Machine Learning the Likelihoods

Rafał Masełek

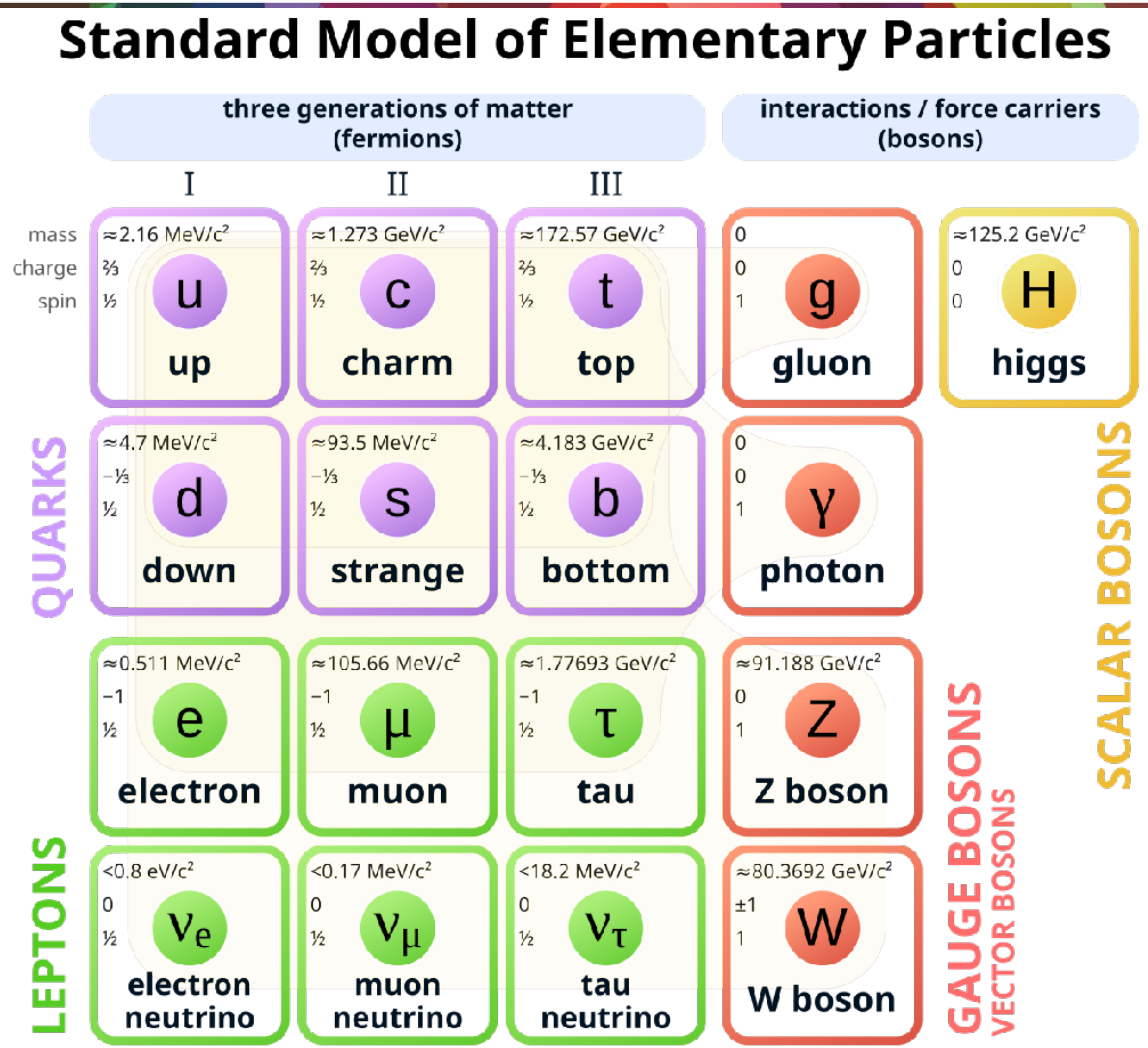
in collaboration with

J. Araz, A. Butter, J. Iturizza, S. Kraml, H. Reyes-Gonzalez, and W. Waltenberger

BRDA 2025: Selected topics in (B)SM, 30-10-2025

This project has received the funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101081355. The operation (SMASH project) is co-funded by the Republic of Slovenia and the European Union from the European Regional Development Fund. The research is supported by IN2P3 theory master project DataMATTER.

Status of the Particle Physics



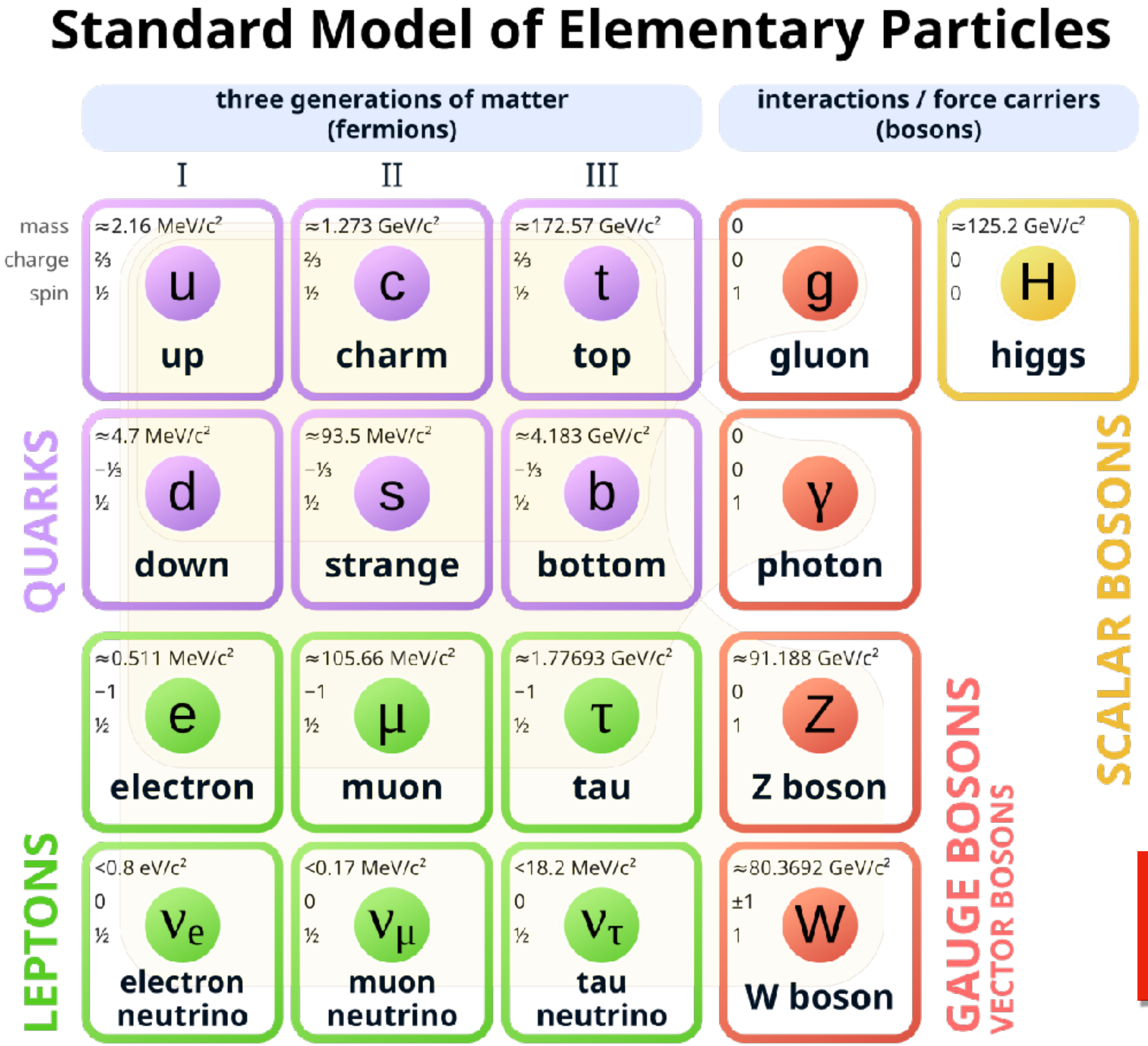
IMG source: https://en.wikipedia.org/wiki/Standard_Model

Status of the Particle Physics

⚠ Dark Matter

⚠ matter-antimatter asymmetry

⚠ neutrino masses



⚠ hierarchy problem

⚠ strong CP problem

⚠ quantum gravity?

IMG source: https://en.wikipedia.org/wiki/Standard_Model

Status of the Particle Physics

⚠ Dark Matter

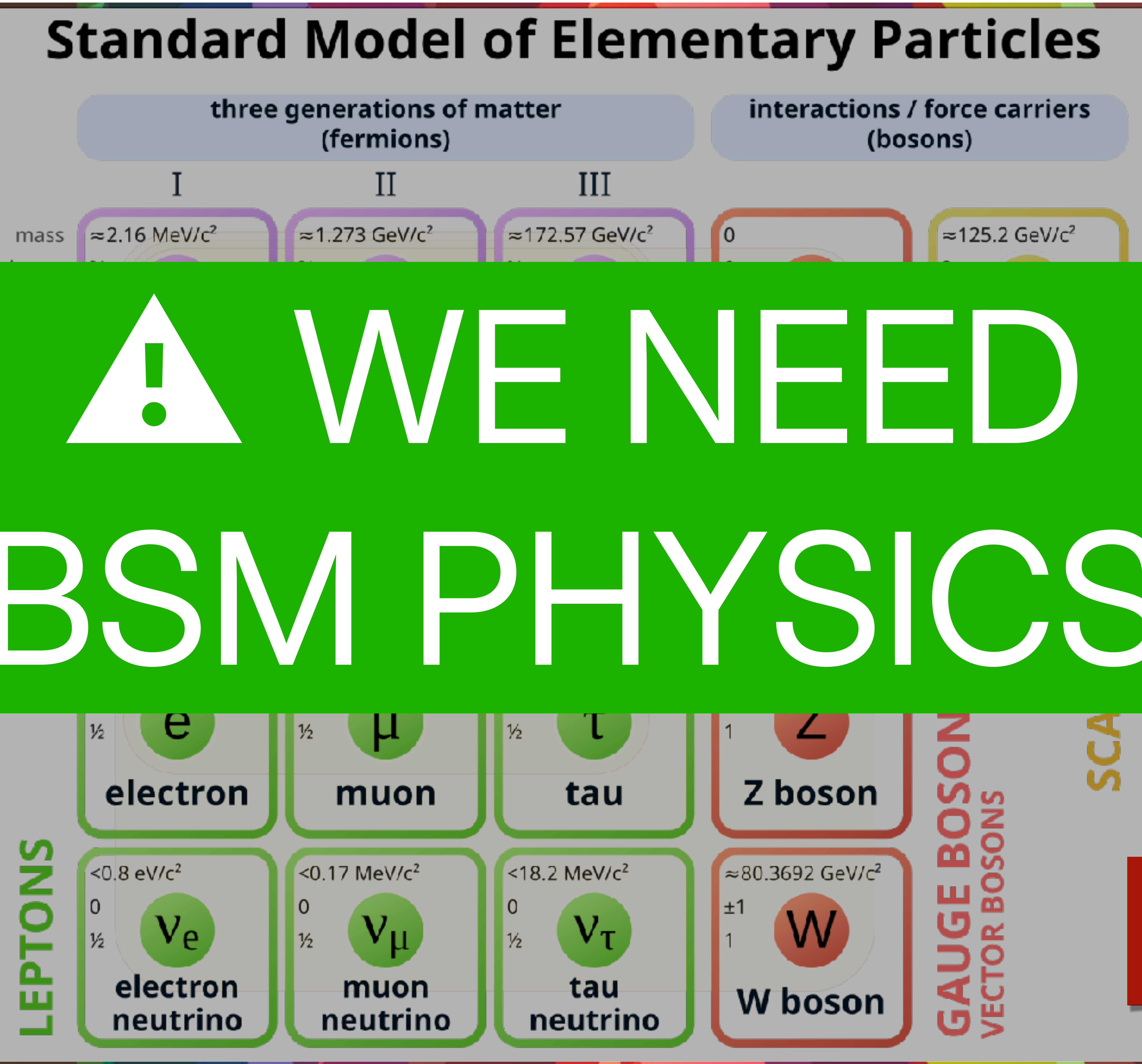
⚠ matter-antimatter asymmetry

⚠ neutrino masses

⚠ hierarchy problem

⚠ strong CP problem

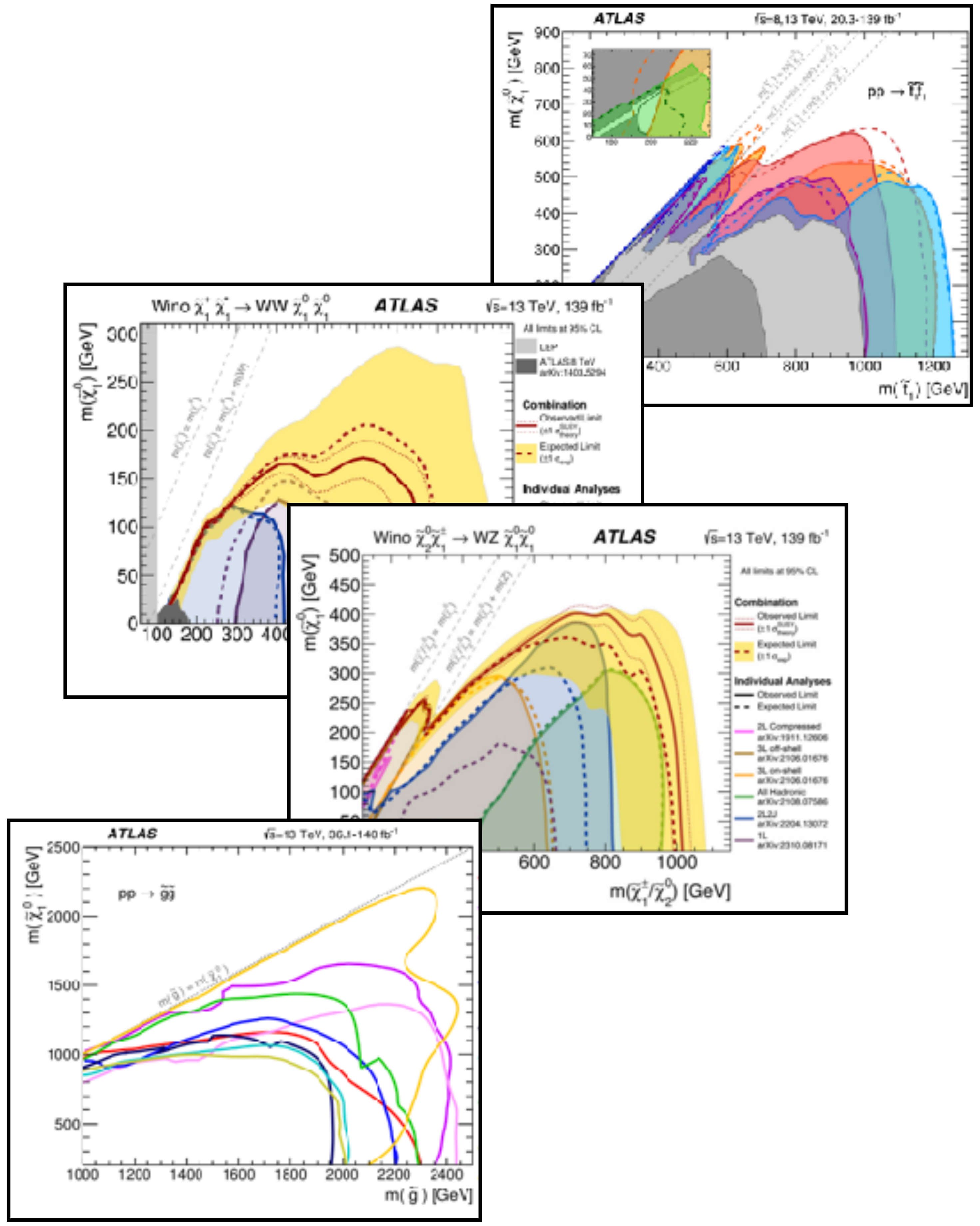
⚠ quantum gravity?

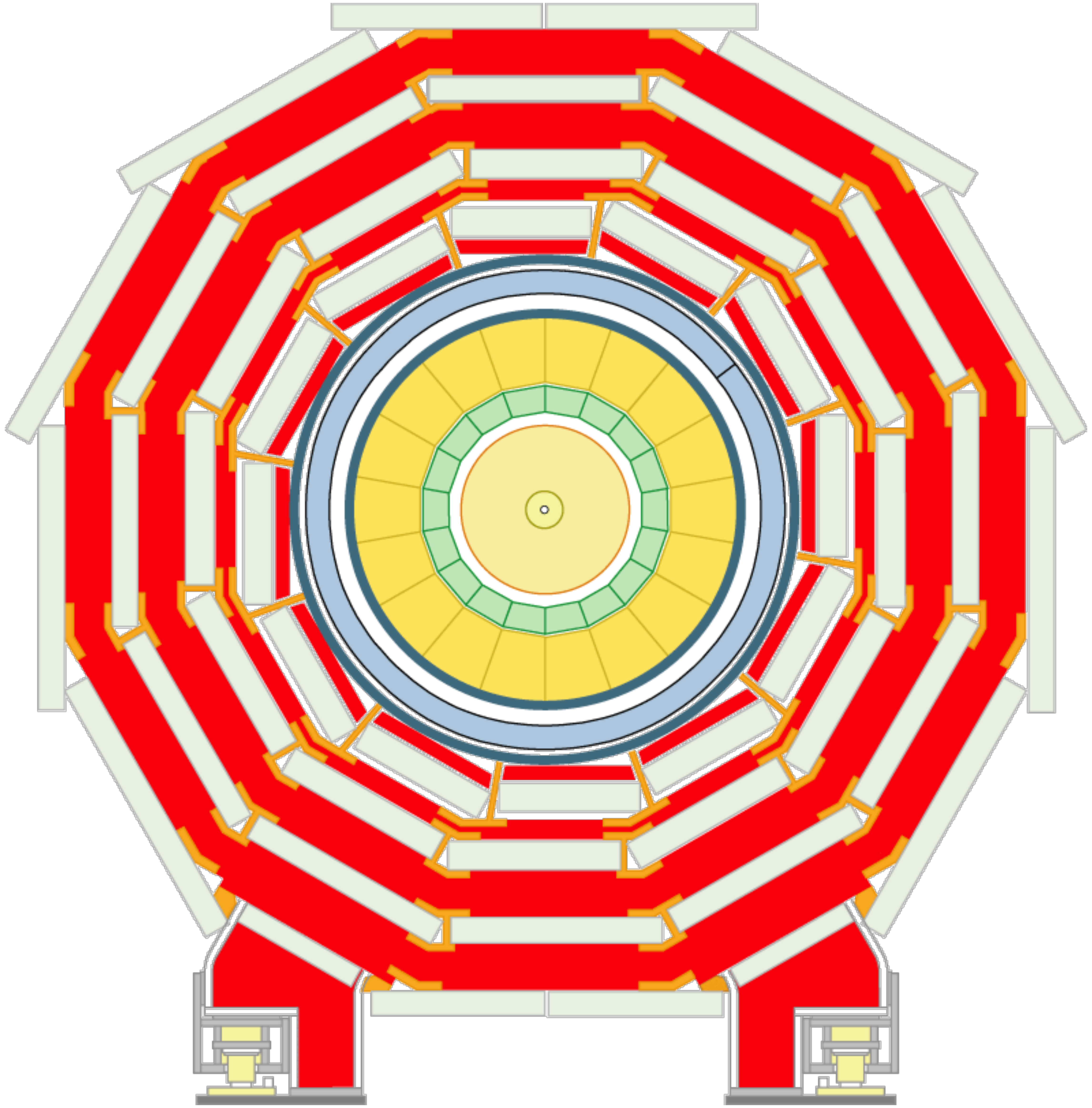


⚠ WE NEED BSM PHYSICS

IMG source: https://en.wikipedia.org/wiki/Standard_Model

Searches for BSM at the LHC

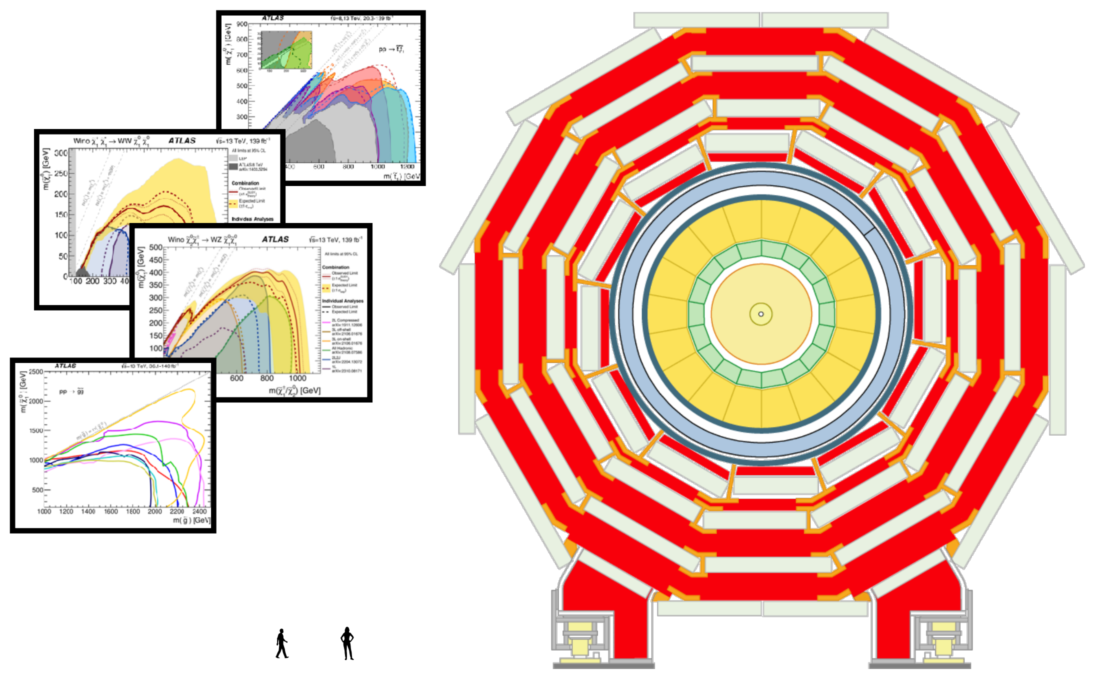




plots from: https://doi.org/10.1016/j.physrep.2024.09.010

Searches for BSM at the LHC

reinterpretation



WIMPs

AXIONS

LEPTOQUARK

DARK SECTOR

LR-SYMMETRY

COMPOSITENESS

SUPERSYMMETRY

NEUTRINO MODELS

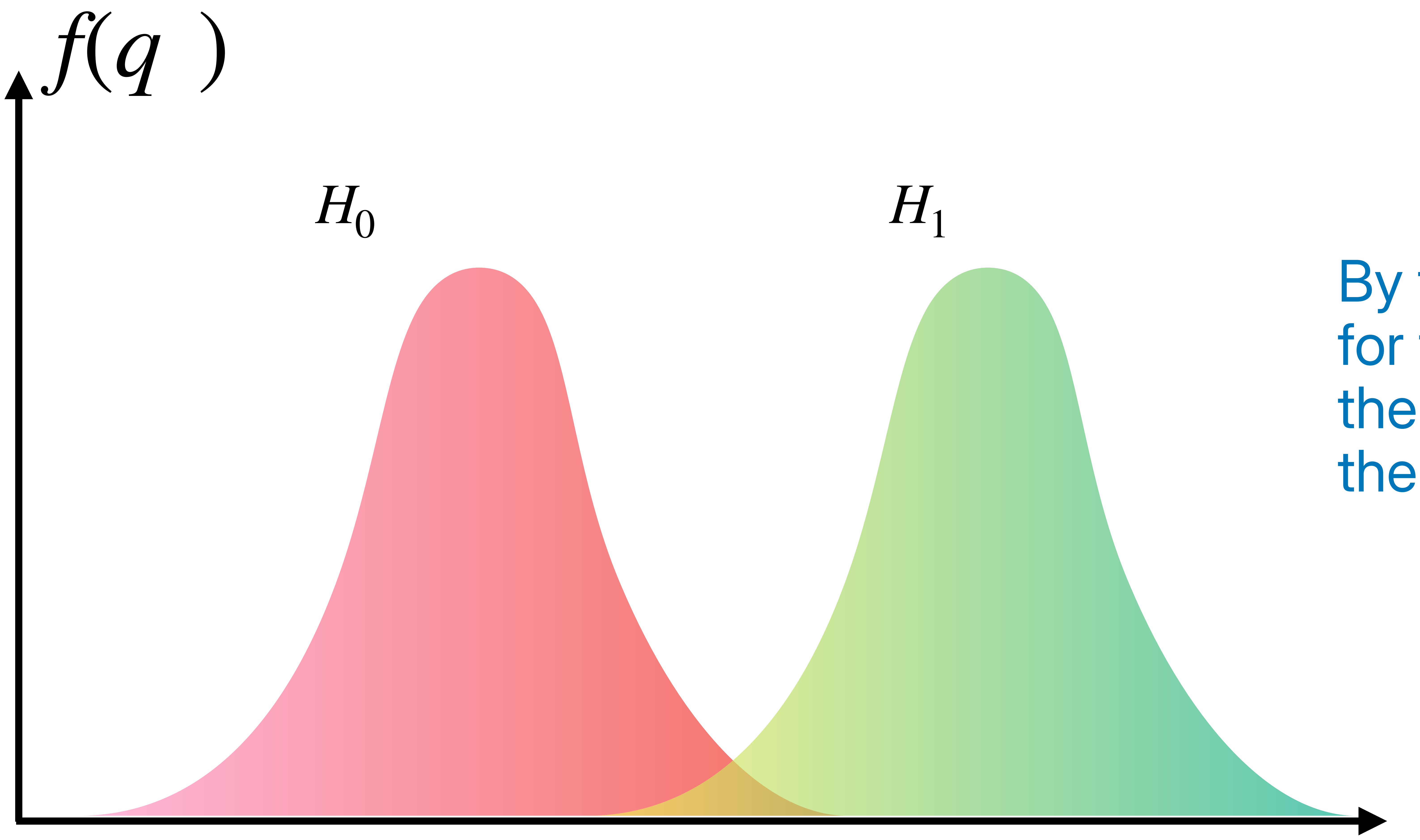
EXTRA DIMENSIONS

TWO HIGGS DOUBLETS

MILICHARGED PARTICLES

Reinterpretation

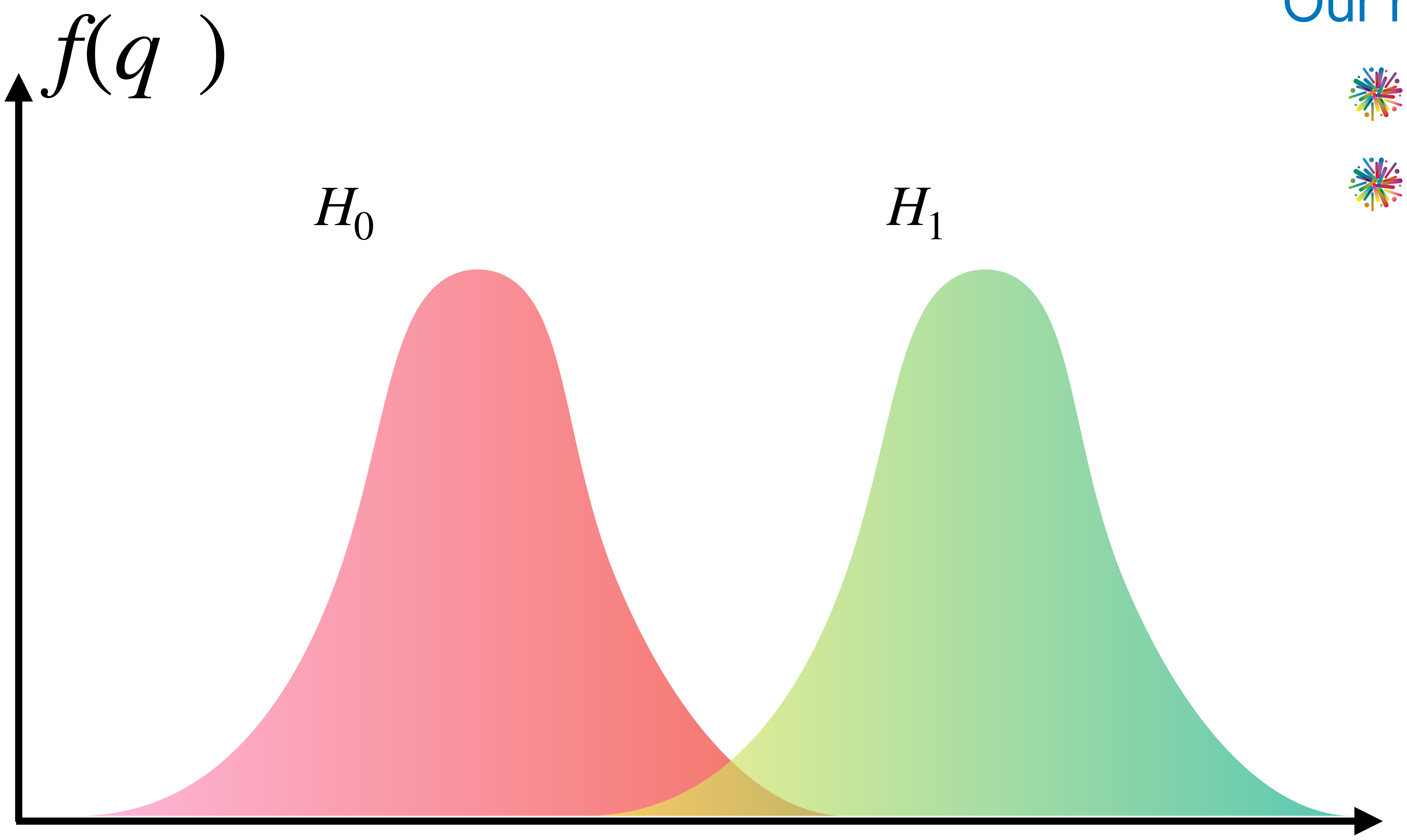
Hypothesis testing



By the Neyman-Pearson lemma, for testing simple hypotheses the most powerful test statistic is the likelihood ratio

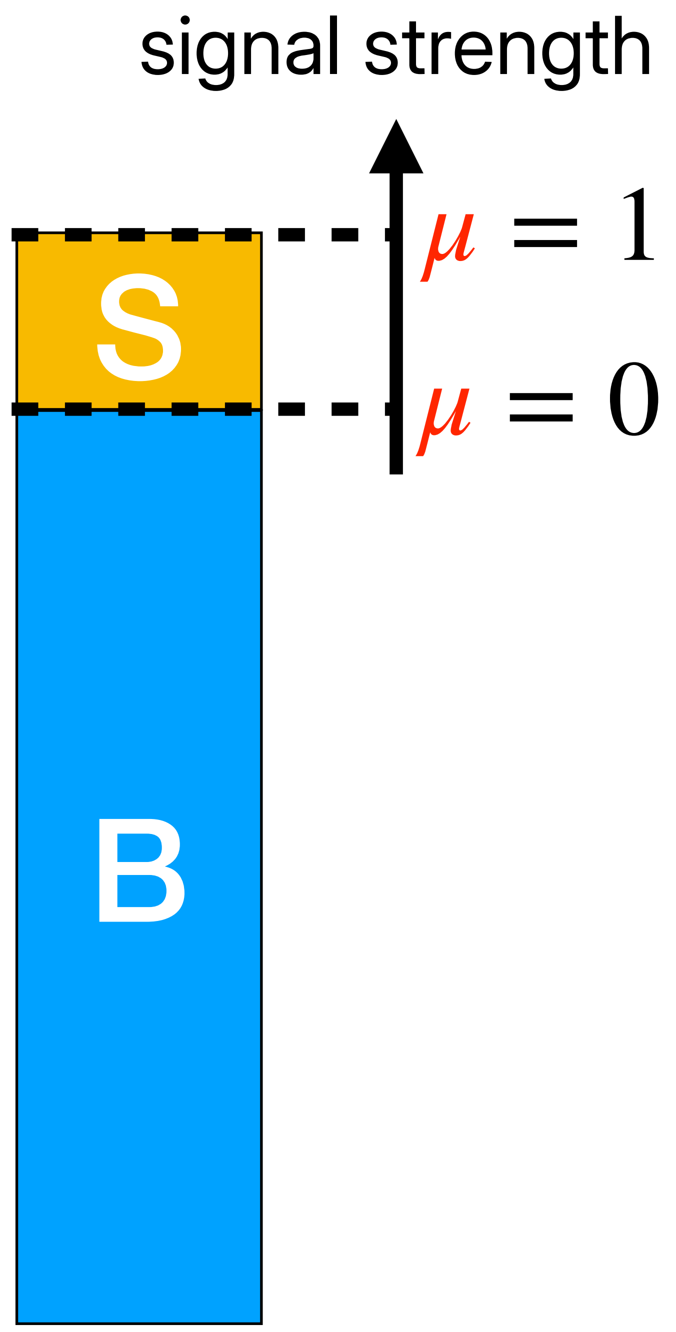
$$q = -2\ln \frac{L\left(\text{data}\,|\,H_0\right)}{L\left(\text{data}\,|\,H_1\right)}$$

Hypothesis testing

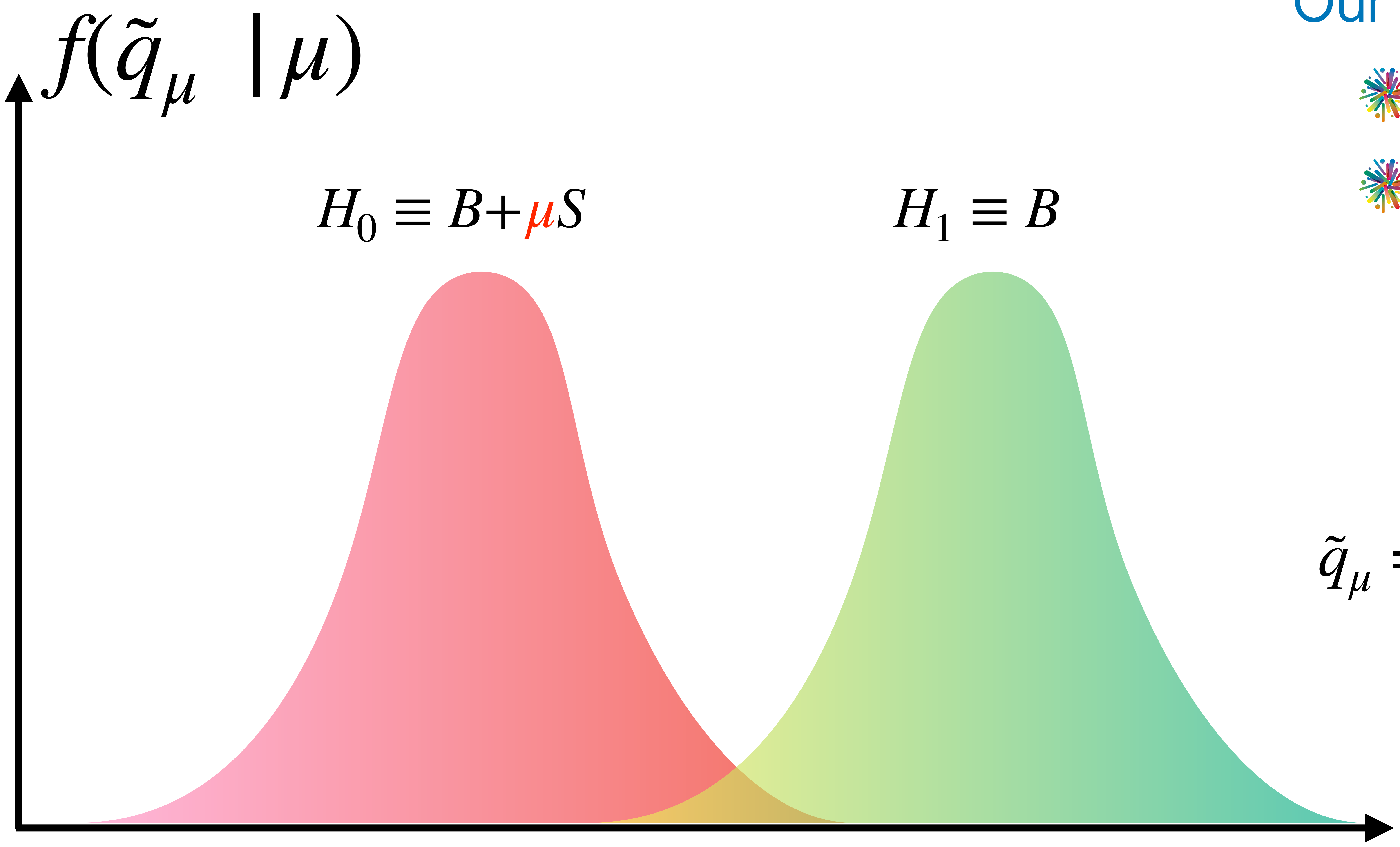


Our hypotheses are composite:

- We have niussance parameters θ
- We have a parameter of interest μ



Hypothesis testing



Our hypotheses are composite:

- We have niussance parameters θ
- We have a parameter of interest μ

$$\tilde{q}_\mu = \begin{cases} -2\ln \dfrac{L\left(\mu, \hat{\hat{\theta}}(\mu)\right)}{L\left(0, \hat{\hat{\theta}}(0)\right)} & \hat{\mu} < 0, \\ -2\ln \dfrac{L\left(\mu, \hat{\hat{\theta}}(\mu)\right)}{L\left(\hat{\mu}, \hat{\theta}\right)} & 0 \le \hat{\mu} \le \mu, \\ 0 & \hat{\mu} > \mu \end{cases}$$

$\hat{\square}$ unconditional MLE

$\hat{\hat{\square}}(\mu)$ conditional MLE

Full likelihood models by ATLAS (Hist Factory)

$$L\left(n, a \,|\, \mu, \theta\right) = \prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \text{Pois}\left(\boxed{n_{cb}} \,|\, \nu_{cb}(\boxed{\mu}, \boxed{\theta})\right) \prod_{\theta} p_\theta\left(\boxed{a_\theta} \,|\, \boxed{\theta}\right)$$

Full likelihood models by ATLAS (Hist Factory)

channel data

auxiliary data

$$L\left(n, a \,|\, \mu, \theta\right) = \underbrace{\prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \mathrm{Pois}\left(\boxed{n_{cb}} \,|\, \nu_{cb}(\boxed{\mu}, \boxed{\theta})\right)}_{\text{simultaneous measurement of multiple channels}} \underbrace{\prod_{\theta} p_{\theta}\left(\boxed{a_{\theta}} \,|\, \boxed{\theta}\right)}_{\text{constraint terms for "auxiliary measurements"}}$$

free parameters

constrained parameters

simultaneous measurement
of multiple channels

constraint terms for
"auxiliary measurements"

Full likelihood models by ATLAS (Hist Factory)

auxiliary data

channel data

$$L\left(n, a \,|\, \mu, \theta\right) = \underbrace{\prod_{c}^{\text{channels}} \prod_{b}^{\text{bins}_c} \mathrm{Pois}\left(\boxed{n_{cb}} \,|\, \nu_{cb}(\boxed{\mu}, \boxed{\theta})\right)}_{\text{simultaneous measurement of multiple channels}} \underbrace{\prod_{\theta} p_\theta\left(\boxed{a_\theta} \,|\, \boxed{\theta}\right)}_{\text{constraint terms for "auxiliary measurements"}}$$

free parameters

constrained parameters

simultaneous measurement of multiple channels

constraint terms for "auxiliary measurements"

$$\nu_{cb}(\mu,\theta) = \sum_{s}^{\text{samples}} \nu_{scb}(\mu,\theta) = \sum_{s}^{\text{samples}} \underbrace{\left(\prod_{\kappa} \kappa_{scb}(\mu,\theta)\right)}_{\textbf{multiplicative modifiers}} \left(\underbrace{\nu^0_{scb}(\mu,\theta)}_{\textbf{const. nominal rate}} + \underbrace{\sum_{\Delta} \Delta_{scb}(\mu,\theta)}_{\textbf{additive modifiers}}\right)$$

multiplicative modifiers **const. nominal rate** **additive modifiers**

Why bother?

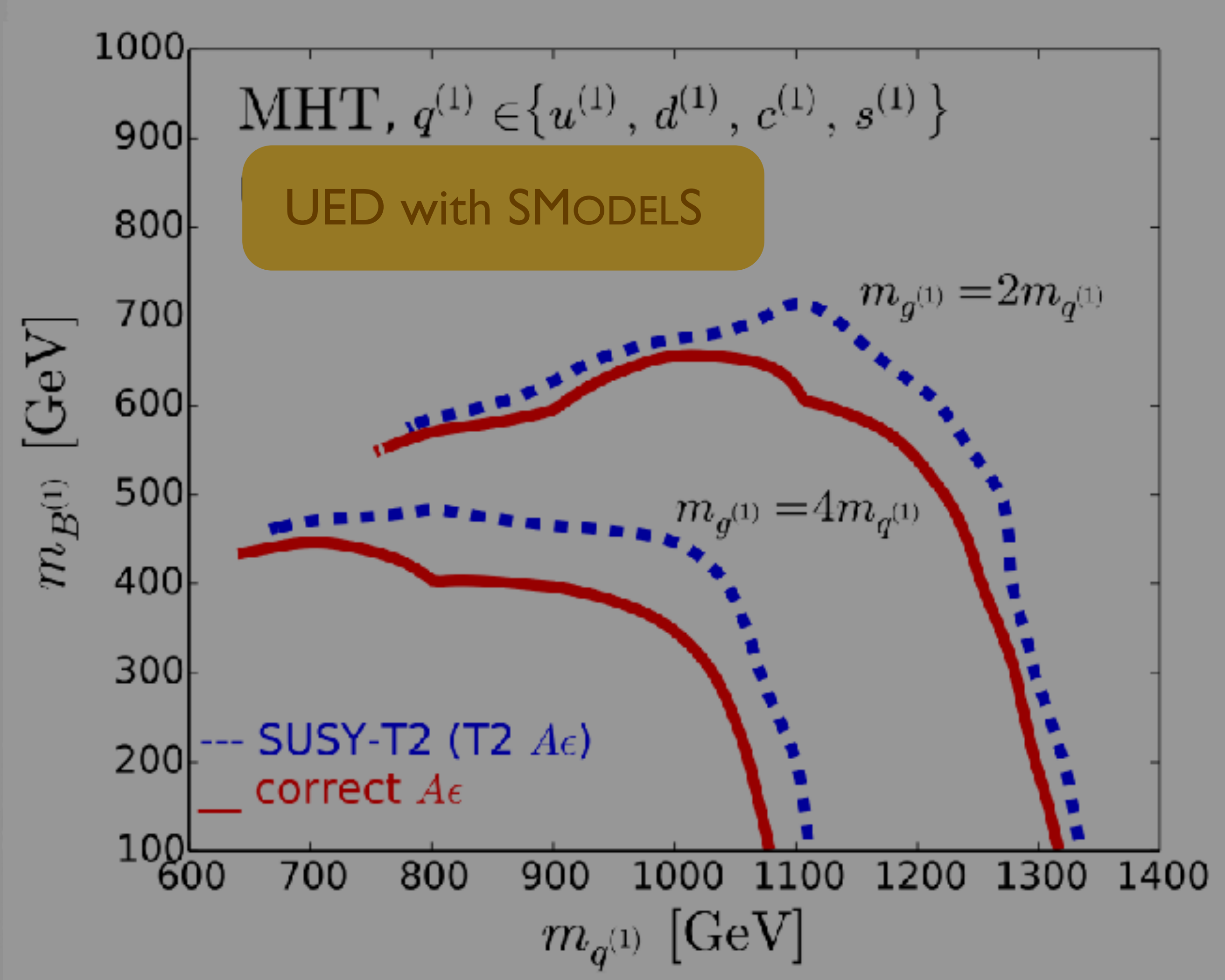
SMS-based *public* tools

SMS-based

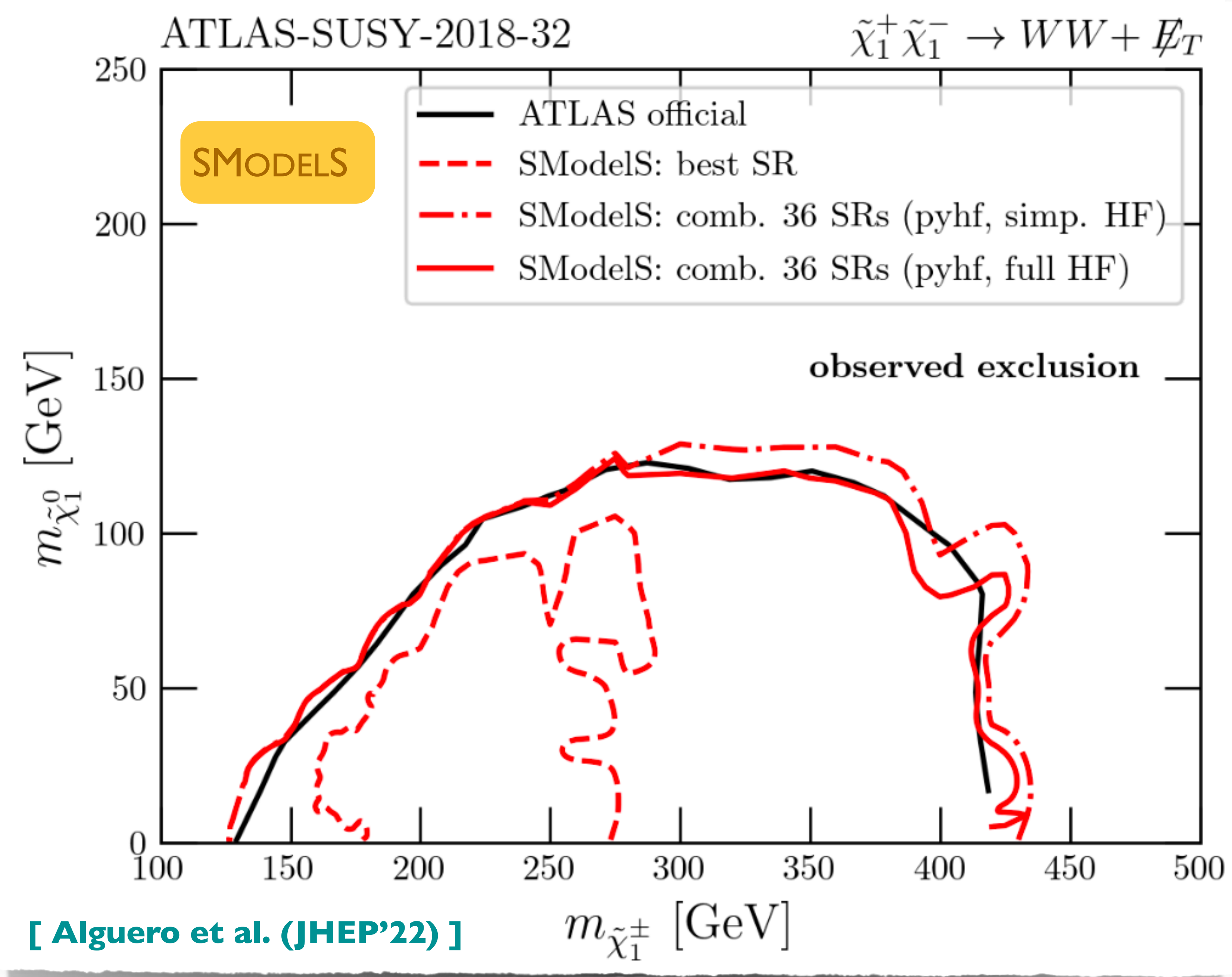
- SMODELS [$O(100)$ analyses, from GITHUB]
  - ⋆ Validation example (ATLAS-SUSY-2018-32)
    - ➙ $\tilde{W}/\tilde{B}$ scenario with 2 leptons and $\not{E}_T$
    - ➙ PYHF model crucial

[Kraml et al. (EPJC'14); Alguero et al. (JHEP'22)]
[Altakach et al. (SciPost'24)]

- Dark photons: DARKCAST [from GITLAB] [Ilten et al. (JHEP'18)]

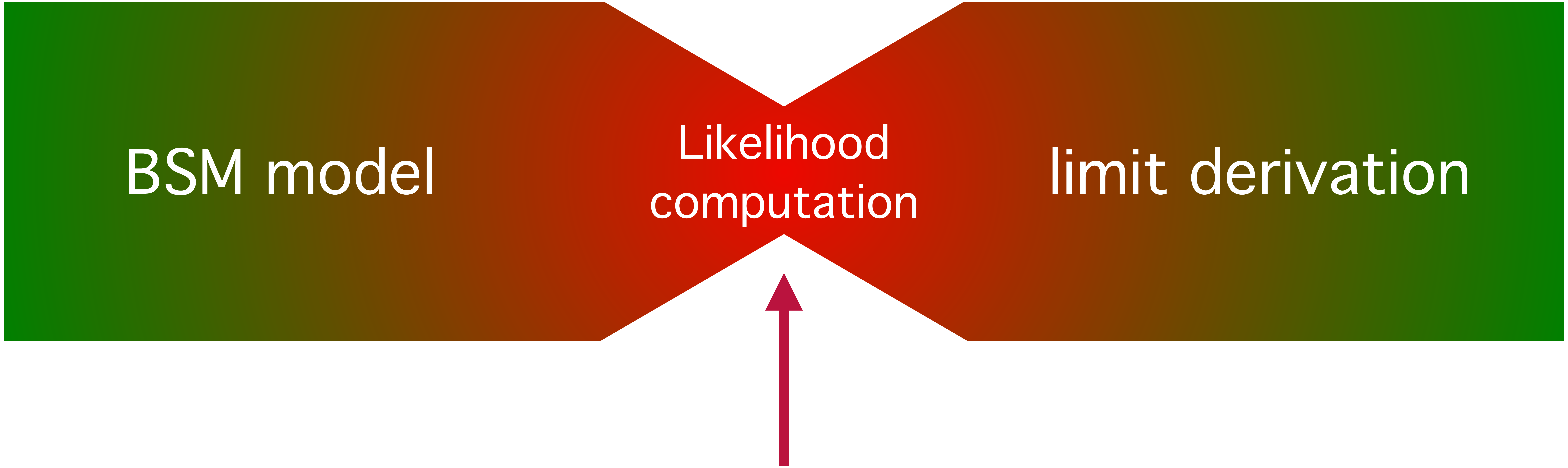


[Edelhäuser, Krämer, Sonneveld (JHEP'15)]



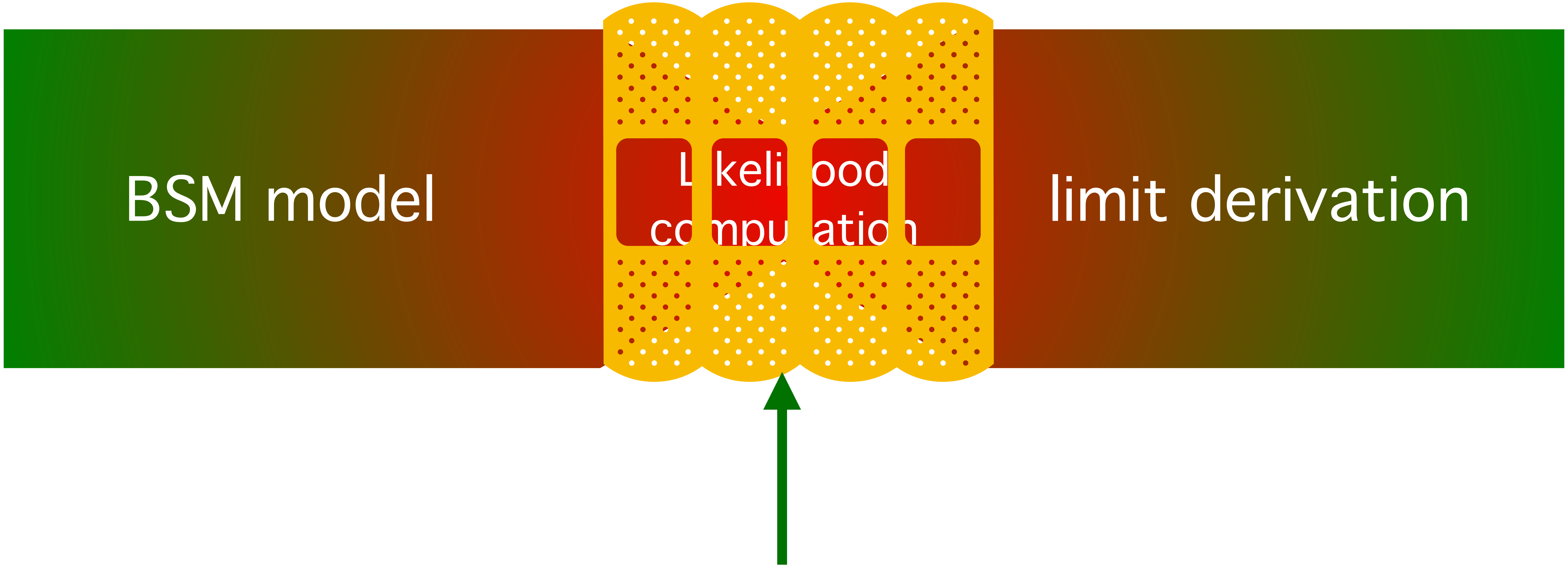
[Alguero et al. (JHEP'22)]

Computational bottleneck



Full statistical model calculations enter here

Computational bottleneck



Machine learning enters here

The goal

Use Machine Learning to create surrogates to Full Statistical Models that would be **accurate** but **much faster** to use.

The approach

If we have a regressor for profiled likelihoods, we can calculate limits.

$$\Delta_{\text{data}}^{(\mu_1,\ldots\mu_N)} \equiv \ln \frac{L\left(\text{data}\,\middle|\,\mu_1,\ldots,\mu_N;\ \hat{\hat{\theta}}(\mu_1,\ldots,\mu_N)\right)}{L\left(\text{data}\,\middle|\,0,\ldots,0;\ \hat{\hat{\theta}}(0,\ldots,0)\right)}$$

The approach

varying signal
in each bin independently

profiling

$$\Delta_{\mathrm{data}}^{(\mu_1,\ldots\mu_N)} \equiv \ln \frac{L\left(\mathrm{data}\,\middle|\,\mu_1,\ldots,\mu_N;\ \hat{\hat{\theta}}(\mu_1,\ldots,\mu_N)\right)}{L\left(\mathrm{data}\,\middle|\,0,\ldots,0;\ \hat{\hat{\theta}}(0,\ldots,0)\right)}$$

Four data sets are used:
1) expected data
2) expected Asimov data
3) observed data
4) observed Asimov data

The approach

regression

yields → likelihoods → limits

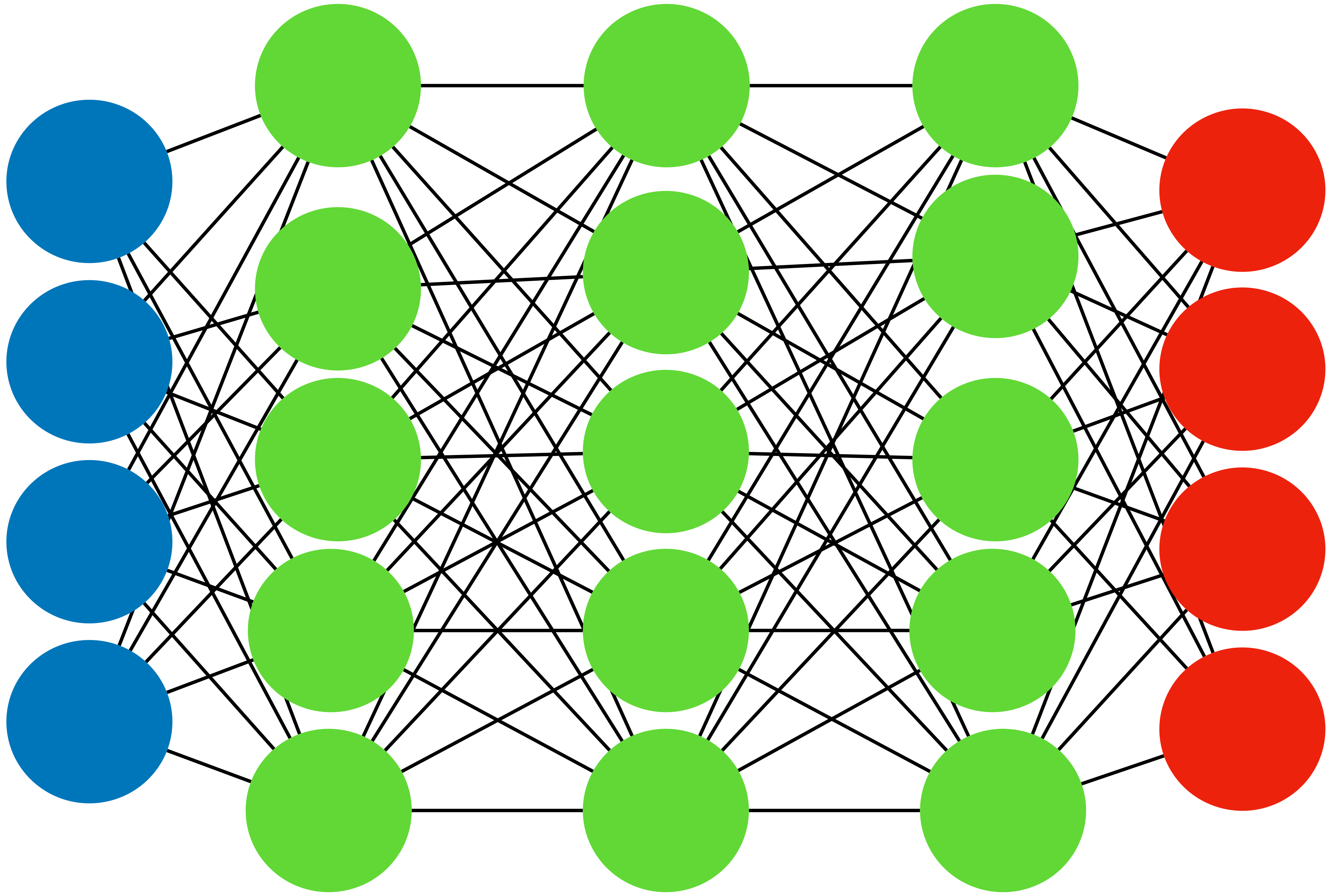
n_1

n_2

n_3

$\vdots$

n_N



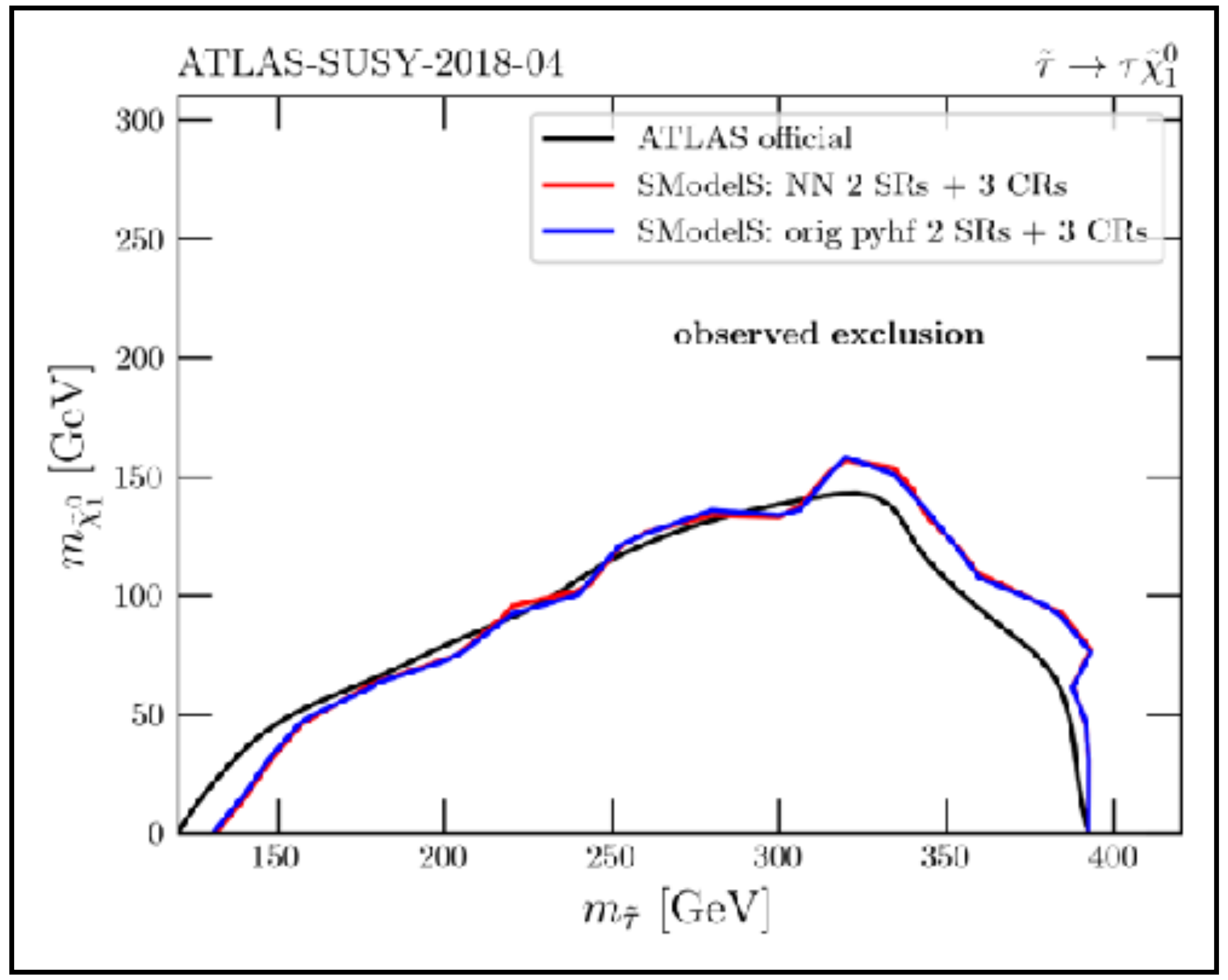
Fully connected network

$\Delta_{\text{exp}}^{(\mu_1,\ldots\mu_N)}$

$\Delta_{\text{exp Asimov}}^{(\mu_1,\ldots\mu_N)}$

$\Delta_{\text{obs}}^{(\mu_1,\ldots\mu_N)}$

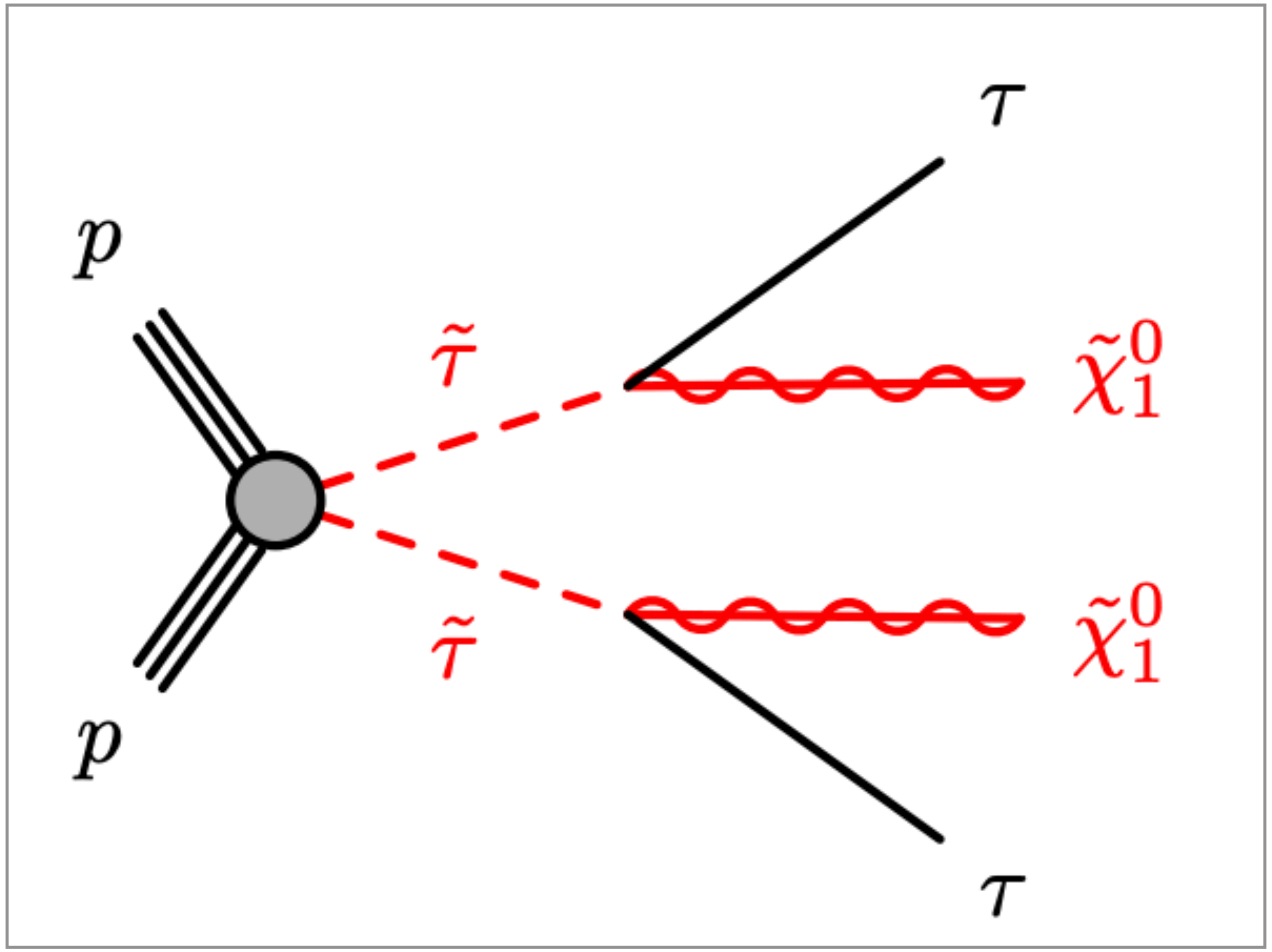
$\Delta_{\text{obs Asimov}}^{(\mu_1,\ldots\mu_N)}$



Moszna Castle, Poland

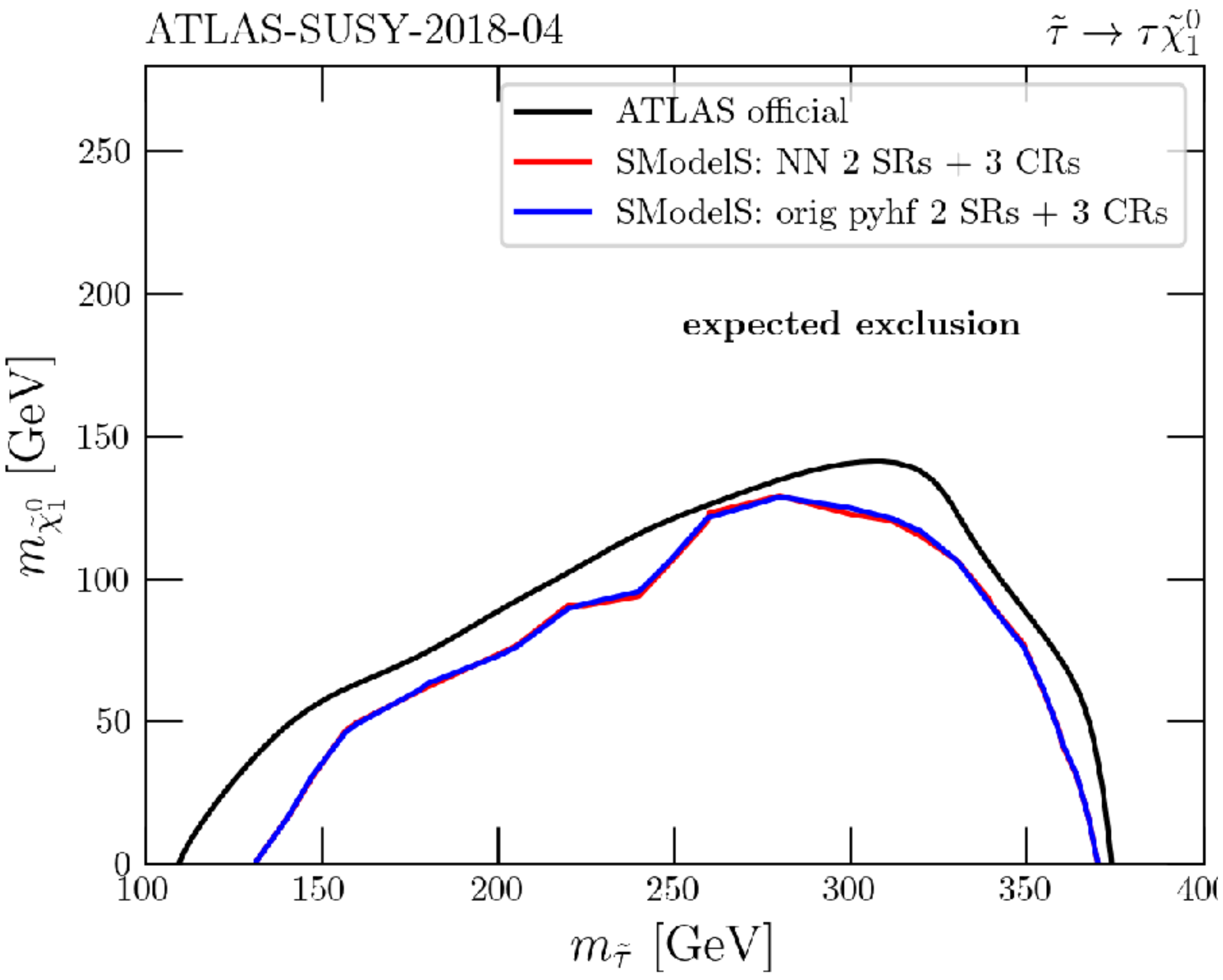
ATLAS-SUSY-2018-04 (1911.0660)

- Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector
- 2 signal bins, 3 control bins
- fit **without** the CRs

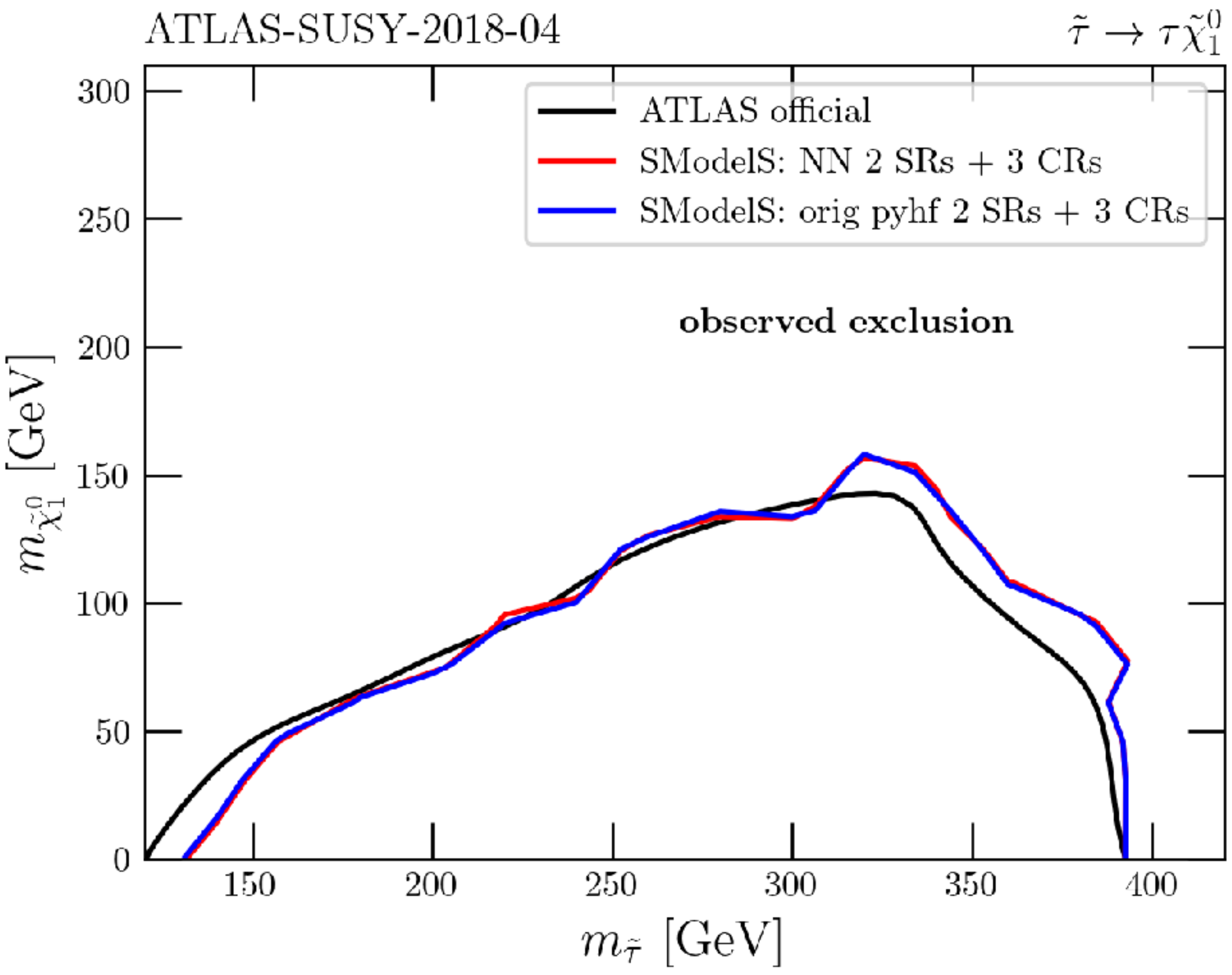


ATLAS-SUSY-2018-04 (1911.0660)

expected

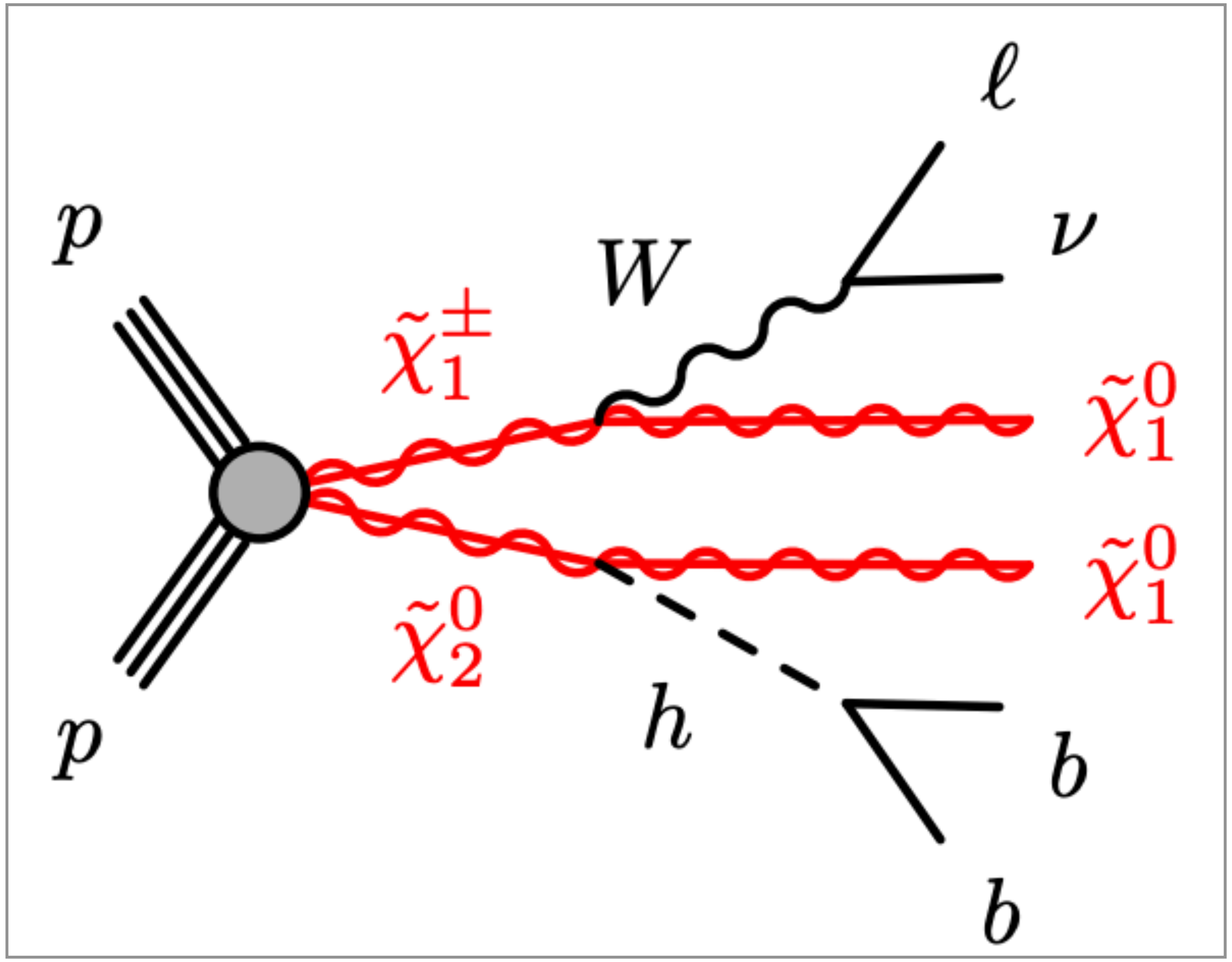


observed



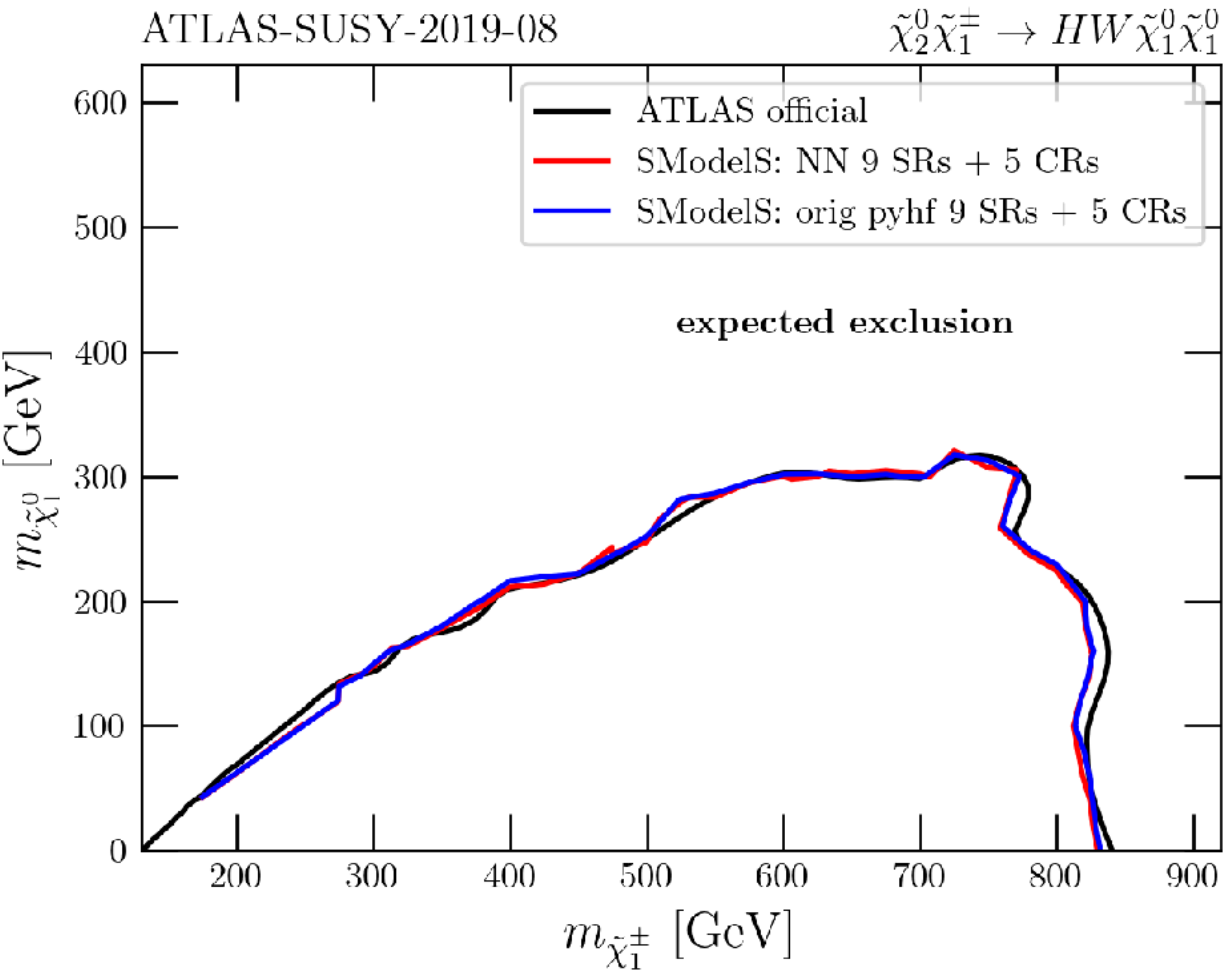
ATLAS-SUSY-2019-08 (1909.09926)

- Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector
- 9 signal bins, 5 control bins

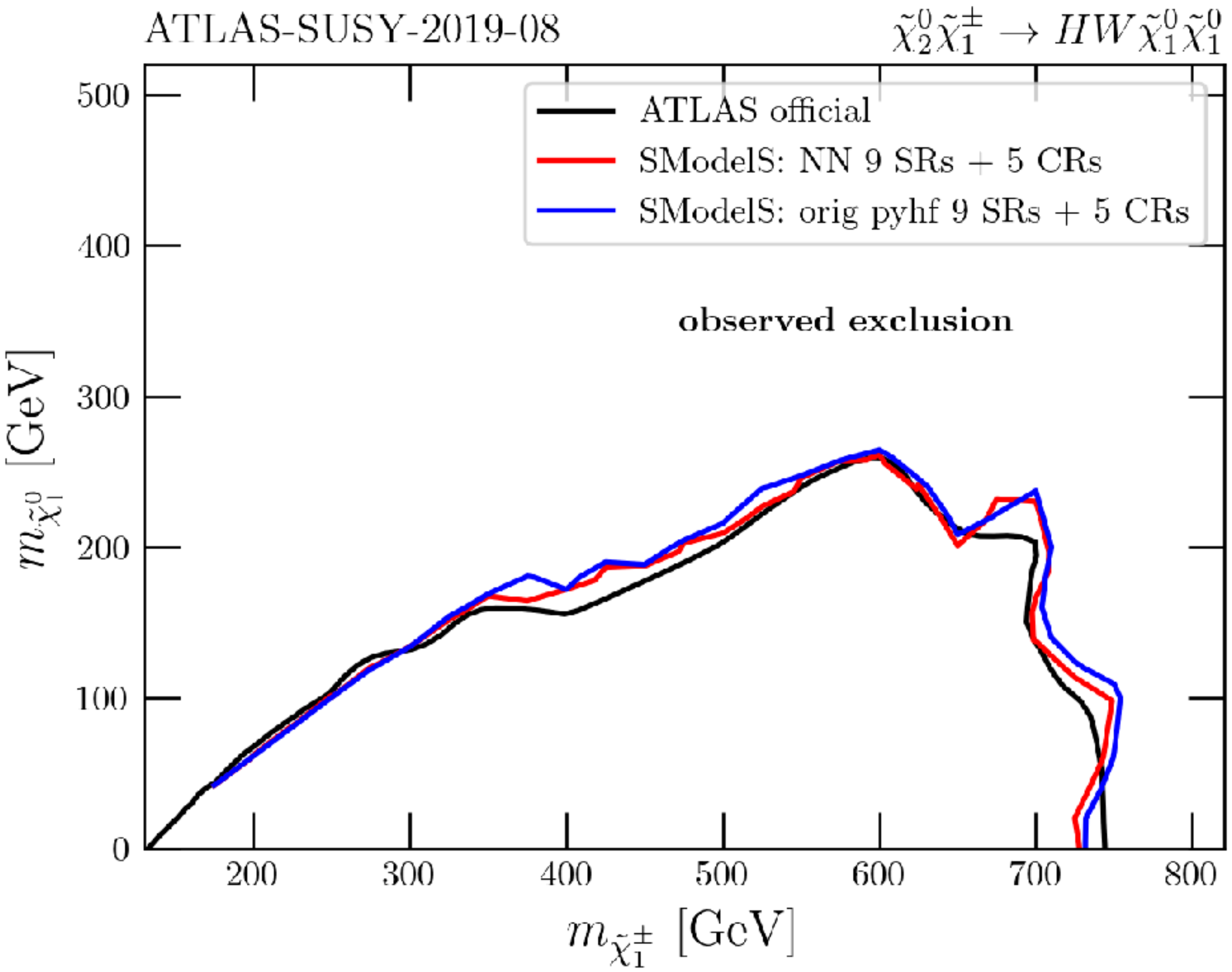


ATLAS-SUSY-2019-08 (1909.09926)

expected



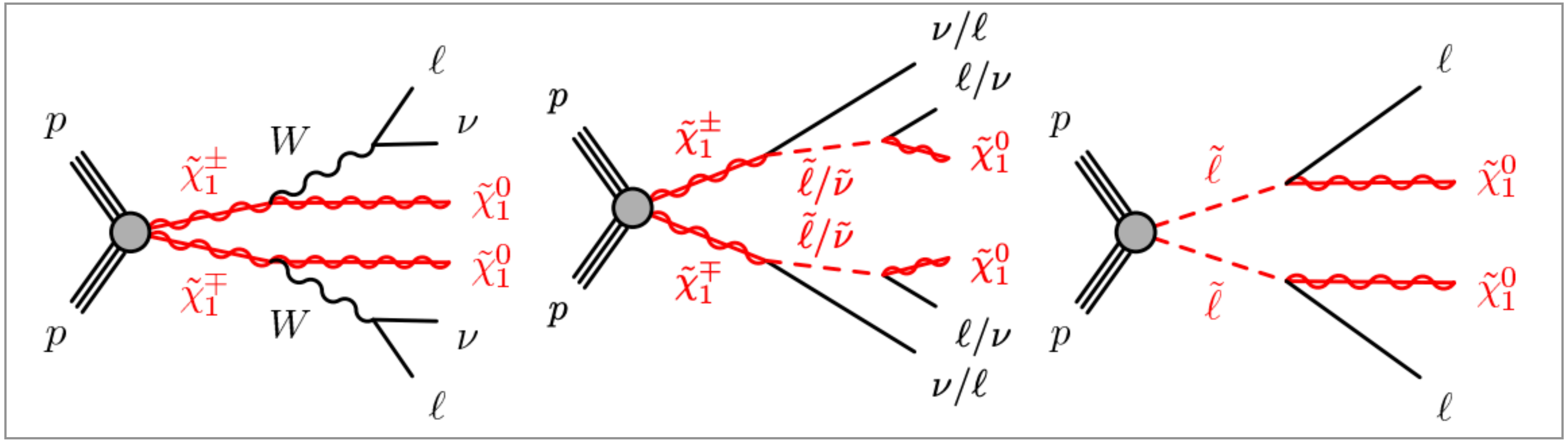
observed



ATLAS-SUSY-2018-32 (1908.08215)

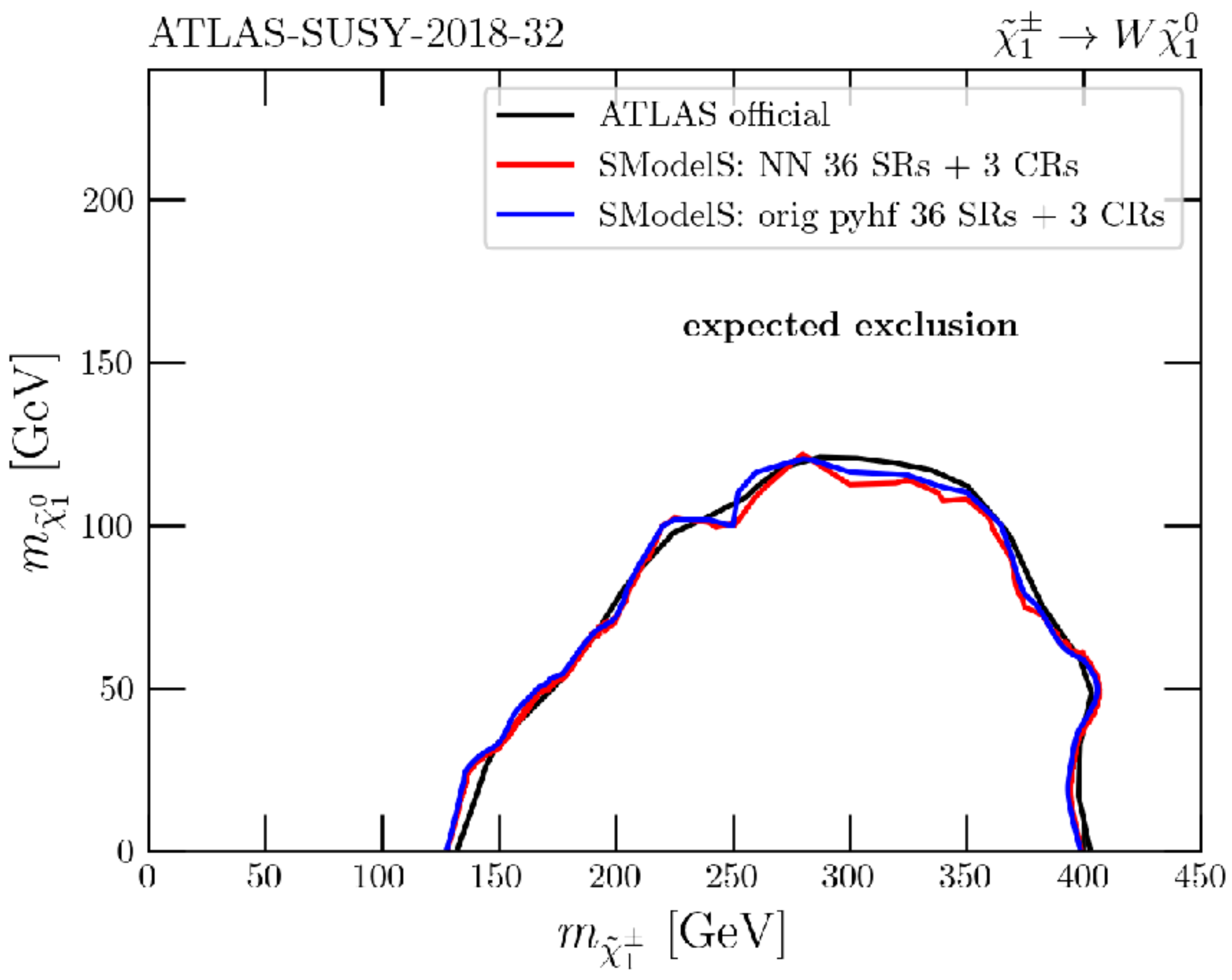
- Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in s=√13 TeV pp collisions using the ATLAS detector

- 36 SRs and 3 CRs

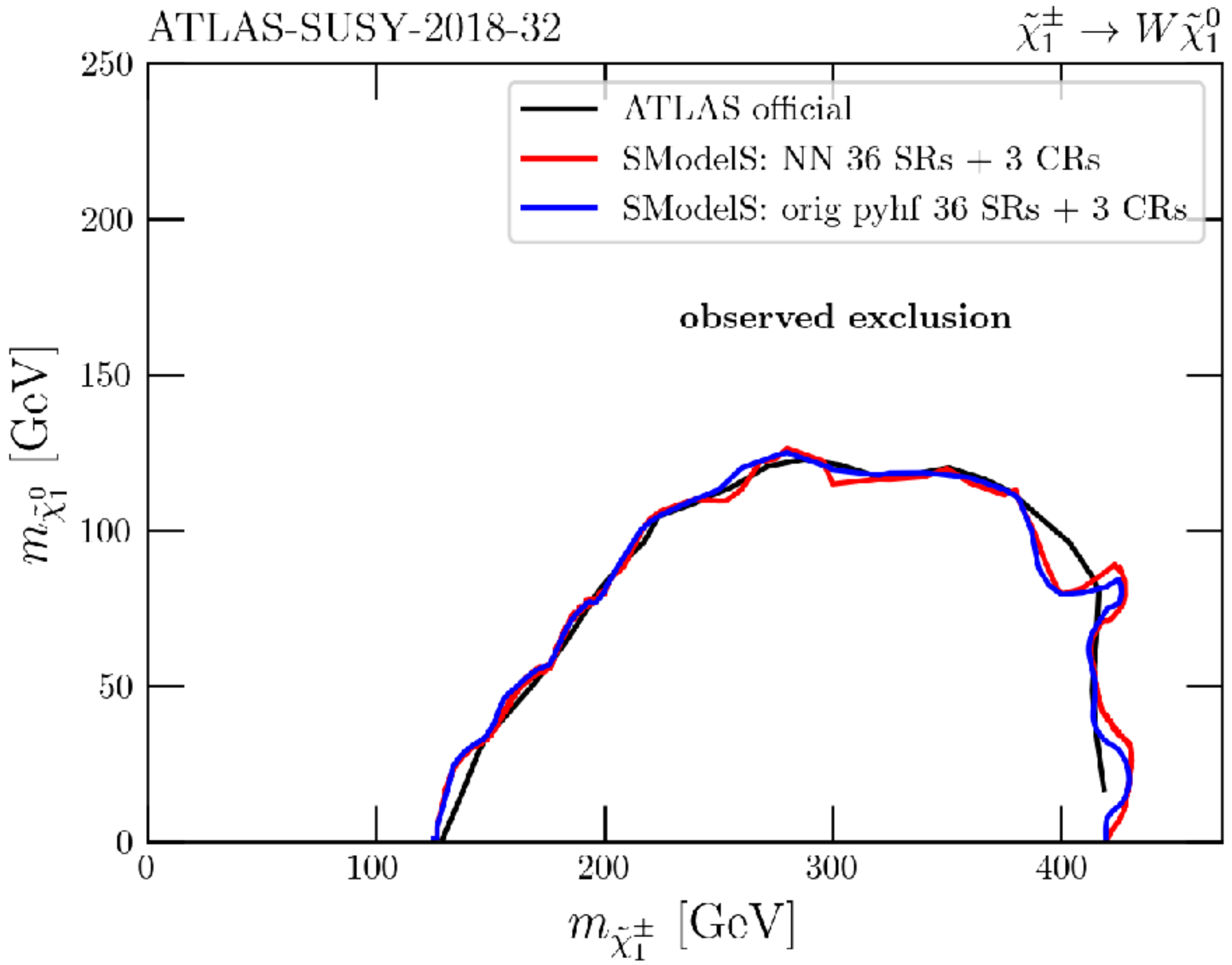


ATLAS-SUSY-2018-32 (1908.08215)

expected

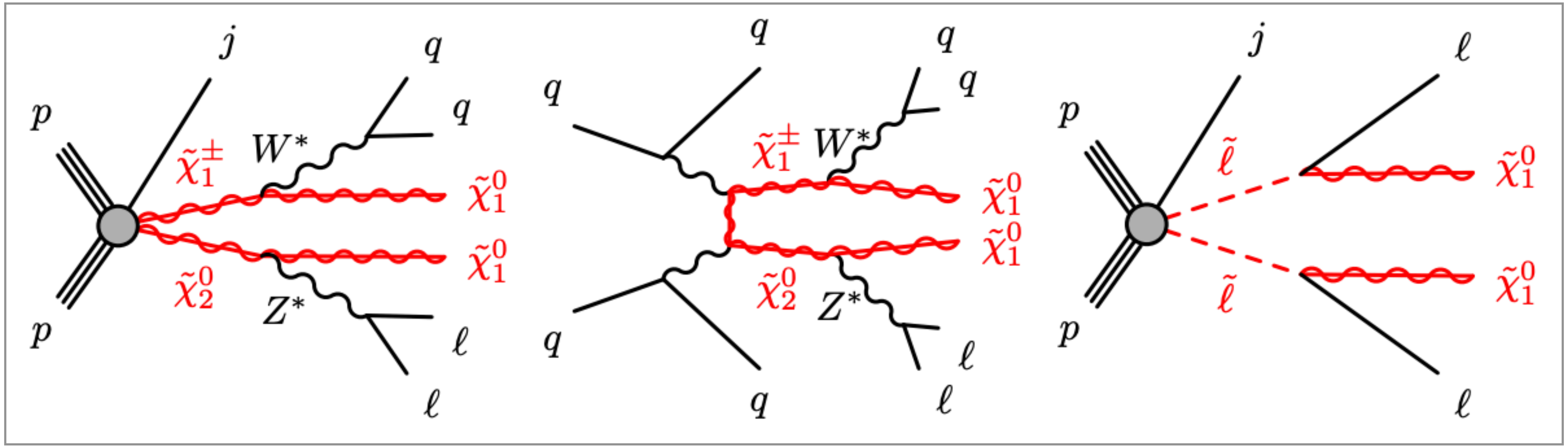


observed



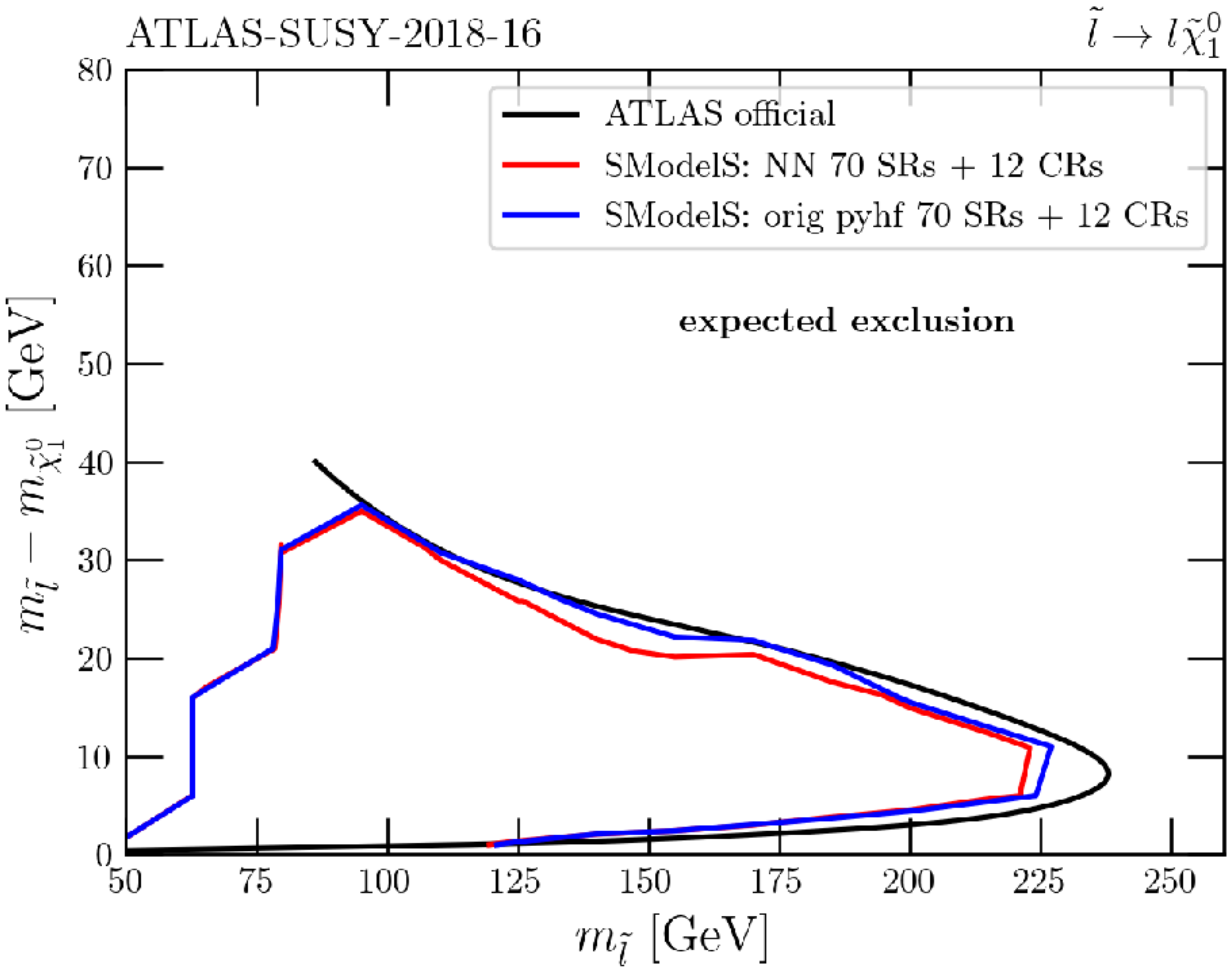
ATLAS-SUSY-2018-16 (1911.12606)

- Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector
- 2 subanalyses: sleptons (32+6) and EWKinos (38+6)
- Subanalyses combination: 70 SRs + 12 CRs (disjoint in this particular case)

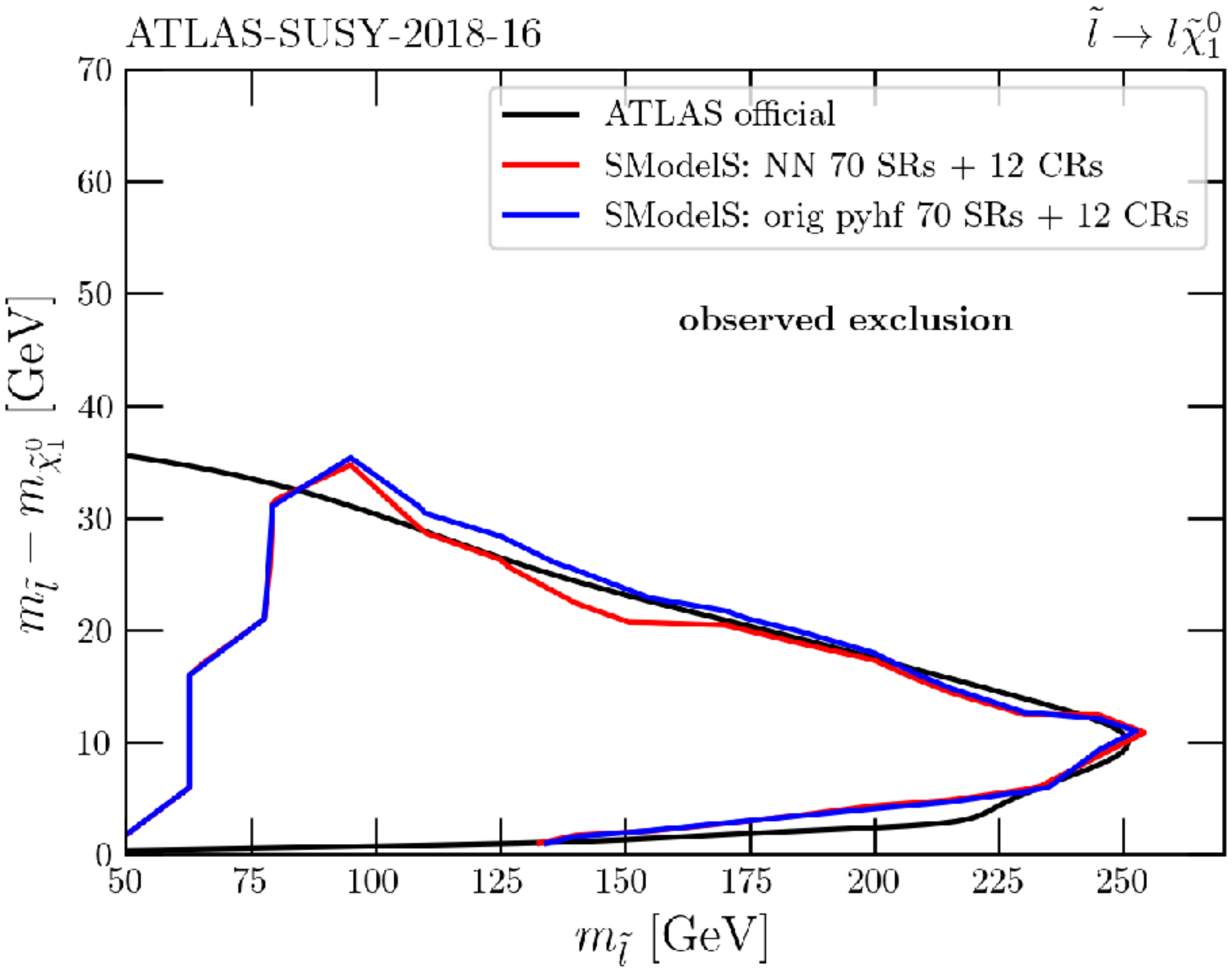


ATLAS-SUSY-2018-16 (1911.12606)

expected (sleptons)



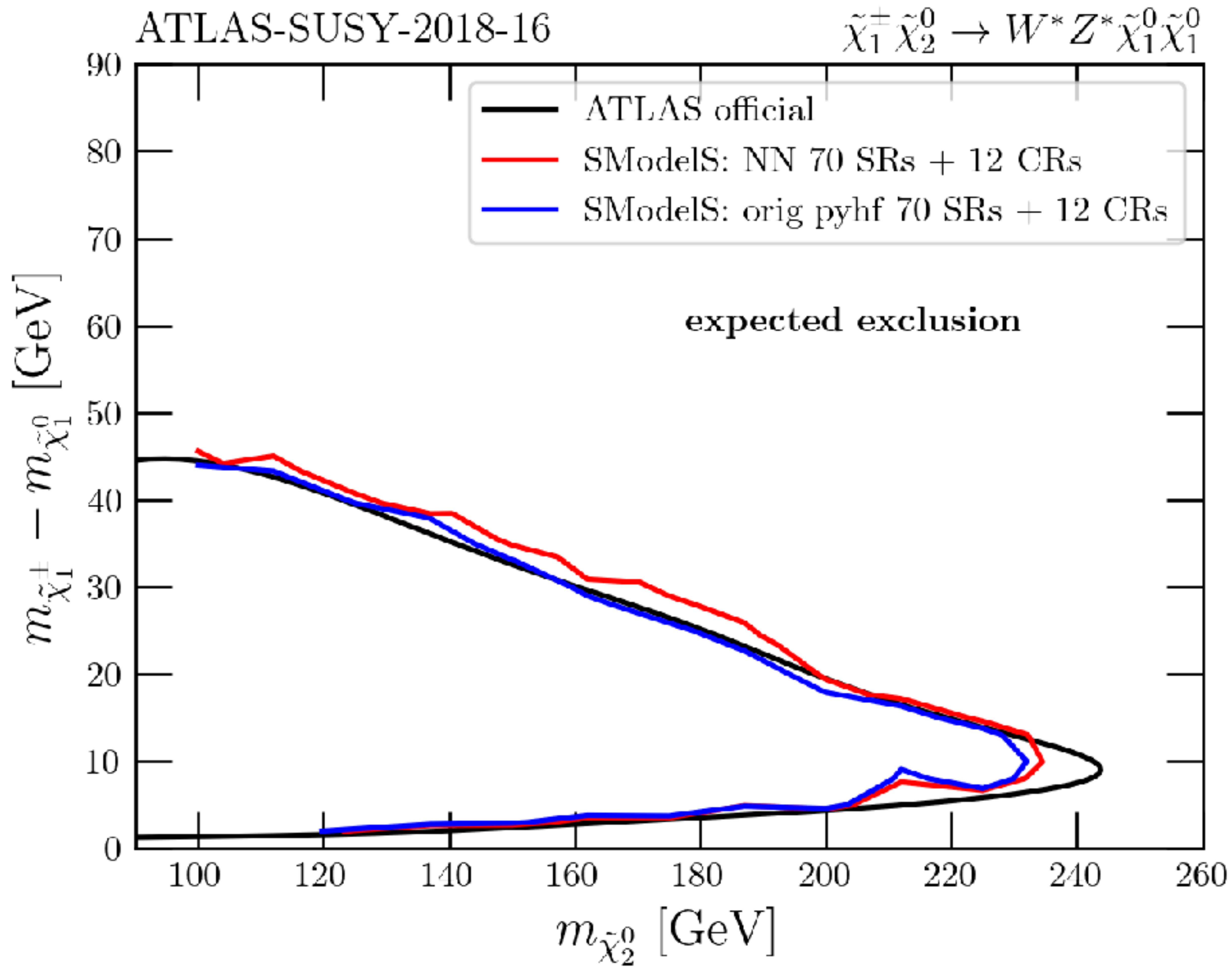
observed (sleptons)

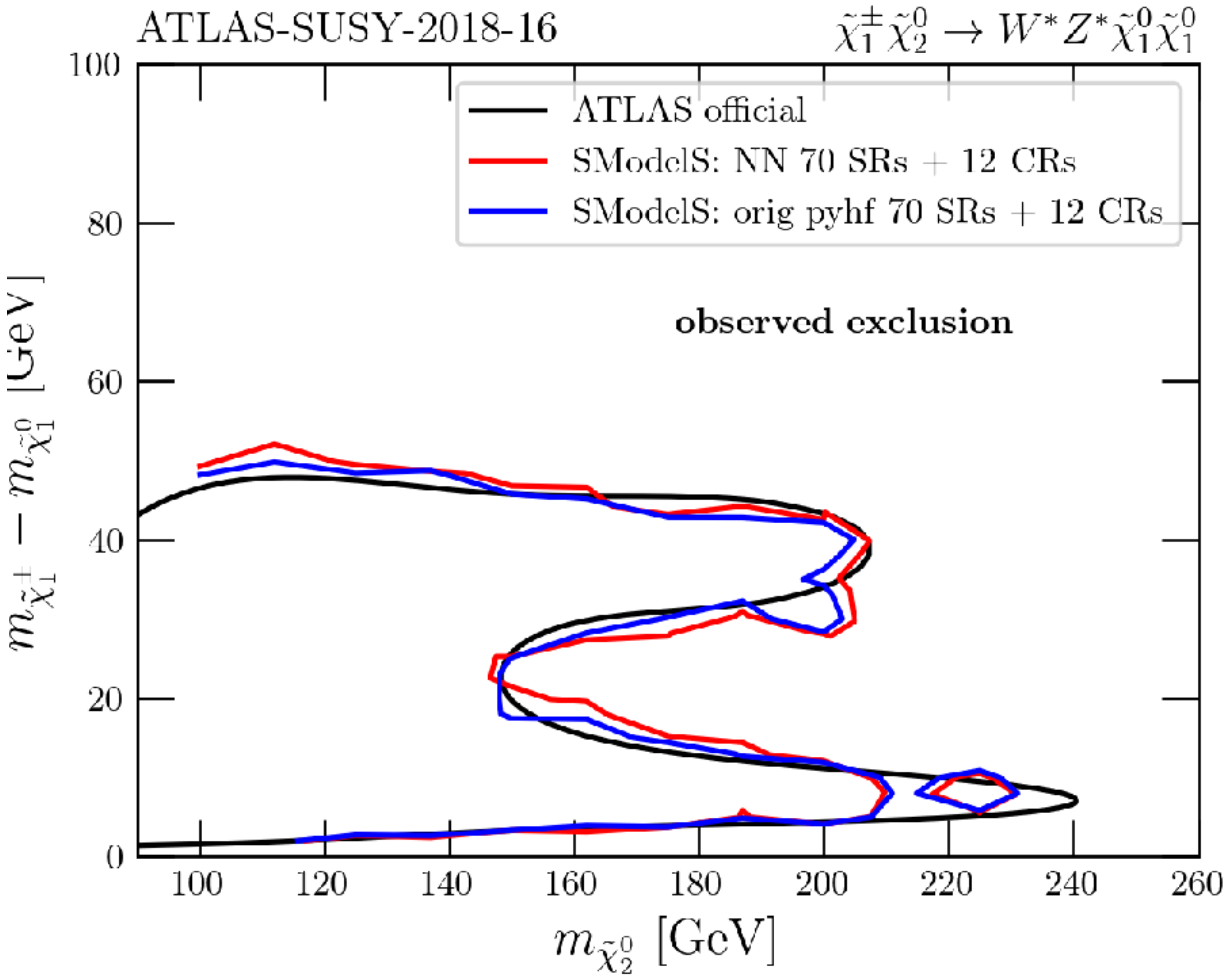


ATLAS-SUSY-2018-16 (1911.12606)

expected (EWKinos)

observed (EWKinos)





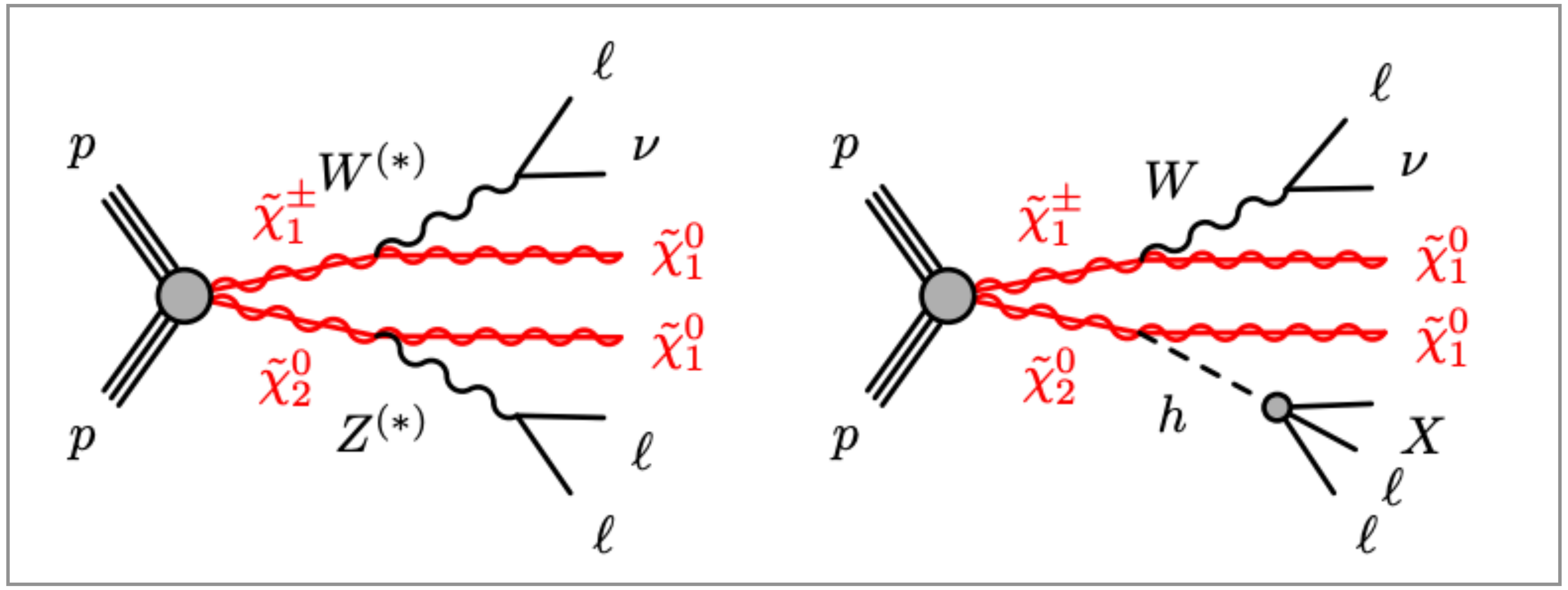
Książ Castle, Poland

ATLAS-SUSY-2019-09 (2106.01676)

- Search for chargino-neutralino pair production in final states with three leptons and missing transverse momentum in √s=13 TeV pp collisions with the ATLAS detector

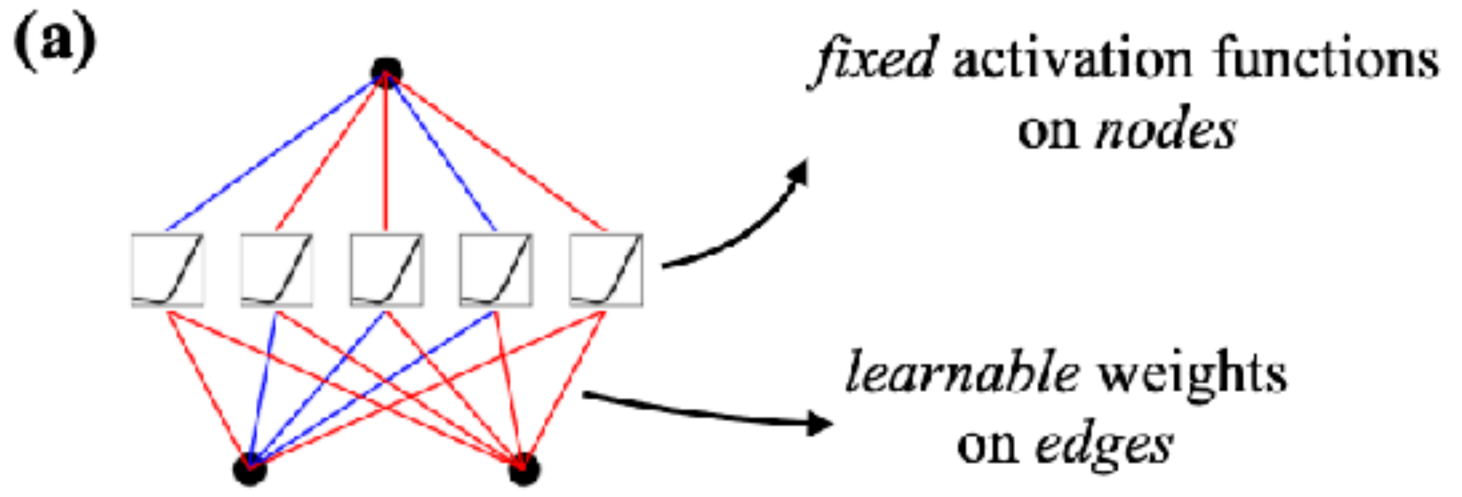
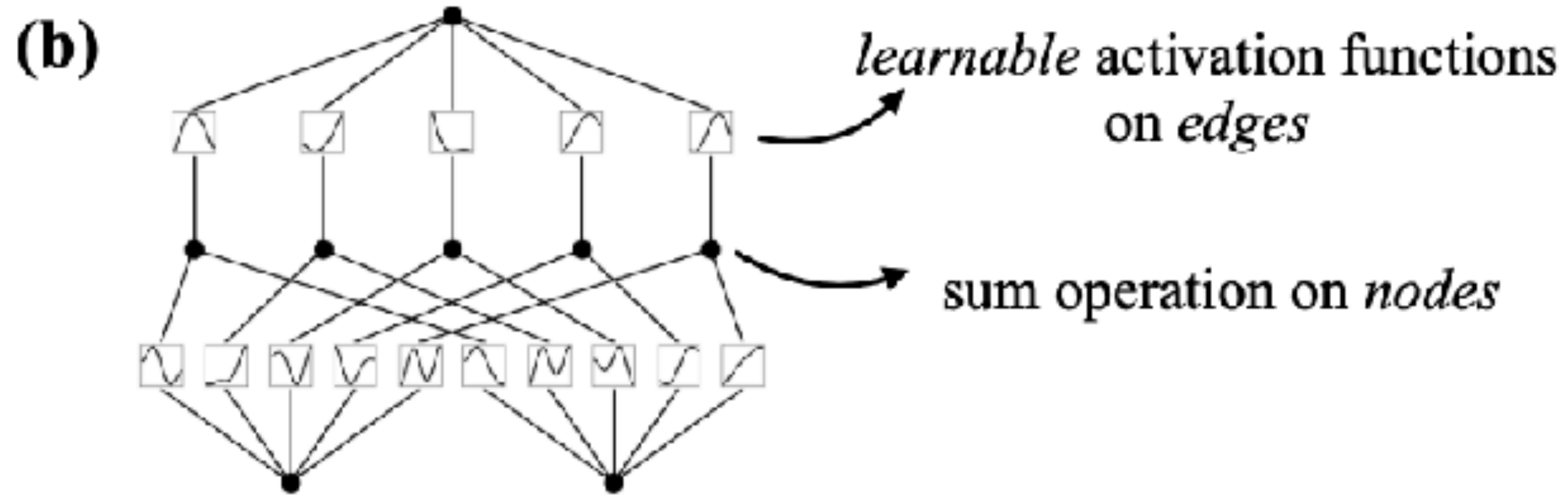
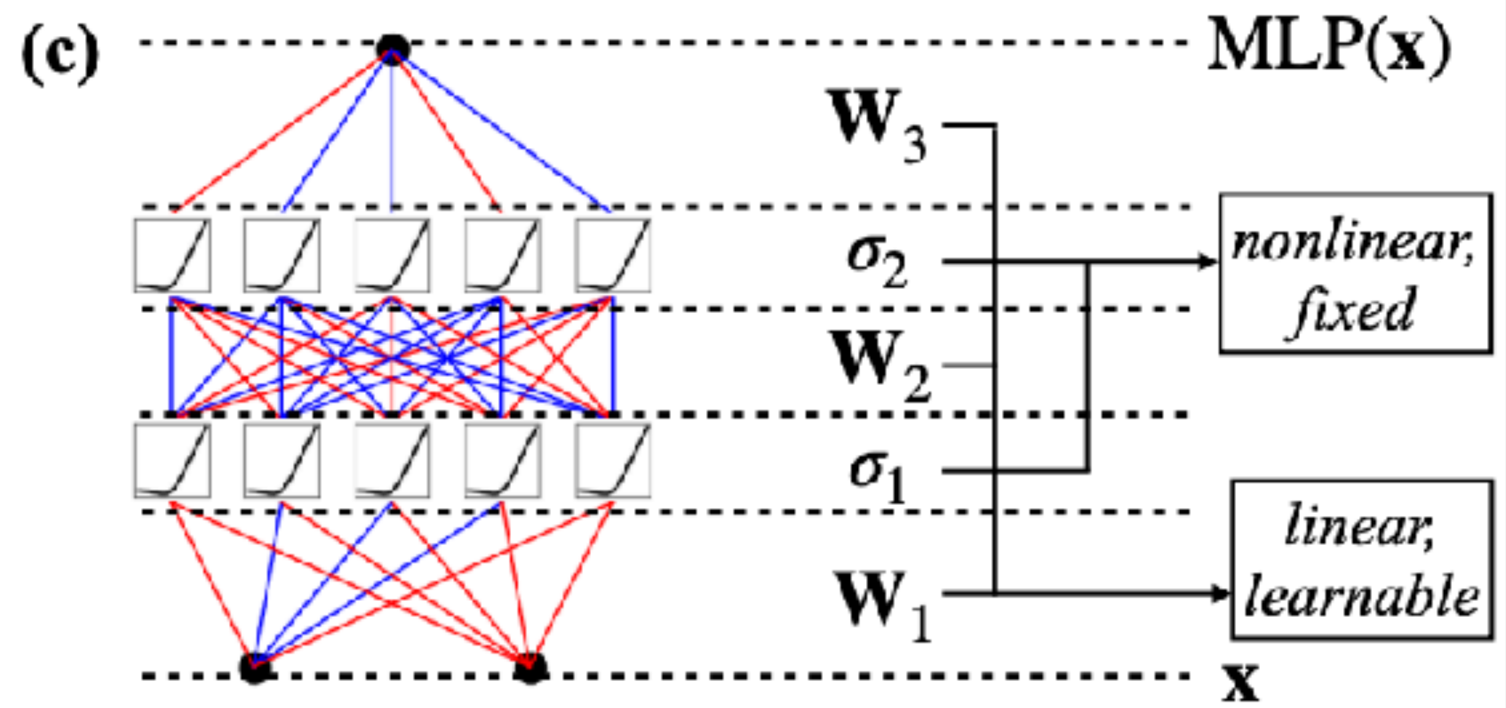
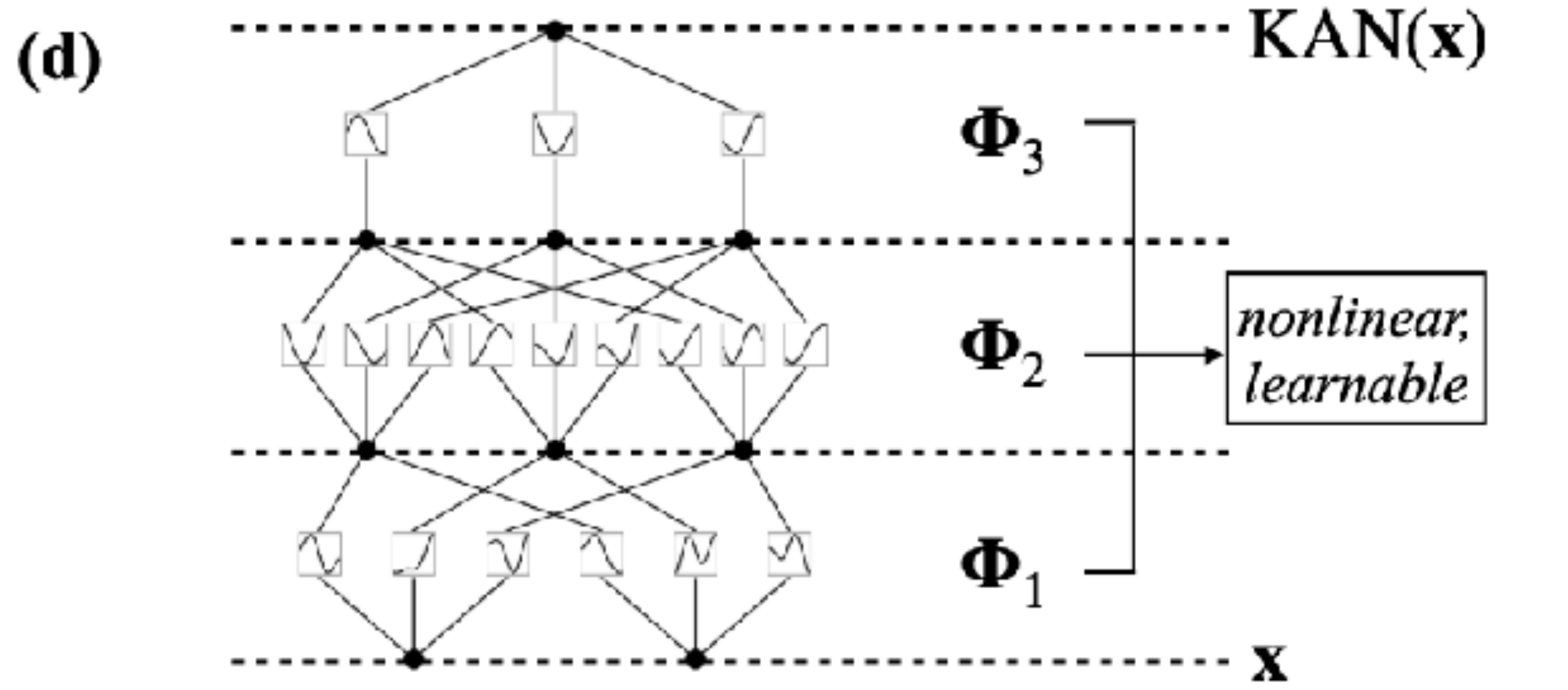
- 4 subanalyses:
  - **onshell** (W/Z when decaying) **wino/bino** (22+2)
  - **offshell wino/bino minus** (sign of the product of two neutralino mass eigenstates) (32+2)
  - **offshell wino/bino plus** (32+2)
  - **offshell higgsino** (32+2)

- Subanalyses combination: 54 SRs + 4 CRs (full overlap for offshell subanalyses)



Kolmogorov-Arnold networks

from Liu et al (2404.19756)

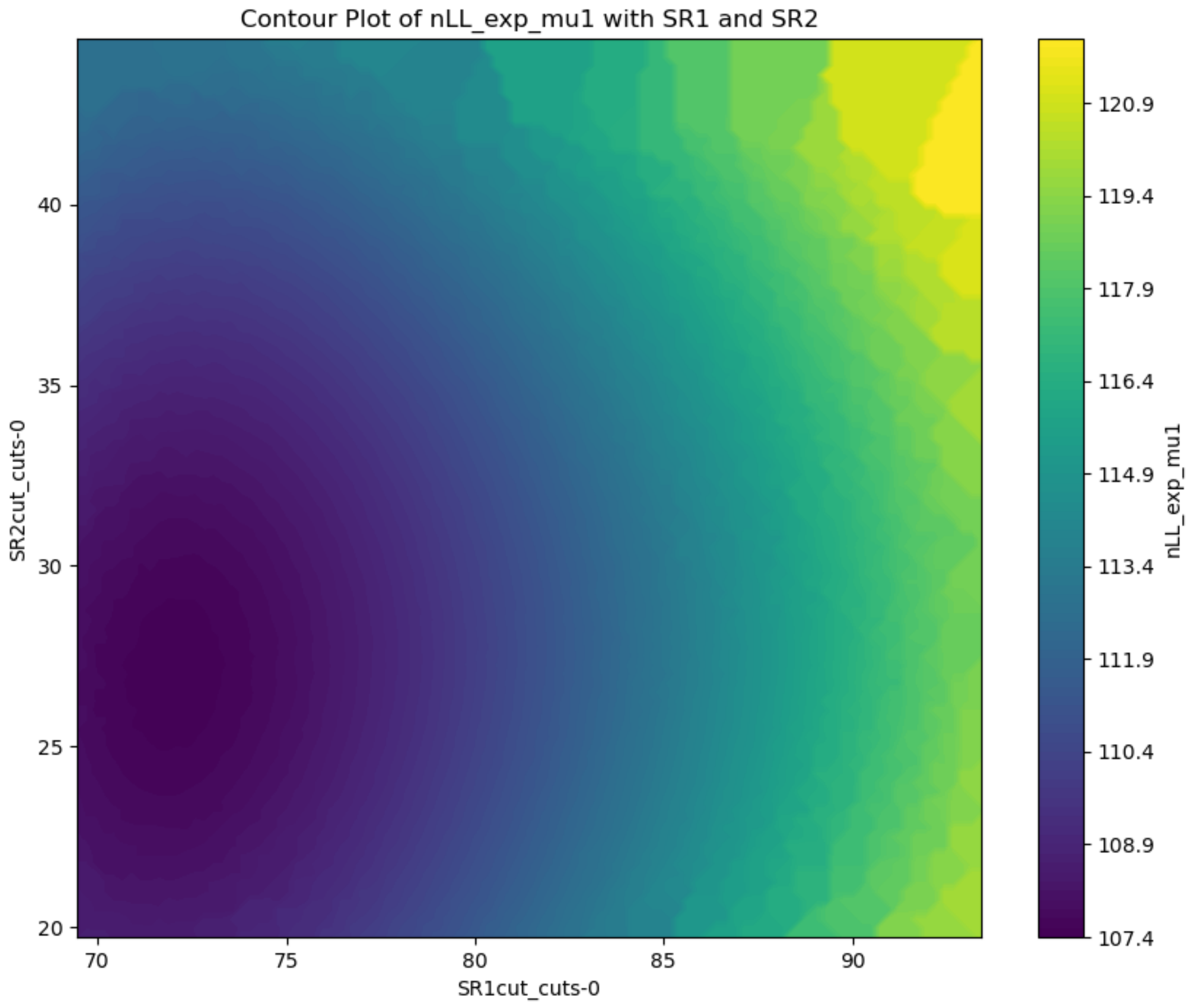
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)		
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)		

Kolmogorov-Arnold networks

- KANs allow to find analytical formulas describing data

- It might be possible, in some cases, to find analytical approximations to likelihood and speed up computations even further

ATLAS-SUSY-2018-04 (1911.0660)



Summary

- Selecting the right BSM model enhances chances of new physics discovery
- Reinterpretation allows to constrain untested models
- Proper hypotheses testing is possible with published Full Statistical Models
- However, using these models constitute a **computational bottleneck**
- Machine Learning surrogates offer **good accuracy with much** (orders of magnitude) **faster computations**
- Preliminary results with MLPs are promising
- The next step will be to try KANs
- Trained models will be published to be used by the community

Tatra mountains, Poland