# SPA-Net for HHH→6b Resolved and Boosted Jet Assignment

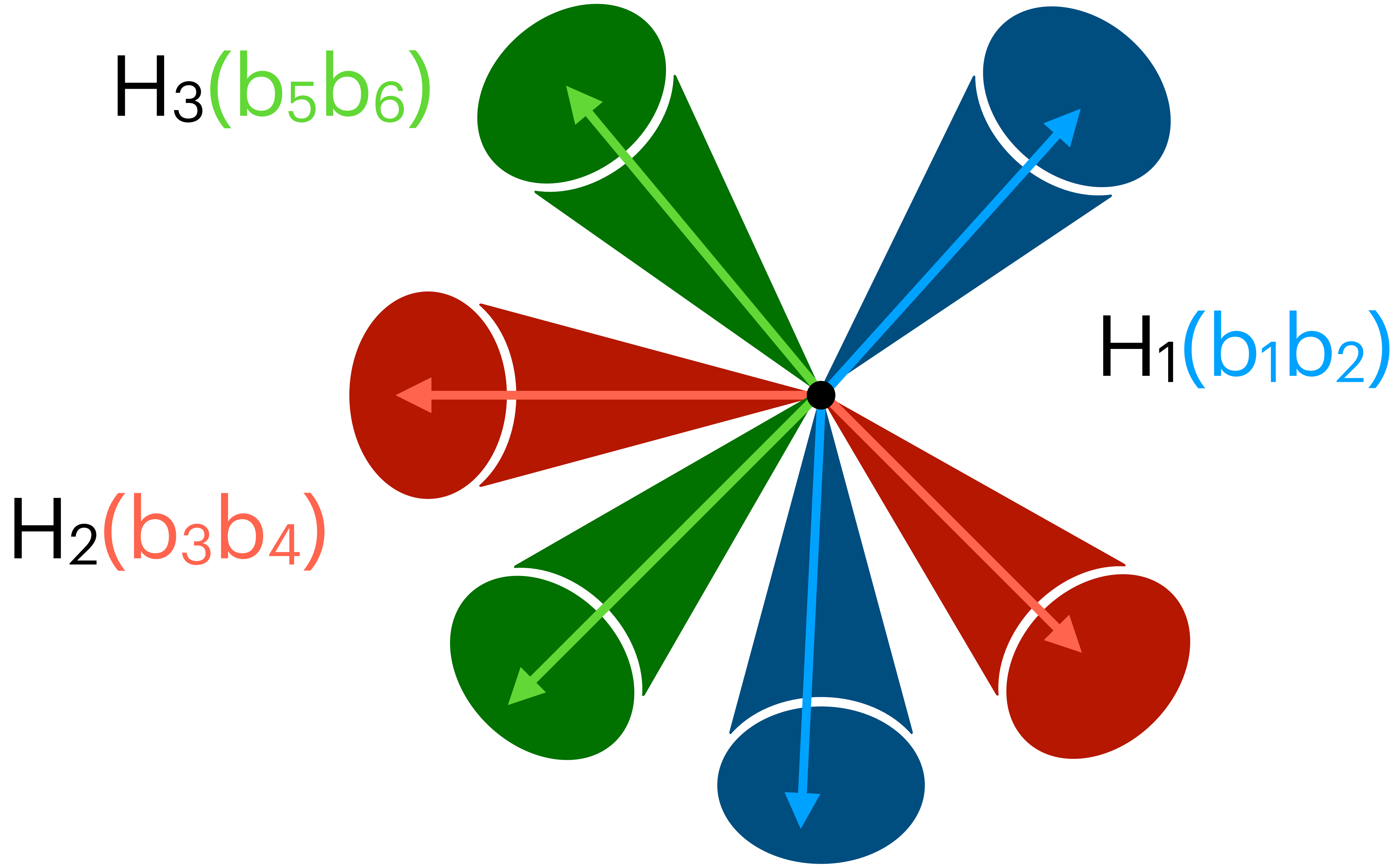
Billy Haoyang Li, Ishaan Kavoori, Melissa Quinnan, Javier Duarte (UCSD)
Alexander Shmakov, Michael Fenton, Daniel Whiteson, Pierre Baldi (UCI)
Cristina Mantilla Suarez (Fermilab)
Jovan Mitic, Predrag Milenovic (Belgrade)
Marko Stamenkovic, Greg Landsberg (Brown)
Caden Mikkelsen, Harvey Newman (Caltech)

14th of July 2023, HHH workshop, Dubrovnik

# Introduction & Outline

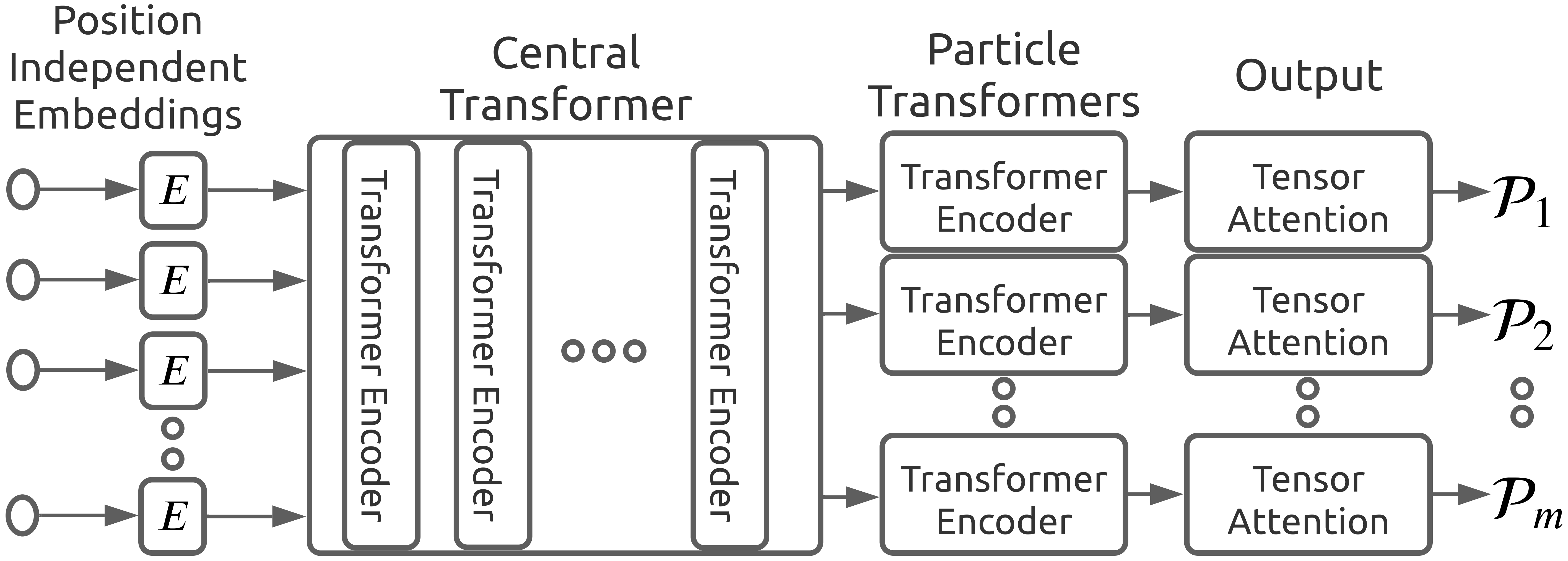
[1] arXiv:2010.09206
[2] arXiv:2106.03898

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



# Symmetry Preserving Attention Networks (SPA-Net)

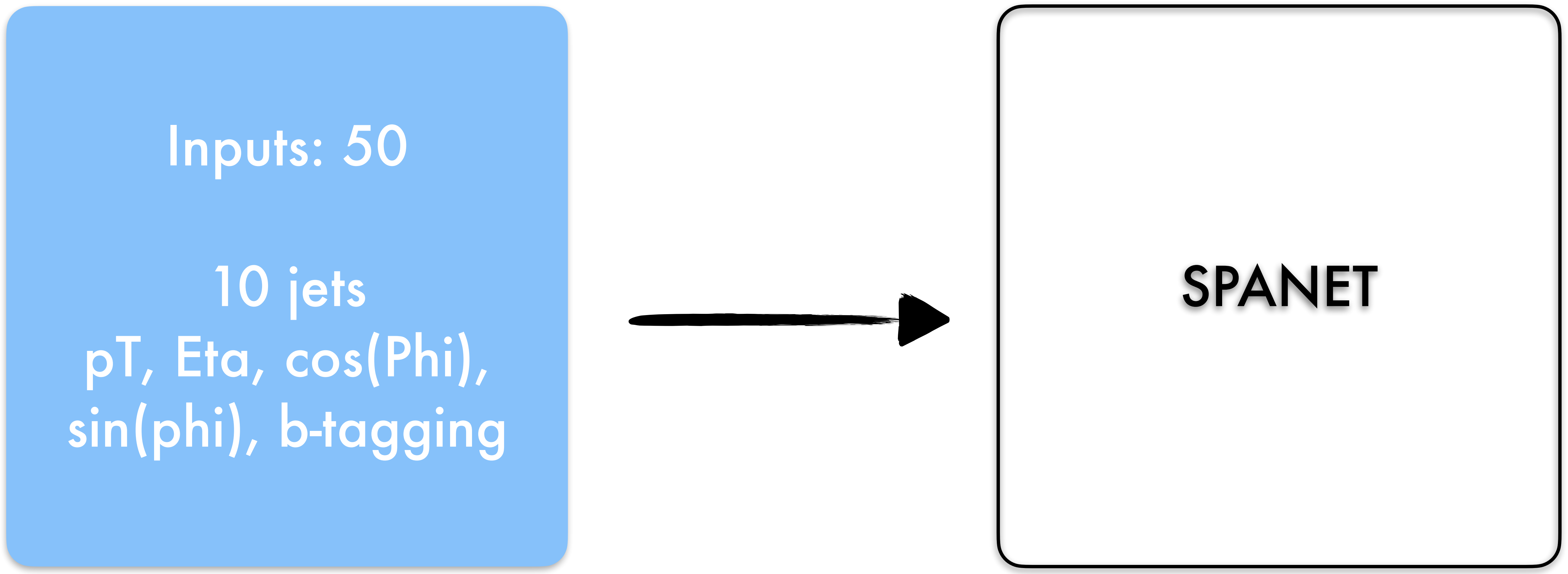
- Consider all valid permutations using *symmetric tensor attention*
  - Resonance particle $p$ (e.g., Higgs) is associated with $k_p$ partons (e.g., 2 b quarks); maximum of $N$ reconstructed jets (e.g., 10)
  - Input: matrix of transformer-encoded jets $X_p \in \mathbb{R}^{N \times D}$
  - Output: rank-$k_p$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

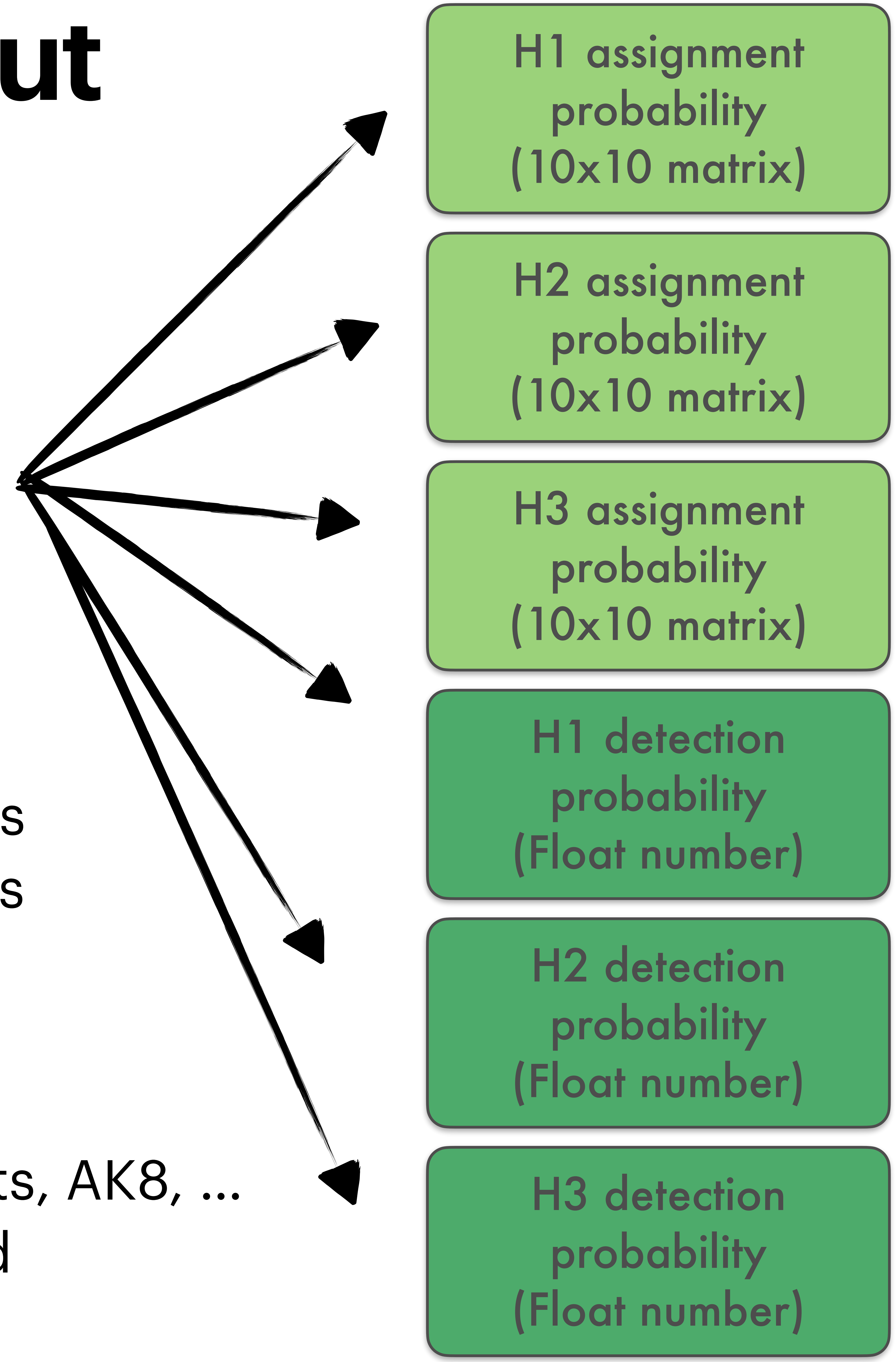


# SPA-Net Output

- [**Detection probability**, **assignment distribution**] x $N_c$ candidates
  - For each particle candidate in $N_c$ 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 SPANet predicts correct:
      - E.g. If $k_p$ = 3 and argmax$(P_{ijk}) = (1,2,3)$, then jet 1, 2, 3 reconstruct this particle candidate. The AD tensor is symmetric by design.
- Optional auxiliary output:
  - Classifications, regressions learned from your customized dataset.
  - MLP classifiers and regressors can take selected or grouped inputs from different depth (event level, particle level, jet level).

# SPA-Net Output



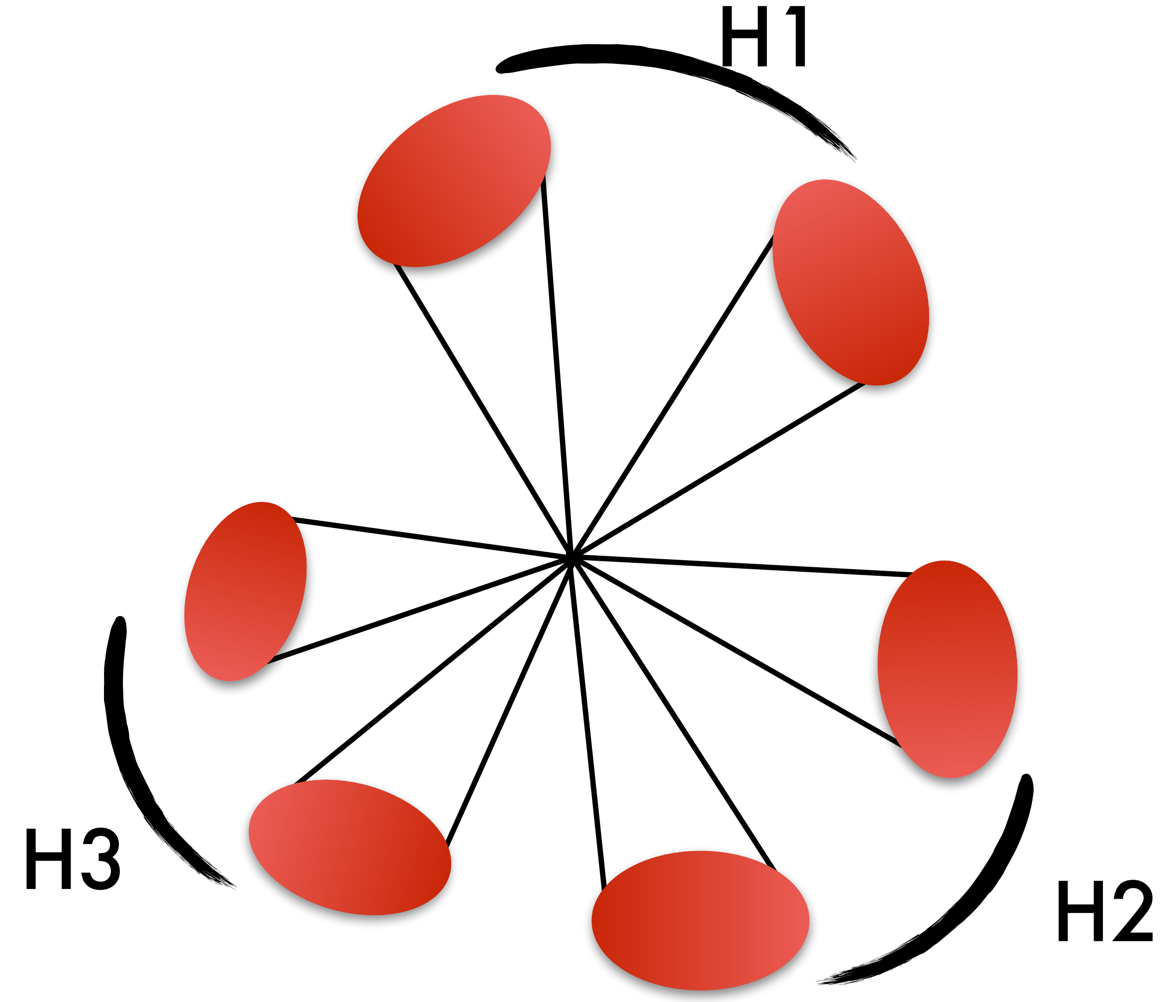

Inputs: start from signal samples HHH used for analysis

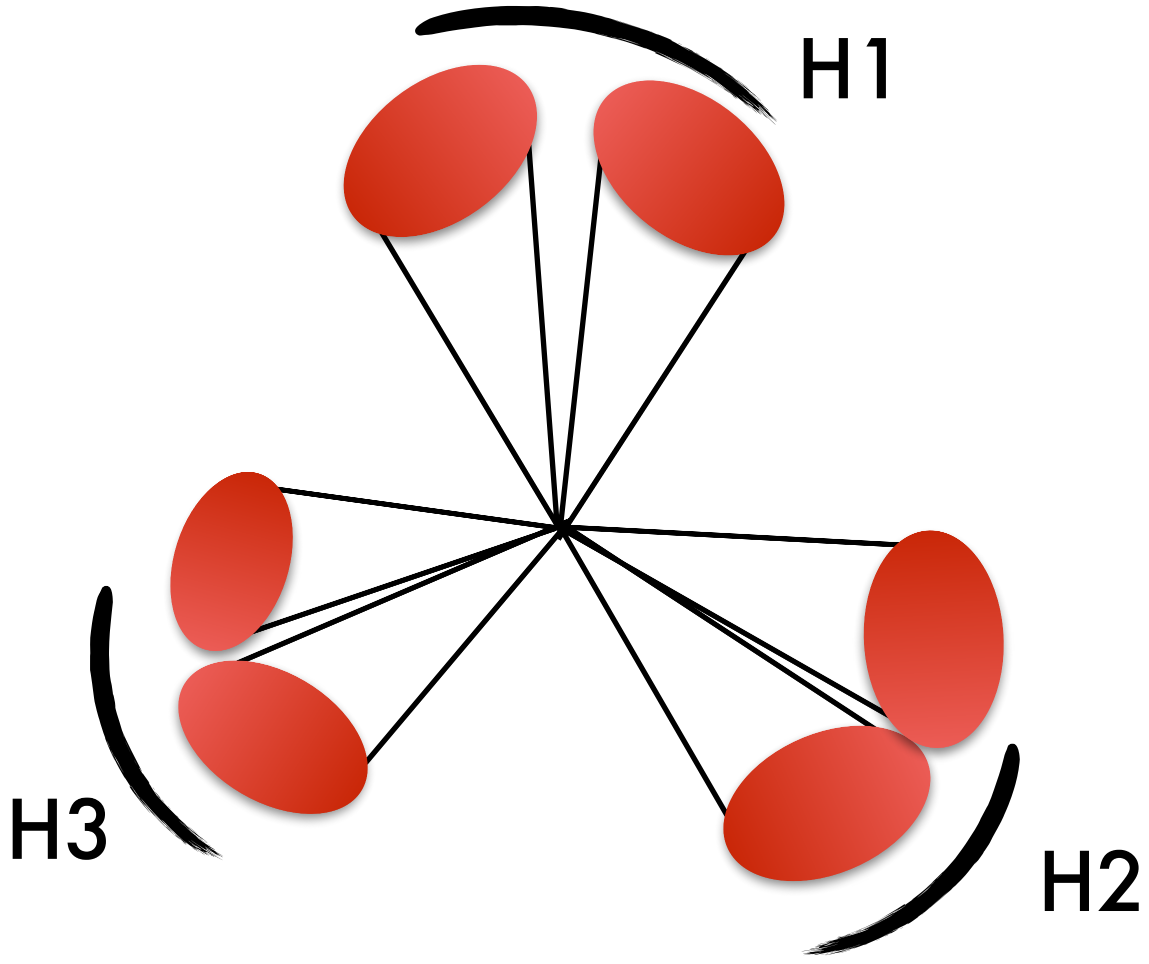
- Inputs: 10 AK5 jets with pT, eta, phi, b-tagging, mass

Advantages of SPANET:

- Exploit kinematic information of all jets (not just 6)
- Dynamical: each event has a different number of jets, AK8, ...
- Interesting interplay between boosted and resolved

# Resolved HHH topologies



# Event Configuration: fully resolved

- To give a flavor of how we use it, we can compare the event configurations for $t\bar{t} \to (bq\bar{q})(\bar{b}q\bar{q})$ (one of the original use cases) and $HHH \to (b\bar{b})(b\bar{b})(b\bar{b})$

```
# ----------------------------------------------------------------------
# REQUIRED - EVENT - Complete list of resonance particles and daughters.
# ----------------------------------------------------------------------
EVENT:
  t1:
    - q1: Jets
    - q2: Jets
    - b:  Jets
  t2:
    - q1: Jets
    - q2: Jets
    - b:  Jets
# ----------------------------------------------------------
# REQUIRED KEY - PERMUTATIONS - List of valid permutations.
# ----------------------------------------------------------
PERMUTATIONS:
    EVENT:
      - [ t1, t2 ]
    t1:
      - [ q1, q2 ]
    t2:
      - [ q1, q2 ]
```

```
# ----------------------------------------------------------------------
# REQUIRED - EVENT - Complete list of resonance particles and daughters.
# ----------------------------------------------------------------------
EVENT:
  h1:
    - b1: Jets
    - b2: Jets
  h2:
    - b1: Jets
    - b2: Jets
  h3:
    - b1: Jets
    - b2: Jets
# ----------------------------------------------------------
# REQUIRED KEY - PERMUTATIONS - List of valid permutations.
# ----------------------------------------------------------
PERMUTATIONS:
    EVENT:
      - [ h1, h2, h3 ]
    h1:
      - [ b1, b2 ]
    h2:
      - [ b1, b2 ]
    h3:
      - [ b1, b2 ]
```

# Dataset & Input Features

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

- Using 14 TeV pp-collisions simulated HHH6b Madgraph+Py8+Delphes:
  - ~1M events for training+validation; ~300k 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 pT)
- Input jet features:
  - $p_{\mathrm{T}}$ (log-normalized), $\eta$ (normalized), $\sin\phi$, $\cos\phi$, and boolean b-tag score

# Partial vs Complete events

| Inclusive >= 6 AK5 jets | Event proportion |
| --- | --- |
| 0 H | 10% |
| 1 H | 20% |
| 2 H | 50% |
| 3 H | 20% |

- SPANet allows to reconstruct partial events as well as complete events
- 10% of events have 0 Higgs Boson reconstructible
- 20% have 1 Higgs reconstructible and 50% with 2 Higgs reconstructible
- 20% have 3 Higgs Boson reconstructible from 2 AK5 jets

# HHH — Baseline Method

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

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

- Estimate SPA-Net performance improvements with respect to Chi2 baseline

# Baseline Method 1 ($m_H = 125$ GeV) — HHH

Note: using top 6 jets in each event ordered by b-tag and pT

| Event Type | Event Purity | H Purity |
|---|---|---|
| 1-3 H | 22% | 39% |
| 3H | 23% | 43% |

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

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

- For 1-3 reconstructible Higgs, **22%** of events correctly reconstructed; **39%** of Higgs correctly reconstructed
- For 3 reconstructible Higgs, **23%** of events correctly reconstructed; **52%** of Higgs correctly reconstructed

# SPA-Net — HHH

| Event Type | Event Purity | H Purity |
|---|---|---|
| 1-3 H | 34% | 52% |
| 3H | 38% | 58% |

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

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

- For 1-3 reconstructible Higgs, **34% (+54%)** of events correctly reconstructed; **52% (+33%)** of Higgs correctly reconstructed
- For 3 reconstructible Higgs, **38% (+65%)** of events correctly reconstructed; **58% (+34%)** of Higgs correctly reconstructed

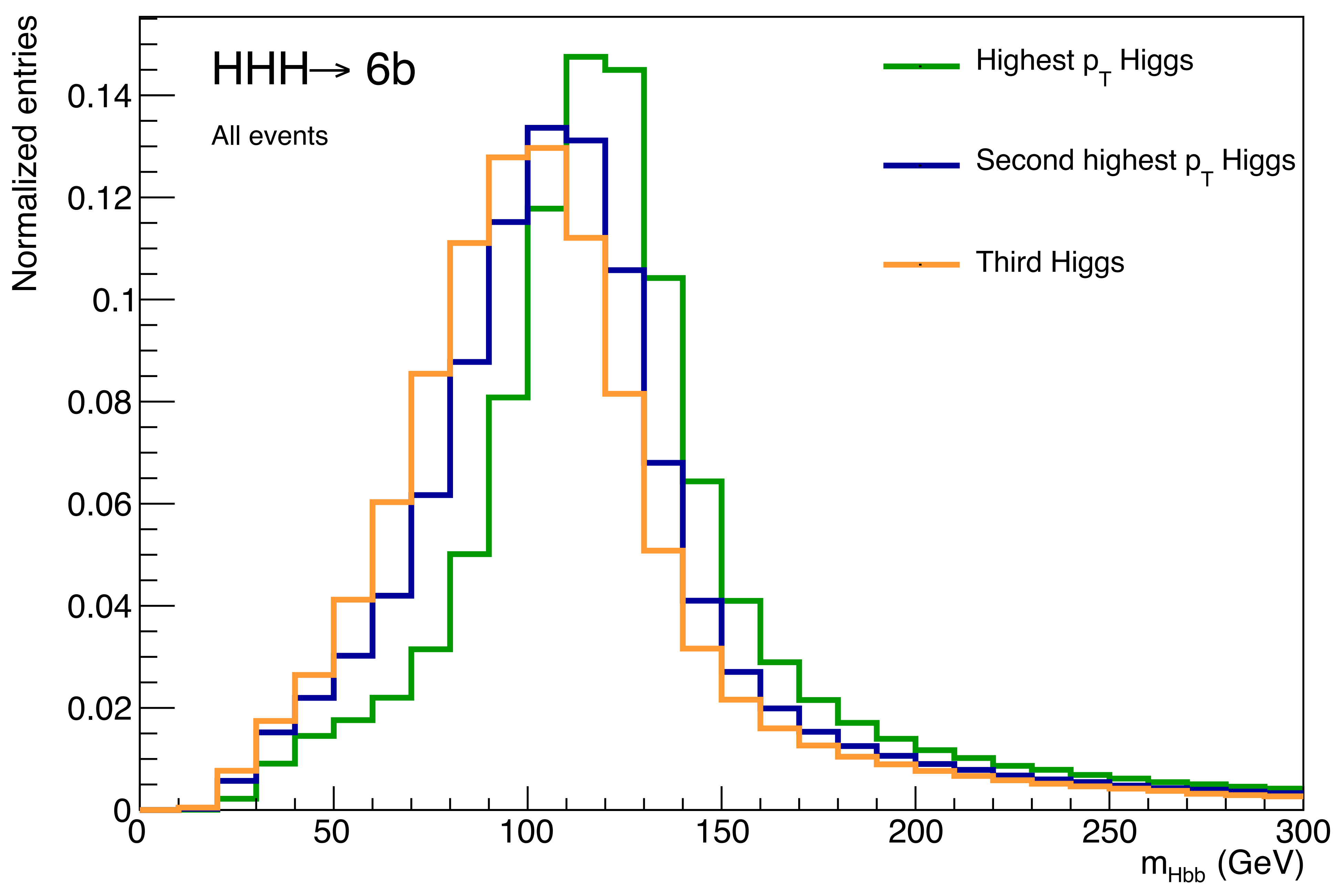
# SPA-Net — HHH

| H purity | 1-3h | | 3h | |
|---|---|---|---|---|
| | Chi2 baseline | SPA-Net | Chi2 baseline | SPA-Net |
| 6 jets | 46.3% | 53.4% (**+15%**) | 64.1% | 67.8% (**+6%**) |
| 7 jets | 40.2% | 53.3% (**+33%**) | 47.0% | 61.9% (**+32%**) |
| >= 8 jets | 33.8% | 49.8% (**+47%**) | 36.4% | 54.2% (**+49%**) |
| Full | 38.6% | 51.7% (**+33%**) | 43.1% | 58.2% (**+35%**) |

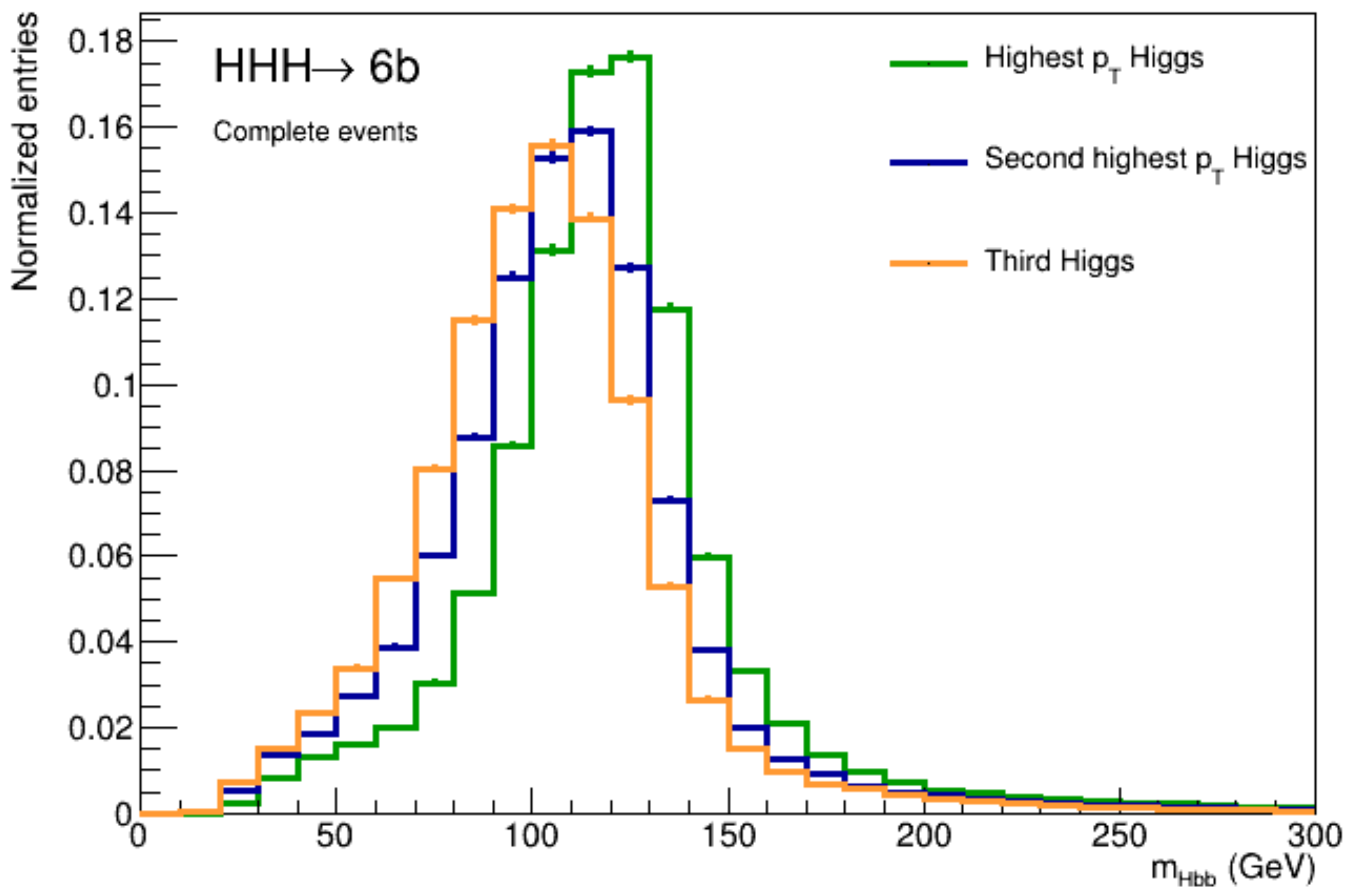
- Performance improvements in H purity:
  - Largest improvements observed with higher number of jets
  - SPA-Net: achieves similar H purity in different jet multiplicity categories

# Mass Reconstruction

1-3h events

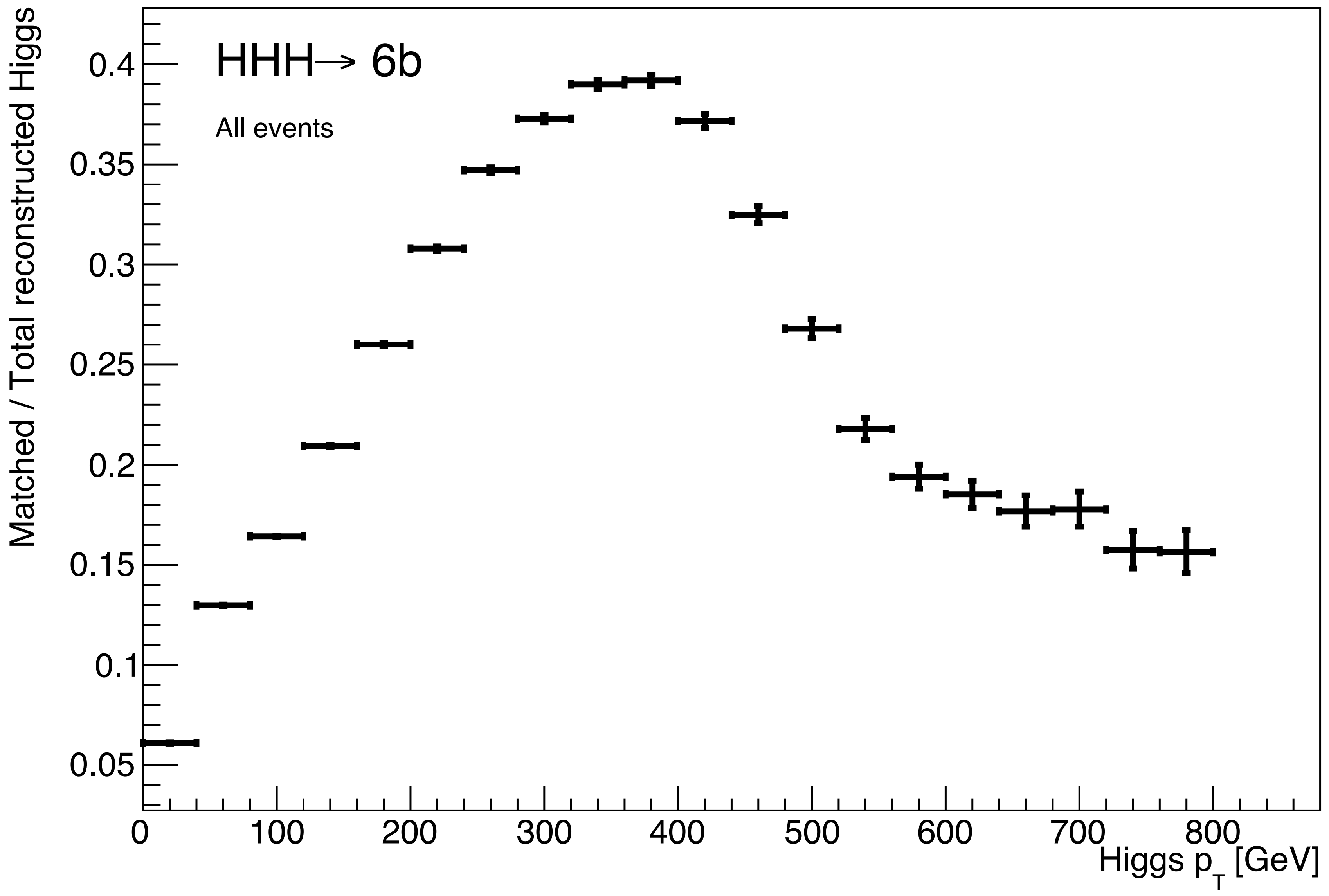



3h events



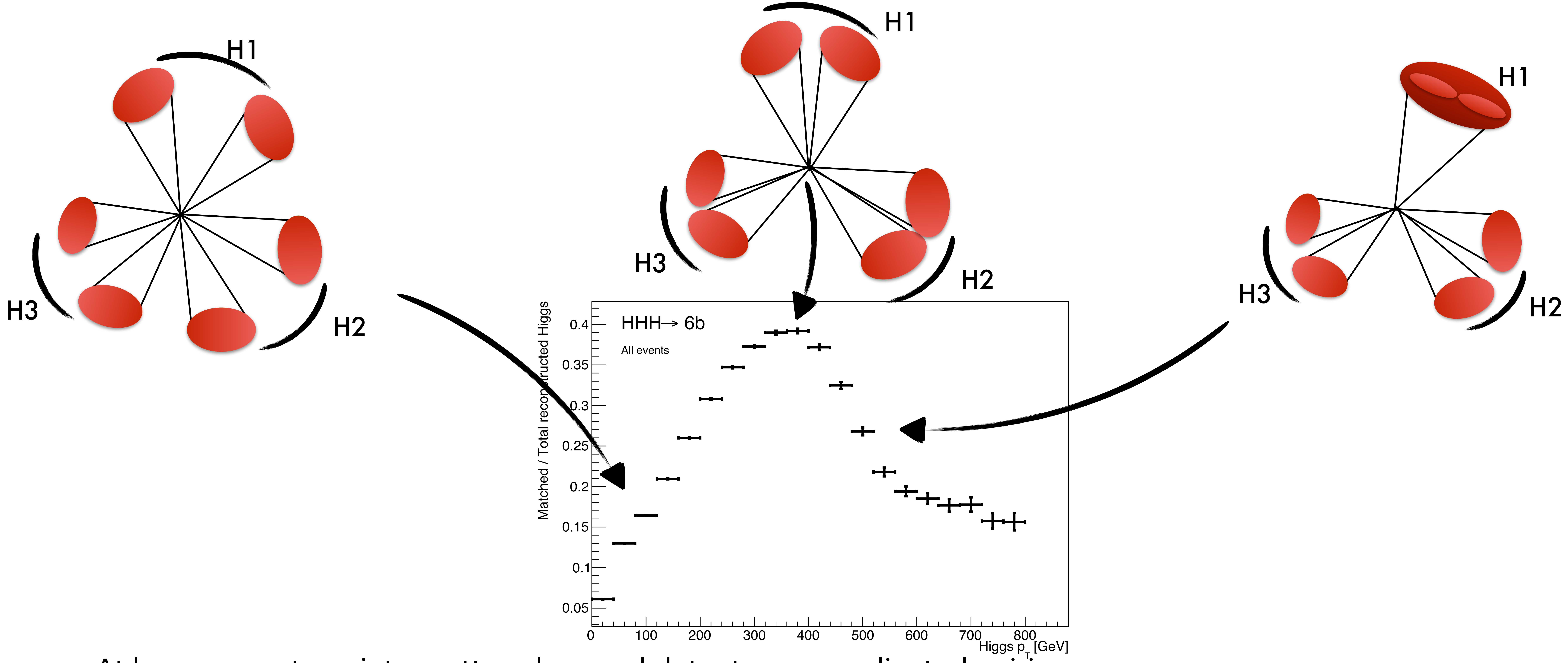

- SPA-Net reconstructs the mass of each Higgs candidate appropriately
- Work in progress: investigating mass sculpting of backgrounds

# Differential matching efficiency



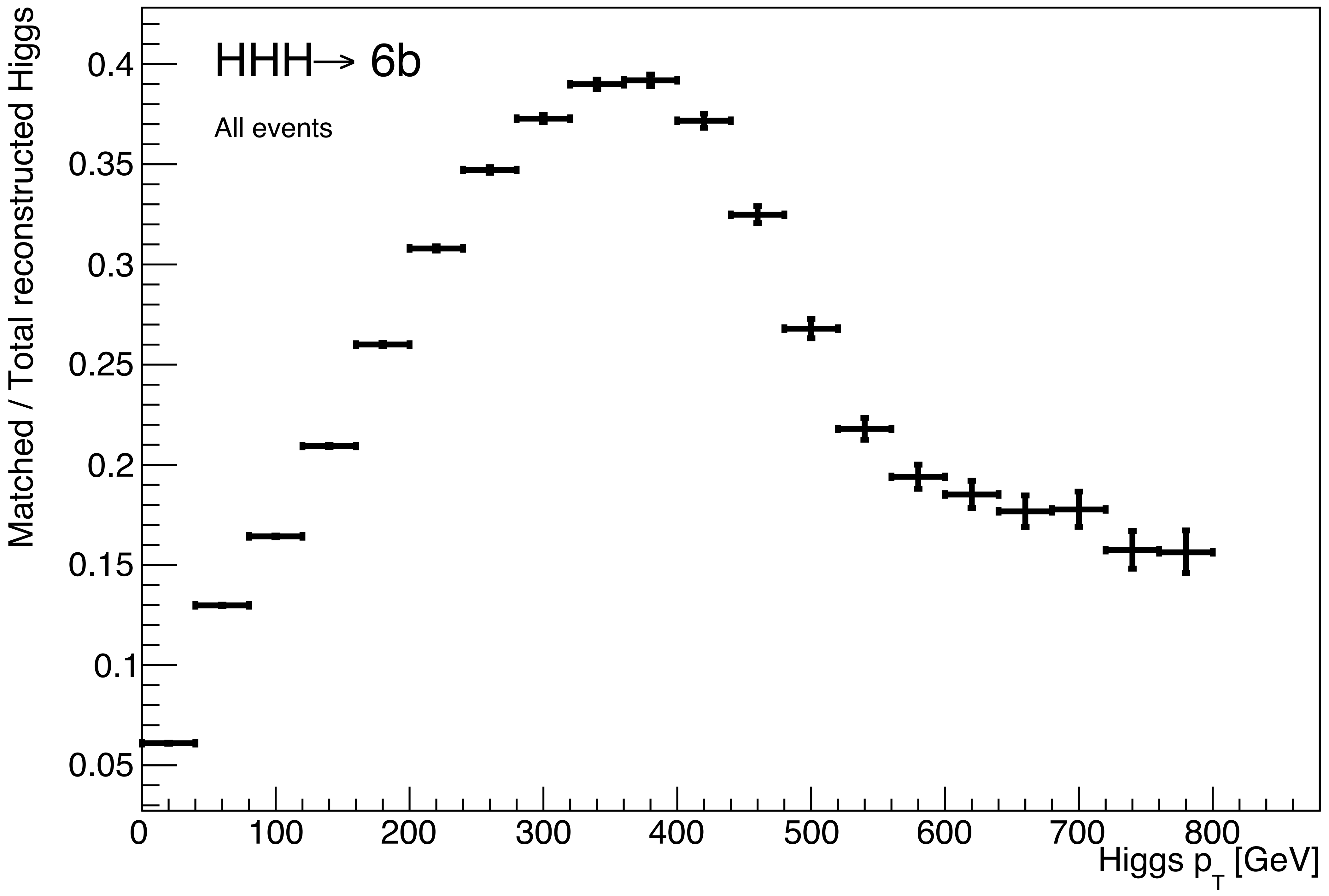

- Matching efficiency: strong dependence on momentum of the Higgs bosons

# Differential matching efficiency

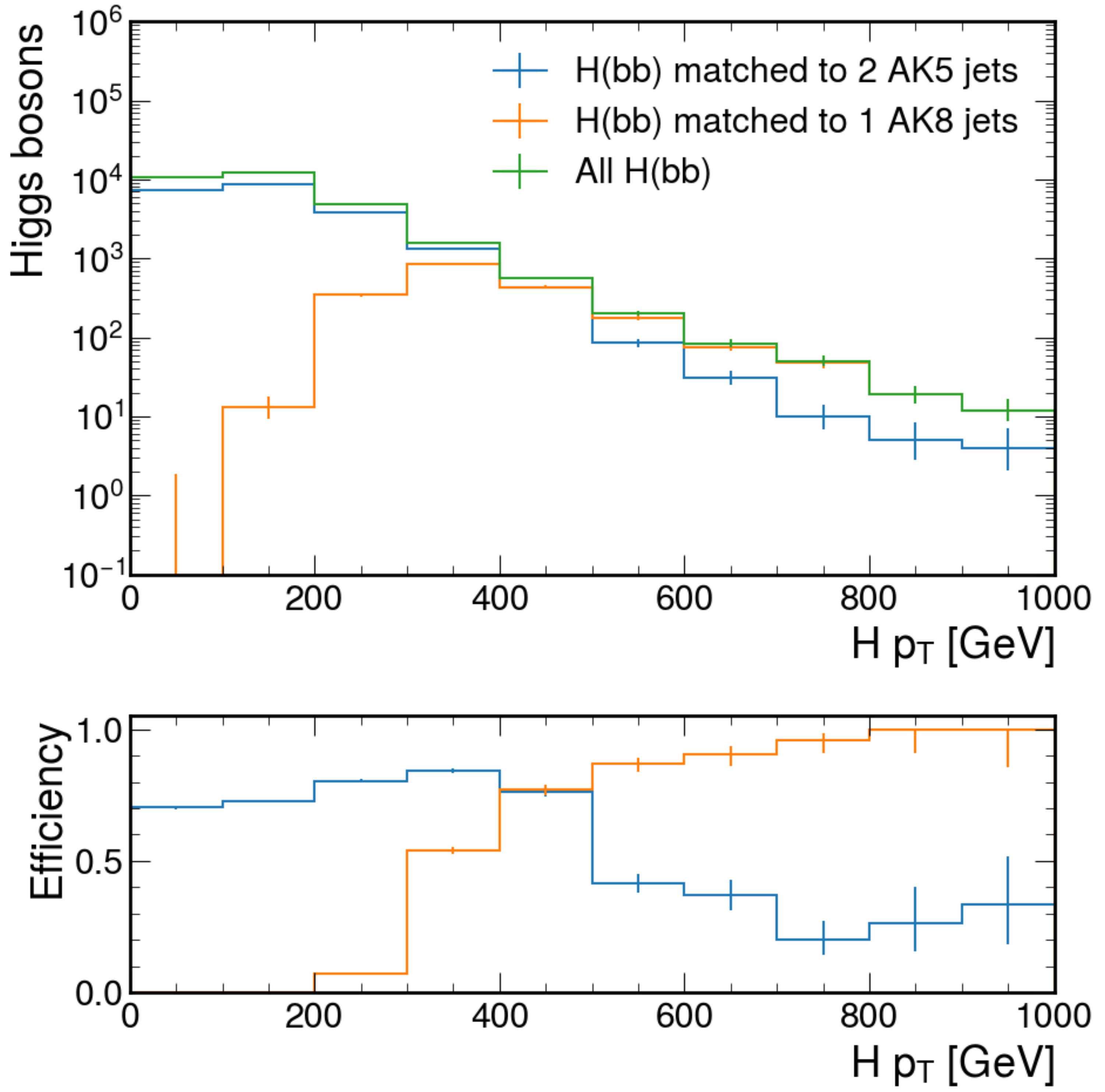



- At low momentum, jets scattered around detector = complicated pairing
- At higher momentum, jets from Higgs boson more and more collimated, clearer correlation
- At very high momentum ( pT > 400 GeV), matching efficiency drops and Higgs reconstructed in AK8 jets

# Differential matching efficiency



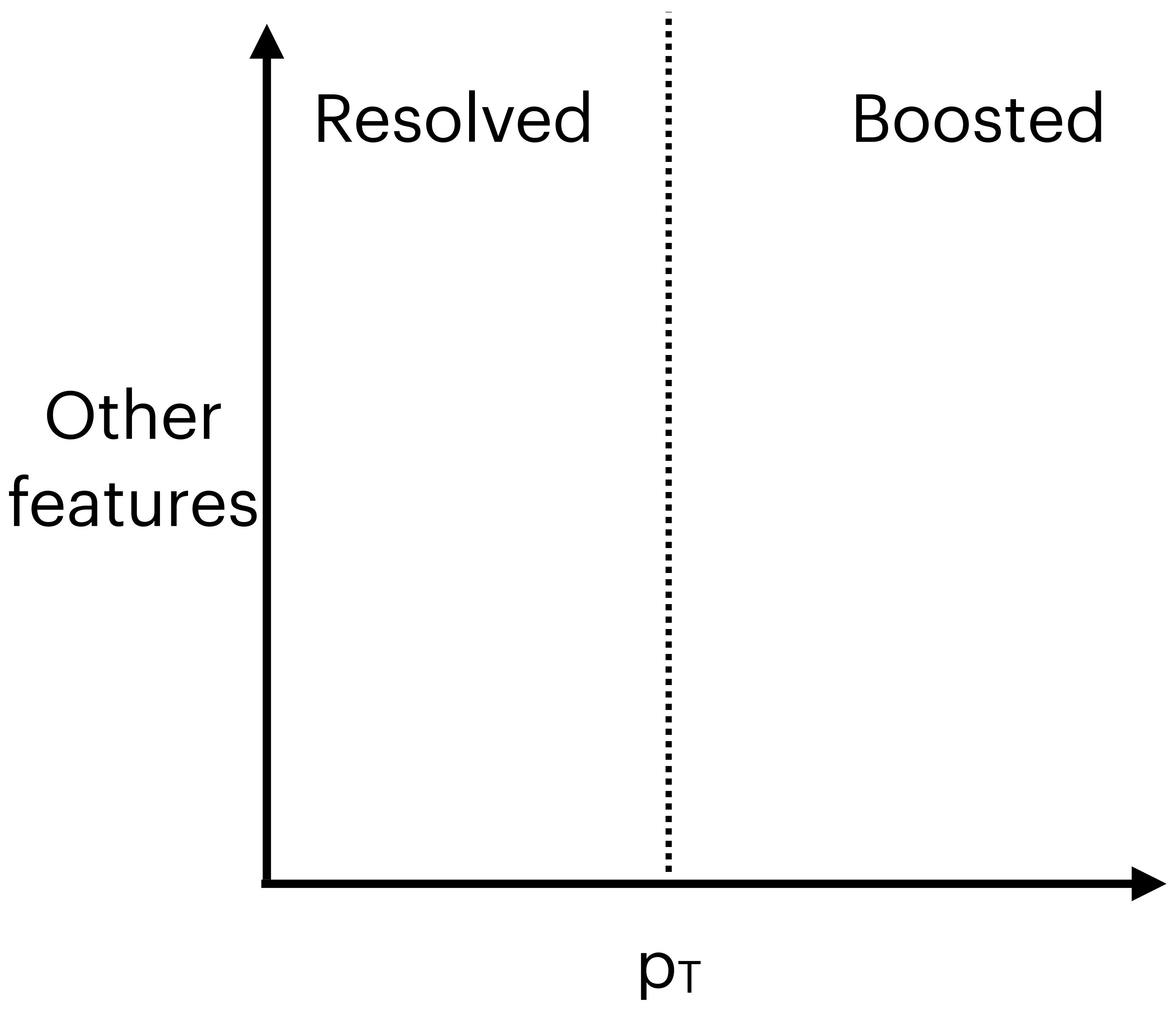

Truth level: Higgs reconstructed in 2 AK5 vs 1 AK8

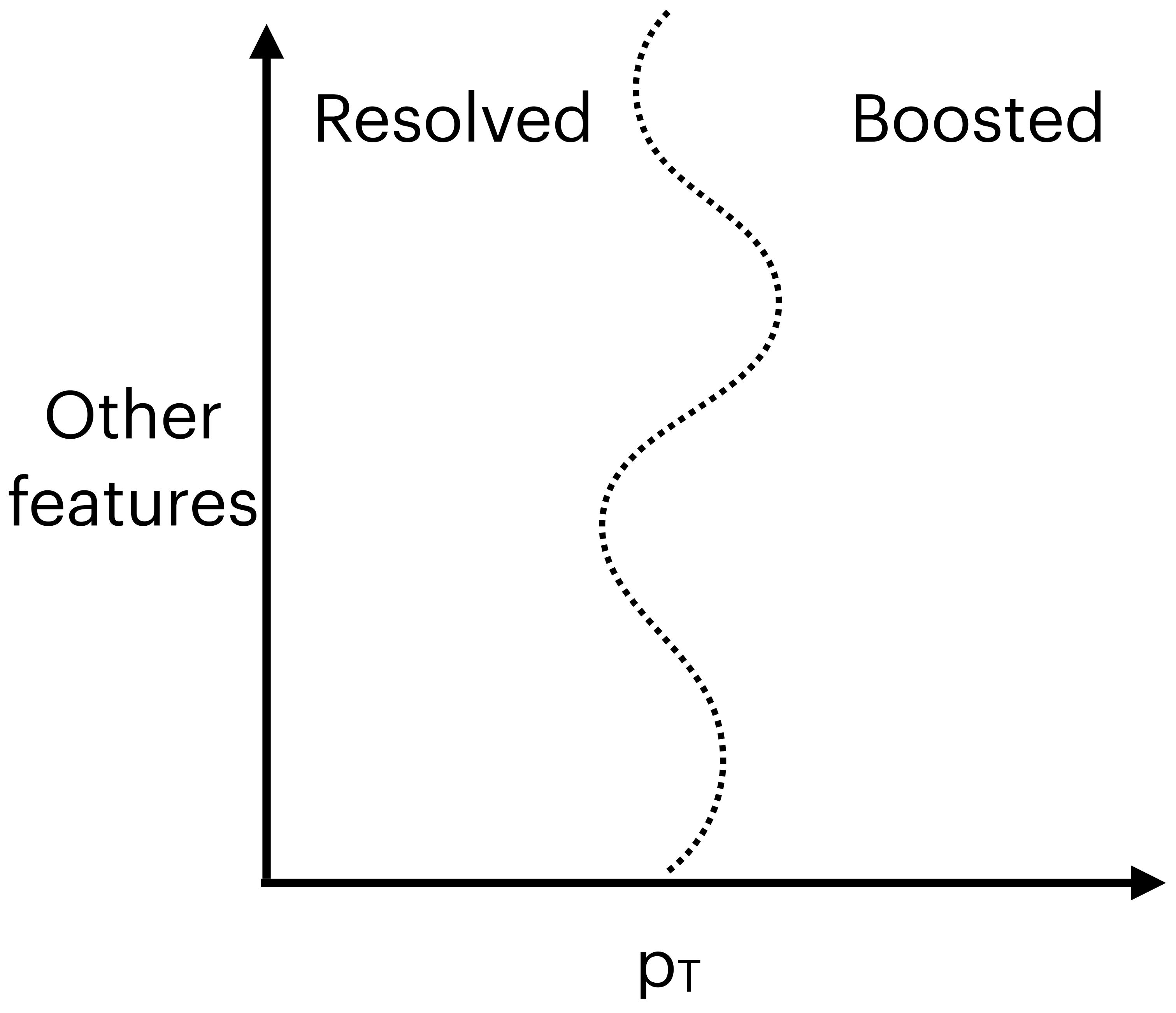



- At 400 GeV, Higgs more likely to be reconstructed in 1 AK8 than 2 AK5
- Optimal performance: generalize approach to both boosted + resolved topologies

# Boosted + Resolved topologies

- Generalizing to **boosted + resolved** topologies
- Idea is to help us divide phase space between boosted and resolved in more intelligent way



# Event Configuration: resolved + boosted

- Comparison boosted+resolved configurations $HHH \to (b\bar{b})(b\bar{b})(b\bar{b})$

```
# ---------------------------------------------------------------------
# REQUIRED - EVENT - Complete list of resonance particles and daughters.
# ---------------------------------------------------------------------
EVENT:
  t1:
    - q1: Jets
    - q2: Jets
    - b:  Jets
  t2:
    - q1: Jets
    - q2: Jets
    - b:  Jets
# ---------------------------------------------------------
# REQUIRED KEY - PERMUTATIONS - List of valid permutations.
# ---------------------------------------------------------
PERMUTATIONS:
    EVENT:
      - [ t1, t2 ]
    t1:
      - [ q1, q2 ]
    t2:
      - [ q1, q2 ]
```

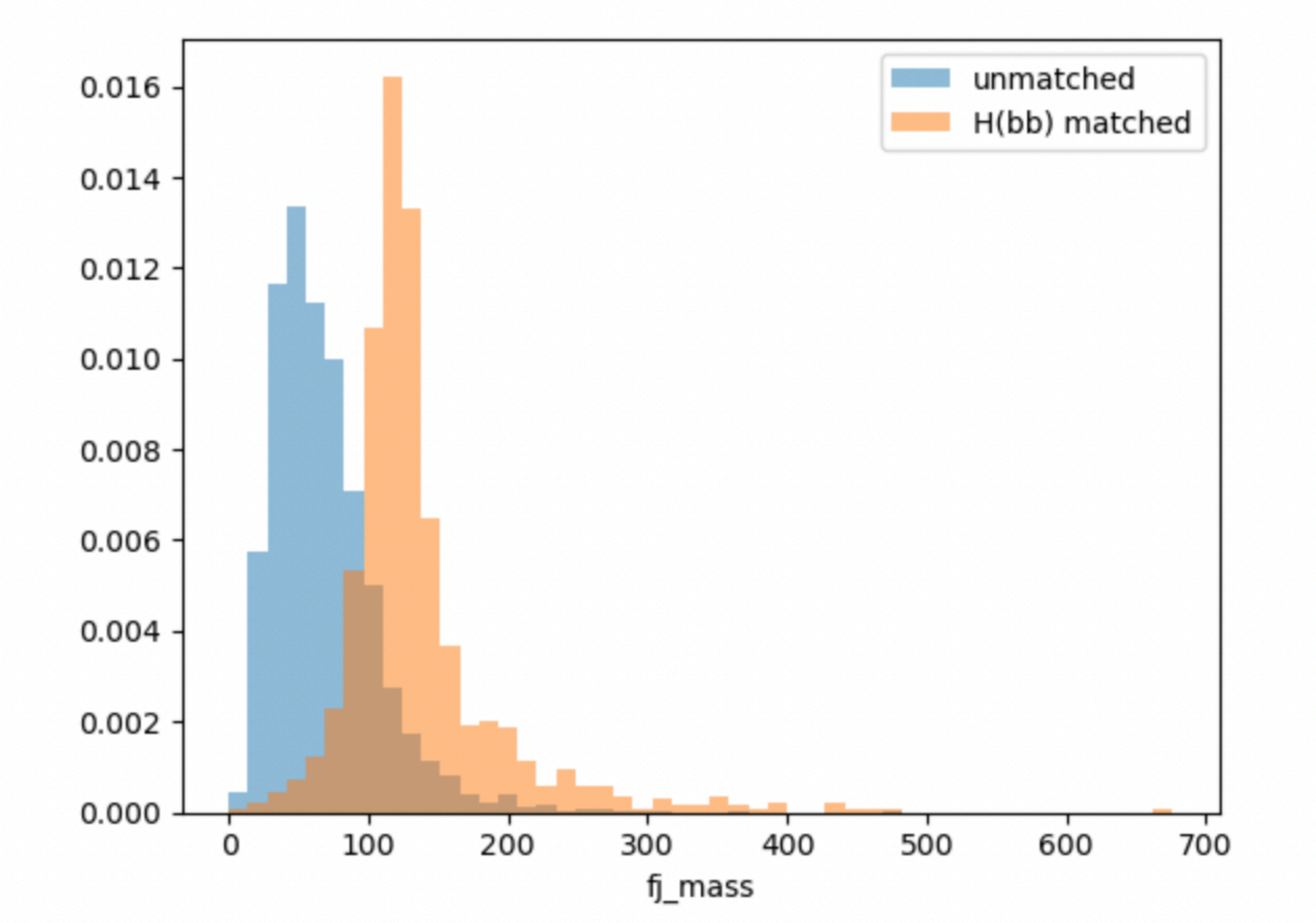
```
# ---------------------------------------------------------------------
# REQUIRED - EVENT - Complete list of resonance particles and daughters.
# ---------------------------------------------------------------------
EVENT:
  h1:
    - b1: Jets
    - b2: Jets
  h2:
    - b1: Jets
    - b2: Jets
  h3:
    - b1: Jets
    - b2: Jets
  bh1:
    - bb: BoostedJets

  bh2:
    - bb: BoostedJets

  bh3:
    - bb: BoostedJets

# ---------------------------------------------------------
# REQUIRED KEY - PERMUTATIONS - List of valid permutations.
# ---------------------------------------------------------
PERMUTATIONS:
    EVENT:
      - [[h1, h2], [bh1,bh2]]

      - [[h1, h3], [bh1,bh3]]

      - [[h2, h3], [bh2,bh3]]

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

**All permutations of AK5 and AK8 jets**

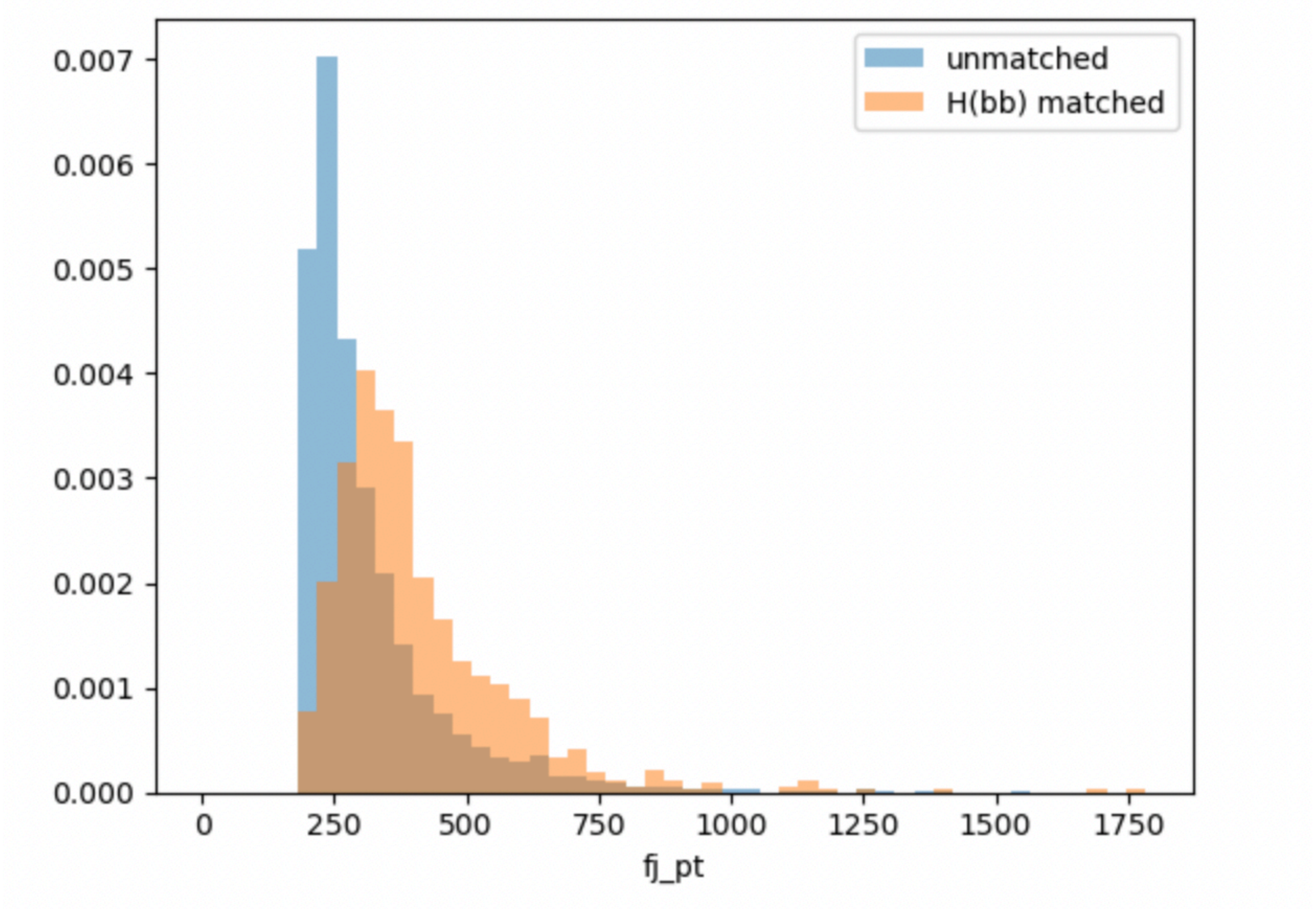
# Boosted variables

- Use boosted jets mass, pT, eta, phi, and subjettiness variables as inputs

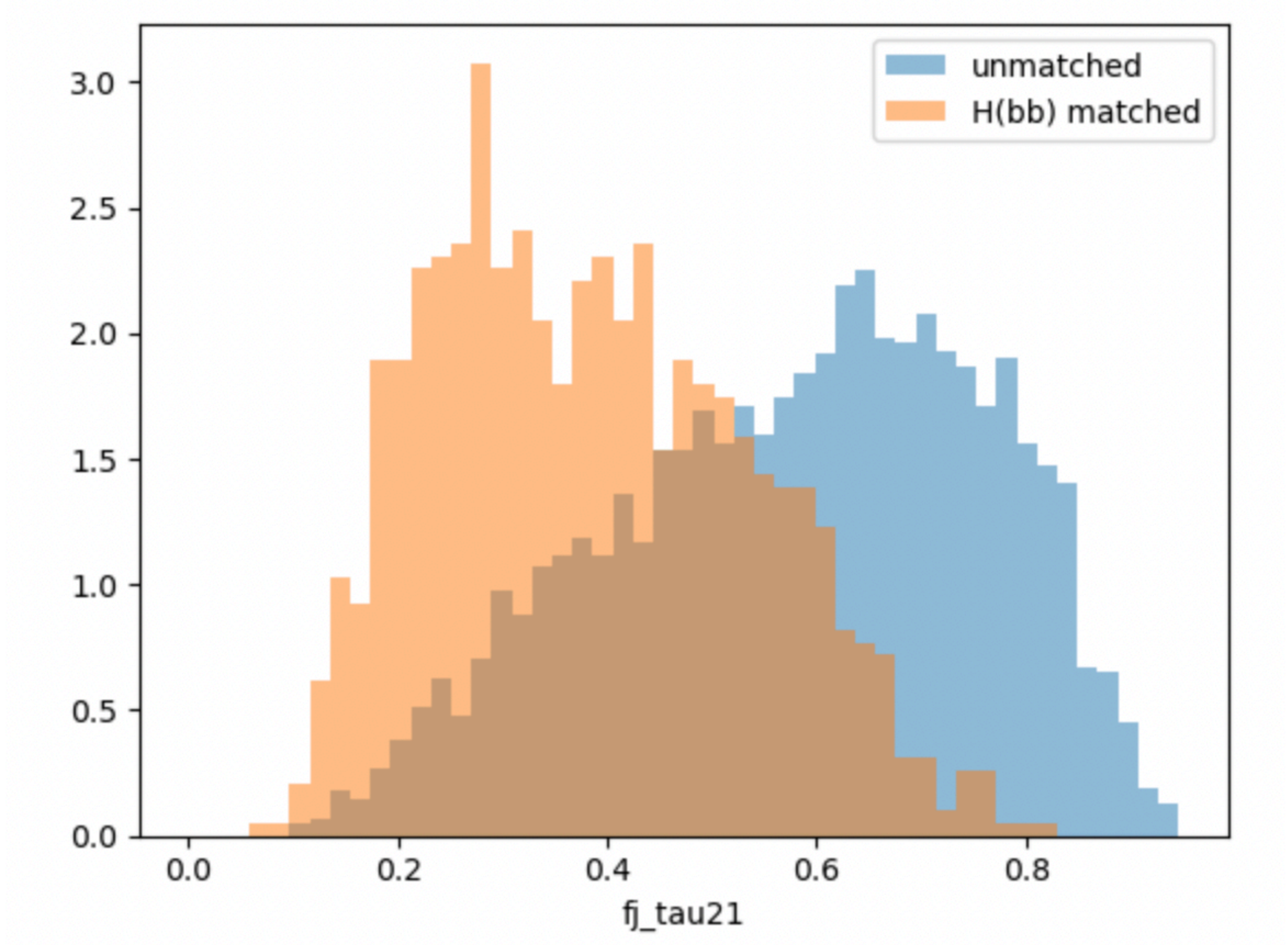
Mass




pT




Subjettiness




- On-going work to include both resolved and boosted jets in one training
    - Currently studying network hyperparameters and metrics to compare to resolved
    - Stay tuned: more results will be published in the paper

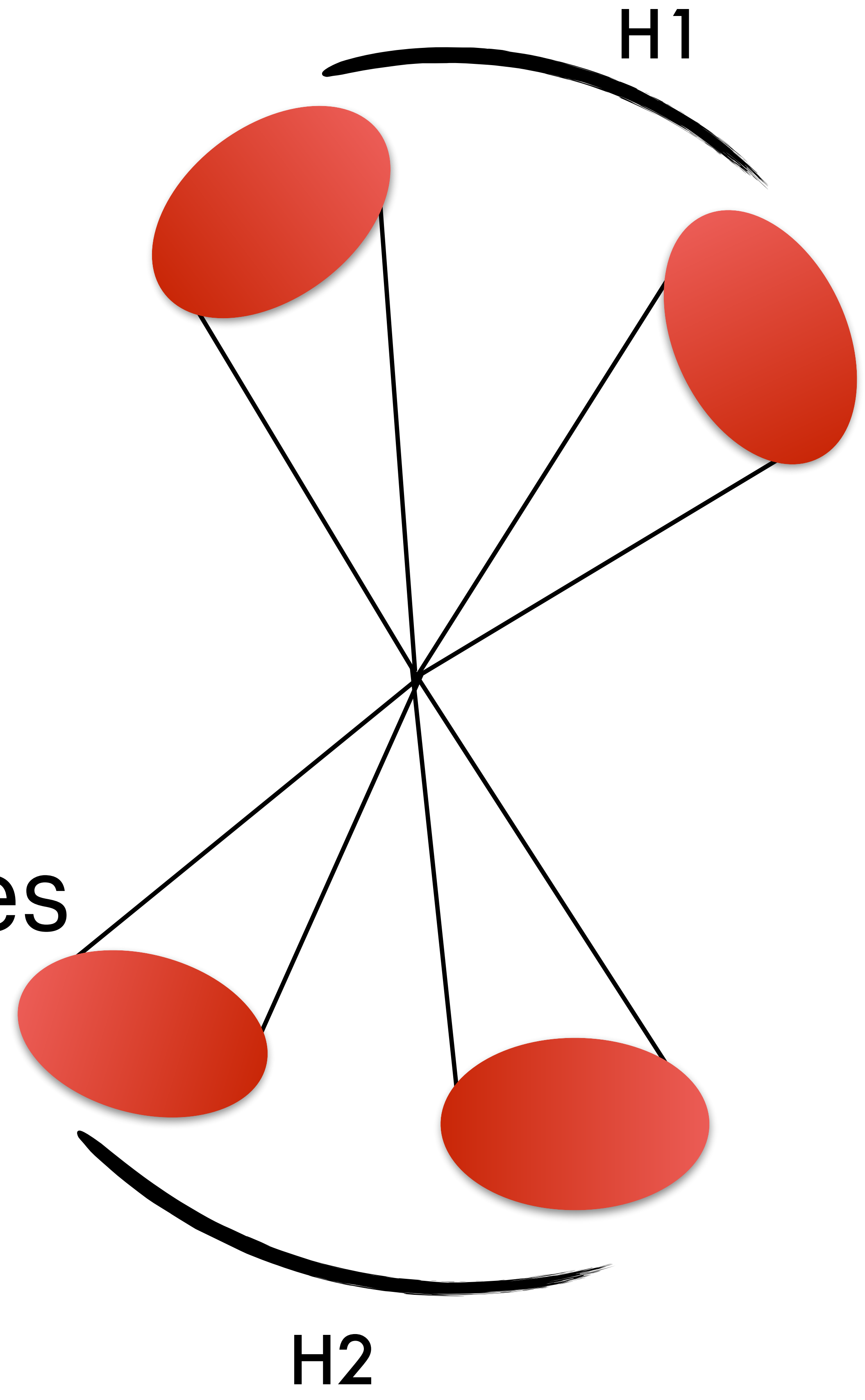
# HH pairing comparisons

- 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:
  - Find pairs based on minimal distance between 2 Higgs masses

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



# Baseline Method 1 ($m_H$ = 125 GeV) — HH

Note: using top 4 jets in each event

| Event Type | Event Purity | H Purity |
|---|---|---|
| 1-2 H | 44% | 57% |
| 2H | 21% | 53% |

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

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

- For 1-2 reconstructible Higgs, **44%** of events correctly reconstructed; **57%** of Higgs correctly reconstructed
- For 2 reconstructible Higgs, **21%** of events correctly reconstructed; **53%** of Higgs correctly reconstructed

# SPA-Net — Resolved HH training

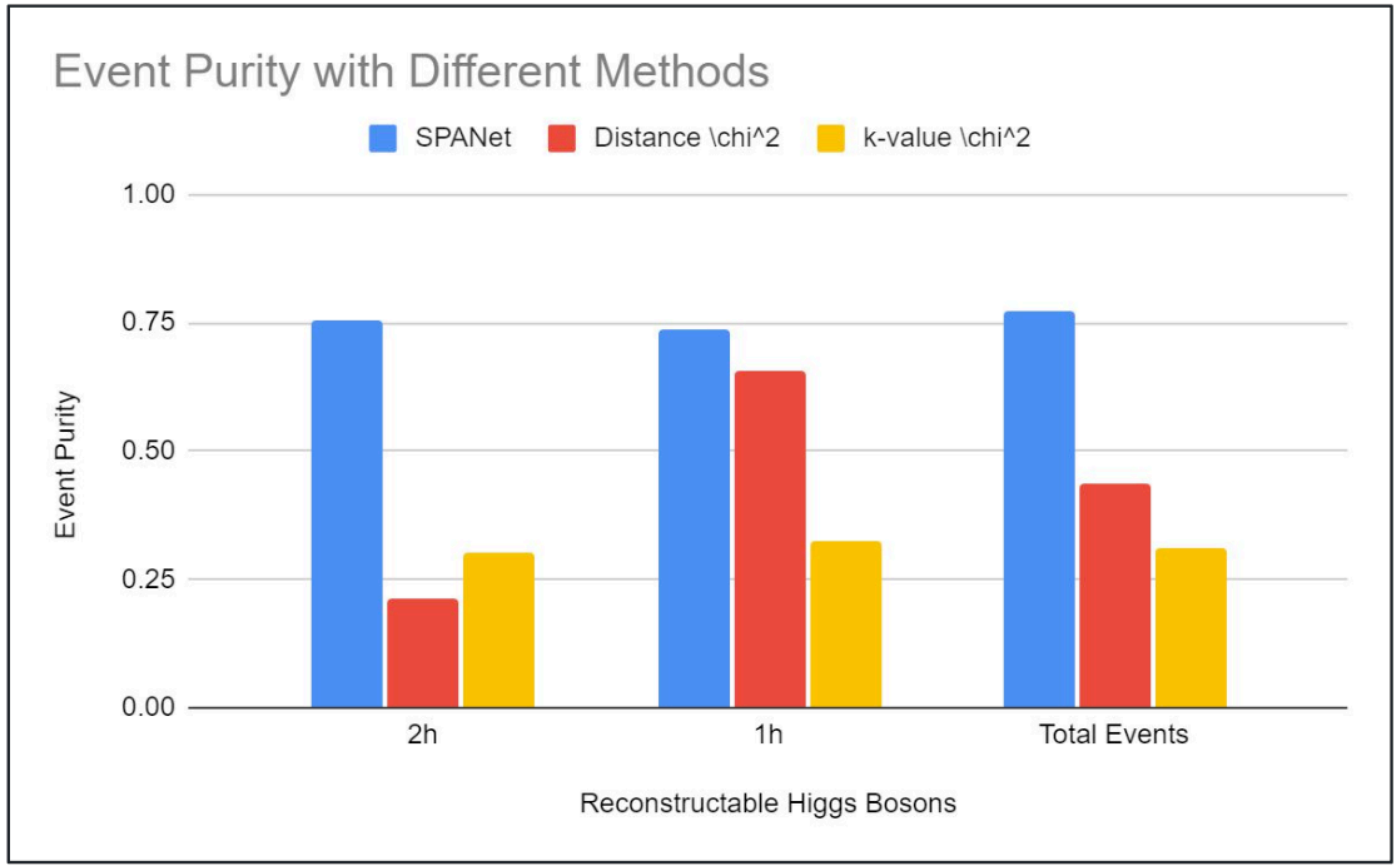
| Event Type | Event Purity | H Purity |
|---|---|---|
| 1-2 H | 76% | 81% |
| 2H | 77% | 84% |

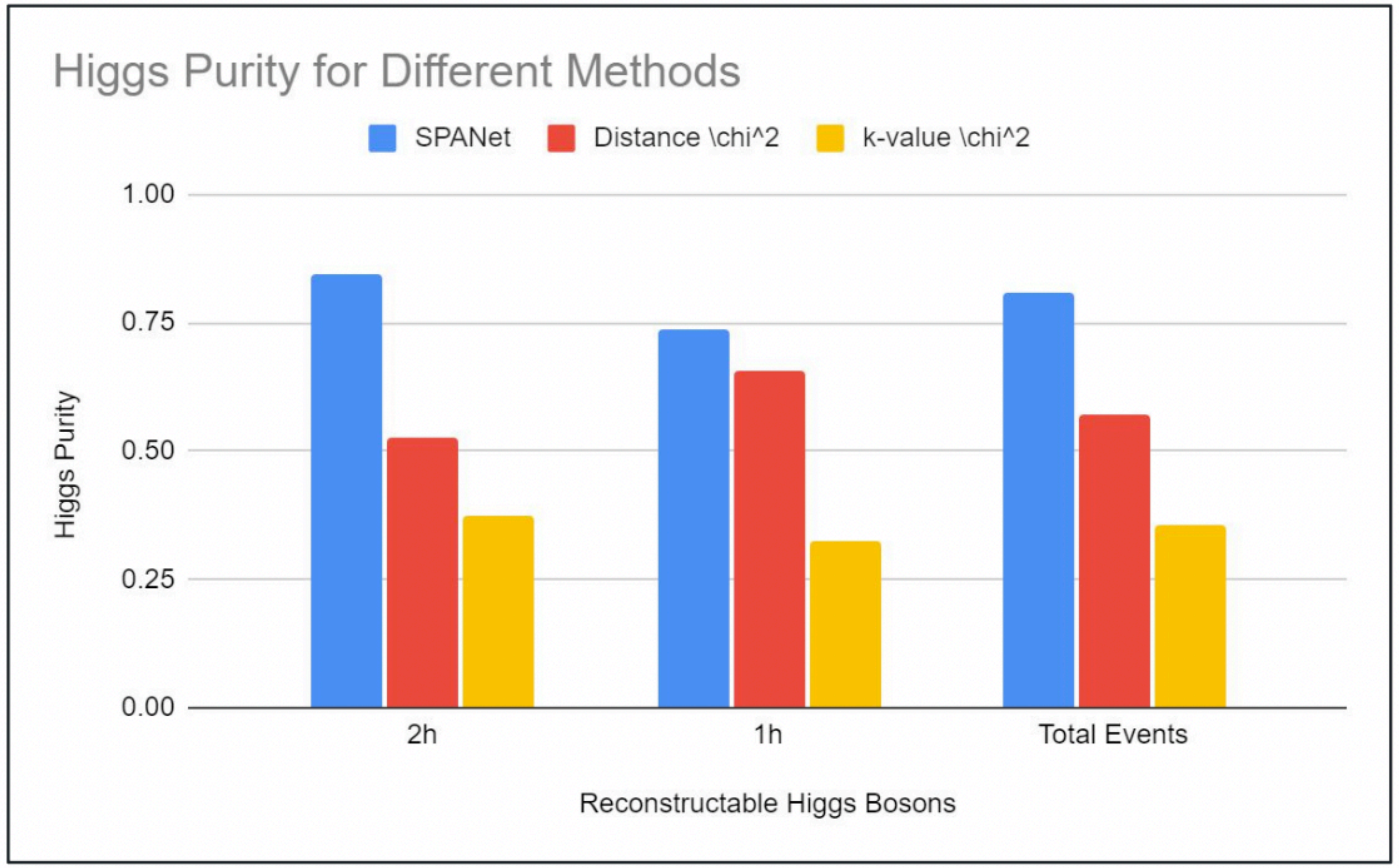
$$\text{Event Purity} = \frac{\text{Number of events that all Higgs are reconstructed}}{\text{Total number of events}}$$

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

- For 1-2 reconstructible Higgs, **76% (+72%)** of events correctly reconstructed; **81% (+42%)** of Higgs correctly reconstructed
- For 2 reconstructible Higgs, **77% (+360%)** of events correctly reconstructed; **84% (+58%)** of Higgs correctly reconstructed - result to be understood in context of analysis, limitations of Delphes…

# Method comparison

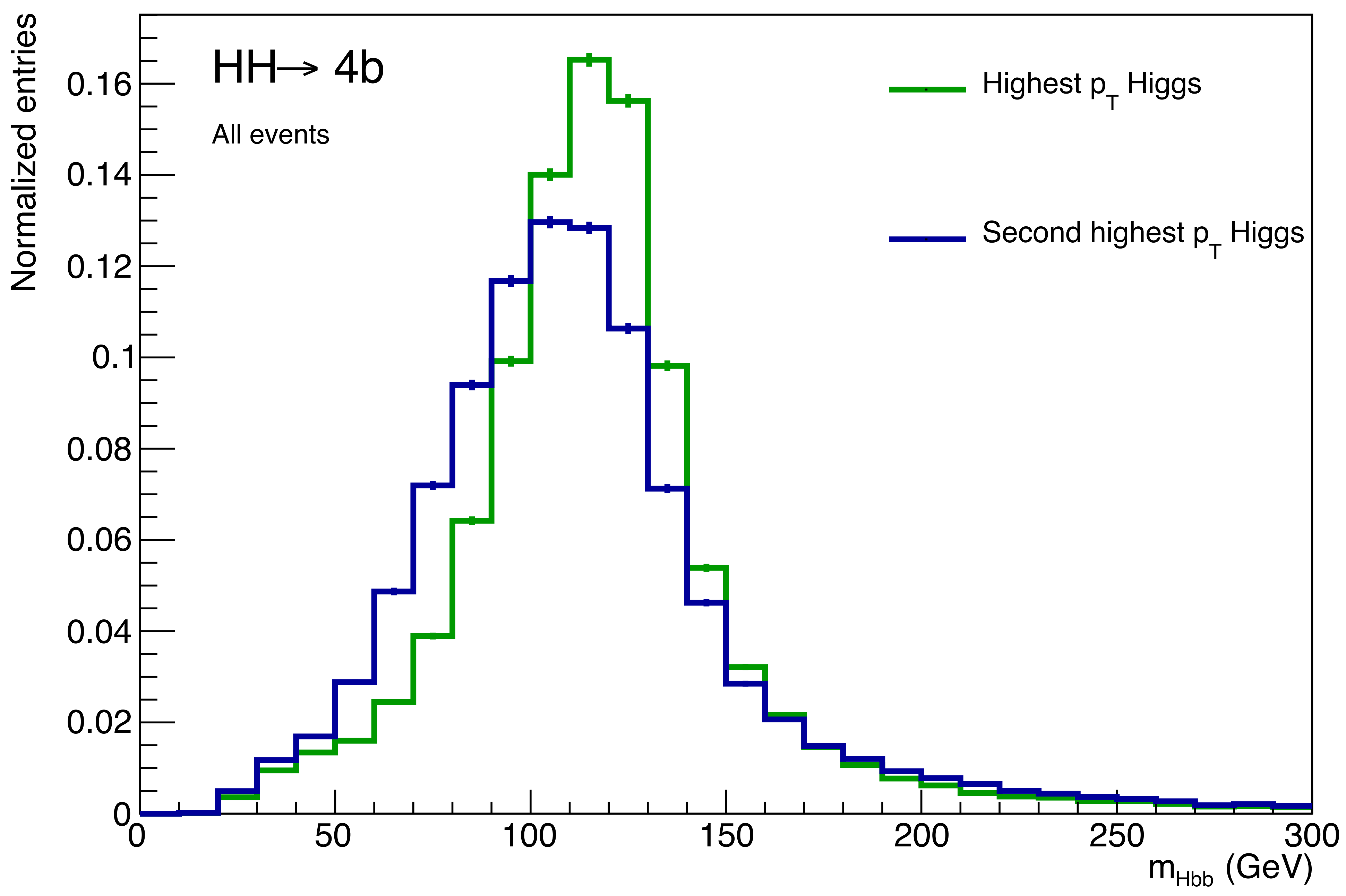


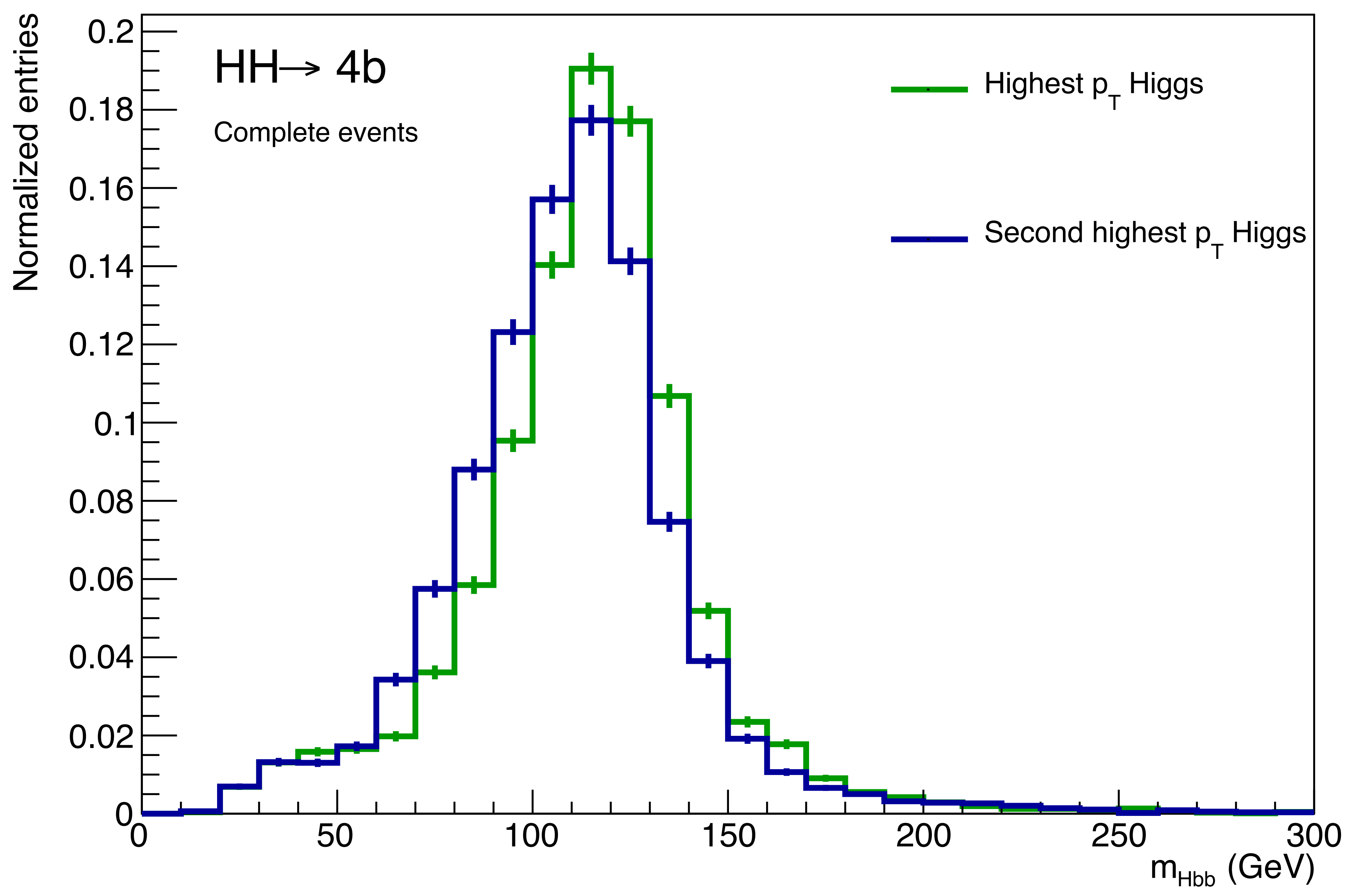

- SPA-Net: clear performance improvements with respect to both baselines
- Short-comings: chi2 methods consider only 4 jets so far, boolean b-tagging
  - Likely lower improvements than in analysis specific reconstruction
  - However, great potential for SPA-Net to improve HH and HHH analyses

# Mass Reconstruction HH

1-2h events



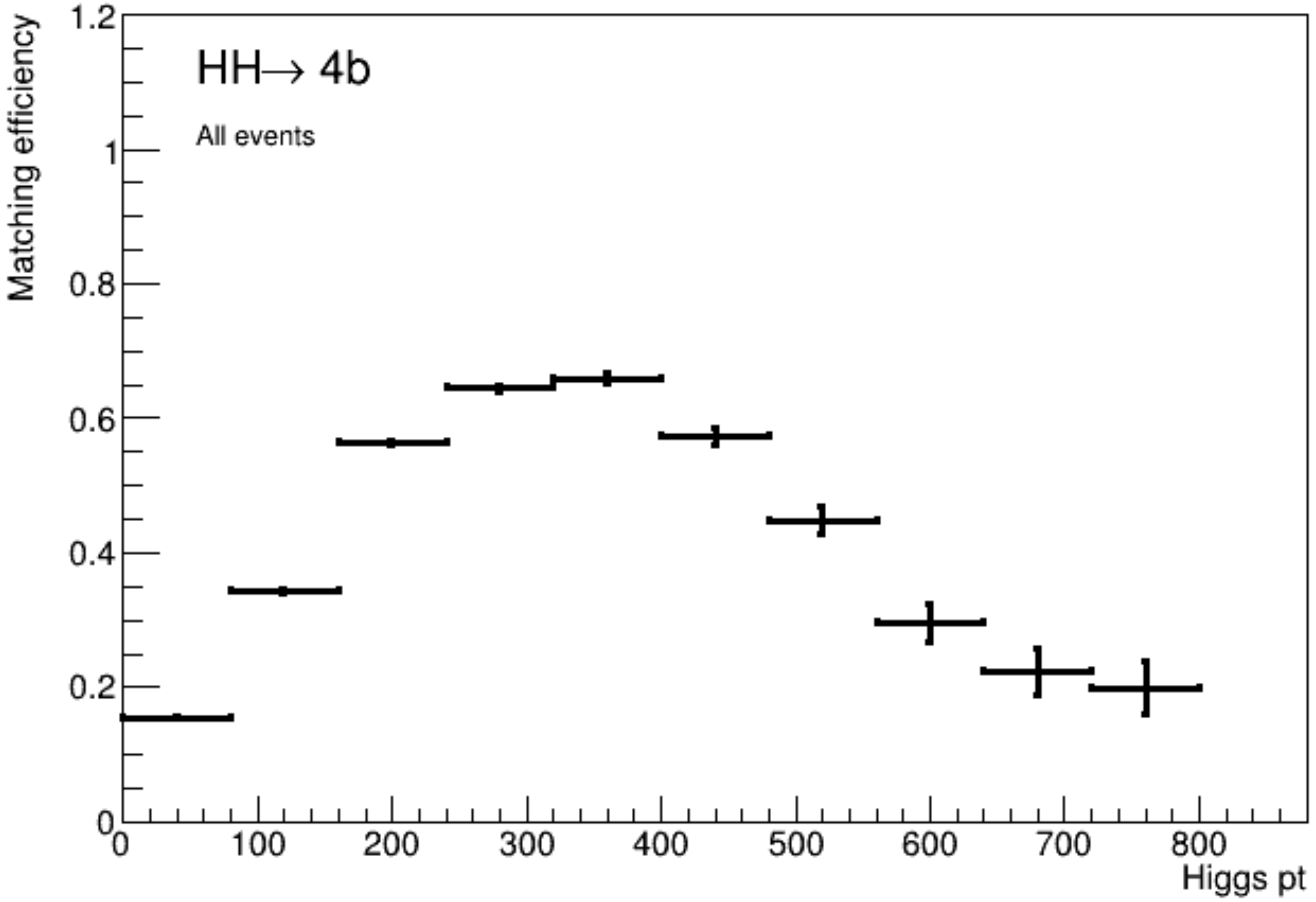

2h events



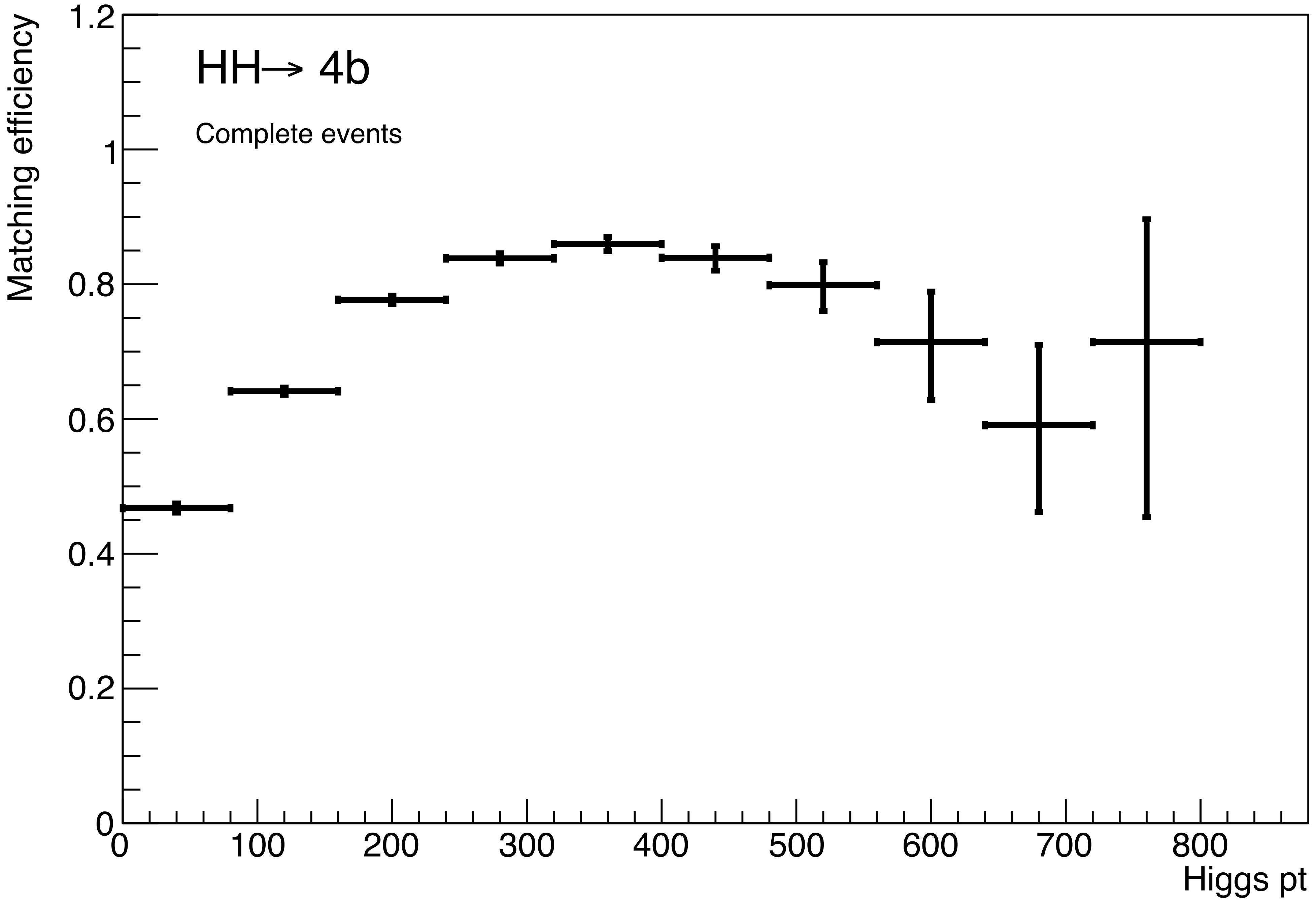

- SPA-Net reconstructs the mass of each Higgs candidate appropriately
- Work in progress: investigating mass sculpting of backgrounds
- Investigating boosted + resolved reconstruction to validate SPA-Net method

# 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

# 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

- Stay tuned: paper being finalised and will be published soon!

# Backup

# 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

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

# Generalization to Boosted & Resolved

- $N_j$ : number of AK4 jets; $N_{fj}$ : number of AK8 jets
- Two sets of reconstruction targets: [h1, h2] and [bh1 (boosted H 1), bh2]
  - Output AK4 jet AD tensor is a $N_j$ x $N_j$ symmetric matrix,
  - Output AK8 jet AD tensor is a vector of size $N_{fj}$.
- `{h_i}` and `{bh_i}` are the same set of particles ordered by pT, so `h_i` and `bh_i` have one-to-one correspondence.
- `bh_i` will be selected if it is detected. If not, `h_i` will be selected if it is detected.

# Model Configuration

- Many hyperparameters to tune!
- We suggested 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: 0.2
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
```