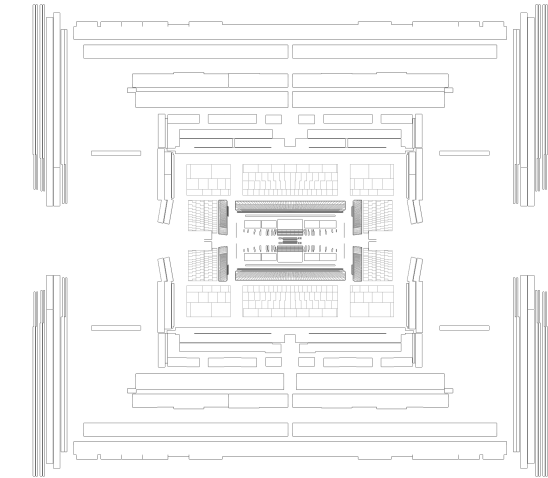
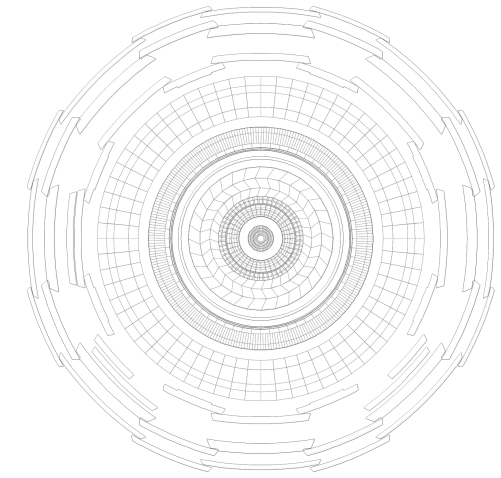


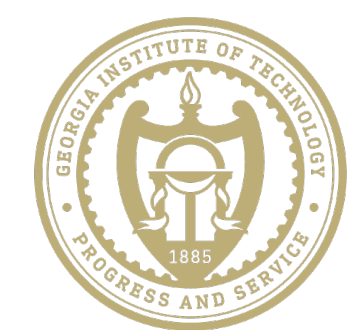


Unlocking Physics in High Dimensions from Small-Sample Data

Aishik Ghosh

PARTIQAL, Daejeon, Korea

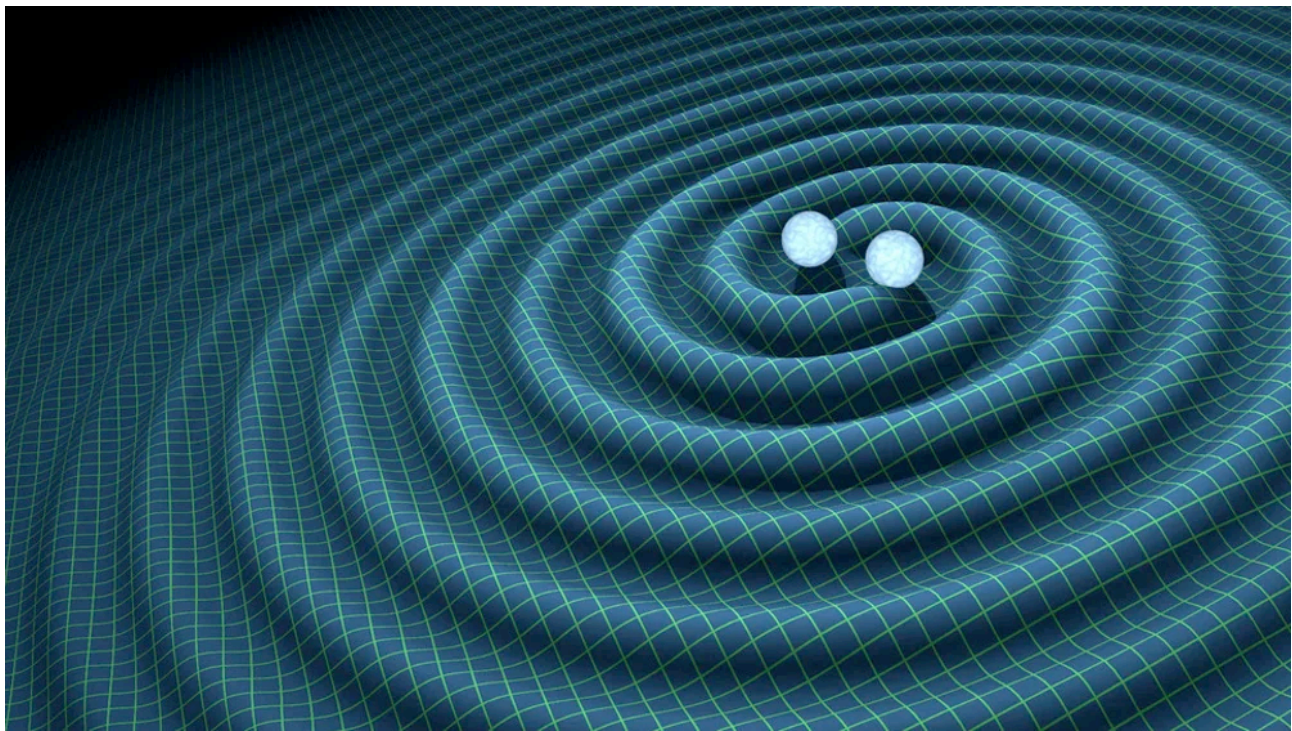
13 July 2026



**Georgia Institute
of Technology**

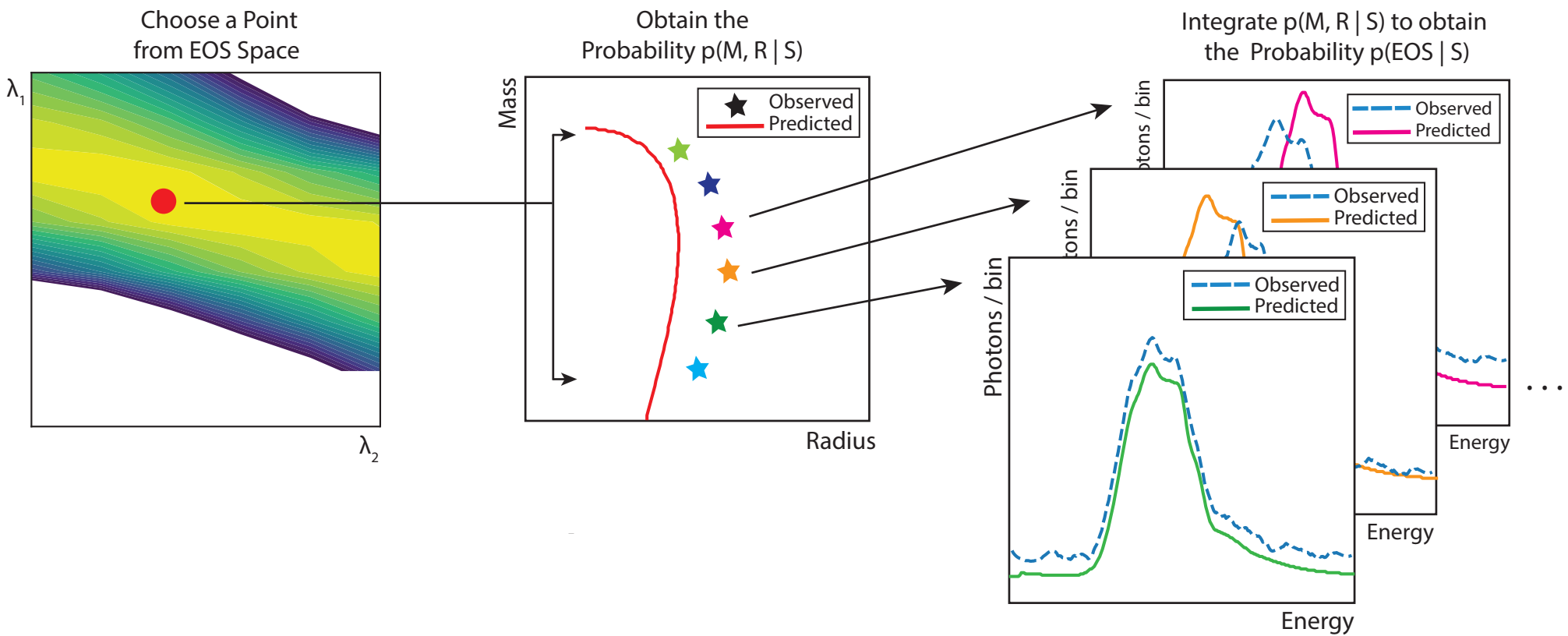


Gravitational Wave discovery (2016)



Astrophysics:

Probing the interior of neutron stars

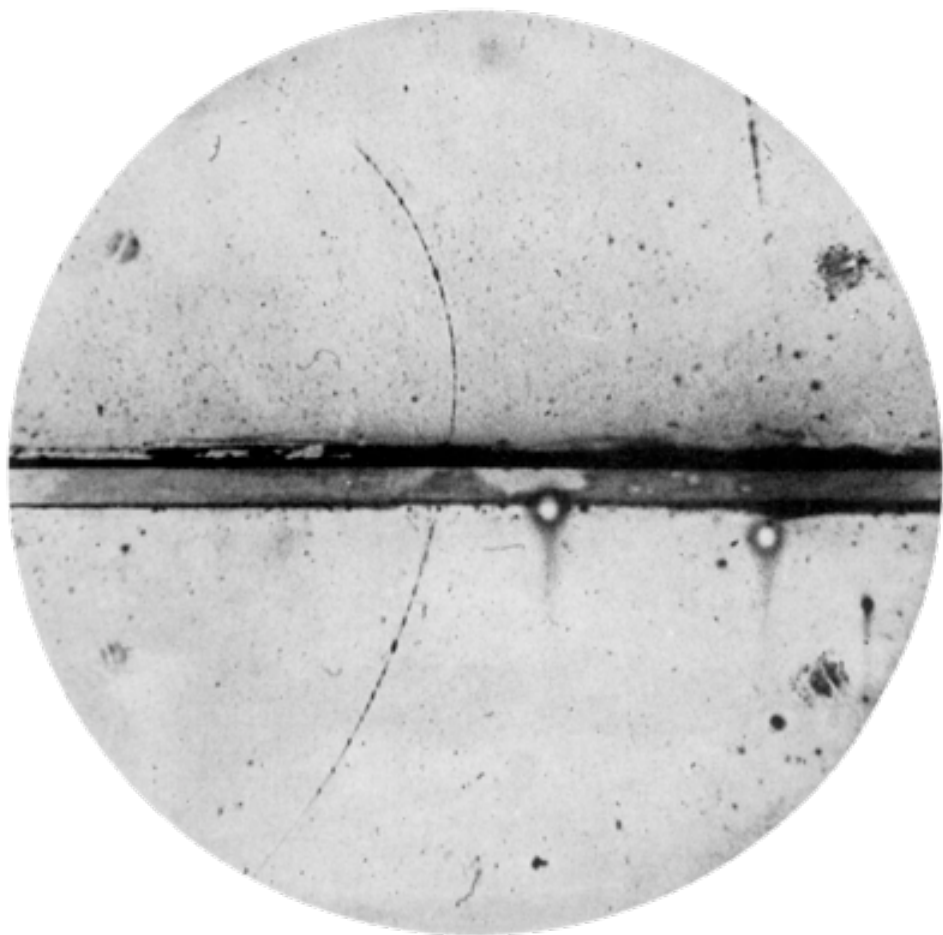


[JCAP12\(2023\)022](#): Farrell et al. (incl. Ghosh)

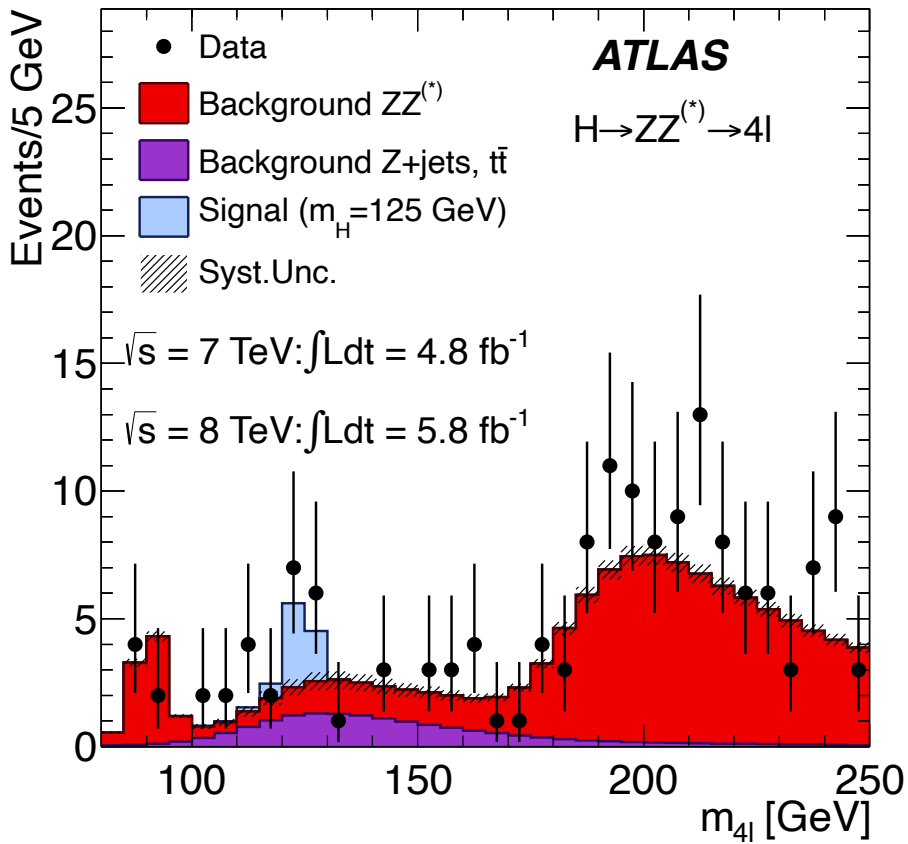
Single event

Collection of events

Particle physics:



Positron discovery (1930s)



Higgs boson discovery (2012)

Statistical inference of physics parameters

$$m_H = 125.25 \pm 0.17 \text{ GeV}$$

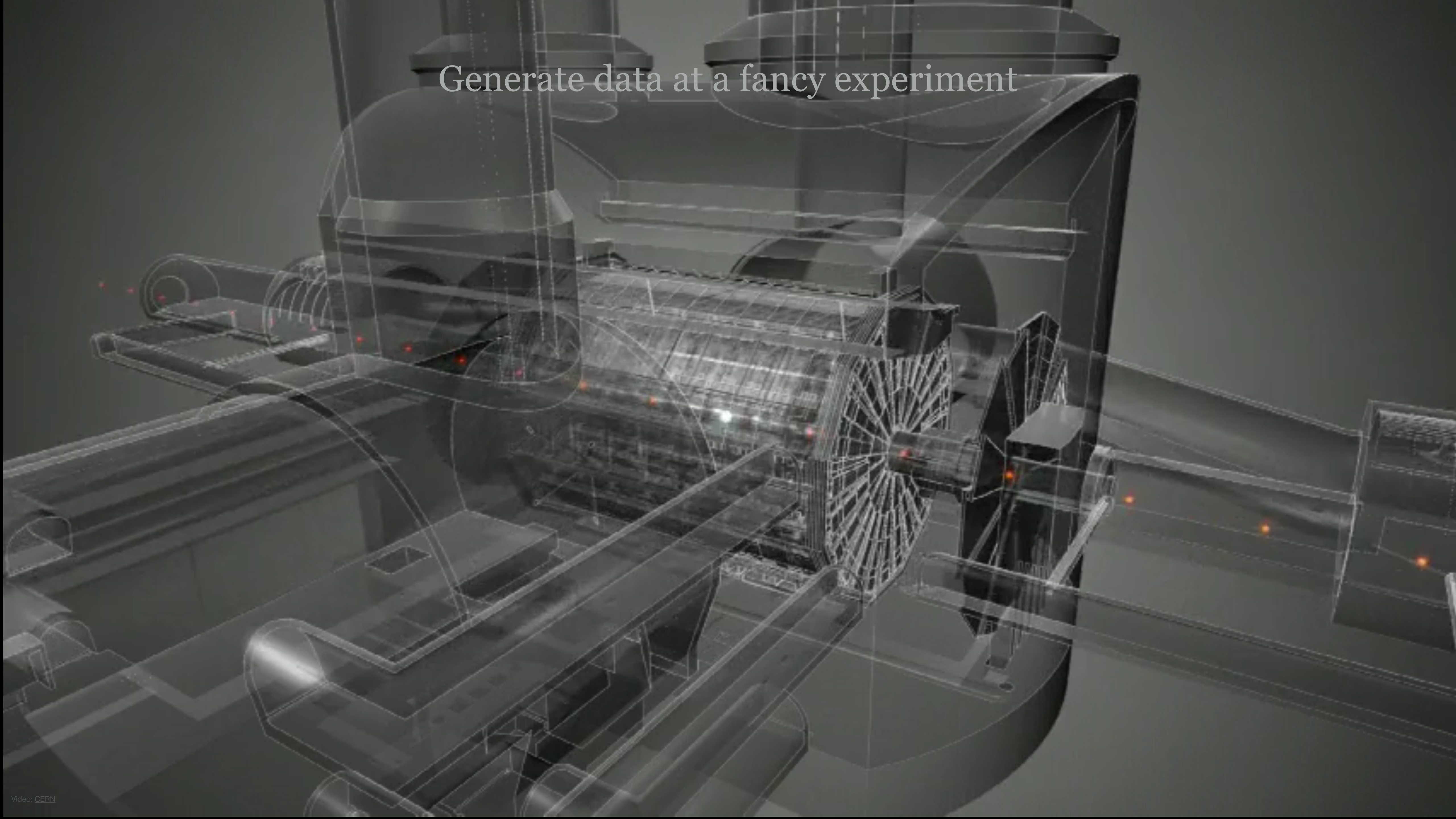
Statistical uncertainties, but also detector systematic, theory systematic, MC stat,



“How sure am I ? How can I reduce my uncertainty ?”

An experimentalist thinks about ‘confidence intervals’

Generate data at a fancy experiment



ML for Detector Simulation

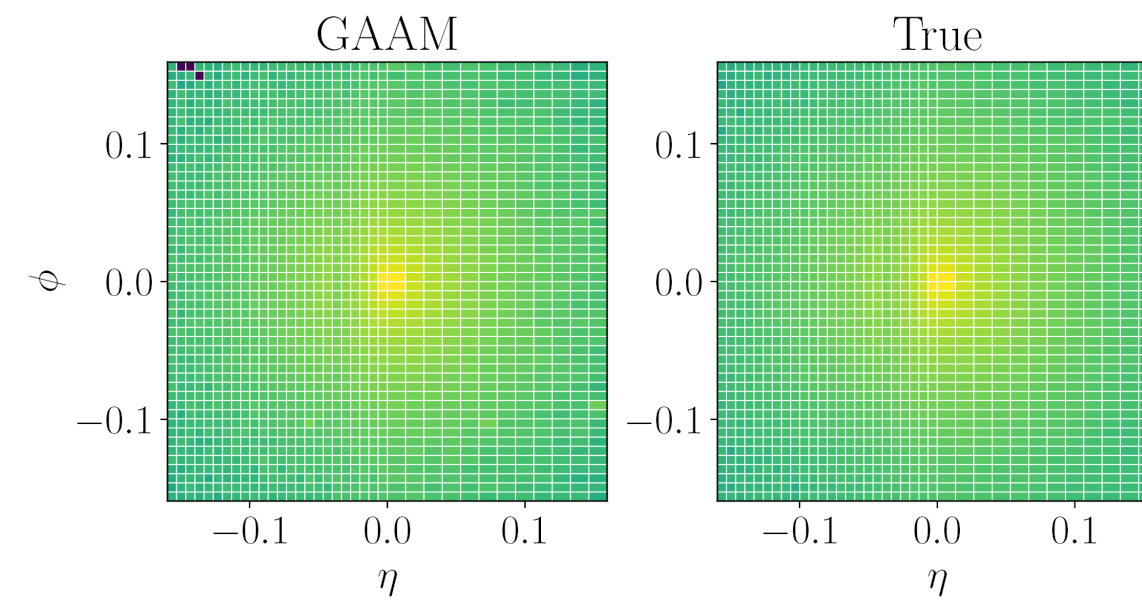


Image: [arXiv:2305.11531](https://arxiv.org/abs/2305.11531)

ML for Reconstruction

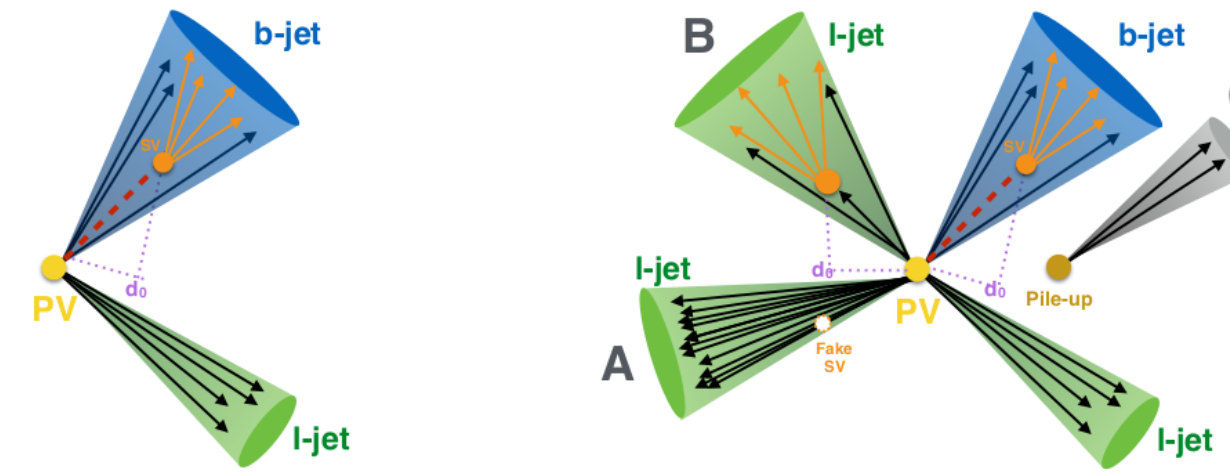


Image: [USiegen group](https://www.uni-siegen.de/physik/)

ML for Signal vs Background

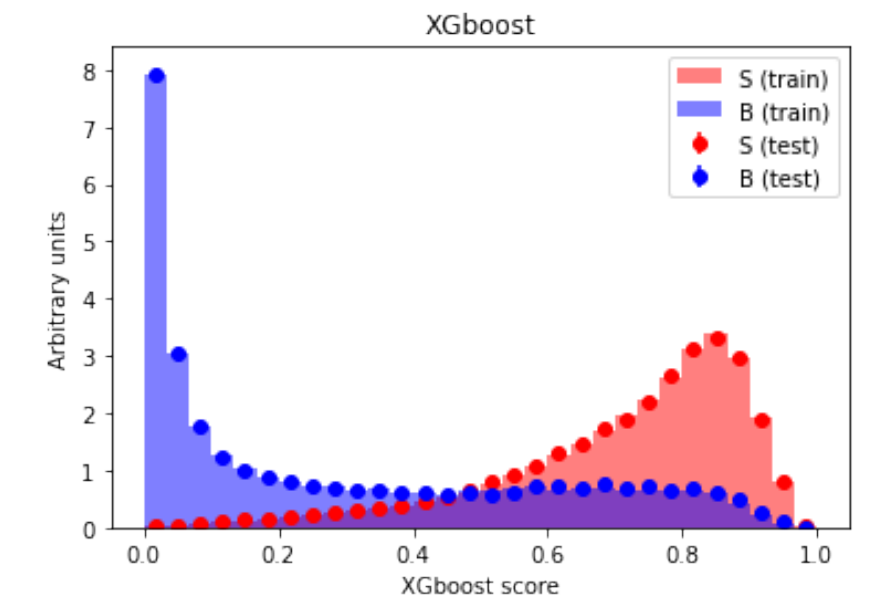
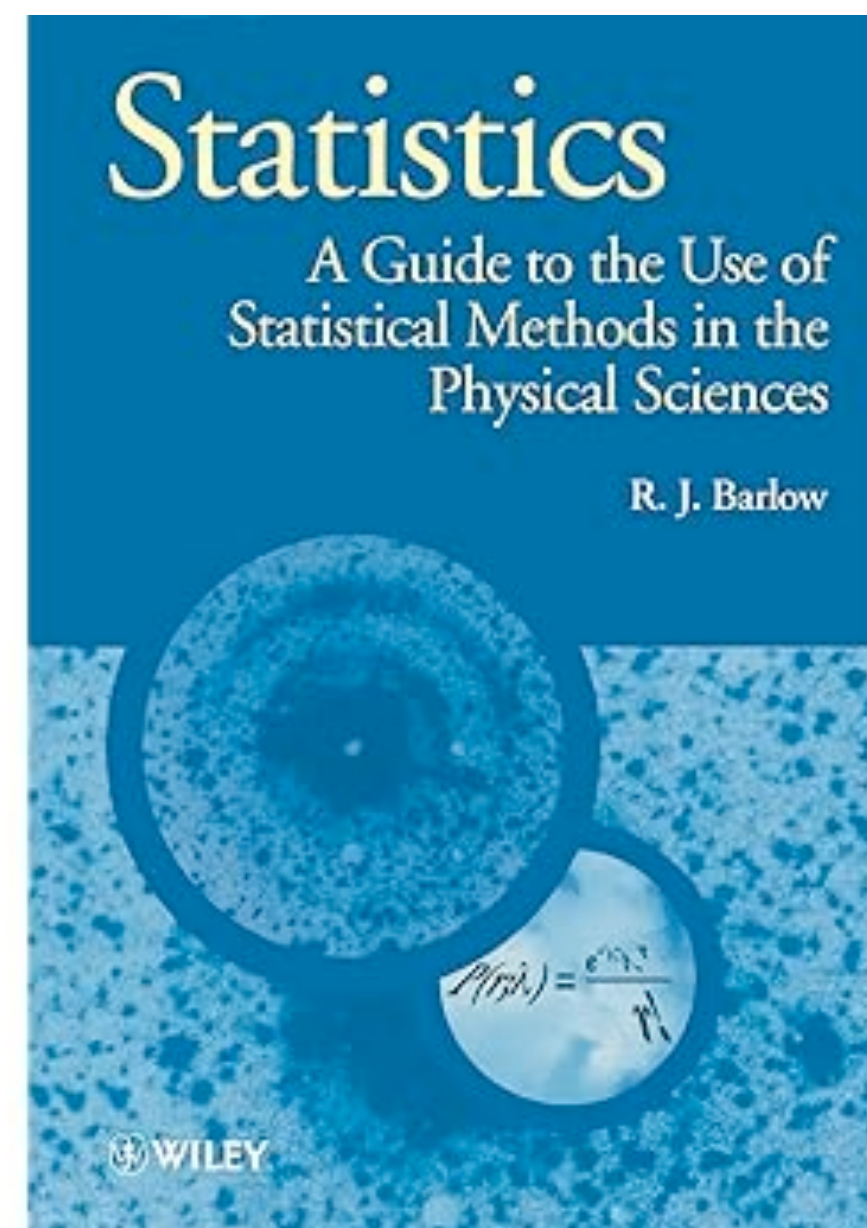


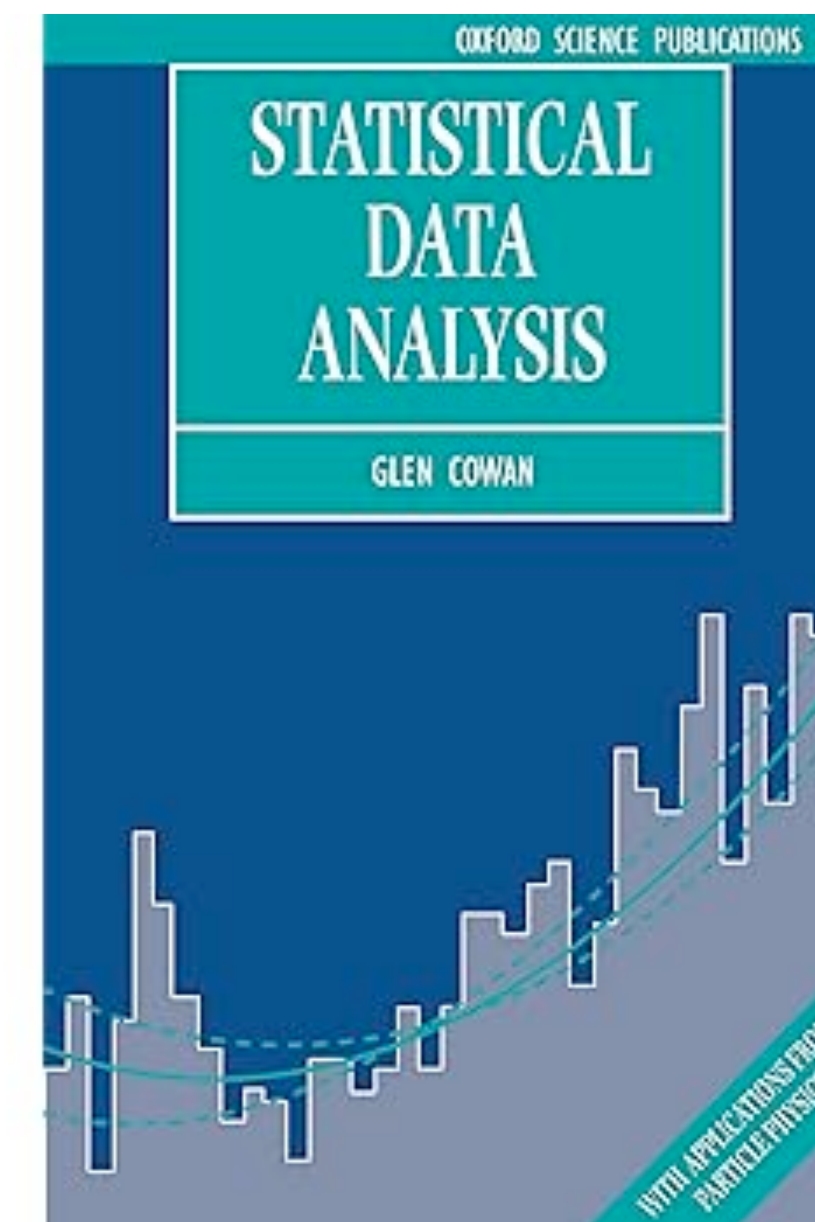
Image: [ML4FP School](https://ml4fp.school/)

The final statistical inference step has remained the same



Publication date : November 1, 1993

Time to modernise?

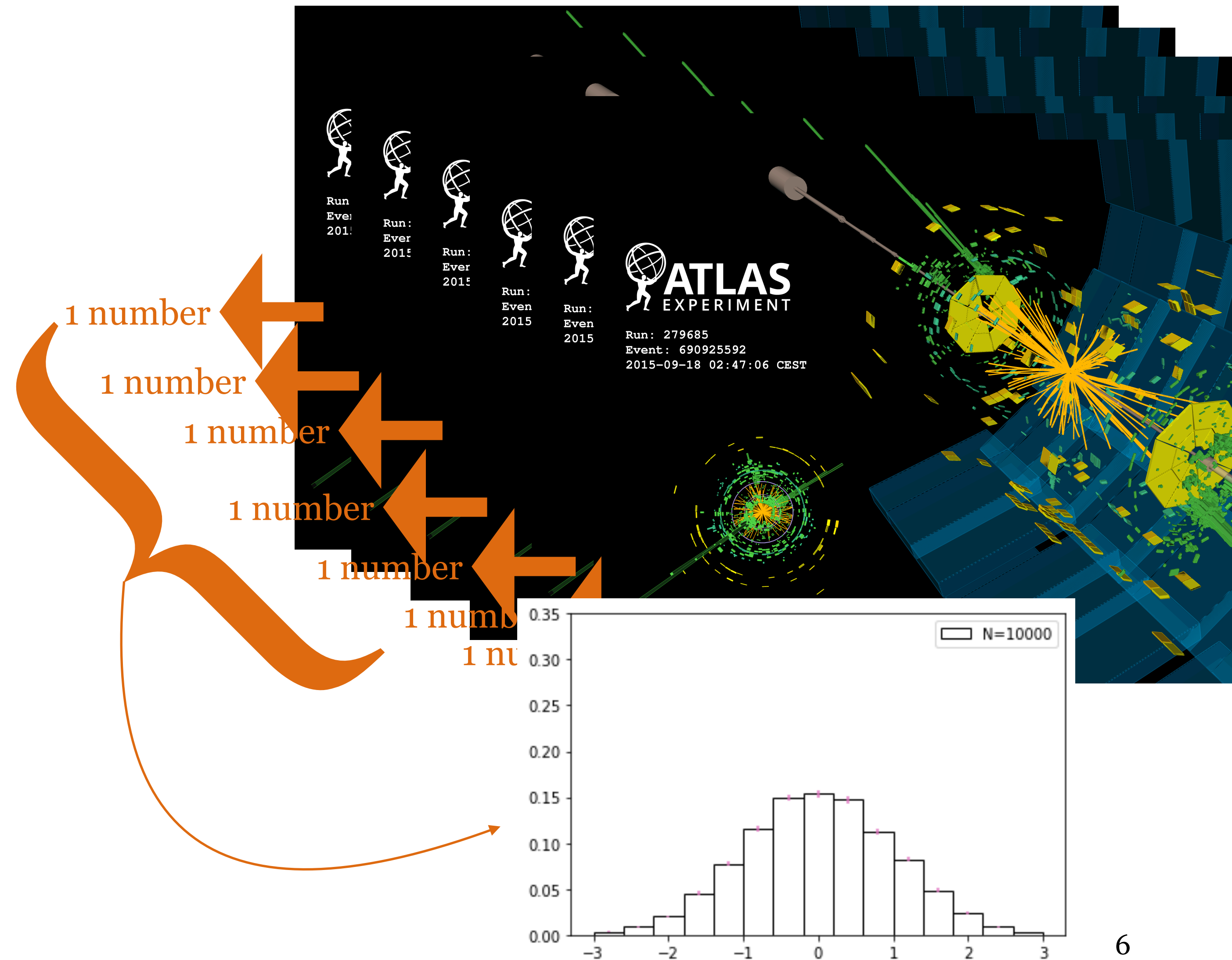


Publication date : December 4, 1997

Traditional Approach: Design one sensitive observable

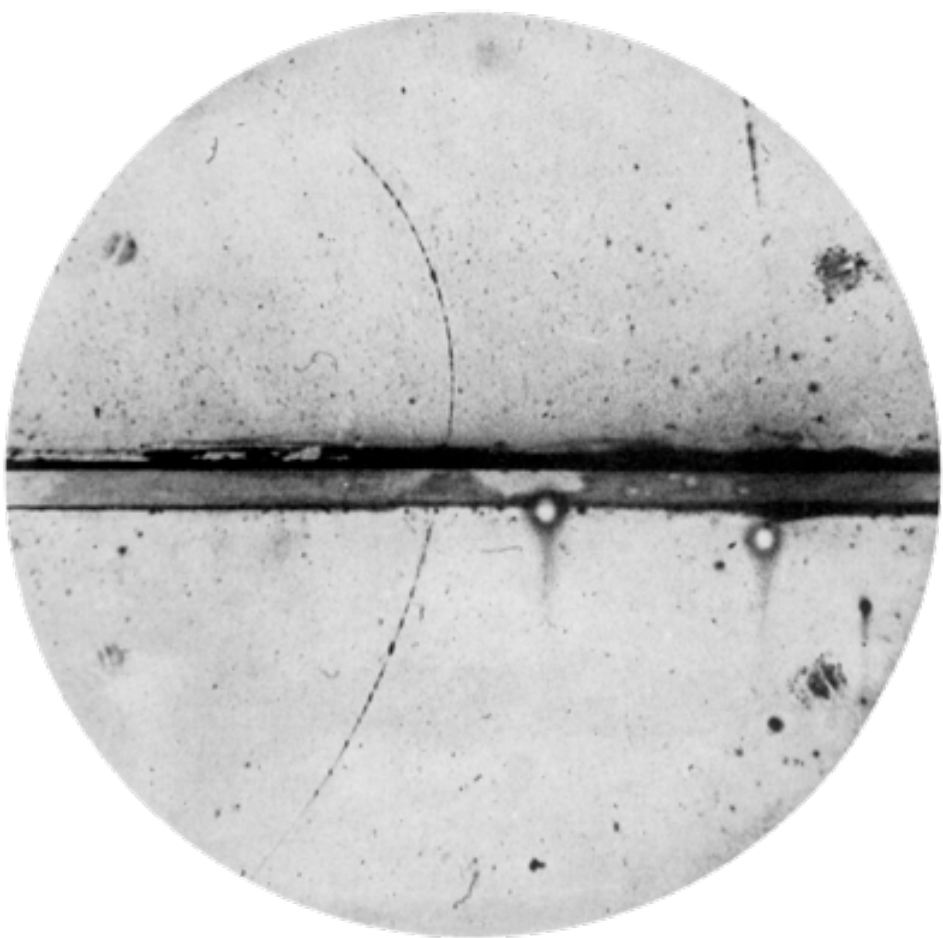
- Detector has $O(100 \text{ million})$ sensors
- Reconstruction pipeline, event selection
- Design one summary variable
 - Compression: $O(100 \text{ million}) \rightarrow 1$
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events



Positron discovery (1930s)

Top quark discovery (1990s)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

What is the probability of seeing 27 events if my null hypothesis predicts 7?

$$p(n_{obs} | n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Counting experiment → Poisson distribution

Single event

Multiple events:
Cut-and-count

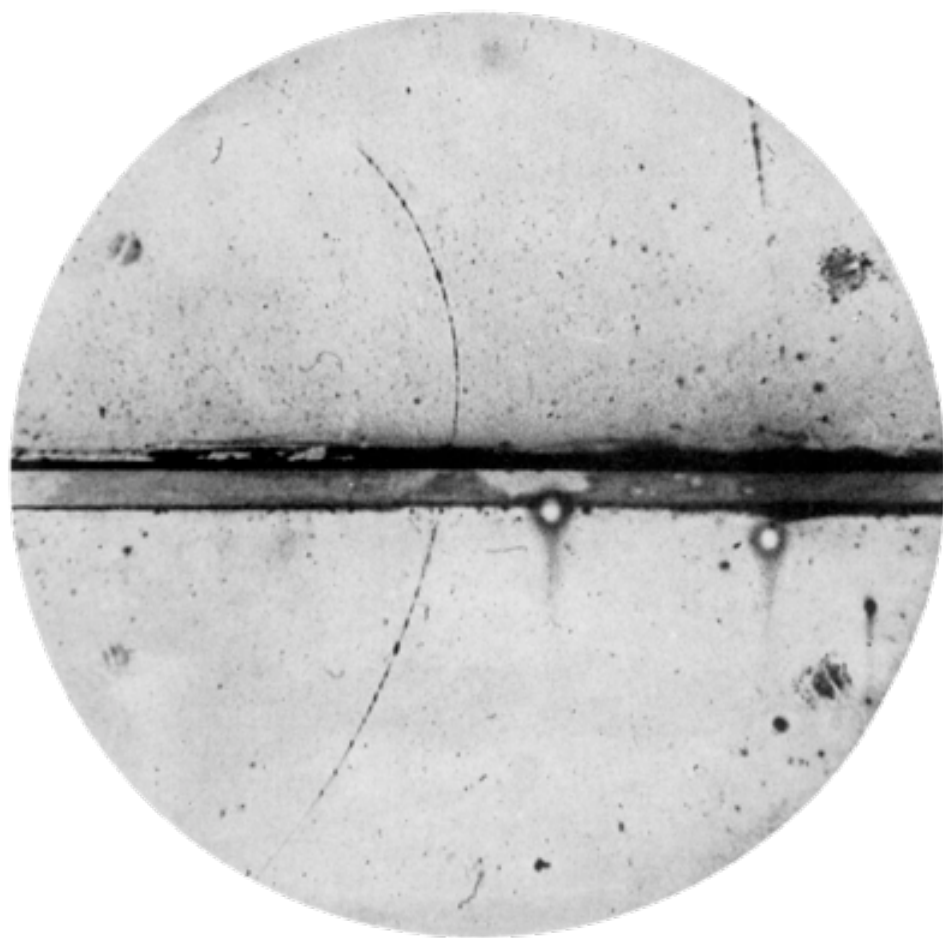
Image: [Wikipedia / PhysRev.43.491](#)

CDF Collaboration: [arXiv:9503002](#)

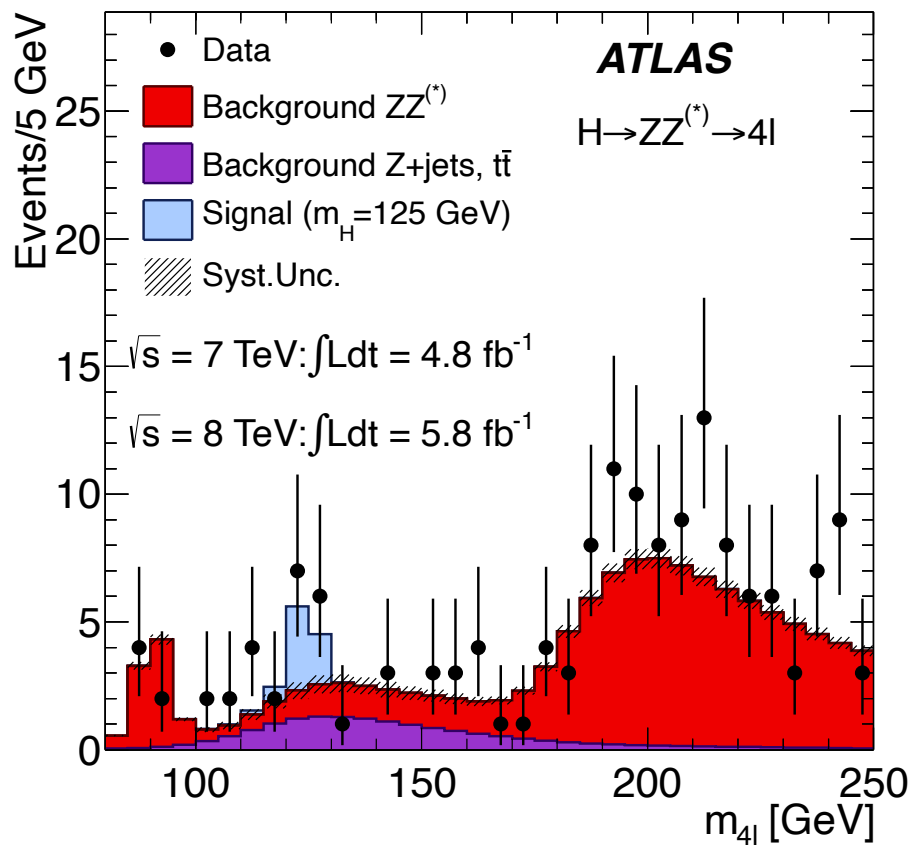
Positron discovery (1930s)

Top quark discovery (1990s)

Higgs boson
discovery (2010s)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}



$$p_{bin}(n_{obs} | n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

$$p(\mathcal{D} | \mu) = \prod p_{bin}(n_{obs} | n_{exp})$$

Histogram bins just a bunch
of independent counting
experiments

Single event

Multiple events:
Cut-and-count

Shape information:
Histogram

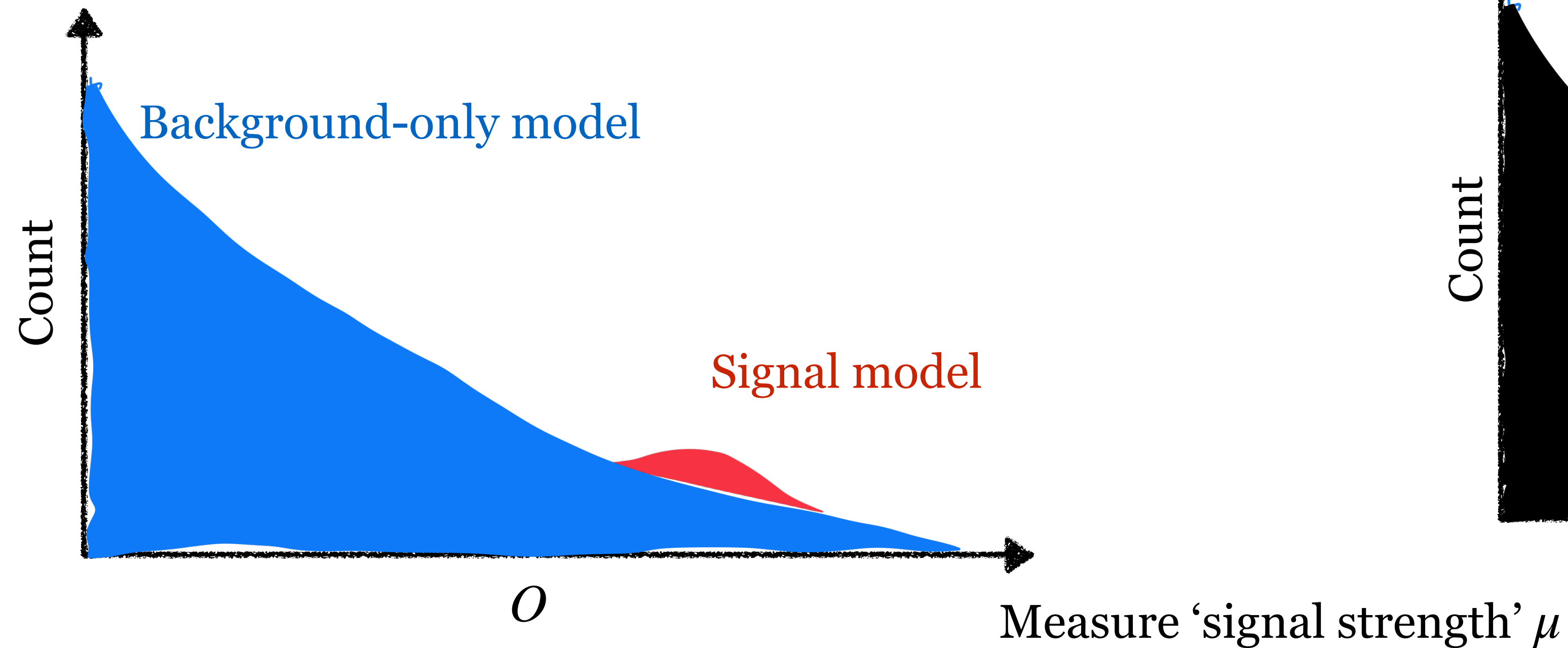
Image: [Wikipedia / PhysRev.43.491](#)

CDF Collaboration: [arXiv:9503002](#)

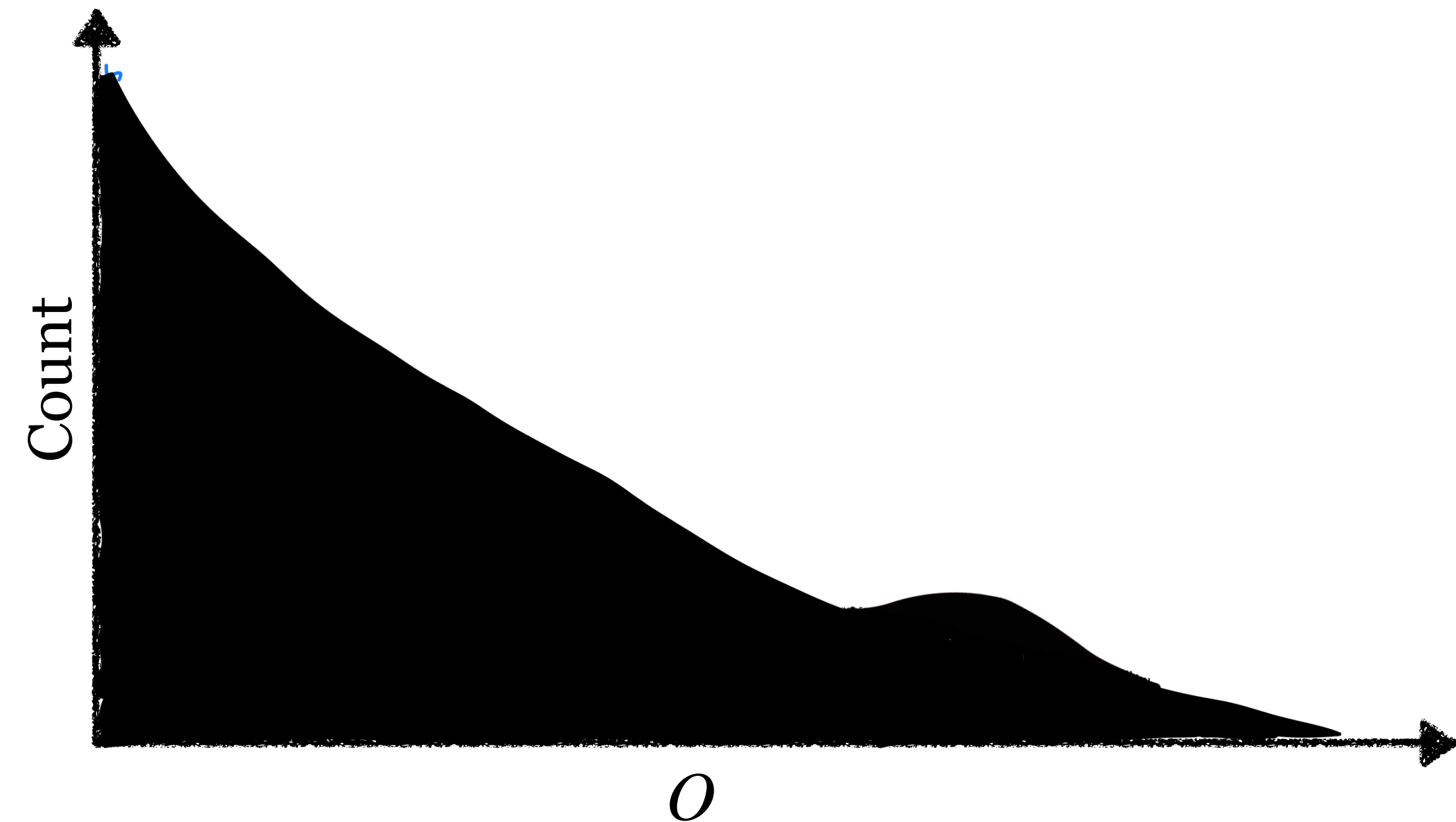
ATLAS Collaboration: [arXiv:1207.7214](#)

Probability density estimation with 1-D histogram

Theory Predictions



Data



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

$\mathcal{D}$ now is a 1D summary of the high-dimensional data

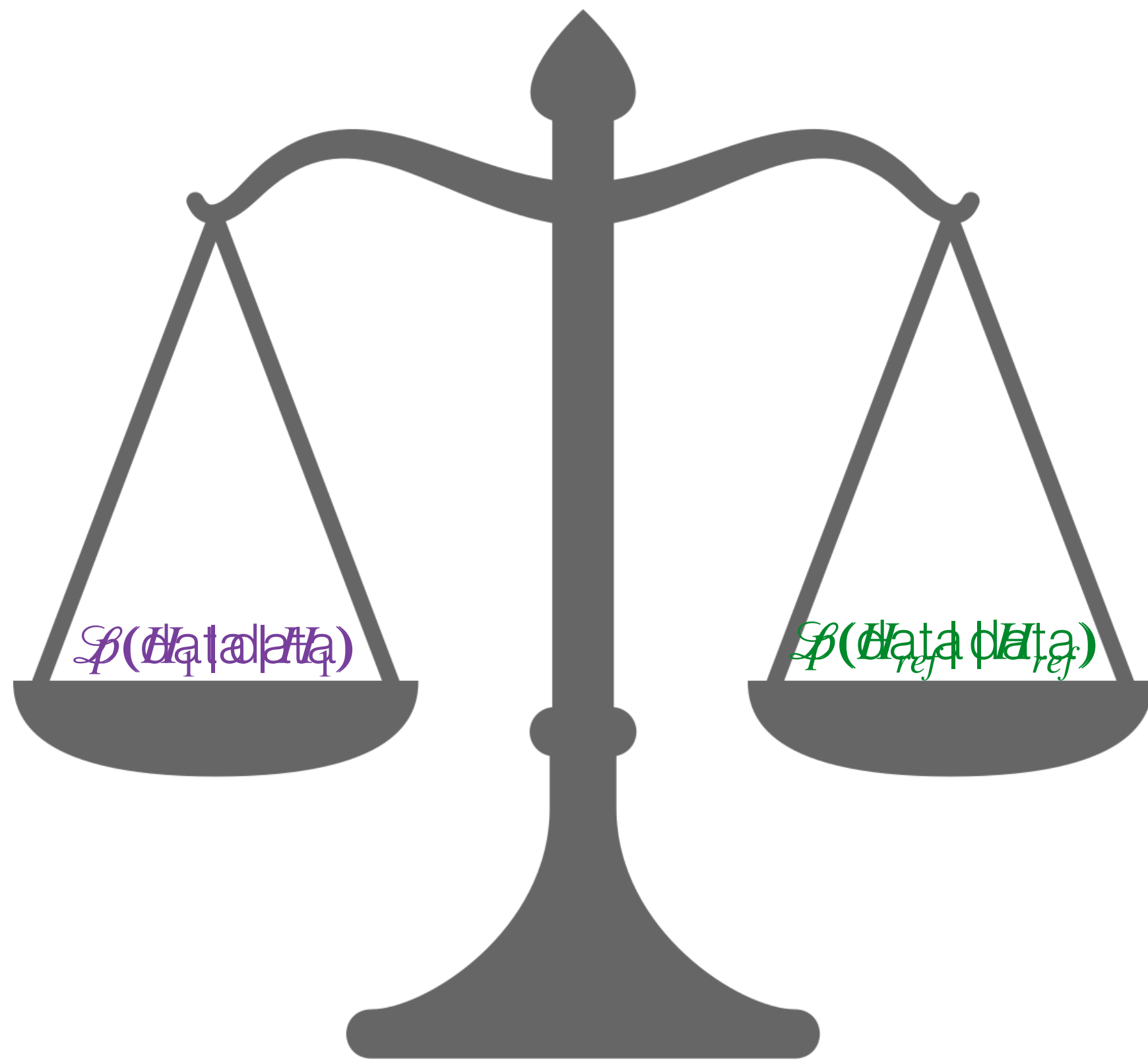
(Frequentist) Hypothesis tests

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

$$H_1 = \{\mu_1\} \quad H_1^1 = \{\mu_1 \alpha_1\}$$

$$\text{Likelihood } \mathcal{L}(\mu, \alpha | \mathcal{D}): \quad p(\mathcal{D} | \mu, \alpha)$$

$$\text{Profile likelihood ratio:} \quad \frac{p(\mathcal{D} | \mu, \hat{\hat{\alpha}})}{p(\mathcal{D} | \hat{\mu}, \hat{\alpha})}$$



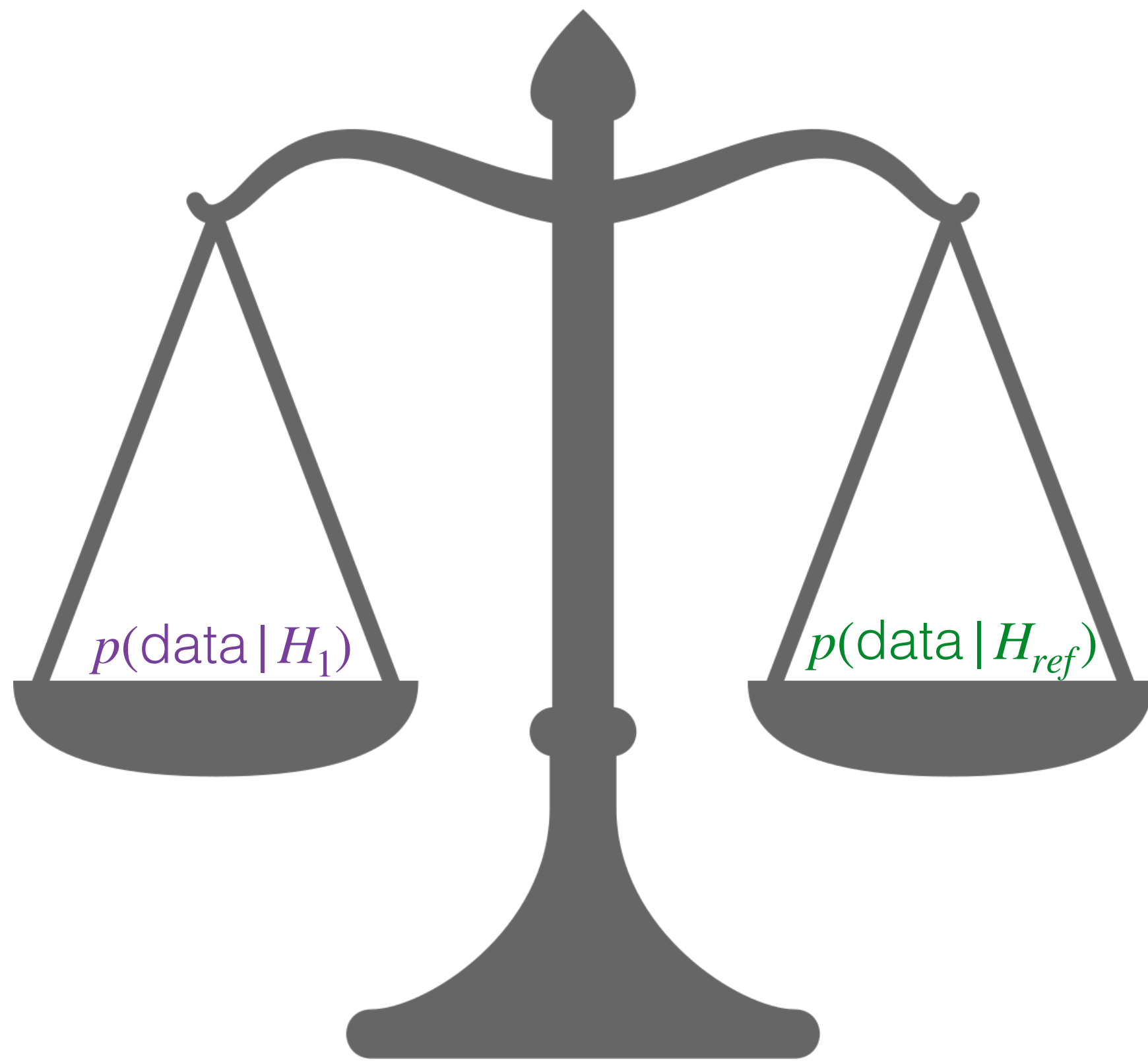
μ : Parameter of interest, you want to measure this

α : Nuisance parameters, you need to measure this to avoid biases. Systematic uncertainties like jet energy scale

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

(Frequentist) Hypothesis tests

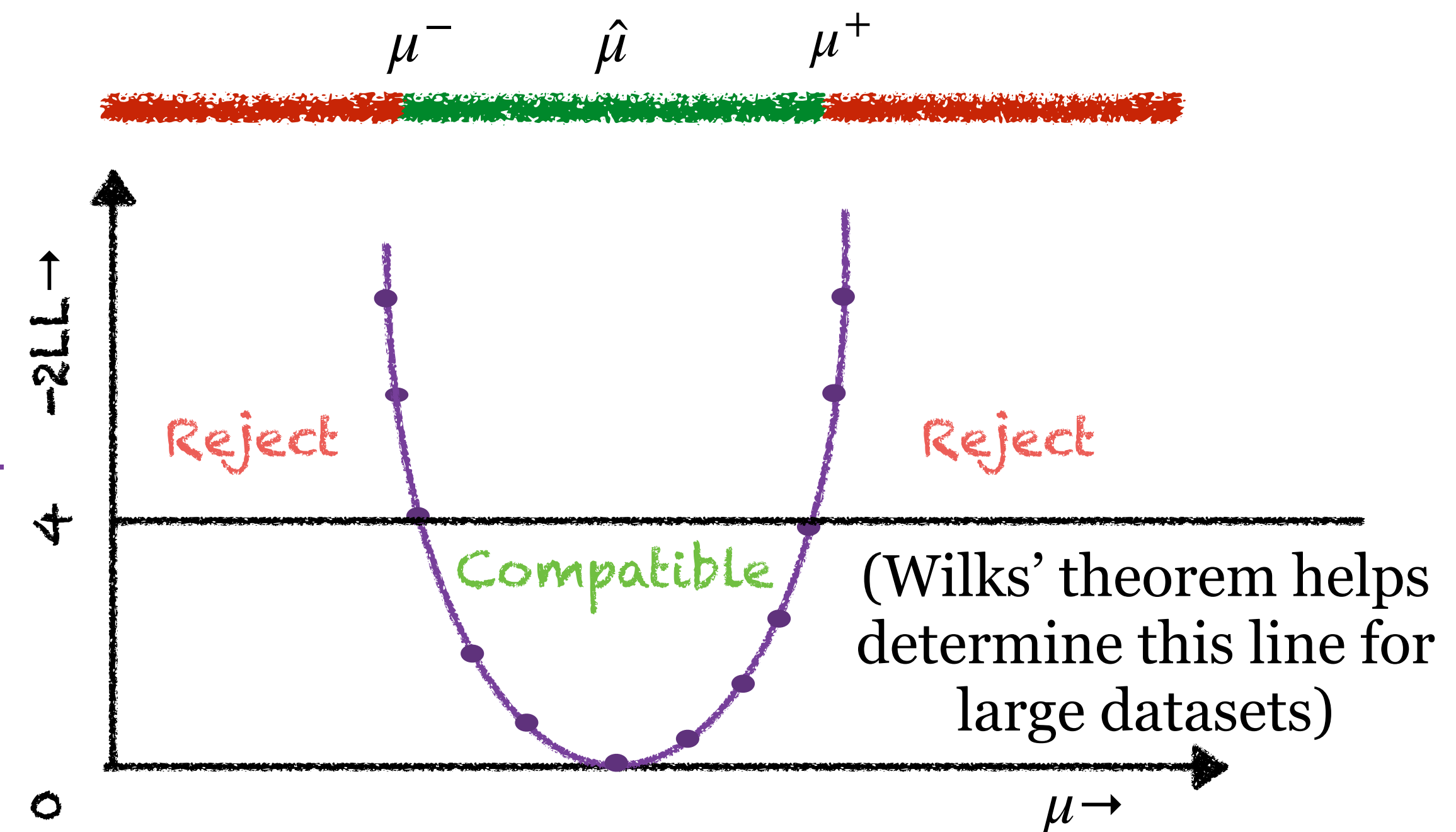
$$\mathcal{L}(H_1 | \text{data}) = p(\text{data} | H_1)$$



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

$$-2 \log \frac{p(\mathcal{D} | \mu, \hat{\alpha})}{p(\mathcal{D} | \hat{\mu}, \hat{\alpha})}$$

Invert the test to get confidence intervals

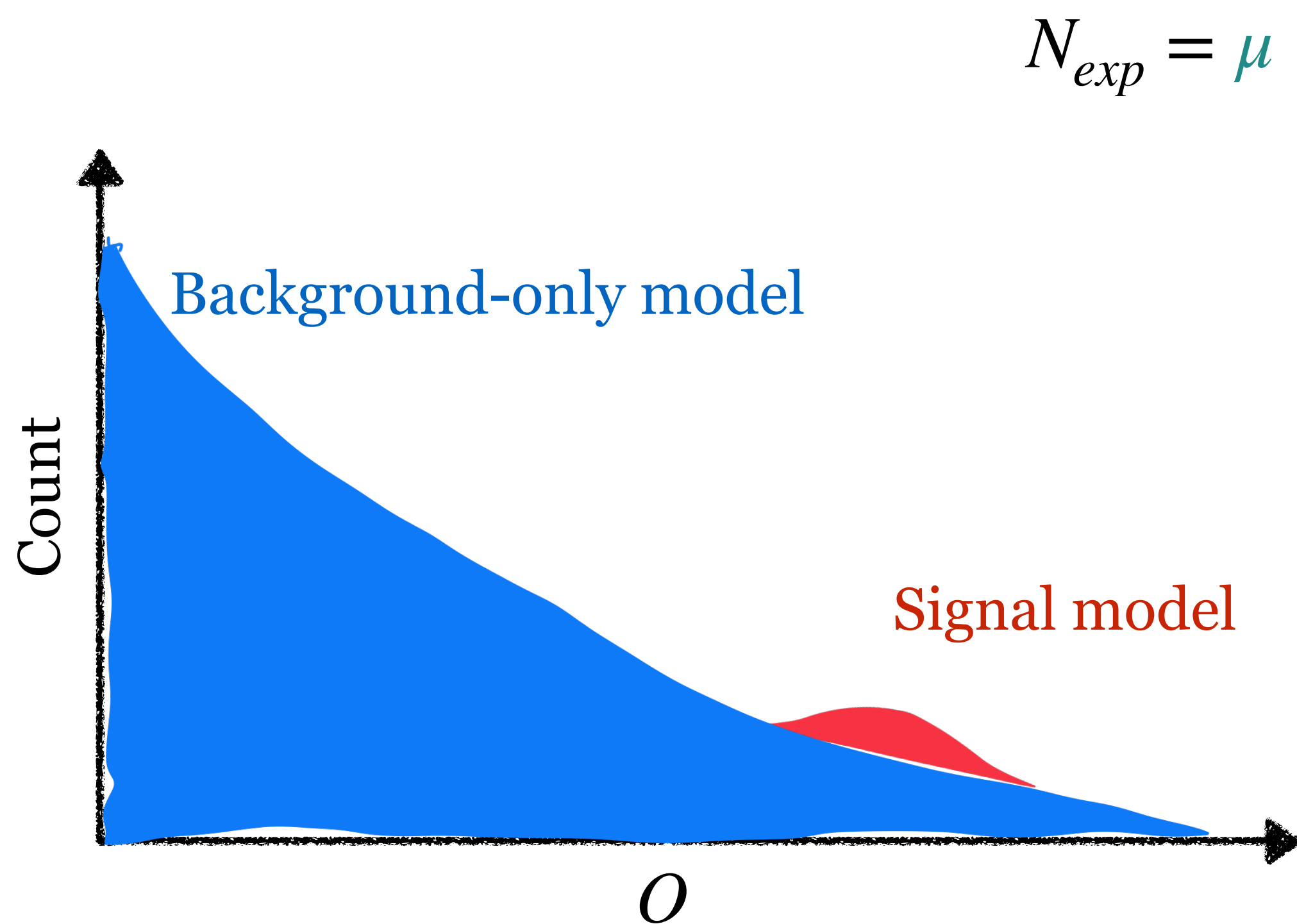


Parameter estimation (infinite hypotheses)

$\mathcal{Z}$ now is a 1D summary of the high-dimensional data

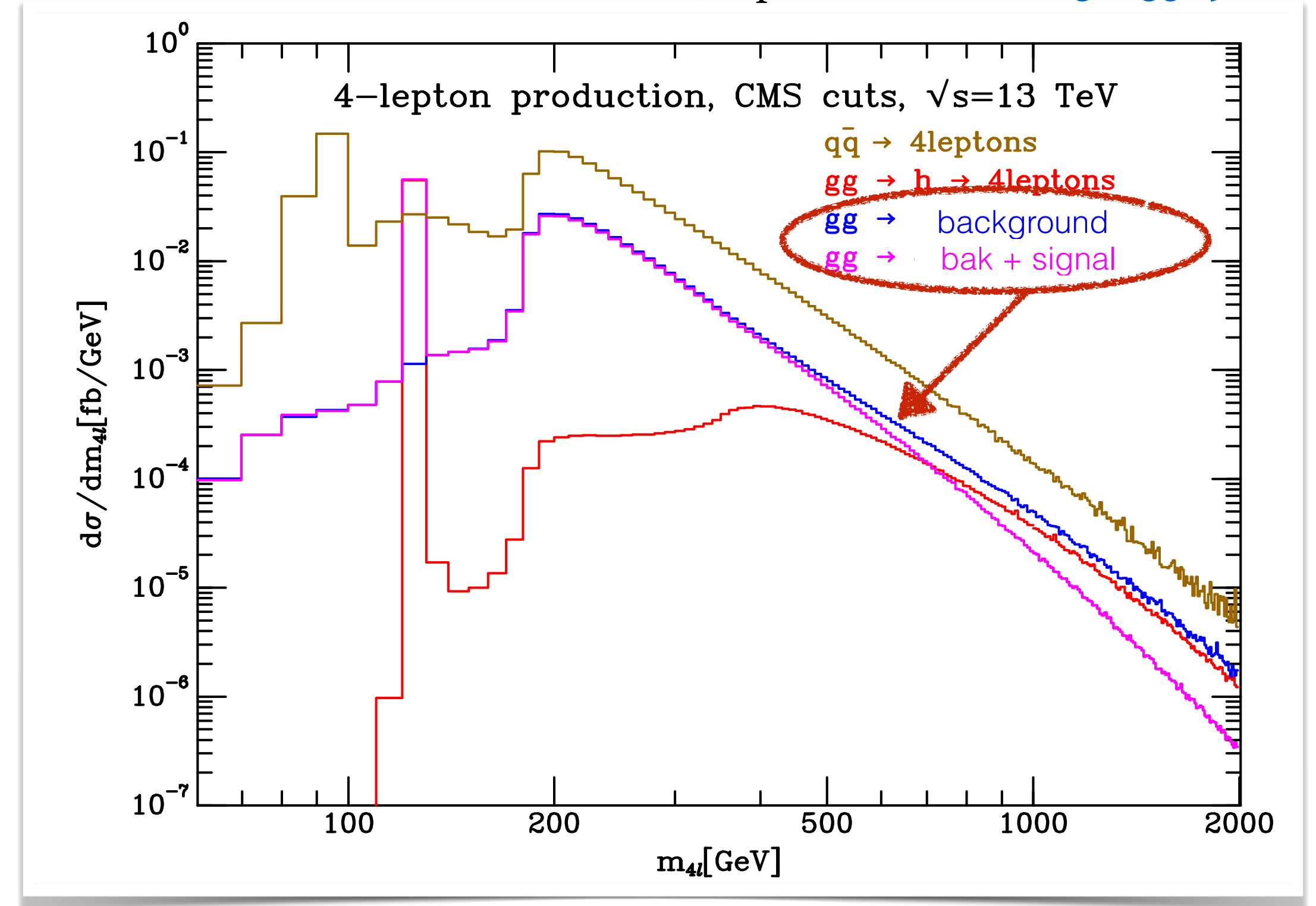
A new challenge: Quantum interference

Non-linear changes in kinematics

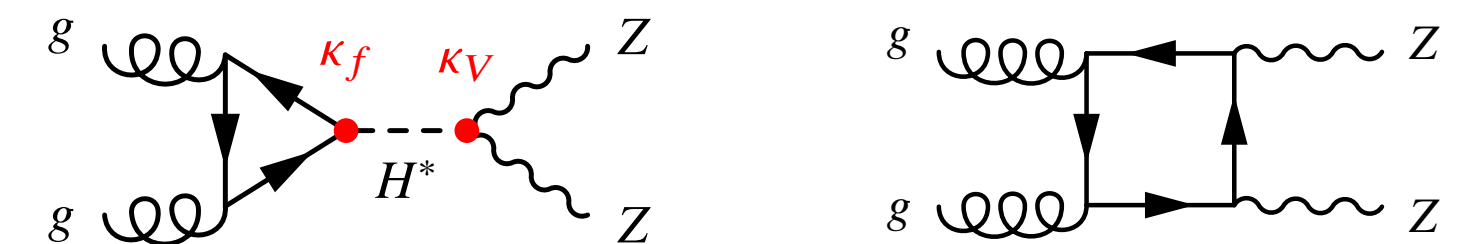


$$N_{exp} = \mu \cdot S + B + \sqrt{\mu} \cdot I$$

Campbell et al: [arXiv:1311.3589](https://arxiv.org/abs/1311.3589)

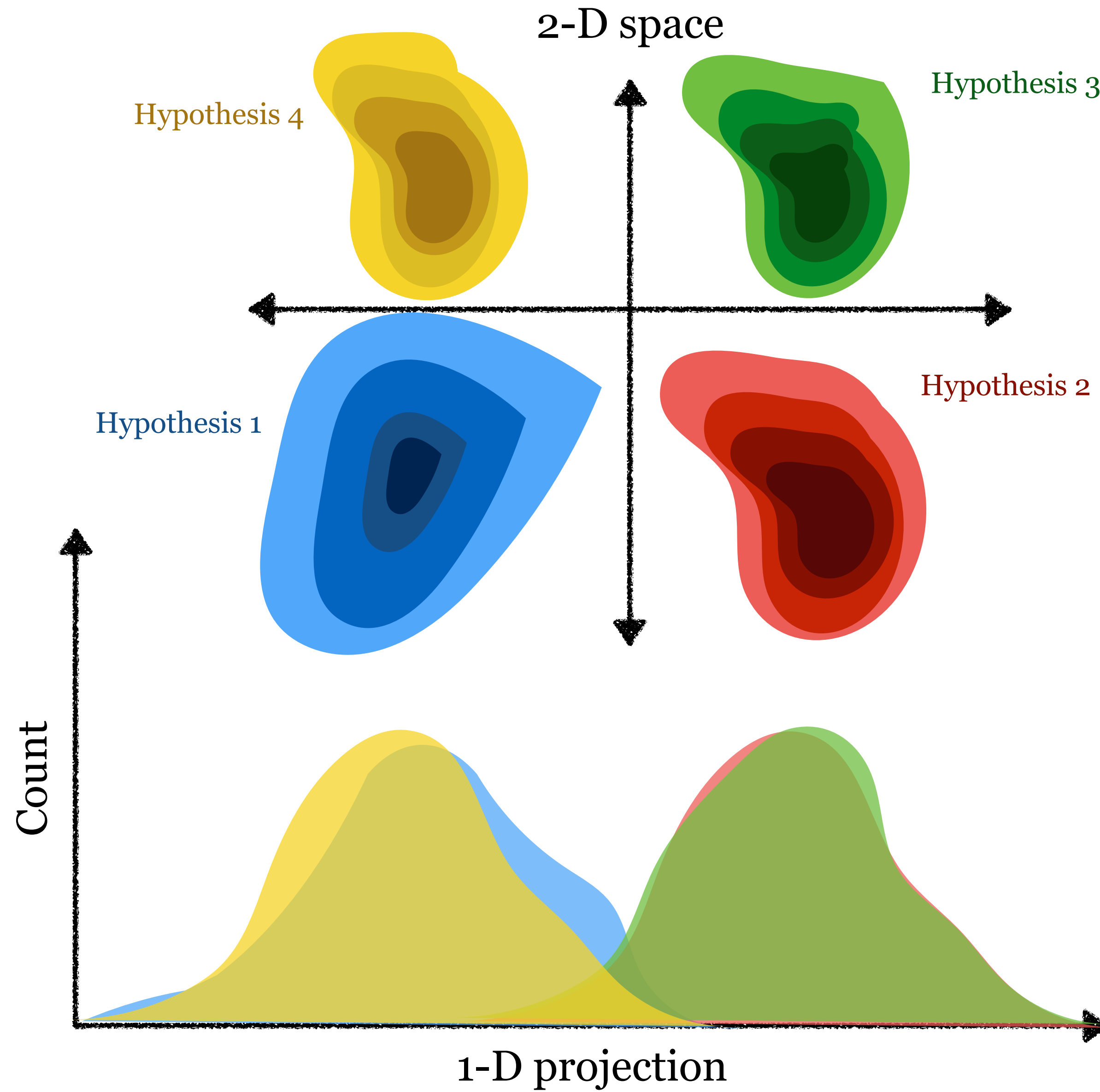


Quantum interference:



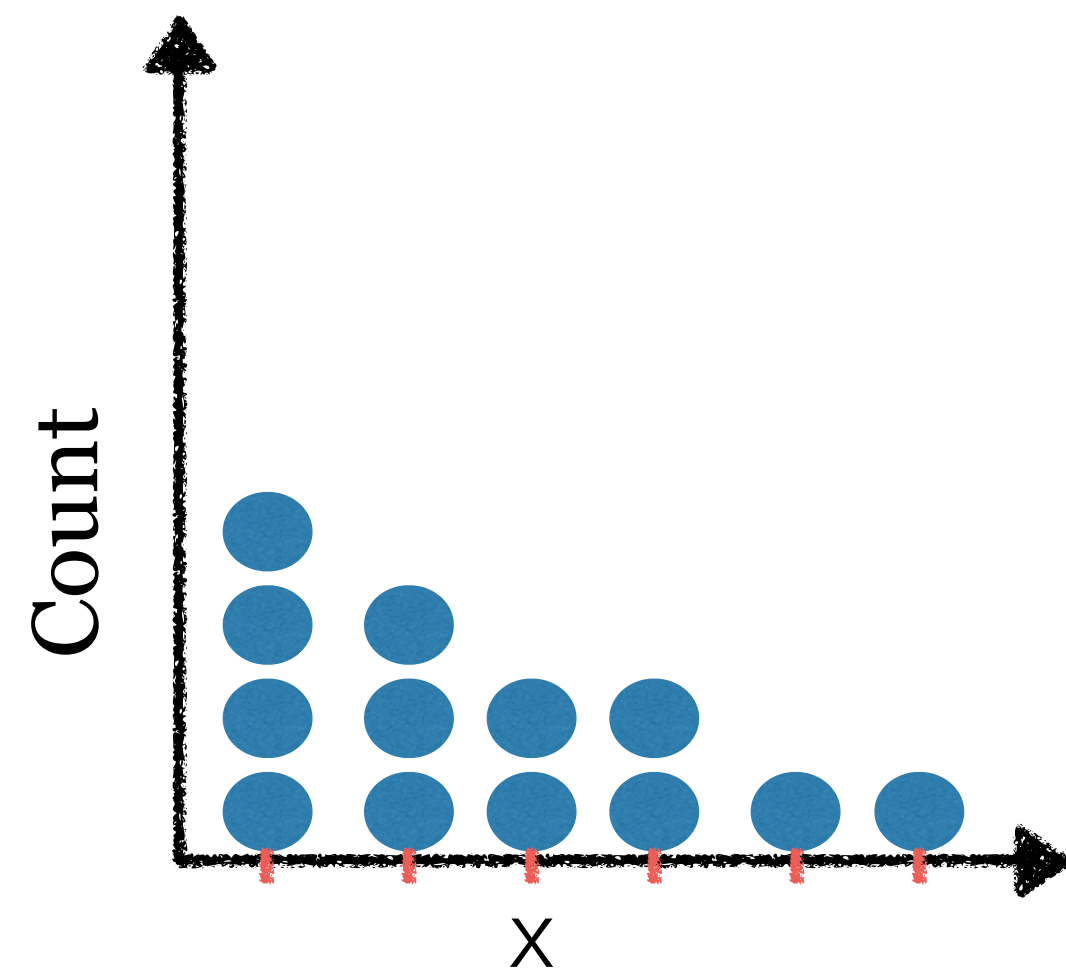
Any 1D histogram will be a lossy summary!

Can we always find a sufficient 1-D summary?

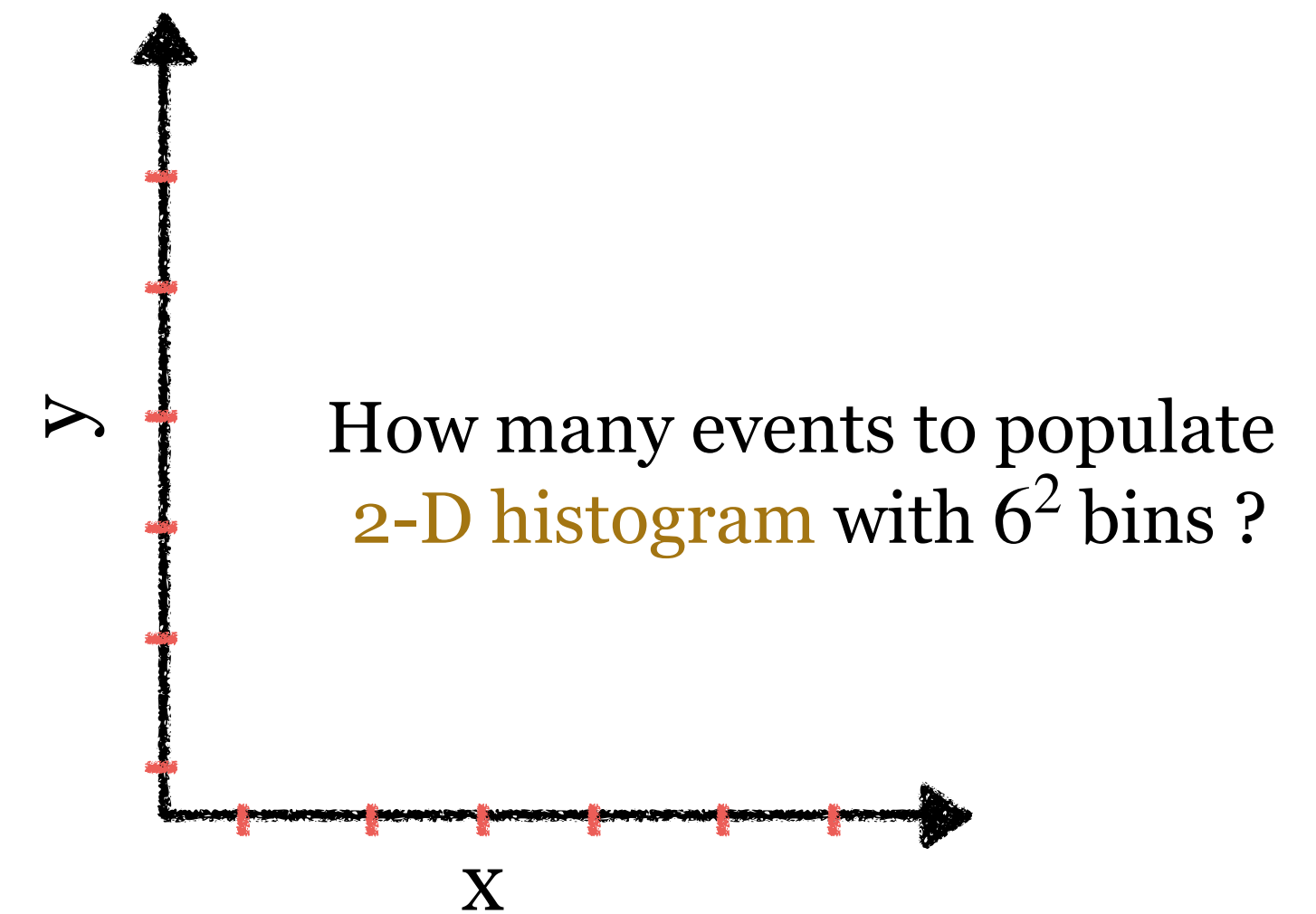


- Clearly separable in 2-D
- A 1-D sufficient summary statistic does not exist for several important particle physics measurements
 - Higgs Width: [hal-02971995\(p172\)](https://arxiv.org/abs/1506.02169): Ghosh et al
 - Higgs Self-coupling: [arXiv:2507.02032](https://arxiv.org/abs/2507.02032): Ghosh et al
 - Systematic uncertainties: [PRD 104.056026](https://arxiv.org/abs/104.056026): Ghosh et al
 - Effective Field Theories: [PRD 98.052004](https://arxiv.org/abs/98.052004): Brehmer et al

But probability density estimation in higher dimensions is hard...



1-D histogram with 6 bins: few events enough to populate it



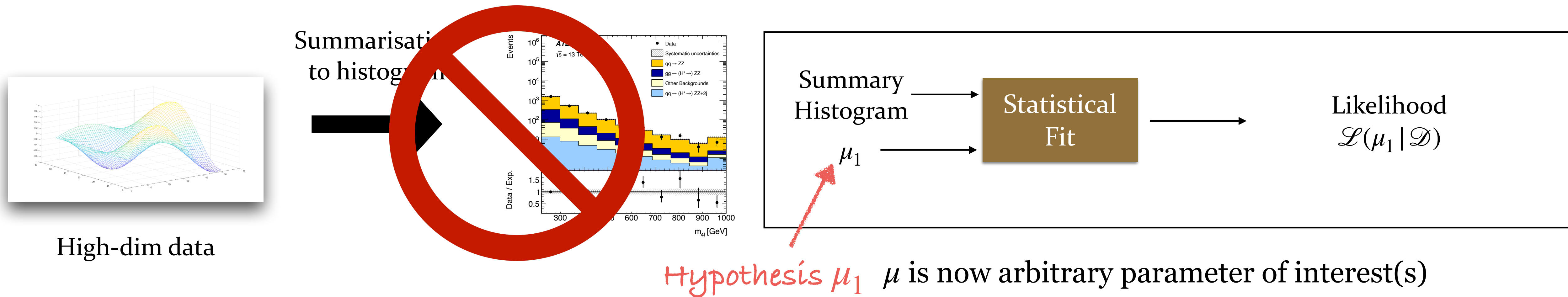
How many events to populate
2-D histogram with 6^2 bins ?

Curse of dimensionality

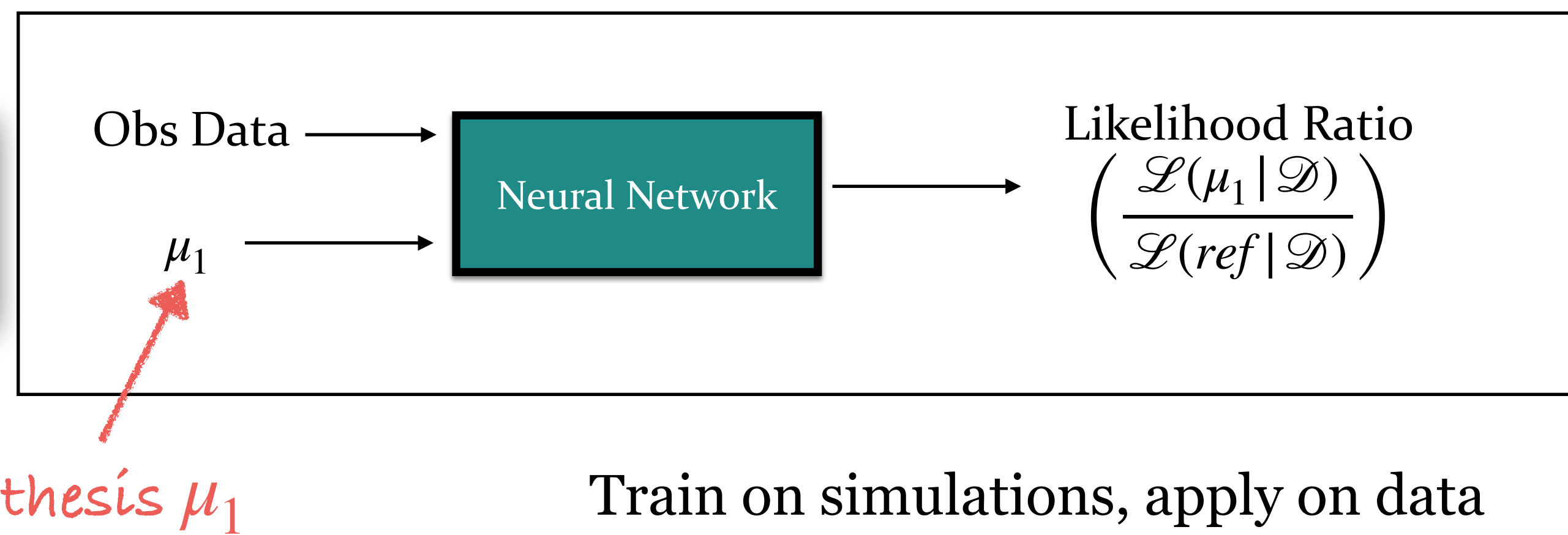
How many events for 50-D histogram
with 6^{50} bins ?

A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:

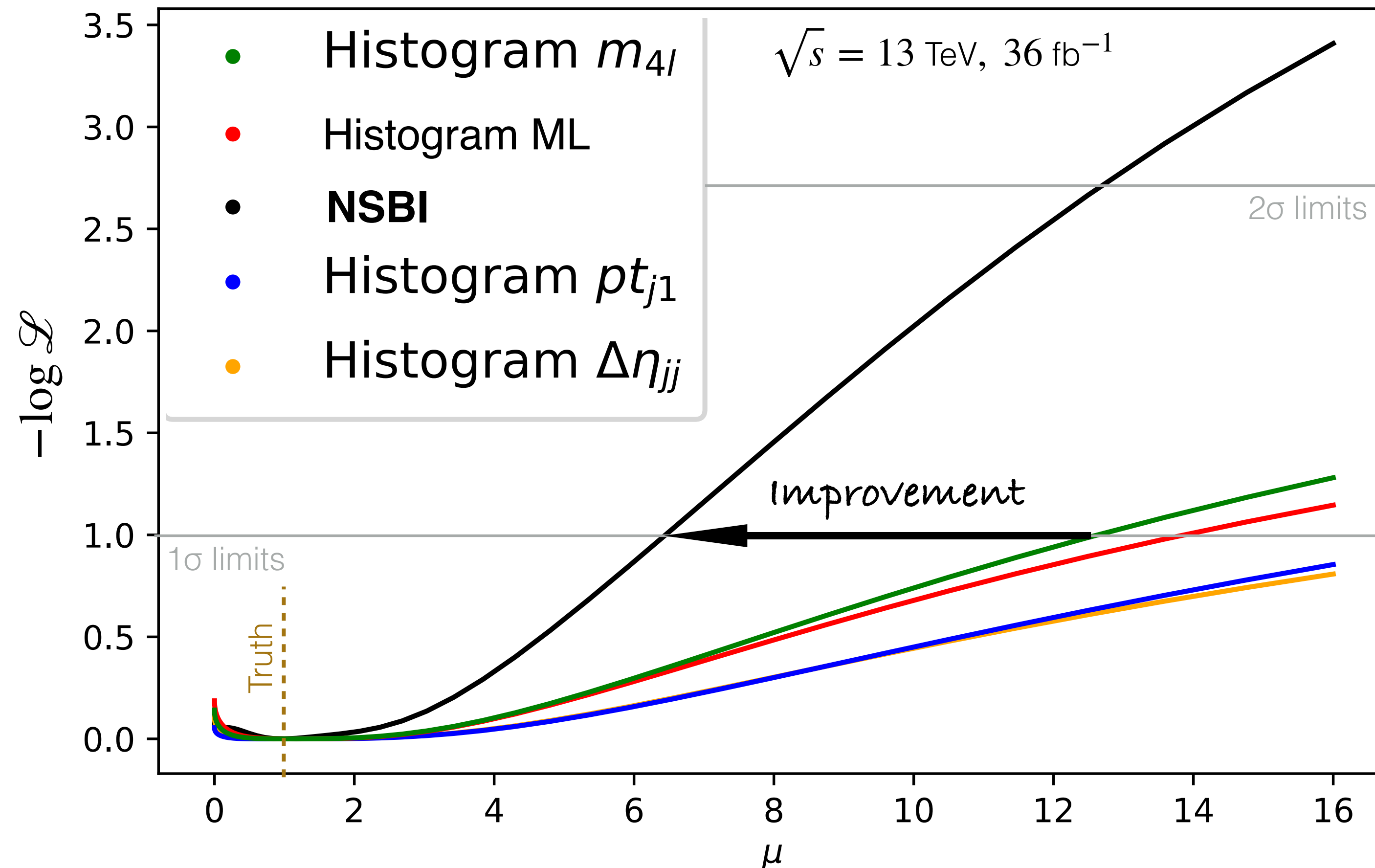


Neural simulation-based inference framework:

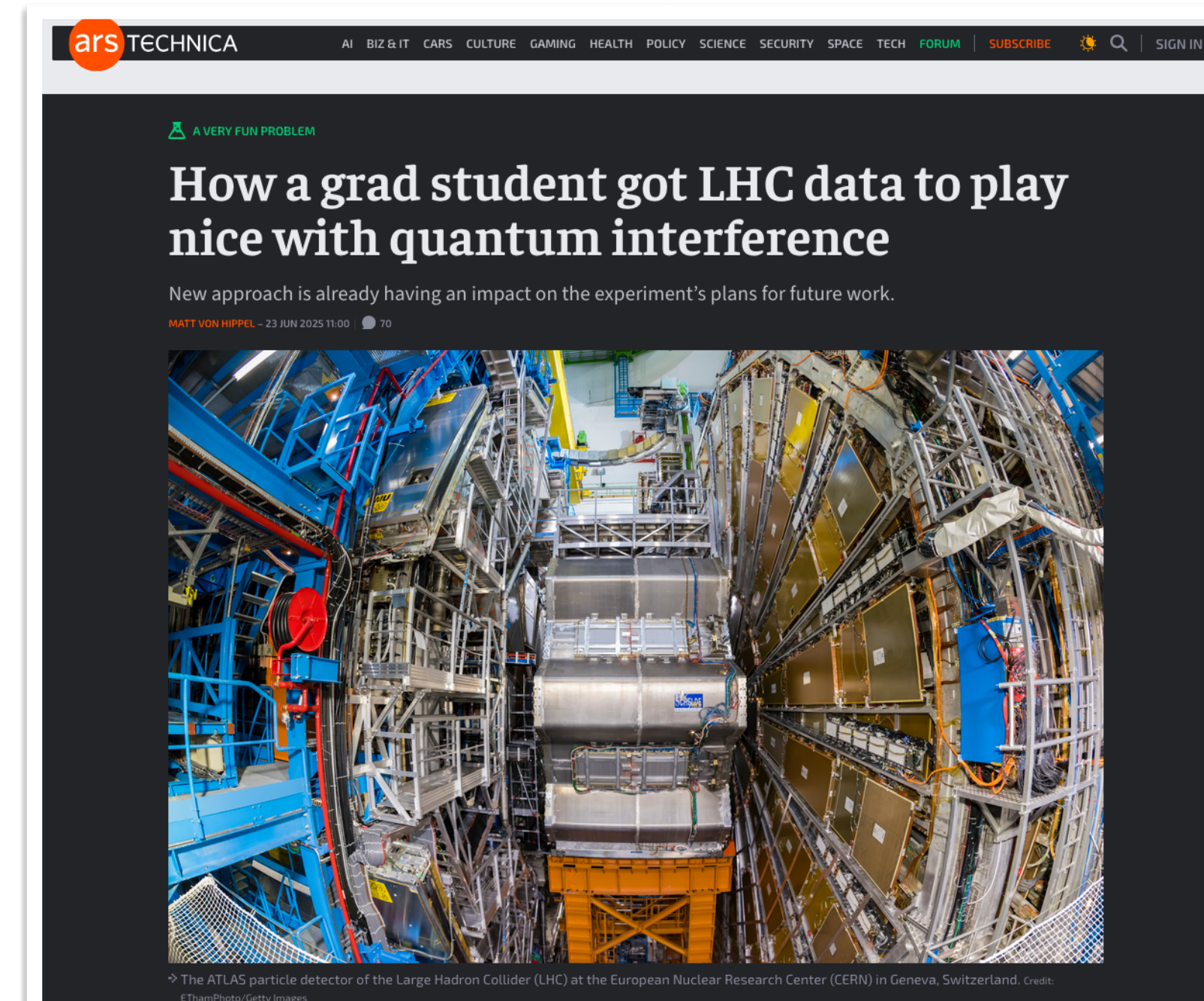


Pheno study of NSBI for Higgs width

[hal-02971995v3](#) (p172): Ghosh & Rousseau, [Thesis](#): Ghosh



Huge improvement possible in a flag-ship Higgs measurement!



Simulations with $\mu = 1$

Is this dream attainable in practice?

- How to ensure robustness? By design and validation
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

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

Reports on Progress in **Physics**



PAPER • OPEN ACCESS

An implementation of neural simulation inference for parameter estimation

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

View the [article online](#) for updates and enhancements.

ATLAS collaboration (incl. Coelho Lopes de Sa, Ghosh, Louppe, Maury, Rousseau, Sandesara, Schaffer)

Reports on Progress in **Physics**



PAPER • OPEN ACCESS

Measurement of off-shell Higgs boson production in the $H^* \rightarrow ZZ \rightarrow 4\ell$ 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 Rouschatzakis, Natalia B Perkins, Qiang Luo et al.

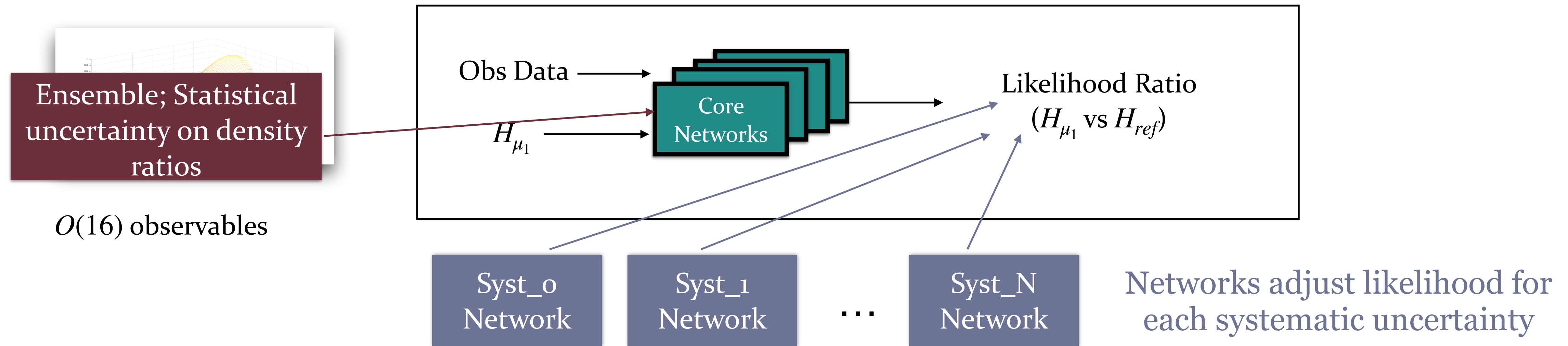
- [70 years of hyperon spectroscopy: a review of strange, baryons, and the spectrum of charmed and bottom baryons](#)
Volker Crede and John Yelton

Is this dream attainable in practice?

- **How to ensure robustness? By design and validation**
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

Combine physics knowledge with AI using ‘semi-parametric’ approach

NN learns probability ratios in high dim



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

Training details

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



Is this dream attainable in practice?

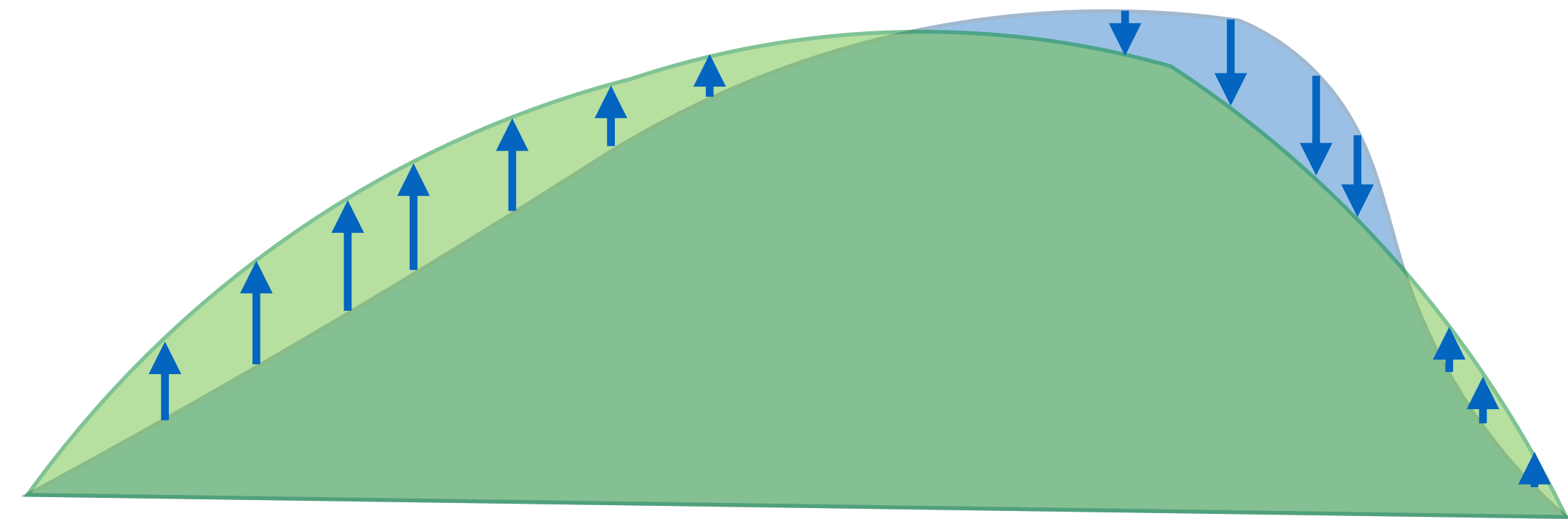
- How to ensure robustness? By design and validation
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

Design a list of diagnostic tests, let's see a few examples

Validate quality of LR estimation with re-weighting task

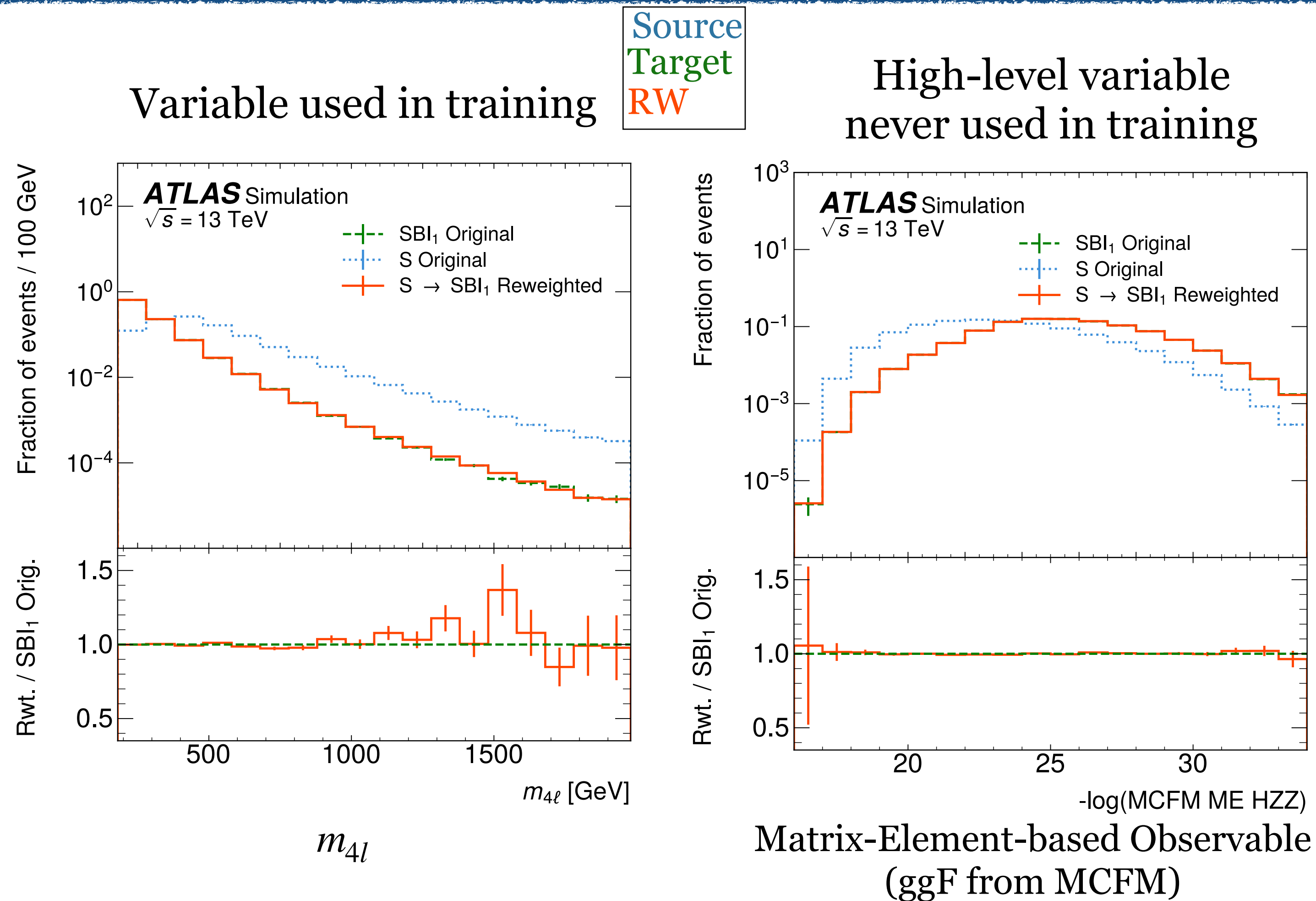
Reweighting: Calculate weights w_i for events x_i in **blue sample** to match **green sample**

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

Already estimated using an ensemble of networks

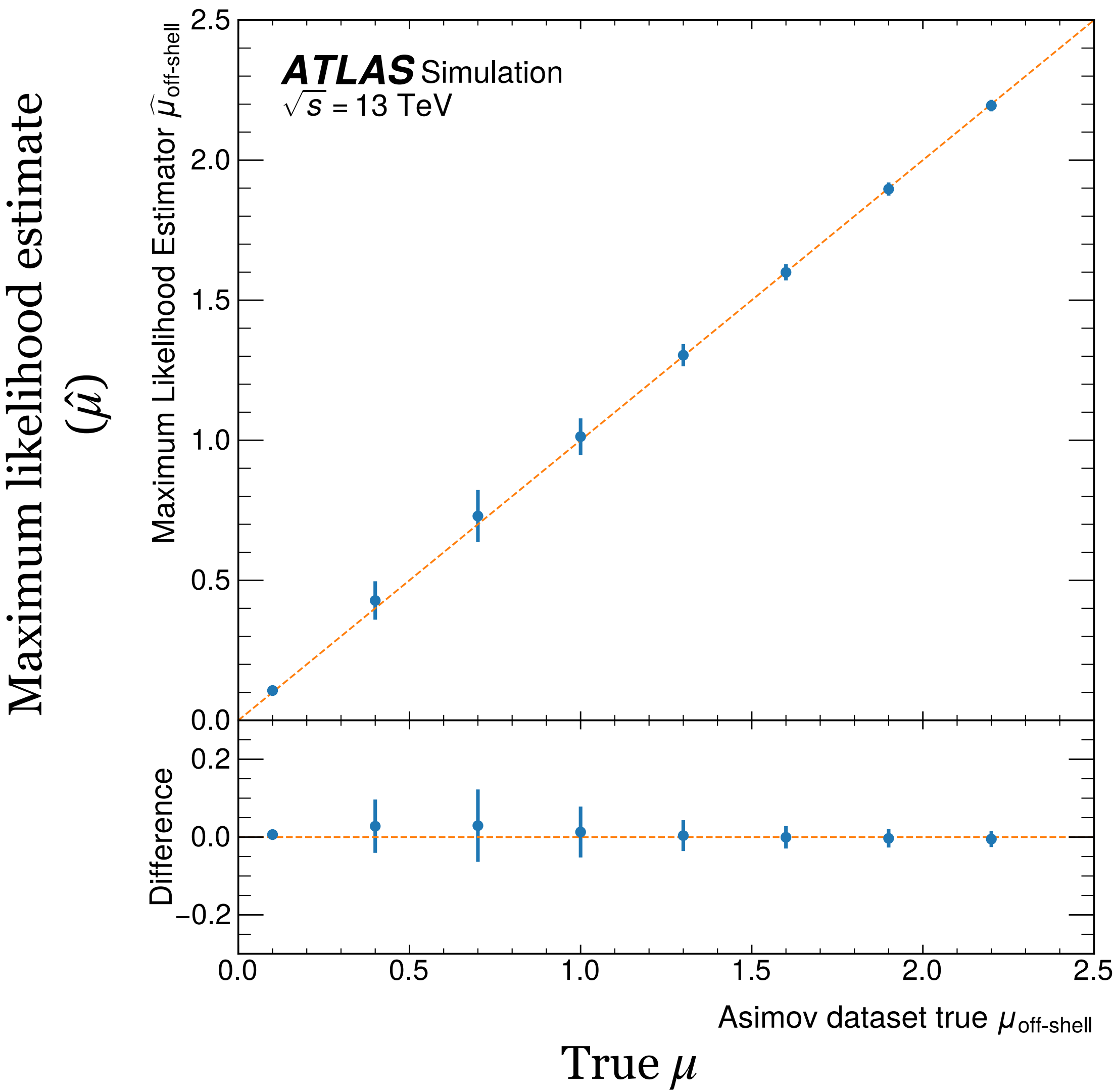
Re-weight closures



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

Hoping to see **Target** and **RW** match

Testing full analysis on simulated samples from different values of μ



Ensure that true value is recovered by NSBI analysis on simulations

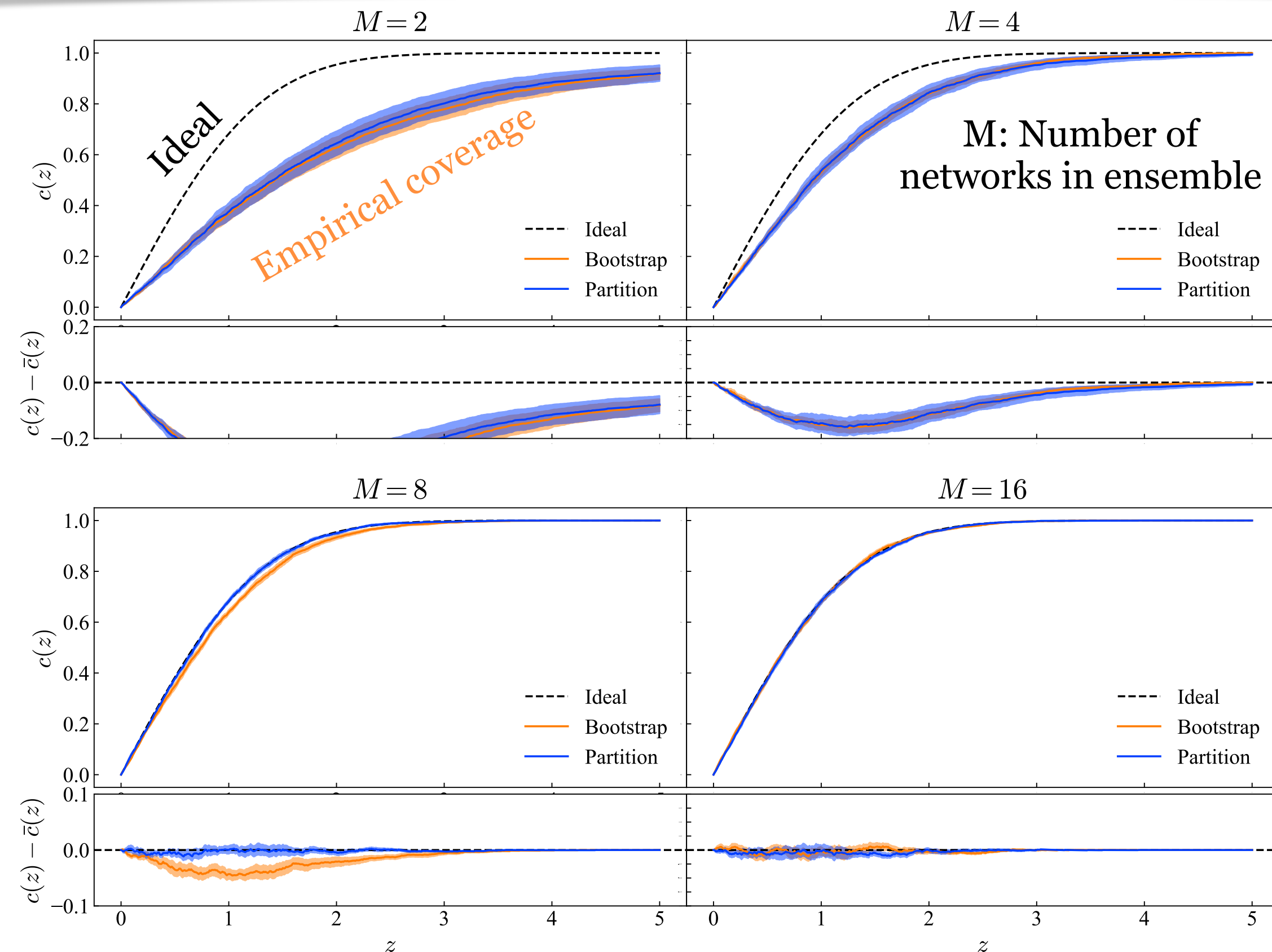
Uncertainties on the NN models themselves

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

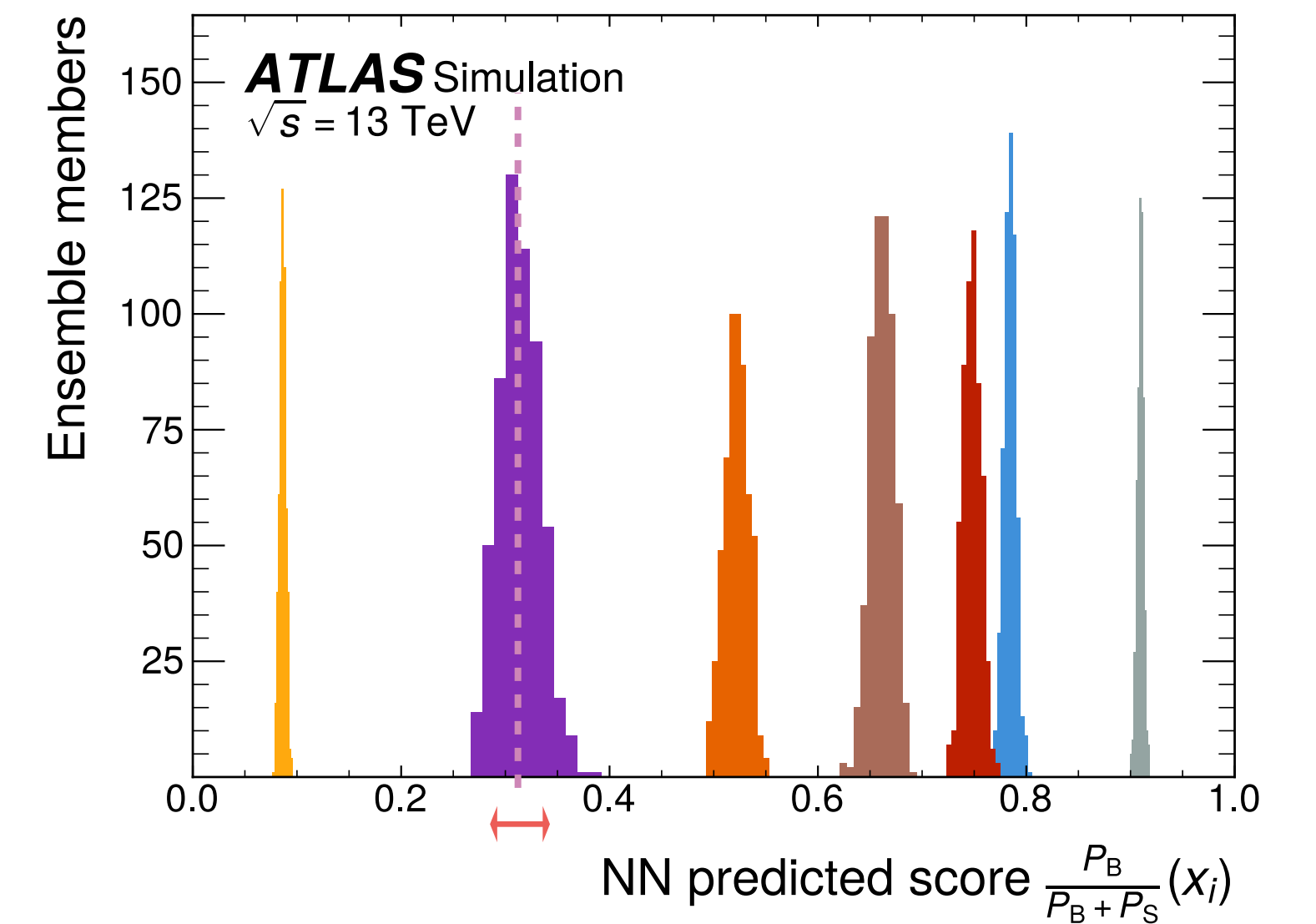
Sean Benevedes^{1,2,*} and Jesse Thaler^{1,2,†}

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

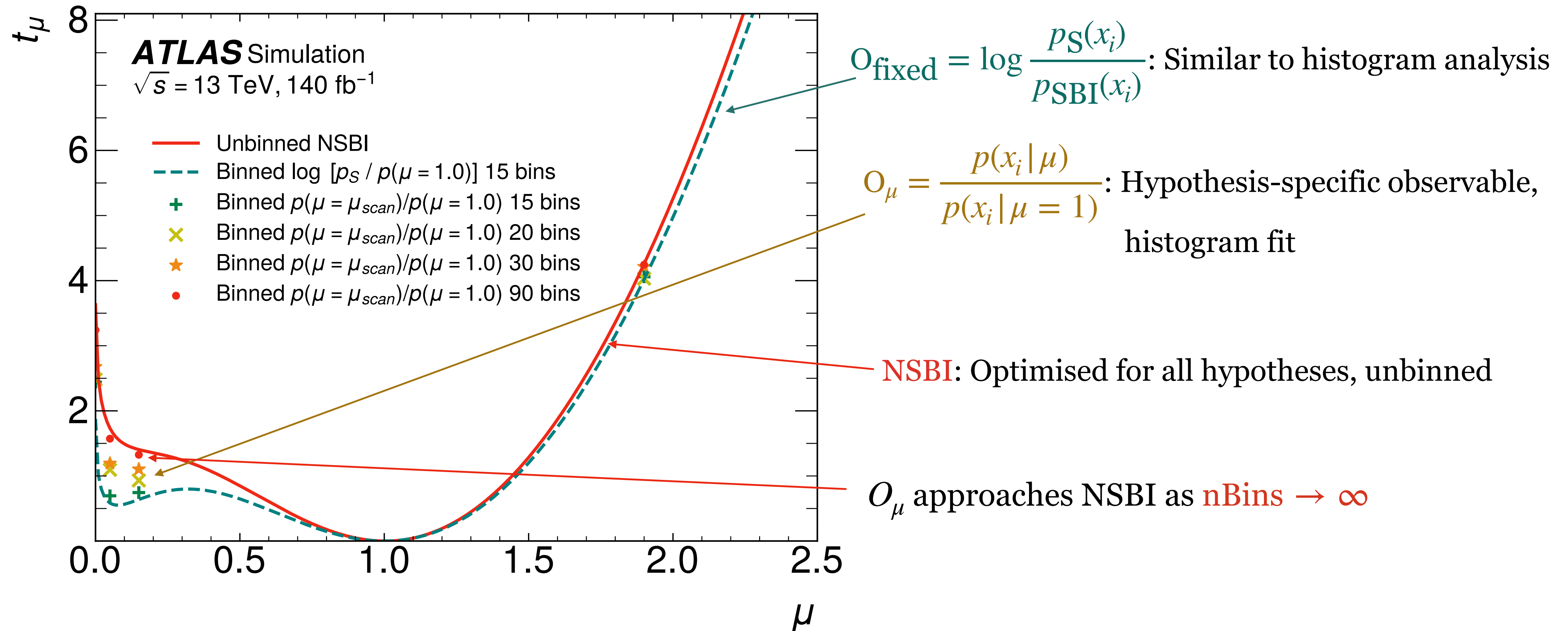


- Genuine 95% coverage, with much smaller ensembles
- Elegant, mathematically motivated method to estimate uncertainties

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



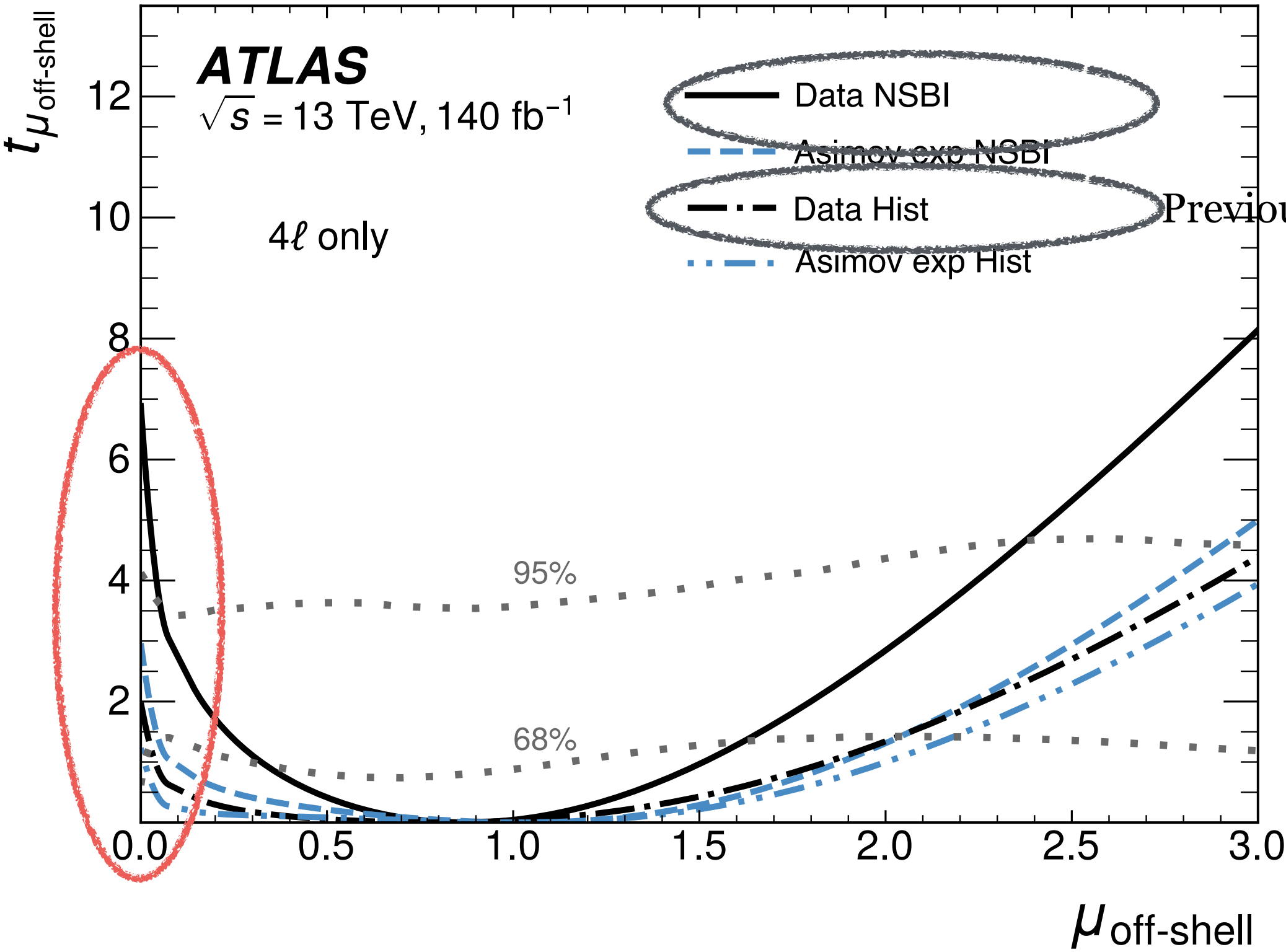
Why does it work better than traditional analyses?



Narrower curve $\rightarrow$ more precise measurement

Application to a flagship Higgs measurement at LHC

NSBI vs (published) histogram analysis



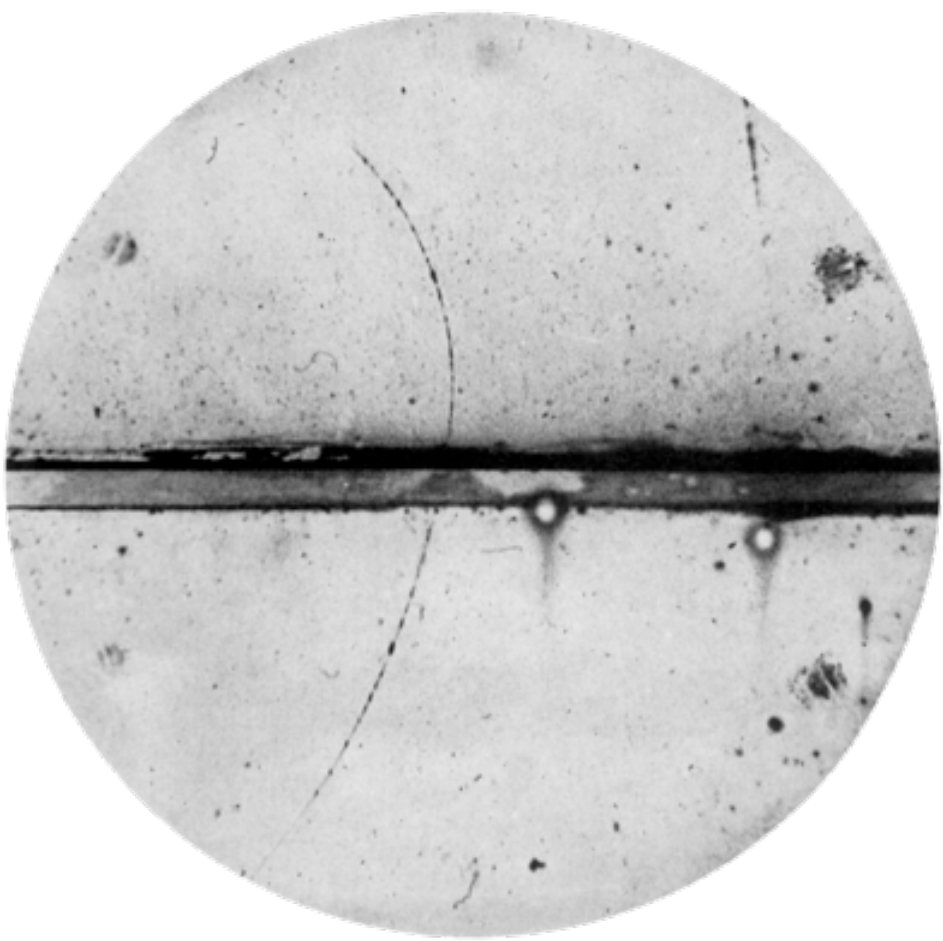
Unprecedented improvement in ability to reject null hypothesis!

Positron discovery (1930s)

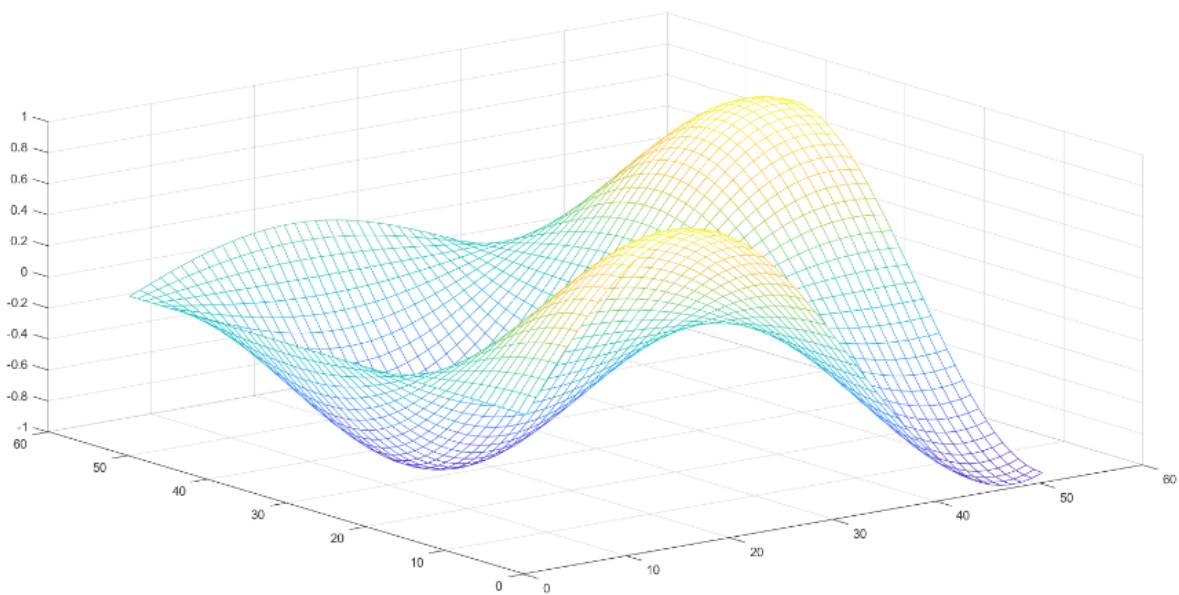
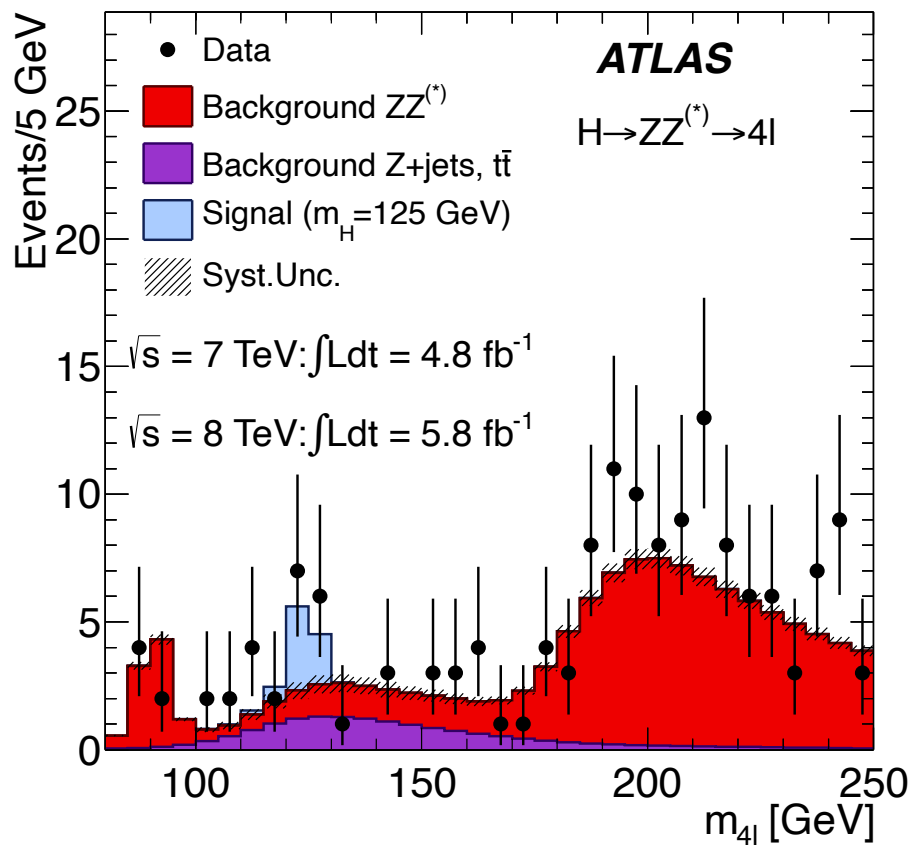
Top quark discovery (1990s)

Higgs boson discovery (2010s)

Future discovery (2020s ?)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}



Single event

Multiple events:
Cut-and-count

Shape information:
Histogram

High-dim shape information,
continuous (i.e. unbinned):
Neural inference

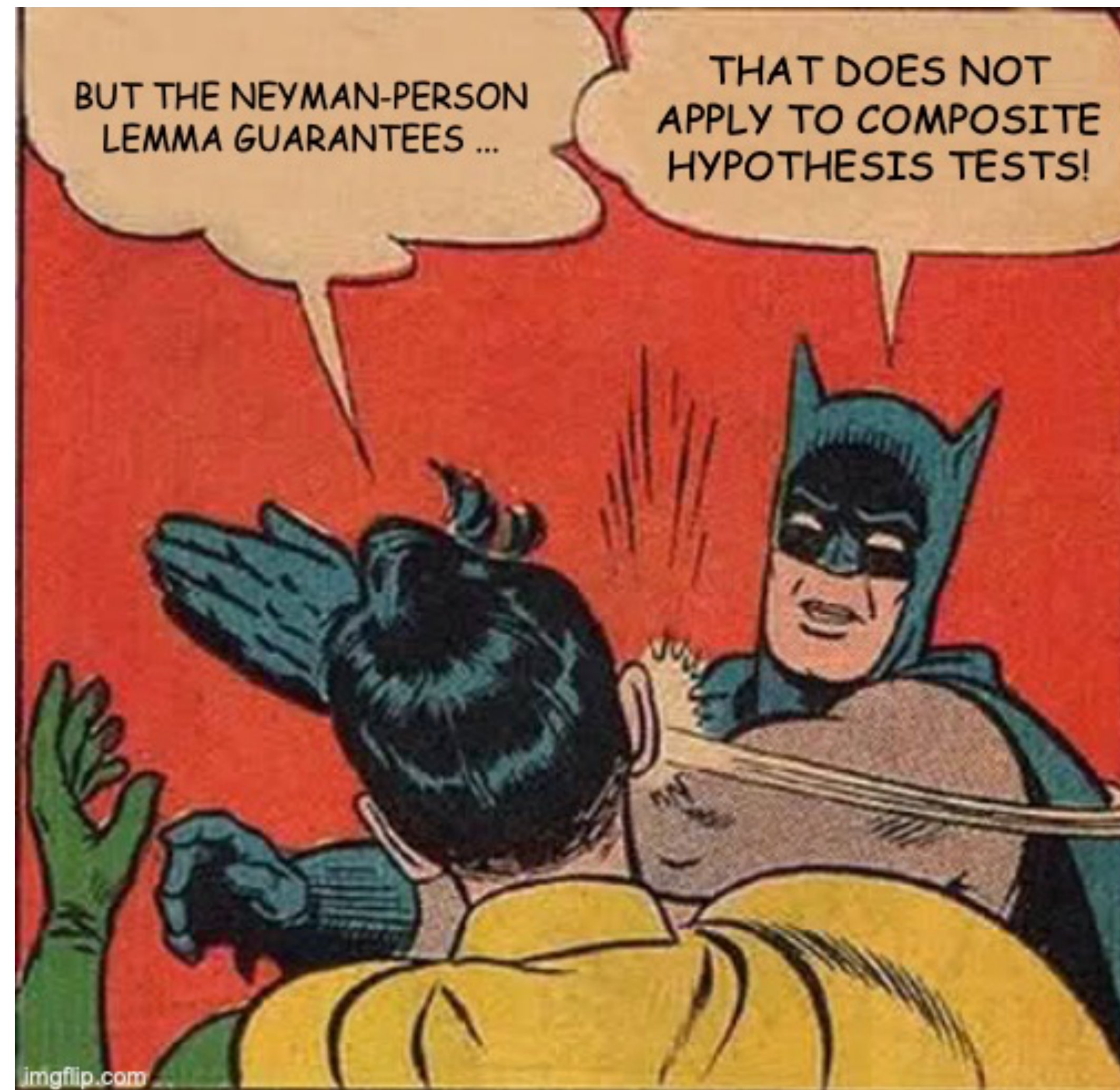
Image: [Wikipedia / PhysRev.43.491](#)

CDF Collaboration: [arXiv:9503002](#)

ATLAS Collaboration: [arXiv:1207.7214](#)

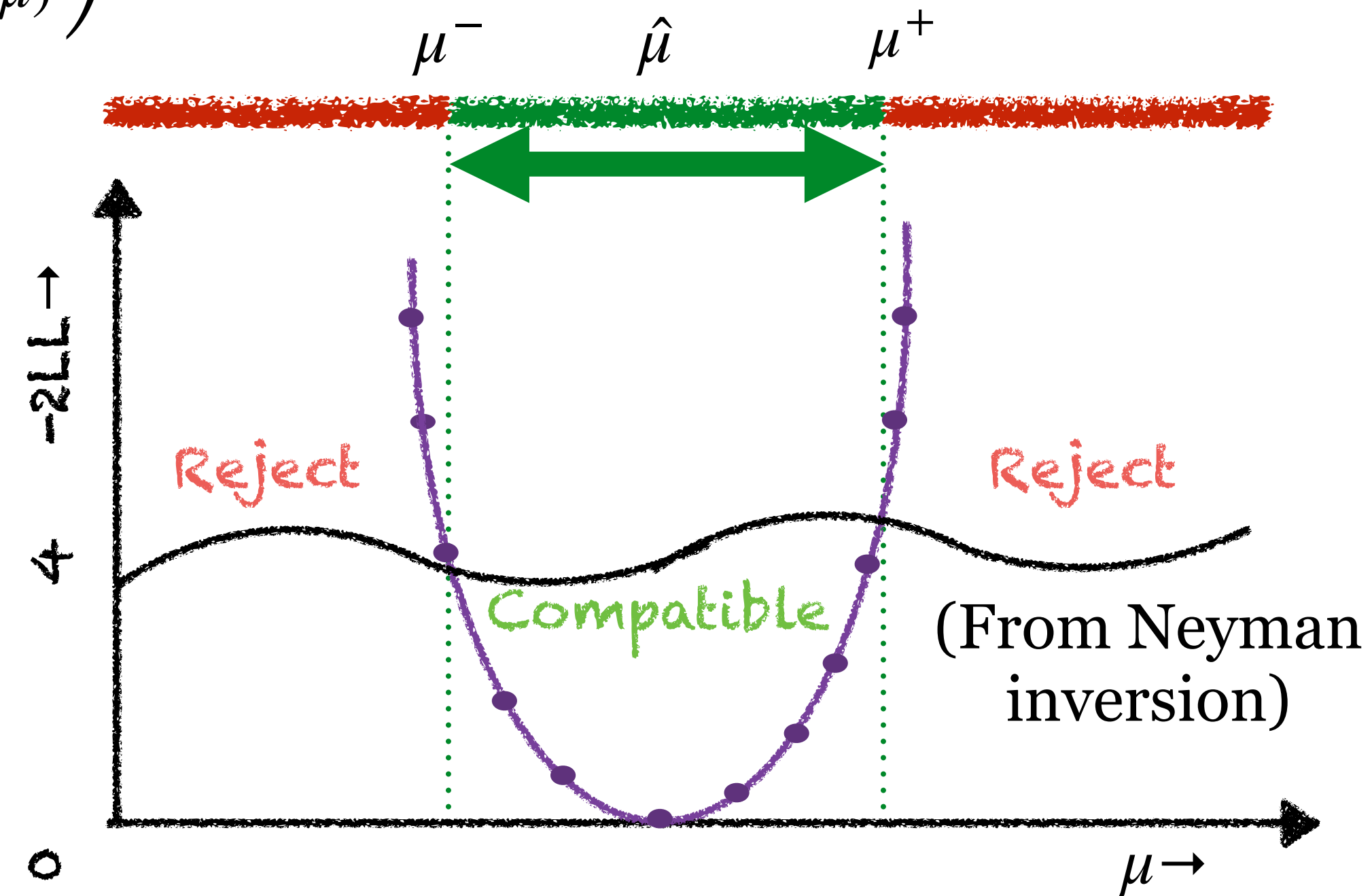
Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?



We care about length of confidence intervals

$$LRS(\mathcal{D}; \mu_0) = -2 \log \left(\frac{p(\mathcal{D} | \mu_0)}{\sup_{\mu \in \Theta} p(\mathcal{D} | \mu)} \right)$$



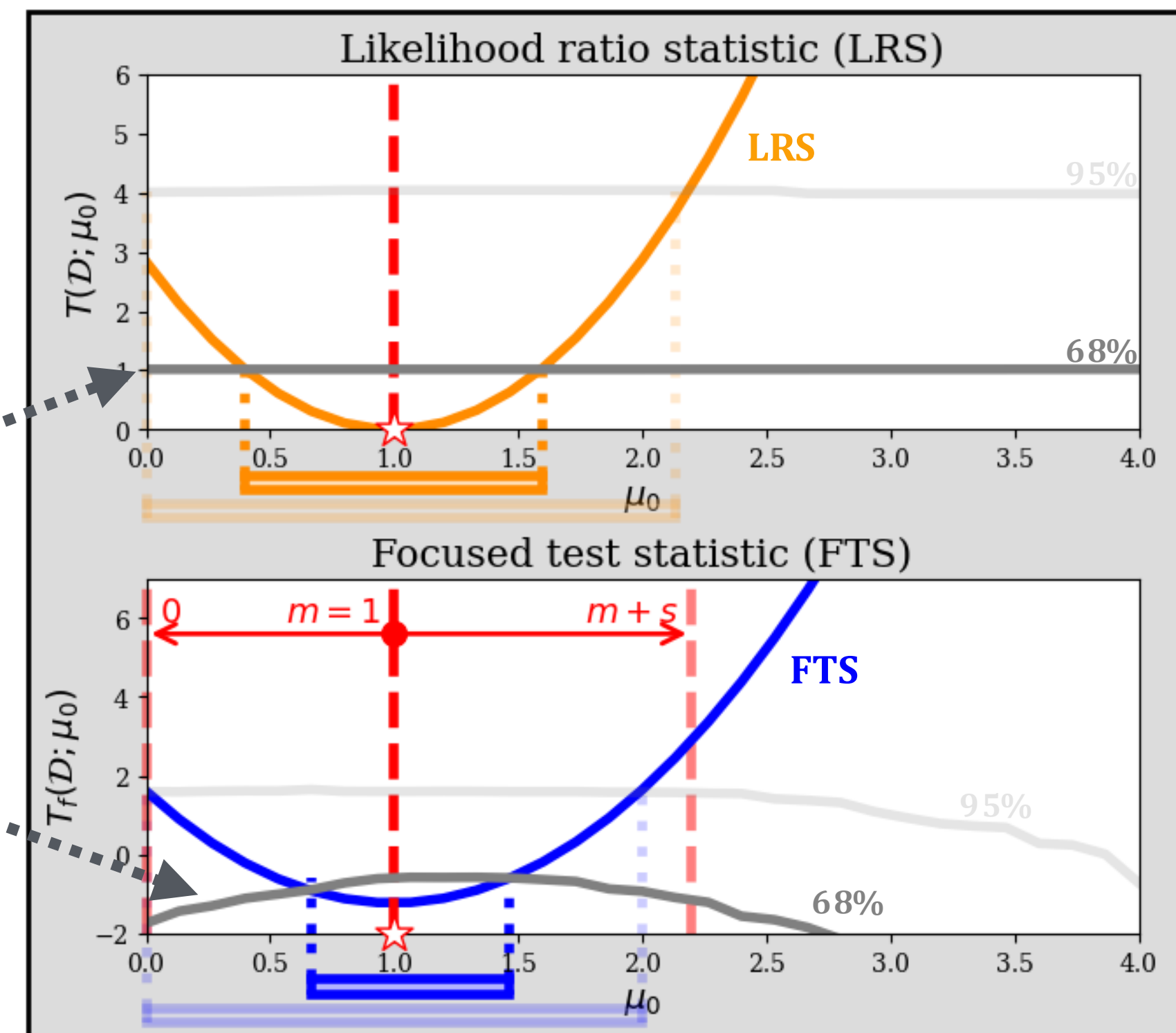
No guarantee that 2σ confidence interval is shortest with likelihood ratio statistic

Challenging a deeply-held belief in particle physics

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

- 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

Ideal for dark matter, neutrino experiments!



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

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

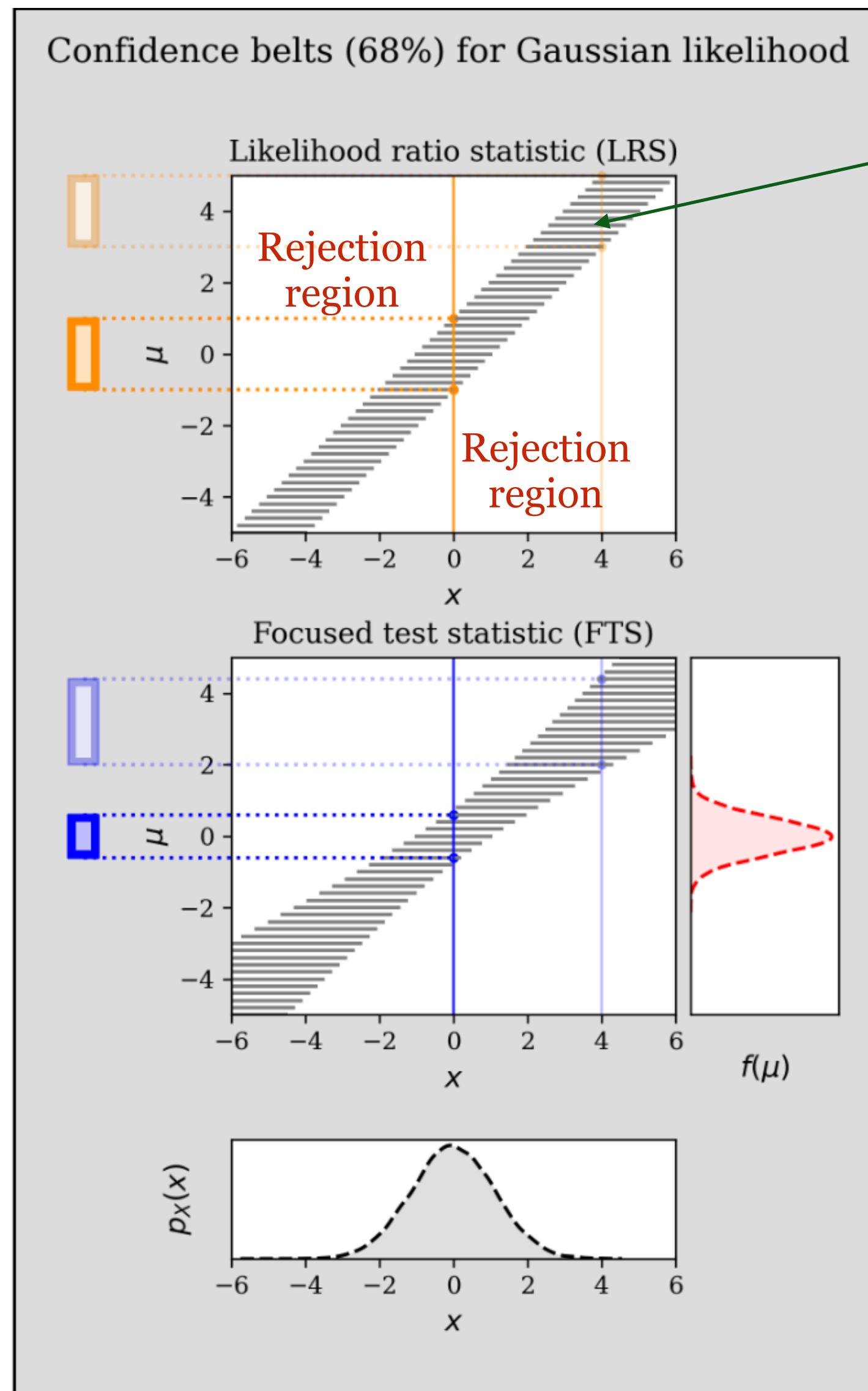
In my group, we test the flammability of plant species using a barbecue. The results can help with fire-mitigation policies and with understanding the evolution of flammability. As part of an outreach activity, we host school-children at the university who haven’t had much exposure to academia before. 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



James Carson

Test inversion for simplest measurement

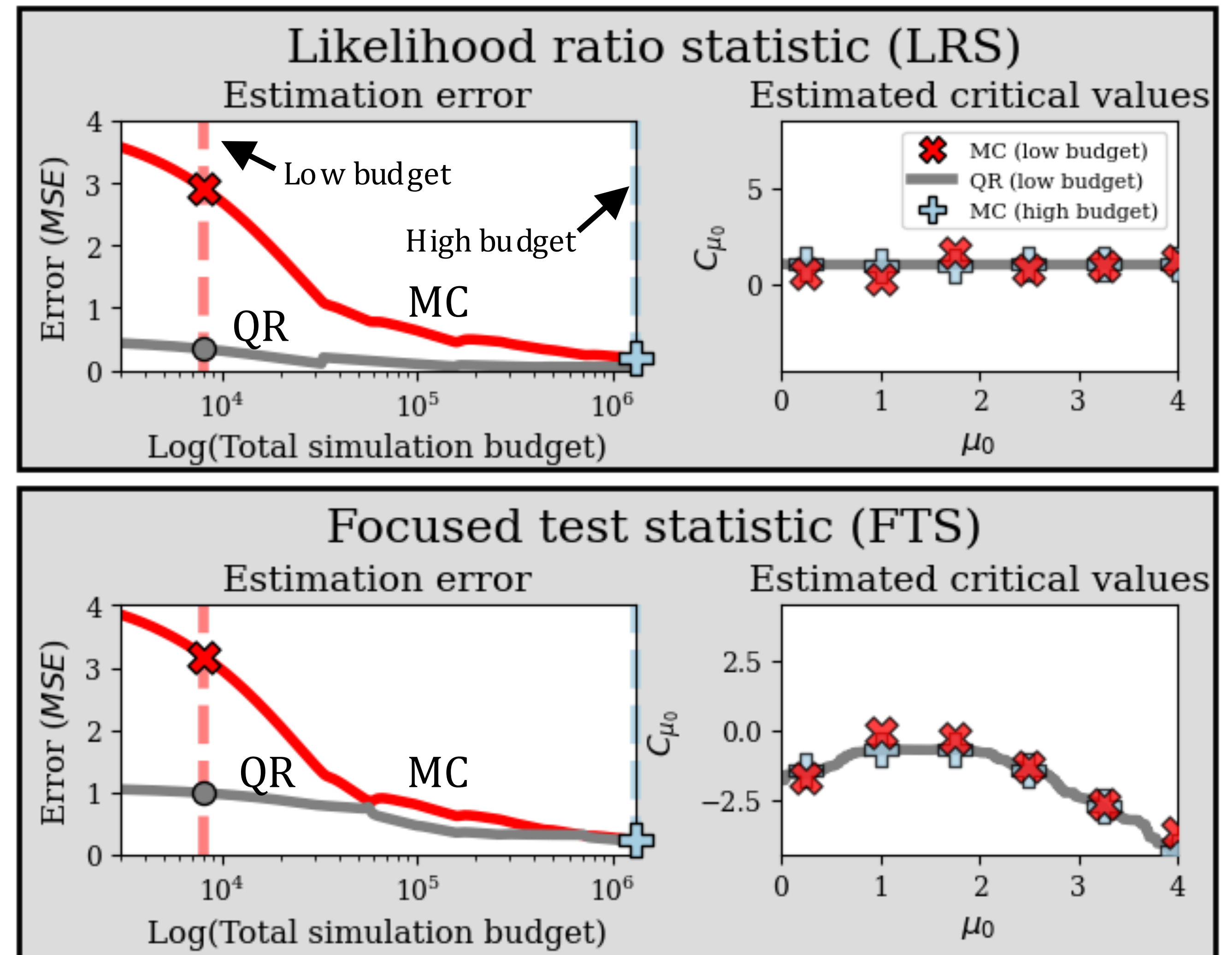


Data comprises single measurement x

- 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

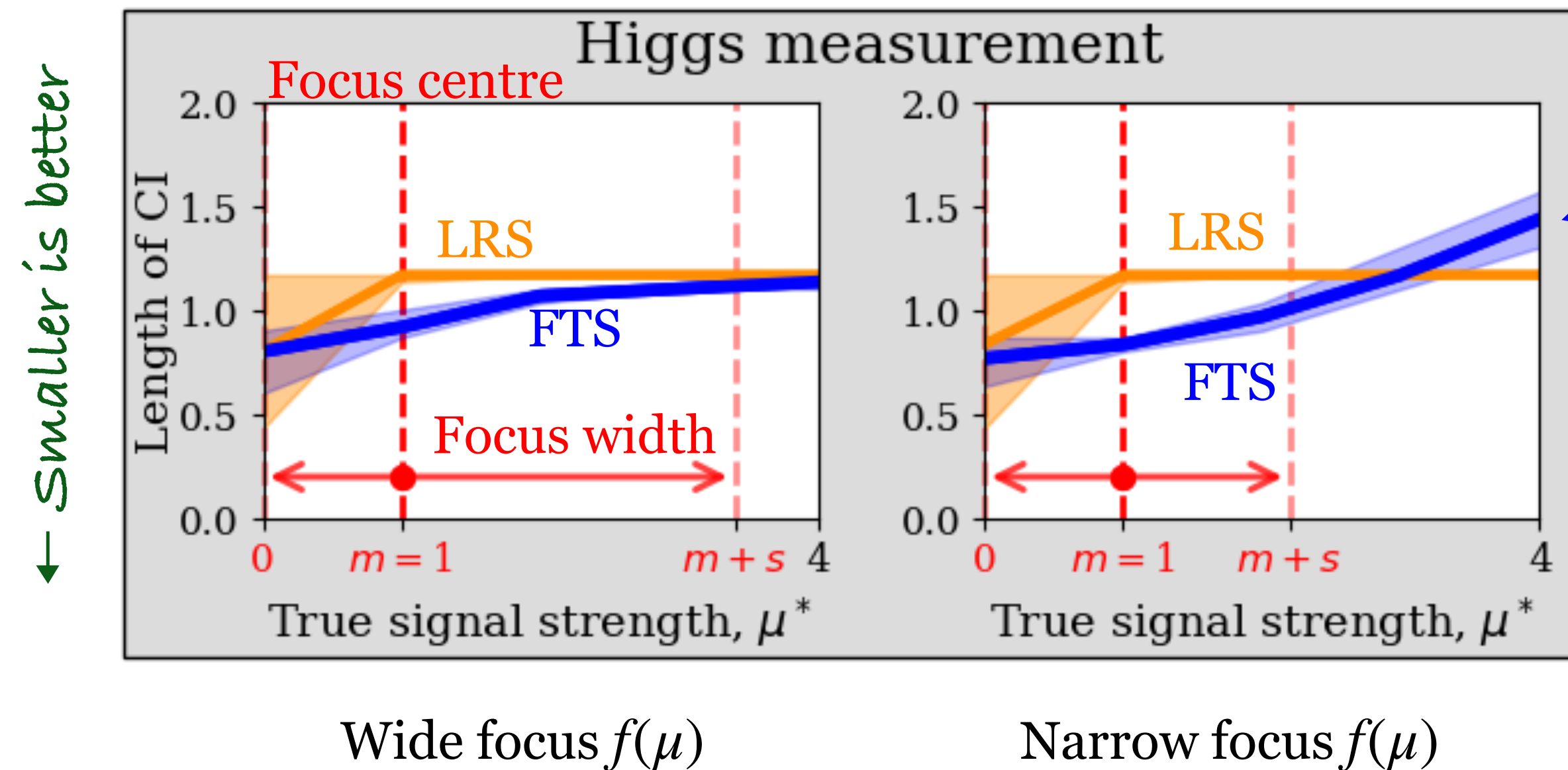
Computational Challenge: Inverting the test, in high dimensions

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



But what if you're wrong about the focus... how is this not a Bayesian analysis?

Compare LRS to FTS under misspecified focus function



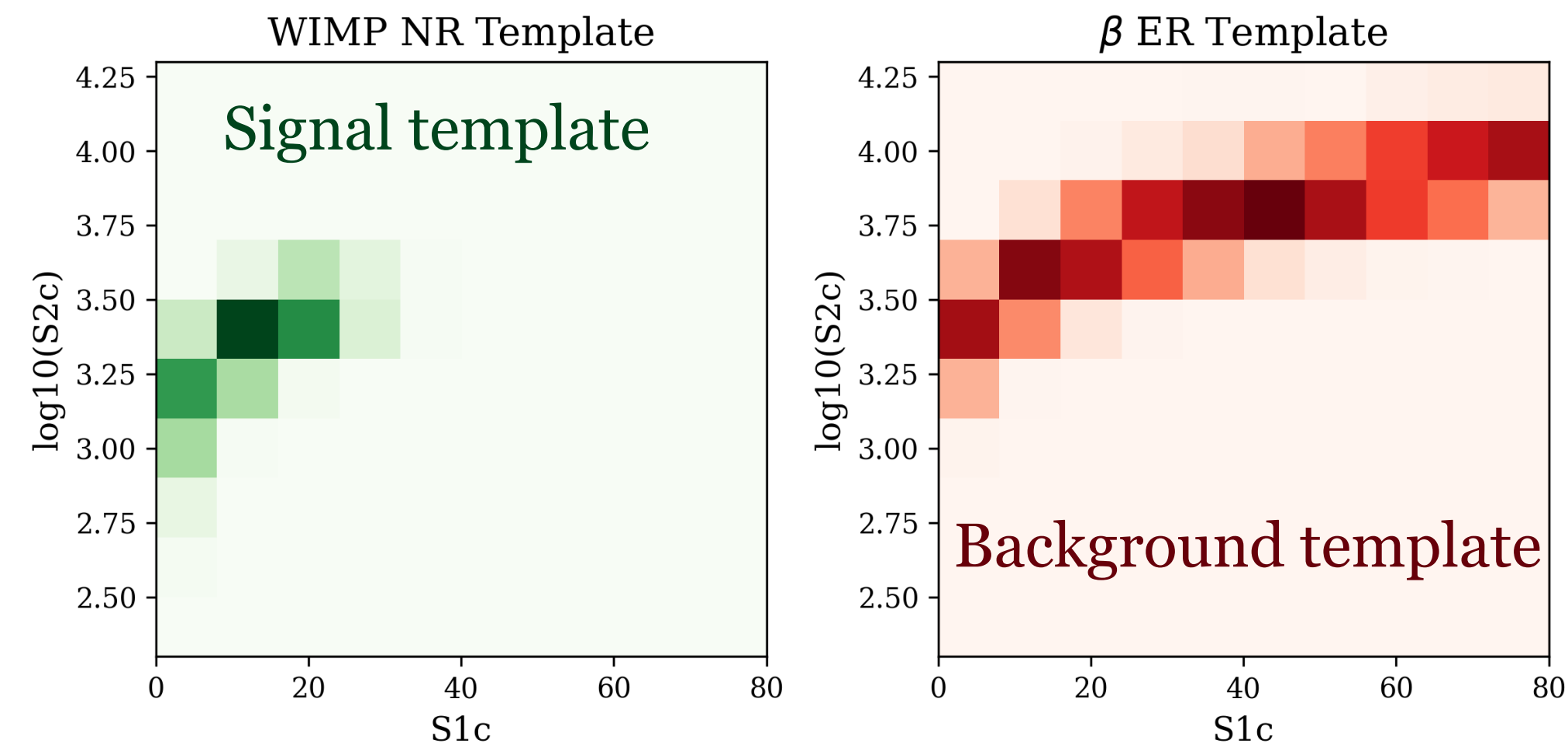
FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \rightarrow \tau\tau$ ATLAS simulated public [benchmark dataset](#)
Focus is around SM point

Which focus would you choose?

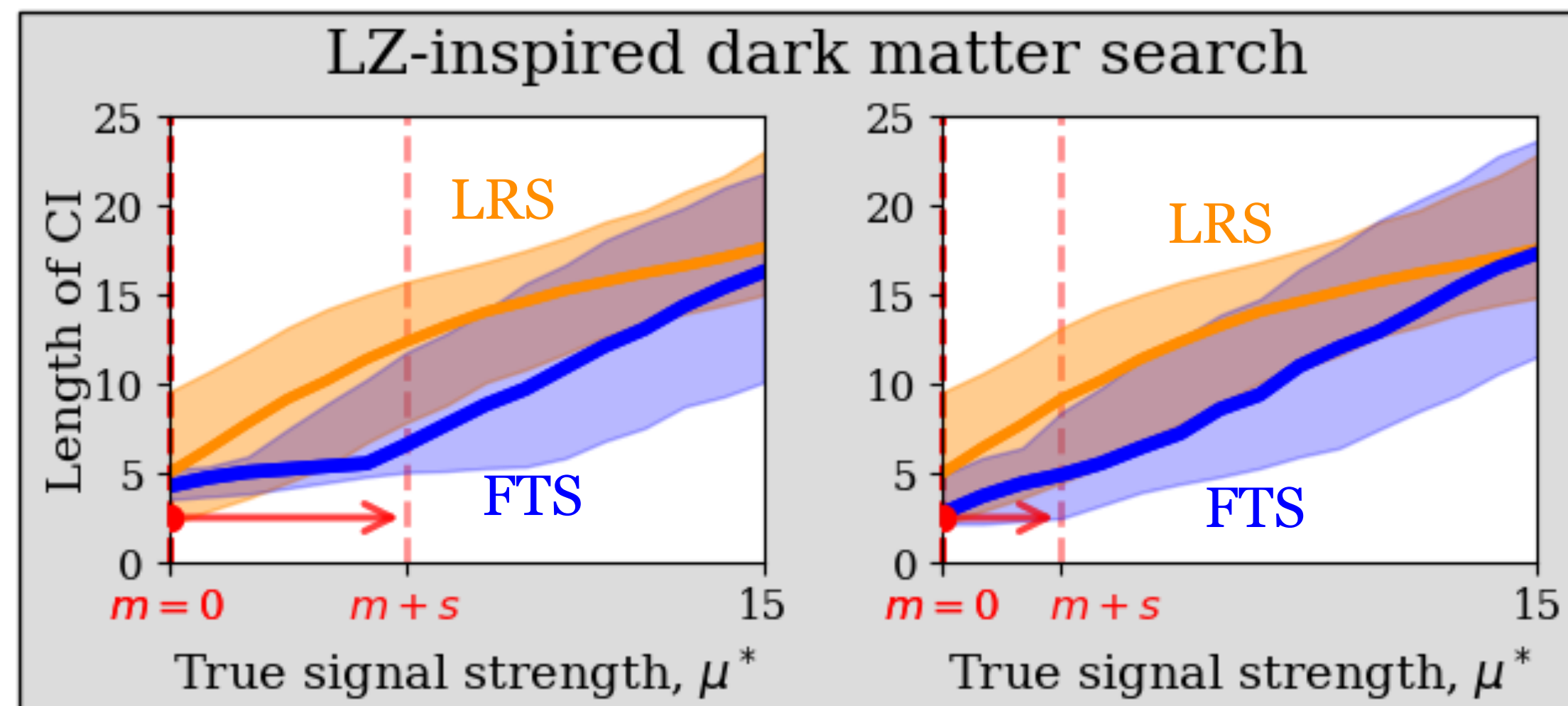
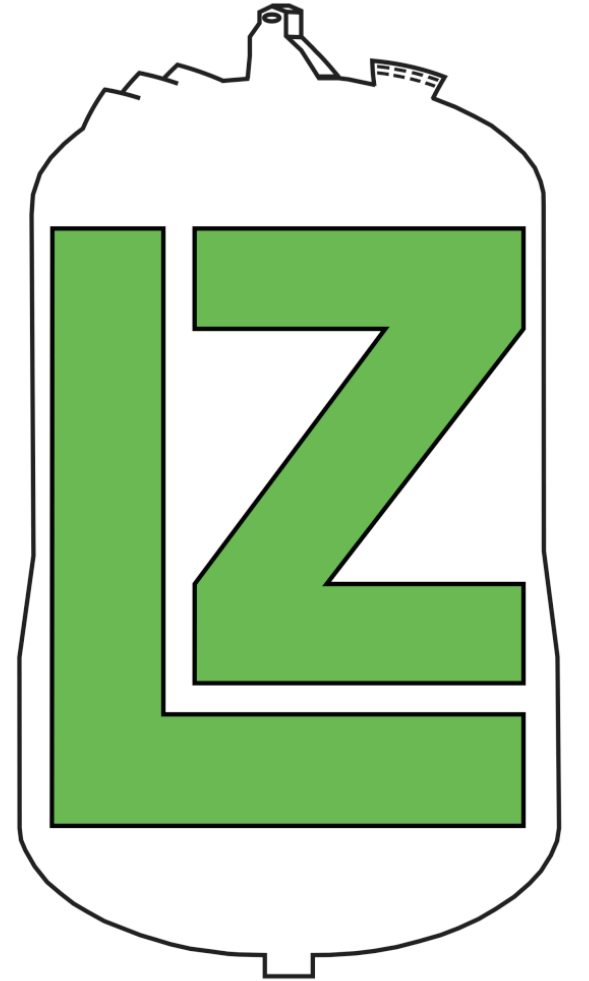
Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

Second case study: Search for dark matter 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

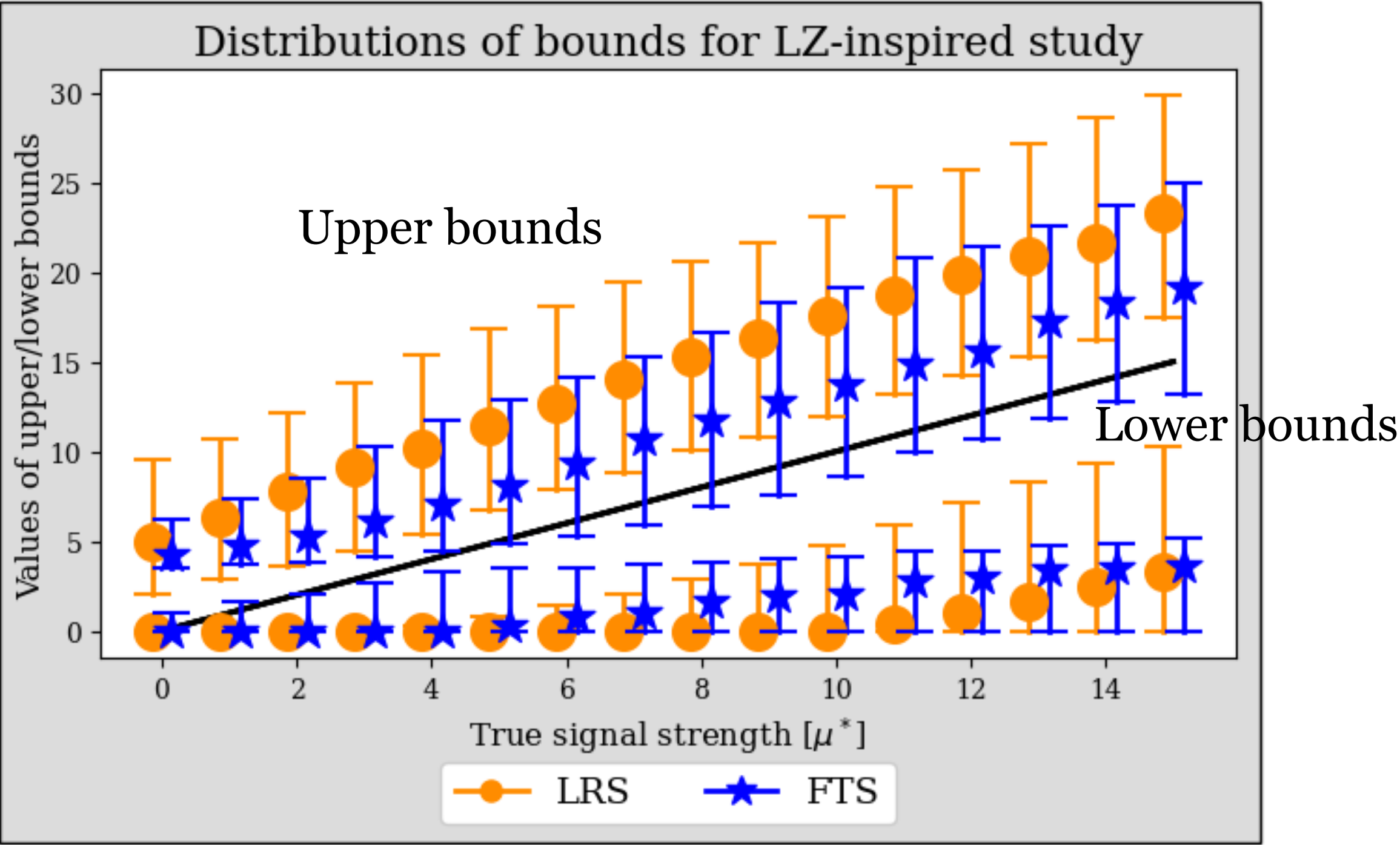
Narrow focus

FTS gains stable over large parameter space for both focus functions

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

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

Consistently larger lower bounds if signal exists, 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)

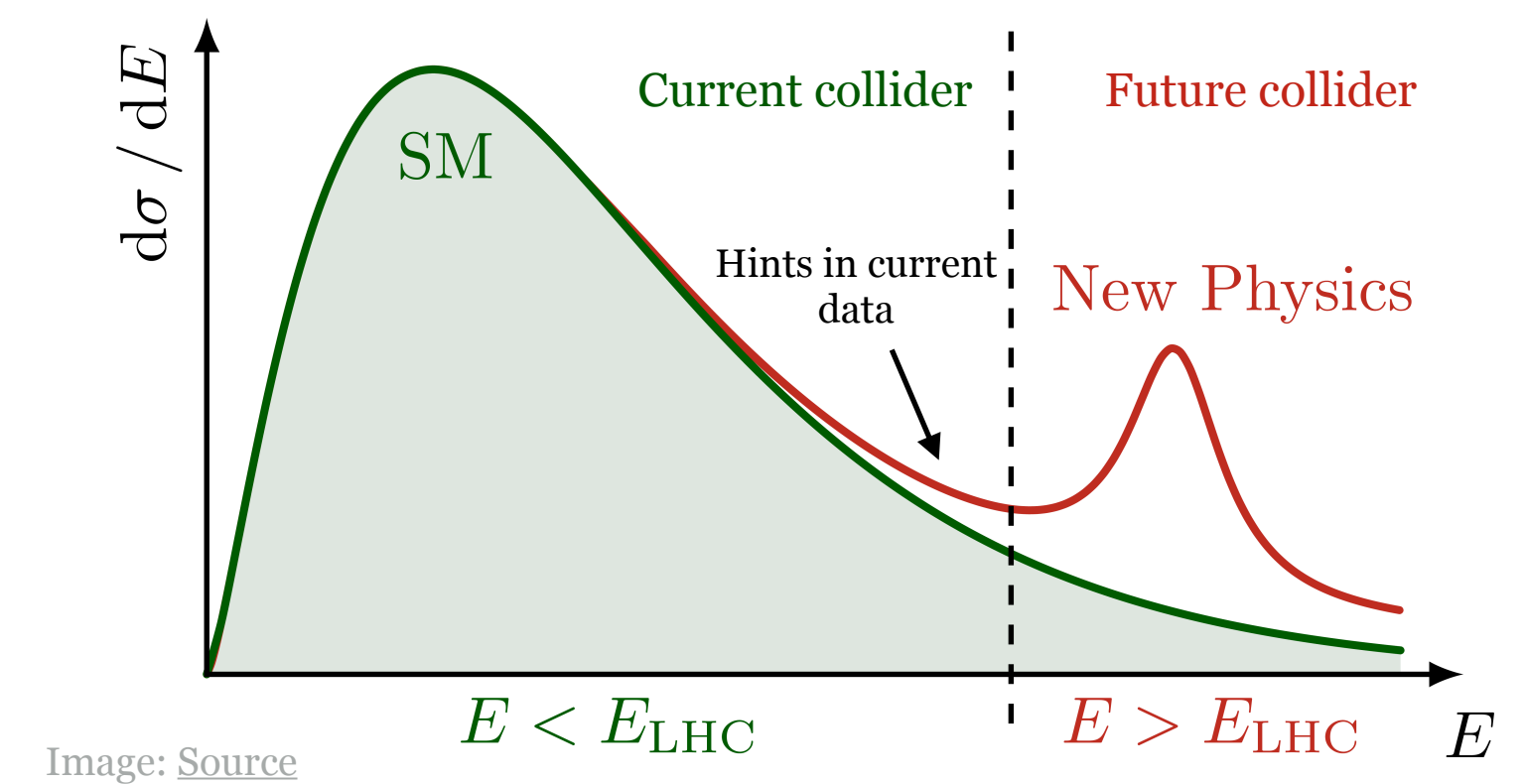
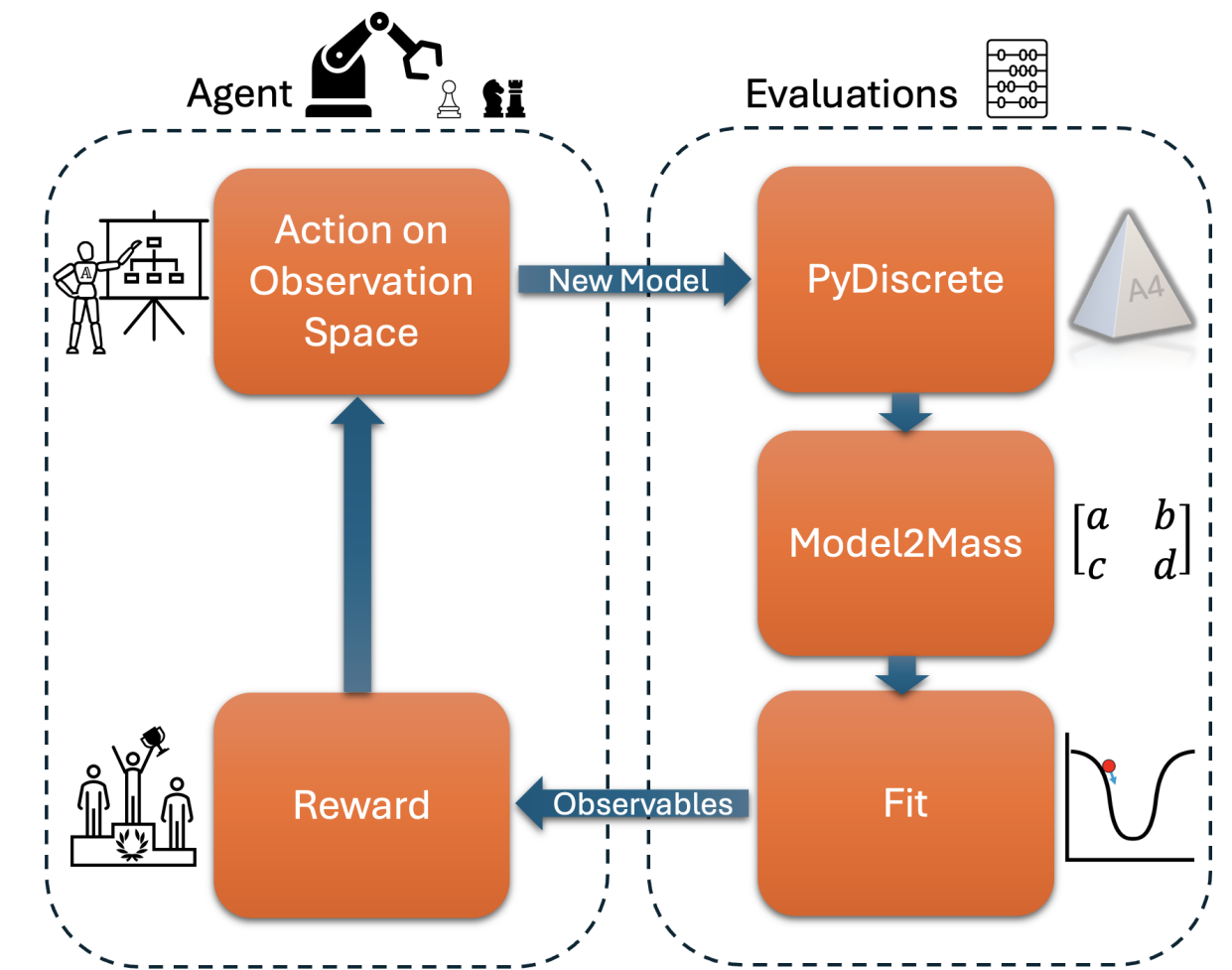
Upper bounds when there is no signal:

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

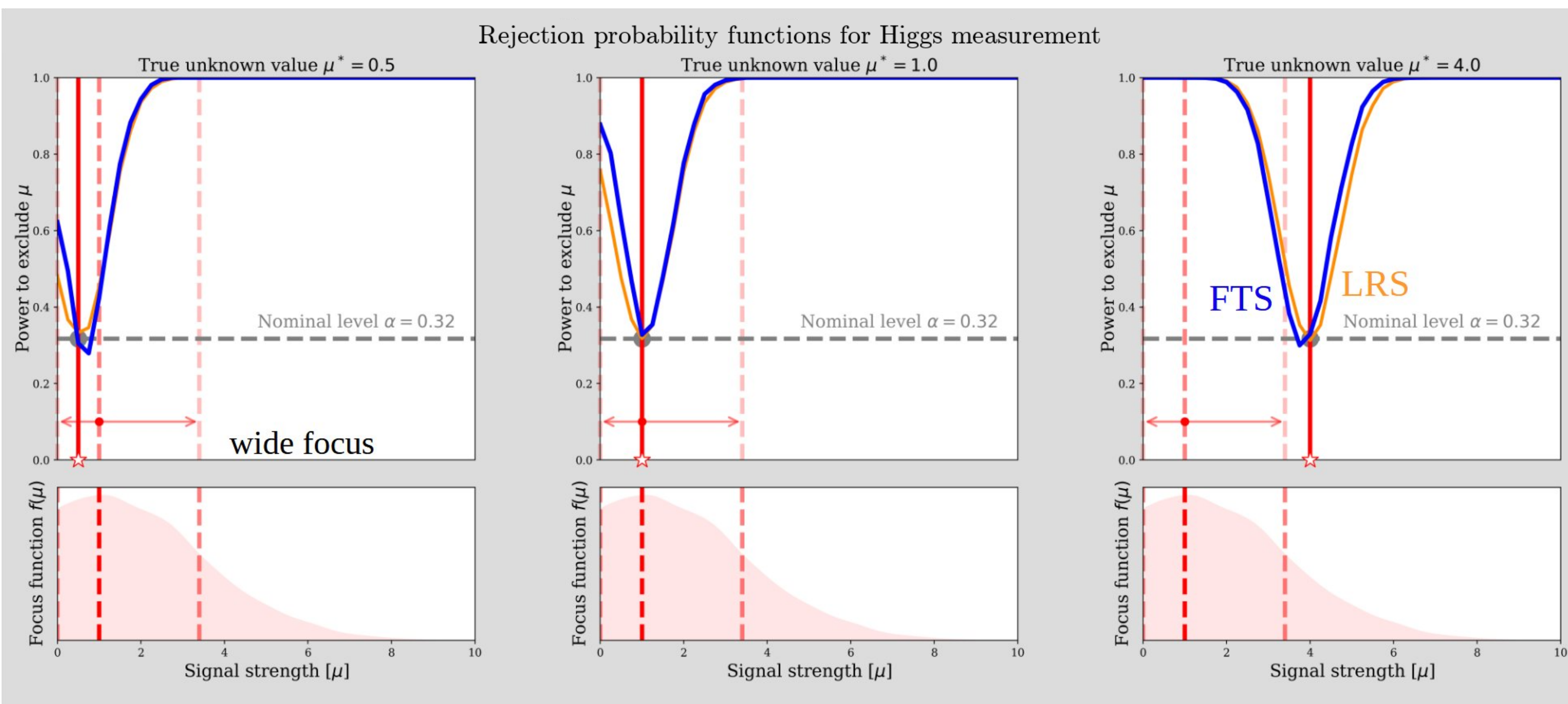
- AI now let's us
 - Access complete information in our **high-dimensional data**
 - Question deeply-held beliefs in statistical inference
 - Explore high-dimensional theoretical spaces
- New statistical methods like FTS particularly useful in low sample size regimes, could help us discover new physics from **tails of distributions!**
 - Applications foreseen across neutrino & dark matter experiments!



Thank you !

FTS backup slides

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

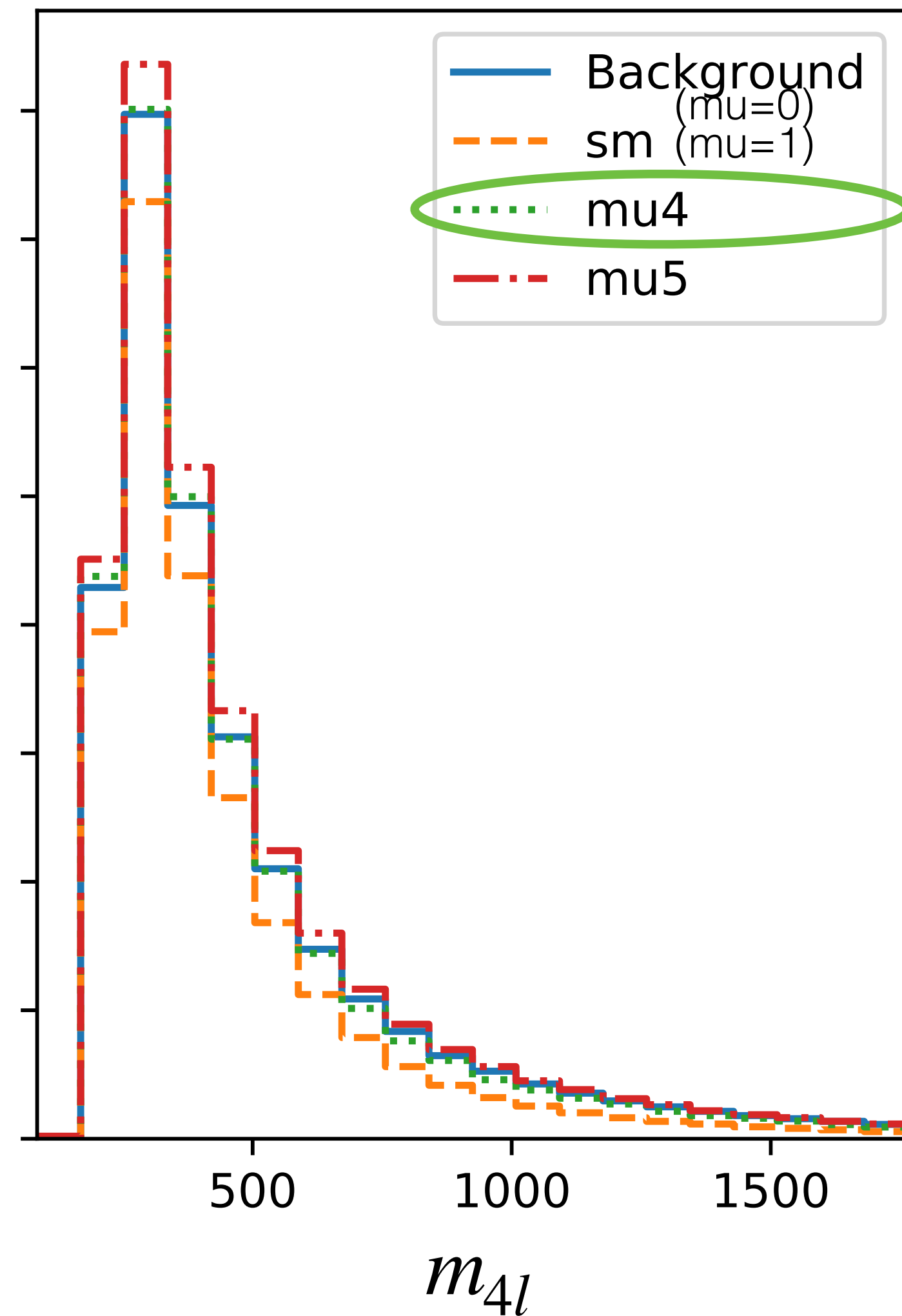
Any history of biased tests in particle physics?:
The CLs method

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

NSBI backup slides

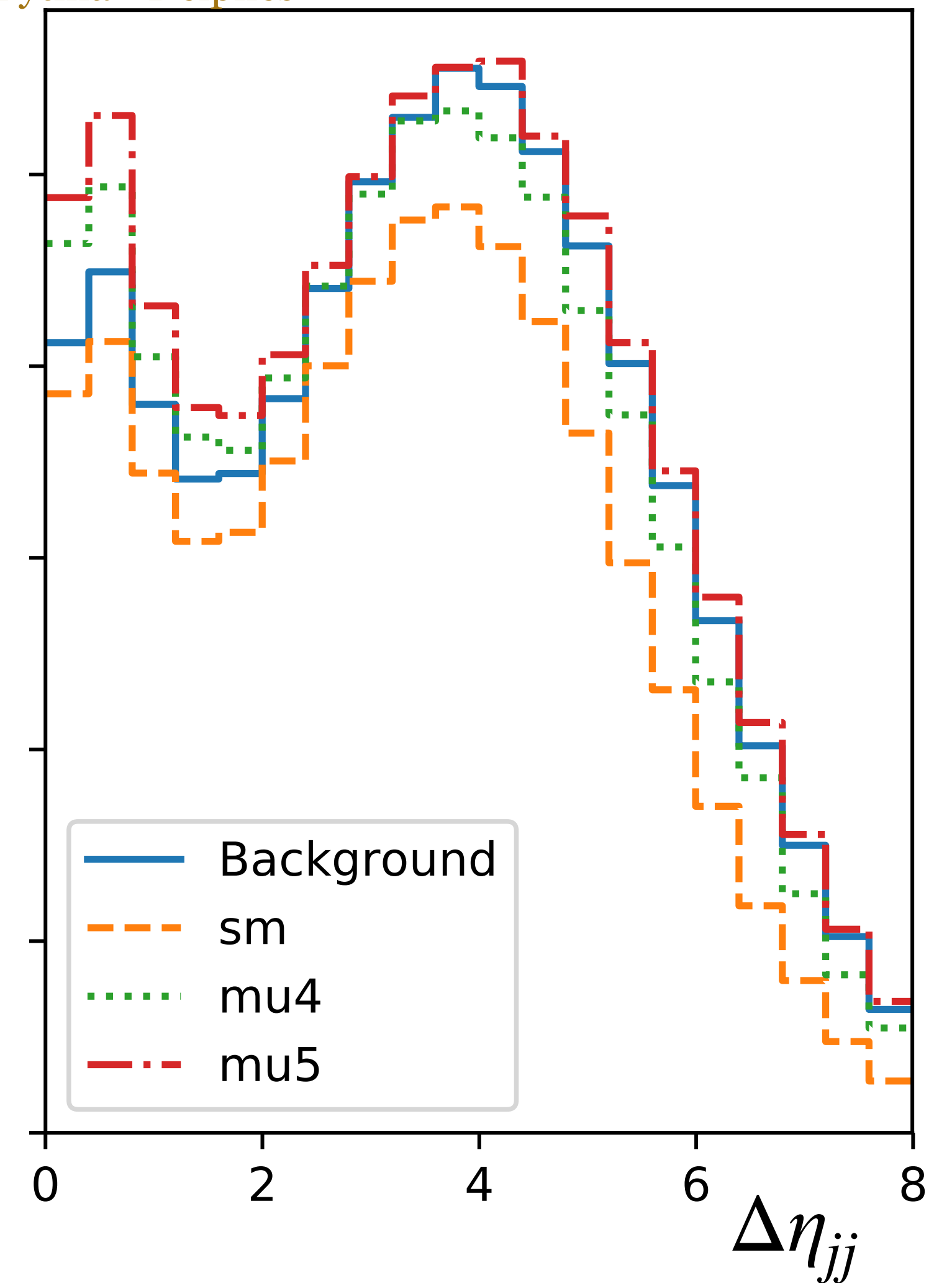
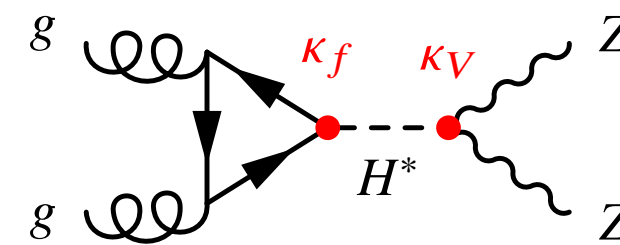
No single observable captures all information in Higgs width study

Signal-background-inference simulations (VBF processes): MG + Pythia+ Delphes



Can you spot the green plot?

$\mu=4$ indistinguishable from $\mu=0$
but other observables can break
the degeneracy



Optimal observable now changes as a function of μ : Cannot collapse problem to 1 dimension

What breaks down?

$$P(X) = |M_s(X) + M_b(X)|^2 = \underbrace{|M_s(X)|^2}_{P_s(X)} + \underbrace{|M_b(X)|^2}_{P_b(X)} + 2 \underbrace{\text{Re}(\overline{M_s(X)} M_b(X))}_{P_i(X)}$$

$$N_{exp} = \mu \cdot S + B + \sqrt{\mu} \cdot I$$

A neural network classifier trained on S vs B, estimates the decision function*: $s(x_i) = \frac{p(x_i|S)}{p(x_i|S) + p(x_i|B)}$

Which contains all the information required for the likelihood ratio:

$$\frac{p(x_i|\mu)}{p(x_i|\mu=0)} = \frac{1}{\mu \cdot \nu_S + \nu_B} \frac{\mu \cdot \nu_S p(x_i|S) + \nu_B p(x_i|B)}{p(x_i|B)} = \frac{\mu}{\mu \cdot \nu_S + \nu_B} \cdot \left(\frac{s(x_i)}{1 - s(x_i)} + \nu_B \right)$$

* Equal class weights

Same observable s is optimal to test all μ hypotheses!

No need to develop separate analysis per hypothesis μ

11

No longer in this convenient special case: The same observable no longer optimal due to non-linear effects coming from quantum interference

Also does not generalise to an arbitrary theory parameter θ , (eg. Effective Field Theory parameters), or systematics

Can we modify the HEP statistical methodology to design near-optimal analyse for the general case?

High-dimensional density estimation is possible with NNs

$$\mathcal{L}(\mu | \mathcal{D}) = p(\mathcal{D} | \mu)$$

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

$$\frac{p(\mathcal{D} | \mu)}{p(\mathcal{D} | ref)}$$

A neural network classifier trained on **simulated samples from μ_1 vs simulated samples from ref** , estimates the decision function:

$$s(x_i) = \frac{p(x_i | \mu_1)}{p(x_i | \mu_1) + p(x_i | ref)}$$

Which contains all the information required for the likelihood ratio:

$$\frac{p(x_i | \mu_1)}{p(x_i | ref)} = \frac{s(x_i)}{1 - s(x_i)}$$

- * Optimal statistic to test each value of μ
- * We get the LR *per event* (unbinned)

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_i^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_k \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Rate term

Prod over events

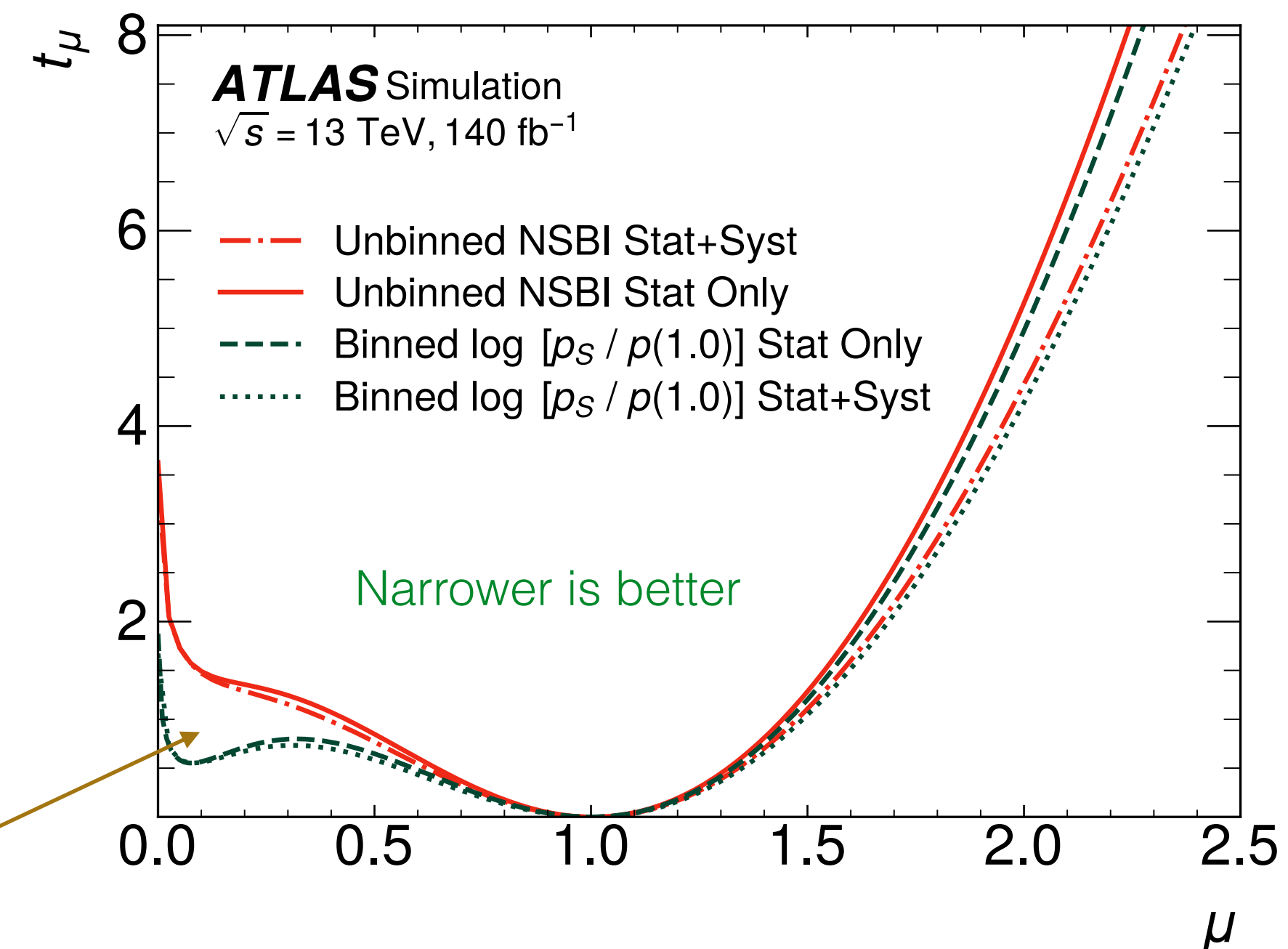
Now includes nuisances α

Constrain term

Profiling:

$$t_\mu = -2 \ln \left(\frac{L_{\text{full}}(\mu, \hat{\hat{\alpha}}) / L_{\text{ref}}}{L_{\text{full}}(\hat{\mu}, \hat{\alpha}) / L_{\text{ref}}} \right)$$

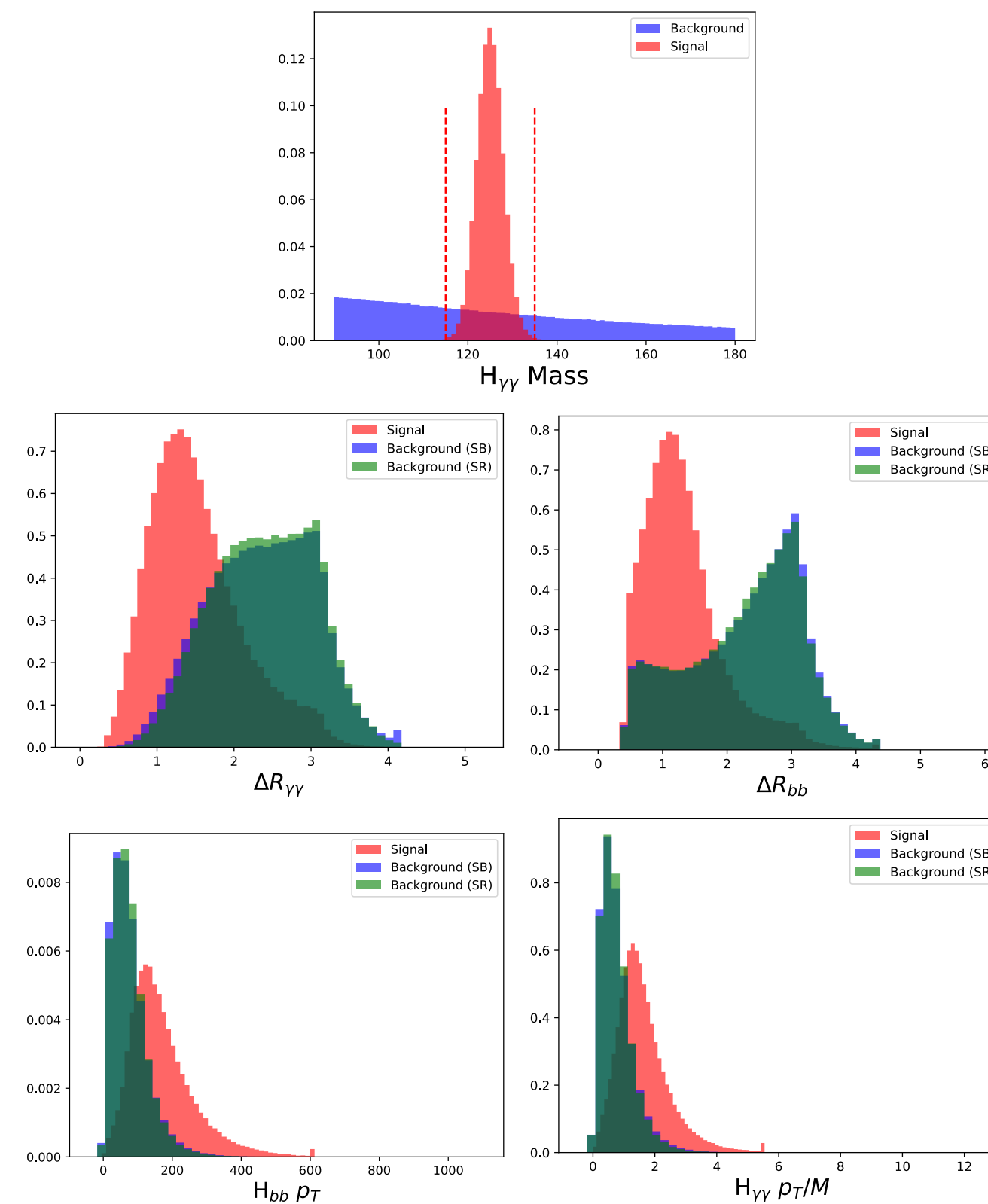
This is why we define p_{ref} to be independent of μ



Non-parabolic shape due to non-linear effects from quantum interference

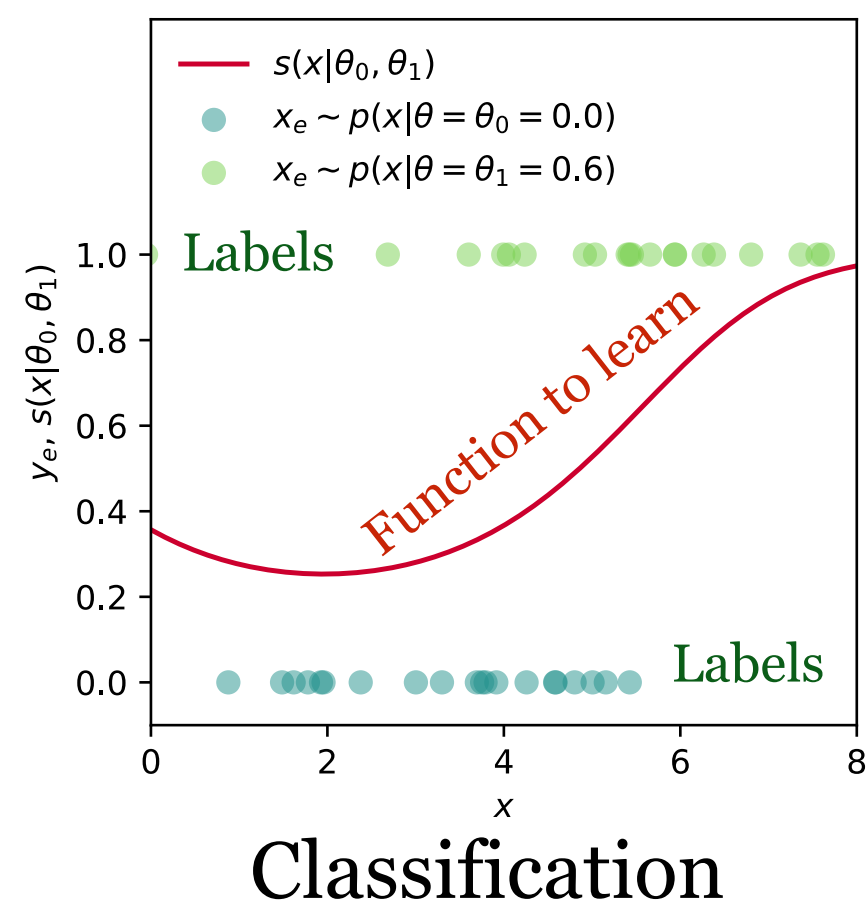
Unreliable simulator?

- Training data can come from control regions in ‘likelihood-free inference’ !
- Port the usual data-driven background estimation techniques eg. in any HH analyses

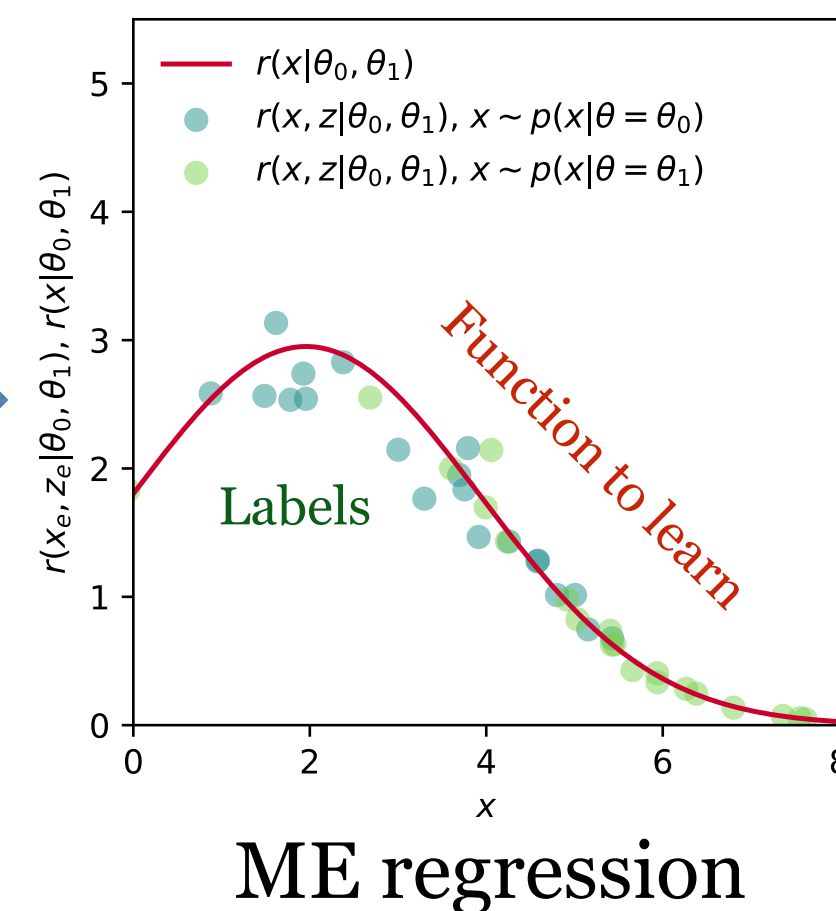


Where would you want to use NSBI?

- If you're annoyed by non-linear effects due to quantum interference
- **Effective field theory:**
 - Highly non-linear with parameters of interest
 - But **challenging to train networks on subtle differences** in phase space!
 - Brehmer et al ([PRD 98.052004](#)) show that you can use matrix-element information to enhance training. But ME typically not available for background processes
 - Tae Park designed a way to combine both ([arXiv:2507.02032](#)), shown for Higgs self-coupling



Change the learning task

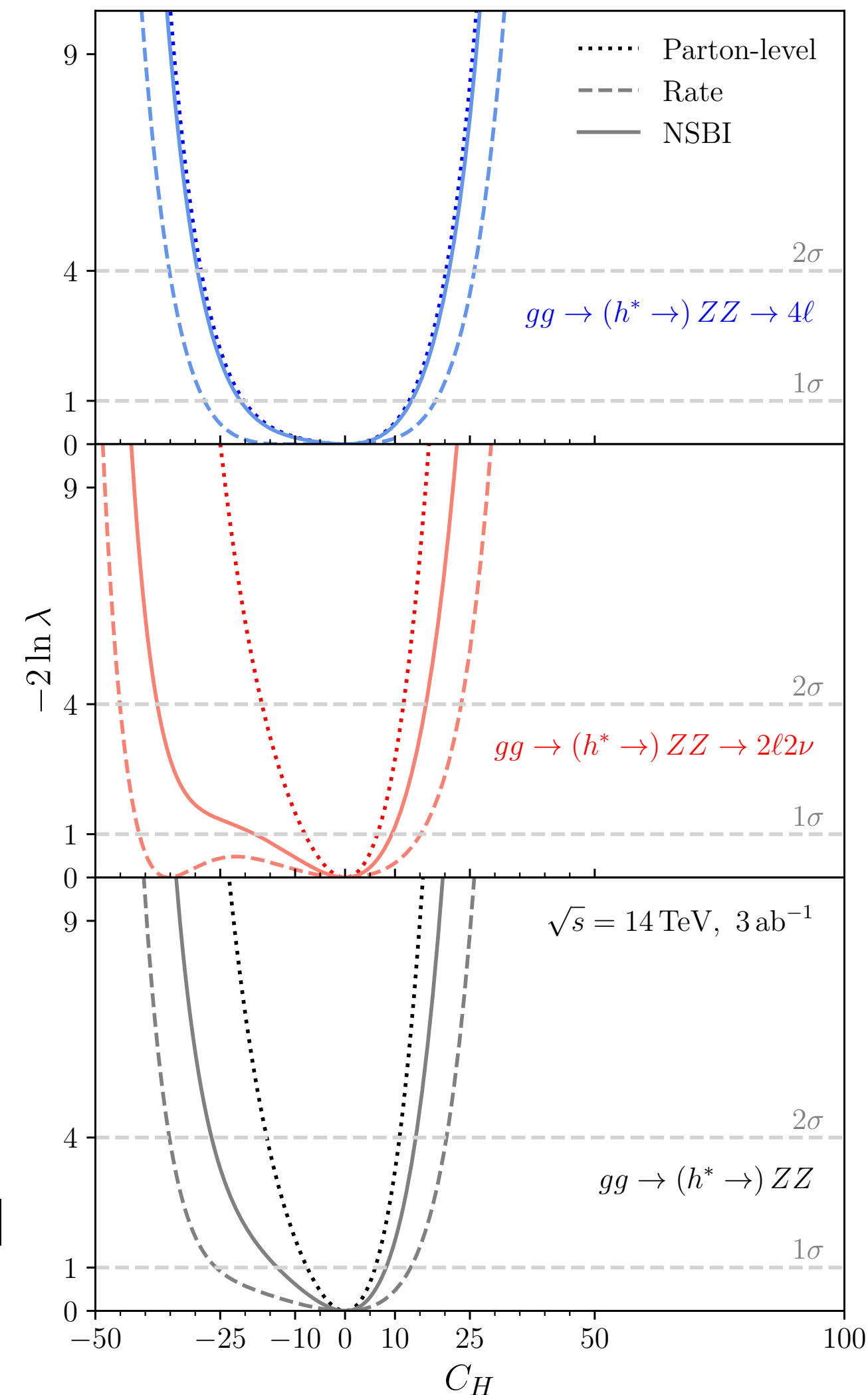
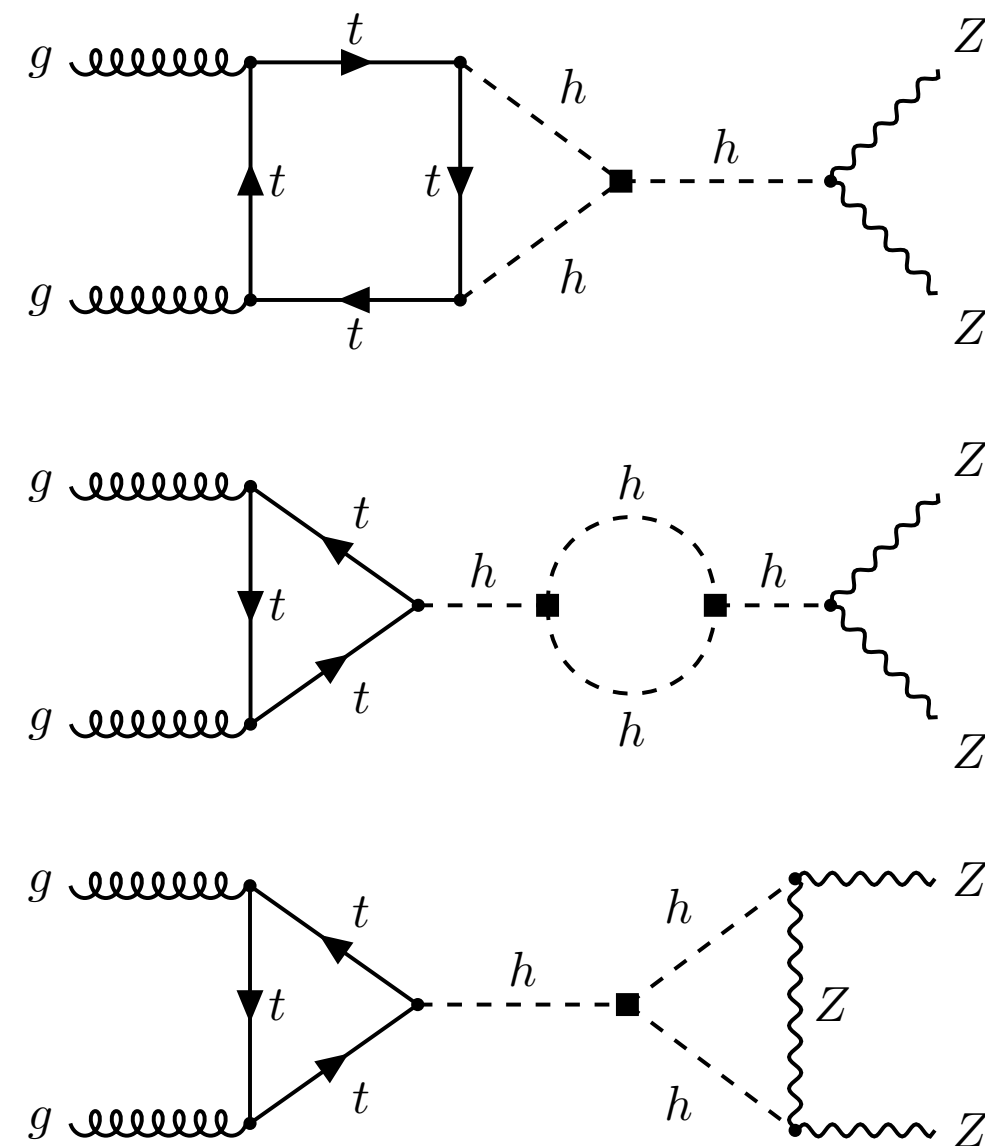


ME approach for
 ggF EFT

Classifier approach
for $q\bar{q}$ bkg

Full LR for pp

NSBI for Higgs self-coupling



Learning subtle differences required matrix-element-enhanced network training + factorisation tricks

$ll\nu\nu$: NSBI stable despite missing neutrinos!

Can be probed in off-shell single Higgs channel

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{v(\mu)} \sum_j^C f_j(\mu) \cdot v_j p_j(x_i) \quad ? \quad \xrightarrow{\text{Estimated using an ensemble of networks}} \frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{v(\mu)} \sum_j^C f_j(\mu) \cdot v_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Event rates estimated from simulations

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

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{v_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu}) v_S p_S(x) + \sqrt{\mu} v_{\text{SBI}_1} p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) v_B p_B(x) \right]$$

$$\xrightarrow{\quad} \frac{p(x|\mu)}{p_S(x)} = \frac{1}{v(\mu)} \left[(\mu - \sqrt{\mu}) v_S + \sqrt{\mu} v_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) v_B \frac{p_B(x)}{p_S(x)} \right]$$

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

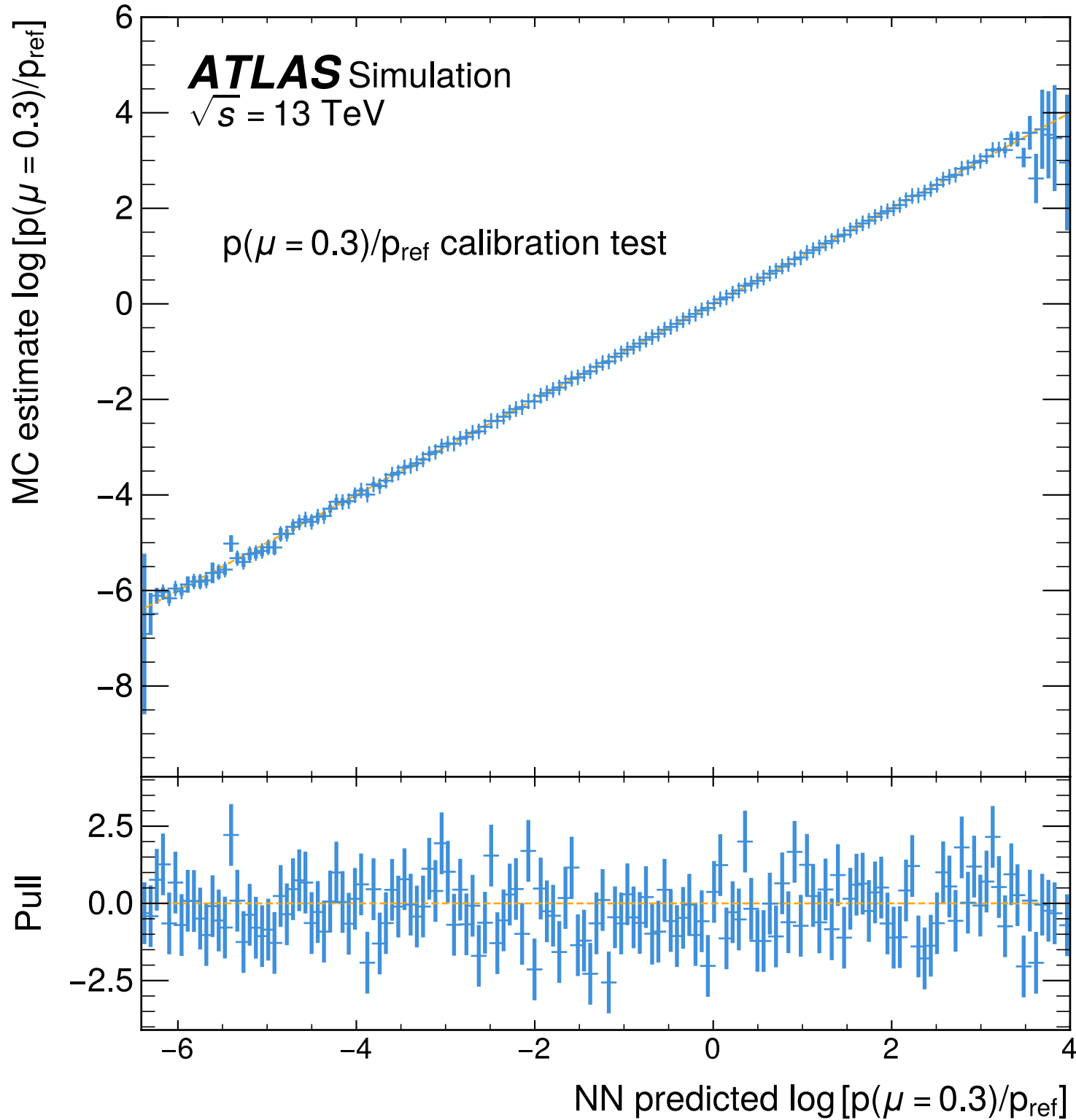
More diagnostics

Calibration curves of probability density ratios

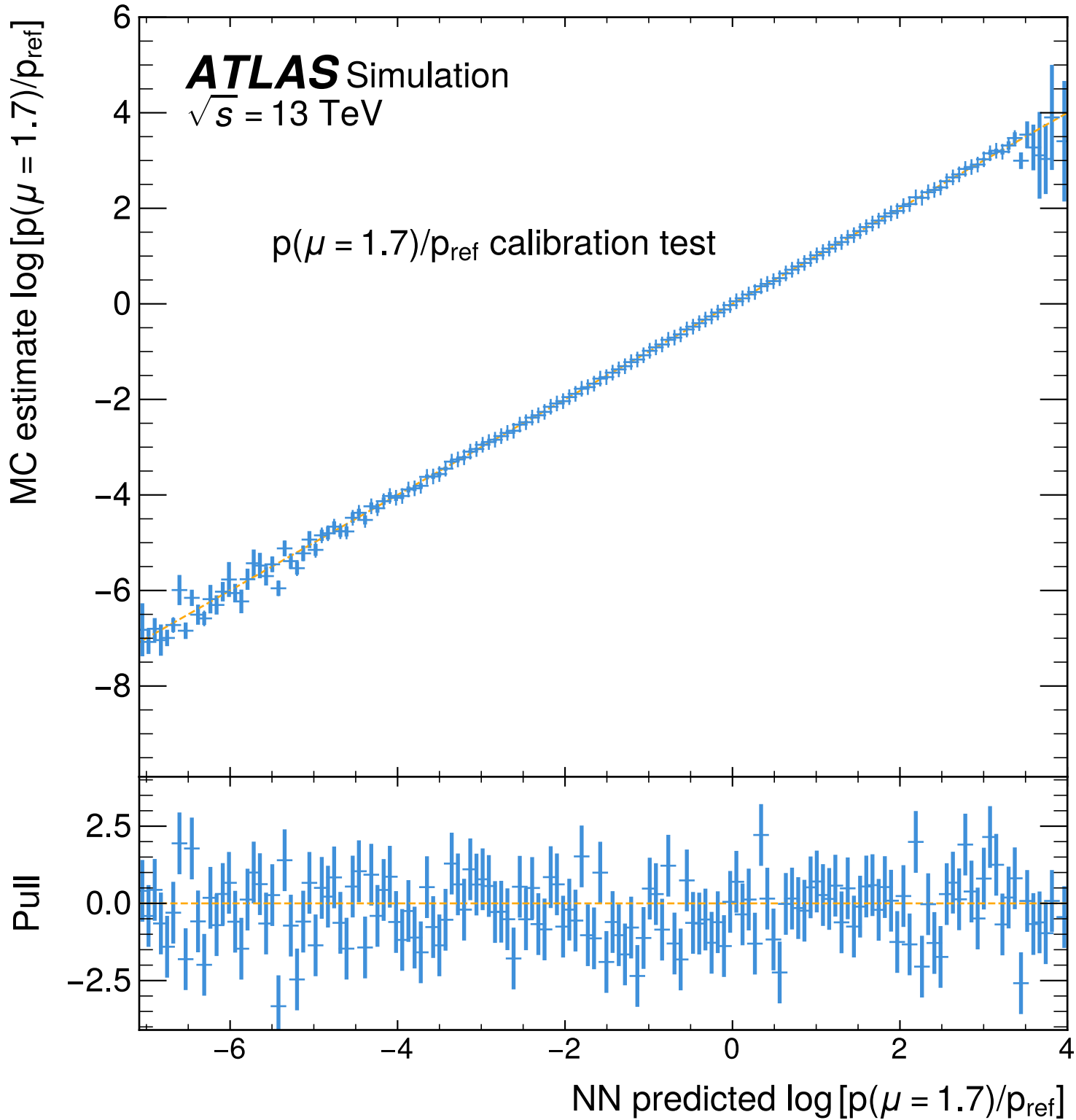
$$\frac{p_{\mu=0.3}(x_i)}{p_{ref}(x_i)}$$

$$\frac{p_{\mu=1.7}(x_i)}{p_{ref}(x_i)}$$

Binned estimate



Ensemble prediction



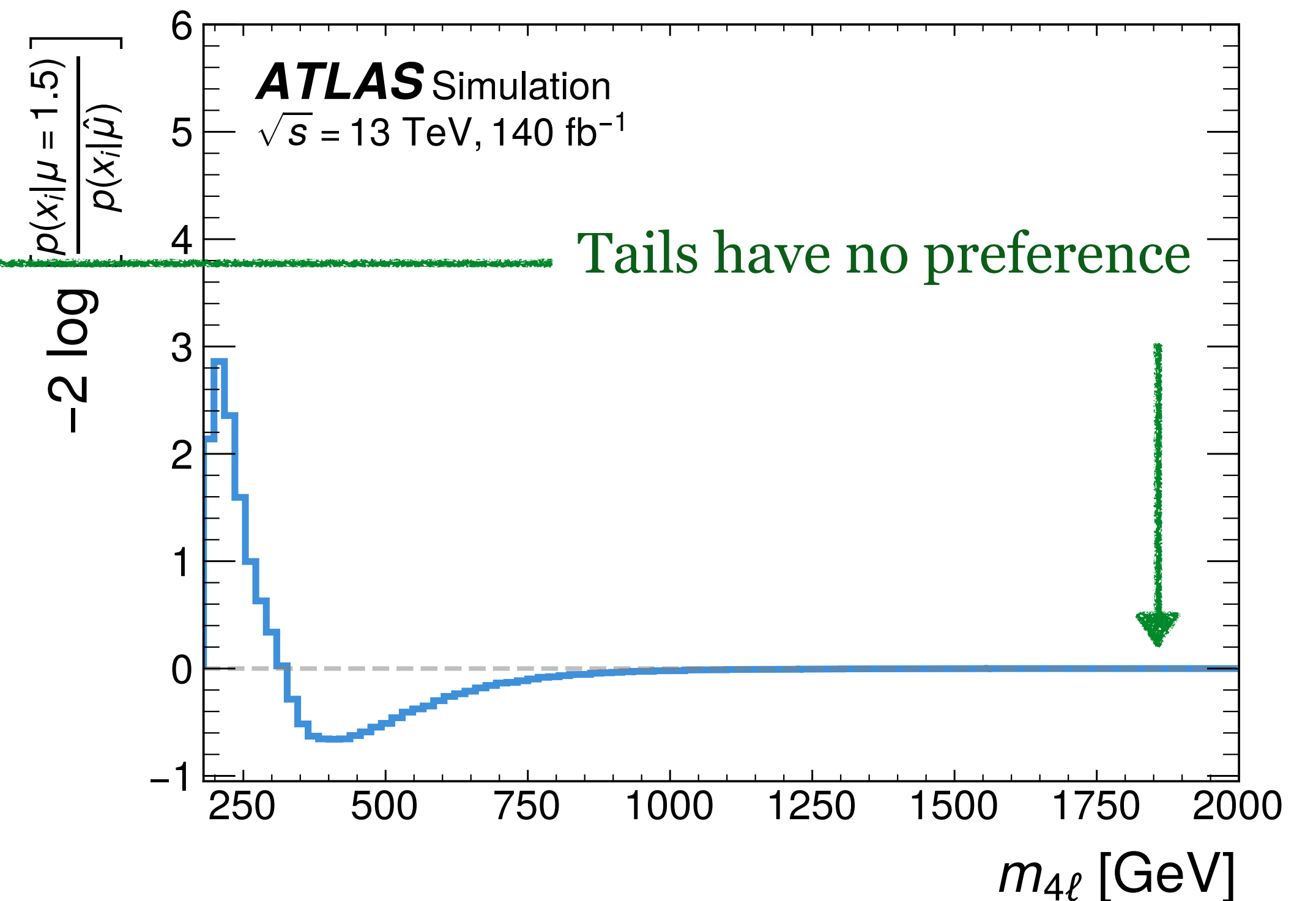
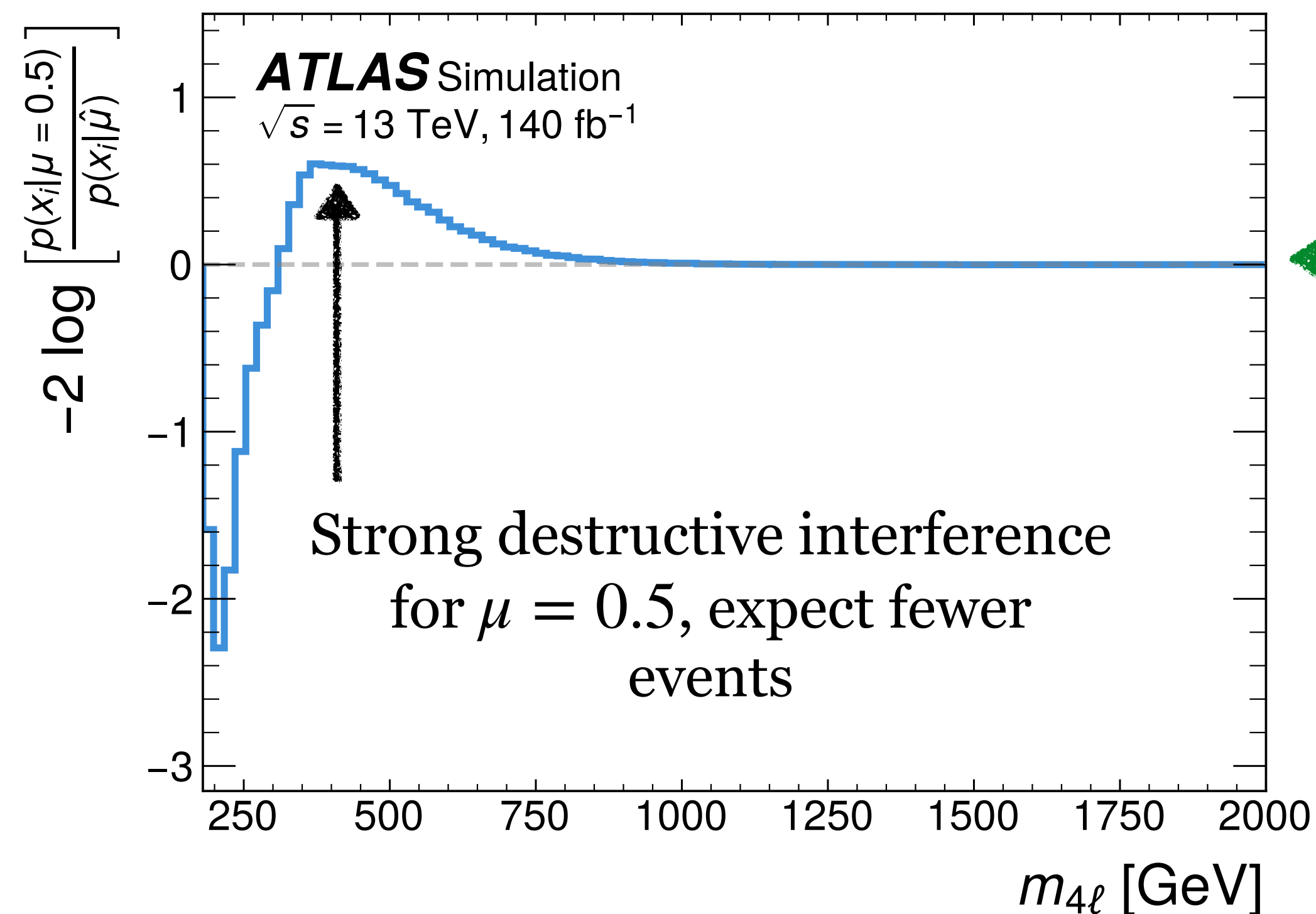
Ensemble prediction

Perfect calibration would give $y = x$

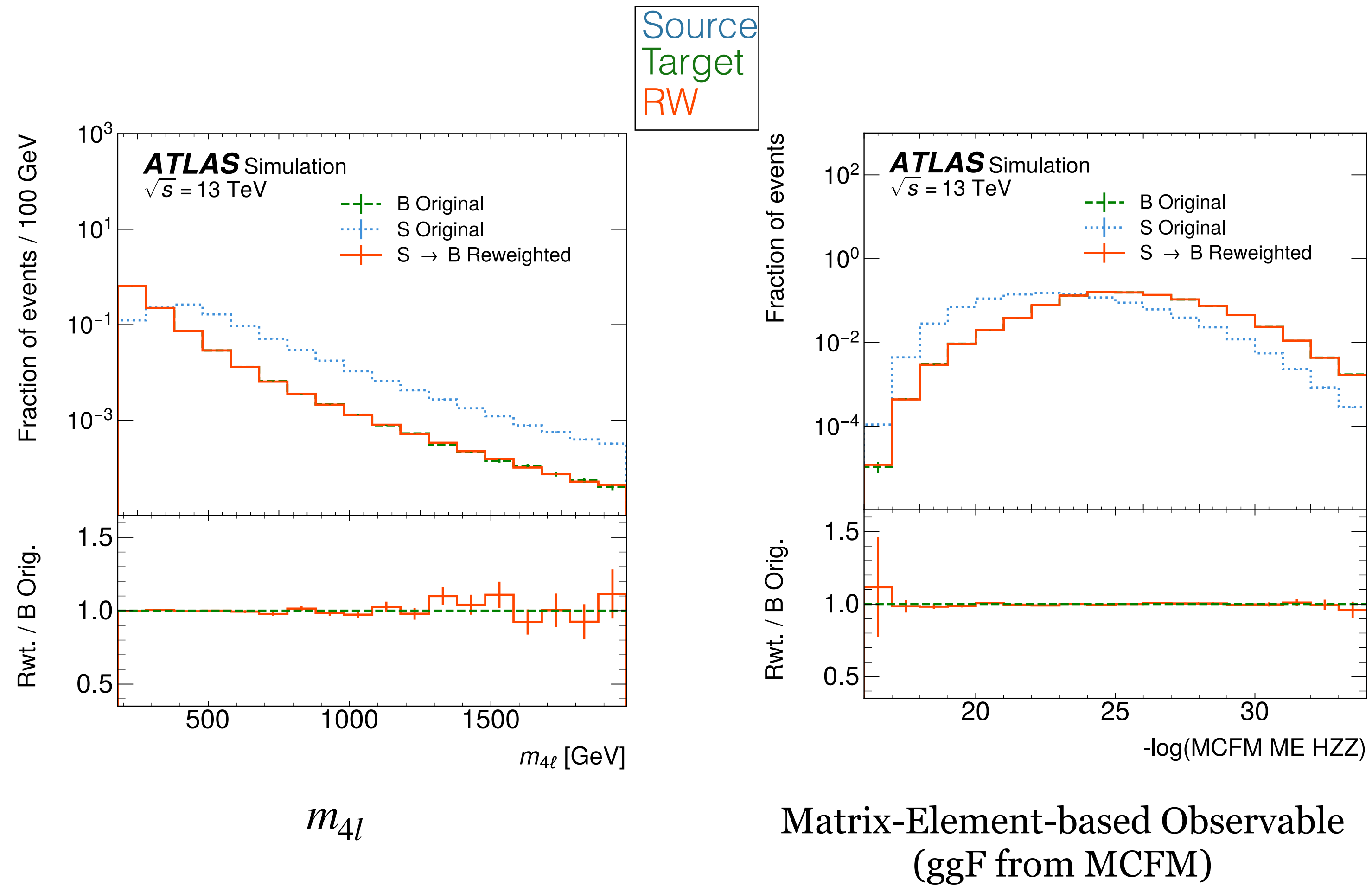
Interpretability: Which phase space favours one hypothesis over another?

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

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

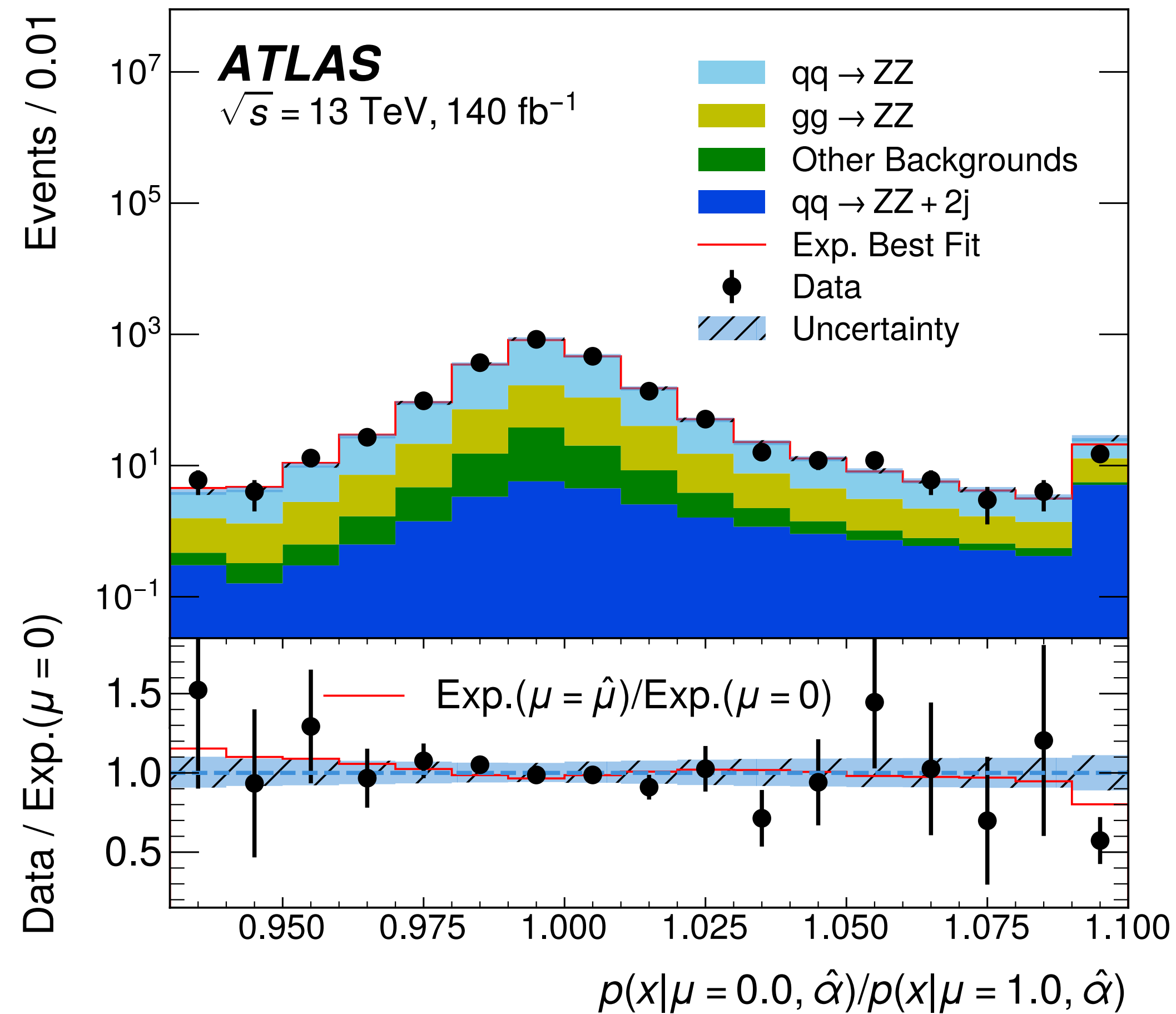
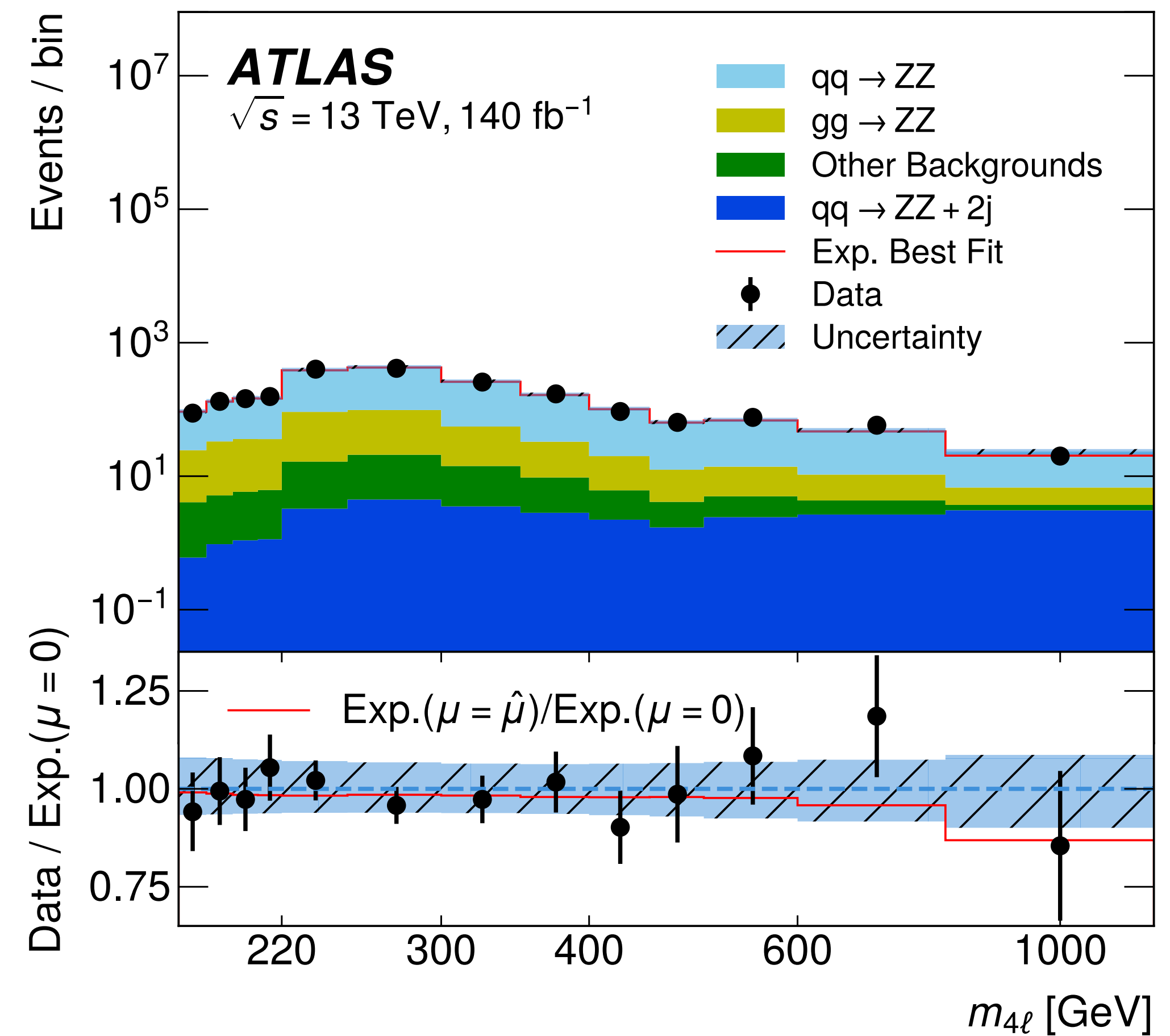


Re-weight closures for B



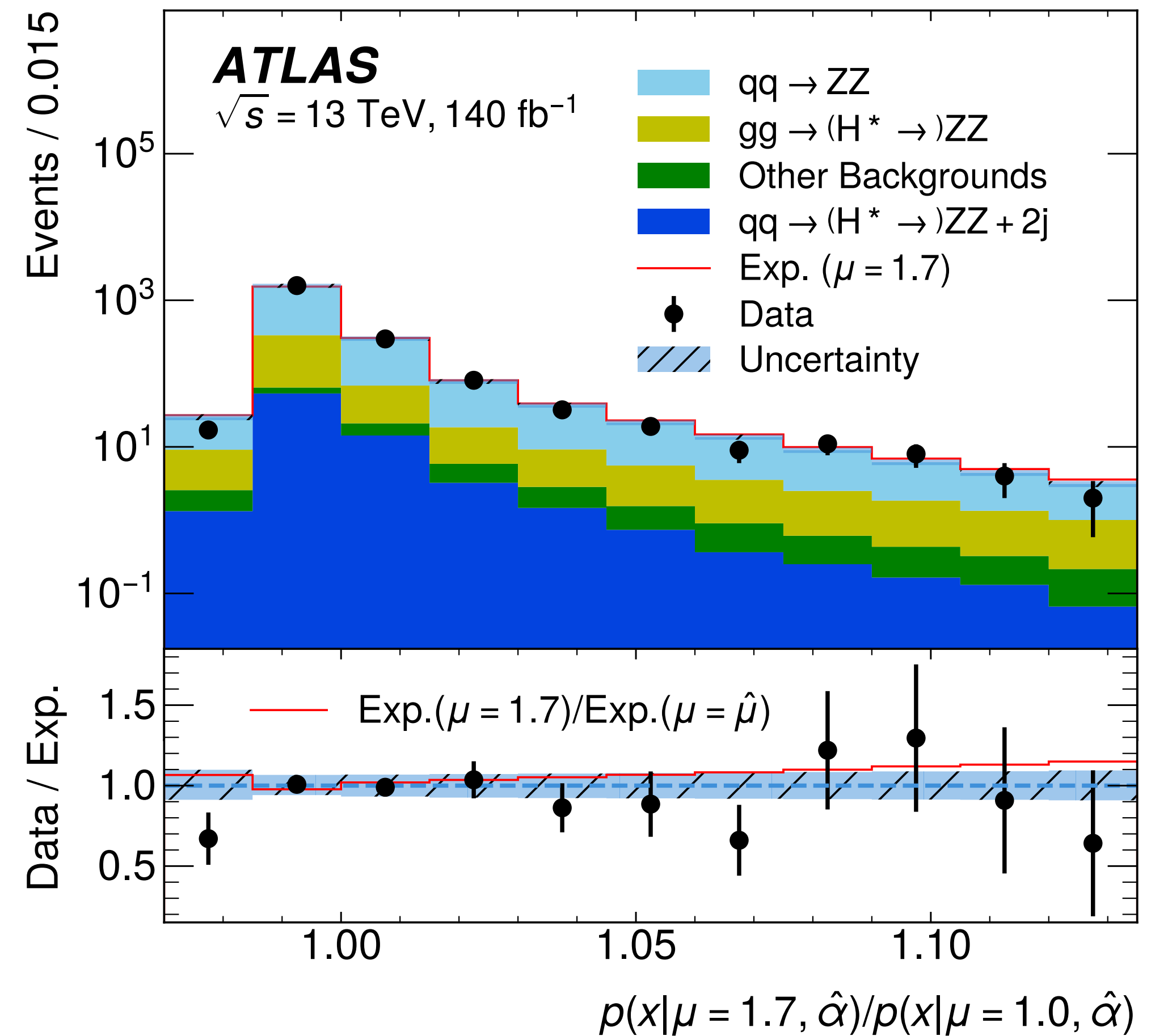
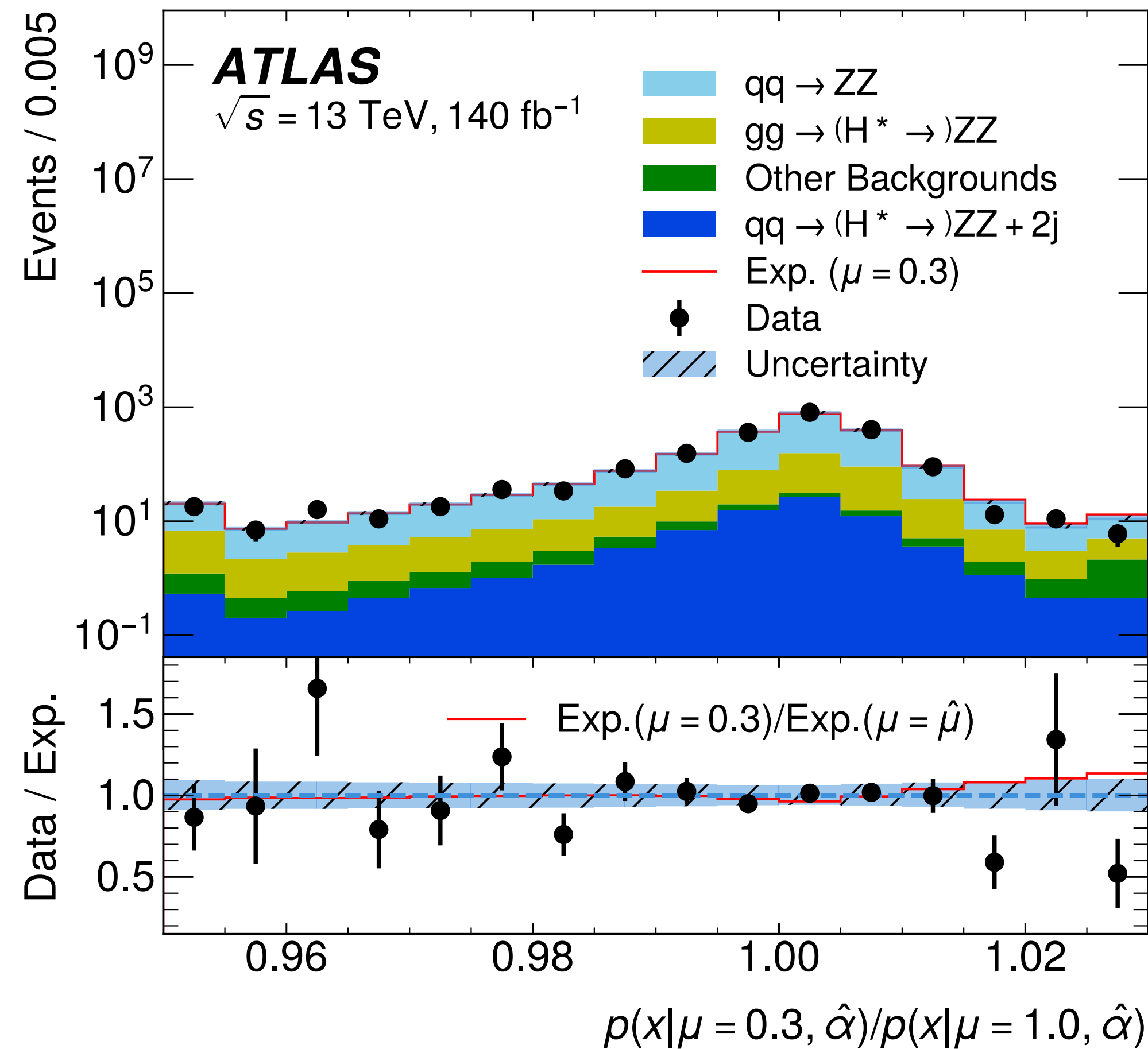
Data-MC validation

NN observable

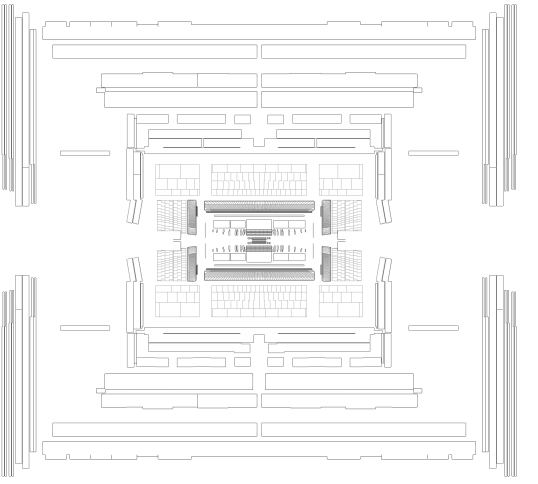
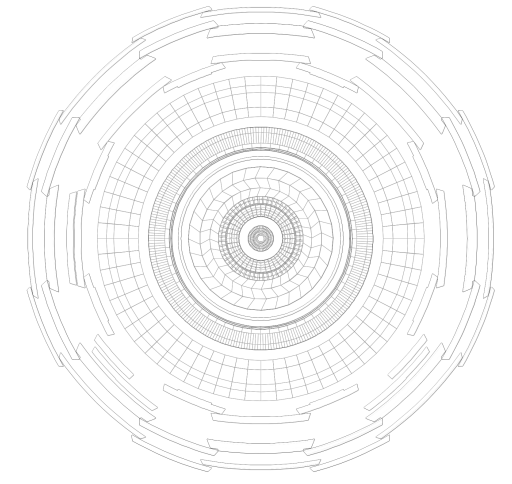
 m_{4l} 

Data-MC validation

Different NN observables



Physics & statistics motivation



A measurement of the Higgs width via off-shell couplings

- Enables the probe of a wide variety of new massive particles, other new physics
- Can't measure directly: SM Higgs width ~ 4 MeV, resolution of detector ~ 1 GeV
- Can be probed indirectly by measuring off-shell couplings (and combining with on-shell measurements)

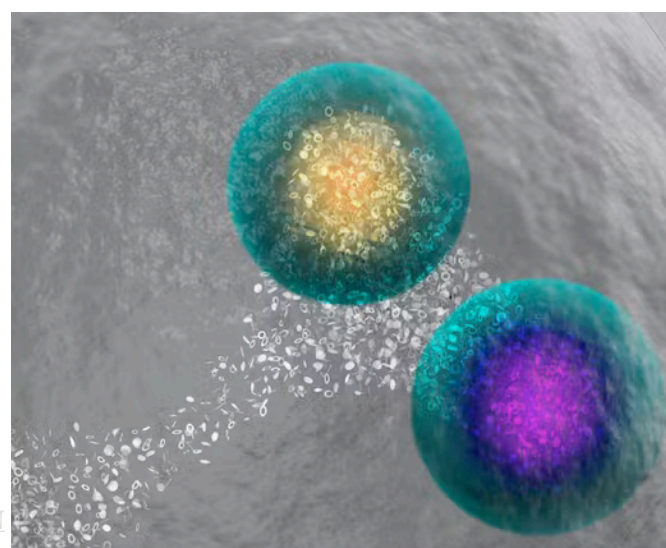
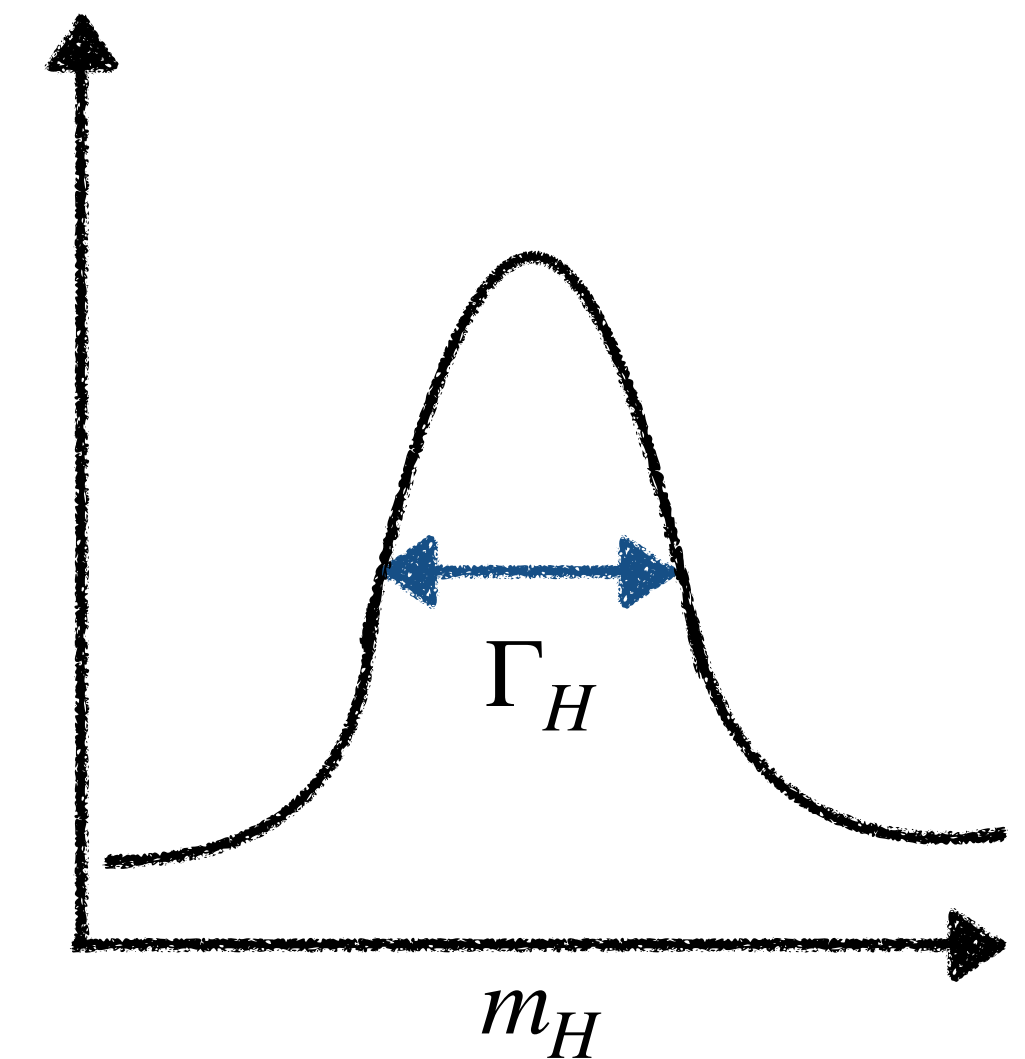


Image: CERN

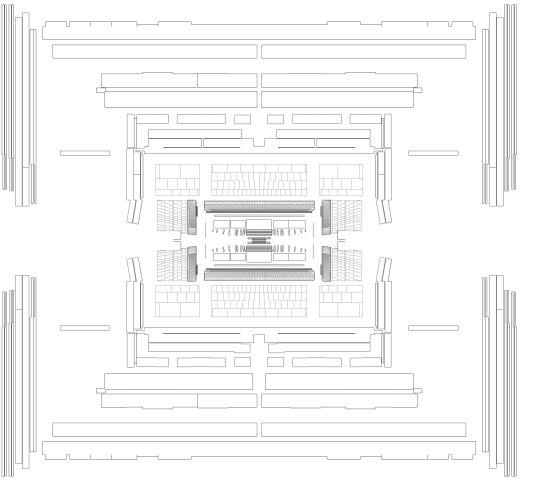
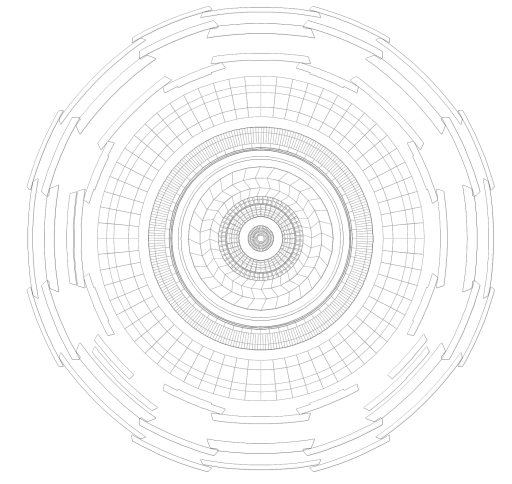


H



Γ_H

Undiscovered massive particles



Higgs Width from off-shell Higgs production

[arXiv:1405.0285](https://arxiv.org/abs/1405.0285) [arXiv:1406.1757](https://arxiv.org/abs/1406.1757)

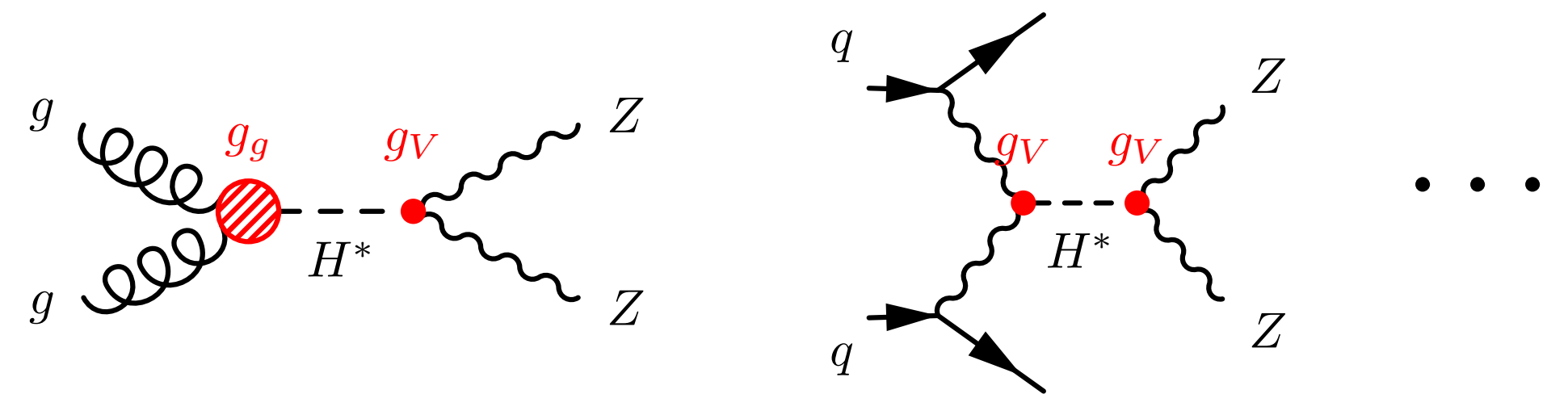
- Off-shell production helps probe Higgs width

$$\frac{\mu_{\text{off-shell}}}{\mu_{\text{on-shell}}} = \frac{\Gamma_H}{\Gamma_H^{SM}}.$$

$$\sigma_{on-shell}^{gg \rightarrow H \rightarrow ZZ} \sim \frac{g_f^2 g_V^2}{m_H \Gamma_H}$$

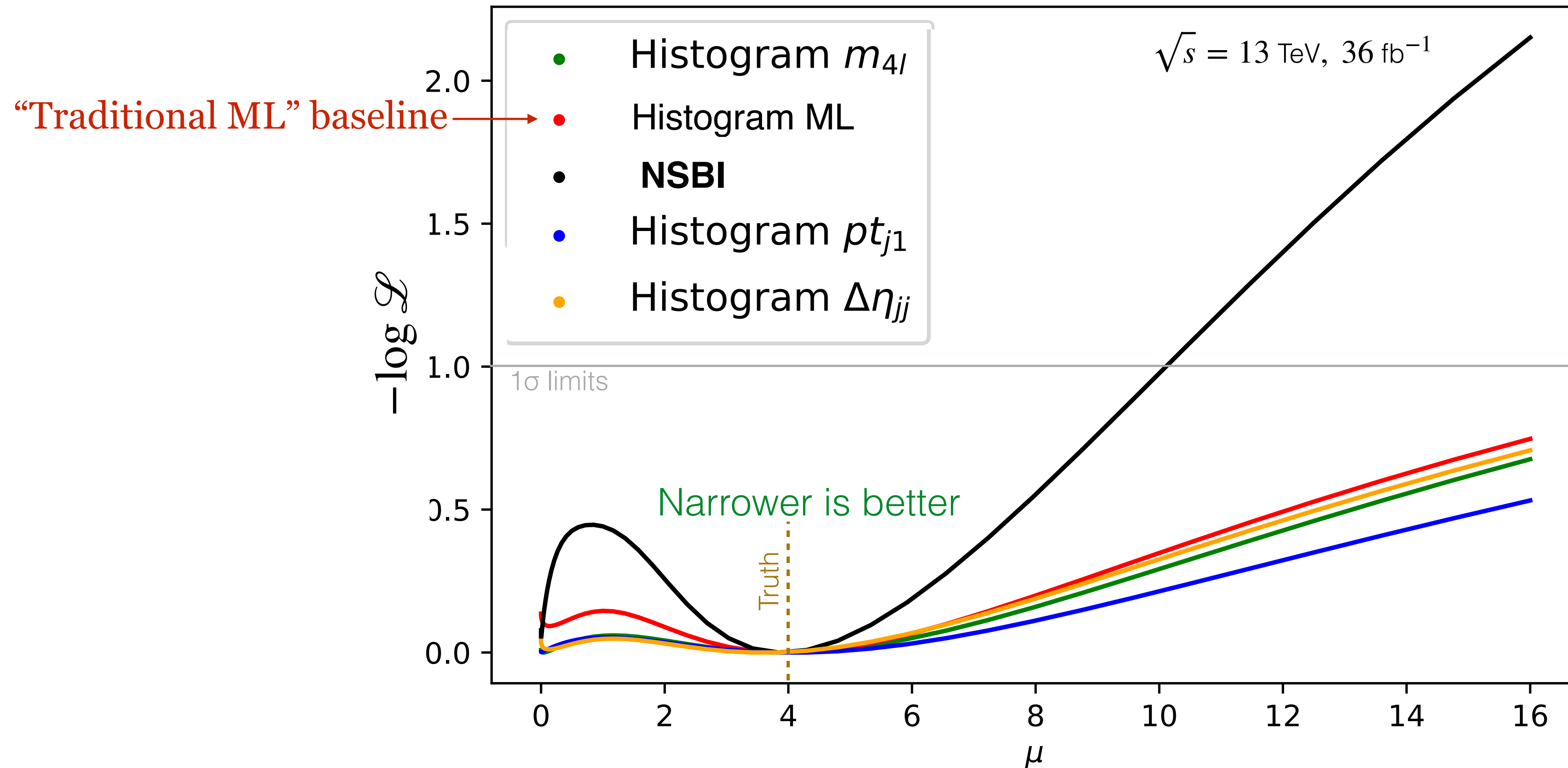
$$\frac{d\sigma_{off-shell}^{gg \rightarrow H \rightarrow VV}}{dm_{VV}} \propto \frac{g_f^2 g_V^2}{(m_{VV}^2 - m_H^2)^2 + m_H^2 \Gamma_H^2} \simeq \frac{g_f^2 g_V^2}{m_{VV}^4}$$

- Interpretation assumes no new physics
- Essential to measure independently in multiple production modes (ggF, VBF) and final states to verify consistent results



NSBI for Higgs width in proof-of-concept study

Expected sensitivity [hal-02971995v3](#) (p172): Ghosh & Rousseau, [Thesis](#): Ghosh



Simulations with $\mu = 4$

More on the method

Reference Sample

A combination of signal samples, to ensure non-zero probability in entire region of analysis
Does not have to be physical!

$$p_{\text{ref}}(x_i) = \frac{1}{\sum_k v_k} \sum_k^{C_{\text{signals}}} v_k \cdot p_k(x_i)$$

$\Rightarrow$ In our dataset, $p_{\text{ref}}(\cdot) = p_S(\cdot)$

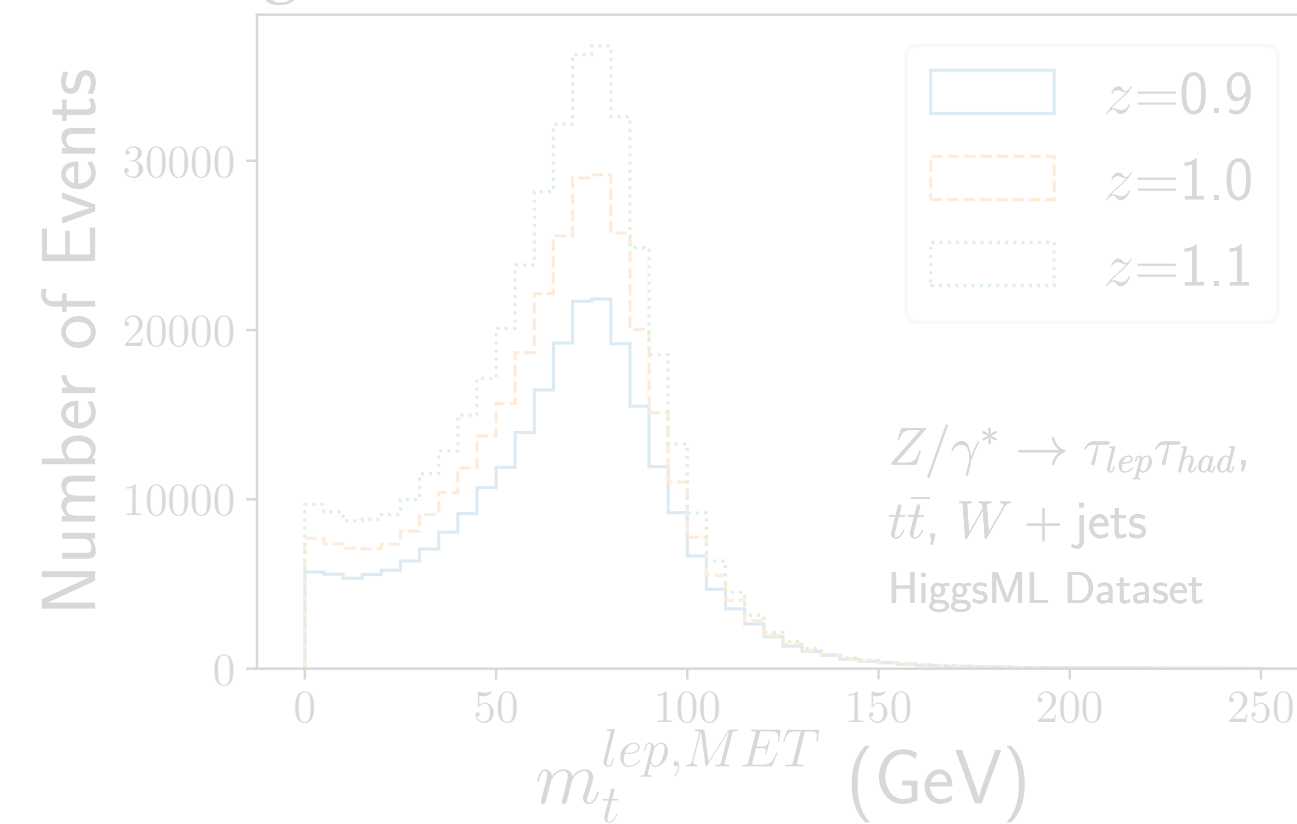
Choice of $p_{\text{ref}}(\cdot)$ can be made purely on numerical stability of training, as it drops out in profile step

$$t_\mu = -2 \ln \left(\frac{L_{\text{full}}(\mu, \hat{\hat{\alpha}}) / \cancel{L_{\text{ref}}}}{L_{\text{full}}(\hat{\mu}, \hat{\alpha}) / \cancel{L_{\text{ref}}}} \right)$$

Systematic uncertainties

Experimental uncertainties:

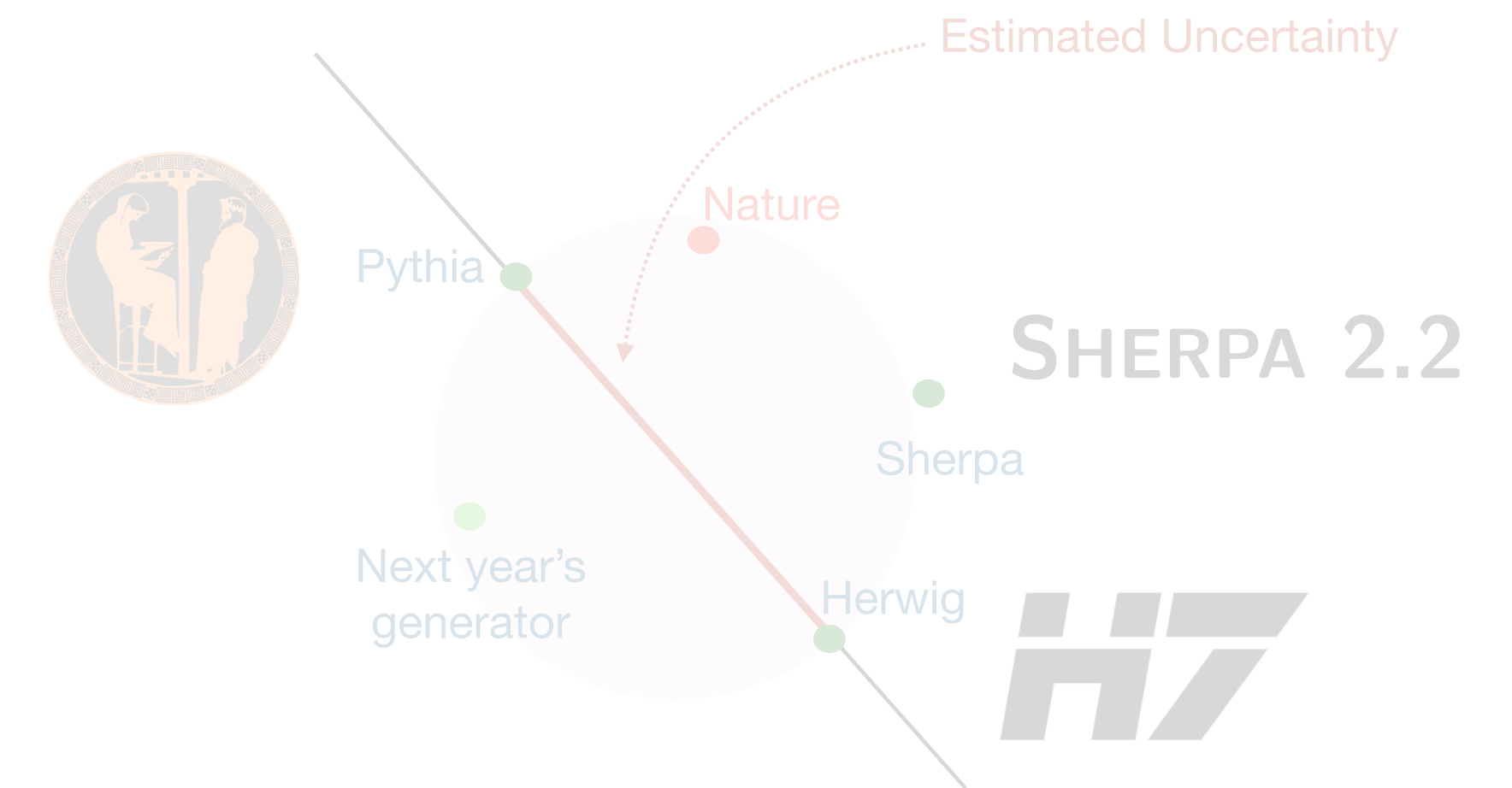
Eg. Inaccuracies in the calibration of our detector



Ghosh et al: [PhysRevD.104.056026](#)

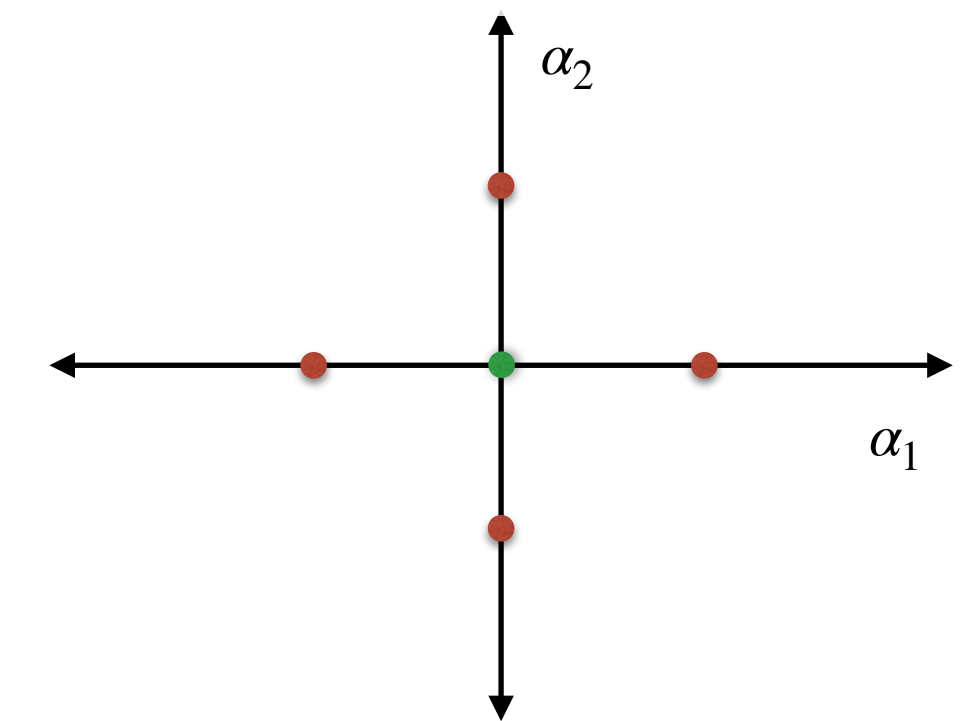
Theory uncertainties:

Eg. Inability to compute QFT to infinite order



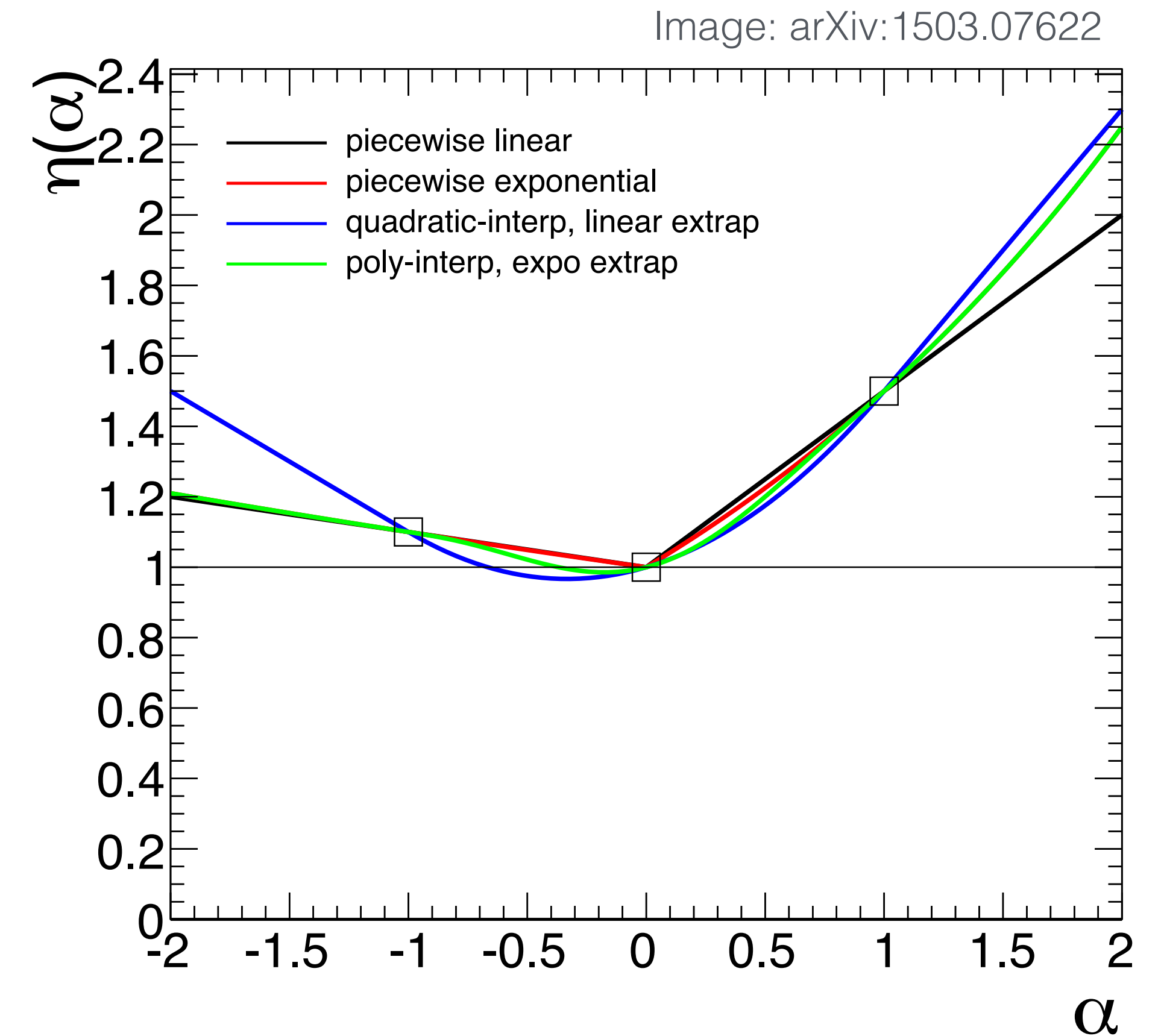
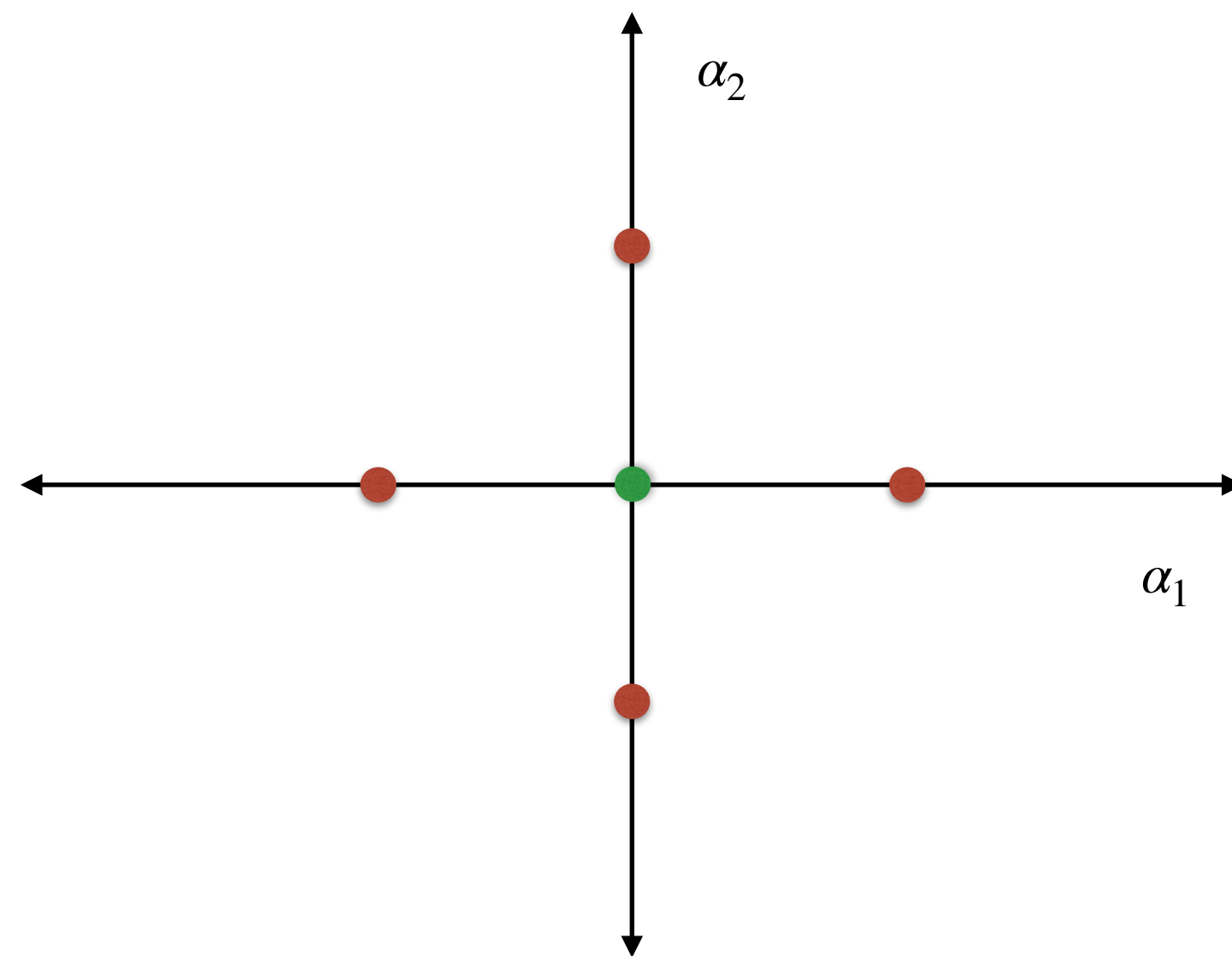
Ghosh and Nachman: [epjc.s10052.022.10012.w](#)

- We only have simulations at 3 variations of each nuisance parameter α_k



Known interpolation strategies

See formula used in [backup](#)



⇒ Combine these traditional interpolation with neural network estimation of per-event likelihood ratios

Probability density ratio including nuisance parameters (α)

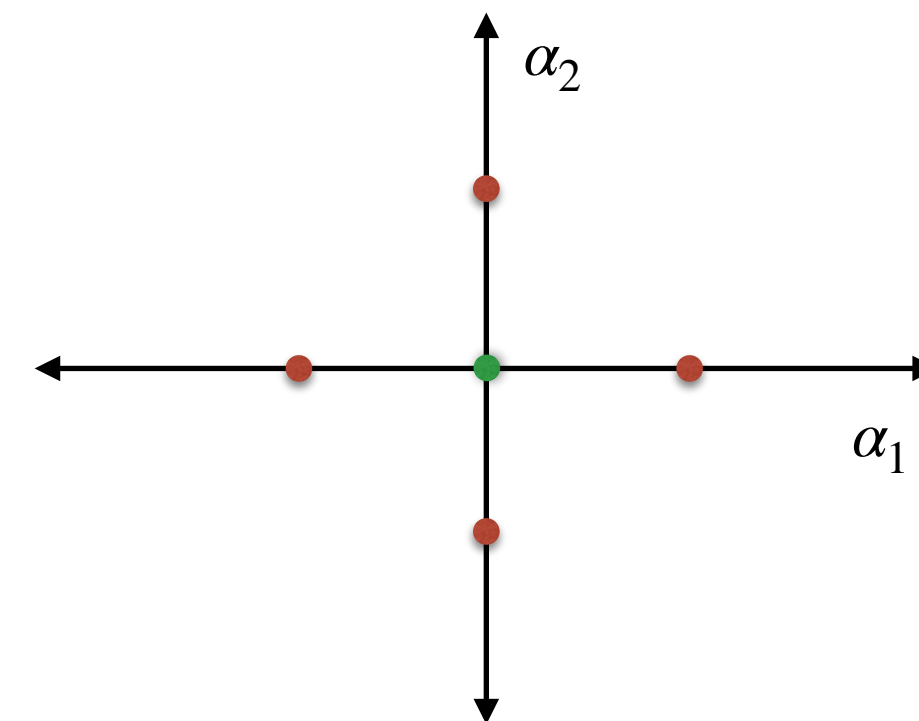
x_i is one individual event

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

We have this already

Per-event terms estimated using another ensemble of networks and interpolation methods

Estimate from simulations and existing interpolation methods

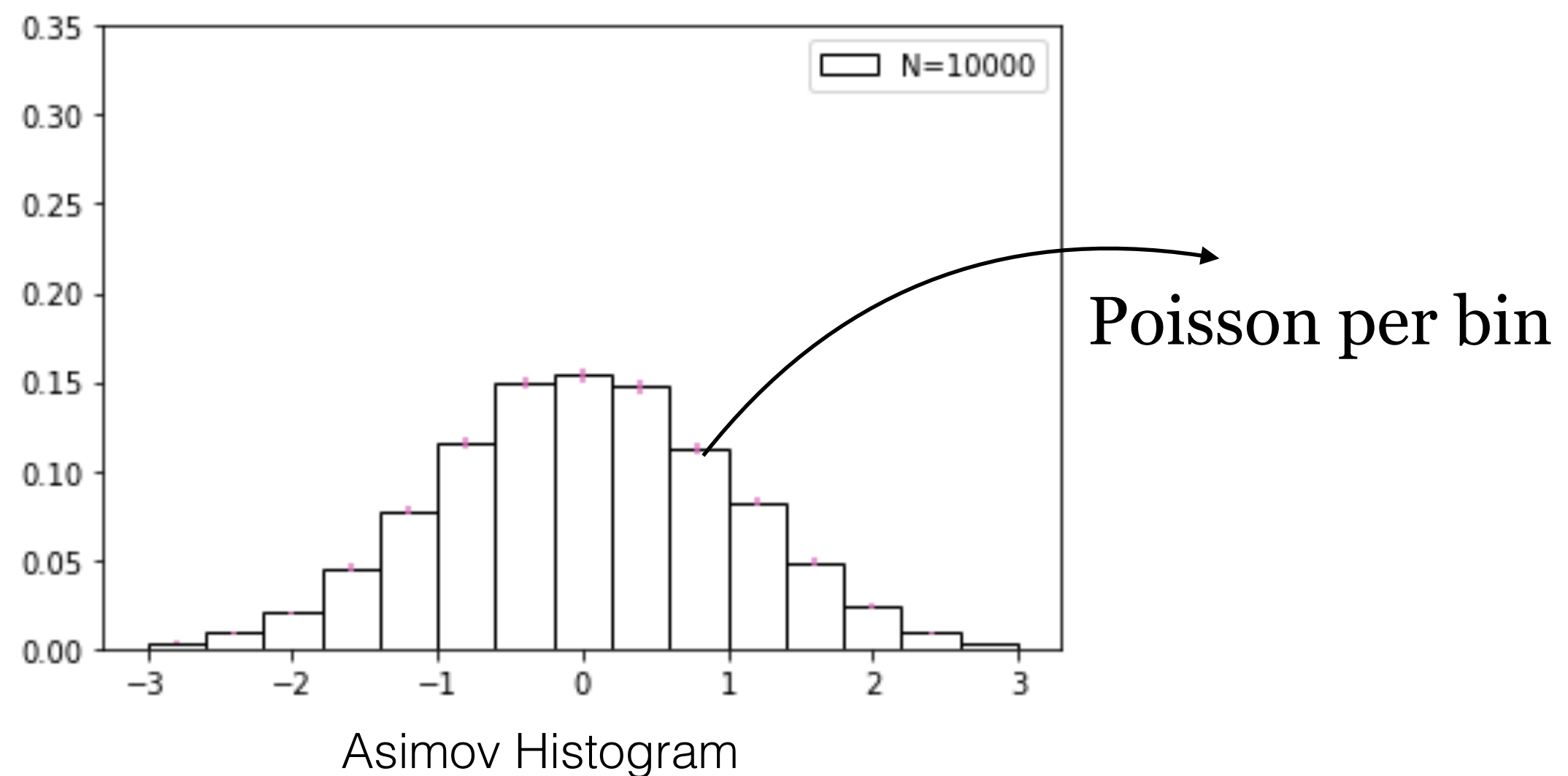


$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

Generating event-level pseudo-experiments

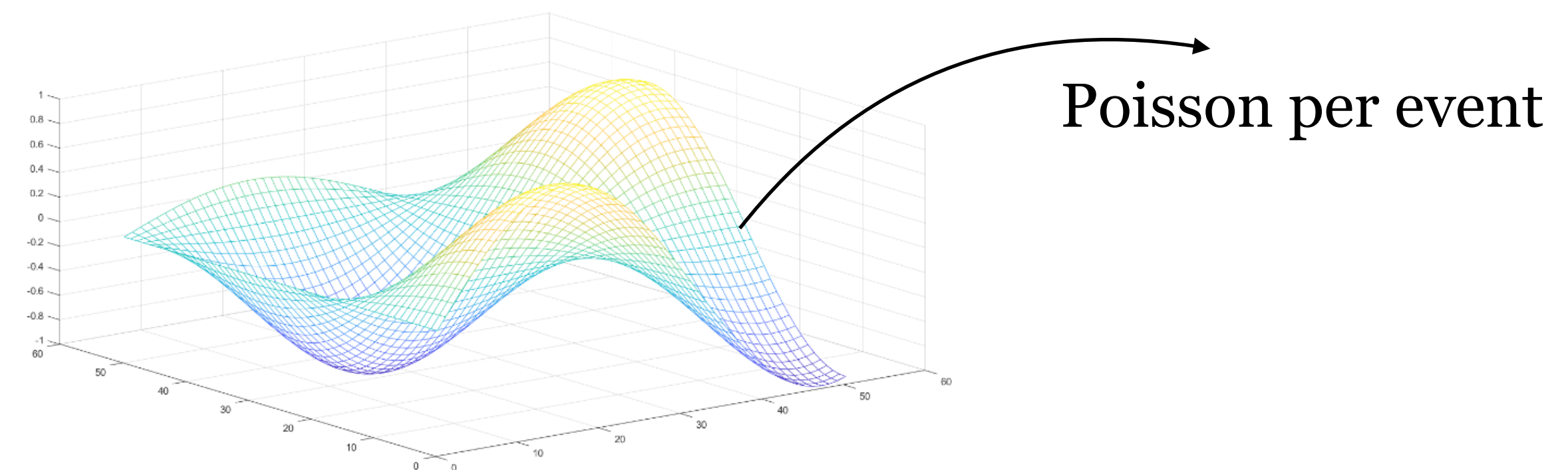
Need to generate random possible datasets we could collect at the LHC

Traditionally:



$$N_i^{toy} = \text{Poisson}(N_i^{Asimov})$$

NSBI:



$$w_i^{toy} = \text{Poisson}(w_i^{Asimov})$$

‘Unweighted’ events, i.e. integer weights

Dealing with negative weighted events

$$w_i^{toy} = \text{Poisson}(w_i^{Asimov})$$

Simulated samples include events with negative weights due to the way we calculate QFT higher order effects

Use a positive weighted sample instead:

1. Start from a positive weighted reference sample
2. Re-weight it to intended parameter point in μ, α
3. Throw toys from this sample

$$w_i^{\text{rwt-ref}} \rightarrow w_i^{\text{Asimov}}(\mu, \alpha) = \frac{\nu(\mu, \alpha)}{\nu_{\text{rwt-ref}}} \cdot \frac{p(x_i | \mu, \alpha)}{p_{\text{rwt-ref}}(x_i)} \cdot w_i^{\text{rwt-ref}}$$

Combination with histogram analyses

$$\frac{L_{\text{comb}}(\mu, \alpha)}{L_{\text{ref}}} = \frac{L_{\text{full}}(\mu, \alpha)}{L_{\text{ref}}} L_{\text{hist}}(\mu, \alpha)$$

Calculating pulls and impacts in JAX

Hessian:

$$C_{nm} = \left[\frac{1}{2} \frac{\partial^2 \lambda}{\partial \alpha_n \partial \alpha_m}(\hat{\mu}, \hat{\alpha}) \right]^{-1}$$

$$\lambda(\mu, \alpha) = -2 \ln(L_{full}(\mu, \alpha)/L_{ref})$$

Pulls:

$$\frac{\hat{\alpha}_k - \alpha_k^0}{\sqrt{C_{kk}}}.$$

Post-fit Impact:

$$\begin{aligned} \Gamma_k &= \frac{\partial \hat{\mu}}{\partial \alpha_k} \times \sqrt{C_{kk}} \\ &= - \left[\frac{\partial^2 \lambda}{\partial^2 \mu}(\hat{\mu}, \hat{\alpha}) \right]^{-1} \frac{\partial^2 \lambda}{\partial \mu \partial \alpha_k}(\hat{\mu}, \hat{\alpha}) \times \sqrt{C_{kk}}, \end{aligned}$$

Vertical interpolation

$$G_j(\alpha_k) = \begin{cases} \left(\frac{\nu_j(\alpha_k^+)}{\nu_j(\alpha_k^0)} \right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^6 c_n \alpha_k^n & -1 \leq \alpha_k \leq 1 \\ \left(\frac{\nu_j(\alpha_k^-)}{\nu_j(\alpha_k^0)} \right)^{-\alpha_k} & \alpha_k < -1 \end{cases} \quad g_j(x_i, \alpha_k) = \begin{cases} \left(g_j(x_i, \alpha_k^+) \right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^6 c_n \alpha_k^n & -1 \leq \alpha_k \leq 1 \\ \left(g_j(x_i, \alpha_k^-) \right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

With some continuity requirements

Physics analysis results

Comparison to previous result on same data

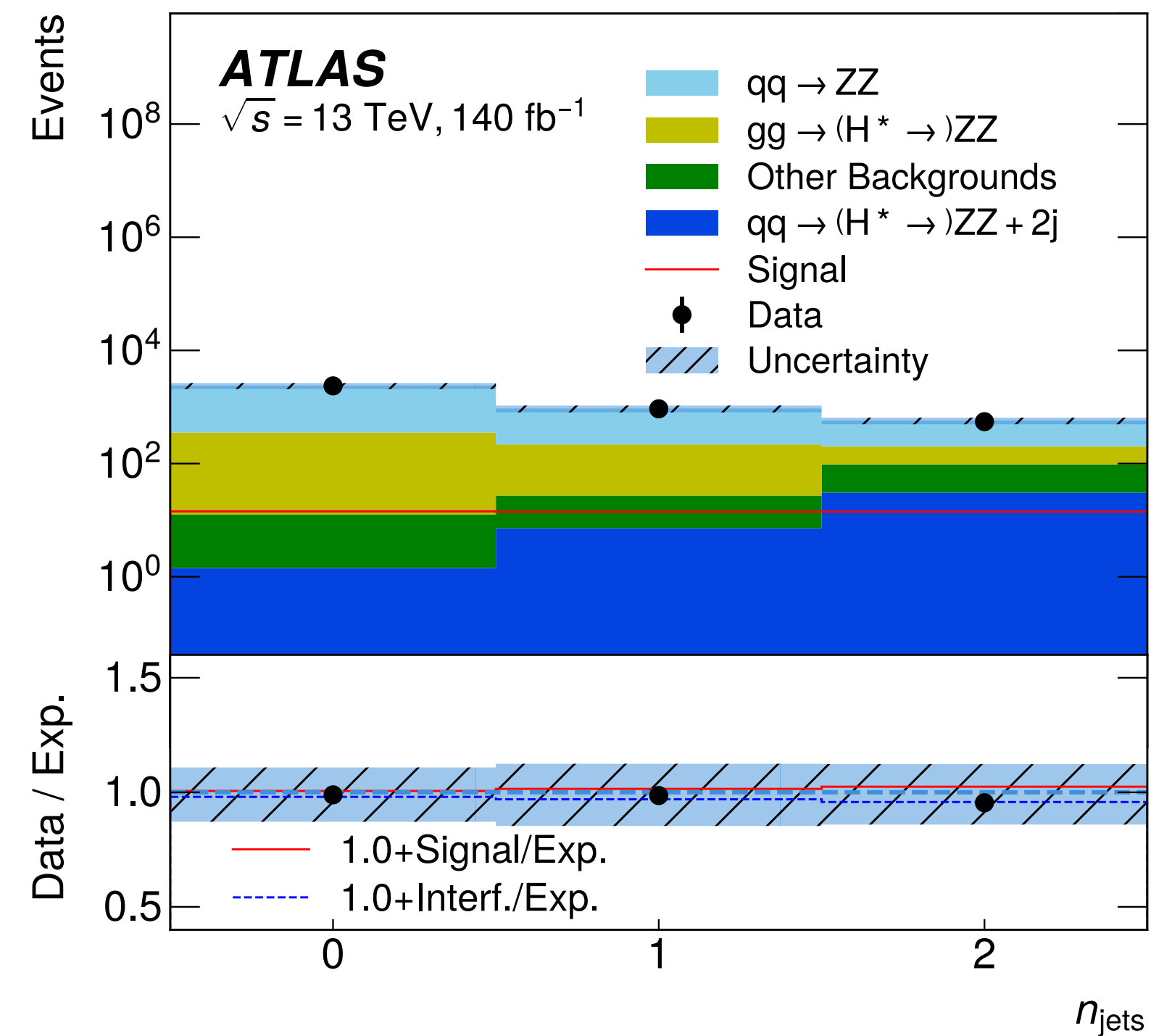
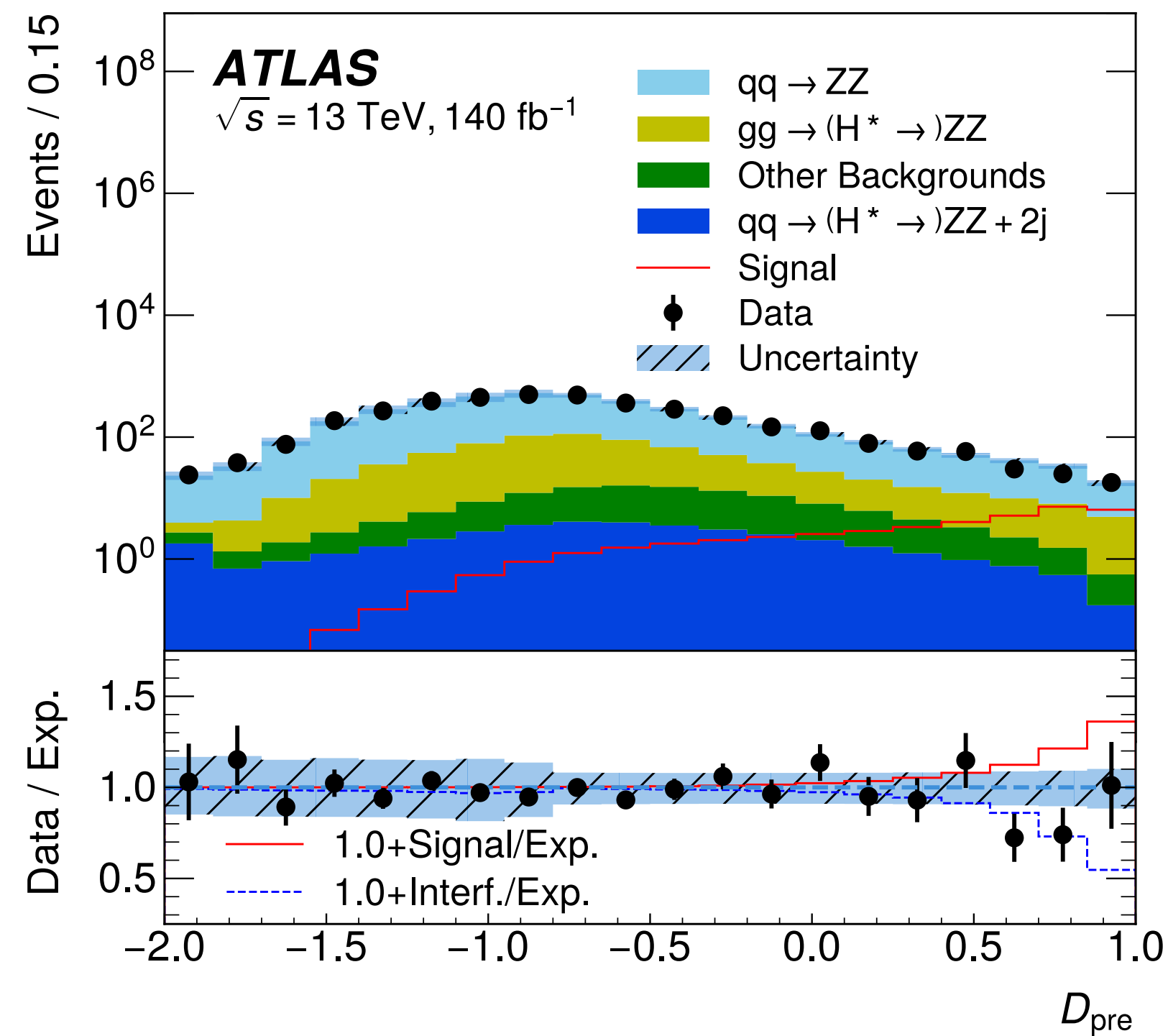
Parameter	Value	68% CL interval		95% CL interval	
		Observed	Expected	Observed	Expected
NSBI analysis					
$\mu_{\text{off-shell}} (4\ell \text{ only})$	0.87	[0.33, 1.62]	[0.05, 2.04]	[0.05, 2.38]	< 2.38
$\mu_{\text{off-shell}}$	1.06	[0.61, 1.67]	[0.17, 1.83]	[0.21, 2.24]	[0.01, 2.42]
Γ_H [MeV] (4ℓ only)	3.43	[1.37, 6.71]	[0.20, 8.25]	[0.18, 9.98]	< 12.09
Γ_H [MeV]	4.29	[2.41, 6.95]	[0.66, 7.61]	[0.76, 9.66]	[0.12, 10.50]
R_{gg}	1.19	[0.53, 2.07]	[0.02, 1.92]	< 2.96	< 2.73
R_{VV}	0.95	[0.61, 1.39]	[0.31, 1.70]	[0.30, 1.86]	[0.06, 2.14]
Histogram-based analysis					
$\mu_{\text{off-shell}} (4\ell \text{ only})$	0.79	[0.02, 2.00]	< 2.14	< 2.97	< 3.10
$\mu_{\text{off-shell}}$	1.09	[0.54, 1.81]	[0.08, 1.90]	[0.10, 2.41]	[0.01, 2.52]
Γ_H [MeV] (4ℓ only)	3.43	[0.10, 8.42]	< 8.89	< 12.48	< 12.89
Γ_H [MeV]	4.37	[2.13, 7.43]	[0.35, 7.94]	[0.39, 10.14]	< 10.79
R_{gg}	1.23	[0.00, 2.20]	< 1.98	< 3.15	< 2.84
R_{VV}	0.95	[0.60, 1.43]	[0.27, 1.74]	[0.26, 1.90]	[0.02, 2.18]

Impact of nuisance parameters

Systematic Uncertainty Fixed	$\mu_{\text{off-shell}}$ Value at which $t_{\mu_{\text{off-shell}}} = 4$	
	NSBI analysis	Histogram-based
All (stat-only)	1.96	2.13
Parton shower uncertainty for $gg \rightarrow ZZ$ (normalization)	2.07	2.26
Parton shower uncertainty for $gg \rightarrow ZZ$ (shape)	2.12	2.29
NLO EW uncertainty for $q\bar{q} \rightarrow ZZ$	2.10	2.27
NLO QCD uncertainty for $gg \rightarrow ZZ$	2.09	2.29
Parton shower uncertainty for $q\bar{q} \rightarrow ZZ$ (shape)	2.12	2.29
Jet energy scale and resolution uncertainty	2.11	2.26
None (full result)	2.12	2.30

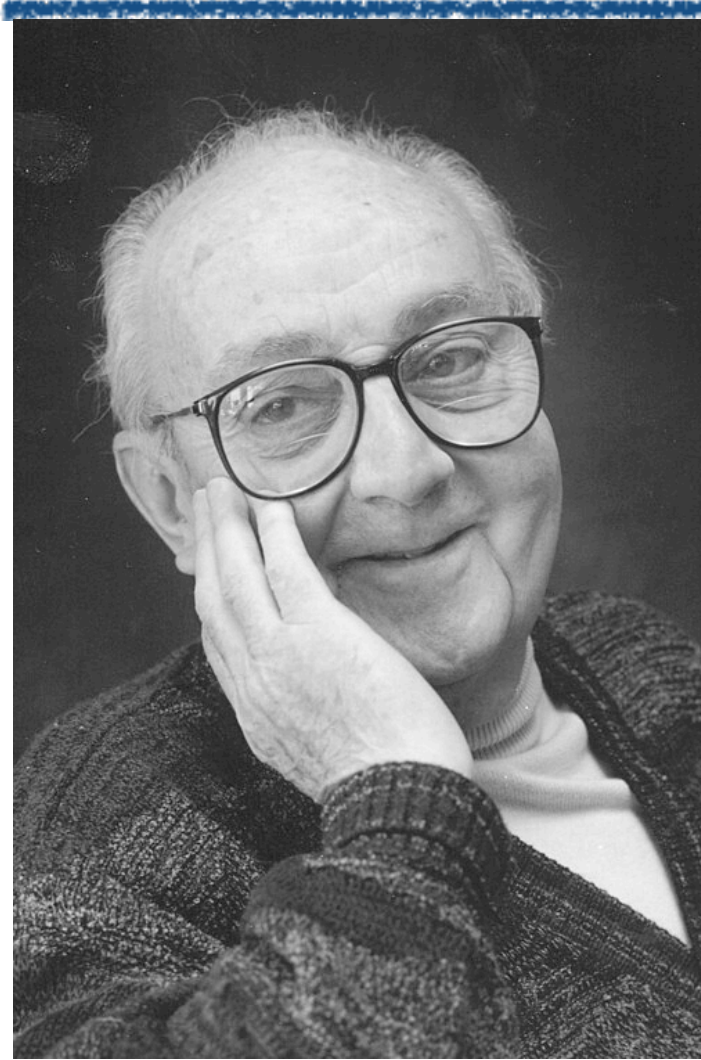
Pre-selection region definition

$$D_{\text{pre}}(x) = \log \frac{s_{\text{pre}, S}^{\text{ggF}}(x) + s_{\text{pre}, S}^{\text{EW}}(x)}{s_{\text{pre}, B}^{\text{ggF}}(x) + s_{\text{pre}, B}^{\text{EW}}(x) + s_{\text{pre}, q\bar{q}ZZ}(x)},$$



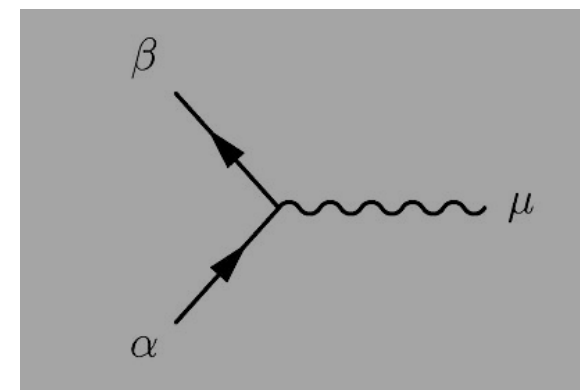
AI for hypothesis generation

Physicists build mathematical models of the Universe



“All models are wrong, but some are useful”

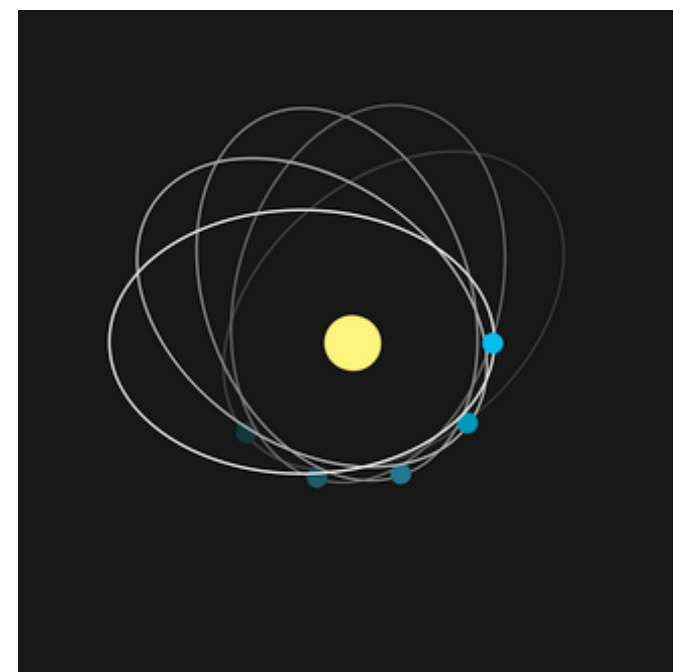
What makes a useful model?: They are predictive



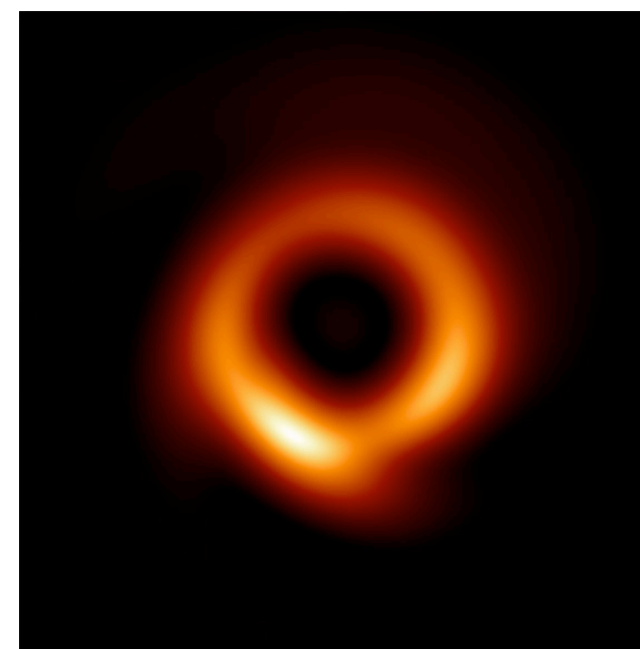
Particle physics is more messy!

General Relativity reproduces
Newtonian gravity in the right
limits

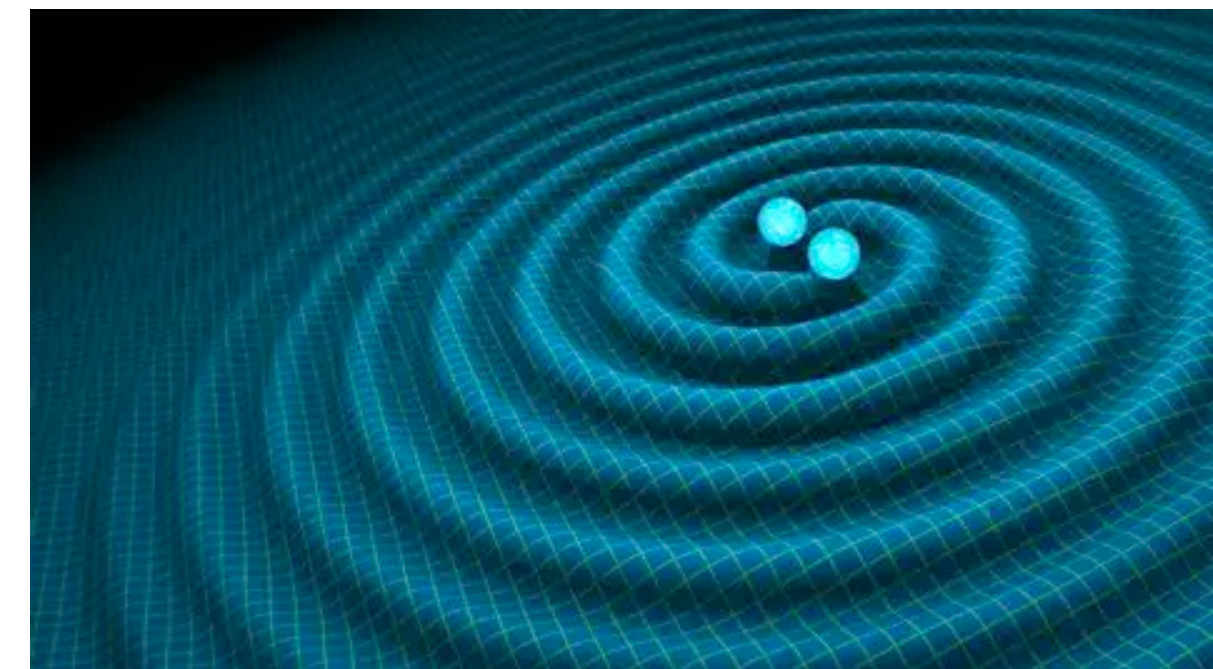
But also predicts:



Perihelion of mercury



Black Holes



Gravitational Waves

Visible universe explained by this 'Lagrangian'

$$\begin{aligned}\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ & + i\bar{\psi}\not{D}\psi + h.c. \\ & + \bar{\psi}_i Y_{ij} \psi_j \phi + h.c. \\ & + |D_\mu \phi|^2 - V(\phi)\end{aligned}$$

The Standard Model of Particle Physics. Highly successful.

But also deeply flawed:

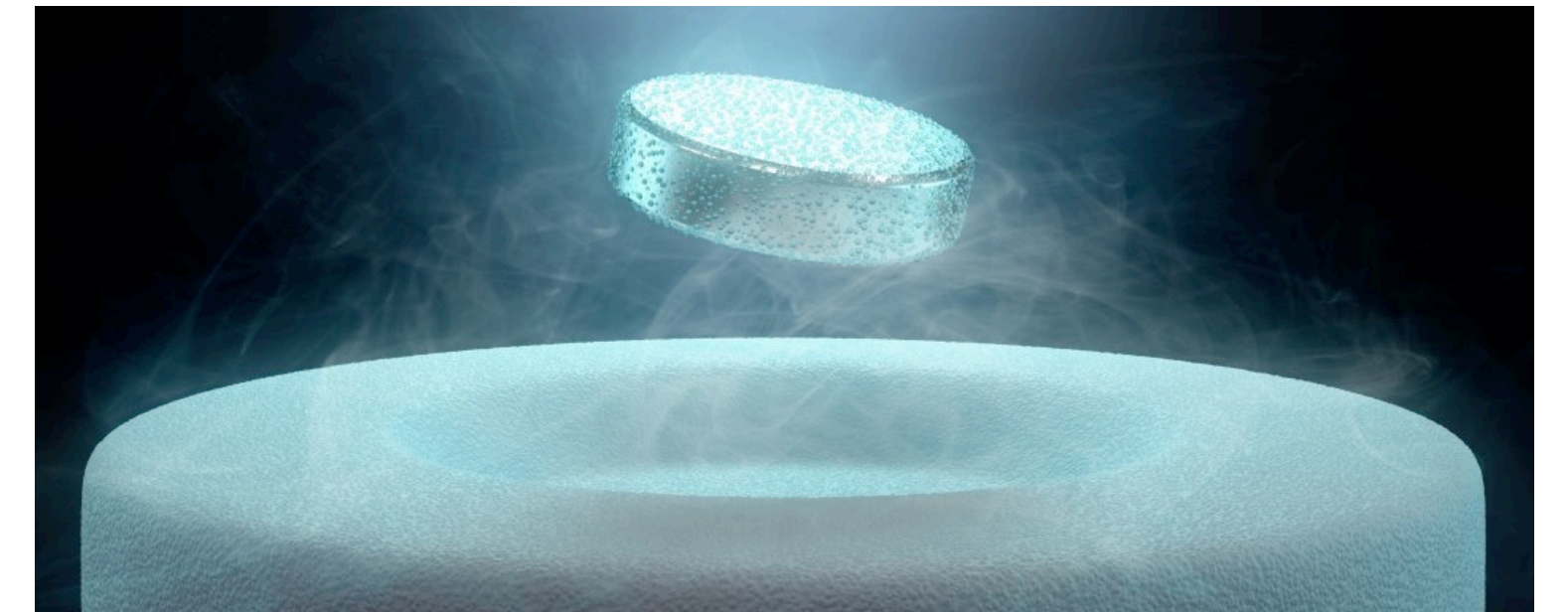
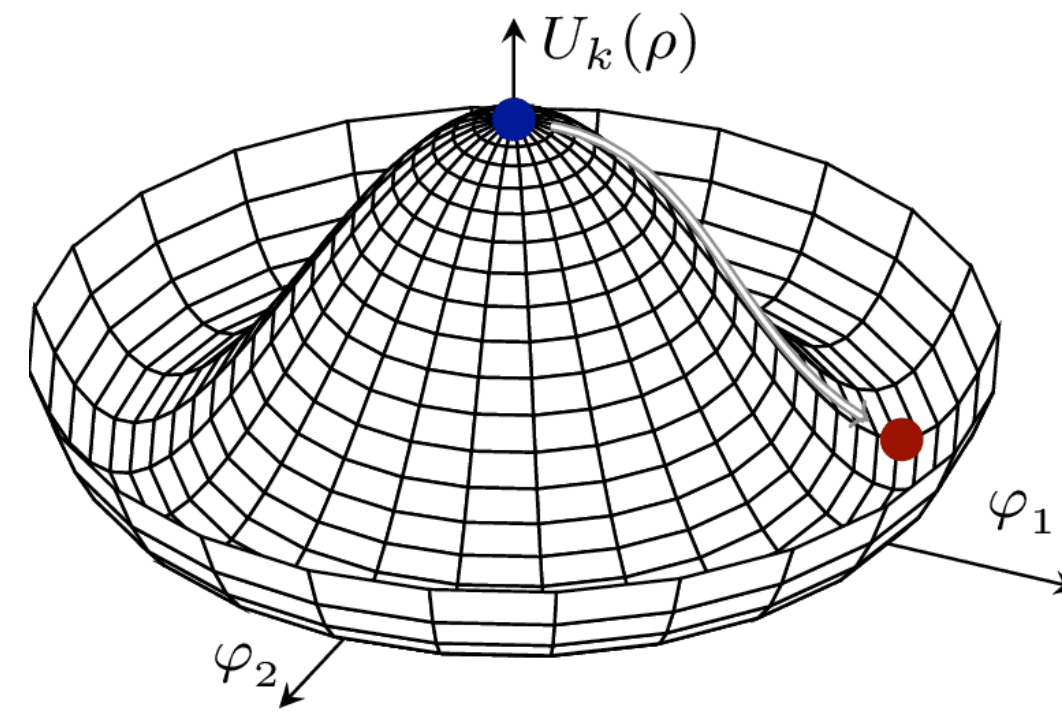
- Can't explain 85% of all matter (dark matter)
- Why more matter than anti-matter?
- Model parameters need incredible fine-tuning
- ...
- **Why do neutrinos have mass and why do they oscillate?**

Could we automatically design better theories and test them against data?

$$\nu_\tau \quad | \quad \nu_e \quad | \quad \nu_\mu$$

The space of mathematical tricks we could use is enormous

In the old electroweak theory, **we didn't know how to give particles like Z and W bosons their mass** without violating important symmetries



Higgs mechanism to the rescue:

- The core **mathematical trick is unintuitive, inspired by revelations from superconductivity research** in the 1950s!
- How long would particle physicists have take to figure it out on their own?

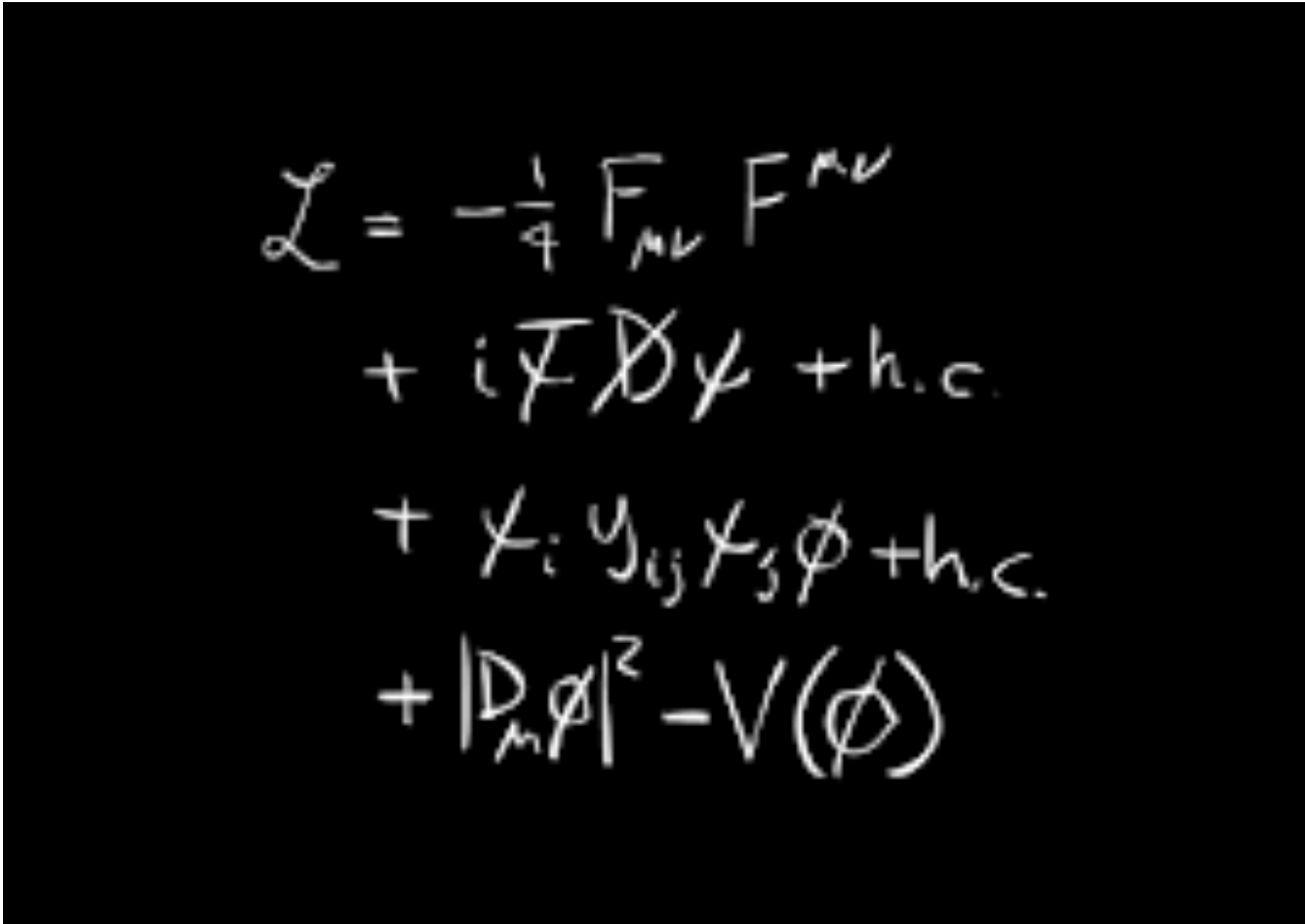
What if solutions to current open problems are in other unintuitive mathematical tricks?

Neutrinos remain massless in Standard Model, can't explain their oscillations

A testing ground for automated theory design

ν_τ | ν_e | ν_μ

How to design your own particle physics theory ?


$$\begin{aligned}\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ & + i \bar{\psi} \not{D} \psi + \text{h.c.} \\ & + \chi_i Y_{ij} \chi_j \phi + \text{h.c.} \\ & + |D_\mu \phi|^2 - V(\phi)\end{aligned}$$

Determined by choice of **fields** and **symmetries**
they must obey

Fields sit inside symmetry group representations

$\mathfrak{su}(2)$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

Adding fields: Akin to adding particles to your universe, gives you more free parameters, more flexibility.
Trivial to explain data with 100s of free parameters

Adding symmetries: Adds structures and restrictions, reduces flexibility but makes the theory richer

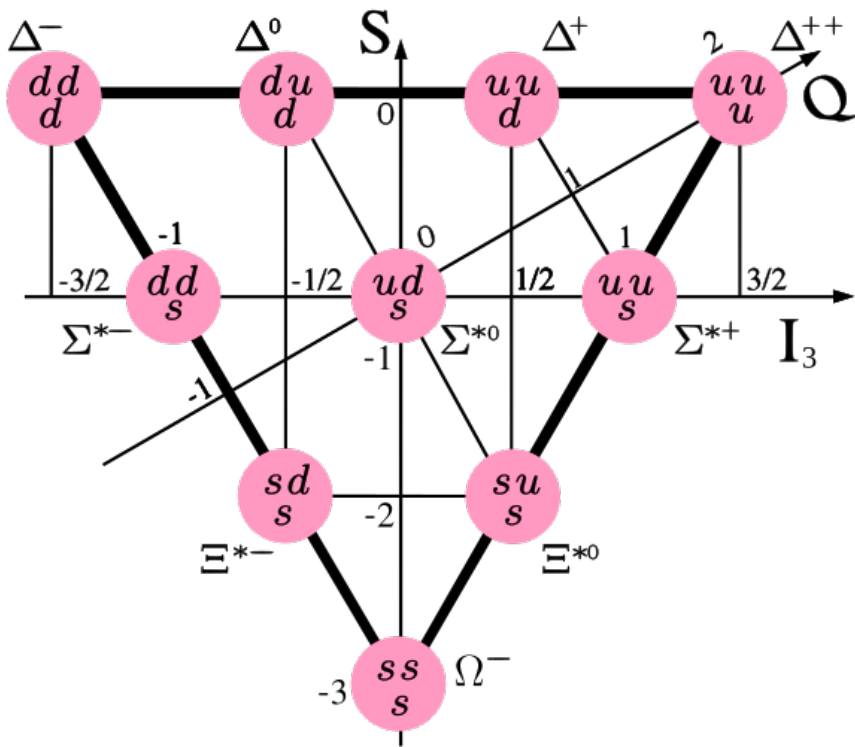
A rich space of **discrete** mathematical choices

$$\mathfrak{su}(3)$$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$



Takes months to build intuition for how to use these symmetry groups and their representations

Risky: After months of study, you might hit a dead-end

$$\mathfrak{su}(2)$$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

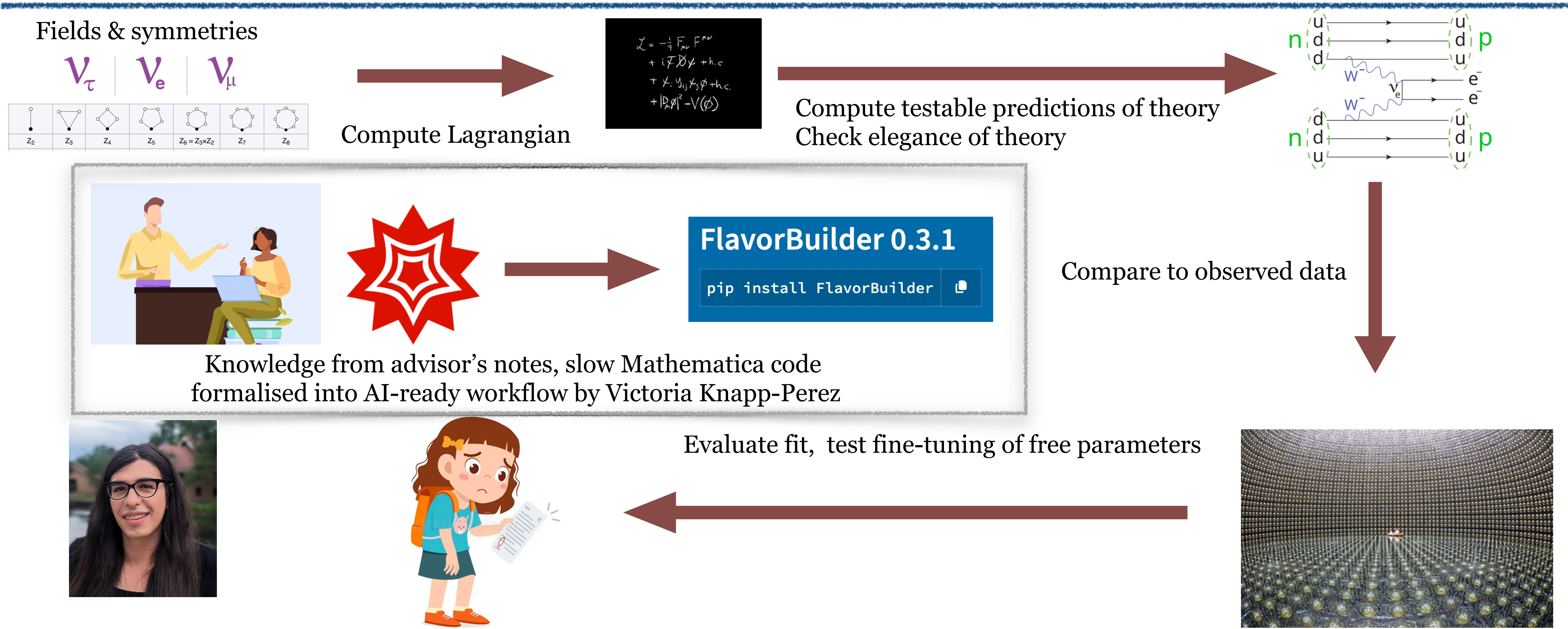
$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

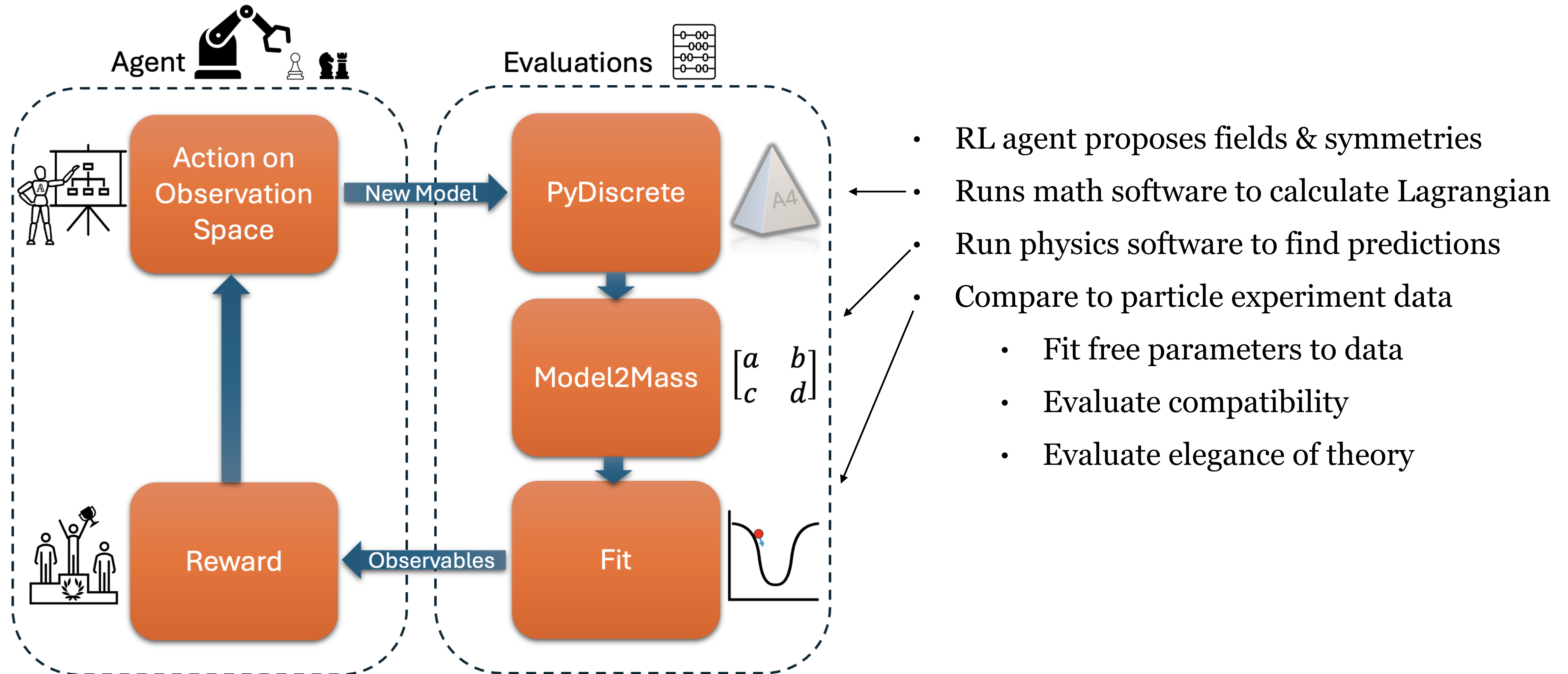
Cycle graphs up to order 24

Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆ = Z ₃ ×Z ₂	Z ₇	Z ₈
Z ₉	Z ₁₀ = Z ₅ ×Z ₂	Z ₁₁	Z ₁₂ = Z ₄ ×Z ₃	Z ₁₃	Z ₁₄ = Z ₇ ×Z ₂	Z ₁₅ = Z ₅ ×Z ₃	Z ₁₆
Z ₁₇	Z ₁₈ = Z ₉ ×Z ₂	Z ₁₉	Z ₂₀ = Z ₅ ×Z ₄	Z ₂₁ = Z ₇ ×Z ₃	Z ₂₂ = Z ₁₁ ×Z ₂	Z ₂₃	Z ₂₄ = Z ₈ ×Z ₃

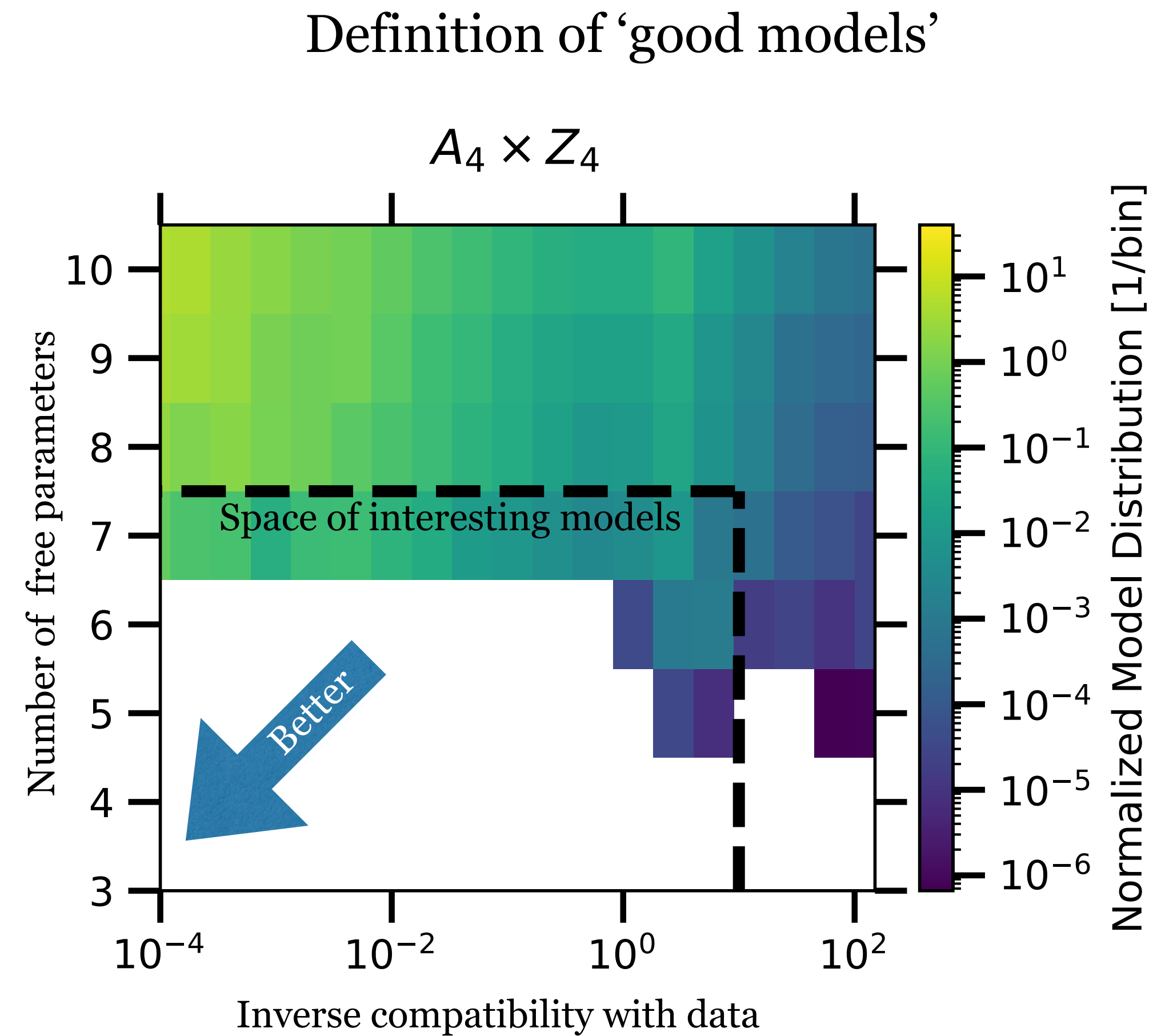
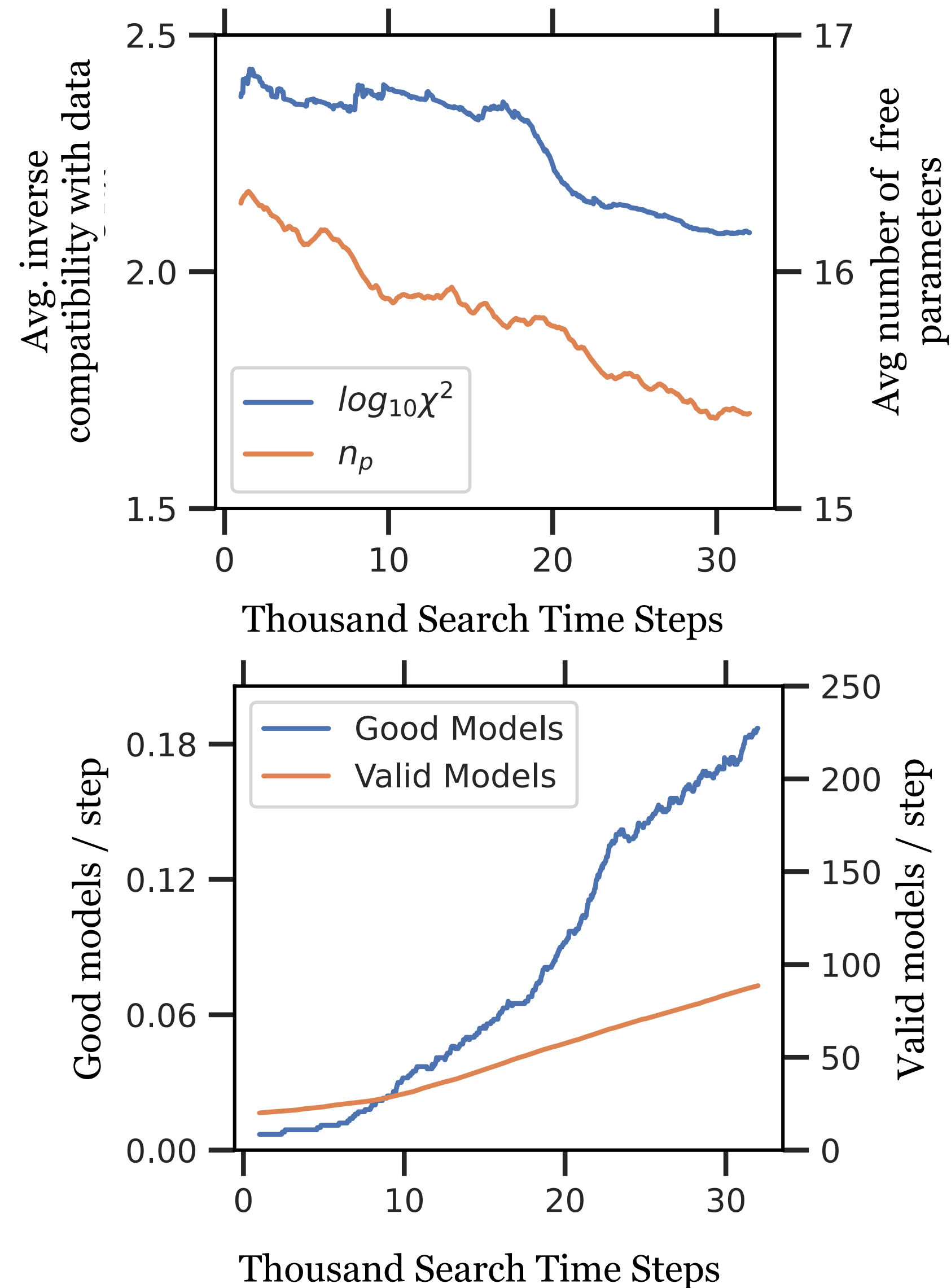
Design end-to-end automated pipeline: Physics software to form building blocks



Reinforcement-learning-assisted theory design



Agent learns to design good theories... too many!



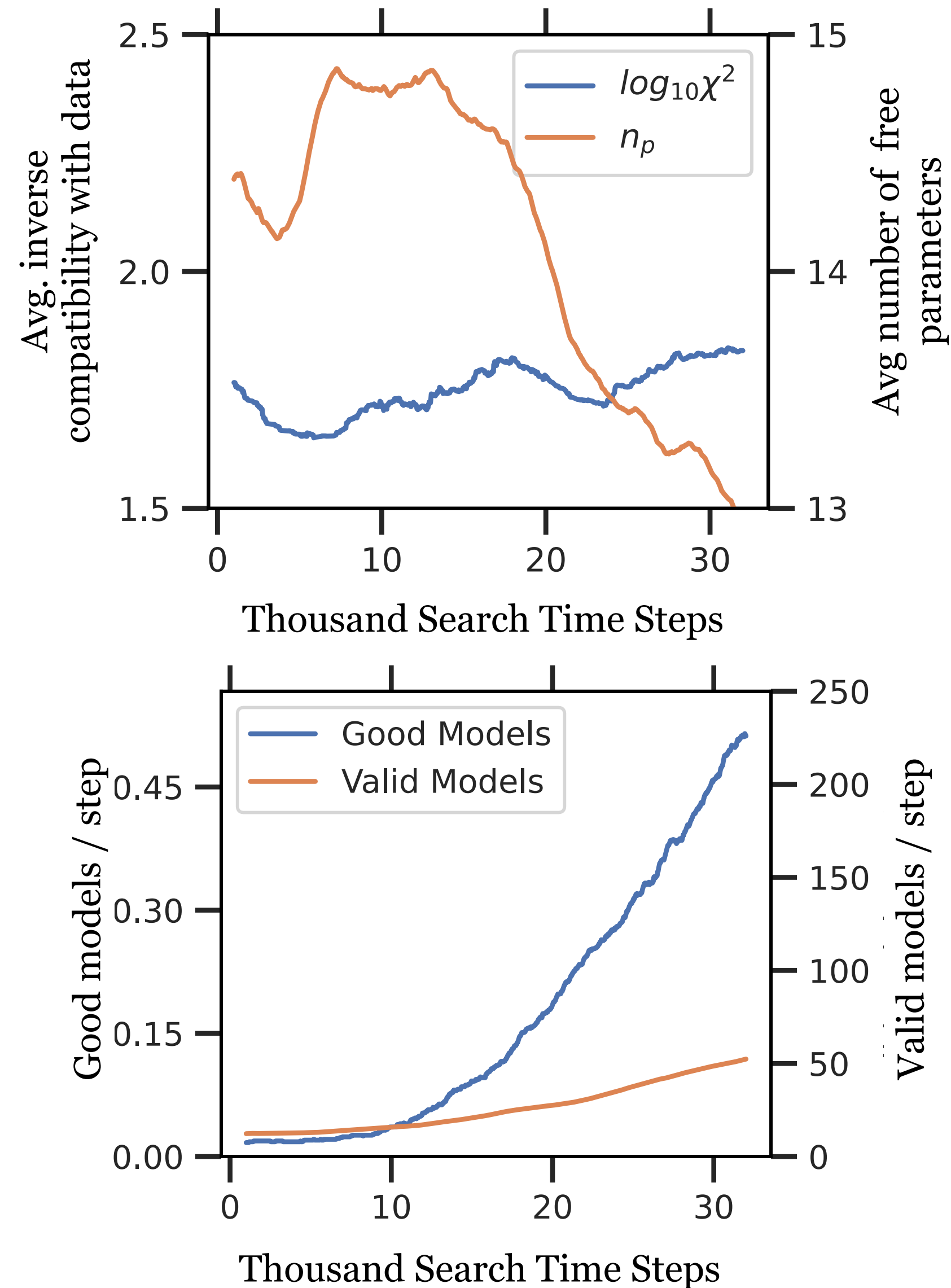
Searching in the most well-studied space of theories: $A_4 \times Z_4$

Re-discovers models of same form as human-designed models

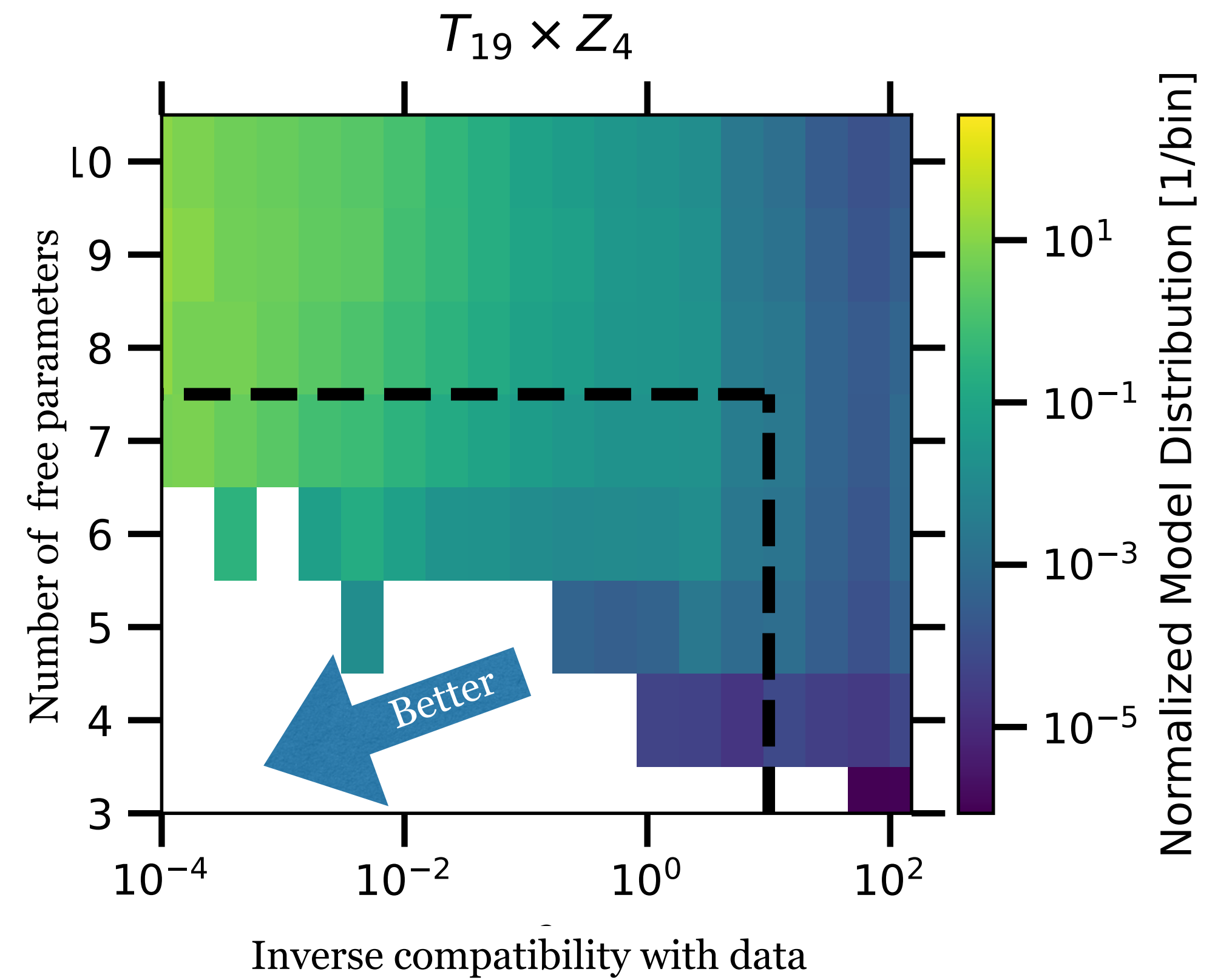
How about exploring new symmetry never previously studied by physicists?

$$T_{19} = \mathbb{Z}_{19} \rtimes \mathbb{Z}_3$$

A new symmetry never before studied for neutrino model building



**Rich landscape of viable neutrino theories,
never previously studied!**



More details studies for these new theory models also performed

Curious what an example new theory look like?

3 Lepton singlets

Lepton doublets (L_1, L_2, L_3)

Right handed neutrinos

Higgs doublet

5 Flavons ϕ_i

	Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Groups	T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
	$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Charges

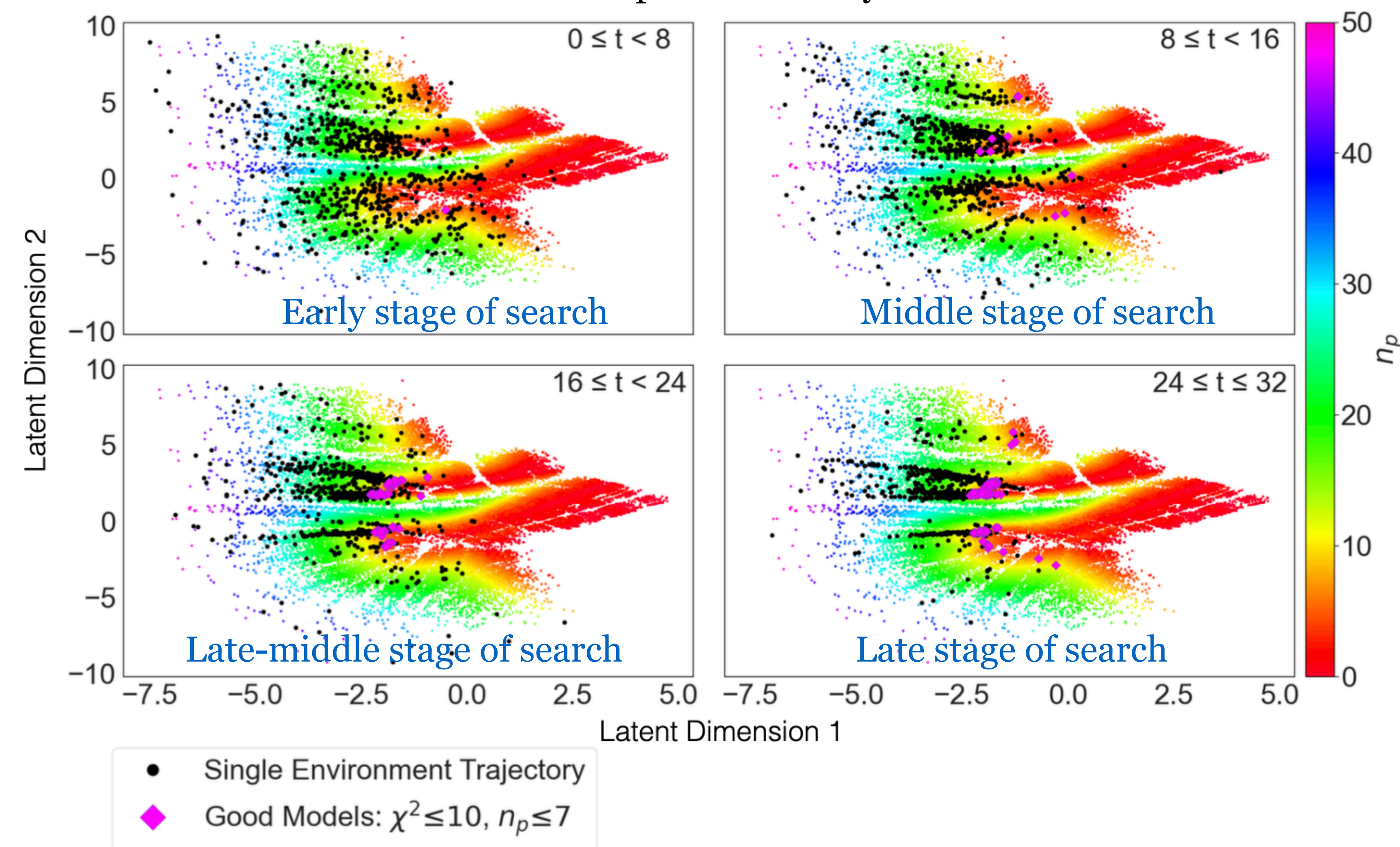
Irreducible representation assigned to fields

LLMs (eg. GPT5) fail at such search tasks!

Excitement in flavour physics community,
including from veterans of the field
([arXiv:2507.19278](#), [arXiv:2510.25495](#))

Deep dive into what the agent is doing over time

Visualised in 2D space created by an auto-encoder



LLMs (eg. GPT5) fail at this!

Black dots are search trajectory for one agent in one environment

Pink are 'good models'

Explores widely in the beginning, but realises that zero-parameter models are useless.

Focuses on few parameter models towards the end