Reinterpretation and Analysis Preservation Infrastructure for LHC Experiments

L Heinrich USLUA Meeting

October 2017













What models are still viable?

How good is simplified model coverage?

What's the best strategy to test as many models as possible at the LHC?









Basic Tension in HEP

Theorists

NEW YORK UNIVERSITY

Experiments









Basic Tension in HEP

- do not have data
- even with data analysis too complex
- have many ideas for still viable / interesting models :)

Theorists

Experiments









Basic Tension in HEP

- do not have data
- even with data analysis too complex
- have many ideas for still viable / interesting models :)

Theorists

- have data
- know how to analyze it
- do not have resources to design dedicated analysis for each model :(

Experiments

Reinterpretation efficient way to resolve the tension:

- re-use existing analyses that have good sensitivity for new model no need for a new, dedicated analysis
- only need to simulate the new signal data/background estimates unchanged
- not optimal, but we learn something about the model if no analysis can exclude it, good motivation for a dedicated search







store backgrounds and data you'll need them for final fit



store data pipeline to transport BSM signal model to a limit result

 only need final ntuples histos



store backgrounds and data you'll need them for final fit

store data pipeline to transport BSM signal model to a limit result

- only need final ntuples histos
- well known, analysisindependent process for MC production
- can scale complexity / accuracy as needed:
 - (very) fast / parametrized simulation O(s) / event
 - traditional fast sim
 - full Geant sim
 - future: GANs?



• only need final ntuples histos

you'll need them for final fit

- well known, analysisindependent process for MC production
- can scale complexity / accuracy as needed:
 - (very) fast / parametrized simulation O(s) / event
 - traditional fast sim
 - full Geant sim
 - future: GANs?

- many frameworks
- one-off code
- distributed knowledge & development
- people move on



store backgrounds and data you'll need them for final fit

store data pipeline to transport BSM signal model to a limit result

- only need final ntuples histos
- well known, analysisindependent process for MC production
- can scale complexity / accuracy as needed:
 - (very) fast / parametrized simulation O(s) / event
 - traditional fast sim
 - full Geant sim
 - future: GANs?

- many frameworks
- one-off code
- distributed knowledge & development
- people move on

systematic analysis preservation enables reinterpretation

Analysis Preservation for re-use

Recent developments in industry make HEP analysis preservation feasible

- capture heterogeneity in industry-standard ^{kubernetes} openstack^{*} containers — sharable, archivable analysis software (-runtimes)
- for analysis re-execution use industry-grade container orchestration tools developed for industry cloud services to execute on cluster







Analysis Presevation for re-use

LHC-wide effort to enable analysis re-use

- CERN Analysis Preservation Portal (CAP):
 - central place for LHC experiments to store software, runtime environments, workflows, data needed to reexecution
- Analysis Re-execution Service (REANA)
 - provide interfaces to pull info from CAP and re-execute analysis in the cloud





Once preserved — re-use analyses for reinterpretation

- prepare pipeline inputs for new model (scan)
- run pipeline to get new limits



450

400

350

300

250

150

100

abcissa mStop

pointRequests

explode

ordinate mNeutralino \$

colormap ClsHeat 🛊



Process All -

Mass Plane

Request Details

with the ATLAS detector

© 2017-08-13

Analysis 4

Reason

Additional Info

Requested Parameter Points

Parameter Point 1					
1 Basic Requests	mStop: 200.0	mNeutralino: 0.0	obs: 1.00	exp: 1.00	

Search for direct production of the top squark in the all-hadronic

ttbar + etmiss final state in 21 fb-1 of p-pcollisions at sqrt(s)=8 TeV

The grid is useful to validate / check third-party (such as CheckMate)

the grid consists of 36 points in the stop / neutralino mass plane. The

input parameters are given in the standard format consisting of

implementations of the ATLAS-CONF-2013-024 as it

model parameters and number of events

Basic Request 1: Format: standard_format Process - 📰 Show Processings Results - 🕑 Upload to RECAS

Current Status

- system was developed during Run-1 ATLAS pMSSM paper
- prototype analysis Run-1 SUSY electroweak 2L production



now preserving to Run-2 analyses — reinterpretations underway





It's the difference between if you had airplanes where you threw away an airplane after every flight, versus if you could reuse them multiple times.

– Elon Musk

HEP analysis HEP analyses testing it against

– Elon Musk



 reinterpretation needed to fully exploit the LHC dataset

- preservation enables fullfidelity reinterpretation within the experiments
- recent IT advances make preservation of *original* HEP analyses feasible