Investigating the Higgs self-couplings through HHH production

Panagiotis Stylianou

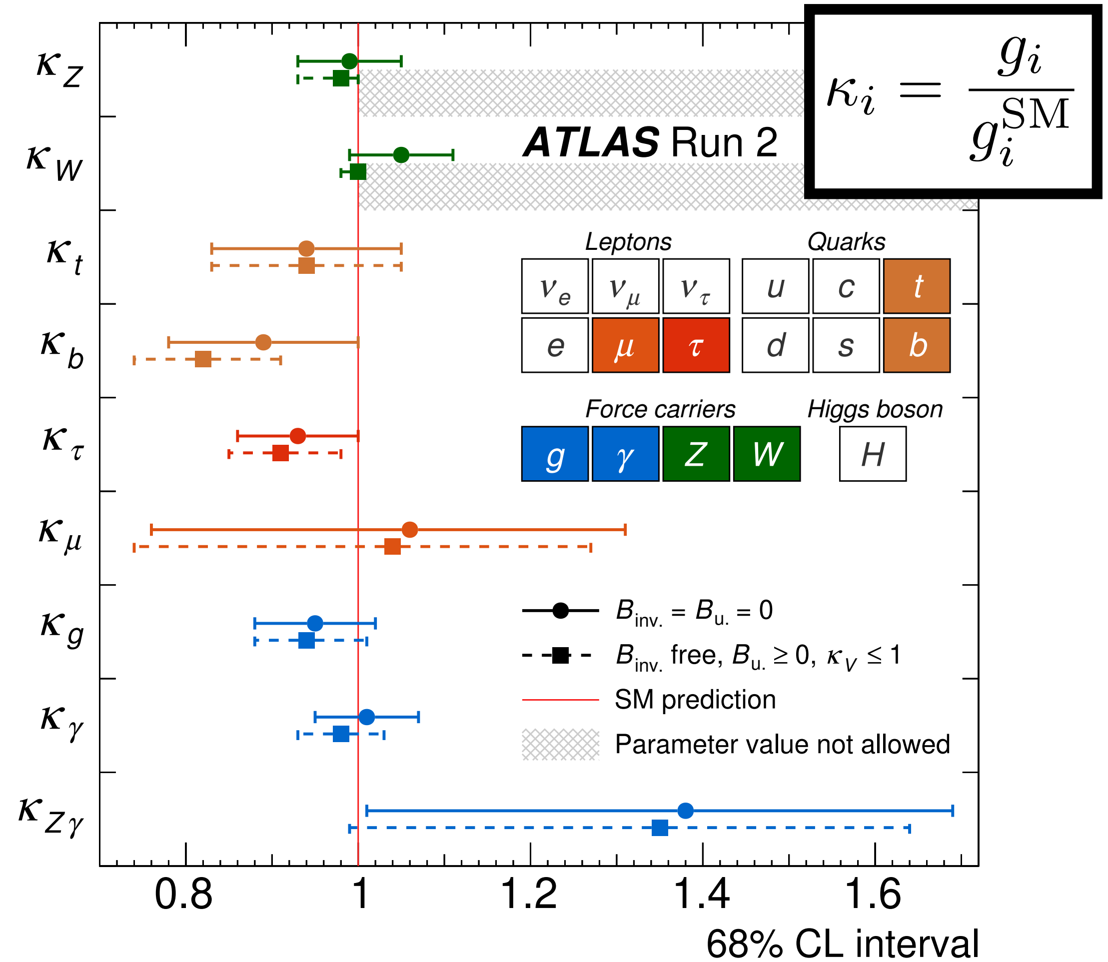
based on work in progress with Georg Weiglein

HHH workshop, Dubrovnik

15 July 2023

Introduction

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

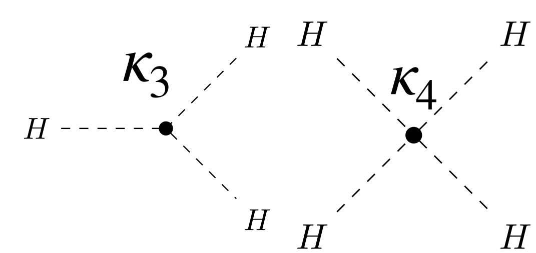


$$\kappa_i = \frac{g_i}{g_i^{\mathrm{SM}}}$$

- **Most challenging:**

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

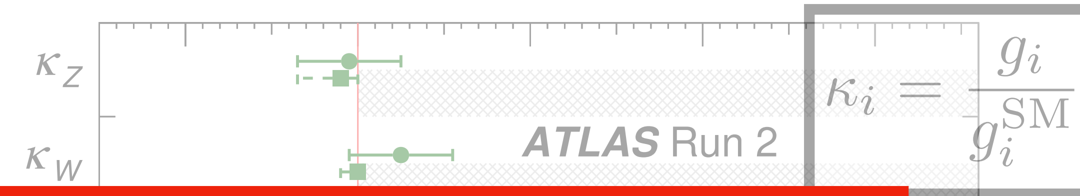
BSM theories → more complicated shapes



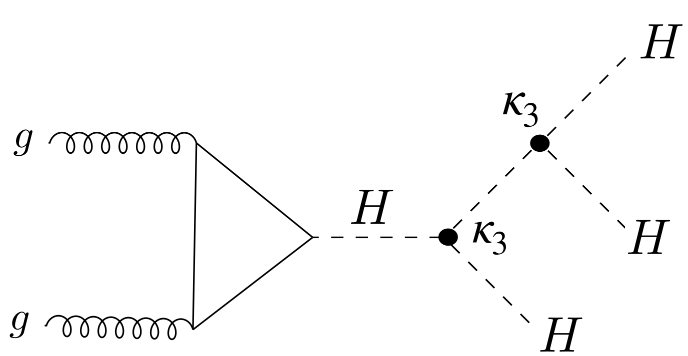
- **First step:**
  - measure κ_3 → double Higgs production
- **But:** κ_3 also appears in triple-Higgs production

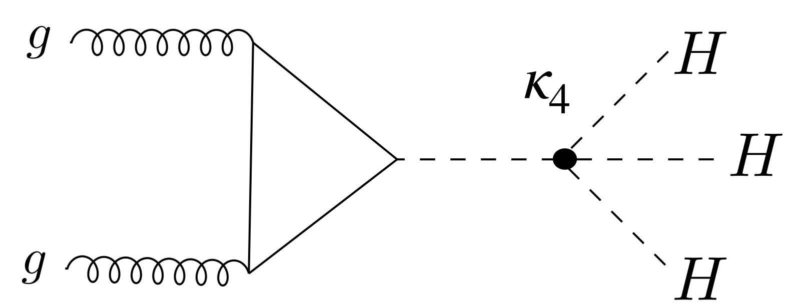
Content

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



- Values of (κ_3, κ_4) theoretically consistent with perturbative unitarity
- 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:

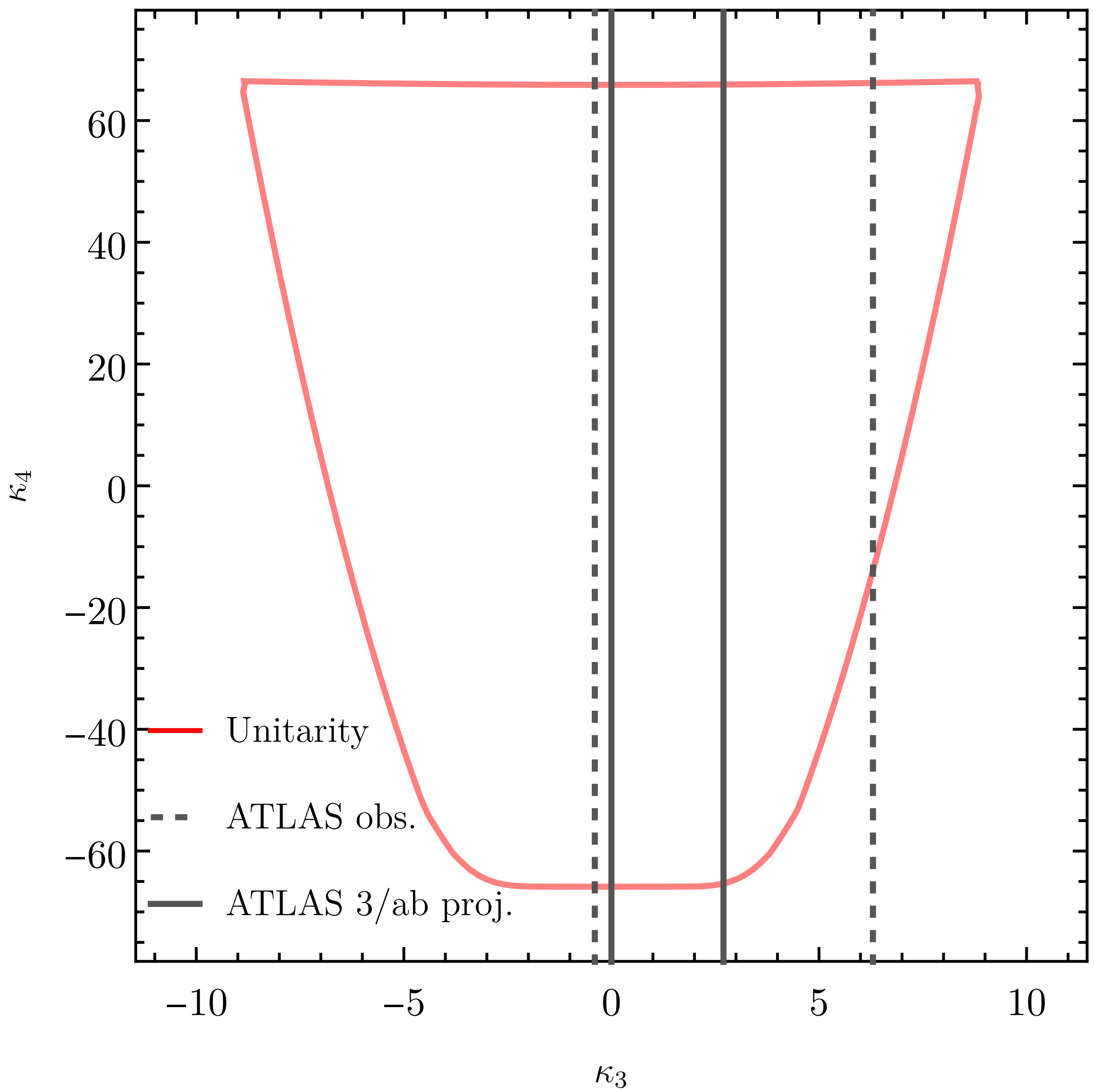
$$\mathrm{Im} a^0_{ii} \geq |a^0_{ii}|^2 \implies |\mathrm{Re} a^0_{ii}| \leq \frac{1}{2}$$

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

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

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

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

Contributions to κ_3, κ_4:

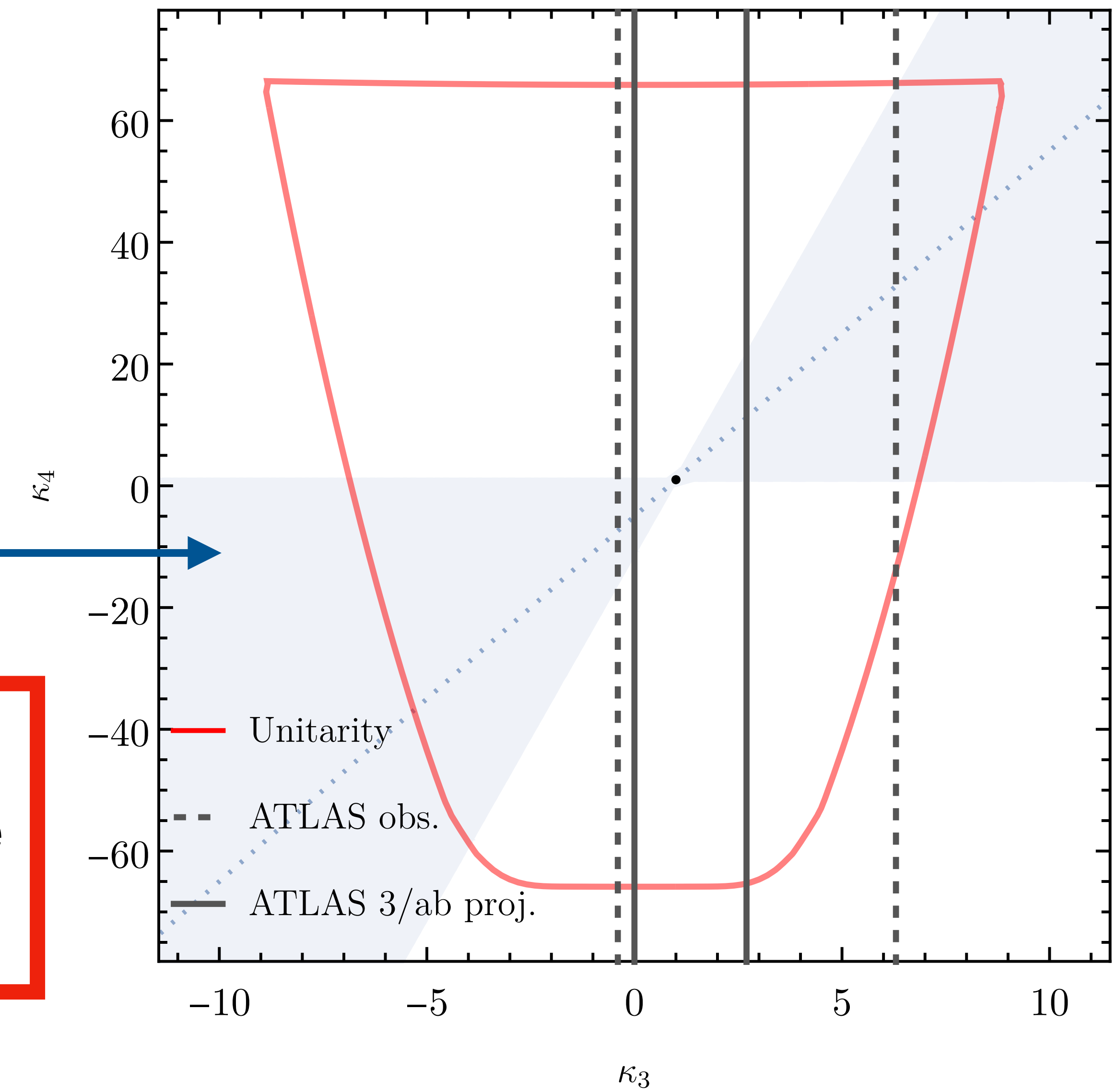
$$(\kappa_3 - 1) = \frac{C_6 v^2}{\lambda\Lambda^2}\,,$$

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

vanishing dimension-8 →

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

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



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

SM Potential higher order terms

Power expansion for higher order terms:

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

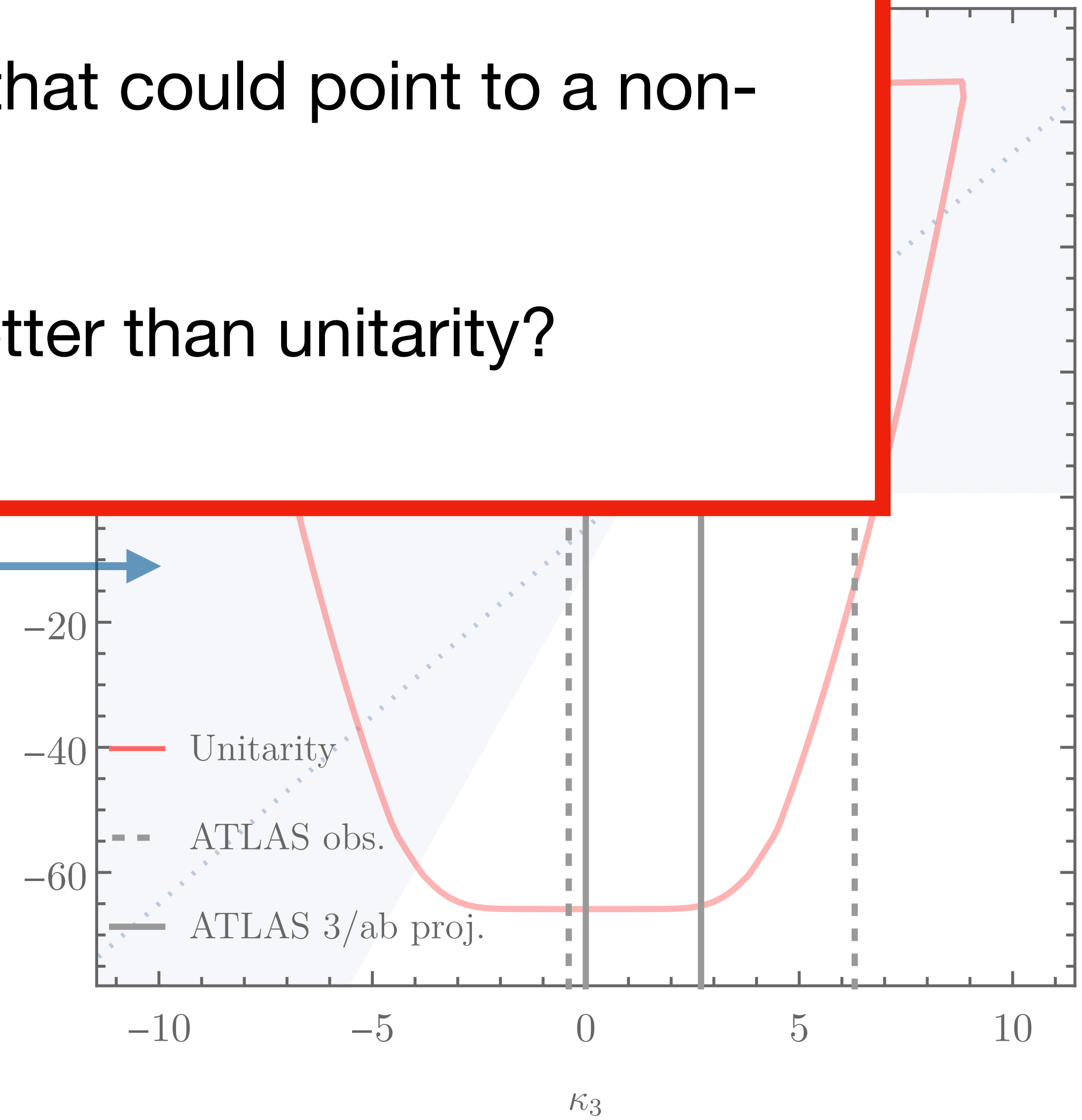
$$V_{\text{BSM}} = \frac{C_6}{\Lambda^2}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^3 + \frac{C_8}{\Lambda^4}\left(\Phi^\dagger\Phi - \frac{v^2}{2}\right)^4 + \ldots$$

Contribution

- Can κ_3 be constrained by HHH production?
- Can HL-LHC probe regions that could point to a non-linear character?
- Can HL-LHC constrain κ_4 better than unitarity?

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**

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])

$4b2\tau$: dominant $\mathcal{O}(\alpha^2)\mathcal{O}(\alpha_S^4)$

subdominant (~15 times smaller) $\mathcal{O}(\alpha^4)\mathcal{O}(\alpha_S^2)$

neglected $\mathcal{O}(\alpha^6)$

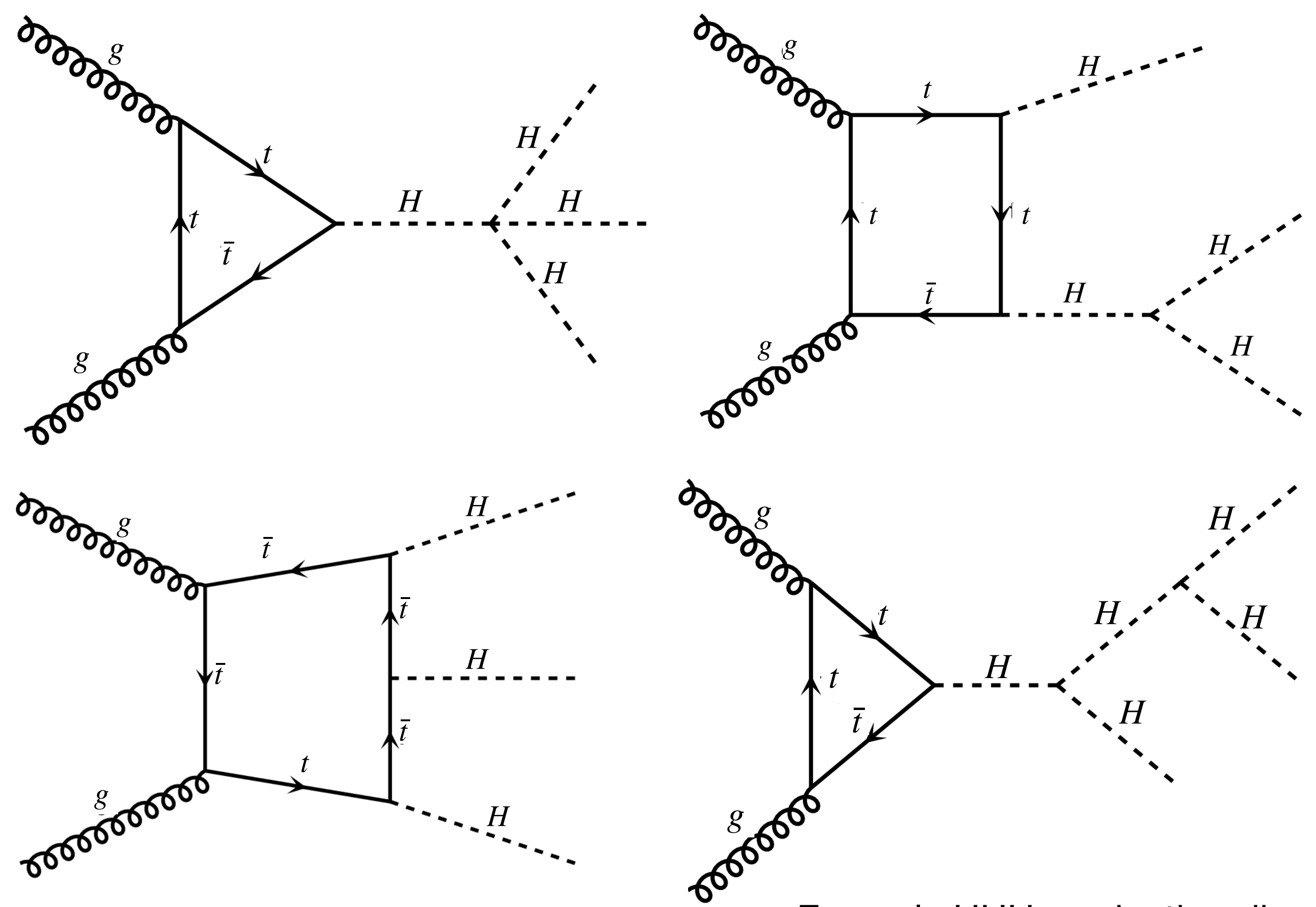
Event generation and pre-selection

- 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 $m_{ij} \in [110, 140]$

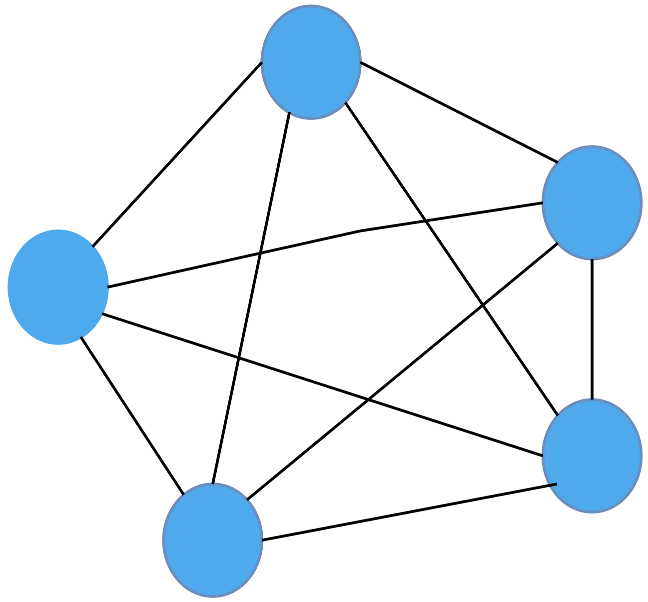
$p_T(b) > 30$ GeV $\quad$ $p_T(\tau) > 10$ GeV

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

Graph Embedding

1.

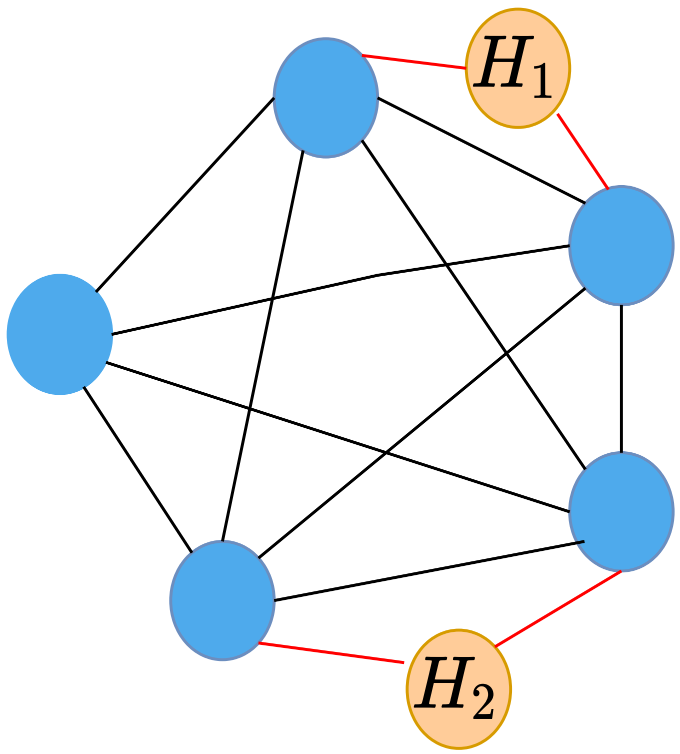
- Fully-connected nodes for b and τ final states
- **Input features**: $[p_T, \eta, \phi, E, m,$ PDGID$]$



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 node H_i



RN: Reconstructed Nodes

Edge Convolution

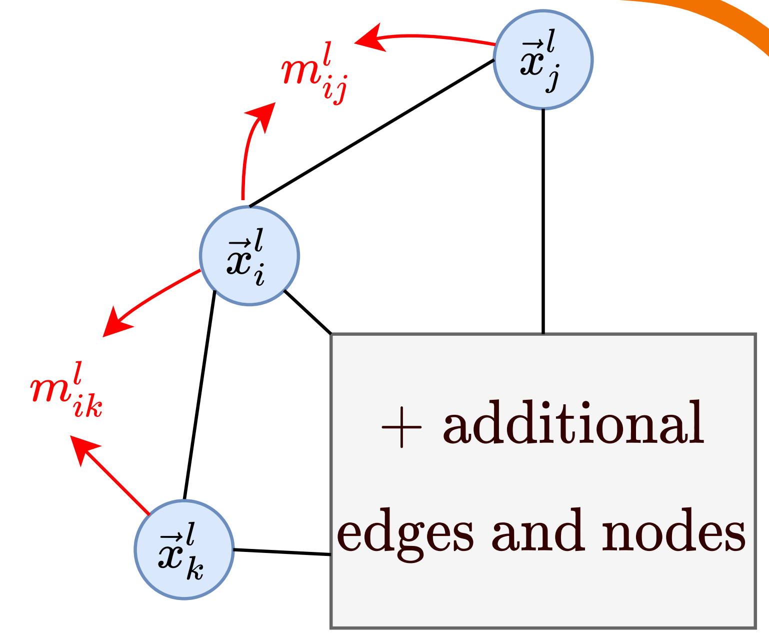
Input features: $\vec{x}_i^{(0)}$ → update iteratively with **Edge Convolution** operation:

Edge Convolution operation

'Message' calculation:

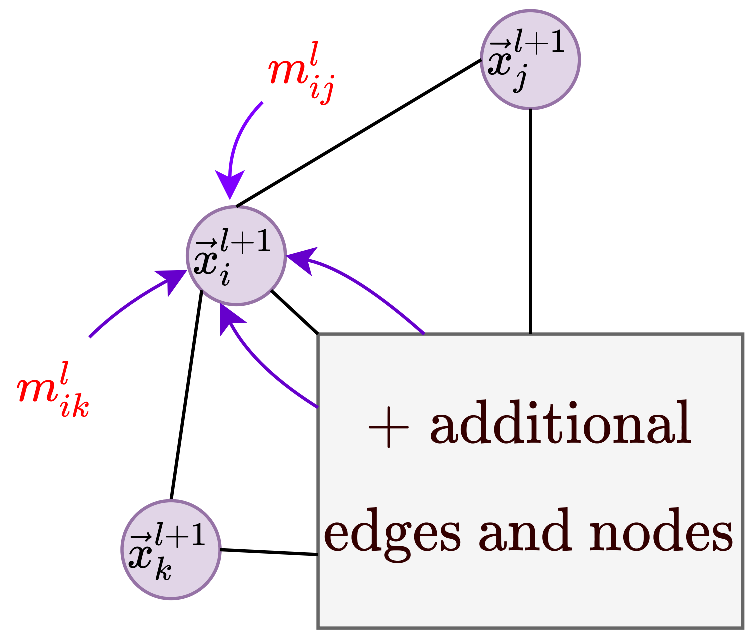
$$m_{ij}^{(l)} = \text{ReLU}\left(\Theta(\vec{x}_j^{(l)} - \vec{x}_i^{(l)}) + \Phi(\vec{x}_i^{(l)})\right)$$

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 efficiencies

Dataset with signal & background graphs → **GNN** (2 EdgeConv layers + linear) → $\Big\{ P(\text{Signal}), P(\text{Background}) \Big\}$

- GNN trained on $(\kappa_3, \kappa_4) = (1,1)$ sample
- Identify NN score threshold with $99\ \%$ background rejection → Care needed! Background should not be depleted

Complex final states

↓

Compare promising channels at parton level

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

'Reconstructed Nodes' case outperforms Fully-Connected

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

GNN efficiencies & significance

Dataset with signal & background graphs → GNN → $\left\{P(\text{Signal}), P(\text{Background})\right\}$

- **Assumption:** Same GNN efficiency for other values of (κ_3, κ_4)
- Flat optimistic $80\ \%$ b-tagging and τ-tagging efficiency
- **Significance:** $Z = \left[2\left((S+B)\log\left(\frac{(S+B)(B+\sigma_B^2)}{(B^2+(S+B)\sigma_B^2)}\right) - \frac{B^2}{\sigma_B^2}\log\left(1+\frac{\sigma_B^2 S}{B(B+\sigma_B^2)}\right)\right)\right]$

362τ | 0.27 | 0.98

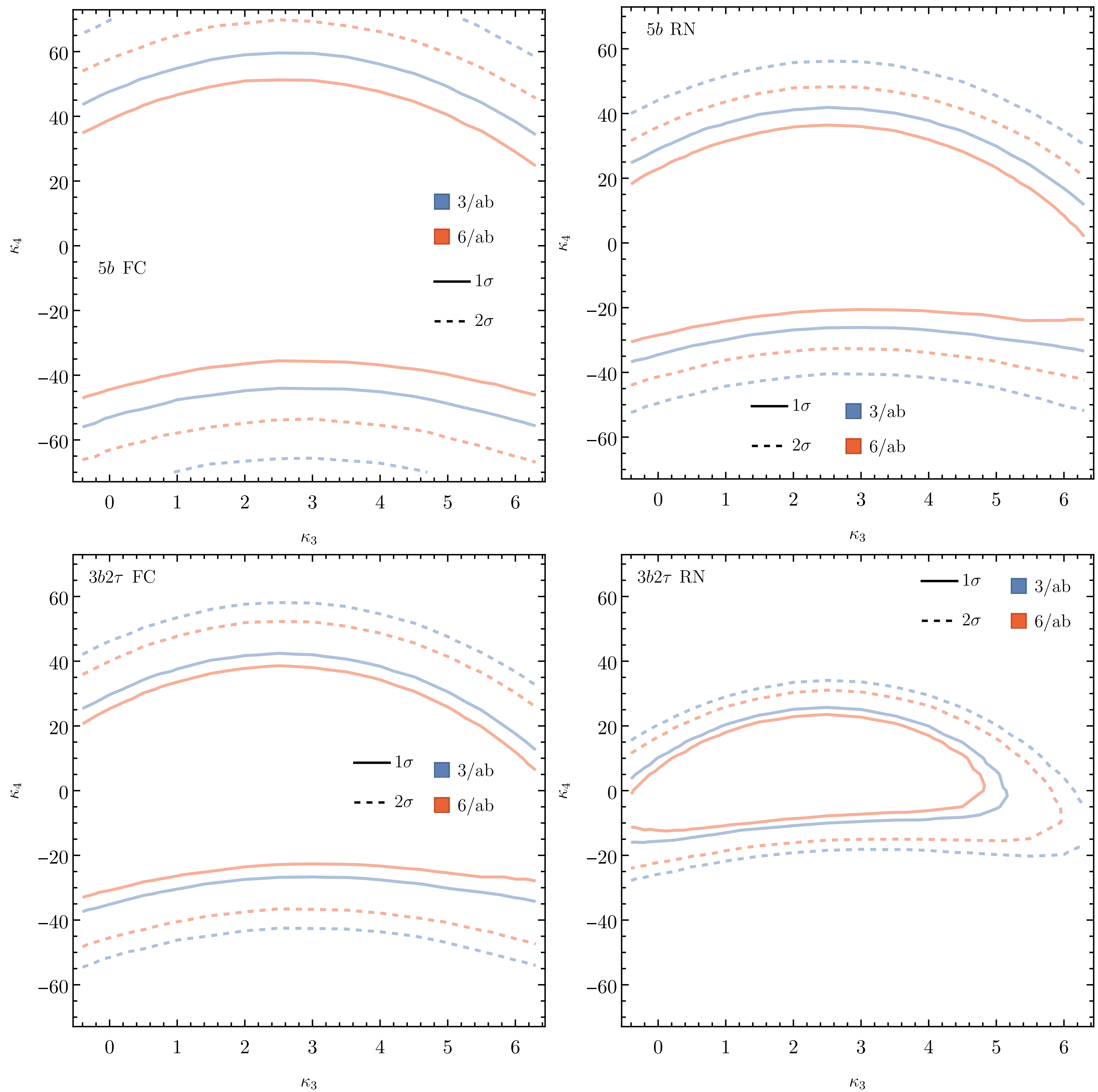
Compare promising channels at parton level

'Reconstructed Nodes' case outperforms Fully Connected

$5b$ vs. $3b2\tau$ at parton-level

Evaluate bounds for:

- HL-LHC 3/ab luminosity
- Combined ATLAS & CMS 6/ab luminosity



- $3b2\tau$ more promising!

BUT are the bounds realistic?

→ should evaluate beyond parton-level

→ will only consider $3b2\tau$

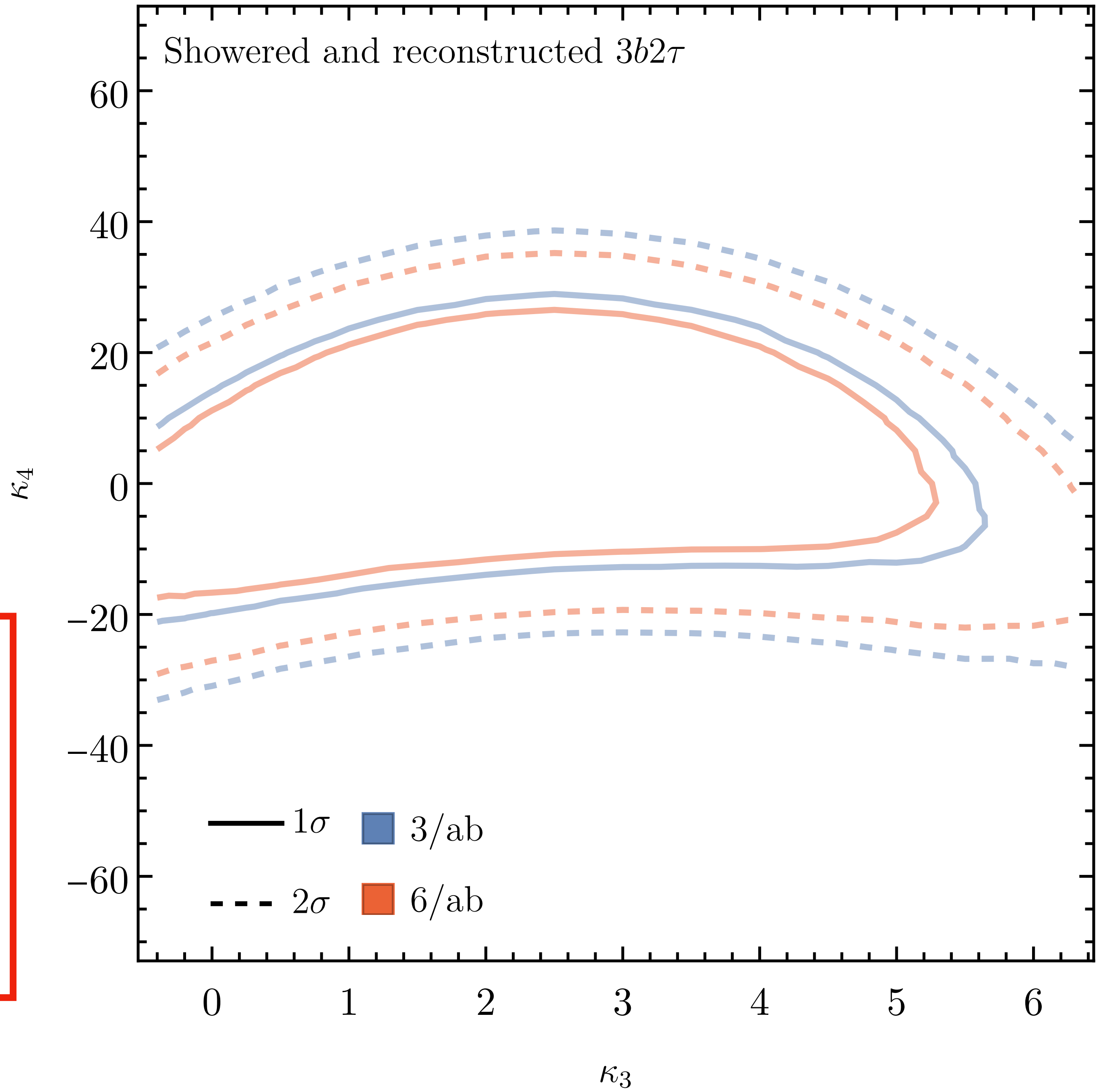
Showered and reconstructed results

- Showering and reconstruction of events: Pythia, FastJet, Rivet
- Include mis-tagging effects for c-quarks efficiency: $\sim 20\,\%$

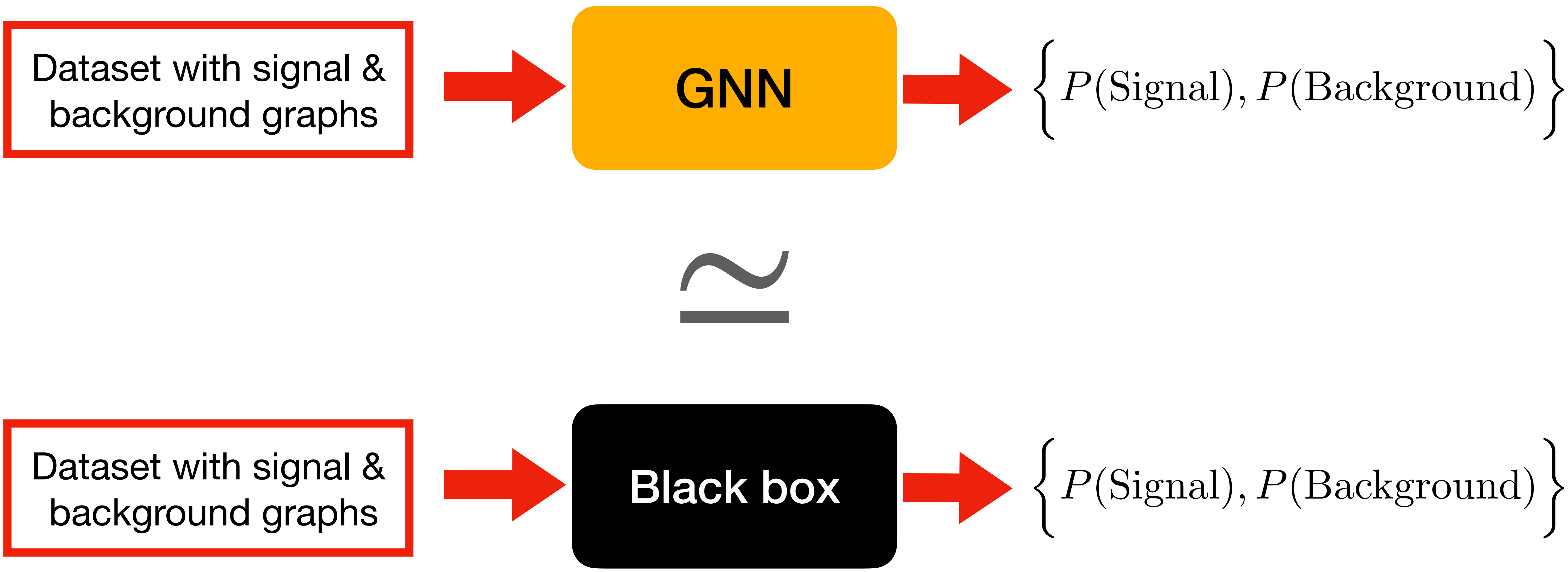
  → included as uncertainty

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

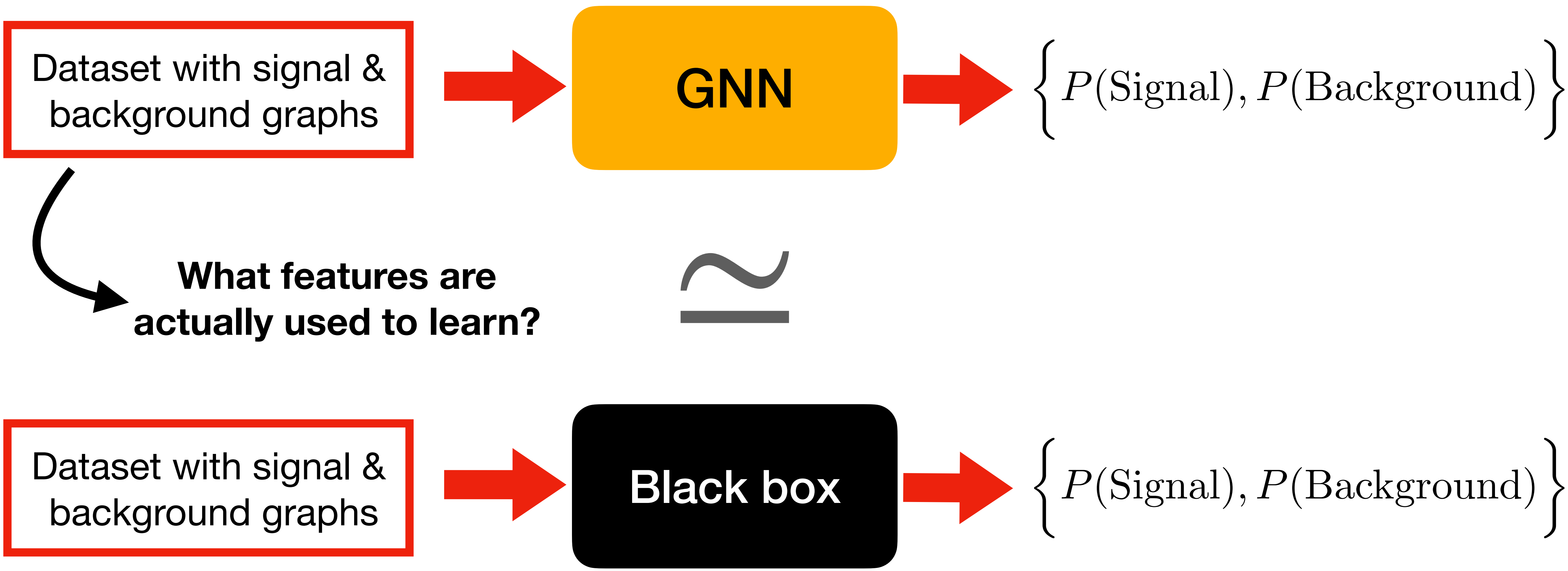
Optimistic results
but more sophisticated reconstruction/tagging techniques and combinations could yield improvements



How do Neural Networks learn?



How do Neural Networks learn?

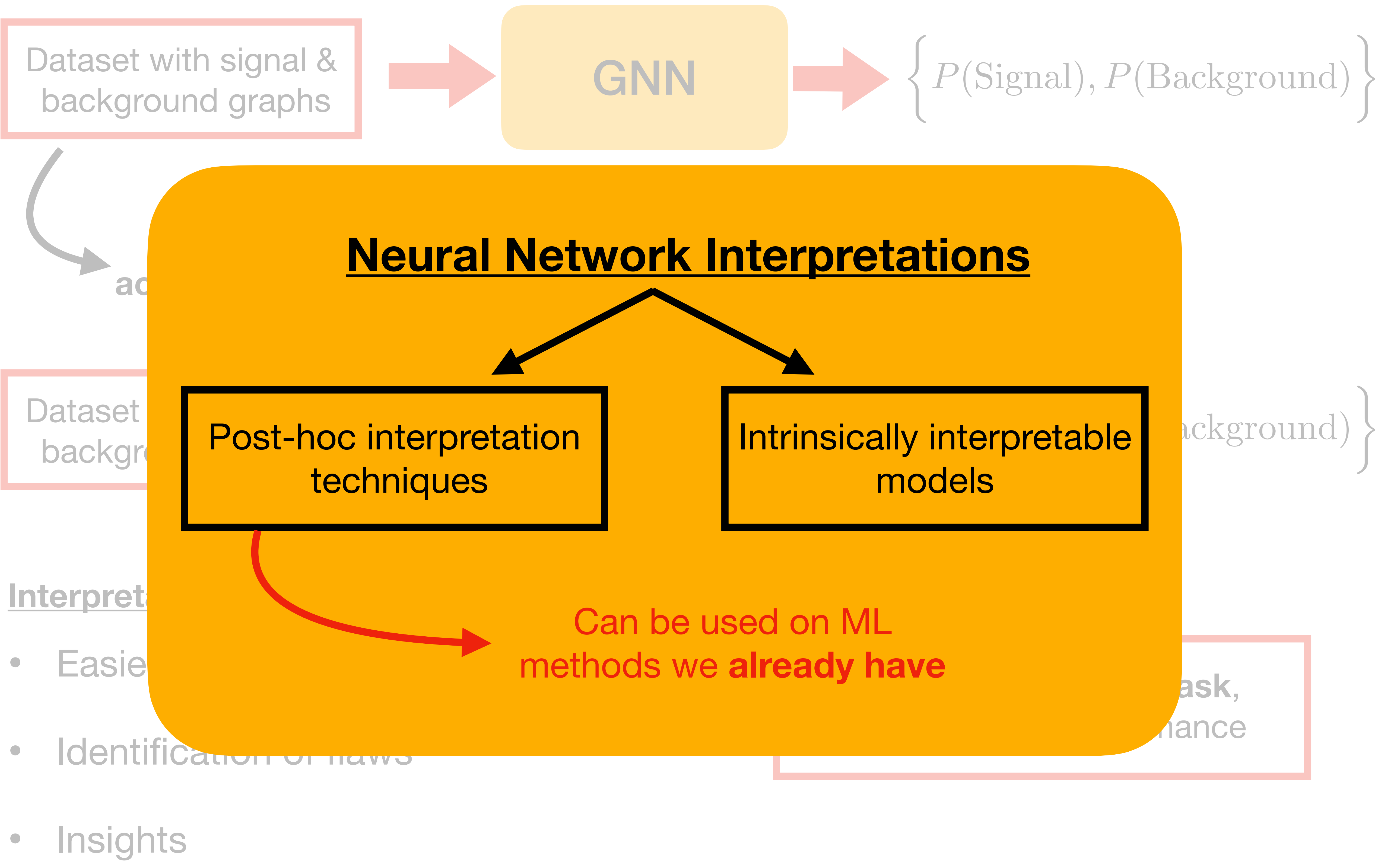


Interpretation of Neural Network results

- Easier debugging
- Identification of flaws
- Insights

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

How do Neural Networks learn?



Integrated Gradients

→ **Integrated Gradients**: [Sundararajan, Taly, Yan 1703.01365]

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

→ **suitable for Neural Networks!**

- Definition:

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

input: x_i; baseline: x'_i

Attribution scores → importance of feature

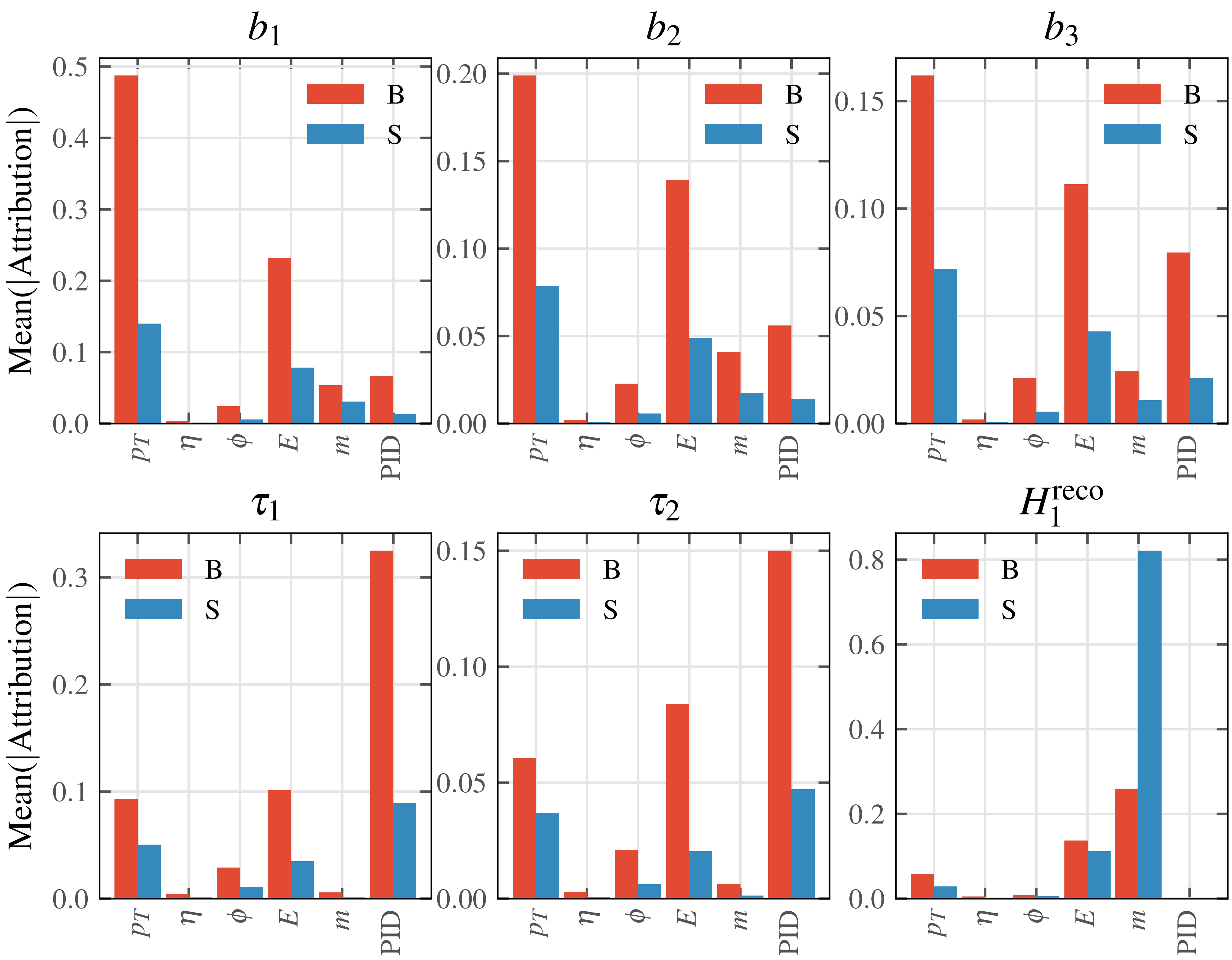
Gradient of Neural Network F

- 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

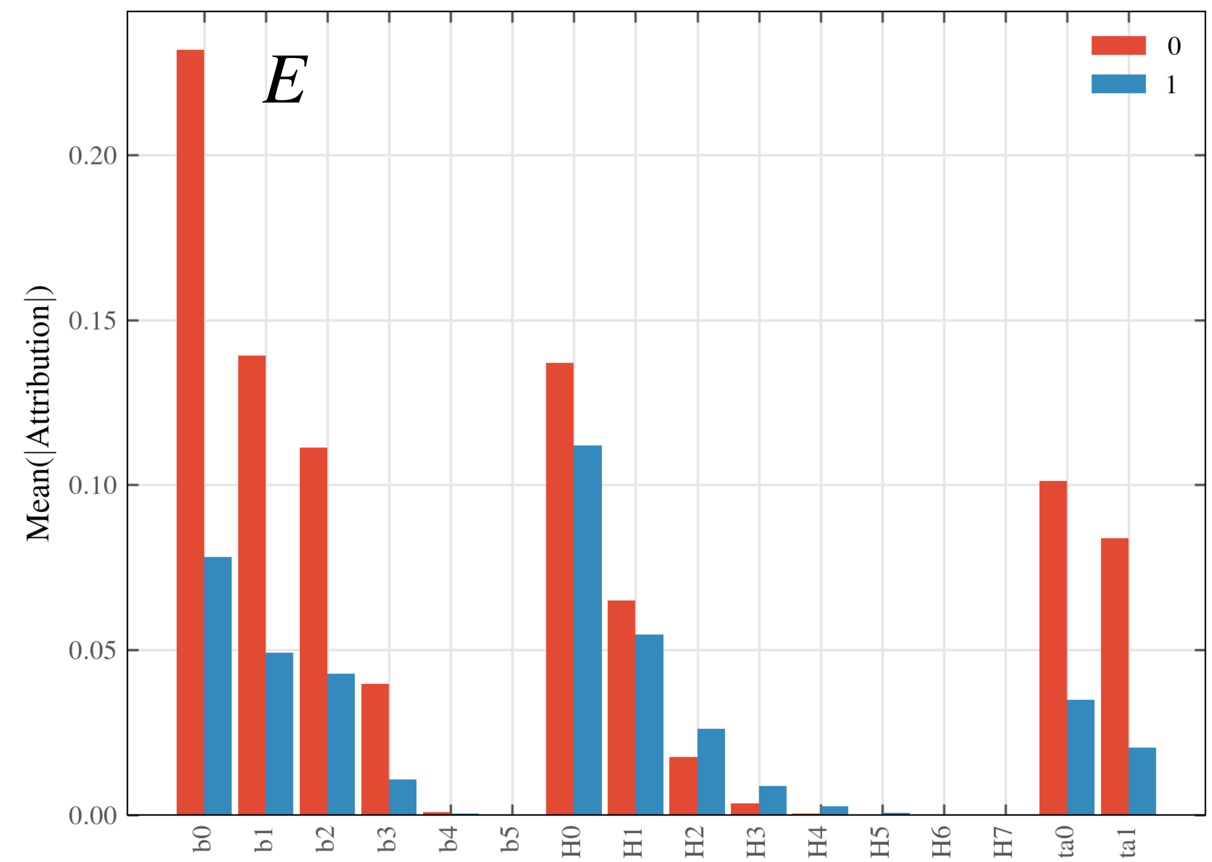
Attributions

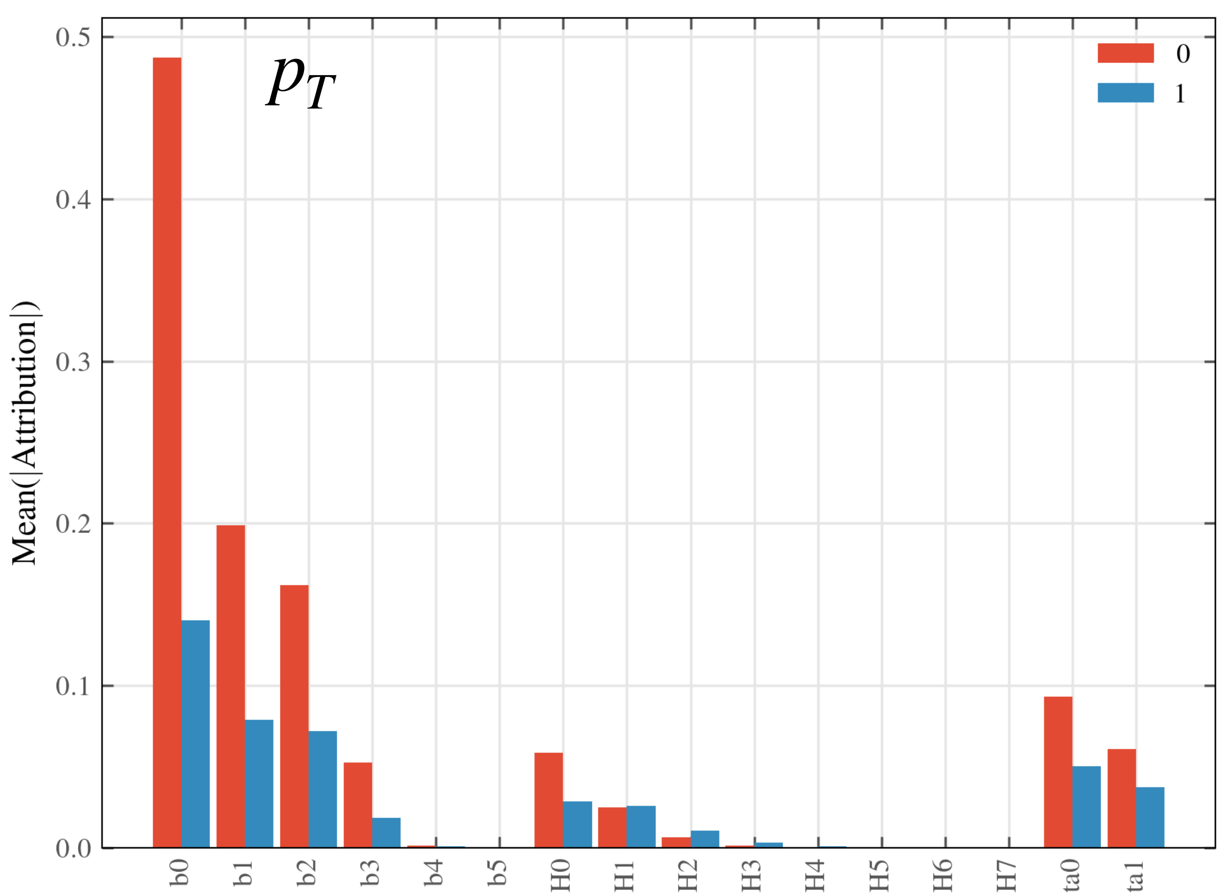
- Tagged b-jets and τ nodes ordered by p_T
- 'Roughly' reconstructed Higgs nodes ordered by 'closeness' to 125 GeV
- p_T , E and PID more important than angular observables
- Higgs masses most important

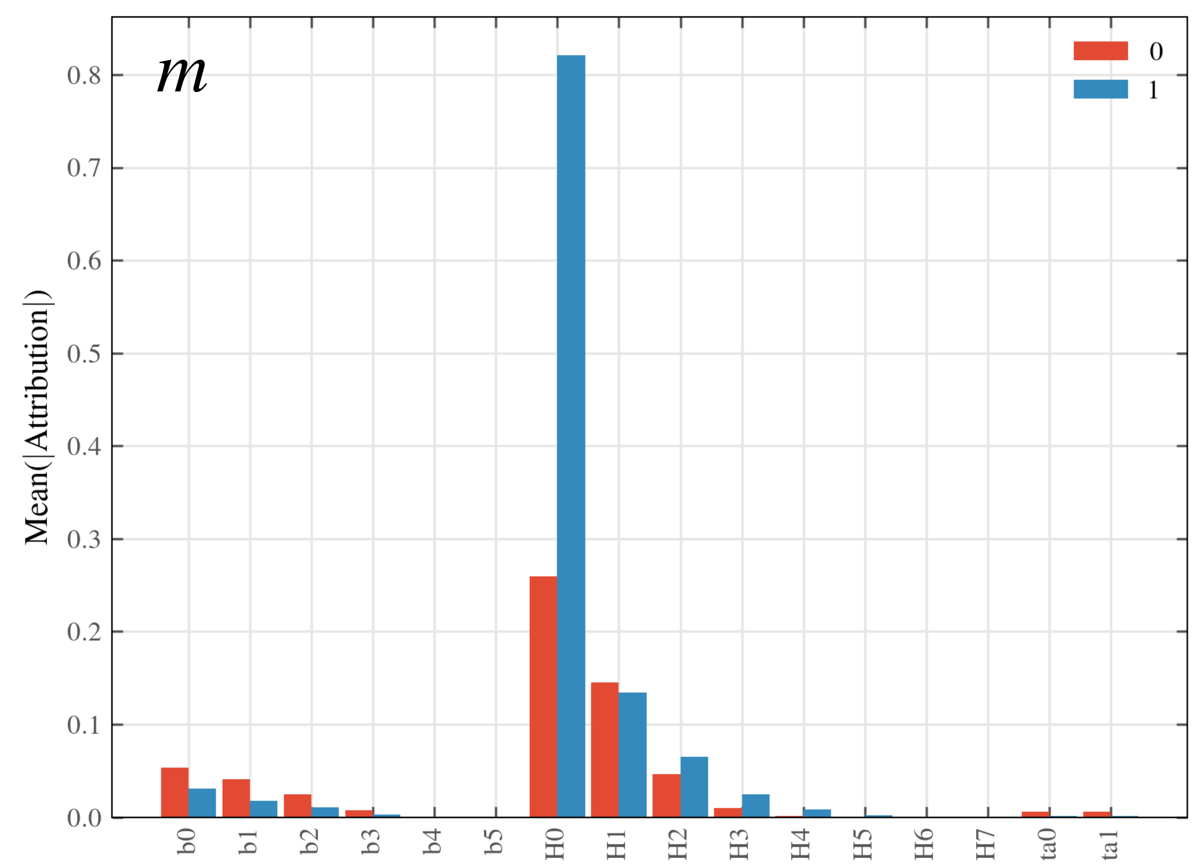


Attribution vs. nodes

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





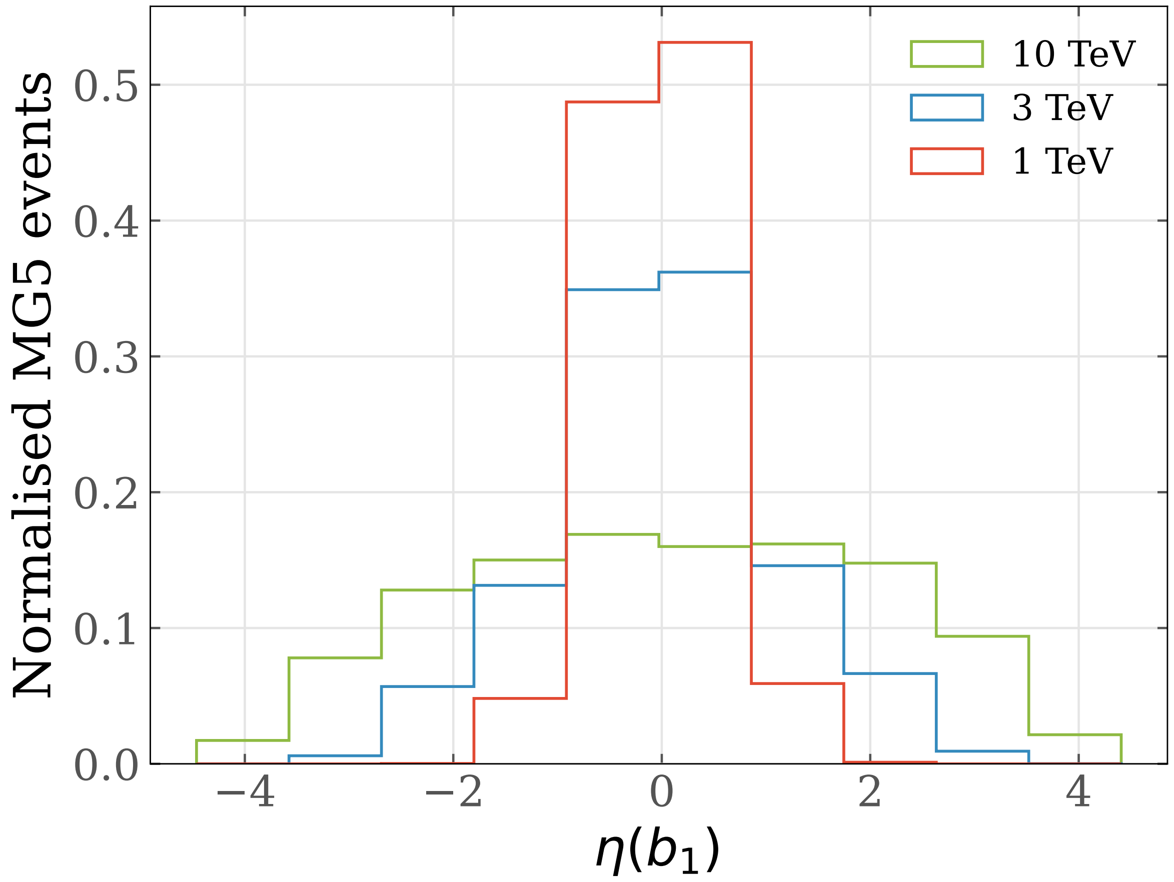


Lepton Colliders

- Complete picture of $(\kappa_3, \kappa_4) \to$ lepton colliders?
- Inclusive $\ell\ell \to HHH + X$ analysis with $H \to b\bar{b}$

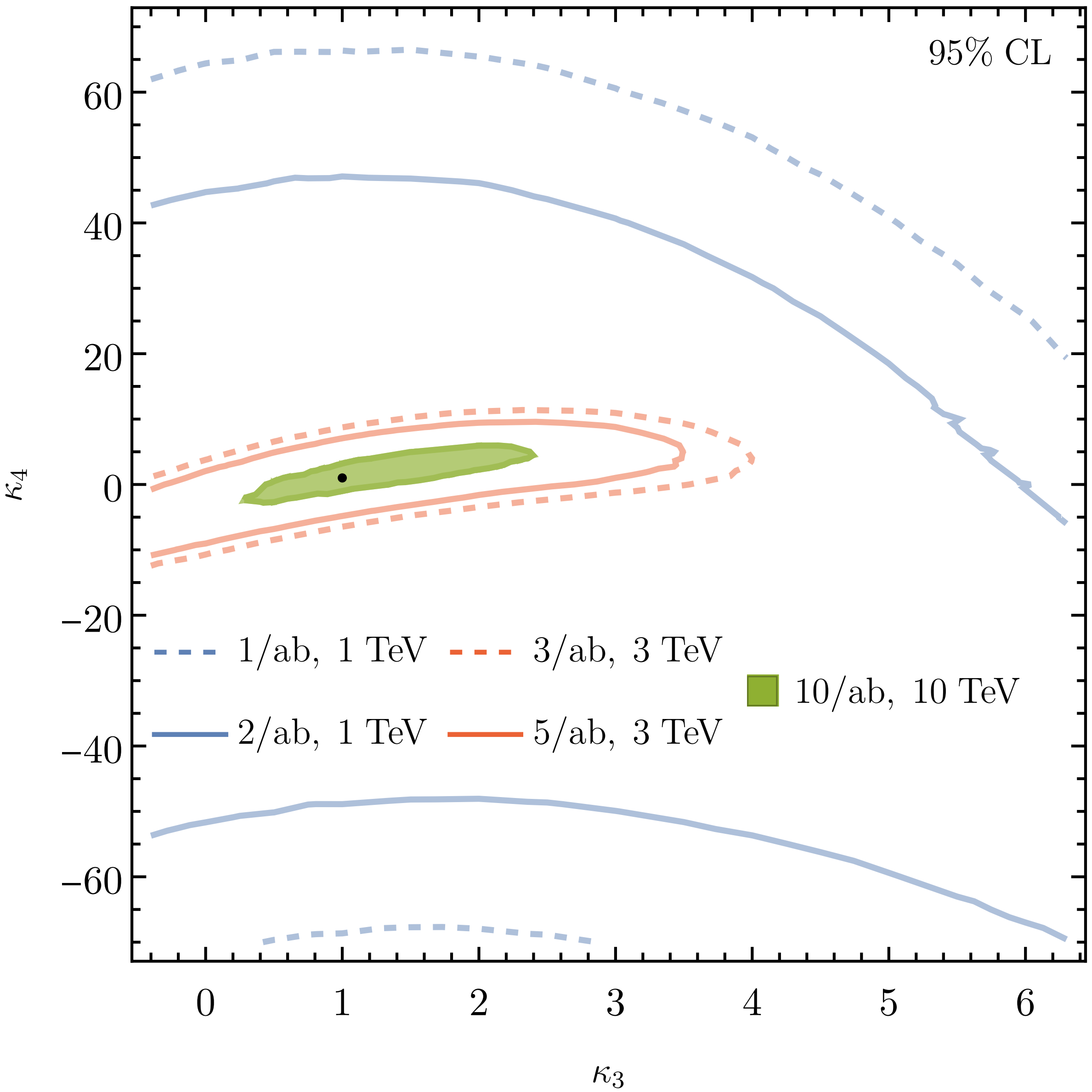
  - At least 5 tagged b-quarks with $p_T(b) > 30$ GeV
  - Tagging efficiency: $80\,\%$

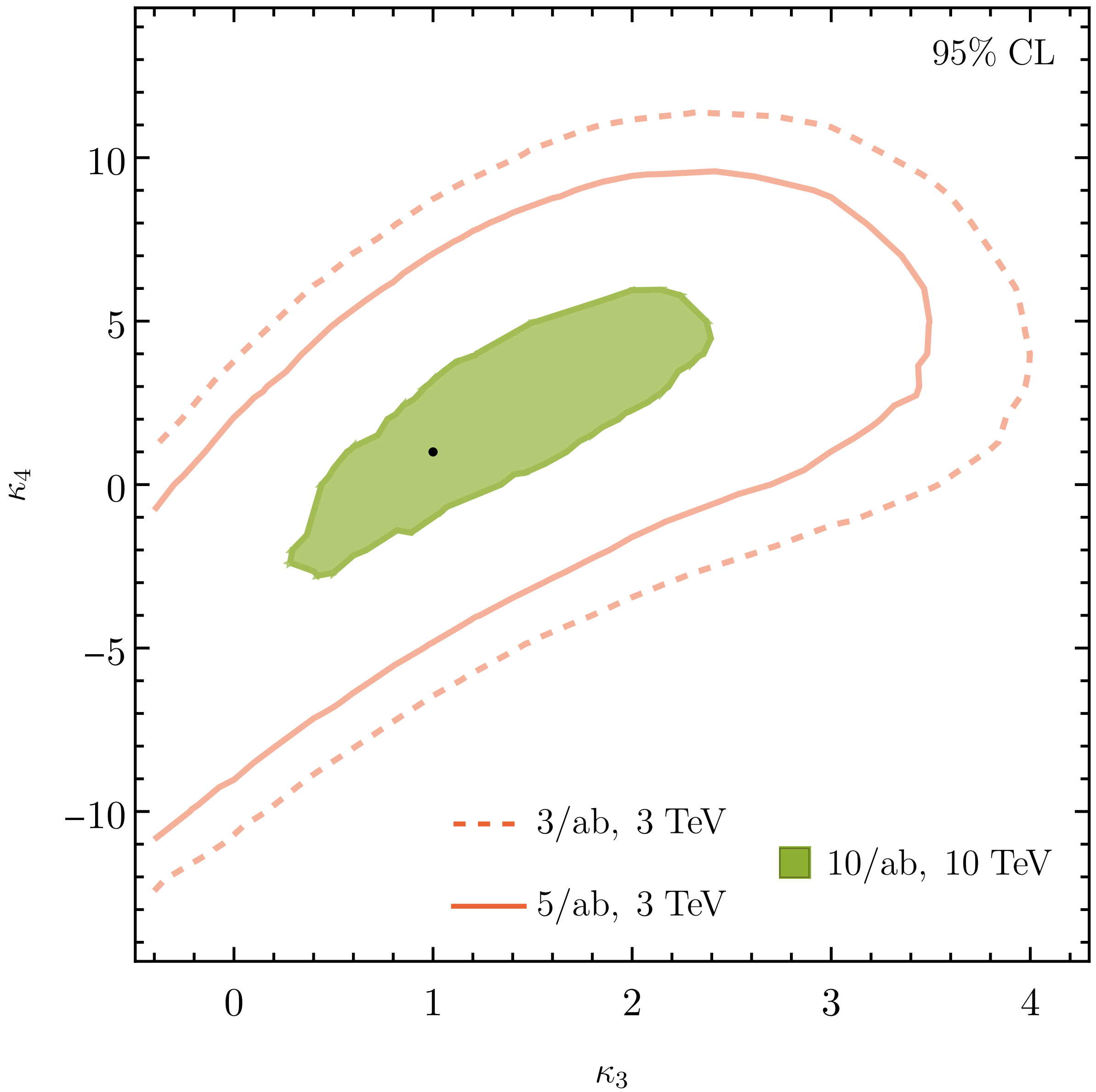
- **Important:** For high energies b-quarks are not only in the central part of detector → requires extended tagging capabilities
- Negligible background from other SM processes



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]

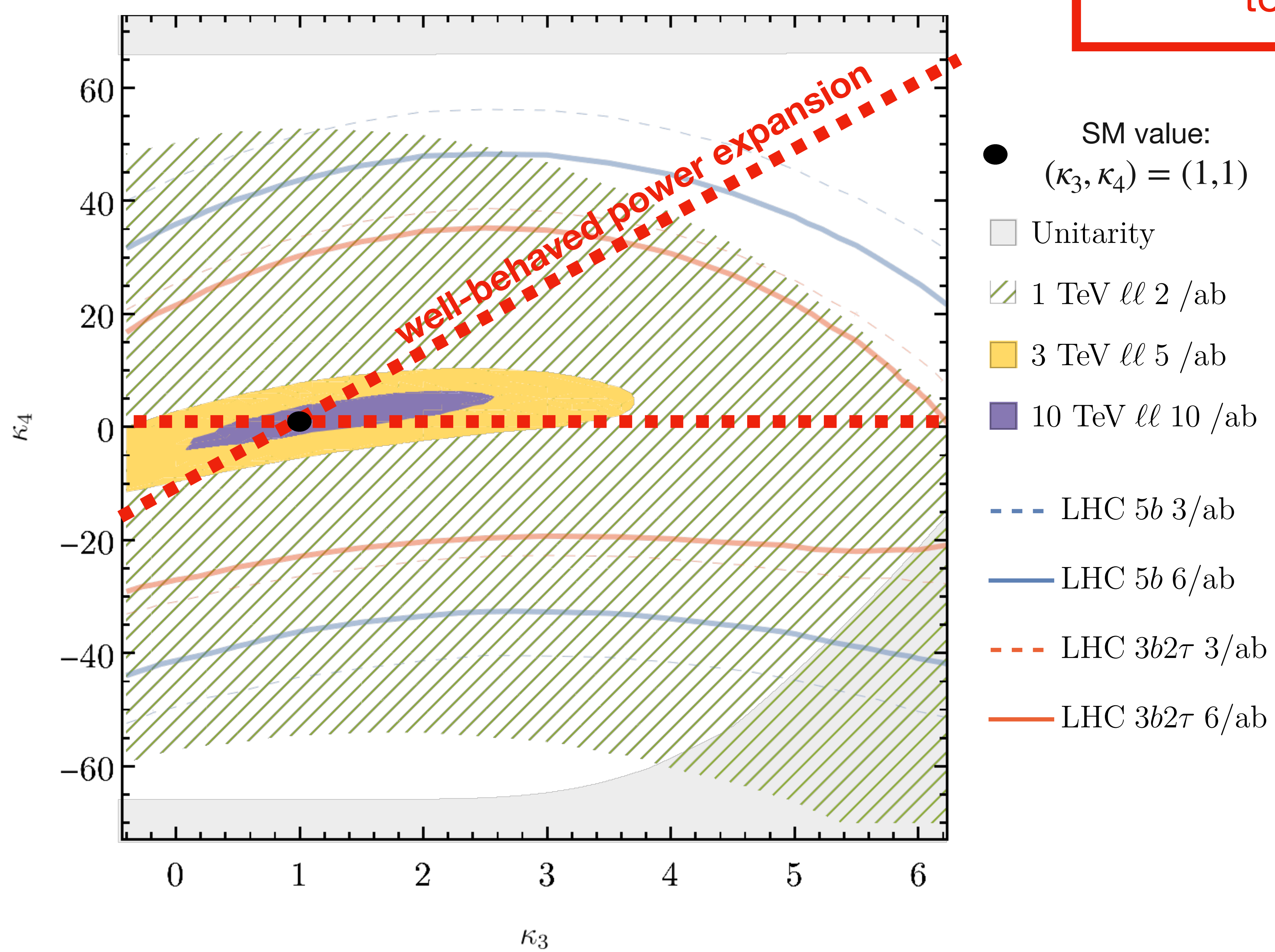




HL-LHC vs. future lepton colliders

- 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
  - HL-LHC will be able to probe interesting regions which could point to linear vs. non-linear prescriptions
  - 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!

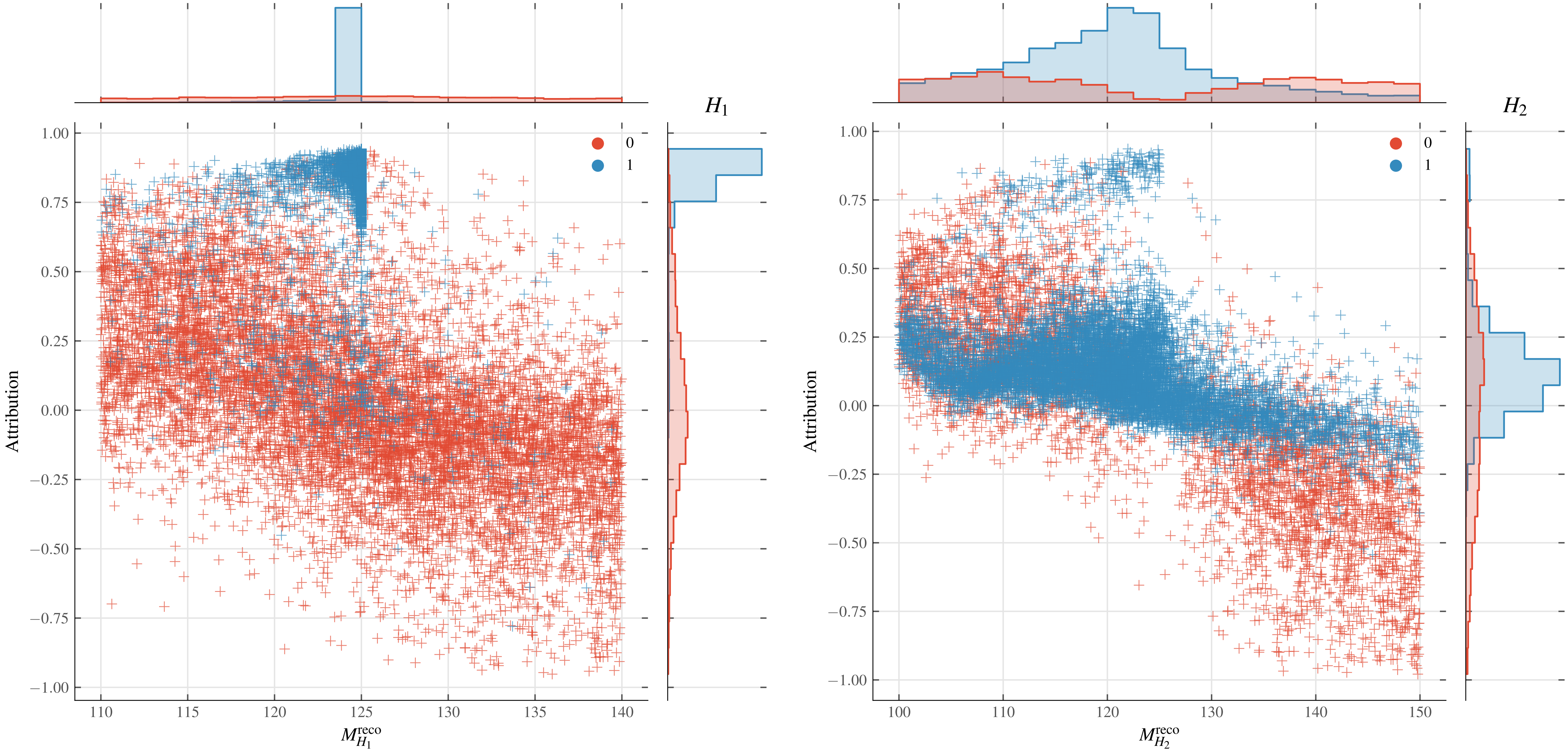
Backup: Interpretation axioms

- **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 $\rightarrow$ **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