Spey
What are we up to these days?

Jack Y. Araz

Thomas Jefferson National Accelerator Facility

LPSC Grenoble, FR

June 18th, 2024

Sales pitch for the talk

- There is not enough information to construct a reliable likelihood
- How can we enhance the accuracy of the simplified likelihoods?
- Full likelihoods can become computationally intensive. Can we simplify them without sacrificing? (Spoiler alert: ML!)
- There are many different software for hypothesis testing; how can we unify them?

Outline

- What on earth is Spey?
- Plans for world domination
- Conclusion

What on earth is Spey?

Jack Y. Araz

Searches for new physics

Lagrangian

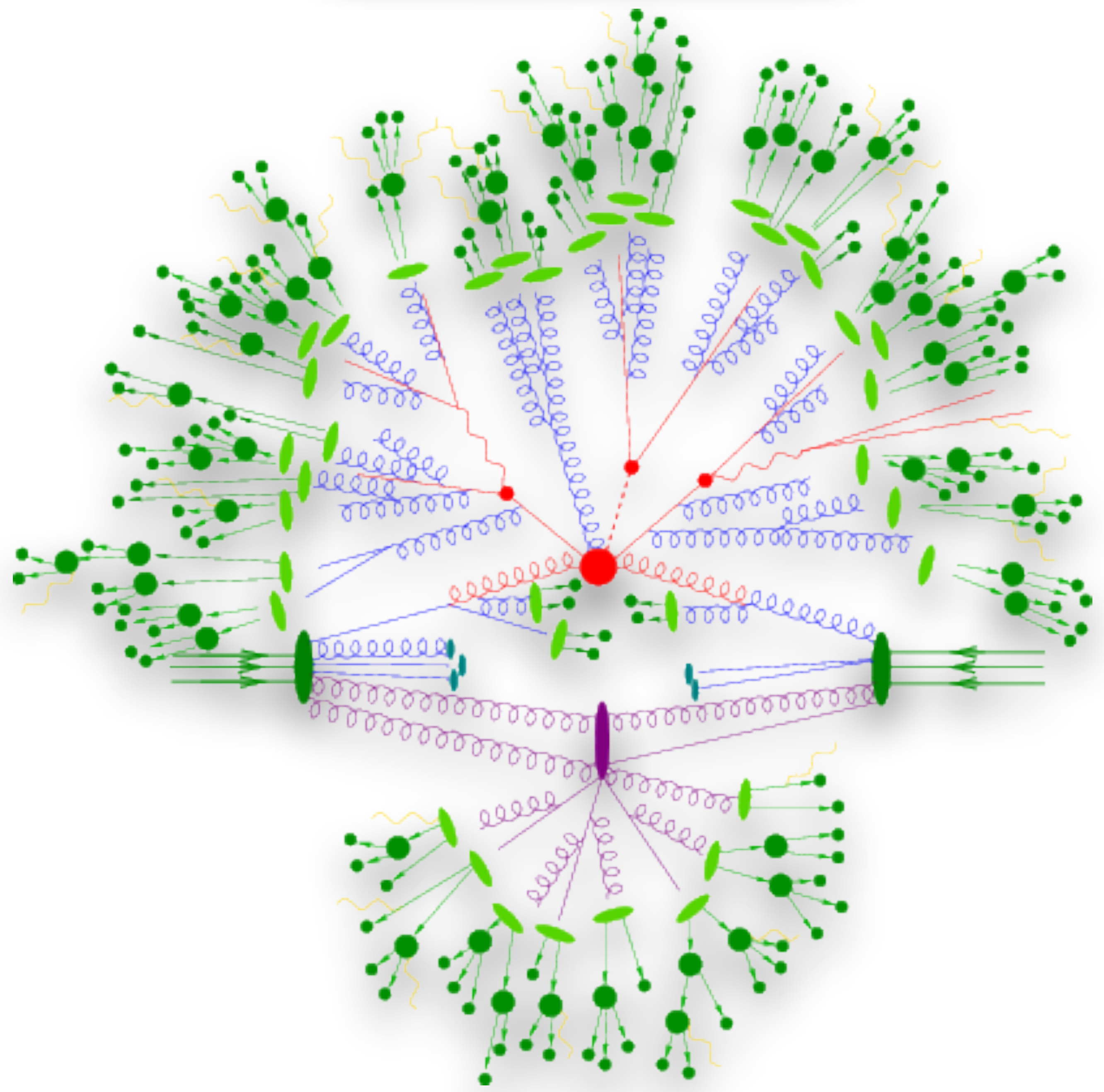


Image credit: Sherpa

Searches for new physics

Lagrangian

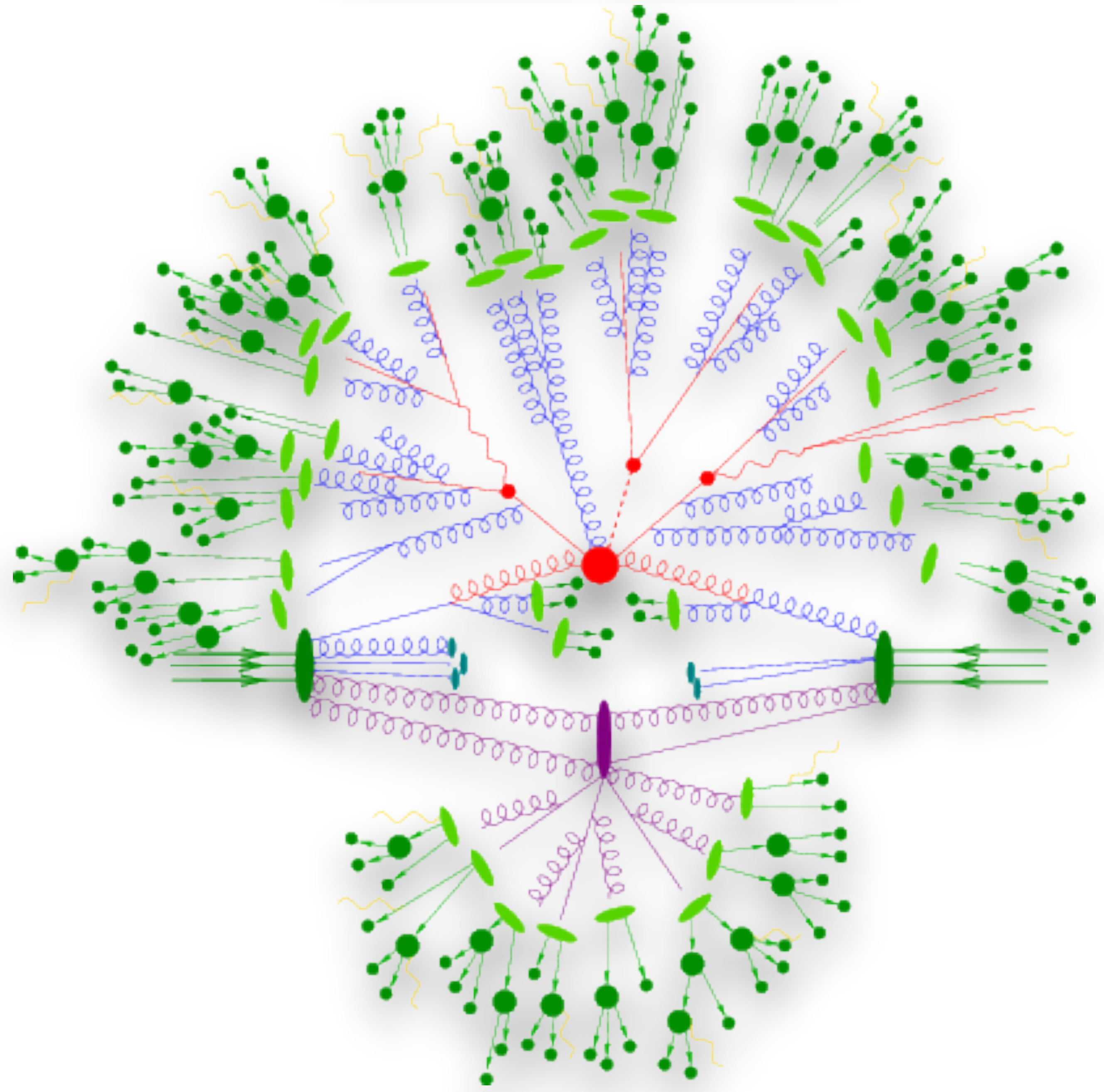
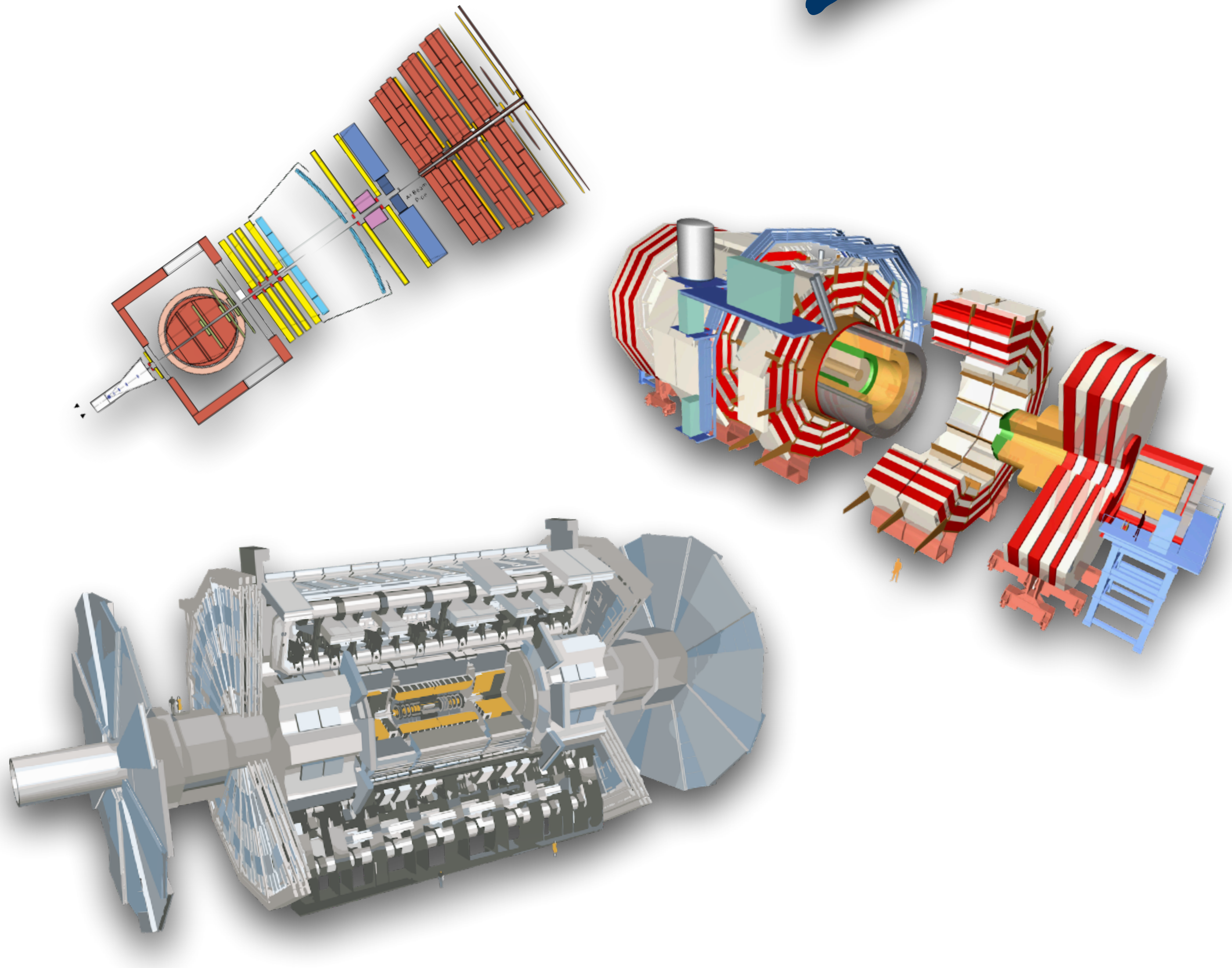
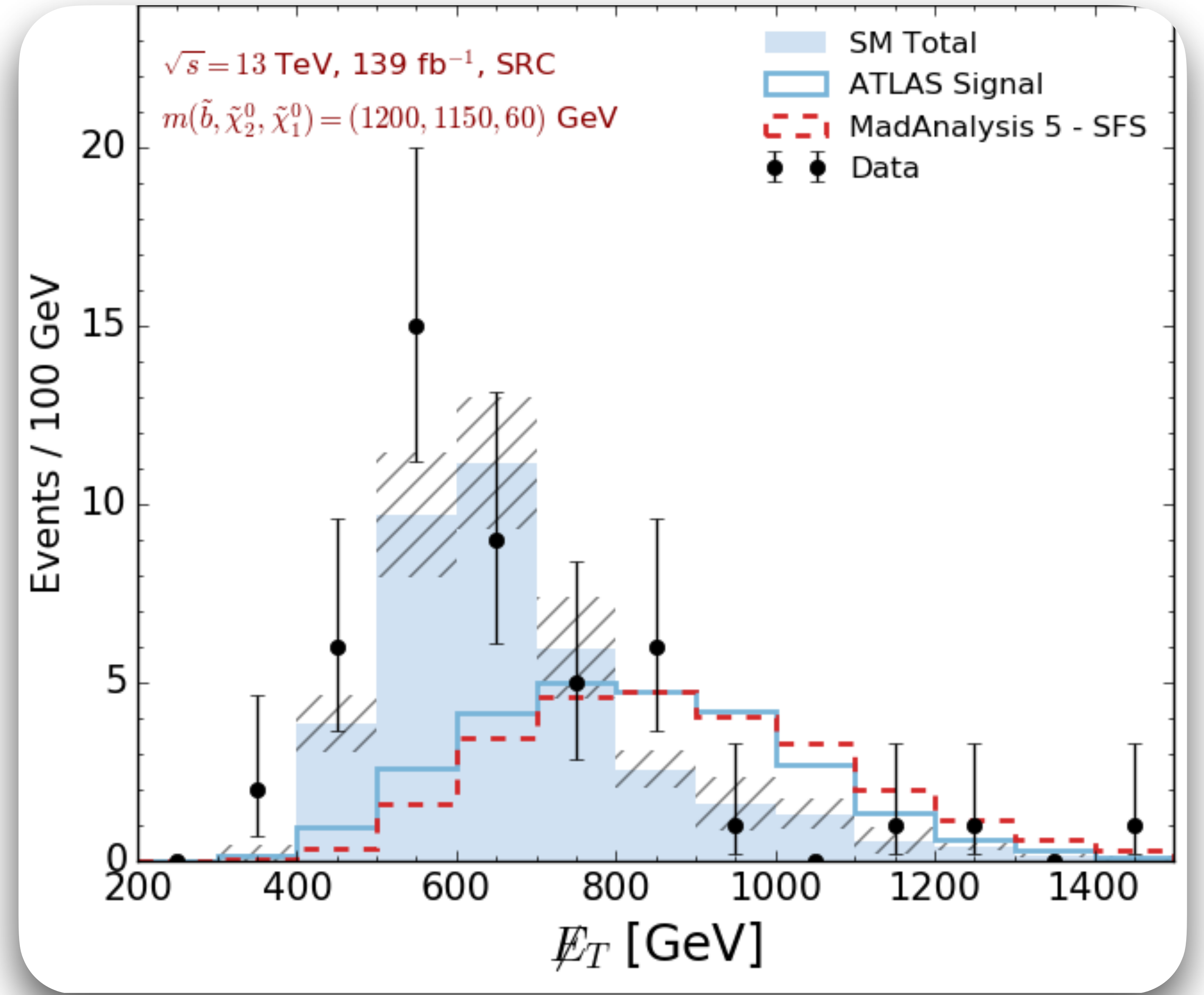


Image credit: Sherpa

Smearing



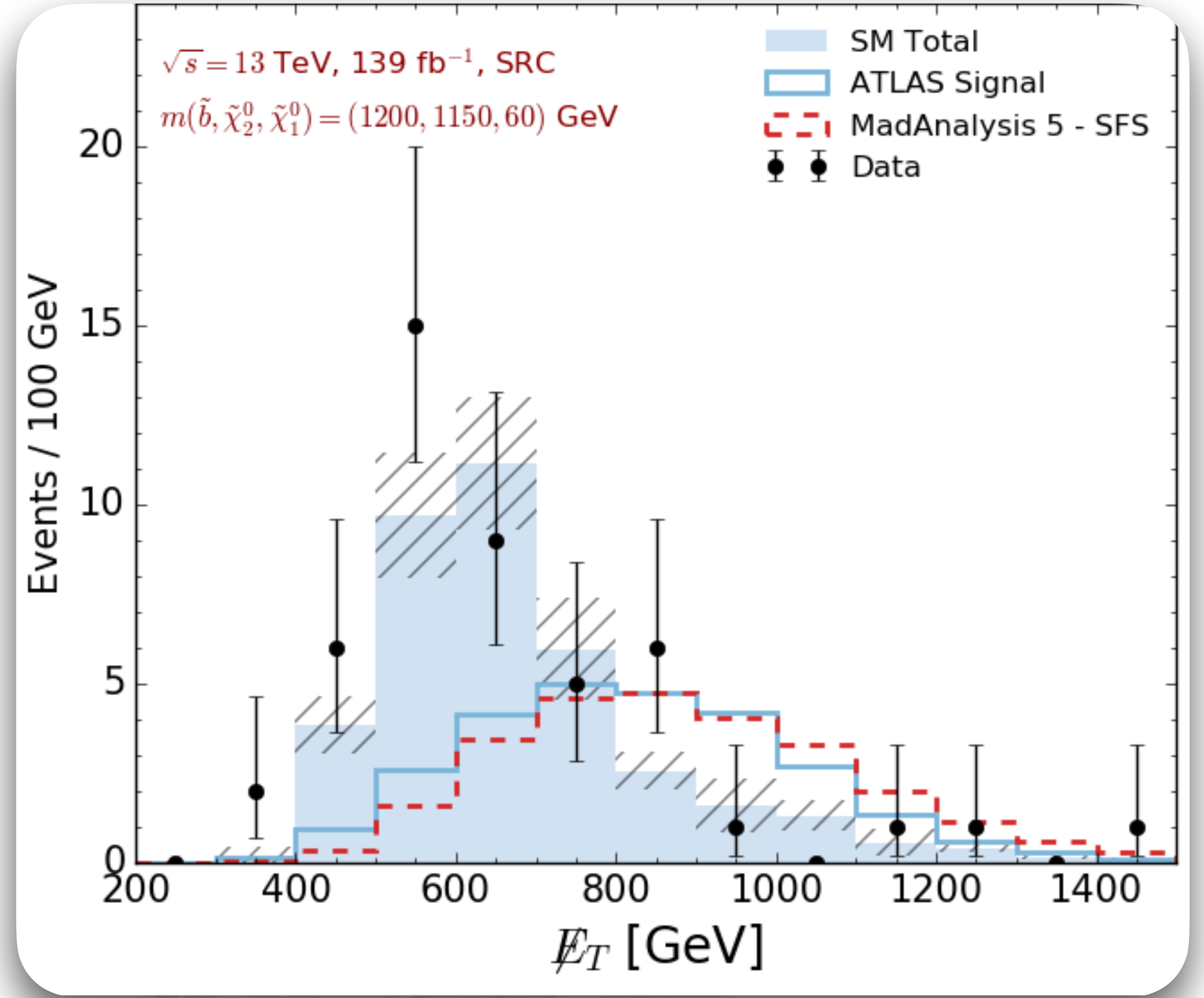


Smeared MC $\oplus$ observed data

Searches for new physics

Lagrangian

Image credit: Sherpa



Smeared MC $\oplus$ observed data

I need to construct a statistical model,

$$\mathscr{L}(\mu, \theta) = \cdots,$$

to test my theory

Spey demystified

`$ pip install spey`

Compute p-values

Generate samples

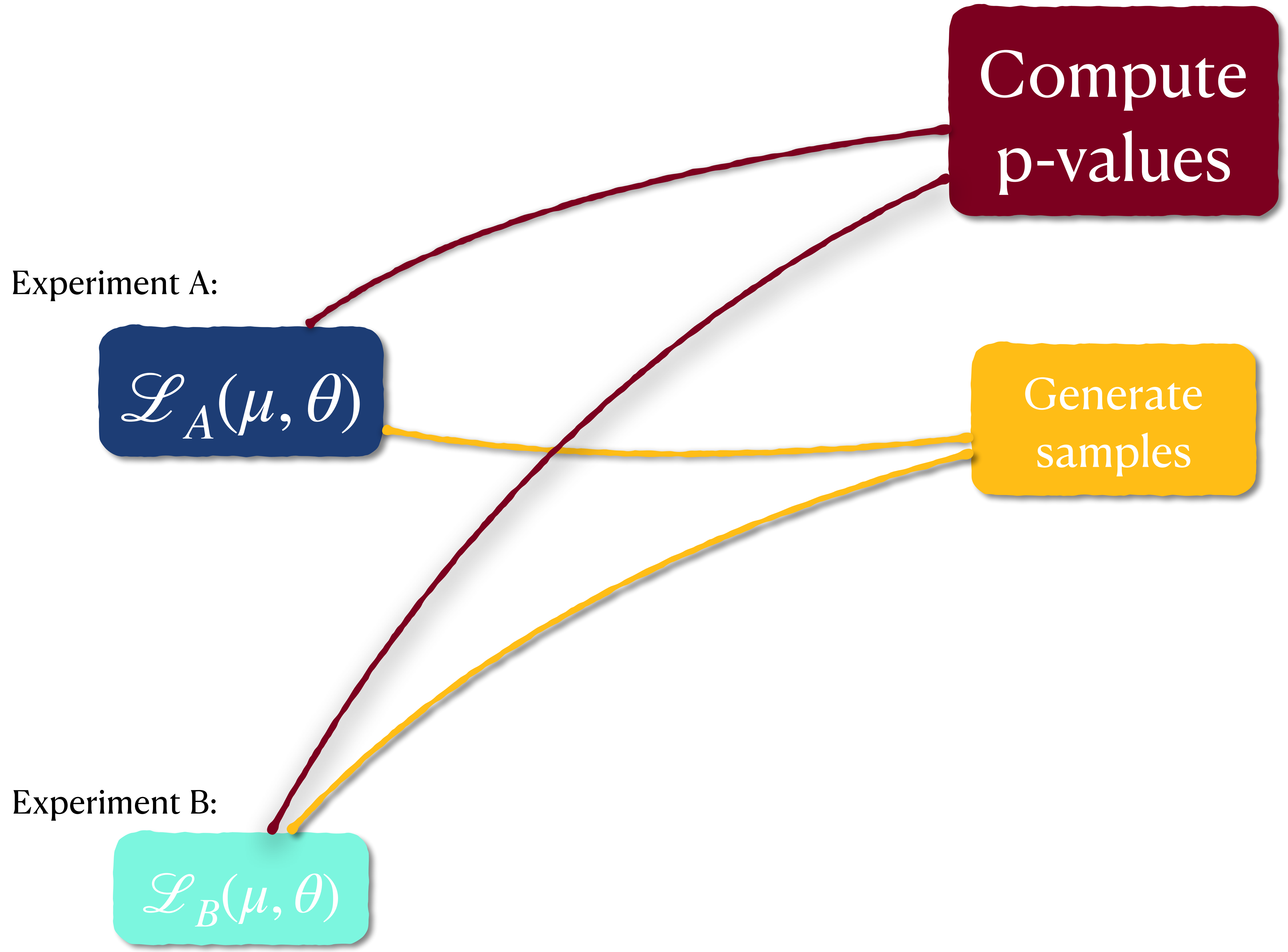
- Exclusion limits
- χ^2 analysis
- POI upper limits

And more…

Spey demystified

JYA, SciPost '24

`$ pip install spey`



- Exclusion limits
- χ^2 analysis
- POI upper limits

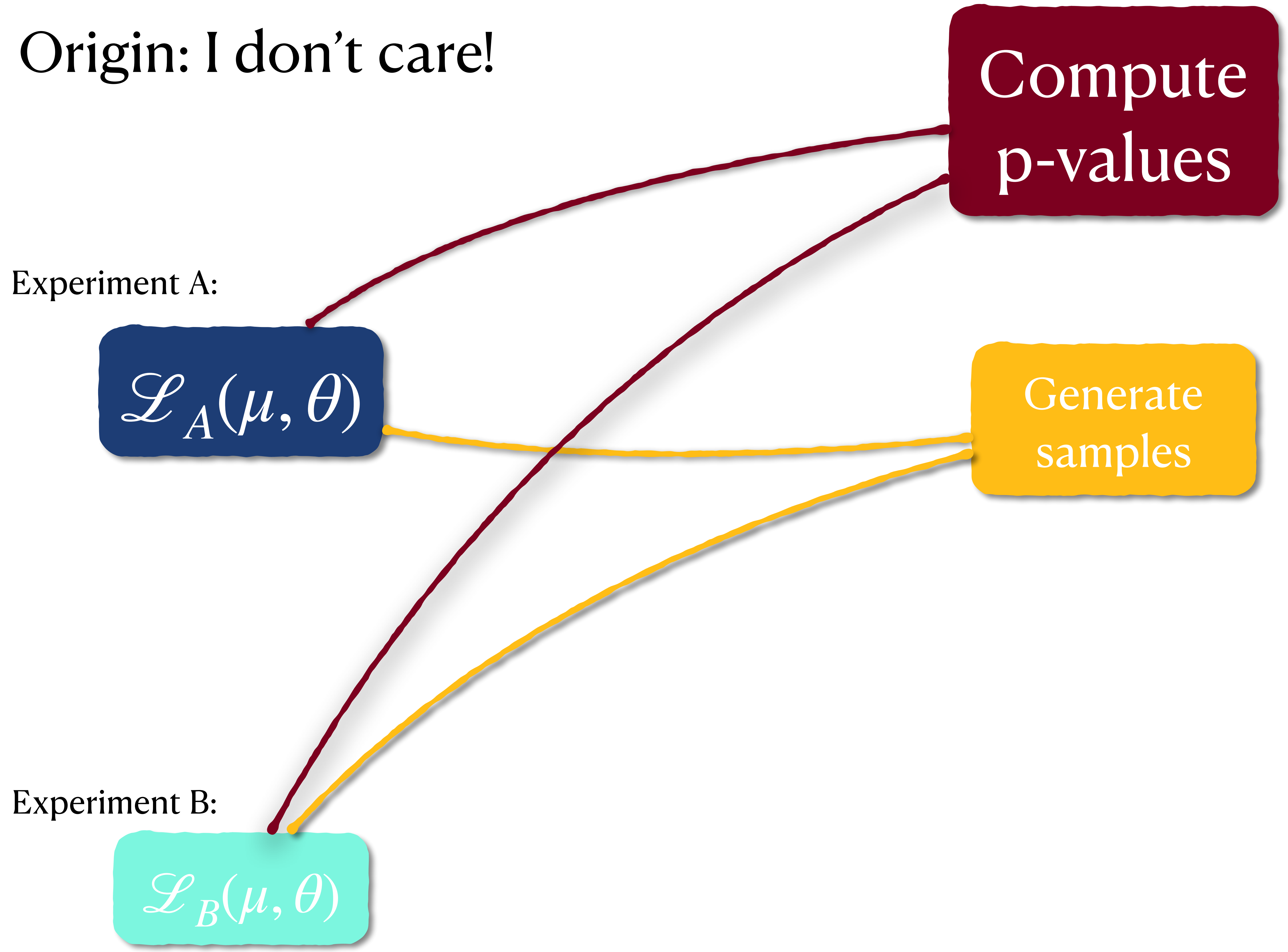
And more...

Spey demystified

JYA, SciPost ’24

`$ pip install spey`

Origin: I don’t care!



- Exclusion limits
- χ^2 analysis
- POI upper limits

And more…

Spey demystified

`$ pip install spey`

Origin: I don't care!

Experiment A:

$\mathcal{L}_A(\mu, \theta)$

Experiment B:

$\mathcal{L}_B(\mu, \theta)$

Compute p-values

Generate samples

- Exclusion limits
- χ^2 analysis
- POI upper limits

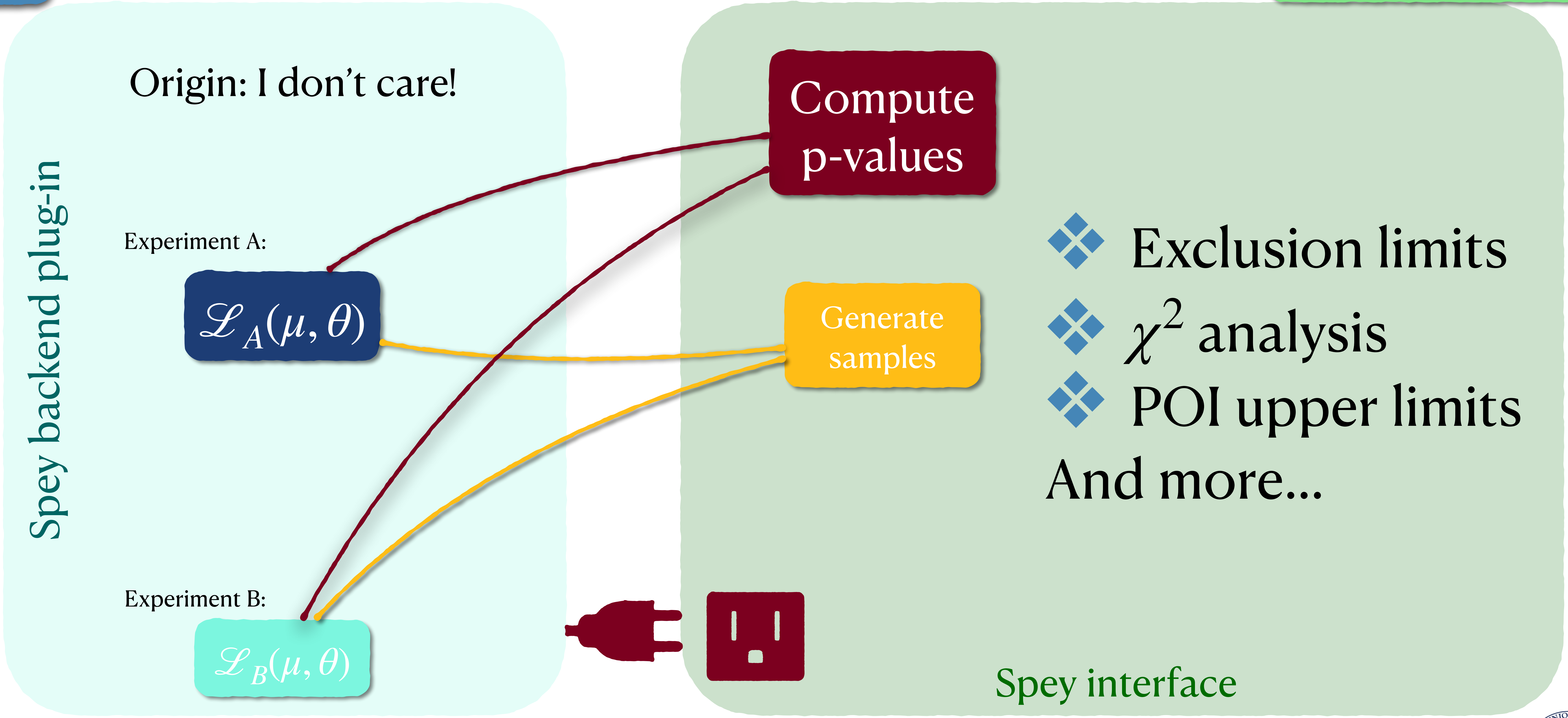
And more...

Spey interface

Spey demystified

JYA, SciPost ’24

`$ pip install spey`



Spey demystified

JYA, SciPost '24

`$ pip install spey`

Spey backend plug-in

Origin: I don't care!

Experiment A:

$\mathscr{L}_A(\mu, \theta)$

Experiment B:

$\mathscr{L}_B(\mu, \theta)$

Compute p-values

Generate samples

Combine

- Exclusion limits
- χ^2 analysis
- POI upper limits

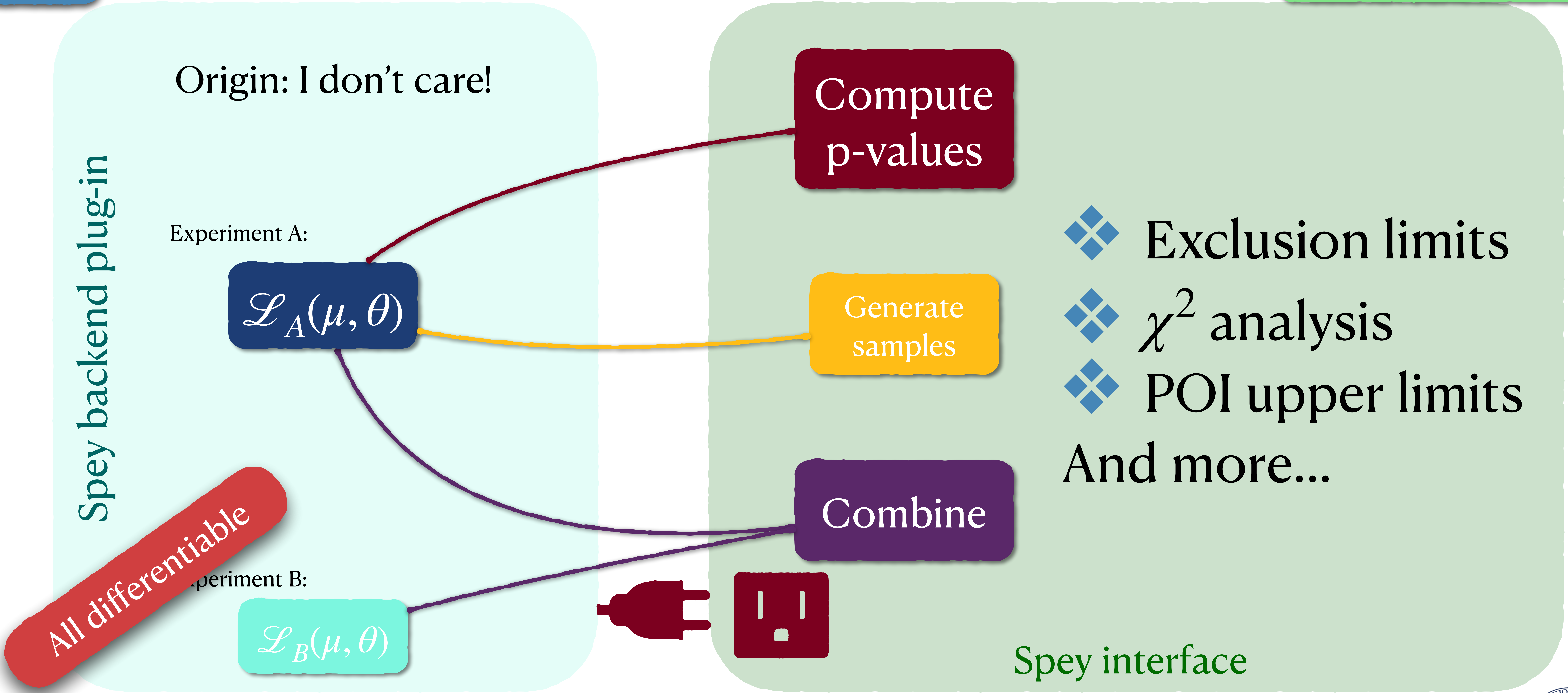
And more...

Spey interface

Spey demystified

JYA, SciPost '24

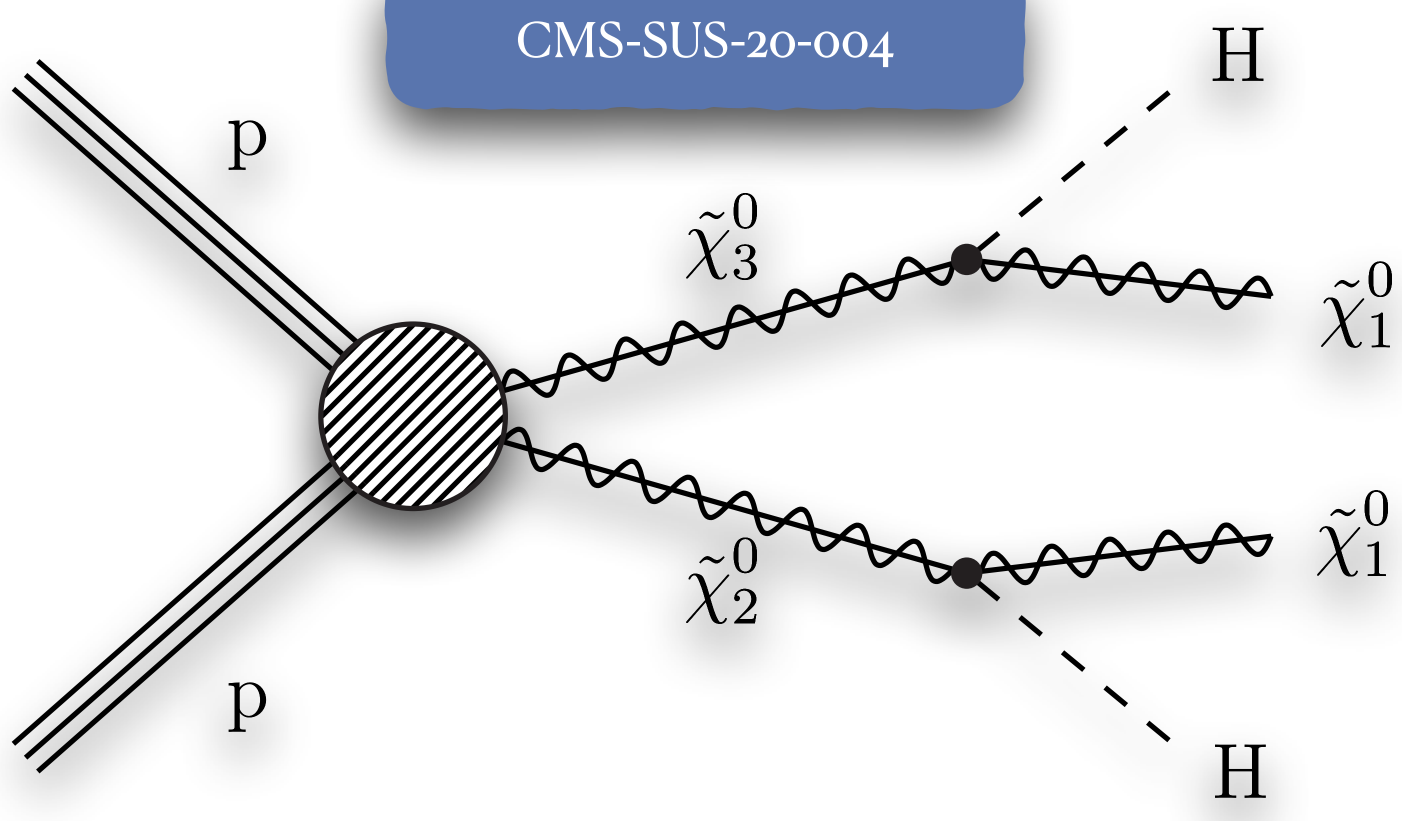
`$ pip install spey`



Improving Simplified Likelihoods

JYA, SciPost '24

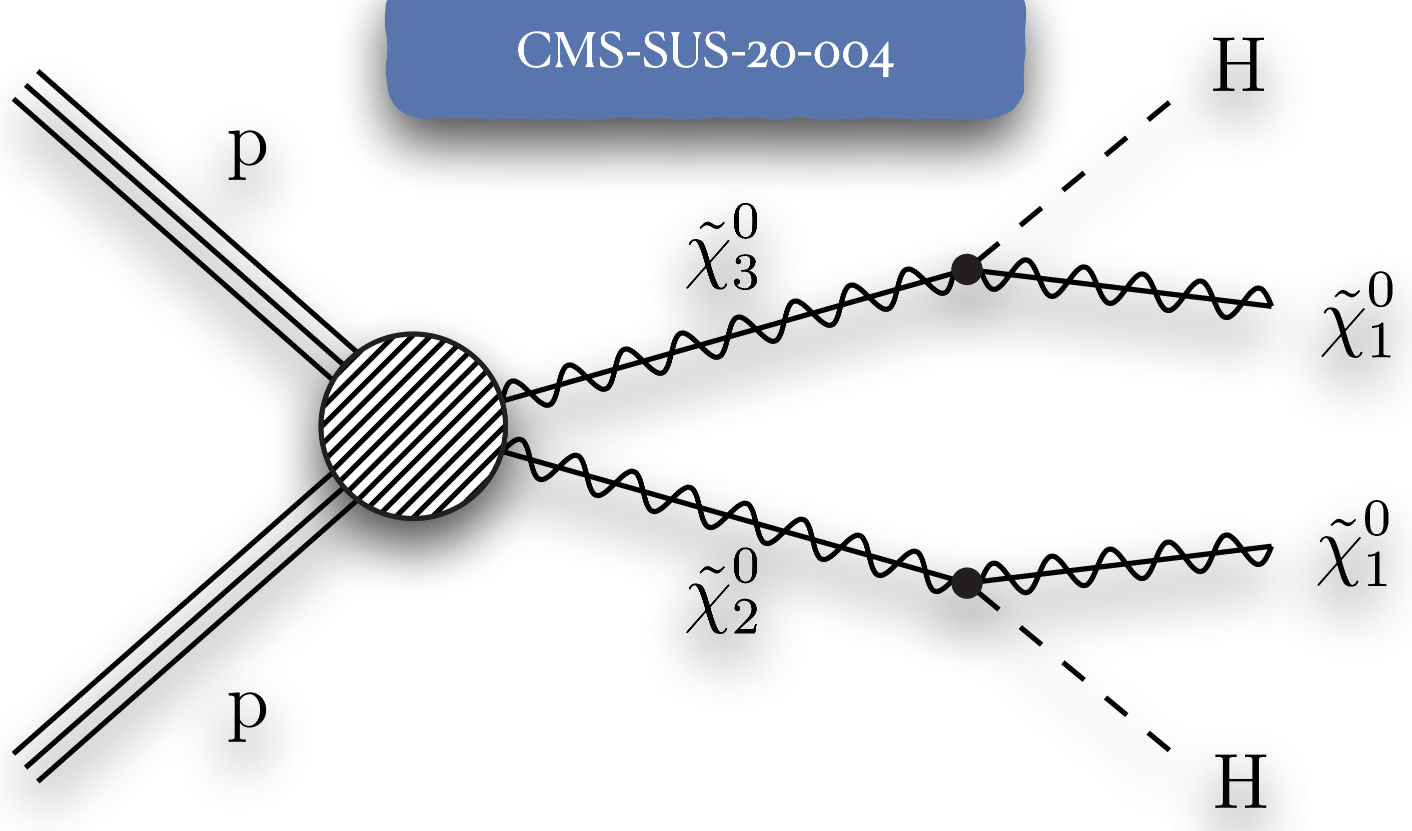
CMS-SUS-20-004



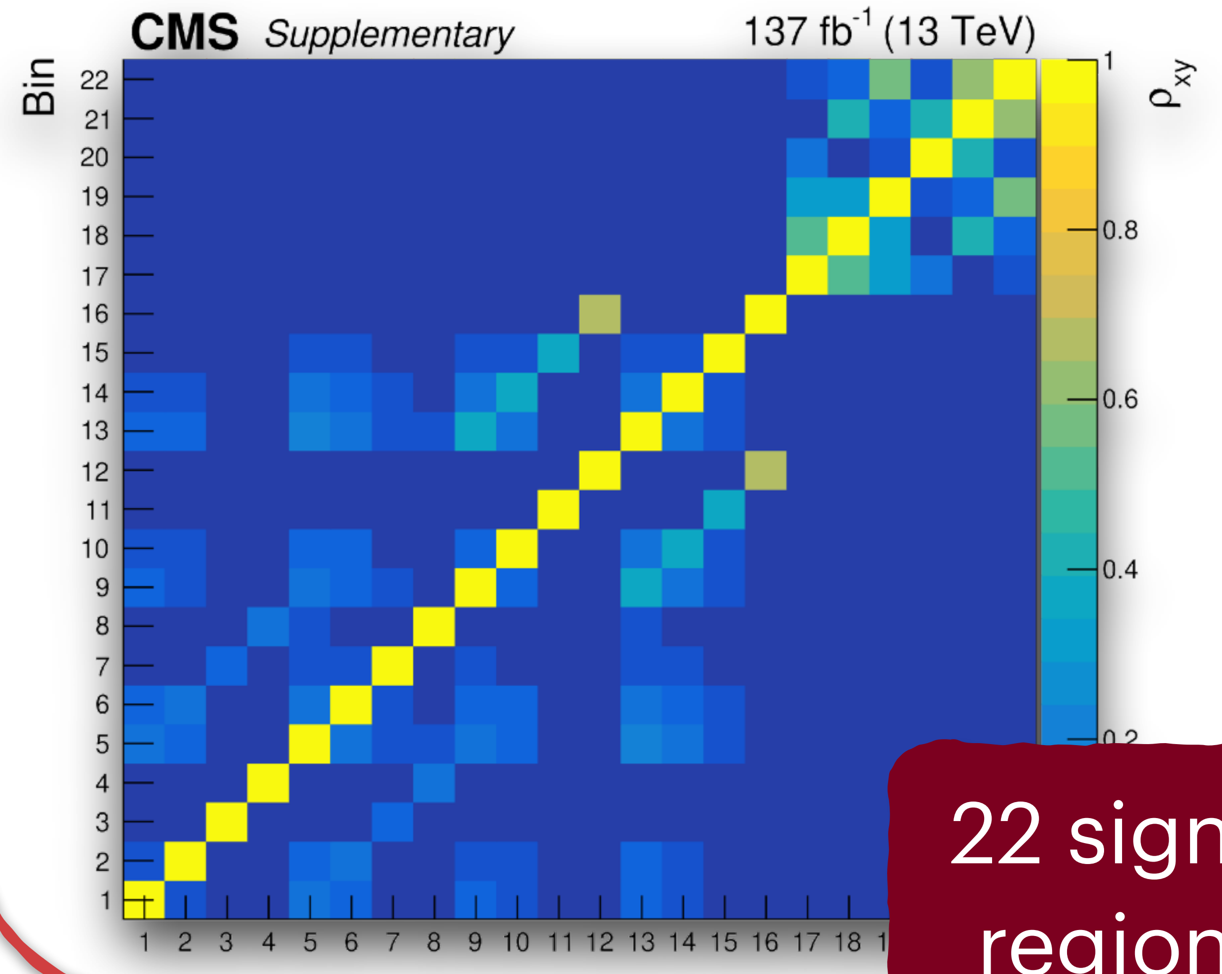
Improving Simplified Likelihoods

JYA, SciPost '24

CMS-SUS-20-004



Correlation Matrix

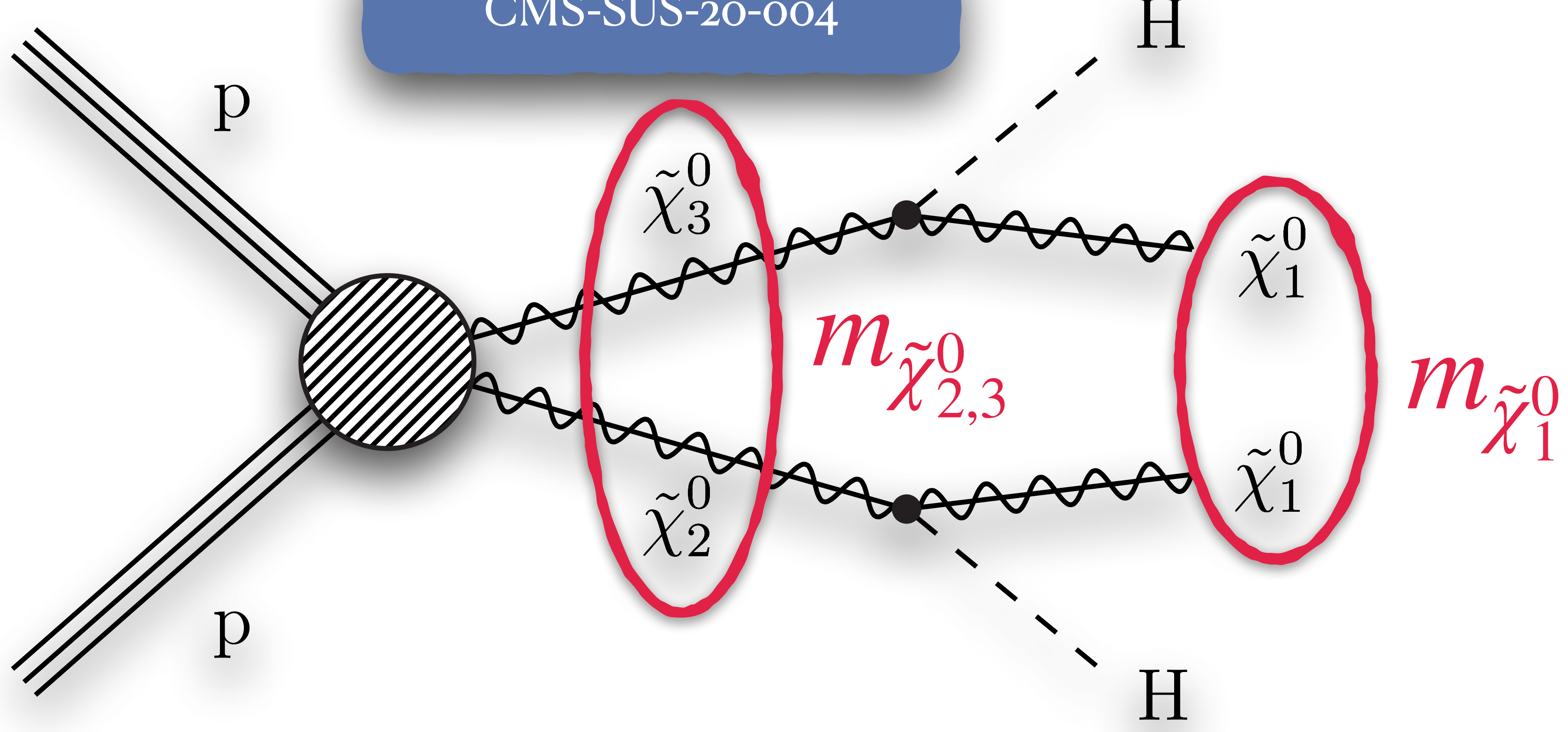


22 signal regions

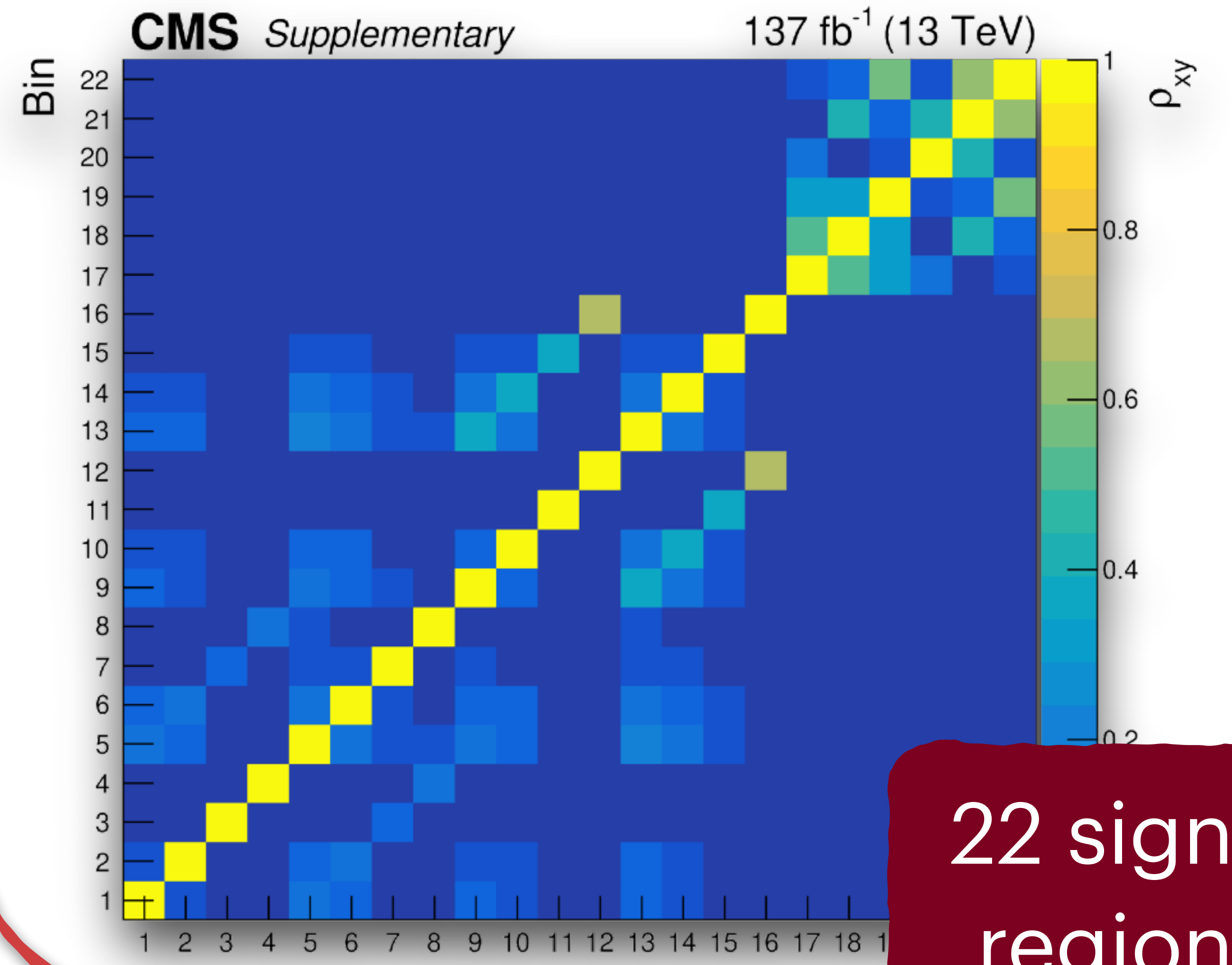
Improving Simplified Likelihoods

JYA, SciPost '24

CMS-SUS-20-004

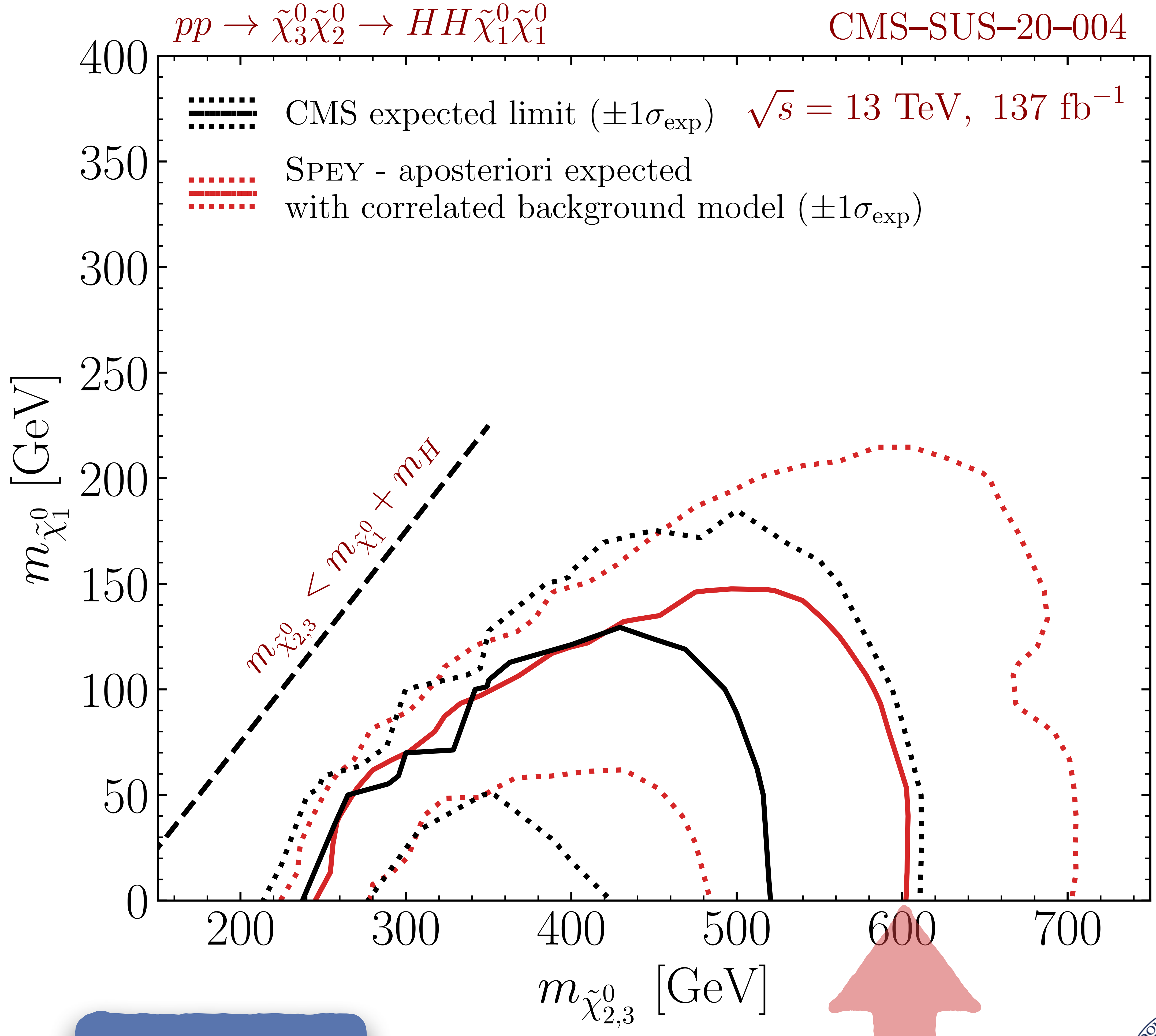


Correlation Matrix



22 signal regions

`backend="default_pdf.correlated_background"`

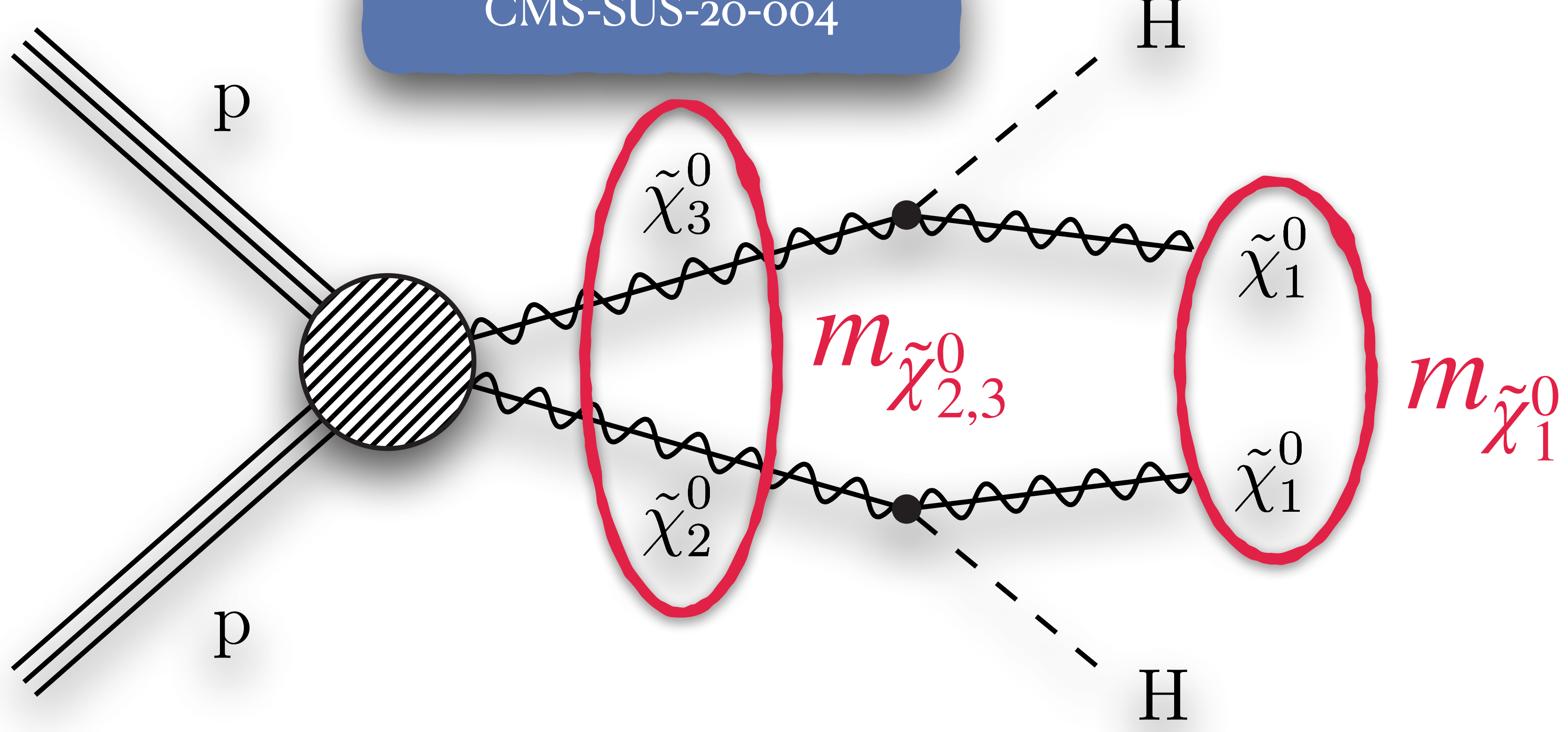


CMS-NOTE-2017-001

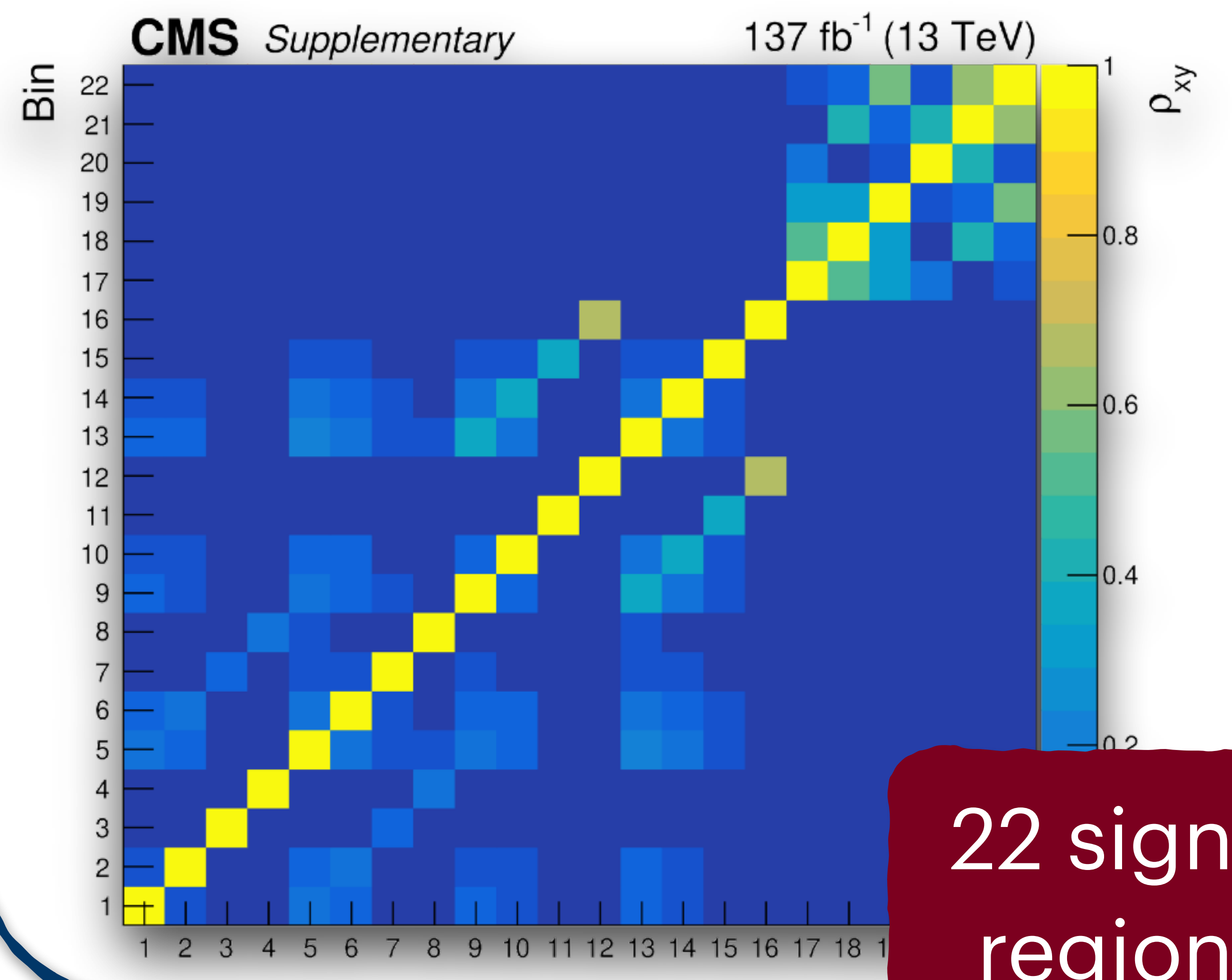
Improving Simplified Likelihoods

JYA, SciPost '24

CMS-SUS-20-004



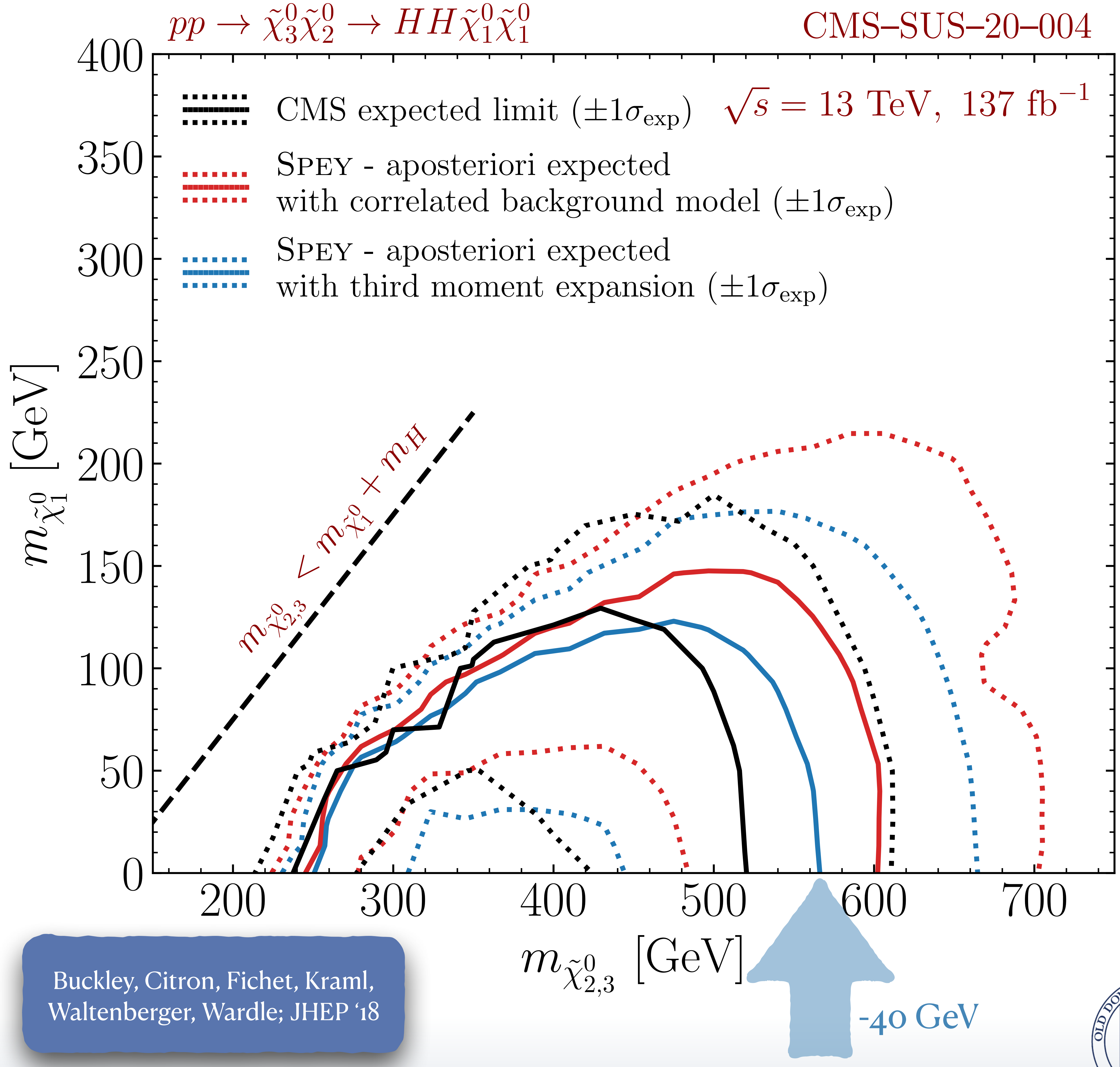
Correlation Matrix



22 signal regions

$\oplus$ Third Moments

`backend="default_pdf.third_moment_expansion"`

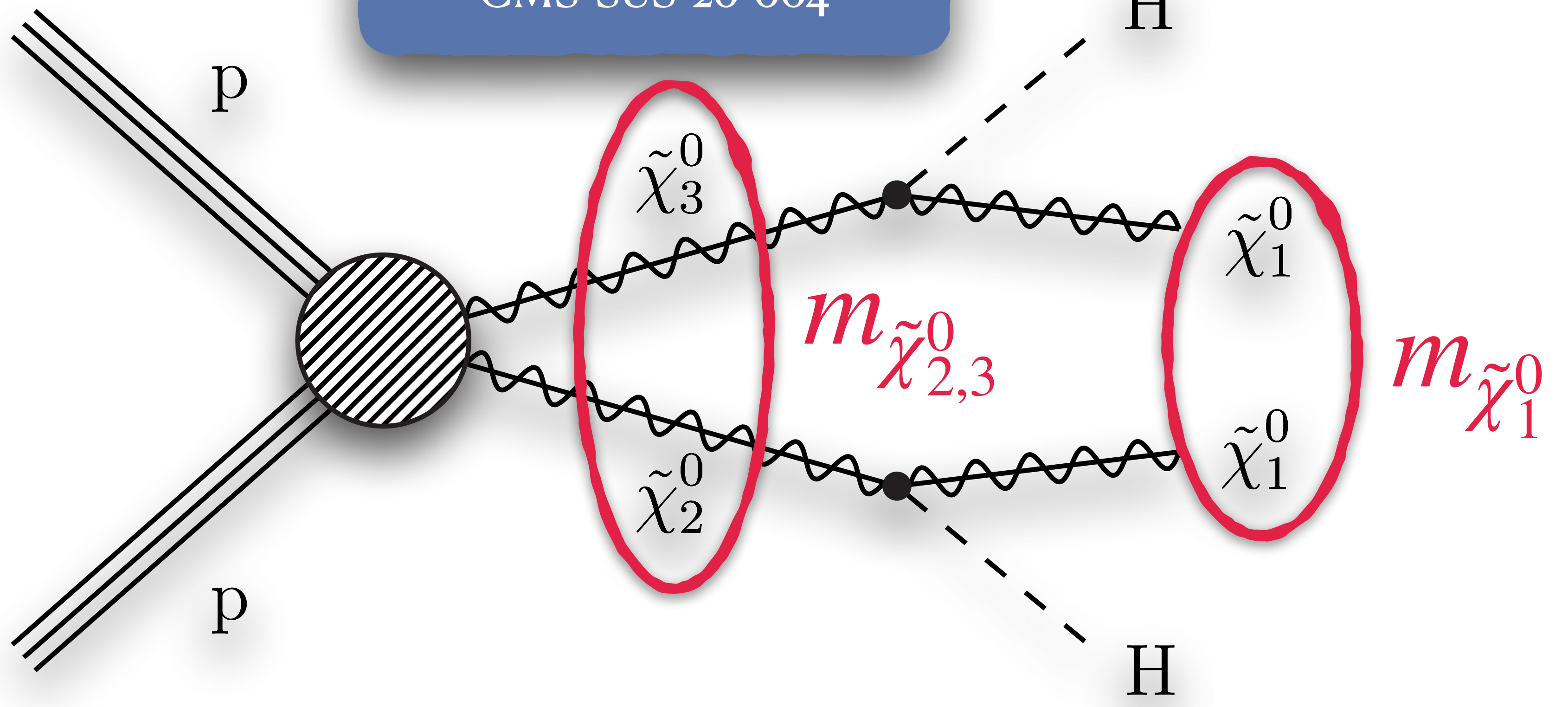


Buckley, Citron, Fichet, Kraml, Waltenberger, Wardle; JHEP '18

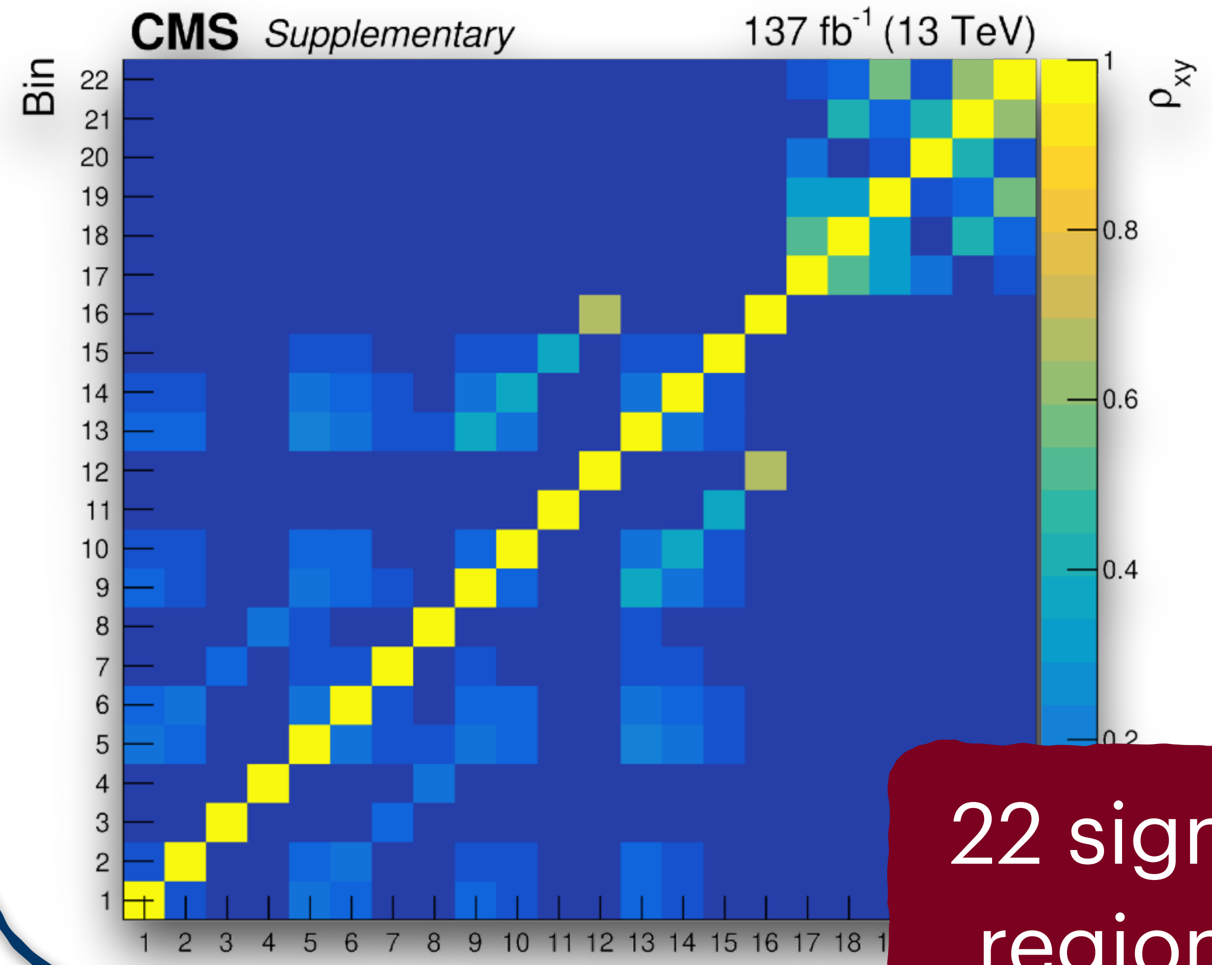
Improving Simplified Likelihoods

JYA, SciPost '24

CMS-SUS-20-004



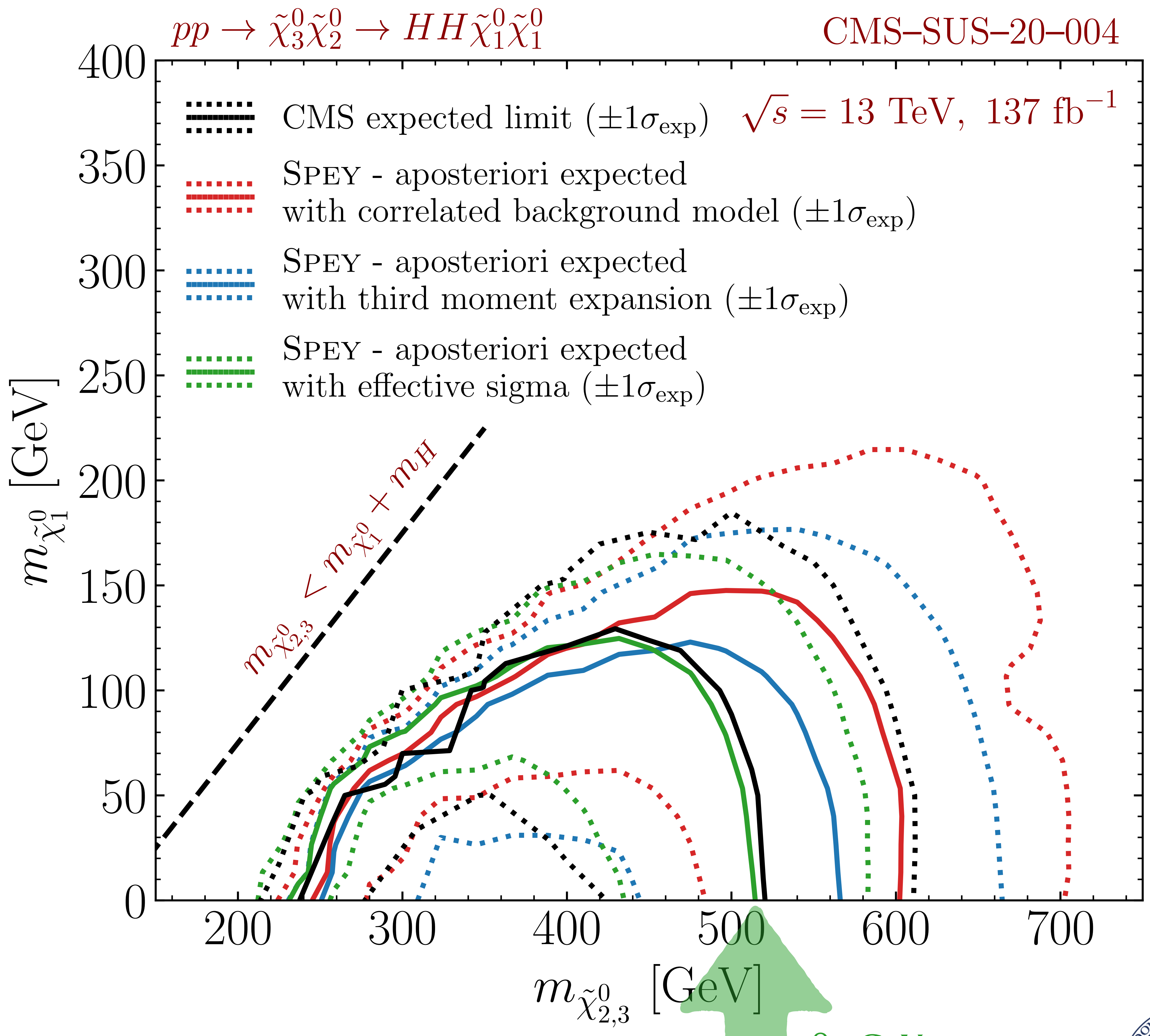
Correlation Matrix



22 signal regions

⊕ Asymmetric uncertainties

`backend="default_pdf.effective_sigma"`

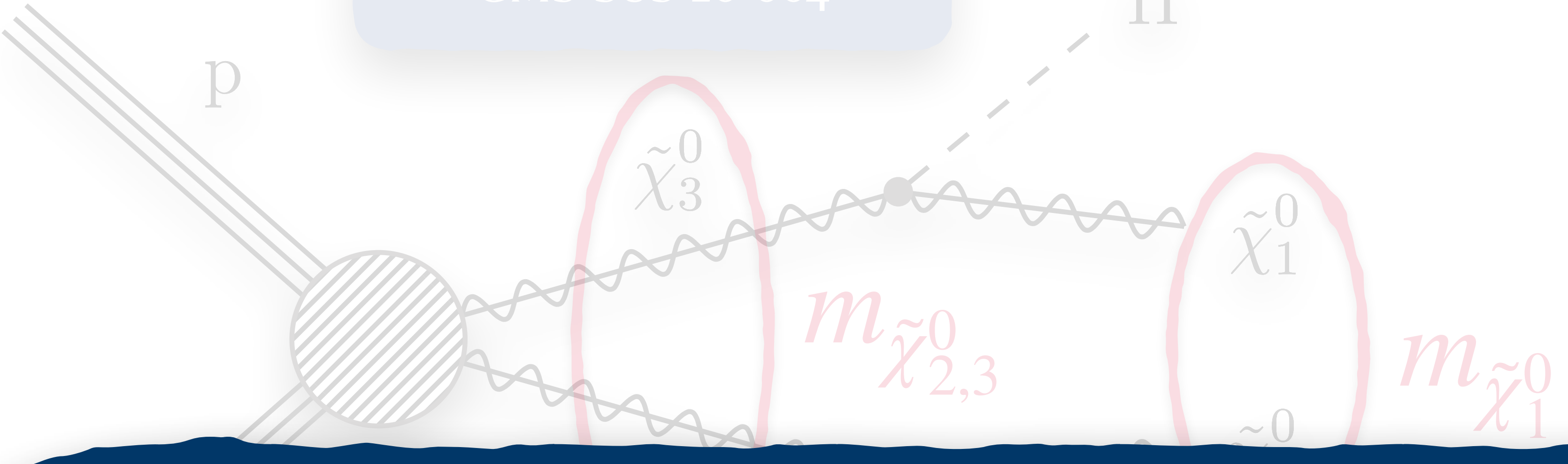


-80 GeV

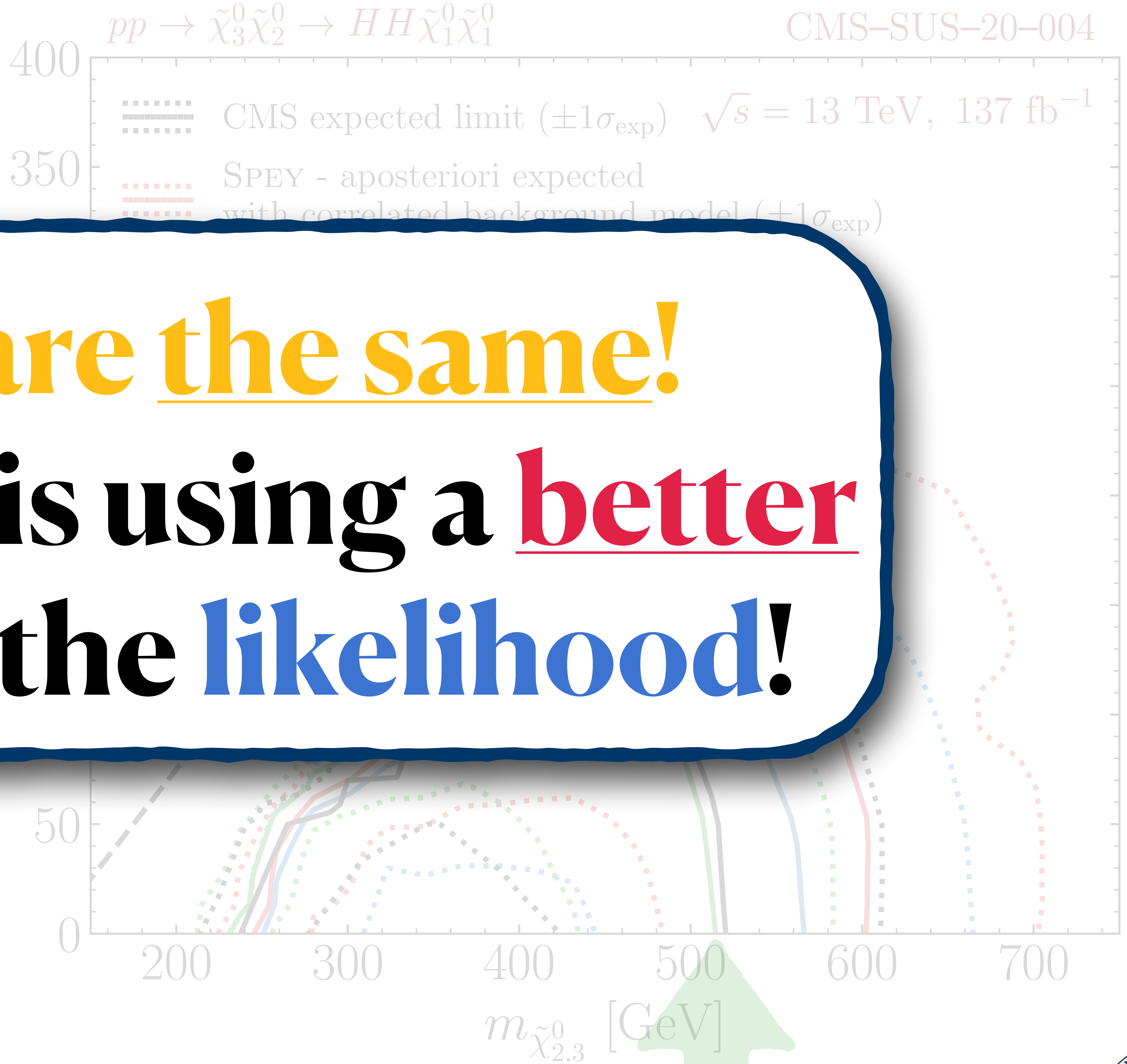
Improving Simplified Likelihoods

JYA, SciPost '24

CMS-SUS-20-004



`backend="default_pdf.effective_sigma"`



Data and signal are the same!

The only difference is using a better approximation for the likelihood!

Correlat...

uncertainties

22 signal regions

-80 GeV

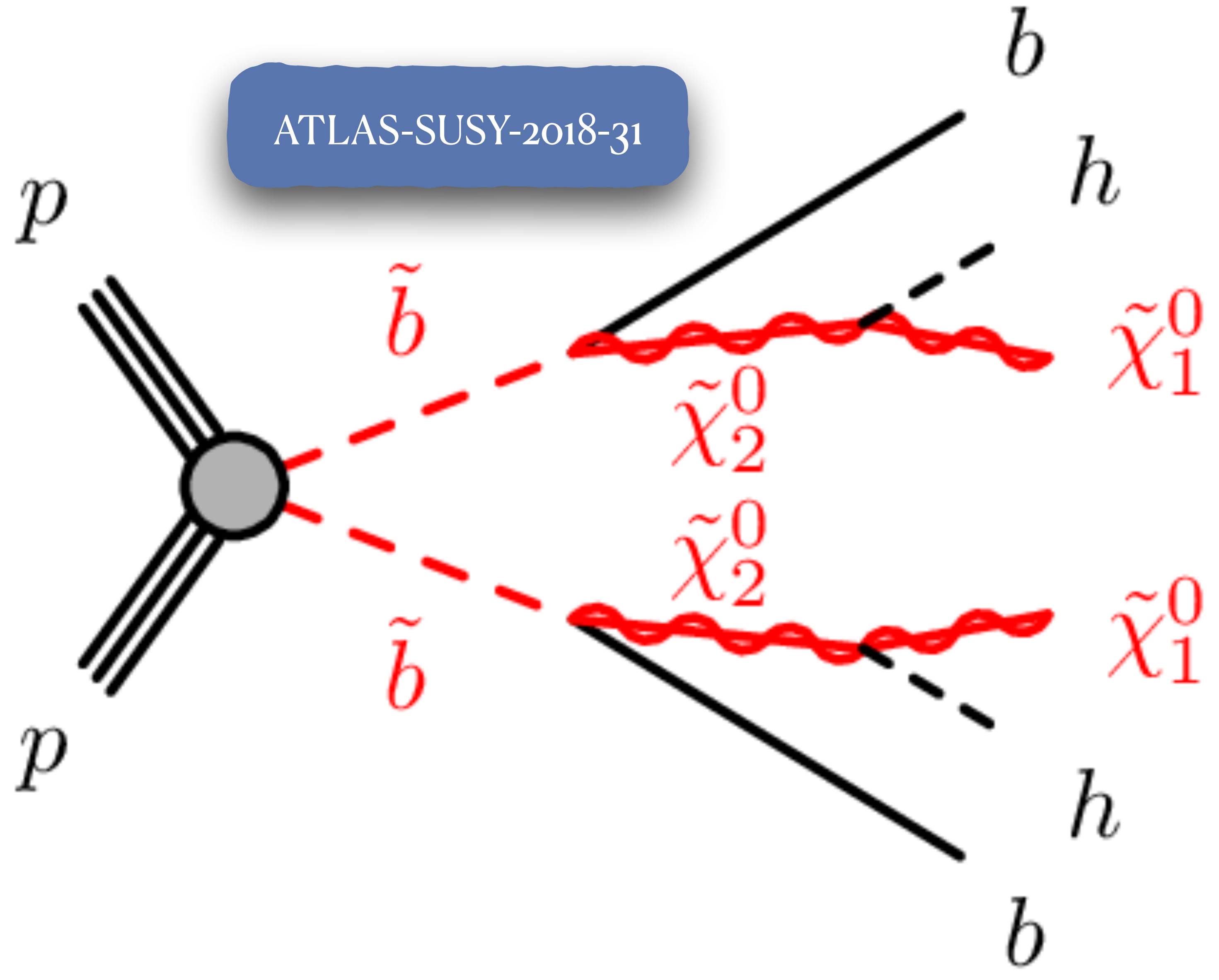
Jefferson Lab
EXPLORING THE NATURE OF MATTER

Full Likelihoods

`$ pip install spey-pyhf`

JYA, SciPost '24

ATLAS-SUSY-2018-31



HEPData

Full likelihoods from ATLAS

ATL-PHYS-PUB-2019-029

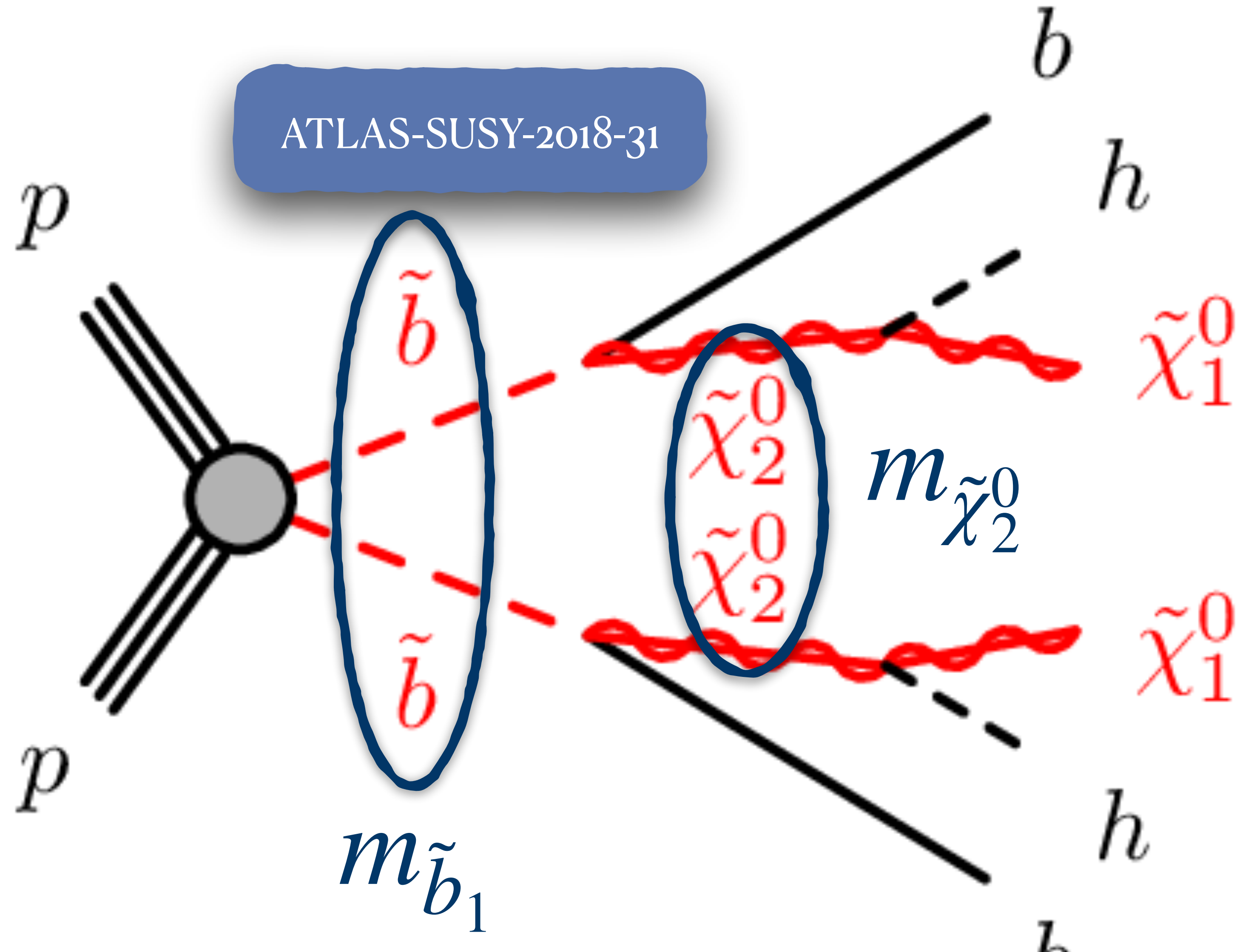
Full Likelihoods

$ pip install spey-pyhf

JYA, SciPost '24

This analysis includes three distinct super regions!

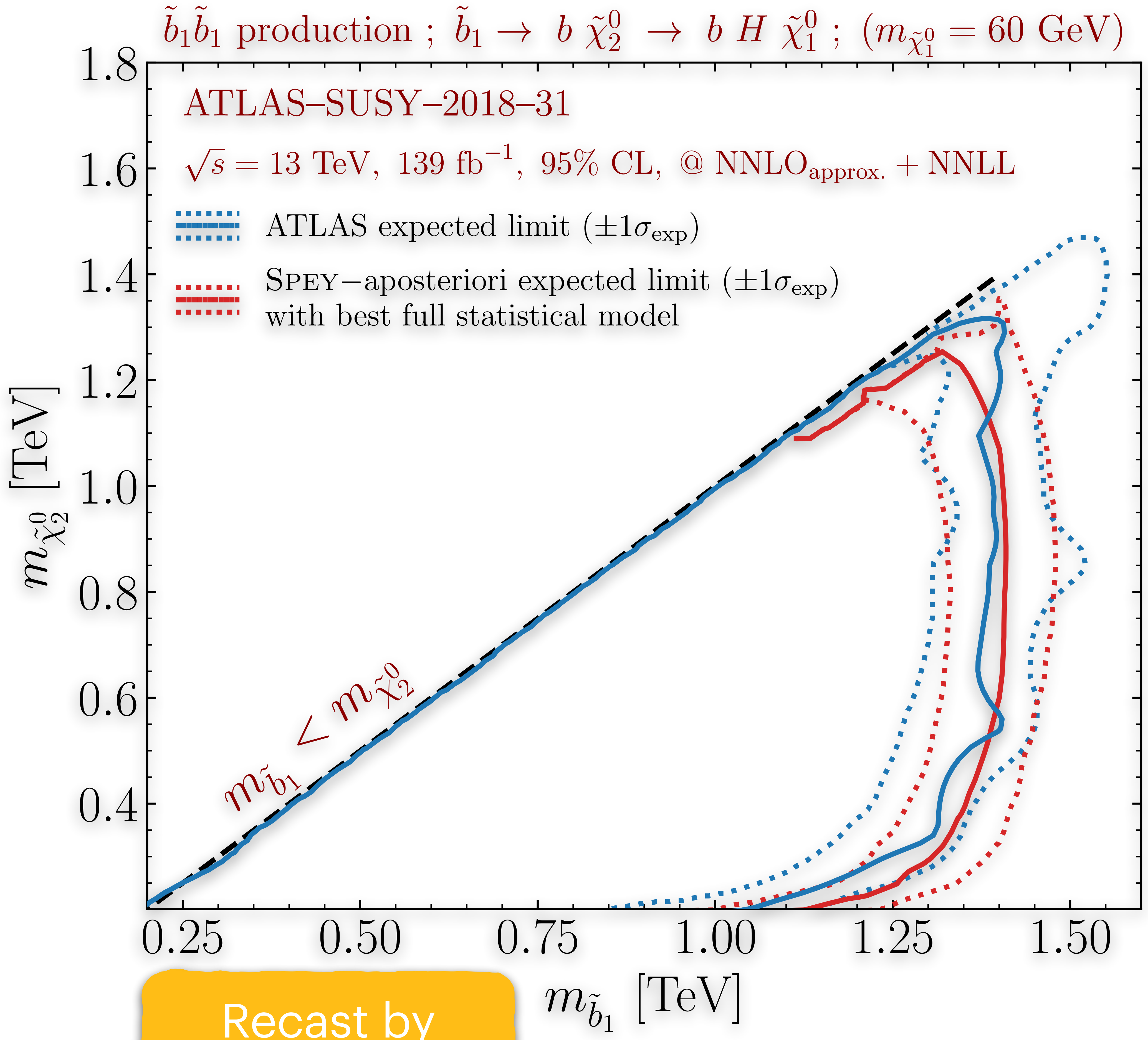
ATLAS-SUSY-2018-31



HEPData

Full likelihoods from ATLAS

ATL-PHYS-PUB-2019-029



Recast by MadAnalysis 5

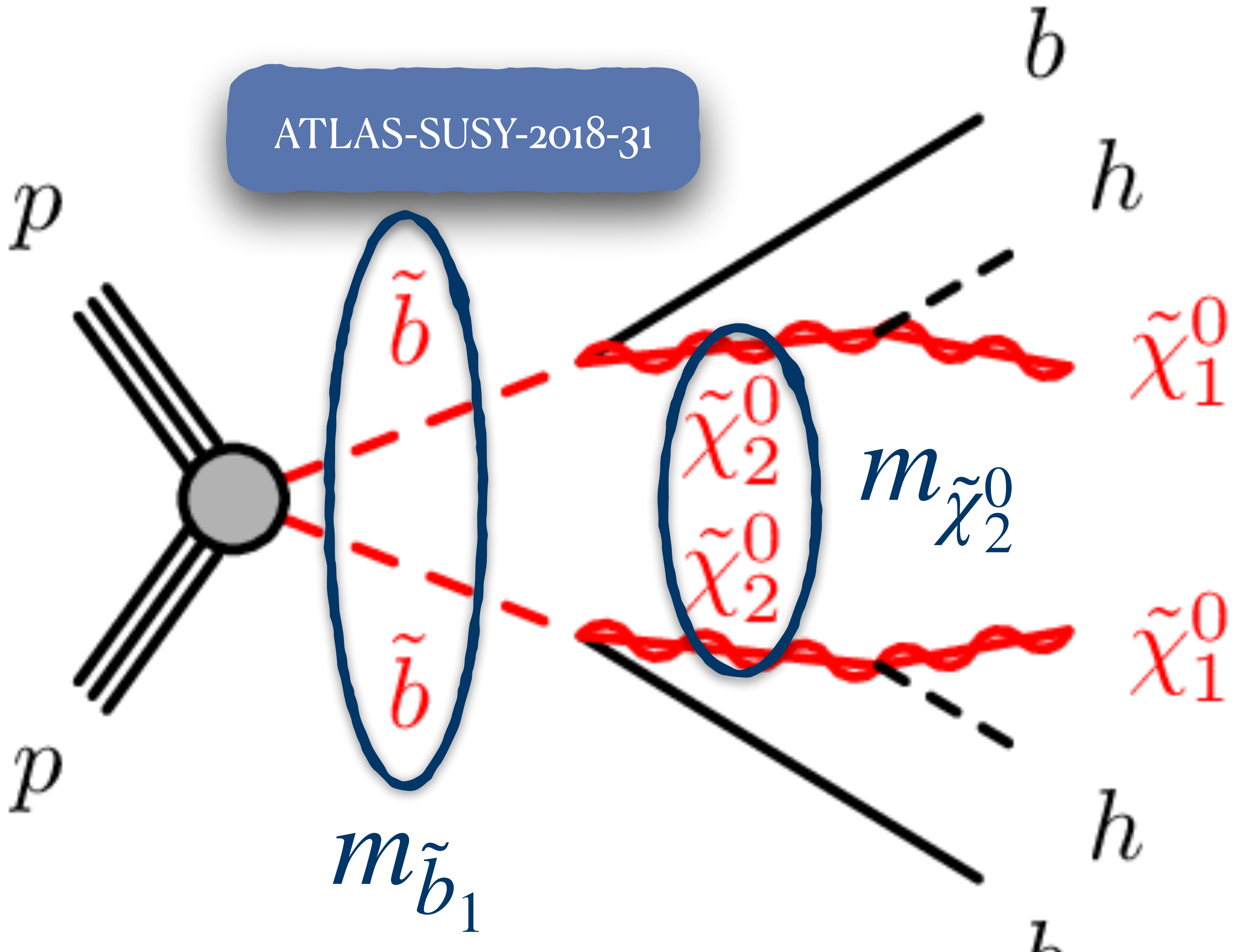
Full Likelihoods

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

JYA, SciPost '24

This analysis includes three distinct super regions!

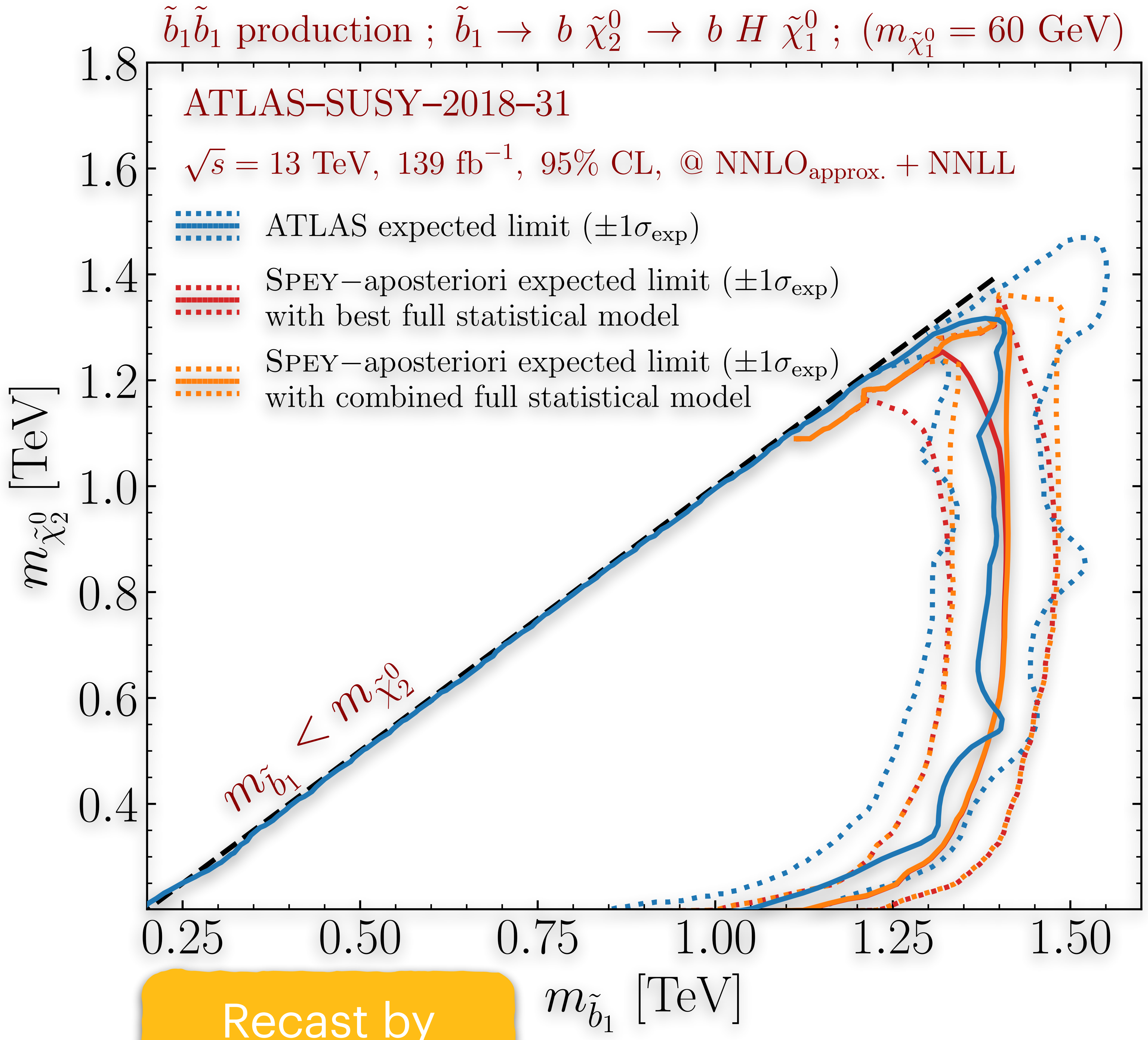
ATLAS-SUSY-2018-31



Full likelihoods from ATLAS

ATL-PHYS-PUB-2019-029

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



Recast by MadAnalysis 5

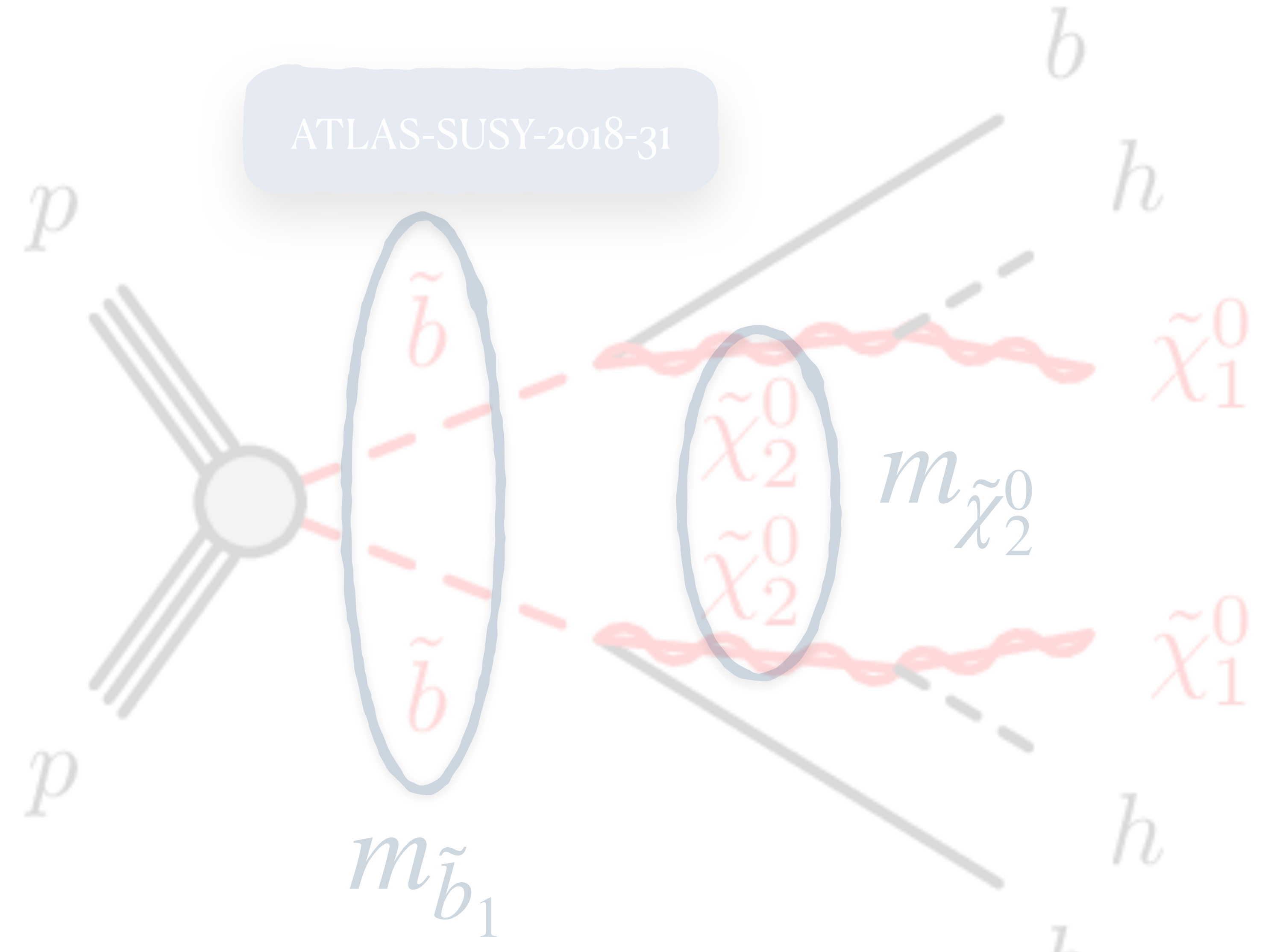
Full Likelihoods

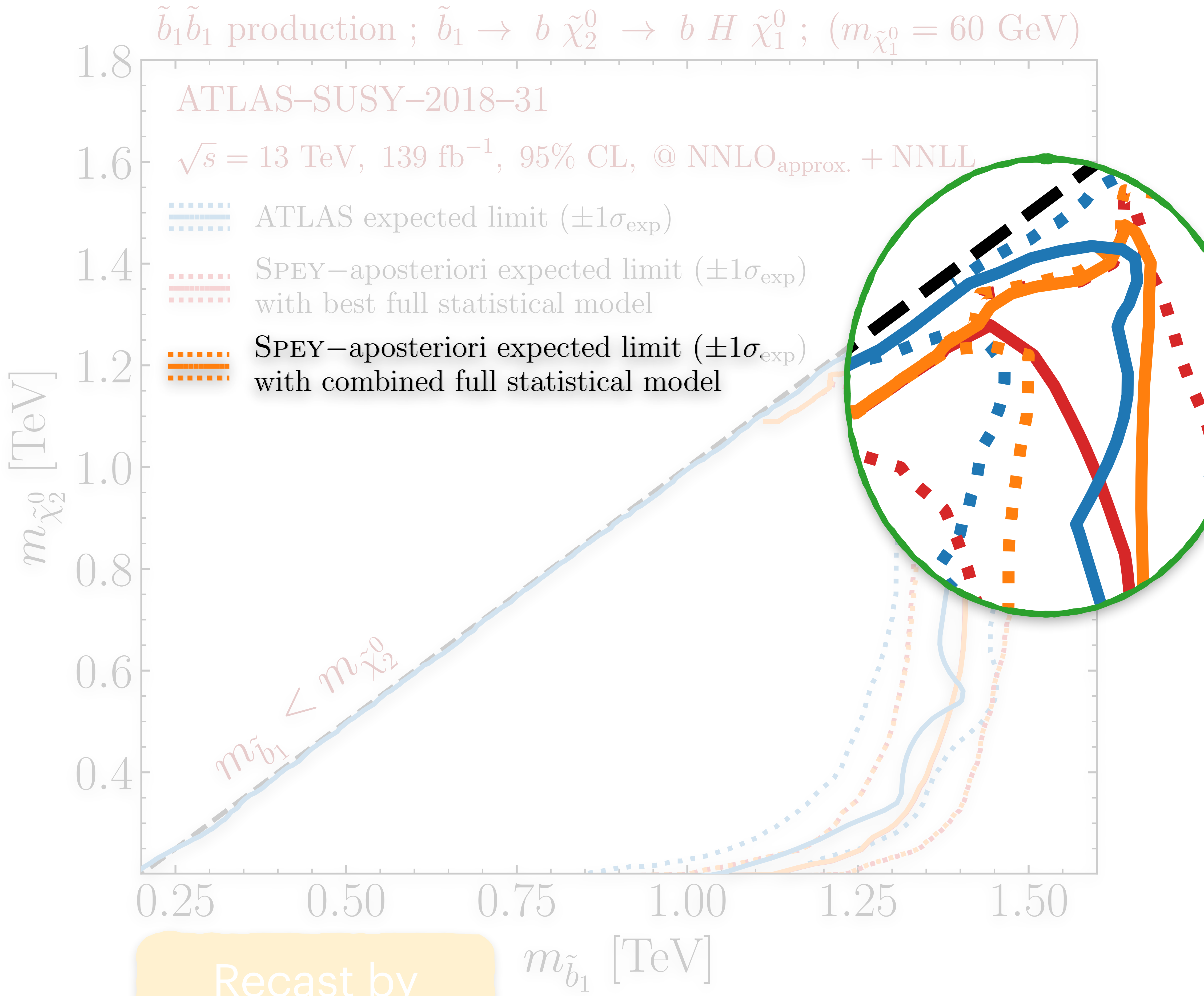
JYA, SciPost '24

`$ pip install spey-pyhf`

This analysis includes three distinct super regions!

ATLAS-SUSY-2018-31





Recast by MadAnalysis 5

HEPData

Full likelihoods from ATLAS

pyhf ∂ifferentiable ℒikelihoods

ATL-PHYS-PUB-2019-029

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

Towards global sensitivity

JYA, SciPost '24

$\mathscr{L}_{\text{ATLAS}}$

$\mathscr{L}_{\text{CMS}}$

Full likelihood

Simplified likelihood with
effective sigma model

Towards global sensitivity

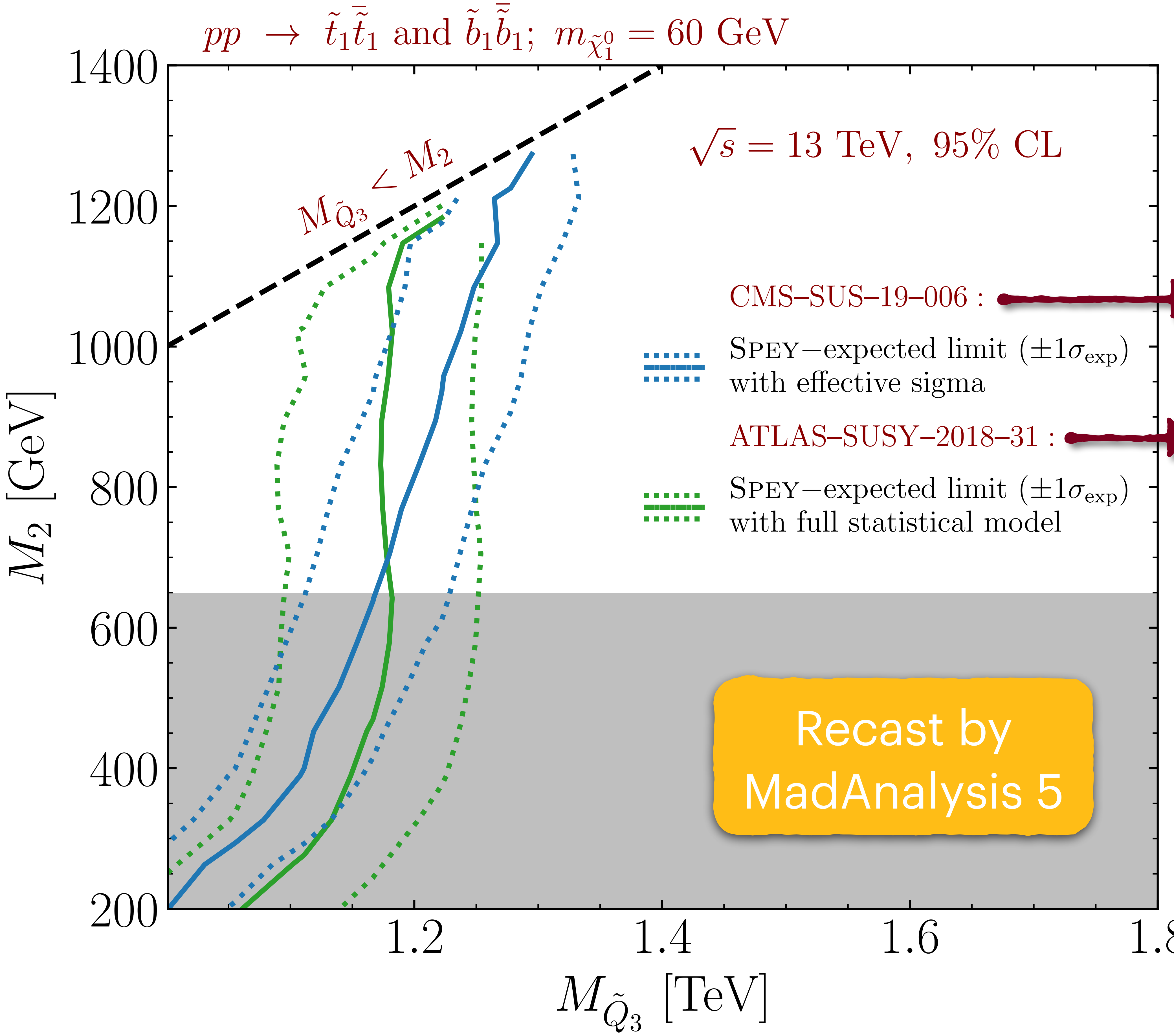
JYA, SciPost '24

$\mathscr{L}_{\text{ATLAS}}$ $\mathscr{L}_{\text{CMS}}$

Full likelihood

Simplified likelihood with effective sigma model

MSSM: $M_1 = M_2 = M_3 = M_{\tilde{Q}}$ at GUT scale



Original analysis:

$pp \to \tilde{g}\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$

$pp \to \tilde{b}_1\bar{\tilde{b}}_1(\to bh\tilde{\chi}_2^0)$

Towards global sensitivity

JYA, SciPost '24

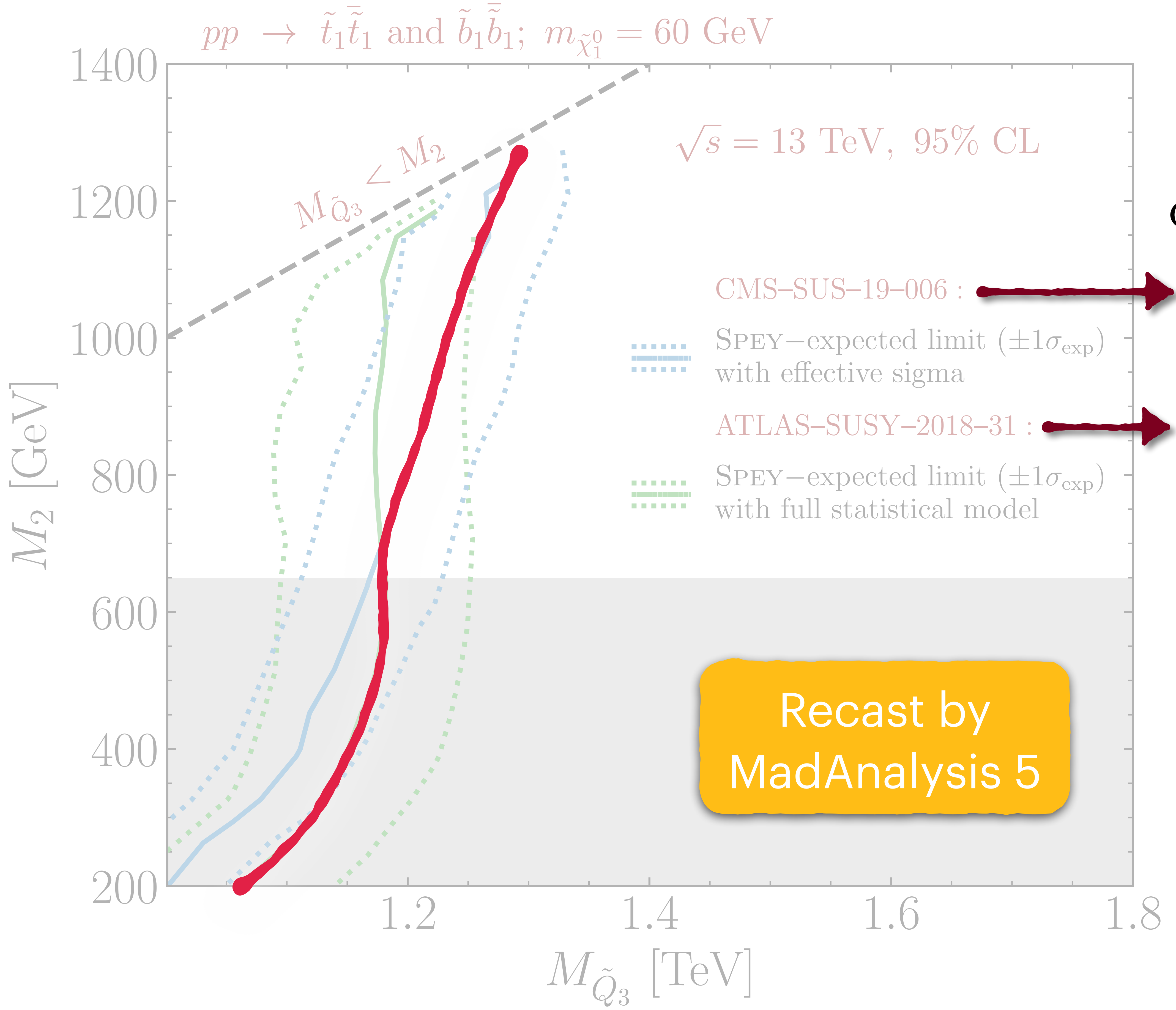
$\mathcal{L}_{\text{ATLAS}}$

$\mathcal{L}_{\text{CMS}}$

Full likelihood

Simplified likelihood with effective sigma model

MSSM: $M_1 = M_2 = M_3 = M_{\tilde{Q}}$ at GUT scale



Original analysis:

$pp \to \tilde{g}\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$

$pp \to \tilde{b}_1\bar{\tilde{b}}_1(\to bh\tilde{\chi}_2^0)$

Towards global sensitivity

JYA, SciPost '24

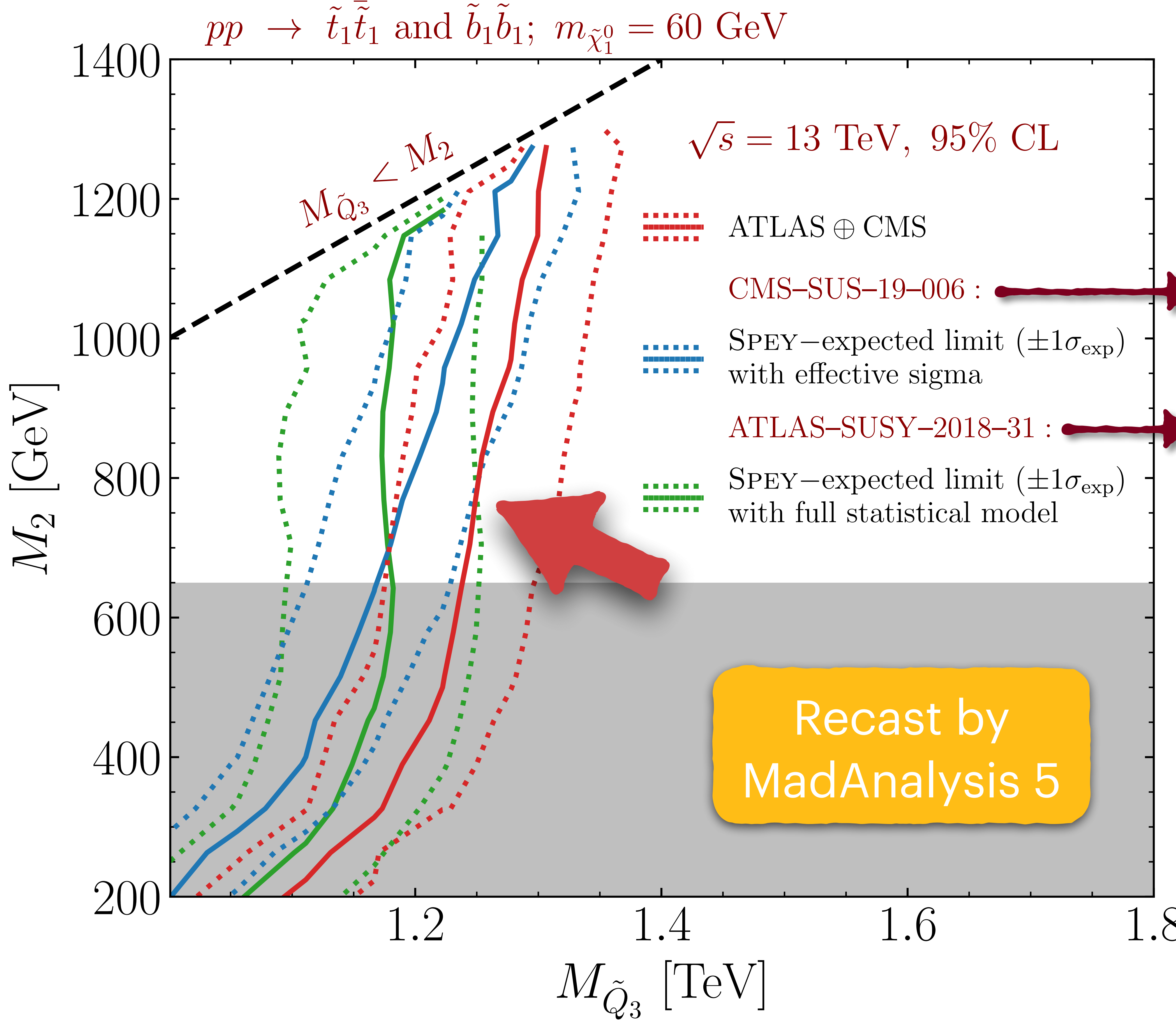
$$\mathscr{L}' = \mathscr{L}_{\text{ATLAS}} \oplus \mathscr{L}_{\text{CMS}}$$

Full likelihood

Simplified likelihood with effective sigma model

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

MSSM: $M_1 = M_2 = M_3 = M_{\tilde{Q}}$ at GUT scale



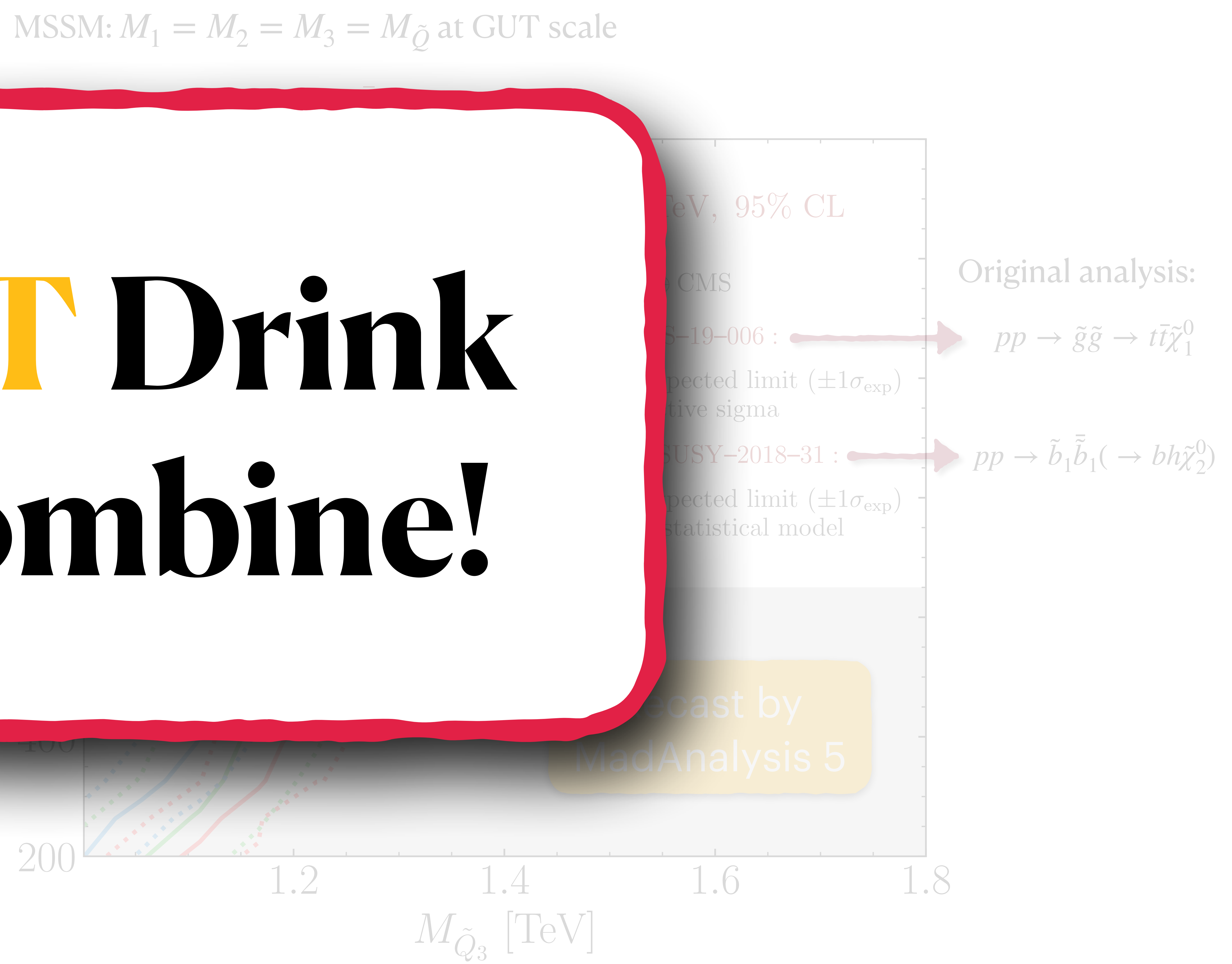
Original analysis:

$pp \to \tilde{g}\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$

$pp \to \tilde{b}_1\bar{\tilde{b}}_1(\to bh\tilde{\chi}_2^0)$

Towards global sensitivity

JYA, SciPost '24



$\mathscr{L}' = \mathscr{L}_{\mathrm{ATL}}$ ...

Full likelihood

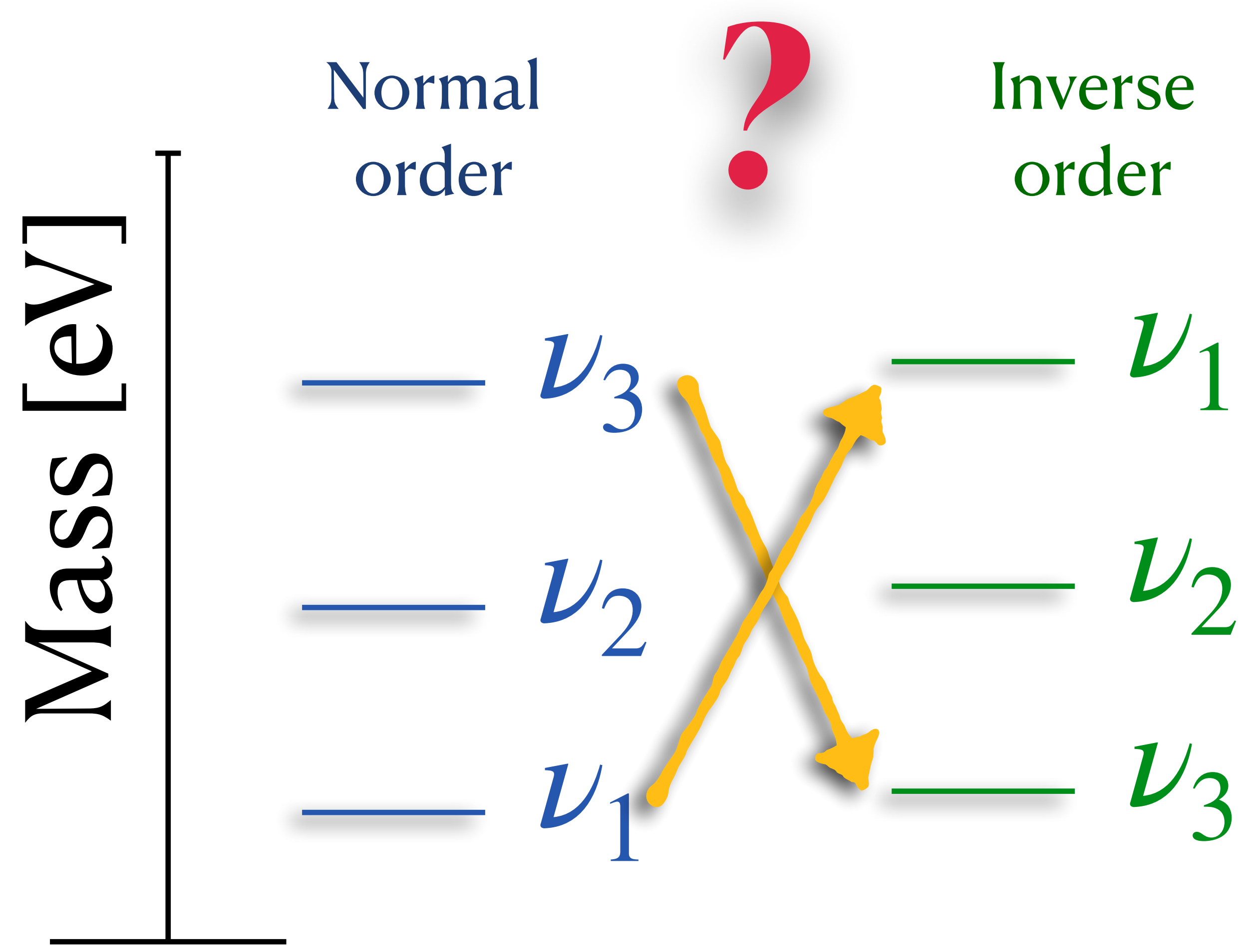
Simp... eff...

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

Do NOT Drink and Combine!

Beyond LHC: Neutrino mass ordering problem

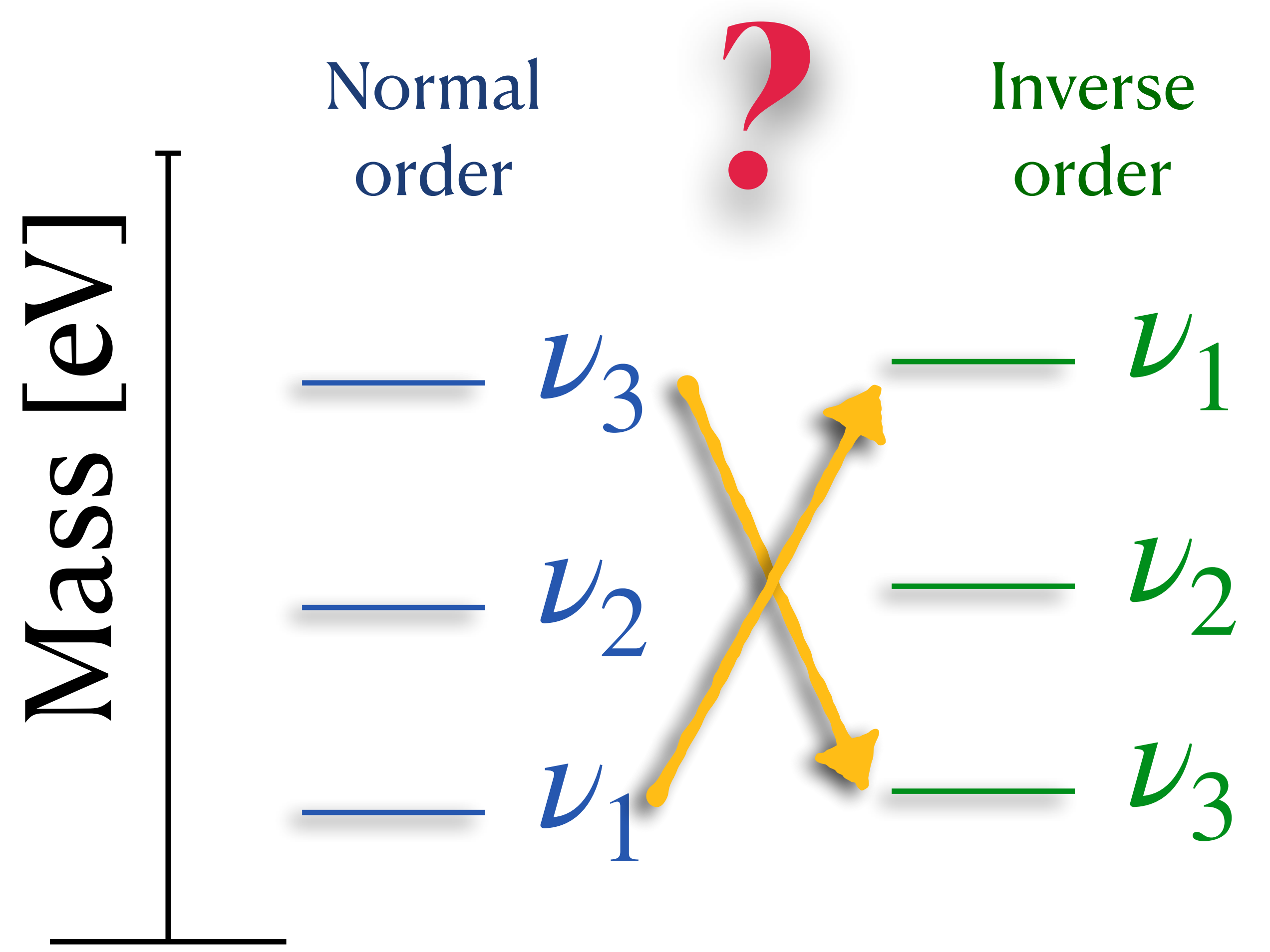
JYA, SciPost '24



$n_s(\delta_{CP}, \sin^2\phi_{23}, \Delta m^2_{31})$

Beyond LHC: Neutrino mass ordering problem

JYA, SciPost '24



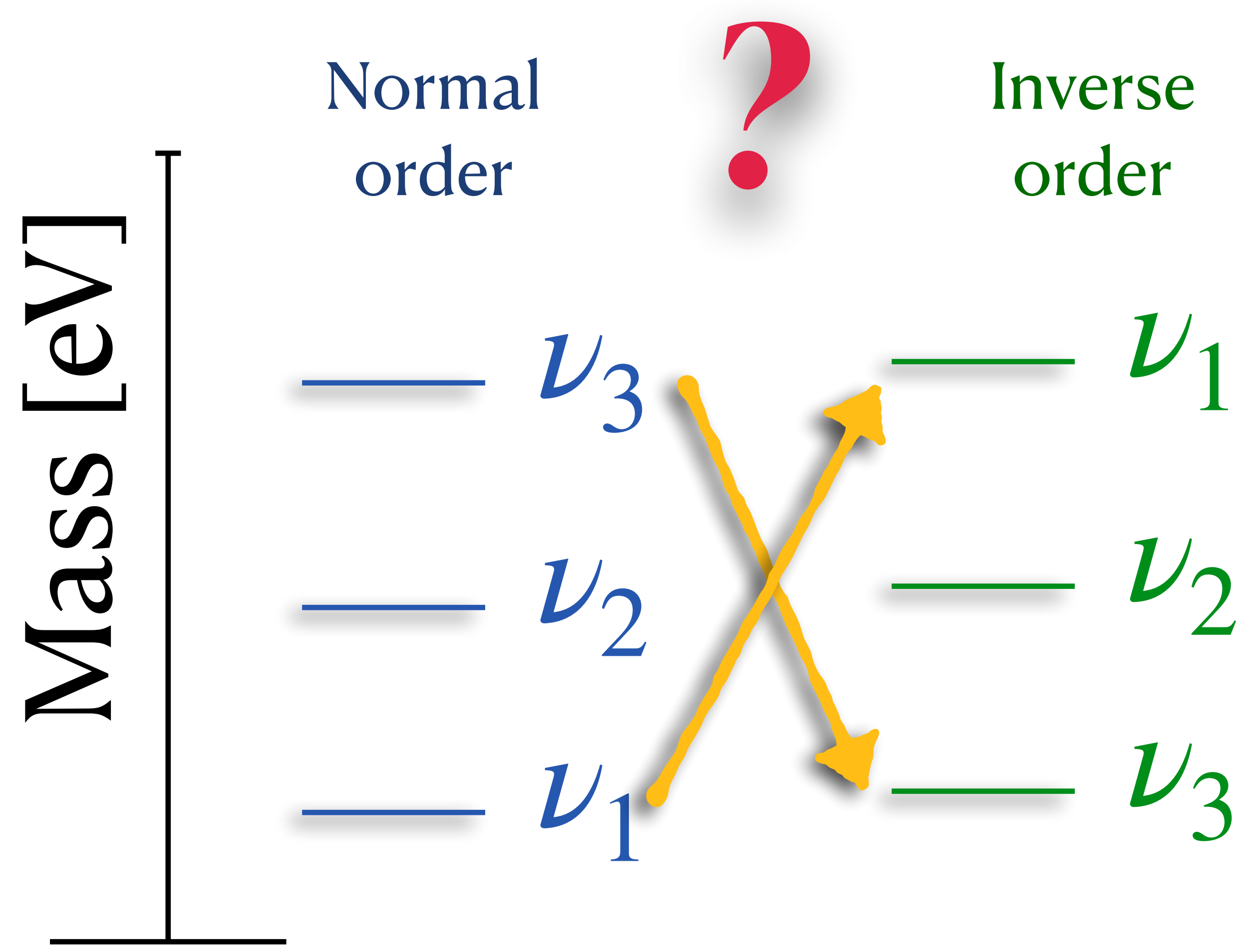
$n_s(\delta_{CP}, \sin^2\phi_{23}, \Delta m^2_{31})$

T2K Experiment
5 channels, 40 bins each

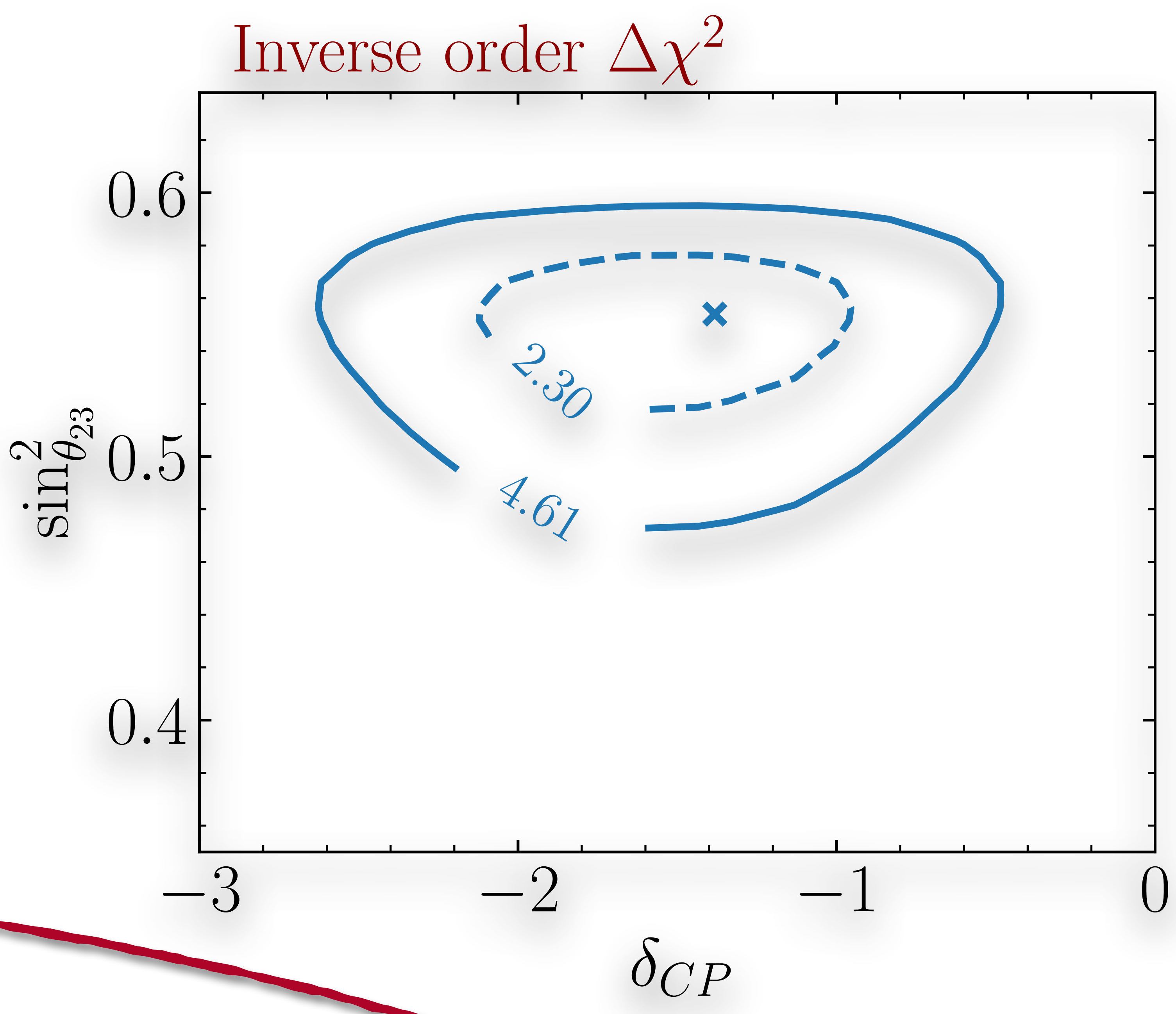
$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{channels}}\prod_{j_i\in\text{bins}} \text{Poiss}\left(\underbrace{n^j}_{\text{Observed}} \middle| \underbrace{(\mu n_s^j + n_b^j)(1+\theta^j\sigma_b^j)}_{\text{Expected}}\right)\right]\cdot\prod_{k\in\text{nuis}}\mathscr{N}\left(\theta^k \middle| 0,1\right)$$

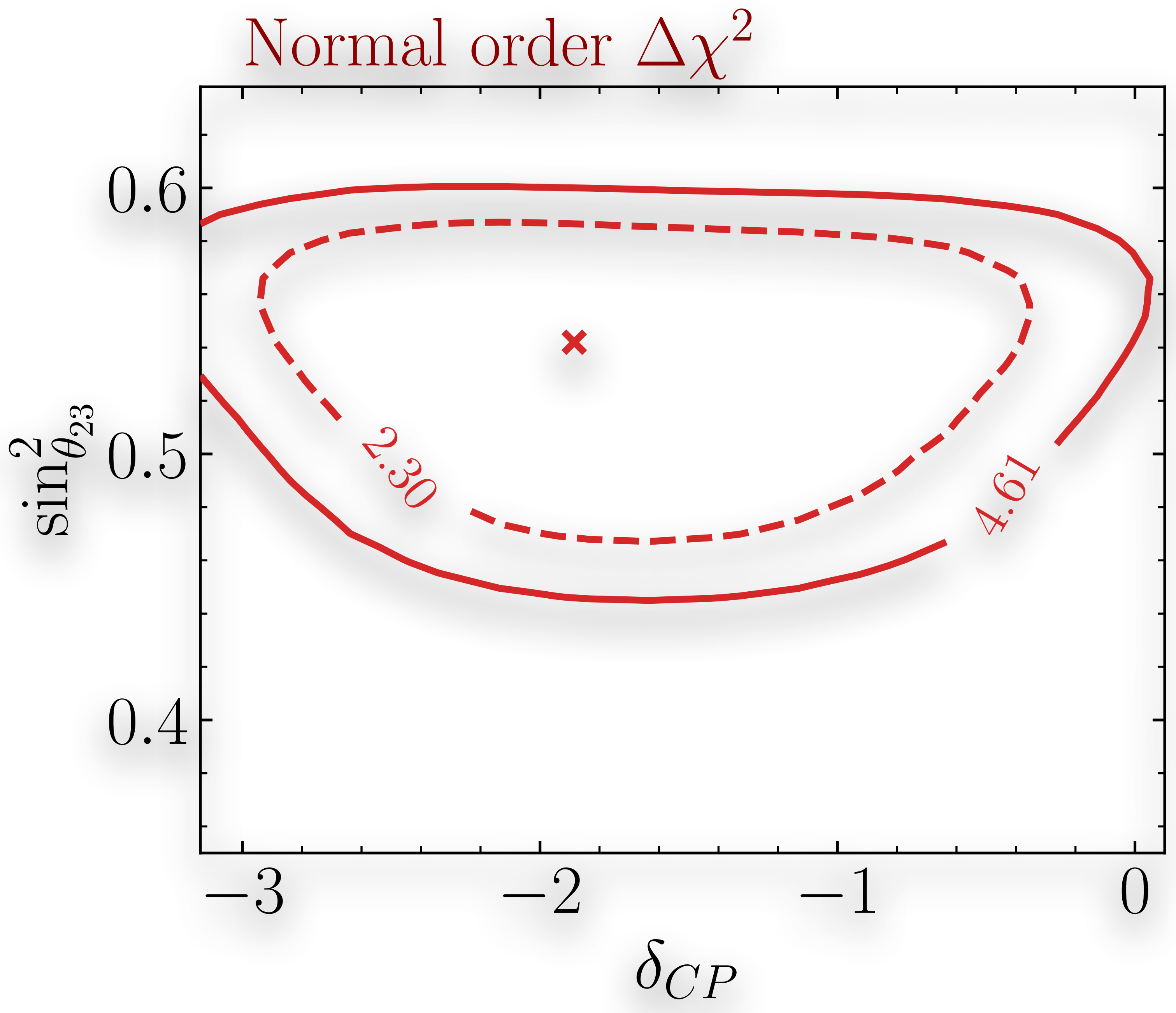
Beyond LHC: Neutrino mass ordering problem

JYA, SciPost '24



$n_s(\delta_{CP}, \sin^2\phi_{23}, \Delta m^2_{31})$





T2K Experiment
5 channels, 40 bins each

$$\mathscr{L}(\mu,\theta) = \left[\prod_{i\in\text{channels}}\prod_{j_i\in\text{bins}} \text{Poiss}\left(\underbrace{n^j}_{\text{Observed}} \middle| (\mu n_s^j + \underbrace{n_b^j}_{\text{Expected}})(1+\theta^j\sigma_b^j)\right)\right]\cdot\prod_{k\in\text{nuis}}\mathcal{N}\left(\theta^k \middle| 0,1\right)$$

Future Plans & Conclusion

Spey in Reinterpretation tools

Contur (→Rivet):

Jon Butterworth & Joe Egan

Spey in Reinterpretation tools

Contur (→Rivet):
Jon Butterworth & Joe Egan

Wolfgang Waltenberger

Spey in Reinterpretation tools

Contur (→Rivet):
Jon Butterworth & Joe Egan

SModelS
Wolfgang Waltenberger

CHECKMATE
Krzysztof Rolbiecki

Spey in Reinterpretation tools

Contur (→Rivet):
Jon Butterworth & Joe Egan

SModelS
Wolfgang Waltenberger

CHECKMATE
Krzysztof Rolbiecki

MA5
Me

Spey in Reinterpretation tools

Contur (→Rivet):
Jon Butterworth & Joe Egan

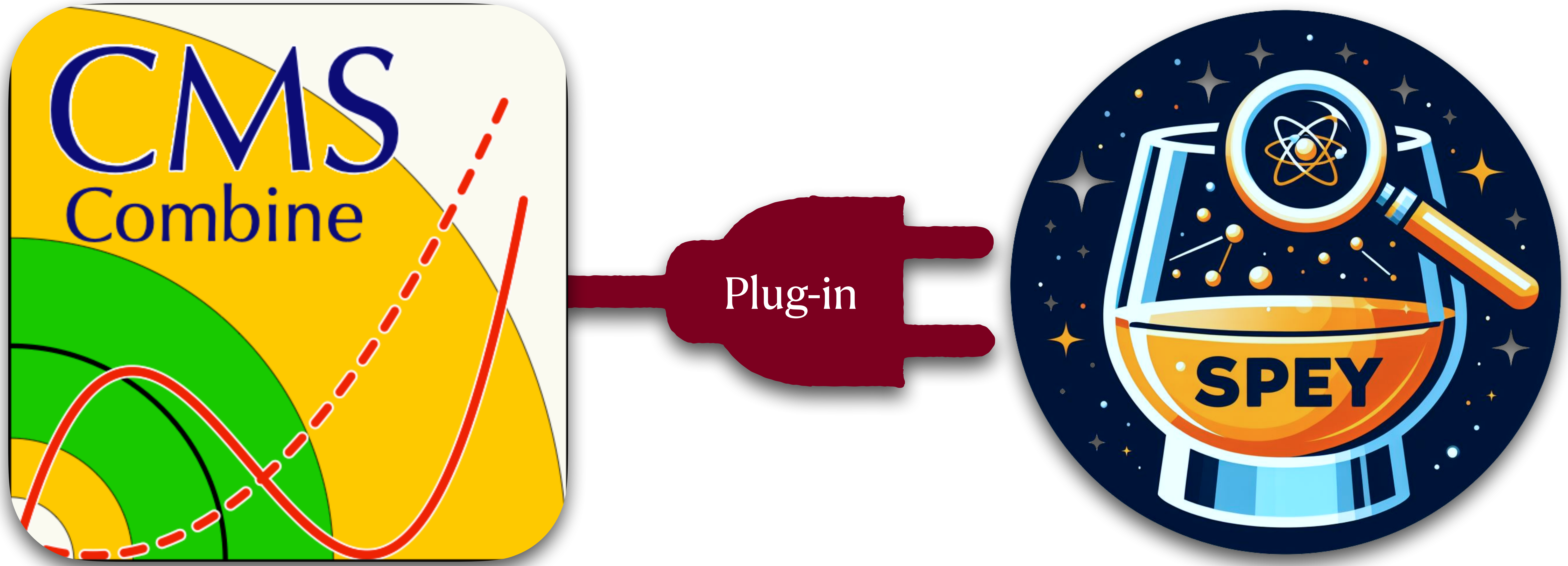
Wolfgang Waltenberger

Krzysztof Rolbiecki

Me

Any analysis, regardless of their origin, can be combined for statistical inference

CMS Full Likelihoods



NEW

See Aliya's talk just after this one!

Citing the proper source

```python
bibtex = spey.get_backend_bibtex("default_pdf.third_moment_expansion")
print(bibtex["inspire"][0])
```

```
@article{Buckley:2018vdr,
    author = "Buckley, Andy and Citron, Matthew and Fichet, Sylvain and Kraml, Sabine and Waltenberger, Wolfgang and Wardle, Nicholas",
    title = "{The Simplified Likelihood Framework}",
    eprint = "1809.05548",
    archivePrefix = "arXiv",
    primaryClass = "hep-ph",
    doi = "10.1007/JHEP04(2019)064",
    journal = "JHEP",
    volume = "04",
    pages = "064",
    year = "2019"
}
```

```python
bibtex = spey.get_backend_bibtex("default_pdf.effective_sigma")
print(bibtex["inspire"][0])
```

```
@inproceedings{Barlow:2004wg,
    author = "Barlow, Roger",
    title = "{Asymmetric statistical errors}",
    booktitle = "{PHYSTAT (2005): Statistical Problems in Particle
    eprint = "physics/0406120",
    archivePrefix = "arXiv",
    reportNumber = "MAN-HEP-04-02",
    pages = "56--59",
    month = "6",
    year = "2004"
}
```

```python
bibtex = spey.get_backend_bibtex("pyhf")
print(bibtex["inspire"][0])
print(bibtex["doi.org"][0])
```

```
@article{Heinrich:2021gyp,
    author = "Heinrich, Lukas and Feickert, Matthew and Stark, Giordon and Cranmer, Kyle",
    title = "{pyhf: pure-Python implementation of HistFactory statistical models}",
    doi = "10.21105/joss.02823",
    journal = "J. Open Source Softw.",
    volume = "6",
    number = "58",
    pages = "2823",
    year = "2021"
}

@misc{https://doi.org/10.5281/zenodo.1169739,
  doi = {10.5281/ZENODO.1169739},
  url = {https://zenodo.org/doi/10.5281/zenodo.1169739},
  author = {Lukas Heinrich,  and Matthew Feickert,  and Giordon Stark, },
  keywords = {physics, statistics, fitting, scipy, numpy, tensorflow, pytorch, jax, auto-differentiation},
  title = {scikit-hep/pyhf: v0.7.6},
  publisher = {Zenodo},
  year = {2024},
  copyright = {Apache License 2.0}
}
```

Possible queries:

- Inspire HEP
- doi.org
- Zenodo

ML integration in Spey

Differentiable programming

`spey.math.value_and_grad(`$\mathscr{L}(\theta)$`)`

`spey.math.hessian(`$\mathscr{L}(\theta)$`)`

$$-\frac{\partial \log \mathscr{L}(\theta)}{\partial \theta}$$

$$-\frac{\partial^2 \log \mathscr{L}(\theta)}{\partial \theta_i \partial \theta_j}$$

- Custom optimiser implementation
- Using likelihoods in AI workflows
- Developing new techniques

ML integration in Spey

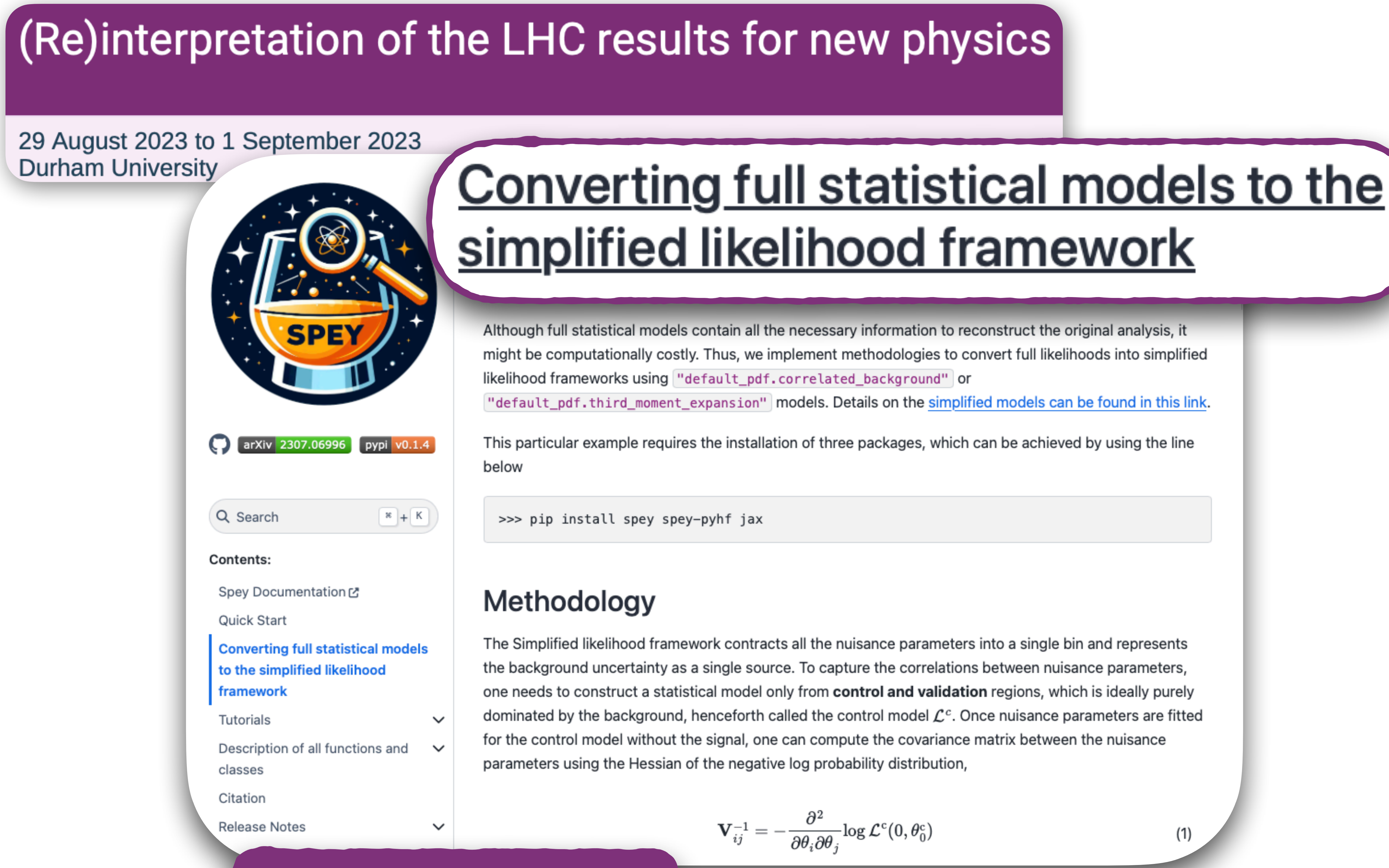
Differentiable programming

```
spey.math.value_and_grad(ℒ(θ))
spey.math.hessian(ℒ(θ))
```

$$-\frac{\partial \log \mathscr{L}(\theta)}{\partial \theta}$$

$$-\frac{\partial^2 \log \mathscr{L}(\theta)}{\partial \theta_i \partial \theta_j}$$

- Custom optimiser implementation
- Using likelihoods in AI workflows
- Developing new techniques



(Re)interpretation of the LHC results for new physics

29 August 2023 to 1 September 2023
Durham University

Converting full statistical models to the simplified likelihood framework

SPEY

arXiv 2307.06996 pypi v0.1.4

Search

Contents:

- Spey Documentation
- Quick Start
- Converting full statistical models to the simplified likelihood framework
- Tutorials
- Description of all functions and classes
- Citation
- Release Notes

Although full statistical models contain all the necessary information to reconstruct the original analysis, it might be computationally costly. Thus, we implement methodologies to convert full likelihoods into simplified likelihood frameworks using `"default_pdf.correlated_background"` or `"default_pdf.third_moment_expansion"` models. Details on the simplified models can be found in this link.

This particular example requires the installation of three packages, which can be achieved by using the line below

```
>>> pip install spey spey-pyhf jax
```

Methodology

The Simplified likelihood framework contracts all the nuisance parameters into a single bin and represents the background uncertainty as a single source. To capture the correlations between nuisance parameters, one needs to construct a statistical model only from **control and validation** regions, which is ideally purely dominated by the background, henceforth called the control model $\mathcal{L}^c$. Once nuisance parameters are fitted for the control model without the signal, one can compute the covariance matrix between the nuisance parameters using the Hessian of the negative log probability distribution,

$$\mathbf{V}^{-1}_{ij} = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathcal{L}^c(0,\theta^c_0) \quad (1)$$

Thanks to Nick Wardle, Sabine Kraml & Wolfgang Waltenberger

ML integration in Spey

Differentiable programming

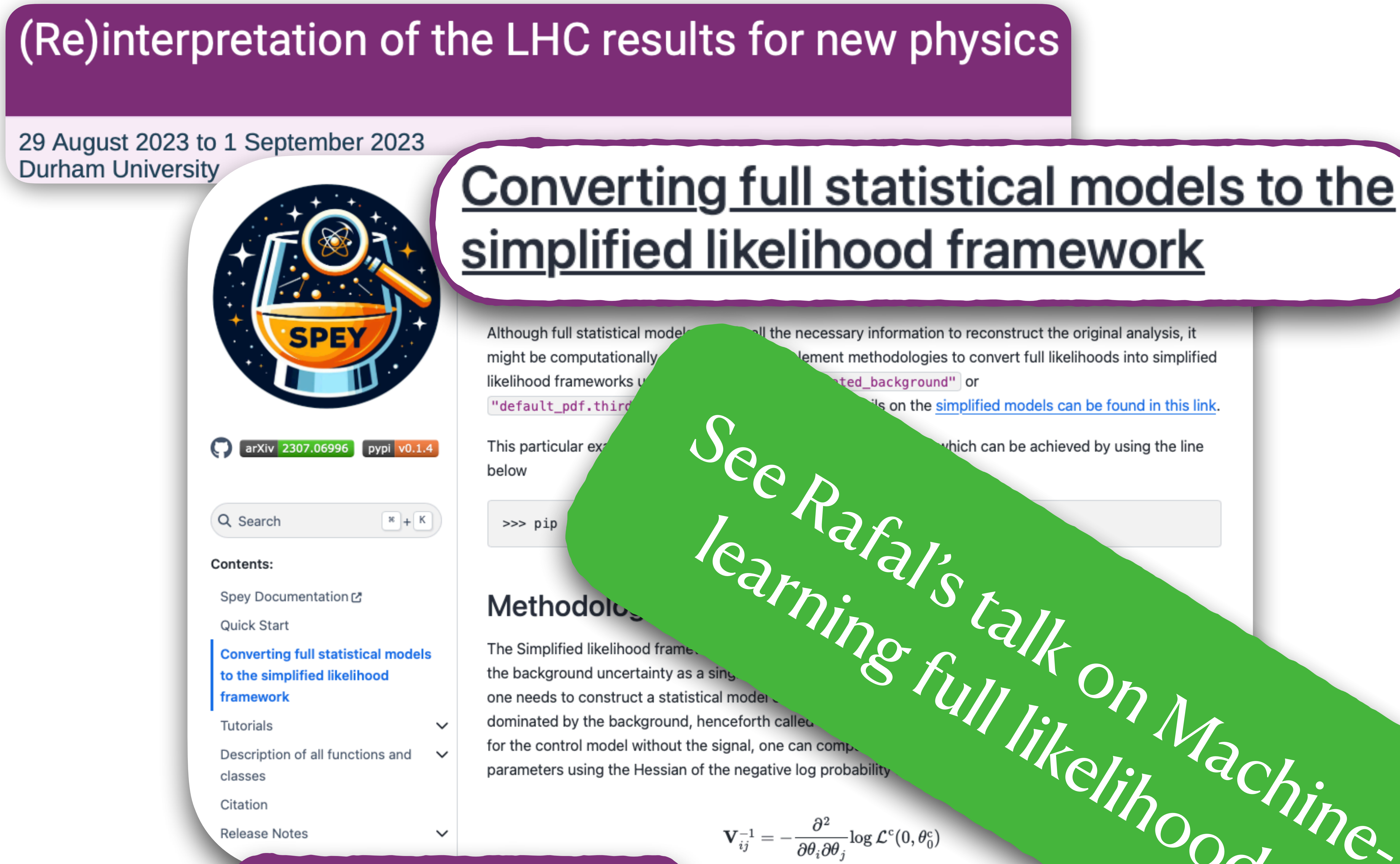
`spey.math.value_and_grad(`$\mathscr{L}(\theta)$`)`

`spey.math.hessian(`$\mathscr{L}(\theta)$`)`

$$-\frac{\partial \log \mathscr{L}(\theta)}{\partial \theta}$$

$$-\frac{\partial^2 \log \mathscr{L}(\theta)}{\partial \theta_i \partial \theta_j}$$

- Custom optimiser implementation
- Using likelihoods in AI workflows
- Developing new techniques



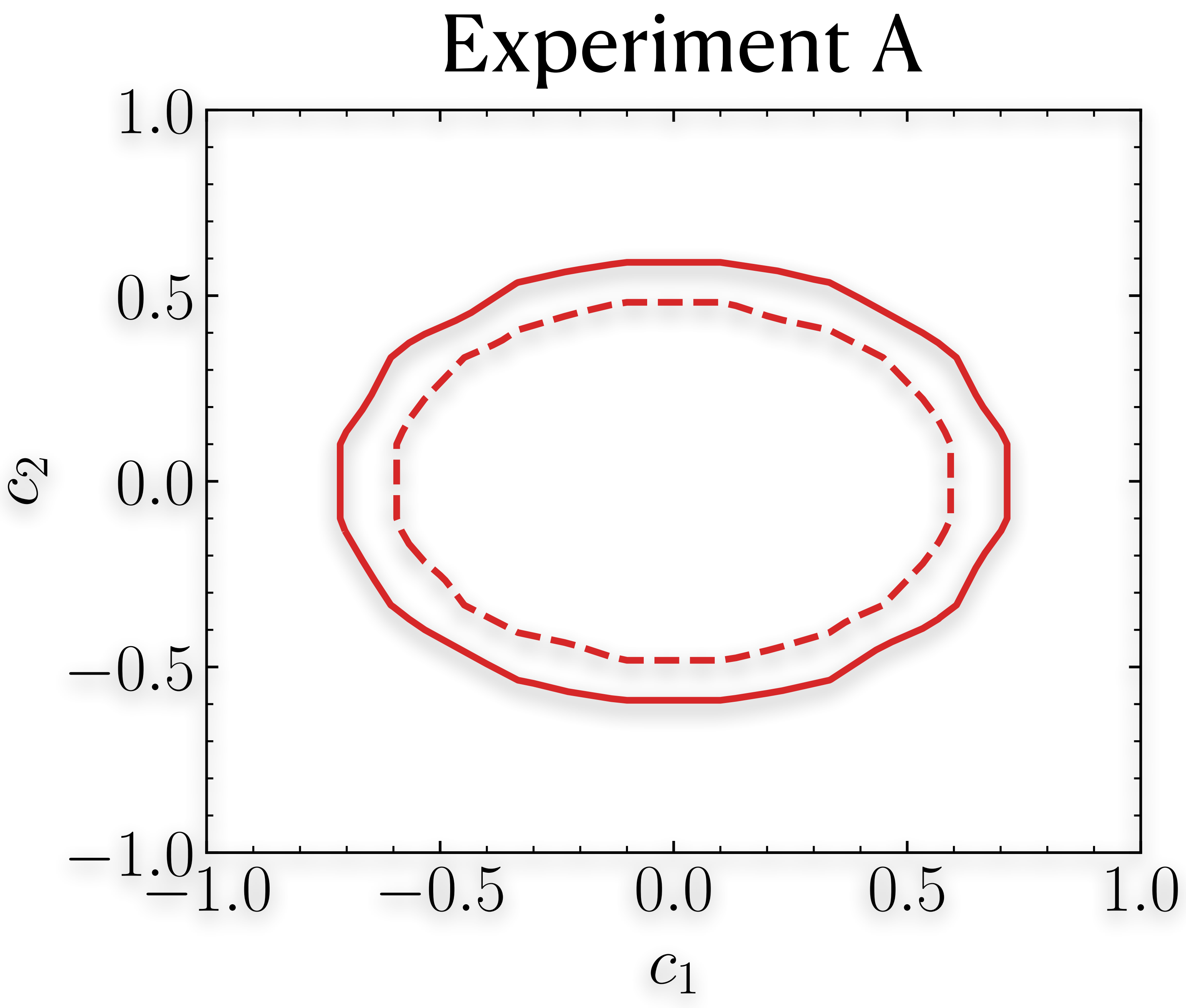
See Rafal's talk on Machine-learning full likelihoods!

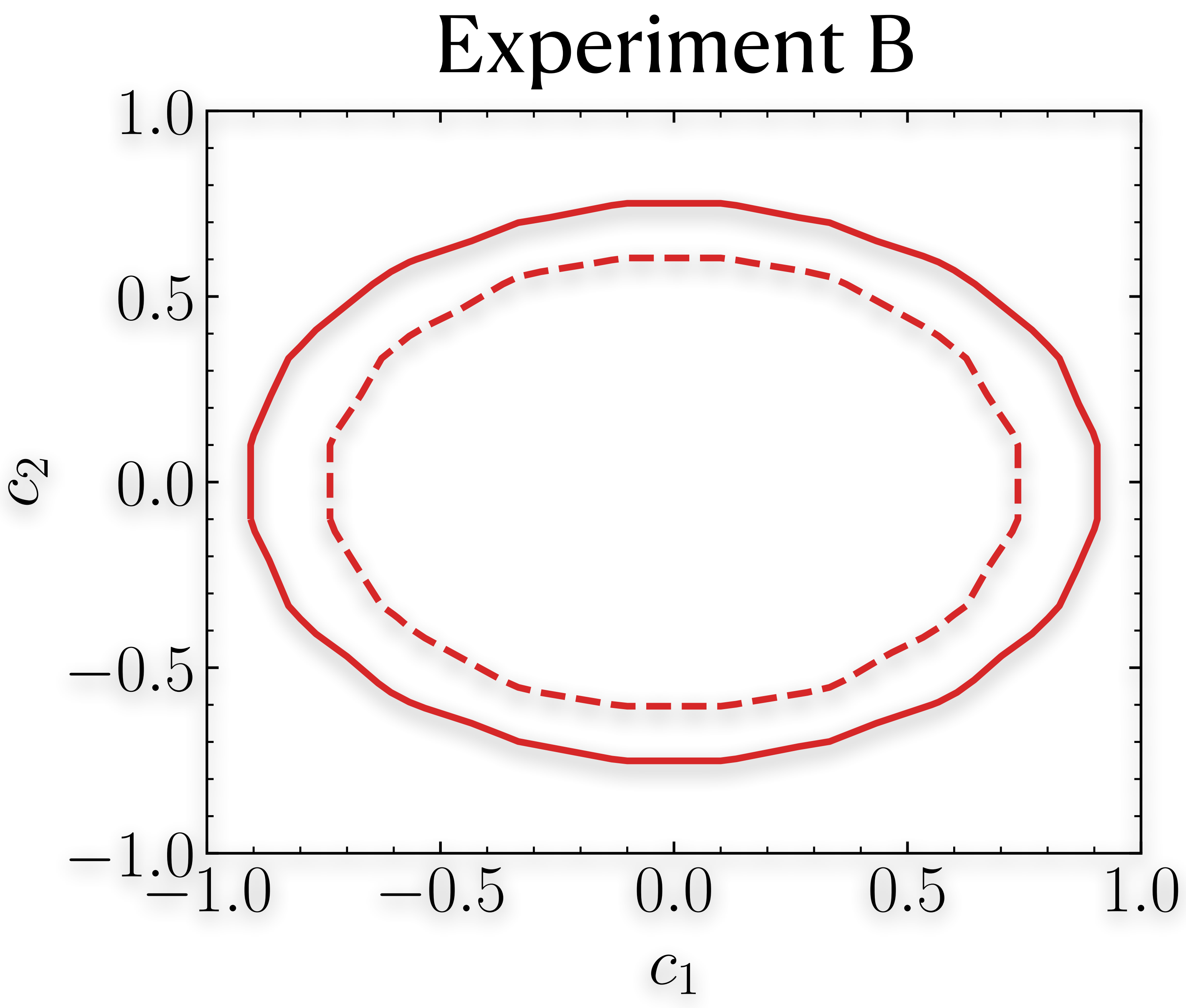
Thanks to Nick Wardle, Sabine Kraml & Wolfgang Waltenberger

Coming Soon: functional signal-based likelihoods

$$n_s \equiv \alpha c_1^2 + \beta c_2^2$$

Click for
the example!
Meow

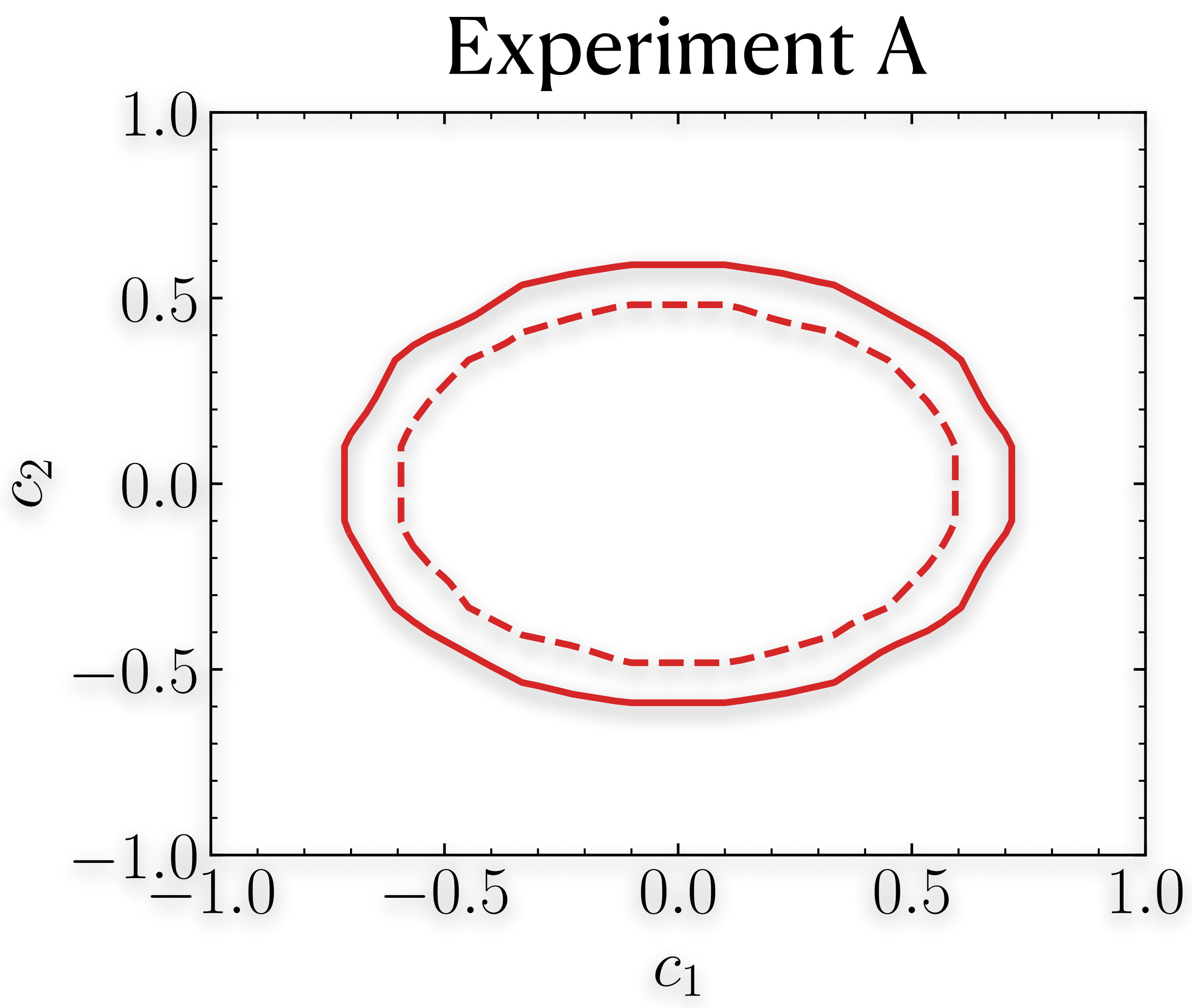




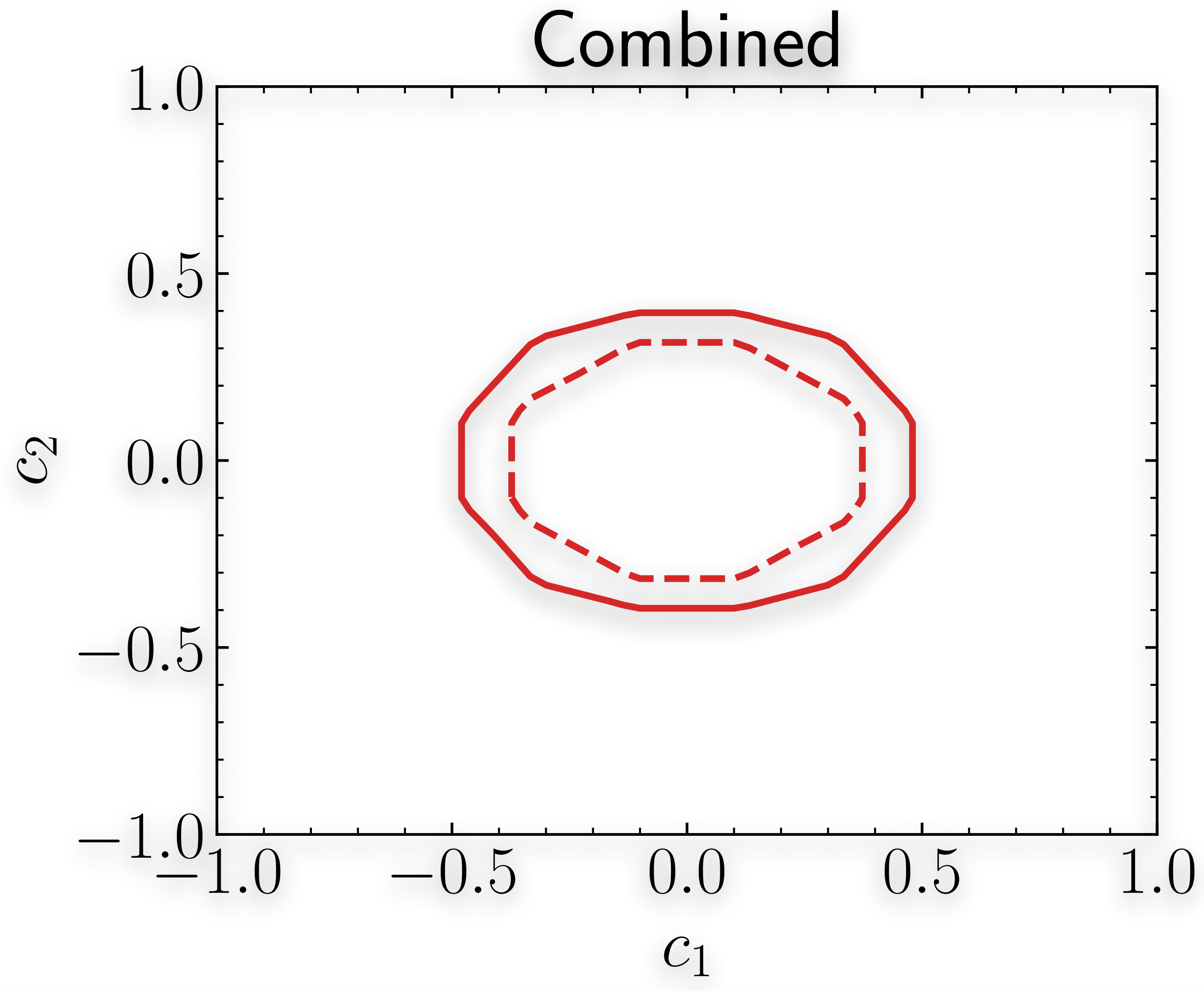
Coming Soon: functional signal-based likelihoods

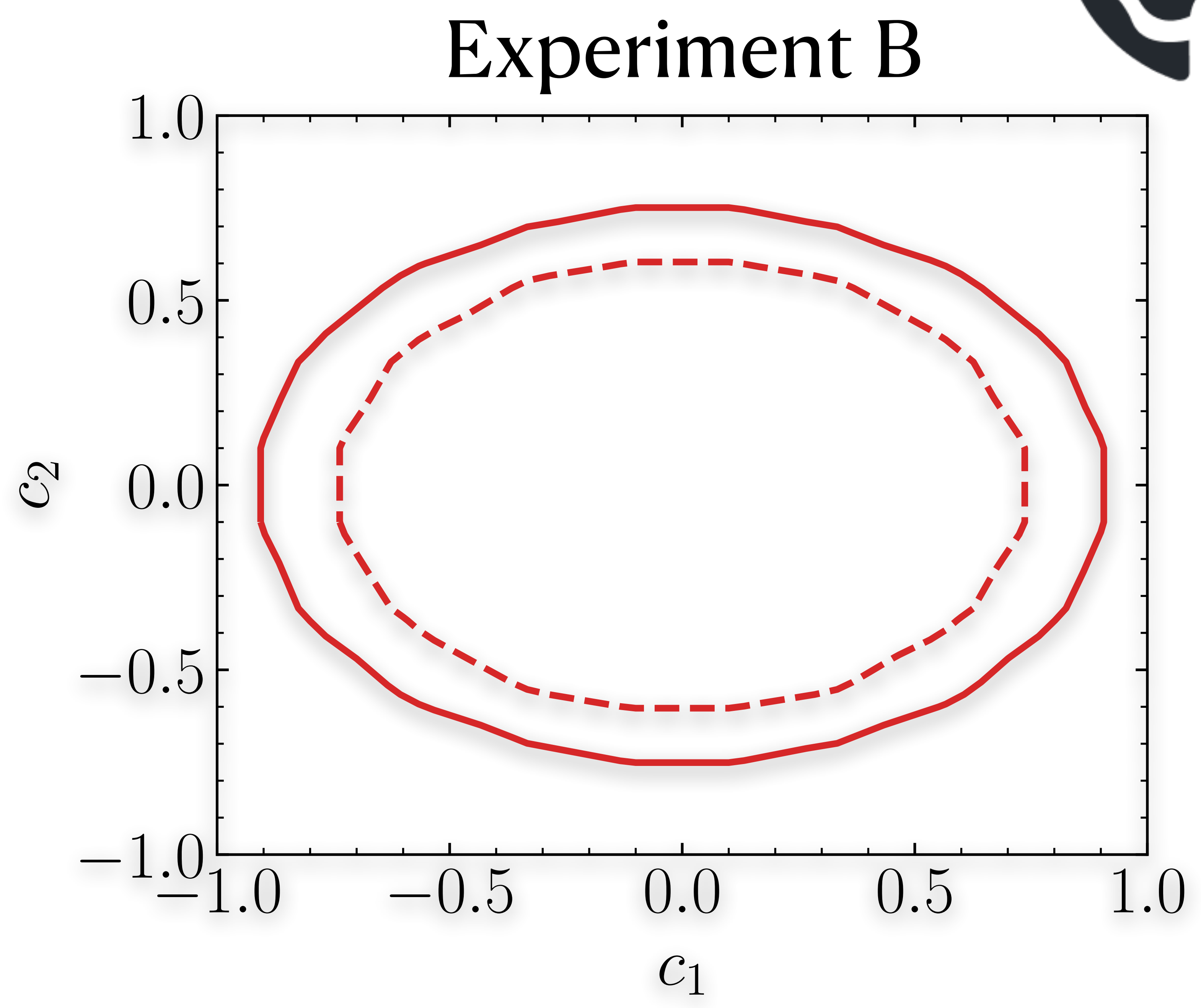
$$n_s \equiv \alpha c_1^2 + \beta c_2^2$$

Click for the example! Meow



$$\oplus$$





Tell me more!

Search

Contents:

Welcome to the documentation of Spey

issues 1 open DOI 10.21468/SciPostPhys.16.1.032 DOI 10.5281/zenodo.10671596

arXiv 2307.06996 pypi v0.1.8

Contents:

spey.readthedocs.io

Tell me more!

Tool description

How to build your own custom likelihood

Tutorials

spey.readthedocs.io


arXiv 2307.06996 pypi v0.1.8

Search

Contents:

Welcome to the documentation of Spey

issues 1 open DOI 10.21468/SciPostPhys.16.1.032 DOI 10.5281/zenodo.10671596

Contents:

Tell m

Tool d … docs.io

How to bu
custom

Tut

arXiv 2307.06996 pypi v0.1.8

Search ⌘ + K

Contents:

Contributing to Spey

We welcome contributions to `spey` via Pull Requests to our GitHub repository. To get started, fork the `main` repository.

For anything beyond minimal fixes that do not require discussion, please first open an issue to discuss your request with the development team.

If your proposed changes are more extensive than a few lines of code, please create a draft pull request. This draft should include the context of the change, a description, and the benefits of the implementation.

- For changes within the Python interface of the program, please run standard tests and write additional tests if necessary.
- Ensure you add examples demonstrating the new implementation.
- Specify any drawbacks of your implementation and offer possible solutions if available.

Pull request procedure

Follow these steps to submit a pull request:

1. Fork the `spey` repository.
2. Open an issue and discuss the implementation with the developers.
3. Commit your changes to a feature branch on your fork and push all your changes.
4. Start a draft pull request and inform the developers about your progress.
5. Pull the `main` branch to ensure there are no conflicts with the current code developments.
6. Modify the appropriate section of `docs/releases/changelog-dev.md`.
7. Once complete, request a review from one of the maintainers.

Come join!
Statistics is fun!

Really!