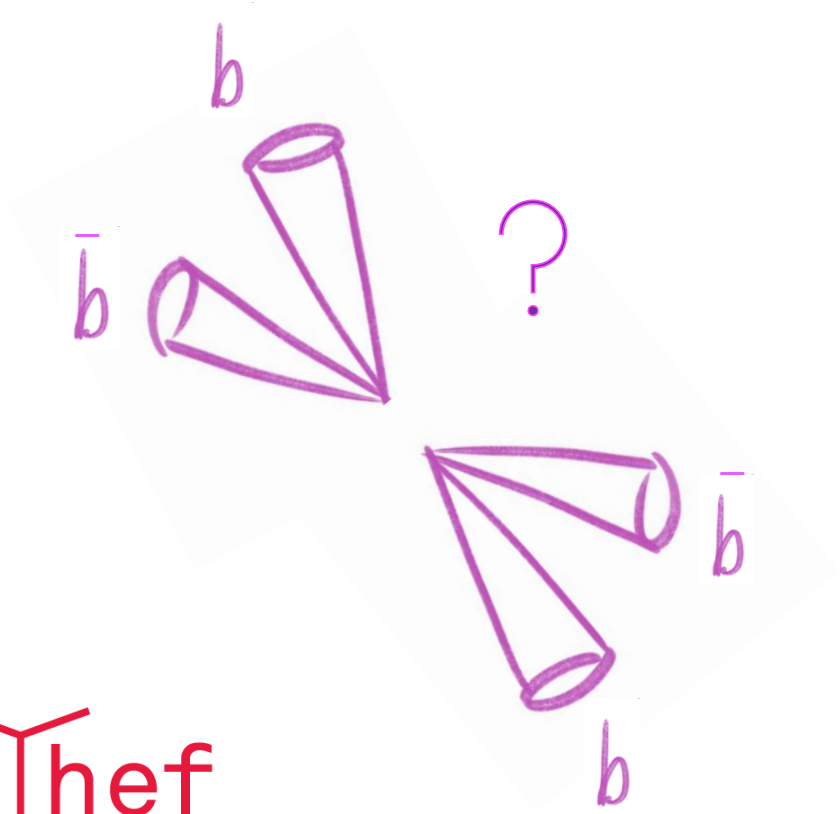
Reconstruction of Higgs Boson Pairs in the 4b Final State

Which Jet Belongs to Which Higgs?



Oza Passot

University of Amsterdam – NIKHEF
DESY Summer Student 2025 – CMS Higgs

Supervisors: Suman Chatterjee, Rainer Mankel

07/11/2025

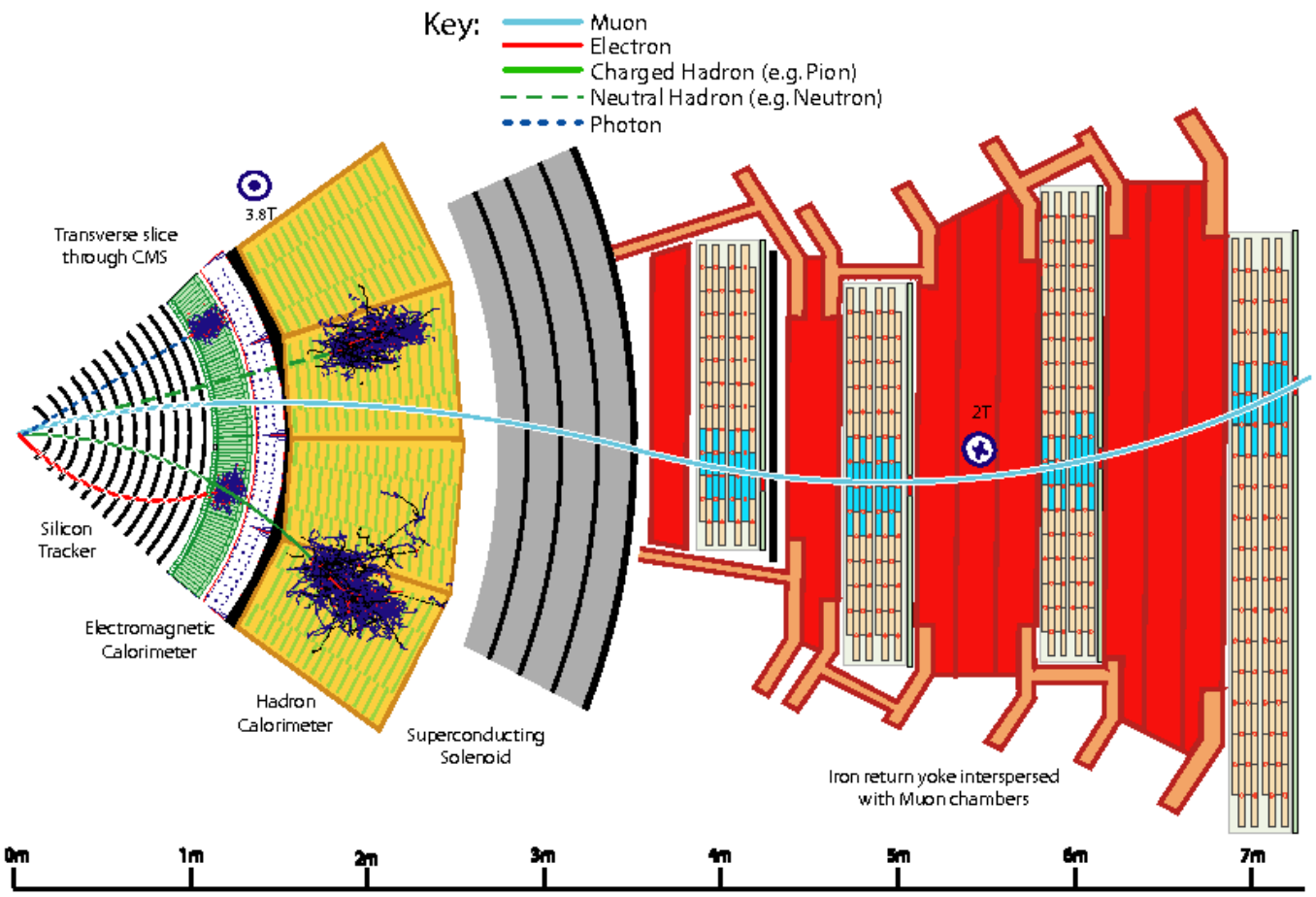
HELMHOLTZ
DESY.

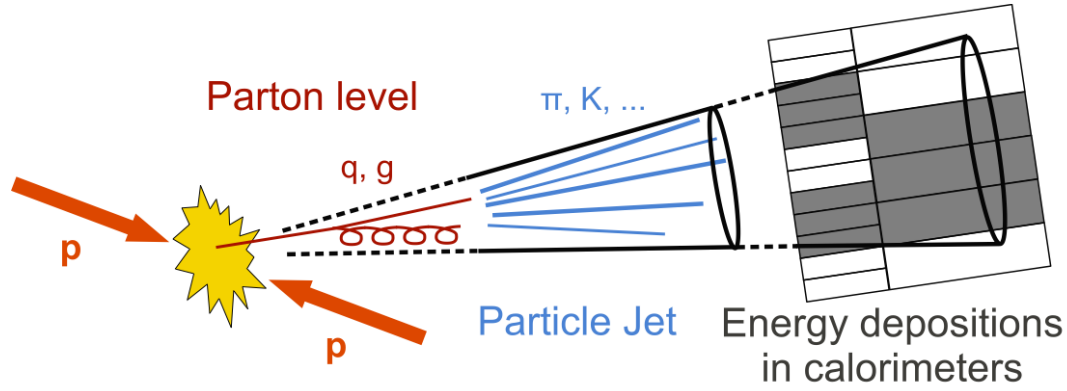
CMS
Compact Muon Solenoid

DESY.

Nikhef

Compact Muon Solenoid (CMS)

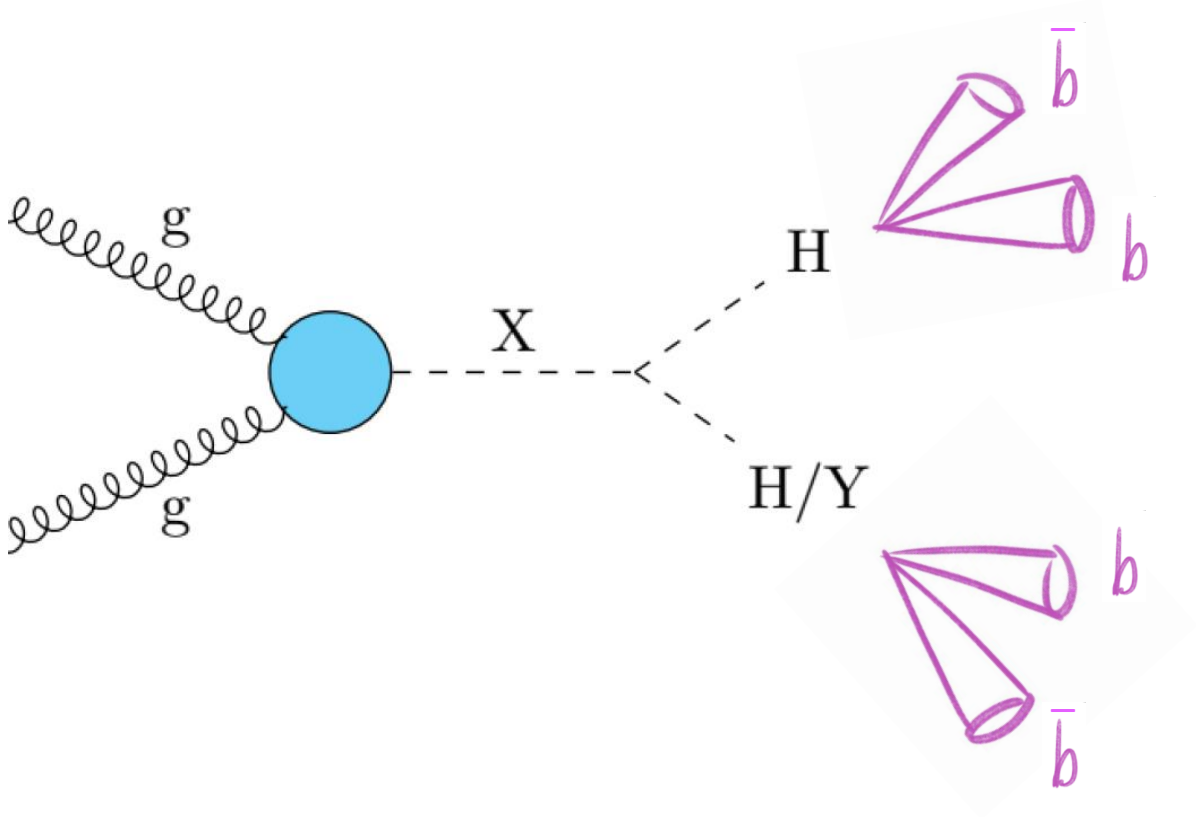


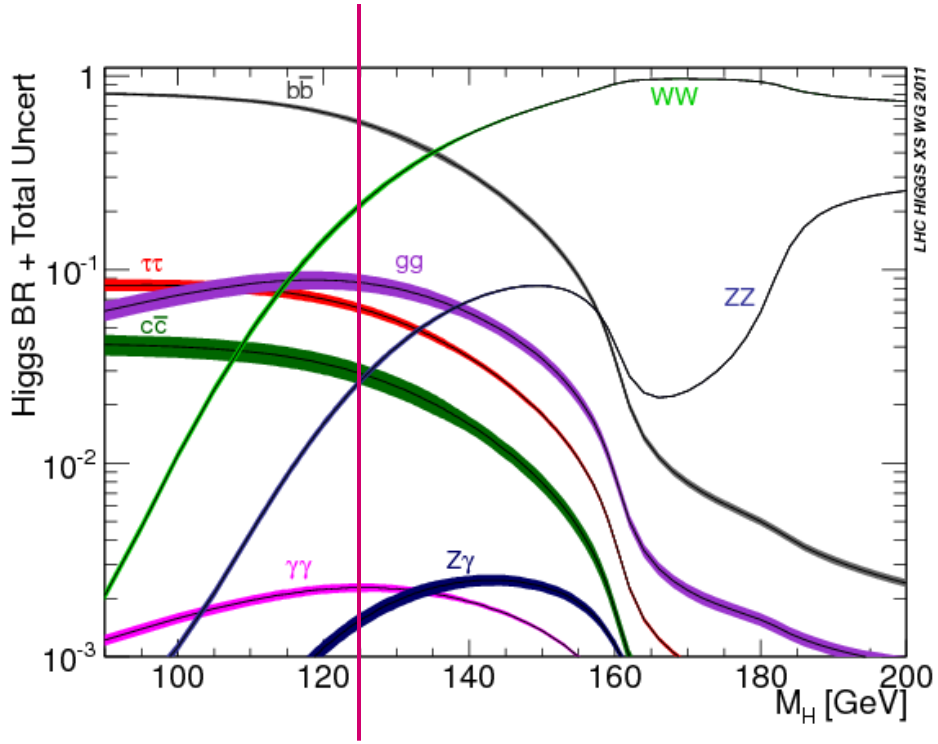


Detector does not see individual quarks but **jets** → shower of particles

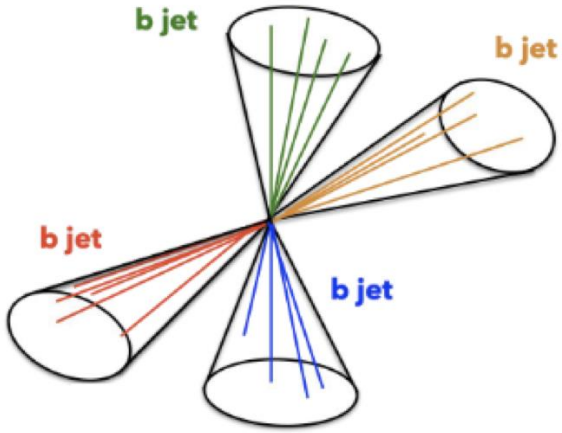
Search for Additional Higgs Bosons

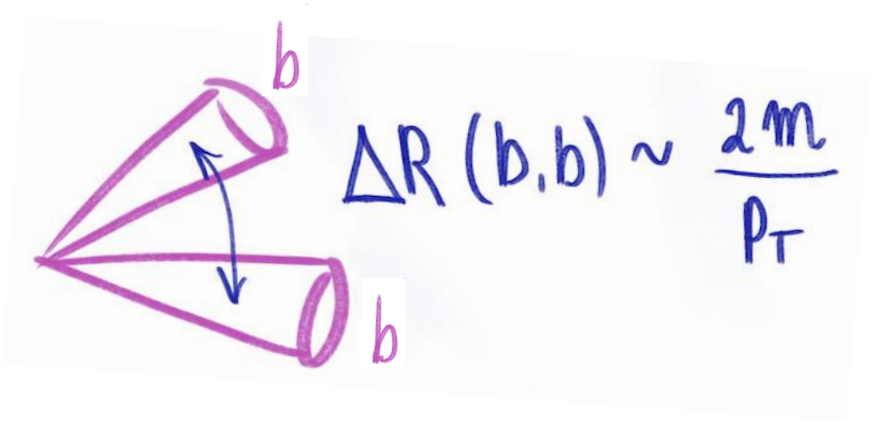
- Search for an extended Higgs sector with 2 additional Higgs bosons: X, Y
- Possible explanation of matter-antimatter asymmetry
- Target signal: X → YH → 4b, with m_H = 125 GeV
- One challenge: which b jet comes from which Higgs boson? → Machine Learning!



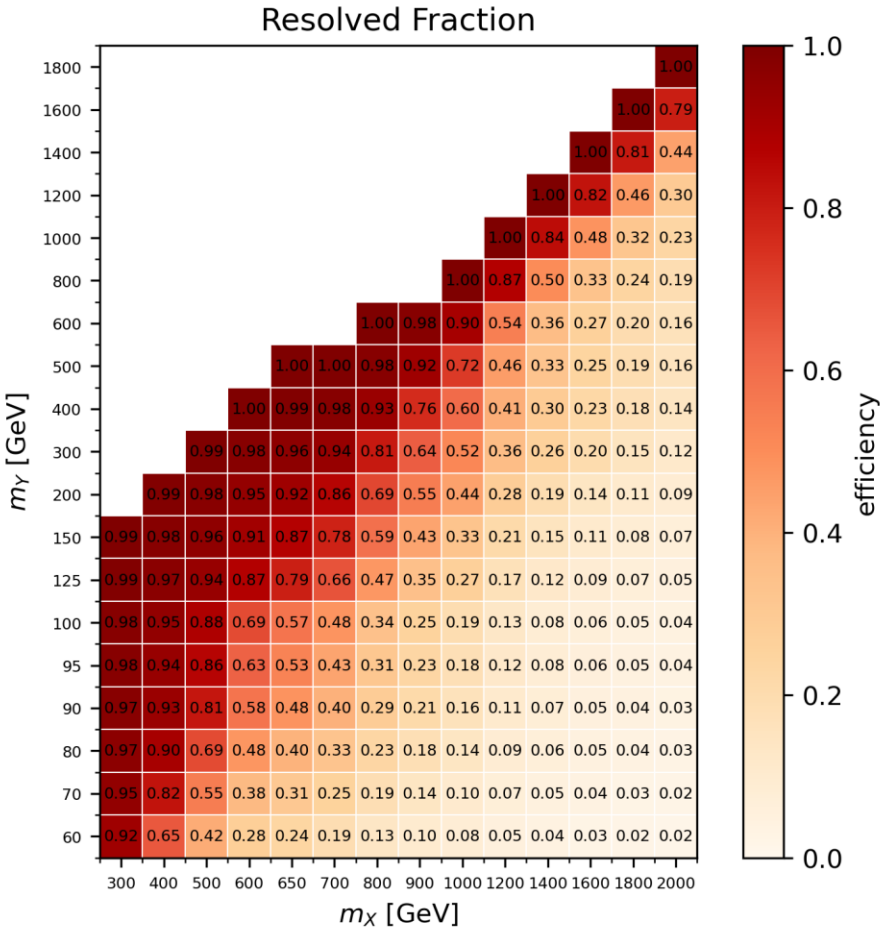


Analysis Phase Space: Resolved Topology



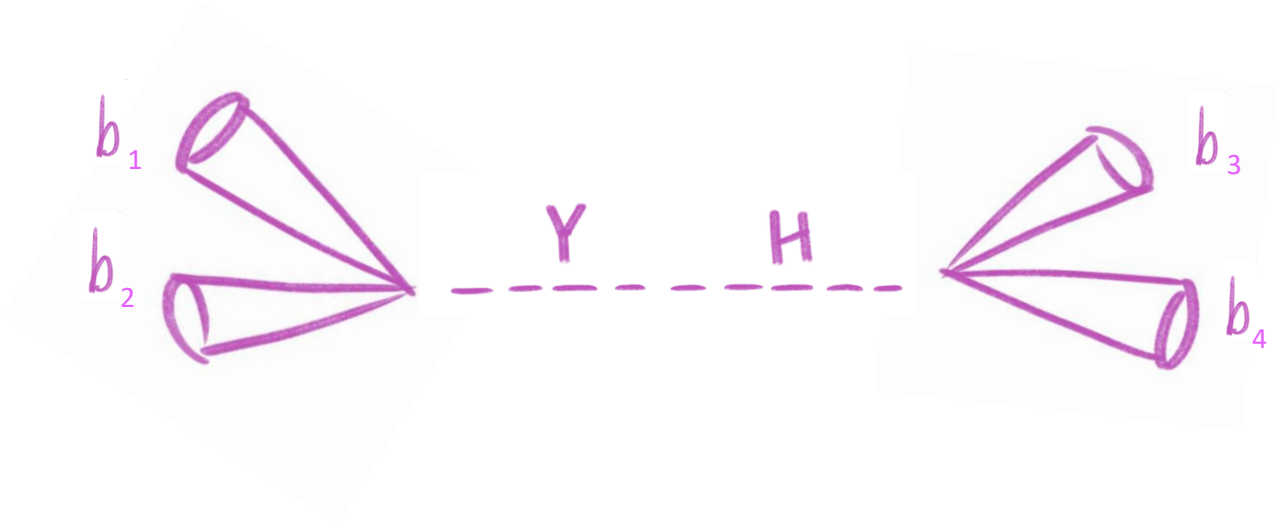


- ΔR(b,b) > 0.8 → resolved topology
- ΔR(b,b) depends on 2 unknowns: m_X and m_Y
- Resolved topology covers a large phase space region



Efficiency = fraction of events where $\Delta R_Y(b,b) > 0.8$ and $\Delta R_H(b,b) > 0.8$

What is a pairing?



3 possible pairings:

- Incorrect → b_1b_3 and b_2b_4
- Incorrect → b_1b_4 and b_2b_3
- Correct → b_1b_2 and b_3b_4

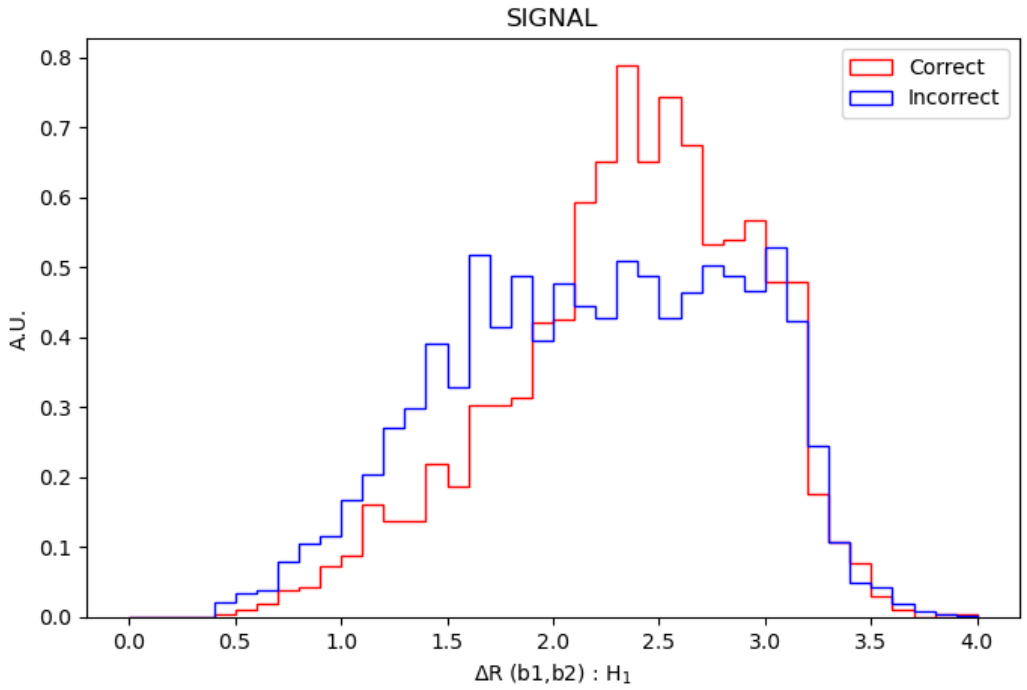
Correct: pairing with di-jet matched to b quarks from the same particle (H or Y)

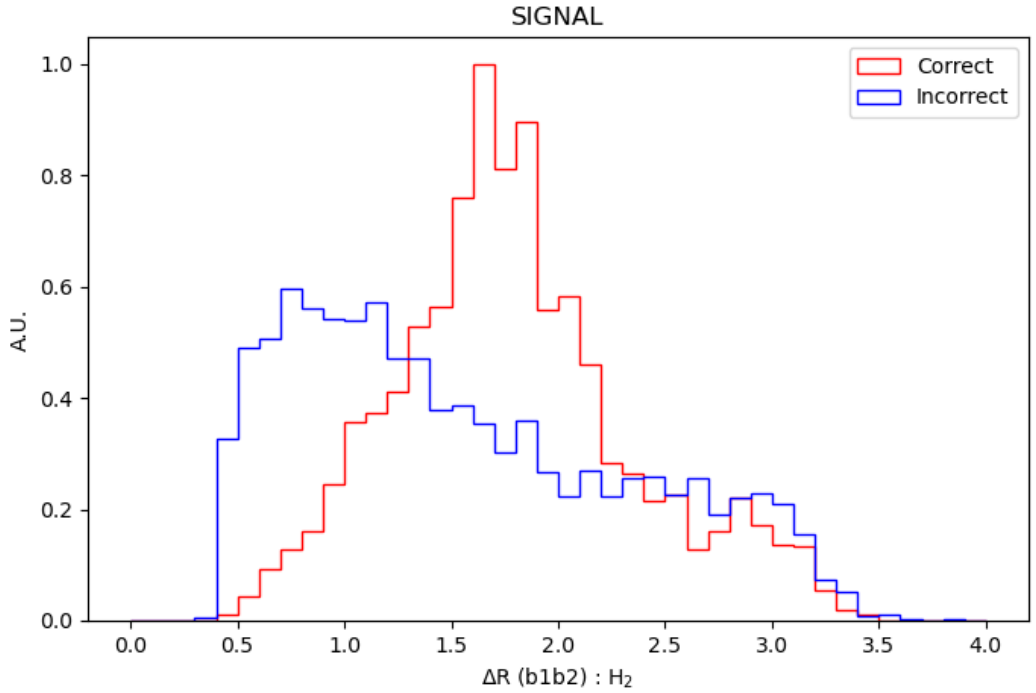
Incorrect: other pairings

B-jet pairings

m_X = 400 GeV
m_Y = 200 GeV

H_1: heavier reconstructed Higgs
H_2: lighter reconstructed Higgs





Correct: pairing with di-jet matched to b quarks from the same particle (H or Y)

Incorrect: other pairings

BDT Score & Overtraining Test

Signal = correct pairings

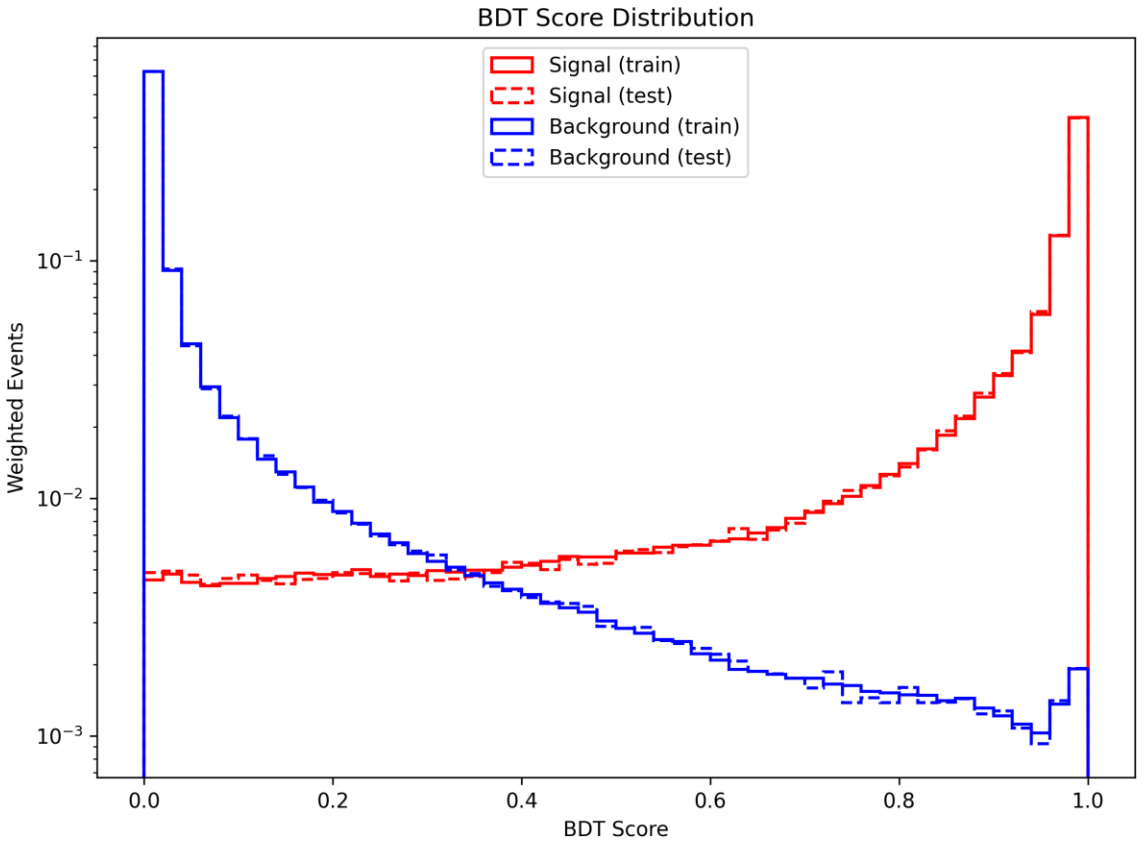
Background = incorrect pairings

- Input features

Category	Variables
Higgs 1 (H1)	$\Delta R_{b_1b_2}^{H_1}, \Delta\eta_{b_1b_2}^{H_1}, \Delta\phi_{b_1b_2}^{H_1}, p_{T\,\text{ratio}}^{b_1b_2,H_1}, \kappa_{0.3,\,\text{prod}}^{b_1b_2,H_1}, \kappa_{0.3,\,\text{sum}}^{b_1b_2,H_1}, \theta_{H_1}$
Higgs 2 (H2)	$\Delta R_{b_1b_2}^{H_2}, \Delta\eta_{b_1b_2}^{H_2}, \Delta\phi_{b_1b_2}^{H_2}, p_{T\,\text{ratio}}^{b_1b_2,H_2}, \kappa_{0.3,\,\text{prod}}^{b_1b_2,H_2}, \kappa_{0.3,\,\text{sum}}^{b_1b_2,H_2}, \theta_{H_2}$
Higgs pair (H1H2)	$\Delta R_{H_1H_2}, \Delta\eta_{H_1H_2}, \Delta\phi_{H_1H_2}, p_{T\,\text{ratio}}^{H_1H_2}$
Mass parameters	m_X, m_Y

- Boosted Decision Tree (BDT) powerful to separate correct/incorrect pairings

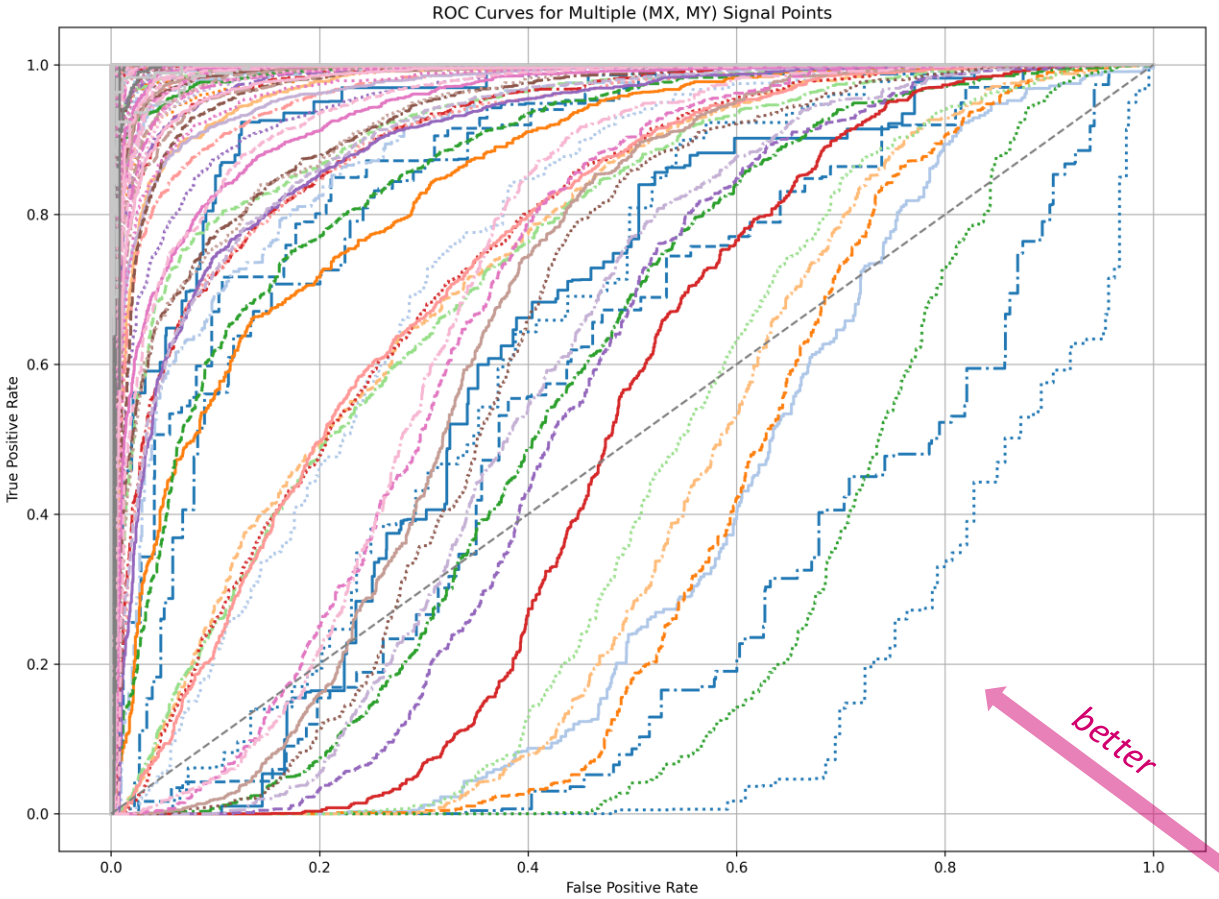
- Train and test in agreement → **no overtraining**



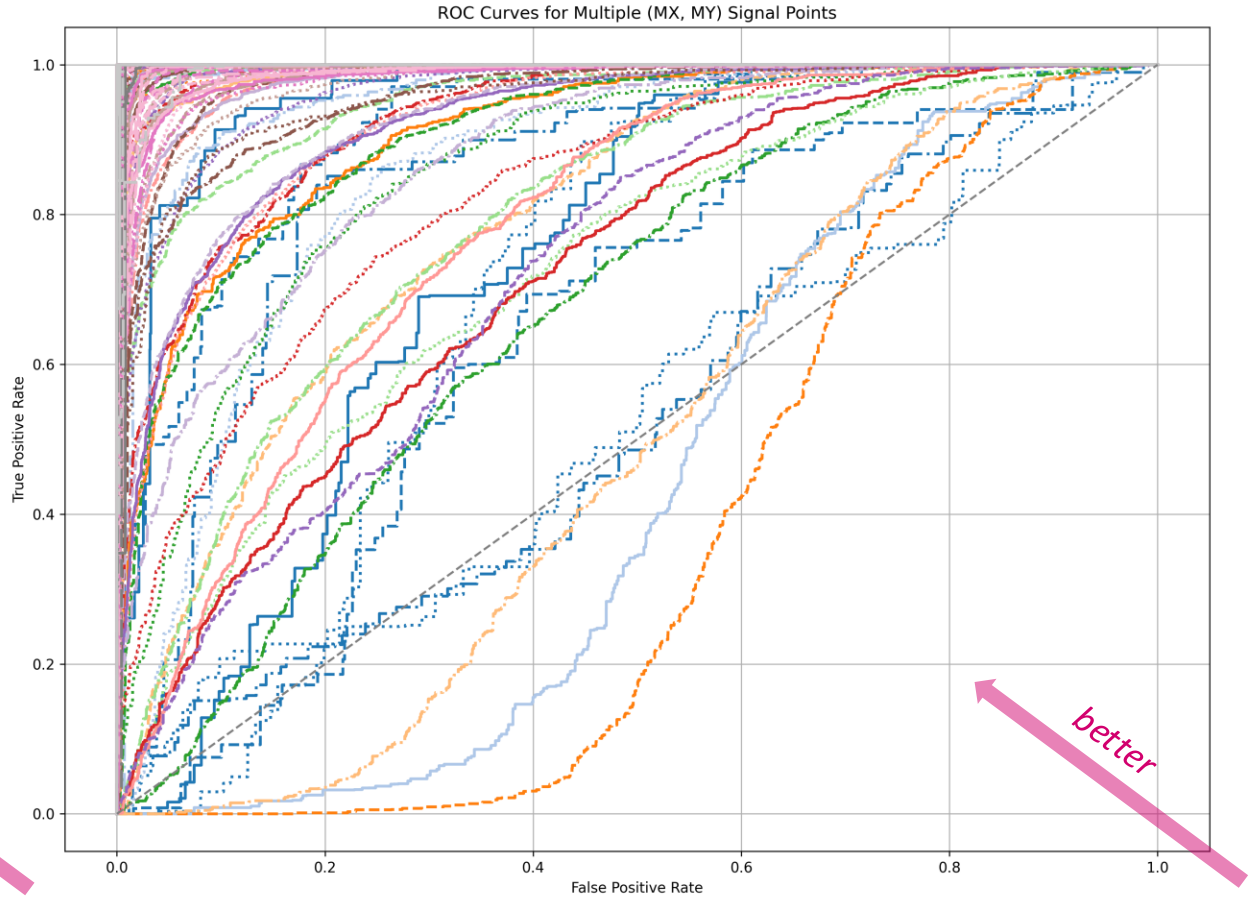
Power of BDT

Mass parametrization = including m_X and m_Y as input features in BDT training

Without mass parametrization



With mass parametrization



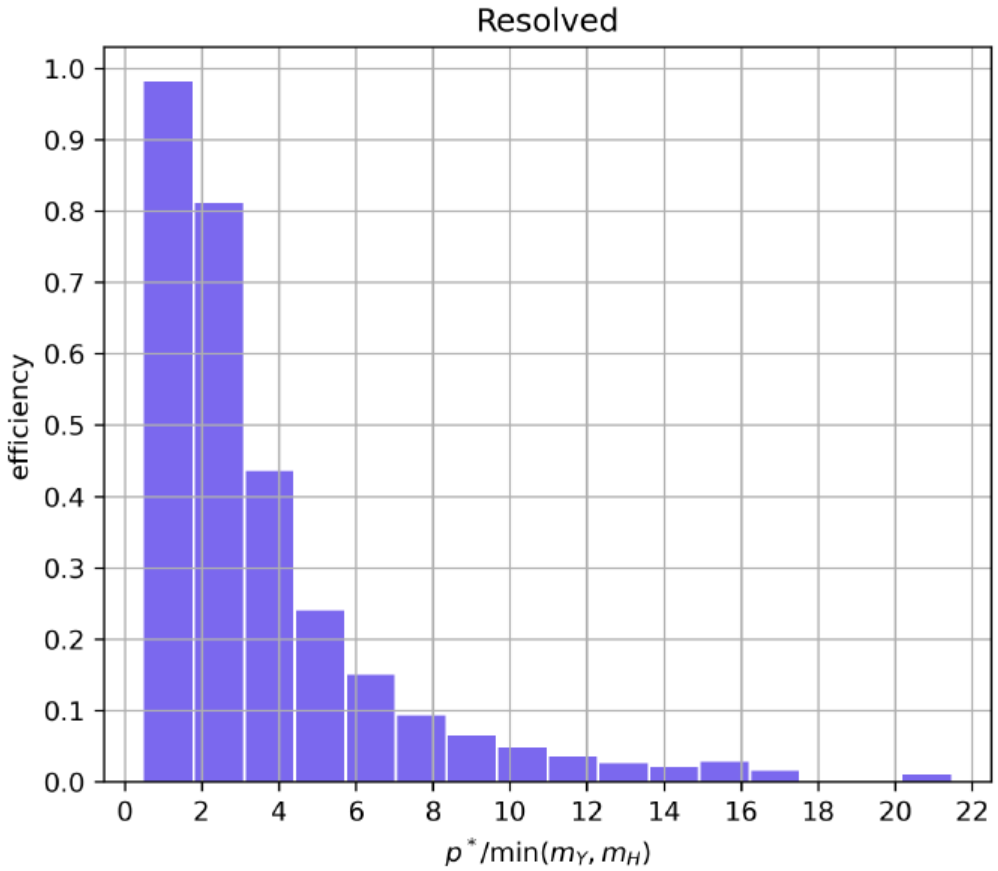
→ Using mass parametrization improves the performance

Derivation of $p^*/\min(m_Y, m_Y)$

Rest frame of X: $p_Y = p_H = p^*$

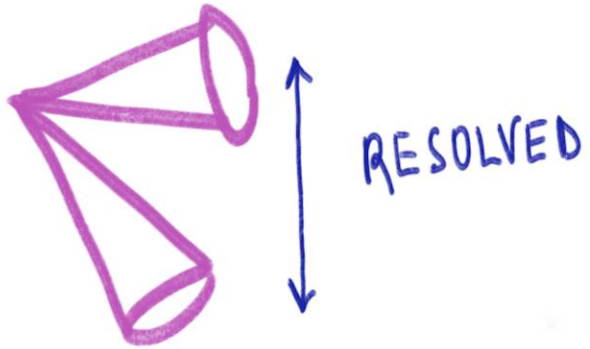
$$p^* = \frac{\sqrt{\left(m_X^2 - (m_Y - m_H)^2\right)\left(m_X^2 - (m_Y + m_H)^2\right)}}{2m_X}$$

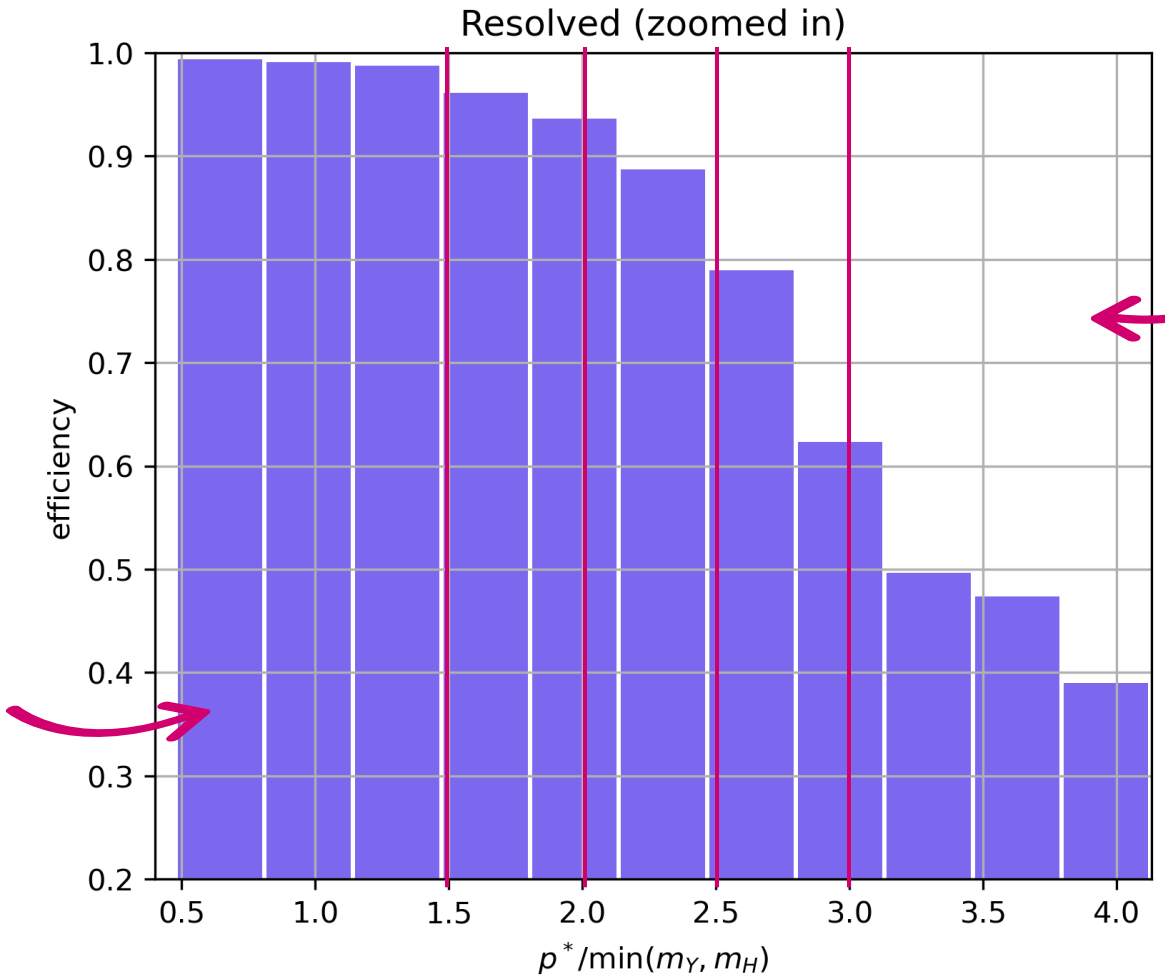
$$\Delta R(b,b) \sim \frac{2m}{p_T} \quad \Rightarrow \quad \frac{p^*}{\min(m_Y, m_H)}, \quad m_H = 125\,\text{GeV}$$

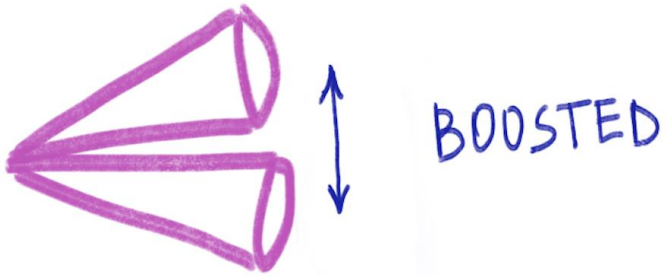


Efficiency = fraction of events where $\Delta R_Y(b,b) > 0.8$ and $\Delta R_H(b,b) > 0.8$

BDT Training Categorization





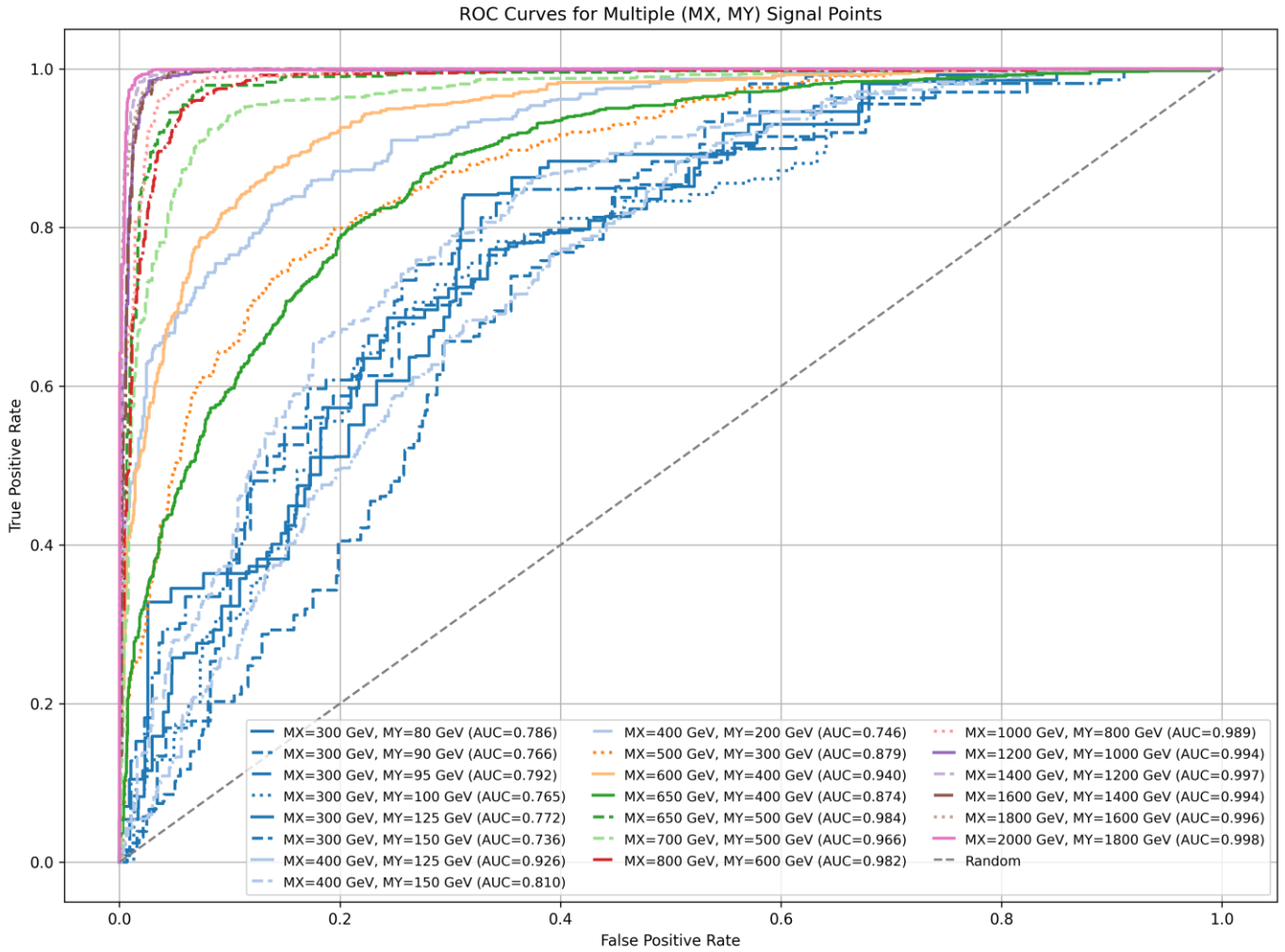


→ 5 BDT training regions

$$\frac{p^*}{\min(m_Y, m_H)} \in [0, 1.5], \quad [1.5, 2.0],$$

$$[2.0, 2.5], \quad [2.5, 3.0], \quad [3.0, \infty)$$

First Training Region



$$\frac{p^*}{\min(m_Y, m_H)} \in [0, 1.5]$$

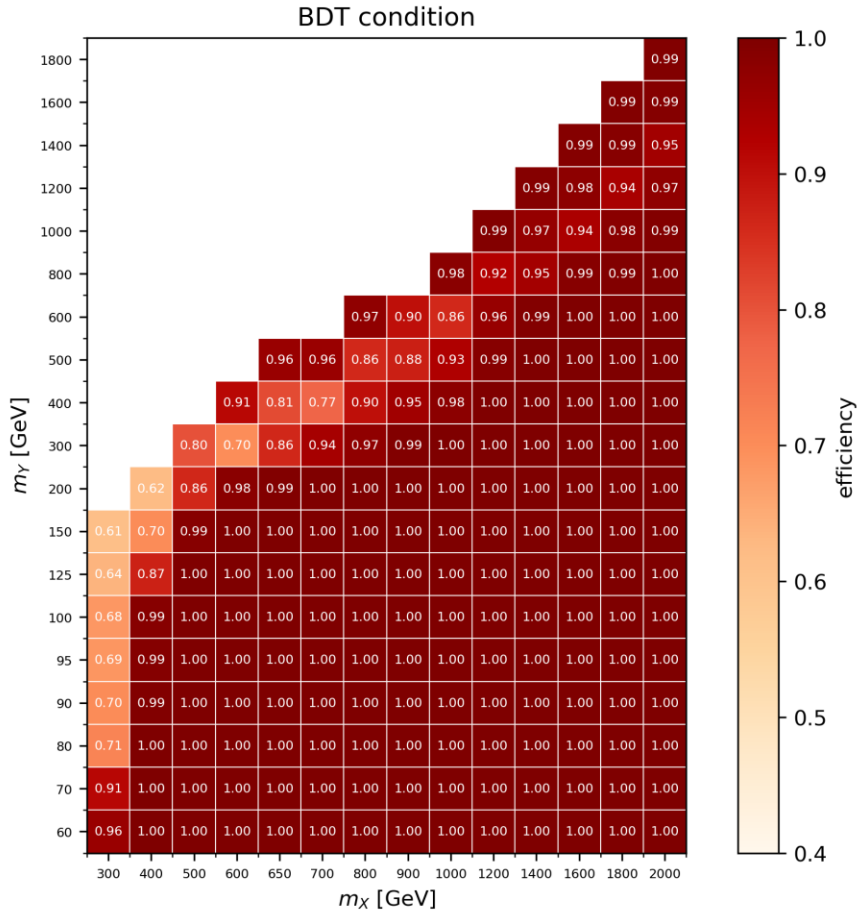
→ Performance gets better with categorization

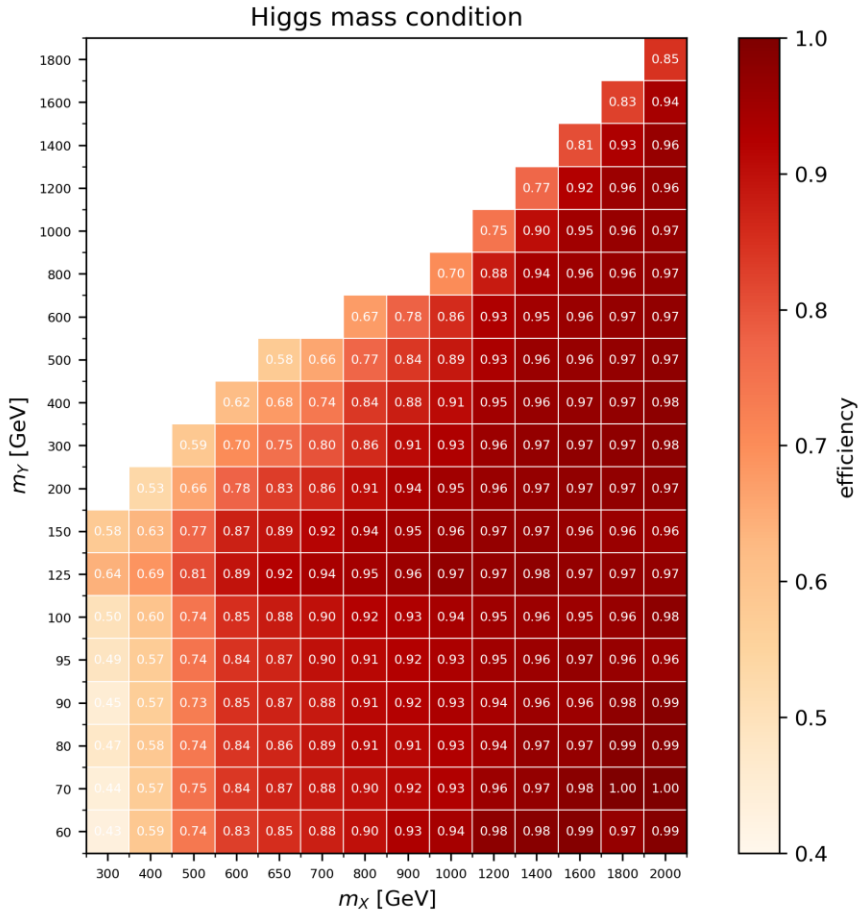
Performance Comparison: Efficiency of Finding True Pairing

BDT condition:
pairing with the highest BDT score

Minimum/Higgs mass condition:
pairing with a di-jet system mass closest to 125 GeV

Efficiency = fraction of true pairings that satisfy the condition



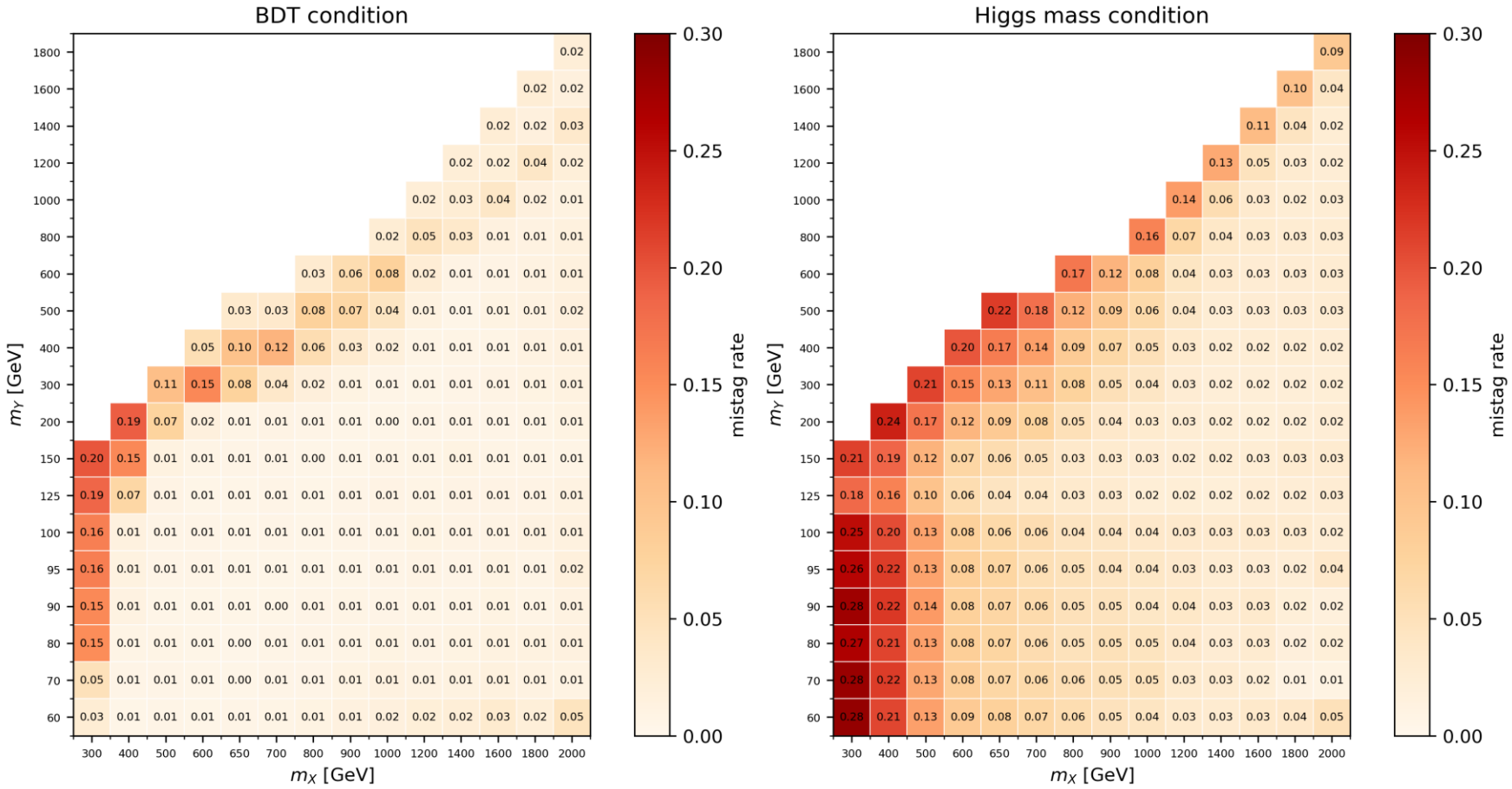


Performance Comparison: Mistag Rate

BDT condition:
pairing with the highest BDT score

Minimum/Higgs mass condition:
pairing with a di-jet system mass closest to 125 GeV

Mistag rate = fraction of wrong pairings that satisfy the condition

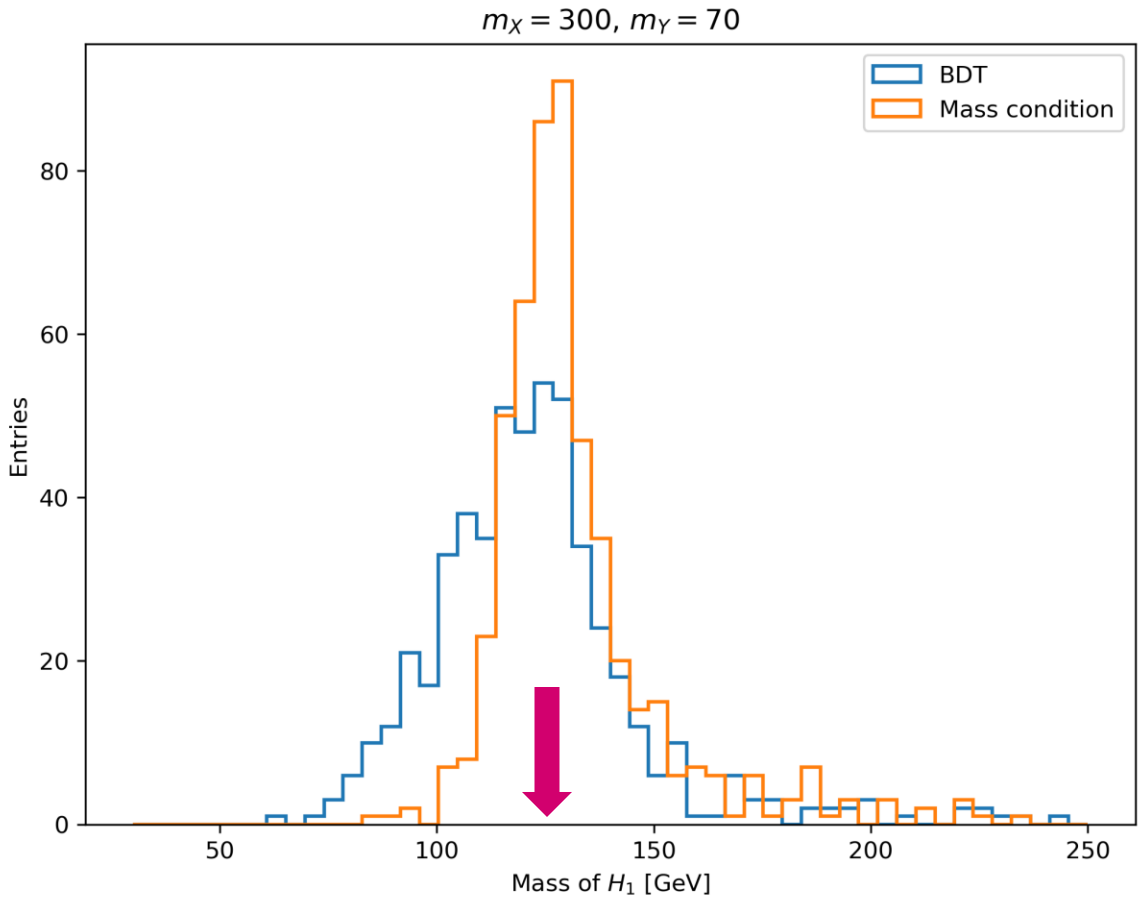


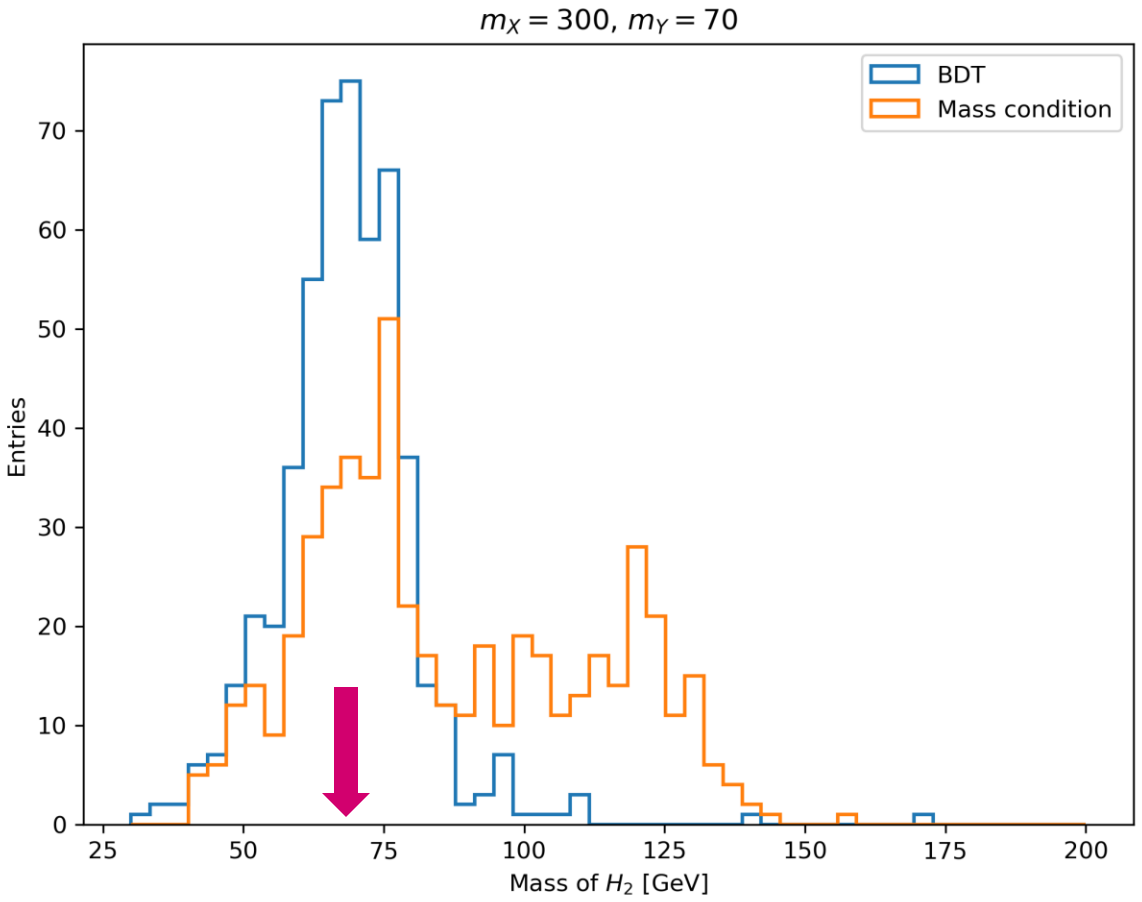
→ **BDT performs better than Higgs mass condition!**

Reconstructed Higgs Boson Mass Distributions

H_1: heavier reconstructed Higgs

H_2: lighter reconstructed Higgs



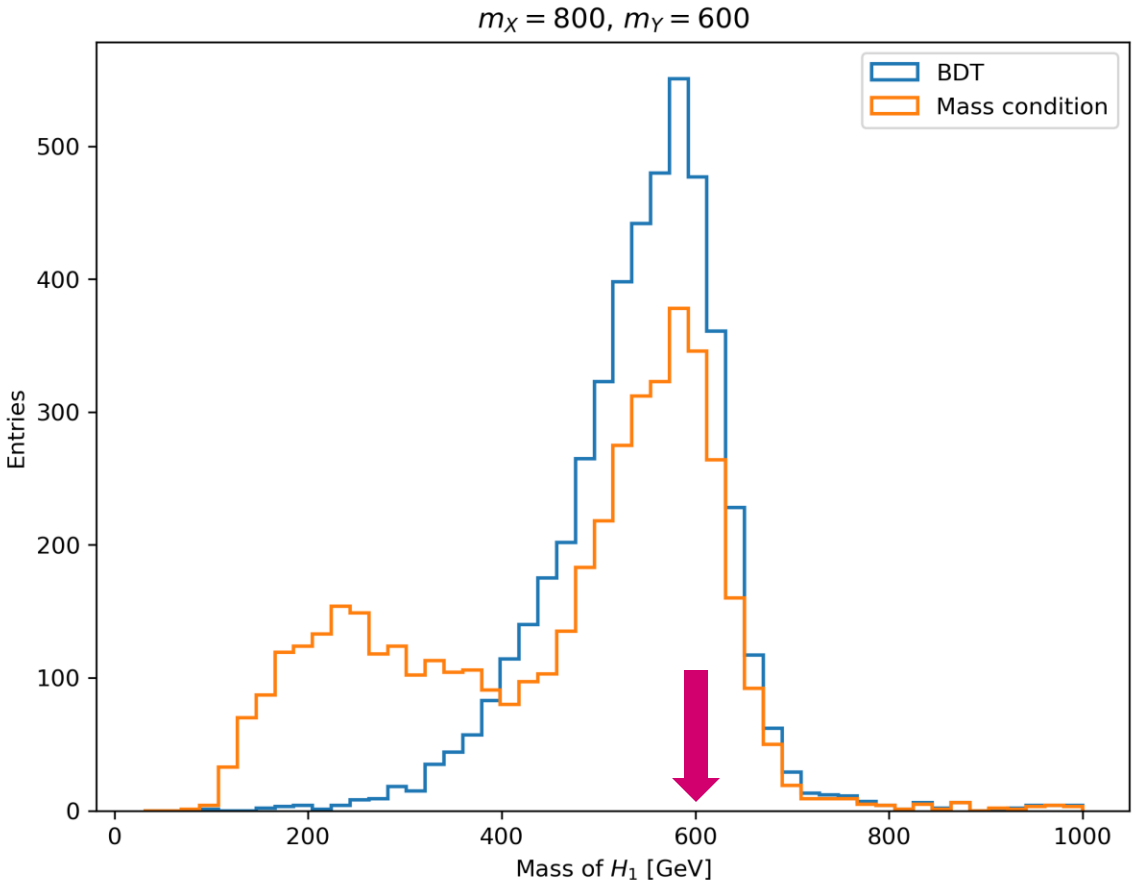


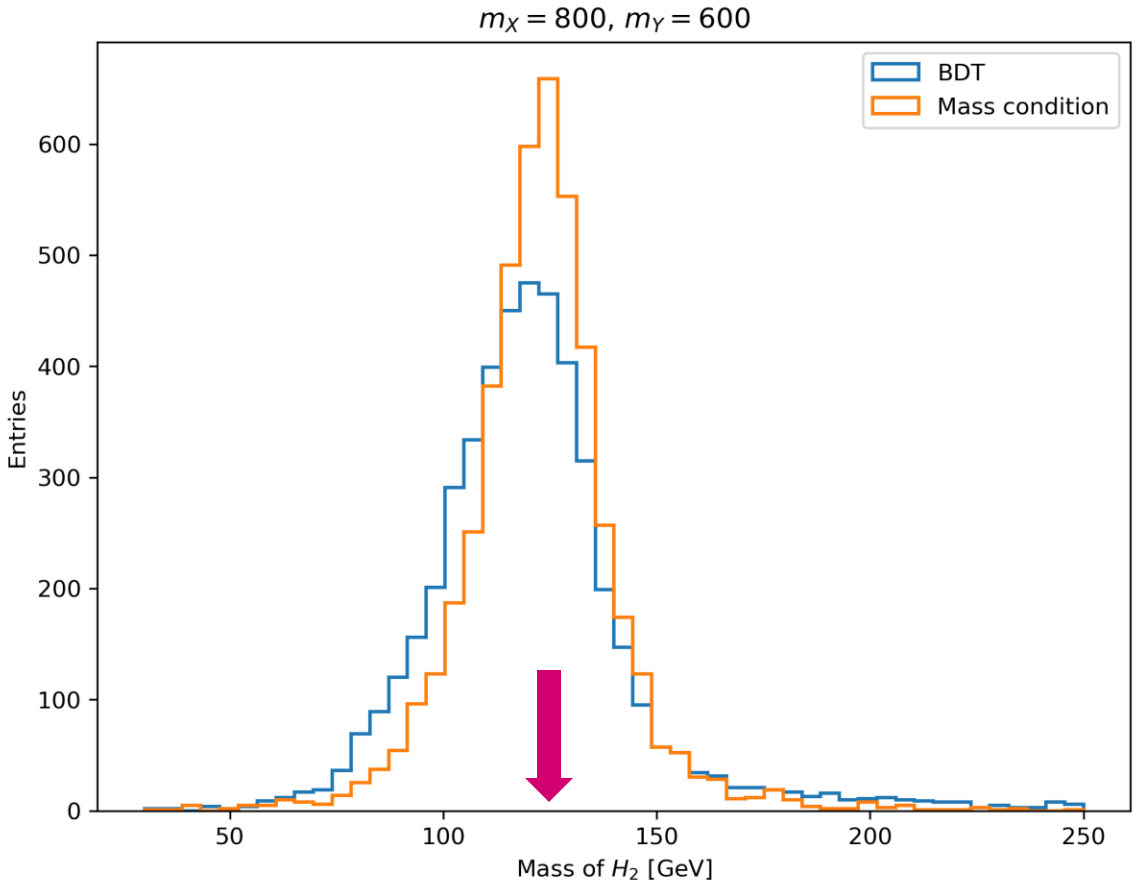
→ BDT gives the right Higgs masses

→ Higgs mass condition has an artificial peak around 125 GeV, due to wrong di-jet pairing

Reconstructed Higgs Boson Mass Distributions

H_1: heavier reconstructed Higgs
H_2: lighter reconstructed Higgs





→ BDT gives the right Higgs masses
→ Higgs mass condition has an artificial peak around 125 GeV, due to wrong di-jet pairing

Summary & Outlook

Finding correct b jet pairing
in X → YH → 4b signature

Focus on resolved topology

BDT is a powerful tool to identify
which b jet comes from which Higgs boson !

Next steps: investigate impact of the method on data and signal sensitivity

Thank you CMS Higgs group!

Back Up Slides

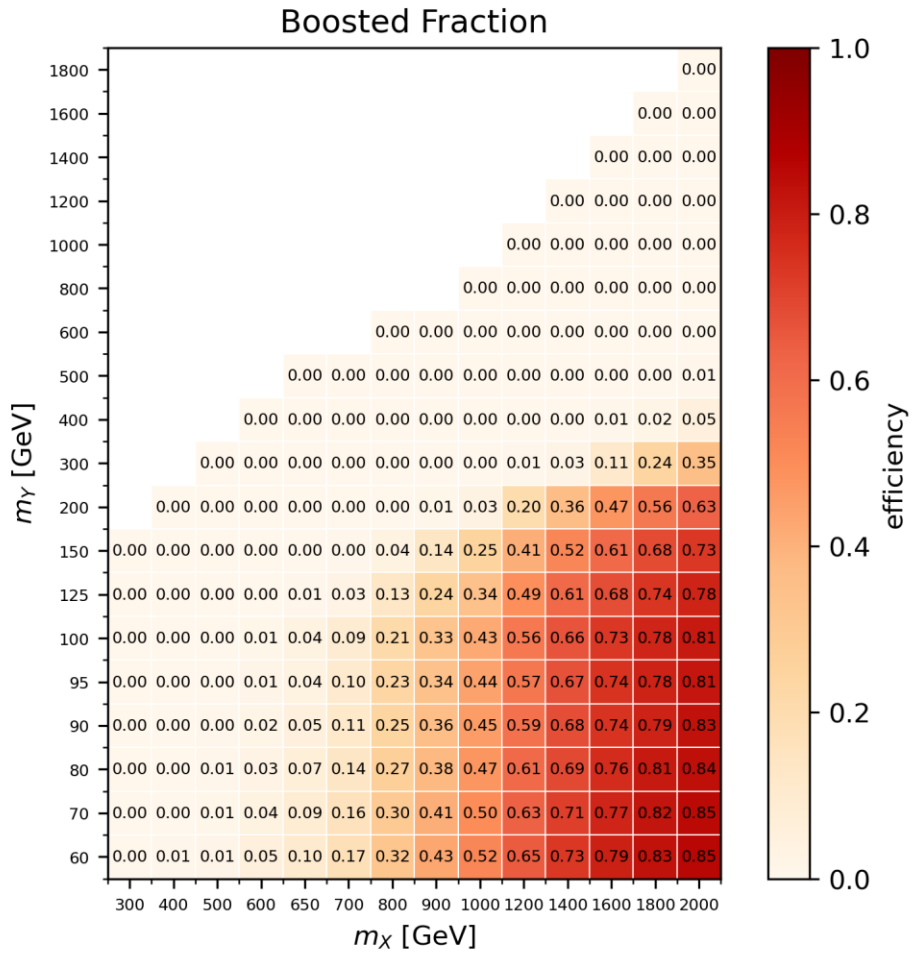
Definition of Jet

Angular distance depends on pseudorapidity **η** and azimuthal angle **ϕ**

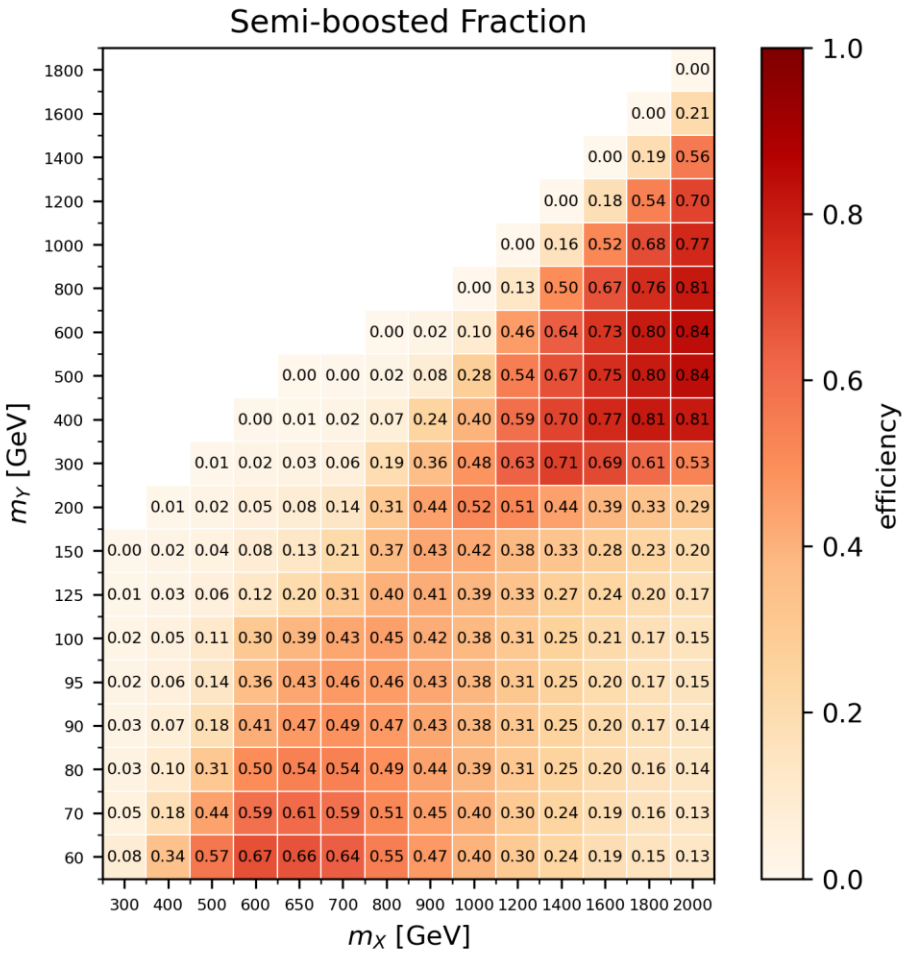
$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

- Use anti-k_t algorithm to reconstruct jets
- Each algorithm has radius parameter **R** which determines how big jet cone is in **(η, ϕ)**
- ➔All particles within circle of radius 0.4 in (η, ϕ) space are grouped into the same jet

Fraction of boosted and semi-boosted events

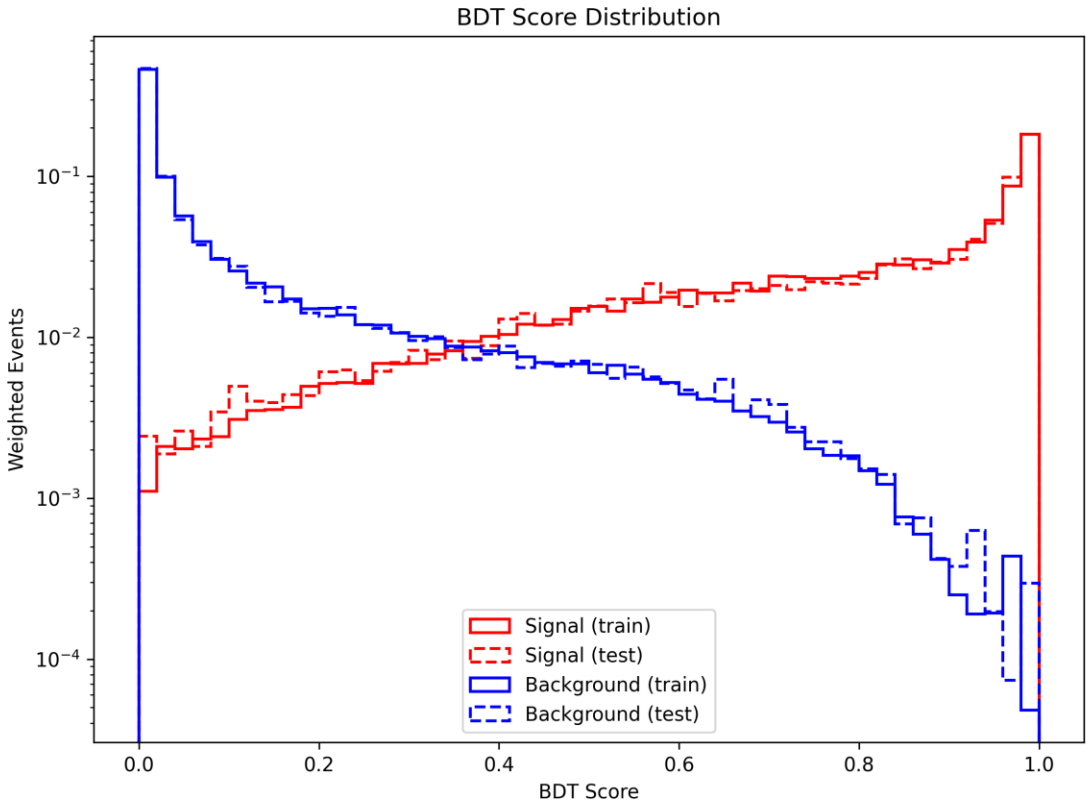


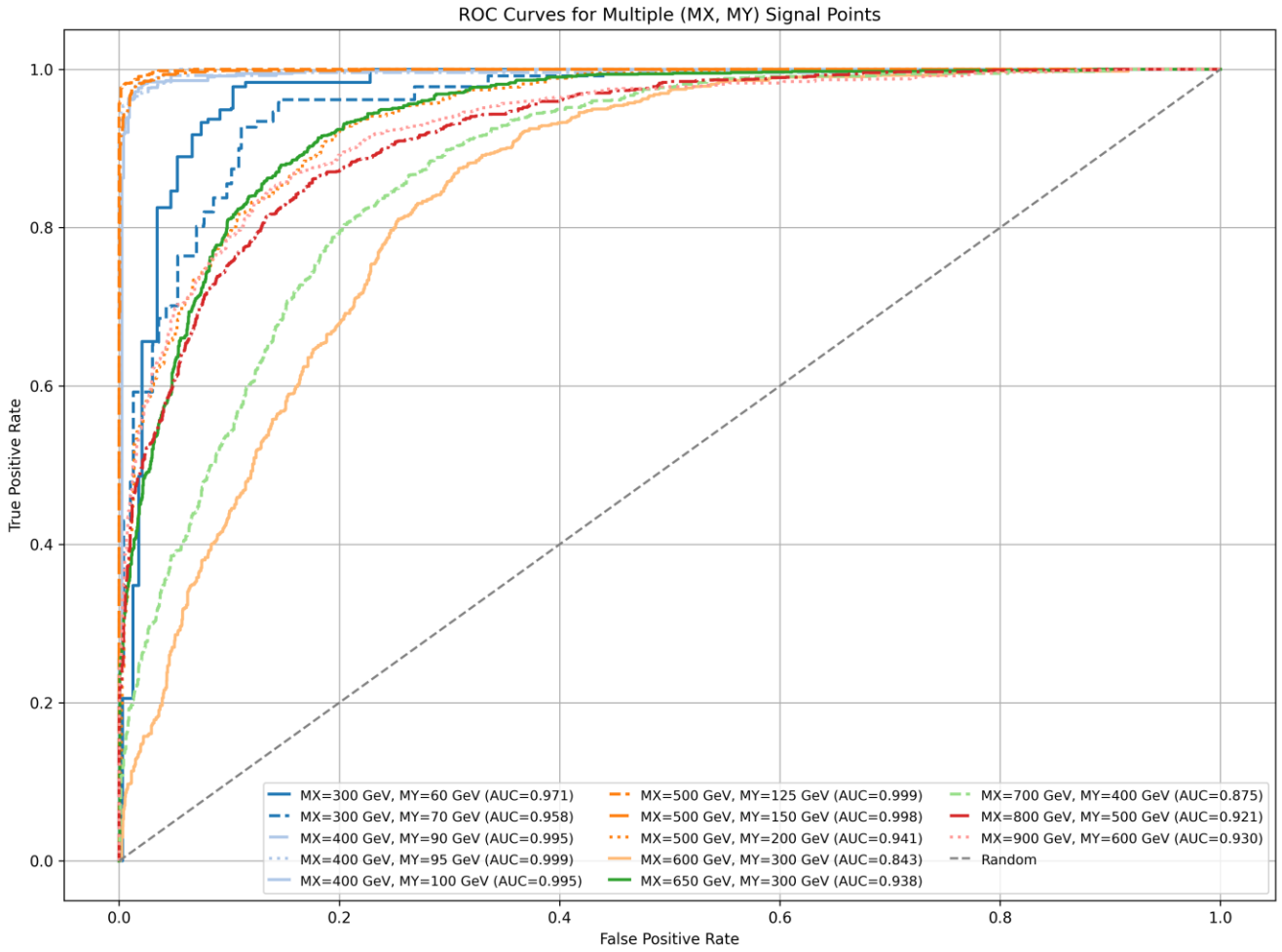
Efficiency = fraction of events where $\Delta R_Y(b,b) < 0.8$ and $\Delta R_H(b,b) < 0.8$



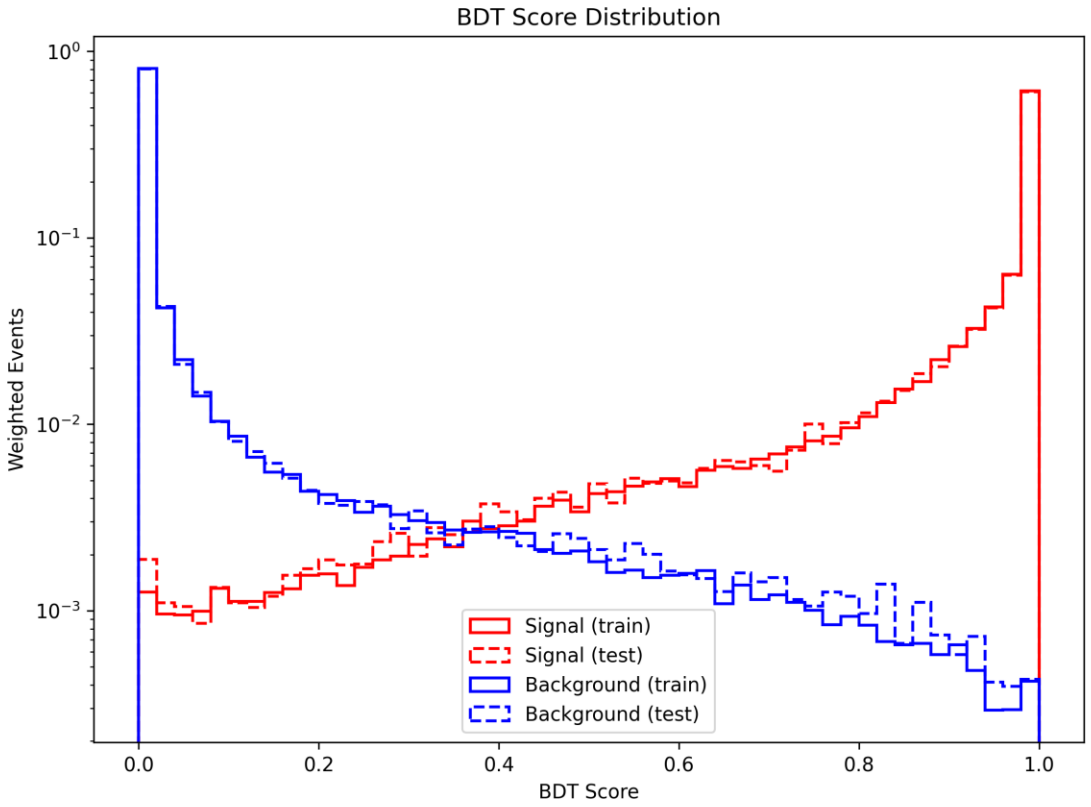
Efficiency = fraction of events where $\Delta R_Y(b,b) < 0.8$ and $\Delta R_H(b,b) > 0.8$ OR $\Delta R_Y(b,b) > 0.8$ and $\Delta R_H(b,b) < 0.8$

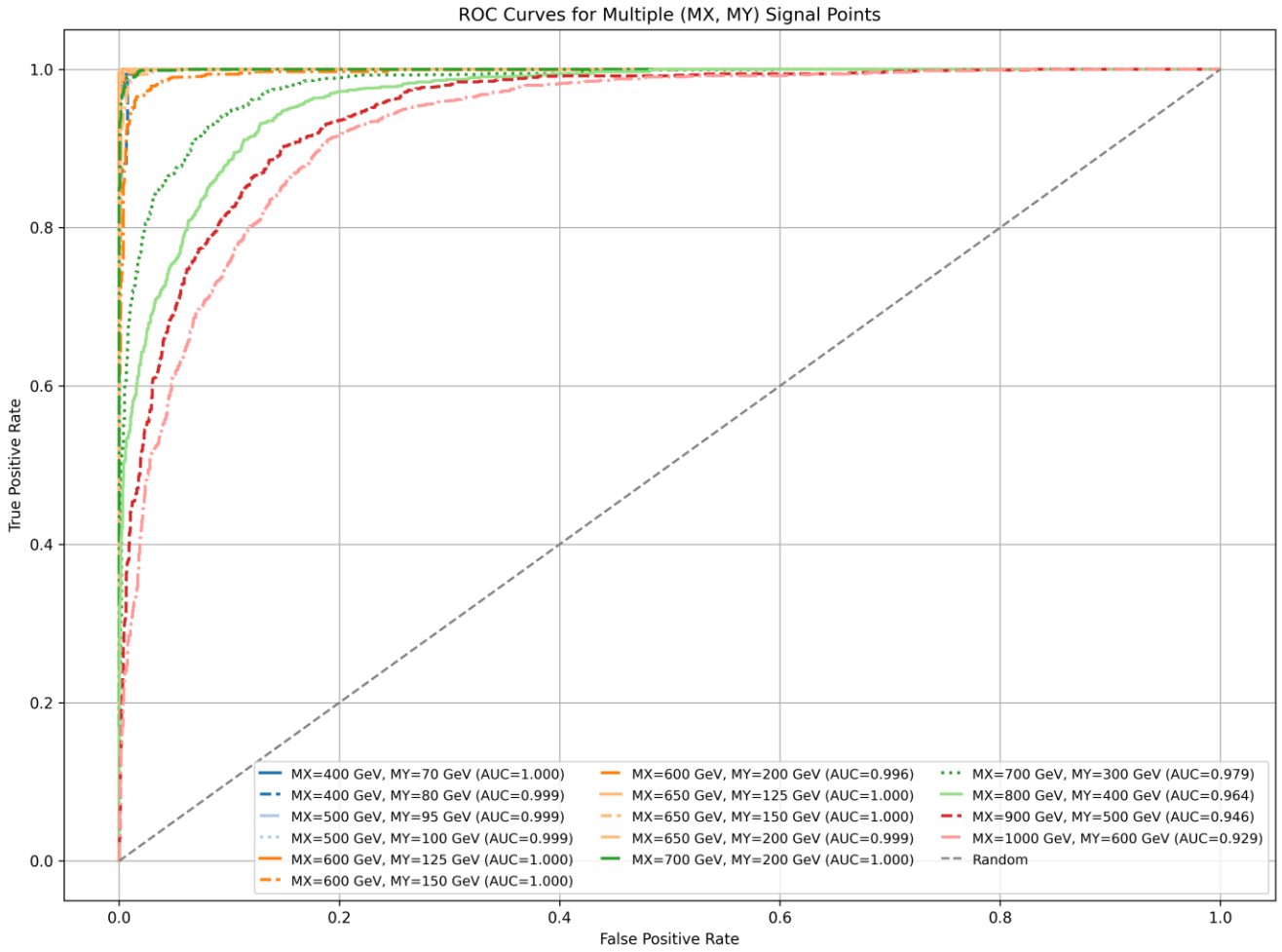
Performance [1.5, 2]



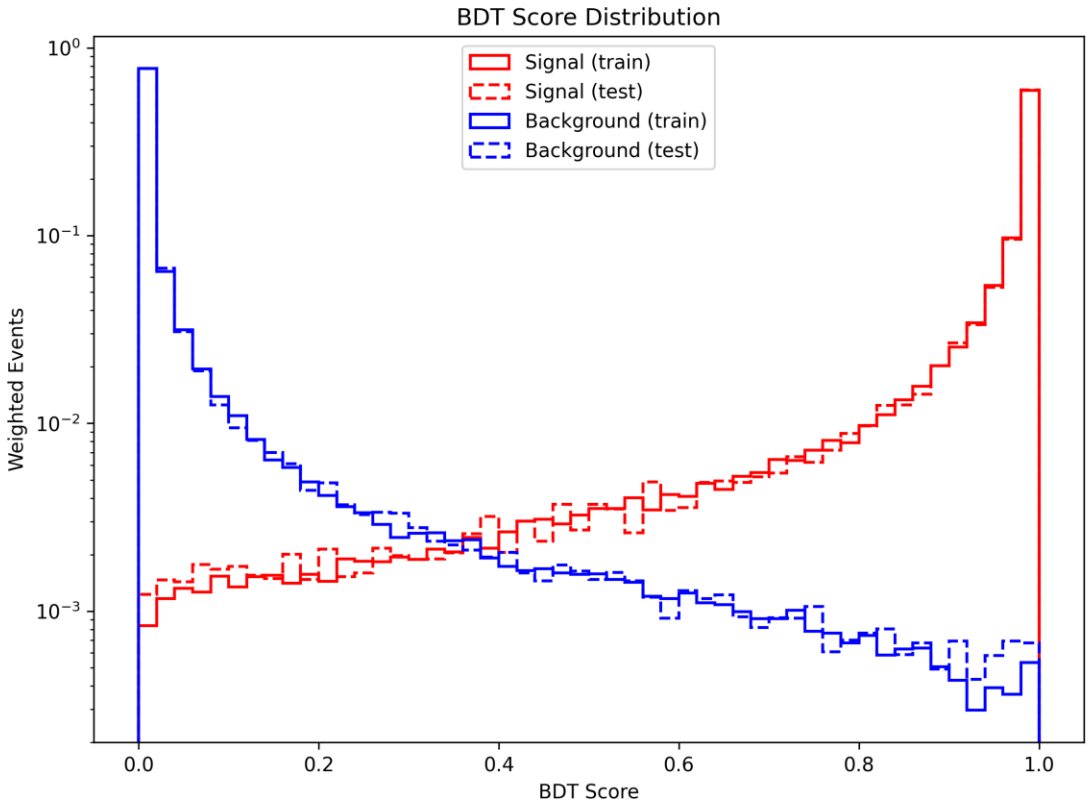


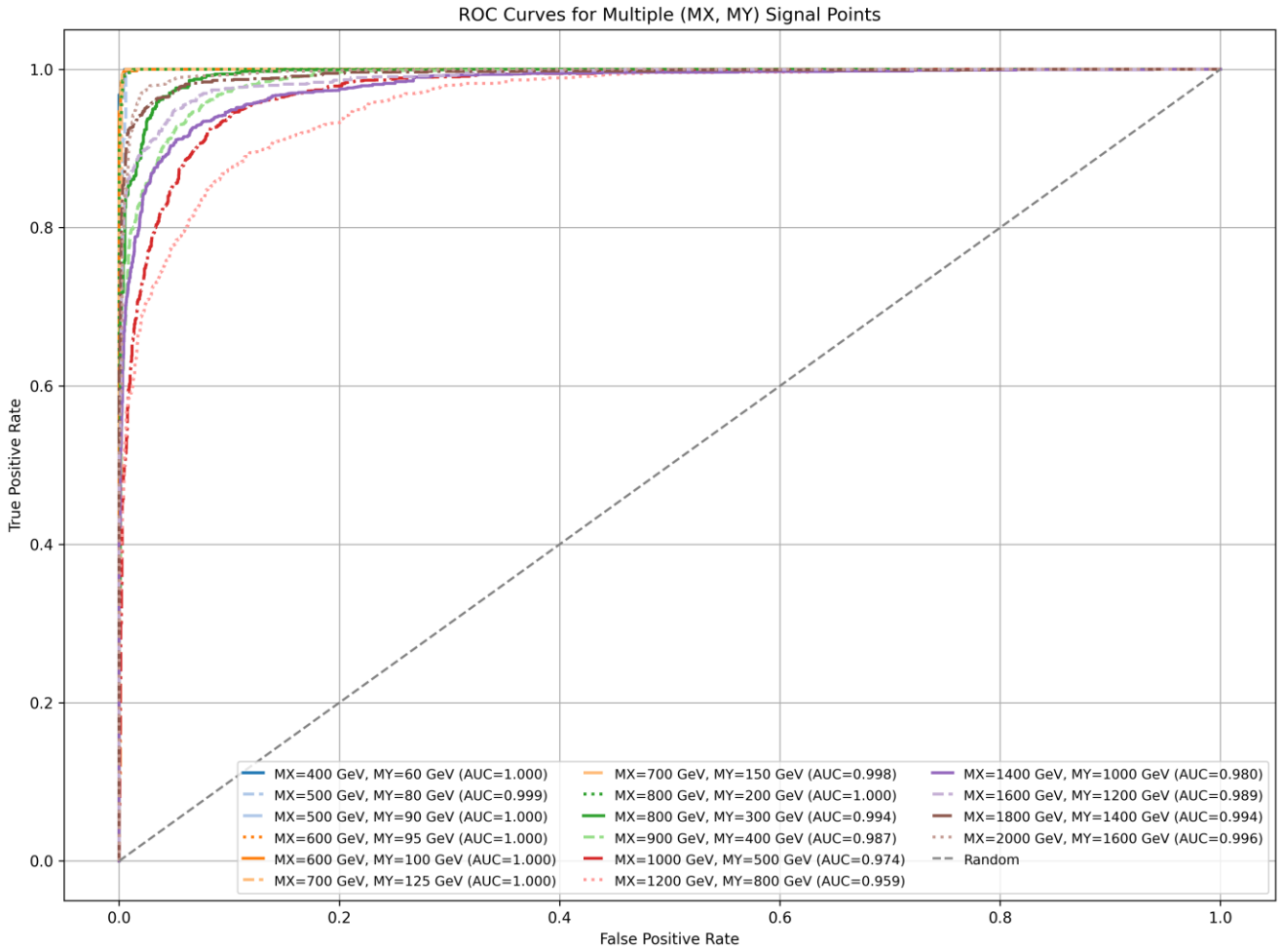
Performance [2, 2.5]



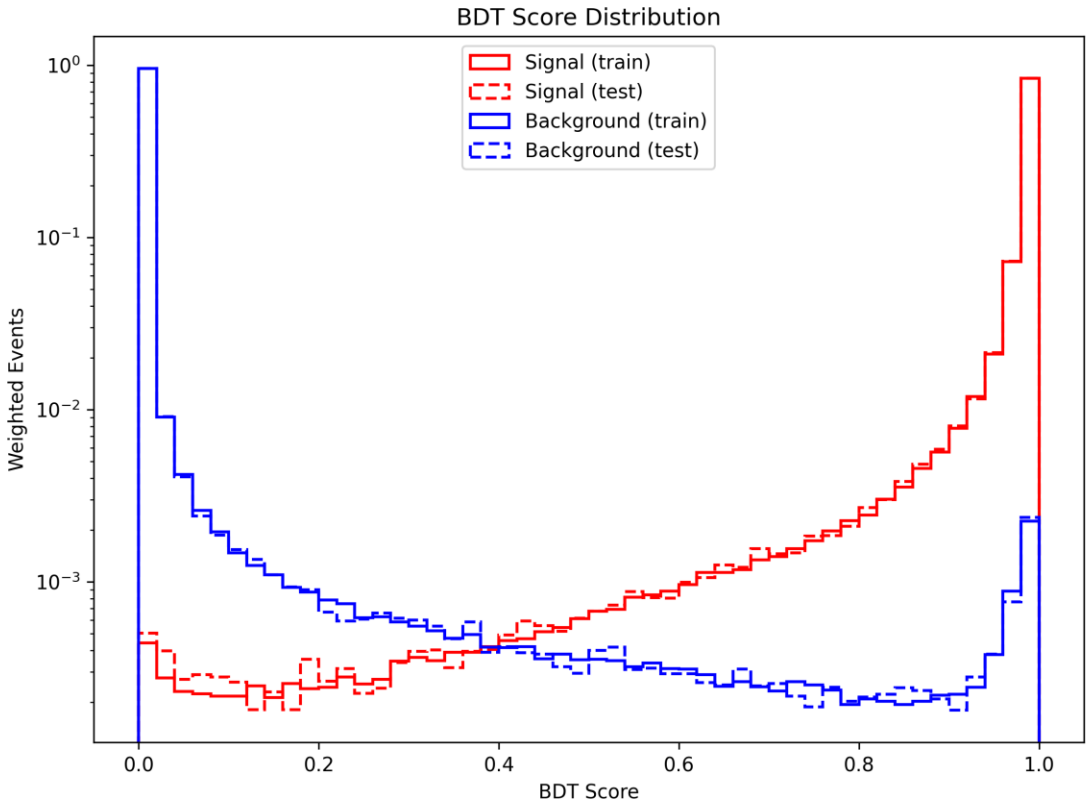


Performance [2.5, 3]

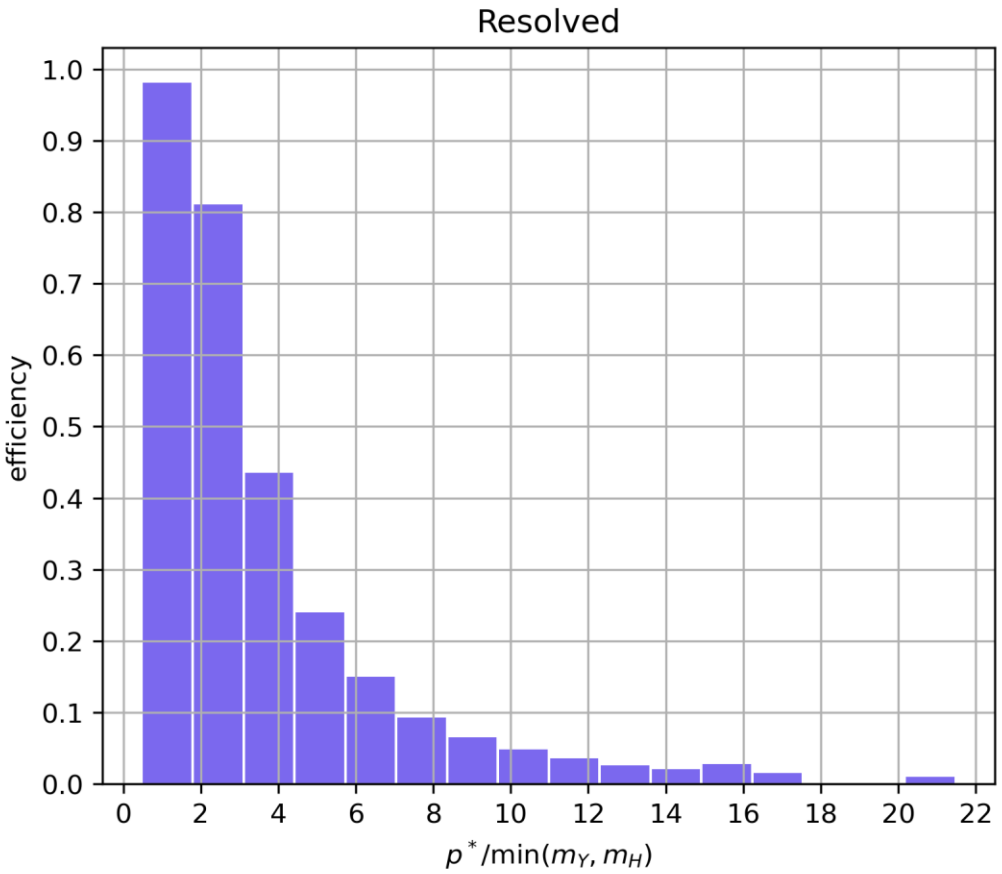


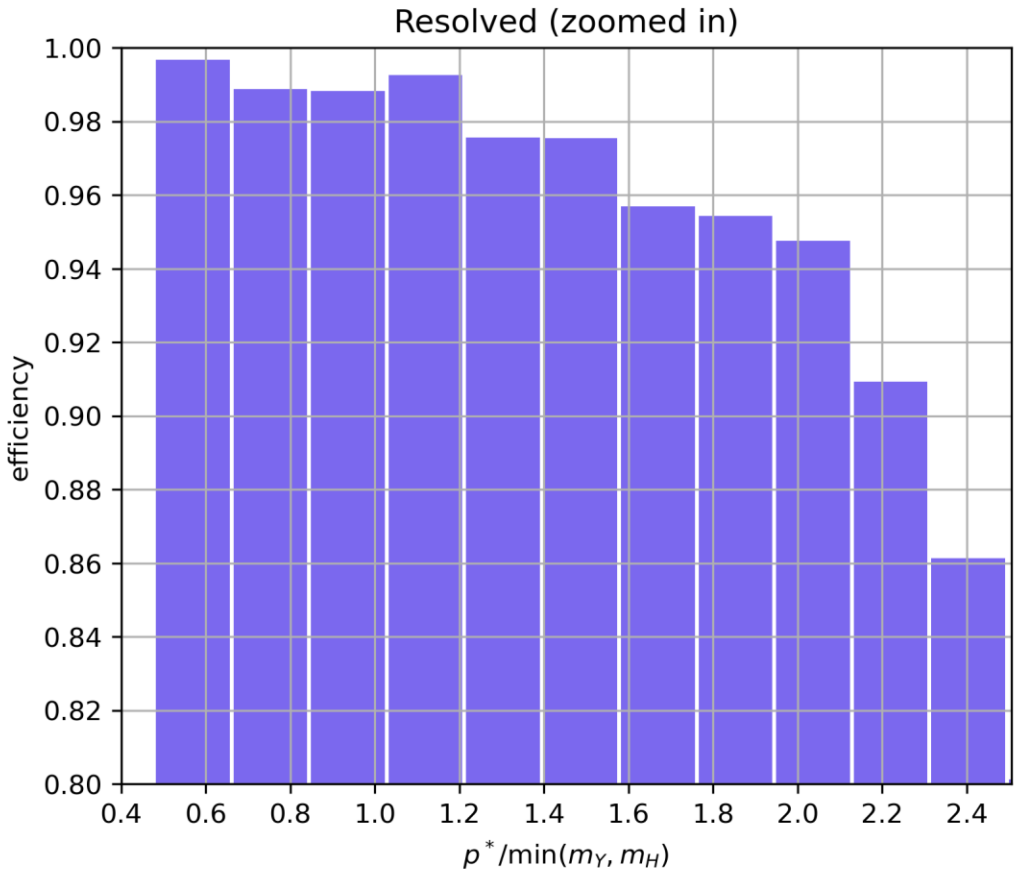


Performance [3, inf]

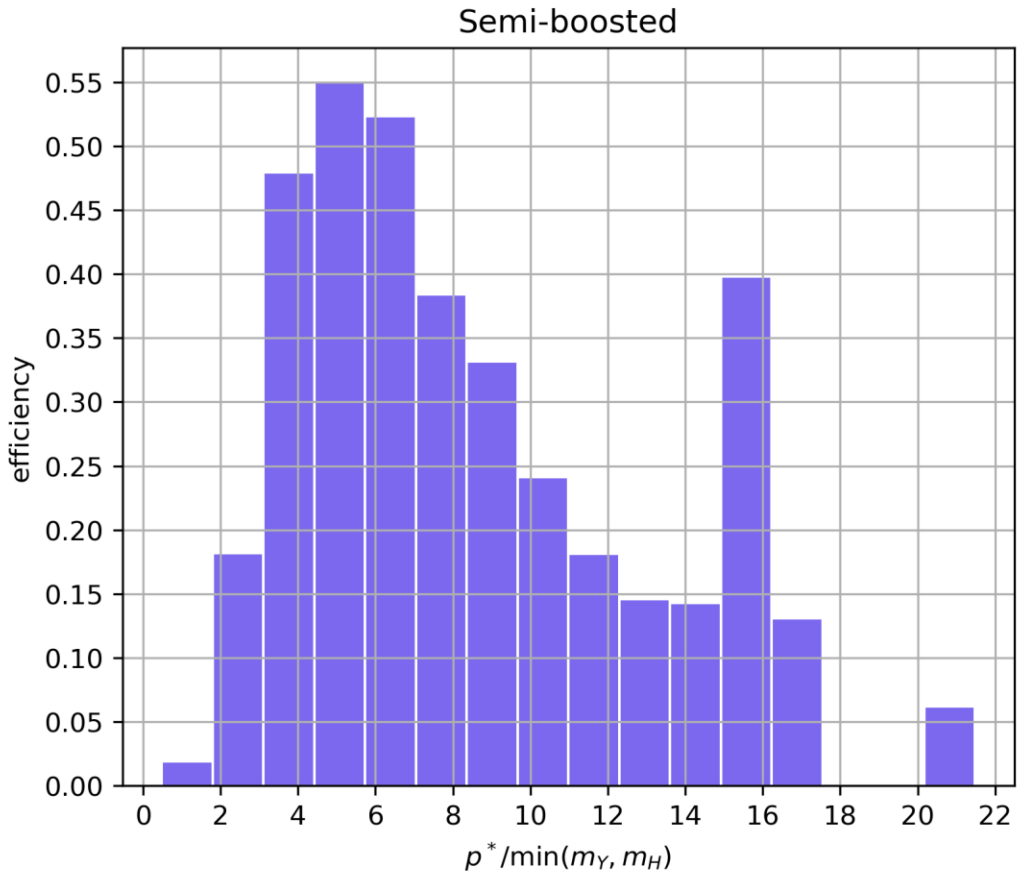


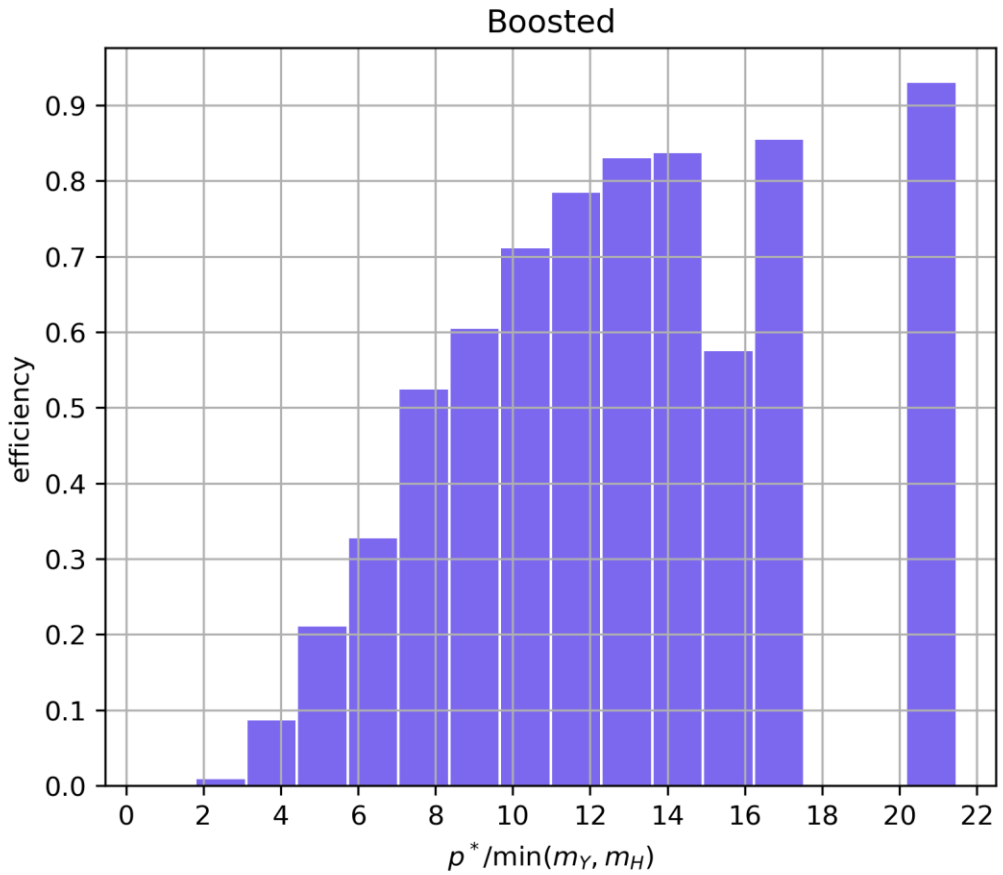
Efficiency Over $p^*/\min(m_y,m_H)$





Efficiency Over $p^*/\min(m_y, m_H)$

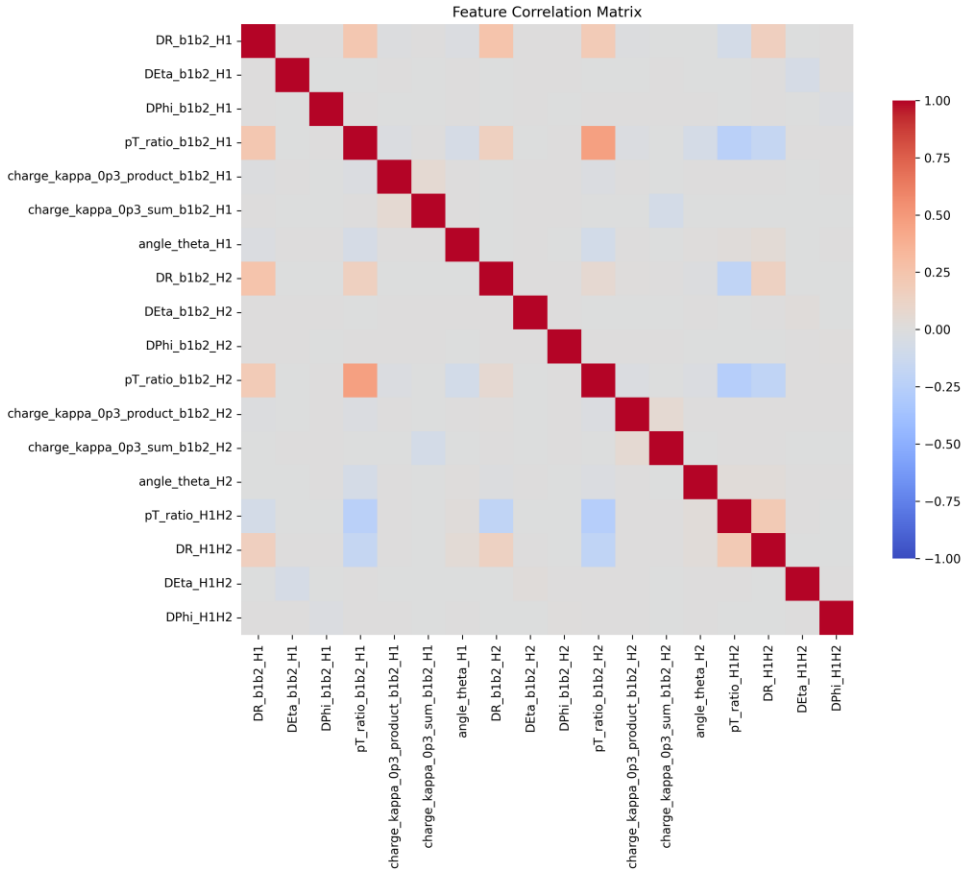




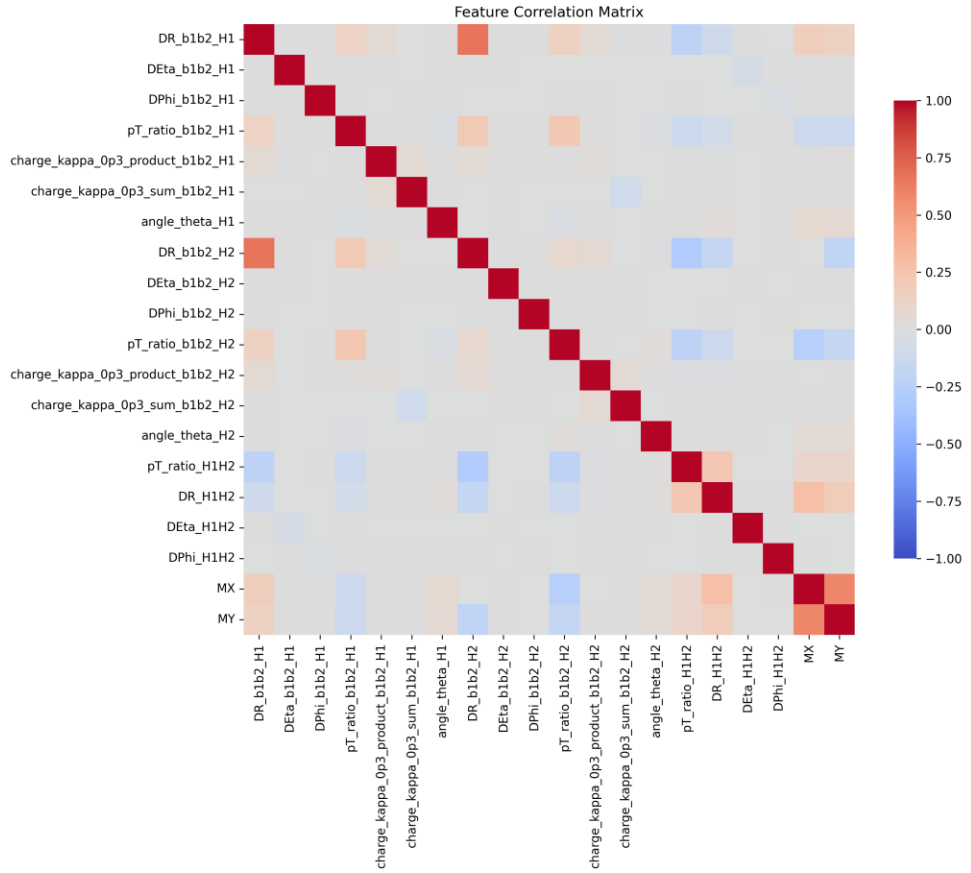
Mass parametrization = including m_X and m_Y as input features in BDT training

Feature Correlation Matrix

Without mass parametrization



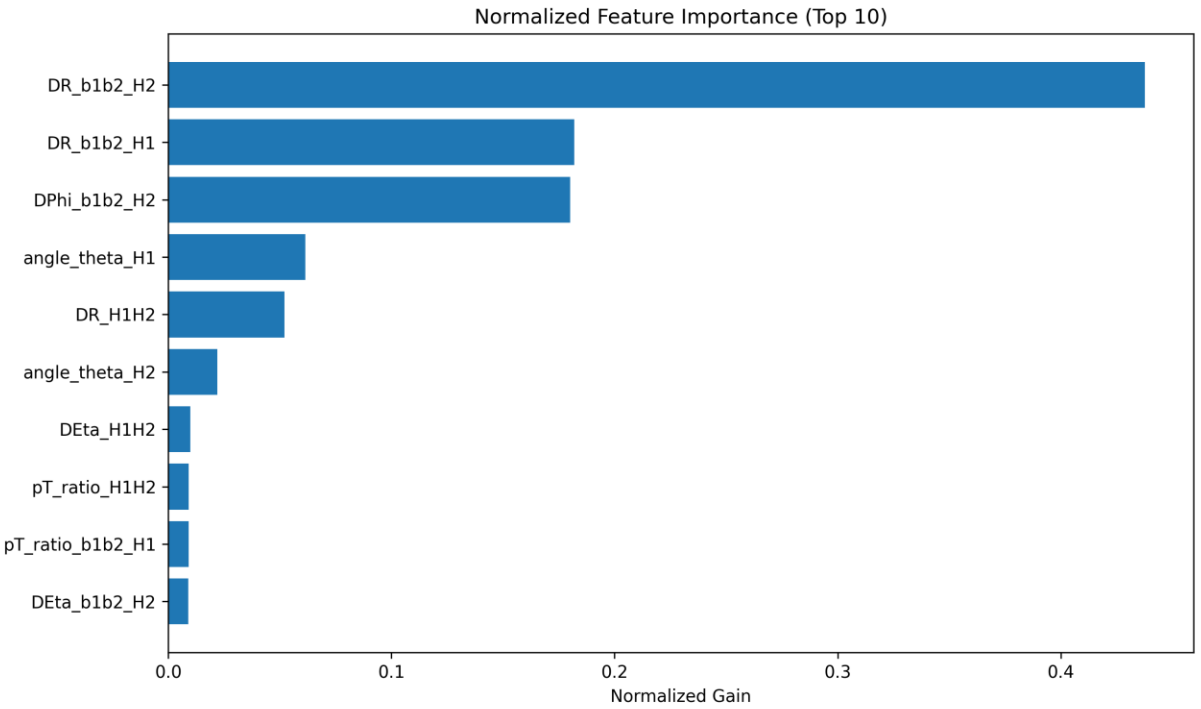
With mass parametrization



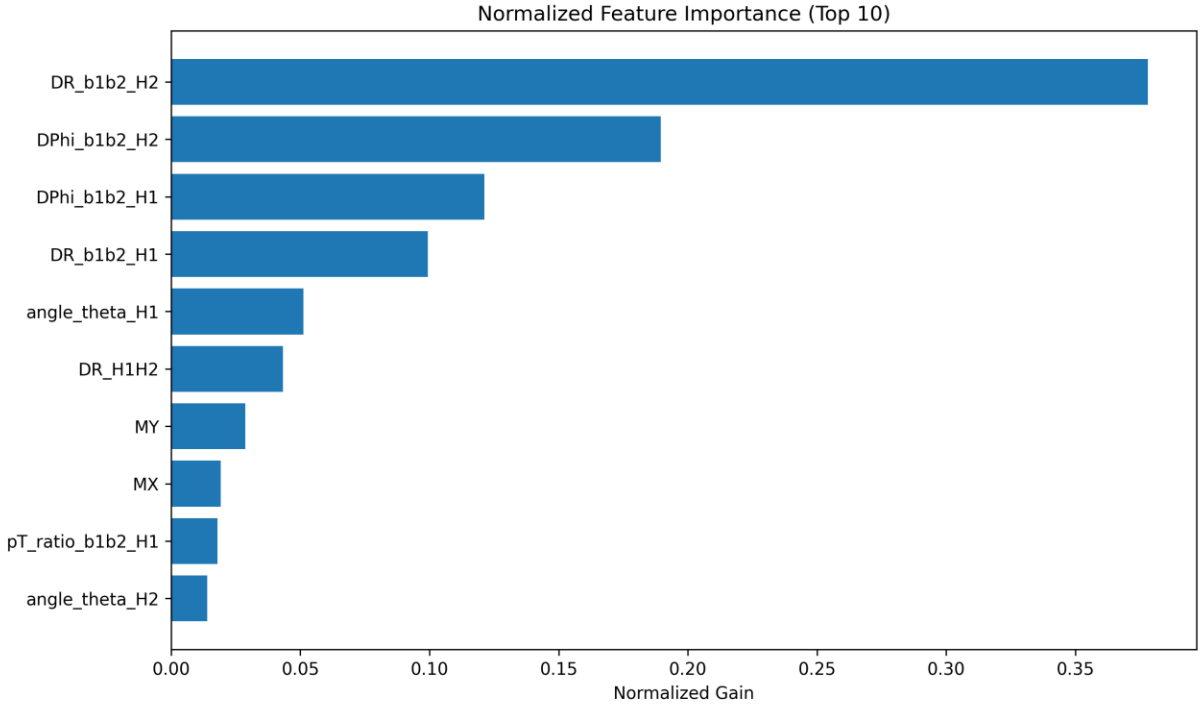
Mass parametrization = including m_X and m_Y as input features in BDT training

Feature Importance

Without mass parametrization



With mass parametrization



Efficiency Over m_y/m_x