Mass Independent Event Classification for $H^{\pm\pm}$ Search at ATLAS

Canadian Association of Physicists Congress 2025
Saskatoon, SK

Adrienne Scott, June 10 2025

Extended Higgs Sector Motivation

- Additional Higgs bosons are not forbidden - why should the Higgs sector be minimal?

Extended Higgs Sector Motivation

- Additional Higgs bosons are not forbidden - why should the Higgs sector be minimal?

Many open problems in physics …

Why is there a matter-antimatter asymmetry?

What is the nature of dark matter?

Do we exist in a metastable vacuum?

Why does QCD appear to preserve CP-symmetry?

Extended Higgs Sector Motivation

- Additional Higgs bosons are not forbidden - why should the Higgs sector be minimal?

Many open problems in physics …

Why is there a matter-antimatter asymmetry?	→	CP violating Higgs decays
What is the nature of dark matter?	→	Dark matter candidates
Do we exist in a metastable vacuum?	→	Enhancement of vacuum stability
Why does QCD appear to preserve CP-symmetry?	→	Axions

… could be addressed by an extended Higgs sector

Extended Higgs Sector Motivation

- Additional Higgs bosons are not forbidden - why should the Higgs sector be minimal?

Many open problems in physics …

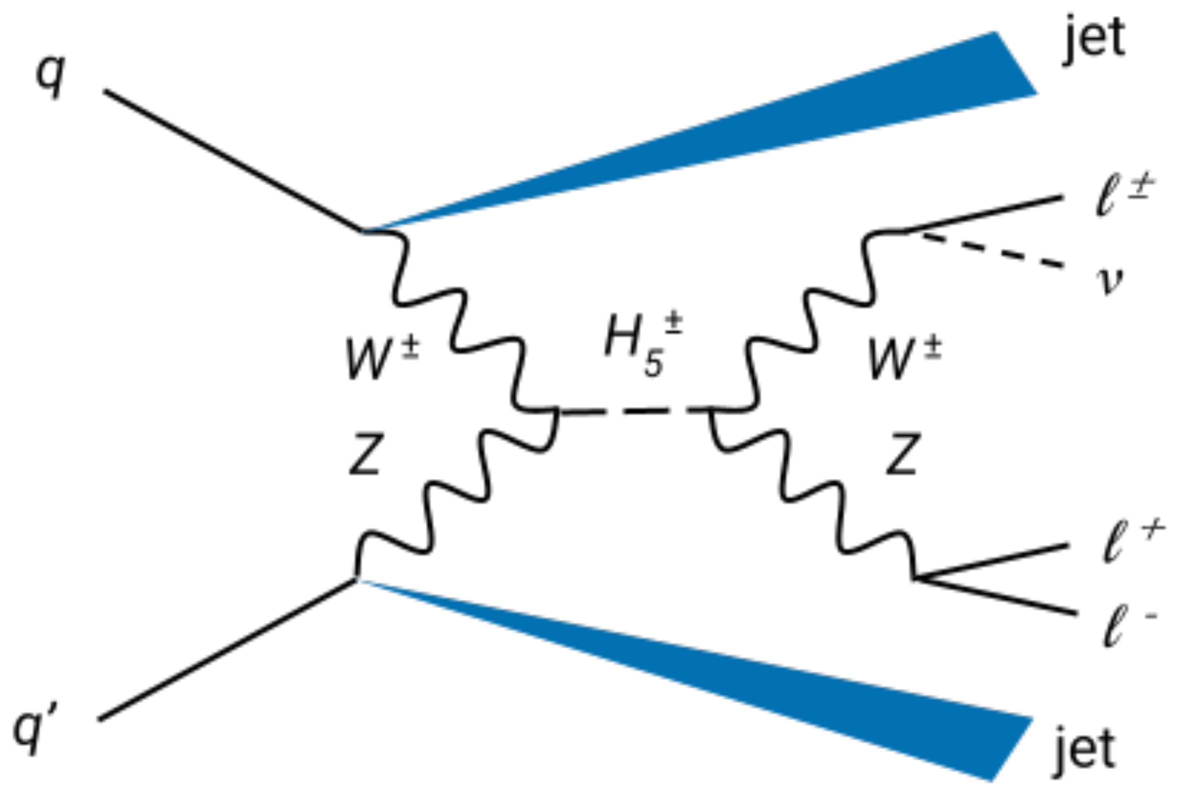
Why is there a matter-antimatter asymmetry?	→	CP violating Higgs decays
What is the nature of dark matter?	→	Dark matter candidates
Do we exist in a metastable vacuum?	→	Enhancement of vacuum stability
Why does QCD appear to preserve CP-symmetry?	→	Axions

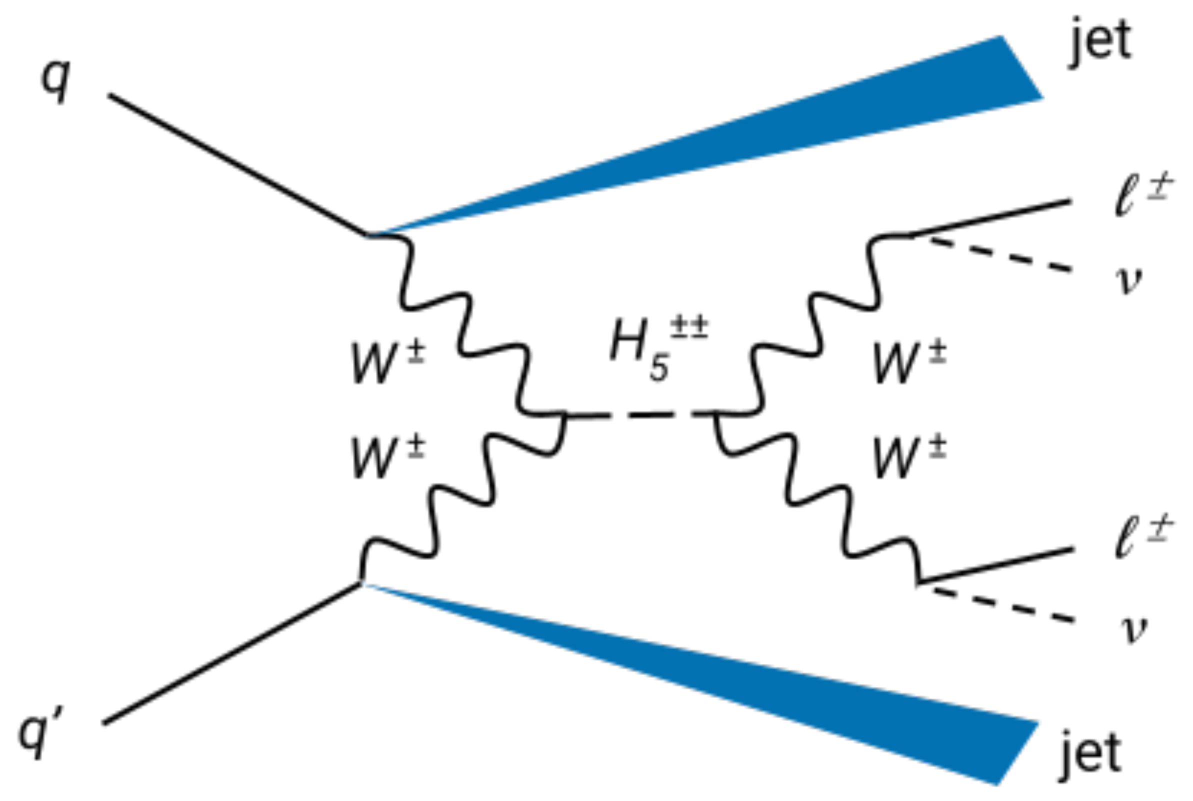
… could be addressed by an extended Higgs sector

- Unprecedented searches for additional Higgs bosons can be performed using *p*-*p* collisions recorded by the ATLAS detector at the LHC
  - Sensitivity continues to improve with increased luminosity

Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector by adding a quintuplet ($H_5^0, H_5^{\pm}, H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is searching for $H_5^{\pm}$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^{\pm}Z$ or $W^{\pm}W^{\pm}$

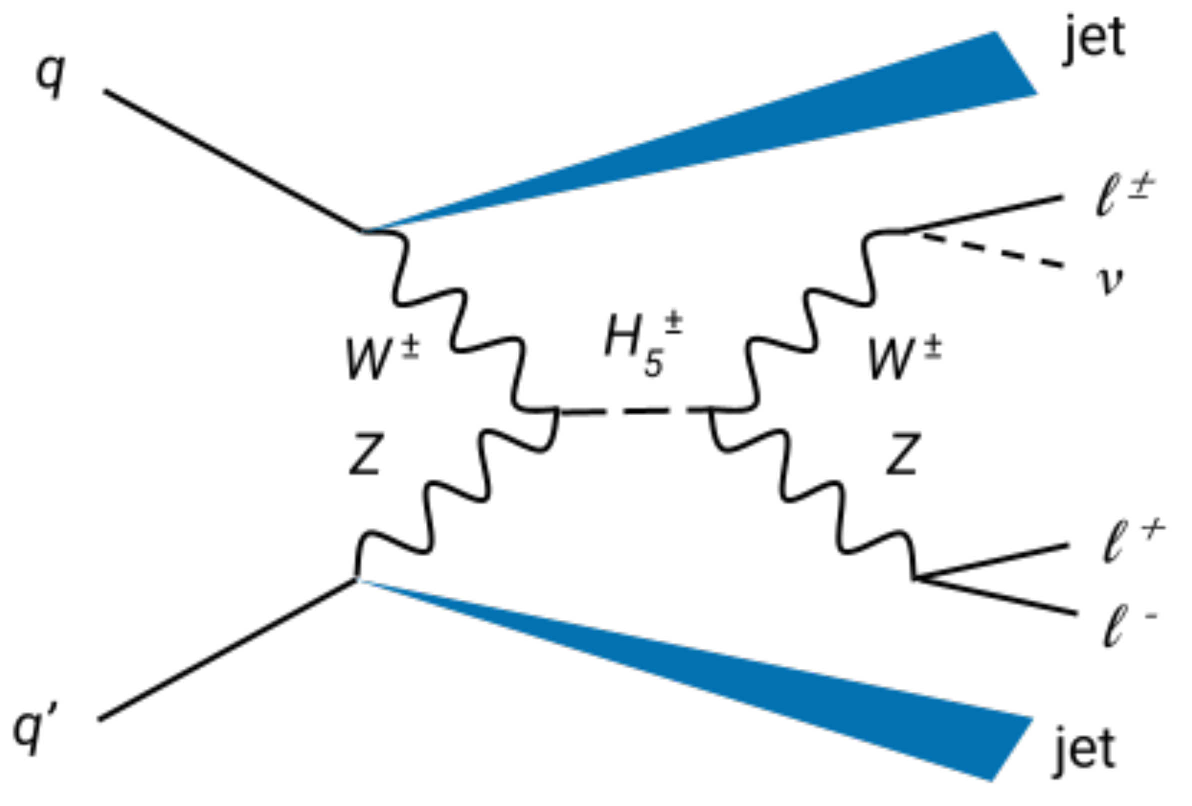


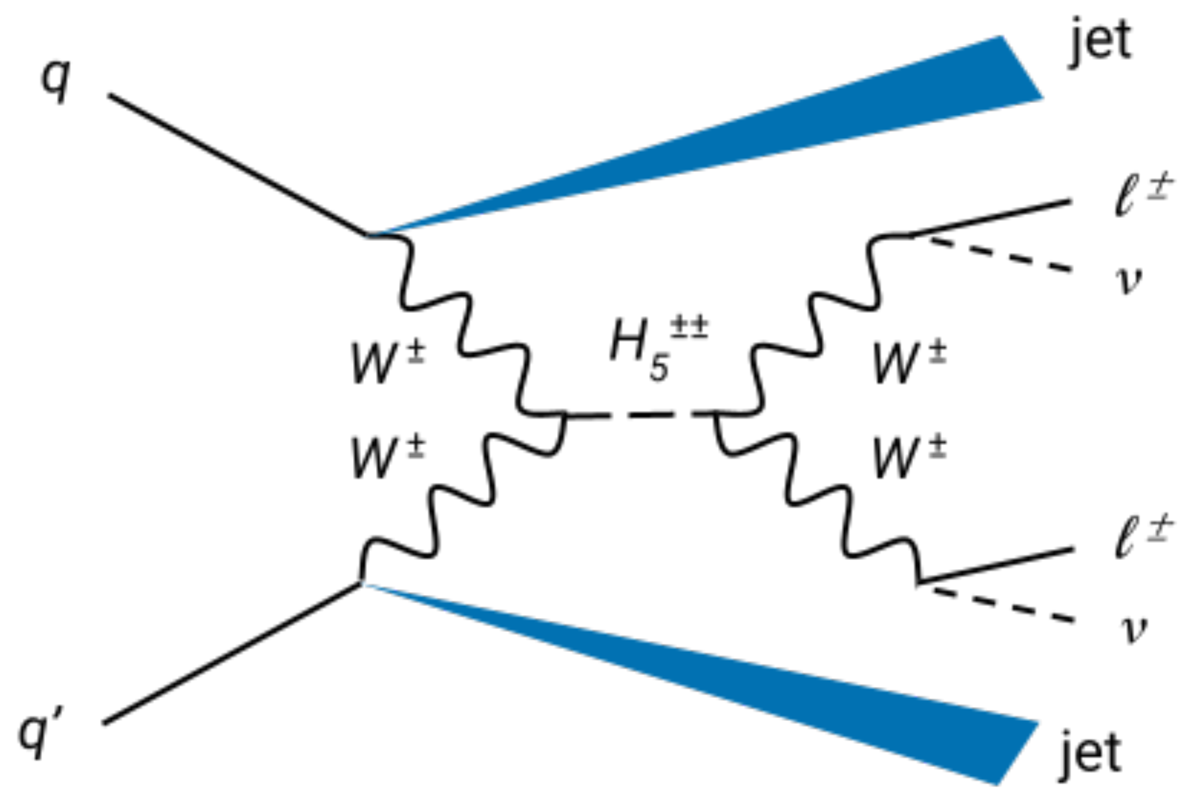


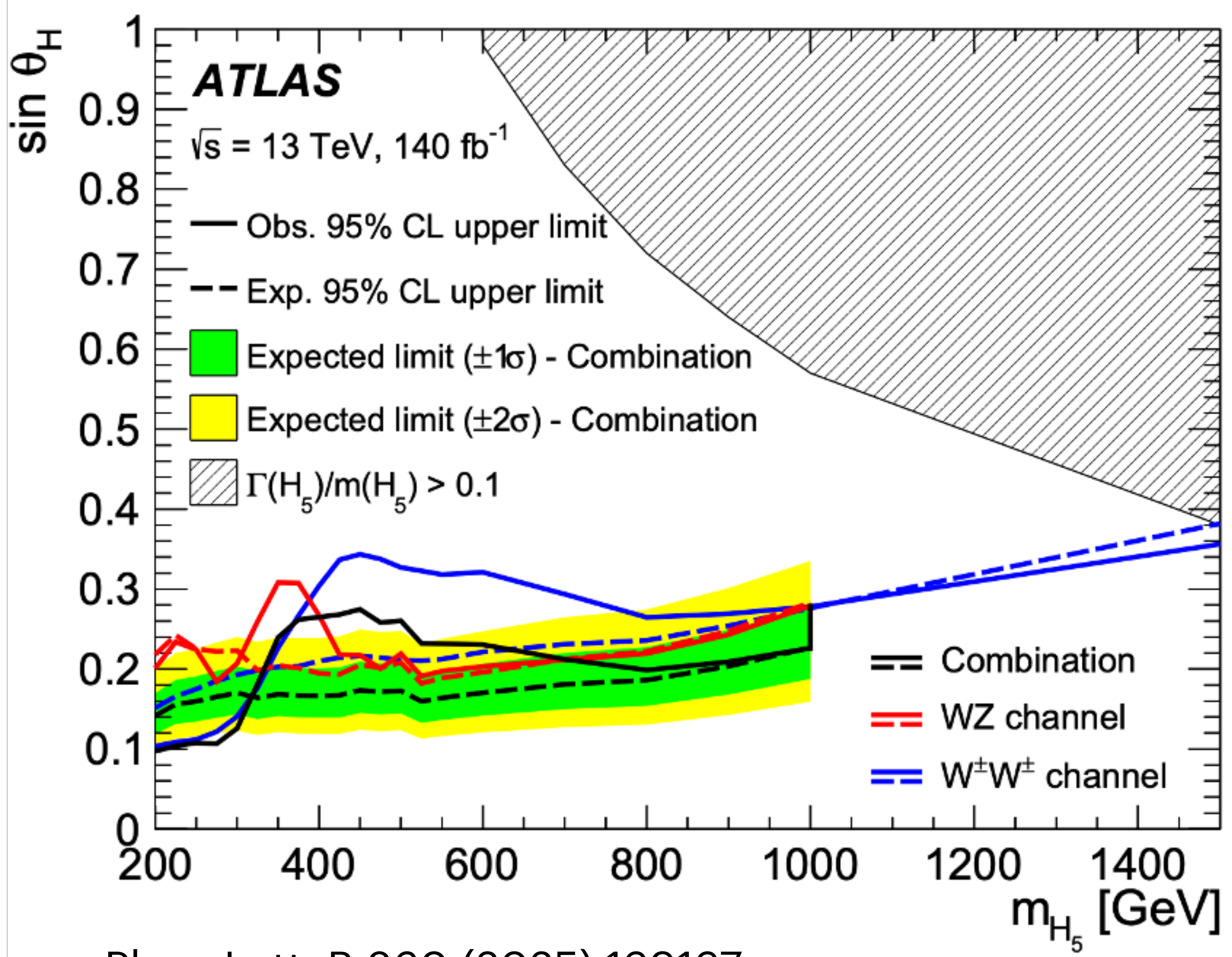
1 Nucl. Phys. B 262 (1985) 463

Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector by adding a quintuplet ($H_5^0, H_5^{\pm}, H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is searching for $H_5^{\pm}$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^{\pm}Z$ or $W^{\pm}W^{\pm}$
- Motivated by excess observed in Run 2 in $W^{\pm}Z$ and $W^{\pm}W^{\pm}$ channels
- Objective is to investigate excess and improve limits by optimizing search for GM model and using Run 2 and partial Run 3 (2022-2023) data







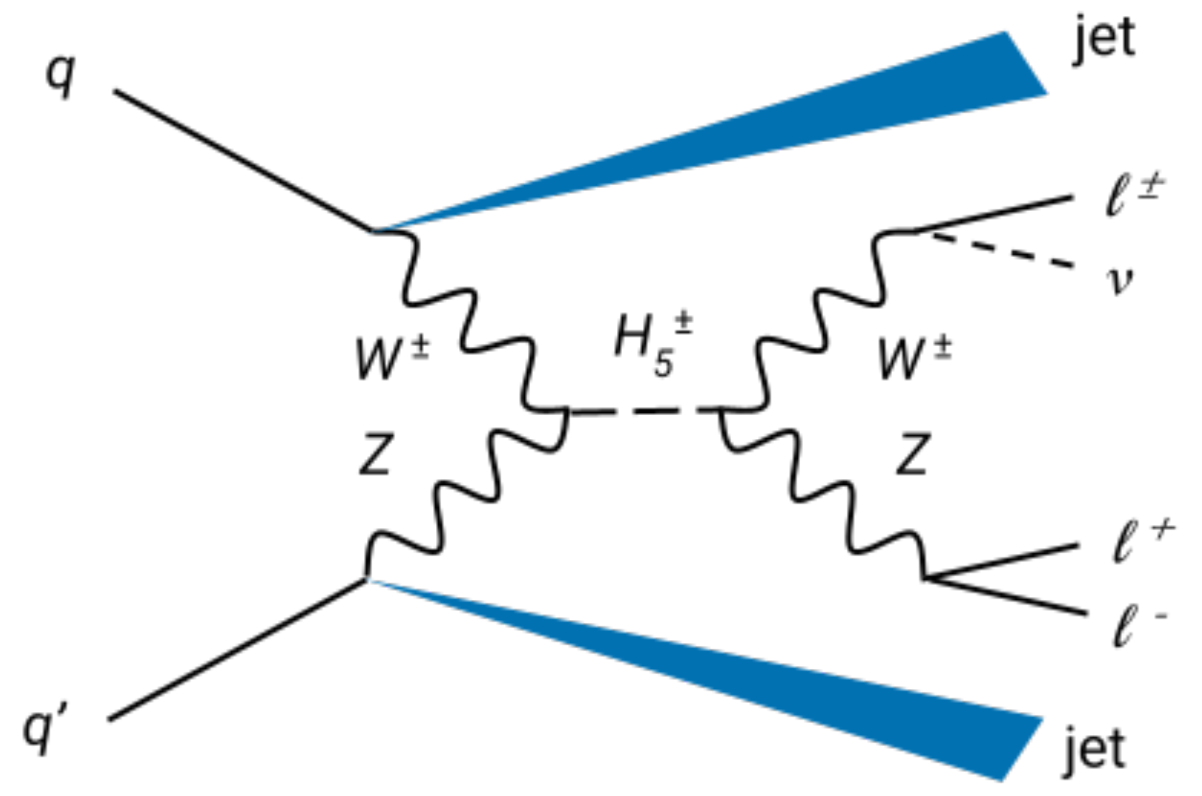
Phys. Lett. B 860 (2025) 139137

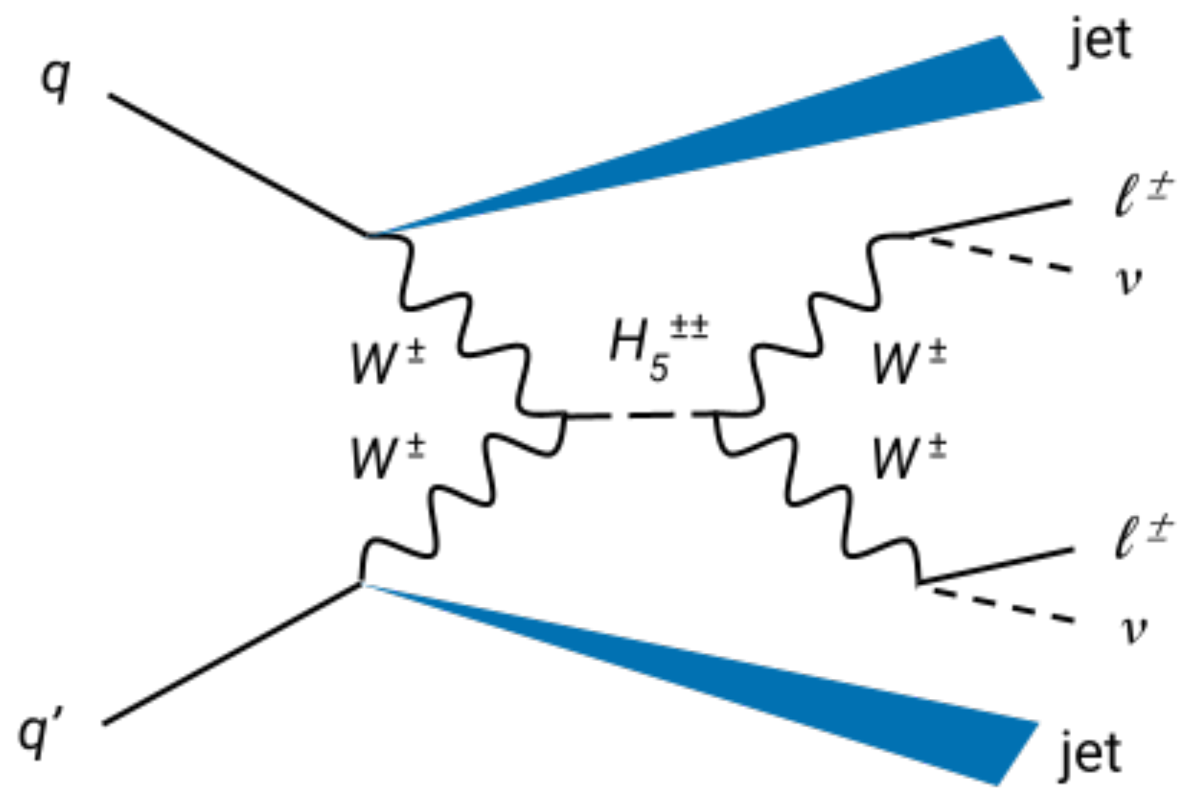
- Local (global) significance of 3.3 (2.5) at 375 GeV

1 Nucl. Phys. B 262 (1985) 463

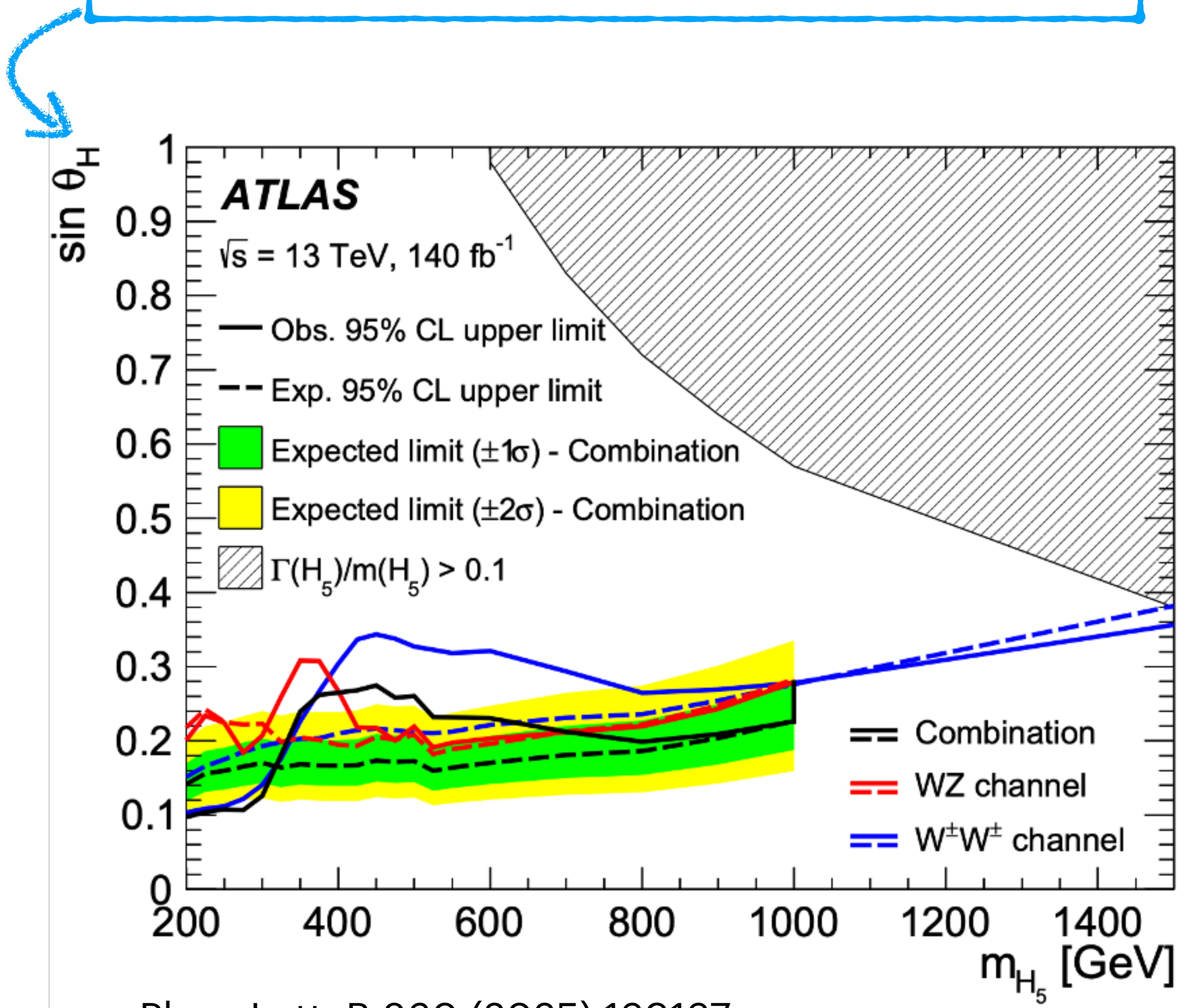
Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector by adding a quintuplet ($H_5^0, H_5^{\pm}, H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is searching for $H_5^{\pm}$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^{\pm}Z$ or $W^{\pm}W^{\pm}$
- Motivated by excess observed in Run 2 in $W^{\pm}Z$ and $W^{\pm}W^{\pm}$ channels
- Objective is to investigate excess and improve limits by optimizing search for GM model and using Run 2 and partial Run 3 (2022-2023) data





$\sin(\theta_H)$ is a free parameter in the GM model; σ for $H_5^{\pm}$ and $H_5^{\pm\pm}$ production $\propto \sin^2(\theta_H)$



Phys. Lett. B 860 (2025) 139137

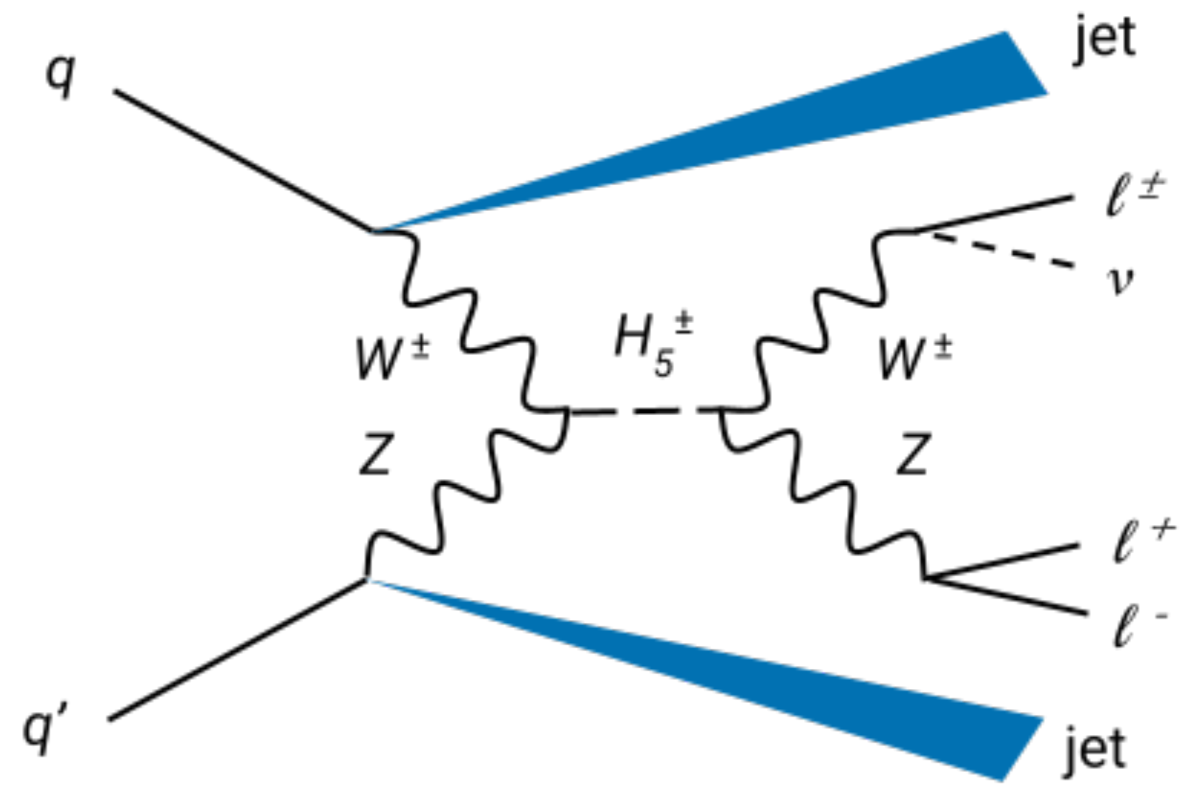
- Local (global) significance of 3.3 (2.5) at 375 GeV

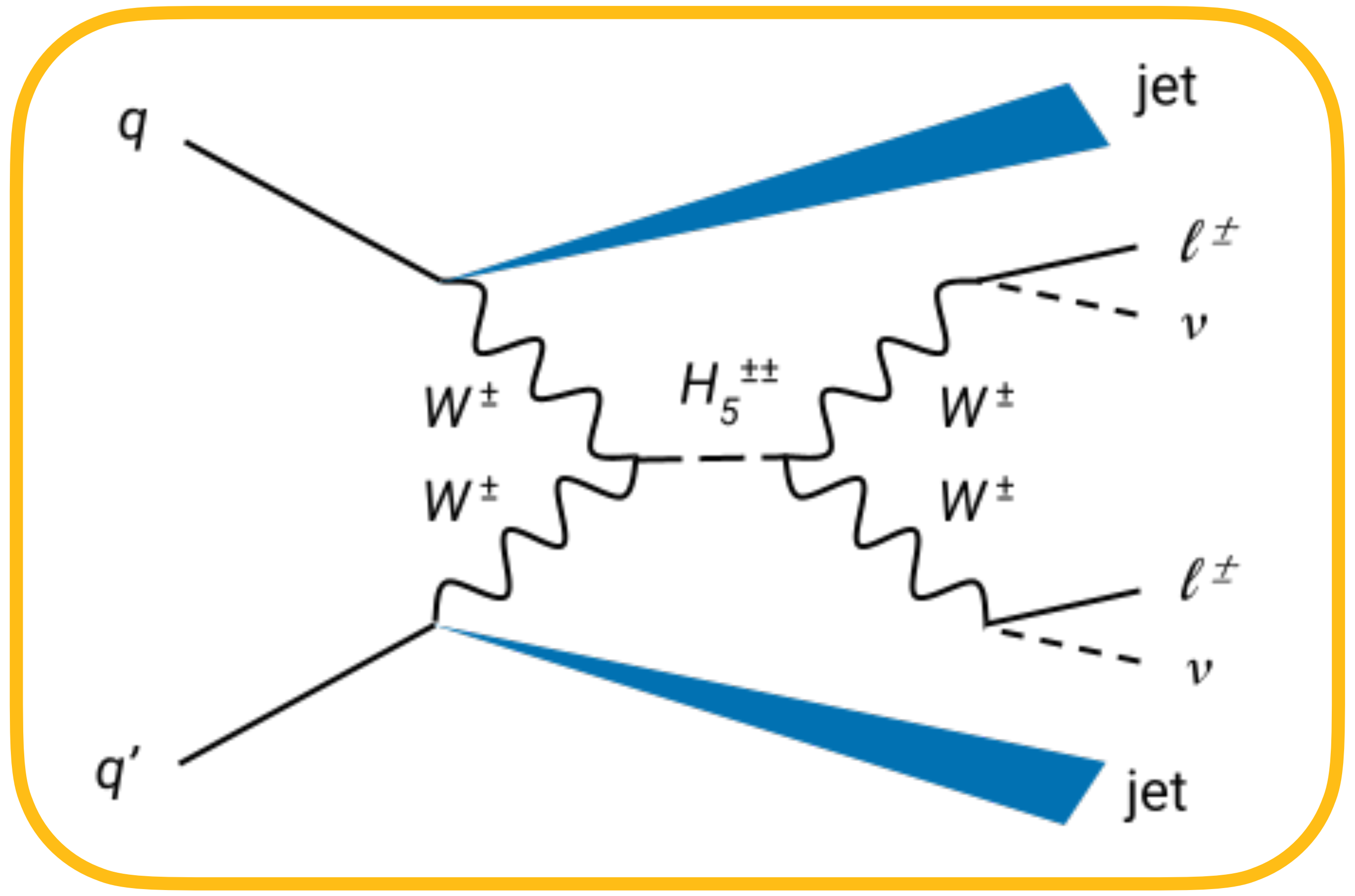
1 Nucl. Phys. B 262 (1985) 463

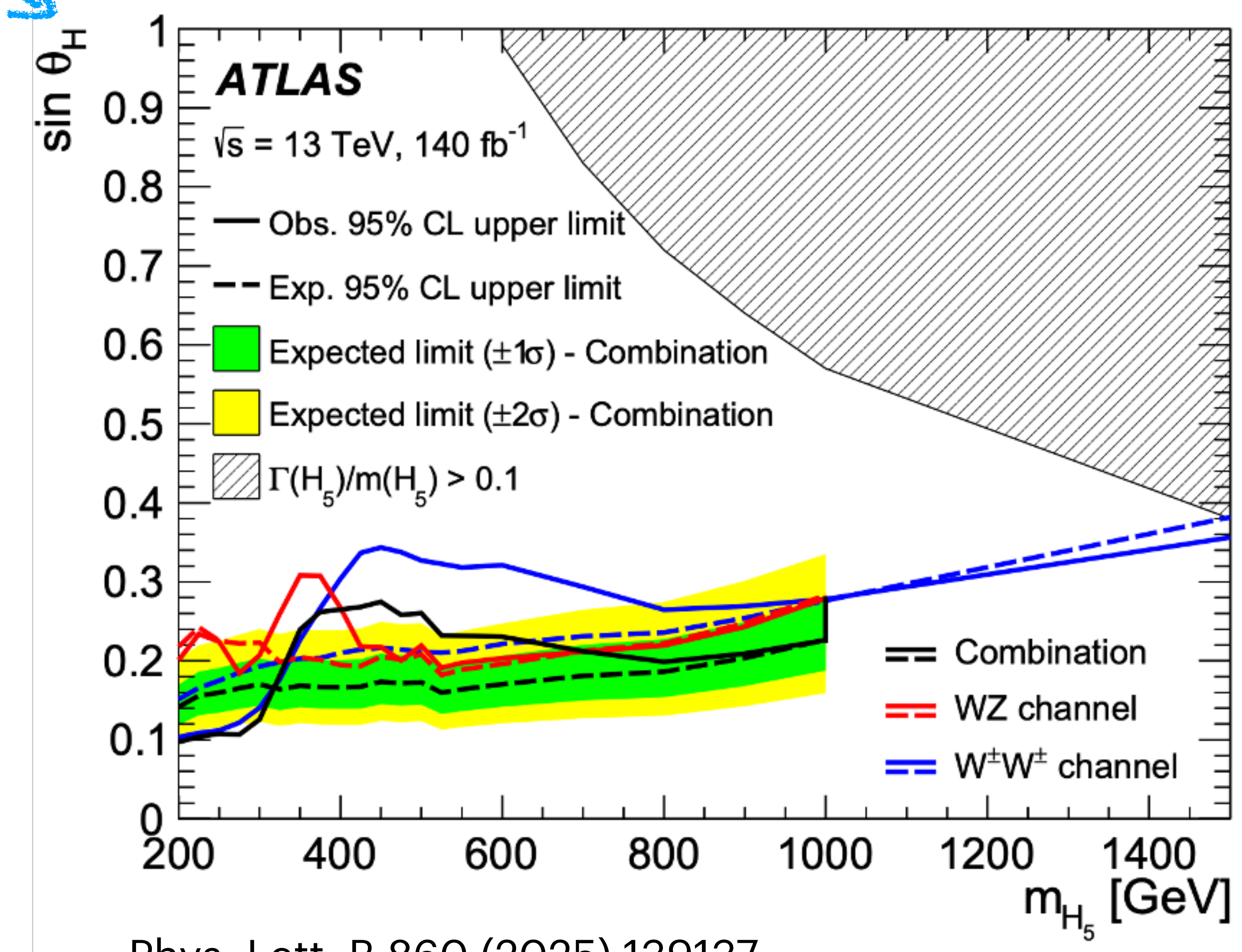
Charged Higgs Search

- The Georgi-Machacek (GM) model extends the Higgs sector by adding a quintuplet ($H_5^0, H_5^\pm, H_5^{\pm\pm}$) which are degenerate in mass (m_{H_5})[1]
- Our analysis is searching for $H_5^\pm$ and $H_5^{\pm\pm}$ produced by vector boson fusion (VBF) and decaying leptonically via $W^\pm Z$ or $W^\pm W^\pm$
- Motivated by excess observed in Run 2 in $W^\pm Z$ and $W^\pm W^\pm$ channels
- Objective is to investigate excess and improve limits by optimizing search for GM model and using Run 2 and partial Run 3 (2022-2023) data

$\sin(\theta_H)$ is a free parameter in the GM model; σ for $H_5^\pm$ and $H_5^{\pm\pm}$ production $\propto \sin^2(\theta_H)$







Phys. Lett. B 860 (2025) 139137

- Local (global) significance of 3.3 (2.5) at 375 GeV

1 Nucl. Phys. B 262 (1985) 463

$H^{\pm\pm}$ Signal Region

Major background sources:

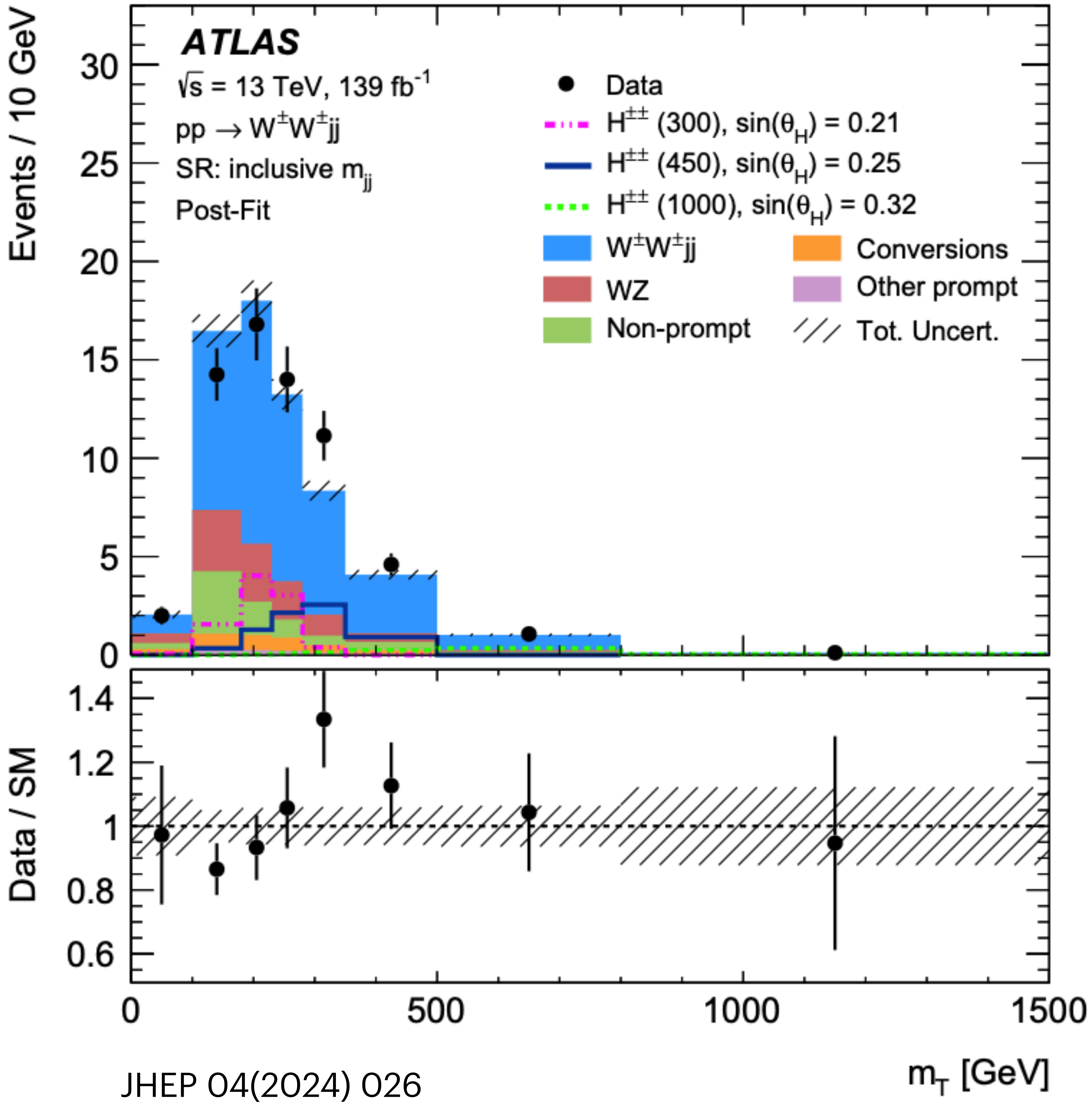
67% $W^{\pm}W^{\pm}$

15% $W^{\pm}Z$

12% Jets reconstructed as leptons i.e. "non-prompt" (V + jets and $t\bar{t}$ semileptonic)

3% Photon conversions ($V\gamma$)

2% Electrons with charge mis-ID (Z, $t\bar{t}$ dileptonic)



JHEP 04(2024) 026

$H^{\pm\pm}$ Signal Region

Major background sources:

67% $W^{\pm}W^{\pm}$

15% $W^{\pm}Z$

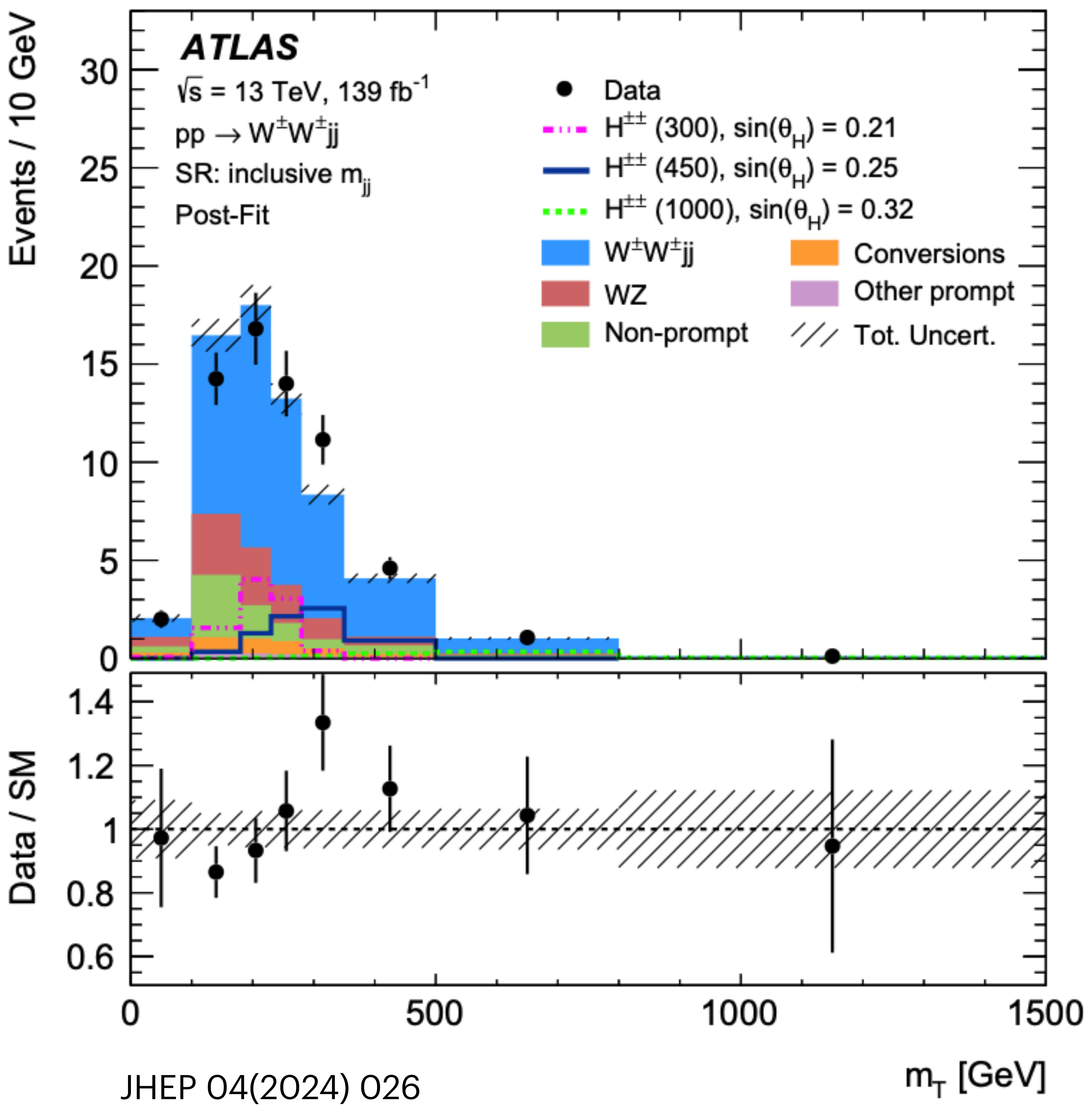
12% Jets reconstructed as leptons i.e. "non-prompt" (V + jets and $t\bar{t}$ semileptonic)

3% Photon conversions ($V\gamma$)

2% Electrons with charge mis-ID (Z, $t\bar{t}$ dileptonic)

✴ Goal: Improve the signal region (SR) selection to improve our sensitivity to $H^{\pm\pm}$

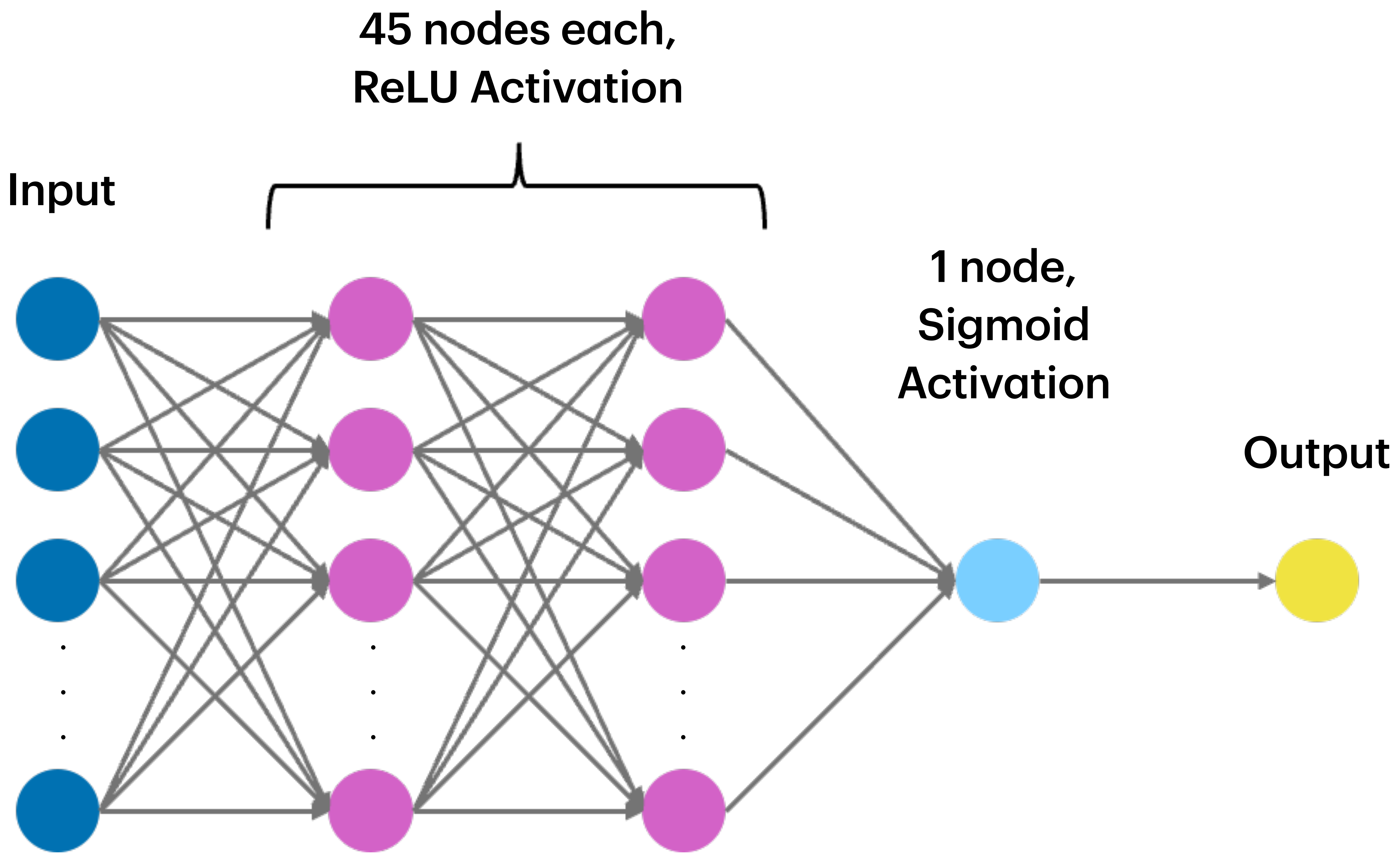
- Complex problem with large number of labelled simulation events is ideal for ML



JHEP 04(2024) 026

Neural Network Model

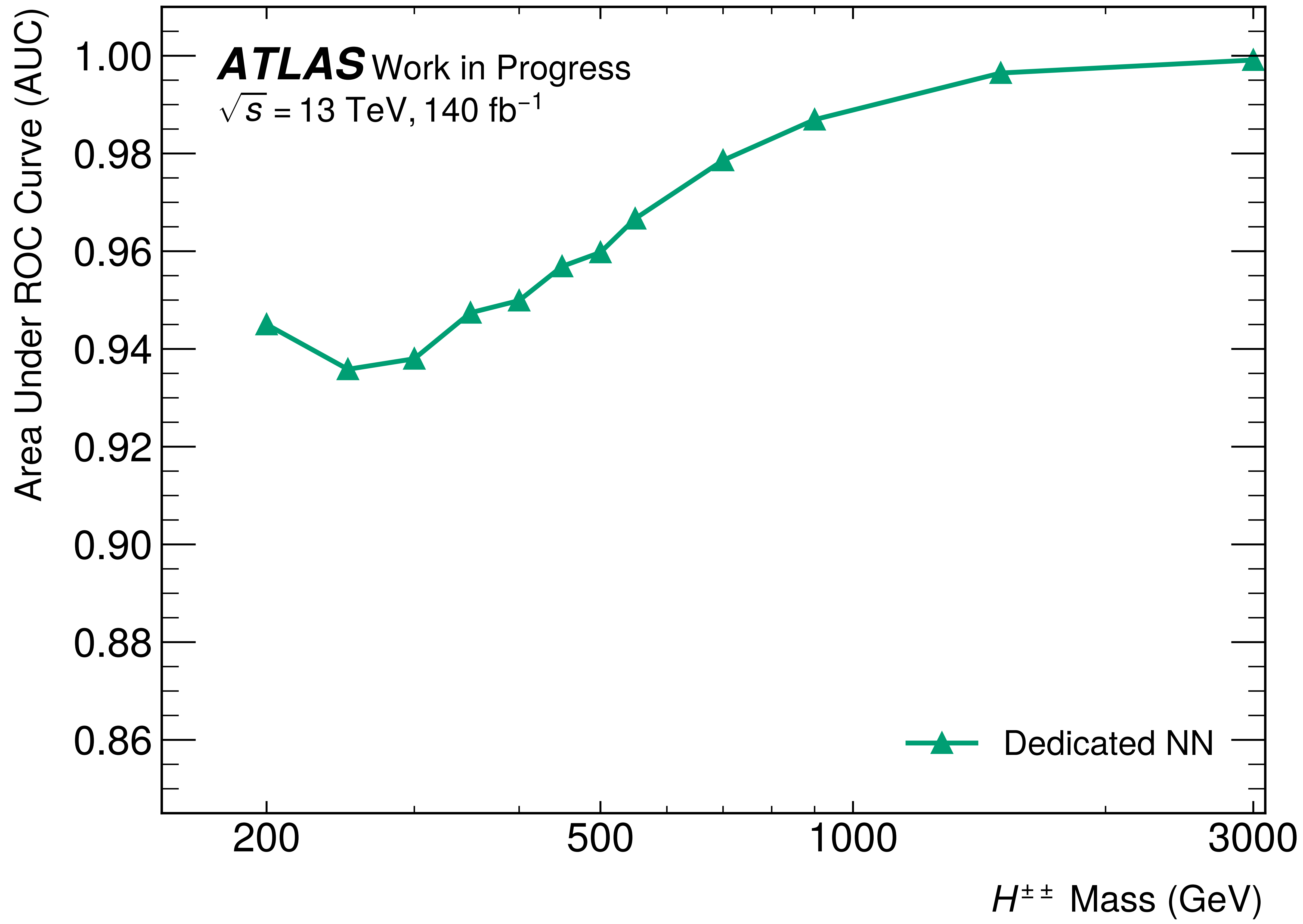
- Neural network (NN) trained to classify each event as signal (1) or background (0)
- Simulated events from Run 2 which pass a basic SR cut prior to training
- Signal – 277,393 events
  - $H^{\pm\pm}$ produced via VBF
- Backgrounds – 879,119 events
  - EW, QCD & interference $W^{\pm}W^{\pm}$
  - EW, QCD & interference $W^{\pm}Z$
- Balance signal & background using class weights



✳Batch normalization between layers

Mass Problem

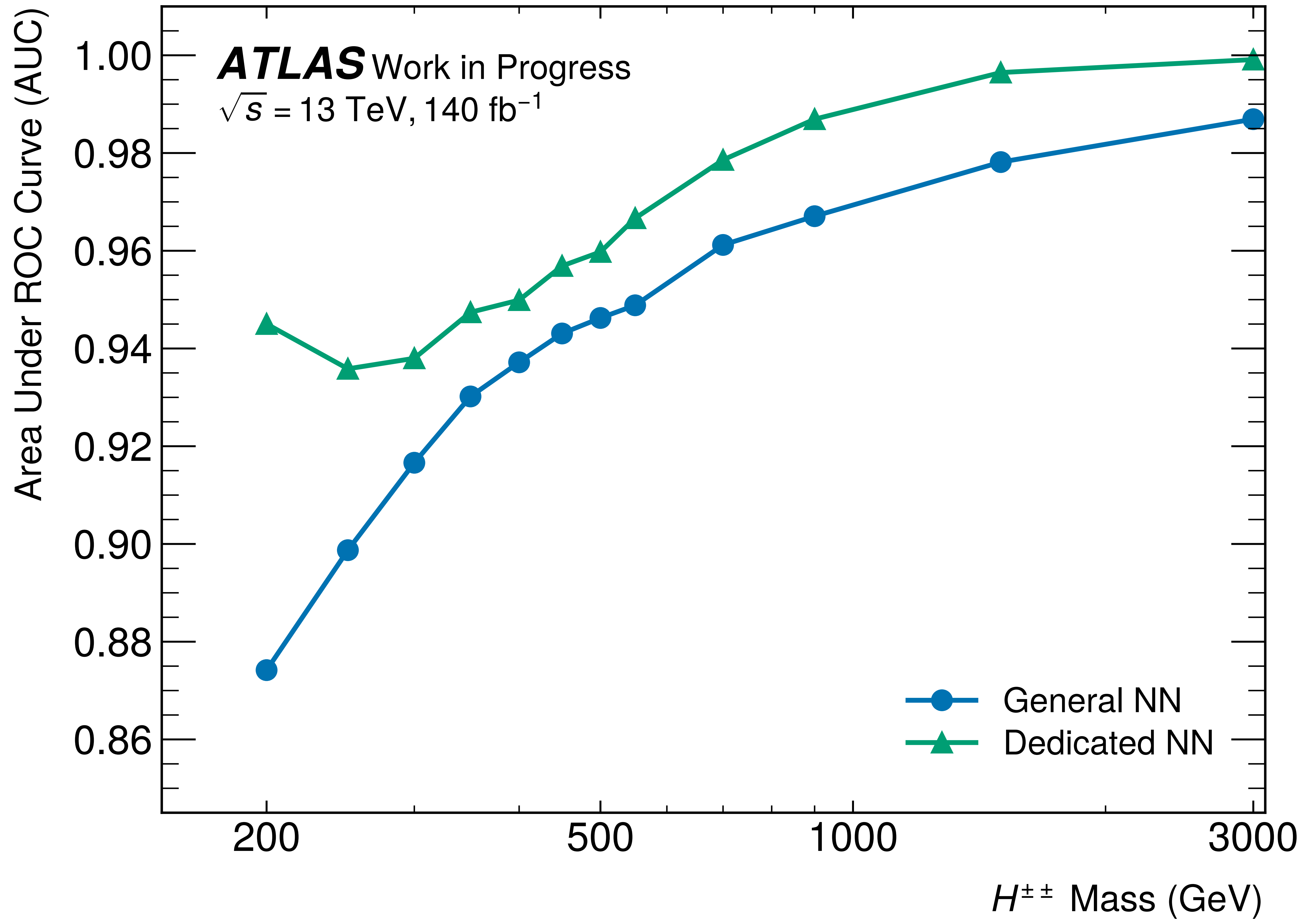
- Ideally, we would train a dedicated NN for every simulated mass point



Mass Problem

- Ideally, we would train a dedicated NN for every simulated mass point

- However, it is much more practical to use a single general NN for the entire mass region

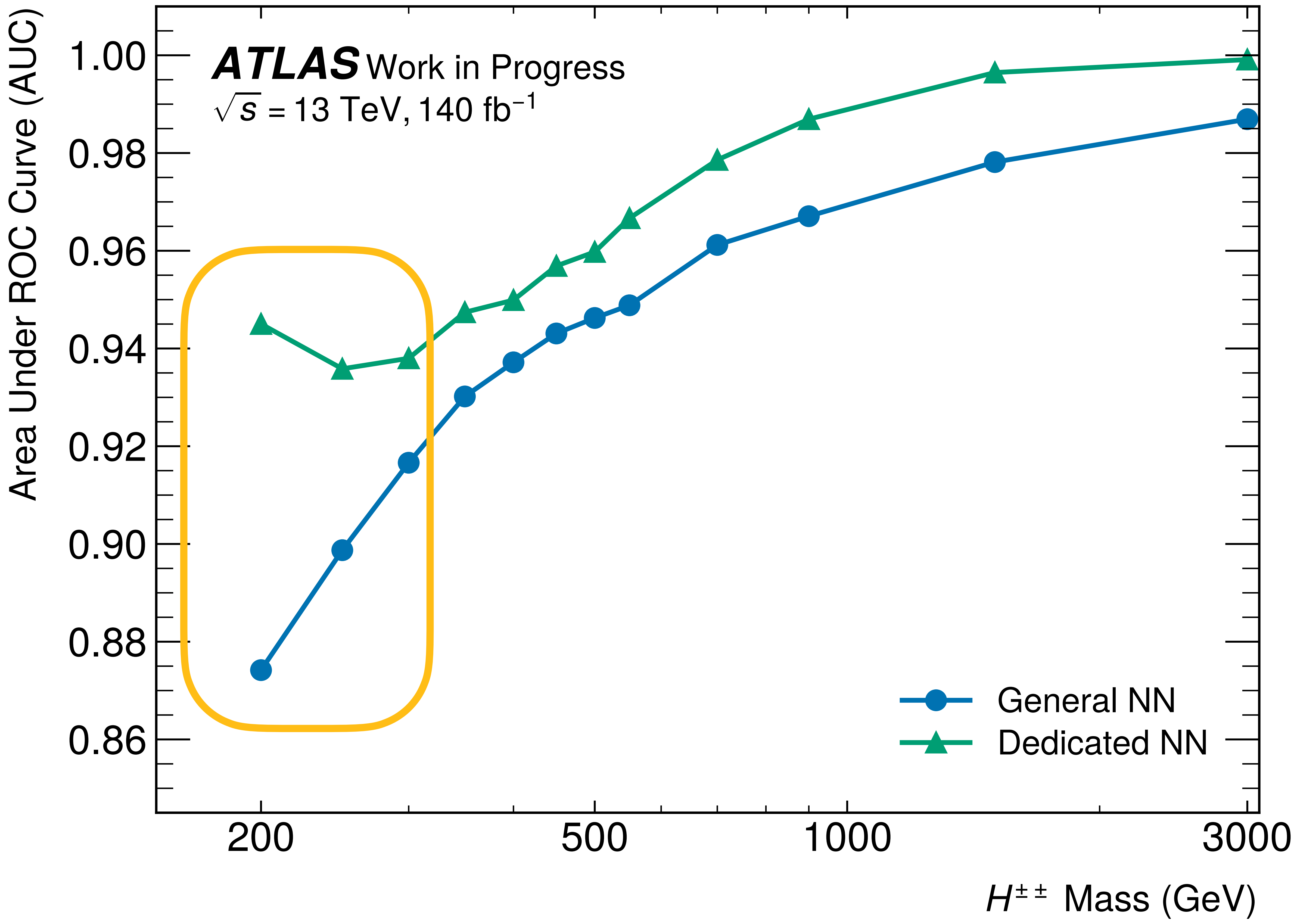


Mass Problem

- Ideally, we would train a dedicated NN for every simulated mass point

- However, it is much more practical to use a single general NN for the entire mass region

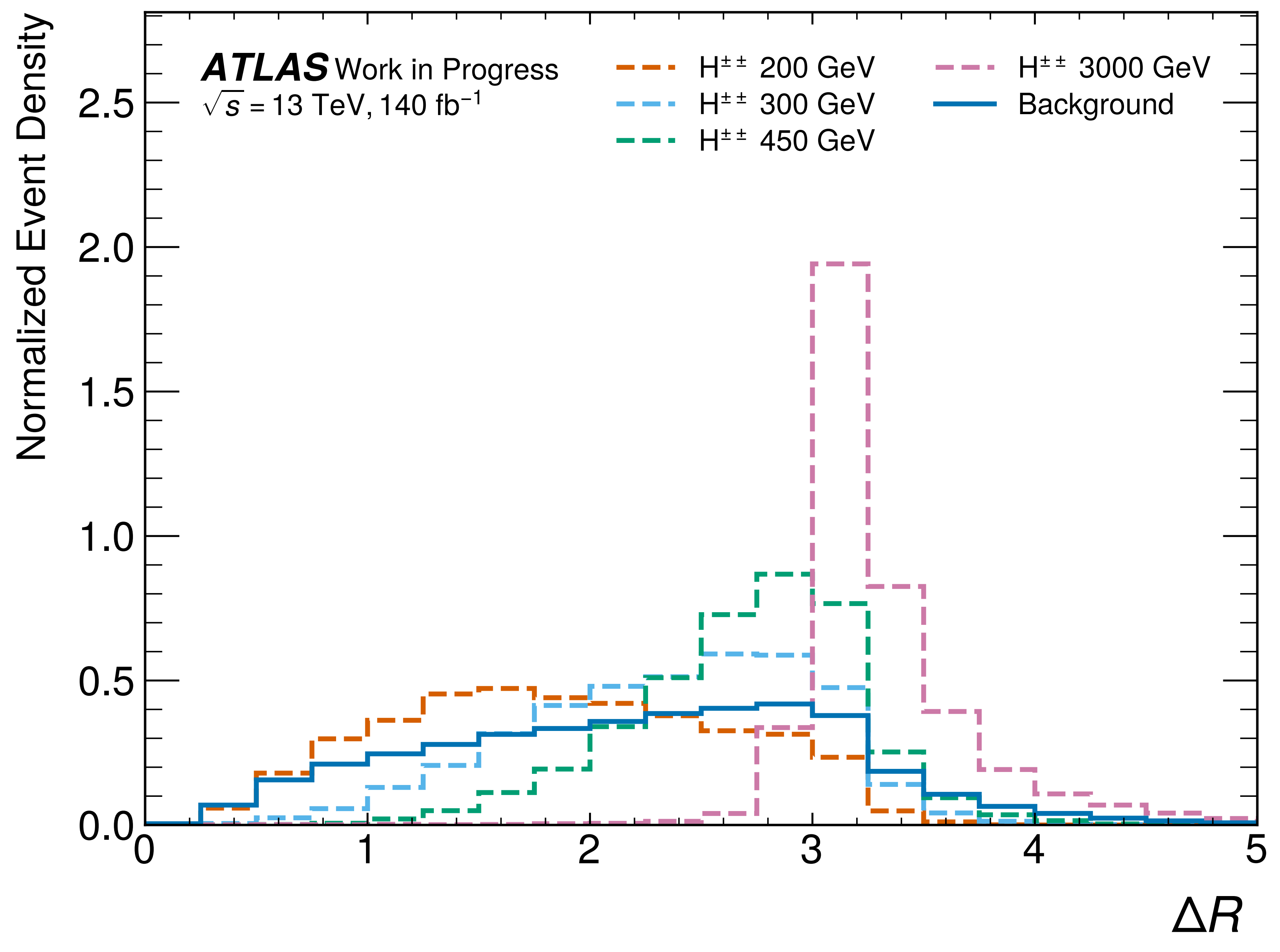
✴ Problem: training a NN with a range of signal samples (200 - 3000 GeV) produces poor results at low mass



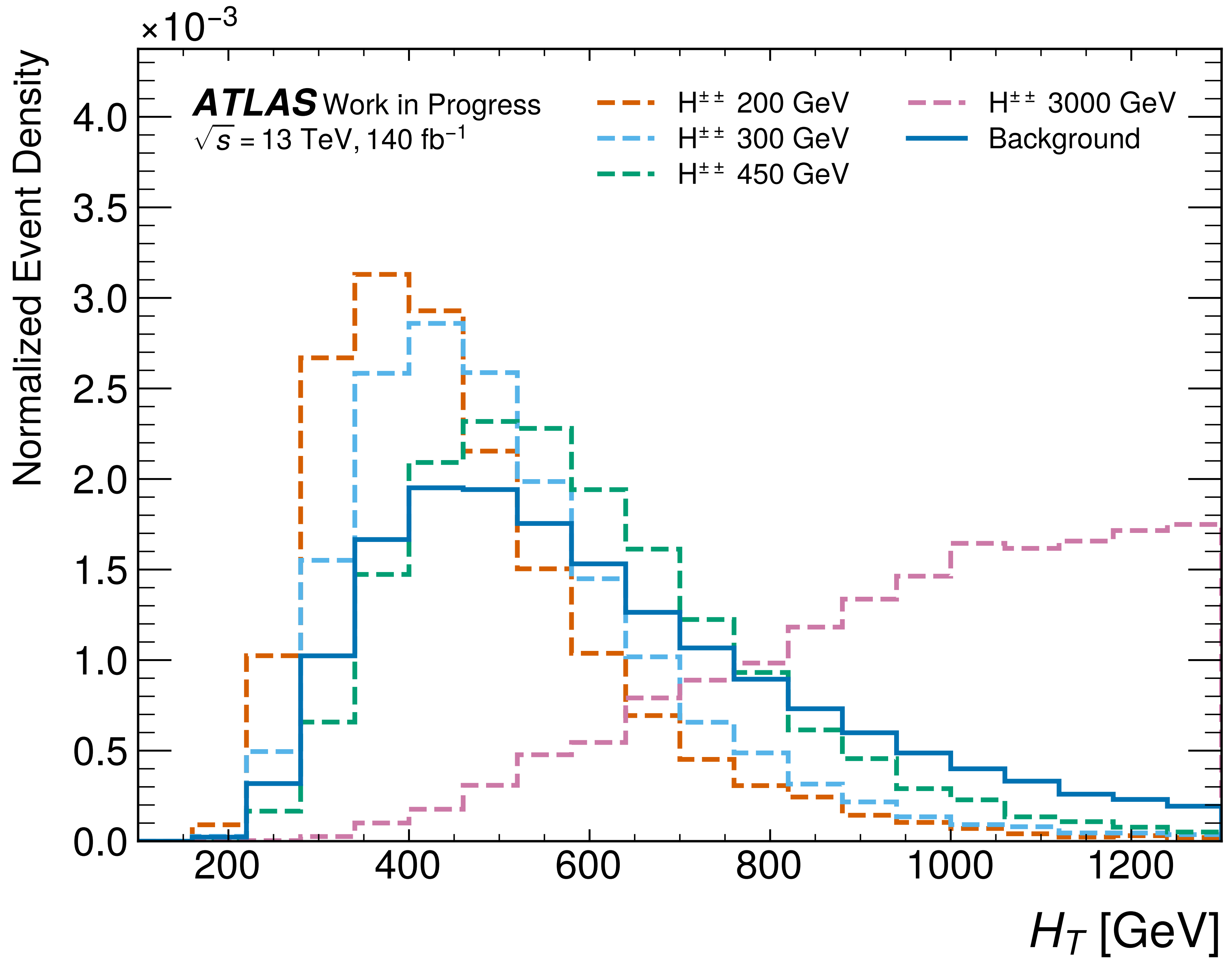
Mass Problem

Why does this happen?

- Kinematics vary significantly with increasing $H^{\pm\pm}$ mass
- For some NN inputs, low mass trends opposite to high mass relative to background
- Low mass signal is more similar to background than high mass signal



$$\Delta R_{ll} = \sqrt{(\Delta\eta_{ll})^2 + (\Delta\phi_{ll})^2}$$



H_T = sum of p_T of all event objects

Event Weighting Solution

- Solution: Modify how different mass points are weighted in the loss function so that the NN performance is more uniform across the mass range

Event Weighting Solution

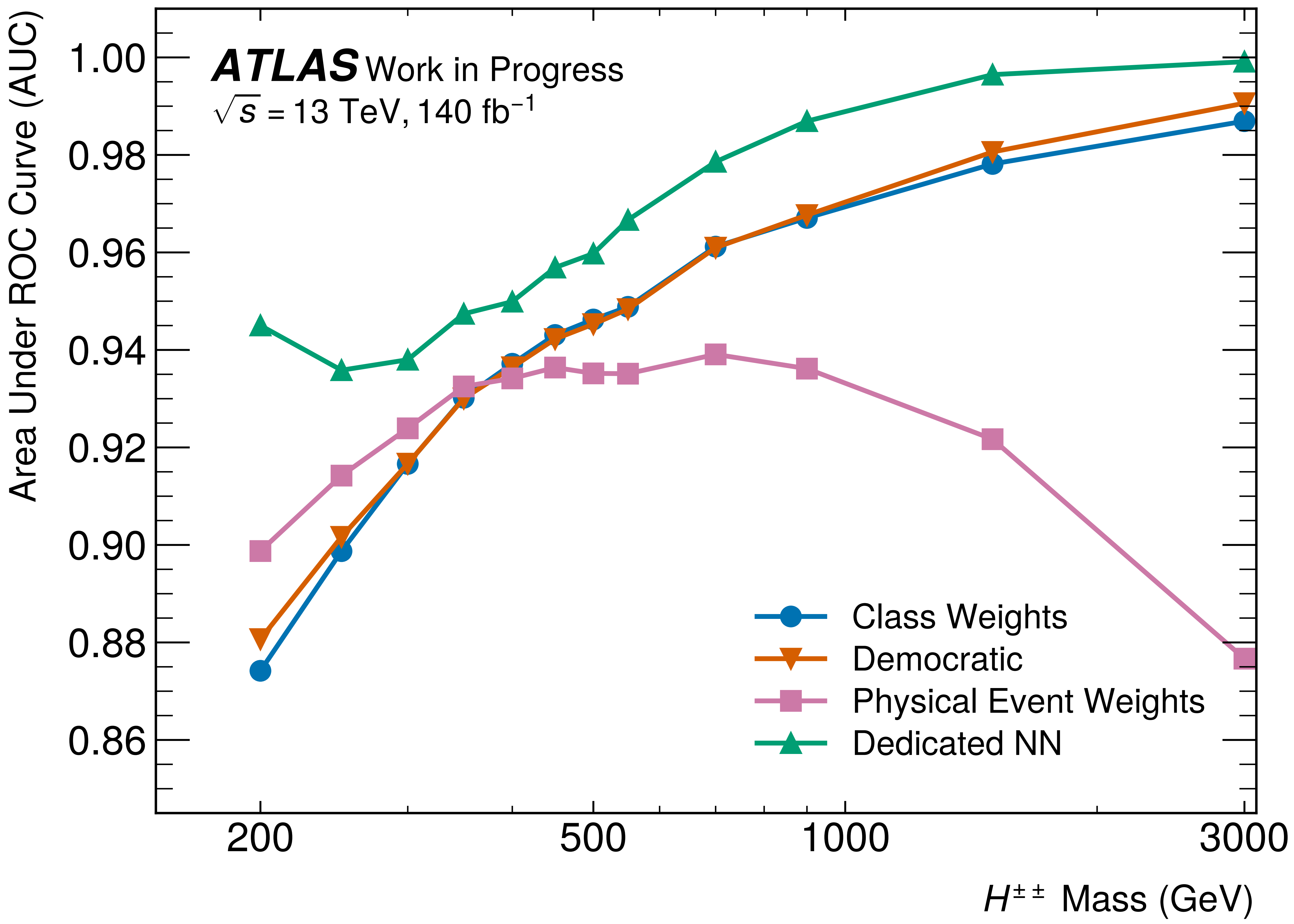
✴Solution: Modify how different mass points are weighted in the loss function so that the NN performance is more uniform across the mass range

Two new approaches:

1. **Democratic Signal Class Weights**
   - Assign signal weights so that **each signal sample has equal sum of weights**
   - Eliminates dependence on number of events per mass point sample
2. **Physical Event Weights**
   - Proportional to **Monte Carlo weight** and **cross section**
   - Encourages network to focus on events which are more likely to be found in data

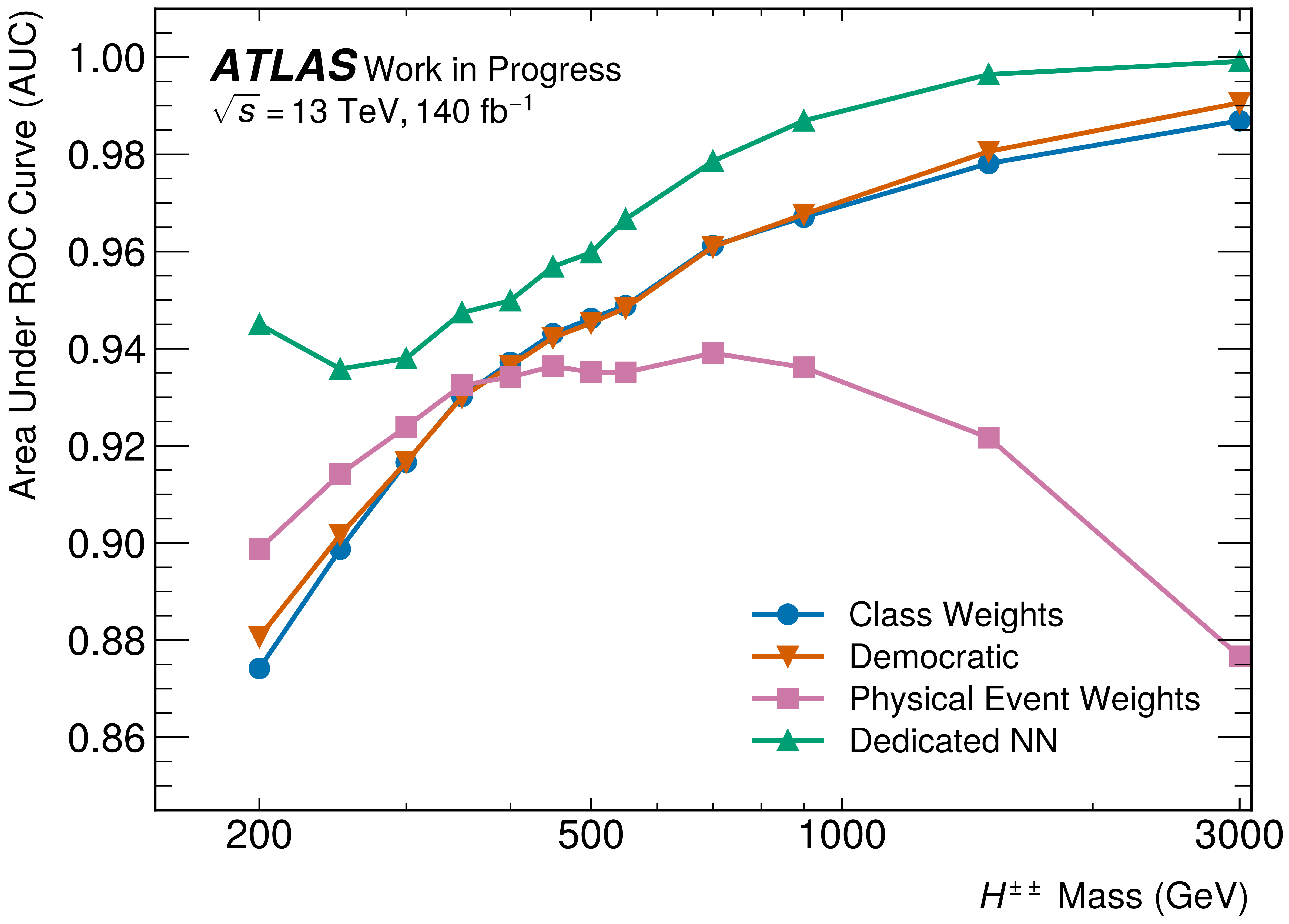
Event Weighting Results

- Democratic class weights slightly improve performance at low mass
  - *Low mass samples have less events*



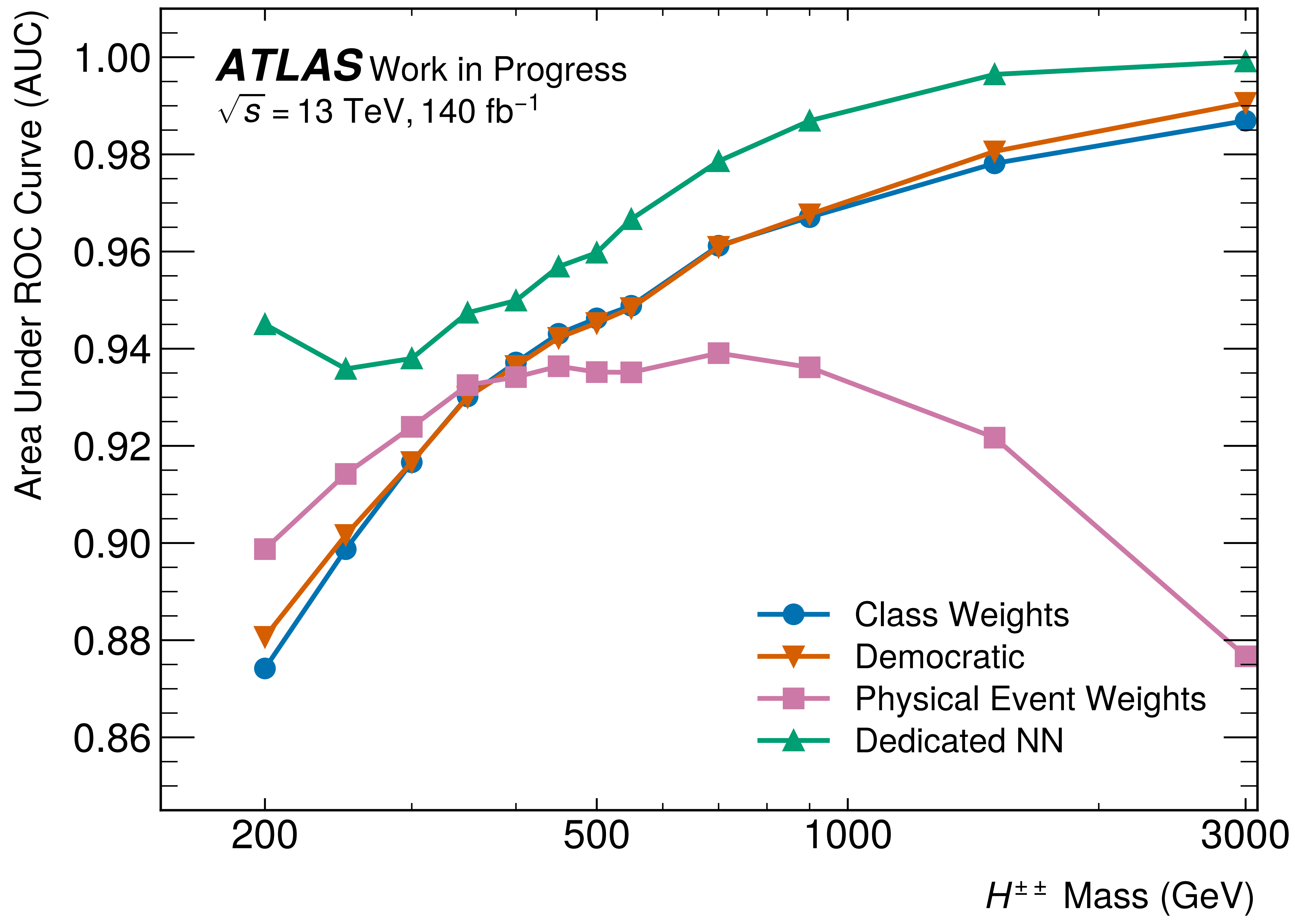
Event Weighting Results

- Democratic class weights slightly improve performance at low mass
  - *Low mass samples have less events*
- Physical event weights significantly improve performance at low mass
  - *Cross-section decreases with mass*



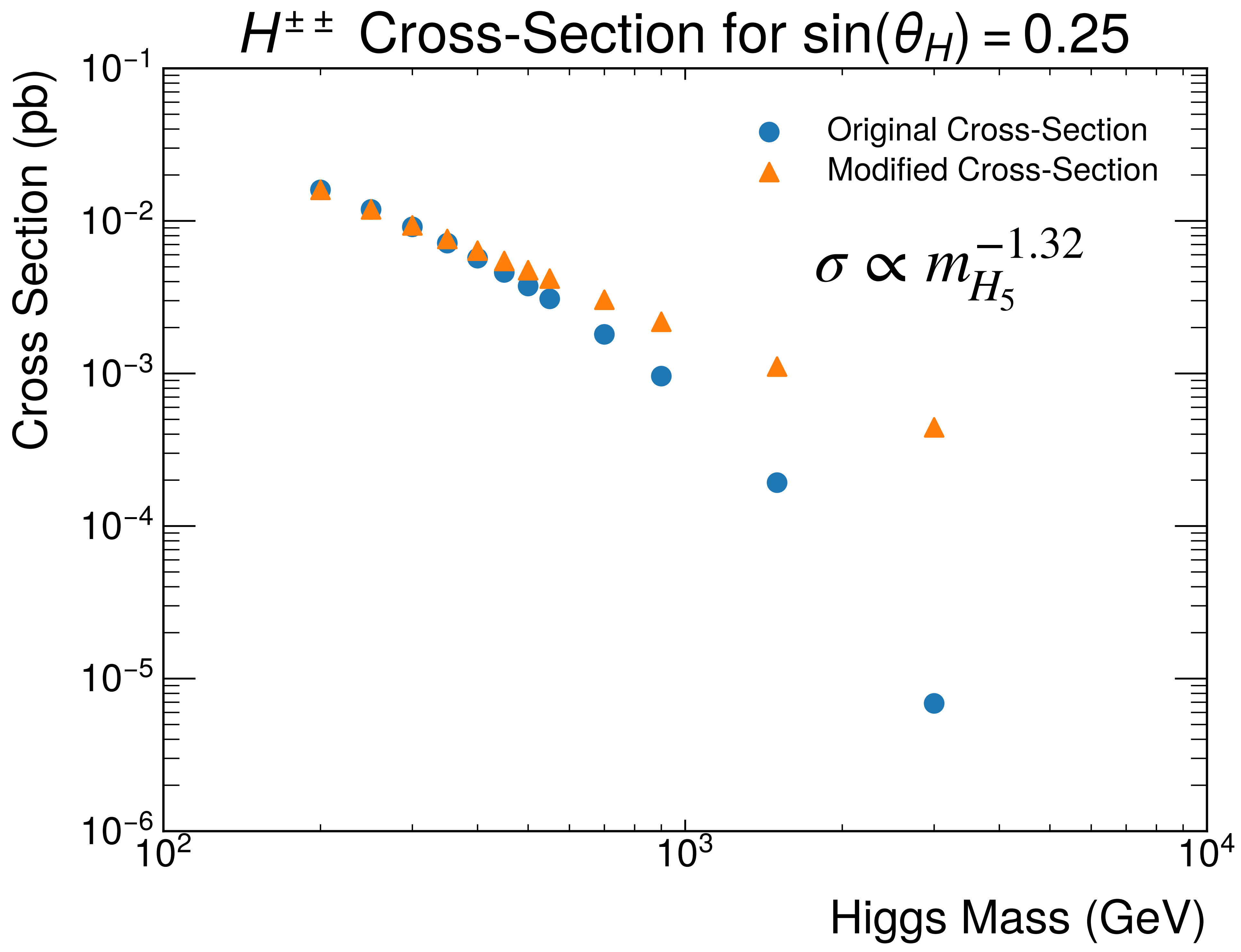
Event Weighting Results

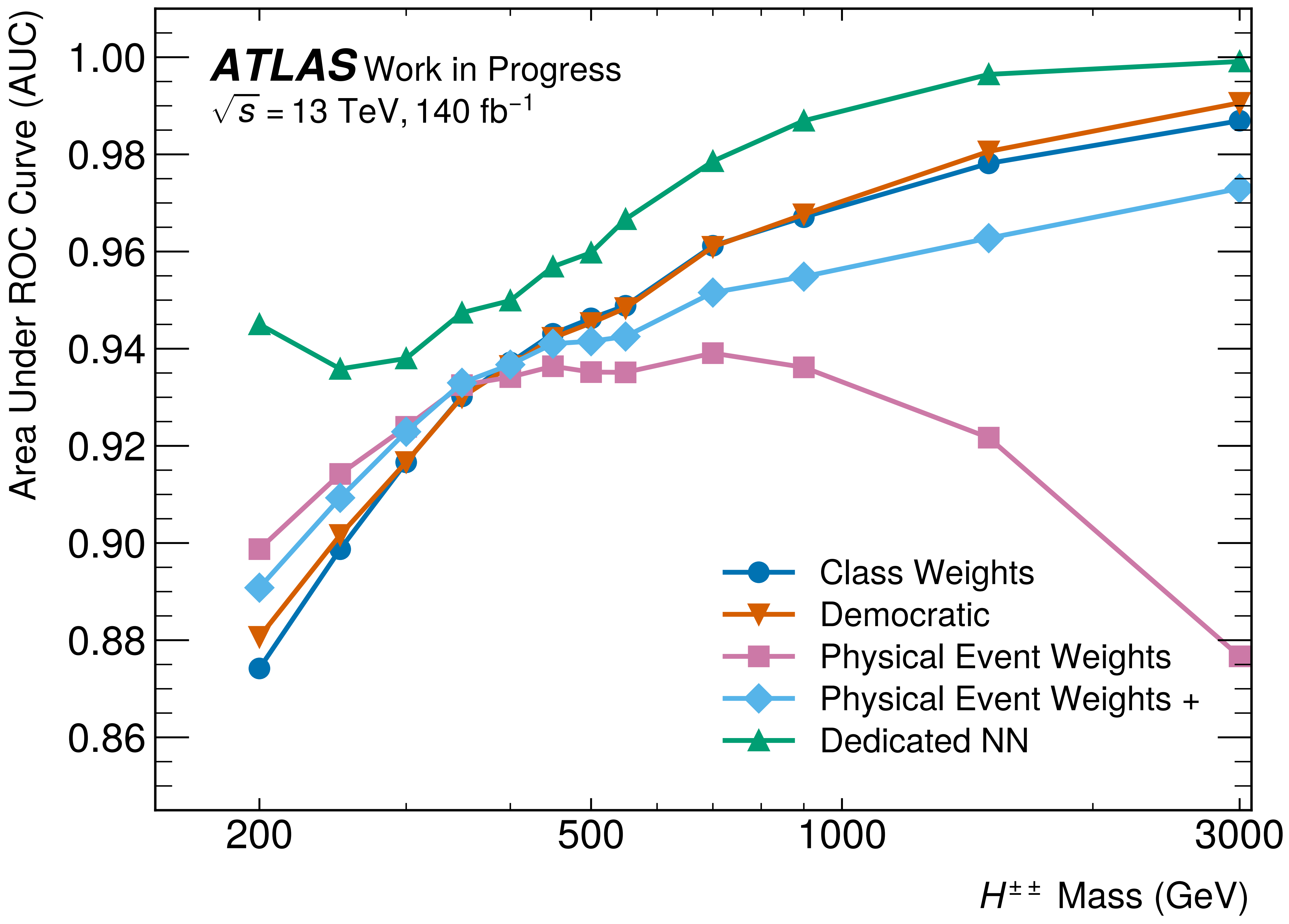
- Democratic class weights slightly improve performance at low mass
  - *Low mass samples have less events*
- Physical event weights significantly improve performance at low mass
  - *Cross-section decreases with mass*
- Physical event weights significantly reduce performance at high mass
  - *Cross-section at 3000 GeV is 10^3 smaller than at 200 GeV*



Modification to Physical Event Weights

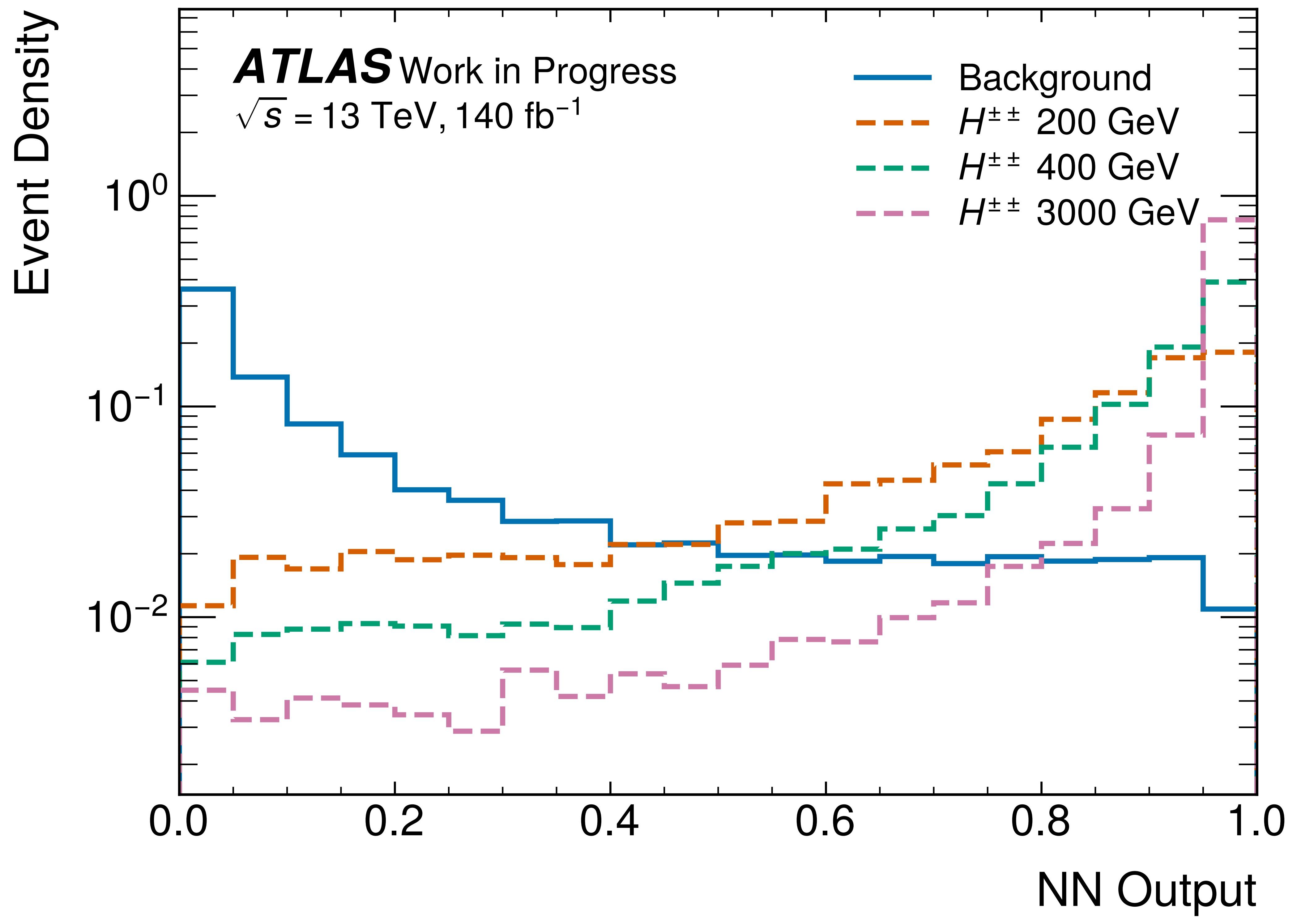
- If we modify the event weight with an effective cross section that decreases with mass through a power law, we recover performance at high mass

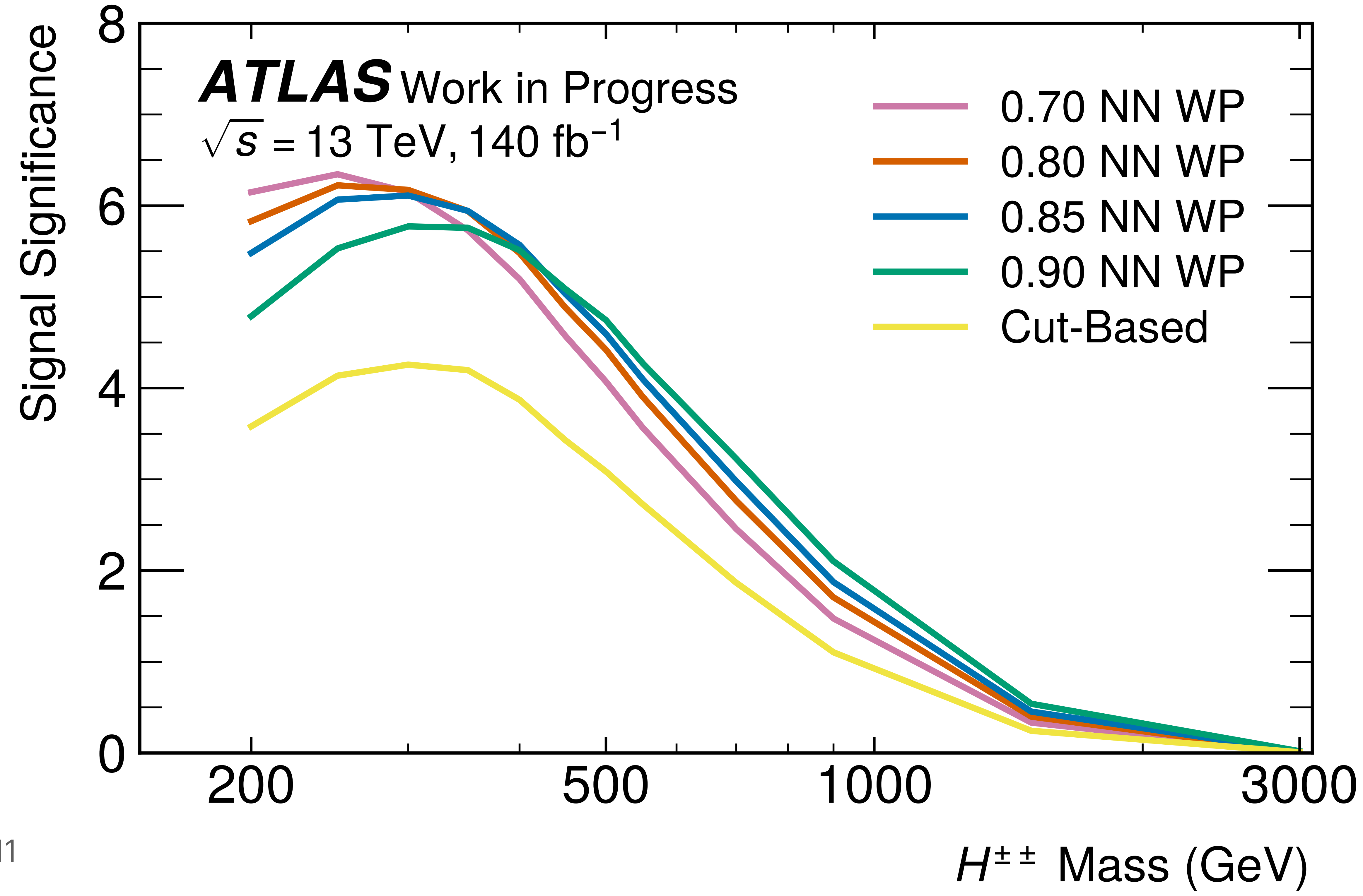




Final NN - Physical Event Weights+

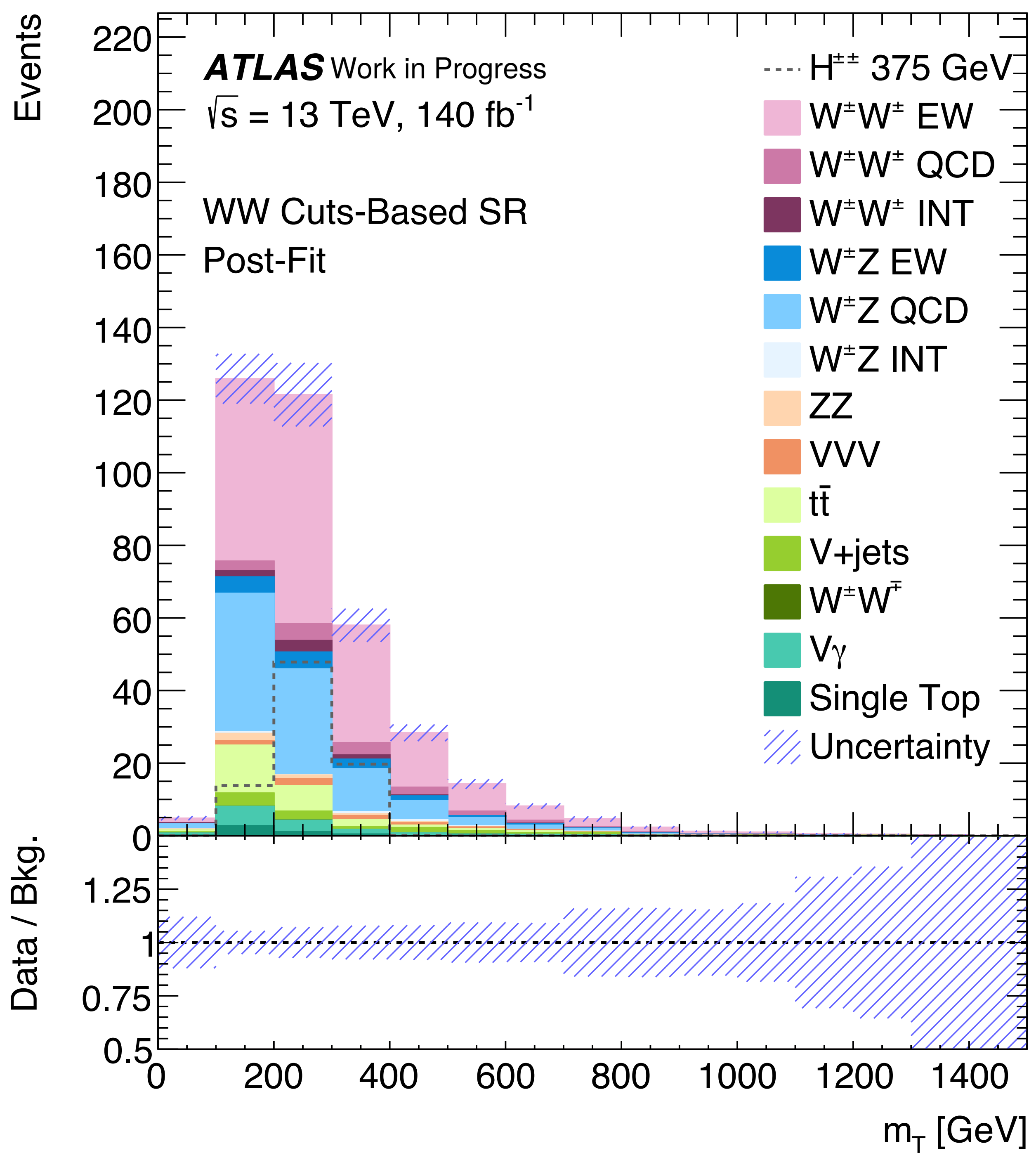
- NN outperforms cuts-based selection at all mass points in terms of SR significance
- Ideal choice of NN working point (WP) depends on mass
  - High mass distribution peaks more strongly at 1

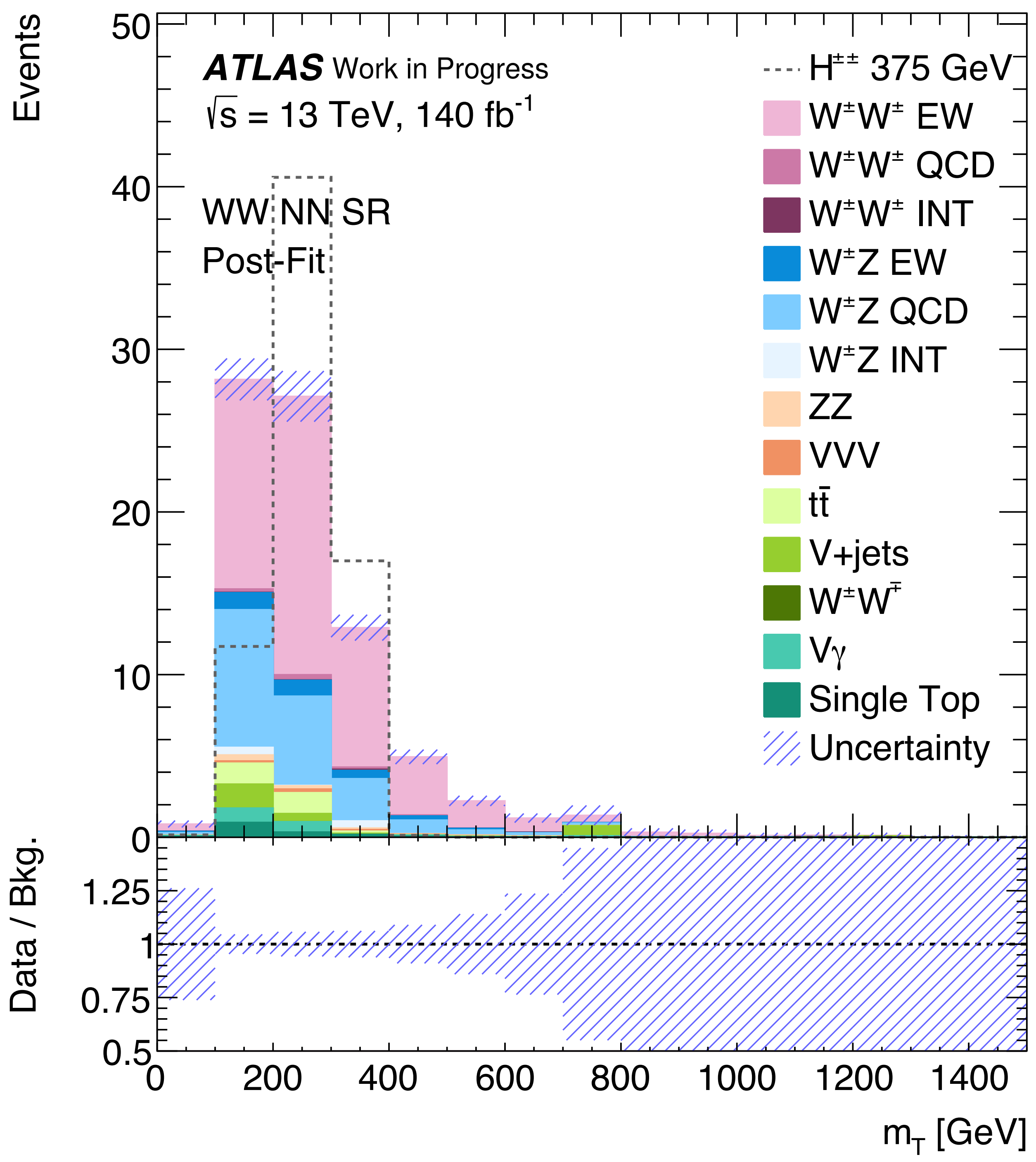




Impact on Limits

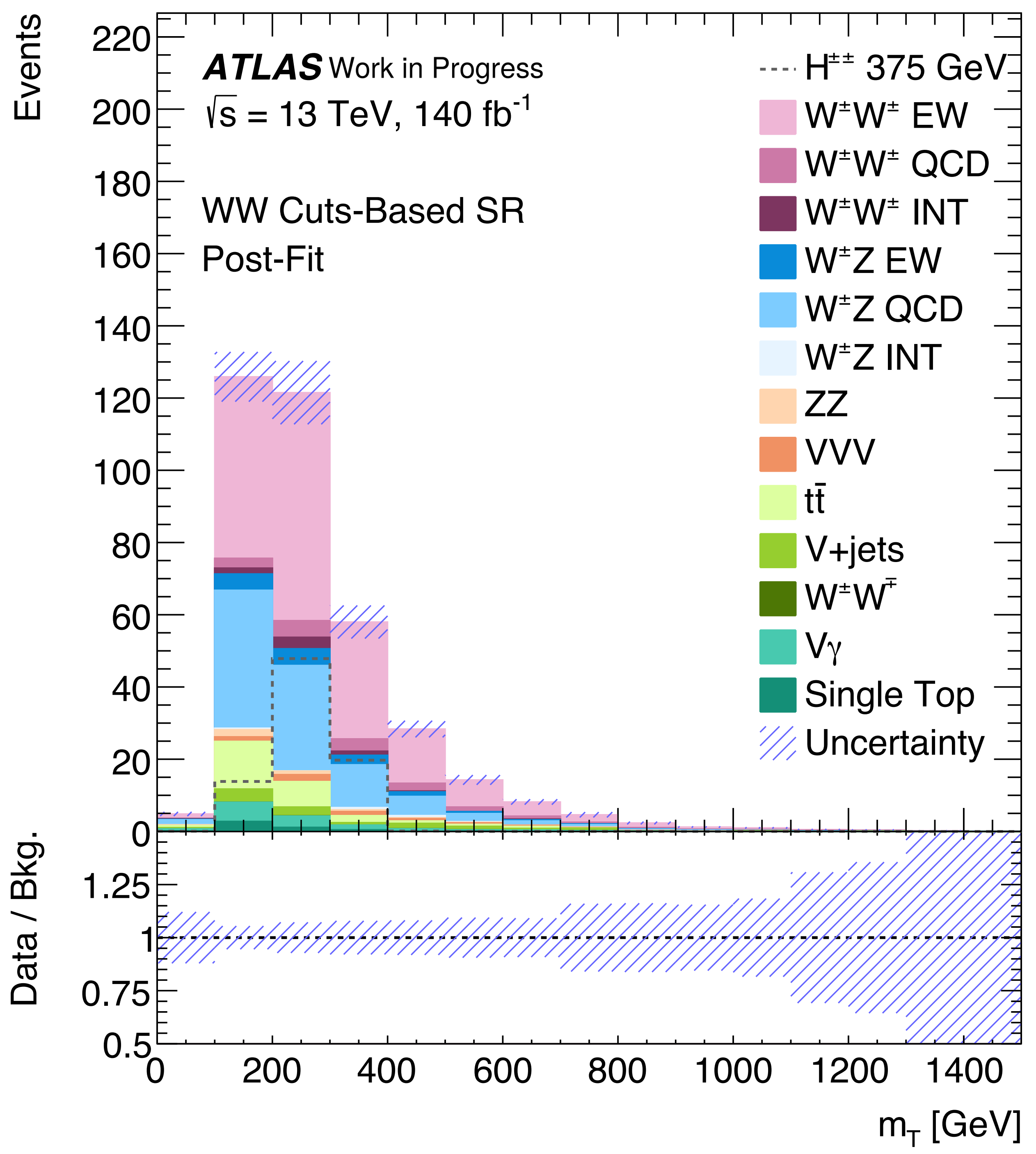
- Asimov fit increases signal region purity in NN SR (NN score > 0.85) compared to the cuts-based SR for $\sin(\theta_H) = 0.25$

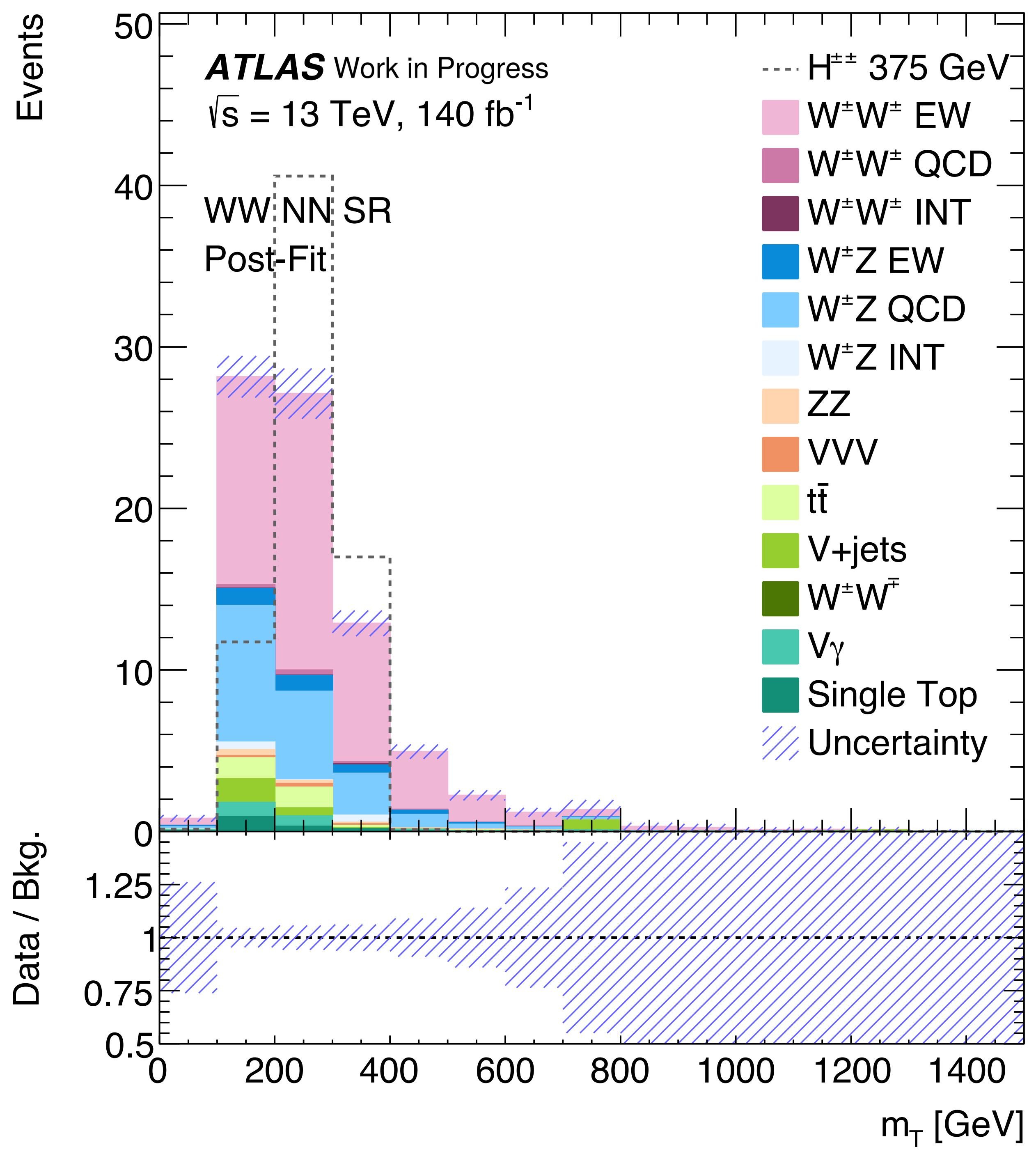


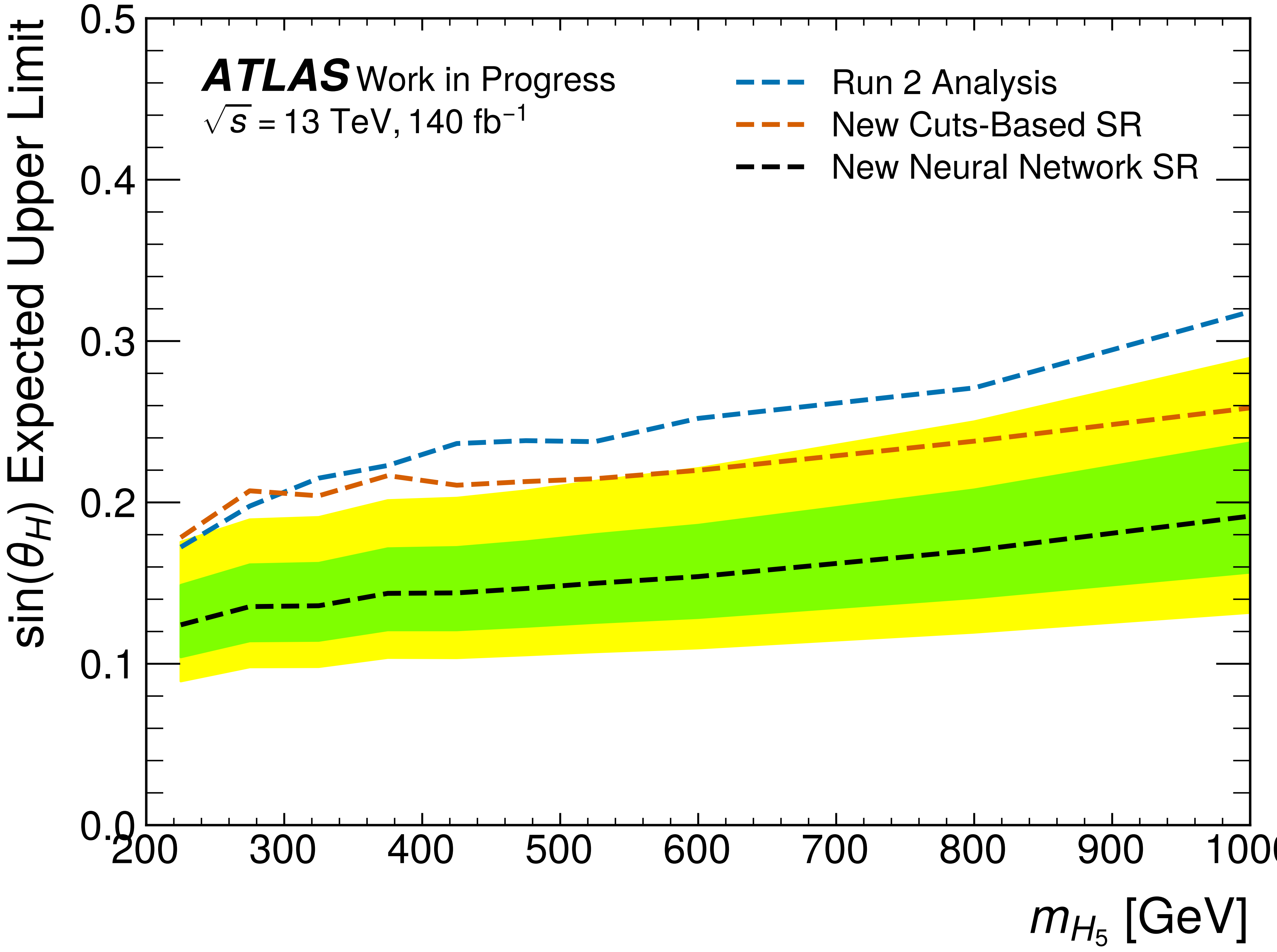


Impact on Limits

- Asimov fit increases signal region purity in NN SR (NN score > 0.85) compared to the cuts-based SR for $\sin(\theta_H) = 0.25$
- NN SR improves expected upper limits on $\sin(\theta_H)$
  - *Evaluating only on mass points not seen in training*







Conclusion

- Classifying events with a single NN across a range of signal masses is challenging due to the variability of signal characteristics
- Strategically weighting mass points can improve the consistency of NN classification across the mass range
- Simple NNs are an invaluable tool for maximizing our sensitivity to extended Higgs sectors and offer a rich environment for exploring ML behaviour

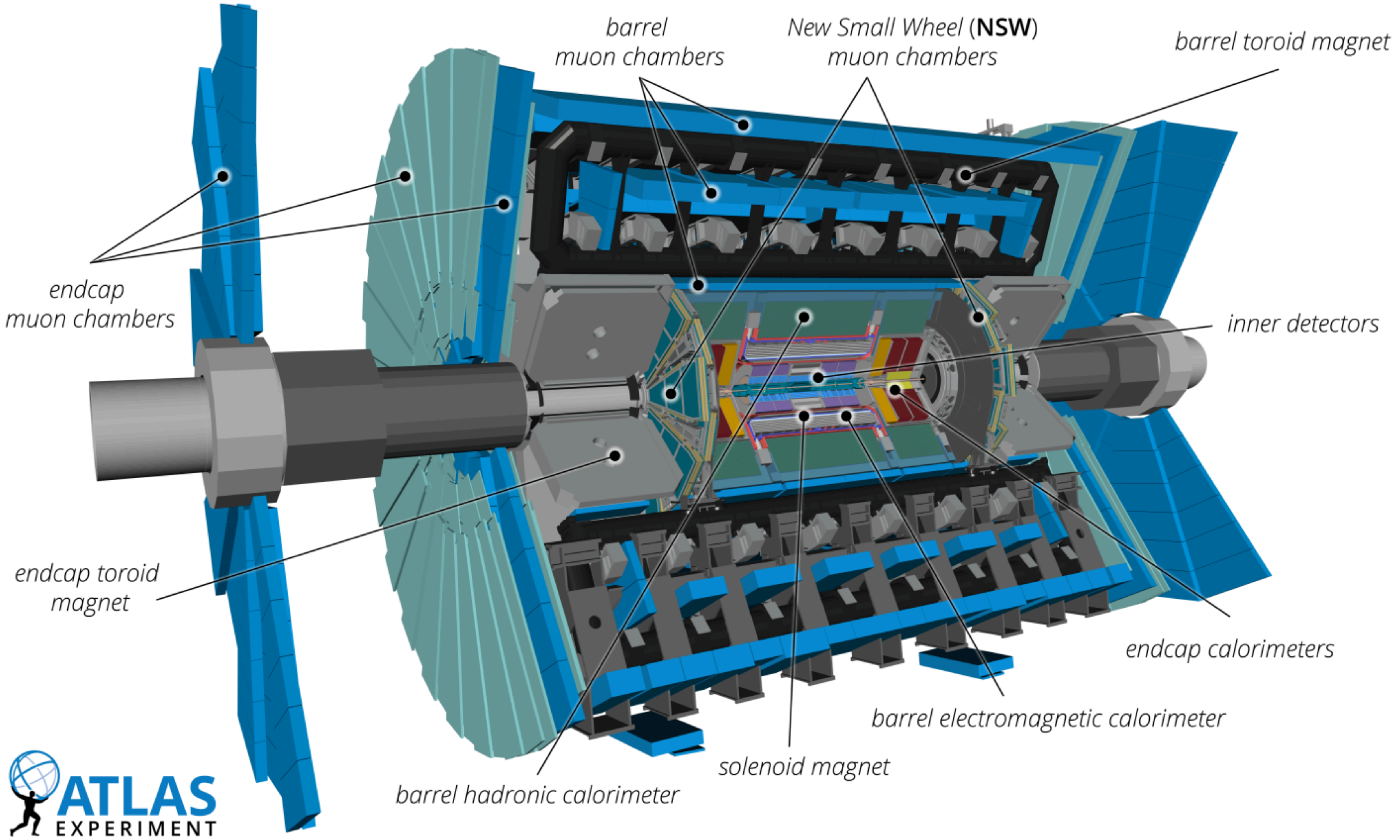
Backup: The ATLAS Experiment

ATLAS collects data from proton-proton collisions at the LHC

- Run 2 (2015-2018): $\sqrt{s}$ = 13 TeV, 140 fb^{-1}
- Run 3 (2022-2026): $\sqrt{s}$ = 13.6 TeV, 183 fb^{-1} so far
- In this talk, only using Run 2 simulated events

Events are reconstructed using:

- Charged particle tracks in the inner detector
- Energy deposits in the calorimeters
- Hits in the muon spectrometer



Backup: Basic SR Selection for NN Training

Basic SR Selection
Exactly two same-sign signal leptons with p_T > 27 GeV (|η| < 1.37 in *ee* channel)
m_{ll} > 20 GeV
Z peak removal (m_{ee} more than 15 GeV separated from m_Z)
At least 2 jets with jet 1 p_T > 65 GeV & jet 2 p_T > 35 GeV *
MET > 30 GeV *
Pass 3 lepton veto (less than 3 baseline leptons)

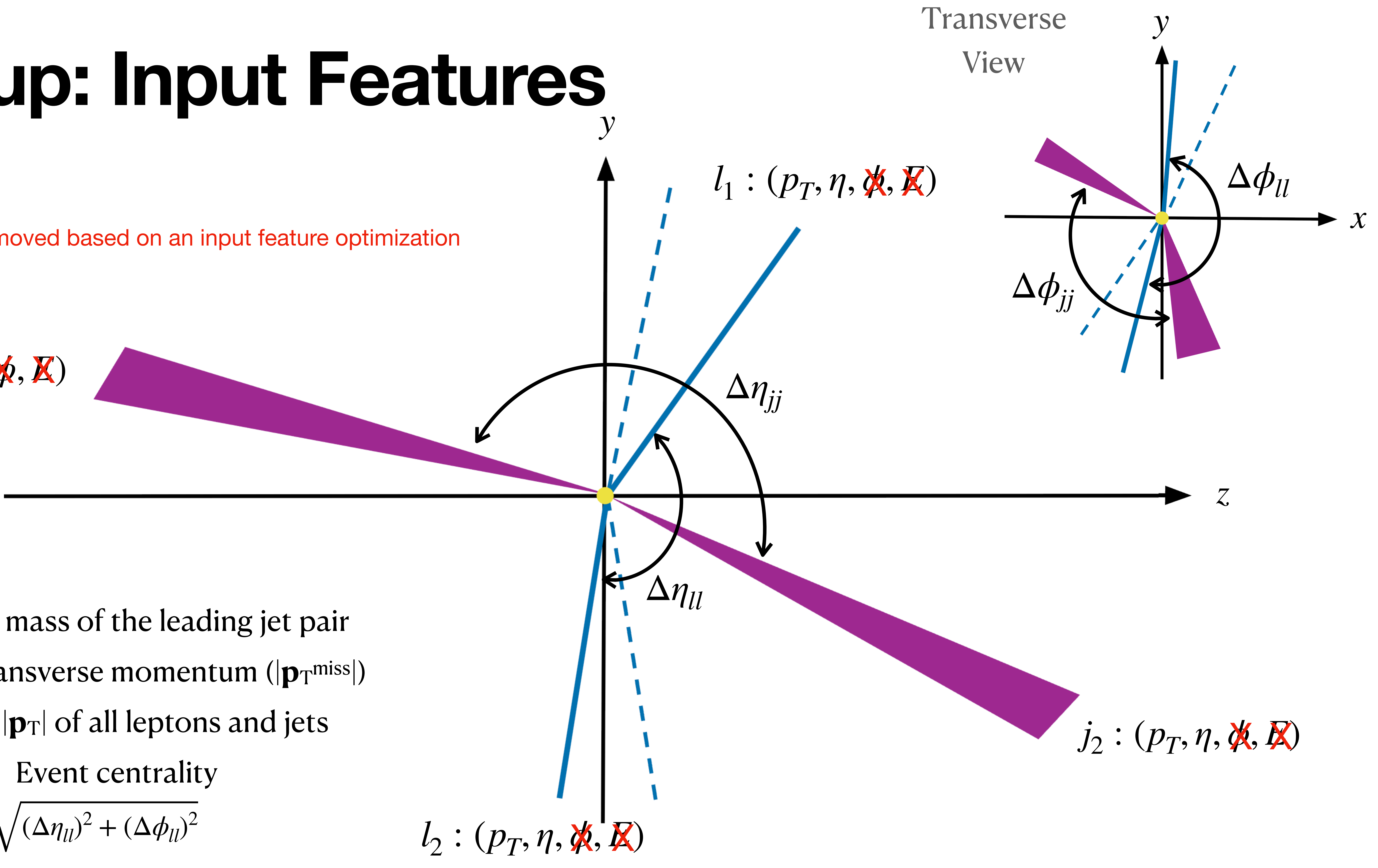
Additional Selections for Cuts-Based SR
$\|\Delta\phi_{ll}\|$ > 1.5
$\|\Delta y_{jj}\|$ > 2
m_{jj} > 500 GeV

Additional Selections for NN SR
NN Score > 0.8

* For the results on slide 12 & 13, the p_T cut for jet 1 was reduced to 50 GeV, the p_T cut for jet 2 was reduced to 30 GeV, the *MET* cut was reduced to 25 GeV, an m_{jj} > 200 GeV cut was applied, and a b-veto was required

Backup: Input Features

ϕ and E were removed based on an input feature optimization



m_{jj}	Invariant mass of the leading jet pair
E_T^{miss}	Missing transverse momentum ($\lvert\mathbf{p}_T^{\text{miss}}\rvert$)
H_T	Sum of $\lvert\mathbf{p}_T\rvert$ of all leptons and jets
ξ	Event centrality
ΔR_{ll}	$\sqrt{(\Delta\eta_{ll})^2 + (\Delta\phi_{ll})^2}$

Backup: Neural Network

The BCE loss function is defined as follows:

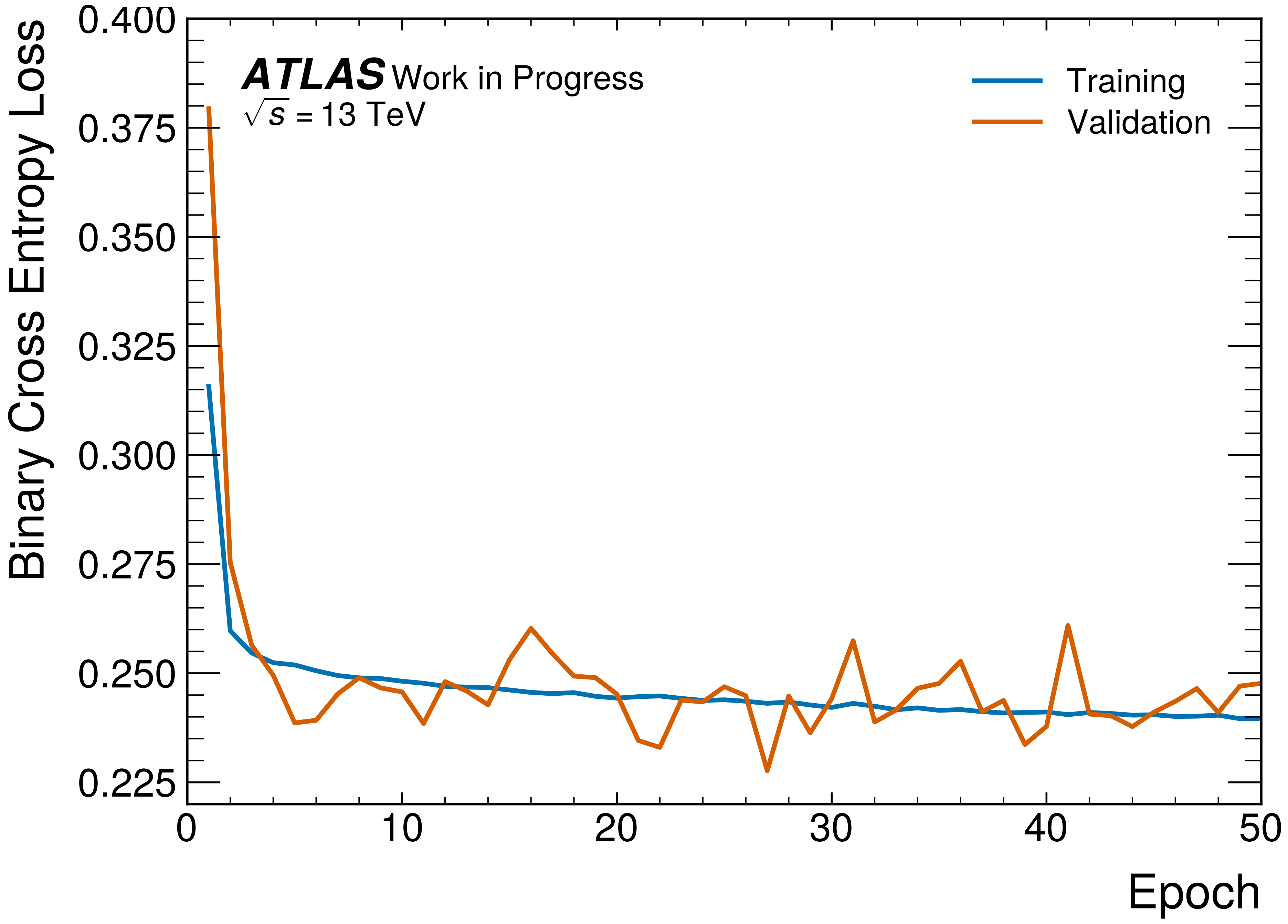
$$BCE(p) = -\sum_{i}^{N} w_i[(1-y_i)\log(1-p(y_i)) + y_i \log(p(y_i))]$$

Where w_i is the event weight, y_i is the label, and $p(y_i)$ is the NN prediction.

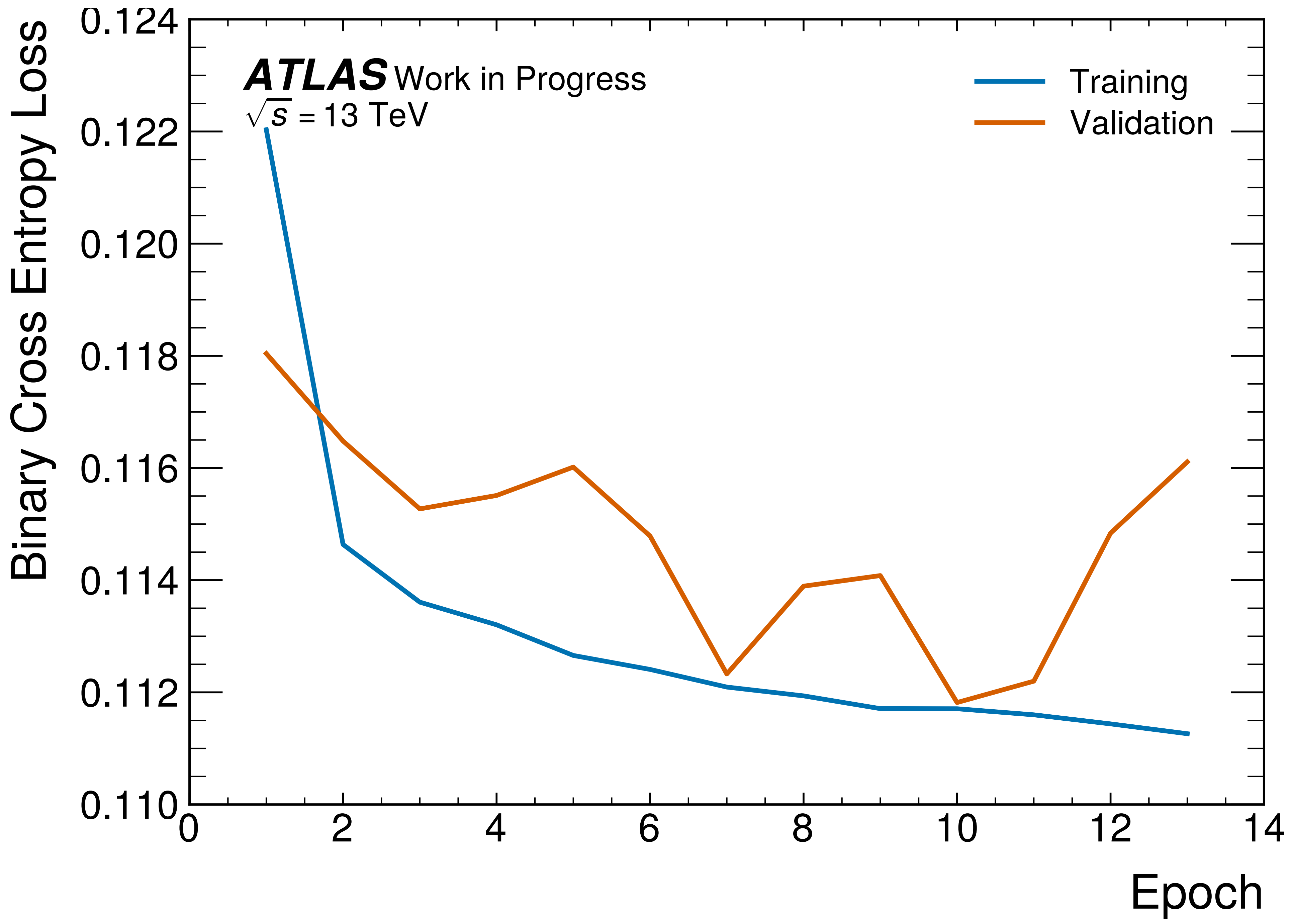
Hyperparameters	
Optimizer	Adam
Learning Rate	0.001
Loss Function	BCE
Epoch #	50
Batch Size	1000
Parameter #	~3000
*Early Stopping min. Δ	0.0002
*Early Stopping Patience	3

* Only used with event weights, not class weights

Loss Curve for NN Trained with Class Weights



Loss Curve for NN Trained with Physical Event Weights +



Backup: Event Weighting Equations

Class Weights:

$$f_s = \frac{N_s + N_b}{2N_s}, f_b = \frac{N_s + N_b}{2N_b}$$

Democratic Signal Class Weights:

$$f_{s,i} = \frac{N_s}{nN_{s,i}} \frac{N_s + N_b}{2N_s}, f_b = \frac{N_s + N_b}{2N_b}$$

Physical Event Weights:

1. For each event the initial weight is MC Weight × Scale Factors × $\sigma \frac{\mathscr{L}}{\Sigma w}$ where Σw is the sum of weights of events for that signal or background sample
2. The 1st percentile event weight is scaled to 0, the 99th percentile event weight is scaled to 1
   - Weights below the 1st percentile are set to 0, weights above the 99th percentile are set to 1
3. Multiply signal and background weights by f_s and f_b respectively to equalize the sum of weights of signal and background

$$f_s = \frac{\Sigma w_s + \Sigma w_b}{2\Sigma w_s}, f_b = \frac{\Sigma w_s + \Sigma w_b}{2\Sigma w_b}$$