

Simple Analysis for LLPs

– Truth-level Analysis Framework

Yvonne Ng @ DESY

2023-6-20

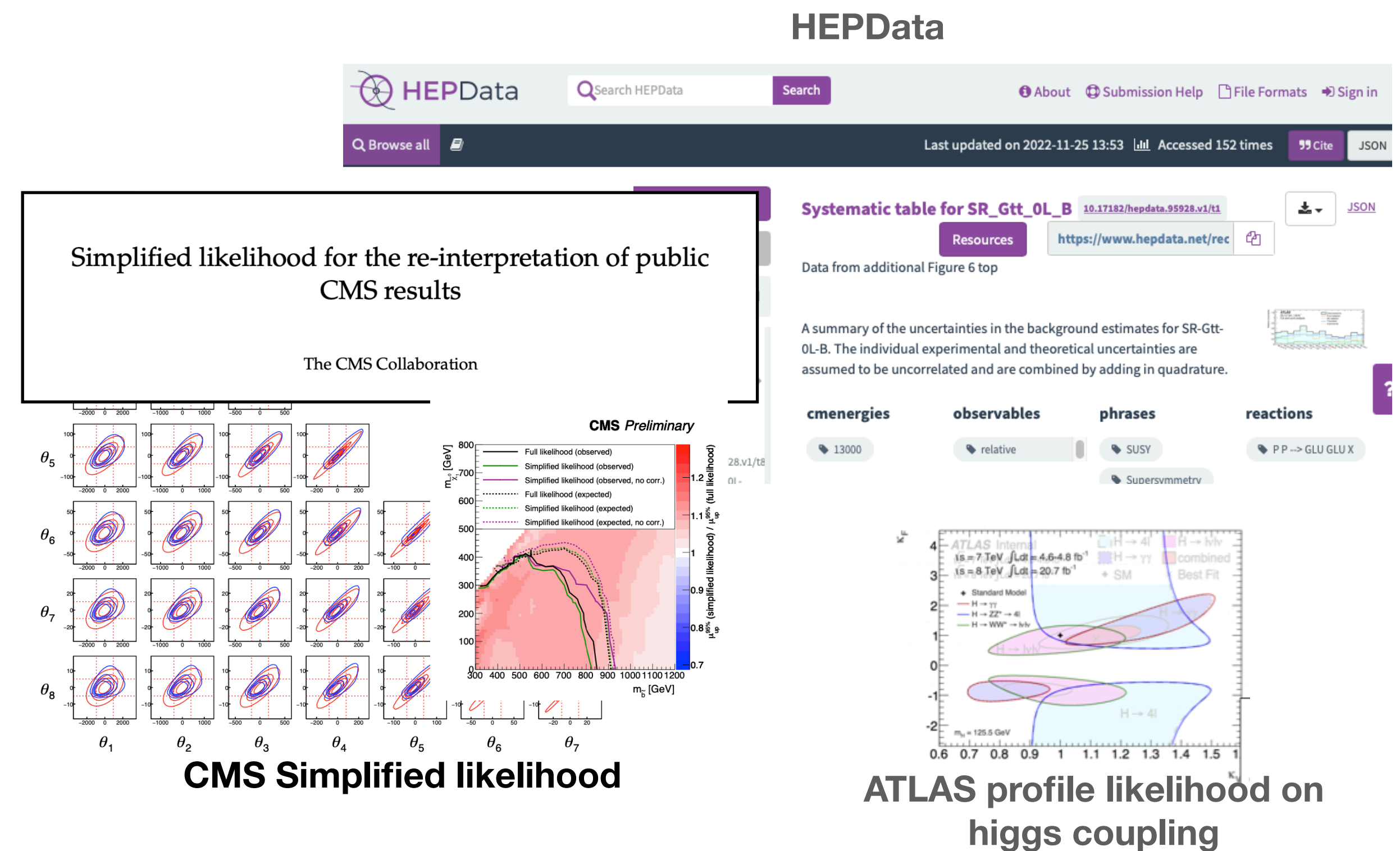
LLP13



A timeline on Reinterpretation

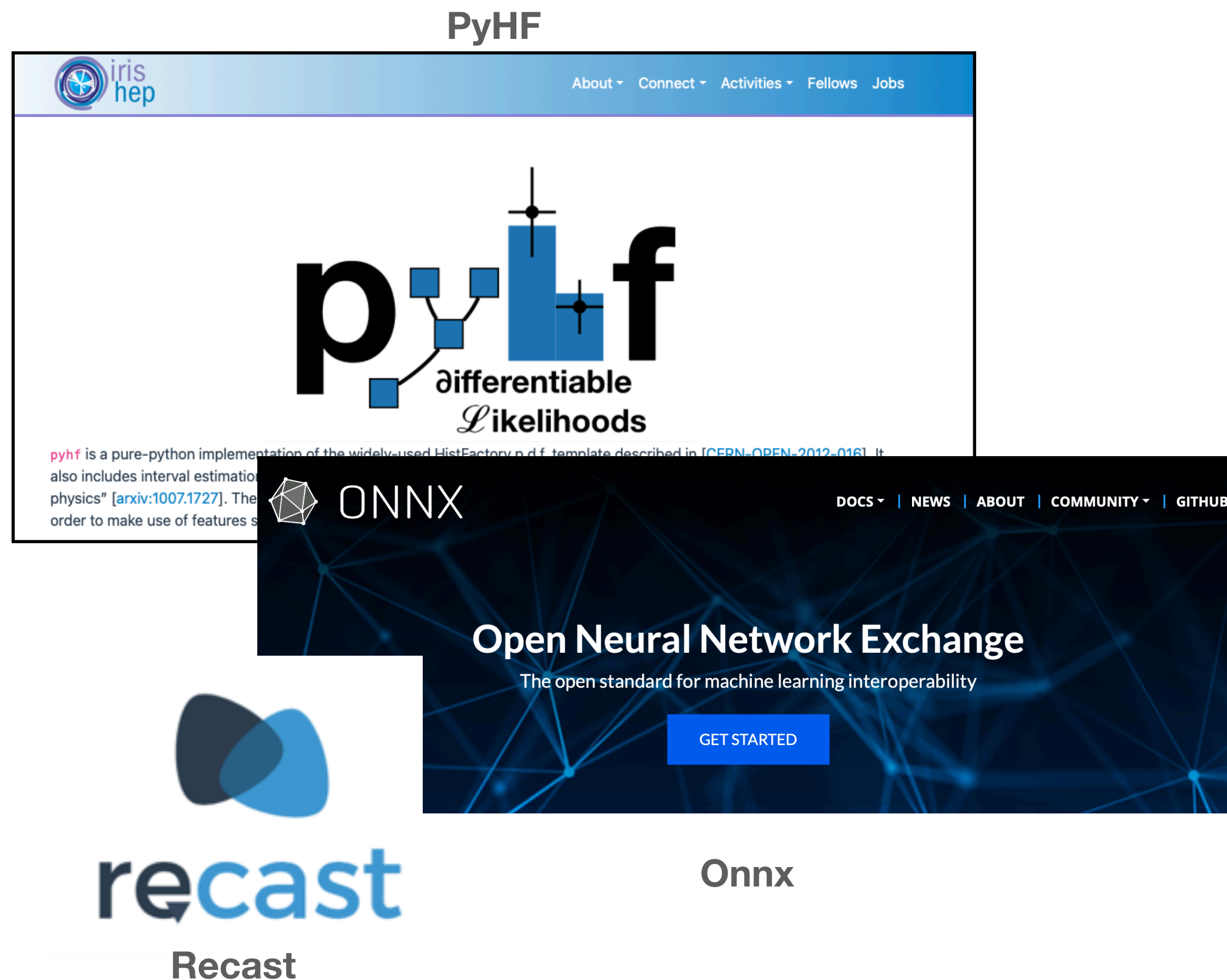
– Development of Tools and the Progression of Collaboration Standard

- 2000: 1st PHYSTAT
- 2010: introduction of Workspaces
- 2010: First proposition of Recast
- 2012: ATLAS profile likelihoods scans released for Higgs coupling
- 2017: CMS Simplified Likelihoods
- 2018: Release of analysis preservation and systematic reinterpretation with Recast within ATLAS



A timeline on Reinterpretation

– Development of Tools and the Progression of Collaboration Standard



- 2019: First full likelihood release in ATLAS for SUSY searches
 - Using json schema of Histfactory class of likelihood
- 2019: ATLAS publication of full statistical likelihoods PUB note released (JSON HF)
- 2020: First full likelihood release regularly in ATLAS for SUSY searches
- 2022: “[Simple Analysis: Generator-level Analysis Framework](#)” pub note released— Framework going public!
- 2022: Snowmass: “Reinterpretation and Long-Term Preservation of Data and Code”

*ATLAS development highlighted in blue

Different Reinterpretation Tools

Less Precision

Level of Complexity

More information/Precision

Experiment specific

Simple Likelihood

Reinterpretation from likelihood

$$p(\text{theory}|\text{data}) \sim p(\text{data}|\text{theory}) * p(\text{theory})$$

New Result!

Theoretical Prior of New Model

[Simple Likelihood Pubnote](#)

Existing Exp. Likelihood (What We Provide)

6

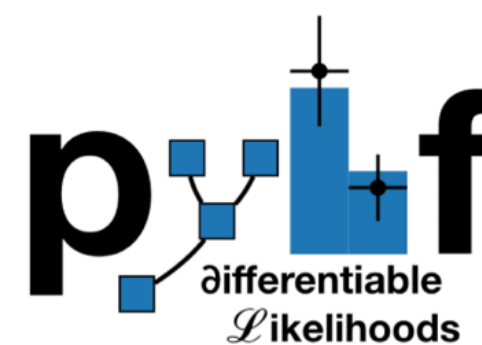
HEPData:

Storing data products from analyses for reinterpretation

HistFactory Json: Storing data products from analyses for reinterpretation



About · Connect · Activities · Fellows · Jobs



pyhf is a pure-python implementation of the widely-used HistFactory p.d.f. template described in [CERN-OPEN-2012-016]. It also includes interval estimation based on the asymptotic formulas of "Asymptotic formulae for likelihood-based tests of new physics" [arxiv:1007.1727]. The aim is also to support modern computational graph libraries such as PyTorch and TensorFlow in order to make use of features such as autodifferentiation and GPU acceleration.

4

Recast:
Storing data products from analyses for reinterpretation

preservation of the original analysis pipeline for collaboration-approvable results

Simple Analysis

– Experiment-independent middle

- Preserve **full analysis pipeline** for manipulation/maximal information
- **Open source**
 - Experiment non-specific, not tie to detector reconstruction
 - Approximation can be added
- Allow for **reinterpretation up to the truth level**
 - Detector effect can be added
- Non-experiment input format
 - **HEPMC, DAOD, ROOT-ntuple, Delphes** etc.
- For Public use: can be used outside of ATLAS
- Additional information for reinterpretation

SimpleAnalysis, v1.1.0

Search

SimpleAnalysis, v1.1.0

Home

ATLAS Internal

Analysis List

N-Tuple Structure

Tutorial

N-Tuple Structure >

Simplified ATLAS SUSY analysis framework

Holds a collections of SimpleAnalysis in different formats:

- DAOD_TRUTH (TRU)
- xAOD (either truth-le)
- HepMC (uncompressed)
- slimmed ntuples (Re)
- DELPHES (Converted)

It provides the analysis of histograms or ntuples with

Analysis List

Analysis Name	ATLAS Public Result page
ZeroLeptonJigsaw2016	ANA-SUSY-2016-07
SbottomMultiB2018	ANA-SUSY-2018-31
DMbb2016	ANA-SUSY-2016-18
DMttZeroLepton2016	ANA-SUSY-2016-18
MultiJets2018	ANA-SUSY-2018-17
EwkThreeLeptonResonance2018	ANA-SUSY-2018-36
EwkThreeLeptonERJR2018	ANA-SUSY-2018-06
PairedDijets2016	ANA-SUSY-2016-09
EwkTwoLeptonTwoJet2016	ANA-SUSY-2016-24
DirectStau2018	ANA-SUSY-2018-04
OneLeptonMultiJets2016	ANA-SUSY-2016-11
DisappearingTrack2016	ANA-SUSY-2016-06
StopZ2016	ANA-SUSY-2016-20
EwkTwoLeptonRJ2016	ANA-SUSY-2017-03

SimpleAnalysis: [Webpage](#)



[SimpleAnalysis PubNote](#)

Available list of Analyses

Simple Analysis Framework

Generator input format

DAOD
(ATLAS format)

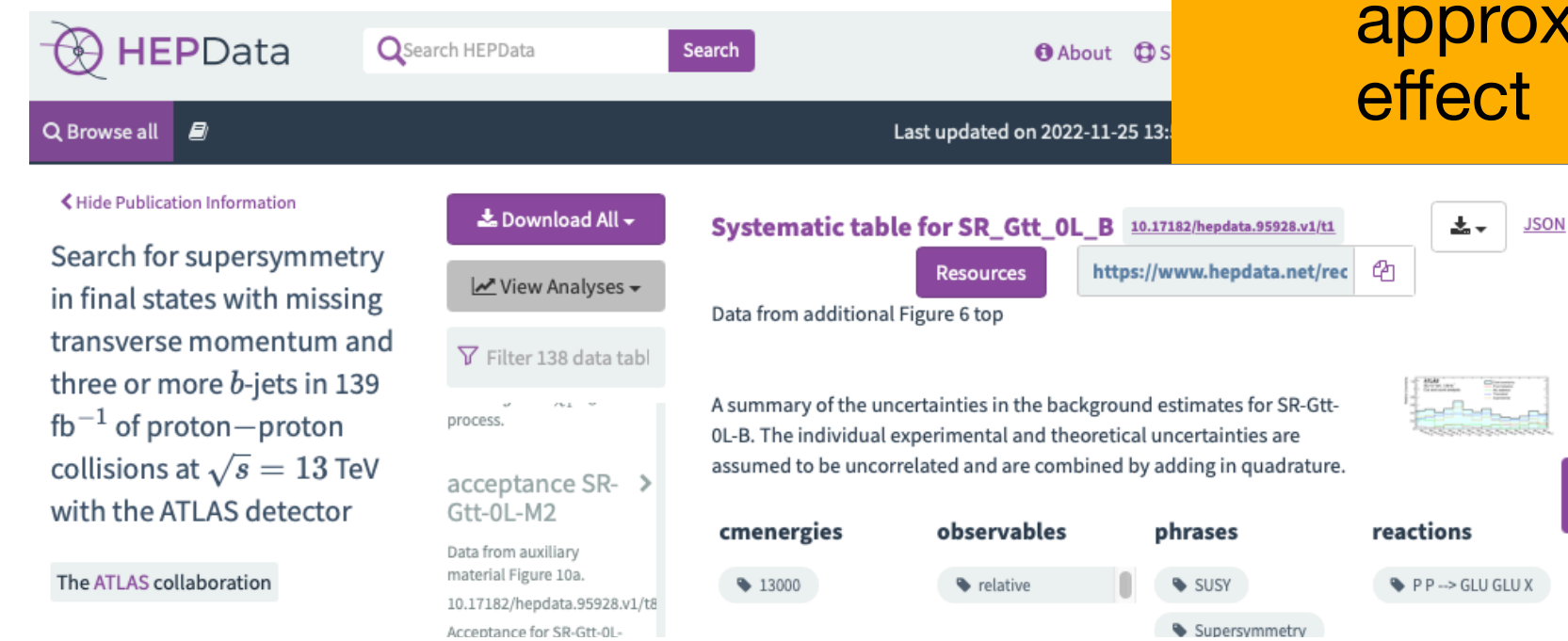
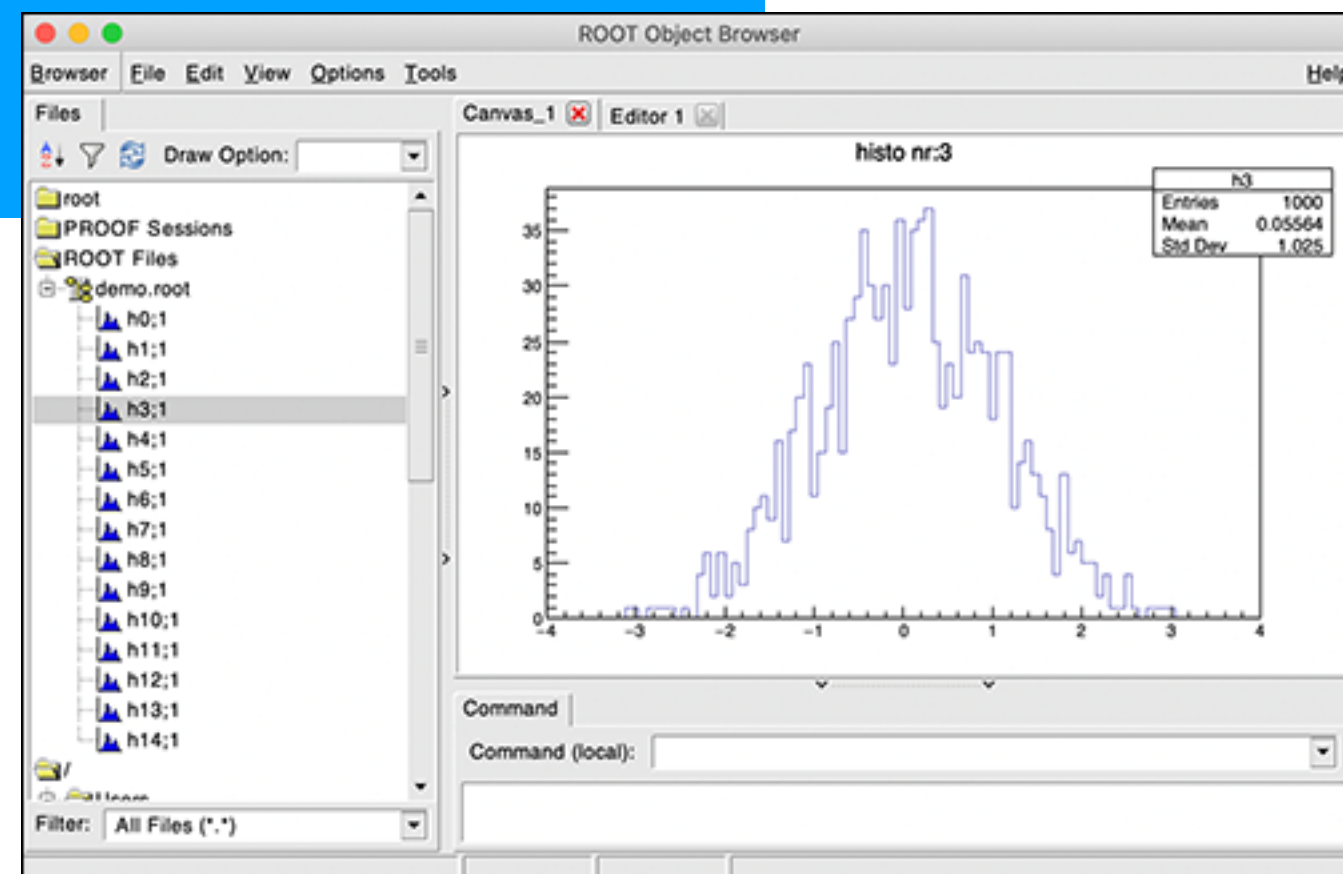
HEPMC
ROOT
DELPHES

Simple Analysis Routine

Complex kinematic variables
Custom efficiency maps
Additional utilities for serializing NN
and multi-variate analysis /BDT

HEP data publishing

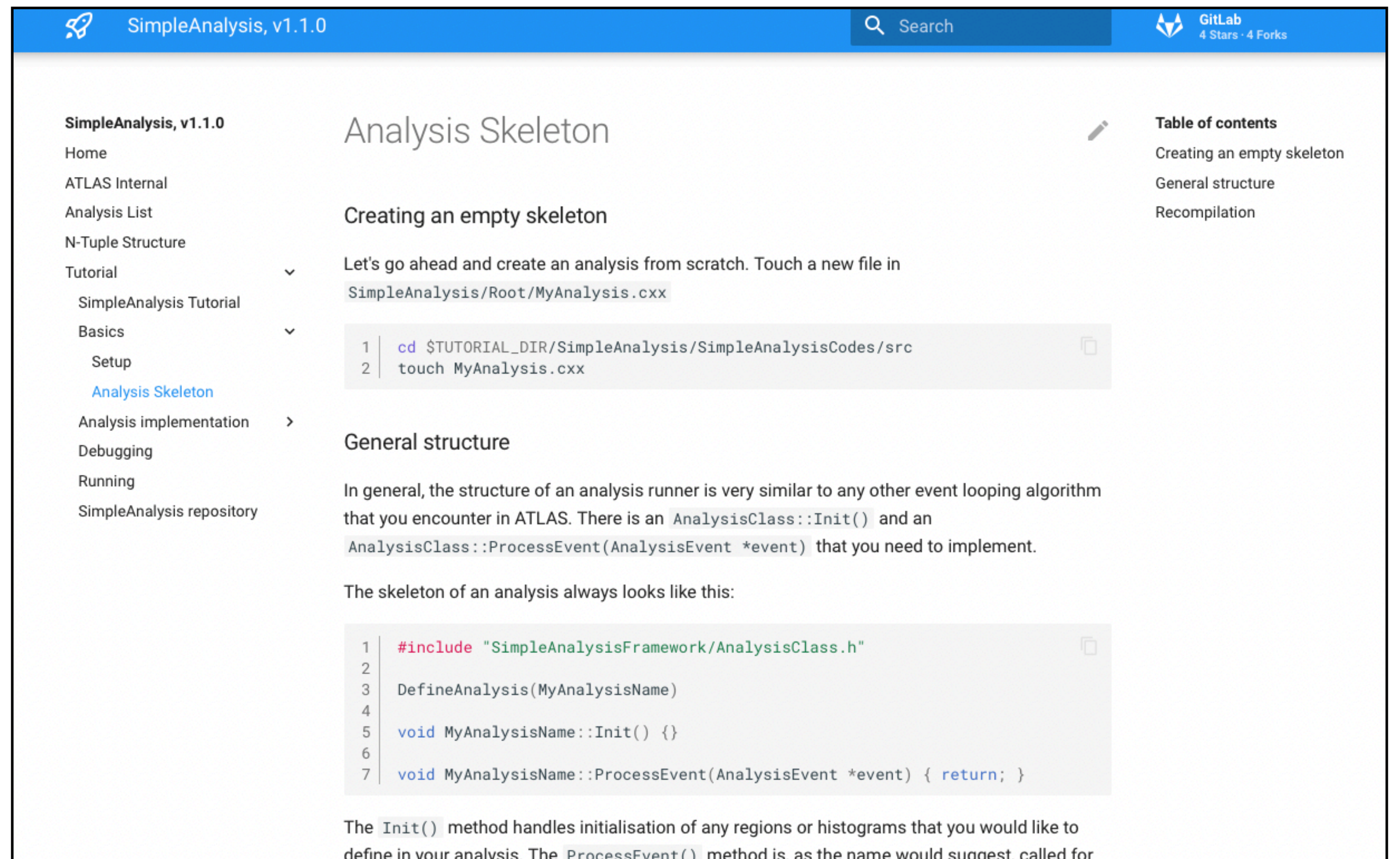
- Truth level acceptance and selection efficiencies for HEPData publishing
- Root flat-ntuple files containing histograms
- Smearing available to approximate detector effect



Simple Analysis

– Tutorial

- Tutorial contains:
 - Building analysis selection
 - Analysis pipeline Preservation
 - Reinterpreting results from previous SA pipelines.
 - Reinterpretation Effort
 - Welcome feedback!



SimpleAnalysis, v1.1.0

Search

GitLab
4 Stars · 4 Forks

SimpleAnalysis, v1.1.0

- Home
- ATLAS Internal
- Analysis List
- N-Tuple Structure
- Tutorial
 - SimpleAnalysis Tutorial
 - Basics
 - Setup
 - Analysis Skeleton
 - Analysis implementation
 - Debugging
 - Running
 - SimpleAnalysis repository

Analysis Skeleton

Creating an empty skeleton

Let's go ahead and create an analysis from scratch. Touch a new file in `SimpleAnalysis/Root/MyAnalysis.cxx`

```
1 | cd $TUTORIAL_DIR/SimpleAnalysis/SimpleAnalysisCodes/src
2 | touch MyAnalysis.cxx
```

General structure

In general, the structure of an analysis runner is very similar to any other event looping algorithm that you encounter in ATLAS. There is an `AnalysisClass::Init()` and an `AnalysisClass::ProcessEvent(AnalysisEvent *event)` that you need to implement.

The skeleton of an analysis always looks like this:

```
1 | #include "SimpleAnalysisFramework/AnalysisClass.h"
2 |
3 | DefineAnalysis(MyAnalysisName)
4 |
5 | void MyAnalysisName::Init() {}
6 |
7 | void MyAnalysisName::ProcessEvent(AnalysisEvent *event) { return; }
```

The `Init()` method handles initialisation of any regions or histograms that you would like to define in your analysis. The `ProcessEvent()` method is, as the name would suggest, called for

Table of contents

- Creating an empty skeleton
- General structure
- Recompilation

Simple Analysis

– Current analysis list

- Most SUSY analyses on ATLAS published analysis selection at the level of generation
- Growing list can be found [here](#).

Analysis List

The screenshot shows the SimpleAnalysis v1.1.0 website. The main heading is "Simplified ATLAS SUSY analysis framework". On the left, there is a navigation menu with "Analysis List" selected. The main content area lists various analysis formats: DAOD_TRUTH (TRU), xAOD (either truth-le), HepMC (uncompressed), slimmed ntuples (Re), and DELPHES (Converted). Below this, it states "It provides the analysis a histograms or ntuples w".

SimpleAnalysis: [Webpage](#)



Source Code

[SimpleAnalysis PubNote](#)

Analysis Name	ATLAS Public Result page
ZeroLeptonRjigsaw2016	ANA-SUSY-2016-07
SbottomMultiB2018	ANA-SUSY-2018-31
DMbb2016	ANA-SUSY-2016-18
DMttZeroLepton2016	ANA-SUSY-2016-18
MultiJets2018	ANA-SUSY-2018-17
EwkThreeLeptonResonance2018	ANA-SUSY-2018-36
EwkThreeLeptonERJR2018	ANA-SUSY-2018-06
PairedDijets2016	ANA-SUSY-2016-09
EwkTwoLeptonTwoJet2016	ANA-SUSY-2016-24
DirectStau2018	ANA-SUSY-2018-04
OneLeptonMultiJets2016	ANA-SUSY-2016-11
DisappearingTrack2016	ANA-SUSY-2016-06
StopZ2016	ANA-SUSY-2016-20
EwkTwoLeptonRJ2016	ANA-SUSY-2017-03

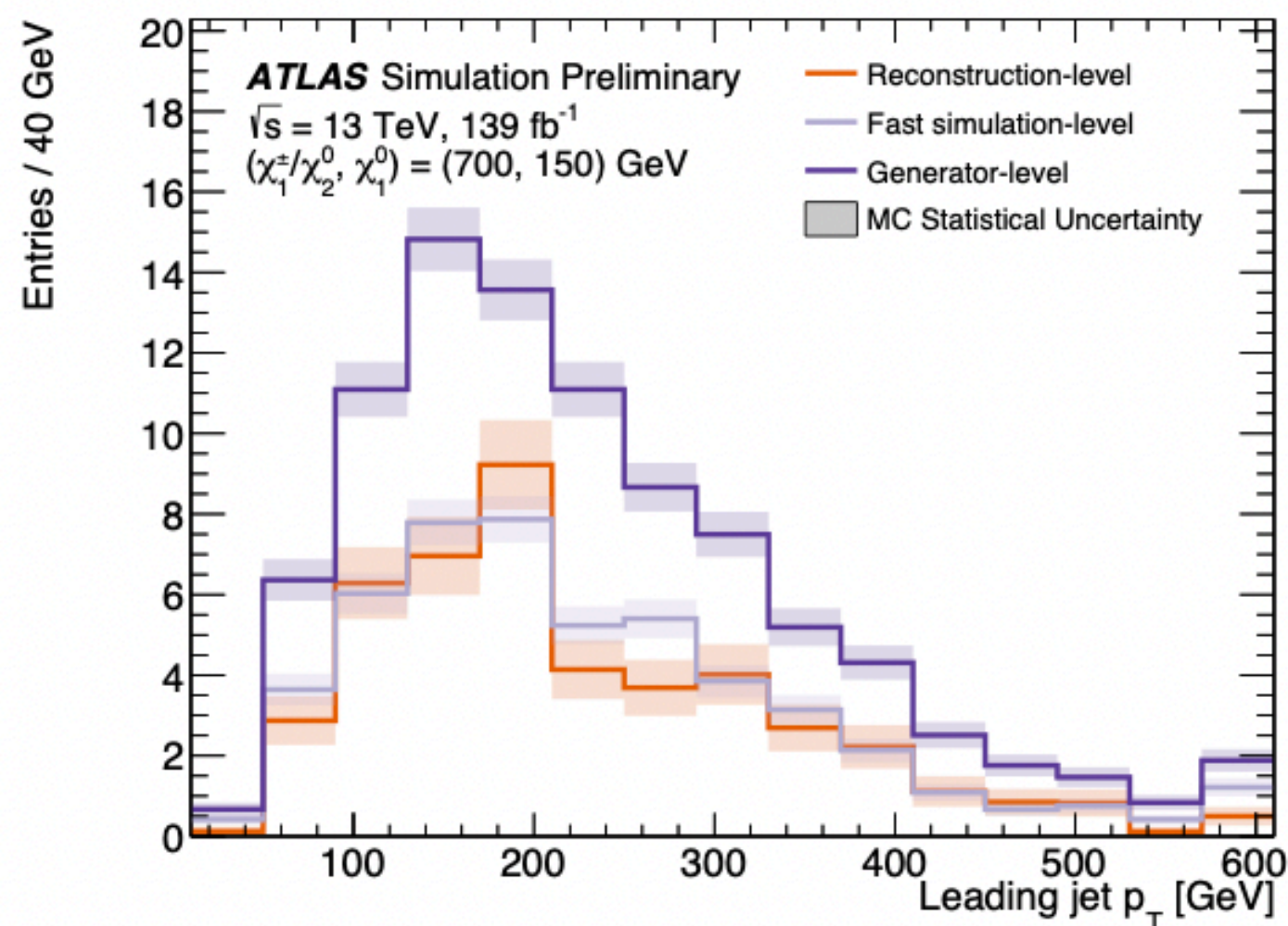
Available list of Analyses

EwkTwoLeptonRJ2016	ANA-SUSY-2017-03
------------------------------------	----------------------------------

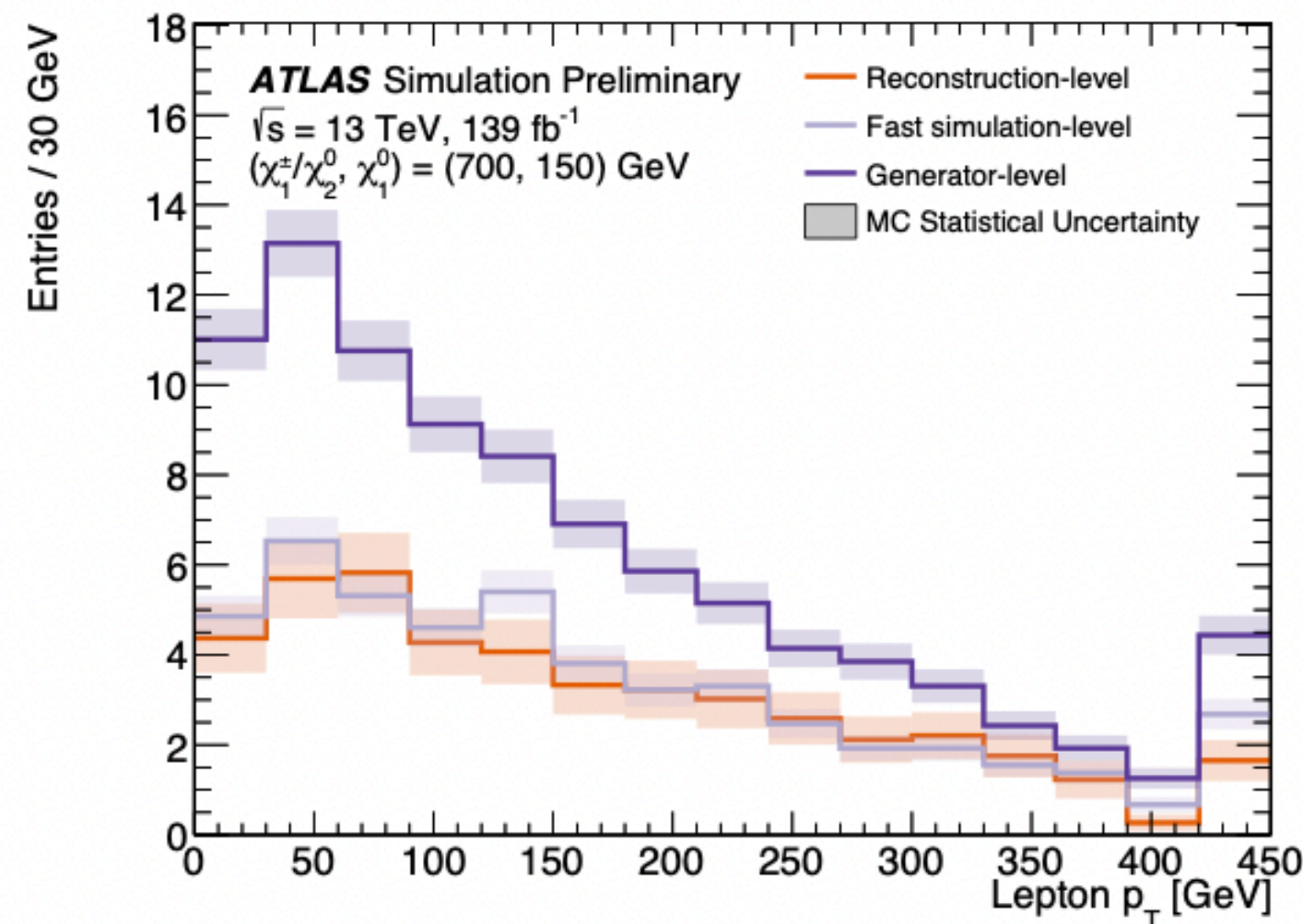
Simple Analysis

– Pub note

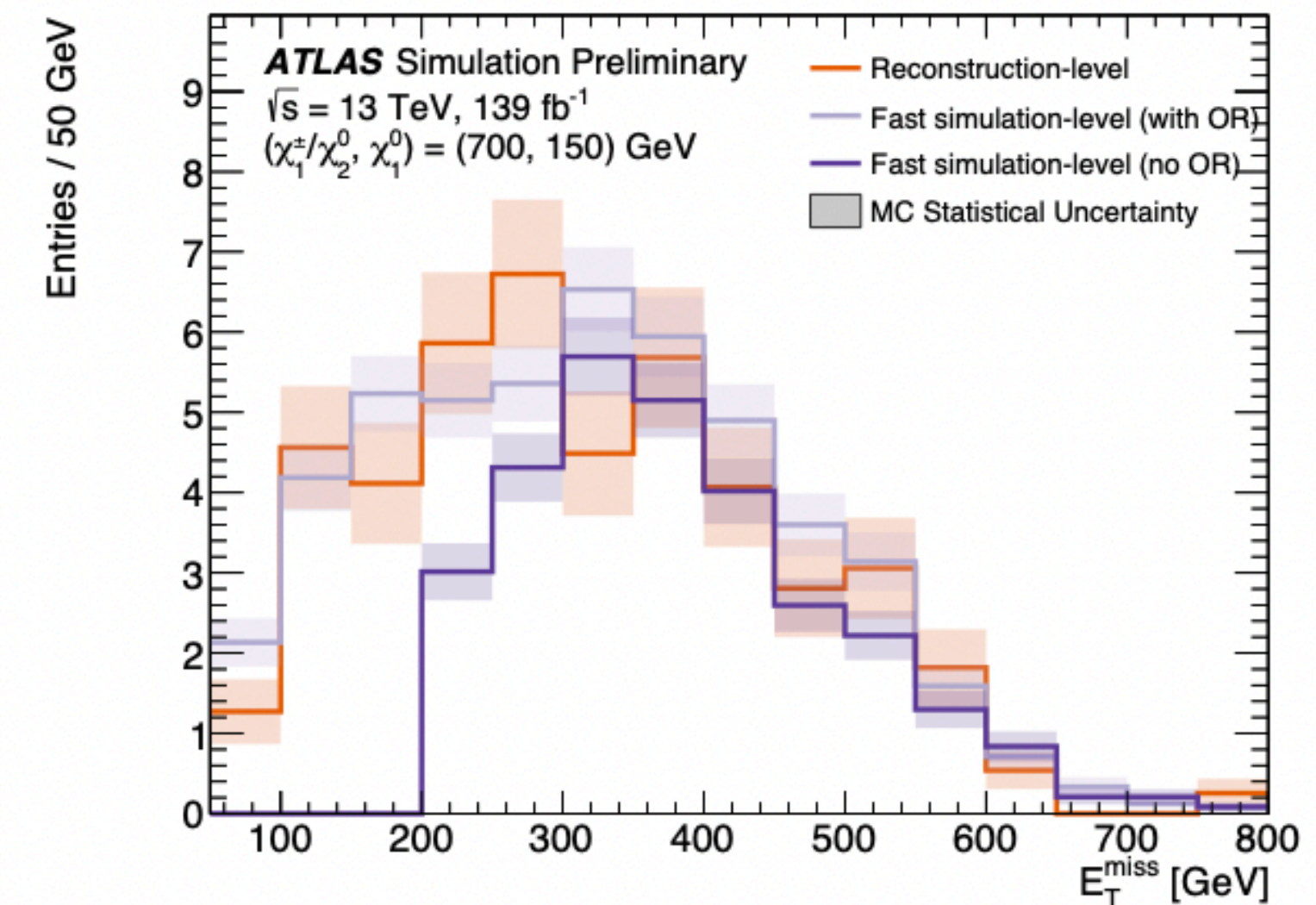
- [Pub note](#) described the tool and analysis pipelines for preservation
- Comparing the generator level analysis by SimpleAnalysis with the Monte Carlo samples at reconstruction level
- One lepton+ two b-jet analysis is used to demonstrate the validation



(a) Leading jet p_T



(b) Leading lepton p_T



(f) Missing transverse momentum

LLP analysis in Simple Analysis

- Currently few LLP analyses are available in SA
- Disappearing track
- ..More to come!
- R22 is [work in progress](#) that could use more person input

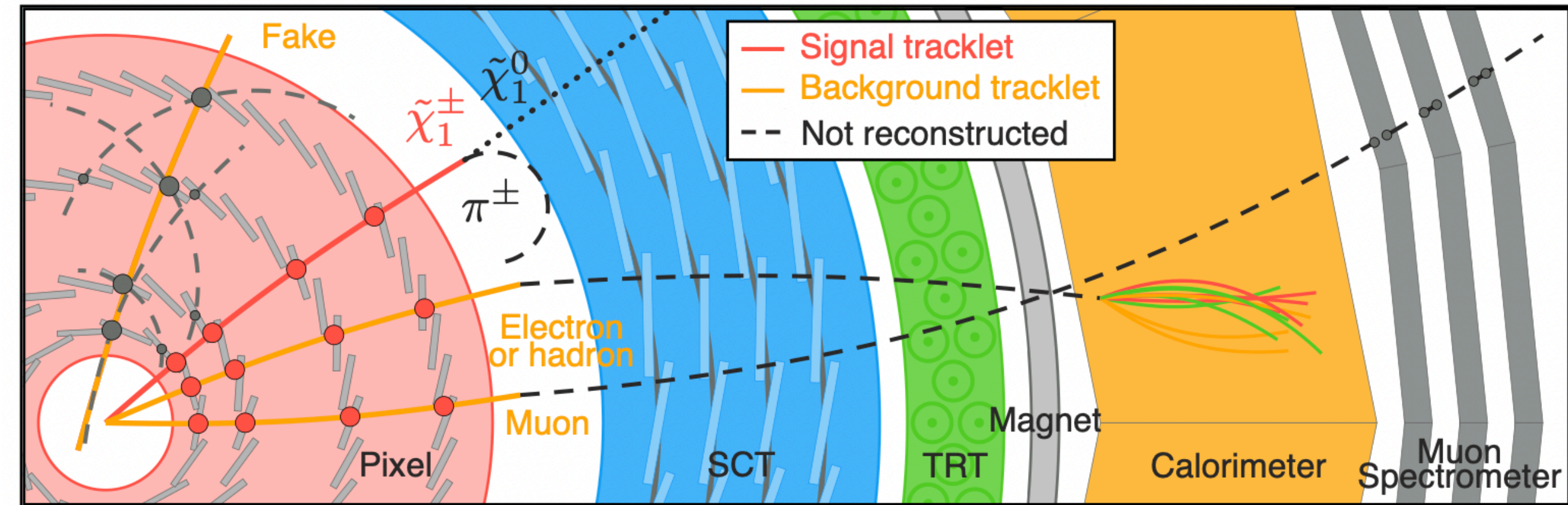
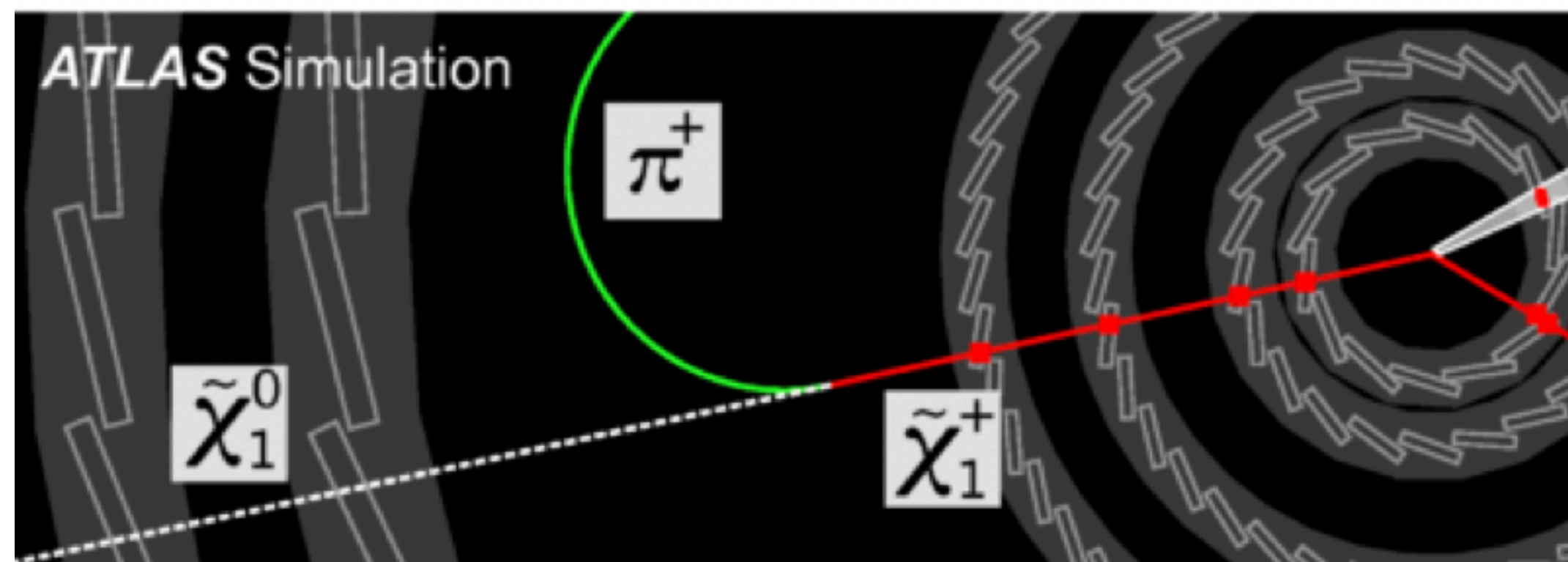


Image from [Toshiaki Kaji](#)





Future of Simple Analysis

-A full reusing pipeline(Mapyde)

- Fresh off the arrive today at 9am

REDUCE, REUSE, REINTERPRET: AN END-TO-END PIPELINE FOR
RECYCLING PARTICLE PHYSICS RESULTS

A PREPRINT

 Giordon Stark^{*1}, Camila Aristimuno Ots^{1,2}, and  Mike Hance^{†,1}

¹University of California, Santa Cruz, Santa Cruz Institute for Particle Physics, 1156 High Street, Santa Cruz, CA 95064

²University of Southern California, Liquid Propulsion Laboratory, 854 Downey Way, Los Angeles, CA 90089

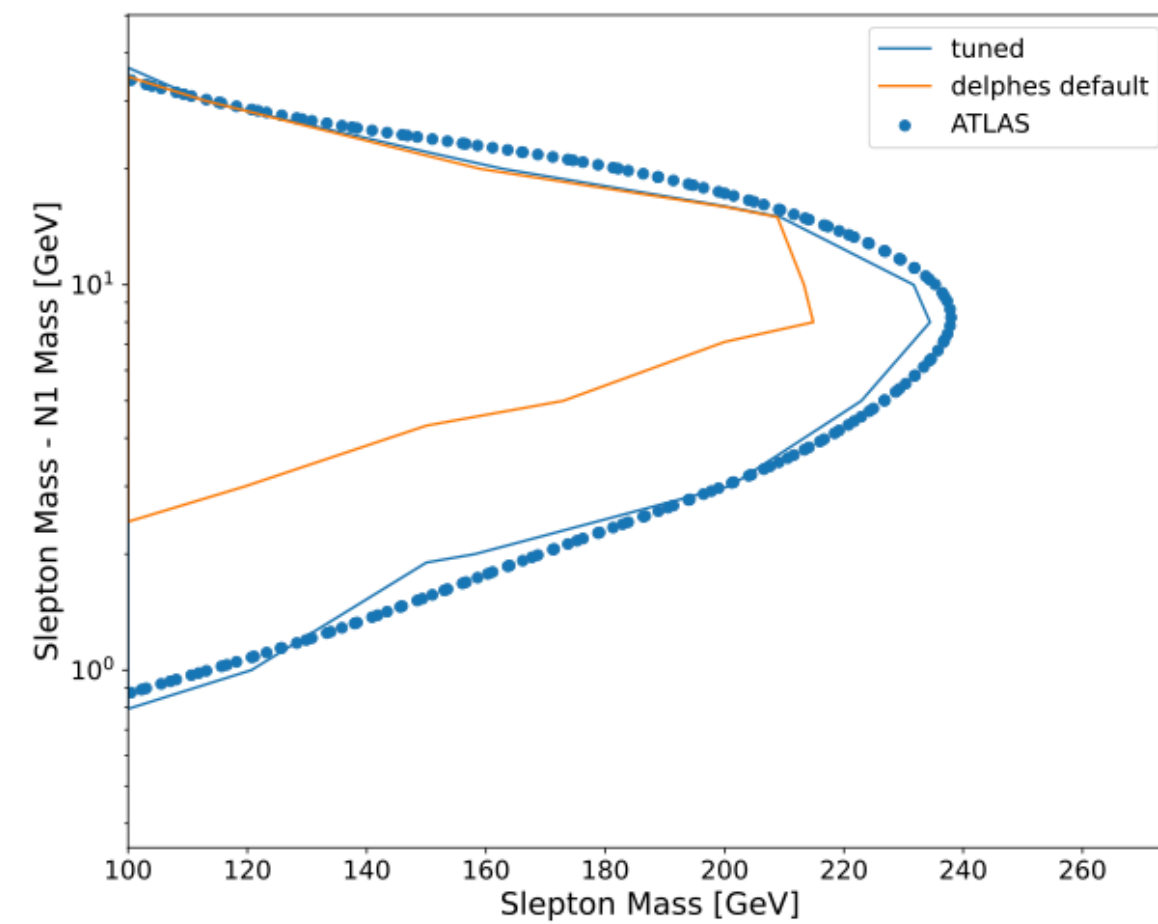
Mapyde

- Mayede: Performing a full pipeline of truth level analysis from event generation to statistical analysis!

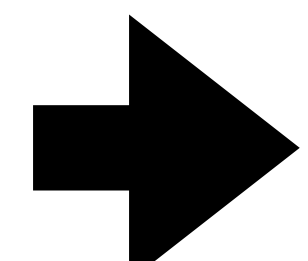
Containing

- MADGRAPH5_AMC@NLO(MADGRAPH) [26, 27] (event generation)
- PYTHIA8 [28] (parton shower, hadronization, decays)
- DELPHES [29–31] (detector simulation)
- SimpleAnalysis [32, 33] (analysis description)
- pyhf [34, 35] (probability model fitting)

Reinterpretation Results from Mapyde



ATLAS Slepton result



Wino-bino reinterpretation from Mayede against ATLAS results

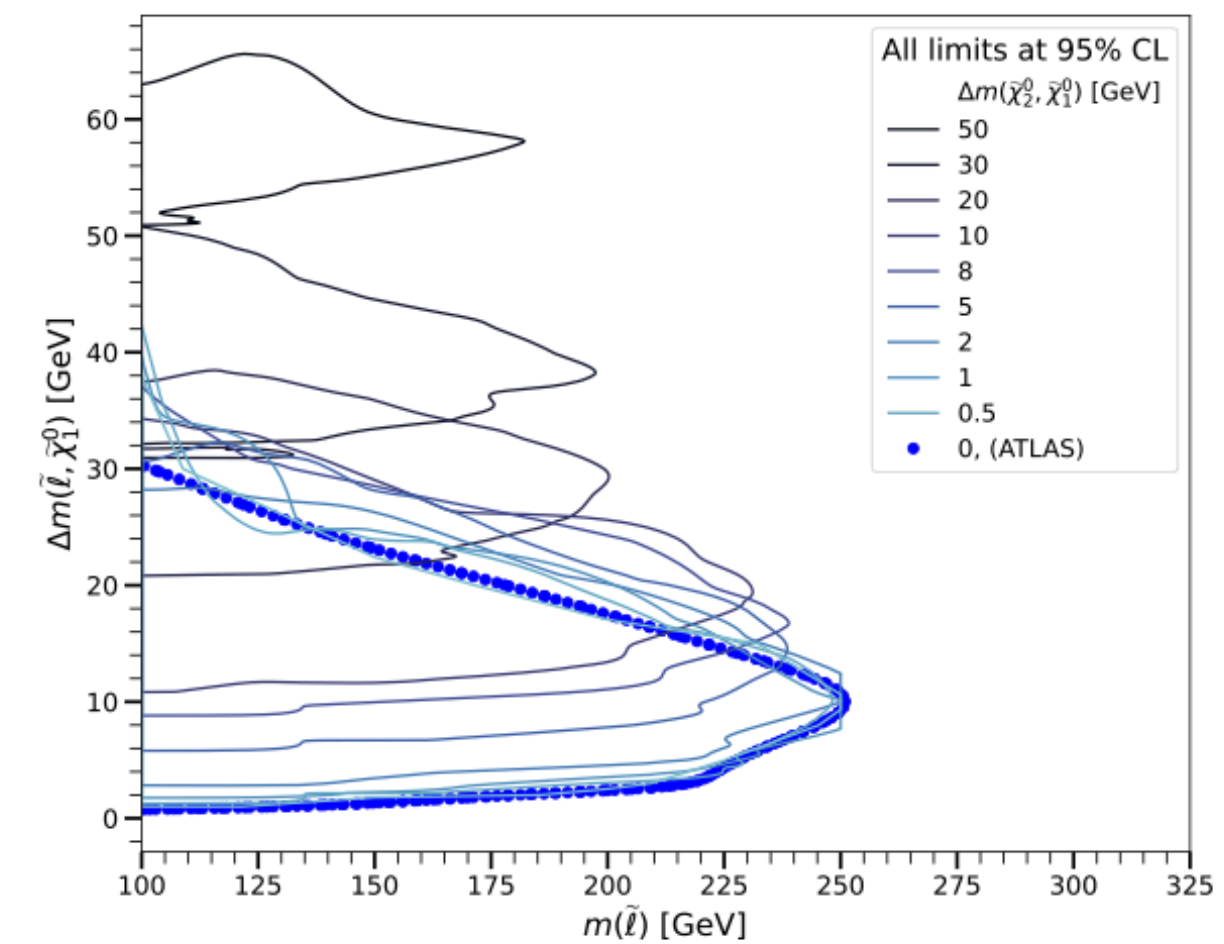
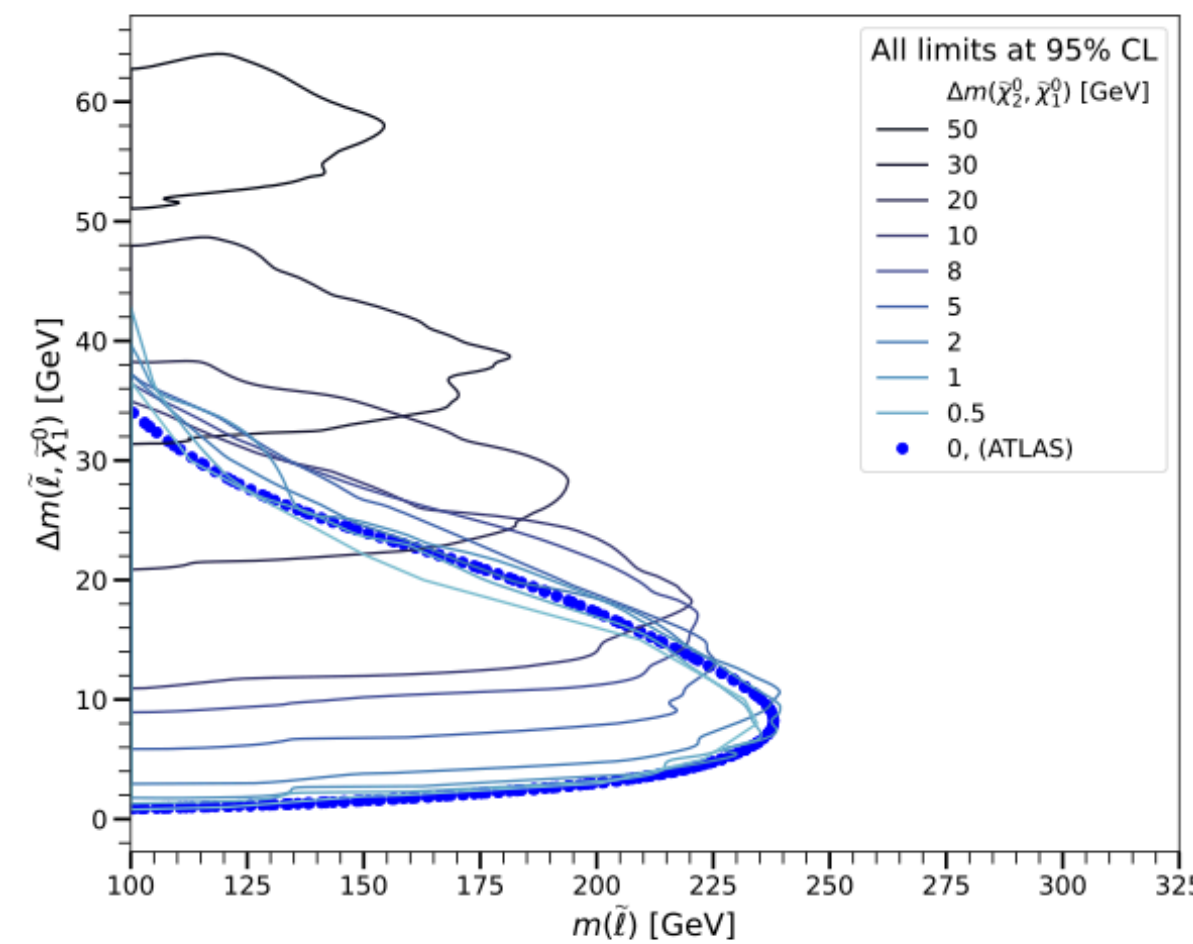
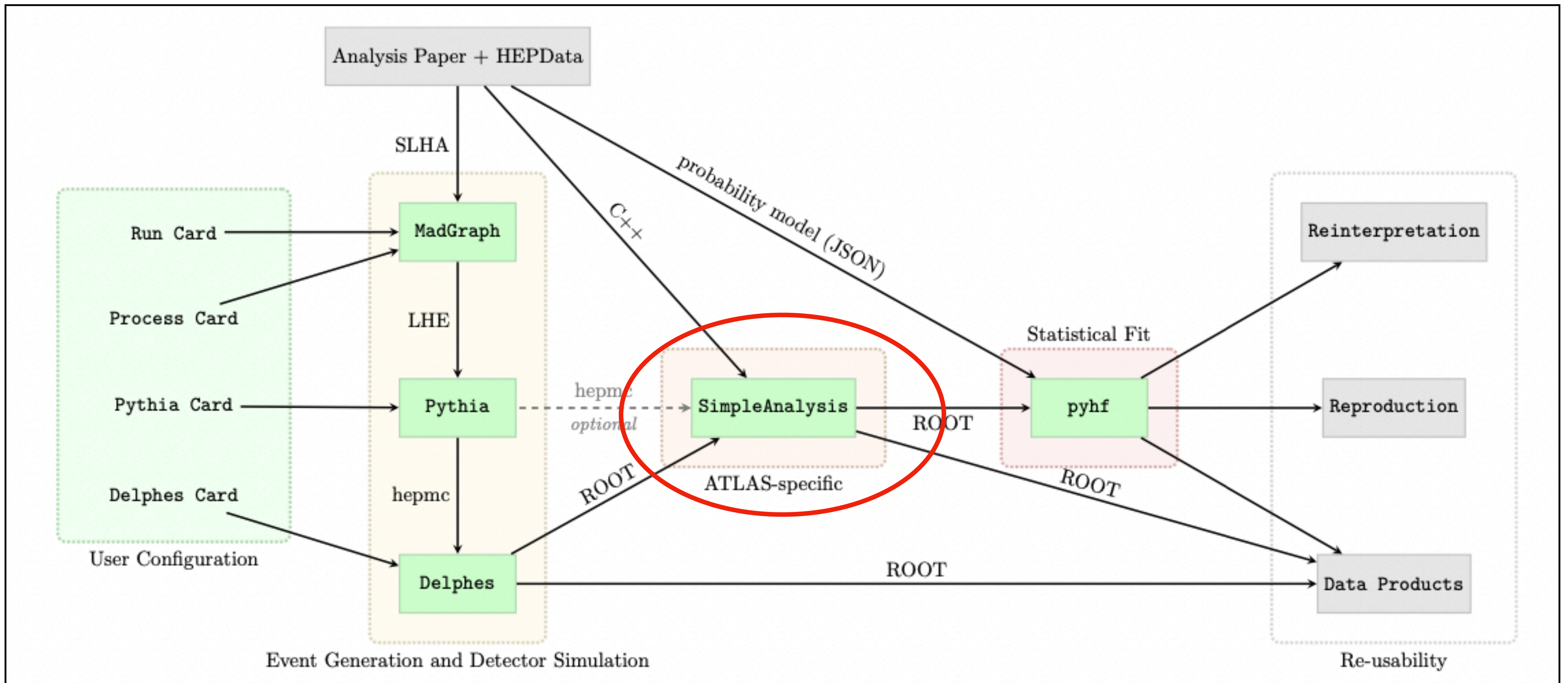


Figure 6: Expected (left) and observed (right) constraints on the slepton-wino-bino model. The model is parameterized by the slepton mass, the slepton- $\tilde{\chi}_1^0$ mass splitting, and the $\tilde{\chi}_2^0$ - $\tilde{\chi}_1^0$ splitting, and compared against the slepton-bino results from Ref. [40].

Simple Analysis pub note summary



Summary

- Simple Analysis is fully **public**, available for theorists/experimentalists alike.
- Generator level pipeline preservation allow for analysis detail to be preserved.
- Reinterpretation made more straight forward.
- Reconstruction/detector effect can be added later on.
- Validation has been made on the generator/reco level kinematics plot.
- Mapyde, a new pipe that that would allow for experimental independent full reinterpretation is available today!

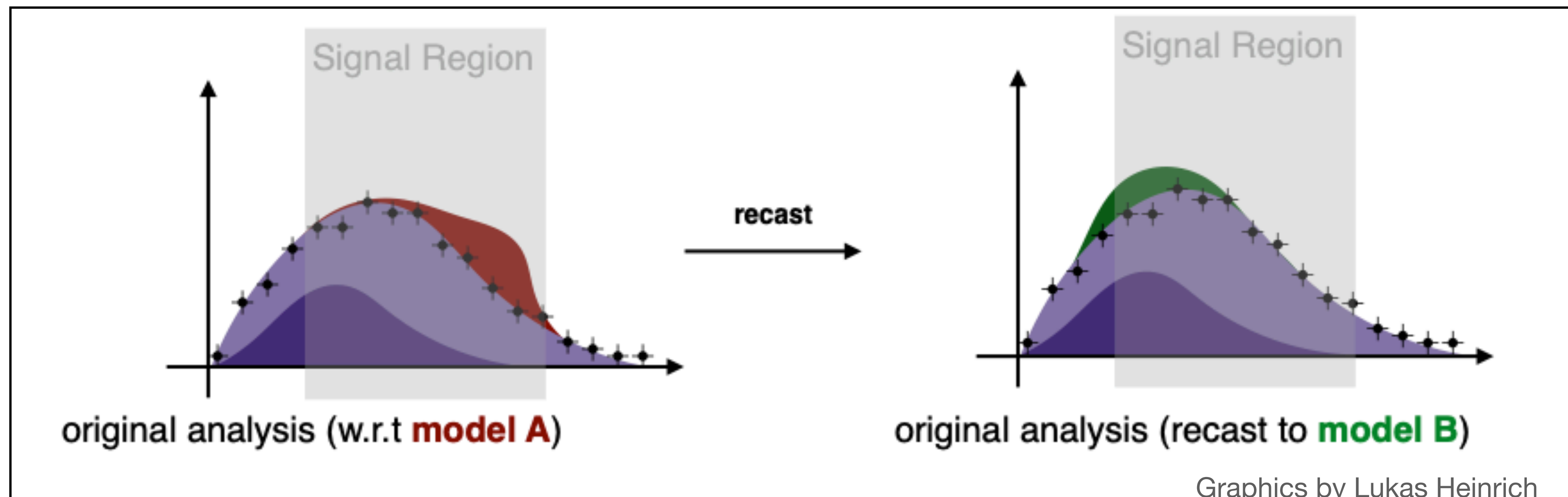


Back up

What is re-interpretation?

—And its importance in the BSM context

- Reinterpretation: The reusing of experimental results for alternative physics hypothesis
- **Advantages:** 1. Reuse of data/workflow/optimization 2. Less labor intensive 3. Time saving



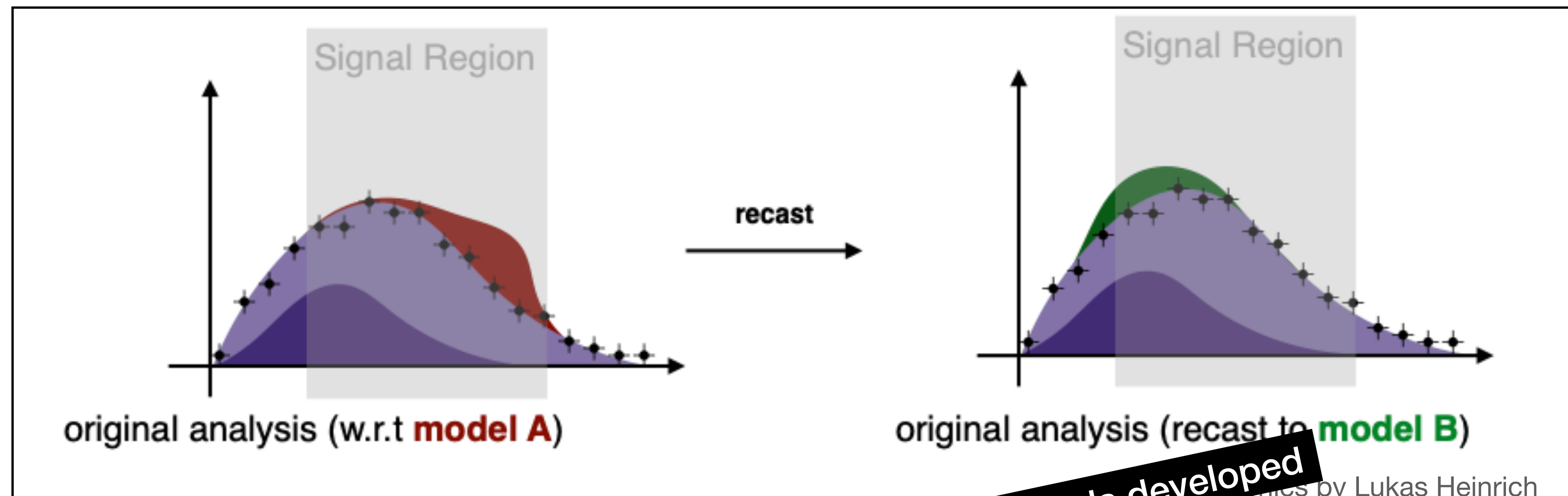
• Two types of Reinterpretation

- > 1. Re-interpretation (Single data-> Many theory)
Abundance of theoretical ideas vs experimental results available
- > 2. Combination (Many data -> Single theory)
“Combination of results (E.G. Higgs combination)”

What is re-interpretation?

—And its importance in the BSM context

- Reinterpretation: The reusing of experimental results for alternative physics hypothesis
- **Advantages:** 1. Reuse of data/workflow/optimization 2. Less labor intensive 3. Time saving



• Two types of Reinterpretation

- > 1. Re-interpretation (Single data -> Many theory)
Abundance of theoretical ideas vs experimental results available
- > 2. Combination (Many data -> Single theory)
“Combination of results (E.G. Higgs combination)”

Available reinterpretation tools

Within Experiment

Outside of Experiment

Unfolded Reinterpretation(Without detector effect) Tools:

MAD Analysis 5

MadAnalysis

Rivet

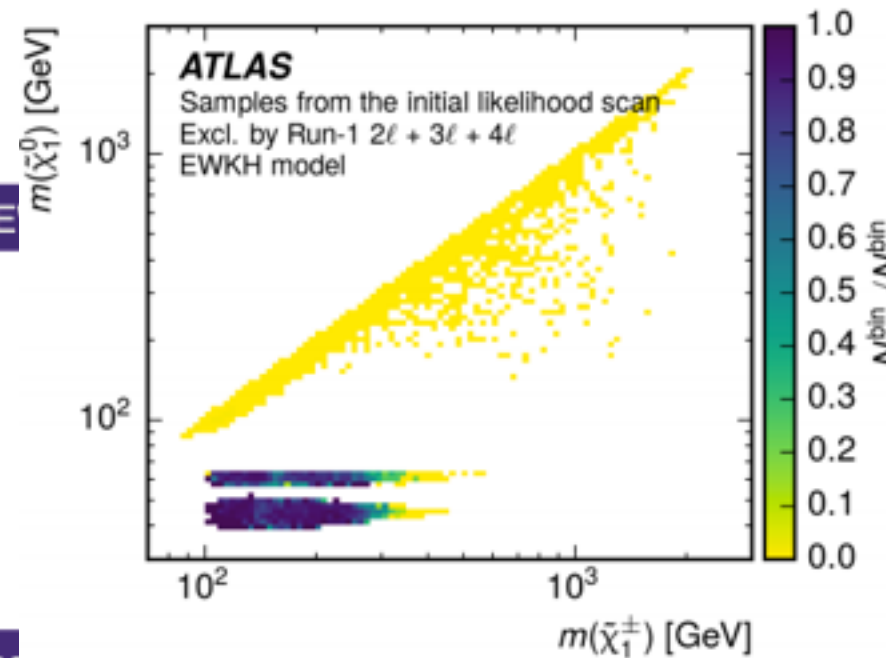
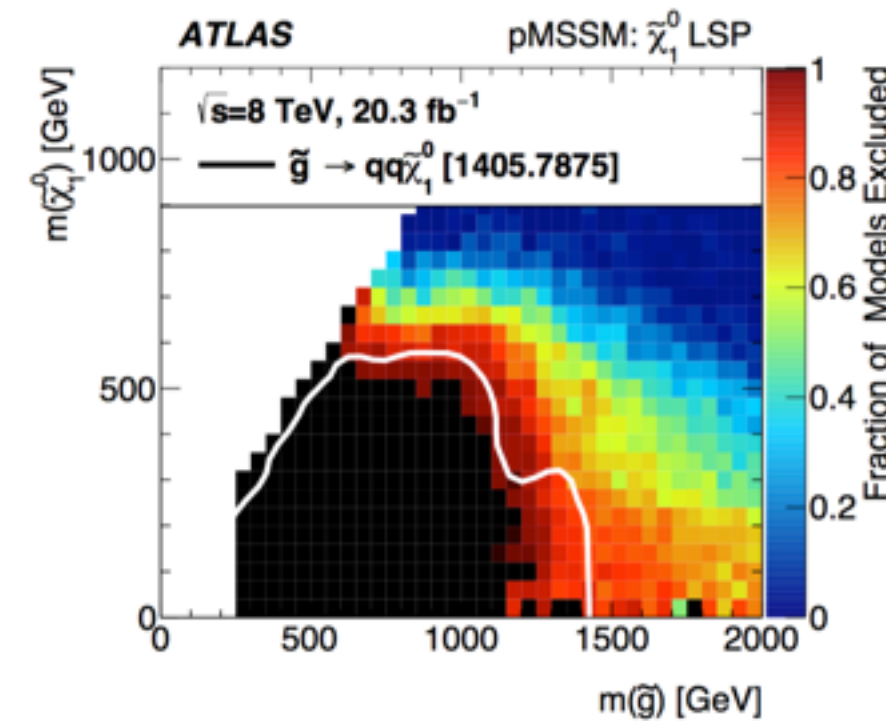
Other Homegrown
HepMC based
Analysis

Folded Reinterpretation(With Detector Effect) Tools:

CHECKIMATE

Homegrown Toy Detector Approximation

Delphes based Analysis



HEP
PUBLISHED FOR SISSA BY SPRINGER
RECEIVED: August 27, 2015
ACCEPTED: September 23, 2015
PUBLISHED: October 21, 2015

Summary of the ATLAS experiment's sensitivity to supersymmetry after LHC Run 1 — interpreted in the phenomenological MSSM

arXiv:1508.06608

19-D(!) pMSSM reinterpretation

HEP
PUBLISHED FOR SISSA BY SPRINGER
RECEIVED: August 3, 2016
ACCEPTED: September 22, 2016
PUBLISHED: September 30, 2016

Dark matter interpretations of ATLAS searches for the electroweak production of supersymmetric particles in $\sqrt{s} = 8$ TeV proton-proton collisions

5-D scan of E

A re-interpretation of $\sqrt{s} = 8$ TeV ATLAS results on electroweak supersymmetry production to explore general gauge mediated models

The ATLAS Collaboration

3-D recast for General Gauge Med

ATLAS-CONF-2016-033

Tools used by ATLAS for reinterpretation

– Introducing the players

Precision Level	Preserved Item	Tools/Service used by ATLAS
0	Data product	HEPData
1	Likelihood	Simple likelihood JSON HistFactory
2	Truth Level Pipeline	SimpleAnalysis/Rivet
10	Full Reco Pipeline	Recast
10+	Preservation of ML Model/optimization	ONNX

++Increased level of Information/Precision

Increase precision enable better reinterpretation from original analysis!

Level 0: HEPData

– Storing of experimental data product from analyses for reinterpretation

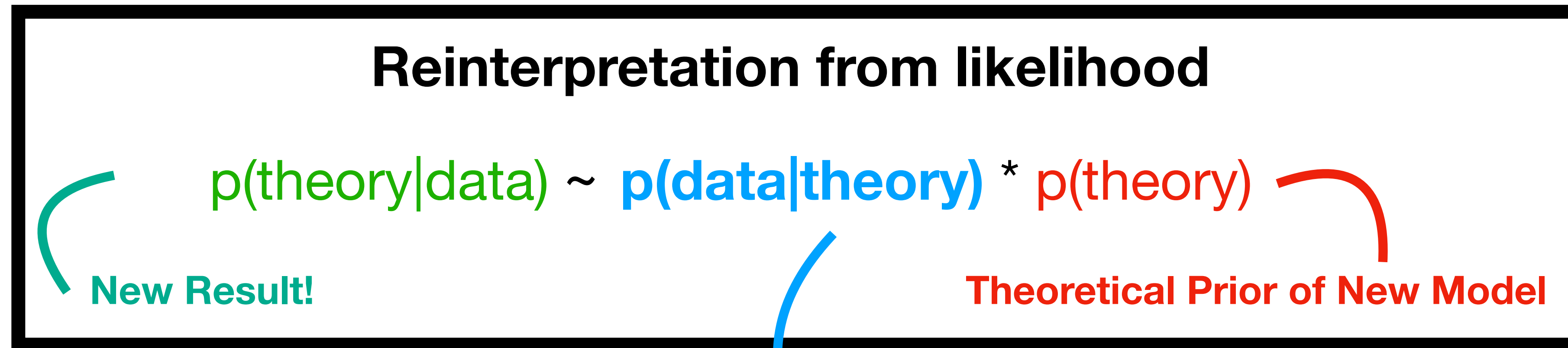
- Previously: Table-based webpage (40 year-old):
- [New digital library](#) : Paper
 - Database: Include Efficiency Map, yields, uncertainties, outflow tables
 - Reinterpretation tools: Rivet and SimpleAnalysis code, preservation of ML
 - Webservice
- Made an ATLAS requirement for publication
- Centralized tool across other HEP experiments



Level 1: Statistical Model (Simple Likelihood)

–Saving the most of our experiment results in one variable

- Previously, often HEPData information is used to construct approximate likelihood
- **Likelihood: Great bang for the buck:** encapsulate information about Limits, data/MC plots, yield take, and systematics. ->Best object that can be provided!
- ATLAS started providing them on HEPData since 2019 (SUSY)



[Simple Likelihood Pubnote](#)

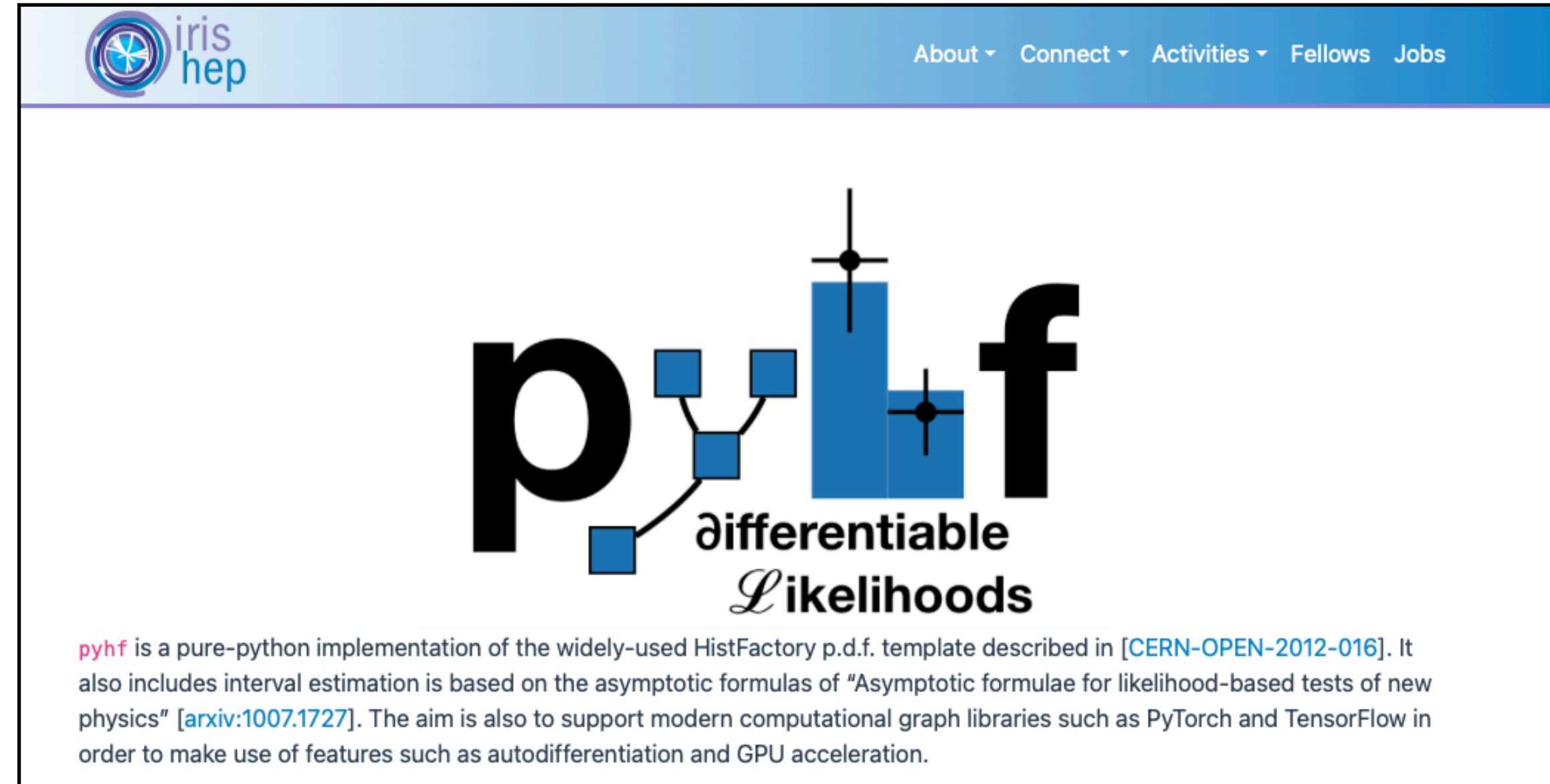
Existing Exp. Likelihood (What We Provide)

Level 1: HistFactory JSON

See ["Publishing statistical models discussion" on Tuesday](#)

–pyhf as a pure-Python implementation of HistFactory with JSON

- [PyHF](#): HistFactory in full python
- Increasingly used by ATLAS analyses ([24 analyses](#) published-up-to-date)
- Results in easy to read json format: [Database](#)
- Reinterpretation:
 - 1. Publishes statistical model (Likelihood) in json file format
 - 2. New version of [SModel](#) and [MadAnalysis5](#) allow for automatic reinterpretation from pyHF json file



The screenshot shows the top part of the pyhf website. At the top left is the 'iris hep' logo. To the right are navigation links: 'About', 'Connect', 'Activities', 'Fellows', and 'Jobs'. The main content area features a large logo for 'pyhf' where the 'y' is a blue graph with nodes. Below the logo is the text 'differentiable Likelihoods'. At the bottom, there is a paragraph of text: 'pyhf is a pure-python implementation of the widely-used HistFactory p.d.f. template described in [CERN-OPEN-2012-016]. It also includes interval estimation is based on the asymptotic formulae of "Asymptotic formulae for likelihood-based tests of new physics" [arxiv:1007.1727]. The aim is also to support modern computational graph libraries such as PyTorch and TensorFlow in order to make use of features such as autodifferentiation and GPU acceleration.'



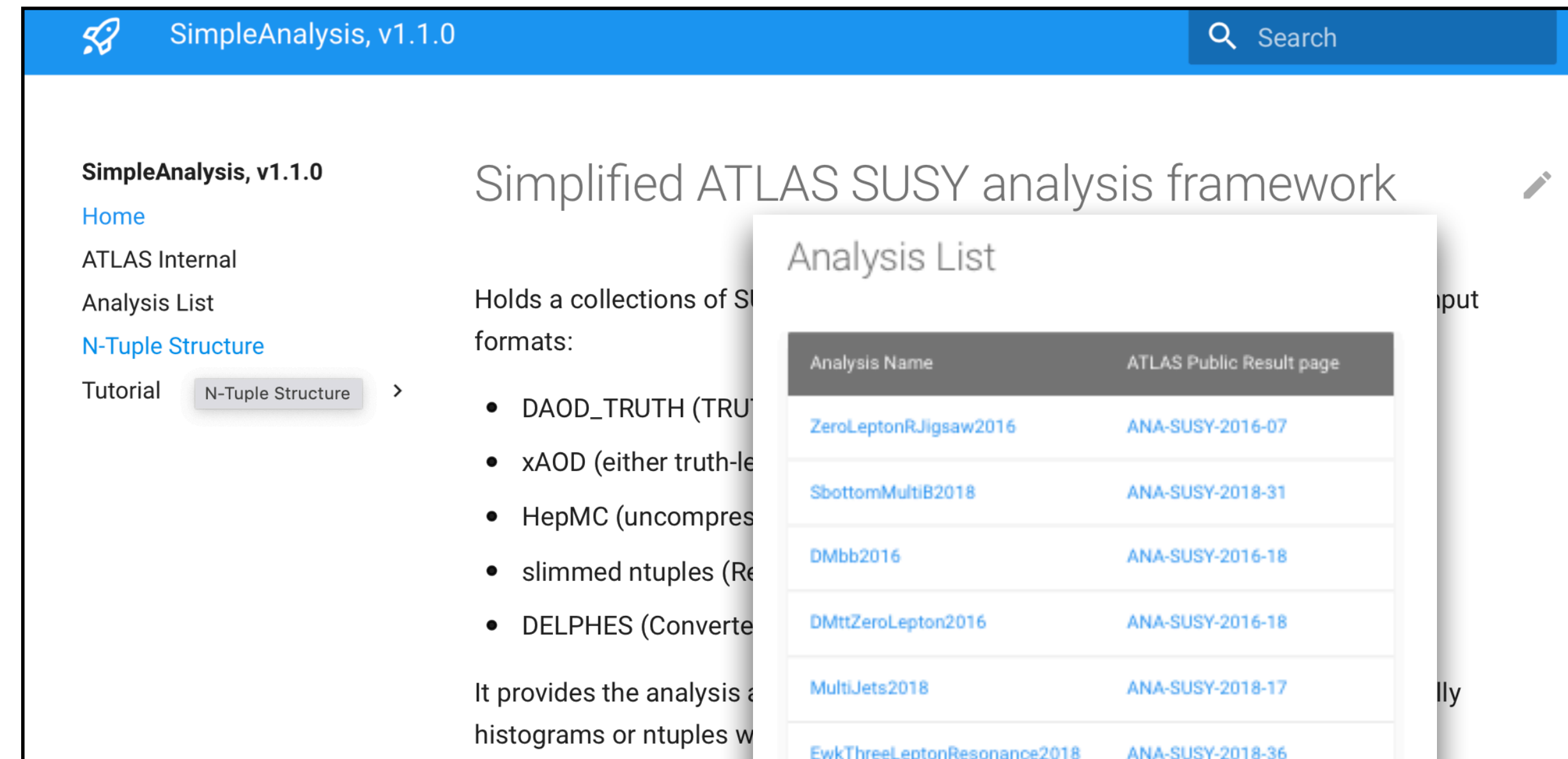
Source Code

Level 2: SimpleAnalysis

– Truth level pipeline preservation

- Preserves the analysis code pipeline
- Allow for reinterpretation up to the truth level
 - Detector effect can be added: public simulation like Delphes or internally by ATLAS fast simulation
- Experimentalist friendly input format (HepMC, DAOD, ROOT n-tuple, DELPHES)
- Large list of [ATLAS SUSY analyses](#) available:
- For **public use**: can be used external to ATLAS

See [tutorial by Giordon Stark](#) Wednesday



The screenshot shows the SimpleAnalysis v1.1.0 website. The header includes the logo and version number, and a search bar. The main content area is titled "Simplified ATLAS SUSY analysis framework" and lists supported input formats: DAOD_TRUTH (TRU), xAOD (either truth-le), HepMC (uncompressed), slimmed ntuples (Re), and DELPHES (Converted). Below this, it states "It provides the analysis of histograms or ntuples w". On the right, a modal window titled "Analysis List" is open, displaying a table of analyses.

Analysis Name	ATLAS Public Result page
ZeroLeptonJigsaw2016	ANA-SUSY-2016-07
SbottomMultiB2018	ANA-SUSY-2018-31
DMbb2016	ANA-SUSY-2016-18
DMttZeroLepton2016	ANA-SUSY-2016-18
MultiJets2018	ANA-SUSY-2018-17
EwkThreeLeptonResonance2018	ANA-SUSY-2018-36
EwkThreeLeptonERJR2018	ANA-SUSY-2018-06
PairedDijets2016	ANA-SUSY-2016-09
EwkTwoLeptonTwoJet2016	ANA-SUSY-2016-24
DirectStau2018	ANA-SUSY-2018-04
OneLeptonMultiJets2016	ANA-SUSY-2016-11
DisappearingTrack2016	ANA-SUSY-2016-06
StopZ2016	ANA-SUSY-2016-20
EwkTwoLeptonRJ2016	ANA-SUSY-2017-03

SimpleAnalysis: [Webpage](#)



Source Code

[SimpleAnalysis PubNote](#)

Available list of Analyses

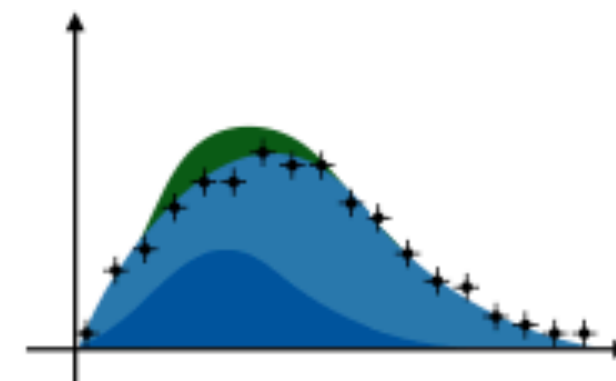
Level 10: Recast

– Full Analysis pipeline preservation

- Preservation of original analysis pipeline for re-interpretation
- Contain full detector simulation information + original code/workflow
- Reinterpretation with full workflow + simulation effect
- Made a requirement for publication in Exotics, HDBS and SUSY
- Implementation Internal to ATLAS: some reinterpretation results published with the tool
- [Software](#) fully open source!
- Future: Allow for an interface to submit reinterpretation request online



preservation of the original analysis pipeline for collaboration-approvable results

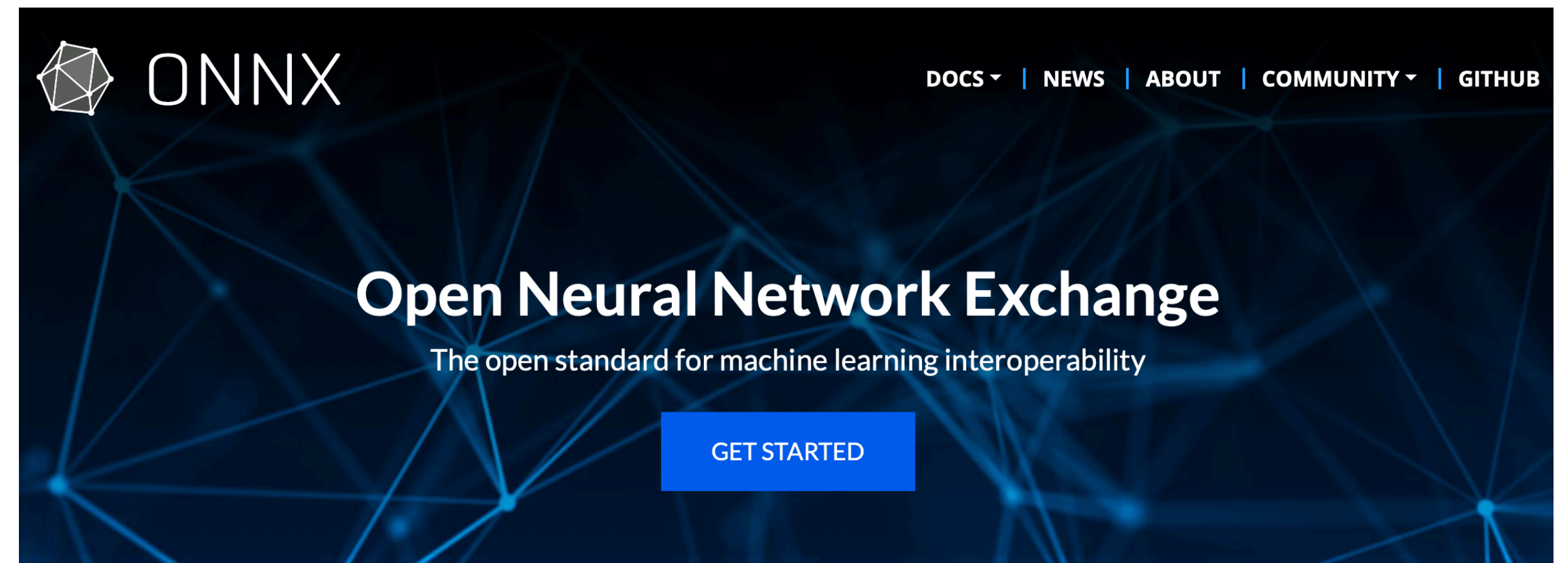


Level 10+: ONNX/NN

See [talk by Dan Guest](#) later today

–Beyond usual pipeline preservation: Reuse of ML training results

- ML development significantly improve analyses sensitivity. Theorists would like to make use of the BDTs/NN.
- ONNX(Open Neural Network Exchange)
- BDT preservation -> convert to standalone c++ in SimpleAnalysis
- Non-BDT NN -> Preservation by Serializing through Onnx
- Can be preserved with SimpleAnalysis to be reused
- 2 analyses already have ML encapsulated in the workflow



3 b-jets+

Search for supersymmetry in final states with missing transverse momentum and three or more b -jets in 139 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector

The ATLAS Collaboration

[ANA-SUSY-2018-30](#)

[SA with Onnx](#)

One Lepton MultiJets

Search for R-parity-violating supersymmetry in a final state containing leptons and many jets with the ATLAS experiment using $\sqrt{s} = 13 \text{ TeV}$ proton–proton collision data

The ATLAS Collaboration

[ANA-SUSY-2019-04](#)

[SA with Onnx](#)

Two existing analyses already contain Onnx



Source Code