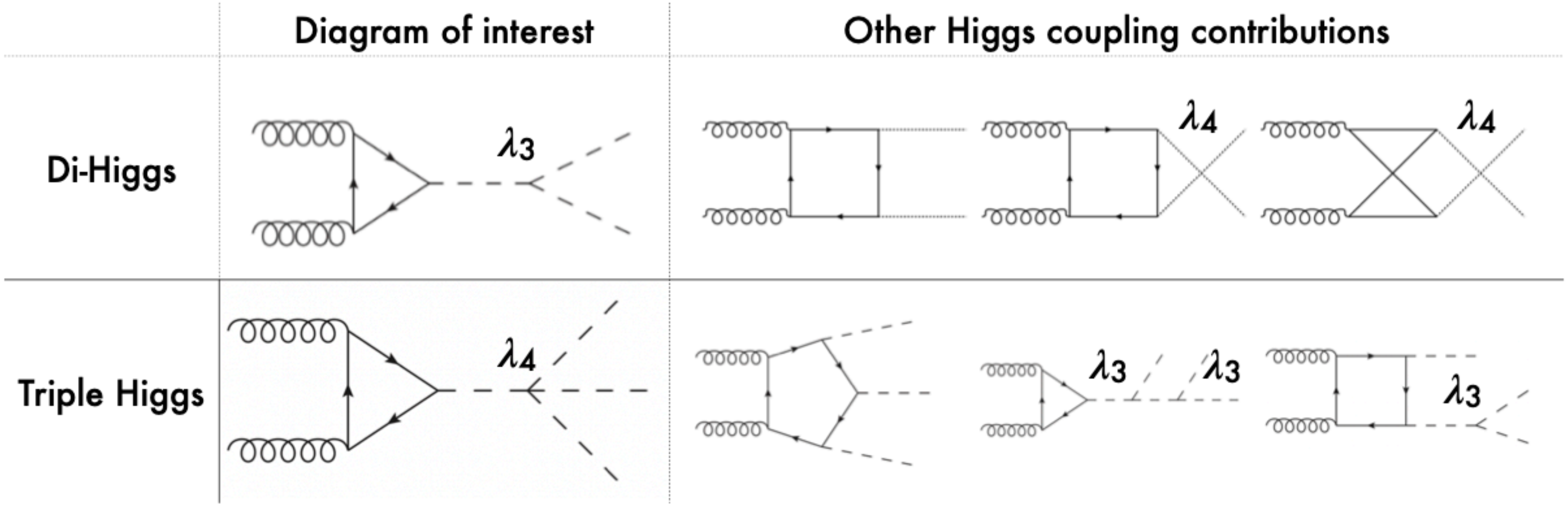
Nonresonant HHH6b: Boosted+Resolved Jet Assignment using Symmetry-Preserving Attention Networks

Haoyang Li, Ishaan Kavoori, Melissa Quinnan, Javier Duarte (UCSD),
Jovan Mitic, Predrag Milenovic (Belgrade),
Marko Stamenkovic (Brown),
Caden Mikkelsen, Harvey Newman (Caltech),
Cristina Mantilla Suarez (Fermilab),
Alexander Shmakov, Michael Fenton, Daniel Whiteson, Pierre Baldi (UCI)

Multi-Boson Interactions 2023, UCSD

Probing self-interaction di-higgs and triple Higgs

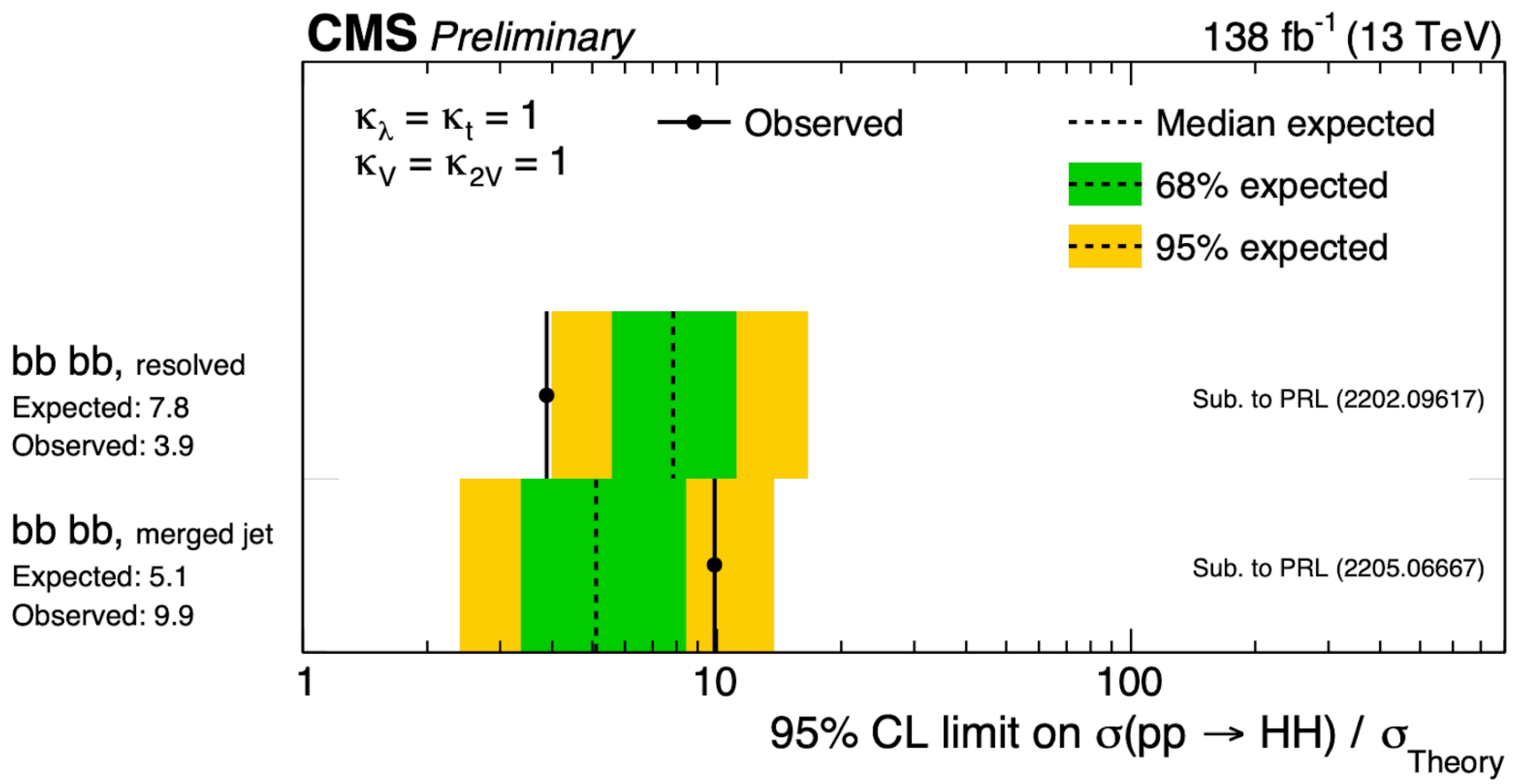


Probing the Higgs self-coupling possible through di-Higgs and triple Higgs measurements:

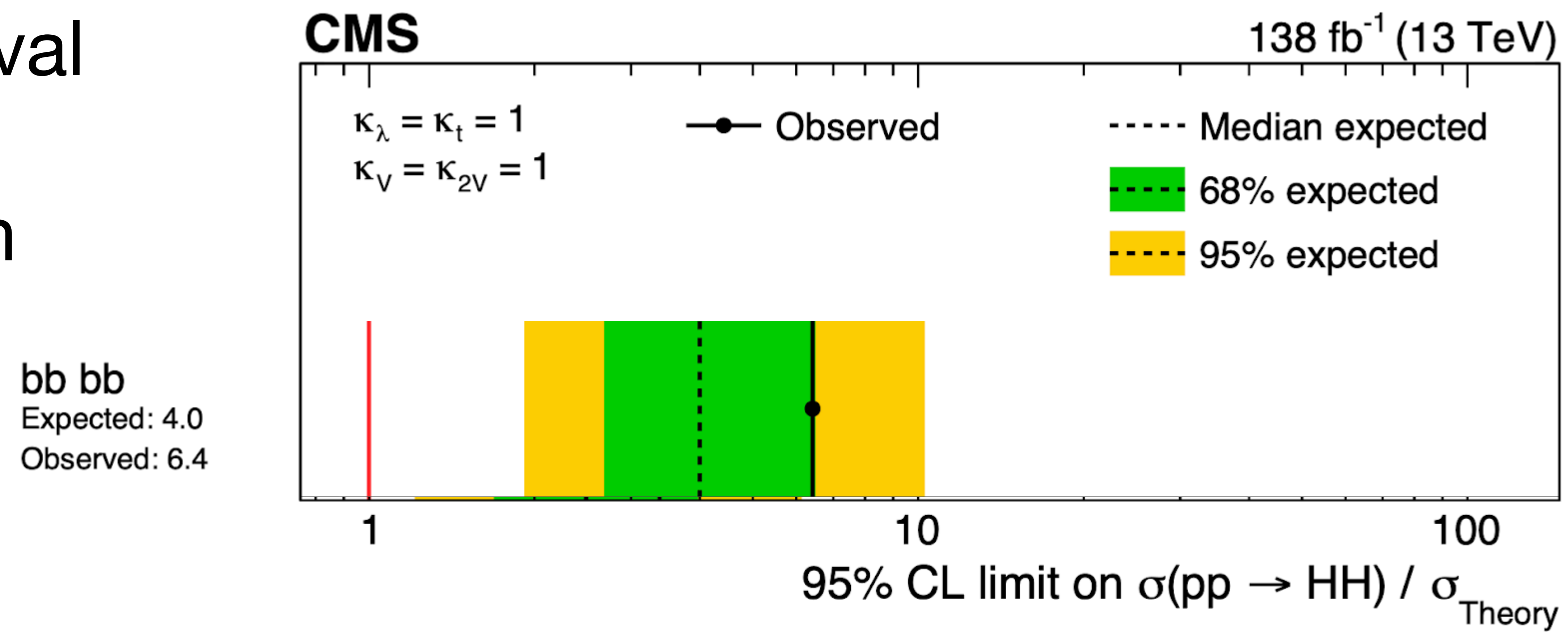
- Di-Higgs: (nearly) exclusively sensitive to λ_3 coupling
  - Small contribution from λ_4
- Triple Higgs: sensitive to both λ_3 and λ_4 coupling

→Full determination of Higgs potential only possible through combined measurement!

Recent CMS HH4b results



Overlap removal
&
combination



- Recent CMS combination of separate bb bb, resolved and bb bb, boosted/merged channels
- Combined after post-hoc overlap removal
- Since overlap removal was performed after the fact, it is potentially not optimal
- Can ML help us to determine whether events should be reconstructed as resolved or boosted?

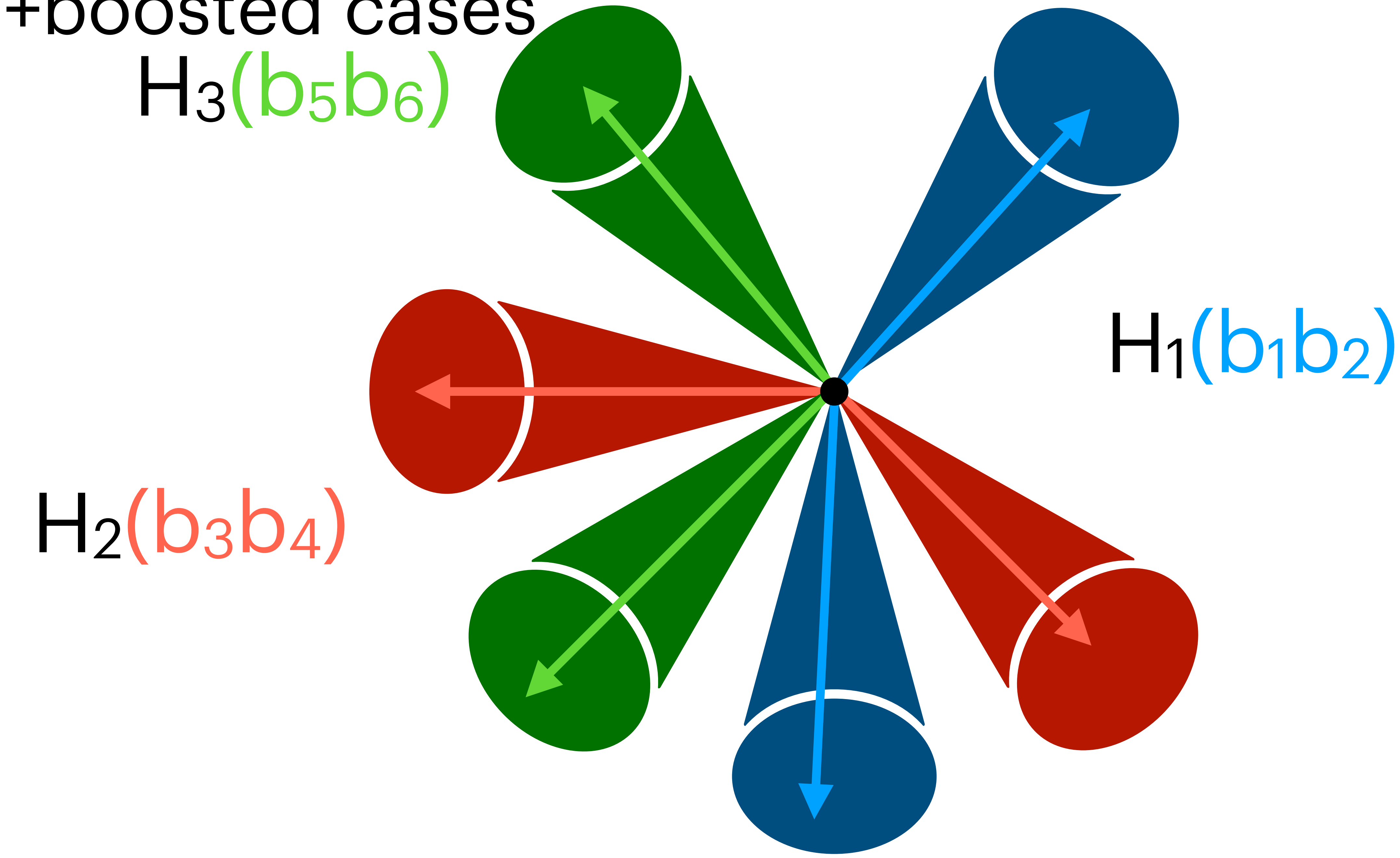
https://twiki.cern.ch/twiki/pub/CMSPublic/SummaryResultsHIG/limits_HH_SM.pdf

https://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-22-001/CMS-HIG-22-001_Figure_005-a.pdf

Introduction & Outline

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

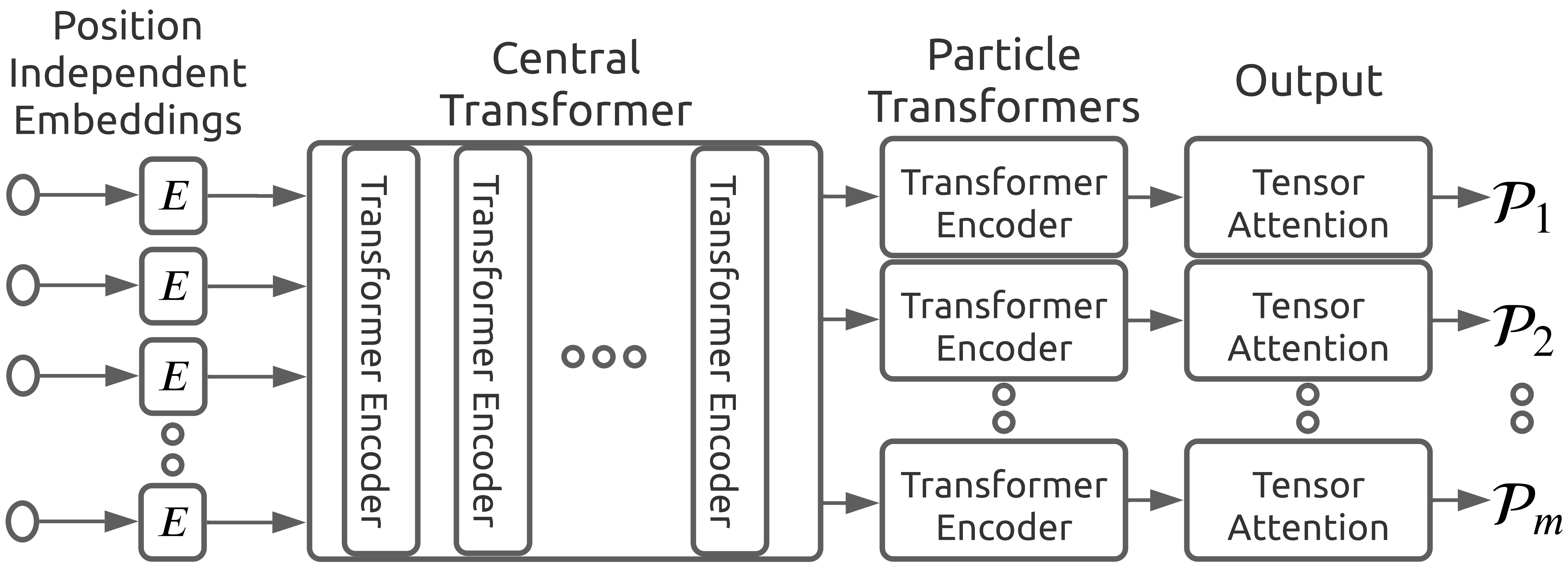
- HHH has a complex final state with many different types of (partial) reconstruction possible, including overlapping small- and large-radius jets
- Idea: use transformer to optimize reconstruction efficiency
    - Fully exploit event topology and kinematic correlations of jets to pair the 3 Higgs correctly
    - 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 resolved+boosted cases



- Outline of rest of talk
    - Overview of SPA-Net
    - Baseline methods
    - Preliminary results with HHH6b and HH4b

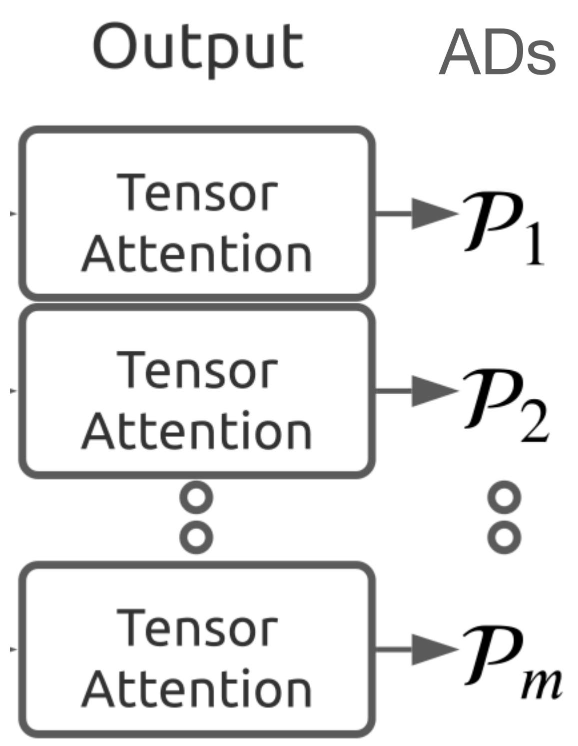
Symmetry Preserving Attention Networks (SPA-Net)

- 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 => $\mathrm{diag}(\mathscr{P}_p) = 0$



SPA-Net Output

- [Detection probability, assignment distribution] x N_p candidates (resonant particles)
  - For each particle candidate in N_p candidates
    - 1. If DP is lower than the threshold, SPANet did not find the particle, and the corresponding AD is ignored.
    - 2. The peak of AD indicates which combination of jets that SPANet predicts to reconstruct the particle



P_1	Jet 1	Jet 2	Jet 3	Jet 4
Jet 1	Mask	0.8	0.03	0.01
Jet 2	0.8	Mask	0.07	0.05
Jet 3	0.03	0.07	Mask	0.04
Jet 4	0.01	0.05	0.04	Mask

argmax$(P_1) = (2,1)$:
Jet 1 and Jet 2 are assigned to Particle 1

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

- Specify a list of resonant particles and their daughters

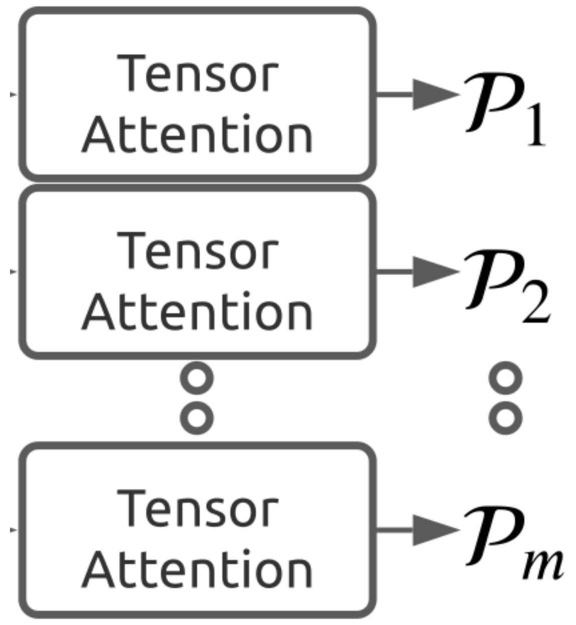
```
EVENT:
  h1:
    - b1: Jets
    - b2: Jets
  h2:
    - b1: Jets
    - b2: Jets
  h3:
    - b1: Jets
    - b2: Jets
```

- 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 all permutations to

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

Dataset & Input Features

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

- Using 14 TeV $pp \to HHH \to 6b$ events simulated with MadGraph+Pythia8+Delphes:
  - ~1M events for training+validation; ~300 events for testing
- Truth matching condition:
  - 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
- Higgs boson is "reconstructible" if both b quark daughters match to AK5 jets
- Up to 10 AK5 Jets are considered per event (ranked by p_{T})
- Input jet features:
  - p_{T} (log-normalized), η (normalized), $\sin\phi$, $\cos\phi$, boolean b-tag score, and jet mass (normalized)

Model Configuration

- Model hyperparameters:

```
num_embedding_layers: 10
position_embedding_dim: 16

transformer_activation: gelu
transformer_dim: 32
transformer_dim_scale: 2.0
transformer_type: Gated
num_attention_heads: 4

linear_activation: gelu
linear_block_type: GRU

hidden_dim: 64
initial_embedding_dim: 16
initial_embedding_skip_connections: 1
skip_connections: 1

num_encoder_layers: 4
num_branch_embedding_layers: 3
num_branch_encoder_layers: 3
num_detector_layers: 2
```

```
num_jet_embedding_layers: 0
num_jet_encoder_layers: 2
num_regression_layers: 3
num_classification_layers: 3

normalization: LayerNorm
normalize_features: 1
split_symmetric_attention: 1
```

HHH — Baseline Methods

- Method 1 (Higgs mass): $m_H = 125\,\mathrm{GeV}$
  - Note: higher efficiency, worse background mass sculpting

$$\chi^2 = (m_{b_1b_2} - m_H)^2 + (m_{b_3b_4} - m_H)^2 + (m_{b_5b_6} - m_H)^2$$

Baseline script

SPA-Net — HHH Performance

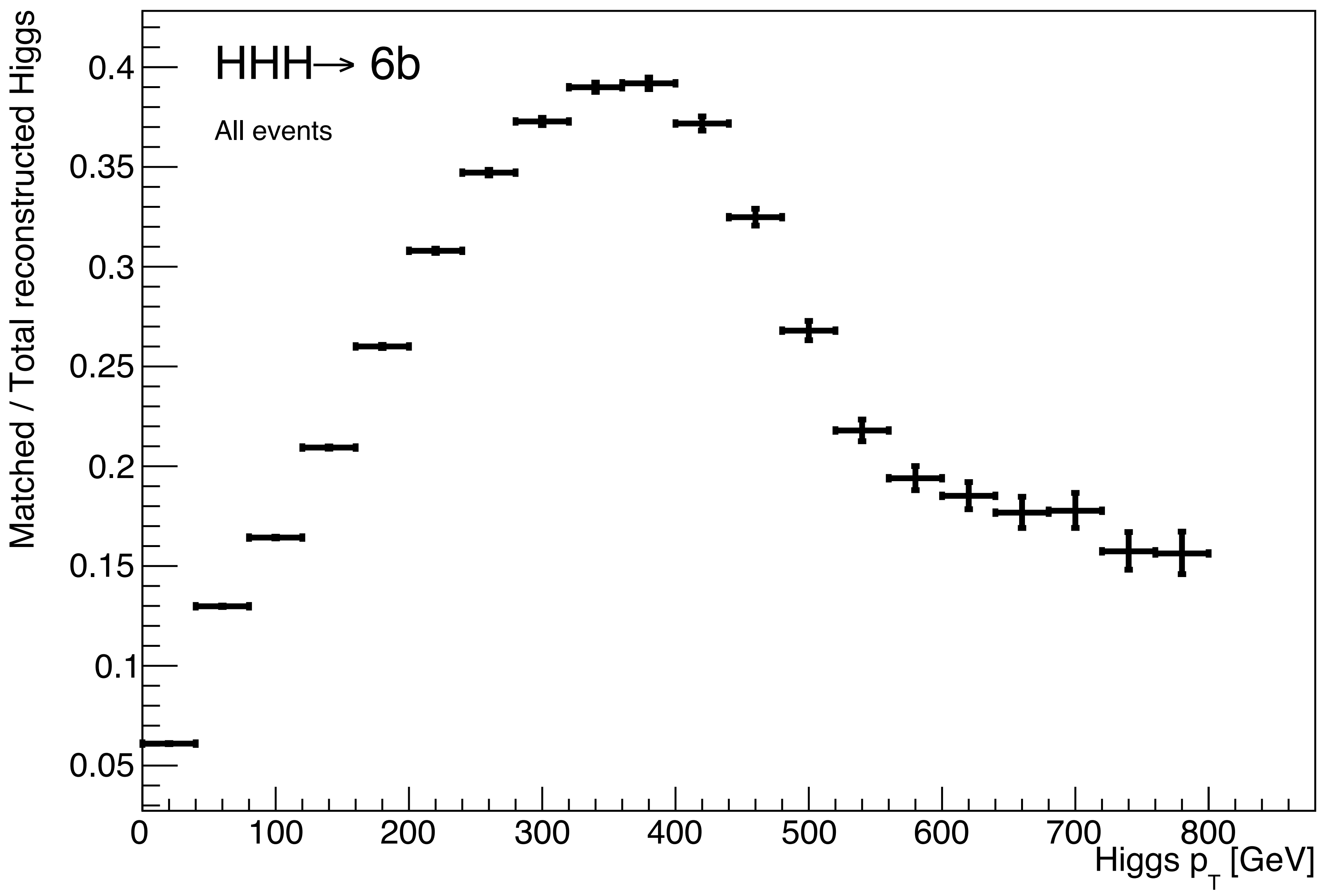
$$\text{Event Purity} = \frac{\text{Number of events that all Higgs are reconstructed}}{\text{Total number of events}}$$

$$\text{H Purity} = \frac{\text{Number of reconstructed Higgs}}{\text{Total number of Higgs}}$$

Event Type	Method	Event Purity	H Purity
1-3 H (98%)	Baseline	22%	39%
	SPANet	**34% (+54%)**	**52% (+33%)**
3 H (29%)	Baseline	23%	43%
	SPANet	**38% (+65%)**	**58% (+34%)**

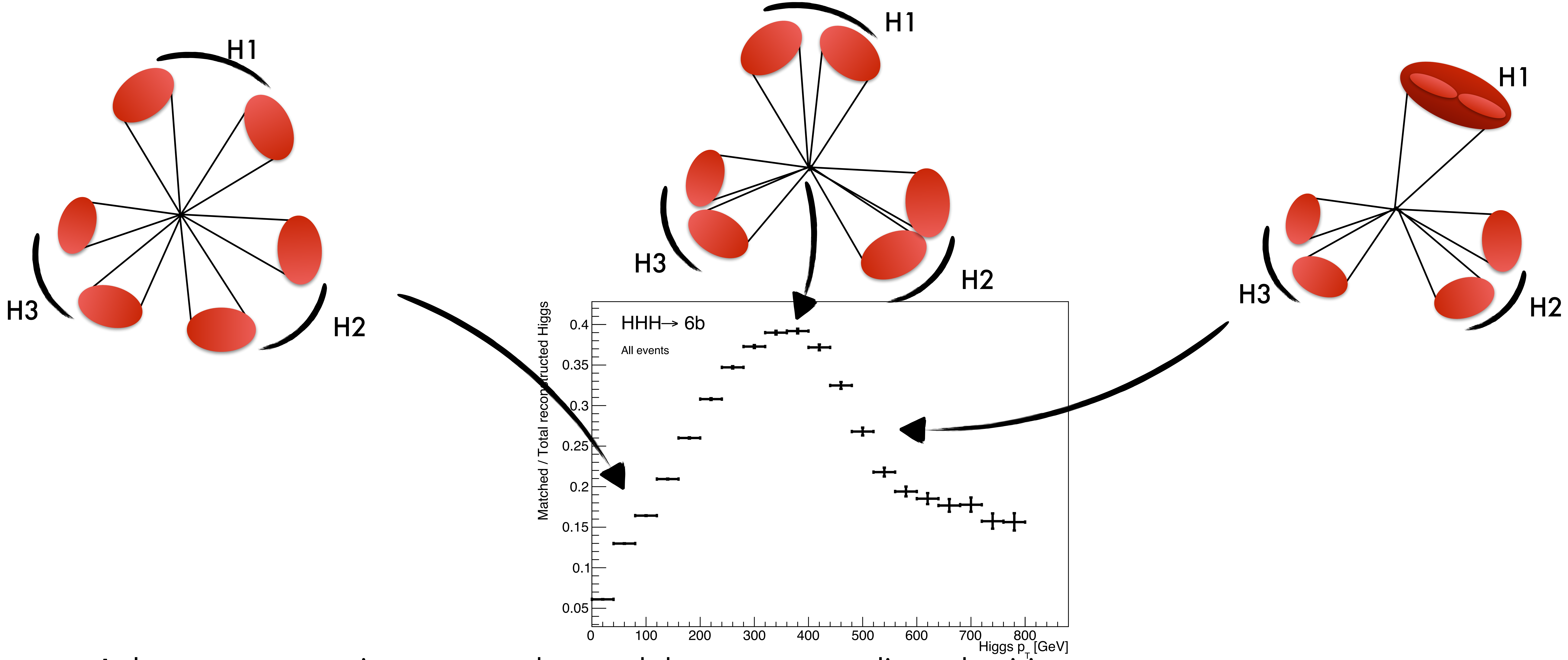
Differential matching efficiency

$$\text{Matching Efficiency} = \frac{\text{Number of predicted Higgs that are matched to gen}}{\text{Total number of Higgs}}$$



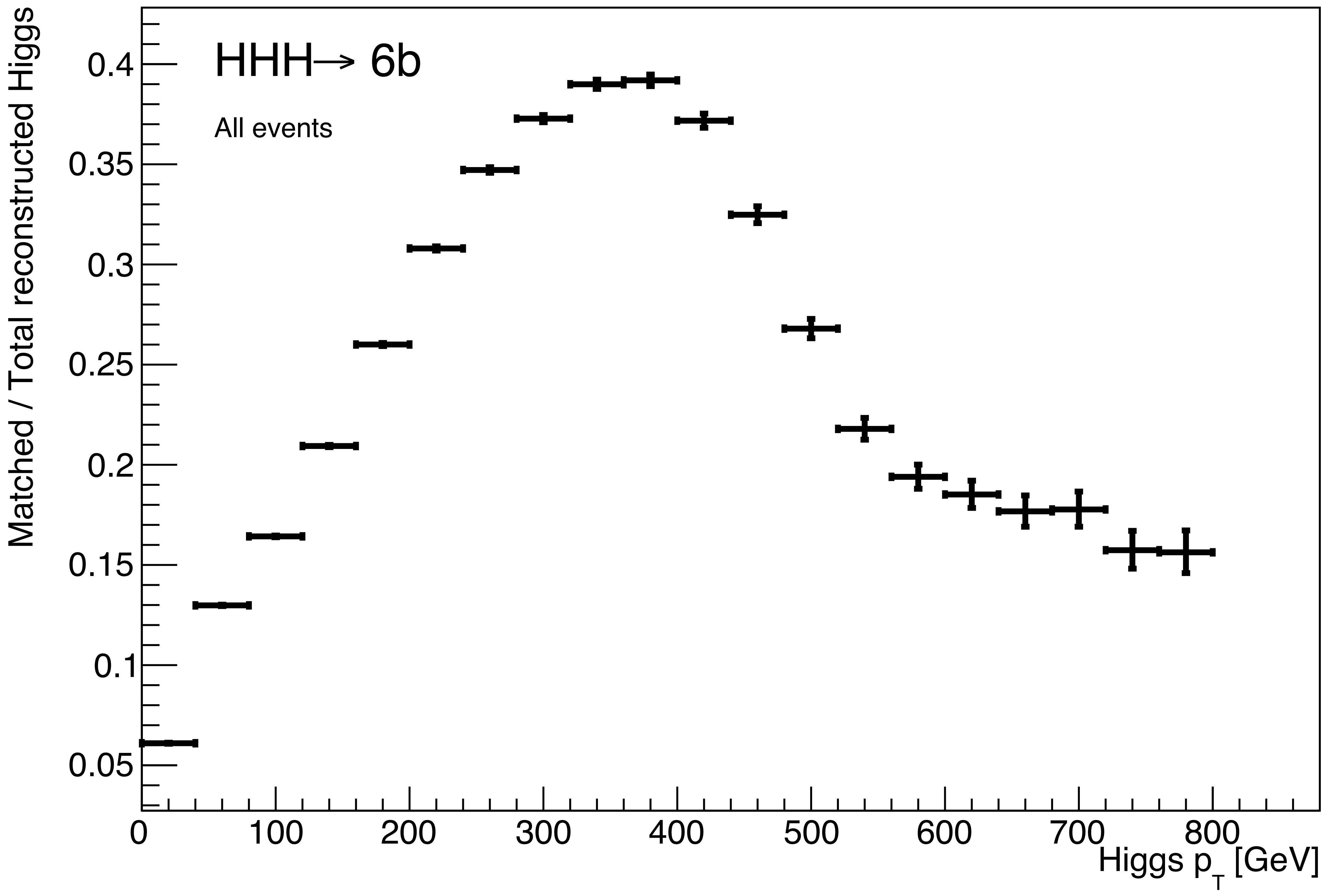
- Matching efficiency: strong dependence on momentum of the Higgs bosons

Differential matching efficiency

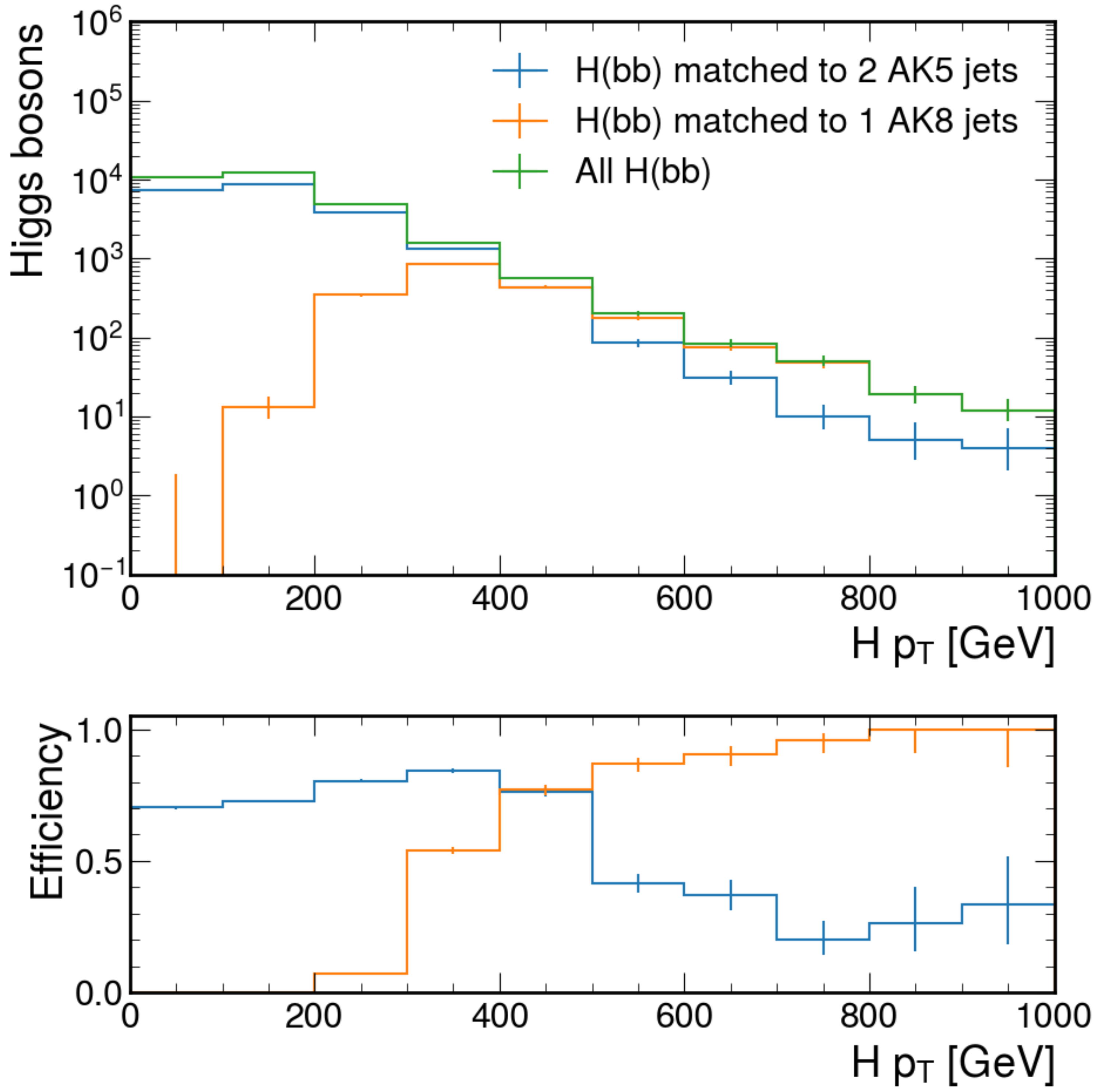


- At low momentum, jets scattered around detector = complicated pairing
- At higher momentum, jets from Higgs boson more and more collimated, clearer correlation
- At very high momentum (pT > 400 GeV), matching efficiency drops and Higgs reconstructed in AK8 jets

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

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.
- Tell SPANet boosted Higgs and resolved Higgs are the same particles by specifying all valid permutations

Input: >= 6 jets with pT > 20 GeV

- When existing, AK8 jets with pT > 250 GeV

Reconstruction algorithm:

- Prioritize boosted AK8 jets over AK5
  - If >= 1 AK8 jets found with high assignment probability, assign Higgs bosons to it
  - Complete remaining Higgses with AK5 pairs obtained by SPANET
- Next goal: let SPANET decide between AK8 and 2 AK5

Preliminary Study of SPANet on mixed HH4b

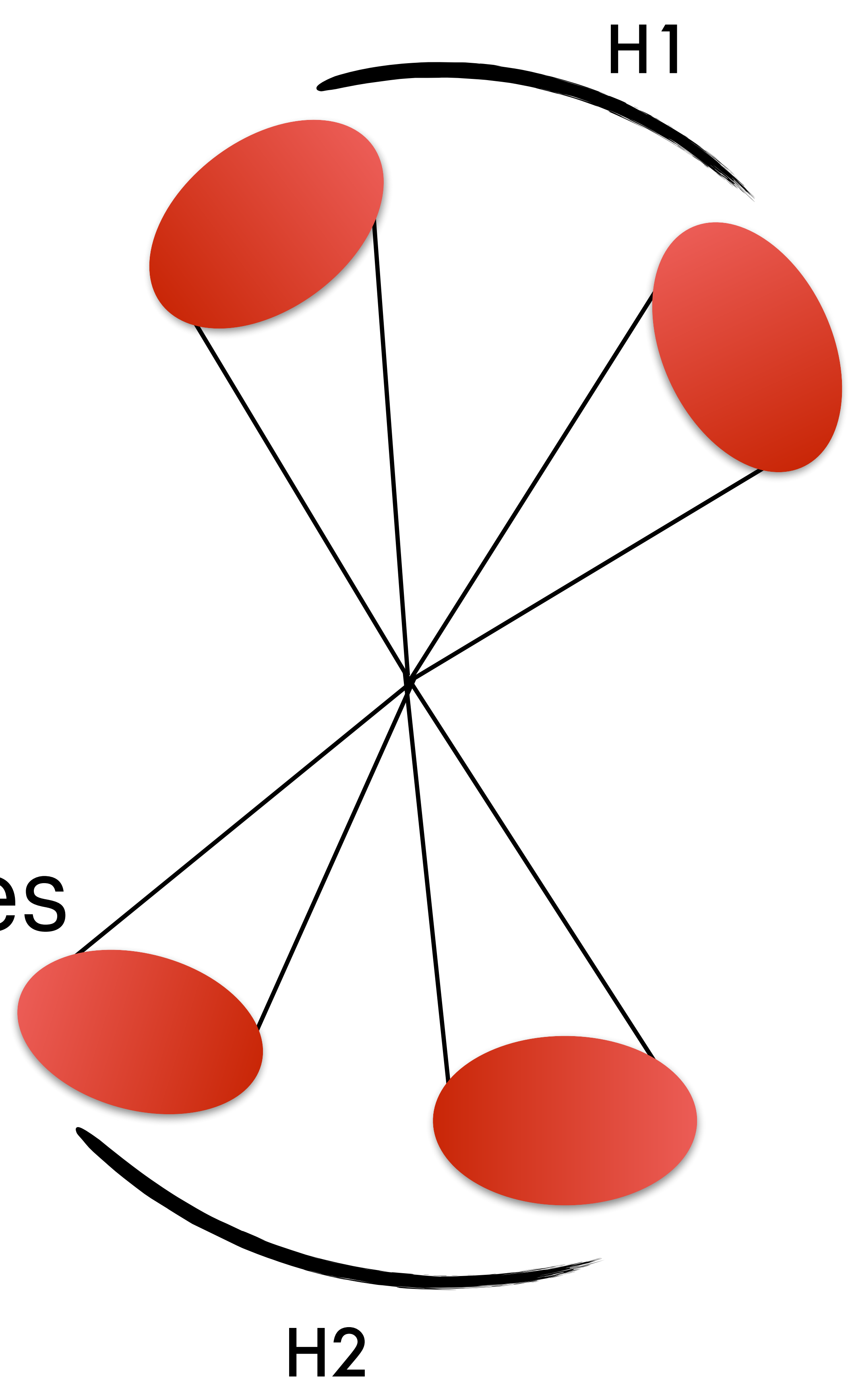
[1] arXiv:2202.09617

- Baseline (Higgs mass): $m_H = 125\,\mathrm{GeV}$
  - Note: background mass sculpting

$$\chi^2 = (m_{b_1b_2} - m_H)^2 + (m_{b_3b_4} - m_H)^2$$

- Mass agnostic distance method [1]:
  - Find pairs based on minimal distance between 2 Higgs masses

$$D = |\,m(b_1, b_2) - k \times m(b_3, b_4)\,| / \sqrt{1 + k^2}, k = 125/120$$



SPA-Net — HH Performance

$$\text{Event Purity} = \frac{\text{Number of events that all Higgs are reconstructed}}{\text{Total number of events}}$$

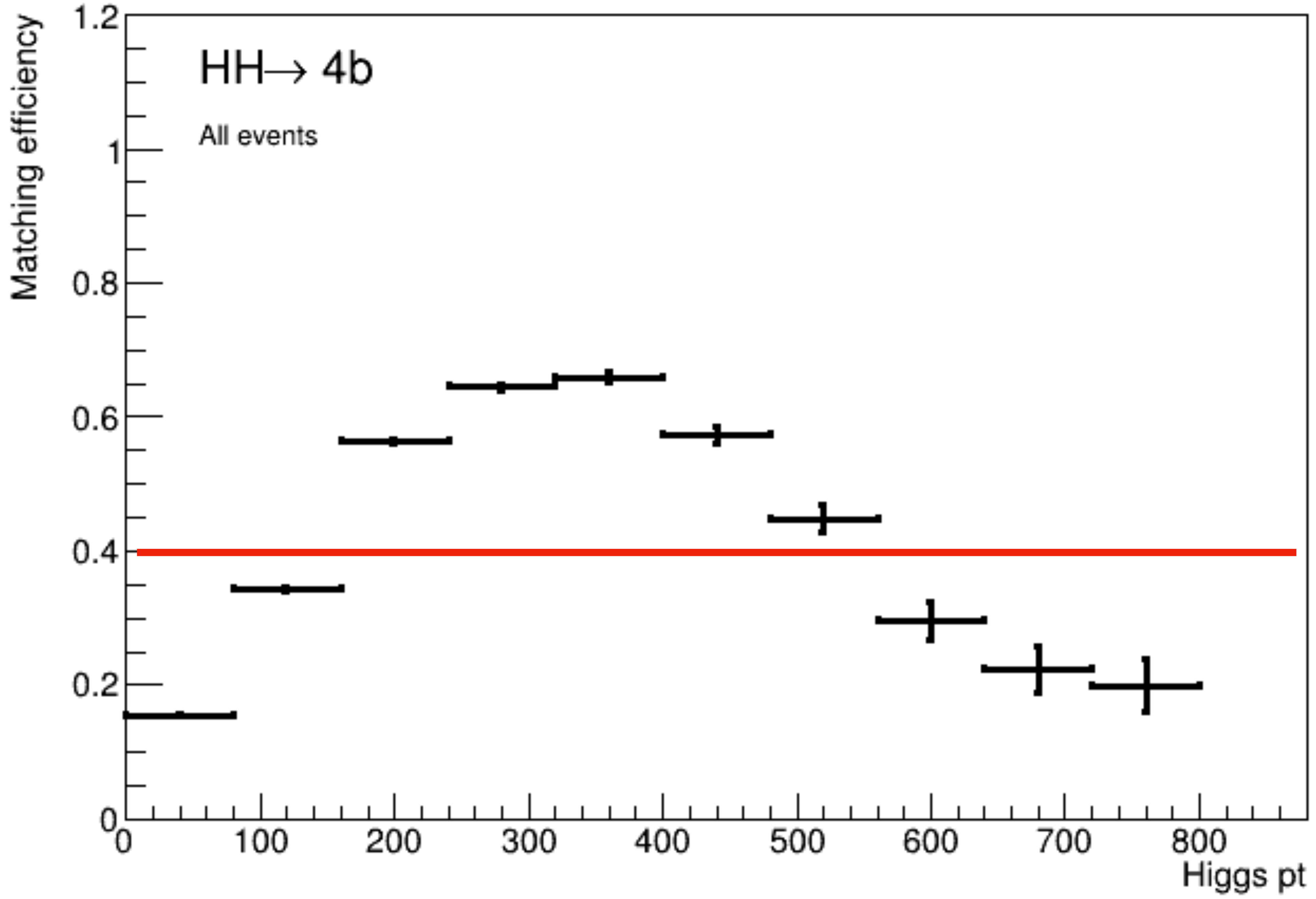
$$\text{H Purity} = \frac{\text{Number of reconstructed Higgs}}{\text{Total number of Higgs}}$$

Event Type	Method	Event Purity	H Purity
1-2 H	Baseline	44%	57%
	SPANet	**76% (+72%)**	**81% (+42%)**
2 H	Baseline	21%	53%
	SPANet	**77% (+360%)**	**84% (+58%)**

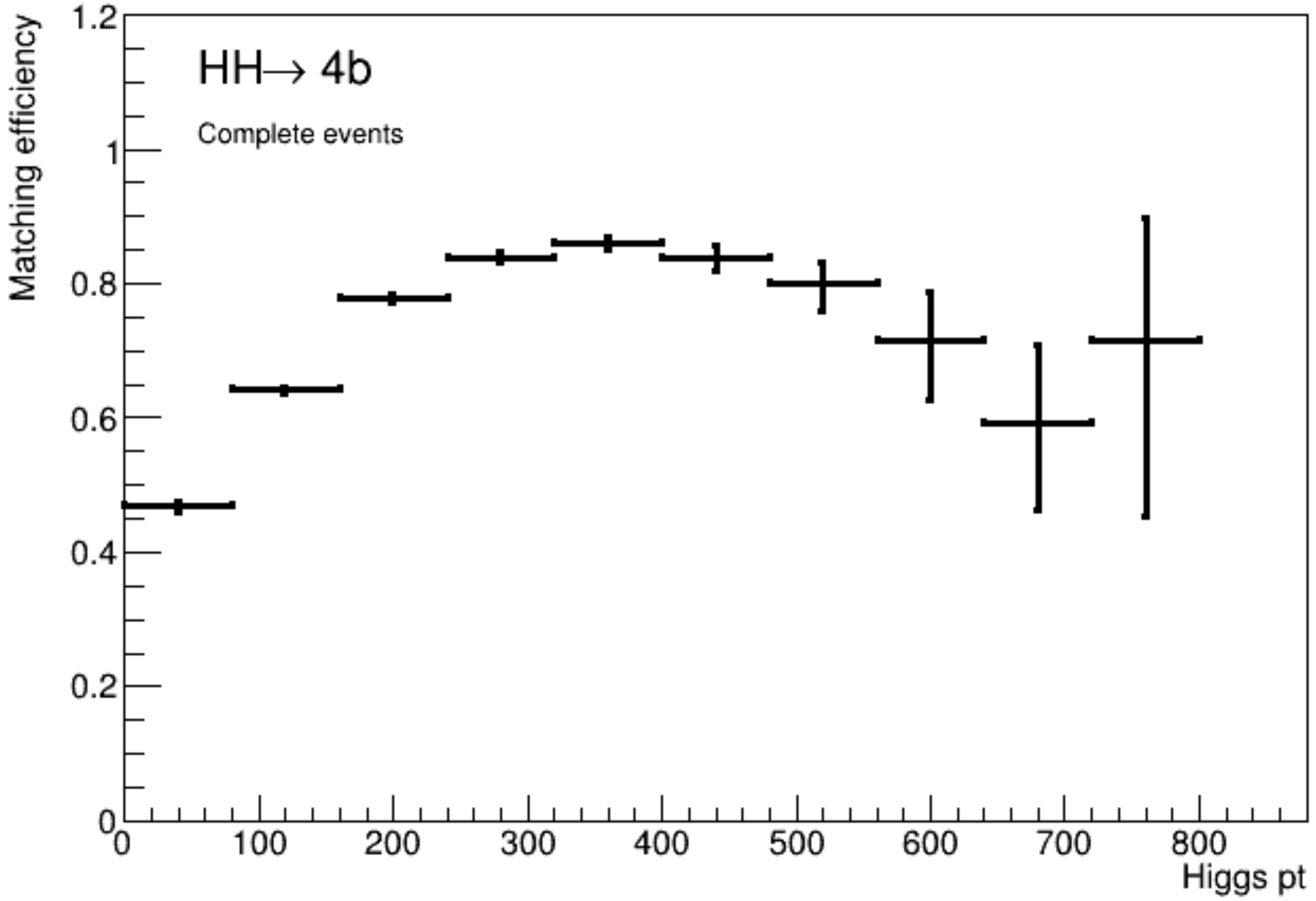
Note: using top 4 jets in each event ordered by pT

Differential matching efficiency

1-2h events



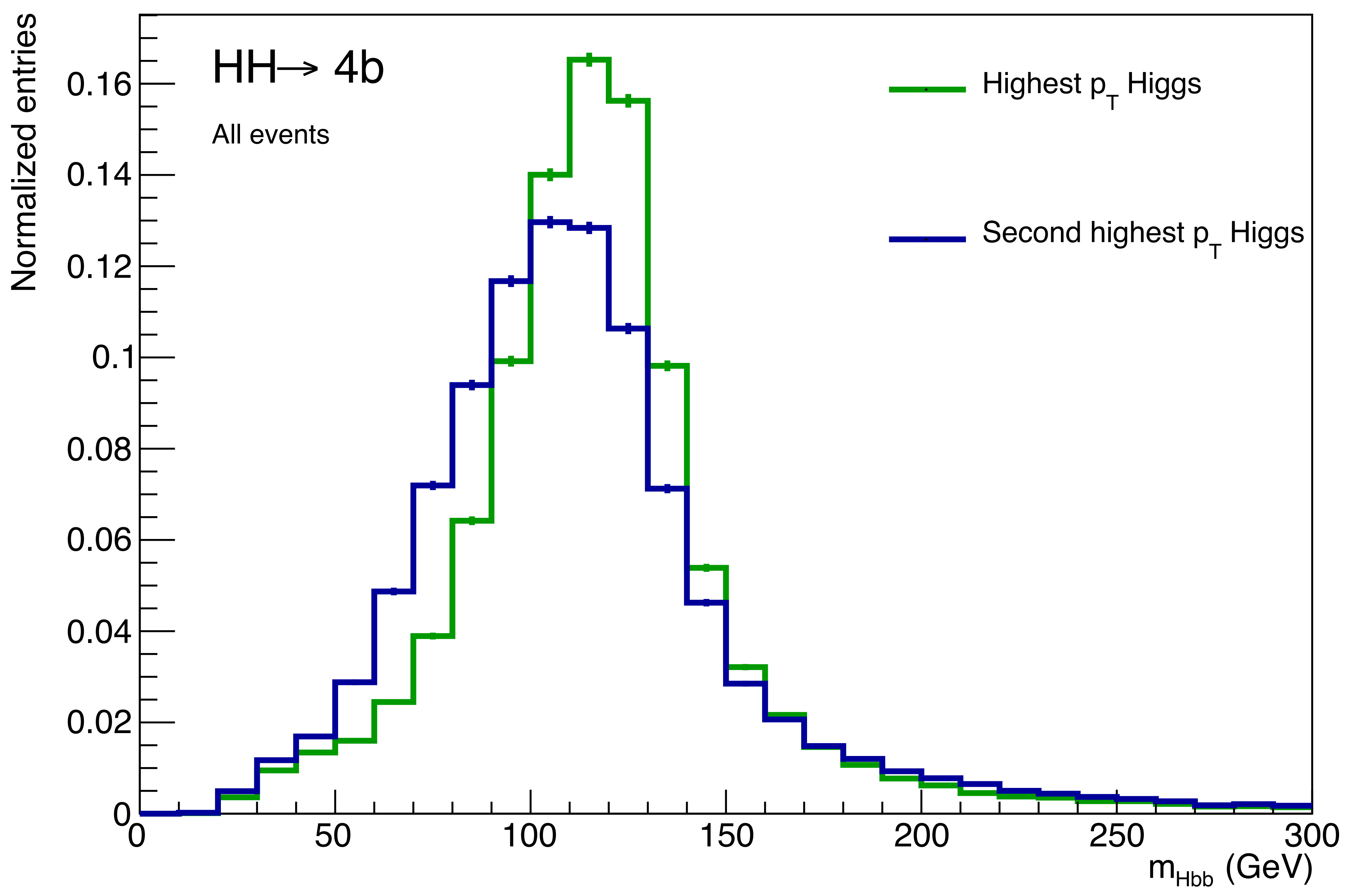
2h events



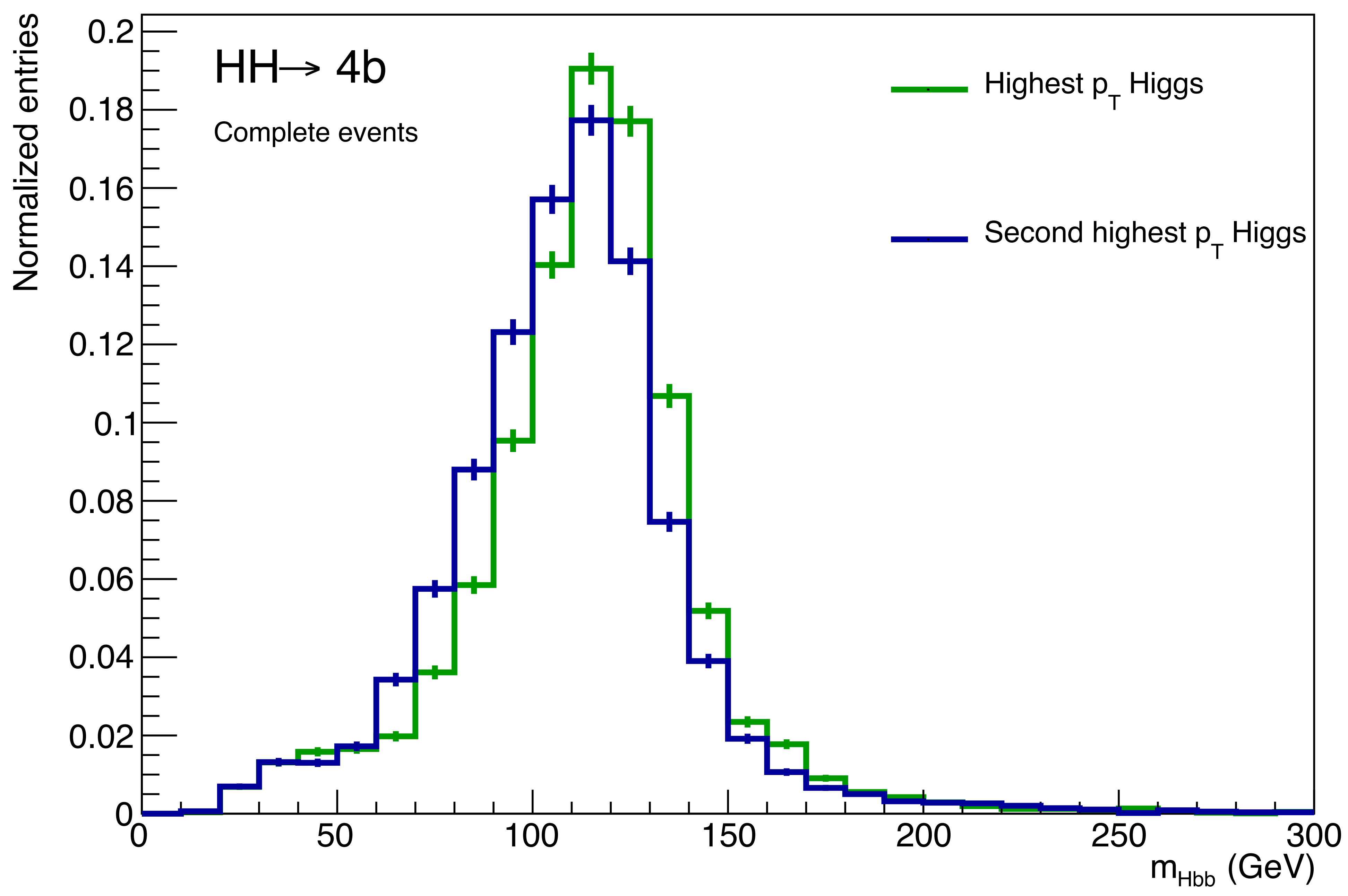
- At 400 GeV, Higgs more likely to be reconstructed in 1 AK8 than 2 AK5
- Optimal performance: generalize approach to both boosted + resolved topologies
- On-going work to define best strategy and compare results with HHH

Mass Reconstruction HH

1-2h events



2h events



- SPA-Net reconstructs the mass of each Higgs candidate appropriately
- Work in progress: investigating mass sculpting of backgrounds
- Investigating using SPANet to do boosted + resolved analysis

Summary

- SPANet: A transformer model for particle reconstruction.
- SPANet shows better performance than chi2 in our preliminary study of HHH6b.
  - Unique algorithm to pair fully resolved, semi-boosted, fully boosted simultaneously
  - Performance improvements validated on HH4b signal too

- SPANet can lead to better reconstruction efficiency and therefore better determination of fundamental parameters in the Higgs sector

Back-up

Detailed Model Configuration

- Many hyperparameters to tune!
- We used the following:


```yaml
assignment_loss_scale: 1.0
balance_classifications: false
balance_jets: 0
balance_losses: true
balance_particles: 1
batch_size: 4096
classification_loss_scale: 0.0
combinatorial_scale: 0.0
combine_pair_loss: min
dataset_limit: 1.0
dataset_randomization: 0
detection_loss_scale: 0.0
dropout: 0
epochs: 250
event_info_file: event_files/hhh_masses.yaml
focal_gamma: 0.0
gradient_clip: 0.0
hidden_dim: 64
initial_embedding_dim: 16
initial_embedding_skip_connections: 1
kl_loss_scale: 0.0
l2_penalty: 0.0002
learning_rate: 0.0015
learning_rate_cycles: 1
learning_rate_warmup_epochs: 1.0
limit_to_num_jets: 0
linear_activation: gelu
linear_block_type: GRU
linear_prelu_activation: true
mask_sequence_vectors: 1
masking: Filling
normalization: LayerNorm
normalize_features: 1
num_attention_heads: 4
num_branch_embedding_layers: 3
num_branch_encoder_layers: 3
num_classification_layers: 3
num_dataloader_workers: 4
num_detector_layers: 2
num_embedding_layers: 10
num_encoder_layers: 4
num_gpu: 1
num_jet_embedding_layers: 0
num_jet_encoder_layers: 2
num_regression_layers: 3
optimizer: AdamW
partial_events: 1
position_embedding_dim: 16
regression_loss_scale: 0.0
skip_connections: 1
split_symmetric_attention: 1
testing_file: ''
train_validation_split: 0.95
training_file: data/hhh_training_masses.h5
transformer_activation: gelu
transformer_dim: 32
transformer_dim_scale: 2.0
transformer_type: Gated
trial_output_dir: ./test_output
trial_time: ''
usable_gpus: ''
validation_file: ''
verbose_output: false
```

Symmetric Tensor Attention

- Note $\mathscr{P}_p$ is an "overparameterization" of the valid jet assignments: many represent the same physical combinations.
  - For example for the $HHH \to 6b$ case, 10 jets maximum
    - Each $\mathscr{P}_p$ has 100 entries
    - But we can swap (b_1, b_2) for each H, and can swap H_1, H_2, H_3
    - In the end we end up with only 3150 unique physical assignments!

$$\forall \sigma \in G_p \ \left(j_1, j_2, \ldots, j_{k_p}\right) \simeq \left(j_{\sigma(1)}, j_{\sigma(2)}, \ldots, j_{\sigma\left(k_p\right)}\right) \iff \mathscr{P}_p^{j_1 j_2 \cdots j_{k_p}} = \mathscr{P}_p^{j_{\sigma(1)} j_{\sigma(2)} \cdots j_{\sigma\left(k_p\right)}}$$

$$\mathcal{S}^{i_1 i_2 \cdots i_{k_p}} = \sum_{\sigma \in G_p} \Theta^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma\left(k_p\right)}},$$

$$\mathcal{O}^{j_1 j_2 \cdots j_{k_p}} = X_{i_1}^{j_1} X_{i_2}^{j_2} \cdots X_{i_{p_k}}^{j_{p_k}} \mathcal{S}^{i_1 i_2 \cdots i_{k_p}},$$

$$\mathscr{P}_p^{j_1 j_2 \cdots j_{k_p}} = \frac{\exp(\mathcal{O}^{j_1 j_2 \cdots j_{k_p}})}{\sum_{j_1, j_2, \ldots, j_{p_k}} \exp(\mathcal{O}^{j_1 j_2 \cdots j_{k_p}})}.$$

Combined Symmetric Loss

- Symmetric attention layers produce solutions $\{\mathscr{P}_1, \mathscr{P}_2, \ldots, \mathscr{P}_m\}$ for each particle's jet-carton assignment sub-problem
- True assignments are delta-distributions containing one possible valid jet assignment $\{\mathscr{T}_1, \mathscr{T}_2, \ldots, \mathscr{T}_m\}$.
- Loss for each sub-problem is the categorical cross entropy for each particle p
- Permutation group G_E induces an equivalence relation over particles: $\forall \sigma \in G_E, (\mathscr{T}_1, \mathscr{T}_2, \ldots, \mathscr{T}_m) \simeq (\mathscr{T}_{\sigma(1)}, \mathscr{T}_{\sigma(2)}, \ldots, \mathscr{T}_{\sigma(m)})$
- Incorporate these symmetries by allowing network to fit any equivalent jet assignment (minimize loss over a given equivalence class)

Partial Event Reconstruction

- Though each parton is usually expected to produce a jet, some particles are impossible to reconstruct
- Mask unreconstrable particles and only include the loss contributed by reconstructable particles
- Also, scale the loss based on the distribution of events present in the training dataset by computing the effective class count for each partial combination

$$\mathscr{L}_{\text{min}}^{\text{masked}} = \min_{\sigma \in G_E} \left(\sum_{i=1}^{m} \frac{\mathscr{M}_{\sigma(i)} CE(\mathscr{P}_i, \mathscr{T}_{\sigma(i)})}{CB\left(\mathscr{M}_{\sigma(1)}, \mathscr{M}_{\sigma(2)}, \ldots, \mathscr{M}_{\sigma(m)}\right)} \right).$$