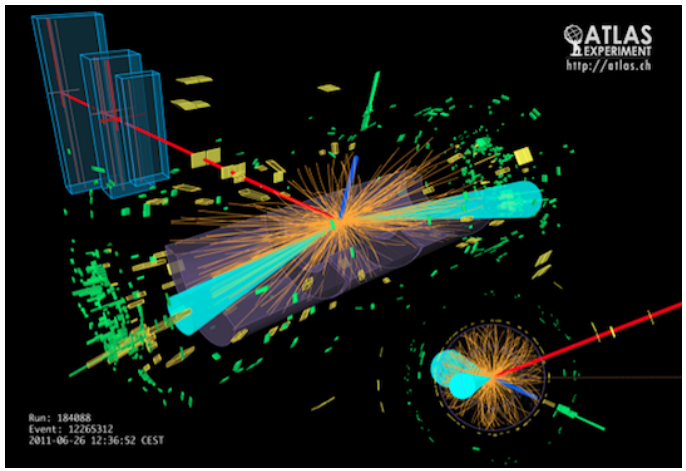
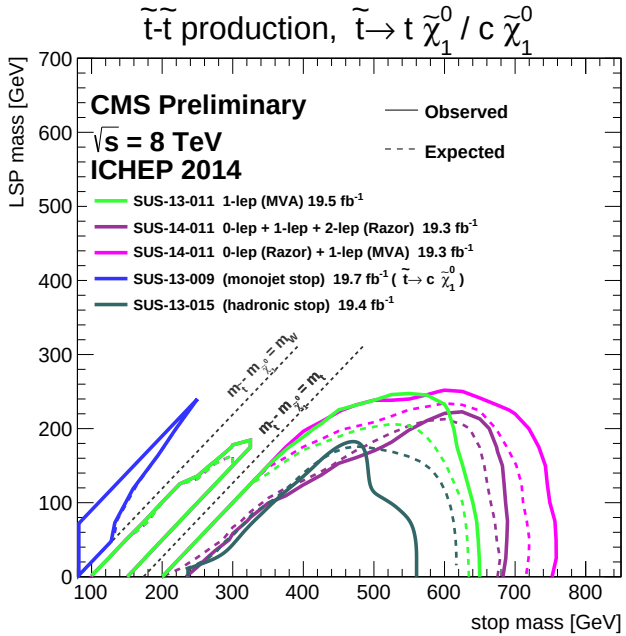
RECASTING AND REINTERPRETING LHC SEARCHES WITH MADANALYSIS 5

Guillaume CHALONS

LPSC Grenoble

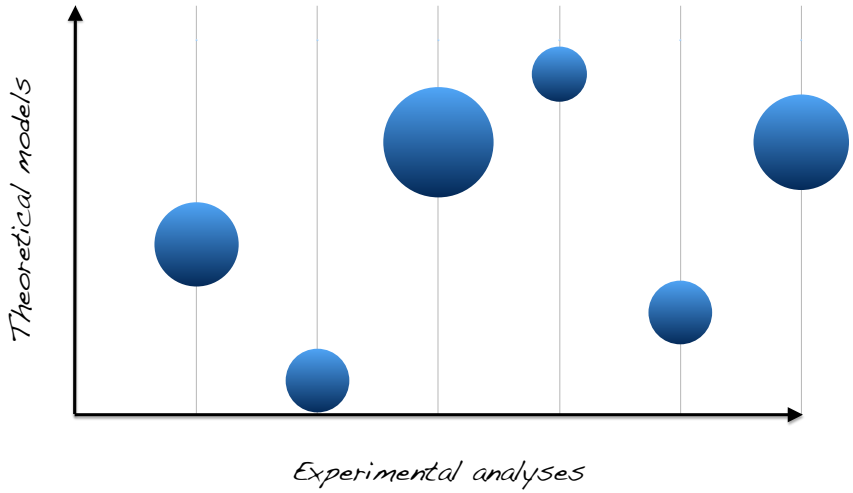




In collaboration with S. Bein, E. Conte, B. Dumont, B. Fuks, S. Kraml S. Kulkarni, D. Sengupta, L. Mitzka, C. Wymant and others

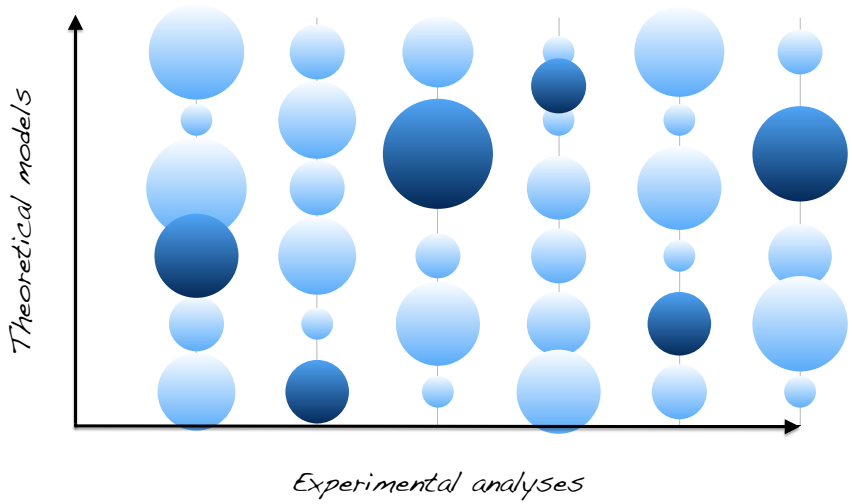
THE NEED FOR REINTERPRETATION STUDIES/TOOLS

☞ # Theo. Models. $\gg$ # Exp. Analyses



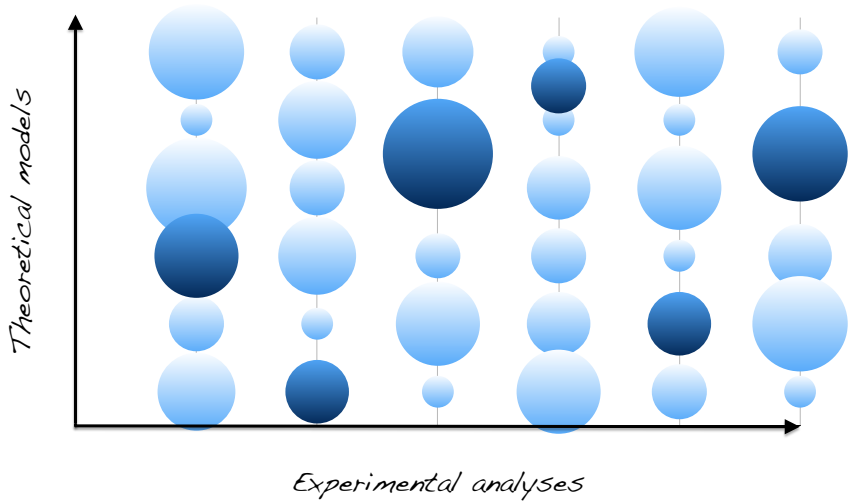
THE NEED FOR REINTERPRETATION STUDIES/TOOLS

- ☞ # Theo. Models. $\gg$ # Exp. Analyses
- ☞ Constrain any model not covered with already existing analyses



THE NEED FOR REINTERPRETATION STUDIES/TOOLS

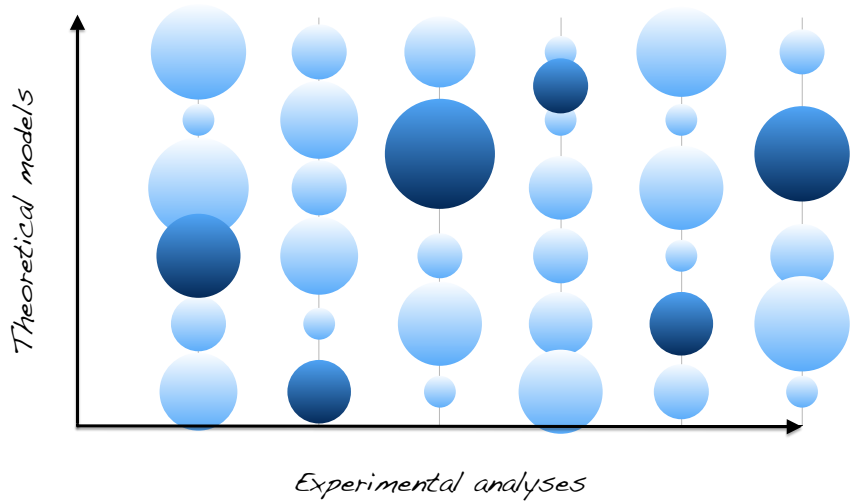
- ☞ # Theo. Models. $\gg$ # Exp. Analyses
- ☞ Constrain any model not covered with already existing analyses



Experiment $\Longleftrightarrow$ Theory

THE NEED FOR REINTERPRETATION STUDIES/TOOLS

- # Theo. Models. $\gg$ # Exp. Analyses
- Constrain any model not covered with already existing analyses



Experiment $\Longleftrightarrow$ Theory

- Improve the documentation
- Preserve the data
- Identify coverage of the existing analyses
- Elaborate new search strategies

TWO CATEGORIES OF REINTERPRETATION TOOLS

- Several groups have been developping private codes for recasting BSM searches
- A number of public tools have become available recently

SIMPLIFIED MODELS (SMS)

- SModelS: generic decomposition into SMS topologies, cross section upper limits from more than 50 ATLAS and CMS SMS results
  [Kraml et al., 1312.4175]
- Fastlim: reconstructs visible cross section for SMS topologies from precalculated efficiency and cross section tables; currently 11 ATLAS analyses implemented
  [Papucci et al., 1402.0492]
- XQCAT:determines the CLs for heavy extra quarks based on efficiency maps, CMS search for top partners plus 2 SUSY searches at 8 TeV
  [Barducci et al., 1405.0737]

EVENT SIMULATION

- CheckMATE: check 95% CL for simulated events of any model; currently 10 ATLAS and 2 CMS SUSY analyses implemented
  [Drees et al., 1312.2591]
- MA5 PAD: public analysis database within the MadAnalysis 5 framework; currently 4 ATLAS and 3 CMS analyses, more in progress
  [Dumont, GC, et al., 1407.3278]

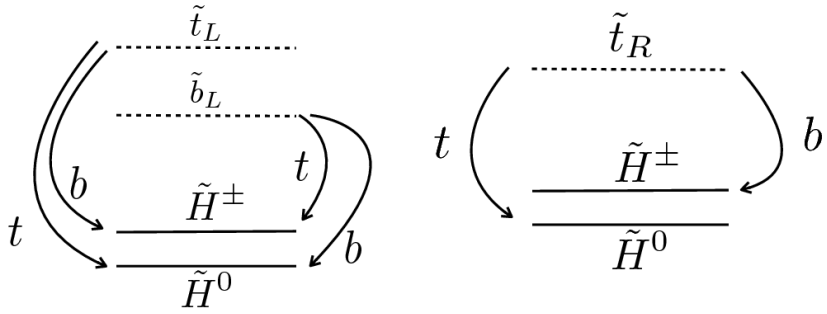
Or resort to the RECAST initiative

The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)

The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)

A LITTLE EXAMPLE:

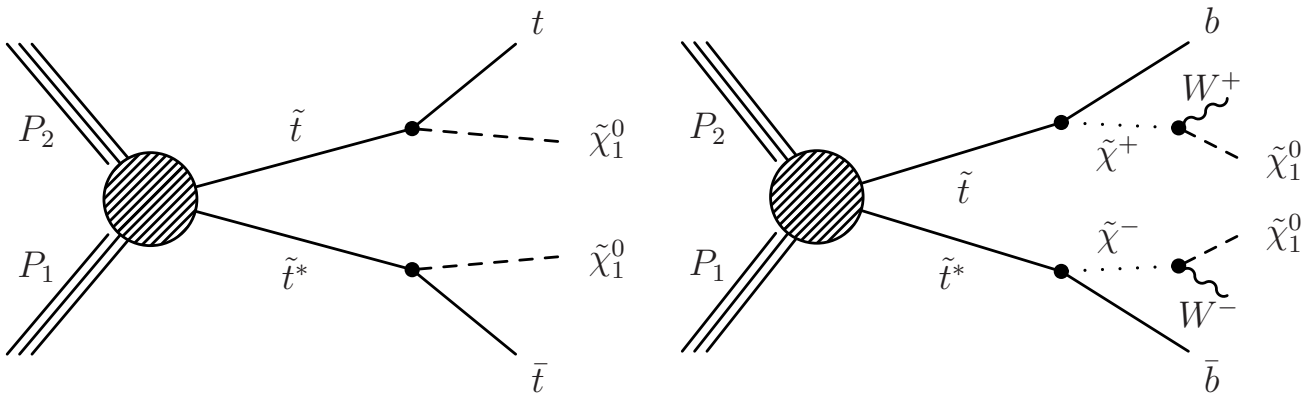
- Natural SUSY scenarios do not predict $BR(\tilde{t}_1, \tilde{b}_1 \rightarrow X) = 100\%$



The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)

A LITTLE EXAMPLE:

- Natural SUSY scenarios do not predict $BR(\tilde{t}_1, \tilde{b}_1 \rightarrow X) = 100\%$

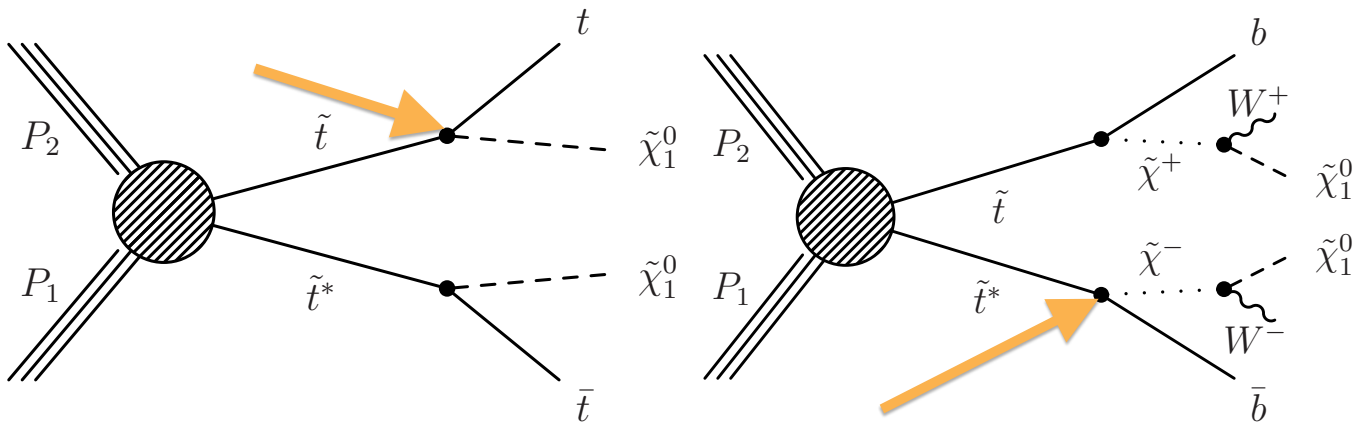


- CMS-SUS-13-011 ($\ell + \not{E}_T$) targets each topology in each SR
- In a typical Natural SUSY scenario the decay chains of $\tilde{t}_1$ can be mixed and would lead to the same final state ($1\ell + 2b + 2j + \not{E}_T$)
- What would be the sensitivity of this analysis to this mixed topology ?

The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)

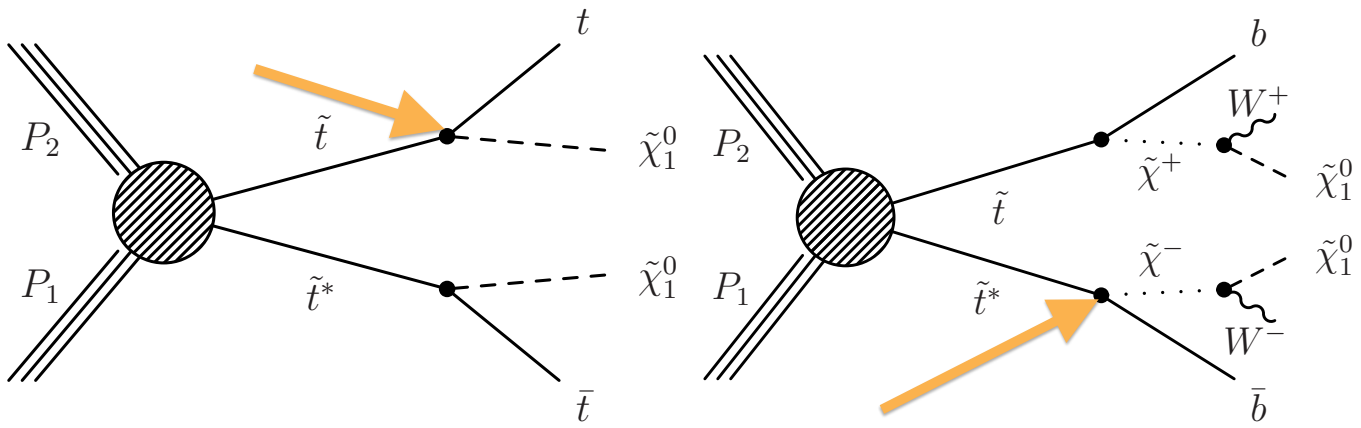
A LITTLE EXAMPLE:

- Natural SUSY scenarios do not predict $BR(\tilde{t}_1, \tilde{b}_1 \rightarrow X) = 100\%$



- CMS-SUS-13-011 ($\ell + \not{E}_T$) targets each topology in each SR
- In a typical Natural SUSY scenario the decay chains of $\tilde{t}_1$ can be mixed and would lead to the same final state ($1\ell + 2b + 2j + \not{E}_T$)
- What would be the sensitivity of this analysis to this mixed topology ?

The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)



- 3rd generation SUSY searches can also be used to constrain vector-like quarks models (Composite Higgs, X-Dim, Little Higgs . . .)
- As soon as one goes **beyond** the SMS (ex: relaxing 100% BR's) one has to **recast** the analysis

THE STATUS

The Simplified Models cover many topologies but have limitations (signal efficiencies depend on the event kinematics, not details of BSM model)

Going beyond the SMS approach requires a fast detector simulation

For a given topology one needs to :

- Scan over parameter space including event generation
- Implement some of the related existing experimental analyses
- Validate the implementations
- Then apply to different frameworks

The task is huge!

- Need to iterate for each topology
- A lot of manpower needed
- Some analyses may have been already implemented by other groups but validation of the implementation not always public.

Instead of reinventing the wheel and to avoid redundancy we may want to share the effort $\Longrightarrow$ **Creation of a publicly available database of reimplemented analyses**

RECASTING AND DESIGNING LHC PHYSICS ANALYSES

Determining the sensitivity of the LHC to a NP model by analyzing specific signatures

RECASTING

1) Read & understand the experimental paper
2) Write the C++ analysis code
3) Get missing info from collab
   - Efficiencies
   - MC config
   - benchmark points
   - cutflows
   - Exp. final # events in each SR
4) Digitize histograms (if not on HEPData, Twiki)
5) Produce your own cutflows/histos, compare,discuss with collab., iterate until reasonable agreement

DESIGNING

1) Generation of bckgd&signal MC samples
2) Analysis by tuning thresholds
3) Increase a factor of merit of choice

LPSC Grenoble

WHY MADANALYSIS 5 ?

Conte, Fuks, Serret, **CPC (2013)**; Conte, Dumont, Fuks, Wymant, **EPJC (2014)**

- Expert mode extended to recast existing LHC analyses
  - C++/ROOT language within the SAMPLEANALYZER framework
  - Supports multiple sub-analyses, a developer-friendly way of handling cuts and histograms

- Results are stored in a so-called SAF file
  - Text based (similar to XML)
  - Easily reprocessable

```
<SAFHEADER>
</SAFHEADER>

<HISTO>
<DESCRIPTION>
"MCT SRA"
# NBINS        XMIN           XMAX
20             0              500
# ASSOCIATED REGIONSELECTIONS
 SRA, HIGHDELTAM, MET > 150                          # REGION NR. 1
</DESCRIPTION>
<STATISTICS>
326 0 # NEVENTS
326 0 # SUM OF EVENT-WEIGHTS OVER EVENTS
326 0 # NENTRIES
326 0 # SUM OF EVENT-WEIGHTS OVER ENTRIES
326 0 # SUM WEIGHTS^2
39046.3 0 # SUM VALUE*WEIGHT
5.28533E+06 0 # SUM VALUE^2*WEIGHT
</STATISTICS>
```

- Modified version of DELPHES 3
  - Isolation information kept

- Recasting Tools
  - Import/Export Analyses
  - A (light) module to provide a comparison with the N_{95} upper limit or a simplified CLs
  - Efficiency map calculation (!!! NEW !!!)

AIM OF THE MA5 PAD

Our aim : reimplement several ATLAS and CMS analyses and developp a public analysis database (PAD) within a common platform for collecting objects definitions, cuts, etc...

To draw limits and/or interpret a deviation from the SM expectation one needs

- # expected bckgd events & # of observed events from physics paper
- # of expected signal events after cuts for a given NP model

MADANALYSIS5 is designed to take care of the last item : takes a simulated event sample, pass it to detector simulation and then analysis code

To validate the analysis we have to rely on our own detector simulation

Try to reproduce the official cutflows and distributions given in the analysis paper

Very tedious (given the available information)

CURRENT STATUS OF IMPLEMENTATION/VALIDATION

"Towards a public analysis database for LHC NP searches using MADANALYSIS 5":
B. Dumont, B. Fuks, S. Kraml, **G.C** *et. al* published in **EPJC** (2015)

- CMS searches:
  - CMS-SUS-13-011 ($\tilde{t}$ search, 1 ℓ-analysis)
  - CMS-SUS-13-016 ($\tilde{g}$ search, 2 OS ℓ, large $\not{E}_T$ & High Jet-multiplicity)
  - CMS-SUS-13-012 ($\tilde{q}, \tilde{g}$ search, multijet + large $\not{E}_T$)
  - CMS-B2G-13-003 (vector-like quarks in multilepton events)
  - CMS-SUS-13-002 ($\geqslant 3\ell +$ $\not{E}_T$)
  - ...

- ATLAS searches:
  - ATLAS-SUSY-13-05 ($\tilde{t}, \tilde{b}$ search, 2 b-jets + $\not{E}_T$)
  - ATLAS-SUSY-13-11 (EWK-inos, 2 $\ell +$ $\not{E}_T$)
  - ATLAS-HIGGS-13-03 ($Z^0 H \rightarrow \ell\ell$ + invisible)
  - ATLAS-EXOT-14-06 (mono-photons + $\not{E}_T$)
  - ATLAS-SUSY-13-21 ($\tilde{t} \rightarrow c +$ $\not{E}_T$, monojet+ctagged)
  - ATLAS-SUSY-13-02 ($0\ell + 2 - 6$ jets + $\not{E}_T$)
  - ATLAS-SUSY-13-19 ($\tilde{t}$ search, 2 OS ℓ analysis)

- This is a non-exhaustive list of ongoing implementations/validations
- Since we started the project there has been a clear **improvement** in the **quality** of the documentation from **both** experiments

MA5 PUBLIC ANALYSIS DATABASE

http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase

Available Analyses

!! please properly cite all the re-implementation codes you are using (see Inspire citation entry) !!

ATLAS analyses, 8 TeV

Analysis	Short Description	Implemented by	Code	Validation note	Status
ATLAS-SUSY-2013-05 (published)	stop/sbottom search: 0 leptons + 2 b-jets	G. Chalons	Inspire	PDF (figures)	done
ATLAS-SUSY-2013-11 (published)	EWK-inos, 2 leptons + MET	B. Dumont	Inspire	PDF (source)	done
ATLAS-HIGG-2013-03 (published)	ZH->ll+invisible	B. Dumont	Inspire	PDF (source)	done
ATLAS-EXOT-2014-06 (published)	mono-photons + MET	D. Barducci	Inspire	PDF	done

Delphes card for ATLAS-SUSY-2013-05 and ATLAS-EXOT-2014-06
Delphes card for ATLAS-SUSY-2013-11

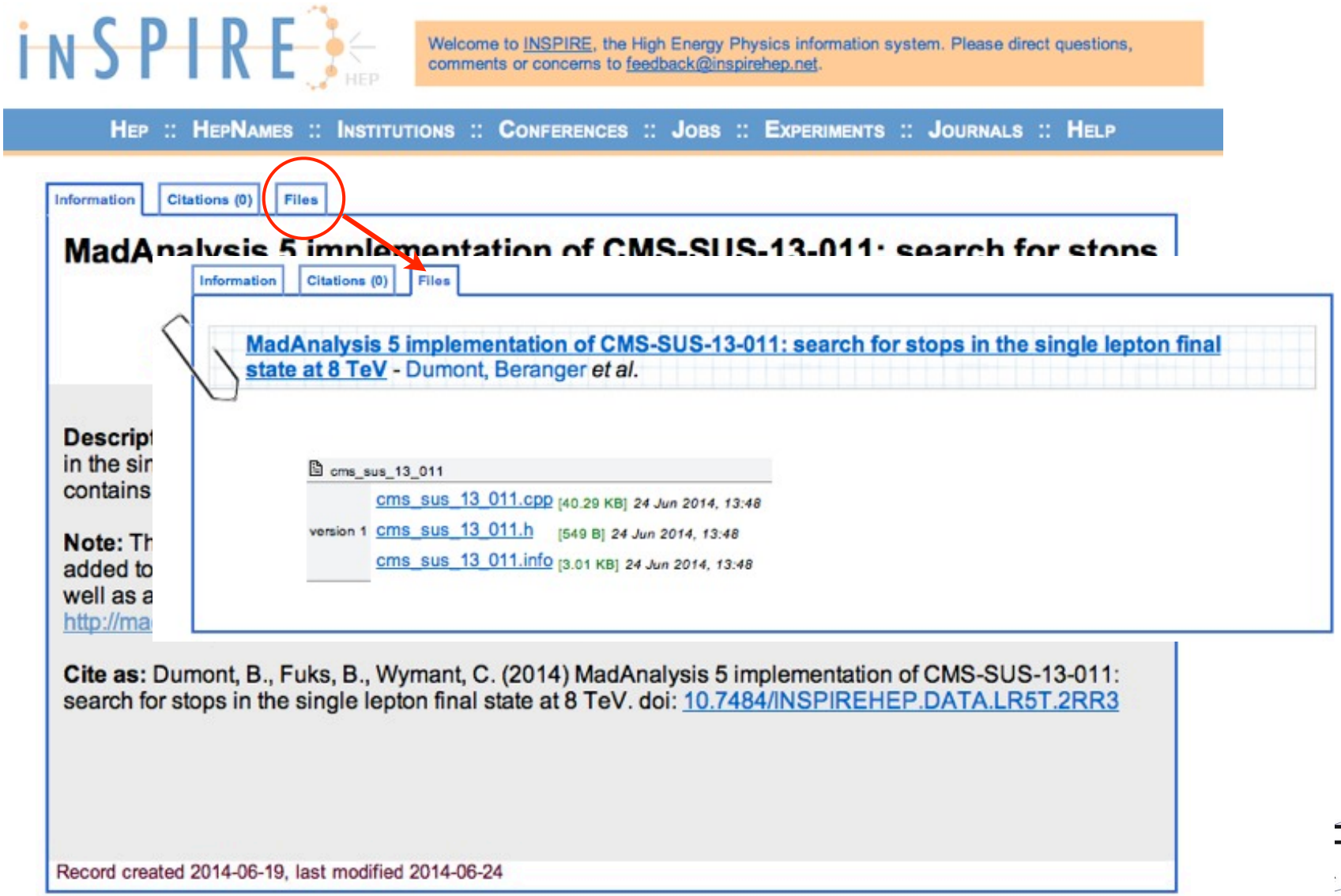
CMS analyses, 8 TeV

Analysis	Short Description	Implemented by	Code	Validation note	Status
CMS-SUS-13-011 (published)	stop search in the single lepton mode	B. Dumont, B. Fuks, C. Wymant	Inspire [1]	PDF (source)	done
CMS-SUS-13-012 (published)	gluino/squark search in jet multiplicity and missing energy	S. Bein, D. Sengupta	Inspire	PDF (source)	done
CMS-SUS-13-016 (PAS)	search for gluinos using OS dileptons and b-jets	D. Sengupta, S. Kulkarni	Inspire	PDF (source)	done

Delphes card for these analyses

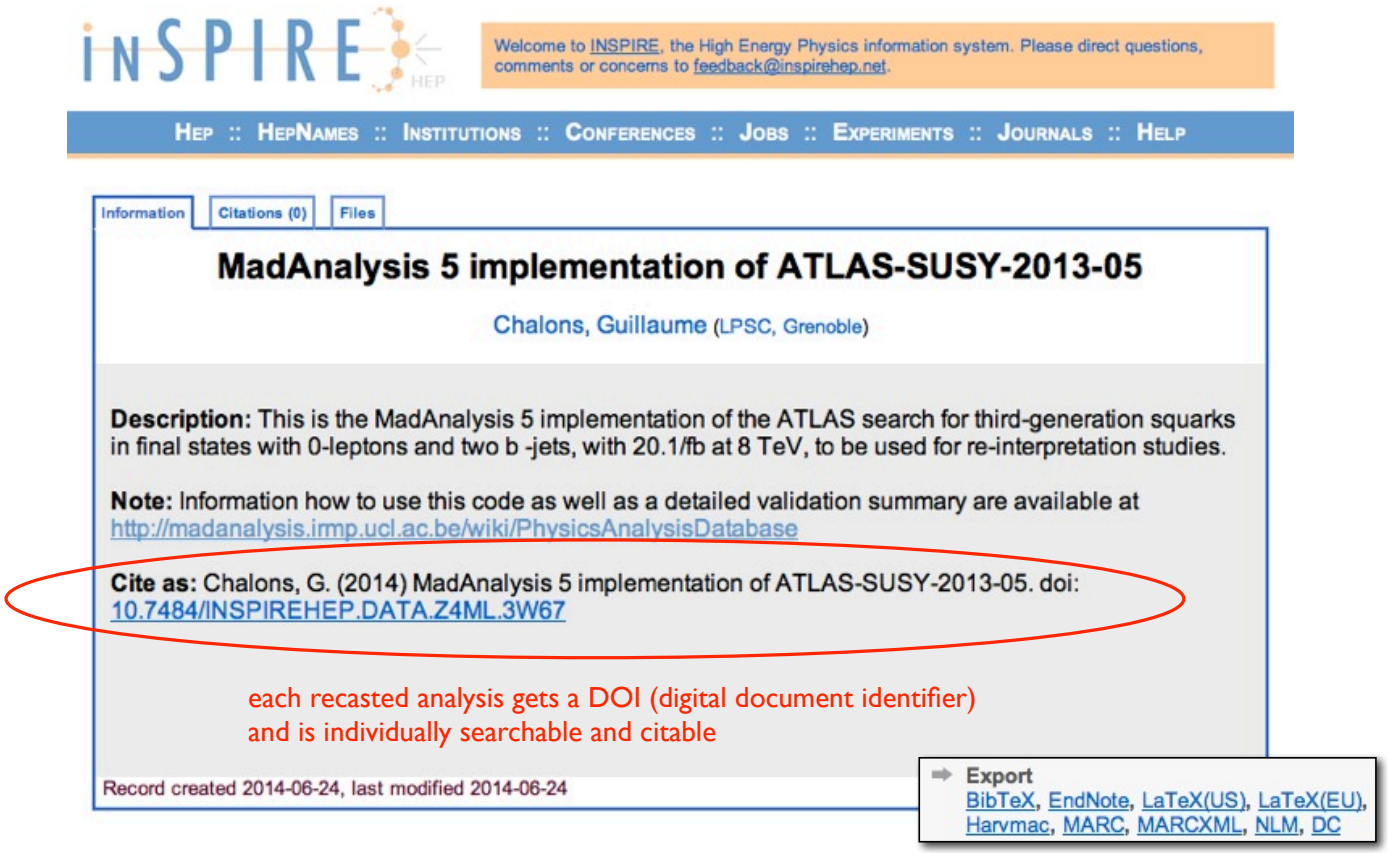
PAD: SUBMIT YOUR CODE

http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase



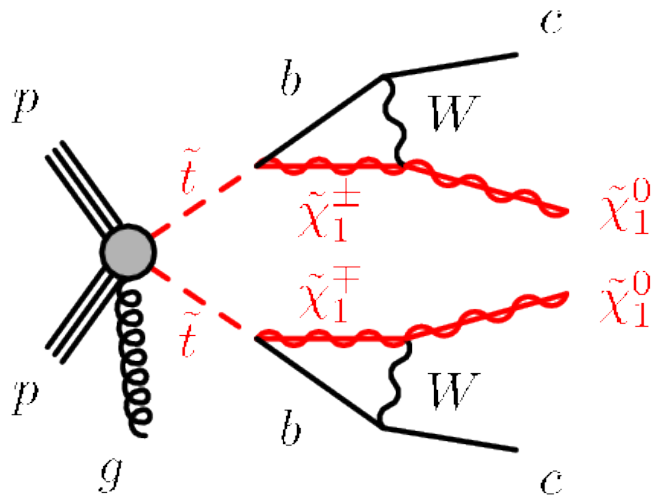
PAD: SUBMIT YOUR CODE

http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase



A VALIDATED ANALYSIS: ATLAS-SUSY-2013-21

- This analysis targets direct $\tilde{t}_1$ pair production in compressed spectra scenarios
- In particular it is optimised for $\tilde{t}_1 \rightarrow c + \tilde{\chi}_1^0$ using a monojet and c-tagged search strategies



- We only implemented the monojet search since
  - It can be reinterpreted in DM or other compressed spectra scenarios
  - We do not have access to the needed charm-tagging information

ATLAS-SUSY-2013-21:CUTFLOW

	$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/125) cutflow		$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/195) cutflow	
cut	# events (scaled to σ and $\mathcal{L}$)	# events (official)	# events (scaled to σ and $\mathcal{L}$)	# events (official)
Initial # of events	376047.3		376047.3	
ALL		181902.0		103191.0

	$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/125) cutflow		$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/195) cutflow	
cut	# events (scaled to σ and $\mathcal{L}$)	# events (official)	# events (scaled to σ and $\mathcal{L}$)	# events (official)
Initial # of events	376047.3		376047.3	
$\not{E}_T > 80$ GeV Filter	192812.8 (−48.7%)	181902.0	104577.6 (−72.2%)	103191.0
$\not{E}_T > 100$ GeV	136257.1 (−29.3%)	97217.0	82619.0 (−21.0%)	64652.0 (−37.3%)

ATLAS-SUSY-2013-21:CUTFLOW

	$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/125) cutflow		$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/195) cutflow	
cut	# events (scaled to σ and $\mathcal{L}$)	# events (official)	# events (scaled to σ and $\mathcal{L}$)	# events (official)
Initial # of events	376047.3		376047.3	
$\not{E}_T > 80$ GeV Filter	192812.8 (−48.7%)	181902.0	104577.6 (−72.2%)	103191.0
$\not{E}_T > 100$ GeV	136257.1 (−29.3%)	97217.0	82619.0 (−21.0%)	64652.0 (−37.3%)
Trigger, ...	-	82131.0 (−15.5%)	-	57566.0 (−30.3%)

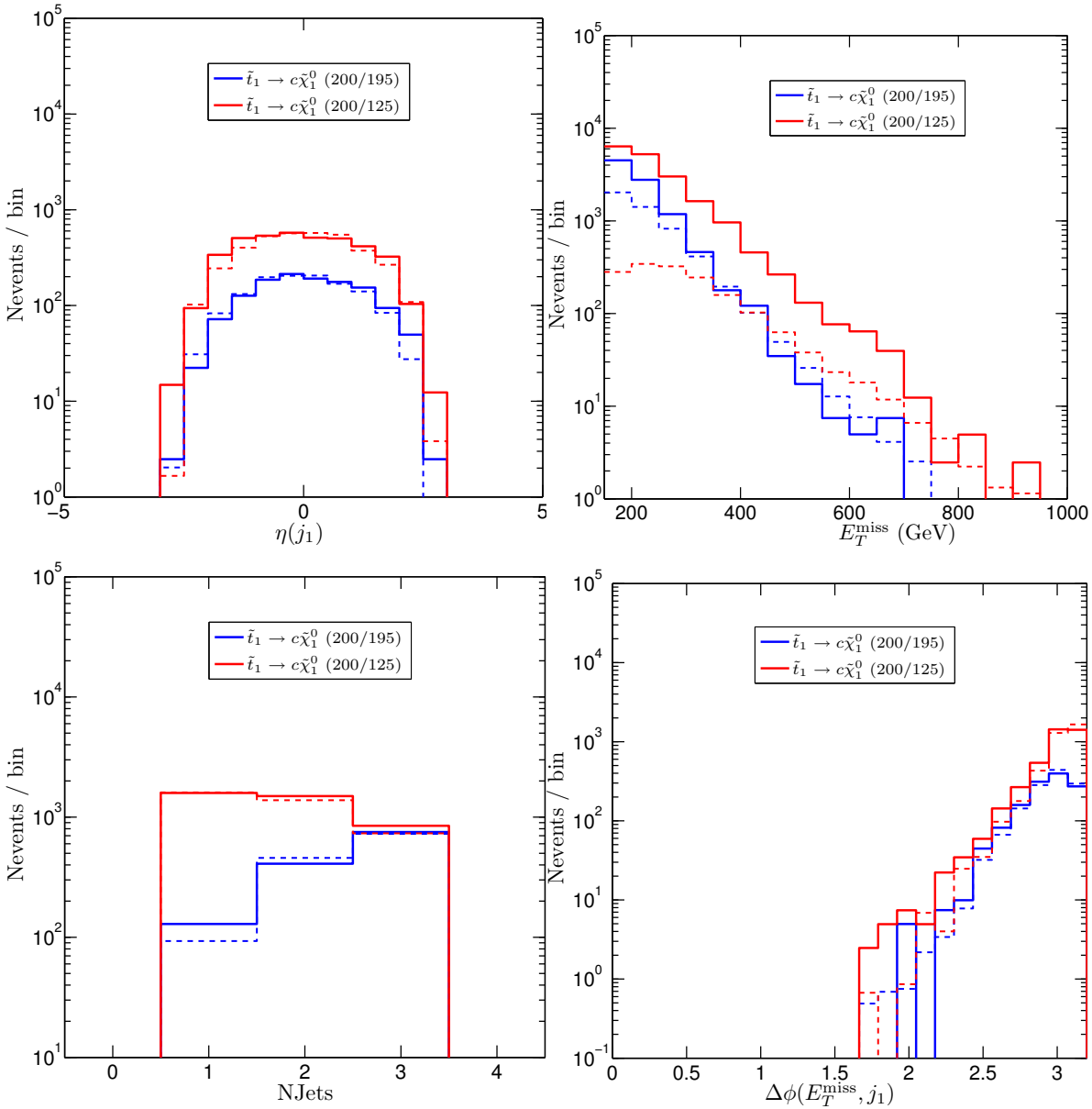
ATLAS-SUSY-2013-21:CUTFLOW

	$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/125) cutflow		$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/195) cutflow	
cut	# events (scaled to σ and $\mathcal{L}$)	# events (official)	# events (scaled to σ and $\mathcal{L}$)	# events (official)
Initial # of events	376047.3		376047.3	
$\not{E}_T > 80$ GeV Filter	192812.8 (−48.7%)	181902.0	104577.6 (−72.2%)	103191.0
$\not{E}_T > 100$ GeV	136257.1 (−29.3%)	97217.0	82619.0 (−21.0%)	64652.0 (−37.3%)
Trigger, ...	-	82131.0 (−15.5%)	-	57566.0 (−30.3%)
Lepton veto	134894.2 (−1.0%)	81855.0 (−15.8%)	82493.9 (−0.2%)	57455.0 (−11.1%)
$N_{\text{jets}} \leq 3$	101653.7 (−24.6%)	59315.0 (−27.5%)	75391.5 (−8.6%)	52491.0 (−8.6%)
$\Delta\phi(\not{E}_T, \text{jets}) > 0.4$	95568.8 (−2.1%)	54295.0 (−8.5%)	70888.1 (−1.2%)	49216.0 (−6.2%)
$p_T(j_1) > 150$ GeV	17282.8 (−81.9%)	14220.0 (−73.8%)	25552.0 (−64.0%)	20910.0 (−57.5%)
$\not{E}_T > 150$ GeV	10987.8 (−36.4%)	9468.0 (−33.4%)	21569.1 (−15.6%)	18297.0(−12.5%)
M1 Signal Region				

ATLAS-SUSY-2013-21:CUTFLOW

	$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/125) cutflow		$\tilde{t} \longrightarrow c\tilde{\chi}_1^0$ (200/195) cutflow	
cut	# events (scaled to σ and $\mathcal{L}$)	# events (official)	# events (scaled to σ and $\mathcal{L}$)	# events (official)
Initial # of events	376047.3		376047.3	
$\not{E}_T > 80$ GeV Filter	192812.8 (−48.7%)	181902.0	104577.6 (−72.2%)	103191.0
$\not{E}_T > 100$ GeV	136257.1 (−29.3%)	97217.0	82619.0 (−21.0%)	64652.0 (−37.3%)
Trigger, ...	-	82131.0 (−15.5%)	-	57566.0 (−30.3%)
Lepton veto	134894.2 (−1.0%)	81855.0 (−15.8%)	82493.9 (−0.2%)	57455.0 (−11.1%)
$N_{\text{jets}} \leq 3$	101653.7 (−24.6%)	59315.0 (−27.5%)	75391.5 (−8.6%)	52491.0 (−8.6%)
$\Delta\phi(\not{E}_T, \text{jets}) > 0.4$	95568.8 (−2.1%)	54295.0 (−8.5%)	70888.1 (−1.2%)	49216.0 (−6.2%)
$p_T(j_1) > 150$ GeV	17282.8 (−81.9%)	14220.0 (−73.8%)	25552.0 (−64.0%)	20910.0 (−57.5%)
$\not{E}_T > 150$ GeV	10987.8 (−36.4%)	9468.0 (−33.4%)	21569.1 (−15.6%)	18297.0(−12.5%)
M1 Signal Region				
$p_T(j_1) > 280$ GeV	2031.2 (−81.5%)	1627.0 (−82.8%)	4922.0 (−77.2%)	3854.0 (−78.9%)
$\not{E}_T > 220$ GeV	1517.6 (−25.3%)	1276.0 (−21.6%)	4628.4 (−6.0%)	3722.0 (−3.4%)
M2 Signal Region				
$p_T(j_1) > 340$ GeV	858.0 (−92.2%)	721.0 (−92.4%)	2509.0 (−88.4%)	1897.0 (−89.6%)
$\not{E}_T > 340$ GeV	344.4 (−59.9%)	282.0 (−60.9%)	1758.9 (−29.9%)	1518.0 (−20.0%)
M3 Signal Region				
$p_T(j_1) > 450$ GeV	204.3 (−98.1%)	169.0 (−98.2%)	773.3 (−96.4%)	527.0 (−97.1%)
$\not{E}_T > 450$ GeV	61.3 (−70.0%)	64.0 (−62.1%)	476.8 (−38.3%)	415.0 (−21.3%)

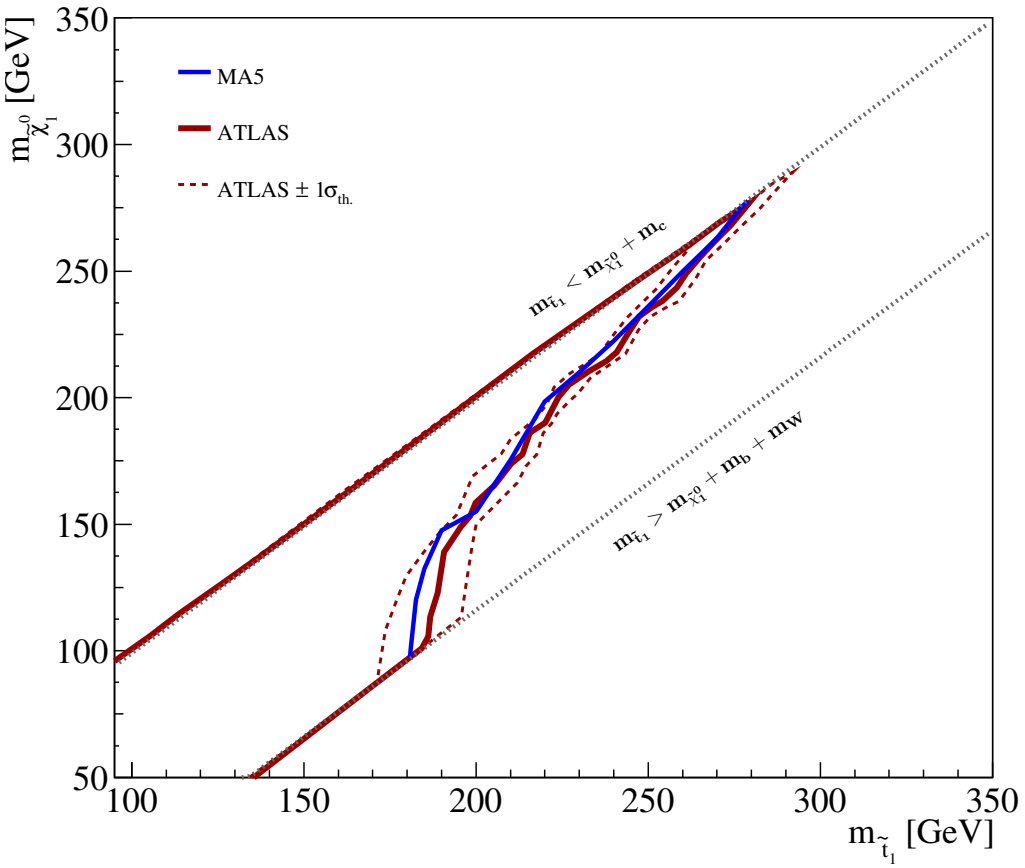
ATLAS-SUSY-2013-21:DISTRIBUTIONS



REPRODUCTION OF THE EXCLUSION PLOT

To derive limits we use a lightweight exclusion code picking the most sensitive SR via

$$\mathcal{L} = \text{poiss}(n_i^{obs\cdot} | n_i^s + n_i^b) \cdot \text{gauss}(n_i^b | n_i^{b,exp}, \Delta n_i^b)$$



- We aim at a precision of order 20-30% on the limit setting

REINTERPRETATION STUDIES USING THE MA5-PAD

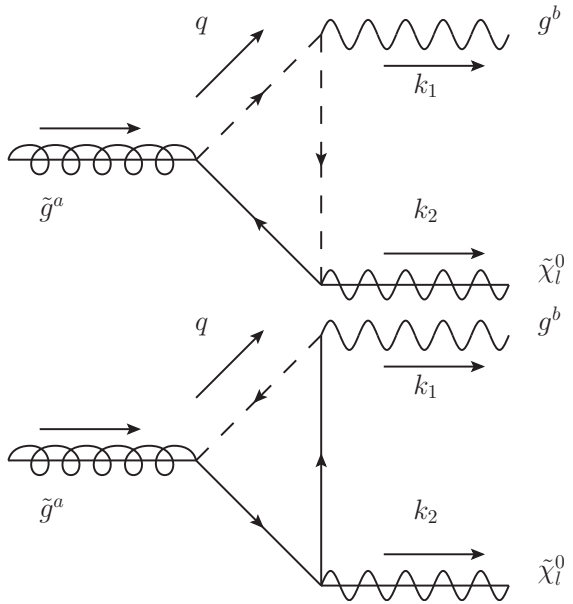
"Probing top-philic sgluons with LHC Run I data ":
L. Beck *et. al* arXiv:1501.07580 [hep-ph] to appear in **PLB** (2015)
"Dilepton constraints in the Inert Doublet Model from Run 1 of the LHC":
G. Bélanger *et. al* arXiv:1503.0736 [hep-ph]

"Probing top-philic sgluons with LHC Run I data ":
L. Beck *et. al* arXiv:1501.07580 [hep-ph] to appear in **PLB** (2015)
"Dilepton constraints in the Inert Doublet Model from Run 1 of the LHC":
G. Bélanger *et. al* arXiv:1503.0736 [hep-ph]

Work in Progress with D. Sengupta

- SMS Exp. limits on $m_{\tilde{g}}$ use $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0, b\bar{b}\tilde{\chi}_1^0, t\bar{b}\tilde{\chi}_1^{\pm}$
- Loss **sensitivity** when threshold is **closed**

- Not considered : $\tilde{g} \to g\tilde{\chi}_i^0$
- Existing ATLAS & CMS analyses having the **same signature**
- Reinterpretation in a SMS model using CMS multijet analysis



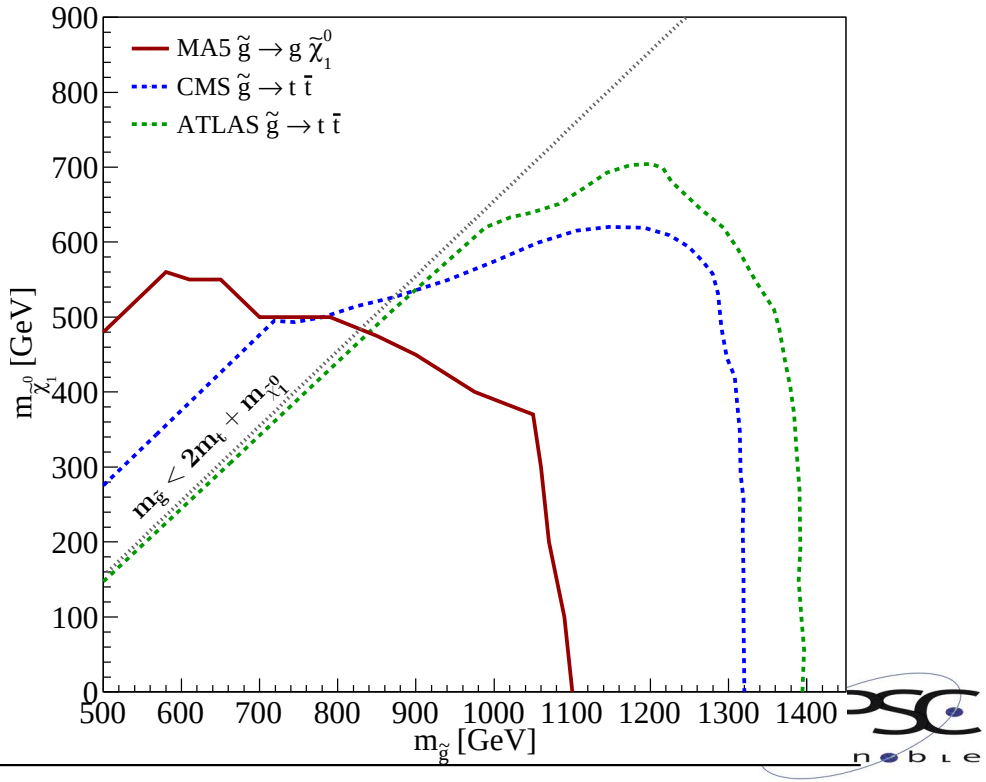
REINTERPRETATION STUDIES USING THE MA5-PAD

"Probing top-philic sgluons with LHC Run I data ":
L. Beck *et. al* arXiv:1501.07580 [hep-ph] to appear in **PLB** (2015)
"Dilepton constraints in the Inert Doublet Model from Run 1 of the LHC":
G. Bélanger *et. al* arXiv:1503.0736 [hep-ph]

Work in Progress with D. Sengupta

- SMS Exp. limits on $m_{\tilde{g}}$ use $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0, b\bar{b}\tilde{\chi}_1^0, t\bar{b}\tilde{\chi}_1^{\pm}$
- Loss **sensitivity** when threshold is **closed**

- Not considered : $\tilde{g} \to g\tilde{\chi}_i^0$
- Existing ATLAS & CMS analyses having the **same signature**
- Reinterpretation in a SMS model using CMS multijet analysis



CONCLUSIONS

- A publicly available analysis database would strengthen LHC legacy
  - Re-use of old analysis
  - in case of NP discovery at 14 TeV → was it already hiding in 7-8 TeV ?
    → consistency with 7-8 TeV run ?
- This is of utmost importance since there is a lot of turnover (PhD, Post-Doc's) in the HEP community
  - some analyses/data of 7 TeV run already lost

LPSC Grenoble

CONCLUSIONS

- A publicly available analysis database would strengthen LHC legacy
  - Re-use of old analysis
  - in case of NP discovery at 14 TeV → was it already hiding in 7-8 TeV ?
    → consistency with 7-8 TeV run ?
- This is of utmost importance since there is a lot of turnover (PhD, Post-Doc's) in the HEP community
  - some analyses/data of 7 TeV run already lost

- Efficient communication between Theo. & Exp crucial
- Common effort to define new strategies to cover "theory" parameter space
- `MadAnalysis5` can be used to generate efficiency maps for Tools like SModels, XQCAT to avoid detector simulation for each model's benchmark.

CONCLUSIONS

- A publicly available analysis database would strengthen LHC legacy
  - Re-use of old analysis
  - in case of NP discovery at 14 TeV → was it already hiding in 7-8 TeV ?
    → consistency with 7-8 TeV run ?
- This is of utmost importance since there is a lot of turnover (PhD, Post-Doc's) in the HEP community
  - some analyses/data of 7 TeV run already lost

- Efficient communication between Theo. & Exp crucial
- Common effort to define new strategies to cover "theory" parameter space
- `MadAnalysis5` can be used to generate efficiency maps for Tools like SModels, XQCAT to avoid detector simulation for each model's benchmark.

- More pragmatically:
  - High Level of information from each experimental analysis desirable
  - `LHE` files available or at least input files for MC (`SLHA+` configuration)
  - preselection informations crucial (trigger/ID/b-tagging efficiencies)
  - Follow more systematically Les Houches Recommandation for presenting public results arXiv:1203.2489
- Since the beginning of the project the quality of the documentation has greatly improved for Run I latest analyses
- We hope this will continue for Run II