SMS scans in CMS

Nadja Strobbe

Universiteit Gent, PhD Fellowship of the FWO-Flanders

LPCC Workshop on Coordinating a simplified models effort
Oct 29-30, 2013

Table of Contents

For additional information, see also talk by M. Pierini at MC4BSM 2013

Introduction – From cMSSM to SMS

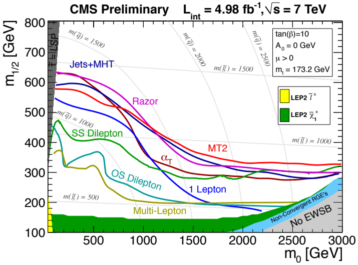
At the startup of the LHC, SUSY analyses interpreted their results mainly in the cMSSM model.

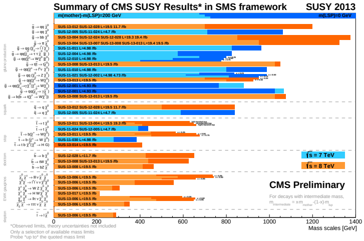
- One scan, used by all
- Using Pythia6

Gradually Simplified Models gained more interest. Now this is by far the most widely used way of interpreting results.

- A large number of topologies are considered
- One scan used by subset of all analyses
- First using Pythia6
  now using MadGraph + Pythia6

Other ways of interpreting results include pMSSM and GMSB models.

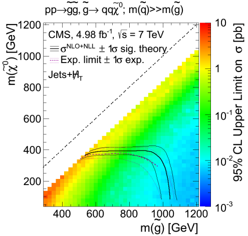




Which SMS's and how to decide?

Natural SUSY and Higgs discovery

1. In the beginning, basic SMS's such as $pp \to \tilde{g}\tilde{g}, \tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ were used
   - Good way to show reach of analyses
   - Provide useful information to theory community

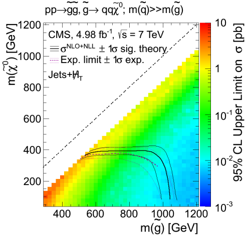


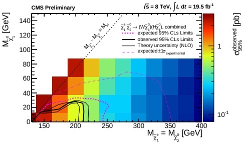
Which SMS's and how to decide?

Natural SUSY and Higgs discovery

1. In the beginning, basic SMS's such as $pp \to \tilde{g}\tilde{g}, \tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ were used
   - Good way to show reach of analyses
   - Provide useful information to theory community
2. Discovery of a Higgs + No SUSY
   $\Rightarrow$ Focus shifts towards "Natural SUSY models"
   - Light third generation sparticles in the spectrum
   - Enhancement of top and bottom quarks and tau leptons in the final state

   $\Rightarrow$ Consider additional SMS's
   - 3^{rd} generation, e.g. $pp \to \tilde{g}\tilde{g},\ \tilde{g} \to b\bar{b}\tilde{\chi}_1^0$
   - Higgs, e.g. $pp \to \tilde{\chi}_2^0\tilde{\chi}_1^\pm, \tilde{\chi}_2^0 \to H\tilde{\chi}_1^0, \tilde{\chi}_1^\pm \to W^\pm\tilde{\chi}_1^0$

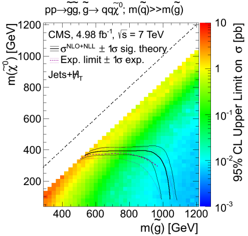




Which SMS's and how to decide?

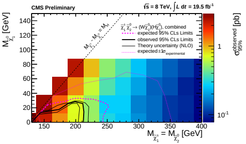
Natural SUSY and Higgs discovery

1. In the beginning, basic SMS's such as $pp \to \tilde{g}\tilde{g}, \tilde{g} \to q\bar{q}\tilde{\chi}_1^0$ were used
    - Good way to show reach of analyses
    - Provide useful information to theory community



2. Discovery of a Higgs + No SUSY
   ⇒ Focus shifts towards "Natural SUSY models"
    - Light third generation sparticles in the spectrum
    - Enhancement of top and bottom quarks and tau leptons in the final state

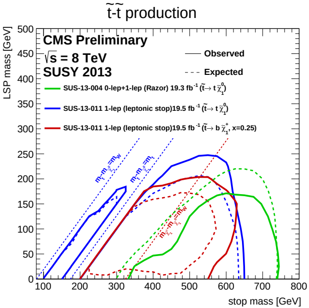
   ⇒ Consider additional SMS's
    - 3^{rd} generation, e.g. $pp \to \tilde{g}\tilde{g},\ \tilde{g} \to b\bar{b}\tilde{\chi}_1^0$
    - Higgs, e.g. $pp \to \tilde{\chi}_2^0\tilde{\chi}_1^{\pm}, \tilde{\chi}_2^0 \to H\tilde{\chi}_1^0, \tilde{\chi}_1^{\pm} \to W^{\pm}\tilde{\chi}_1^0$



3. Now: preparing for the final word on SUSY with Run1 data
    - Review what we have and do not have, talk with theory community
    - Try to be as comprehensive as possible
      ⇒ Produce missing scans, e.g. $pp \to \tilde{g}\tilde{g},\ \tilde{g} \to t\bar{b}\tilde{\chi}_1^-$

Which SMS's and how to decide?

SMS as gauge of our sensitivity



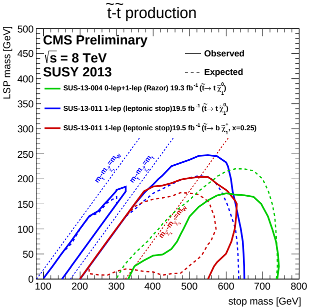
Observation:

- No sensitivity if $m_{\tilde{t}_1} - m_{LSP} \approx m_t$
- Reduced sensitivity for $m_{\tilde{t}_1} - m_{LSP} \to 0$

How to deal with this?

Which SMS's and how to decide?

SMS as gauge of our sensitivity



Observation:

- No sensitivity if $m_{\tilde{t}_1} - m_{LSP} \approx m_t$
- Reduced sensitivity for $m_{\tilde{t}_1} - m_{LSP} \to 0$

How to deal with this?

$\Rightarrow$ Look at complementary topologies, with one extra sparticle in the decay chain, providing a handle to close the sensitivity gap

- $pp \to \tilde{t}_2 \tilde{t}_2^*, \tilde{t}_2 \to \tilde{t}_1 Z$

$\Rightarrow$ Generate SMS for this topology, with the assumption that $\tilde{t}_1 \to t \tilde{\chi}_1^0$ and $m_{\tilde{t}_1} - m_{LSP} \approx m_t$

Which SMS's and how to decide?

Practically...

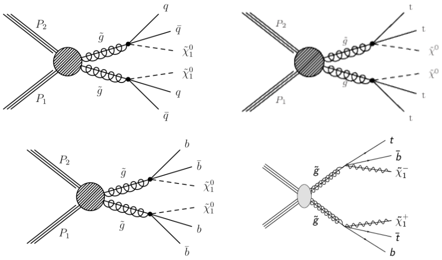
Within the CMS SUSY group there is a dedicated MC subgroup whose main task is to coordinate the SMS effort.

If analysts have suggestions, they are presented and discussed in the SUSY MC meeting. Scan details are also decided at that point:

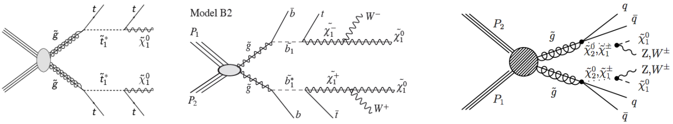
- Scan range
  → driven by reference cross section and expected luminosity
  `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections`
  → scans often stop at the boundary where a particle goes off-shell
- Granularity of the scan
  → do certain regions need to be more finely binned?
- Number of events per point
  → driven by expected efficiency and branching ratio into leptons when applicable; usually $\mathcal{O}(100k)$ events per point
- What to do with intermediate sparticles in the decay chain
  → one mass value or several?

All information on a specific SMS is documented on a dedicated twiki.

No intermediate sparticles:



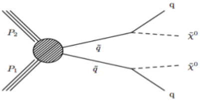
With intermediate sparticles:

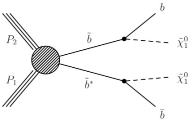


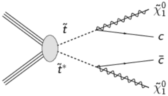
Produced SMS's

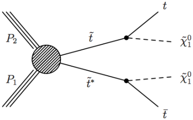
Squark production

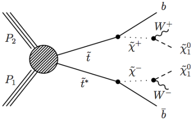
No intermediate sparticles:



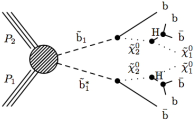


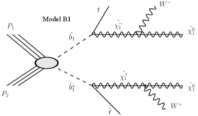


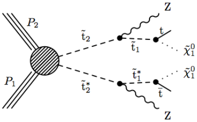




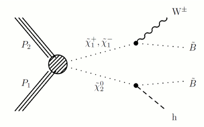
With intermediate sparticles:

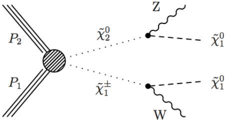


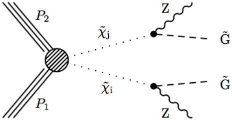




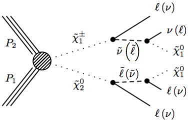
No intermediate sparticles:

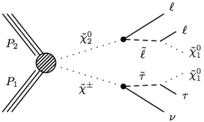


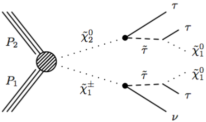




With intermediate sleptons:







Workflow overview

Production of SMS's in CMS currently goes through several consecutive steps:

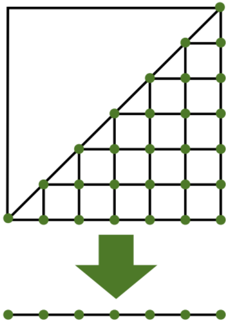
1. Production of initial sparticles with MadGraph
   No decays are done at this step
2. Decay of sparticles with Pythia6
3. Injection in CMS official production
4. Available for use in analyses

Step 1: Production of initial sparticles with MadGraph

- Switched from Pythia to MadGraph to better model hard ISR jets
- We generate the sparticle production in standalone MadGraph, with up to 2 extra partons
- xqcut and qcut value for matching/merging determined per process and for different mass ranges
- All other parameters as close as possible to background samples (mainly $t\bar{t}$+jets)
- Sparticles are left undecayed

  ⇒ 2D scan is reduced to 1D line, only depending on mass of the initial sparticle
  ⇒ MadGraph step doesn't need to be redone when new SMS is added

```
import model mssm
define p u u~ d d~ s s~ c c~ b b~ g
define j p
generate p p > go go  @1
add process p p >  go go j  @2
add process p p > go go j j @3
```



Step 1: Production of initial sparticles with Madgraph

LHE file at this stage looks like:

```
<event>
 6   3  0.2305800E-05  0.8719852E+03  0.7957747E-01  0.9571844E-01
       21   -1    0    0  505  501  0.00000000000E+00  0.00000000000E+00  0.26527458742E+04  0.26527458742E+04  0.00000000000E+00 0. -1.
       21   -1    0    0  506  502  0.00000000000E+00  0.00000000000E+00 -0.46740398862E+03  0.46740398862E+03  0.00000000000E+00 0. -1.
  1000021    1    1    2  503  504 -0.20603613134E+03 -0.28228983763E+03  0.12528160604E+04  0.15269860907E+04  0.80000000000E+03 0.  1.
  1000021    1    1    2  504  501  0.16767445981E+03  0.30192564736E+03  0.10208545500E+04  0.13421690777E+04  0.80000000000E+03 0.  1.
       21    1    1    2  505  502 -0.47948663097E+02 -0.34081756567E+02  0.56638803298E+02  0.81661462530E+02  0.00000000000E+00 0.  1.
       21    1    1    2  506  503  0.86310334618E+02  0.14445946843E+02 -0.14496752808E+03  0.16933323194E+03  0.00000000000E+00 0. -1.
#  0.8719852E+03  0.8719852E+03  0.5882721E+02  0.6979681E+02
</event>
```

Available processes:

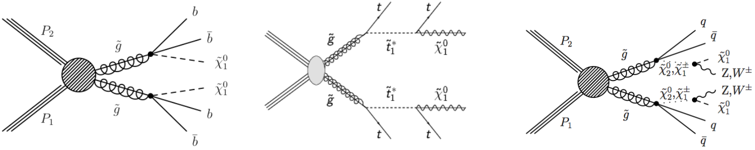
- $pp \to \tilde{g}\tilde{g}$: 400k events per mass point, 50 GeV mass spacing
- $pp \to \tilde{t}_1\tilde{t}_1^*$: $>$ 1M events per mass point, 25 GeV mass spacing; reused for $pp \to \tilde{b}_1\tilde{b}_1^*$

$\rightarrow$ Undecayed LHE files stored on EOS for easy access and use

- $pp \to \tilde{\chi}_2^0\tilde{\chi}_1^\pm$: Produced by analysts and not (yet) stored on EOS

Step 2: Decay of sparticles with Pythia6

The MadGraph LHE files are reused for multiple decay modes, e.g.



Pythia6 is fully embedded within CMSSW

- Configuration through python modules
- Read in undecayed MadGraph LHE file
- Pass SLHA file with details of the SMS to Pythia6
- Only do the decay in this step, considering a flat Matrix Element
- Output a new LHE file to be passed to CMS central production

```
PythiaParameters = cms.PSet(
    pythiaUESettingsBlock,
    processParameters = cms.vstring('MSEL=0','IMSS(1)=11','MSTP(161)=67','MSTP(162)=68','MSTP(163)=69',),
    SLHAParameters = cms.vstring('SLHAFILE = ./SLHACARD'),
```

Step 2: Decay of sparticles with Pythia6

After the decay the LHE files take the following form:

```
<event>
    12      4  1.000000E+00  1.079599E+03  7.957747E-02  9.338893E-02
      21    -1     0     0   101   102  0.0000000000E+00  0.0000000000E+00  2.3398452272E+03  2.3398452272E+03  0.0000000000E+00 0. 9.
      21    -1     0     0   103   104  0.0000000000E+00  0.0000000000E+00 -8.1302109705E+02  8.1302109705E+02  0.0000000000E+00 0. 9.
 1000021     2     1     2   101   104 -4.7167880532E+02 -6.7757210035E+02  3.5564618729E+02  1.2033574104E+03  8.0000000000E+02 0. 9.
 1000021     2     1     2   103   105  2.0351150472E+02  5.8540903390E+02  6.4728251334E+02  1.2012890250E+03  8.0000000000E+02 0. 9.
       1     1     1     2   105     0  2.7331335761E+02 -8.7946876532E+01  1.3021204411E+02  3.1526182923E+02  0.0000000000E+00 0. 9.
      -1     1     1     2     0   102 -5.1460570058E+00  1.8010994298E+02  3.9368338537E+02  4.3295805961E+02  0.0000000000E+00 0. 9.
 1000022     1     3     0     0     0 -4.0964035429E+02 -6.0490373530E+02  3.0542082421E+02  9.9347653656E+02  6.0000000000E+02 0. 9.
      -5     1     3     0     0   104 -1.0554577939E+01 -7.1082827336E+01 -5.6328172542E+01  9.1433421020E+01  4.8000000000E+00 0. 9.
       5     1     3     0   101     0 -5.1483873088E+01 -1.5855377174E+00  1.0655353562E+02  1.1844745279E+02  4.8000000000E+00 0. 9.
 1000022     1     4     0     0     0  7.9257413809E+01  3.4903125804E+02  5.4457882006E+02  8.8581637374E+02  6.0000000000E+02 0. 9.
      -5     1     4     0     0   105  1.9546868616E+01  1.7444863072E+02  4.8316475224E+01  1.8213161892E+02  4.8000000000E+00 0. 9.
       5     1     4     0   103     0  1.0470722230E+02  6.1929145137E+01  5.4387218060E+01  1.3334103234E+02  4.8000000000E+00 0. 9.
#  0.1079599E+04  0.1079599E+04  0.2808574E+03  0.1801834E+03
# model T1bbbb_800_600  0.0742372
</event>
```

Step 2: Decay of sparticles with Pythia6

After the decay the LHE files take the following form:

```
<event>
    12      4  1.000000E+00  1.079599E+03  7.957747E-02  9.338893E-02
      21    -1     0     0   101   102  0.0000000000E+00  0.0000000000E+00  2.3398452272E+03  2.3398452272E+03  0.0000000000E+00 0. 9.
      21    -1     0     0   103   104  0.0000000000E+00  0.0000000000E+00 -8.1302109705E+02  8.1302109705E+02  0.0000000000E+00 0. 9.
 1000021     2     1     2   101   104 -4.7167880532E+02 -6.7757210035E+02  3.5564618729E+02  1.2033574104E+03  8.0000000000E+02 0. 9.
 1000021     2     1     2   103   105  2.0351150472E+02  5.8540903390E+02  6.4728251334E+02  1.2012890250E+03  8.0000000000E+02 0. 9.
       1     1     1     2   105     0  2.7331335761E+02 -8.7946876532E+01  1.3021204411E+02  3.1526182923E+02  0.0000000000E+00 0. 9.
      -1     1     1     2     0   102 -5.1460570058E+00  1.8010994298E+02  3.9368338537E+02  4.3295805961E+02  0.0000000000E+00 0. 9.
 1000022     1     3     0     0     0 -4.0964035429E+02 -6.0490373530E+02  3.0542082421E+02  9.9347653656E+02  6.0000000000E+02 0. 9.
      -5     1     3     0     0   104 -1.0554577939E+01 -7.1082827336E+01 -5.6328172542E+01  9.1433421020E+01  4.8000000000E+00 0. 9.
       5     1     3     0   101     0 -5.1483873088E+01 -1.5855377174E+00  1.0655353562E+02  1.1844745279E+02  4.8000000000E+00 0. 9.
 1000022     1     4     0     0     0  7.9257413809E+01  3.4903125804E+02  5.4457882006E+02  8.8581637374E+02  6.0000000000E+02 0. 9.
      -5     1     4     0     0   105  1.9546868616E+01  1.7444863072E+02  4.8316475224E+01  1.8213161892E+02  4.8000000000E+00 0. 9.
       5     1     4     0   103     0  1.0470722230E+02  6.1929145137E+01  5.4387218060E+01  1.3334103234E+02  4.8000000000E+00 0. 9.
#  0.1079599E+04  0.1079599E+04  0.2808574E+03  0.1801834E+03
# model T1bbbb_800_600  0.0742372
</event>
```

During central production, events from all mass points are put together in one big dataset.

In order to retrieve the mass point information, we add an extra comment string in the LHE file.

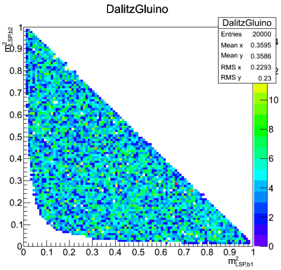
This comment string is read by CMSSW and stored in the final AOD output, allowing users to reconstruct the scan.

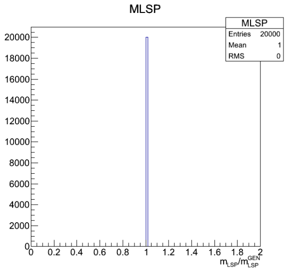
Step 2: Decay of sparticles with Pythia6

To ensure uniformity across the samples, the production of these decayed LHE files is done only by:

1. the SUSY LHE production team
2. the requesters of the sample, in so-called "assisted production mode"

At this stage – before injecting the LHE files in central processing – the LHE files also undergo some basic validation.





Step 3: Injection in CMS official production

Once step 2 has been successfully completed, including the validation, the LHE files are ready to be injected in the CMS official MC production.

The official production takes care of:

1. Parton shower
2. Hadronization
3. Detector simulation
   $\rightarrow$ FastSim, with a few benchmark points in FullSim
4. Reconstruction

Parton shower and hadronization is done using Pythia6, configured by python modules that are passed to CMSSW. Parameters for matching the matrix element to the parton shower are also passed in this way.

```
jetMatching = cms.untracked.PSet(
    scheme = cms.string("Madgraph"),
    mode = cms.string("auto"),
    MEMAIN_nqmatch = cms.int32(5),
    MEMAIN_etaclmax = cms.double(5),
    MEMAIN_qcut = cms.double(44),
    MEMAIN_minjets = cms.int32(0),
    MEMAIN_maxjets = cms.int32(2),
    MEMAIN_showerkt = cms.double(0),
    MEMAIN_excres = cms.string(""),
    )
```

Summary of used resources for SMS production

Production of SMS's takes up a lot of resources
AND requires a lot of time spent on bookkeeping!

Each dataset needs to be acted upon by hand:

- Tested and entered into the official MC production database by the SUSY MC contact
- Approved by the GEN conveners
- Handled by DataOps
- Distributed to Tier2's

This overhead is reduced by combining mass points into a single dataset.

Consequences:

- Less datasets to run over on the grid by analysts
- Need to do decay outside of official production as it is not possible (yet) to pass different SLHA files for one request
- Needed to devise a way to associate events to the correct mass point

Summary of used resources for SMS production

As illustration, some numbers for the ongoing 8TeV SMS production using MadGraph:

- $\mathcal{O}(30)$ SMS's with up to 10 datasets per SMS
- $\mathcal{O}(60M)$ events on average per SMS (100-150k per mass point)
- Total number of datasets = 150 and counting
- Total number of reconstructed events = 1025M and counting
  Multiply this by 3 to get total number of LHE events produced!
- Total amount of storage currently used for one copy of the final AOD output datasets = 375 TB

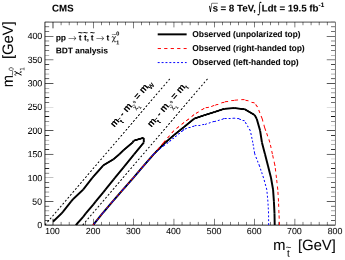
$\Rightarrow$ We cannot produce every possible scenario!

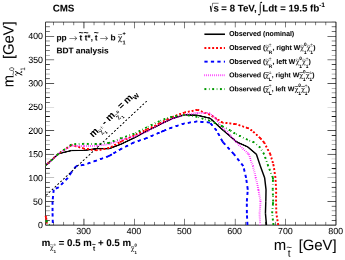
Reweighting events to study different scenarios

The number of samples to be generated can be reduced by performing event reweighting.

A good example of this is the possible polarization of the top quark or chargino in stop decays.

- Generate decays according to a flat ME
  $\Rightarrow$ can reweight afterwards for any specific scenario, e.g. fully-lefthanded tops, or some admixture, ...
- Code was made available to the analysts to perform this reweighting, and documented in a CMS Analysis Note
- See also arXiv:1304.0491 for more details





Modeling of ISR/FSR – MC validation

arXiv:1308.1586 [hep-ex]

The acceptance for signal events depends on the amount of initial-state radiation. This is especially important for compressed spectra as ISR boost is necessary to pass the analysis event selection criteria (e.g. jet multiplicity, E_T^{miss}, lepton p_T)

$\Rightarrow$ Need to validate the ISR modeling in the simulation.

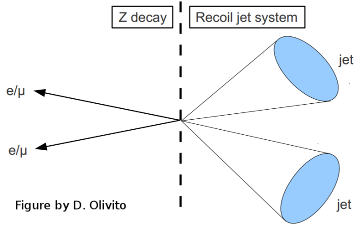
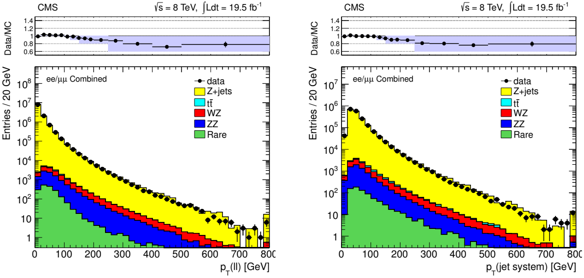


Figure by D. Olivito

- Select data sample of Z+jets, $t\bar{t}$+jets and WZ using dilepton triggers
- Compare MadGraph prediction of the p_T-spectrum of the system recoiling against ISR jets to spectrum observed in data
- For Z+jets, compute $p_T(Z)$ from the dilepton system or from the jet recoil system

Modeling of ISR/FSR – Reweighting recipe



All comparisons show the same behavior within statistical uncertainties:

- good agreement for the low p_T part of the spectrum
- for higher p_T the simulation exceeds the data (10 to 20%)

$\Rightarrow$ Weights are derived, correcting the MC prediction as a function of the p_T of the recoiling system.

- The full values of these weights are used as systematic uncertainty
- This recipe should be applied to all of our MadGraph SMS scans

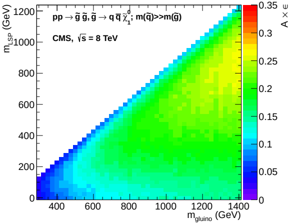
Treatment of systematic uncertainties

When using SMS scans in an analysis, a set of systematic uncertainties has to be taken into account. Most of these are common for any use of MC samples:

- Pileup reweighting: $< 1\%$
- Trigger efficiencies: $\mathcal{O}(5\%)$
- Luminosity: 2.6%
- Jet energy scale and resolution: $\mathcal{O}(2-20\%)$ depending on the model and analysis
- Lepton identification and isolation efficiencies: $\mathcal{O}(5\%)$
- B-tagging: $\mathcal{O}(1-4\%)$
- PDF uncertainties
- ISR modeling

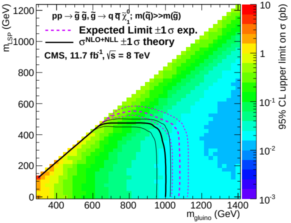
$\Rightarrow$ Total uncertainty is usually $\mathcal{O}(10-30\%)$ depending on the model and analysis

Setting limits



Acceptance maps

$\Rightarrow$



95% CL Upper limits

Ingredients needed to calculate limits

- Acceptance and associated uncertainty per mass point
- Background estimate and its uncertainty
- Code to run LHC-style CLs, e.g. Combine, LandS or custom code
- Depending on the details of the analysis (event counts, amount of nuisances): a lot of time

Summary

- Shift from cMSSM to SMS's tailored for "Natural SUSY"
- Changed from Pythia6 to MadGraph + Pythia6 to improve the accuracy of the modeling of the associated jet production
- SMS production in CMS is a three-step process
  1. Production in MadGraph
  2. Decay in Pythia6
  3. Official CMS FastSim
- SMS production takes up a lot of resources, so we cannot produce every scenario
- Flat Matrix Elements are used, so we can reweight afterwards

Questions?

Comments?

Suggestions?

BACKUP

Matching/Merging parameters

Process	xqcut	qcut
$\tilde{g}\tilde{g}$, $m_{\tilde{g}} < 800$ GeV	30	45
$\tilde{g}\tilde{g}$, $m_{\tilde{g}} \geq 800$ GeV	30	50
$\tilde{t}\tilde{t}^*$, $m_{\tilde{t}} < 500$ GeV	30	44
$\tilde{t}\tilde{t}^*$, $m_{\tilde{t}} \geq 500$ GeV	30	46
$\tilde{\chi}\tilde{\chi}$	15	23