

MACHINE LEARNING THE LIKELIHOODS

Rafał Masełek

in collaboration with

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

From imaging algorithms to quantum methods Seminar, 19-01-2026

Status of the Particle Physics

Standard Model of Elementary Particles

three generations of matter (fermions)				interactions / force carriers (bosons)	
	I	II	III		
mass	$\approx 2.16 \text{ MeV}/c^2$	$\approx 1.273 \text{ GeV}/c^2$	$\approx 172.57 \text{ GeV}/c^2$	0	$\approx 125.2 \text{ GeV}/c^2$
	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
charge	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
quarks	u up	c charm	t top	g gluon	H higgs
	$\approx 4.7 \text{ MeV}/c^2$	$\approx 93.5 \text{ MeV}/c^2$	$\approx 4.183 \text{ GeV}/c^2$	0	
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	
leptons	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.77693 \text{ GeV}/c^2$	$\approx 91.188 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
electron	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.77693 \text{ GeV}/c^2$	$\approx 91.188 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
muon	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.77693 \text{ GeV}/c^2$	$\approx 91.188 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
tau	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.77693 \text{ GeV}/c^2$	$\approx 91.188 \text{ GeV}/c^2$	
	-1	-1	-1	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
neutrino	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
electron neutrino	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
muon neutrino	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
tau neutrino	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
Z boson	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
W boson	$< 0.8 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$	$\approx 80.3692 \text{ GeV}/c^2$	
	0	0	0	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	

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

Status of the Particle Physics

Standard Model of Elementary Particles

! Dark Matter

! matter-
antimatter
asymmetry

! neutrino
masses

three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III	
QUARKS	$\approx 2.16 \text{ MeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ u up	$\approx 1.273 \text{ GeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ c charm	$\approx 172.57 \text{ GeV}/c^2$ $\frac{2}{3}$ $\frac{1}{2}$ t top	0 0 1 g gluon
	$\approx 4.7 \text{ MeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ d down	$\approx 93.5 \text{ MeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ s strange	$\approx 4.183 \text{ GeV}/c^2$ $-\frac{1}{3}$ $\frac{1}{2}$ b bottom	0 0 1 γ photon
	$\approx 0.511 \text{ MeV}/c^2$ -1 $\frac{1}{2}$ e electron	$\approx 105.66 \text{ MeV}/c^2$ -1 $\frac{1}{2}$ μ muon	$\approx 1.77693 \text{ GeV}/c^2$ -1 $\frac{1}{2}$ τ tau	$\approx 91.188 \text{ GeV}/c^2$ 0 1 Z Z boson
LEPTONS	$< 0.8 \text{ eV}/c^2$ 0 $\frac{1}{2}$ ν_e electron neutrino	$< 0.17 \text{ MeV}/c^2$ 0 $\frac{1}{2}$ ν_μ muon neutrino	$< 18.2 \text{ MeV}/c^2$ 0 $\frac{1}{2}$ ν_τ tau neutrino	$\approx 80.3692 \text{ GeV}/c^2$ ± 1 1 W W boson
				SCALAR BOSONS
				$\approx 125.2 \text{ GeV}/c^2$ 0 0 0 H higgs
				GAUGE BOSONS VECTOR BOSONS

! hierarchy
problem

! strong CP
problem

! quantum
gravity?

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

Status of the Particle Physics

Standard Model of Elementary Particles

! Dark Matter

! matter-
antimatter
asymmetry

! WE NEED
BSM PHYSICS

! hierarchy
problem

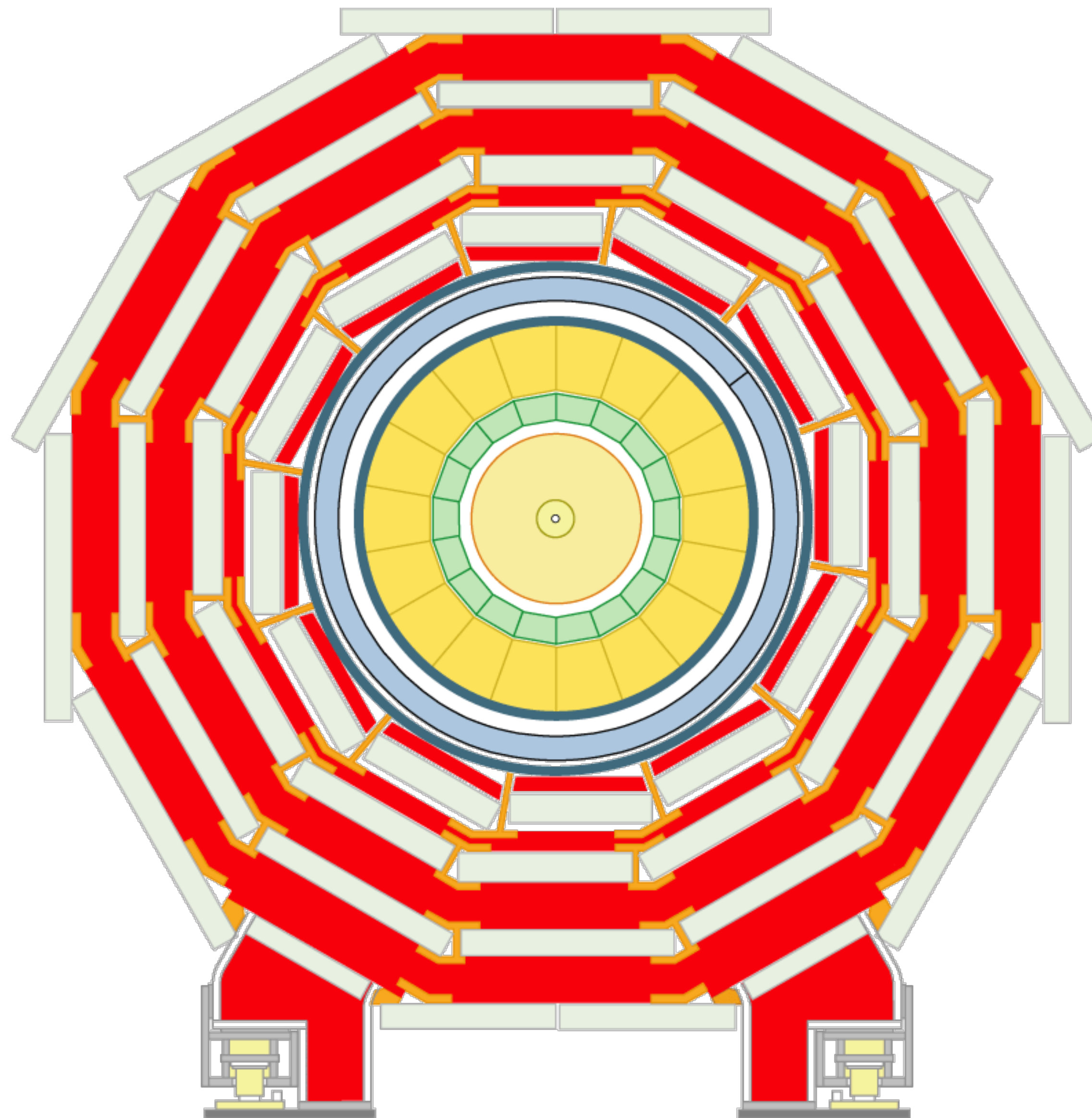
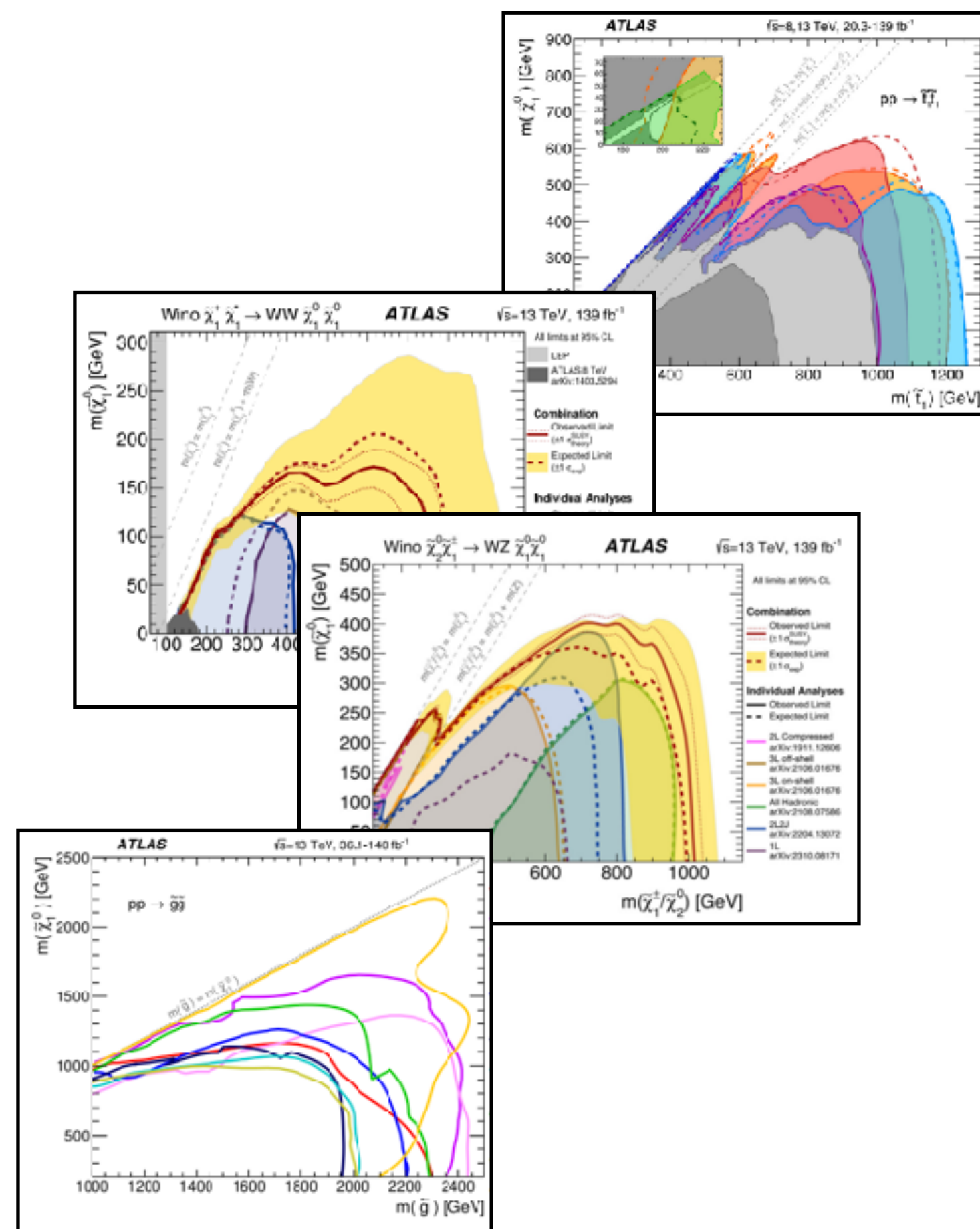
! strong CP
problem

! neutrino
masses

! quantum
gravity?

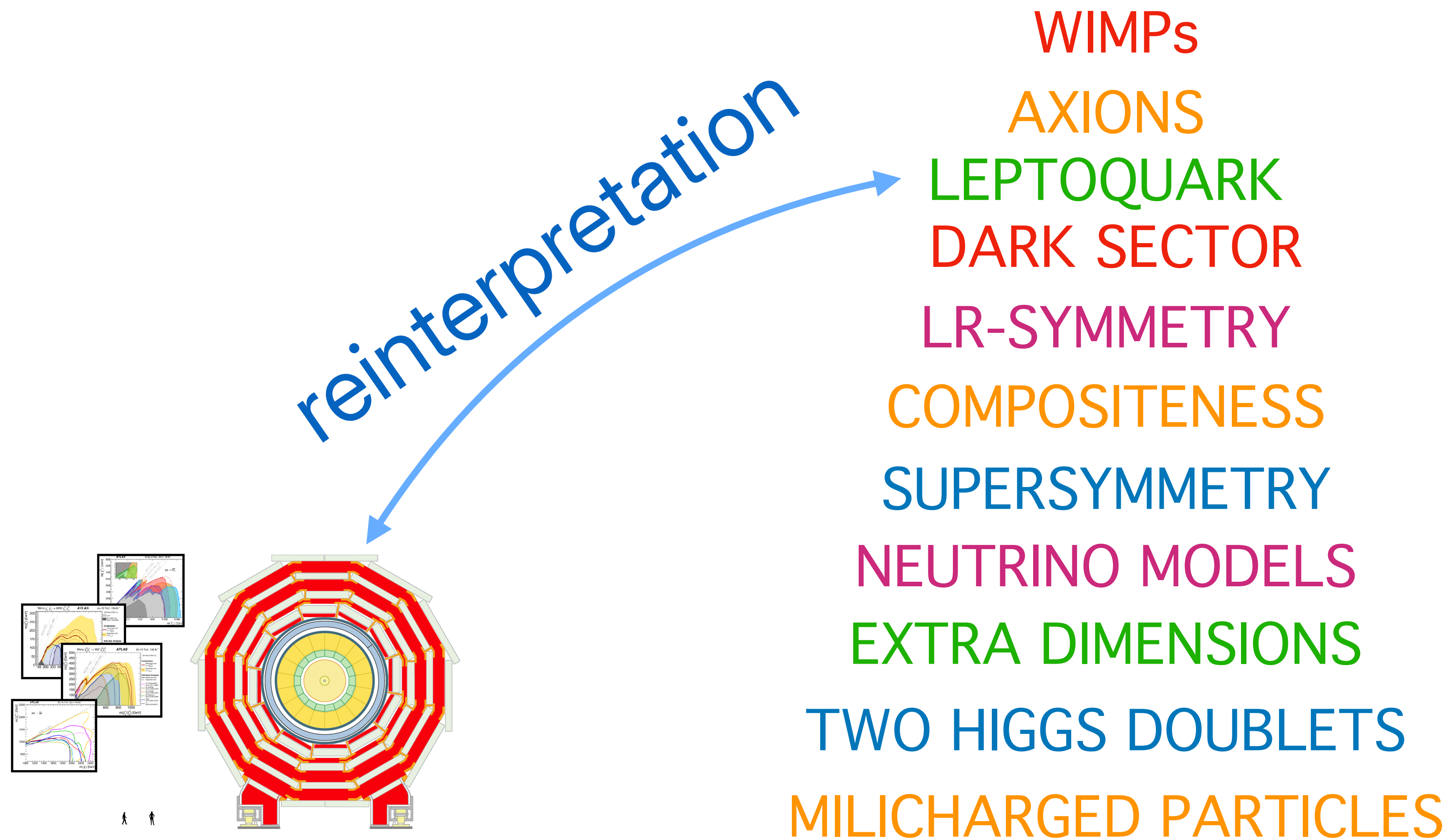


Searches for BSM at the LHC



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

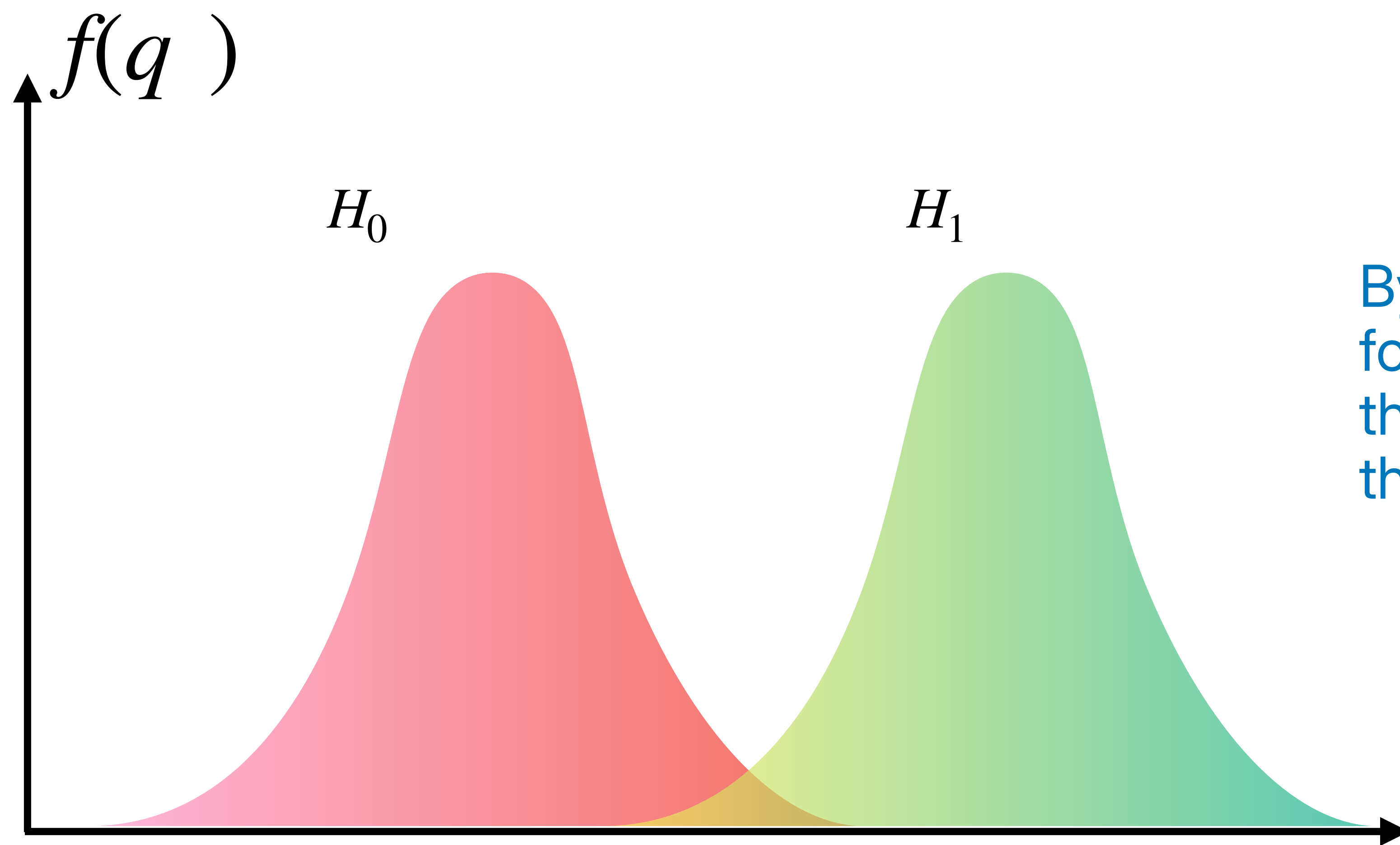
Searches for BSM at the LHC



Reinterpretation



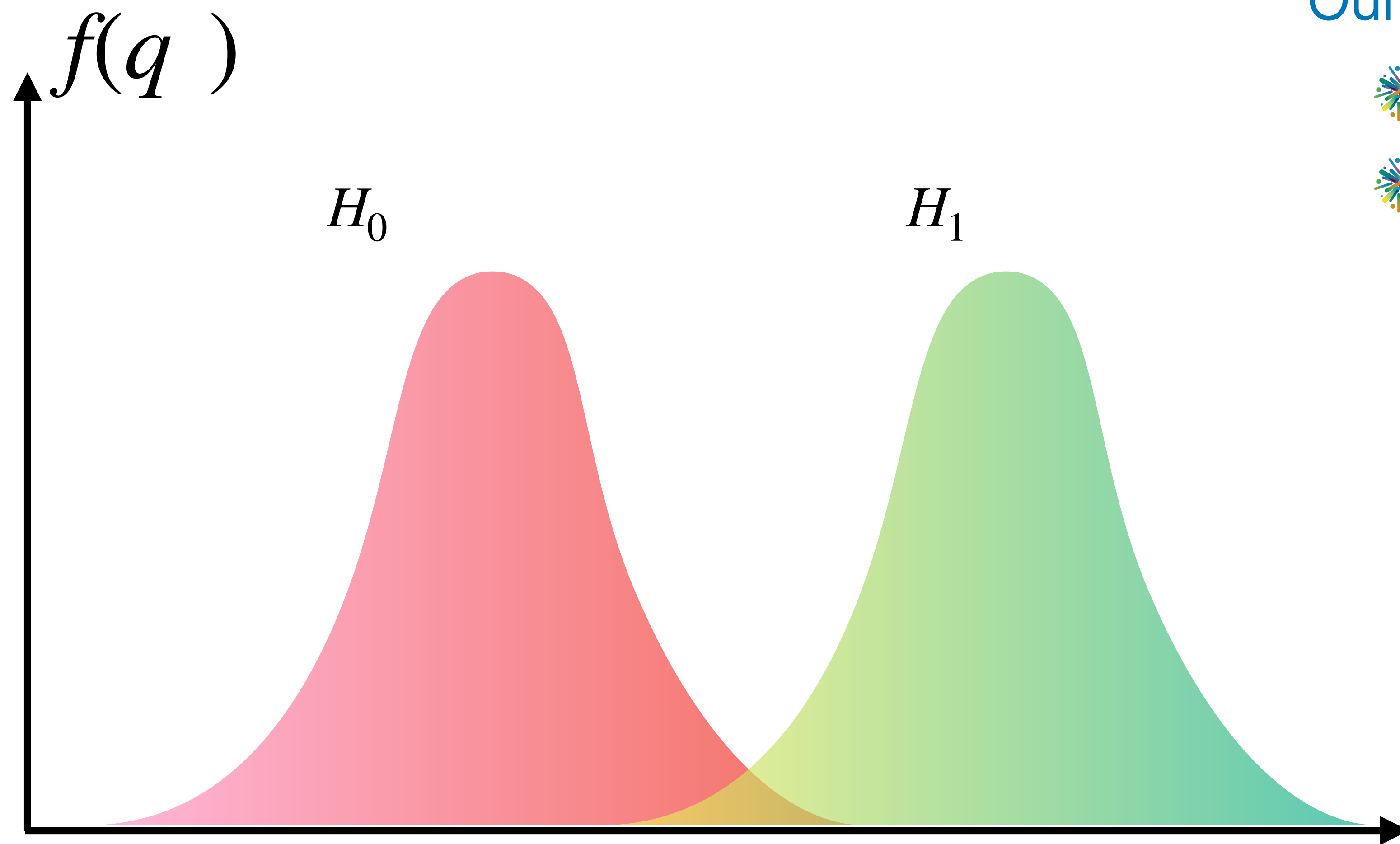
Hypothesis testing



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

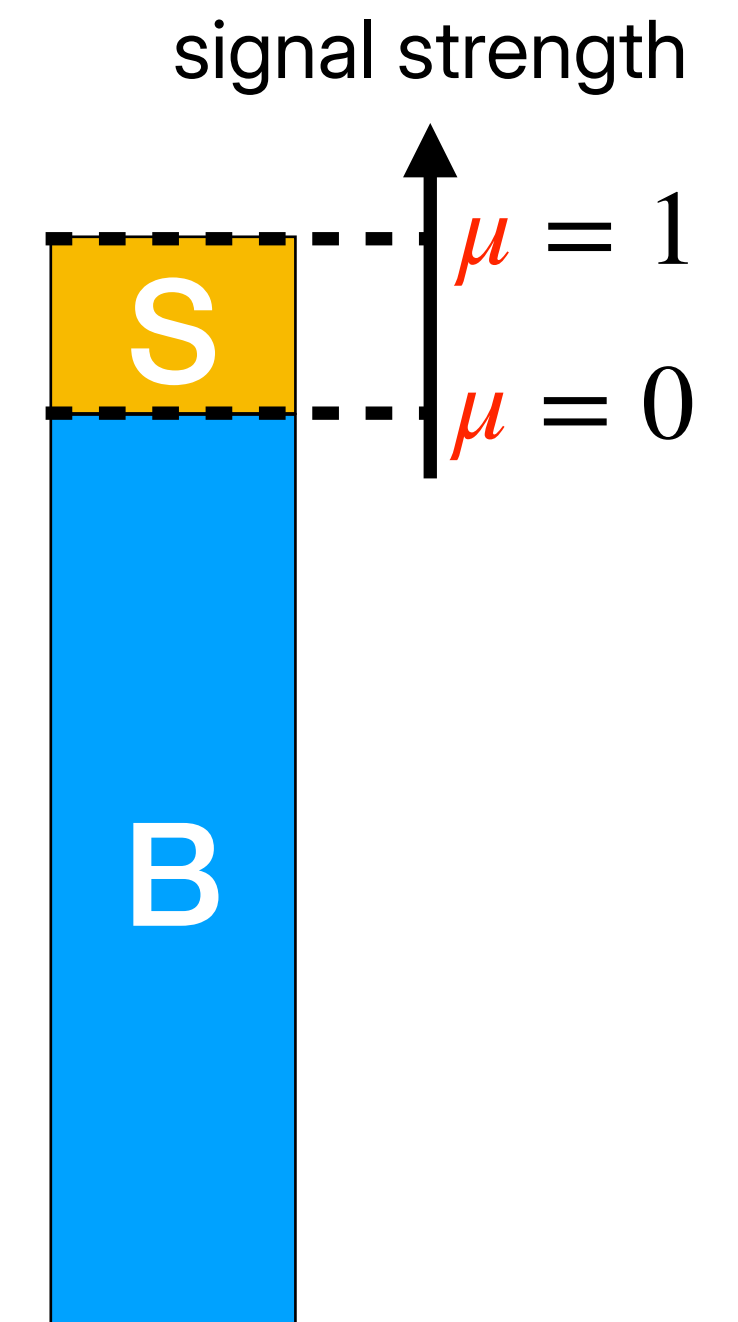
$$q = -2 \ln \frac{L(\text{data} | H_0)}{L(\text{data} | H_1)}$$

Hypothesis testing

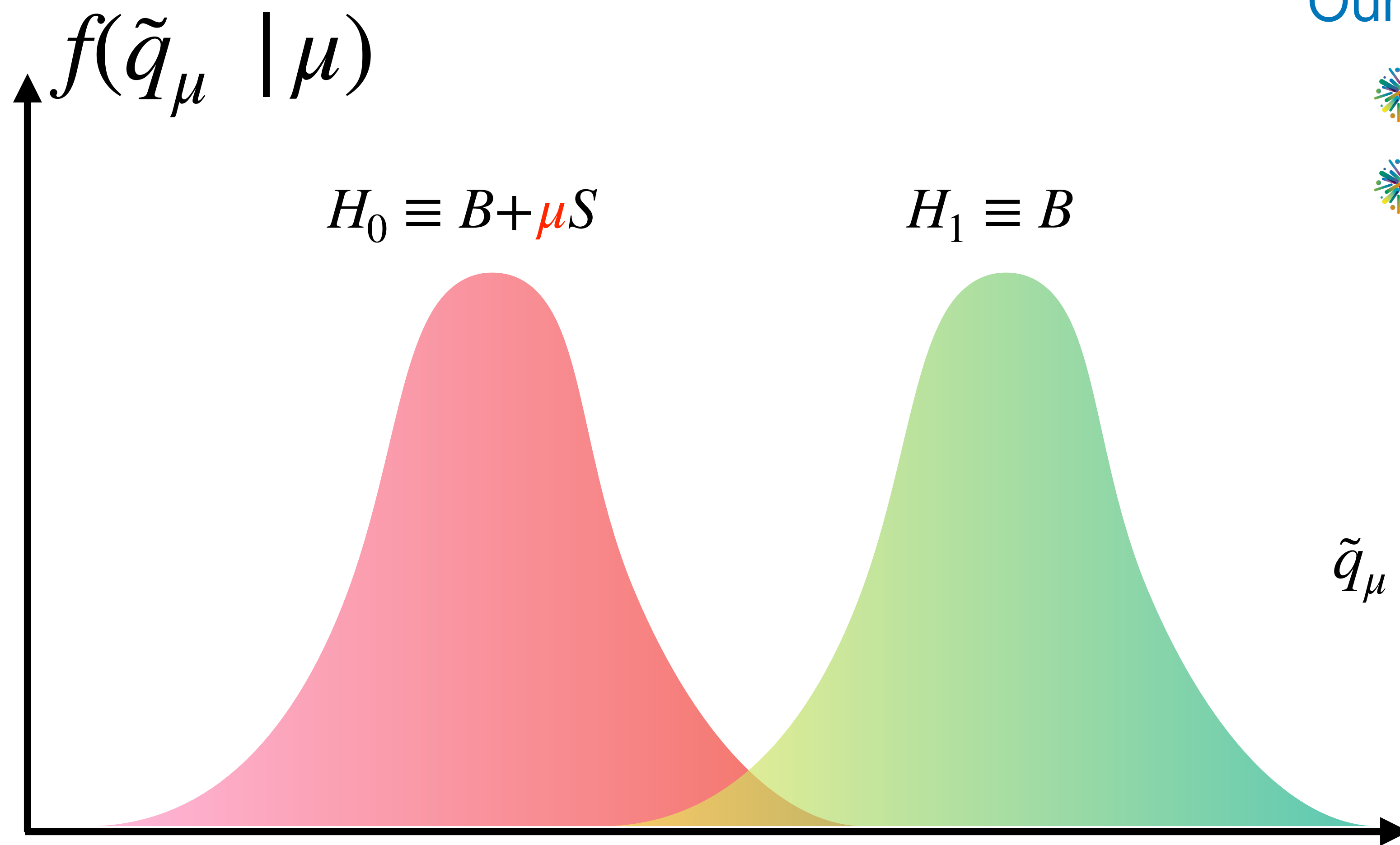


Our hypotheses are composite:

- ✿ We have nuisance parameters θ
- ✿ We have a parameter of interest μ



Hypothesis testing



Our hypotheses are composite:

✿ We have nuisance parameters θ

✿ We have a parameter of interest μ

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

$\hat{\square}$ unconditional MLE

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

Full likelihood models by ATLAS (Hist Factory)

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

Full likelihood models by ATLAS (Hist Factory)

channel data

auxiliary data

free parameters

constrained parameters

$$L(n, a | \mu, \theta) = \prod_c \prod_{b \in \text{bins}_c} \text{Pois}(n_{cb} | \nu_{cb}(\mu, \theta)) \prod_{\theta} p_{\theta}(a_{\theta} | \theta)$$

simultaneous measurement of multiple channels

constraint terms for "auxiliary measurements"

Full likelihood models by ATLAS (Hist Factory)

channel data auxiliary data

$$L(n, a | \mu, \theta) = \prod_c \prod_{b \in \text{bins}_c} \text{Pois}(n_{cb} | \nu_{cb}(\mu, \theta)) \prod_{\theta} p_{\theta}(a_{\theta} | \theta)$$

free parameters constrained parameters

simultaneous measurement of multiple channels constraint terms for "auxiliary measurements"

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

multiplicative modifiers const. nominal rate additive modifiers

Full likelihood models by ATLAS (Hist Factory)

- Full statistical models by ATLAS are available on HEPData
- They are provided as JSON files
- There are background files and signal patches
- Each patch corresponds to some signal point and contains modifiers to the background files
- There can be hundreds of modifiers
- Spey/PyHF can load and process these files

```
"patch": [  
  {  
    "op": "add",  
    "path": "/channels/0/samples/0",  
    "value": {  
      "data": [  
        2.3051342964172363  
      ],  
      "modifiers": [  
        {  
          "data": null,  
          "name": "lumi",  
          "type": "lumi"  
        },  
        {  
          "data": [  
            0.6571804118166927  
          ],  
          "name": "staterior_QCR1cut_cuts",  
          "type": "staterior"  
        },  
        {  
          "data": {  
            "hi": 1.06675,  
            "lo": 0.911403  
          },  
          "name": "PRW_DATASF",  
          "type": "normsys"  
        }  
      ]  
    }  
  ]  
}
```

Why bother?

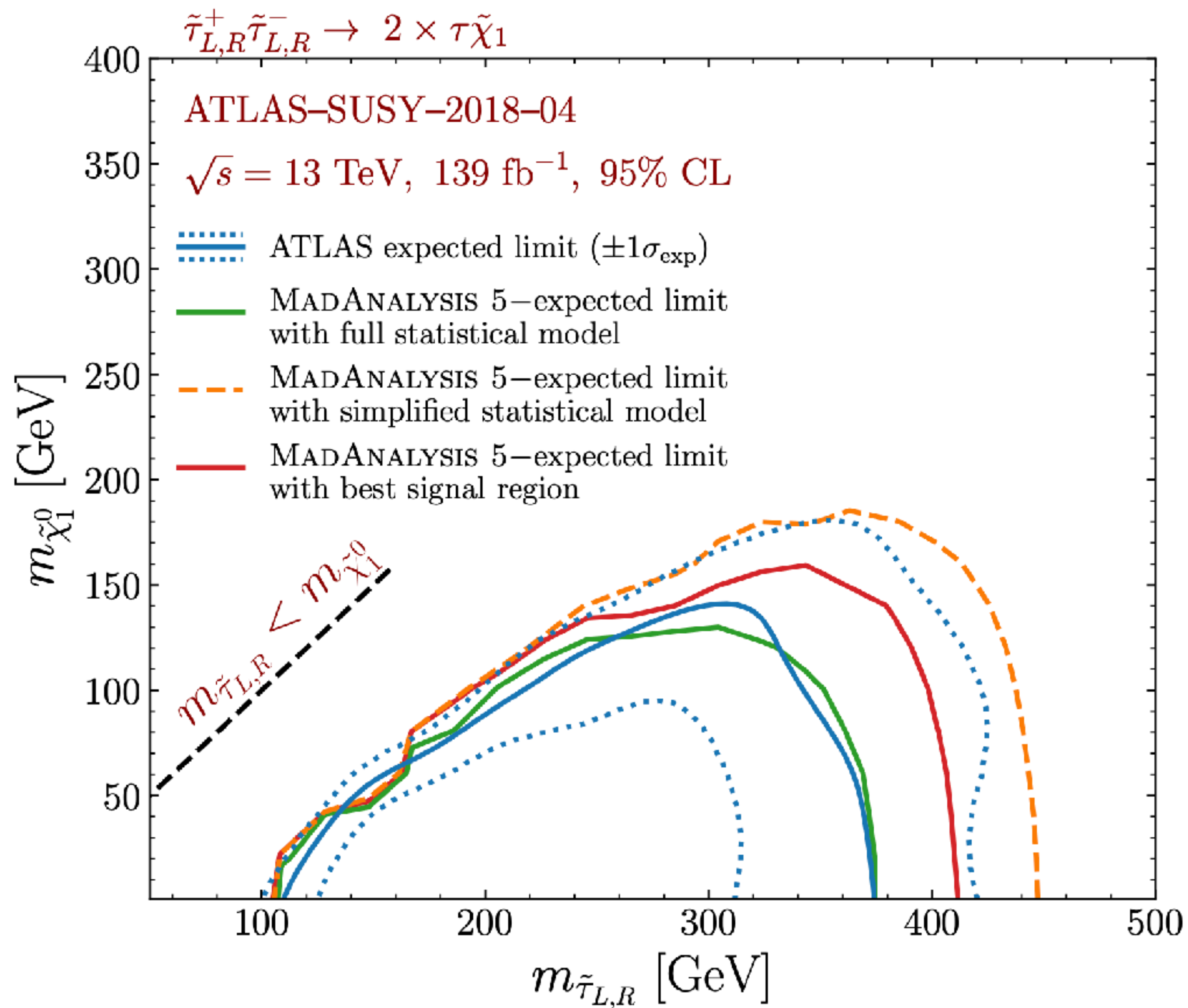
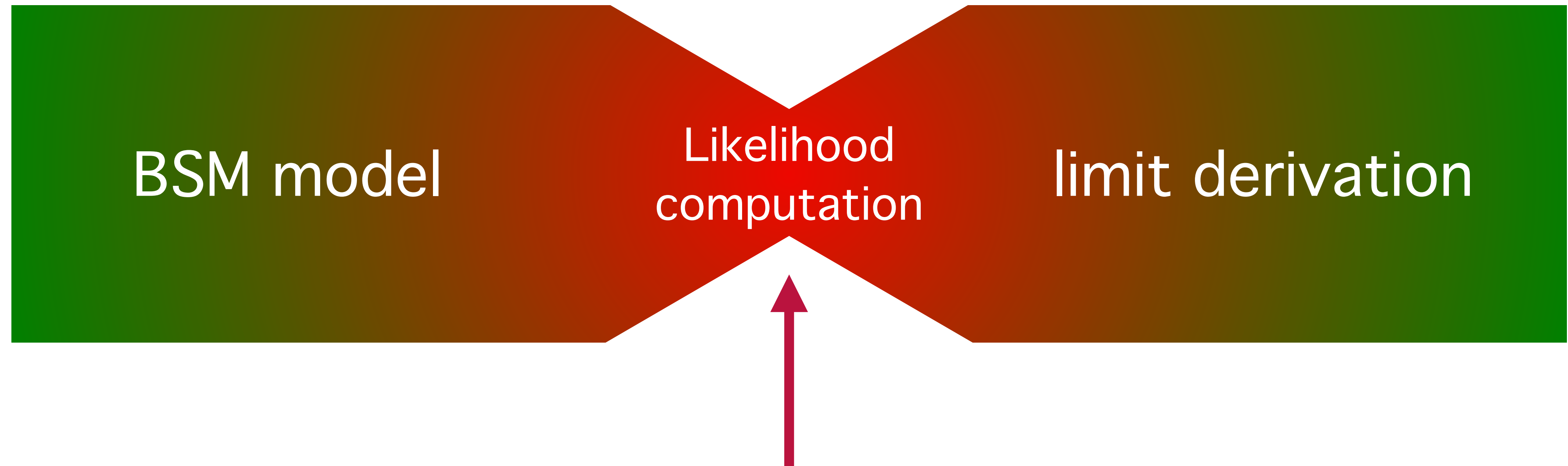


Figure from Gael Alguero's PhD thesis
<https://theses.hal.science/tel-04076269v1>



Full statistical model calculations enter here

Computational bottleneck

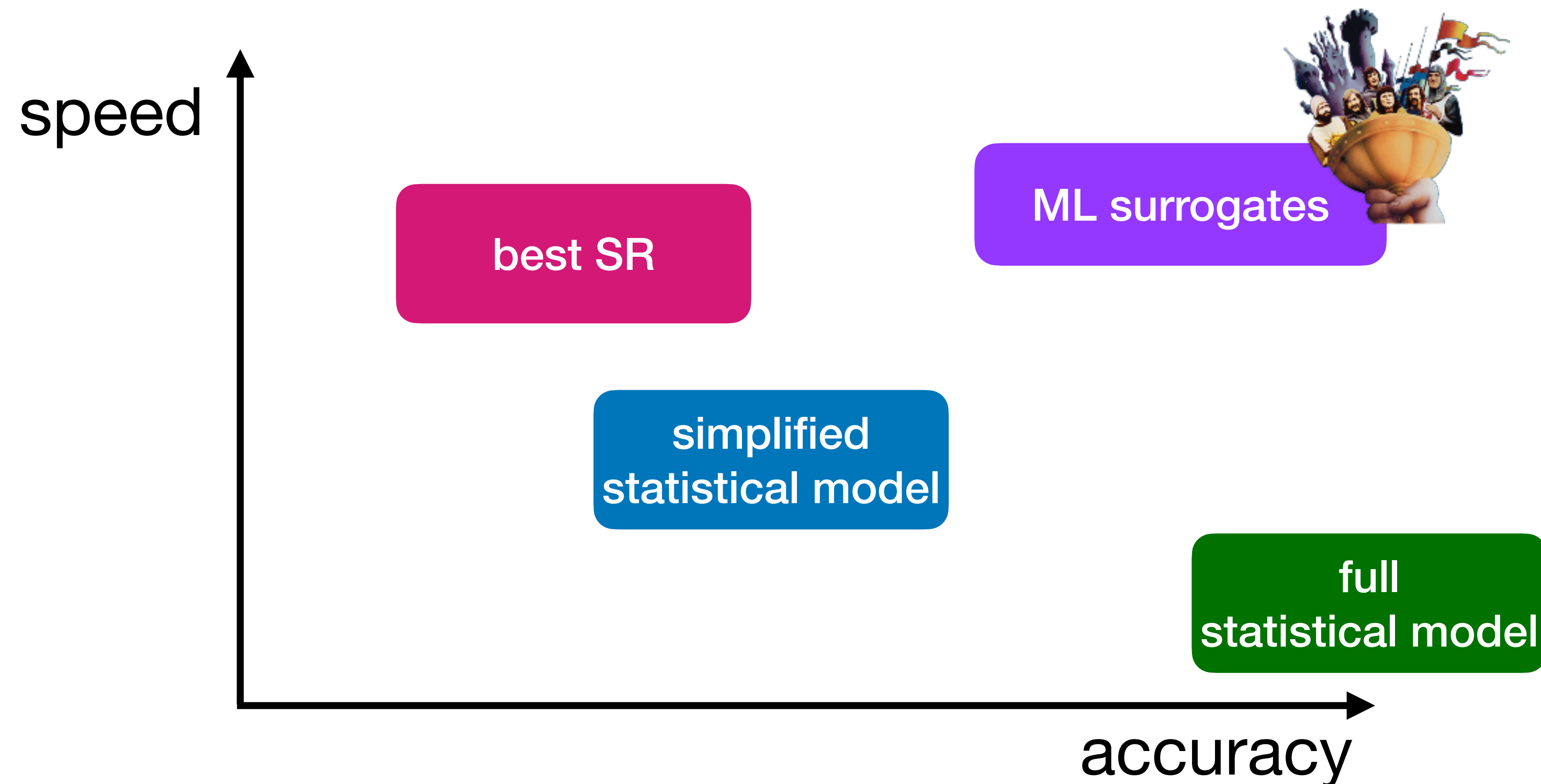


Machine learning enters here

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

The goal

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





Lublin Castle, Poland



The approach

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

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



The approach

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

Diagram annotations:

- Three pink arrows point from the text "varying signal in each bin independently" to the parameters $\mu_1, \dots, \mu_N$ in the numerator's likelihood function.
- A green arrow points from the text "profiling" to the profile likelihood $\hat{\hat{\theta}}(\mu_1, \dots, \mu_N)$ in the numerator.
- Two blue arrows point from the text "Four data sets are used:" to the "data" terms in the numerator and denominator likelihood functions.

Four data sets are used:

- 1) expected data
- 2) expected Asimov data
- 3) observed data
- 4) observed Asimov data



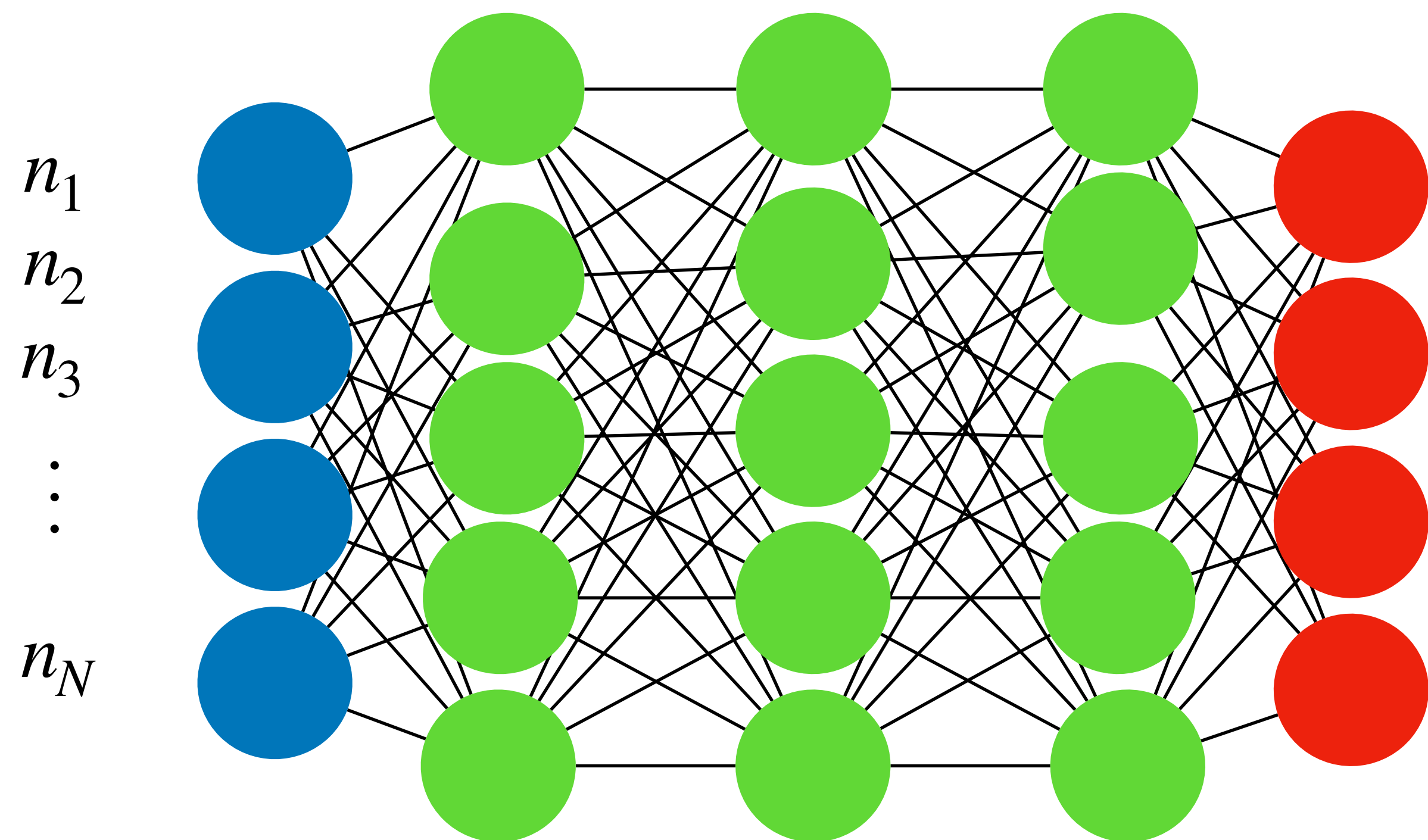
The approach

yields

regression

likelihoods

limits



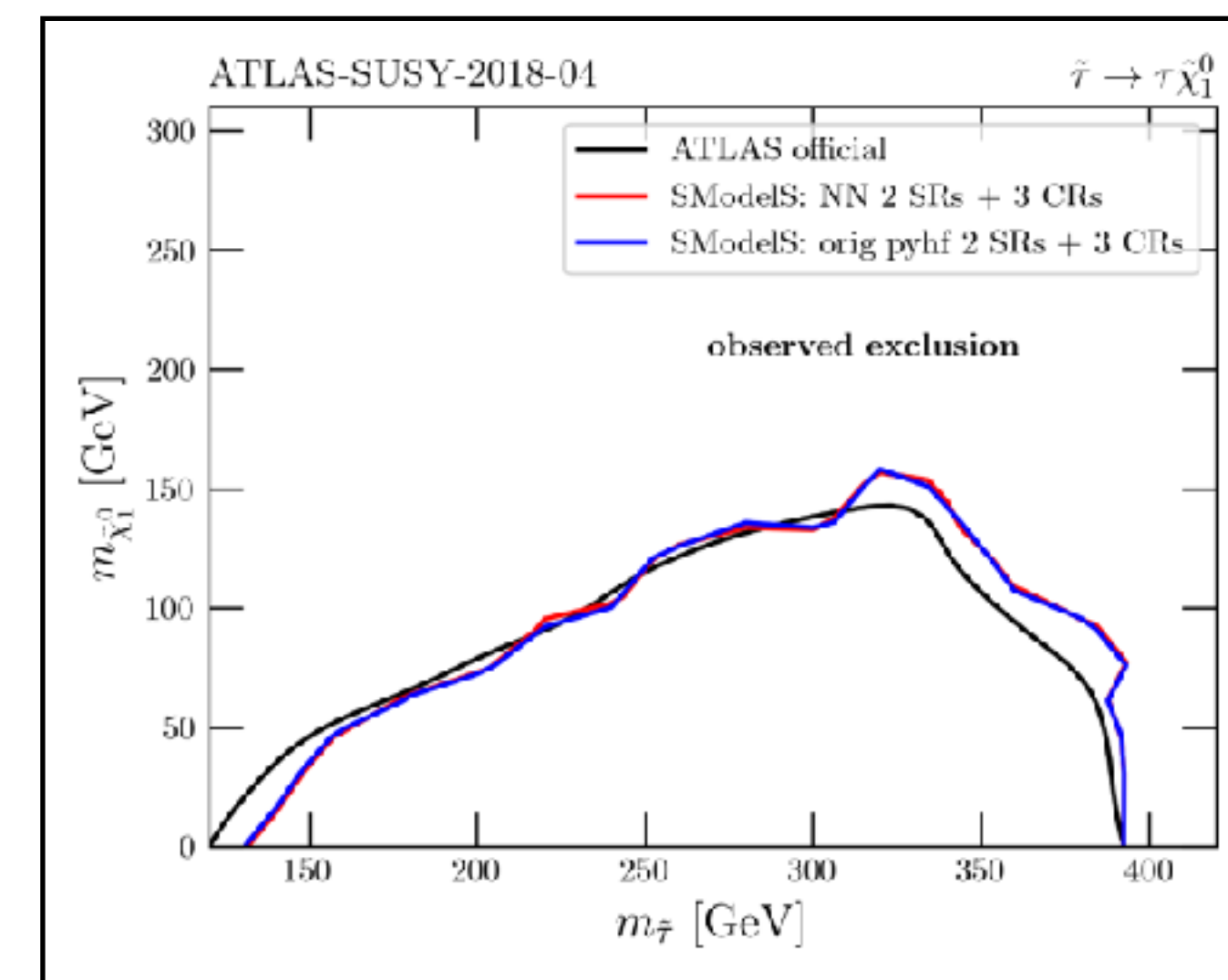
Fully connected network

$$\Delta(\mu_1, \dots, \mu_N)_{\text{exp}}$$

$$\Delta(\mu_1, \dots, \mu_N)_{\text{exp Asimov}}$$

$$\Delta(\mu_1, \dots, \mu_N)_{\text{obs}}$$

$$\Delta(\mu_1, \dots, \mu_N)_{\text{obs Asimov}}$$



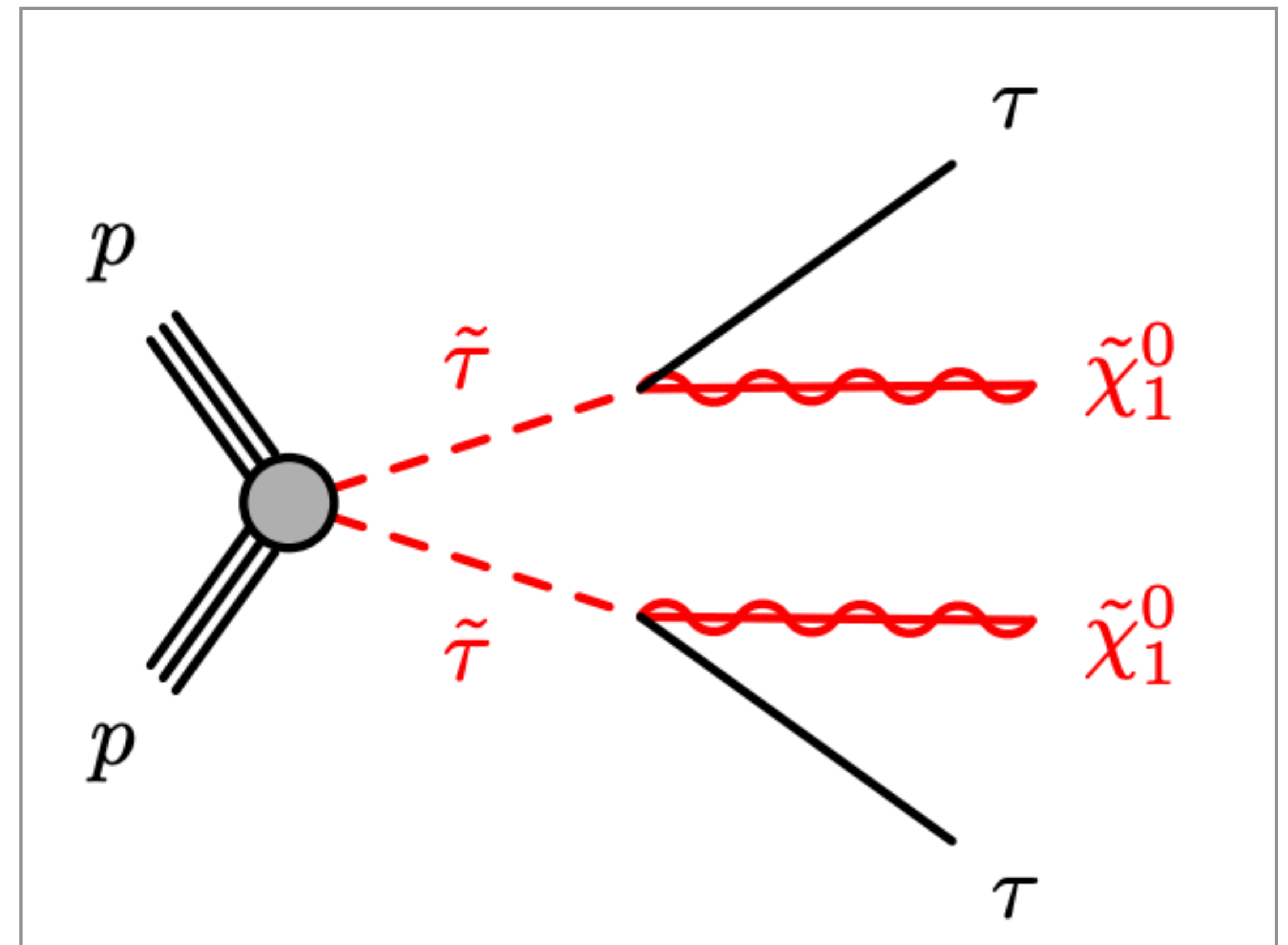


Moszna Castle, Poland

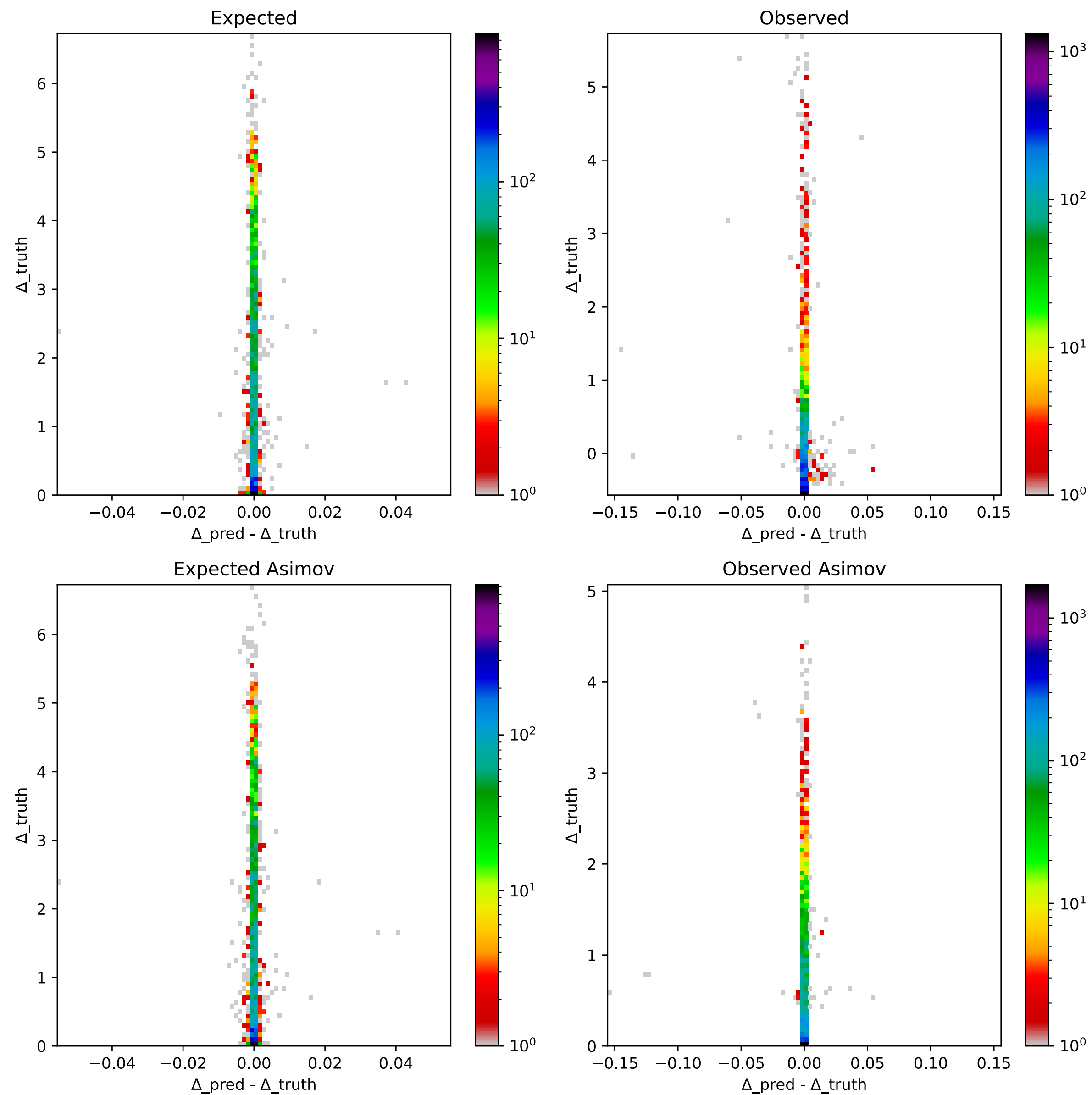
Search for direct stau production in events with two hadronic τ -leptons in $\sqrt{s}=13$ TeV pp collisions with the ATLAS detector

2 signal bins, 3 control bins

fit **without** the CRs

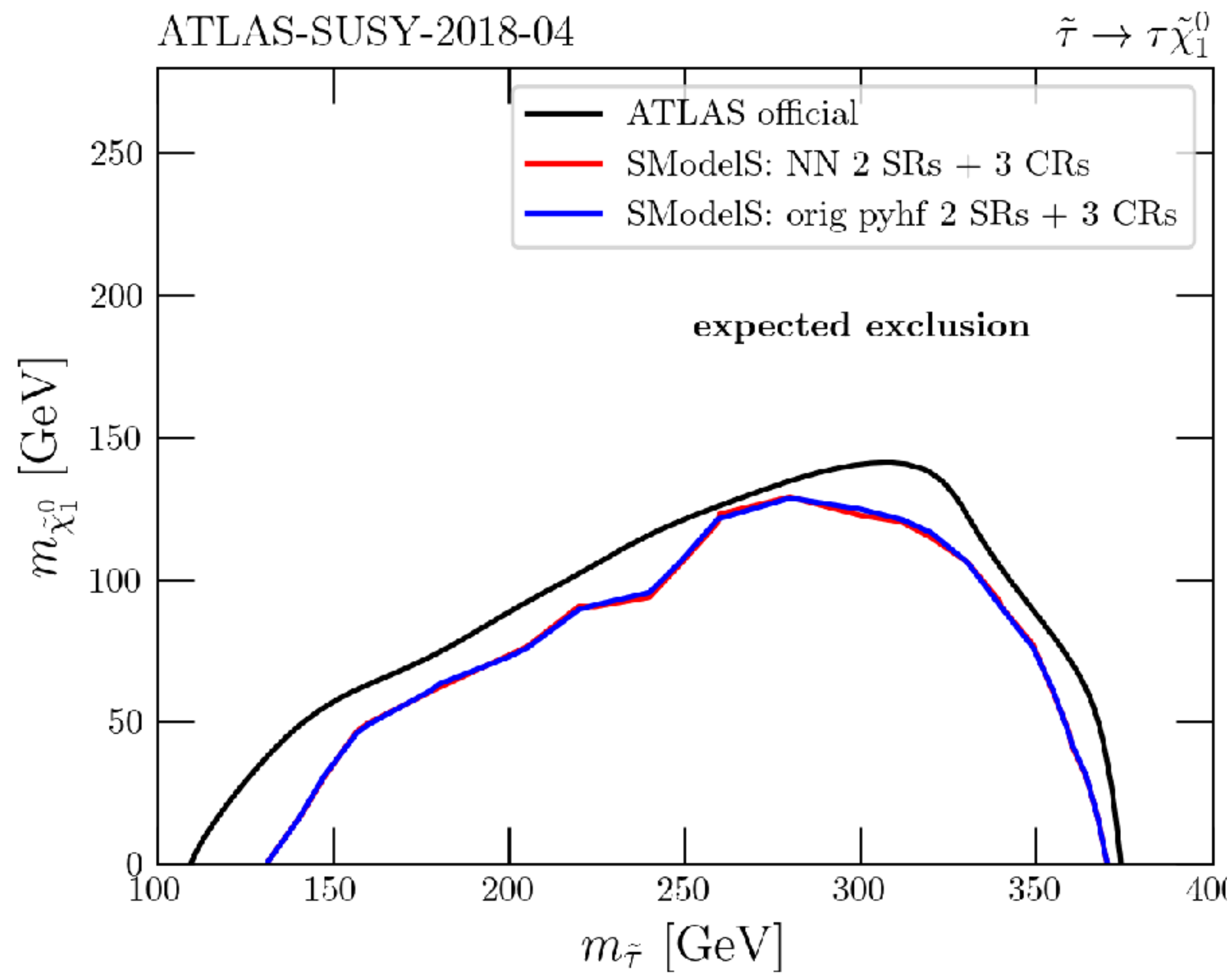


ATLAS-SUSY-2018-04 (1911.06660)

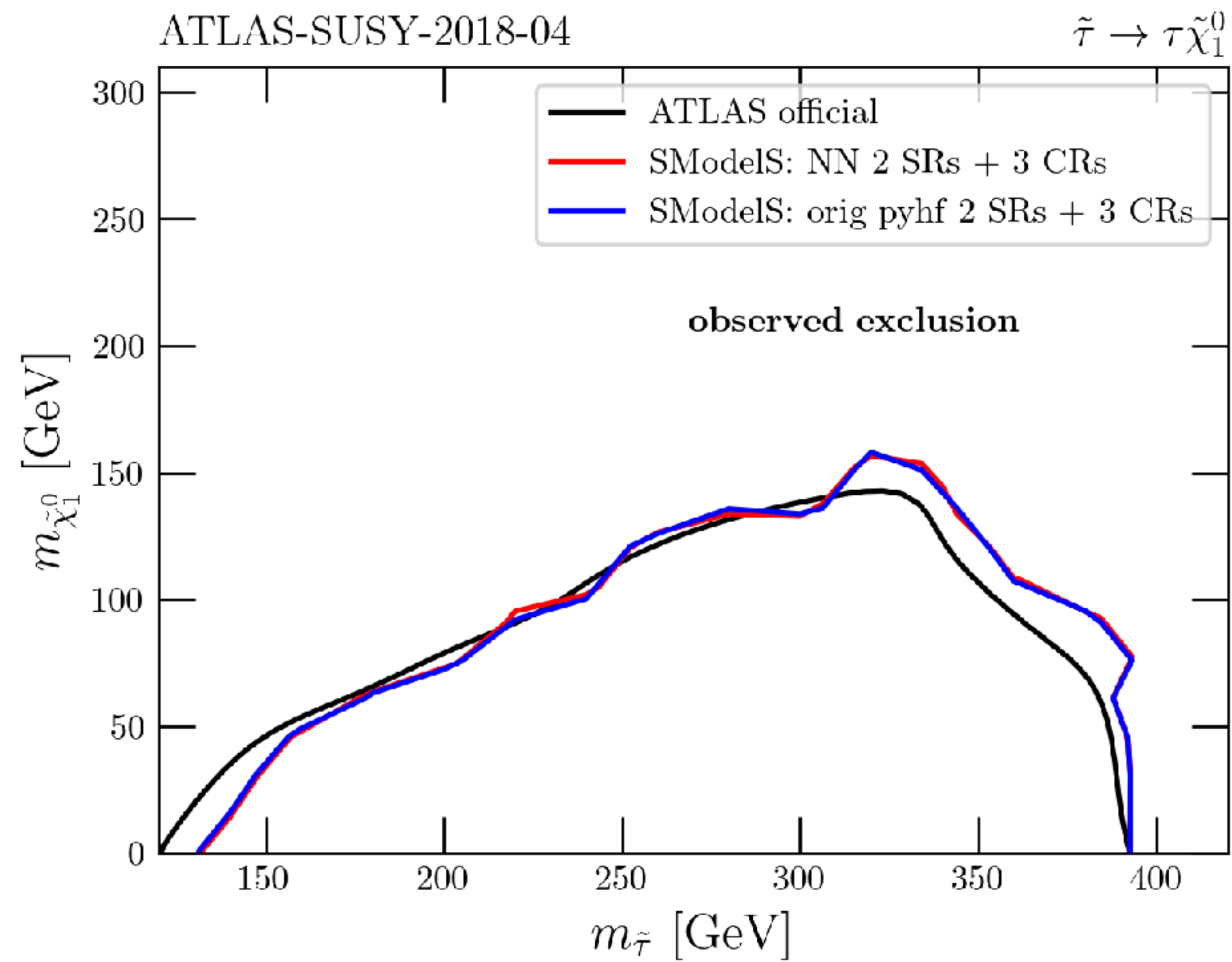


ATLAS-SUSY-2018-04 (1911.06660)

expected

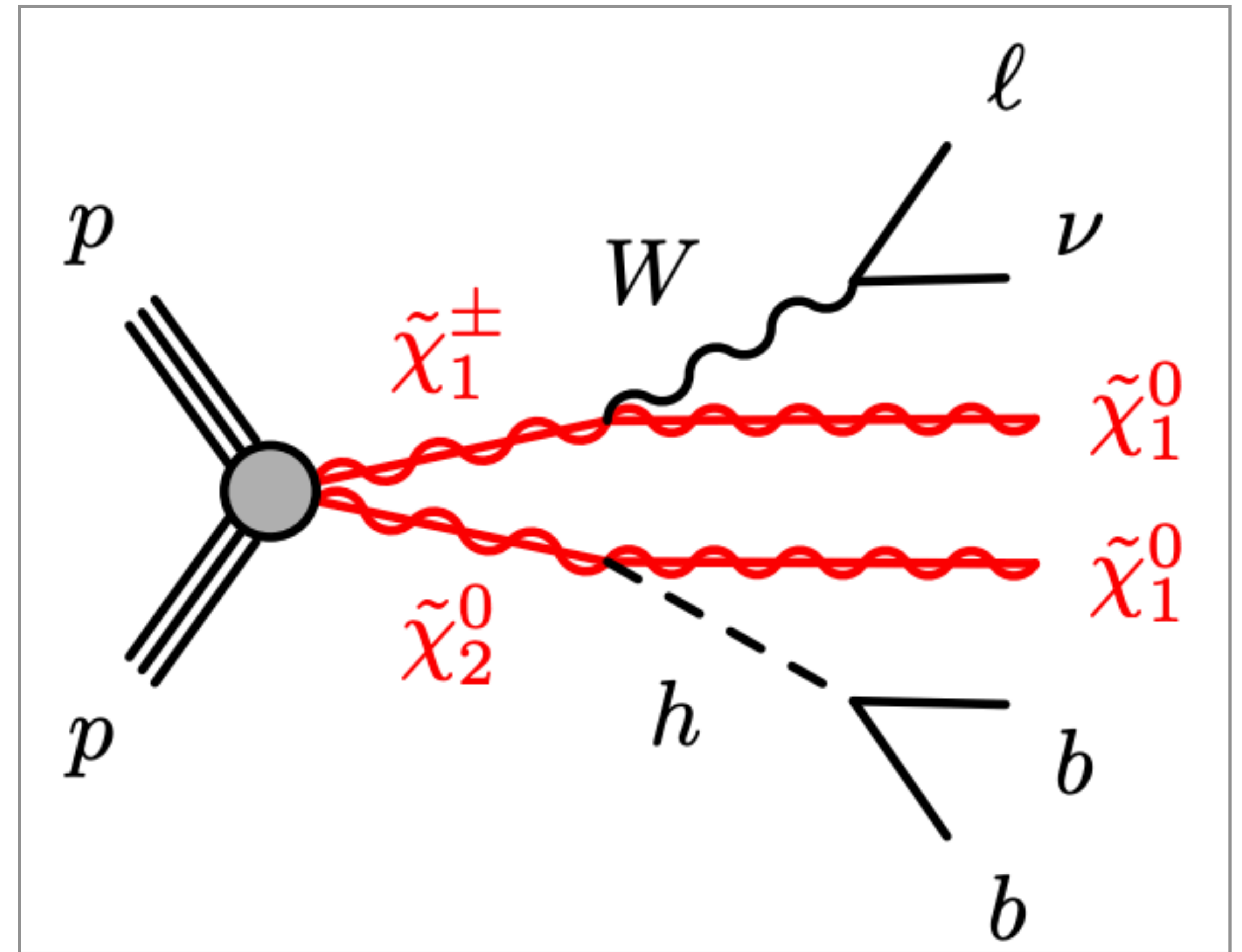


observed



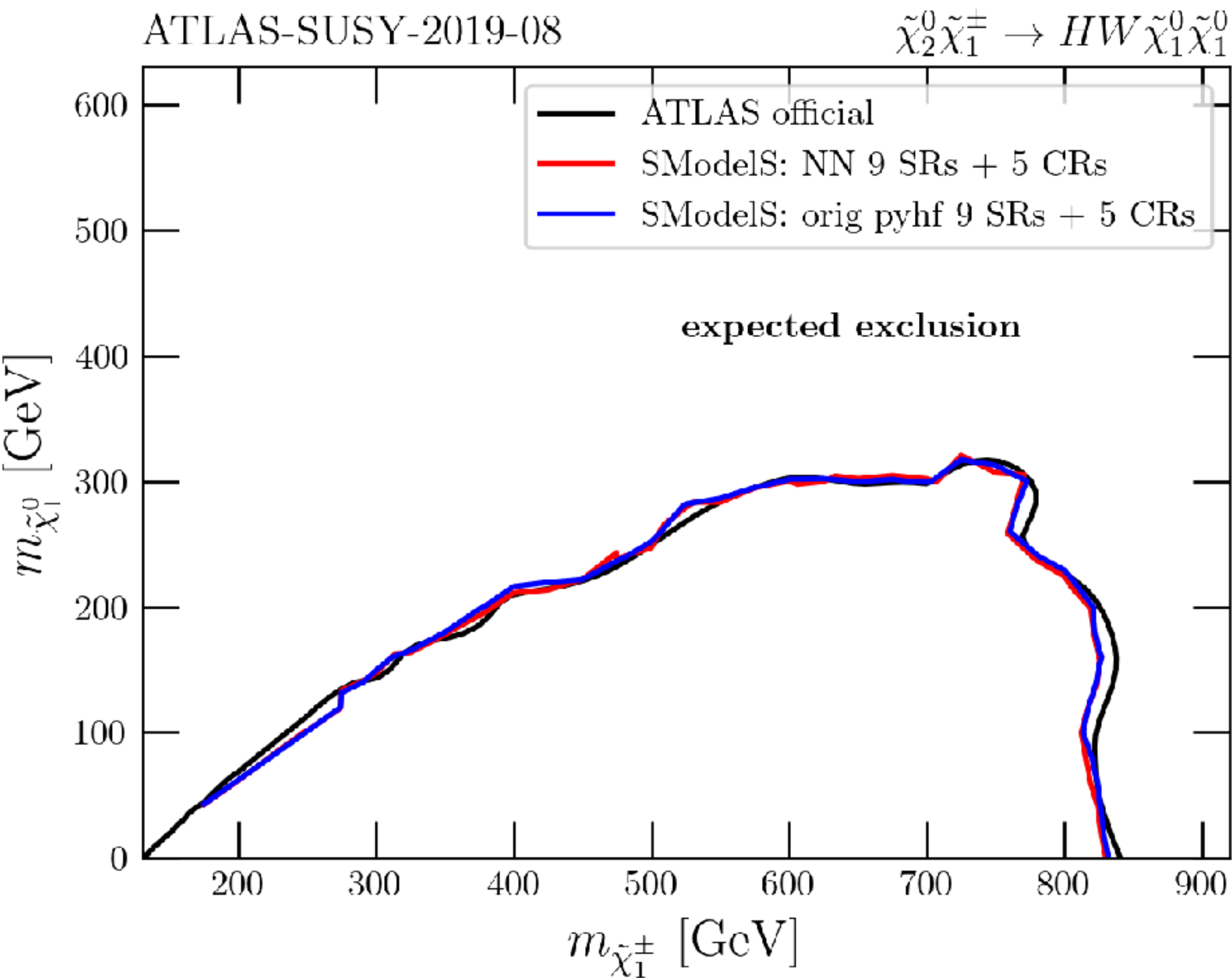
Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector

9 signal bins, 5 control bins

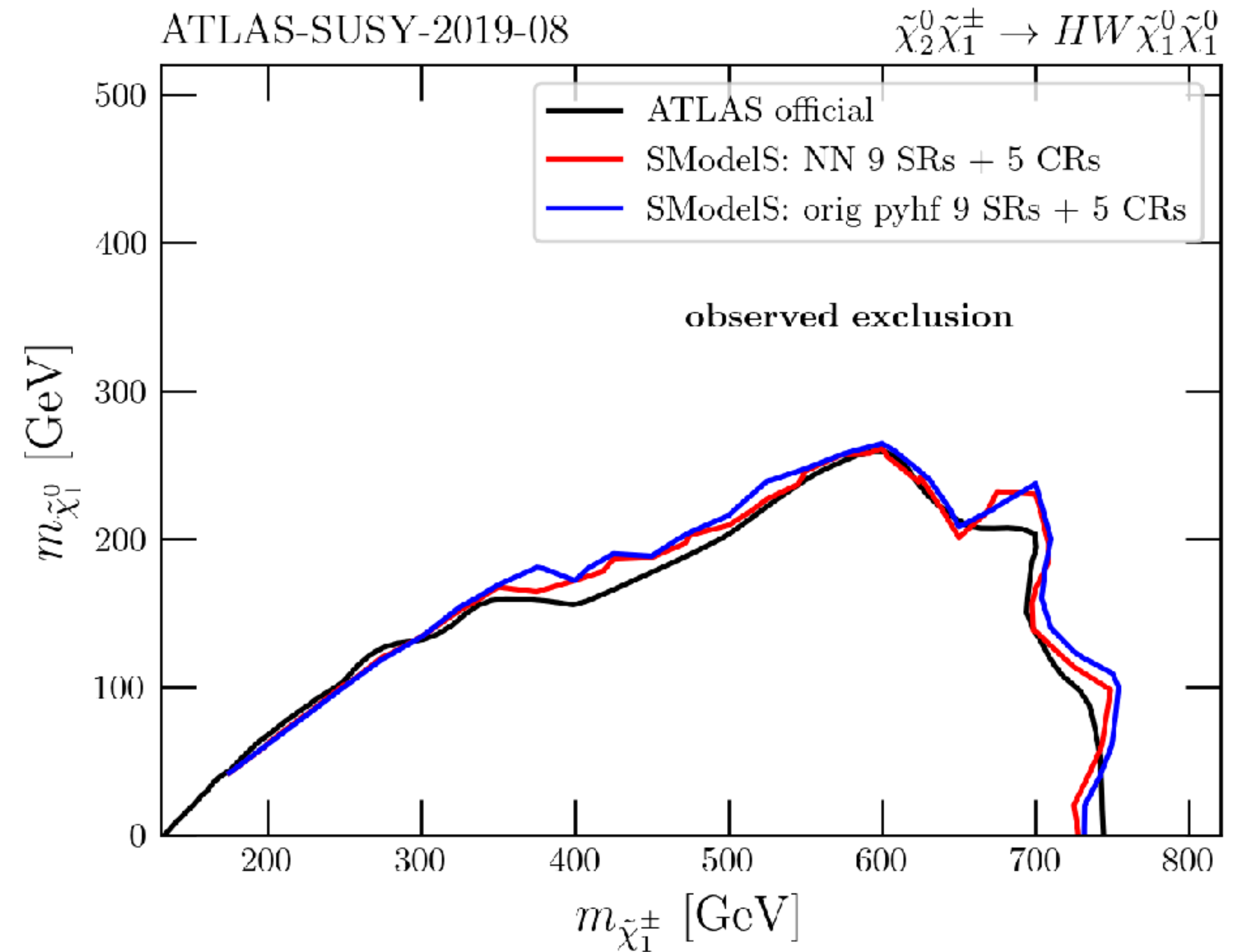


ATLAS-SUSY-2019-08 (1909.09926)

expected

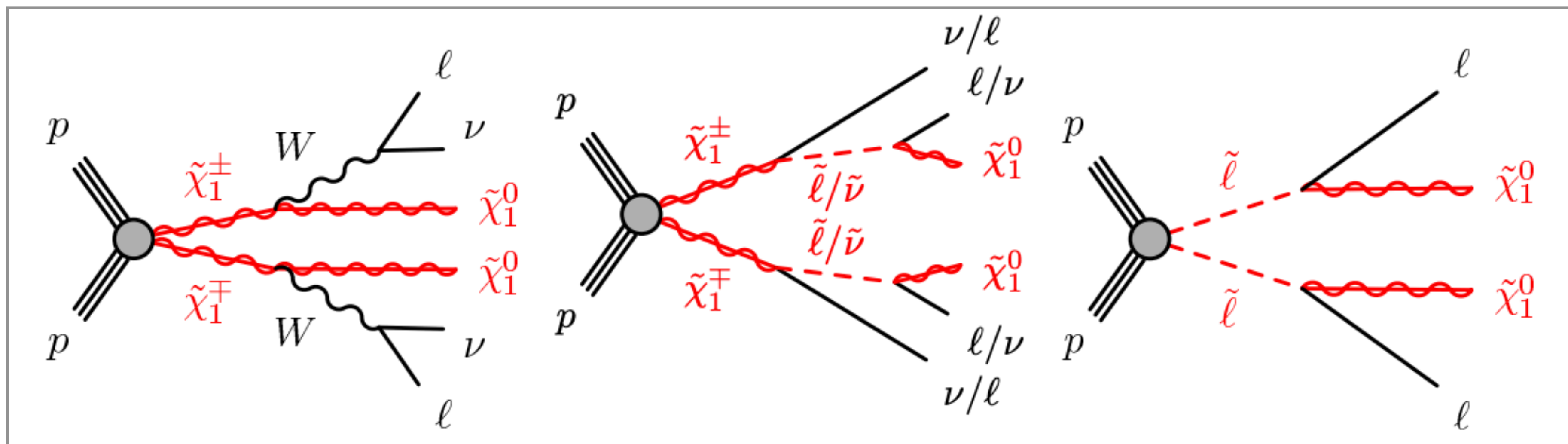


observed



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

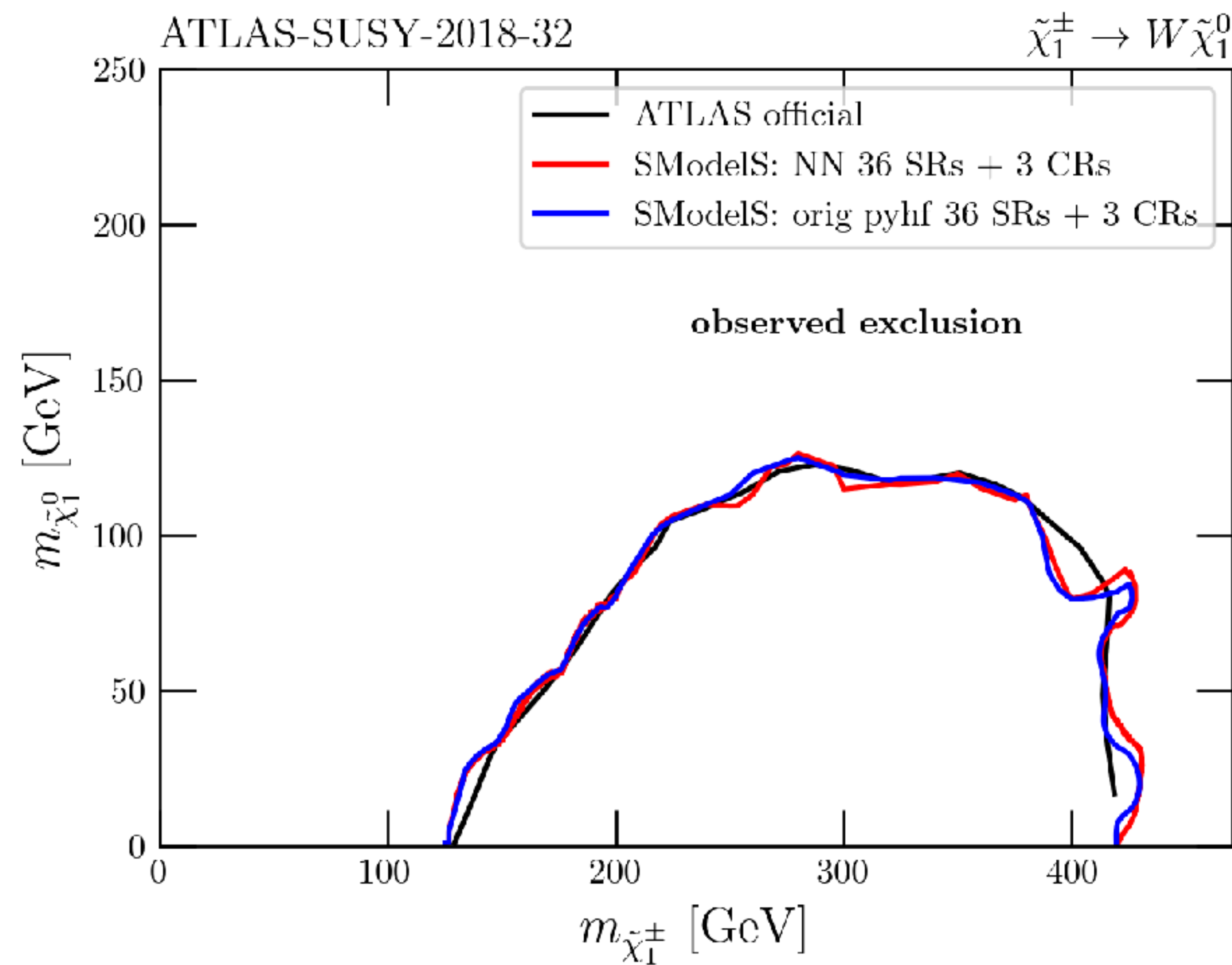
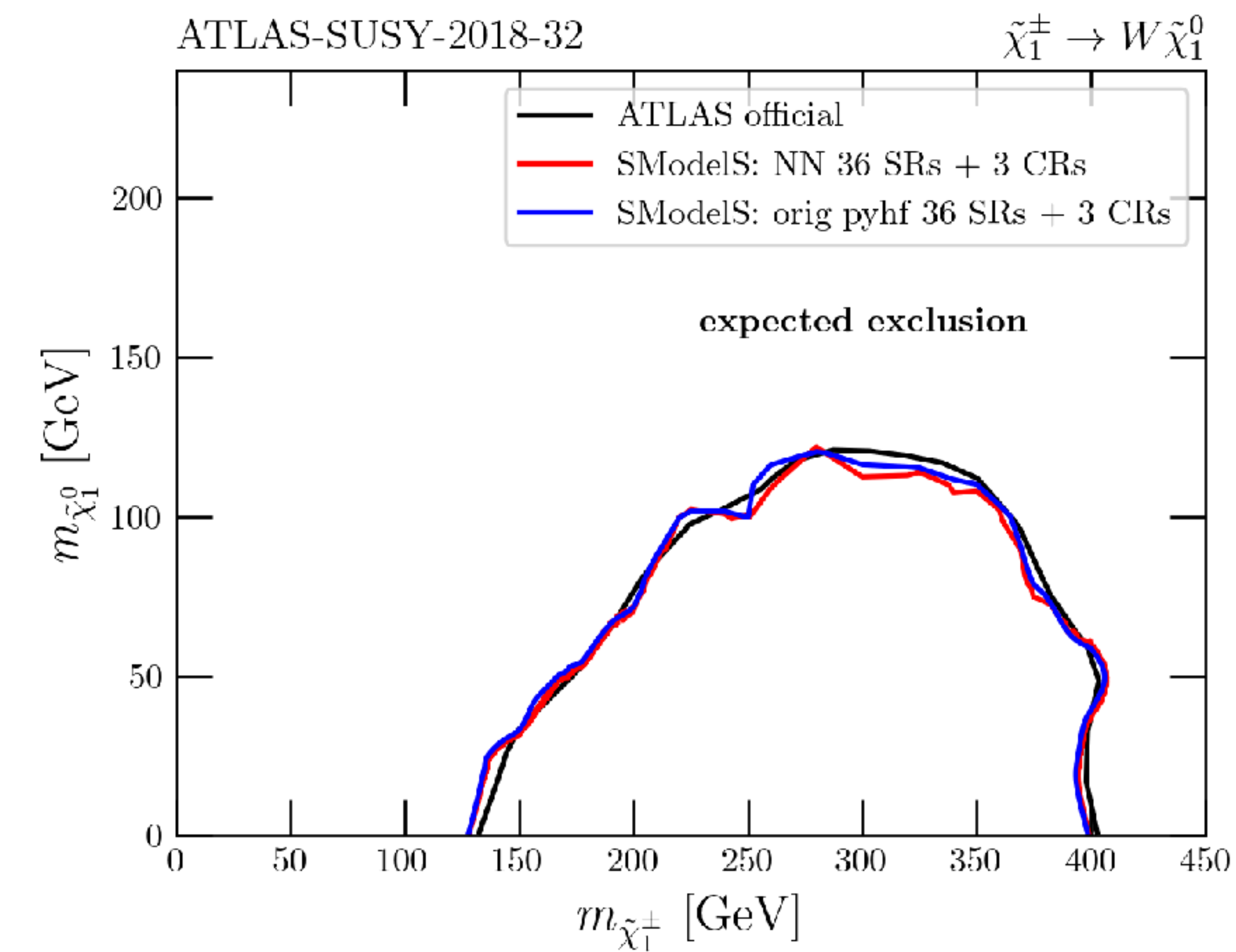
36 SRs and 3 CRs



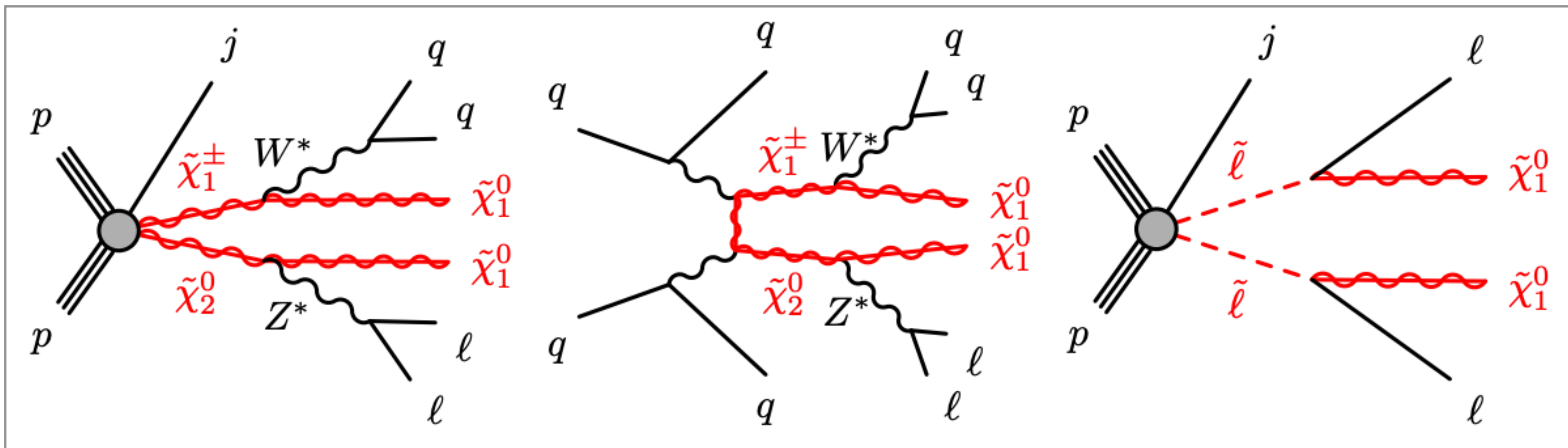
ATLAS-SUSY-2018-32 (1908.08215)

expected

observed

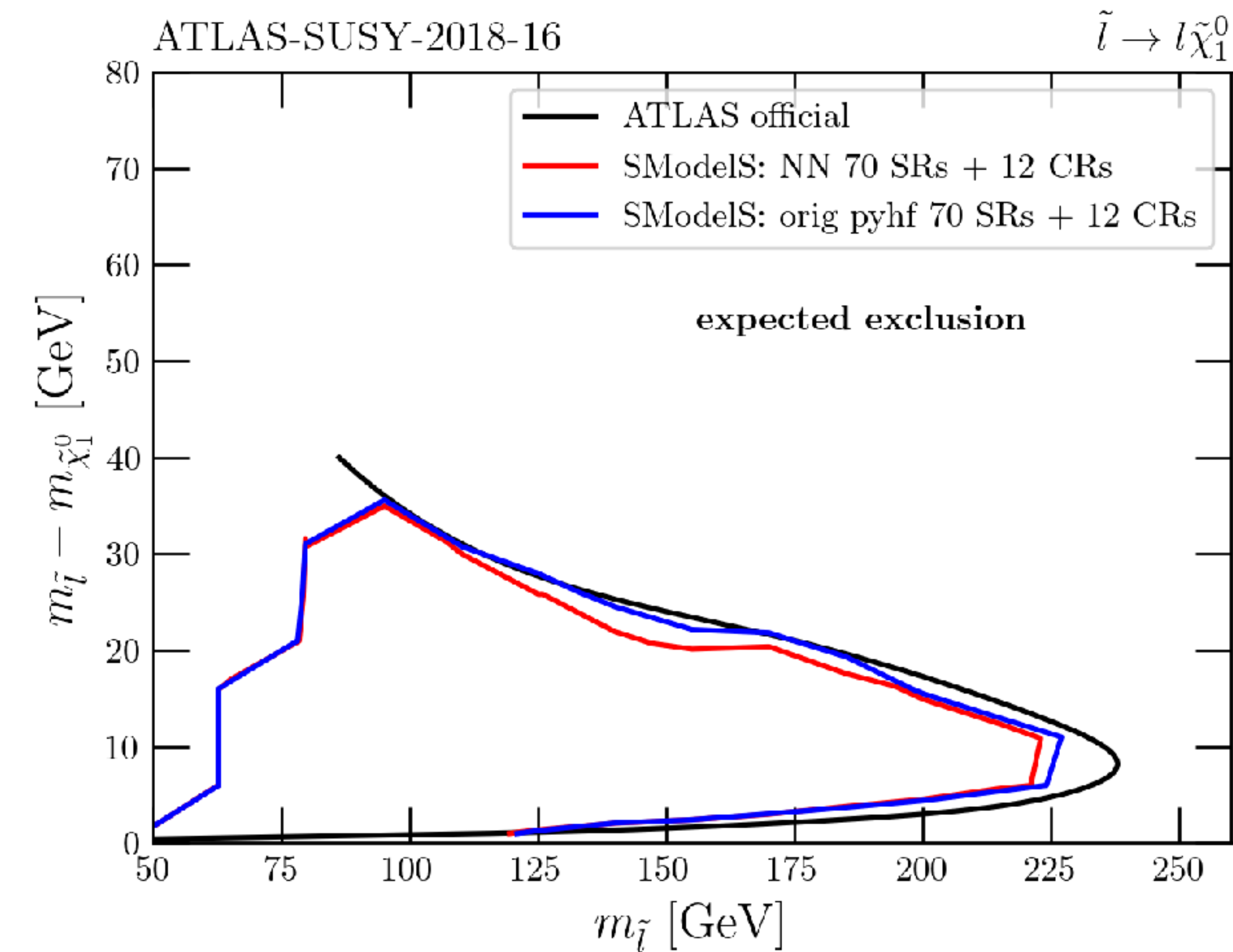


- Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s}=13$ TeV pp collisions with the ATLAS detector
- 2 subanalyses: sleptons (32+6) and EWKinos (38+6)

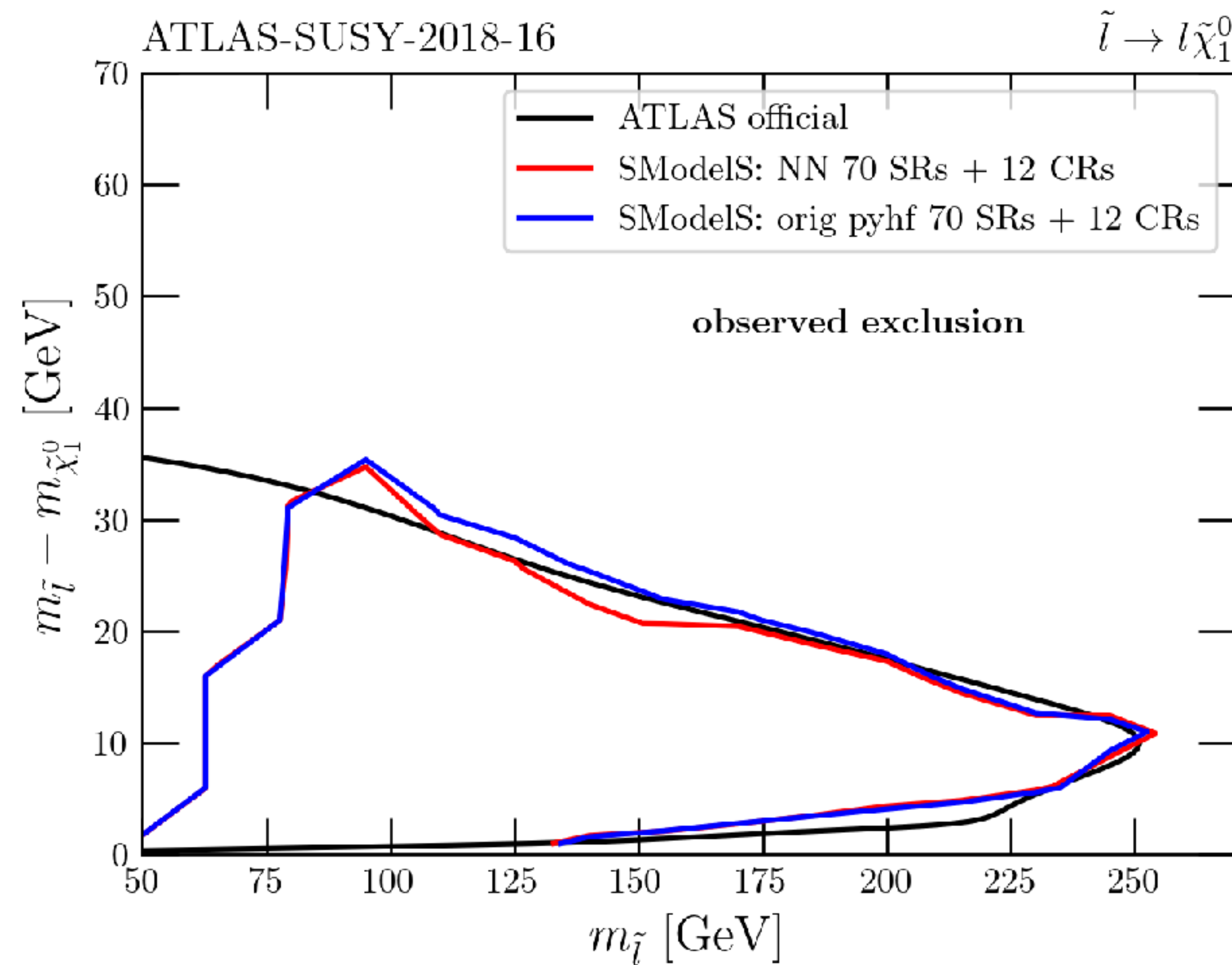


ATLAS-SUSY-2018-16 (1911.12606)

expected (sleptons)

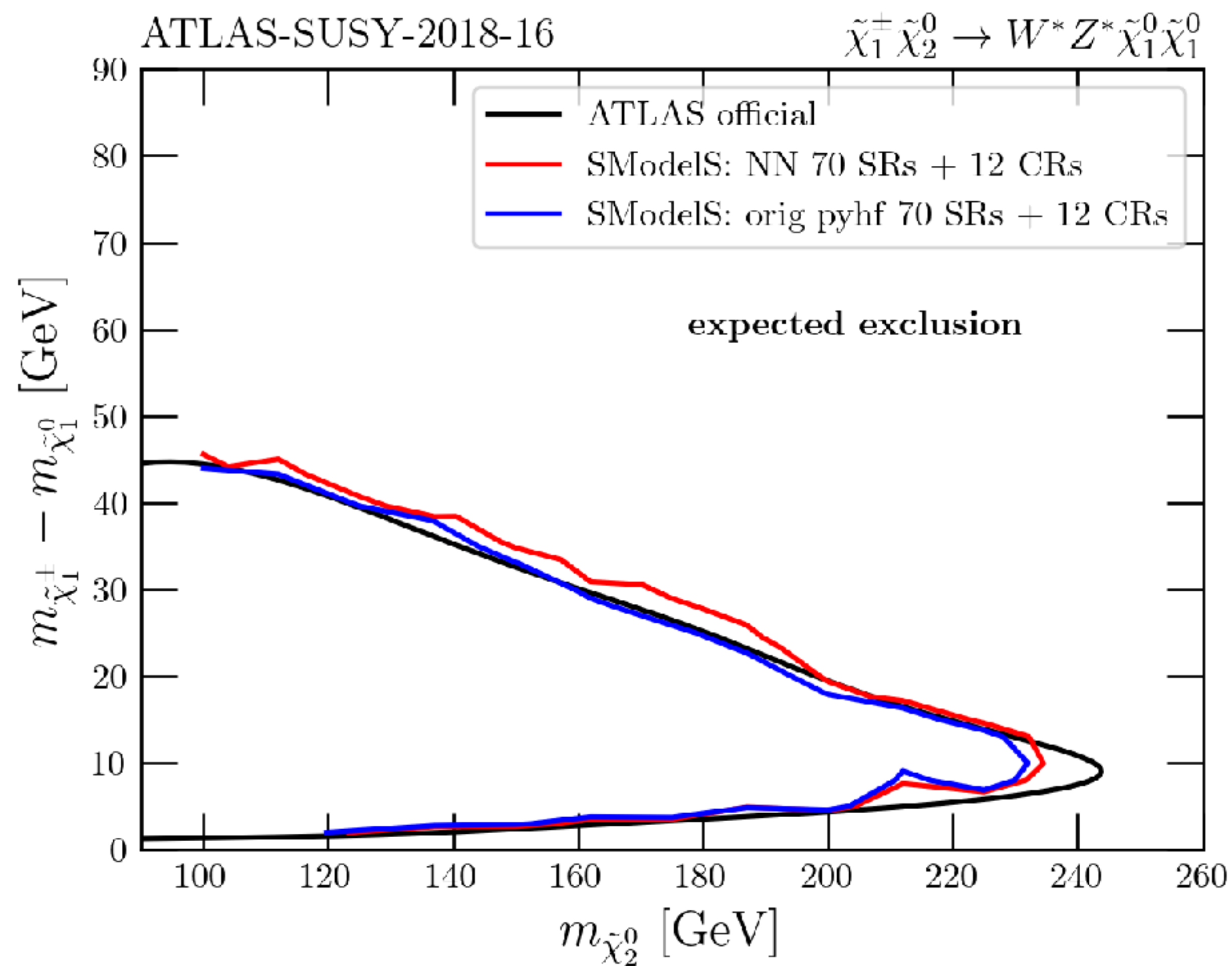


observed (sleptons)

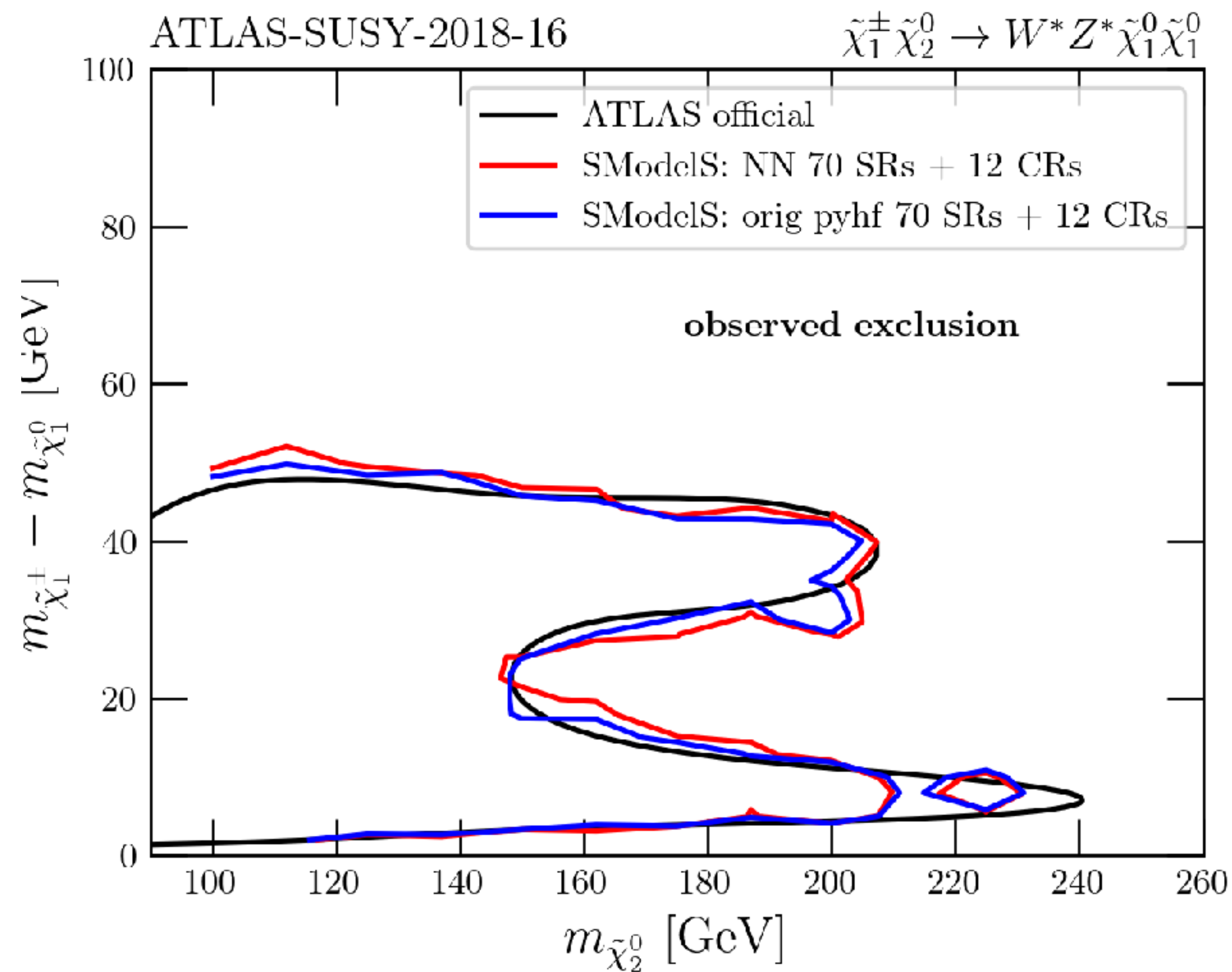


ATLAS-SUSY-2018-16 (1911.12606)

expected (EWKinosh)



observed (EWKinosh)



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

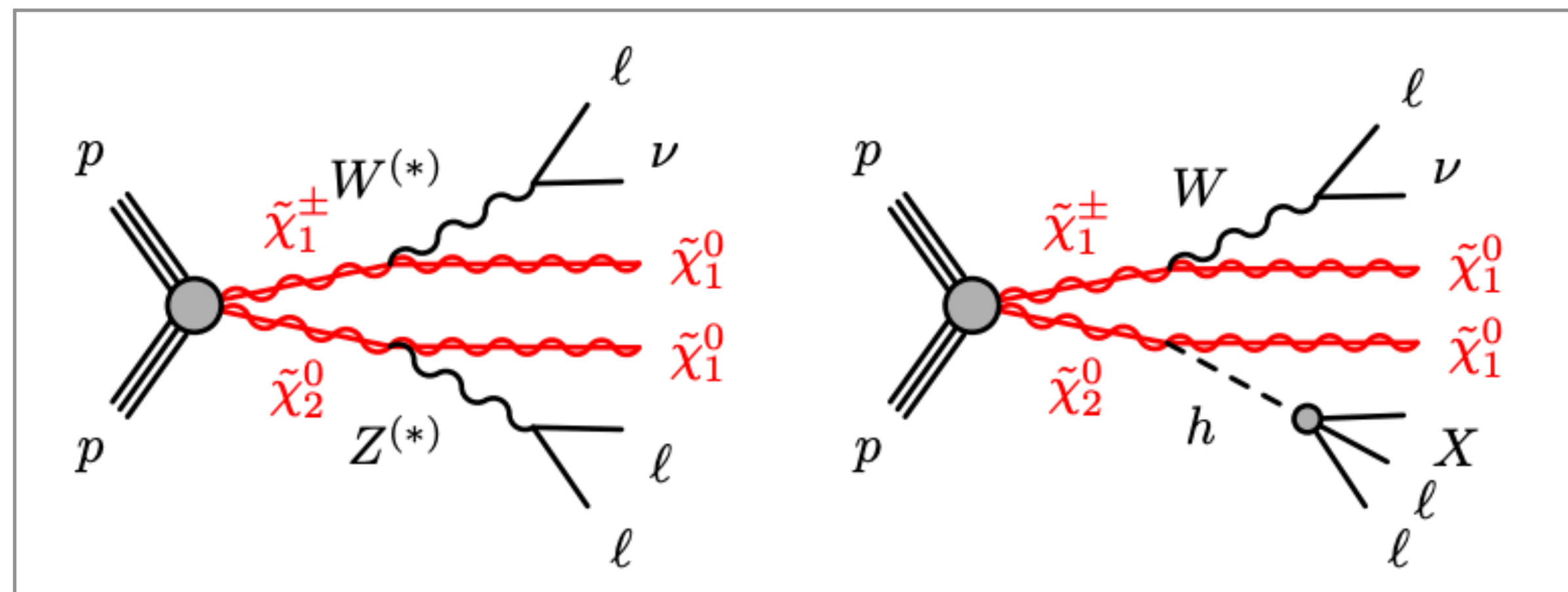
4 subanalyses:

onshell (W/Z when decaying) **wino/bino** (22+2) ❌

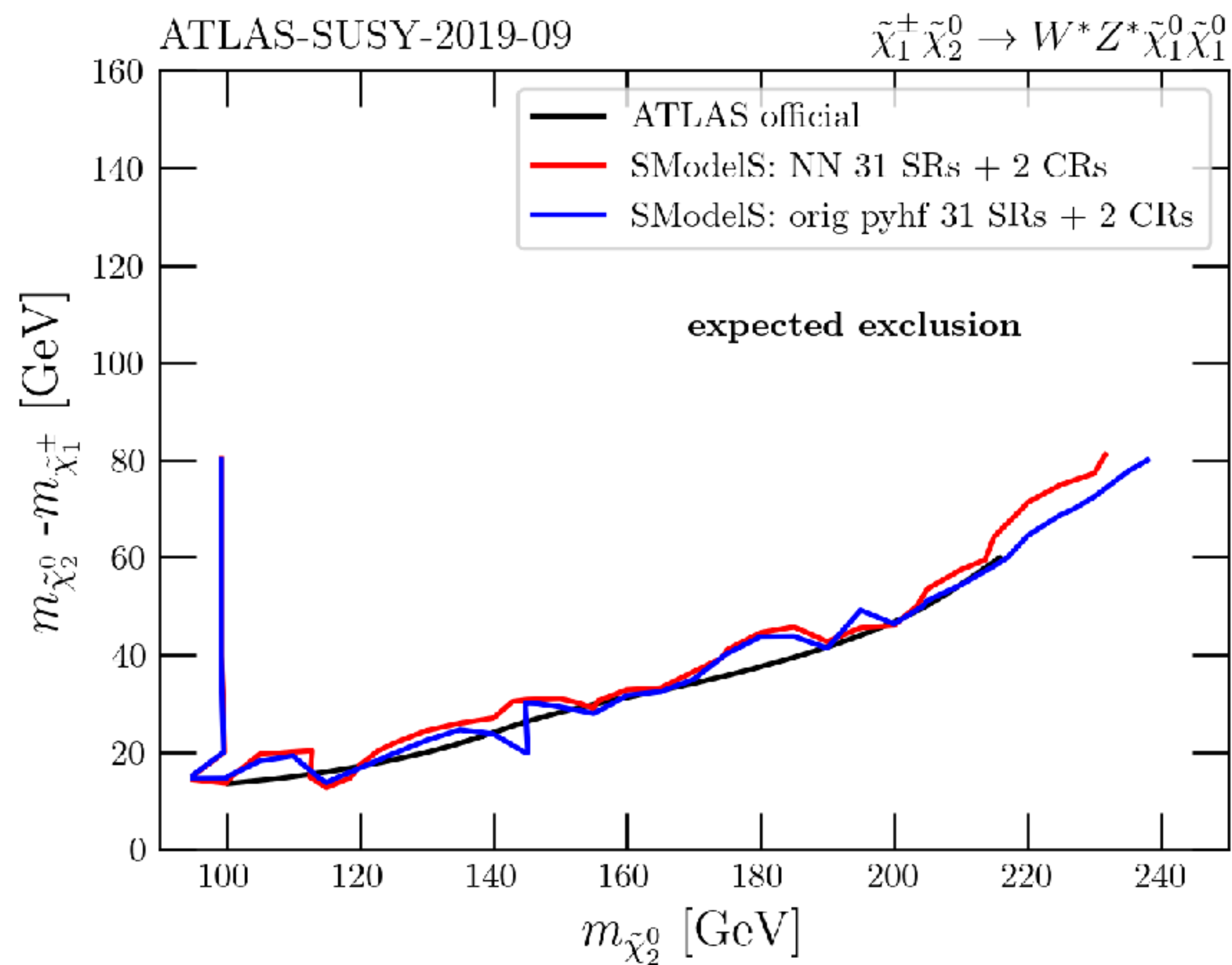
offshell wino/bino minus (sign of the product of two neutralino mass eigenstates) (32+2) ✔️

offshell wino/bino plus (32+2) ✔️

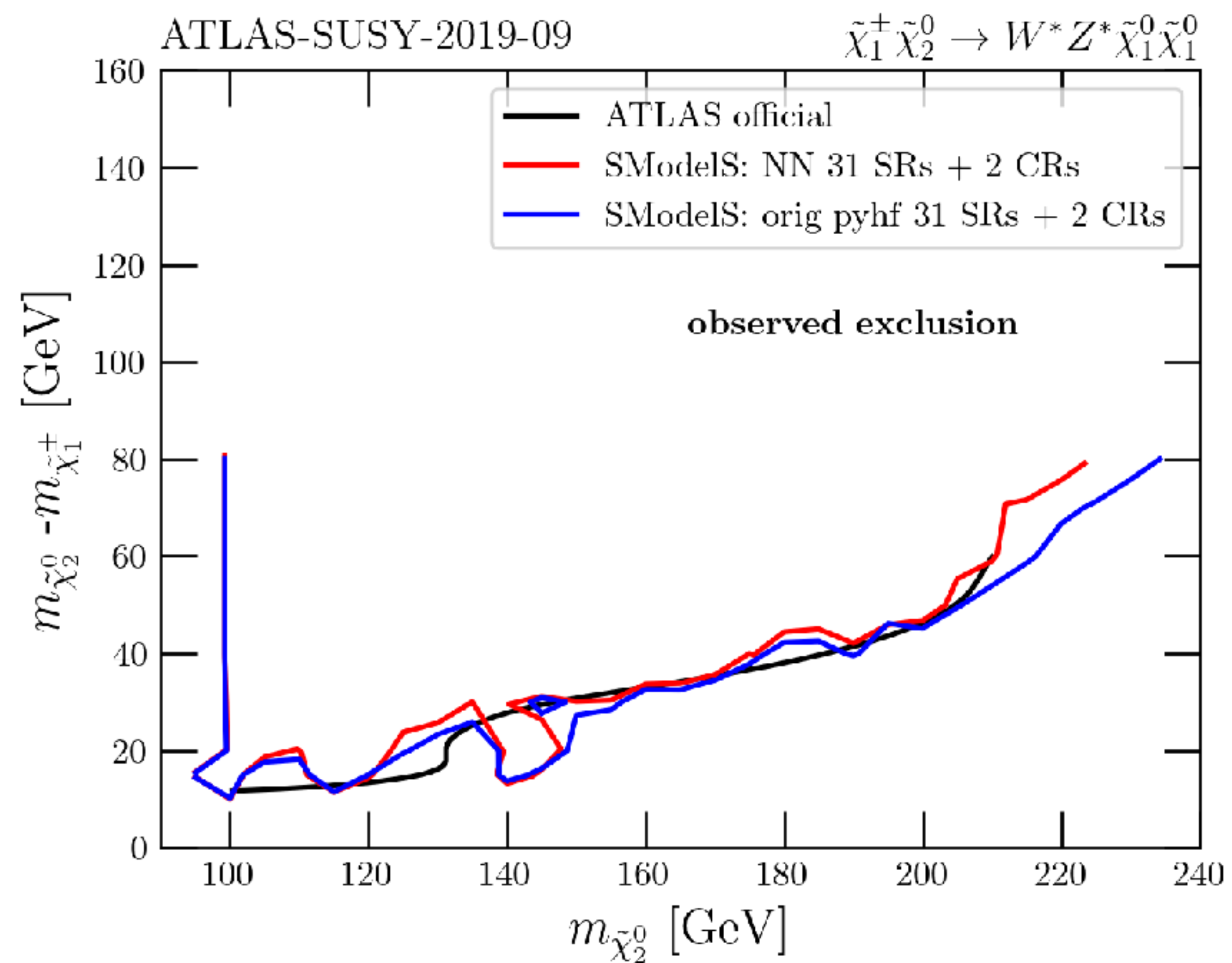
offshell higgsino (32+2) ✔️



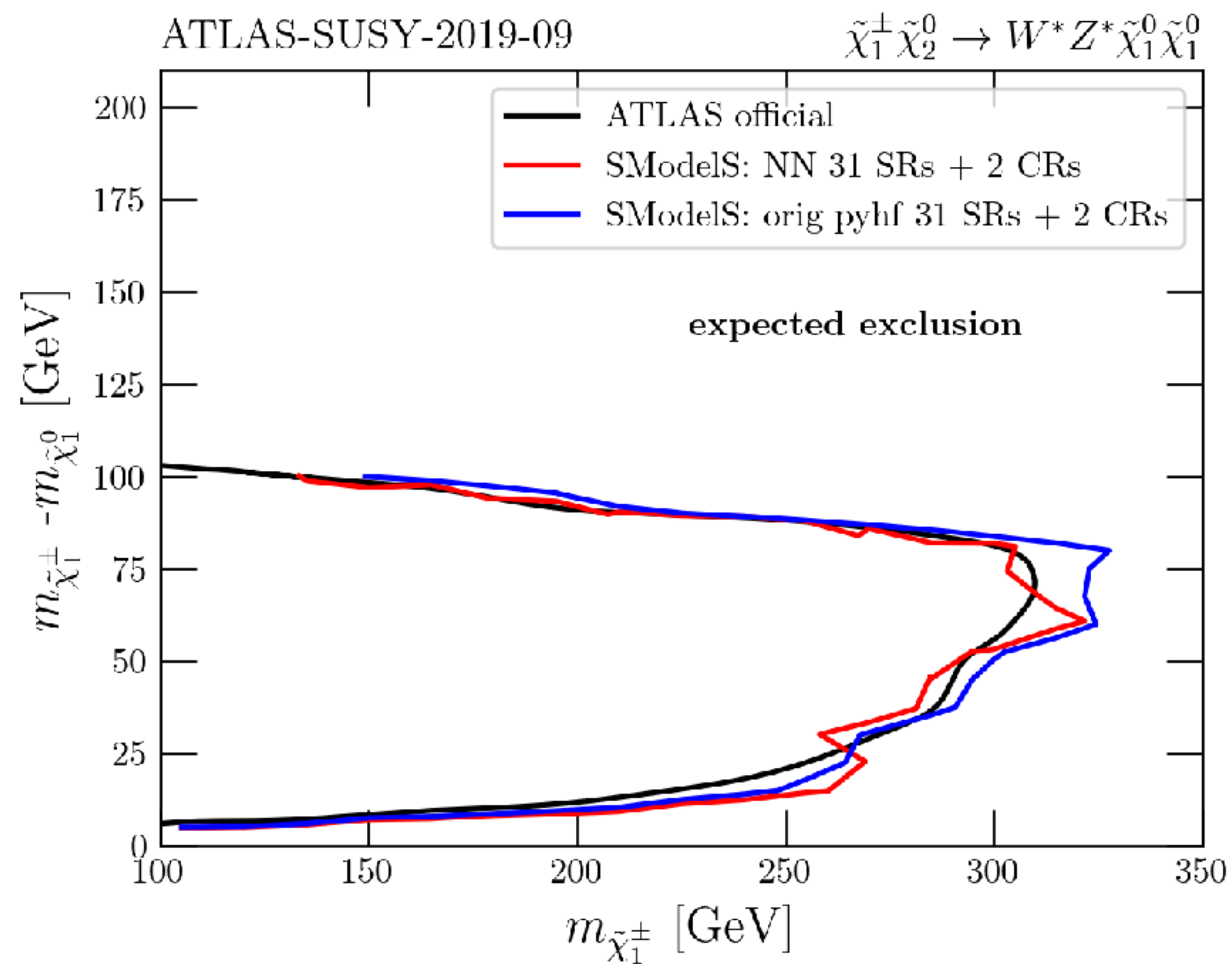
expected (offshell Higgsino)



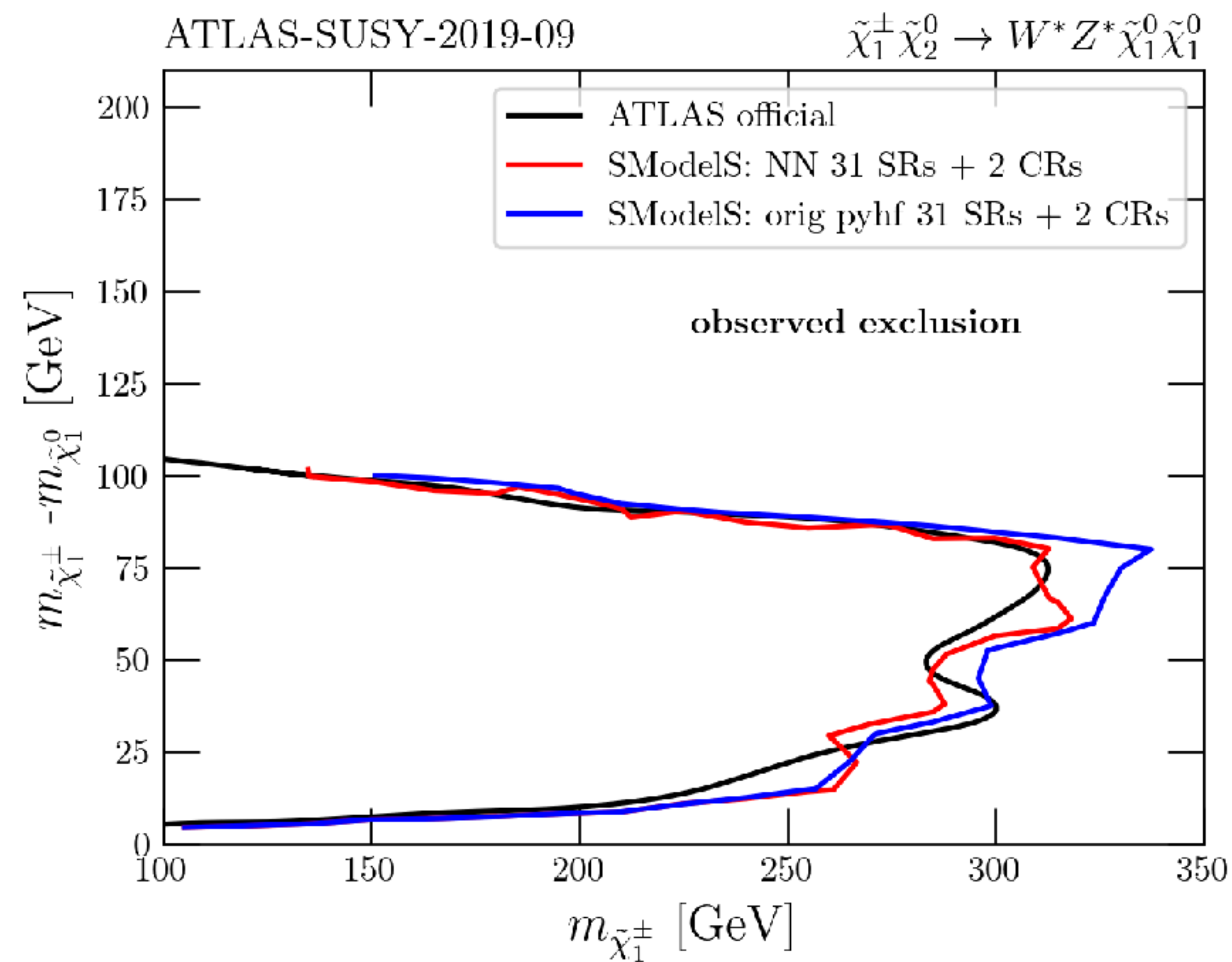
observed (offshell Higgsino)



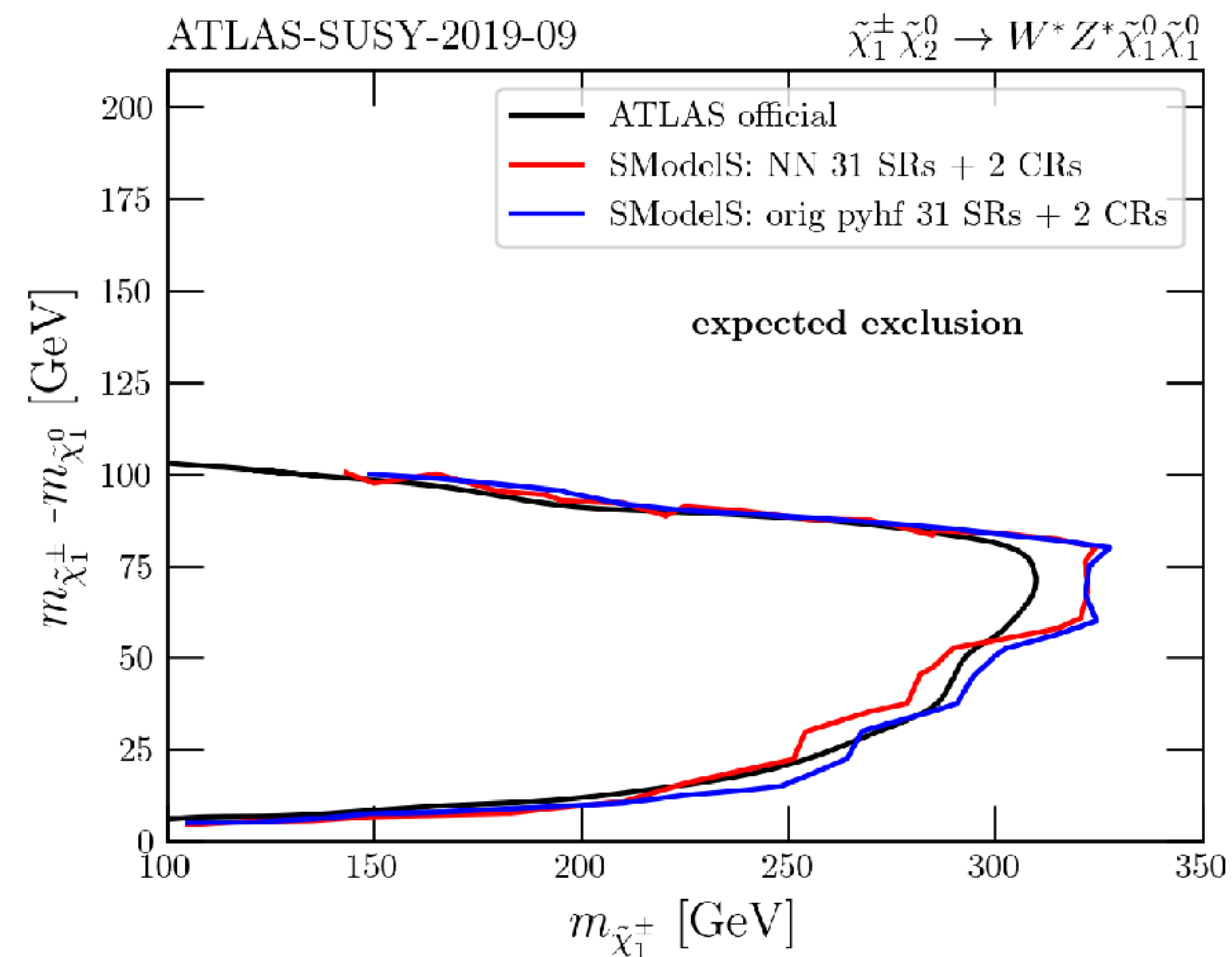
expected (offshell wino/bino -)



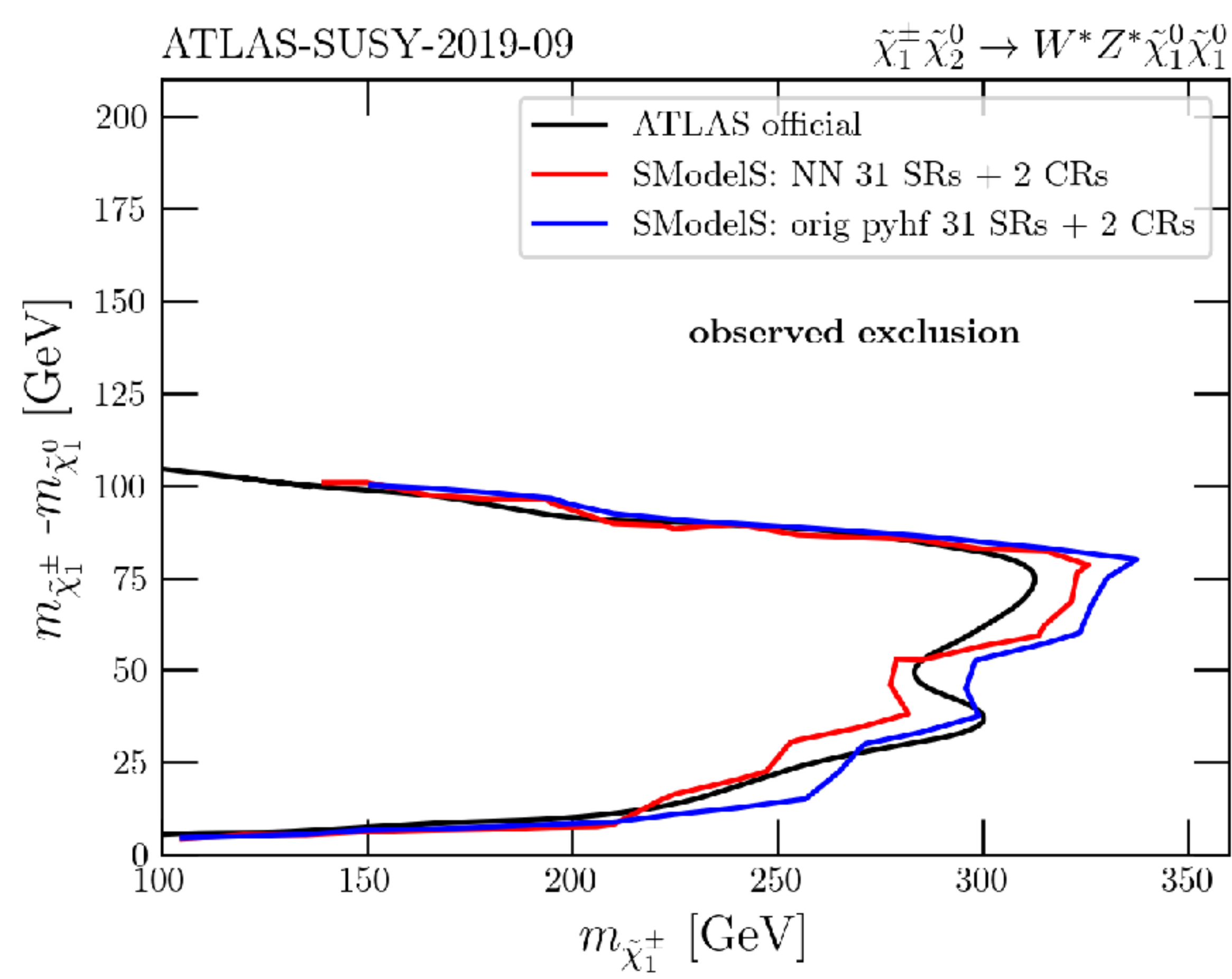
observed (offshell wino/bino -)



expected (offshell wino/bino +)



observed (offshell wino/bino +)

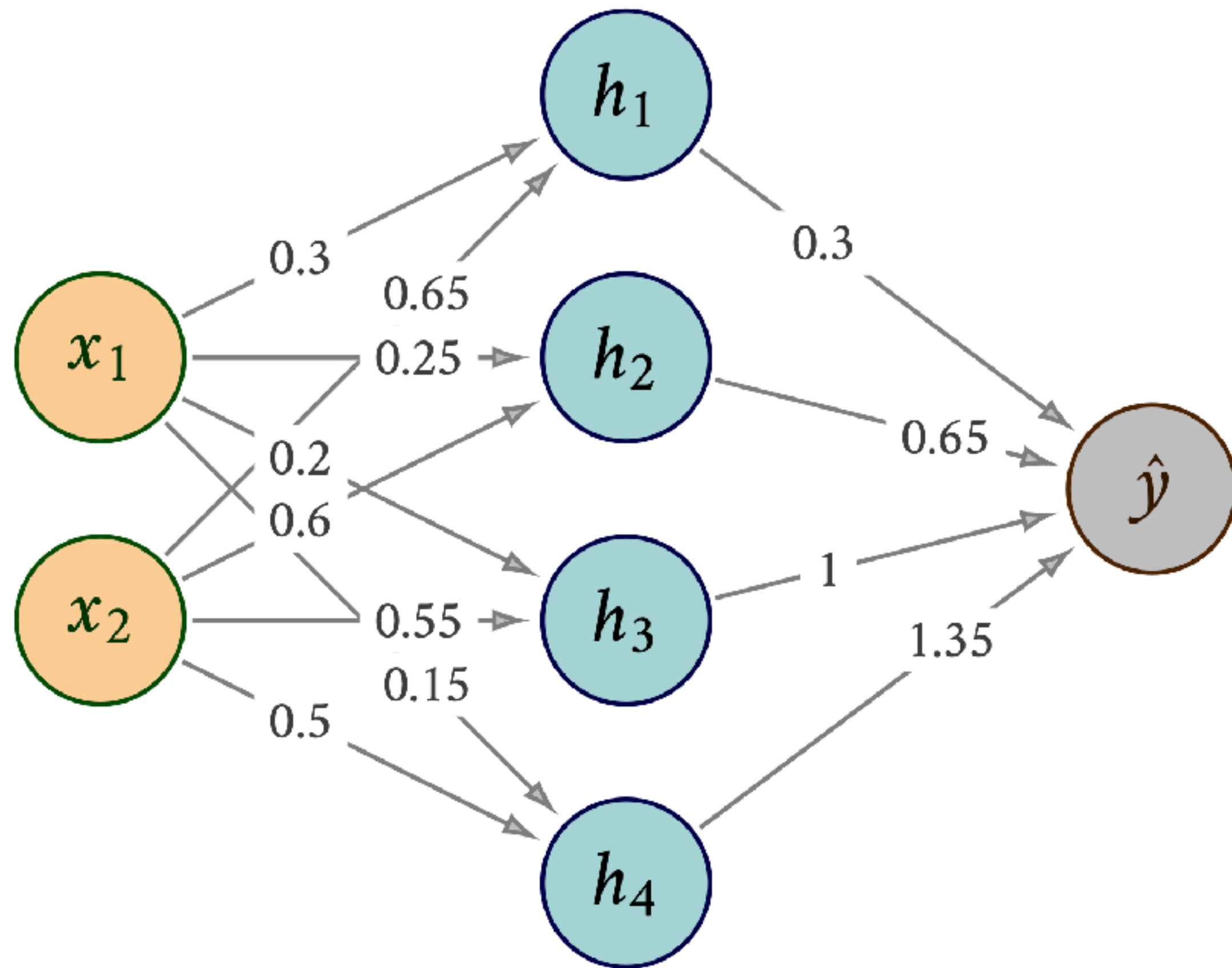




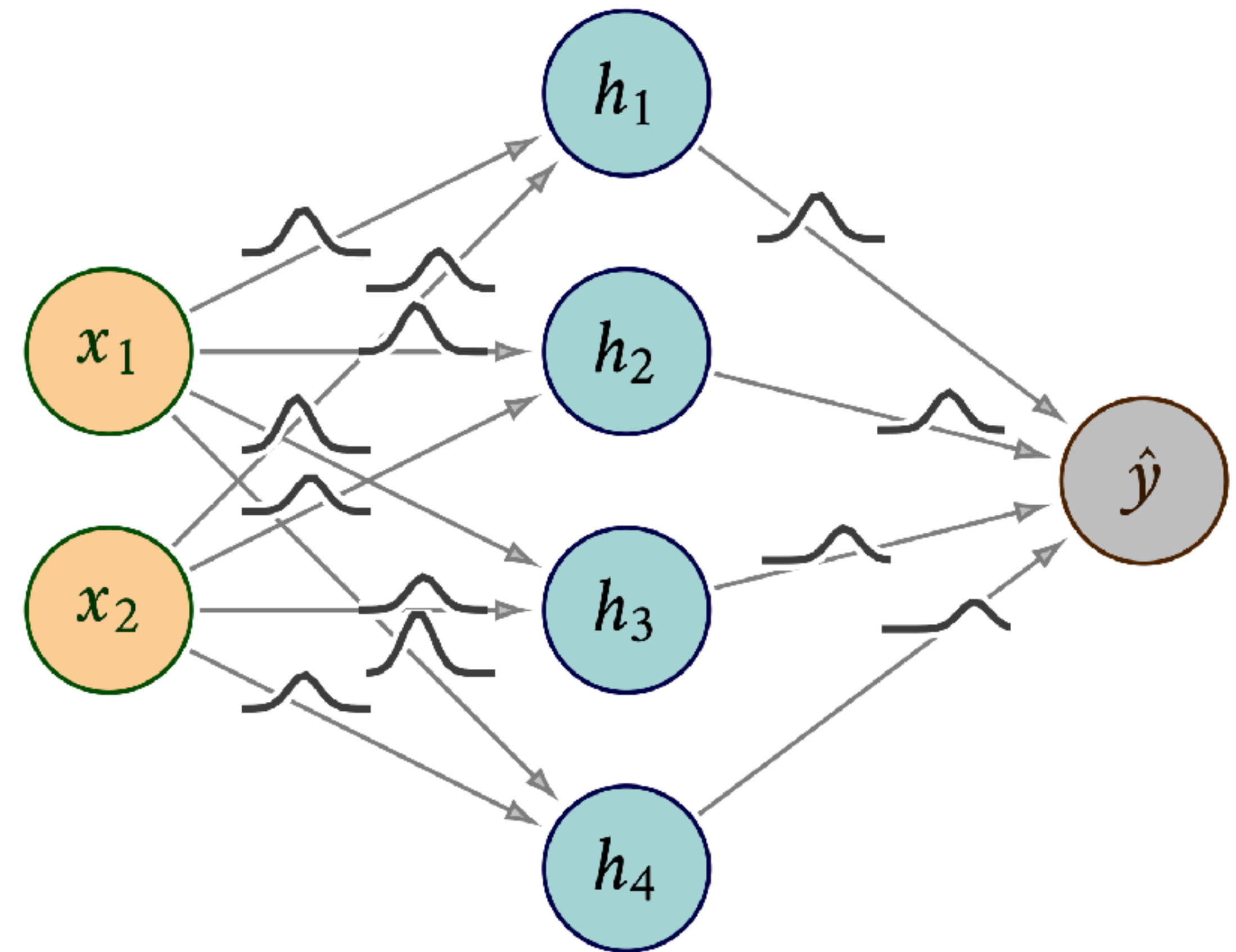
Książ Castle, Poland

Bayesian Neural Networks

Regular Neural Network

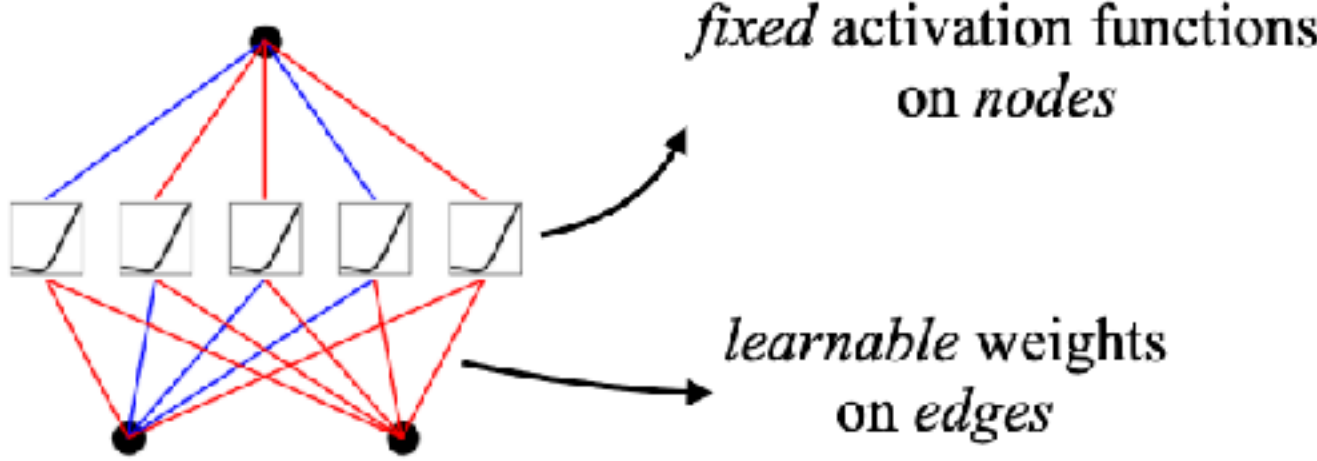
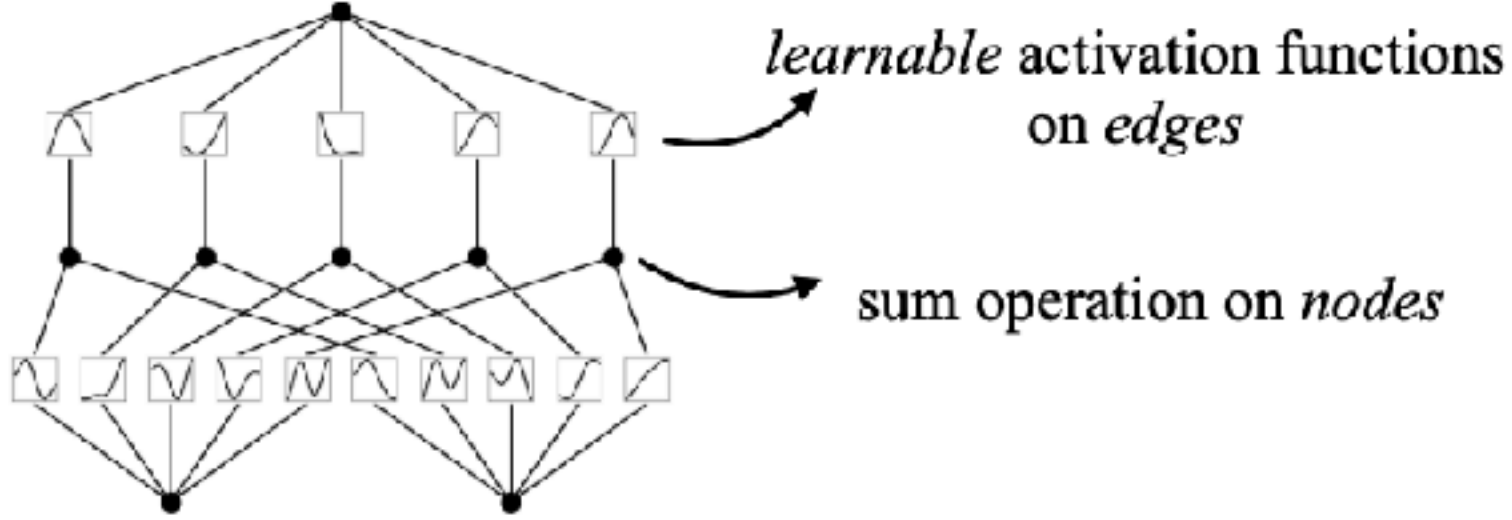
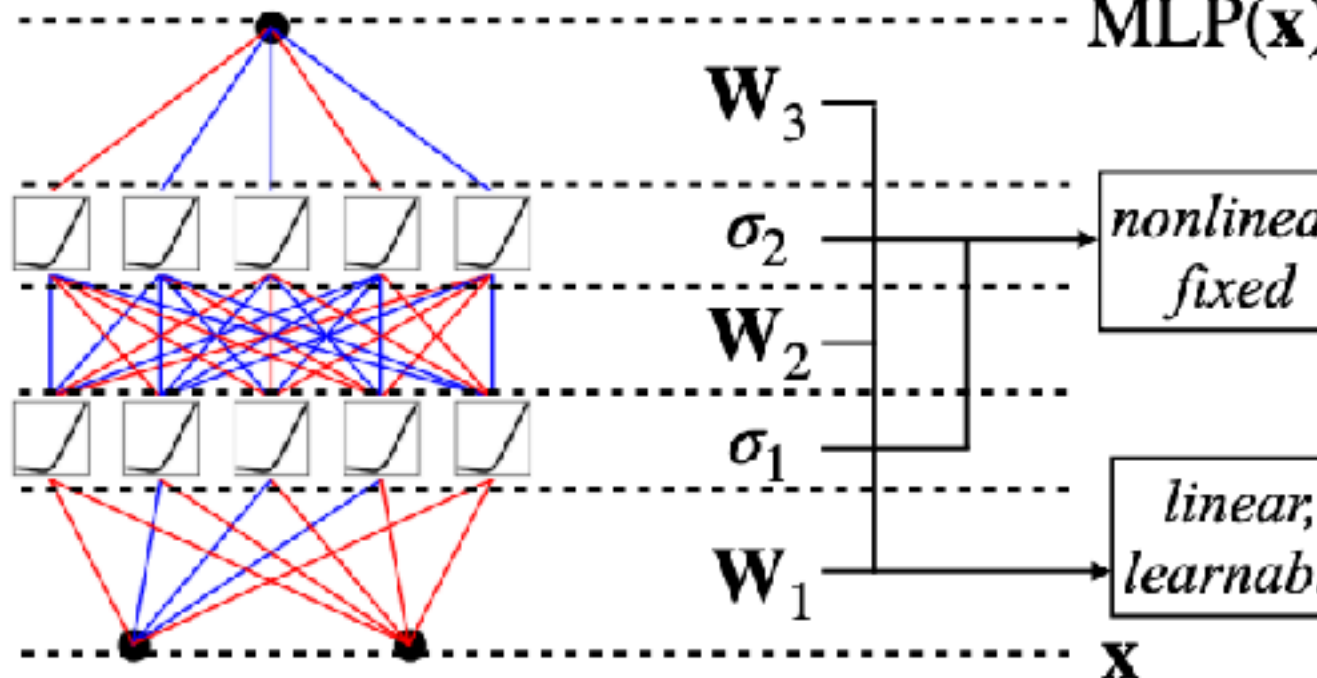
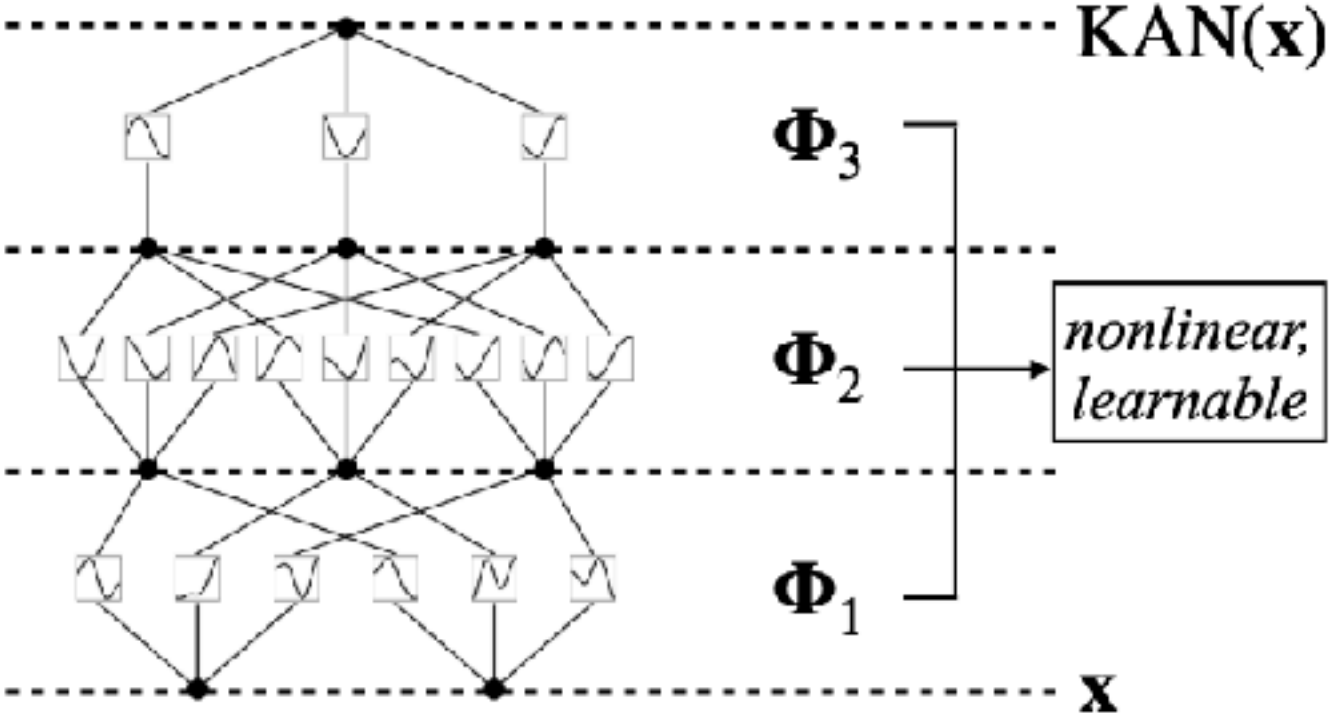


Bayesian Neural Network



Kolmogorov-Arnold networks

from Liu et al (2404.19756)

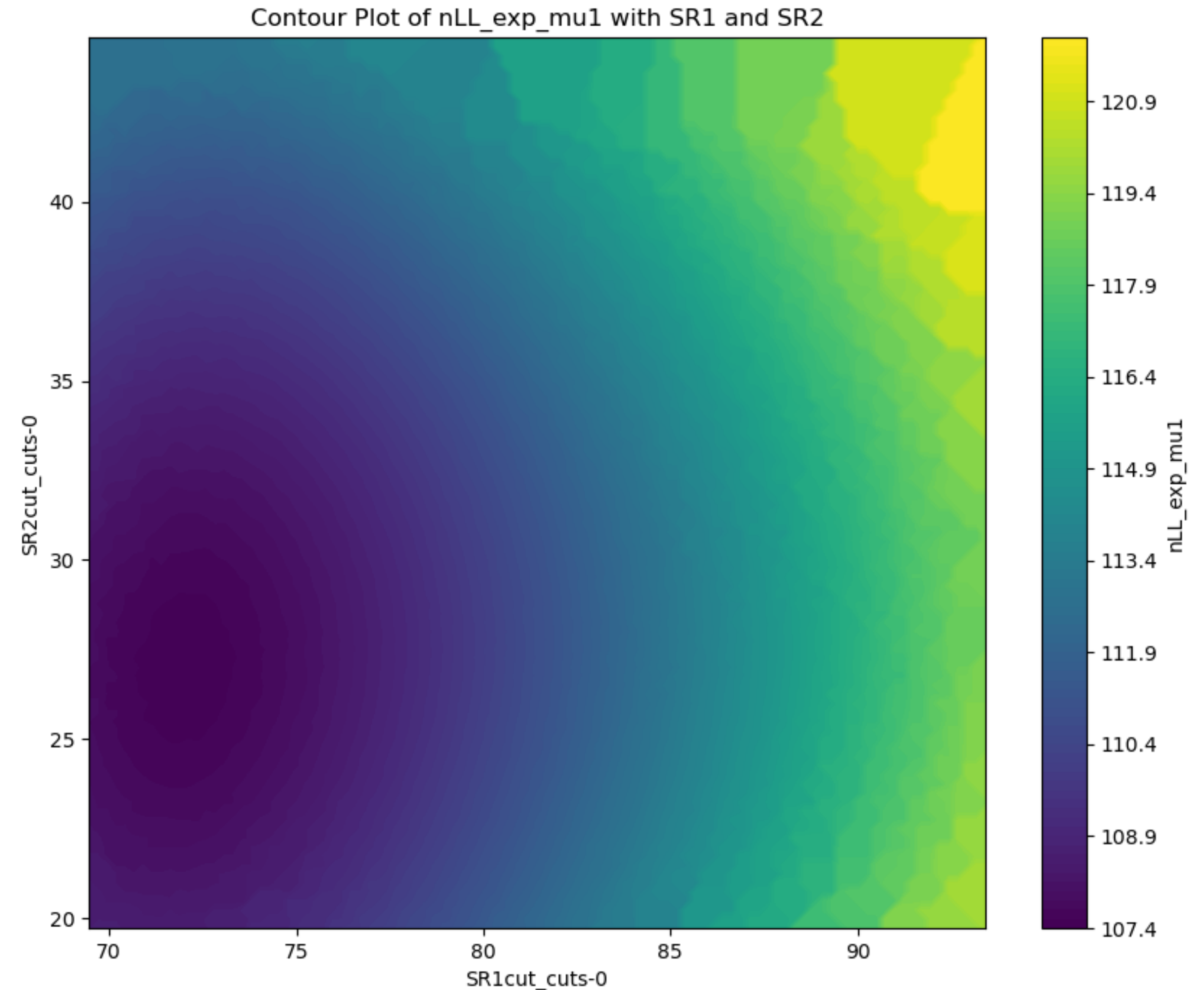
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(e)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a)  fixed activation functions on nodes learnable weights on edges	(b)  learnable activation functions on edges sum operation on nodes
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c)  MLP(x) $\mathbf{W}_3$ σ_2 $\mathbf{W}_2$ σ_1 $\mathbf{W}_1$ $\mathbf{x}$ nonlinear, fixed linear, learnable	(d)  KAN(x) Φ_3 Φ_2 Φ_1 $\mathbf{x}$ nonlinear, learnable

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

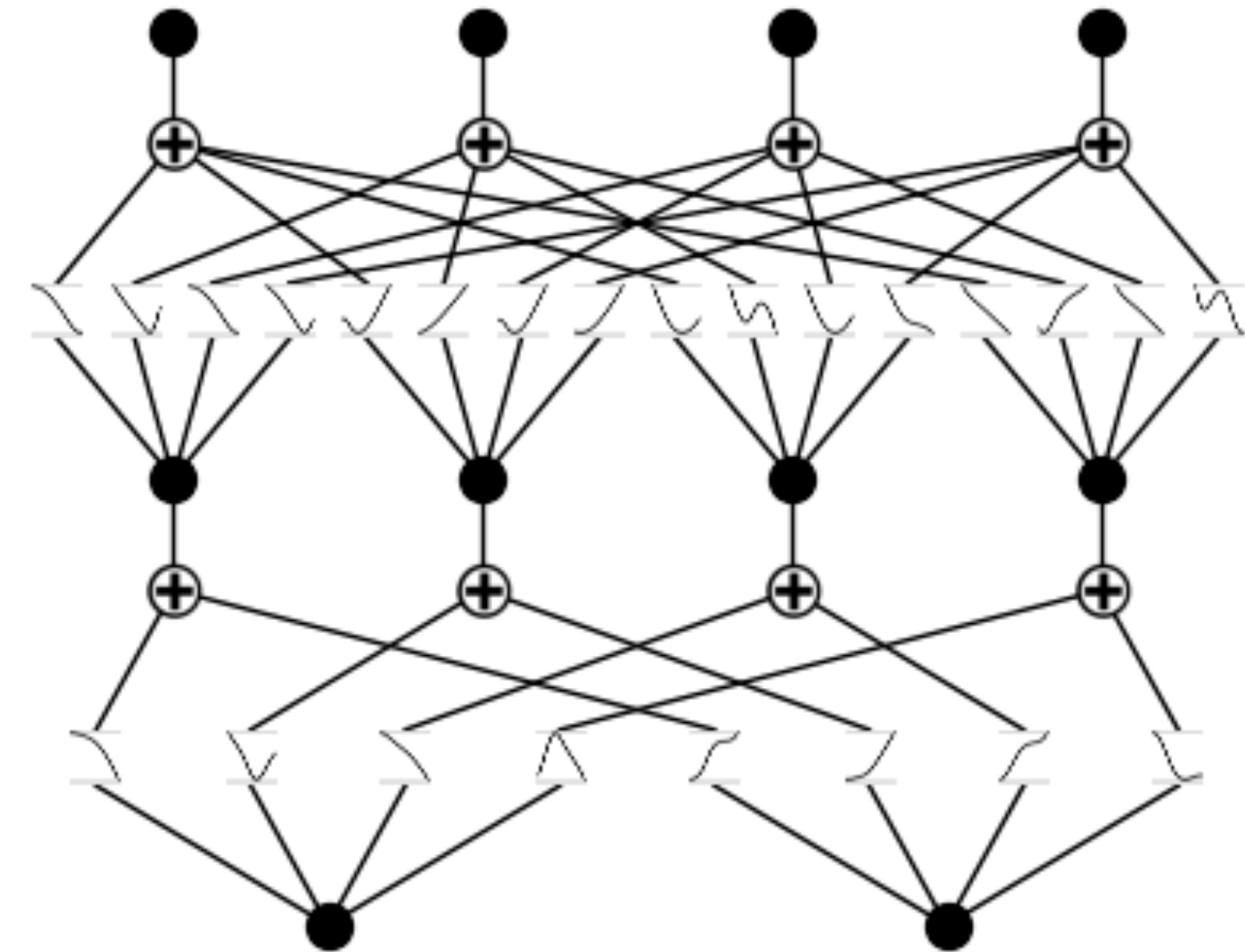
ATLAS-SUSY-2018-04 (1911.0660)



Kolmogorov-Arnold networks

ATLAS-SUSY-2018-04 (1911.0660)

- 🌸 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

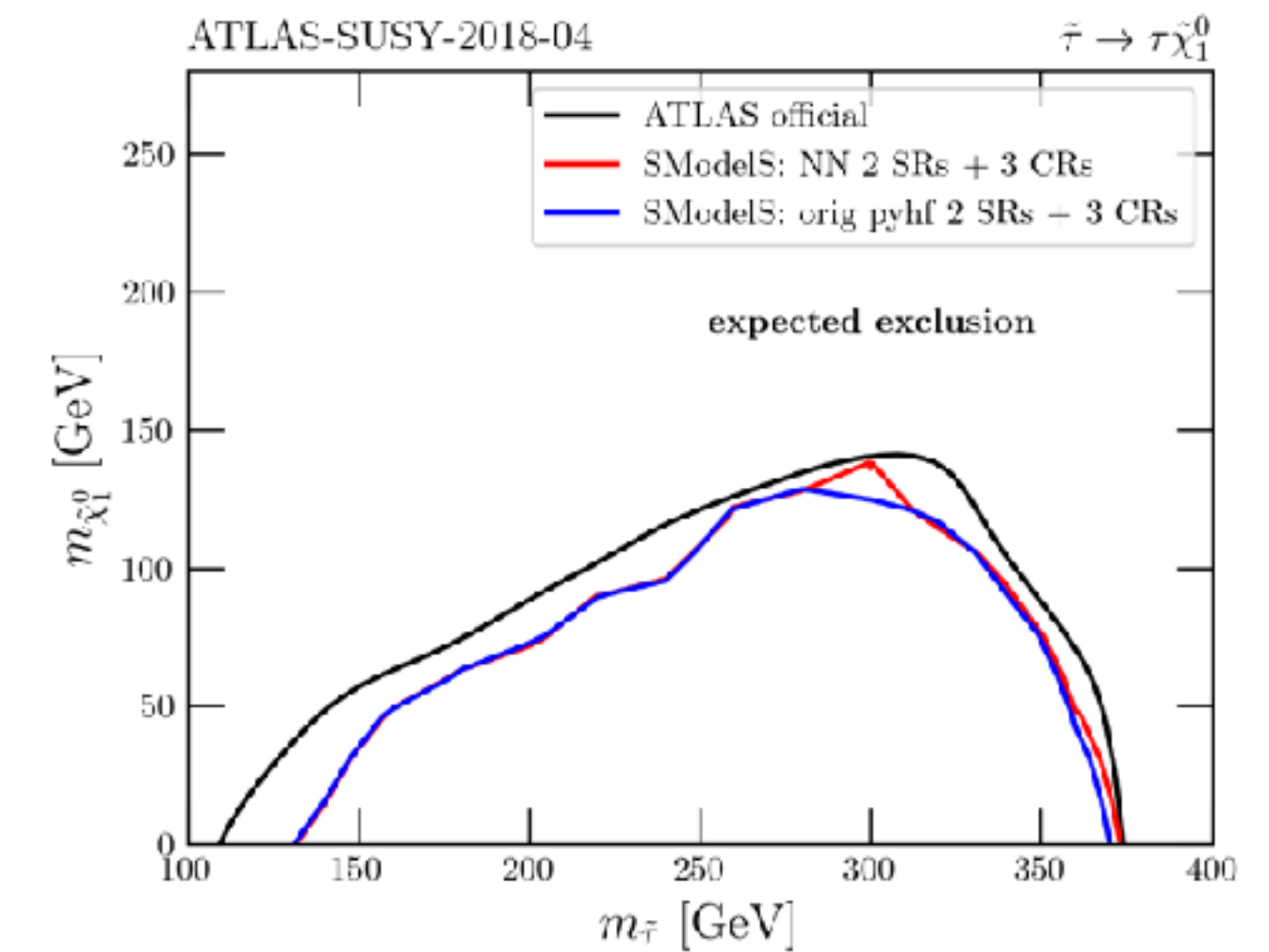
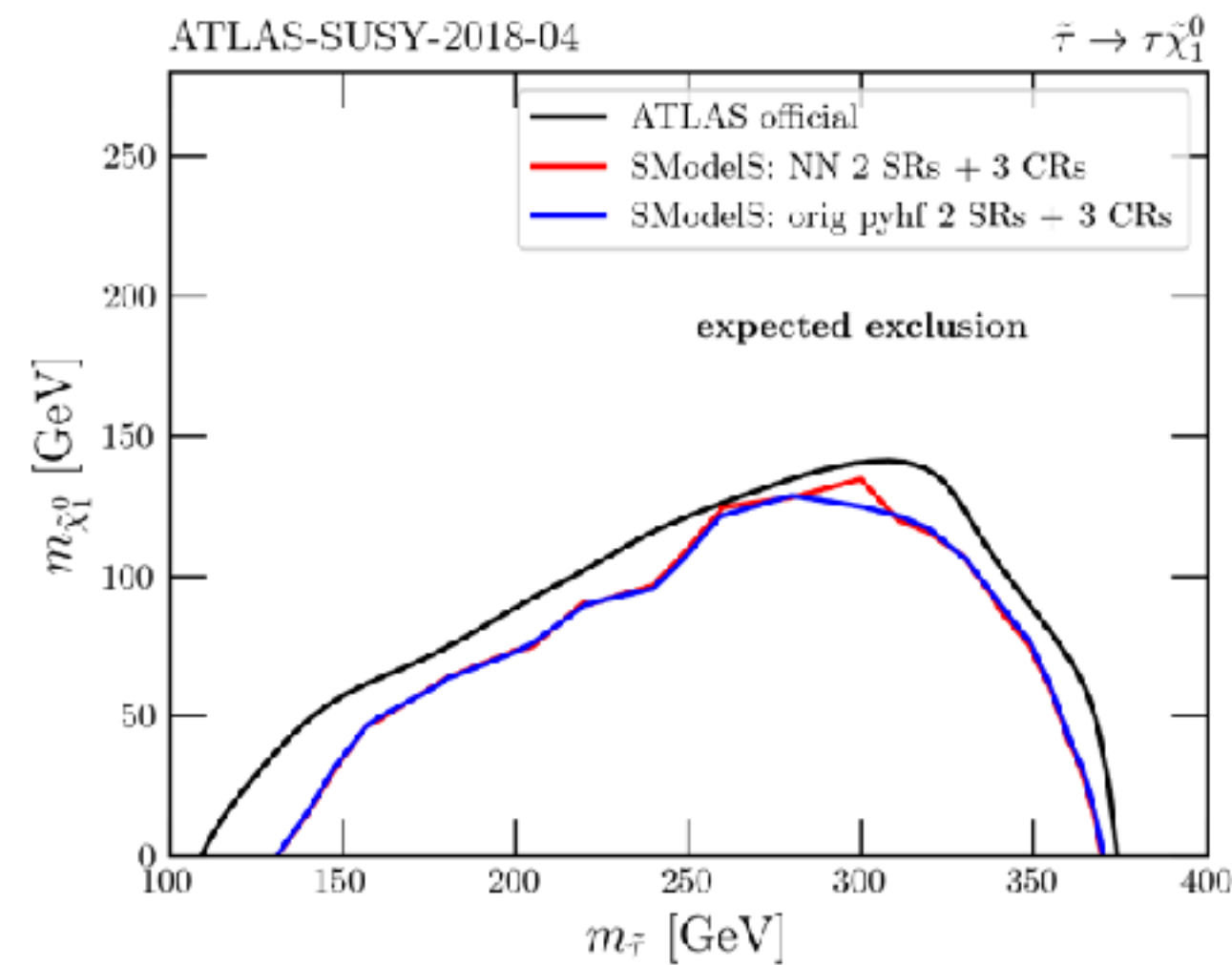
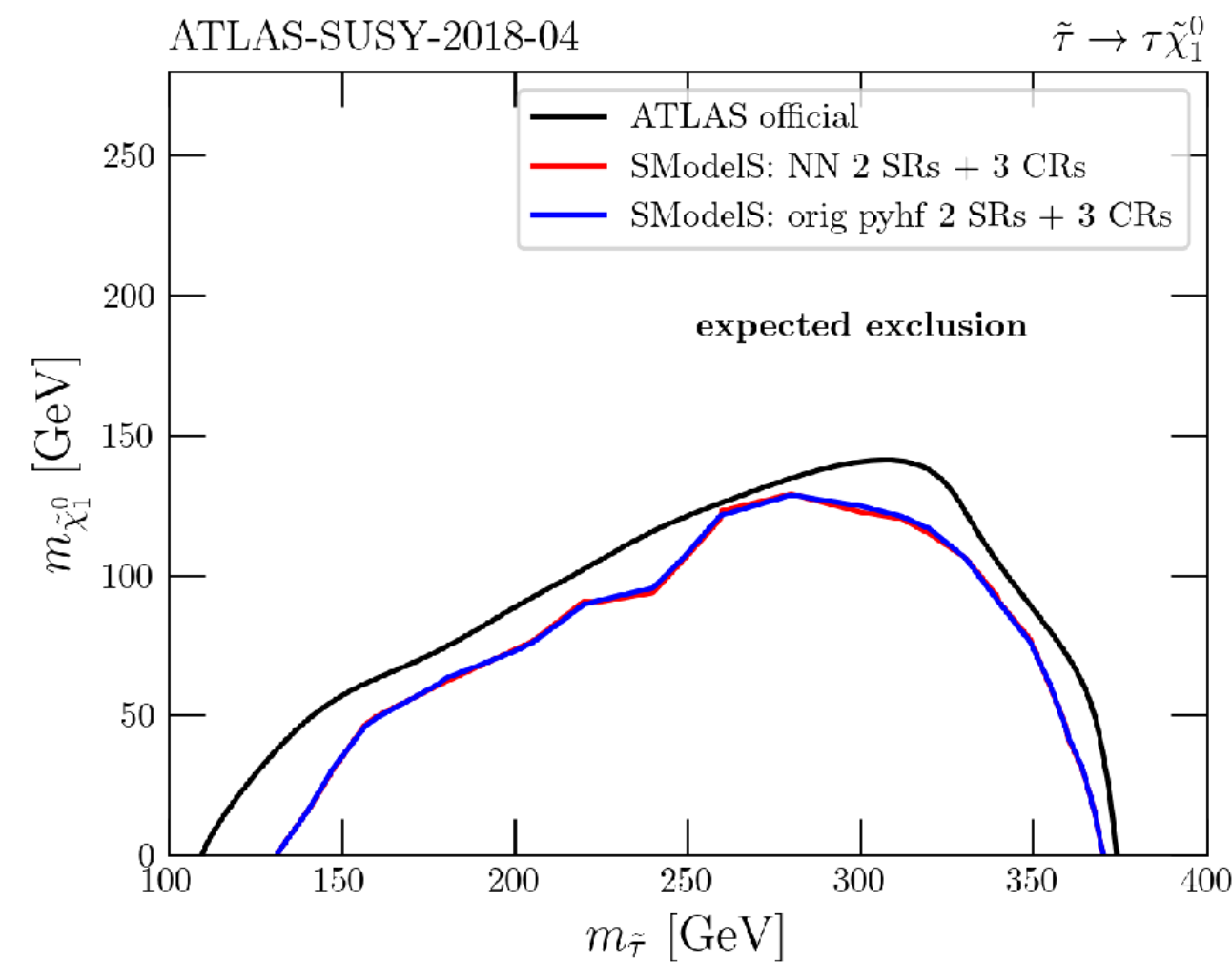


MLP

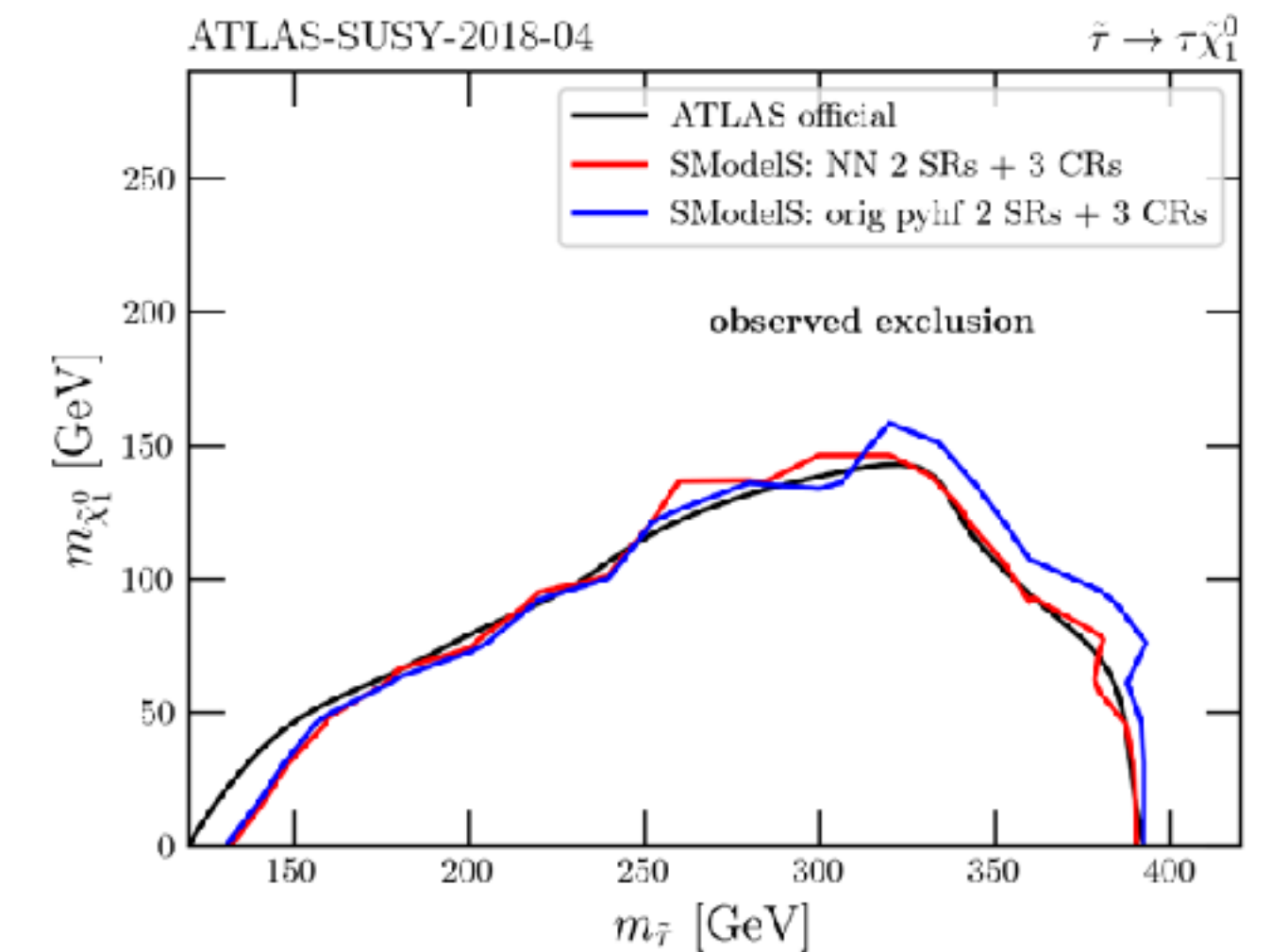
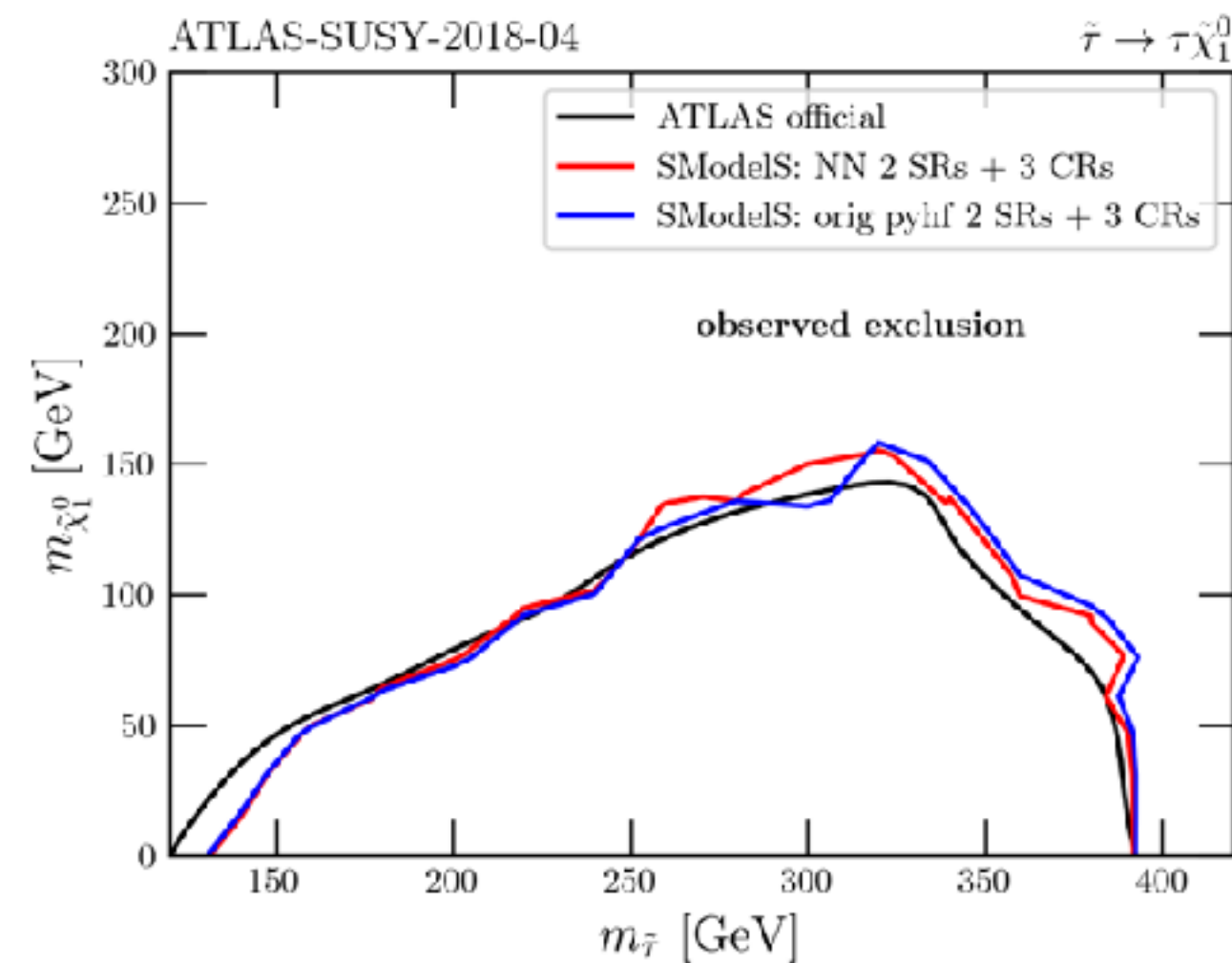
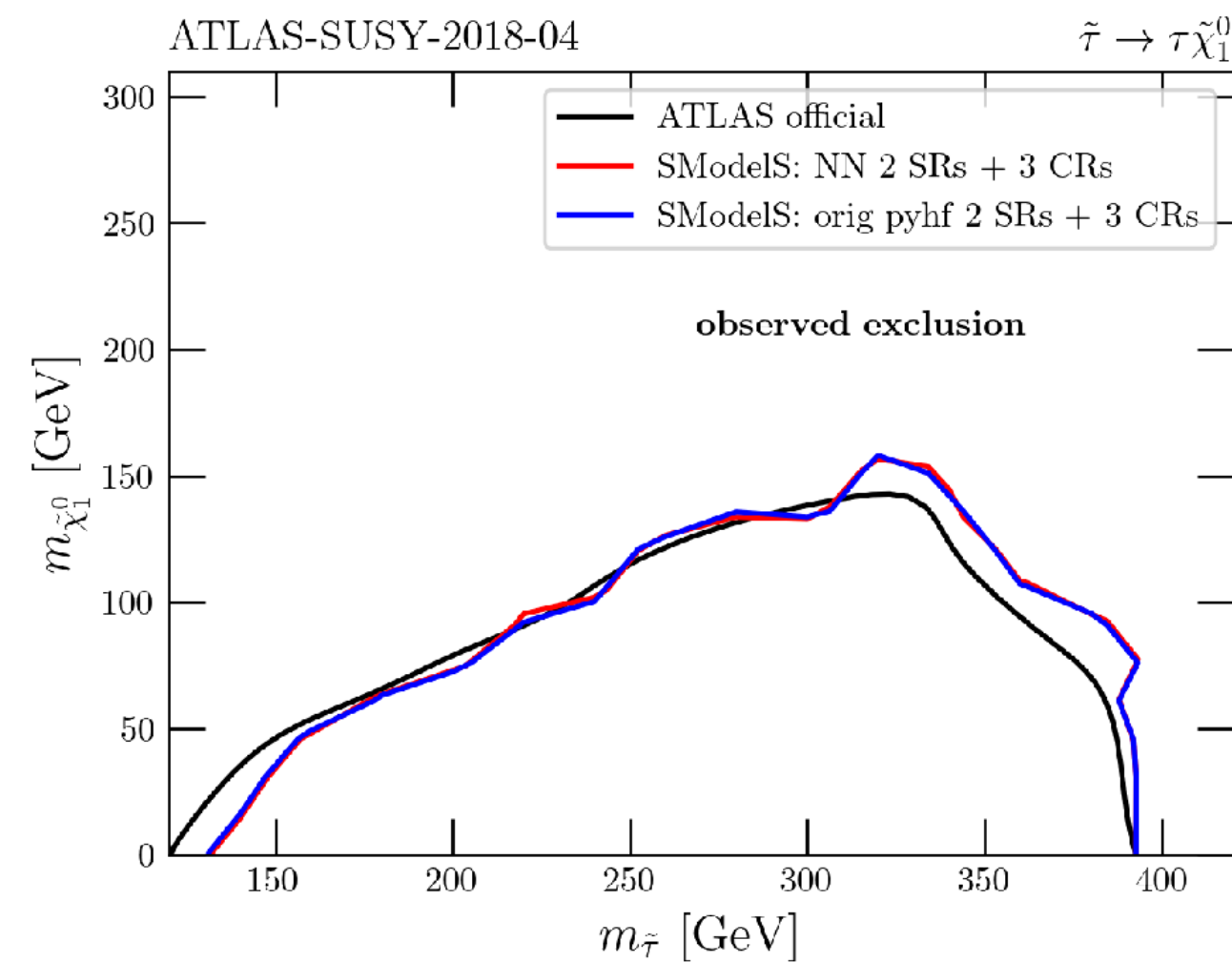
KAN

KAN- symbolic

expected



observed



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