VBF for Higgs to BSM at the future Muon collider

Yongik Jang,
Kyungpook National University

19, August 2025
Summer Institute 2025

In collaboration with
- Kyu Jung Bae, Kyungpook National University
- Kyoungchul Kong, University of Kansas
- Myeonghun Park, Seoultech

Motivation / Introduction

- Testing the direct coupling of Higgs to heavy SM neutral particle:

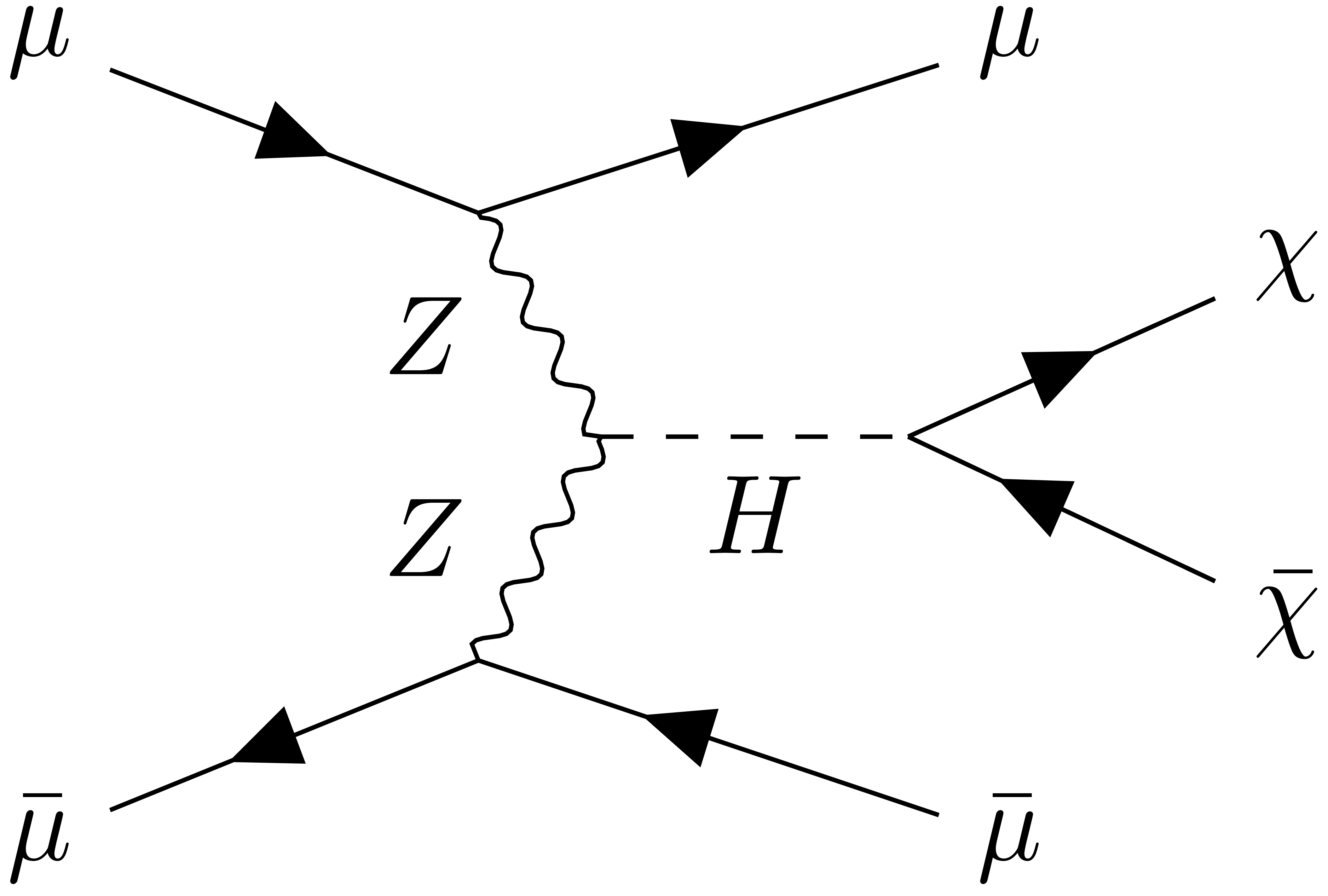
$$\mathscr{L}_{\text{int.}} \supset \frac{1}{\Lambda}|H|^2\bar{\chi}\chi,\ \frac{\lambda}{2}|H|^2\phi^2,\ \ldots$$

- The muon collider is an ideal place to test such couplings using the forward muon detector, owing to the characteristics of vector boson fusion.

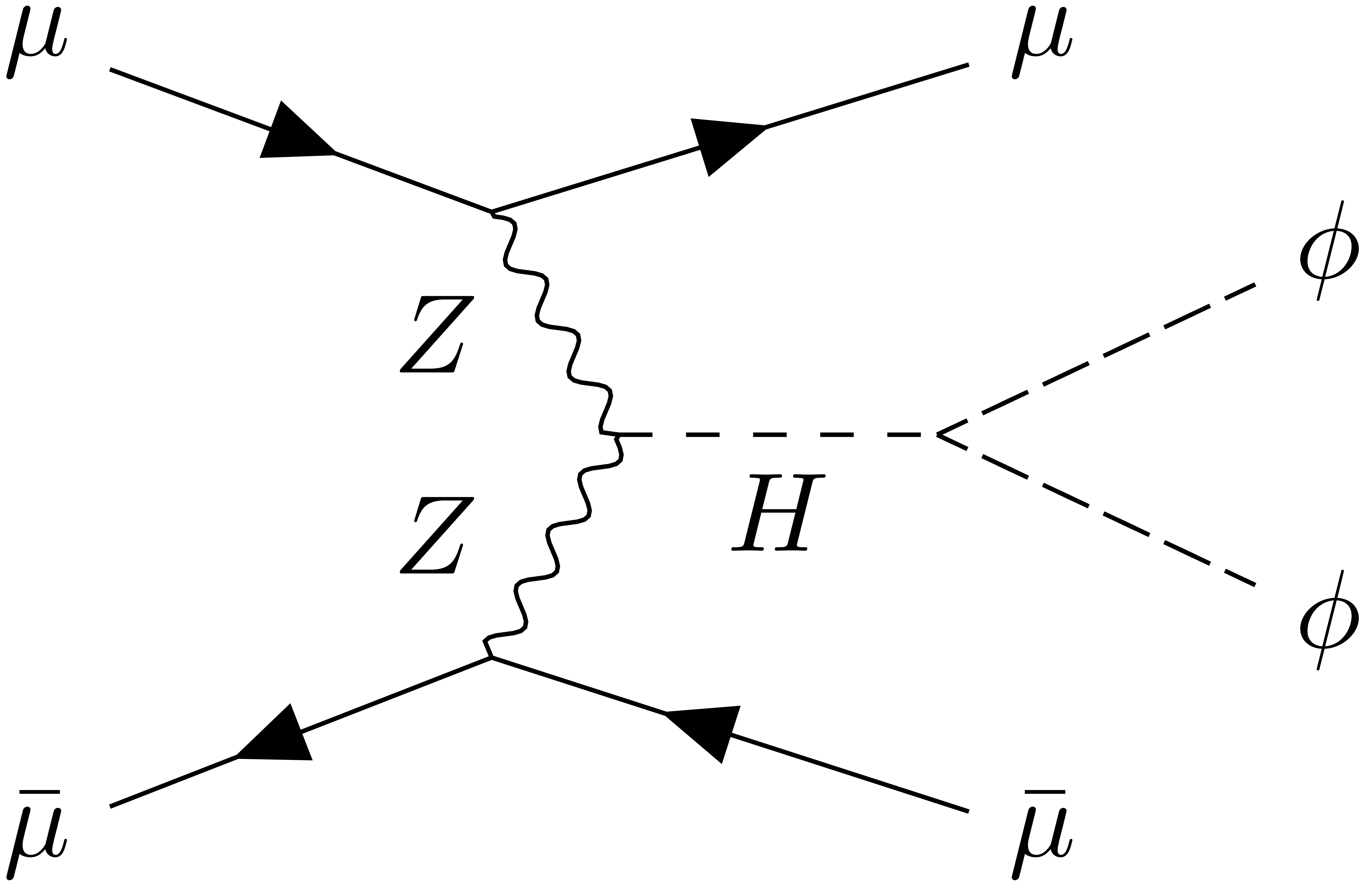
- Neural networks allow effective evaluation of the sensitivity.

- A neural network-based hypothesis test can be used to verify if newly discovered physics is truly a result of Higgs production.

Signal processes

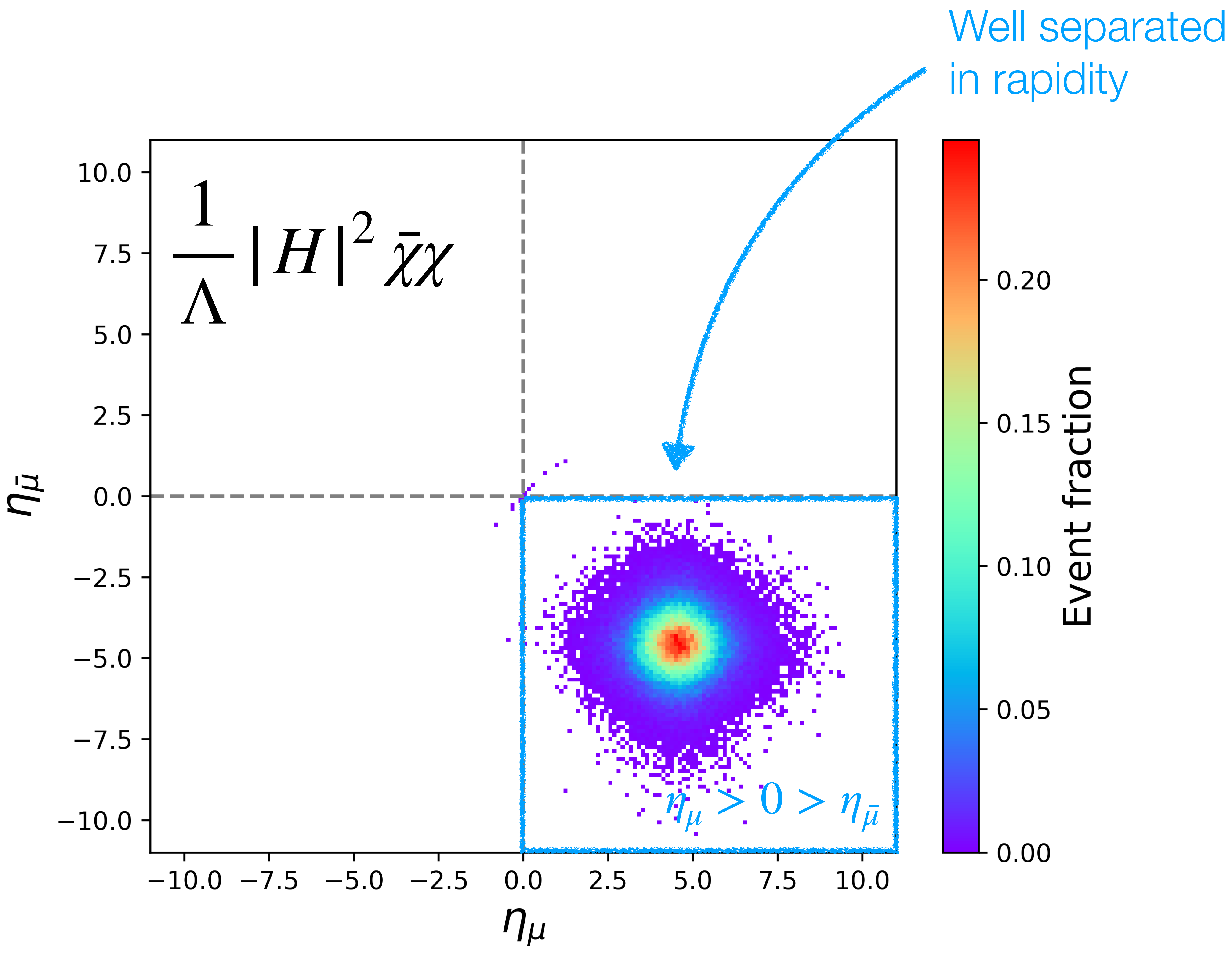


$$\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$$



$$\frac{\lambda}{2}|H|^2\phi^2$$

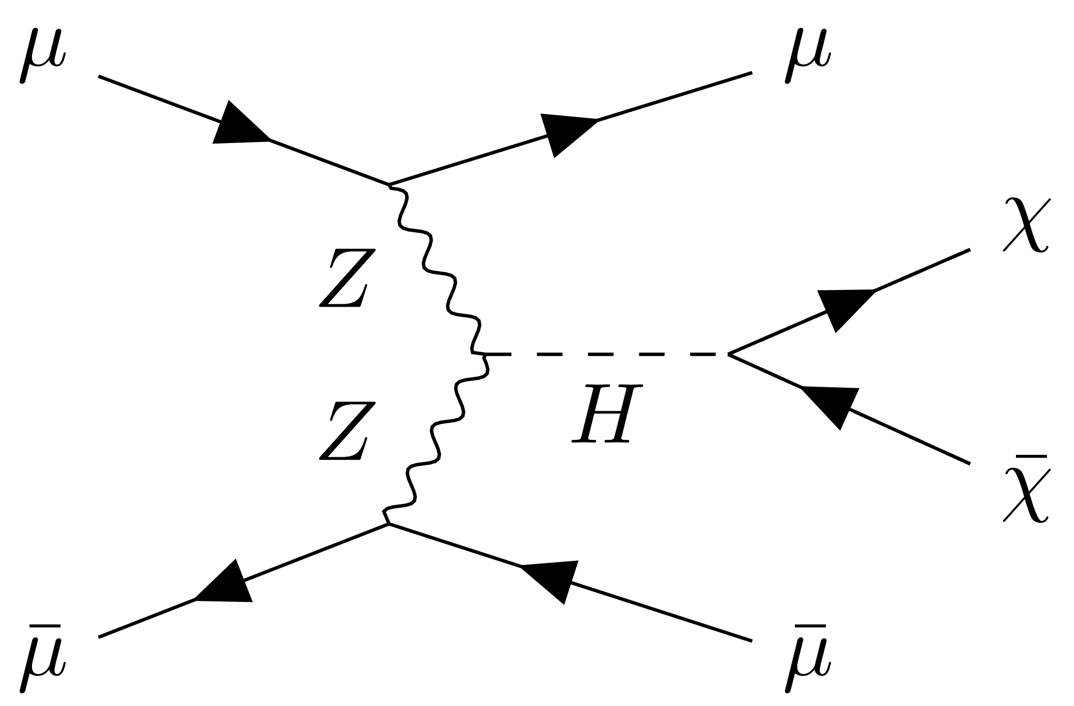
Signal processes: Vector boson fusion



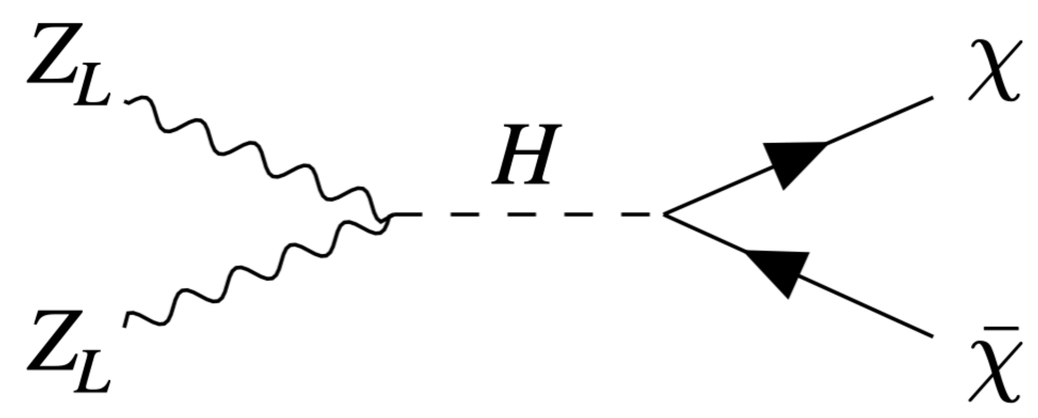
- Approximately, 98% of muon pairs are located in $|\eta| > 2.4$.
- The forward muon detector can effectively capture the signal.

Normalized parton level distribution of $M_\chi = 200$ GeV for full phase space.

Signal processes: Effective vector boson approximation



$\approx f_{Z_L} f_{Z_L}$



$\equiv \hat{\sigma}$

Leading order μ PDF

$$\hat{\sigma}(Z_L Z_L \to \chi\bar{\chi}) = \frac{1}{8\pi}\frac{1}{\Lambda^2}\frac{1}{(1-m_H^2/\hat{s})^2}\left(1-\frac{4M_\chi^2}{\hat{s}}\right)\left(1-\frac{2m_Z^2}{\hat{s}}\right)^2\sqrt{\frac{1-4M_\chi^2/\hat{s}}{1-4m_Z^2/\hat{s}}}$$

$$\hat{\sigma}(Z_L Z_L \to \phi\phi) = \frac{\lambda^2}{8\pi}\frac{1}{\hat{s}}\frac{1}{(1-m_H^2/\hat{s})^2}\left(1-\frac{2m_Z^2}{\hat{s}}\right)^2\sqrt{\frac{1-4M_\phi^2/\hat{s}}{1-4m_Z^2/\hat{s}}}$$

$s \gg M_{\chi(\phi)} > m_H, m_Z$

$$\sigma_\chi(VBF) \propto \frac{1}{\Lambda^2}\ln^2\left(\frac{s}{4M_\chi^2}\right) \qquad \sigma_\phi(VBF) \propto \left(\frac{\lambda}{M_\phi}\right)^2\ln^2\left(\frac{s}{4M_\phi^2}\right)$$

SM Background

1. $\mu\bar{\mu} \to \mu\bar{\mu}\nu\bar{\nu}$
2. $\mu\bar{\mu} \to \mu\bar{\mu}\gamma$
3. $\mu\bar{\mu} \to \mu\bar{\mu}f\bar{f}, \quad f \in \{l, q\}$
4. $\mu\bar{\mu} \to \mu\bar{\mu}W^-W^+, \quad W \to l\nu$ or $q\bar{q}$
5. $\mu\bar{\mu} \to W^-W^+\nu\bar{\nu}, \quad W \to \mu\nu$
6. $\mu\bar{\mu} \to \tau\bar{\tau}, \quad \tau \to \mu\nu\nu$

- Muon colliders use tungsten nozzle shields around the beam pipe to suppress beam-induced background.
- Particles outside the main detector → considered as background

Brief overview of the simulation settings

- Detector setting:

  $\sqrt{s} = 10$ TeV, $|\eta_{\rm main}| < 2.44(\theta_{\rm min} \approx 10°)$, $|\eta_{\rm max}| = 6.0$,

  $\mathscr{L} = 10$ ab^{-1}, $\delta E_{\rm res} = 10\,\%$

- To implement energy resolution, Gaussian smearing is applied on forward muons, $\dfrac{\Delta E}{E} = 10\,\%$.

Signal selection

- Well separated energetic muon pairs in high rapidity regions:

$\eta_\mu > 0 > \eta_{\bar{\mu}},\ \Delta R_{\mu\bar{\mu}} > 0.4,\ 6 > |\eta_{\mu(\bar{\mu})}|,\ E_{\rm min.} > 500$ GeV

- Large contribution from elastic scattering ($\mu\bar{\mu} \to \mu\bar{\mu}$):

$p_T^{\mu\bar{\mu}} > 50$ GeV

Normalized kinematic distribution of the signal and background.

S to BG discrimination

- Cut-flow analysis allows a 5σ discovery of $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$ up to $\Lambda = 360$ GeV for $M_\chi = 200$ GeV.

- The **neural network** gives better results, as we will see…

Input features:

$$\log\left(\frac{p_T^{\mu(\bar{\mu})}}{20\ \text{GeV}}\right),\ \log\left(\frac{p_T^{\mu\bar{\mu}}}{50\ \text{GeV}}\right),\ \frac{\eta_{\mu(\bar{\mu})}}{6},\ \frac{\Delta\eta_{\mu\bar{\mu}}}{12},\ \frac{|\Delta\phi_{\mu\bar{\mu}}|}{\pi},$$

$$\frac{E_{\min}}{\sqrt{s}/2},\ \frac{M_{\mu\bar{\mu}}}{\sqrt{s}},\ \frac{M^2_{\chi\bar{\chi}}}{s}$$

S to BG discrimination

- Construct separate networks for each mass point $(100,\ 200,\ \ldots,\ 1000)$ to determine the optimal suppression scale Λ.

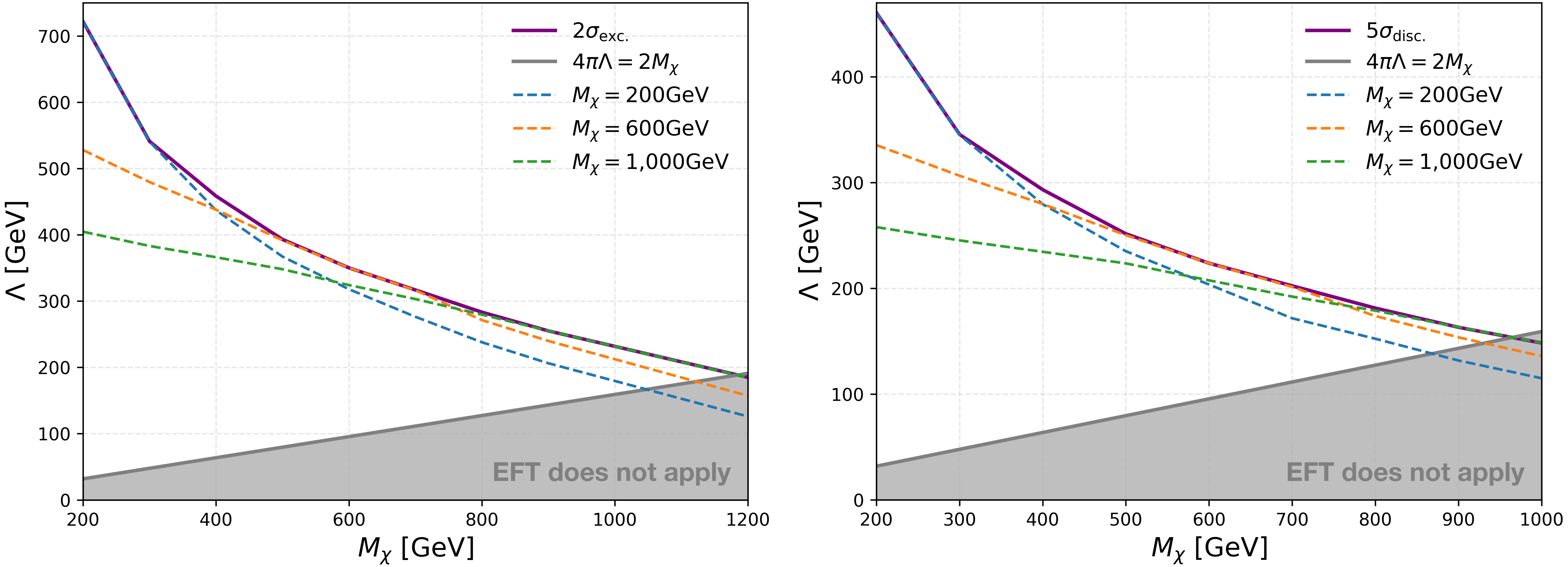
- Calculate 2σ exclusion and 5σ discovery limit

- Use networks trained on different mass points to confirm the optimal results (e.g., calculate Λ for $M_\chi = 300$ GeV using NN_{200}).

- NN_M denotes the optimally trained neural network for a given mass M.

S to BG discrimination: $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$

$$\sigma_\chi(VBF) \propto \frac{1}{\Lambda^2}\ln^2\left(\frac{s}{4M_\chi^2}\right)$$



Left figure show 2σ exclusion limit and the right figure show 5σ discovery limit. The gray shaded region indicates where effective field theory is not valid. Each dotted lines are calculated using NN_{200}(blue), NN_{600}(orange), NN_{1000}(green).

S to BG discrimination: $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$

The NN_{200} selection for $M_\chi = 200$ GeV

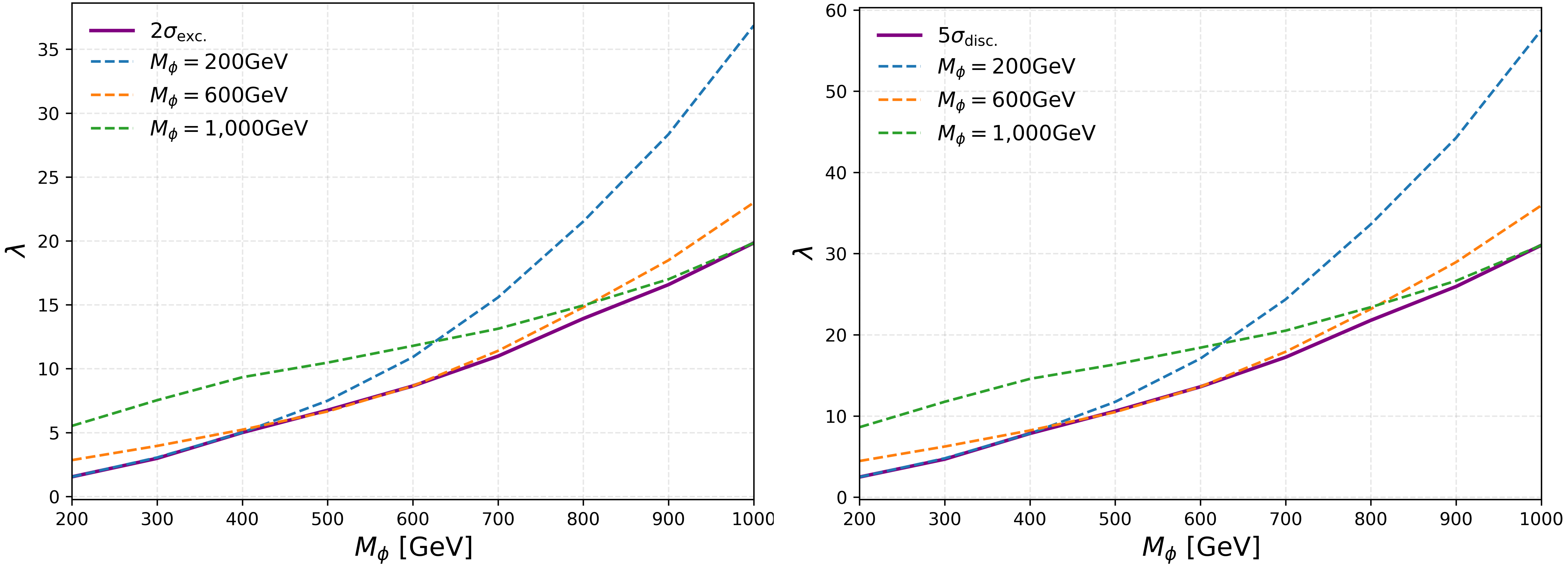
	$\mu\bar{\mu}\chi\bar{\chi}$	$\mu\bar{\mu}\nu\bar{\nu}$	$\mu\bar{\mu}\gamma$	$\mu\bar{\mu}f\bar{f}$	$\mu\bar{\mu}W^-W^+$	$W^-W^+\nu\bar{\nu}$	$\tau\bar{\tau}$
Baseline	$5.9\times10^2\cdot(\mathrm{TeV}/\Lambda)^2$	1.3×10^6	2.4×10^7	1.4×10^6	3.0×10^5	2.5×10^3	75
NN selection	$1.6\times10^2\cdot(\mathrm{TeV}/\Lambda)^2$	9.3×10^3	3.1×10^3	2.1×10^3	6.9×10^3	11	0

$\sqrt{s} = 10$ TeV, $\mathscr{L} = 10\mathrm{ab}^{-1}$, $|\eta_{\mathrm{main}}| < 2.44$, $\delta E_{res.} = 10\,\%$

- 5σ discovery at $\Lambda = 460$ GeV

S to BG discrimination: $\frac{\lambda}{2}|H|^2\phi^2$

$$\sigma_\phi(VBF) \propto \left(\frac{\lambda}{M_\phi}\right)^2 \ln^2\left(\frac{s}{4M_\phi^2}\right)$$



Left figure show 2σ exclusion limit and the right figure show 5σ discovery limit. Each dotted lines are calculated using NN_{200}(blue), NN_{600}(orange), NN_{1000}(green).

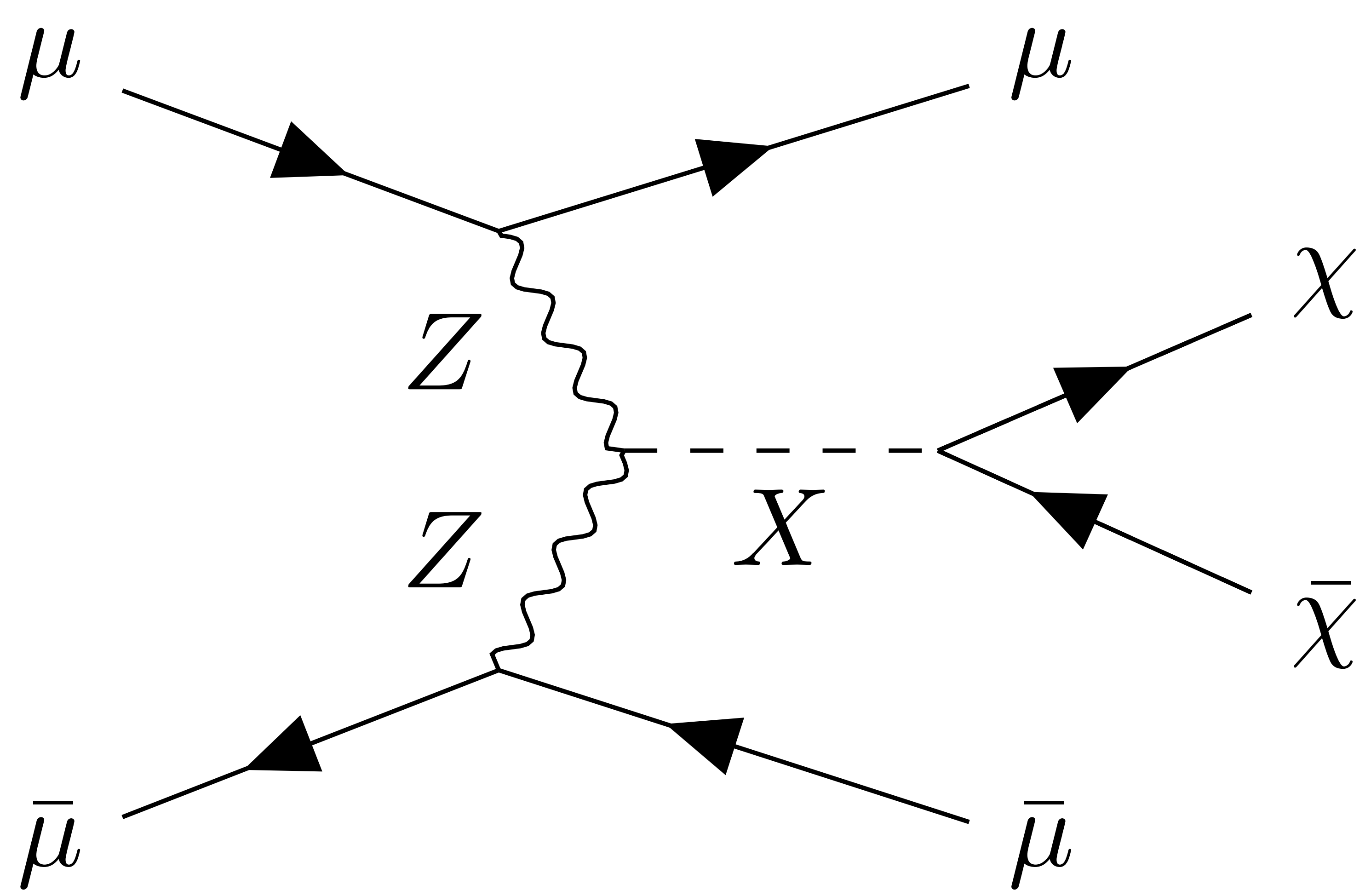
S to BG discrimination: $\frac{\lambda}{2}|H|^2\phi^2$

The NN_{200} selection for $M_\phi = 200$ GeV

	$\mu\bar{\mu}\phi\phi$	$\mu\bar{\mu}\nu\bar{\nu}$	$\mu\bar{\mu}\gamma$	$\mu\bar{\mu}f\bar{f}$	$\mu\bar{\mu}W^-W^+$	$W^-W^+\nu\bar{\nu}$	$\tau\bar{\tau}$
Baseline	$3.1\times10^2\cdot\lambda^2$	1.3×10^6	2.4×10^7	1.4×10^6	3.0×10^5	2.5×10^3	75
NN selection	$1.1\times10^2\cdot\lambda^2$	1.1×10^4	2.6×10^3	1.3×10^3	3.6×10^3	4	0

- 5σ discovery at $\lambda = 2.5$

Really Higgs?



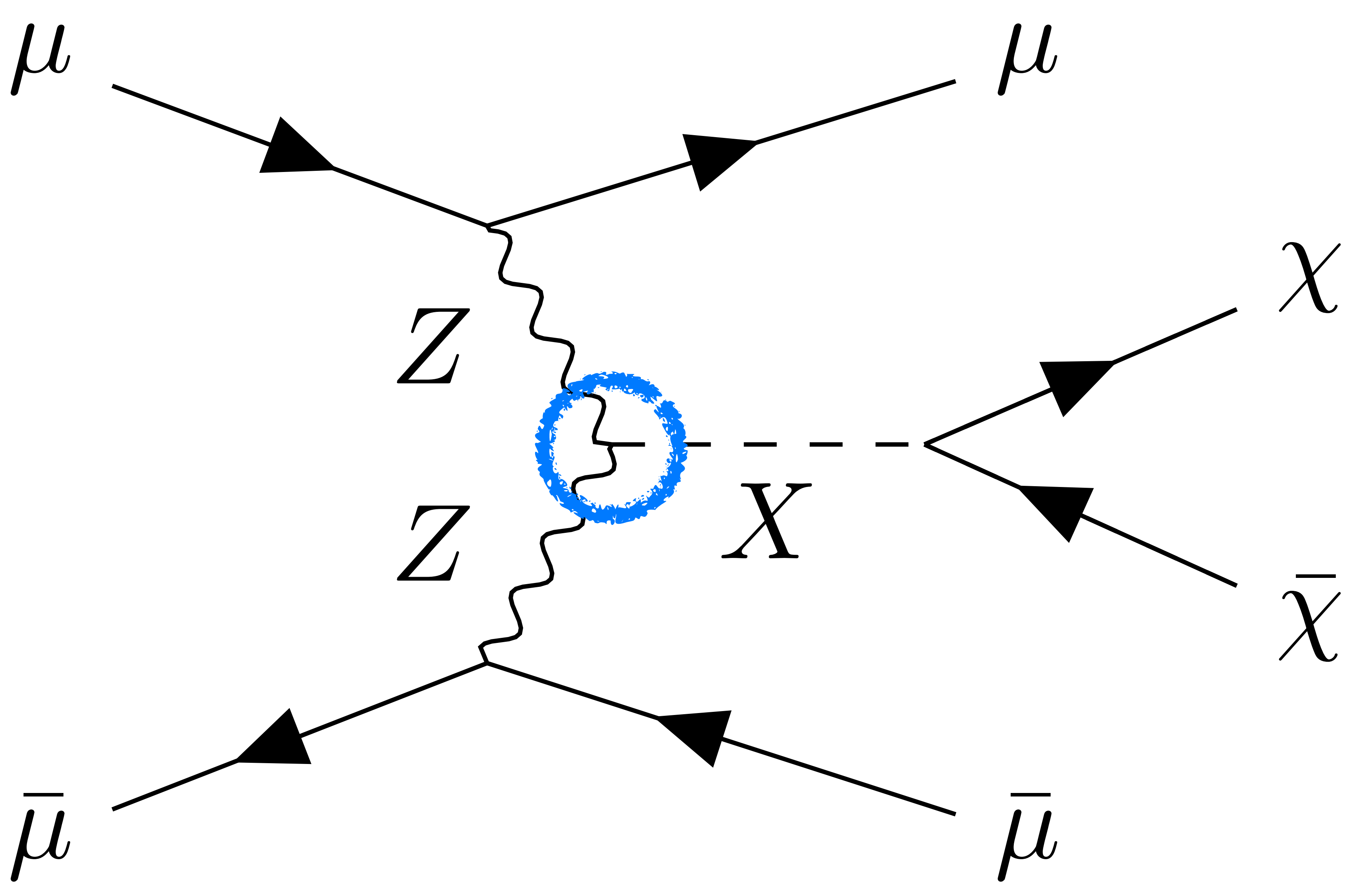
$$X \overset{?}{=} H$$

- However, what if the process is not mediated through the Higgs boson?

$$\mathscr{L}_S \supset \frac{1}{\Lambda_S} S Z^{\mu\nu} Z_{\mu\nu} + g_{S\chi\chi} S \bar{\chi}\chi \quad \text{or} \quad \mathscr{L}_A \supset \frac{1}{\Lambda_A} A \widetilde{Z}^{\mu\nu} Z_{\mu\nu} + g_{A\chi\chi} A \bar{\chi}(i\gamma^5)\chi$$

- How can we discriminate?

Really Higgs?: Angular Correlation

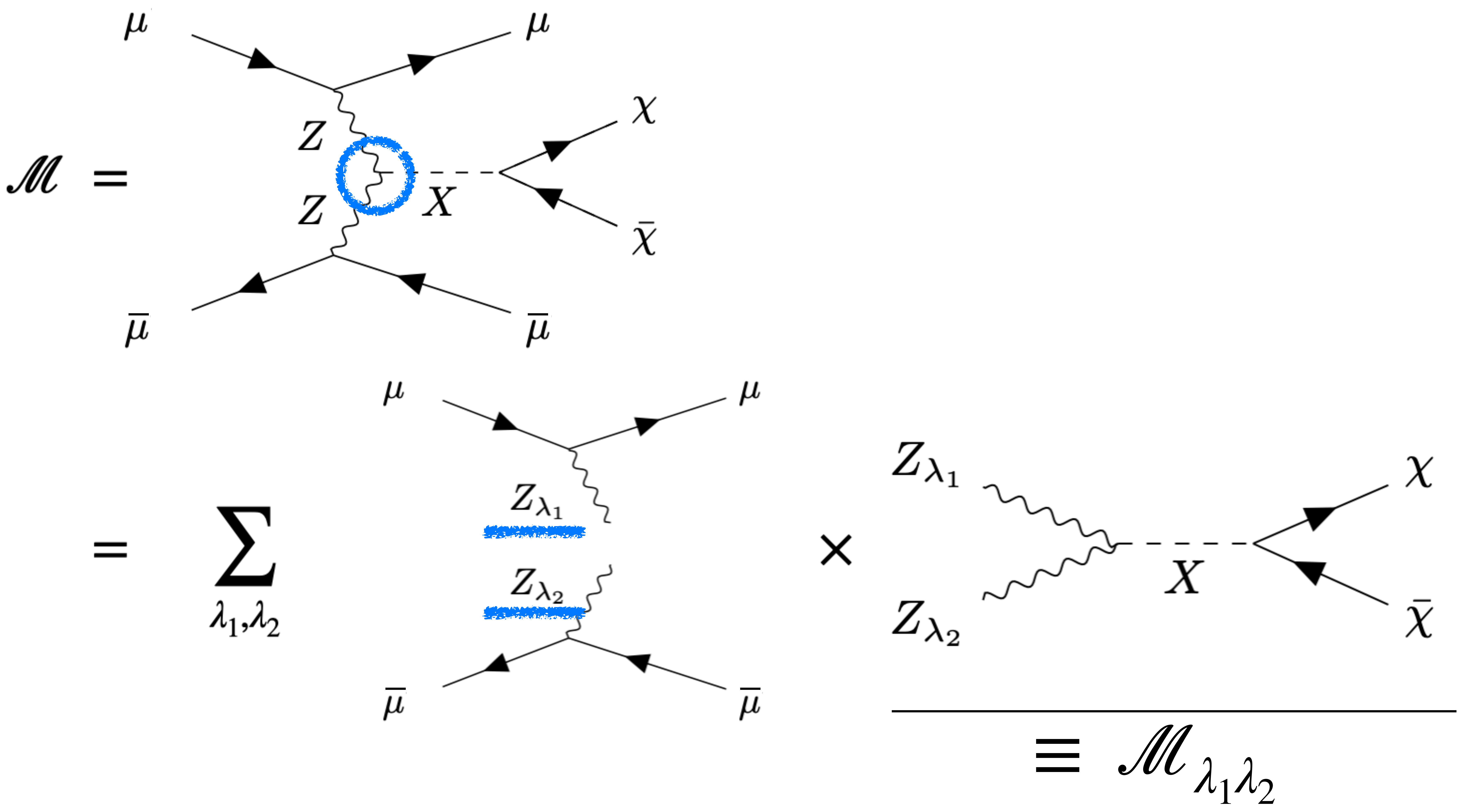


Z_λ \ X	H	S	A
Transverse		**Dominant**	**Dominant**
Longitudinal	**Dominant**		X

- Is there kinematic variable represents this? → $\Delta\phi_{\mu\bar{\mu}} = \phi_\mu - \phi_{\bar{\mu}}$

$$\Delta\phi \equiv \phi_+ - \phi_- - 2\pi\,\Theta\big(|\phi_+ - \phi_-| - \pi\big)\ \mathrm{sgn}\big(\phi_+ - \phi_-\big)$$

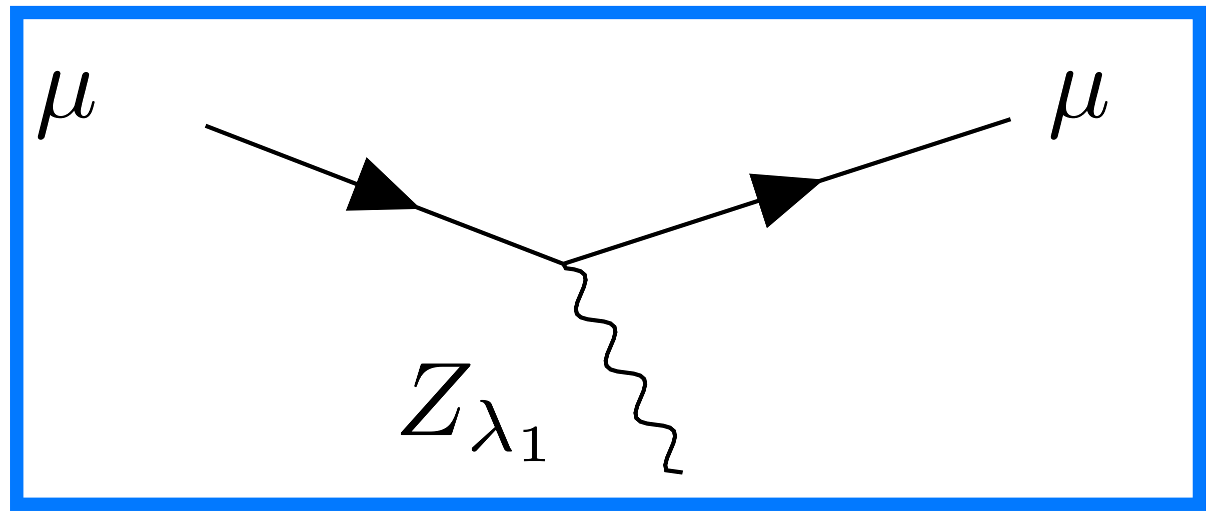
Angular Correlation: The helicity formalism



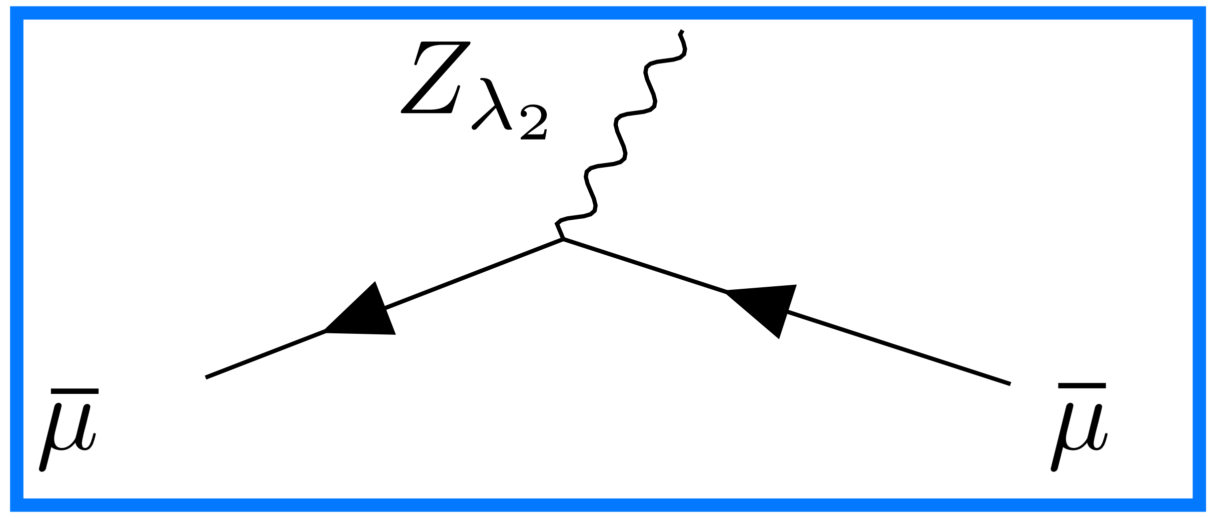
Kaoru Hagiwara et al JHEP07(2009)101

Angular Correlation: The helicity formalism

- In the scenario, by the helicity selection rule only $\lambda_1 = \lambda_2$ contribute. $\mathscr{M}_{\lambda_1\lambda_2} \to \mathscr{M}_{\lambda}$
- Under the EVA limit,



$$\propto \exp(-i\lambda\phi_{\mu})$$



$$\propto \exp(i\lambda\phi_{\bar{\mu}})$$

$$\sum_{(\lambda_1=\lambda_2)} \mathscr{M}_{\lambda_1}\mathscr{M}^*_{\lambda_2} \to \text{const.}$$

$$\sum_{(\lambda_1,\lambda_2)=(\pm,0),(0,\pm)} \mathscr{M}_{\lambda_1}\mathscr{M}^*_{\lambda_2} \to \cos(\Delta\phi)$$

$$\sum_{(\lambda_1,\lambda_2)=(\pm,\mp)} \mathscr{M}_{\lambda_1}\mathscr{M}^*_{\lambda_2} \to \cos(2\Delta\phi)$$

$$\Rightarrow \; d\sigma(\text{VBF}) = C_0 + C_1\cos(\Delta\phi) + C_2\cos(2\Delta\phi)$$

- The sign and relative magnitude of the coefficient function are determined by $\hat{\sigma}(\lambda_1\lambda_2)$.
- This does not depend on the final state particles.

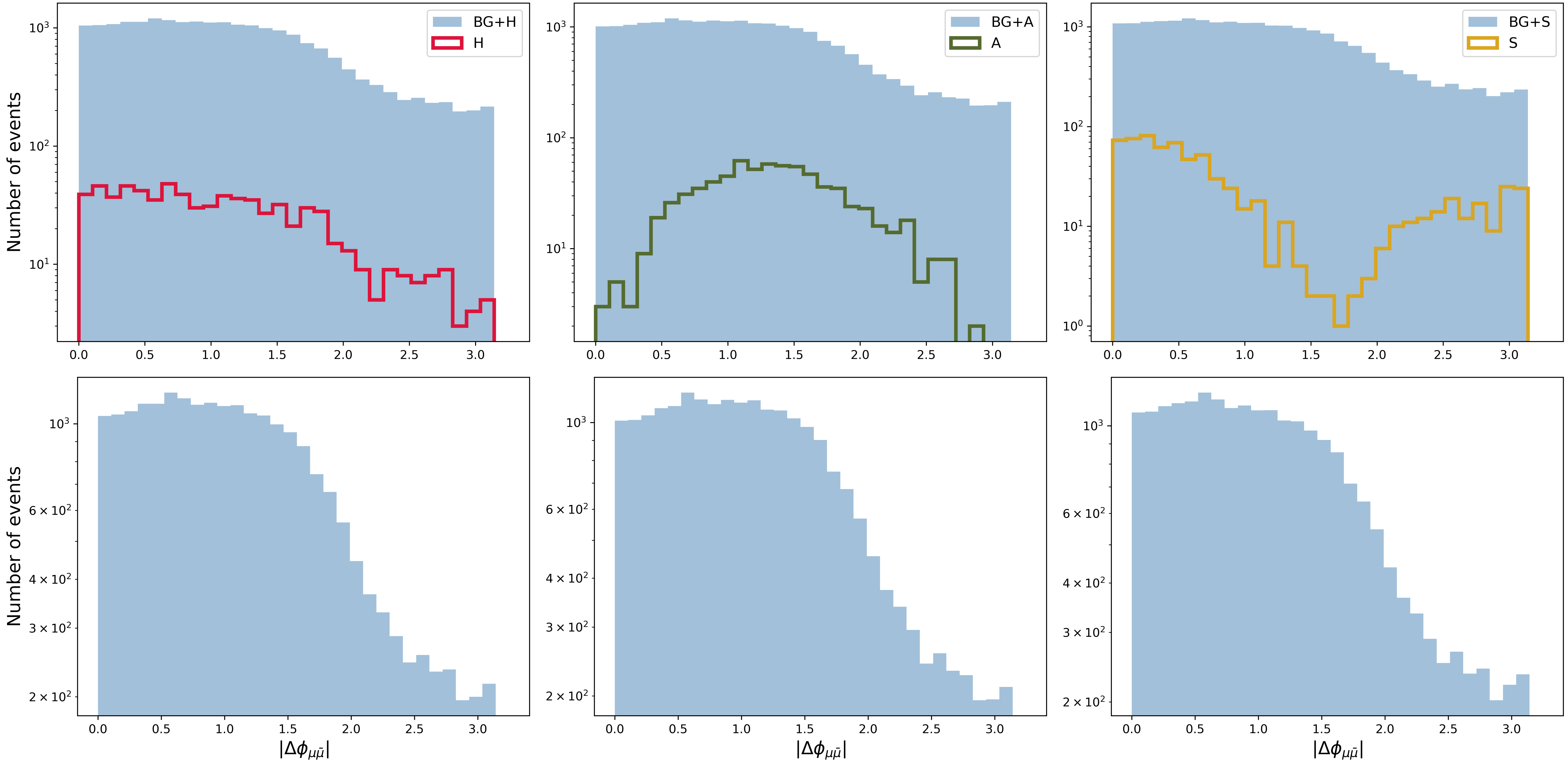
Angular Correlation: The helicity formalism

$$\text{H} \sim C_0$$

$$\text{S} \sim C_0 + C_2 \cos(2\Delta\phi_{\mu\bar{\mu}})$$

$$\text{A} \sim C_0 - C_2 \cos(2\Delta\phi_{\mu\bar{\mu}})$$

Really Higgs?: Hypothesis test



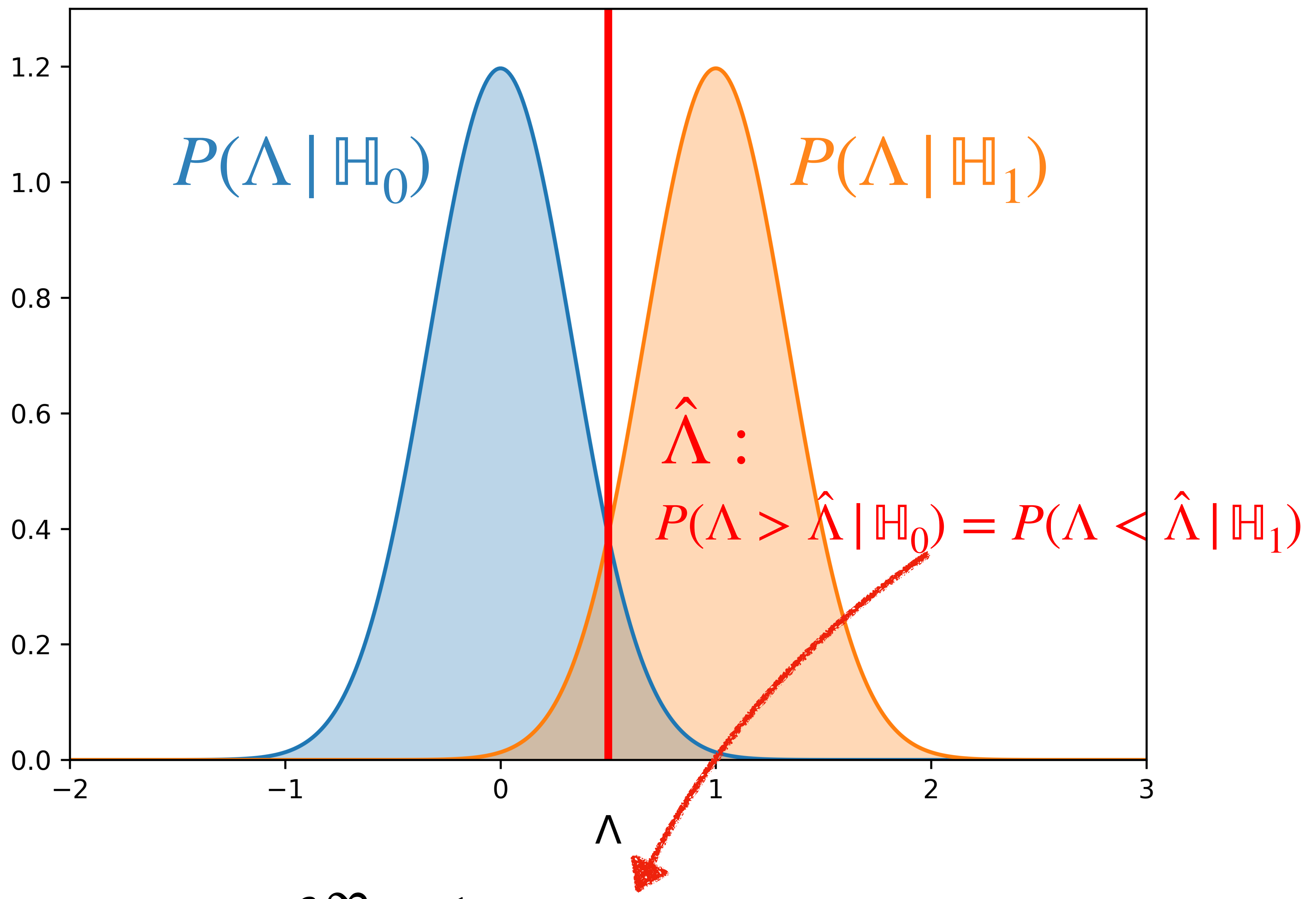
BG covers the signal → Hard to distinguish with only the $\Delta\phi_{\mu\bar{\mu}}$ distribution

→ Need to perform hypothesis test → **Neural network** makes it easier

Really Higgs?: Hypothesis test

Test statistic: Log-likelihood ratio

$$\Lambda = \ln \frac{\mathscr{L}(H_1)}{\mathscr{L}(H_0)} = \ln \frac{\Pi_i P_{H_1}(\overrightarrow{X_i})}{\Pi_i P_{H_0}(\overrightarrow{X_i})} = \sum_{i=1}^{N} \ln \frac{P_{H_1}(\overrightarrow{X}_i)}{P_{H_0}(\overrightarrow{X}_i)}$$



$$= \int_{\tilde{Z}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$$

$$Z = 2\tilde{Z} \ [\sigma]$$

Really Higgs?: Hypothesis test

- No quantum interference between the signal and background

  →Construct neural network that classifies H and S(A)

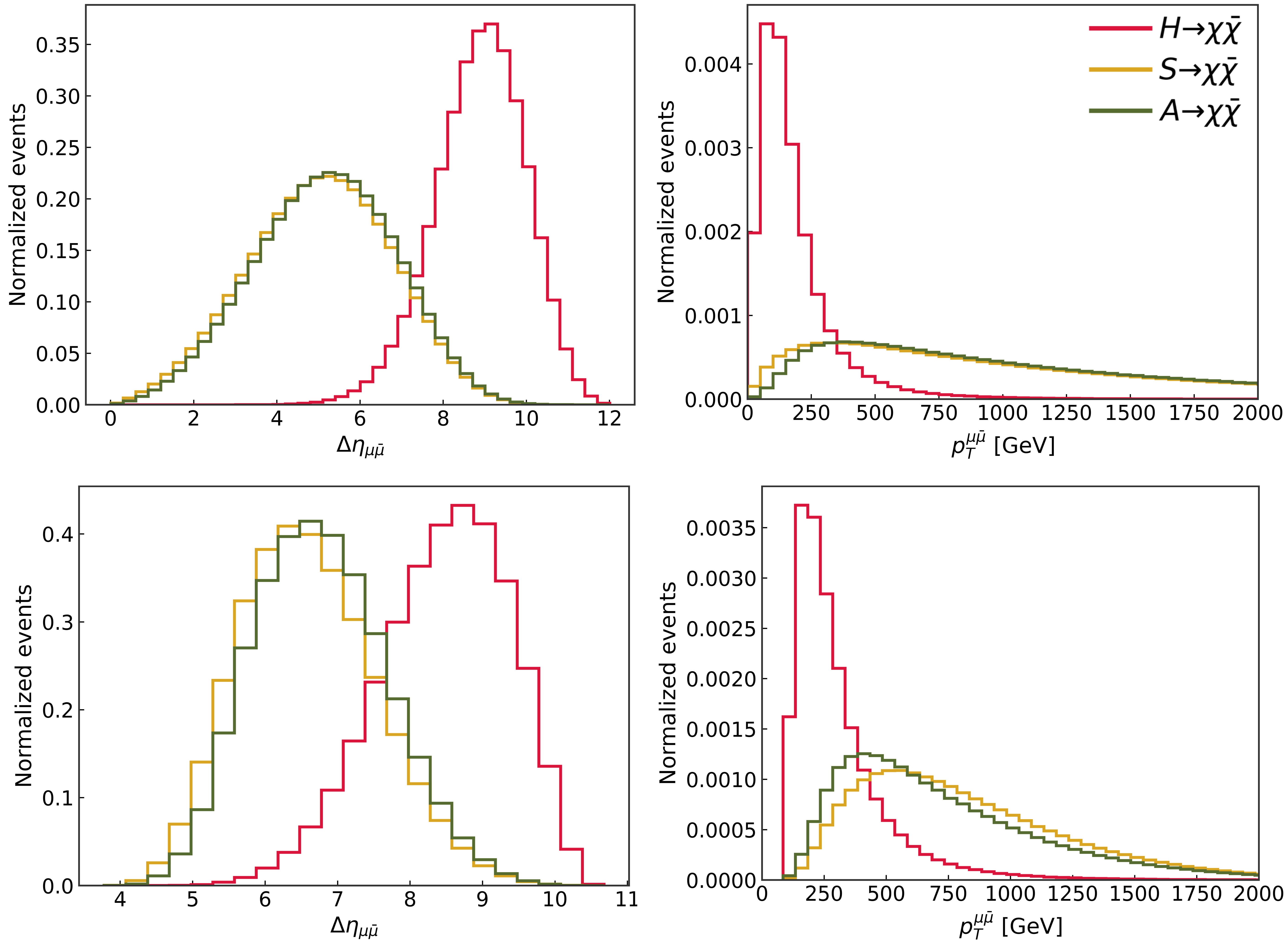
- Approximate score ratio to PDF ratio:

$$\frac{P_{H_1}(\vec{X}_i)}{P_{H_0}(\vec{X}_i)} \xrightarrow{\text{Likelihood ratio trick}} \frac{s(\vec{X}_i)}{1 - s(\vec{X}_i)}$$

- Five input features, encoding ZZX coupling:

$$\log\left(\frac{p_T^{\mu(\bar{\mu})}}{20\ \text{GeV}}\right),\ \log\left(\frac{p_T^{\mu\bar{\mu}}}{50\ \text{GeV}}\right),\ \frac{\Delta\eta_{\mu\bar{\mu}}}{12},\ \frac{|\Delta\phi_{\mu\bar{\mu}}|}{\pi}$$

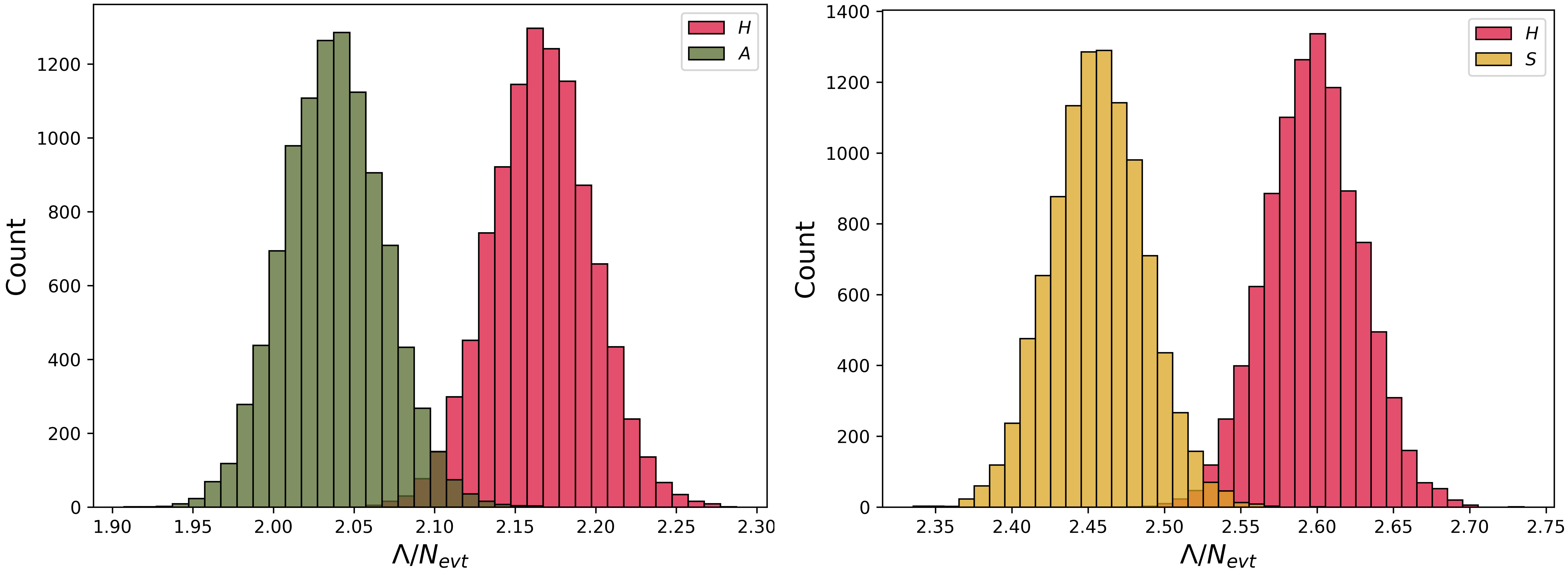
Really Higgs?: Hypothesis test



Normalized kinematic distribution of parton level after selection.

Same as upper line, but for after NN_{200} selection.

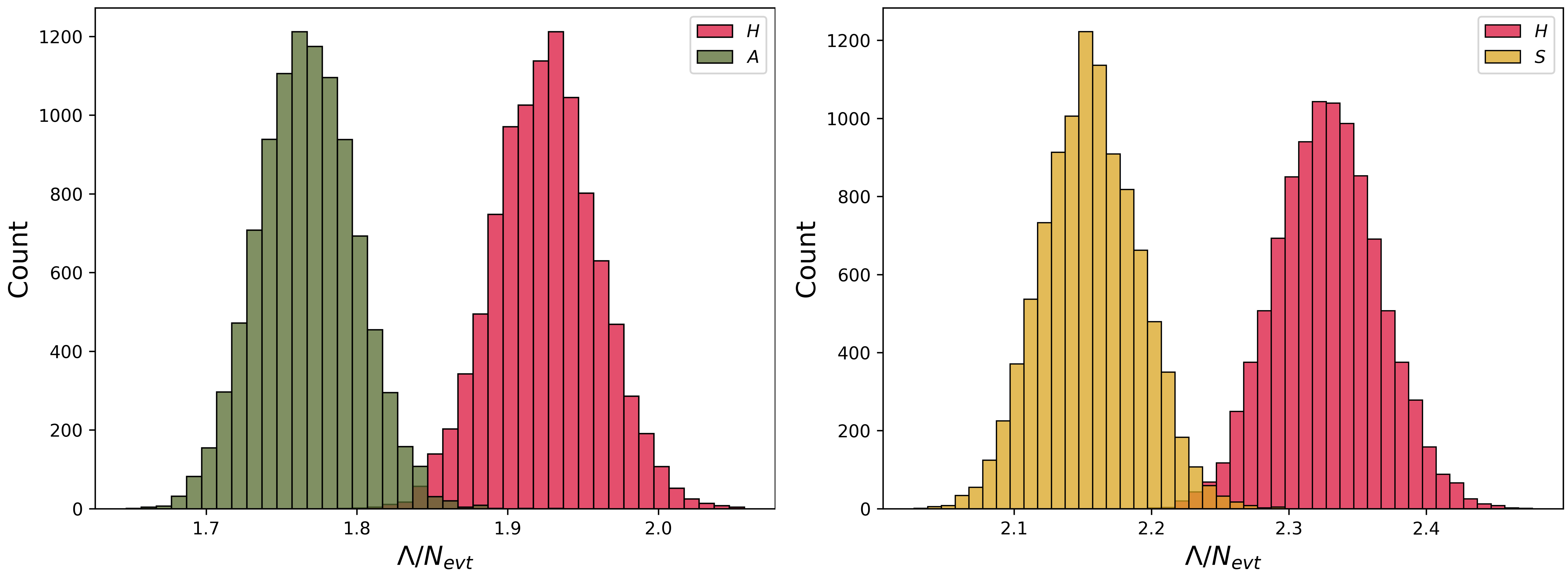
Hypothesis test: $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$



Results of 10,000 pseudo-experiments: left, H vs. A; right, H vs. S.

	H vs A	H vs S
Z	4.1σ	4.5σ

Hypothesis test: $\frac{\lambda}{2}|H|^2\phi^2$

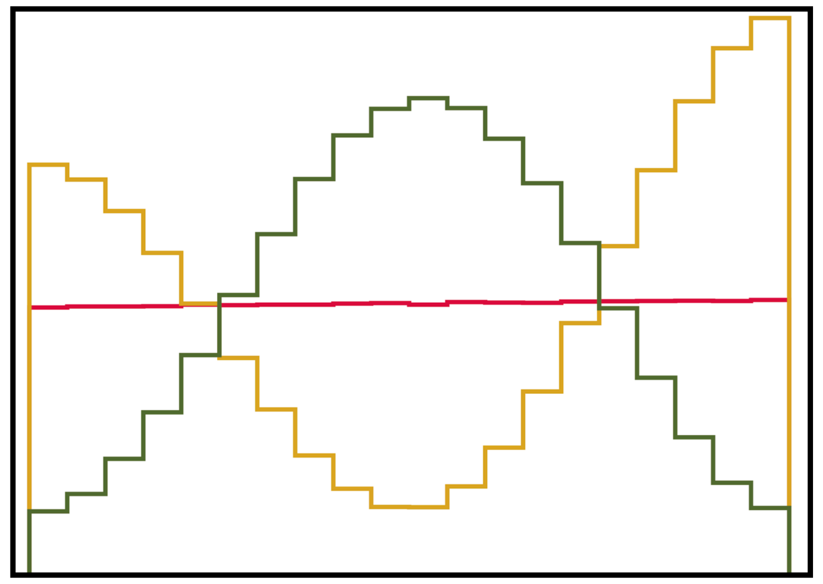


Results of 10,000 pseudo-experiments: left, H vs. A; right, H vs. S.

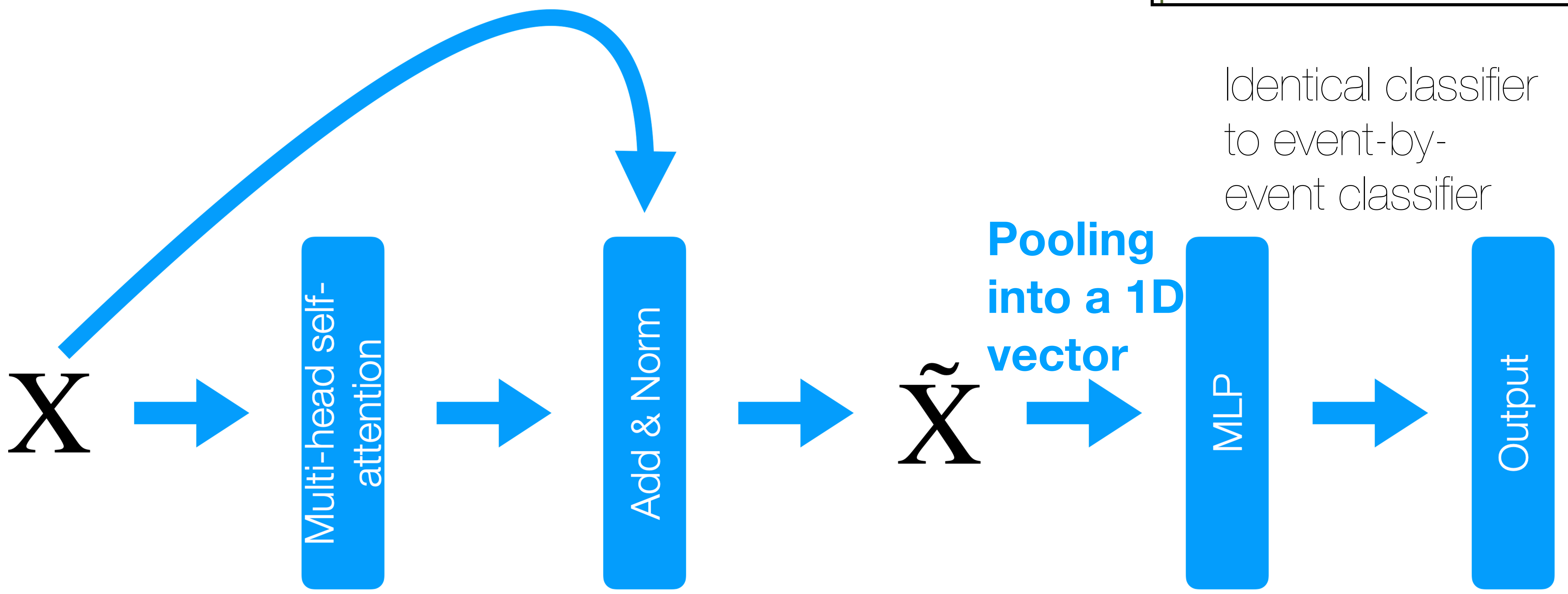
	H vs A	H vs S
Z	4.7σ	4.7σ

Hypothesis test - Transformer

- Not just encoding each event, but the shape



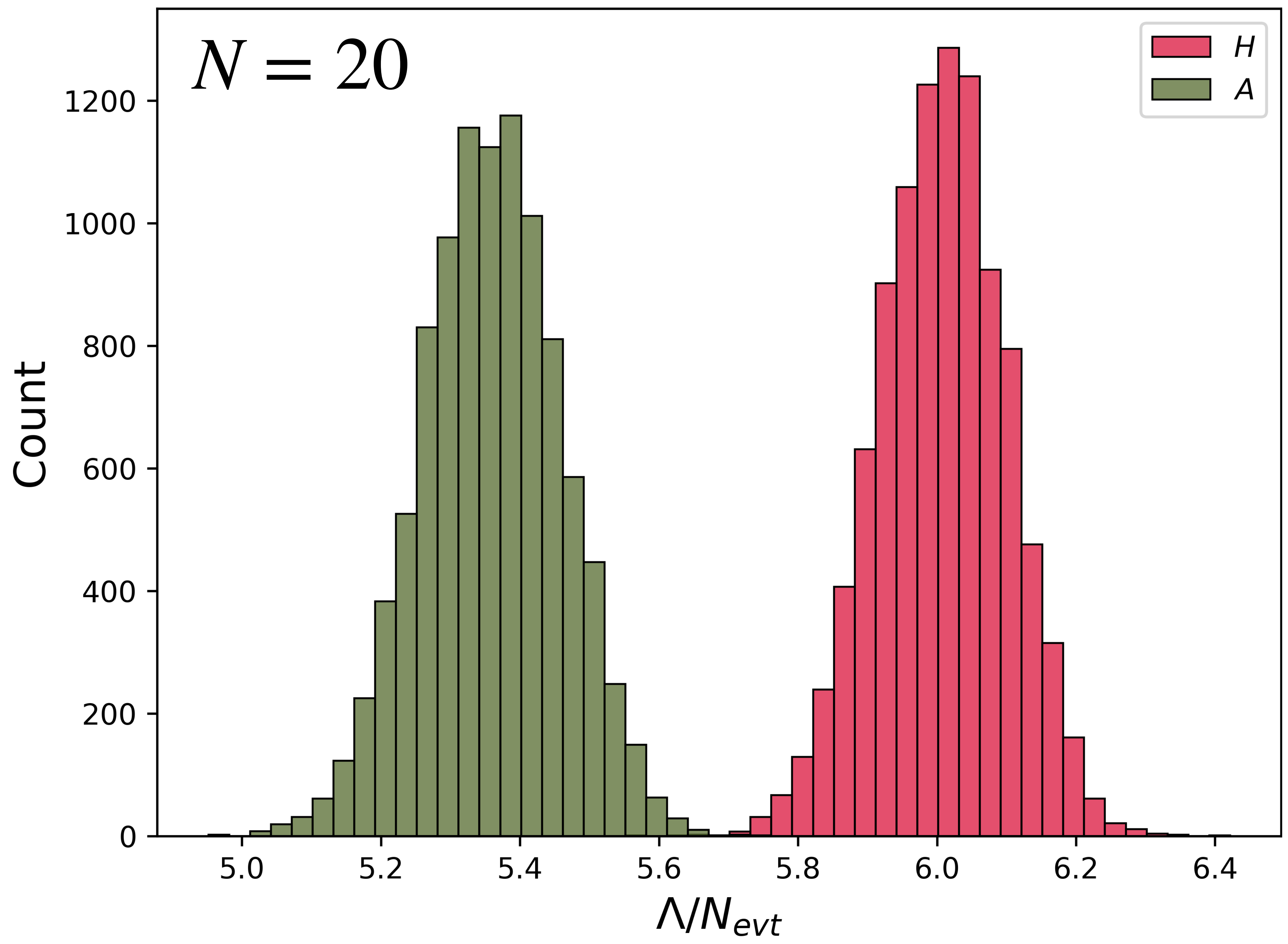
$p_\mu^1,\, p_{\bar\mu}^1$
$p_\mu^2,\, p_{\bar\mu}^2$
$\vdots$
$p_\mu^N,\, p_{\bar\mu}^N$

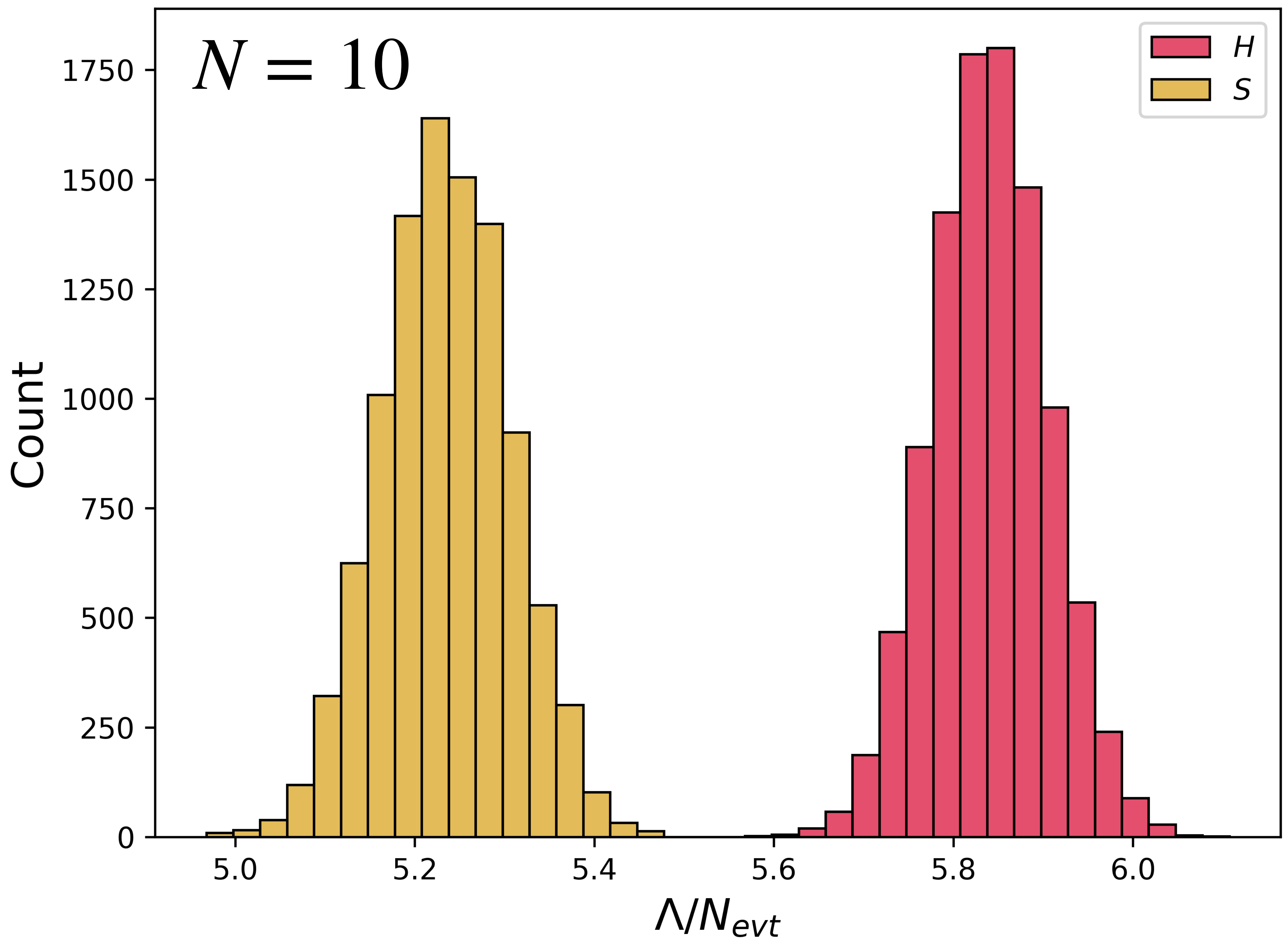


Full architecture of classifier with transformer layer.

- Use $p_\mu,\, p_{\bar\mu}$ as input vectors, so that the machine can learn hidden correlations without prior bias
- The number of input events, N, is important hyperparameter.

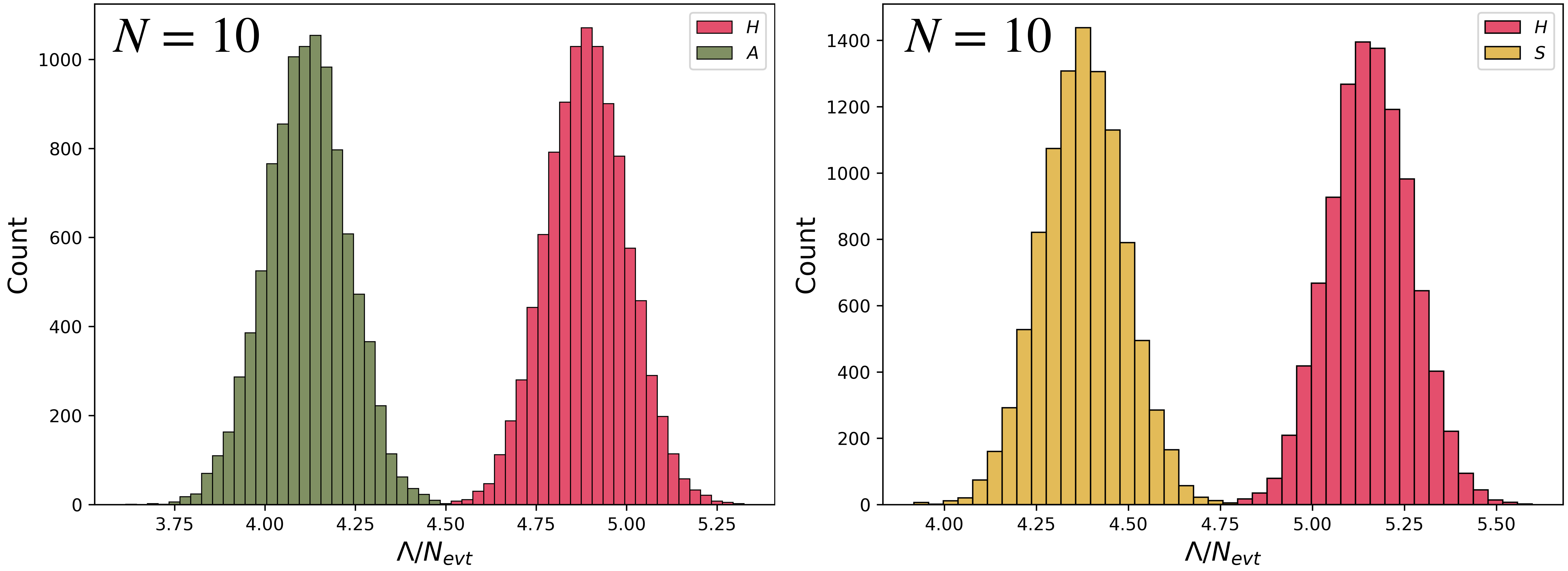
Hypothesis test - Transformer: $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$





	H vs A	H vs S
Z	6.7σ	14σ

Hypothesis test - Transformer: $\frac{\lambda}{2}|H|^2\phi^2$

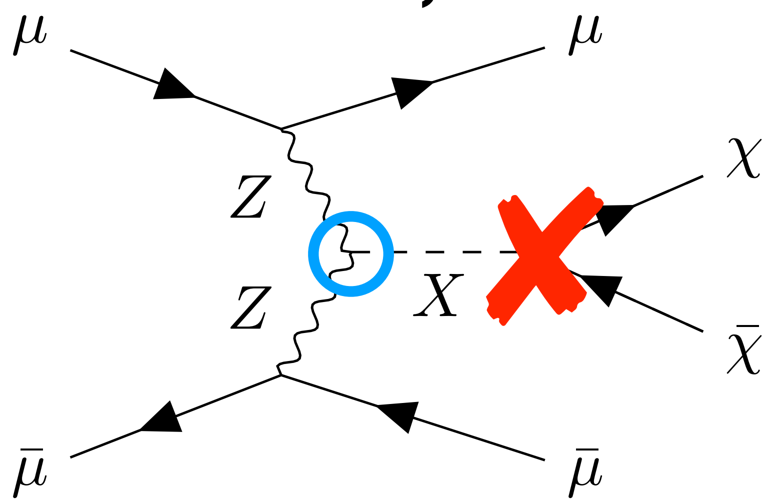


Results of 10,000 pseudo-experiments: left, H vs. A; right, H vs. S. Using classifier with transformer layer.

	H vs A	H vs S
Z	7.0σ	7.0σ

Hypothesis test - Generalization

- Recall that the angular correlation depends on ZZX, not on X-something.



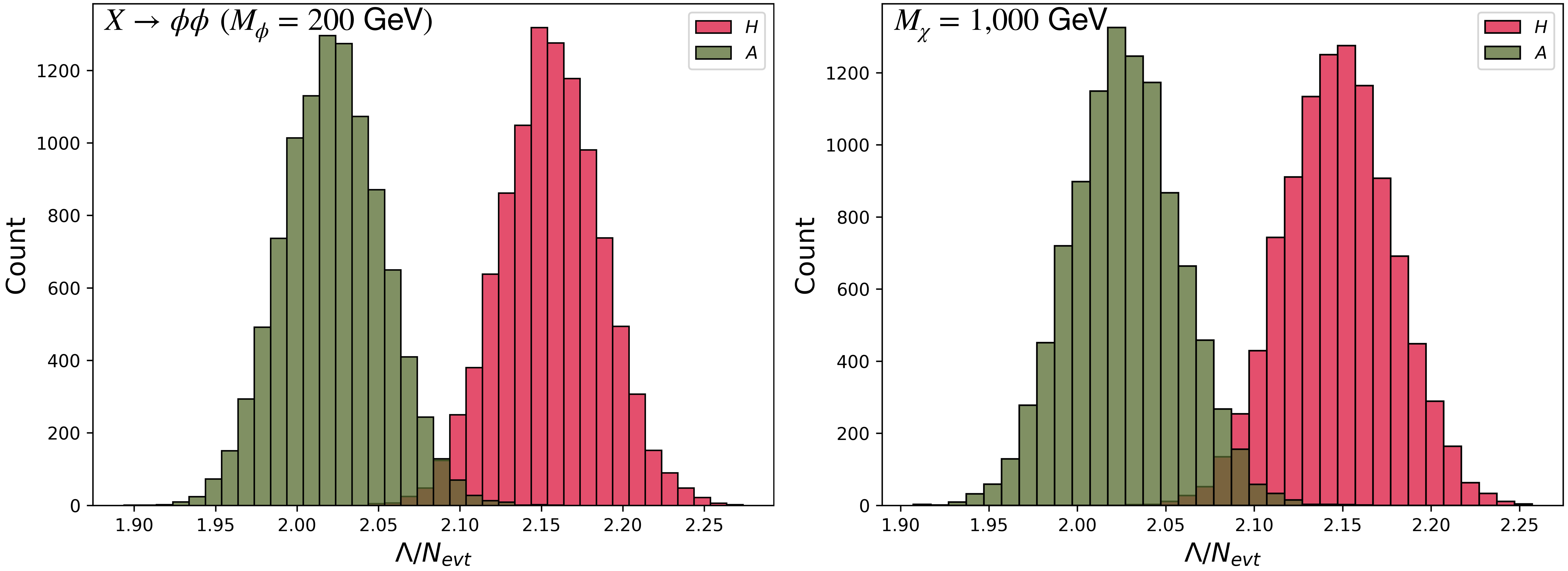
- The same holds $f_{\mu \to Z_\lambda}$.

- Therefore, even if the mediator discriminator is trained with specific mass point ($M_\chi = 200$ GeV in our case), we expect it can classify other cases as well (e.g., different mass points or final-state couplings).

Hypothesis test - Generalization

Event-by-event classifier

H vs. A



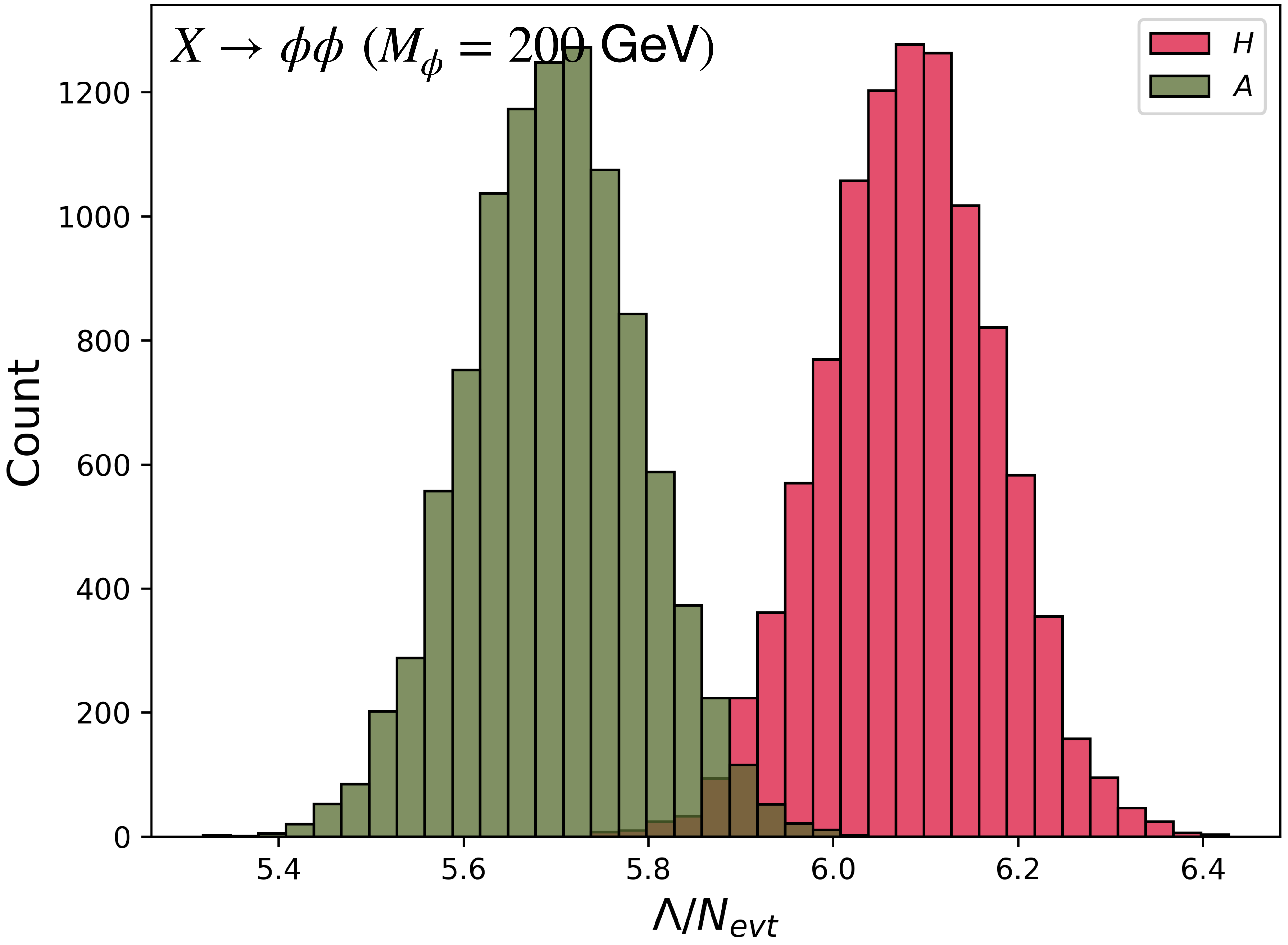
Left: different spin (ϕ), Right: different mass ($M_\chi = 1{,}000$ GeV). Using **event-by-event** classifier

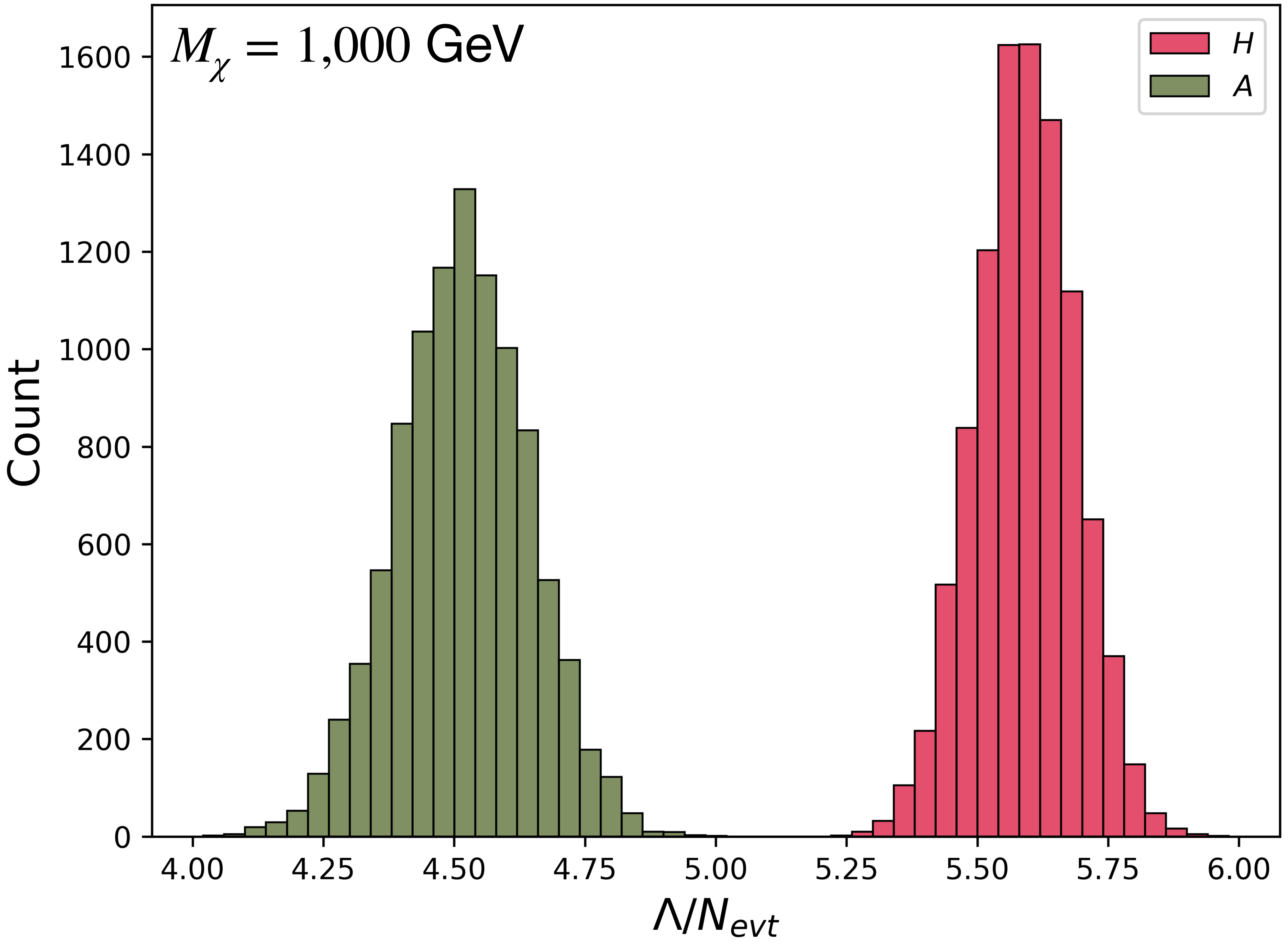
	$M_\phi = 200$ GeV	$M_\chi = 1000$ GeV
Z	4.2σ	3.9σ

Hypothesis test - Generalization

Classifier with transformer

H vs. A





Left: different spin (ϕ), Right: different mass ($M_\chi = 1{,}000$ GeV). Using classifier with **transformer layer.**

	$M_\phi = 200$ GeV	$M_\chi = 1000$ GeV
Z	4.1σ	17σ

Summary and Conclusion

- The Higgs direct coupling to SM neutral particles dominated by VBF process can be effectively tested at a muon collider using the forward detector.
- With high confidence, it can be verified whether it is actually produced via the Higgs, regardless of its mass or coupling

Thank you for your attention

S to BG discrimination - Cut-flow: $\frac{1}{\Lambda}|H|^2\bar{\chi}\chi$

Cut flow table for $M_\chi = 200$ GeV

	$\mu\bar{\mu}\chi\bar{\chi}$	$\mu\bar{\mu}\nu\bar{\nu}$	$\mu\bar{\mu}\gamma$	$\mu\bar{\mu}f\bar{f}$	$\mu\bar{\mu}W^-W^+$	$W^-W^+\nu\bar{\nu}$	$\tau\bar{\tau}$
Baseline	$5.9\times10^2\cdot(\text{TeV}/\Lambda)^2$	1.3×10^6	2.4×10^7	1.4×10^6	3.0×10^5	2.5×10^3	75
$\text{sign}(M^2_{\chi\bar{\chi}})\cdot\sqrt{\lvert M^2_{\chi\bar{\chi}}\rvert} > 180$ GeV	$4.9\times10^2\cdot(\text{TeV}/\Lambda)^2$	6.5×10^5	9.9×10^6	6.6×10^5	2.5×10^5	2.5×10^3	75
$\Delta\eta_{\mu\bar{\mu}} > 8.2$	$3.5\times10^2\cdot(\text{TeV}/\Lambda)^2$	1.8×10^5	5.0×10^6	4.0×10^5	1.3×10^5	72	0
$\lvert\Delta\phi_{\mu\bar{\mu}}\rvert < 2.2$	$2.8\times10^2\cdot(\text{TeV}/\Lambda)^2$	1.2×10^5	1.2×10^6	2.9×10^5	8.6×10^4	55	0
$p_T^{\mu\mu} > 150$ GeV	$1.2\times10^2\cdot(\text{TeV}/\Lambda)^2$	1.6×10^4	7.9×10^4	3.4×10^4	1.9×10^4	8	0
$M_{\mu\bar{\mu}} > 6.0$ TeV	$1.1\times10^2\cdot(\text{TeV}/\Lambda)^2$	1.1×10^4	5.9×10^4	2.8×10^4	1.4×10^4	5	0
$E_{\min} > 4.1$ TeV	$38\cdot(\text{TeV}/\Lambda)^2$	1.3×10^3	4.8×10^2	3.8×10^2	8.5×10^2	1	0

$\sqrt{s} = 10$ TeV, $\mathscr{L} = 10\text{ab}^{-1}$, $|\eta_{\text{main}}| < 2.44$, $\delta E_{res.} = 10\,\%$

- 5σ discovery at $\Lambda = 360$ GeV

Maximum likelihood ratio

$$\sigma_{\text{exc.}} = \sqrt{-2\ln\left(\frac{L(B\,|\,B)}{L(S+B\,|\,B)}\right)}$$

$$\sigma_{\text{disc.}} = \sqrt{-2\ln\left(\frac{L(B\,|\,S+B)}{L(S+B\,|\,S+B)}\right)}$$

$$L(x\,|\,y) = \frac{x^y}{y!}e^{-x}$$

- Exclusion significance by excluding the signal plus background hypothesis
- Discovery significance by excluding the background only hypothesis

Details on NN: SIG-BG classifier

Input-Layer (10-dim) $\rightarrow$
Hidden-Layer 1 (64-dim) / Batch normalization / ReLU / Dropout($p = 0.3$) $\rightarrow$
Hidden-Layer 2 (64-dim) / Batch normalization / ReLU / Dropout($p = 0.3$) $\rightarrow$
Output-Layer (1-dim)

Network architecture

	200	300	400	500	600	700	800	900	1,000
χ	128	256	128	128	128	256	256	512	256
ϕ	512	128	1,024	256	128	128	512	256	512

Mini-batch size of each NN_M

Optimizer: Adam

Loss function: Binary cross entropy

- Training / Validation
  - 20k events / background, 160k for signal → split 8:2
- Test
  - Same # of events as training/validation
- Weighting
  - Background events weighted by cross section *(loss calculation)*
- Limit
  - Calculated on validation set → verified on test set *(per mass point)*

Details on NN: Mediator discrimination (event-by-event)

Input-Layer (5-dim) $\to$
Hidden-Layer 1 (32-dim) / Batch normalization / ReLU / Dropout($p = 0.3$) $\to$
Hidden-Layer 2 (32-dim) / Batch normalization / ReLU / Dropout($p = 0.3$) $\to$
Output-Layer (1-dim)

Network architecture

Optimizer: Adam

Loss function: Binary cross entropy

Mini-batch size: 512

- Training / Validation / Test
  - 40k events / model → split 8:1:1
- Pseudo-experiment
  - Multiple sets
    - 15 sets / signal & background (for statistical independence)
  - Variation
    - Gaussian $\pm 5\ \%$ on # of passing events (for statistical fluctuations)
  - 10k runs → construct Λ distribution

Details on NN: Mediator discrimination (transformer)

Input: $\mathbf{X} \in \mathbb{R}^{B\times N\times D} \;\rightarrow$

Multi-Head Self-Attention (8 heads, each with $d = 1$) $\rightarrow$

Transpose to $\mathbb{R}^{B\times D\times N} \;\rightarrow$

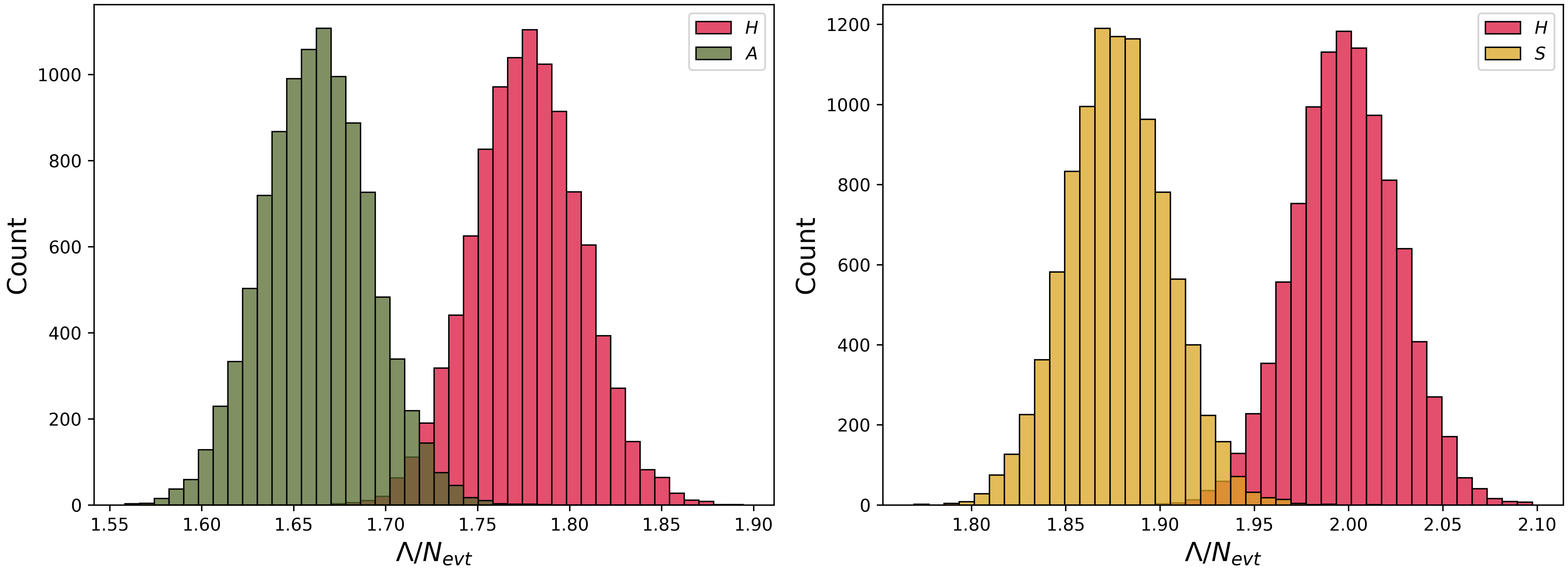
1D Pooling over sequence dimension $(N \rightarrow 1)$

Transformer layer architecture

Classifier part is identical to event-by-event classifier

- Training / Validation
  - 500k events / model → grouped into input matrices → split 8:2
- Pseudo-experiment
  - Input matrix construction
    - Total events may not divide evenly → **last event duplicated** to fill the group
  - Multiple sets
    - 15 sets / signal & background (for statistical independence)
  - Variation
    - Gaussian $\pm 5\,\%$ on # of passing events (for statistical fluctuations)
  - 10k runs → construct Λ distribution

Is the improved performance truly due to the transformer?



Event-by-event classifier, but using four-momentum, $p_\mu, p_{\bar{\mu}}$, as input features.

Better results aren't due to different input features.

	H vs A	H vs S
Z	3.9σ	4.4σ