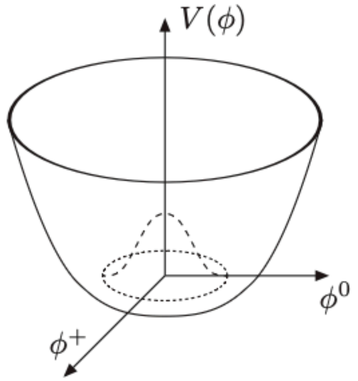
Nonresonant HHH6b: Boosted+Resolved Jet Assignment using Symmetry-Preserving Attention Networks

Haoyang “Billy” Li, Darius Chao, Kaitlyn White, Melissa Quinnan, Javier Duarte (UCSD),
Jovan Mitic (Belgrade),
Marko Stamenkovic, Greg Landsberg (Brown),
Caden Mikkelsen, Harvey Newman (Caltech),
Cristina Mantilla Suarez (Fermilab),
Alexander Shmakov, Michael Fenton, Daniel Whiteson, Pierre Baldi (UCI)

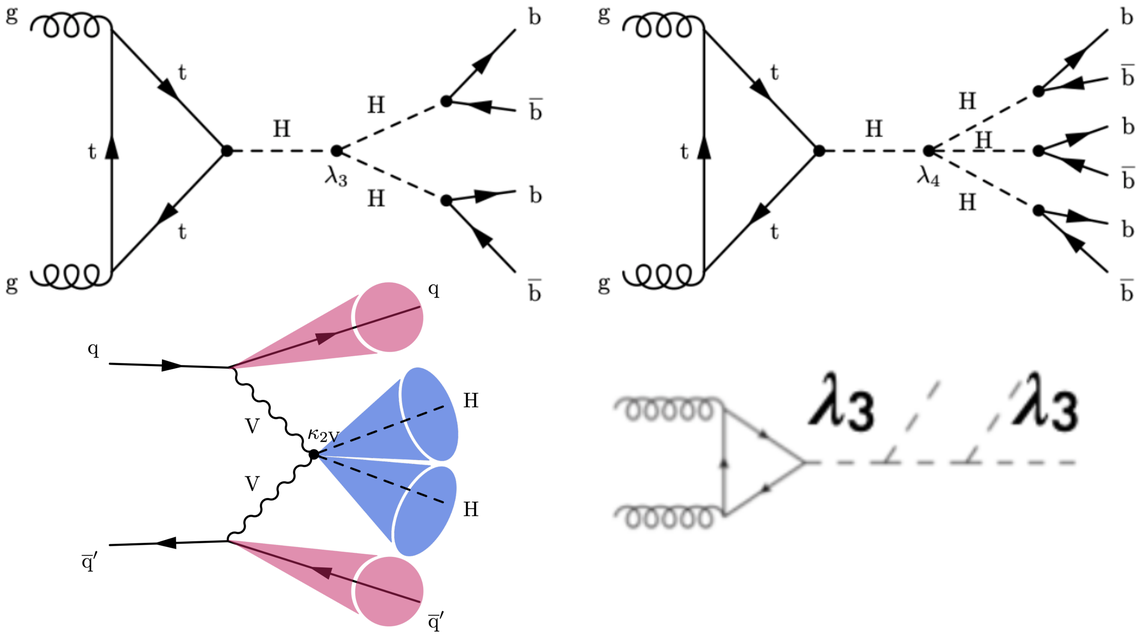
A3D3 Meeting, 2025

Di-higgs and Tri-Higgs Search



Motivation

- Probing the Higgs self-coupling possible through di-Higgs and triple Higgs measurements:
- Higgs self-coupling is pivotal in shaping the potential energy of the Higgs boson and studying the Electroweak Symmetry Breaking
- Higgs self-coupling is important for understanding early universe (electroweak baryogengesis) and stability of the vacuum
- κ_{2V} is sensitive to BSM physics



The EWSB potential:

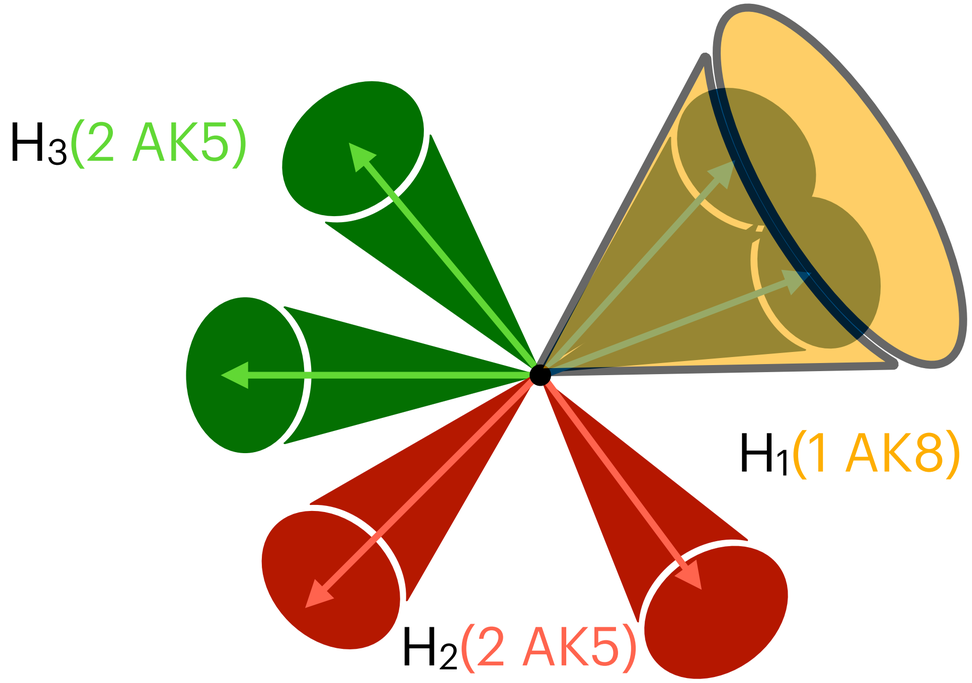
$$V(H) = \frac{1}{2!}\lambda_{HH}H^2 + \frac{1}{3!}\lambda_{HHH}H^3 + \frac{1}{4!}\lambda_{HHHH}H^4$$

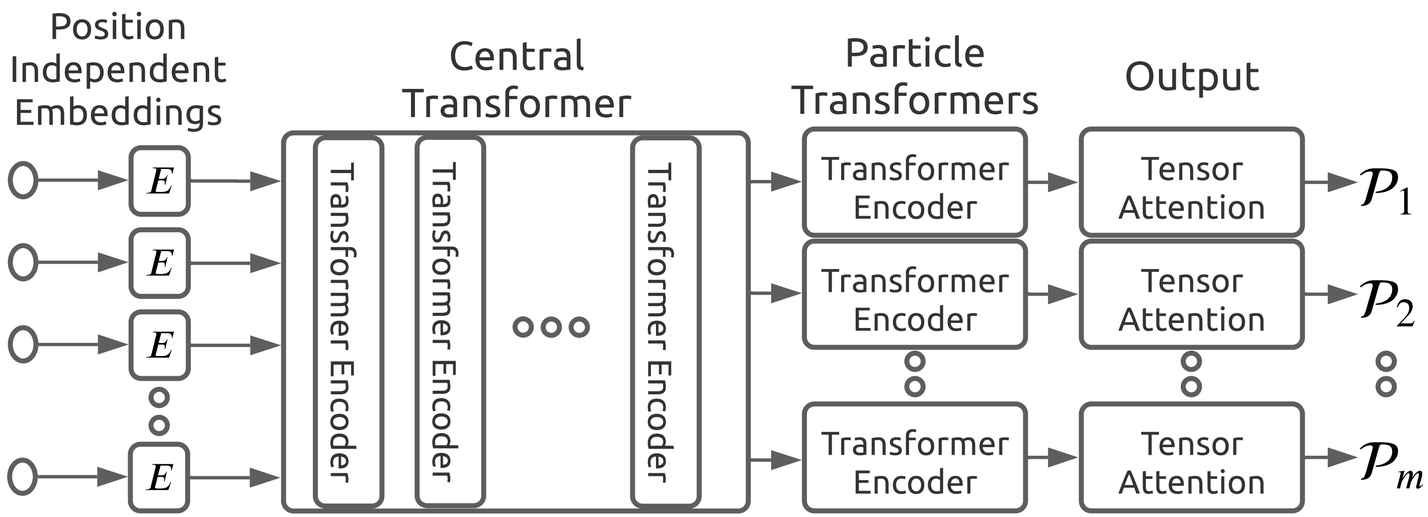
Trilinear coupling	$\lambda_{HHH} = 3\frac{M_H^2}{v}$	
Quartic coupling	$\lambda_{HHHH} = 3\frac{M_H^2}{v^2}$	

SPANet for Resolved HHH

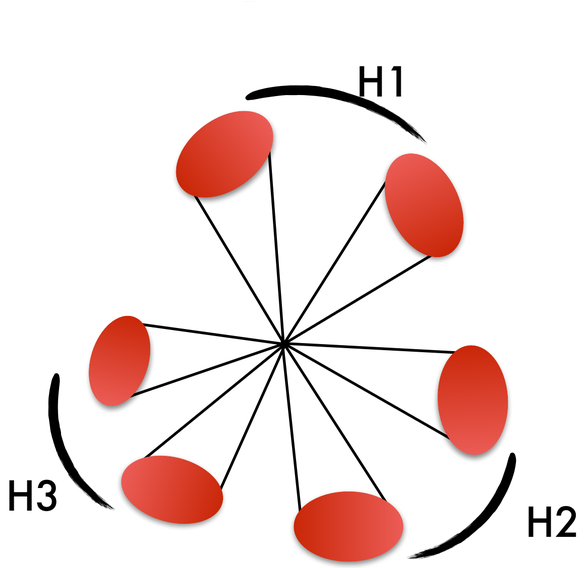
[1] arXiv:2010.09206
[2] arXiv:2106.03898

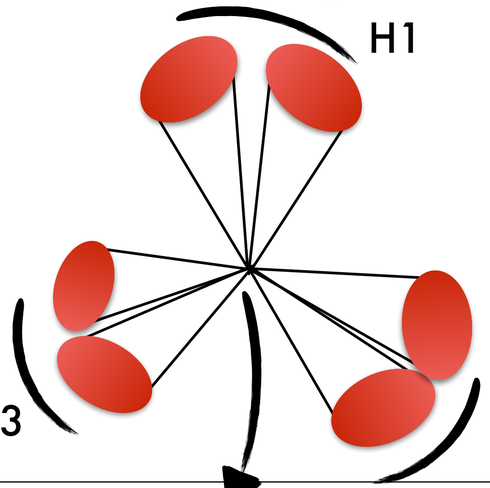
- HHH has a complex final state with many possible permutations and different types of (partial) reconstruction possible, including overlapping small- and large-radius jets
- Idea: use transformer to optimize reconstruction efficiency
  - Fully exploit event topology and kinematic correlations of jets to pair the 3 Higgs correctly
  - Existing approach (SPA-Net [1, 2]) works out of the box for fully-resolved case (6 small-radius b jets), would like to generalize to mixed resolved/boosted cases

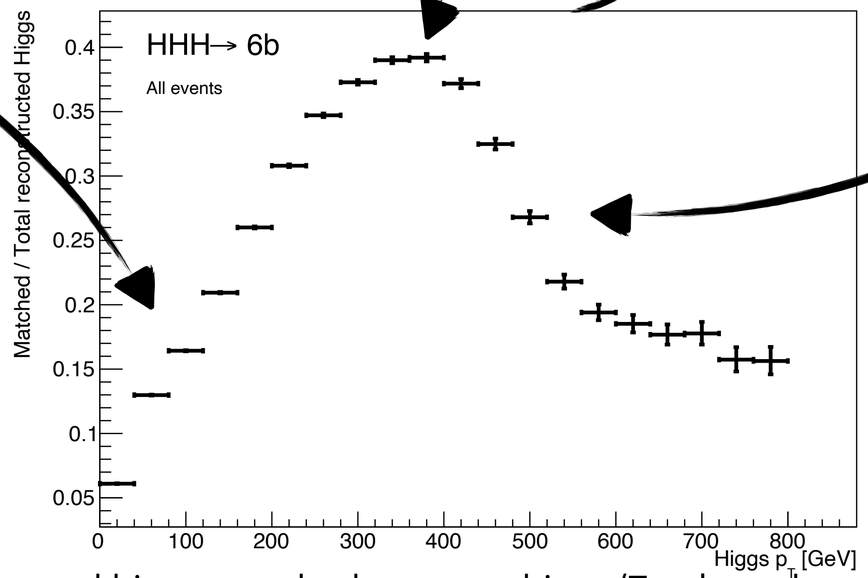


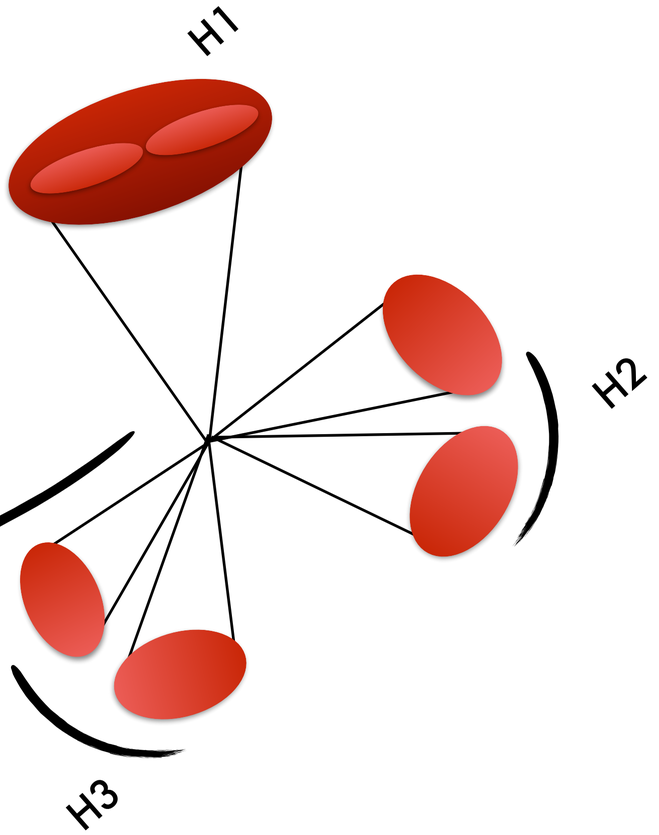


Dependence of Reco. Purity on Higgs Momentum







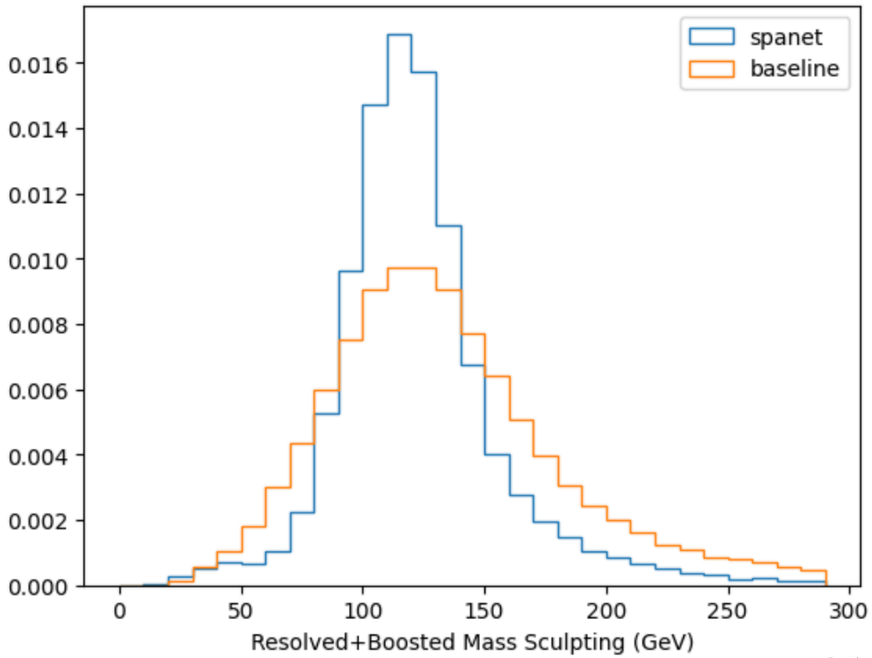


- Reconstruction Purity: Number of reconstructed higgs matched to target higgs/Total number of reconstructed higgs
- At low momentum, jets scattered around detector = complicated pairing
- At higher momentum, jets from Higgs boson more and more collimated, clearer correlation
- At very high momentum (pT > 400 GeV), reconstruction purity drops and Higgs reconstructed in AK8 jets

SPANet Mass Sculpting

From a preliminary HHH SPANet study

- Mass Sculpting: artificial creation of peaks in the background invariant mass distributions, can misleadingly appear as false signals of particles.
- We inferred SPANet and baseline to QCD backgrounds and plotted the outputs

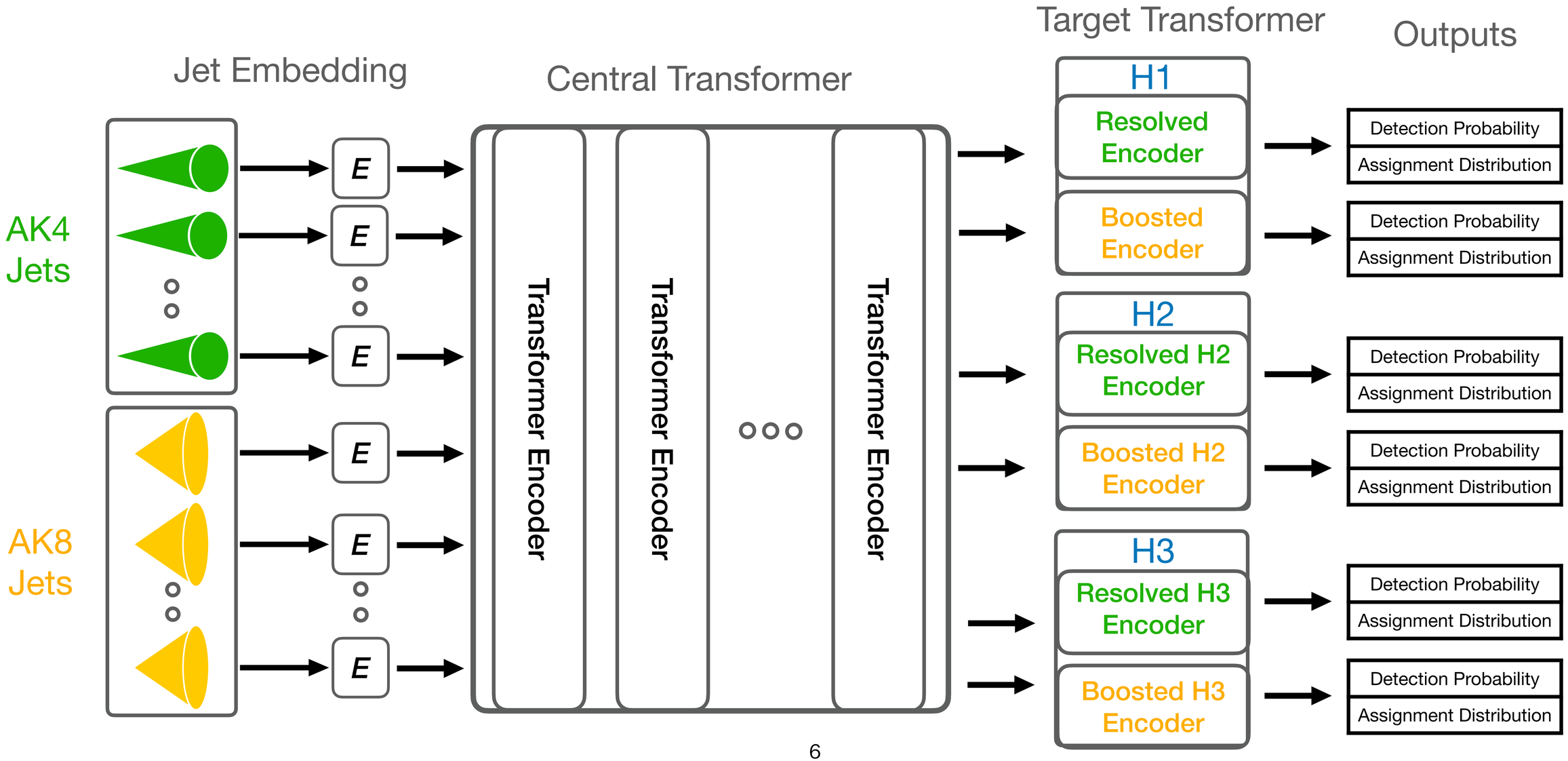


Preliminary Study of SPANet trained on SM Higgs ($m_H = 125\text{GeV}$)

Baseline is $\chi^2 + \text{BDT}$ (details in later slides)

Upgraded SPANet Architecture

Reconstructing Higgs bosons from both AK5 jets and AK8 jets

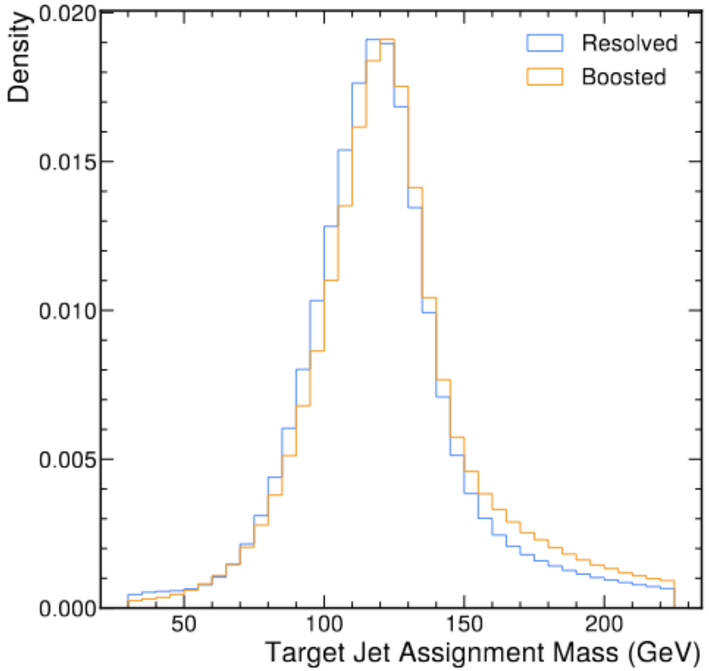


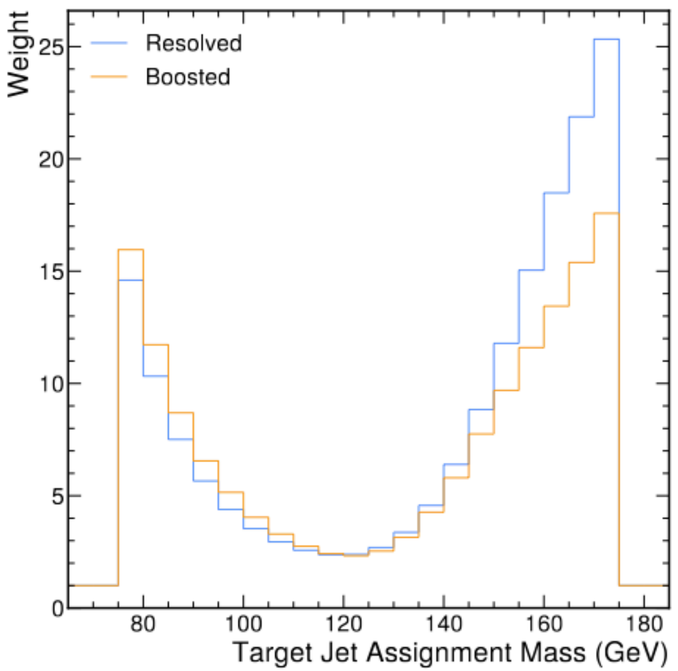
Target Higgs Boson Mass Reweighting

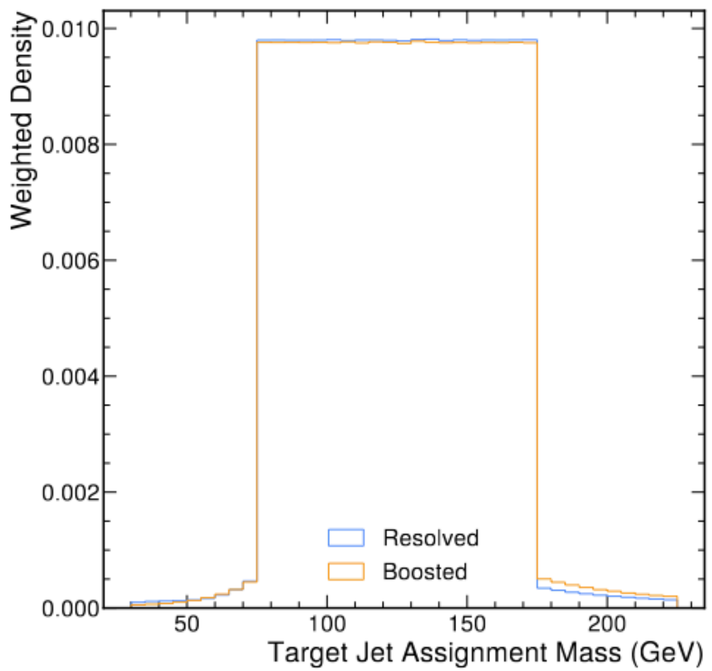
Calculate target weights and incorporate them into losses

$$w(m) = \begin{cases} \dfrac{p_{\text{uniform}}(m_i)}{p_{\text{mass}}(m_i)}, & 75 \le m_i \le m < m_{i+1} \le 175\,\text{GeV} \\ 1, & m \notin [75, 175]\,\text{GeV} \end{cases},$$

$$(\textit{Simplified}:) \quad \mathscr{L}_{\text{new}} = \sum_{i=1}^{N_{\text{target}}} w_i \mathscr{L}_i$$

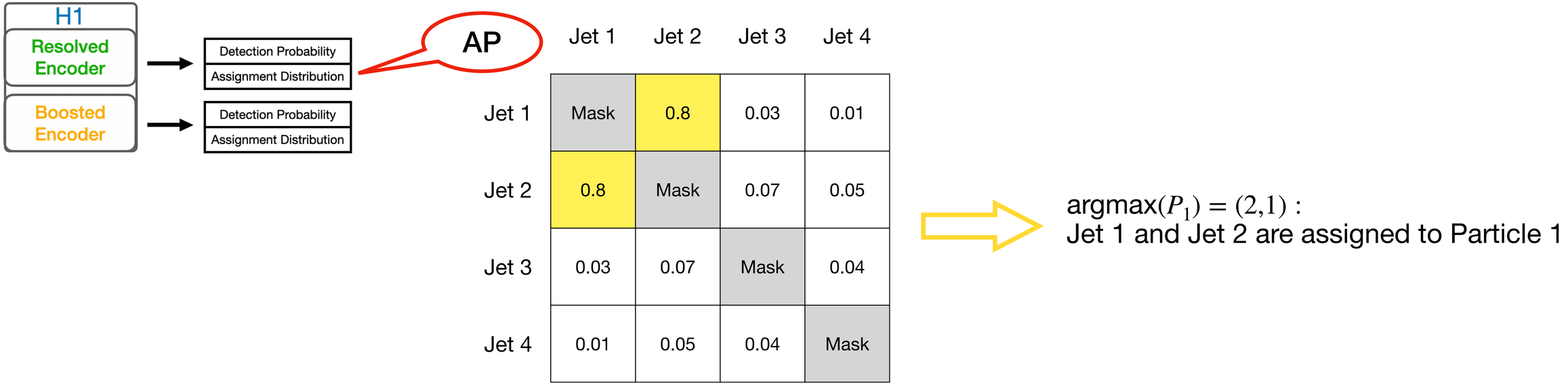






SPA-Net Assignment Output

- All valid permutations are evaluated using *symmetric loss*, and only the one with the minimal loss is selected.
- Output: Assignment Distribution x N_p candidates (resonant particles)
- Each entry of the tensor represents the likelihood that a set of objects reconstructs the target.
- The indices of the max tensor entry indicate which objects are to be assigned to the target



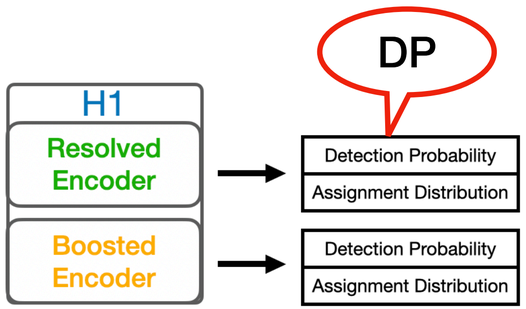
SPANet Post-processing and Overlap Removal

Prioritizing Reconstructions from AK8 Jets

for event in all_events:

for reconstruction_topology in [boosted, resolved]:

1. Calculate the most likely number ($\hat{N}_H$) of Higgs boson by DP (detection probability).



$$p(N_{\mathrm{H}}) = \begin{cases} (1-\mathrm{DP}_1)(1-\mathrm{DP}_2)(1-\mathrm{DP}_3) & N_{\mathrm{H}}=0 \\ \mathrm{DP}_1(1-\mathrm{DP}_2)(1-\mathrm{DP}_3)+\mathrm{DP}_2(1-\mathrm{DP}_1)(1-\mathrm{DP}_3) & N_{\mathrm{H}}=1 \\ \quad +\mathrm{DP}_3(1-\mathrm{DP}_1)(1-\mathrm{DP}_2) & \\ \mathrm{DP}_1\mathrm{DP}_2(1-\mathrm{DP}_3)+\mathrm{DP}_2\mathrm{DP}_3(1-\mathrm{DP}_1)+\mathrm{DP}_3\mathrm{DP}_1(1-\mathrm{DP}_2) & N_{\mathrm{H}}=2 \\ \mathrm{DP}_1\mathrm{DP}_2\mathrm{DP}_3 & N_{\mathrm{H}}=3 \end{cases}.$$

$$\hat{N}_{\mathrm{H}} = \arg\max_{N_{\mathrm{H}}} p(N_{\mathrm{H}}),$$

2. Reconstruct $\hat{N}_H$ targets by picking the top $\hat{N}_H$ assignment by DP x AP

Remove the resolved reco Higgs that have any AK5 constitute overlapped with a boosted reco Higgs ($\Delta R \leq 0.5$)

HHH — Baseline Methods

- Resolved Baseline: $m_H = 125\,\mathrm{GeV}$, pick the permutation minimizing χ^2
  - Note: higher efficiency, worse background mass sculpting

$$\chi^2 = (m_{b_1 b_2} - m_H)^2 + (m_{b_3 b_4} - m_H)^2 + (m_{b_5 b_6} - m_H)^2$$

- Boosted Baseline: A BDT as AK8-jet tagger
  - Input fearures:
    - pT,
    - η,
    - mass m,
    - soft-drop mass mSD,
    - fraction of charged energy,
    - number of charged particles,
    - and the N -subjettiness jet substructure variables τ32 and τ21
  - Output:
    - Probability that an AK8 jet is a Hbb jet vs. QCD jet

Baseline script

Dataset & Input Features

Madgraph model provided by
A. Papaefstathiou, T. Robens, G. Tetlalmatzi-Xolocotzi
JHEP 05 (2021) 193

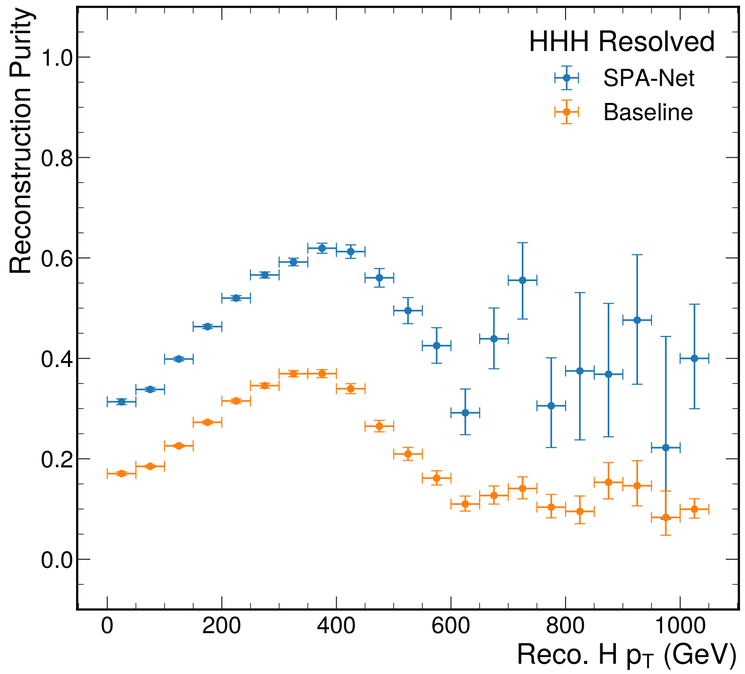
- Using 14 TeV $pp \rightarrow HHH \rightarrow 6b$ events simulated with MadGraph+Pythia8+Delphes:
  - ~10M events for training+validation+testing.
  - Additionally, ~1M $pp \rightarrow HH \rightarrow 4b$ events
- Truth matching condition:
  - Resolved Higgs: Gen b-quark from Higgs boson decay is within $\Delta R \leq 0.5$ of AK5 jet. Added hadron “b” flavor requirement on AK5 jet.
  - Boosted Higgs: Gen b-quarks from Higgs boson decay and the Higgs boson is within $\Delta R \leq 0.8$ of AK8 jet.
- A resolved Higgs boson is “reconstructible” if two b quark daughters match to AK5 jets
- A boosted Higgs boson is “reconstructible” if two b quark daughters and the mother Higgs match to a AK8 jet.
- Up to 10 AK5 and 3 AK8 Jets are considered per event (ranked by p_{T})
- Input jet features:
  - AK4 jets: p_{T} (log-normalized), η (normalized), $\sin\phi$, $\cos\phi$, boolean b-tag score, and jet mass (normalized)
  - AK8 jets: p_{T} (log-normalized), η (normalized), $\sin\phi$, $\cos\phi$, and jet mass (normalized)

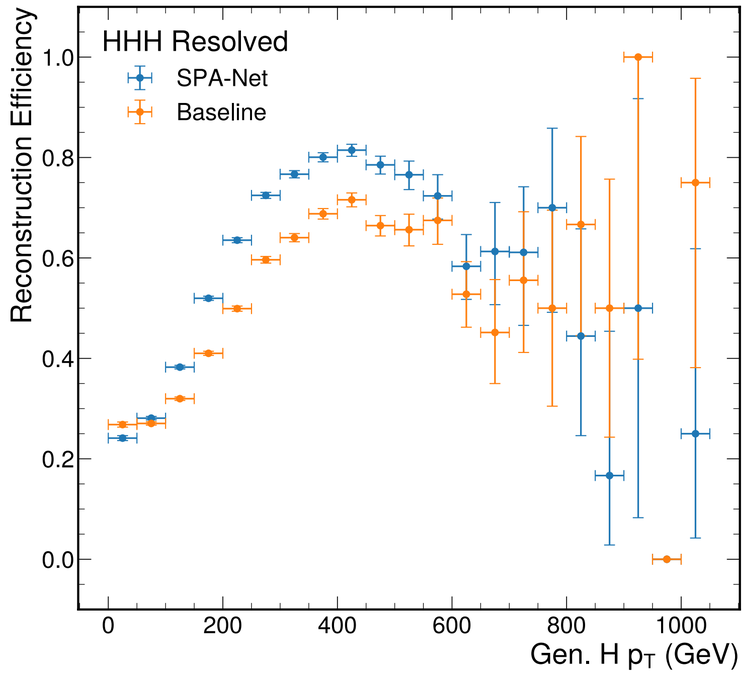
Comparing Resolved SPANet with Baseline on HHH6b

$$\text{Reconstruction Purity} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Reco Higgs}}$$

$$\text{Matching Efficiency} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Target Higgs}}$$

Matching condition: Target and reco have the same jet assignment





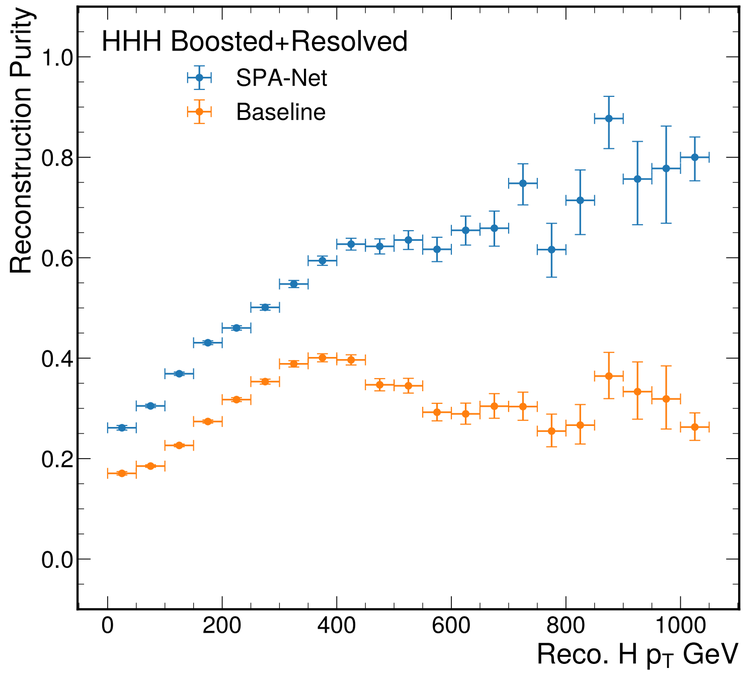
- SPANet outperforms baseline's in most pT bin of each metric.

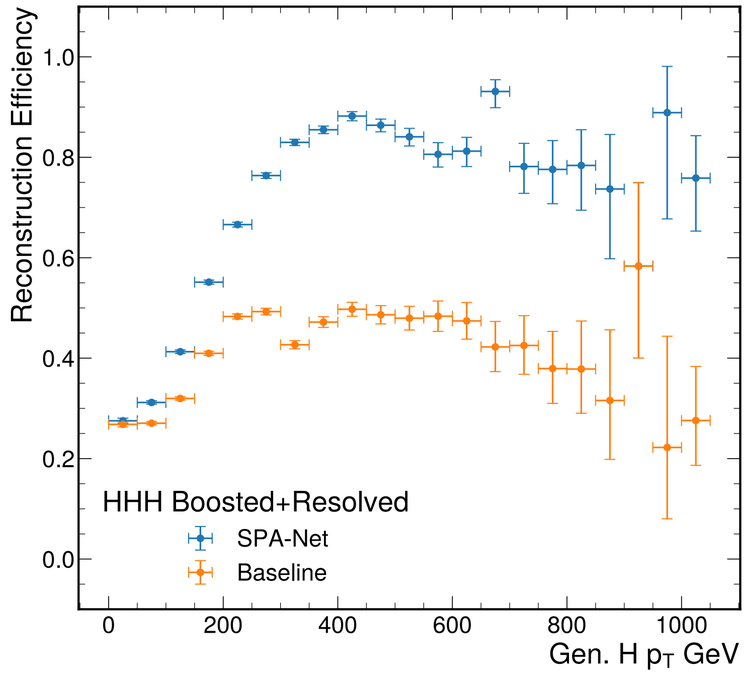
SPANet HHH6b Performance Improves after Including AK8 Jets

$$\text{Reconstruction Purity} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Reco Higgs}}$$

$$\text{Matching Efficiency} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Target Higgs}}$$

Matching condition: Target and reco have the same jet assignment



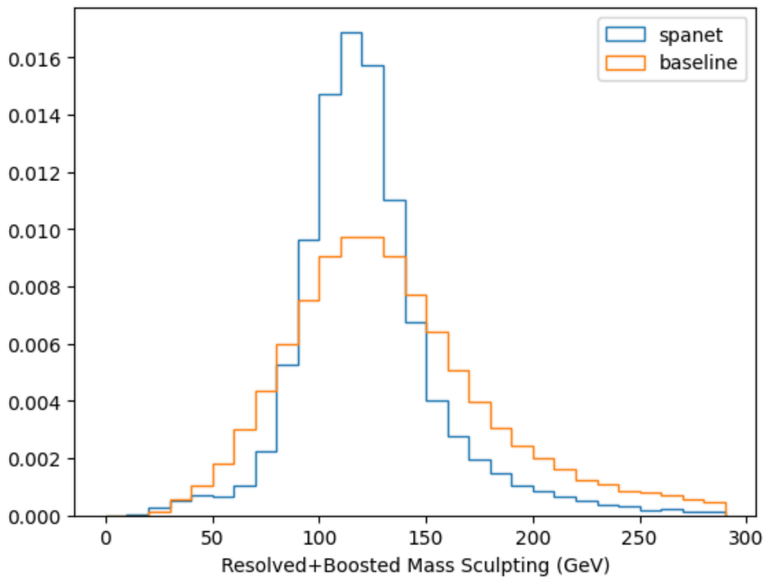


- SPANet performance outperform baseline's in any pT bin of each metric.
- SPANet performance is overall better than the baseline method's performance.

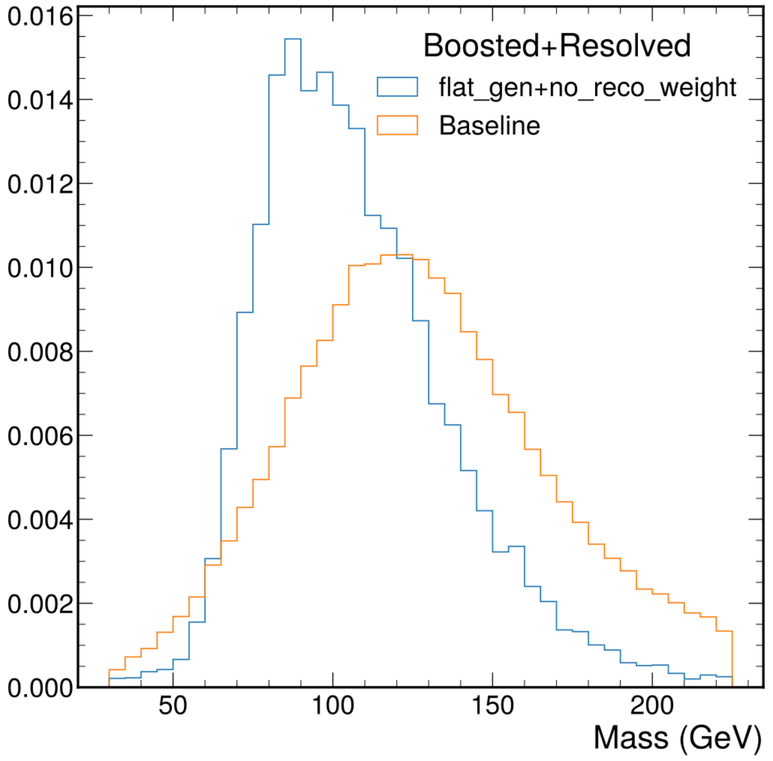
SPANet HHH6b Mass Sculpting

- Mass Sculpting: artificial creation of peaks in the background invariant mass distributions, can misleadingly appear as false signals of particles.
- We inferred SPANet and baseline to QCD backgrounds and plotted the outputs
- Got a less peak-like distribution than the baseline's

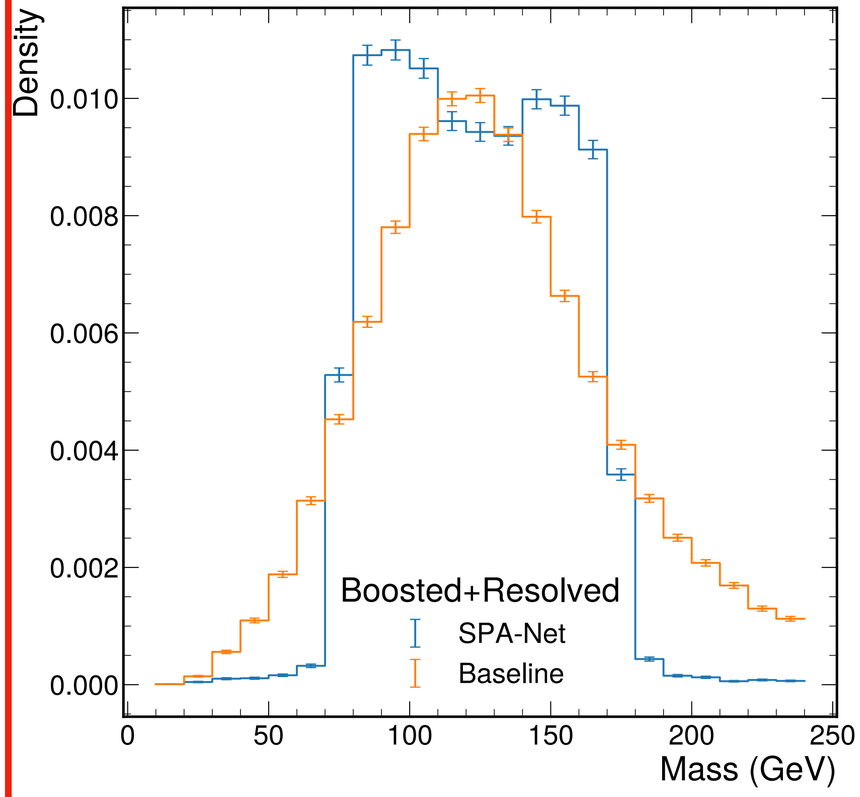
Trained on SM H



Trained on a flat gen. H distribution



Trained on a flat reco. H distribution



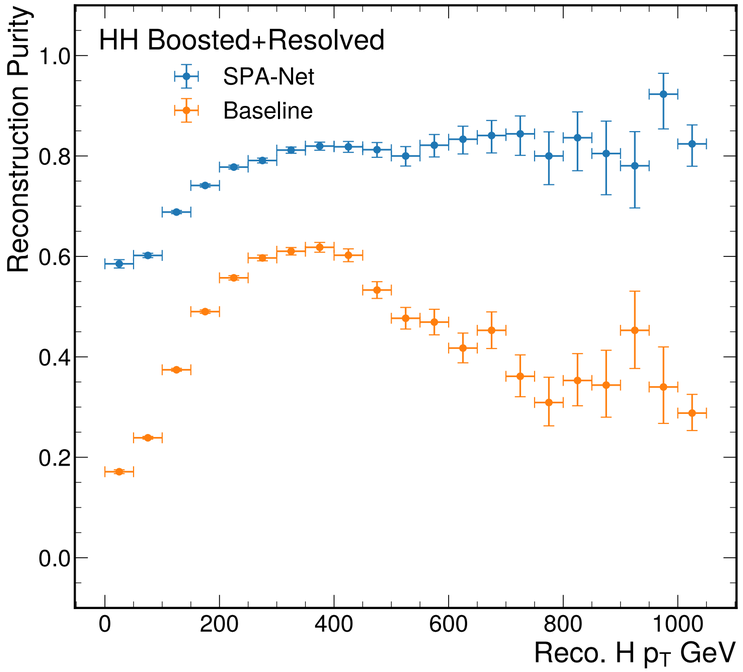
Gen Higgs simulated at m=[80, 90, … 170]

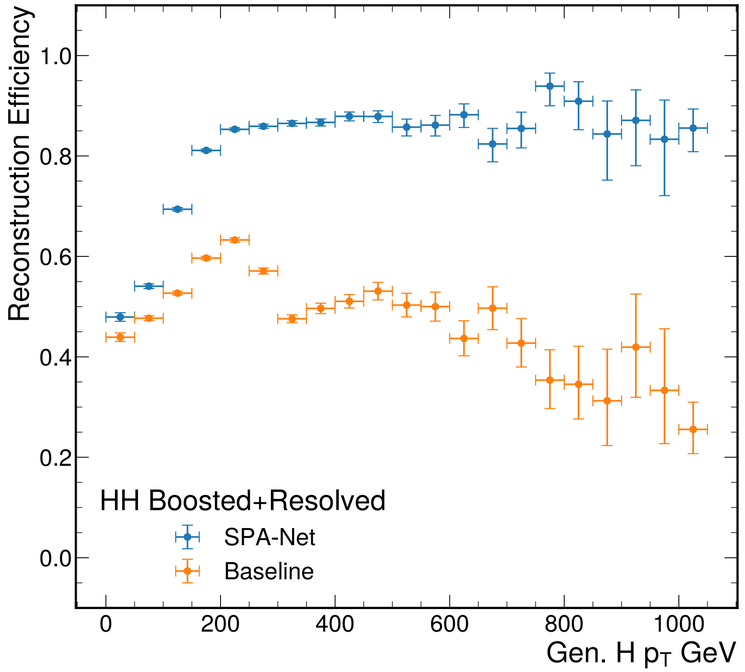
Comparing SPANet with Baseline on HH4b

$$\text{Reconstruction Purity} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Reco Higgs}}$$

$$\text{Matching Efficiency} = \frac{\text{Number of Reco Higgs matched to Target Higgs}}{\text{Number of Target Higgs}}$$

Matching condition: Target and reco have the same jet assignment





Similar to HHH, a comparison of boosted+resolved SPA-Net training with the χ2+BDT baseline shows that SPA-Net consistently outperforms the baseline across both evaluation metrics.

Summary

- SPANet: A transformer model for particle reconstruction
- The upgraded SPANet shows better performance than chi2-BDT baseline in our study of HHH6b and HH4b.
    - Unique algorithm to pair fully resolved, semi-boosted, fully boosted simultaneously
    - Improved performance in high-pT regions by including the boosted higgs
- SPANet trained with our crafted weights has less pronounced peak-like distribution compared to the baseline
- SPANet can lead to better reconstruction efficiency and therefore faster multi-Higgs discovery as well as better determination of fundamental parameters in the Higgs sector

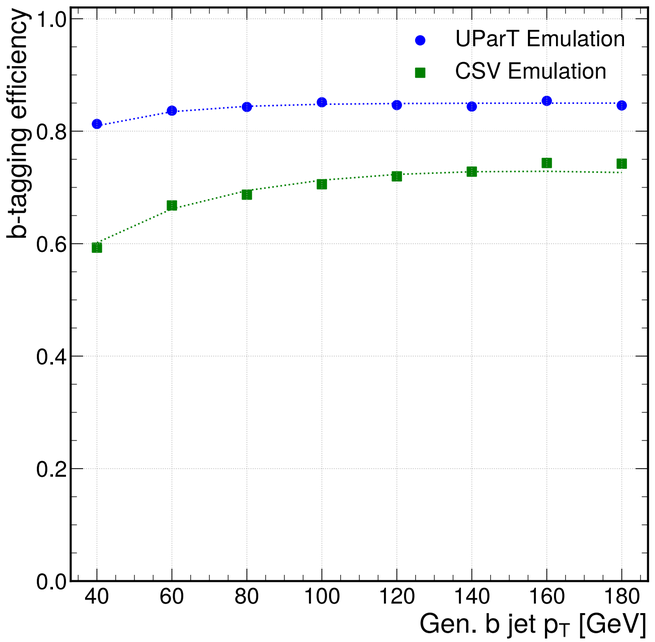
Submitted to JHEP, on arxiv (2412.03819)

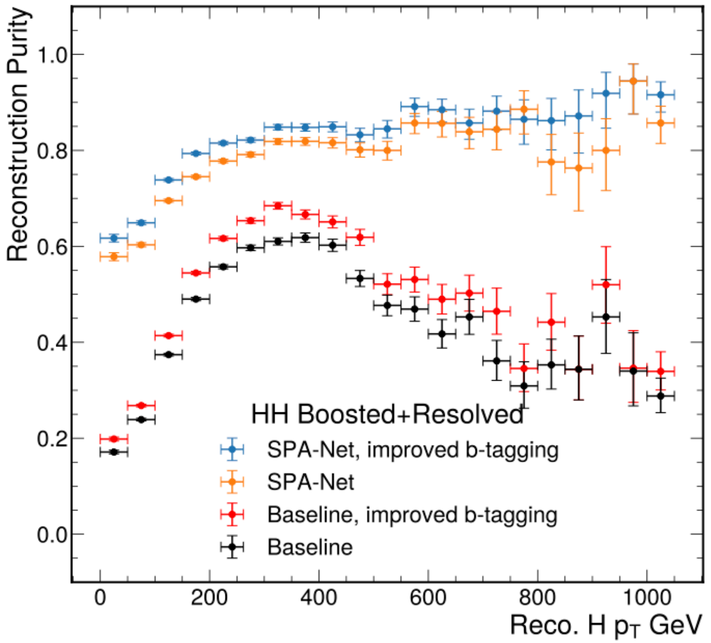
Back-up

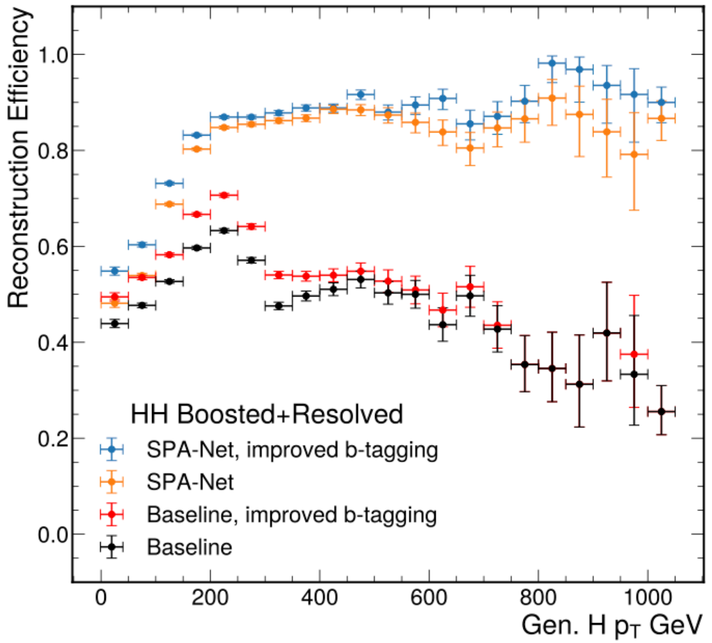
Study on Improved b-tagging

From the Combined Secondary Vertex (CSV) to Unified Particle Transformer (UParT)

- The b-tagging efficiency improves from approximately 70% to 85%.



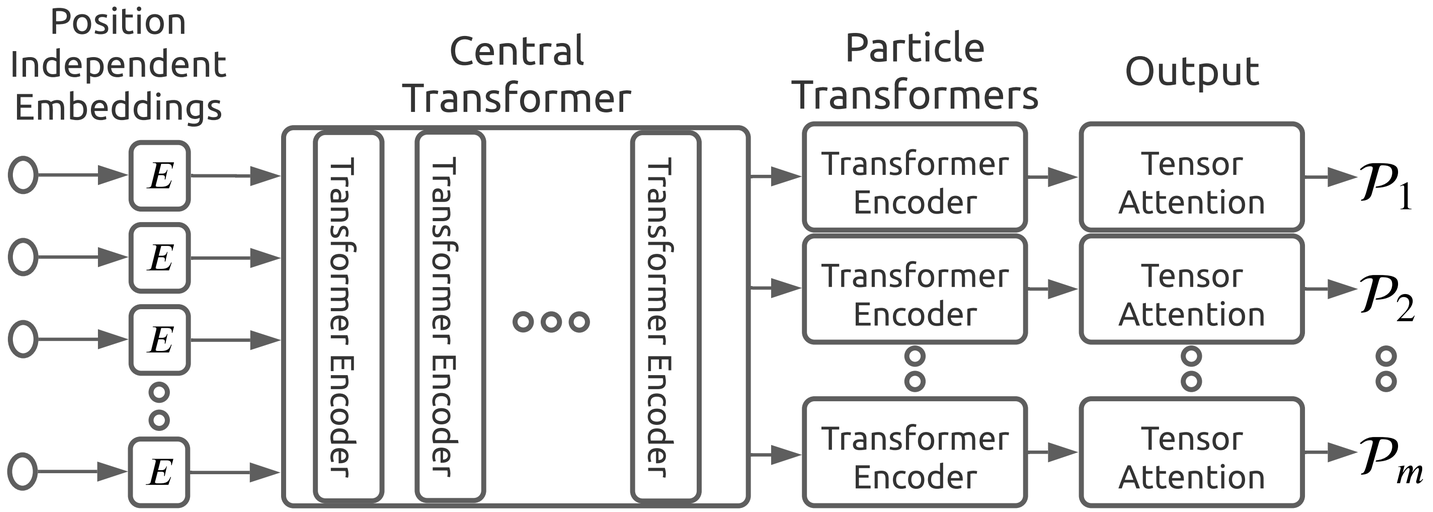




- Both SPA-Net and baseline benefit from the improved b-tagging
- The relative performance gain of SPA-Net remains stable across different b-tagging configurations.

The Original Symmetry Preserving Attention Networks (SPA-Net)

- Consider all valid permutations using *symmetric tensor attention*
  - Resonance particle p (e.g., Higgs) is associated with k_p partons (e.g., 2 b quarks); maximum of N reconstructed jets (e.g., 10)
  - Input: matrix of transformer-encoded jets $X_p \in \mathbb{R}^{N \times D}$
  - Output: rank-k_p tensor $\mathscr{P}_p \in \mathbb{R}^{N \times N \times \cdots \times N}$ the joint distribution over k_p-jet assignments
    - $\sum \mathscr{P}_p = 1$
    - Valid solutions => $\mathrm{diag}(\mathscr{P}_p) = 0$



Training on Resolved+Boosted Dataset: Event Selection

SPANet configuration:

- 6 reconstruction targets: Resolved Higgs 1,2&3, and Boosted Higgs 1,2&3.
- Tell SPANet boosted Higgs should be reconstructed from AK8 jets and specify all valid permutations

Input: >= 6 jets with pT > 20 GeV

- When existing, AK8 jets with pT > 250 GeV

Dataset details

Sample	N_{events}	Training Fraction	Validation Fraction	Testing Fraction
HHH, $m_{\text{H}} = 125\,\text{GeV}$	855,254	90.25%	4.75%	5%
HHH, $m_{\text{H}} \in \{120, 122.5, 125, 127.5, 130\}\,\text{GeV}$	9,580,718	95%	5%	—
HH, $m_{\text{H}} = 125\,\text{GeV}$	1,209,807	90.25%	4.75%	5%
QCD multijet	127,876	80%	—	20%

we doubled the SM HHH dataset and increased the SM HHH training+validation dataset by simulating additional nonresonant HHH events with Higgs boson masses set to 120, 122.5, 127.5, and 130 GeV. Each additional mass point contains the same number of events with the SM HHH training and validation dataset before pre-selection. These datasets were combined prior to training the boosted+resolved HHH SPA-Net models. This approach aims to produce a flatter Higgs boson mass distribution and reduce mass sculpting effects.

Event Configuration of HHH $\rightarrow (b\bar{b})(b\bar{b})(b\bar{b})$

- Specify a list of resonant particles and their daughters

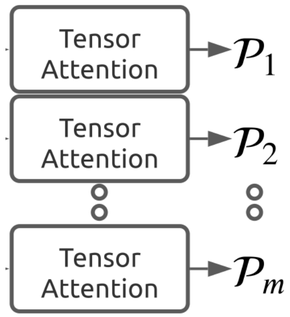
```
EVENT:
  h1:
    - b1: Jets
    - b2: Jets
  h2:
    - b1: Jets
    - b2: Jets
  h3:
    - b1: Jets
    - b2: Jets
```

- Provide a list of permutations to tell SPANet which particles are of the same kind.

```
PERMUTATIONS:
    EVENT:
      - [ h1, h2, h3 ]
    h1:
      - [ b1, b2 ]
    h2:
      - [ b1, b2 ]
    h3:
      - [ b1, b2 ]
```

apply each permutation to output

Output



[P1, P2, P3]
[P1, P3, P2]
[P2, P1, P3]
[P2, P3, P1]
[P3, P1, P2]
[P3, P2, P1]

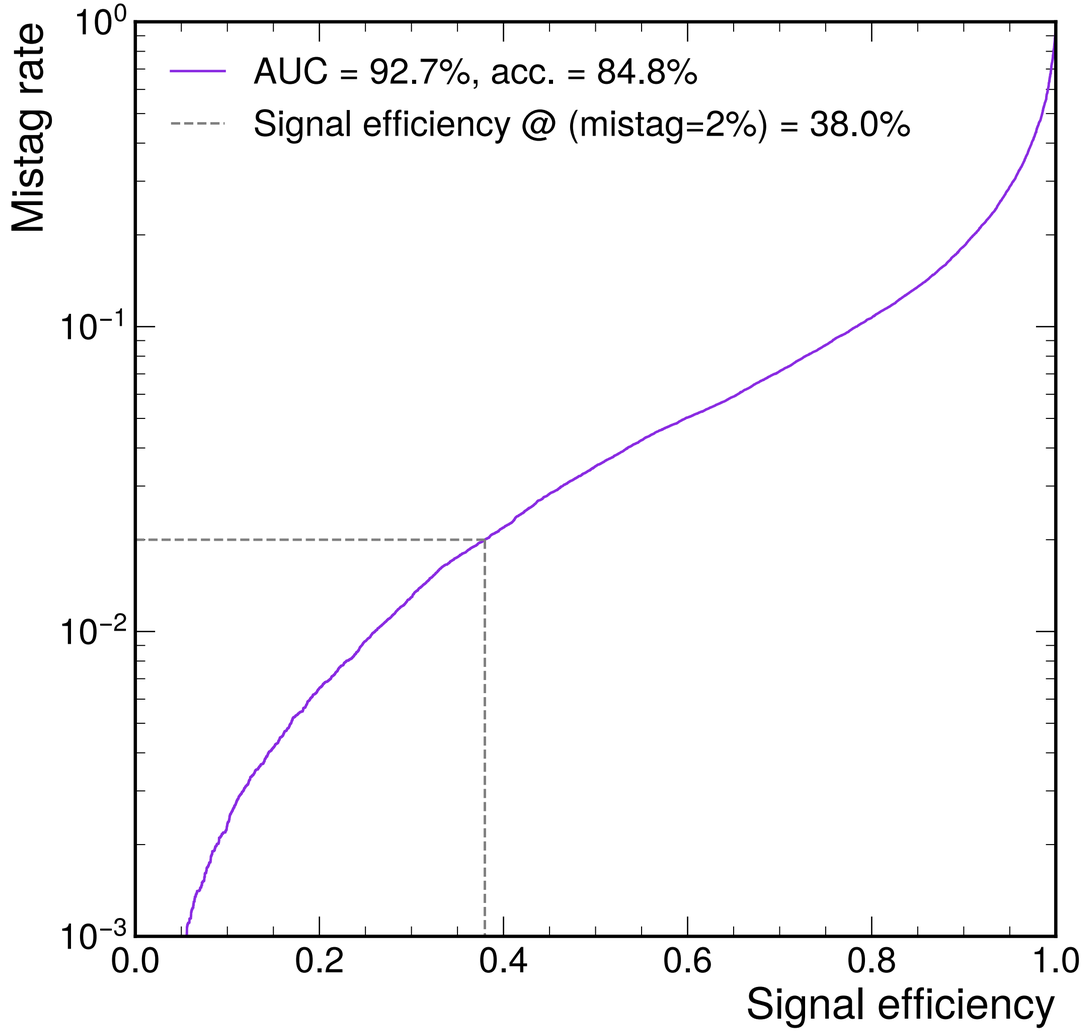
Calculate loss for each permutation

Targets: [T1, T2, T3]

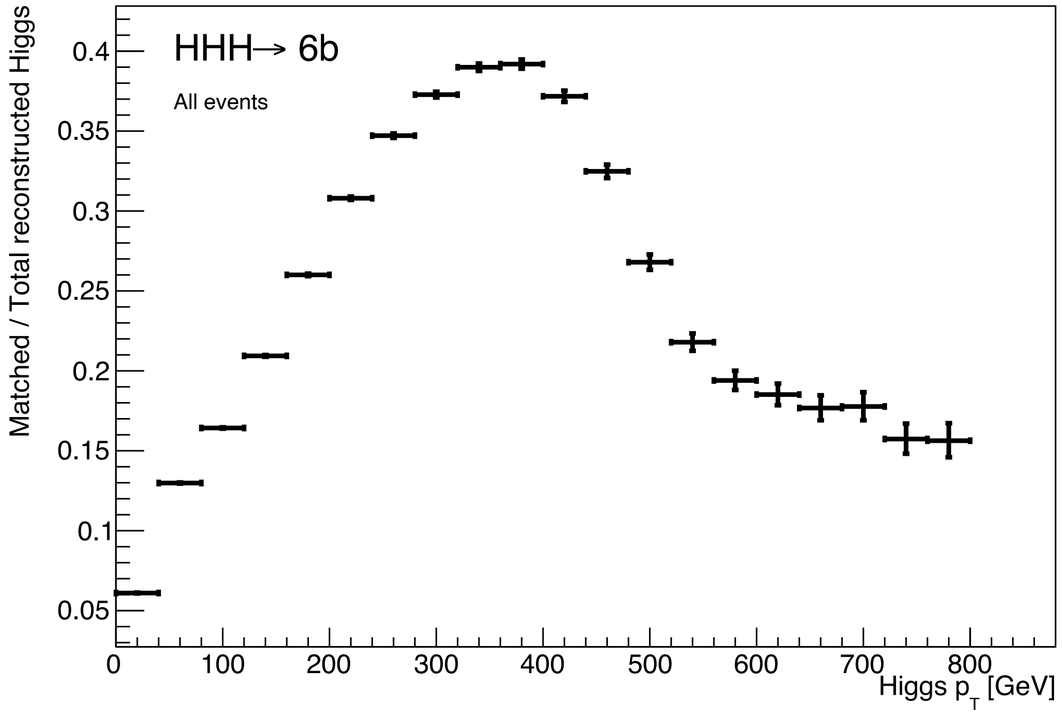
loss_1
loss_2
loss_3
loss_4
loss_5
loss_6

Choose the minimum to backpropagate and Update weights

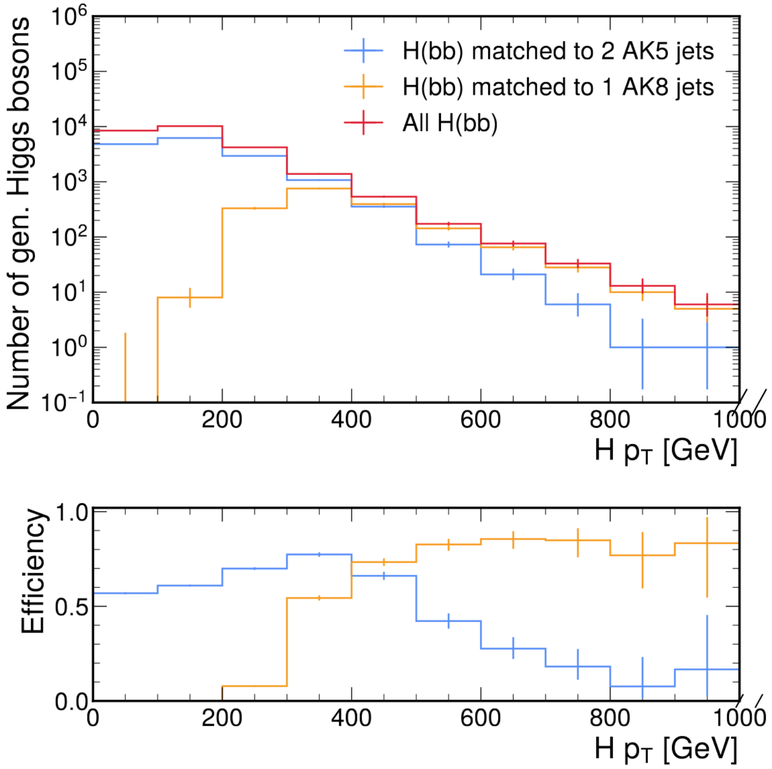
BDT Boosted Baseline



Differential matching efficiency



Truth level: Higgs reconstructed in 2 AK5 vs 1 AK8



- At 400 GeV, Higgs more likely to be reconstructed in 1 AK8 than 2 AK5
- Optimal performance: generalize approach to both boosted + resolved topologies

Target Higgs Boson Mass Reweighting

Modified losses and validation metrics

$$\mathscr{L}_{\text{assignment}} = \min_{\sigma \in G_t} \sum_{i=1}^{N_{\text{target}}} \mathscr{M}_{\sigma(i)} \text{CCE}(P_i, T_{\sigma(i)})$$

$$\mathscr{L}_{\text{detection}} = \min_{\sigma \in G_t} \sum_{i=1}^{N_{\text{target}}} \text{BCE}(\text{DP}_i, \mathscr{M}_{\sigma(i)})$$

$$\mathscr{L} = \alpha_0 \mathscr{L}_{\text{assignment}} + \alpha_1 \mathscr{L}_{\text{detection}}$$

where CCE represents the categorical cross-entropy loss, G_t denotes the permutation group of all targets, restricted to permutations within the same reconstruction topology, σ is an element of the permutation group, N_{target} represents the maximum number of the targets that the SPA-Net model can output, P_i is the ith predicted jet assignment, $T_{\sigma(i)}$ is the ith target jet assignment under permutation σ, and $\mathcal{M}_{\sigma(i)}$ is the target mask associated with $T_{\sigma(i)}$. The assignment outputs are trained only on examples in which the event contains all detector objects necessary for a correct target assignment, i.e. the target is *reconstructible*.

$$\mathscr{L}_{\text{assignment}} = \min_{\sigma \in G_t} \sum_{i=1}^{N_{\text{target}}} w_{\sigma(i)} \mathscr{M}_{\sigma(i)} \text{CCE}(P_i, T_{\sigma(i)}),$$

where, in contrast to Eq. (3.2), the assignment loss term for each target is scaled by the corresponding target-wise weight $w_{\sigma(i)}$. Similarly, the detection loss is modified by multiplying the loss term for each target in the event by $w_{\sigma(i)}$.

Detailed Model Configuration

- Many hyperparameters to tune!
- Here is a short summary

Parameter	**Benchmark Models**			
	HHH Boosted+Resolved	HHH Resolved	HH Boosted+Resolved	HH Resolved
Training Epochs	500	500	500	500
Learning Rate	0.0005	0.0005	0.0015	0.0015
Batch Size	4096	4096	4096	4096
Dropout	0.2	0.2	0.2	0.2
L2 Gradient Clipping	1.0	1.0	1.0	1.0
L2 penalty	0.0002	0.0002	0.0002	0.0002
Hidden Dimension	64	32	32	32
Central Encoders	8	8	8	8
Branch Encoders	6	6	6	6
Number of heads	8	8	8	8
Partial Event Training	Yes	Yes	Yes	Yes
Cosine Annealing Cycles	5	1	1	1

Summary of Model Performance

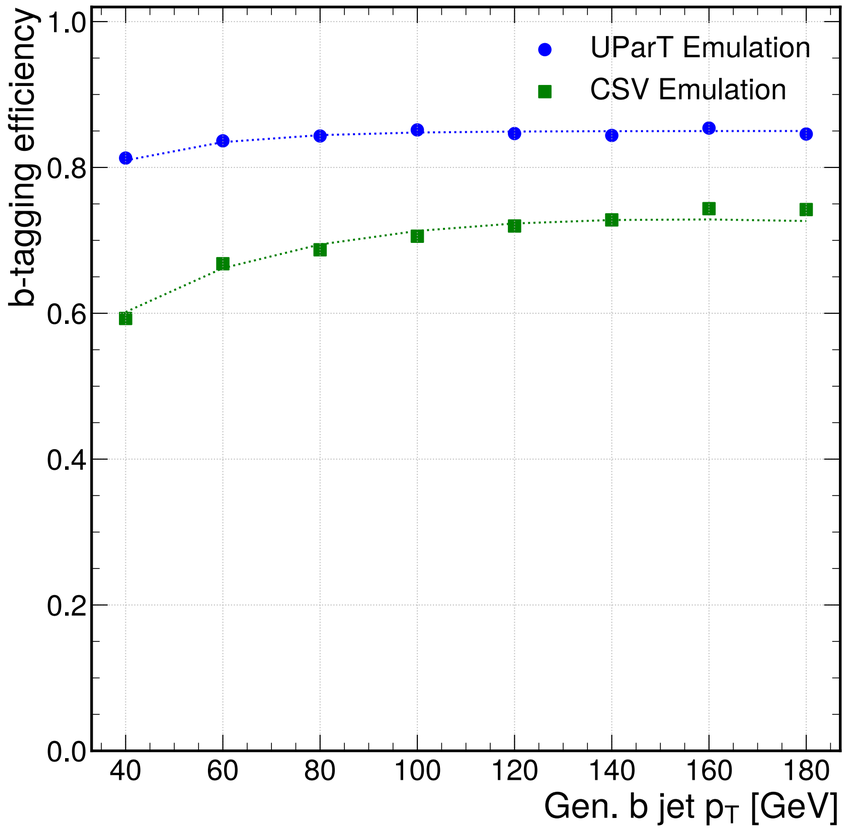
Particle metrics

Metrics	HHH				HH			
	Resolved		Boosted+Resolved		Resolved		Boosted+Resolved	
	Baseline	SPA-Net	Baseline	SPA-Net	Baseline	SPA-Net	Baseline	SPA-Net
H Purity	0.251	**0.444**	0.261	**0.409**	0.390	**0.702**	0.402	**0.719**
H Efficiency	0.394	**0.464**	0.367	**0.507**	0.578	**0.769**	0.547	**0.752**
Reco. Target H	32,223	**37,922**	30,787	**42,486**	47,039	**62,562**	45,481	**62,522**
Target H	81,810	81,810	83,858	83,858	81,330	81,330	83,178	83,178

Study on Improved b-tagging

Unified Particle Transformer (UParT) and the Combined Secondary Vertex (CSV)

- The b-tagging efficiency improves from approximately 70% to 85%.
- Markers: efficiencies calculated directly from the dataset
- Dotted lines: the parameterized equations.



Summary of Model Performance

Event confusion metrics