Investigating the Higgs self-couplings through HHH production

Panagiotis Stylianou

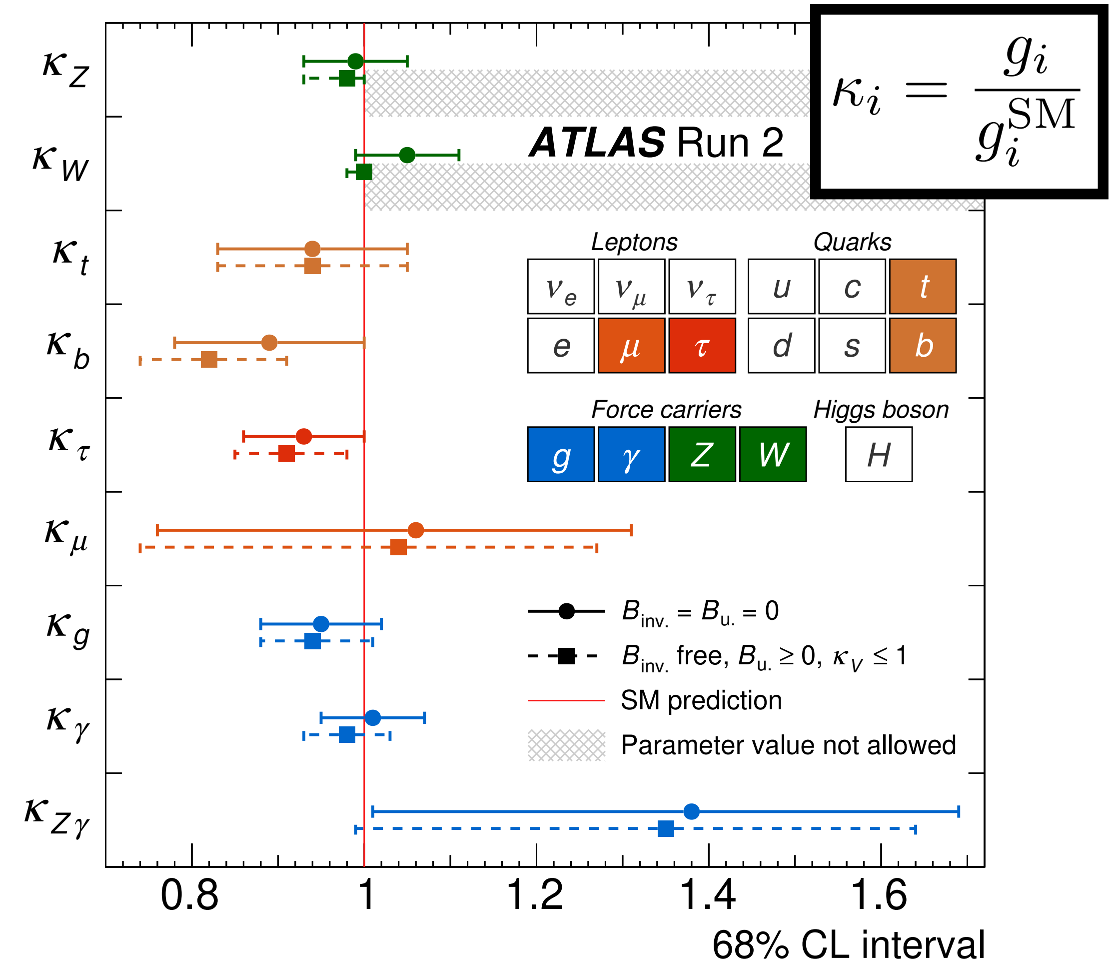
based on work in progress with Georg Weiglein

DESY.

3 August 2023

Introduction

- Higgs discovery at the LHC
- Tremendous efforts from experiments to pinpoint consistency with SM Higgs

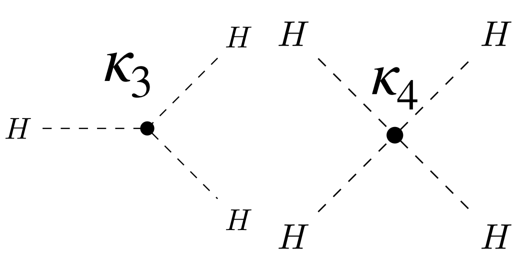


- **Most challenging:**

<u>SM Potential:</u>
$$V(\Phi) = \lambda(\Phi^\dagger\Phi)^2 - \mu^2\Phi^\dagger\Phi$$
$$\supset -\lambda v H^3 - \frac{\lambda}{4}H^4$$

<u>BSM theories</u> → more complicated shapes

- <u>**First step:**</u>
  - measure κ_3 → double-Higgs production

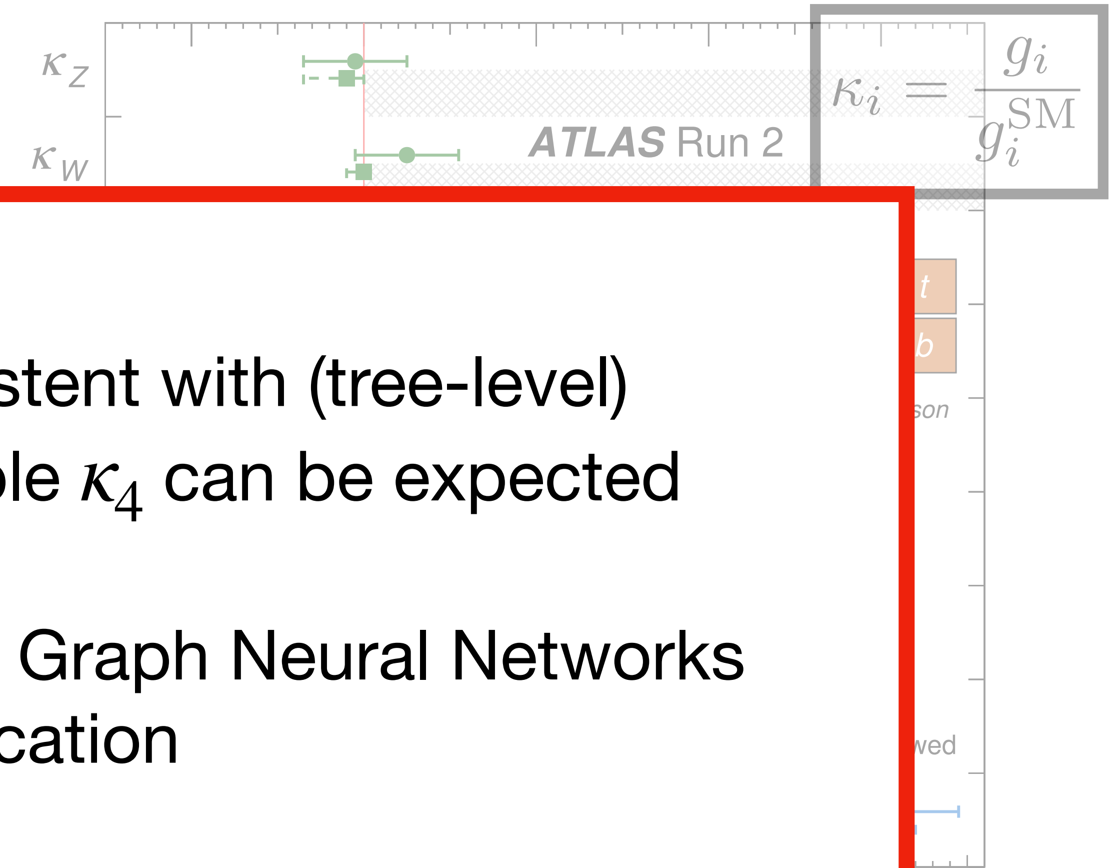


- <u>**But:**</u> κ_3 also appears in triple-Higgs production along with κ_4

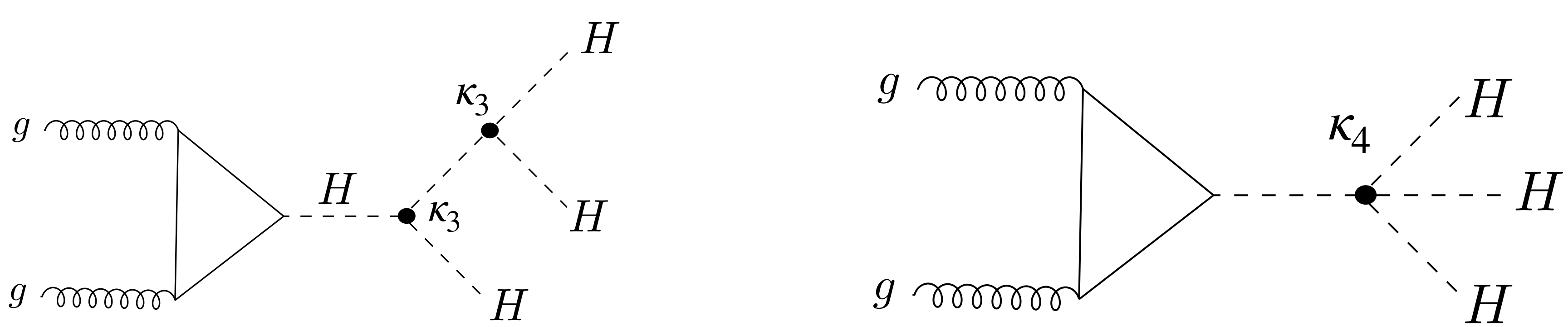
small cross sections

Content

- Higgs discovery at the LHC
- Tre
  pi
- M
  SM F
  BSM



- Values of (κ_3, κ_4) theoretically consistent with (tree-level) perturbative unitarity and why sizeable κ_4 can be expected
- HL-LHC exclusions for (κ_3, κ_4) using Graph Neural Networks (GNN) for signal-background classification
- Interpreting GNN results
- Comparison with future linear lepton collider concepts



Perturbative unitarity and Higgs couplings

- Process relevant for κ_3, κ_4 is $HH \to HH$ scattering (see also [Liu et al ` 18])
- Jacob-Wick expansion allows to extract partial waves

$$\beta(x,y,z) = x^2 + y^2 + z^2 - 2xy - 2yz - 2xz$$

Wigner functions

$$a^J_{fi} = \frac{\beta^{1/4}(s, m^2_{f_1}, m^2_{f_1})\beta^{1/4}(s, m^2_{i_1}, m^2_{i_1})}{32\pi s} \int_{-1}^{1} \mathrm{d}\cos\theta\, \mathcal{D}^J_{\mu_i \mu_f}\, \mathcal{M}(s, \cos\theta)$$

- Tree level unitarity:

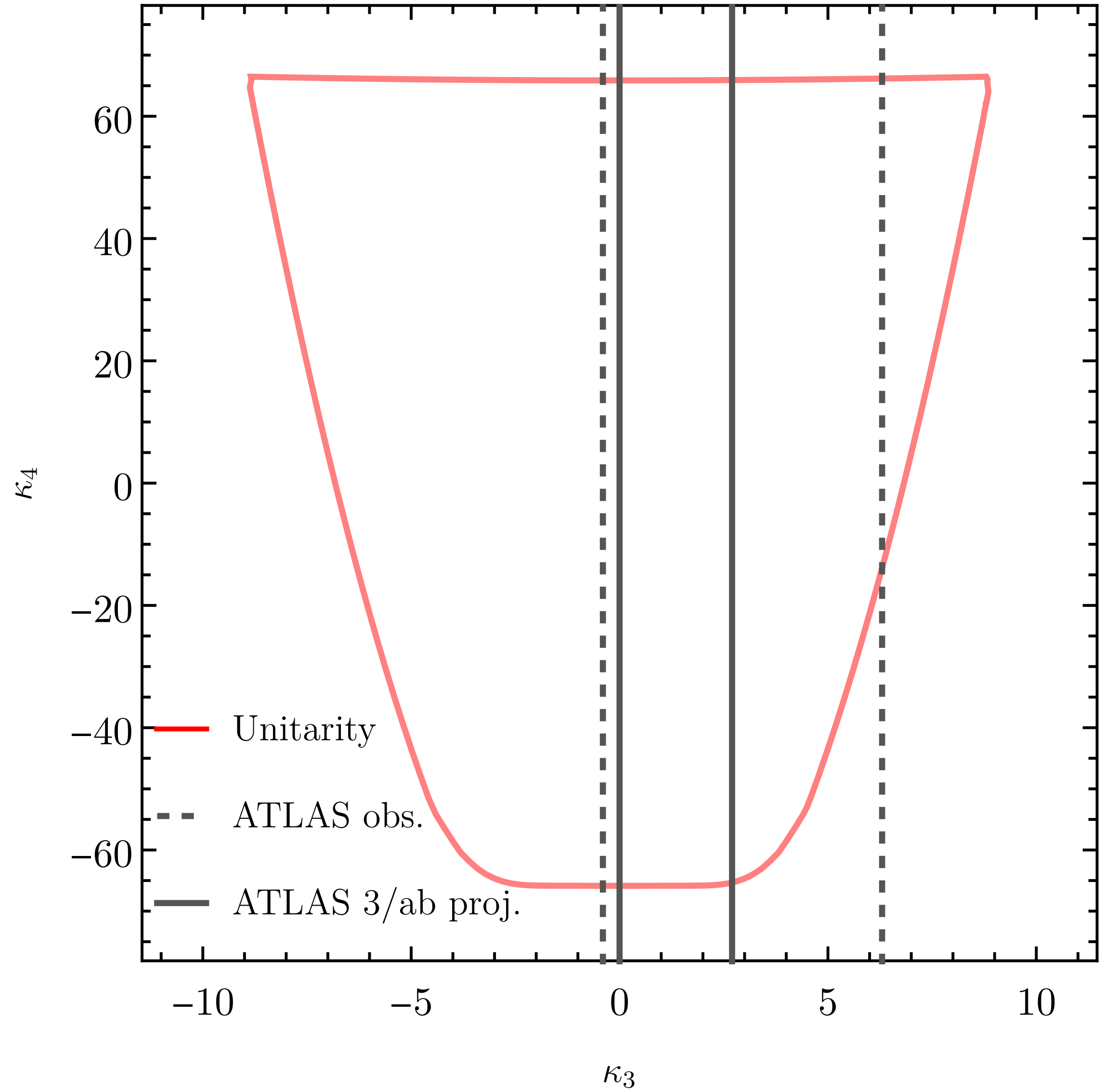
$$\mathrm{Im} a^0_{ii} \geq |a^0_{ii}|^2 \implies |\mathrm{Re} a^0_{ii}| \leq \frac{1}{2}$$

95 % CL

ATLAS current bounds: $[-0.4, 6.3]$

CMS & ATLAS HH projections: $[0.1, 2.3]$

[ATLAS 2211.01216]
[CERN Yellow Rep. 1902.00134]



Extension of SM potential by operators

Linear power expansion for higher order terms in Λ^{-1} orders:

[Boudjema, Chopin `96]
[Maltoni, Pagani, Zhao `18]

$$V_{\rm BSM} = \frac{C_6}{\Lambda^2}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^3 + \frac{C_8}{\Lambda^4}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^4 + \dots$$

Contributions to κ_3, κ_4:

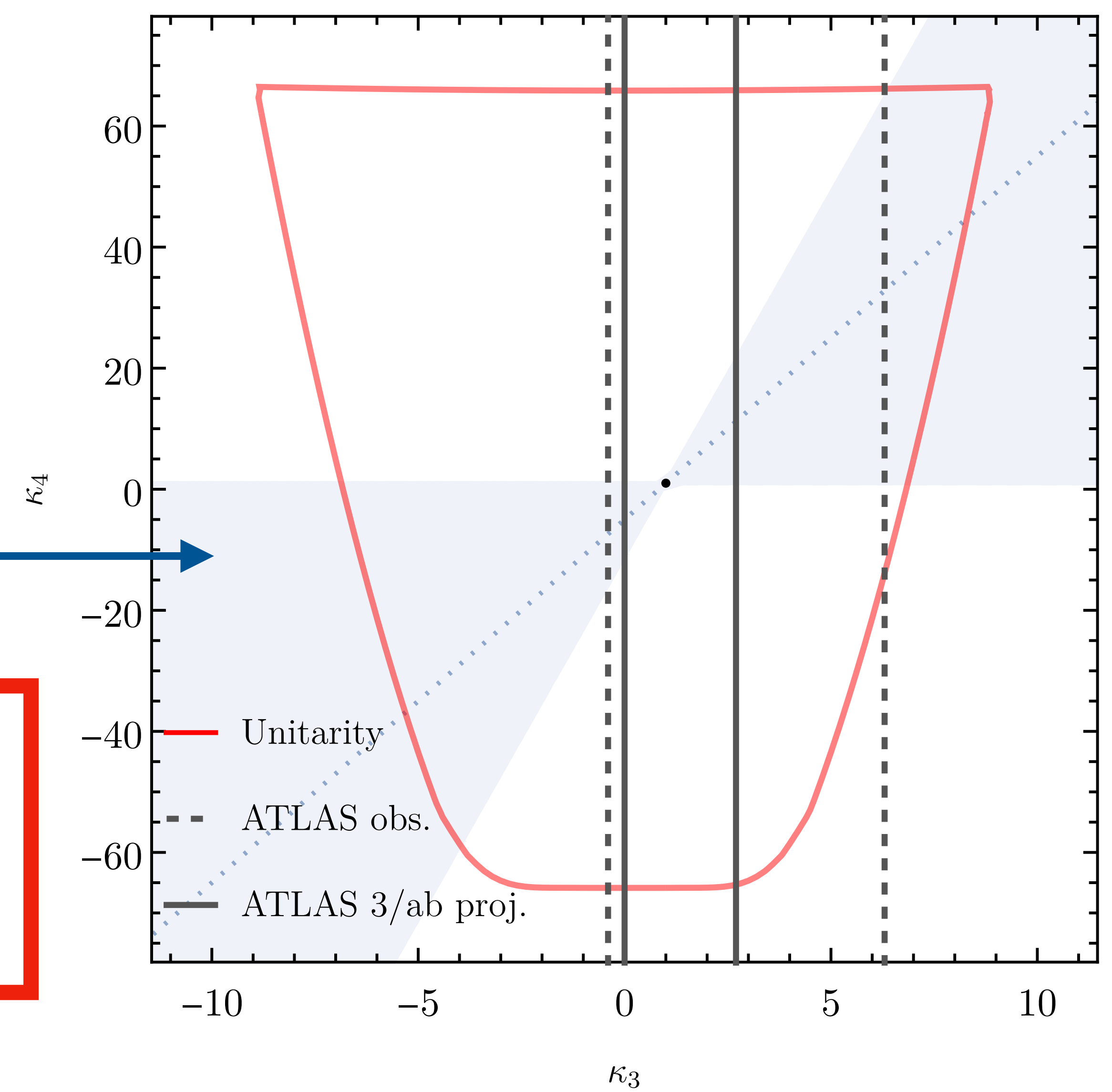
$$(\kappa_3 - 1) = \frac{C_6 v^2}{\lambda\Lambda^2}\,,$$

$$(\kappa_4 - 1) = \frac{6C_6 v^2}{\lambda\Lambda^2} + \frac{4C_8 v^4}{\lambda\Lambda^4}$$

vanishing dimension-8 →

$$\simeq 6(\kappa_3 - 1) + \mathcal{O}\left(\frac{1}{\Lambda^4}\right)$$

Shaded region: $\frac{4C_8 v^4}{\lambda\Lambda^4} < \frac{6C_6 v^2}{\lambda\Lambda^2}$ →



Electroweak Chiral Lagrangian (HEFT):
Higgs introduced as singlet and κ_3 and κ_4 are **free parameters** →probes **non-linearity**

Extension of SM potential by operators

Linear power expansion for higher order terms in Λ^{-1} orders:

[Boudjema, Chopin `96]
[Maltoni, Pagani, Zhao `18]

$$V_{\rm BSM} = \frac{C_6}{\Lambda^2}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^3 + \frac{C_8}{\Lambda^4}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^4 + \ldots$$

Contributio

$$(\kappa_3 - 1) = \frac{C_6 v^2}{\lambda\Lambda^2}\,,$$

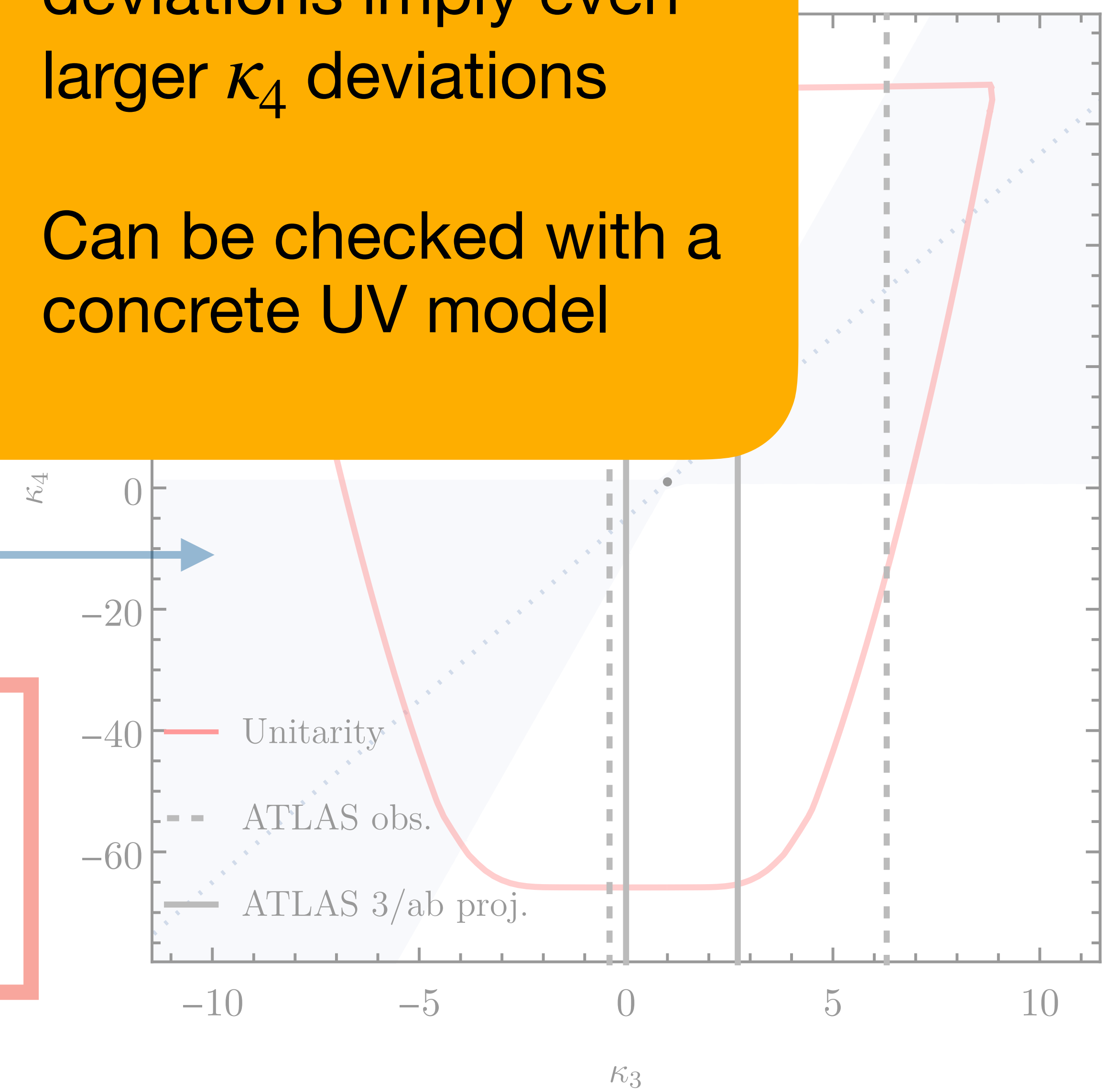
$$(\kappa_4 - 1) = \frac{6C_6 v^2}{\lambda\Lambda^2} + \frac{4C_8 v^4}{\lambda\Lambda^4}$$

$$\simeq 6(\kappa_3 - 1) + \mathcal{O}\left(\frac{1}{\Lambda^4}\right)$$

- indicates that κ_3 deviations imply even larger κ_4 deviations
- Can be checked with a concrete UV model

vanishing dimension-8

Shaded region: $\frac{4C_8 v^4}{\lambda\Lambda^4} < \frac{6C_6 v^2}{\lambda\Lambda^2}$



Electroweak Chiral Lagrangian (HEFT):

Higgs introduced as singlet and κ_3 and κ_4 are **free parameters** → probes **non-linearity**

Model example: 2HDM

- Consider the 2HDM as an example

- Two-Higgs Doublet Model (2HDM) → a second doublet: $\Phi_1 = \begin{pmatrix} \phi_1^+ \\ \frac{1}{\sqrt{2}}(v_1 + \rho_1 + i\eta_1) \end{pmatrix}, \; \Phi_2 = \begin{pmatrix} \phi_2^+ \\ \frac{1}{\sqrt{2}}(v_2 + \rho_2 + i\eta_2) \end{pmatrix}$

$$V_{2\mathrm{HDM}} = m_{11}^2(\Phi_1^\dagger\Phi_1) + m_{22}^2(\Phi_2^\dagger\Phi_2) - m_{12}^2(\Phi_1^\dagger\Phi_2 + \Phi_2^\dagger\Phi_1) + \frac{\lambda_1}{2}(\Phi_1^\dagger\Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger\Phi_2)^2$$
$$+\lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) + \lambda_4(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) + \frac{\lambda_5}{2}\left((\Phi_1^\dagger\Phi_2)^2 + (\Phi_2^\dagger\Phi_1)^2\right)$$

- Free parameters: $m_H, m_{H'}, \; m_A, \; m_{H^\pm}, \; m_{12}^2, \; v, \; \cos(\beta-\alpha), \; \tan\beta$

Scalar Particle content:
Neutral scalars: $H', H(m_H = 125$ GeV)
Neutral pseudoscalars: A
Charged scalars: $H^\pm$

Alignment limit → couplings of light Higgs same as SM
$\cos(\beta - \alpha) = 0$

- Sizeable contributions to κ_3 at loop-level

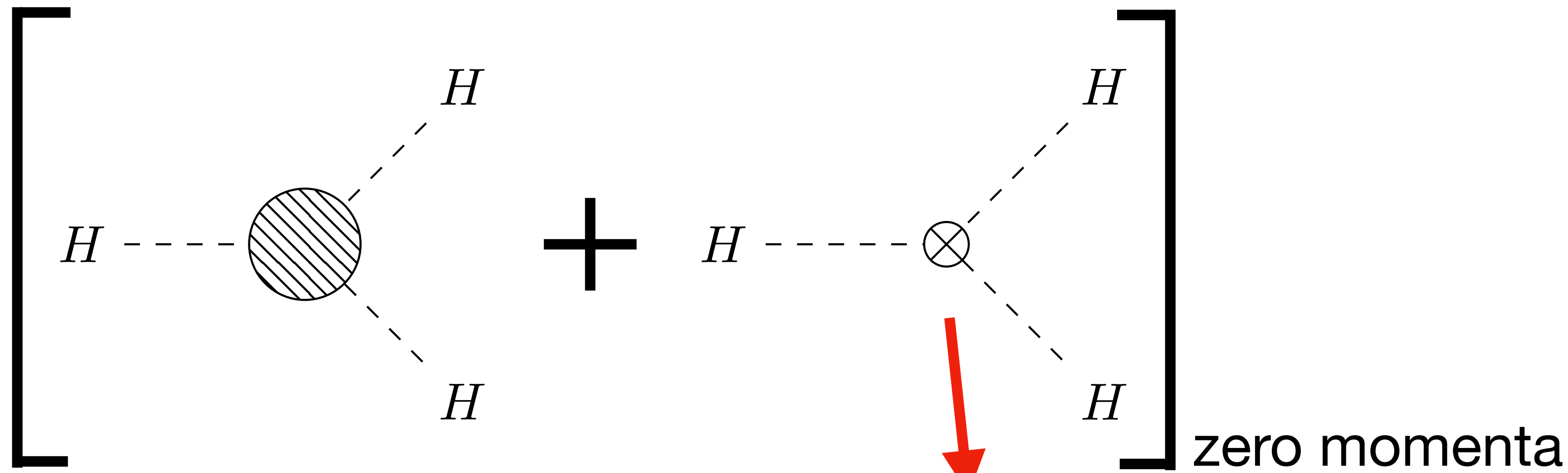
reproduce 1-loop results

2-loop calculations: [Braathen, Kanemura `20], [Bahl, Braathen, Weiglein `22]
anyH3: [Bahl, Braathen, Gabelmann, Weiglein `23]

Model example: 2HDM - calculation

- 1-loop calculation for κ_3, κ_4 with FeynArts, FormCalc, LoopTools in alignment limit

$$\hat{\Gamma}^{(1)}_{3H} =$$

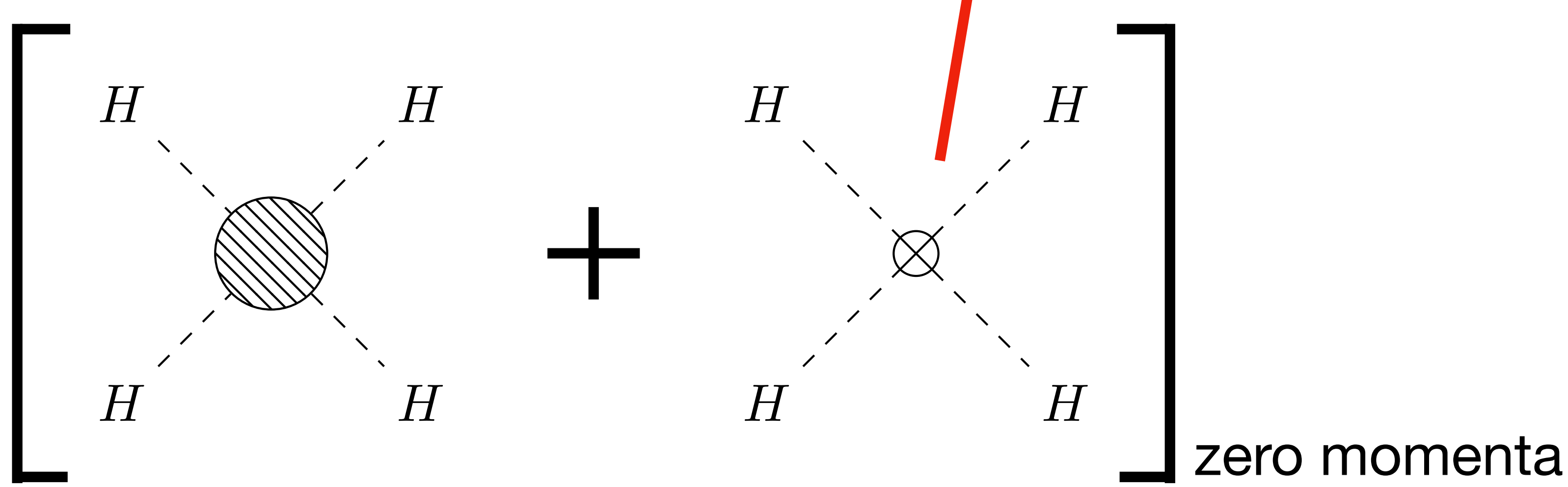


- 2HDM renormalisation constants calculated with on-shell conditions
- divergence cancellation checked analytically

$$\Gamma^{\mathrm{CT}}_{3H} = -\frac{3}{v}\left[\delta M_H^2 + \frac{\delta T_H}{v} + M_H^2\left(\delta Z_e + \frac{3}{2}\delta Z_H - \frac{\delta M_W^2}{2M_W^2} - \frac{\delta s_W}{s_W}\right)\right]$$

$$\Gamma^{\mathrm{CT}}_{4H} = -\frac{3}{v^2}\left[\delta M_H^2 + \frac{\delta T_H}{v} + 2M_H^2\left(\delta Z_e + \delta Z_H - \frac{\delta M_W^2}{2M_W^2} - \frac{\delta s_W}{s_W}\right)\right]$$

$$\hat{\Gamma}^{(1)}_{4H} =$$

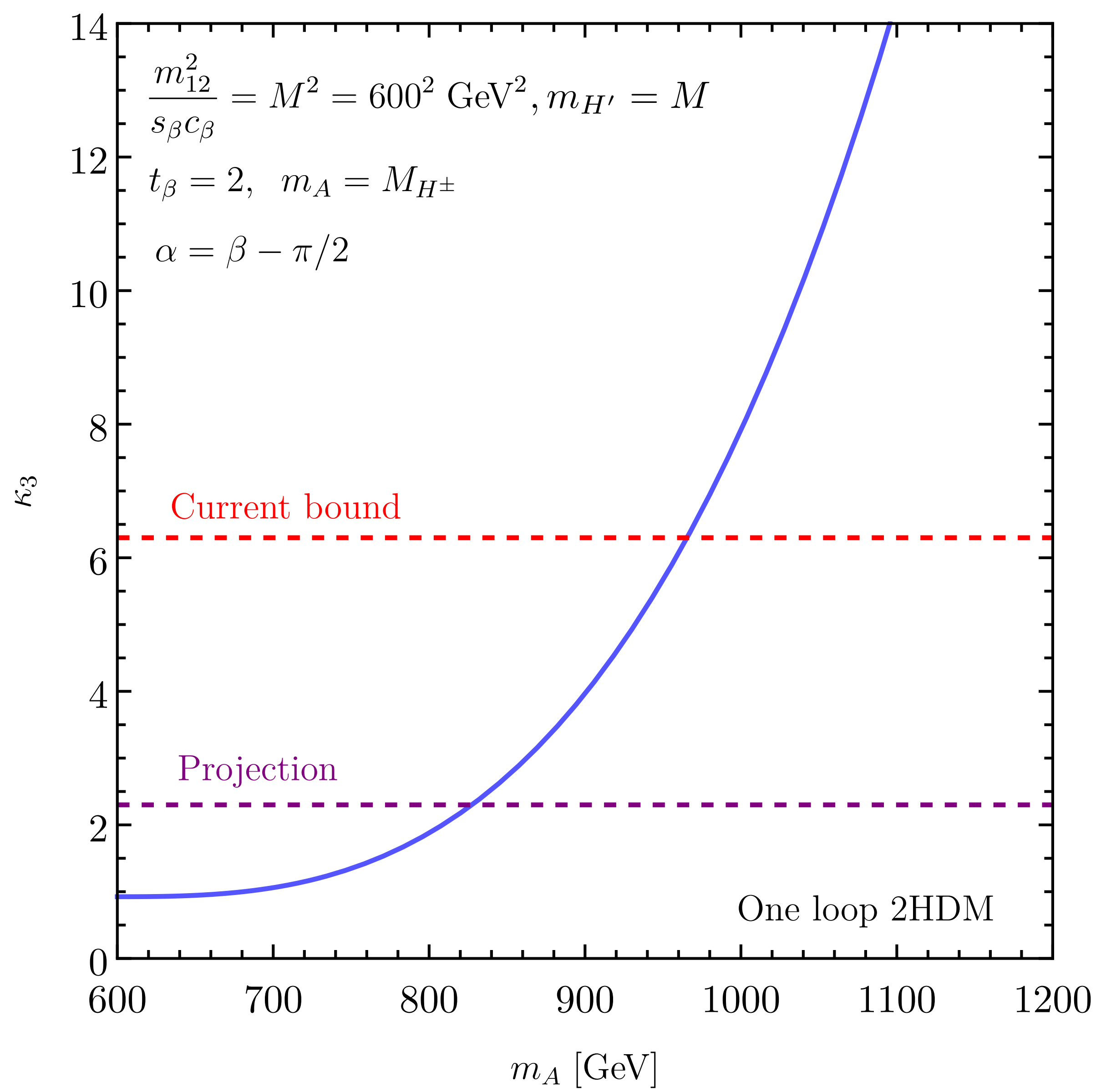


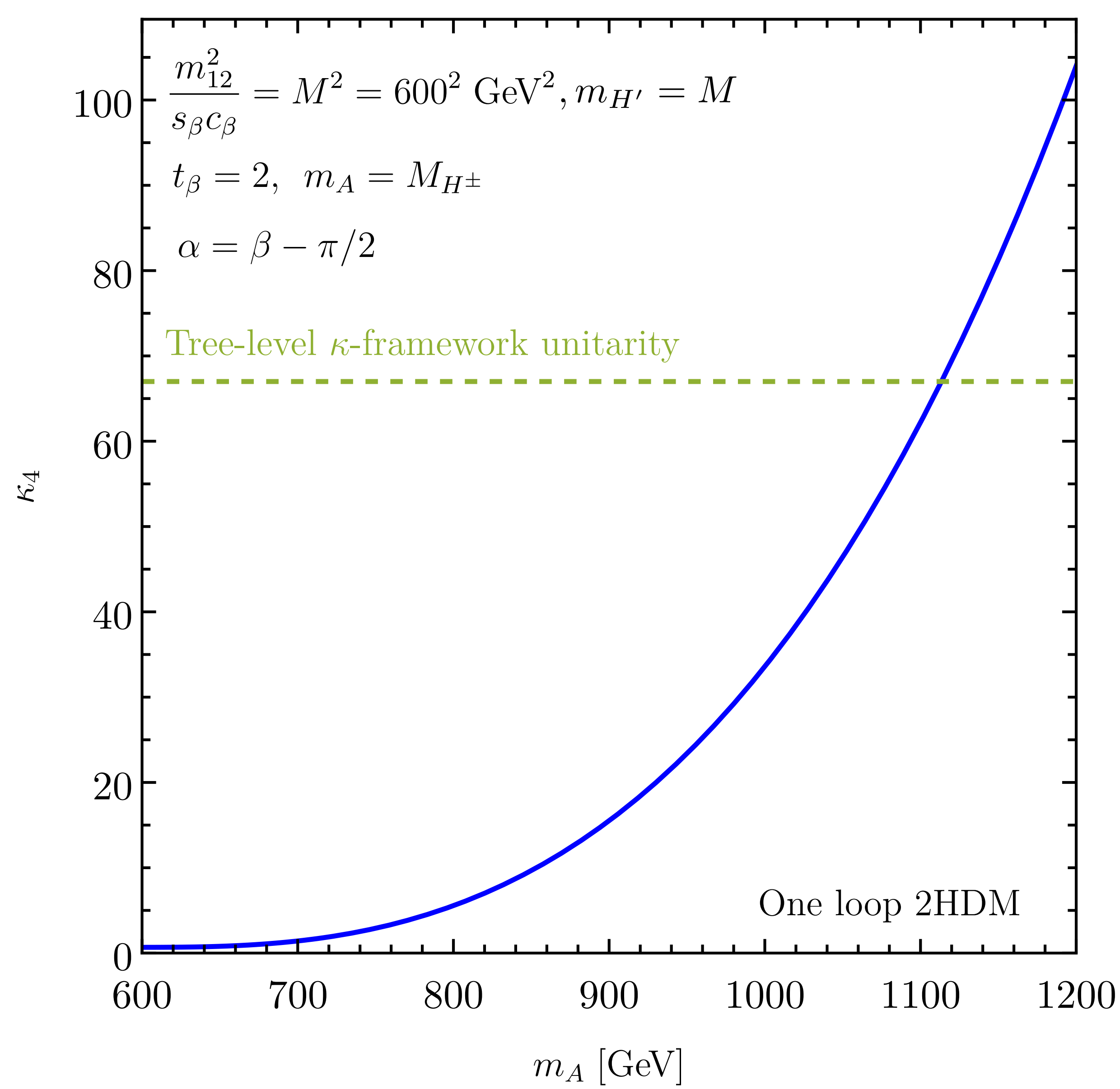
Model example: 2HDM - trilinear vs quartic

- Benchmark Point of [Bahl, Braathen, Weiglein `22] → cross-check κ_3 result (also with anyH3)
- Expectedly deviations in κ_3 induce sizeable deviations in κ_4

$$\kappa_i = \frac{\Gamma_i^{(0)} + \hat{\Gamma}_i^{(1)}}{\Gamma_{\mathrm{SM},i}^{(0)}}$$

$$i \in \{3H, 4H\}$$





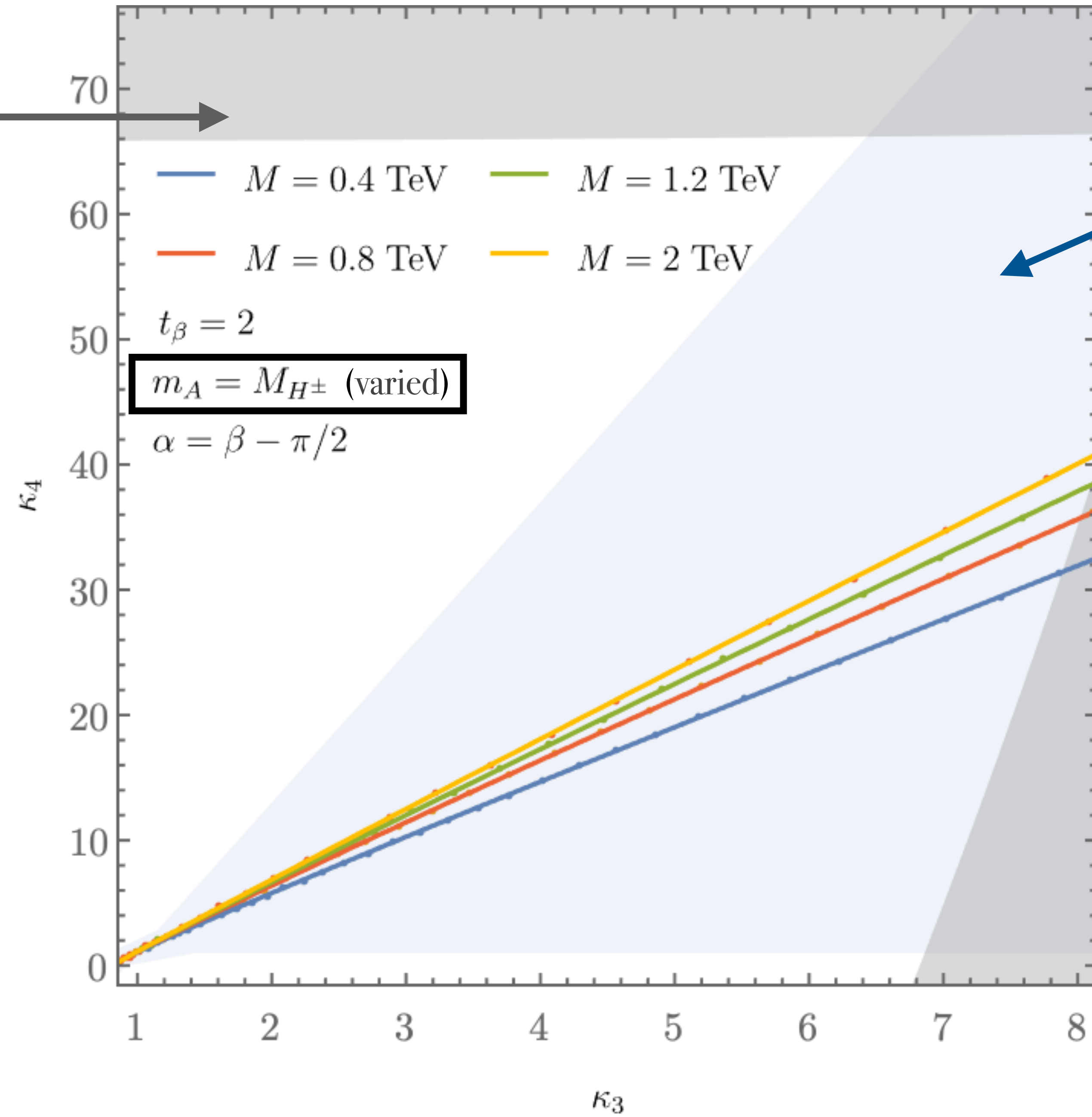
Model example: 2HDM - trilinear vs quartic

- Benchmark Point of [Bahl, Braathen, Weiglein `22] → cross-check κ_3 result (also with anyH3)
- Expectedly deviations in κ_3 induce sizeable deviations in κ_4

$$\kappa_i = \frac{\Gamma_i^{(0)} + \hat{\Gamma}_i^{(1)}}{\Gamma_{\mathrm{SM},i}^{(0)}}$$

$$i \in \{3H, 4H\}$$

tree-level κ-framework perturbative unitarity



well-behaved perturbative expansion from before

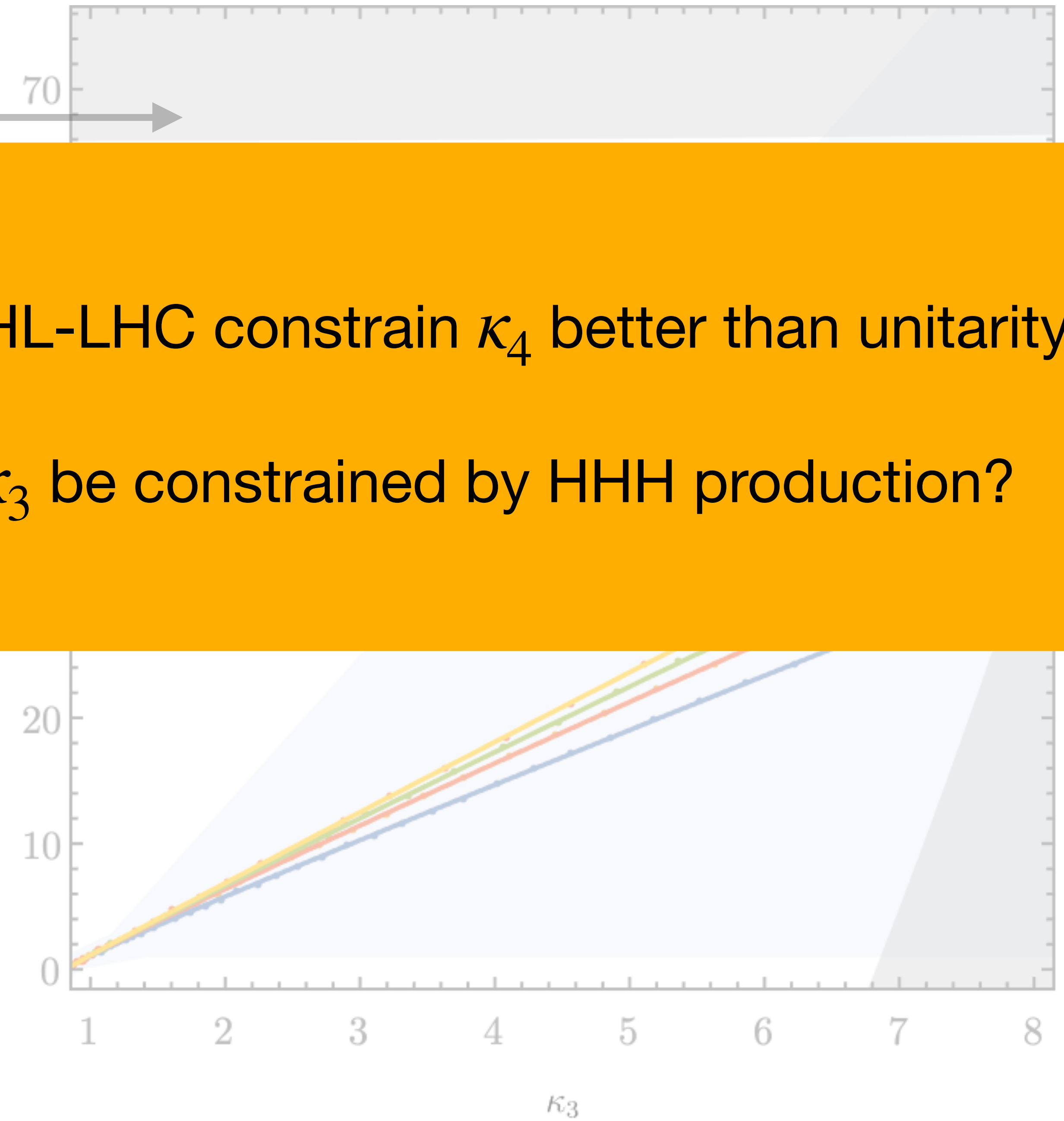
Model example: 2HDM - trilinear vs quartic

- Benchmark Point of [Bahl, Braathen, Weiglein `22] → cross-check κ_3 result (also with anyH3)
- Expectedly deviations in κ_3 induce sizeable deviations in κ_4

$$\kappa_i = \frac{\Gamma_i^{(0)} + \hat{\Gamma}_i^{(1)}}{\Gamma_{\mathrm{SM},i}^{(0)}}$$

$$i \in \{3H, 4H\}$$

tree-level κ-framework
perturbative u...



well-behaved
...bative
...sion
...efore

- Can HL-LHC constrain κ_4 better than unitarity?
- Can κ_3 be constrained by HHH production?

Relevant channels at LHC

- Small rates at LHC

Need dominant production & decays

- gluon fusion
- BRs:

$$\mathrm{BR}(H \to b\bar{b}) = 0.584$$
$$\mathrm{BR}(H \to \tau^+\tau^-) = 6.627 \times 10^{-2}$$
$$\mathrm{BR}(H \to \gamma\gamma) = 2.26 \times 10^{-3}$$

$2b4\tau$ and $4b2\gamma$ produce relatively few events even for large $\kappa_3 \gtrsim 4.5, \ \kappa_4 \gtrsim 30$

- Focus on $6b$ and $4b2\tau$ final states with 5 and 3 tagged b-quarks, respectively

Backgrounds:

$6b$: dominant QCD contributions (see also [Papaefstathiou, Robens, Xolocotzi ` 21])

$4b2\tau$: $W^+W^-b\bar{b}b\bar{b}, \ Zb\bar{b}b\bar{b},$
$t\bar{t}(H \to \tau\tau), \ t\bar{t}(H \to b\bar{b}),$
$t\bar{t}(Z \to \tau\tau), \ t\bar{t}(Z \to b\bar{b}), \ t\bar{t}t\bar{t}$

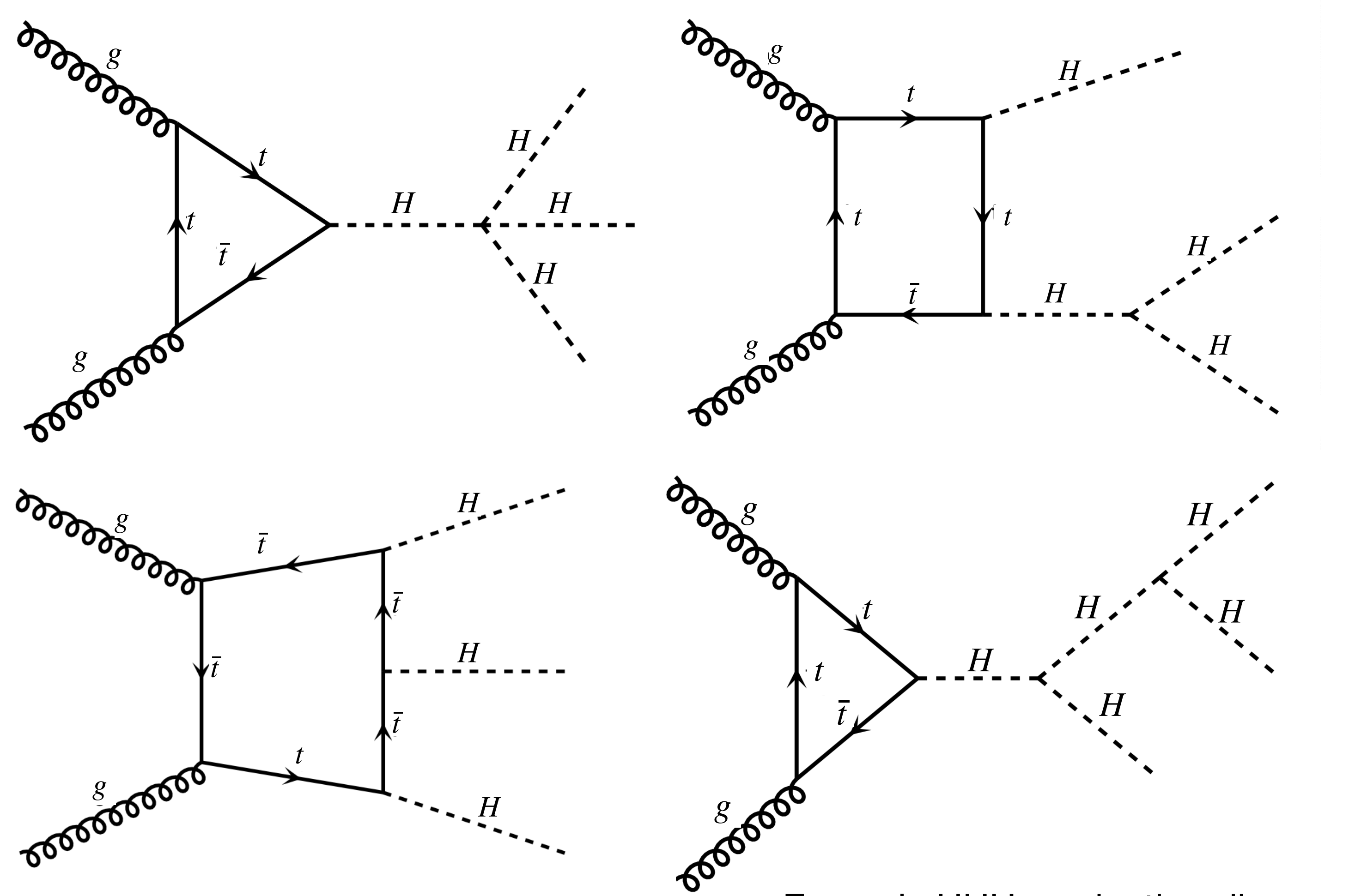
Event generation and pre-selection

- Events generated with MadGraph5_aMC@NLO
- Higgs states decayed with MadSpin

(conservative) background K-factor of 2

signal K-factor of 1.7

[Florian, Fabre, Mazzitelli`20]



Example HHH production diagrams

Pre-selection cuts:

Invariant mass of final states: $\gtrsim 350$ GeV

At least one pair of tagged states with $m_{ij} \in [110,140]$

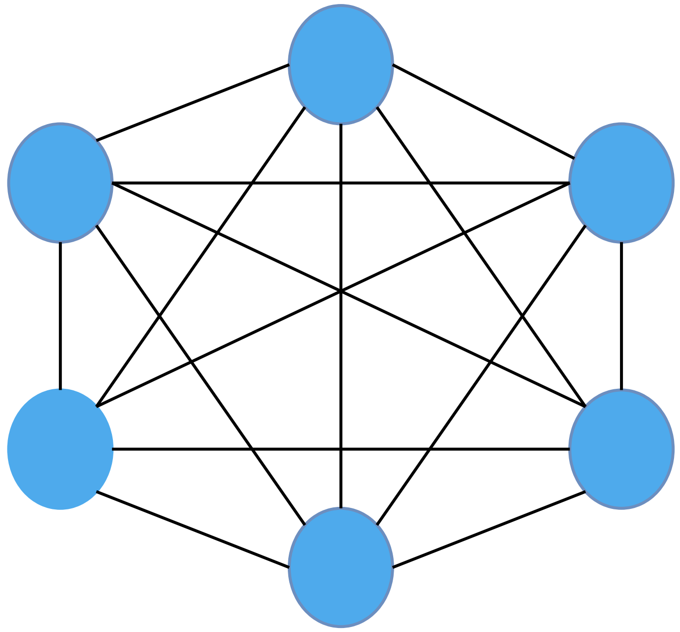
$p_T(b) > 30$ GeV $\quad p_T(\tau) > 10$ GeV

$|\eta(\tau)| < 2.5 \quad |\eta(b)| < 2.5$

Graph Embedding

1.

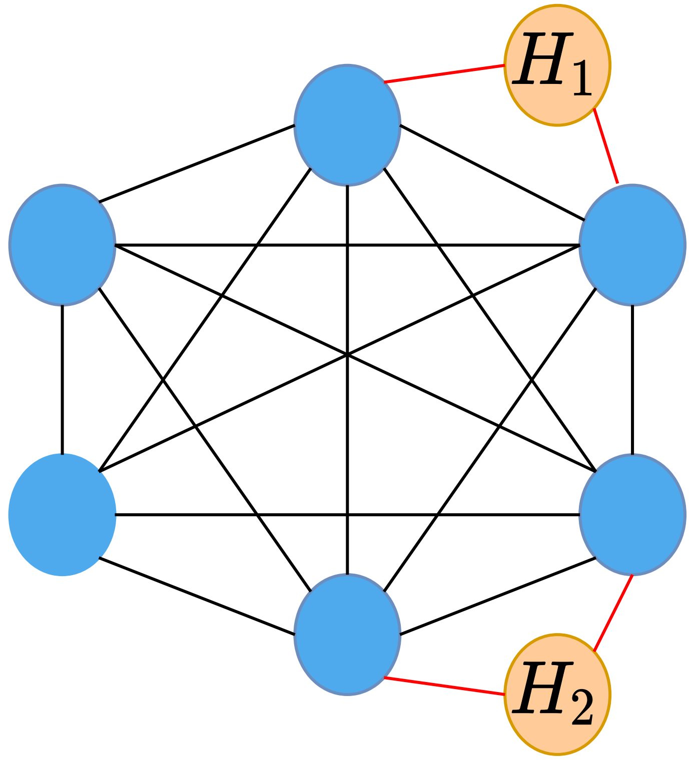
- Fully-connected nodes for b and τ final states
- **<u>Input features</u>**: $[p_T, \eta, \phi, E, m,$ PDGID$]$
- Additional node for Missing Transverse Momentum (MTM) in showered & reconstructed events



FC: Fully-Connected

2.

- Consider combinations of b-quarks and τ with reconstructed four-momentum $(p_i + p_j)$
- If $m_{ij} \in [100,150]$ (GeV) add extra node H_i



RN: Reconstructed Nodes

Edge Convolution

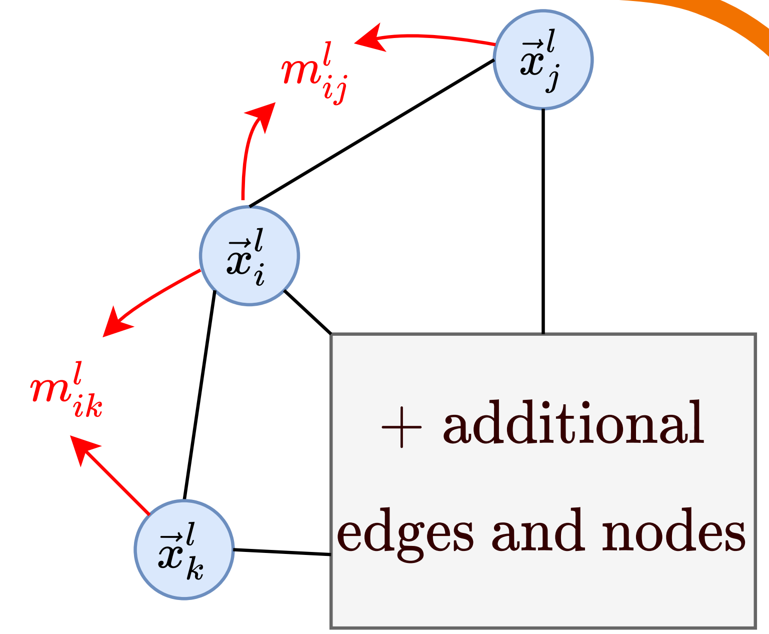
Input features: $\vec{x}_i^{(0)}$ → update iteratively with **Edge Convolution** operation:

Edge Convolution operation

'Message' calculation:

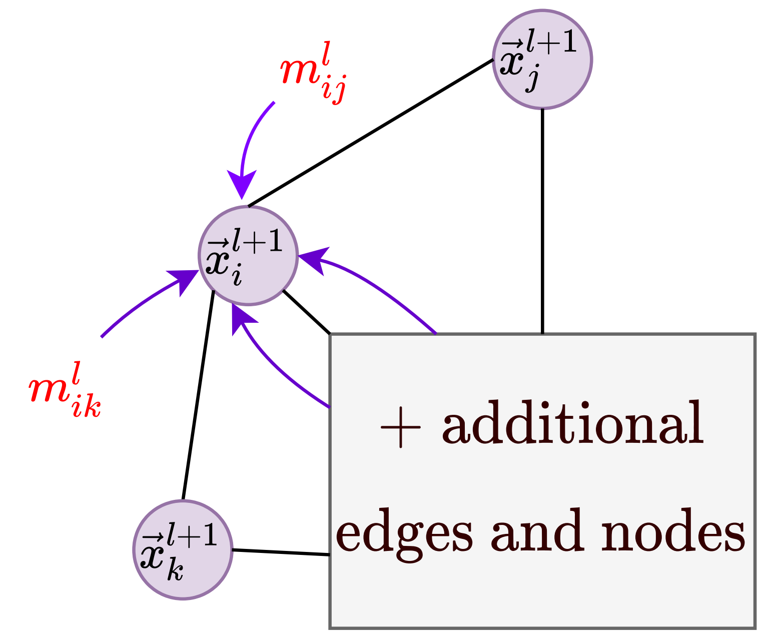
$$m_{ij}^{(l)} = \textsc{ReLU}\left(\Theta(\vec{x}_j^{(l)} - \vec{x}_i^{(l)}) + \Phi(\vec{x}_i^{(l)})\right)$$

linear layers



Aggregation: update node features

$$\vec{x}_i^{(l+1)} = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} m_{ij}^{(l)}$$



GNN embedding efficiencies

Dataset with signal & background graphs → **GNN** (2 EdgeConv layers + linear) → $\left\{ P(\mathbf{S}), P(\mathbf{B}_1), \ldots, P(\mathbf{B}_N) \right\}$

- GNN trained on $(\kappa_3, \kappa_4) = (1,1)$ sample

Complex final states

↓

- Compare embeddings at parton level with only parts of background
- Check signal efficiencies for 99% background rejection

* for parton-level considerations H_i nodes have only $[p_T, E]$ as input features, with additional noise introduced

	FC efficiency	RN efficiency
$5b$	0.21	0.47
$3b2\tau$	0.27	0.98

‘Reconstructed Nodes’ case outperforms Fully-Connected

↓

focus only on this embedding for full analysis

GNN embedding efficiencies

Dataset with signal & background graphs → **GNN** (2 EdgeConv layers + linear) → $\left\{ P(\mathbf{S}), P(\mathbf{B}_1), \ldots, P(\mathbf{B}_N) \right\}$

- GNN trained on $(\kappa_3, \kappa_4) = (1,1)$ sample

- **Assumption:** Same GNN efficiency for other values of (κ_3, κ_4)
- Flat optimistic $80\,\%$ b-tagging and τ-tagging efficiency
- **Significance:**

$$Z = \sqrt{2\left((S+B)\ln\left(1 + \frac{S}{B}\right) - S \right)}$$

from [Cowan, Cranmer, Gross, Vitells `10]

- Check signal efficiencies for 99% background rejection

* for parton-level considerations H_i nodes have only $[p_T, E]$ as input features, with additional noise introduced

outperforms Fully Connected

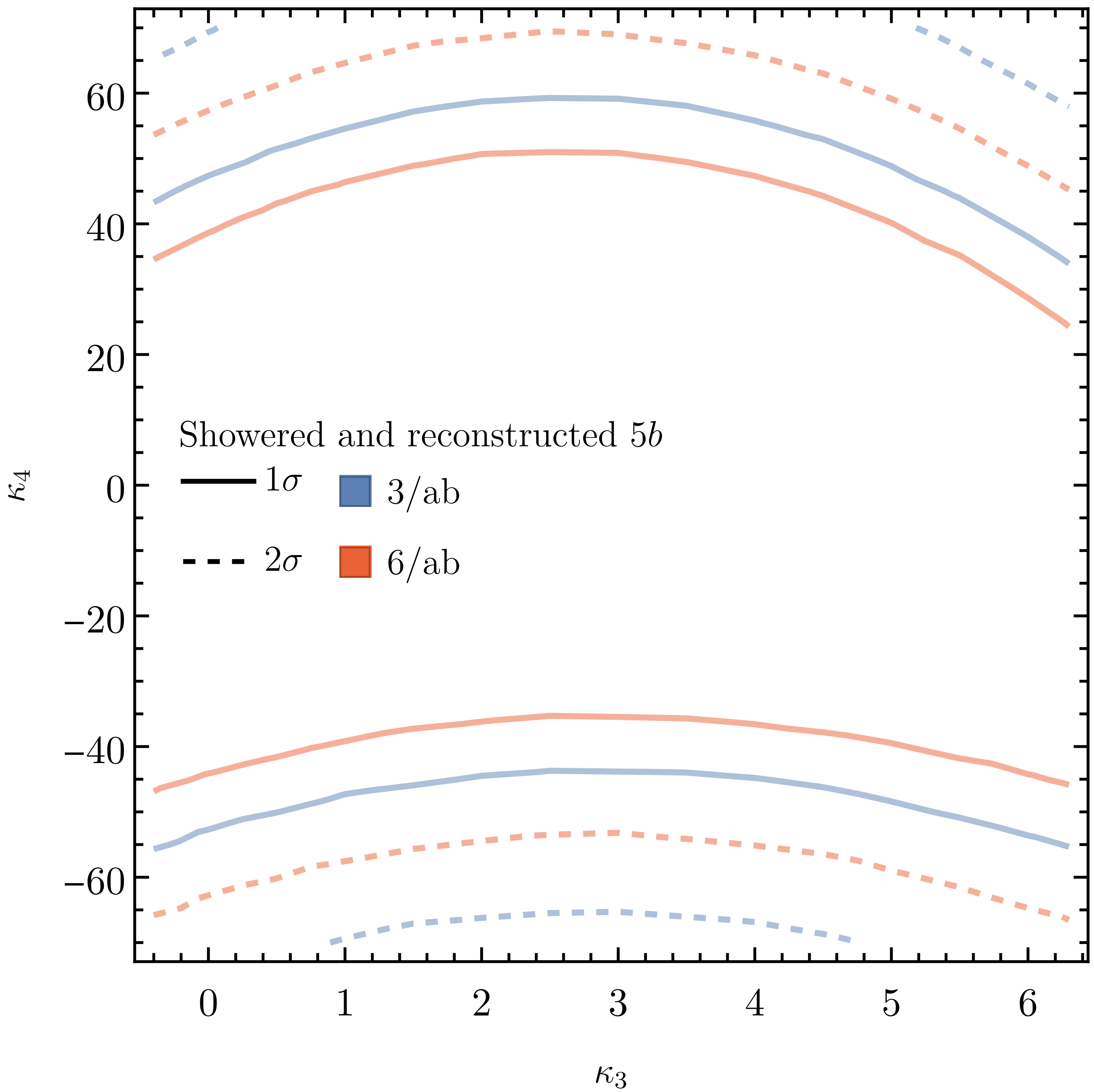
↓

focus only on this embedding for full analysis

Showered and reconstructed results $5b$

- Showering and reconstruction of events: Pythia, FastJet, Rivet
- HL-LHC luminosity of $3/$ab and ATLAS-CMS combined luminosity of $6/$ab

- ‘Reconstructed Nodes’ embedding
- Re-train GNN and re-identify threshold corresponding to background rejection of ~ $99\ \%$



Showered and reconstructed results $3b2\tau$

- $3b2\tau$ more complicated due to multiple backgrounds → multi-class classification
- Train on backgrounds: $W^+W^-b\bar{b}b\bar{b},\ Zb\bar{b}b\bar{b},\ t\bar{t}(H\to\tau^+\tau^-)$

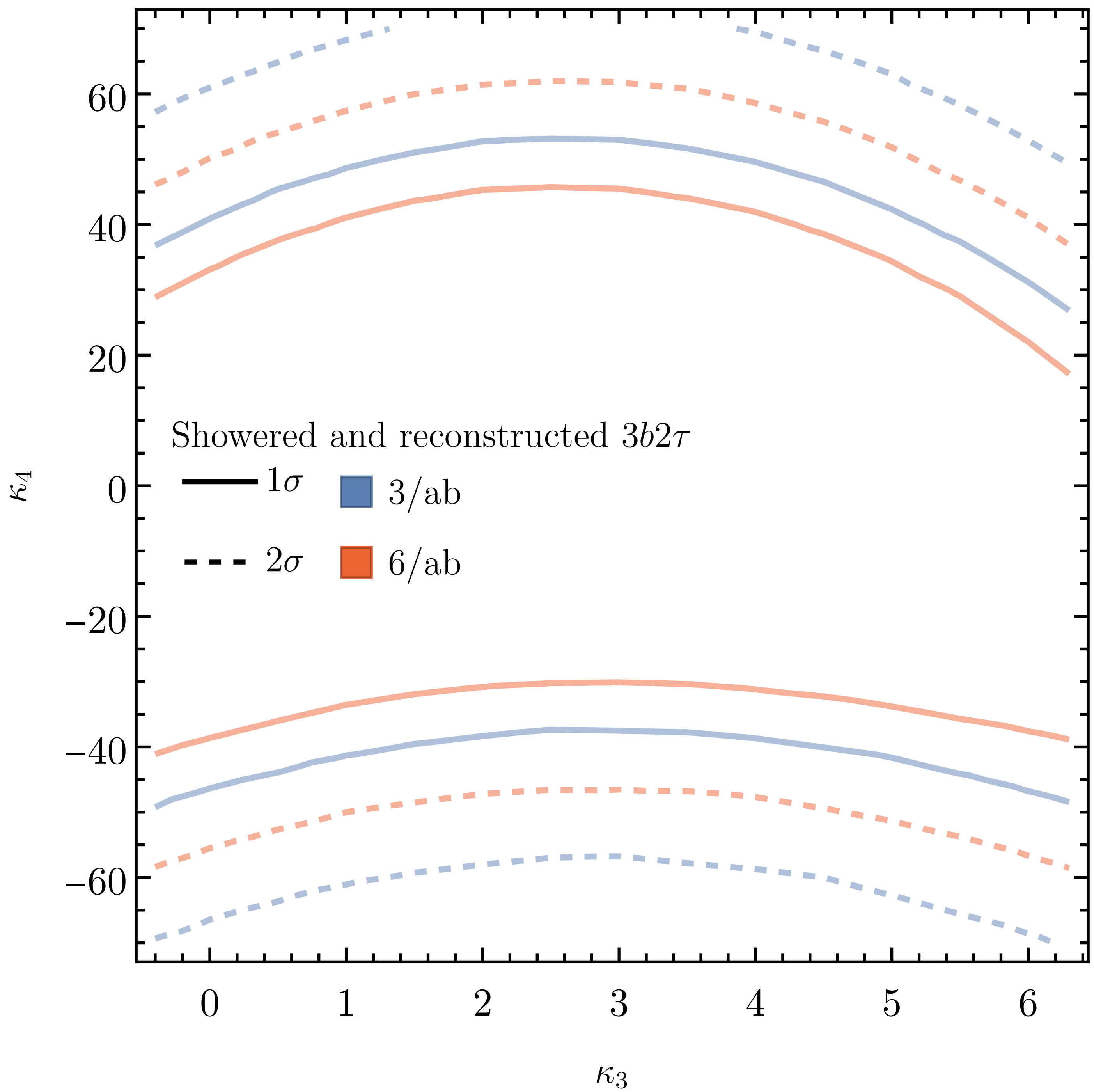
- Impose cuts on NN scores to reduce backgrounds:

$$P[W^+W^-b\bar{b}b\bar{b}] < 0.03,\quad P[Zb\bar{b}b\bar{b}] < 0.1,\quad P[t\bar{t}(H\to b\bar{b})] < 0.3$$

	σ(gen.)(fb)	σ(sel.)(fb)	σ(NN)(fb)
$tt(H\to\tau\tau)$	3.8	0.17	0.010
$WWbbbb$	31	4.6	7.0×10^{-3}
$tt(H\to bb)$	3.5	0.89	3.6×10^{-3}
$Zbbbb$	4.3	0.45	3.3×10^{-4}
$tt(Z\to bb)$	0.77	0.15	2.9×10^{-4}
$tttt$	0.38	0.091	2.1×10^{-4}
$tt(Z\to\tau\tau)$	4.7	0.080	1.1×10^{-4}

Showered and reconstructed results $3b2\tau$

- $3b2\tau$ more complicated due to multiple backgrounds → multi-class classification
- Train on backgrounds: $W^+W^-b\bar{b}b\bar{b},\ Zb\bar{b}b\bar{b},\ t\bar{t}(H \to \tau^+\tau^-)$

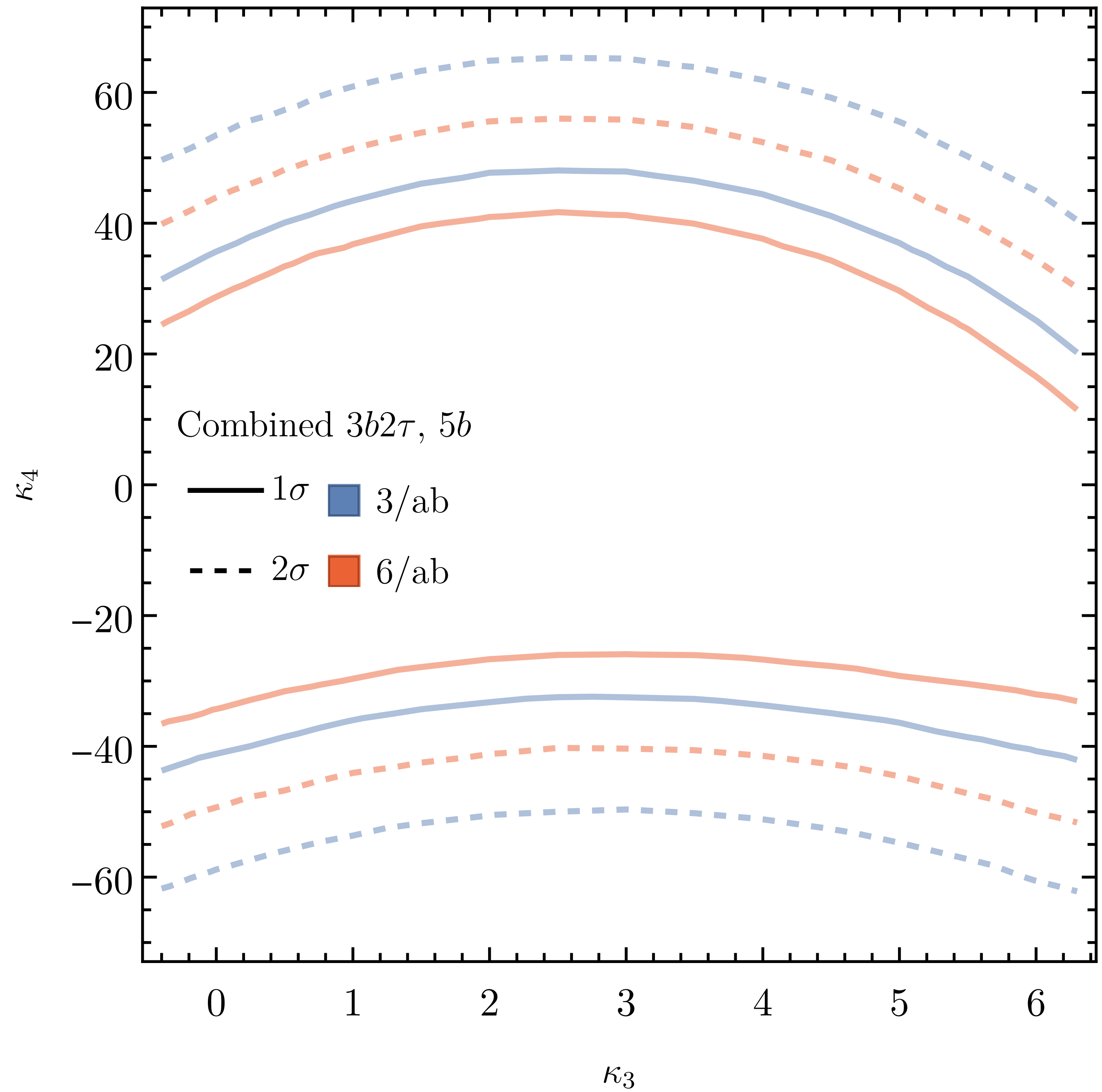


Combined Results

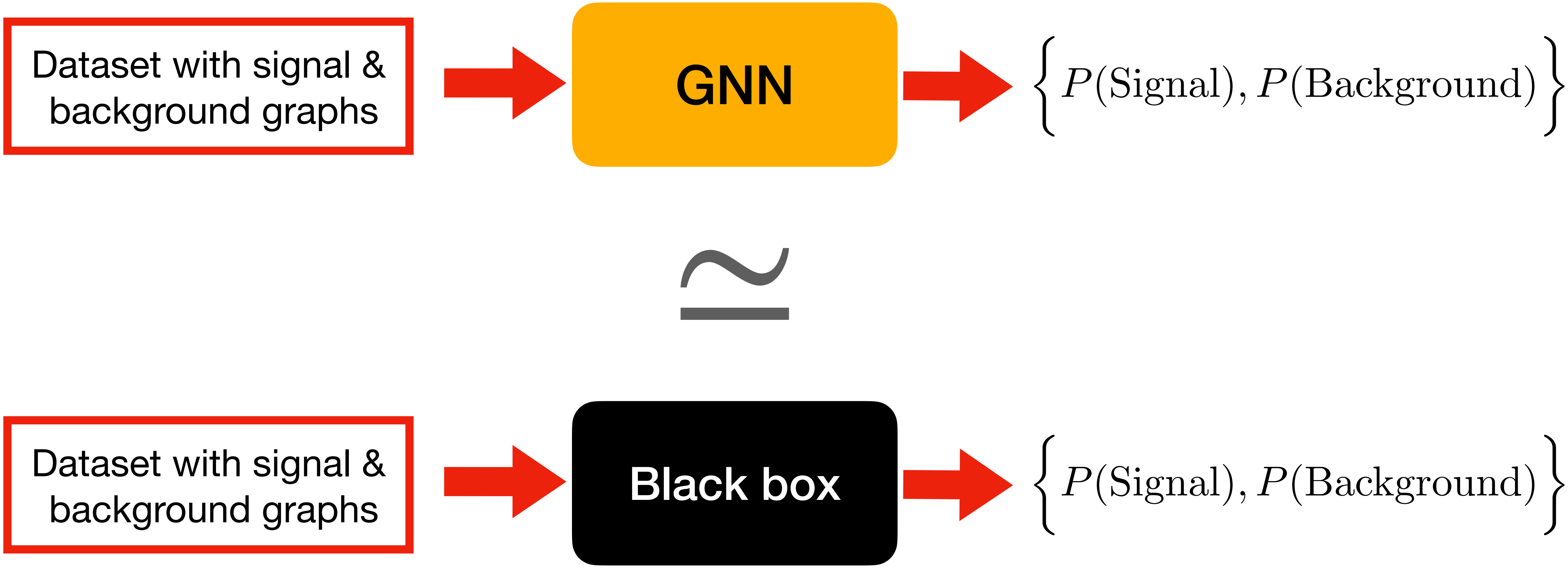
- **Assumption:** No correlations
- Simplified combination of significances (Stouffer method)

$$Z_{\text{comb.}} = \frac{Z_{3b2\tau} + Z_{5b}}{\sqrt{2}}$$

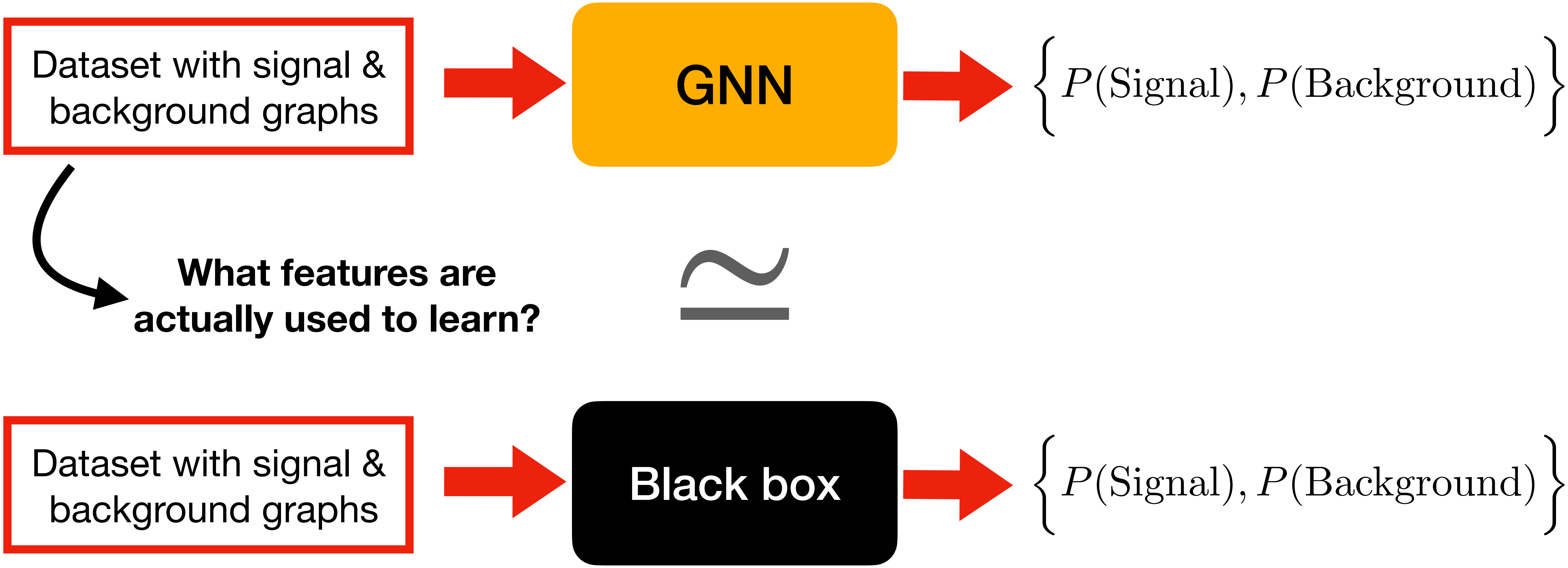
Combination of further channels and improvements of **tagging/reconstruction** methods could enhance results further



How do Neural Networks learn?



How do Neural Networks learn?

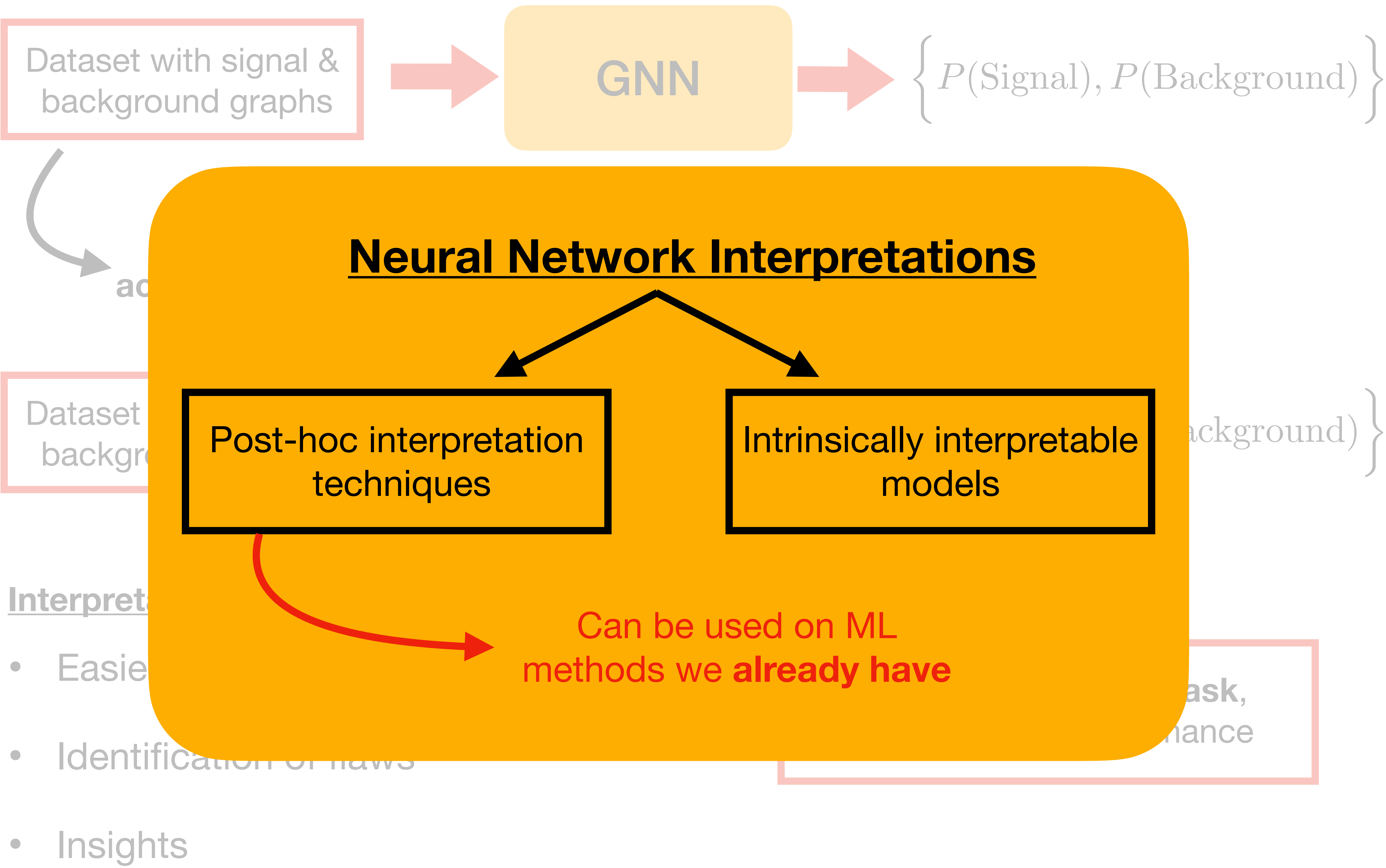


Interpretation of Neural Network results

- Easier debugging
- Identification of flaws
- Insights

Usually seen as a **secondary task**, main focus is performance

How do Neural Networks learn?



Integrated Gradients

→ <u>**Integrated Gradients**</u>: [Sundararajan, Taly, Yan 1703.01365]

- axiomatic method
- uses Neural Network gradients → **fast!**
- requires a differentiable model → **suitable for Neural Networks!**

- Definition:

input / baseline

$$\mathcal{I}_i(x) = (x_i - x'_i) \int_0^1 d\alpha \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i}$$

Attribution scores
→ importance of feature

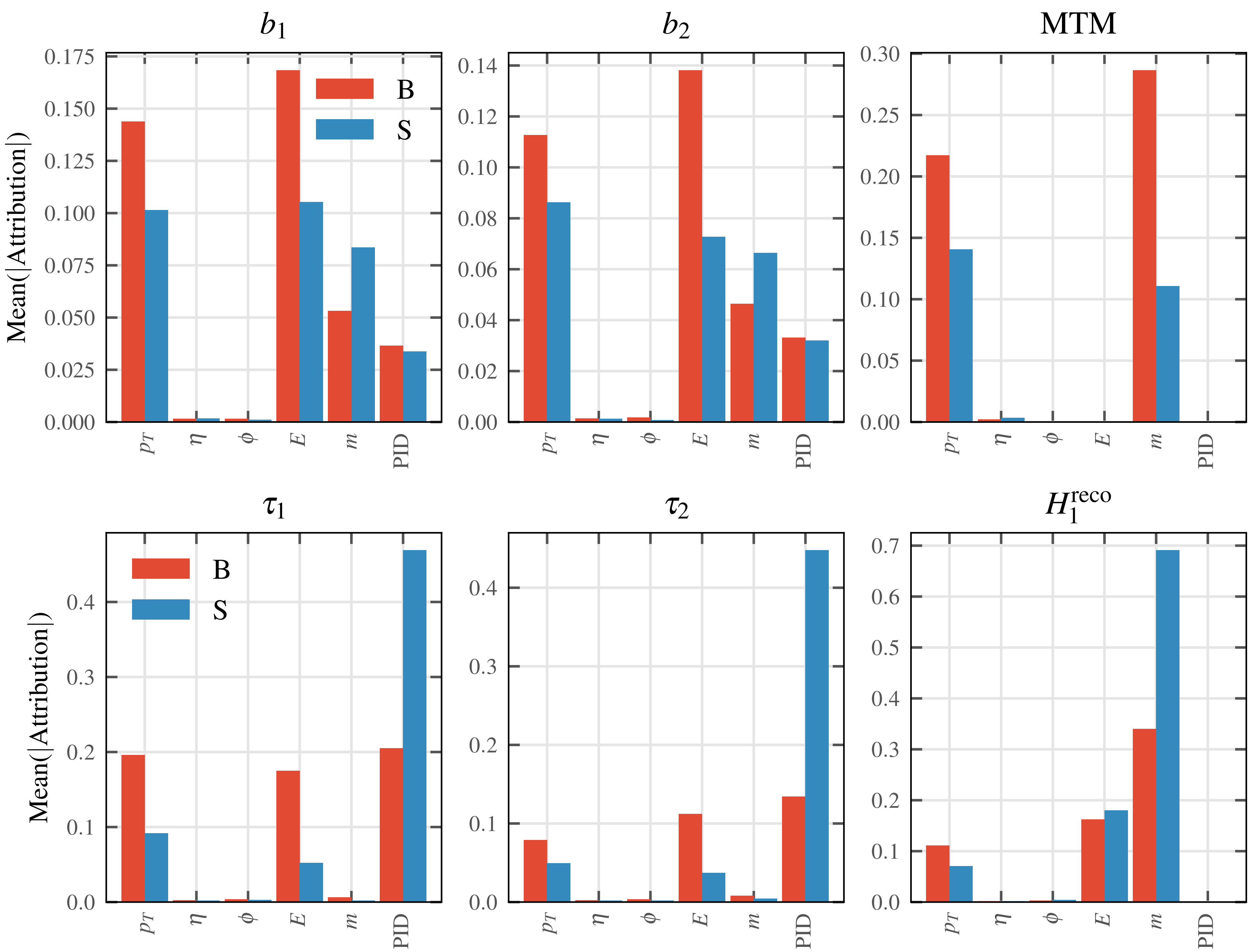
Gradient of Neural Network F

- Easy to implement for Graph Neural Networks as well

- Does **not** take into account graph structures → work in progress in Deep Learning community
- Viable to understand important features → expect mass of reconstructed Higgs to be important

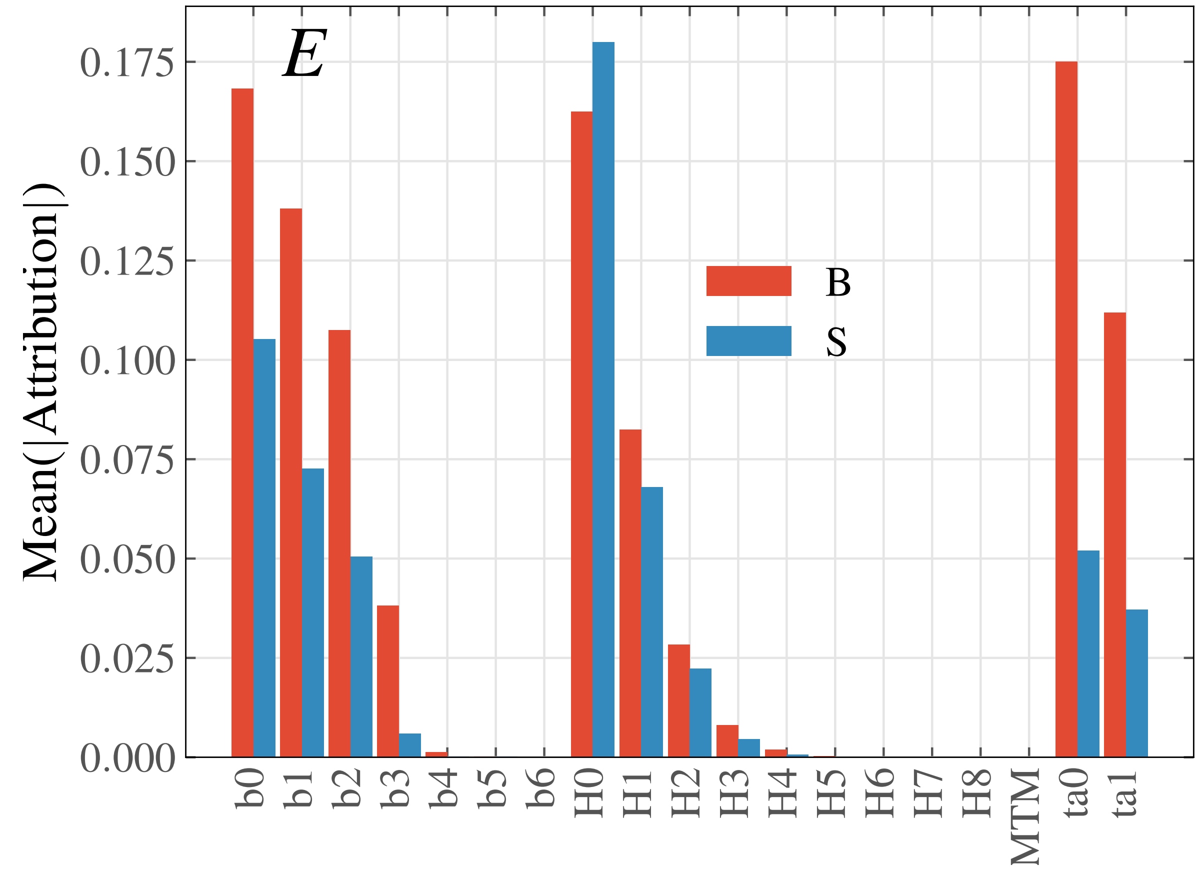
Attributions

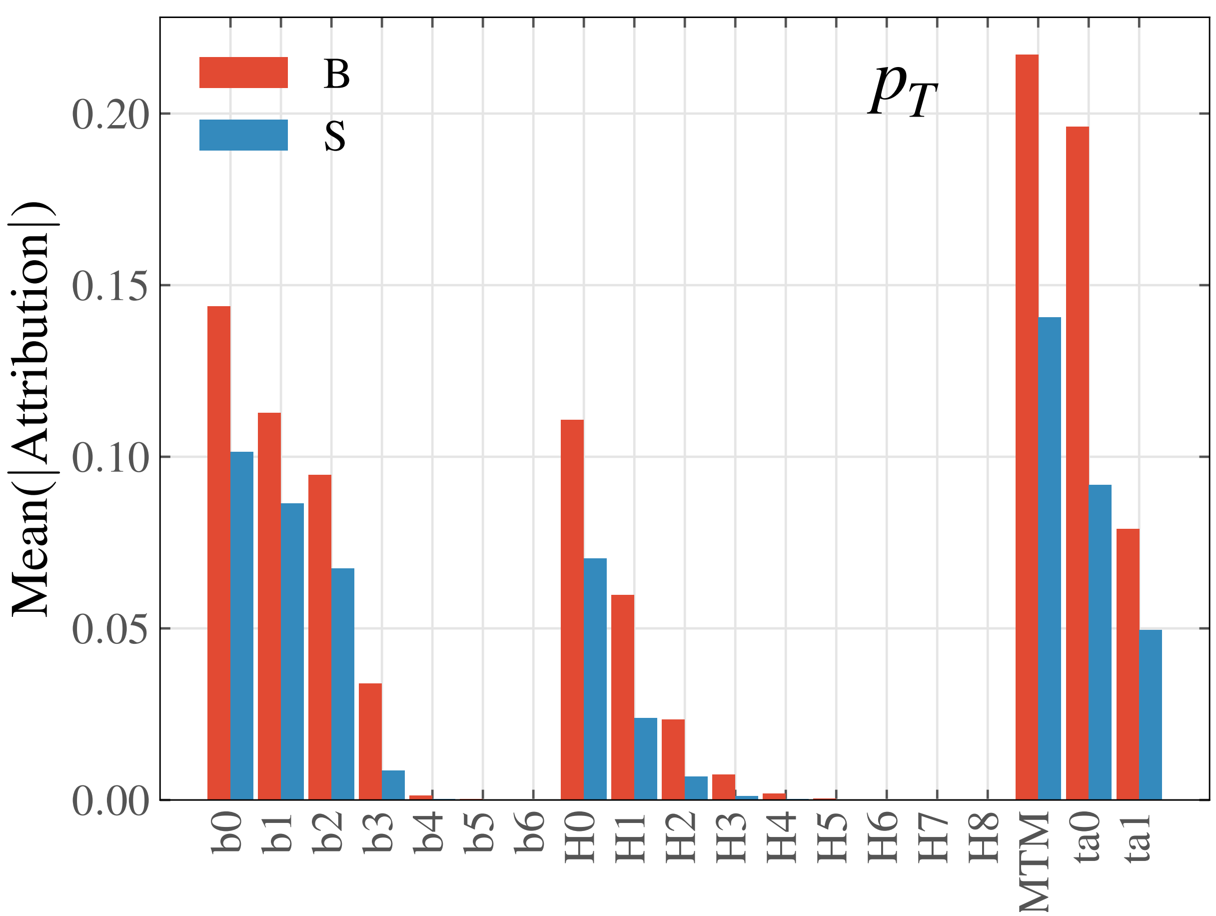
- Tagged b-jets and τ nodes ordered by p_T
- 'Roughly' reconstructed Higgs nodes ordered by 'closeness' to 125 GeV
- p_T , E and PID more important than angular observables
- Higgs masses most important

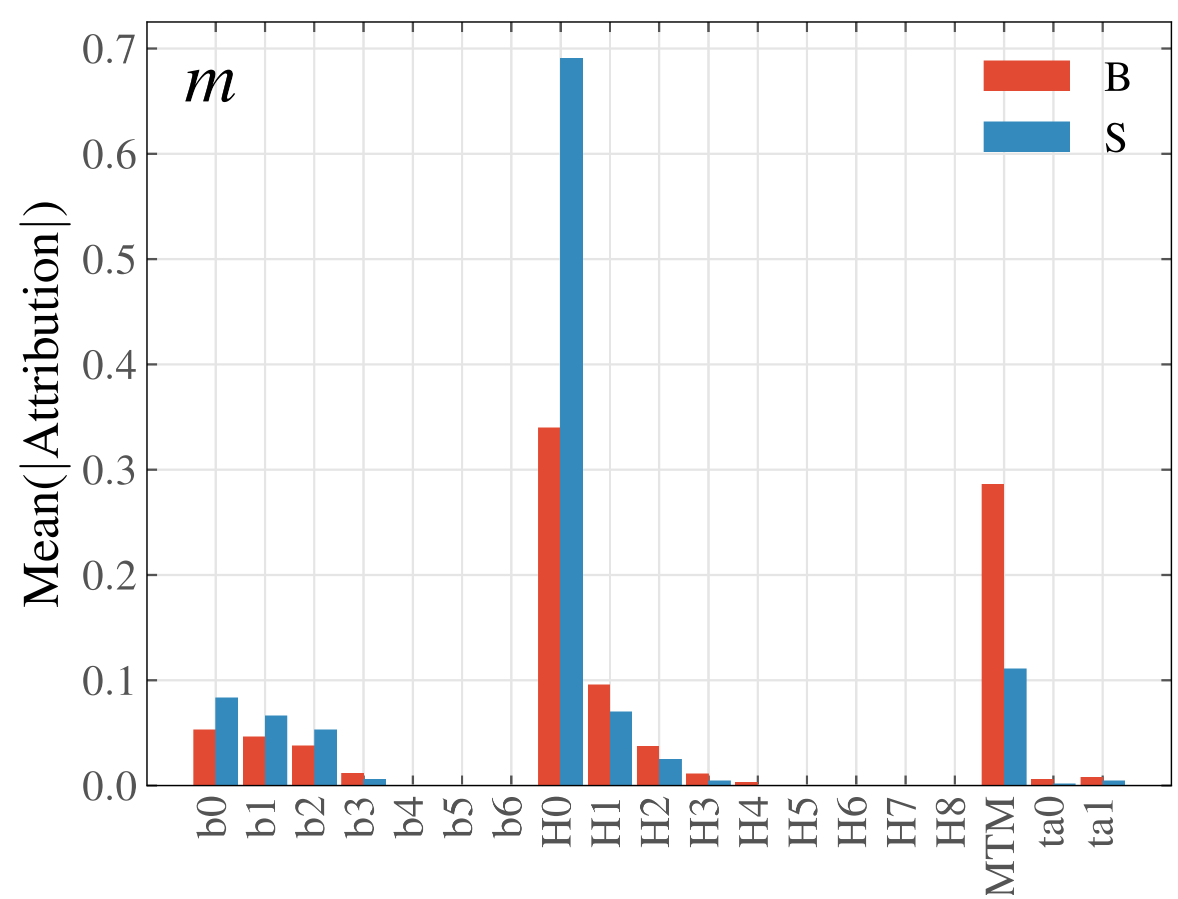


Attribution vs. nodes

- E and p_T from leading order particles is more important
- m is more important for the reconstructed Higgs closest to the SM mass value



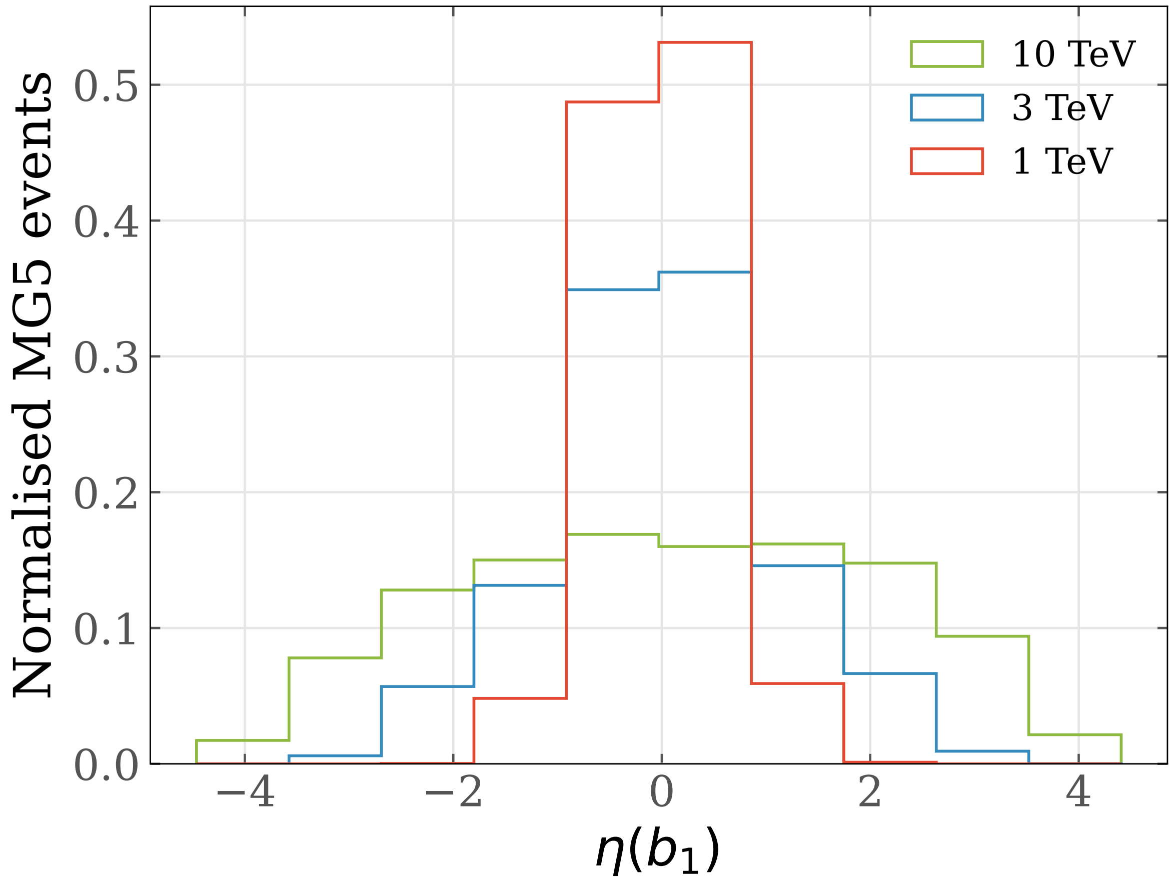




Lepton Colliders

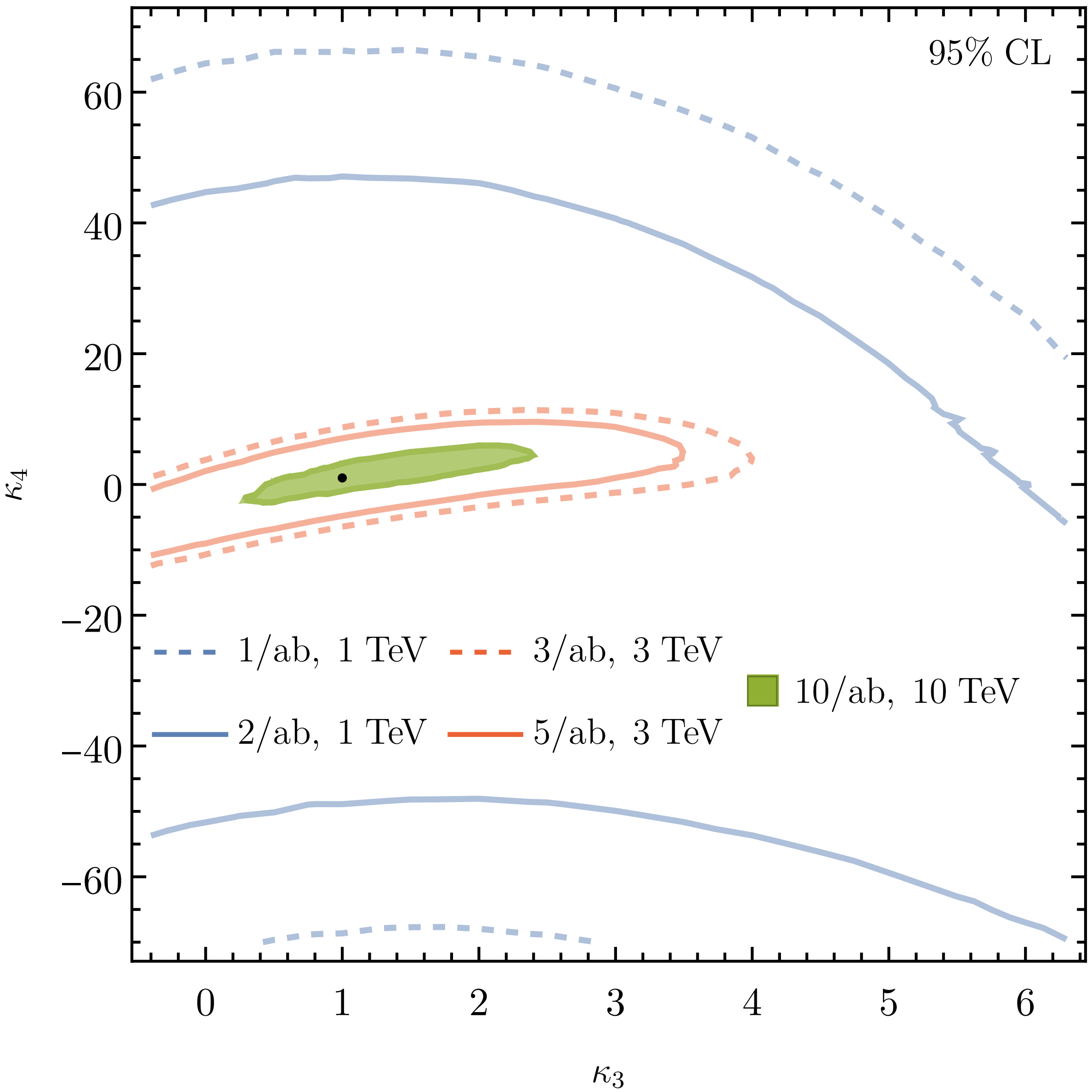
- Complete picture of $(\kappa_3, \kappa_4) \rightarrow$ lepton colliders?
- Inclusive $\ell\ell \rightarrow HHH + X$ analysis with $H \rightarrow b\bar{b}$
    - At least 5 tagged b-quarks with $p_T(b) > 30$ GeV
    - Tagging efficiency: $80\ \%$

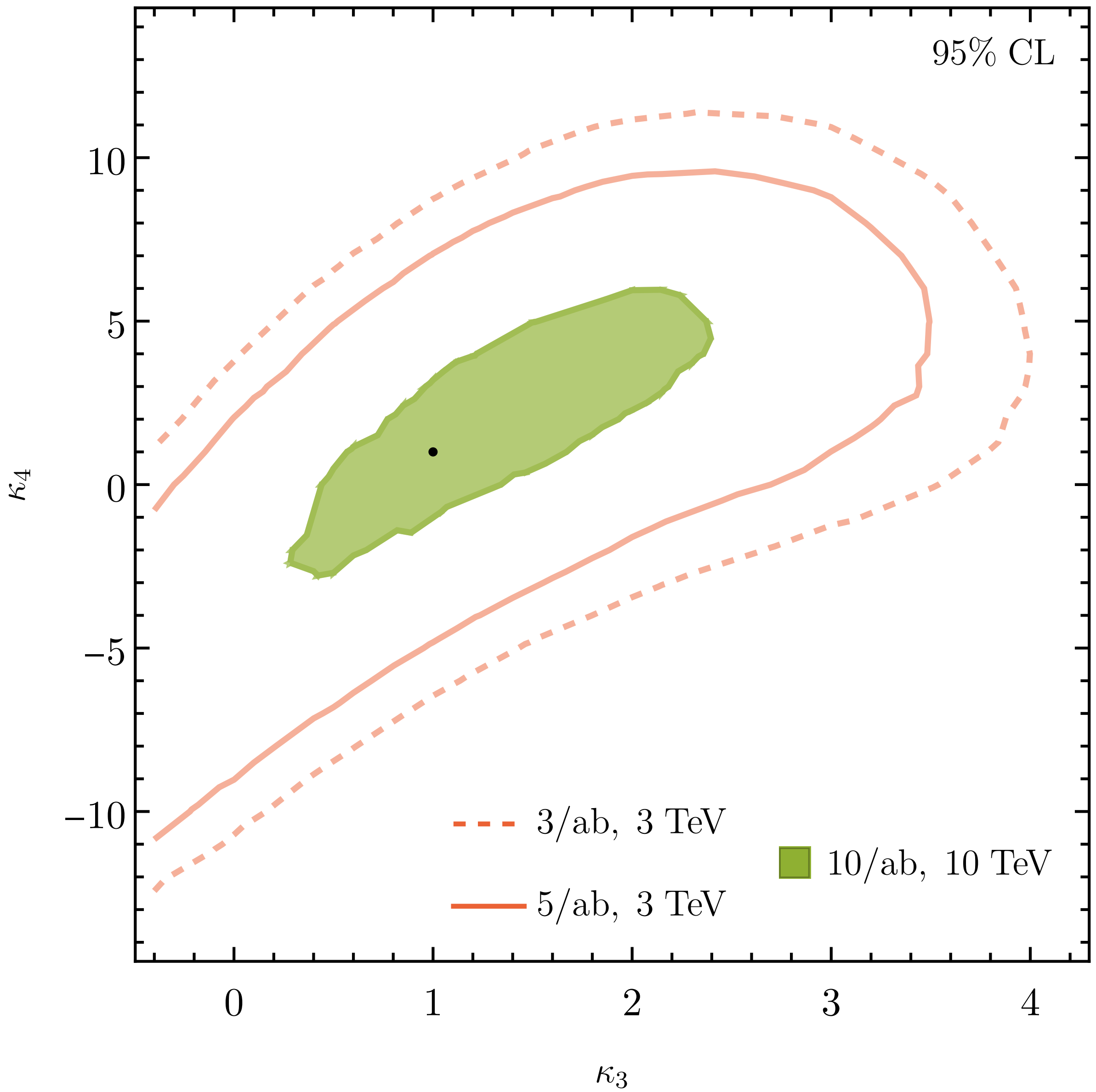
- **Important:** For high energies b-quarks are not only in the central part of detector $\rightarrow$ requires extended tagging capabilities
- Negligible background from other SM processes



Lepton Collider Results

- Poissonian analysis: $\mu_{\text{up}} = \frac{1}{2} F^{-1}_{\chi^2}\left[2(n+1); \text{CL}\right]$
- Results similar to other works with dedicated analyses, e.g. [Maltoni, Pagani, Zhao ` 18]

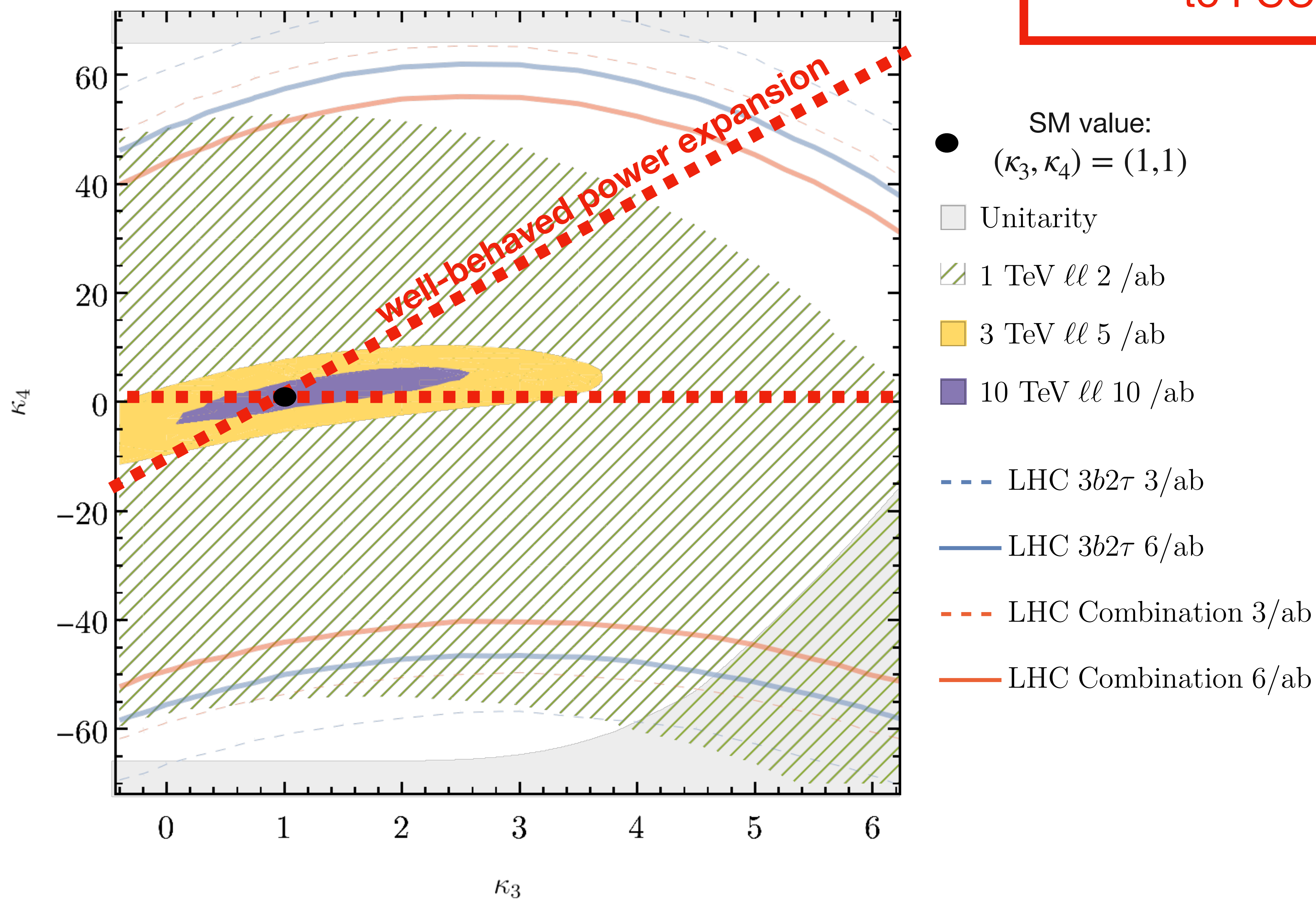




HL-LHC vs. future lepton colliders

- HL-LHC can provide competitive results compared to 1 TeV collider
- High energy lepton collisions way more sensitive

But such machines more comparable to FCC



Conclusions

- If there is a sizeable deviation in κ_3, an even larger deviation in κ_4 is not unreasonable

  sizeable κ_4 deviations allowed by unitarity

- **GNNs** provide enhanced results at HL-LHC
  - HL-LHC should be able to probe regions allowed by unitarity
  - HHH not powerful enough to constrain κ_3 as well as di-Higgs bounds

  BUT can provide complementary information and be used in combination with di-Higgs

- HL-LHC competitive with 1 TeV lepton colliders but higher energies more sensitive

- Neural Network interpretations useful for understanding ML techniques

Thank you!

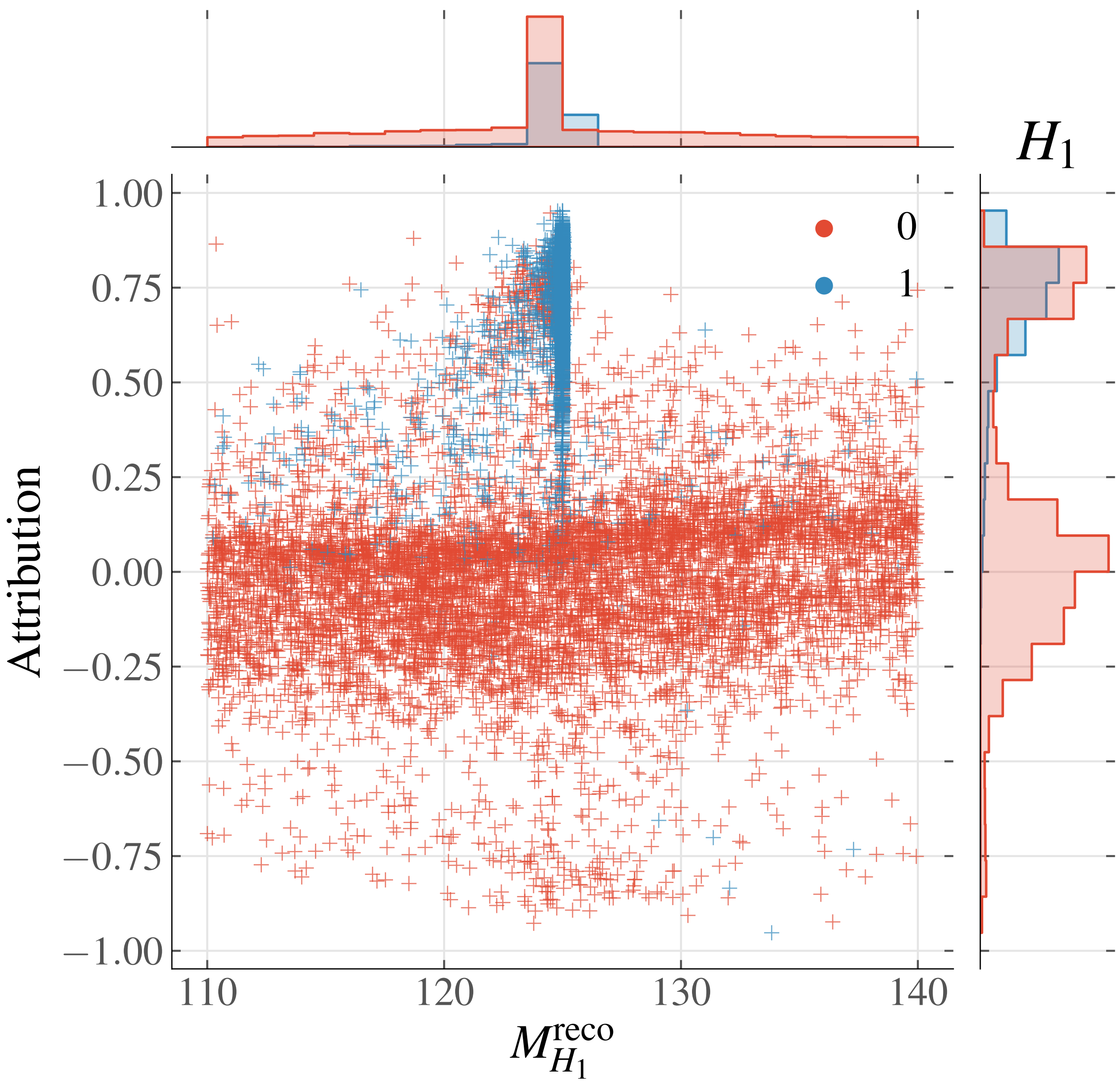
Backup: Interpretation axioms

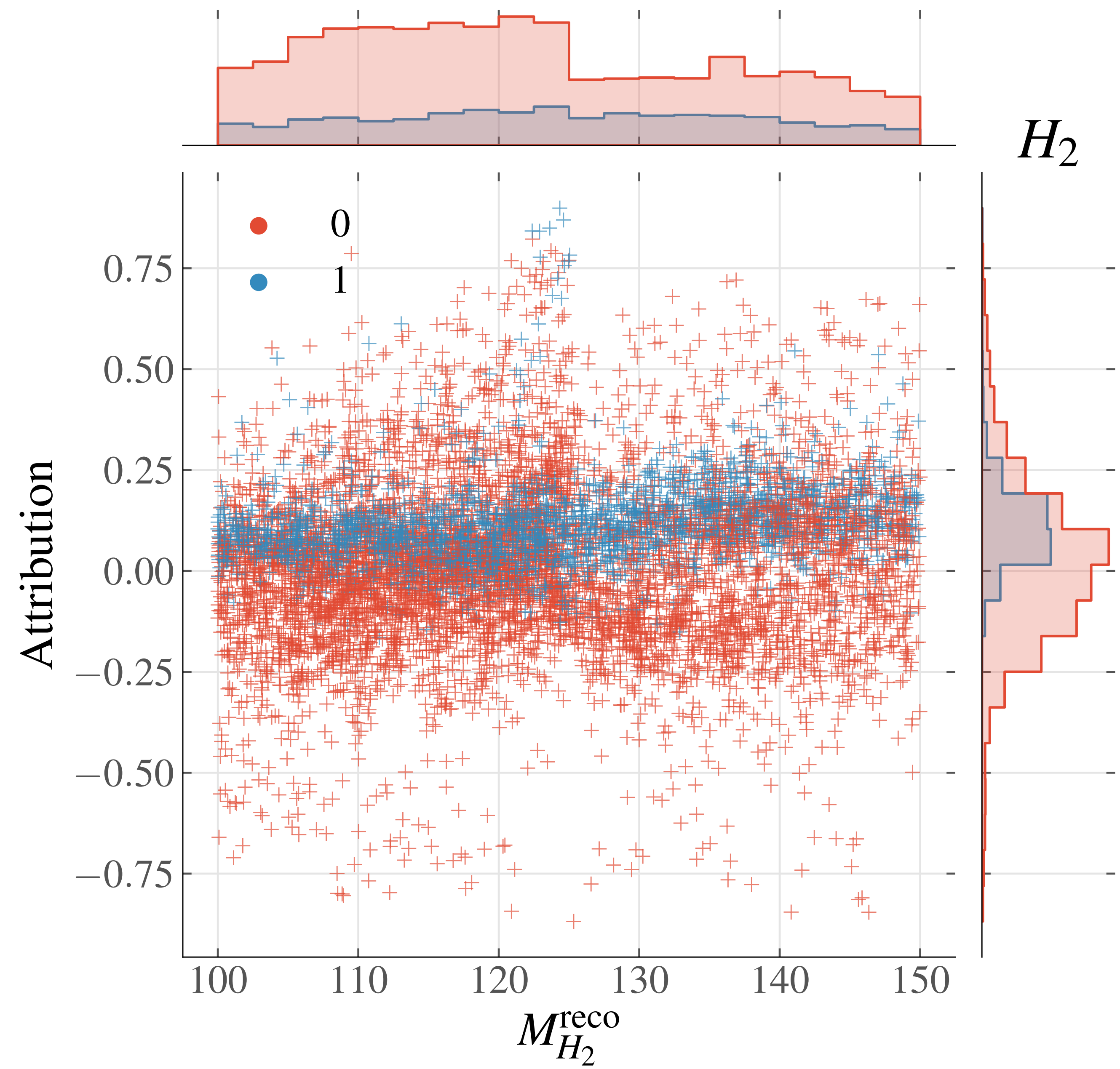
- **Axioms:**
  - **Completeness**: sum of attributions equal to difference of network output for input and baseline values
  - **Sensitivity**: when baseline and input have different values and different NN outputs, attributions should also be different
  - **Dummy**: A zero input should yield no attribution
  - **Implementation Invariance**: If two methods are equivalent (i.e. yield same scores for all inputs despite being different) then attributions should be identical
  - **Linearity**: Attributions should be linear for linear combinations of networks $aF_1 + bF_2$
  - **Symmetry**: For a network symmetric for two variables $F(x, y) = F(y, x)$, the attributions should be the same

Backup: Reconstructed Higgs Mass

- **Interpretation as expected:**
  If a Higgs close to 125 GeV can be found $\Longrightarrow$ signal

- Complete understanding would require to study correlations between observables → **future work**





Backup: Lepton collider cross sections

- Inclusive $\ell\ell \to HHH + X$ analysis with $H \to b\bar{b}$
- Cross sections small below 1 TeV
- **Note:** $\mu^+\mu^-$ vs. e^+e^- collider at 10 TeV has difference of less than $5\,\%$ on cross sections