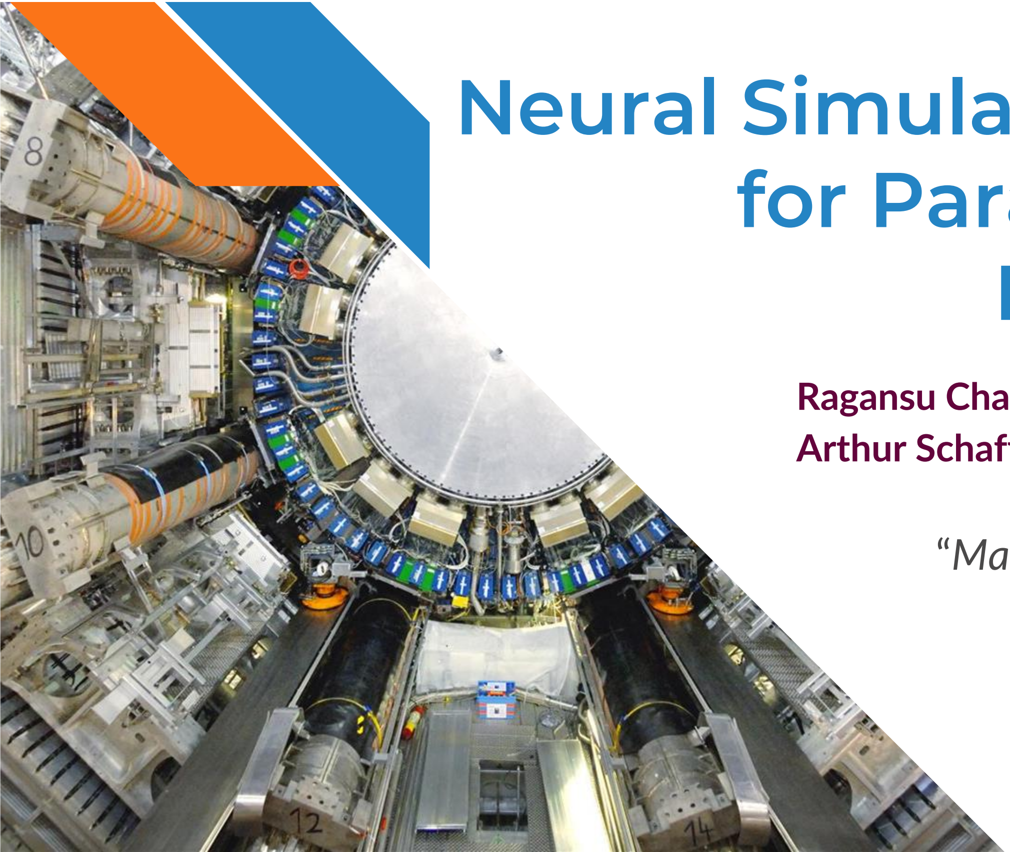


Neural Simulation Based Inference for Parameter Estimation in Higgs boson physics"

Ragansu Chakkappai, Arnaud J Maury, David Rousseau, Arthur Schaffer

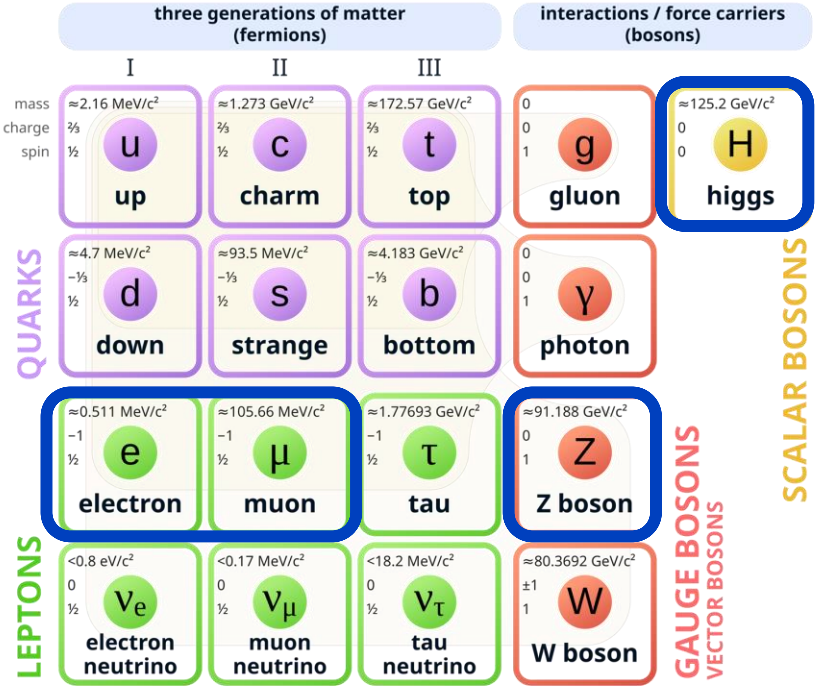
25th March 2025

"*Mathematical Foundations of AI*" day

The Standard Model of Particle Physics

Fermions: ½-spin

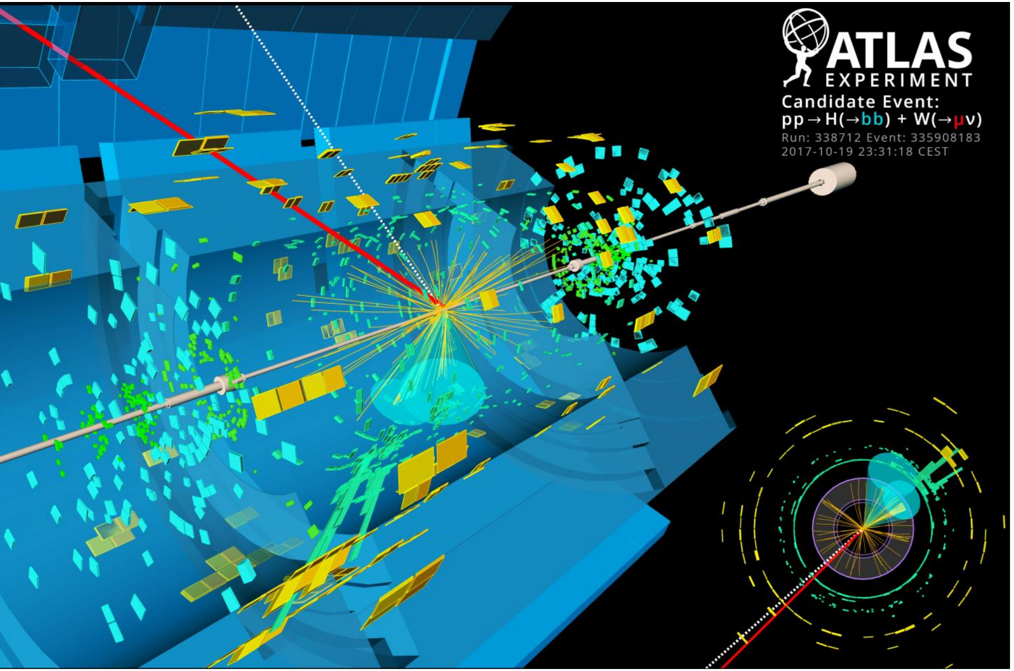
- **Quarks** : Strong (and weak) force
- **Leptons** : Weak force



Bosons: 0- or 1-spin

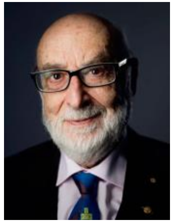
- **Gauge bosons** : mediate interactions
- **Higgs boson** : mass term for fermions

Higgs Boson



- ➔ The Higgs mechanism give mass of fundamental particles
- ➔ Proposed in 1964
- ➔ Discovered in 2012 at ATLAS and CMS experiment in LHC
- ➔ 2013 Nobel Physics Prize

Peter W. Higgs

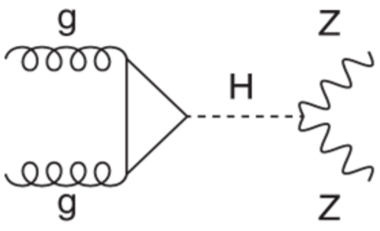


François Englert

Off-shell Higgs Boson cross-section

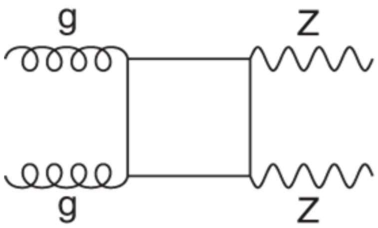
Signal S

$gg \to H \to ZZ$:



Background B

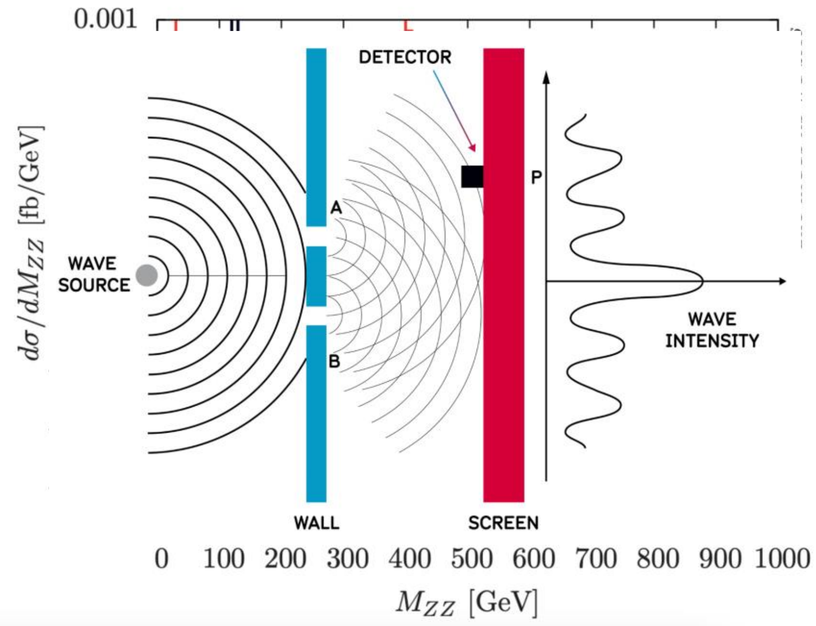
$gg \to ZZ$:



background $\quad < 0$

$$|\sqrt{\mu}\, M_S + M_B|^2 = \mu |M_S|^2 + |M_B|^2 + 2\sqrt{\mu}\, \Re\left[M_S M_B^*\right]$$

signal $\propto \mu$ $\quad$ *interference* $\propto \sqrt{\mu}$



Yield : Number of events counted.

$$\nu_{ggF}(\mu) = (\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{SBI} + (1 - \sqrt{\mu})\, \nu_B$$

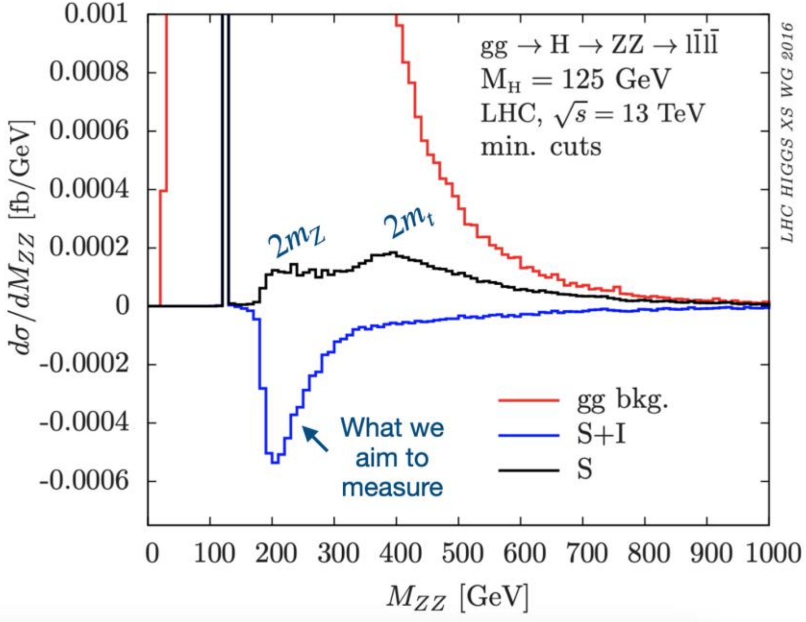
Off-shell Higgs Boson cross-section

Signal Strength

$$\mu = \frac{\nu_{obs}}{\nu_{exp}}$$

In Standard Model

$$\mu = 1$$



$$|\sqrt{\mu}\, M_S + M_B|^2 = \underbrace{\mu |M_S|^2}_{\text{signal} \propto \mu} + \overbrace{|M_B|^2}^{\text{background}} + \underbrace{\overbrace{2\sqrt{\mu}\, \Re\, [M_S M_B^*]}^{<0}}_{\text{interference} \propto \sqrt{\mu}}$$

Yield : Number of events counted.

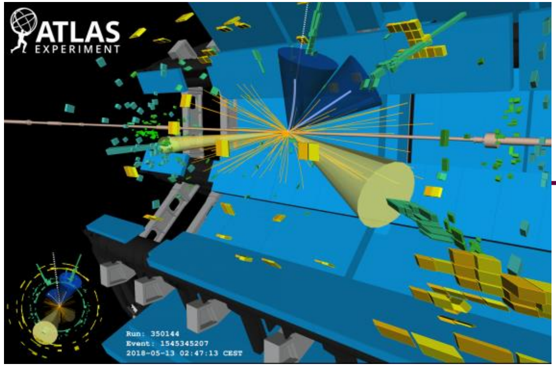
$$\nu_{ggF}(\mu) = (\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{SBI} + (1 - \sqrt{\mu})\, \nu_B$$

Statistical framework

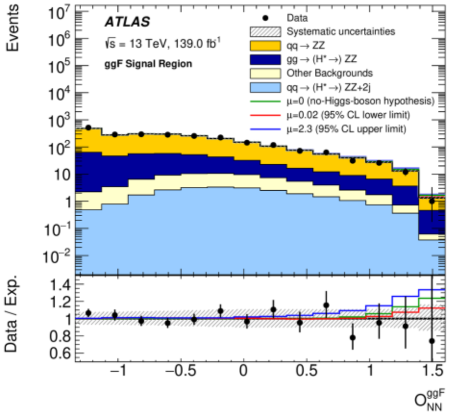
- Probability of a dataset D: $p(\mathcal{D}|\mu) = \mathrm{Pois}(N_{\text{data}}|\nu_{pp\to 4\ell}(\mu)) \prod_i^{N_{\text{data}}} \frac{1}{\sigma_{pp\to 4\ell}(\mu)} \frac{d\sigma_{pp\to 4\ell}}{dx}(x_i|\mu)$
- Likelihood: $\mathcal{L}(\mu|\mathcal{D}) = \frac{e^{-\nu(\mu)}\nu(\mu)^{N_{\text{data}}}}{N_{\text{data}}!} \prod_i^{N_{\text{data}}} \frac{1}{\sigma_{pp\to 4\ell}(\mu)} \frac{d\sigma_{pp\to 4\ell}}{dx}(x_i|\mu)$
- Negative Log Likelihood: $\text{NLL}(\mu) = -2\log\left(L(\mu|\mathcal{D})\right)$
- Neyman-Pearson lemma → (optimal) test statistic: $t_\mu = -2\log\left(\frac{L(\mu|\mathcal{D})}{L(\hat{\mu}|\mathcal{D})}\right) = \text{NLL}(\mu) - \text{NLL}(\hat{\mu})$

with $\hat{\mu} = \underset{\mu}{\operatorname{argmin}}\, \text{NLL}(\mu)$

Classical Histogram Method



High dimensional data



Summary Histogram
(**Neural Network classifier Score**)

Parameter of interest : $\boldsymbol{\mu}$

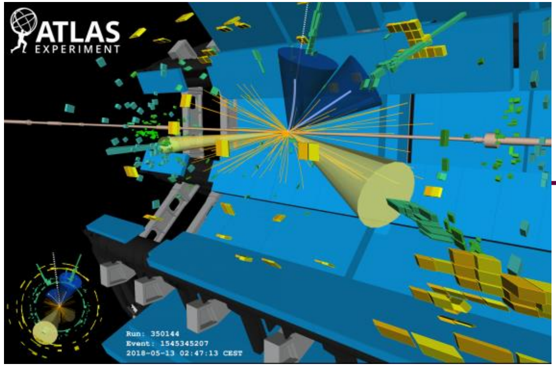
Statistical Fit

Inference:

$$\text{NLL}(\mu) = \sum_{j}^{N_{\text{bins}}} \left[-2\, N_{\text{data}}^{(j)} \left(\log \nu^{(j)}(\mu) \right) + 2 \left(\nu^{(j)}(\mu) \right) \right]$$

Here the likelihood is computed per **bin**

Unbinned high-dimensional neural inference



High dimensional data

$\boldsymbol{\mu}$

Parameter of interest

Neural Networks

Inference:

$$t_\mu = -2\log\frac{\mathcal{L}(\mu|\mathcal{D})}{\mathcal{L}(\text{ref}|\mathcal{D})}$$

$$= -2\left(N_{\text{data}}\log(\nu(\mu)) - \nu(\mu)\right) - 2\sum_{i=1}^{n} w_i \log\left(\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)}\right)$$

Here the likelihood is computed per **event**

Search-oriented Mixture model

Let $p(x_i|\mu)$ be the probability of finding any event x_i for a given parameter μ

Event rates known from simulation

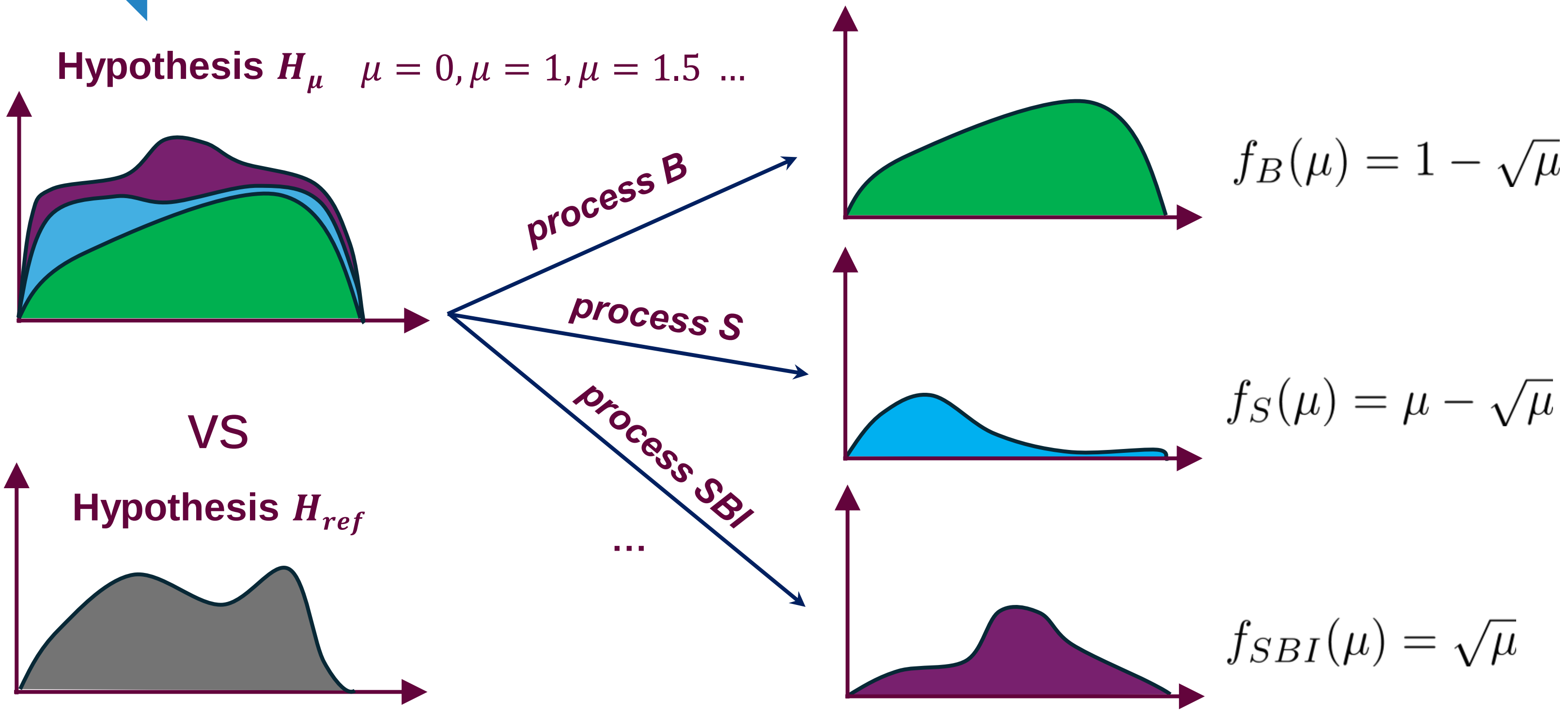
Known analytically from theory

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_X^C f_X(\mu) \cdot \nu_X \cdot p_X(x_i)$$

Estimated using an ML Classifier

$$\frac{p(x_i|\mu)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu)} \sum_X^C f_X(\mu) \cdot \nu_X \cdot \frac{p_X(x_i)}{p_{ref}(x_i)}$$

NSBI for Off-shell Higgs Boson



$$\frac{p(x_i|\mu)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu)}\left[(\mu - \sqrt{\mu}) \cdot \nu_S \cdot \frac{p_S(x_i)}{p_{ref}(x_i)} + \sqrt{\mu} \cdot \nu_{SBI} \cdot \frac{p_{SBI}(x_i)}{p_{ref}(x_i)} + (1 - \sqrt{\mu}) \cdot \nu_B \cdot \frac{p_B(x_i)}{p_{ref}(x_i)}\right]$$

Density ratio estimation using NNs

Dataset:
- Tabular datasets with 13 features
- 4ℓ system decay kinematic variables: $\cos(\theta^*)$, $\cos(\theta_1)$, $\cos(\theta_2)$, ϕ_1, ϕ, m_{Z_1} , m_{Z_2}
- Higgs production kinematic variables: $p_{T_{4\ell}}$, $\Upsilon_{4\ell}$
- Jet kinematics variables (if applicable): n_{jets} , m_{jj} , ϕ_{jj} , η_{jj}

Neural Networks (X vs Ref)

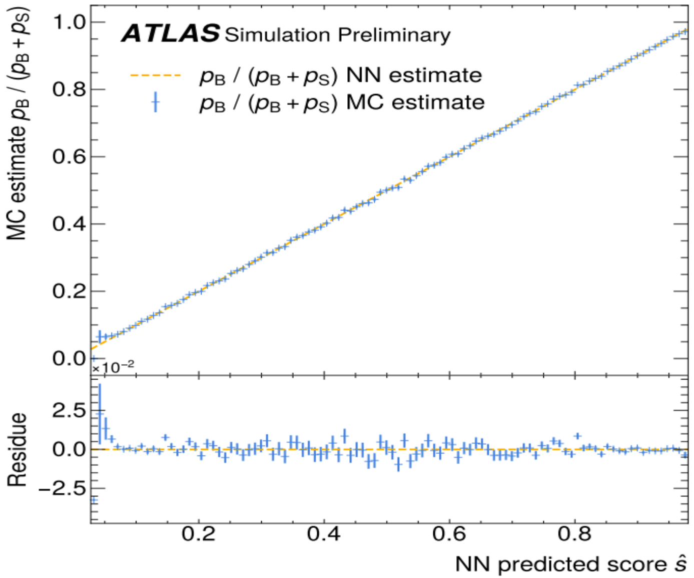
$$\hat{s}_{\text{pred}}(x_i) = \frac{p_X(x_i)}{p_X(x_i) + p_{\text{ref}}(x_i)}$$

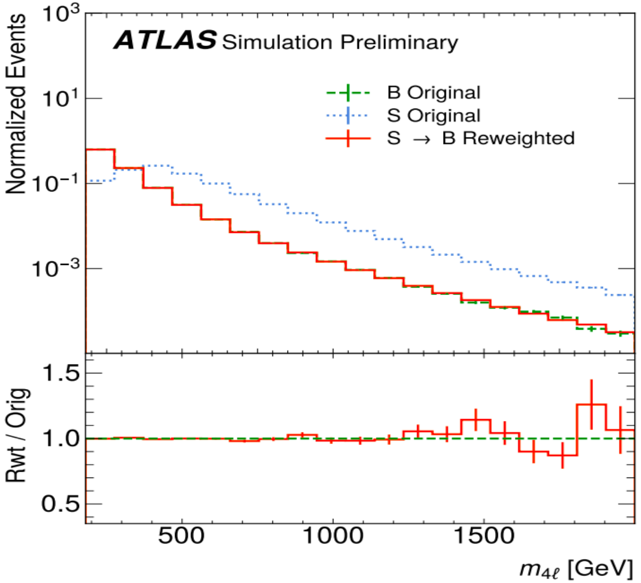
- Simple DNN classifier with 5 hidden layers of 1000 nodes
- Binary Cross Entropy loss
- Ensembles of DNNs to reduce bias and variance:
  - ~50 NNs per process

$$\frac{p_X(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{\hat{s}_{\text{pred}}(x_i)}{1 - \hat{s}_{\text{pred}}(x_i)}$$

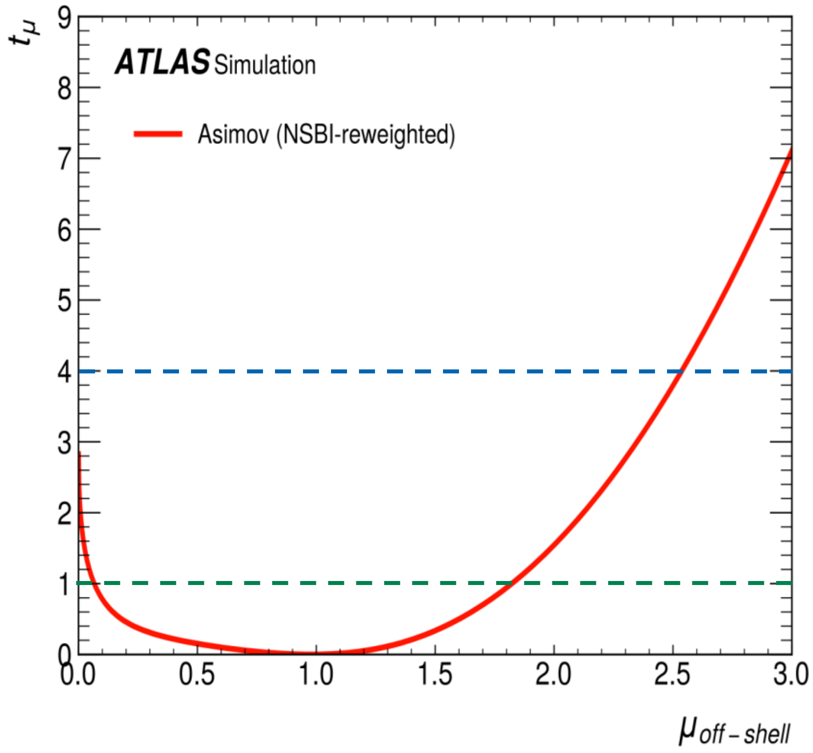
Classifier checks

- Ideally, score should reflect the $\frac{p_X(x_i)}{p_X(x_i)+p_{ref}(x_i)}$
  - Calibration Plots (left) checks this.
- The rewriting plot(right) are signal events reweighted with $\frac{p_X(x_i)}{p_{ref}(x_i)}$





NLL of NSBI H*4ℓ analysis



H^*4l NLL is non-quadratic

1σ and 2σ Confidence Intervals might not be at 1 and 4…

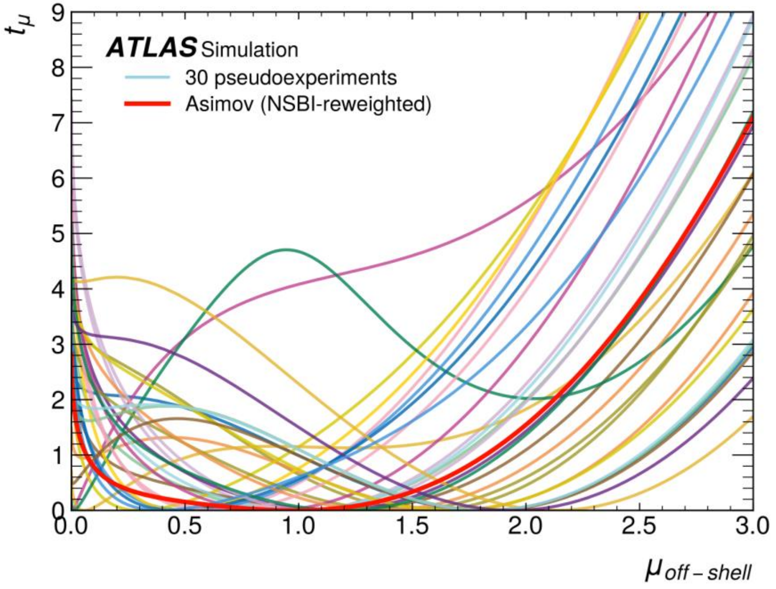
Asimov dataset is the average dataset

NLL minimisation for each pseudo-experiments

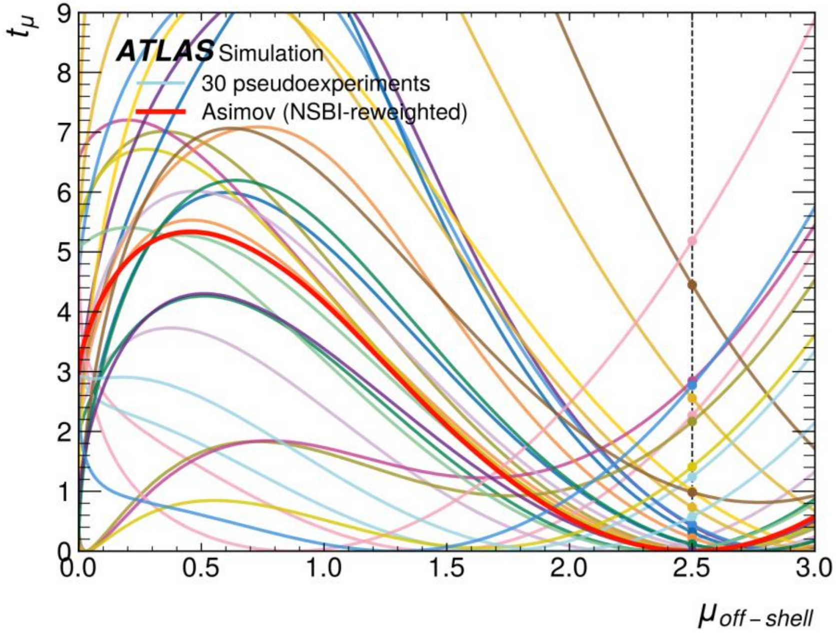
Hypothesis $\mu_{truth} = 1$

Pseudo-experiments with bootstrapped datasets

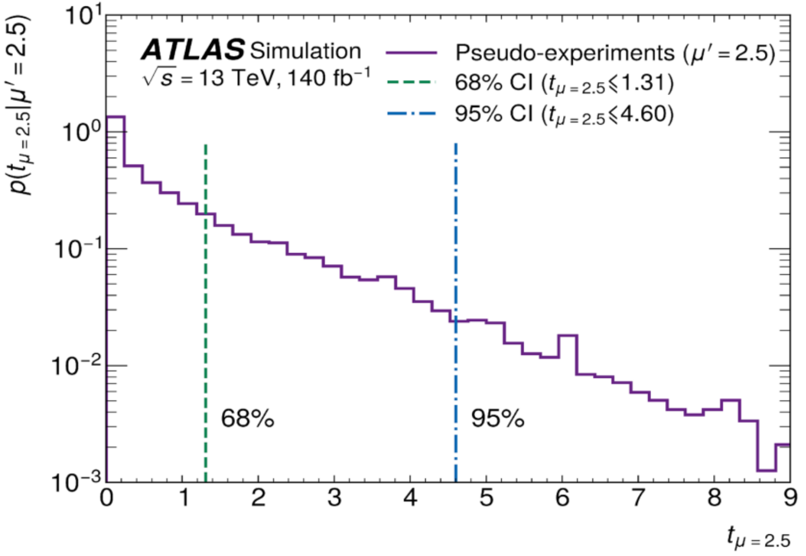
- ~2000 events sampled **with replacement** from an Asimov dataset of ~10M
- ~10K pseudo-experiments for each hypothesis μ_{truth}
  - millions of pseudo-experiments overall
- Due to interference, each pseudo experiment behaves differently from each other



Neyman construction

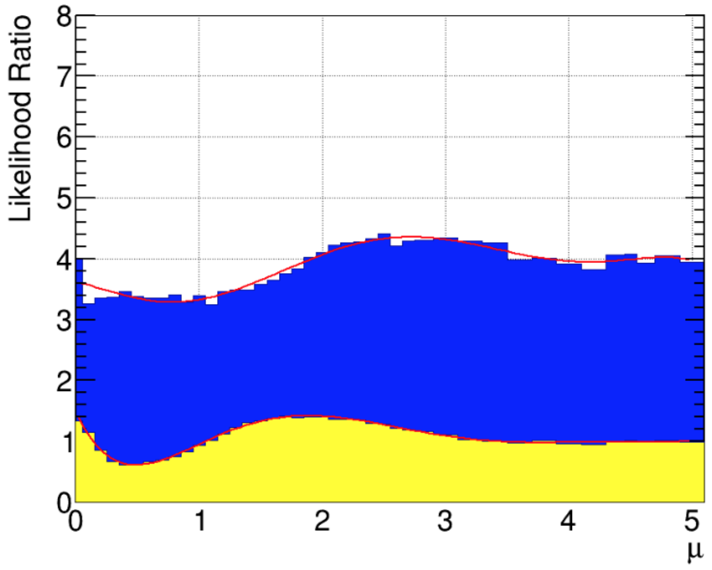


- Compute $t_{\mu=\mu_{truth}}$ → the test statistic at $\mu = \mu_{truth}$
- Integrate up to 68.27% (95.45%) to determine the 1σ (2σ) CI for hypothesis μ_{truth}



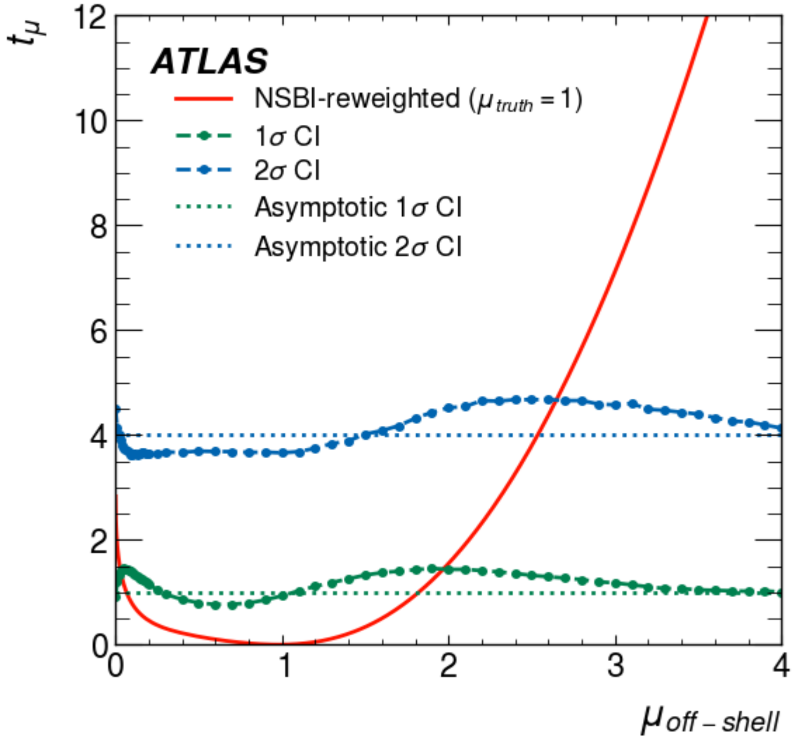
Repeat this study for several hypotheses μ_{truth} …

Results of Neyman construction for Stat-only



(a) Histogram-based $H^*4\ell$ analysis [126]

Plot credit: Michiel Jan Veen



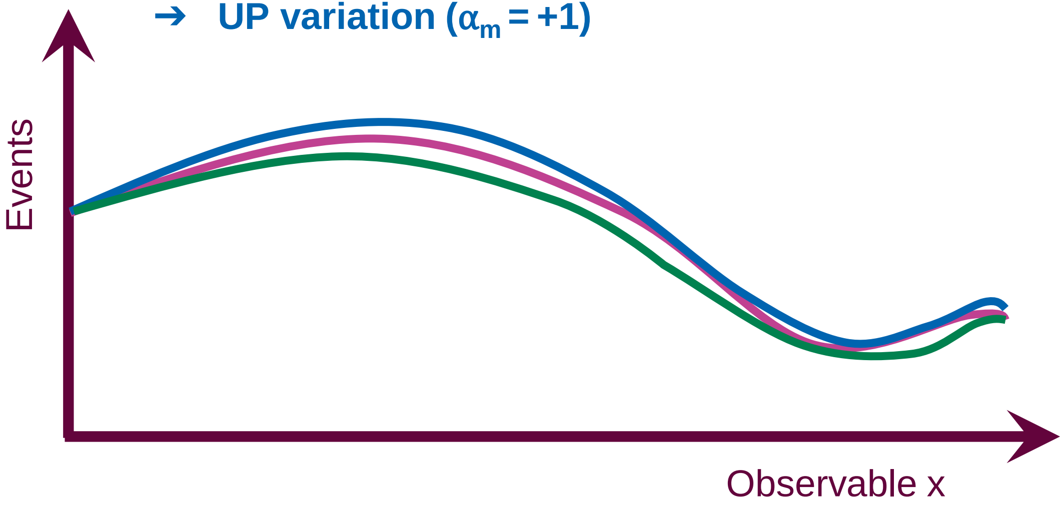
Systematic Uncertainties (Epistemic)

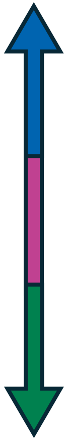
- There are experimental and theoretical uncertainties.
- In our probability model, systematic uncertainties are modelled by Nuisance Parameters α

➔ **DOWN variation ($\alpha_m = -1$)**

➔ **Nominal ($\alpha_m = 0$)**

➔ **UP variation ($\alpha_m = +1$)**





UP variation

$$W_X(\alpha_m^+) = \frac{\nu_X(\alpha_m = +1)}{\nu_X}$$

Nominal

DOWN variation

$$W_X(\alpha_m^-) = \frac{\nu_X(\alpha_m = -1)}{\nu_X}$$

Systematic Uncertainties (Epistemic)

Yield term: gives the factor of change in yield due to the NP

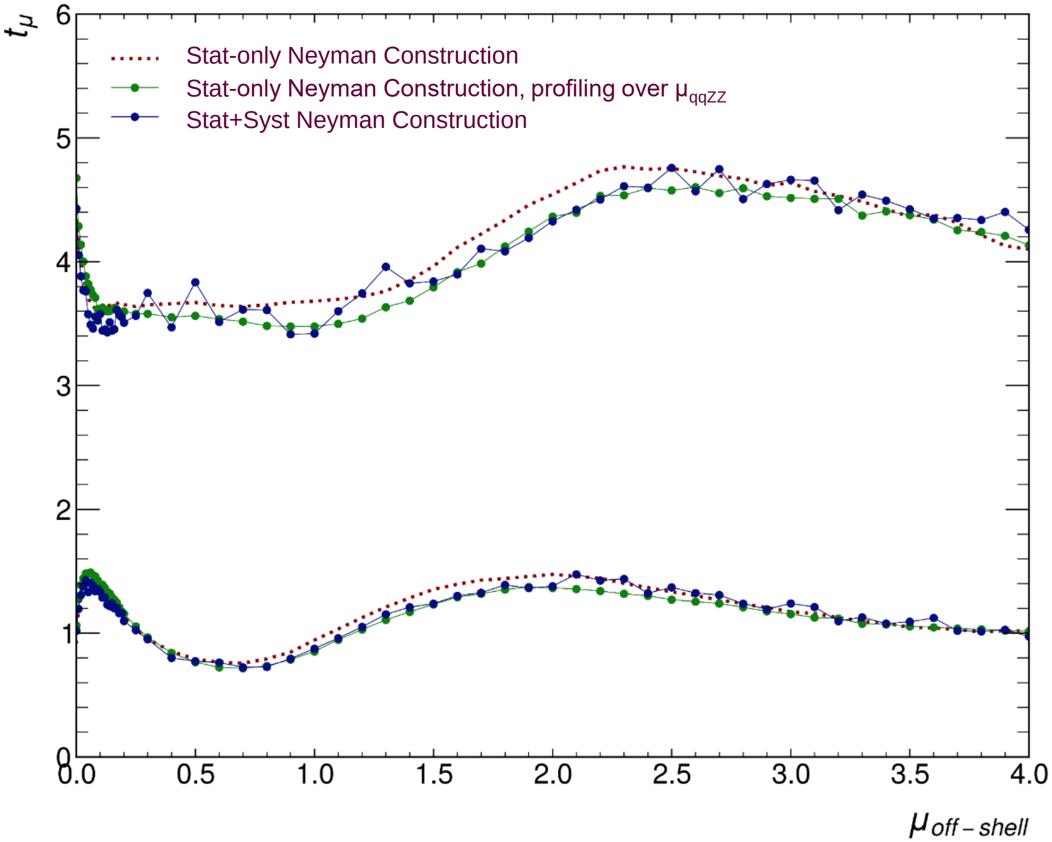
The density ration can be modified with Nuisance parameter NP,

Which then can be profiled

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_{X}^{C} f_X(\mu) \cdot \nu_X \cdot \frac{p_X(x_i)}{p_{ref}(x_i)} \prod_{k}^{N_{syst}} W_X(\alpha_m) g_X(x_i,\alpha_m)$$

Gives the shape change due to the NP

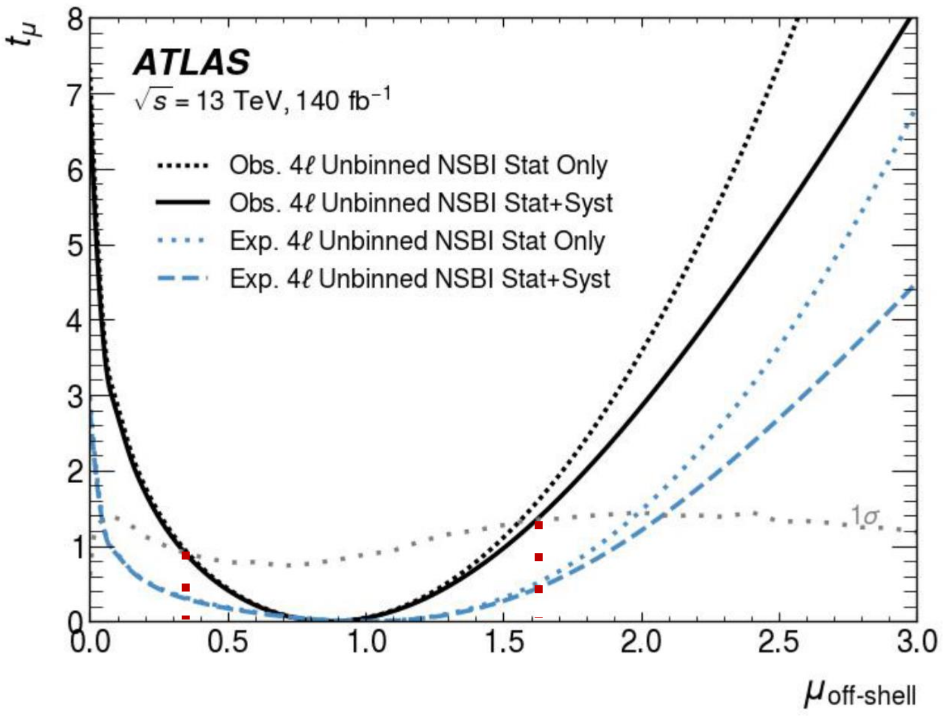
Results of Neyman construction for Stat+Syst



- Similar Confidence Intervals
- Stat-only CIs are sufficient
- Stat+Syst verification needed

Results of NSBI H*4l analysis

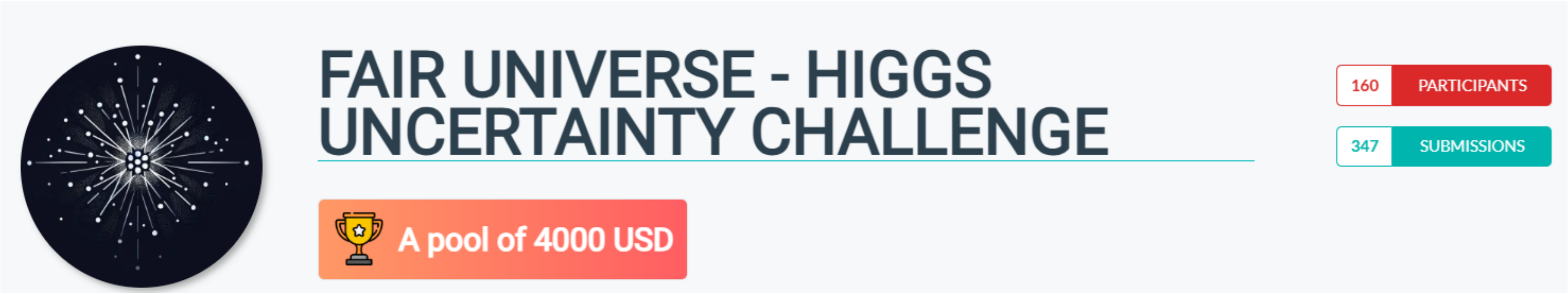
Observed off-shell signal strength $\mu_{\text{off-shell}} = 0.87^{+0.75}_{-0.54}$



Conclusion and Discussion

- The first implementation of Neural Simulation Based Inference (NSBI) in Particle Physics
- 2 ATLAS papers made public in Nov 2024 (submitted to Report on Progress on Physics)
    - Methods Paper : arXiv:2412.01600 [hep-ex]
    - Physics Analysis Paper: arXiv:2412.01548 [hep-ex]
- NSBI: well-suited for similar problems with strong (negative) quantum interferences

Fair Universe: HiggsML Uncertainty Challenge



- HiggsML Uncertainty Challenge Ran from September 12 to March 14th
- Accepted as NeurIPS competition 2024
- Dedicated workshop at NeurIPS 2024 Conference , Vancouver

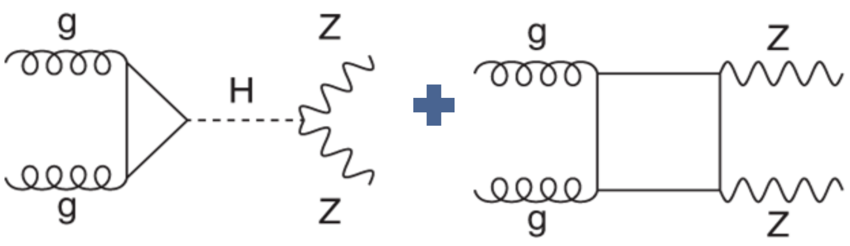
Thank you for your attention!

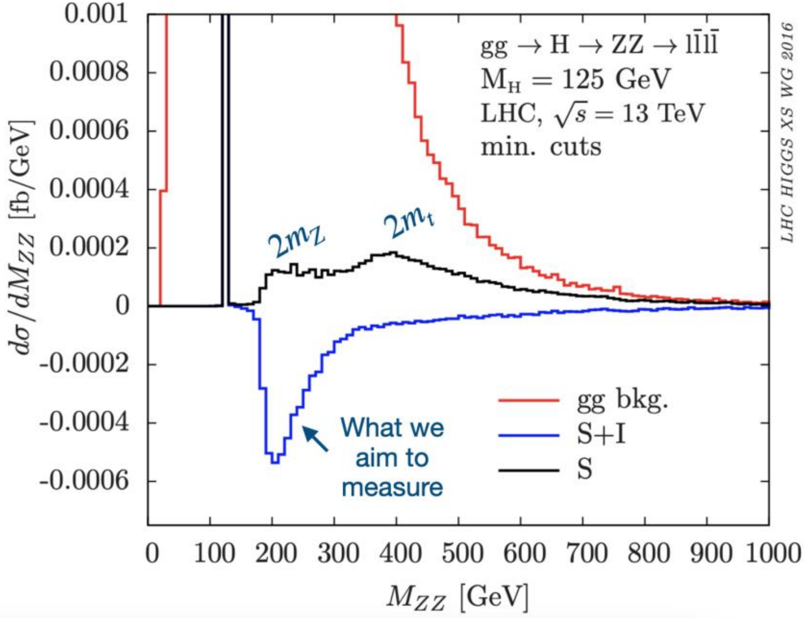
Back Up

Off-shell Higgs Boson cross-section

Signal S | **Background B**

$gg \to H \to ZZ$: | $gg \to ZZ$:



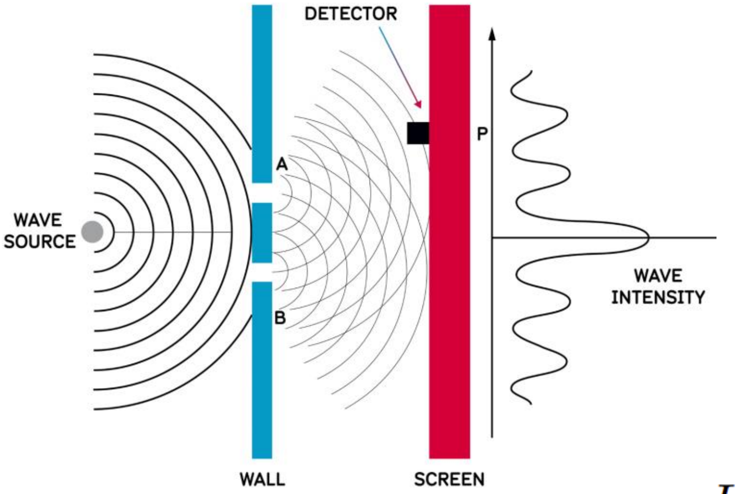


$$|\sqrt{\mu}\, M_S + M_B|^2 = \underbrace{\mu |M_S|^2}_{\text{signal} \,\propto\, \mu} + \overbrace{|M_B|^2}^{\text{background}} + \underbrace{\overbrace{2\sqrt{\mu}\, \Re\left[M_S M_B^*\right]}^{<0}}_{\text{interference} \,\propto\, \sqrt{\mu}}$$

Yield : Number of events counted.

$$\nu_{ggF}(\mu) = (\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{SBI} + (1 - \sqrt{\mu})\, \nu_B$$

Quantum Interference



$$E_1 = Ae^{i(kx-\omega t)}$$

$$E_2 = Be^{i(kx-\omega t+\phi)}$$

$$E_{\text{total}} = E_1 + E_2 = Ae^{i(kx-\omega t)} + Be^{i(kx-\omega t+\pi)}$$

$$E_{\text{total}} = e^{i(kx-\omega t)}(A - B)$$

$$I \propto |E_{\text{total}}|^2$$

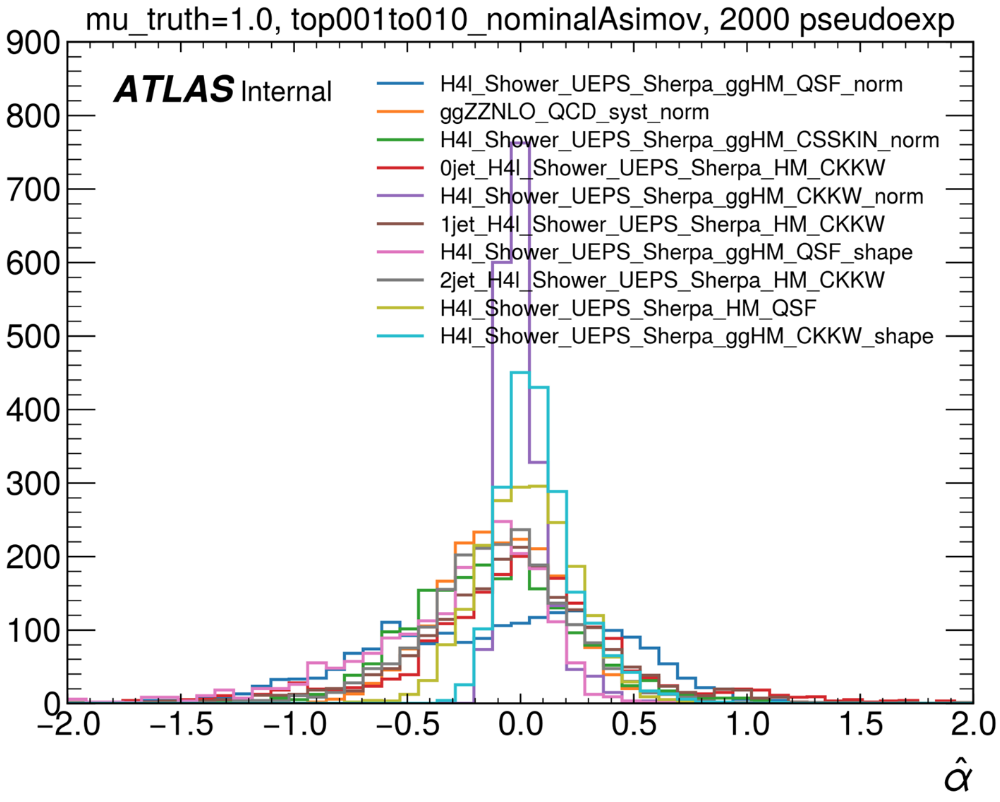
$$I \propto |E_{\text{total}}|^2 = A^2 + B^2 - 2AB$$

Quantum Interference Decreases the intensity

Neyman construction with systematics (Nuisance Parameters)

Maximum Likelihood Estimates (MLE)

$$(\hat{\mu}, \hat{\alpha}) = \underset{\mu,\alpha}{\mathrm{argmax}}\, \mathcal{L}_{\mathsf{Full}}(\mu, \alpha)$$



Systematics - Interpolation

- The G(α) is the then fitted from ratio of the yields
- gx(α +) and gx(α -) is density ratio from NN Classifiers (+/- vs nominal)
- g(α) is a polynomial fit between them

$$W_X(\alpha_m) = \begin{cases} \left(\frac{\nu_X({\alpha_m}^{(+)})}{\nu_X({\alpha_m}^{0})}\right)^{\alpha_m}, & \text{if } \alpha_m > 1 \\ 1 + \sum_{n=1}^{3} c_n {\alpha^n}_k, & \text{if } -1 \leq \alpha_m \leq 1 \\ \left(\frac{\nu_X({\alpha_m}^{(-)})}{\nu_X({\alpha_m}^{0})}\right)^{-\alpha_m}, & \text{if } \alpha_m < -1 \end{cases}$$

Estimated using a NN Classifier

$$g_X(x_i, \alpha_m) = \begin{cases} g_X(x_i, {\alpha_m}^{(+)})^{\alpha_m}, & \text{if } \alpha_m > 1 \\ 1 + \sum_{n=1}^{3} c_n {\alpha^n}_k, & \text{if } -1 \leq \alpha_m \leq 1 \\ g_X(x_i, {\alpha_m}^{(-)})^{-\alpha_m}, & \text{if } \alpha_m < -1 \end{cases}$$