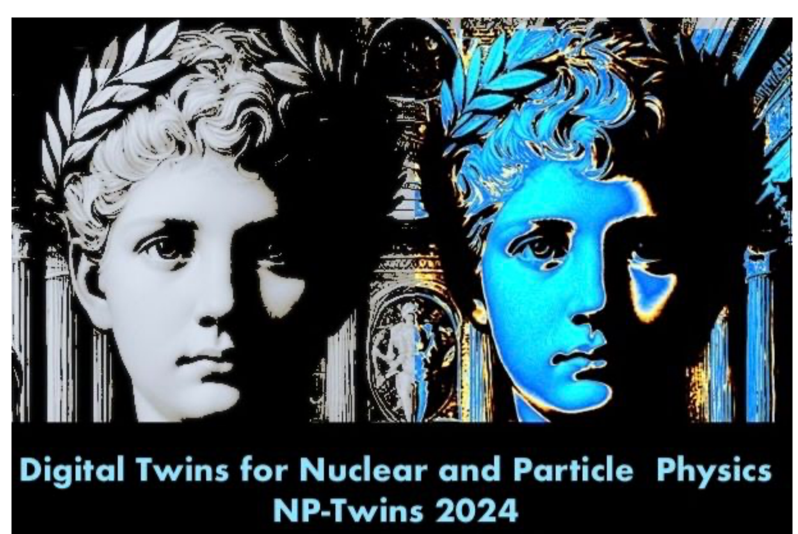
Neural Simulation-Based Inference for Parameter Estimation at the LHC

Jay Sandesara

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON



UNIVERSITY OF MASSACHUSETTS
1863

Introduction

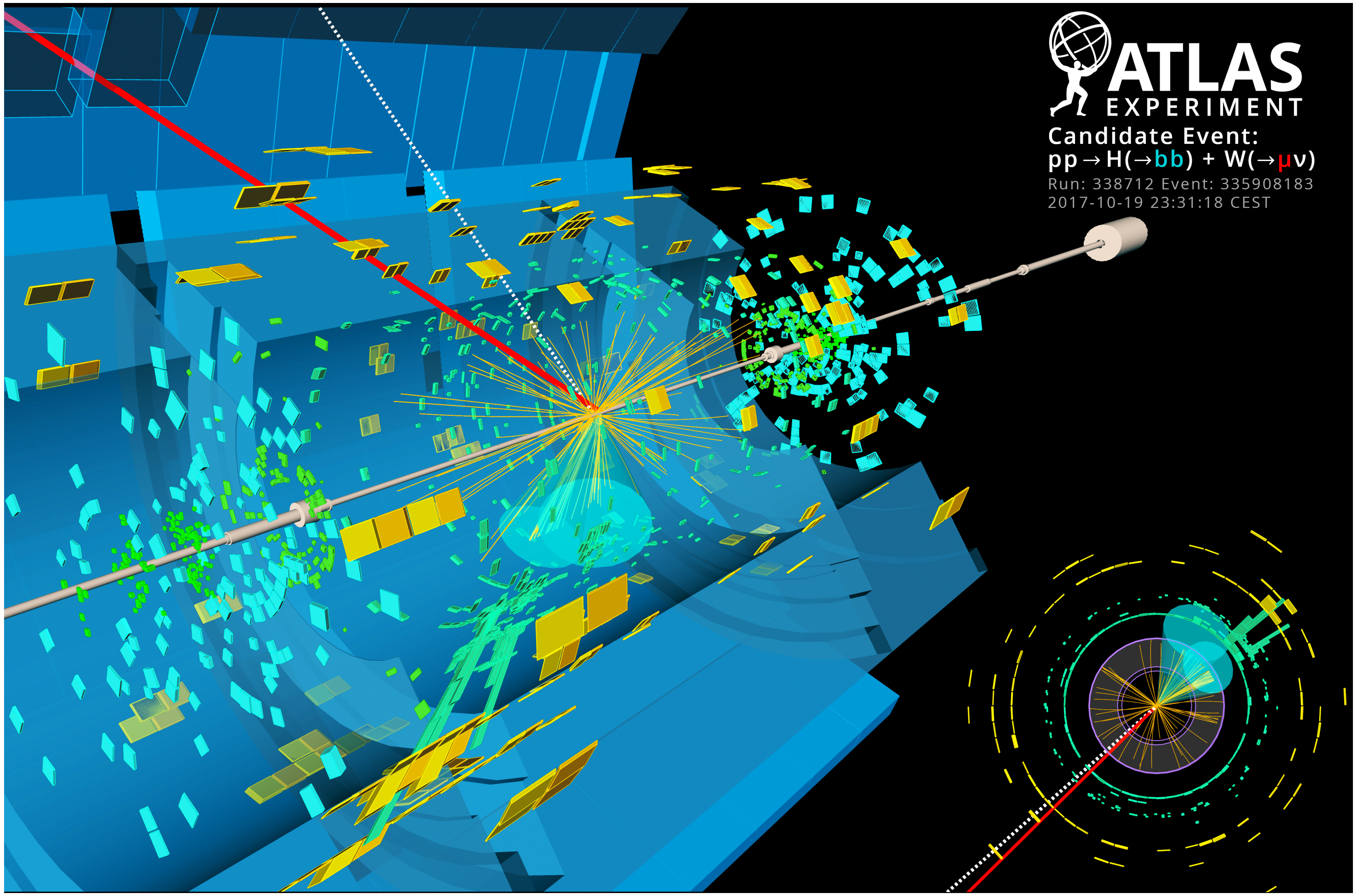
- Simulation-Based Inference (or Neural Simulation-Based Inference or Likelihood-Free Inference) covers a broad range of techniques.

- General idea is to use Machine Learning techniques for powerful statistical inference in the presence of intractable likelihoods (e.g. LHC analysis), or are slow to compute analytically (e.g. gravitational wave analysis).

- The focus of this talk is on a practical application of these methods to LHC analysis, with a particular example of Higgs boson width measurement at the ATLAS experiment.

- For more general overview see the excellent review paper by Cranmer et al: https://www.pnas.org/doi/10.1073/pnas.1912789117

Measurements at the LHC

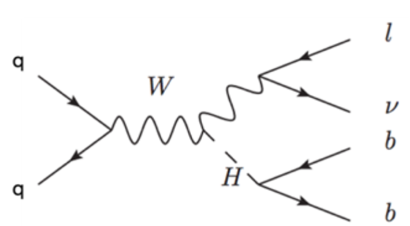


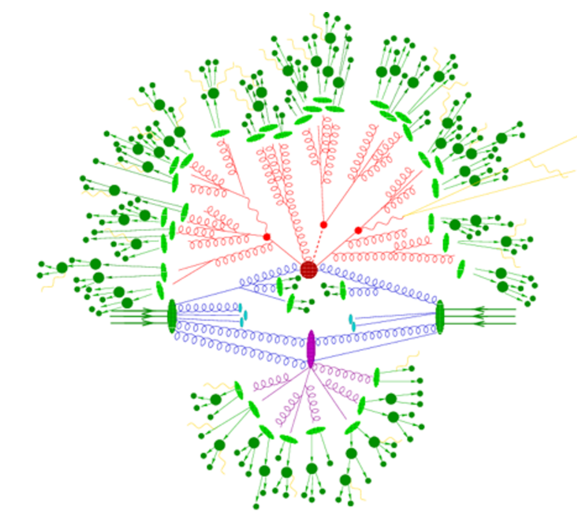
How a typical $p-p$ collision event looks like in the ATLAS detector which has $O(100M)$ sensors

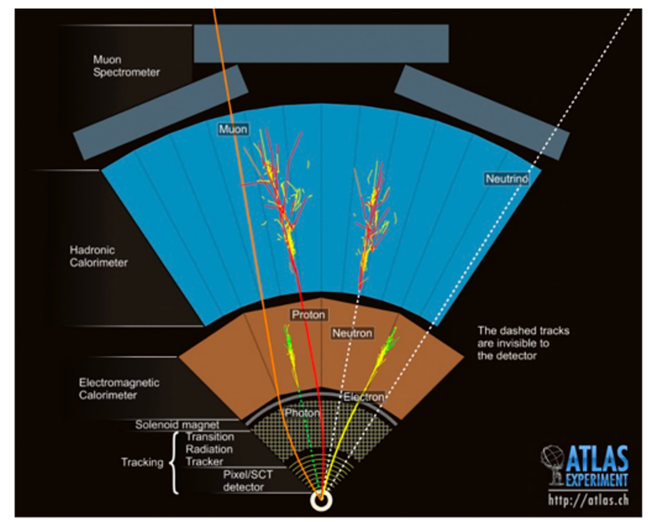
Measurements at the LHC

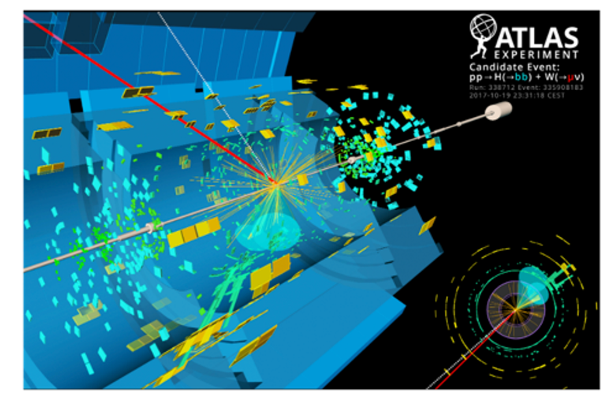
Physics model of interest → **What we observe**

z z' z'' x









Hard-scattering

$$p(z|\theta) = \frac{1}{\sigma(\theta)}\frac{d\sigma}{dz}(z|\theta)$$

Calculable using perturbative QFT

Parton-Showers

Soft QCD processes modelled using non-perturbative physics

Detector Interaction Simulation

The particles are simulated to interact with detector electronics.

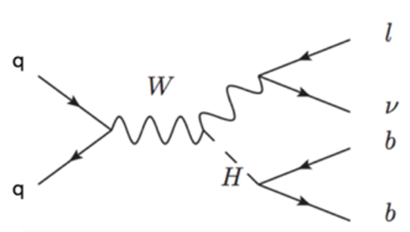
Reconstructed output

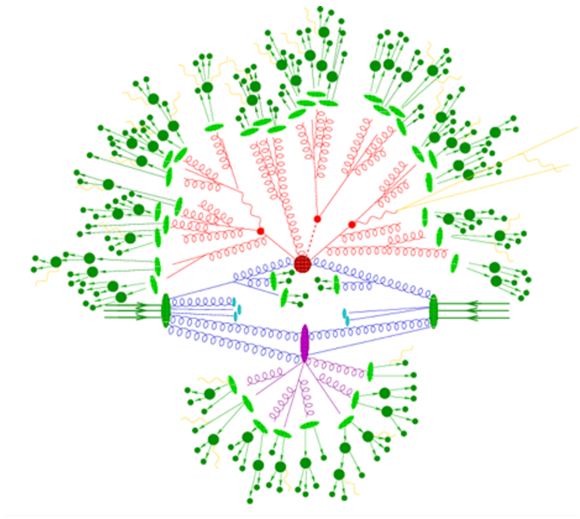
The final output with signal from the particle interactions $O(100M)$ sensors

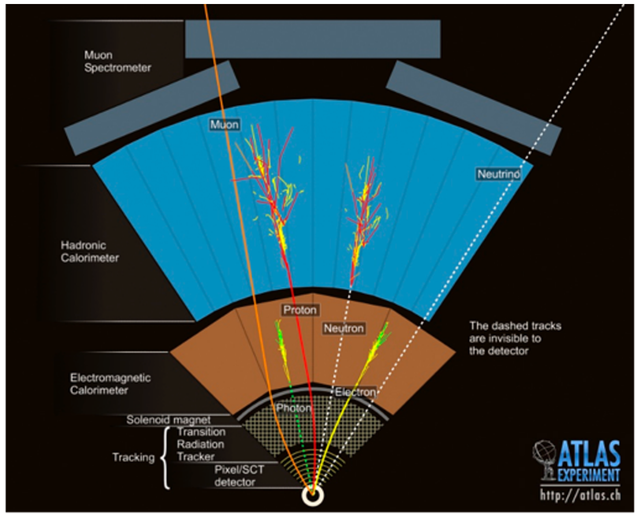
Measurements at the LHC

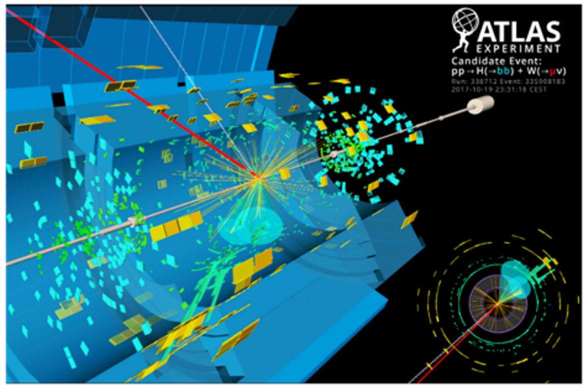
Physics model of interest → **What we observe**

z | z' | z'' | x









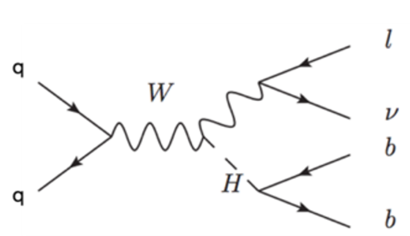
Hard-scattering | **Parton-Showers** | **Detector Interaction Simulation** | **Reconstructed output**

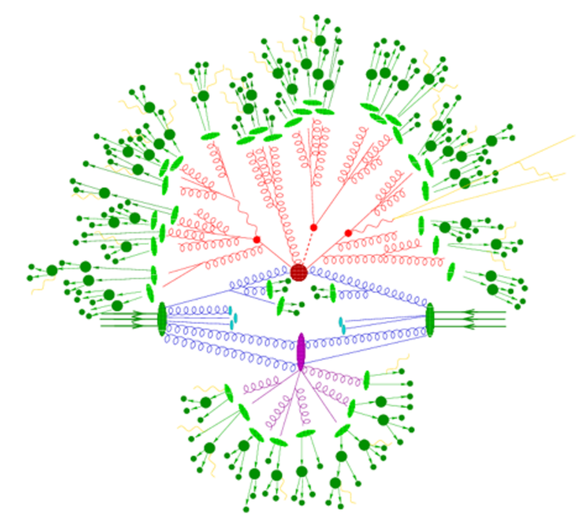
$$p(x\,|\,\theta) = \int dz\, dz'\, dz''\, p(x\,|\,z'')\, p(z''\,|\,z)\, p(z'\,|\,z)\, p(z\,|\,\theta)$$

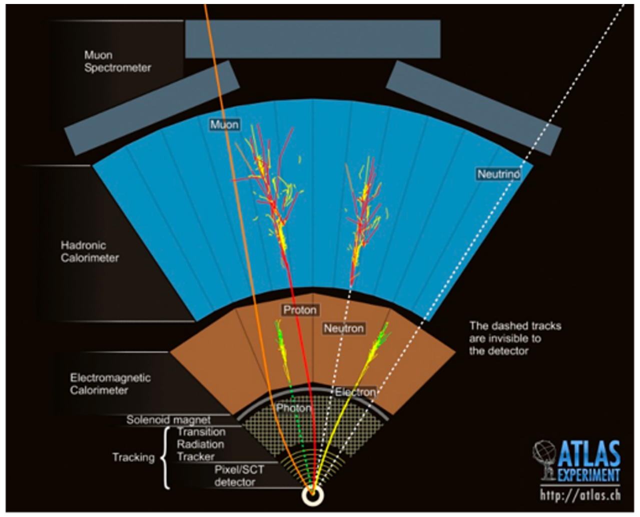
Measurements at the LHC

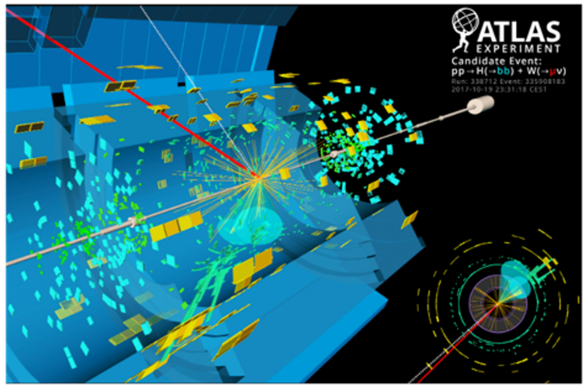
Physics model of interest → **What we observe**

z z' z'' x









Hard-scattering **Parton-Showers** **Detector Interaction Simulation** **Reconstructed output**

$$p(x\,|\,\theta) = \int dz\, dz'\, dz''\, p(x\,|\,z'')\, p(z''\,|\,z)\, p(z'\,|\,z)\, p(z\,|\,\theta)$$

Large dimensional integral - needs Monte Carlo techniques

Measurements at the LHC

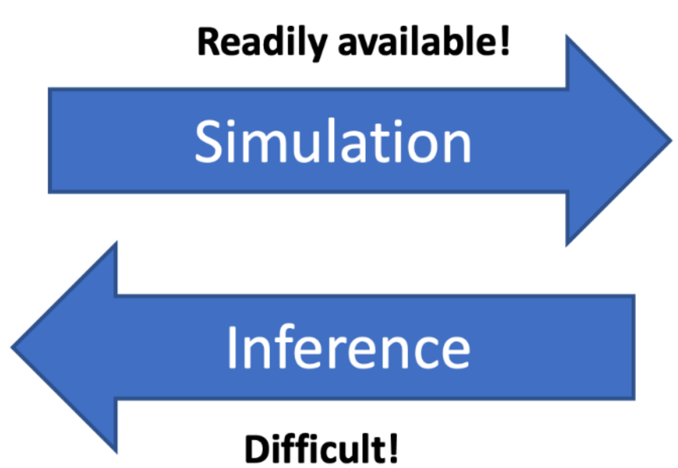
Physics model of interest → **What we observe**

The statistical model for LHC parameter inference is built using Monte Carlo techniques

$$p(x\,|\,\theta) = \frac{1}{\sigma(\theta)}\frac{d\sigma}{dx}(x\,|\,\theta)$$

This is only a forward model - we can sample events from it, but going in the other direction, building $p(x\,|\,\theta)$ from x is the challenging part.

$p(x|\theta)$

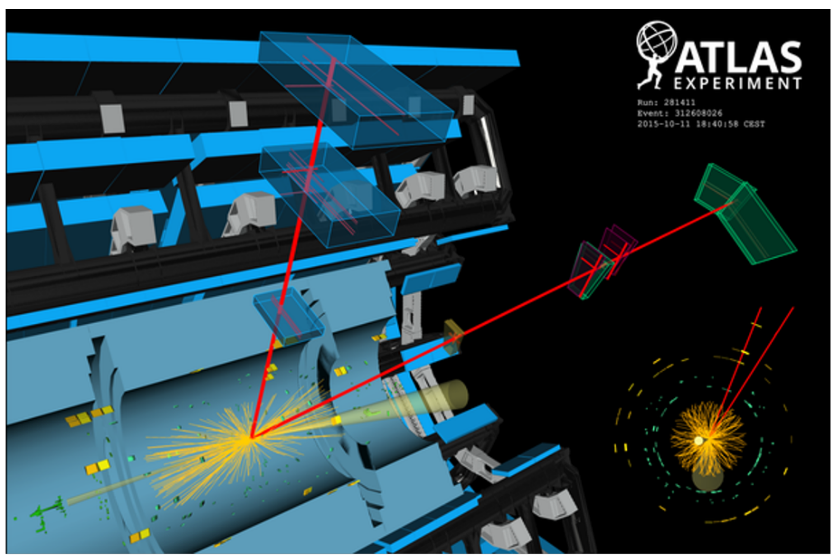


$x \sim p(x|\theta)$

Physics model of interest ← **What we observe**

Measurements at the LHC

Physics model of interest ← What we observe



Raw $O(100M)$-dimensional output from the detector

→

Object Reconstruction

↓

Object Selections

↓

High-Level Features

$$x = \left(E, p_x, p_y, p_z\right)$$

↓

Event Selections

← Four momenta (plus flavour and charge information) of the particles in the final state.

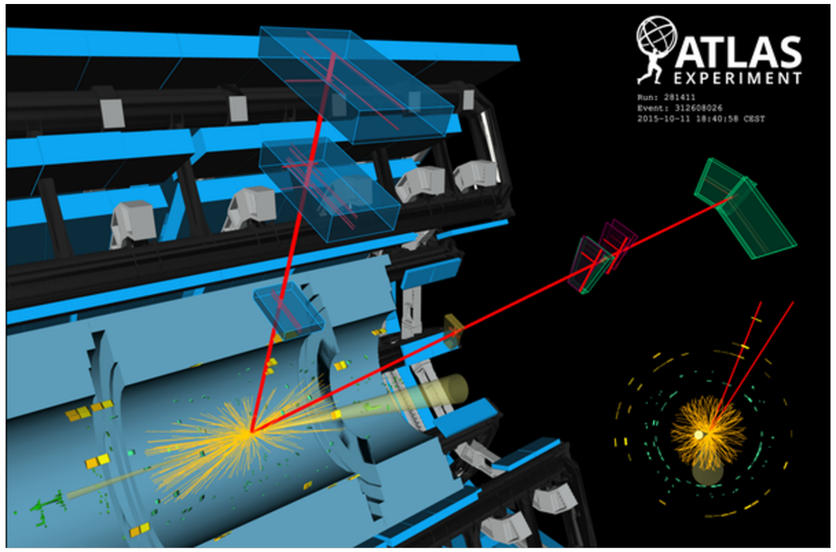
$O(100M) \rightarrow O(100)$ dimensional output

The large dimension of the output makes the inference a difficult task!

We reduce the dimensionality of the data using a combination of physics-motivated and ML-based algorithms

Measurements at the LHC

Physics model of interest ← **What we observe**



→

Object Reconstruction

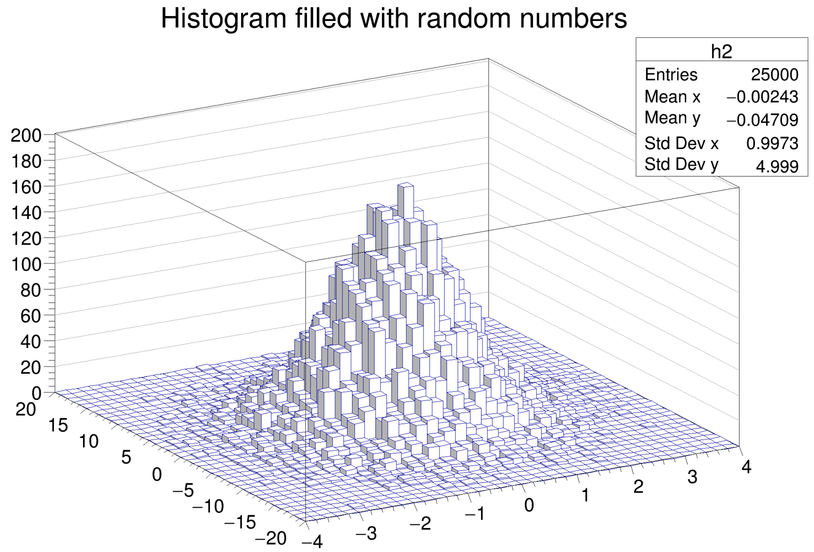
Object Selections

High-Level Features

$$x = \left(E, p_x, p_y, p_z\right)$$

Event Selections

→



Raw $O(100M)$-dimensional output from the detector

The large dimension of the output makes the inference a difficult task!

We reduce the dimensionality of the data using a combination of physics-motivated and ML-based algorithms

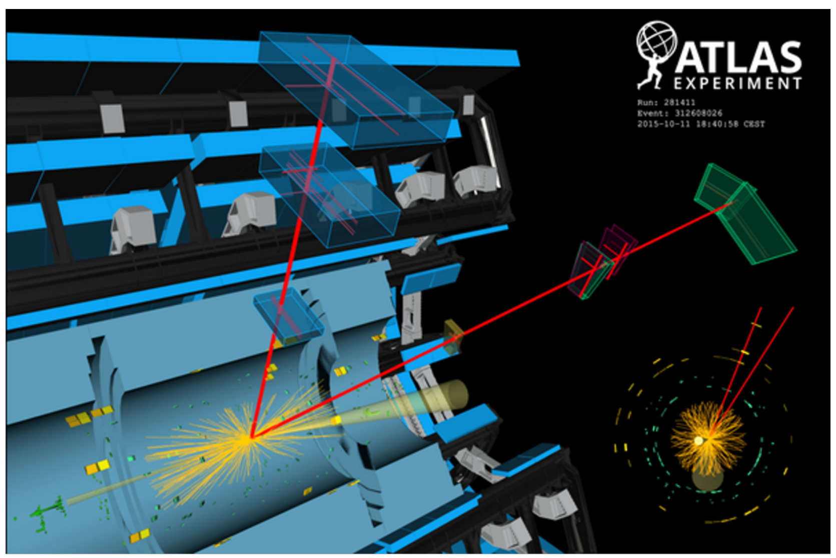
Build $O(100)$- dimensional histograms to model $p(x\,|\,\theta)$?

Impractical - Monte Carlo statistics needed grows exponentially with each additional dimension.

Simulation is very expensive!

Measurements at the LHC

Physics model of interest ← **What we observe**



Raw $O(100M)$-dimensional output from the detector

→

Object Reconstruction

↓

Object Selections

↓

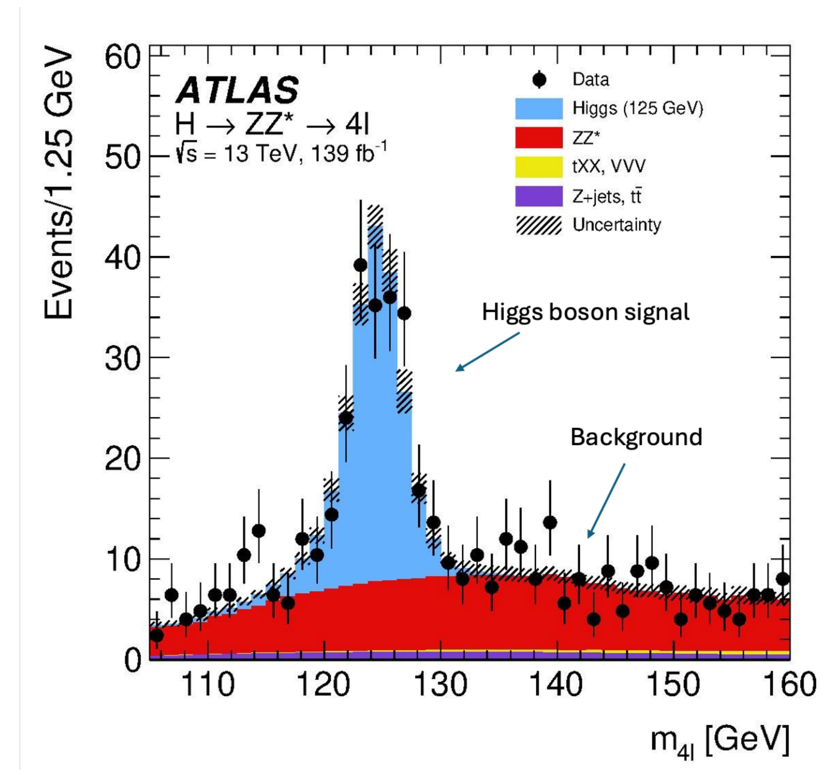
High-Level Features

$$x = \left(E, p_x, p_y, p_z\right)$$

↓

Event Selections

→



"Summary Statistic" used for statistical inference on physics models.

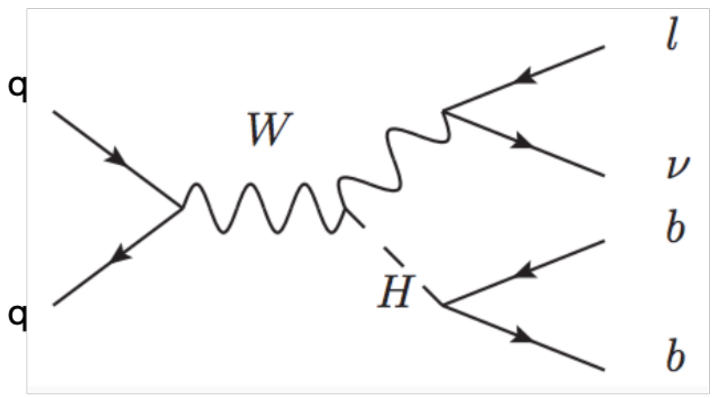
Summarizes the important information from the multi-dimensional output to a low-dimensional representation.

The large dimension of the output makes the inference a difficult task!

We reduce the dimensionality of the data using a combination of physics-motivated and ML-based algorithms

Frequentist Hypothesis Tests

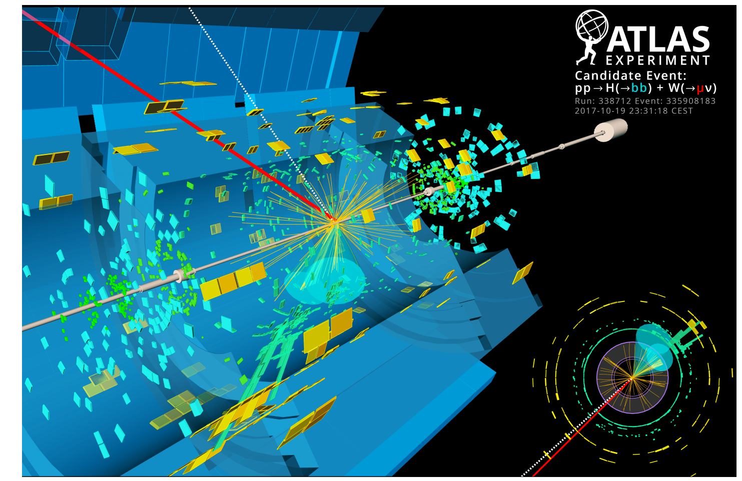
$x \sim p(x \mid \theta)$



Parton Shower

Hadronization

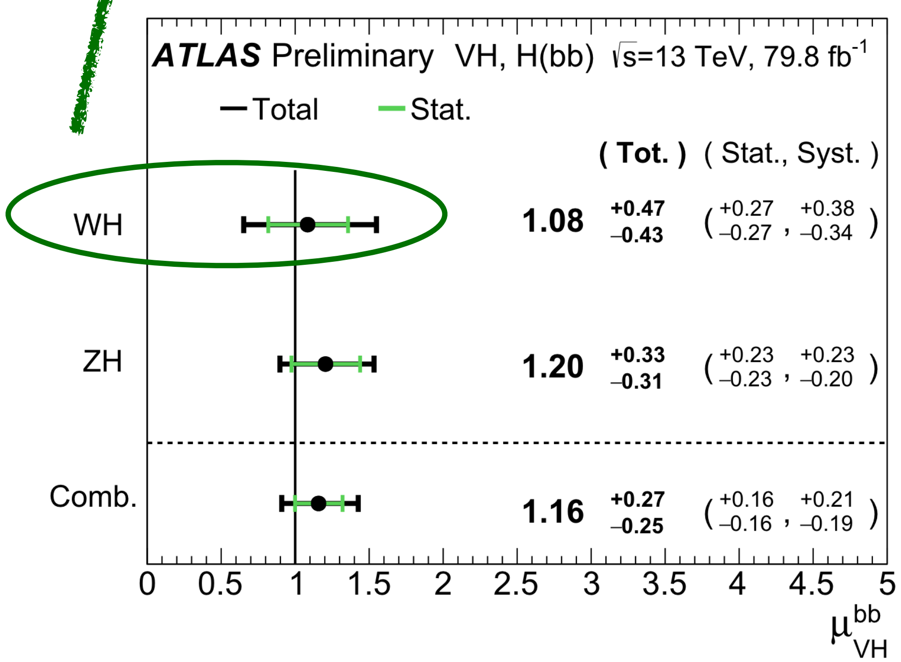
Detector Reconstruction



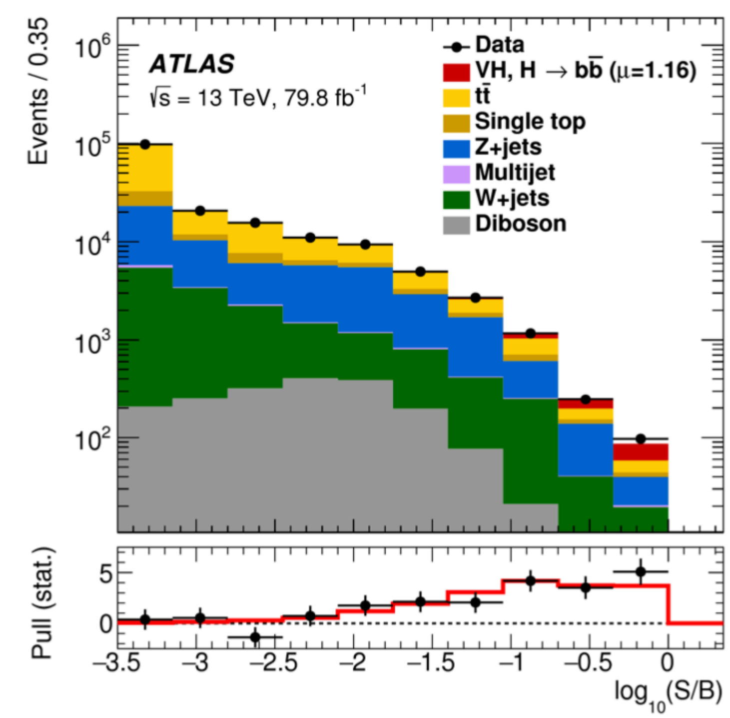
Inference

$$L(\theta \mid x) = \prod_{bin} \frac{e^{-\nu_{bin}(\theta)} \cdot \nu_{bin}(\theta)^{\mathcal{N}_{bin}}}{\mathcal{N}_{bin}!}$$

Binned Poisson Likelihood



Profile Negative Log-Likelihood Fit



Summary Statistic $u(x)$

Limitations

- The dimensional reduction of the full final state information to a low-dimensional summary statistic $x \to u(x)$ can result in a loss of information.

- The summary statistic $u(x)$ is independent of the parameters θ corresponding to hypothesis being tested. $u(x)$ can be optimal locally, but not globally.

- All the events in a bin have the same probability. Information from rare possibly signal-like events can be lost in this approximation.

$$p(x_i | \theta) = \boxed{\frac{\nu_{I(x_i|u)}(\theta)}{\nu(\theta)}}$$

← Fraction of events in the bin of x_i
(Same likelihood for every event in the bin)

Limitations

- The dimensional reduction of the full final state information to a low-dimensional summary statistic $x \to u(x)$ can result in a loss of information.

- The summary statistic $u(x)$ is independent of the parameters θ corresponding to hypothesis being tested. $u(x)$ can be optimal locally, but not globally.

- All the events in a bin have the same probability. Information from rare possibly signal-like events can be lost in this approximation.

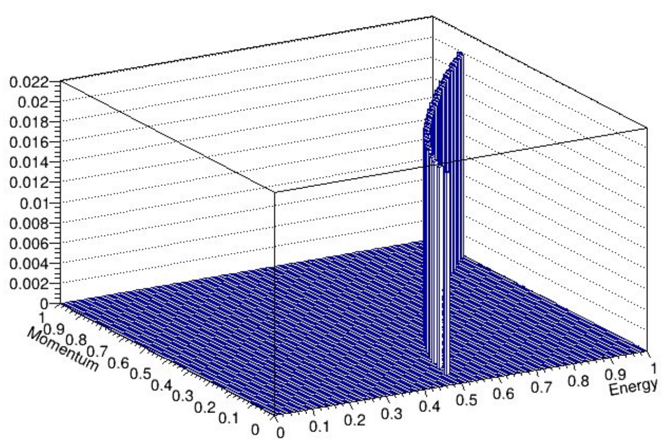
$$p(x_i \mid \theta) = \boxed{\frac{\nu_{I(x_i|u)}(\theta)}{\nu(\theta)}}$$

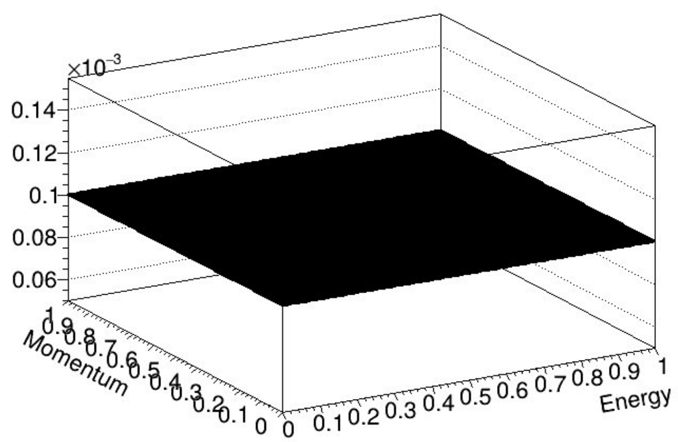
← Fraction of events in the bin of x_i
(Same likelihood for every event in the bin)

Measurements at the LHC

Typical measurements in ATLAS consist of a small "signal" of interest in a mountain of backgrounds!

Objective: find the optimal reduction $x \rightarrow u(x)$, from the multi-dimensional high-level feature space x, that optimally isolates the signal from background.

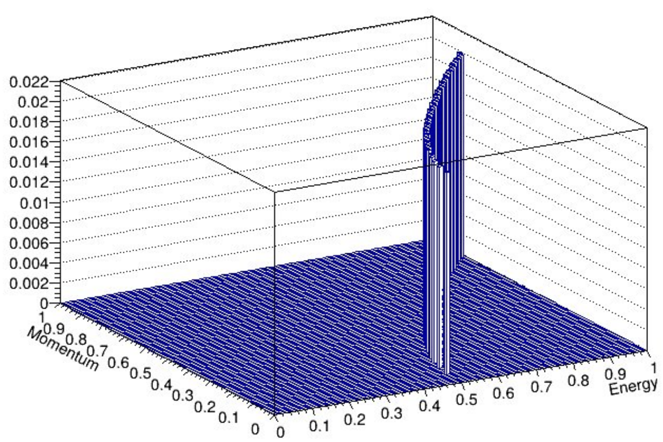


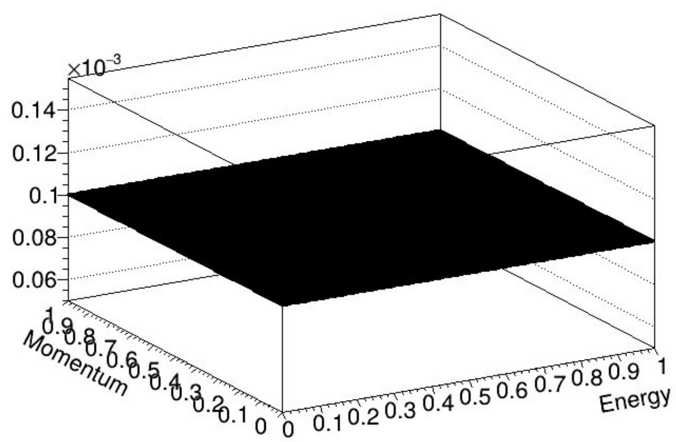


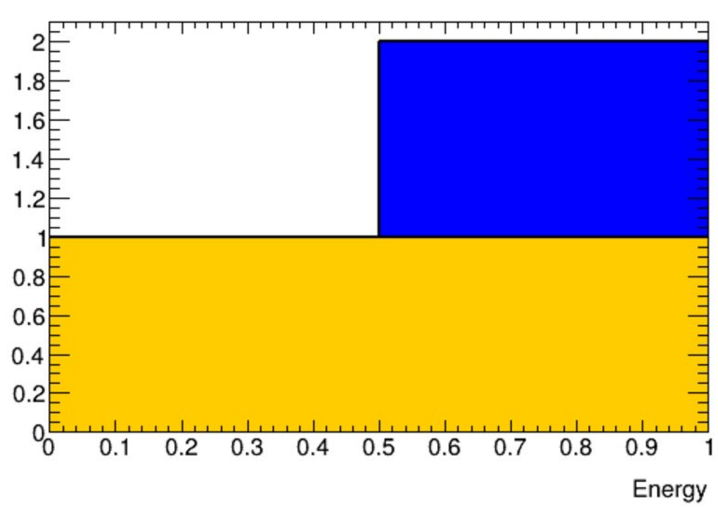
Measurements at the LHC

Typical measurements in ATLAS consist of a small "signal" of interest in a mountain of backgrounds!

Objective: find the optimal reduction $x \rightarrow u(x)$, from the multi-dimensional high-level feature space x, that optimally isolates the signal from background.



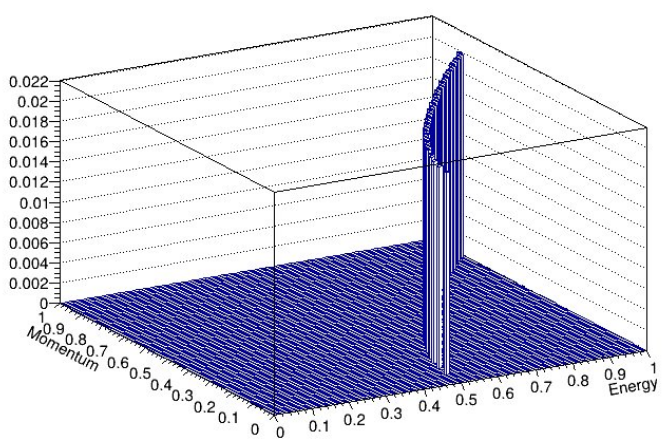


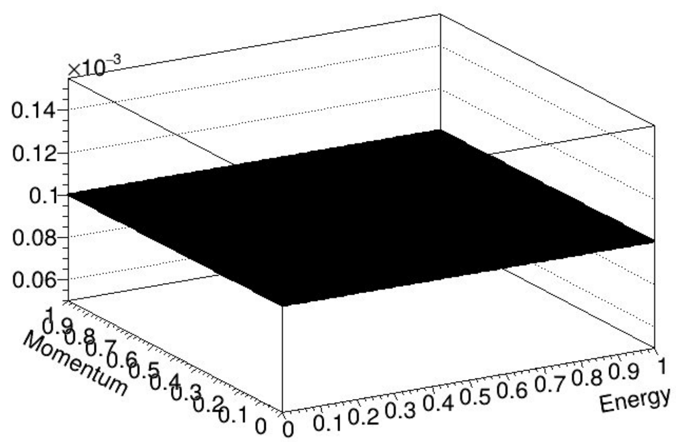


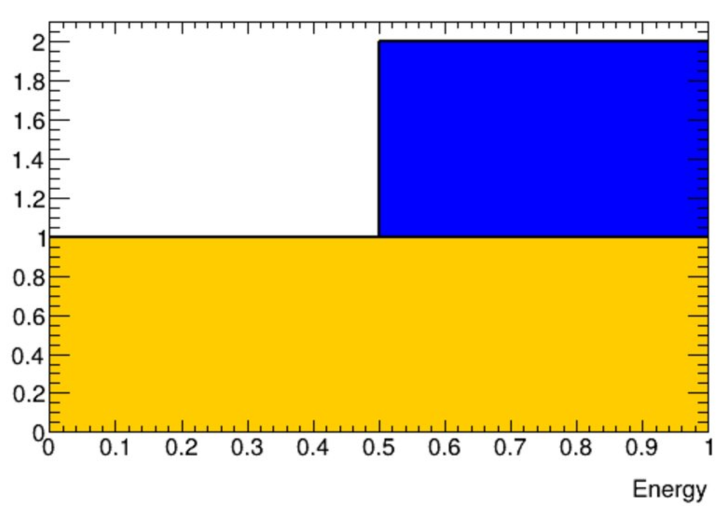
Measurements at the LHC

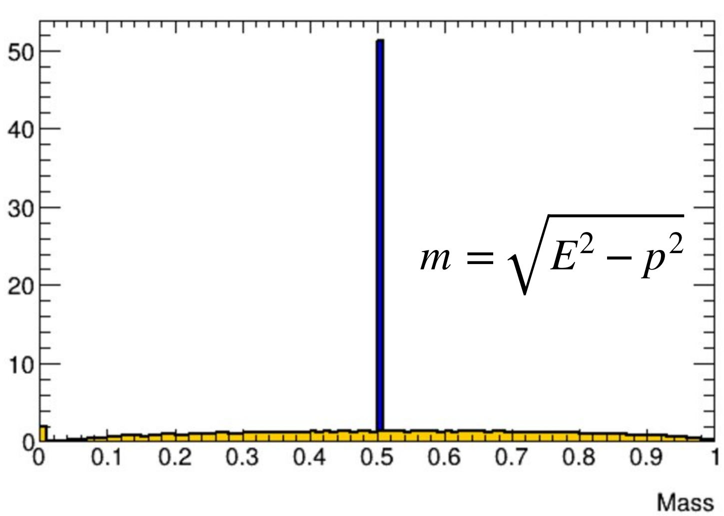
Typical measurements in ATLAS consist of a small "signal" of interest in a mountain of backgrounds!

Objective: find the optimal reduction $x \rightarrow u(x)$, from the multi-dimensional high-level feature space x, that optimally isolates the signal from background.









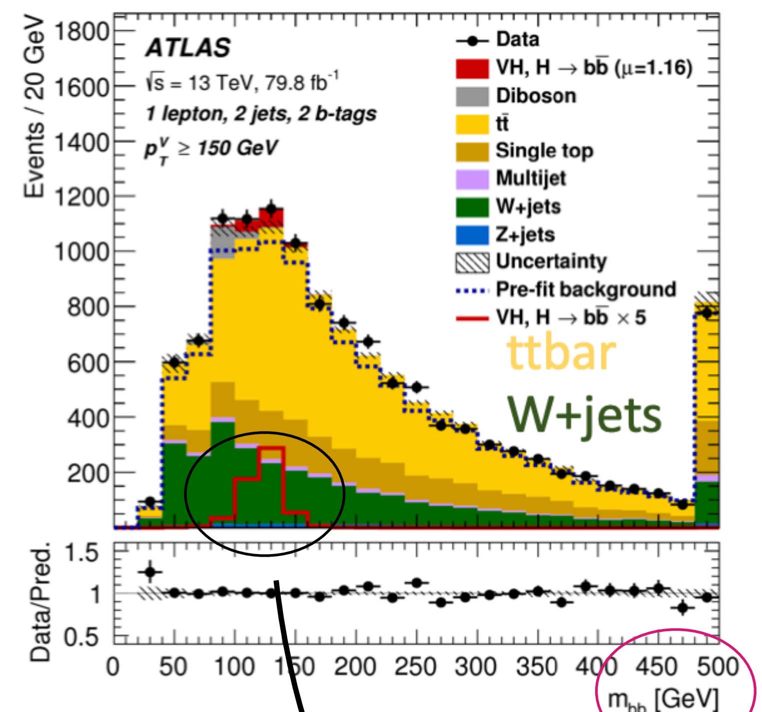
← **Use of "domain" knowledge to build summary statistic**

Measurements at the LHC

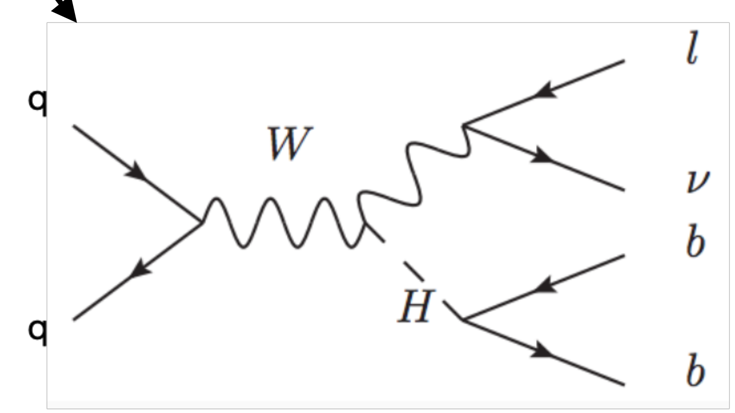
Typical measurements in ATLAS consist of a small "signal" of interest in a mountain of backgrounds!

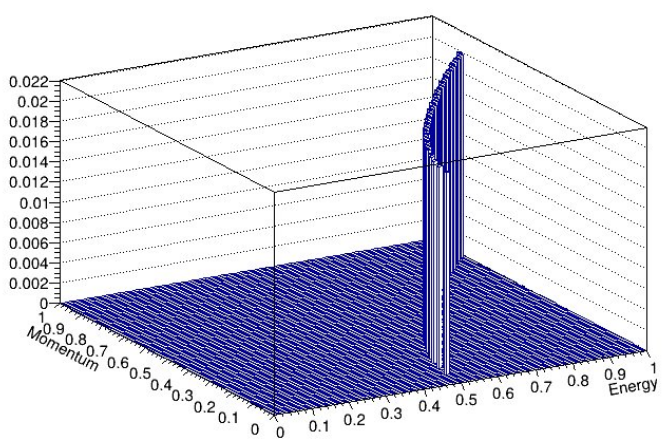
Objective: find the optimal reduction $x \rightarrow u(x)$, from the multi-dimensional high-level feature space x, that optimally isolates the signal from background.

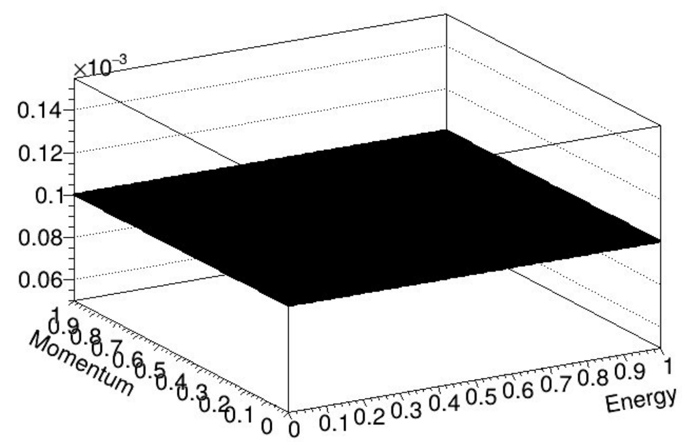
Example:

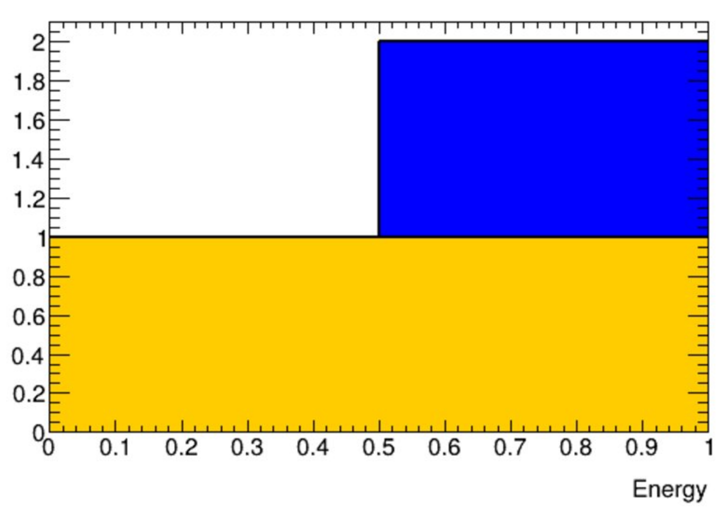


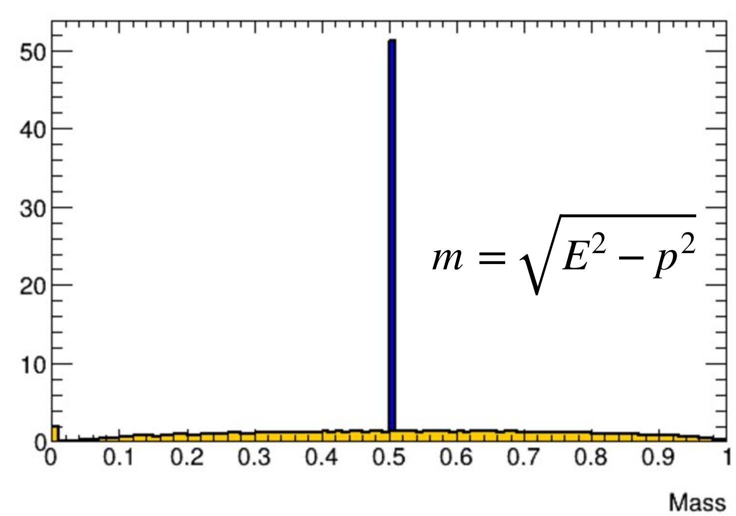
Invariant mass (reconstruction of Higgs boson mass)







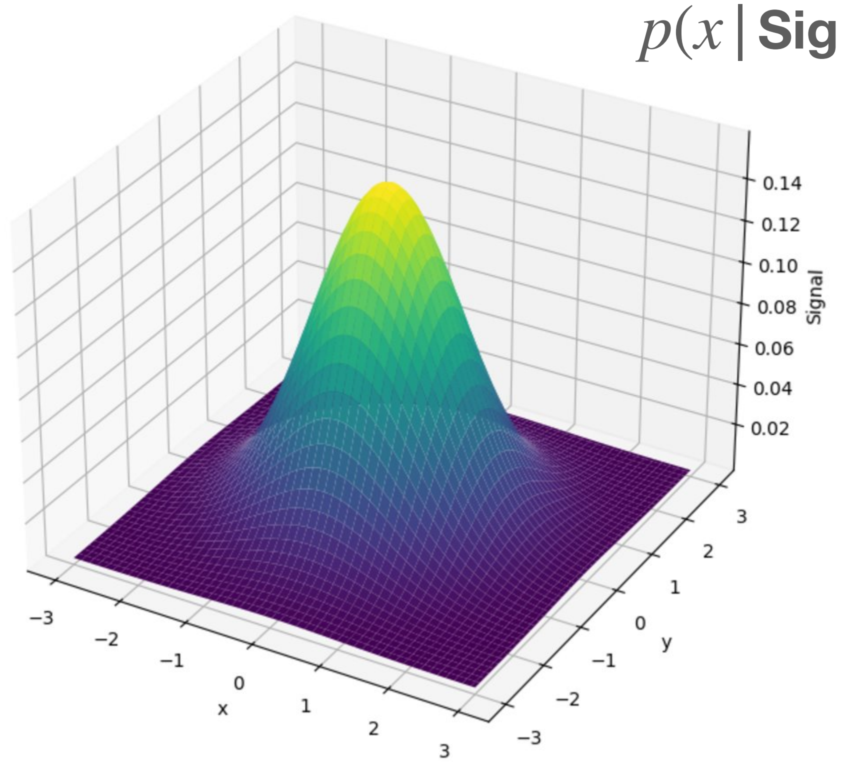




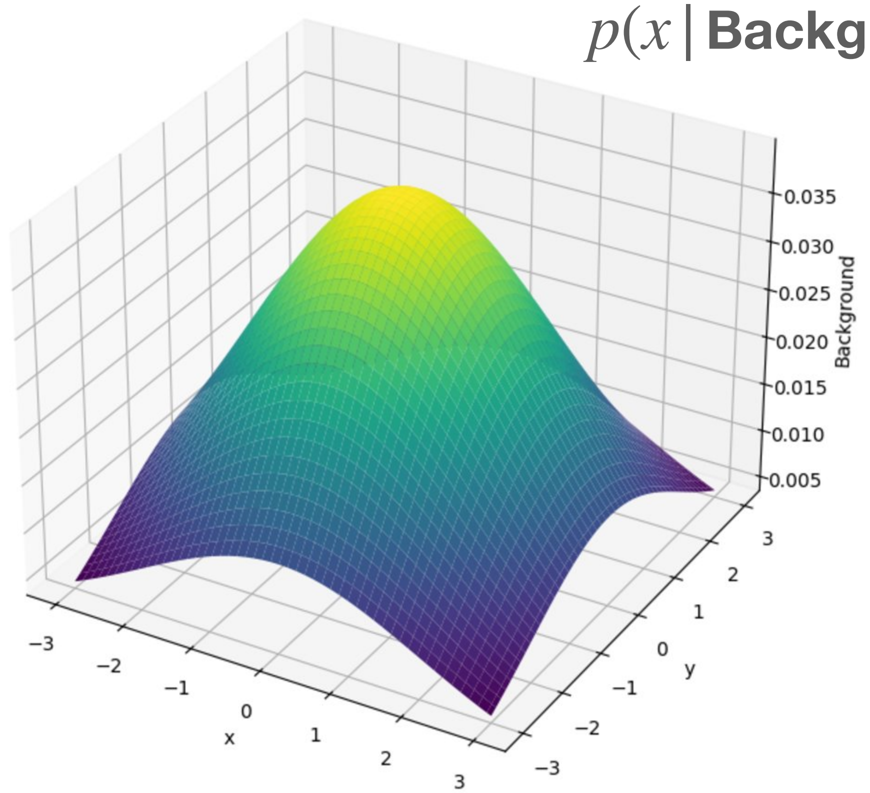
Use of "domain" knowledge to build summary statistic

Building "Optimal" Observables

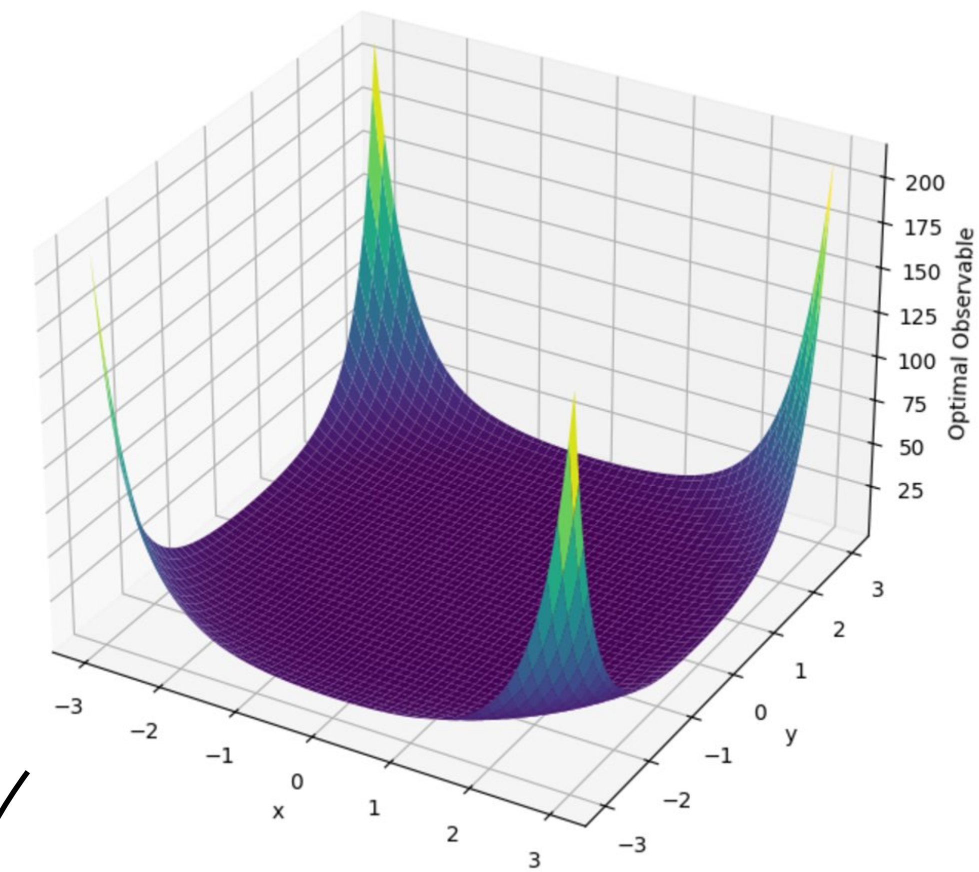
$p(x \mid \textbf{Signal})$



$p(x \mid \textbf{Background})$



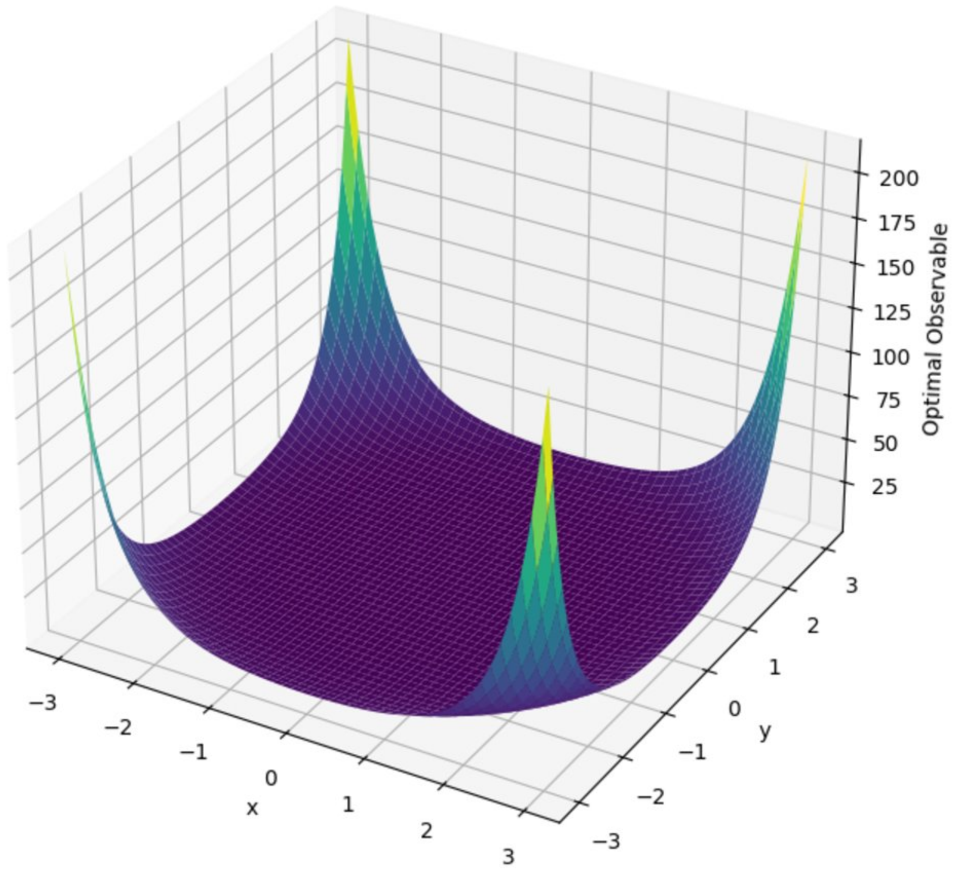
$$\frac{p(x \mid \textbf{Signal})}{p(x \mid \textbf{Background})}$$



Optimal for reducing Type-II errors (false negatives) and thus maximizing the power of the hypothesis test (inspired by Neyman-Pearson lemma)

Building "Optimal" Observables

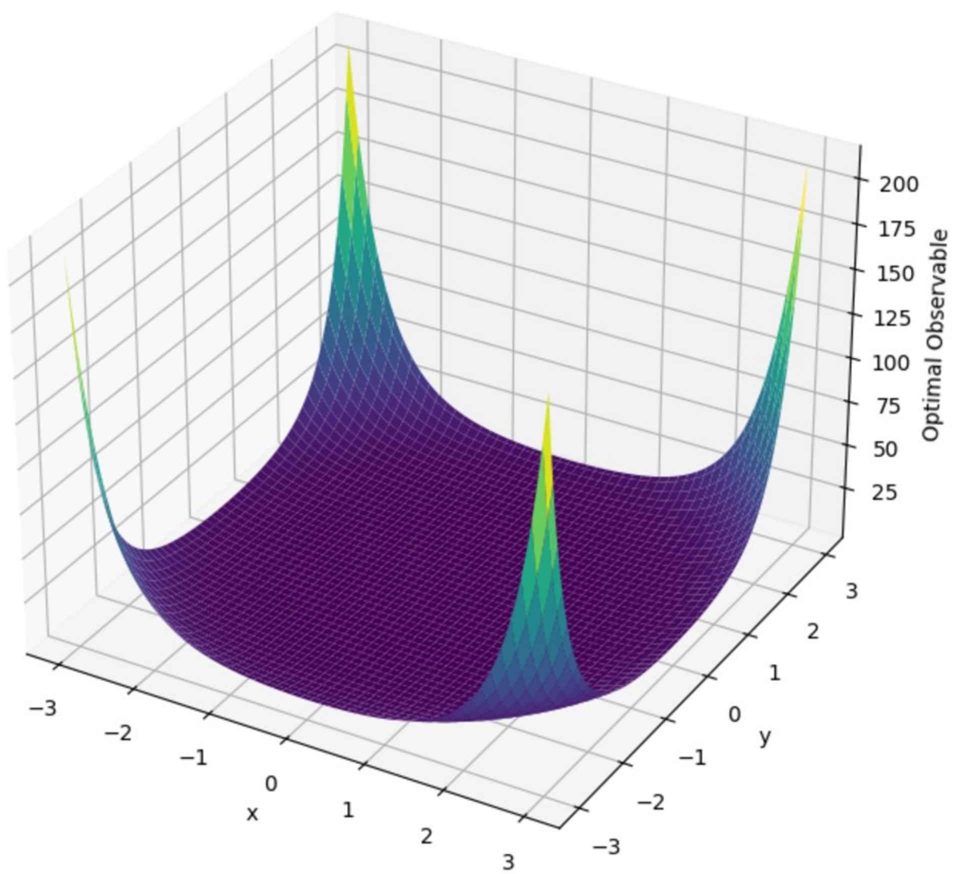
$$\frac{p(x\,|\,\textbf{Signal})}{p(x\,|\,\textbf{Background})}$$



How to build this optimal observable in the presence of multi-dimensional input feature space x?

Building "Optimal" Observables

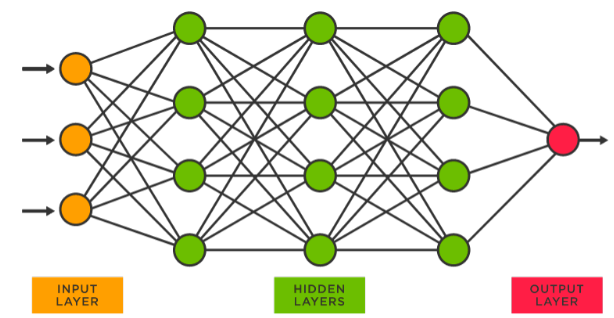
$$\frac{p(x\,|\,\textbf{Signal})}{p(x\,|\,\textbf{Background})}$$



Simulated samples form the forward model

$x \sim p(x\,|\,\textsf{S})$

$x \sim p(x\,|\,\textsf{B})$



$$\frac{p(x\,|\,\textsf{S})}{p(x\,|\,\textsf{B})}$$

How to build this optimal observable in the presence of multi-dimensional input feature space x?

Neural Networks, able to handle high-dimensional information efficiently, can serve as the optimal observable "surrogates" -> Their output to be used as a sufficient summary statistic.

Limitations

- The dimensional reduction of the full final state information to a low-dimensional summary statistic $x \rightarrow u(x)$ can result in a loss of information.

- The summary statistic $u(x)$ is independent of the parameters θ corresponding to hypothesis being tested. $u(x)$ can be optimal locally, but not globally.

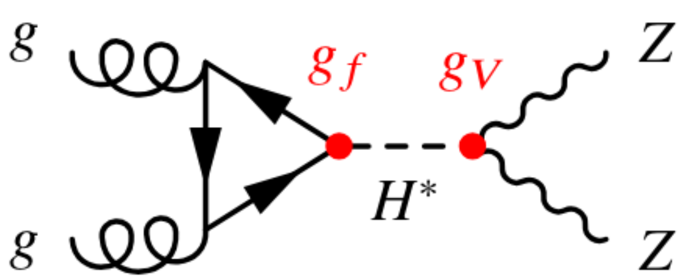
- All the events in a bin have the same probability. Information from rare possibly signal-like events can be lost in this approximation.

$$p(x_i | \theta) = \boxed{\frac{\nu_{I(x_i|u)}(\theta)}{\nu(\theta)}}$$

← Fraction of events in the bin of x_i
(Same likelihood for every event in the bin)

Optimal Measurements at the LHC

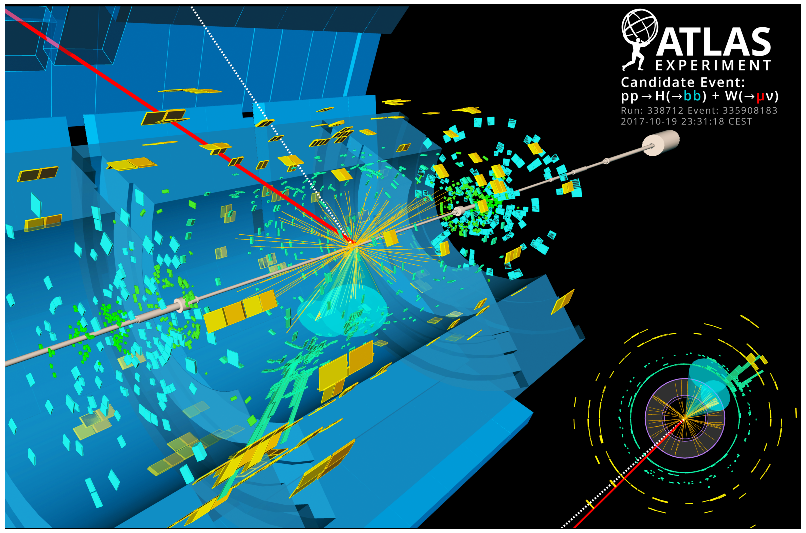
$$x \sim p_S(x\,|\,\theta)\,,p_B(x\,|\,\theta)$$



Parton Shower

Hadronization

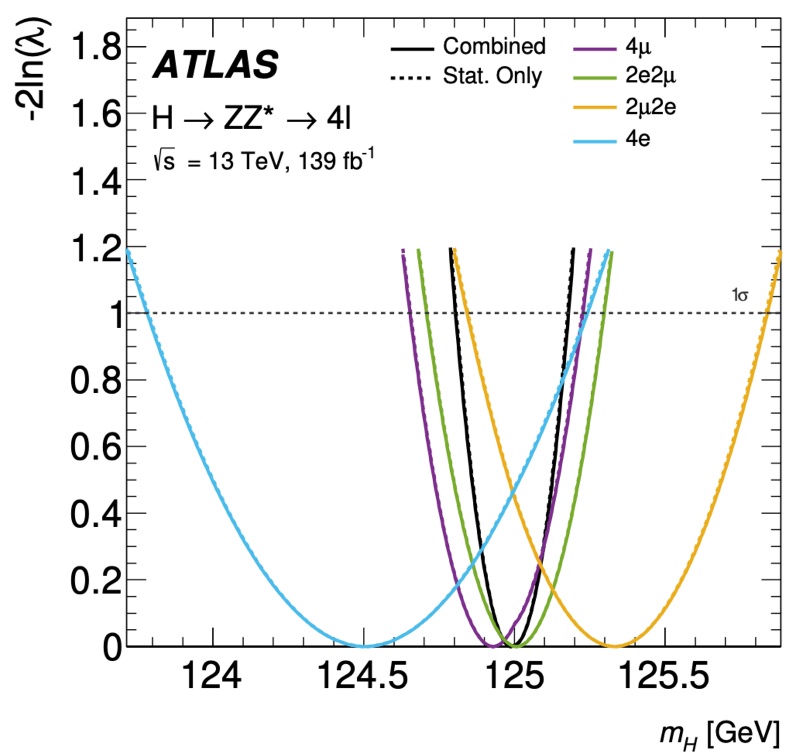
Detector
Reconstruction



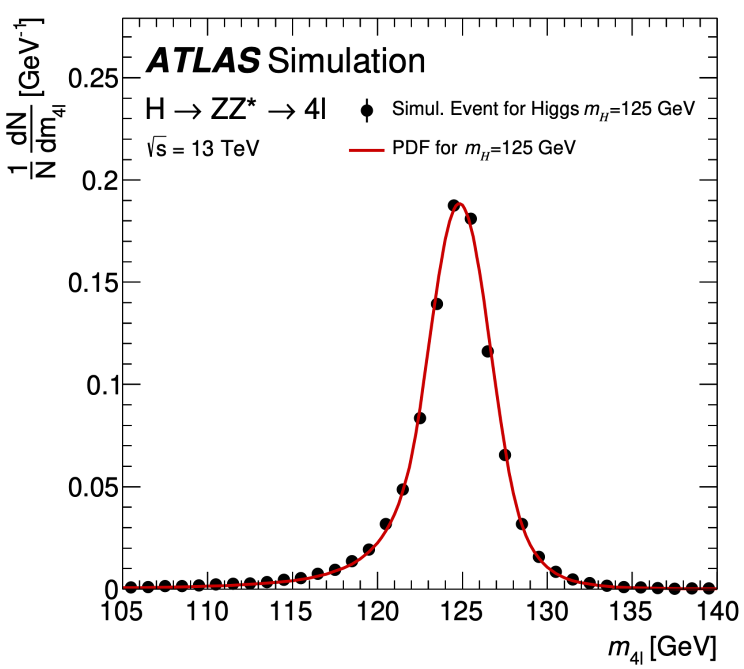
Inference

$$\lambda(m_H) = \frac{\mathcal{L}(m_H, \hat{\hat{\theta}}(m_H))}{\mathcal{L}(\hat{m}_H, \hat{\theta})}$$

Event-by-event parameterized likelihood ratios



Profile Negative Log-Likelihood Fit

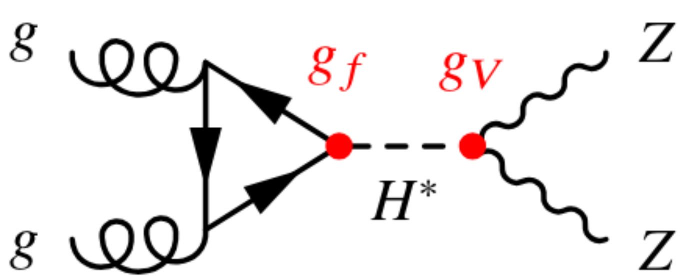


Analytical "surrogate" model

2207.00320

Neural Simulation-Based Inference

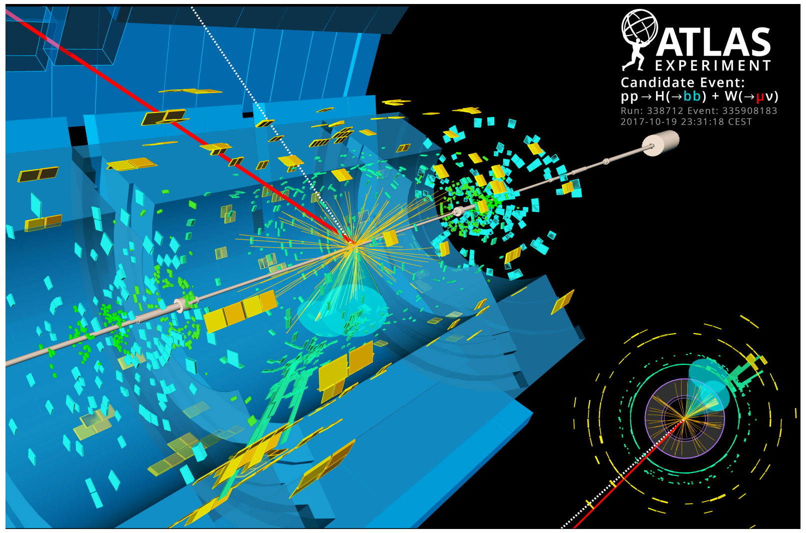
$$x \sim p_S(x \mid \theta)\,, p_B(x \mid \theta)$$



Parton Shower

Hadronization

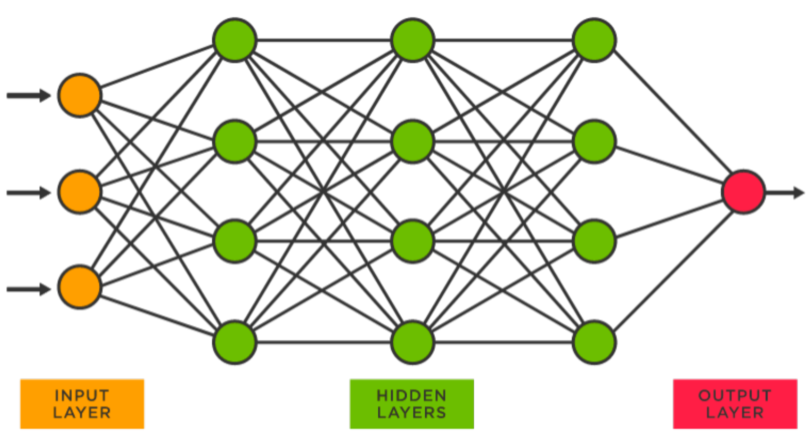
Detector Reconstruction



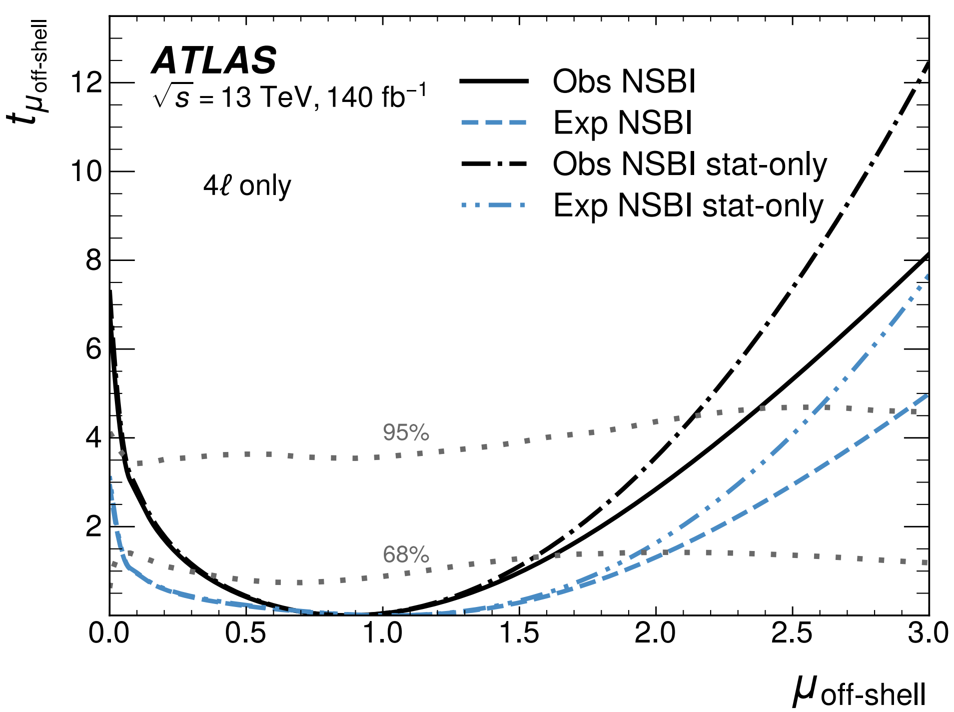
Inference

$$-2 \cdot \sum_{i \in events} \log \frac{p(x_i \mid \mu)}{p(x_i \mid \hat{\mu})}$$

Event-by-event parameterized likelihood ratios



NN Surrogate Model

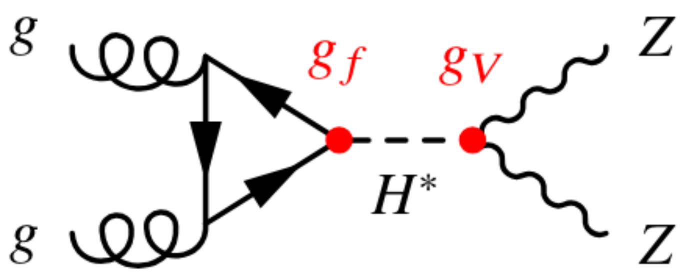


Profile Negative Log-Likelihood Fit

2412.01548

Neural Simulation-Based Inference

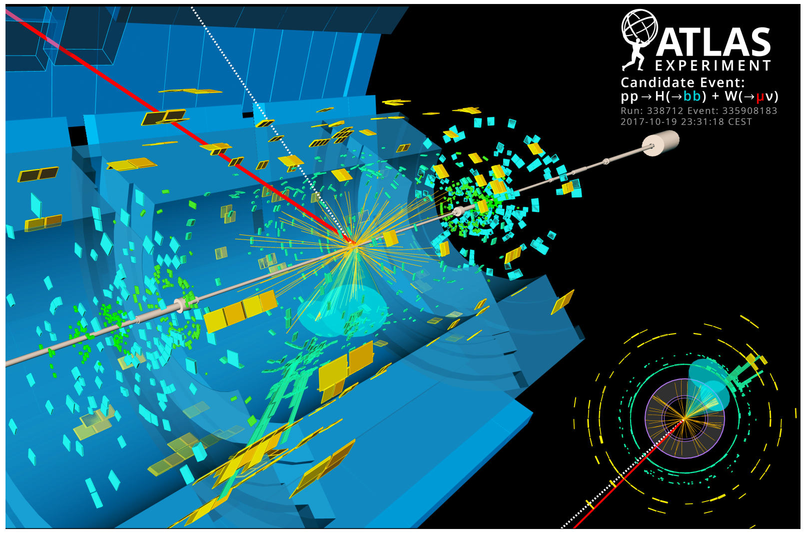
$$x \sim p_S(x\,|\,\theta)\,, p_B(x\,|\,\theta)$$



Parton Shower

Hadronization

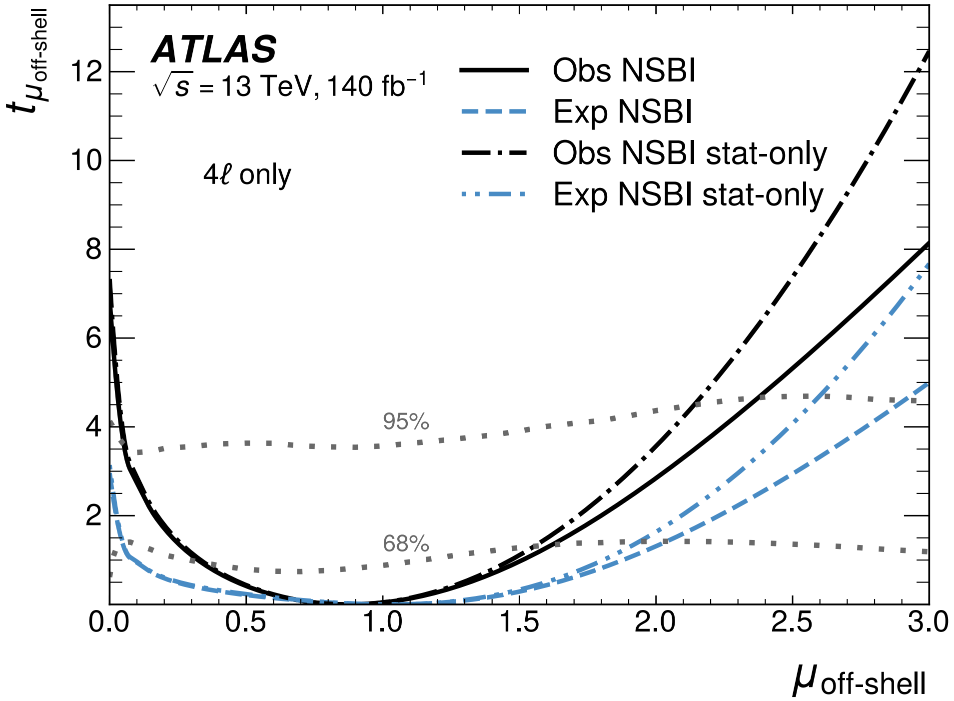
Detector Reconstruction



Inference

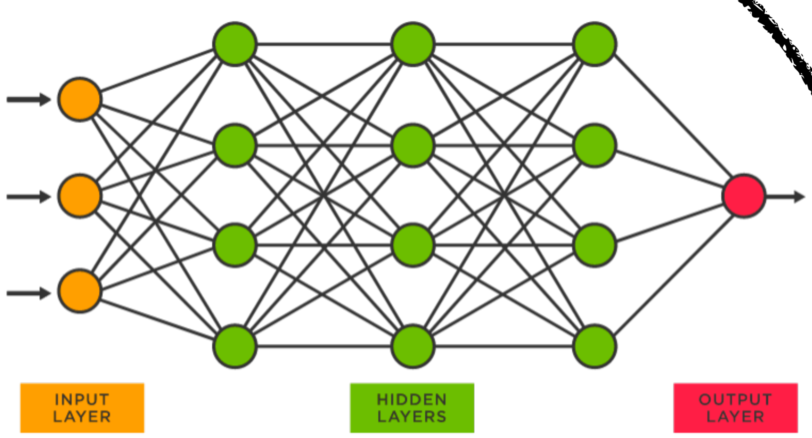
$$-2 \cdot \sum_{i \in events} \log \frac{p(x_i\,|\,\mu)}{p(x_i\,|\,\hat{\mu})}$$

Event-by-event parameterized likelihood ratios



Profile Negative Log-Likelihood Fit

2412.01548



NN Surrogate Model

Need unbiased estimates of likelihoods/likelihood ratios. Very difficult!

Papers

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

[hep-ex] 2 Dec 2024

2412.01548

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

-ex] 2 Dec 2024

2412.01600

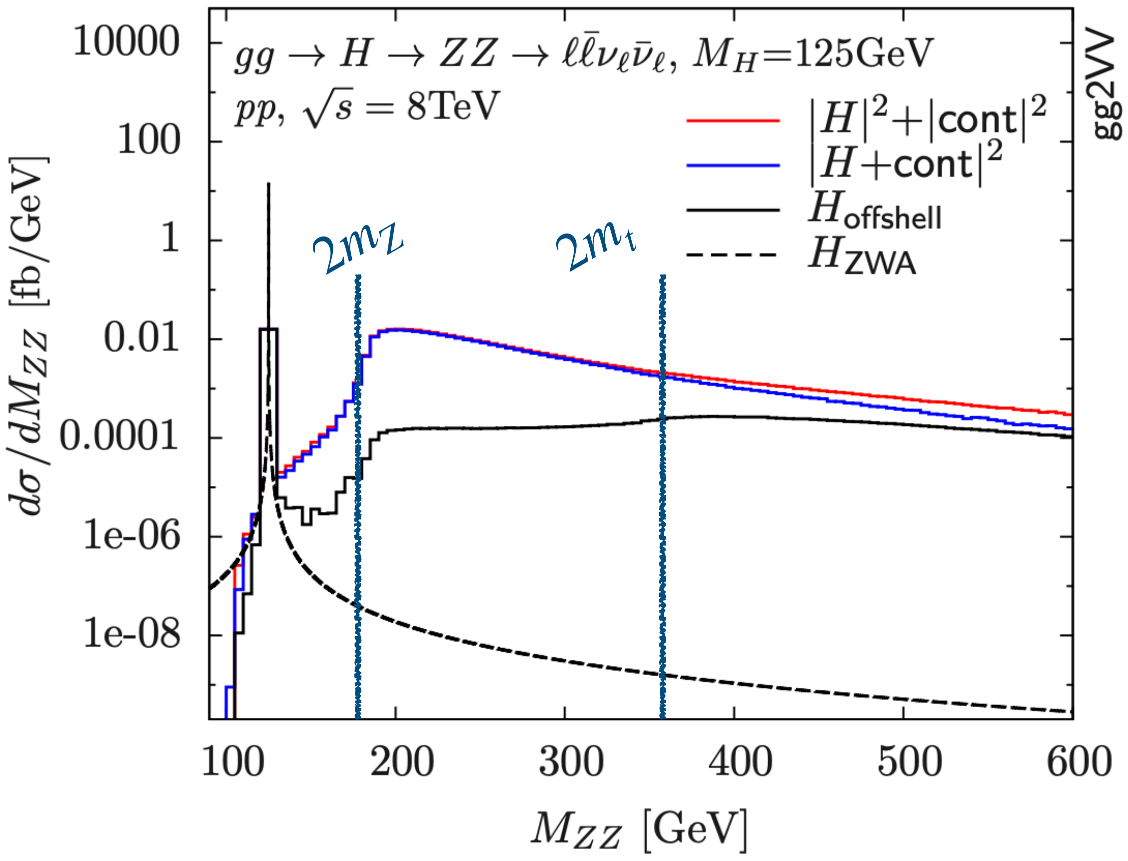
ATLAS Data Analysis using a Parallel Workflow on Distributed Cloud-based Services with GPUs

Jay Sandesara[1,*], *Rafael* Coelho Lopes de Sa[1], *Verena* Martinez Outschoorn[1], *Fernando* Barreiro Megino[2], *Johannes* Elmsheuser[3], and *Alexei* Klimentov[3] on behalf of the ATLAS Computing Activity

[1]University of Massachusetts Amherst, Amherst, MA, USA
[2]University of Texas at Arlington, Arlington, TX, USA
[3]Brookhaven National Laboratory, Upton, NY, USA

ATL-SOFT-PROC-2023-023

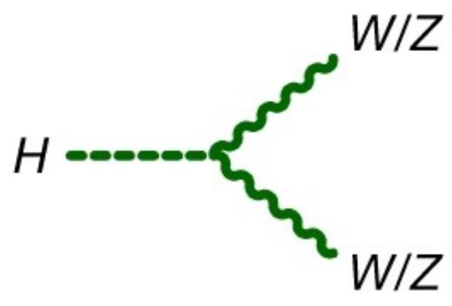
The Off-shell Higgs boson



$$\frac{d\sigma^{H\to VV}_{\text{off-shell}}}{dm^2_{VV}} \propto \frac{g^2_{\text{prod}}(\hat{s})\, g^2_{\text{decay}}(\hat{s})}{(m^2_{VV} - m^2_H)^2}$$

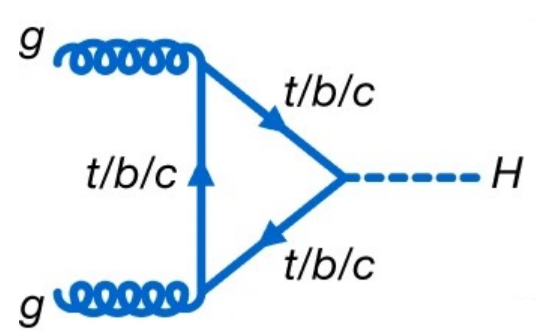
The down-scaling is compensated by the numerator terms in $m_{ZZ} \gtrsim 180$ GeV phase space

$$g_{\text{HZZ}}(\hat{s}) \sim \hat{s}^2 \quad \text{for} \quad \hat{s} \gtrsim 2m_Z$$



The on-shell Z-bosons become energetically accessible

$$g_{t\bar{t}H}(\hat{s}) \sim m_t\, \hat{s} \quad \text{for} \quad \hat{s} \gtrsim 2m_t$$



The on-shell top quarks become energetically accessible

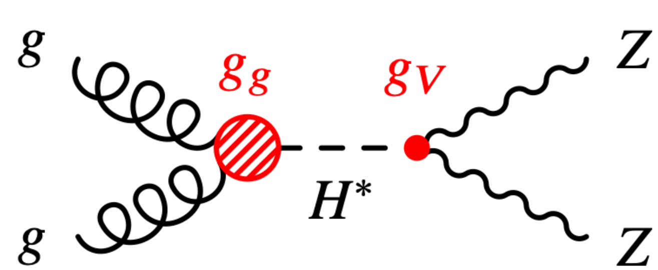
The off-shell Higgs boson

The probability model of the off-shell Higgs boson:

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

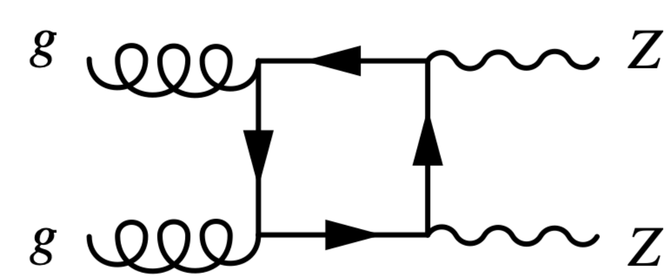
$p_S(x)$ $p_B(x)$

$p_I(x) = 2 \cdot Re \,\Big|$



ggF Signal

$\times$



ggF Background

$\Big|$

The off-shell Higgs boson

The probability model of the off-shell Higgs boson:

$$p(\mathbf{x}|\mu_{\text{off-shell}}) = \frac{1}{\nu(\mu_{\text{off-shell}})} \times$$

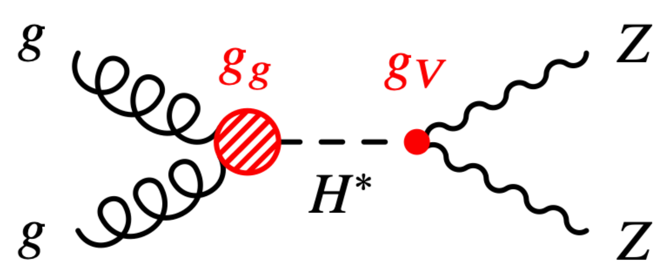
$$\Big[\mu_{\text{off-shell}}\nu_{\text{S}}^{\text{ggF}}p_{\text{S}}^{\text{ggF}}(\mathbf{x}) + \sqrt{\mu_{\text{off-shell}}}\nu_{\text{I}}^{\text{ggF}}p_{\text{I}}^{\text{ggF}}(\mathbf{x}) + \nu_{\text{B}}^{\text{ggF}}p_{\text{B}}^{\text{ggF}}(\mathbf{x}) +$$

$$\mu_{\text{off-shell}}\nu_{\text{S}}^{\text{EW}}p_{\text{S}}^{\text{EW}}(\mathbf{x}) + \sqrt{\mu_{\text{off-shell}}}\nu_{\text{I}}^{\text{EW}}p_{\text{I}}^{\text{EW}}(\mathbf{x}) + \nu_{\text{B}}^{\text{EW}}p_{\text{B}}^{\text{EW}}(\mathbf{x}) + \boxed{\nu_{\text{NI}}p_{\text{NI}}(\mathbf{x})}\Big]$$

NI → Non-interfering background

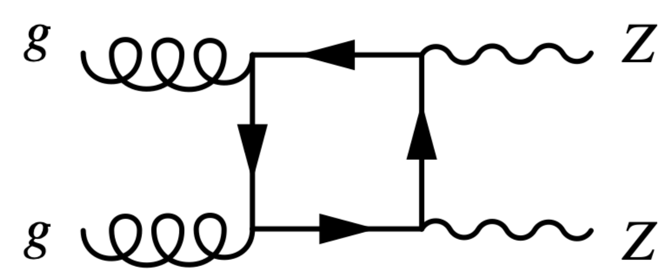
$p_S(x)$ $p_B(x)$

$p_I(x) = 2 \cdot Re \,\Big|$



ggF Signal

$\times$

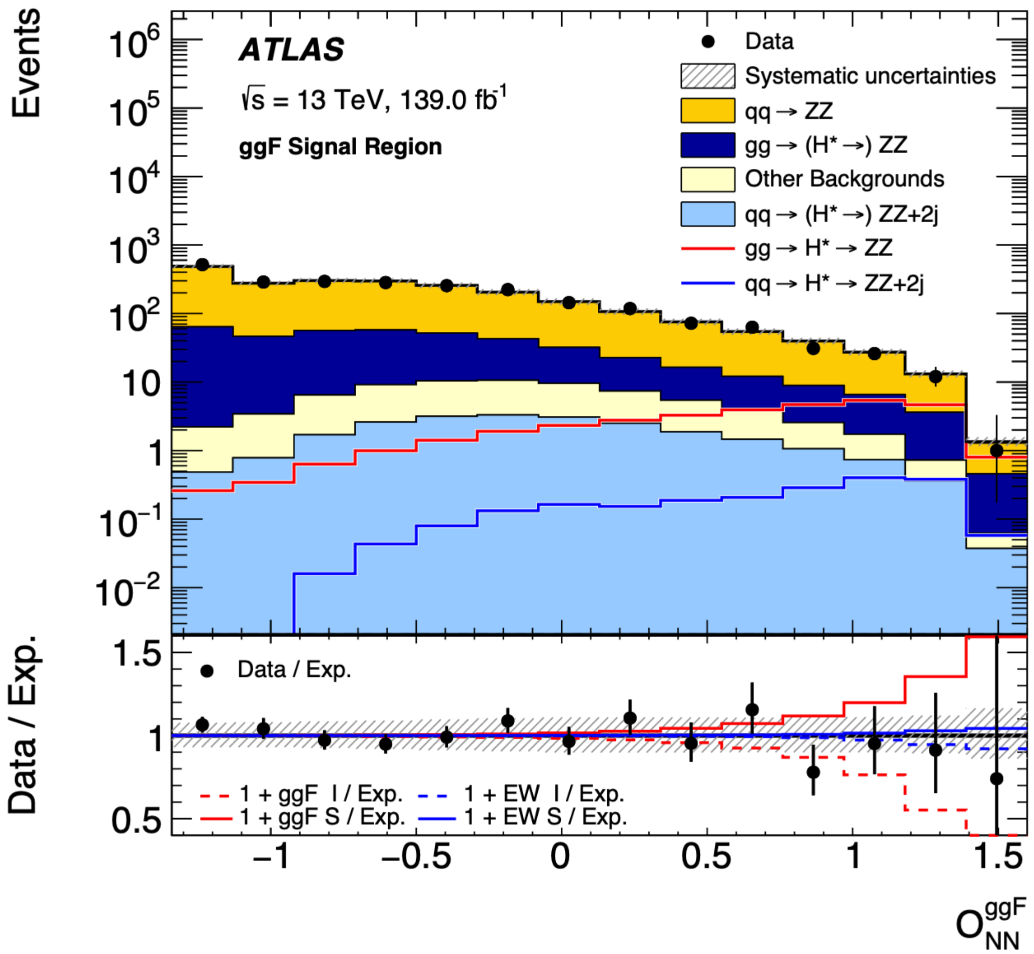


ggF Background

$\Big|$

Previous Measurement

Signal vs Background discriminant optimal ONLY when the probability model can be made linear in POI μ using a smooth transformation $f(\mu) = \mu$



$$\frac{p(x\,|\,\mu)}{\sum p_B(x)} = \mu \cdot \frac{\sum p_S(x)}{\sum p_B(x)} + \frac{\sum p_B(x)}{\sum p_B(x)}$$

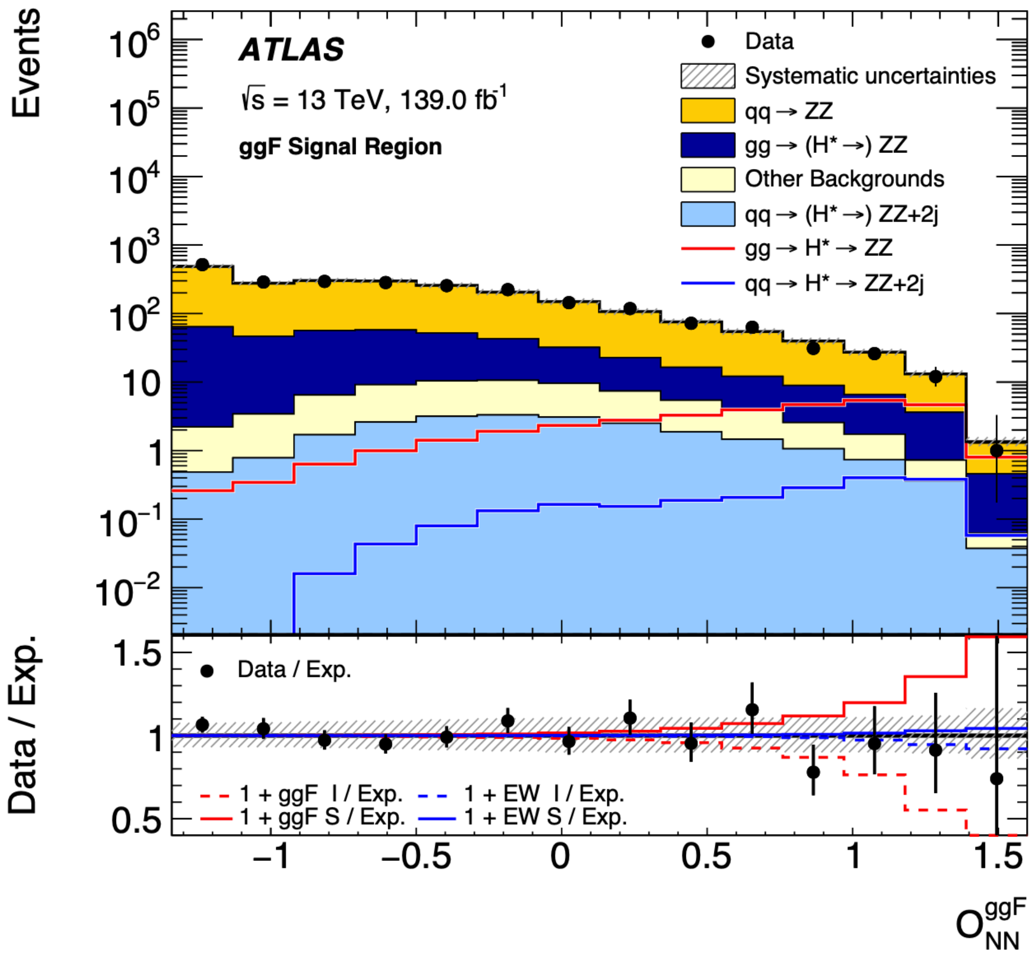
Neyman pearson lemma ✓

i.e. optimal across the parameter range (consider point-by-point testing)

$$O_{NN} = \log \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

NN-based Signal vs Background classification

Previous Measurement



$$O_{NN} = \log \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

NN-based Signal vs Background classification

Signal vs Background discriminant optimal ONLY when the probability model can be made linear in POI μ using a smooth transformation $f(\mu) = \mu$

$$\frac{p(x|\mu)}{\sum p_B(x)} = \mu \cdot \frac{\sum p_S(x)}{\sum p_B(x)} + \frac{\sum p_B(x)}{\sum p_B(x)}$$

Neyman pearson lemma ✔

What if the probability model is non-linear in POI?

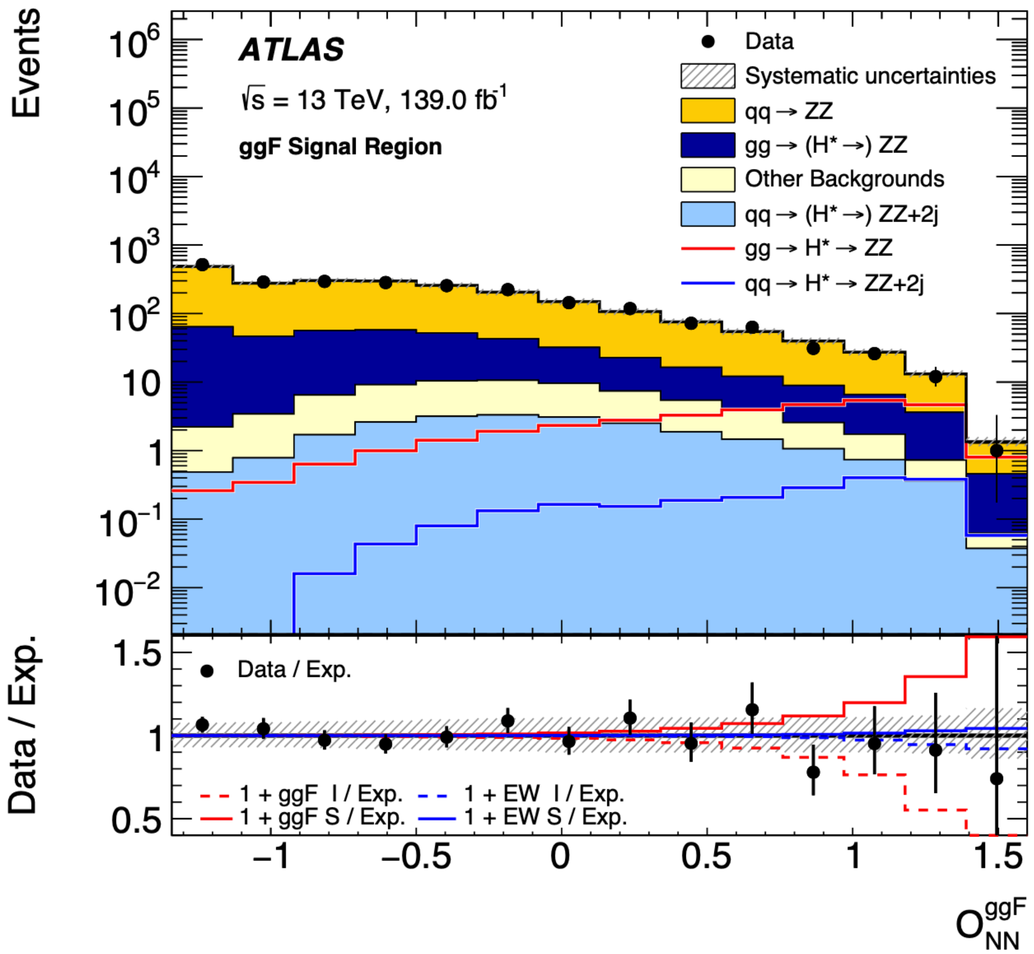
E.g.: interference effects of off-shell Higgs boson production.

$$\frac{p(x|\mu)}{\sum p_B(x)} = \mu \cdot \frac{\sum p_S(x)}{\sum p_B(x)} + \sqrt{\mu} \cdot \frac{\sum p_I(x)}{\sum p_B(x)} + \frac{\sum p_B(x)}{\sum p_B(x)}$$

Neyman pearson lemma ✘

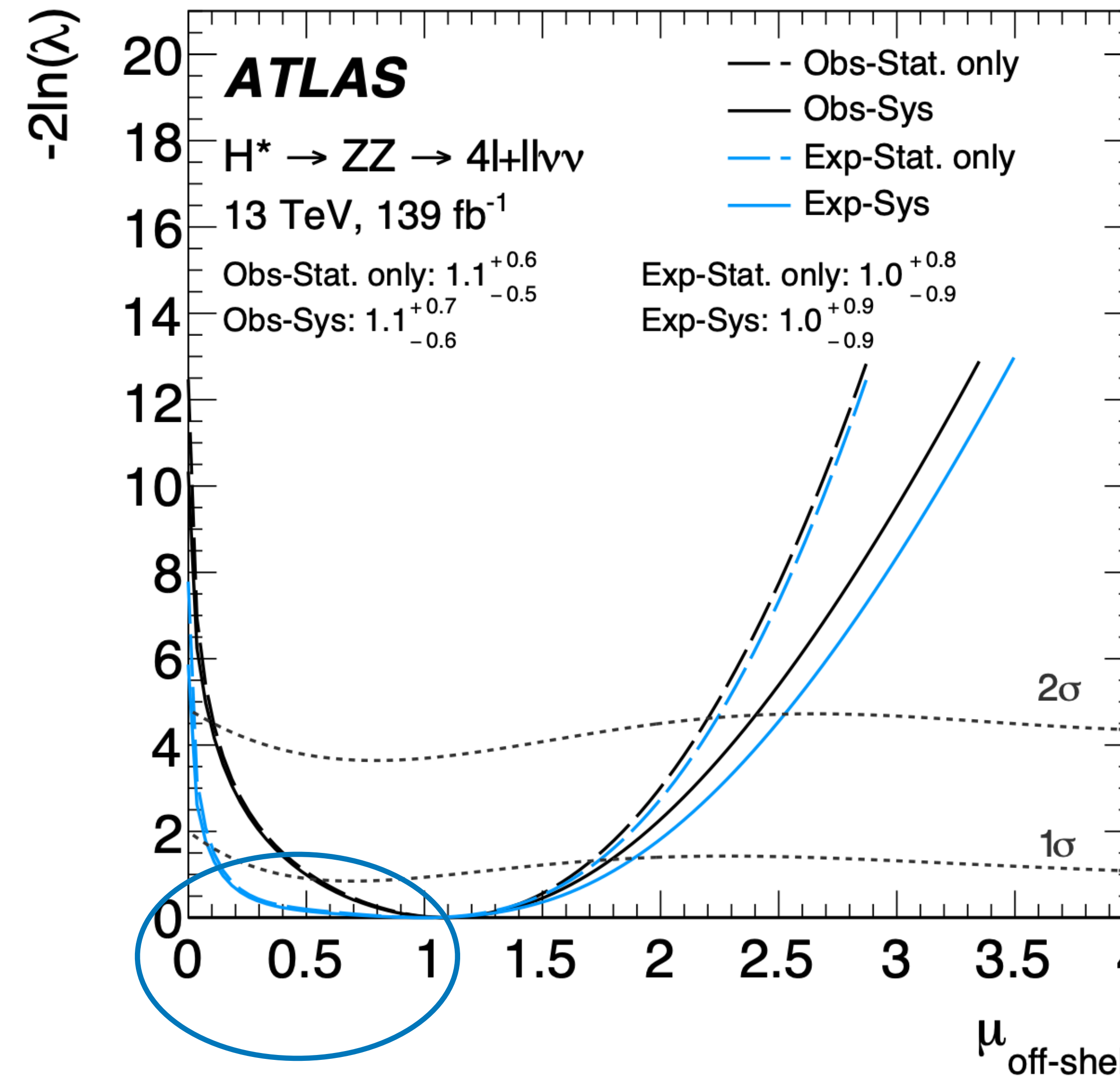
What about optimally discriminating interference from background for different μ- values?

Previous Measurement



$$O_{NN} = \log \frac{p_S}{p_B + 0.1 \cdot p_{NI}}$$

NN-based Signal vs Background classification



Flat NLL region implies sub-optimality in regions with $\sqrt{\mu} \cdot p_I \gg \mu \cdot p_S$

New Measurement

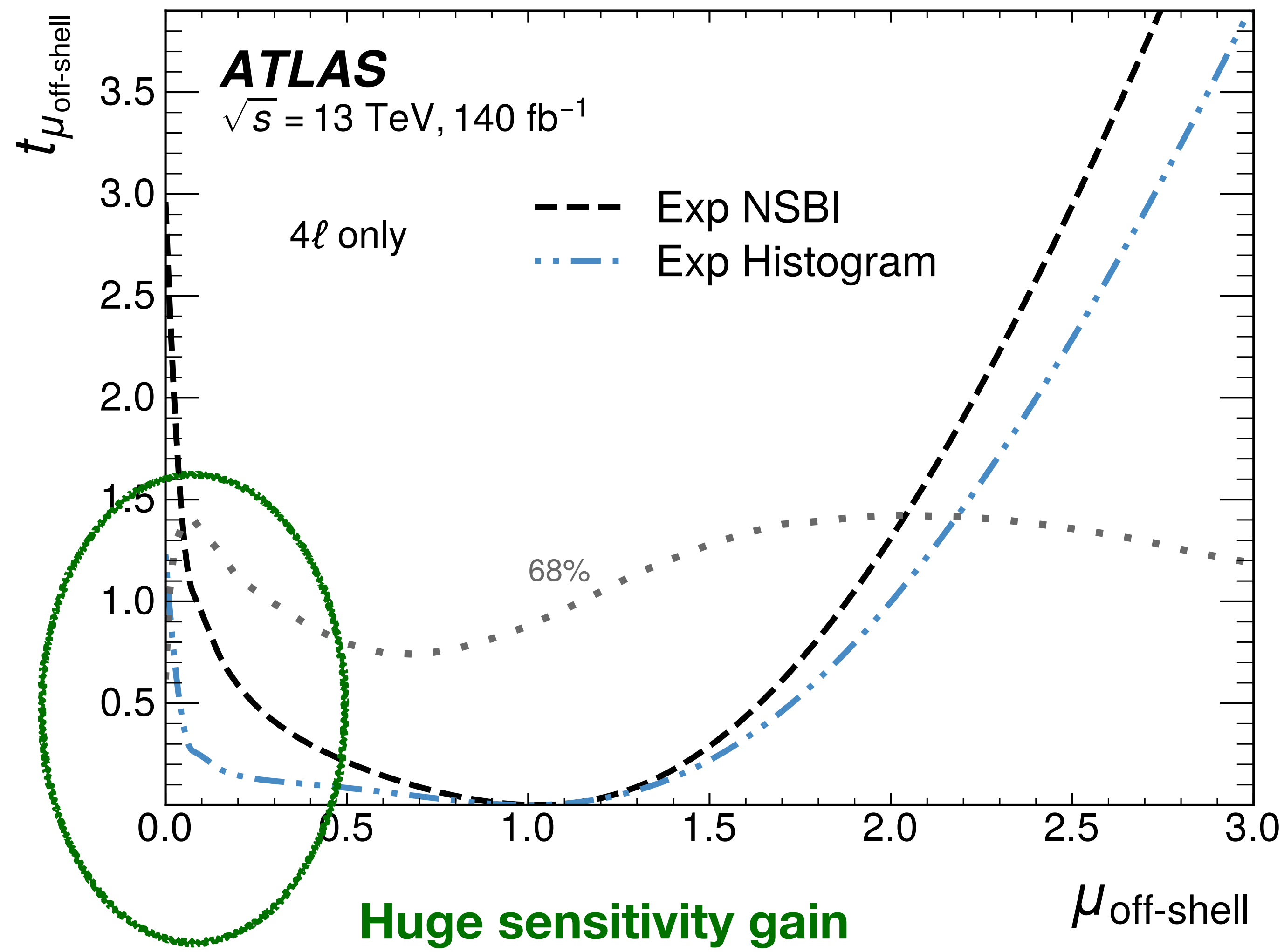
Carefully trained **parameterized per-event density ratios** are now used to build the test statistic:

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu)}{p(x_i \,|\, \hat{\mu})}$$

No fixed observable - optimality throughout μ space.

Neyman pearson lemma ✓

Additional sensitivity from unbinned nature (no Poisson fits) ✓



Huge sensitivity gain expected in interference rich regions

$$\sqrt{\mu} \cdot p_I(x) \gg \mu \cdot p_S(x)$$

Note: we use the same pre-selections, Monte Carlo samples, background normalization, and systematic uncertainty model as the previously published analysis [link to paper for details]

New Measurement

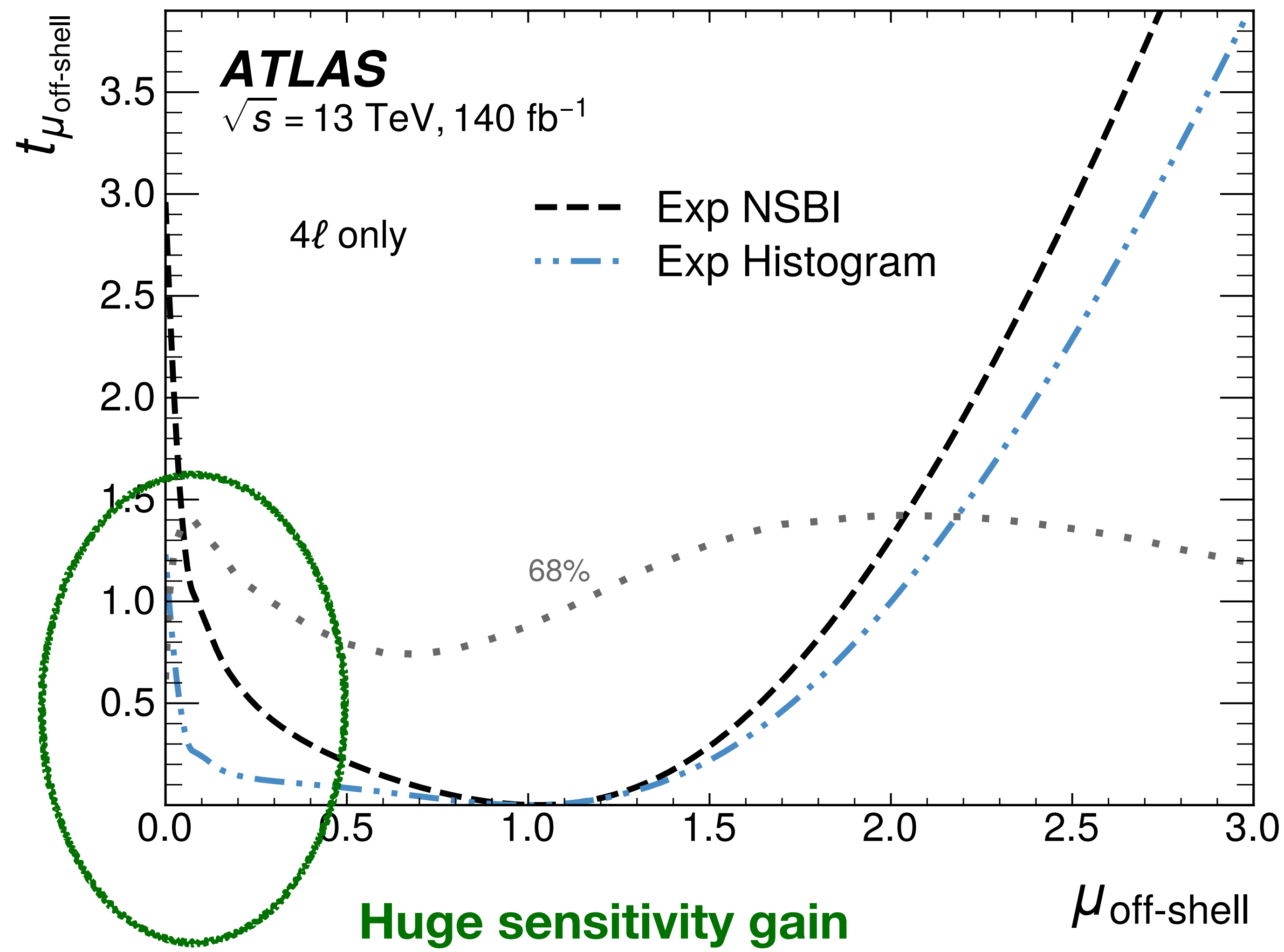
Carefully trained **parameterized per-event density ratios** are now used to build the test statistic:

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i | \mu)}{p(x_i | \hat{\mu})}$$

No fixed observable - optimality throughout μ space.

Neyman pearson lemma ✓

Additional sensitivity from unbinned nature (no Poisson fits) ✓



Huge sensitivity gain expected in interference rich regions

$$\sqrt{\mu} \cdot p_I(x) \gg \mu \cdot p_S(x)$$

Exploiting the known analytical formula - we break down the parameterized ratio into simpler parts:

$$\frac{p(x|\mu)}{p(x|\hat{\mu})} = \frac{p(x|\mu)/p_{ref}(x)}{p(x|\hat{\mu})/p_{ref}(x)} \longrightarrow \frac{p(x|\mu)}{p_{ref}(x)} = \mu \cdot \frac{p_S(x)}{p_{ref}(x)} + \sqrt{\mu} \cdot \frac{p_I(x)}{p_{ref}(x)} + \frac{p_B(x)}{p_{ref}(x)} + \frac{p_{NI}(x)}{p_{ref}(x)}$$

p_{ref} is a carefully chosen **parameter-independent** hypothesis

We learn everything, including interference effects

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \,|\, \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \,|\, \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

Extended Poisson term

Sum of event-by-event log-likelihood ratios

Constraint terms

$N_{obs} \rightarrow$ total observed events

$p_{subs} \rightarrow$ likelihood from subsidiary measurements of the nuisance parameters

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs}\,|\,\mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs}\,|\,\hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i\,|\,\mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}(x_i)}}{p(x_i\,|\,\hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}(x_i)}} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

sum over processes $c = S,\, B,$ **etc.**

parameter-independent ratio

$$\frac{p(x_i\,|\,\mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i\,|\,\alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Parameterized per-event ratios

Factorized **nuisance parameter** α-dependence:

$$g_c(x\,|\,\alpha) = \prod_m \frac{p_c(x\,|\,\alpha_m)}{p_c(x)}$$

Parameter dependancies are factorized out (see slide 31)

Overview: Neural Simulation-Based Inference

Full test statistic function with nuisance parameters α:

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \,|\, \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \,|\, \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

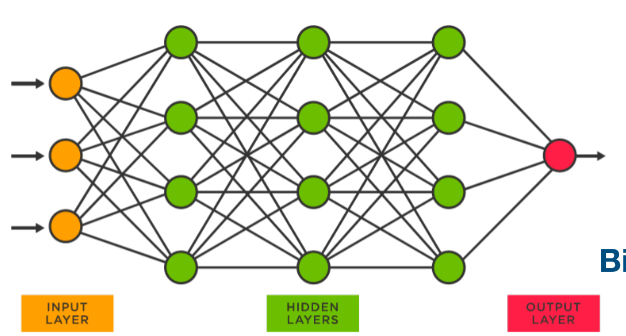
sum over processes
$c = S, B,$ **etc.**

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i \,|\, \alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

$x \sim p_c$
$S = 1$

$x \sim p_{ref}$
$S = 0$

Two hypothesis:
p_c and p_{ref}



Classification NN

$$\underset{\omega}{\text{argmin}}\ \text{L} \longrightarrow \hat{s}(x) = \frac{p_c}{p_{ref} + p_c}(x) \longrightarrow \frac{p_c}{p_{ref}}(x) = \frac{\hat{s}(x)}{1.0 - \hat{s}(x)}$$

Binary Cross-Entropy loss

"Likelihood ratio trick"

Many examples in ATLAS - HH4b background estimation, Omnifold, etc.

Probability Calibration Test

The NN ratios are meticulously trained to be **true representations of the density ratios**

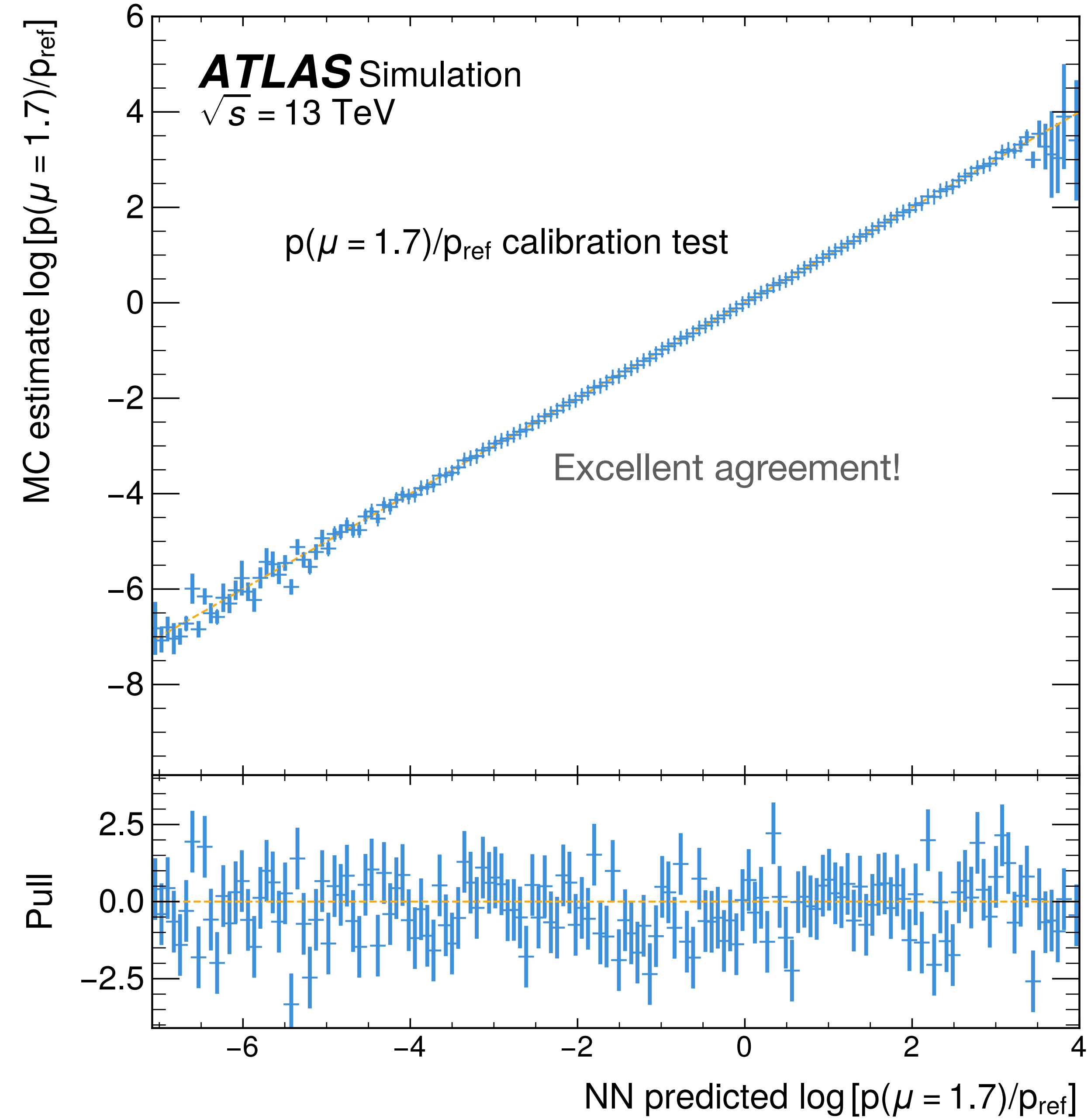
Diagnostic:

Does the NN output correspond to real probabilities?

$$\log \frac{p(x|\mu)}{p_{ref}(x)}(x) \leftrightarrow \log \frac{N^{bin}(x|\mu)}{N^{bin}_{ref}(x)} \ ?$$

MC estimate

$$\log \frac{N^{bin}(x|\mu)}{N^{bin}_{ref}(x)}$$



NN predicted $\log \frac{p(x|\mu)}{p_{ref}(x)}(x)$

The tests are performed for the full range of μ being scanned over to **ensure robust statistical interpretation of the final test statistic.**

Bias Test with DR reweighting

Do the ratios capture the full un-biased dependence of the multi-dimensional feature space x ?

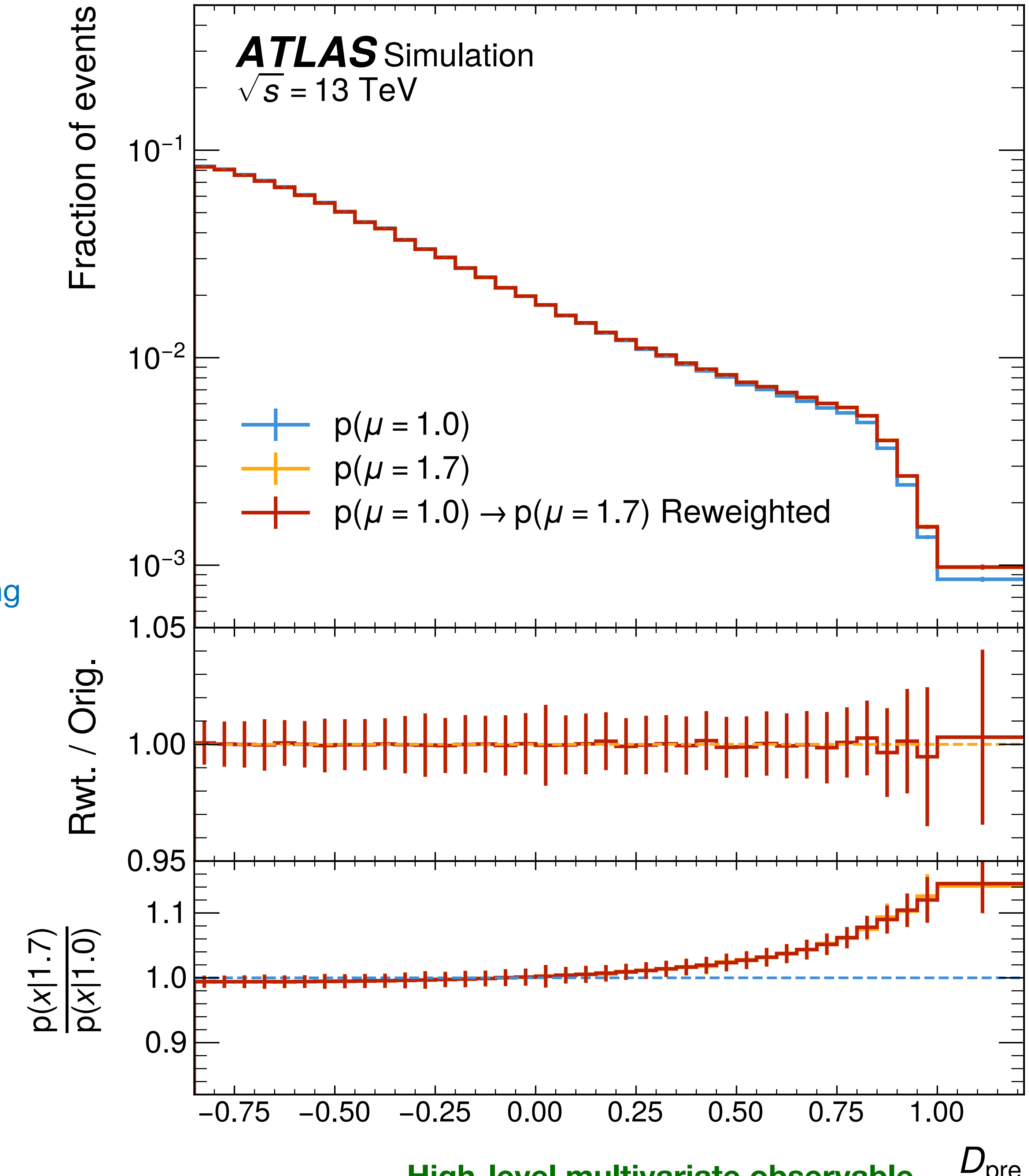
To test the modelling of the likelihood ratios, we can use 1D reweighting tests:

$$\boxed{\frac{p(x|\mu)}{p(x|\mu=1)}} \times p(x|\mu=1) \sim p(x|\mu)$$

NN prediction

MC sample at $\mu = 1$

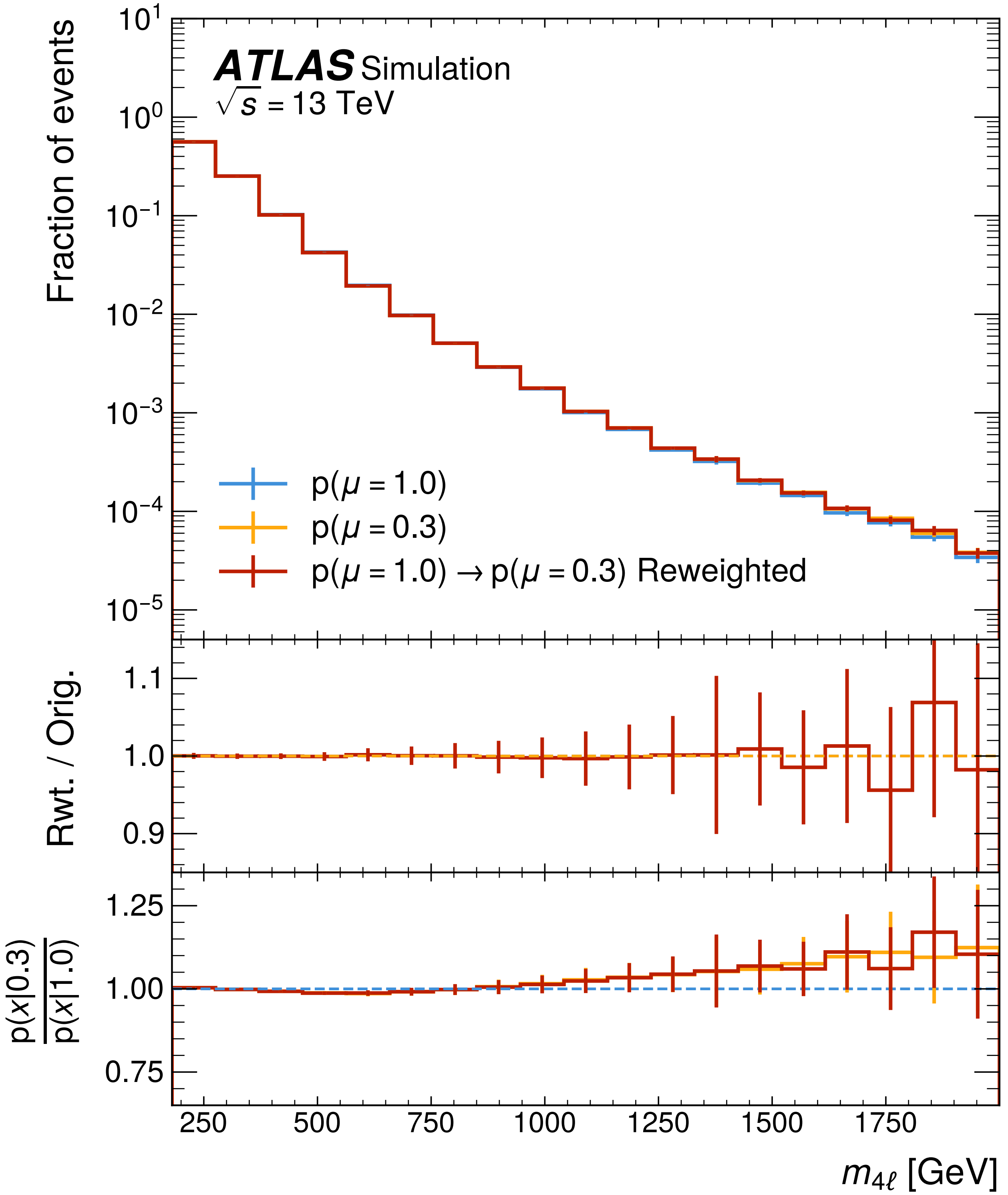
NN reweighted sample resembling MC at μ

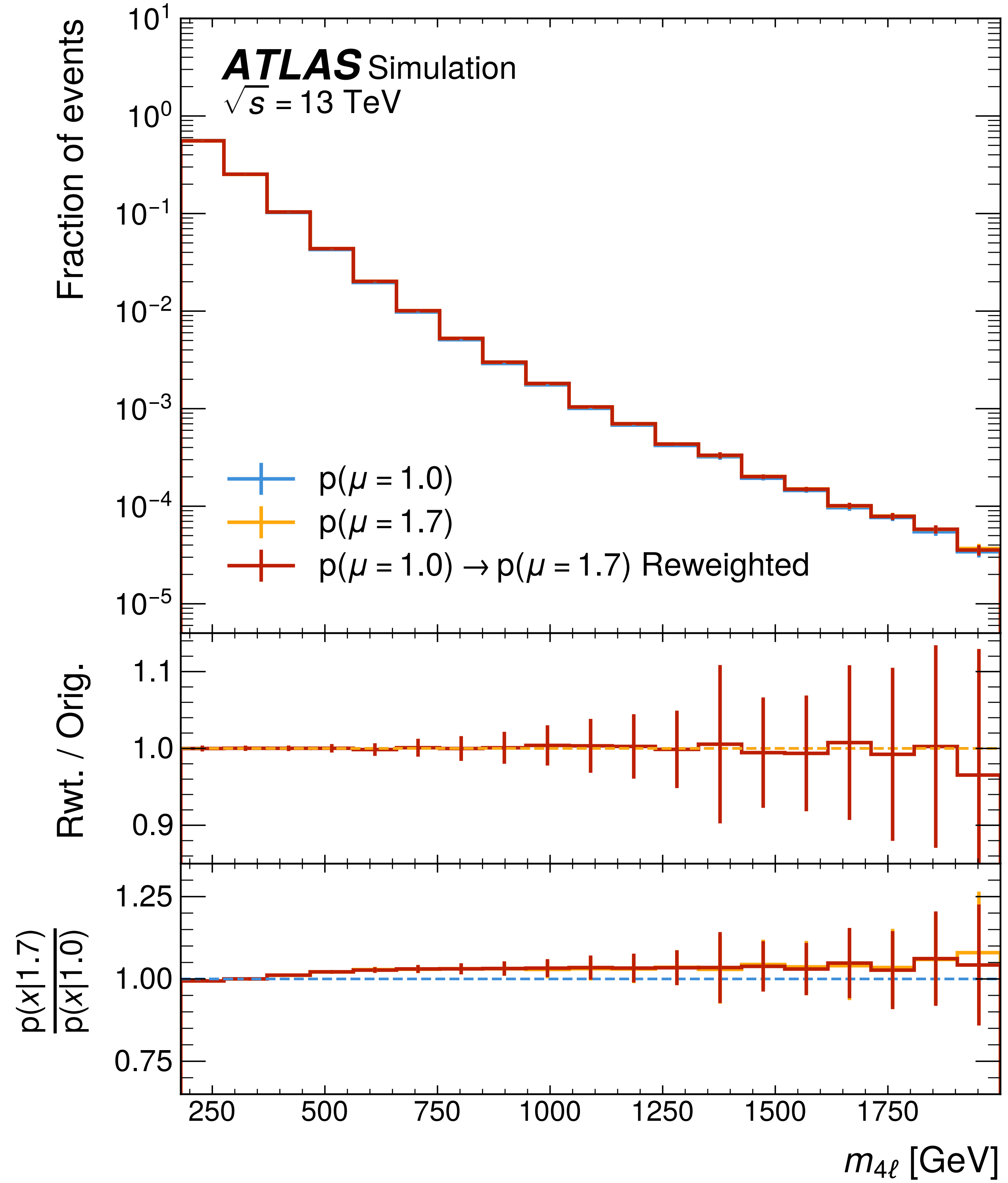


The kinematic dependence of the original $\mu = 1.7$ sample is perfectly captured by the NN reweighted $\mu = 1.7$ sample

High-level multivariate observable

Bias Test with DR reweighting

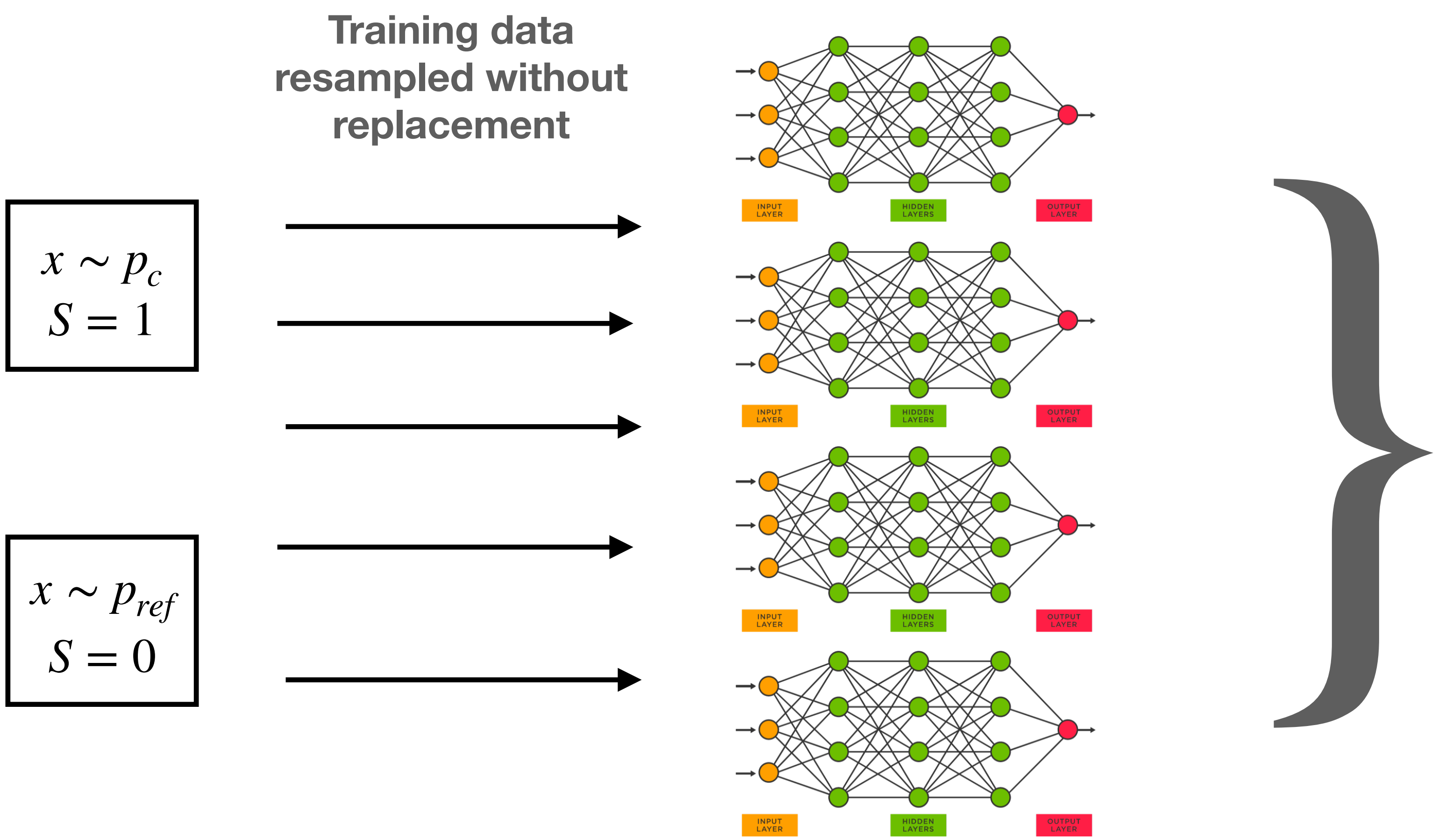




Challenges: Density Ratio Estimation

Challenge: The best fit value from a profile likelihood fit $\hat{\mu}$ with a single NN per p_c/p_{ref} is biased.

Solution: An ensemble of $O(100)$ or more NNs were trained to be robust against this bias.



$$\frac{p_c}{p_{ref}}(x) = \frac{1}{N_{ens}} \sum_{a}^{N_{ens}} \left[\frac{p_c}{p_{ref}}(x) \right]_a$$

Surrogate model from an ensemble of trained NNs

By building an ensemble of NNs per p_c/p_{ref} we become **robust against the bias** in the fit value:

$$\hat{\mu} \rightarrow \mu_{truth}$$

Uncertainty Parameterization

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c} \sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **yield** α-dependence:

$$G_c(\alpha) = \prod_k \frac{\nu_c(\alpha_k)}{\nu_c}$$

with $\nu_c(\alpha_k)/\nu_c$ estimated using **analytic interpolation techniques:**

Available from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

$$\frac{\nu_c(\alpha_k)}{\nu_c} = \begin{cases} \left(\boxed{\frac{\nu_c(\alpha_k^+)}{\nu_c}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \leq \alpha_k \leq 1, \\ \left(\boxed{\frac{\nu_c(\alpha_k^-)}{\nu_c}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Ref: HistFactory

Uncertainty Parameterization

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c} \sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **yield** α-dependence:

$$G_c(\alpha) = \prod_k \frac{\nu_c(\alpha_k)}{\nu_c}$$

with $\nu_c(\alpha_k)/\nu_c$ estimated using **analytic interpolation techniques:**

Available from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

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

Per-event analog of standard techniques

Factorized **per-event** α-dependence:

$$g_c(x|\alpha) = \prod_k \frac{p_c(x|\alpha_k)}{p_c(x)}$$

with $p_c(x|\alpha_k)/p_c(x)$ estimated using a **mix of NNs and analytic interpolation techniques:**

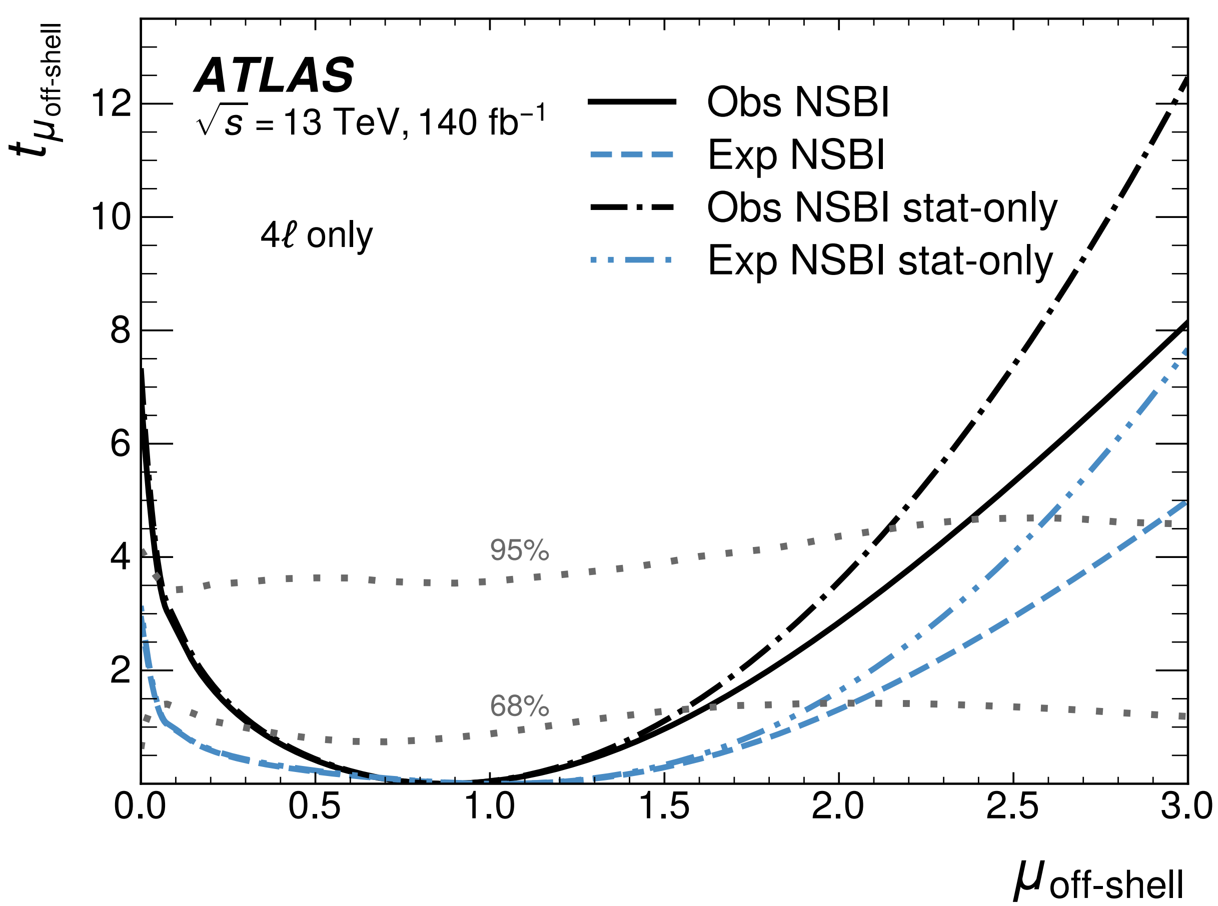
Density ratios trained using NNs from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

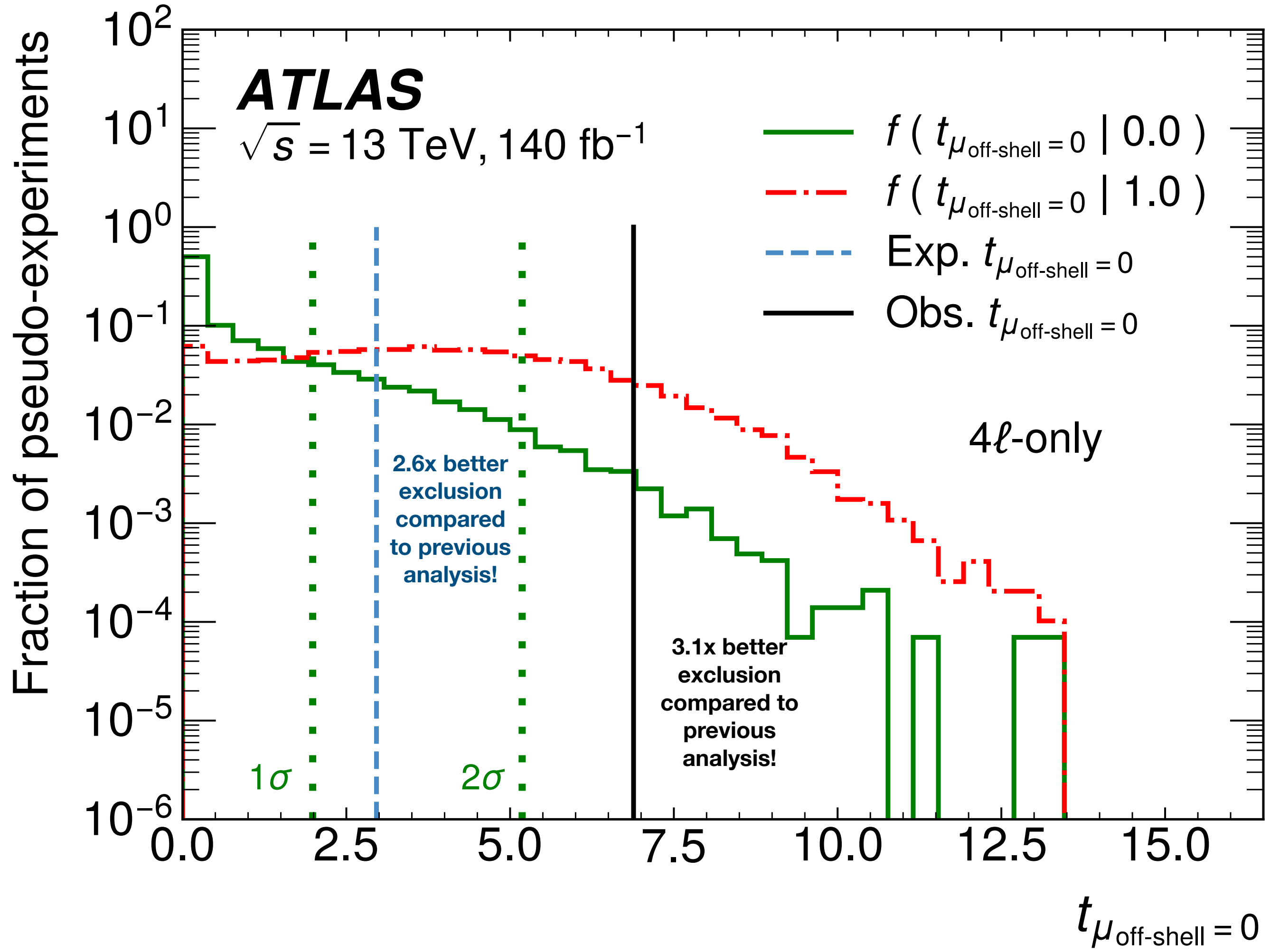
$$\frac{p_c(x|\alpha_k)}{p_c(x)} = \begin{cases} \left(\boxed{\dfrac{p_c(x|\alpha_k^+)}{p_c(x)}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1\ . \\ \left(\boxed{\dfrac{p_c(x|\alpha_k^-)}{p_c(x)}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Ref: HistFactory

Unblinded Results - Parameter scans

Having validated the parameterized density ratios we build the test statistic scan for $\mu_{offshell}$





Pseudo-experiments sampled using the newly developed techniques developed have been used to calculate the exact confidence intervals and background exclusion significance.

The Non-Linear Models

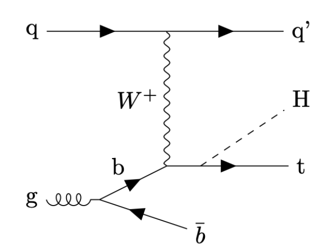
Not all LHC parameter measurements have a linear model. Quantum Intereference can play a major role in the production cross-section in several measurements:

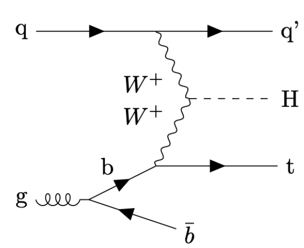
$$p(x\,|\,\mu) = \frac{1}{\nu(\mu)}\left[\mu \cdot \nu_S \cdot \sum p_S(x) + \sqrt{\mu} \cdot \nu_I \cdot \sum p_I(x) + \sum \nu_B \cdot p_B(x)\right]$$

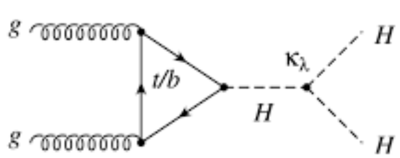
$I \rightarrow$ Signal-Background Interference Hypothesis

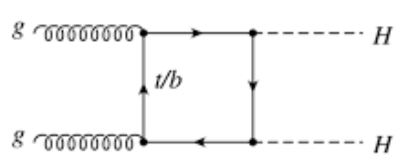
Examples of typical LHC analysis with large quantum interference include :

$pp \rightarrow tHq$



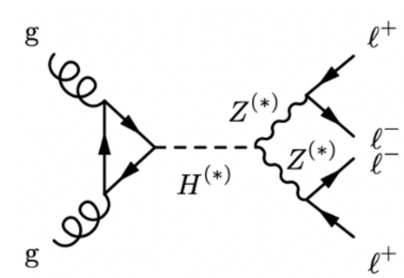


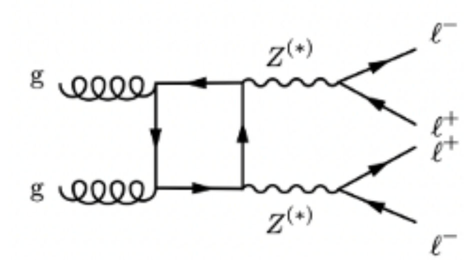




$pp \rightarrow HH$

$pp \rightarrow H^*$



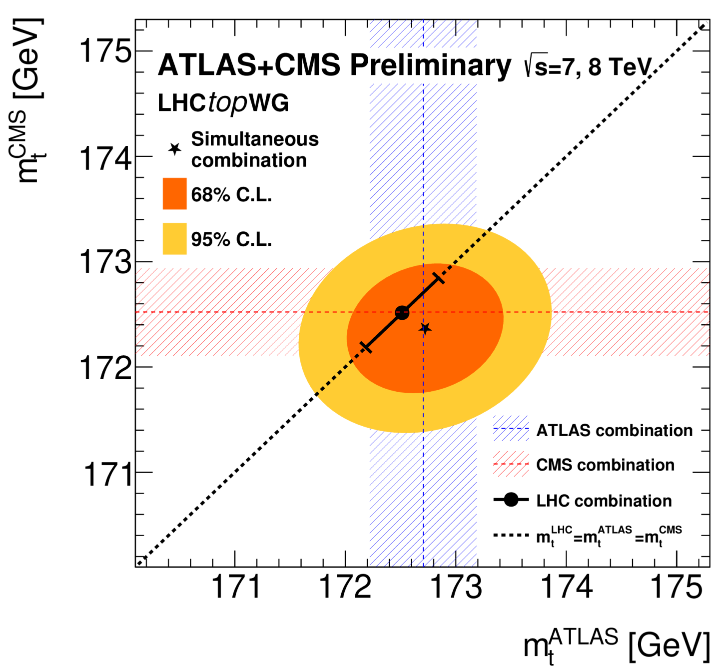


... and every EFT parameter measurements

The Non-Linear Models

There are also examples where the non-linearity comes from a latent variable dependence, e.g. the mass of a resonance or the CP violation parameter:

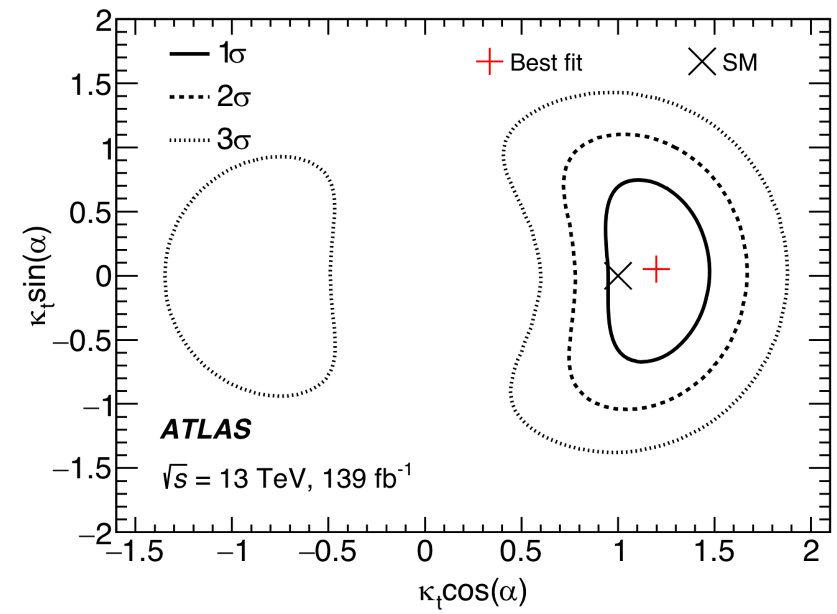
Top mass m_t measurement

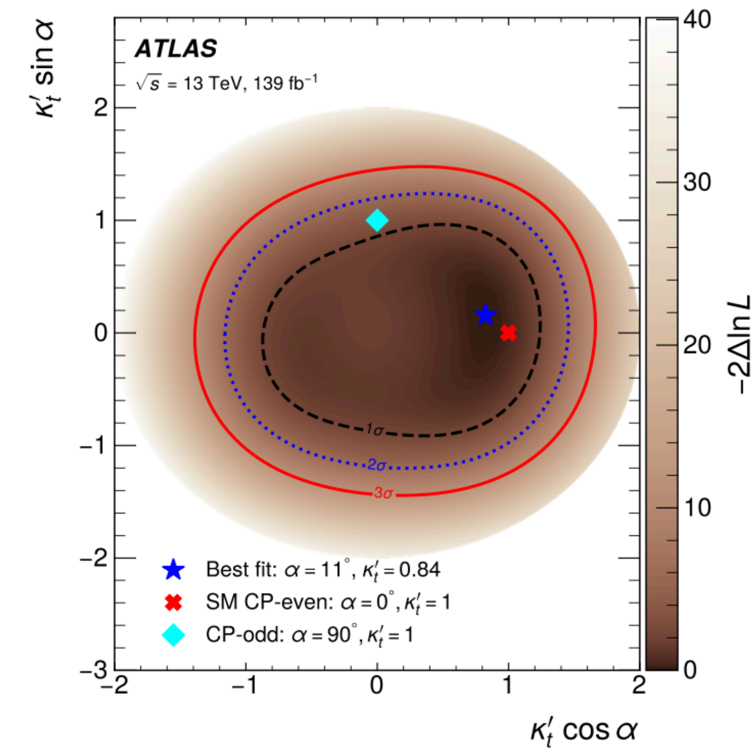


In some cases it is not even possible to write out an analytic dependence on the parameter of interest.

$$p(x \mid m_t) = ?$$

CP mixing angle α measurement





$$\mathcal{L}_{t\bar{t}H} = -\kappa_t' y_t \phi \bar{\psi}_t (\cos\alpha + i\gamma_5 \sin\alpha)\psi_t$$

Non-linearity comes from the cosine and sine dependence. Analytical dependence known.

Many precision measurements at the LHC (and beyond) can gain significantly from using an NSBI approach

Conclusions and Outlook

- Neural Simulation-Based Inference is a powerful statistical framework that can bring dramatic improvements in sensitivity for several measurements with non-linear models. Originally proposed by Cranmer et al, in three publications: [1805.00020, 1805.12244, 1805.00013]

- Several novel developments were done, like systematic uncertainty parameterization, robust diagnostics, Neyman Construction, efficient computing workflow, etc. to make the new workflow practical for a full analysis using the ATLAS experiment.

- A precise measurement of the off-shell Higgs boson and the Higgs boson decay width was performed using the ATLAS experiment data, with the new NSBI techniques.

- Hope to see wider adoption, accelerating the physics discovery potential.

Backup

Full workflow of the SBI Analysis