Probing Higgs Self-coupling
At TeV Muon Colliders

K.C., Jae Sik Lee, Soojin Lee, Chen Wang 2608.01888

FLASY 2026

Motivations

1. Since the Higgs discovery in 2012, there have not been any breakthroughs in BSM.

2. Probing the Higgs self-couplings is vital in understanding the EWSB, related to origin of mass.

3. The LHC currently constrains:
   $-0.71 < \kappa_3 < 6.1$ at 68% C.L. $(\kappa_3 \equiv \lambda_{3h}/\lambda_{3h}^{\rm sm})$

4. Even the HL-LHC is projected to constrain
   $0.5 < \kappa_3 < 1.6$ at 68% C.L.

5. Muon colliders at TeV could give a sensitivity at (our new results)
   $0.8 < \kappa_3 < 1.3$ at 68% C.L. (3 TeV, 1 ab^{-1})
   $0.96 < \kappa_3 < 1.05$ at 68% C.L. (10 TeV, 10 ab^{-1})

Understand Higgs Interactions

- Basically the Higgs boson couples to massive particles, proportional to the mass.
- So the Higgs boson mainly interacts with W, Z bosons, top quark. Because they are heavy.

- $H \to WW, ZZ$: The couplings of H to WW and ZZ are

$$\mathcal{L} = g m_W \, H \, W^{+\mu} \, W^{-}_{\mu} + \frac{1}{2} g_z m_Z H \, Z \, Z$$

- $H \to f \bar{f}$: The decay into a fermion pair is given by

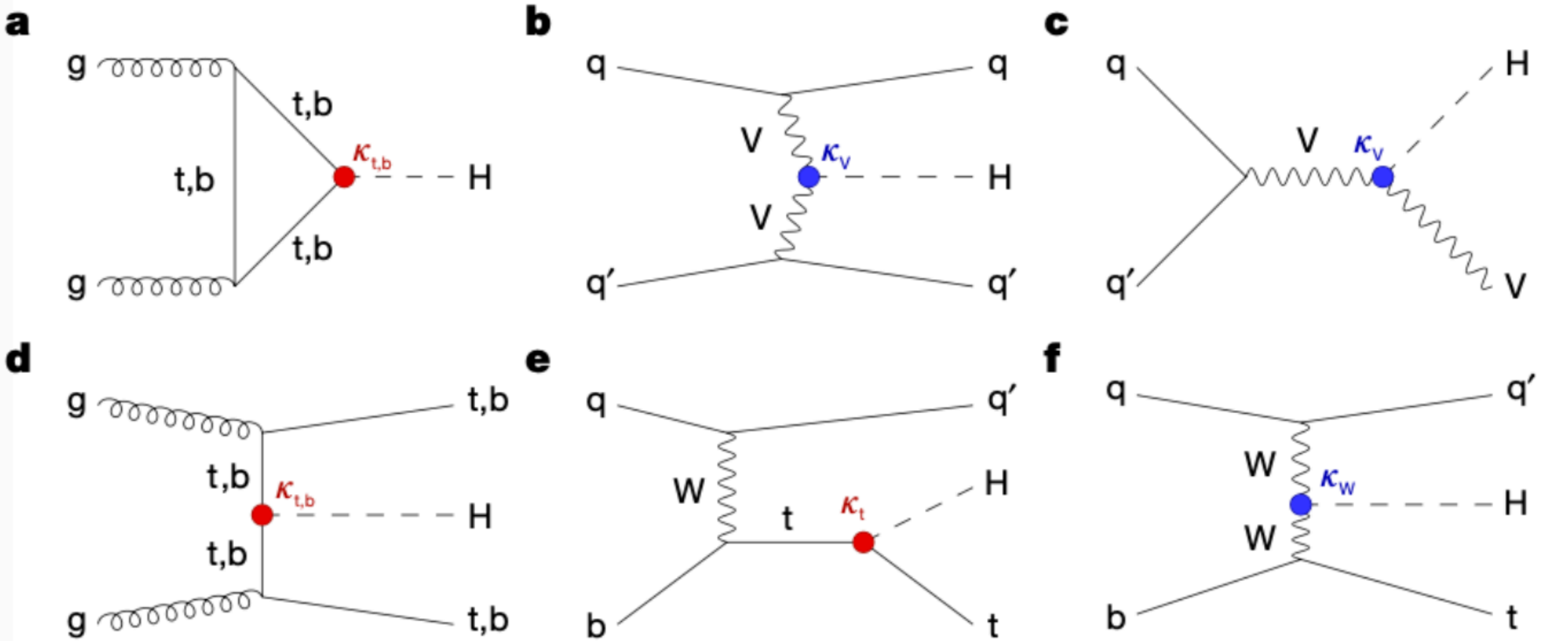
$$\mathcal{L} = -\frac{g m_f}{2 m_W} H \, \bar{f} \, f$$

- $H \to gg$: Higgs decays into a pair of gluons via a triangular loop described by an effective Lagrangian

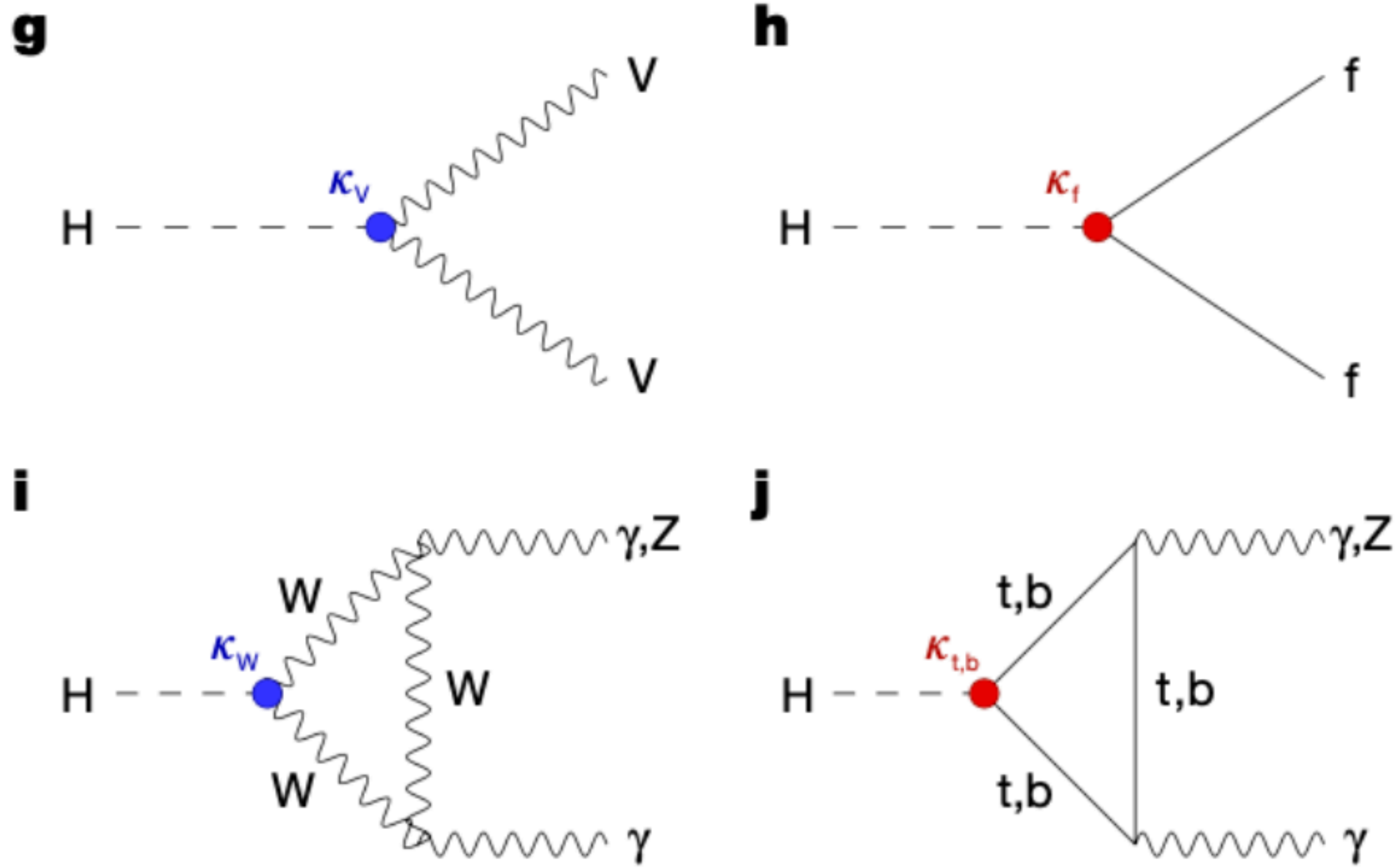
$$\mathcal{L} = -\frac{g^2}{2 m_W} \frac{\alpha_s(m_H)}{12\pi} \, I \, G^a_{\mu\nu} \, G^{a\,\mu\nu} \, H$$

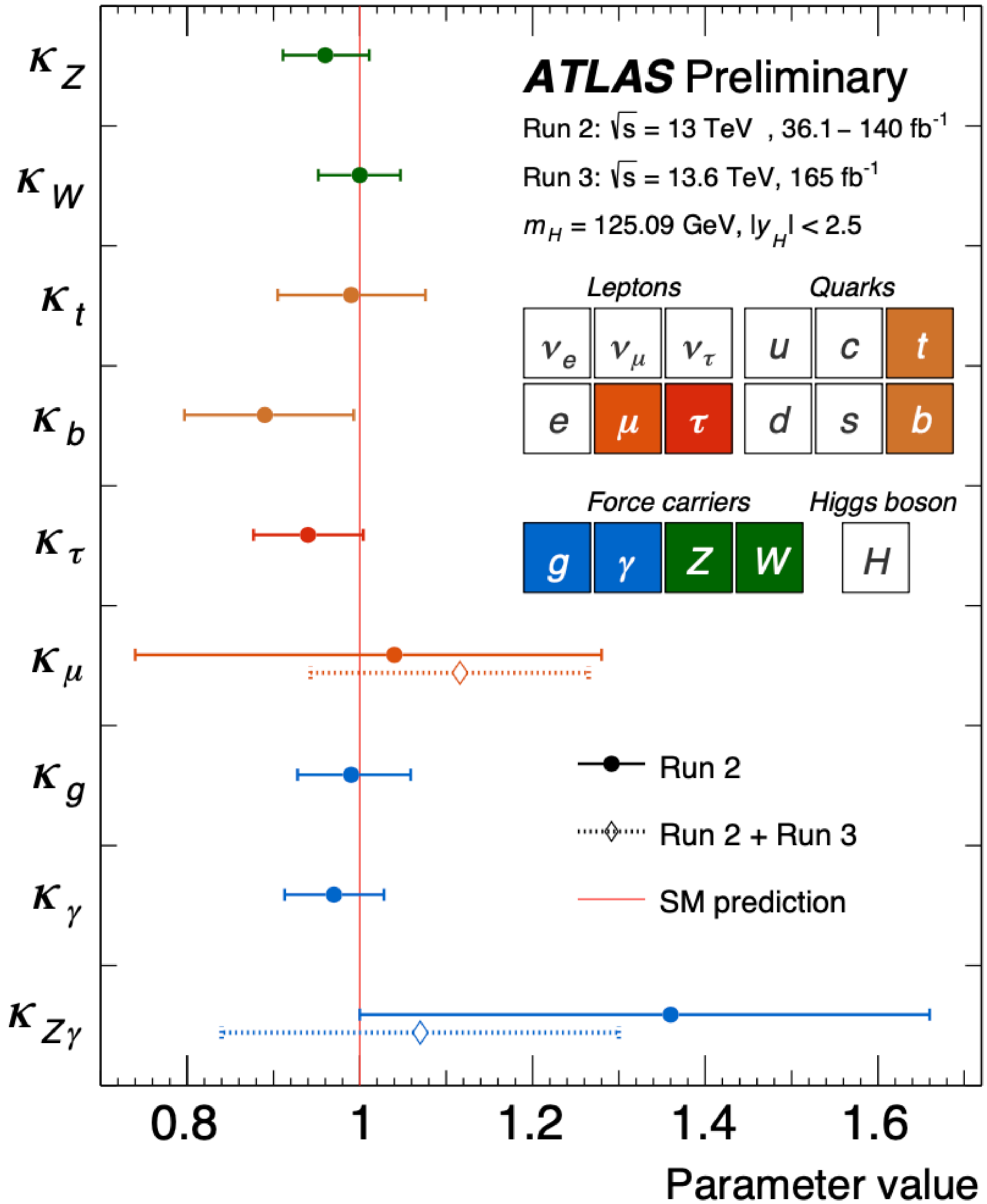
- $H \to \gamma\gamma, Z\gamma$:

Higgs boson production modes



Higgs boson decay channels





ATLAS-CONF-2025-006

Higgs Sector Itself

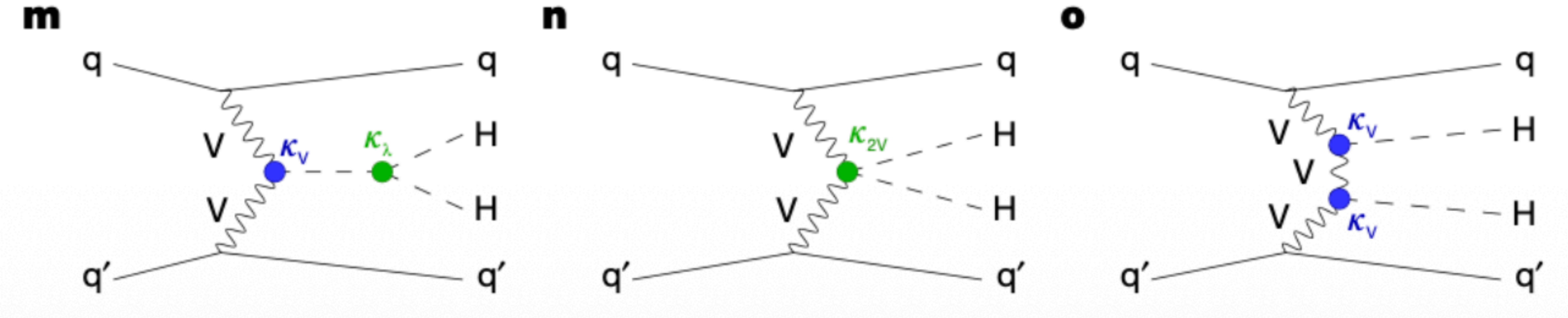
We have no information about $V(\Phi)$ except that it gives a nontrivial VEV. In the SM,

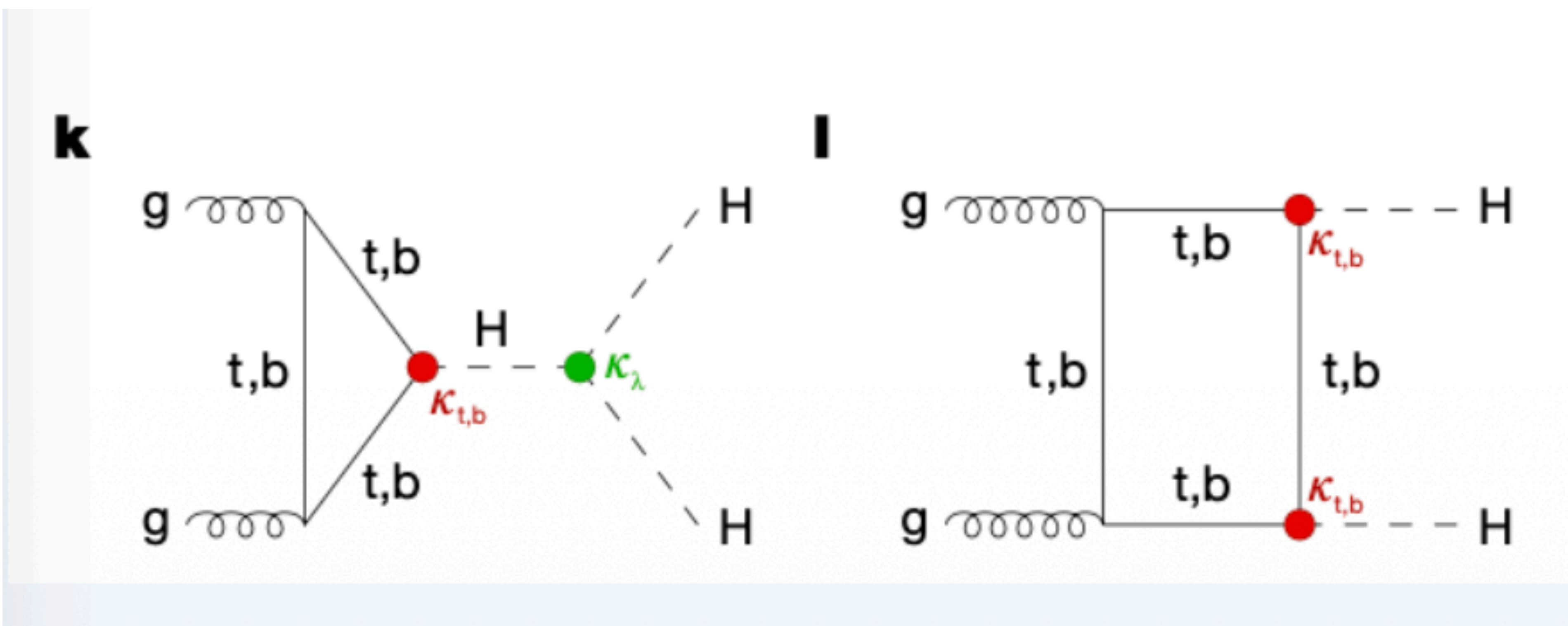
$$V(\phi) = -\frac{\lambda}{4}v^4 + \frac{1}{2}m_H^2 H^2 + \frac{m_H^2}{2v}H^3 + \frac{\lambda}{4}H^4$$

This is the simplest structure. The self couplings are fixed. But for extended Higgs sector it is not the case.

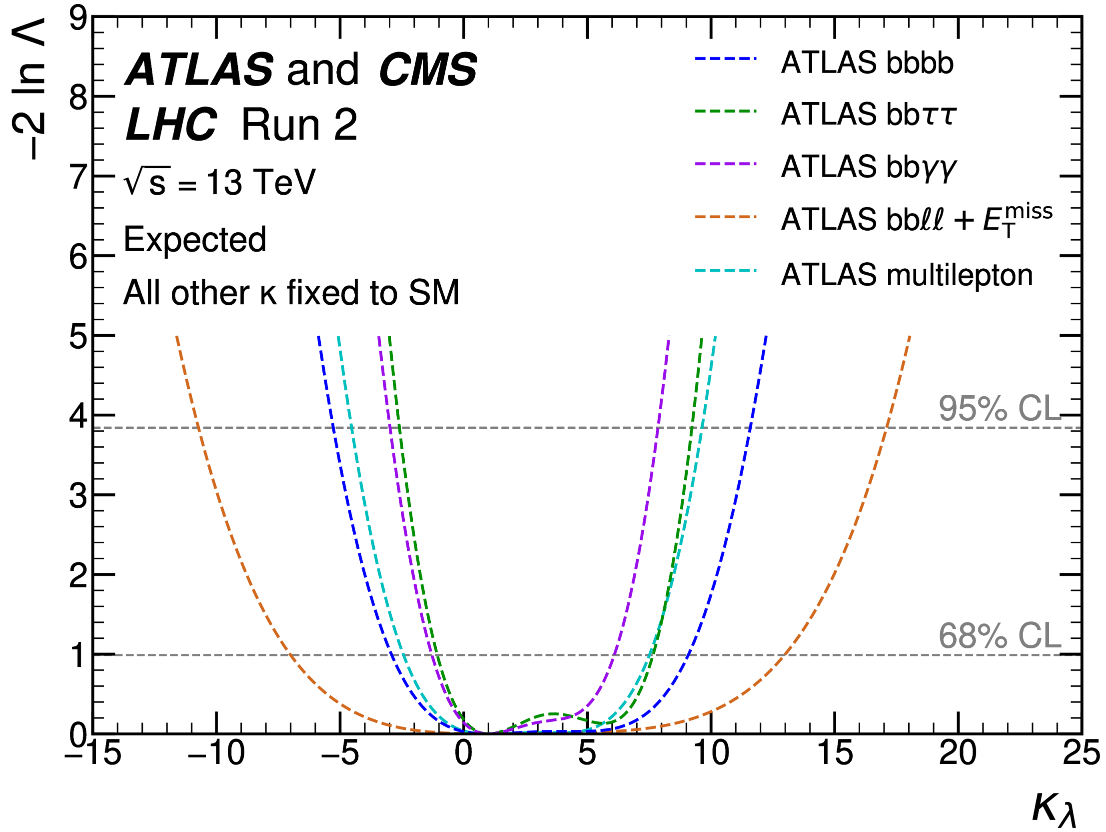
Probing self interactions of the Higgs boson becomes an important avenue to understand the Higgs sector.

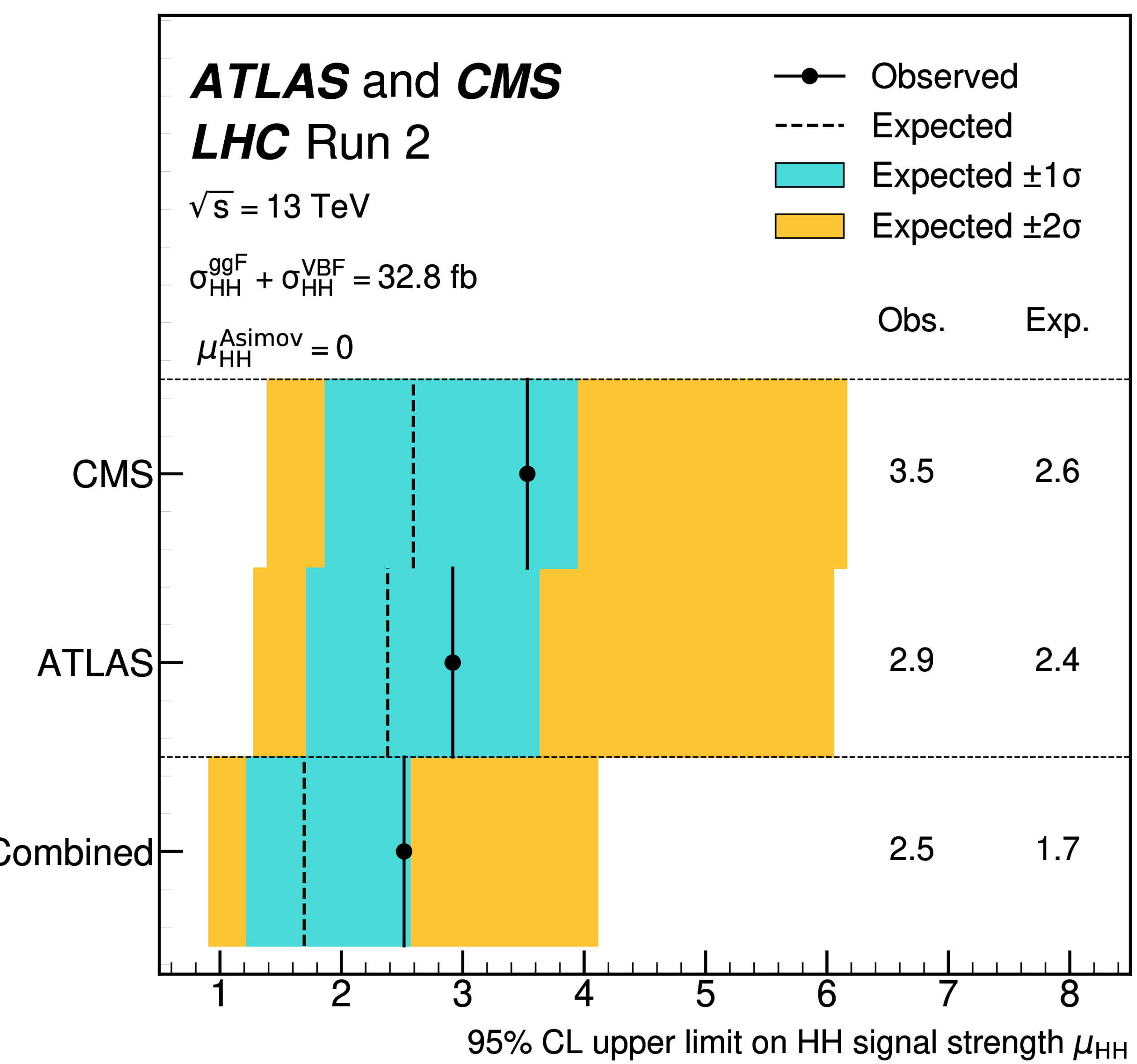
Higgs boson pair production

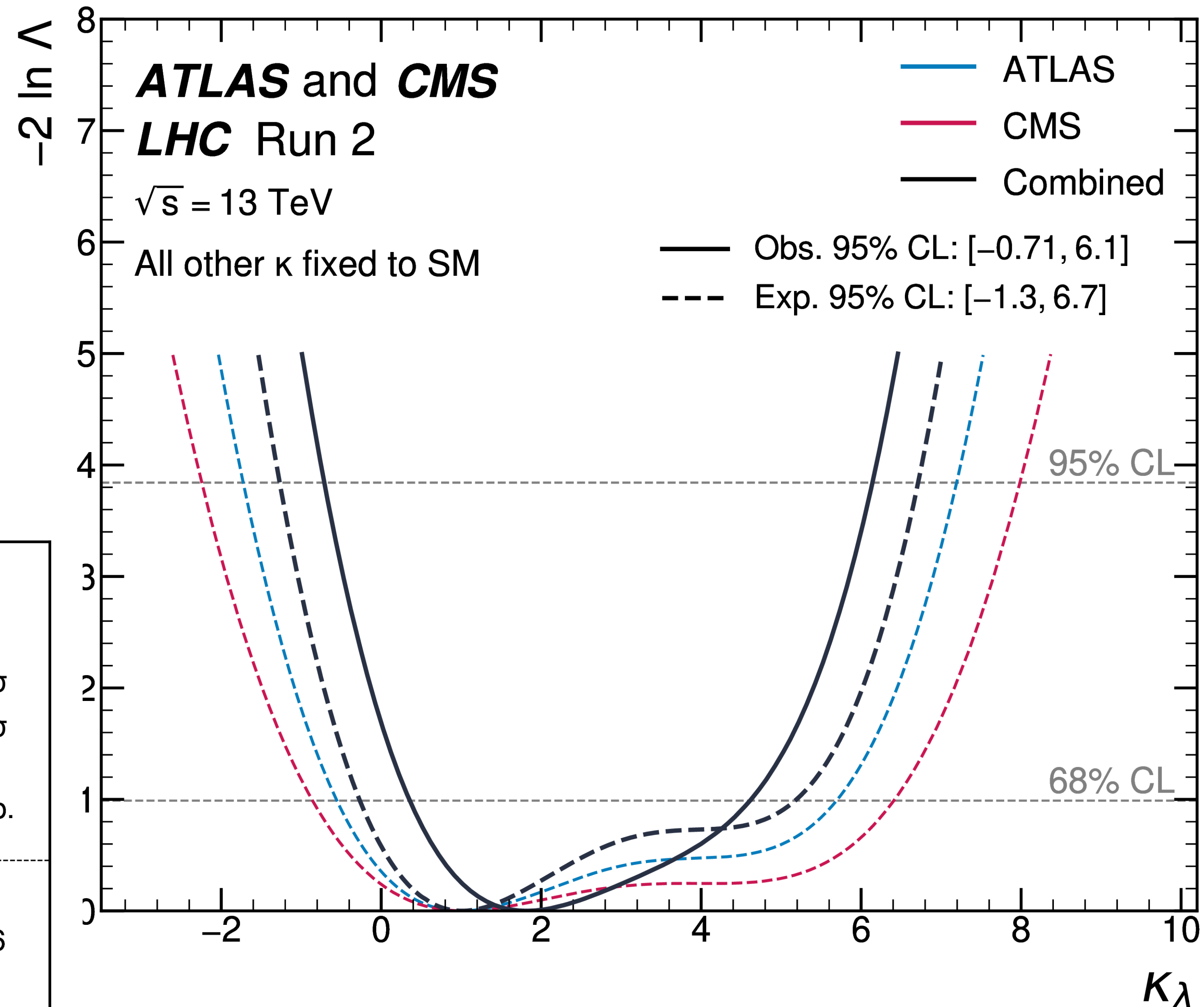




	Channel	Details	Expected			Observed		
			μ_{HH}	κ_λ	κ_{2V}	μ_{HH}	κ_λ	κ_{2V}
ATLAS	bbbb		< 8.1	[−5.3, 12]	[0.38, 1.7]	< 5.4	[−3.4, 11]	[0.55, 1.5]
ATLAS	$bb\tau\tau$		< 3.3	[−2.6, 9.2]	[−0.24, 2.4]	< 5.8	[−3.3, 9.1]	[−0.51, 2.7]
ATLAS	$bb\gamma\gamma$		< 5.2	[−3.0, 7.9]	[−1.1, 3.3]	< 4.1	[−1.6, 7.0]	[−0.48, 2.7]
ATLAS	$bb\ell\ell + E_T^{miss}$		< 14	[−11, 17]	[−0.50, 2.7]	< 9.6	[−6.5, 13]	[−0.19, 2.4]
ATLAS	Multilepton		< 11	[−4.5, 9.6]	[−1.9, 4.1]	< 17	[−6.4, 12]	[−2.5, 4.7]
ATLAS	Combined		< 2.4	[−1.7, 7.2]	[0.41, 1.7]	< 2.9	[−1.3, 7.2]	[0.57, 1.5]
CMS	bbbb		< 4.3	[−4.6, 12]	[0.63, 1.4]	< 7.0	[−5.0, 12]	[0.66, 1.4]
CMS	$bb\tau\tau$		< 5.4	[−4.2, 11]	[−0.64, 2.8]	< 3.5	[−1.9, 8.9]	[−0.32, 2.5]
CMS	$bb\gamma\gamma$		< 5.7	[−3.5, 8.8]	[−0.94, 3.1]	< 8.7	[−3.5, 8.0]	[−1.4, 3.6]
CMS	bbWW		< 19	[−9.4, 16]	[−1.4, 3.5]	< 15	[−6.1, 13]	[−1.0, 3.1]
CMS	Multilepton		< 20	[−8.0, 12]	[−2.5, 4.6]	< 22	[−5.8, 10]	[−3.4, 5.6]
CMS	Combined		< 2.6	[−2.2, 8.0]	[0.63, 1.4]	< 3.5	[−1.4, 6.6]	[0.66, 1.4]
ATLAS + CMS combined			< 1.7	[−1.3, 6.7]	[0.66, 1.4]	< 2.5	[−0.71, 6.1]	[0.73, 1.3]

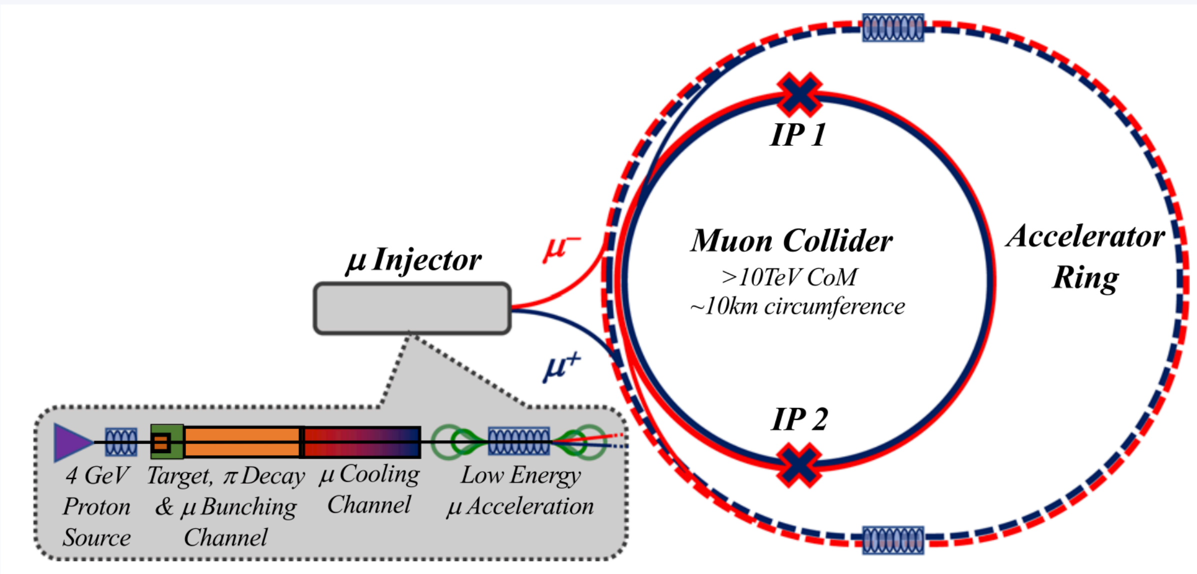






HL-LHC (3 ab^{-1}) projection: $0.5 < \kappa_3 < 1.6$ (68% CL)

Why a Muon Collider?



This work: two energy stages

Stage 1	Stage 2
$\sqrt{s}$ = 3 TeV	$\sqrt{s}$ = 10 TeV
$\mathcal{L}$ = 1 ab^{-1}	$\mathcal{L}$ = 10 ab^{-1}
Ring: ~4.5 km	Ring: ~10 km

Advantages over LHC for κ_3 measurement

1. **Point-like muon**

   Full $\sqrt{s}$ available; no PDF dilution from parton distributions

2. **Clean lepton-collider environment**

   No underlying event, no QCD multijet — but BIB mitigation required

3. **QCD suppression**

   EW backgrounds only (WW, ZZ, $t\bar{t}$) — no QCD multijet flood

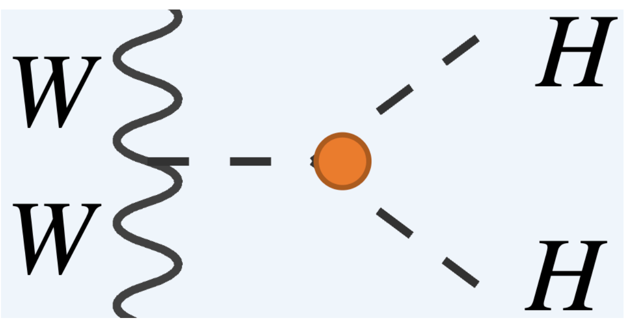
4. **No synchrotron loss**

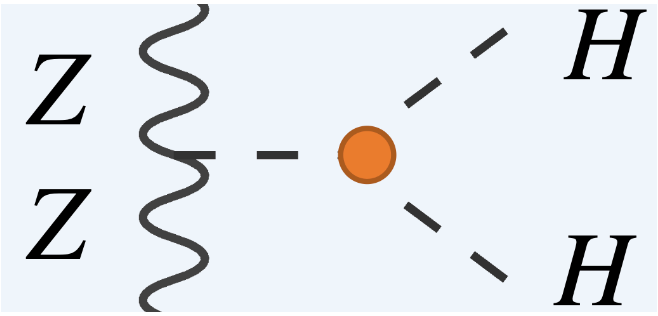
   200× heavier than e^+e^- → compact multi-TeV ring feasible

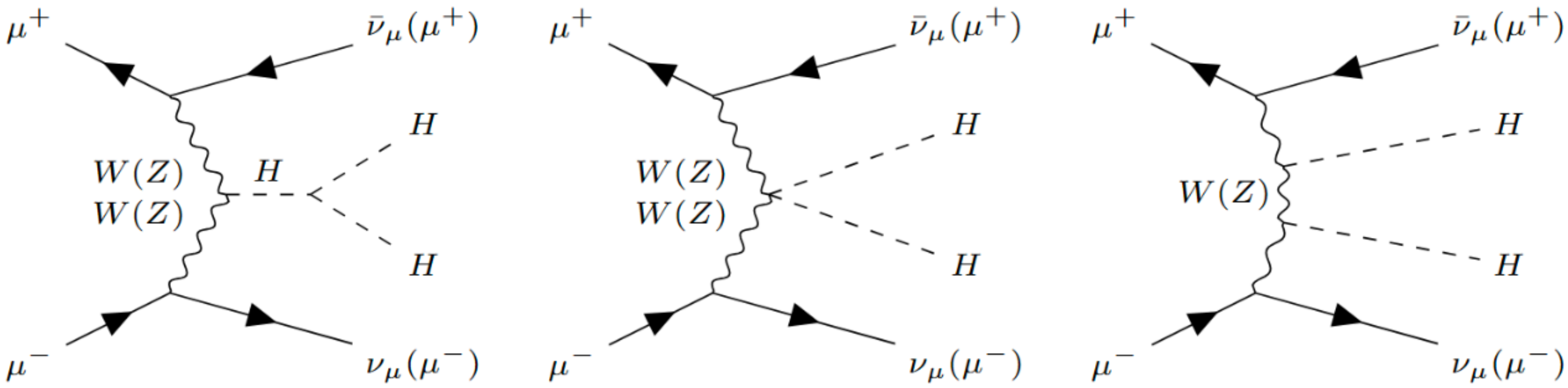
Special thanks to Soojin for using some of her slides

Probing Trilinear Higgs Coupling at Muon Collider

Central part λ_3







Involve $\kappa_3 \equiv \lambda_3/\lambda_3^{\rm sm}, \kappa_{2V}, \kappa_V$

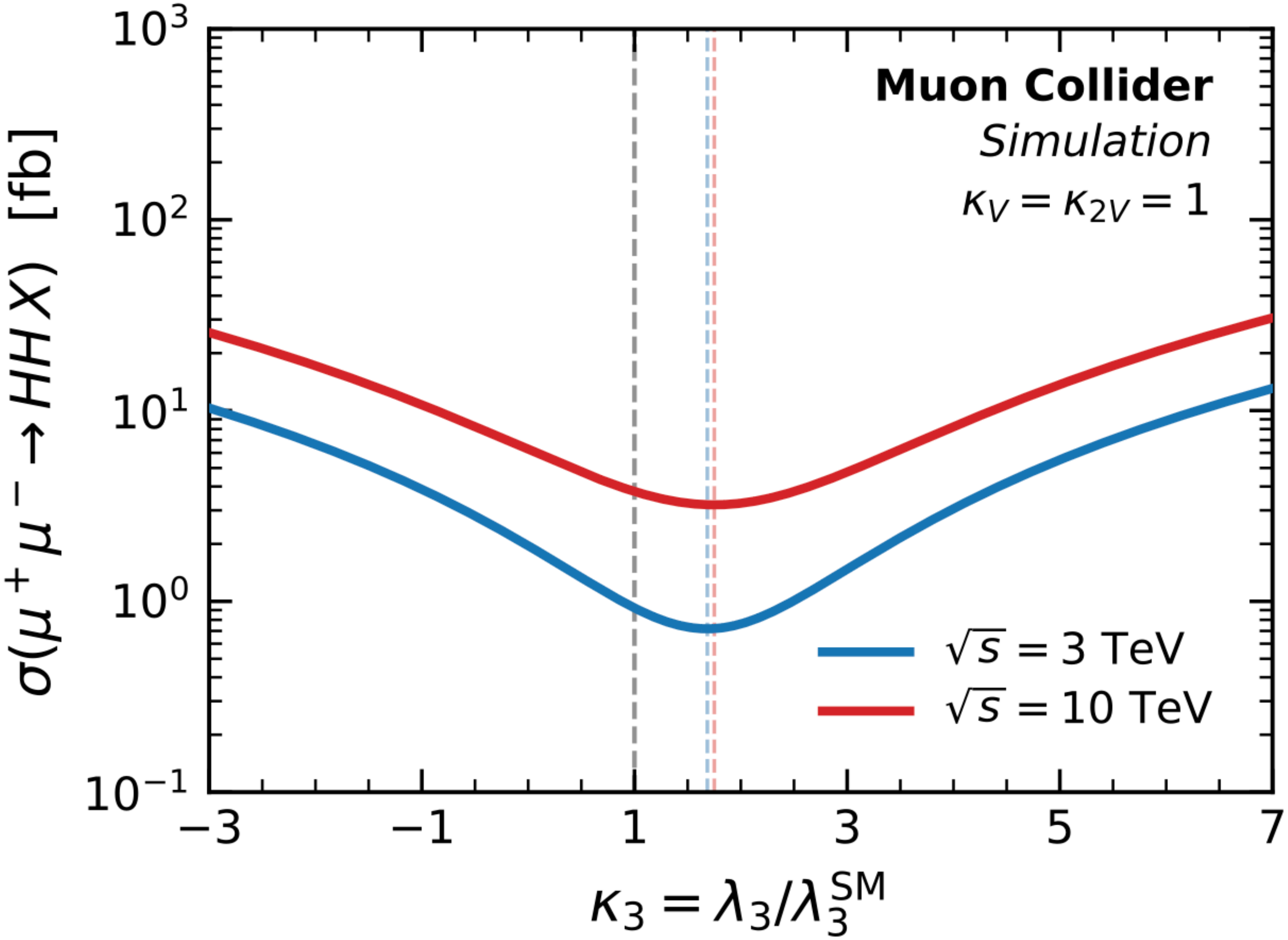
- In the SM Higgs potential: $V(H) = -\mu^2 H^\dagger H + \lambda(H^\dagger H)^2$ the trilinear coupling is fixed:
  $\lambda_3 = m_H^2/(2v) = \lambda v$

- Current ranges of κ's:
  $\kappa_W = 1.02 \pm 0.08, \qquad \kappa_Z = 1.04 \pm 0.07$
  $0.73 < \kappa_{2V} < 1.3 \ (95\,\%\ CL)$, CMS + ATLAS
  $-0.71 < \kappa_3 < 6.1 (95\,\%\ CL)$, CMS + ATLAS

- Channels:
  $\mu^-\mu^+ \to \nu_\mu \bar{\nu}_\mu HH\ , \quad \mu^-\mu^+ \to \mu^-\mu^+ HH$,
  followed by
  $HH \to b\bar{b}b\bar{b}$

- We fix $\kappa_V = \kappa_{2V} = 1$

- Scan $0.2 \leq \kappa_3 \leq 1.8$



- For muons in the neutral current final state: $|\eta_\mu| > 2.5$ to escape detection
- $\kappa_3^{\min} = 1.68\ (1.75)$ at 3 TeV (10 TeV).
- For $\kappa_3 = 1$, $\sigma(HH) \times [B(H \to b\bar{b})]^2 = 0.312\ (1.27)\,\mathrm{fb}$ at 3 TeV (10TeV)

Strategies

- Goals:
  case 1 — Observability of HH production above the SM bkgd.
  null hypothesis :: SM background
  test hypothesis :: SM background + HH (various κ_3)

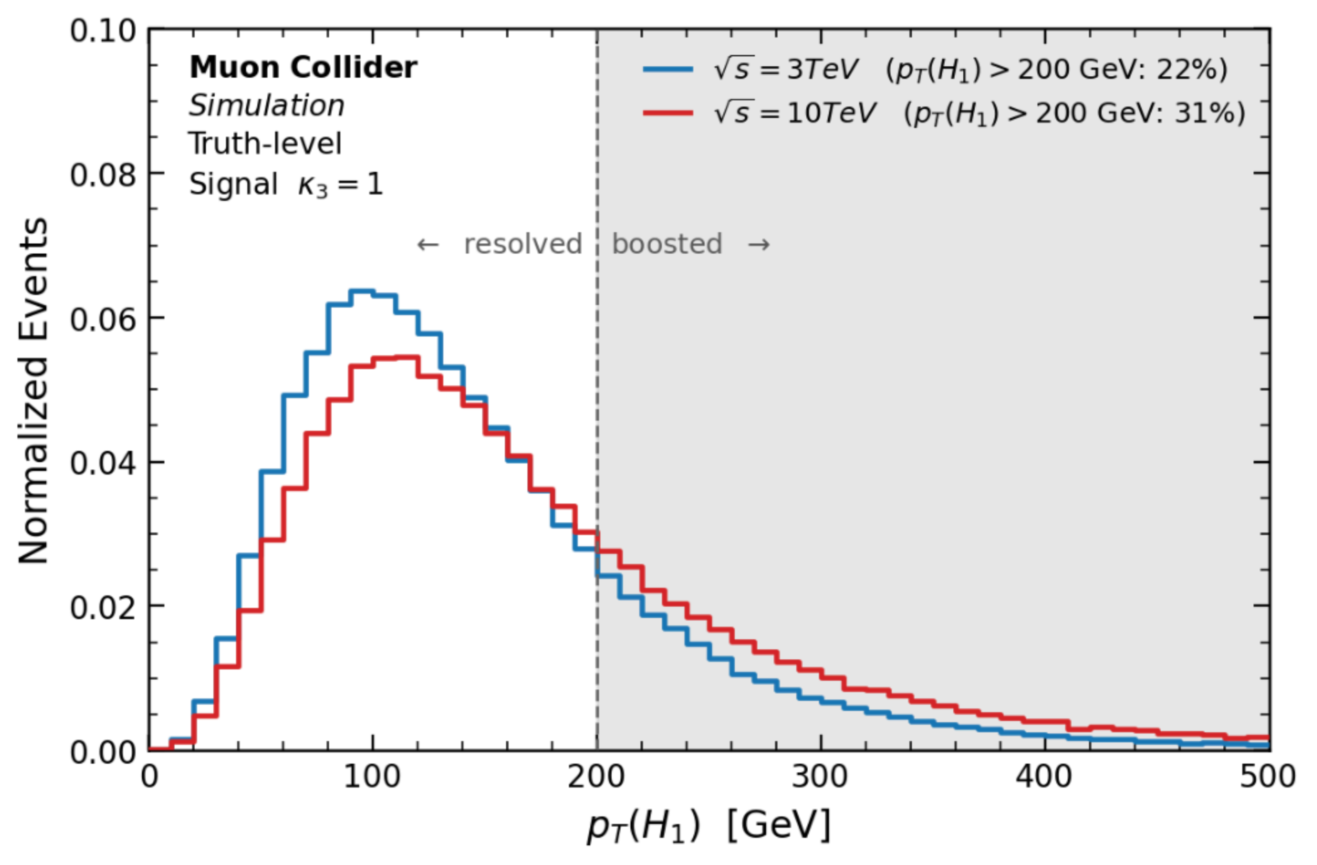
  case 2 — Discrimination of HH with $\kappa_3 \neq 1$ from $\kappa_3 = 1$
  null hypothesis :: SM background + HH ($\kappa_3 = 1$)
  test hypothesis :: SM background + HH ($\kappa_3 \neq 1$)

- Divide the kinematic regions exclusively with
  $p_T(H) < 200\,\mathrm{GeV}$ for resolved region
  $p_T(H) > 200\,\mathrm{GeV}$ for boosted region

  because $H \to b\bar{b}$ behaves quite differently, as

$$\Delta R(b,\bar{b}) \sim \frac{2m_H}{p_T} \simeq 1 \quad \text{for} \quad p_T(H) = 200\,\mathrm{GeV}$$

Two Complementary Regions: Resolved and Boosted



Non-overlapping kinematic regions → statistically independent → combined at likelihood level

Resolved

- pT(H) < 200 GeV (for both Higgs candidates)
- 4 small-R jets (VLC, R=0.5)
- 78% (69%) of SM signal @ 3 TeV (10 TeV)

Boosted

- pT(H) > 200 GeV (for both Higgs candidates)
- 2 large-R fat jets (VLC, R=1.0)
- 7% (10%) of SM signal @ 3 TeV (10 TeV)

Special thanks to Soojin for using some of her slides

Signal and Background Processes

Resolved (4 small-R jets, R=0.5)

S	**$\mu^+\mu^- \rightarrow HH\, \nu\bar{\nu}$ (W-fusion)**	4b + MET
S	**$\mu^+\mu^- \rightarrow HH\, \mu^+\mu^-$ (Z-fusion)**	4b + μμ (forward)
B	$\mu^+\mu^- \rightarrow H\, q\bar{q}\, \nu\bar{\nu}$	2b2j + MET
B	$\mu^+\mu^- \rightarrow W^+W^-\, \nu\bar{\nu}$	4j + MET
B	$\mu^+\mu^- \rightarrow ZZ\, \nu\bar{\nu}$	4j + MET
B	$\mu^+\mu^- \rightarrow t\bar{t}\, \nu\bar{\nu}$	4j + MET (hadronic top)
B	$\mu^+\mu^- \rightarrow W^+W^-$	4j (s-channel)
B	$\mu^+\mu^- \rightarrow ZZ$	4j (s-channel)
B	$\mu^+\mu^- \rightarrow t\bar{t}$	4j (s-channel)

VBF-like ($\nu\bar{\nu}$) topologies dominate after basic selection

Boosted (2 fat jets, R=1.0)

S	**$\mu^+\mu^- \rightarrow HH\, \nu\bar{\nu}$ (W-fusion)**	2 fat b-jets + MET
S	**$\mu^+\mu^- \rightarrow HH\, \mu^+\mu^-$ (Z-fusion)**	2 fat b-jets + μμ (fwd)
B	$\mu^+\mu^- \rightarrow jj$	2 fat jets
B	$\mu^+\mu^- \rightarrow HZ\, \nu\bar{\nu}$	fat b-jet + fat Z-jet + MET
B	$\mu^+\mu^- \rightarrow b\bar{b}$	2 fat b-jets
B	$\mu^+\mu^- \rightarrow W^+W^-\, \nu\bar{\nu}$	2 fat jets + MET
B	$\mu^+\mu^- \rightarrow W^+W^-$	2 fat jets
B	$\mu^+\mu^- \rightarrow ZZ\, \nu\bar{\nu}$	2 fat jets + MET
B	$\mu^+\mu^- \rightarrow ZZ$	2 fat jets
B	$\mu^+\mu^- \rightarrow t\bar{t}$	2 fat jets

Special thanks to Soojin for using some of her slides

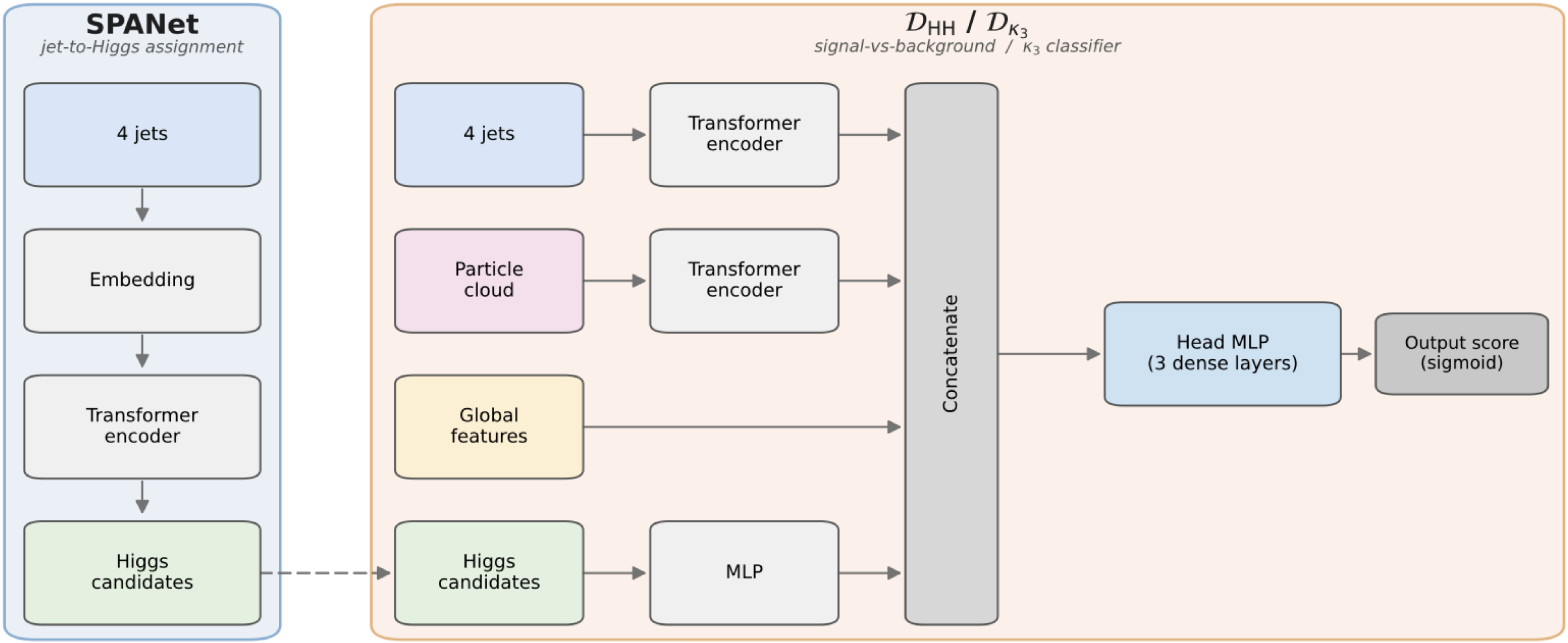
Resolved Region

- Pre-selection:
  1. Four b jets with $R = 0.5, p_T(j) > 20\,\mathrm{GeV}, |\eta_j| < 2.5$
  2. No isolated leptons
  3. Both Higgs $p_T(H) < 200\,\mathrm{GeV}$

	$\sqrt{s} = 3\,\mathrm{TeV}$			$\sqrt{s} = 10\,\mathrm{TeV}$		
Process	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}
HHX	0.312	42.4	132	1.273	27.2	3,467
$Hq\bar{q}\nu\bar{\nu}$	3.164	45.2	1,431	8.520	32.9	27,989
$ZZ\nu\bar{\nu}$	12.16	45.3	5,514	23.93	36.7	87,860
$W^+W^-\nu\bar{\nu}$	19.52	47.6	9,292	33.97	40.6	137,855
$t\bar{t}\nu\bar{\nu}$	2.228	58.8	1,310	6.982	45.4	31,726
ZZ	0.755	14.4	109	0.0004	3.7	0.1
W^+W^-	9.191	22.2	2,043	0.0026	4.4	1.1
$t\bar{t}$	7.792	1.6	122	0.704	0.3	23.4
Total background	54.81	—	19,821	74.11	—	285,455

	$\sqrt{s} = 3\,\text{TeV}$			$\sqrt{s} = 10\,\text{TeV}$		
κ_3	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}
0.2	0.567	45.5	258	1.886	30.8	5,811
0.4	0.485	44.9	218	1.689	30.1	5,086
0.6	0.417	43.8	182	1.527	29.3	4,479
0.8	0.358	43.1	154	1.383	28.3	3,913
0.9	0.333	42.8	143	1.324	27.9	3,697
1.0	0.312	42.4	132	1.273	27.2	3,467
1.1	0.294	41.9	123	1.225	27.0	3,308
1.2	0.278	41.3	115	1.184	26.5	3,135
1.4	0.255	40.6	104	1.127	26.0	2,925
1.6	0.245	40.2	98	1.092	25.5	2,783
1.8	0.246	40.0	98	1.085	25.6	2,774

Machine Learning Model



SPANet

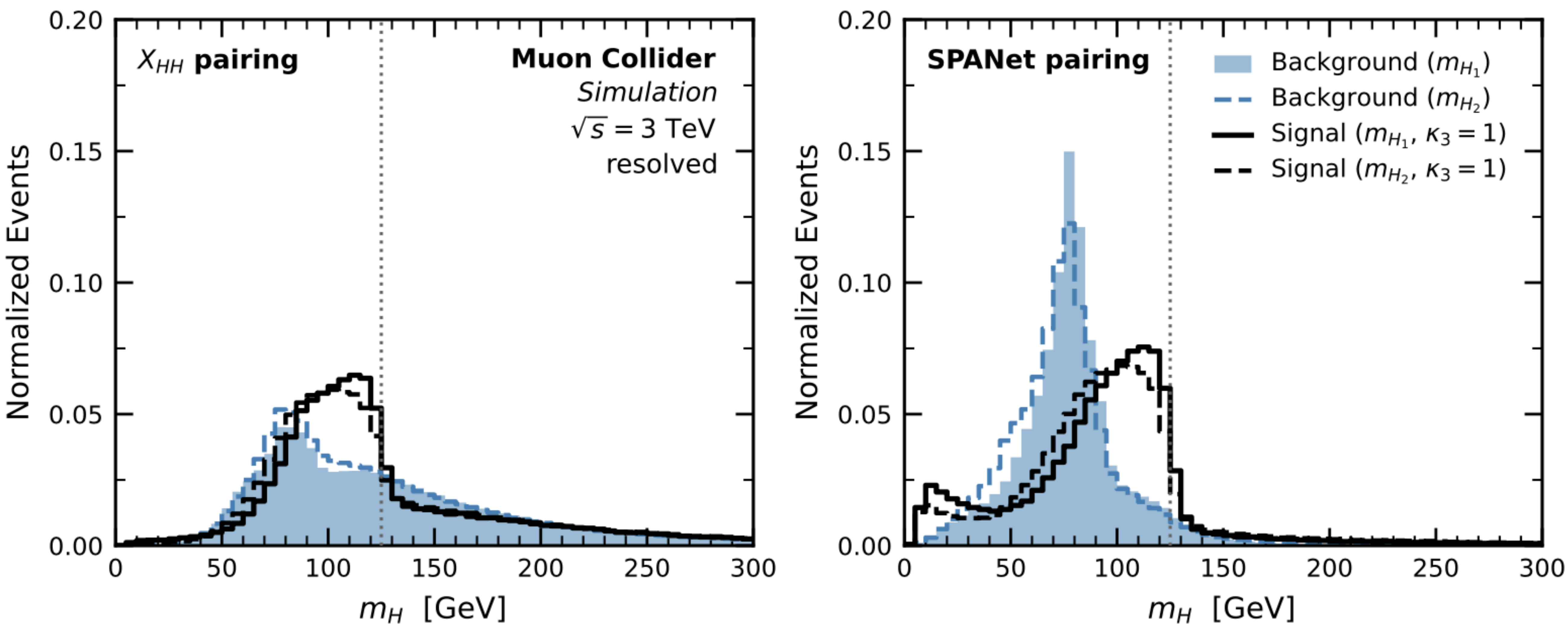
Multi-stream classifier (right) used in the resolved analysis.

- Pairing of the 4 jets into the correct Higgs Boson, traditionally done by minimizing the variable:

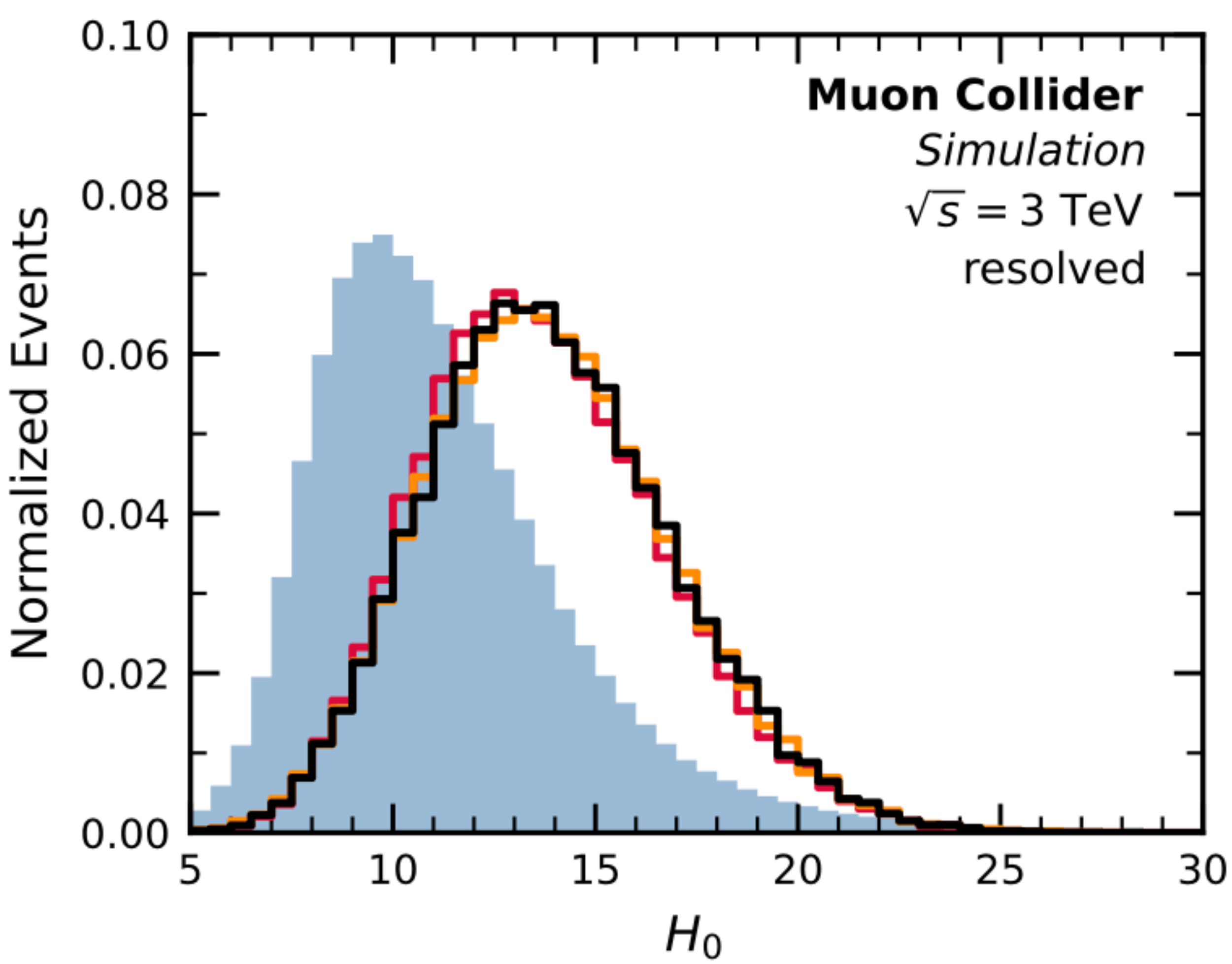
$$X_{HH} = \sqrt{\left(\frac{m_{H_1} - M_1}{0.1 M_{H_1}}\right)^2 + \left(\frac{m_{H_2} - M_2}{0.1 M_{H_2}}\right)^2}$$

with choices of M_1, M_2

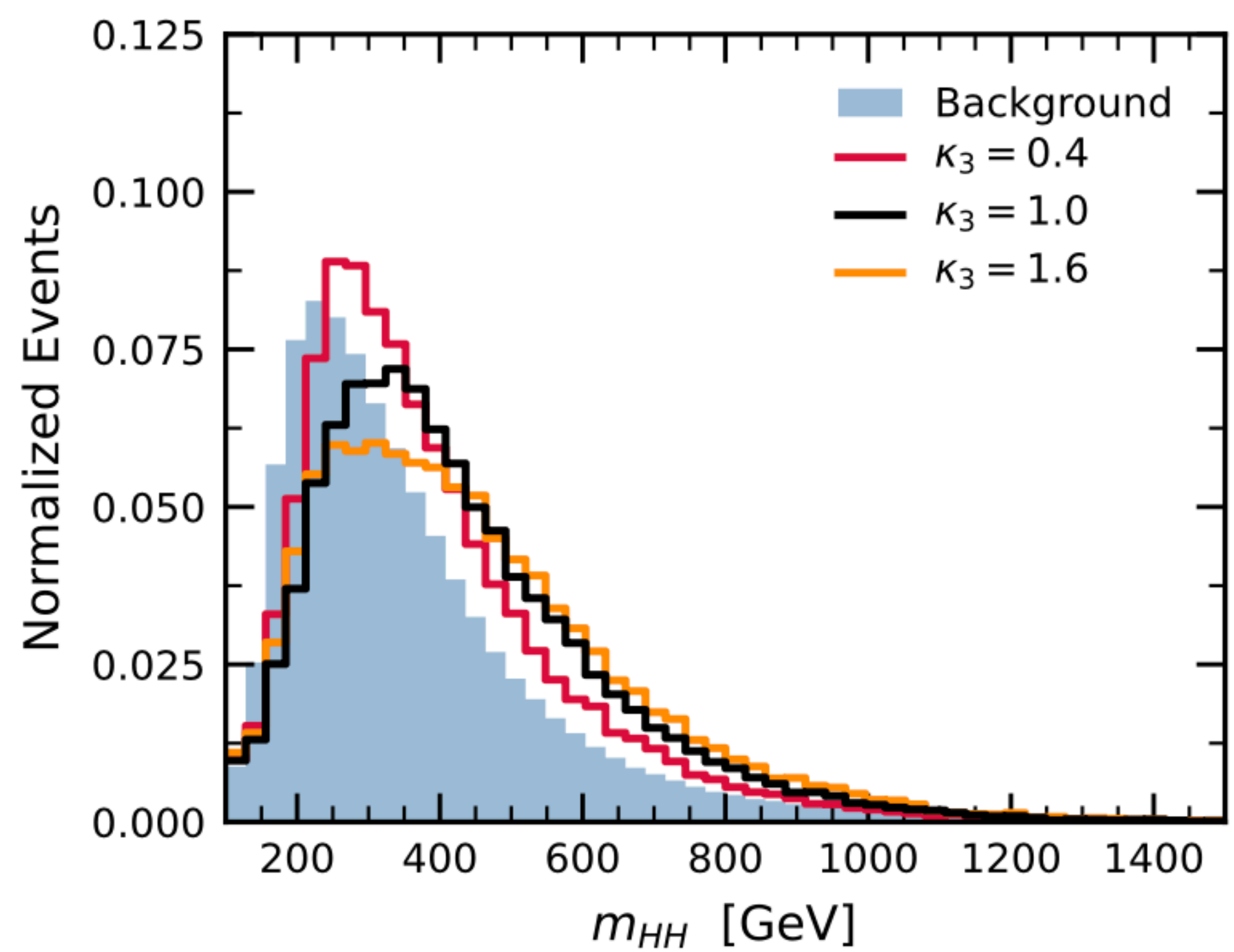
- But the jet-to-Higgs assignment in this work performed by SPANet network, which classifies the three possible pairings of the four jets directly and keeps only the symmetry-preserving structure.



- Four input streams:
  1. 4 jets: $\ln p_T^j, \eta_j, \sin\phi_j, \cos\phi_j, \ln(1+m_j/m_0), |\Delta\phi(j, E_T^{\text{miss}})|, b-tag$
  2. Higgs candidates:
  3. Global features: $m_{HH}, p_T^{HH}, X_{HH}, \Delta\phi(HH, E_T^{\text{miss}}), \Delta R_{H_1,H_2}, b-tag, E_T^{\text{miss}}$, scalar sum of energies, five TDA descriptors H_0, H_1, S_0, S_1, LB_1
  4. Particle Cloud: for the ten leading EFlow constituents of each jet, particle type code



Separates the $H \to 4b$ signal from the bulk of the background

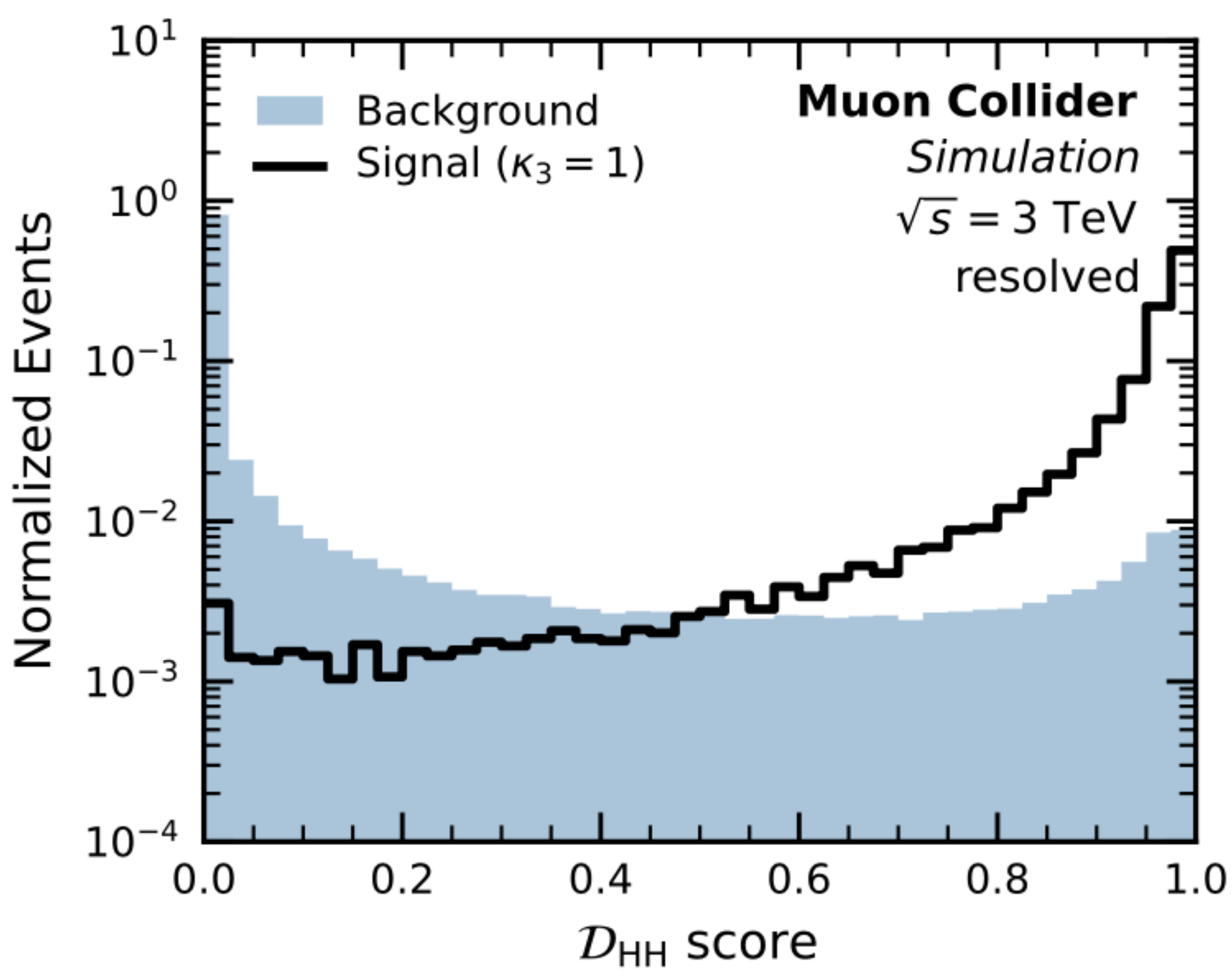


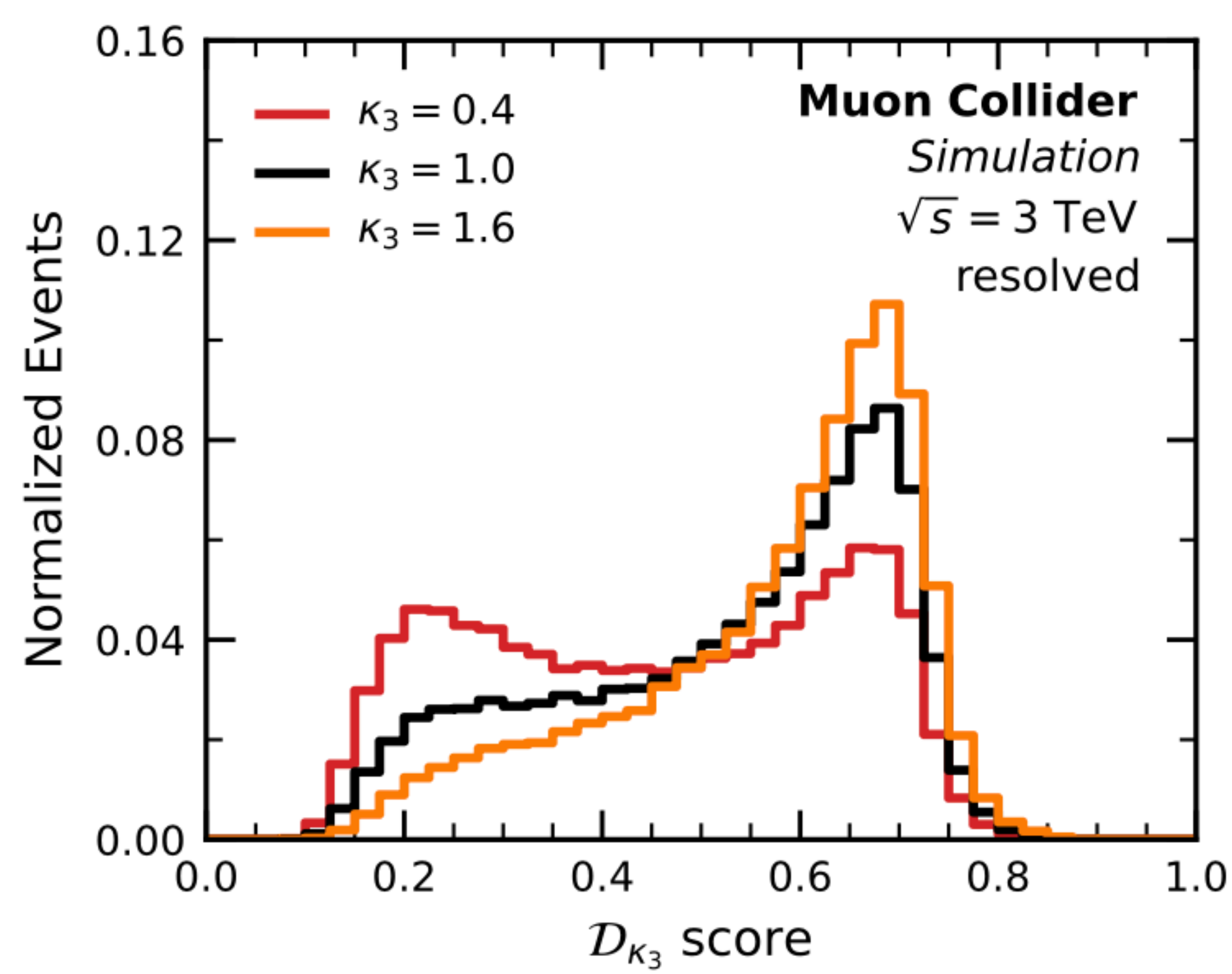
m_{HH} clearly discriminates various κ_3

Two Machine Learning Approaches

- ML1: HH signal vs SM background :
  Null hypothesis :: SM background
  test hypothesis :: SM background + HH (various κ_3)
  –> Classifier D_{HH}

- ML2 : κ_3 shape discriminant :
  null hypothesis :: SM background + HH ($\kappa_3 = 1$)
  test hypothesis :: SM background + HH ($\kappa_3 \neq 1$)
  Trained on signal only with $\kappa_3 = 0.4$ and $\kappa_3 = 1.6$
  –> Classifier D_{κ_3}





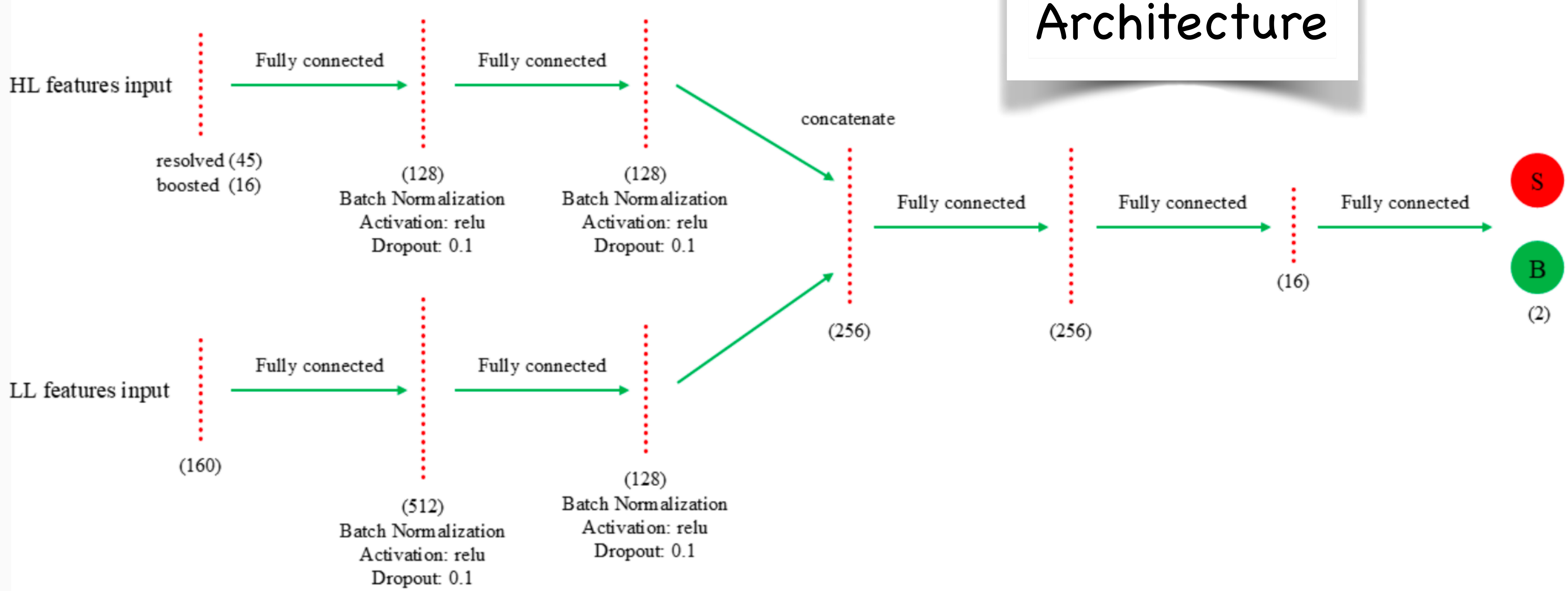
Boosted Region

- Pre-selection:
  1. Two fat jets with $R = 1.0, |\eta_J| < 2.5$
  2. $200\,\mathrm{GeV} < P_T(J_1) < 800\,\mathrm{GeV}$
  3. $200\,\mathrm{GeV} < P_T(J_2) < 600\,\mathrm{GeV}$
  4. $65\,\mathrm{GeV} < M(J_{1,2}) < 150\,\mathrm{GeV}$

	$\sqrt{s} = 3\,\mathrm{TeV}$			$\sqrt{s} = 10\,\mathrm{TeV}$		
Process	$\sigma_{\rm gen}$ [fb]	$\epsilon_{\rm pre}$ [%]	$N_{\rm pre}$	$\sigma_{\rm gen}$ [fb]	$\epsilon_{\rm pre}$ [%]	$N_{\rm pre}$
HHX	0.0208	29.6	6.15	0.128	23.0	293.42
$HZ\nu\bar{\nu}$	0.5196	26.6	138.22	2.8644	13.8	3,953
$ZZ\nu\bar{\nu}$	6.469	24.2	1,567	39.84	14.8	59,022
$W^+W^-\nu\bar{\nu}$	12.23	20.0	2,451	70.54	11.6	81,832
ZZ	9.617	12.2	1,168	1.336	0.0036	0.48
W^+W^-	150.7	10.1	15,287	22.01	0.0014	3.08
$t\bar{t}$	4.424	7.2	316.25	0.0354	3.7	12.97
$b\bar{b}$	2.326	7.8	181.78	0.0187	2.1	3.87
$q\bar{q}$ $(q \neq b)$	13.45	7.1	949.81	0.108	1.9	20.50
Total background	199.7	—	22,059	136.8	—	144,848

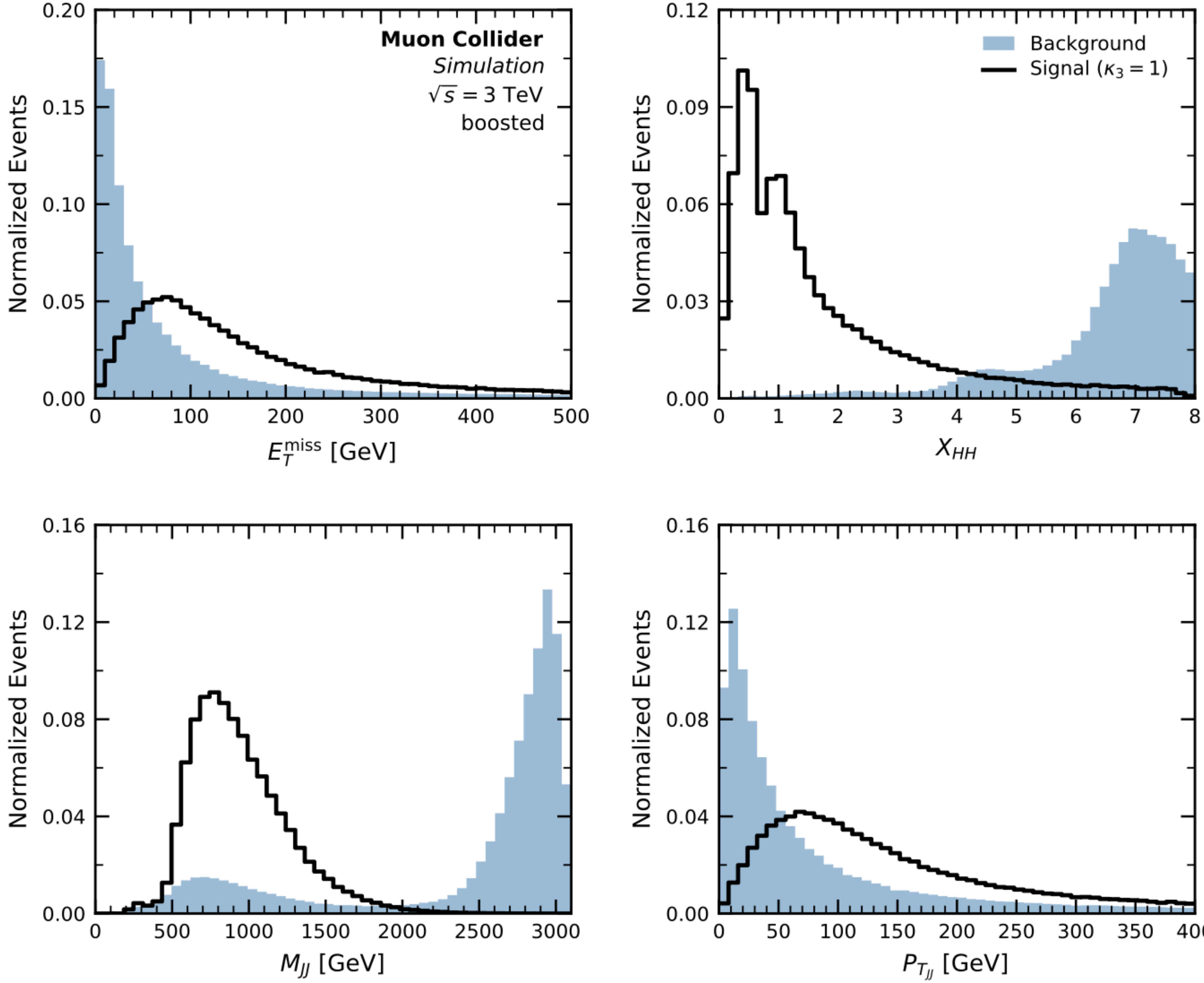
	$\sqrt{s} = 3\,\mathrm{TeV}$			$\sqrt{s} = 10\,\mathrm{TeV}$		
κ_3	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}	σ_{gen} [fb]	ϵ_{pre} [%]	N_{pre}
0.2	0.0455	31.4	14.29	0.221	27.0	596.25
0.4	0.0375	31.0	11.62	0.192	26.2	501.76
0.6	0.0310	30.6	9.48	0.166	25.2	418.39
0.8	0.0251	30.4	7.64	0.145	24.5	353.79
0.9	0.0229	29.9	6.85	0.136	23.8	323.26
1.0	0.0208	29.6	6.15	0.128	23.0	293.42
1.1	0.0190	29.0	5.52	0.121	22.3	269.32
1.2	0.0176	28.6	5.01	0.115	21.8	249.96
1.4	0.0154	27.4	4.22	0.106	20.6	218.19
1.6	0.0145	26.5	3.84	0.101	19.6	197.46
1.8	0.0147	25.7	3.78	0.100	19.2	192.64

Architecture

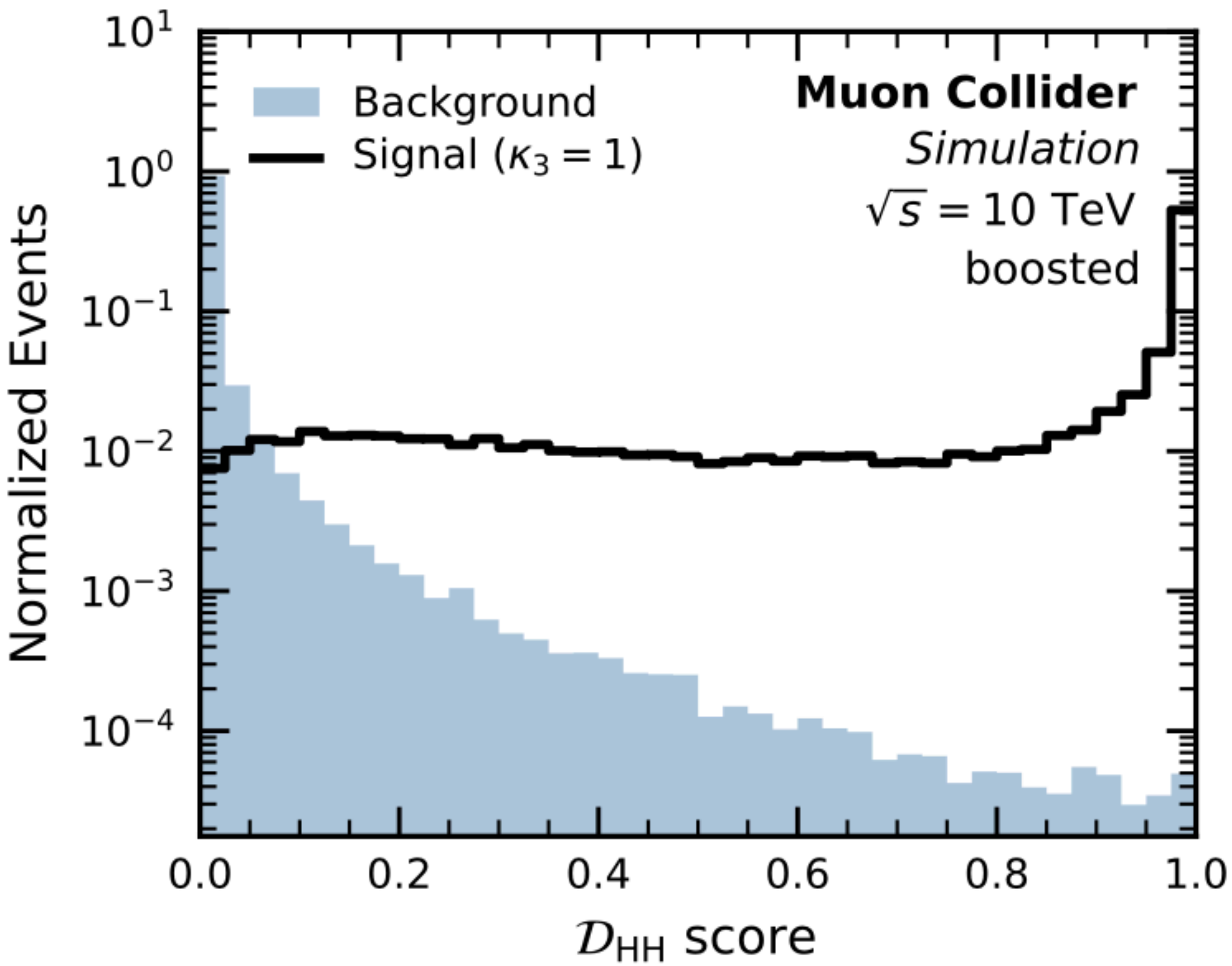


- Two branches: high-level and low-level features

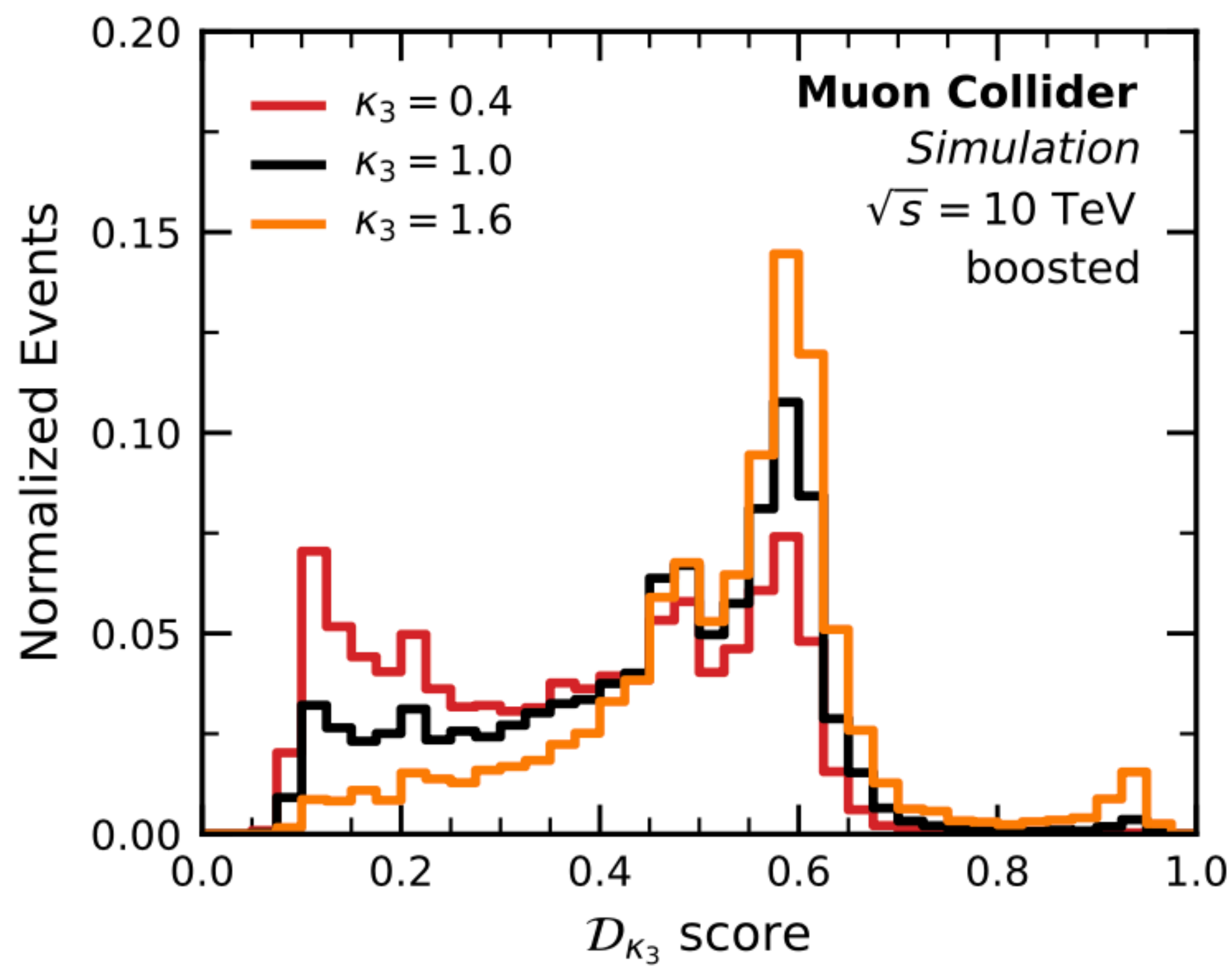
- Fat jets: $M(J), p_T(J), \mathrm{Btag}(J)$
  Global features: $E_T^{\mathrm{miss}}, X_{HH}, M(JJ), P_T(JJ), \Delta\eta_{JJ}$, 5 TDA H_0, H_1, S_0, S_1, LB_1
  Particle Cloud: low-level features of the two fat jets



Muon Collider
Simulation
$\sqrt{s} = 3$ TeV
boosted
Normalized Events
E_T^{miss} [GeV]
Background
Signal ($\kappa_3 = 1$)
X_{HH}
M_{JJ} [GeV]
$P_{T_{JJ}}$ [GeV]



D_{HH} score separates the $H \to 4b$ signal from the bulk of the background



D_{κ_3} discriminates various κ_3

Statistical Analysis

- We define the likelihood as the product of Poisson probabilities of each bin i with signal $s_i(\kappa_3)$ and bkgd b_i , observed number n_i

$$L = \prod_i \frac{[\mu_i(\kappa_3)]^{n_i} \, e^{-\mu_i(\kappa_3)}}{n_i!} \qquad \mu_i(\kappa_3) = s_i(\kappa_3) + b_i$$

- Expected significance is given by $Z = \sqrt{-2\Delta \ln L}$
The resolved and boosted values for $-\Delta \ln L(\kappa_3)$ can be simply added to see combined effects.

- Case 1 : Observability of HH production versus SM background only

$$-\Delta \ln L(\kappa_3) = \ln L(s_i = 0) - \ln L(\kappa_3)$$

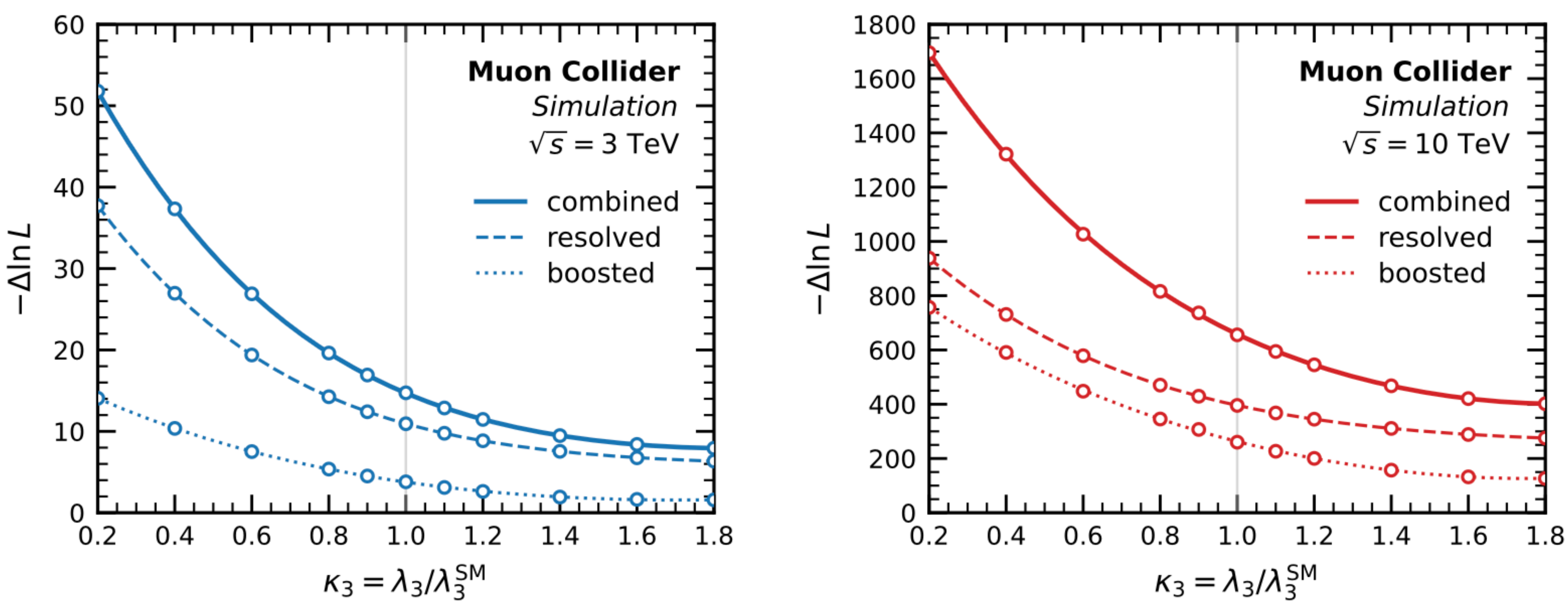
- Expected significance is given by $Z = \sqrt{-2\Delta \ln L}$

- For resolved region only, the SM value $\kappa_3 = 1$ achieves 4.7σ (28σ) at 3 (10) TeV.

- The boosted region achieves 2.8σ (23σ) at 3 (10) TeV.

- Combined, he SM value $\kappa_3 = 1$ achieves 5.4σ (36σ) at 3 (10) TeV.

- Minimum occurs at $\kappa_3 \approx 1.7$, which gives 4σ (28σ) at 3 (10) TeV.

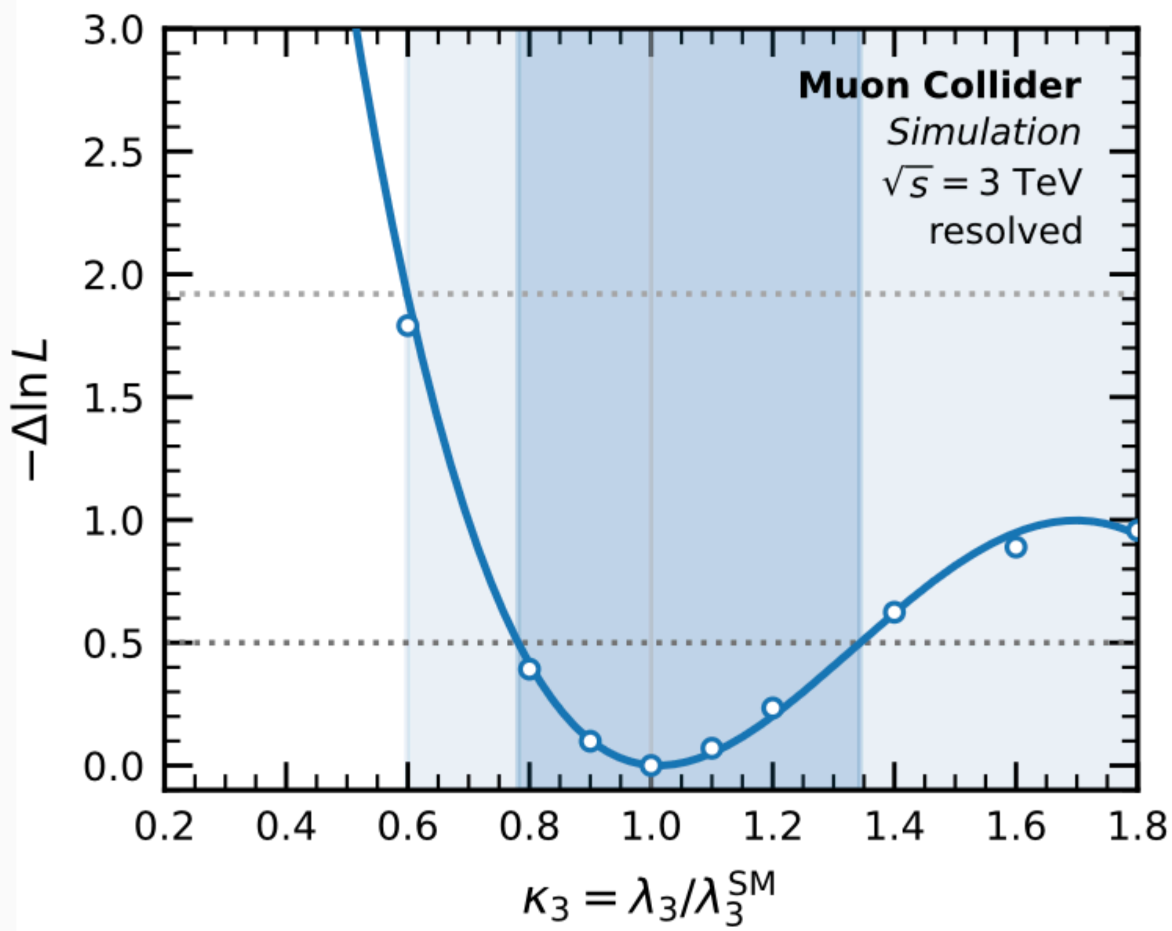


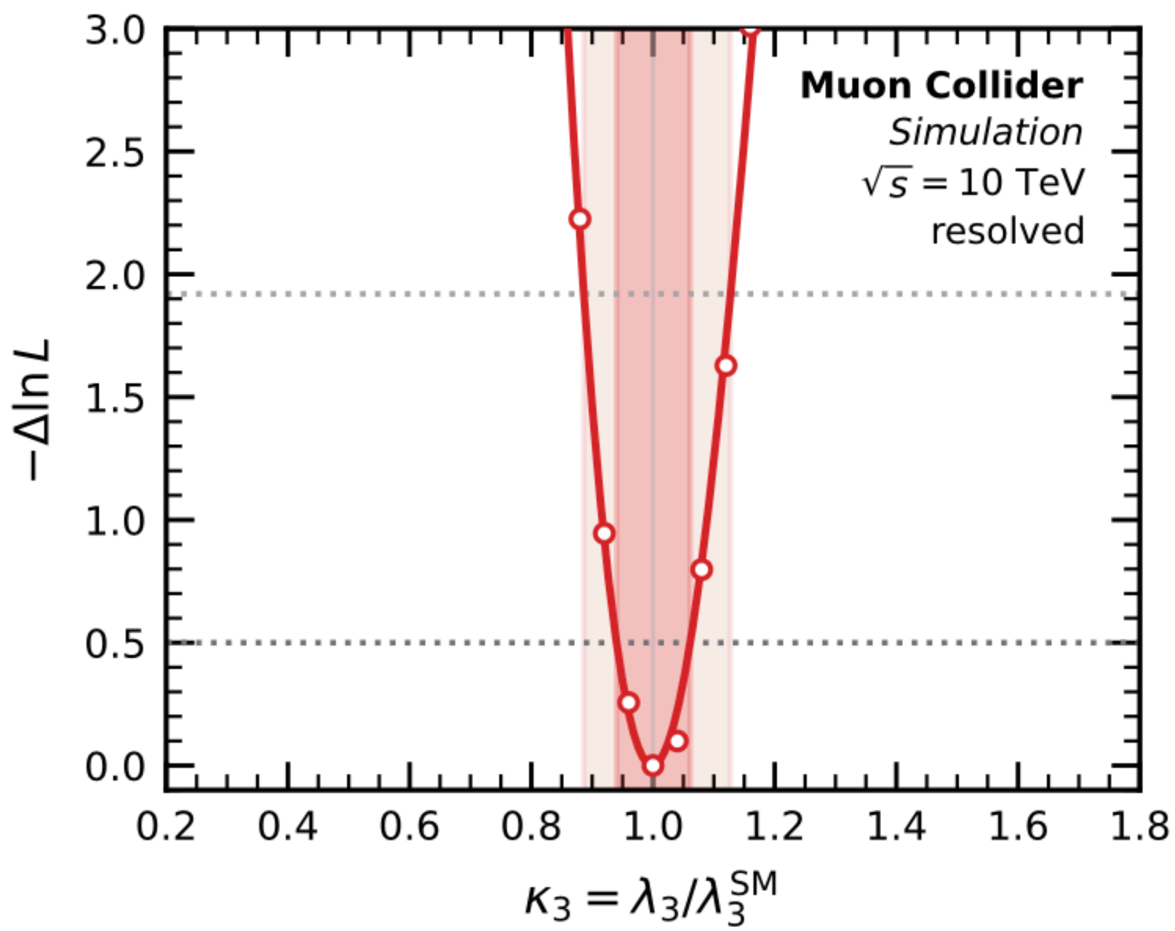
- Case 2: Discrimination of HH with $\kappa_3 \neq 1$ from $\kappa_3 = 1$

$$-\Delta \ln L(\kappa_3) = \ln L(\kappa_3 = 1) - \ln L(\kappa_3)$$

- Minimum at $\kappa_3 = 1$, the profile is fitted to a 4th order polynomial.

Resolved Region

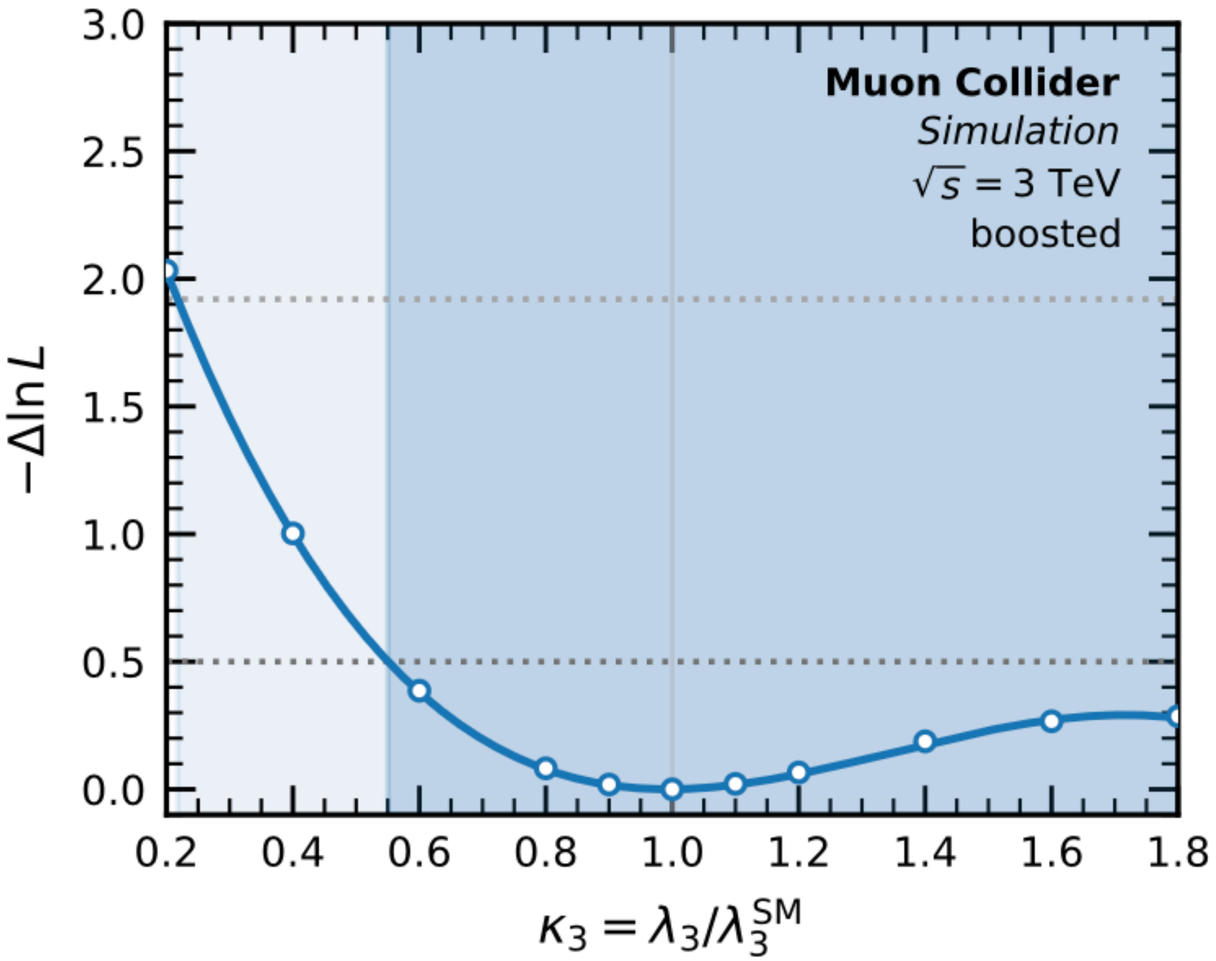


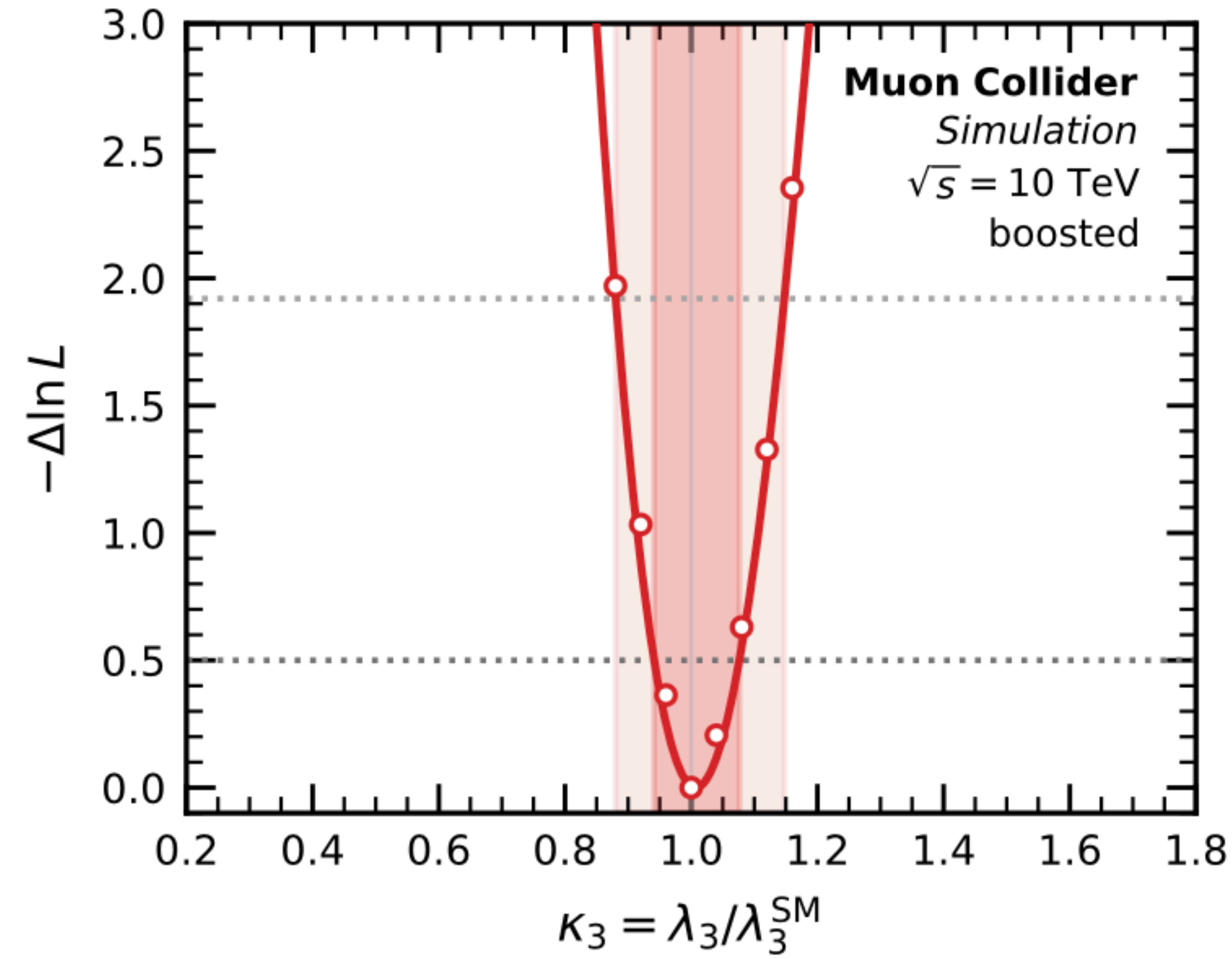


$0.78 < \kappa_3 < 1.34 (68\,\%\, CL)$, $0.60 < \kappa 3 < - (95\,\%\, CL)$ at 3 TeV

$0.94 < \kappa_3 < 1.06 (68\,\%\, CL)$, $0.89 < \kappa 3 < 1.13 (95\,\%\, CL)$ at 10 TeV

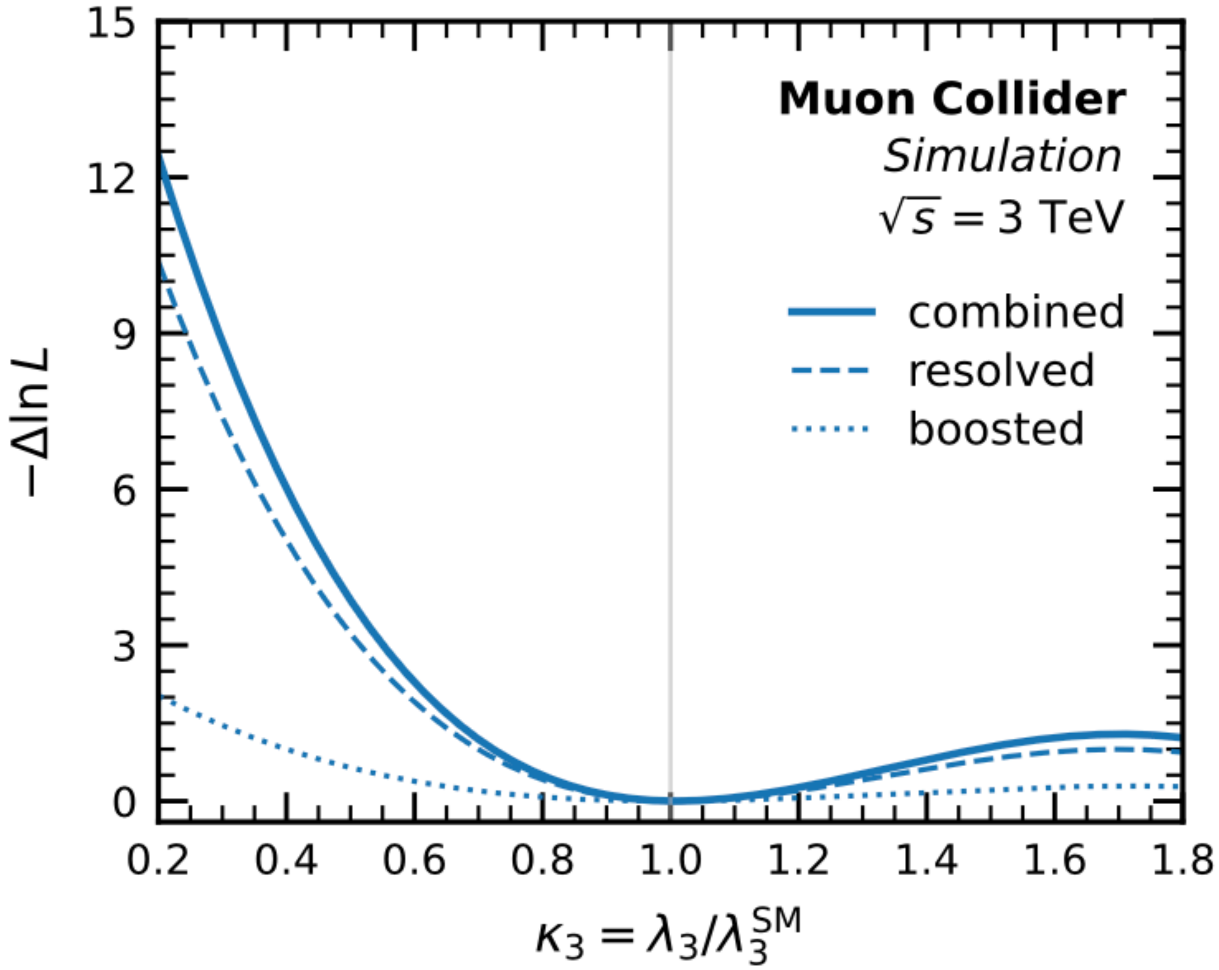
Boosted Region

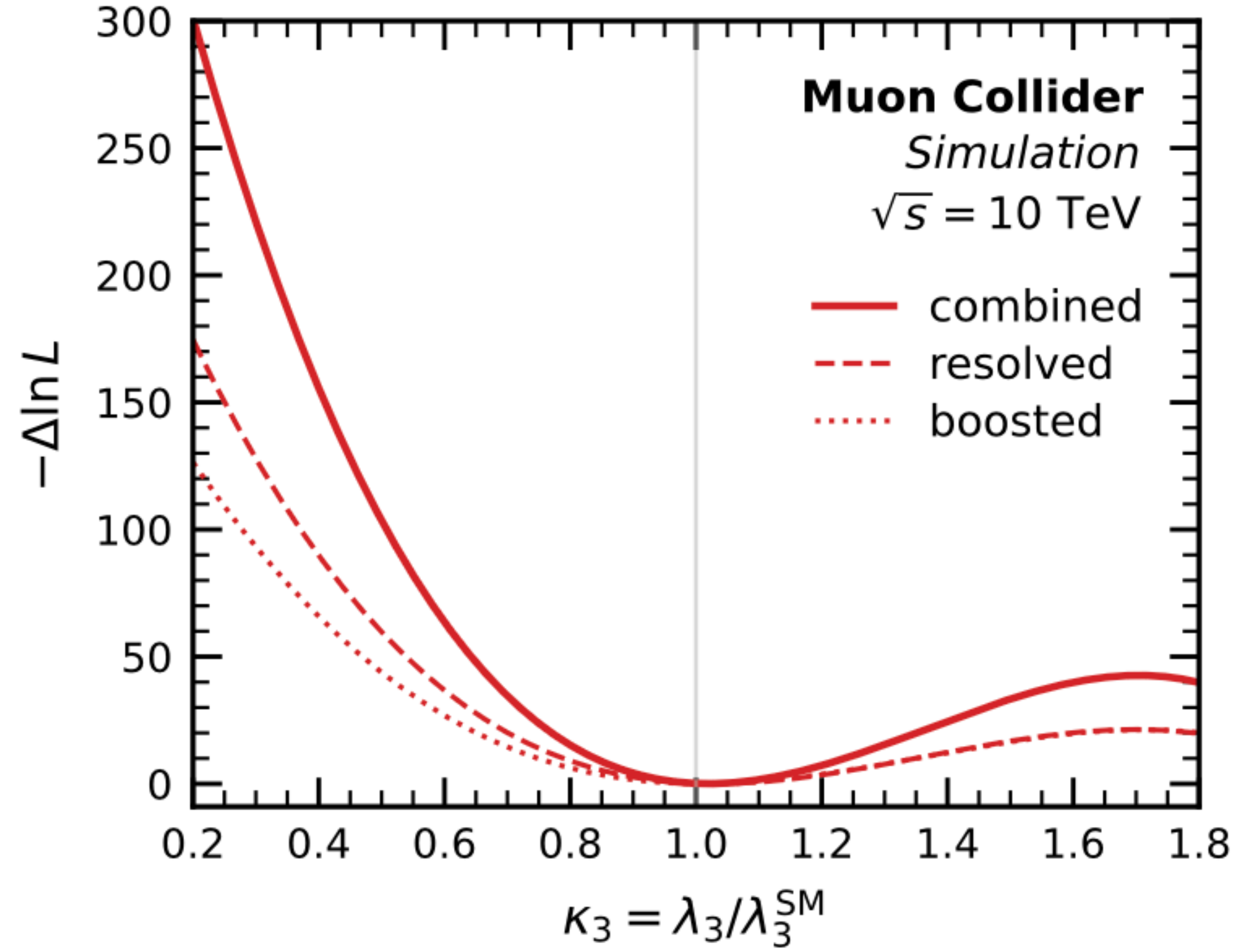




$0.55 < \kappa_3 < -(68\,\%\,CL)$, $\quad 0.22 < \kappa 3 < -(95\,\%\,CL)$ at 3 TeV

$0.94 < \kappa_3 < 1.08(68\,\%\,CL)$, $\quad 0.88 < \kappa 3 < 1.15(95\,\%\,CL)$ at 10 TeV





Combined

$0.80 < \kappa_3 < 1.2, (68\,\%\,CL)$, $0.63 < \kappa3 < - (95\,\%\,CL)$ at 3 TeV, 1 ab^{-1}

$0.96 < \kappa_3 < 1.05, (68\,\%\,CL)$, $0.92 < \kappa3 < 1.10\,(95\,\%\,CL)$ at 10 TeV, 10 ab^{-1}

% level

Conclusions

- Muon collider is a useful machine in probing Higgs self couplings.
- SPANet Higgs b-jet pairing improves the conventional X_{HH} method.
- Topological Data Analysis (TDA) variables capture multi-scale geometry of energy flow.
- Two neural networks, two classifiers, D_{HH}, D_{κ_3} trained on orthogonal targets.

1. Observability of HH production above the SM bkgd :: D_{HH}
2. Discrimination of HH with $\kappa_3 \neq 1$ from $\kappa_3 = 1$ :: D_{κ_3}

- Resolved $\oplus$ Boosted: independent kinematic regions combined
- Combined results at 68% CL: $0.8 < \kappa_3 < 1.3$ $(0.96 < \kappa_3 < 1.05)$ at $\sqrt{s}$ = 3 TeV, 1 ab^{-1} ($\sqrt{s}$ = 10 TeV, 10 ab^{-1})

Back-up Slides

Resolved Region

ML Pipeline Overview

SPANet motivated by arXiv: 2106.03898
ML1 / ML2 motivated by arXiv:2405.19314

① SPANet

Jet → Higgs pairing
(which 2 jets form H_1, H_2?)

② ML1

Signal vs Background
(binary classifier)

③ ML2

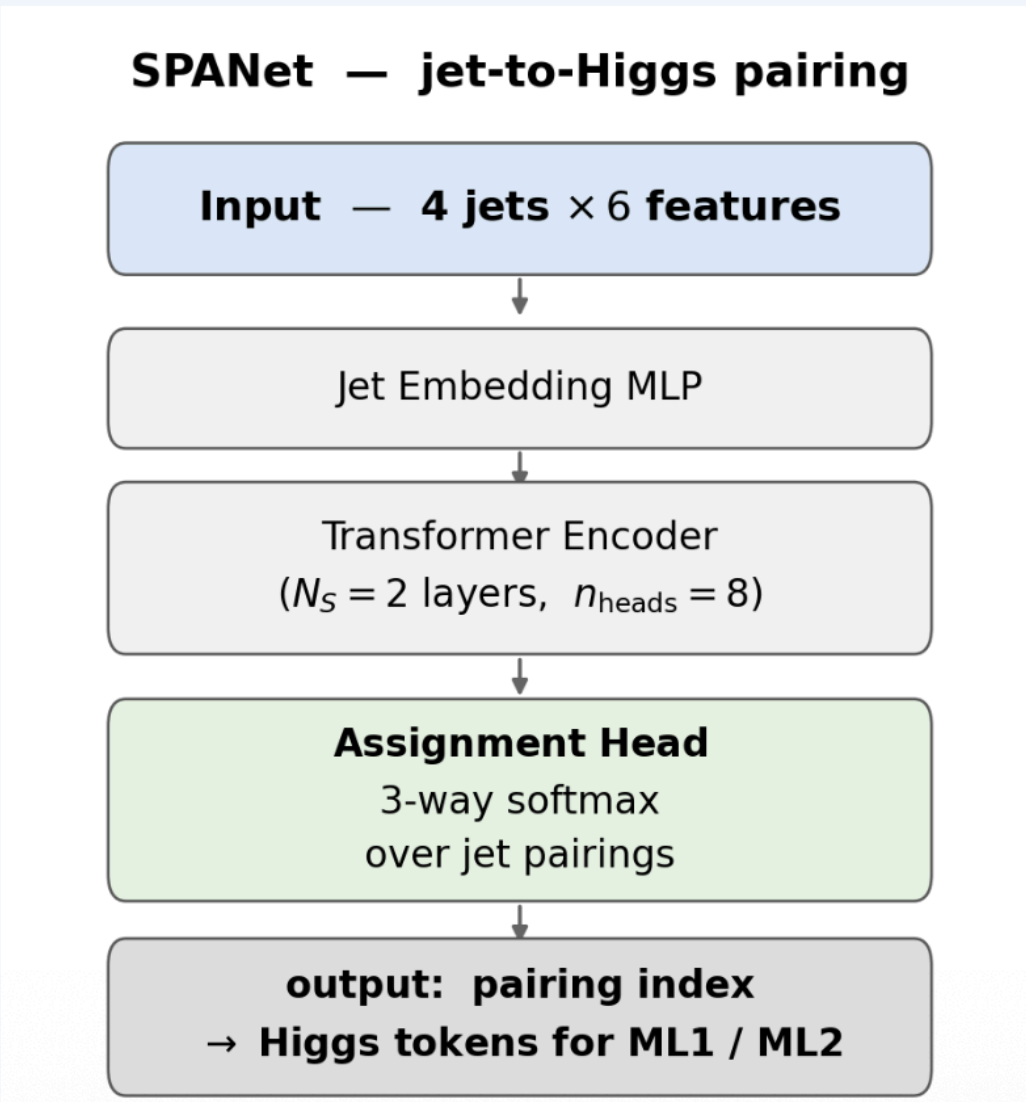
κ_3 shape discriminant
(κ=0.4 vs κ=1.6, signal only)

4 b-jets → SPANet pairing → (ML1 score, ML2 score) → 2D template → Asimov DLL → poly-4 fit → κ_3 sensitivity

- Backbone: PyTorch (SPANet) · TensorFlow / Keras (ML1, ML2)
- Hyperparameters tuned with Optuna (anti-overparameterisation cap on truth-valid subset)
- 70 / 15 / 15 stratified train / val / test split · seed 42 · test rescaled ×6.67 to recover full luminosity

Special thanks to Soojin for using some of her slides

SPANet — jet-to-Higgs pairing



Symmetry-Preserving Attention Network

- Job: **choose which 2 jets form H1, which 2 form H2** — 3 pairings to score
- Input: **4 jets × 6 features** (log pT, η, sin φ, cos φ, log mass, b-tag)
- Backbone: per-jet MLP + small Transformer (2 layers, 8 heads)
- Assignment head: enumerate 3 pairings → MLP scorer → softmax(3)
- Built-in symmetry: mean-pool inside a Higgs candidate;
  H1 ↔ H2 absorbed by the enumeration
- Classification accuracy of SPANet: **95% for both 3 TeV and 10 TeV**
- Output: argmax pairing index {0,1,2} → forwarded as **Higgs tokens to ML1/ML2**

Special thanks to Soojin for using some of her slides

Topological Data Analysis: variable definitions

$$H_0 = \sum_{k \in \mathrm{dgm}_0} \ell_k$$

Total angular scale over which separate energy "clusters" merge into one.

Large: angularly spread-out, isotropic events
small: compact, low-multiplicity events

$$H_1 = \sum_{k \in \mathrm{dgm}_1} \ell_k$$

Total weight of angular "loops" enclosed by particle rings.

Large: many particle rings enclosing empty regions
small: few loops

$$S_i = -\sum_k \frac{\ell_k}{H_i} \log_2 \frac{\ell_k}{H_i}$$

Shannon entropy of the normalised H0 / H1 lifetime spectra.

Large: many features of comparable persistence (multi-scale)
small: one feature dominates

$$LB_1 = \sum_{k \in \mathrm{dgm}_1} b_k \ell_k$$

Birth-scale–weighted sum of 1-D loops.

Large: big, late-forming loops
small: small, early-forming loops

1 · Build the filtration

Vietoris–Rips filtration on detector EFlow constituents in the (η, ϕ) plane, using the cylindrical angular distance

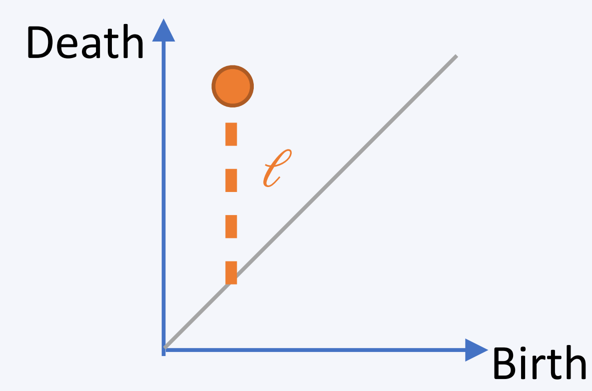
$$d_{ij} = \sqrt{(\Delta\eta_{ij})^2 + \left(2 \sin \frac{\Delta\phi_{ij}}{2}\right)^2}$$

2 · Track birth & death

As the scale grows, features appear and merge. For each homology class k with birth b_k and death d_k, define the persistence lifetime

$$\ell_k \equiv d_k - b_k$$

H0: connected components
H1: loops



Special thanks to Soojin for using some of her slides