Exotica Monte Carlo and formats for reporting results

Maurizio Pierini
CERN

Exotica Monte Carlo and formats for reporting results

Maurizio Pierini
CERN

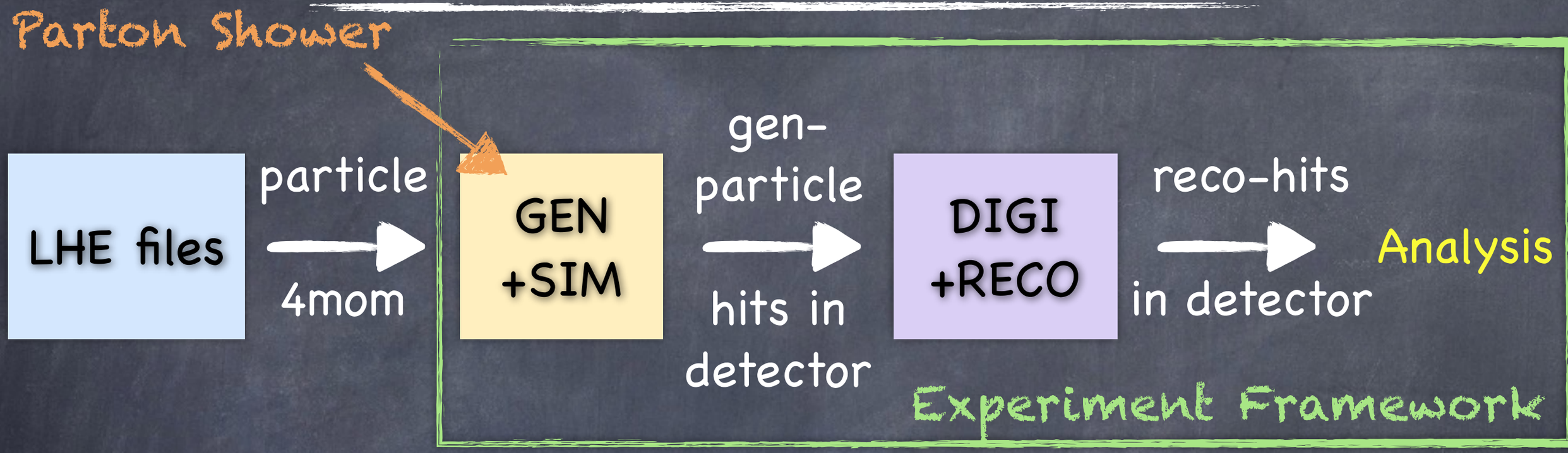
Disclaimer

- What I am going to say is mainly CMS specific
  - Personal bias: This is what I am familiar with
  - Details about Evt Generation workflows are usually internal to a Collaboration and very specific
  - Not sure if what I will say applies 100% to ATLAS, but to some extent the picture should be similar
- Three messages to convey
  - Common repository for MC samples (a new MCDB)
  - Official detector simulation to fold detector effects
  - (Even better) supporting RECAST to use official MC/analysis tool (also good for long-term preservation)

BSM results and Monte Carlo Generation

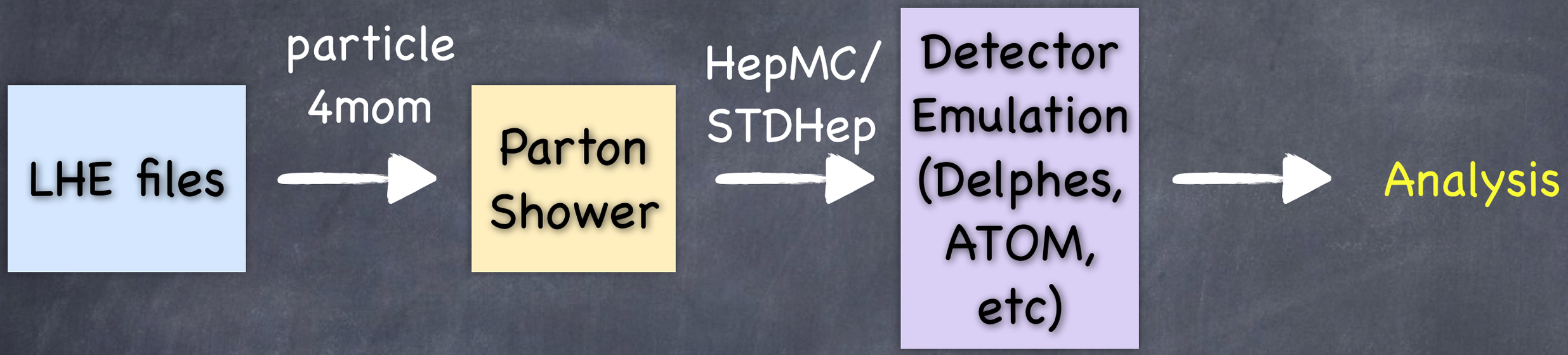
- For experiments to design a search:
  - benchmark models as motivation/guidance/interpretation of an experimental result
  - a complete set of benchmarks important to highlight weaknesses in search strategy (see mSugra vs Simplified Models in SUSY)
- For everybody to interpret the results:
  - experimental results have implications beyond benchmark models
  - established workflow for experimental analyses (e.g. Madgraph/ Pythia/Delphes) cover many cases. With which accuracy?
  - Sometimes situation more complicated (e.g. exotic signatures). A step up in accuracy comes from using official tools (e.g. detector simulation)

The easy part: Experiments



- MC can be generated only inside the Experiment Software Framework
- Using LHE files helps a lot MC Generation
  - easy bridge between MC generators & experimental frameworks
  - library of reusable LHE samples (important now that we might stay @ 13 TeV for long time)
- Why not a common repository (ATLAS/CMS/theorists) of samples or gridpacks, or UFOs. Something like what MCDB used to be

The easy part: Pheno studies

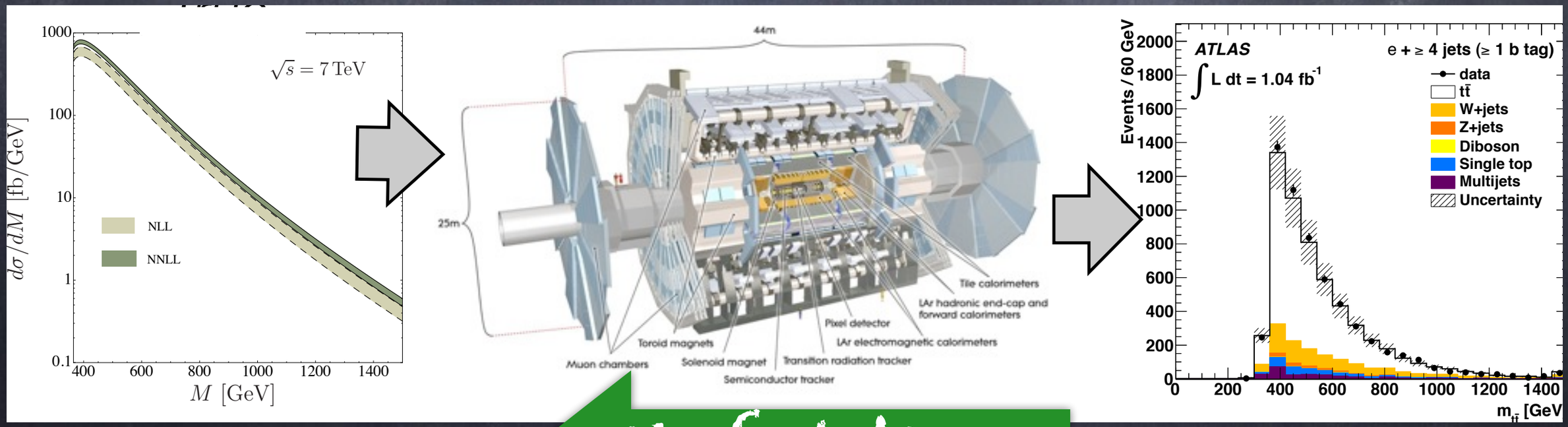


- A similar workflow is used in pheno studies, with simplified assumptions
  - simple (i.e. in average) detector response
  - optimistically clean collision (beam bkg, detector noise, out-of-time pileup)
- Sometimes, the picture is oversimplified
  - parametric detector emulations have some limitation (signal efficiency & resolution).
- Sometimes, the issue is more deep (i.e. bkg estimate, correlation between uncertainties, etc)

Detector effects: Folding vs Unfolding

- In many cases (mainly SM-related measurements) detector effects are unfolded, to go back to the "true" distribution before detector effects
  - pheno-usable distributions are provided: best possible format to present result
  - useful for PDF studies, tuning of the underlying event, parton showering, etc

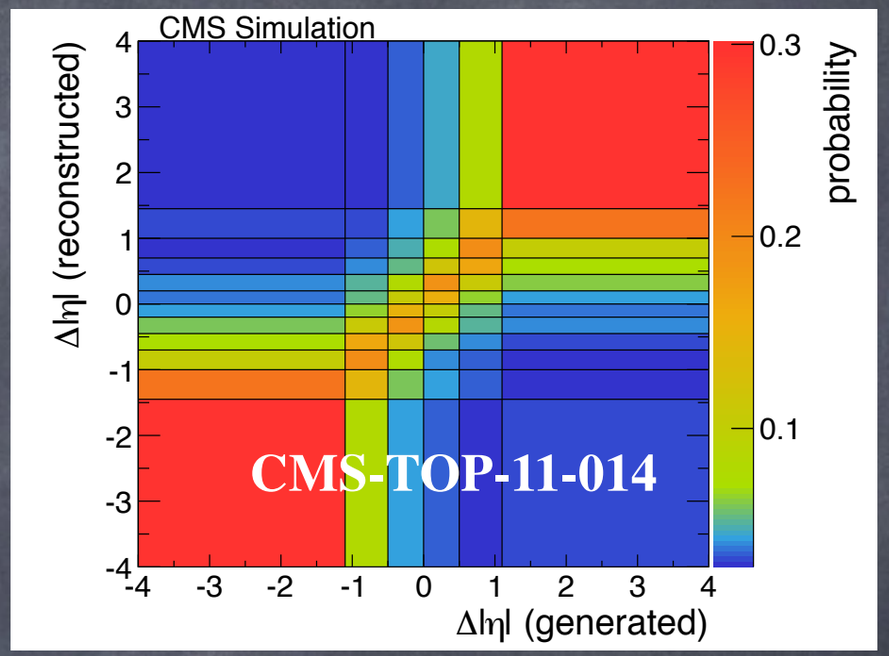
See for instance this instructive talk by

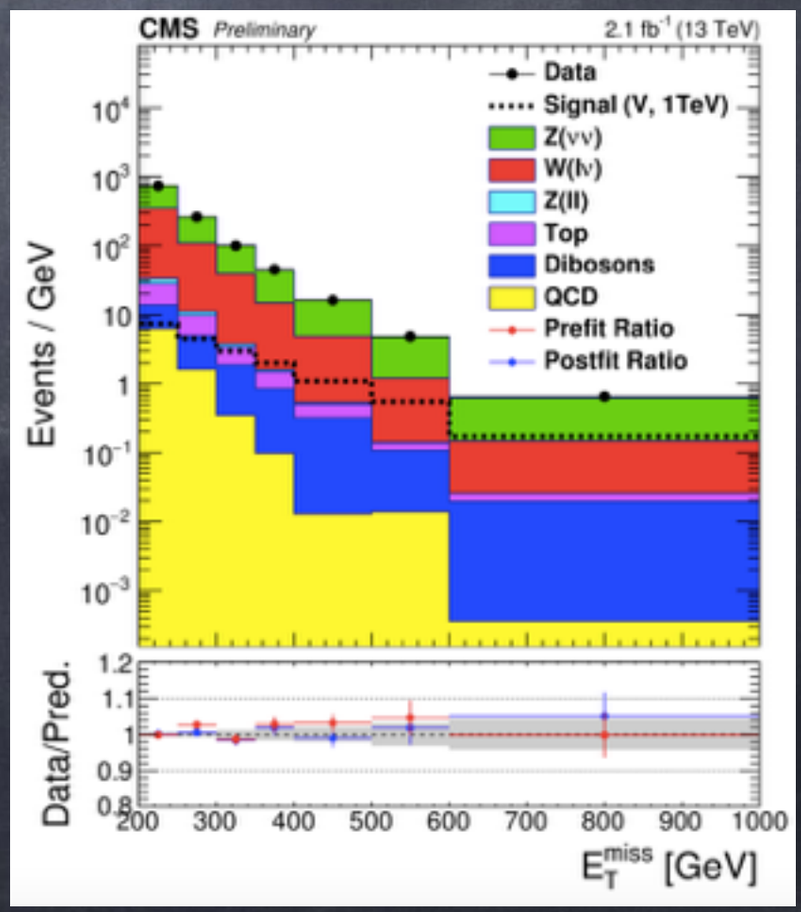


Unfolding

Detector effects: Folding vs Unfolding

- Unfolding has some delicate aspects
  - introduce additional systematic uncertainties
  - relies more on Monte Carlo samples
  - computational complexity limits number of bins
- Problematic for searches
  - many signal regions + different processes difficult to disentangle
  - low-statistics bins, which come with numerical instability

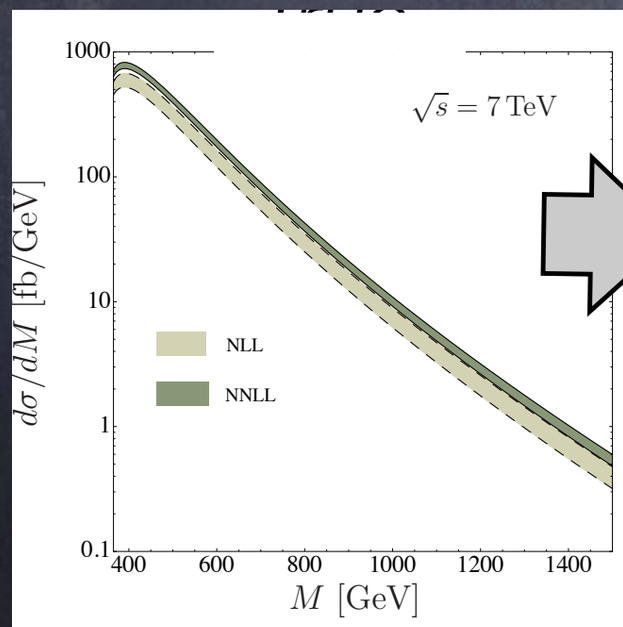


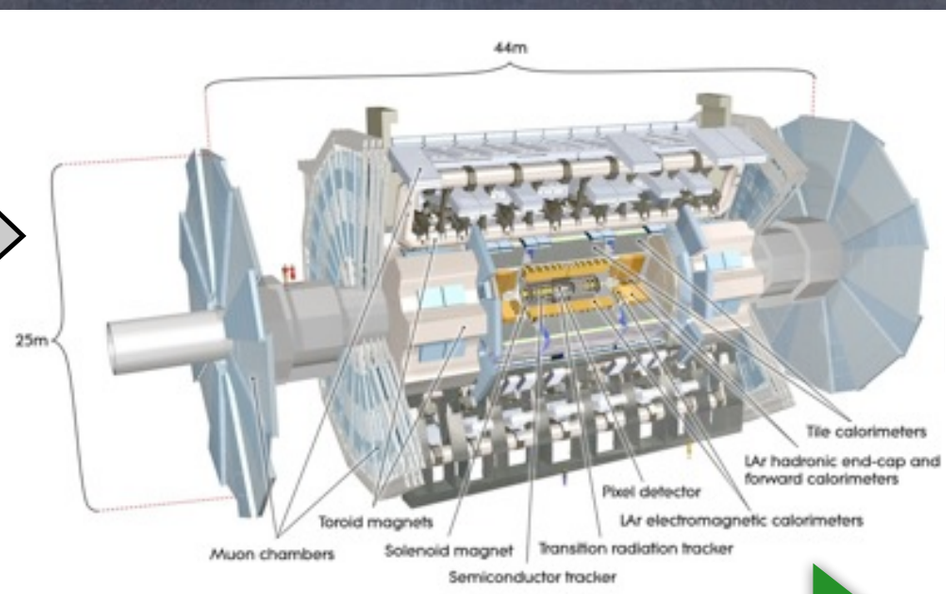


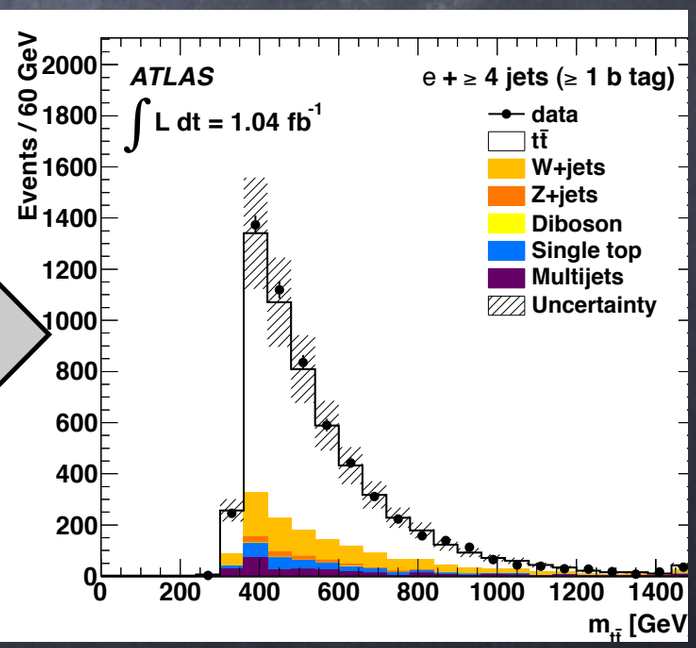
Detector effects: Folding vs Unfolding

- In many cases (mainly SM-related measurements) detector effects are unfolded, to go back to the "true" distribution before detector effects
  - pheno-usable distributions are provided: best possible format to present result
  - useful for PDF studies, tuning of the underlying event, parton showering, etc

See for instance this instructive talk by





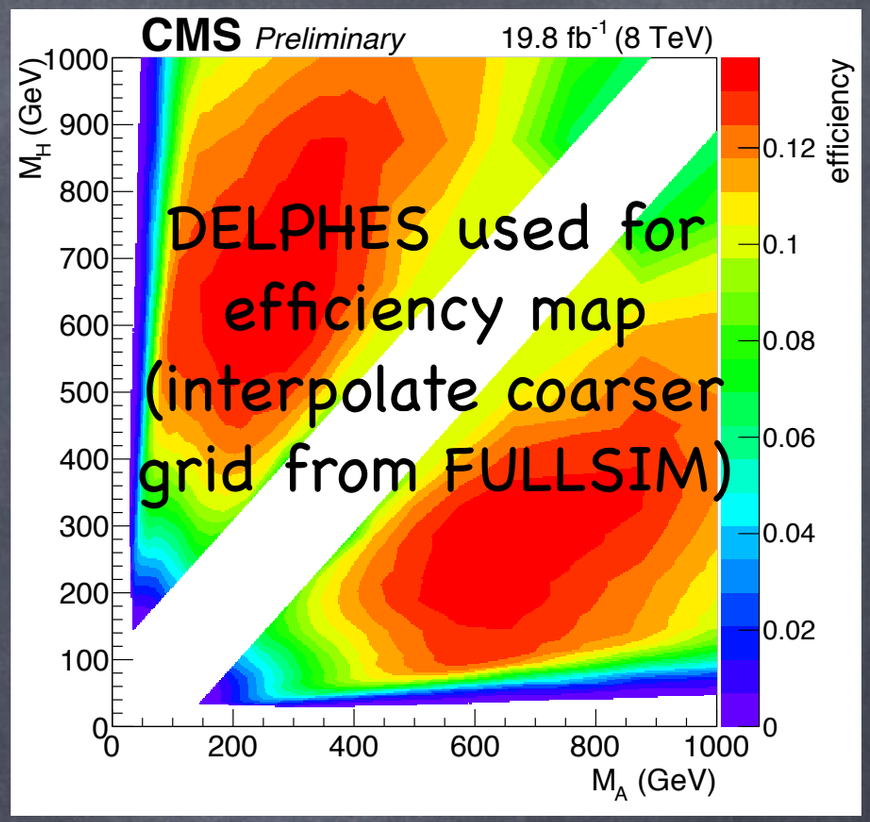


Unfolding

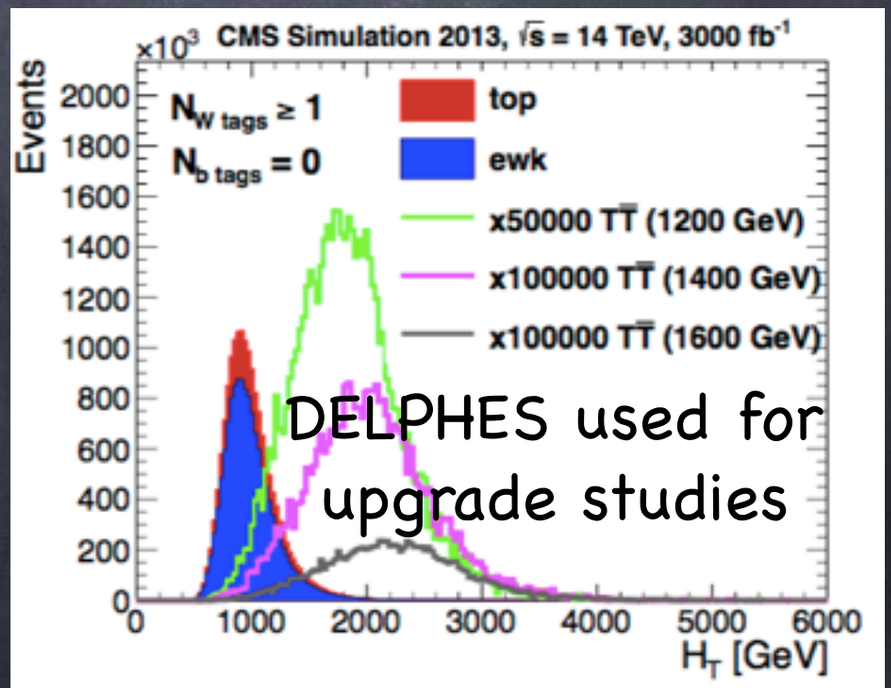
An Official Detector Parameterisation?

- Huge effort by many groups to define a parametric description of ATLAS and CMS
  - used in many pheno papers
  - also employed in official experimental work
- Detector performances change with time, which complicates the effort
  - new detector components in next years (started with ATLAS IBL)
  - new identification criteria, adapting to new running conditions (e.g., new PU scenarios, new bunch spacing, etc)
- Why not releasing and maintaining an official version of "detector cards"?

CMS-PAS-HIG-15-001

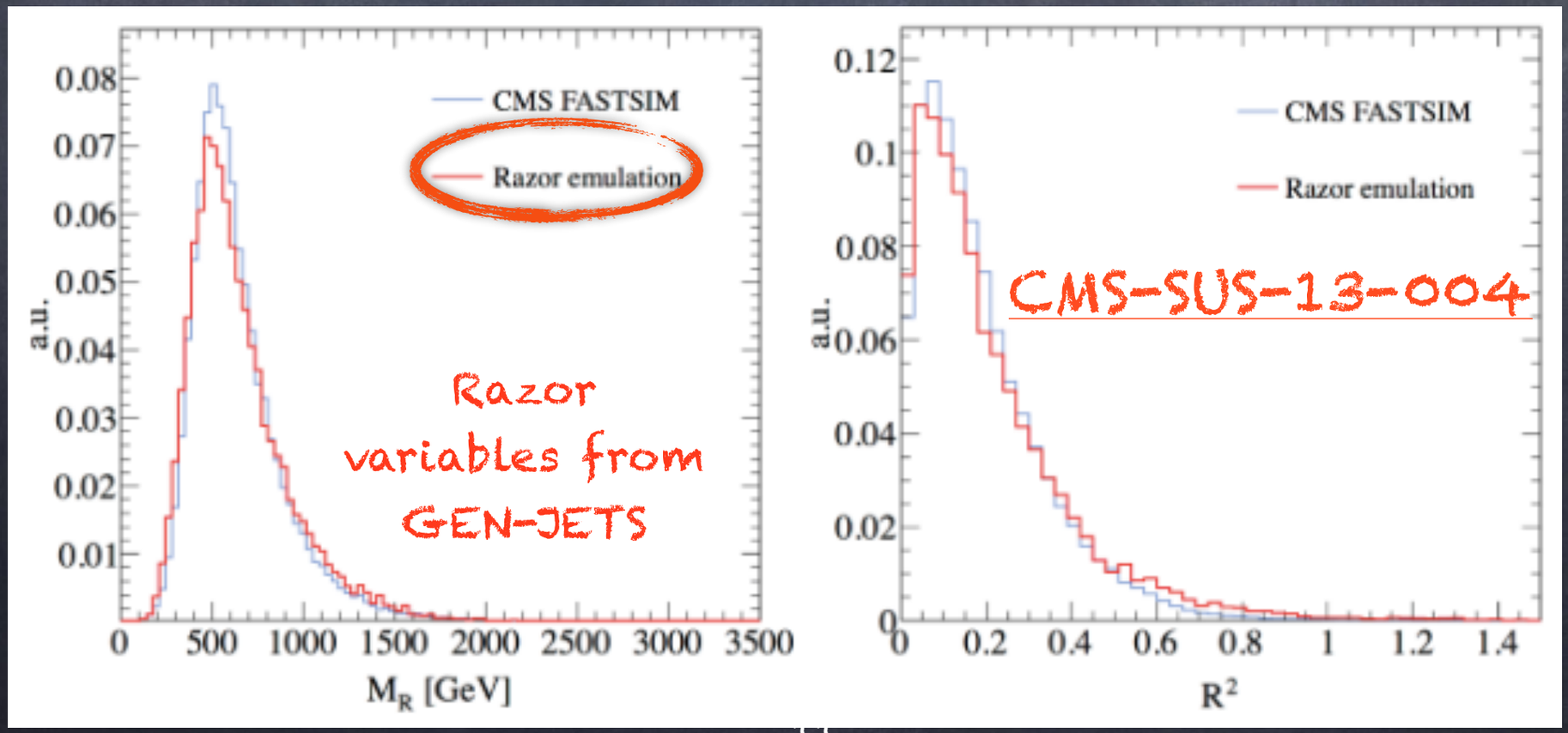


CMS-PAS-FTR-13-026



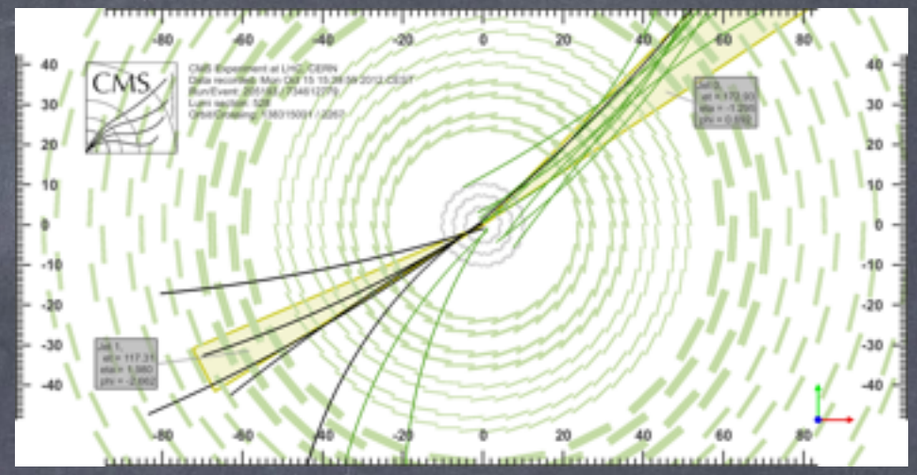
Are Detector Effects so Crucial?

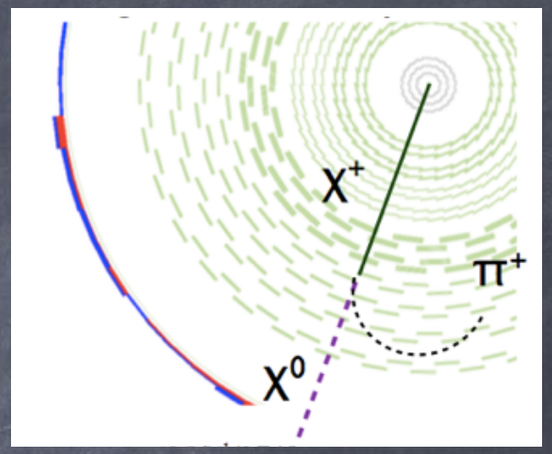
- Detector resolution is a limitation when dealing with clean signals (diphoton, dilepton, etc)
- When jets & MET involved, kinematic quantities have some "resolution" even @GEN level (e.g., jets vs partons)
- Detector resolution becomes a perturbation
- Often, inaccuracy on detector resolution is the perturbation on the perturbation

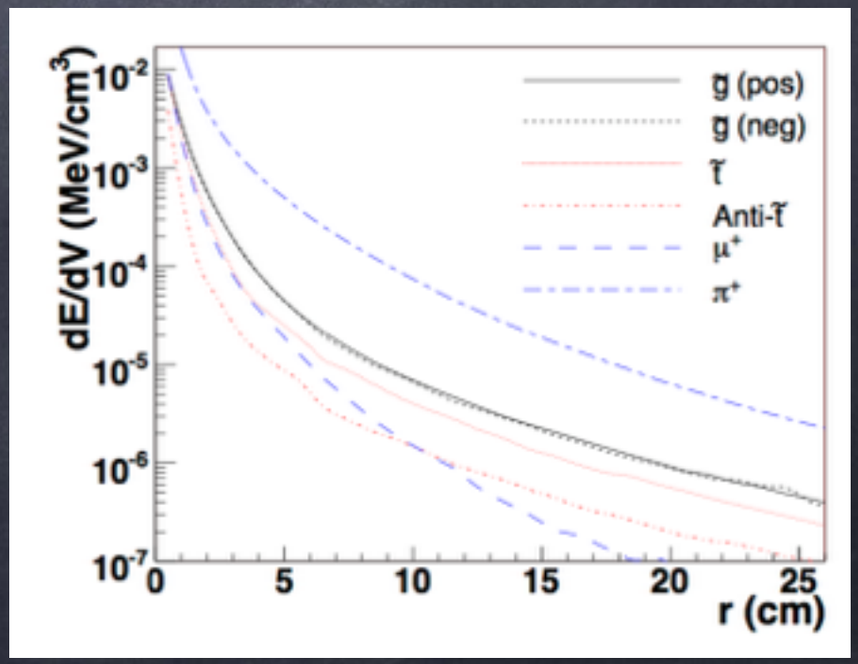


Difficult corners: exotic signatures

- What said above works OK for ballpark of searches
- Sometimes the signature is more complex and detector specific
  - long-living charged particles (dE/dx, TOF, disappearing tracks)
  - long-living neutral particles (displaced vertices)
- These signatures imply special workflows
  - For pheno: DELPHES out-of-the-box has troubles with these (e.g., no tracking running). But could be used with extra information
  - For experiments: workflow standardised with work of many (e.g. R-hadron package in GEANT from the paper by Mackeprang and Rizzi)

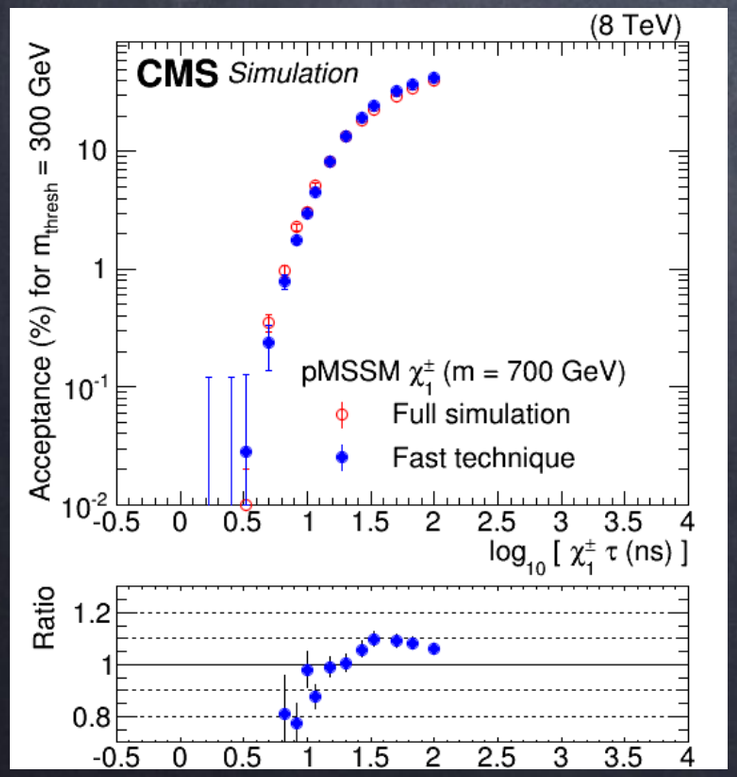


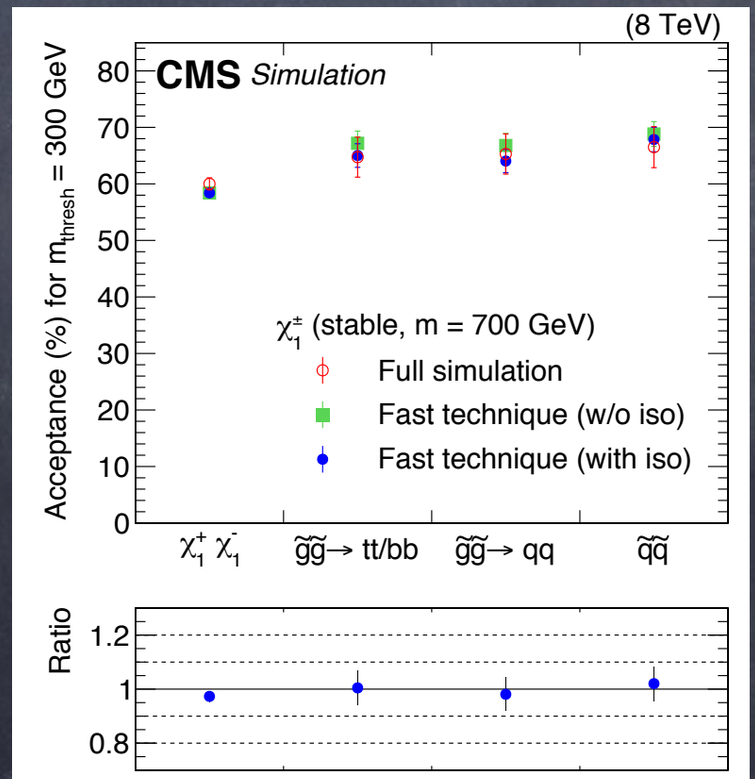


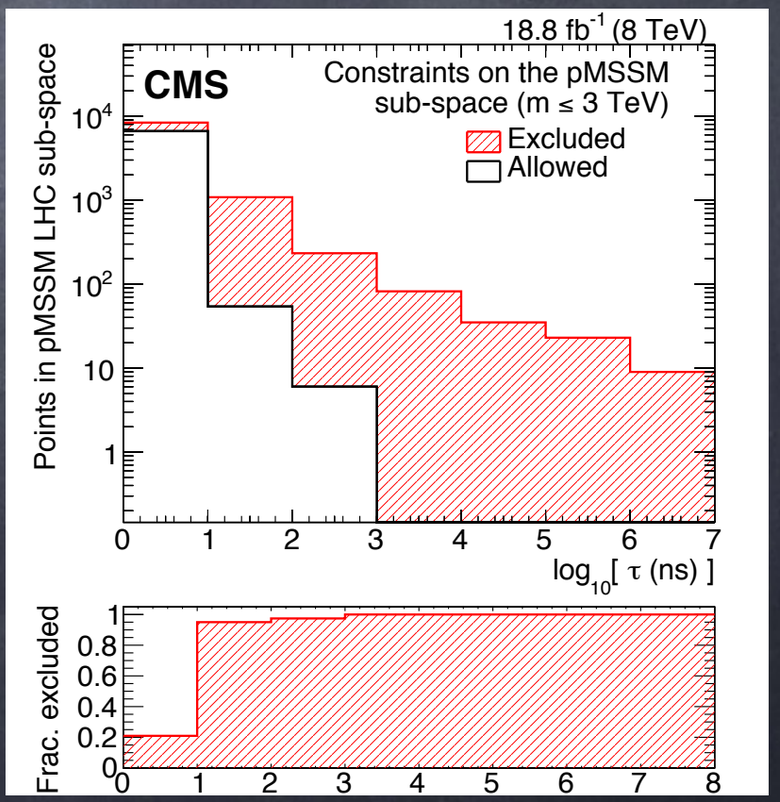


An analysis-specific shortcut

- EXAMPLE: pMSSM analysis in CMS
  - put generic bounds on SUSY from many SUSY searches, using CMS FASTSIM
  - identify parameter space to which searches are blind
  - long-living LSP happened to be in the list of surviving signatures
  - As DELPHES, FASTSIM has no dE/dx information
- The dE/dx response is well controlled in data
- Adding this as an external parameterization, FASTSIM could be used
- Result is not FASTSIM specific, but parameterisation is analysis specific



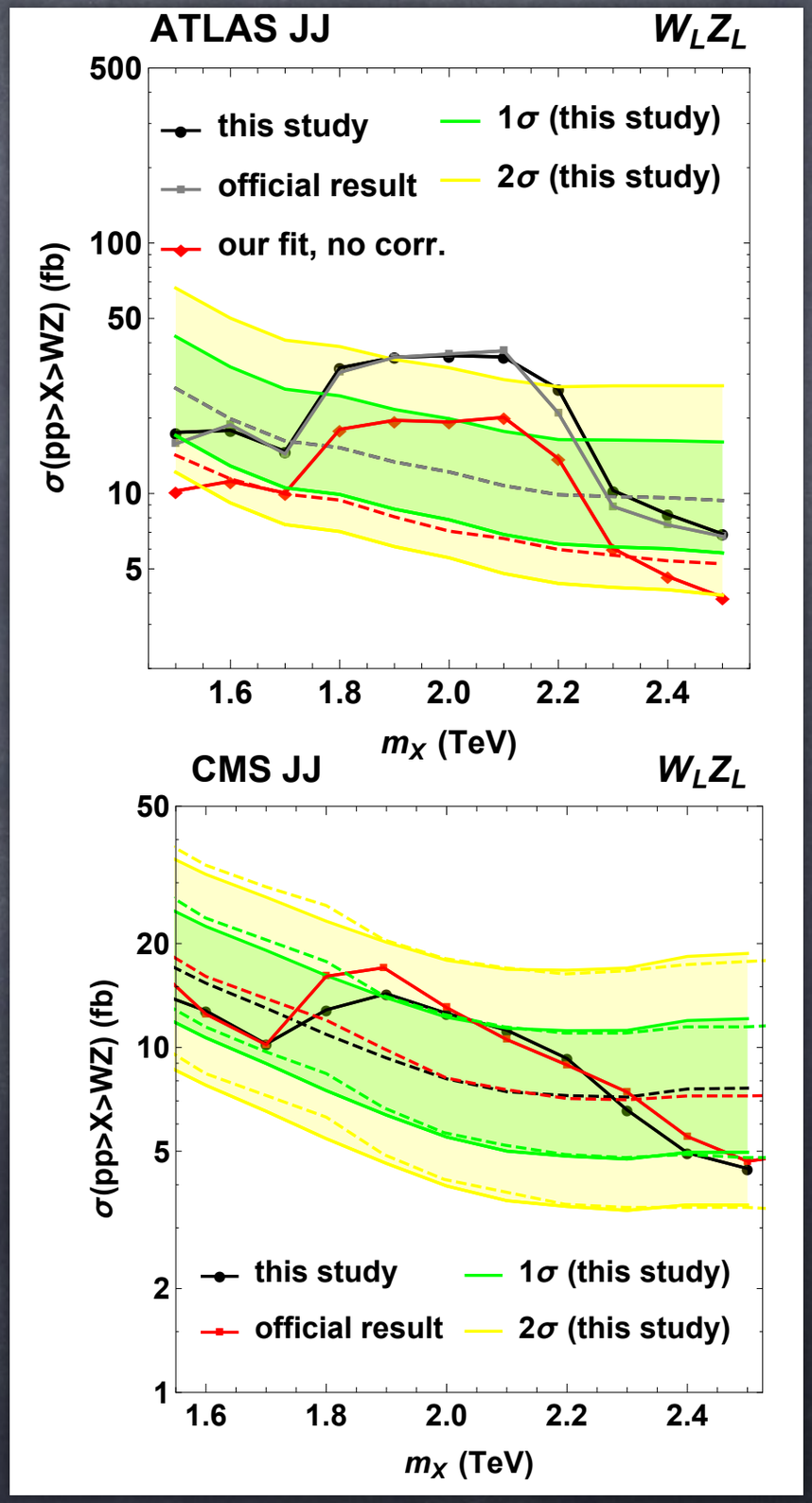




EXO-13-006

A deeper problem: reproducibility

- Diboson combination from ATLAS&CMS public results
  - use bkg estimate provided
  - use benchmark signal distribution provided
  - use signal efficiency provided
- Event Generation not even needed for this
- Agreement can go from good to bad, depending on the analysis

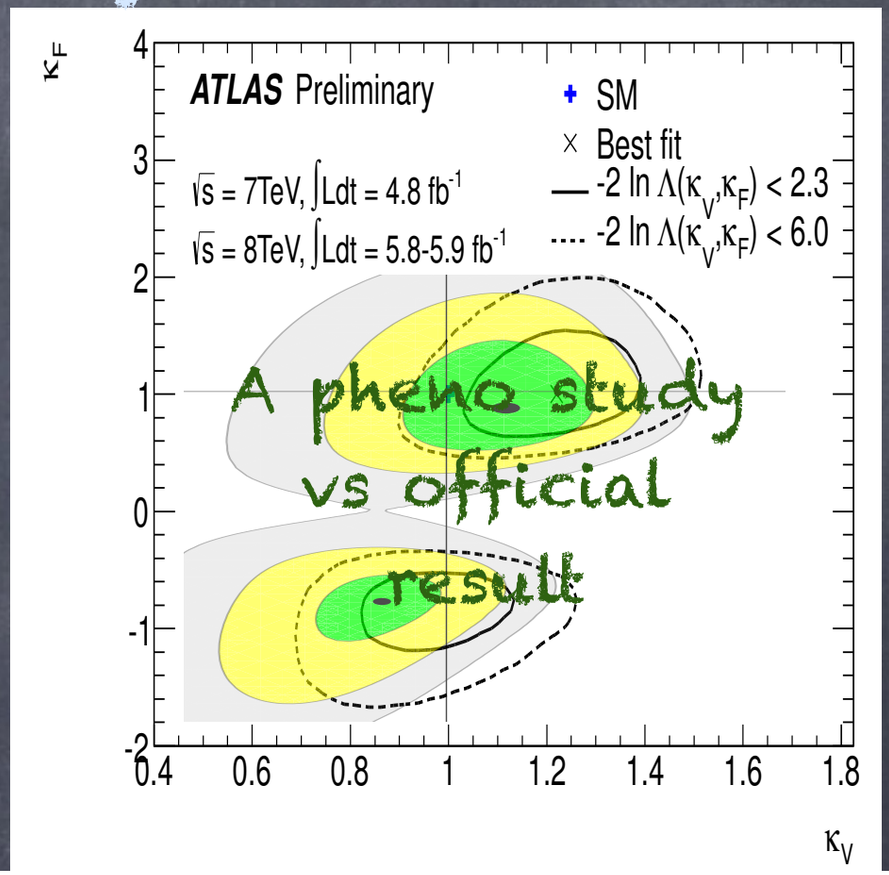


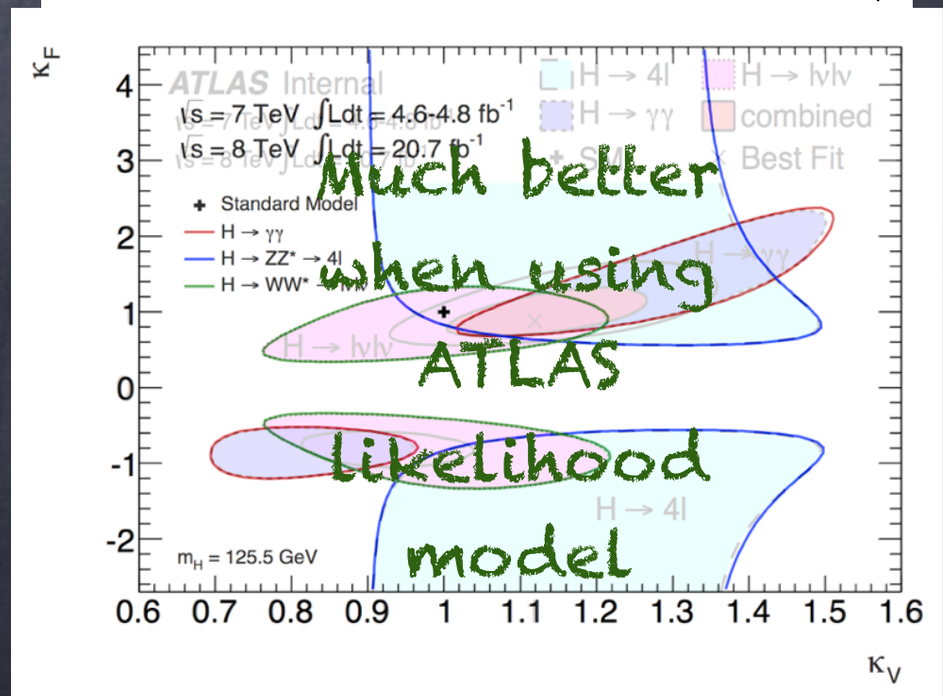
http://arxiv.org/abs/1512.03371

What about correlations/systematics?

- BSM searches are becoming more complex than cut&count
- They involve fits to signal regions +sidebands, multiple datasets, etc
- These fits include
  - correlations among different measurements (e.g. common detector effects)
  - systematics as nuisance parameters
- **Not just a problem for searches: see the Higgs couplings**

Talk by K. Cranmer DS@LHC





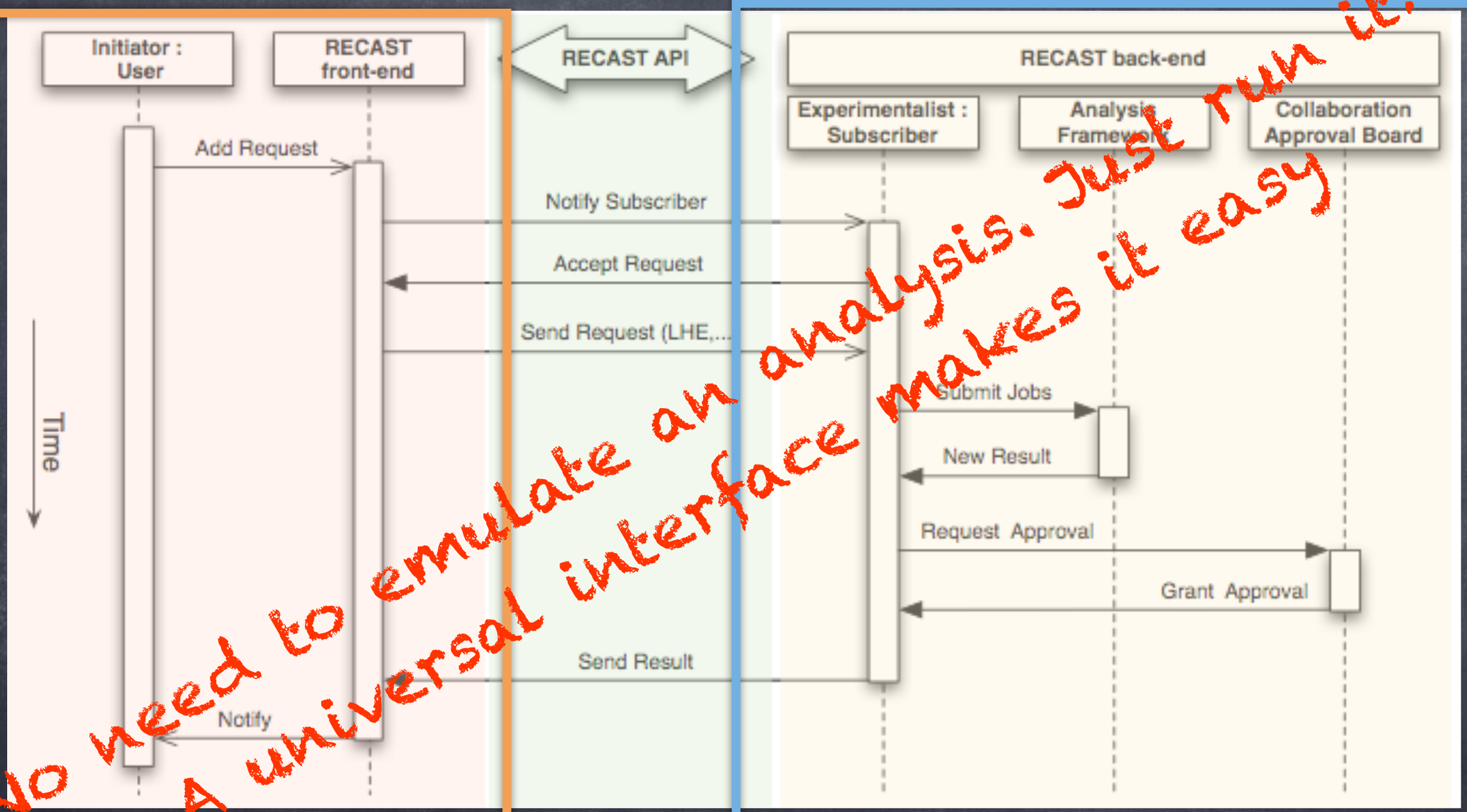
What about correlations/systematics?

- The ultimate information is in the Likelihood model
  - including correlation model for systematics, usually coming with some degree of arbitrariness
- **Having the likelihood, more than detector effects, would be IMPORTANT**
- Experiments are reluctant in releasing likelihoods
  - they could share statistics tools + data cards
- Releasing the Likelihood is not really needed: **RECAST allows to make a full analysis usable**

A Step ahead: RECAST

- RECAST (K. Cranmer, L. Heinrich, et al.) was proposed to solve reproducibility/reinterpretation issues in a clean way
  - Running the official analysis is much better than having to emulate it
  - Allows to run "official code" without having it (i.e., compliant with ATLAS/CMS rules)
- Solves many problems in one step
  - no need for customised detector simulation
  - no need for guessed likelihood models
  - not attached to one experiment/detector
  - also work with pheno tools (RIVET, etc)
  - It helps making the workflow standard (and not related to ATLAS vs CMS vs RIVET vs ...)

RECAST: a win-win solution?



No need to emulate an analysis. Just run it!
A universal interface makes it easy

theorist submit the request providing the needed ingredients (i.e. theory to theorists)

Analysis runs within the experimental collaboration (i.e. experiment to the experimentalists)

RECAST development status

- Working prototype has a few real ATLAS analyses that can rerun full chain on new signal
  - (UFO ->) LHEF -> full simulation -> reconstruction -> event selection -> limit setting
  - used internally for some channels in ATLAS pMSSM scan
  - runs original analysis code as in original paper
  - runs on CERN open-stack, uses docker, full integration to ATLAS software, authentication, ...
- working on integrations with Cern Analysis Preservation framework
  - idea is CAP would preserve analysis ingredients, RECAST would provide reinterpretation service
  - generic infrastructure, generic wrapper for Rivet analyses + can wrap other pheno recasting tools: Checkmate, MadAnalysis, Atom, ...
  - can provide uniform interface for several recasting tools
  - can have pheno, fast sim, full sim versions for the same analysis
  - most code is generic, not tied to particular experiment

https://github.com/recast-hep

Conclusions

- For the Experiments:
  - use of LHE files makes easier to integrate generators in experimental frameworks
  - **A common repository** (of LHE files? Gridpacks? UFOs?) would be very beneficial (for analyses and phone studies). A new life for MCDB?
- For Pheno studies:
  - for plain signatures, fastsim MC codes exist
  - integrated with **official detector tuning** by ATLAS & CMS
  - BUT sometimes a "good-enough" fastsim is not enough (e.g. exotic signatures)
- The ultimate solution:
  - a system like **RECAST** would solve the problem of re-producing and re-interpreting results.
  - With a complete RECAST library and enough CPU resources all phone needs would be covered
- Our community should (in my opinion) push **in these directions** more