Spey: smooth inference for reinterpretation studies

Jack Y. Araz
Thomas Jefferson National Accelerator Facility

Based on arXiv: 2307.06996 [hep-ph]

PyHEP 2023 workshop
October 12th, 2023

Jefferson Lab
EXPLORING THE NATURE OF MATTER

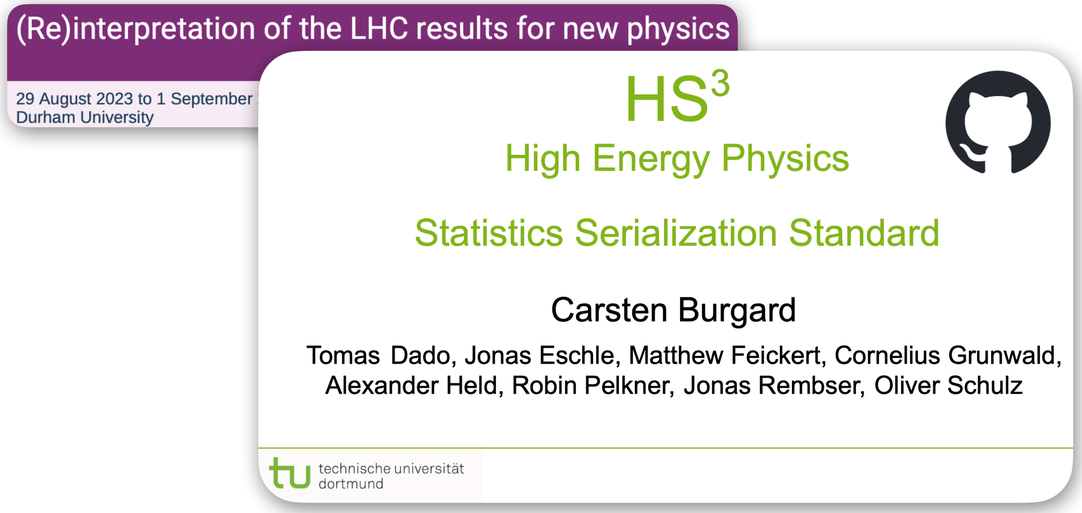
U.S. DEPARTMENT OF ENERGY

MCnet

OLD DOMINION UNIVERSITY

Sales pitch of the talk

- Not enough information to construct a reliable likelihood
- How can we improve the accuracy of the simplified likelihoods?
- Full likelihoods can become computationally intensive; can we simplify them without sacrifice? (Spoiler alert: ML!)
- There are many different software for hypothesis testing; how can we unite them?



Outline

- Introduction
- (Better) simplified likelihoods
- Including existing packages
- Combining likelihoods
- Conclusion

Introduction

Statistical Models

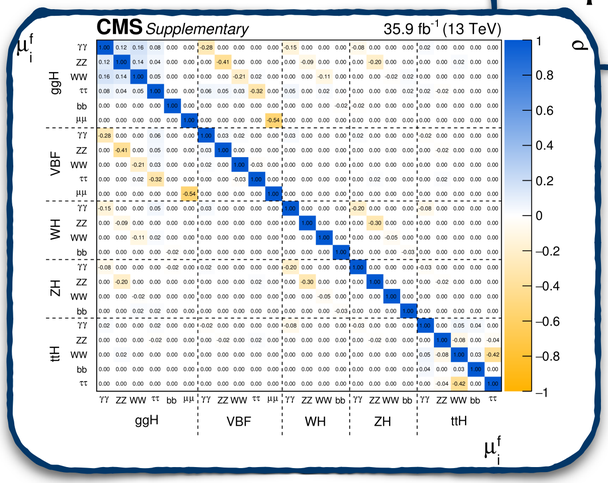
Why important?

ATLAS SUSY and Exotics workshop
S. Kraml '20

- The mathematical description of the analysis is provided within its statistical model.
- A likelihood enables the standard statistical approaches to extract information.
- *i.e.* how reasonably aligned the theoretical predictions with the experimental observations?

Les Houches Recommendations (EPJC '12)

Simplified likelihoods from CMS



CMS-NOTE-2017-001

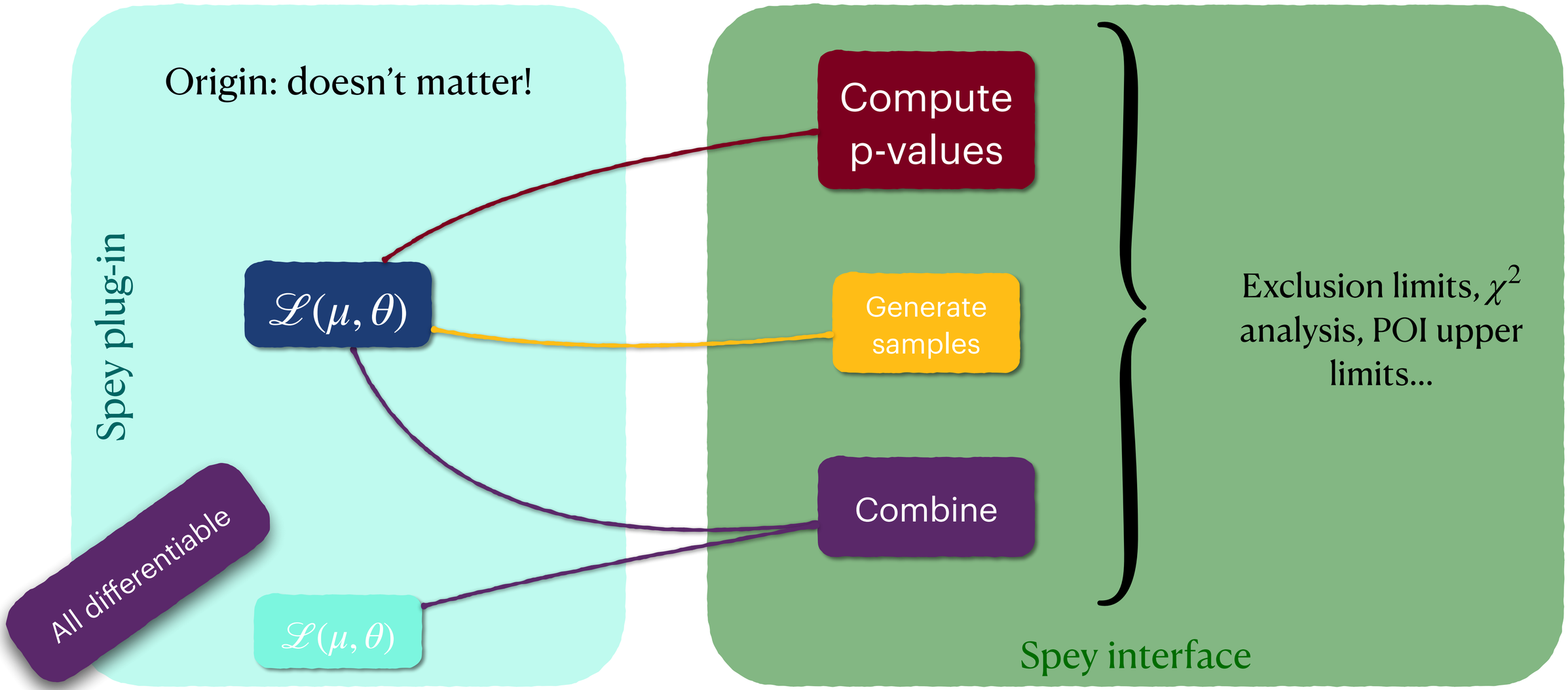
Full likelihoods from ATLAS

pyhf
differentiable Likelihoods

ATL-PHYS-PUB-2019-029

Spey demystified

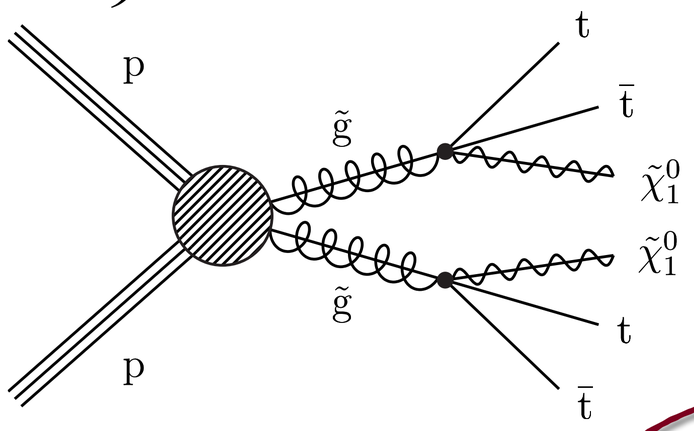
```
$ pip install spey
```



(Better) Simplified Likelihoods

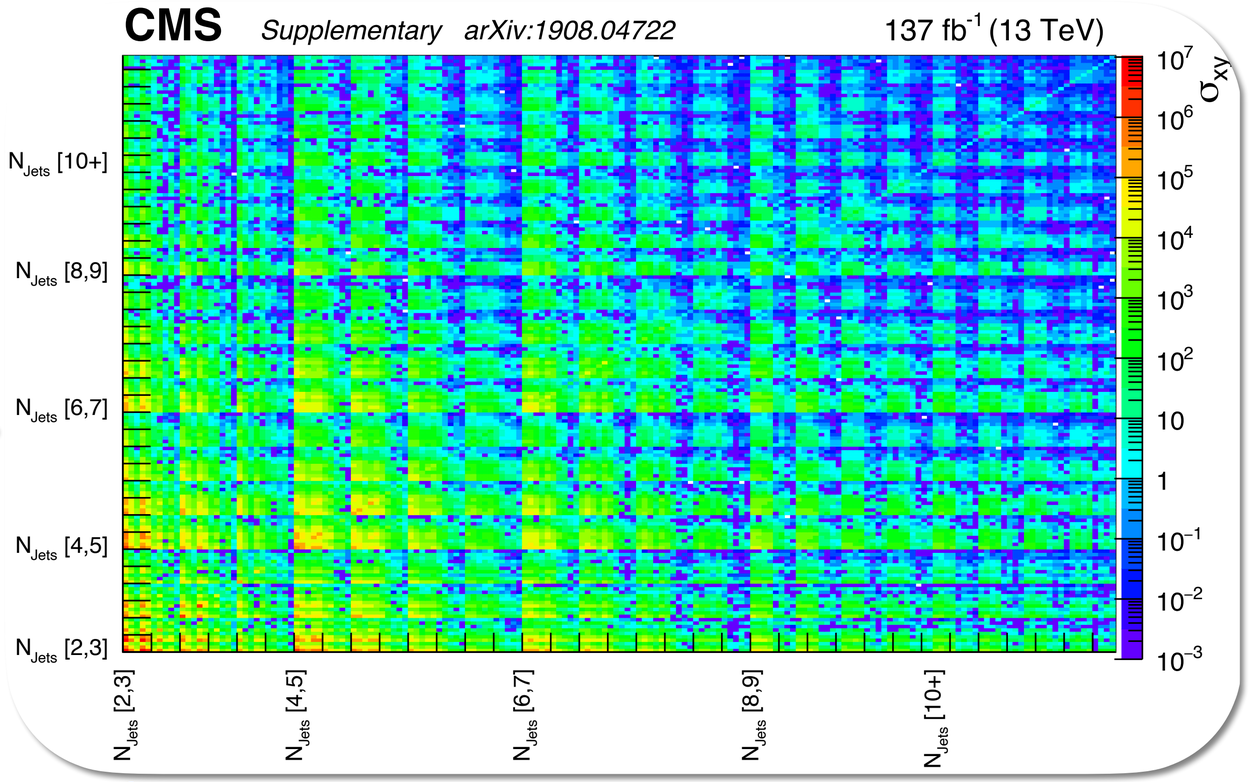
Simplified likelihoods

CMS-SUS-19-006



$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{bins}} \text{Poiss}\left(n^i \;\; \mu n_s^i + n_b^i + \theta^i\sigma_b^i\right)\right] \cdot \mathcal{N}\left(\theta \;\; 0,\rho\right)$$

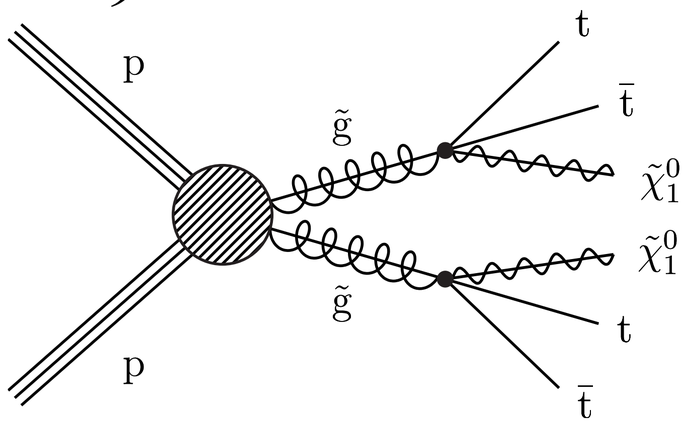
CMS-NOTE-2017-001



186 signal regions!

Simplified likelihoods

CMS-SUS-19-006



$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{bins}} \text{Poiss}\left(n^i \;\; \mu n_s^i + n_b^i + \theta^i \sigma_b^i\right)\right] \cdot \mathcal{N}\left(\theta \;\; 0,\rho\right)$$

CMS-NOTE-2017-001

Third Moment Expansion

$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{bins}} \text{Poiss}\left(n^i \;\; \mu n_s^i + \bar{n}_b^i + A_i\theta_i + C_i\theta_i^2\right)\right] \cdot \mathcal{N}\left(\theta \;\; 0,\bar{\rho}\right)$$

$\bar{n}_b^i$:= the central value of the background
A_i := the effective sigma of the background uncertainty
C_i := asymmetry of the background uncertainty

Buckley, Citron, Fichet, Kraml, Waltenberger, Wardle; JHEP '18

Effective Sigma

$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{bins}} \text{Poiss}\left(n^i \;\; \mu n_s^i + n_b^i + \theta^i \sigma_{\text{eff}}^i(\theta^i)\right)\right] \cdot \mathcal{N}\left(\theta \;\; 0,\rho\right)$$

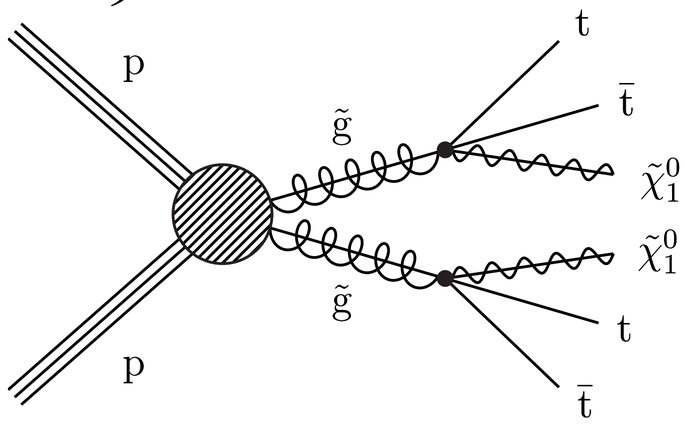
$$\sigma_{\text{eff}}^i(\theta^i) = \sqrt{\sigma_i^+\sigma_i^- + (\sigma_i^+ - \sigma_i^-)(\theta^i - n_b^i)}$$

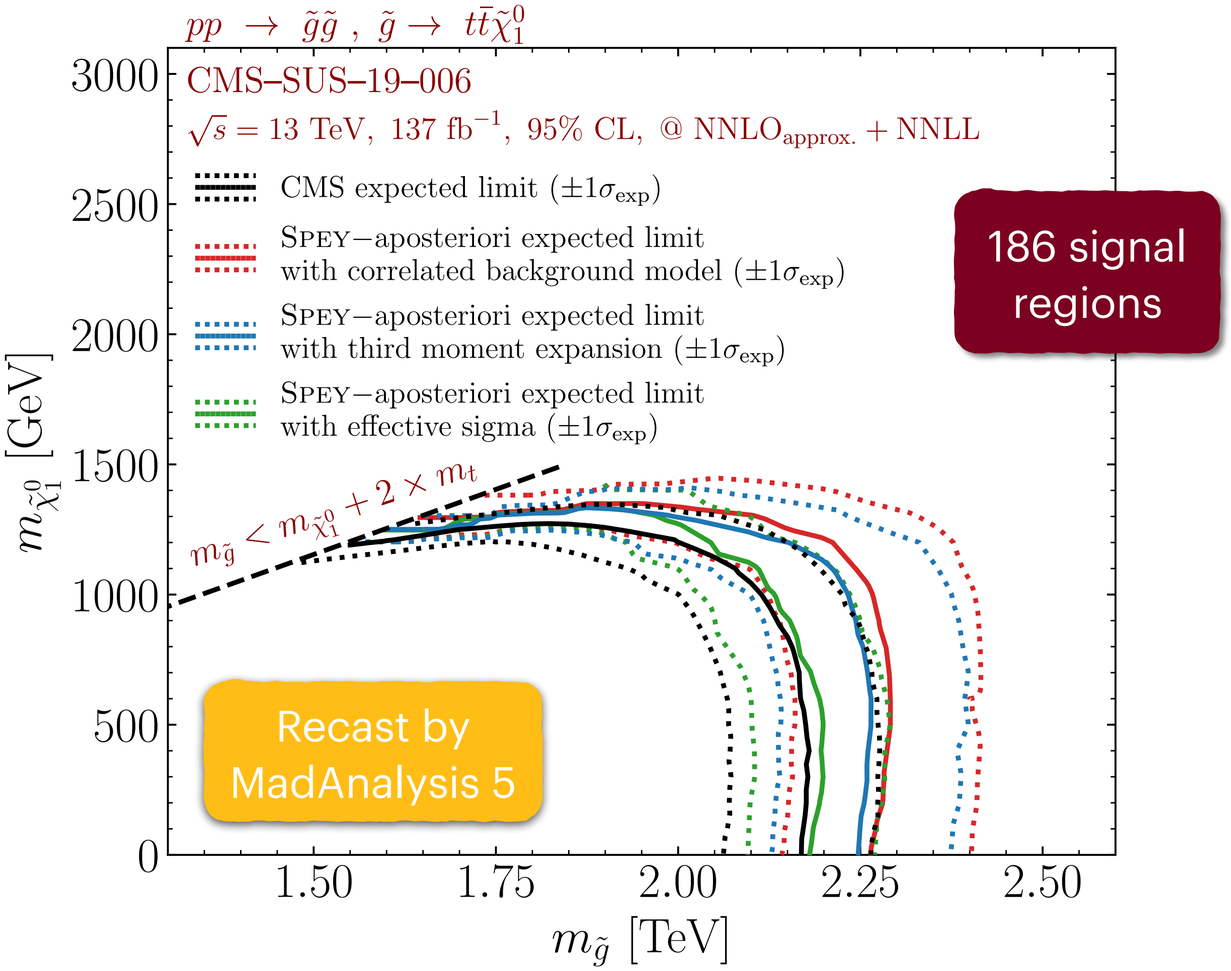
Barlow, '04

New

Simplified likelihoods

CMS-SUS-19-006





186 signal regions

Recast by MadAnalysis 5

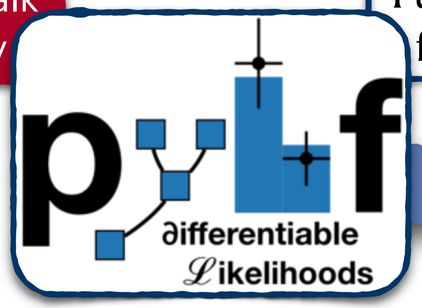
Including existing packages

Full likelihoods

No need to reinvent the wheel!

See Matthew's talk from yesterday

Full likelihoods from ATLAS



ATL-PHYS-PUB-2019-029

- ATLAS shares HistFactory-like JSON files to form full profile likelihoods.
- Each file includes detailed information on backgrounds and corresponding nuisance parameters.

```
$ pip install spey-pyhf
```

HEPData

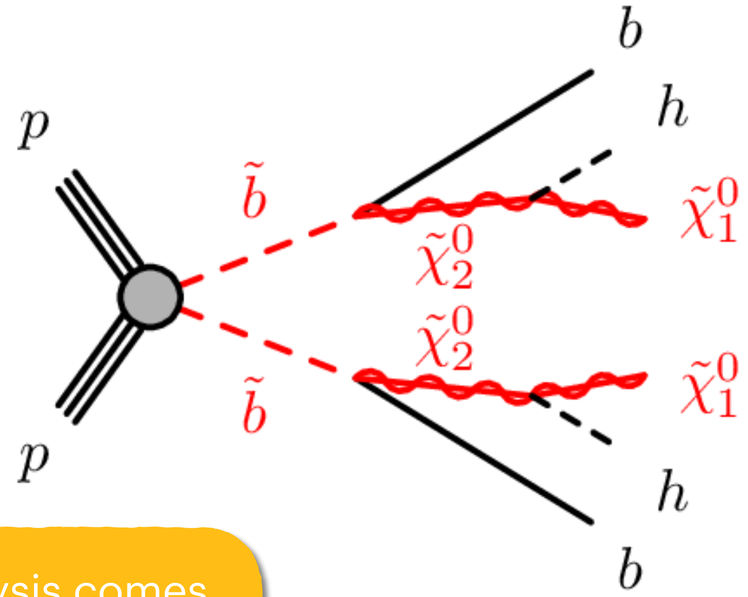
gz File

Archive of full likelihoods in the HistFactory JSON format described in CERN-EP-2019-188. For each signal point the background-only model is found in the file named BkgOnly.json. All jsonpatches are contained in the file patchset.json. Each patch is identified in patchset.json by the metadata field "name": "C1N2_Wh_hbb_[m1]_[m2]" where m1 is the mass of both the lightest chargino and the next-to-lightest neutralino (which are assumed to be nearly mass degenerate) and m2 is the mass of the lightest neutralino.

Download

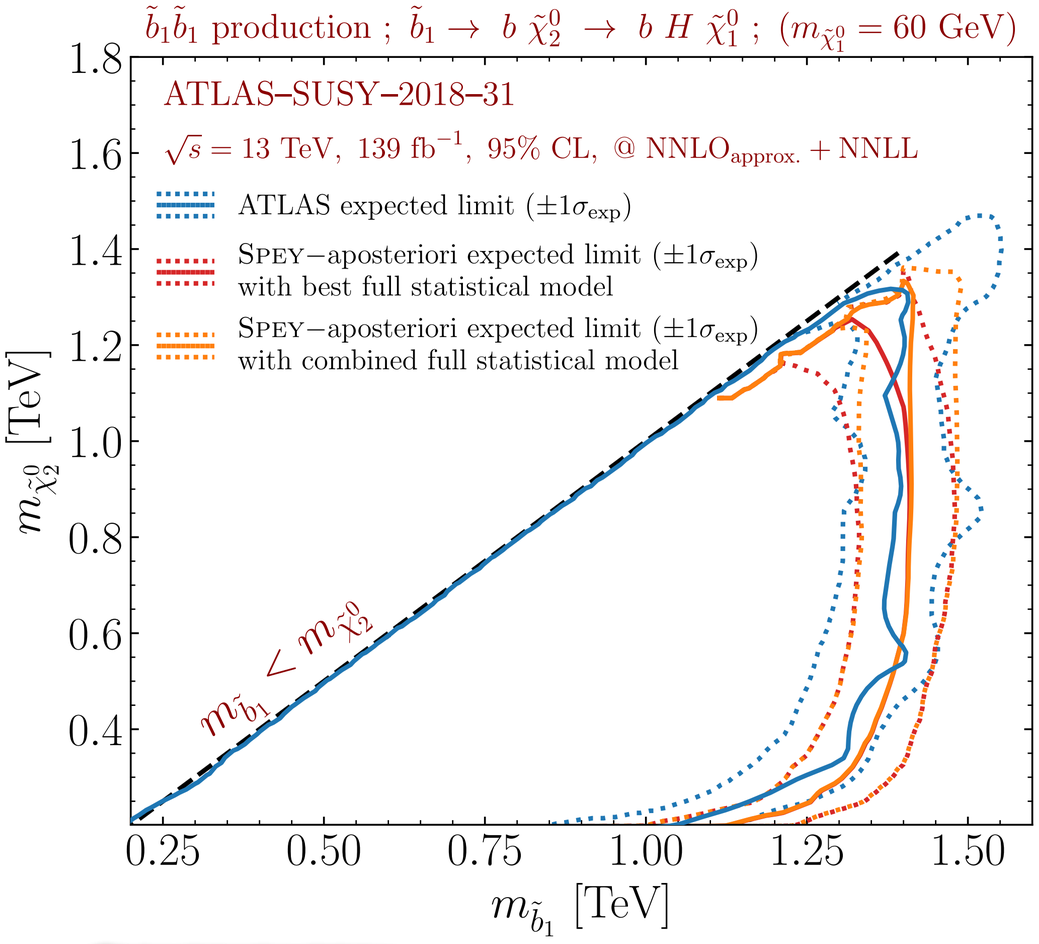
Full likelihoods

```
$ pip install spey-pyhf
```



This analysis comes with three different super regions!

Full likelihoods include all the necessary information to mix and match nuisance parameters to combine them!



Recast by MadAnalysis 5

Combining likelihoods

An MSSM example

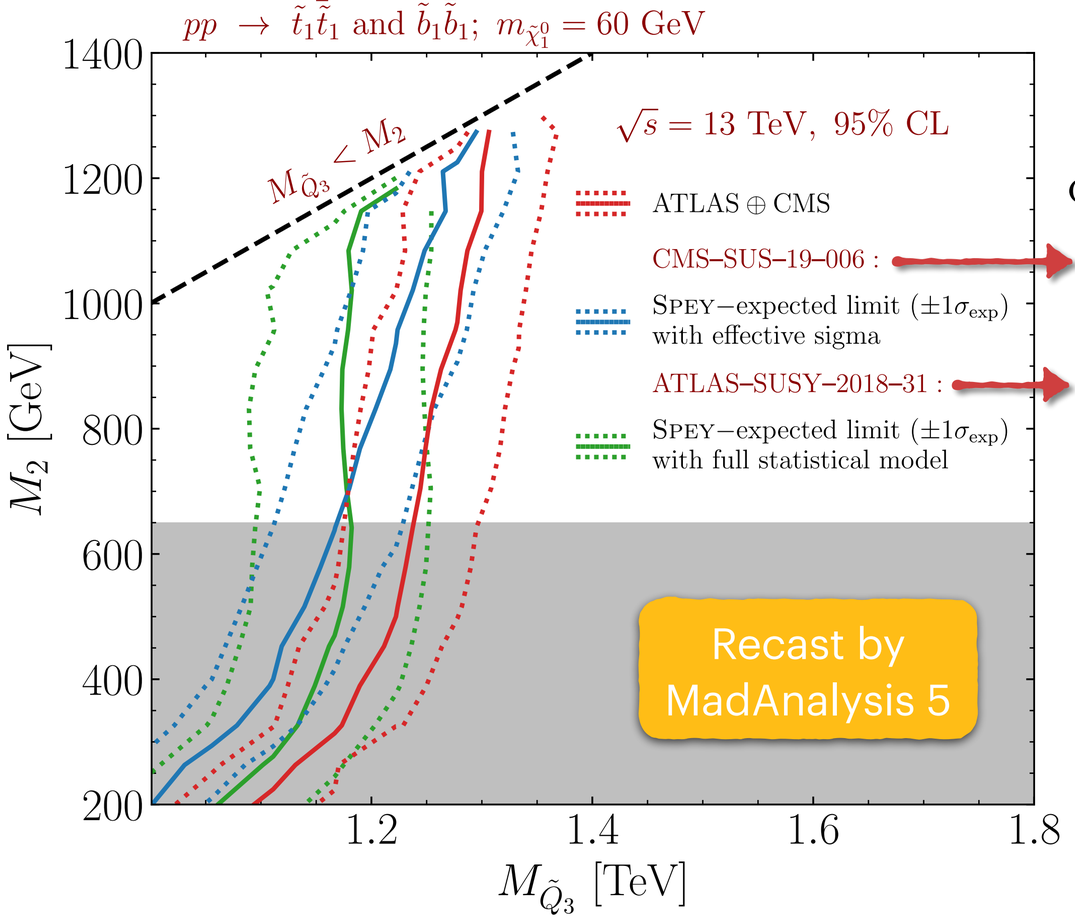
$$\mathscr{L}' = \mathscr{L}_{\text{ATLAS}} \oplus \mathscr{L}_{\text{CMS}}$$

Full likelihood

Simplified likelihood with effective sigma model

A combination of analyses, rather than regions, contains much more information!

! Assumes that likelihoods are not correlated !



Original analysis:

CMS–SUS–19–006 : $pp \to \tilde{g}\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$

ATLAS–SUSY–2018–31 : $pp \to \tilde{b}_1\bar{\tilde{b}}_1(\to bh\tilde{\chi}_2^0)$

Conclusion

Conclusion

Spey online documentation

Spey arXiv:2307.06996

- Choosing the correct likelihood is as crucial as proper recast implementation!
- Spey acts as a hub for various likelihood implementations for hypothesis testing.
- The ability to study likelihoods in a backend-agnostic way opens up various possibilities, such as combinations.
- Spey offers an easily expandable plug-in system for future likelihood implementations.

Smarter simplified likelihood implementations

Machine Learned likelihoods

Dedicated custom likelihoods for other experiments