LHC recasting & reinterpretation tools

Eric Conte (IPHC-GRPHE)

Tools 2017: Tools for the SM and the New Physics

10-13 September 2017 @ Corfu (Greece)

Outlines

1. Motivations
2. Recasting based on SMS
3. Recasting based on detector simulation
4. Summary & opened questions

1. **Motivations**

2. Recasting based on SMS

3. Recasting based on detector simulation

4. Summary & opened questions

BSM researches @ LHC

Intensive search of BSM is ongoing at the LHC:

- Looking for promising signatures (SMS, bottom-up approach, …)
- Setting limits

  → Several benchmarks are chosen for showing the performance the analysis

  Ex: SUSY analysis, VLQ research, …

Difficulty to be exhaustive in interpretation:

- covering all the parameter space of a given model
- testing all the existing models
- testing all the new models which could be conceived after the analysis

→ We must be able to launch an existing analysis, **tomorrow or in few years,** with a **different signal** benchmark and to compute a limit.

Reinterpretation strategy

How to reinterpret LHC results?

Method 1

- Experimentalists keep and maintain their code internally.
- Phenomenologists ask to the authors to test a new model

→ Need manpower, time consuming for experimentalists

Method 2

- Use a framework which :
  - Captures the analysis code, the data, …
  - Allows people to upload they own MC samples
  - Launch automatically the codes and store results

→ **RECAST** project

Method 3

- Experimentalists provide all useful information to phenomenologists.
- Developing an external code which mimics the analysis results.

→ Approximations but much faster (useful for scan over parameter-space)

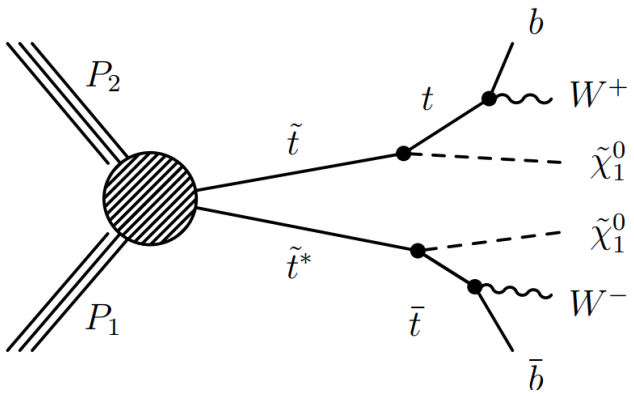
→ Identification of topologies or region not tested by experimentalists

→ Feedback to experimentalists

Type of recasting tools

Based on Simplified Models

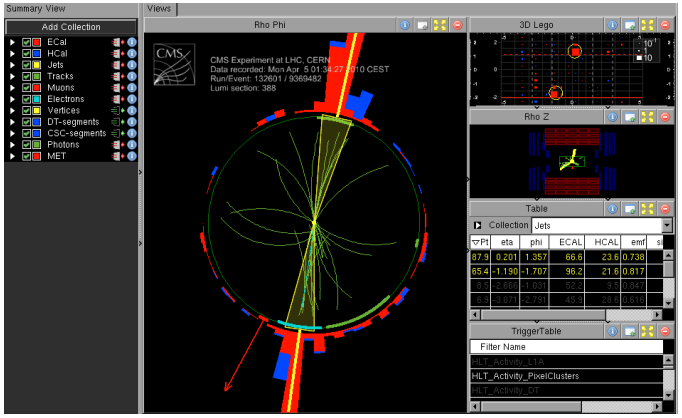
T2tt



- Work with event topology
- Properties are reduced to mass spectrum, xsection and BR

Extremely fast
Moderately accurate / general

Based on Detector simulation



- Mimicking simulation + reconstruction + selection acheived by CMS or ATLAS
- « Very-fast » simulation

Very-Fast
Accurate / general

1. Motivations

2. Recasting based on SMS

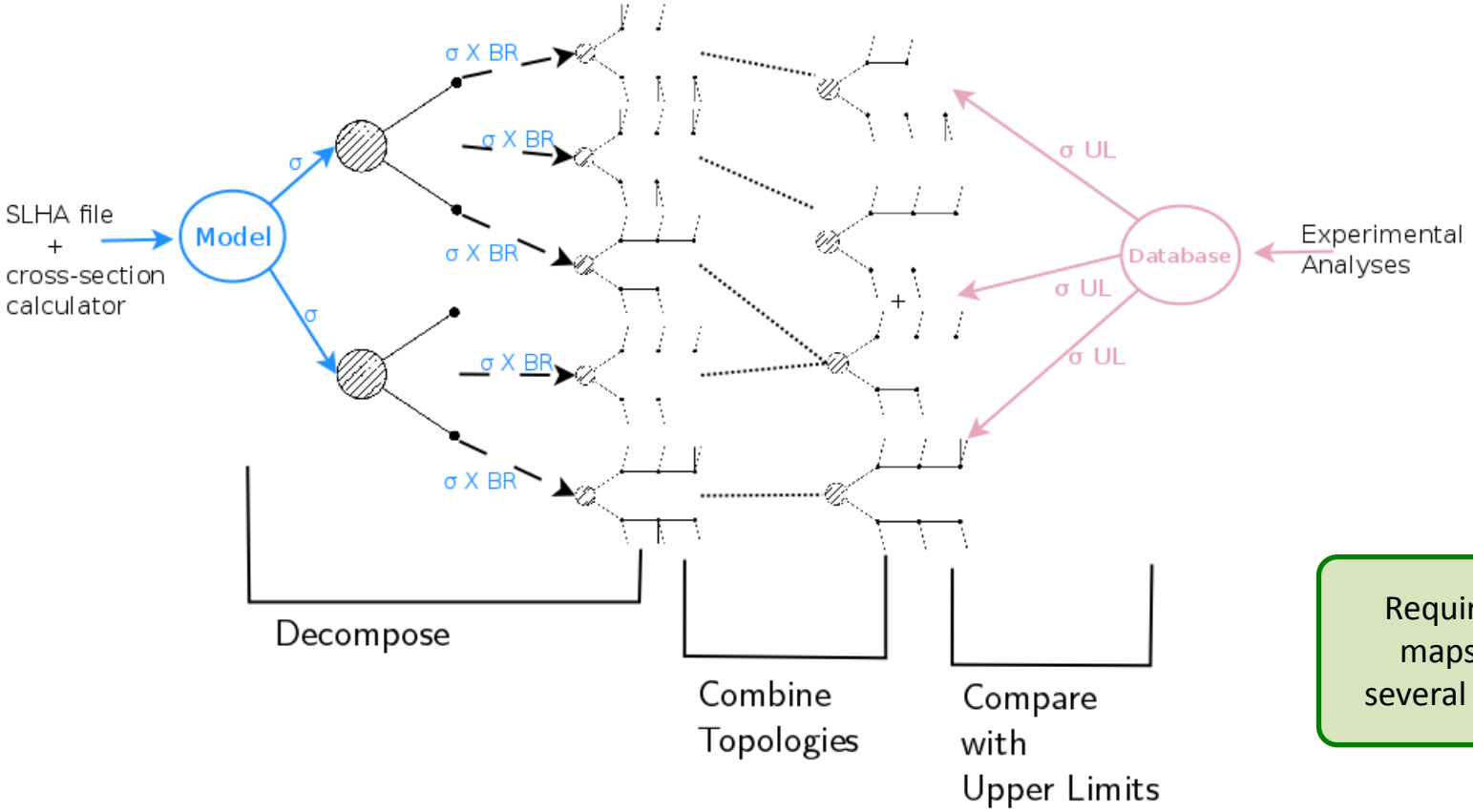
3. Recasting based on detector simulation

4. Summary & opened questions

SModelS & Fastlim: overview

For generic BSM models with a Z2 symmetry

For MSSM (up to now)



Required efficiency maps if there are several signal regions.

SModelS & Fastlim: data base

Fastlim

What are the required experimental inputs?

- Upper limits from CMS/ATLAS paper
- Efficiency maps from simulation package (Fastlim + ATOM)
- Efficiency maps from CMS/ATLAS paper (since SModels 1.1)

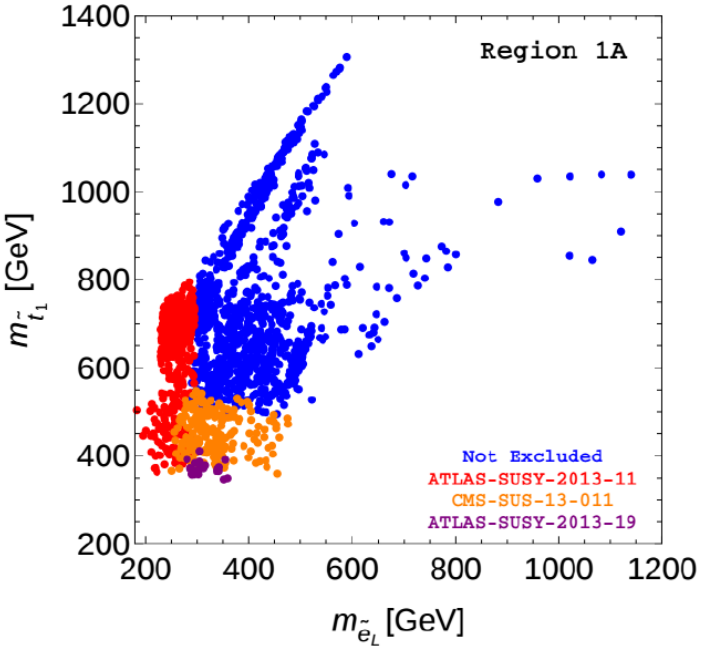
SModels data base (including Fastlim inputs)

Exp	$\sqrt{s}$	Upper limit	Efficiency map
ATLAS	8 TeV	15	18 (of which 9 Fastlim)
	13 TeV	3	2
CMS	8 TeV	15	7
	13 TeV	7	0

SModelS & Fastlim: examples of results

See Wolfgang Waltenberger's talk for details

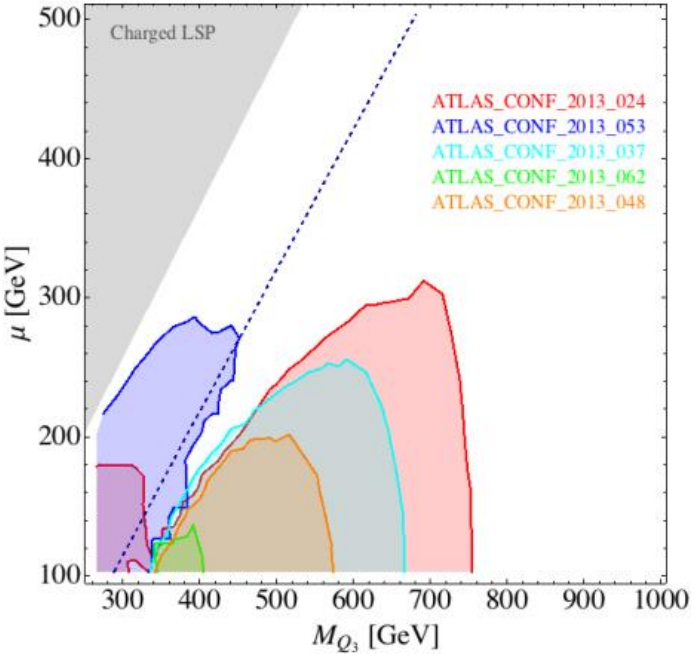
[arXiv:1510.00246v2]



Constraints on NMSSM with 1A benchmarck point

Fastlim

[arXiv:1402.0492v1]

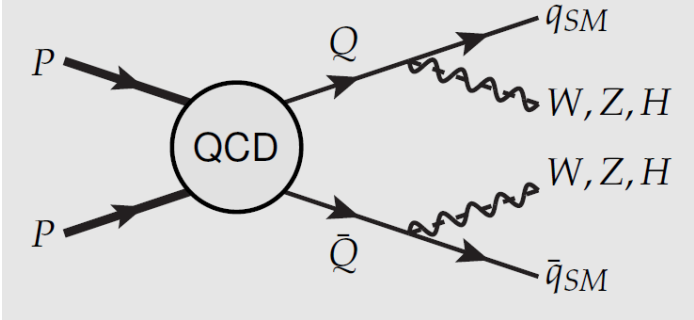


$m_{\tilde{g}} = M_{U_3} = M_{D_3}$ =3 TeV, $\tan\beta$ =10, $X_t = 0$

XQCAT

XQCAT = eXtra Quark Combined Analysis Tool
designed for heavy extra quarks

- **Reinterpreting in terms of pair-production of VLQ in the NWA**

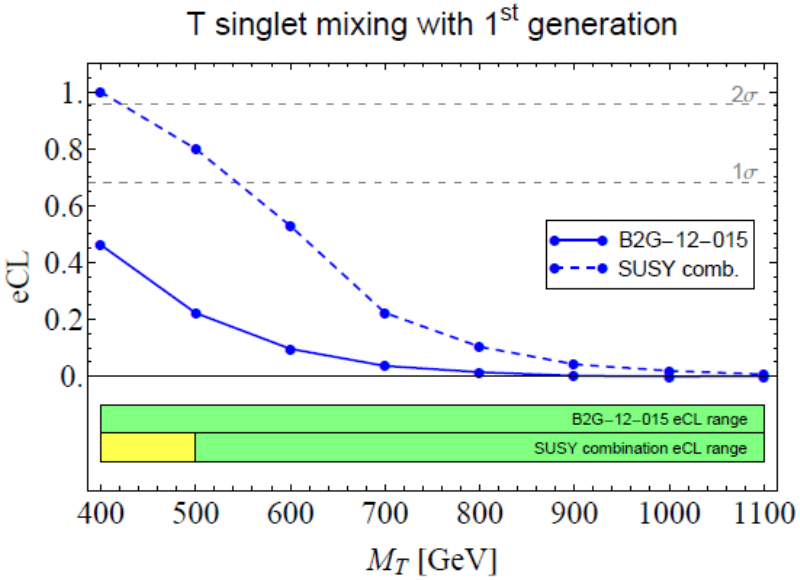


Efficiency maps are extracted from Delphes simulation

- **Example of results:** [arXiv:1304.2185]

Combining :
- VLQ direct research (BG2-12-015)
- SUSY combination @ 7 & 8 TeV
  (α_T, monolepton, SS dilepton, OS dileptons)

$$BR(Zq) = BR(Hq) = 25\% \qquad BR(Wq) = 50\%$$



1. Motivations

2. Recasting based on SMS

3. Recasting based on detector simulation

4. Summary & opened questions

Detector simulation

detector simulation

Generic & global simulation

Full simulation of a detector with:

- Algorithms: calorimetry, energy-flow, …
- Parametrizations: smearing & efficiency
- Full pile-up simulation

Analysis-specific and light simulation

- Simulating only the main detector effects useful for the analysis
- Based essentially on smearing & efficiency ~ transfer functions

- **CheckMate**
- **MadAnalysis 5**

- BuckFast from **Gambit**
- Detector effects in **Rivet** and **ATOM**

Required experimental inputs for recasting

S. KRAMEL et al, *Les Houches Recommendations for the Presentation of LHC Results*, [arXiv:1203.2489v2]

- **Clear description of the selection in the paper:**
  - Definition of the reconstructed objects
  - Each step of the selection
  - Source code of specific and sophisticated of observables (SUSY transverse observable)

- **Clear description of the results:**
  - Crucial numbers
  - Final likelihood expression

- **Detector modeling:**
  - Resolution & efficiency plots for reconstruction of exotic objects, trigger?
  - Efficiency maps

- **Materials for validation:**
  - Cut-flow chart
  - Description of the signal benchmarks and its generation (the best is to have the LHE files)
  - Plots of key observables

CheckMate & MadAnalysis5: a brief overview

(**Check** **M**odels **A**t **T**erascale **E**nergies)

Designed for recasting:

- Choose the objects of interest
- Filter objects
- Check event vetoes
- Check various signal region criteria
- Count number of input events that fall into each signal region

Tune of Delphes:

- Improvement of ATLAS simulation
- Add isolation flags
- Add object definition flags

DELPHES
fast simulation

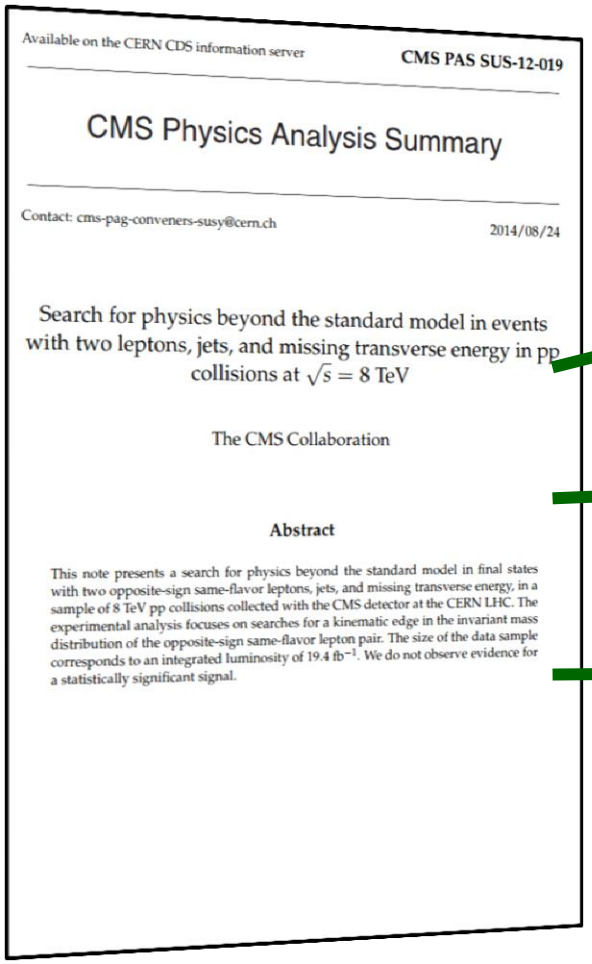
Multipurpose tool:

- Monte-Carlo validation
- Phenomenological analysis with 2 levels:
  - Intuitive metalanguage
  - C++ development
- Recasting with the Physics Data Base

Tune of Delphes:

- Isolation defined at the analysis level
- Produces compact and generic ROOT files
  → avoids as much as possible from launching Delphes

Strategy of CheckMate & MadAnalysis 5



Available on the CERN CDS information server CMS PAS SUS-12-019

CMS Physics Analysis Summary

Contact: cms-pag-conveners-susy@cern.ch 2014/08/24

Search for physics beyond the standard model in events with two leptons, jets, and missing transverse energy in pp collisions at $\sqrt{s} = 8$ TeV

The CMS Collaboration

Abstract

This note presents a search for physics beyond the standard model in final states with two opposite-sign same-flavor leptons, jets, and missing transverse energy, in a sample of 8 TeV pp collisions collected with the CMS detector at the CERN LHC. The experimental analysis focuses on searches for a kinematic edge in the invariant mass distribution of the opposite-sign same-flavor lepton pair. The size of the data sample corresponds to an integrated luminosity of 19.4 fb^{-1}. We do not observe evidence for a statistically significant signal.

Is my model excluded or not?

Signal events

Reconstruction info

DETECTOR VERY-FAST-SIMULATION

Selection / validation material

Recast selection

Numbers of data and background events

Limit computation

CLs

Limits

List of (public) recast analyses

Exp	$\sqrt{s}$	Already validated	Not yet validated
ATLAS	7 TeV	0	1
	8 TeV	21	10
	13 TeV	12	2
	HL 14 TeV	6	0
CMS	7 TeV	0	1
	8 TeV	6	4
	13 TeV	0	1
	HL 14 TeV	0	0

Exp	$\sqrt{s}$	Already validated	Not yet validated
ATLAS	8 TeV	8	0
	13 TeV	2	0
CMS	8 TeV	9	1
	13 TeV	0	0

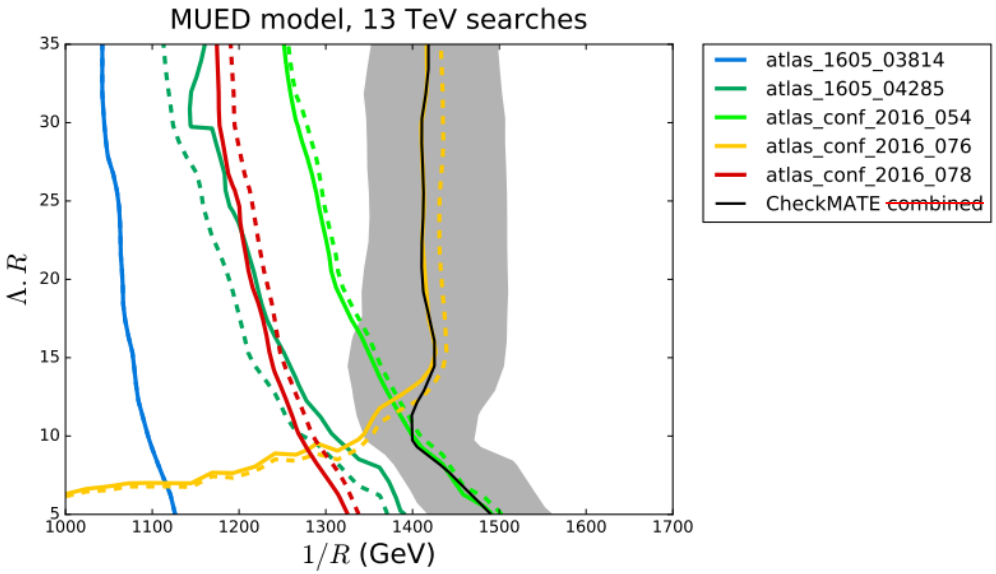
- All the recast analyses are validated and a validation note is released.
- Analyses covered by CheckMate and MadAnalysis 5 are not the same.

List of (public) recast analyses

CHECKMATE

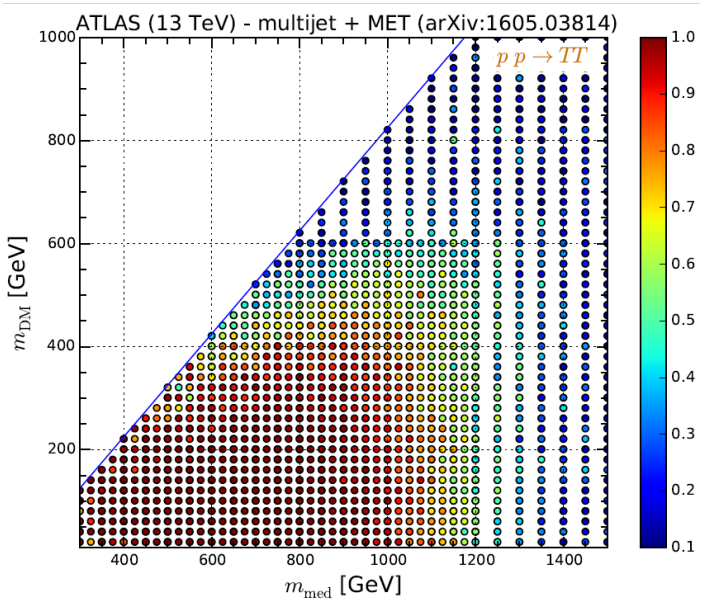
MADAnalysis 5

[arXiv:1702.00410]



Constraining MUED from existing SUSY searches

[coming soon]



Constraining pair of VLQ from SUSY searches

Tools based on machine learning algorithm

- Getting LHC constraints could be time consuming in particular for a scan over SUSY-like parameter space: (one point with CheckMATE / MadAnalysis 5 ≈ 10 min).
- Machine learning algorithm could be used to scan in a clever way:
  - training an ML algorithm with plethora of plethora of points
  - then making a prediction for a new point in few milliseconds

SUSY-AI

- Scan over SUSY parameter-space
- ML algorithm = **random forest**
- Algorithm tested with pMSSM-19 and trained with ~400,000 points

ScyNet

=

Susy Calculating Yields NET

- Scan over SUSY parameter-space
- ML algorithm = **neural network**
- Algorithm validated with pMSSM-11 and trained with ~200,000 points

Still a private tool

Rivet for BSM

RIVET = Robust Independent Validation of Experiment and Theory

Since the release 2.5, Rivet deals with the BSM analyses.

- Use efficiencies and smearing for modeling the detector response
- List of implemented analyses (BSM only):

Exp	$\sqrt{s}$	Already validated	Preliminary	Not yet validated
ATLAS	7 TeV	8	1	6
	8 TeV	0	1	1
	13 TeV	2	0	4
CMS	7 TeV	0	0	0
	8 TeV	0	0	1
	13 TeV	1	0	0

Other tools based on Rivet

CONTUR

CONTUR = Constraints on new theories using Rivet

- Rivet was initially designed for SM (generator validation)
- Unfolded fiducial cross section measurements to constrain new physics

→ Tool chain: FeynRules →Herwig 7 → RIVET

Atom

ATOM = Automated Testing Of Models

- Forked from Rivet "for SM"
- Detector simulation : smearing/efficiency
- Associated to FastLim for limits
- Under development

Still a private tool

Gambit

GAMBIT = Global and Modular BSM Inference Tool

Global statistical fit from different sources of data (frequentists & bayesian): low physics, astrophysics, colliders, …

Modular architecture
→ Module devoted to LHC analyses: **ColliderBit**

- **Generation of events:**
  - ColliderBit works with parallelized Pythia
  - Detector effects: **BuckFast** (efficiencies/smearing) or **Delphes**
  - Multithread (OpenMP) available with BuckFast

See Pat Scott's talk for details

- **Experimental inputs:**
  - LHC analyses (SUSY)
  - Constraint from LEP searches on SUSY particles.
  - Likelihoods from experimental searches for Higgs bosons via HiggsBound & HiggsSignal

Exp	$\sqrt{s}$	# analyses
ATLAS	8 TeV	7
	13 TeV	1
CMS	8 TeV	4
	13 TeV	1

1. Motivations

2. Recasting based on SMS

3. Recasting based on detector simulation

4. Summary & opened questions

Summary

- **Importance of recasting & reinterpretation for:**
  - Identifying holes in the ATLAS/CMS research program
  - Preserving LHC analyses

- **Different kinds of tools:**
  - Constraints from SM measurements: **Contur**
  - Recasting based on simplified models: **SModelS**, **Fastlim**, **XQCAT**
  - Recasting based on detector simulation:
    - With Delphes: **CheckMate**, **MadAnalysis5**
    - With smearing/efficiency: **Rivet**, **ATOM**, **ColliderBit** from **Gambit**
  - And scanners (**SUSY-AI**, **ScyNet**, **ScannerBit**), fitters (**Gambit**, ….)

- **Tools are still in development:** new recast analyses, new functionalities are coming soon

- **Join the effort?**

  https://twiki.cern.ch/twiki/bin/view/LHCPhysics/InterpretingLHCresults

  Announcement: new workshop 16-17 October 2017 at Fermilab
  https://indico.cern.ch/event/639314/

List of discussed tools & projects

Name	Public / Private	Home site	Reference
Aeacus & RHADAManTHUS	Public	http://joelwalker.net/code/	
ATOM	Private	http://fastlim.web.cern.ch/fastlim/	See Fastlim papers currently
CheckMate	Public	https://checkmate.hepforge.org/	arXiv:1312.2591, arXiv:1611.09856
Contur	Public	https://contur.hepforge.org/	arXiv:1606.05296
Fastlim	Public	http://fastlim.web.cern.ch/fastlim/	arXiv:1402.40492 , EPJC74 (2014) 11
Gambit	Public	https://gambit.hepforge.org/	arXiv:1705.07908, arXiv:1705.07919, arXiv:1705.07920, arXiv:1705.07933, arXiv:1705.07959, arXiv:1705.07936
LHADA	Public	https://indico.cern.ch/event/572170/	arXiv:1605.02684, section 16 & 17

List of discussed tools & projects

Name	Public / Private	Home site	Reference
MadAnalysis5	Public	https://launchpad.net/madanalysis5	arXiv:1206.1599, arXiv:1405.3982, arXiv:1407.3278
Recast	Public	http://recast.perimeterinstitute.ca/ https://github.com/recast-hep	arXiv:1010.2506
Rivet	Public	http://rivet.hepforge.org/	arXiv:1003.0694
ScyNET	Private		arXiv:1703.01309
SModelS	Public	http://smodels.hephy.at/wiki	arXiv:1701.06586, arXiv:1312.4175
Susy-AI	Public	http://susyai.hepforge.org/	arXiv: 1605.02797
XQCAT	Public	https://launchpad.net/xqcat	JHEP 1412 (2014) 080,arXiv:1405.0737, arXiv:1409:3116

Some opened questions

- **Can we design a universal language for describing an analysis?**
  - Some metalanguages have been developed:
    - MadAnalysis5: intuitive but too much simple
    - Aeacus & RHADAManTHUS: advanced metalanguage but not enough
  - Framework-independent language:

    Towards an analysis description accord for the LHC [arXiv:1605.02684]

    One proprosal: LHADA (Les Houches Analysis Description Accord for the LHC) in development

- **Can we recast multi-variate analyses?**
  - No: experimentalists must provide also cut-and-count selections
  - Yes: experimentalists must publish all material for training a ML algorithm

- **May we combine recast analyses?**
  - To be correct: required correlation matrices

Back-up slides

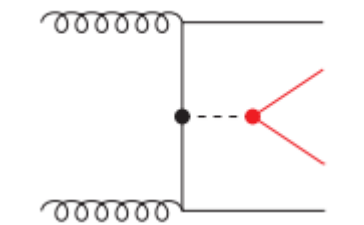
CheckMate & MadAnalysis 5 performance

Extract from: C. Arina et al, *A comprehensive approach to dark matter studies: exploration of simplified top-philic models*, JHEP04(2015)029, arXiv:1605.09242v1

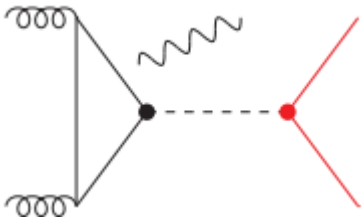
Simplified top-philic dark matter model
→ Fermionic dark matter candidate X
→ Scalar mediator Y_0

$$\mathcal{L}_{t,X}^{Y_0} = -\left(g_t \frac{y_t}{\sqrt{2}} \bar{t}t + g_X \bar{X} X\right) Y_0$$

2 relevant topologies with large MET signature at collider experiments



tt + MET



X + MET
avec X=Z,j,H

4 recast analyses

- CMS-B2G-14-004: tt + MET
- CMS-EXO-12-048: monojet
- CMS-EXO-12-054: mono-Z
- ATLAS-EXOT-2014-20 mono-Higgs

MADAnalysis 5

Home made recasting