Overcoming challenges of quantum interference at LHC with neural simulation-based inference and a full implementation in ATLAS

Aishik Ghosh

CERN Data Science Seminar
04 December 2024

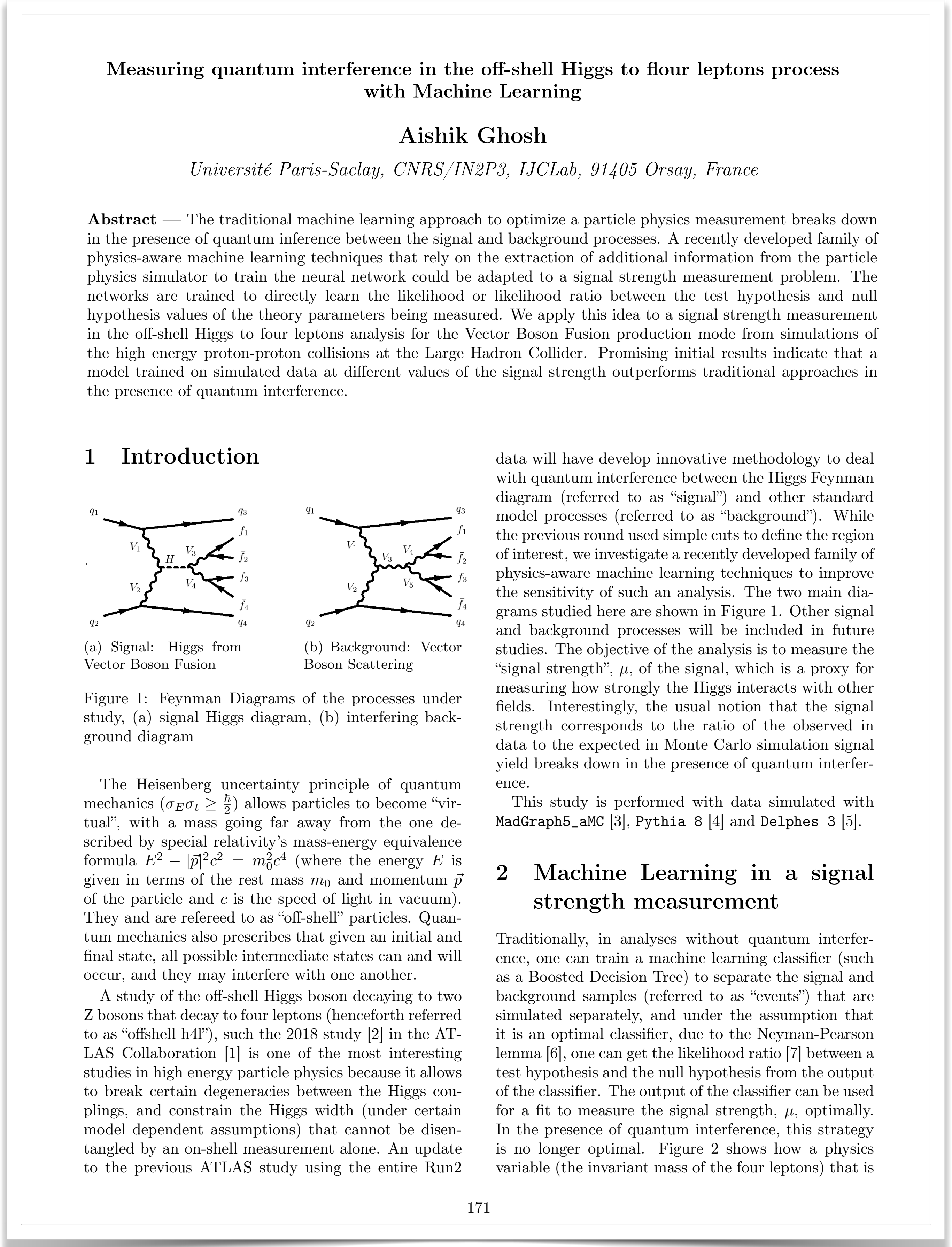
aishikghosh.bsky.social
@Aishik_Ghosh_

What's to come

Statistical inference methods developed for LHC analyses
Option to follow technical details or intuitive explanations

What’s to come

Statistical inference methods developed for LHC analyses
Option to follow technical details or intuitive explanations



Measuring quantum interference in the off-shell Higgs to flour leptons process with Machine Learning

Aishik Ghosh

Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France

Abstract — The traditional machine learning approach to optimize a particle physics measurement breaks down in the presence of quantum inference between the signal and background processes. A recently developed family of physics-aware machine learning techniques that rely on the extraction of additional information from the particle physics simulator to train the neural network could be adapted to a signal strength measurement problem. The networks are trained to directly learn the likelihood or likelihood ratio between the test hypothesis and null hypothesis values of the theory parameters being measured. We apply this idea to a signal strength measurement in the off-shell Higgs to four leptons analysis for the Vector Boson Fusion production mode from simulations of the high energy proton-proton collisions at the Large Hadron Collider. Promising initial results indicate that a model trained on simulated data at different values of the signal strength outperforms traditional approaches in the presence of quantum interference.

1 Introduction

(a) Signal: Higgs from Vector Boson Fusion (b) Background: Vector Boson Scattering

Figure 1: Feynman Diagrams of the processes under study, (a) signal Higgs diagram, (b) interfering background diagram

The Heisenberg uncertainty principle of quantum mechanics ($\sigma_E \sigma_t \geq \frac{\hbar}{2}$) allows particles to become “virtual”, with a mass going far away from the one described by special relativity’s mass-energy equivalence formula $E^2 - |\vec{p}|^2 c^2 = m_0^2 c^4$ (where the energy E is given in terms of the rest mass m_0 and momentum $\vec{p}$ of the particle and c is the speed of light in vacuum). They and are refereed to as “off-shell” particles. Quantum mechanics also prescribes that given an initial and final state, all possible intermediate states can and will occur, and they may interfere with one another.

A study of the off-shell Higgs boson decaying to two Z bosons that decay to four leptons (henceforth referred to as “offshell h4l”), such the 2018 study [2] in the ATLAS Collaboration [1] is one of the most interesting studies in high energy particle physics because it allows to break certain degeneracies between the Higgs couplings, and constrain the Higgs width (under certain model dependent assumptions) that cannot be disentangled by an on-shell measurement alone. An update to the previous ATLAS study using the entire Run2 data will have develop innovative methodology to deal with quantum interference between the Higgs Feynman diagram (referred to as “signal”) and other standard model processes (referred to as “background”). While the previous round used simple cuts to define the region of interest, we investigate a recently developed family of physics-aware machine learning techniques to improve the sensitivity of such an analysis. The two main diagrams studied here are shown in Figure 1. Other signal and background processes will be included in future studies. The objective of the analysis is to measure the “signal strength”, μ, of the signal, which is a proxy for measuring how strongly the Higgs interacts with other fields. Interestingly, the usual notion that the signal strength corresponds to the ratio of the observed in data to the expected in Monte Carlo simulation signal yield breaks down in the presence of quantum interference.

This study is performed with data simulated with `MadGraph5_aMC` [3], `Pythia 8` [4] and `Delphes 3` [5].

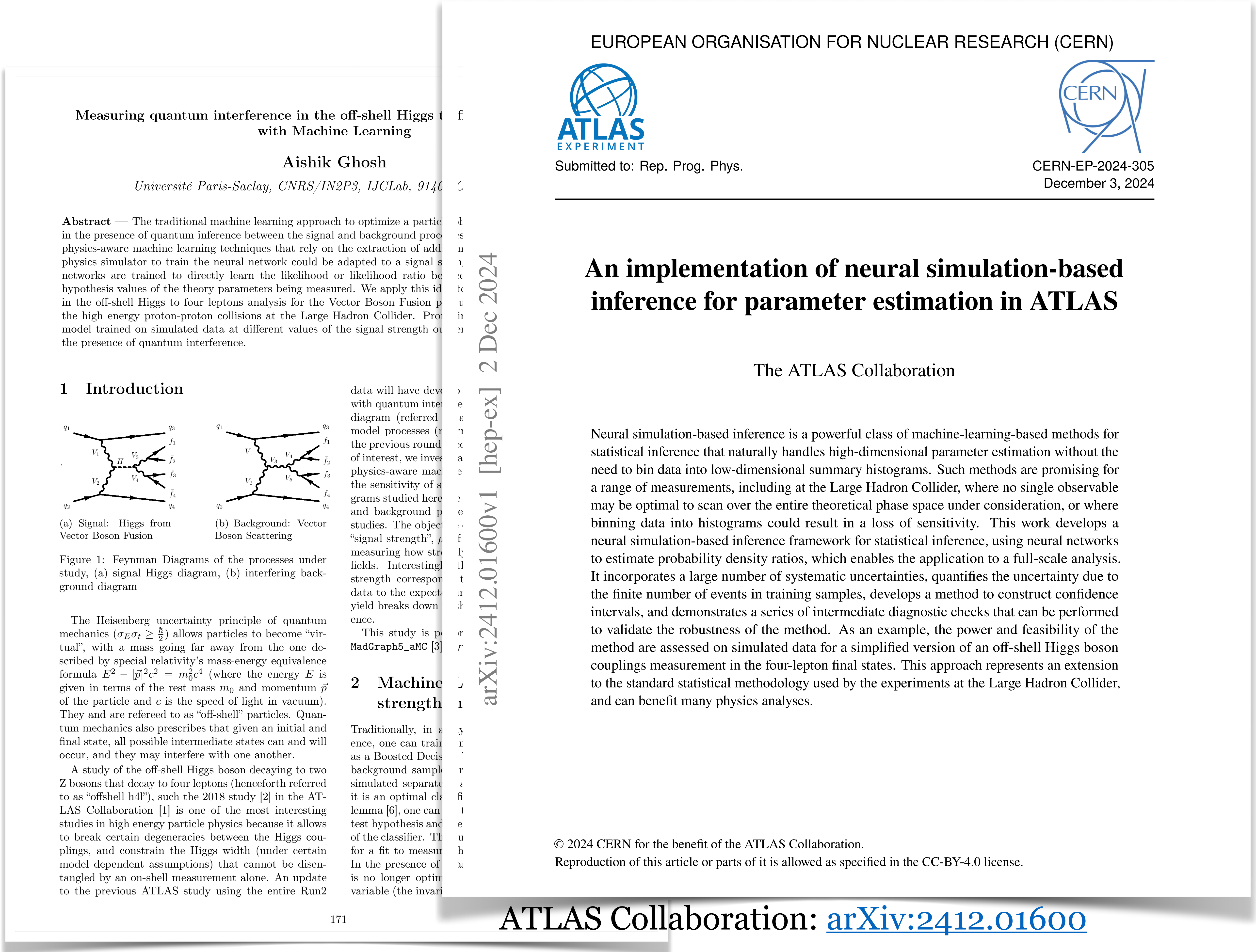
2 Machine Learning in a signal strength measurement

Traditionally, in analyses without quantum interference, one can train a machine learning classifier (such as a Boosted Decision Tree) to separate the signal and background samples (referred to as “events”) that are simulated separately, and under the assumption that it is an optimal classifier, due to the Neyman-Pearson lemma [6], one can get the likelihood ratio [7] between a test hypothesis and the null hypothesis from the output of the classifier. The output of the classifier can be used for a fit to measure the signal strength, μ, optimally. In the presence of quantum interference, this strategy is no longer optimal. Figure 2 shows how a physics variable (the invariant mass of the four leptons) that is

Ghosh et al:
hal-02971995(p172)

What's to come

Statistical inference methods developed for LHC analyses
Option to follow technical details or intuitive explanations



Measuring quantum interference in the off-shell Higgs [...] with Machine Learning

Aishik Ghosh

Université Paris-Saclay, CNRS/IN2P3, IJCLab, 914[...]

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Rep. Prog. Phys.

CERN-EP-2024-305
December 3, 2024

arXiv:2412.01600v1 [hep-ex] 2 Dec 2024

An implementation of neural simulation-based inference for parameter estimation in ATLAS

The ATLAS Collaboration

Neural simulation-based inference is a powerful class of machine-learning-based methods for statistical inference that naturally handles high-dimensional parameter estimation without the need to bin data into low-dimensional summary histograms. Such methods are promising for a range of measurements, including at the Large Hadron Collider, where no single observable may be optimal to scan over the entire theoretical phase space under consideration, or where binning data into histograms could result in a loss of sensitivity. This work develops a neural simulation-based inference framework for statistical inference, using neural networks to estimate probability density ratios, which enables the application to a full-scale analysis. It incorporates a large number of systematic uncertainties, quantifies the uncertainty due to the finite number of events in training samples, develops a method to construct confidence intervals, and demonstrates a series of intermediate diagnostic checks that can be performed to validate the robustness of the method. As an example, the power and feasibility of the method are assessed on simulated data for a simplified version of an off-shell Higgs boson couplings measurement in the four-lepton final states. This approach represents an extension to the standard statistical methodology used by the experiments at the Large Hadron Collider, and can benefit many physics analyses.

© 2024 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

Ghosh et al: hal-02971995(p172)

ATLAS Collaboration: arXiv:2412.01600

What’s to come

Statistical inference methods developed for LHC analyses
Option to follow technical details or intuitive explanations

Measuring quantum interference in the off-shell Higgs
with Machine Learning

Aishik Ghosh

Université Paris-Saclay, CNRS/IN2P3, IJCLab, 914

1 Introduction

EUROPEAN ORGANISATION FOR NUCLEAR R

Submitted to: Rep. Prog. Phys.

arXiv:2412.01600v1 [hep-ex] 2 Dec 2024

An implementation of neural sim
inference for parameter estimati

The ATLAS Collaboration

© 2024 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Rep. Prog. Phys.

CERN-EP-2024-298
December 3, 2024

arXiv:2412.01548v1 [hep-ex] 2 Dec 2024

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

The ATLAS Collaboration

A measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel is presented. The measurement uses 140 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV collected by the ATLAS detector at the Large Hadron Collider and supersedes the previous result in this decay channel using the same dataset. The data analysis is performed using a neural simulation-based inference method, which builds per-event likelihood ratios using neural networks. The observed (expected) off-shell Higgs boson production signal strength in the $ZZ \to 4\ell$ decay channel at 68% CL is $0.87^{+0.75}_{-0.54}$ ($1.00^{+1.04}_{-0.95}$). The evidence for off-shell Higgs boson production using the $ZZ \to 4\ell$ decay channel has an observed (expected) significance of 2.5σ (1.3σ). The expected result represents a significant improvement relative to that of the previous analysis of the same dataset, which obtained an expected significance of 0.5σ. When combined with the most recent ATLAS measurement in the $ZZ \to 2\ell 2\nu$ decay channel, the evidence for off-shell Higgs boson production has an observed (expected) significance of 3.7σ (2.4σ). The off-shell measurements are combined with the measurement of on-shell Higgs boson production to obtain constraints on the Higgs boson total width. The observed (expected) value of the Higgs boson width at 68% CL is $4.3^{+2.7}_{-1.9}$ ($4.1^{+3.5}_{-3.4}$) MeV.

© 2024 CERN for the benefit of the ATLAS Collaboration.
Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

Ghosh et al: [hal-02971995(p172)]

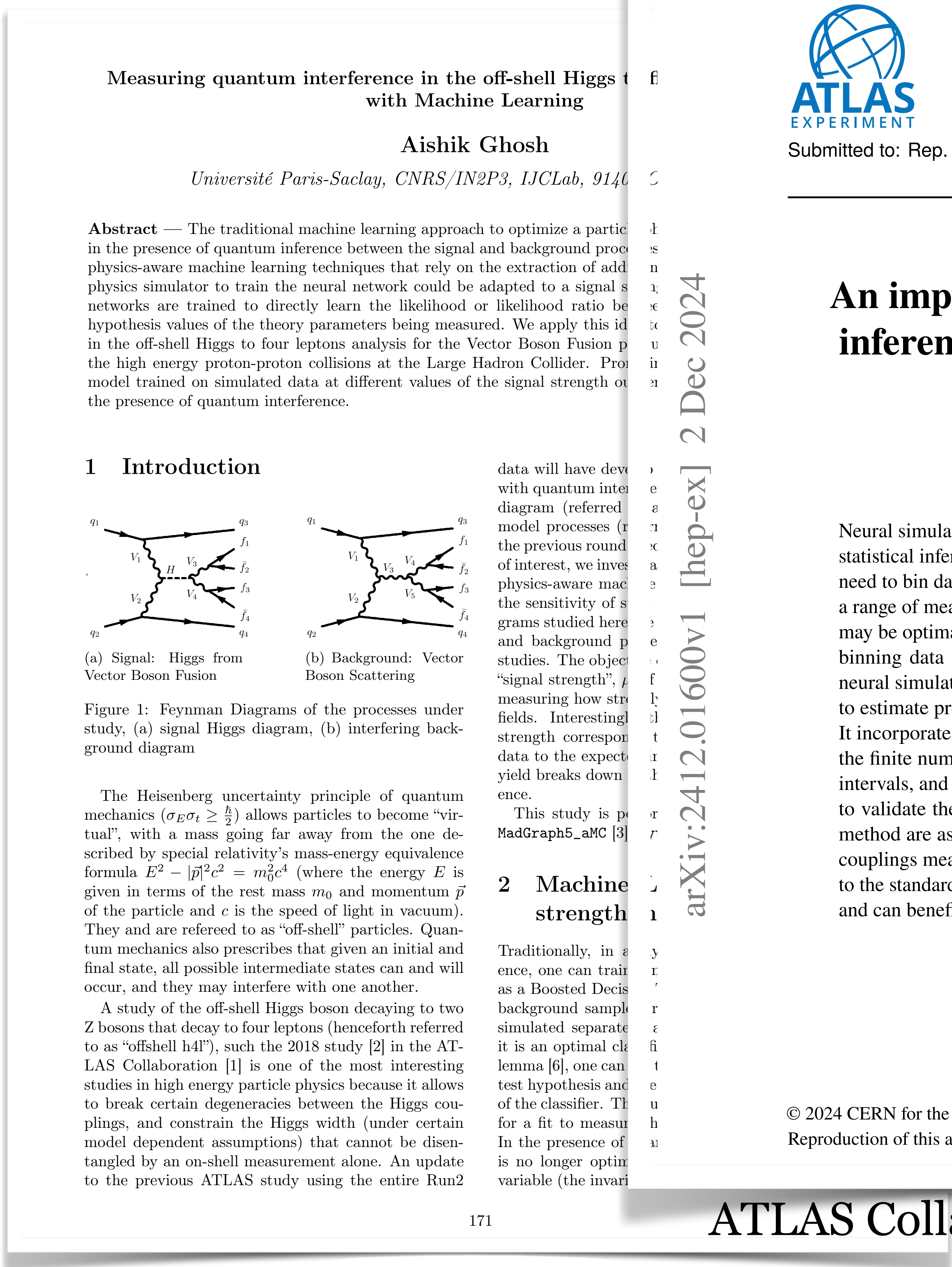
ATLAS Collaboration: [arXiv:2412.01600]

ATLAS Collaboration: [arXiv:2412.01548]

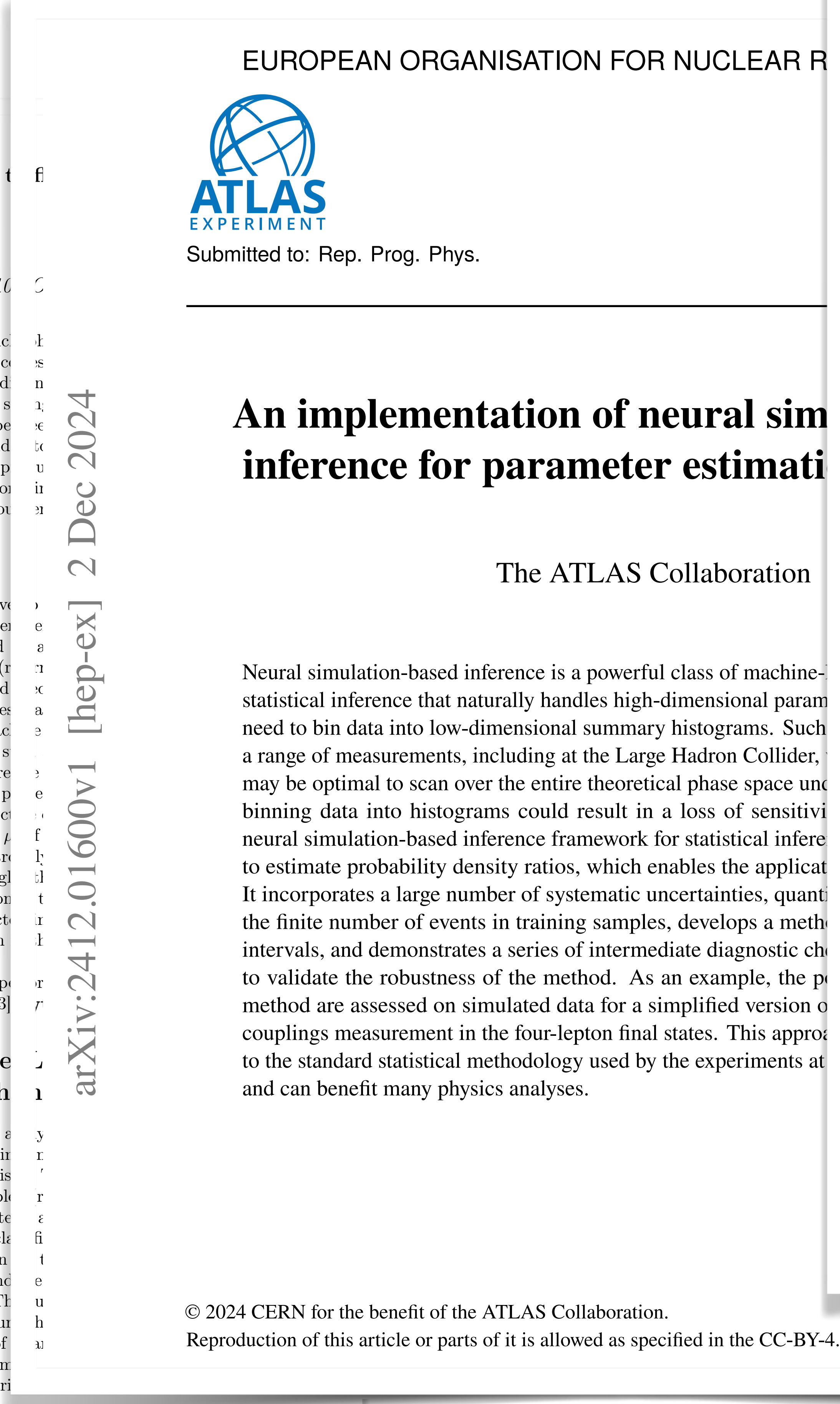
What's to come

Statistical inference methods developed for LHC analyses
Option to follow technical details or intuitive explanations

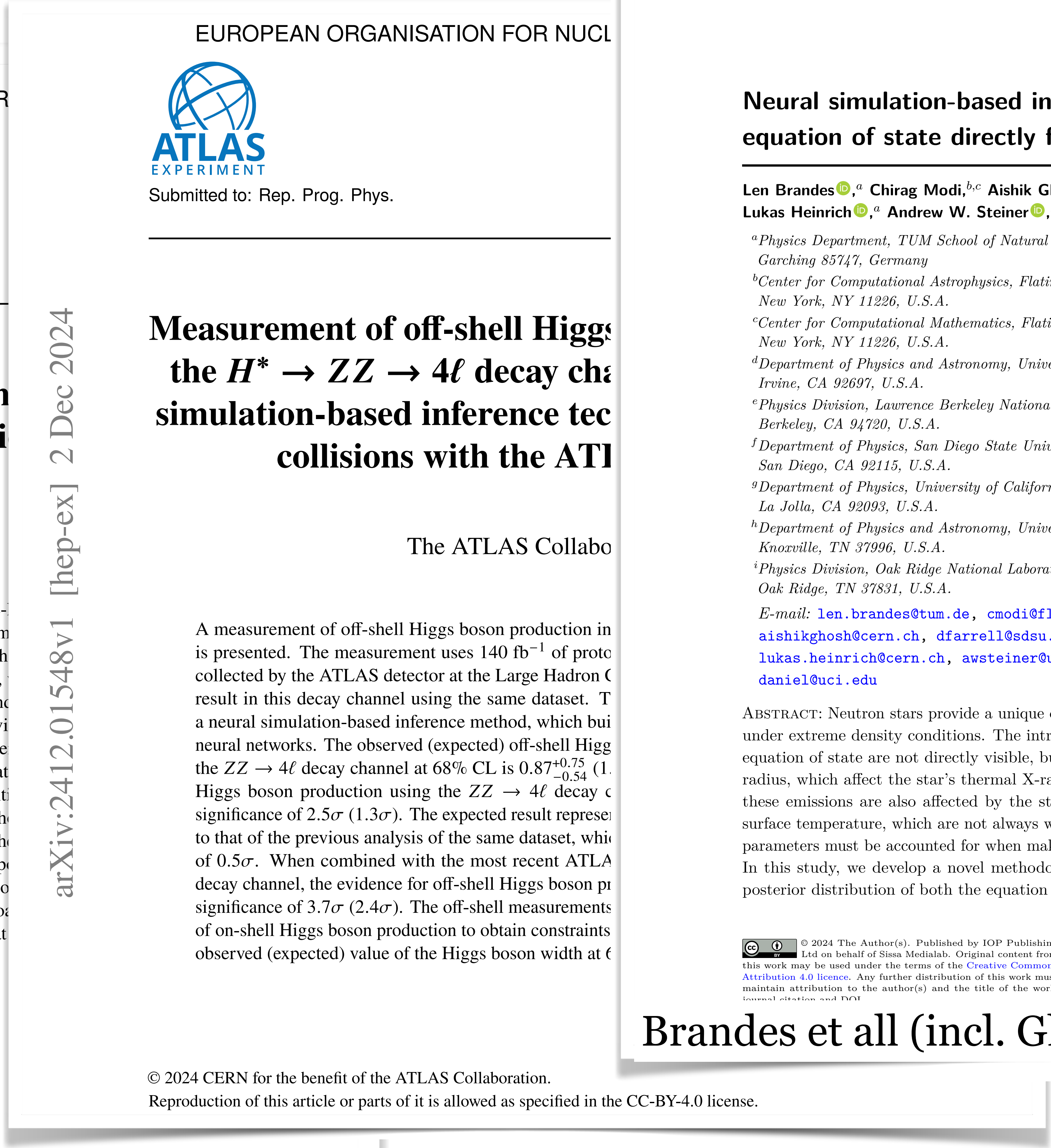
Similar story for neutron star astrophysics



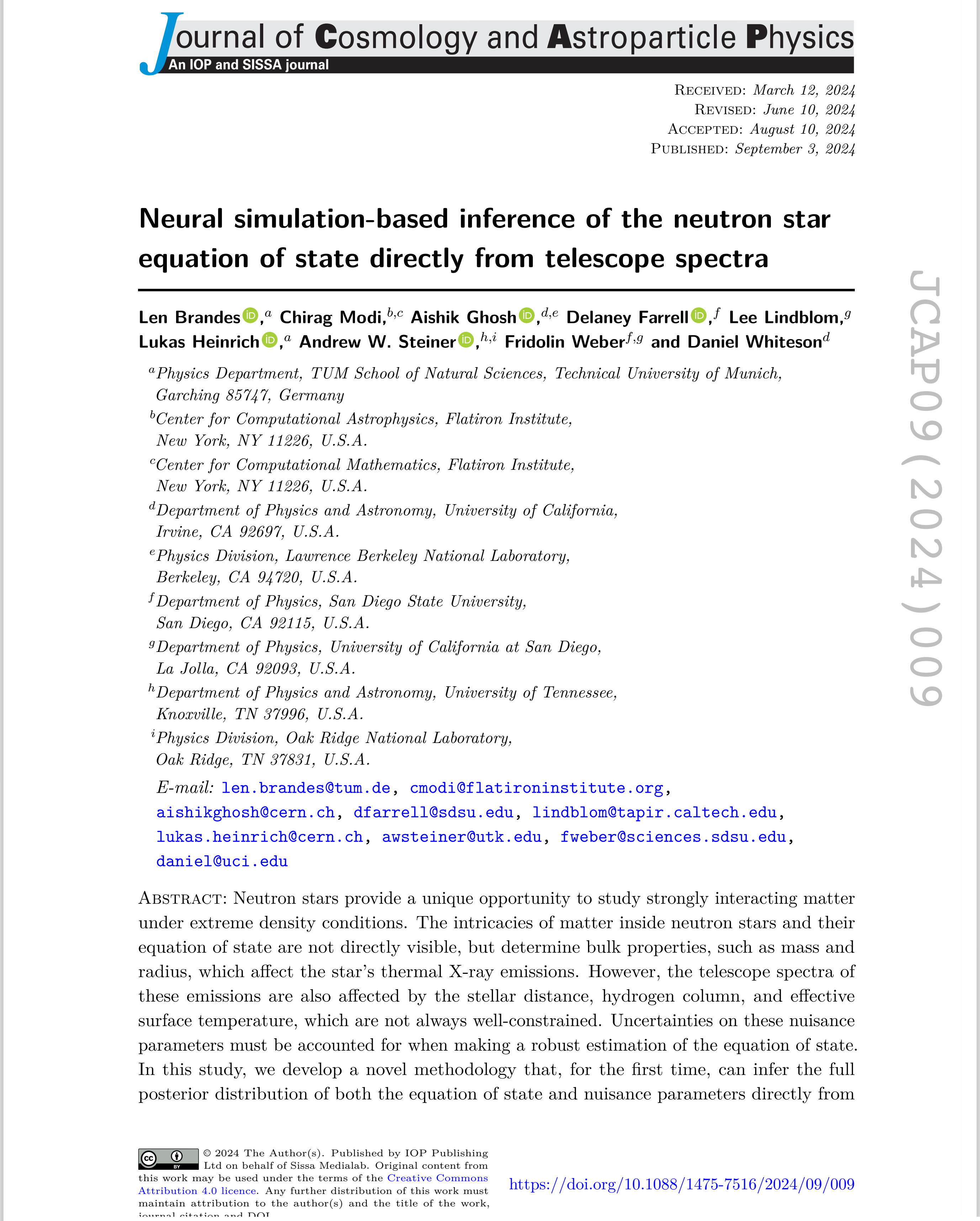
Ghosh et al: hal-02971995(p172)



ATLAS Collaboration: arXiv:2412.01600



ATLAS Collaboration: arXiv:2412.01548



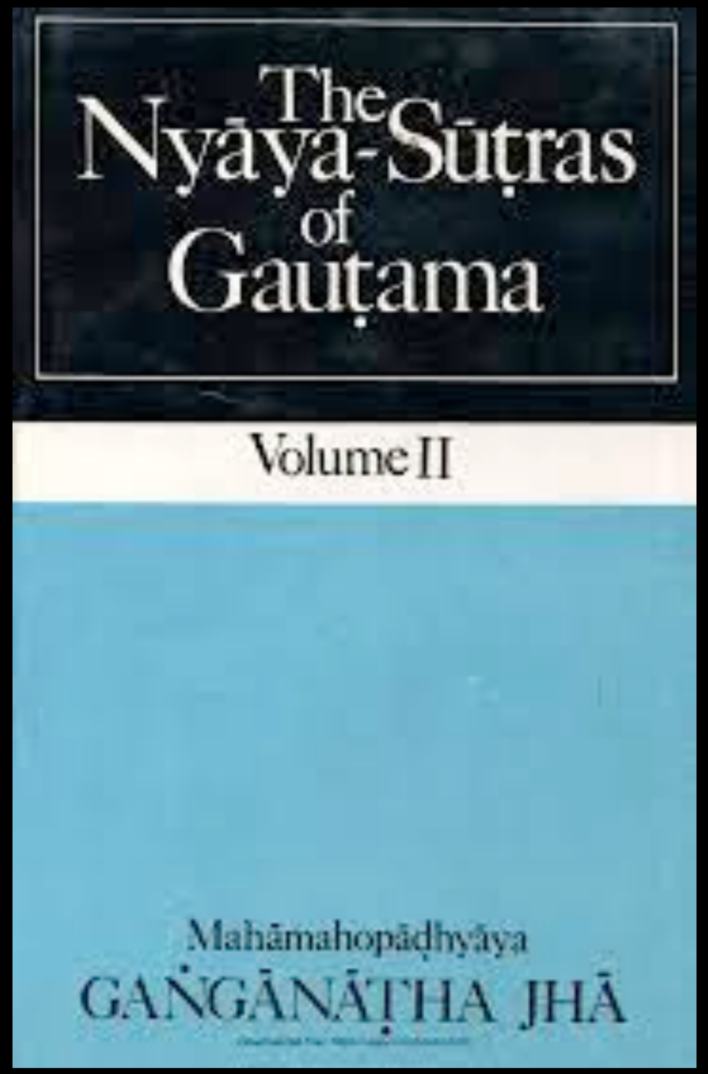
Brandes et all (incl. Ghosh): JCAP 09(2024)009

Some of the oldest questions

Image: DALL-E

What elements make up the universe ?
(5 century BCE)

How sure are we?
Theory of Errors & Empirical Knowledge
(6 century BCE)



Some of the oldest questions

Image: DALL-E

What elements make up the universe ?
(5 century BCE)

How sure are we?
Theory of Errors & Empirical Knowledge
(6 century BCE)

Theorists

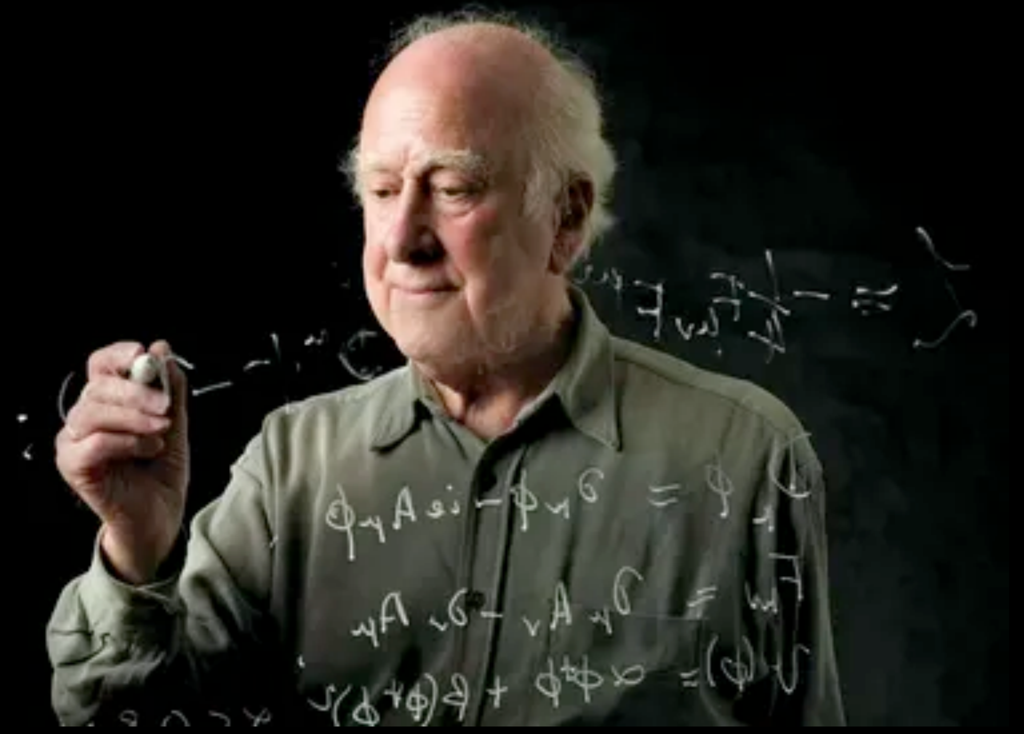
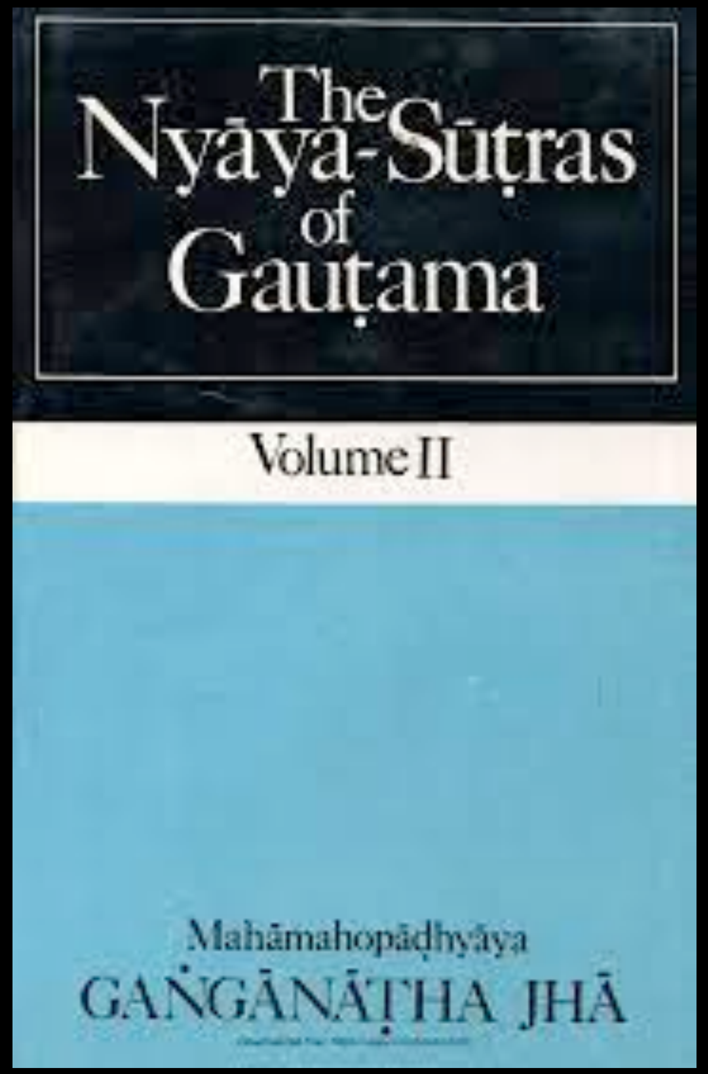


Image: CERN



Some of the oldest questions

Image: DALL-E

What elements make up the universe ?
(5 century BCE)

Theorists

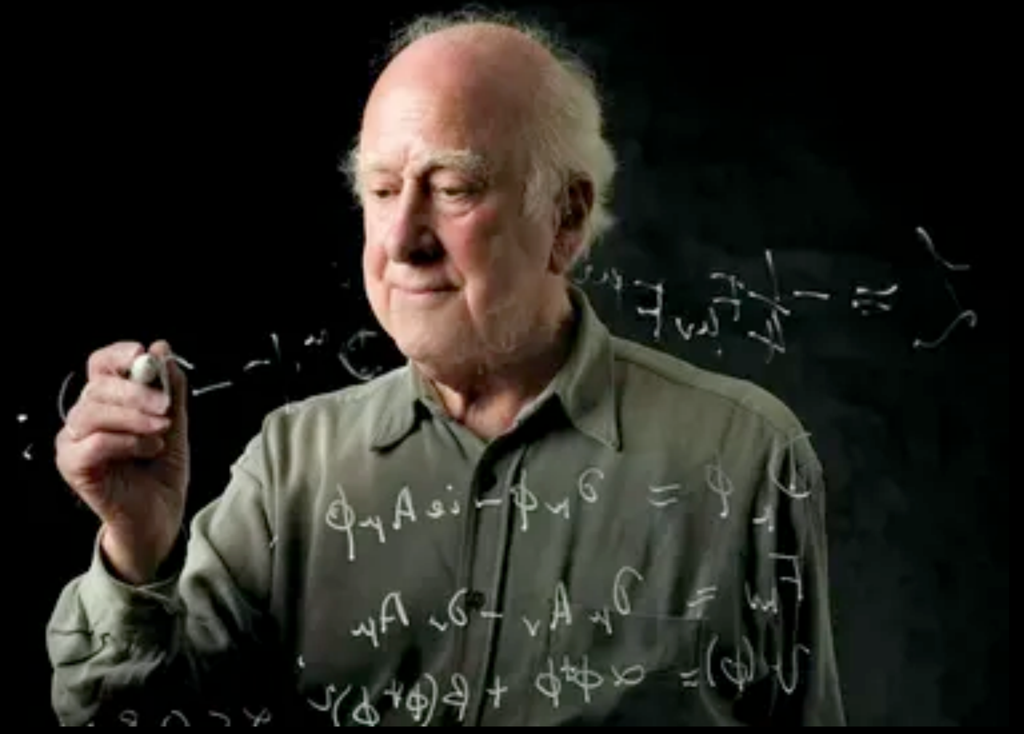
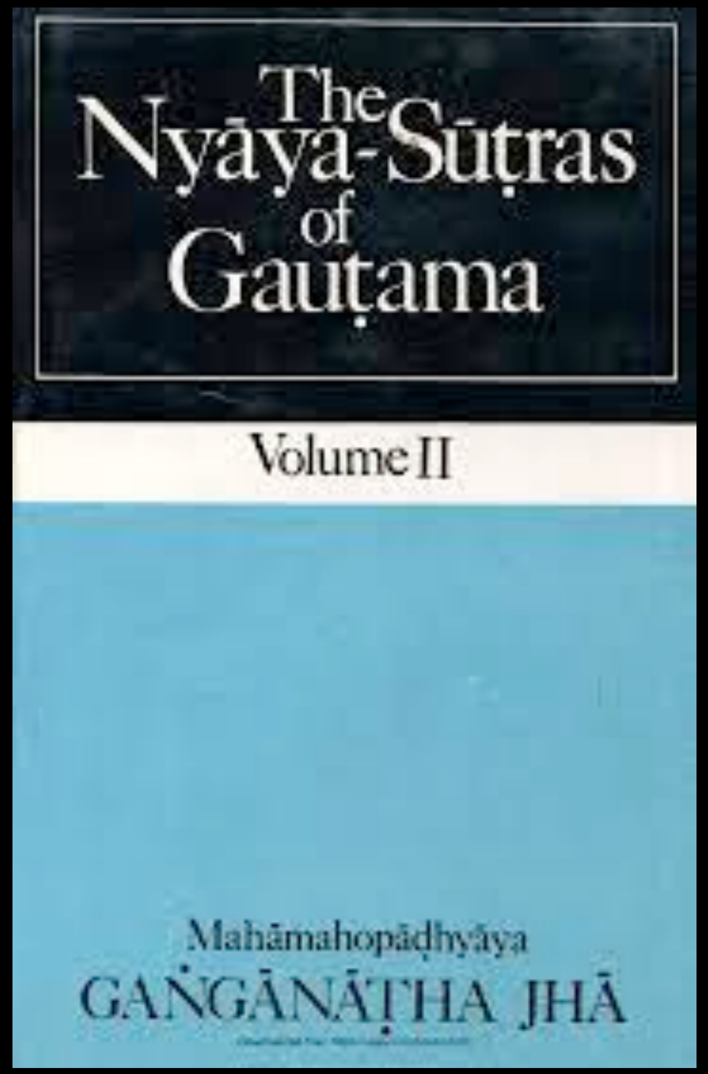


Image: CERN

How sure are we?
Theory of Errors & Empirical Knowledge
(6 century BCE)

Image: CERN

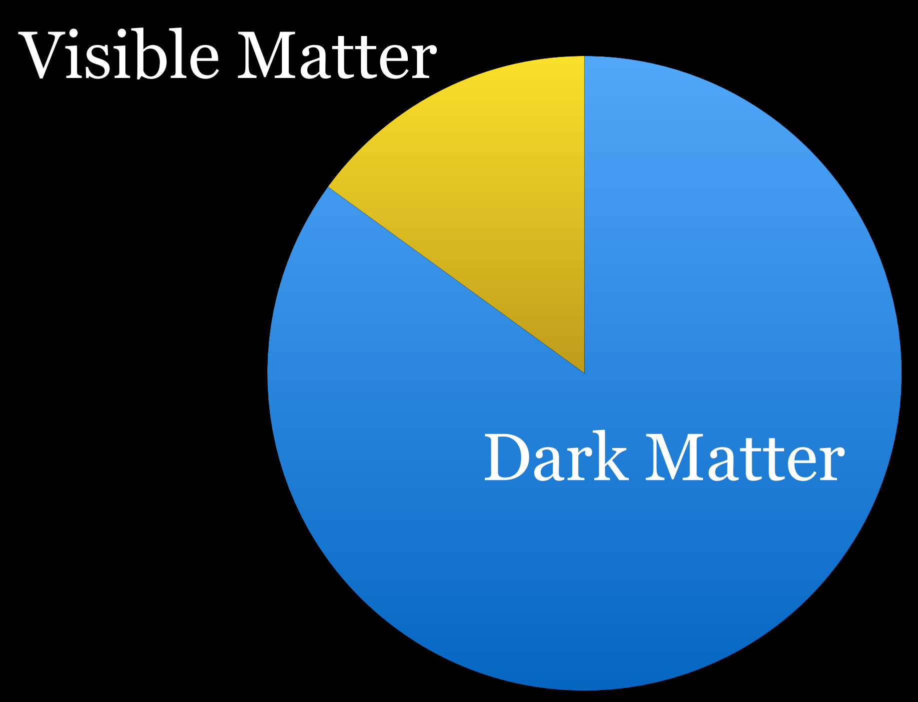
Experimentalists



Questions about the universe ...

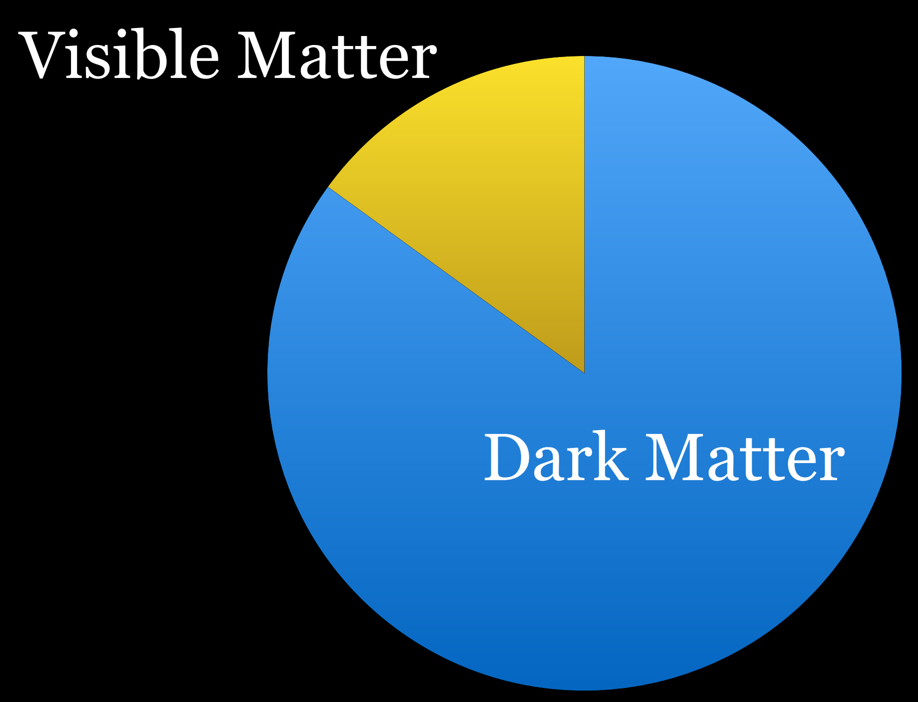
Questions about the universe …

There's so much more dark matter than visible matter in the universe. What is it ?



Questions about the universe …

There's so much more dark matter than visible matter in the universe. What is it ?



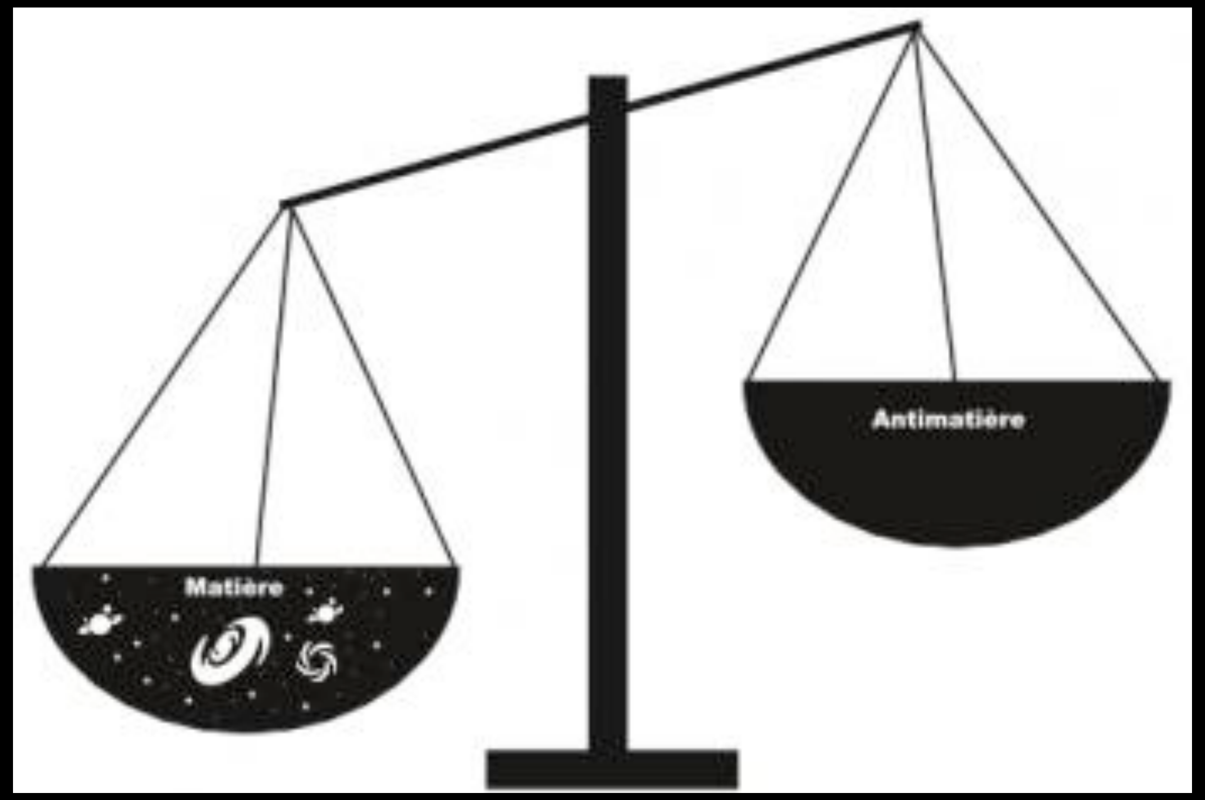
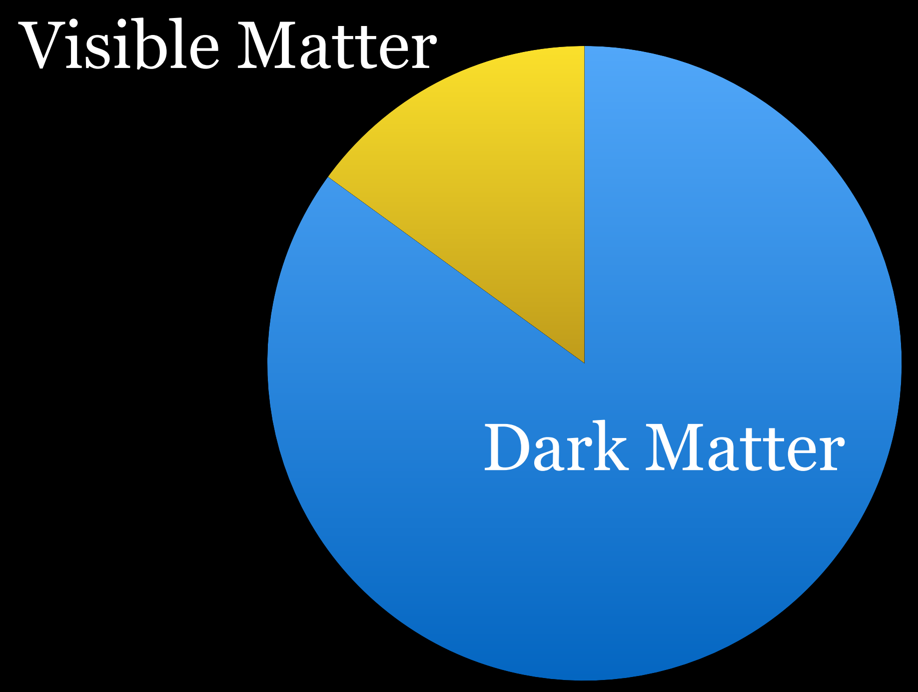


Image: GANIL

Why more matter than anti-matter ?

Questions about the universe …

There's so much more dark matter than visible matter in the universe. What is it ?



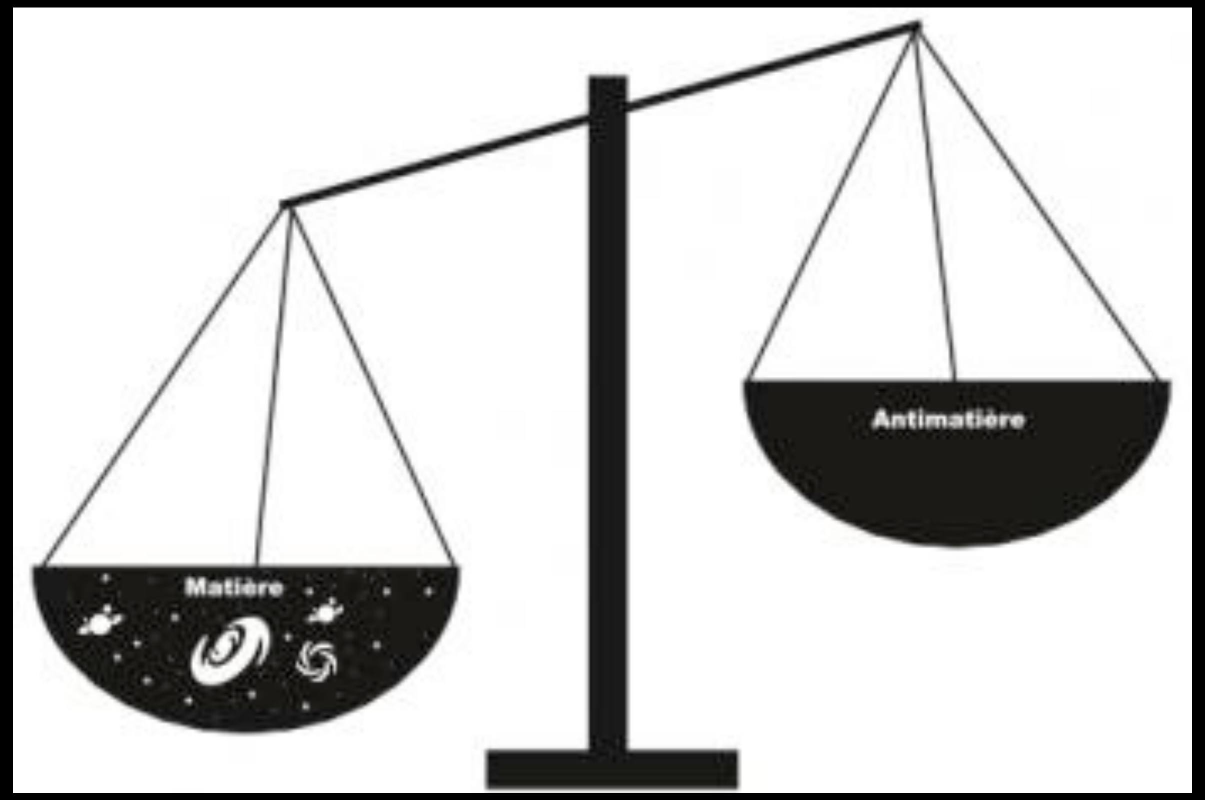


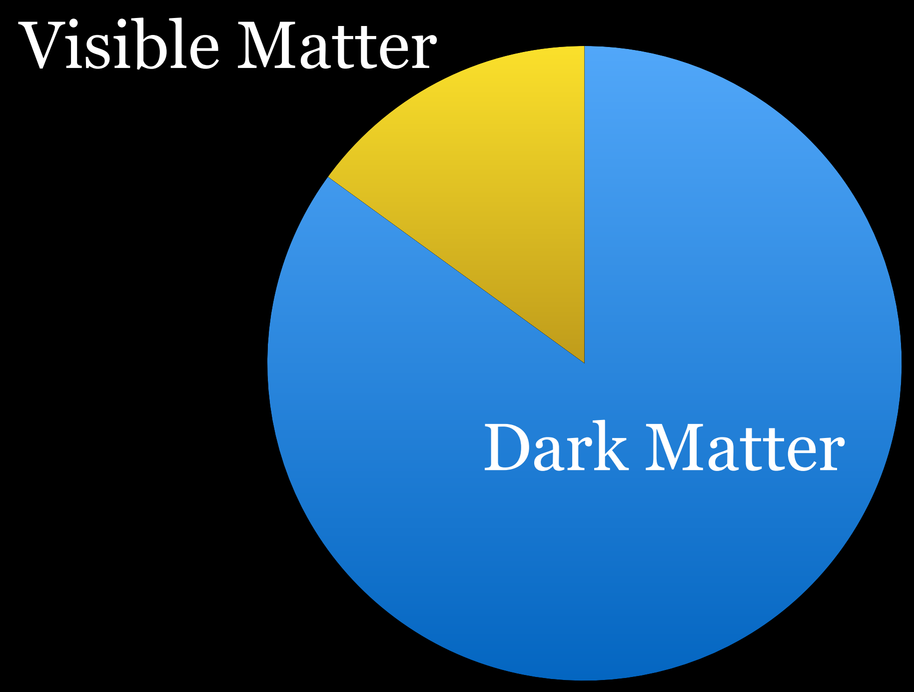
Image: GANIL

Why more matter than anti-matter ?

Are there new forces ?

Questions about the universe …

There's so much more dark matter than visible matter in the universe. What is it ?



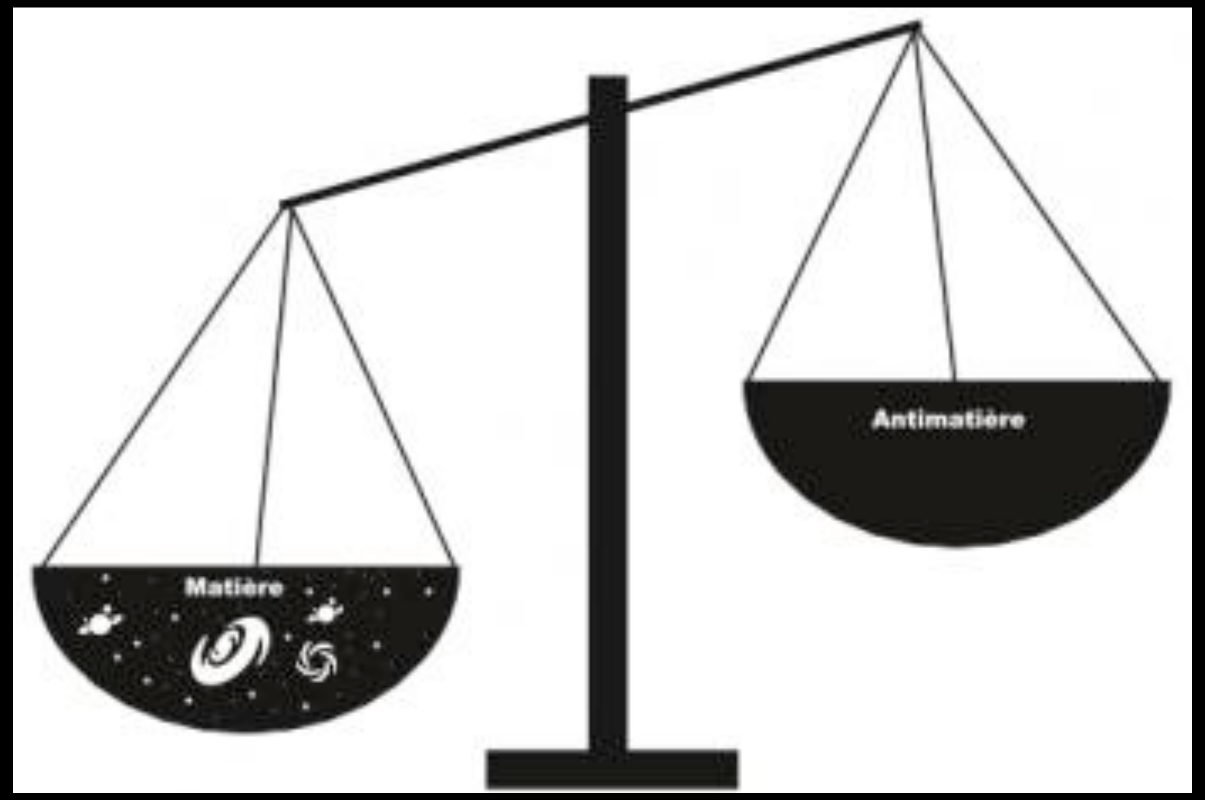


Image: GANIL

Why more matter than anti-matter ?

Are there new forces ?

Image: CERN

New theories often predict new particles yet to be discovered

The most fundamental constituents of matter (that we know of…)

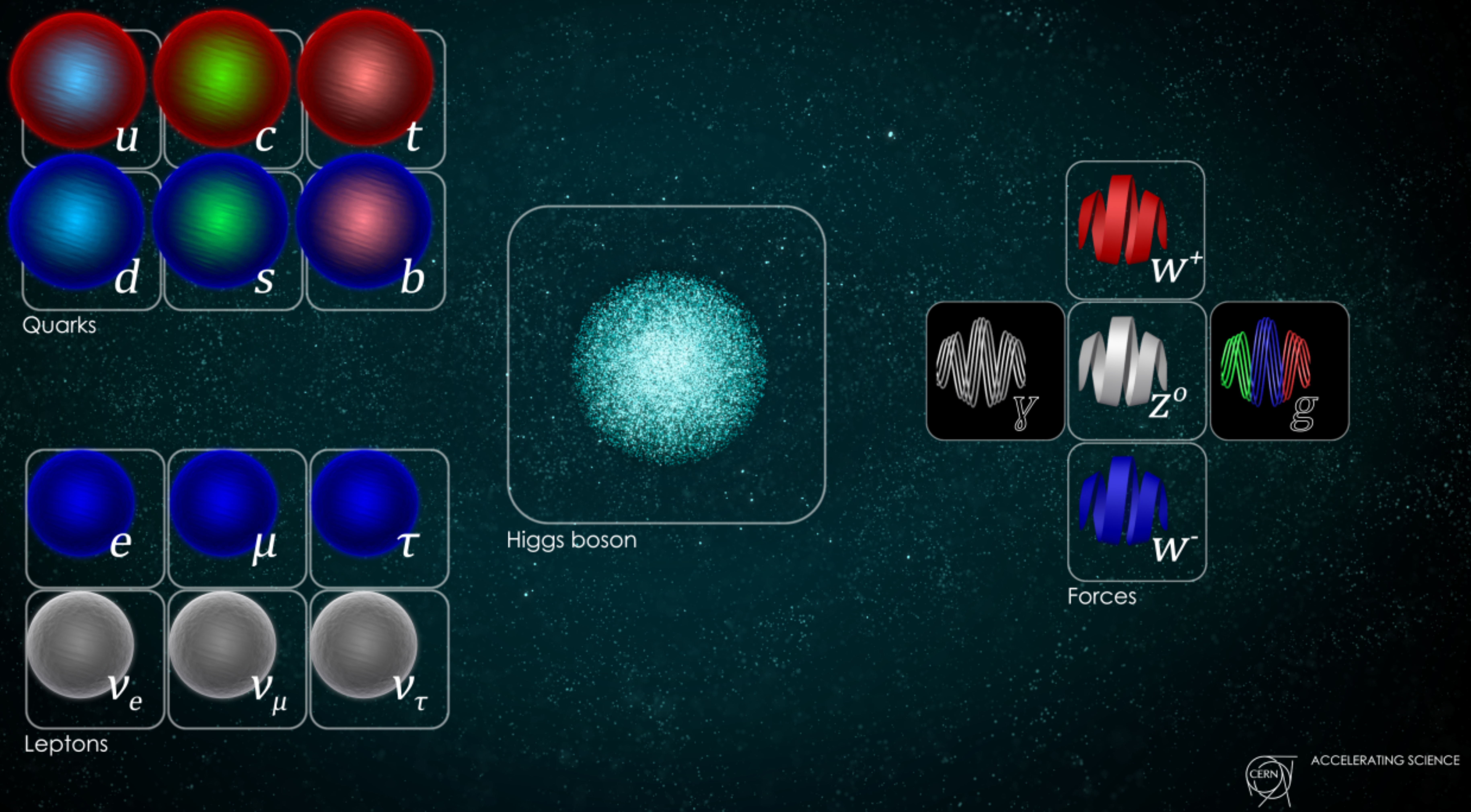
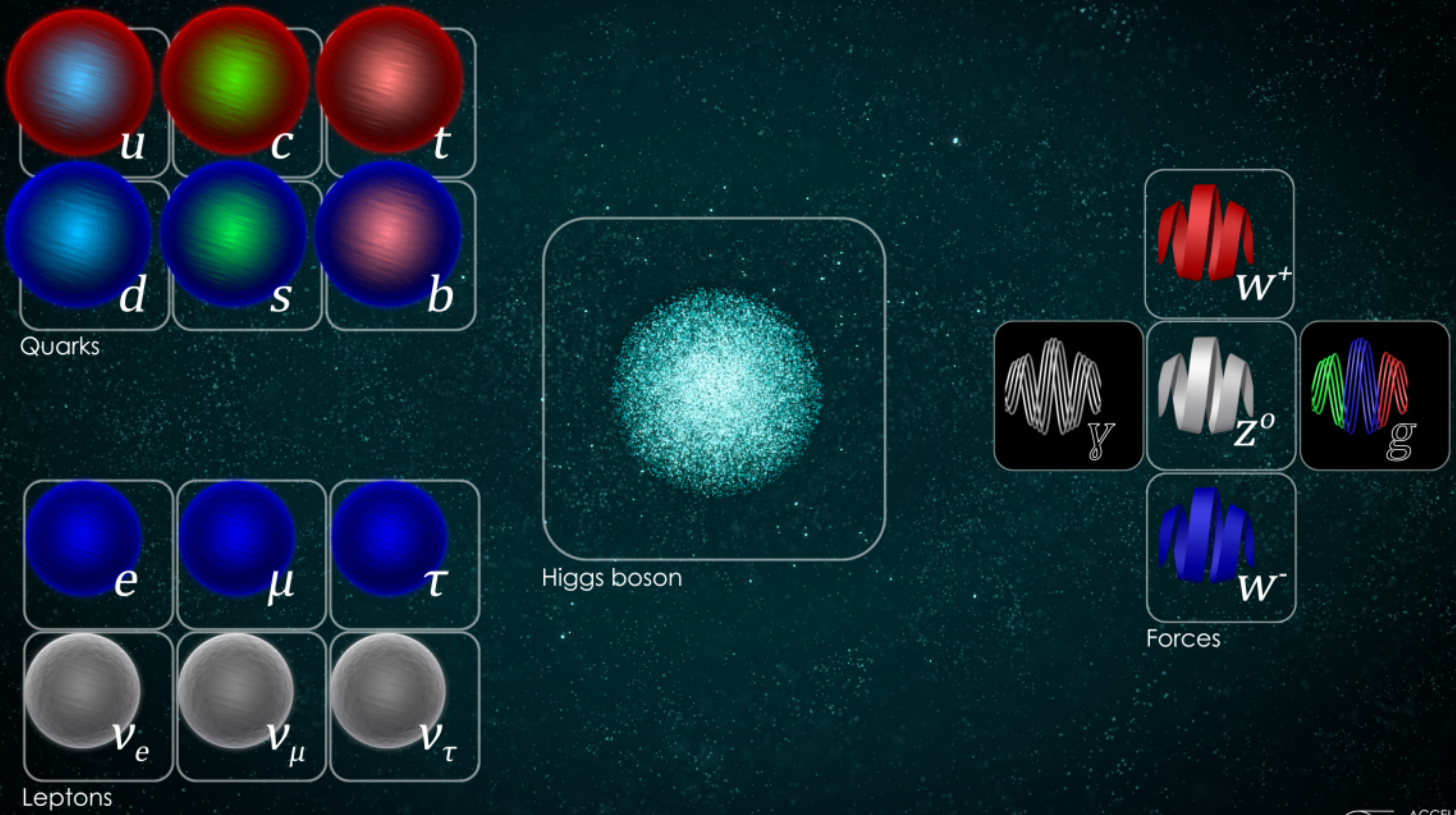


Image: CERN

The most fundamental constituents of matter (that we know of...)

Protons are made up of u and d



ACCELERATING SCIENCE

Image: CERN

The most fundamental constituents of matter (that we know of…)

Protons are made up of u and d

That's the electron and its cousins

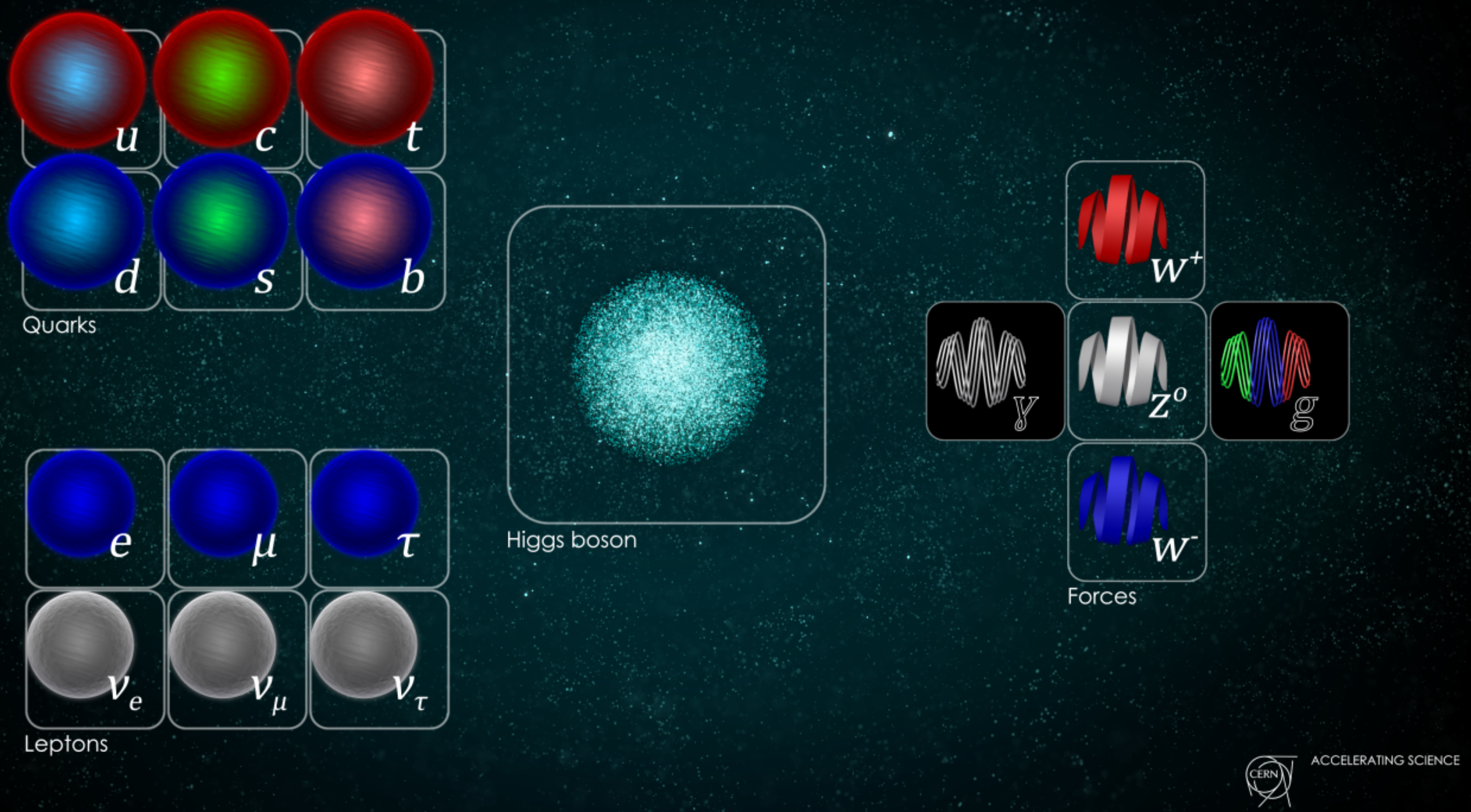
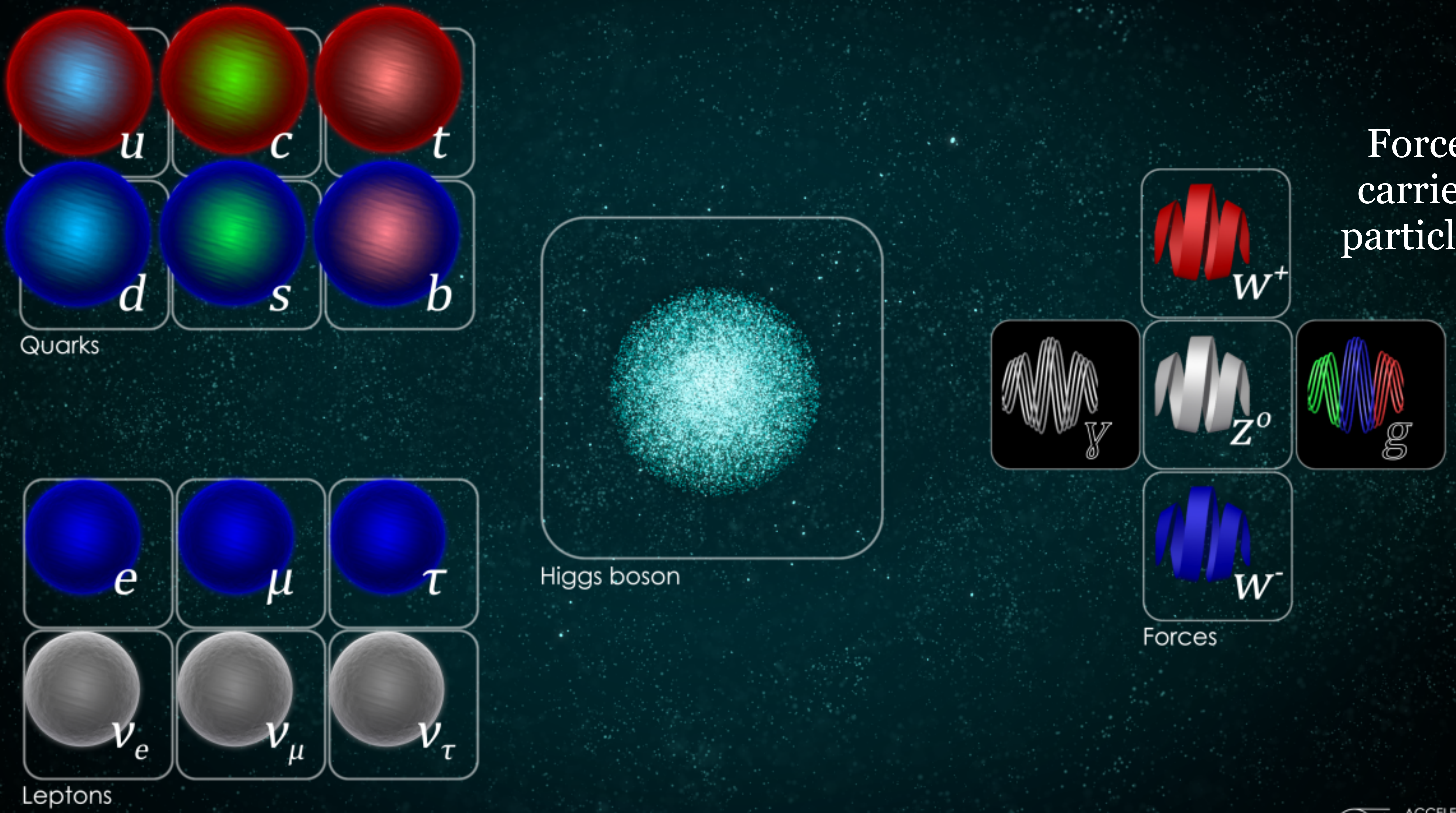


Image: CERN

The most fundamental constituents of matter (that we know of...)

Protons are made up of u and d

That's the electron and its cousins



Force carrier particles

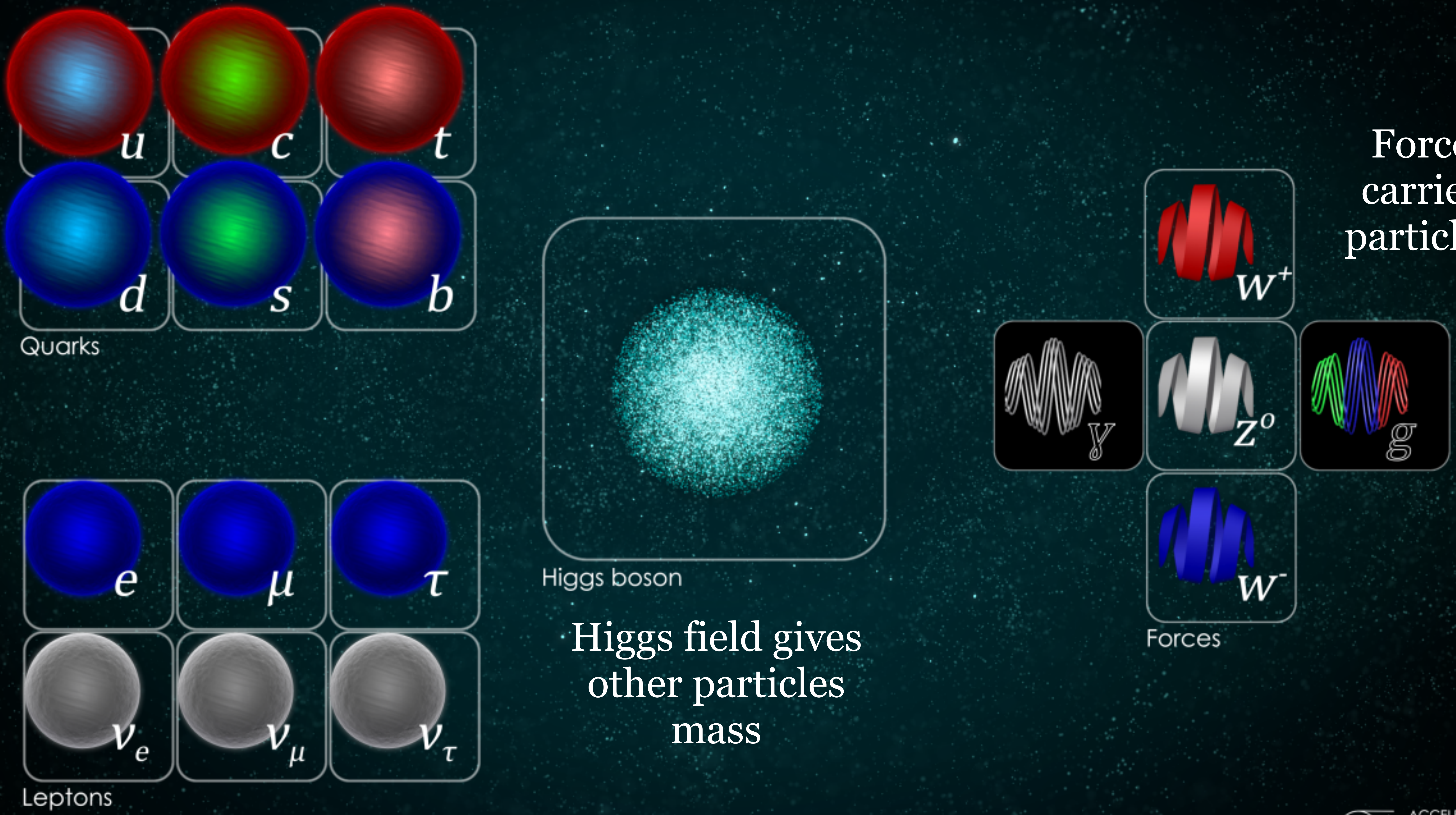
ACCELERATING SCIENCE

Image: CERN

The most fundamental constituents of matter (that we know of...)

Protons are made up of u and d

That's the electron and its cousins



Higgs field gives other particles mass

Force carrier particles

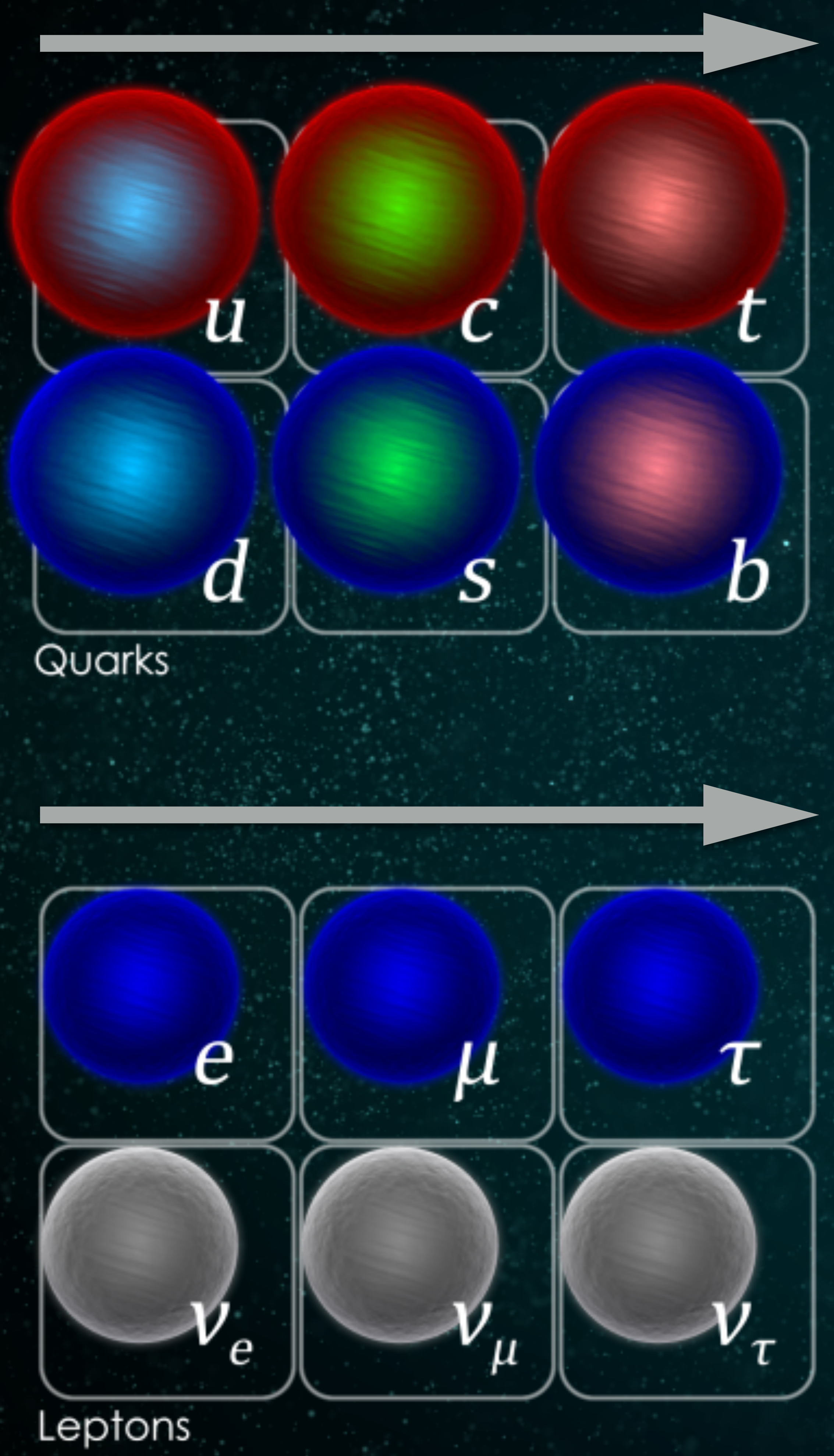
ACCELERATING SCIENCE

Image: CERN

The most fundamental constituents of matter (that we know of...)

Heavier particles, difficult to create

Protons are made up of u and d



That’s the electron and its cousins

Higgs field gives other particles mass

Force carrier particles

ACCELERATING SCIENCE

Image: CERN

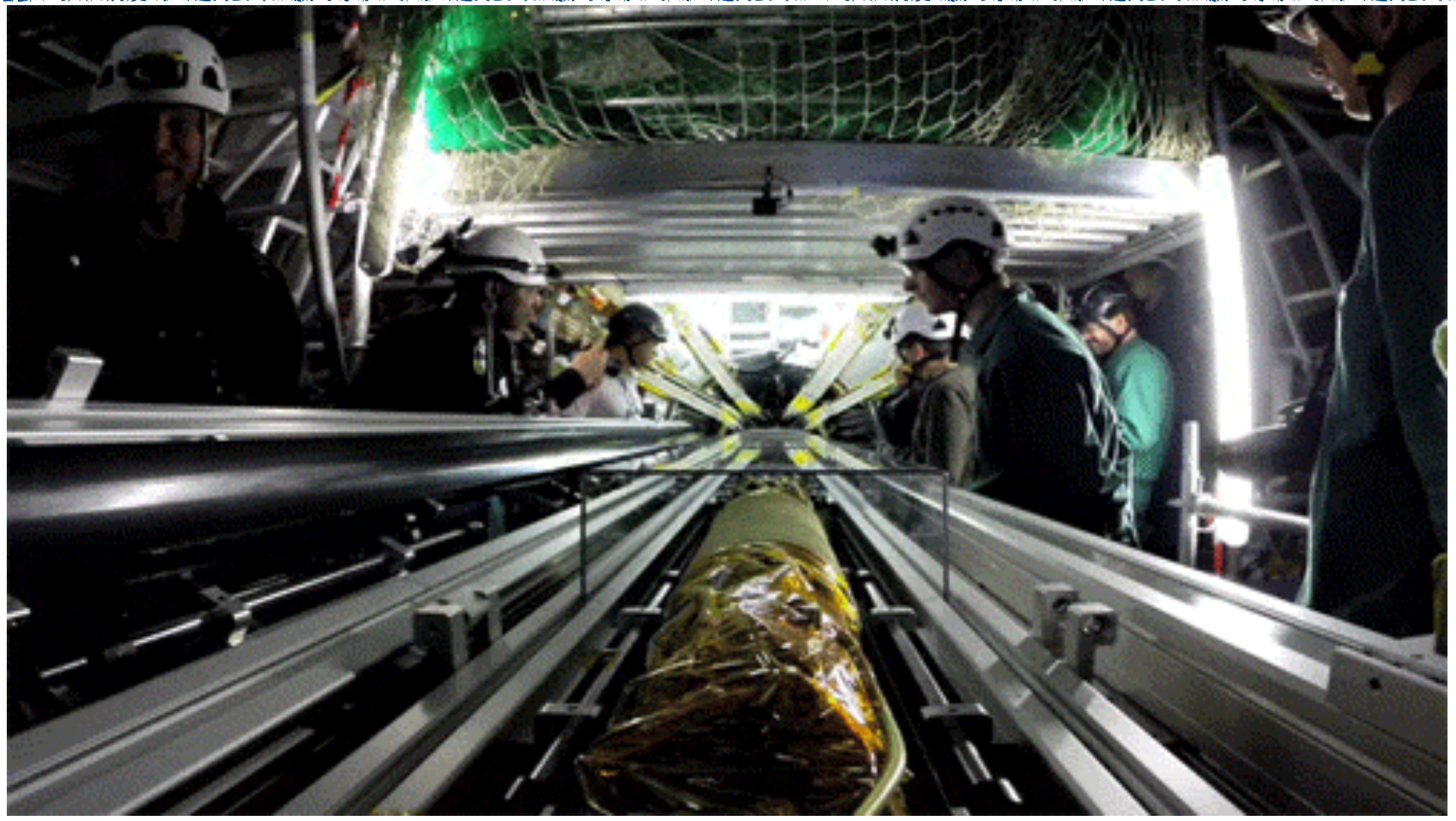
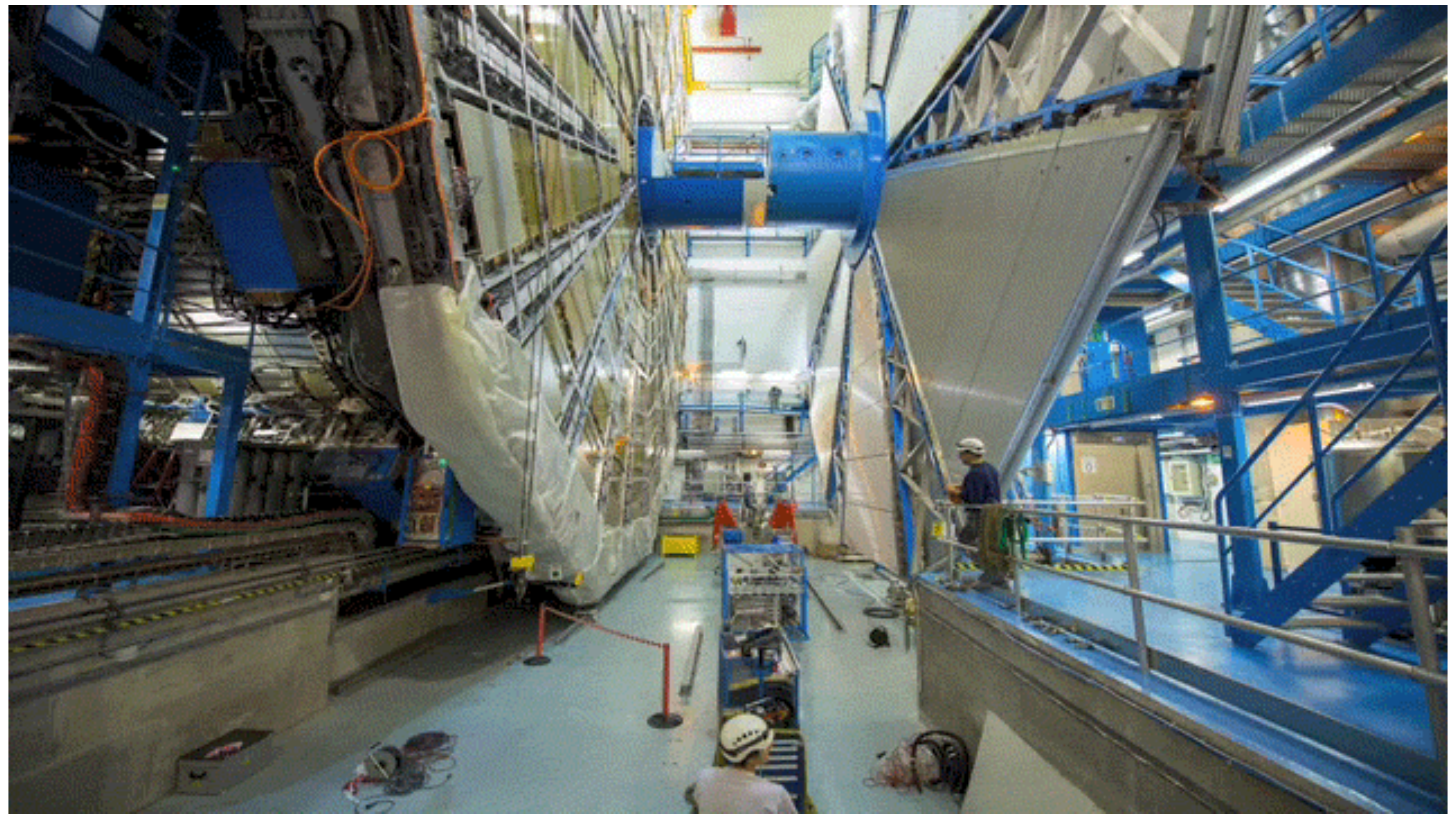
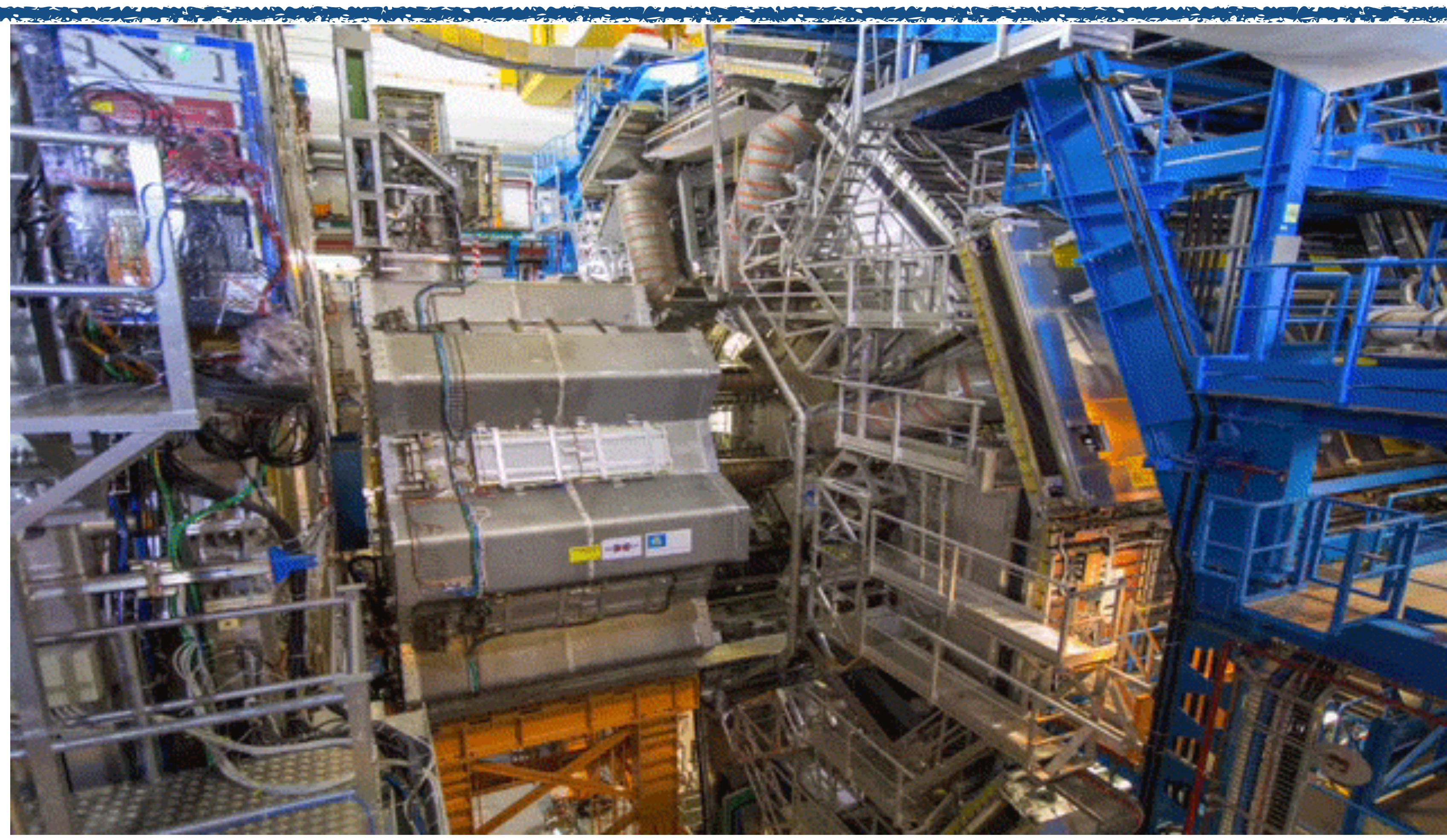
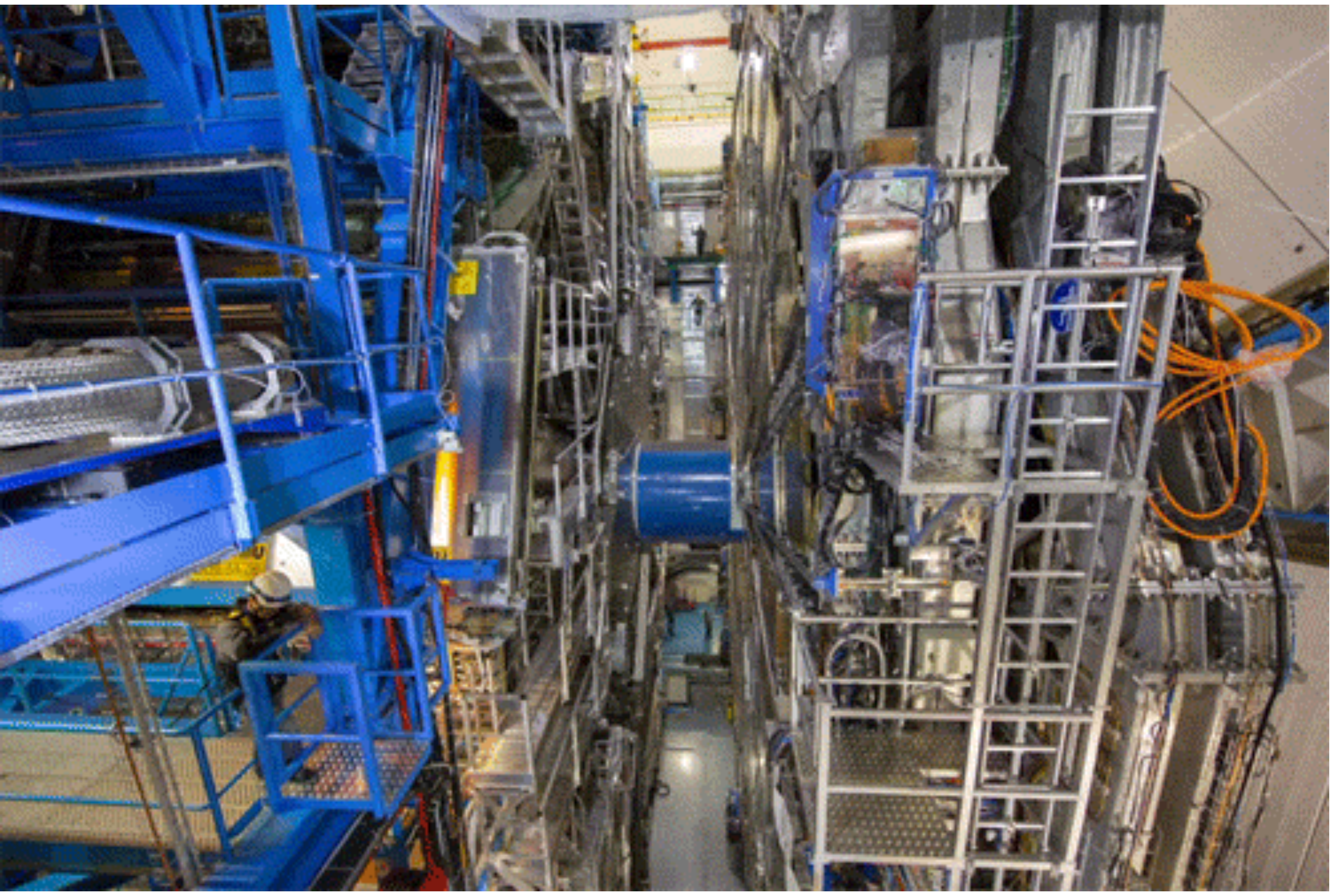
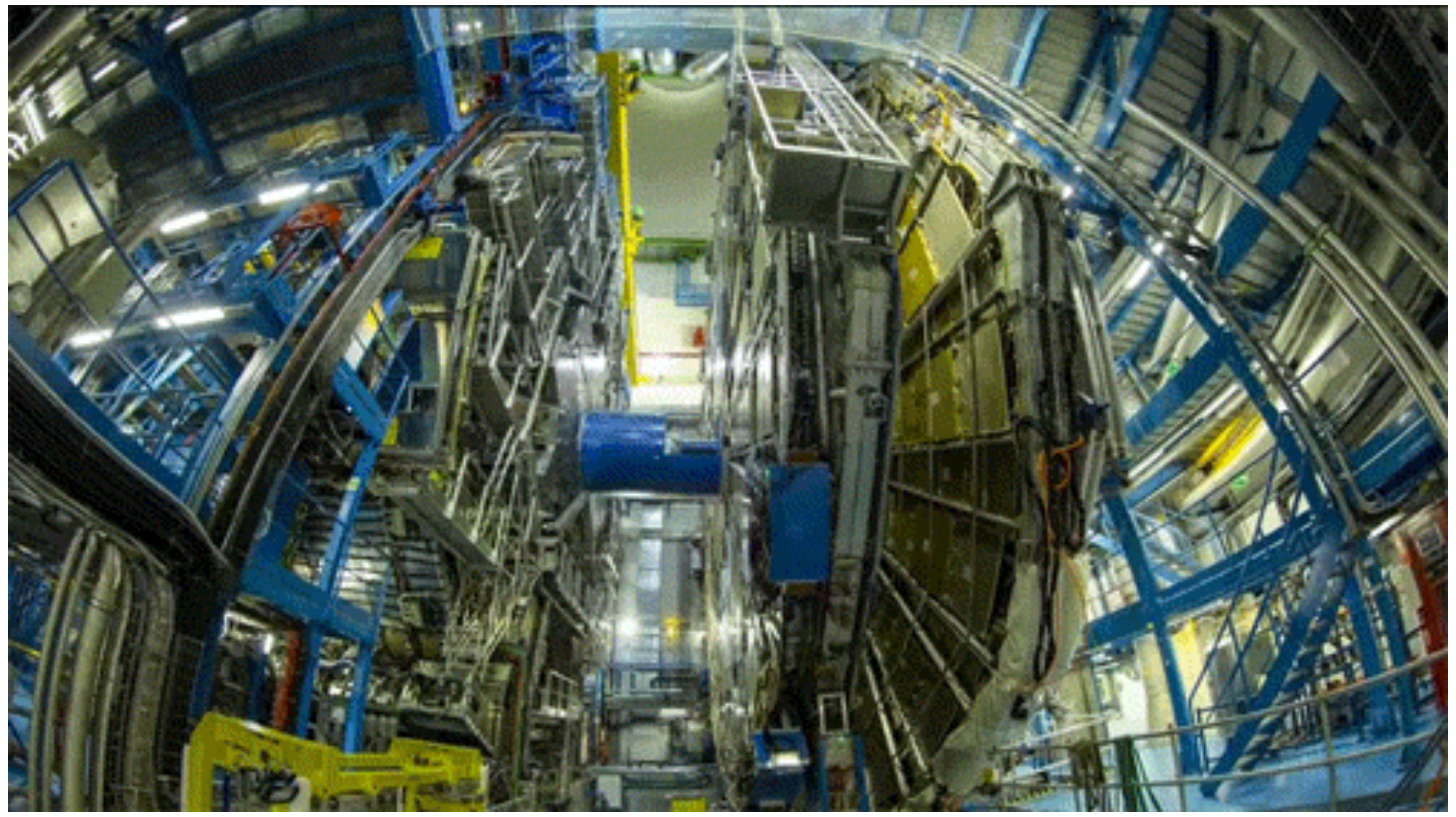
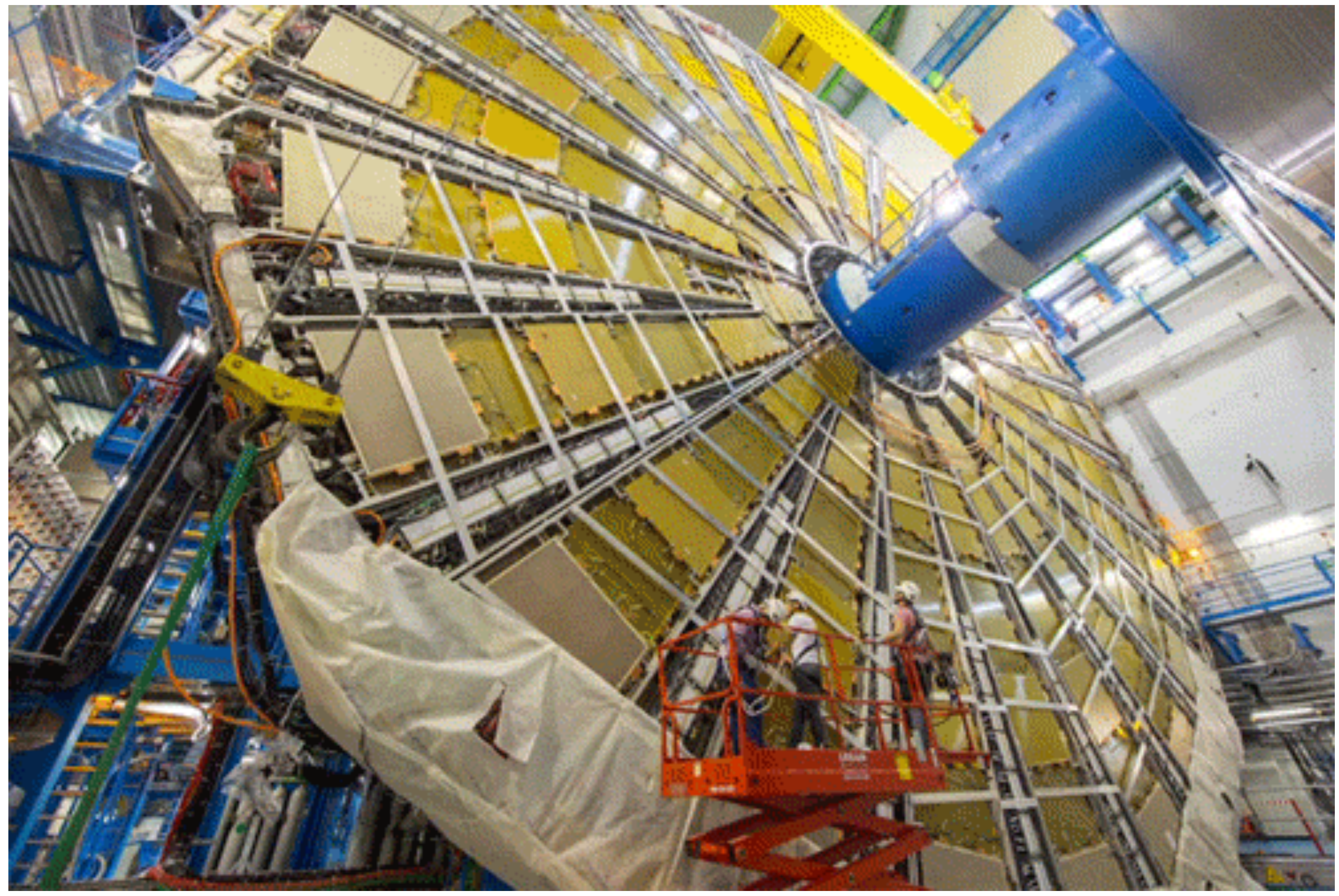
Smash particles at Large Hadron Collider

Video: CERN

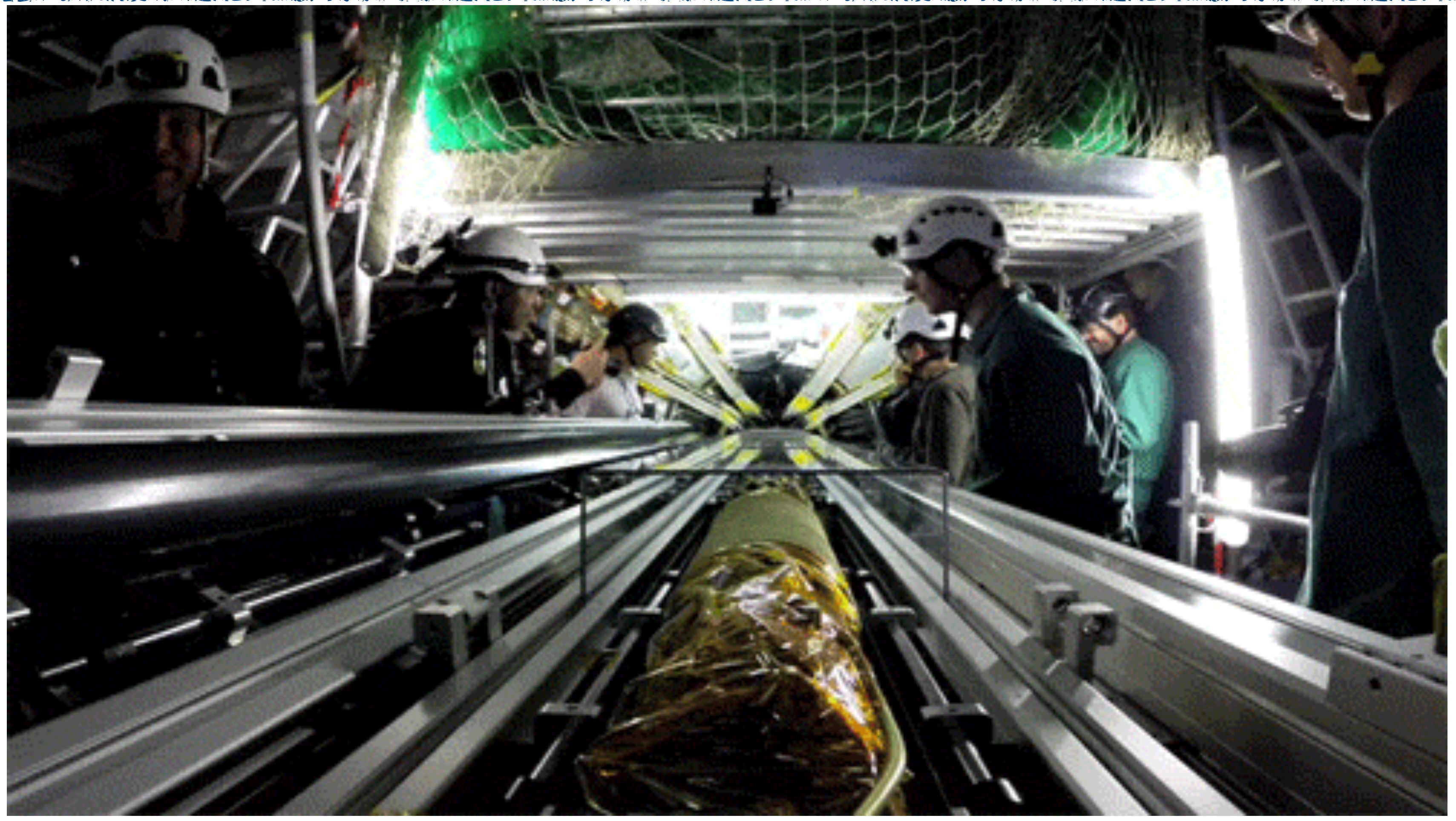
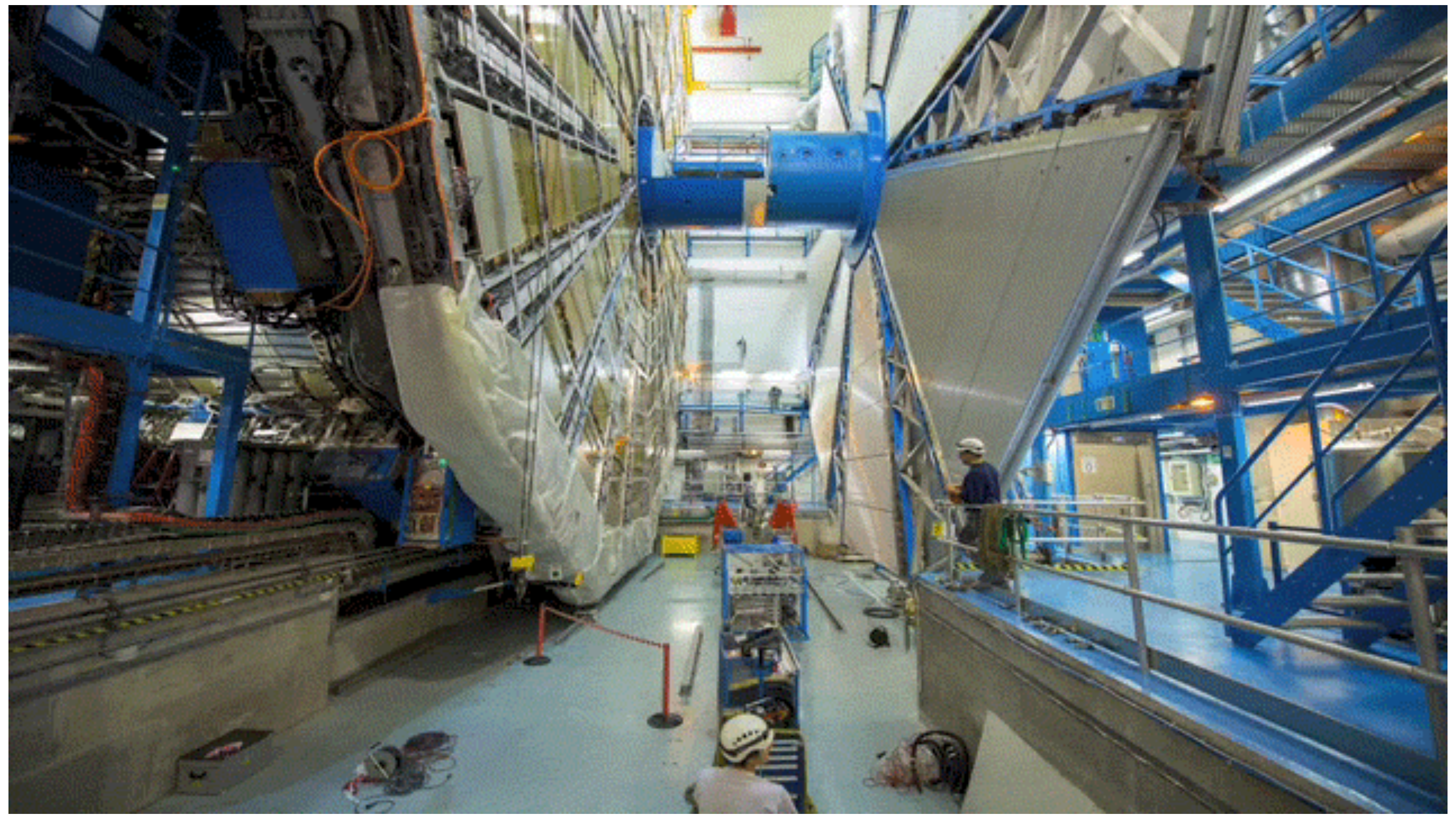
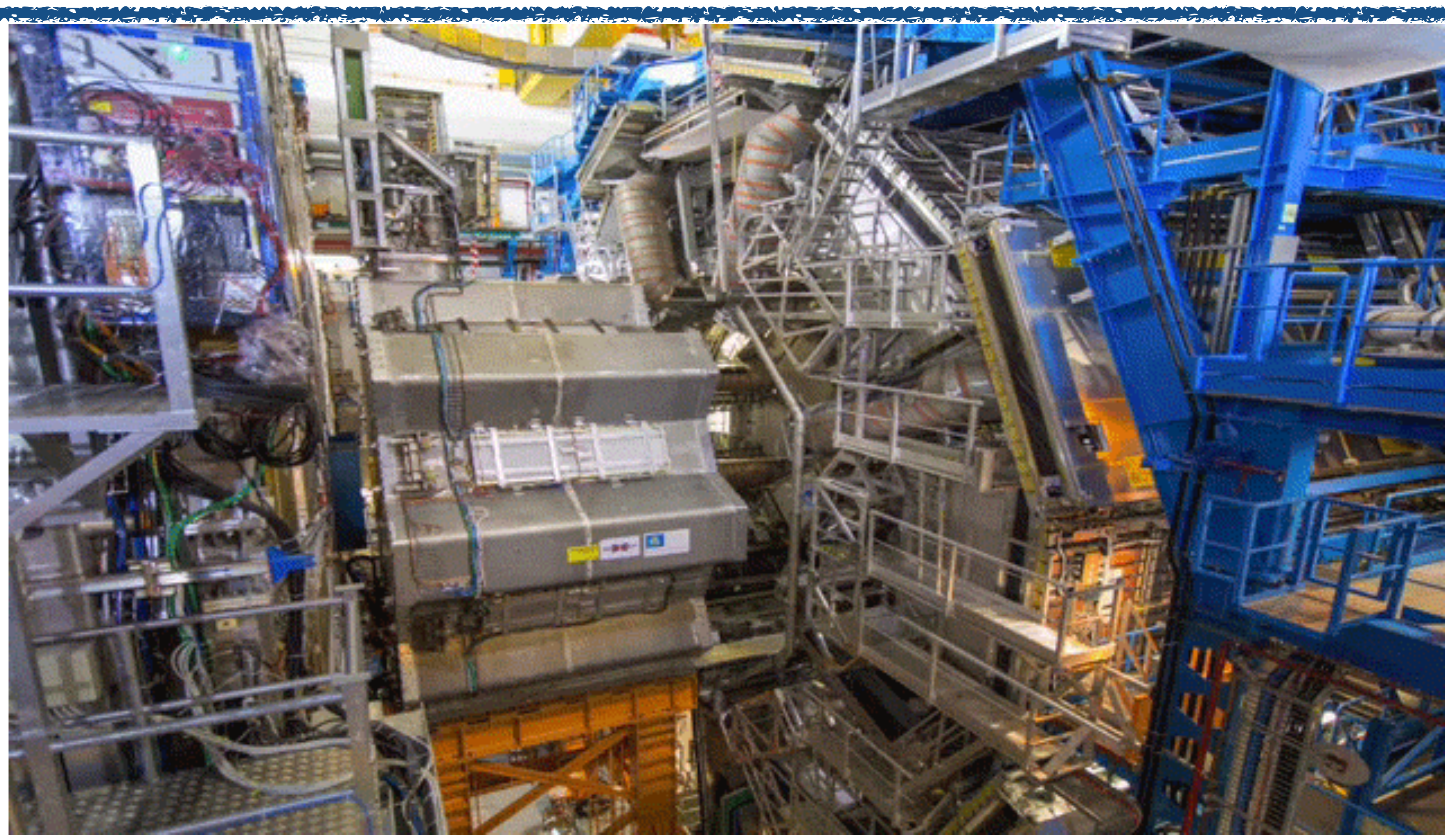
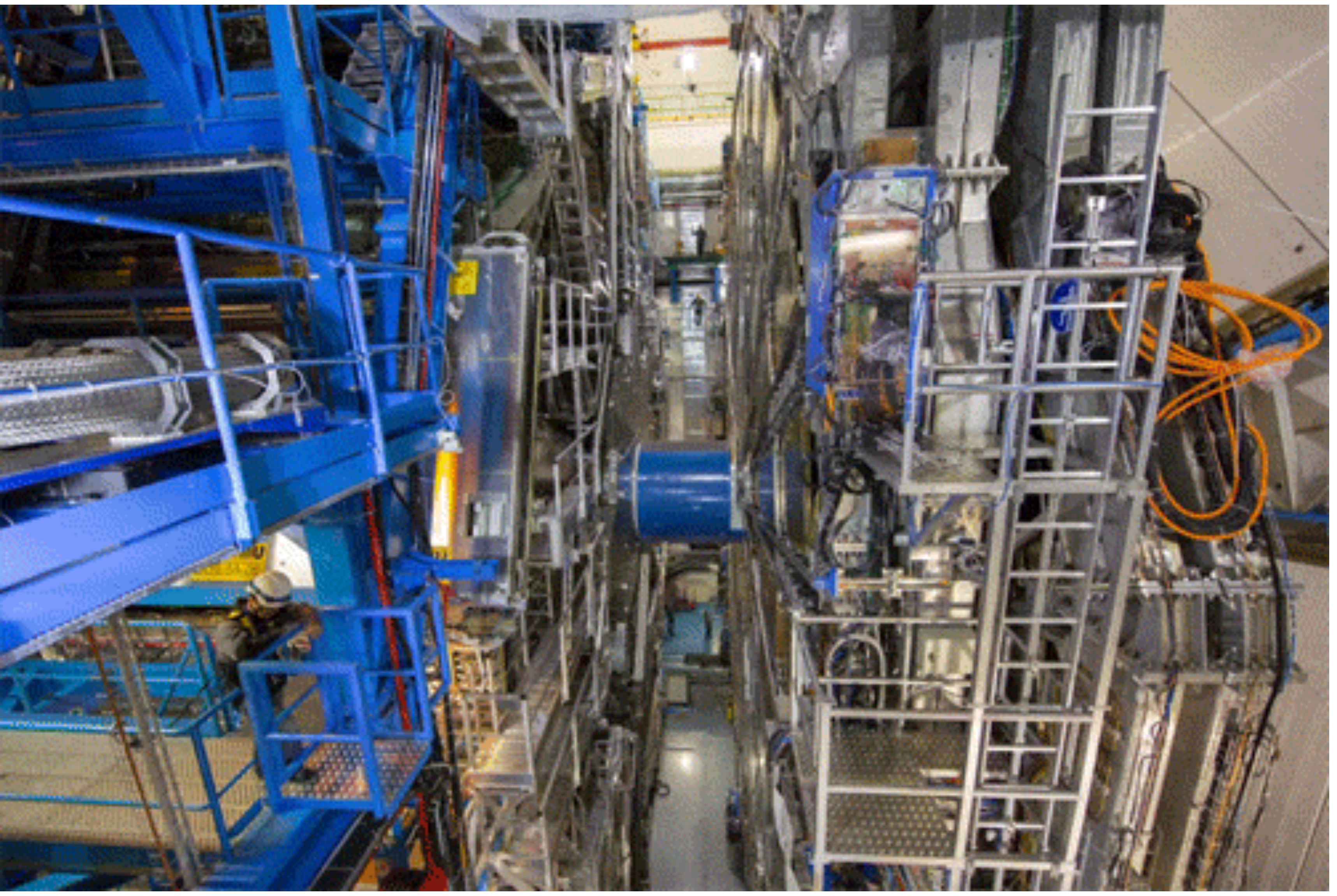
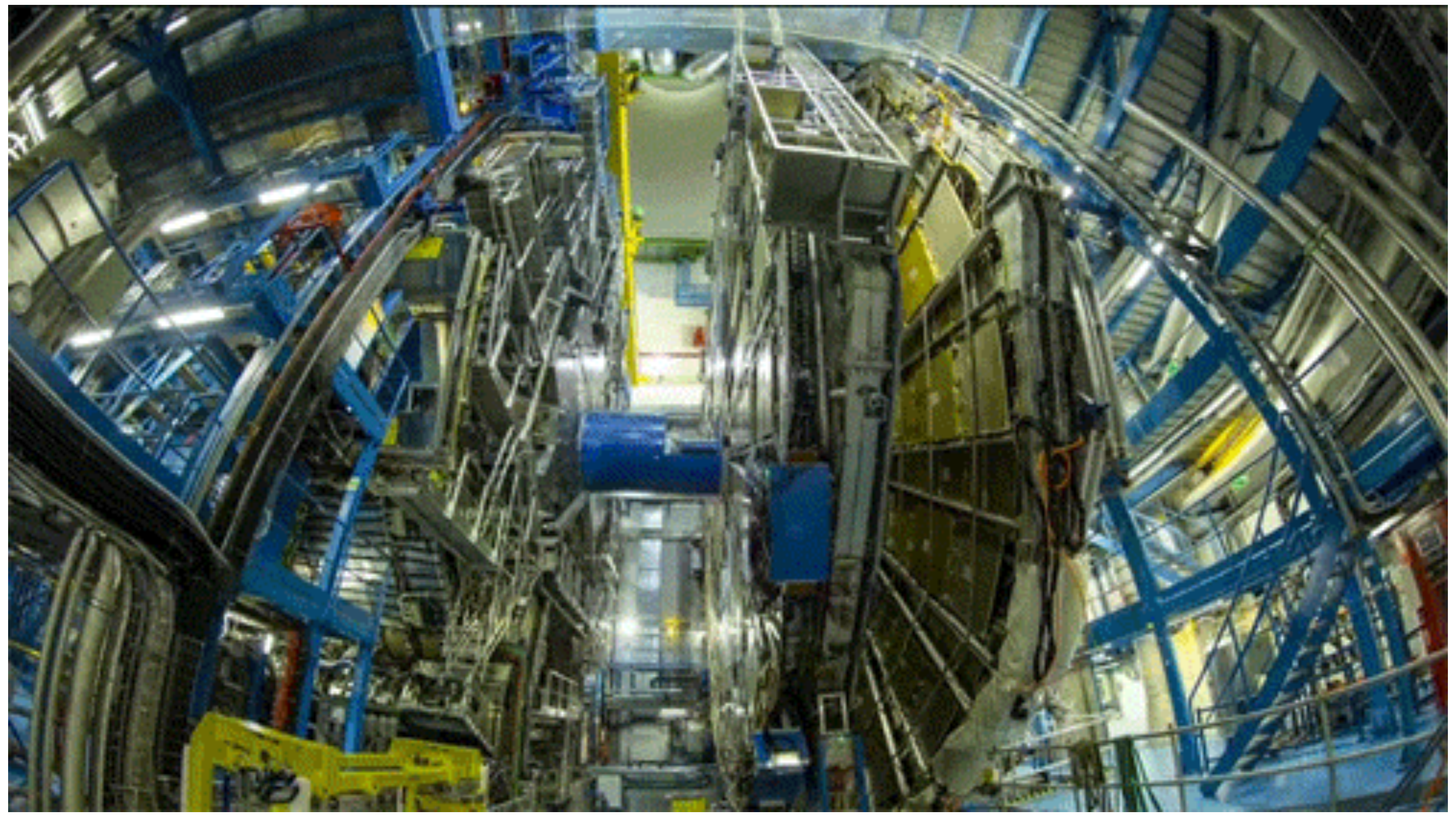
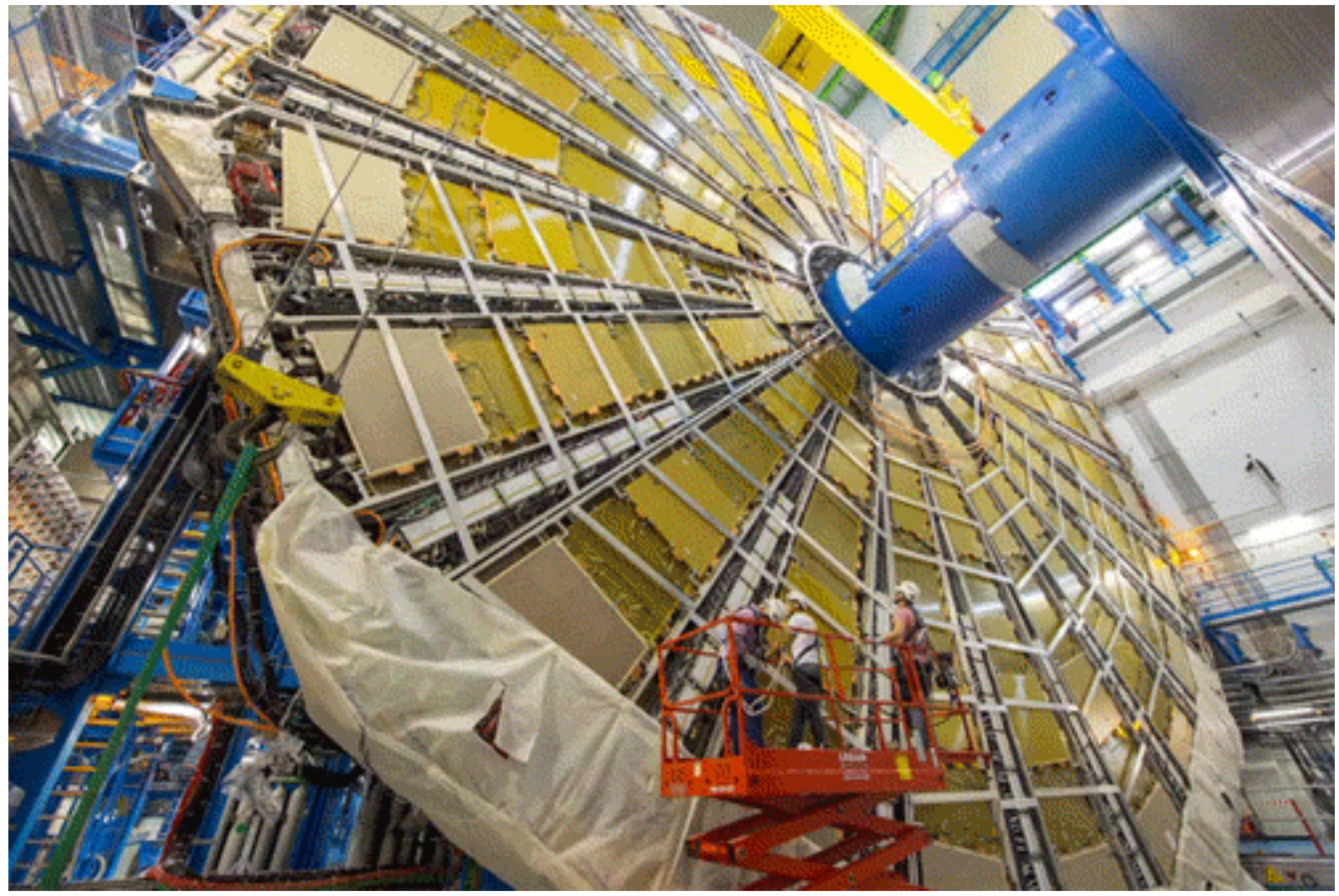
Smash particles at Large Hadron Collider

Video: CERN

The detectors

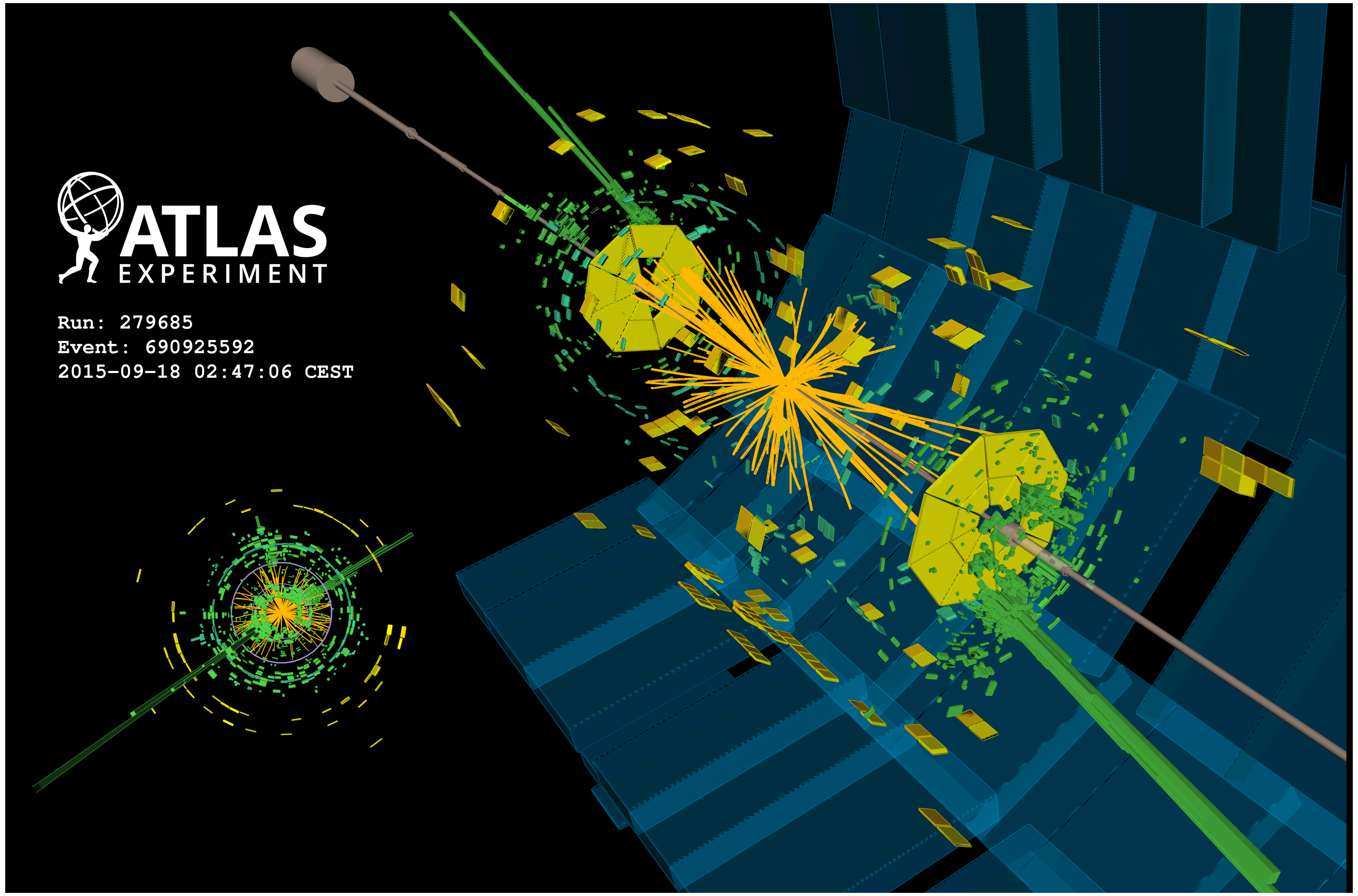
The detectors

Summarise in low dimensions

- Detector has O(100 million) sensors
- Can't build 100M dimensional histogram

- Reconstruction pipeline, event selection
- Design sensitive one-dimensional observable

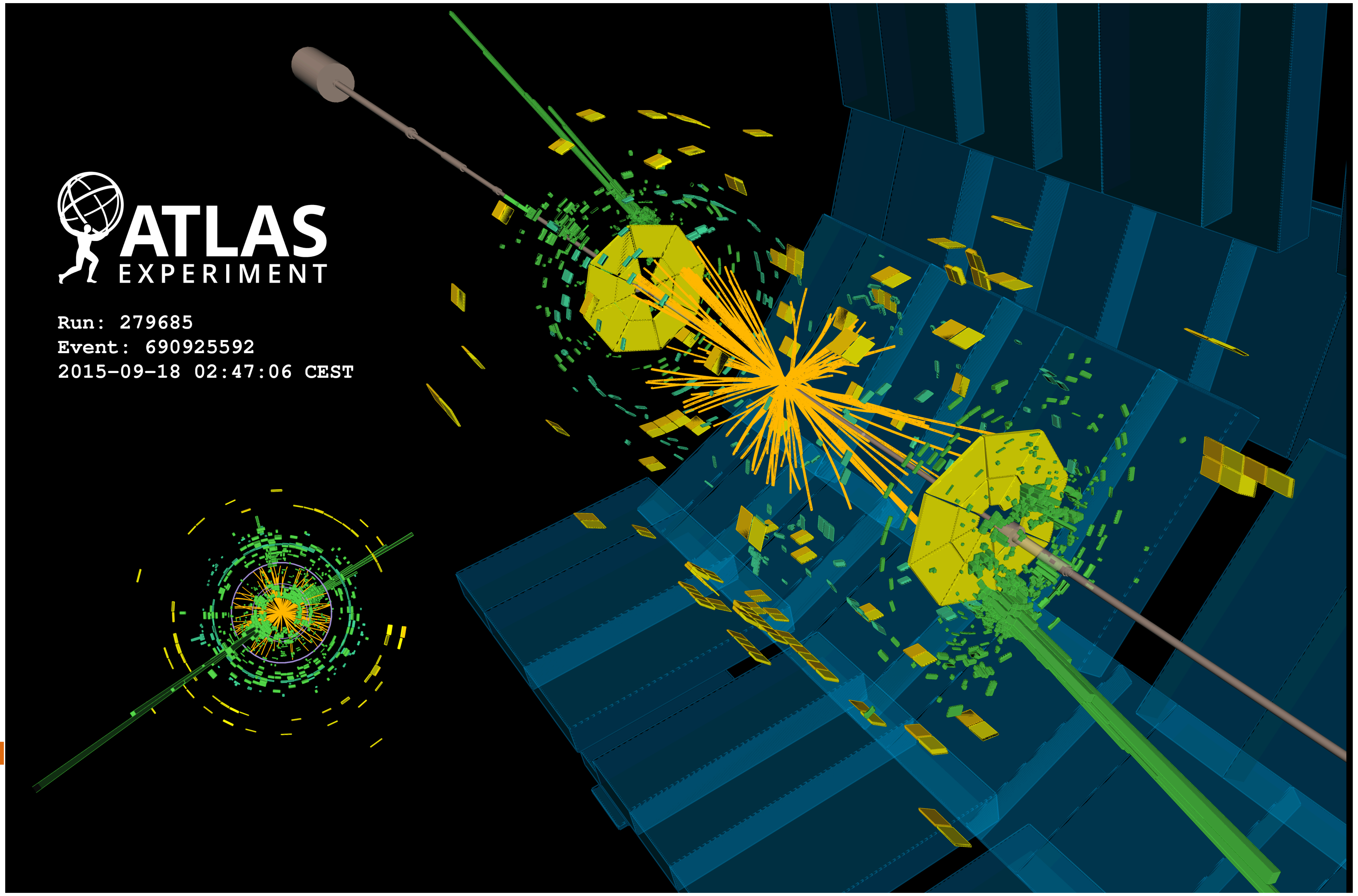


Summarise in low dimensions

- Detector has O(100 million) sensors
- Can't build 100M dimensional histogram

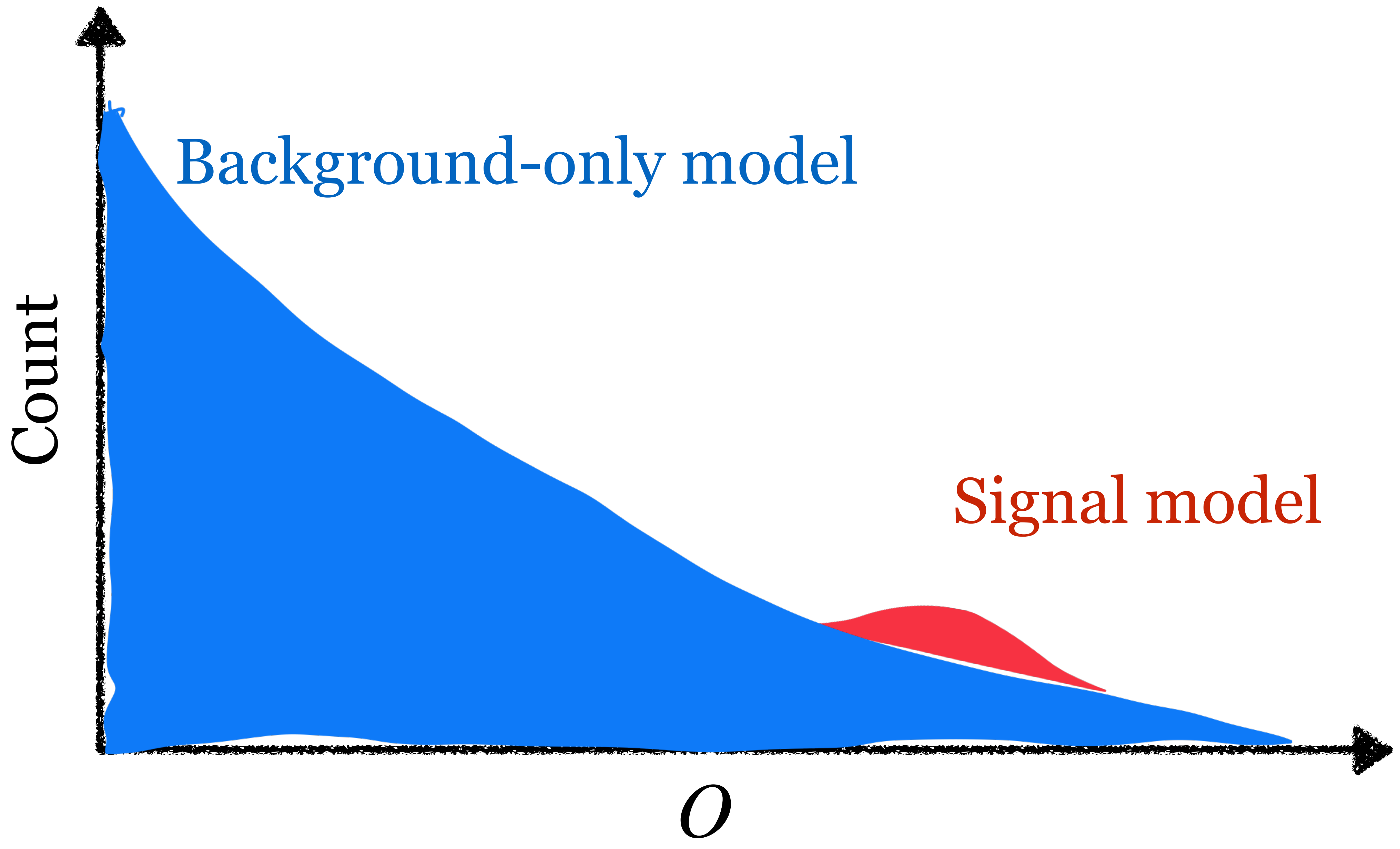
- Reconstruction pipeline, event selection
- Design sensitive one-dimensional observable

1 number

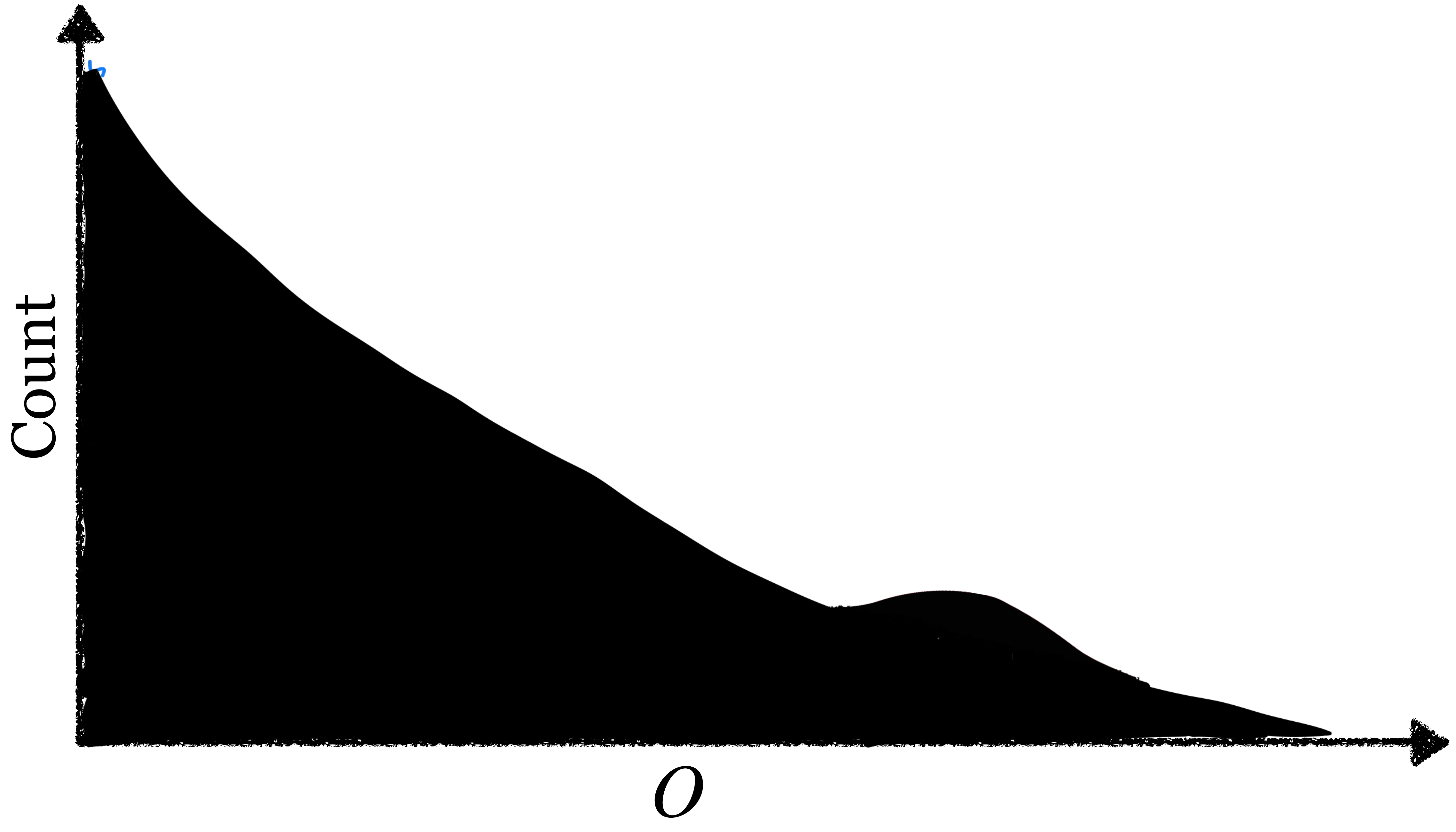


Probability Density Estimation: What we're used to doing..

Theory Predictions

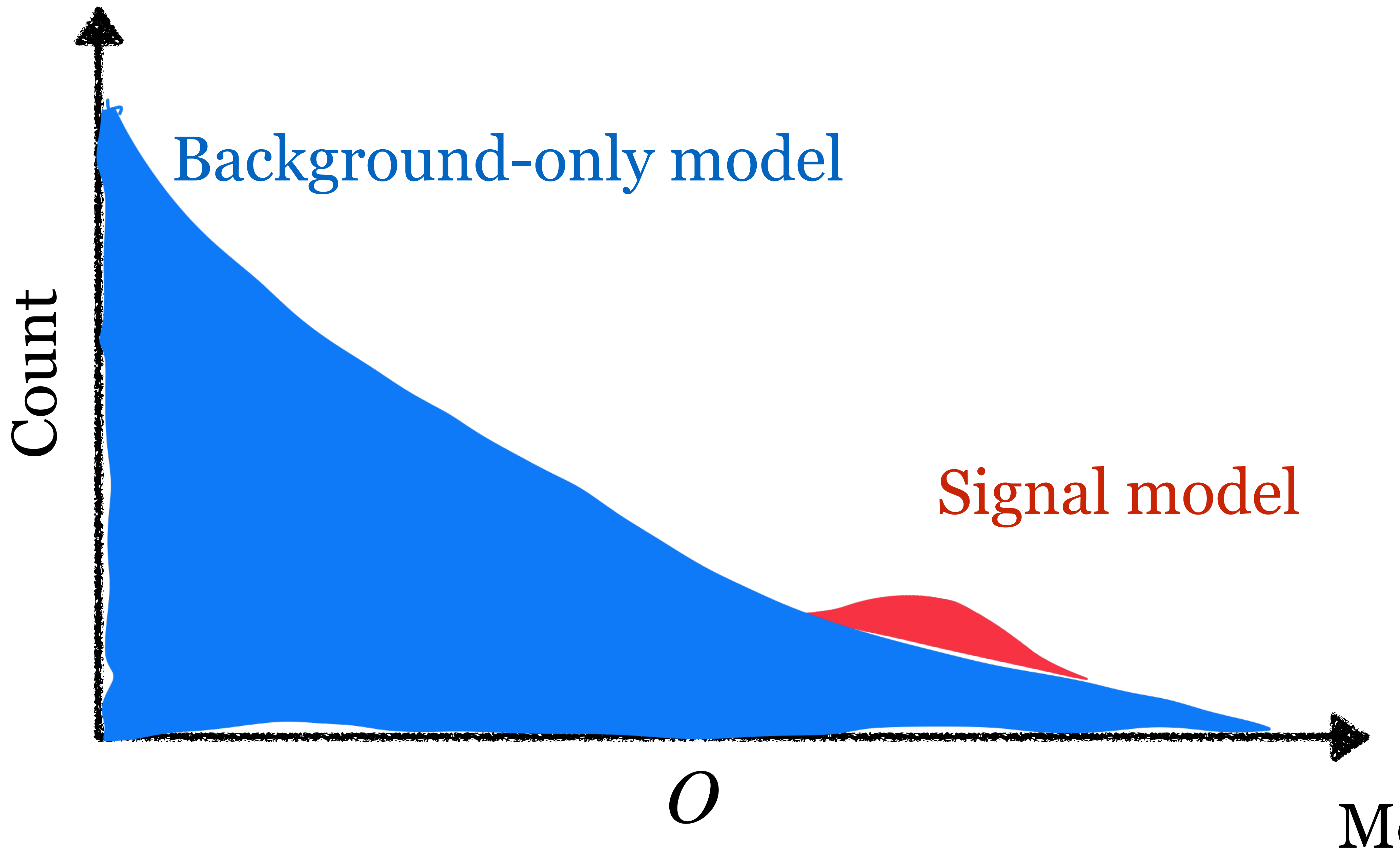


Data

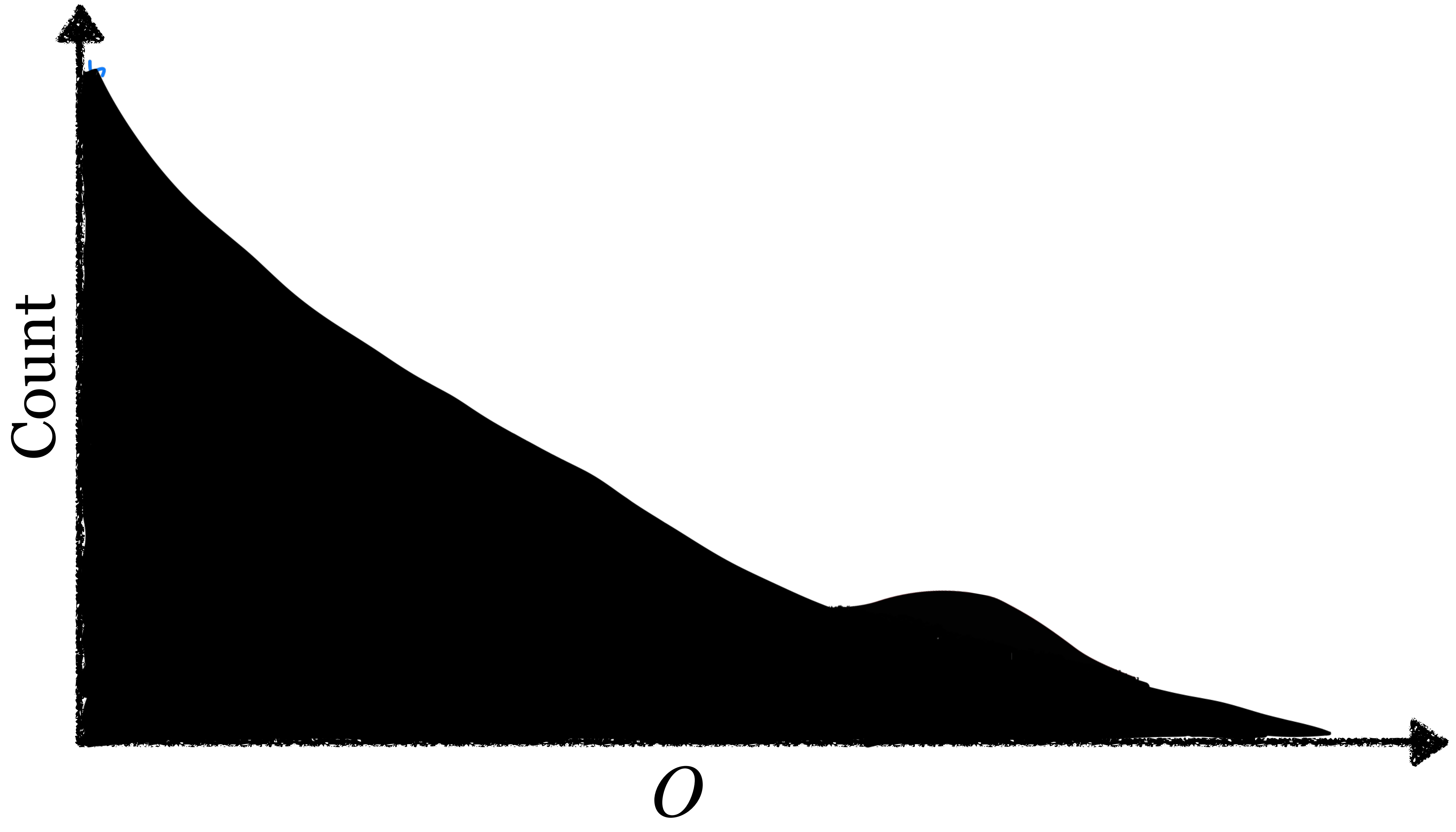


Probability Density Estimation: What we’re used to doing..

Theory Predictions



Data

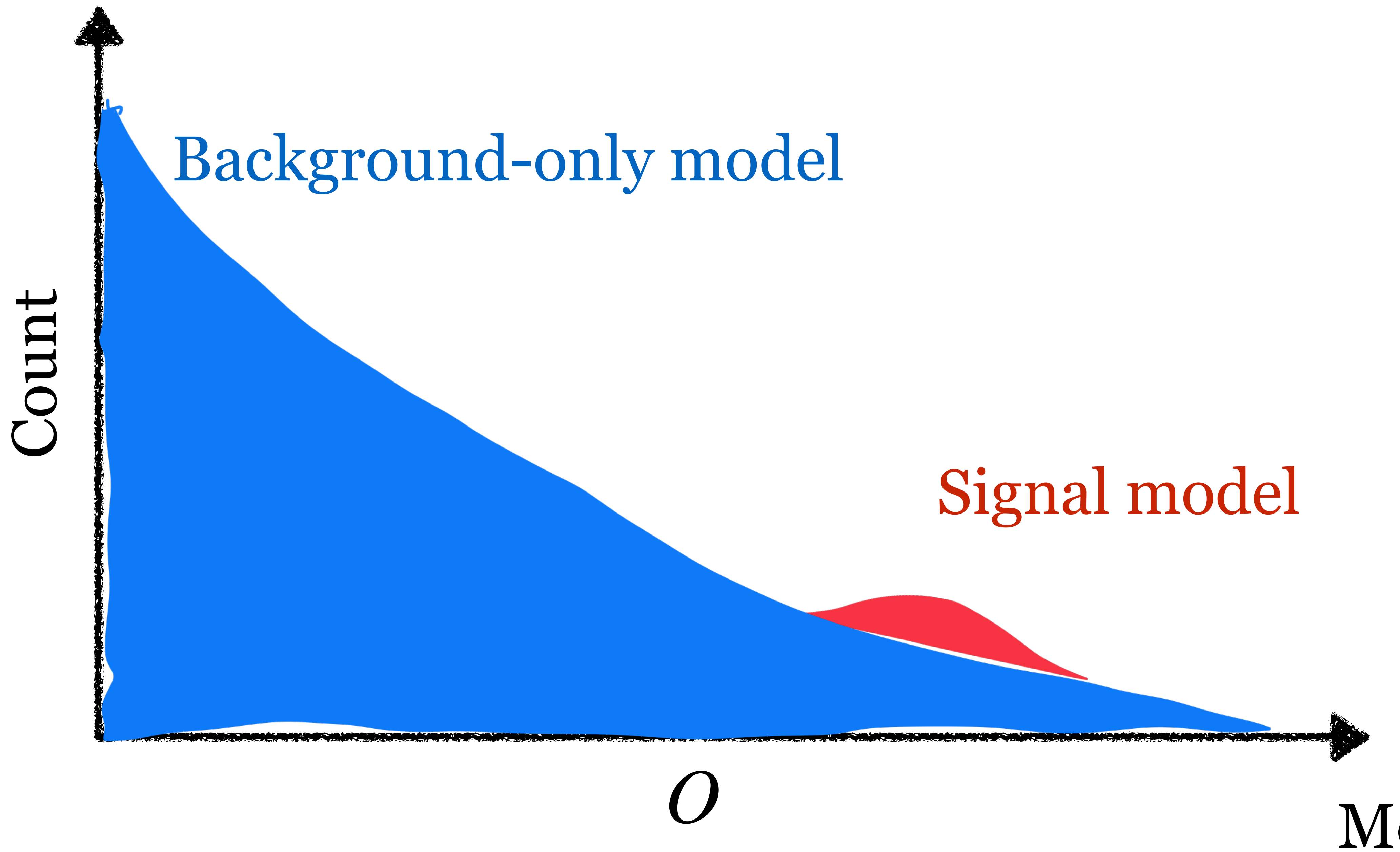


Measure signal strength μ

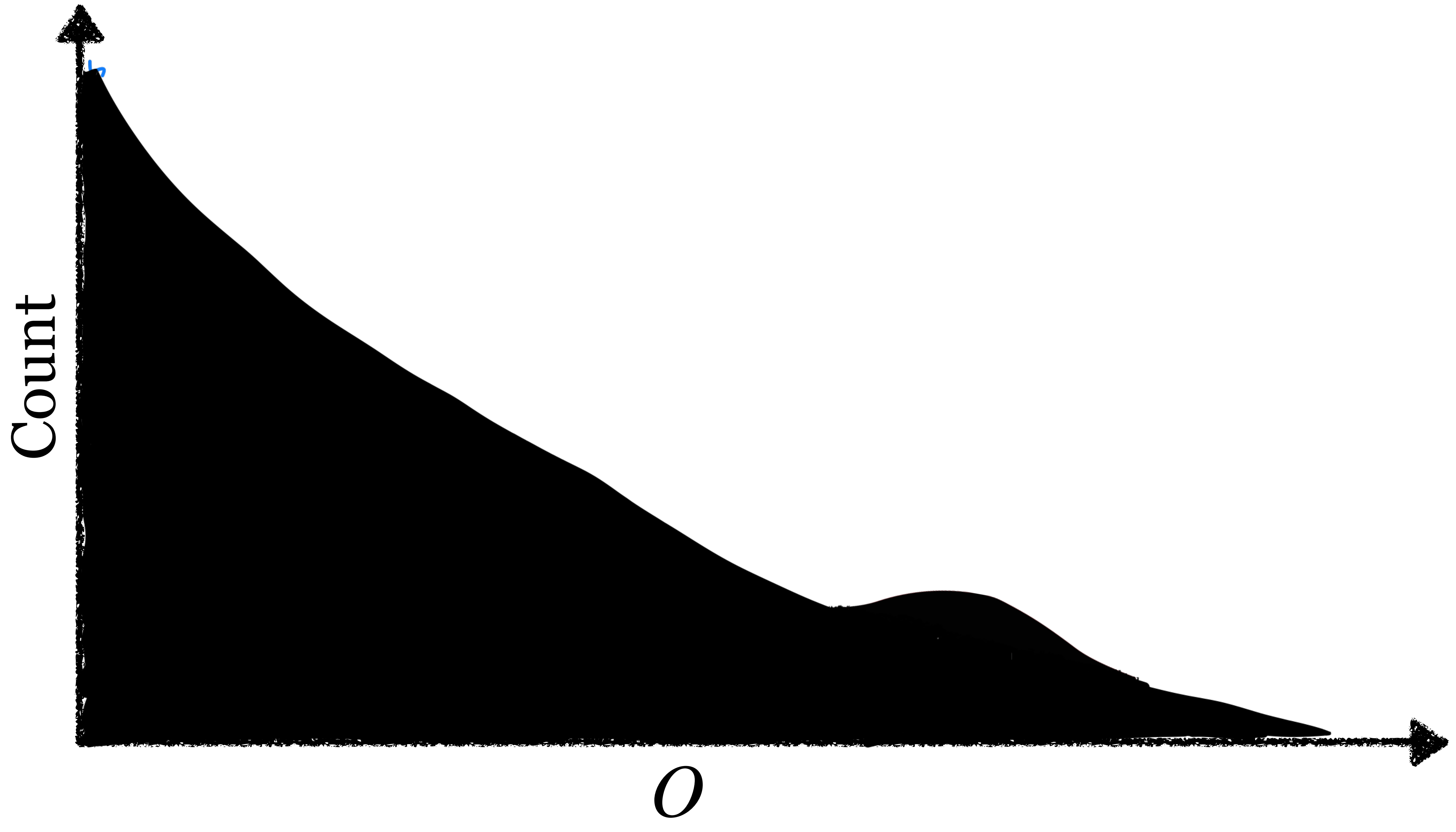
With histograms we can ask “Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?”

Probability Density Estimation: What we're used to doing..

Theory Predictions



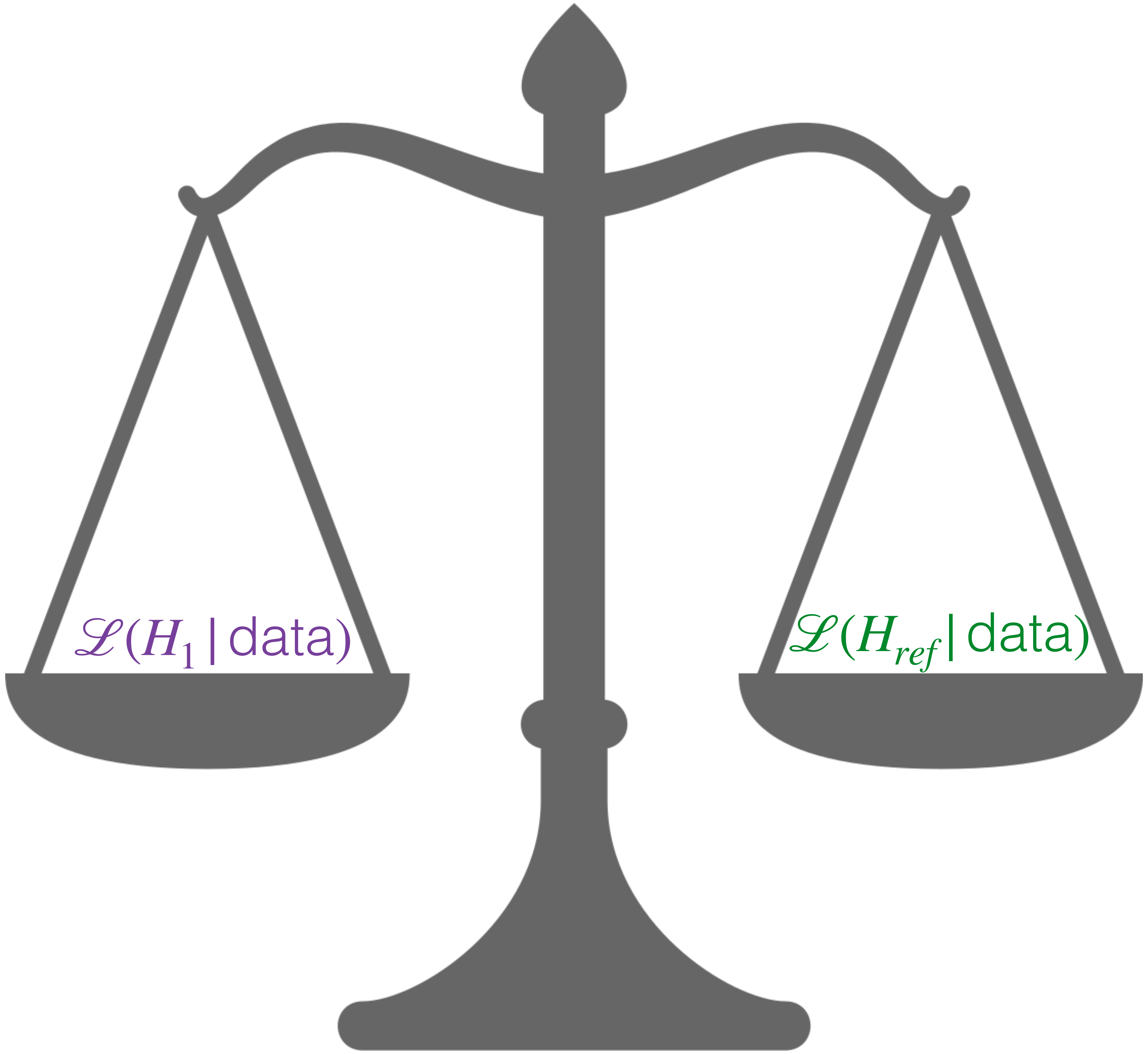
Data



Measure signal strength μ

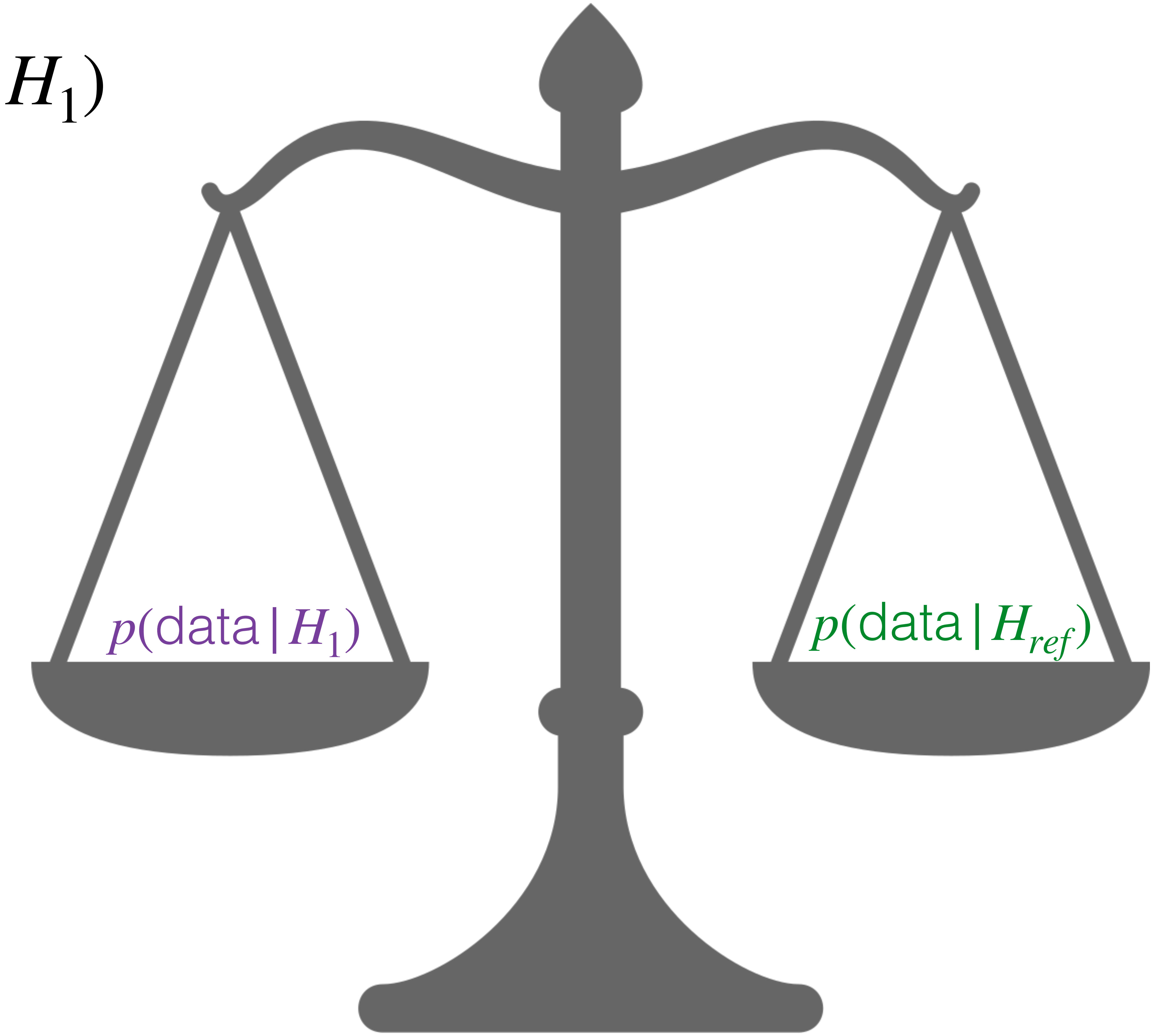
With histograms we can ask "Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?"

(Frequentist) Hypothesis tests



(Frequentist) Hypothesis tests

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



Why we can summarise data down to a single observable for typical analysis

Why we can summarise data down to a single observable for typical analysis

$$\mathscr{L}(\mu \mid \mathscr{D}) = p(\mathscr{D} \mid \mu)$$

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

We want to compare likelihoods: $$\frac{p(\mathscr{D} \mid \mu)}{p(\mathscr{D} \mid \mu_0)}$$

Why we can summarise data down to a single observable for typical analysis

$$\mathscr{L}(\mu \mid \mathscr{D}) = p(\mathscr{D} \mid \mu)$$

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

We want to compare likelihoods: $$\frac{p(\mathscr{D} \mid \mu)}{p(\mathscr{D} \mid \mu_0)}$$

Why we can summarise data down to a single observable for typical analysis

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

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

We want to compare likelihoods: $$\frac{p(\mathcal{D} \mid \mu)}{p(\mathcal{D} \mid \mu_0)}$$

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

* Equal class weights

Why we can summarise data down to a single observable for typical analysis

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

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

We want to compare likelihoods: $$\frac{p(\mathcal{D} \,|\, \mu)}{p(\mathcal{D} \,|\, \mu_0)}$$

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:

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

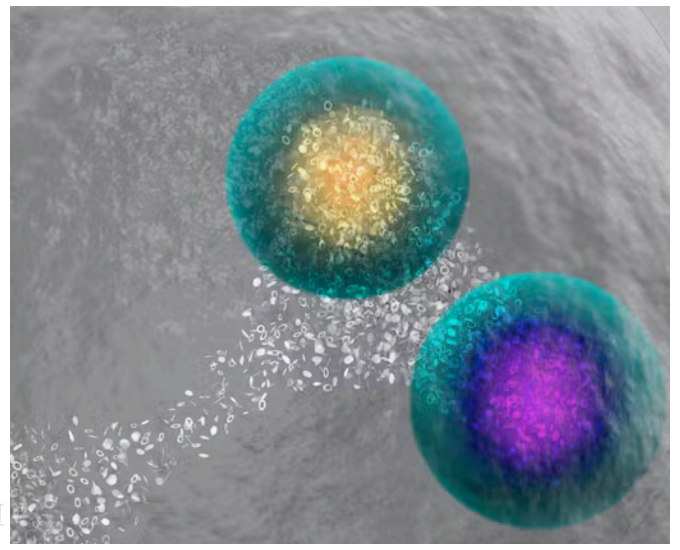
Same observable s is optimal to test all μ hypotheses!

No need to develop separate analysis per hypothesis μ

* Equal class weights

A measurement of the Higgs width

Image: CERN



Undiscovered massive particles

$\longrightarrow H \longrightarrow \Gamma_H$

A measurement of the Higgs width

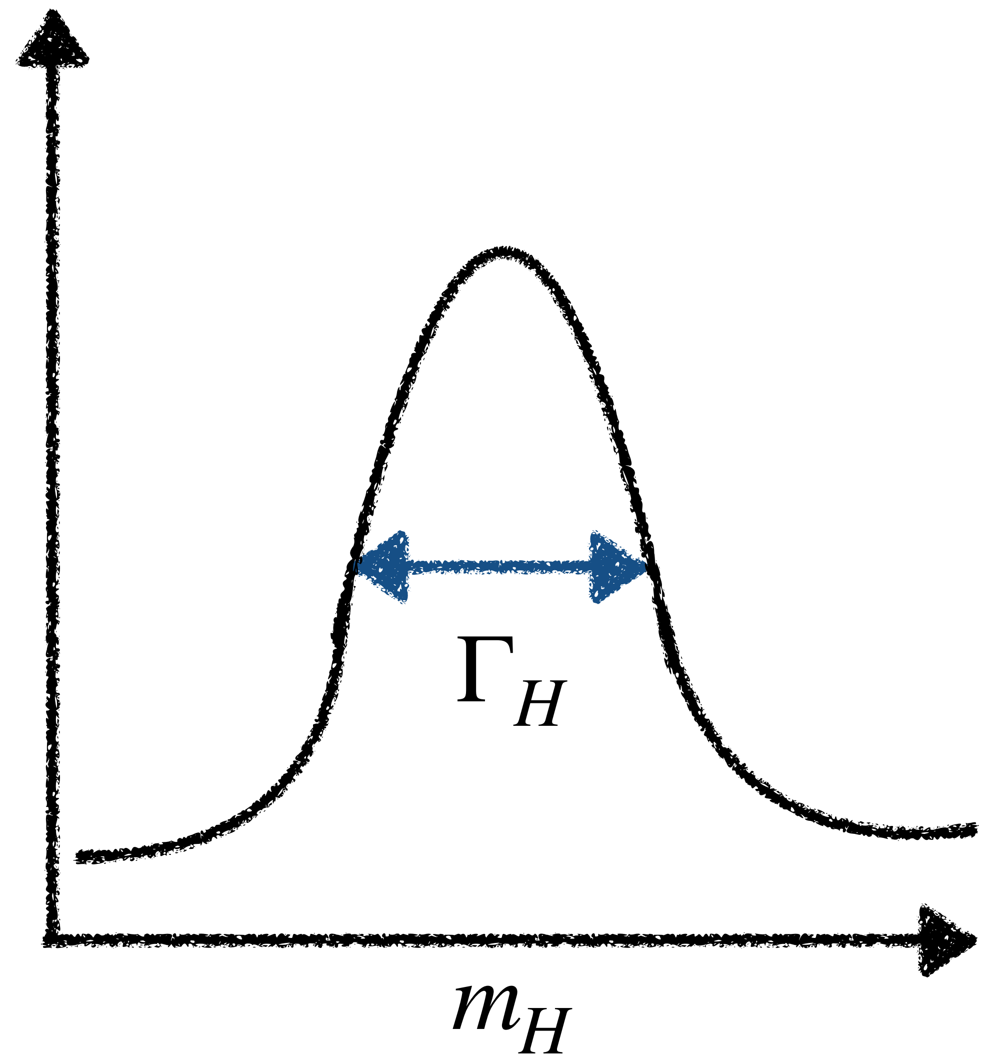
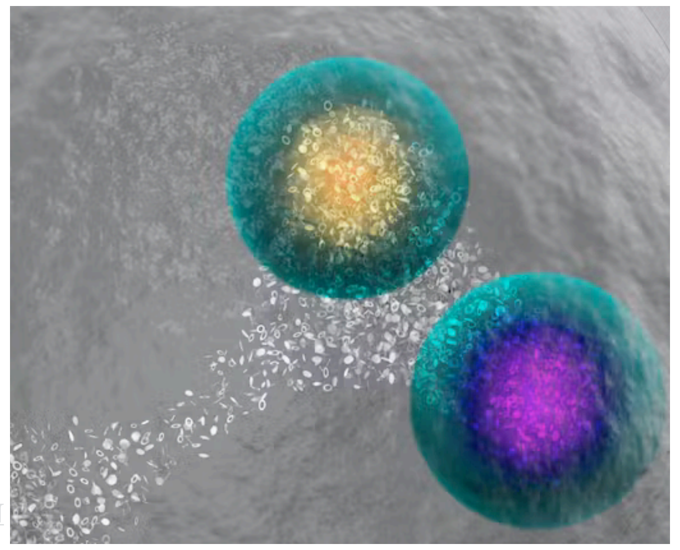
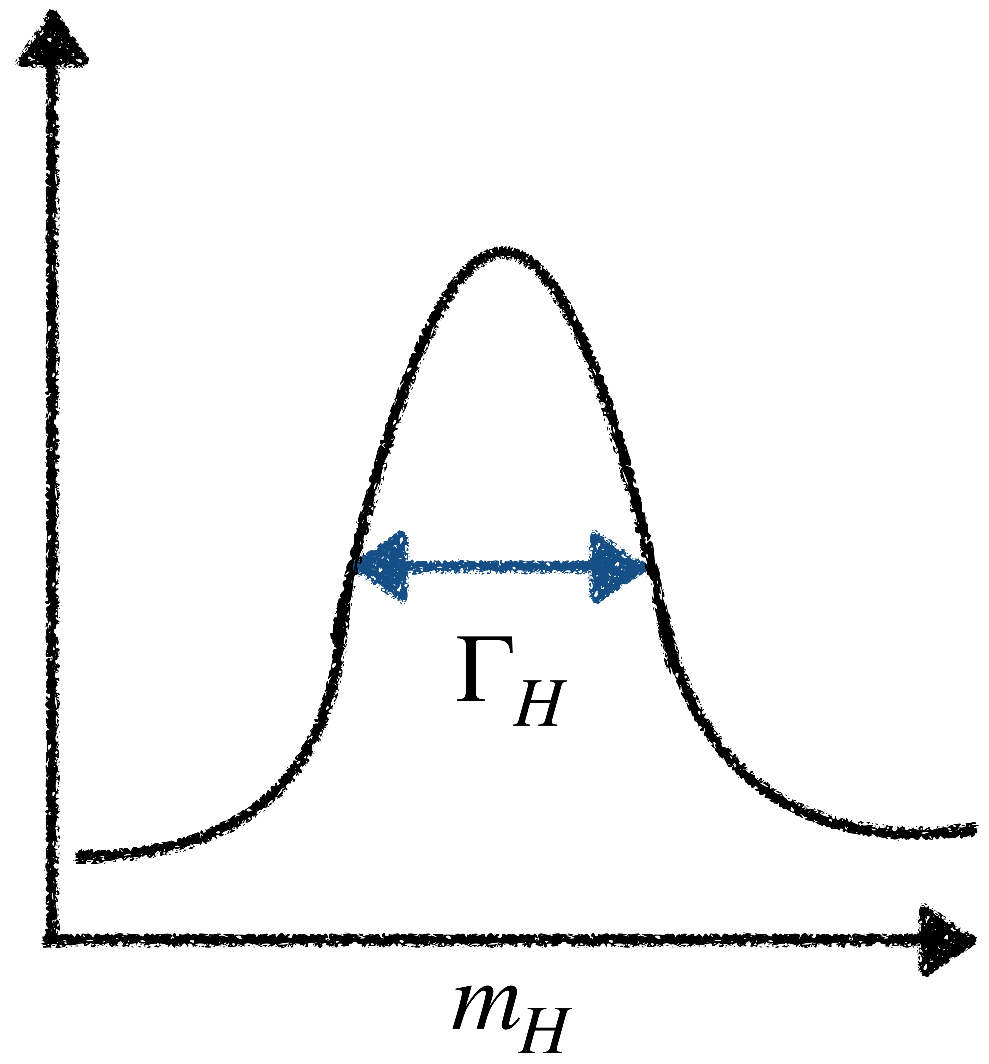


Image: CERN


Undiscovered massive particles → H → Γ_H

A measurement of the Higgs width

- 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
- Central topic for future colliders



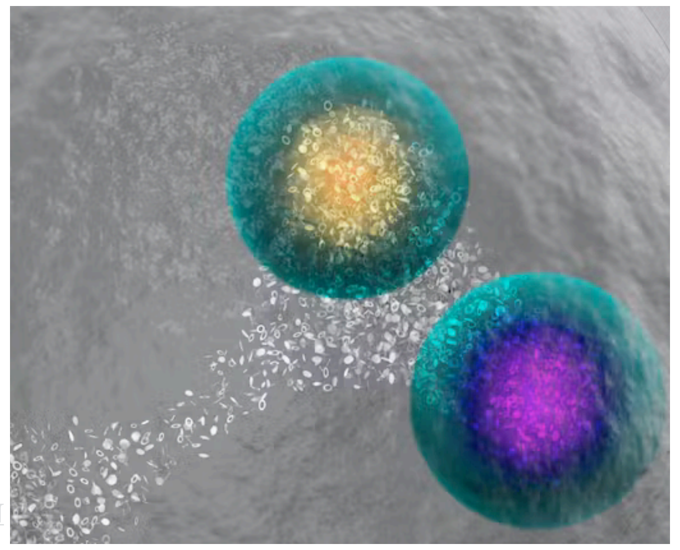


Image: CERN

Undiscovered massive particles → H → Γ_H

Higgs Width from off-shell Higgs production

arXiv:1405.0285 arXiv:1406.1757

- Off-shell production helps probe Higgs width

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

$$\sigma^{gg\to H\to ZZ}_{on-shell} \sim \frac{g_f^2 g_V^2}{m_H \Gamma_H}$$

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

Higgs Width from off-shell Higgs production

arXiv:1405.0285 arXiv:1406.1757

- Off-shell production helps probe Higgs width

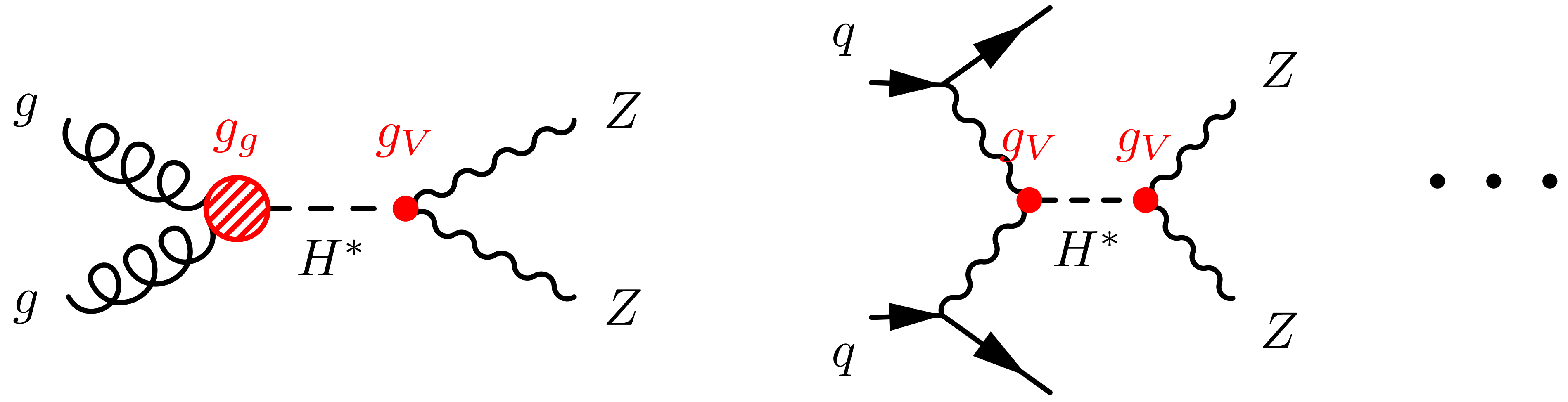
$$\boxed{\frac{\mu_{\text{off-shell}}}{\mu_{\text{on-shell}}} = \frac{\Gamma_H}{\Gamma_H^{SM}}.}$$

$$\sigma^{gg\to H\to ZZ}_{on-shell} \sim \frac{g_f^2 g_V^2}{m_H \Gamma_H}$$

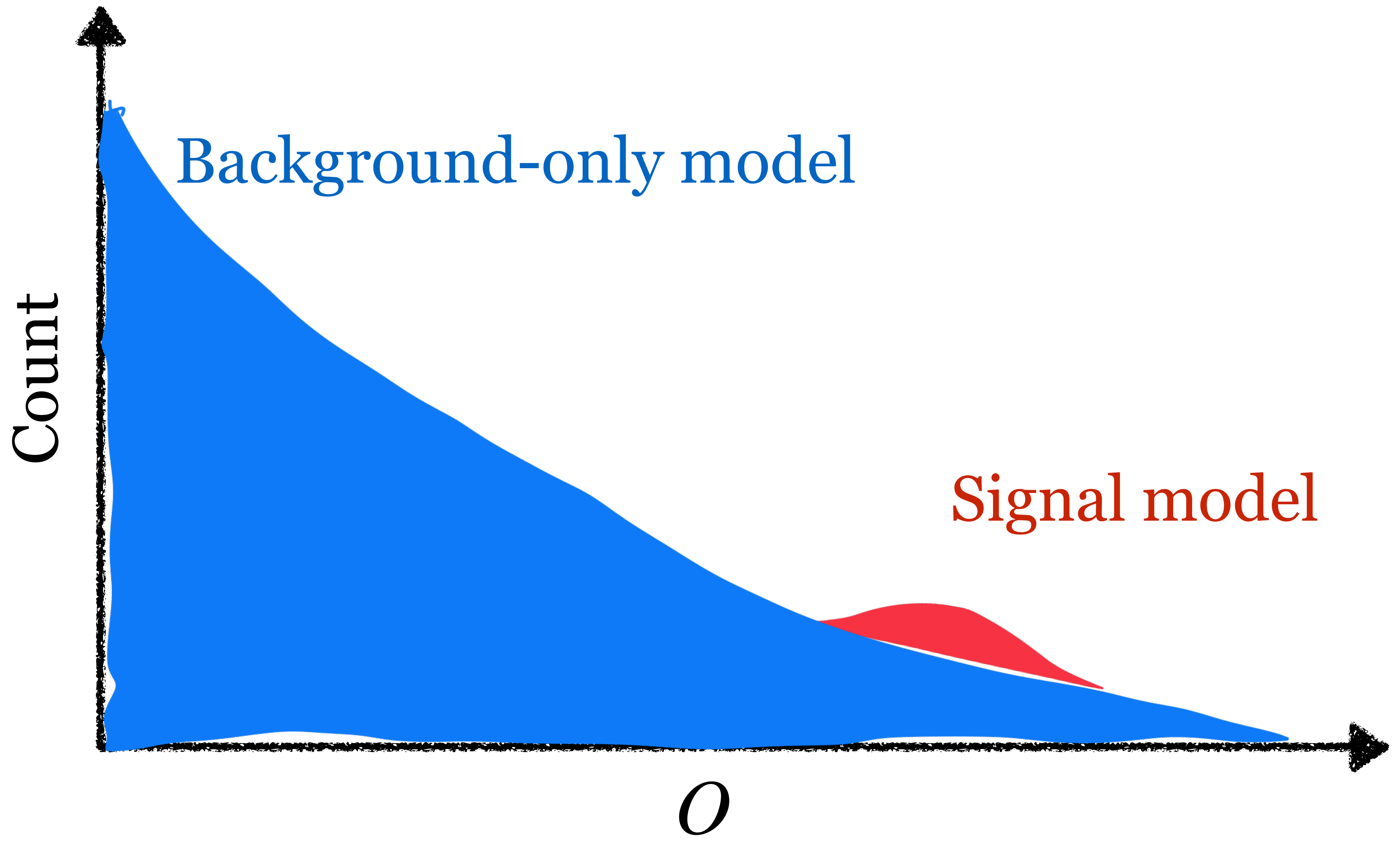
$$\frac{d\sigma^{gg\to H\to VV}_{\text{off-shell}}}{dm_{VV}} \propto \frac{g_f^2 g_V^2}{(m_{VV}^2 - m_H^2)^2 + m_H^2\Gamma_H^2} \simeq \frac{g_f^2 g_V^2}{m_{VV}^4}$$

- Interpretation assumes no new physics

- Essential to measure independently in multiple production modes (ggF, VBF) and final states to verify consistent results

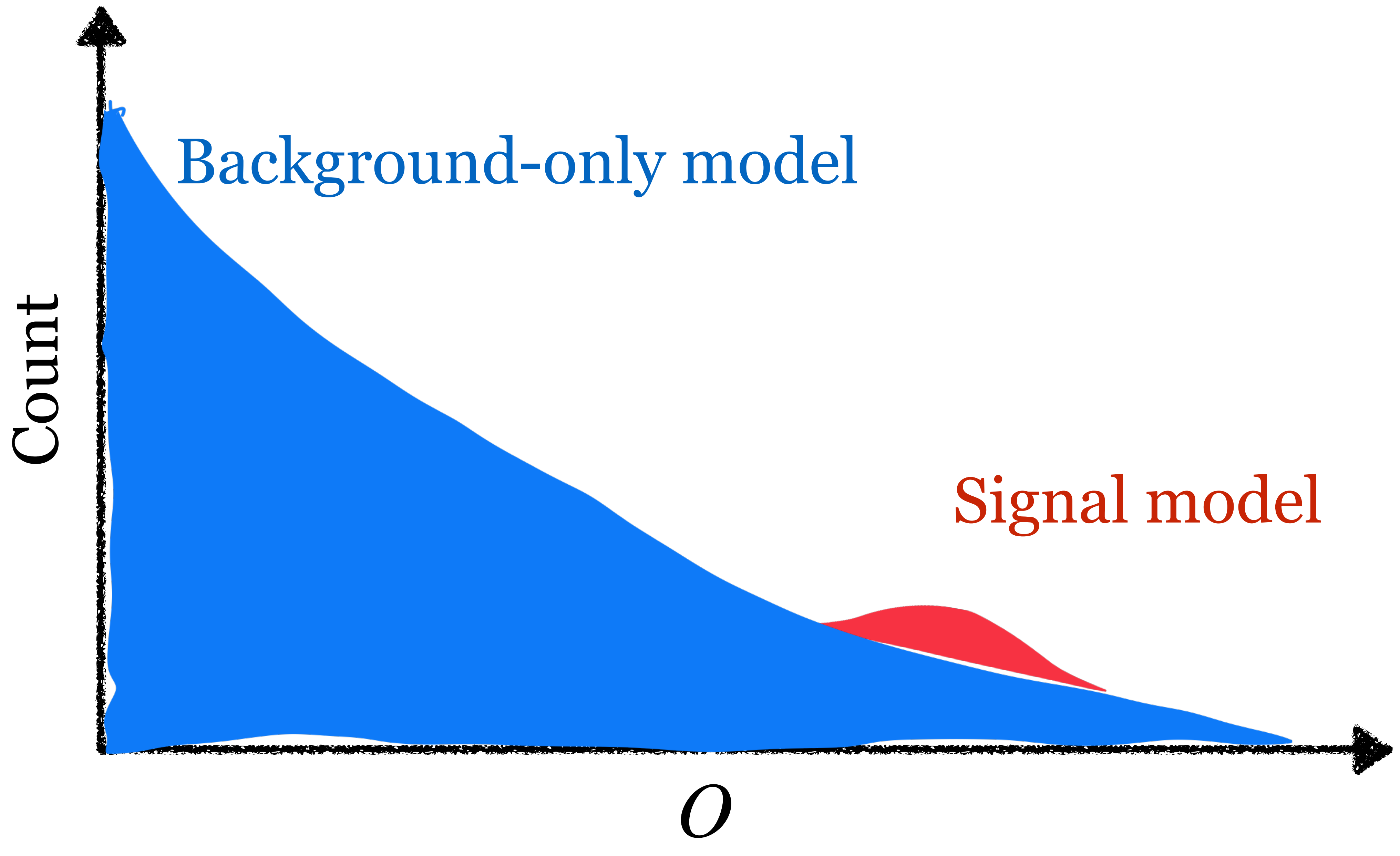


New challenge: Quantum interference
Non-linear changes in kinematics

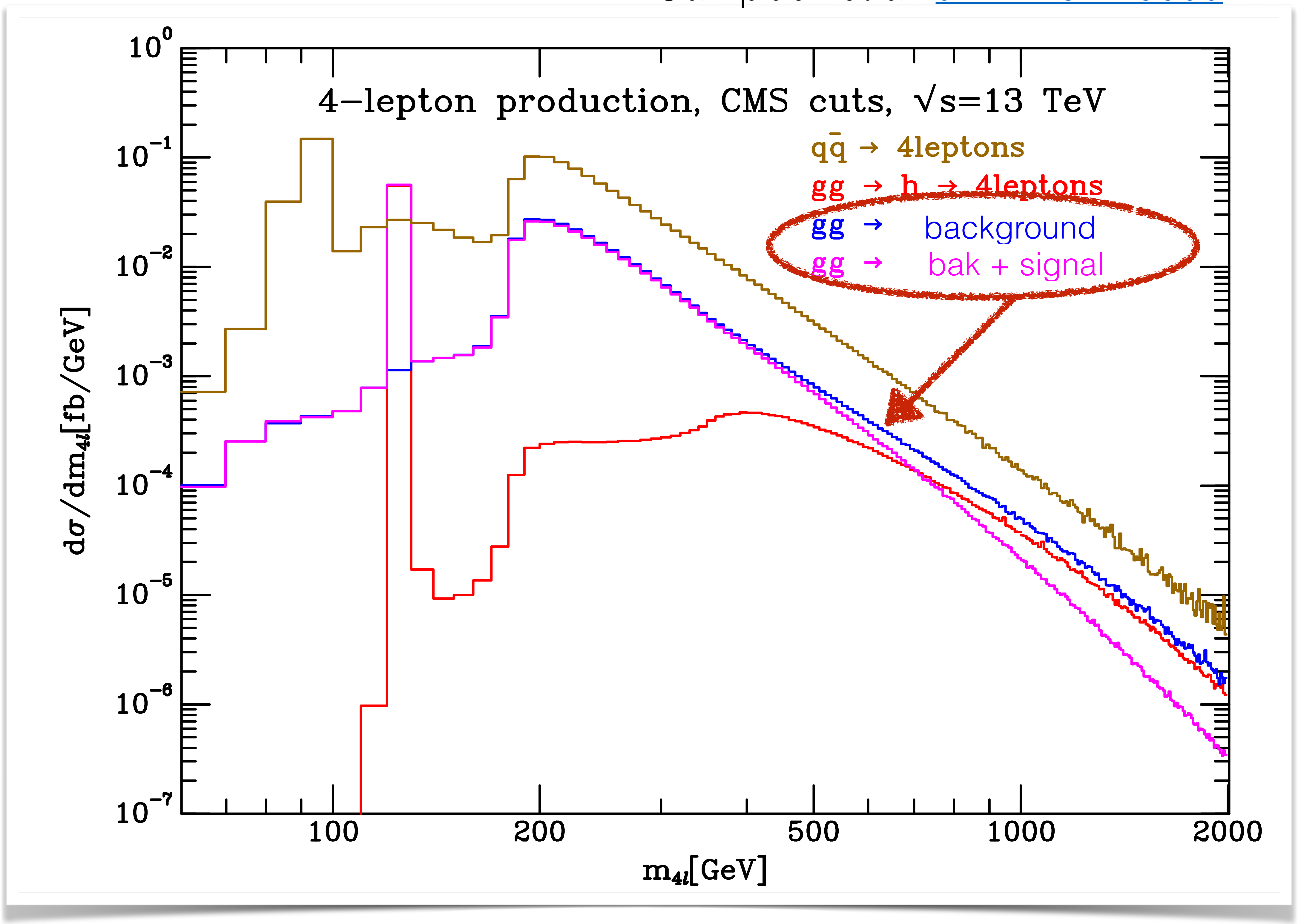


Data can no longer be summarised in 1D histogram (see Ghosh et al: [hal-02971995(p172)](hal-02971995(p172))) !

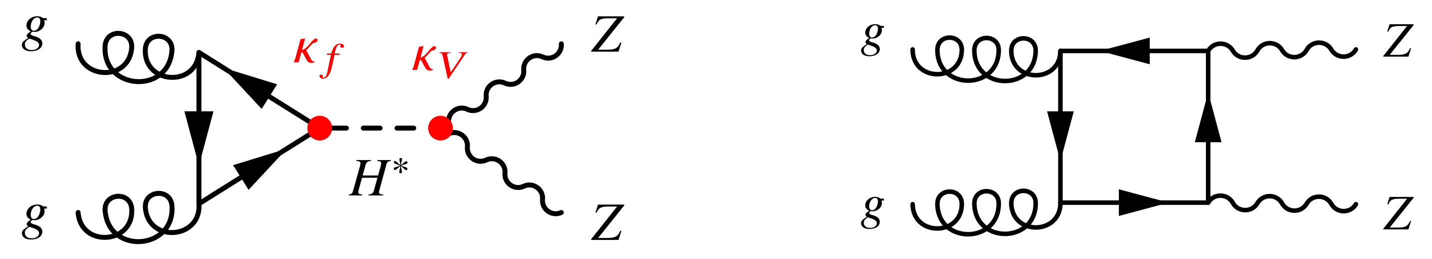
New challenge: Quantum interference
Non-linear changes in kinematics



Campbell et al: arXiv:1311.3589



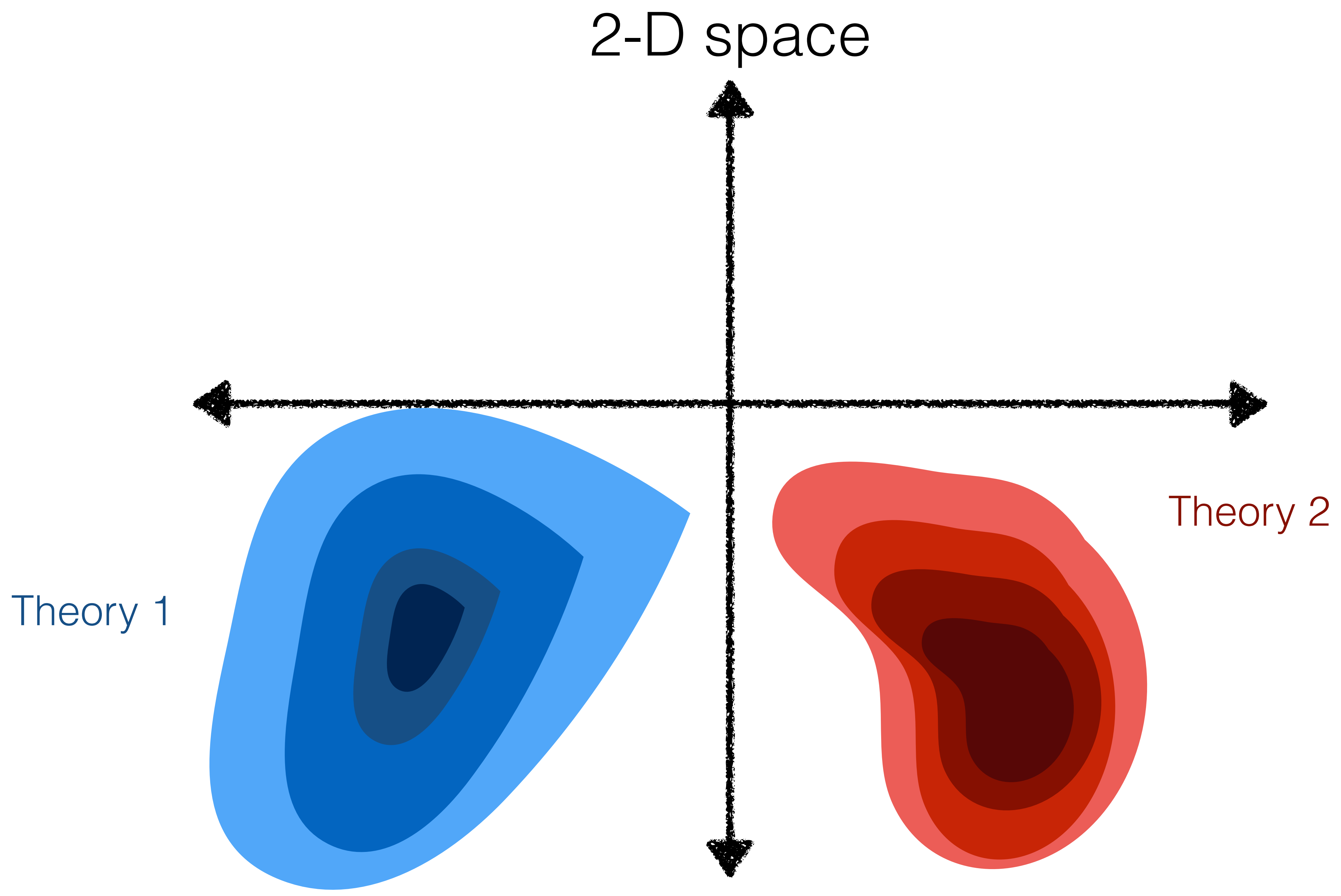
Quantum interference:



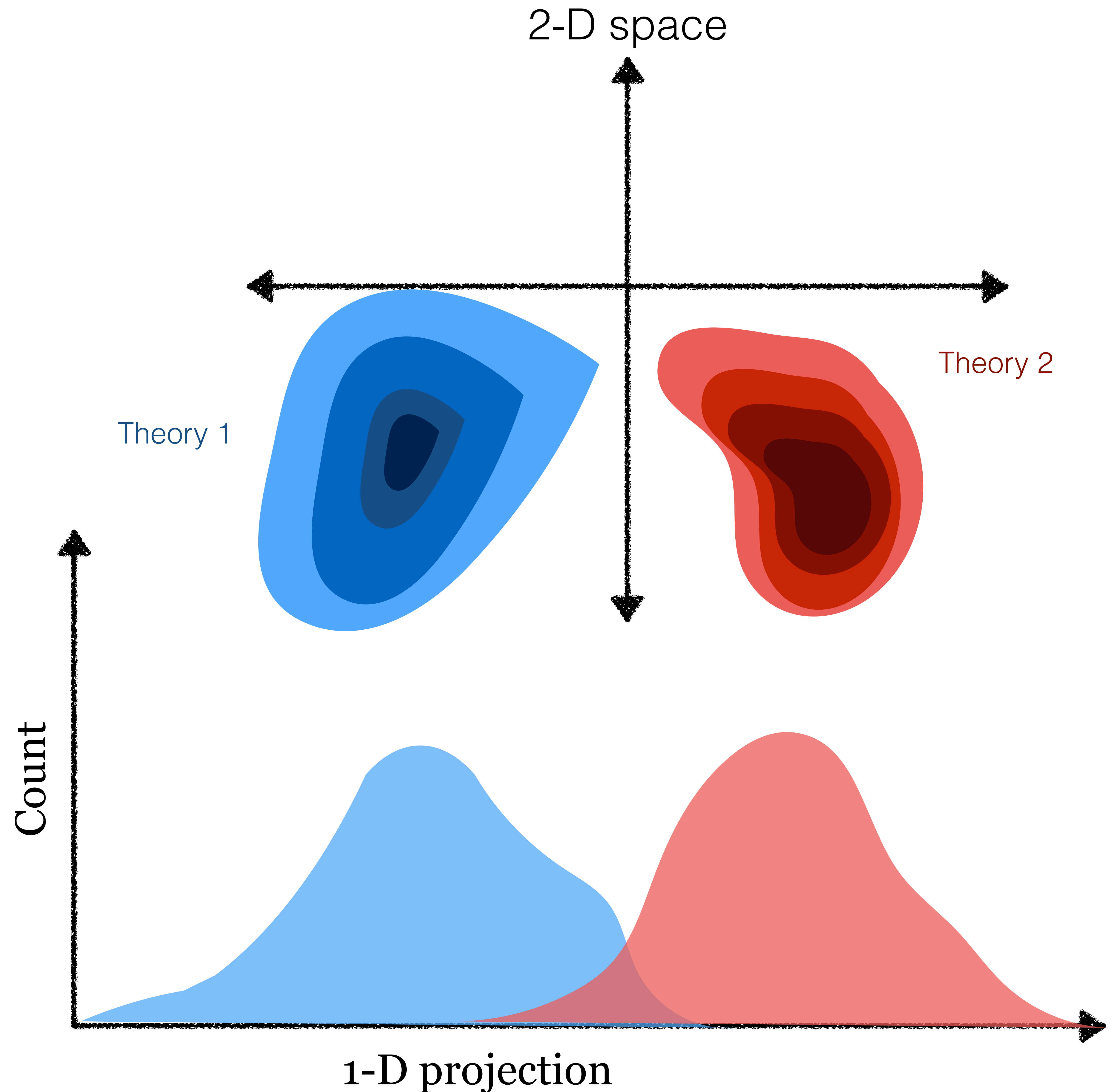
Data can no longer be summarised in 1D histogram (see Ghosh et al: hal-02971995(p172)) !

The problem with one-dimensional summaries…

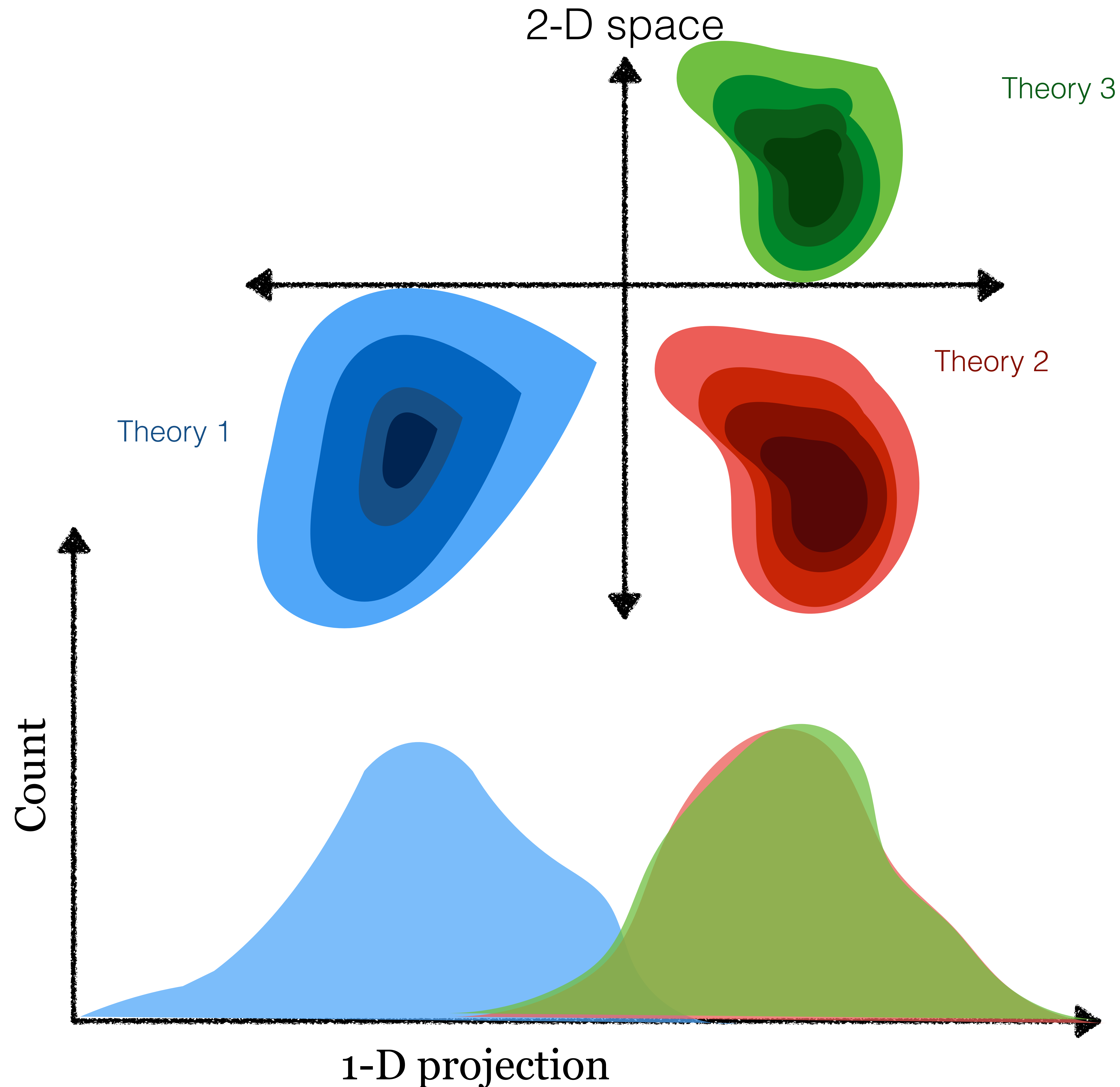
The problem with one-dimensional summaries…



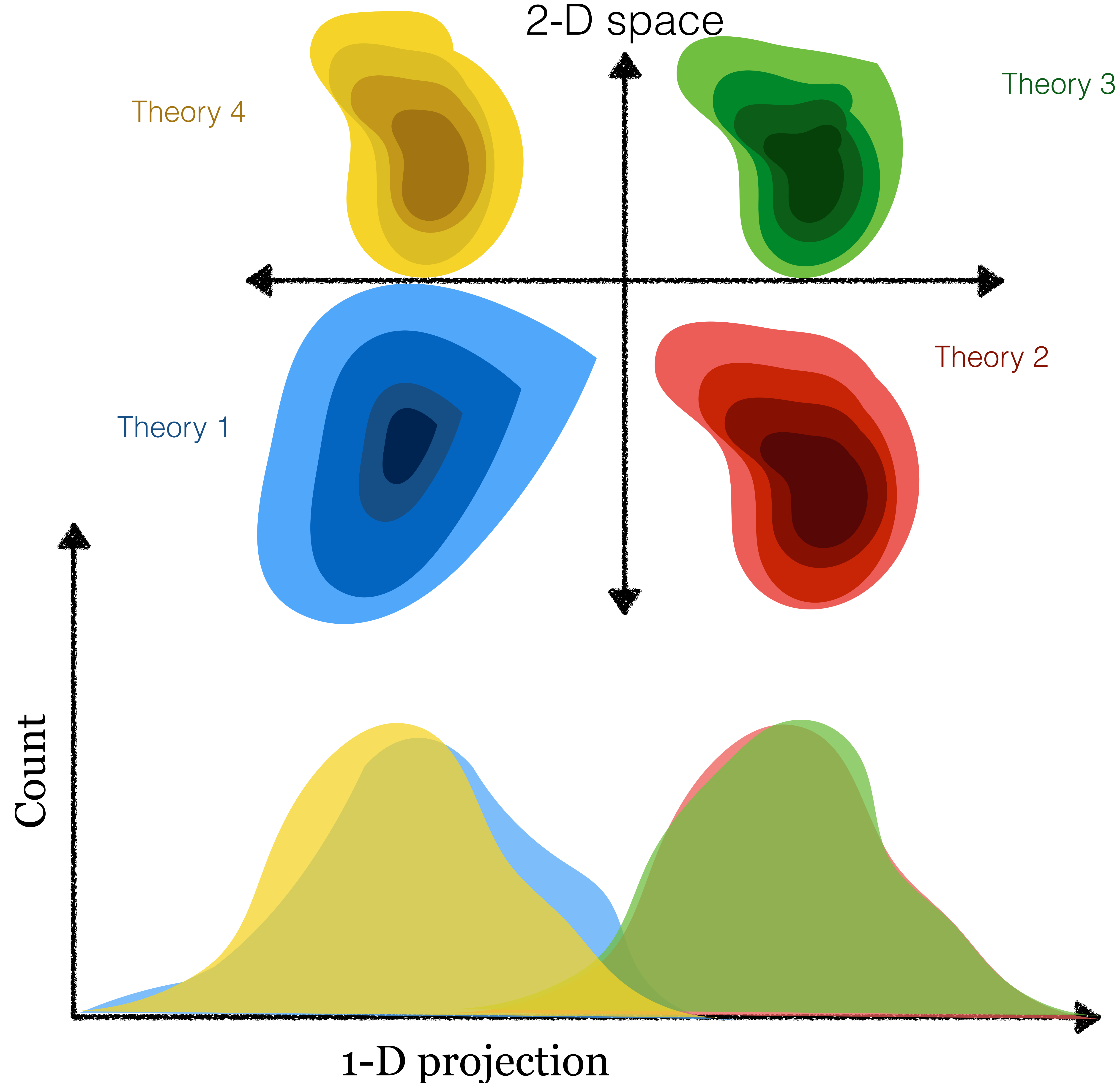
The problem with one-dimensional summaries…



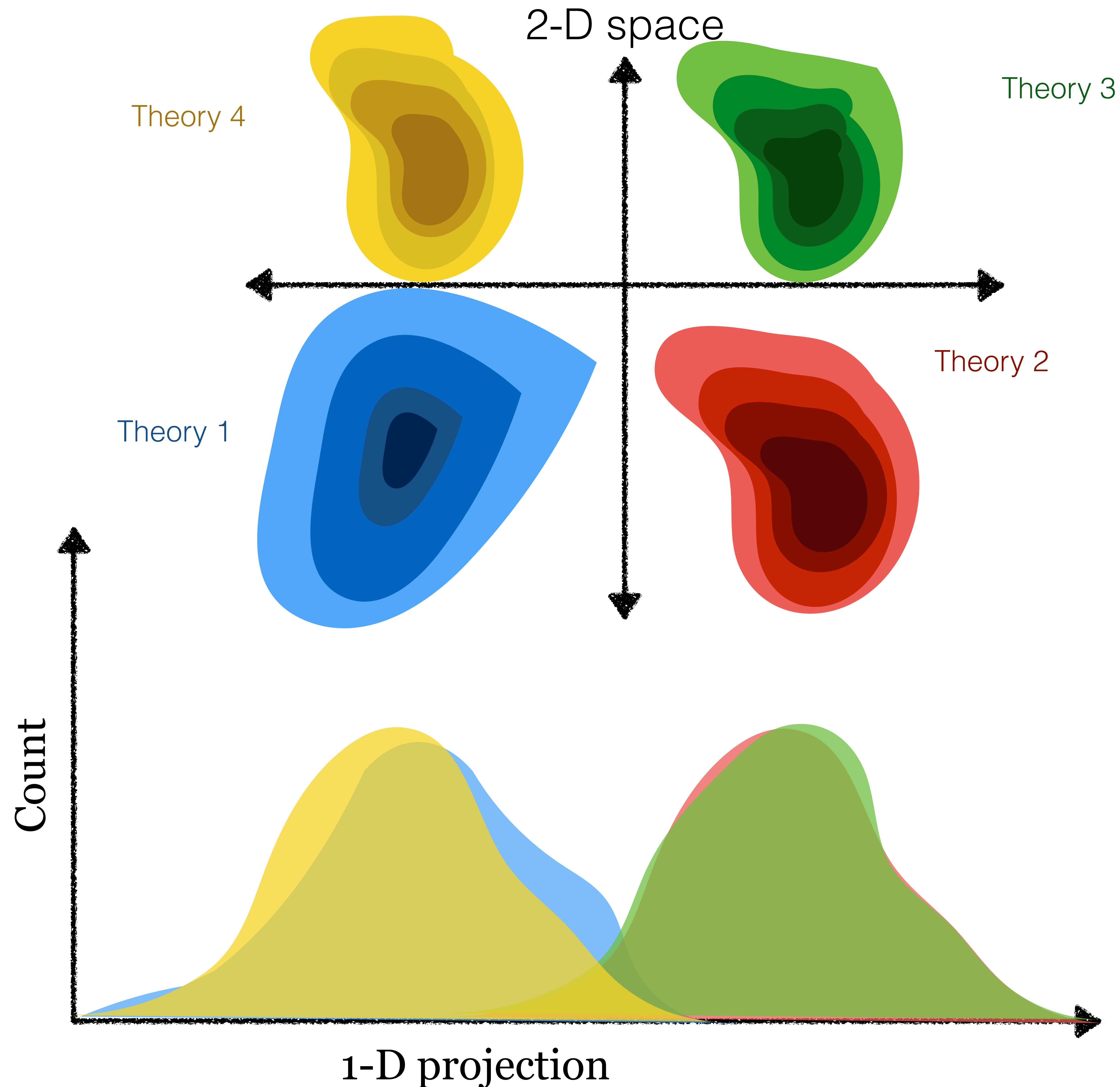
The problem with one-dimensional summaries…



The problem with one-dimensional summaries…



The problem with one-dimensional summaries...



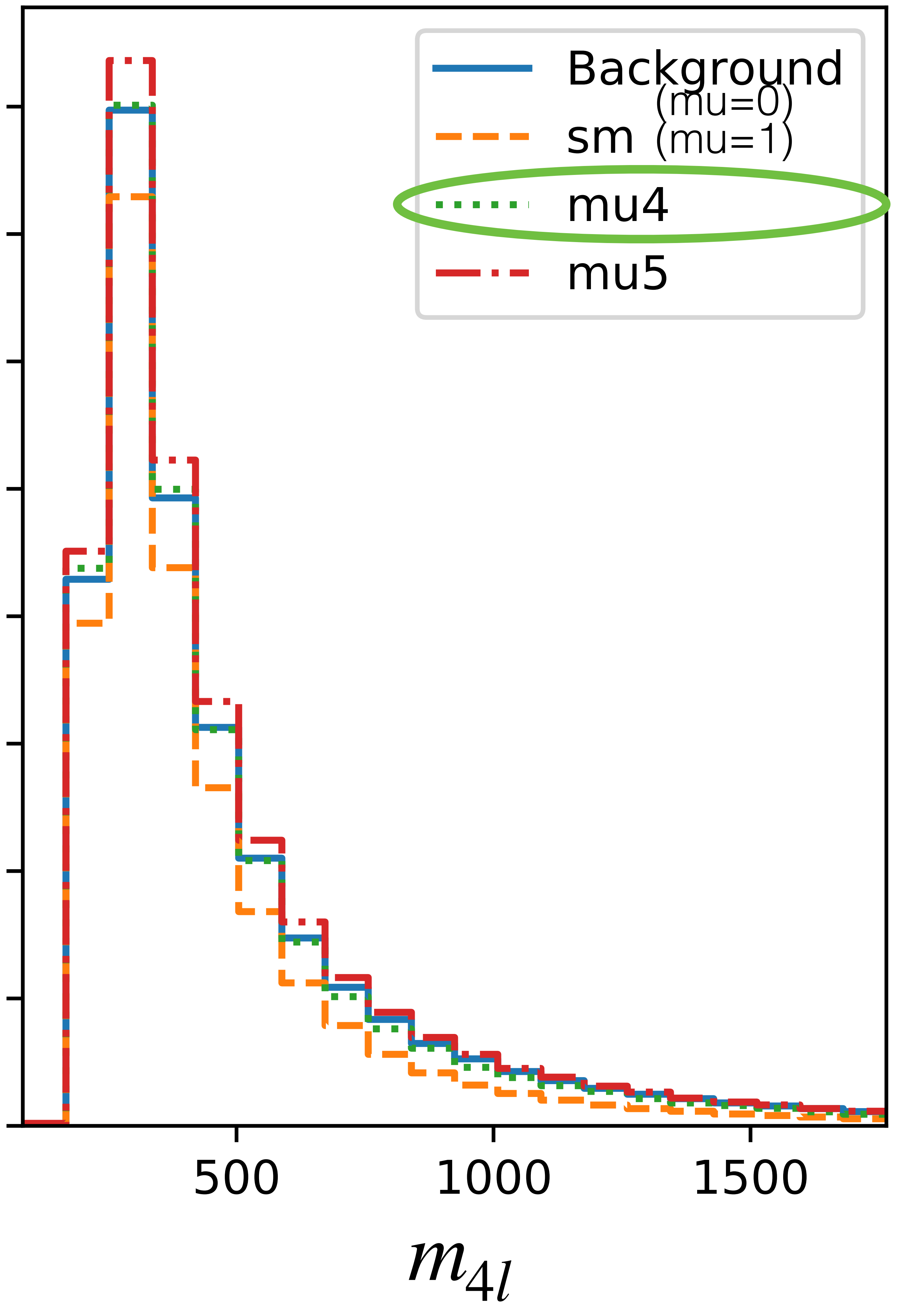
- Clearly separable in 2-D
- No 1-D summary statistic may contain all the information needed to optimally test all theory hypotheses!
- Valuable to have high-dimensional view of data

No single observable captures all information in Higgs width study

Signal-background-inference simulations: MG + Pythia

No single observable captures all information in Higgs width study

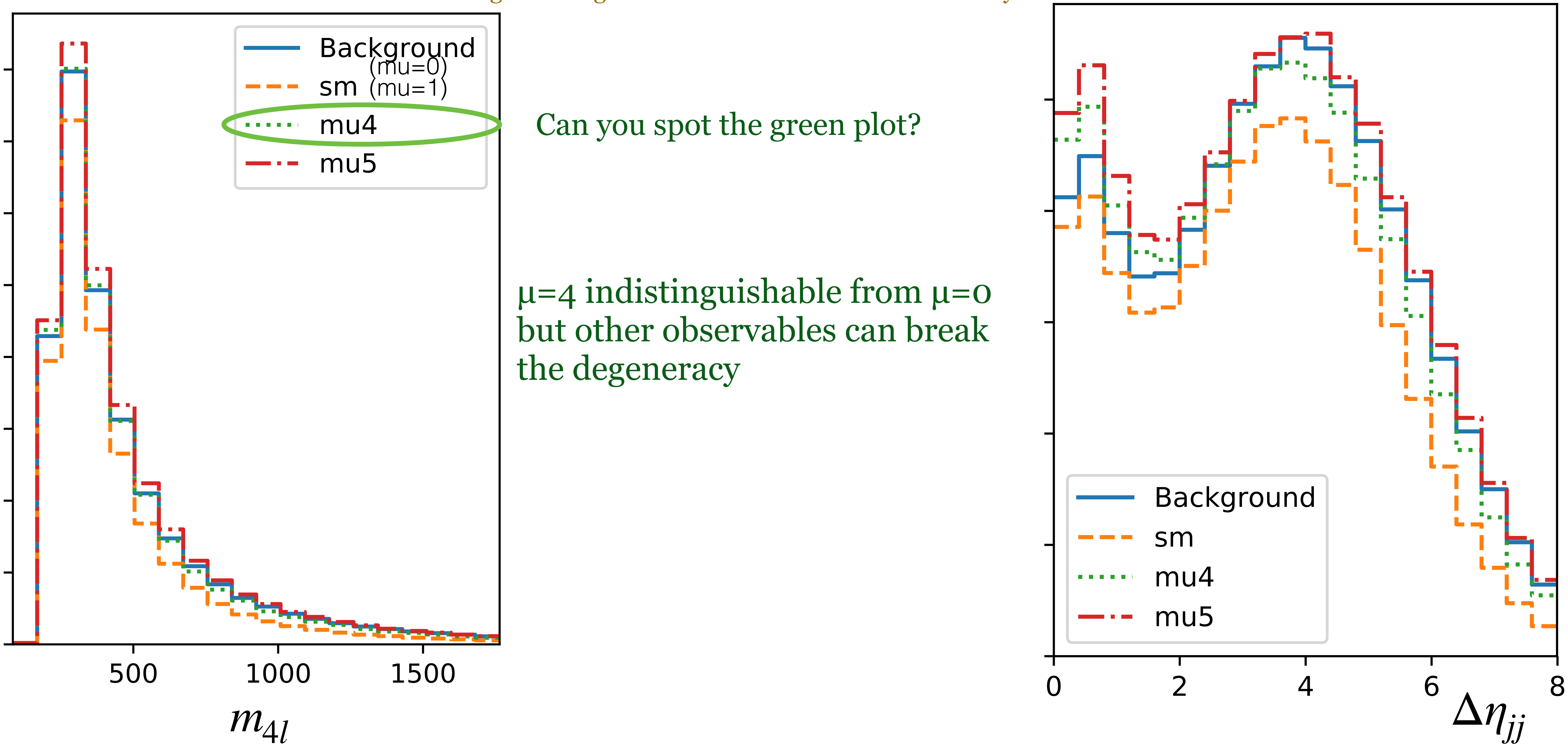
Signal-background-inference simulations: MG + Pythia



Can you spot the green plot?

No single observable captures all information in Higgs width study

Signal-background-inference simulations: MG + Pythia



Can you spot the green plot?

μ=4 indistinguishable from μ=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?

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

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

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

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

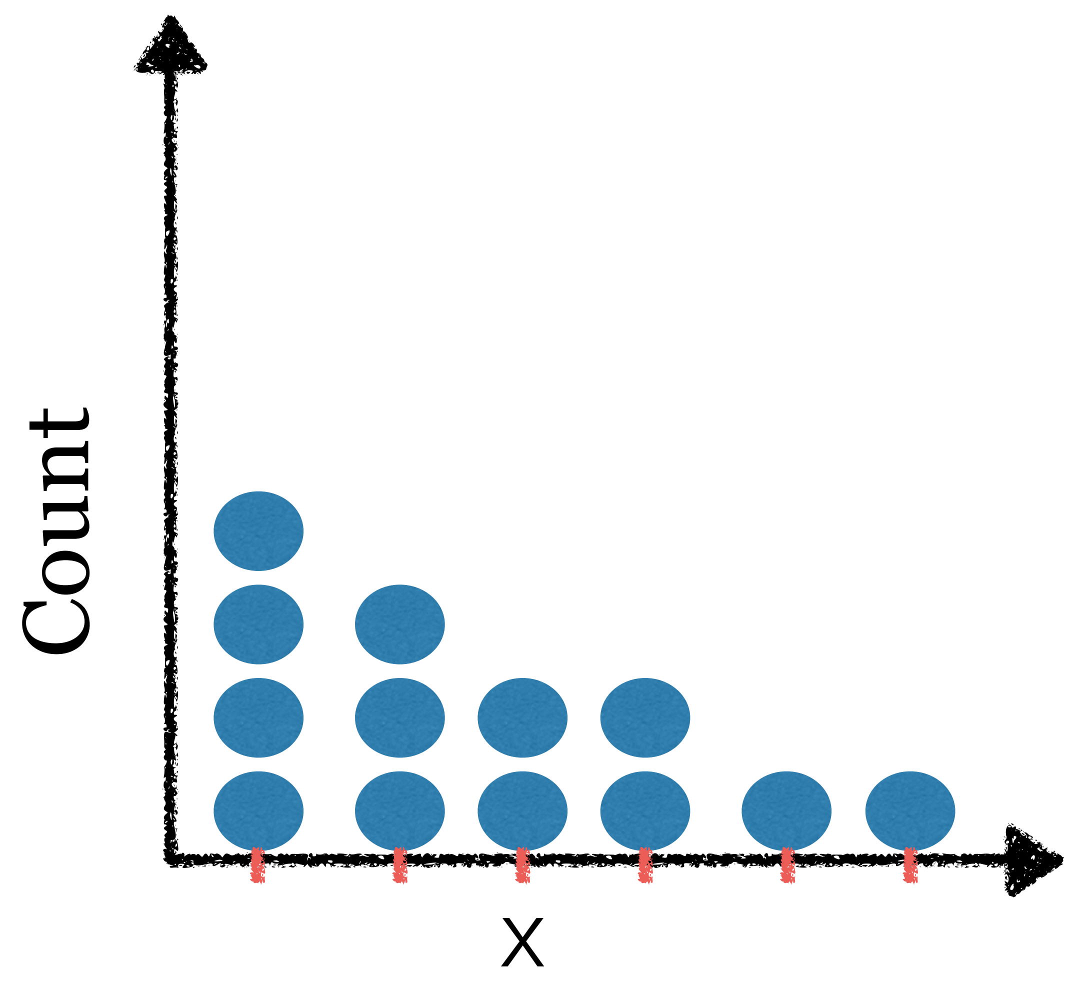
11

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

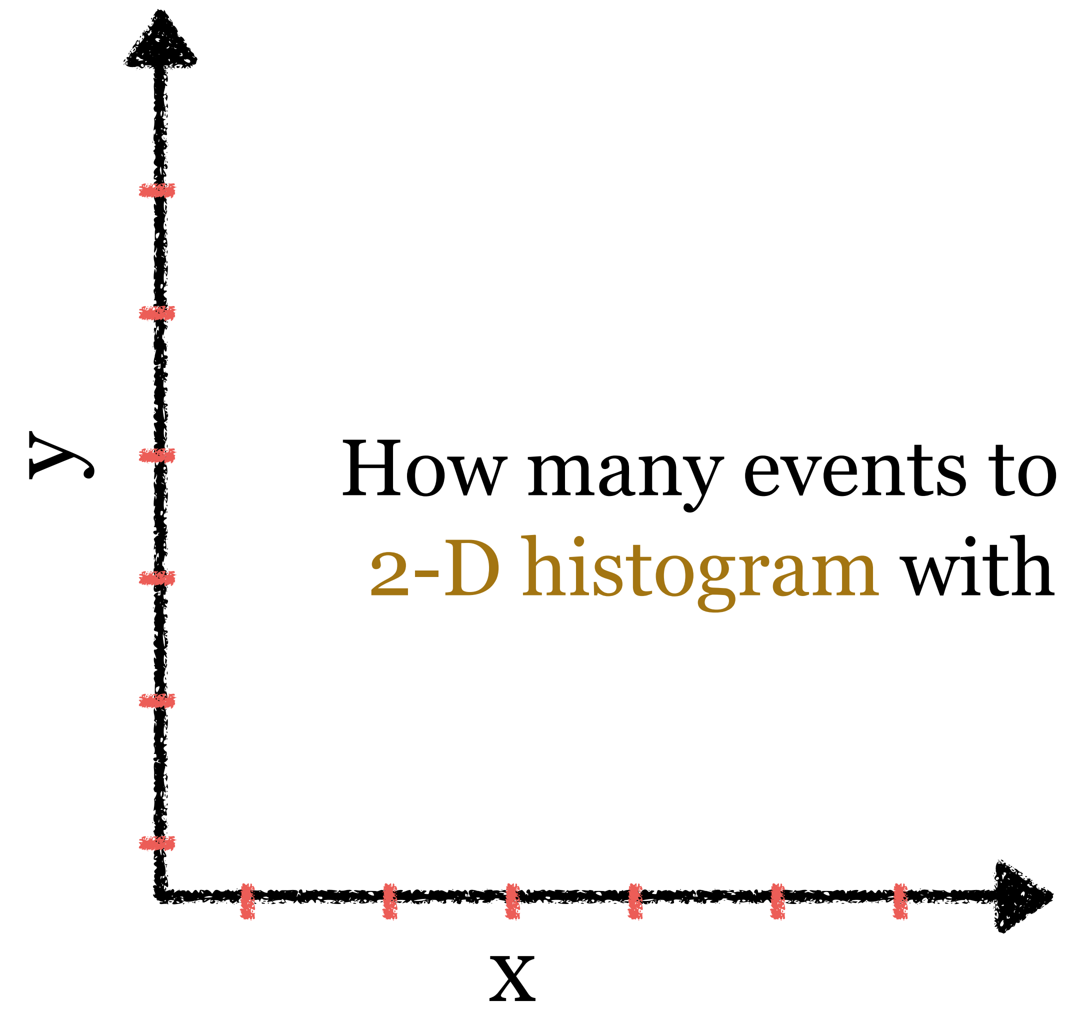
Also does not generalise to an arbitrary theory parameter θ, (eg. Effective Field Theory parameters)

Can we modify the LHC analysis methodology to design near-optimal analyse for the general case?

But probability density estimation in higher dimensions is hard…



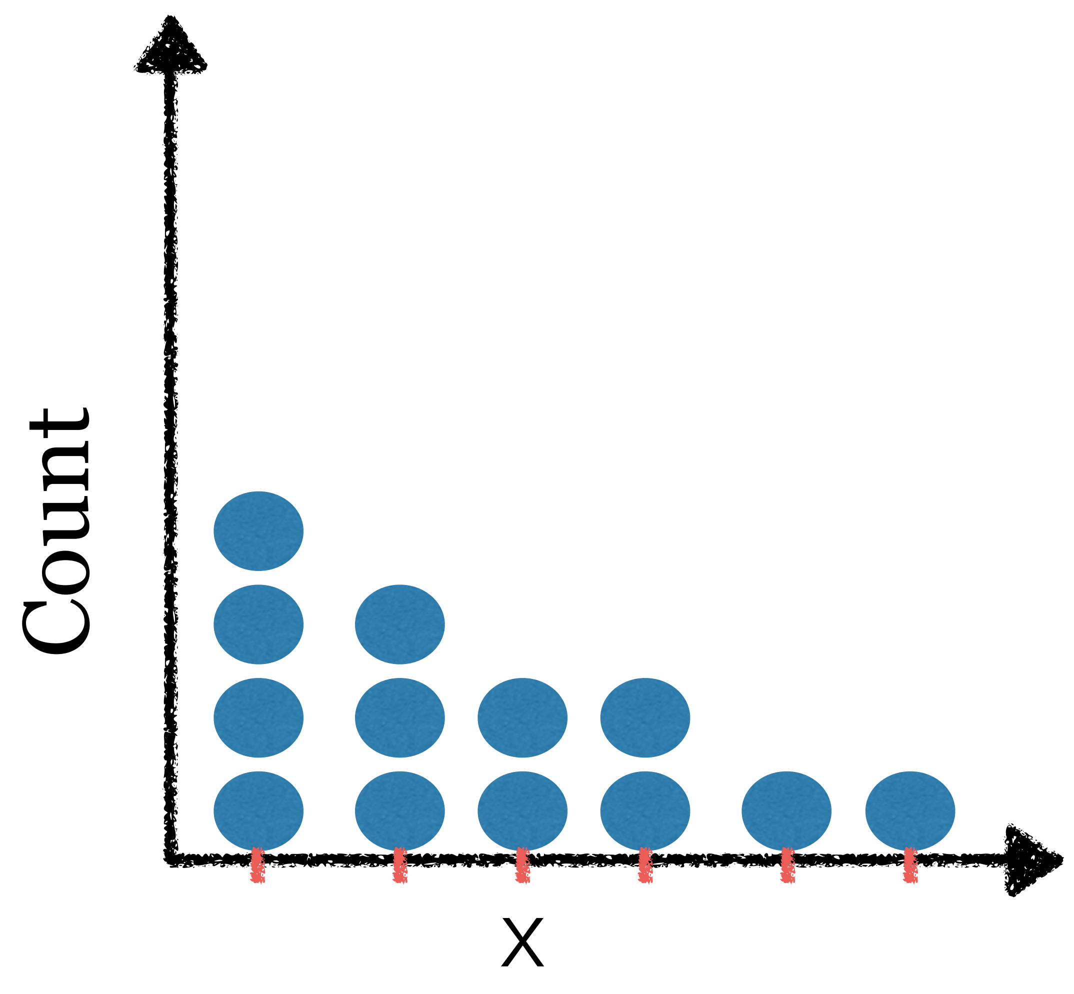
1-D histogram with 6 bins: few events enough to populate it



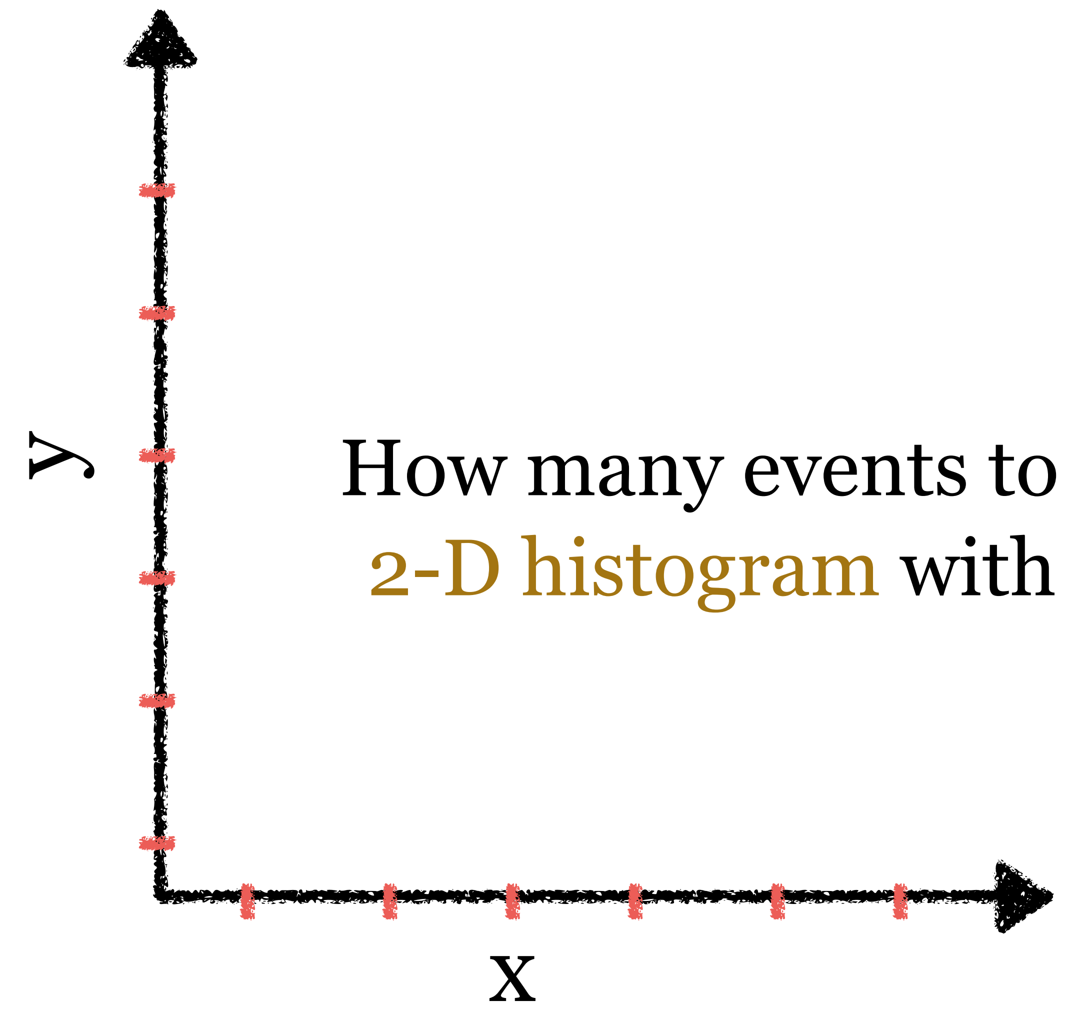
How many events to populate 2-D histogram with 6^2 bins ?

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

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 ?

Cranmer et al: arXiv:1506.02169

Neural networks can give us the likelihood ratios we need

Approximating Likelihood Ratios with Calibrated Discriminative Classifiers

Kyle Cranmer[1], Juan Pavez[2], and Gilles Louppe[1]
[1]New York University
[2]Federico Santa María University

March 21, 2016

Abstract

In many fields of science, generalized likelihood ratio tests are established tools for statistical inference. At the same time, it has become increasingly common that a simulator (or generative model) is used to describe complex processes that tie parameters θ of an underlying theory and measurement apparatus to high-dimensional observations $\mathbf{x} \in \mathbb{R}^p$. However, simulator often do not provide a way to evaluate the likelihood function for a given observation $\mathbf{x}$, which motivates a new class of likelihood-free inference algorithms. In this paper, we show that likelihood ratios are invariant under a specific class of dimensionality reduction maps $\mathbb{R}^p \mapsto \mathbb{R}$. As a direct consequence, we show that discriminative classifiers can be used to approximate the generalized likelihood ratio statistic when only a generative model for the data is available. This leads to a new machine learning-based approach to likelihood-free inference that is complementary to Approximate Bayesian Computation, and which does not require a prior on the model parameters. Experimental results on artificial problems with known exact likelihoods illustrate the potential of the proposed method.

Keywords: likelihood ratio, likelihood-free inference, classification, particle physics, surrogate model

Cranmer et al: arXiv:1506.02169

Neural networks can give us the likelihood ratios

$$\mathscr{L}(\mu \mid \mathscr{D}) = p(\mathscr{D} \mid \mu)$$

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

We want to compare likelihoods: $$\frac{p(\mathscr{D} \mid \mu)}{p(\mathscr{D} \mid ref)}$$

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

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

Which contains all the information required for the likelihood ratio:

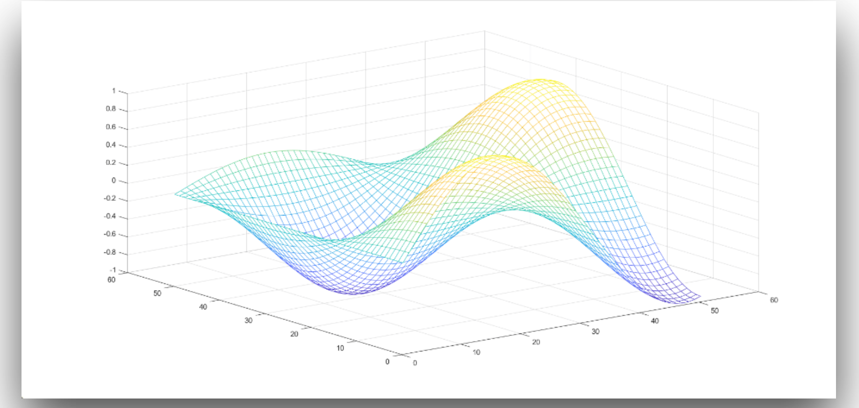
$$\frac{p(x_i \mid \mu_1)}{p(x_i \mid ref)} = \frac{s(x_i)}{1 - s(x_i)}$$

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

Cranmer et al: arXiv:1506.02169

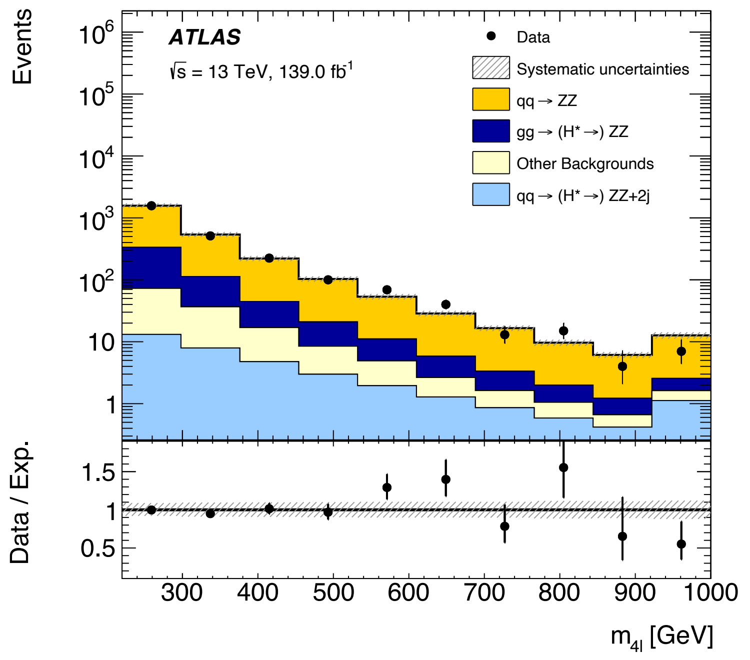
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram

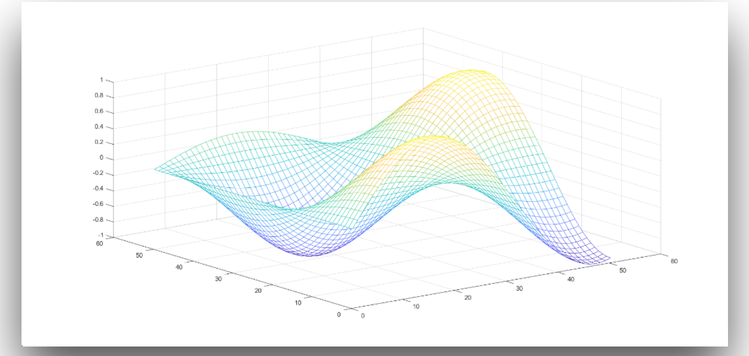


Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

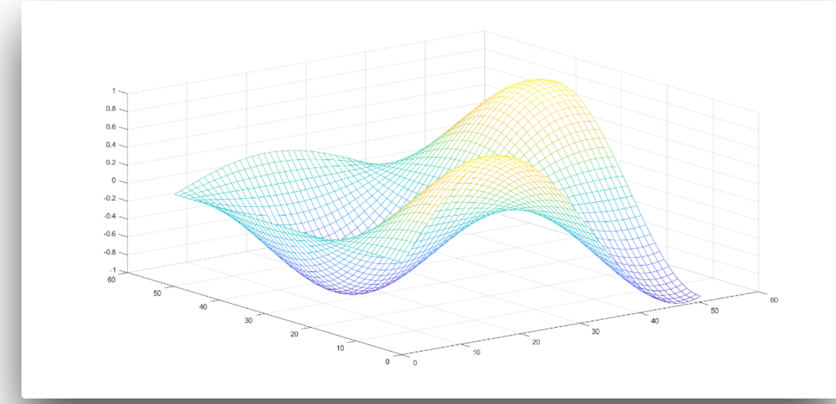
Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

Cranmer et al: [arXiv:1506.02169](https://arxiv.org/abs/1506.02169)

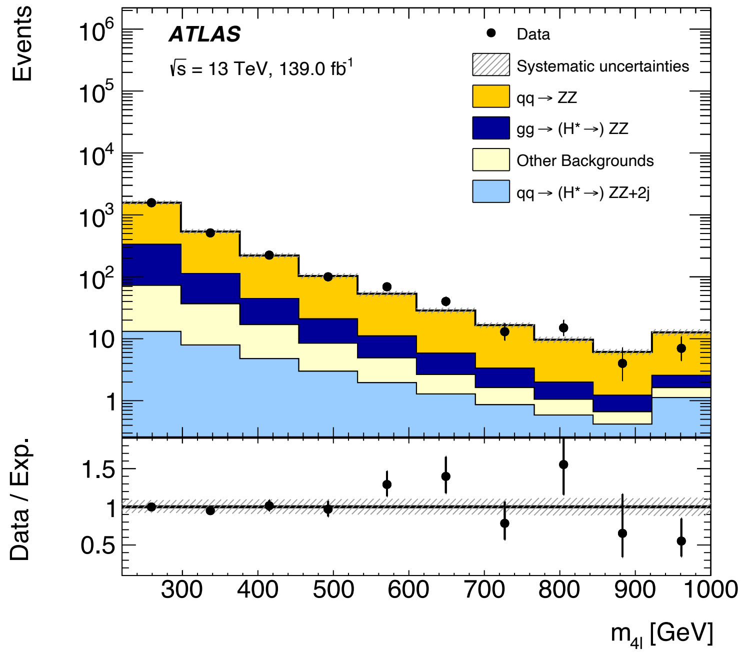
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram



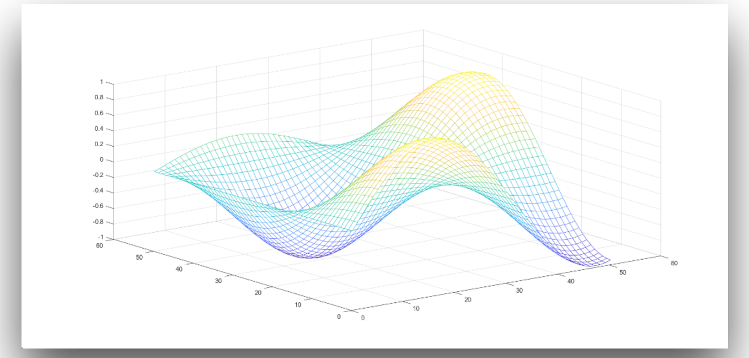
Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

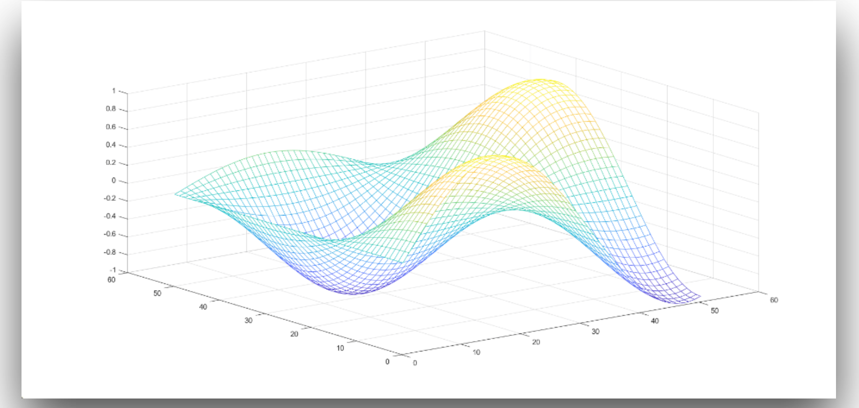
Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

Cranmer et al: arXiv:1506.02169

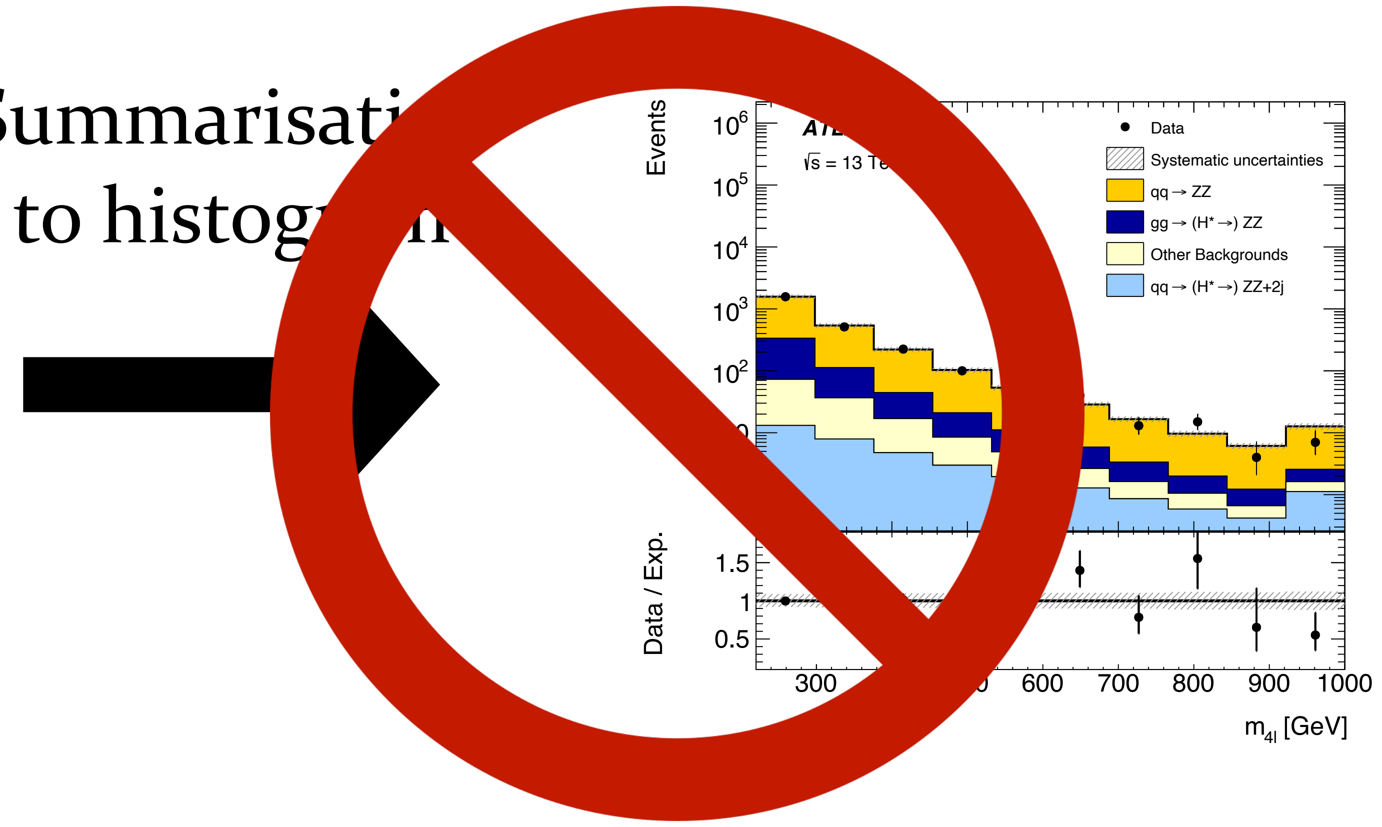
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisati
to histog



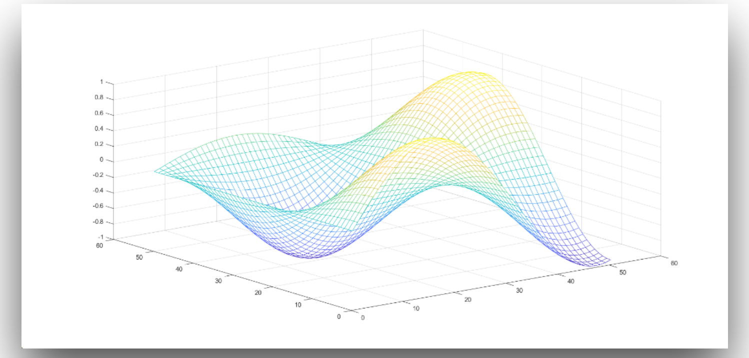
Summary Histogram → Statistical Fit → Likelihood $\mathcal{L}(\mu_1 | \mathcal{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

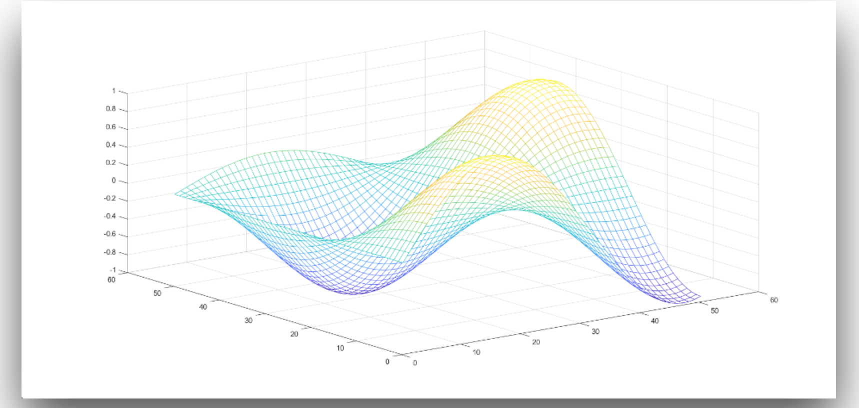
Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathcal{L}(\mu_1 | \mathcal{D})}{\mathcal{L}(ref | \mathcal{D})}\right)$

μ_1 → Neural Network

Cranmer et al: arXiv:1506.02169

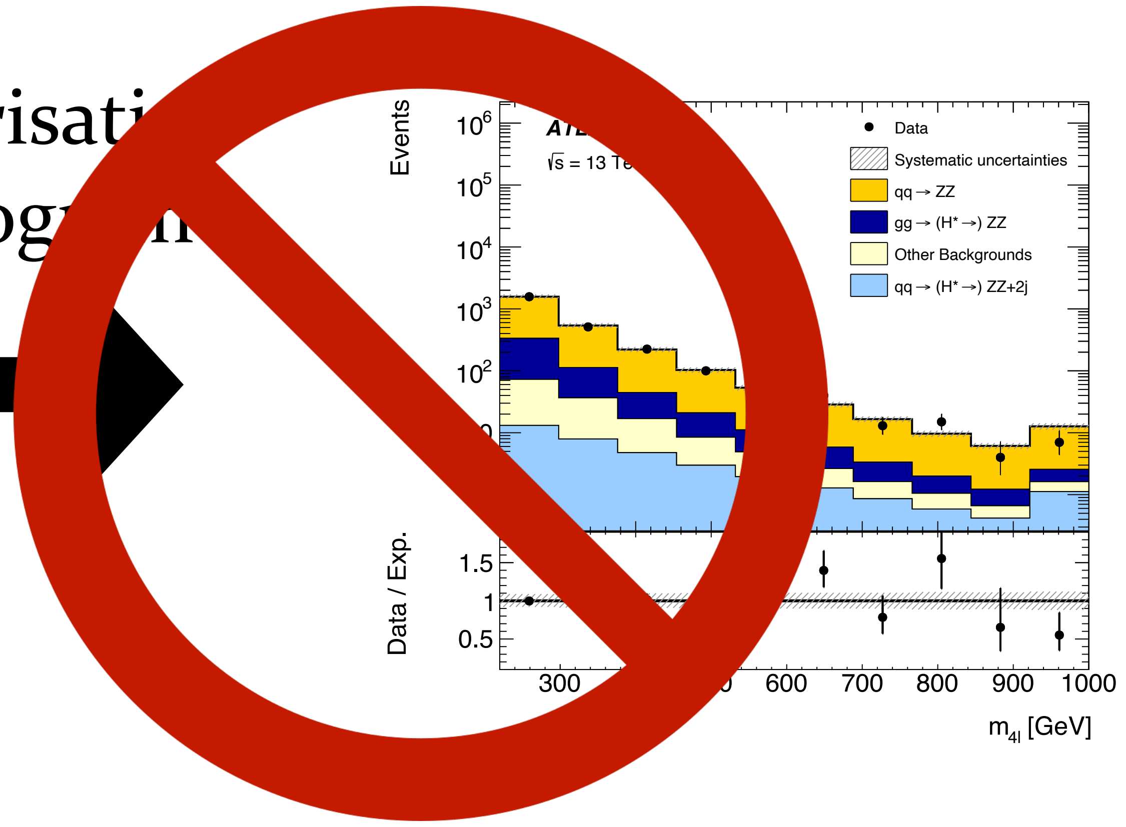
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisati
to histog



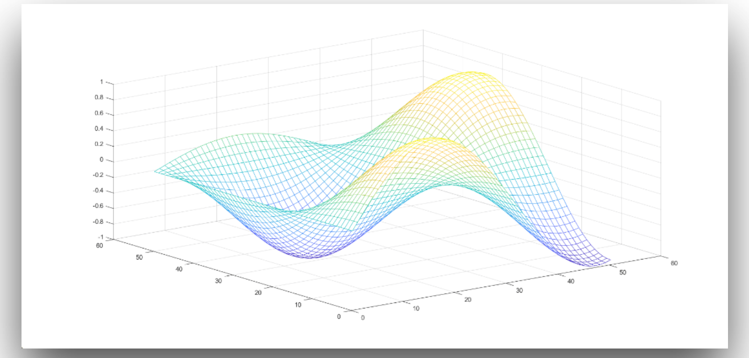
Summary Histogram → **Statistical Fit** → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → **Statistical Fit**

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

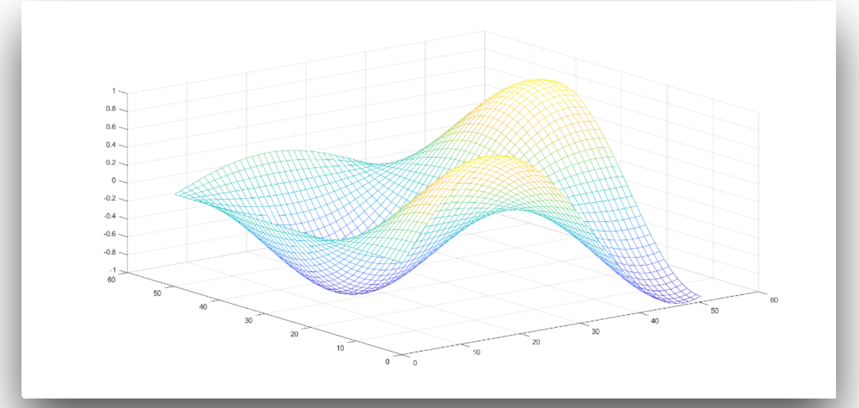
Obs Data → **Neural Network** → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → **Neural Network**

Hypothesis μ_1

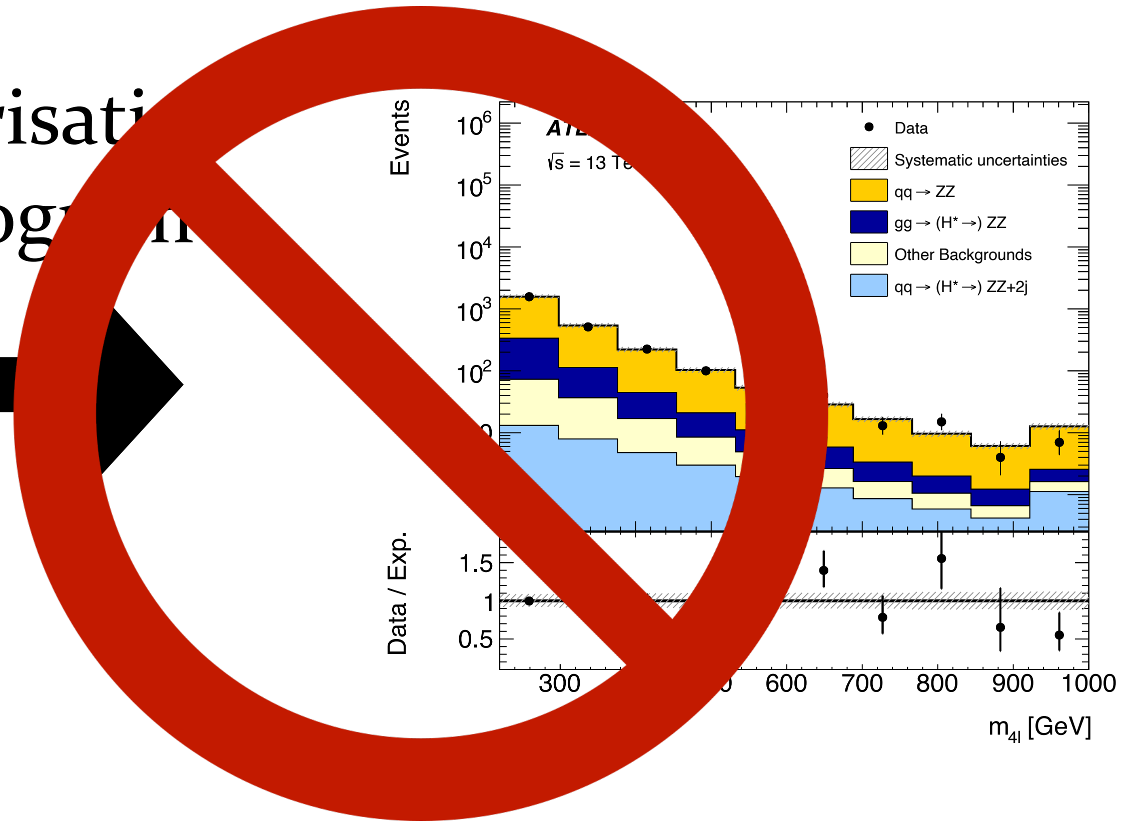
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisati
to histog

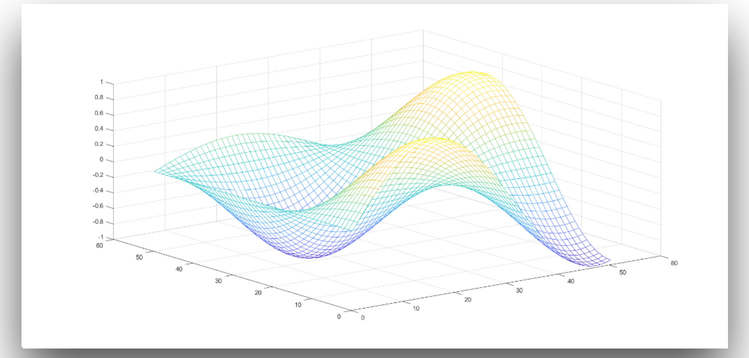


Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1 μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

Hypothesis μ_1

Train on simulations, apply on data

NSBI for Higgs width in proof-of-concept phenomenology study

Expected sensitivity

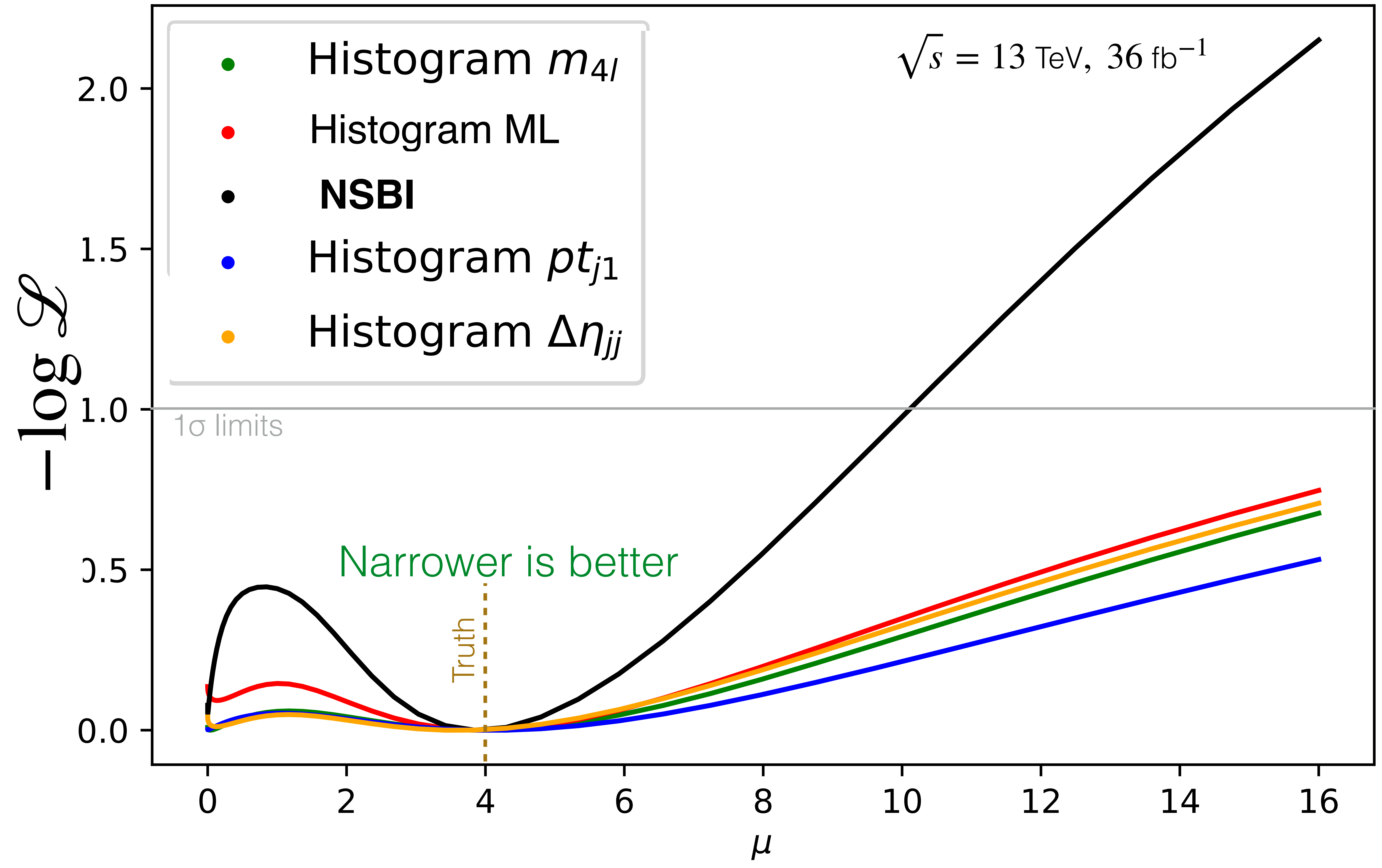
hal-02971995v3 (p172): Ghosh & Rousseau, Thesis: Ghosh

(Beyond Standard Model value)

NSBI for Higgs width in proof-of-concept phenomenology study

Expected sensitivity

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



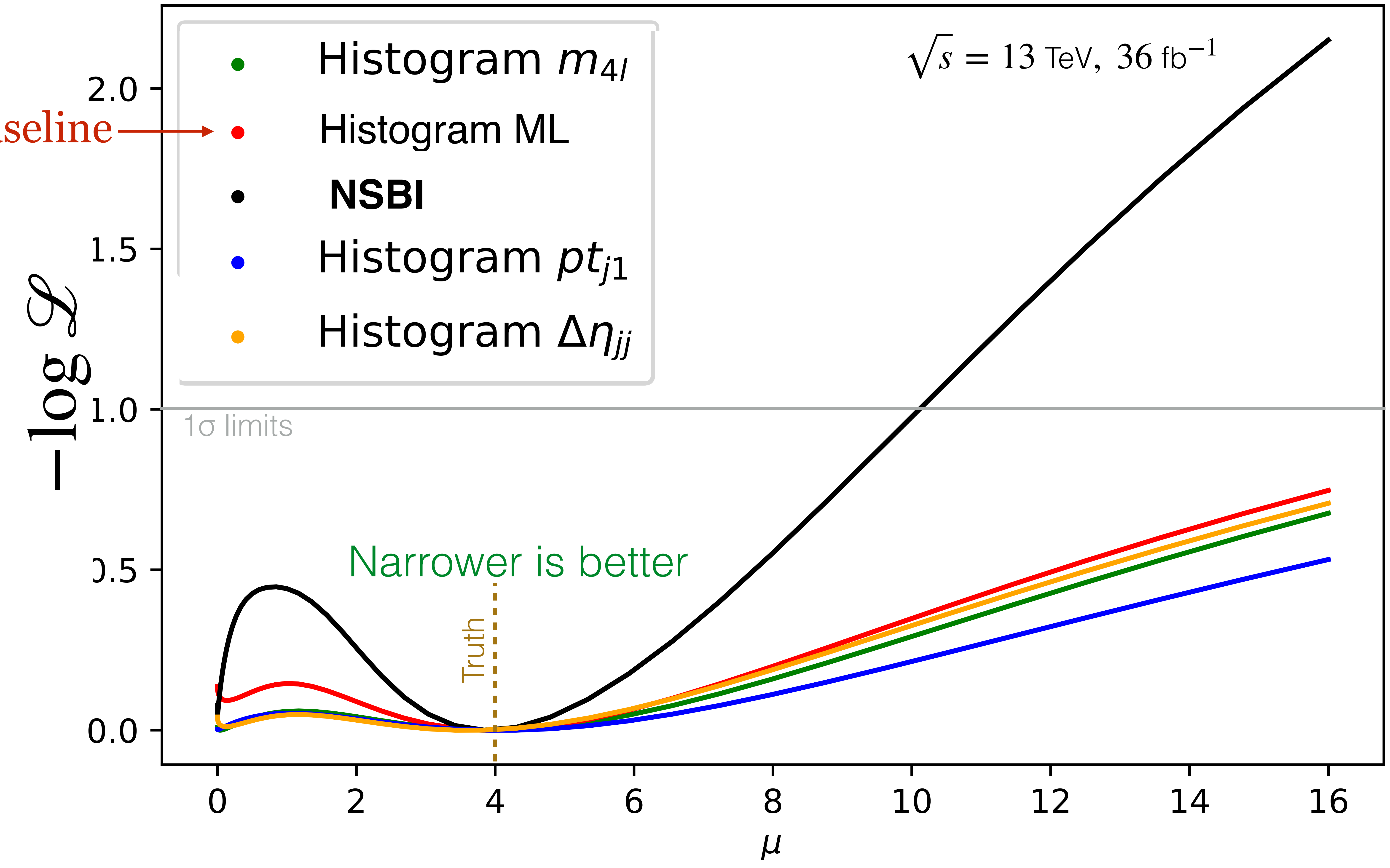
(Beyond Standard Model value) $\mu = 4, \text{ without rate}$

NSBI for Higgs width in proof-of-concept phenomenology study

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

<u>Expected sensitivity</u>

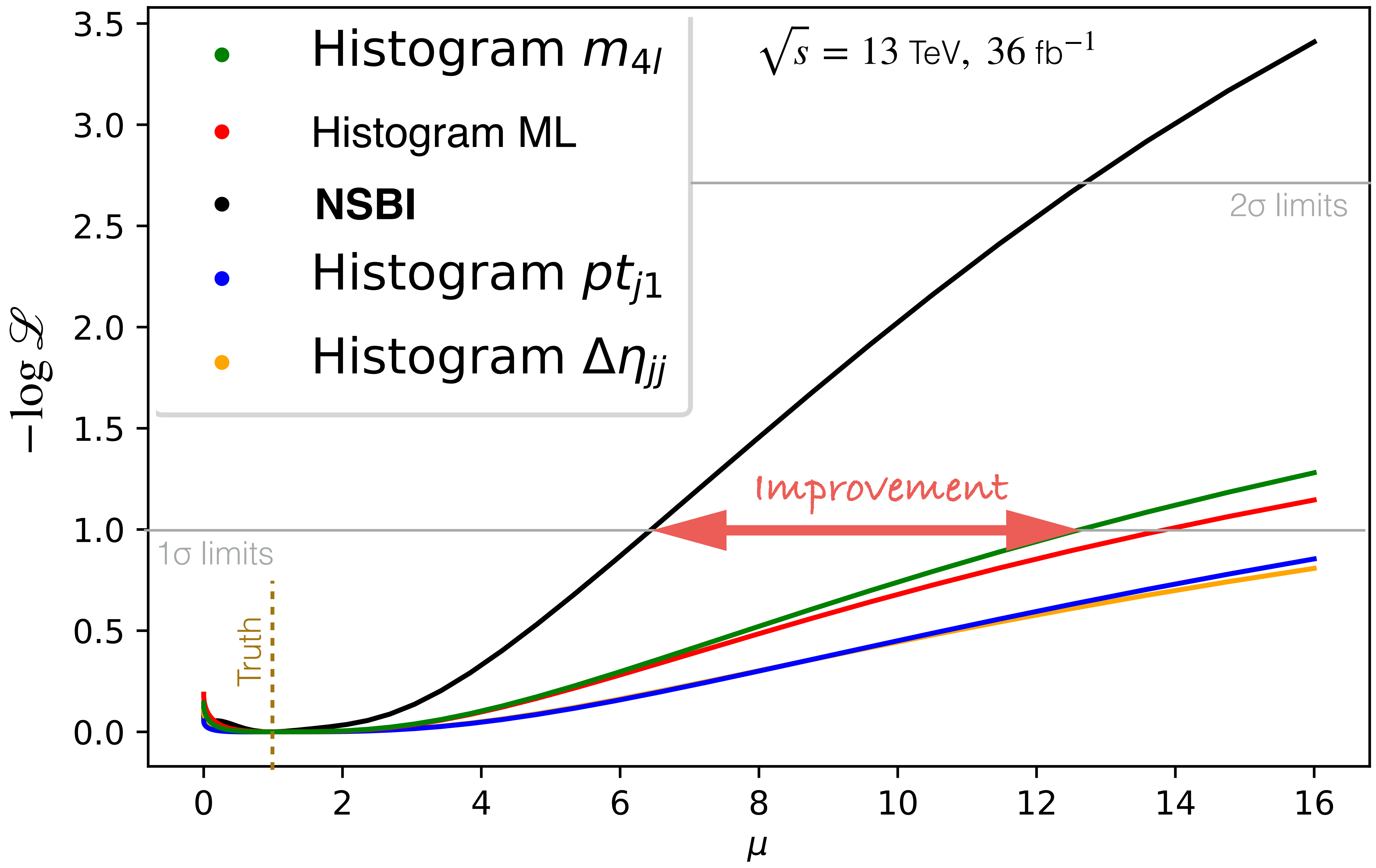
“Traditional ML” baseline



(Beyond Standard Model value) $\mu = 4, \text{ without rate}$

Expected improvement for Standard Model

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



Exciting gains promised!

SM, without rate

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Throwing toys in a per-event analysis

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Throwing toys in a per-event analysis

How frequentists ensure coverage

Solved!

Open problems to extend to full ATLAS

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Rep. Prog. Phys.

CERN-EP-2024-305
December 3, 2024

An implementation of neural simulation-based inference for parameter estimation in ATLAS

The ATLAS Collaboration

Neural simulation-based inference is a powerful class of machine-learning-based methods for statistical inference that naturally handles high-dimensional parameter estimation without the need to bin data into low-dimensional summary histograms. Such methods are promising for a range of measurements, including at the Large Hadron Collider, where no single observable may be optimal to scan over the entire theoretical phase space under consideration, or where binning data into histograms could result in a loss of sensitivity. This work develops a neural simulation-based inference framework for statistical inference, using neural networks to estimate probability density ratios, which enables the application to a full-scale analysis. It incorporates a large number of systematic uncertainties, quantifies the uncertainty due to the finite number of events in training samples, develops a method to construct confidence intervals, and demonstrates a series of intermediate diagnostic checks that can be performed to validate the robustness of the method. As an example, the power and feasibility of the method are assessed on simulated data for a simplified version of an off-shell Higgs boson couplings measurement in the four-lepton final states. This approach represents an extension to the standard statistical methodology used by the experiments at the Large Hadron Collider, and can benefit many physics analyses.

Presented at CHEP 2024, Higgs 2024

Solved!

Open problems to extend to full ATLAS

Applied on Run2 data, superseding previous ATLAS paper on same data !

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Rep. Prog. Phys.

CERN-EP-2024-305
December 3, 2024

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Submitted to: Rep. Prog. Phys.

CERN-EP-2024-298
December 3, 2024

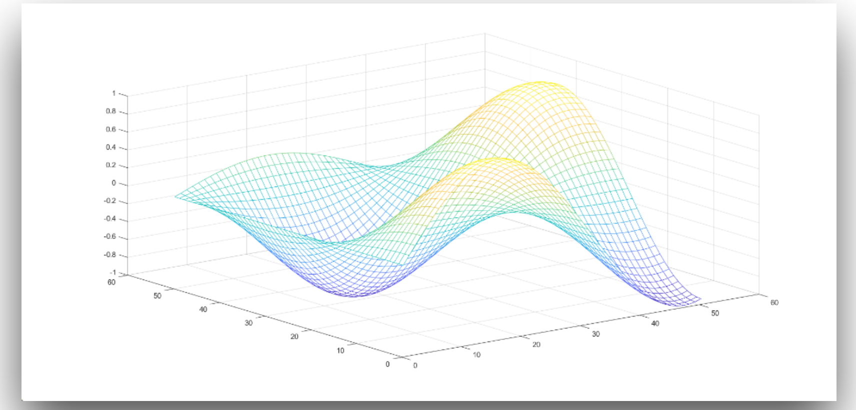
Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV pp collisions with the ATLAS detector

The ATLAS Collaboration

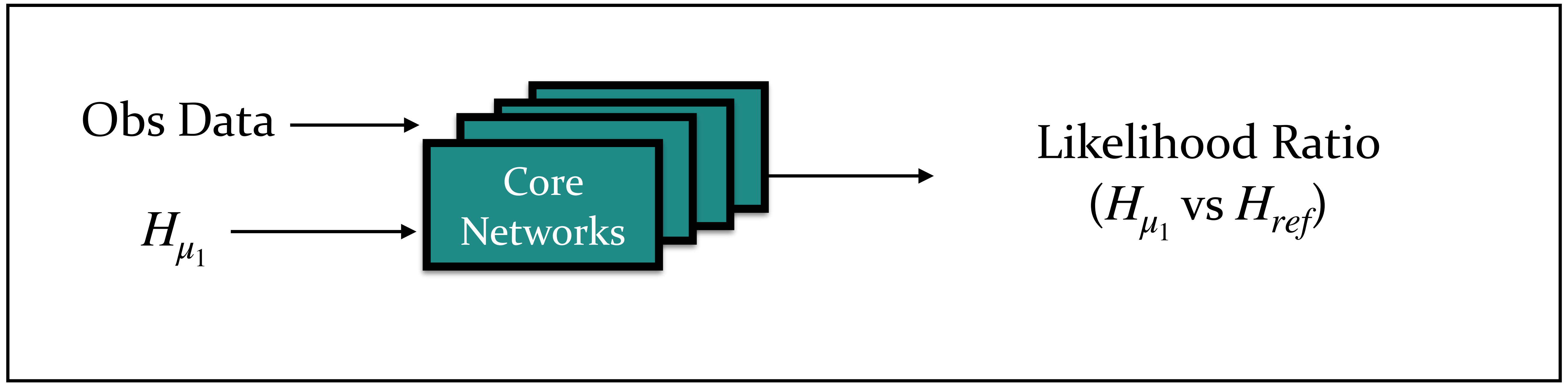
A measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel is presented. The measurement uses 140 fb^{-1} of proton–proton collisions at $\sqrt{s}$ = 13 TeV collected by the ATLAS detector at the Large Hadron Collider and supersedes the previous result in this decay channel using the same dataset. The data analysis is performed using a neural simulation-based inference method, which builds per-event likelihood ratios using neural networks. The observed (expected) off-shell Higgs boson production signal strength in the $ZZ \to 4\ell$ decay channel at 68% CL is $0.87^{+0.75}_{-0.54}$ ($1.00^{+1.04}_{-0.95}$). The evidence for off-shell Higgs boson production using the $ZZ \to 4\ell$ decay channel has an observed (expected) significance of 2.5σ (1.3σ). The expected result represents a significant improvement relative to that of the previous analysis of the same dataset, which obtained an expected significance of 0.5σ. When combined with the most recent ATLAS measurement in the $ZZ \to 2\ell 2\nu$ decay channel, the evidence for off-shell Higgs boson production has an observed (expected) significance of 3.7σ (2.4σ). The off-shell measurements are combined with the measurement of on-shell Higgs boson production to obtain constraints on the Higgs boson total width. The observed (expected) value of the Higgs boson width at 68% CL is $4.3^{+2.7}_{-1.9}$ ($4.1^{+3.5}_{-3.4}$) MeV.

Presented at CHEP 2024, Higgs 2024

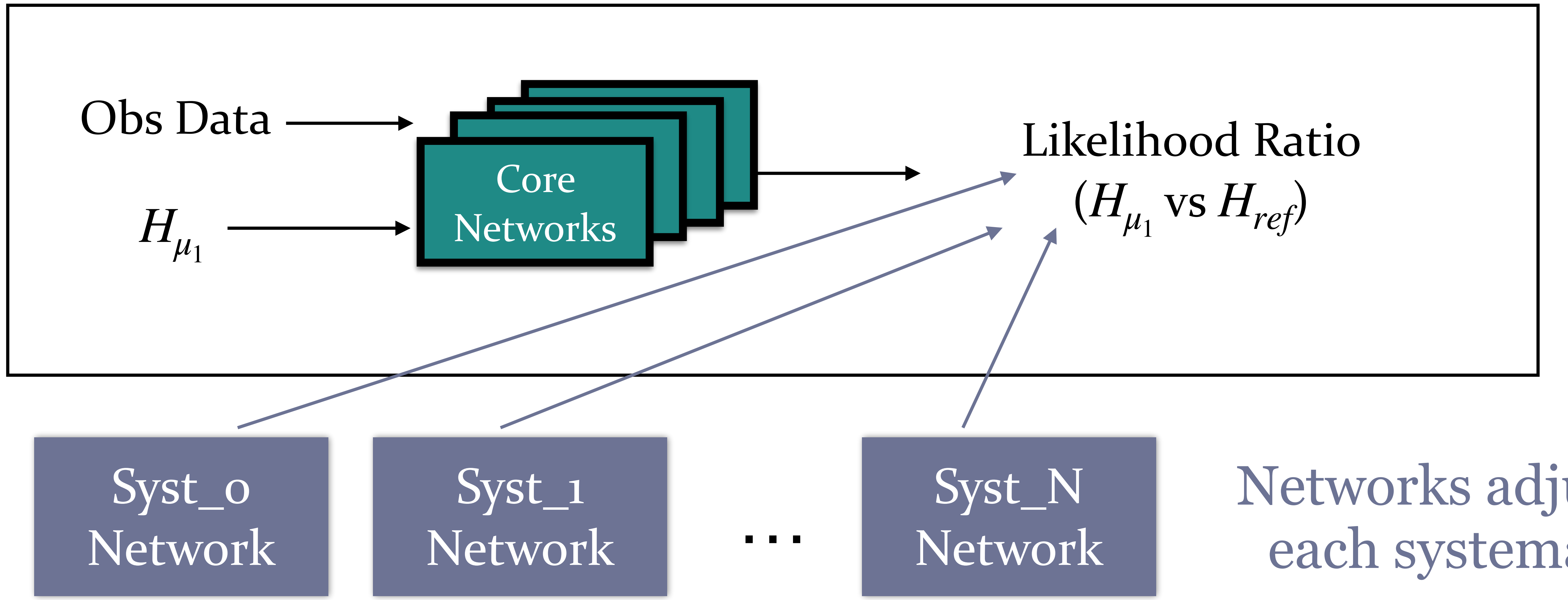
Big picture of full solution developed in ATLAS



$O(16)$ observables

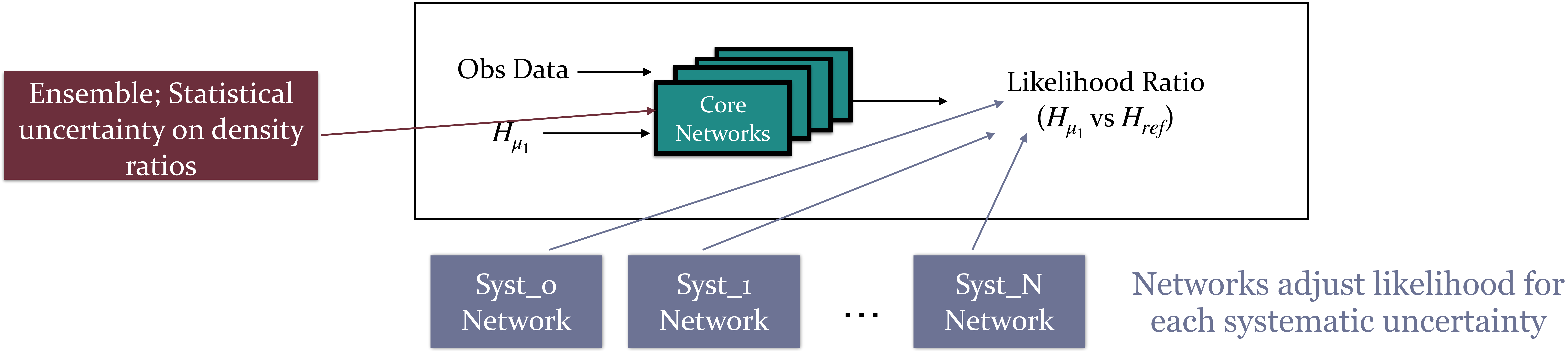


Big picture of full solution developed in ATLAS

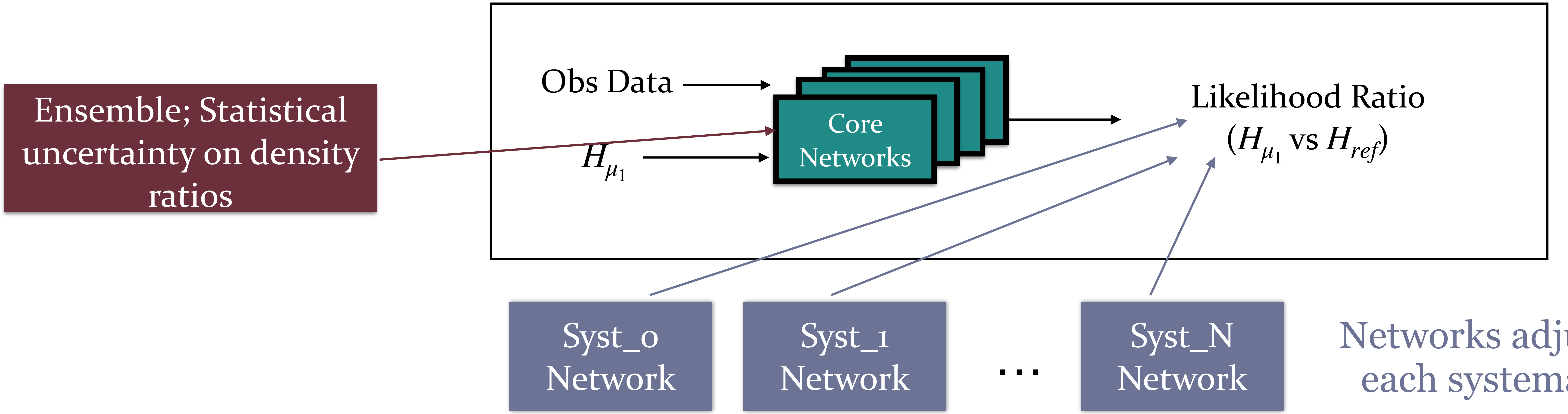


Networks adjust likelihood for each systematic uncertainty

Big picture of full solution developed in ATLAS



Big picture of full solution developed in ATLAS



Networks adjust likelihood for each systematic uncertainty

Training details

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

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Throwing toys in a per-event analysis

Next 2 slides gets a bit technical

Search-Oriented Mixture Model

x_i is 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. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Example use case

Search-Oriented Mixture Model

x_i is 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 is 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

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

Search-Oriented Mixture Model

x_i is 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

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

Search-Oriented Mixture Model

x_i is 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

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

Search-Oriented Mixture Model

x_i is 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

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

Search-Oriented Mixture Model

x_i is 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

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

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

Search-Oriented Mixture Model

x_i is one individual event

General Formula

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

?

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

Estimated using an ensemble of networks

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

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

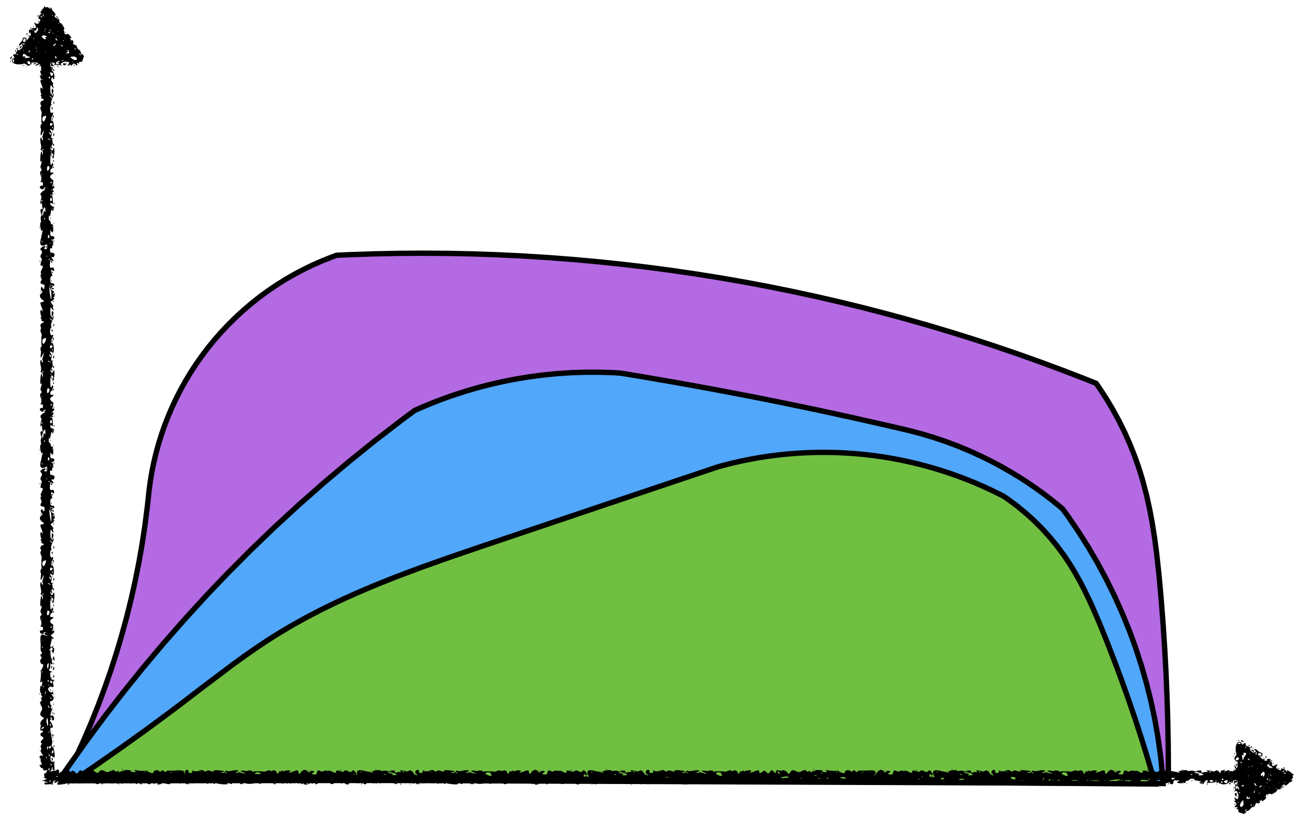
$$\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{p_S(x)} \right]$$

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

H_μ

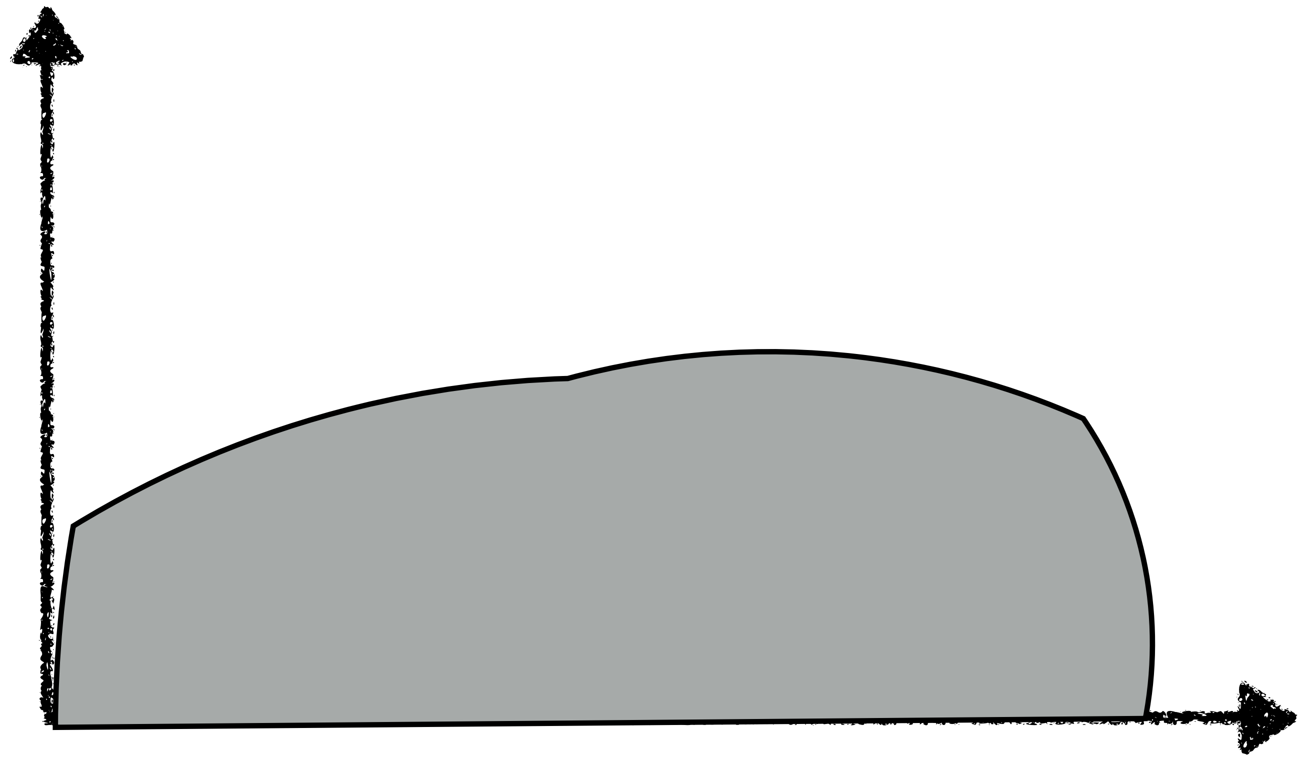
Hypothesis



VS

H_{ref}

Reference hypothesis



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

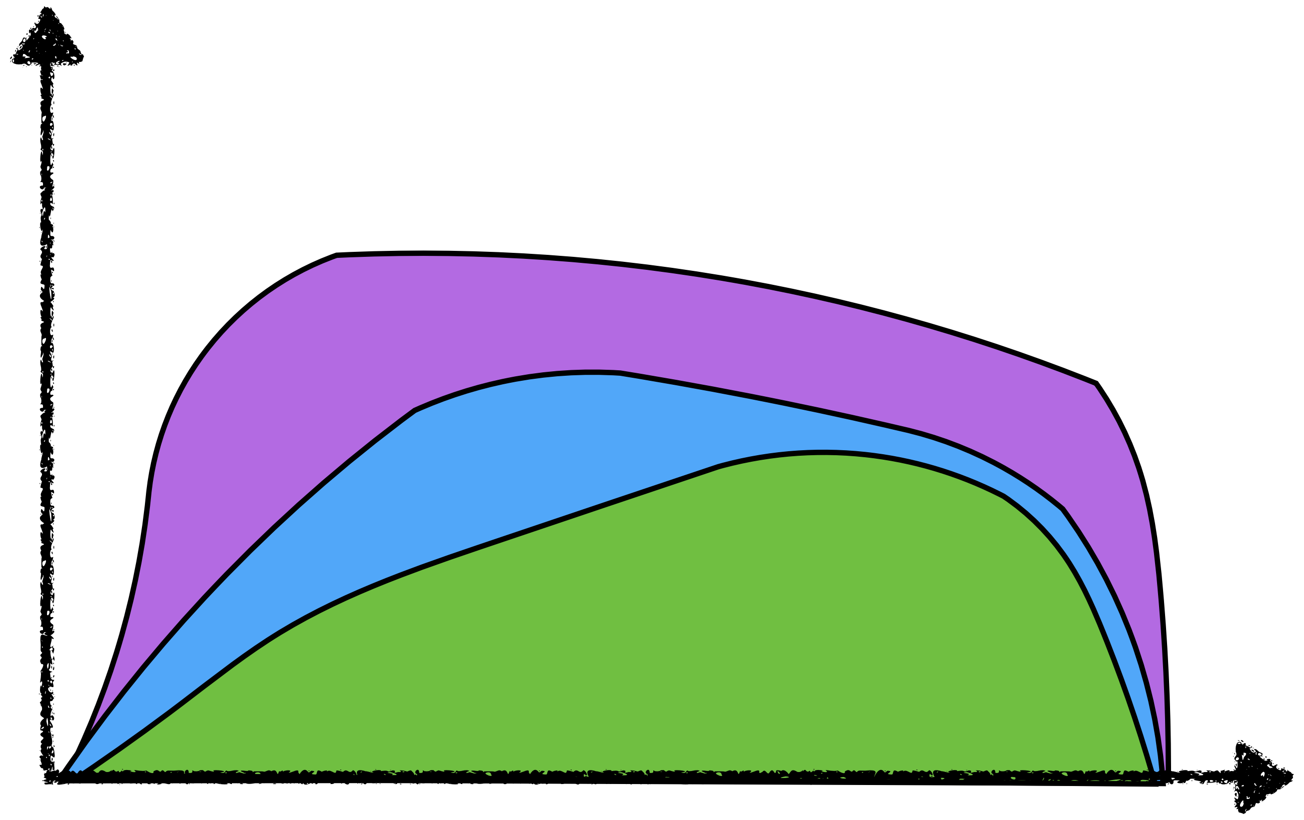
A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

H_μ

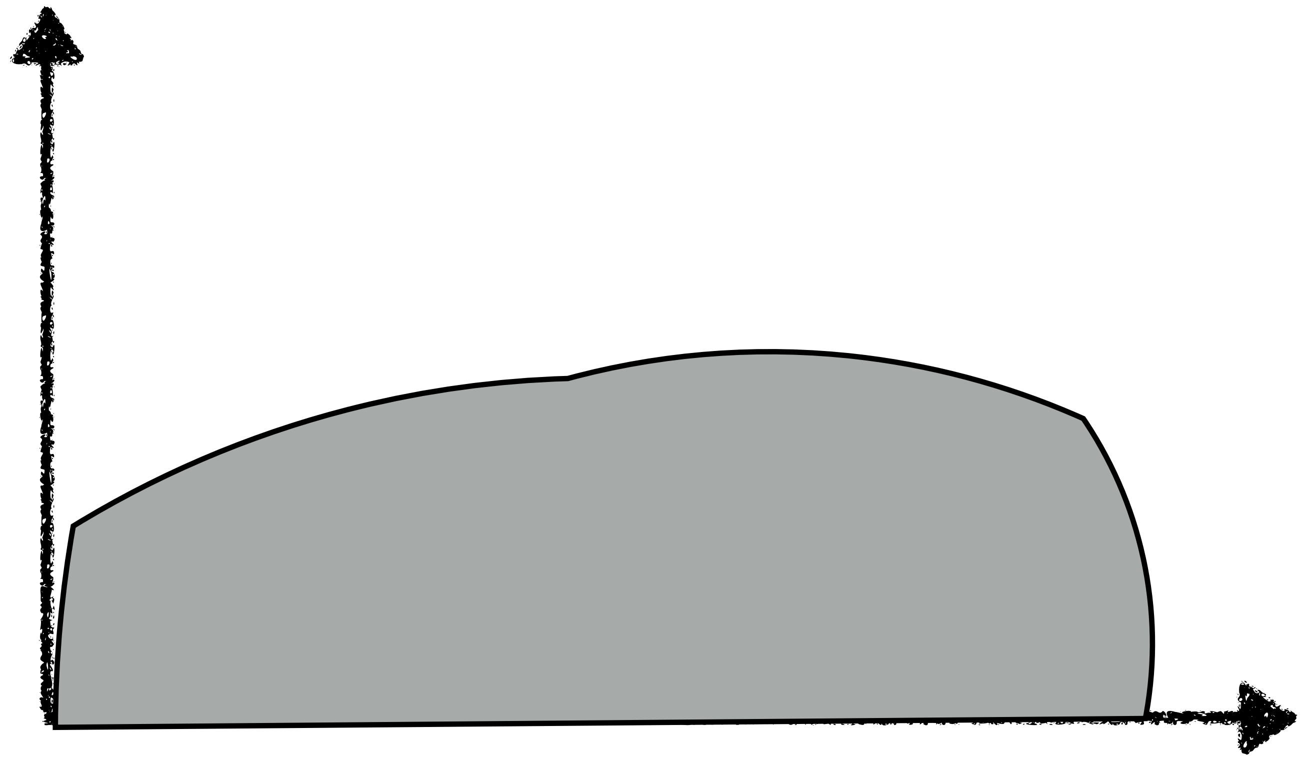
Hypothesis



VS

H_{ref}

Reference hypothesis



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

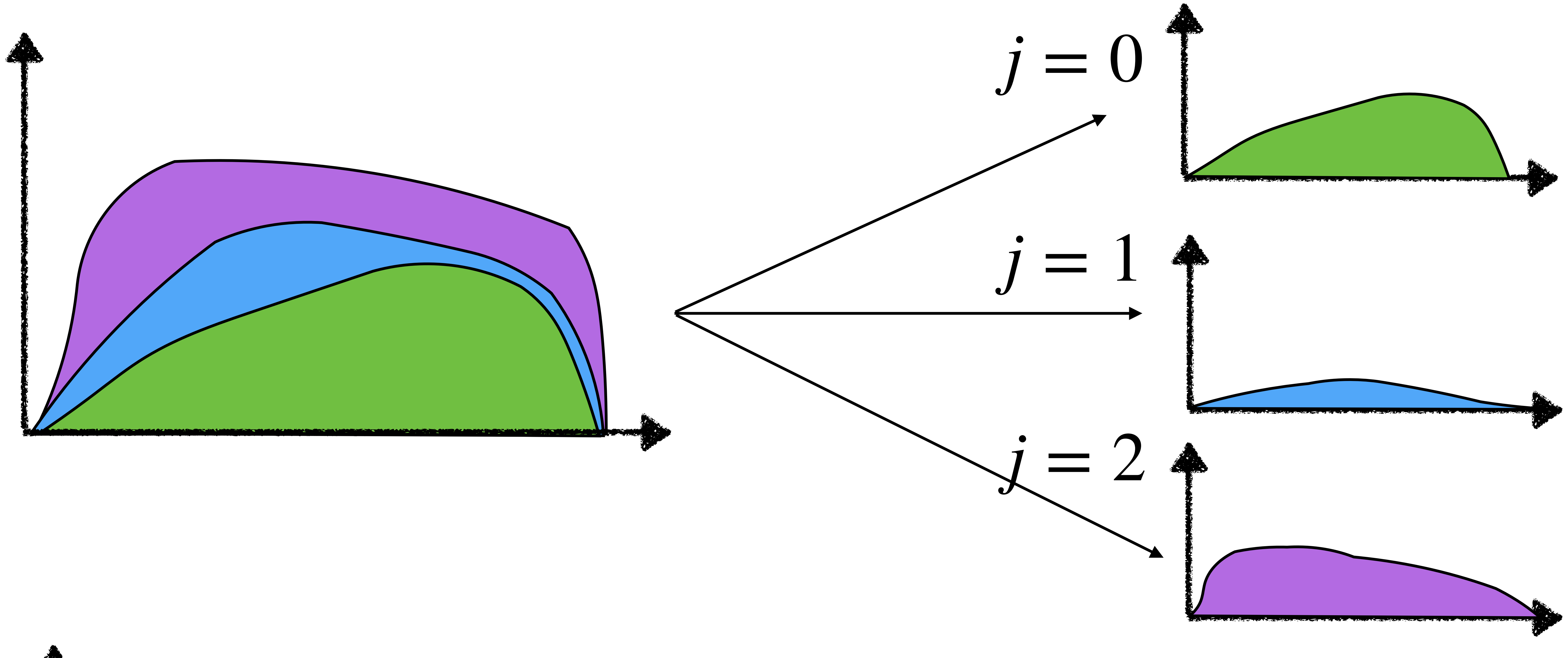
A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Robust, parameterised classifier without parameterising

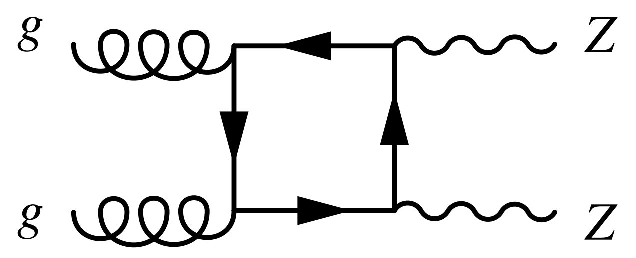
H_{ref} : Reference hypothesis

H_μ

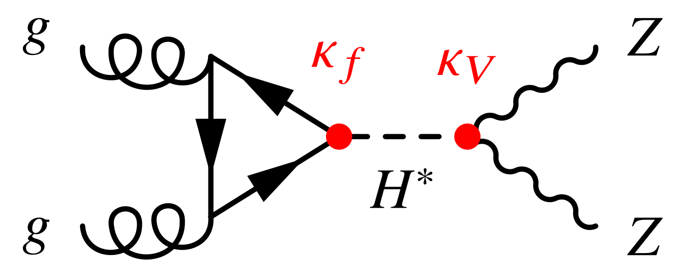
Hypothesis



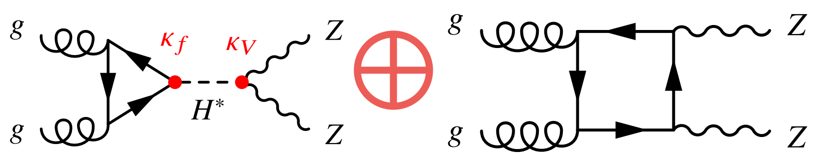
$f_0(\mu) = 1$



$f_1(\mu) = \mu$



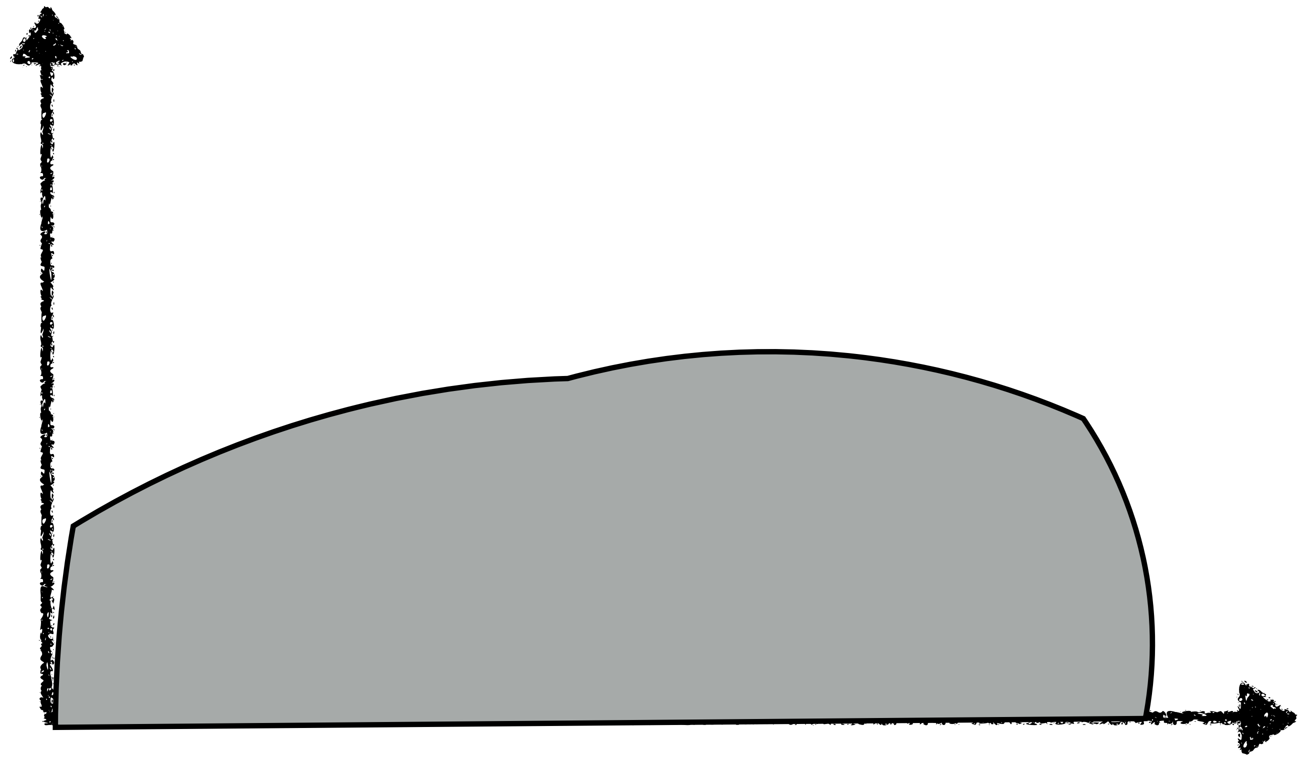
$f_2(\mu) = \sqrt{\mu}$



VS

H_{ref}

Reference hypothesis



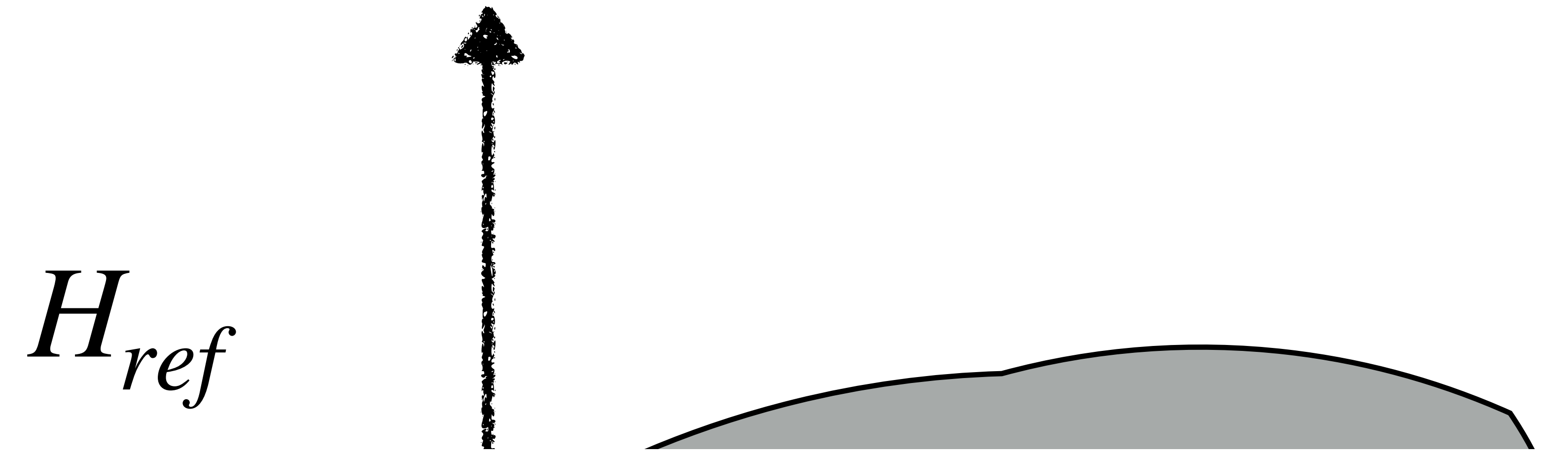
$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

A separate classifier per physics process j

(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis



Analytically parameterised in μ, allows to get LR for any hypothesis μ without training parameterised networks !

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

A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

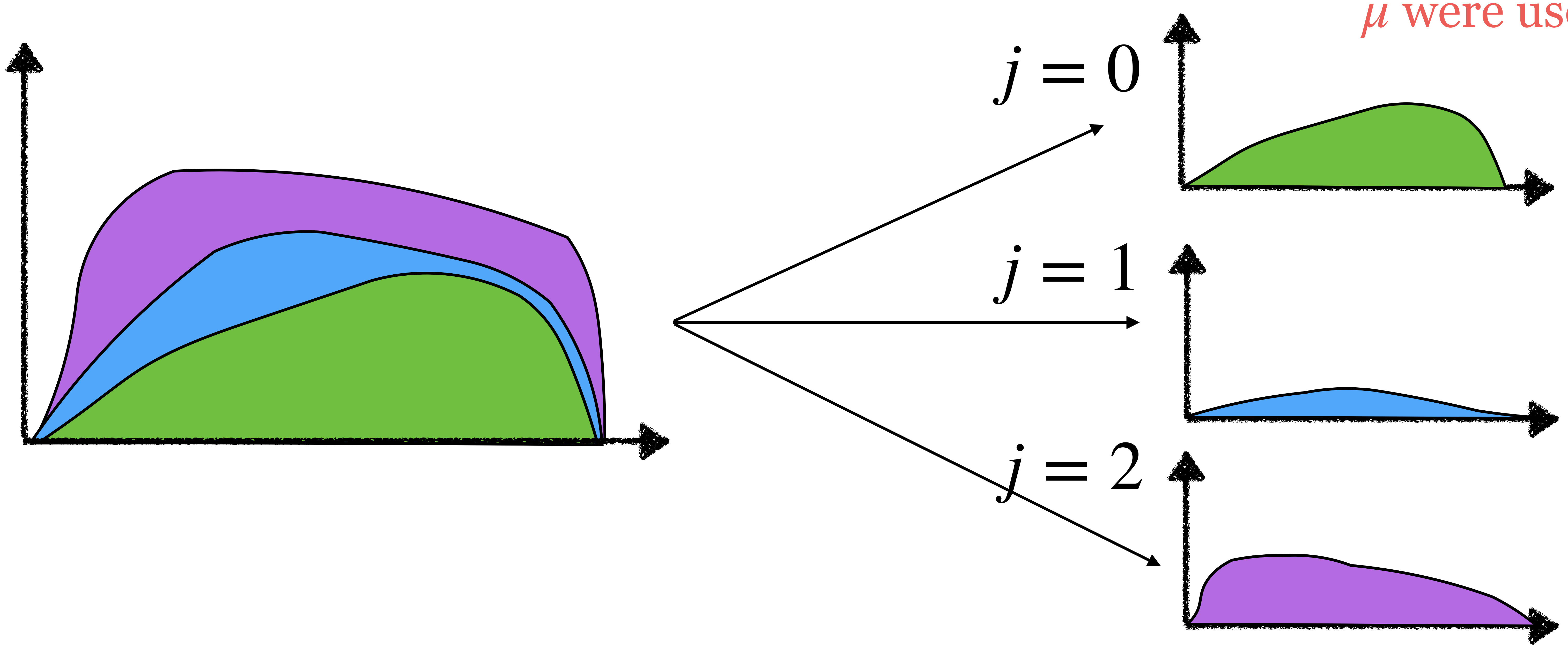
Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

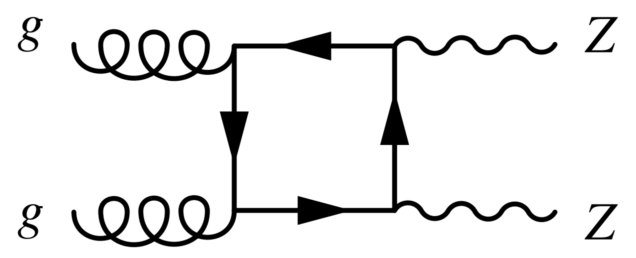
$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

H_μ

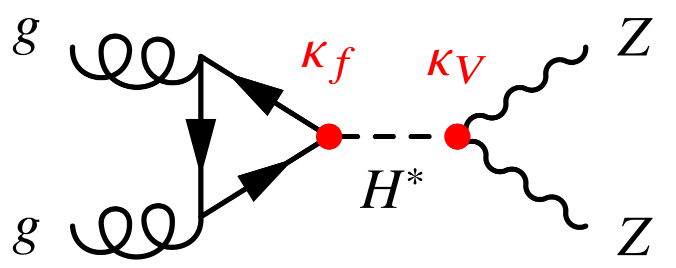
Hypothesis



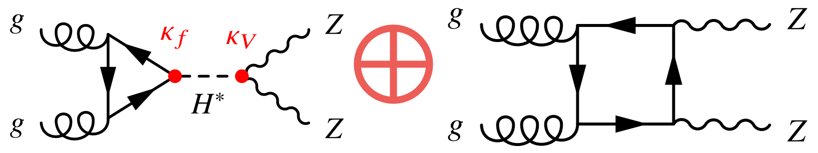
$f_0(\mu) = 1$



$f_1(\mu) = \mu$

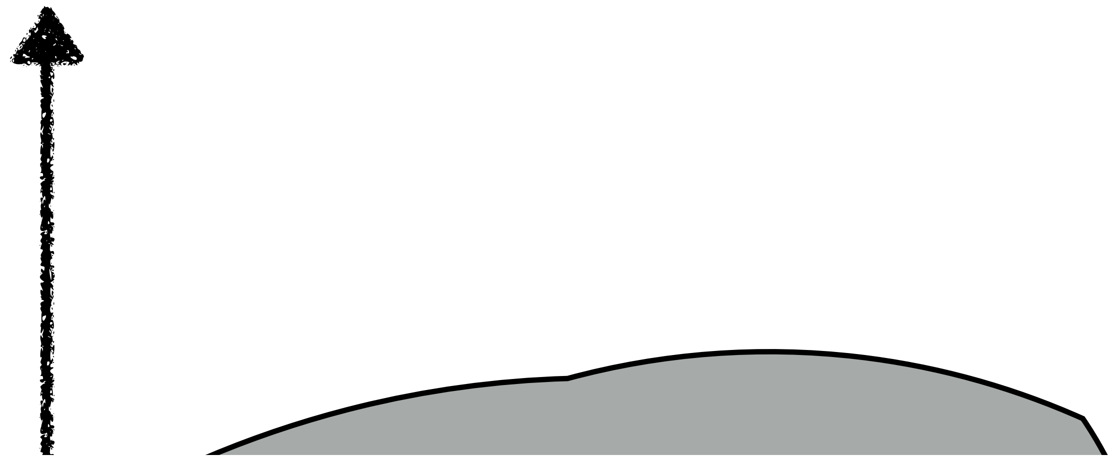


$f_2(\mu) = \sqrt{\mu}$



VS

H_{ref}



Analytically parameterised in μ, allows to get LR for any hypothesis μ without training parameterised networks !

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

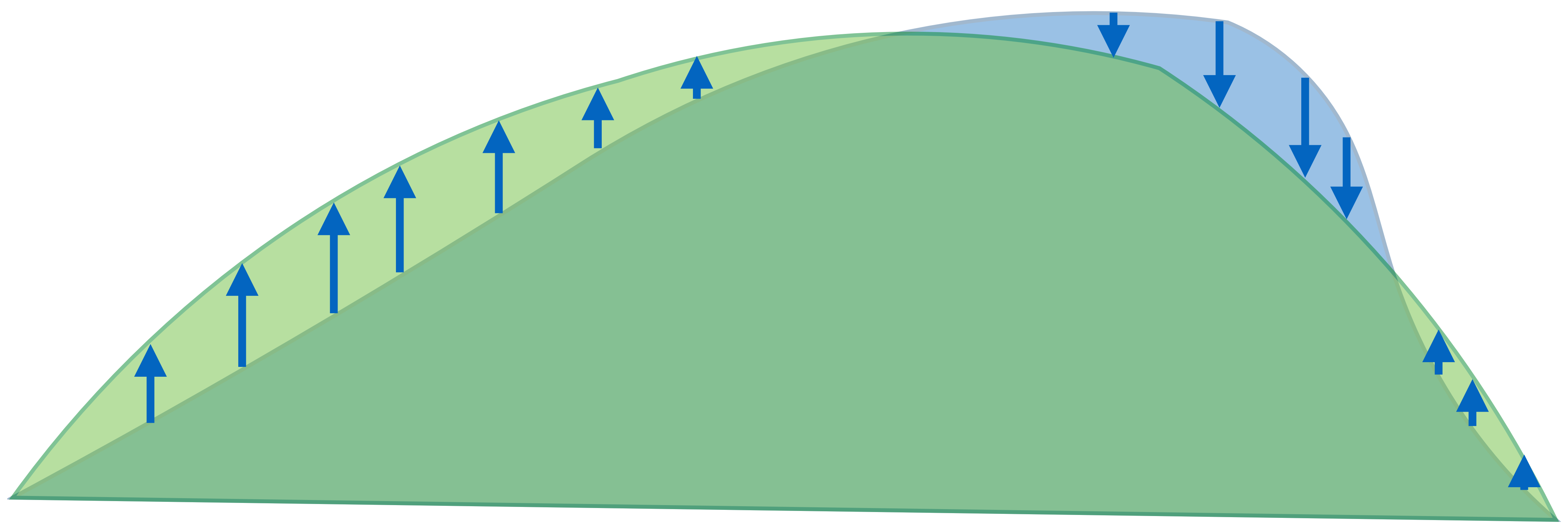
A separate classifier per physics process j (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Throwing toys in a per-event analysis

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

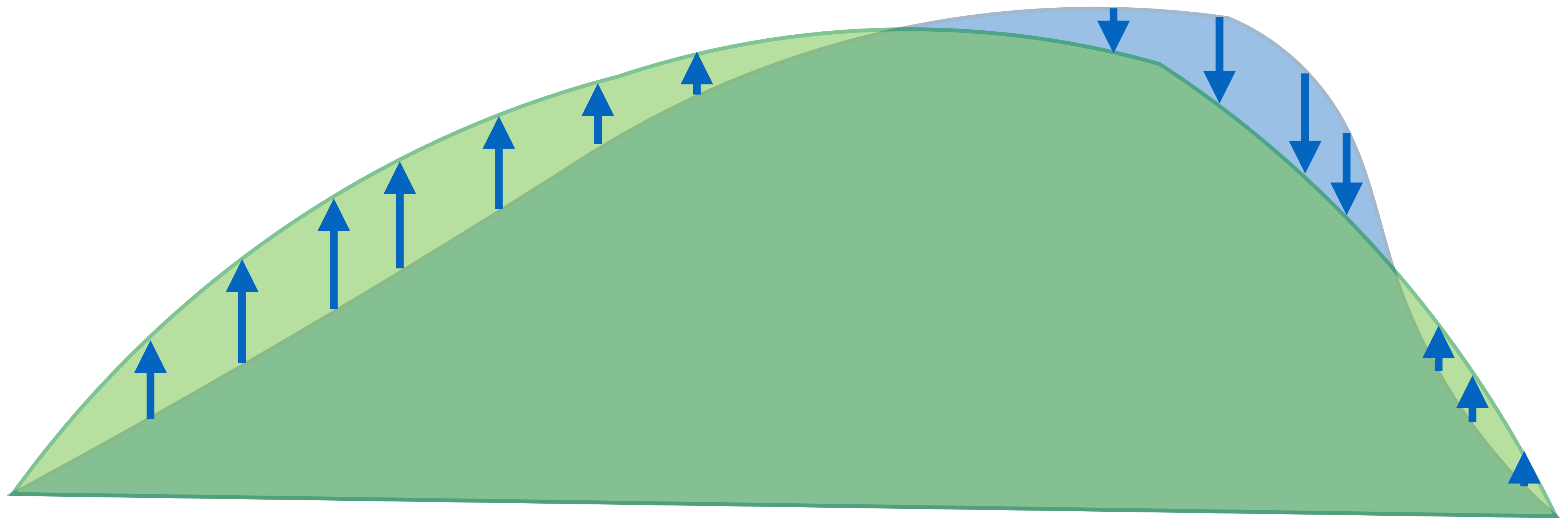


Validate quality of LR estimation with re-weighting task

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

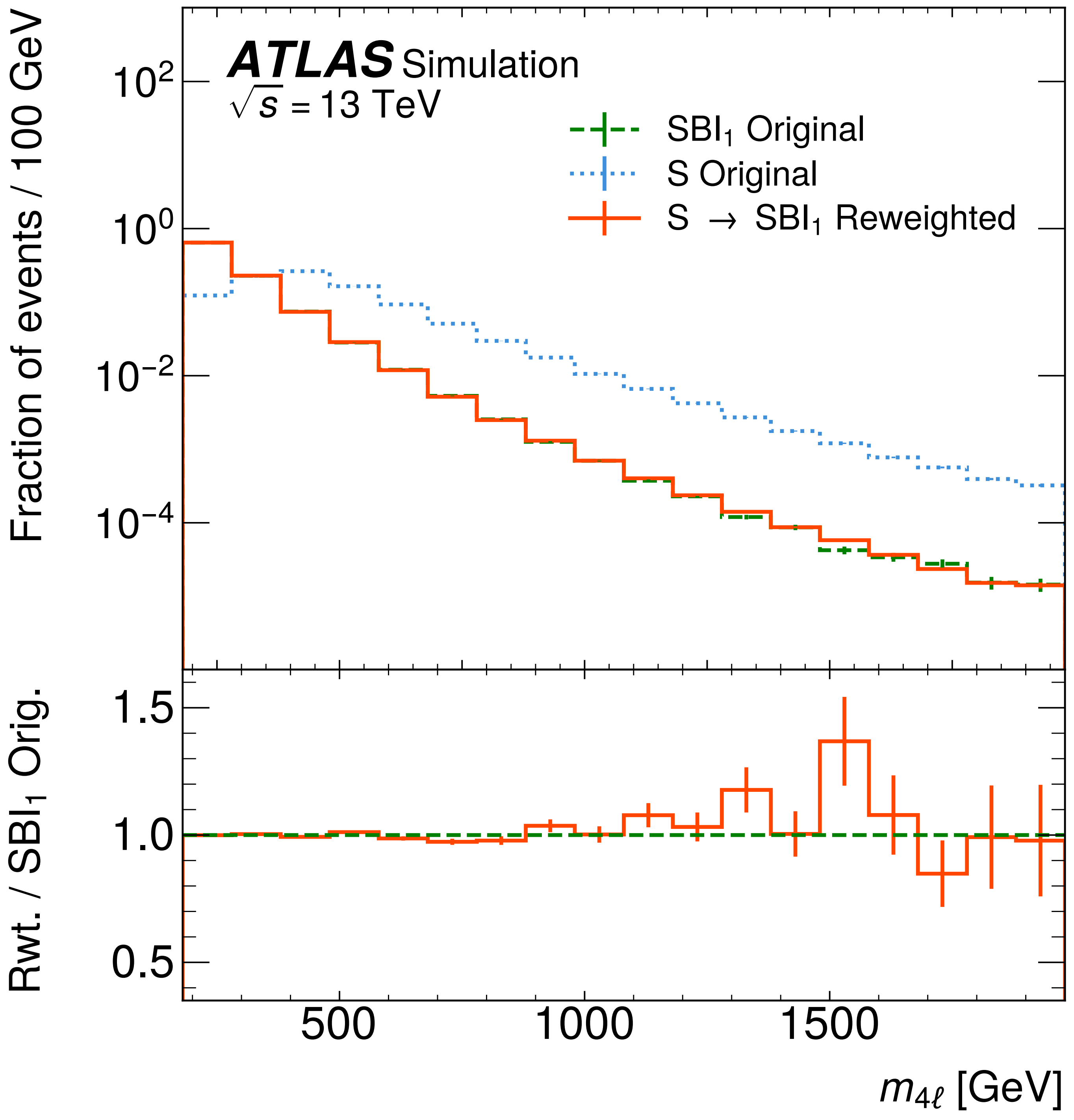


Re-weight closures

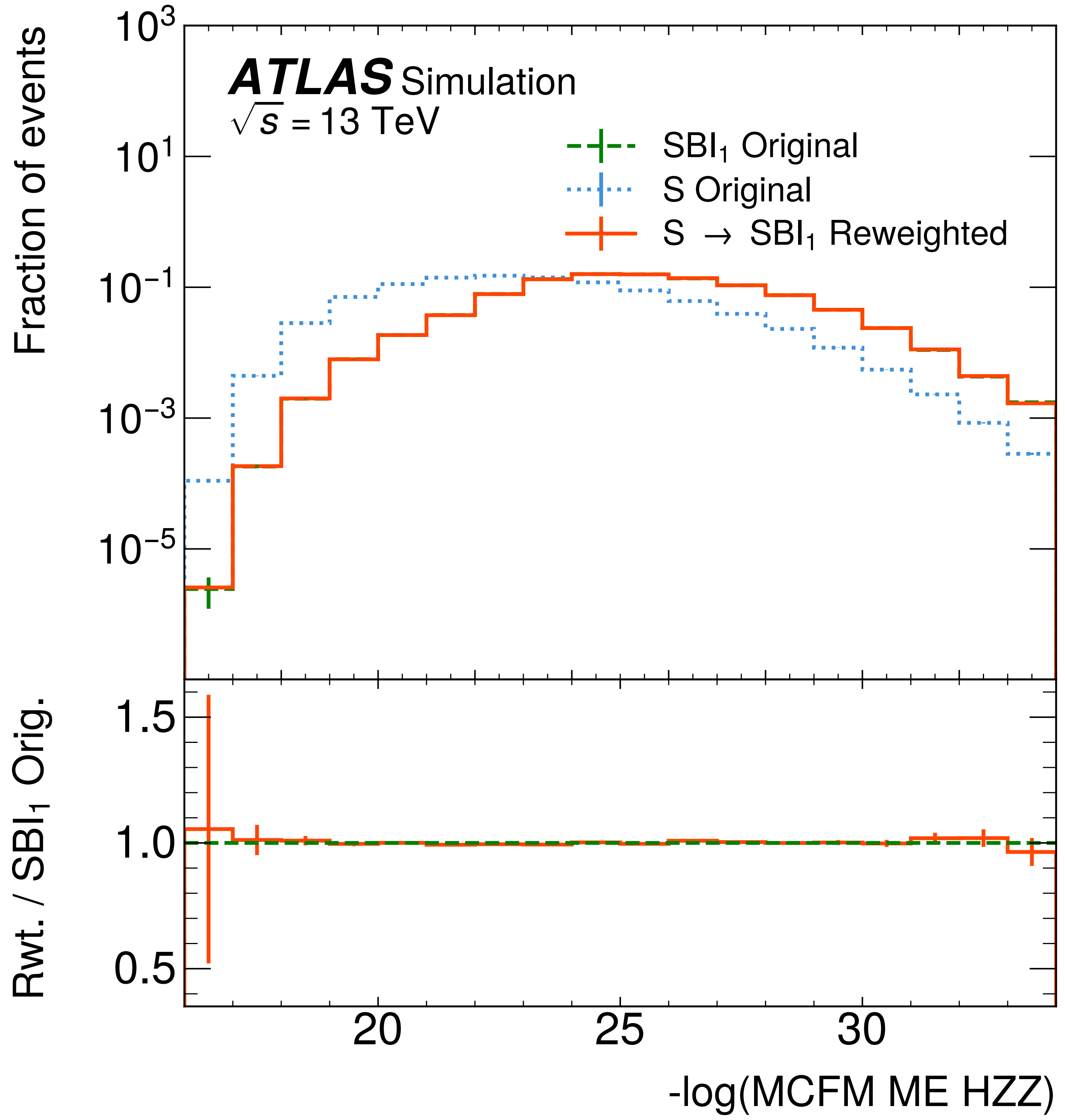
Variable used in training

Source
Target
RW

High-level variable
never used in training



m_{4l}



Matrix-Element-based Observable
(ggF from MCFM)

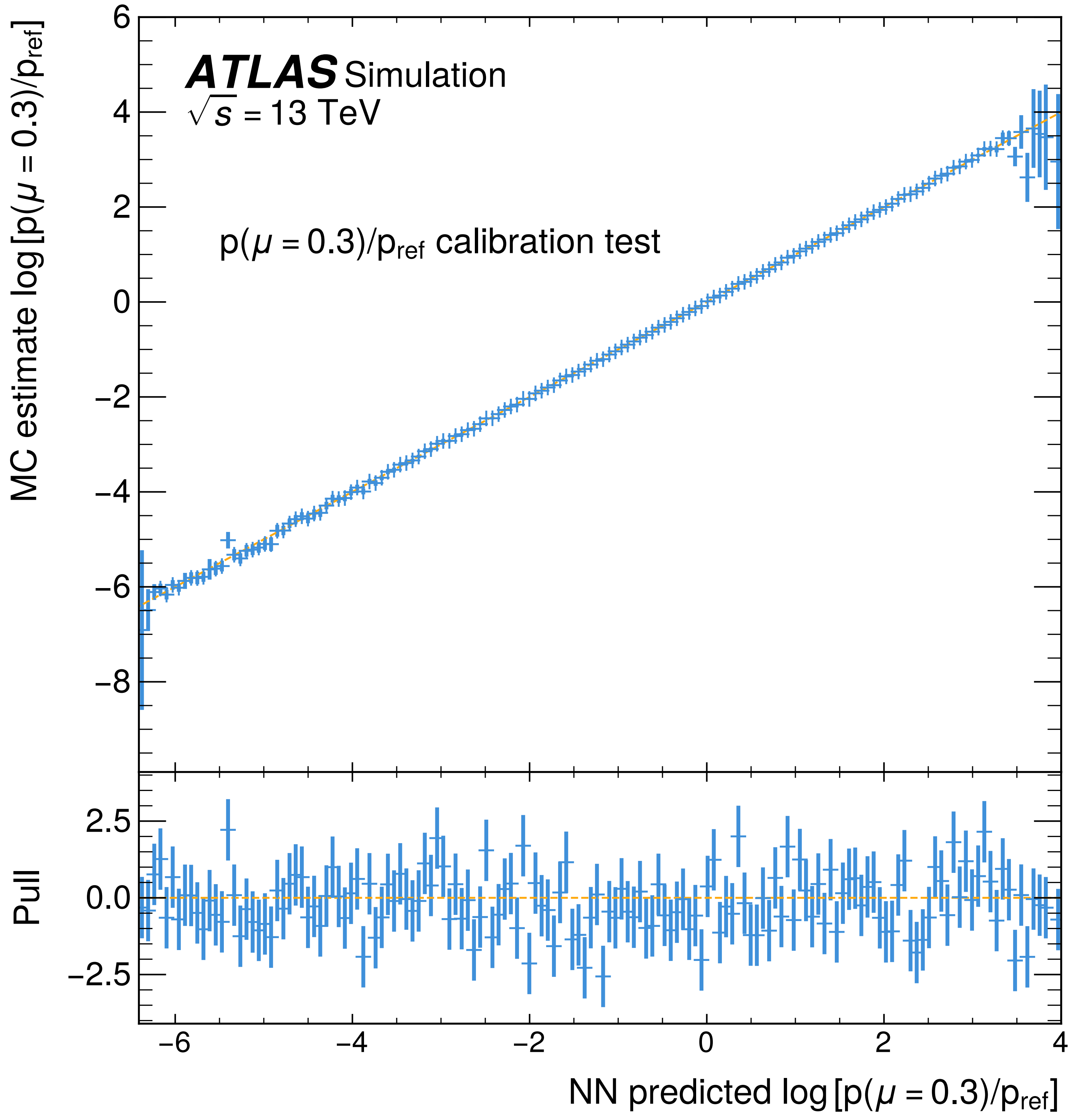
High-Dim Classifier Test:
Train independent classifier on RW vs Target,
AUC=0.5 ⇒ LRs well estimated

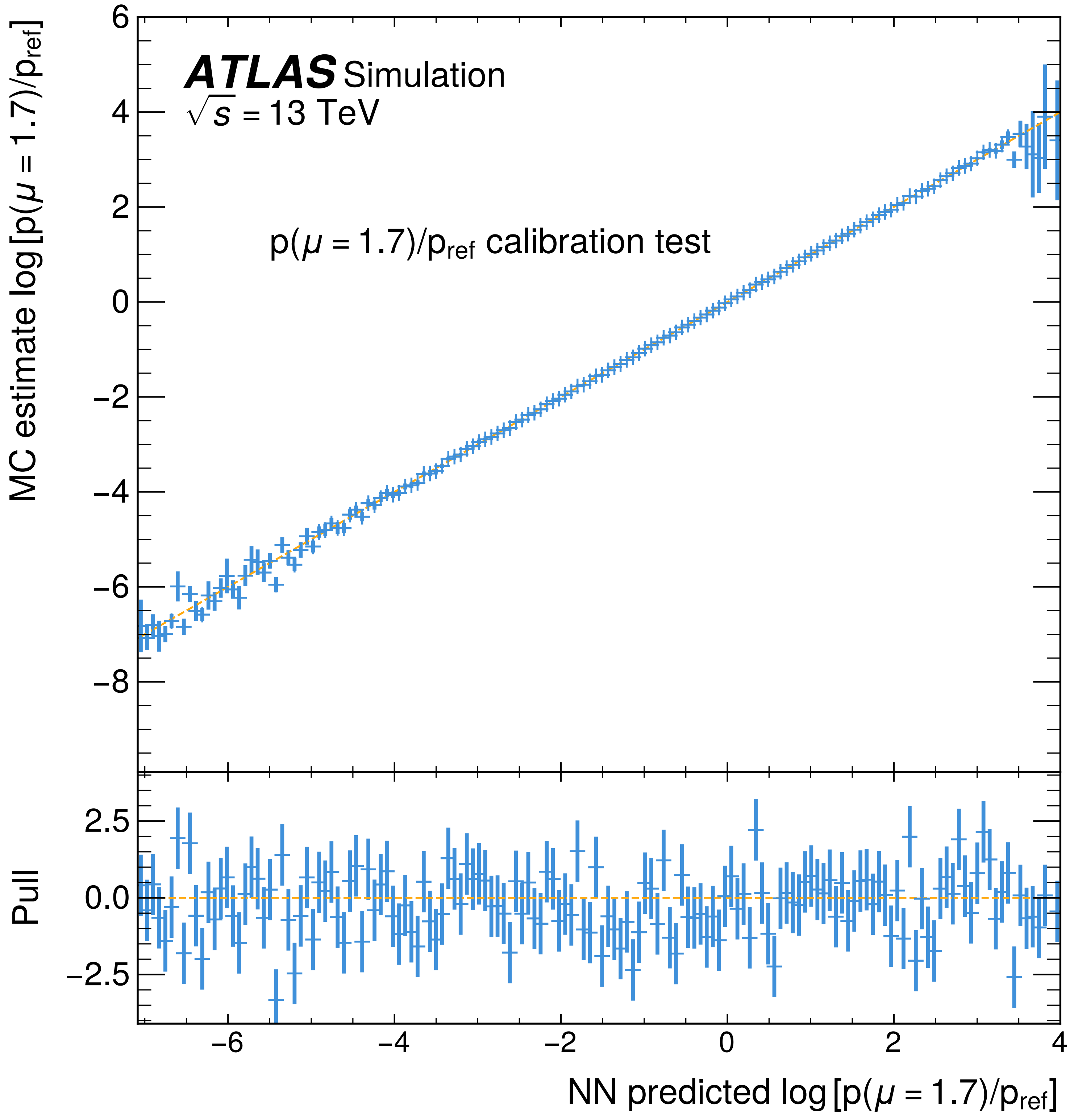
Calibration curves of probability density ratios

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

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

Binned estimate



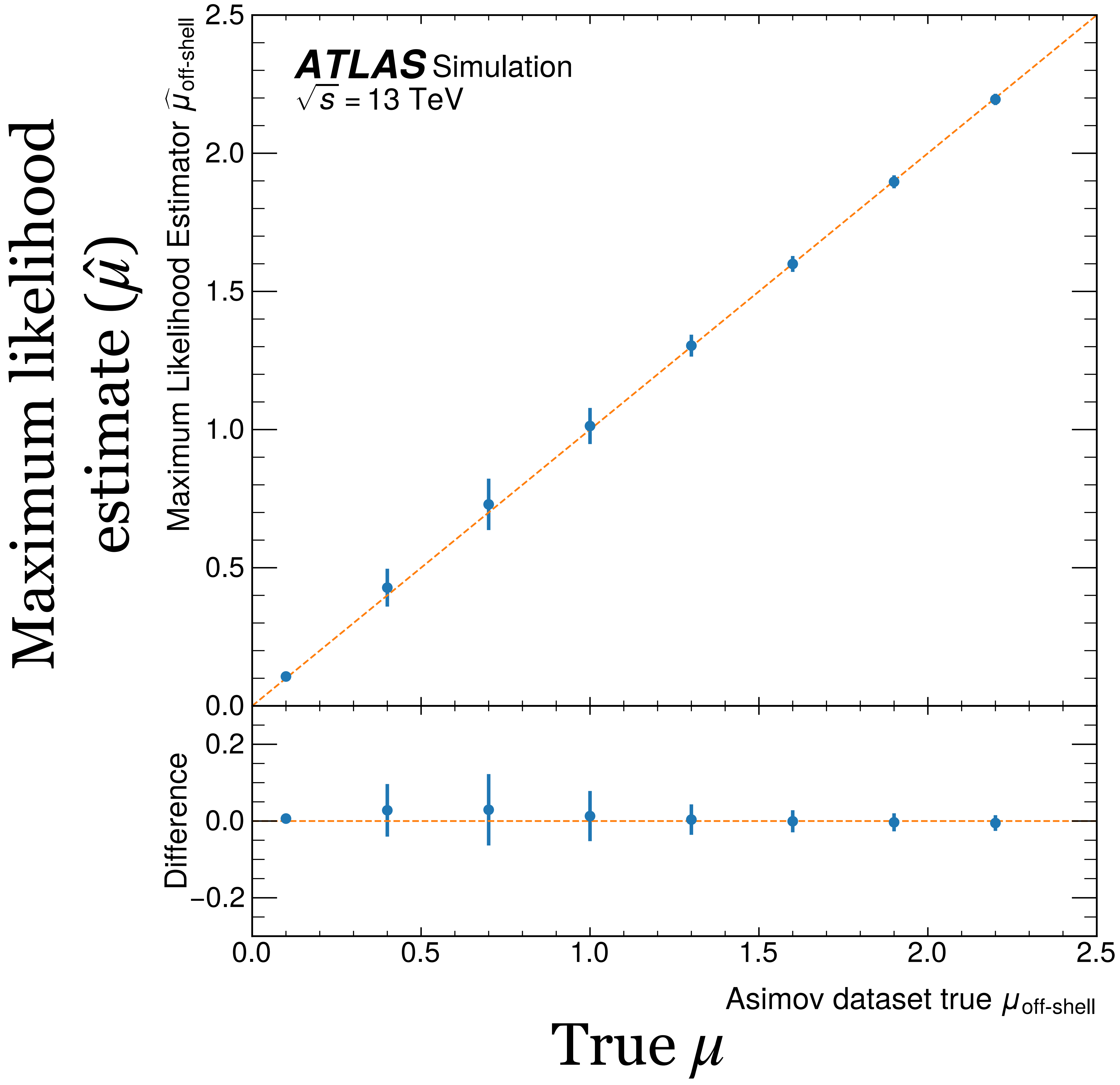


Perfect calibration would give $y = x$

Ensemble prediction

Ensemble prediction

Testing full analysis on samples from different values of μ



No bias: Method recovers correct value of μ on average

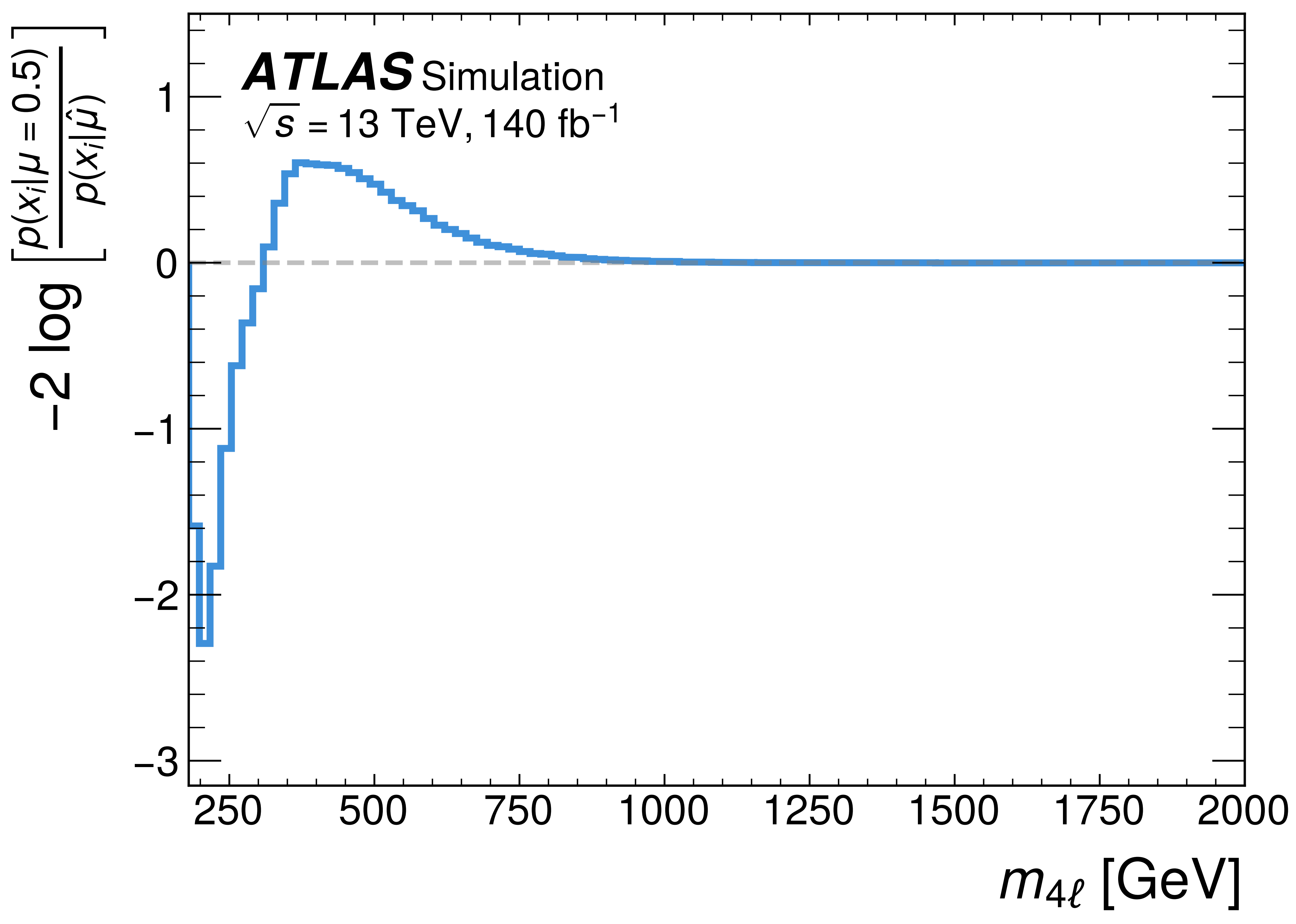
(Correct value when tested on the median 'Asimov dataset')

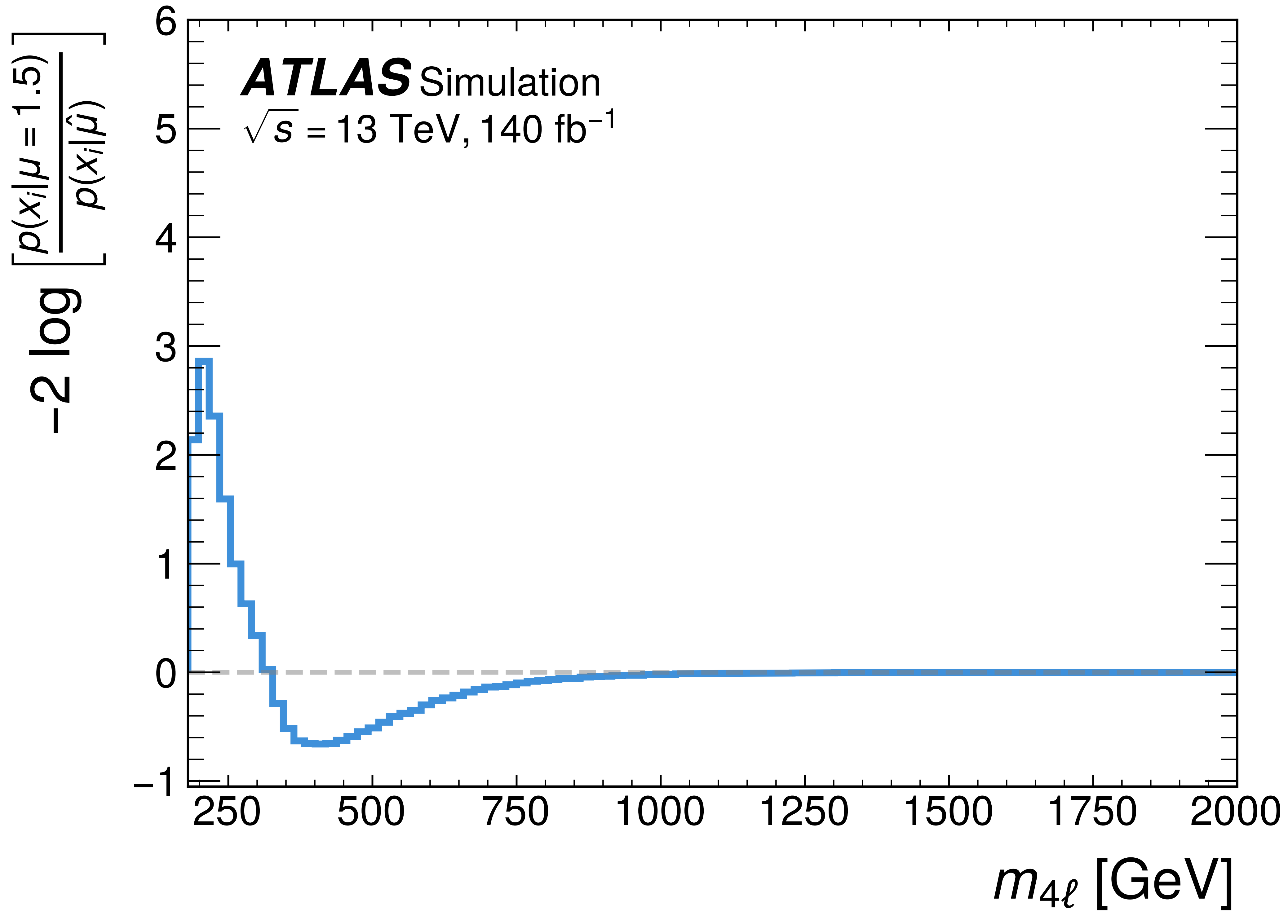
And many more diagnostics (see backup)

Interpretability:
Which phase space favours one hypothesis over another?

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

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

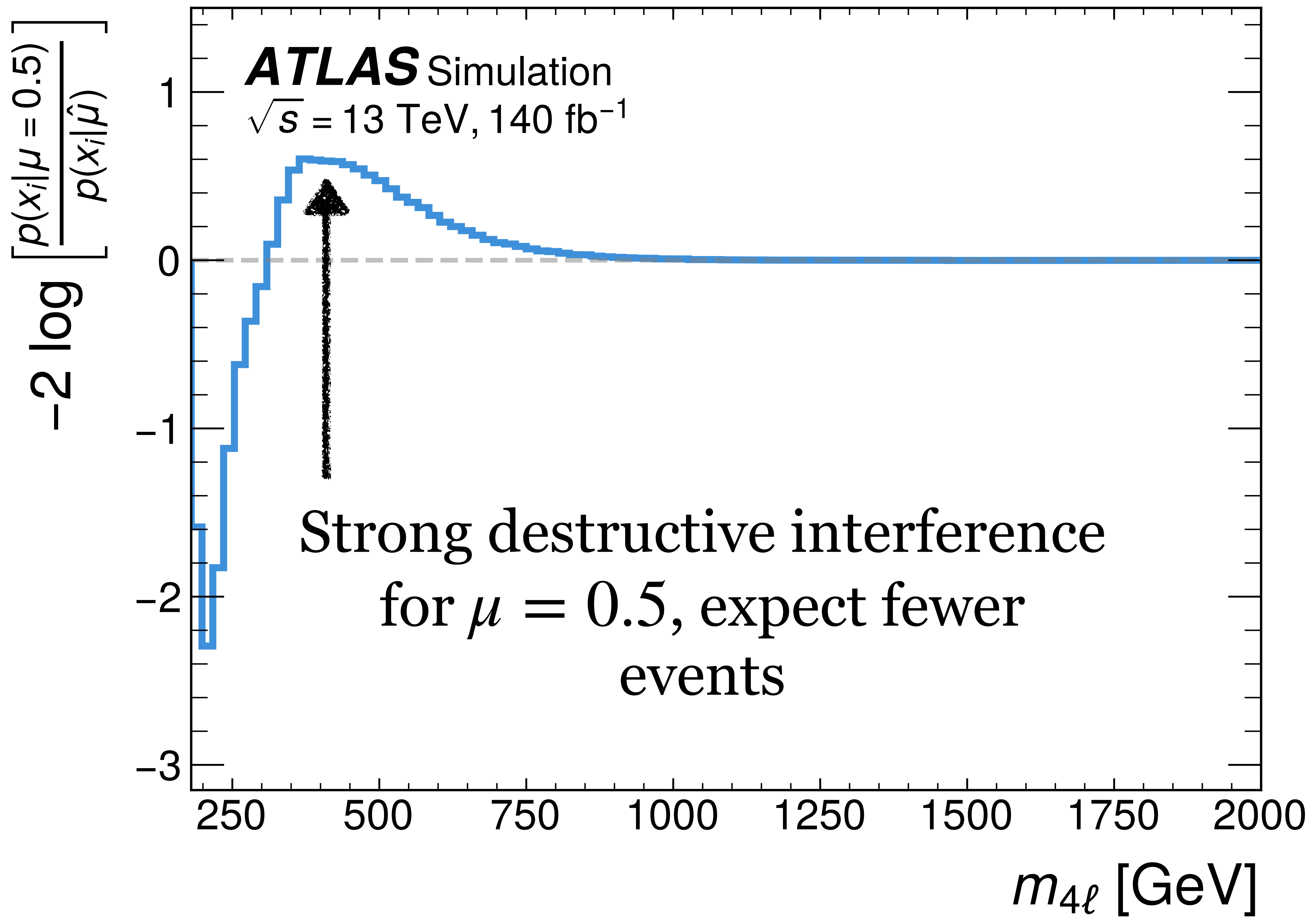


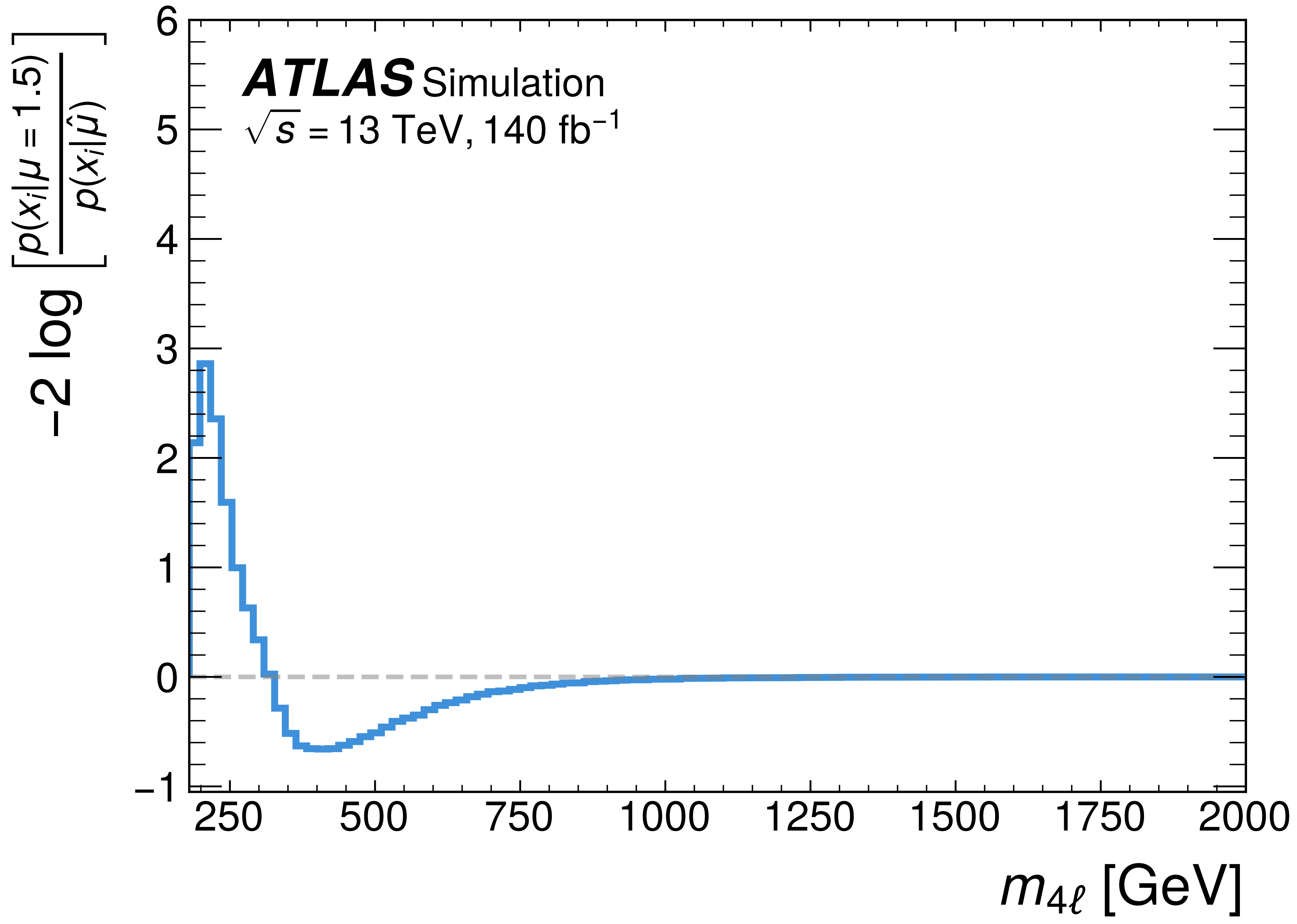


Interpretability:
Which phase space favours one hypothesis over another?

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

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

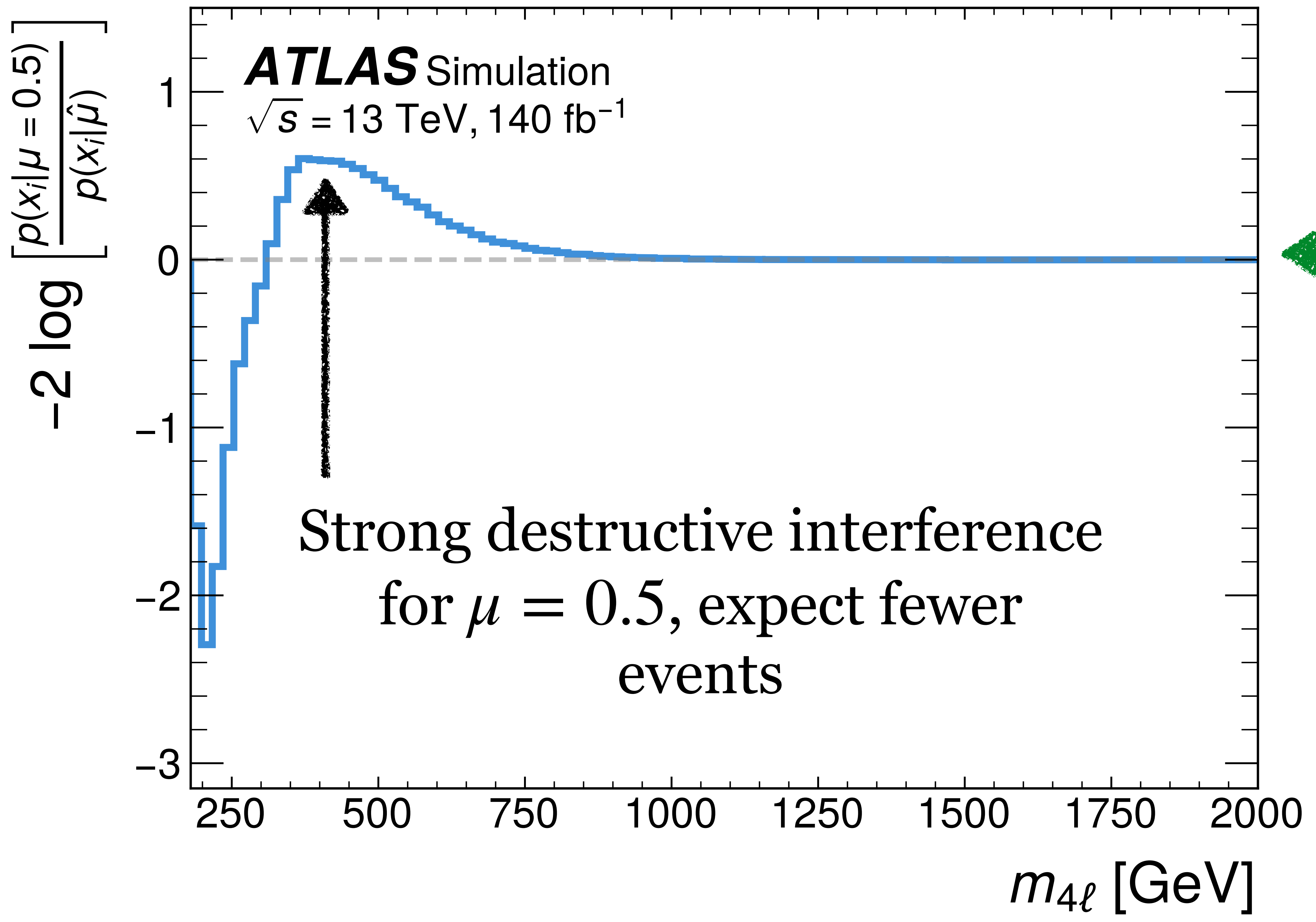


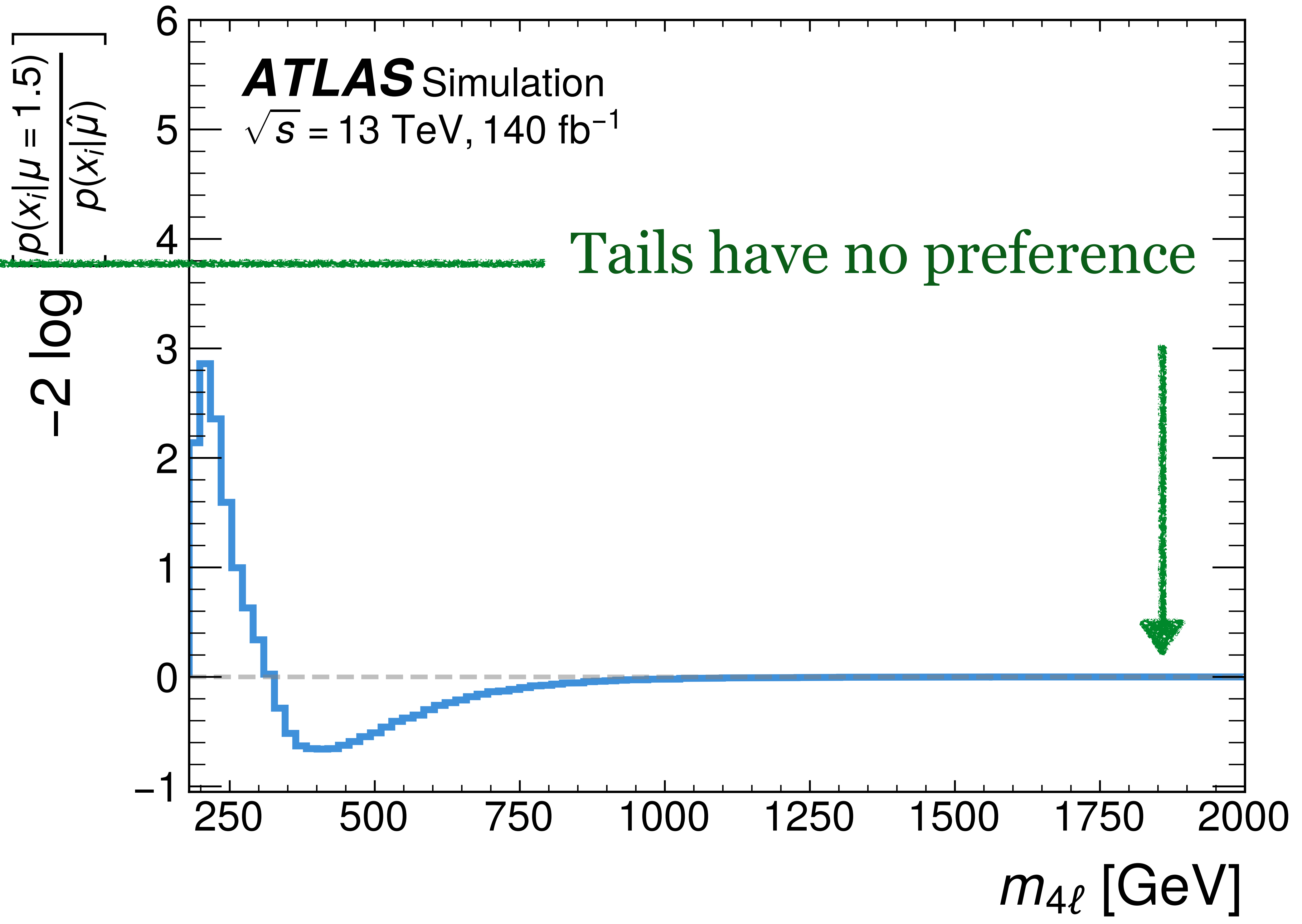


Interpretability:
Which phase space favours one hypothesis over another?

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

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





Open problems to extend to full ATLAS analysis:

- ✓ Robustness: Design and validation
- ‣ Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- • Neyman Construction: Throwing toys in a per-event analysis

Systematic uncertainties

Experimental uncertainties:

Eg. Inaccuracies in the calibration of our detector

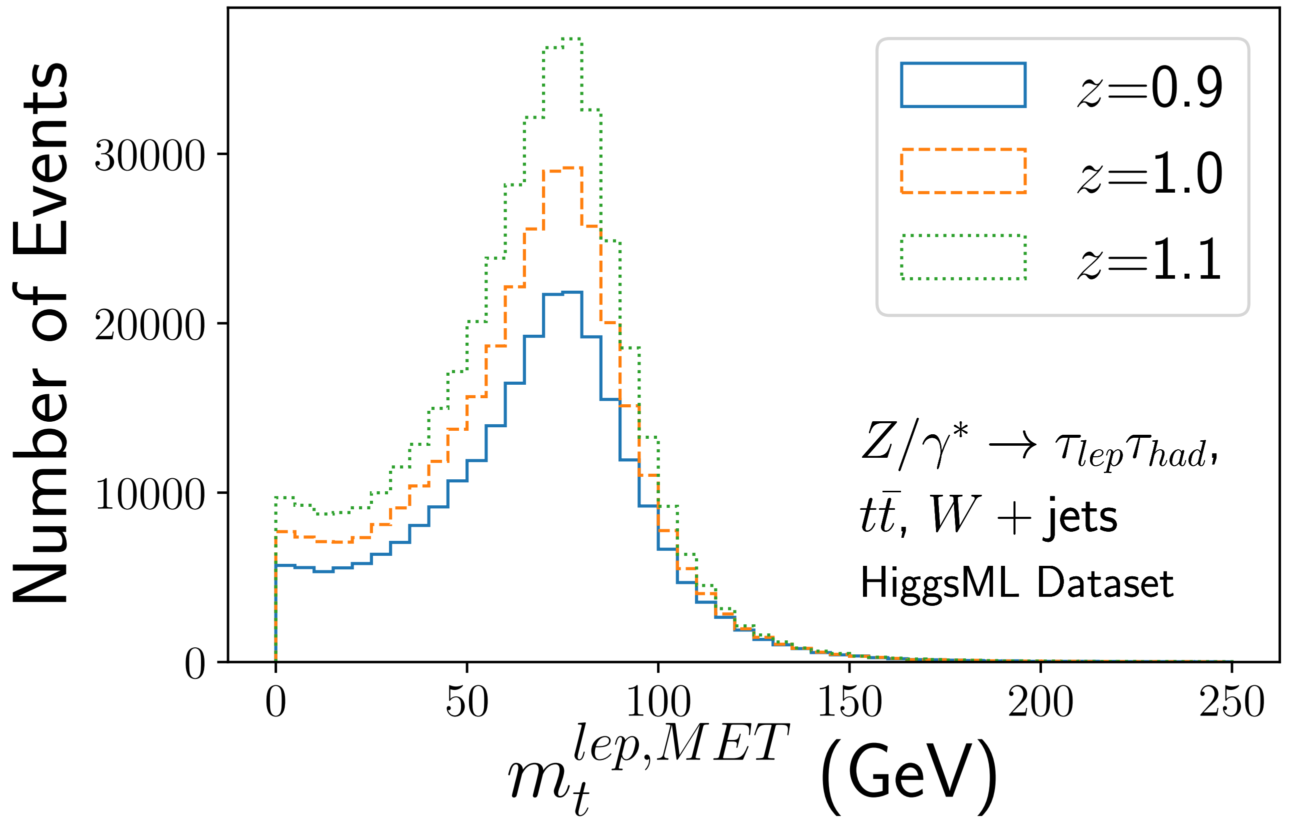


Image: arXiv:2105.08742

Theory uncertainties:

Eg. Inability to compute QFT to infinite order

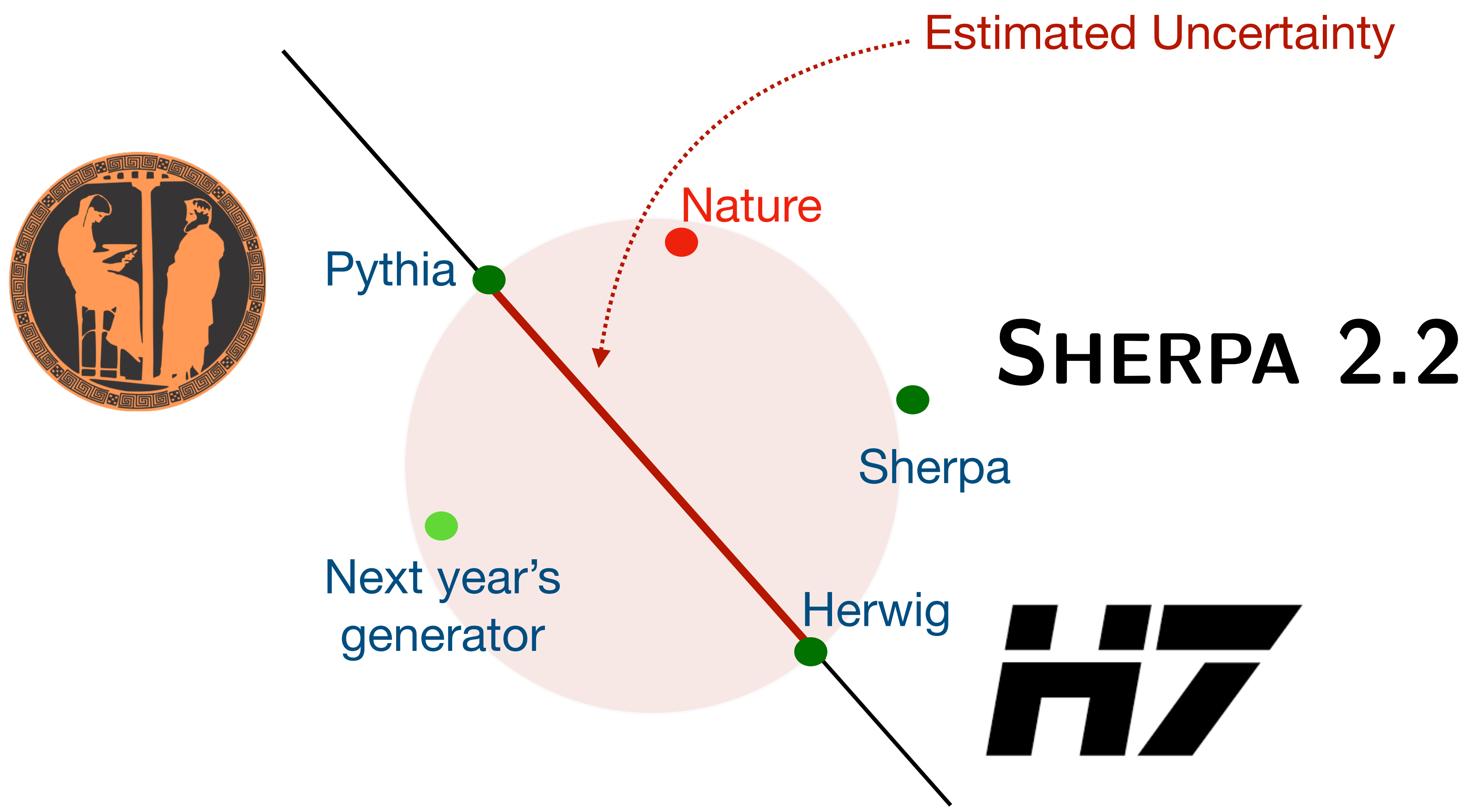


Image: arXiv:2109.08159

Systematic uncertainties

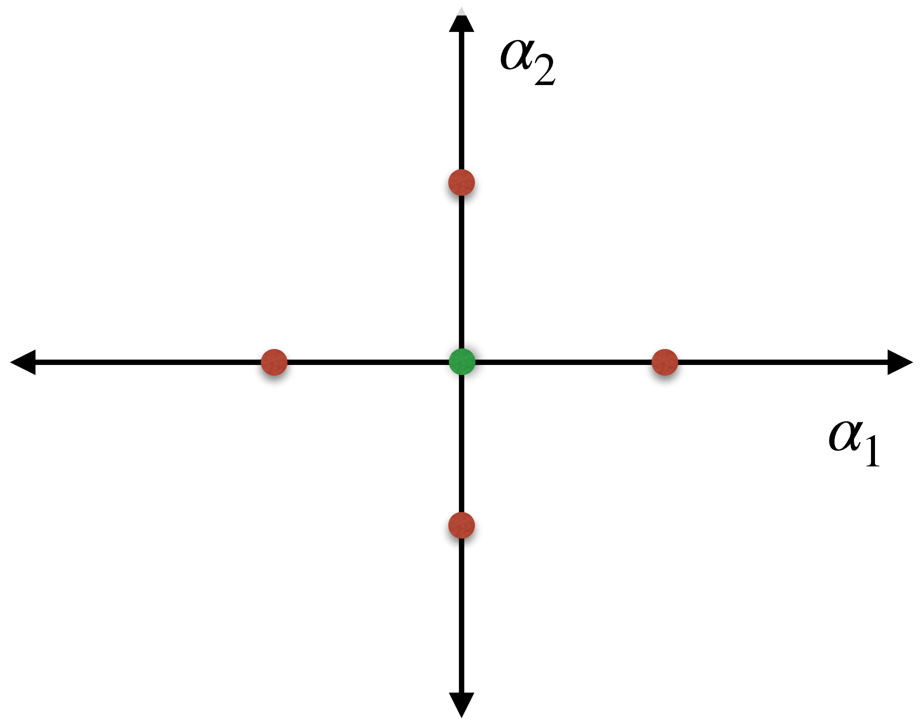
Experimental uncertainties:
Eg. Inaccuracies in the calibration of our detector

Image: arXiv:2105.08742

Theory uncertainties:
Eg. Inability to compute QFT to infinite order

Image: arXiv:2109.08159

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



Known interpolation strategies

See formula used in backup

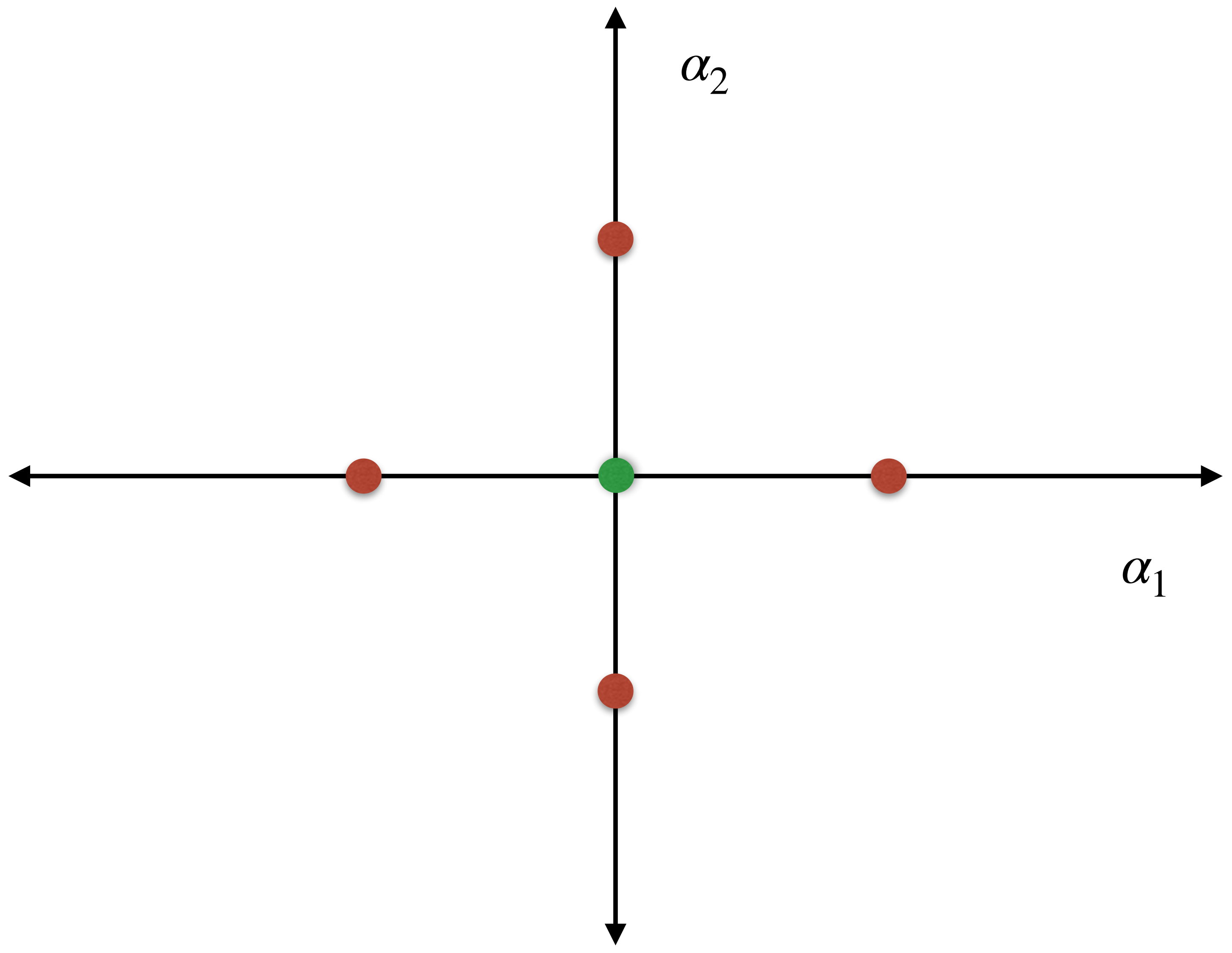
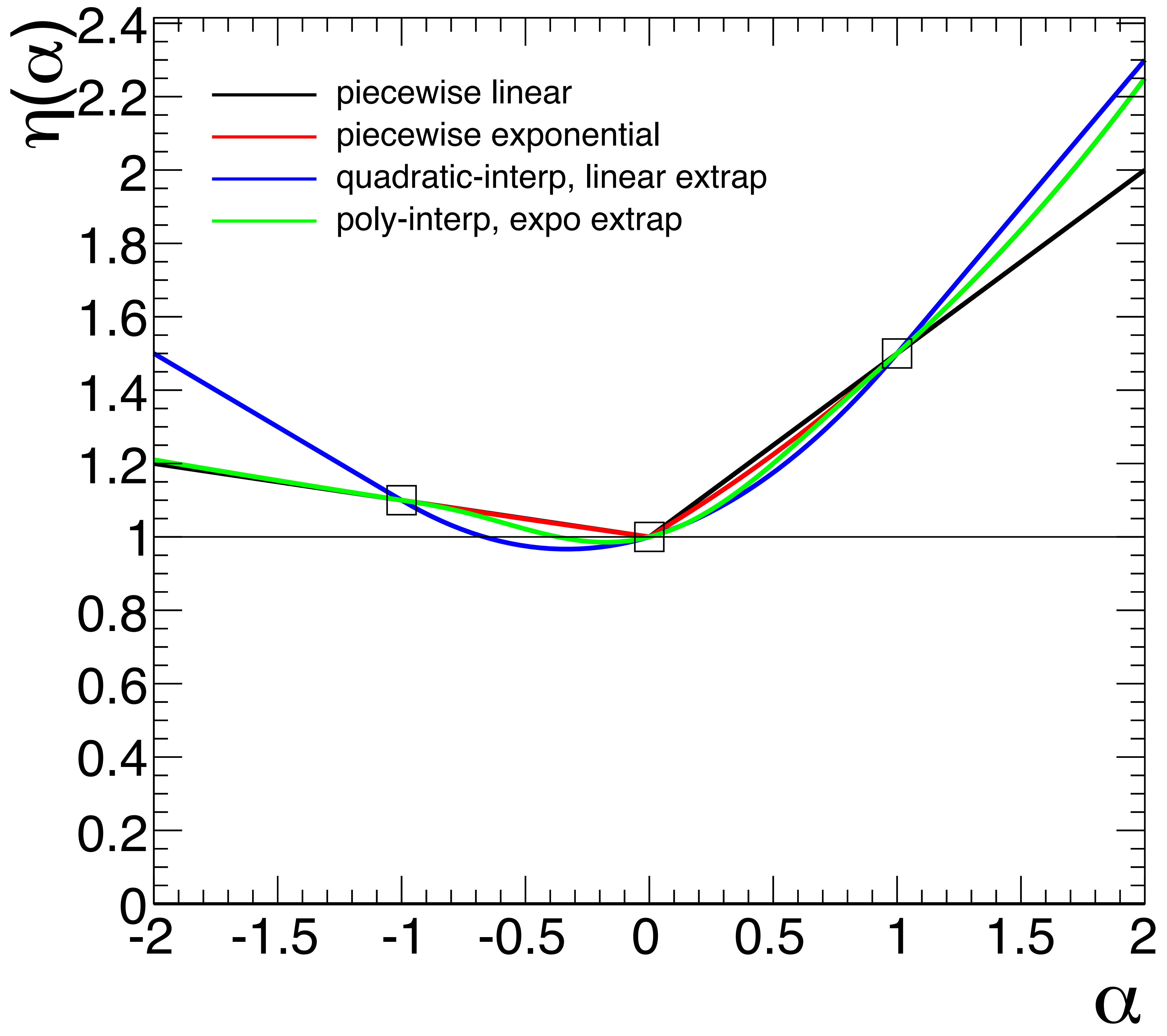


Image: arXiv:1503.07622

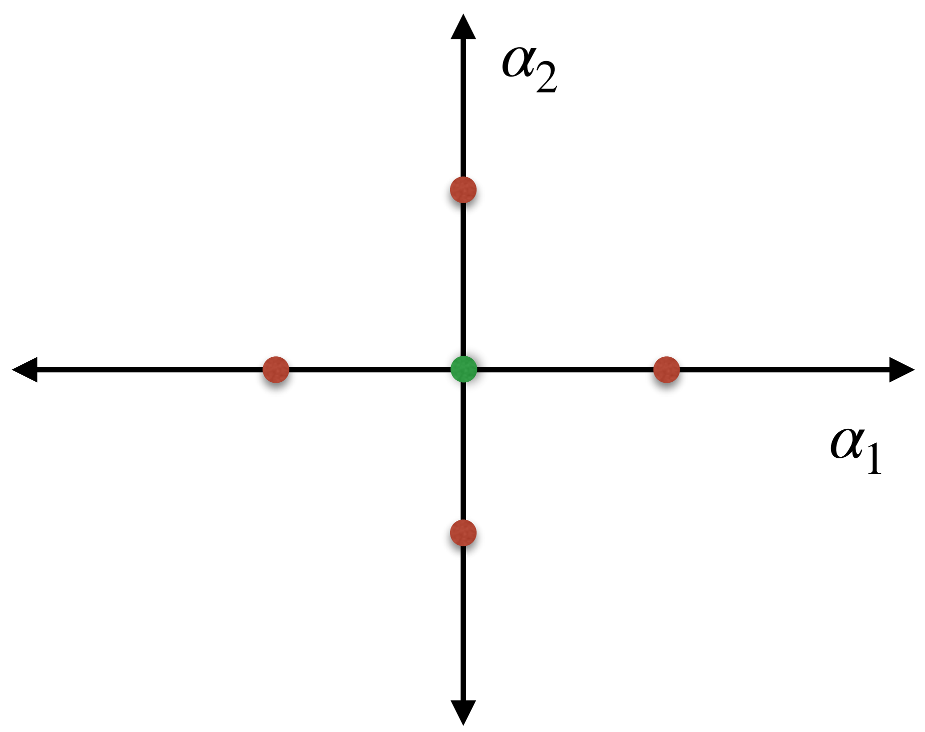


⇒ Combine these traditional interpolation with neural network estimation of per-event likelihood ratios

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} =$$

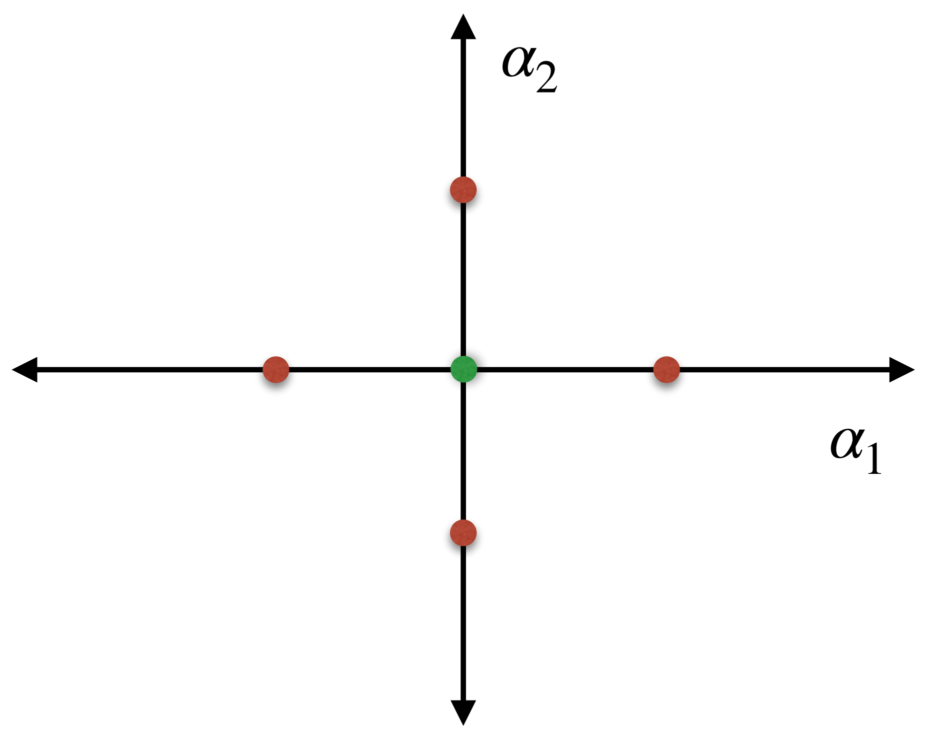


See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

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



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

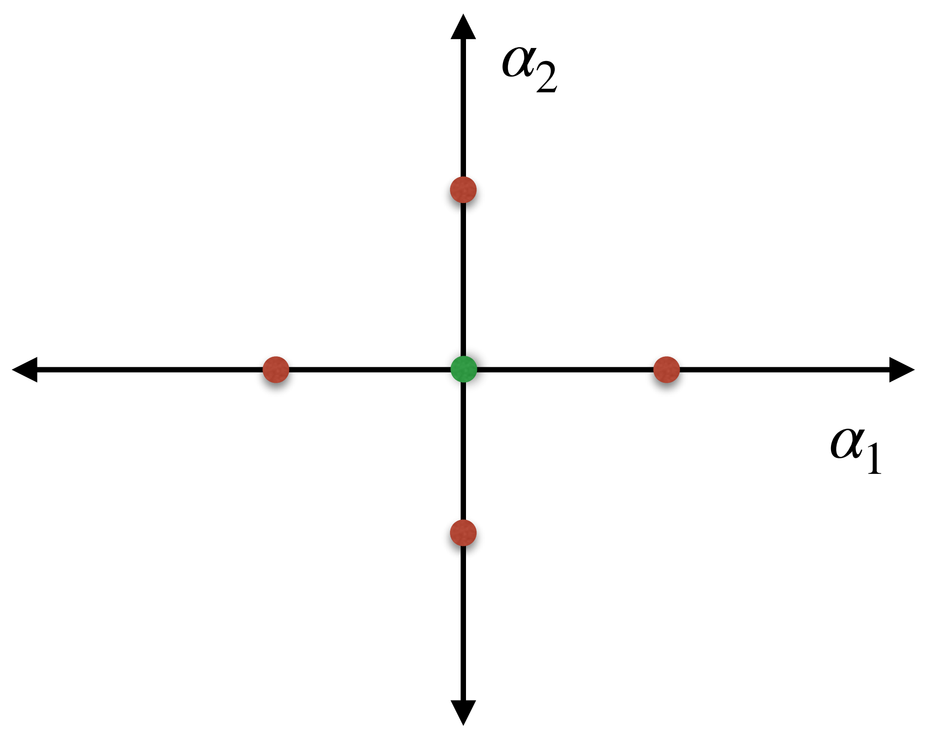
See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

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

We have this already



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

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

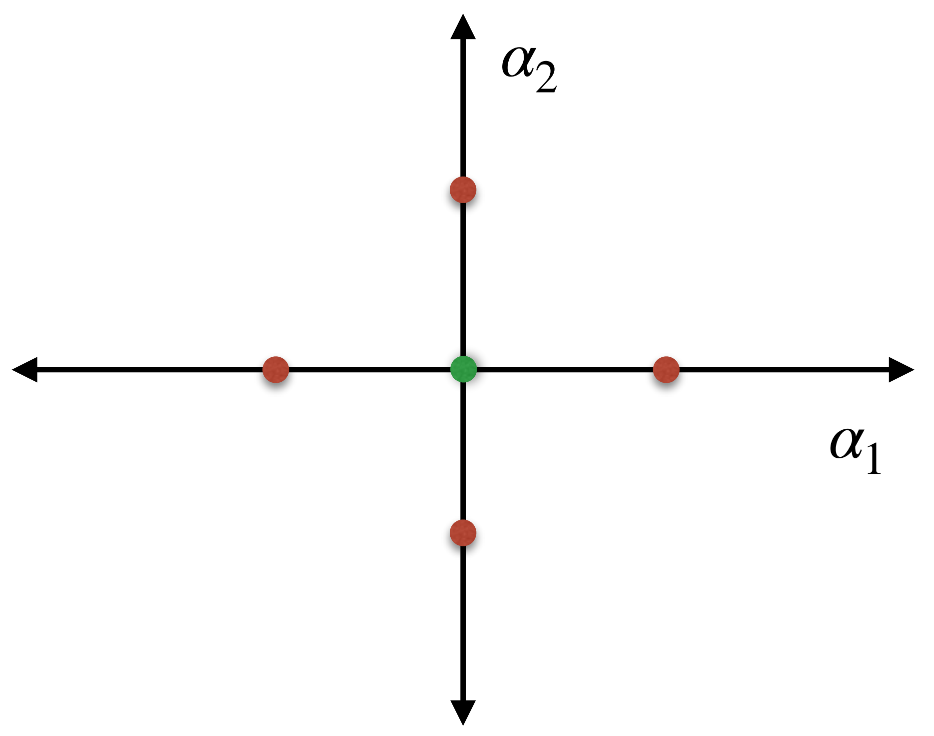
Probability density ratio including nuisance parameters (α)

x_i is one individual event

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

We have this already

Estimate from simulations and existing interpolation methods



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

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

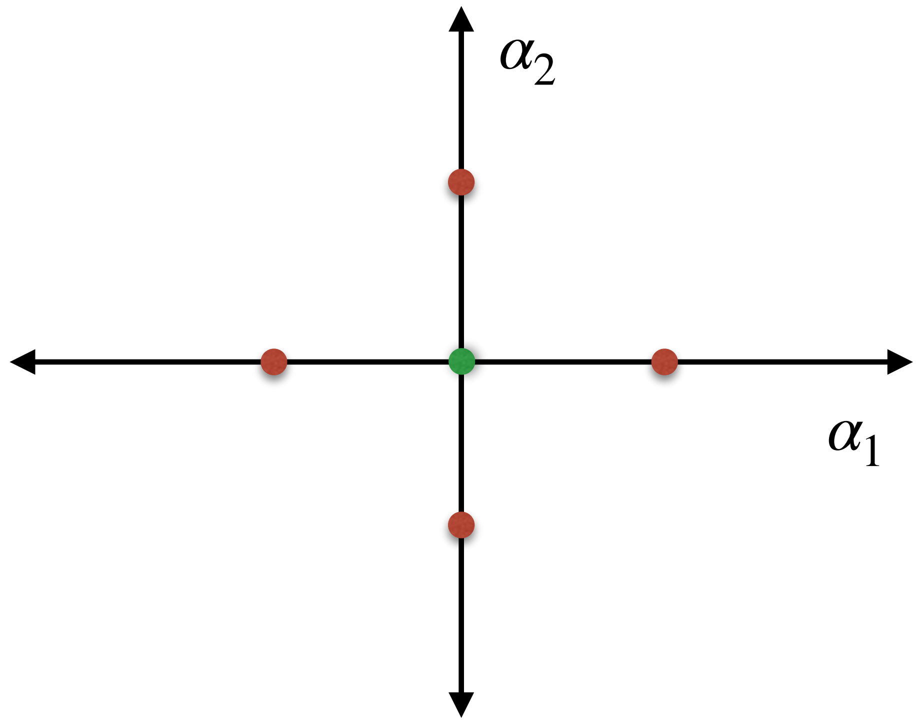
$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$

We have this already

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

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

Estimate from simulations and existing interpolation methods



See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Final test statistic

x_i is one individual event

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

Final test statistic

x_i is one individual event

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

From previous slide

Final test statistic

x_i is one individual event

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

From previous slide

Prod over events

Final test statistic

x_i is one individual event

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

From previous slide

Rate term

Prod over events

Final test statistic

x_i is one individual event

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

From previous slide

Rate term

Prod over events

Constrain term

Final test statistic

x_i is one individual event

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

From previous slide

Rate term

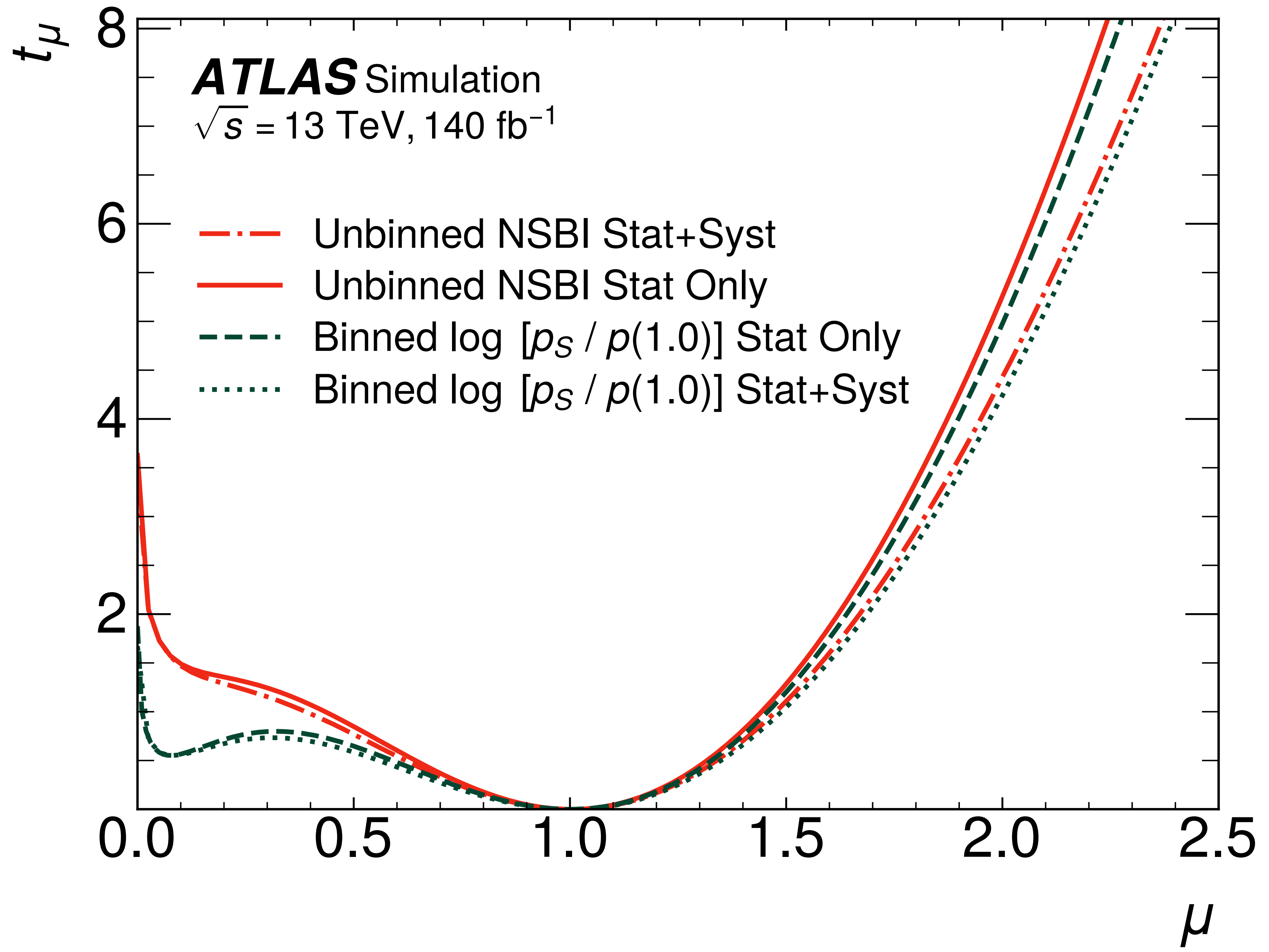
Prod over events

Constrain term

Profiling:

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

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



Final test statistic

x_i is one individual event

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

From previous slide

Rate term

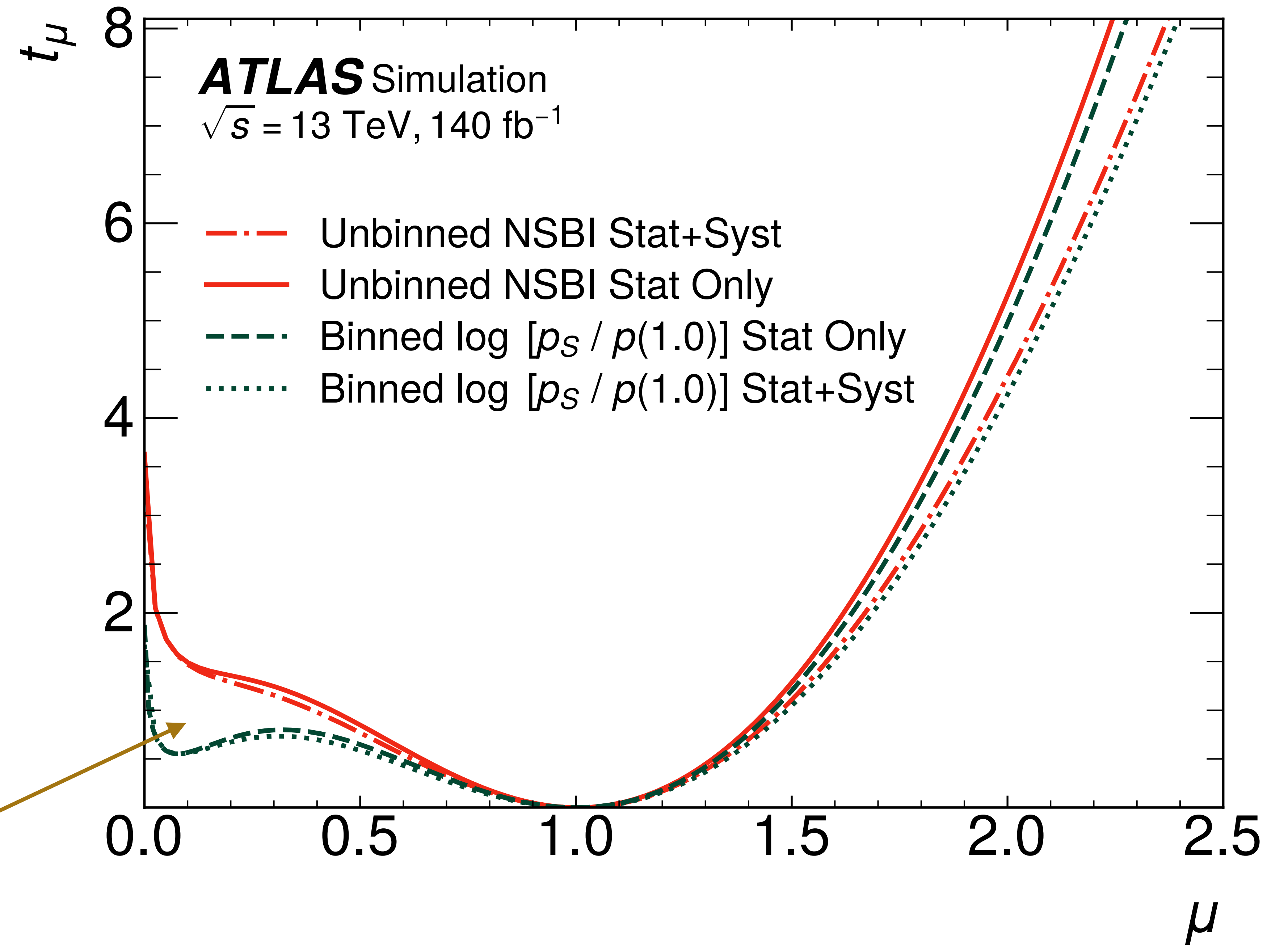
Prod over events

Constrain term

Profiling:

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

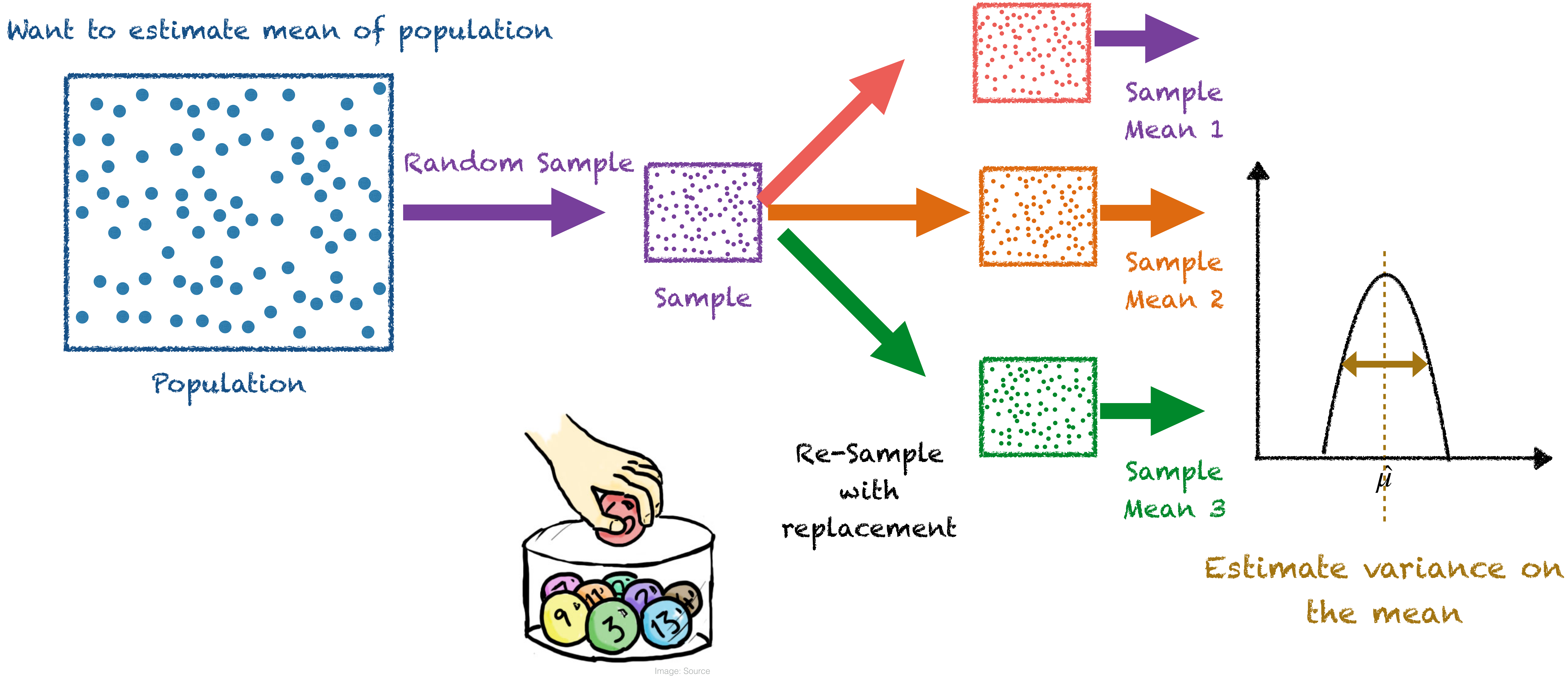
This is why we define p_{ref} to be independent of μ



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

Uncertainty from finite training samples

Estimating the variance on mean: Bootstrapping

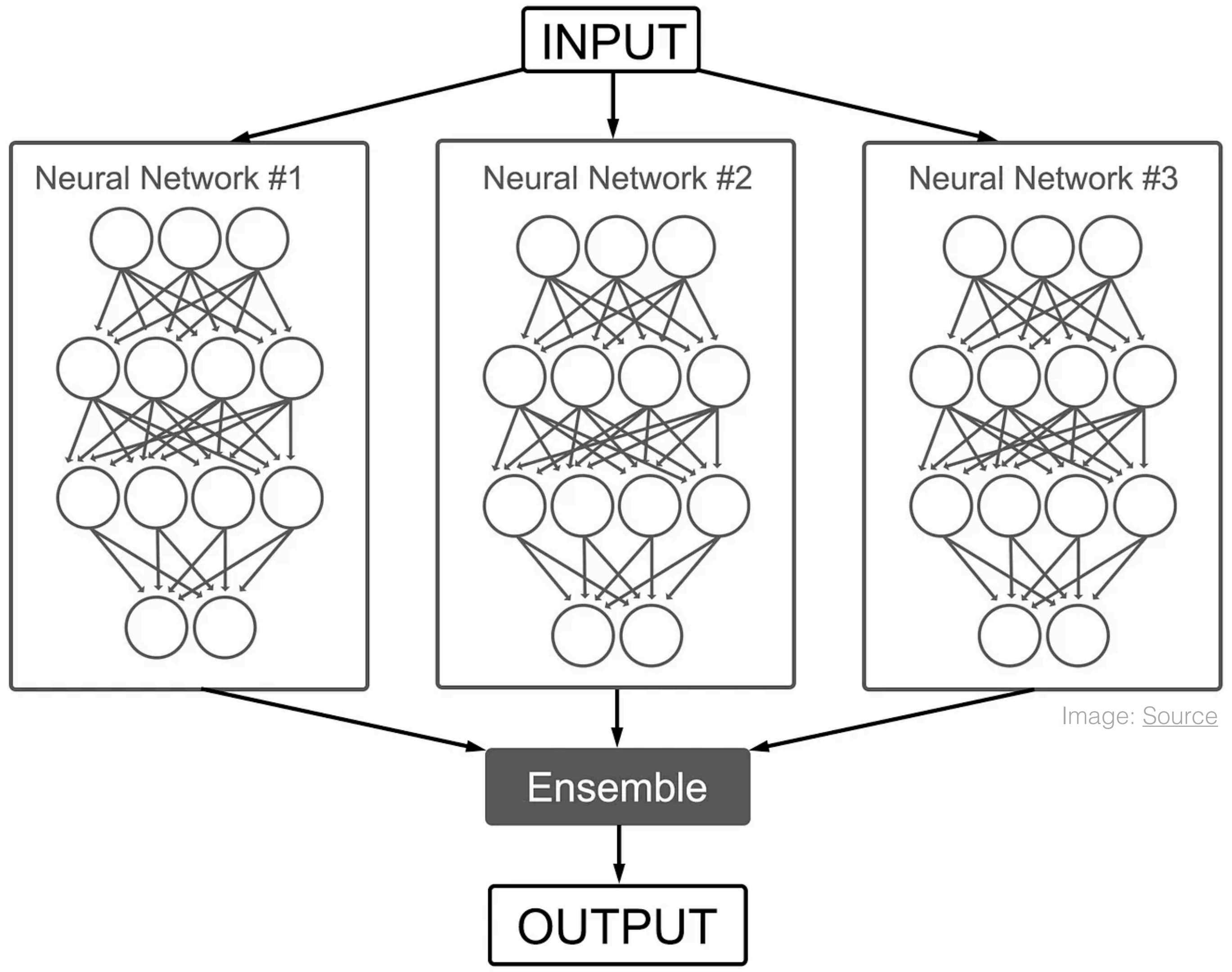


Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset

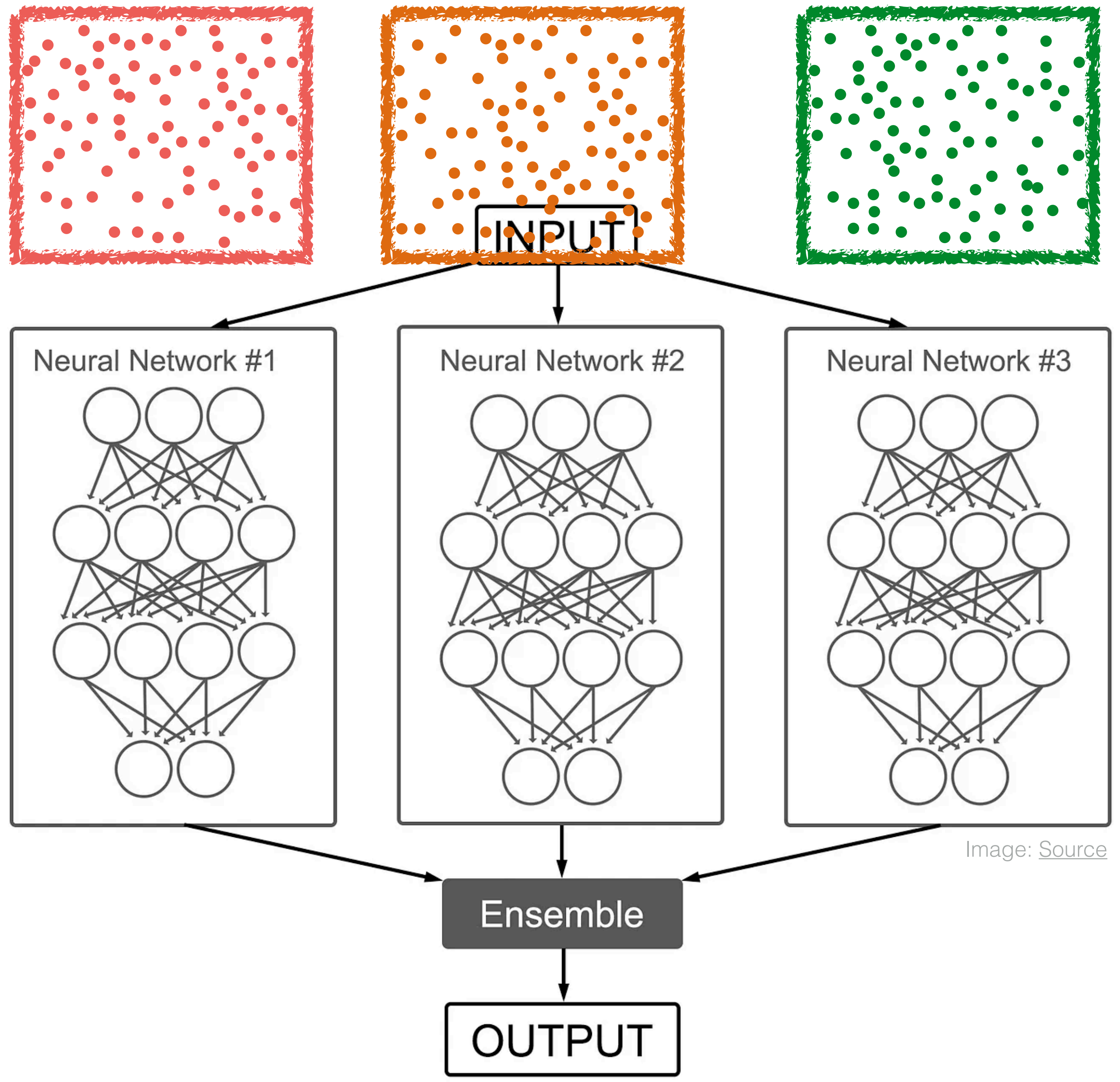
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles



Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

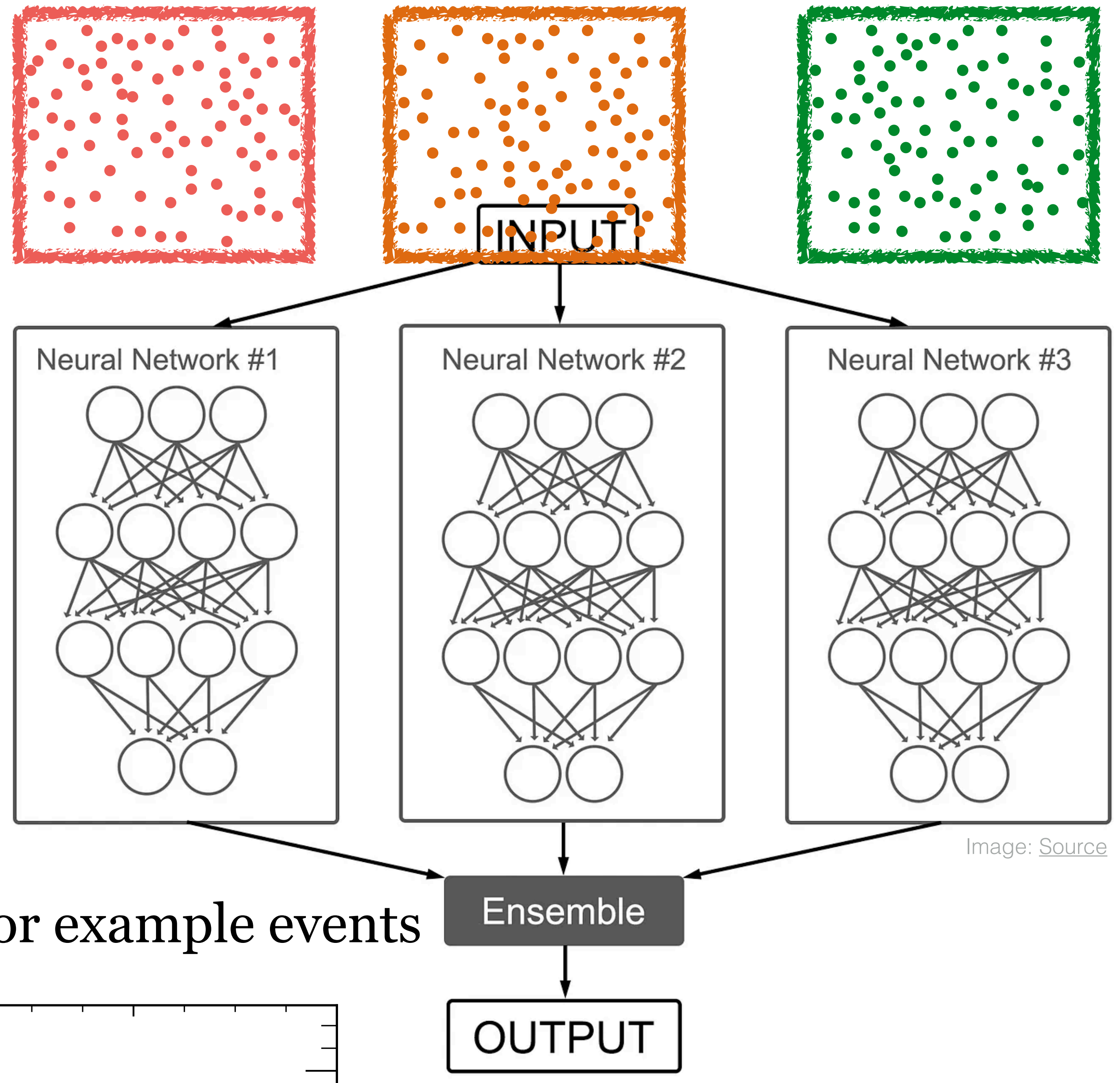
- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles



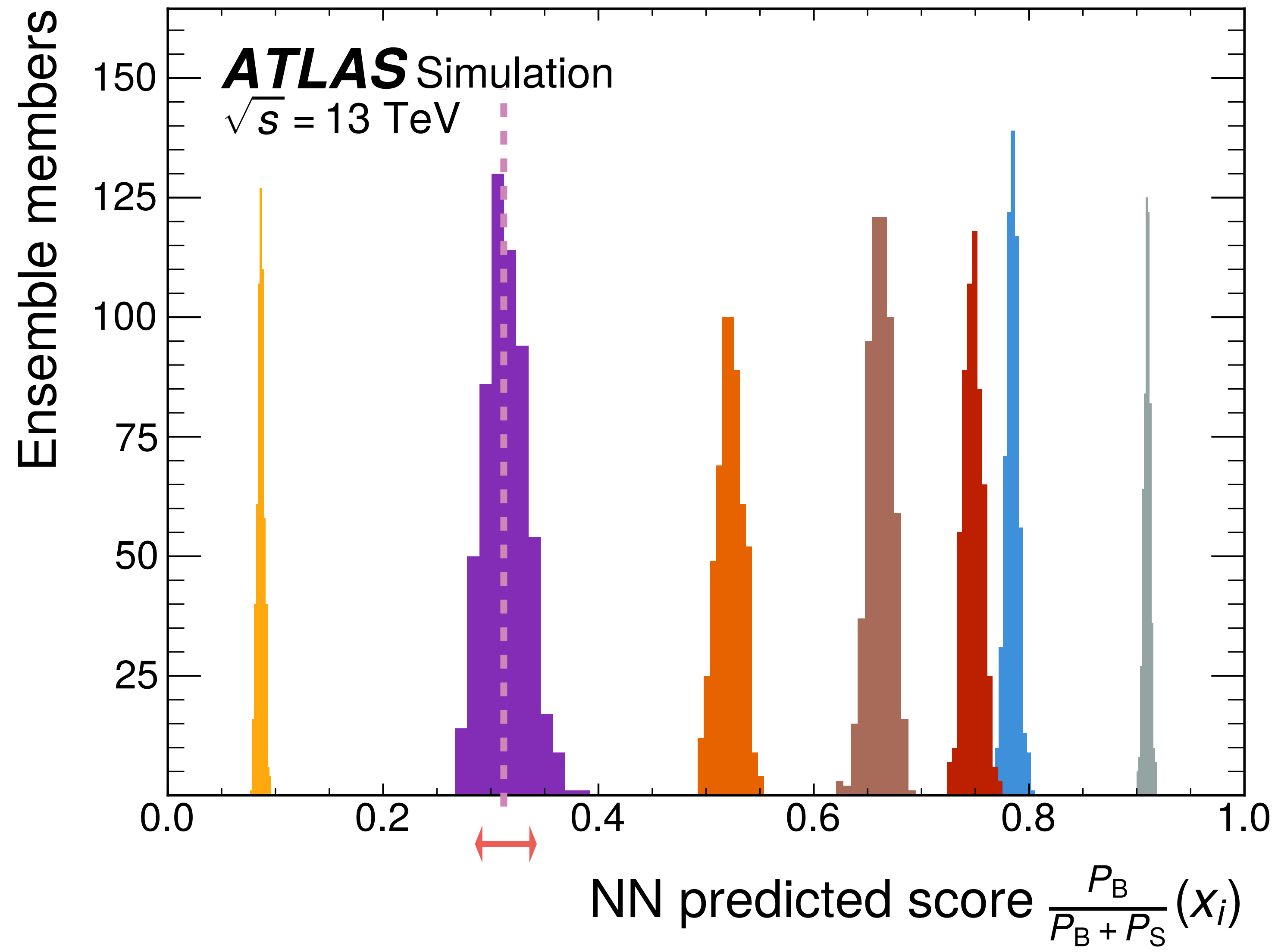
Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles



Distribution of NN predictions for example events



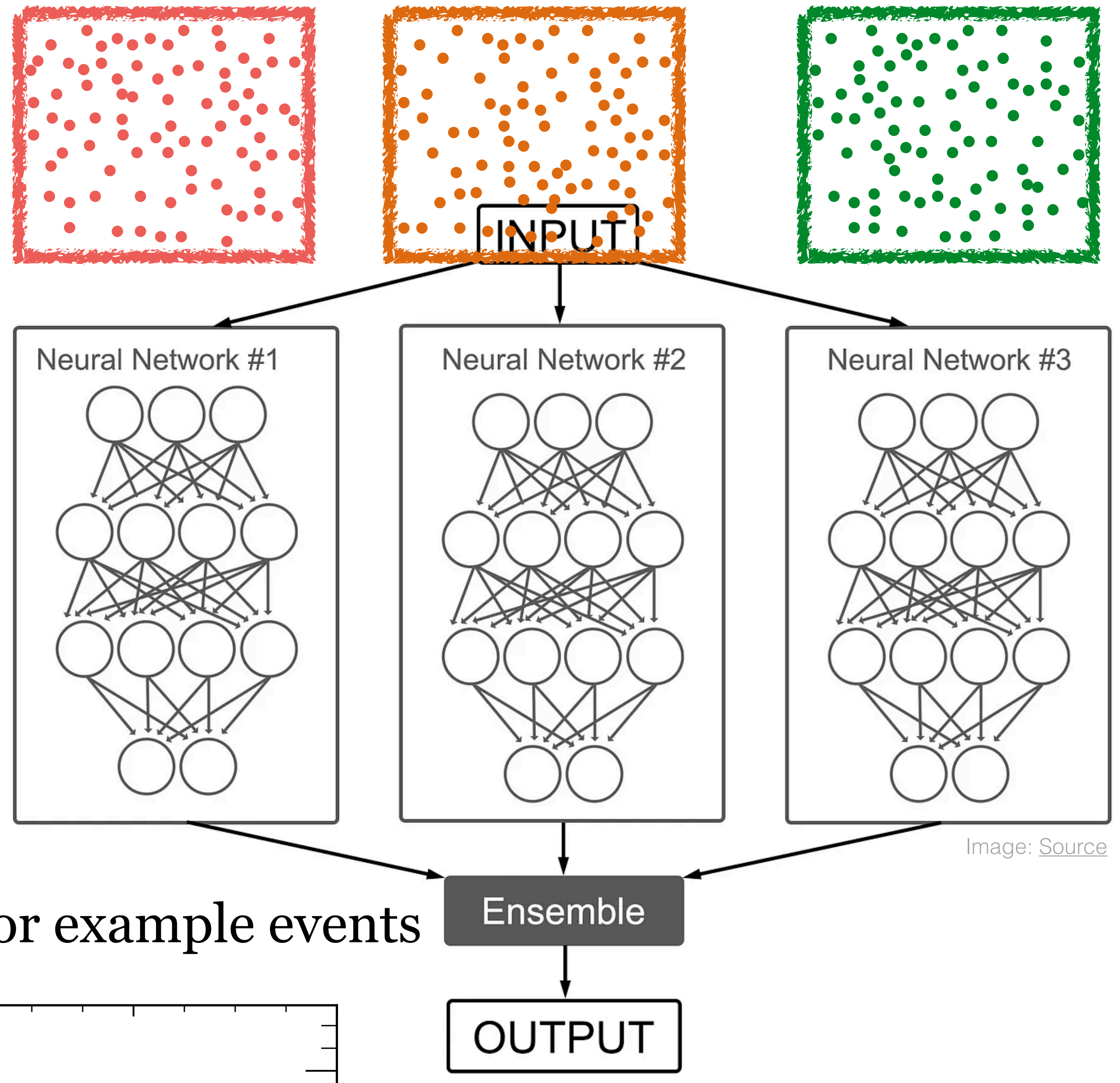
Quantifying uncertainty on estimated density ratio

$$w_i \to w_i \cdot Pois(1)$$

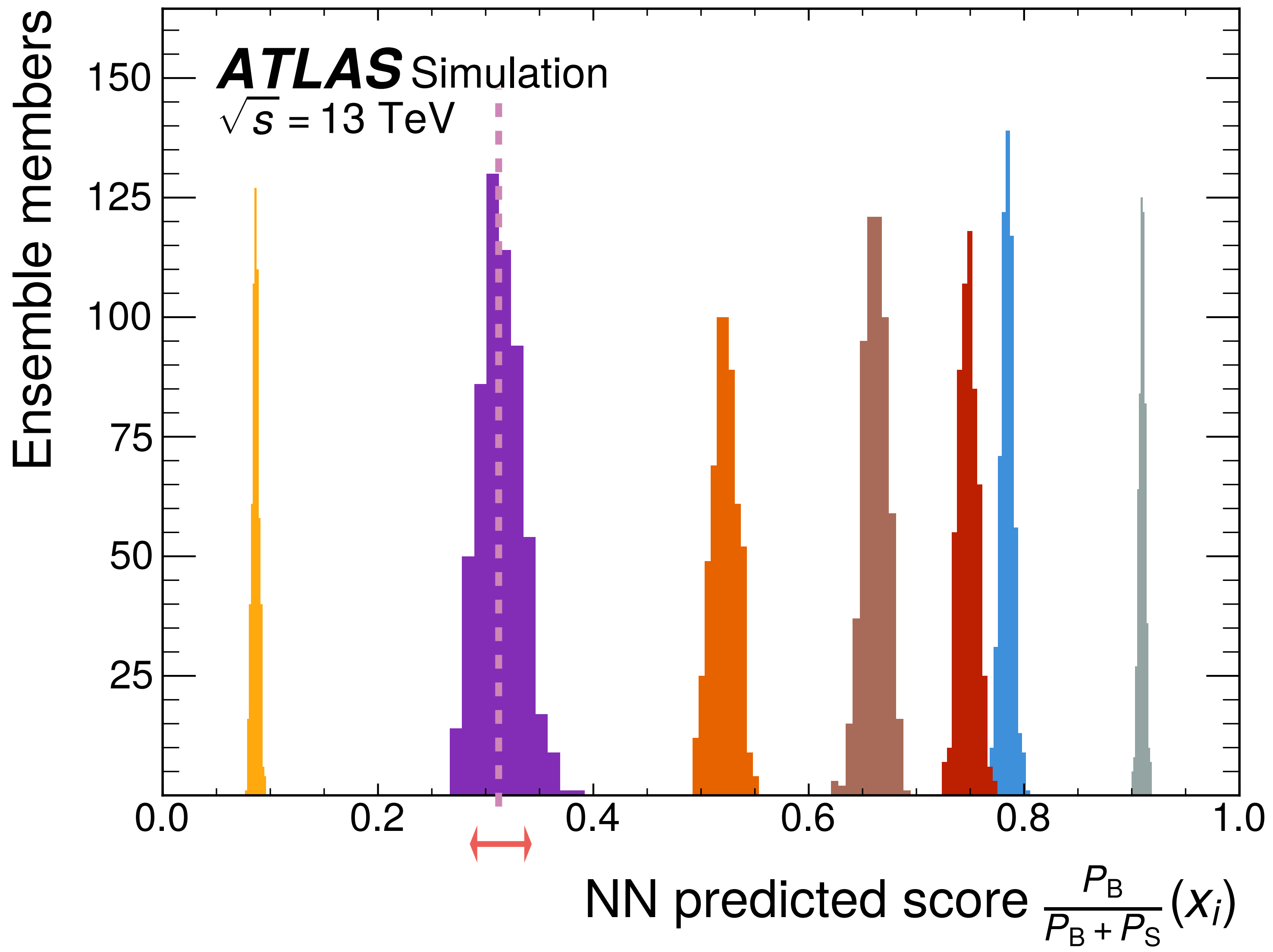
- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Ensemble average used as final prediction, estimate the variance on mean from bootstrapped ensembles
- Propagate with spurious signal method

$$f_j(\mu) \to f_j(\mu + \alpha \cdot \Delta\hat{\mu}(\mu))$$

Constraint term: $Gauss(0,1)$



Distribution of NN predictions for example events



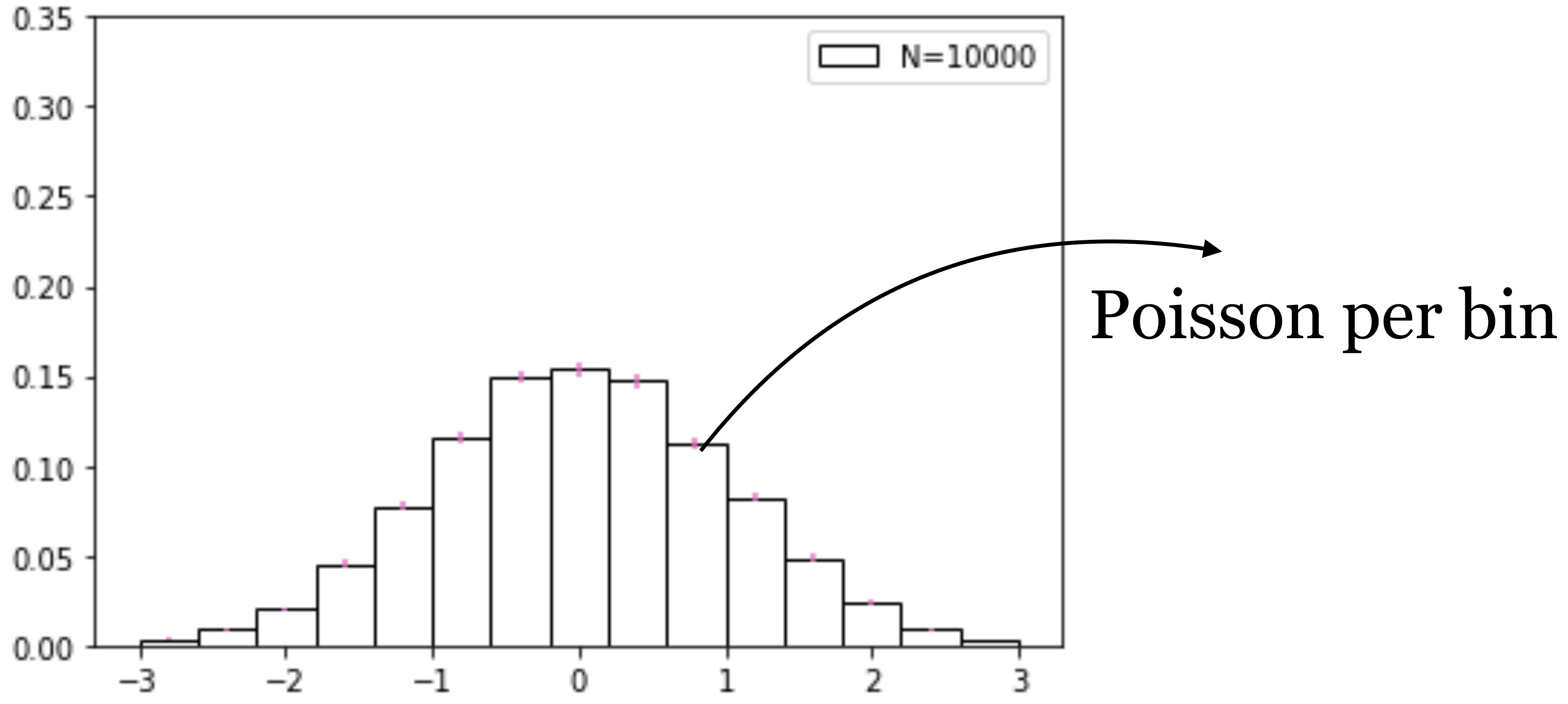
Open problems to extend to full ATLAS analysis:

- ✓ Robustness: Design and validation
- ✓ Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- ‣ Neyman Construction: Throwing toys in a per-event analysis

Generating event-level pseudo-experiments

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

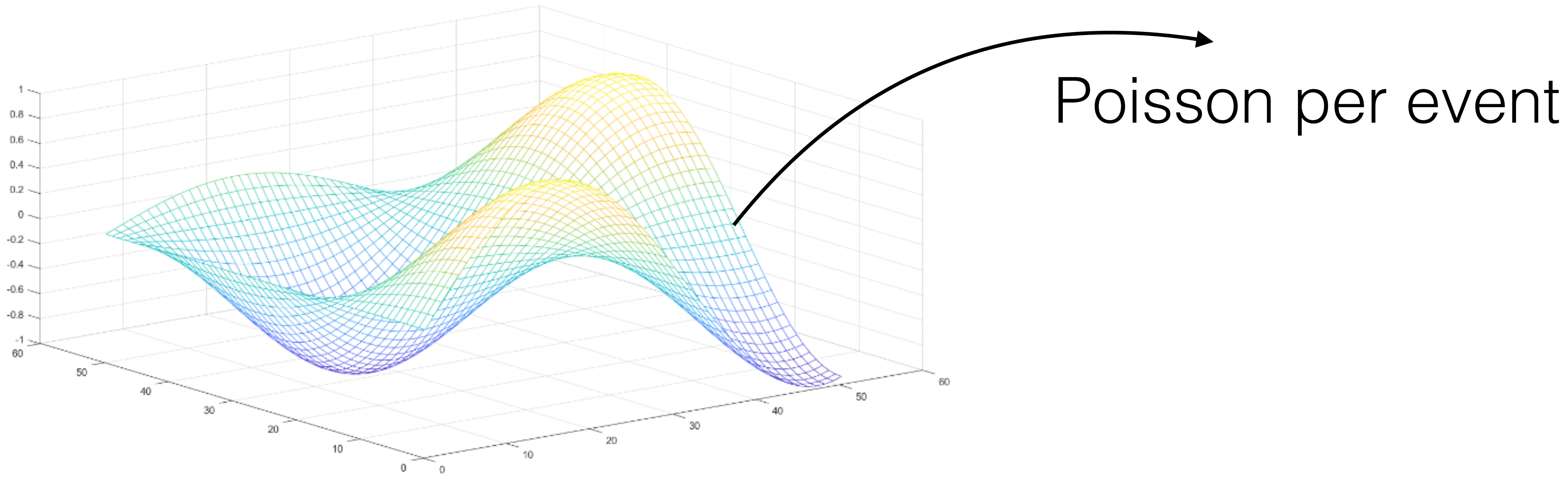
Traditionally:



Asimov Histogram

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

NSBI:



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

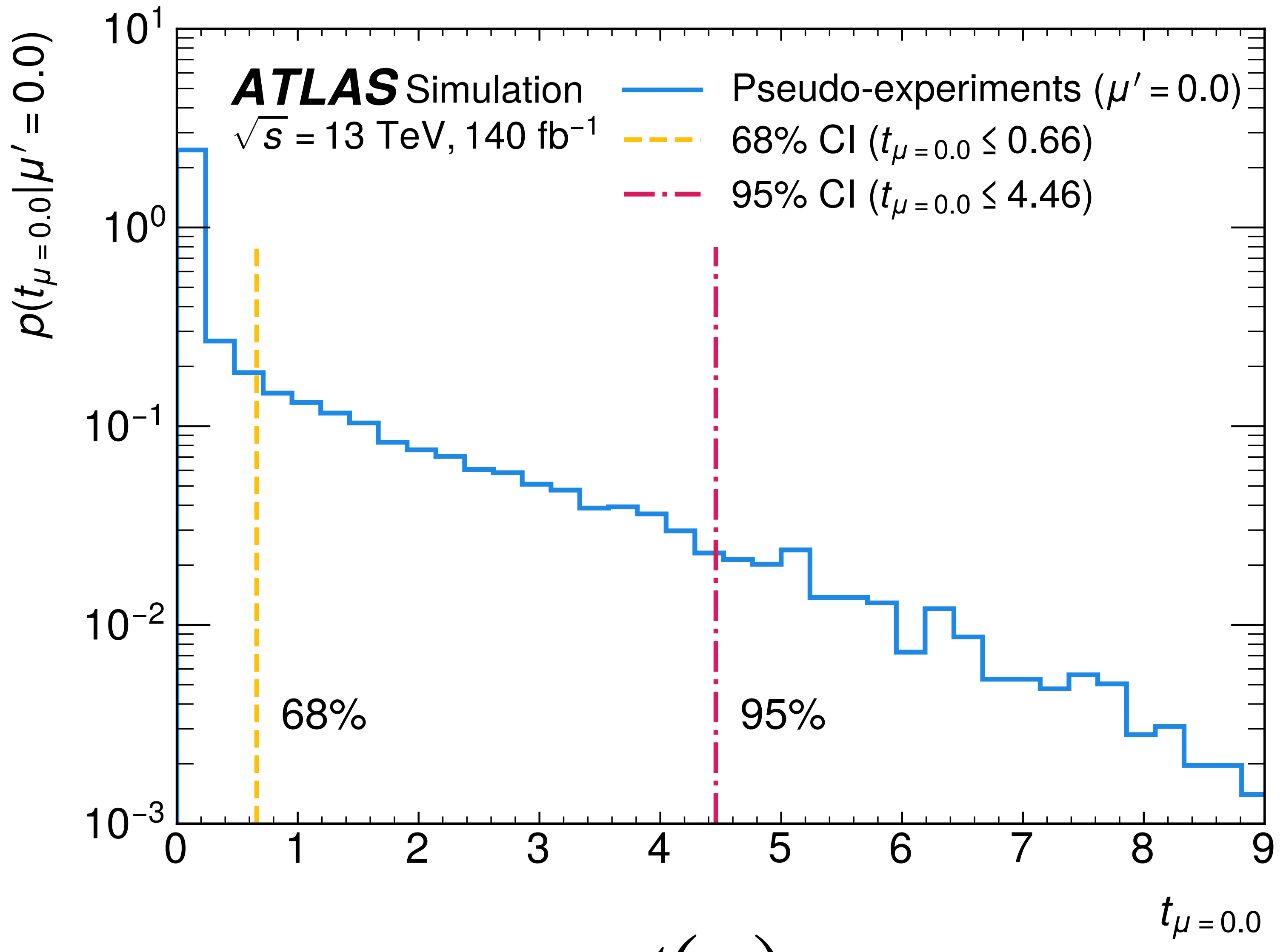
'Unweighted' events, i.e. integer weights

Negative weights? See backup

Neyman Construction

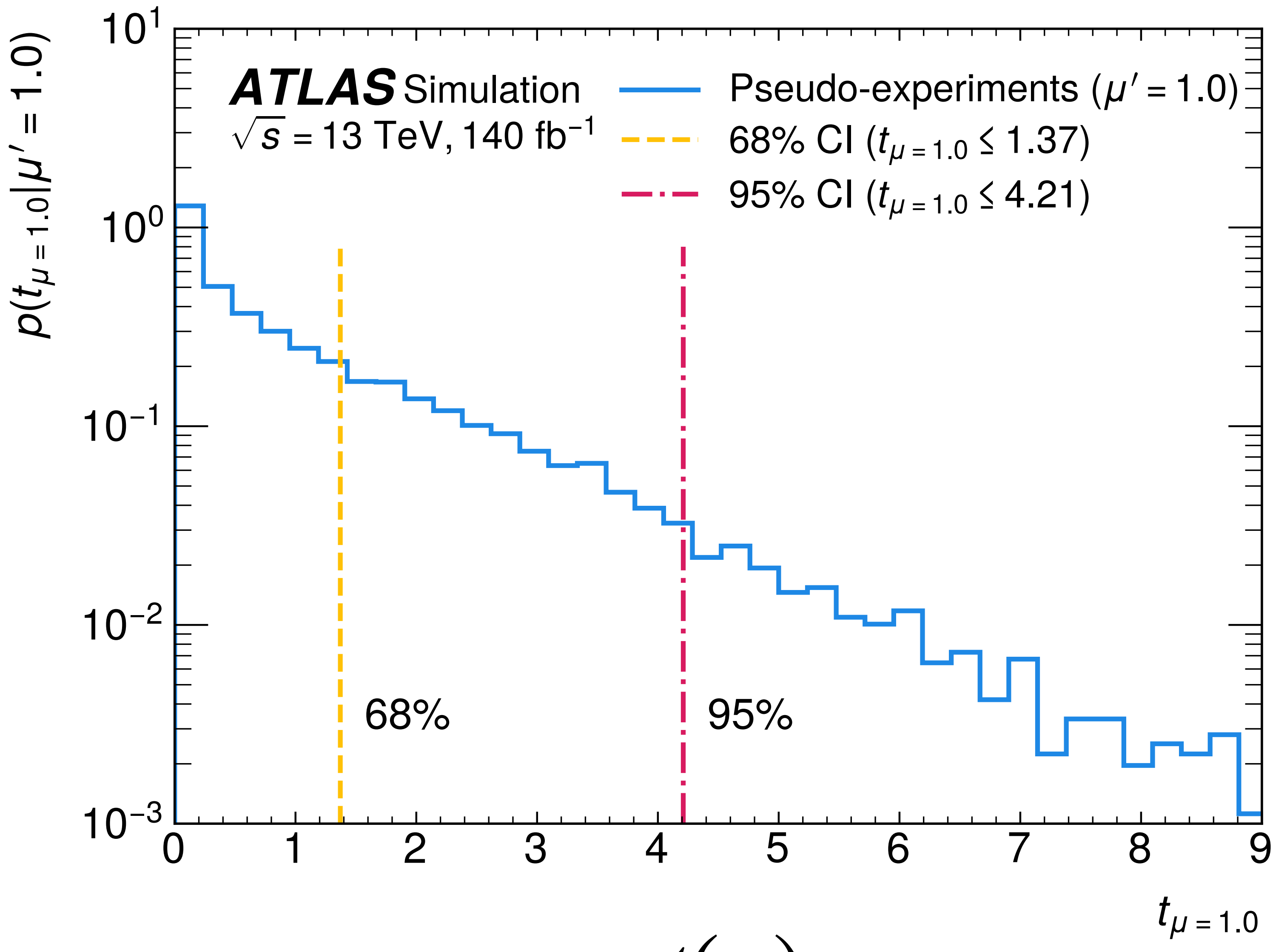
- To build confidence intervals, we need to 'invert the hypothesis test'
- Generate pseudo-experiments ('toys') and determine 68 % & 95 % CI as a function of parameter of interest

True $\mu = 0$



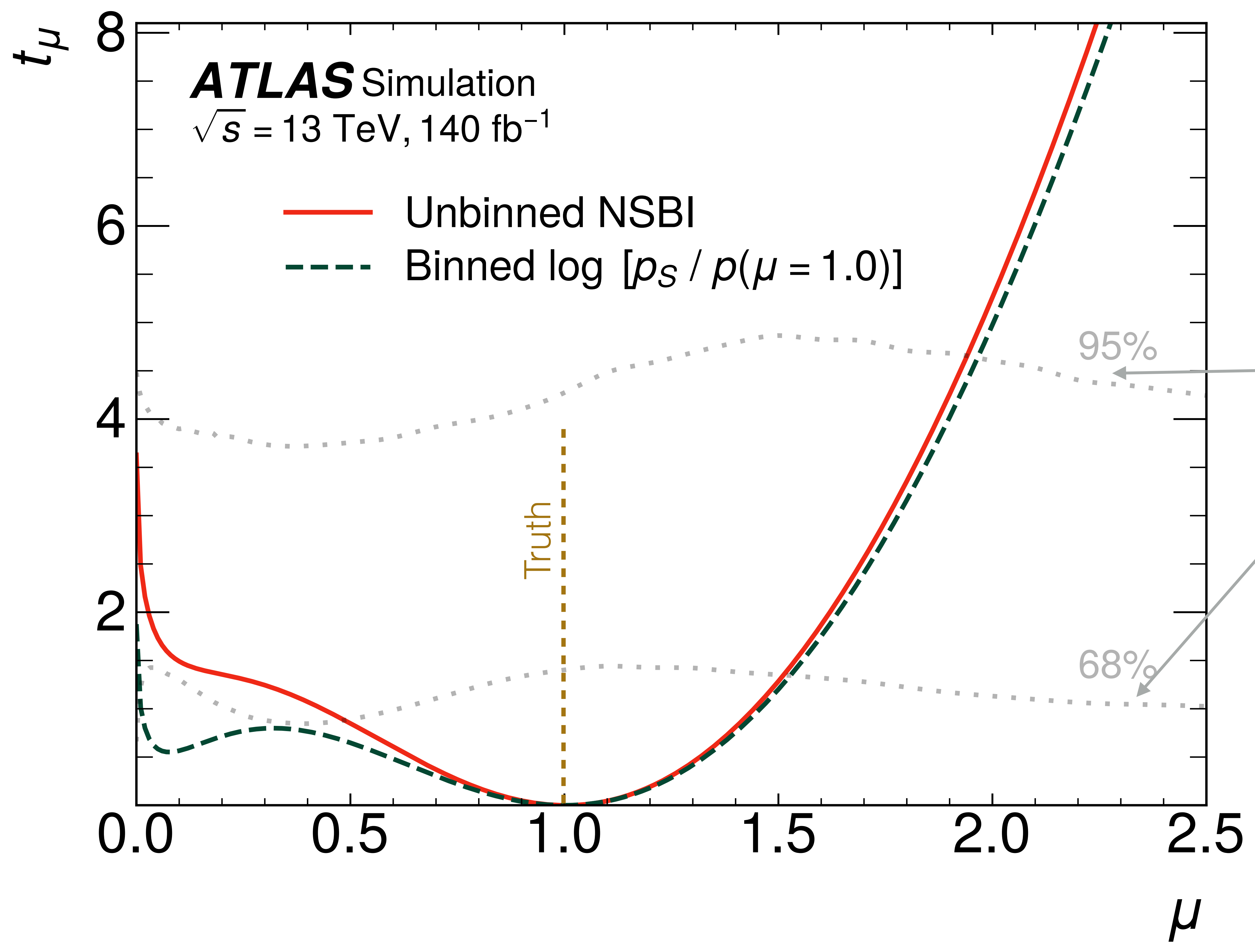
$t(\mu)$

True $\mu = 1$



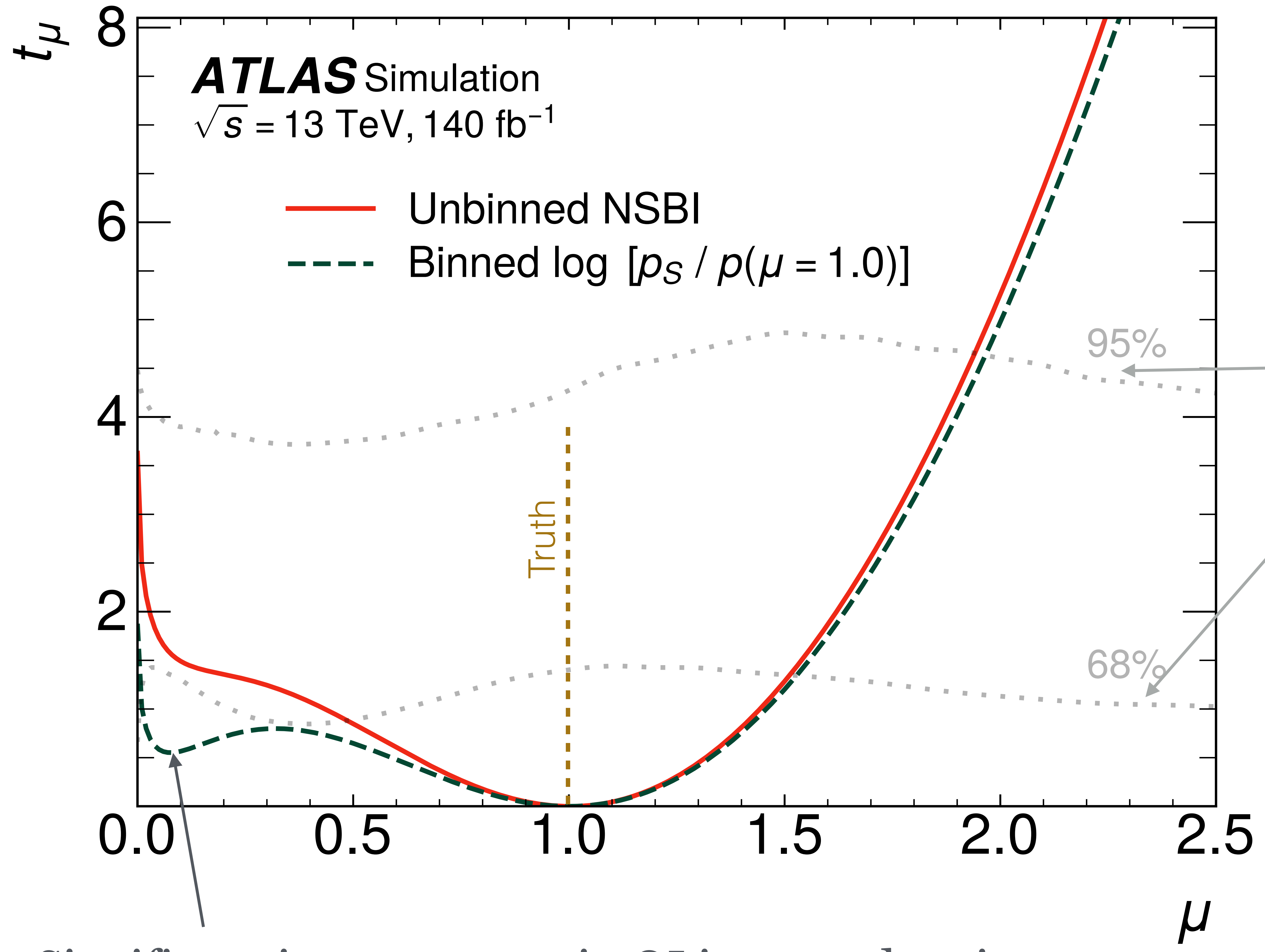
$t(\mu)$

Confidence belts



Similar to structure seen in histogram analysis

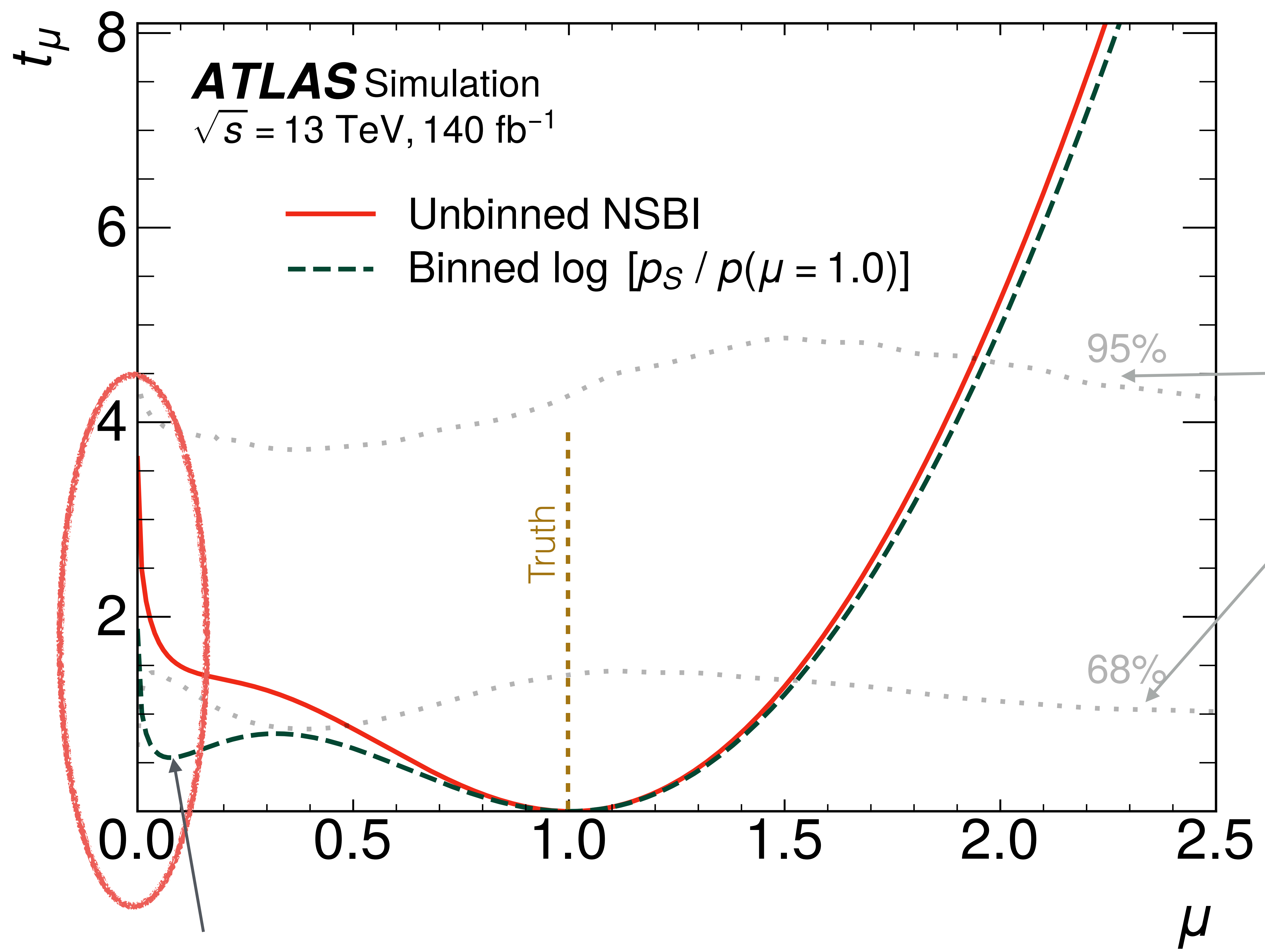
Confidence belts



Similar to structure seen in histogram analysis

Significant improvement in QI impacted region

Confidence belts



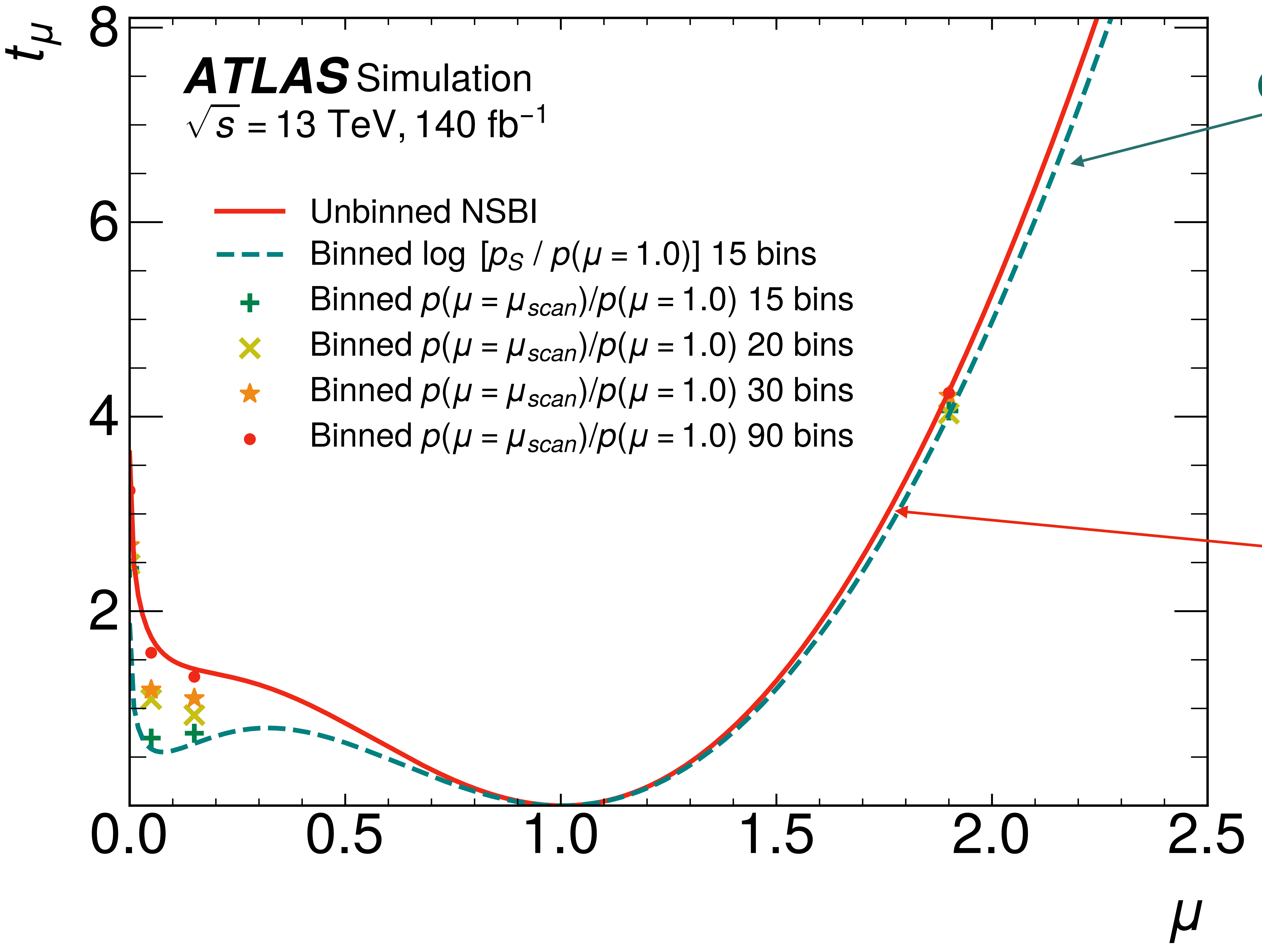
Similar to structure seen in histogram analysis

Significant improvement in QI impacted region

Expect a dramatic improvement in ability to reject null hypothesis

Why does NSBI work better than traditional analyses?

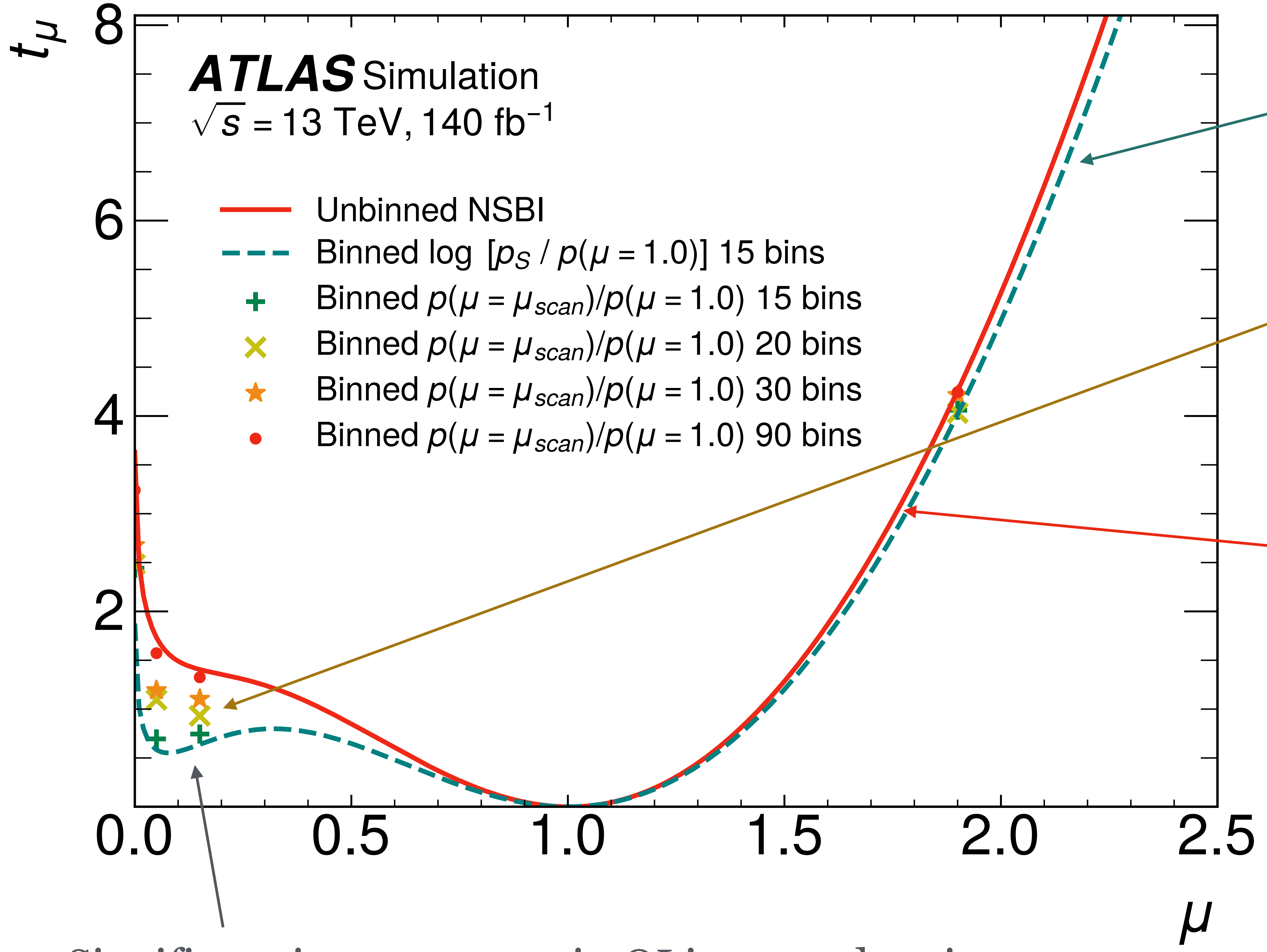
Why does it work better than traditional analyses?



$O_{\text{fixed}} = \log \frac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

NSBI: Parameterised, unbinned

Why does it work better than traditional analyses?



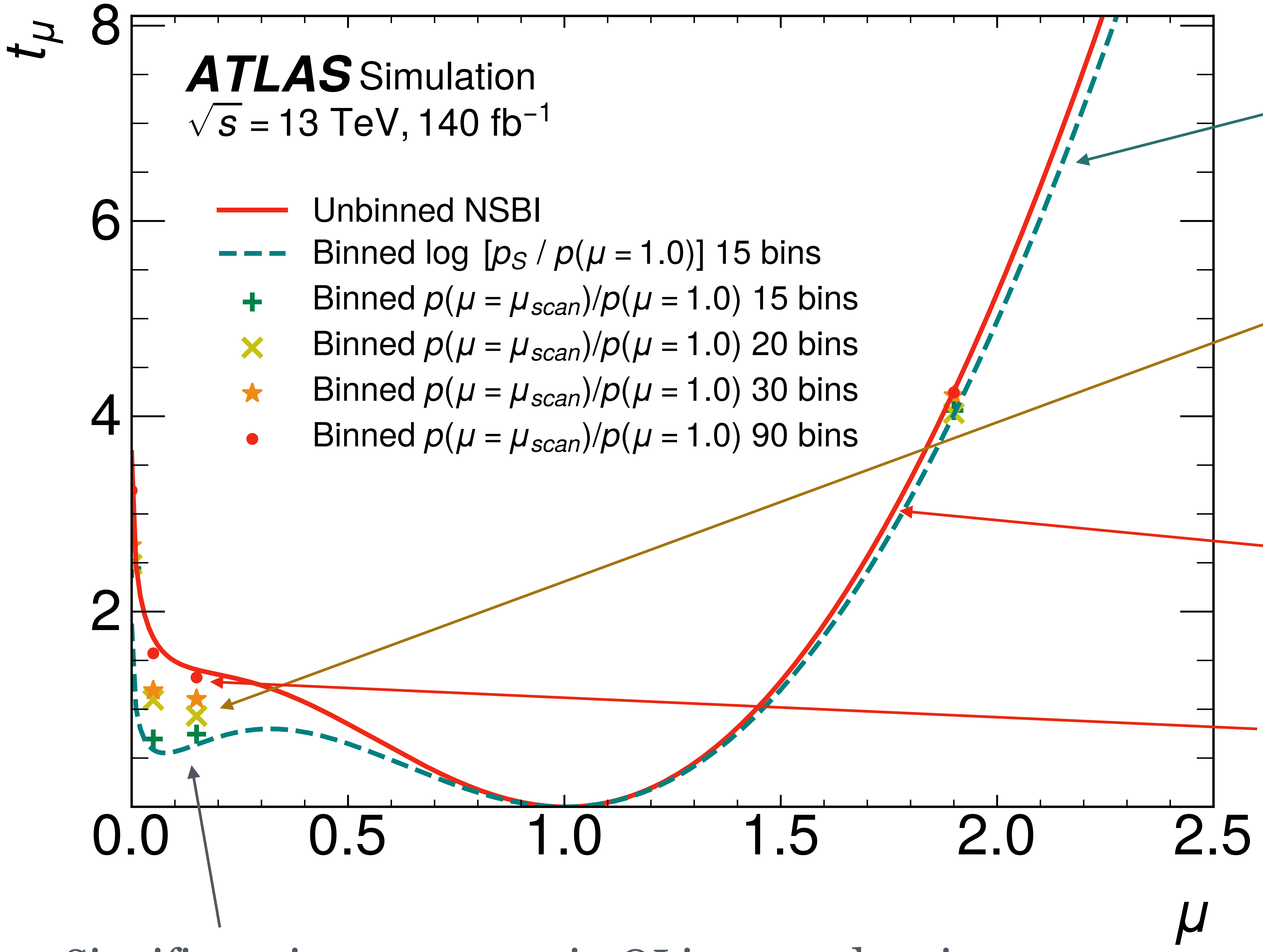
$O_{\text{fixed}} = \log \dfrac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

$O_\mu = \dfrac{p(x_i | \mu)}{p(x_i | \mu = 1)}$: Parameterised observable, histogram fit

NSBI: Parameterised, unbinned

Significant improvement in QI impacted region

Why does it work better than traditional analyses?



$O_{\text{fixed}} = \log \dfrac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

$O_\mu = \dfrac{p(x_i \mid \mu)}{p(x_i \mid \mu = 1)}$: Parameterised observable, histogram fit

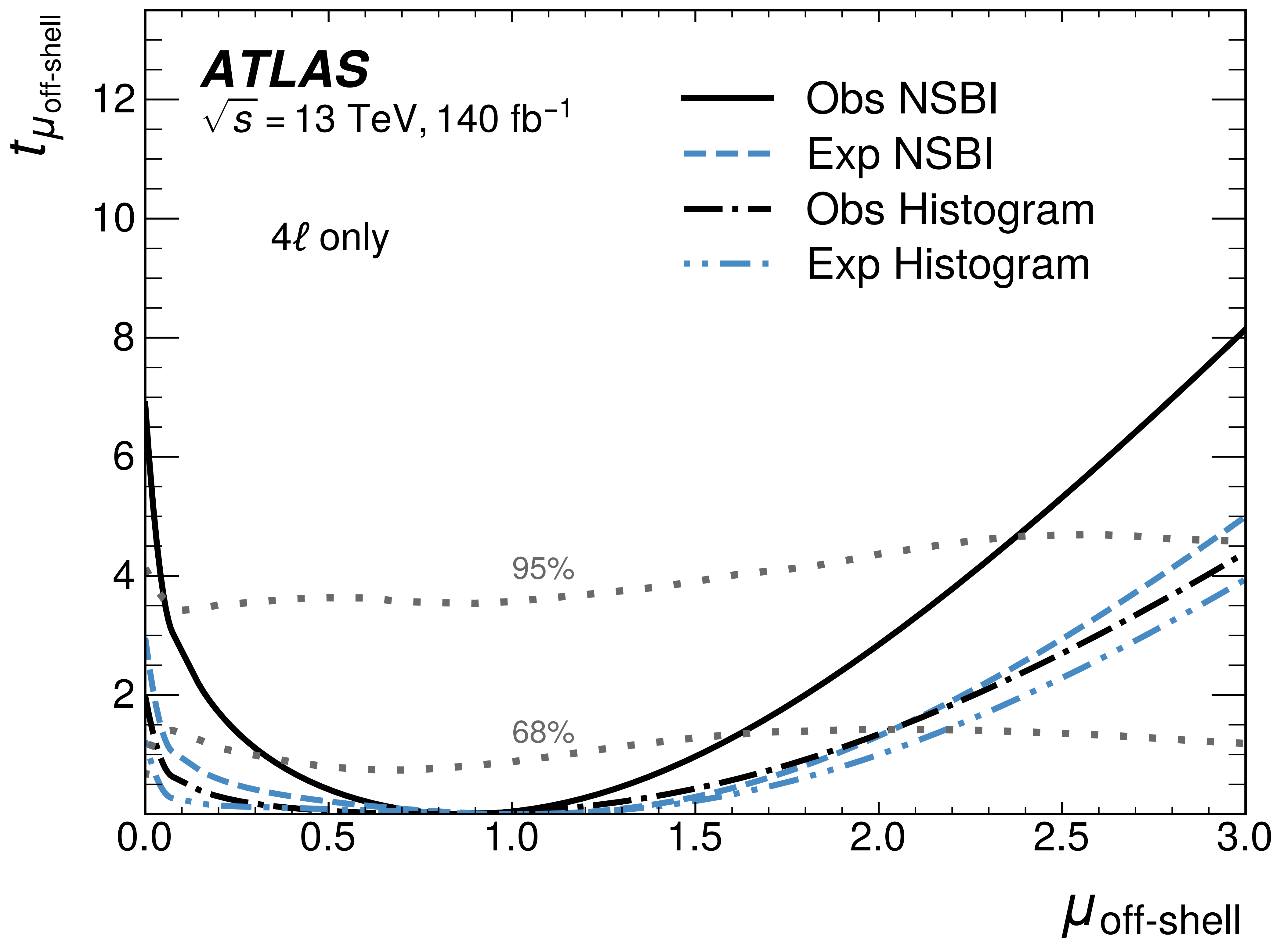
NSBI: Parameterised, unbinned

O_μ approaches NSBI as nBins $\to \infty$

Significant improvement in QI impacted region

Final results: Apply on real data and supersede previous Run2 paper!

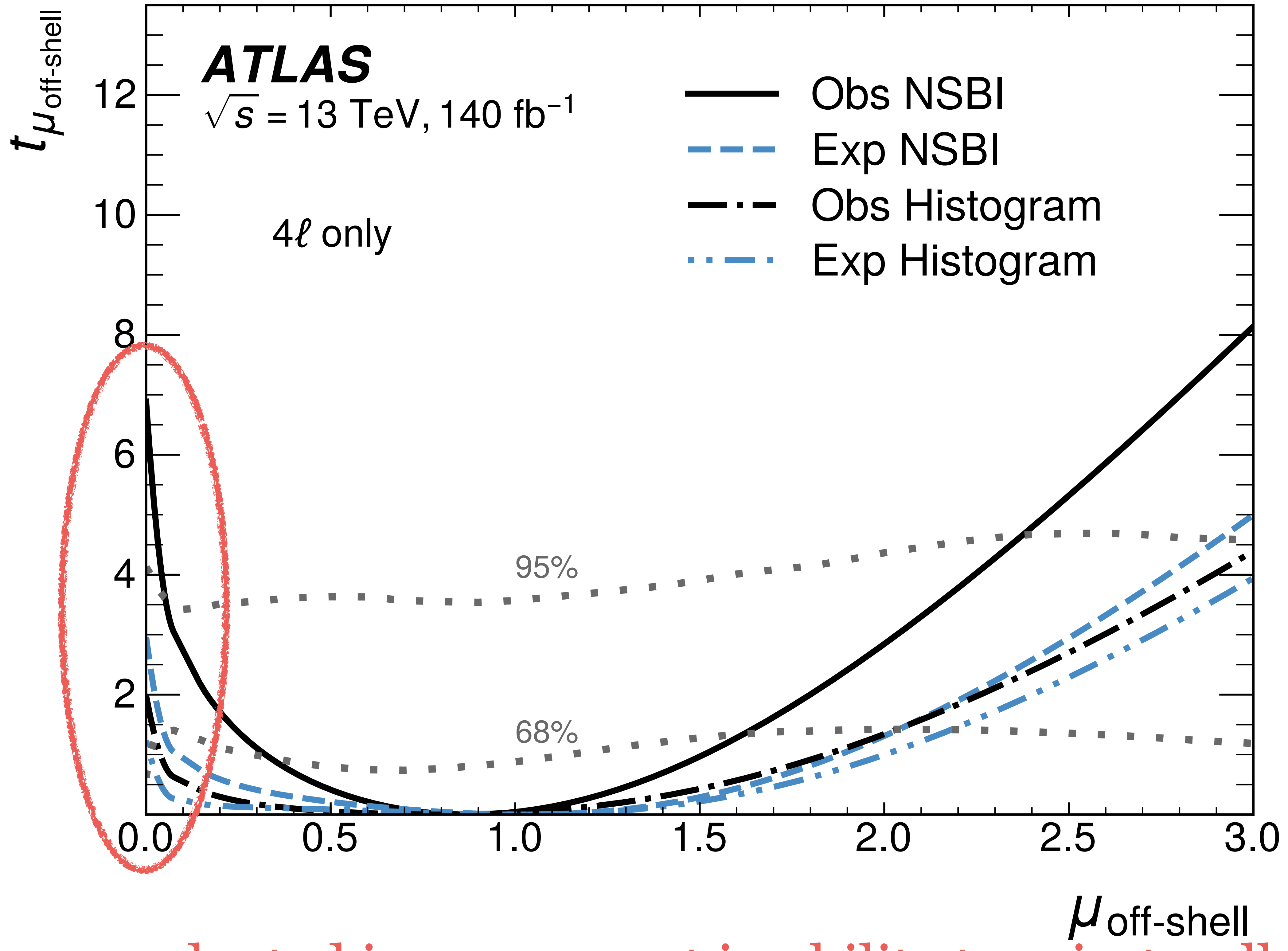
NSBI vs histogram analysis



Observed data happens to provide stronger than expected constrains for both hist and NSBI (consistent)

Final results: Apply on real data and supersede previous Run2 paper!

NSBI vs histogram analysis

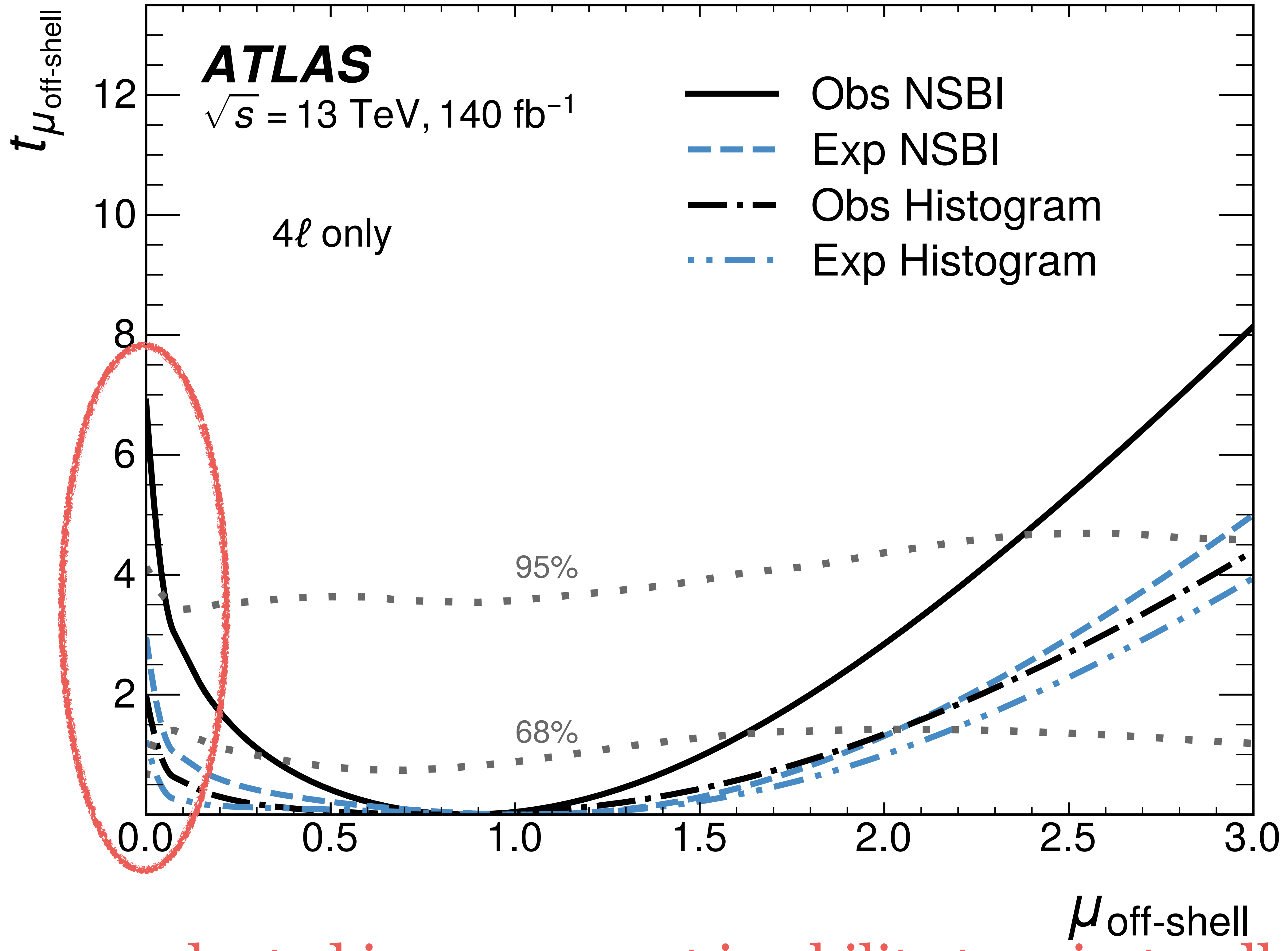


Unprecedented improvement in ability to reject null hypothesis! (2.6x gain over previous method)

Observed data happens to provide stronger than expected constrains for both hist and NSBI (consistent)

Final results: Apply on real data and supersede previous Run2 paper!

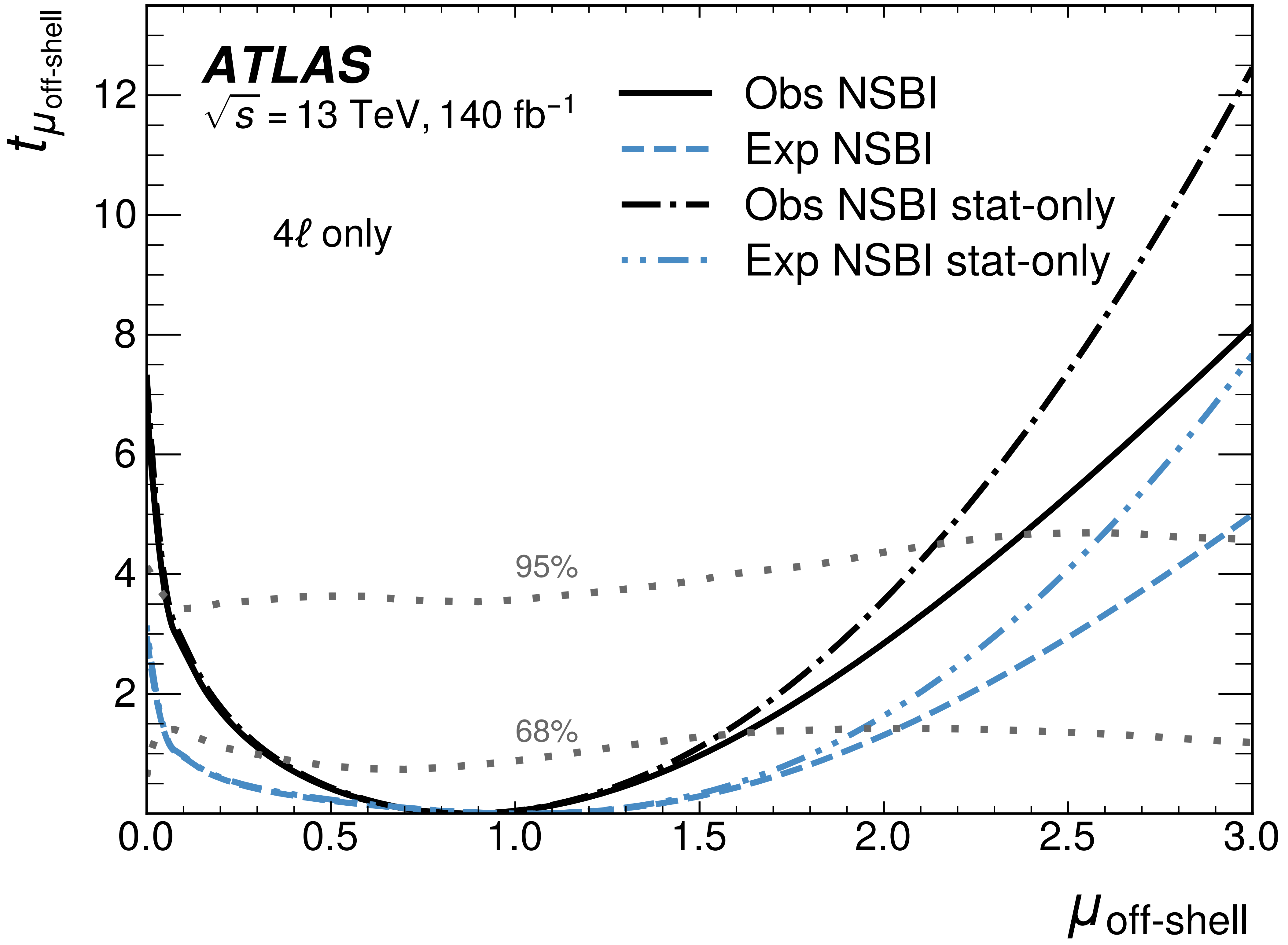
NSBI vs histogram analysis



Unprecedented improvement in ability to reject null hypothesis! (2.6x gain over previous method)

Observed data happens to provide stronger than expected constrains for both hist and NSBI (consistent)

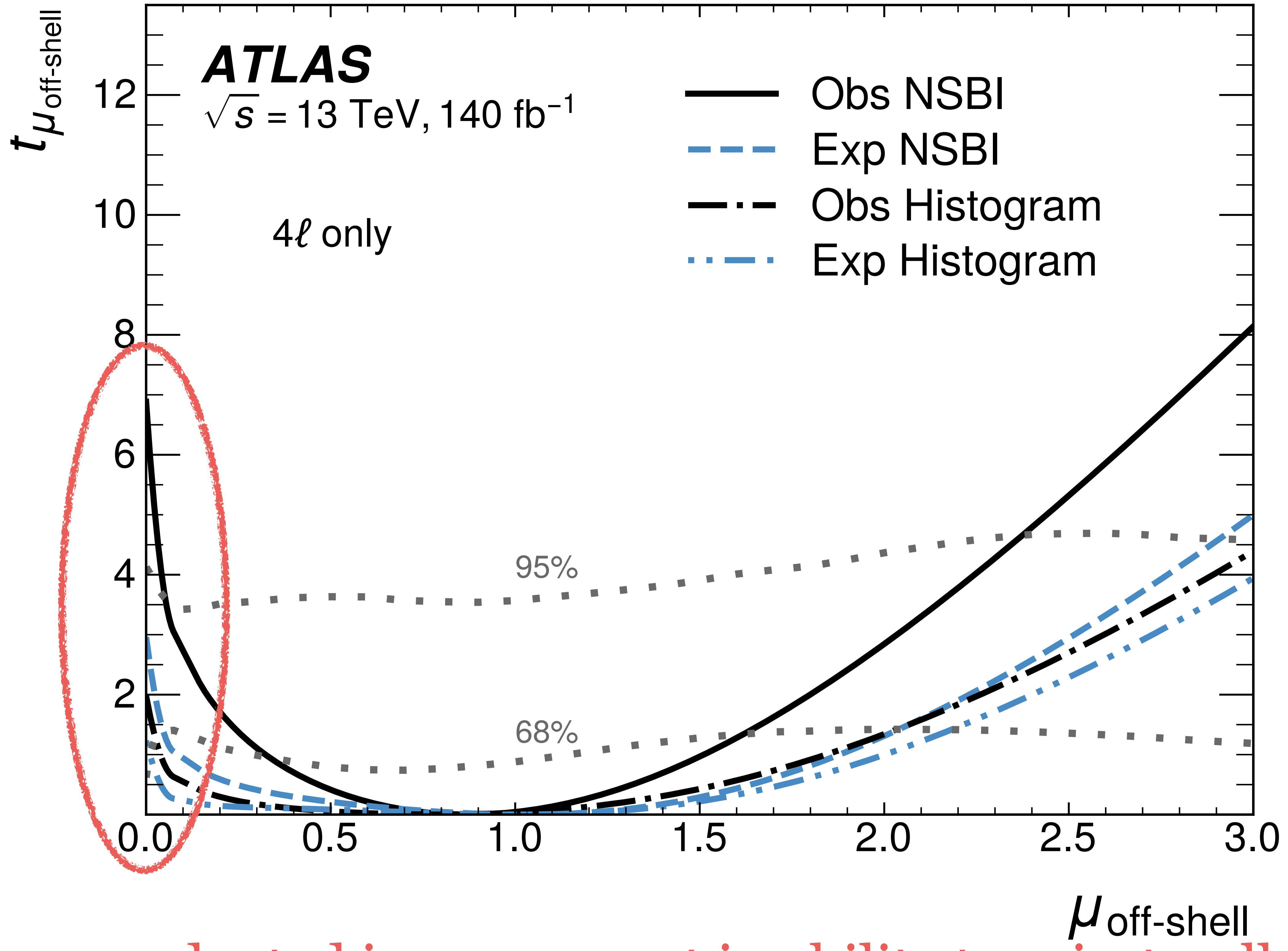
Stat-only vs Stat+Syst uncertainties



Nuisance parameters decrease sensitivity, as expected

Final results: Apply on real data and supersede previous Run2 paper!

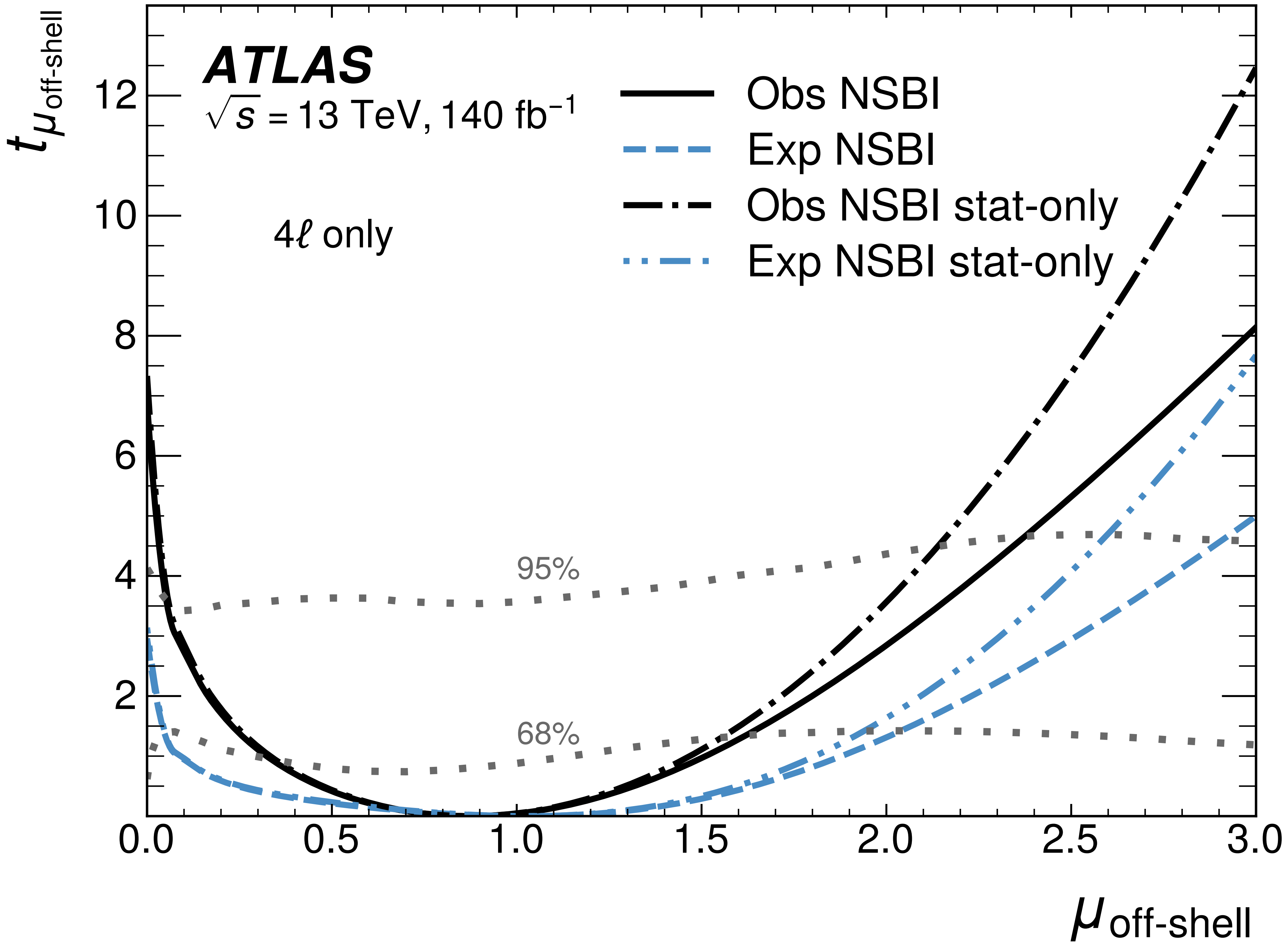
NSBI vs histogram analysis



Unprecedented improvement in ability to reject null hypothesis! (2.6x gain over previous method)

Observed data happens to provide stronger than expected constrains for both hist and NSBI (consistent)

Stat-only vs Stat+Syst uncertainties

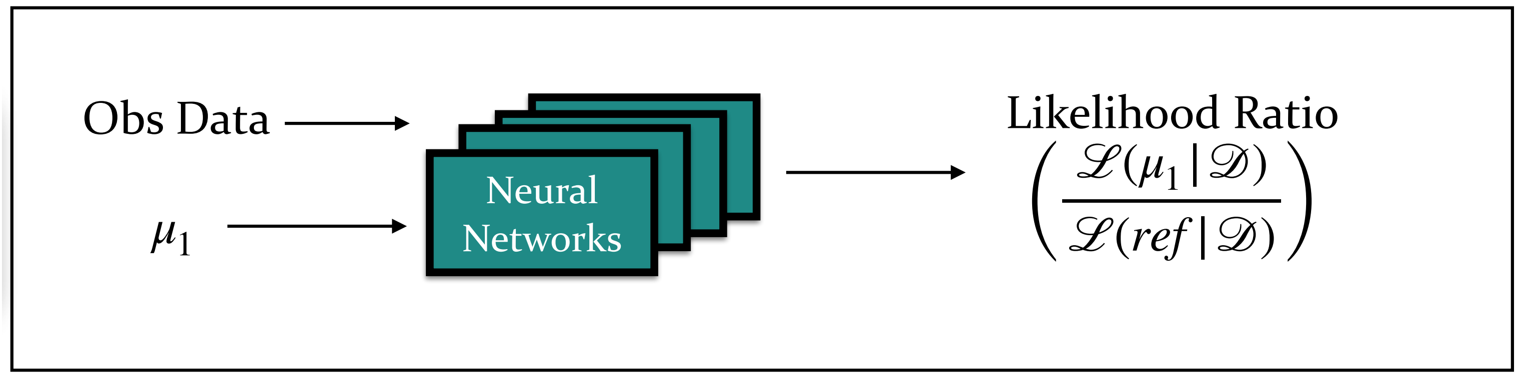


Full results in backup

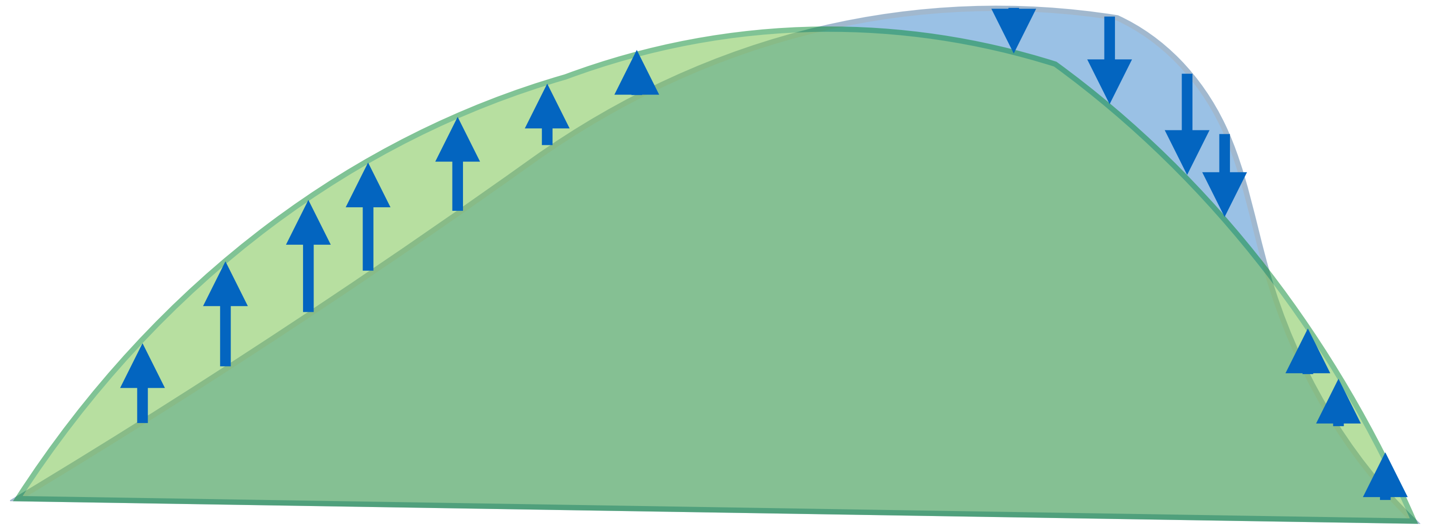
Nuisance parameters decrease sensitivity, as expected

Summary

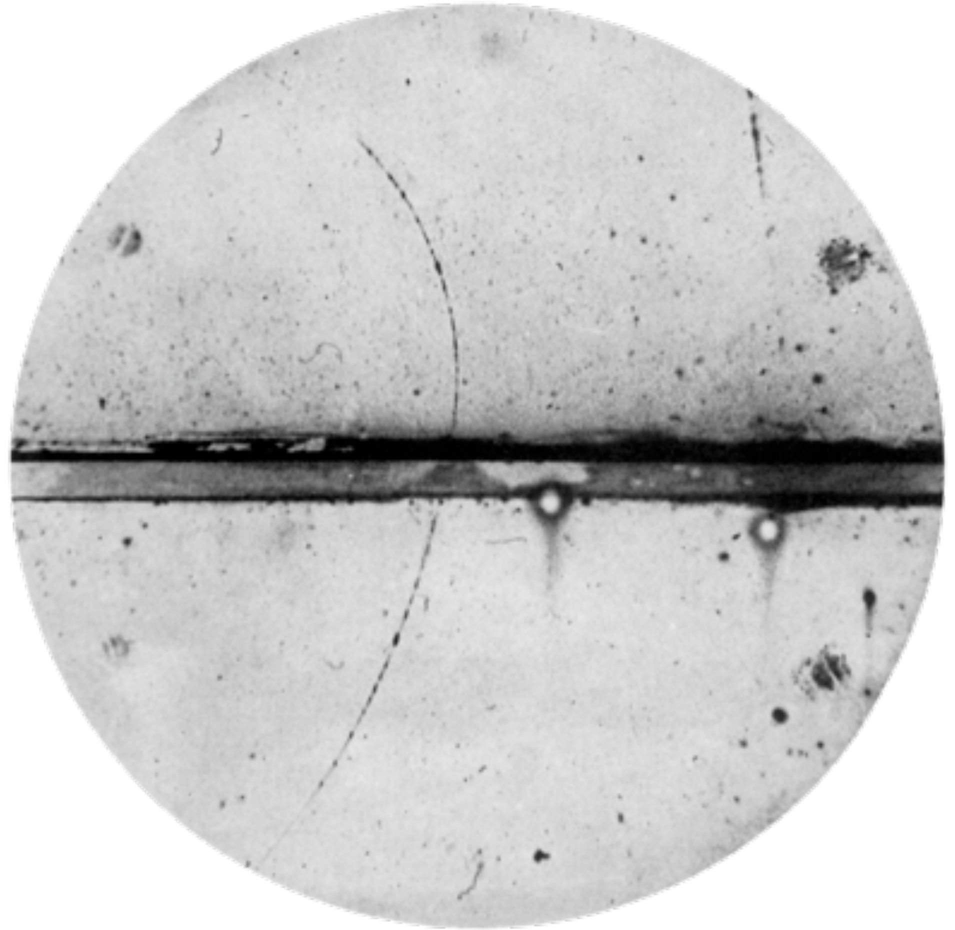
- Quantum interference breaks assumptions in traditional statistical methods at LHC
- Neural inference can optimally handle these challenges:
  - Shown in phenomenology study
  - ATLAS developed method for deployment including systematics
  - Re-analysed Run 2 data and achieved a dramatic improvement in sensitivity ($H \to 4l$)



- NSBI has wide-ranging applications, in particle physics, astrophysics and beyond!
- Weaknesses: Same as traditional analyses (systematics, training statistics)
  - Developed diagnostic tools to identify issues



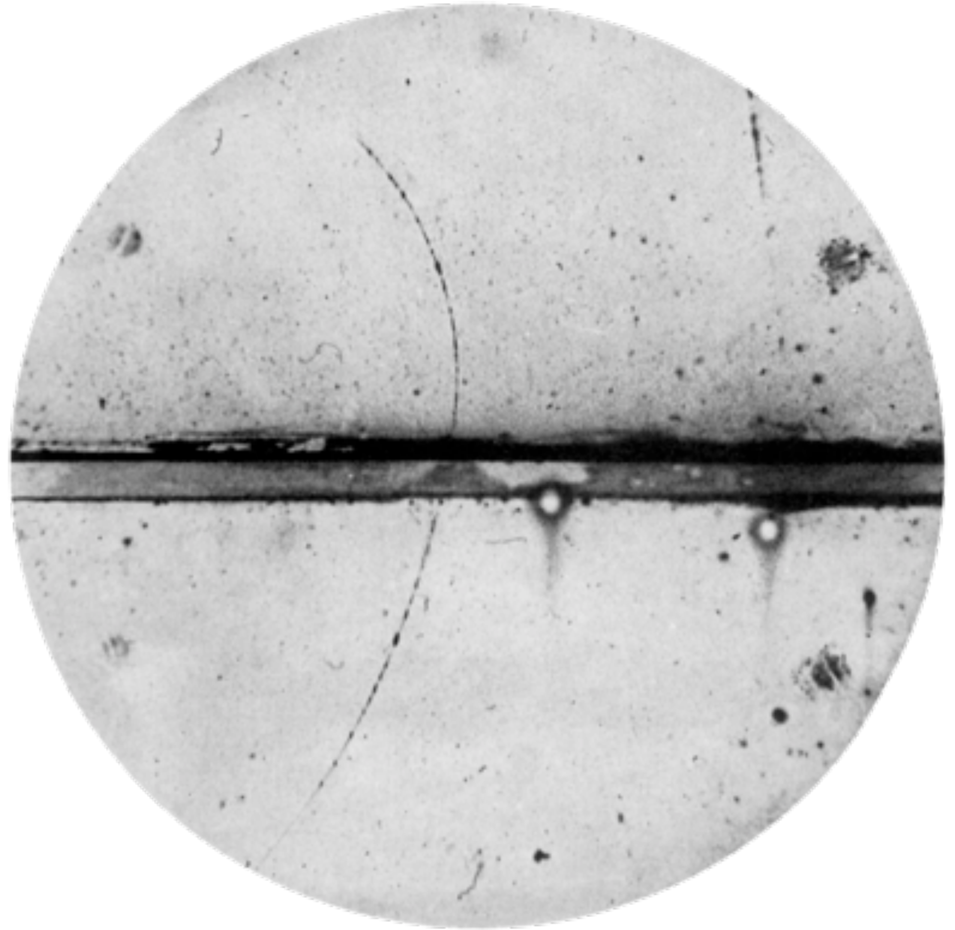
Positron discovery (1930s)



Single event

Positron discovery (1930s)

Top quark discovery (1990s)



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

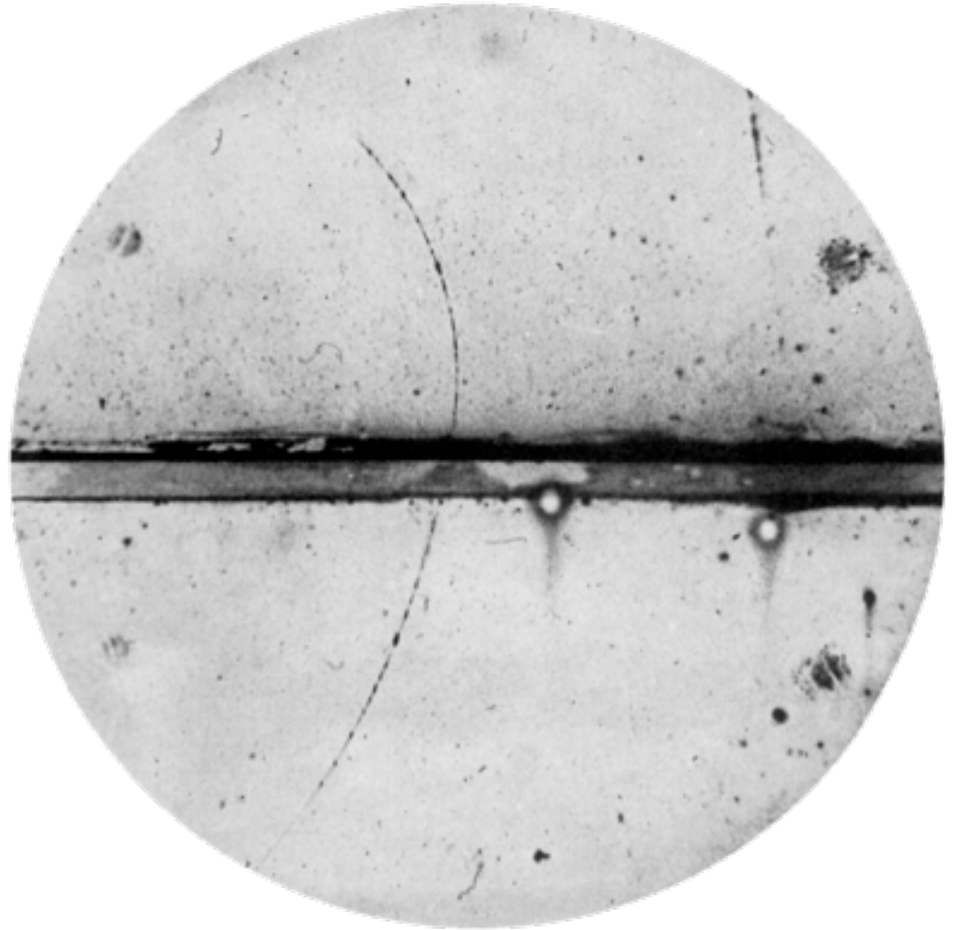
Single event

Multiple events:
Cut-and-count

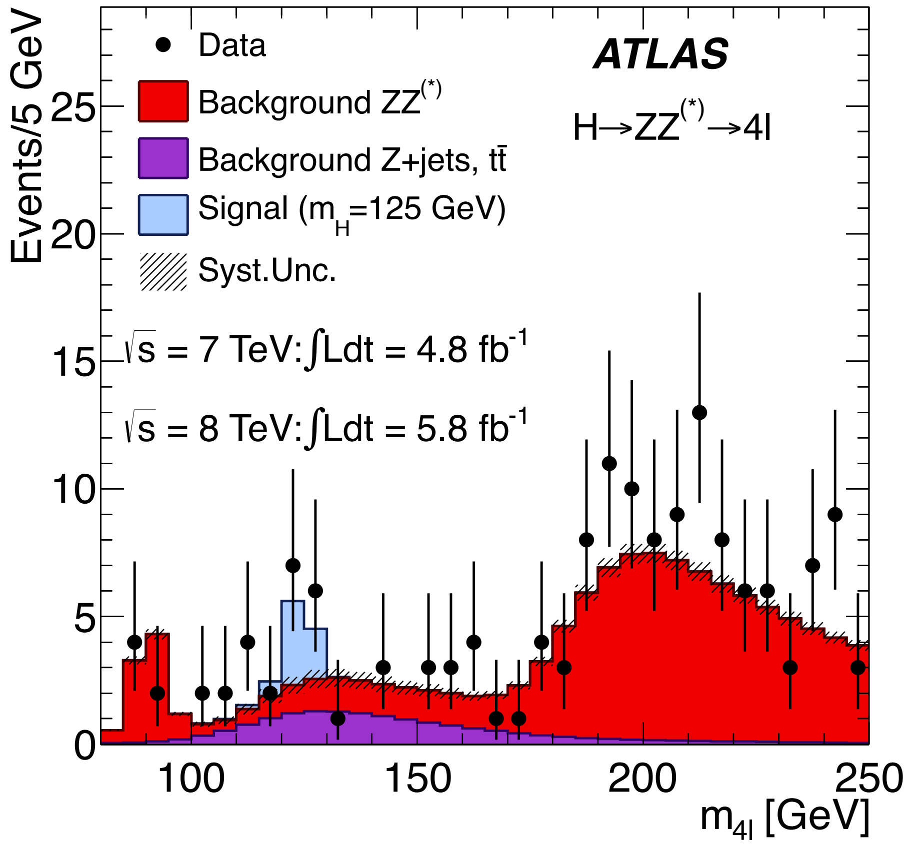
Positron discovery (1930s)

Top quark discovery (1990s)

Higgs boson discovery (2010s)



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

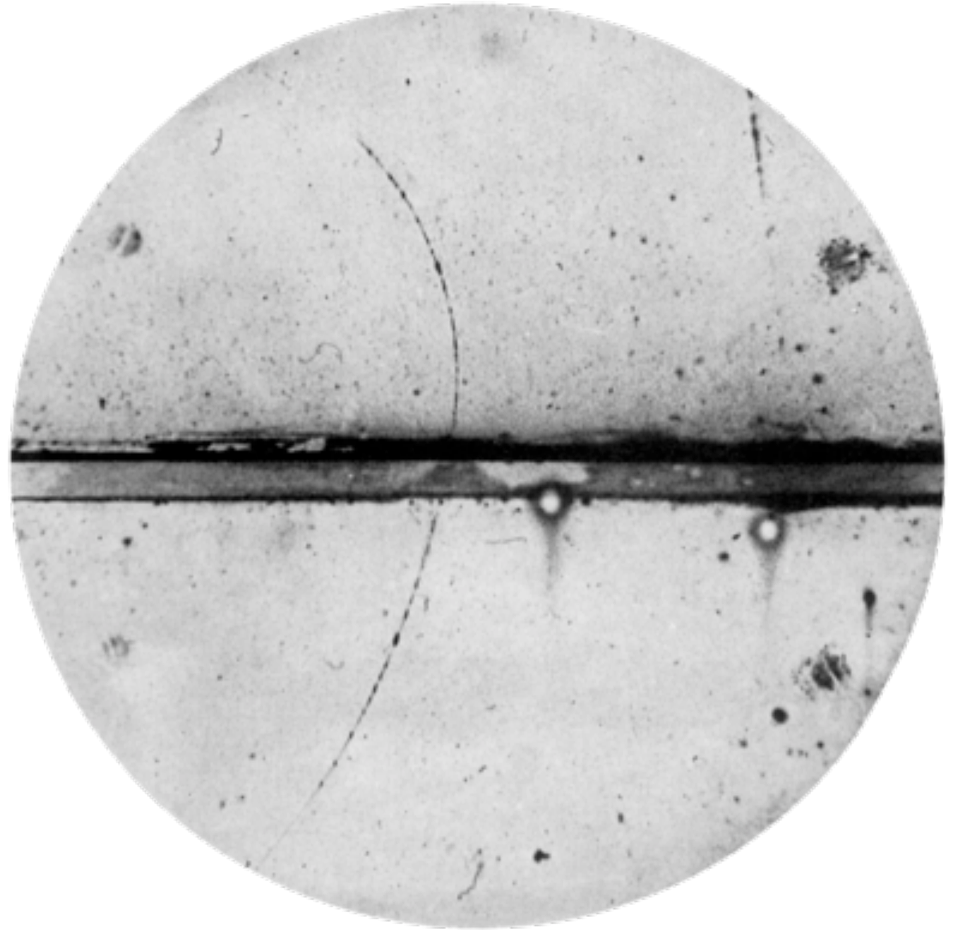


Single event

Multiple events:
Cut-and-count

Shape information:
Histogram

Positron discovery (1930s)



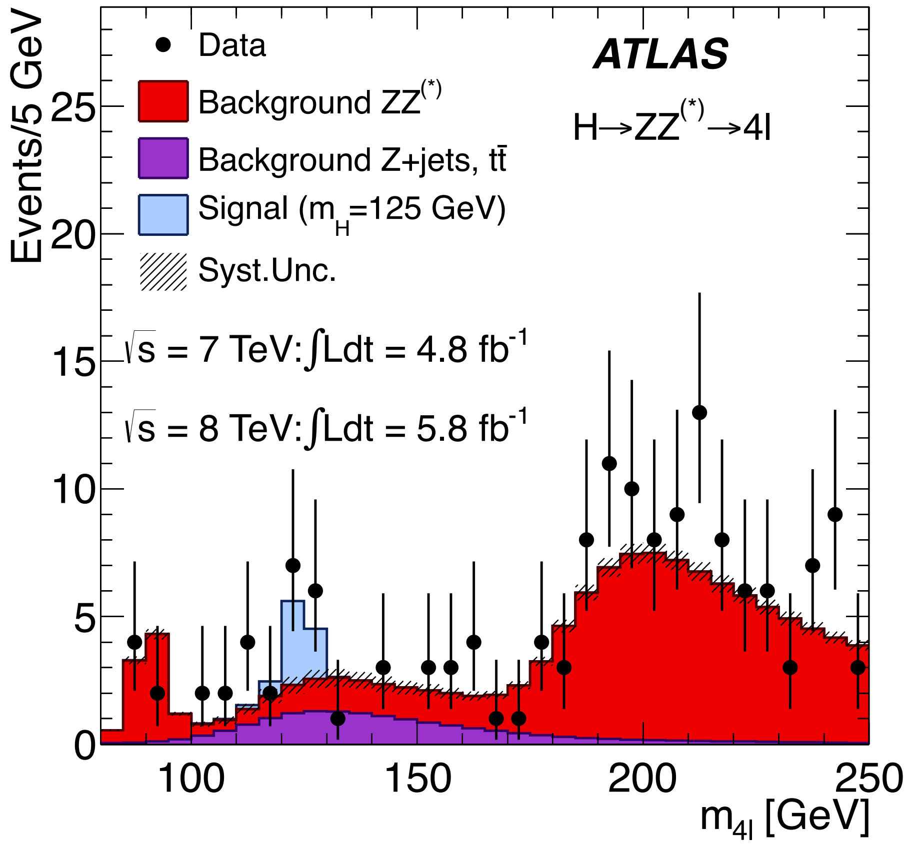
Single event

Top quark discovery (1990s)

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

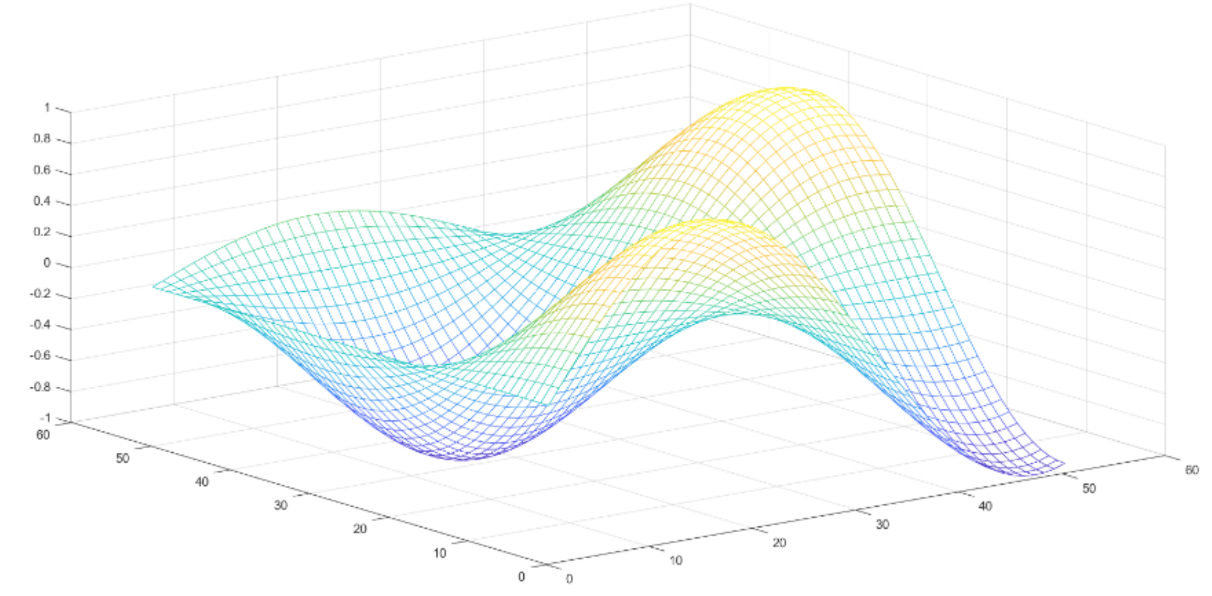
Multiple events:
Cut-and-count

Higgs boson
discovery (2010s)



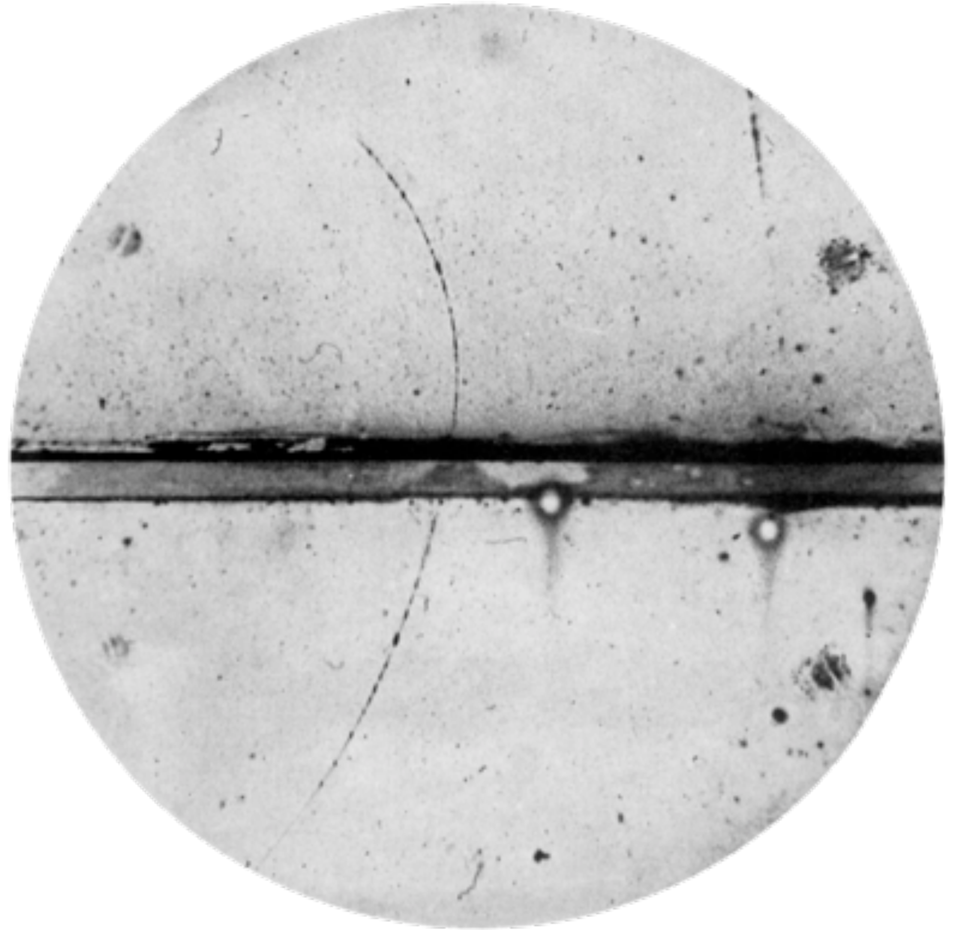
Shape information:
Histogram

Future discovery
(2020s ?)



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

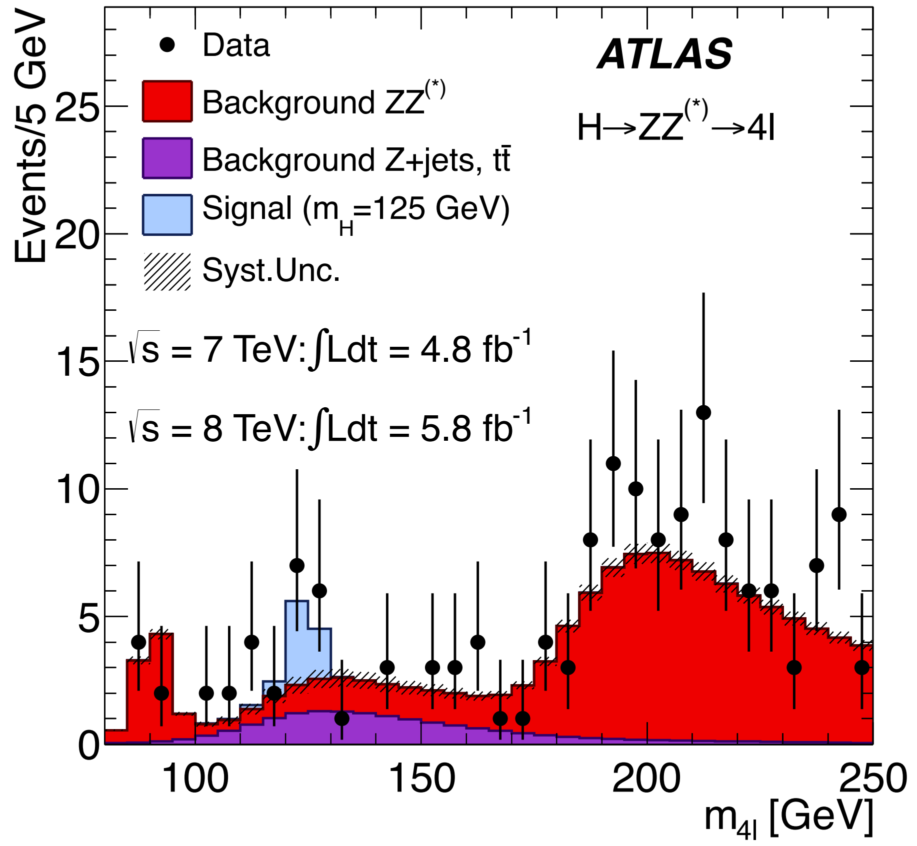
Positron discovery (1930s)



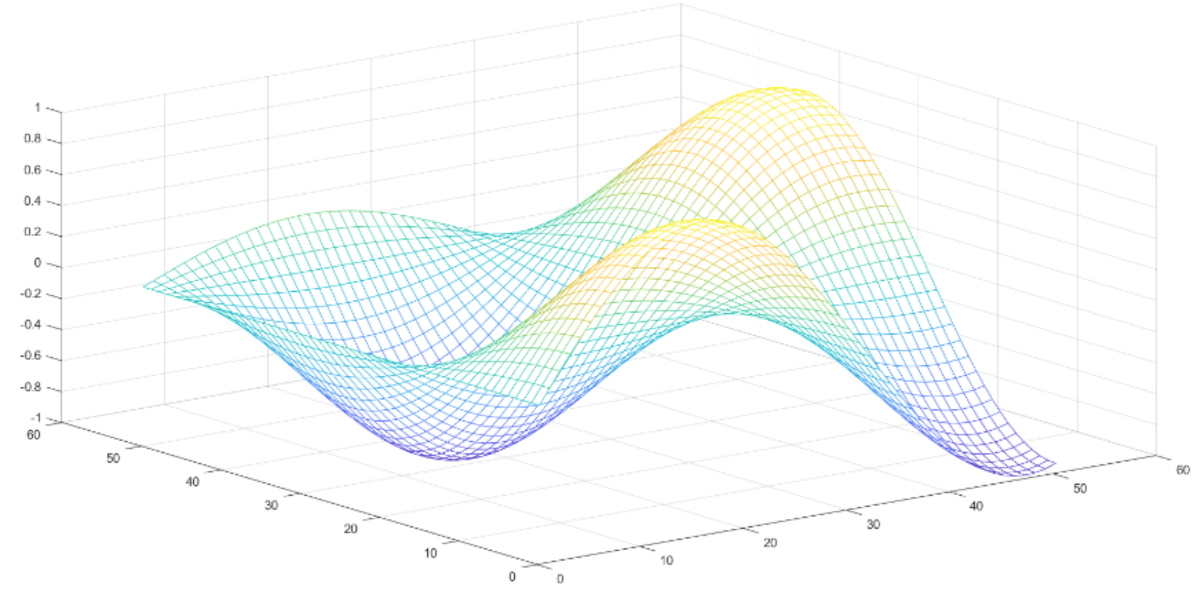
Top quark discovery (1990s)

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

Higgs boson discovery (2010s)



Future discovery (2020s ?)



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

Thank you !

Reference Sample

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

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

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

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

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

Dealing with negative weighted events

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

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

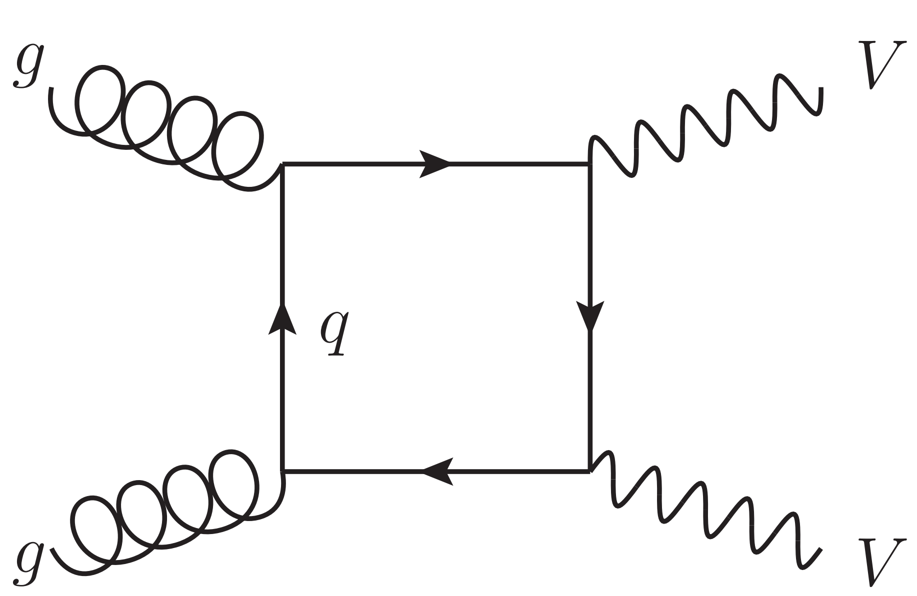
$$w_i^{\text{rwt-ref}} \to w_i^{\text{Asimov}}(\mu, \alpha) = \frac{\nu(\mu, \alpha)}{\nu_{\text{rwt-ref}}} \cdot \frac{p(x_i|\mu, \alpha)}{p_{\text{rwt-ref}}(x_i)} \cdot w_i^{\text{rwt-ref}}$$

Non-linear problem

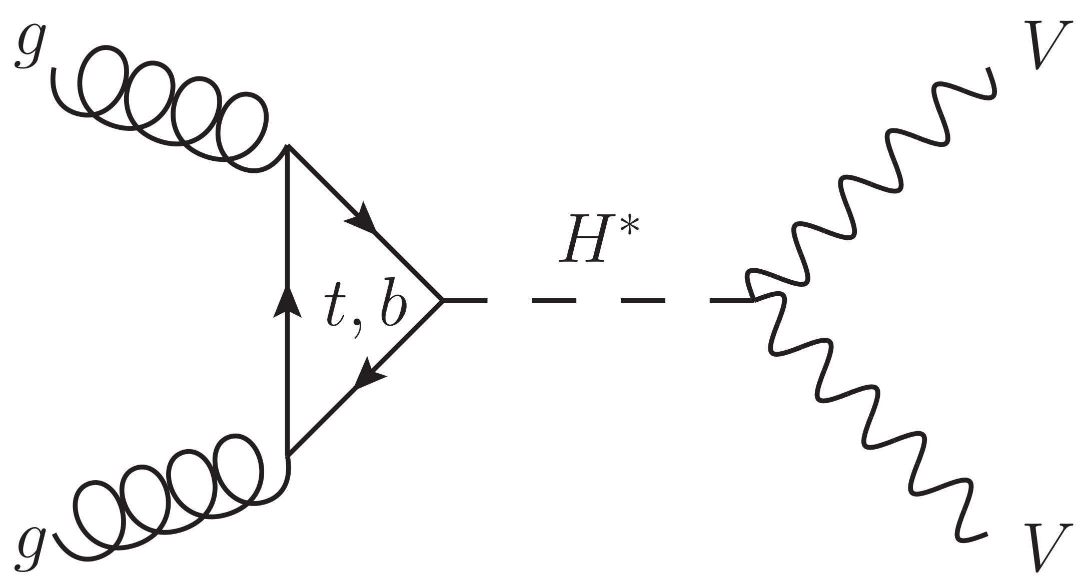
hal-02971995v3: Ghosh et al.

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

This term is negative



gg Background



ggF Signal

Scale by signal strength μ:

$$|M_s(X)|^2 \to |\sqrt{\mu} \cdot M_s(X)|^2,$$

$$P_{\text{scaled}}(X) = \mu \cdot P_s(X) + P_b(X) + \sqrt{\mu} \cdot P_i(X).$$

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

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

Hessian:

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

Pulls:

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

Post-fit Impact:

$$\Gamma_k = \frac{\partial \hat{\mu}}{\partial \alpha_k} \times \sqrt{C_{kk}}$$

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

Vertical interpolation

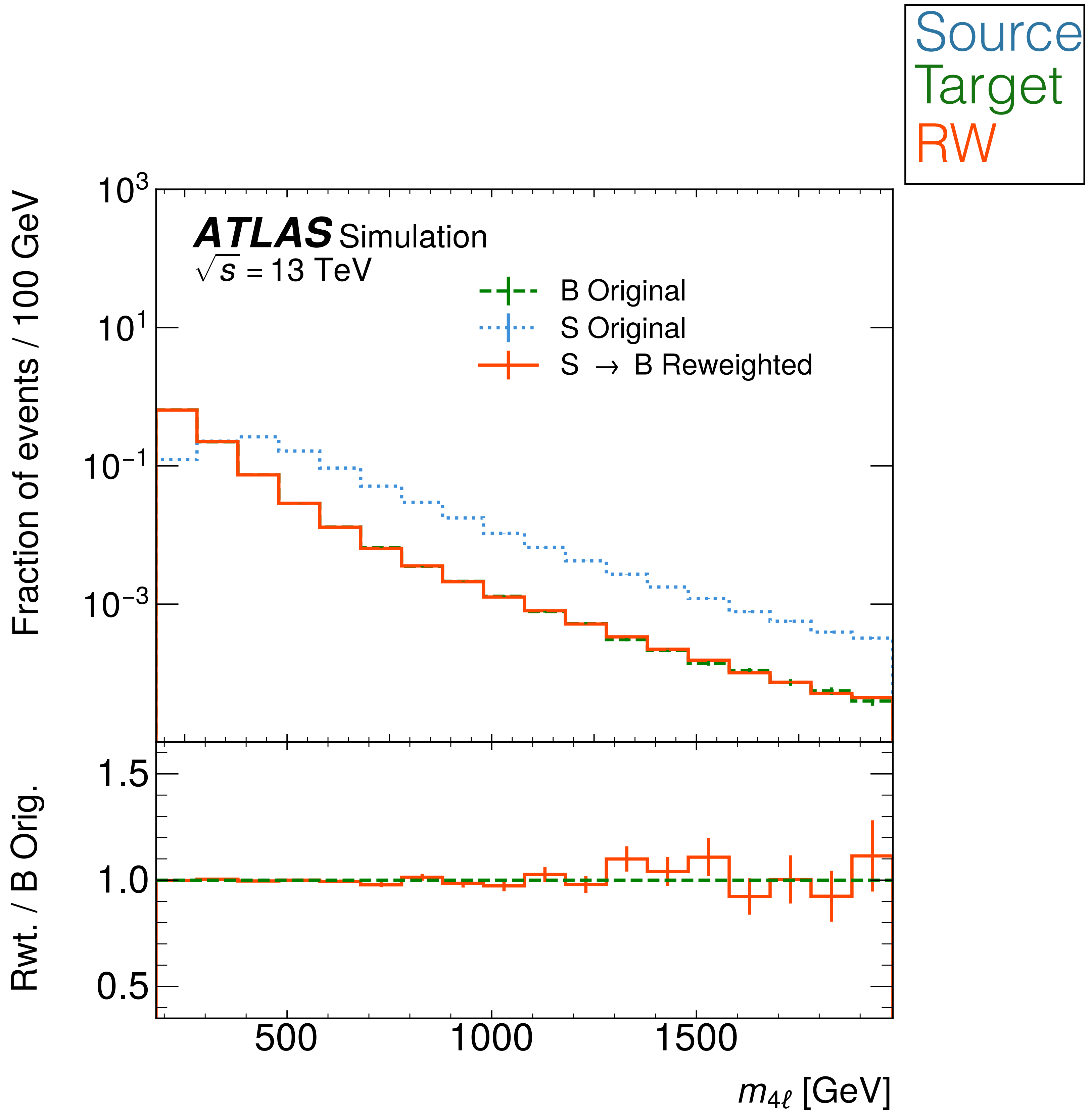
$$G_j(\alpha_k) = \begin{cases} \left(\frac{\nu_j(\alpha_k^+)}{\nu_j(\alpha_k^0)}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1 \\ \left(\frac{\nu_j(\alpha_k^-)}{\nu_j(\alpha_k^0)}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

$$g_j(x_i, \alpha_k) = \begin{cases} \left(g_j(x_i, \alpha_k^+)\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1 \\ \left(g_j(x_i, \alpha_k^-)\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

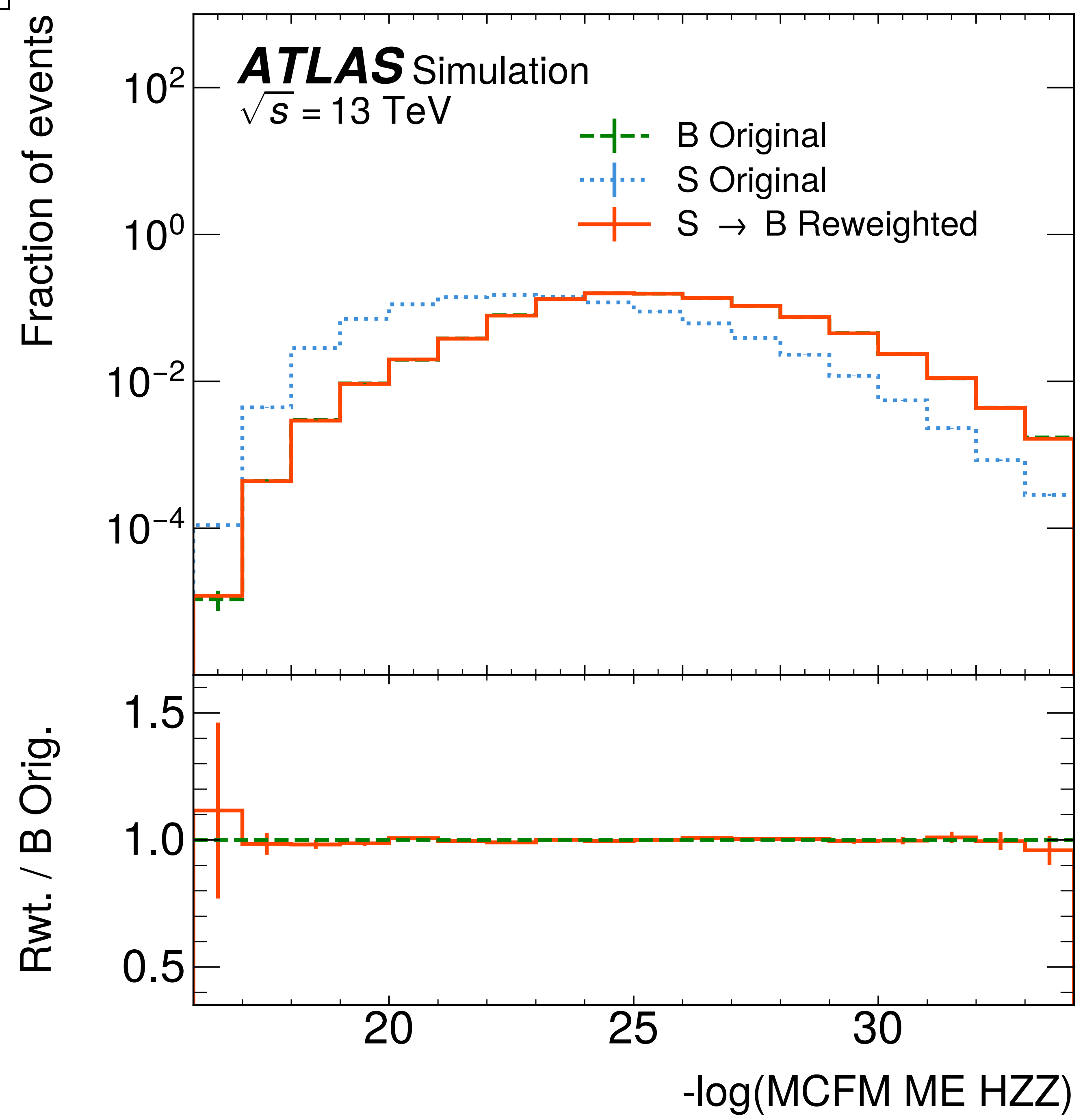
With some continuity requirements

More diagnostics

Re-weight closures for B



m_{4l}

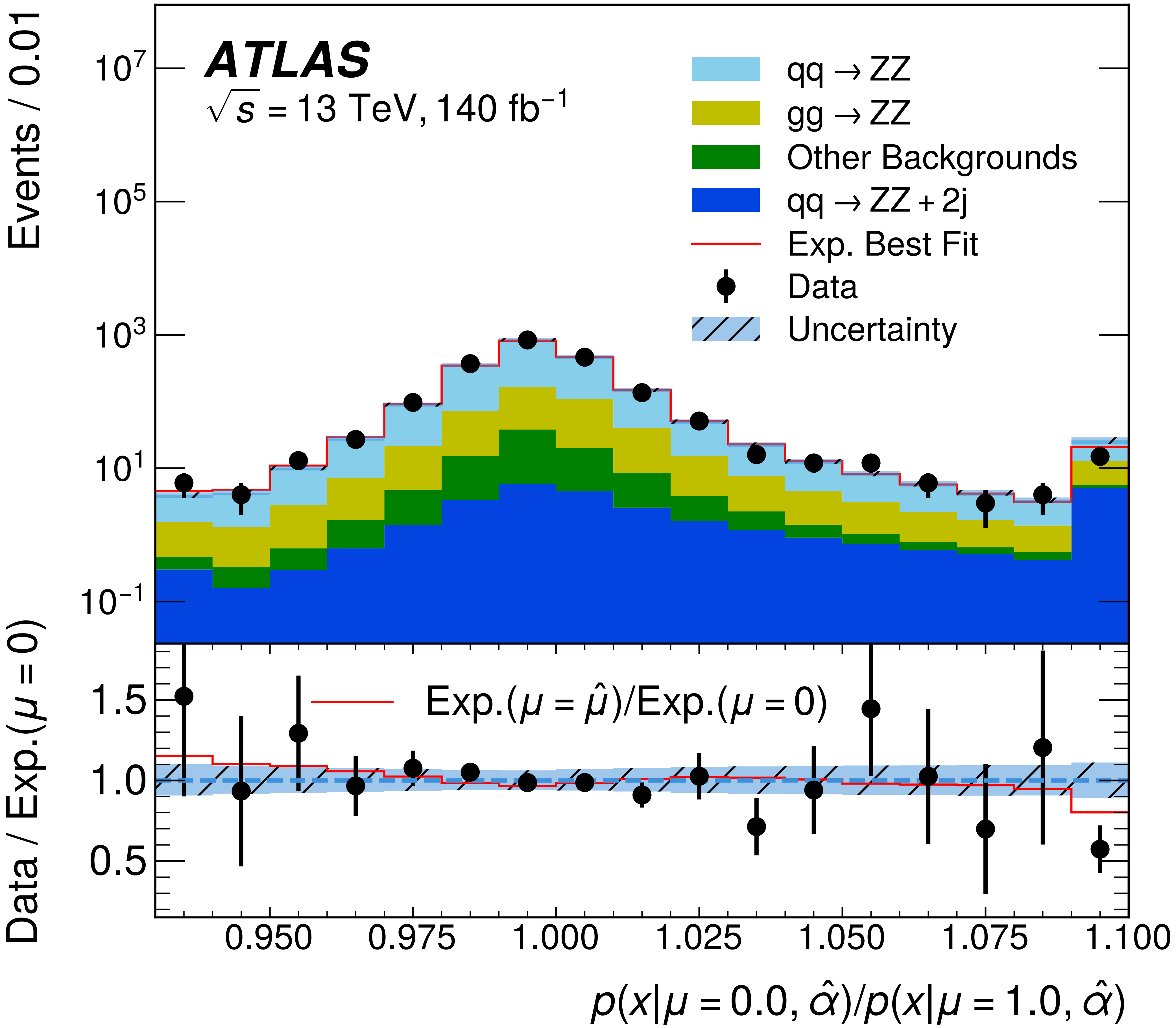


Matrix-Element-based Observable
(ggF from MCFM)

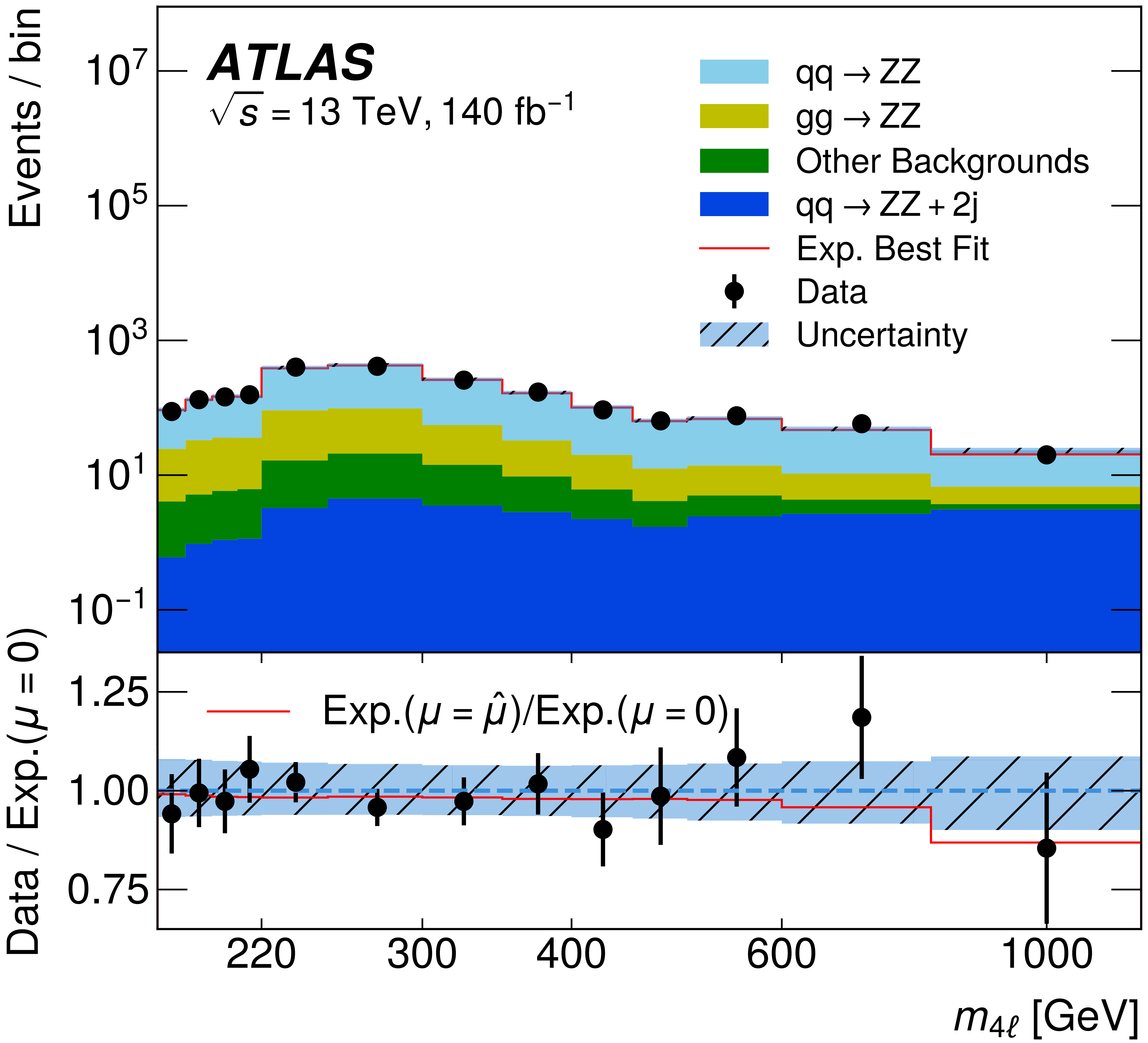
ATLAS Higgs width analysis paper: [arXiv:2412.01548](https://arxiv.org/abs/2412.01548)

Data-MC validation

NN observable

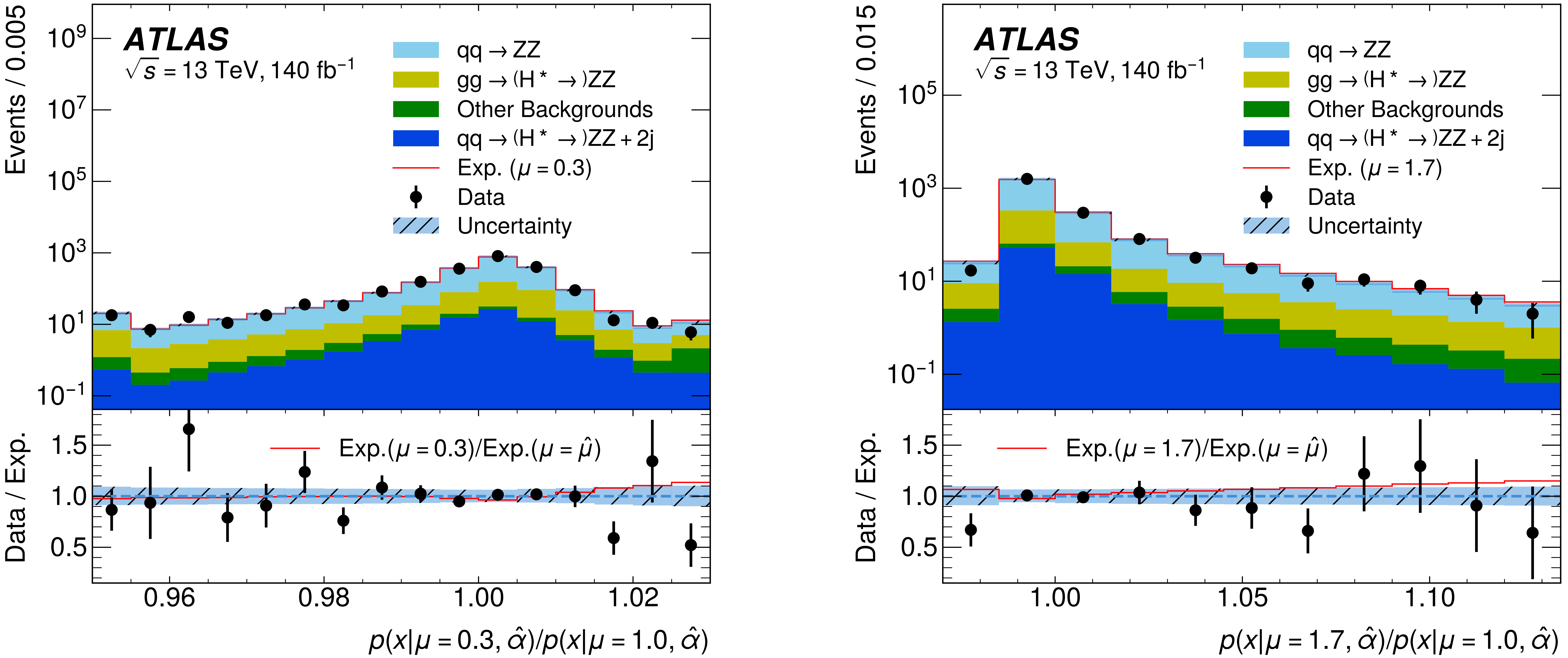


m_{4l}



Data-MC validation

Different NN observables



Physics analysis results

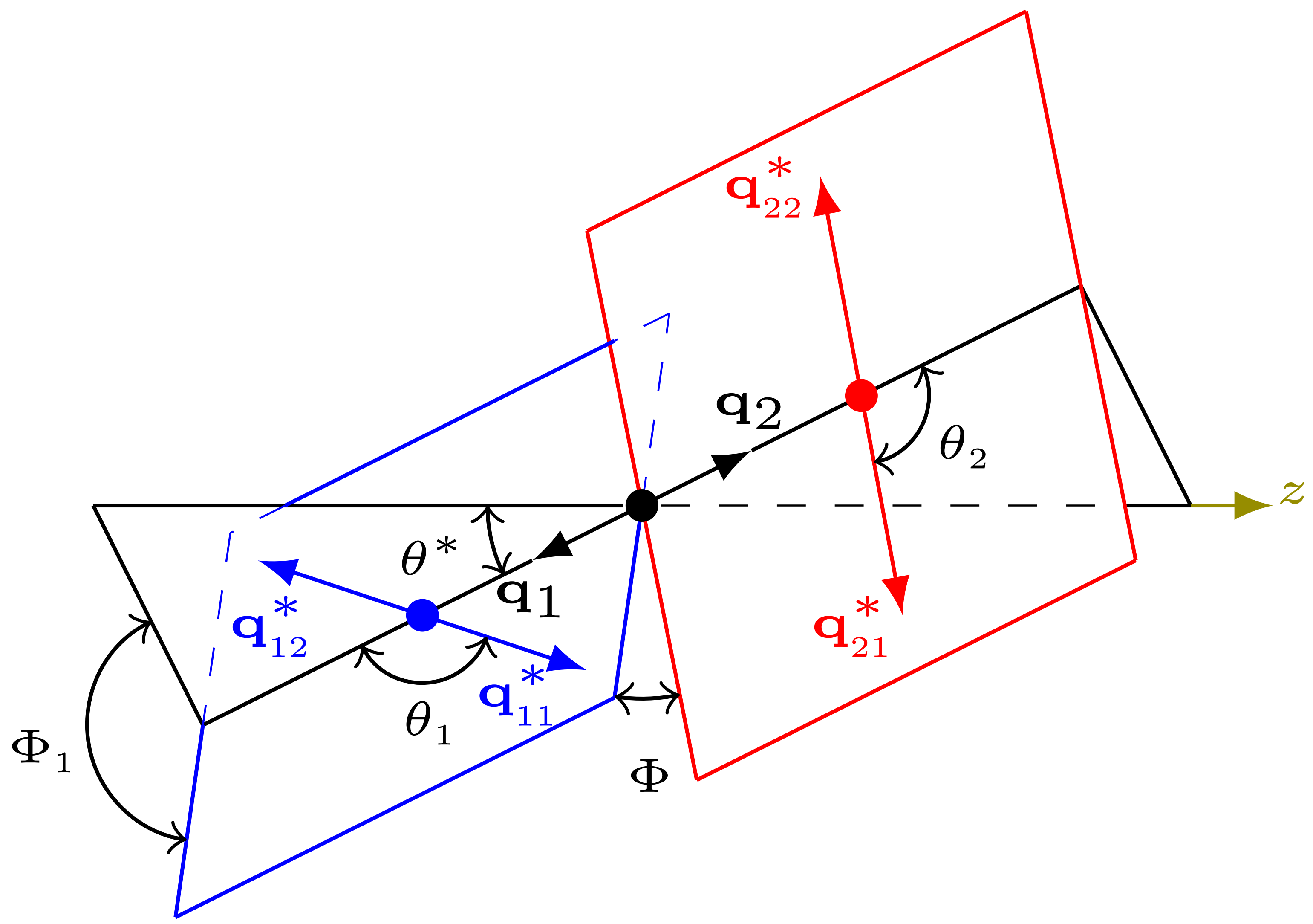
Impact of nuisance parameters

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 $gg \to ZZ$ (normalization)	2.07	2.26
Parton shower uncertainty for $gg \to ZZ$ (shape)	2.12	2.29
NLO EW uncertainty for $q\bar{q} \to ZZ$	2.10	2.27
NLO QCD uncertainty for $gg \to ZZ$	2.09	2.29
Parton shower uncertainty for $q\bar{q} \to ZZ$ (shape)	2.12	2.29
Jet energy scale and resolution uncertainty	2.11	2.26
None (full result)	2.12	2.30

Full probability model, input variables

$$p(x|\mu^{\text{ggF}}_{\text{off-shell}}, \mu^{\text{EW}}_{\text{off-shell}}) = \frac{1}{\nu(\mu^{\text{ggF}}_{\text{off-shell}}, \mu^{\text{EW}}_{\text{off-shell}})} \times \Big[\mu^{\text{ggF}}_{\text{off-shell}} \nu^{\text{ggF}}_{\text{S}} p^{\text{ggF}}_{\text{S}}(x) + \sqrt{\mu^{\text{ggF}}_{\text{off-shell}}} \nu^{\text{ggF}}_{\text{I}} p^{\text{ggF}}_{\text{I}}(x) + \nu^{\text{ggF}}_{\text{B}} p^{\text{ggF}}_{\text{B}}(x) + \mu^{\text{EW}}_{\text{off-shell}} \nu^{\text{EW}}_{\text{S}} p^{\text{EW}}_{\text{S}}(x) + \sqrt{\mu^{\text{EW}}_{\text{off-shell}}} \nu^{\text{EW}}_{\text{I}} p^{\text{EW}}_{\text{I}}(x) + \nu^{\text{EW}}_{\text{B}} p^{\text{EW}}_{\text{B}}(x) + \nu_{\text{NI}} p_{\text{NI}}(x)\Big],$$

Variable	Definition
$m_{4\ell}$	quadruplet mass
m_{Z1}	Z_1 mass
m_{Z2}	Z_2 mass
$\cos\theta^*$	cosine of the Higgs boson decay angle $[\mathbf{q}_1 \cdot \mathbf{n}_z/|\mathbf{q}_1|]$
$\cos\theta_1$	cosine of the Z_1 decay angle $[-(\mathbf{q}_2)\cdot\mathbf{q}_{11}/(|\mathbf{q}_2|\cdot|\mathbf{q}_{11}|)]$
$\cos\theta_2$	cosine of the Z_2 decay angle $[-(\mathbf{q}_1)\cdot\mathbf{q}_{21}/(|\mathbf{q}_1|\cdot|\mathbf{q}_{21}|)]$
Φ_1	Z_1 decay plane angle $[\cos^{-1}(\mathbf{n}_1\cdot\mathbf{n}_{\text{sc}})\,(\mathbf{q}_1\cdot(\mathbf{n}_1\times\mathbf{n}_{\text{sc}})/(|\mathbf{q}_1|\cdot|\mathbf{n}_1\times\mathbf{n}_{\text{sc}}|)]$
Φ	angle between Z_1, Z_2 decay planes $[\cos^{-1}(\mathbf{n}_1\cdot\mathbf{n}_2)\,(\mathbf{q}_1\cdot(\mathbf{n}_1\times\mathbf{n}_2)/(|\mathbf{q}_1|\cdot|\mathbf{n}_1\times\mathbf{n}_2|)]$
$p_T^{4\ell}$	quadruplet transverse momentum
$y^{4\ell}$	quadruplet rapidity
n_{jets}	number of jets in the event
m_{jj}	leading dijet system mass
$\Delta\eta_{jj}$	leading dijet system pseudorapidity
$\Delta\phi_{jj}$	leading dijet system azimuthal angle difference



Network architecture

Feed-forward dense networks

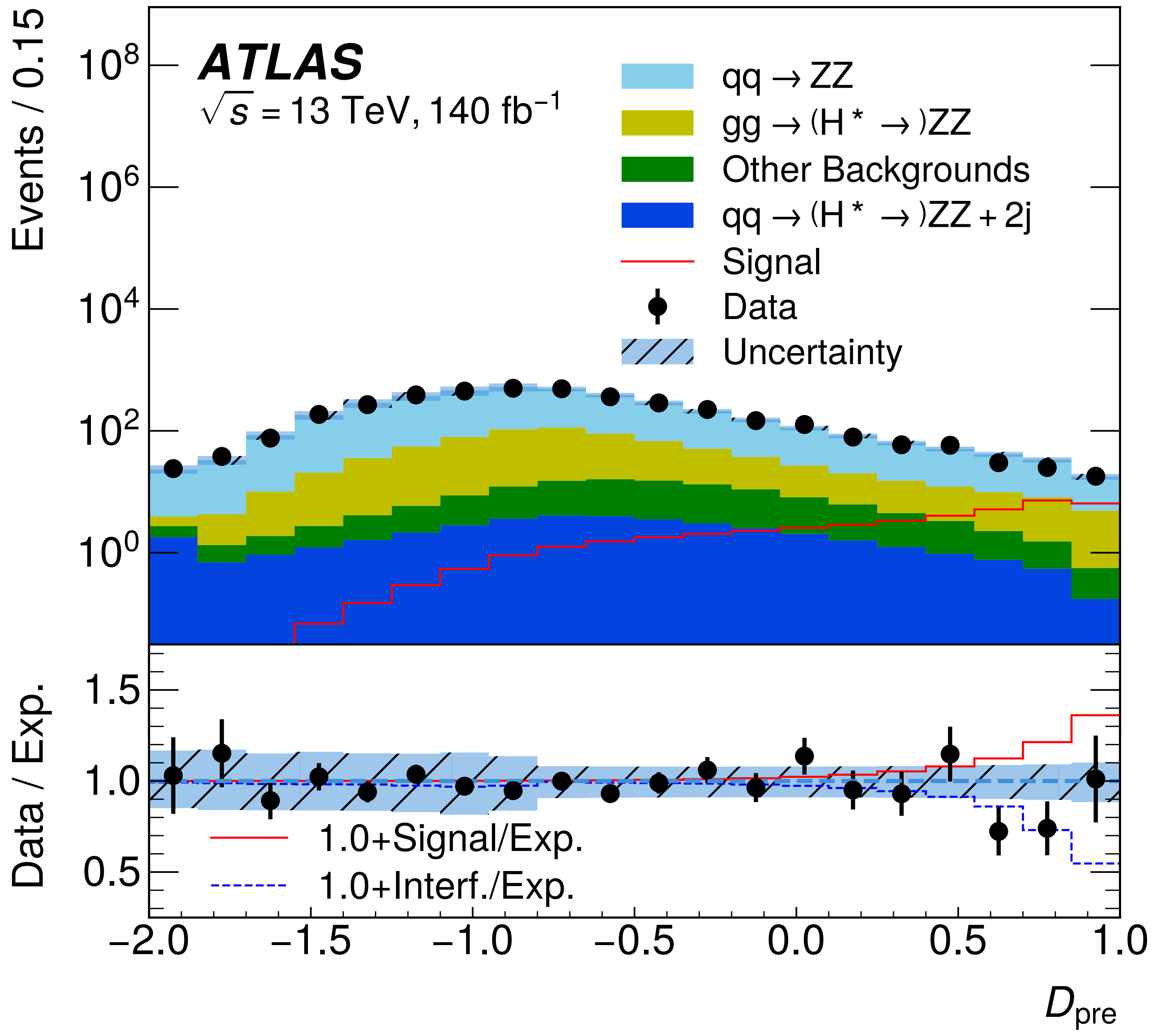
- 5 hidden layers with 1000 nodes
- Swish activation
- Single node output layer with sigmoid activation

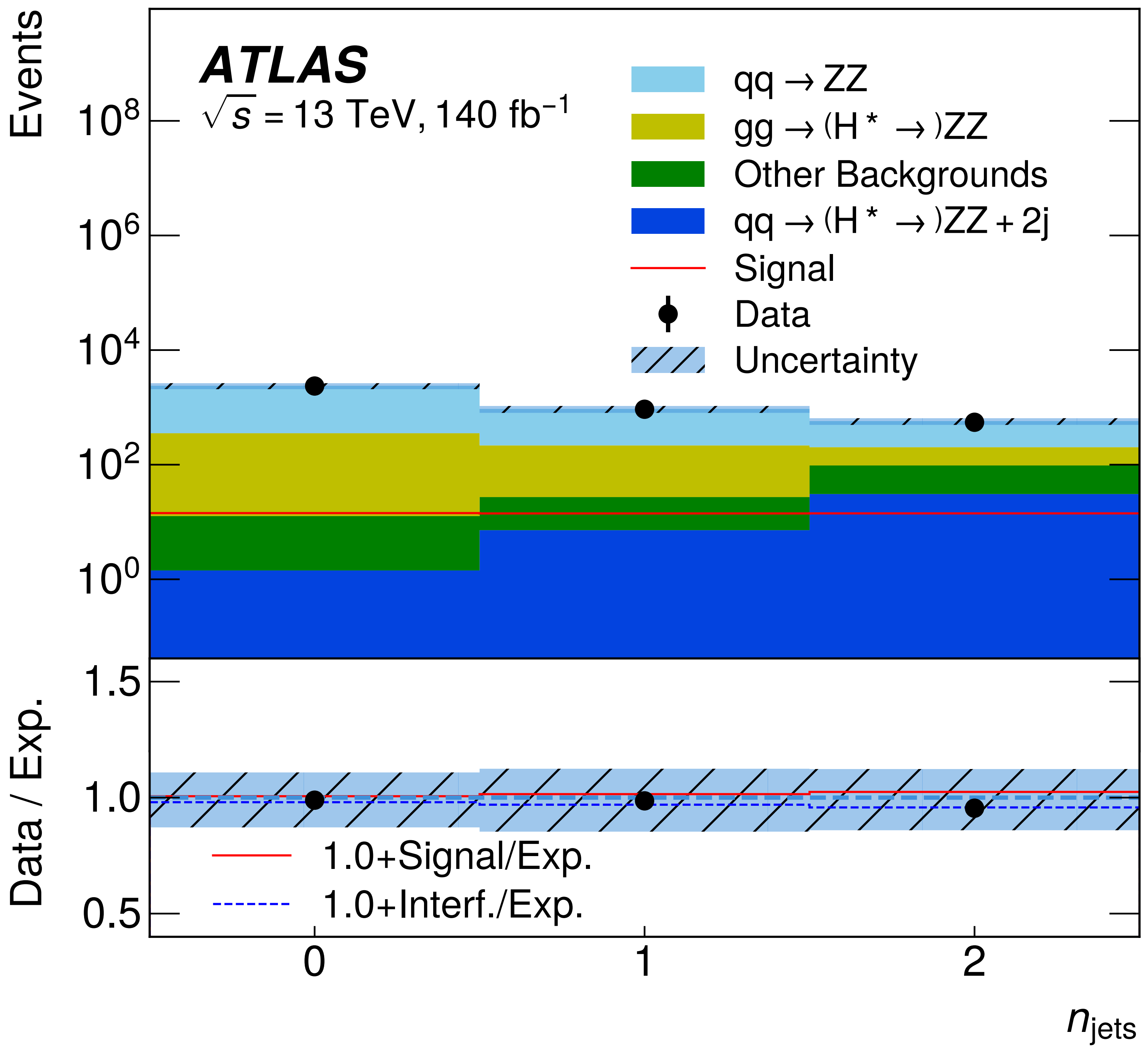
Loss: Weighted binary cross-entropy

$O(10^4)$ networks takes approx 4000 GPU hours to train

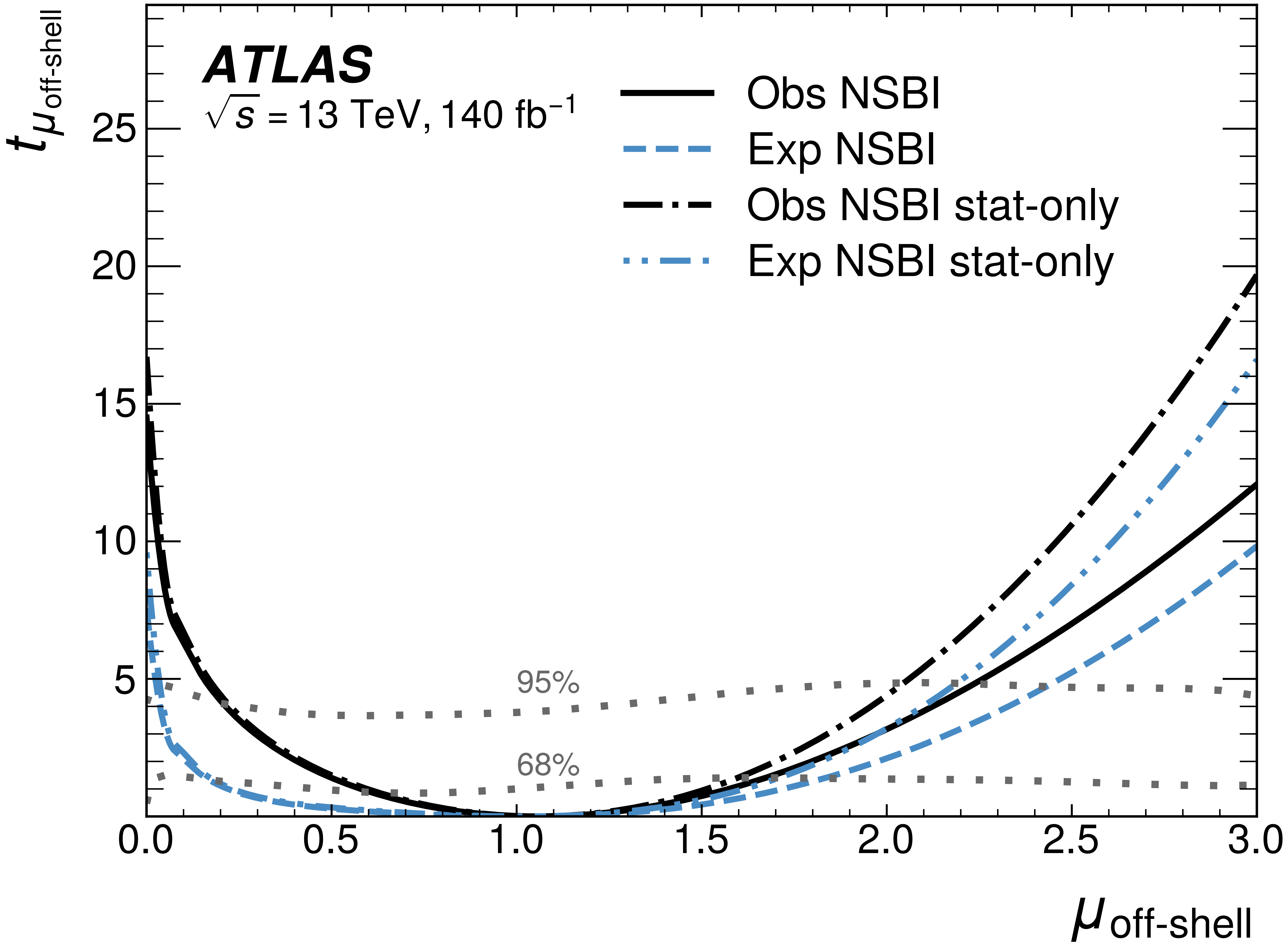
Pre-selection region definition

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



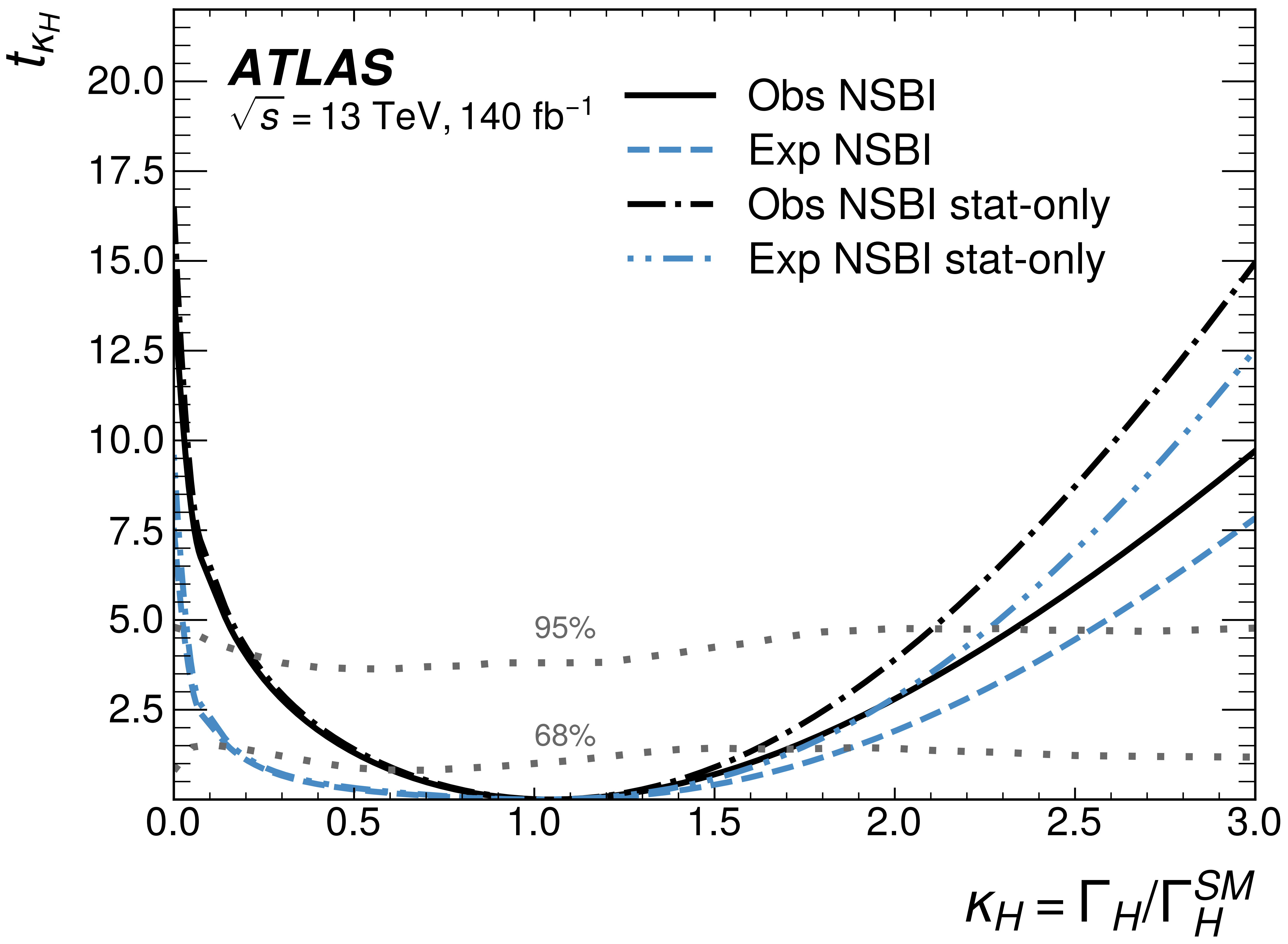


Result after combination with $ll\nu\nu$



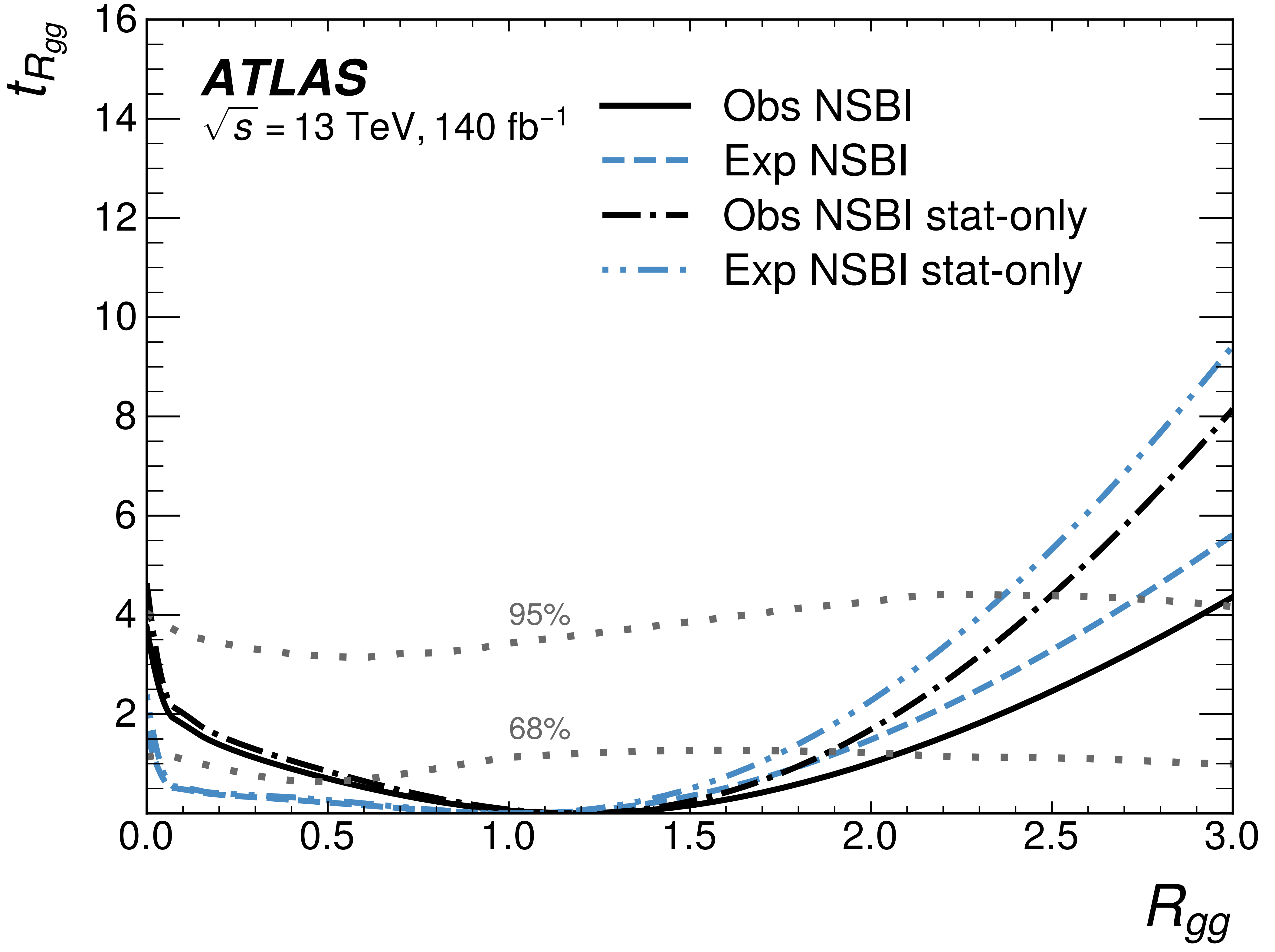
$ll\nu\nu$ dominates sensitivity

Width interpretation

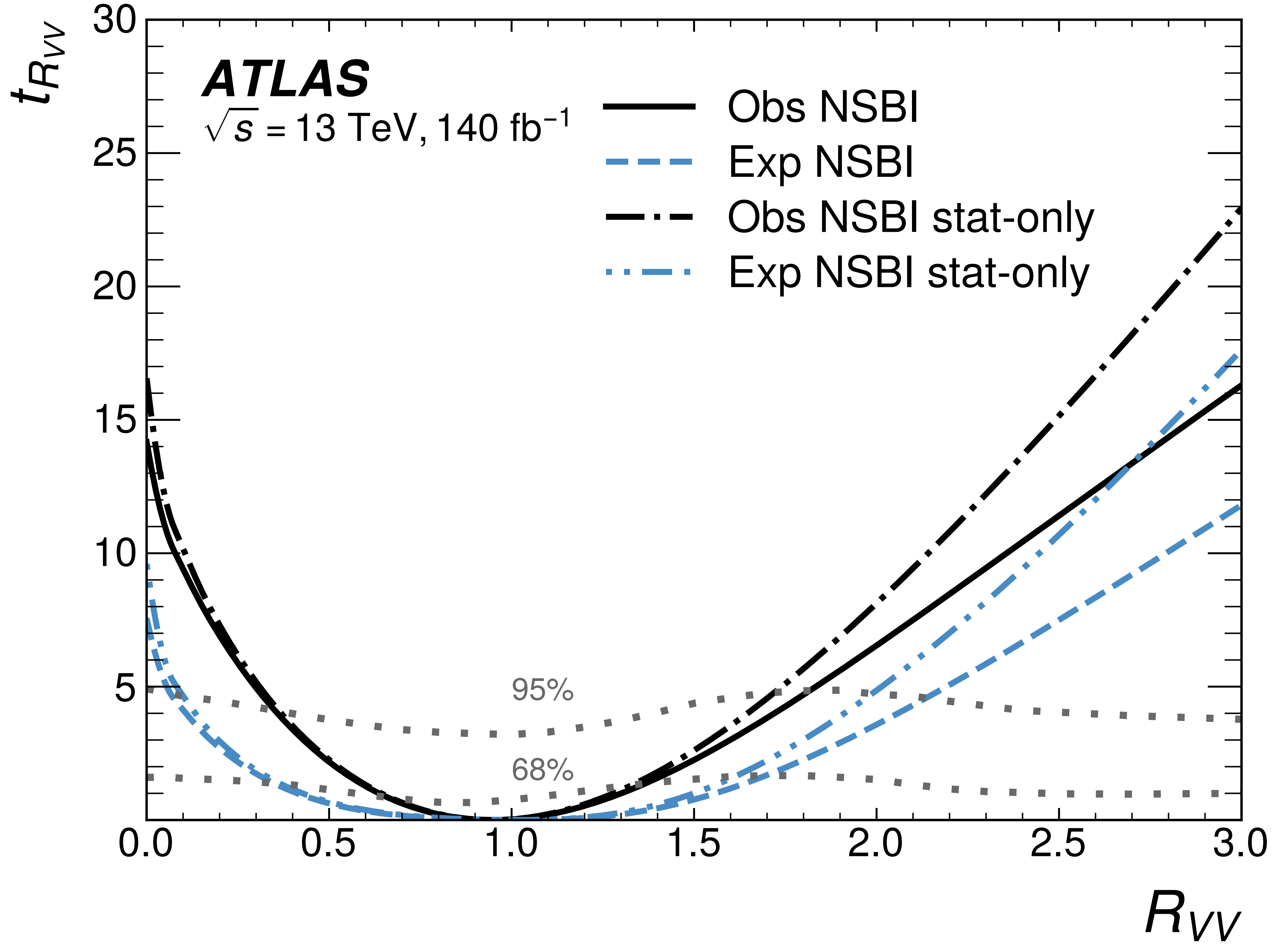


CI obtained from Neyman construction

Width sensitivity in ggF and VBF



$$R_{gg} = \kappa^2_{g,\text{on-shell}} / \kappa^2_{g,\text{off-shell}}$$



$$R_{VV} = \kappa^2_{V,\text{on-shell}} / \kappa^2_{V,\text{off-shell}}$$

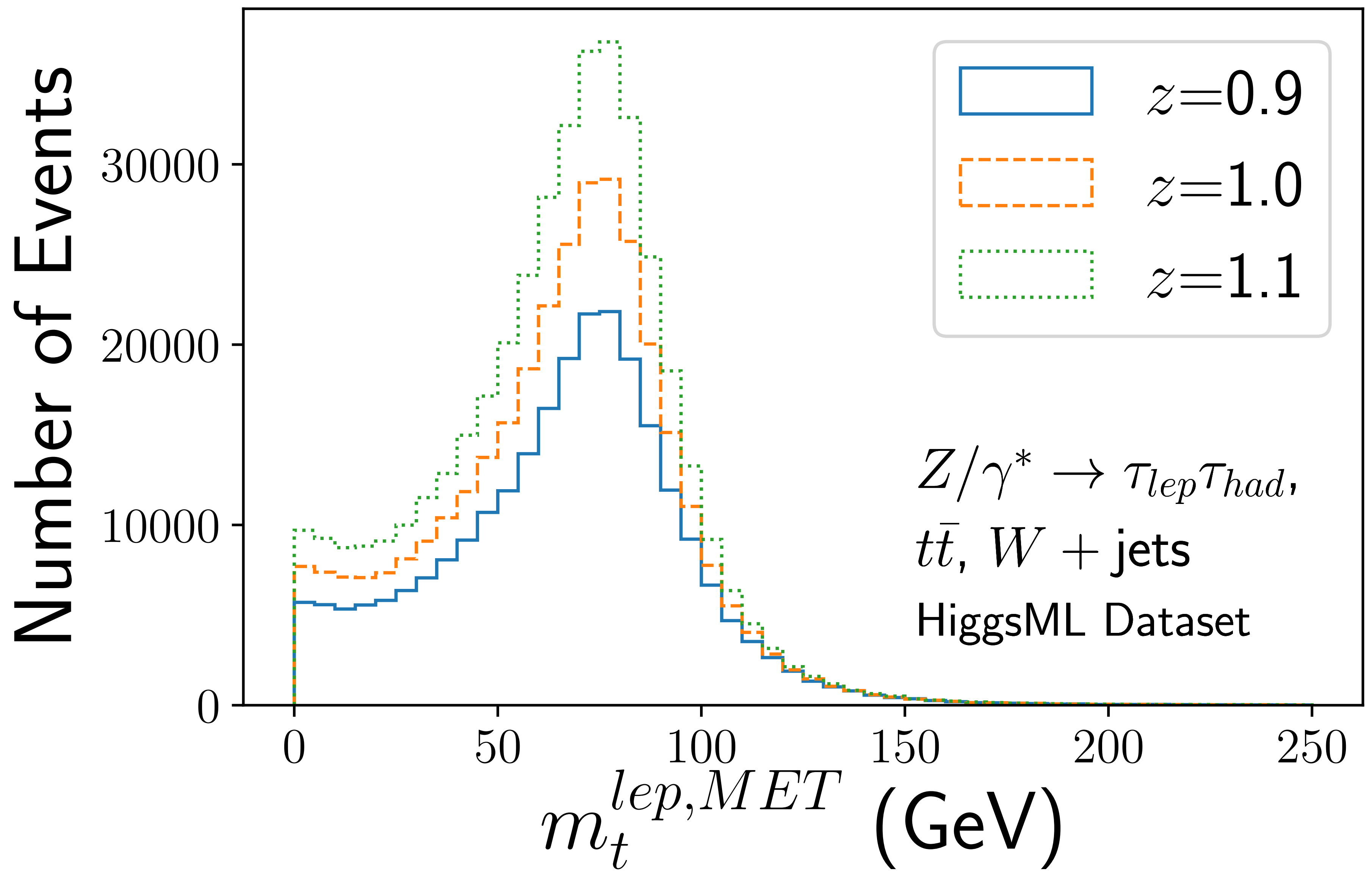
Comparison to previous result in same data

Parameter	Value	68% CL interval Observed	68% CL interval Expected	95% CL interval Observed	95% CL interval Expected
NSBI analysis					
$\mu_{\text{off-shell}}$ (4ℓ 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					
$\mu_{\text{off-shell}}$ (4ℓ 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]

Uncertainty-aware analysis optimisation

PRD.104.056026: Aishik Ghosh, Benjamin Nachman, and Daniel Whiteson

Experimental uncertainties:
Eg. Inaccuracies in the calibration of our detector

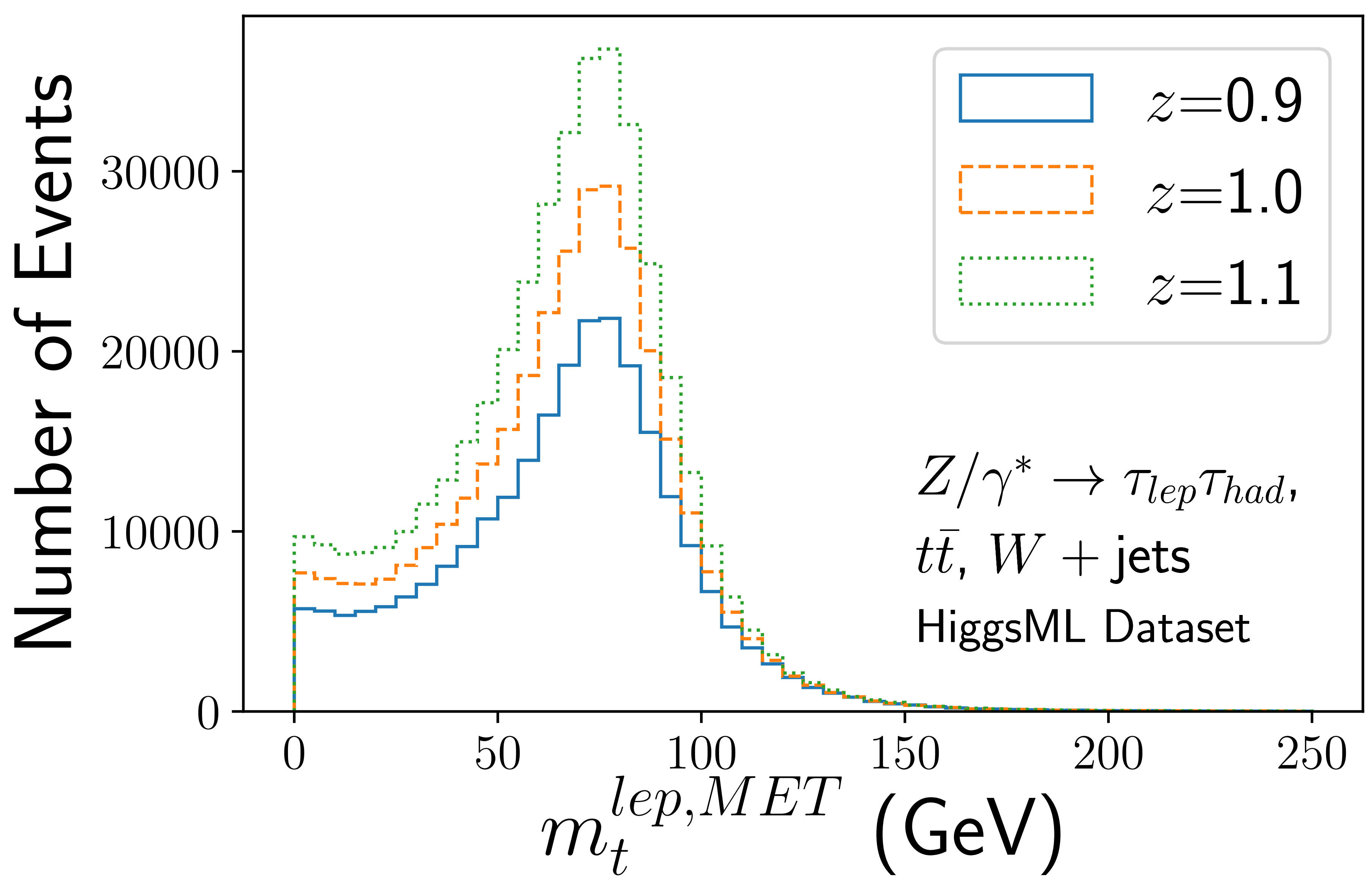


- Current analyses strategies optimised while ignoring systematic uncertainties
- Added in post-facto
- Leads to loss in sensitivity compared to uncertainty-aware optimisation (see details)

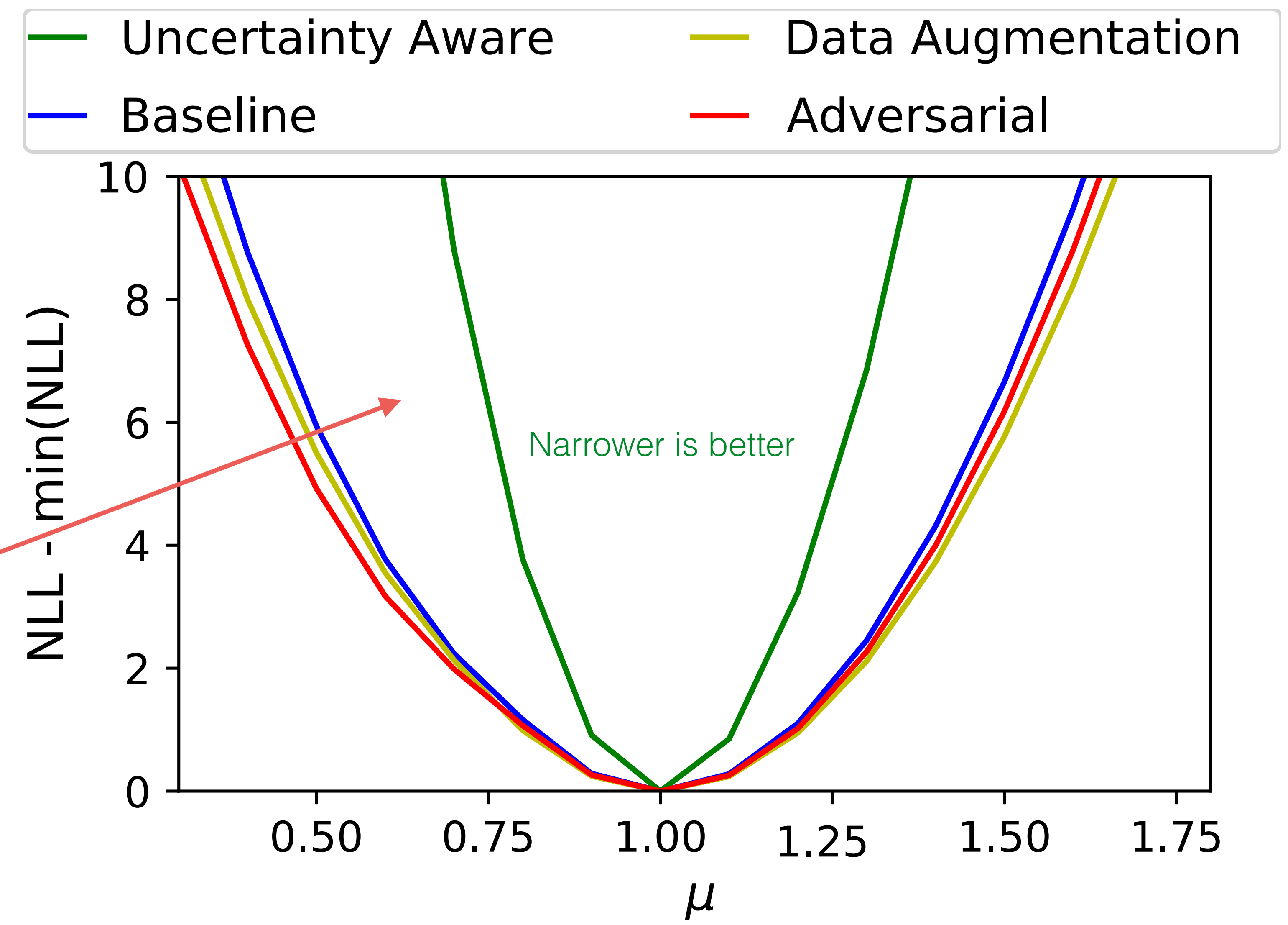
Uncertainty-aware analysis optimisation

[PRD.104.056026](PRD.104.056026): Aishik Ghosh, Benjamin Nachman, and Daniel Whiteson

Experimental uncertainties:
Eg. Inaccuracies in the calibration of our detector

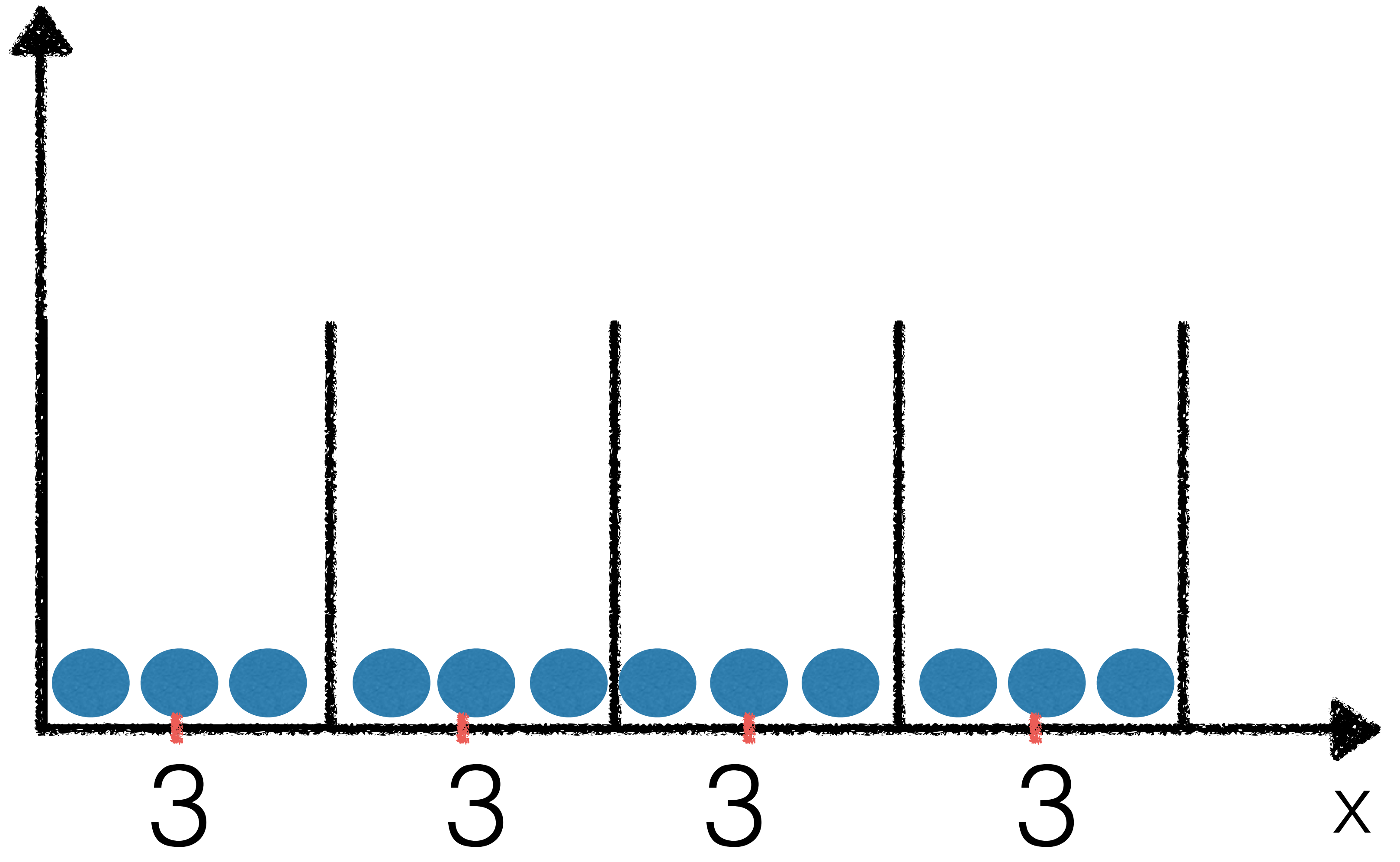


- Current analyses strategies optimised while ignoring systematic uncertainties
- Added in post-facto
- Leads to loss in sensitivity compared to uncertainty-aware optimisation (see details)

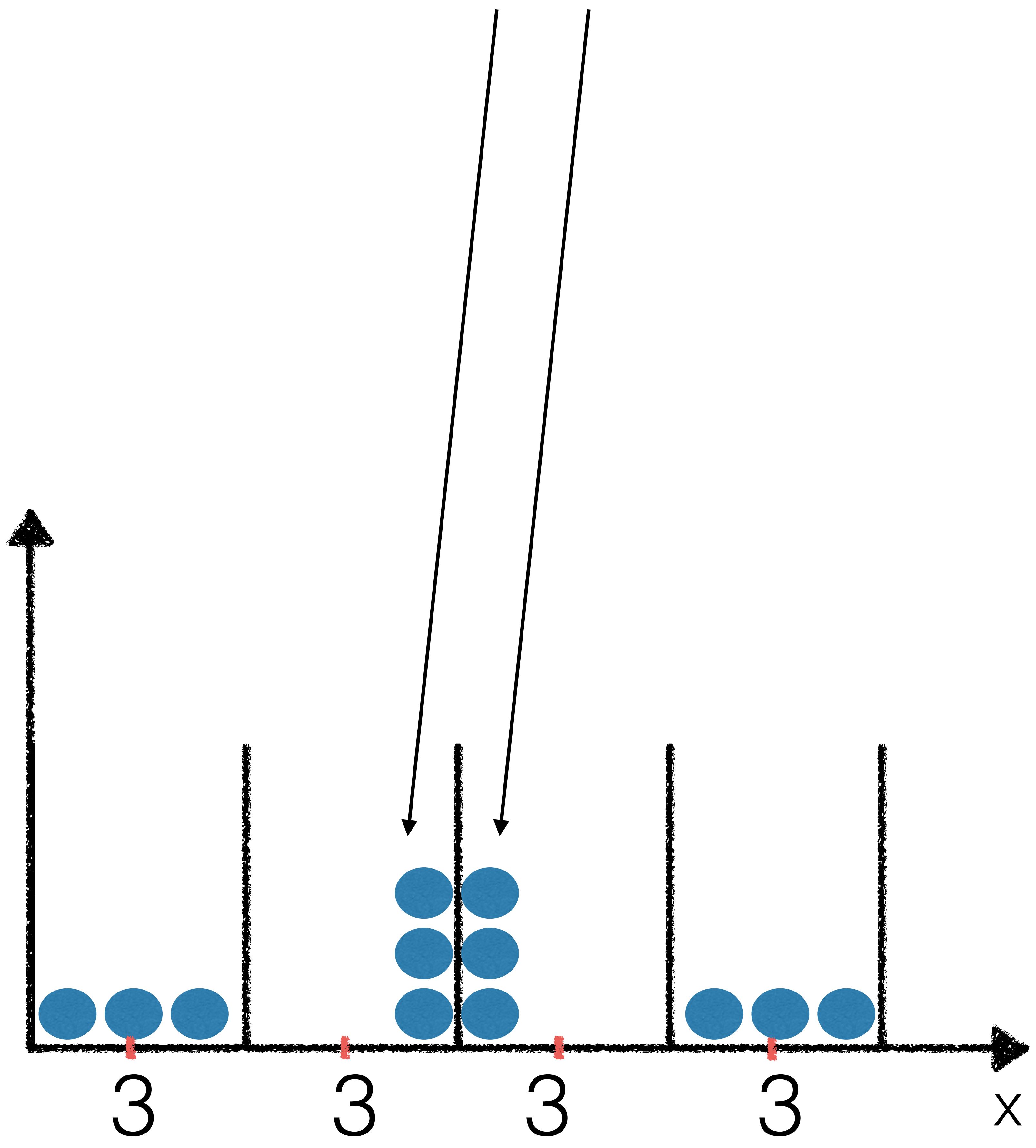


Difference b/w post-facto and uncertainty-aware

Avoids binning data into histograms, which is another lossy compression



Information on individual events lost!



Avoids binning data into histograms, which is another lossy compression

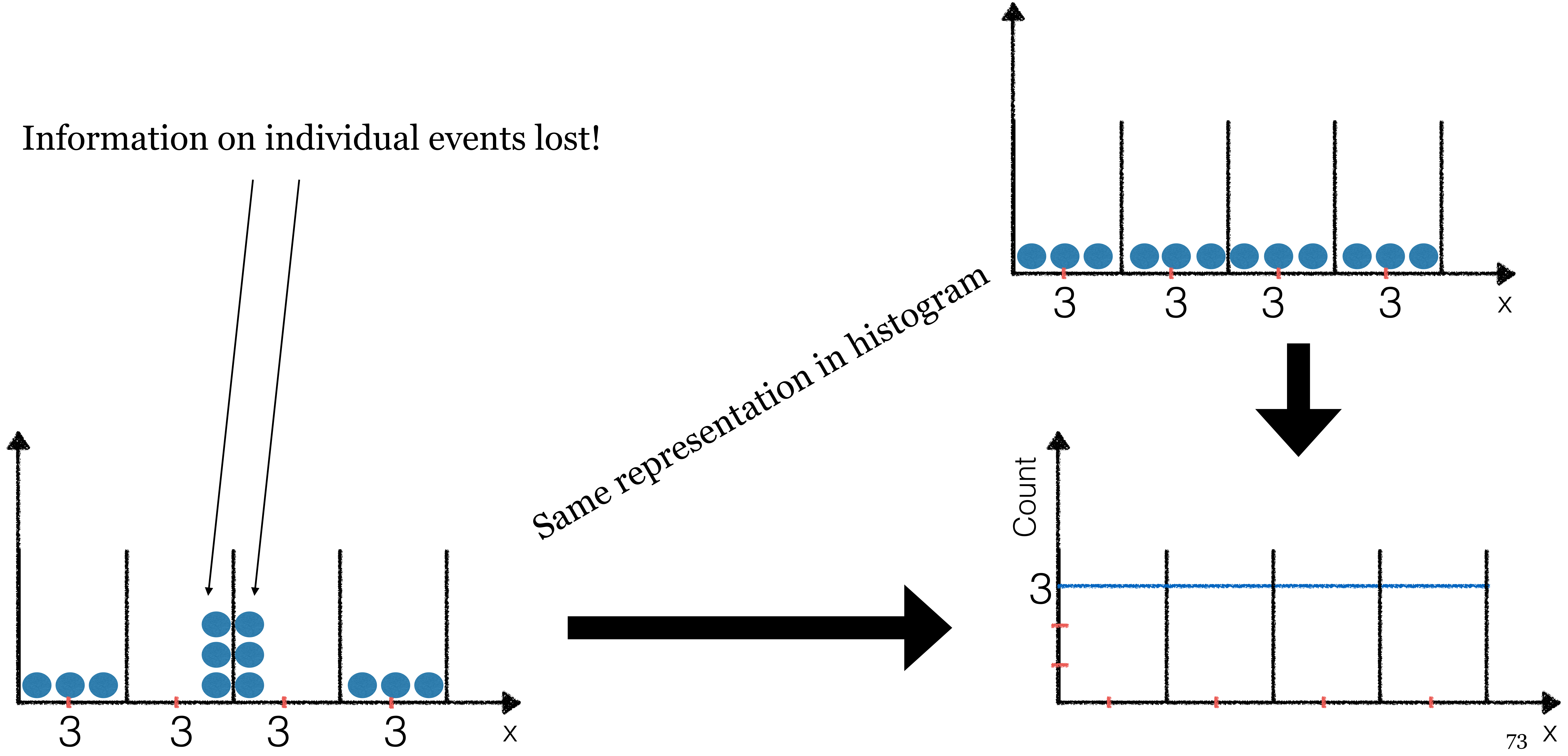
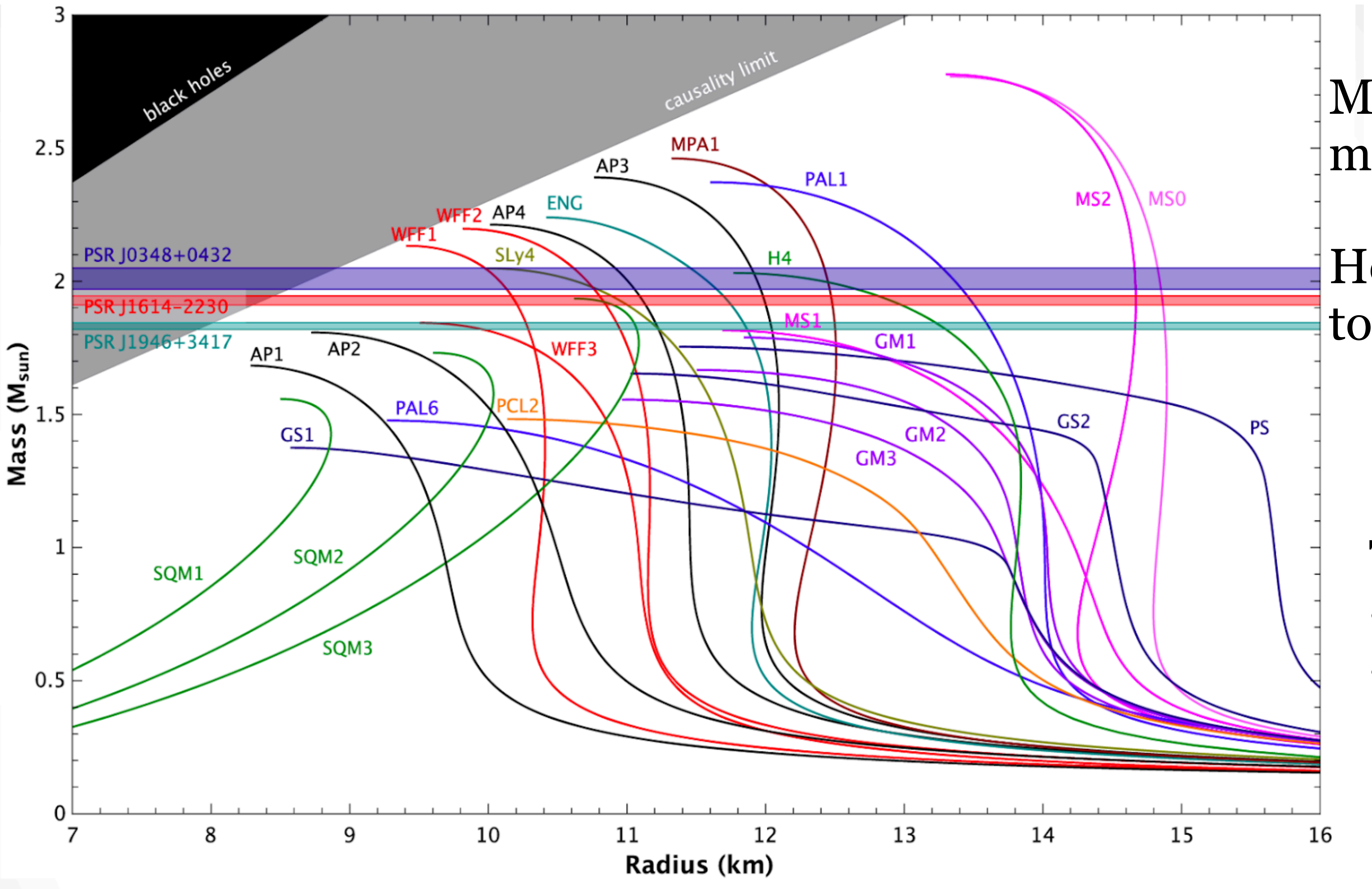


Image: NASA

Image: NASA

Telescope measurements of energy spectra of neutron stars



Mass-radius curves created by different equation of state (EoS) models

Horizontal bars show massive neutron star observations used to "rule out" EoS models.

Two communities:

- Astrophysicists measure mass/radius from telescope
- Nuclear theorists measure EoS from mass/radius

Figure from Lattimer J. M., Prakash M., 2001, The Astrophysical Journal, 550, 426–442

Telescope measurements of energy spectra

Probe the interior: Equation of State parameters
λ_1, λ_2

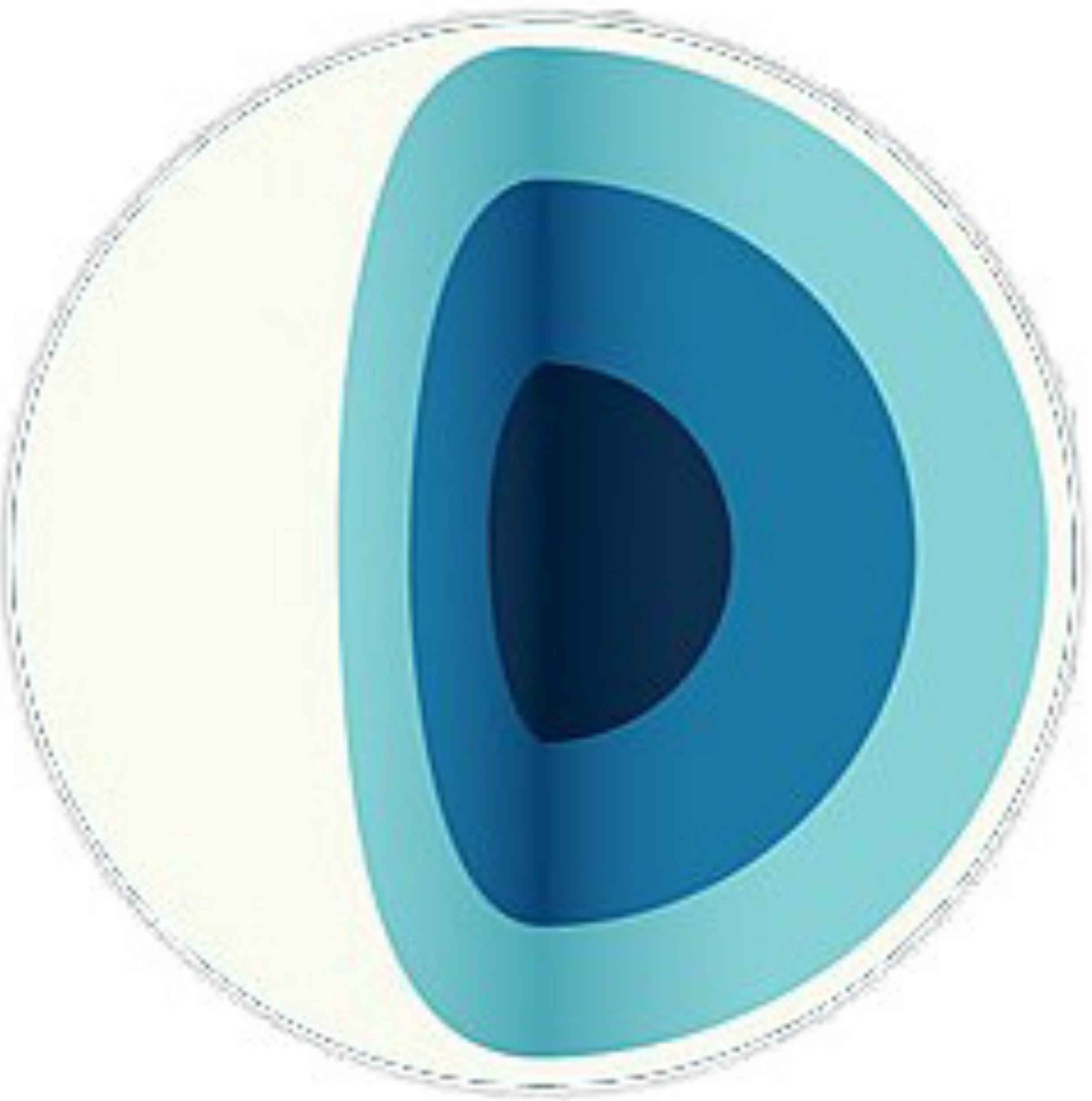
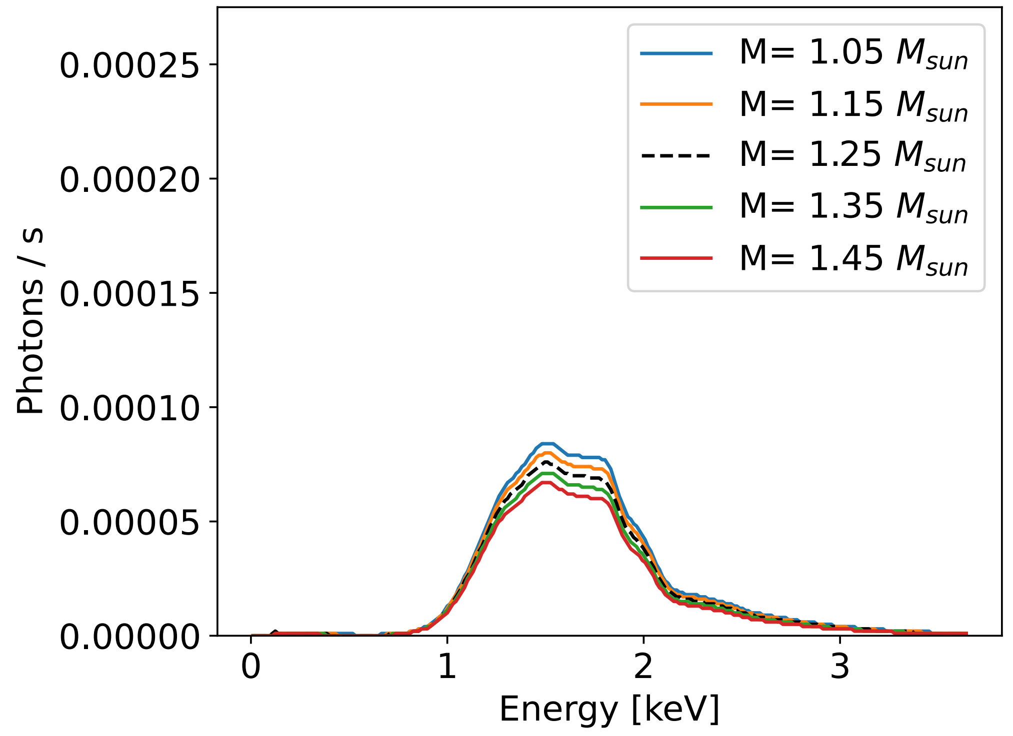
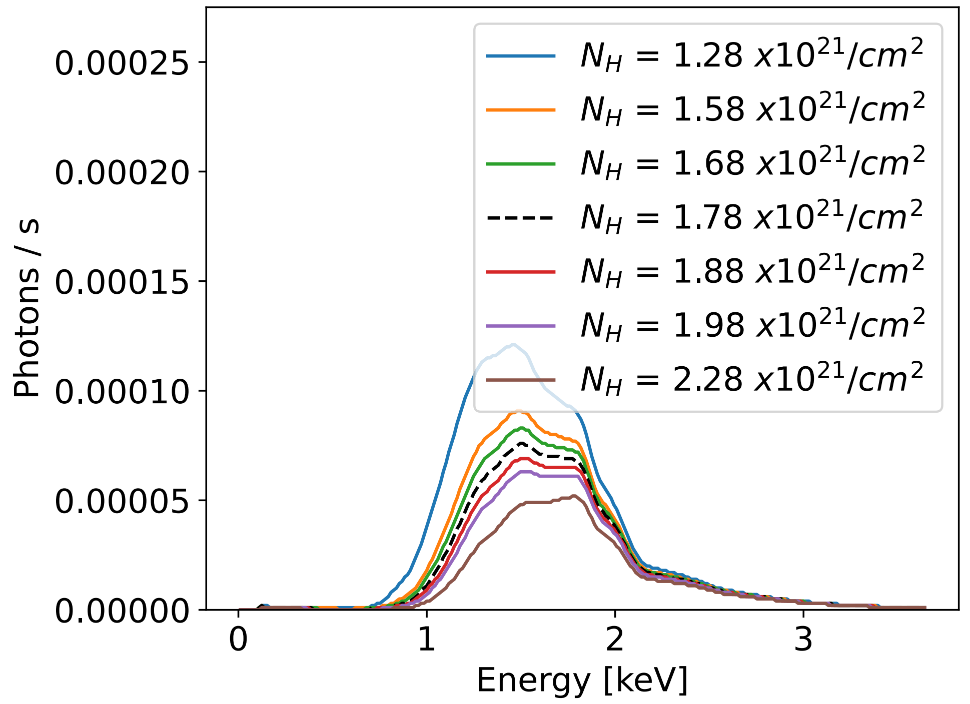


Image: Wikimedia/NASA

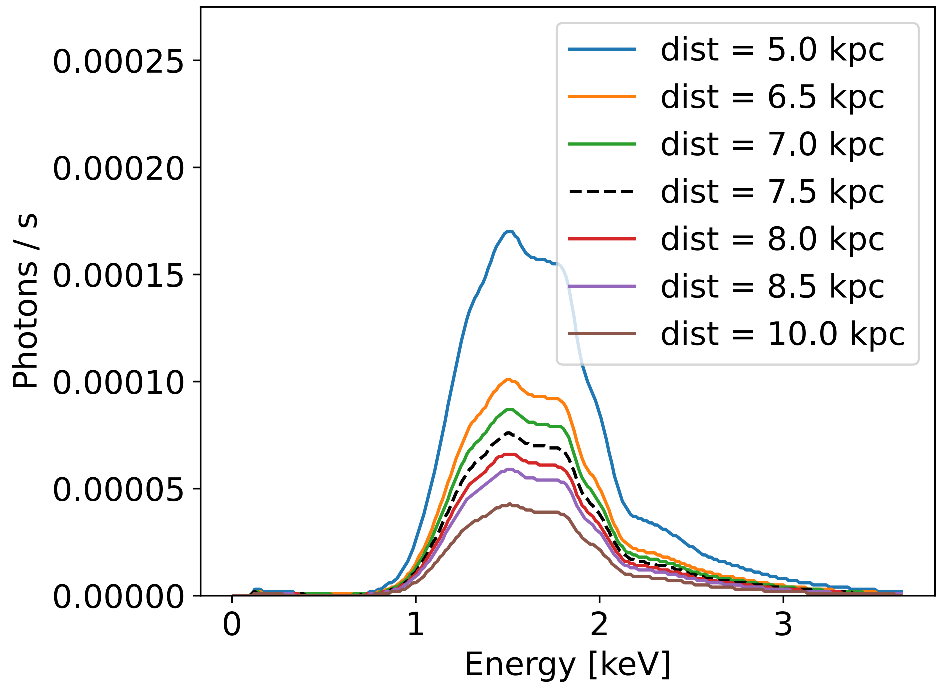
Mass

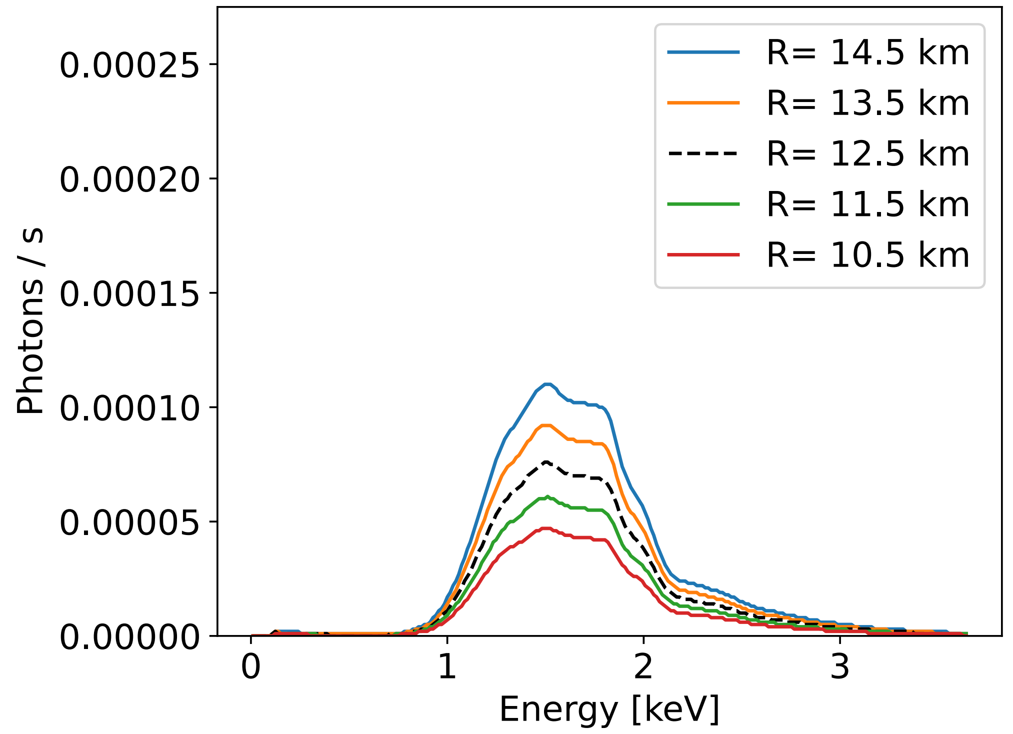


Hydrogen Column

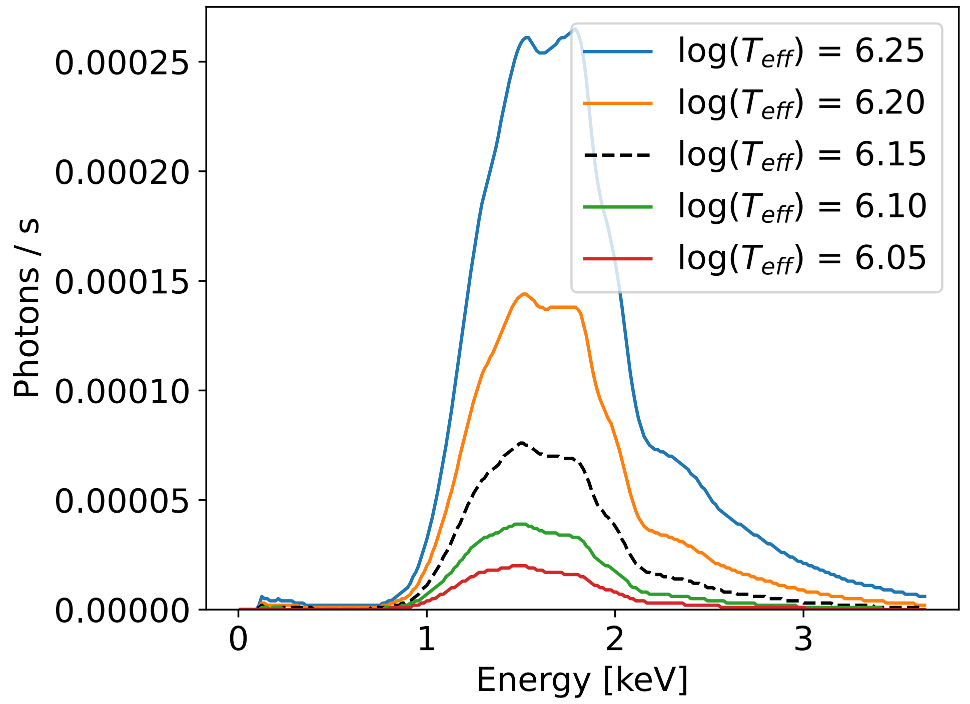


Distance





Radius

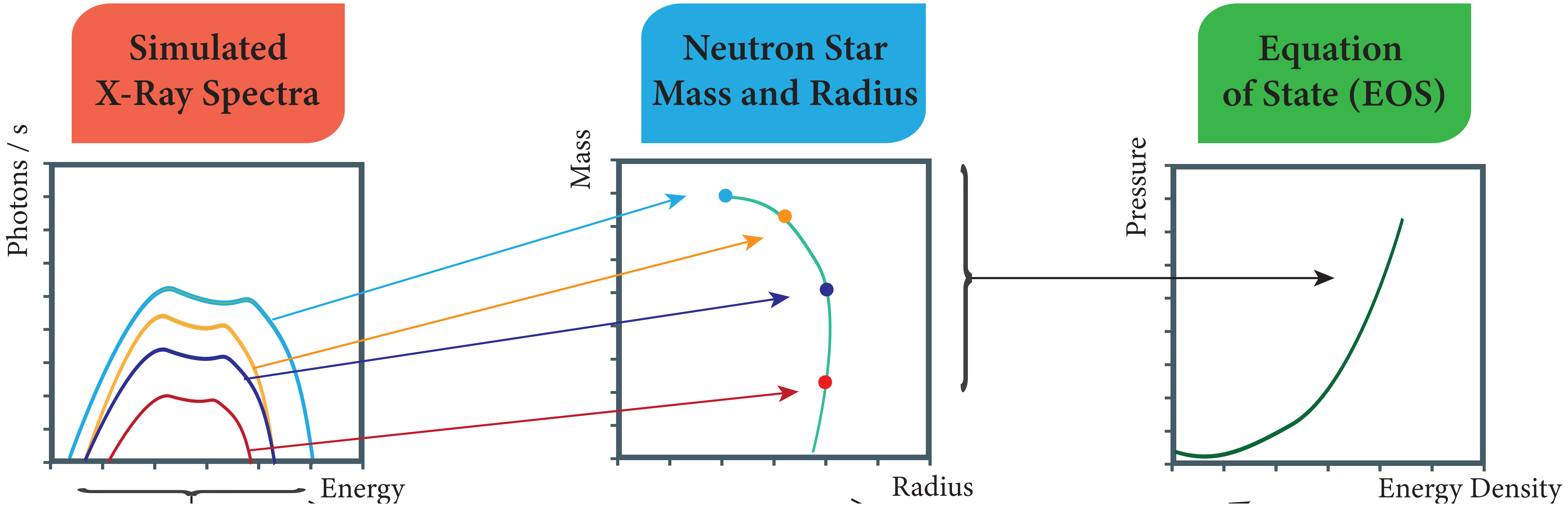


Effective Temperature

JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al

Traditional method: Two-step inference

Neutron star in sky



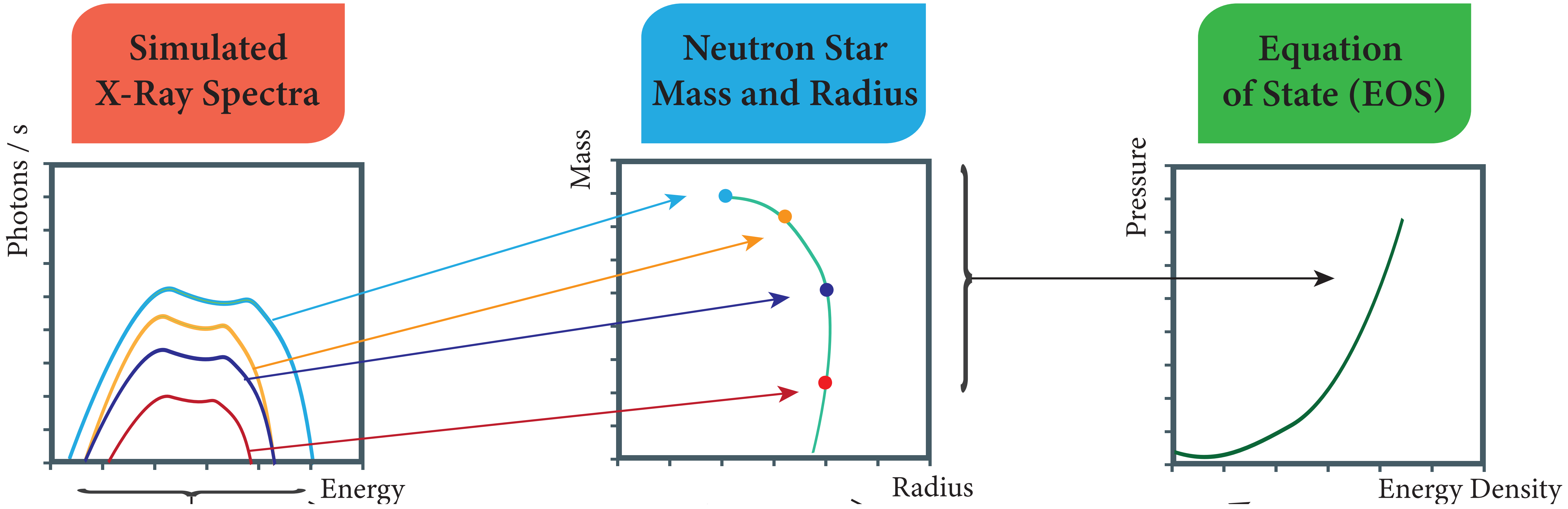
Astrophysicists

Nuclear theorists

JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al

Traditional method: Two-step inference

Neutron star in sky



Astrophysicists

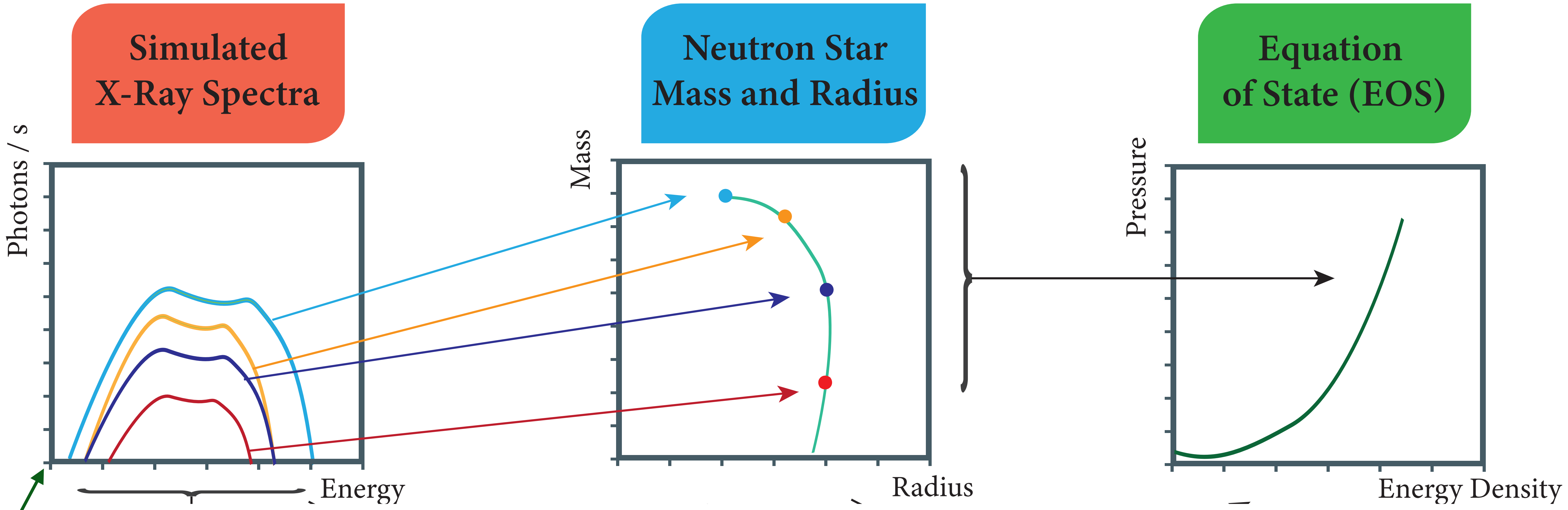
Nuclear theorists

Leak some information on uncertainties in the handover

JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al

Traditional method: Two-step inference

Neutron star in sky



Astrophysicists

Nuclear theorists

NPs

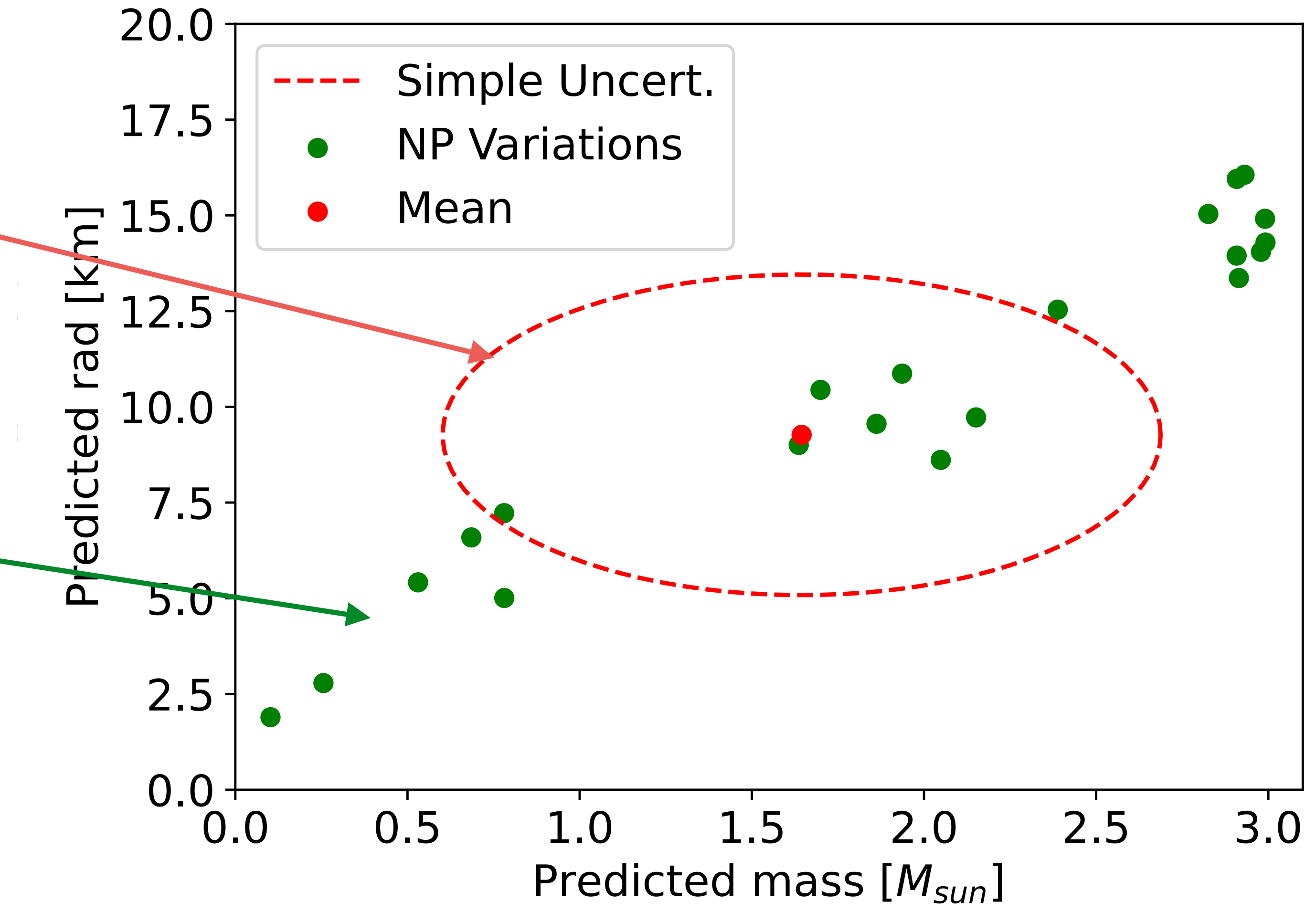
Leak some information on uncertainties in the handover

JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al

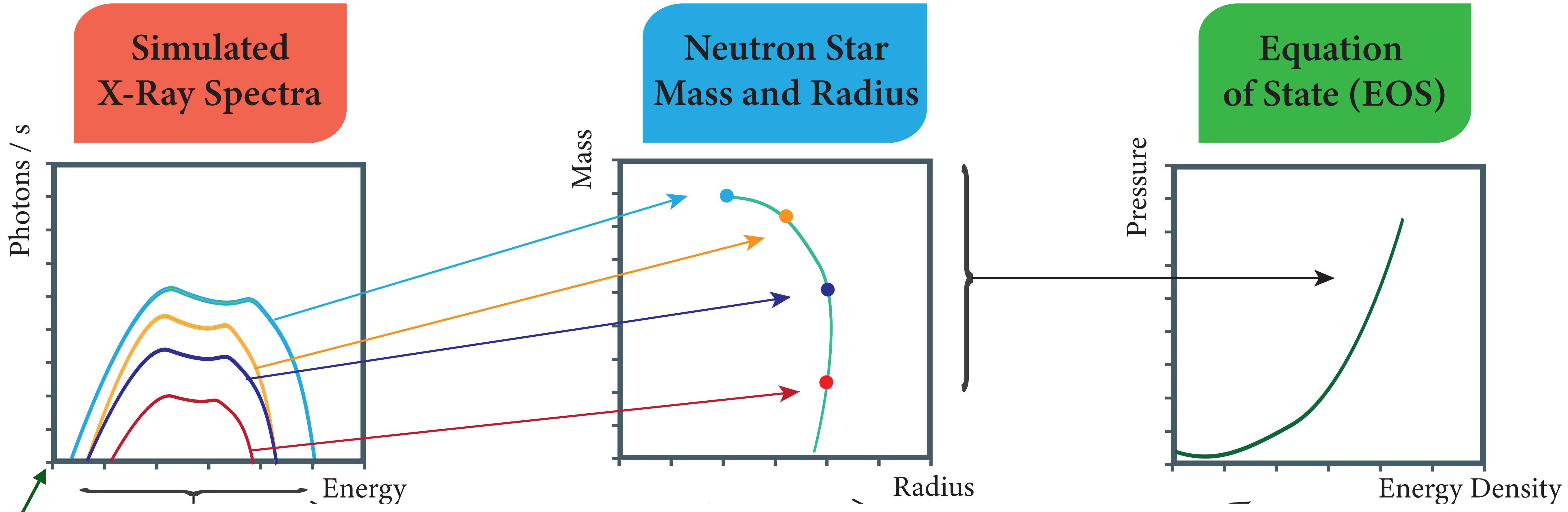
Traditional method: Two-step inference

SOTA collapsed information into 2 numbers + assumed uncorrelated Gaussian uncertainties

Real uncertainties look quite different

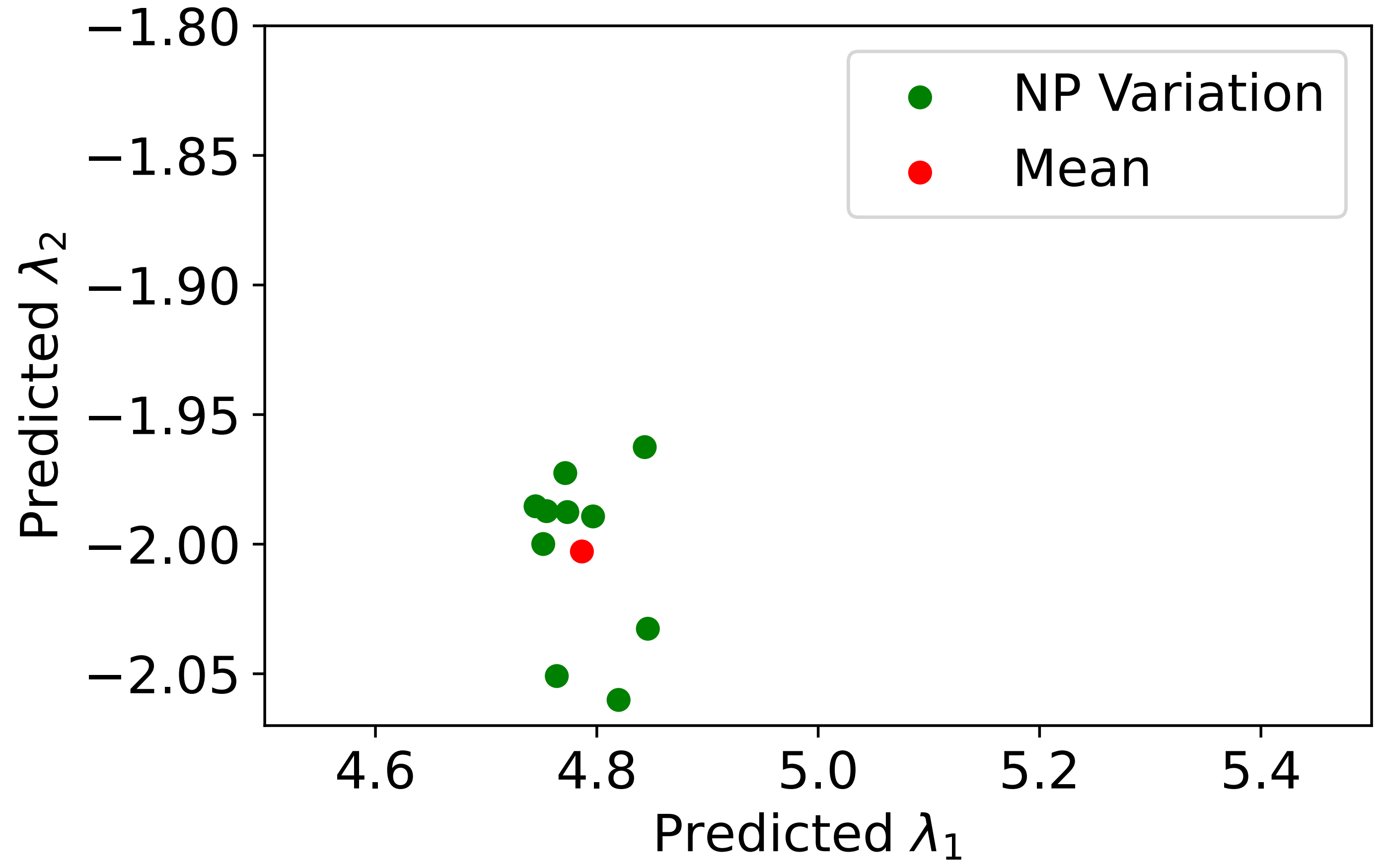


Neutron star in sky



Astrophysicists

Nuclear theorists



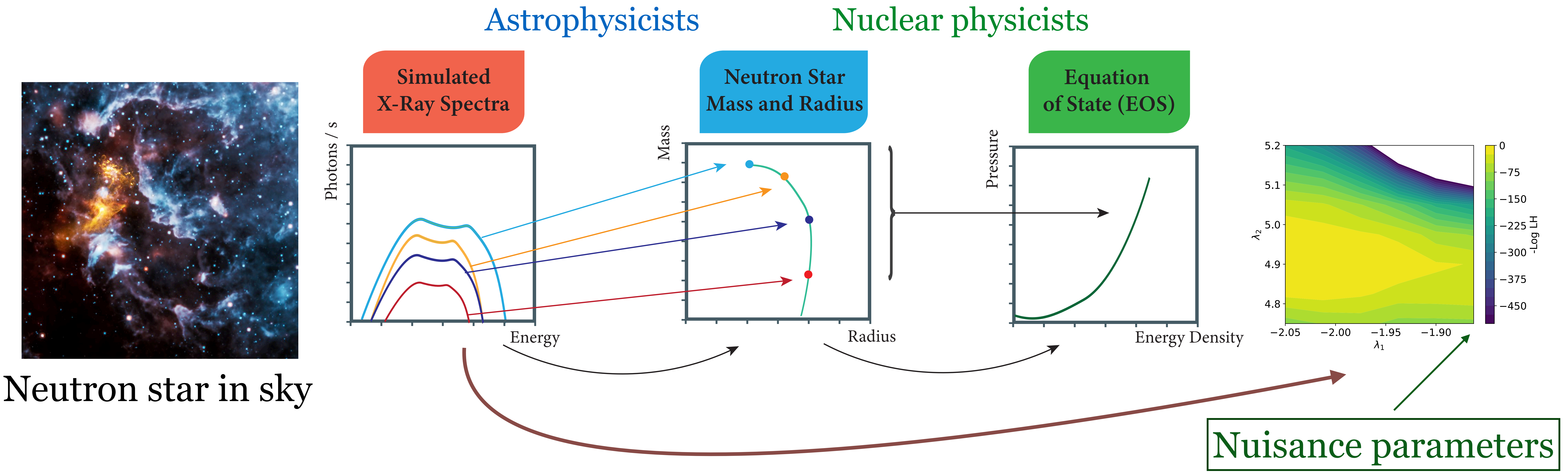
NPs

Leak some information on uncertainties in the handover

JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al

Inferring neutron star EoS parameters with NSBI

Recover the likelihood of EoS + NPs directly from the raw high-dimensional telescope spectra!

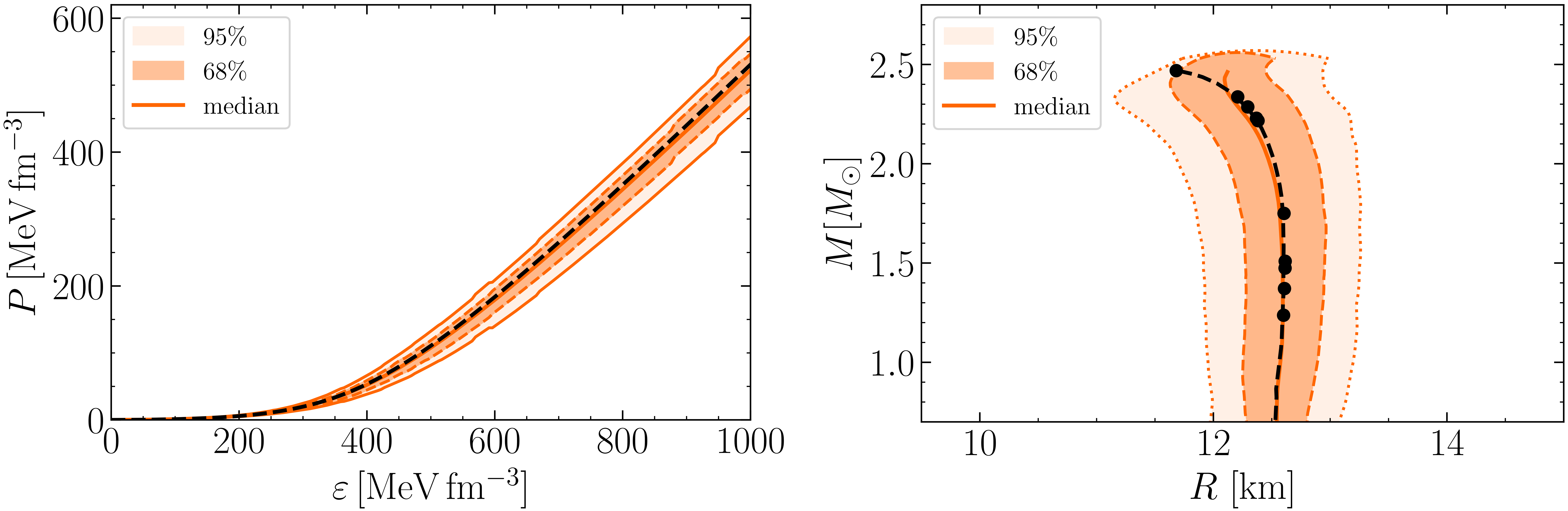


Direct estimation of likelihood from high-dimensional raw data allows more reliable uncertainty propagation and better measurements!

CAP09(2024)009: Brandes, Modi, Ghosh, et al
JCAP12(2023)022: Farrell, Baldi, Ott, Ghosh, et al
JCAP02(2023)016: Farrell, Baldi, Ott, Ghosh, et al

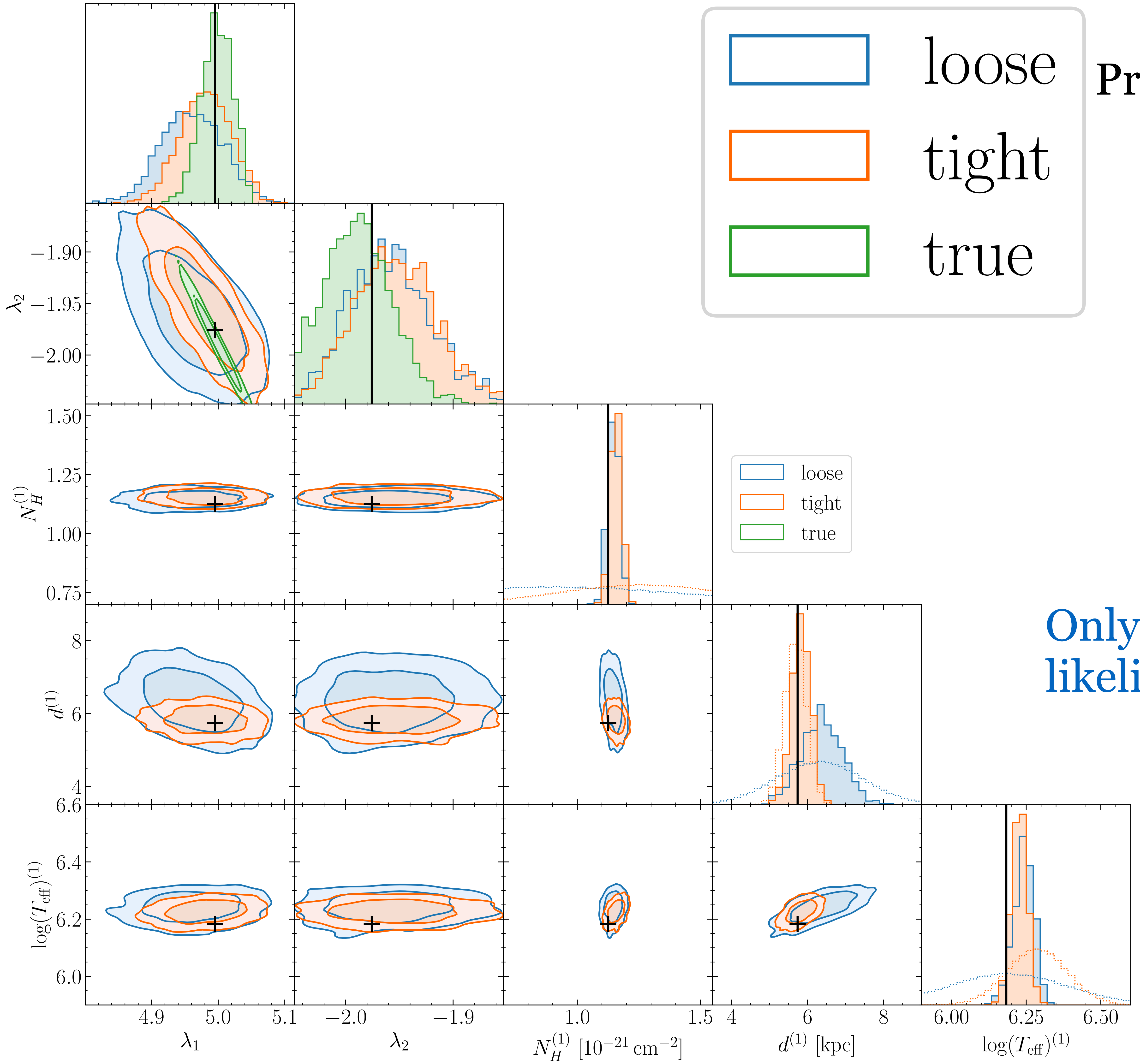
Meaningful posteriors, most sensitive method!

Bayesian Posteriors and credible intervals



CAP09(2024)009: Brandes, Modi, Ghosh, et al

Enhanced Interpretability: Effect of nuisance parameters

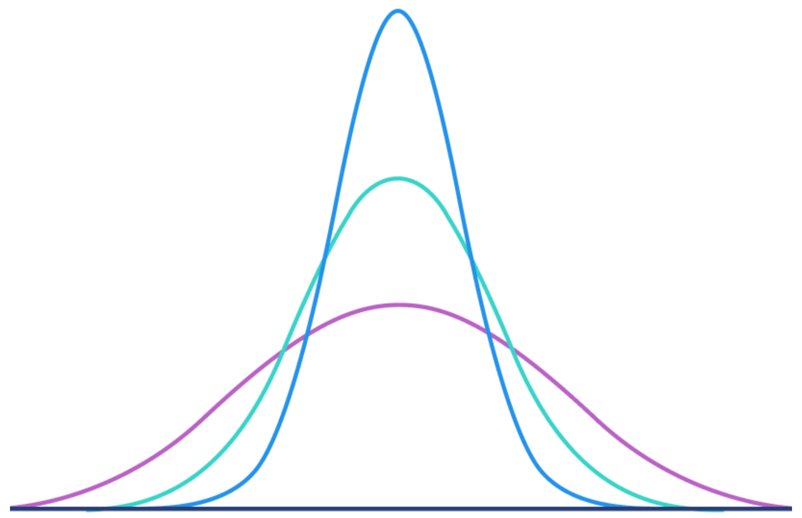


Prior knowledge on nuisance parameters

Only possible to visualise these due to the fast and differentiable likelihood from networks

Most sensitive method for EoS inference to date!

CAP09(2024)009: Brandes, Modi, Ghosh, et al



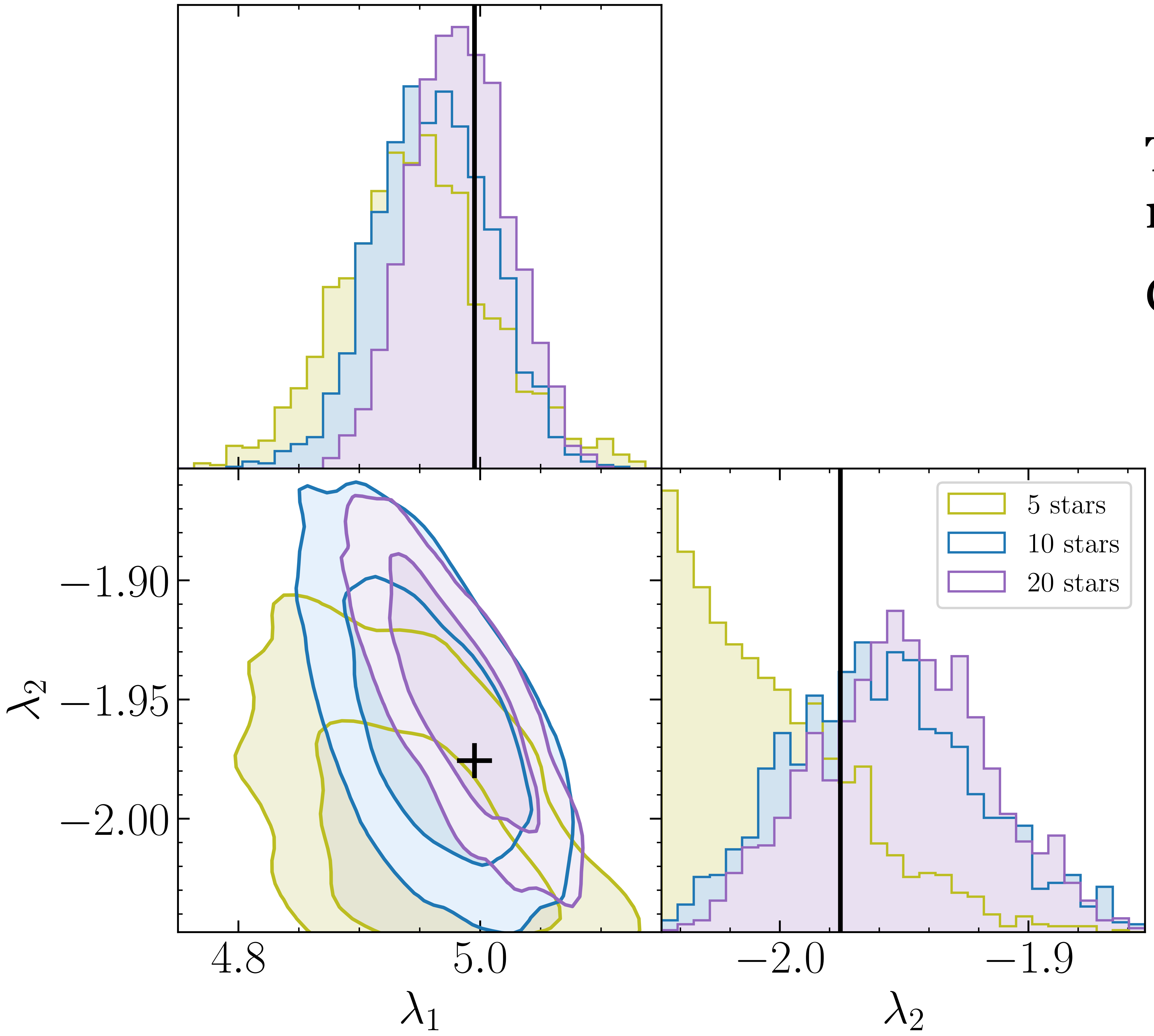
NP priors

		$\lambda_{1,\mathrm{pred}} - \lambda_{1,\mathrm{truth}}$		$\lambda_{2,\mathrm{pred}} - \lambda_{2,\mathrm{truth}}$		Combined
$p(\nu)$	Method	μ	σ	μ	σ	σ_{tot}
true	ML-Likelihood$_{\mathrm{EOS}}$	-0.02	0.066	**0.01**	0.070	0.096
	NN(Spectra)	-0.02	0.066	**0.01**	0.075	0.099
	NN(M, R via XSPEC)	-0.03	0.065	**0.01**	**0.055**	**0.085**
	NLE	**0.00**	**0.056**	**-0.01**	0.070	0.090
tight	ML-Likelihood$_{\mathrm{EOS}}$	-0.02	0.078	0.03	0.081	0.112
	NN(Spectra)	0.02	0.085	-0.02	0.077	0.115
	NN(M, R via XSPEC)	-0.03	0.081	**0.01**	**0.056**	0.098
	NLE	**0.00**	**0.066**	-0.02	0.071	**0.097**
loose	ML-Likelihood$_{\mathrm{EOS}}$	-0.04	0.089	0.03	0.081	0.120
	NN(Spectra)	-0.03	0.131	**-0.01**	0.078	0.152
	NN(M, R via XSPEC)	-0.03	0.123	**0.01**	**0.058**	0.136
	NLE	**0.00**	**0.085**	**-0.01**	0.074	**0.113**

Pretend that nuisance parameters known exactly

Realistic scenarios:

Which neutron stars should we measure next ?



Test potential improvement in sensitivity coming from new measurements

Could inform decisions on which stars to measure next!

CAP09(2024)009: Brandes, Modi, Ghosh, et al

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})\,p(\text{theory})}{p(\text{data})}$$

What we all want (Posterior)

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})\, p(\text{theory})}{p(\text{data})}$$

What we all want (Posterior)

Likelihood

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})\, p(\text{theory})}{p(\text{data})}$$

What we all want (Posterior)

Likelihood

Prior

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})\, p(\text{theory})}{p(\text{data})}$$

Evidence