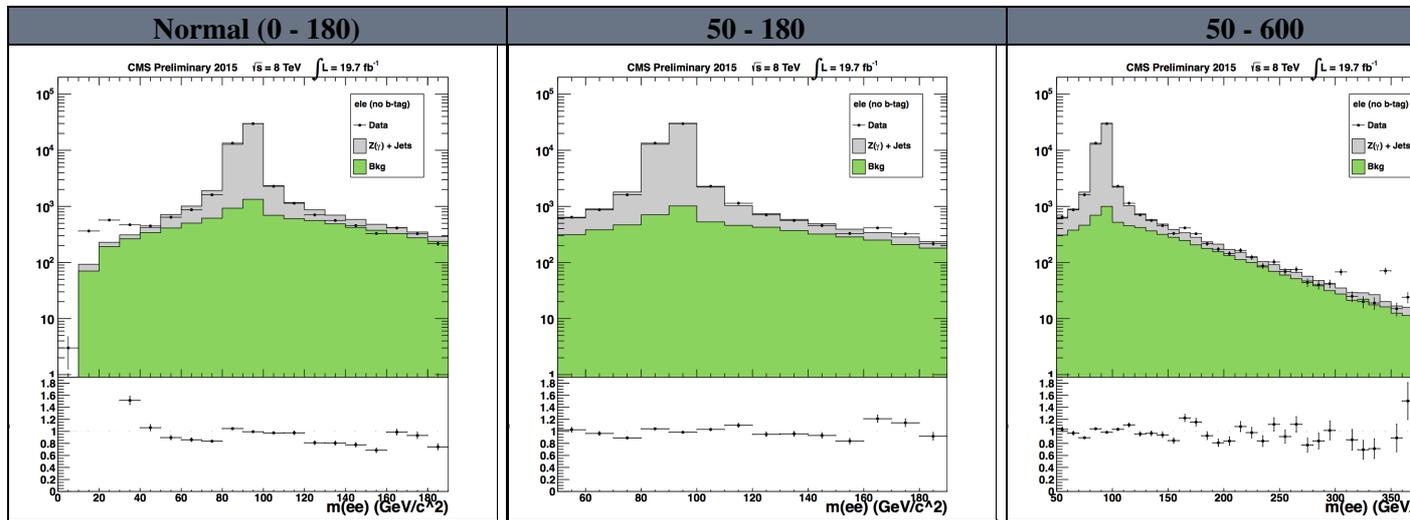
♦ Sec. 4.4.1, bottom of p. 21: how is the overall scale adjustment taken into account in the analysis? From Fig. 15 is seems to be a good 10% effect.

In lines 357-360 and 376-378 explain, this scale adjustment is not actually applied to the final result. The goal of this section is to ask: if we *were* to adjust the photon purity with this scale factor, would the distribution of MET change noticeably? In isolating only the shape of MET in the final evaluation, the extra 100% systematic on background normalization would wash away this overall 10% effect, but would not wash away a change in the shape. You can also see this is a 10% effect from the scale factors in Table 16.

♦ Tab. 15: the discrepancy between Fit and MC seems to be bigger in sigma_IetaIeta? Why not just using chHadIso? Or at least having a systematics that using only one or the other?

Back to the previous answer, neither is actually used in the final results so a systematic reflecting the difference isn't warranted. As for the discrepancy in sigma_IetaIeta here, the tt+gamma cross section measurement also encountered this and treated it in the same way. As you say, the way both analysis handled this was to just use chHadIso. The low sigma_IetaIeta is seen to be from some error in the shower evolution of photons in GEANT4.
If you look in the PAS in lines 147-151 indirectly touch on this question, because the 5% variation is from a very maximal case where you completely replace the MET from ttjets with tt+gamma's shape, or vice versa -- ie, if you were to perform a template fit like chHadIso or sigma_IetaIeta and find a maximal disagreement, the effect on MET would just be 5% bin-by-bin variations.

♦ Tab. 17 & 18: There is a significant excess in the data compared to the total background prediction - in CR1 and if I take the background errors at face value, also in CR2. I assume this came up in the pre-approval. What was decided then?

In the HN I noted these tables did not include the correct uncertainties, so in short this did not come up in pre-approval and nothing was decided. To further temper this issue, compare to Figure 29 to see that the event counts are well within uncertainties for most channels. Related to a previous question of yours, you can also

look at the photon purity measurement in Section 4.4, which in simplified terms can be considered a normalization of the tt+gamma/jet rate to data: it is roughly a 10% effect, which is about the order of the differences you speak of in Tables 17-20. You also might consider the public CMS measurement of the tt+gamma cross section (Public Twiki, CDS) which was higher than predictions by about 30% ± 30% for a similar (but not exactly the same) selection as this. Also, the uncertainties on the theoretical cross section of tt+gamma used here is 50%, and when all combined the theory systematics for ttbar-related rates alone are ~25%, well past the differences of which we're speaking. Lastly, the differences in CR1 are close to the systematic uncertainties therein (see Figure 16), and are used conservatively as an additional systematic in the signal regions -- ignoring the unfortunate presentation of uncertainties in the tables, the variations in all channels are fairly consistent with a tt+gamma rate that is slightly higher than predicted, an effect that in Section 4.4 we found to have minimal effect on the shape of the MET distribution.

- ◆ Tab. 19 & 20: Same comment for SR1 and certainly for muon SR2. What conclusion did the discussion about this data excess come to during the pre-approval?

See the previous answer for SR1. The table uncertainties seemed to have been overlooked in pre-approval and it simply did not come up. As for the muon channel in SR2, this was briefly touched upon in pre-approval as only an interesting notice. As a shape-only comparison however, this did not drive the limits as it was not compatible with the high-MET signal nor compatible with the other channels. The conclusion in pre-approval was that with higher statistics this might be good to explore, and with that CMS should be able to precisely measure the tt+gamma+gamma cross section and not rely on the shape-only provision. A significantly different (mu+jets):(ele+jets) ratio in tt+gg events would be exciting to see but this dataset is not powerful enough to approach that, and with the overall method isolating the MET shape we feel it's best not to address this in the PAS.

- ◆ Sec. 7, p. 44/45: why do you use all MET bins in your definition of your signal region? I thought the low MET bins were used for background normalization? Wouldn't it make sense to start the signal region at moderate MET, say > 50 GeV or so? From Fig. 29, the data-bg discrepancy seems to be at low MET. I think restricting the signal region to not include the low MET bins will also help in getting a better agreement between the data and bg predictions in Tab. 19&20. Was this discussed?

The reason for including these background-dominated bins, especially in SR1, is to allow the limit-setting machinery to constrain these backgrounds (with the 100% log-uniform "float" parameter, this makes it basically a normalization) in the high MET bins. For SR2, removing the low MET bins could be very dangerous for this analysis because if you only have 1-3 bins, you lose most of the "shape" information and you just have a log-uniform free-floating +/- 100% estimate, giving you no sensitivity.
As for "double-using" the low MET (< 50) SR1 region, recall from a previous question that the photon purity scale factor method is not applied to the final estimate. You can consider that method to be simply a check that if you were to change the composition of tt+jets and tt+gamma, would it indeed just be a normalization and not a big change in the MET shape? With that independant check giving a fairly flat 10% effect, you can just allow the limit-setting tool to fit the normalization for you using the log-uniform 100% float parameter and find that the post-fit value is very similar. Once again for Tables 19 and 20, if you include the correct uncertainties there is reasonable agreement and the discrepancy is of order 10% like all these effects. This was discussed in our group also in the context of avoiding "double-using" this low-MET region, and is why the photon purity scale factor is only a check.

Manfred Paulini on PAS v0:

- ◆ use CMS convention of GeV for mass and momentum and remove all GeV/c^2

Done.

- ◆ pp collisions: use pp in roman and not italic

Done.

- ◆ do not use 'fake ...' or fakes and replace all with misidentified or similar

All references replaced with "misidentified photon"

- ◆ look up PubComm recommendations for use of hyphens in b quark, b jet but b-quark jet ... and correct all

I went through the whole text and made many corrections governed by the PubComm hyphen rules.

- ◆ I know we talked about this ... this is just a reminder about the plot beautification and CMS figure standards ...

All plots have been recreated as closely as possible to the recommended style macros ⬀.

- ◆ title: it is not good to have an abbreviation such as GMSB in the title. My suggestion: Search for natural supersymmetry in events with top quark pairs and photons in 8 TeV pp collision data (or: ... in pp collisions at sqrt(s) = 8 TeV)

I agree, I believe the original title was a place-holder of sorts until the ARC began. It has been changed to your suggestion.

- ◆ abstract: We need to add that we don't find an access and set some limits. My suggestion for the abstract wording:
  We present a search for a natural gauge-mediated supersymmetry breaking scenario with the stop squark as the lightest squark and the gravitino as the lightest supersymmetric particle. The strong production of stop quark pairs and their decays would produce events with pairs of top quarks and neutralinos, with each decaying to photon and gravitino. This search is performed with the CMS experiment using pp collision data at sqrt(s) = 8 TeV, corresponding to an integrated luminosity of 19.7 fb-1, in the electron + jets and muon + jets channel, requiring one or two photons in the final state. We compare the missing transverse energy of these events against the expected spectrum of standard model processes. No excess of events is observed beyond background predictions and the result of the search is interpreted in the context of a general model of gauge-mediated supersymmetry breaking deriving limits on the mass of stop quarks up to 750 GeV.

I agree with the comment and have made the abstract to be very similar to your suggestion.

- ◆ Fig. 1: Since this is not a real Feynman diagram where time arrows play a role and need to be correct, I would remove all arrows and just show lines

Okay.

- ◆ Fig. 1 caption: What is GMM? My suggestion for a less redundant caption: Feynman diagram of the GMSB scenario of interest. With stop quarks as the lightest squark, their pair-production would be the dominant production mechanism for SUSY in pp collisions at the LHC. Assuming a bino-like neutralino NLSP, each stop would decay to a top quark and a neutralino, with the neutralino decaying to a gravitino and a photon. Shown above is the electron+jets or muon+jets final state of the top pair decay.

Prefer to call them "top squarks" as they are not quarks. Rewritten as:

"Feynman diagram of the GMSB scenario of interest. With top squarks (stops) as the lightest squark, the pair production of stops would be the dominant production mechanism for SUSY in pp collisions at the LHC. Assuming a bino-like neutralino NLSP, each stop would decay to a top quark and a neutralino, with the neutralino decaying primarily to a photon and gravitino. Shown above the the electron~+~jets or muon~+~jets final state of the top pair decay."

- l 6: what is "a new little Hierarchy problem"? How does it differ from the known 'regular' hierarchy problem? Can you explain or give a reference?

The 'regular' hierarchy problem is the required precision of order $1:10^{34}$ from the loop corrections to achieve a 125 GeV higgs mass, and with SUSY this is down to $1:10^2$. There is still a 'little' tuning between SUSY sparticle masses to keep EWKSB unchanged, and it is related to the stop mass. In literature I've seen from CMS this has gone un-referenced, but a suitable reference might be http://arxiv.org/abs/hep-ph/0007265 .

- l 7: and are left largely un-explored at the LHC. --> and have been left largely unexplored at the CERN LHC.

Okay.

- l 11: CMS now frowns upon the use of the expression "missing transverse energy" since energy is a scalar and a transverse component does not make sense. You can still use the symbol ET^miss but not the wording transverse energy. I would write here: ... contributes to large missing transverse momentum (\vec{p}_T^miss) in the detector, where the magnitude of \vec{p}_T^miss is referred to as ET^miss. [and then you are good to use ET^miss for the rest of the paper]

Done. This also infers a change in the title where 'missing transverse energy' was used; it is now 'missing transverse momentum'.

- l 14: in pp collisions --> in pp collisions at the LHC.

Done.

- l 24: in each signal region ... the reader might wonder what the signal region might be?

Adjusted this to: "in the one and two photons signal regions."

- l 26/27: don't take away the thunder of the paper and already reveal that no excess was found. I would remove "No significant excess ... of Standard Model processes, and" and just write: "... shape-based comparison. The results are compared to a range of stop and bino masses ..."

The first clause as you recommend is removed.

- l 30-32: Since this is a short paper, we do not need to state the "organization". I would remove l 30-32.

Agreed.

- l 42: I would remove "arising from the H -> gg decay." as it doesn't seem to be relevant here.

Done.

♦ l 44-46: Since you are using only barrel photons, there is no need to talk about the endcaps and then say we use barrel only. Remove lines 44-46.

Done, however the sentence saying only barrel photons are used and the reference should be informative.

♦ l 58: you start by saying that all objects are reconstructed using PF and then say that the PF algorithm clusters all particles into jets. I'm not sure whether this is correct. I would write: "... (PF) algorithm [33 35]. Jets are constructed by clustering particles using the anti-kT (note antikT typo) ..."

Done.

♦ l 63: 1.4442 -> 1.44 (that's good enough as precision)

Okay. I also see other papers using 1.44.

♦ l 64: remove the repetition of 'are required'

Done.

♦ l 66: after 'A photon-like shower shape is required.' reference the 8 TeV ECAL performance paper JINST 10 (2015) 08 (arXiv:1502.02702)

Okay.

♦ l 67: remove = sqrt(d-phi^2...) since already defined in l 60

Done.

* l 68: I would omit 'sub-detector dependently.'

Removed.

♦ l 71: to be --> be; same in l 75
♦ l 72: and an isolation energy --> and have an isolation energy; same in l 76

Done. Better parallel construction after changes.

♦ l 77: not sure but I think criteria is plural while we need singular here. Maybe say 'requirement' ?

I agree, and also this sentence is largely reproduced in lines 90-91 in more detail. For now this is changed to: "An additional, looser requirement for each lepton provides a veto against dileptonic backgrounds."

♦ l 86 - 100: a lot is repeated here from Sec. 3. Please remove repetitions such as "Photons are required to be tightly isolated ..." and just give final cuts if not yet done so and state the SR's and CR's

Made several changes to streamline this section.

- l 111: and the control region selection is designed to highlight this. How? Why? Can you explain and motivate this a bit?

If you recall the diphoton inclusive MET search (gamma gamma + X), the MET resolution is very different

for events with 'fake' photons (really jets). What this sentence should convey is that if you also have a semileptonic ttbar decay in the event, the effect of the reduced energy resolution for one object on the total MET is pretty small compared to all the other activity in the event. I've re-written this section to read:

"The control region definition is chosen to be orthogonal to the signal regions, to have very low signal acceptance, and to greatly enhance the population of the photon-like jets contributing the most to the background estimate in the signal regions. The control regions allow for the study of the performance of \MET simulation with the most poorly reconstructed photon-like objects expected in the signal region; the presence of a semileptonic \ttbar decay is expected to be a much larger effect on the \MET resolution than the photon energy resolution."

I think this way, the part about the ttbar system is just a statement of what's expected of these control regions.

- ◆ l 117: of simulated --> of the simulated

Okay.

- ◆ l 117/118: is generator level info used to reject the 0.6% of tt + jets events or how are these 0.6% identified and rejected?

Yes, only generator-level info is used here. If a tt+jets event has a generated photon within the tt+gamma sample definition for photons, it's rejected. I've added a word to clarify: "... the simulated \ttjets events contain generator-level photon radiation falling into..."

*l 119: maybe I missed it but V-gamma should be defined as W/Z + gamma

Changed in several places.

- ◆ l 123: calculated at at least NLO. --> calculated at least at NLO.

Fixed.

- ◆ l 128-138: it was hard for me to follow what was done in order to get the 2 scale factors. Can you try to rephrase and make a bit more clear?

Extensively re-written. This was broken into two paragraphs to hopefully make it easier to follow. I've also changed the "k factor" language to just "scale factor" as this is not at all a correction for loop order diagrams.

- l 139: the removal of the b-tag requirement applies to the MC data under discussion here, right?

For both MC and the data that's being fit.

- ◆ Tab. 1 caption: ... only the one is applied --> which one? Explain. Errors shown are fit+statistical only. --> What is fit+statistical? The error returned from the fit? This is usually considered a statistical error.

Yes, the "fit" error is just the error rerturned on the post-fit parameters. I've changed this to simply say statistical.

- ◆ l 141: of photon purity --> of the photon purity

Done.

♦ l 147: no difference is found ... where is this found? In MC? In which MC?

Changed this sentence to read: "... no difference in the overall distribution of simulated \MET is found when altering the purity of selected photons."

♦ l 147/148: The maximimal difference bin by-bin --> The maximimal bin by-bin difference

Fixed.

♦ l 148: found to be 5%, and when their --> found to be 5%. When their

Done.

♦ Fig. 2 caption: ... channels, and the template fit --> channels. The template fit ... ... b-tag requirement removed. Errors ... --> ... b-tag requirement removed is shown on the right. Errors ...

Done.

♦ l 152: insensitive to source --> insensitive to the source

Fixed.

♦ l 153: To eliminate dependance --> To eliminate the dependance

"To eliminate any dependance" might be better as it is less definite, the dependance on tt+gamma+gamma rate has not been quantified anywhere.

♦ l 155: to effect a completely --> to result in a completely

Done.

♦ l 158: I don't think A x eps is not defined

Changed to be spelled out: "(acceptance times efficiency less than 1%)"

♦ l 160: enhances --> enhance

Fixed.

♦ l 170: 1 - -8% shape systematic ??? 1 - 8% ?

Fixed. Purely typographical, should be "1 - 8%"

♦ l 175: so that they are completely --> and are treated as completely

Done.

♦ Tab. 2: explain what the check marks for 'shape' refer to

Done.

♦ Tab. 2: Control Region Delta --> Control Region Discrepancy

Done.

- ♦ l 183: remove 'corresponding to an integrated luminosity of 19.7 fb-1'

Done.

- ♦ l 185: in each signal region, across the entire range of MET ... See my comment above on Sec. 7 of the AN. Same for the data - prediction issues with Table 3.

As from my comments for the AN, Table 3 is formatted incorrectly and includes only the uncertainty from limited MC statistics, which is more correctly labeled a systematic. By including the correct uncertainties this table is much less confusing. Also as the photon purity method is used only as a check and not included as a normalization in the final background estimate, it is not a 'double-use' of the low MET bins, and the reason for including them is to allow the limit setting tool to constrain the total background using these bins.

- ♦ l 187: we need a bit more description of how the GMSB sample was produced.

I've broken this paragraph up for clarity, and to include a mention of SuSpect/SDECAY spectrum generation and PROSPINO NLO cross section calculations.

- ♦ Fig. 4 caption: The control region-derived uncertainties are not included in the systematic uncertainties shown above. --> Why?

Initially for technical reasons. Now that I've made the plots including these, they are not so different so I've included them in the PAS.

- ♦ l 200/201: observed indicating the presence --> observed that would indicate the presence

Fixed.

- ♦ References:
- ♦ check all references to have the proper way to put only the volume number in bold: Phys. Lett. B 70 (only 70 is in bold). See for example [17]-[19]
- ♦ there are problems with ref's [10], [11], [34], [35], [38], [39], [46]
- ♦ check [40]
- ♦ some refs have no doi: e.g. [2], [3], ...
- ♦ check arXiv only pubs are published by now

Should be much improved. I wasn't able to find DOI links for some of them, as the inspire BibTex doesn't have them. For these few I've tried to be as complete as I can.

Anthony Barker on AN v3:

Anthony Barker on PAS v0:

- • Abstract:

  just lightest squark, not squark/gluino.
  Include a statement in the abstract about what limit you set.

Done.

- • Figure 1:

The first line of text in the body of the paper involves an undefined acronym: GMM. In fact, GMM is not explained anywhere in the paper. Please define GMM.
Second sentence starts with a preposition and has no verb. "With stop squarks...
"Assuming a very bino-like neutralino NLSP" is a caveat better put in the introduction than in the figure.
"Shown above is.." Should be the words starting the figure caption, not the start of the 4th sentence.
Please cleaning up the grammar and flow of this caption.

GMM is removed.
I'm no expert in grammar, but isn't "with x as y" here a dependant clause modifying "production would be the ..."? "would be" should be the predicate.
The "bino-like" comment appears in both the caption and the introduction (line 11-12), as it should to keep the caption self-explanatory.
All captions begin with just a statement of what the object is, rather than "Shown here...".
Caption now reads:
"Feynman diagram of the GMSB scenario of interest. With top squarks (stops) as the lightest squark, the pair production of stops would be the dominant production mechanism for SUSY in pp collisions at the LHC. Assuming a bino-like neutralino NLSP, each stop would decay to a top quark and a neutralino, with the neutralino decaying primarily to a photon and gravitino. Shown above is the electron~+~jets or muon~+~jets final state of the top pair decay."

- Introduction:

  line 4: use double quotes on "natural".

Done.

  line 6, If "little higherarchy problem" is meant to be a proper name, capitalize "Little".

Made all lower case to match other CMS papers.

  line 8 should say "motivated by models of Gauge-Mediates Symmetry Breaking". GMSB is not a single model. Also change the next clause accordingly.

Done.

  line 9: neutralino NLSP is not the only case in GMSB (See http://arxiv.org/pdf/hep-ph/9801271v2.pdf pg 36) Maybe you mean to say "...we describe a GMSB motivated model in which the neutralino is ..."

Agree, changed to be your suggestion.

  line 12 awkward sentence. Suggestion: "...neutralino case whose dominant decay, x->yG, produces a photon in the final state."

Now reads: *"This search considers the case of a very bino-like neutralino, where photons in the final state would originate from its dominant $\chi z\_1 \rightarrow \gamma \tilde{G}$ decay."*

  line 13: get rid of the word the in "conserved, the pair-production"

Done.

  line 19: The second clause lacks a subject; should be "and it defines"

The sentence on lines 18-20 should be split into two sentences. This sentence is also confusing due to the word "isolate". At this point it is unclear whether the analysis is simply requiring a lepton that is assumed to come from tt or whether it is able to identify that the lepton came from the tt system rather than from qcd or gjet. Then it's ambiguous how many photon categories are there? Is it two categories: 1 or 2+? Or is it three categories: 1, 2, more than 2?

line 21: they're not "poorly isolated". You didn't do a bad job of isolating them, as though isolation were some active process that you could mess up. They are loosely isolated photons with tight photons excluded. This sentence has two totally unrelated things going on. Split it in to two sentences at the comma.

Cleaned up considerably. Removed mention of 'poorly isolated' in favor of 'photons that fail the nominal requirements...'.

line 28: you mean in *this" GGM scenario.

"This" GGM scenario isn't defined until later, and "a GGM scenario" is not incorrect.

- Section 3

line 59: "anti-kt" not "antikt" (Example: http://arxiv.org/pdf/0802.1189.pdf ⧉)

Fixed.

- Section 4:

line 104-105 the term "electromagnetically fluctuated jets" is needlessly opaque and will only be understood by other photon experts. Maybe say "jets that hadronized predominantly to electromagnetic objects", "Jets with a predominantly electromagnetic final state", or something that can be understood by a generic physicist. Also, say here that these type of objects are "photon-like jets" so that the term can be understood on line 109.
Something line 104-105 should read something like "These objects, which we will refer to as photon-like jets, are dominantly jets that hadronized predominantly to electromagnetic objects"

Now *"These objects are predominantly jets with large electromagnetic fluctuations in their hadronization and are used..."*

line 110-111: This sentence makes no sense. I don't even have a good guess of what it's attempting to communicate. What effect of the tt system on met resolution do you have in mind? As in, what tt-free baseline are you comparing this to? Why does it make sense to compare met resolution and photon resolution. Which photon resolution are you tailing about? Energy resolution? Of one photon or both? Or do you mean the contribution of the photon energy mis-measurement on met? What sort of highlighting are you talking about and why would you ever want to? Which control region selection? The definition of fake photons or the boundaries of the two control regions? The same confusion occurs in the caption of Table 2.

This section has been rewritten in response to Manfred's comment. My comment to him:
"If you recall the diphoton inclusive MET search (gamma gamma + X), the MET resolution is very different for events with 'fake' photons (really jets). What this sentence should convey is that if you also have a semileptonic ttbar decay in the event, the effect of the reduced energy resolution for one object on the total MET is pretty small compared to all the other activity in the event."

Anthony Barker on PAS v0:   12

line 113: many photons? in the tt+gamma case it's 0, and you see no more than 2 photons. So "many" is 0,1, or 2.

"Many" here refers to all selected photons from all background events. I've added the word "selected" to clarify. For example in the electron SR1 channel, you still expect 901 (many) ttbar + jet events as having one reconstructed, selected photon.

line 116: I don't know what a 2-7 configuration is. Please add a reference to explain that.

This is only a configuration of MadGraph and will not have a reference; the (pp \rightarrow bbjj\ell\nu\gamma) should be the explanation.

line 127: It seems you are using the Z to ee resonance to understand the electron to photon fake rate. This sentence hints at that, but there is no mention of the Z. Please make it explicit.

Done.

line 131: second clause has no subject. Add the word "it": "and it is formed..." or break into two sentences.

Rewritten extensively from a comment by Manfred.

Table 1: please reference the column headers SF_Z(g) and SF_e->g in the preceding paragraph and in the caption.

Done.

Table 1 caption last sentence: Please make this a proper sentence rather than short-hand. Also use the word "uncertainties" rather than "errors": "The shown uncertainties include only the statistical uncertainty and the uncertainty in the integrals of the fits."

Typographical error, this should be "only the first one is applied". Fixed.

Figure 2 caption: say "uncertainty" not "error".**Also, mention that the e-gamma plot is the plot on the right.

Done.

Figure 4: use larger fonts. Make the horizontal axis read Et^miss instead of E-slash to be consistent with the rest of the paper.

Axis label changed; these plots are recreated to be more in line with PubComm recommendations.

Figure 5: the plot header e/mu + >bjj+gg suggests that this is CS * acceptance going into SR2 only. Is it? If so, shouldn't there be a second plot for SR1?

Now this reads e\mu + >= bjj+gamma(gamma). Information dense, but more correct. It is all four channels so some events here are "gamma" and some "gamma gamma".

Figure 5 :Is it possible to show cross section limits for the diagonal region above mStop-mBino < mt? It would be good to show this for the sake of understanding how close to the diagonal can be excluded. Someday someone may want to do a study to exclude part of the diagonal and that information will be useful.

Anthony Barker on PAS v0:   13

This was discussed before pre-approval quite a bit, and what we concluded was that this area needs a much finer mass binning than was available. Furthermore it wasn't clear that the MC was handling off-shell top decays or the possibility of charms (stop --> charm + bino) correctly, whereas in the rest of the mass grid these aren't concerns. This analysis has very low acceptance in this region due to requiring high-PT leptons and b-jets, so we decided it was best just to not report results in this region.

In the future I agree it would be interesting, although triggering on leptons and requiring b-jets isn't ideal. The inclusive di-photon MET search could perform well here, and when the chance arises I like to remind that group of this. A specialized search would be needed to have significant sensitivity here.

> Figure 2: significant figure clean up needed: Larger fonts everywhere. The horizontal axis label interferes with the number markings (hist->GetYaxis()->SetTitleOffset(0.5)). Make the figures either all log or all linear. For the log plots, make a more sensible choice of vertical scale, going no lower than 10^2. In the horizontal axis title. make the c^2 into a superscript or drop it completely. Remove the legend titles or make them make sense instead of having both the left and right figures labeled "ele" when one is ee and the other is e-gamma.

Plots are remade for PubComm suggestions. Also axis range and labels. Third plot is also log-scale to match.

> Figure 3: use larger fonts for everything except the CR2 labels. In the legend, equalize the length of the two columns of entries. This will create vertical space in the legend which should then be used to make the font larger. Use a larger marker (In Anthony's CMSStyle file: PrettyMarker(hist, kBlack, 3 or 4);) CR2 ele looks to not be using Poisson error bars. Always use Poisson error bars for the data: hist->SetBinErrorOption(TH1::kPoisson).

Plots are remade for PubComm suggestions, and are somewhat different now. Not quite sure how you mean the error bars appear, but in CR2 the y-axis is log-scale and is quite low in value, giving the error bars the appearance of being asymmetric -- they are displayed the same as in all other plots.

> Table 2: fill down the Notes rather than using tick marks.

Okay.

- Physics comments*
- Section 1:

  4-5 The statement that SUSY keeps particles with large couplings to the higgs boson light seems to be a self-contradictory statement, as well as not indicating the third generation squarks.
  rewrite -- commas separating "particles (of) light" --> "particles (of things coupling high) are light"

Agree, interesting semantic ambiguity here. Fixed a bit, now reads:
"... models of SUSY in which, for Standard Model particles with large couplings to the Higgs boson, their sparticle partners are kept light, namely the third-generation..."

- Section 2:

  line 66 or 97: please specify the shower shape requirement. I think you mean a sinin cut, in which case, please specify the cut value.

You are correct, but it is not common practice to include the term "\sigma_{i\eta i\eta}" or to define precise cut values. The phrase used is standard in CMS publications.

- Section 3:

Anthony Barker on PAS v0:   14

line 72: Please specify which type of isolation is being applied to the muons?

As this is all isolation types, the current wording seems to be the most efficient as it parallels the wording for photons (line 67) and electrons (line 76).

- Section 4:

line 103: please specify the cut parameters of the fake photon definition.

As from a previous comment, CMS papers don't generally define sigmaIetaIeta and leave it as a "photon-like shower shape". Therefore simply saying these must fail the cut seems most appropriate.

line 118: which version of MadGraph?

Included version 5.1.3

Table 1: Why do the electron and muon channels have different k-factors (SF_Z)? If it's the cross section that the monte carlo gets wrong, we should expect a single k-factor.

Some of these factors were incorrectly named as k-factors. They are now appropriately named just 'scale factors', and so the difference between electron and muon channels less confusing.

line 136-137. I'm confused whether this fit is going on in data or in monte carlo. Presumably this template fit is being made in data. So why is a k-factor being applied?

This section has been extensively rewritten for clarity. The fit is done of data to MC backgrounds, and a scale factor is derived for the MC so as to better describe the data. The k-factor language has been removed.

lines 136 to 149: The description of the electron to photon fake rate scale factor is very unclear and confusing. Here's my confusion: Is the fit described on line 136 done on data or monte carlo? If it is done on data, why are you applying a k-factor to it. If it is being done on monte carlo, then are you applying this MC truth matching on top of the template? That would also make no sense since you could simply eliminate the fit entirely and just do MC truth matching to determine the fake rate. The use of a k-factor on the template fit indicates that this fit cannot possibly be done on data, and yet in order to make a scale factor there has to a measurement somewhere, what is it? Then, the b-tag requirement is removed but the fit is done in SR1, so it's not SR1, but instead some new loosened version of SR1. Finally, how is this second scale factor applied?

Again, now called more correctly a 'scale factor' which is an expression of both MC and data. The fit is done as a fit of two MC background templates to the data, and the truth-matching is done to the templates to better separate them. Truth-matching alone on MC, without a fit to data, only gives a measurement of the fake rate in MC and not a scale factor adjustment to data. As for 'not SR1', the current language should be more expedient than introducing an entirely new control region definition. As said, this section has been largely rewritten.

Line 167: it appears that CR2 is never used, so why not strike it from the PAS and just mention that it was considered and has too little statistics to be of any use.

There still seems value in showing CR2, since the agreement is fair and the message is that it's unused for reasons only of statistics and not an obvious failure in the method. It might be more questionable to leave it out entirely, since its construction is pretty natural compared to the rest of the analysis.

Lines 168-172: How is this additional shape systematic determined? Table 3 shows

significant excesses in three out of four channels. This strongly suggests that the background is inadequately modeled. SR1-ele constitutes a 3 sigma excess; SR1-muon a 2 sigma excess, and SR2-muon a 2.3 sigma excess.

Expanded on the additional shape systematic: it is calculated by the ratio of shapes of MET in signal and control regions. In more detail, if you suppose the agreement of data and MC is perfect in a CR, you have the verification that the MC was able to simulate the shape of MET in that CR perfectly. If however the shape of MET in a CR is different in one bin from the signal region shape by 5%, then you only have a verification that the MET was able to simulate something 5% different from a signal region correctly. So in this simple example, you would add an extra 5% uncertainty to that one bin due to this. In reality it is not just one bin, and the scale on this uncertainty is fairly small.

For Table 3 and the excess comment, again this table clearly did not show the correct uncertainties -- there is a sqrt(N) everywhere that is not explicitly shown. New tables showing the correct uncertainties have been made.

> line 143,153: why doe the need for a tt+gg sample depend on the photon purity? You say the cross section is "exceedingly small" so either the cross section makes it negligible or it needs to be accounted for somehow. This might make sense if you are bumping up the tt+gamma or tt+jet cross section to account for the tt+gg and then saying that it doesn't matter where the apparent photons come from. But that should have different contributions to SR1 and SR2. Why is met shape important rather than the relative contributions to SR1 and SR2?

There are tt+gg events simulated in the samples used; what is not used is an explicit, specialized tt+gg sample where both photons are high-PT and with large radiation angles. The cross section for such events is very very small. You often have high-PT photons (a single one) which is why the specialized sample of tt+gamma is needed, but you do not expect both photons to be in the tails of quickly falling PT distributions when the total yield itself is very small. What most of the selected SR2 events contain are mis-identified jets (as photons) or prompt photons where only one or zero of them are considerably high in PT or radiation angle. Remember that in all samples there are additional jets simulated, so "tt+gamma" is in reality "tt+gamma+jets" where some of those "+jets" do include photons ("+a" in MadGraph parlance).

The real question then is: do the "+jets/a" from MadGraph have the right number of photons, the right amount of electromagnetic fluctuation in jet hadronization? The resolution the analysis came to and that was discussed in pre-approval was that without a precise measurement of the SM tt+gg cross section, you could not be sure -- however if the MET distribution is the same between "+jets/a" contributions, then you could do a shape-based analysis independant of the absolute rate. What the analysis should accomplish is an estimate of the MET from the **selected** photons, and not try to pin down how much is due to **actual** photons.

This is the crux of the shape-based analysis and why the background normalizations are allowed to float. If it were possible to measure the SM tt+gg cross section with the 2012 dataset (ie real ttbar + prompt + prompt with a complete di-photon purity measurement), it would be possible to do this analysis with absolute background normalizations.

> Line 155-156: These lines are completely unclear; I have no idea what this is attempting to communicate. Here's my confusion: The upper limit determination of what? Shape based interpretation of what the shape of what? What result?

Now: *"backgrounds are allowed to float freely in the upper limit calculations so that the interpretation of the results is completely shape-based. This is accomplished by givin the total background in each channel and signal region a 100\% systematic uncertainty, flat in \MET."* I don't think it necessary to explain to the reader again how the results of the analysis are interpreted.

- Conclusion:

> line 200-201: "No significant excesses are observed..." you have no right to make such a claim while showing strong excesses in Table 3.

Table 3 does not show a significant excess when the counting statistics and systematics are included. What is more on this point, Table 3 as-is clearly does not match Figure 4. What is different between Table 3 and Figure 4 is that it's customary to include sqrt(N) uncertainties on the observed data points, but it's not typical to quote observed data as N +/- sqrt(N) in tables. That was the source of the mistake in creating Table 3, and the re-made versions of the table now include counting statistics on the background estimates.

> Figure 6: How is it possible that the observed limit exceeds the expected limit while Table 3 shows 2 and 3 sigma excesses in 3 of the 4 channels?

See above comments about uncertainties in Table 3. Furthermore the wording of the conclusion has been adjusted to include shape, saying "No significant excess in the shape of the MET distribution is observed that would indicate the presence of new physics."

> Figure 6: It would be good to see how the expected limit curves fall at the diagonal.

As from another comment, I would agree. However special care must be taken for this region, and a very fine mass binning of models (ie many samples) is needed that was not available. Furthermore, requiring high-PT leptons and b-jets severely limits acceptance in this region. A completely different analysis, or even the inclusive di-photon MET search, would be much better suited for this region.

-- BrianFrancis - 2015-12-10

This topic: Main > BFrancisNotes
Topic revision: r9 - 2016-01-18 - BrianFrancis