CMS experiences with MC for BSM searches

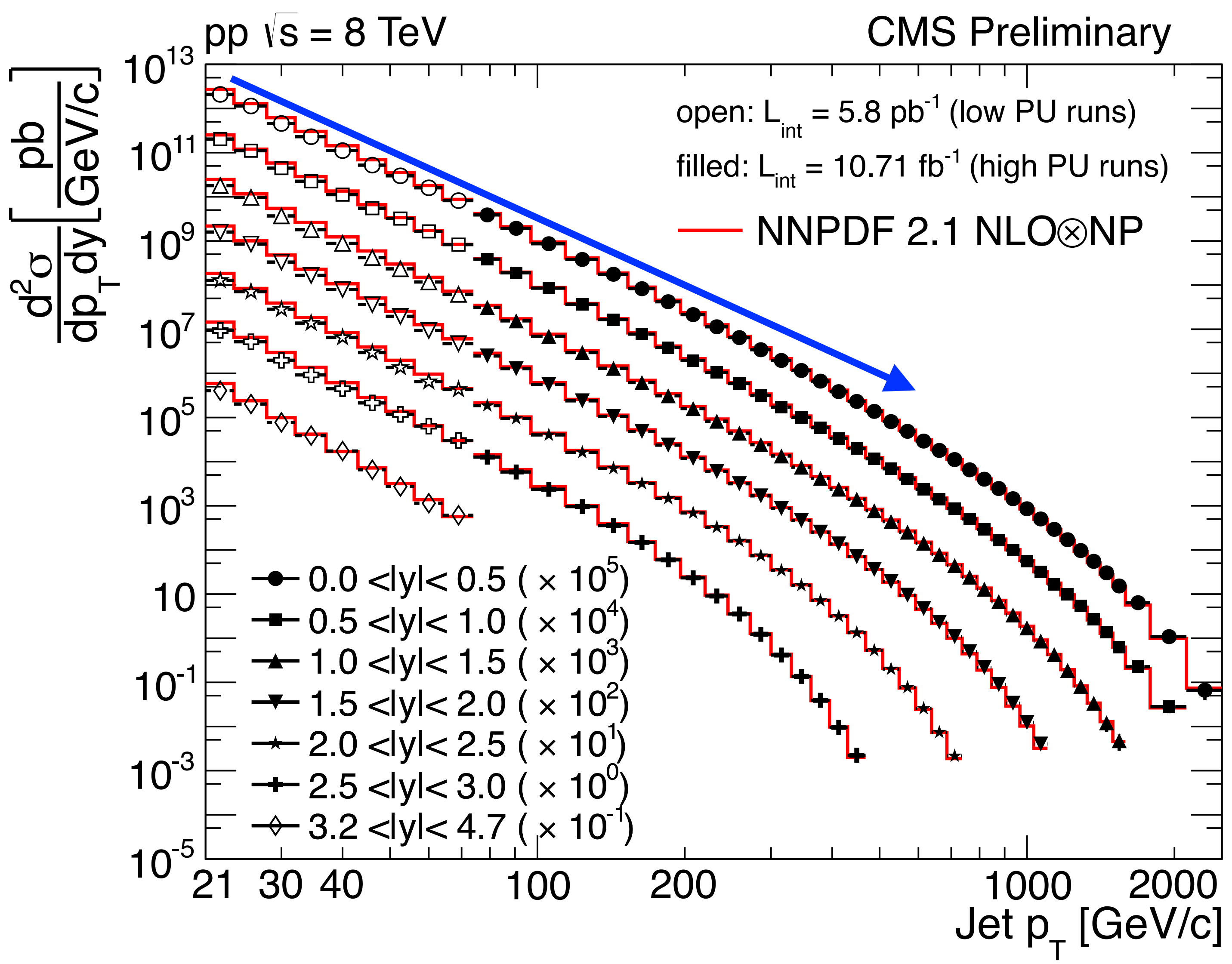
Frank Golf (UCSB)

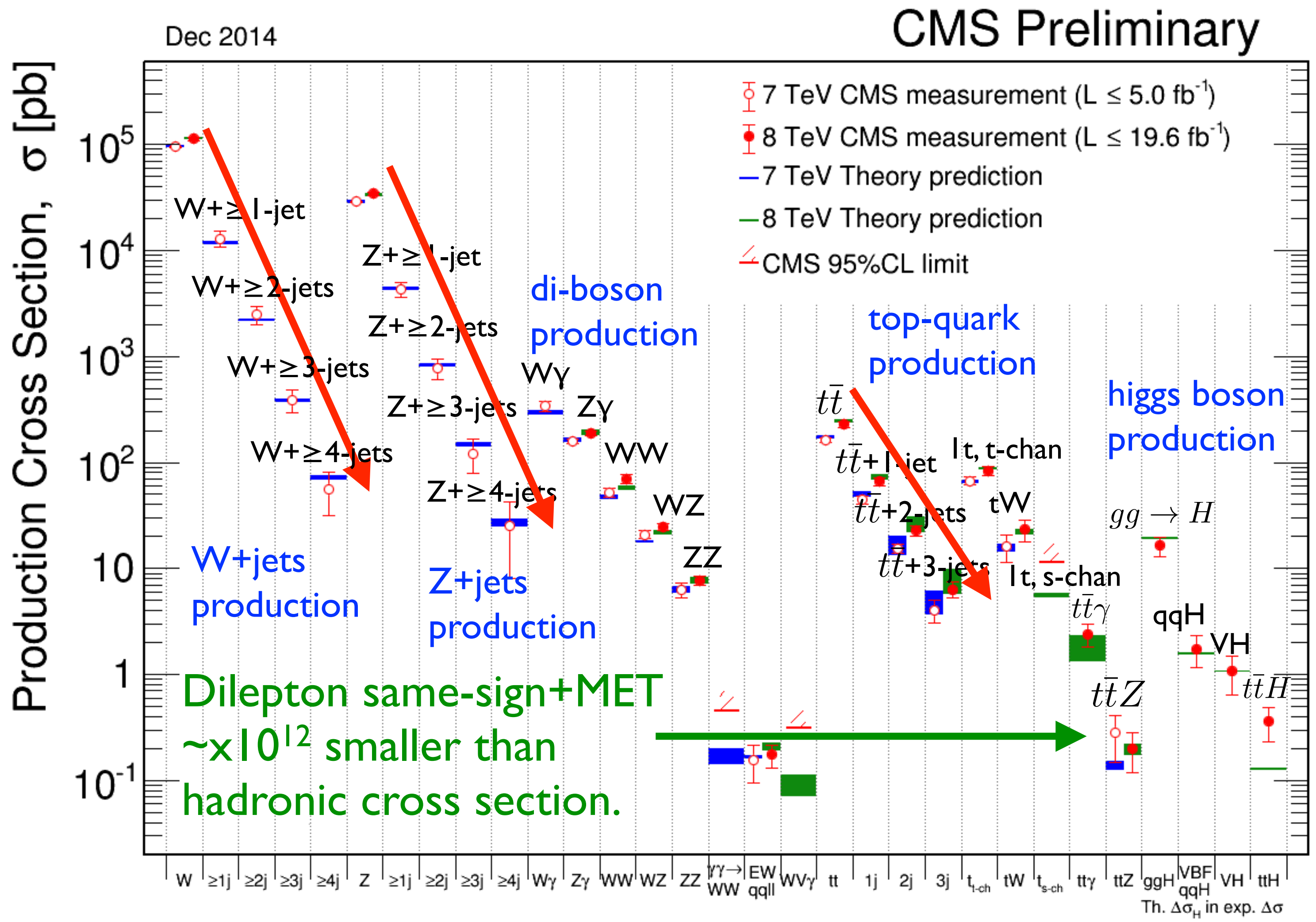
on behalf of the CMS collaboration

Since Korea 1 year ago…

- New results based on Run 1 data have continued to appear, albeit at a slowing pace as much effort over the past year has turned to preparations for 13 TeV data.
- Having discovered a boson with properties consistent with those of a SM Higgs, the focus will probably be on BSM more than ever.
- Run 2 brings higher energy collisions → the approach to NP searches will likely bare many similarities to the beginning of Run 1, returning first to generic signatures that become progressively more targeted over time.
- However, one important difference is that we now begin with better understood tools for generating and simulating SM and BSM processes.
- Today: highlight some past and current experiences with MC from the perspective of an experimentalist working on searches.

Success of MC in Run 1

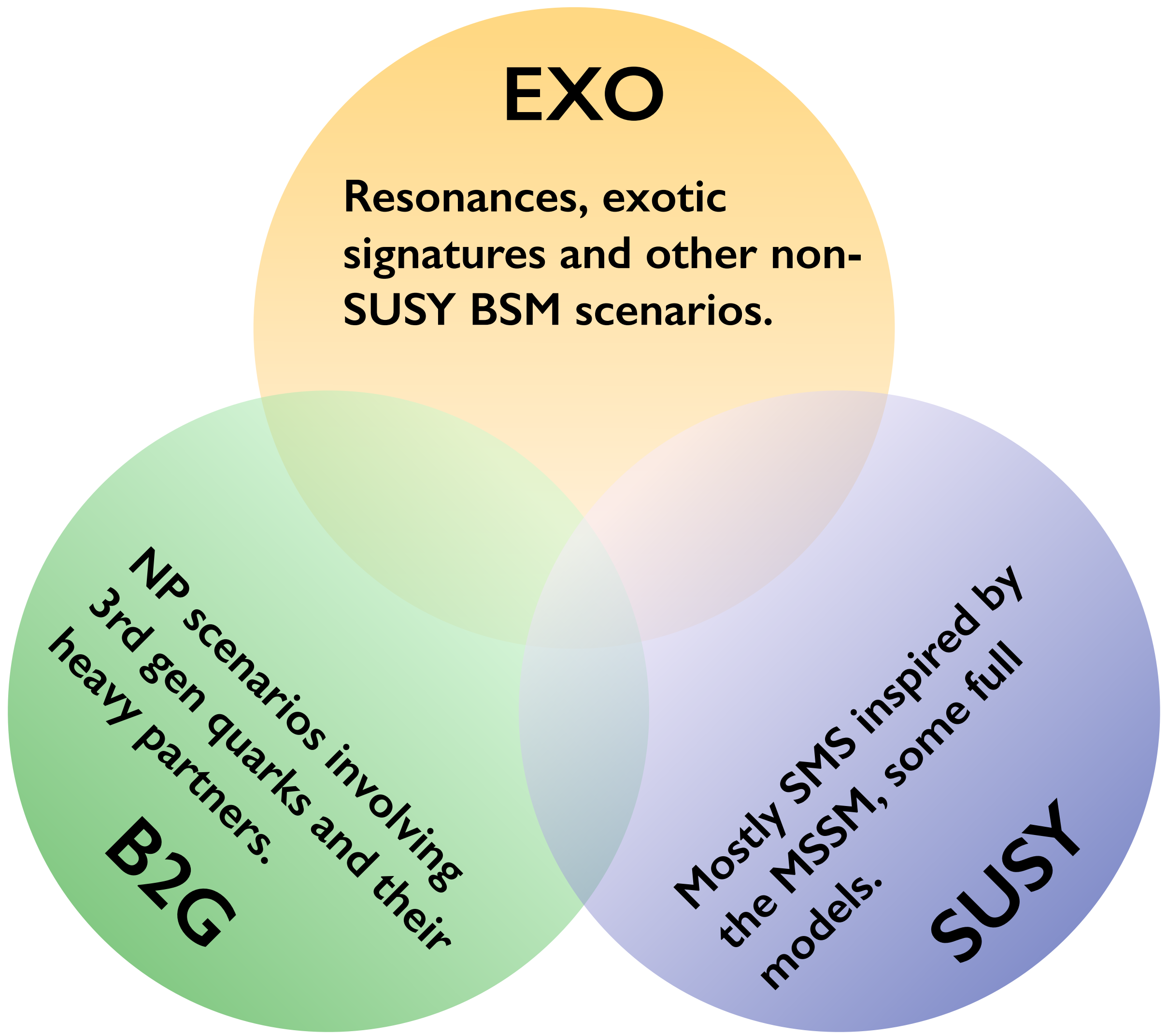




Excellent agreement between theory and experiment.

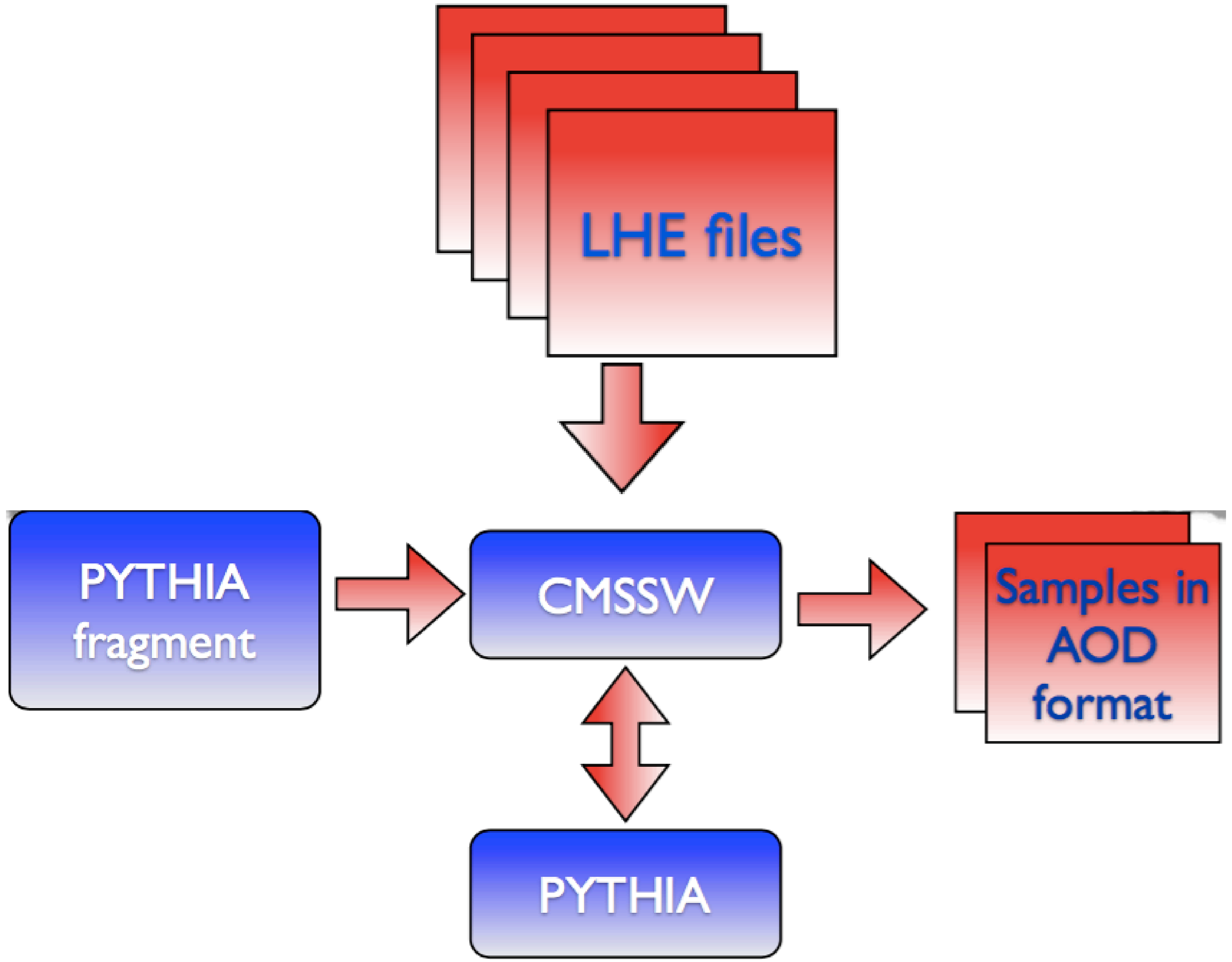
CMS Program of BSM Searches

- BSM searches driven by experimental signature on the one hand, by theory on the other.
- Frequently, multiple theory models give rise to similar experimental signatures.
  - e.g. dark matter, SUSY → jet(s)+MET final states
- Even from a theory perspective, many scenarios equivalent: same effective operators, different coefficients.
- **Aim for systematic approach that covers full spectrum of experimental signatures.**



MC Generation in CMS

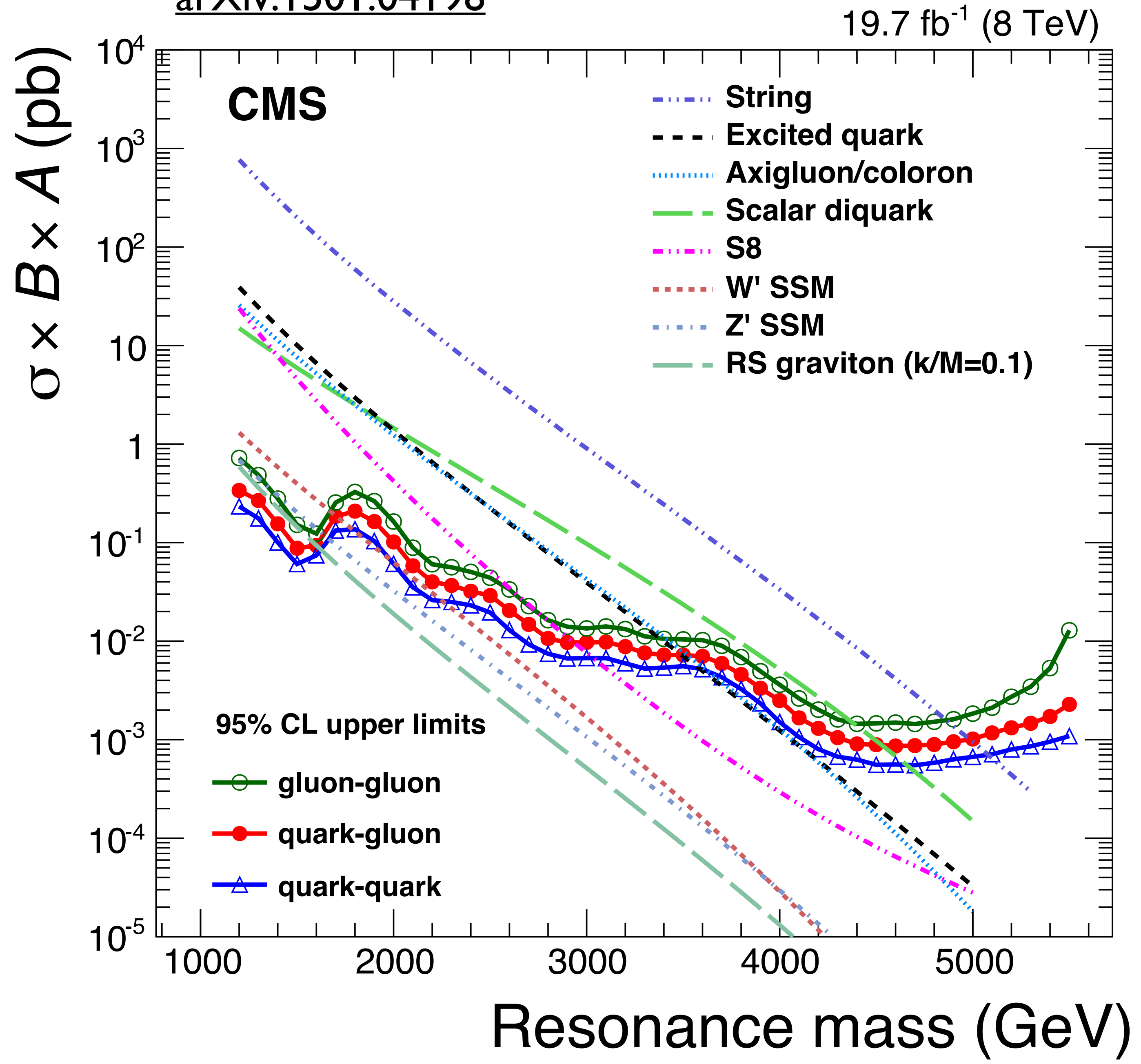
- By design, CMSSW supports Pythia in a trivial way, allowing users to pass a fragment through a python module.
  - Whenever possible, this is the easiest choice.
- CMS also fully supports the use of LHE files, allowing the use of any generator that can write output in the LHE format.



BSM MC in CMS

EXO-12-059
arXiv:1501.04198

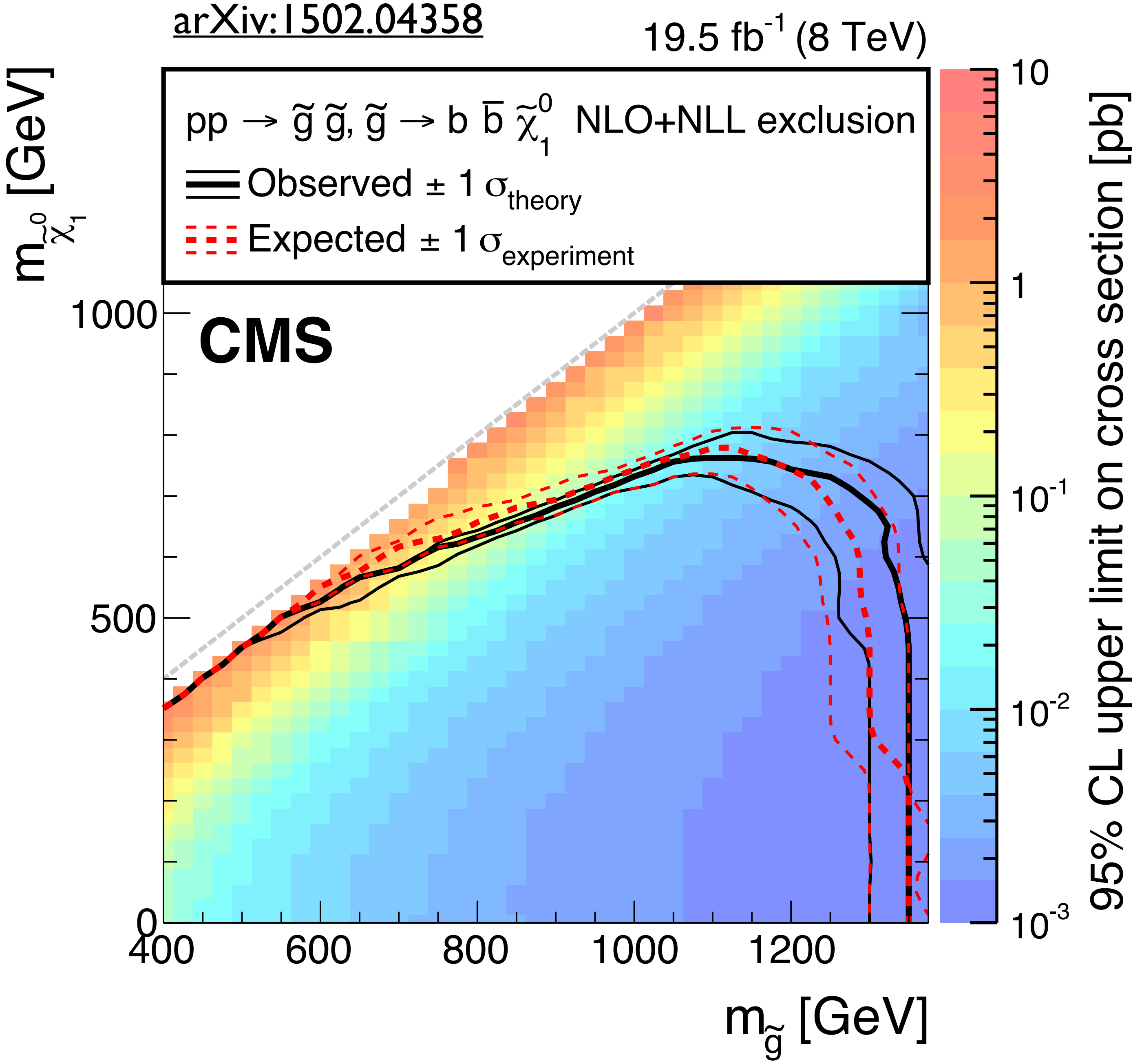
- Generating BSM MC is then just like generating any other process.
- For interpretations, it is usually sufficient to generate a 1D scan containing a small number of points and interpolate.



BSM MC in CMS

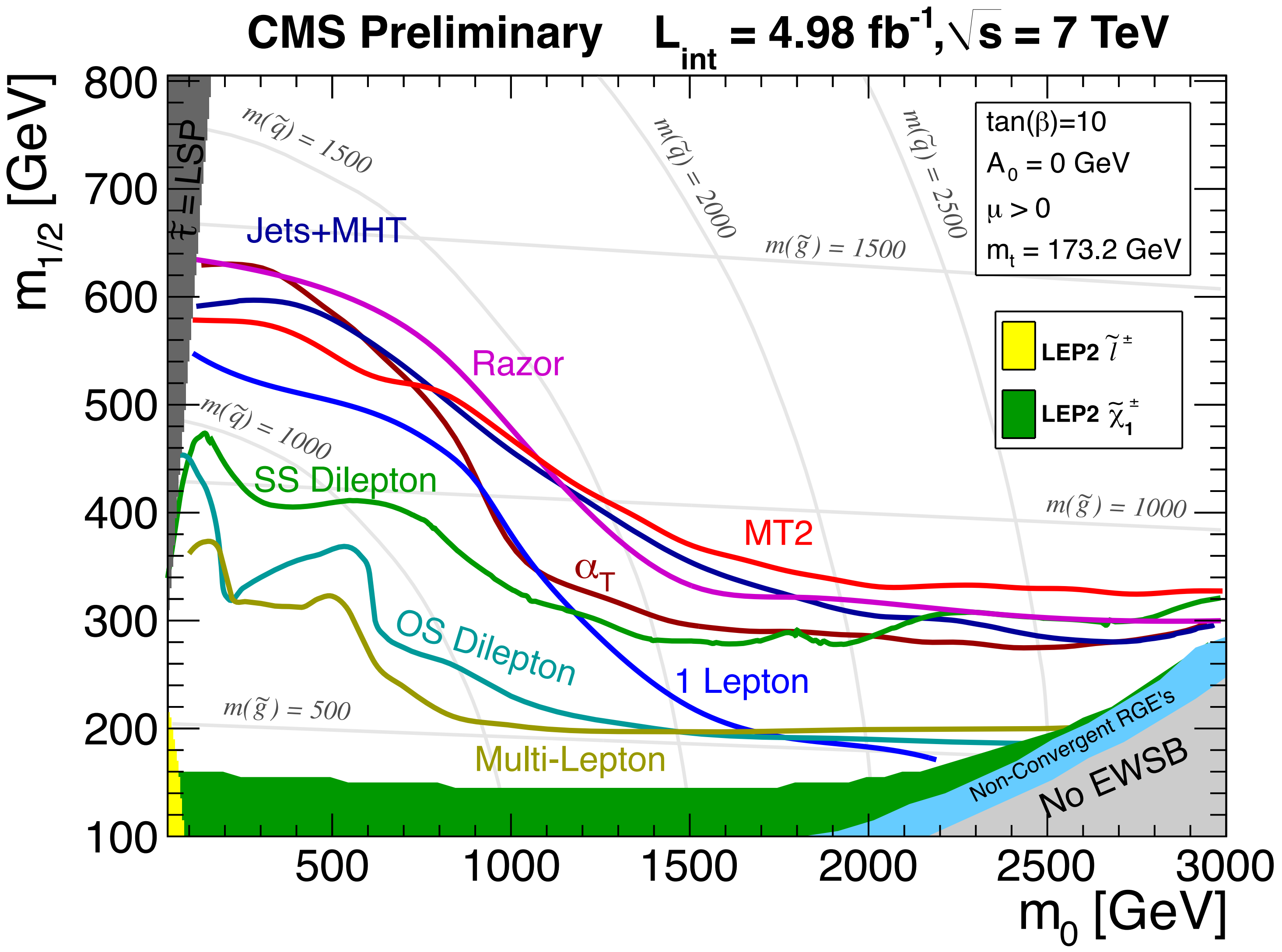
- The main exception in Run 1 was for the interpretation of SUSY results, the large number of free parameters requiring the generation of (at least) 2D mass scans with fine granularity.

SUS-13-019
arXiv:1502.04358



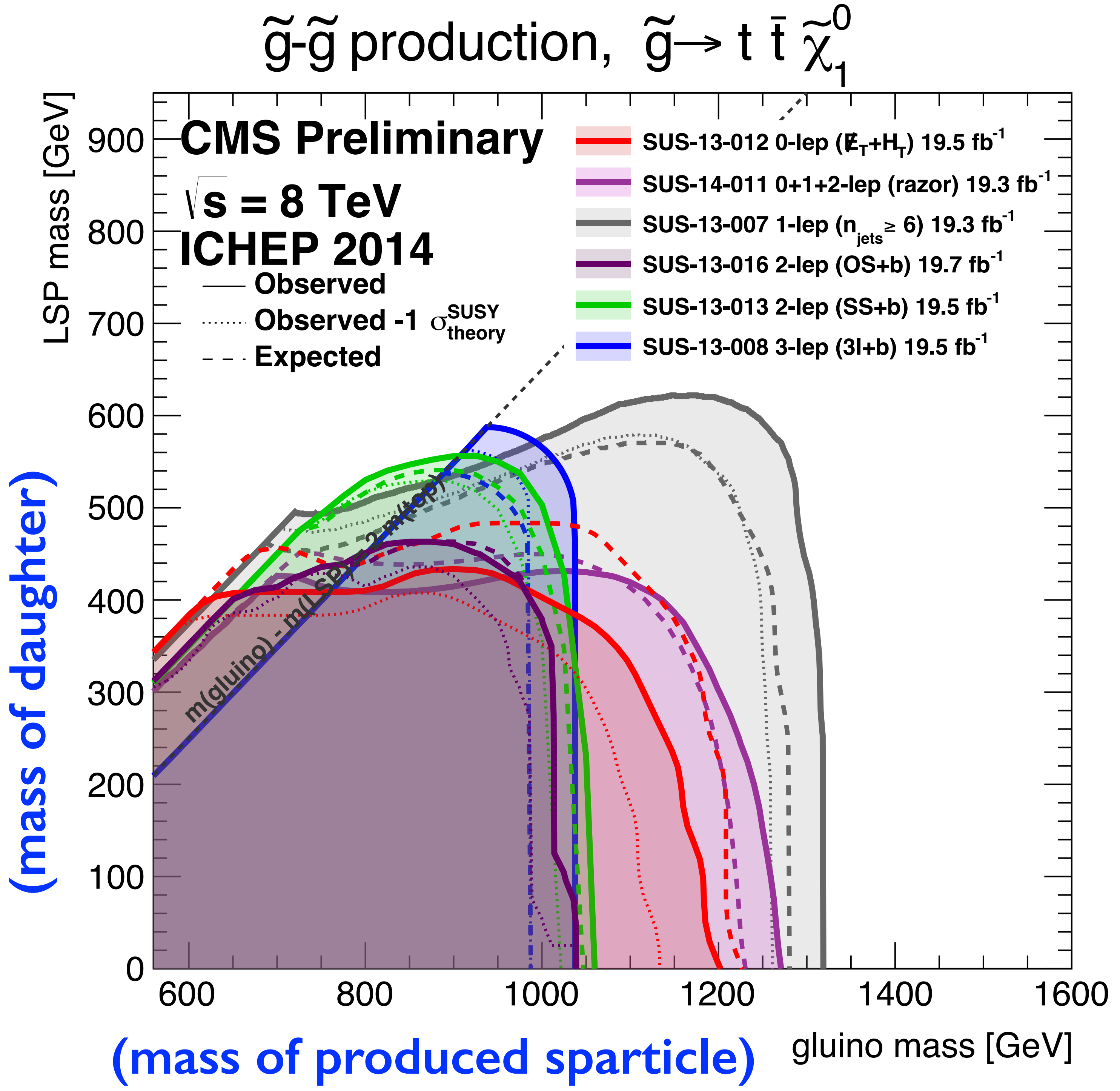
Interpreting SUSY Results: Historical Perspective

- At start of Run 1, searches mainly interpreted their results in the context of the cMSSM model.
  - One scan, used by all searches, generated with Pythia6.
  - Technically simple, required limited resources.



Broader Interpretation of Results with Simplified Models

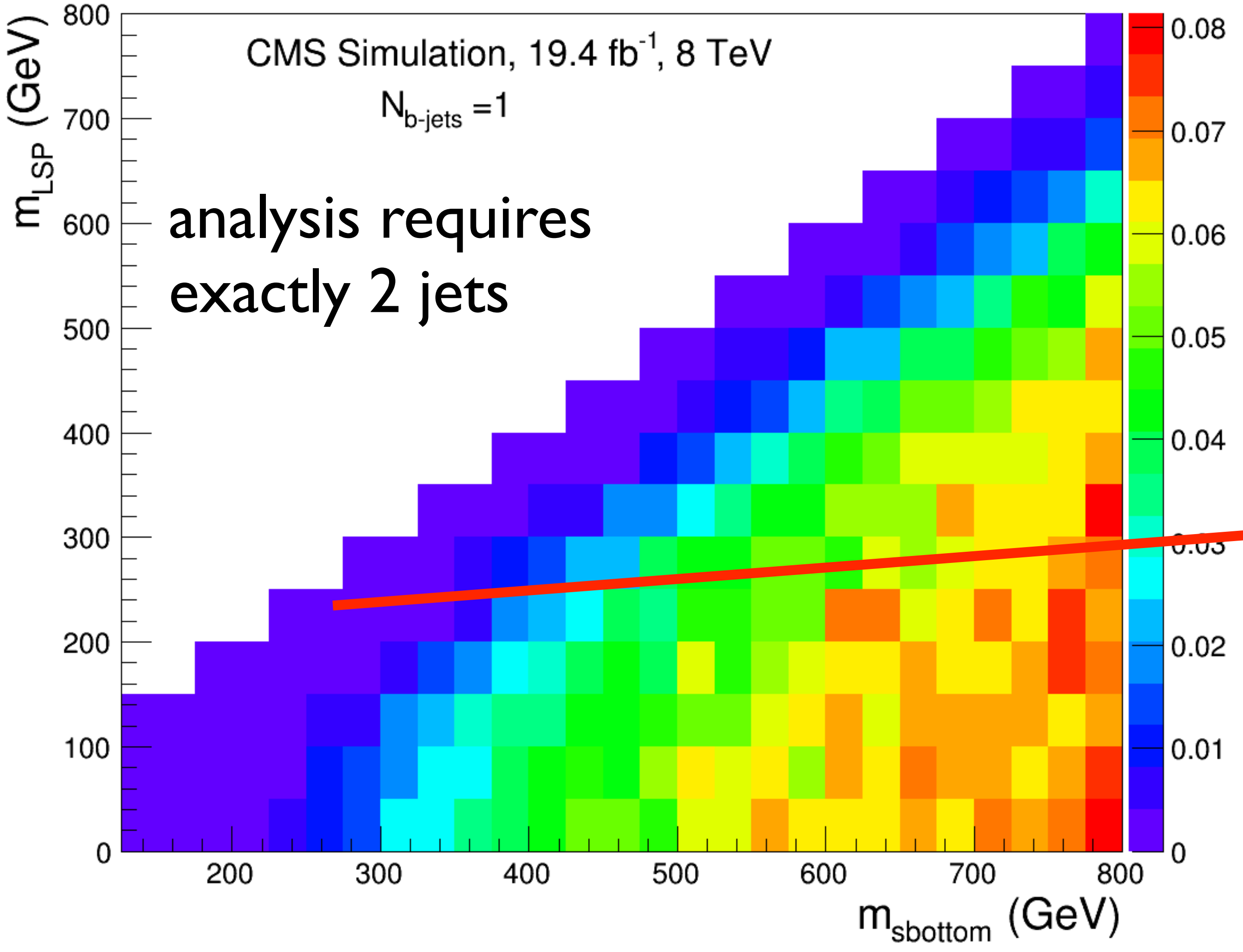
- Over time, simplified models (SMS) gained more interest and by now are the primary means by which we interpret the results of searches for SUSY.
- Large number of topologies considered.
- Each topology can give rise to a variety of experimental signatures → same topology used to interpret the results of multiple searches.
- At the same time, the results of each search are typically interpreted with a variety of topologies.
- Many topologies, fine granularity, (at least) 2D mass scan → significant resources required.



Sparticles and their decay not the full story

The efficiency to trigger and select events decreases as the diagonal is approached. These events only enter our searches thanks to associated jet production — i.e. ISR boost.

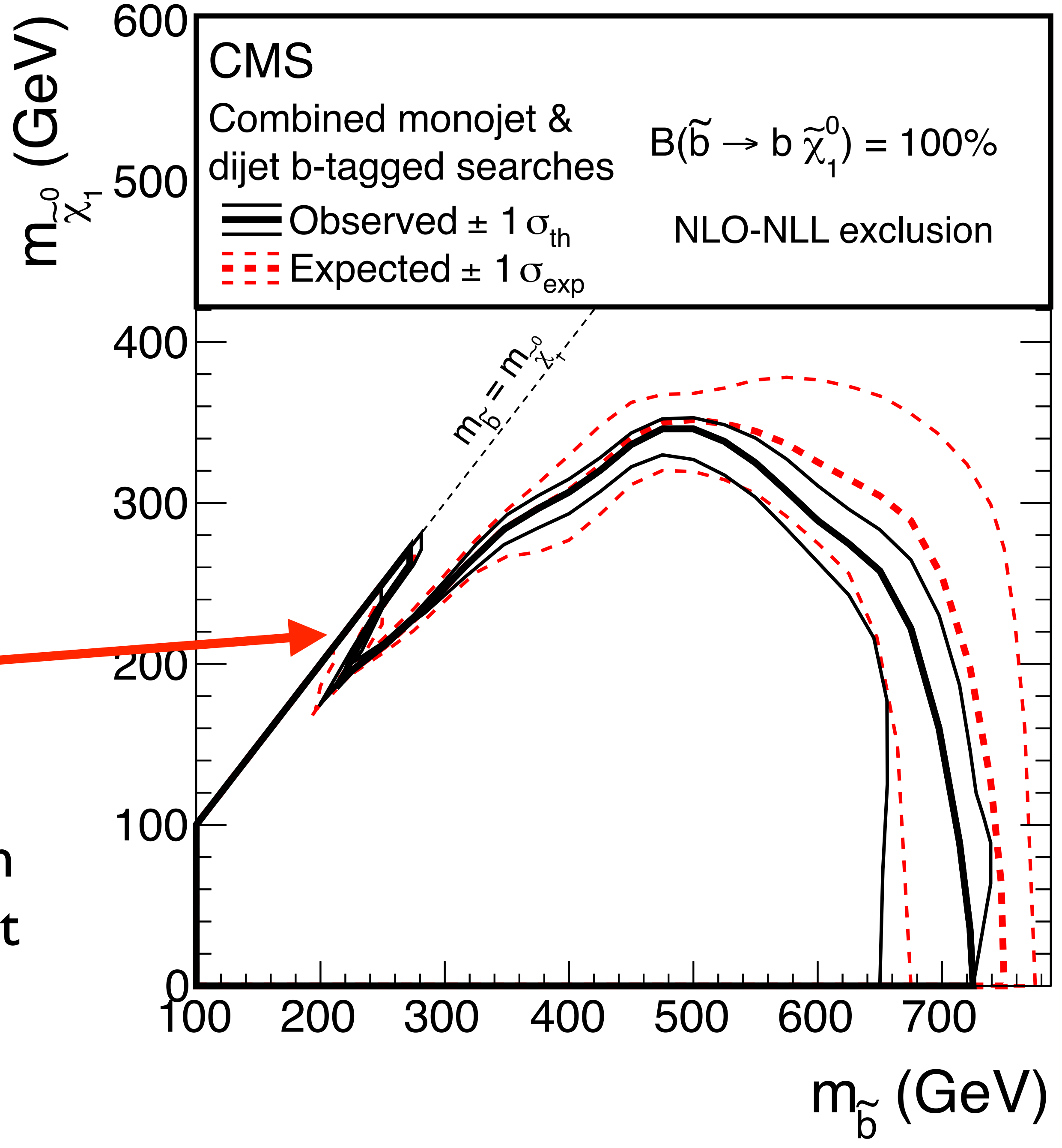
SUS-13-018 $pp \to \tilde{b}\tilde{b}^*,\ \tilde{b} \to b\tilde{\chi}_1^0$



adding monojet search that makes use of the ISR jet enhances sensitivity near the diagonal

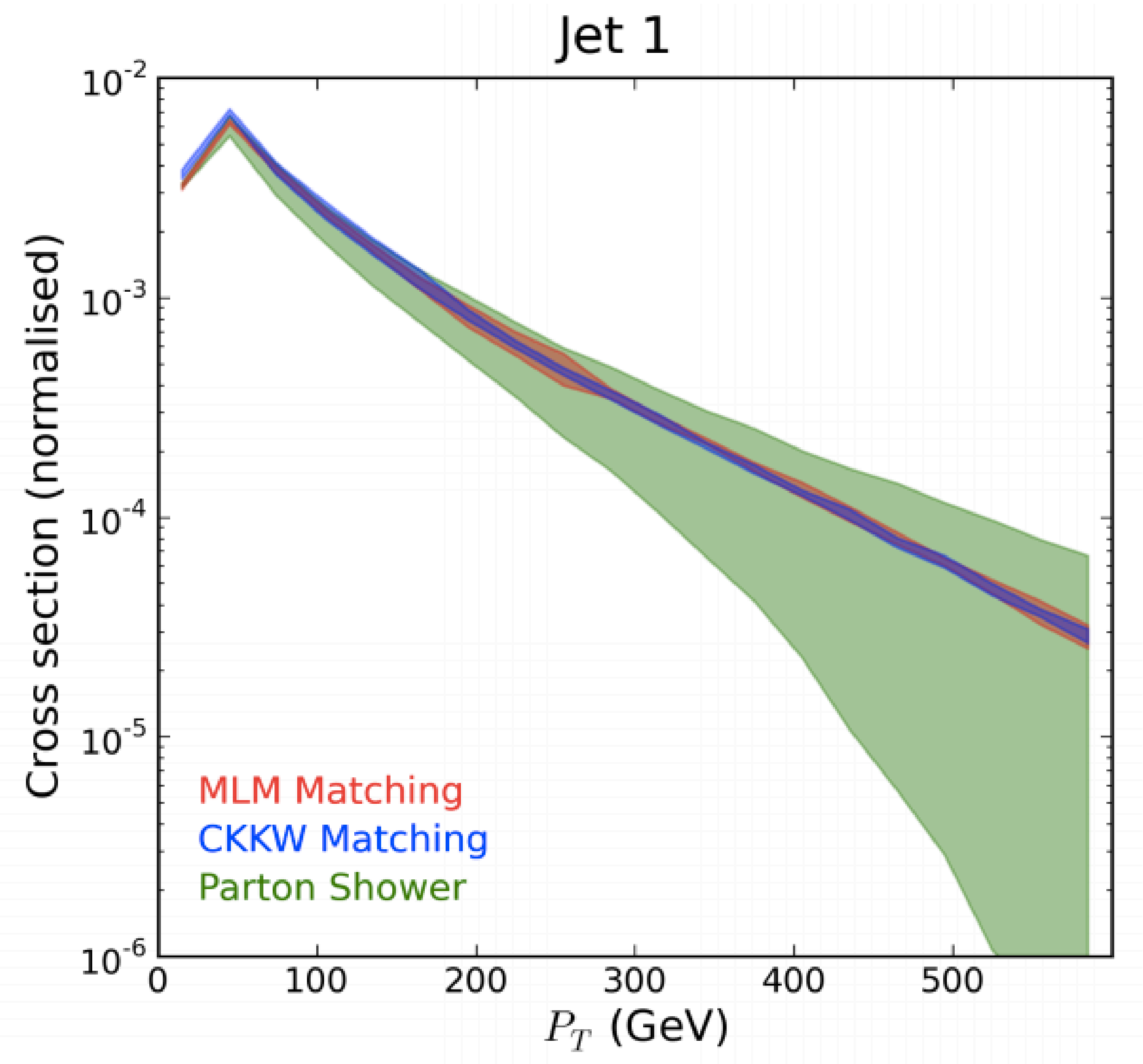
similarly, rely on the modeling of jet production in searches with an explicit jet veto (e.g. ewkino)

SUS-14-001
arXiv:1503.08037



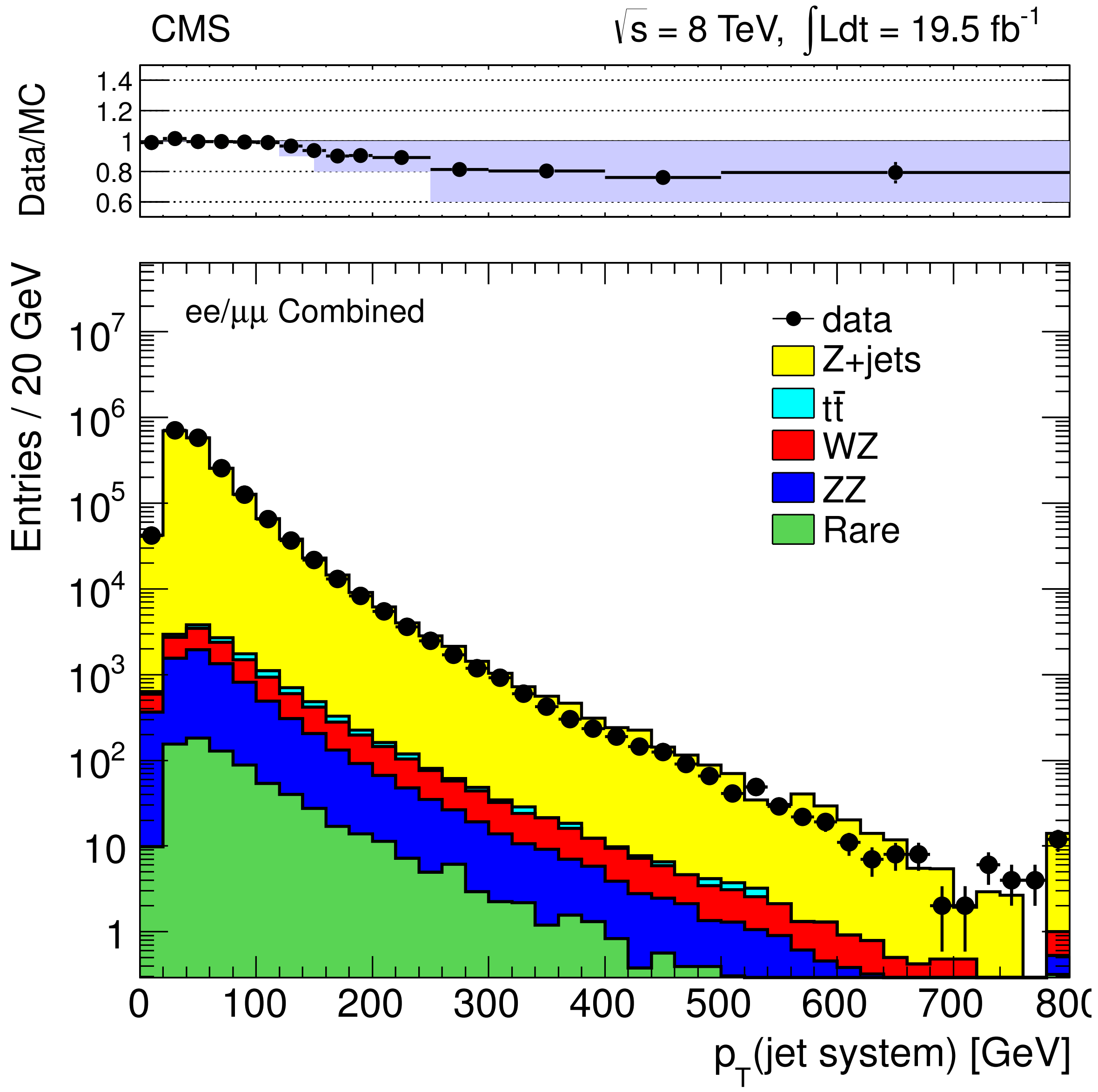
Modeling Jet Production

Going to ME+PS was a first step in this direction.



http://arxiv.org/pdf/1207.1613.pdf

But still find that simulation doesn't represent the data well at high p_T (where we want to work).



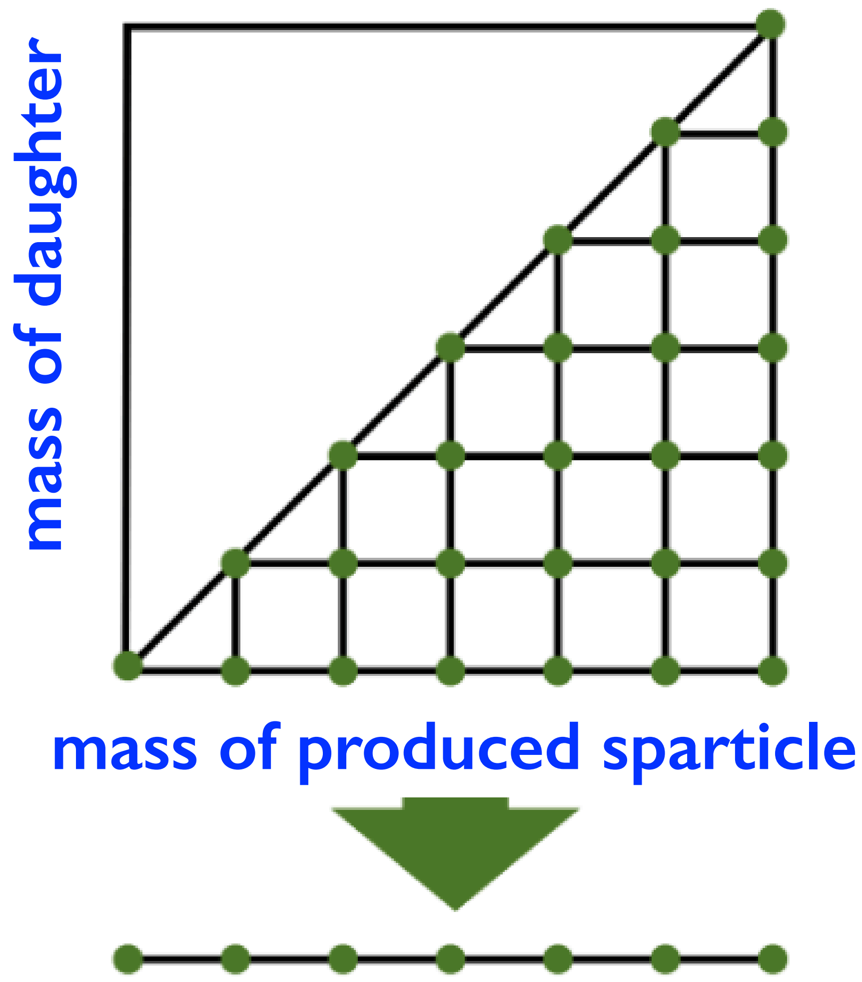
SUS-13-011
arXiv:1308.1586

Going to NLO this may improve, but we cannot (yet) generate signal MC at NLO….

Production of undecayed particles with MG

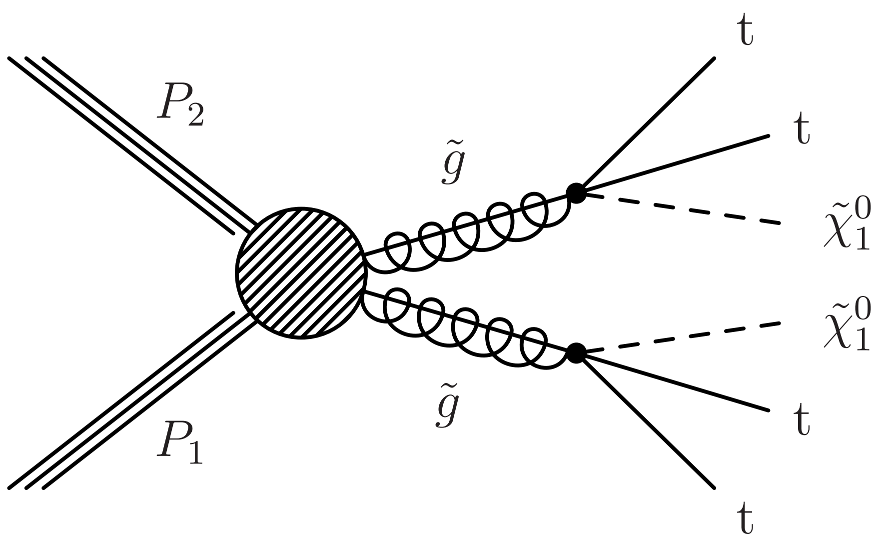
```
import model mssm
define p u u~ d d~ s s~ c c~ b b~ g
define j p
generate p p > go go  @1
add process p p >  go go j  @2
add process p p > go go j j @3
```

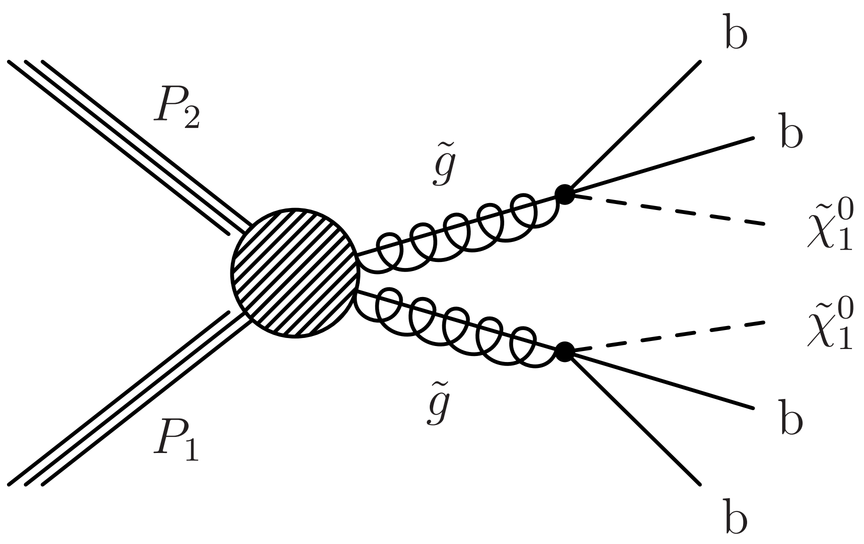
- Run madgraph at LO to produce sparticle pairs with up to two extra partons in the ME.
- xqcut and qcut values for matching/merging determined per process, for different mass ranges
- Sparticles are left undecayed at this step.
  - 2D scan reduced to 1D scan, only free parameter is mass of produced particle
  - Reuse 1D scan for multiple SMS, just need to perform different sparticle decay with Pythia

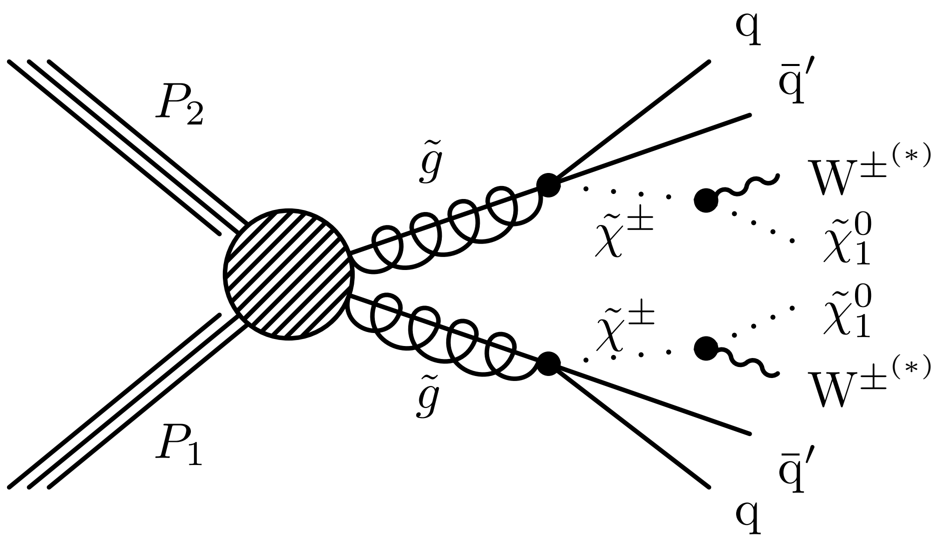


Decaying sparticles with Pythia

Reuse madgraph LHE files for multiple decay modes: e.g. pp→$\tilde{g}\tilde{g}$



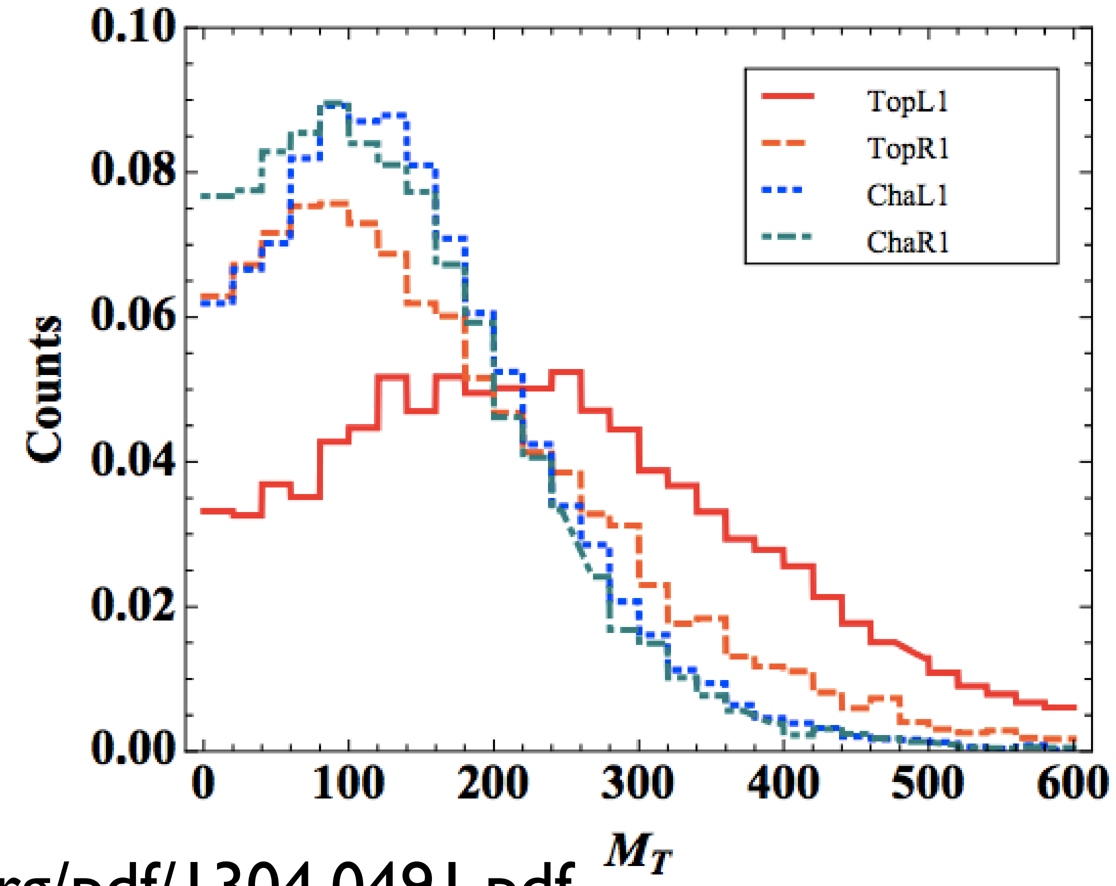




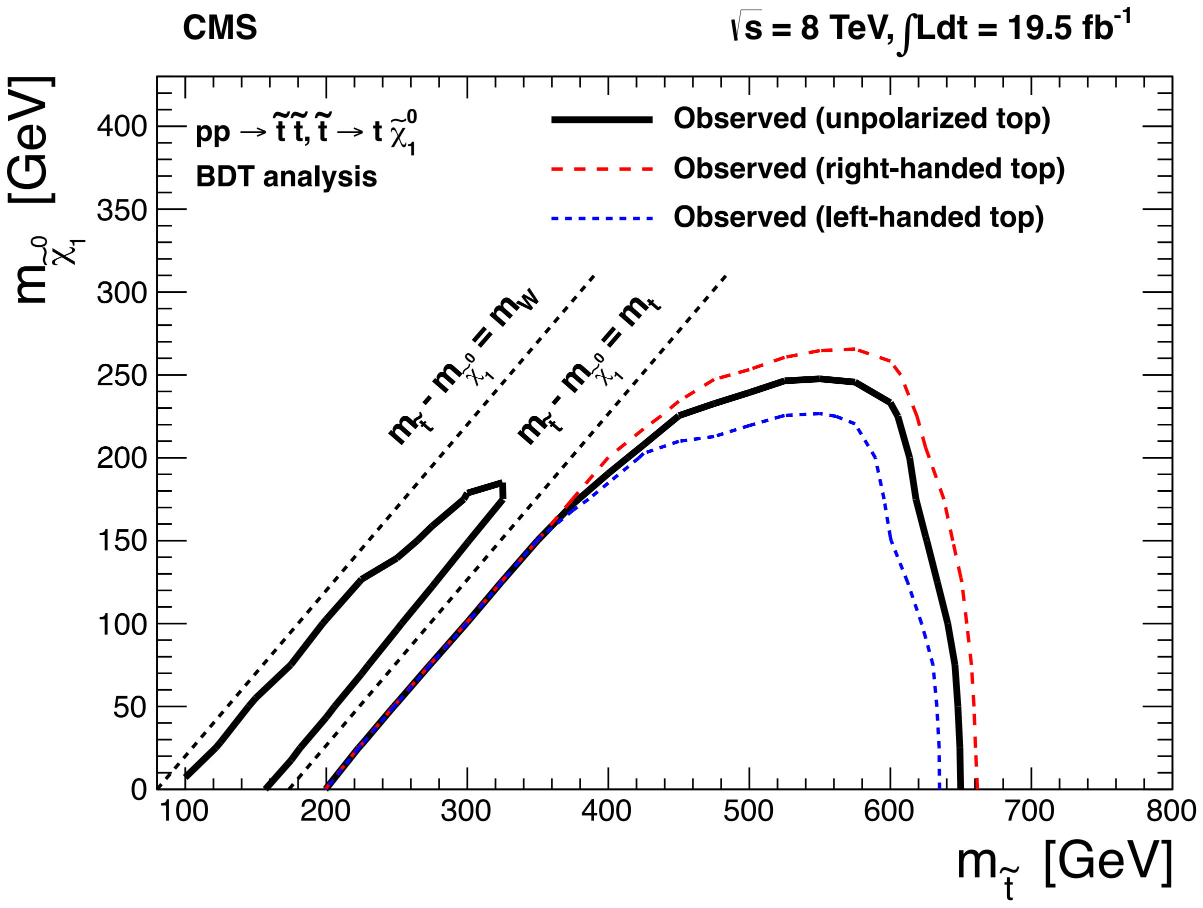
- Read in undecayed LHE file and SLHA with details of the SMS (decay, masses, etc.)
- Output new LHE file that is passed to central CMS production for simulation/reconstruction
- Decay done with flat matrix element → lose information about polarization, for example.

Event Re-weighting

- We decided not to generate any feature of the model (e.g. squark mixing).
- Instead, generate flat ME and then re-weight events for specific cases (e.g. LH/RH top).
- In case of stop production, worked out re-weighing procedure as top polarization was found to have a discernible impact through the lepton spectrum.
- Such re-weighting generally not considered though. Need to find a generic way to do this.



http://arxiv.org/pdf/1304.0491.pdf



SUS-13-011
arXiv:1308.1586

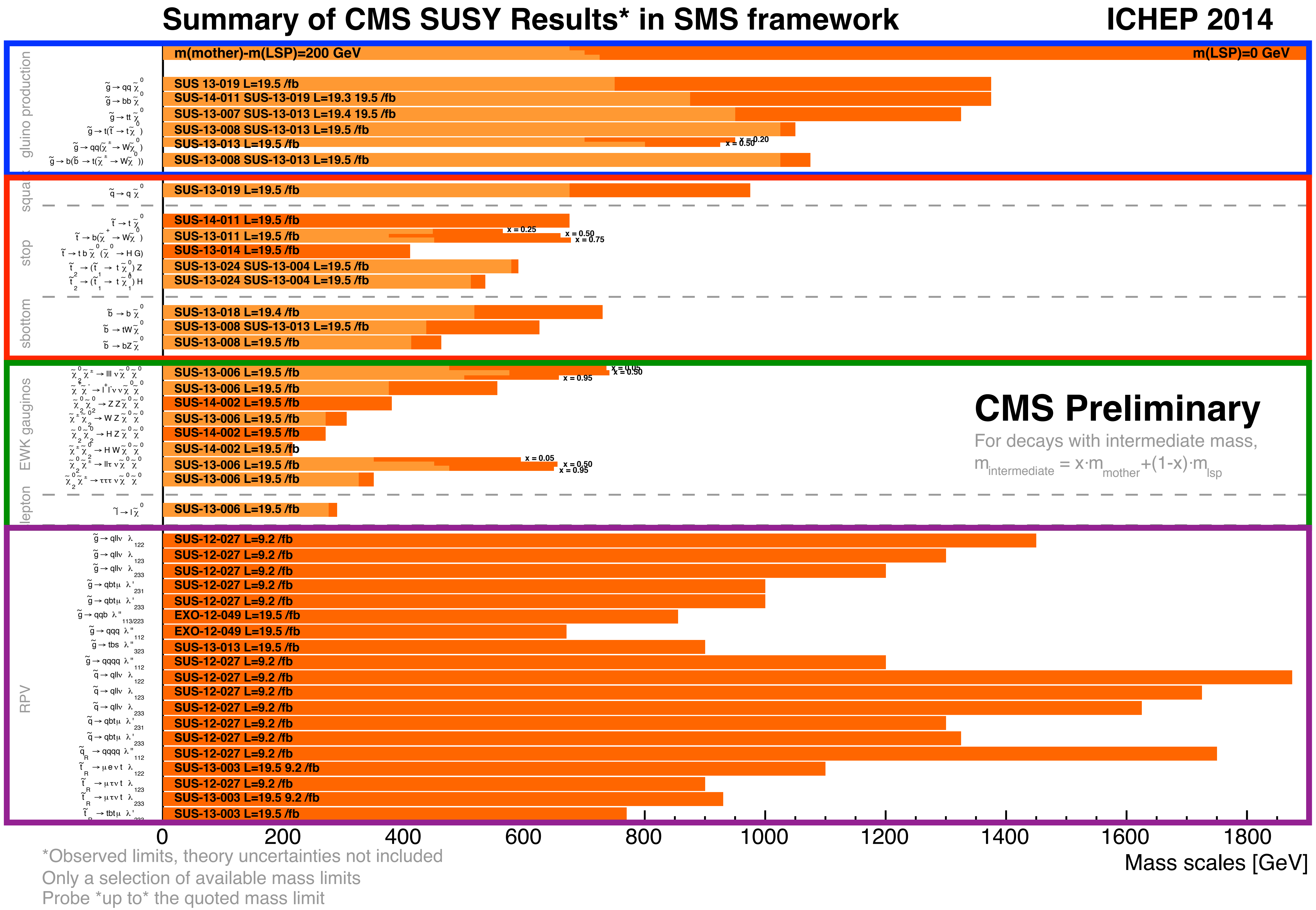
No evidence for SUSY in Run 1 Searches

gluino

squark

chargino
neutralino
slepton

RPV
gluino/squark

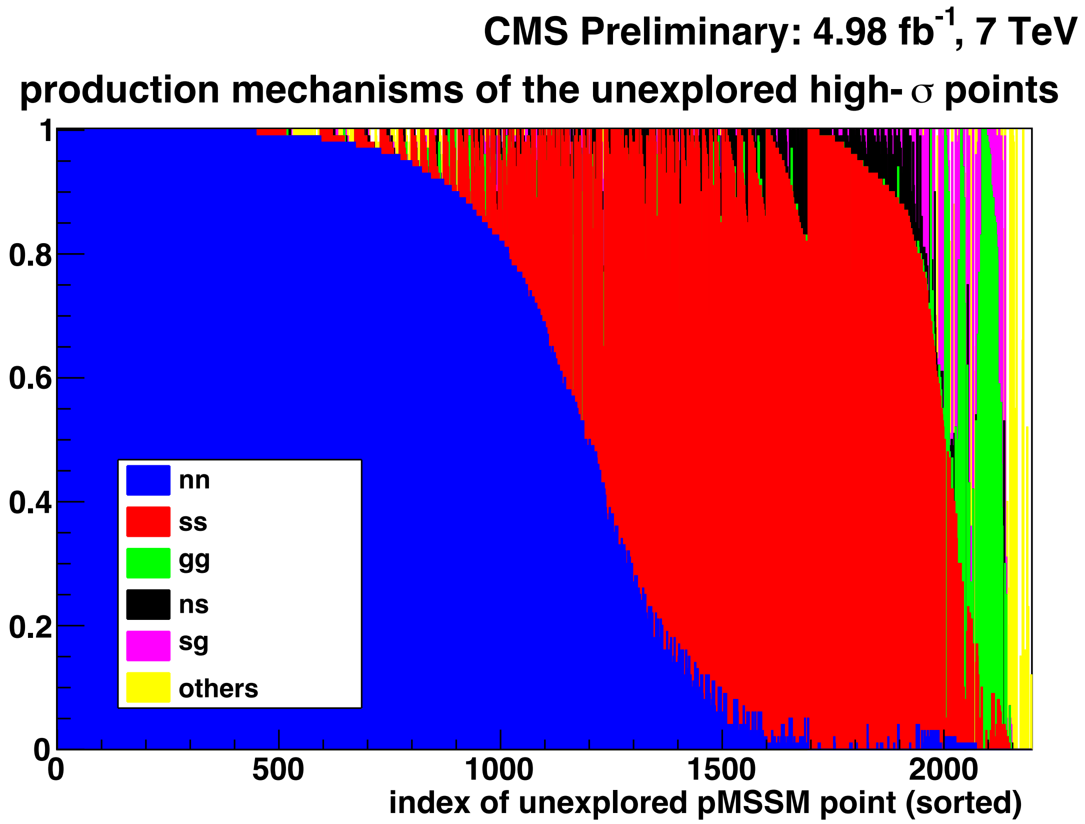


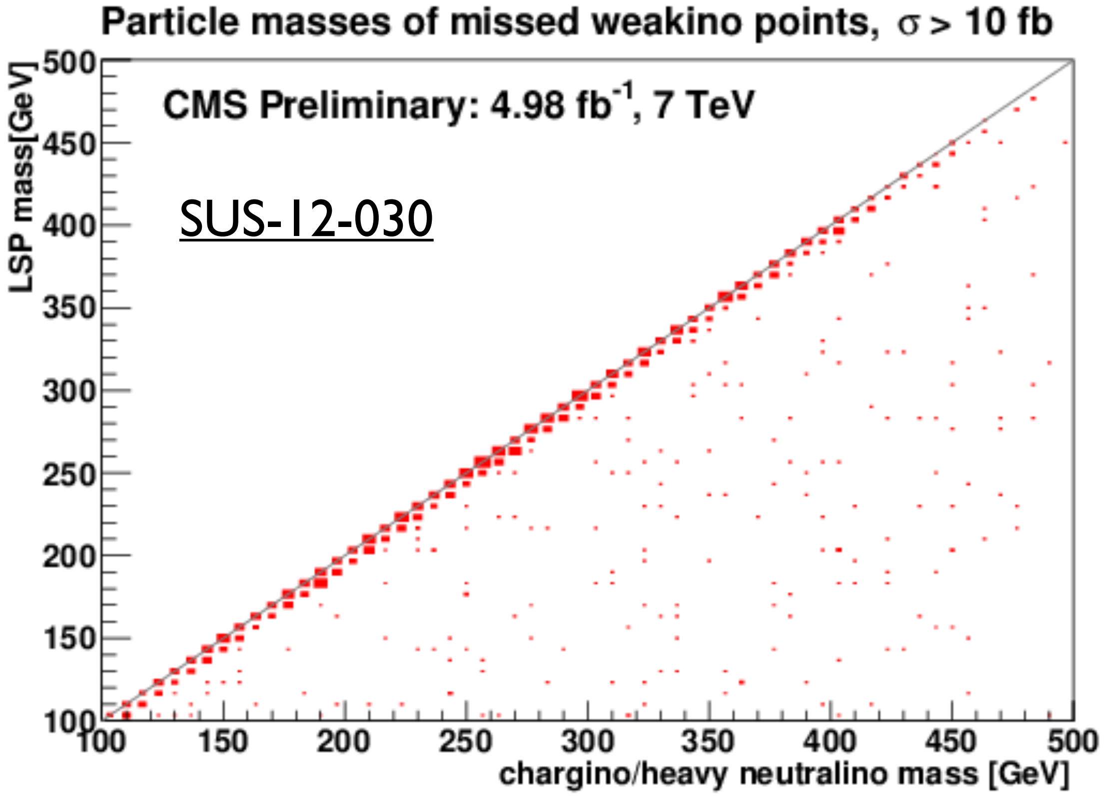
Aside: some numerology

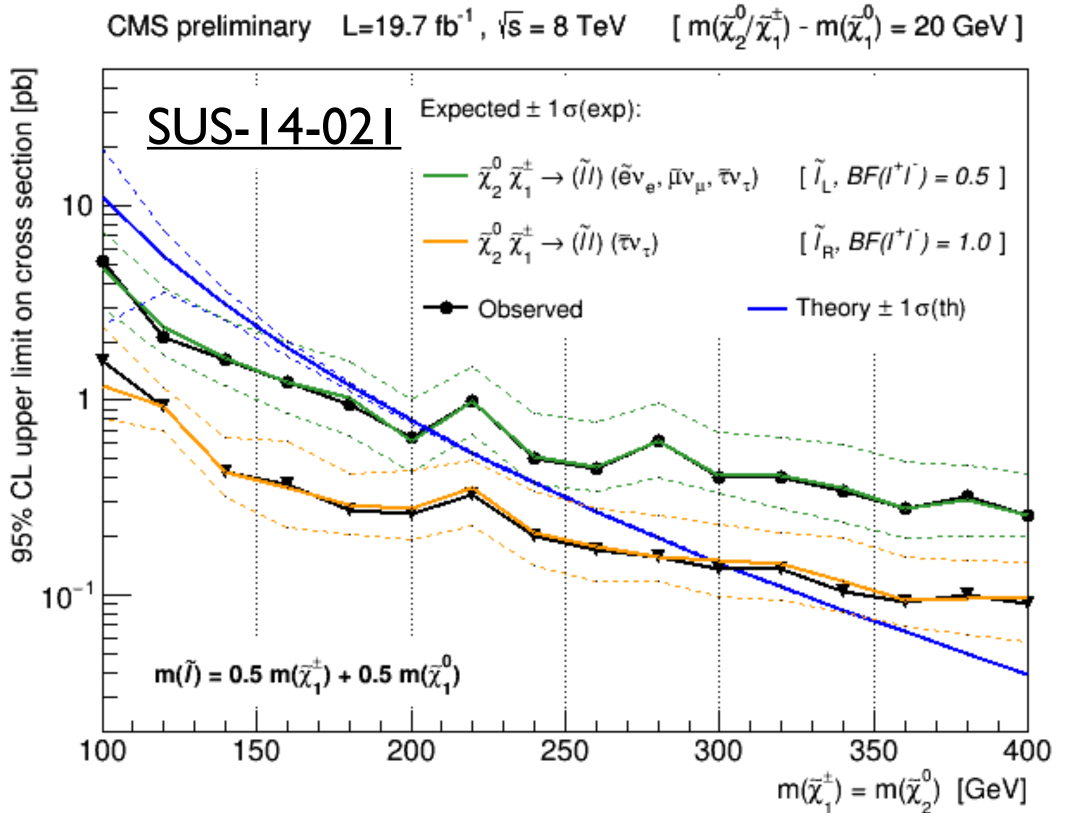
- Major challenge to supporting the results on the previous slide is the sheer scale.
- Populating 2D scans containing O(100) mass points with sufficient statistics requires producing O(10-100M) events.
  - Number of events produced with MG is ~x3 larger due to matching loss.
- Over 50 published or soon to be CMS SUSY searches interpret their results using a similar number of simplified model scans.
  - Over 300 samples totaling ~3B events → O(10B) signal events generated with MG.
- How does this scale as we collect larger datasets at higher energy? NLO?

Complement SMS with full models

- SMS have served us well and will likely continue to be primary way SUSY results are interpreted. However, it is important to not become myopic. Full models can be complementary, help us ensure that we are not overlooking signatures.





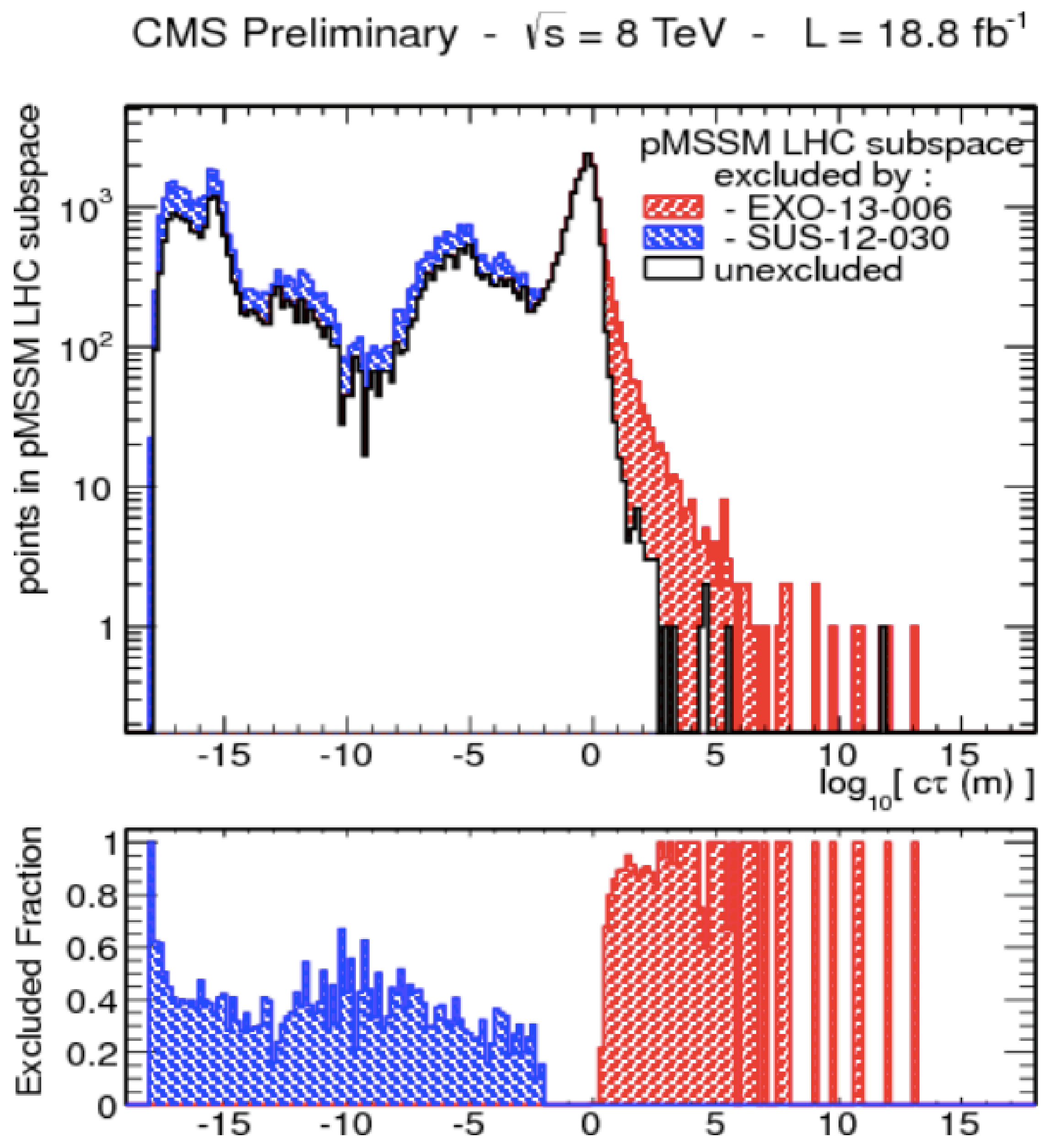


~30% of the sampled points have charginos that live long enough to make it to the detector.
We had to skip those points because FastSim cannot handle those charginos.
No complete study of what standard SUSY searches can say about these signatures.

Rich Program of Non-traditional Searches

Signature

stopped in muon system	delayed muon	Track in muon system at least 1 BX out-of-time from a collision.
stopped in calorimeter	stopped particle	Energy deposition in calorimeter at least 1 BX out-of-time from a collision.
	disappearing track	Isolated high p_T track with missing outer hits and appreciable MET.
displaced vertex in tracker	displaced leptons displaced stops	Tracks or muons with displaced vertex. Trigger on photons for electrons.
prompt	HSCP fractional charge	Track with dE/dx signature.



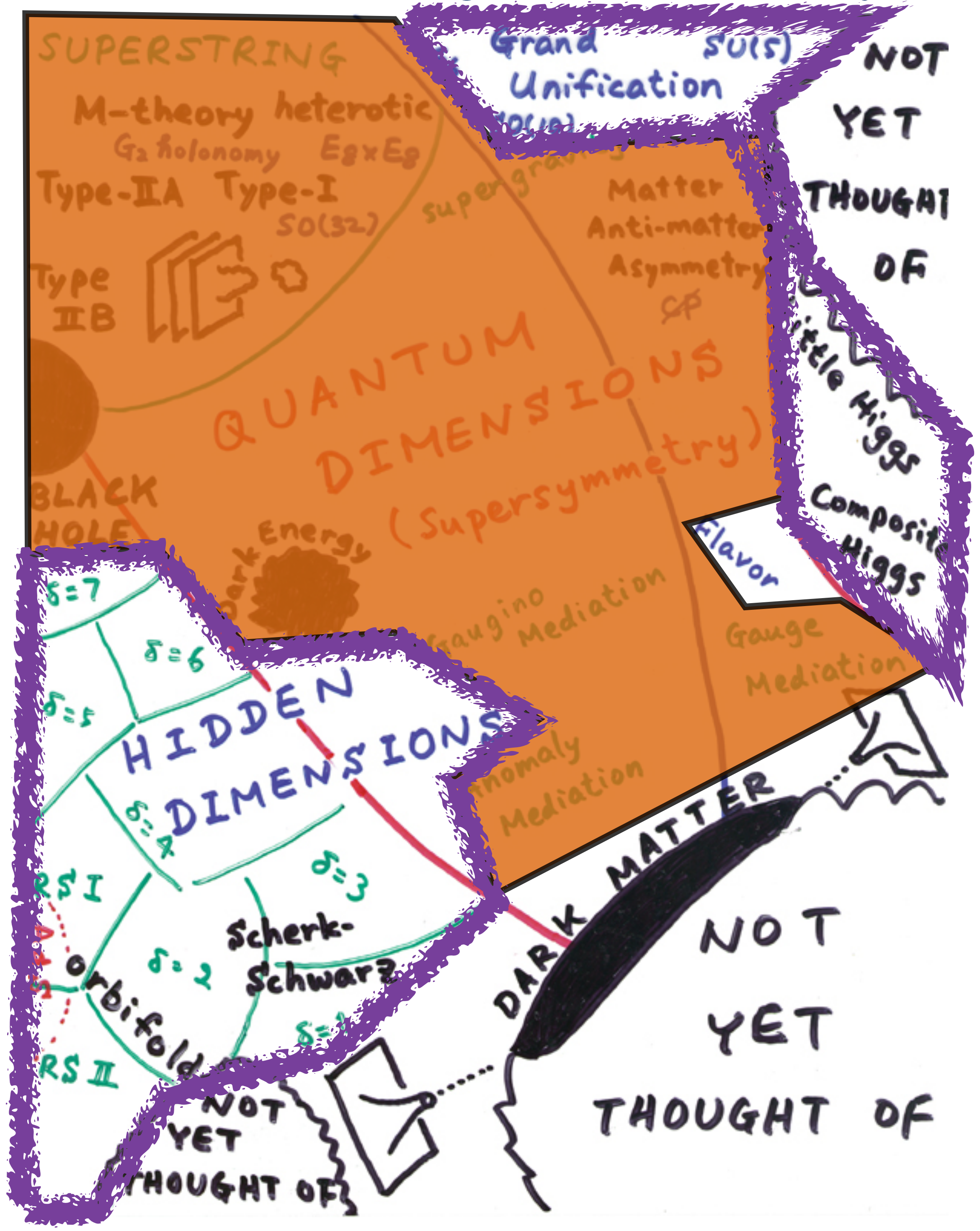
Process events generated with Pythia using full simulation and reconstruction with features developed especially for these novel signatures.

MC Challenges for Non-traditional Searches

- Simulating signals with long-lived particles requires special steps at each stage. Some examples:
  - stopped particles: Simulate performed in 2 steps. In "stage 1", simulate the R-hadrons to determine the position where they stop. Then in "stage 2" do the full GEN-SIM-RECO with the position for the decay based on stage 1.
  - displaced leptons: For long-lived particle decays simulated by Pythia, the long-lived particle's energy loss will not be simulated by Geant if it is a status 2 particle without a decay vertex. Is there an easy fix for this?
- New models often require R&D to produce — e.g. currently don't know how to simulate quirks.
- Signatures with a pair produced particle with multiple decay modes, where some are long-lived and some are not, is currently an unexplored opportunity. What else are we missing?

Naturalness beyond SUSY

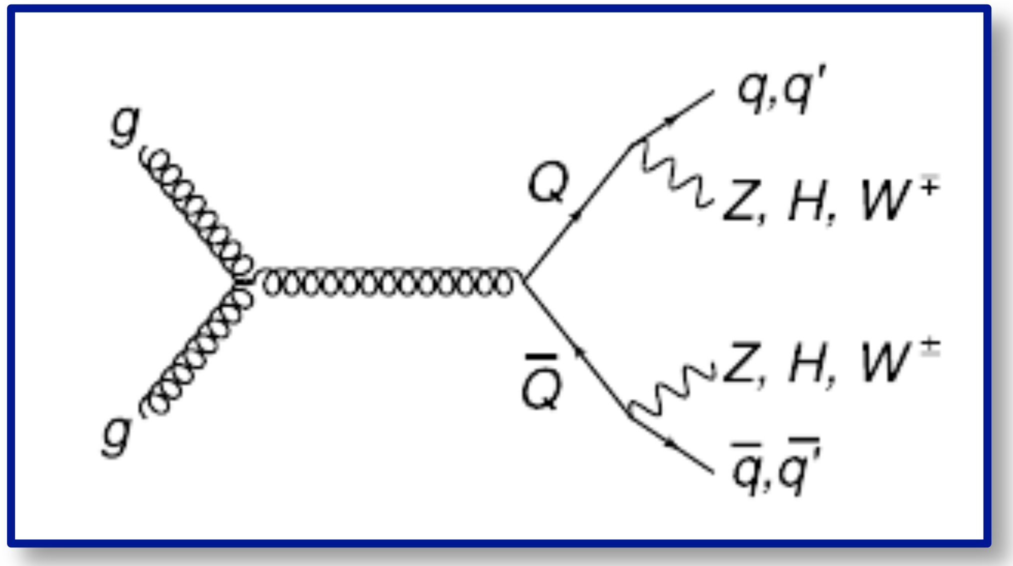
image : Hitoshi Murayama



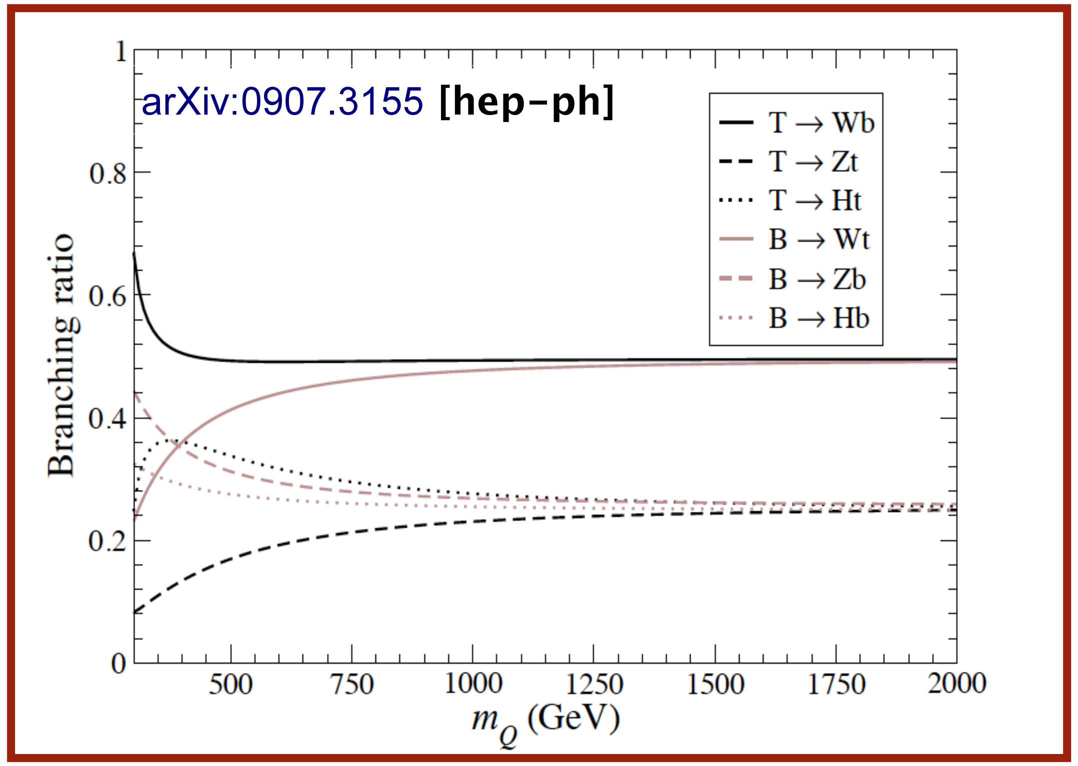
- Searches motivated by phenomenology from models with strong dynamics, RS KK particles, little Higgs, compositeness, etc.
- These often leads to signatures with (boosted) third generation quarks and SM bosons (W,Z,Higgs).
  - vector-like quarks (VLQ)
  - resonances decaying to W,Z,tops
  - DM produced in association with top quarks

VLQ in Run 1

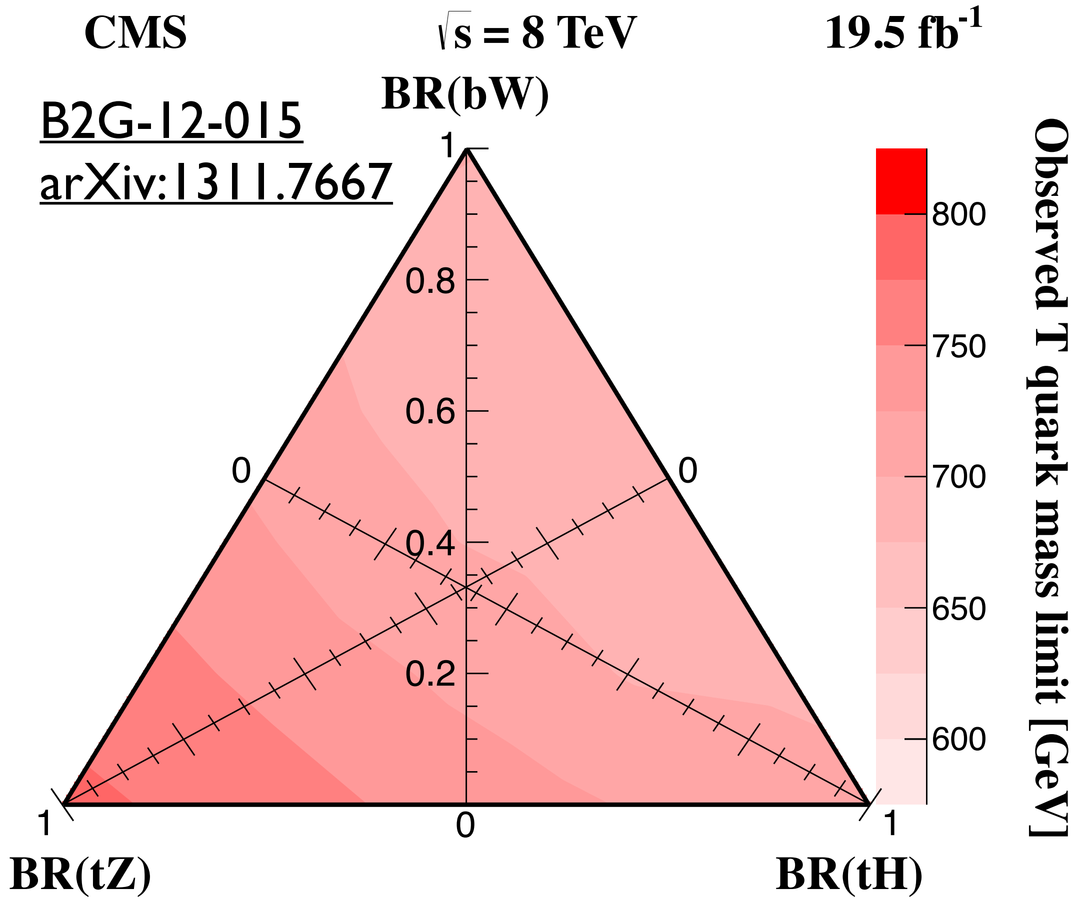
- Pair produced VLQ results in a large number of potential final states, <u>details are model-dependent</u>.
- Use MG (+0-2 partons)+Pythia6 to generate grid of mass points.
- Can re-weight to arbitrary decay fractions and interpret as a mass limit in the triangular plane of the branching ratios.



e.g. BR(bW)+BR(tZ)+BR(tH) = 1

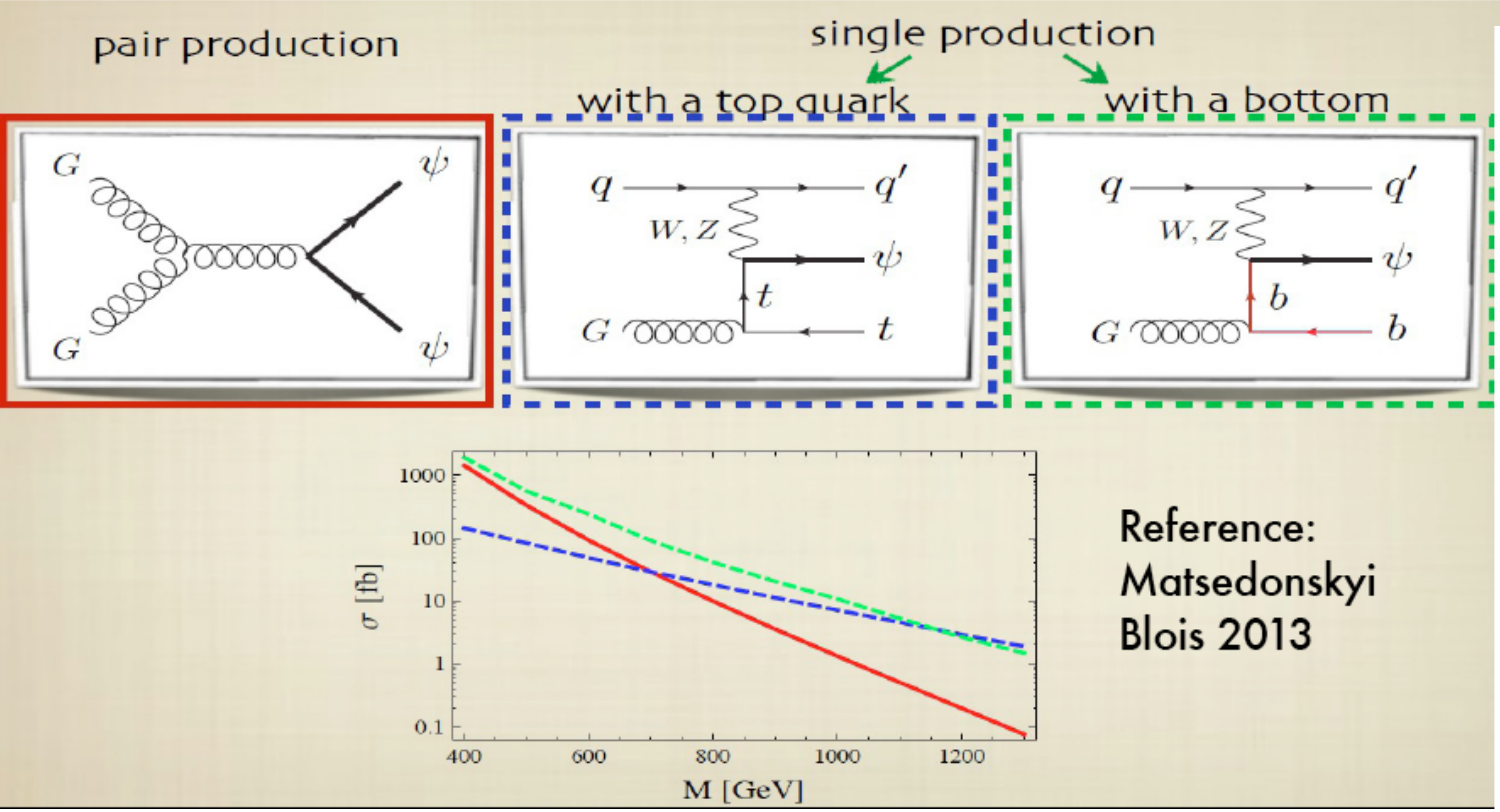


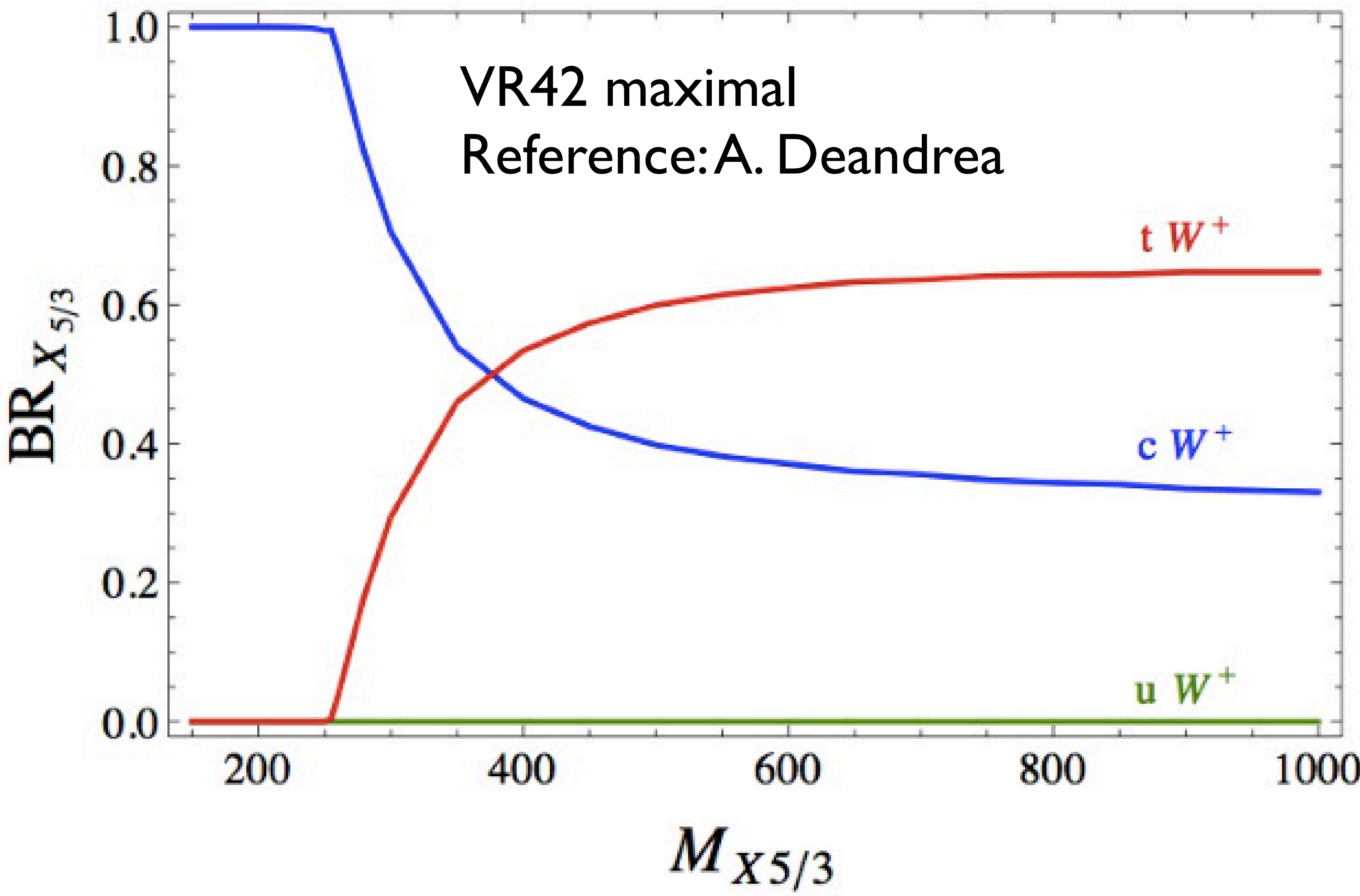
<u>B2G-12-015</u>
<u>arXiv:1311.7667</u>



Expanded program of VLQ searches in Run 2

- As searches push to higher mass sensitivity, single VLQ may become dominant production mode.
  - Weak coupling introduces model-dependence into VLQ production mechanism.
  - Mixing between VLQ and SM generations allow/forbid additional production and decay modes.



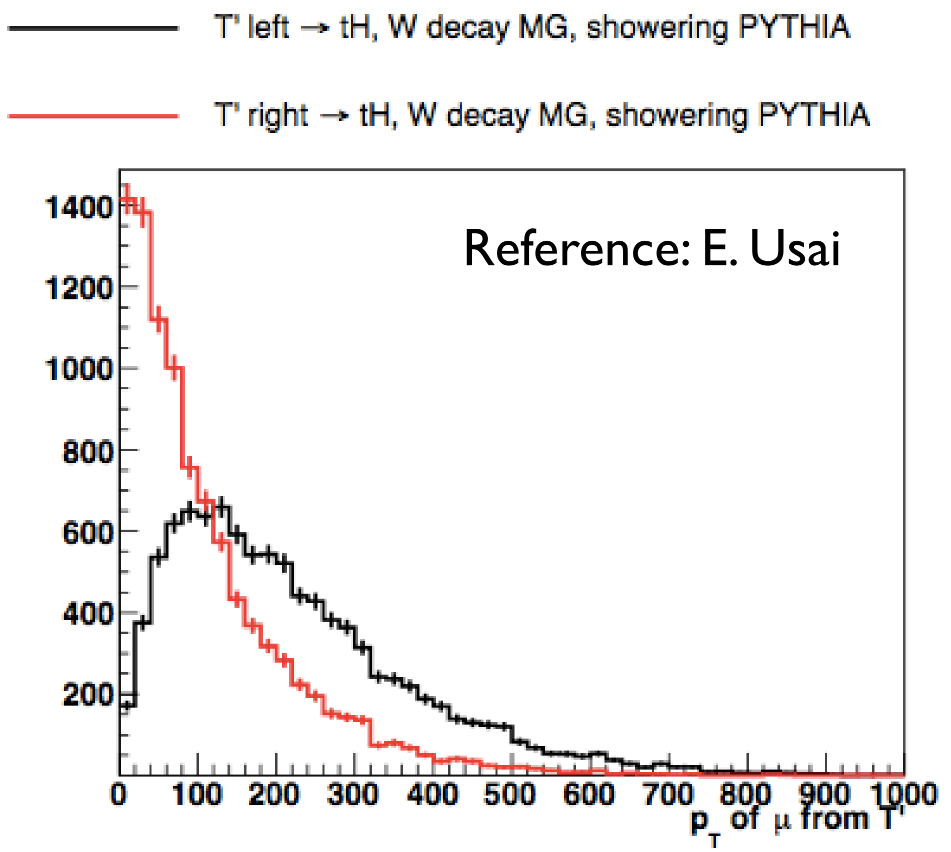


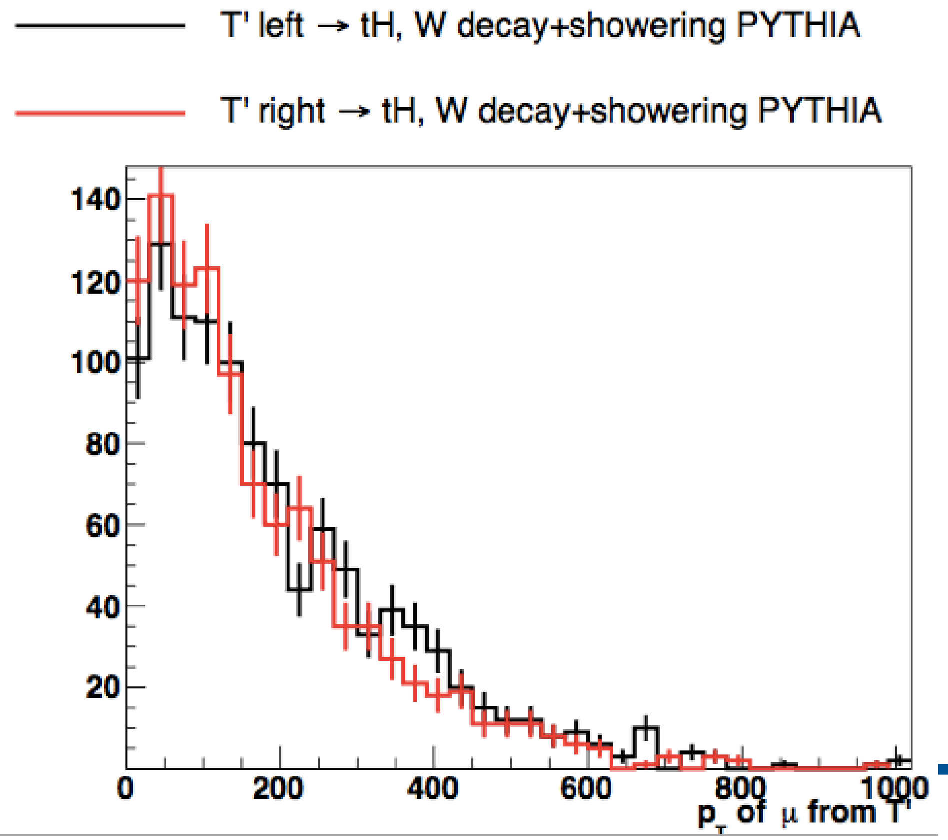
Producing VLQ MC for Run 2

- Generate pair and single VLQ production with MG, 4 flavor scheme, 3rd gen couplings only (for now).

- Most distributions show little sensitivity to chirality, but dependence seen in some cases, especially for daughters of top decays when the latter are boosted. Have not found a consistent way to treat this.
  - Ideal: consistent handling of correlations by Pythia, under development with authors.
  - Interim solution: decay VLQ, top, W with MG (or Madspin), rest with Pythia8.
  - For now, generate LH and a few mass points of RH and work to develop procedure to re-weight to obtain arbitrary mixing in important cases (if possible).

- Aim to produce MC in a model-independent way to allow interpretation of results in many scenarios → much larger scale effort than Run 1, significant resources (human and computing) required.





What is being produced?

pair-produced VLQ

(TT): 6 combinations of 3 decay modes (bW,tZ,tH)
scan masses from 800-1800 GeV

(BB): 6 combinations of 3 decay modes (tW,bZ,bH)
scan masses from 800-1800 GeV

($X_{5/3}X_{5/3}$): decay to tWtW
scan masses from 800-1800 GeV

Z'→TT: 6 combinations of 3 decay modes
2 widths: narrow, wide (~x2 exp. resol.)
2D scan in mass of Z',T

single VLQ production

(Ttj, Tbj): 3 decay modes (bW,tZ,tH)
scan masses from 700-1800 GeV

(Btj, Bbj): 3 decay modes (tW,bZ,bH)
scan masses from 700-1800 GeV

($X_{5/3}$tj): decay to tW
scan masses from 700-1800 GeV

Z'→Tt: 3 decay modes (bW,tZ,tH)
2 widths: narrow, wide (~x2 exp. resol.)
2D scan in mass of Z',T

For a process such as pp→Xtj, the cross section scales as the coupling squared:

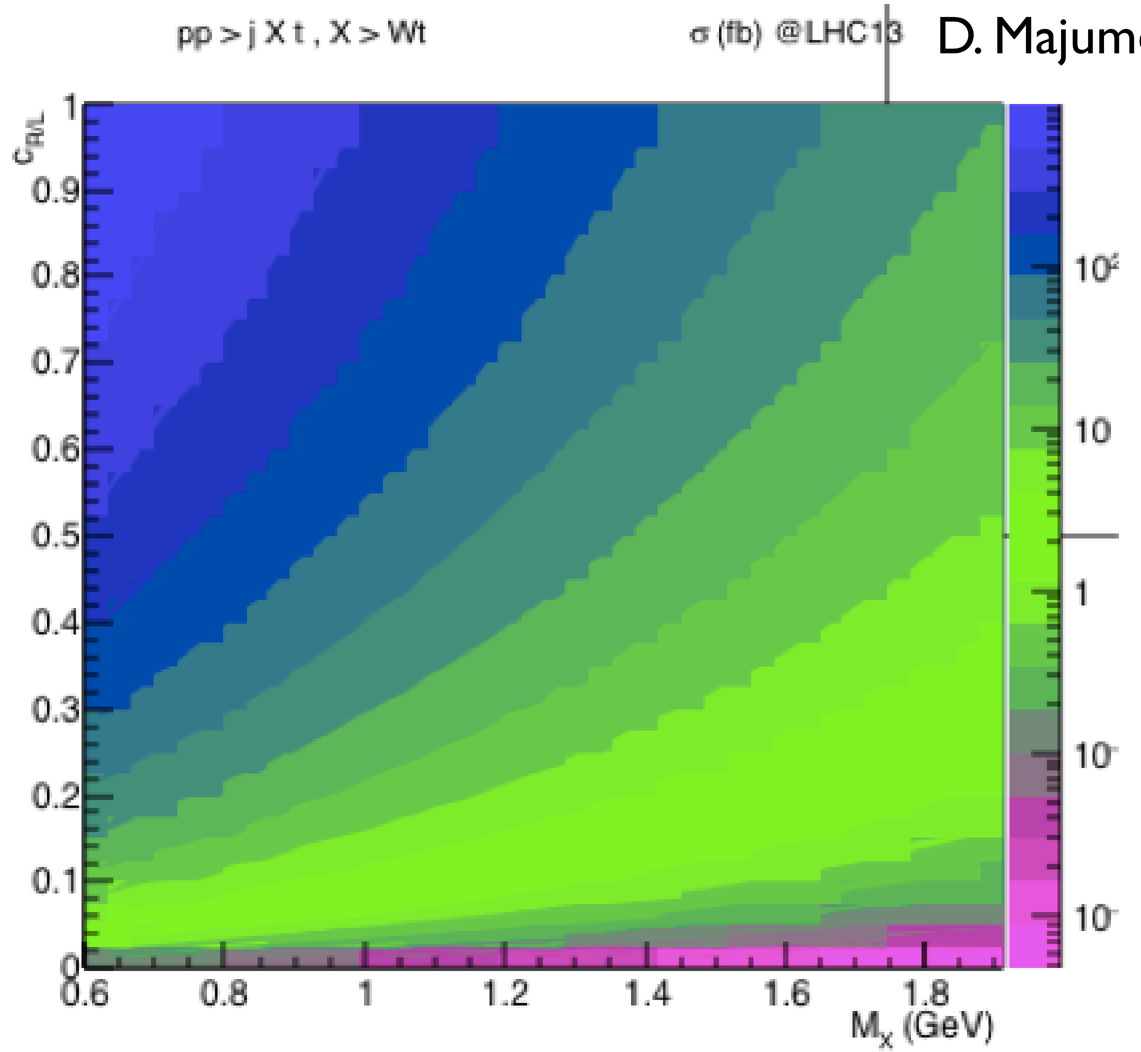
$$\sigma_{\rm sing}\left(X\bar{t}\right) = \left[\left(c_L^{XV}\right)^2 + \left(c_R^{XV}\right)^2\right]\sigma_{V\bar{t}}\left(M_X\right)$$

(assume $C_L \to 0$)

A. Wulzer
(see arXiv:1211.5663)

partner (MG name)	Q	couplings			
		$W^\pm$	Z	h	$W^\pm W^\pm$
$T_{2/3}$ (T23)	2/3	$c_L^{TW},\ c_R^{TW}$	$c_L^{TZ},\ c_R^{TZ}$	$c_L^{Th},\ c_R^{Th}$	—
$B_{1/3}$ (B13)	-1/3	$c_L^{BW},\ c_R^{TW}$	$c_L^{BZ},\ c_R^{BZ}$	$c_L^{Bh},\ c_R^{Bh}$	—
$X_{5/3}$ (X53)	5/3	$c_L^{XW},\ c_R^{XW}$	—	—	—
$Y_{4/3}$ (Y43)	-4/3	$c_L^{YW},\ c_R^{YW}$	—	—	—
$V_{8/3}$ (V83)	8/3	—	—	—	$c_L^{VW},\ c_R^{VW}$

D. Majumder



Interpret results in coupling-mass plane.

Constraints on couplings can then be recast as limits on favorite model.

Dark Matter Searches with CMS

- In Run 1, many analyses interpreted using effective field theory (EFT), but with known limitations.
- For Run 2, complement with simplified models.
- Still using EFT where UV completions not available, but with truncation scheme to remove problematic kinematic regions.
- To facilitate this, have established a joint ATLAS/CMS/Theory working group where discussions on models, parameter scans, etc. can take place.

Producing DM MC for Run 2

- Define simplified models with (minimum) 4 parameters: mass of mediator, mass of dark matter candidate, coupling of mediator to SM particles, coupling of mediator to DM.
- Consider four types of DM particles: Dirac/Majorana fermion, Real/complex scalar.
- Four types of mediator: vector, axial, scalar, pseudoscalar.

Final Sate	Scalar	Pseudoscalar	Vector	Axial
Mono jet	MCFM	MCFM	Powheg(NLO) MCFM backup	Powheg(NLO) MCFM backup
Mono-W	JHUGen	JHUGen	Madgraph	Madgraph
Mono-Z	JHUGen	JHUGen	Madgraph	Madgraph
TT+DM	Madgraph	Madgraph	N/A	N/A
BB+DM	Madgraph	Madgraph	N/A	N/A
Mono-Higgs	Madgraph	Madgraph	Madgraph	Madgraph

Reference: M. Trovato

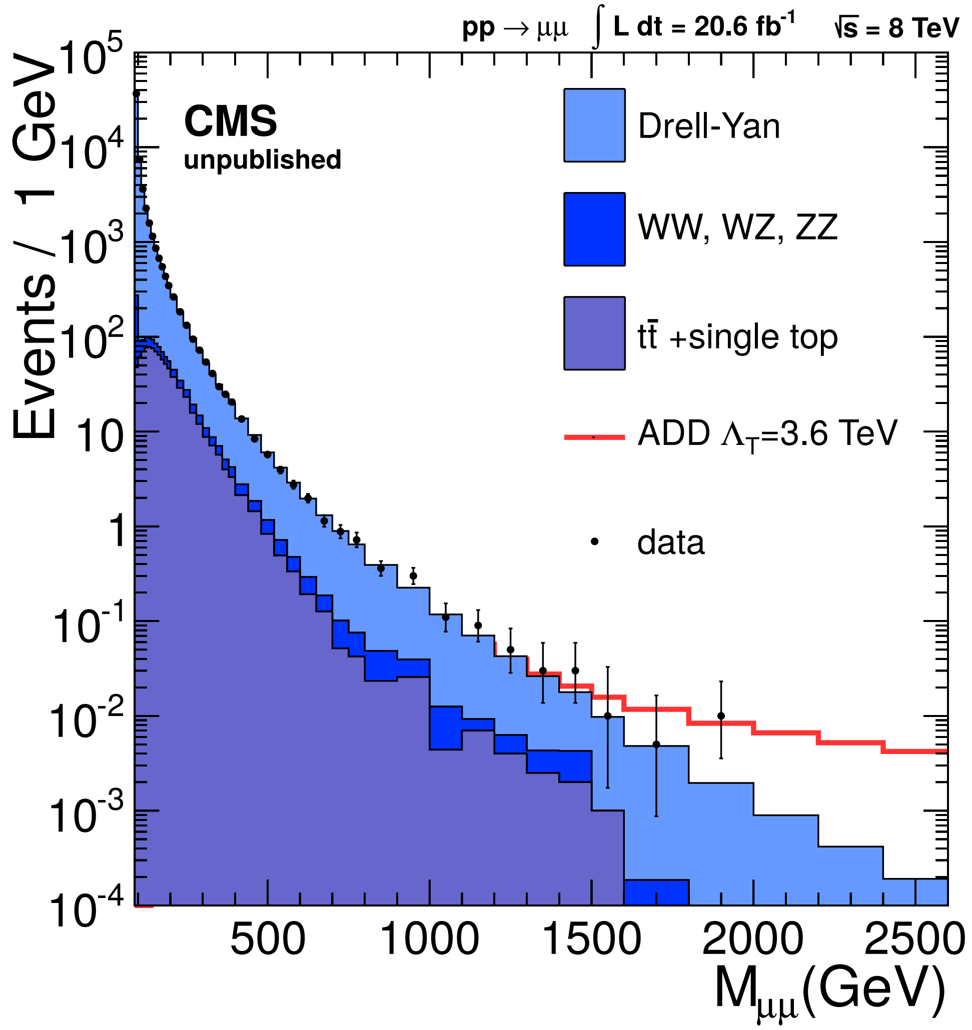
Challenges also exist in the backgrounds

- In the past, we have tried to use data-driven methods to estimate backgrounds to BSM searches, but not always possible as searches sometimes:
  - Probe kinematic regions where the measurement of a particular SM process is difficult or not possible.
  - Probe signatures with significant background from SM processes that have never been measured.
- While the potential robustness of data-driven methods to estimate the background to searches from SM processes will continue to be an important part of the BSM program, the performance of tools for the generation and simulation of known processes has enabled us to rely on MC for searches to an unprecedented degree.
  - As experimentalists, need to understand physics and technical limitations.

Golden Signature: high mass dilepton pairs

- Expect significant increase in reach going to 13 TeV — interesting results already with ~1 fb^{-1}.
- Counting experiments — background dominated by calculable SM processes, taken from simulation.
- Sensitive to higher order corrections, enhancement/suppression of tails from EWK Sudakov logs, etc.

EXO-12-061
arXiv:1412.6302



Straightforward signal generation with Pythia.

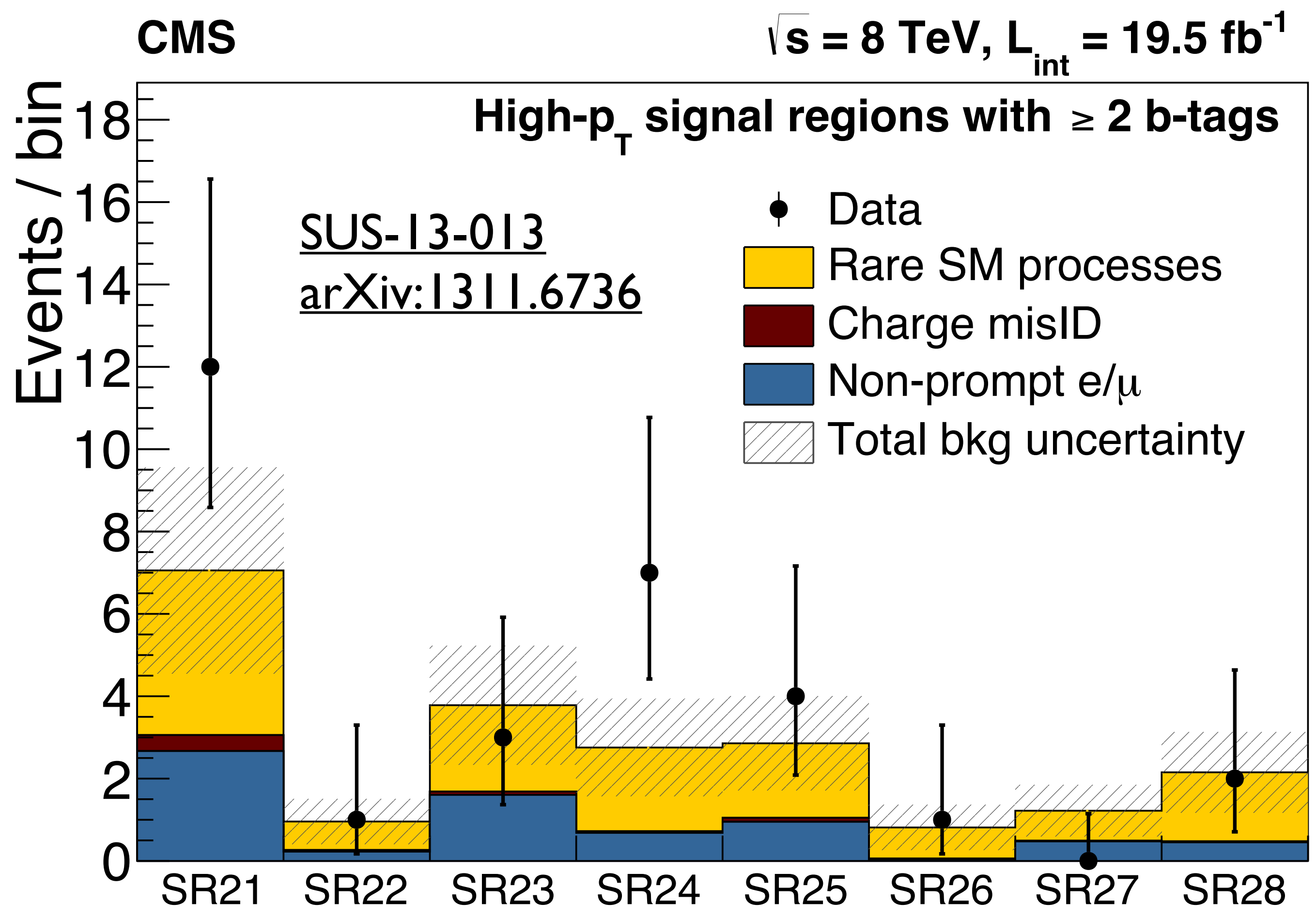
Use Powheg to take advantage of higher order corrections, when generating Z, WW, ttbar backgrounds.

To fill corners of phase space requires either generating very large number of events (not feasible) or generating events in bins of relevant variable.

More flexible binning helpful for searches — e.g. ability to specify upper and lower $M_{\ell\ell}$ cuts.

Main backgrounds to searches with same-sign lepton pair

- SM processes with a fake lepton — i.e. a jet mis-identified as a lepton, including b/c→ℓ.
- "rare" SM processes with a genuine same-sign lepton pair — e.g. ttW/Z, ttH, $qqW^{\pm}W^{\pm}$, WZ



Background from "rare" SM processes taken from simulation, corrected for known differences between data and MC in selection of leptons, jets, b-tagged jets.

Took a flat 50% systematic uncertainty on the total rare background estimate.

Improvements with evolution of MC generators

- Event weights for scale, pdf variations written into LHE files will now allow for the evaluation of better motivated uncertainties for each process and each search region in a trivial way.
- Possibility to generate processes at NLO → reduced scale uncertainty.

Reference: A. George

ttW, NLO, 13 TeV

	Fixed Scales		
σ	$662.69 \pm 0.54\%$	$^{+26.70\%}_{-21.88\%}$	$\pm 1.70\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR1	$16.34 \pm 3.34\%$	$^{+9.45\%}_{-16.82\%}$	$\pm 2.10\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR2	$2.88 \pm 8.23\%$	$^{+21.57\%}_{-21.67\%}$	$\pm 1.96\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR3	$7.52 \pm 5.49\%$	$^{+5.87\%}_{-15.42\%}$	$\pm 1.72\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR4	$10.67 \pm 4.23\%$	$^{+56.19\%}_{-30.25\%}$	$\pm 1.20\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR5	$33.93 \pm 2.38\%$	$^{+10.03\%}_{-18.12\%}$	$\pm 1.97\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR6	$10.58 \pm 4.20\%$	$^{+41.62\%}_{-27.89\%}$	$\pm 1.76\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR7	$16.39 \pm 3.57\%$	$^{+24.54\%}_{-21.72\%}$	$\pm 1.62\%$
$\sigma \cdot \varepsilon \cdot \text{lumi} \cdot \text{norm}$, SR8	$35.40 \pm 2.27\%$	$^{+79.27\%}_{-36.42\%}$	$\pm 1.56\%$

qqW±W±, LO, 13 TeV

	Fixed Scales		
σ	$148.34 \pm 0.29\%$	$^{+76.91\%}_{-37.15\%}$	$^{+2.25\%}_{-1.70\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR1	$5.60 \pm 4.19\%$	$^{+60.50\%}_{-32.89\%}$	$^{+2.12\%}_{-1.56\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR2	$4.38 \pm 4.90\%$	$^{+76.11\%}_{-37.23\%}$	$^{+2.27\%}_{-1.66\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR3	$0.24 \pm 20.75\%$	$^{+70.17\%}_{-34.46\%}$	$^{+2.01\%}_{-1.53\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR4	$1.06 \pm 9.50\%$	$^{+96.28\%}_{-41.16\%}$	$^{+2.15\%}_{-1.55\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR5	$13.80 \pm 3.00\%$	$^{+72.09\%}_{-36.28\%}$	$^{+2.14\%}_{-1.52\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR6	$15.79 \pm 2.89\%$	$^{+89.45\%}_{-40.50\%}$	$^{+2.66\%}_{-1.84\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR7	$0.59 \pm 13.61\%$	$^{+81.69\%}_{-38.14\%}$	$^{+1.86\%}_{-1.39\%}$
$\sigma \cdot \varepsilon \cdot \text{lumi}$, SR8	$4.78 \pm 5.20\%$	$^{+108.56\%}_{-43.85\%}$	$^{+2.43\%}_{-1.66\%}$

$\not\!\!E_T$	# jets	H_T[200-400]	H_T[> 400]
50-120	2-3	SR1	SR2
	≥ 4	SR3	SR4
> 120	2-3	SR5	SR6
	≥ 4	SR7	SR8

Generator-level study.

pred±stat±scale±PDF

Vary fact/renorm. scale by x4.

Flat 50% uncertainty on total background taken from MC was overly conservative.
Could potentially benefit from qqW±W± at NLO, but need support for merging different parton multiplicities as many search regions require at least 4 jets.

Summary

- Run I results were a confirmation of the success our MC tools with agreement between theory and observation in almost every measurement.

- As a result of this success and the continued improvement of the tools by the theory community, MC tools are likely to play an even larger role in the BSM search program for Run 2 than previously.

- Generating BSM MC today is often no different than that for any typical SM process. As the scope of our MC use grows and we continue to explore more exotic signatures, there is a continued need to evolve the tools and maintain efforts between the communities.

- We're all looking forward with anticipation to once again having physics collisions this summer.