Before asymptopia: High-dimensional inference in small sample size regimes in particle physics

Aishik Ghosh

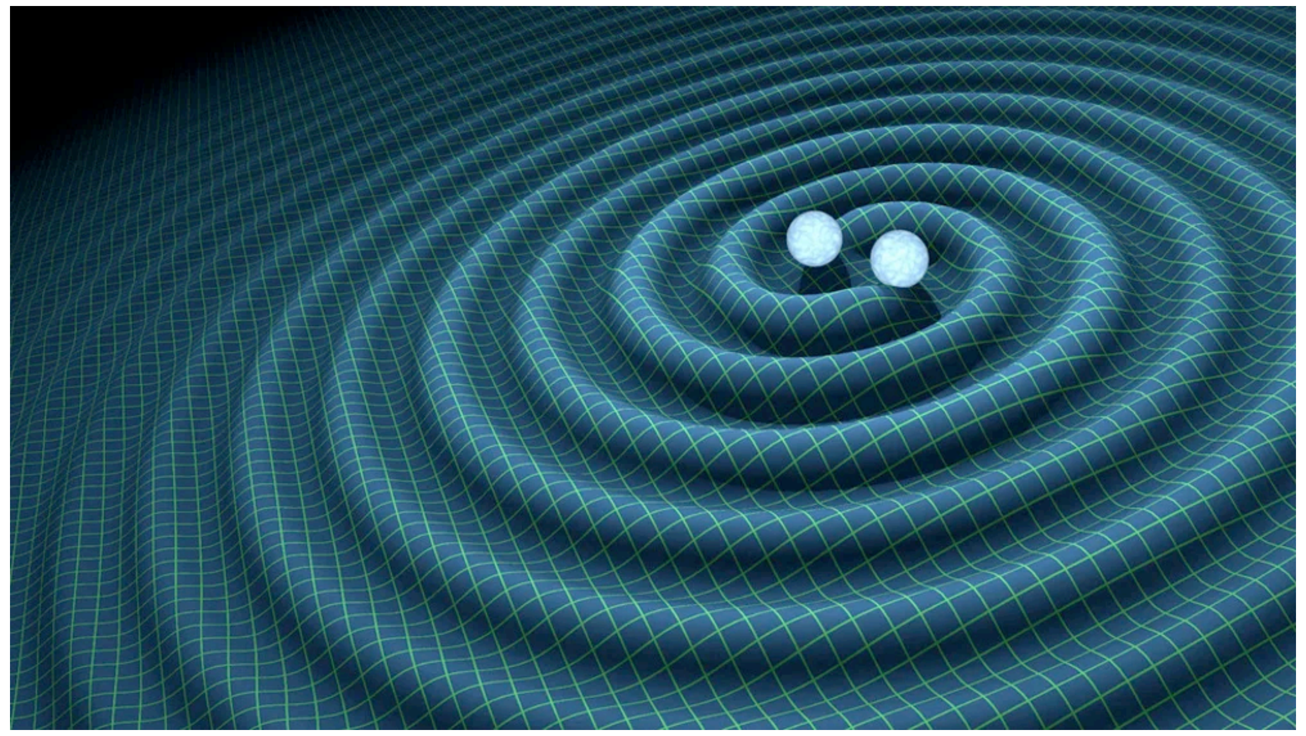
STAMPS Workshop
14 May 2026

Georgia Institute of Technology

Prof-Aishik-Ghosh
aishikghosh.bsky.social
@Aishik_Ghosh_

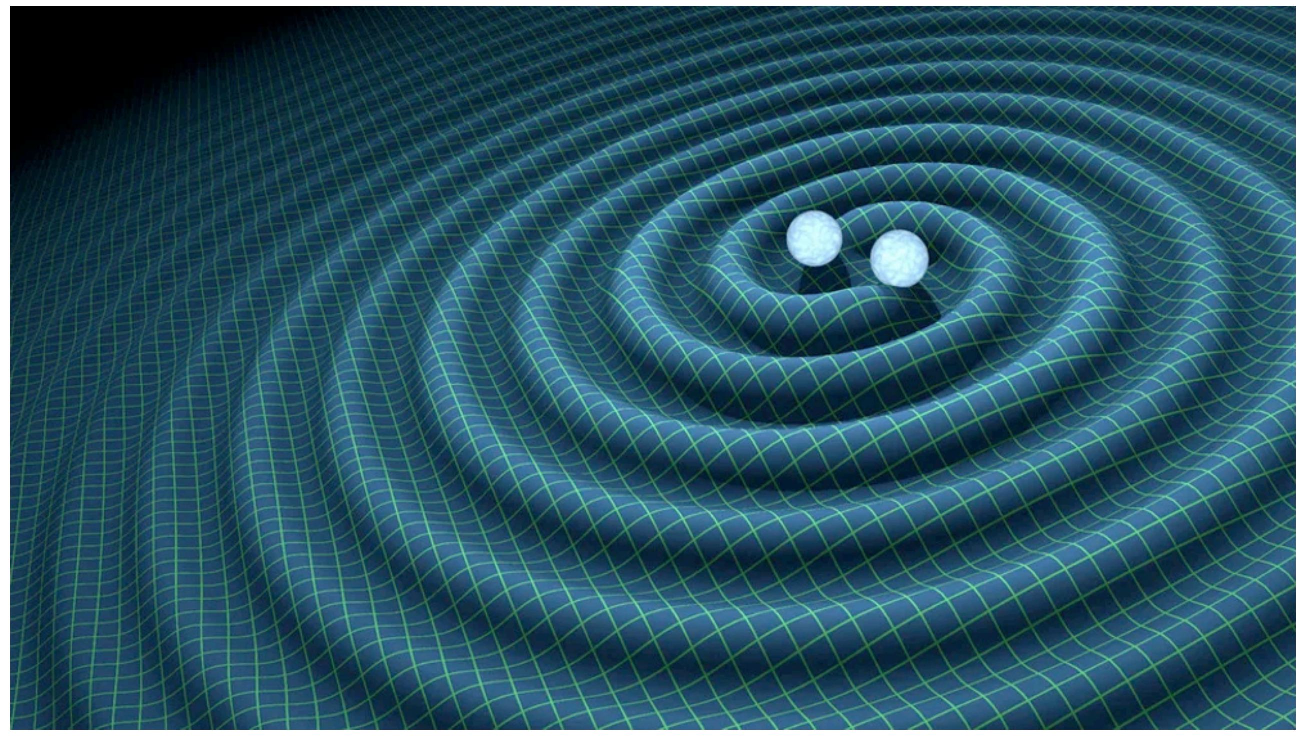
Gravitational Wave discovery (2016)

Astrophysics:

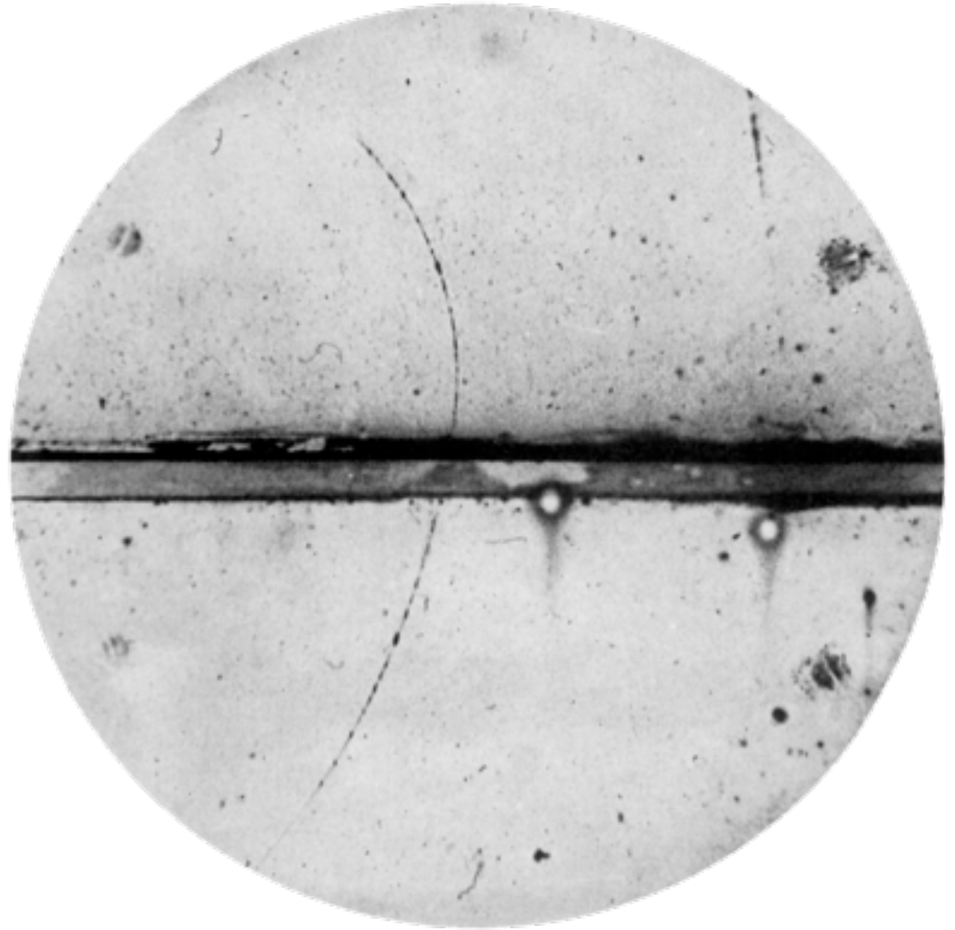


Gravitational Wave discovery (2016)

Astrophysics:



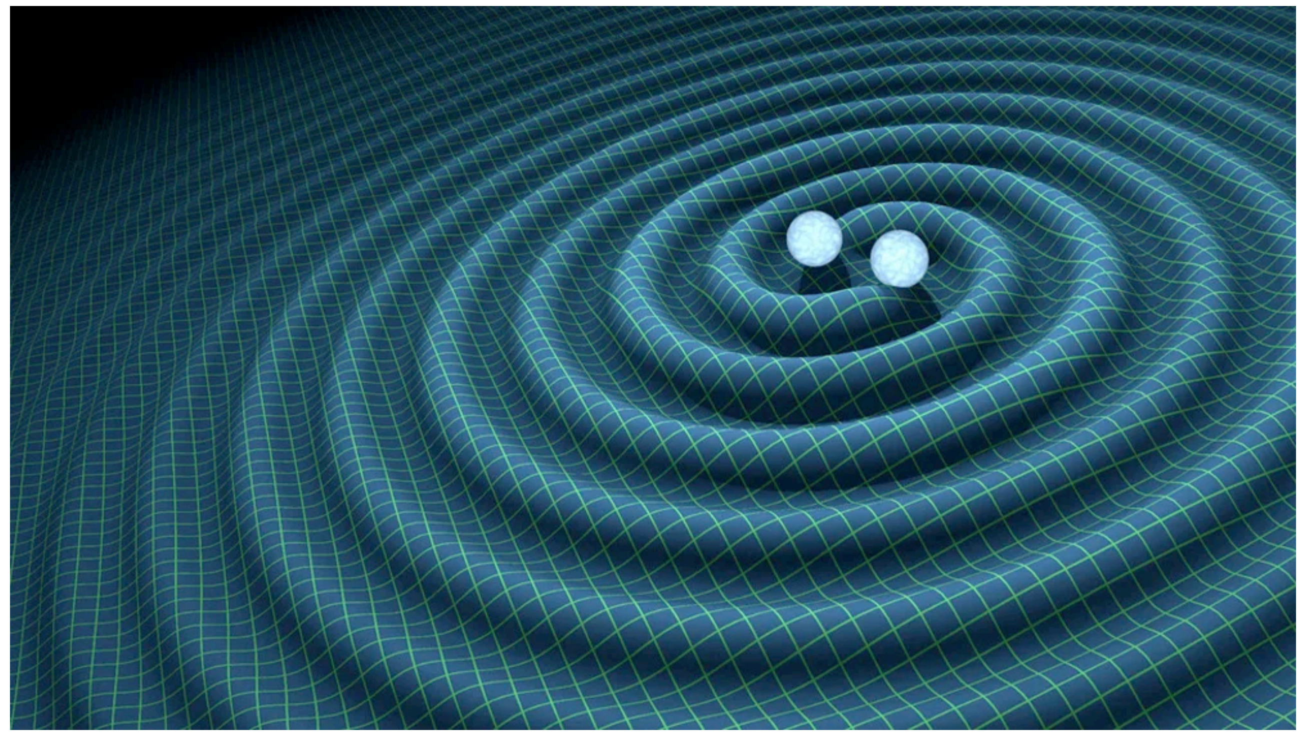
Particle physics:



Positron discovery (1930s)

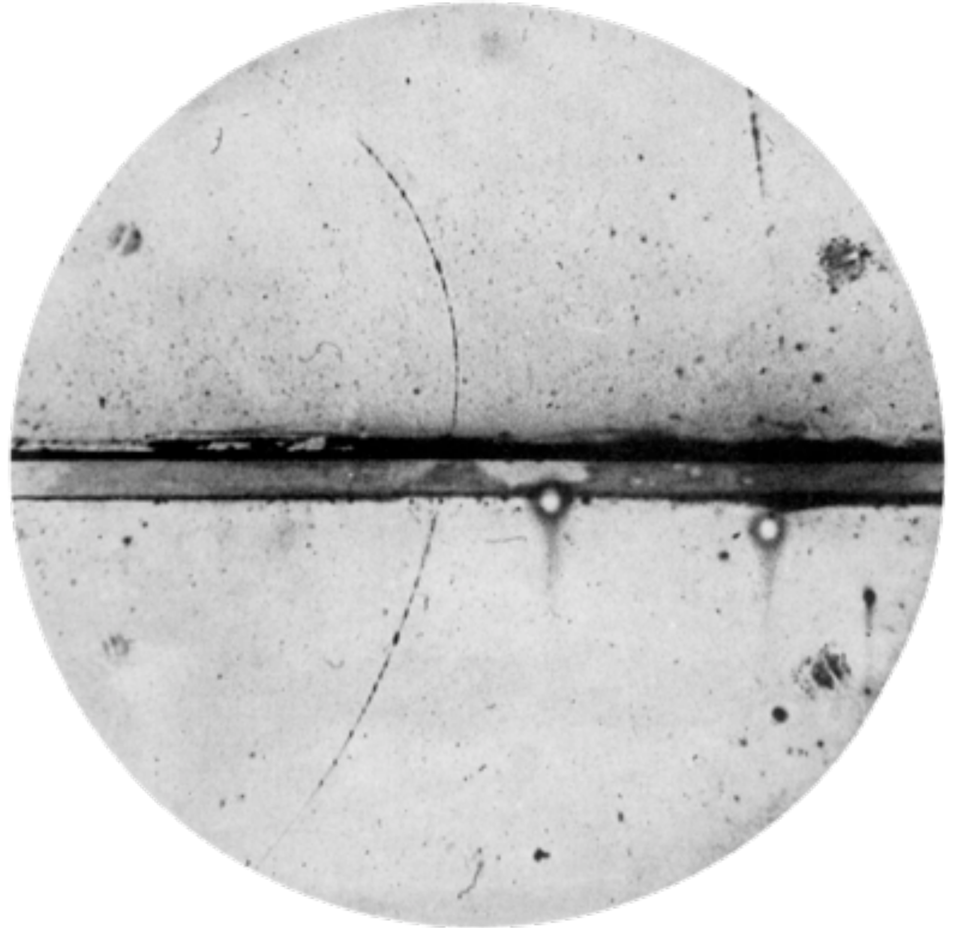
Gravitational Wave discovery (2016)

Astrophysics:



Single event

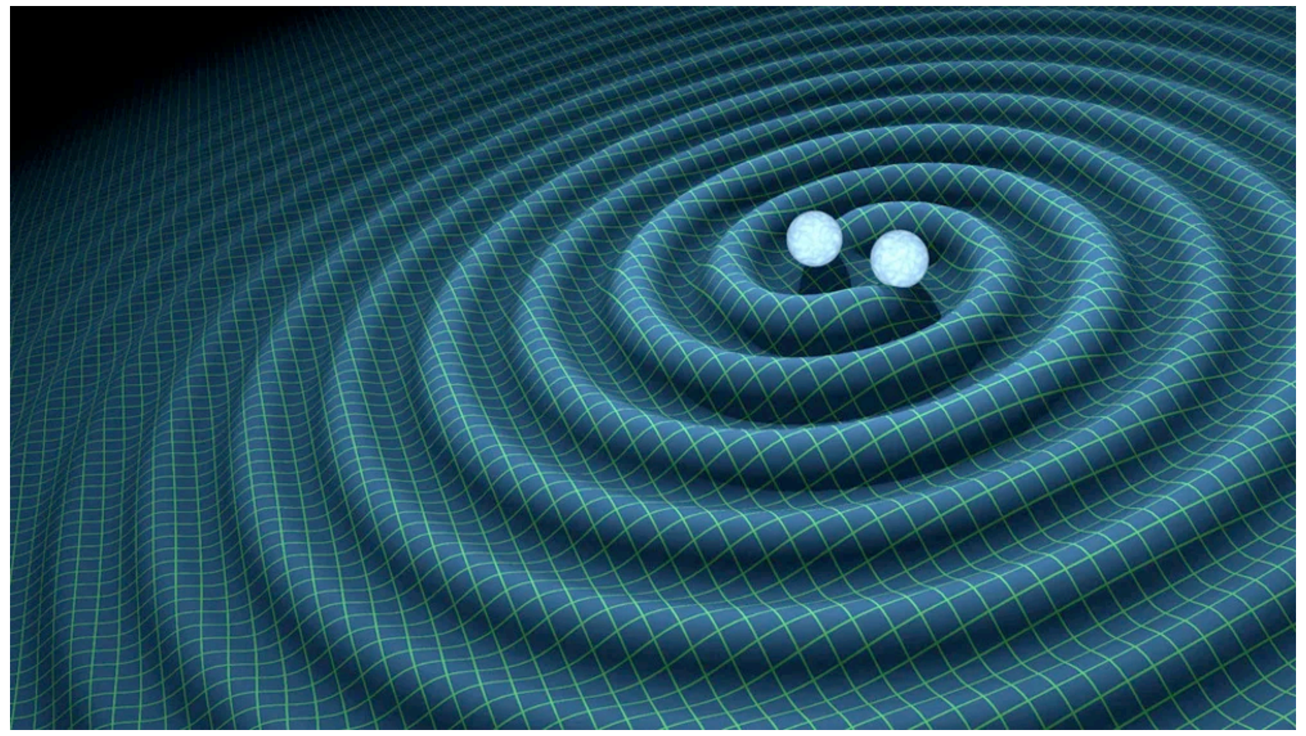
Particle physics:



Positron discovery (1930s)

Gravitational Wave discovery (2016)

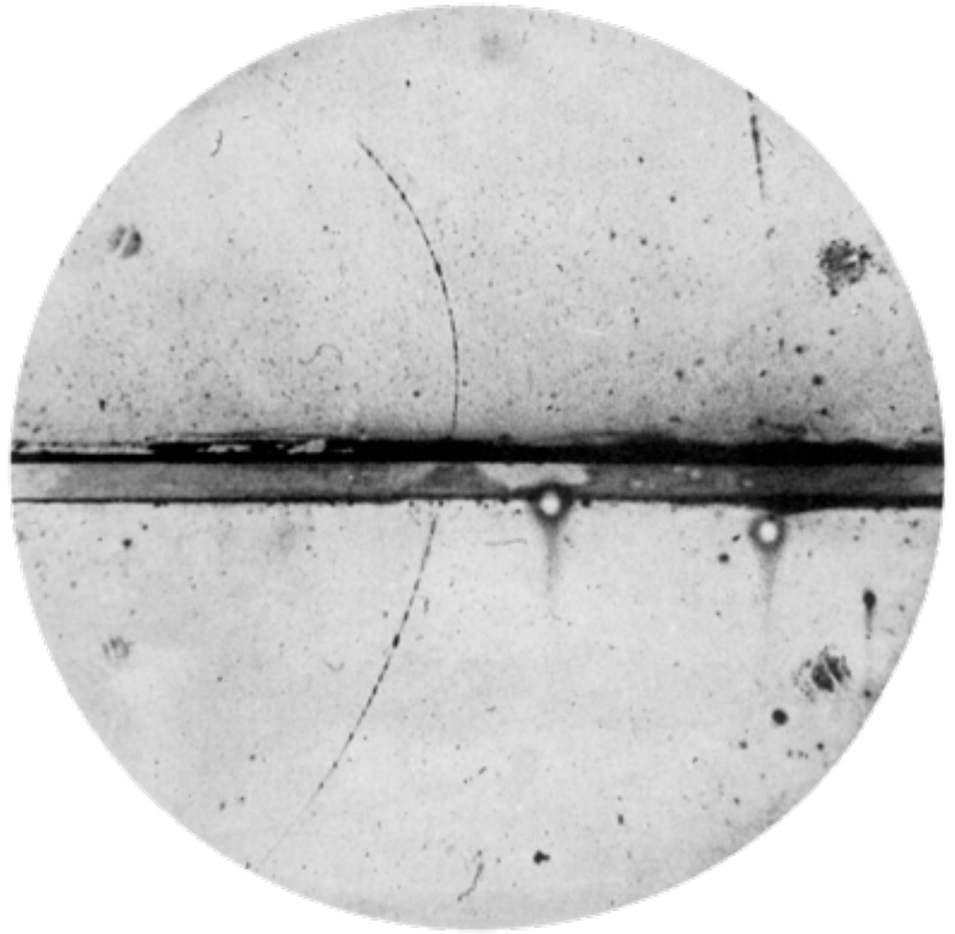
Astrophysics:



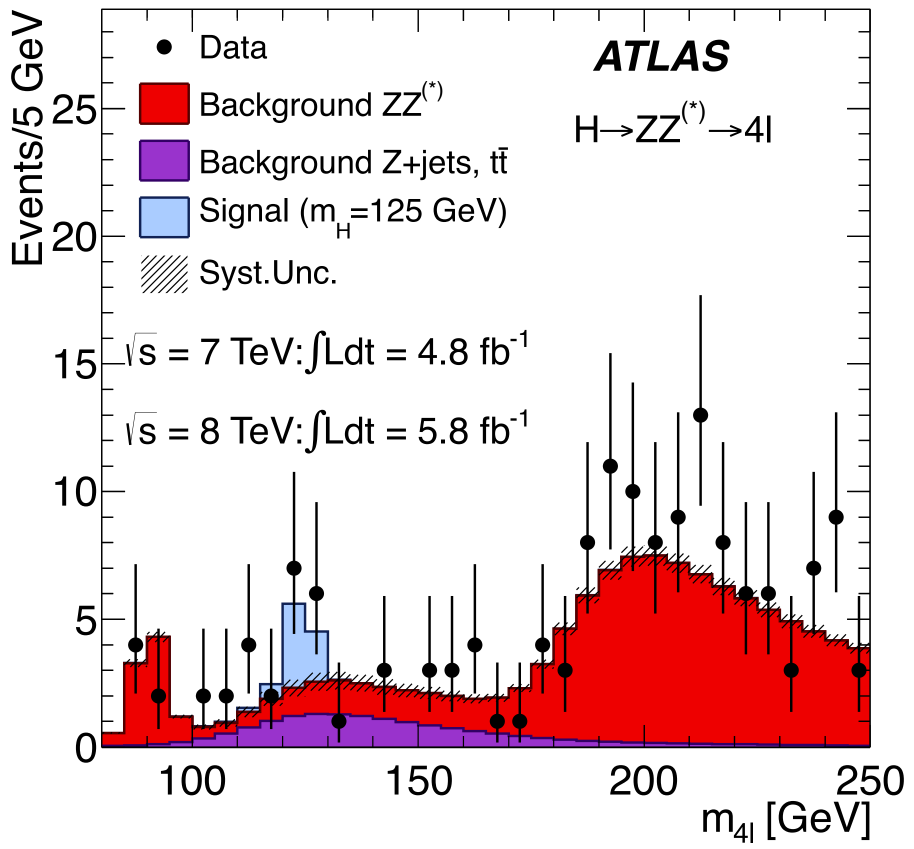
Single event

Collection of events

Particle physics:



Positron discovery (1930s)

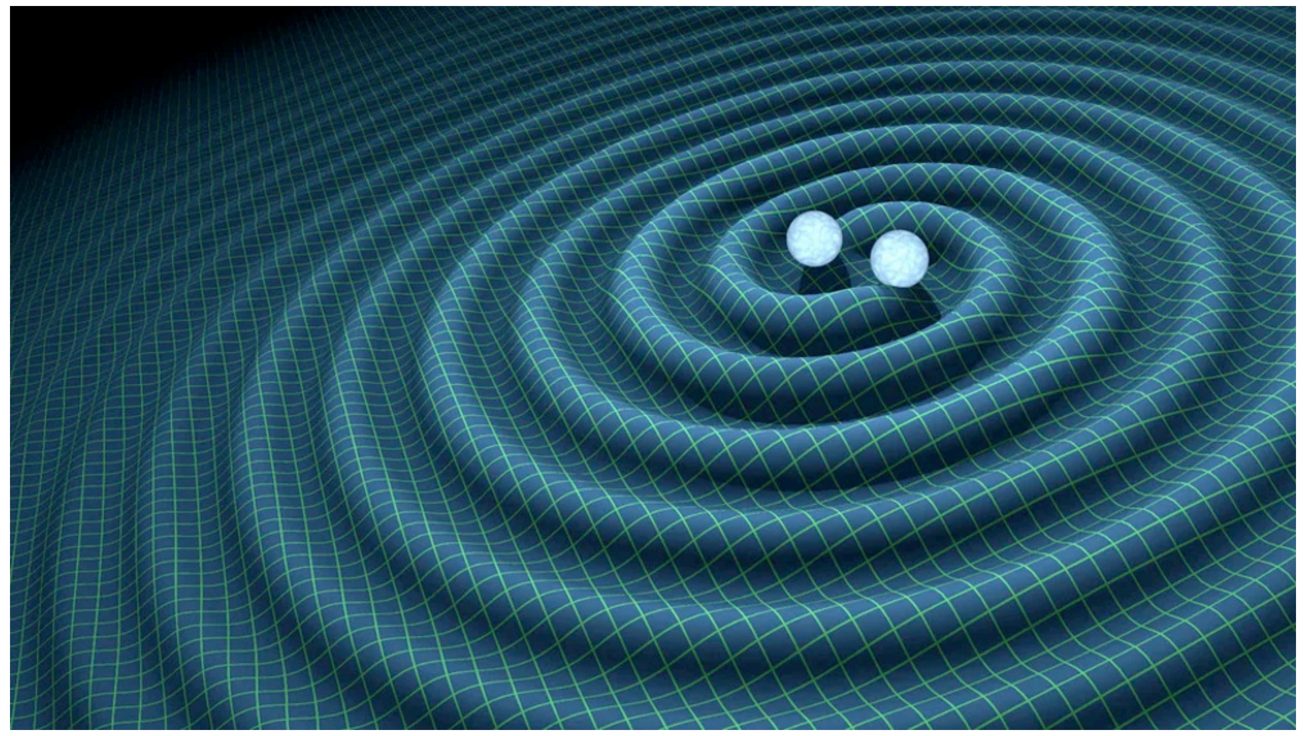


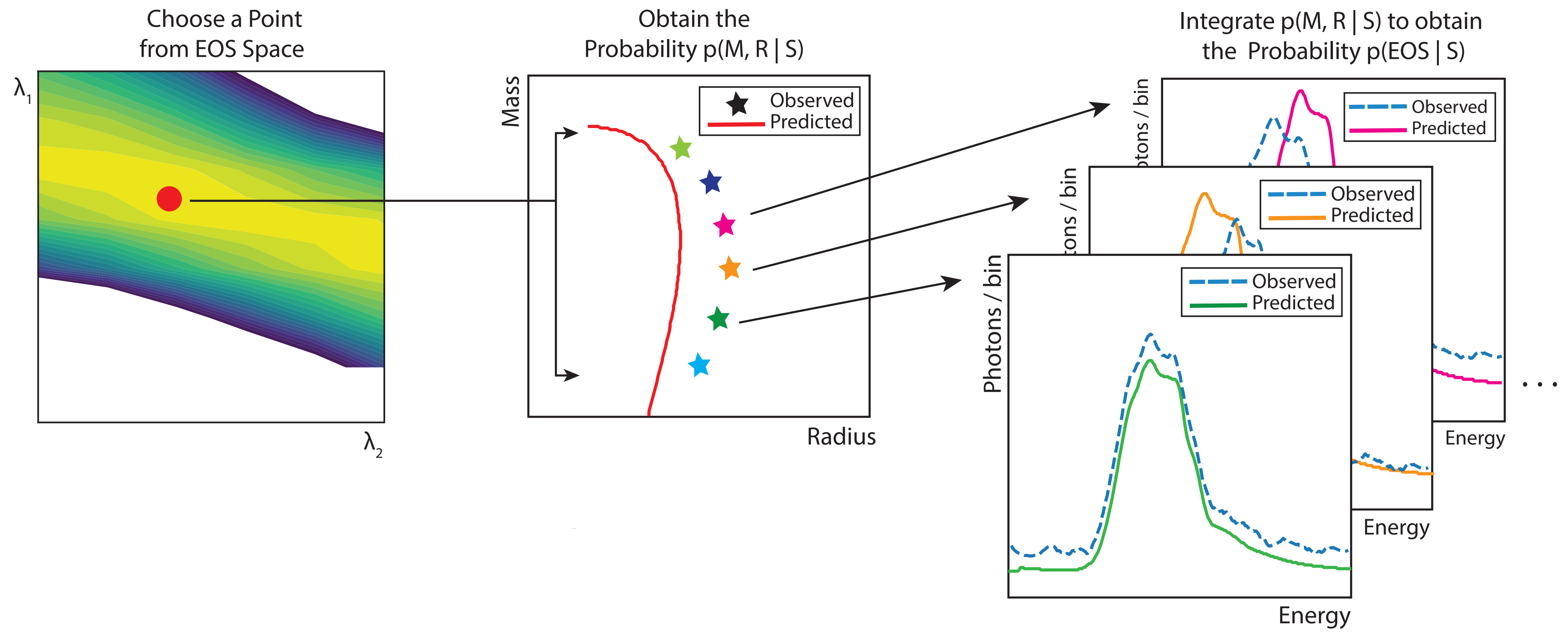
Higgs boson discovery (2012)

Gravitational Wave discovery (2016)

Probing the interior of neutron stars

Astrophysics:

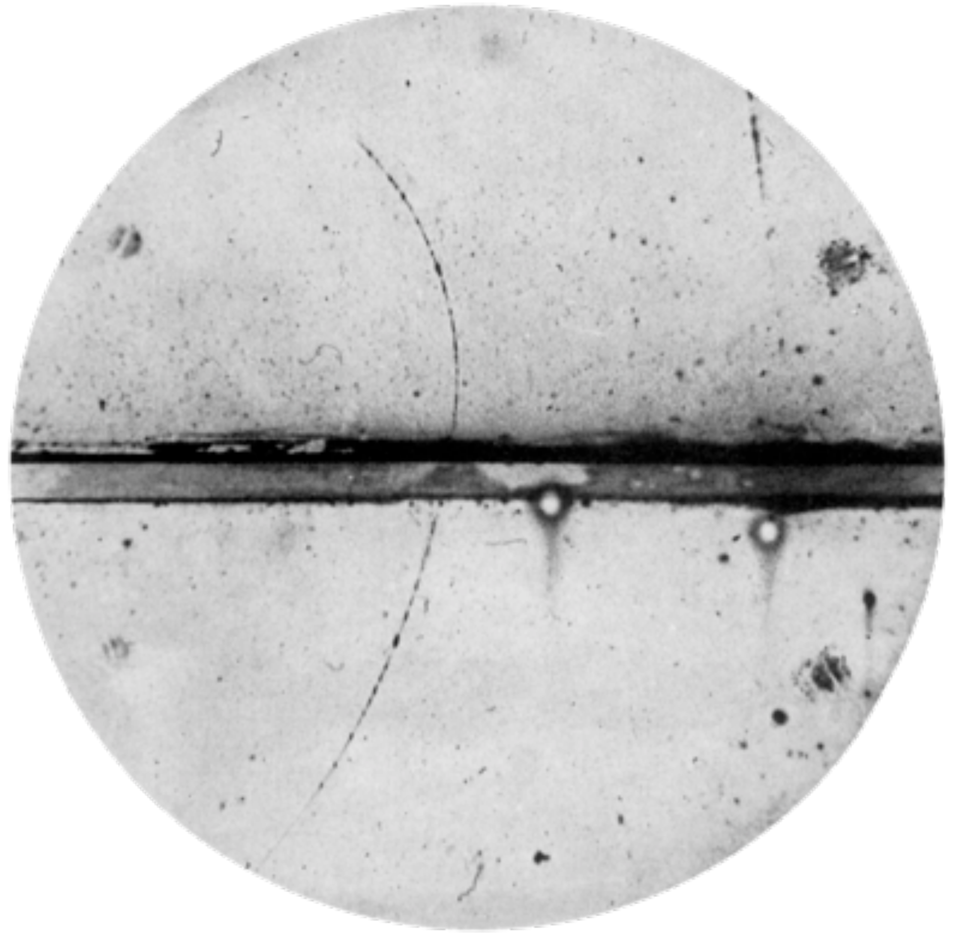




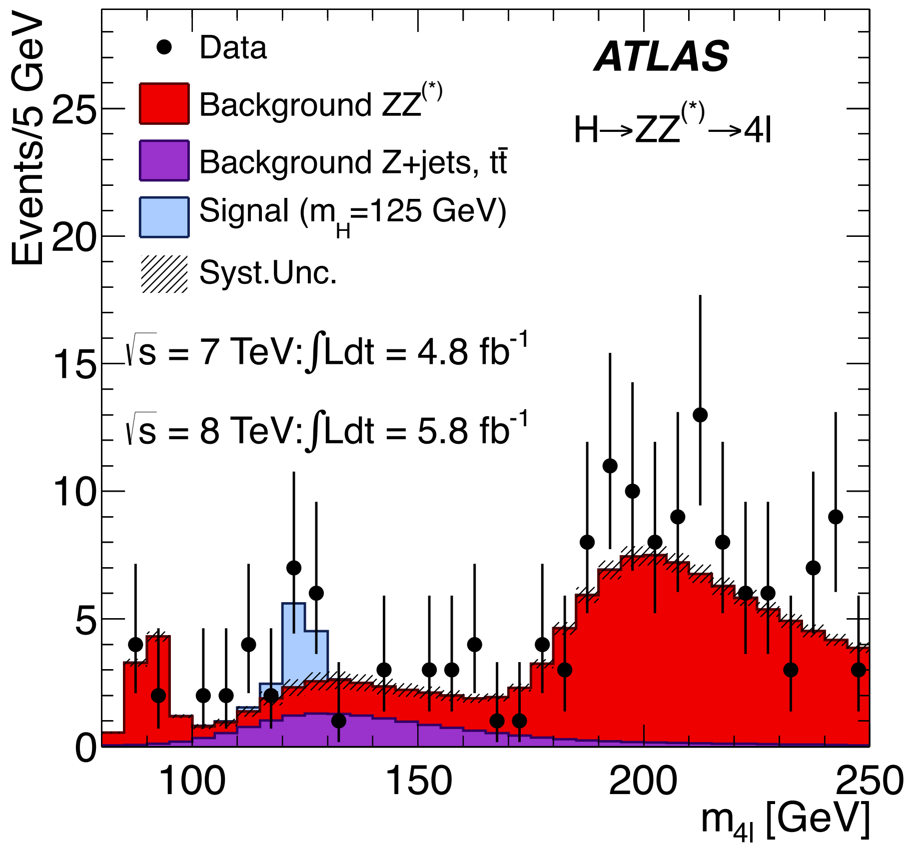
JCAP12(2023)022: Farrell et al. (incl. Ghosh)

Single event — Collection of events

Particle physics:



Positron discovery (1930s)



Higgs boson discovery (2012)

ML for Detector Simulation

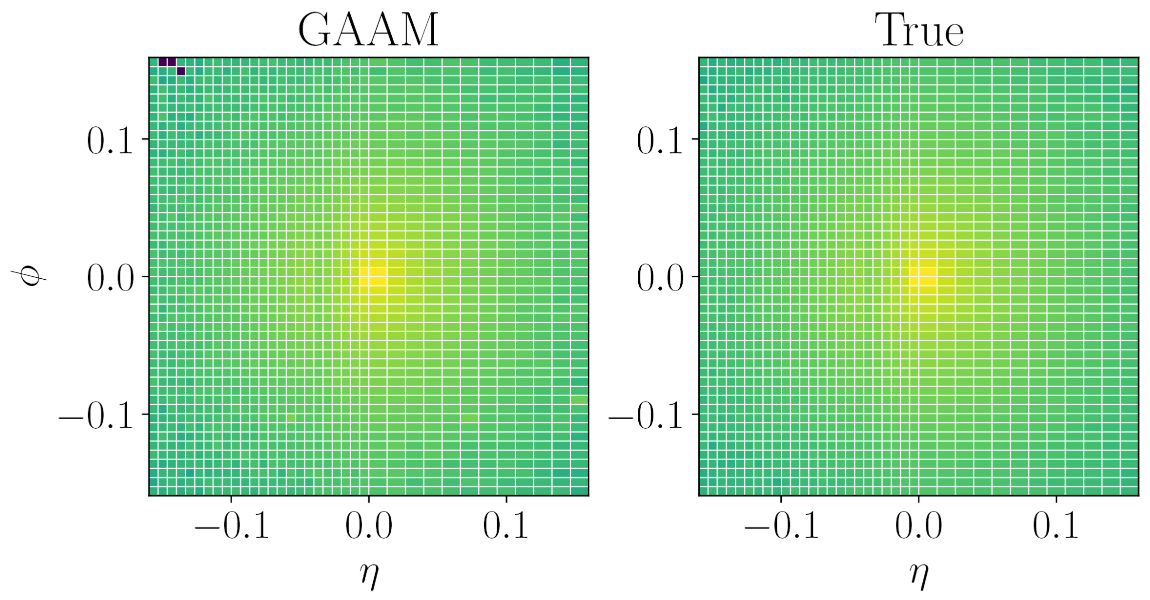


Image: arXiv:2305.11531

ML for Reconstruction

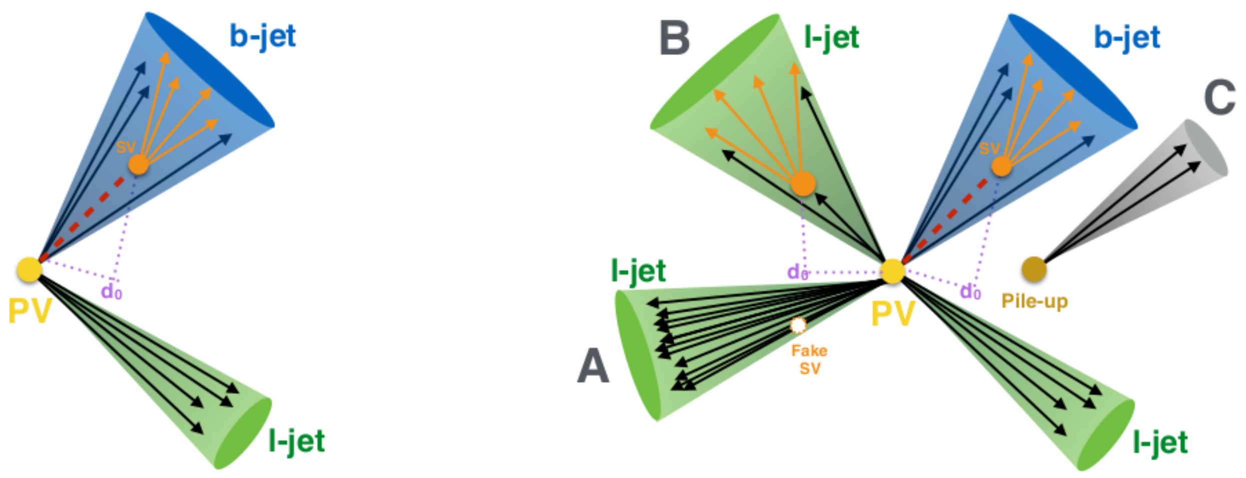


Image: USiegen group

ML for Signal vs Background

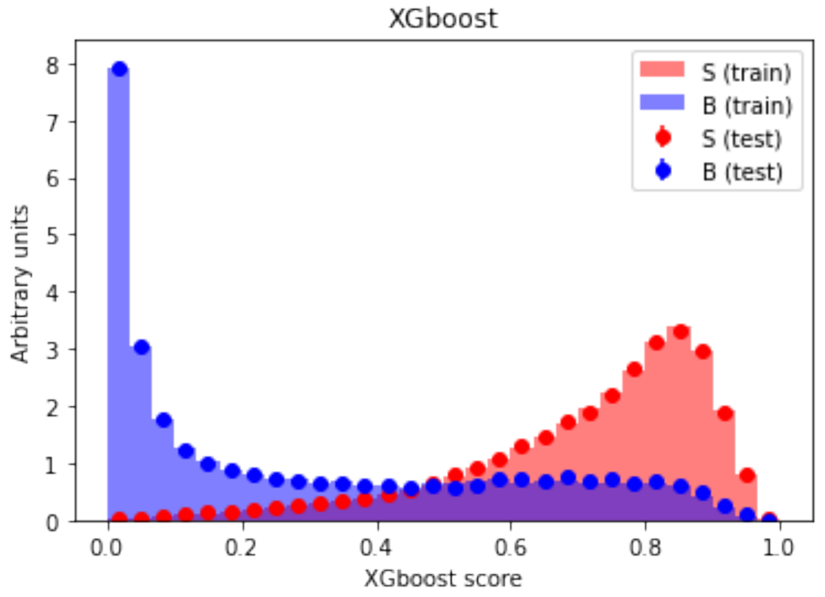


Image: ML4FP School

ML for Detector Simulation

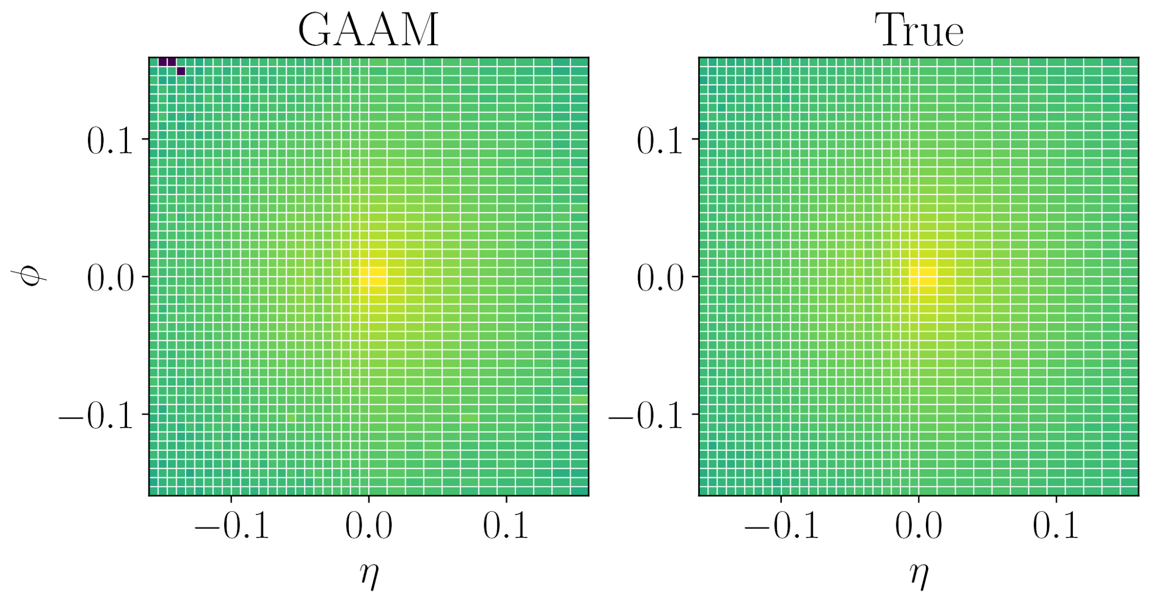


Image: arXiv:2305.11531

ML for Reconstruction

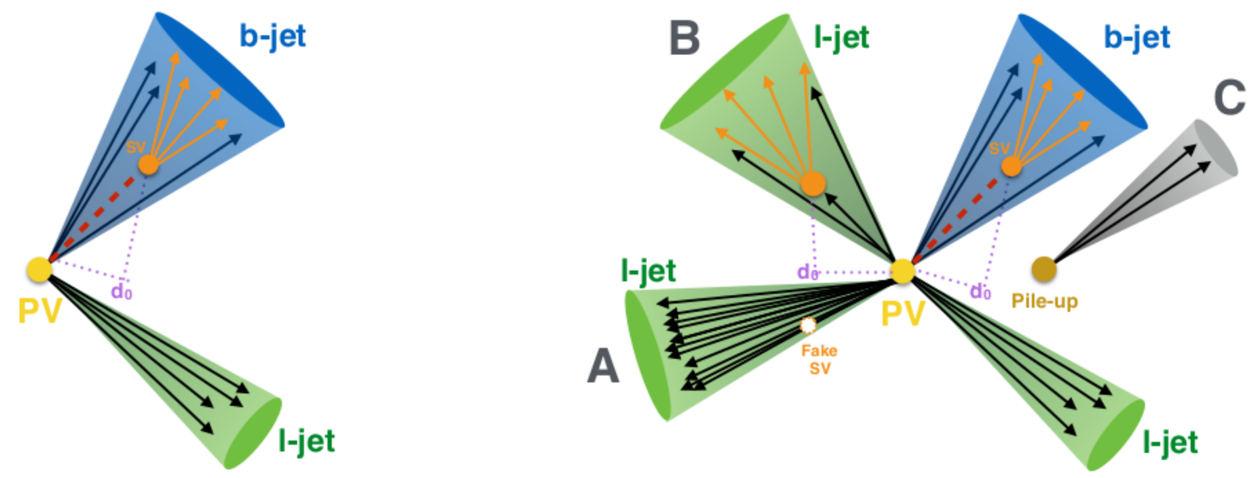


Image: USiegen group

ML for Signal vs Background

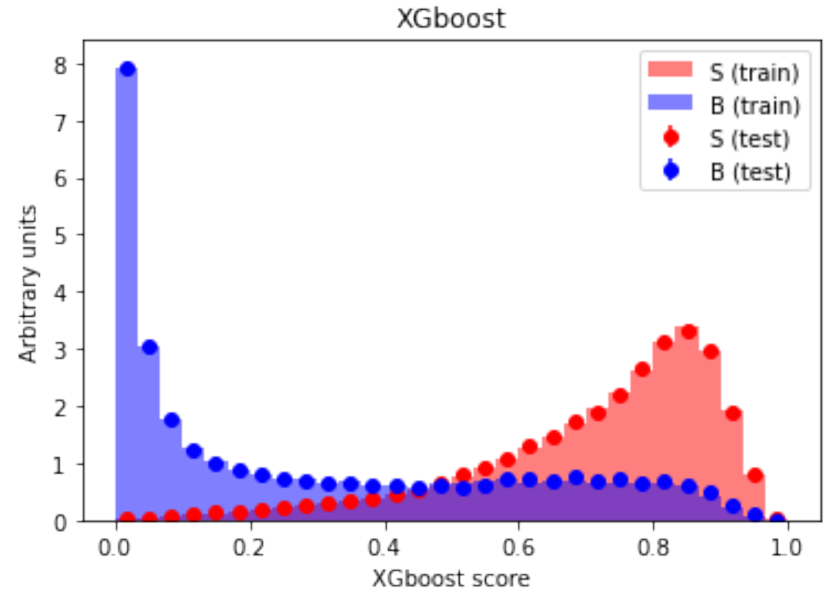


Image: ML4FP School

The final statistical inference step has remained the same

ML for Detector Simulation

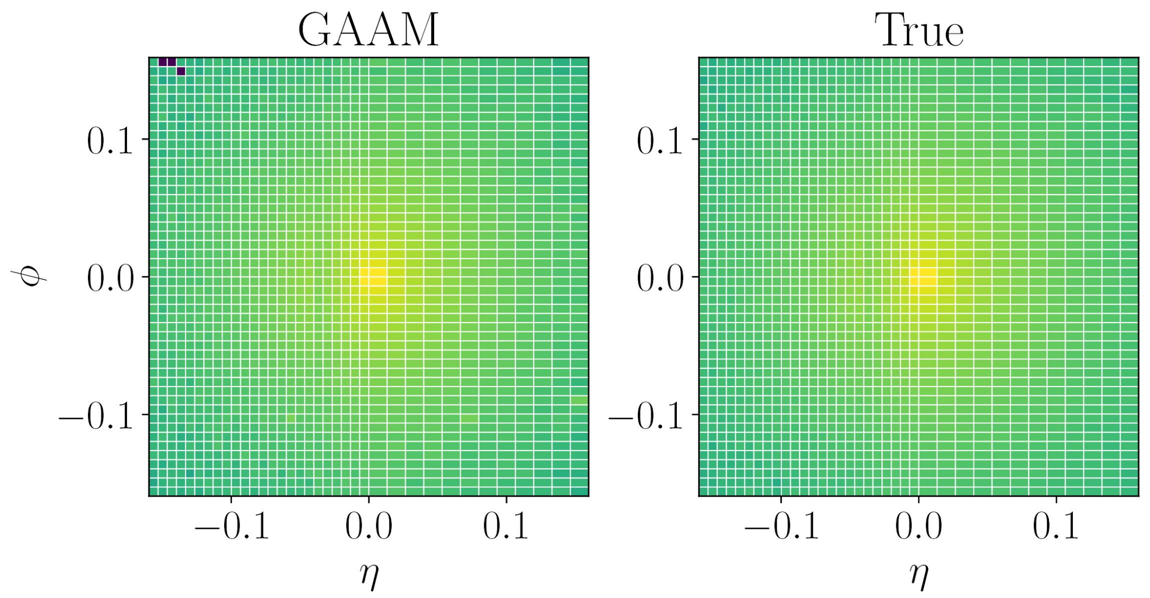


Image: arXiv:2305.11531

ML for Reconstruction

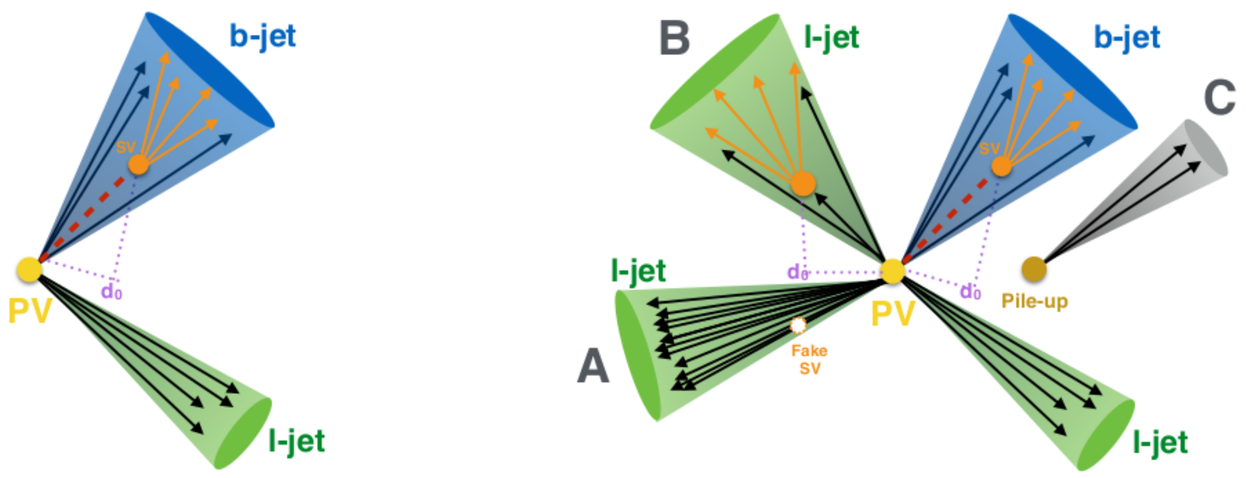


Image: USiegen group

ML for Signal vs Background

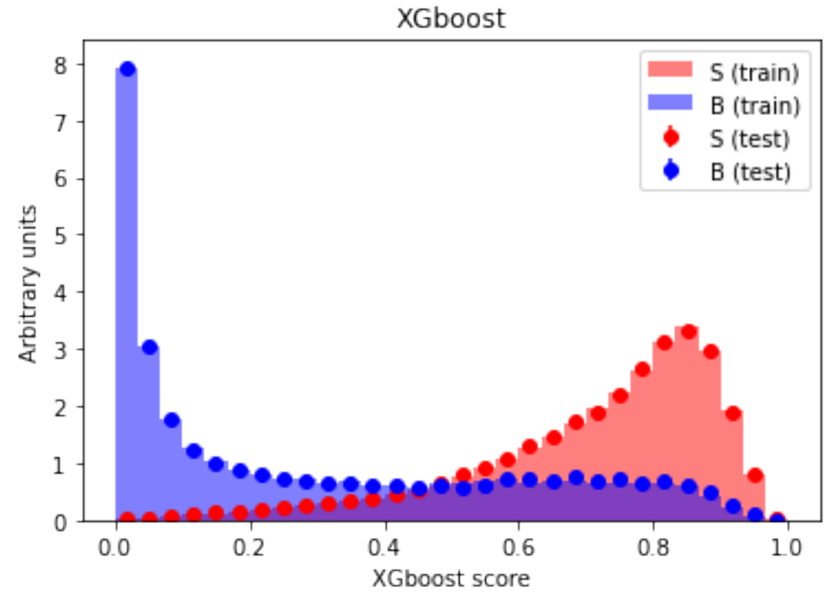
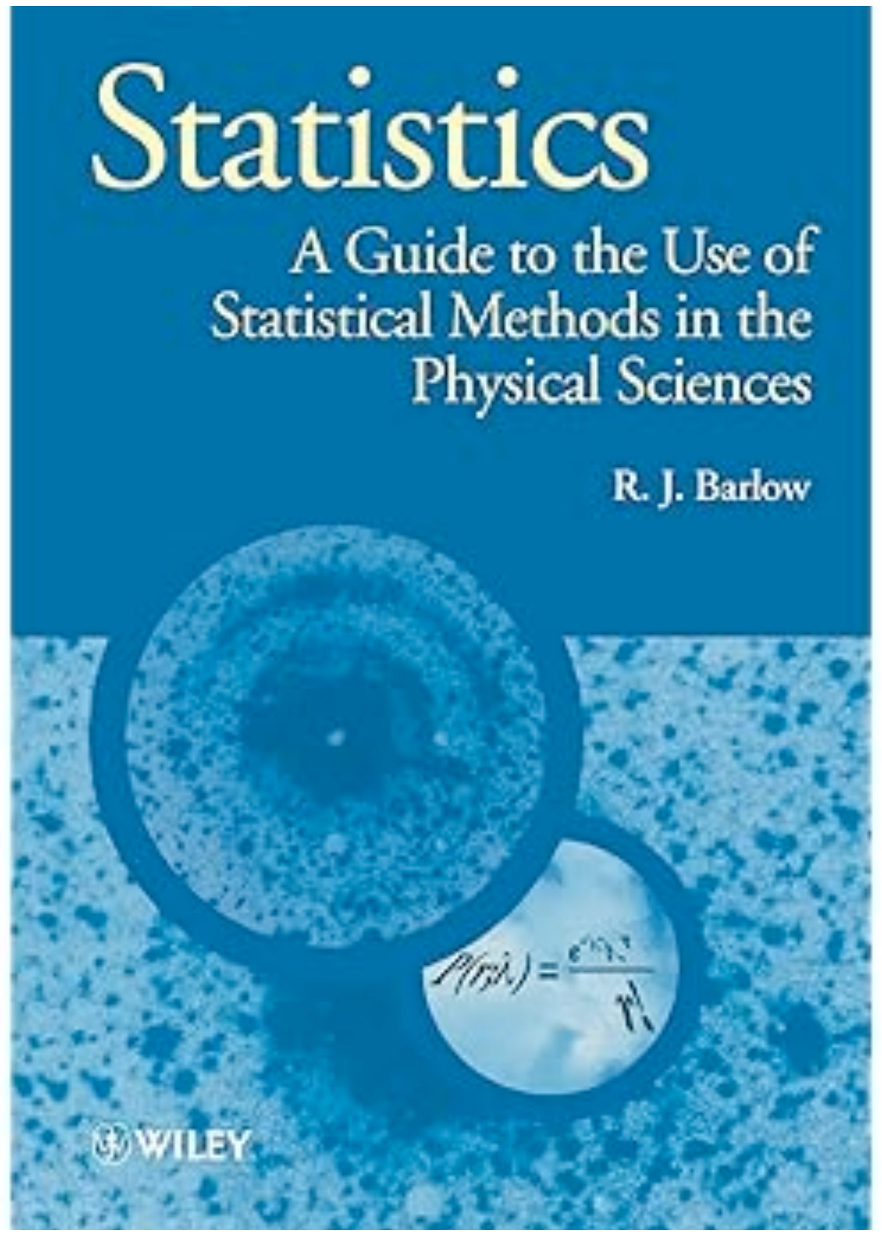
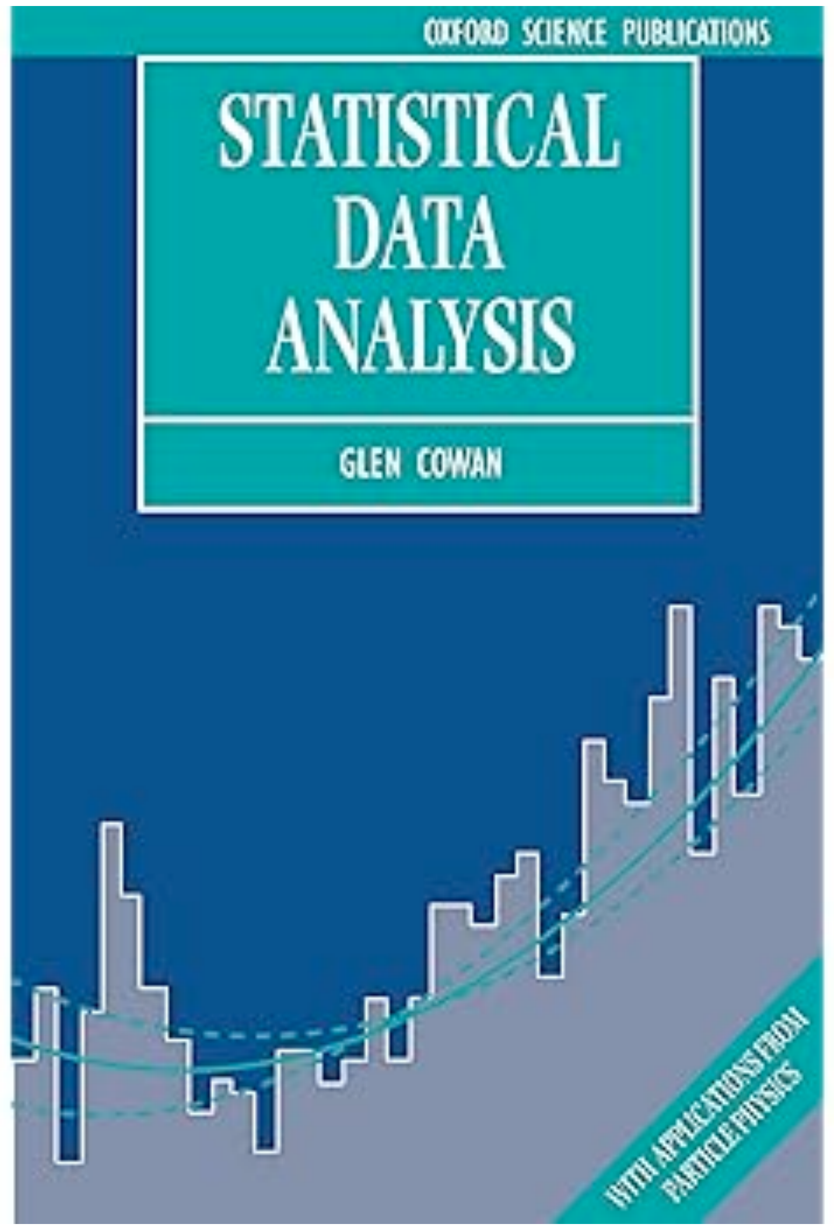


Image: ML4FP School

The final statistical inference step has remained the same



Publication date : November 1, 1993



Publication date : December 4, 1997

ML for Detector Simulation

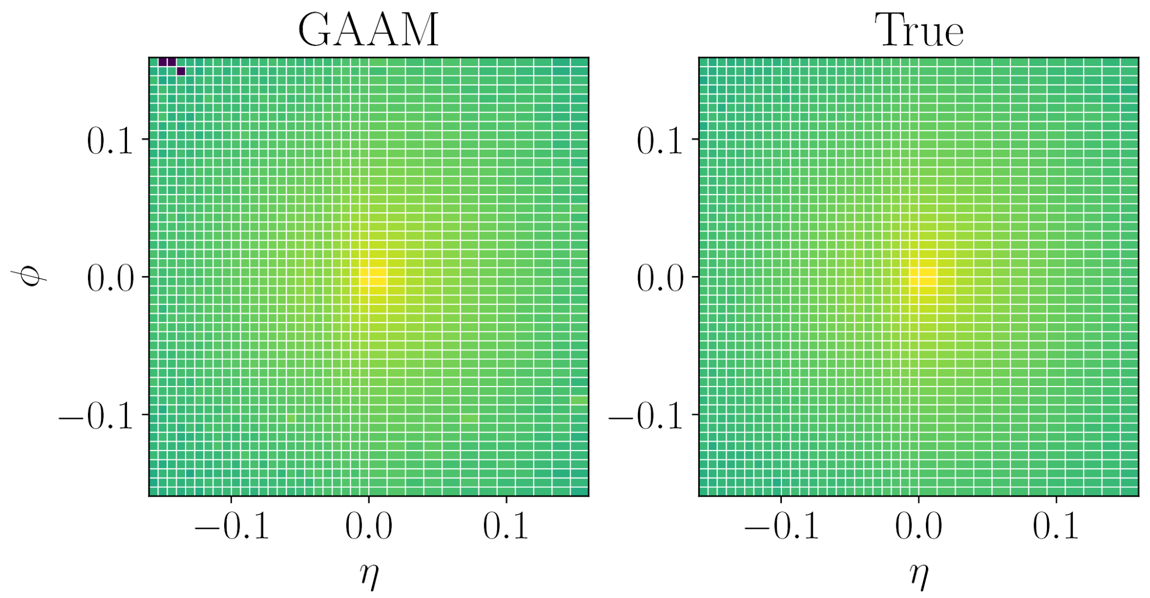


Image: arXiv:2305.11531

ML for Reconstruction

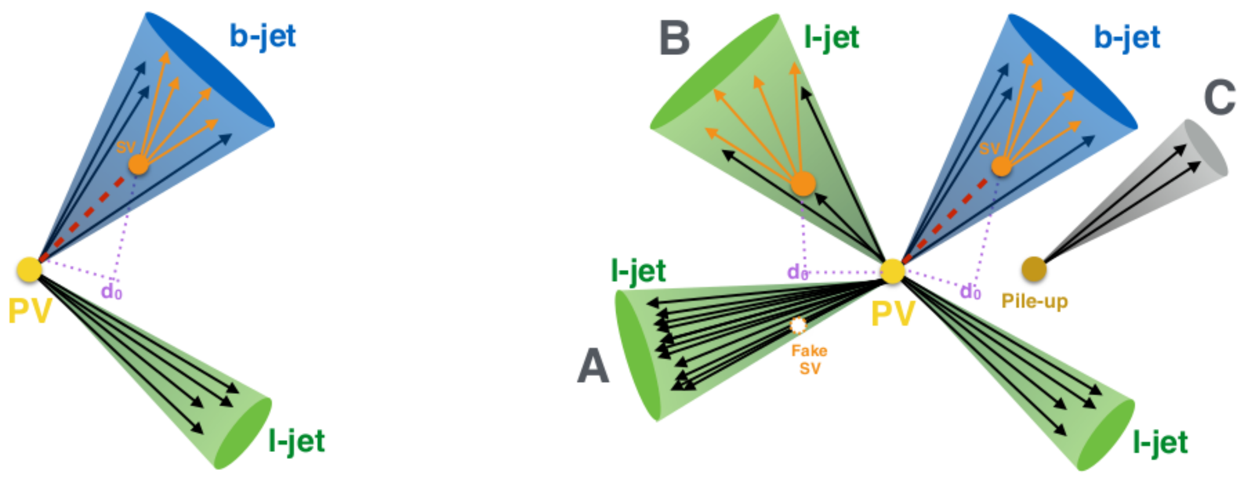


Image: USiegen group

ML for Signal vs Background

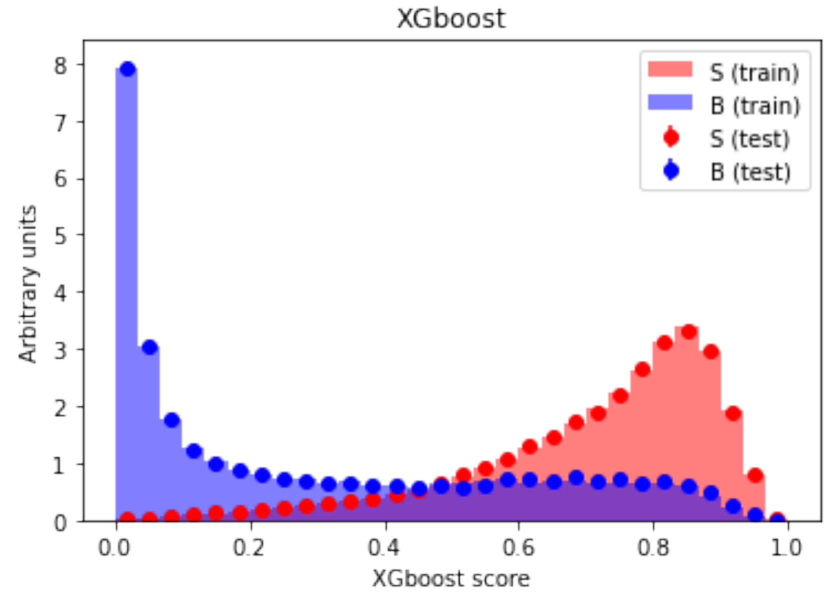
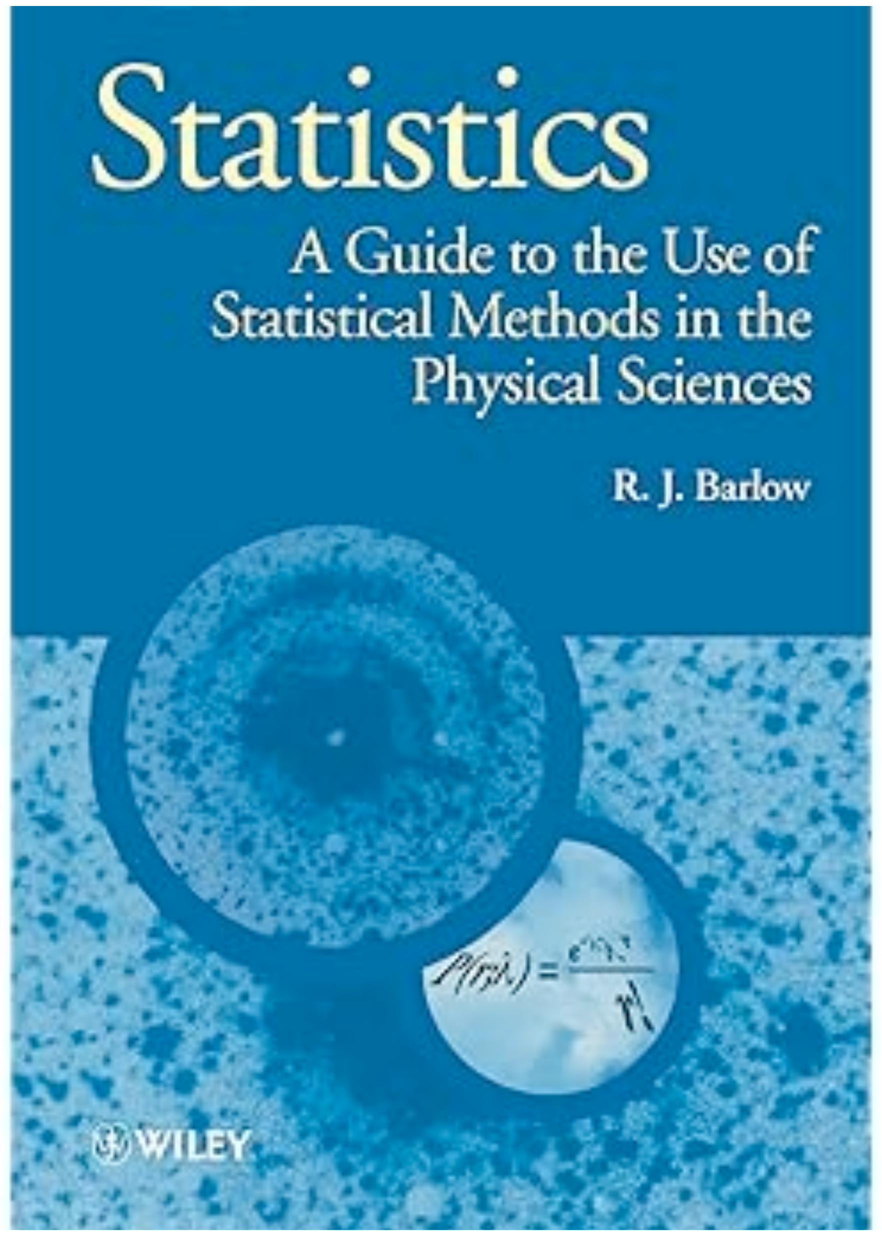


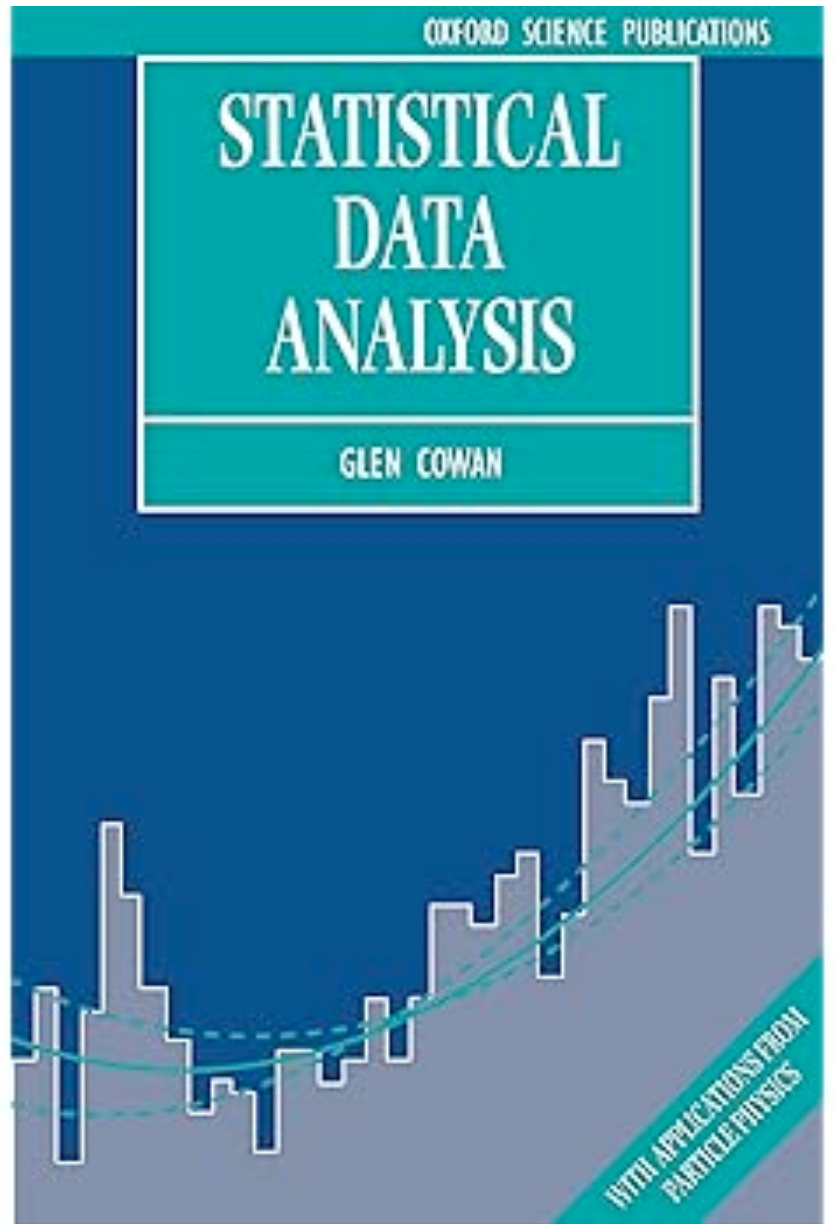
Image: ML4FP School

The final statistical inference step has remained the same



Publication date : November 1, 1993

Time to modernise?

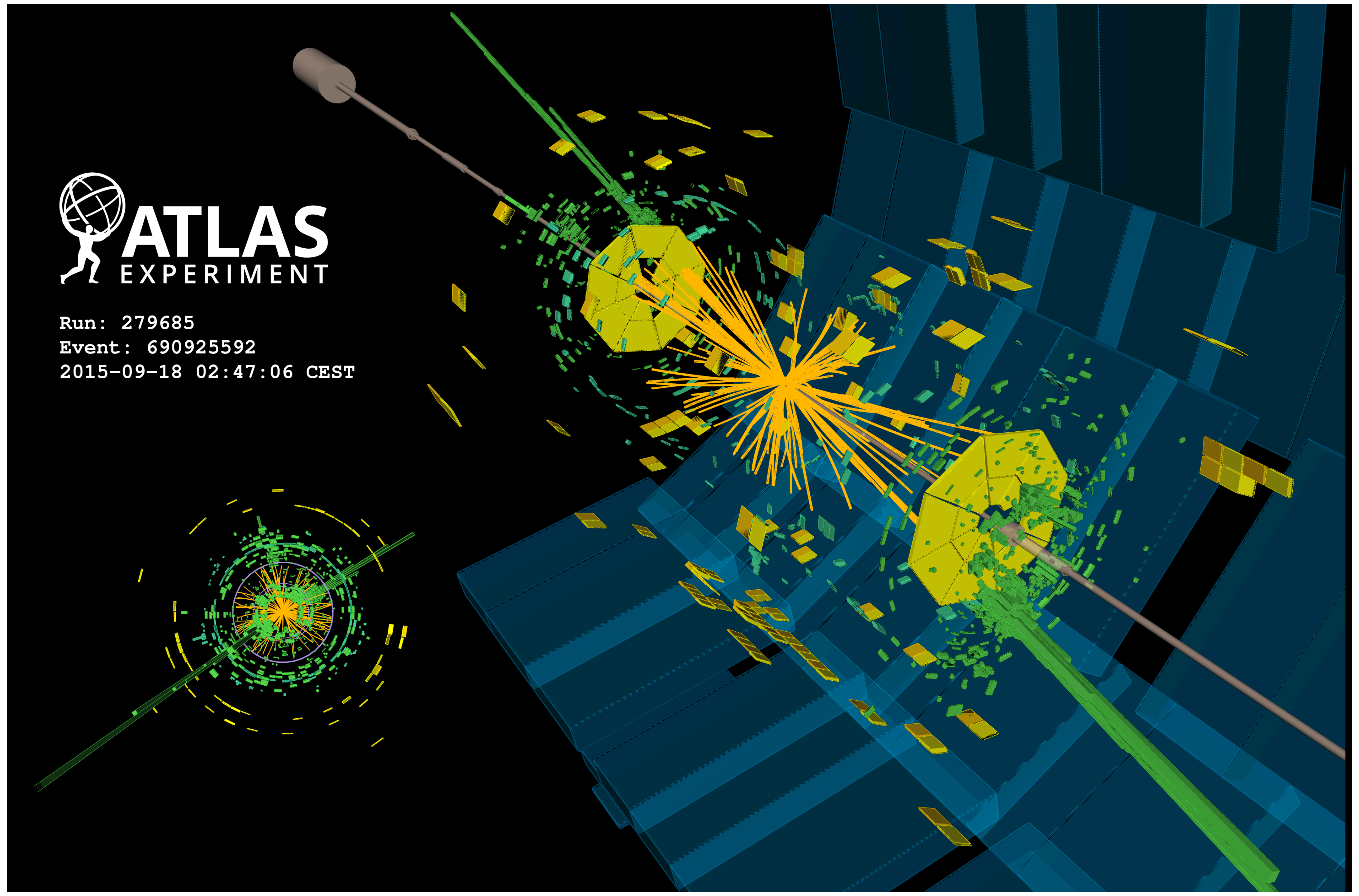


Publication date : December 4, 1997

Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

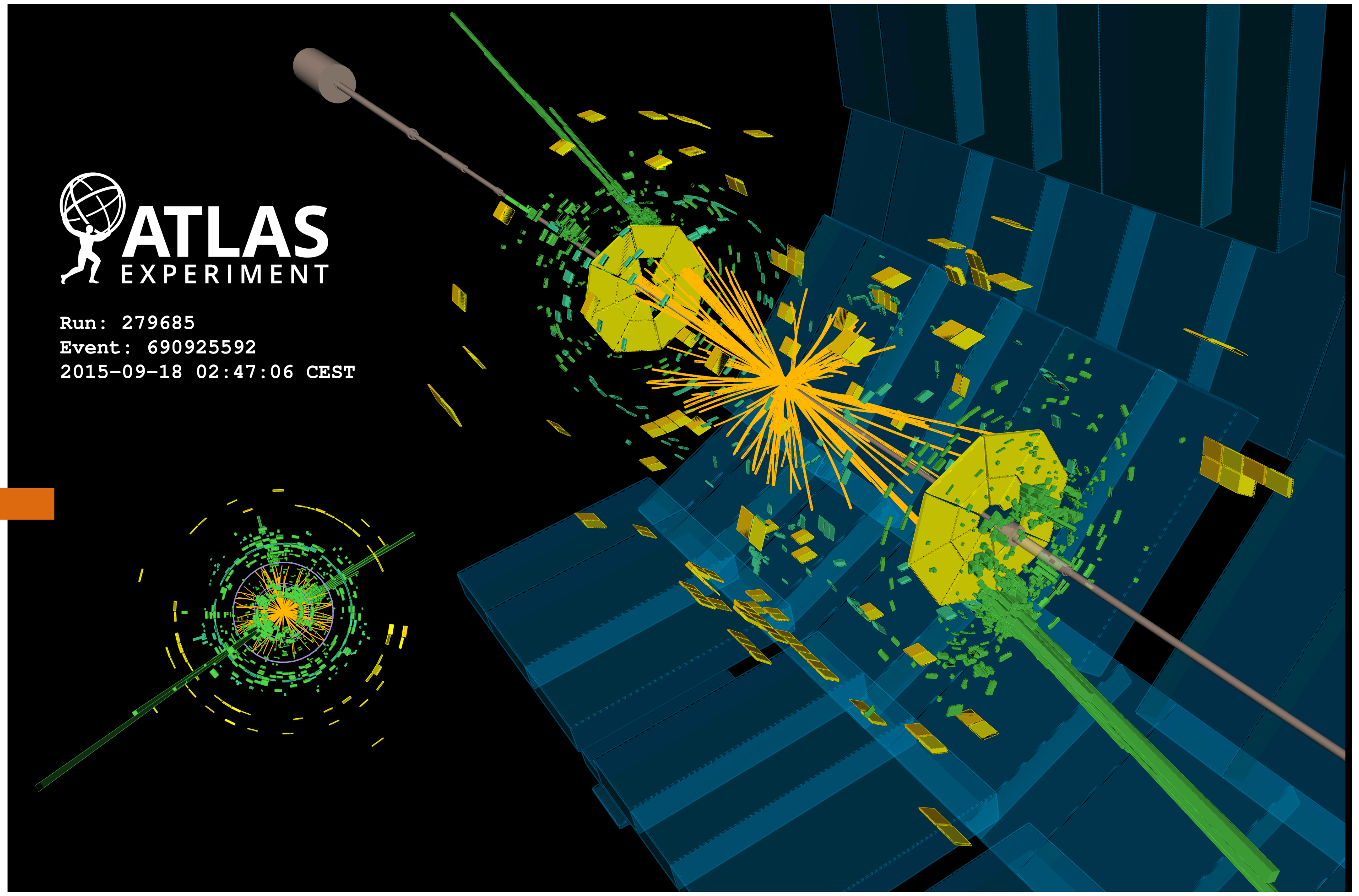


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

1 number



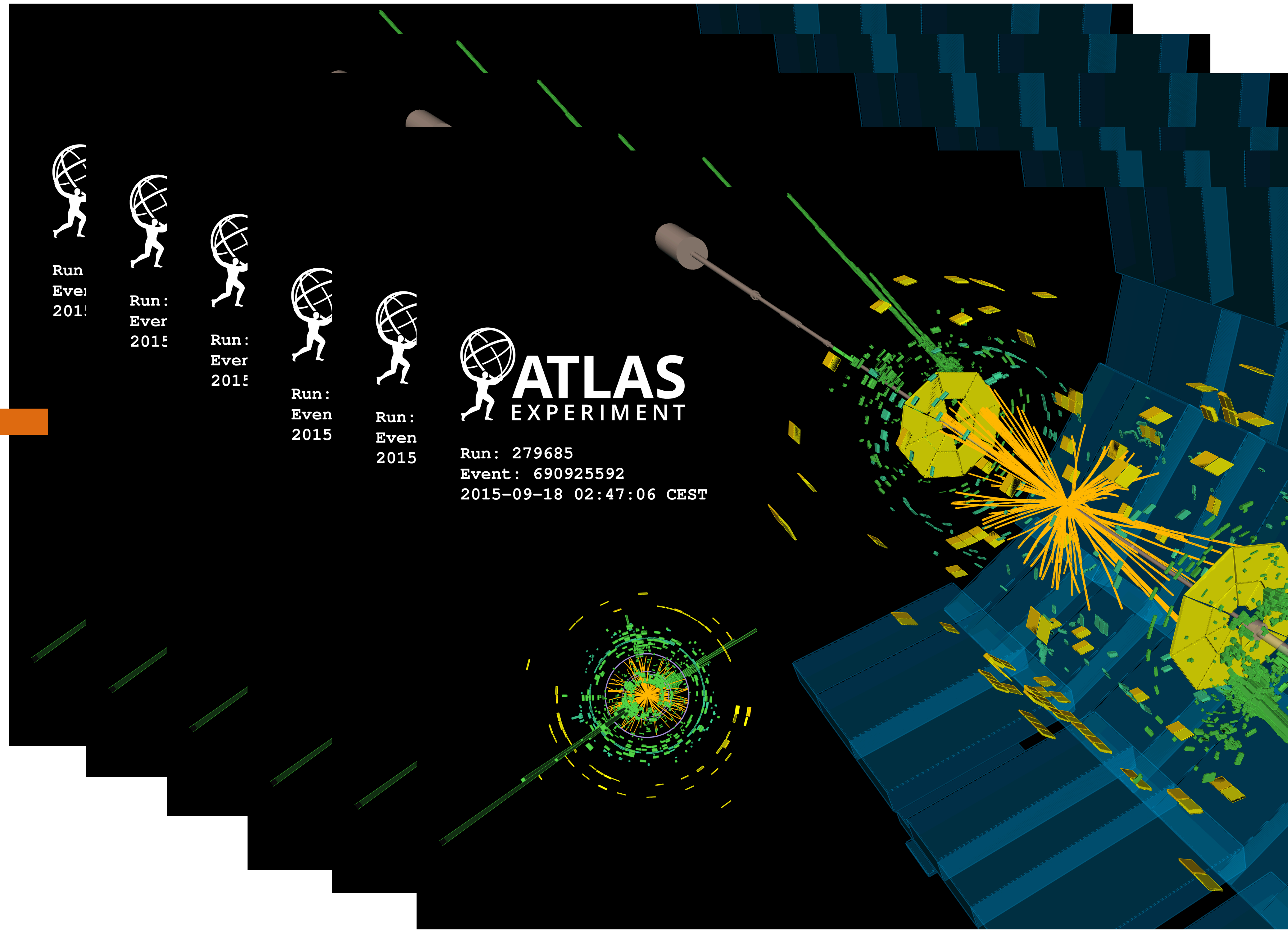
Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

1 number

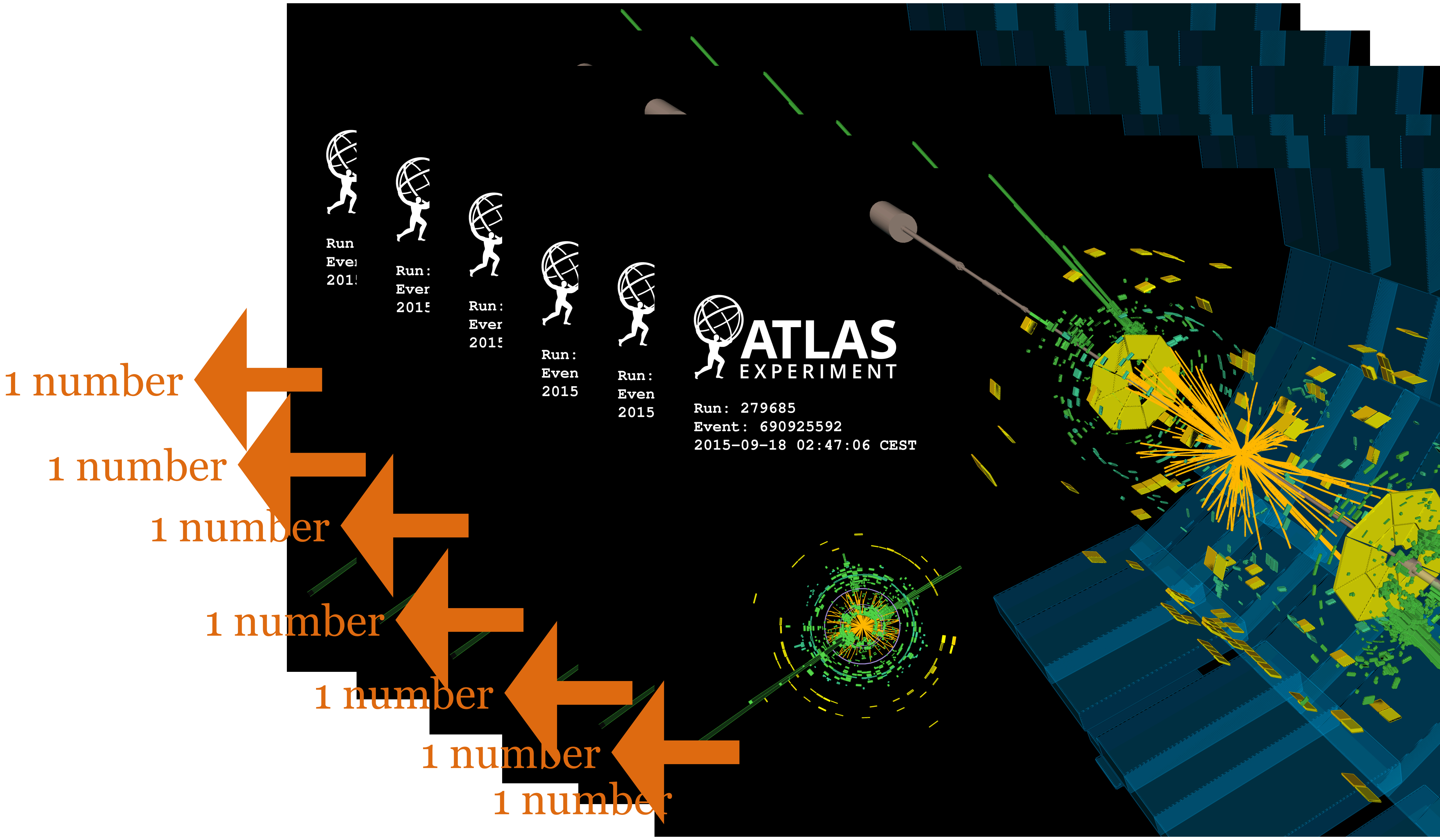


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

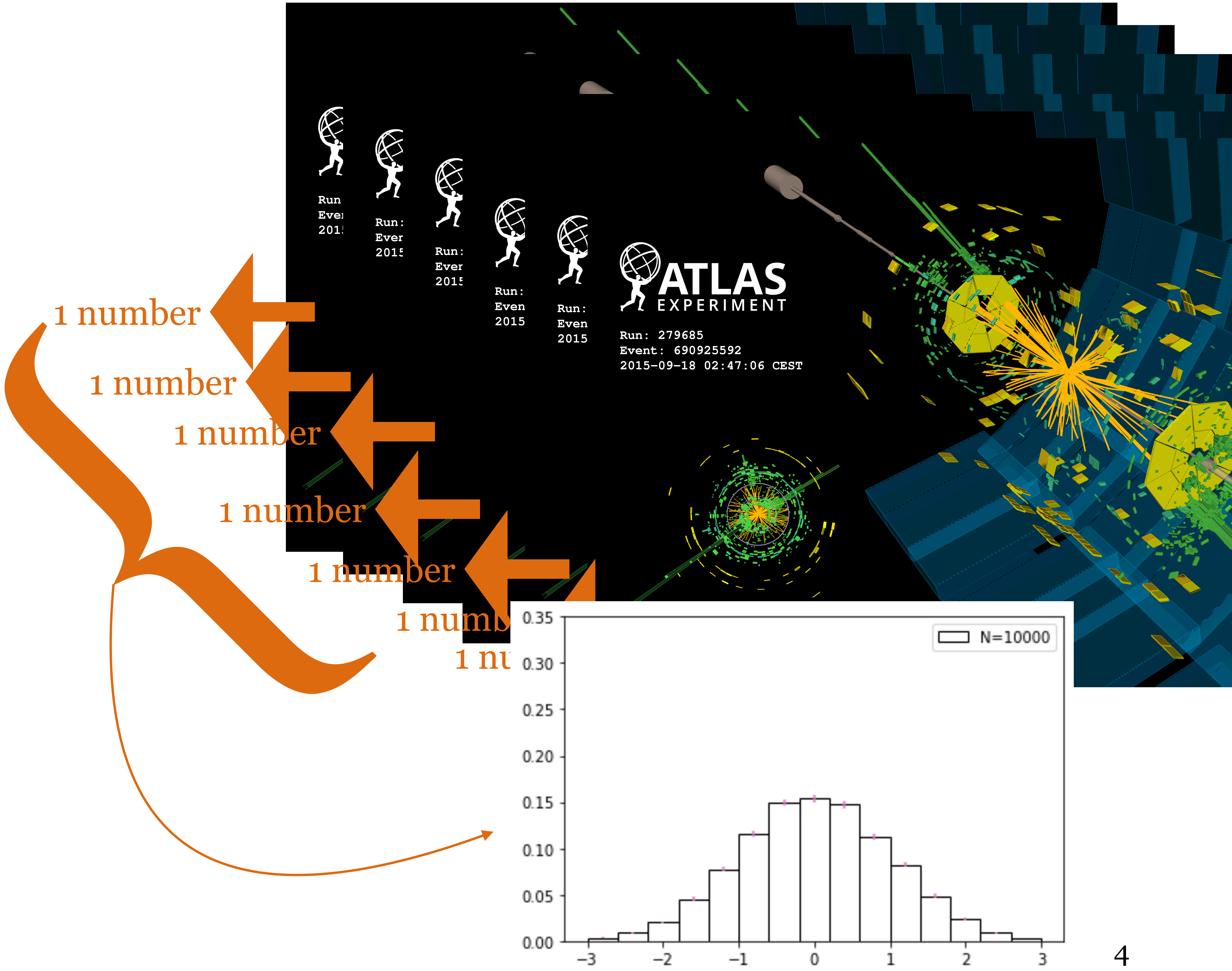


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

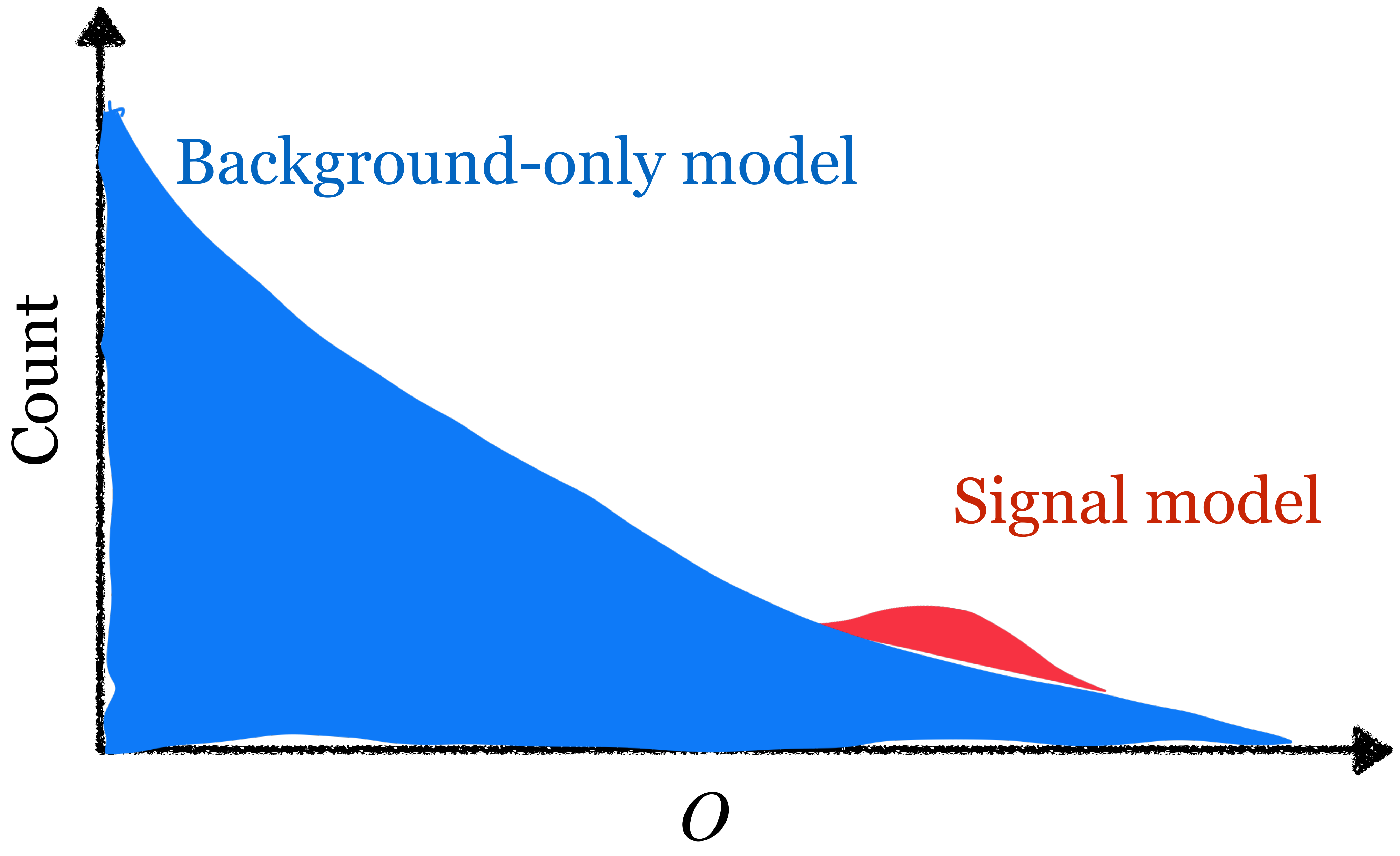
- Reconstruction pipeline, event selection
- Design one summary variable
  - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

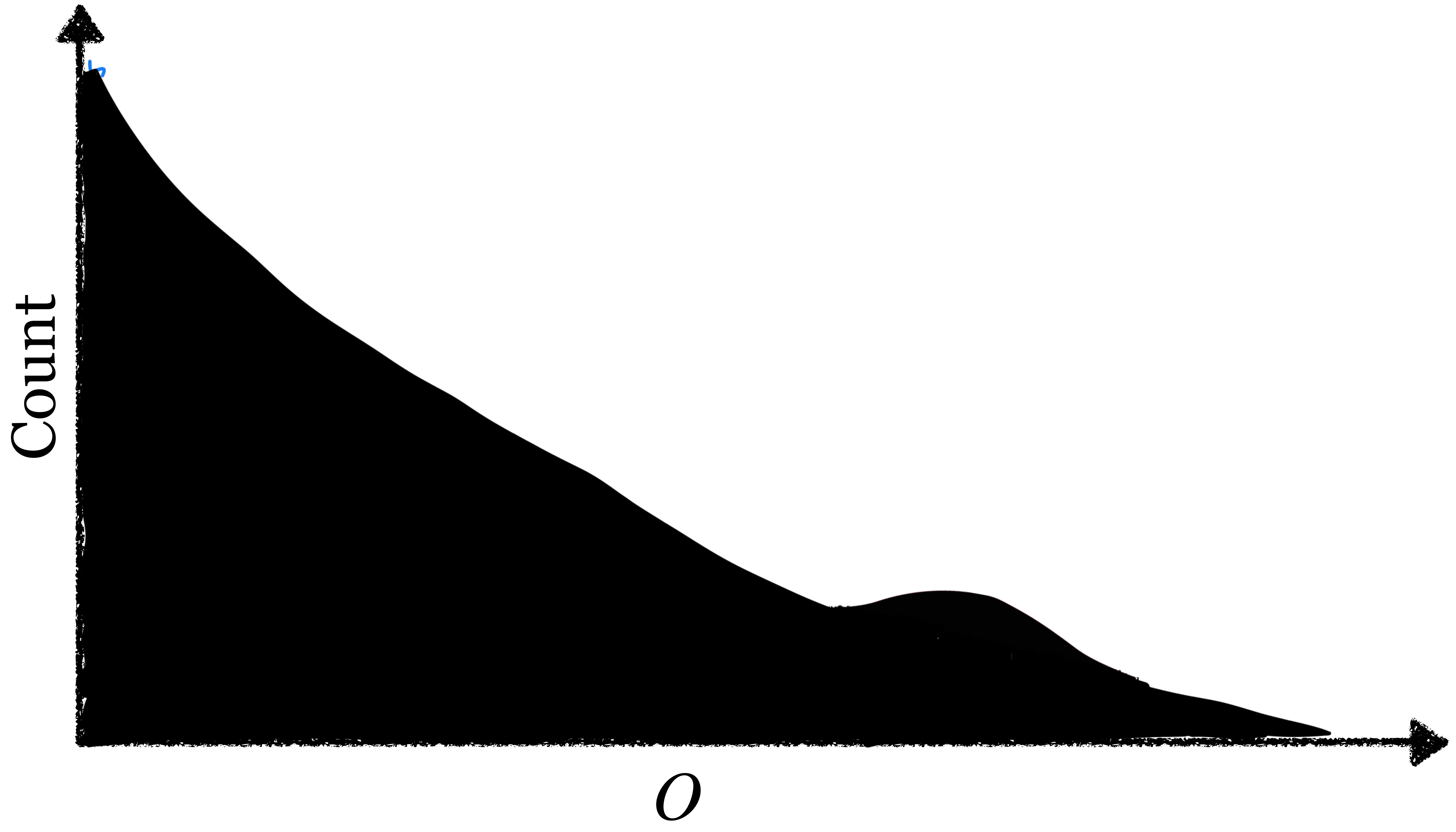


Probability density estimation with 1-D histogram

Theory Predictions

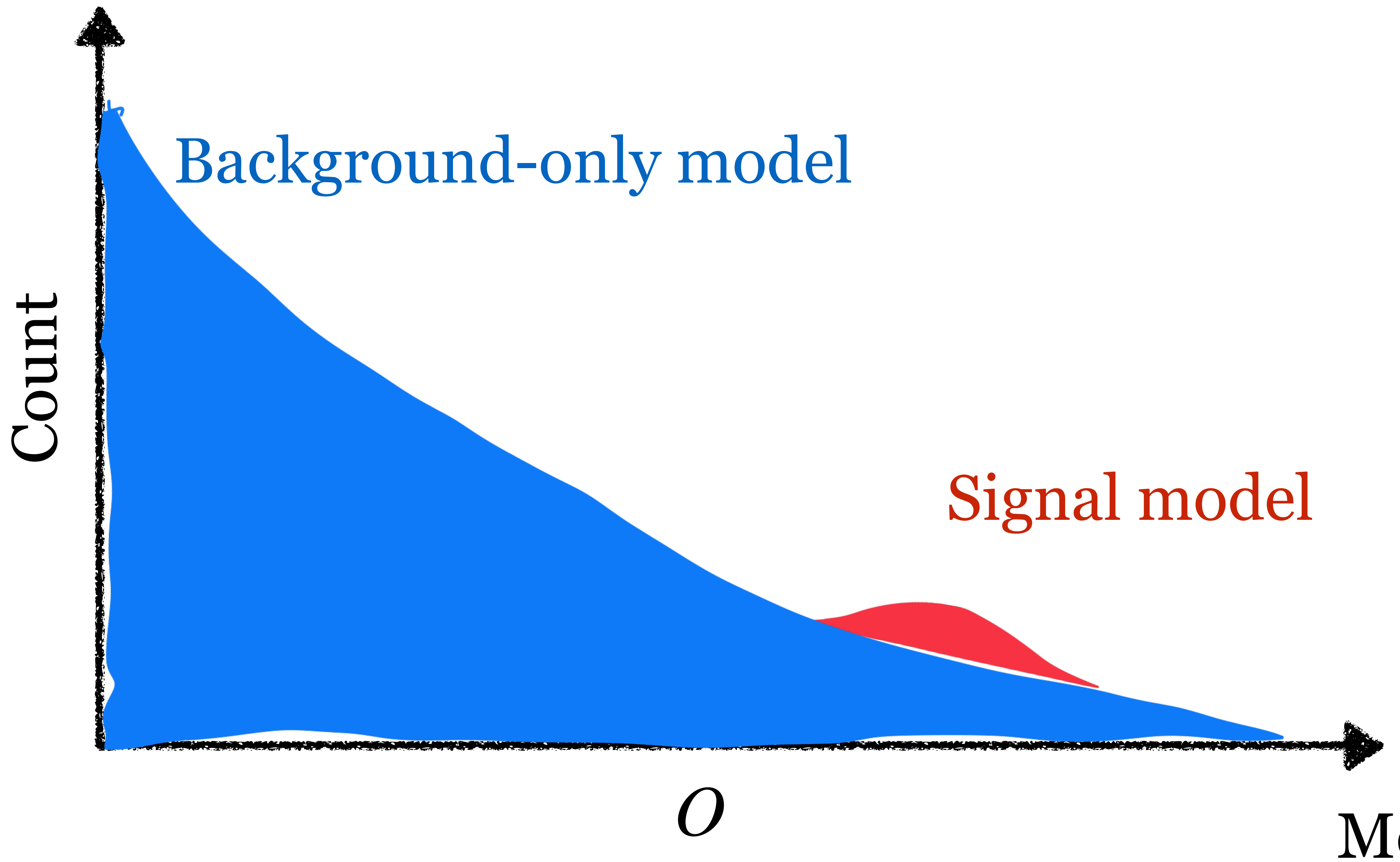


Data

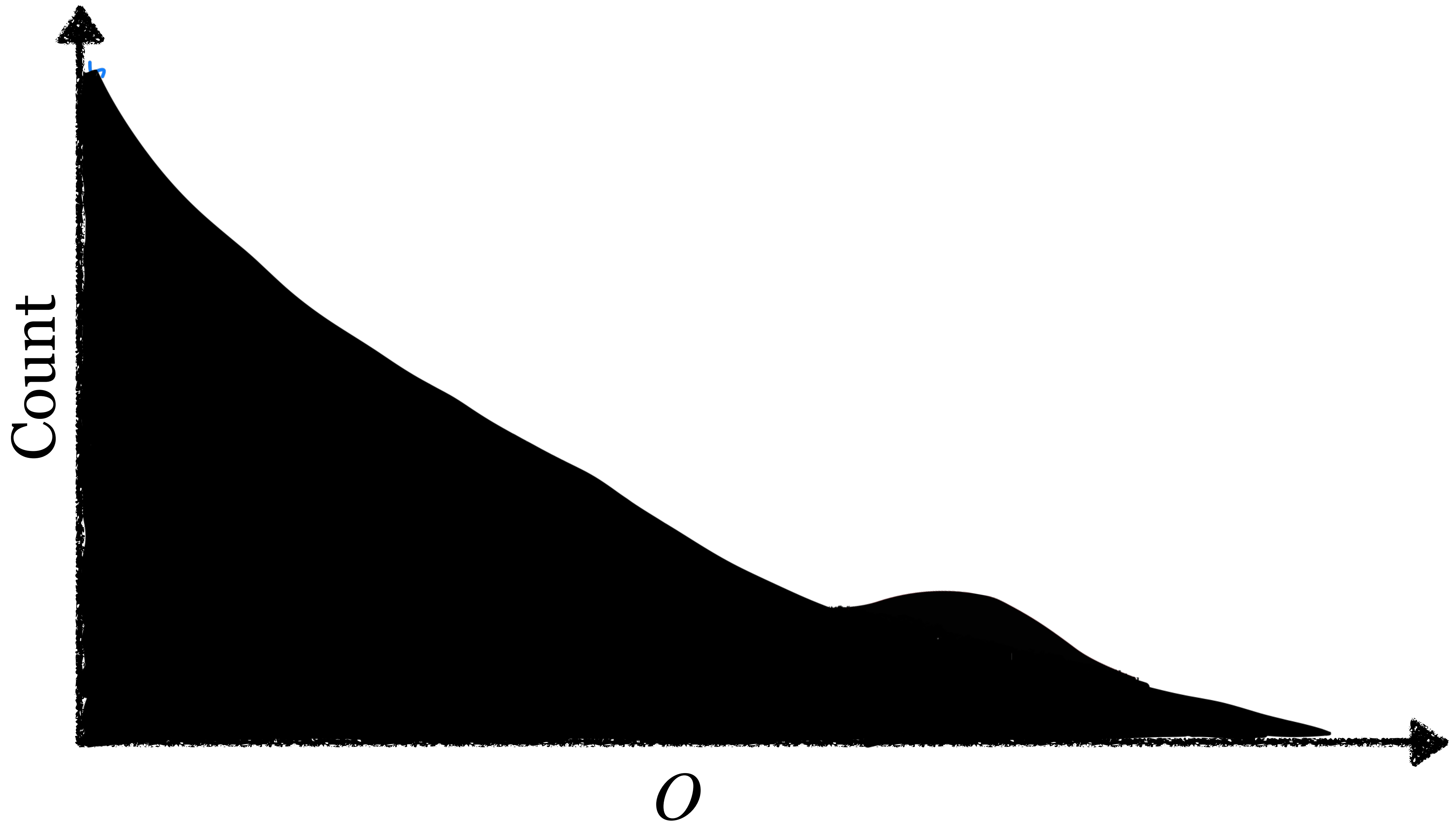


Probability density estimation with 1-D histogram

Theory Predictions



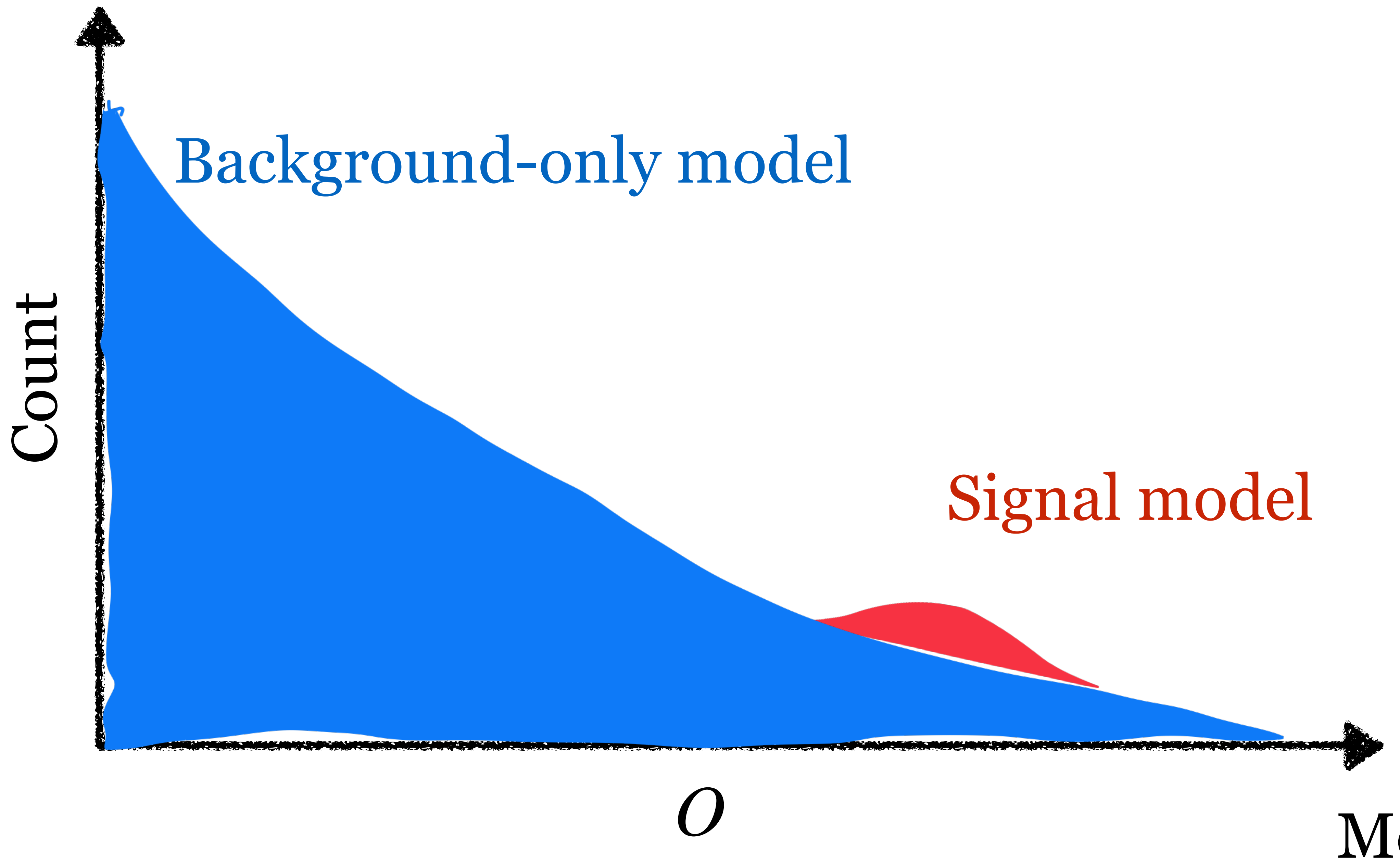
Data



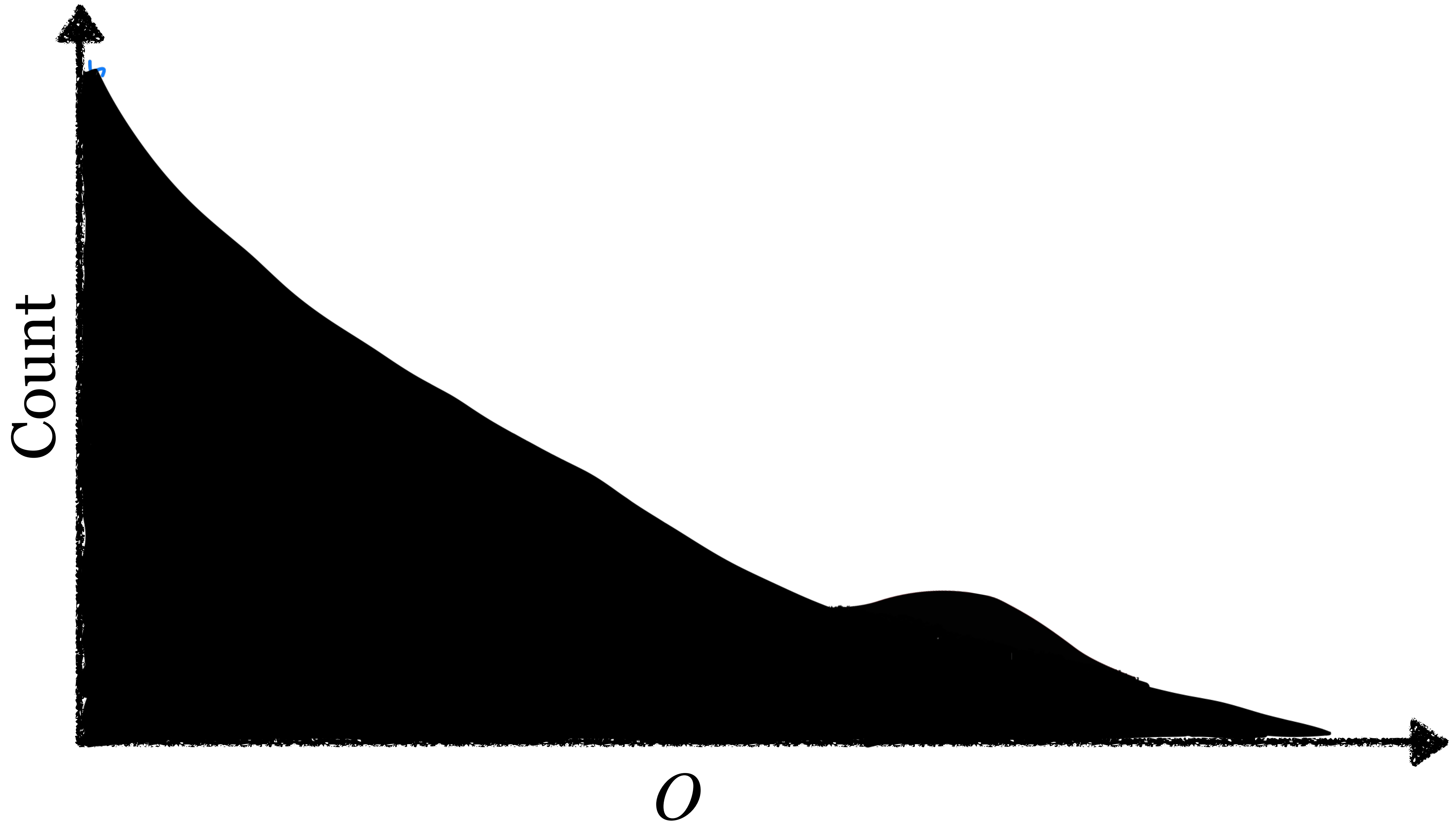
Measure ‘signal strength’ μ

Probability density estimation with 1-D histogram

Theory Predictions



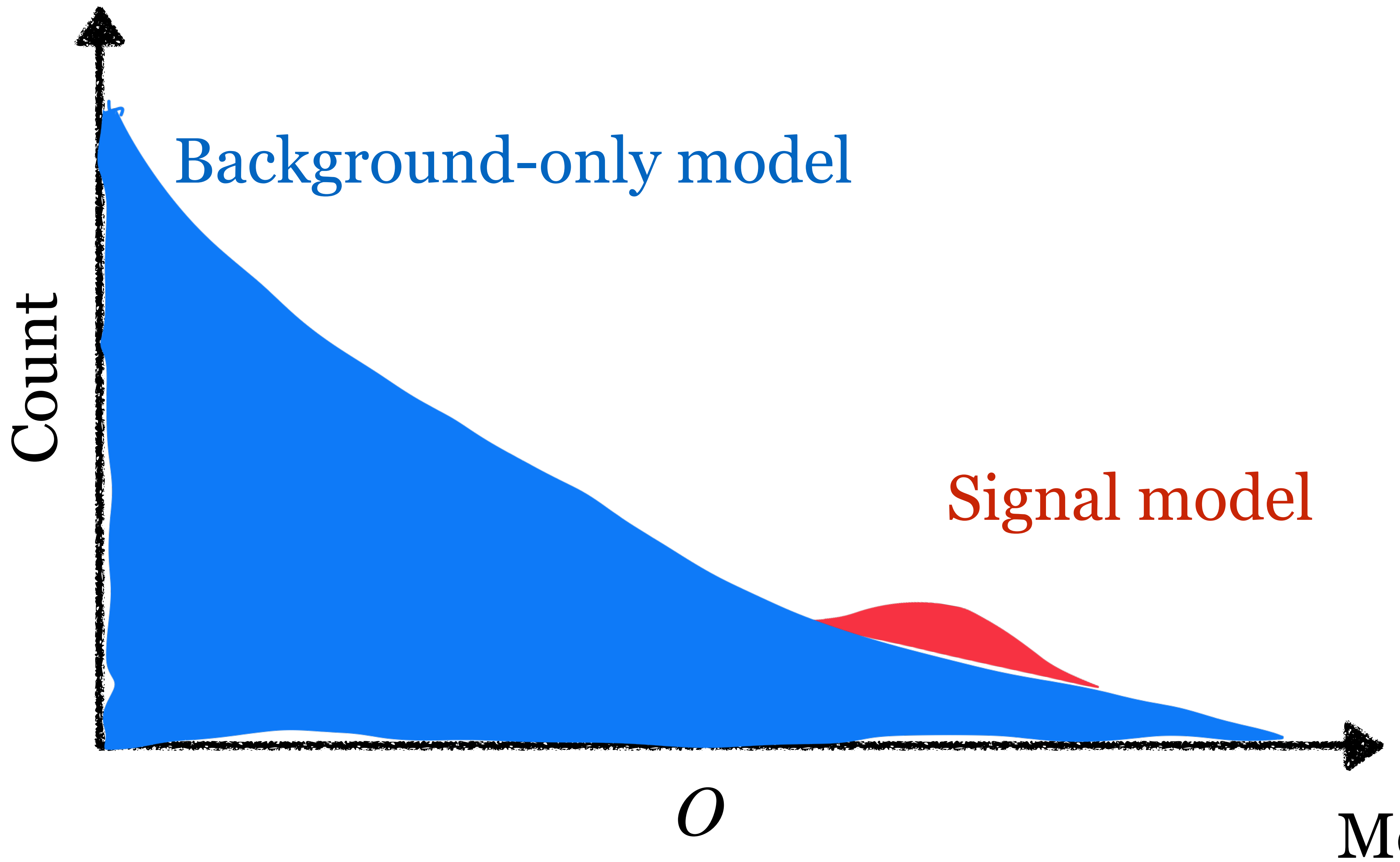
Data



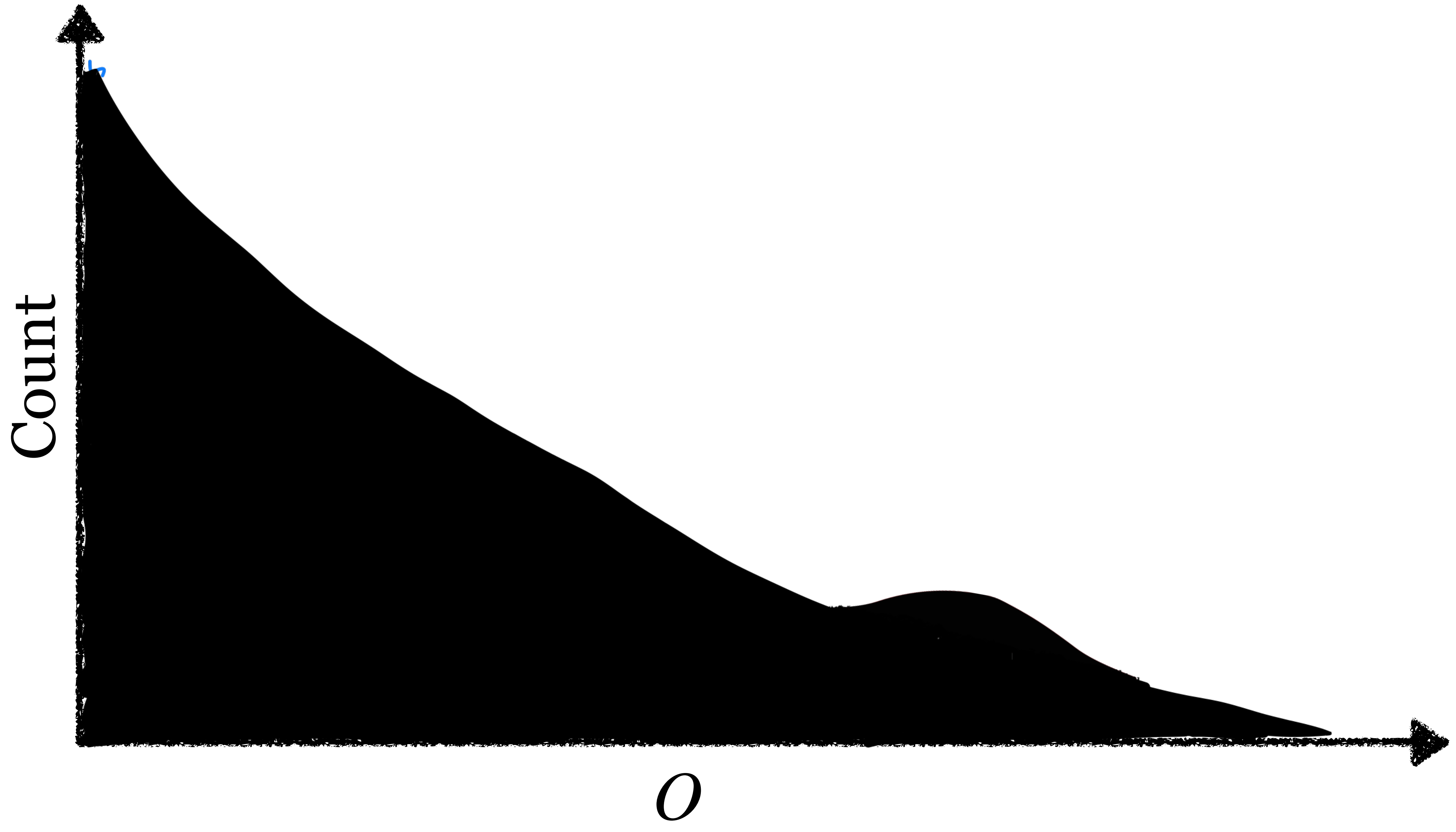
Measure ‘signal strength’ μ

Probability density estimation with 1-D histogram

Theory Predictions



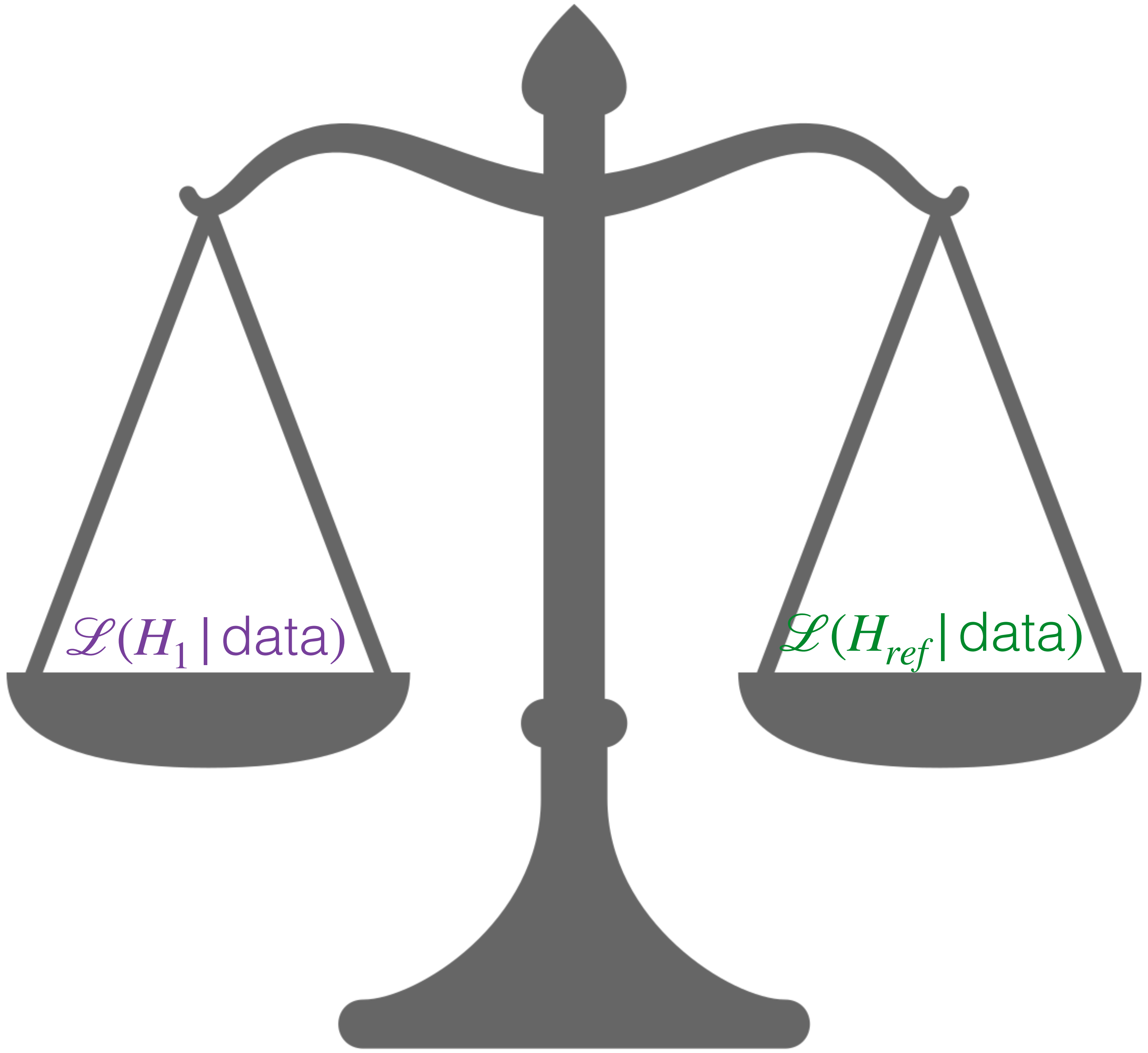
Data



Measure 'signal strength' μ

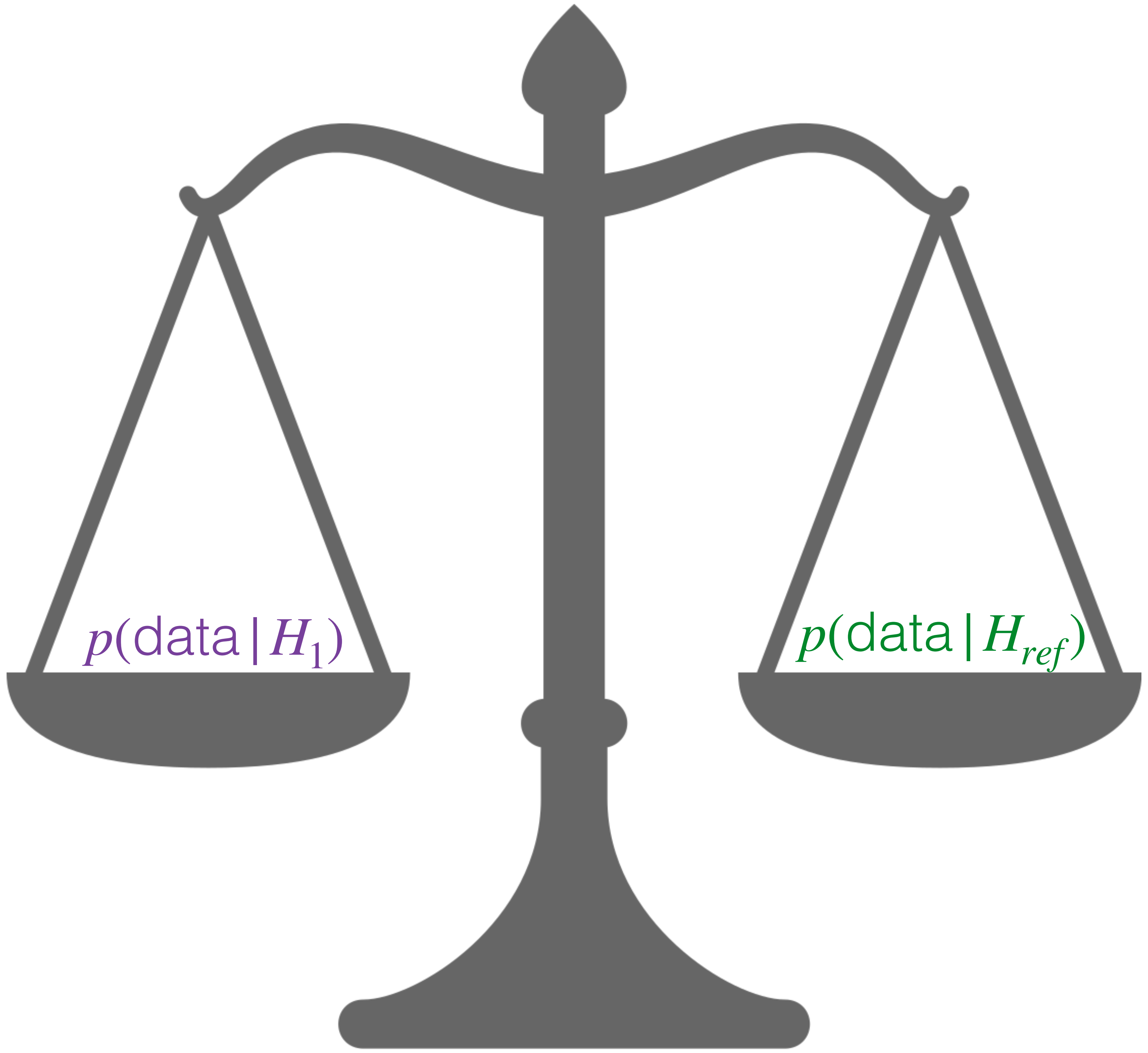
With histograms we can ask "Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?"

Hypothesis tests



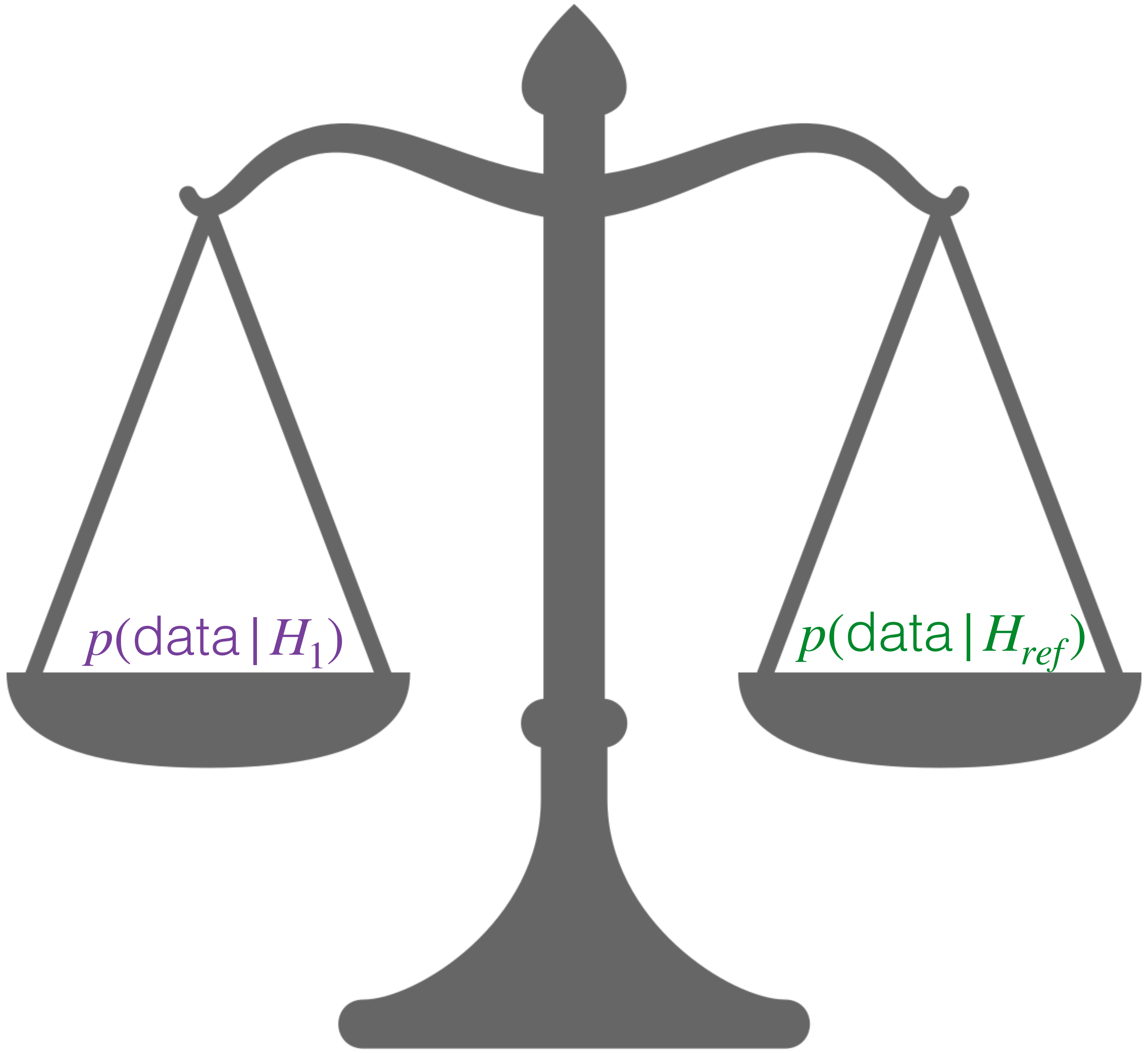
Hypothesis tests

$$\mathscr{L}(H_1 \mid \text{data}) = p(\text{data} \mid H_1)$$



Hypothesis tests

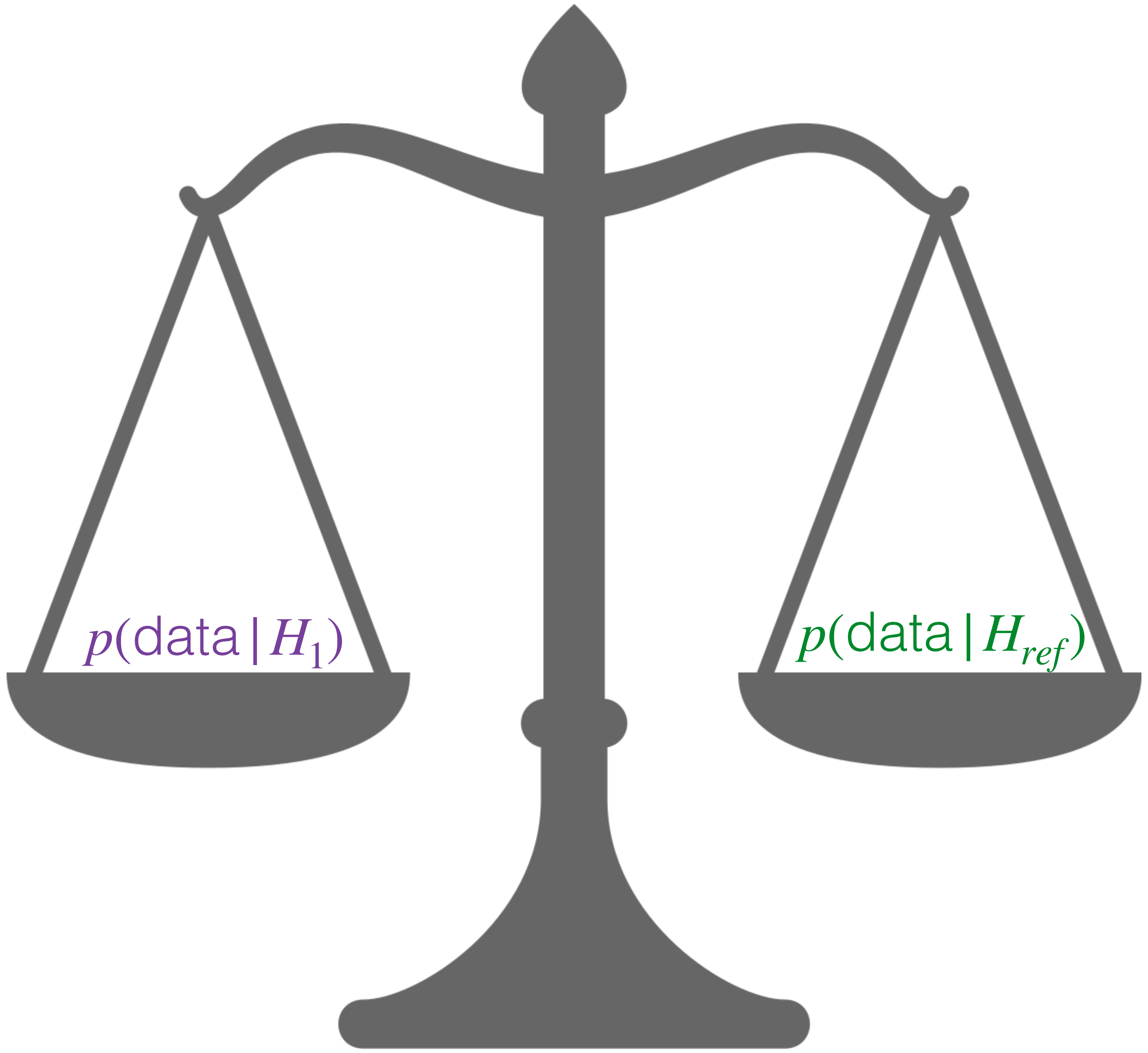
$$\mathscr{L}(H_1\,|\,\text{data}) = p(\text{data}\,|\,H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

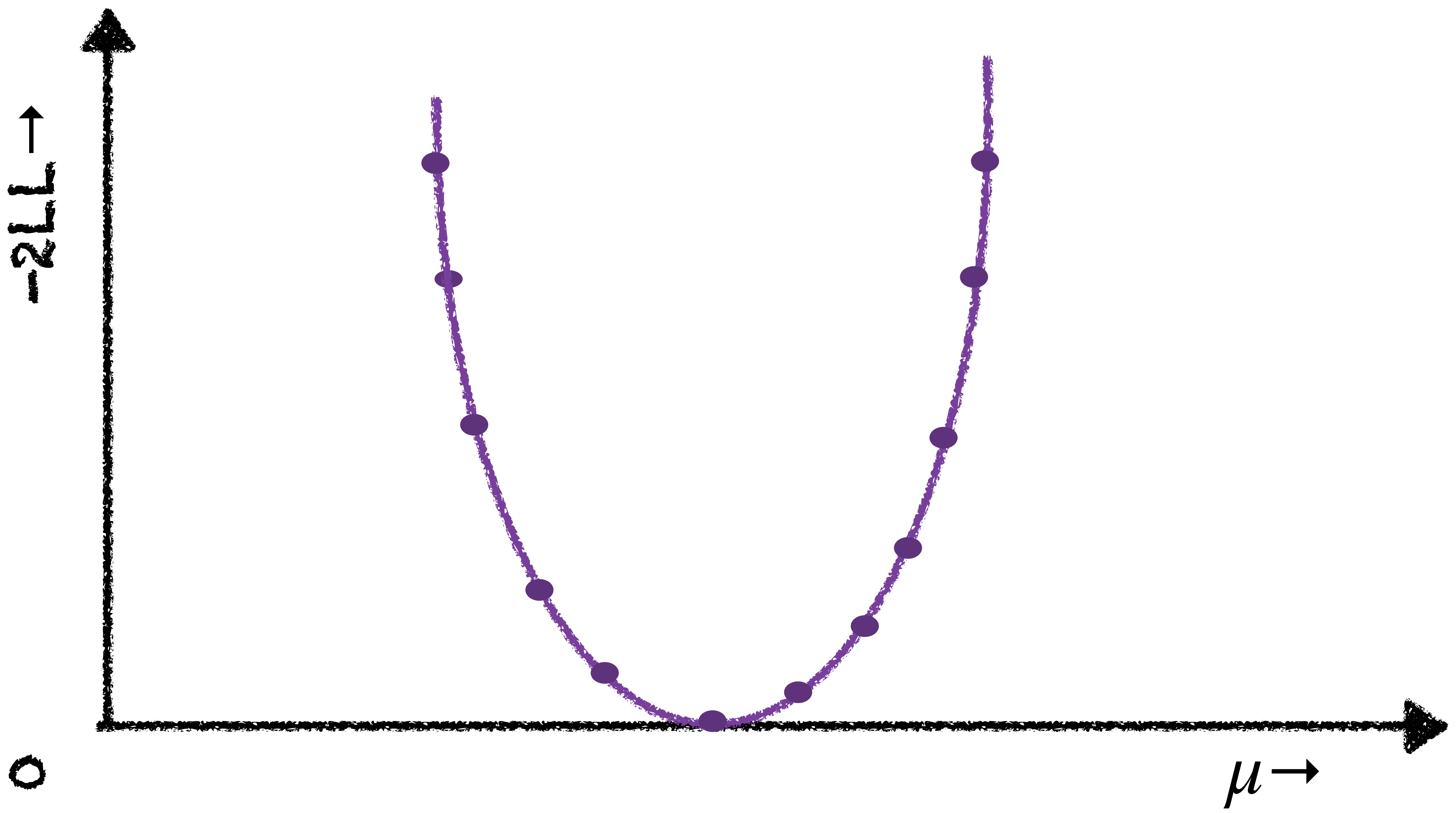
Hypothesis tests

$$\mathscr{L}(H_1 \mid \text{data}) = p(\text{data} \mid H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

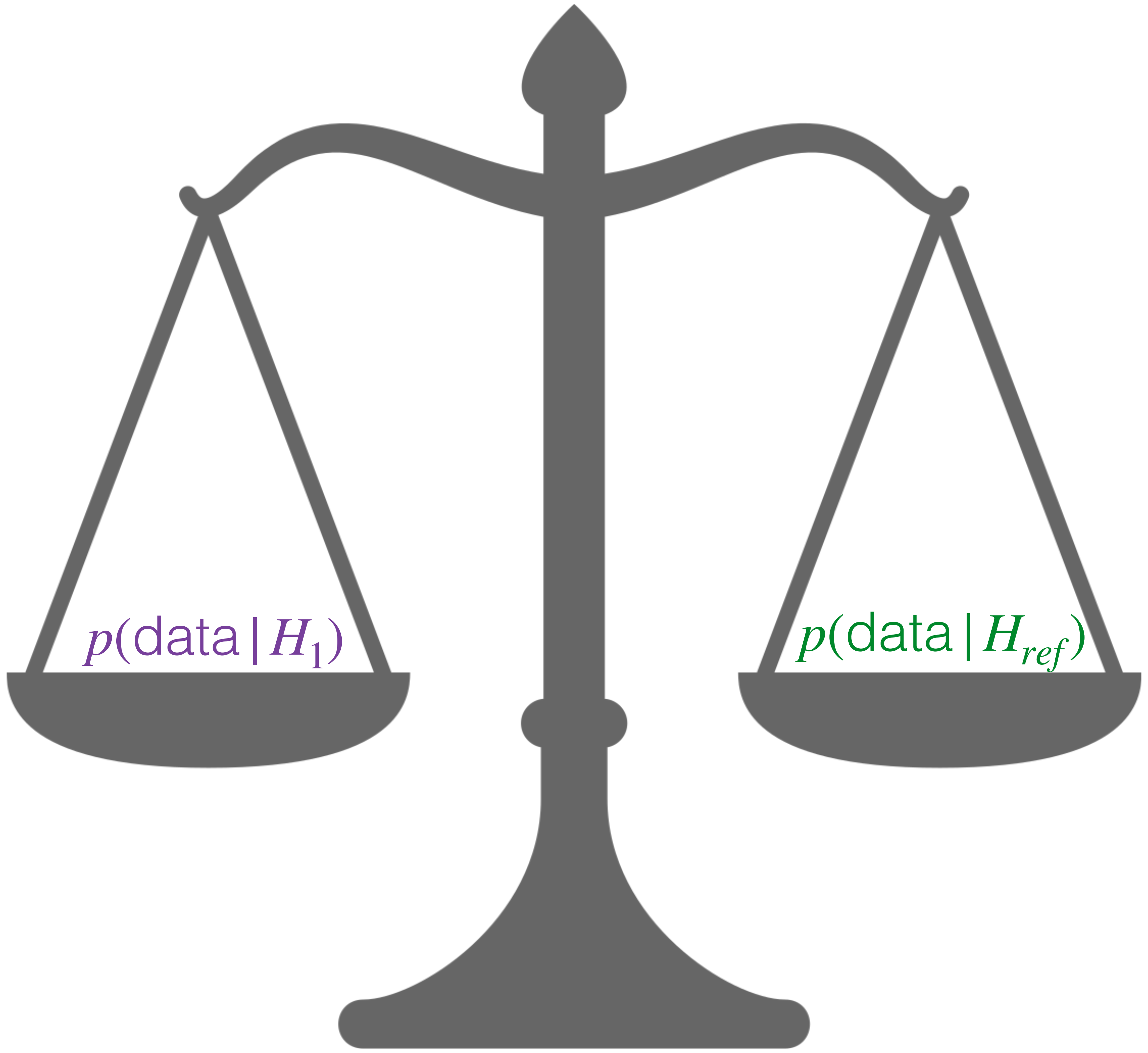
Invert the test for confidence intervals



Parameter estimation (infinite hypotheses)

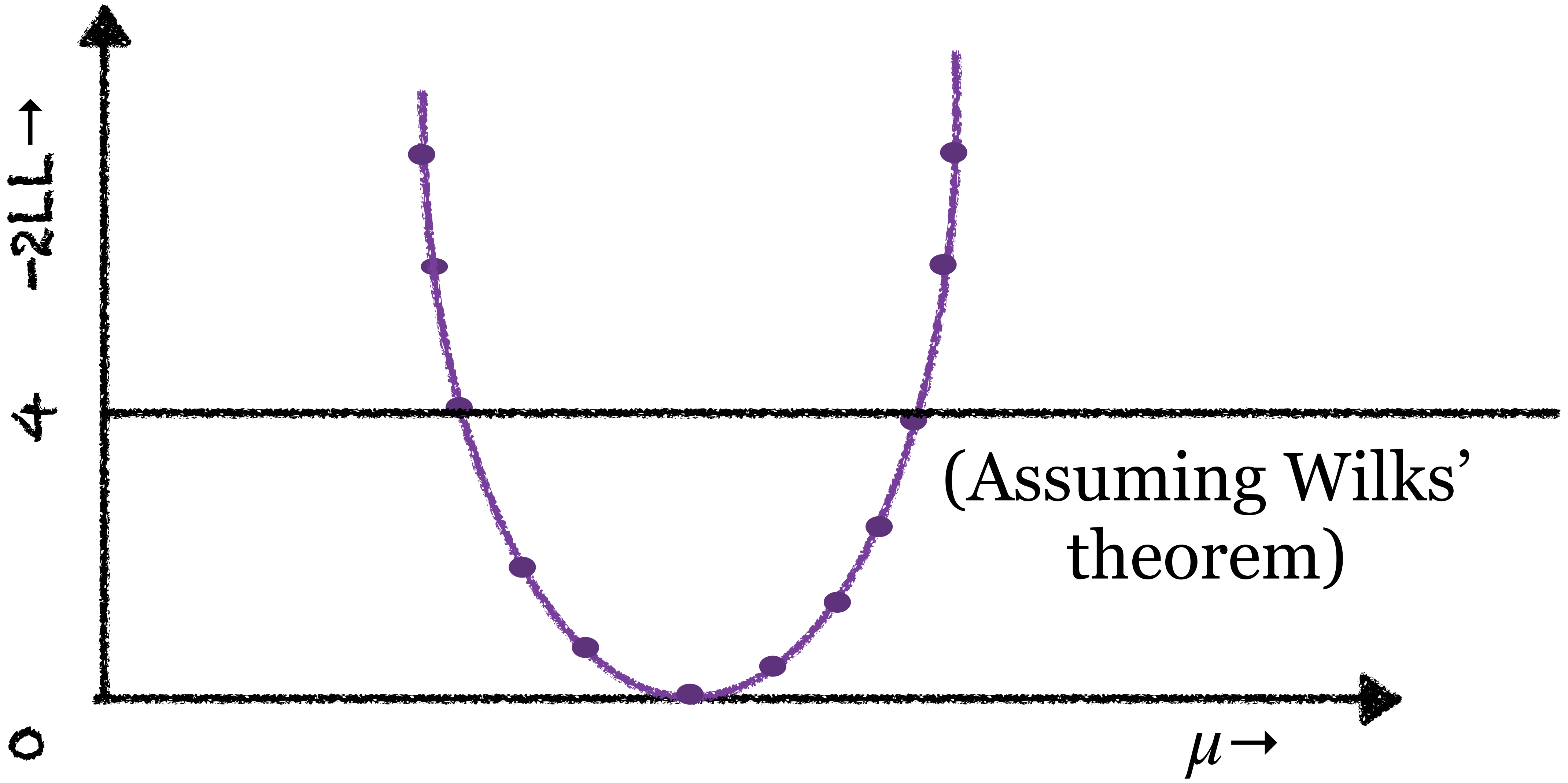
Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

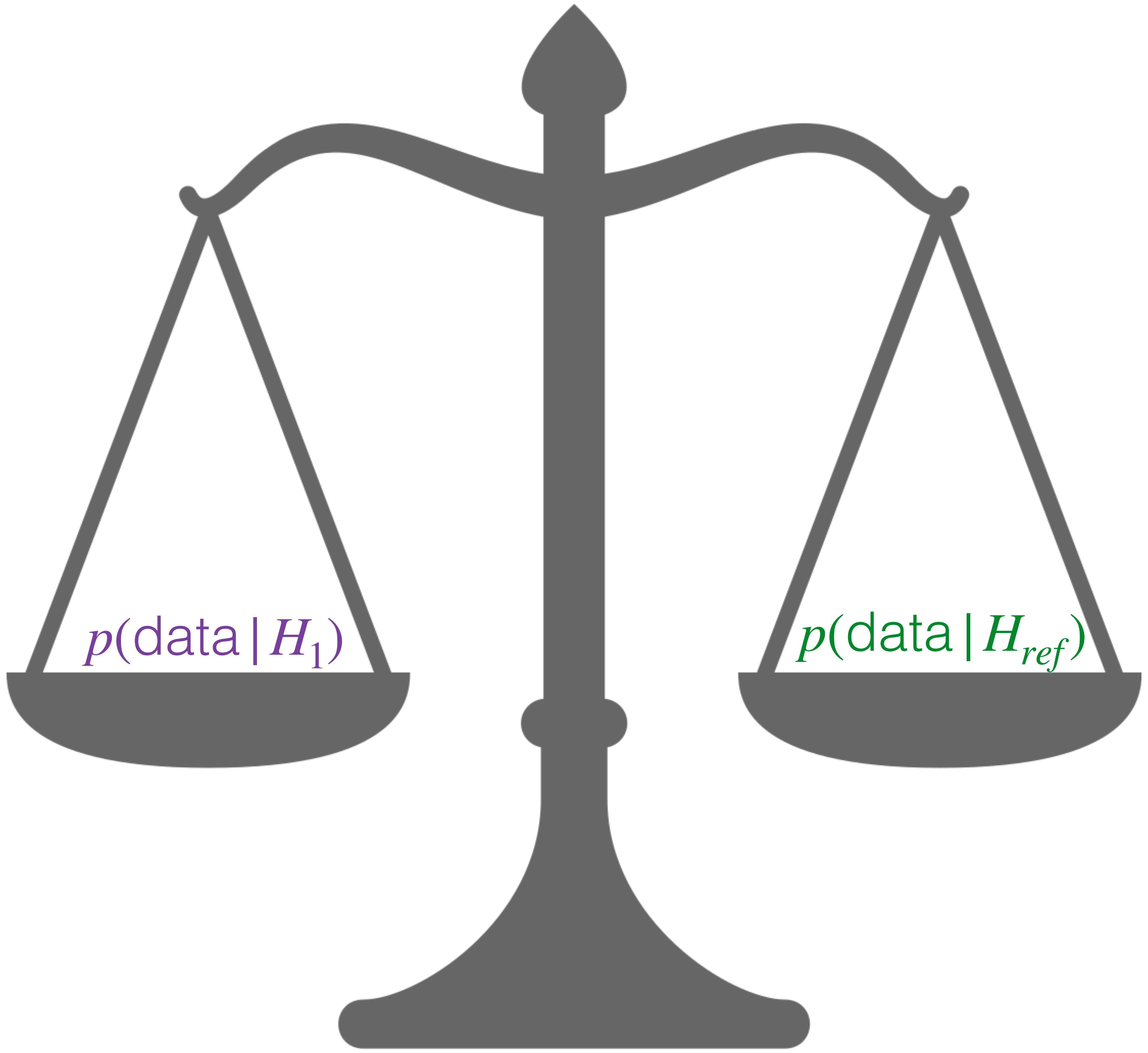
Invert the test for confidence intervals



Parameter estimation (infinite hypotheses)

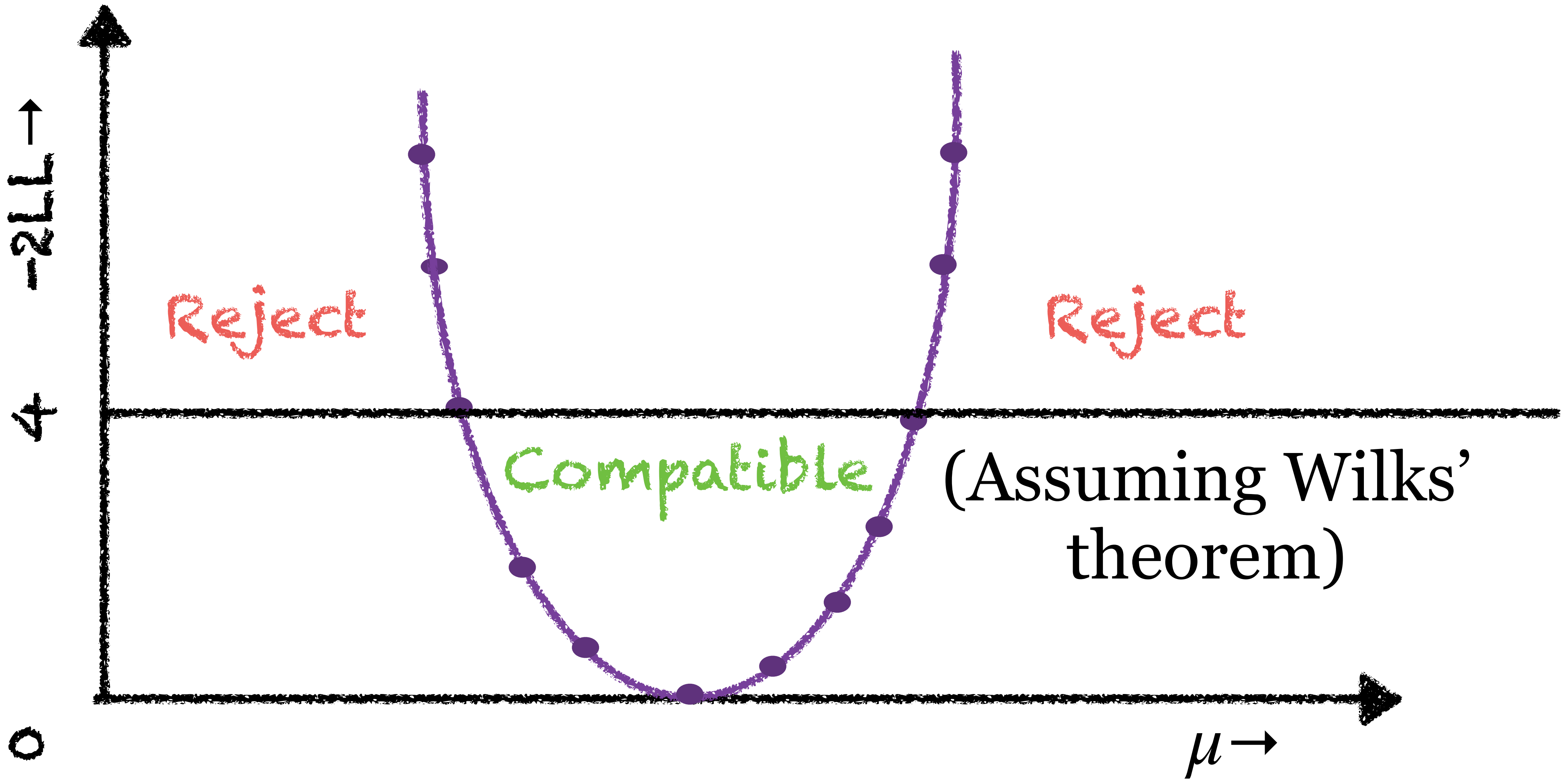
Hypothesis tests

$$\mathscr{L}(H_1\,|\,\text{data}) = p(\text{data}\,|\,H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

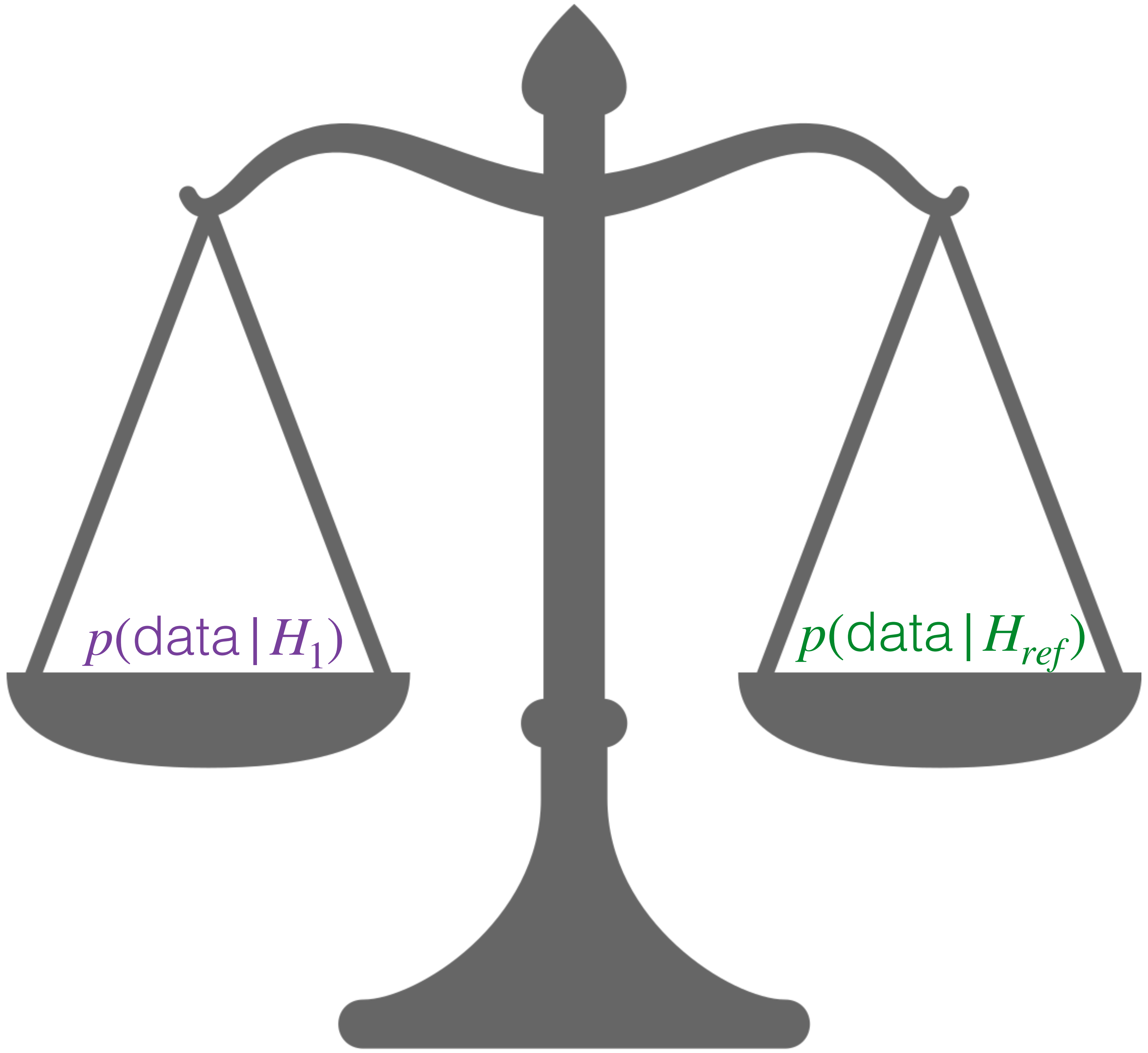
Invert the test for confidence intervals



Parameter estimation (infinite hypotheses)

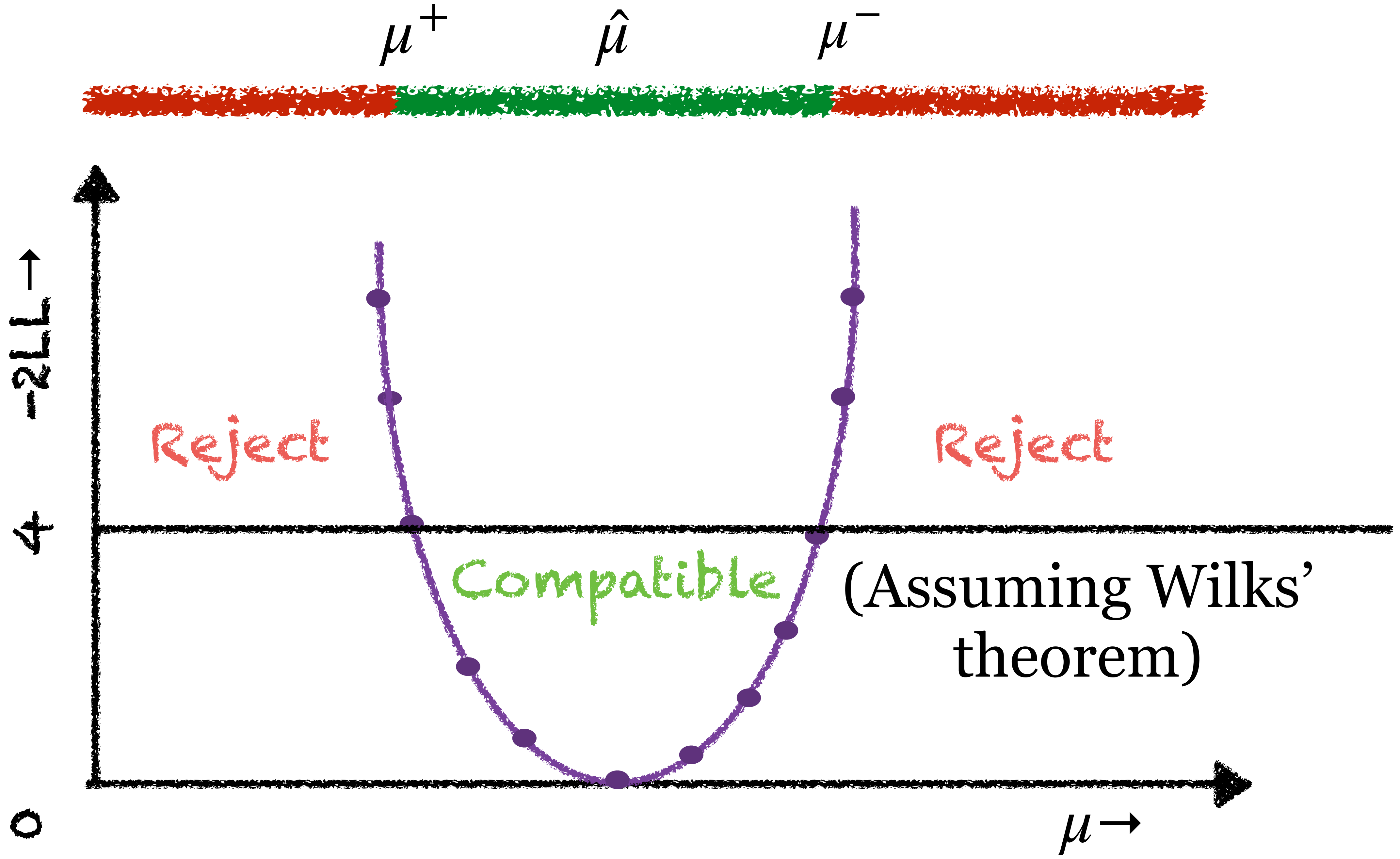
Hypothesis tests

$$\mathscr{L}(H_1\,|\,\text{data}) = p(\text{data}\,|\,H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

Invert the test for confidence intervals



Parameter estimation (infinite hypotheses)

"Well justified assumptions"

<u>Claim</u>

- You can summarise data into one variable
- You can rely on asymptotic approximations
- Impact of nuisance parameters can be factored
- We have thousands of samples, live in asymptopia, and likelihood ratio test is optimal

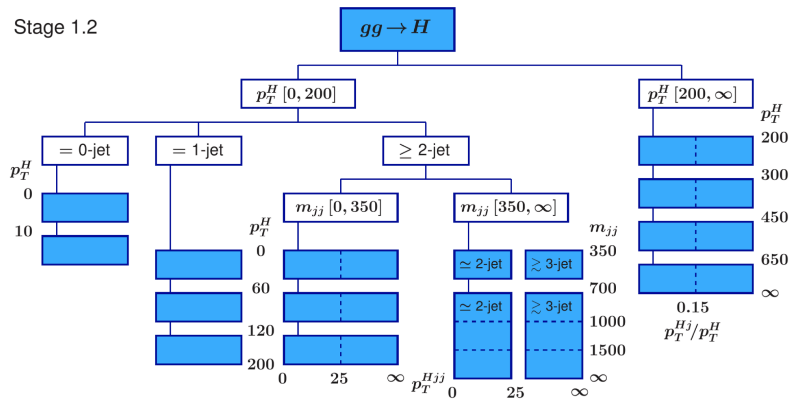
<u>Self-fulfilment</u>

“Well justified assumptions”

Claim

- You can summarise data into one variable
- You can rely on asymptotic approximations
- Impact of nuisance parameters can be factored
- We have thousands of samples, live in asymptopia, and likelihood ratio test is optimal

Self-fulfilment



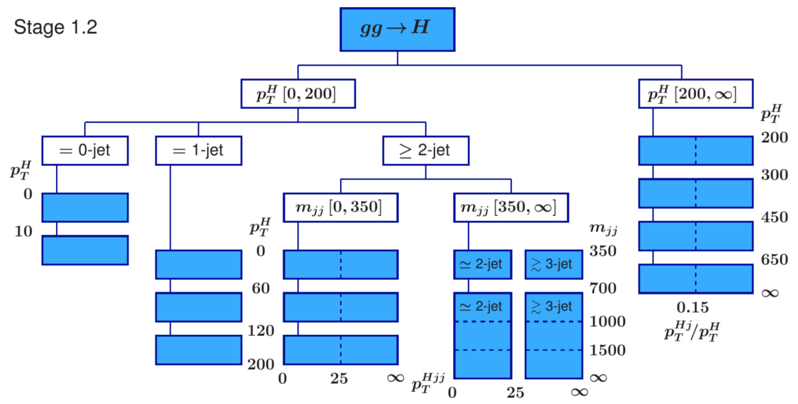
- We will split data into independent analyses until claim becomes effectively true (shrinking statistics per analysis)

“Well justified assumptions”

Claim

- You can summarise data into one variable
- You can rely on asymptotic approximations
- Impact of nuisance parameters can be factored
- We have thousands of samples, live in asymptopia, and likelihood ratio test is optimal

Self-fulfilment



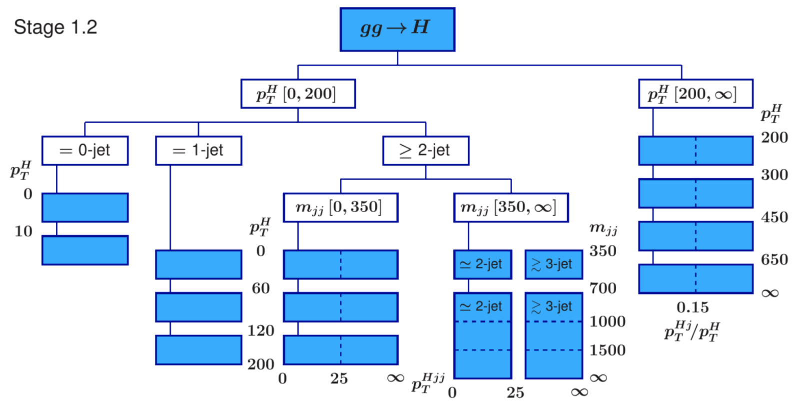
- We will split data into independent analyses until claim becomes effectively true (shrinking statistics per analysis)
- We will only make analysis choices that let asymptotics kick in (sacrificing fine-grain information)

“Well justified assumptions”

Claim

- You can summarise data into one variable
- You can rely on asymptotic approximations
- Impact of nuisance parameters can be factored
- We have thousands of samples, live in asymptopia, and likelihood ratio test is optimal

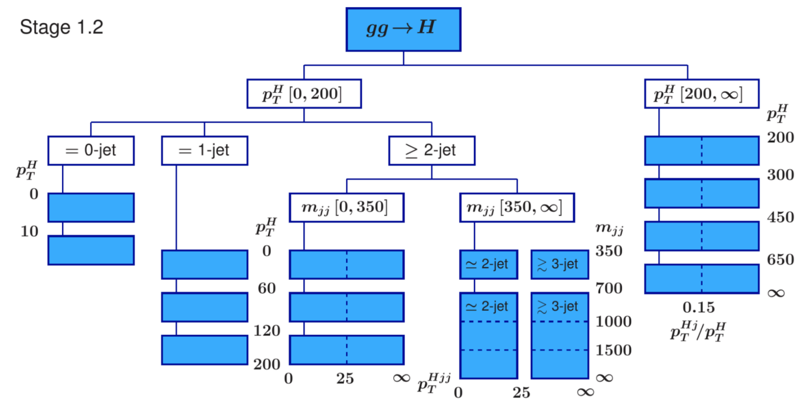
Self-fulfilment



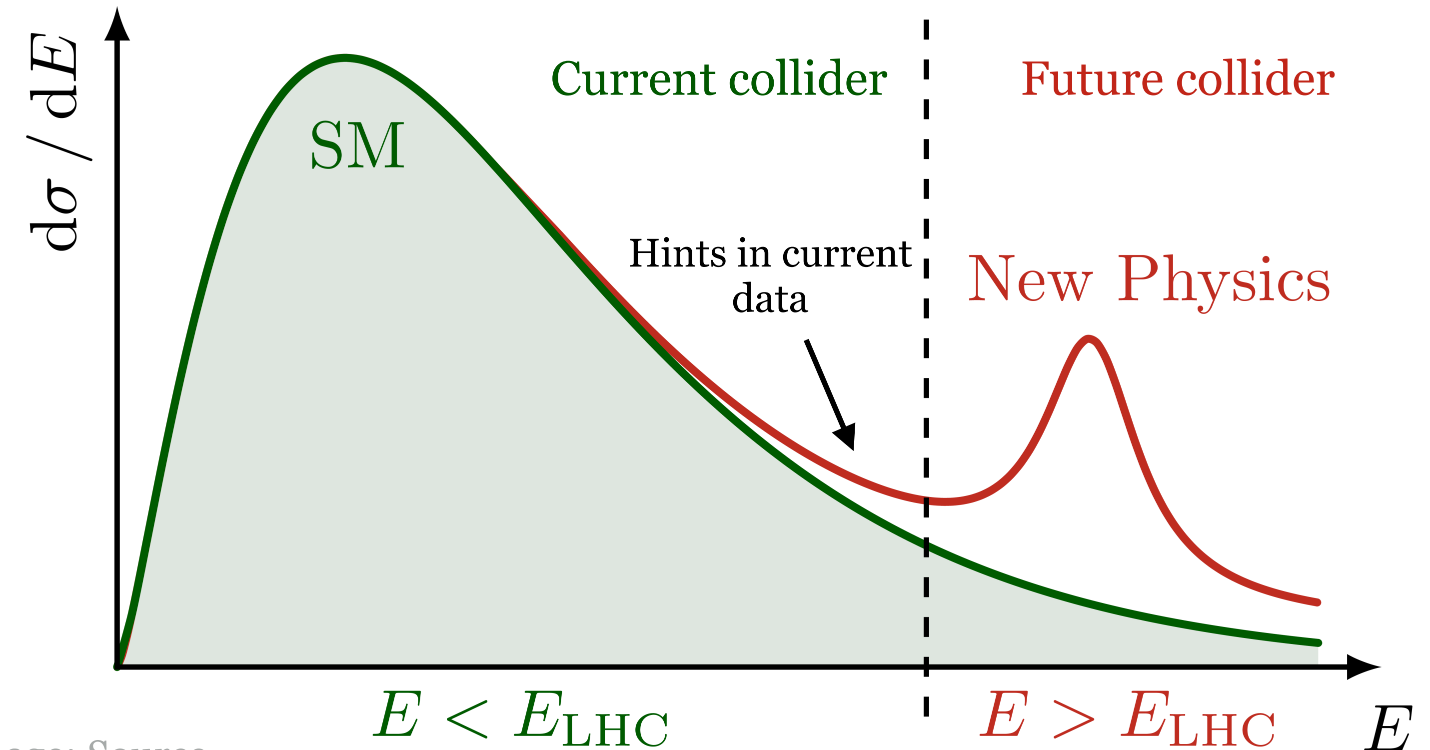
- We will split data into independent analyses until claim becomes effectively true (shrinking statistics per analysis)
- We will only make analysis choices that let asymptotics kick in (sacrificing fine-grain information)
- We re-defined nuisances in the dirtiest way to satisfy this claim (sacrificing power, or worse)

“Well justified assumptions”

Claim	Self-fulfilment
• You can summarise data into one variable	• We will split data into independent analyses until claim becomes effectively true (shrinking statistics per analysis)
• You can rely on asymptotic approximations	• We will only make analysis choices that let asymptotics kick in (sacrificing fine-grain information)
• Impact of nuisance parameters can be factored	• We re-defined nuisances in the dirtiest way to satisfy this claim (sacrificing power, or worse)
• We have thousands of samples, live in asymptopia, and likelihood ratio test is optimal	



But our entire field is about studying the tails!



Open problems to neural SBI application to LHC data:

- Robustness: Design and validation
- Systematic Uncertainties: Propagate them through network
- Neyman Construction: Generating pseudo-experiments to invert the test

Open problems to neural SBI application to LHC data:

- Robustness: Design and validation
- Systematic Uncertainties: Propagate them through network
- Neyman Construction: Generating pseudo-experiments to invert the test

→ Addressed with semi-parametric approach and lots of compute!

Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER · OPEN ACCESS

An implementation of neural sin
inference for parameter estimat

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog.*

View the article online for updates and enhancements.

ATLAS collaboration (incl. Ghosh)

Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER · OPEN ACCESS

Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV *pp* collisions with the ATLAS detector

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 057803

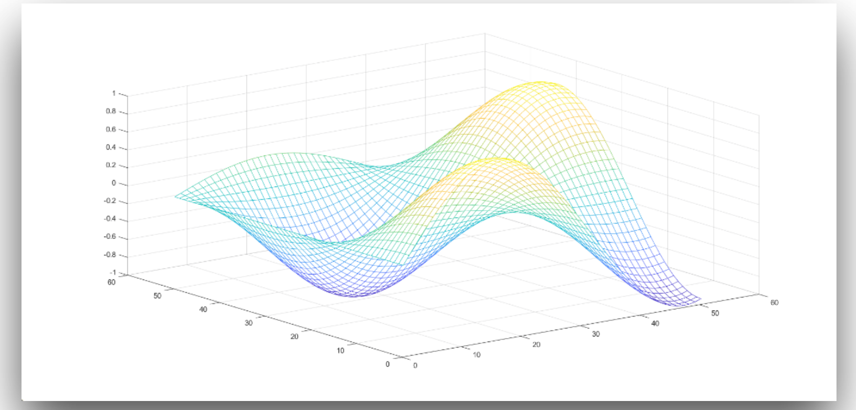
View the article online for updates and enhancements.

You may also like

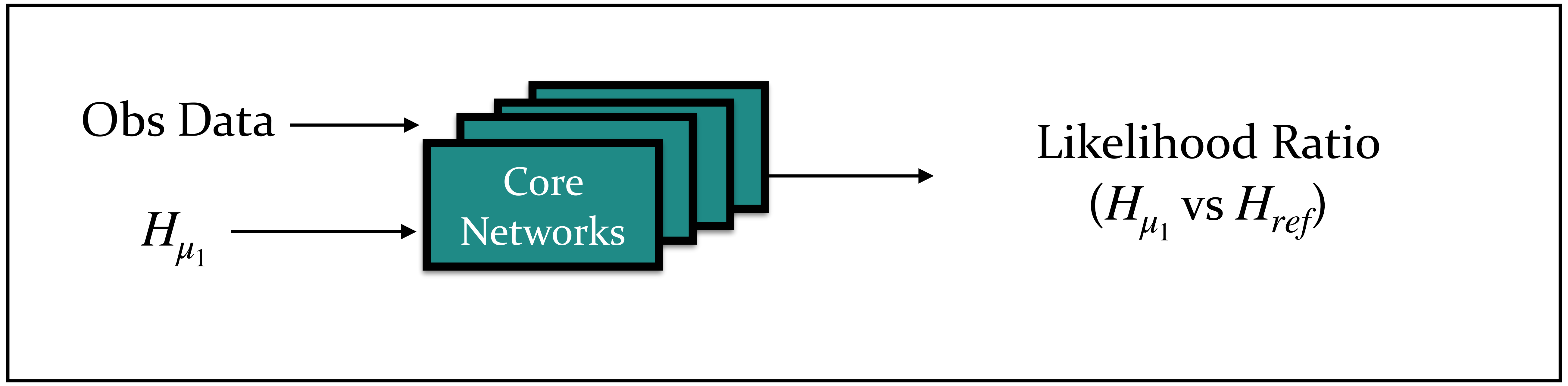
- Dark matter search with a resonantly-coupled hybrid spin system
Kai Wei, Zitong Xu, Yuxuan He et al.
- Beyond Kitaev physics in strong spin-orbit coupled magnets
Ioannis Rousochatzakis, Natalia B Perkins, Qiang Luo et al.
- 70 years of hyperon spectroscopy: a review of strange , baryons, and the spectrum of charmed and bottom baryons
Volker Crede and John Yelton

And took 6 years of my own life to make it a reality..

Semi-parametric NSBI implementation



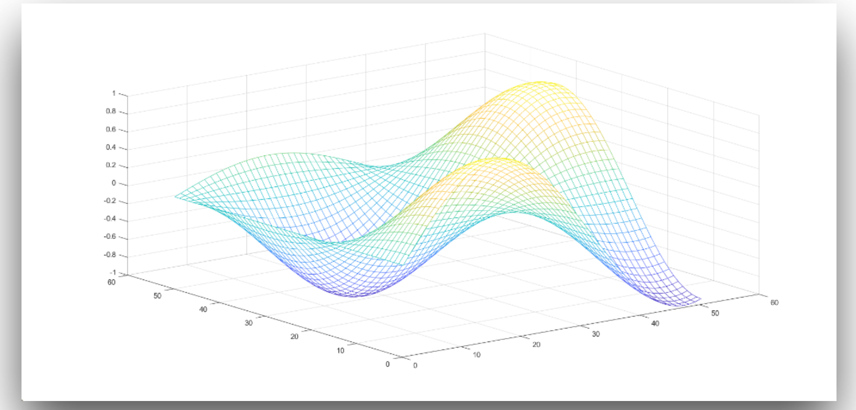
$\mathcal{O}(16)$ observables



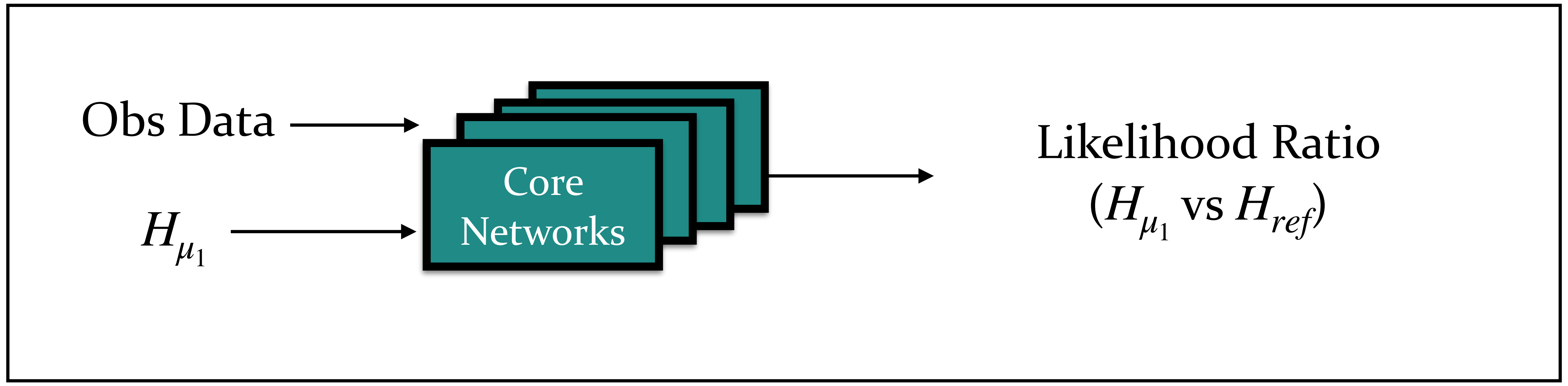
$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_j^C f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_k^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Semi-parametric NSBI implementation



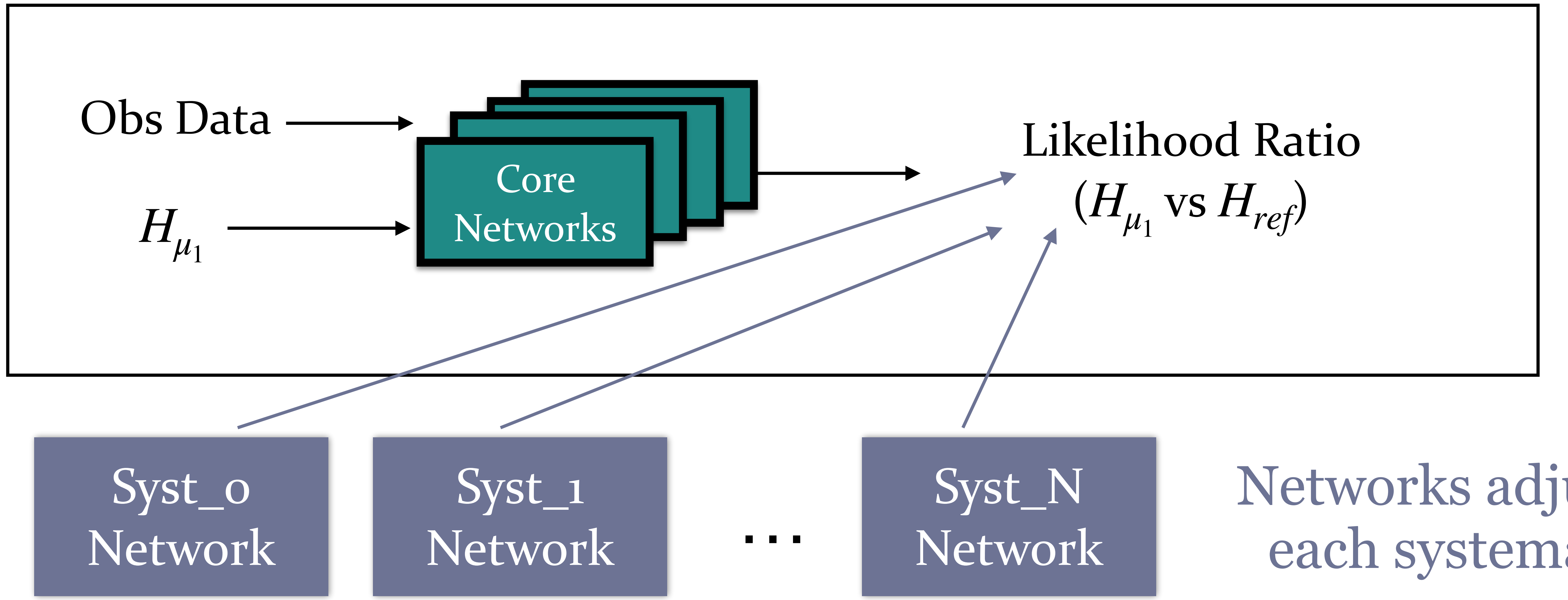
$\mathcal{O}(16)$ observables



$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Semi-parametric NSBI implementation

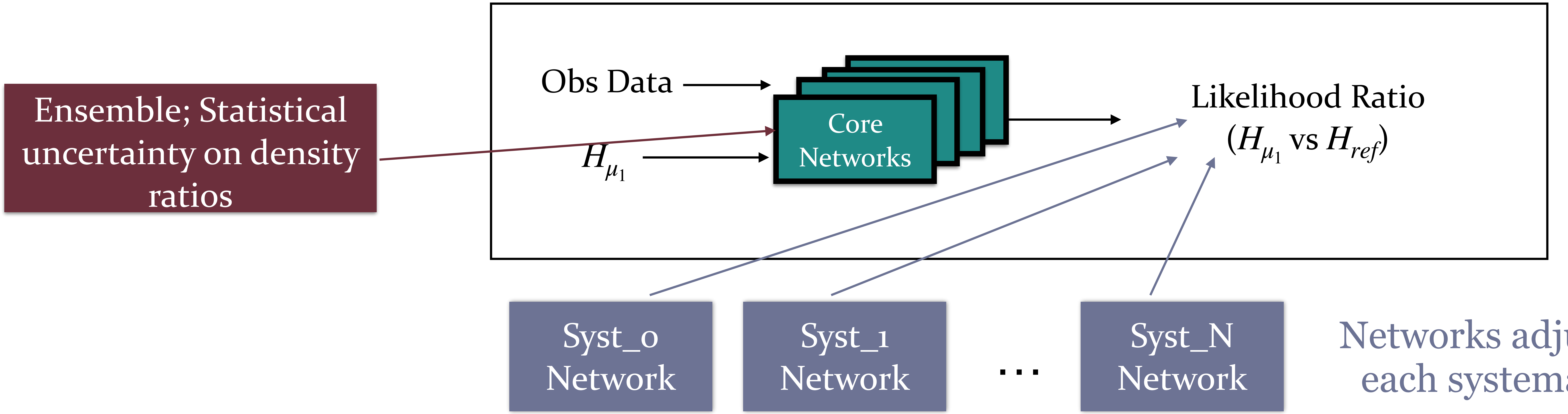


Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

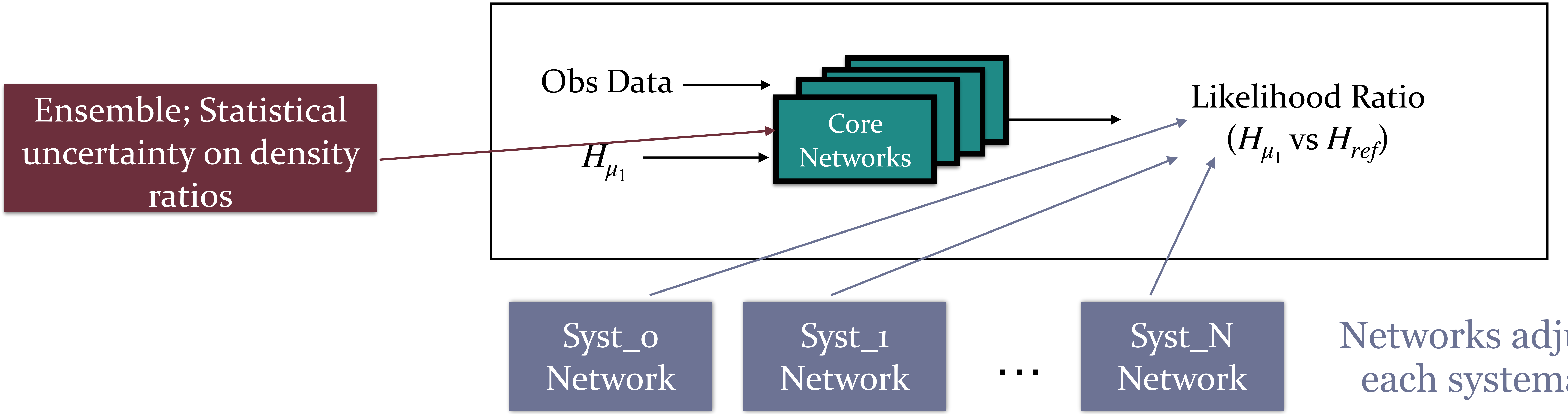
Semi-parametric NSBI implementation



Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^C f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

Semi-parametric NSBI implementation



Networks adjust likelihood for each systematic uncertainty

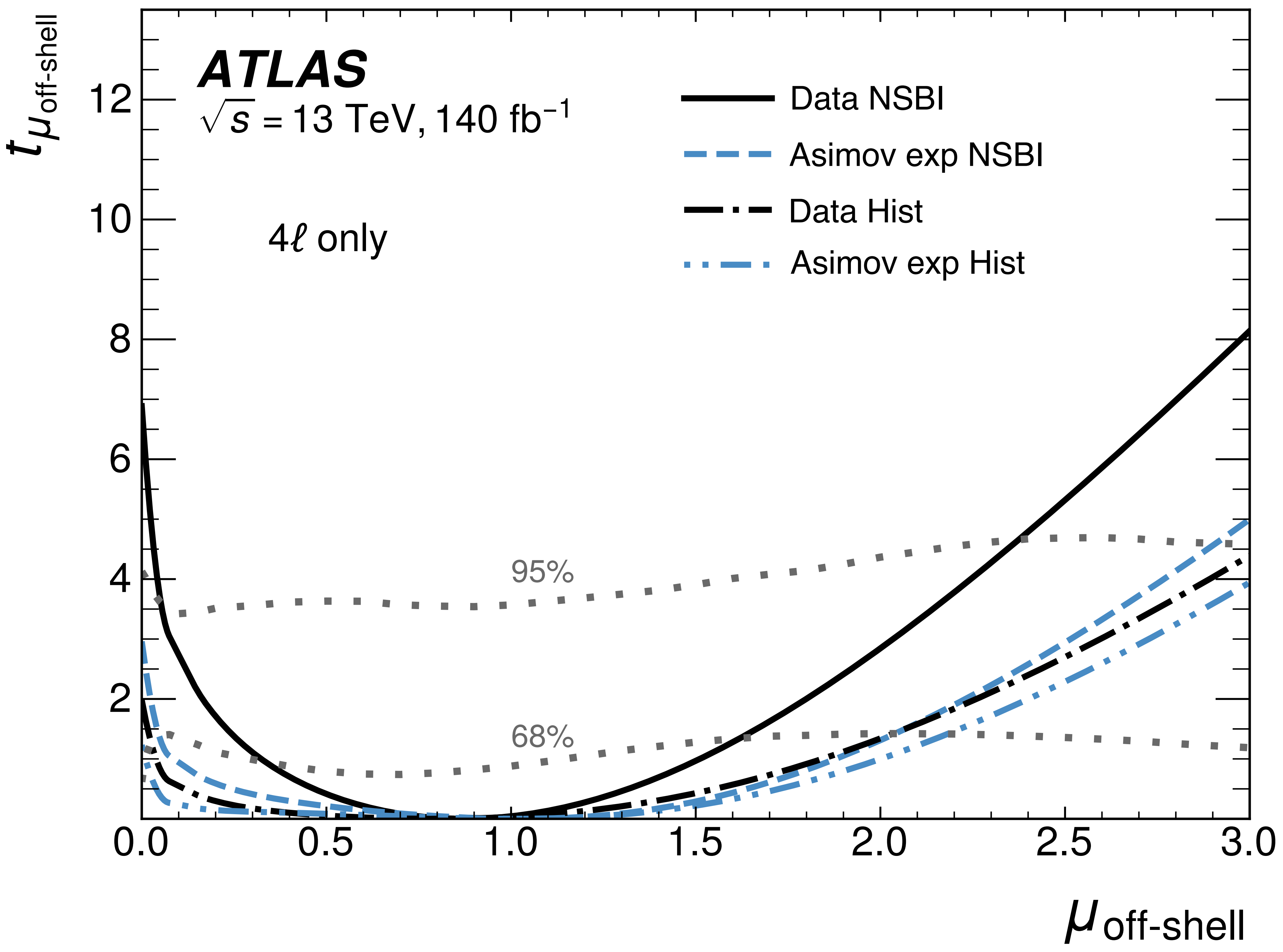
$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

Training details

- ✦ Train $O(10^4)$ networks on TensorFlow
- ✦ Computing resources provided by Google, SMU, other HPC clusters
- ✦ Fits with JAX

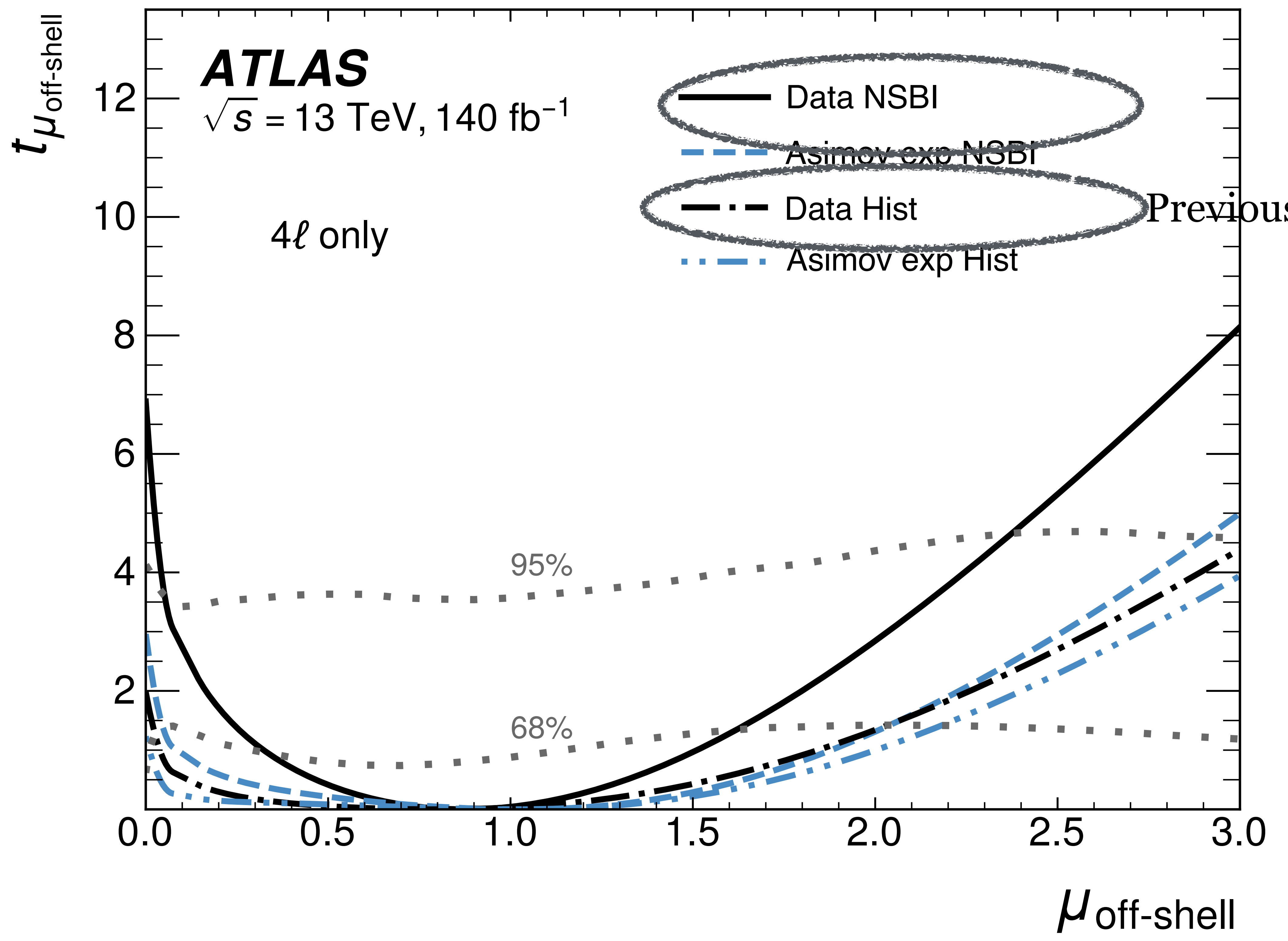
Application to a flagship Higgs measurement at LHC

NSBI vs (published) histogram analysis



Application to a flagship Higgs measurement at LHC

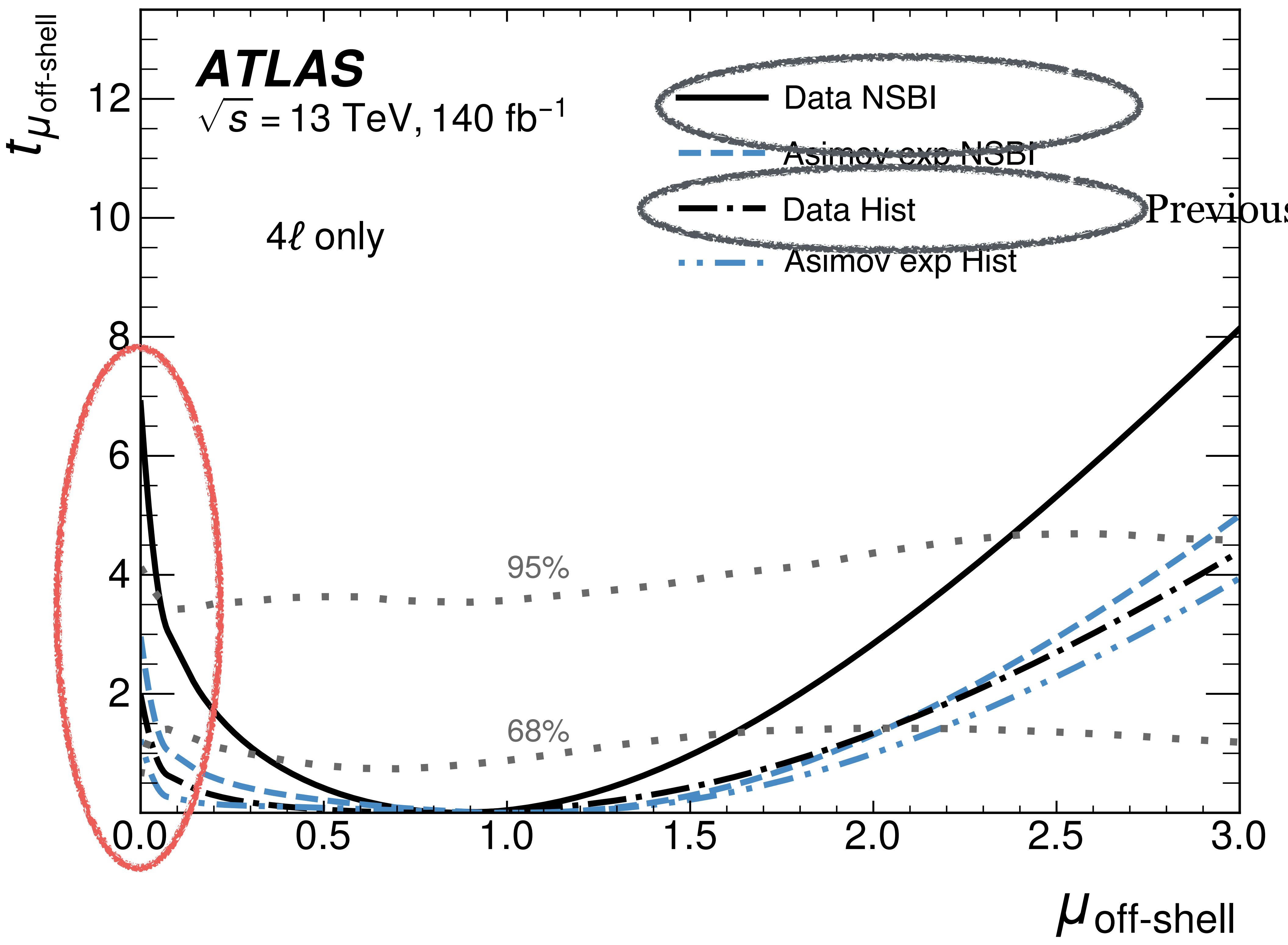
NSBI vs (published) histogram analysis



Previously published histogram paper

Application to a flagship Higgs measurement at LHC

NSBI vs (published) histogram analysis



Previously published histogram paper

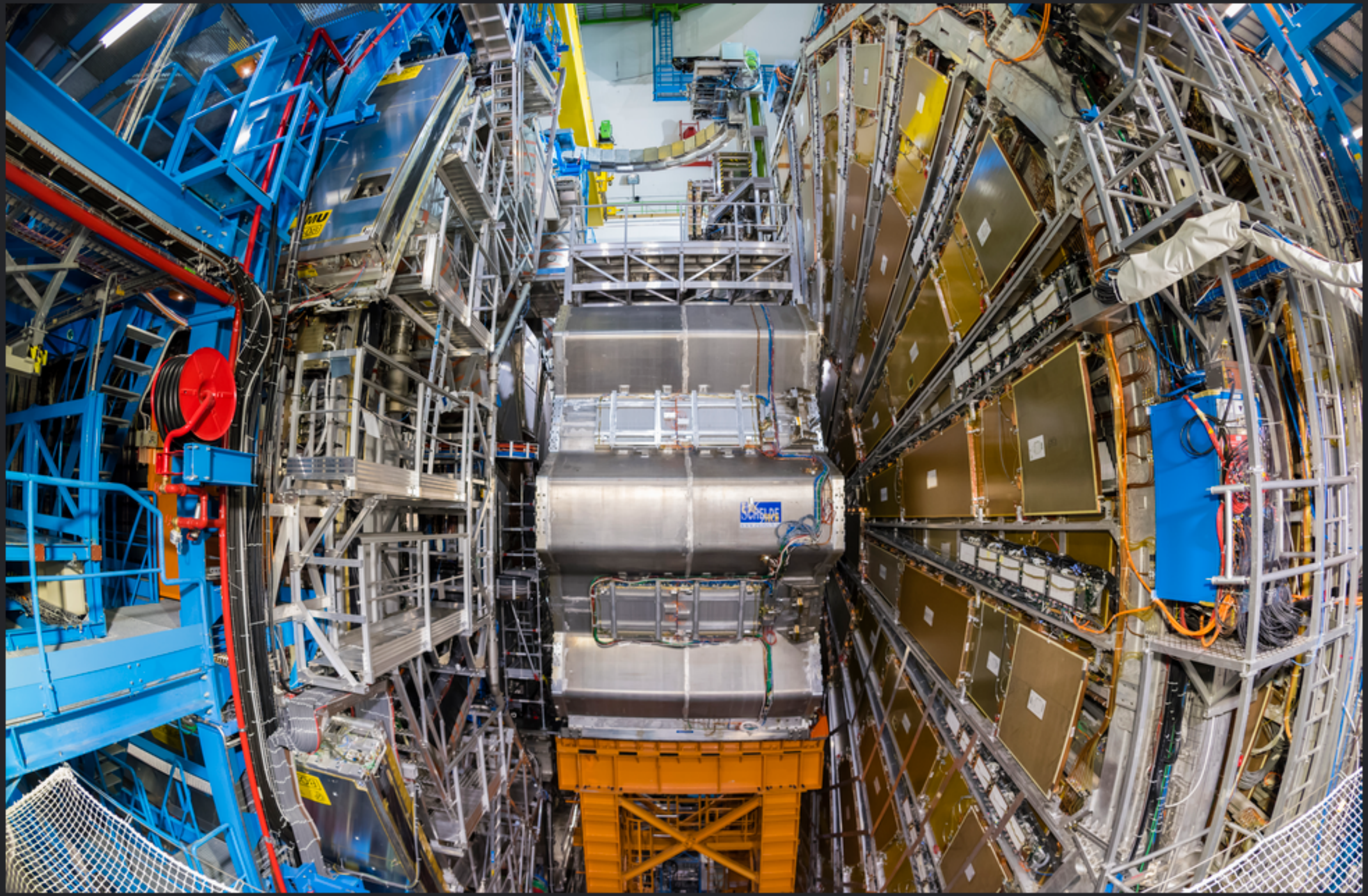
Unprecedented improvement in ability to reject null hypothesis!

A VERY FUN PROBLEM

How a grad student got LHC data to play nice with quantum interference

New approach is already having an impact on the experiment's plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 | 70



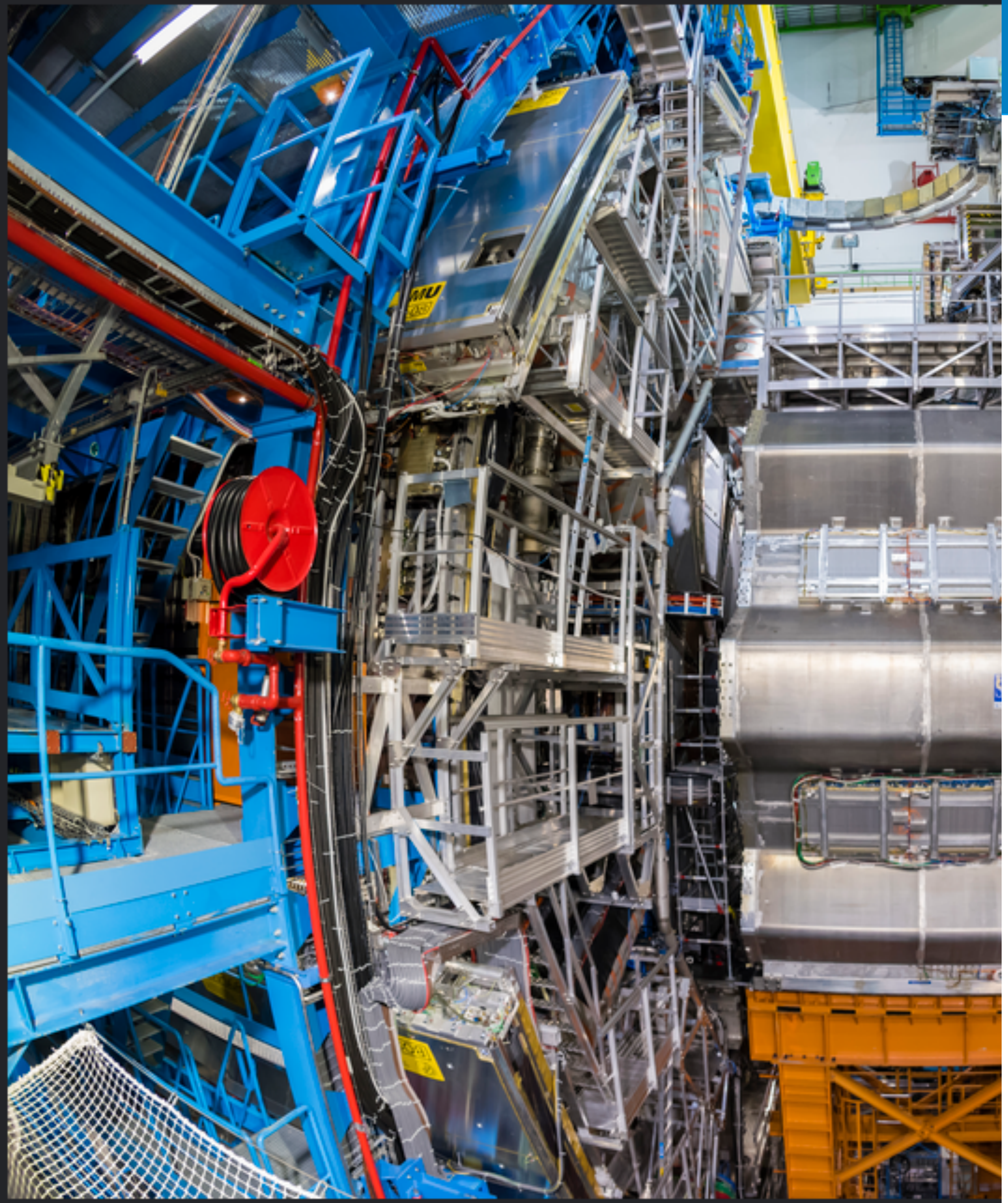
The ATLAS particle detector of the Large Hadron Collider (LHC) at the European Nuclear Research Center (CERN) in Geneva, Switzerland. Credit: EThamPhoto/Getty Images

A VERY FUN PROBLEM

How a grad student got LHC data to play nice with quantum interference

New approach is already having an impact on the experiment's plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 | 70



The ATLAS particle detector of the Large Hadron Collider (LHC) at the European N

EThamPhoto/Getty Images

Updates > Briefing > Cracking open the Higgs shell: new ATLAS measurement of "off-shell" production uses AI techniques

Physics Briefing

Tags:
Higgs boson,
physics results

Cracking open the Higgs shell: new ATLAS measurement of "off-shell" production uses AI techniques

6 November 2024 | By ATLAS Collaboration

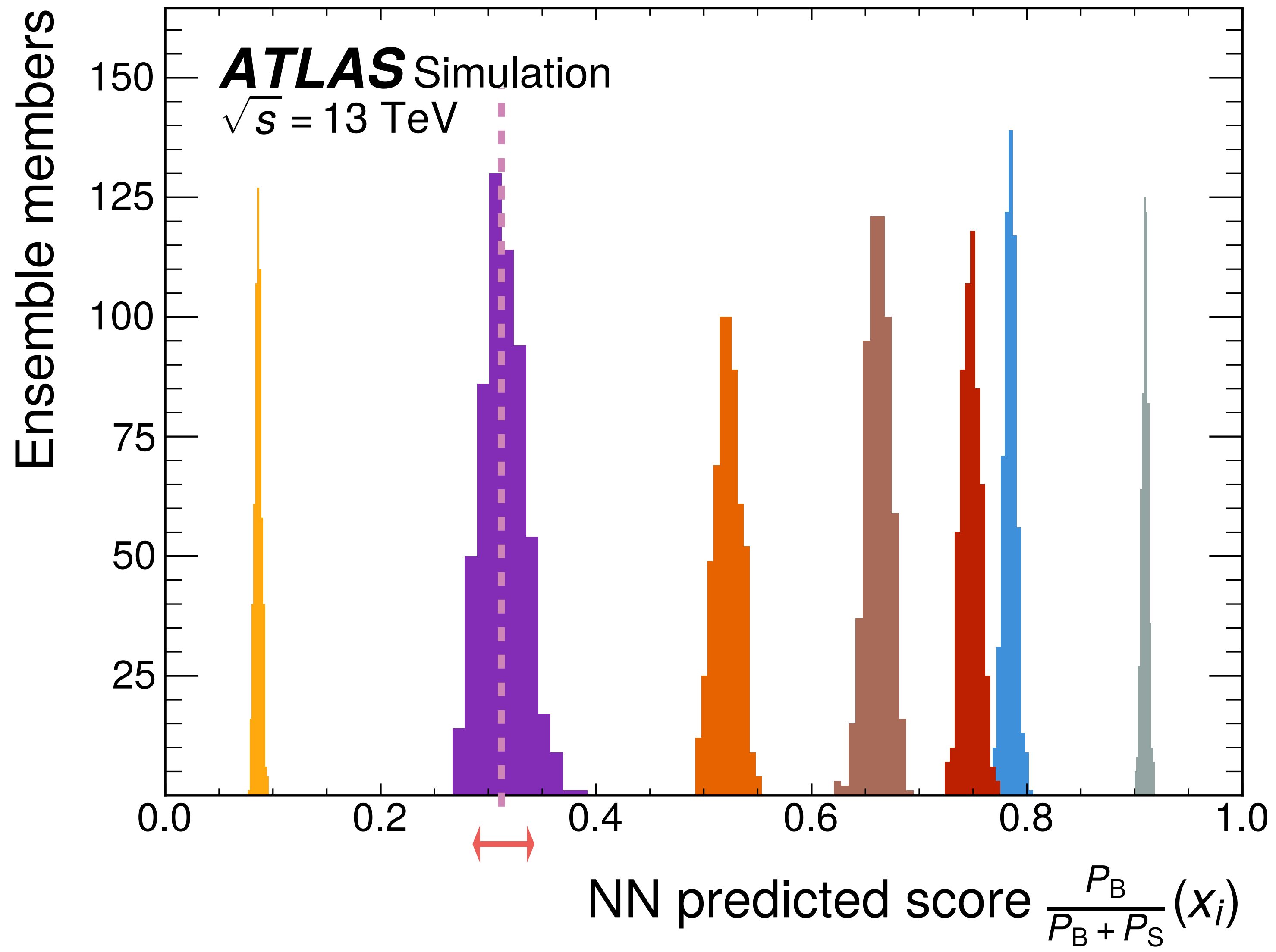
In 2012, scientists from the ATLAS and CMS Collaborations detected a "bump" in their data. This peak in the mass distribution, around 125 GeV, revealed the presence of the Higgs boson. While most Higgs bosons are observed at the LHC with this "on shell" mass, about 15% of Higgs bosons have a "virtual" mass well above 125 GeV – known as the "off shell" mass. This is one of the quirks of quantum mechanics, which allows particles to fluctuate their mass for an extremely short time.

By comparing the production rates of on-shell and off-shell Higgs bosons, physicists can study key properties of the Higgs boson, such as the width of the peak in the mass distribution. This width is related to the Higgs boson's lifetime and its decay rate to other particles. As it is predicted to be just 4 MeV – more than a hundred times smaller than the resolution of the ATLAS detector – this is the only way to set constraints on its value at the LHC. Additionally, the rate of off-shell Higgs-boson production is sensitive to potential contributions from Beyond the Standard Model (BSM) physics at high energies – making it an important tool for exploring new physics.

As off-shell production doesn't result in the characteristic "bump" of on-shell production in the mass spectrum, it can be difficult to distinguish from background processes with identical signatures and larger rates. To overcome these challenges, the initial ATLAS measurement of the Higgs-boson width with LHC Run-2 data (collected in 2015-2018) used a machine-learned neural network to distinguish signal events from background. This approach, however, is not optimal for a measurement of off-shell Higgs production due to quantum interference between the signal and background.

Uncertainties on the network estimates themselves

Distribution of NN predictions for example events



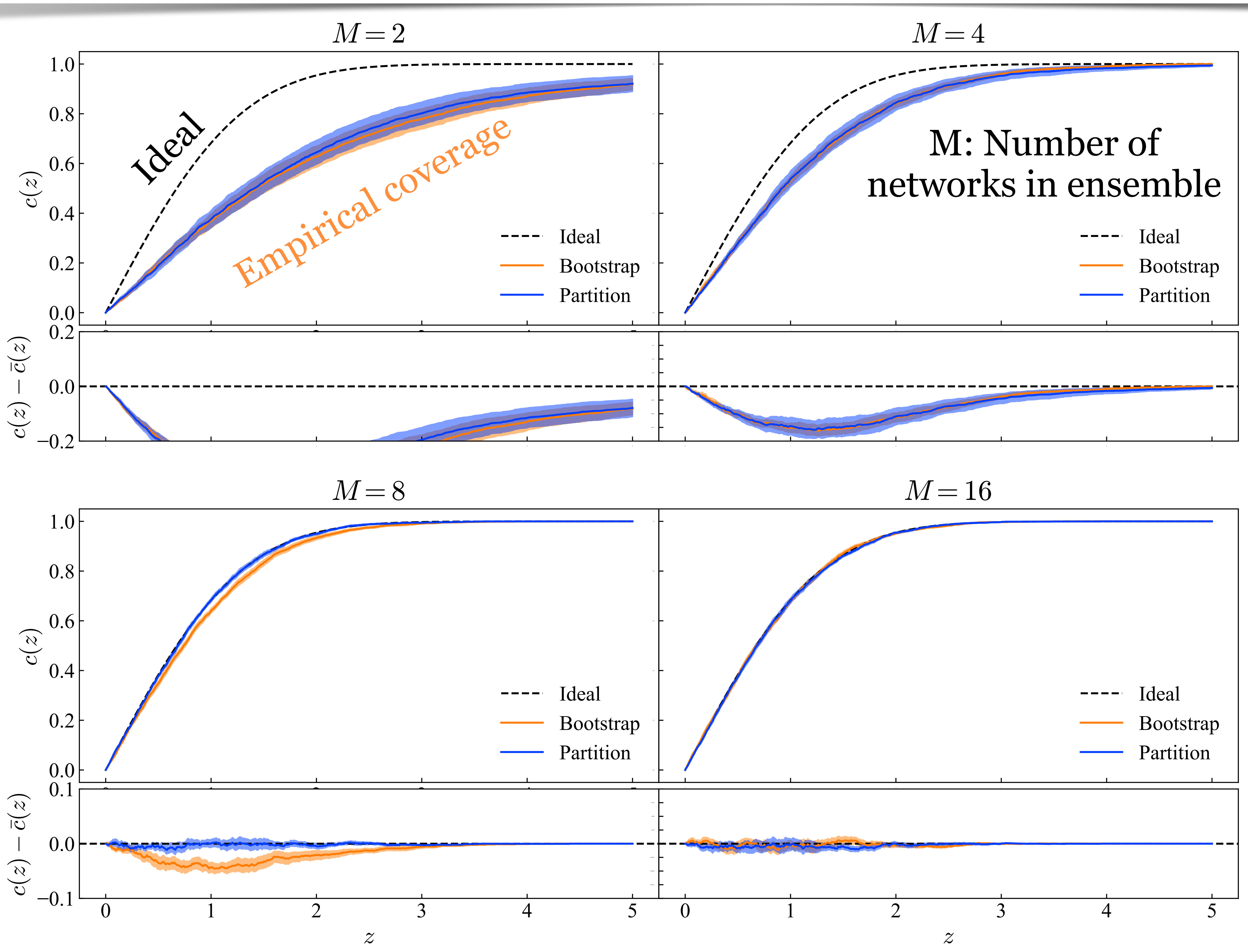
Uncertainties on the network estimates themselves

Frequentist Uncertainties on Neural Density Ratios with $w_i f_i$ Ensembles

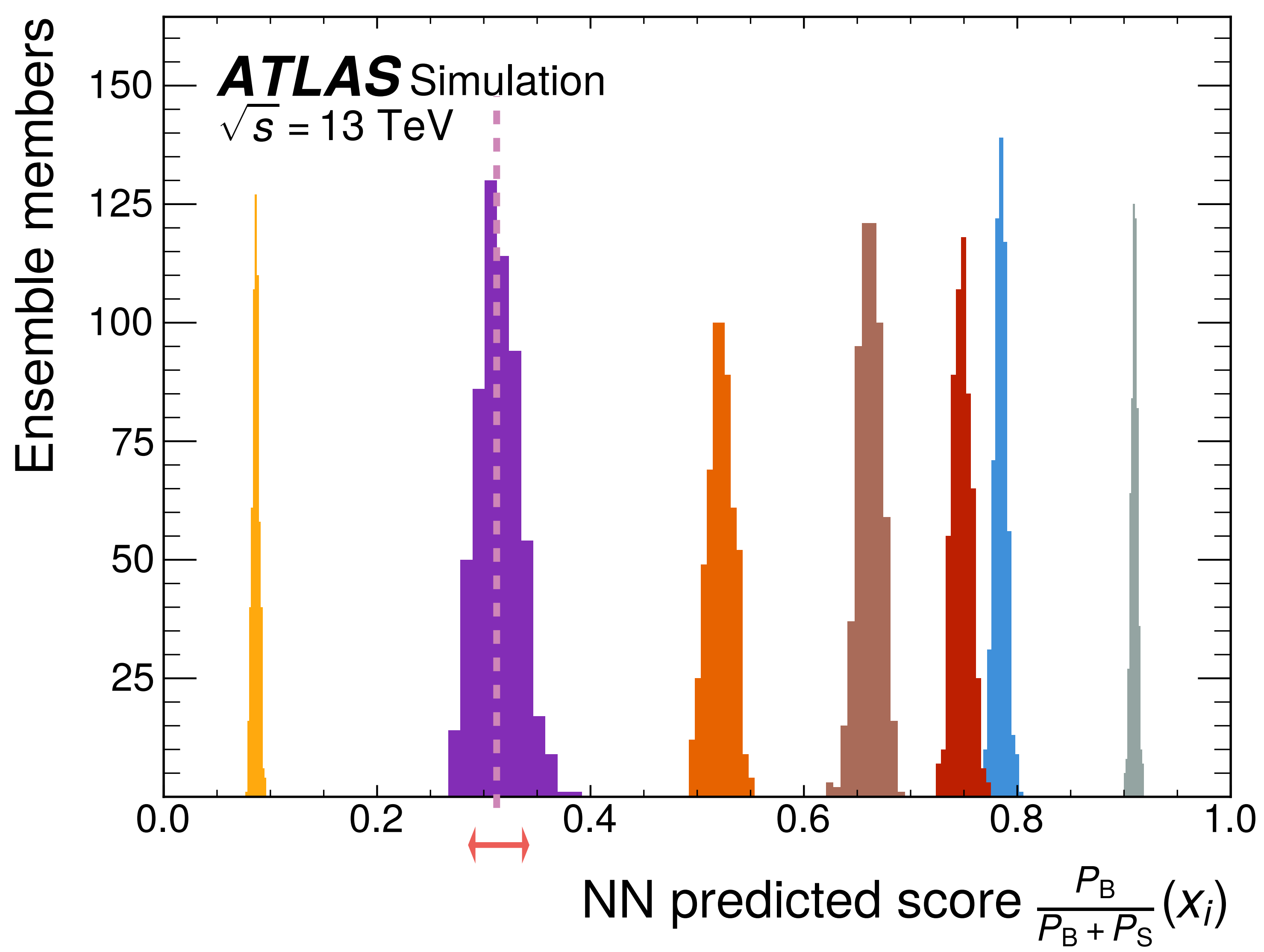
Sean Benevedes[1,2,*] and Jesse Thaler[1,2,†]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States*

[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*



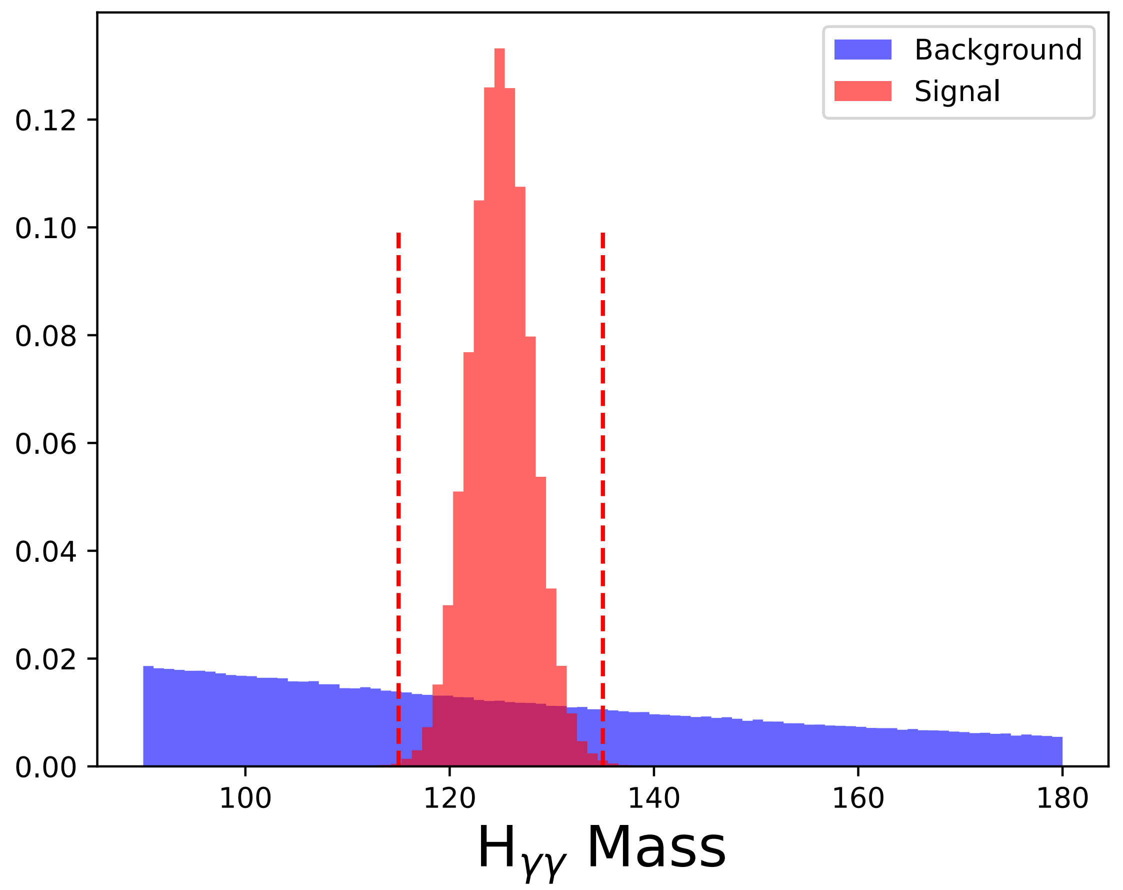
Distribution of NN predictions for example events

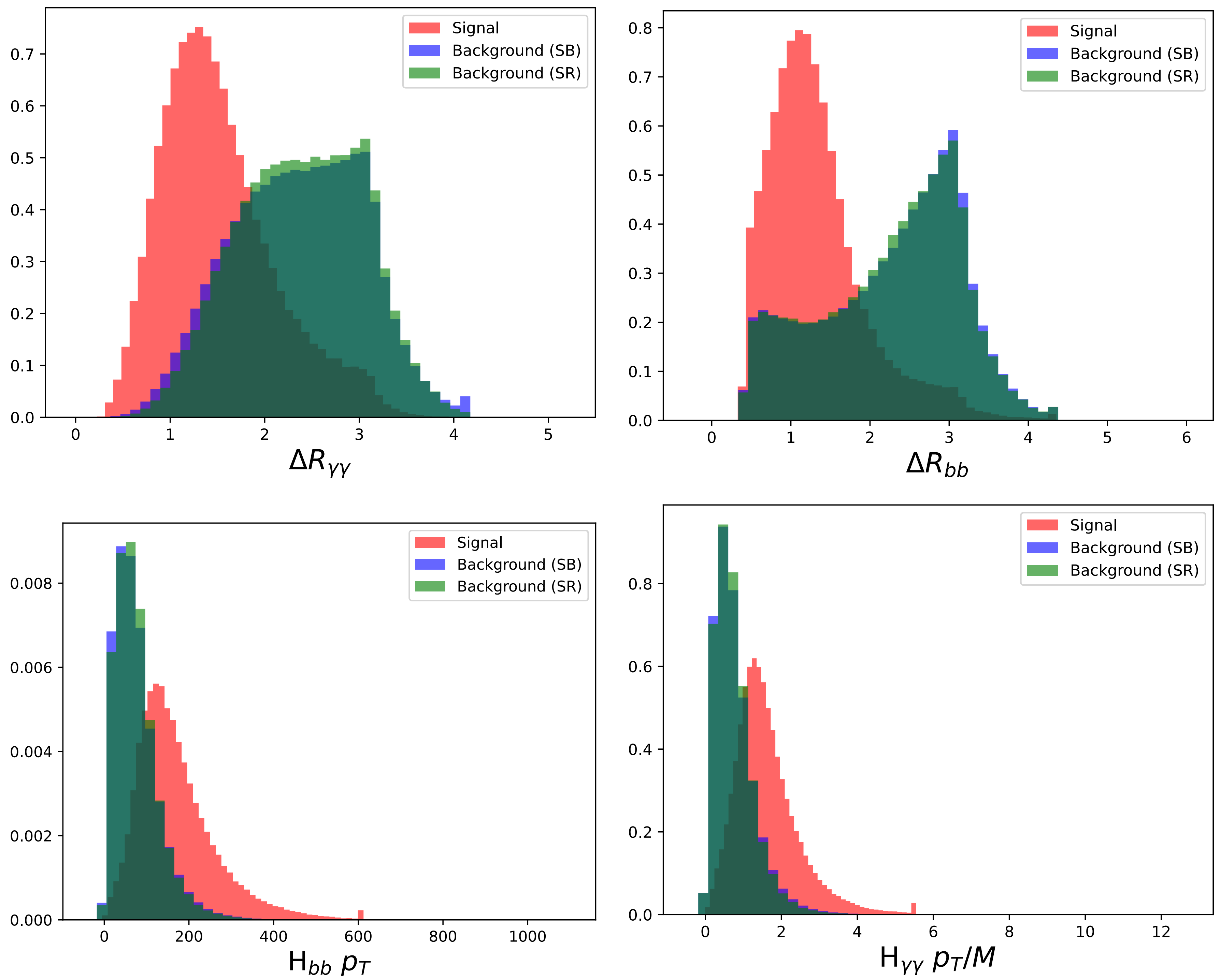


- Genuine 95% coverage, with much smaller ensembles
- Elegant, mathematically motivated method to estimate uncertainties

[arXiv:2506.00113](https://arxiv.org/abs/2506.00113): Benevedes & Thaler

Unreliable simulator ?





- Training data can come from **control regions of real data**
- Data-driven background estimation techniques work with NSBI
- Amram & Szewc uses generative models instead of classifiers for NSBI

arXiv:2506.06438: Amram & Szewc

Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

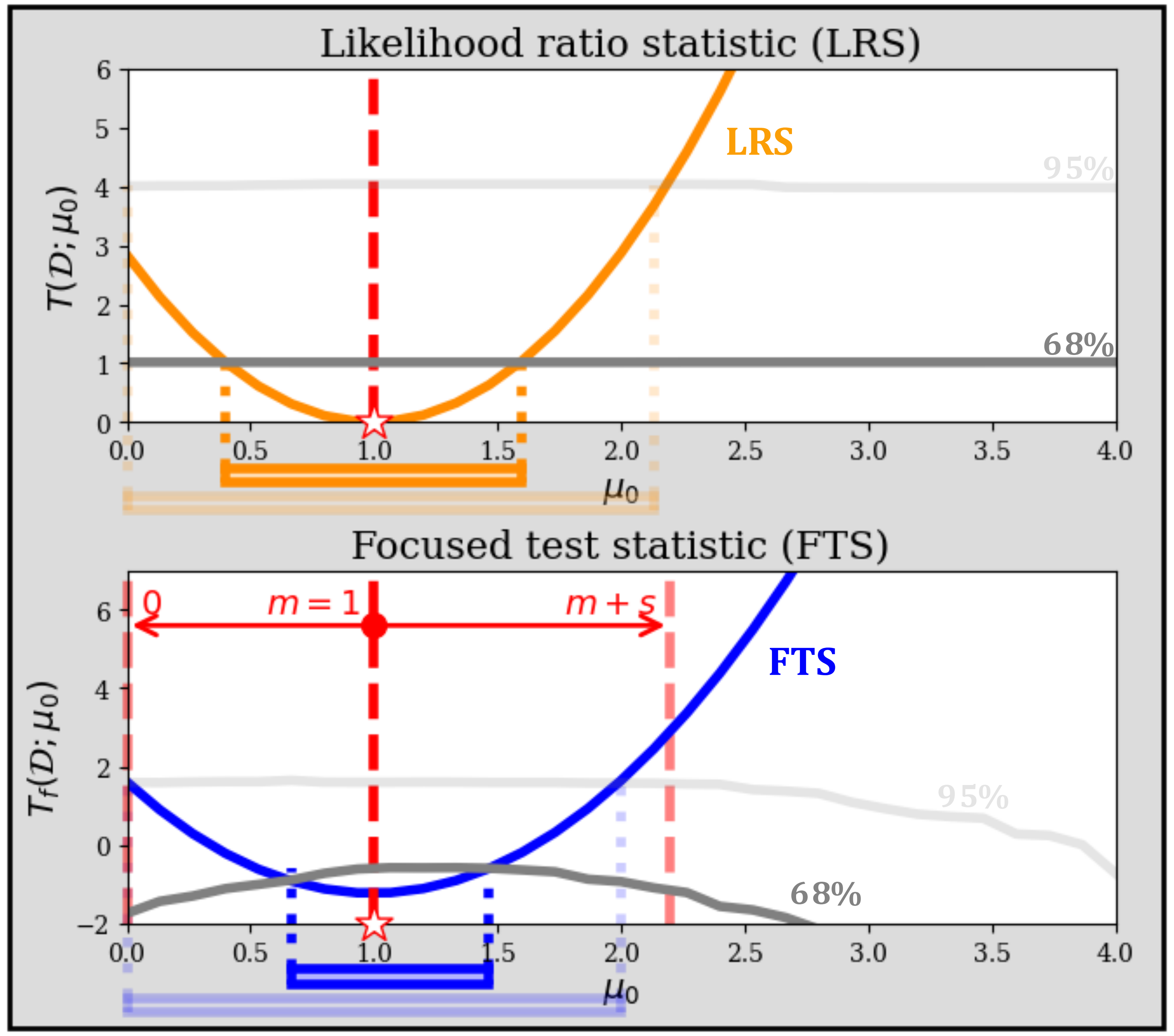
Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \Longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

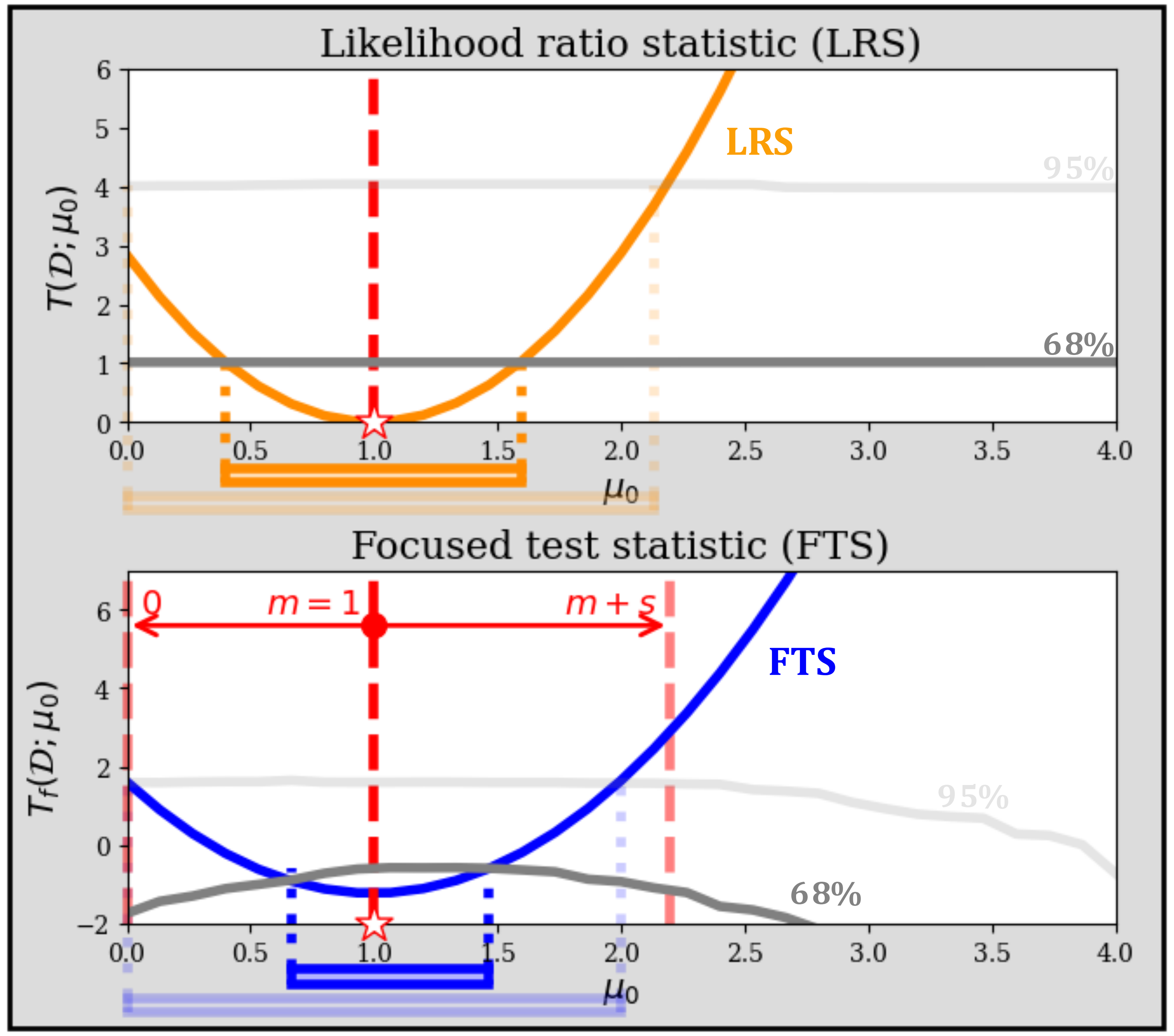
- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses



Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses

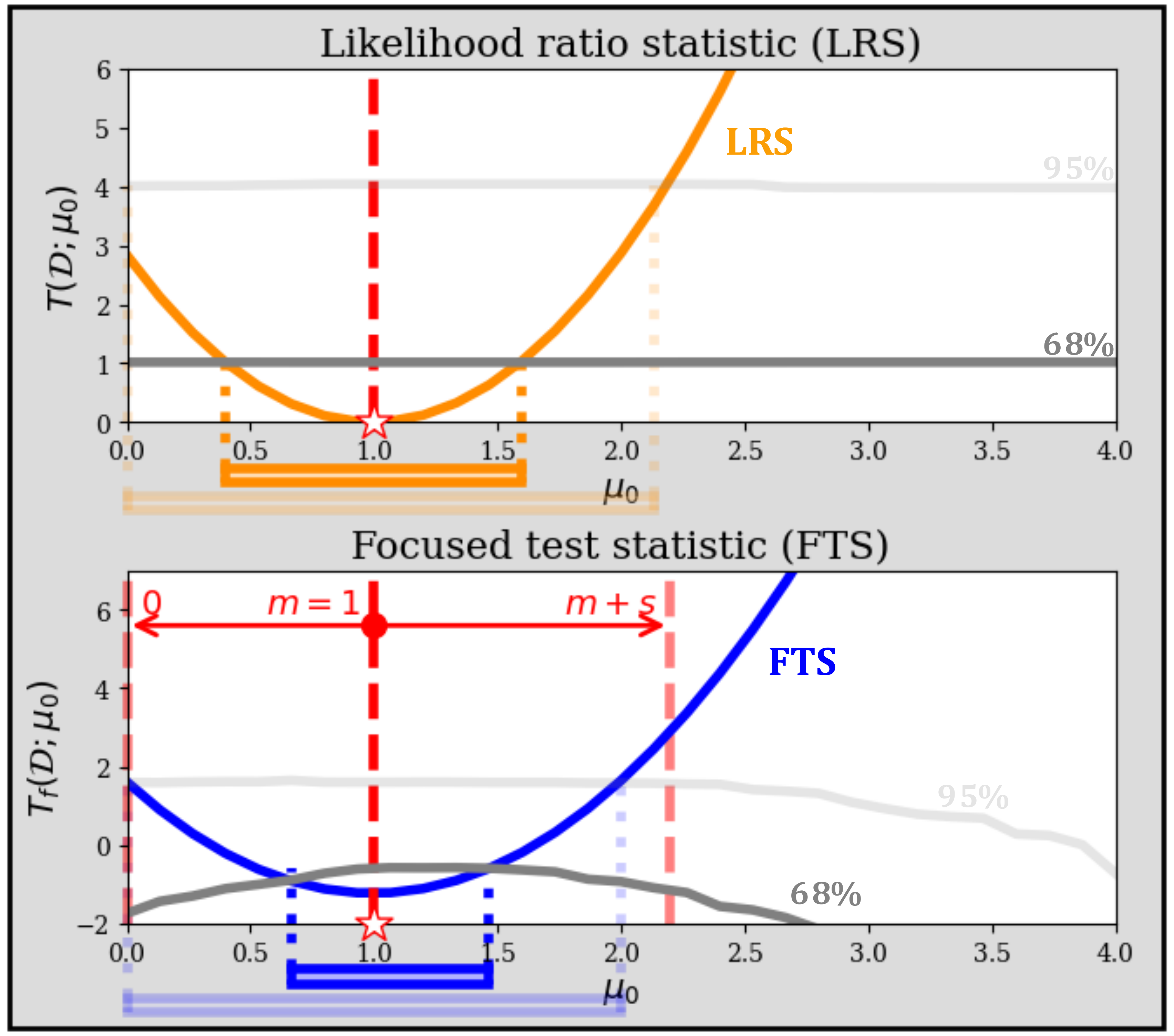


Shorter median length for confidence intervals with FTS even in 'asymptotic regime' where Wilks' applies

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
    - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML

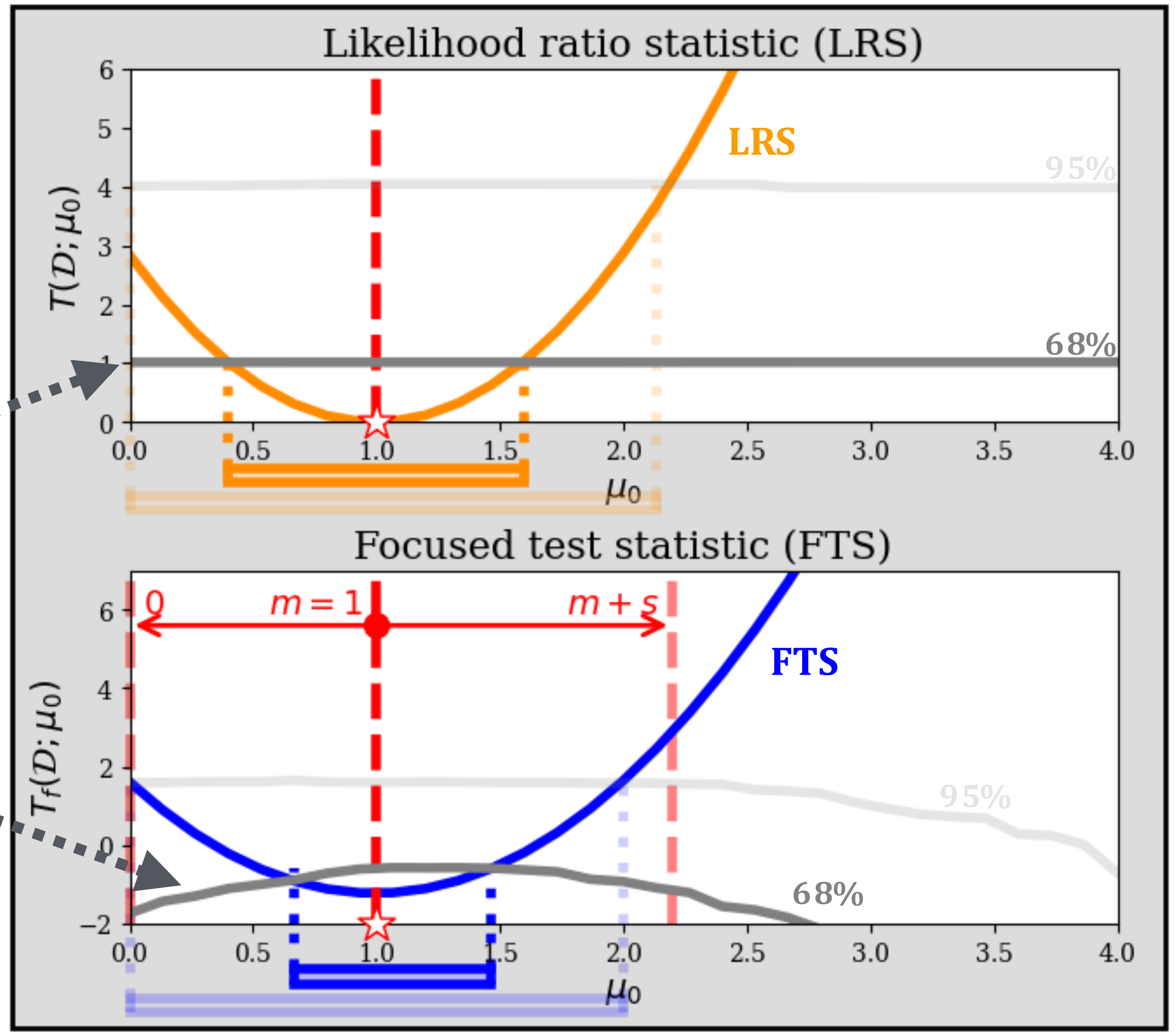


Shorter median length for confidence intervals with FTS even in ‘asymptotic regime’ where Wilks’ applies

Challenging a deeply-held belief in particle physics

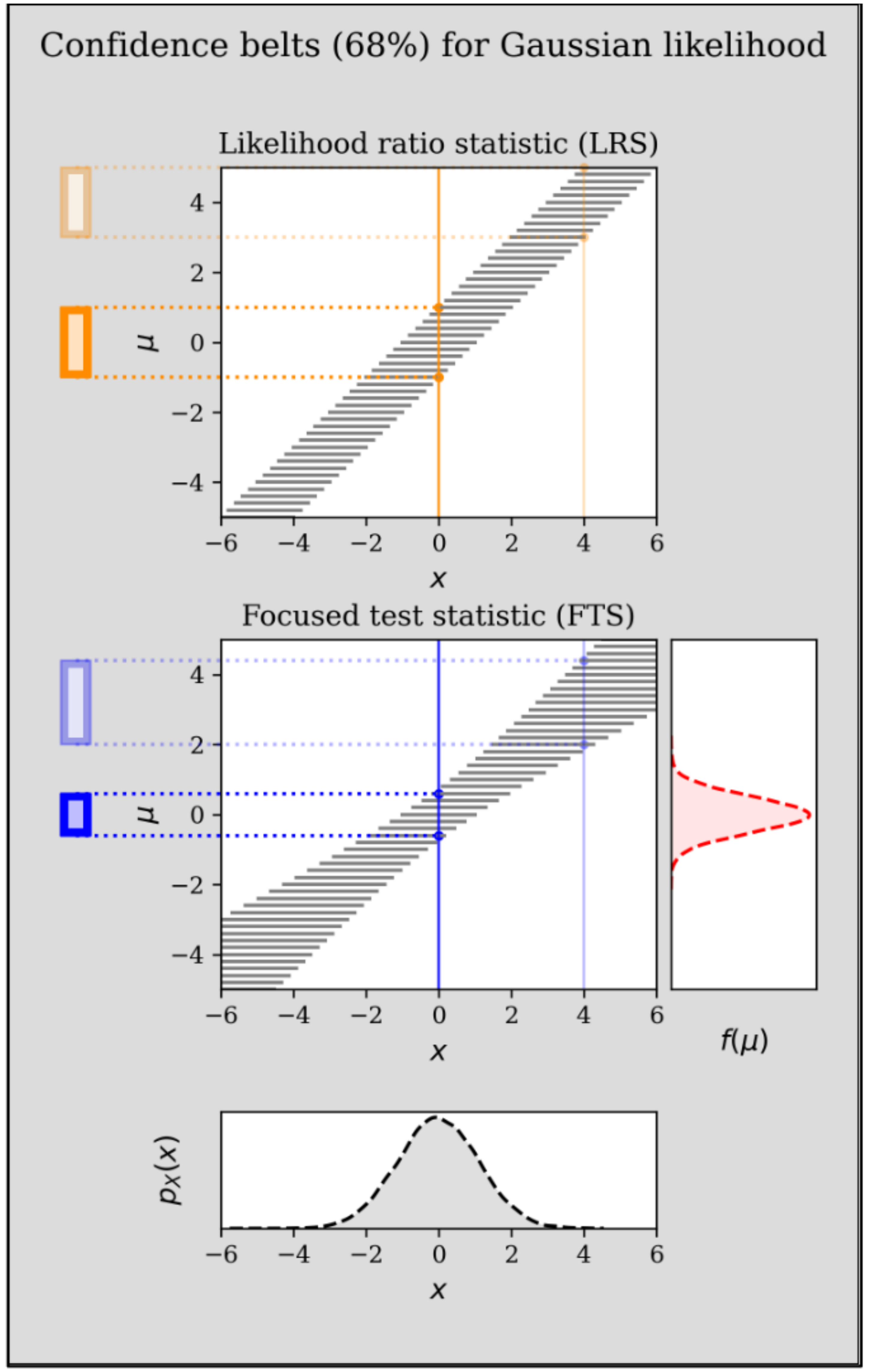
$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
  - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



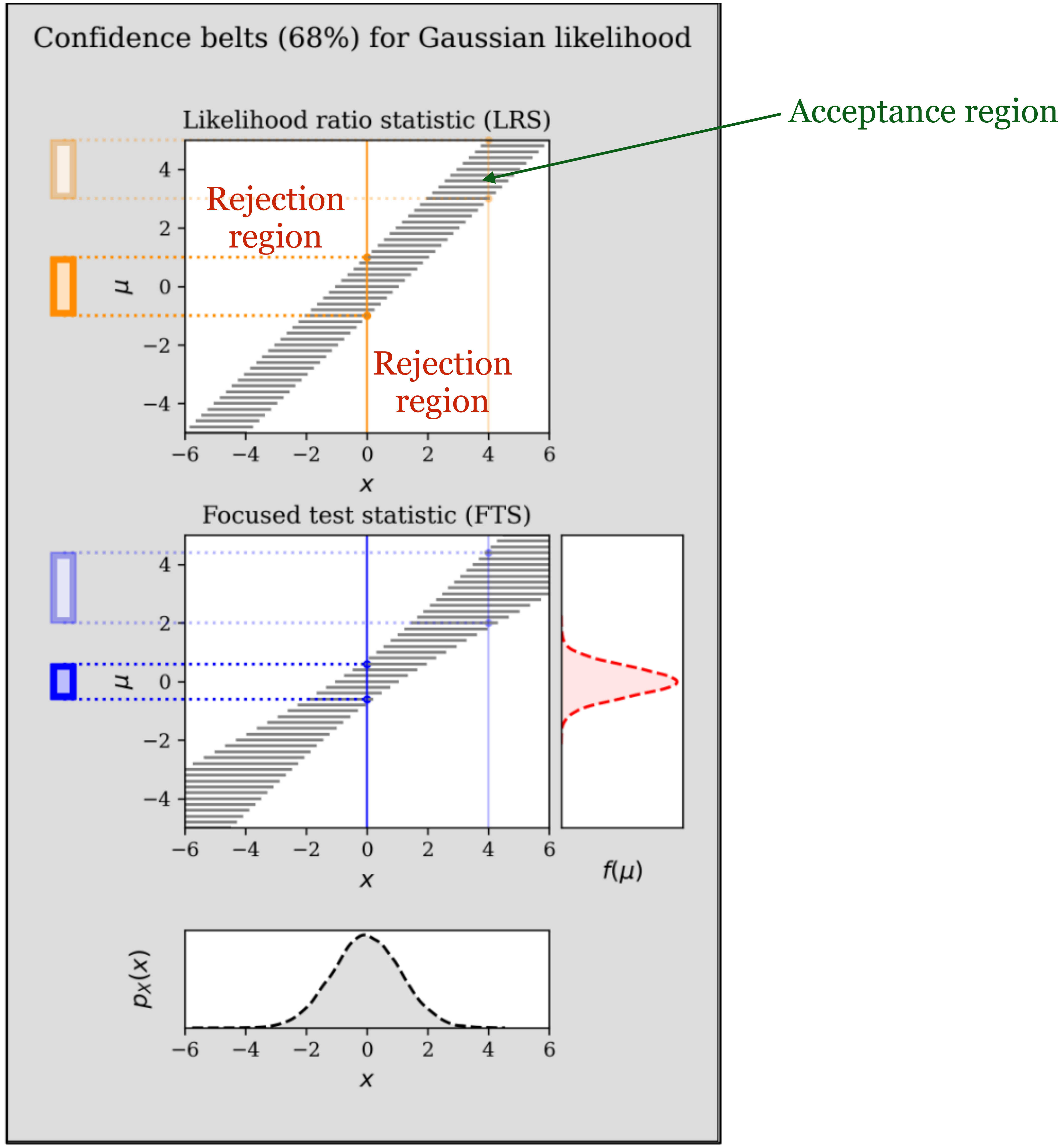
Shorter median length for confidence intervals with FTS even in 'asymptotic regime' where Wilks' applies

Test inversion for simplest measurement



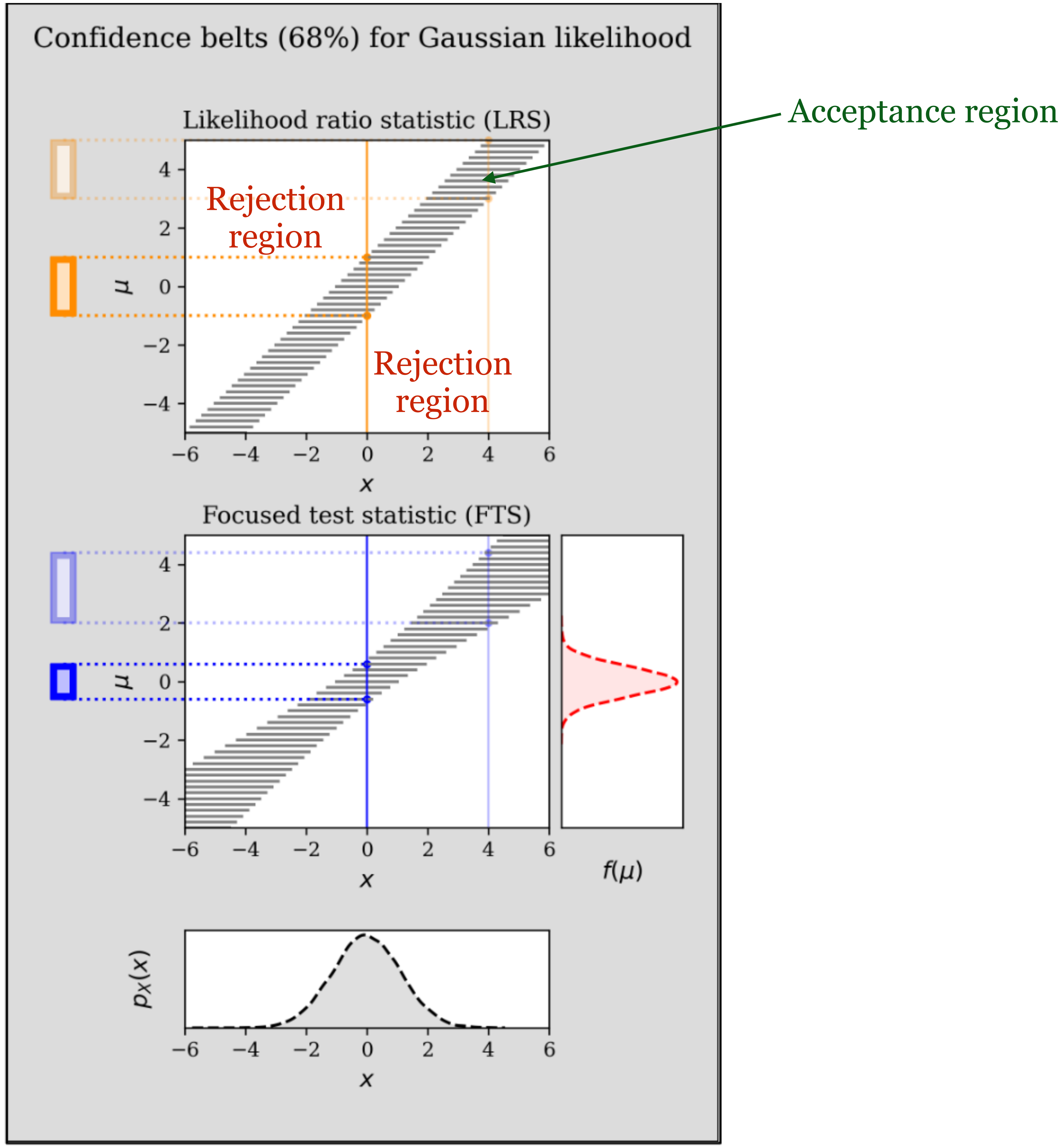
Data comprises single measurement x

Test inversion for simplest measurement



Data comprises single measurement x

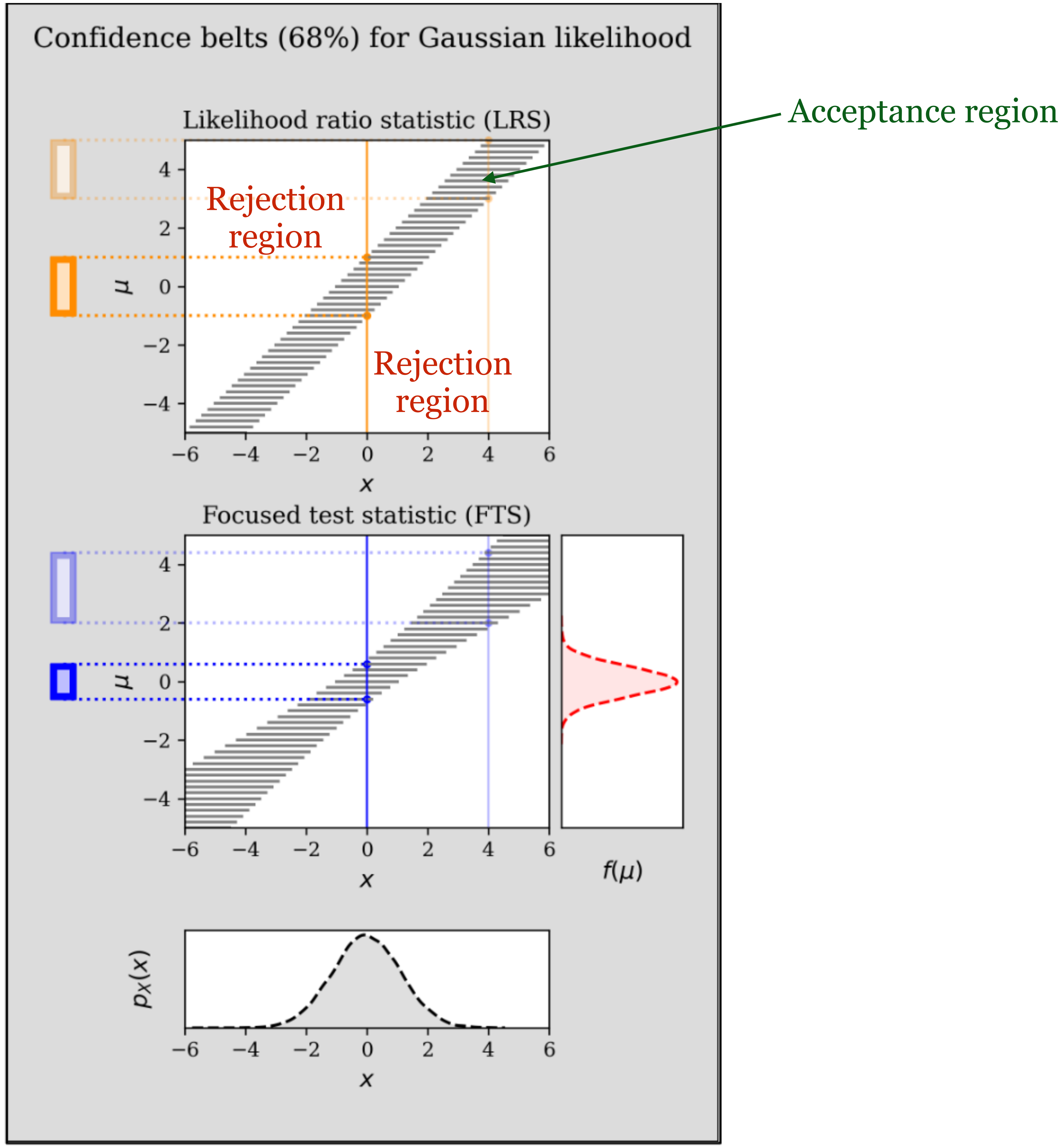
Test inversion for simplest measurement



Data comprises single measurement x

- LRS gives constant CI sizes
- FTS gives narrower near focus region by sacrificing power away from focus region

Test inversion for simplest measurement



Data comprises single measurement x

- LRS gives constant CI sizes
- FTS gives narrower near focus region by sacrificing power away from focus region

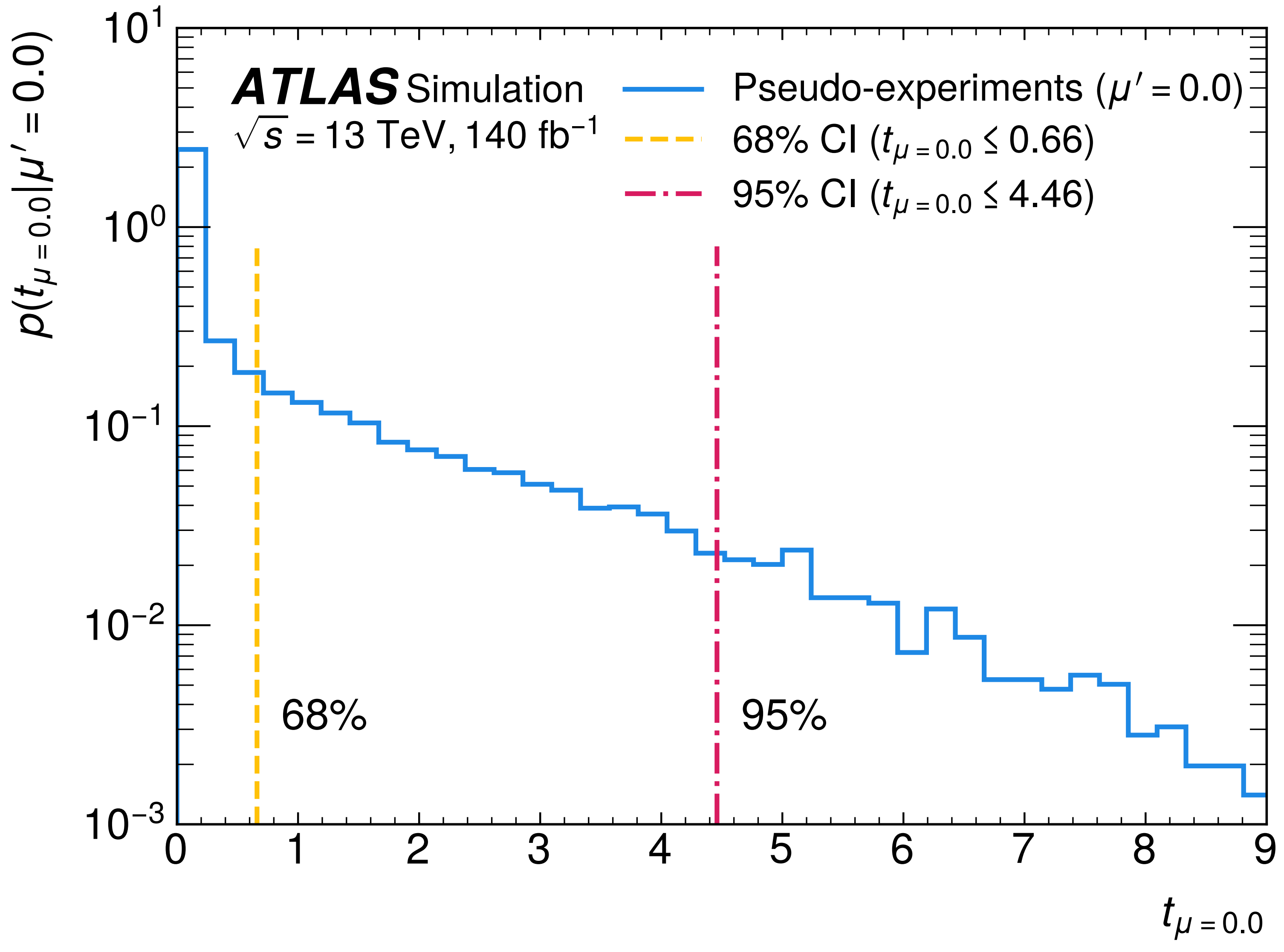
- You have full freedom to **choose any focus function** that gives you the **best expected sensitivity** in physics-motivated regions using simulated samples

Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

t_μ

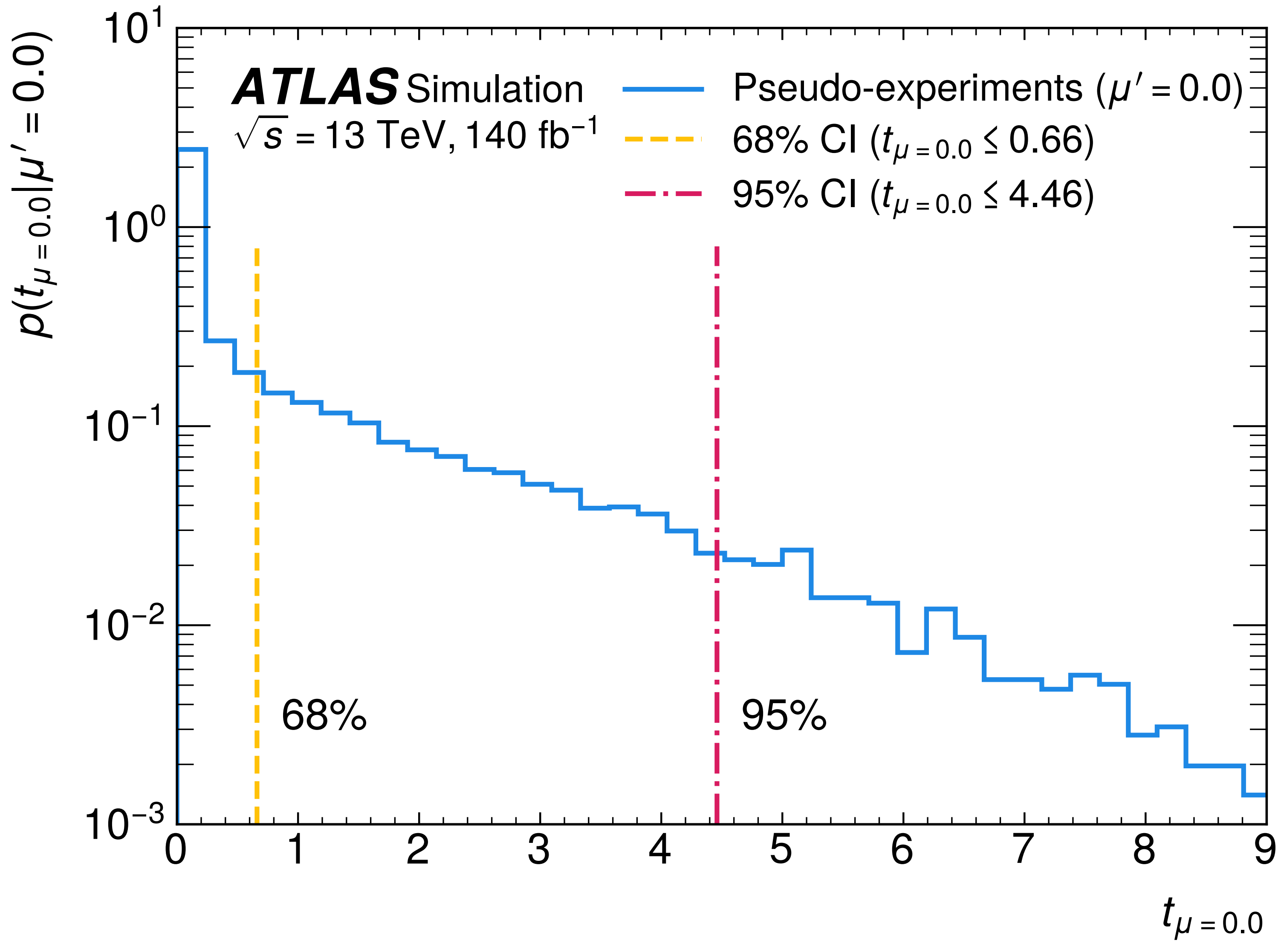
ATLAS NSBI paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

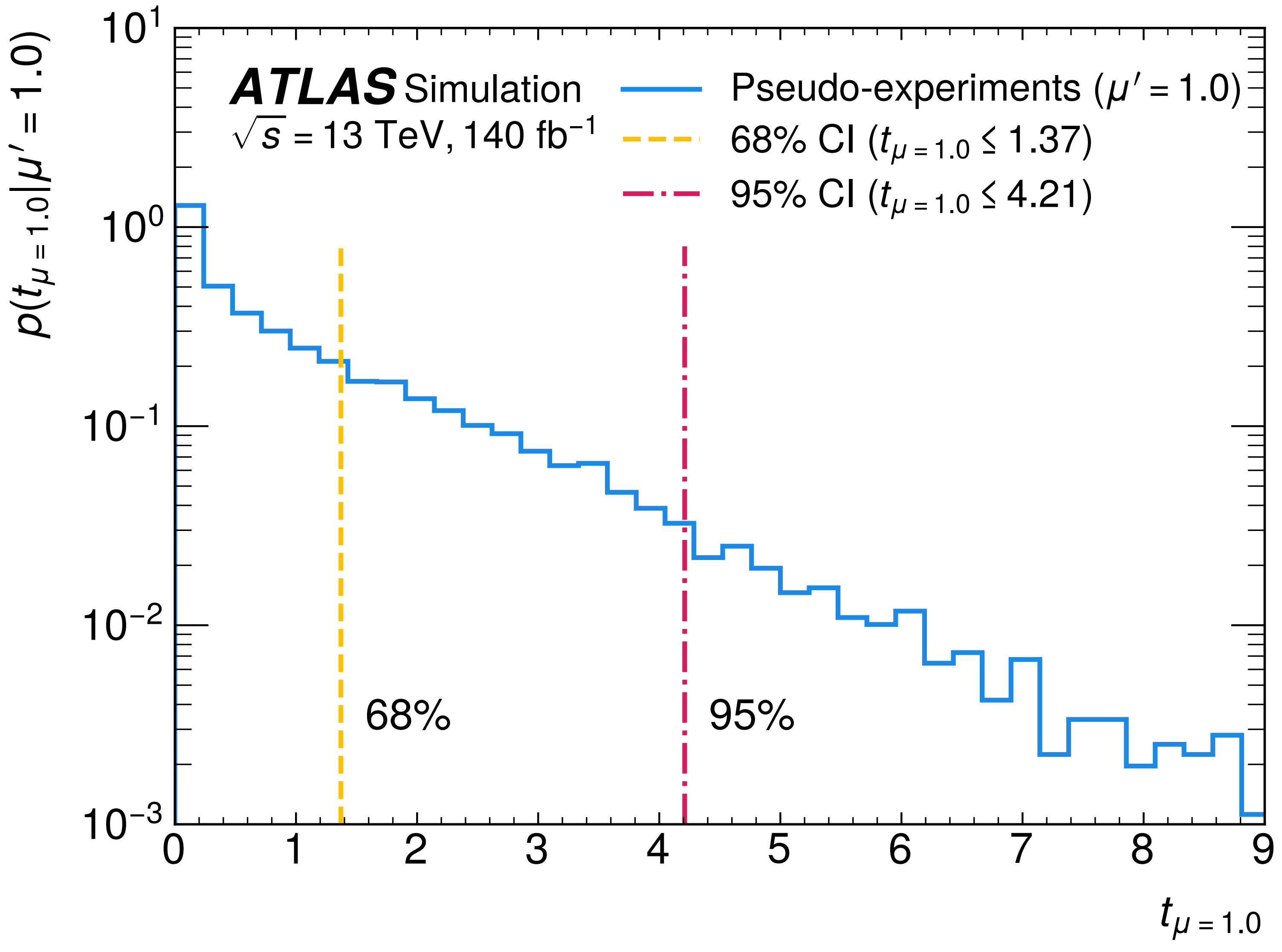
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

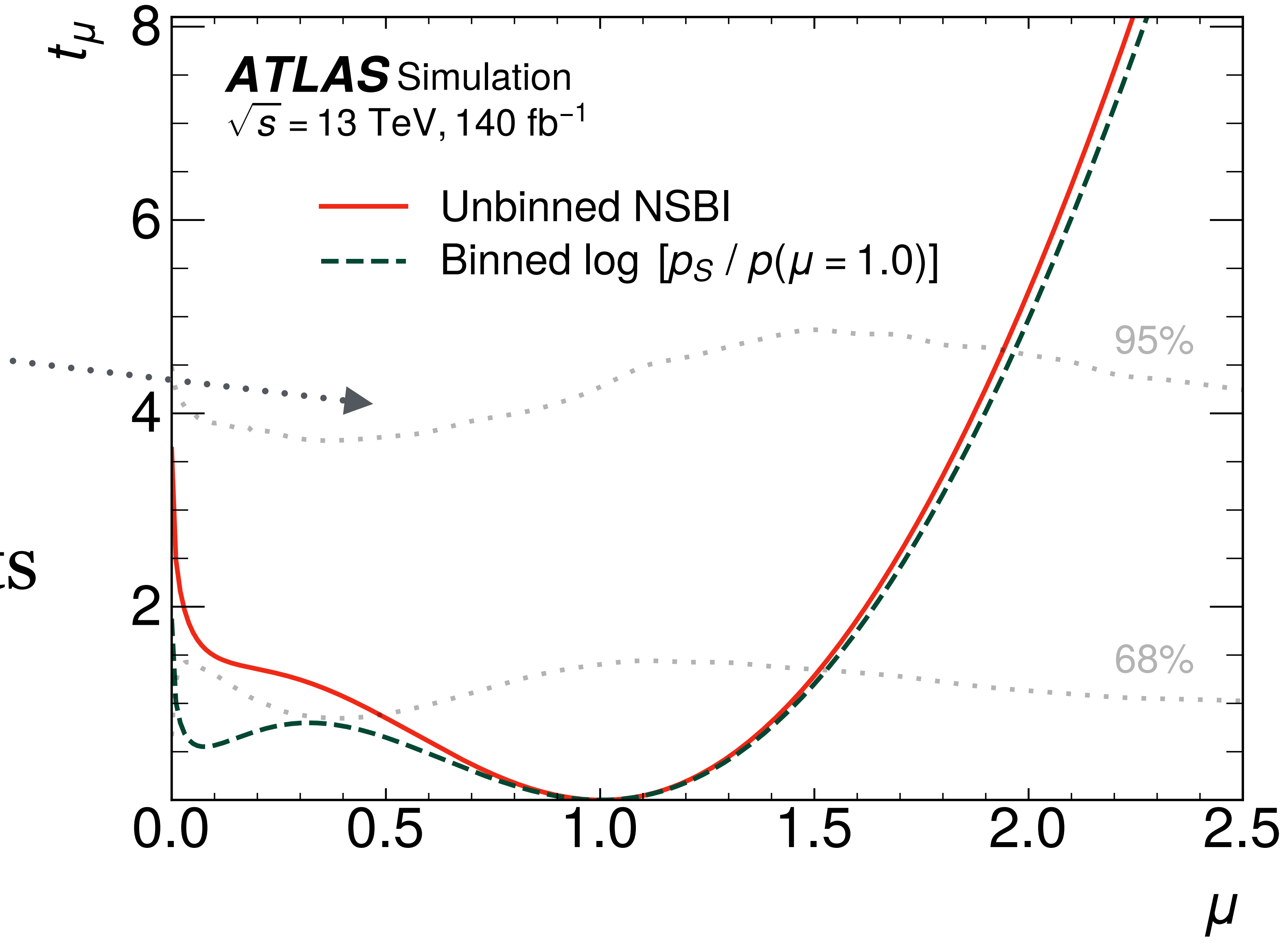
True $\mu = 1$



t_μ

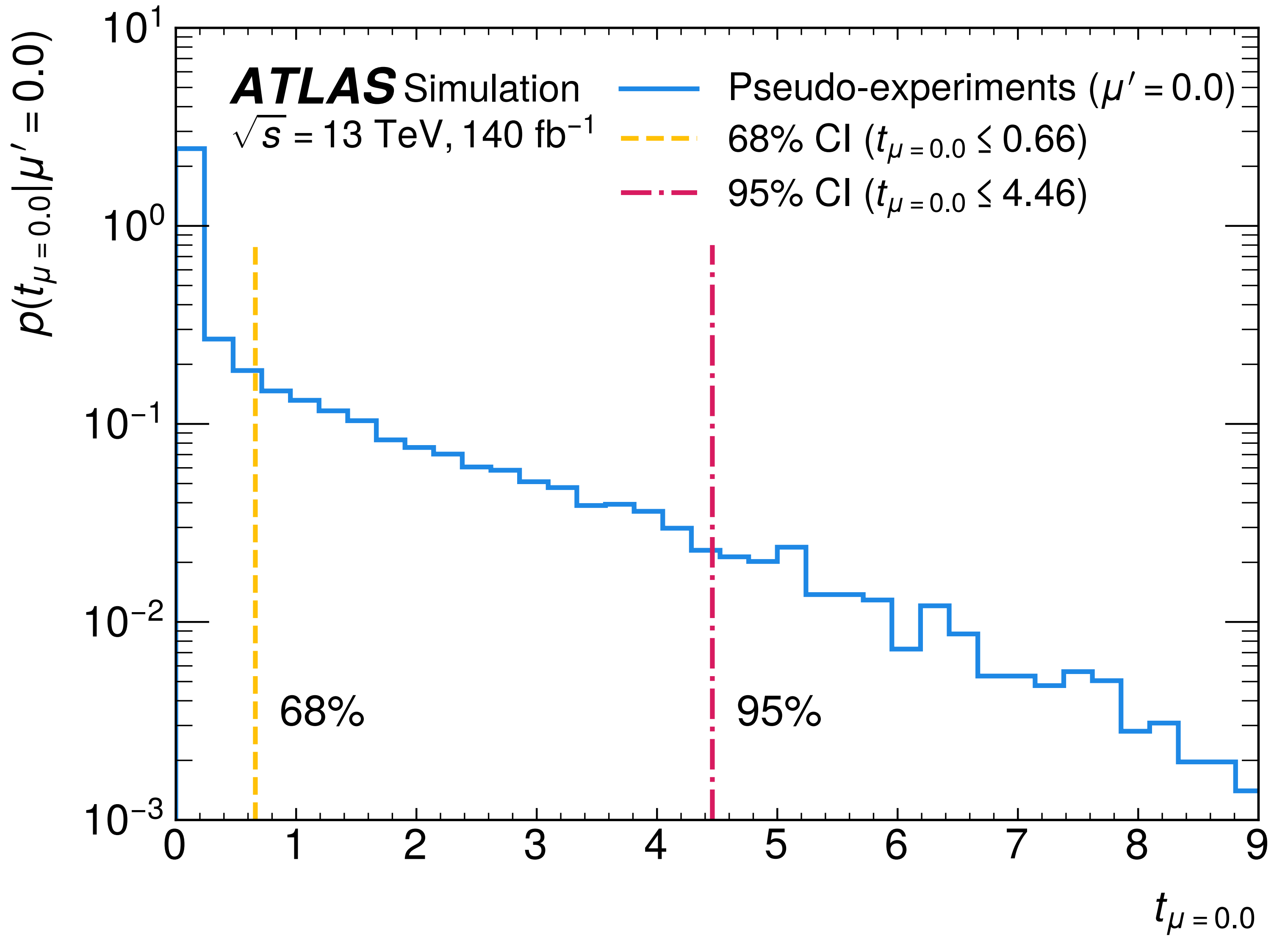
Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



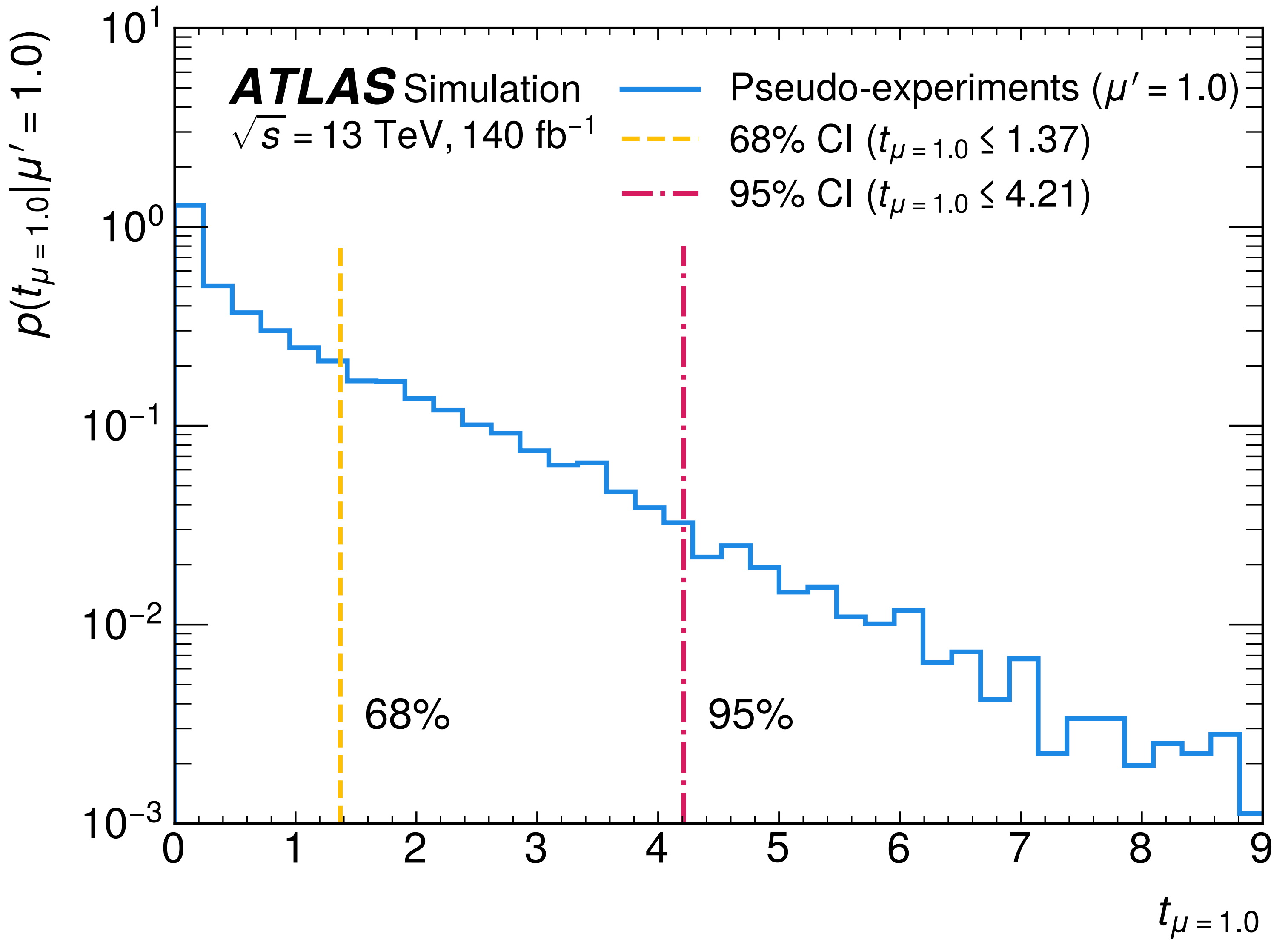
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

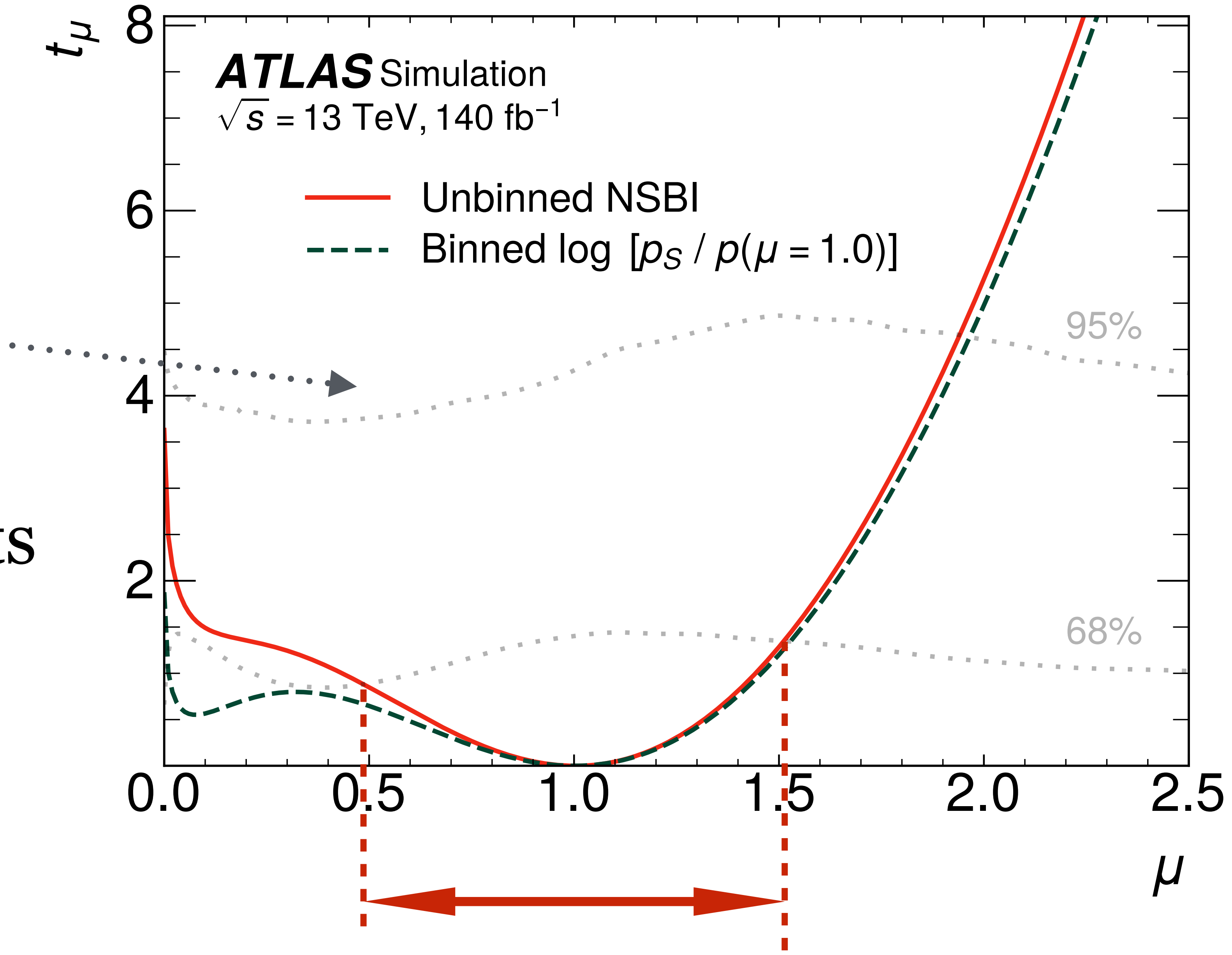
True $\mu = 1$



t_μ

Computational Challenge: Inverting the test

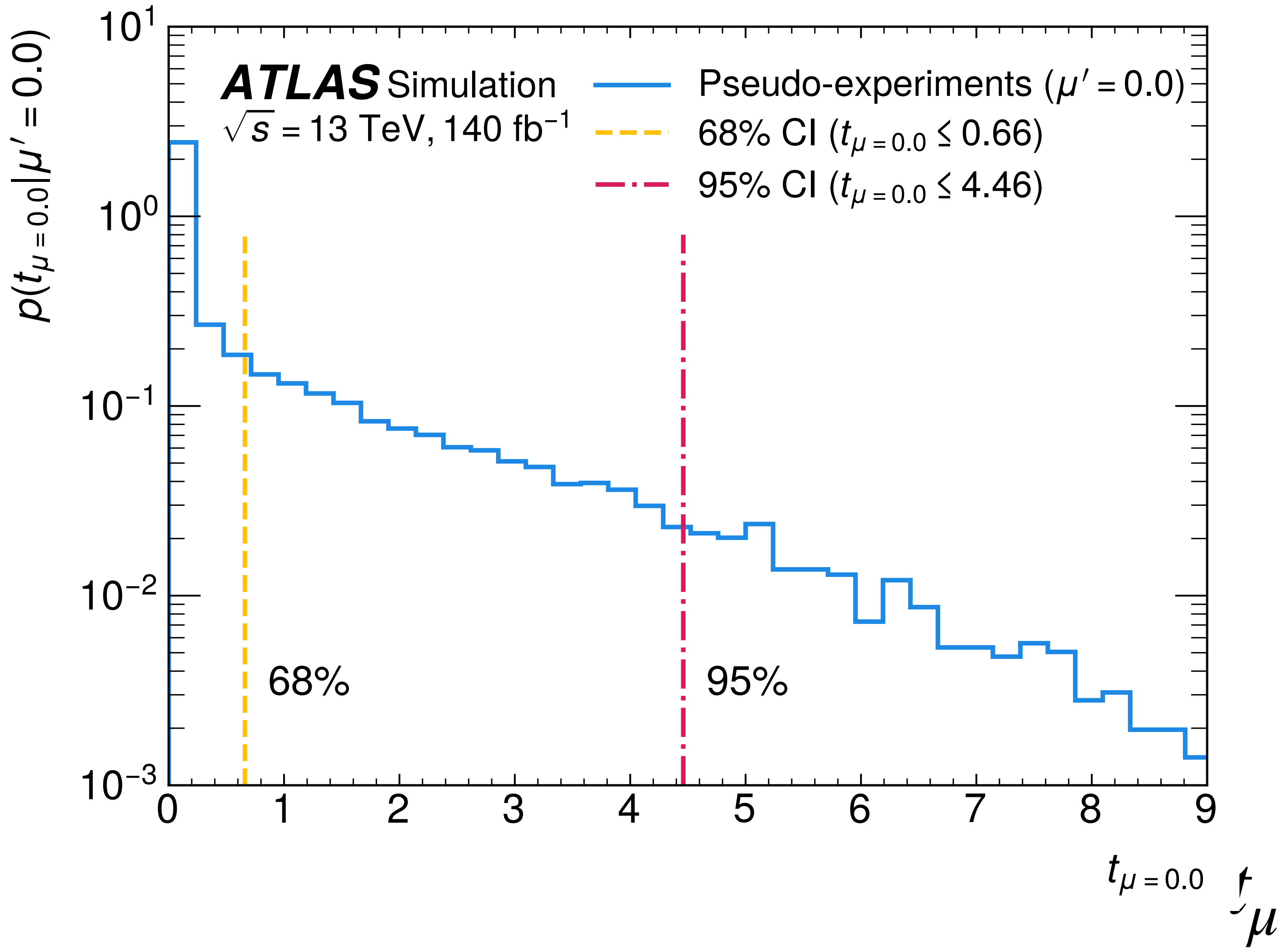
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



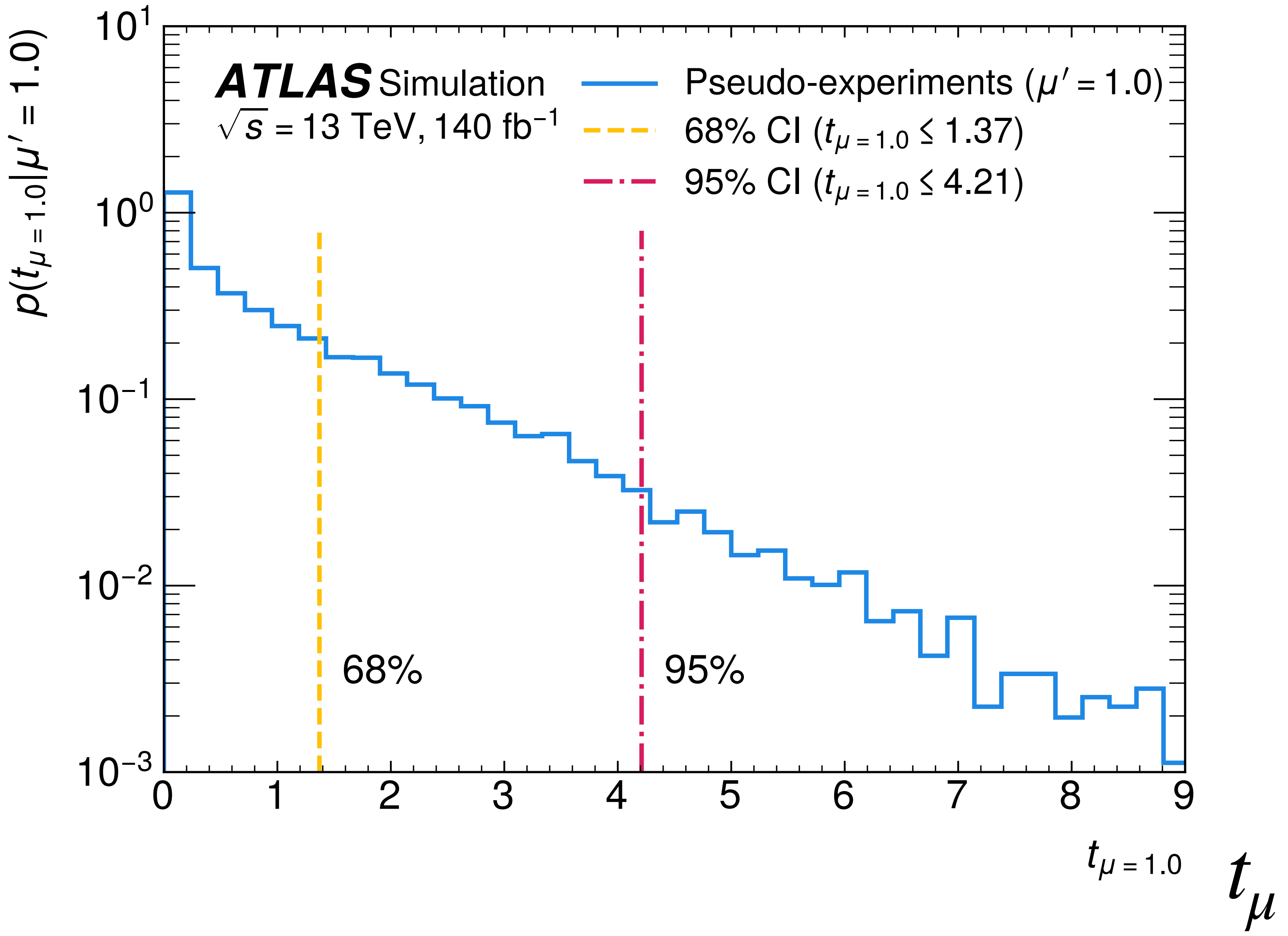
Calibrated confidence interval

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$

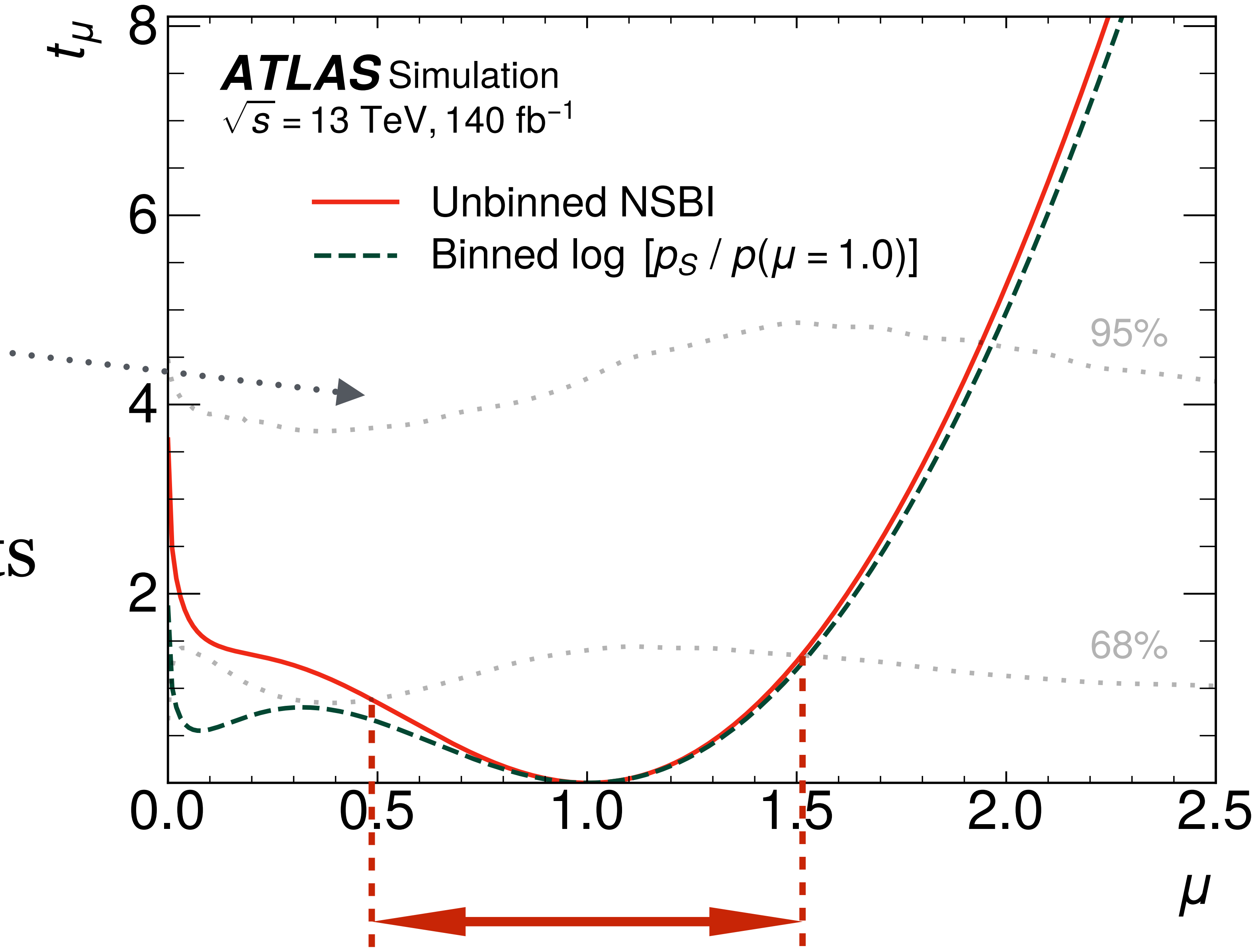


True $\mu = 1$



Computational Challenge: Inverting the test

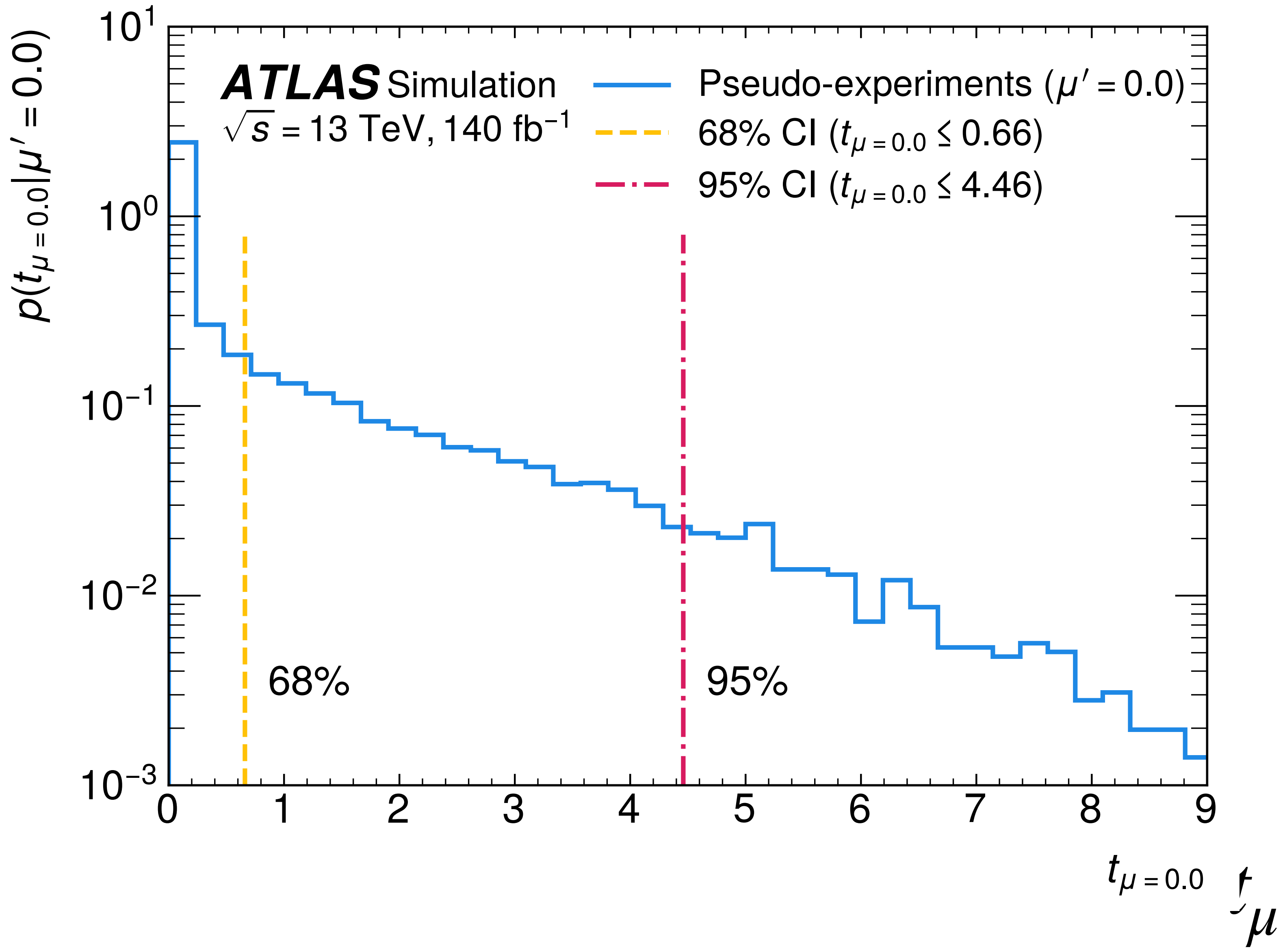
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



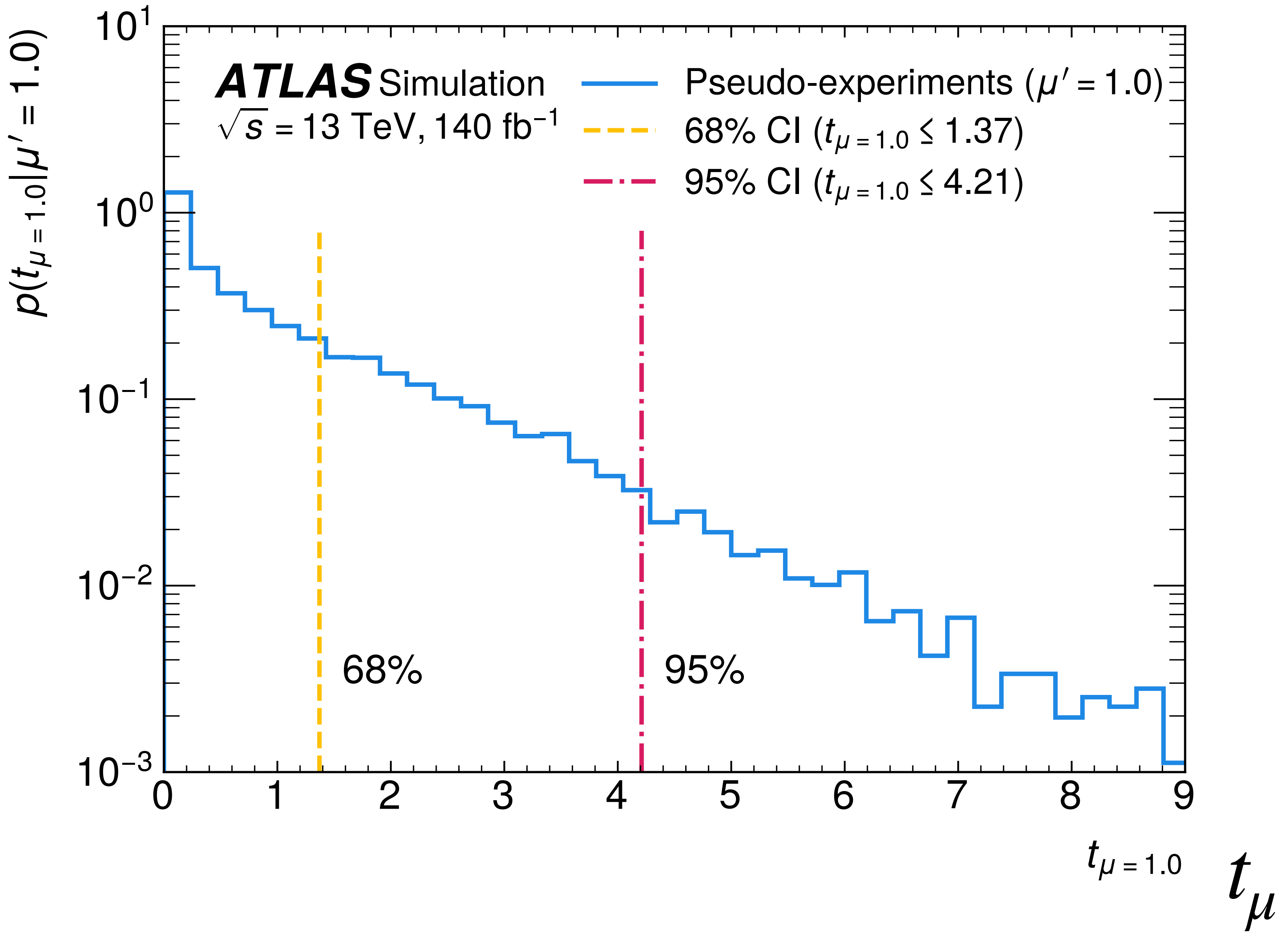
Calibrated confidence interval

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$

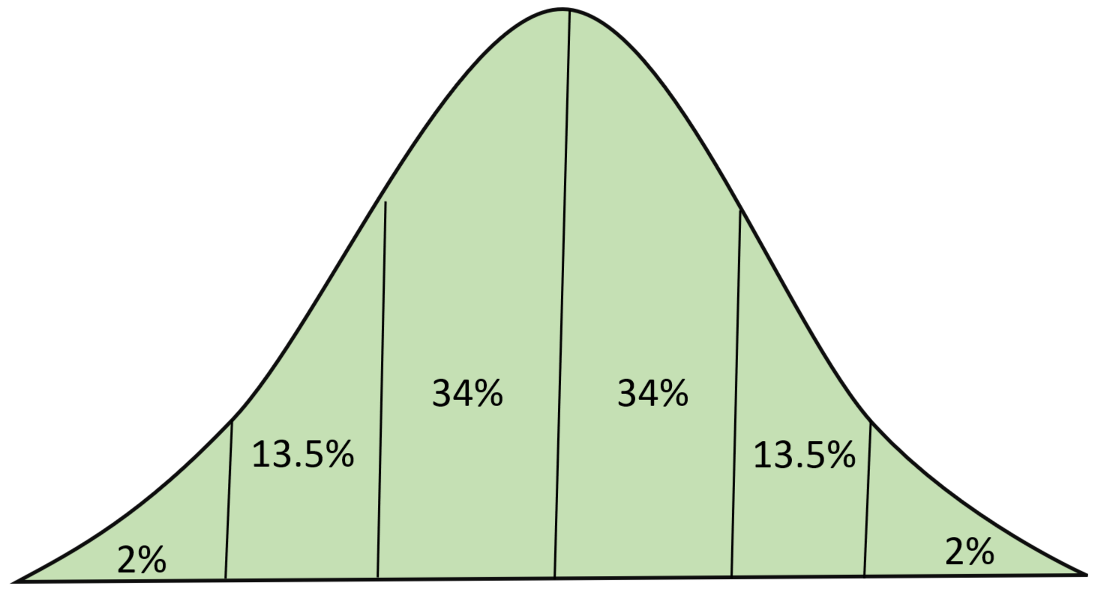


True $\mu = 1$



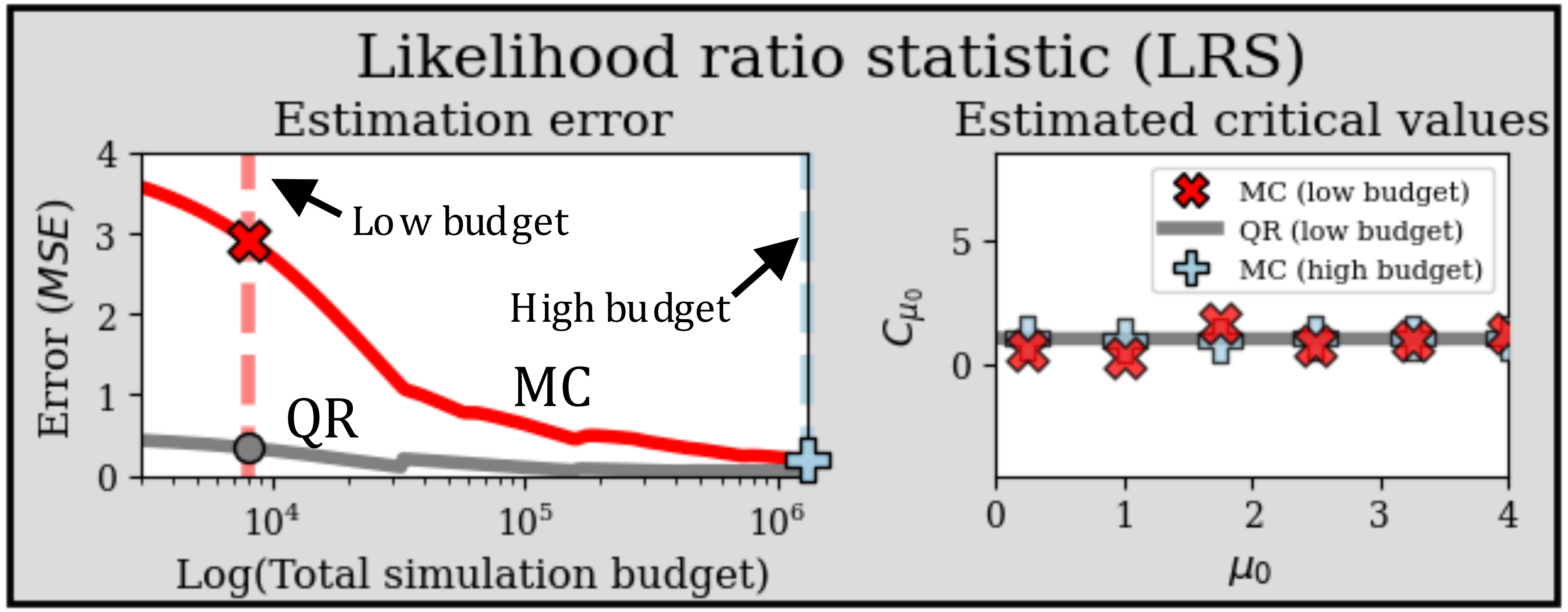
t_μ

Task is to determine quantiles



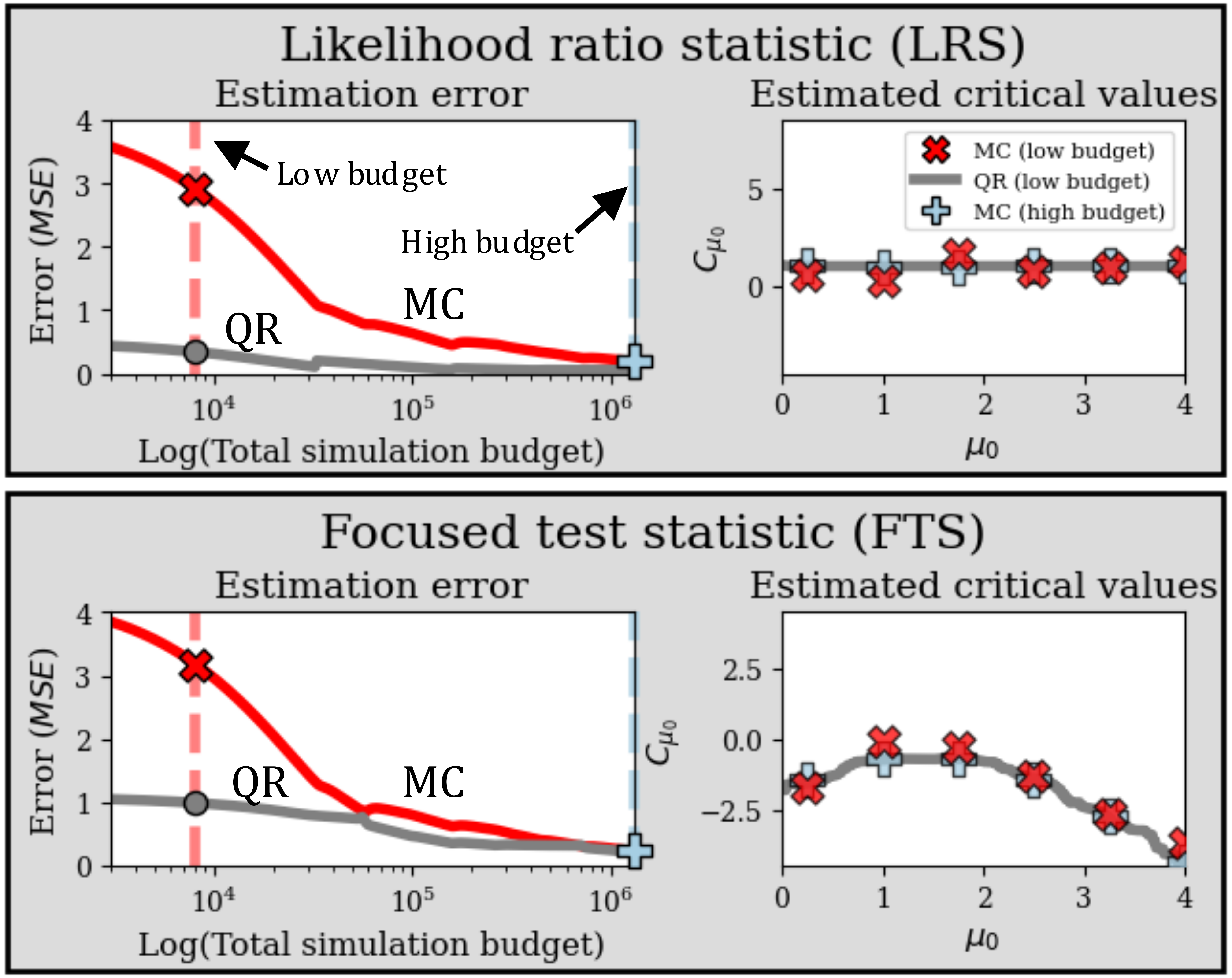
Perform efficiently with quantile regression!

Can do it faster and more accurately with ML technique known as ‘quantile regression’ (QR)



Perform efficiently with quantile regression!

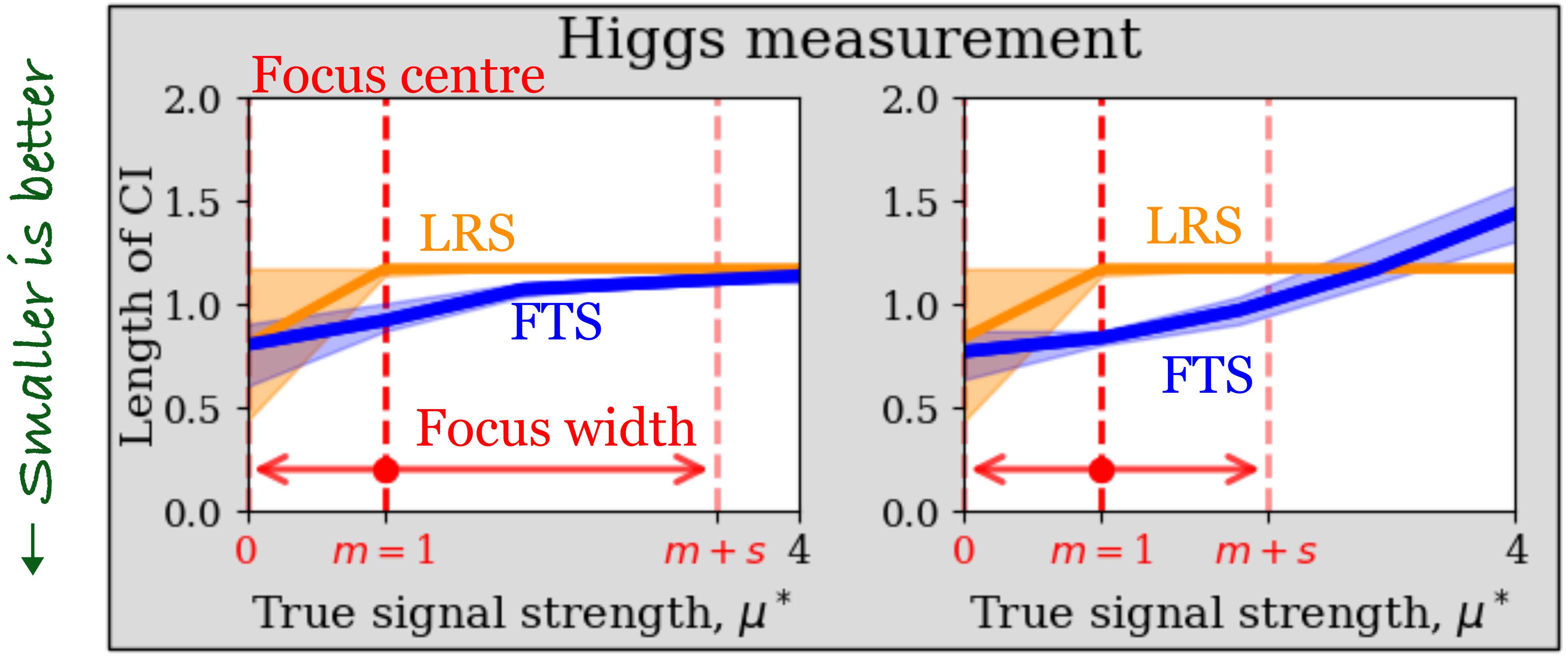
Can do it faster and more accurately with ML technique known as 'quantile regression' (QR)



Now ready to compare LRS to FTS on HiggsML benchmark dataset

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset



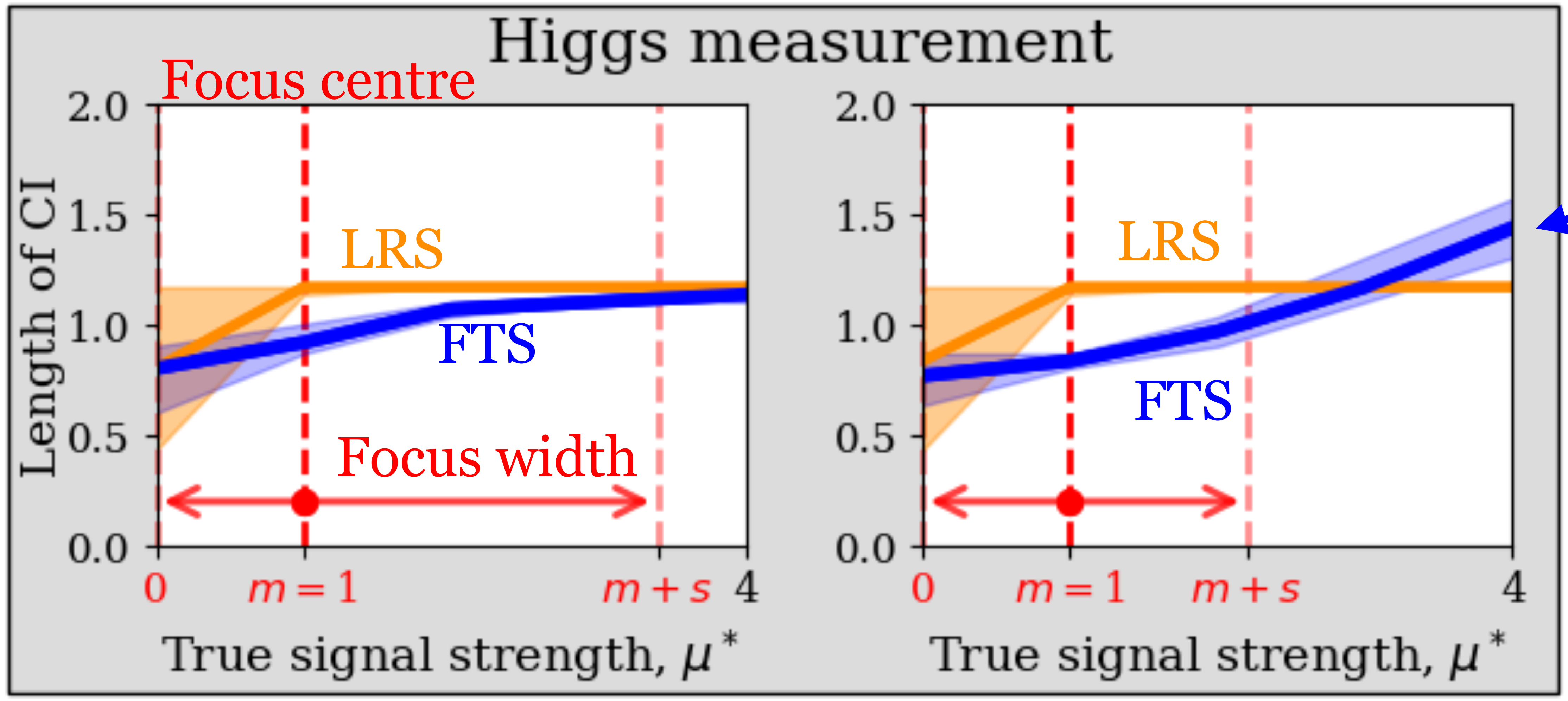
Wide focus $f(\mu)$

Narrow focus $f(\mu)$

$H \to \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



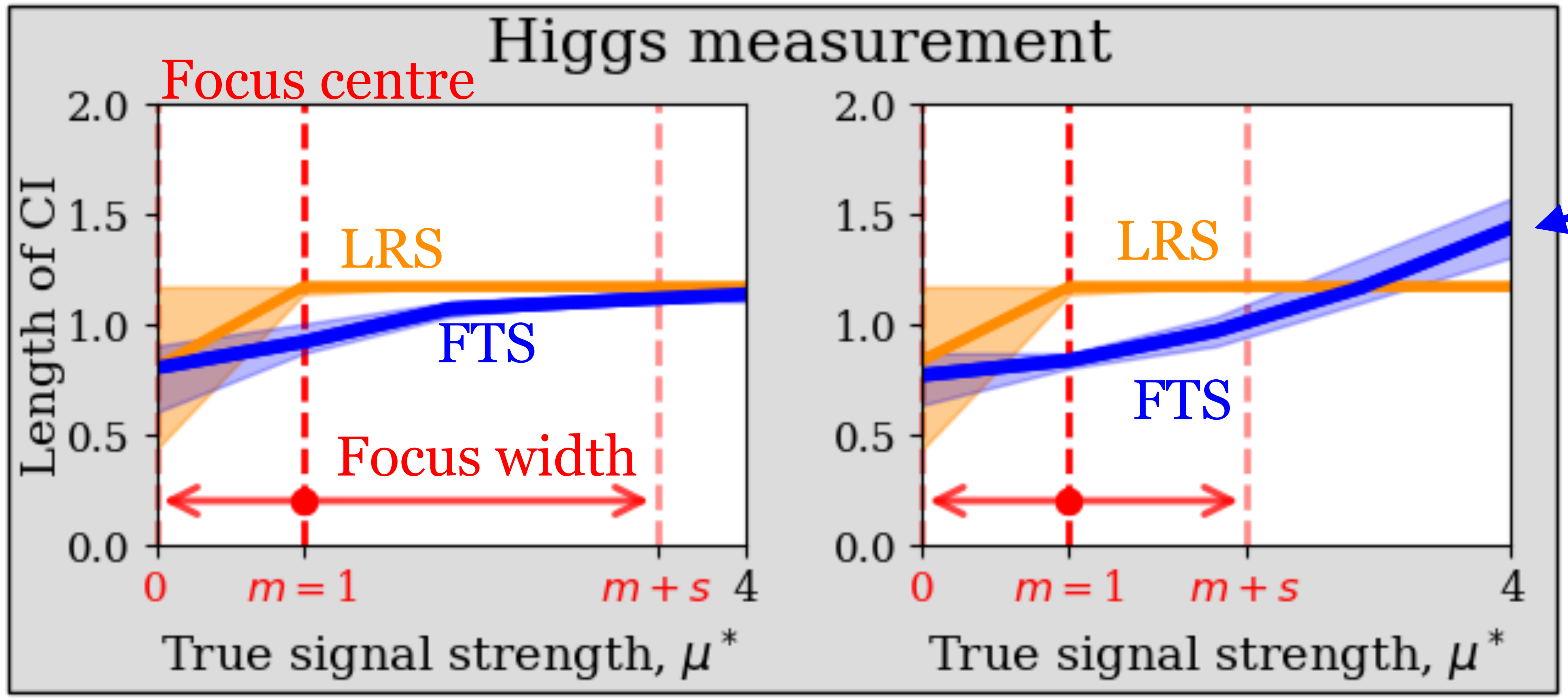
Wide focus $f(\mu)$ Narrow focus $f(\mu)$

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

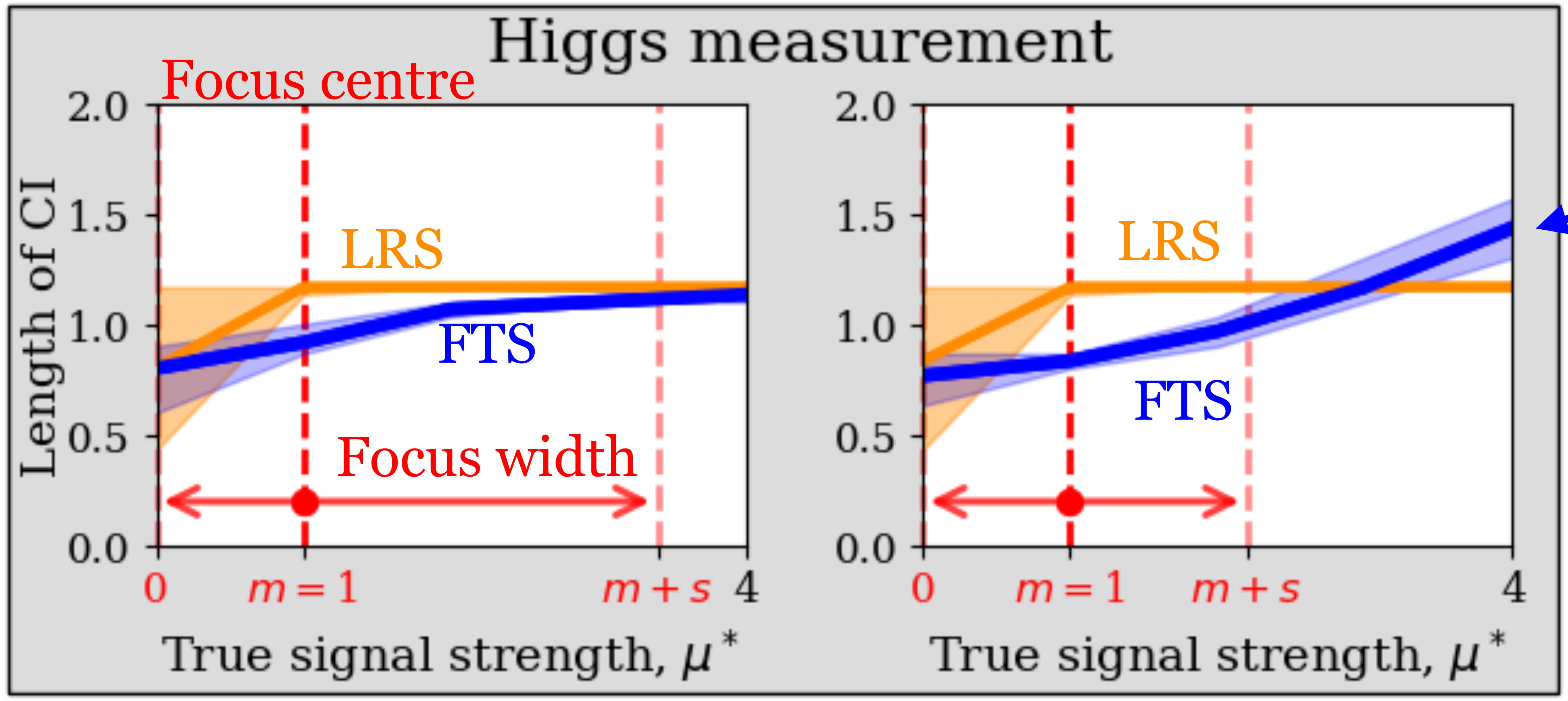
FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \to \tau\tau$ ATLAS simulated public benchmark dataset

Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

Which focus would you choose?

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \to \tau\tau$ ATLAS simulated public benchmark dataset

Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

funding gets harder to secure, principal investigators are in their office writing grants while the trainees get to do the cool stuff.

Bryan W. Jones is a retinal neuroscientist at the University of Pittsburgh in Pennsylvania.

AISHIK GHOSH
STUDENTS OVERTURN LONG-HELD ASSUMPTION

I have worked on experimental particle physics since 2015, searching for Higgs bosons at CERN, Europe's particle-physics lab near Geneva, Switzerland, and now also working on the Deep Underground Neutrino Experiment (DUNE) in the United States. For this research, there's one statistical test we've used for decades to confirm the existence of a new particle – the generalized likelihood ratio test (GLRT). This compares two models – a simple null hypothesis, which includes no new particle or matter being discovered, and a more complicated alternative model, which includes a new particle with many possible values of strength.

In December 2024, a couple of PhD students working with my collaborator, Ann Lee, a data scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania, were confident they could disprove the assumption that the GLRT was optimal. In the corner of my mind, I hoped they would prove us wrong. I gave them one of the most famous Higgs boson data sets to play around with. By early 2025, they showed that, although our previous physics results weren't wrong, our use of the GLRT wasn't ideal because it assumed large sample sizes are always generated, which is often not the case. Instead, the test left valuable information on the table. That day was special. I was still sceptical and I went through a battery of checks because I had to go back to my community and defend the PhD students' work, but it was all correct. The paper is currently in review, receiving a great deal of scrutiny.

Together, we produced a statistical test that will drastically improve our ability to make discoveries in particle physics, for example in searches for a new particle such as dark matter, where we expect to see only a few signal events at best. As a scientist, I want deeply held beliefs to be questioned. It was a real shock to the particle-physics community. Young people find it exciting. Senior members are still highly sceptical, as they should be, but they are coming around. As the DUNE experiment comes online, with this new statistical model in place, we hope to make precise measurements about neutrinos much sooner than anticipated.

Aishik Ghosh is a fundamental physicist at the Georgia Institute of Technology in Atlanta.

RAFIK TAREK NEME GARRIDO
SHOCKING CORAL FIND

A couple of years ago, after a day of pouring rain, the water on the Caribbean coast of Colombia was crystal clear and my master's student, Jorge Mareno, managed to take pictures of corals that no one knew existed here. We could find no scientific reports of corals in the area. Typically, the water is pretty turbid because the Magdalena River, which flows from the south of the country to the Caribbean Sea, brings chemicals and pollutants. It's an ongoing ecological and social challenge, but these corals must be adapting to these conditions. We did a sampling campaign across three days with a boat, using environmental DNA to find areas where corals, sponges and fish successfully survive the conditions. Most of the records are completely new for the region. It's super gratifying.

Rafik Tarek Neme Garrido is an evolutionary biologist at the University of the North in Barranquilla, Colombia.

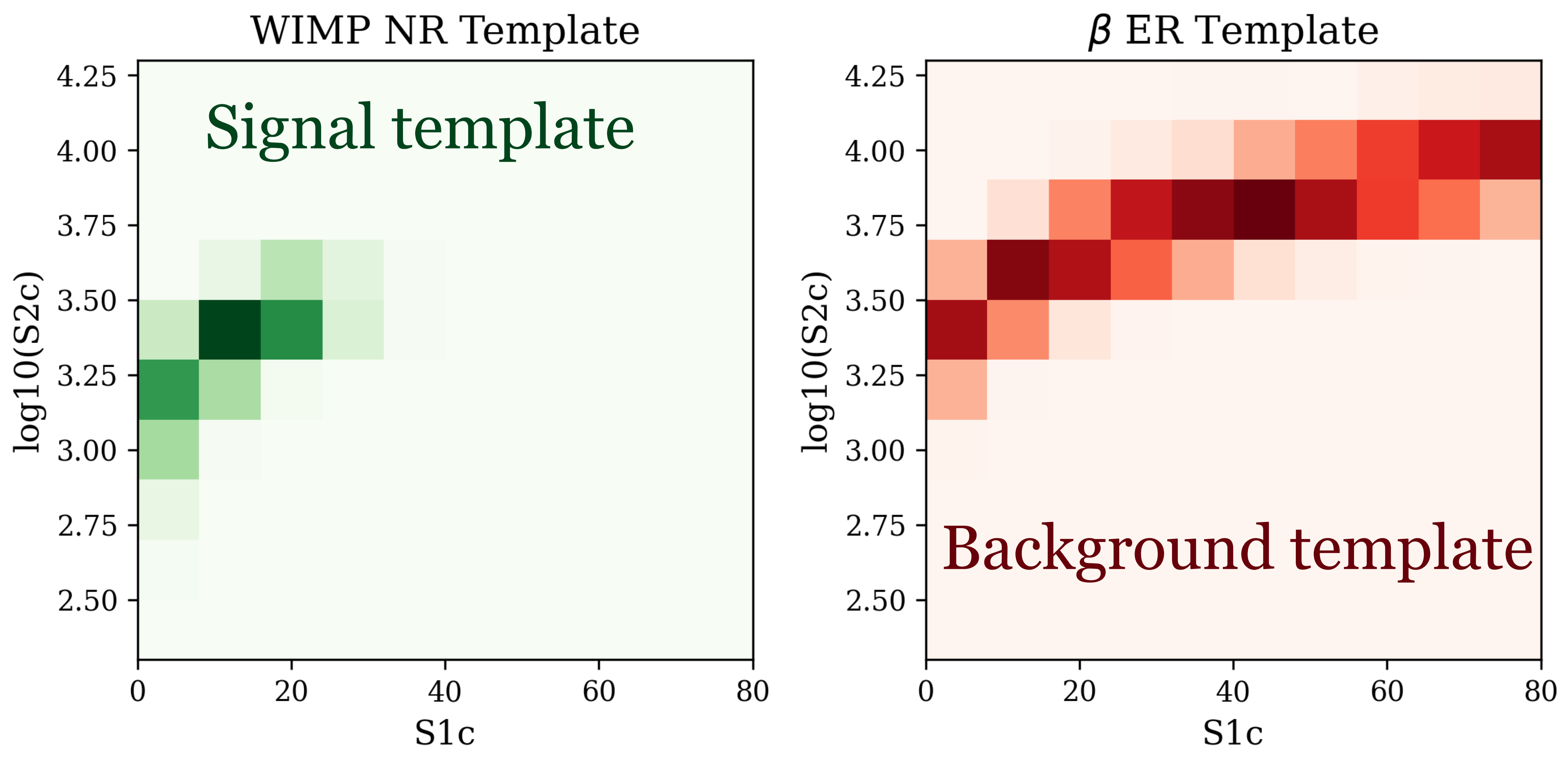
TIM CURRAN
BURN PREDICTIONS

In my group, we test the flammability of plant species using a barbecue. The results can help with fire-mitigation policies and with understanding the evolution of flammability. As part of an outreach activity, we host schoolchildren at the university who haven't had much exposure to academia before. We ask the kids to predict how a particular plant species will behave – for example, what characteristics will make it burn less or more – and then we see who is right. The kids get really into it. They ask amazing questions, the same kind that peer reviewers have asked us, including questioning our methodological assumptions, such as "why do you only blowtorch them for ten seconds?"

Most of the really good days doing science have been associated with young students having a light-bulb moment. In the rather

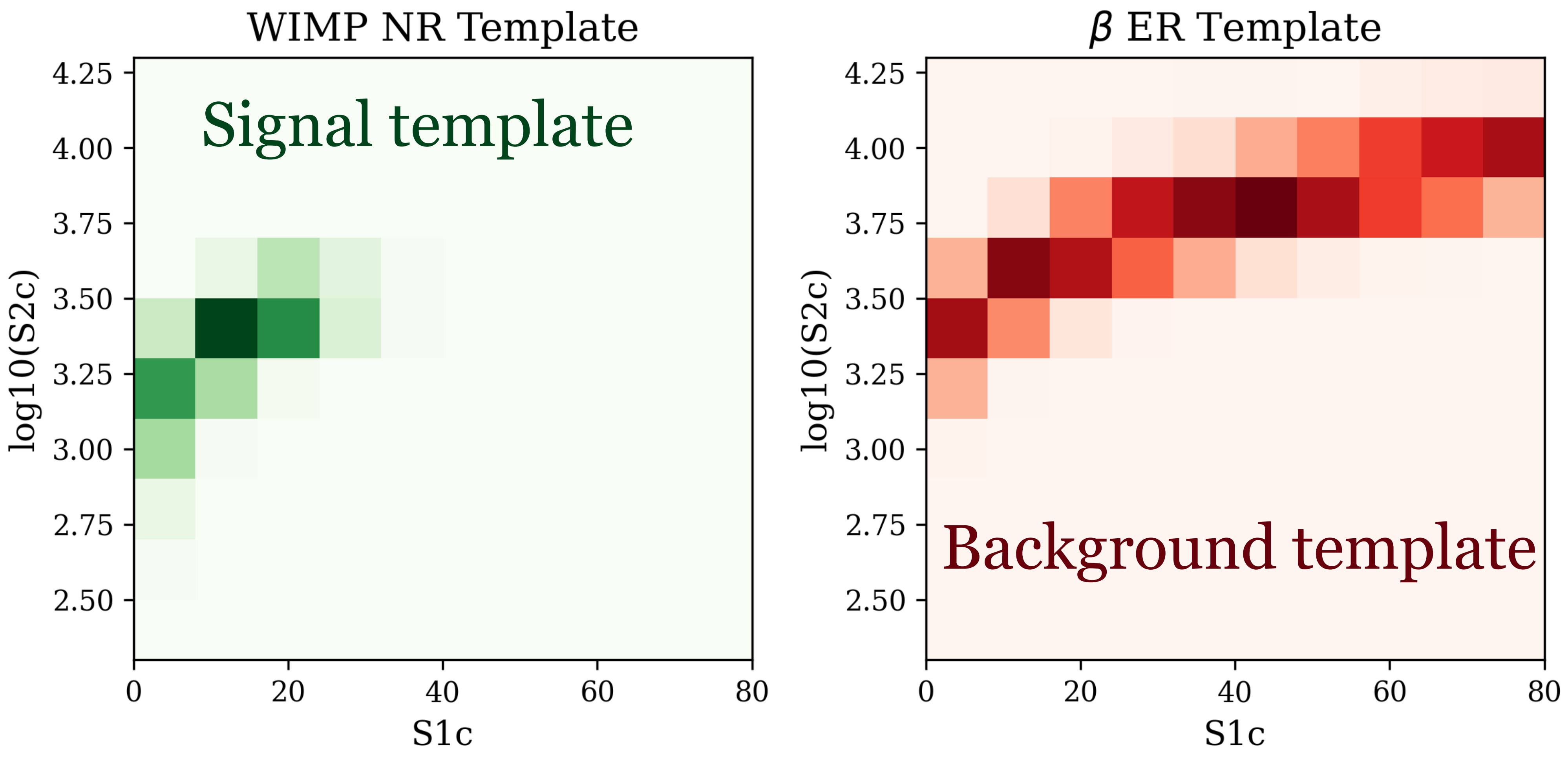
https://www.nature.com/articles/d41586-026-01479-8?utm_source=x&utm_medium=social&utm_campaign=nature&linkId=61827825

Second case study: Dark matter search in LZ experiment



Simulated to mimic [LZ data]

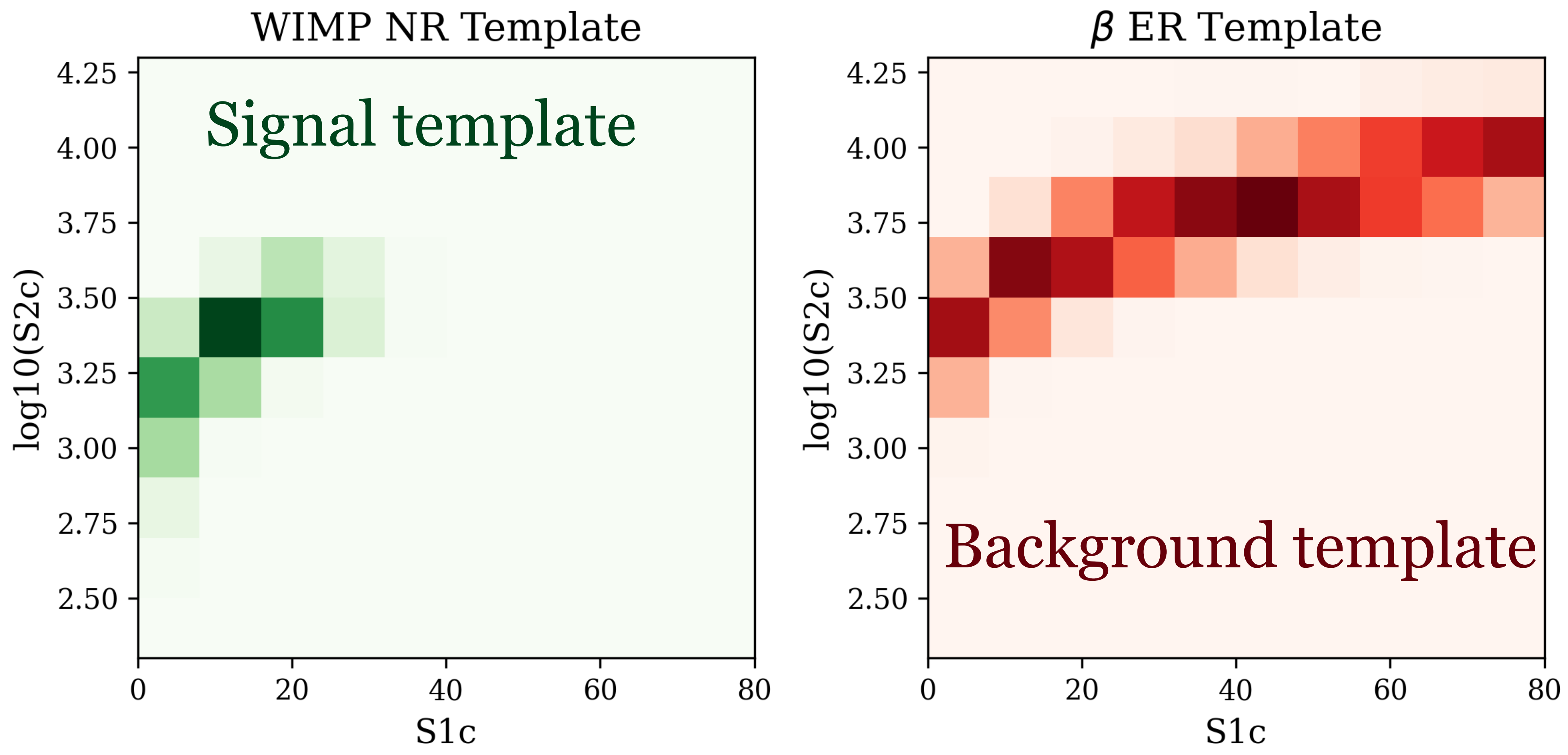
Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

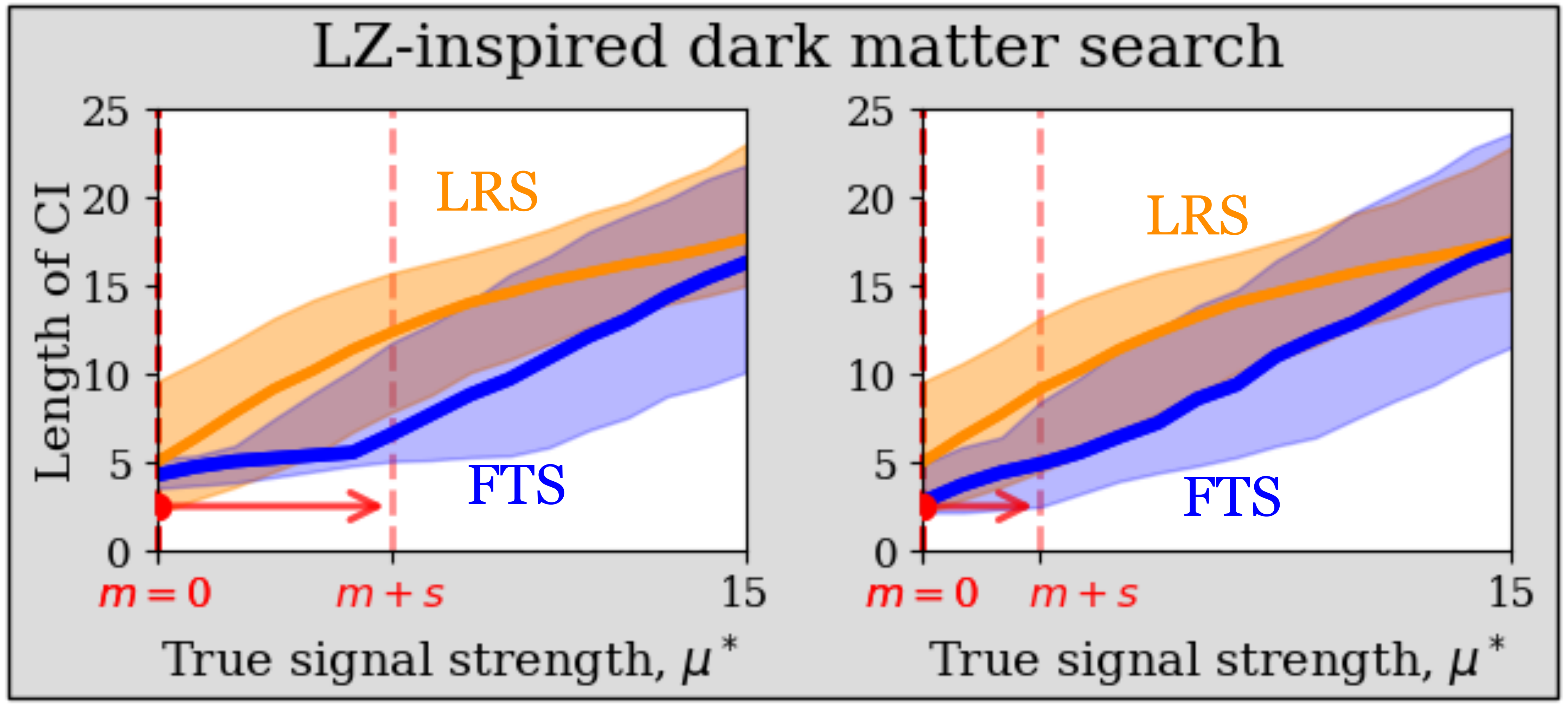
Focus now placed at $\mu = 0$, we want to set upper limits on DM signal

Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

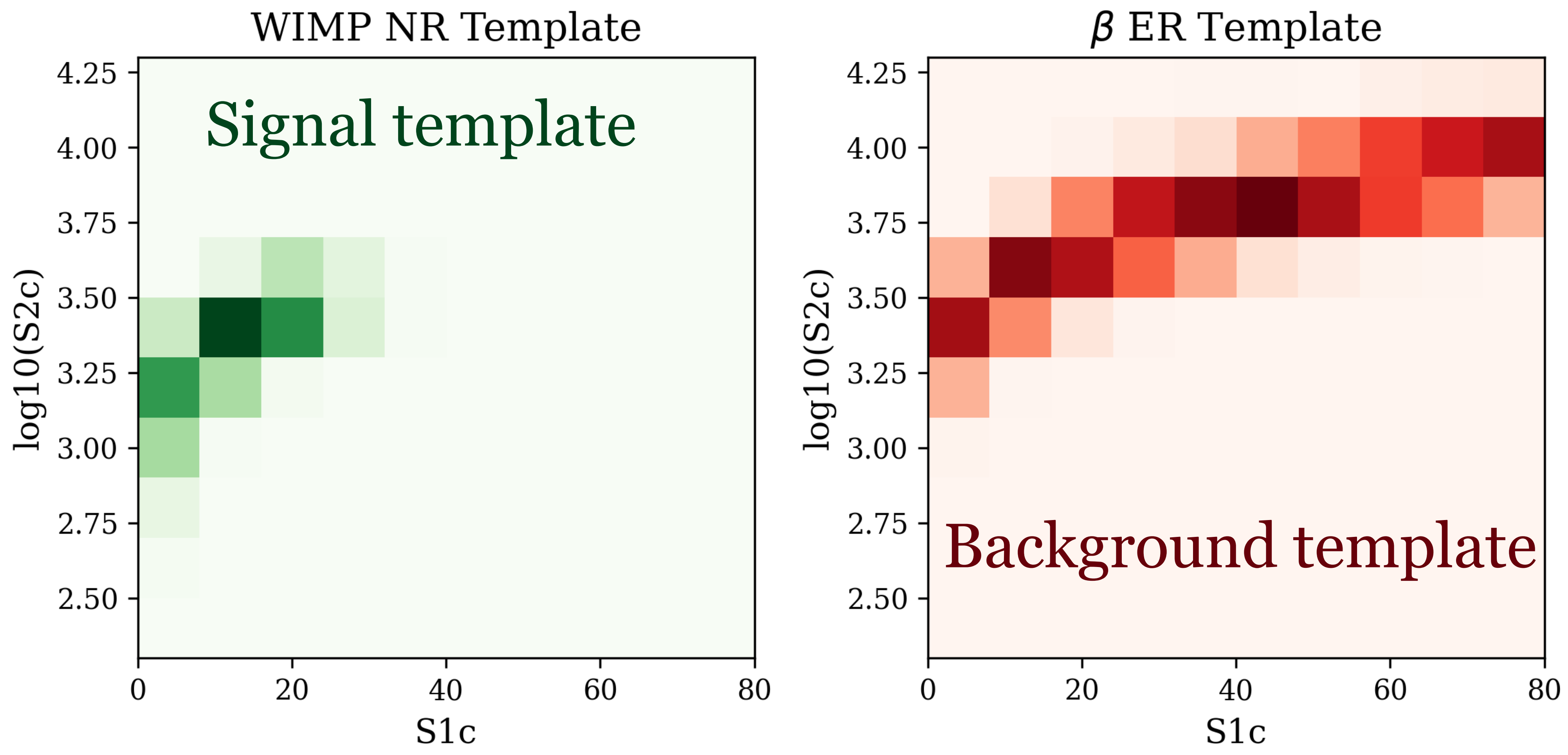
Focus now placed at $\mu = 0$, we want to set upper limits on DM signal



Wide focus

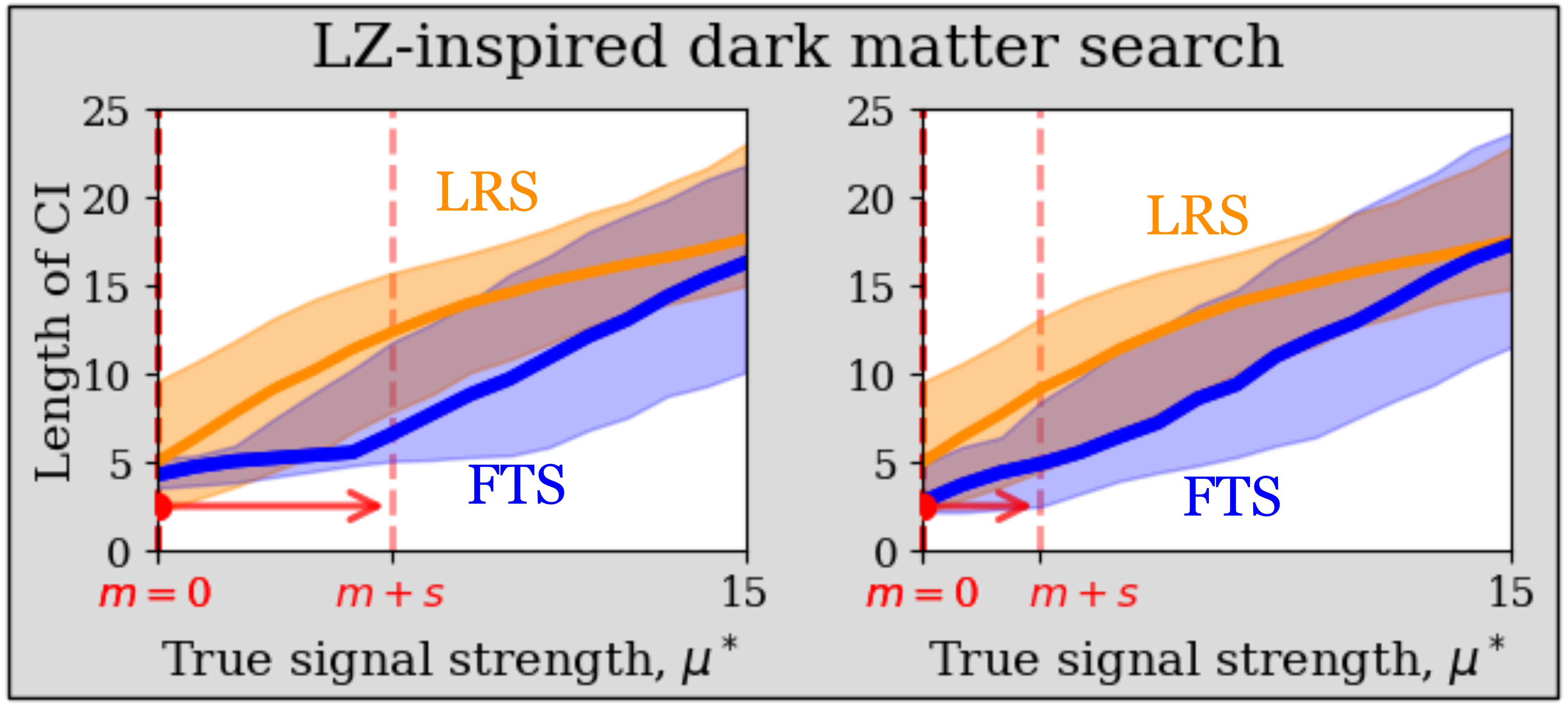
Narrow focus

Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on DM signal



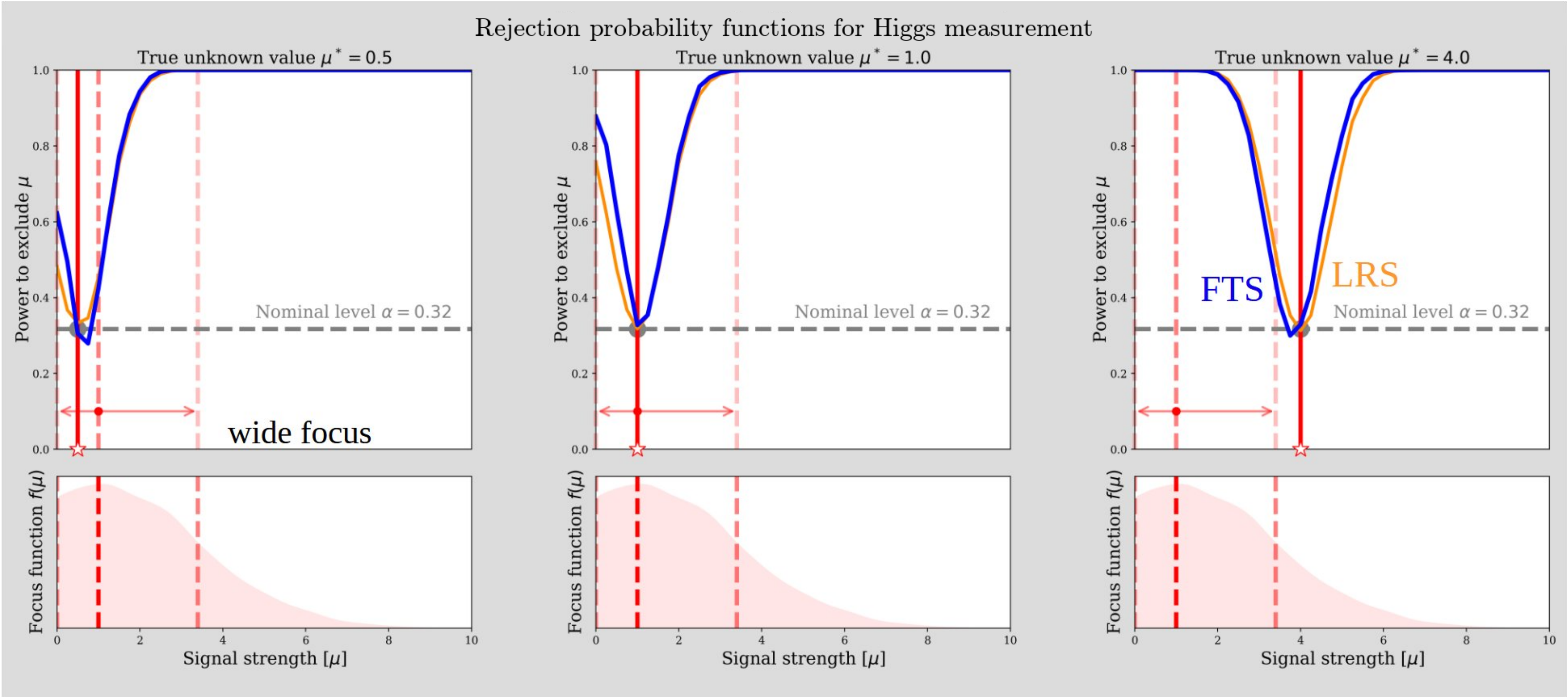
Wide focus

Narrow focus

FTS gains stable over large parameter space for both focus functions

More precision even when there is actually a signal in the observed data ($\mu > 0$)

Introducing a technical bias does not compromise coverage



Bias: FTS could accept some other value of μ more frequently than the true value

Coverage: CIs have correct local coverage because we enforced it by construction

Any history of biased tests in particle physics?: The CLs method

Even for point estimation in high dimensions, one may prefer a low variance estimator over an unbiased estimator

FTS sets better lower bounds even with a focus centred at $\mu = 0$

FTS sets better lower bounds even with a focus centred at $\mu = 0$

Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn't

Lower bounds when there is a signal:

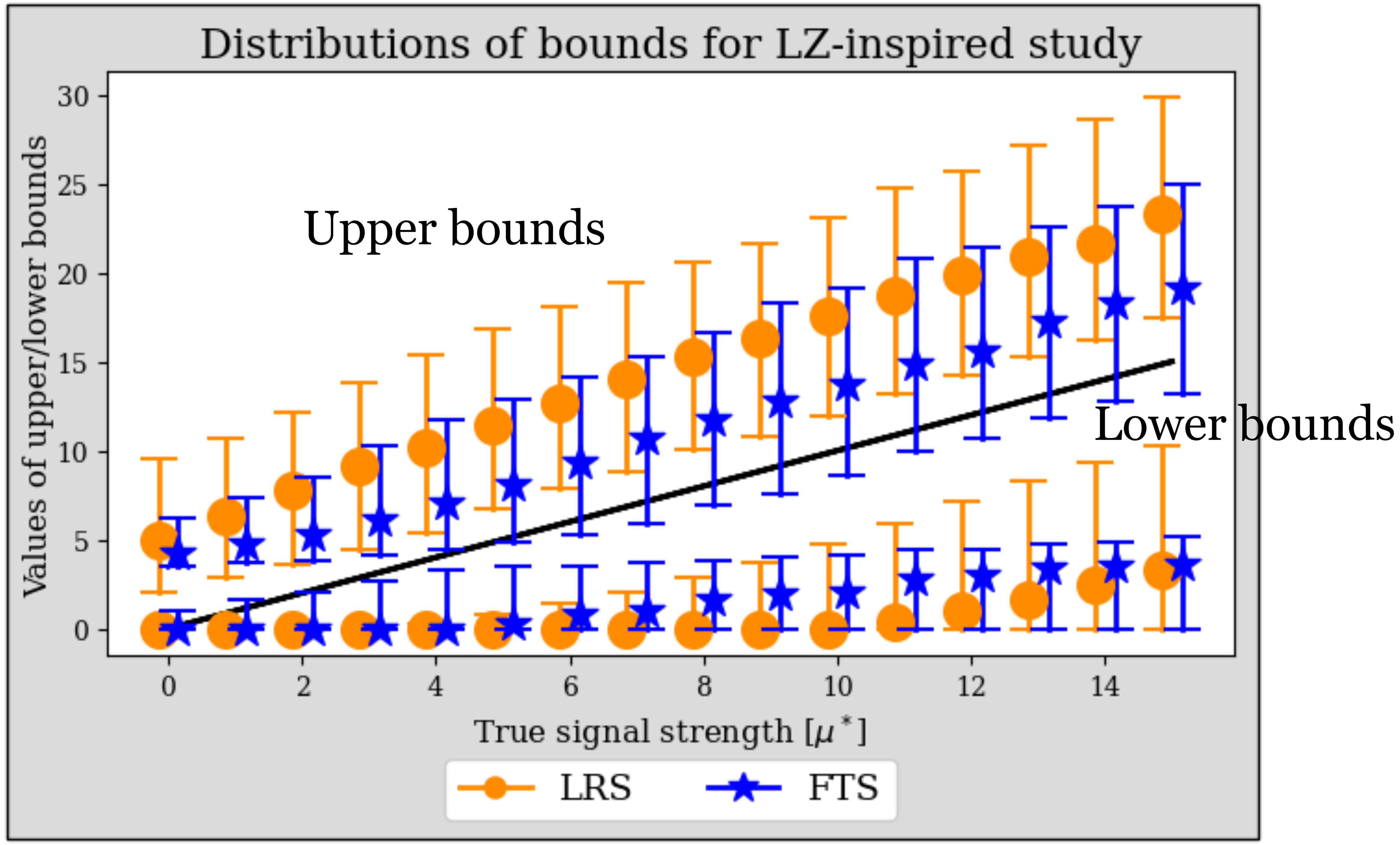
μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

Upper bounds when there is no signal:

μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

Numbers represent length of confidence intervals for 1σ (2σ)

FTS sets better lower bounds even with a focus centred at $\mu = 0$



Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn't

Lower bounds when there is a signal:

μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

Upper bounds when there is no signal:

μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

Numbers represent length of confidence intervals for 1σ (2σ)

Conclusion

- Semi-parametric NSBI let's us maximally extract information in high-dimensional unbinned data
- FTS + QR let's us robustly **outperform the likelihood ratio test,** by focusing power in physics-motivated regions
- NSBI+FTS gives us control over high-dimensional problems in particle physics

$$FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- 'Asymptopia' a fundamental misconception about our field, new physics is always just out of reach!

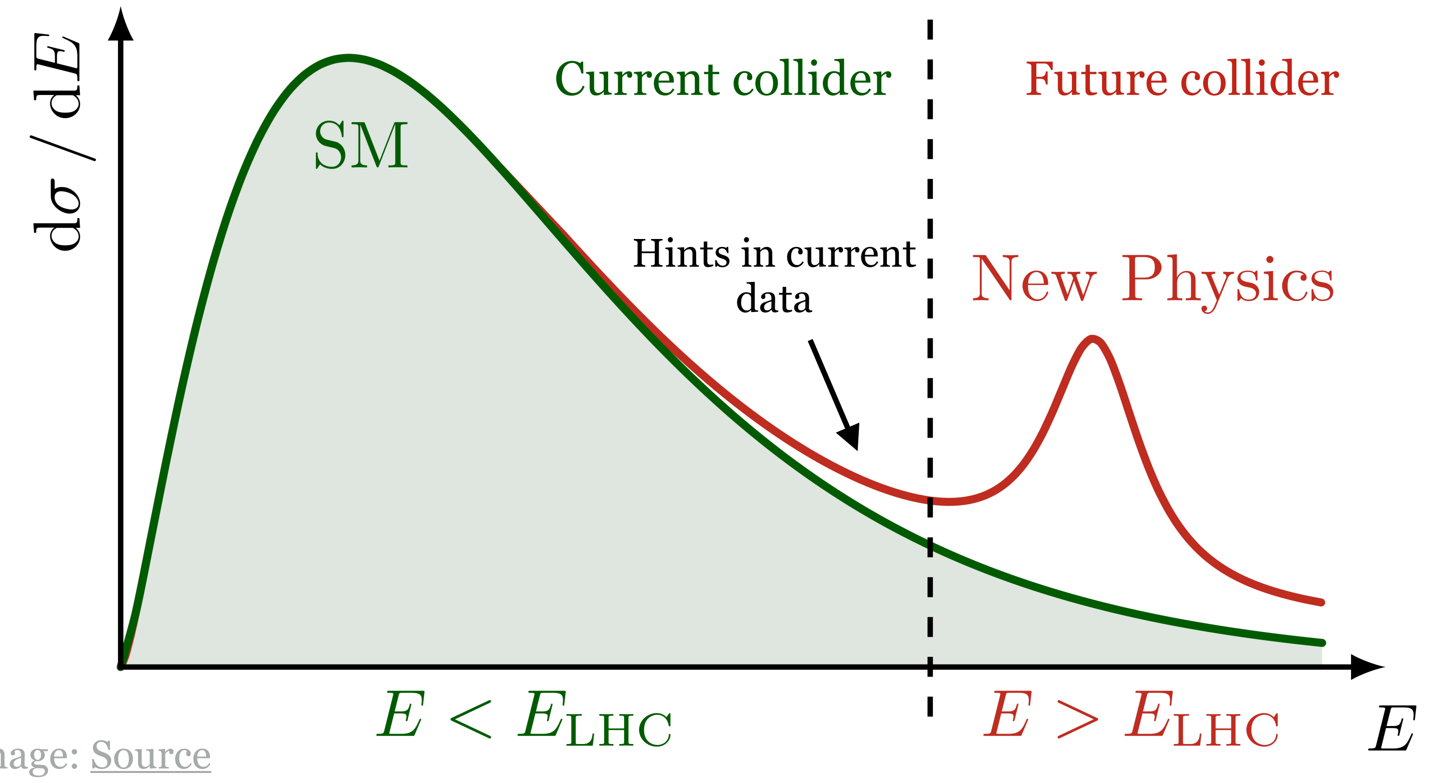


Image: Source

Thank you!

Backup

Quantitative improvements with FTS

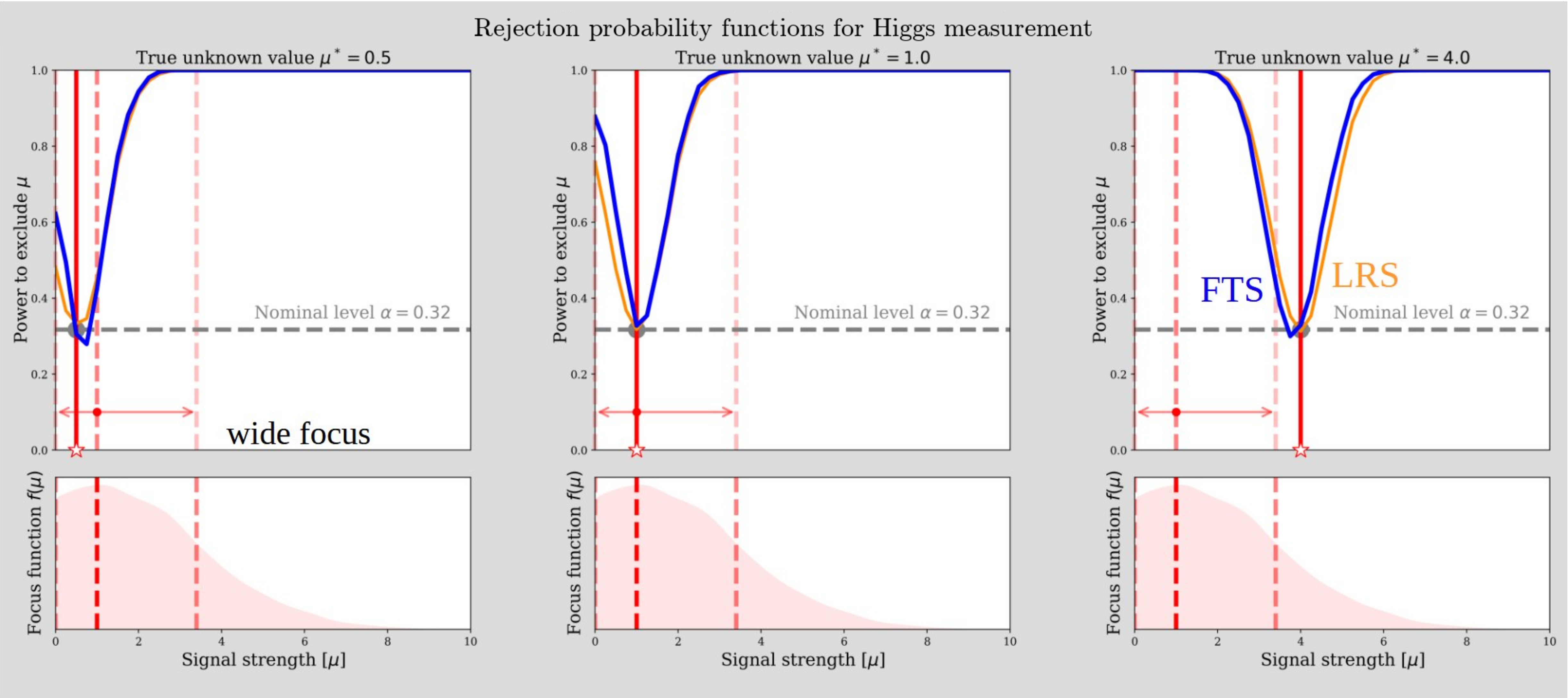
Using reconstructed Higgs mass as observable to construct histogram

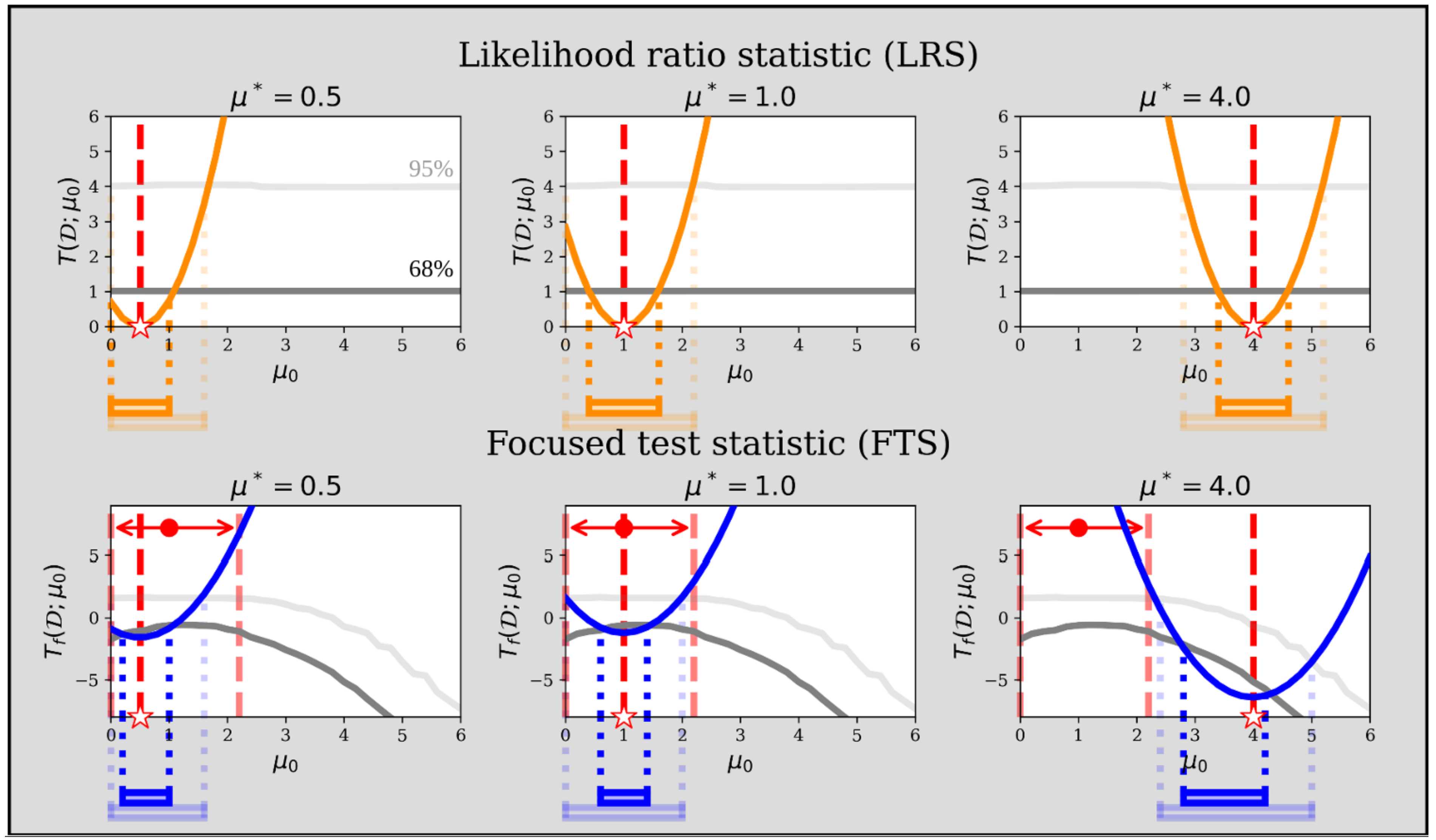
Using visible energy as observable

Setting		Test statistic		
Experiment	μ^*	LRS	FTS-wide	FTS-narrow
Higgs (mass)	1.0	1.08 (2.01)	0.94 (1.89)	**0.85 (1.79)**
Higgs (vis. mass)	1.0	1.26 (2.31)	1.10 (2.17)	**0.98 (2.08)**
LZ-inspired	0.0	5.99 (13.87)	4.68 (11.47)	**3.89 (11.29)**
LZ-inspired	1.0	7.08 (15.18)	5.29 (12.82)	**4.68 (12.75)**

Numbers represent length of confidence intervals for 1σ (2σ)

Introducing a technical bias does not compromise coverage





Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \; p_{\mathrm{S}}(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \; p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu})\nu_{\mathrm{B}} \; p_{\mathrm{B}}(x)\right]$$

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[\underline{(\mu - \sqrt{\mu})}\, \nu_S \; p_S(x) + \underline{\sqrt{\mu}}\, \nu_{\mathrm{SBI}_1} \; p_{\mathrm{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})}\, \nu_{\mathrm{B}} \; p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j\, p_j(x_i)$$

Event rates estimated from simulations

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S\, p_S(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1}\, p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_{\mathrm{B}}\, p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

Event rates estimated from simulations

Comes from theory model chosen to interpret data

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1}\, p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu})\nu_{\mathrm{B}}\, p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

$$\frac{p(x_i|\mu)}{p_{\mathrm{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, \frac{p_j(x_i)}{p_{\mathrm{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \, p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$$\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \frac{p_{\mathrm{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{p_S(x)} \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

Estimated using an ensemble of networks

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1}\, p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu})\nu_B \, p_B(x) \right]$$

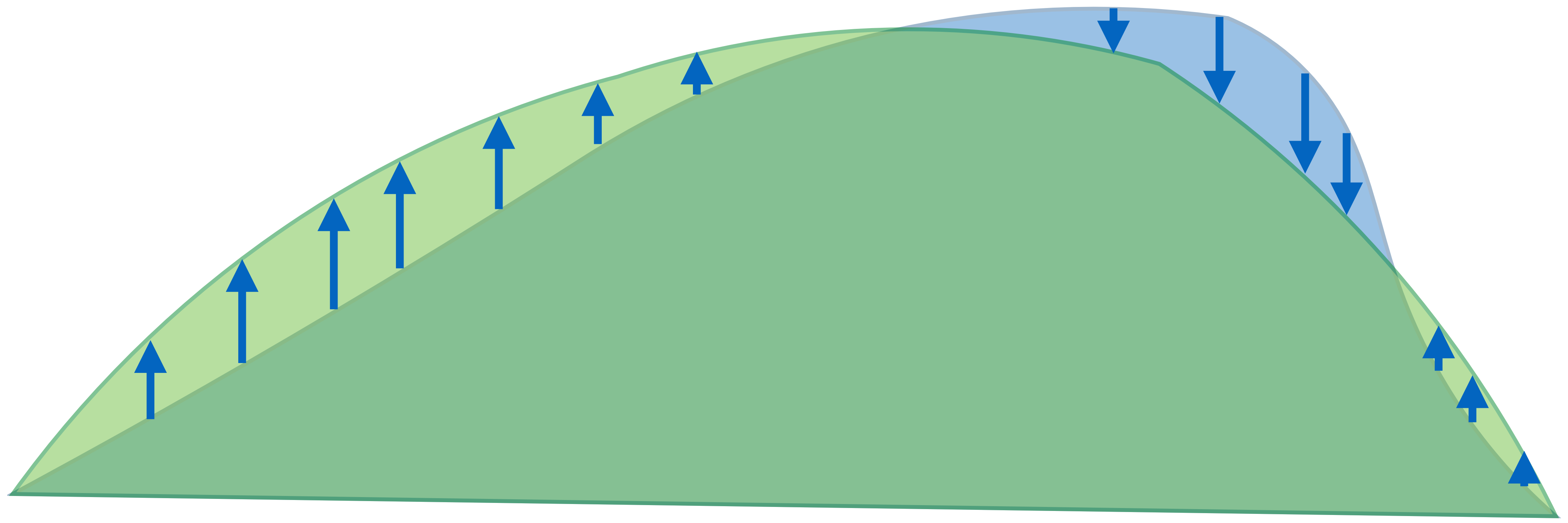
$$\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu})\nu_B \frac{p_B(x)}{p_S(x)} \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

More NSBI diagnostics

Validate quality of LR estimation with re-weighting task

Reweighting: Calculate weights w_i for events x_i in blue sample to match green sample

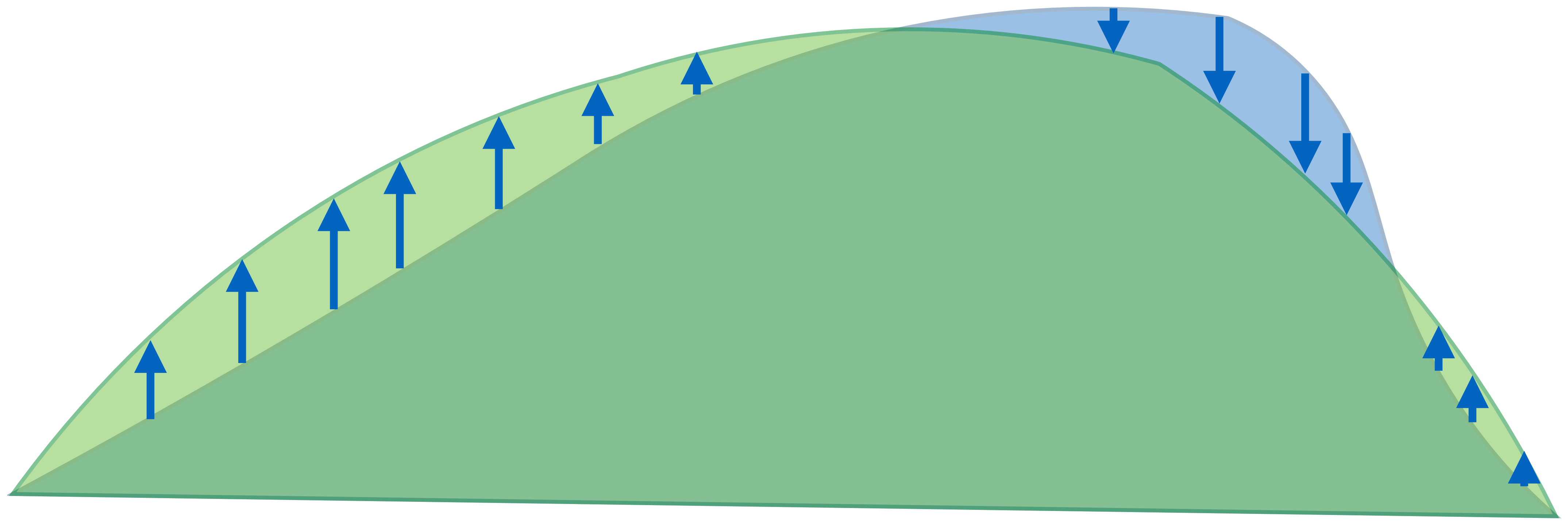


Validate quality of LR estimation with re-weighting task

Reweighting: Calculate weights w_i for events x_i in blue sample to match green sample

$$w_i = r(x_i, \mu_0, \mu_1) = \frac{p(x_i \mid \mu_0)}{p(x_i \mid \mu_1)}$$

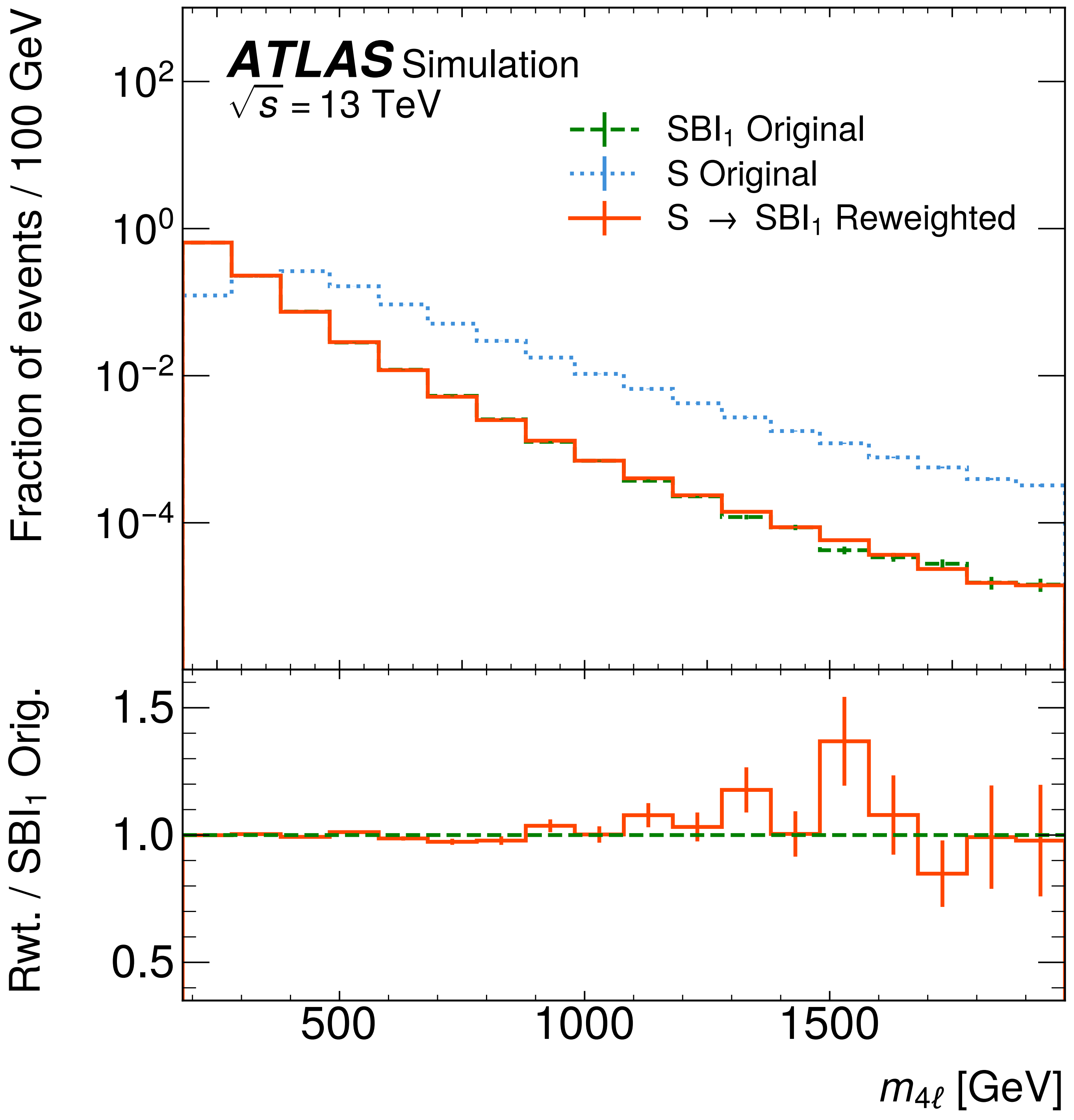
Already estimated using an ensemble of networks



Re-weight closures

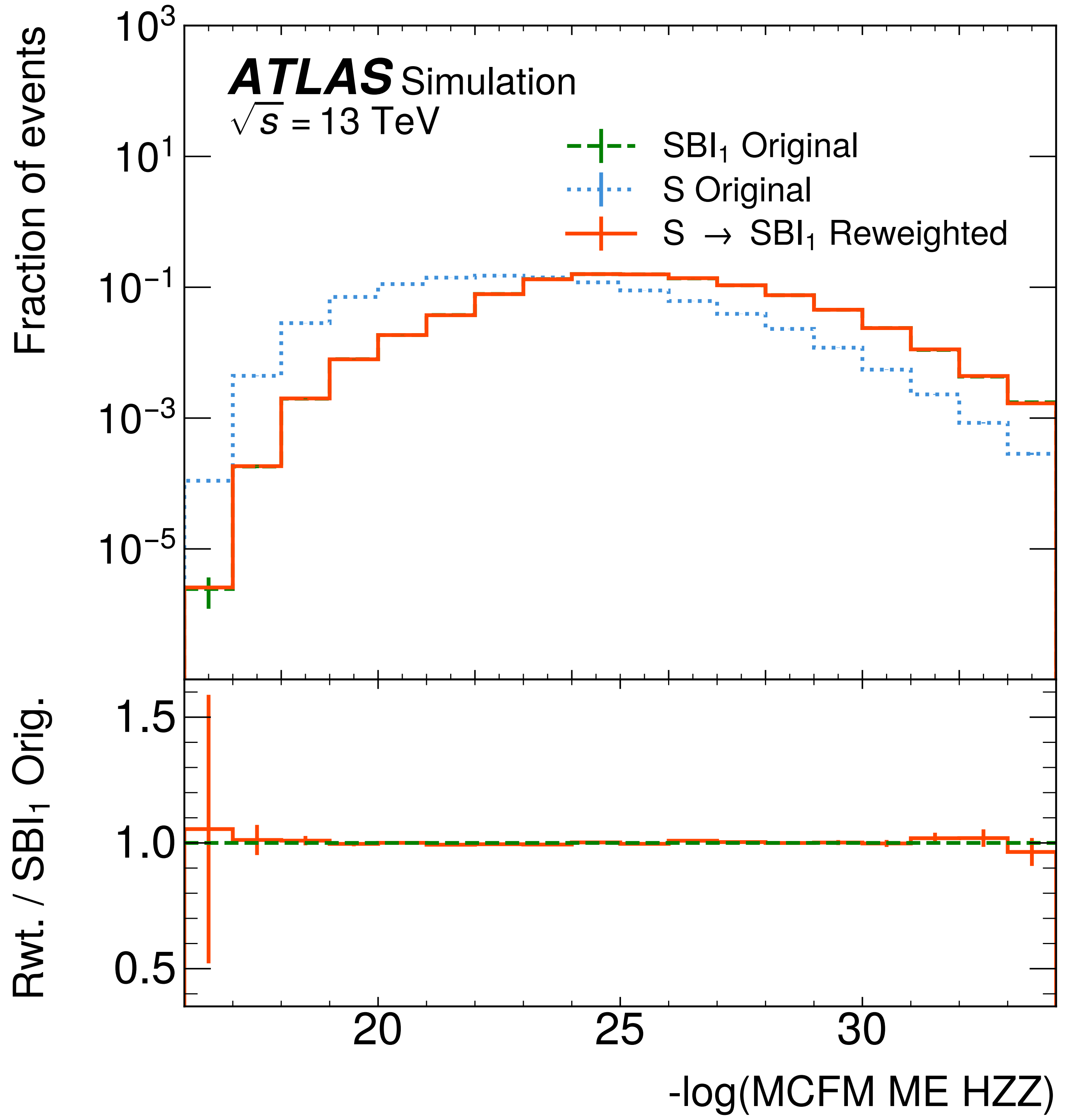
Source
Target
RW

Variable used in training



m_{4l}

High-level variable
never used in training



Matrix-Element-based Observable
(ggF from MCFM)

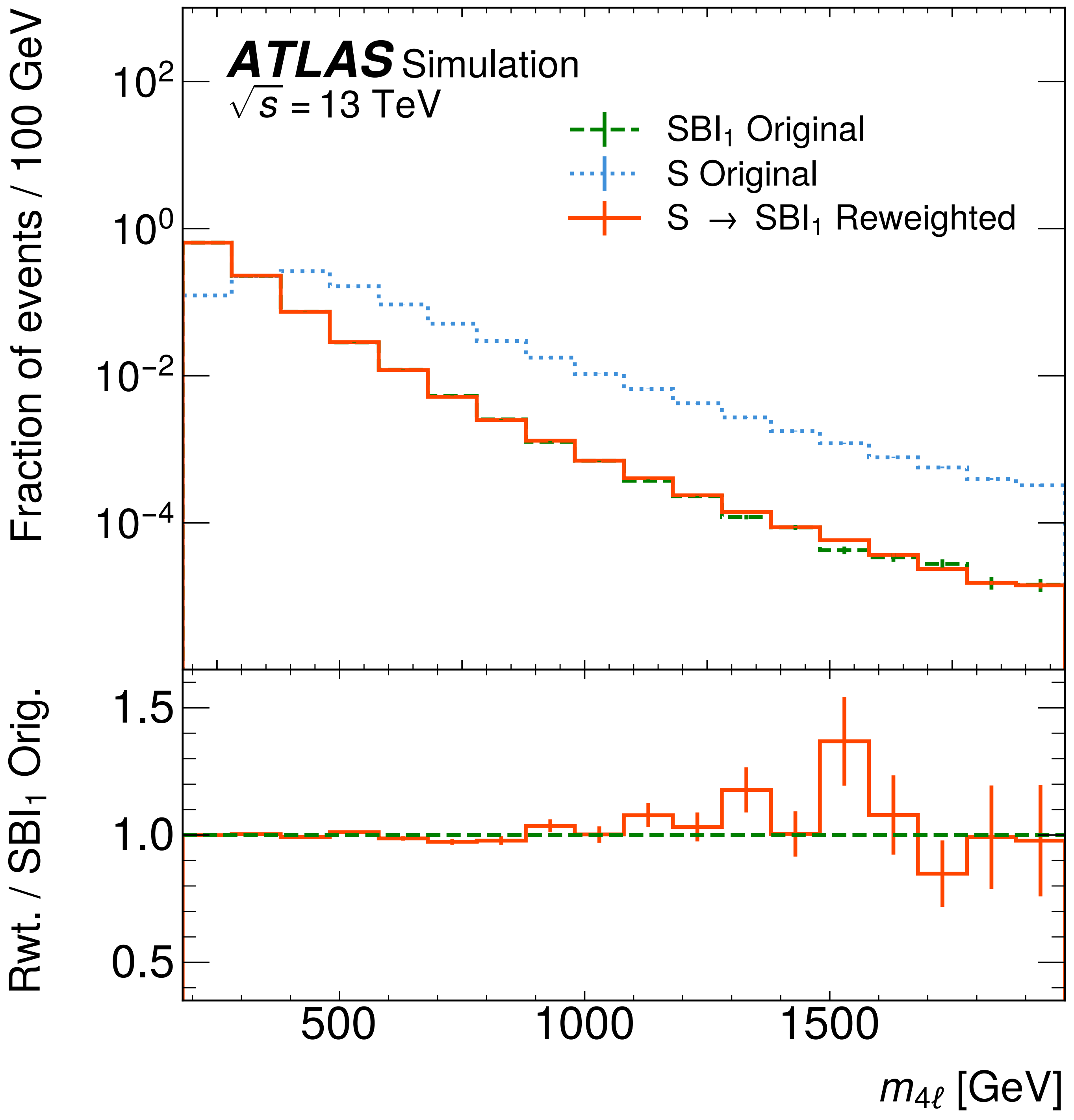
Hoping to see Target and RW match

ATLAS methods paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Re-weight closures

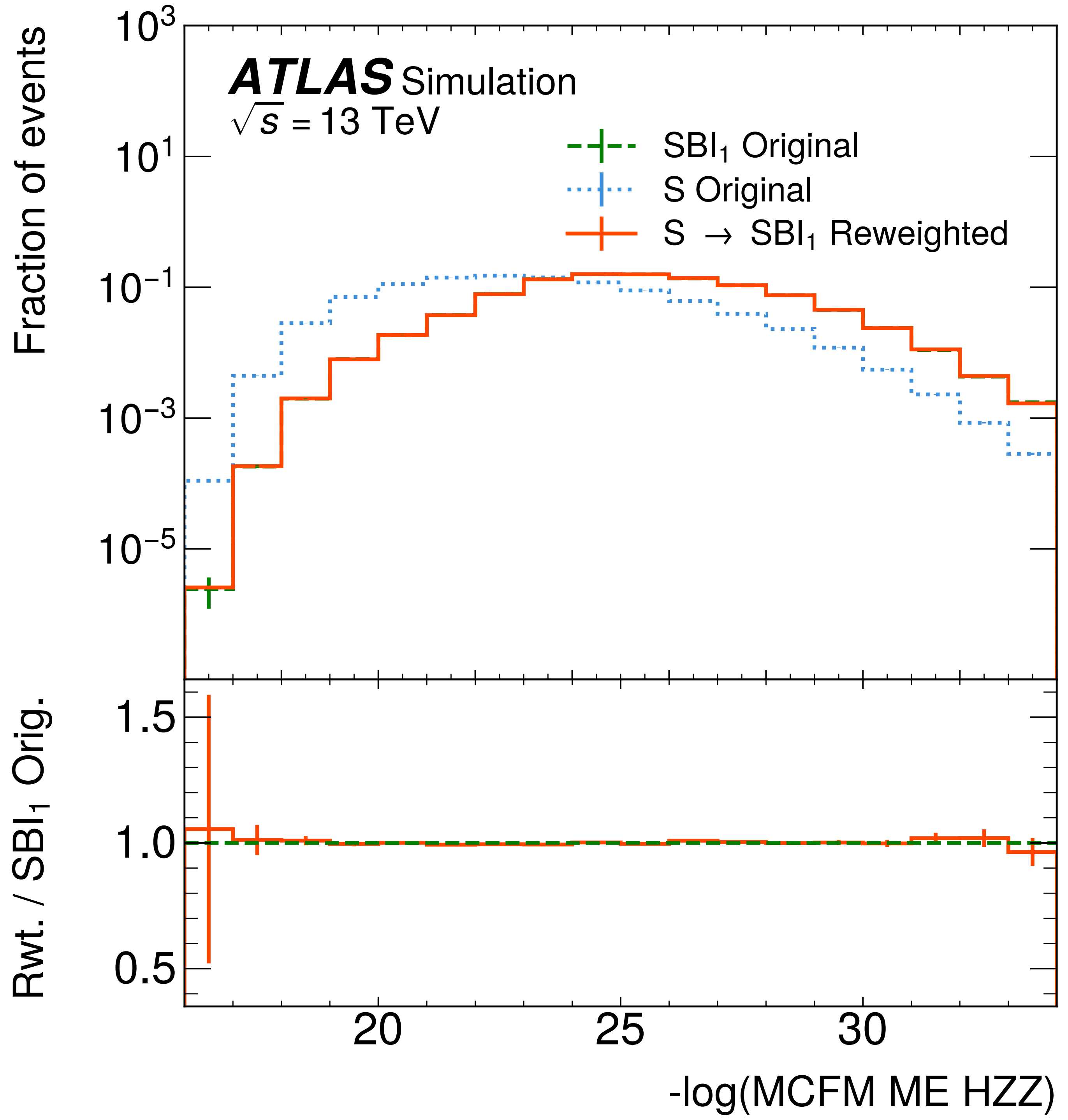
Source
Target
RW

Variable used in training



m_{4l}

High-level variable
never used in training



Matrix-Element-based Observable
(ggF from MCFM)

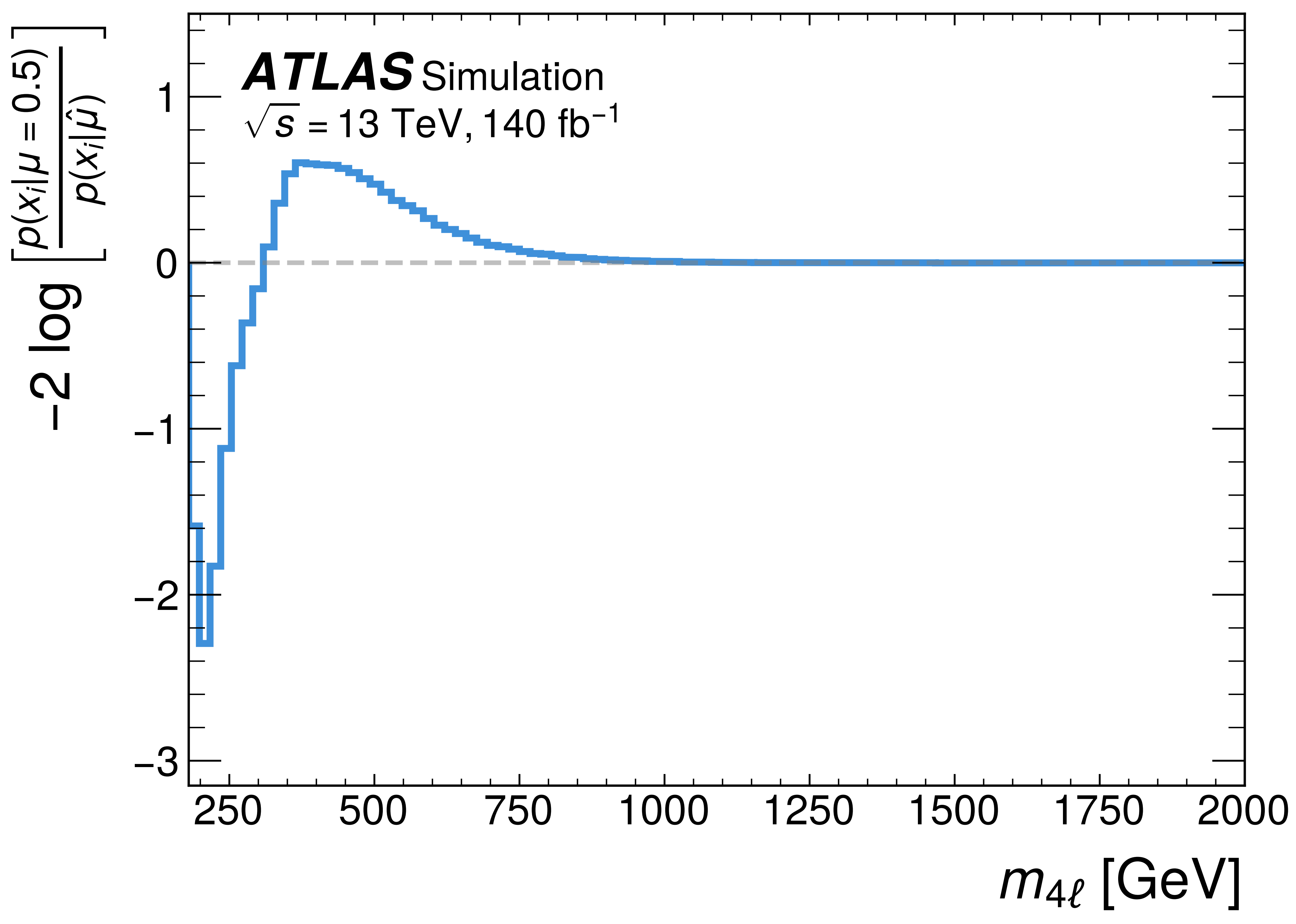
High-dim goodness-of-fit test:
Train independent classifier on RW vs Target
AUC=0.5 ⇒ LRs well estimated

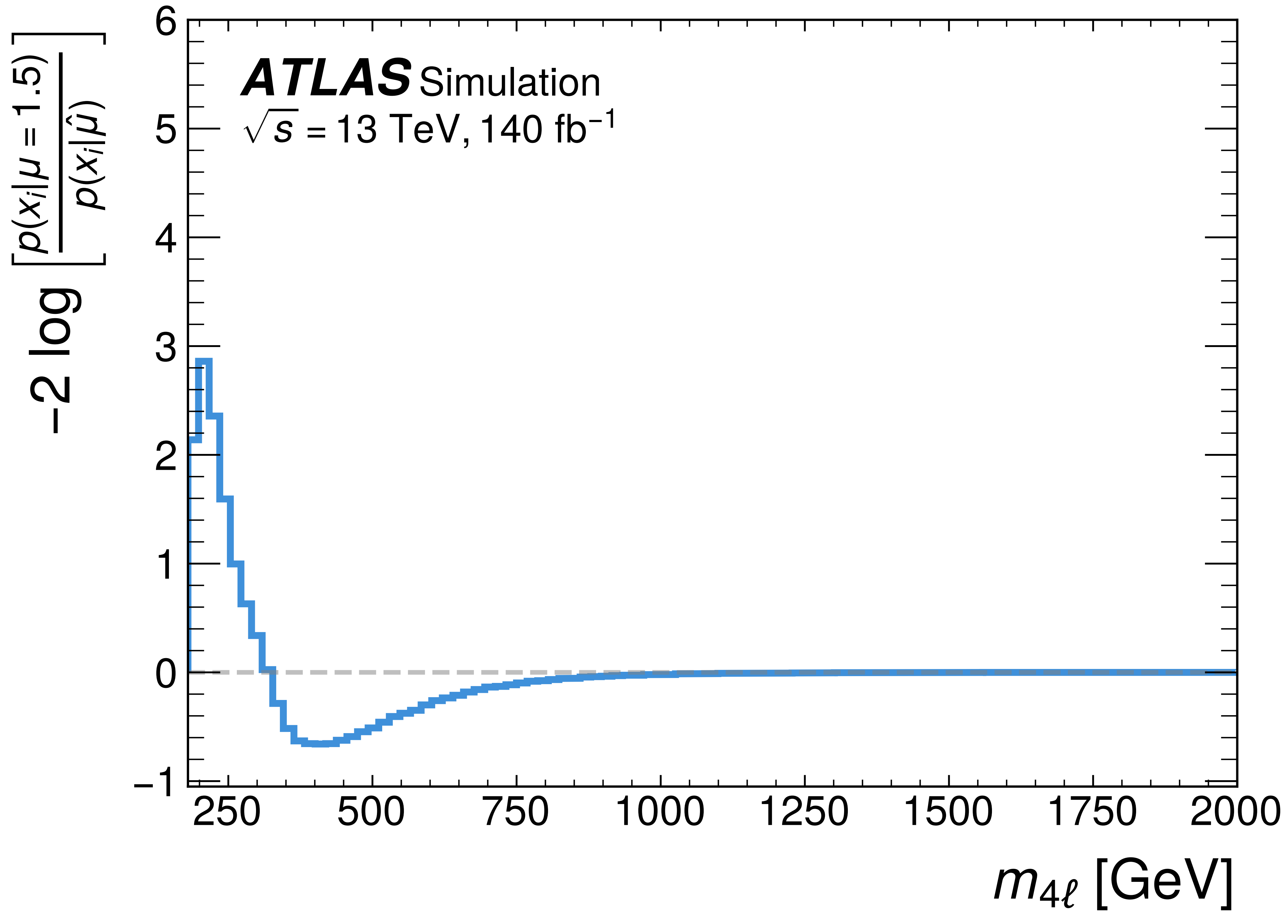
Hoping to see Target and RW match

Interpretability: Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 0.5)}{P(x_i \,|\, \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 1.5)}{P(x_i \,|\, \mu = 1)}$$

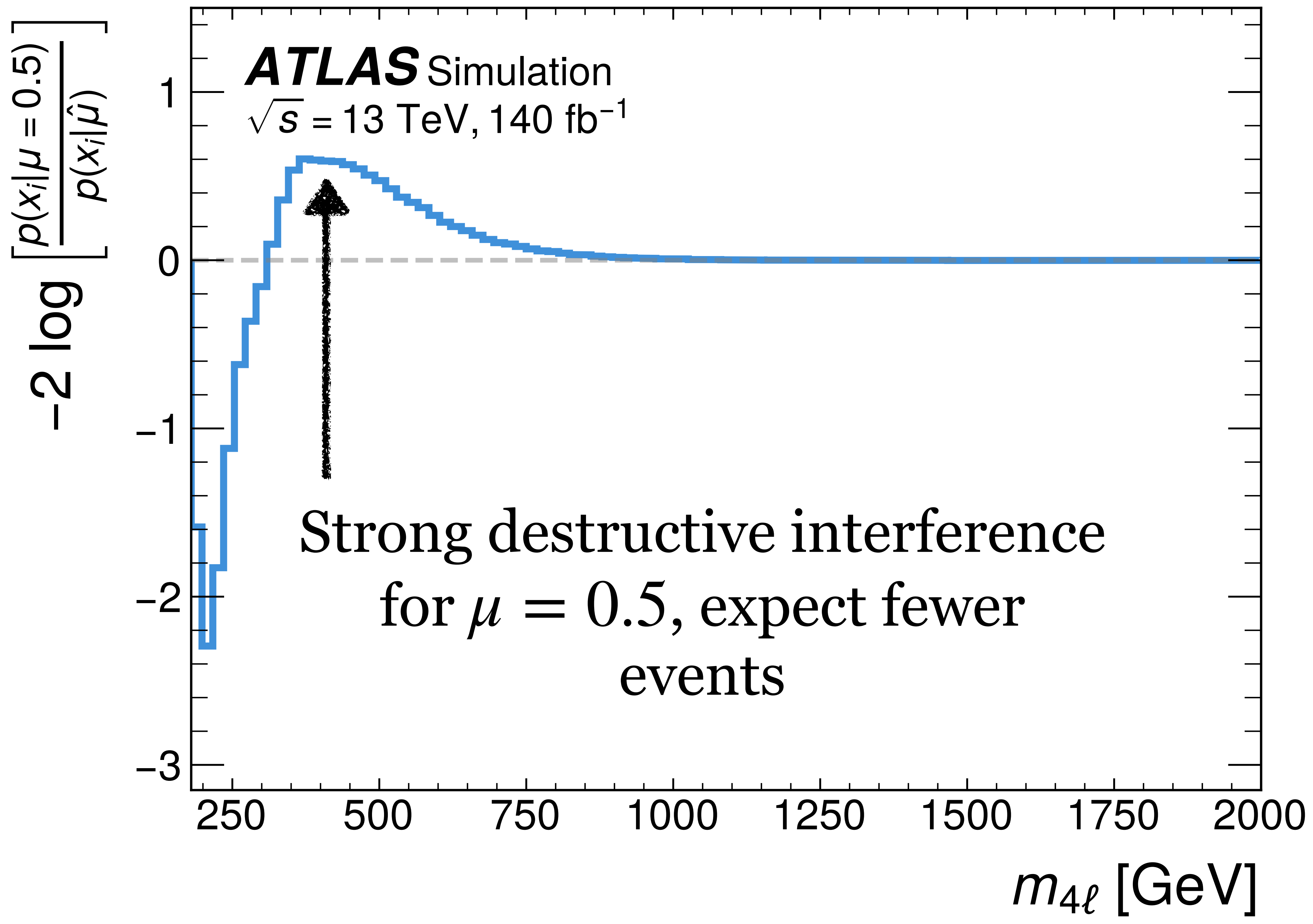




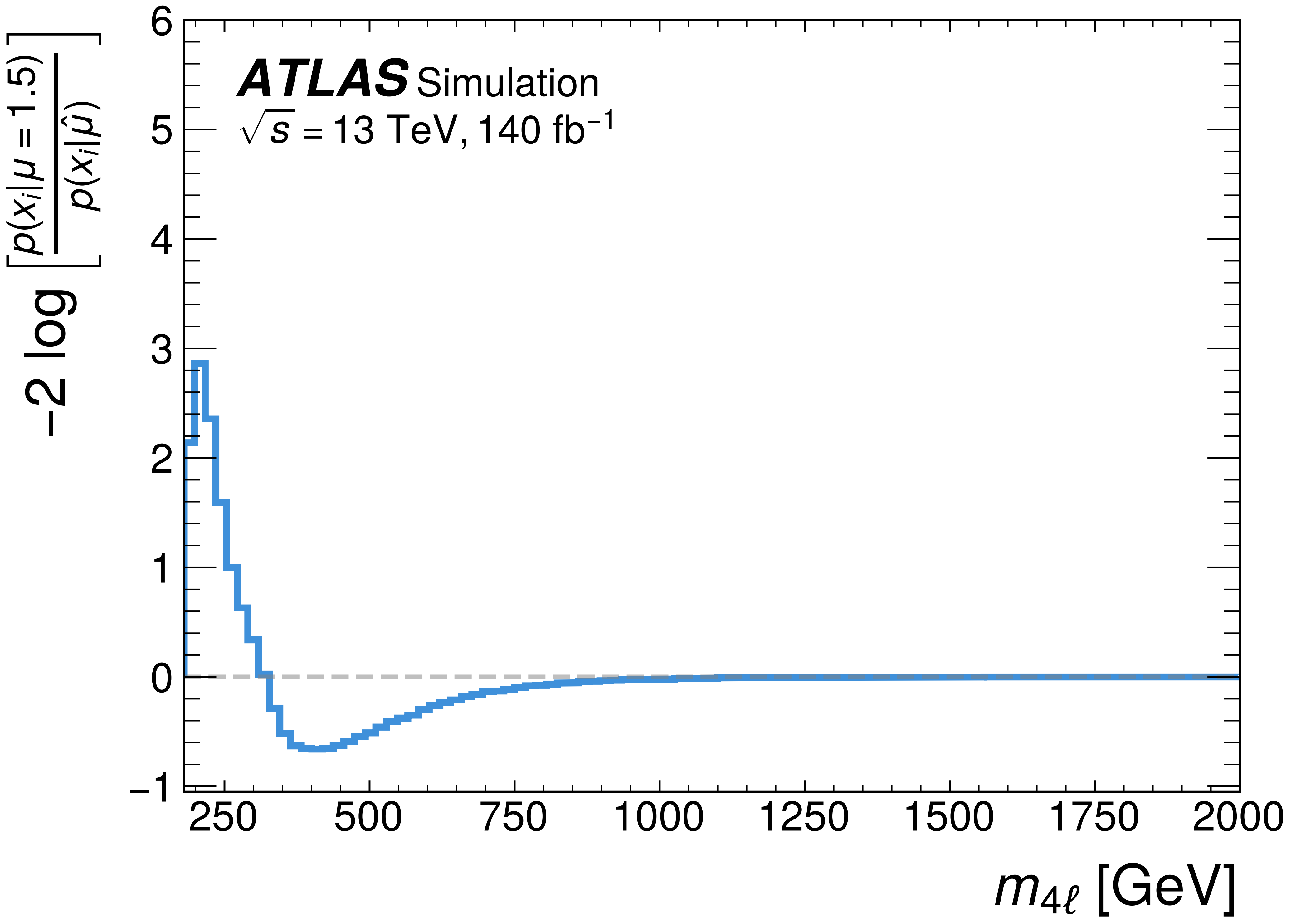
Interpretability:
Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 0.5)}{P(x_i \,|\, \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 1.5)}{P(x_i \,|\, \mu = 1)}$$



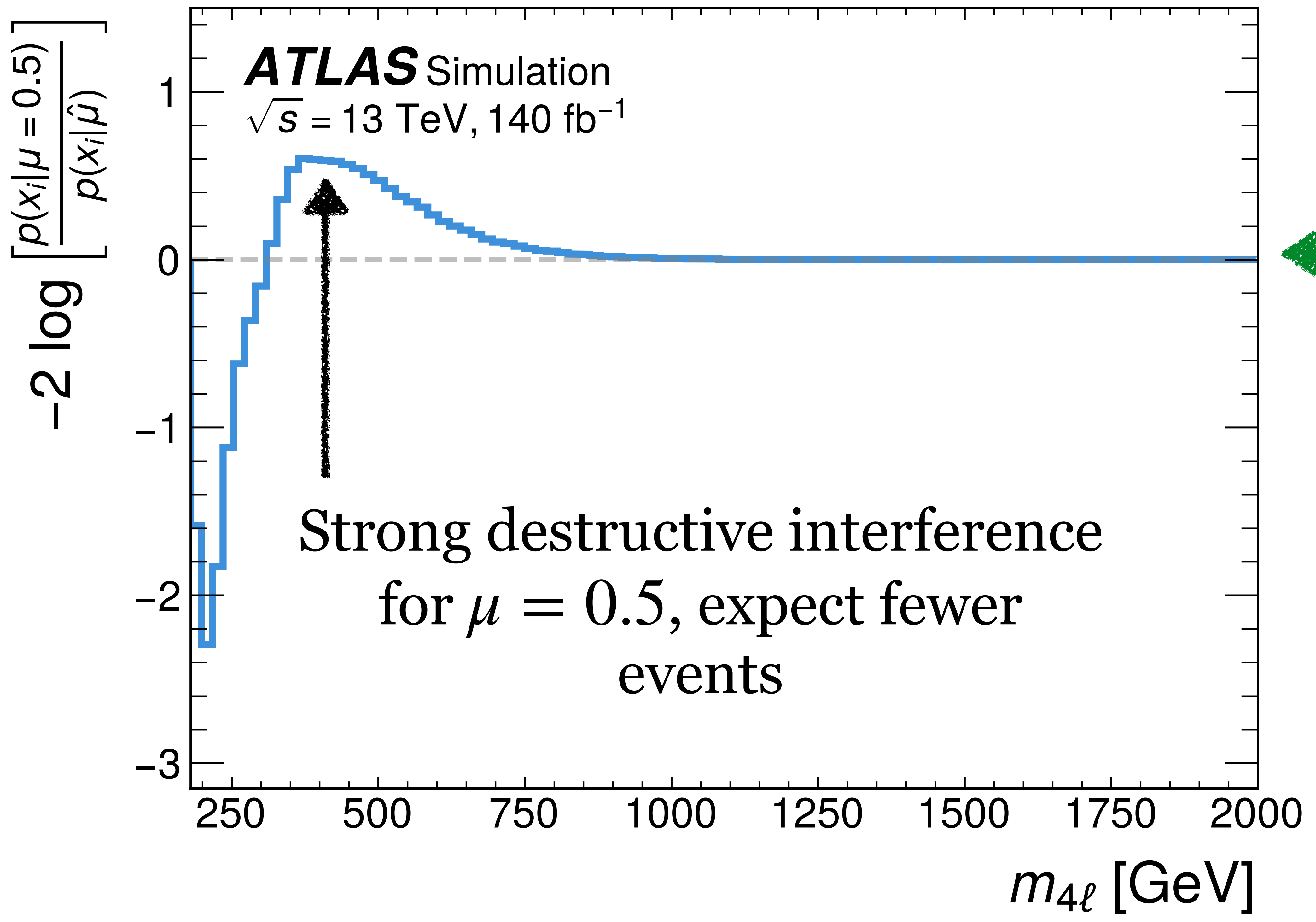
Strong destructive interference for $\mu = 0.5$, expect fewer events

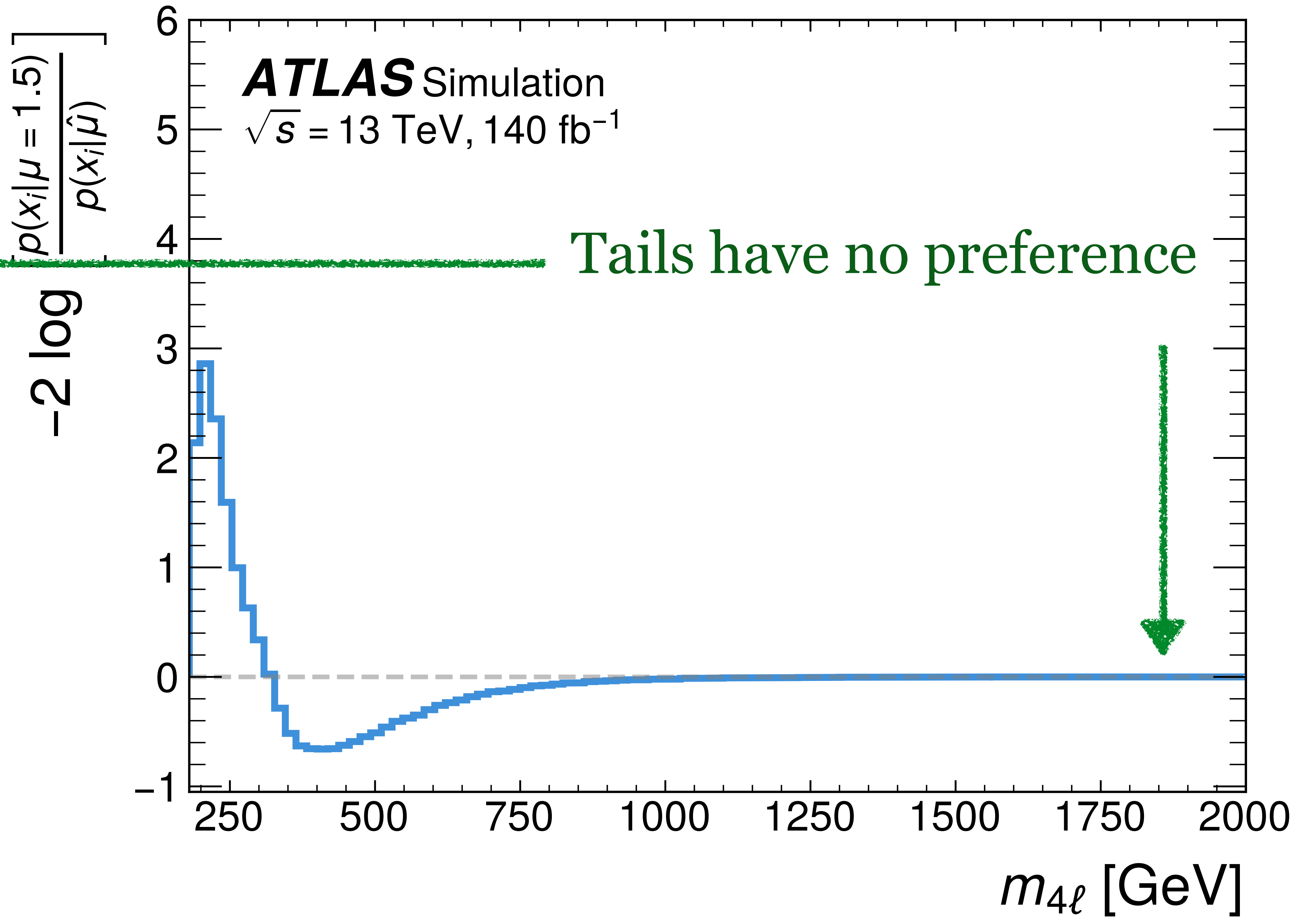


Interpretability:
Which phase space favours one hypothesis over another?

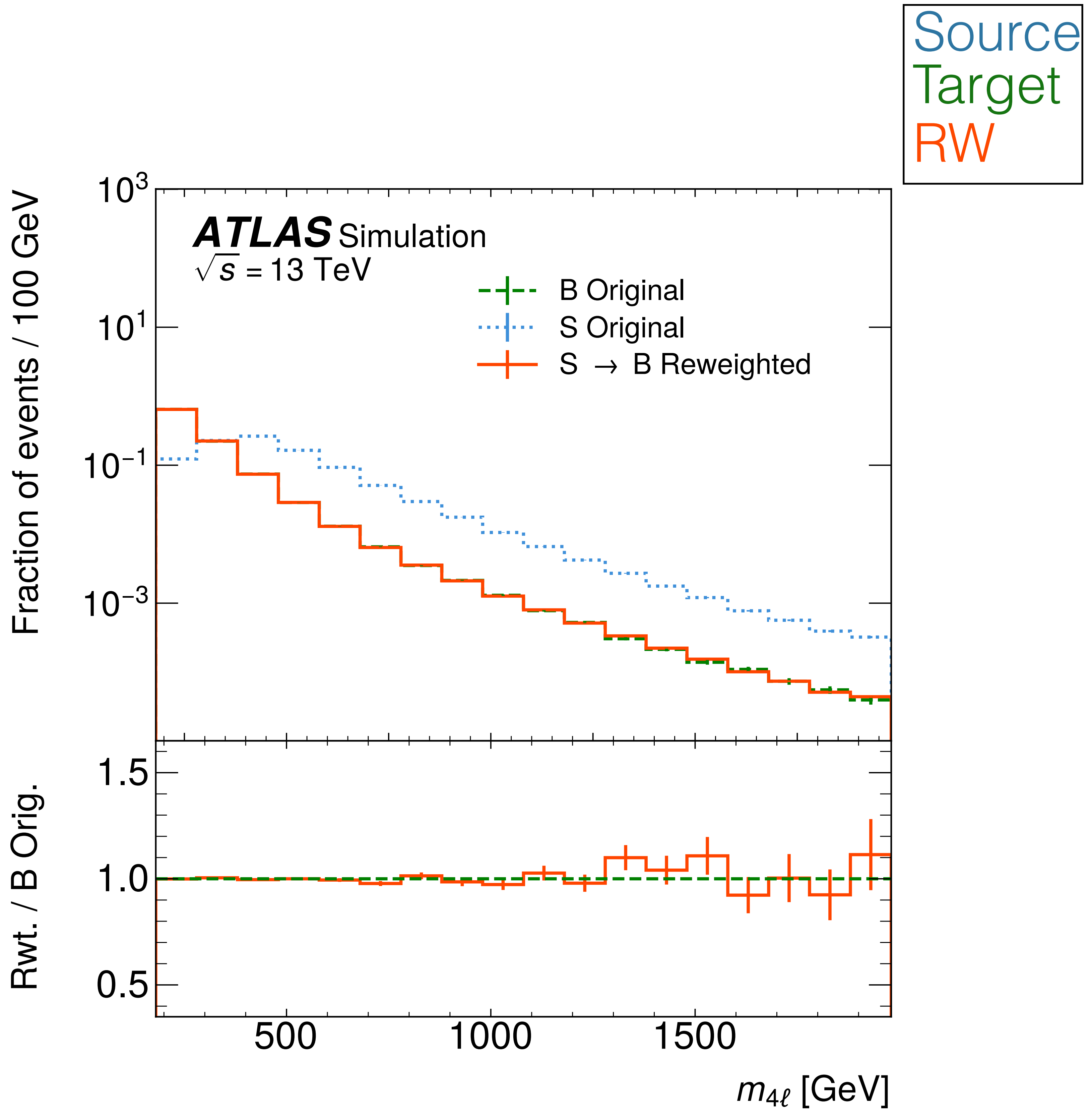
$$-2 \cdot log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}$$

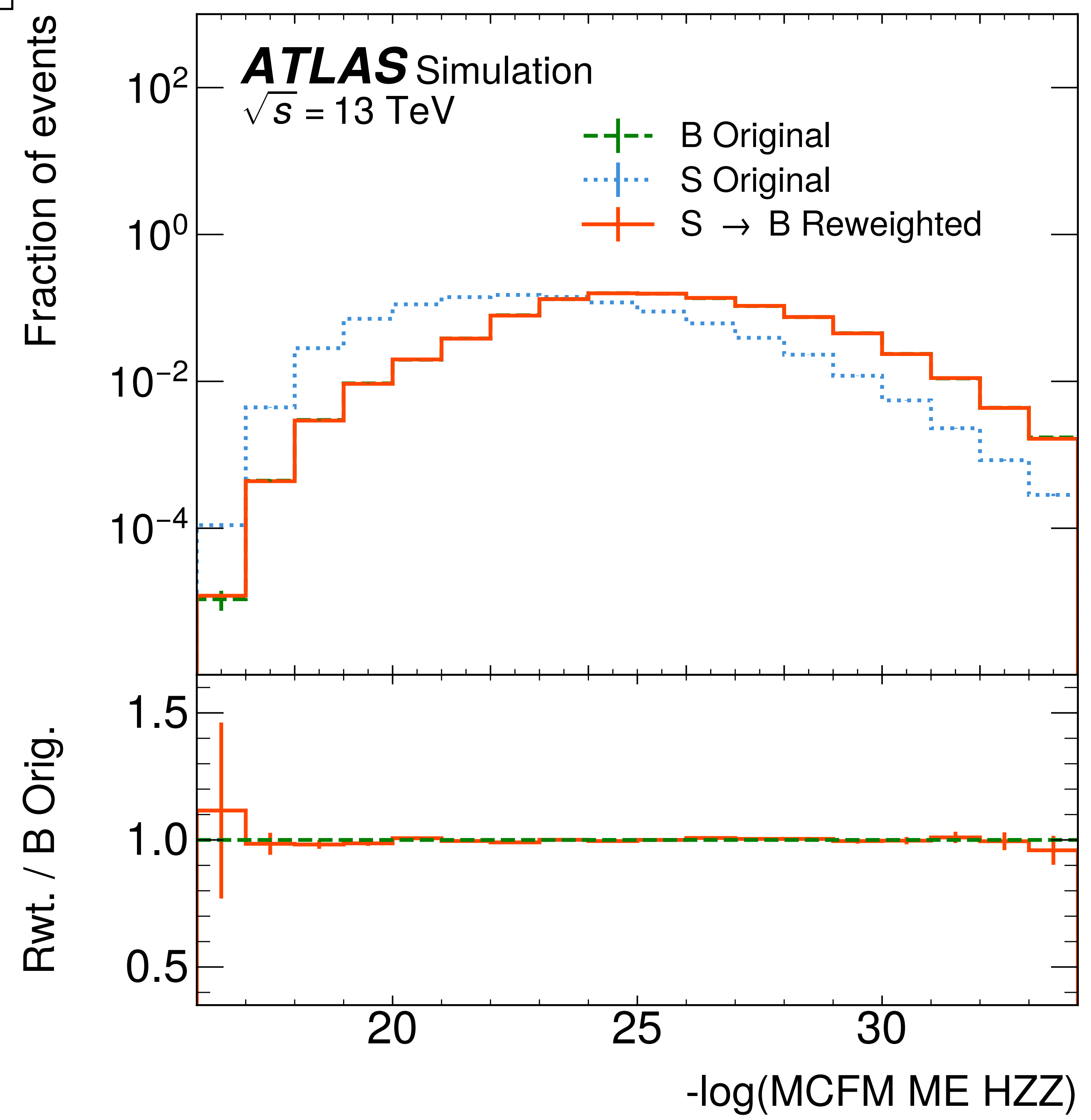




Re-weight closures for B



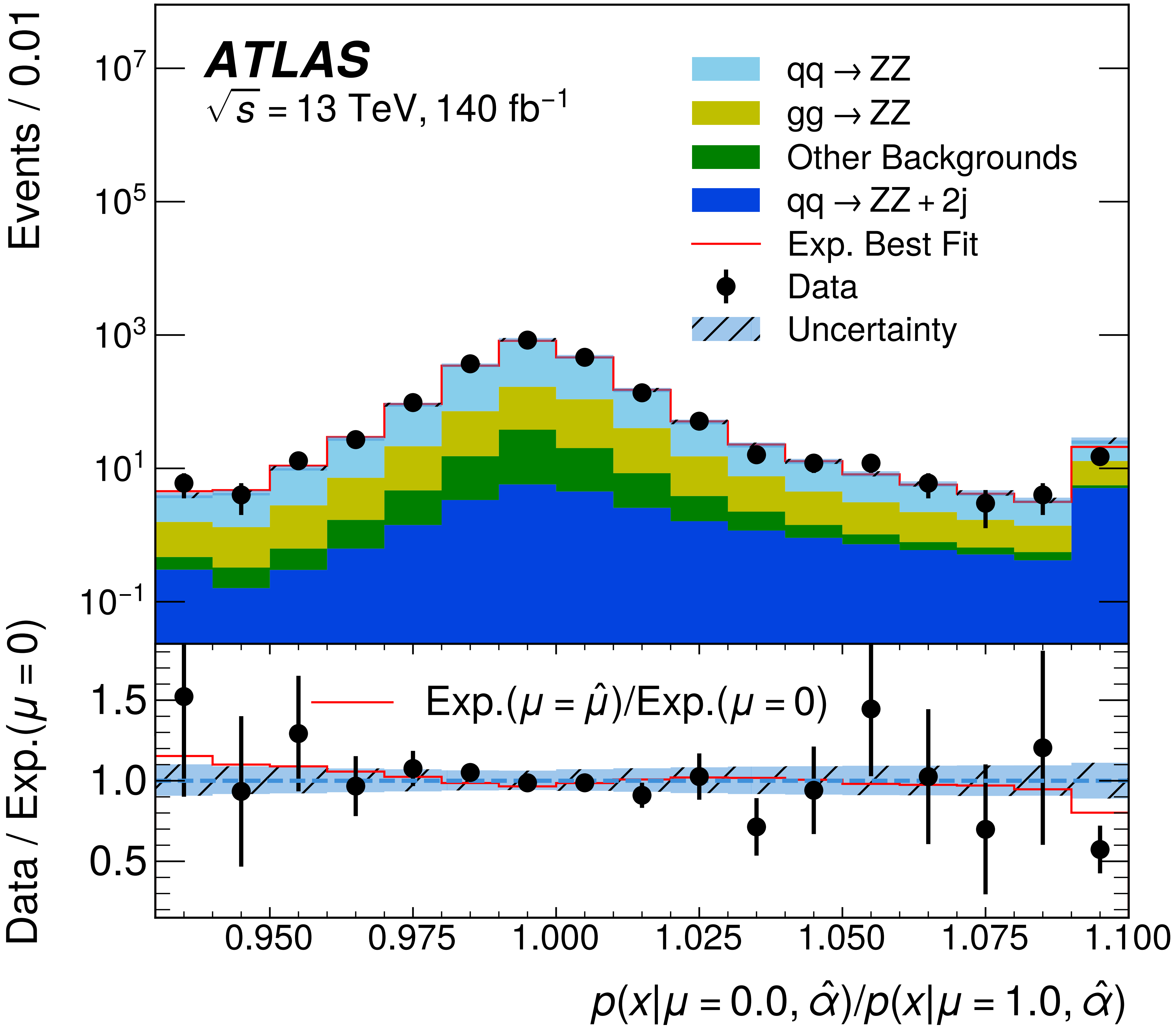
m_{4l}



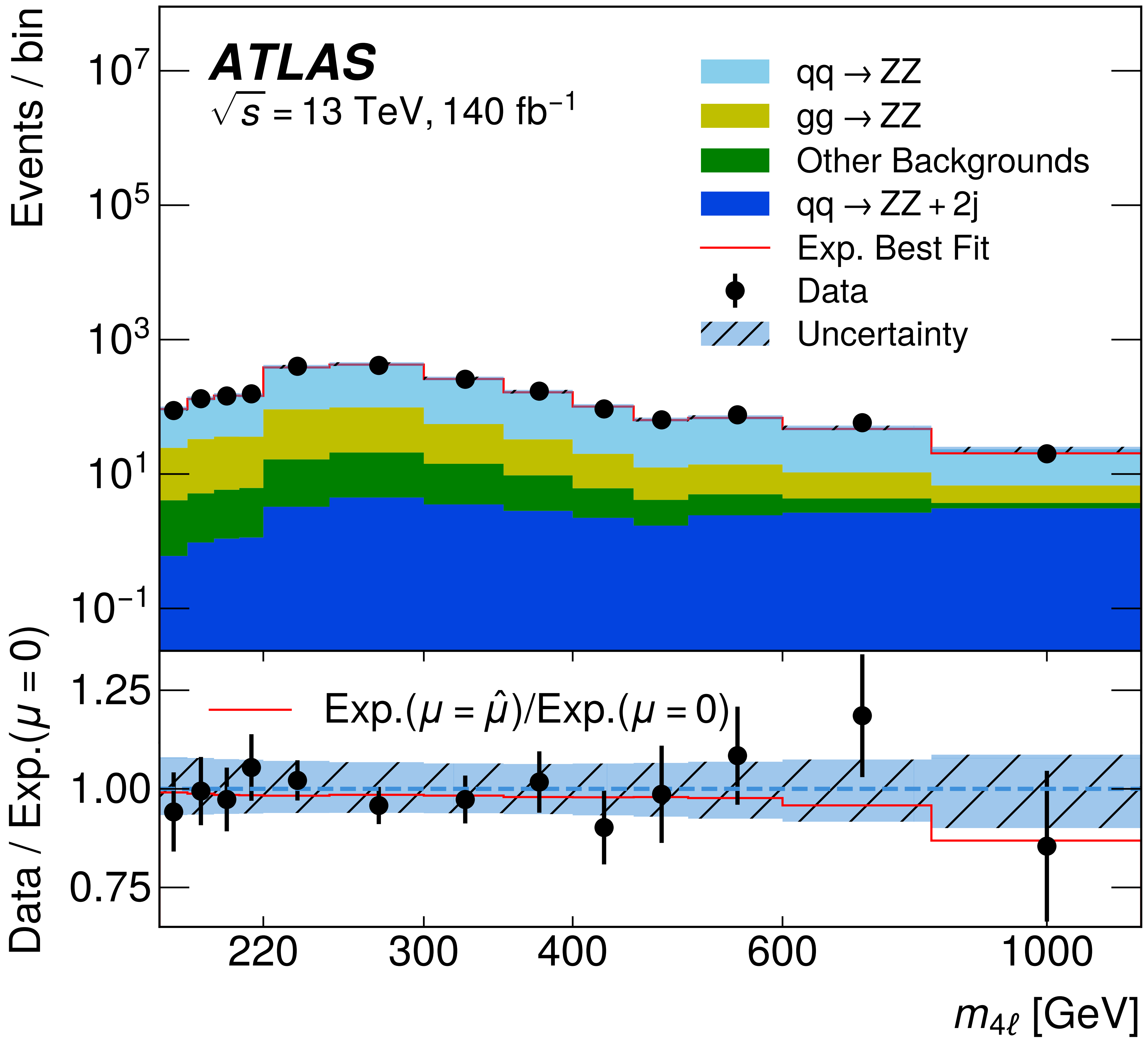
Matrix-Element-based Observable
(ggF from MCFM)

Data-MC validation

NN observable



m_{4l}



Data-MC validation

Different NN observables