Study of the b tagging efficiency on boosted Higgs boson

Yanxi Gu

2021.9.24

Outlines

- **Purpose of study**

- **Efficiency comparison**

  using resolved/unresolved jets to reconstruct Higgs and study the b tagging efficiency

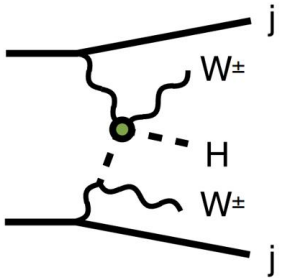
- **Sample choosing: 1 lep+boosted Higgs**

  show signal vs background cutflow tables and show the bkg reduction rate.

Purpose of Study

- To study the HWW signal, we think that at high energy, boosted Higgs boson acts more like unresolved jet instead of two resolved jets.

- The better signal efficiency of using fatjet double b tagging at high energy motivates us to study the boundary —— below which we can treat 2 jets seperately, while otherwise, we use one fatjet.



- In fact, not only HWW events can be studied. Double b tagging can be applied to other signals with boosted Higgs in general.

First round of Selections

- **Pre-selections on the skim**
  - at least one e or μ with $P_T > 35GeV$
  - at least two e,μ or τ with $P_T > 35GeV$
  - if only 2 e,μ or τ, must be same sign
  - at least 2 jets with $P_T > 30GeV$, $\Delta R(jet, lep) > 0.4$

    also, we can need 2 AK4 jets or 1 AK8 jet

- **Event selection cuts**

AK4 channel

- kinematic cuts on jets:
  Jet_cleanmask=1
  Jet_puId=7
  Jet_jetId=6
  Jet $P_T > 20GeV$
  $|\eta_{\text{jet}}| < 2.5$
- take the 2 highest b tagging score jet
  Jet_btagDeepFlavB: L,M,T

AK8 channel

- kinematic cuts on fatjets:
  $|\eta_{fat\text{jet}}| < 2.5$
  fatjet $P_T > 250GeV$
  fatjet $M > 50GeV$
  Fatjet_msoftdrop > 40GeV
- take the leading jet as fatjet
- FatJet_btagDDBvL: L,M,T

Efficiency vs Higgs Pt in region L

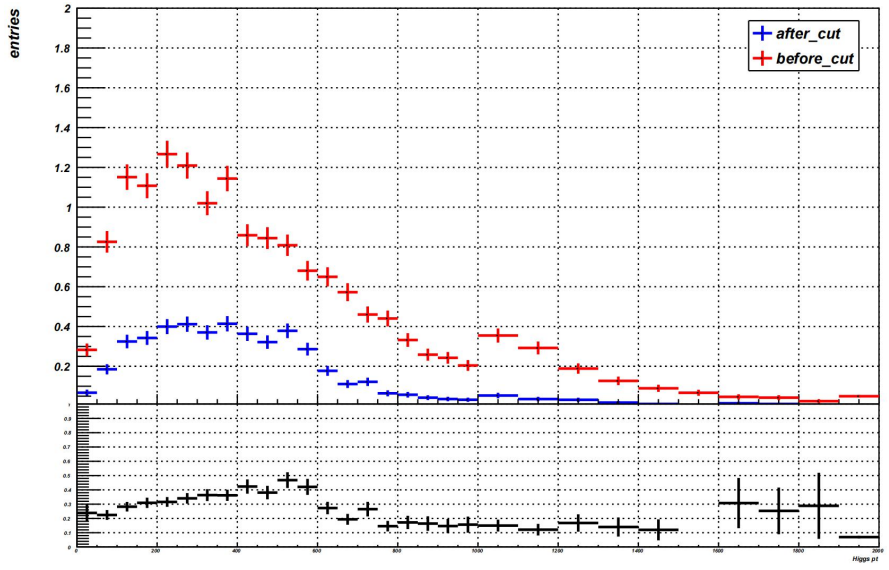


Fig 1. C2V_4p5, using AK4 jet

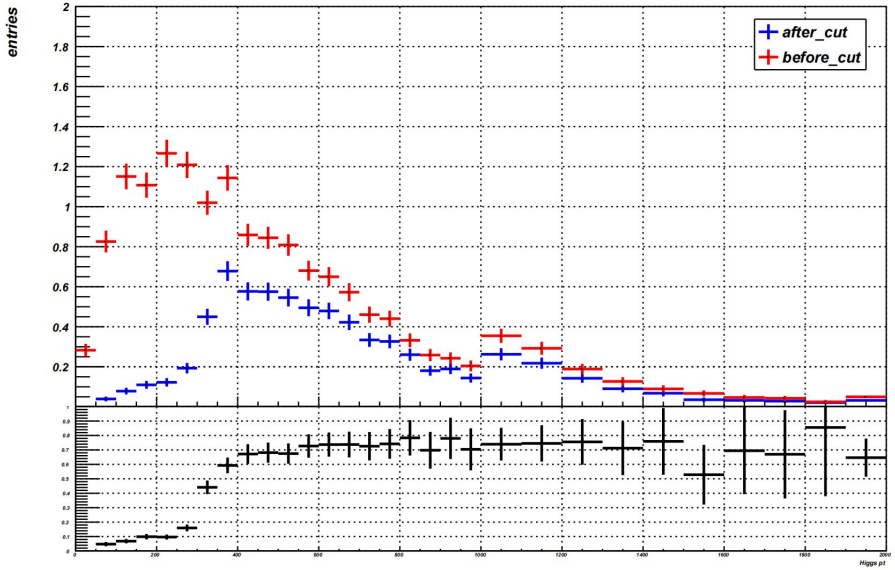


Fig 2. C2V_4p5, using AK8 jet

These plots show how the efficiency is calculated. The red line is the entries after pre-selection skimming, while the blue line is the entries after acceptance cuts and b tagging requirements. The black line is the ratio=blue/red, which means the efficiency.

The X axis of the plot is the generator level Higgs Pt. Binning is changed at 1000GeV, so there exist a twist.

Efficiency for comparison

This is the efficiency comparison for AK4 and AK8 channel signal, with larger bins on the tail

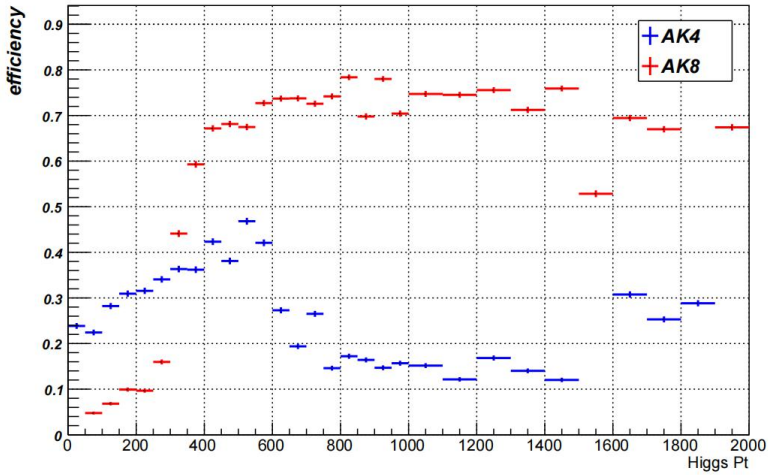


Fig 1. efficiency in region L

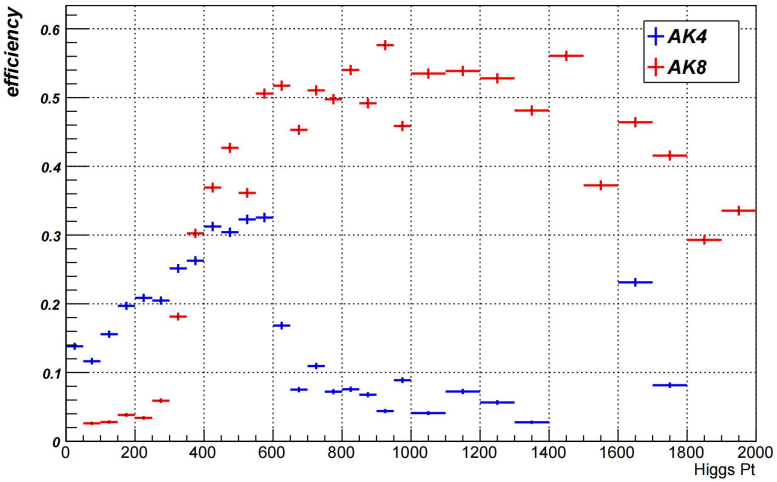


Fig 2. efficiency in region M

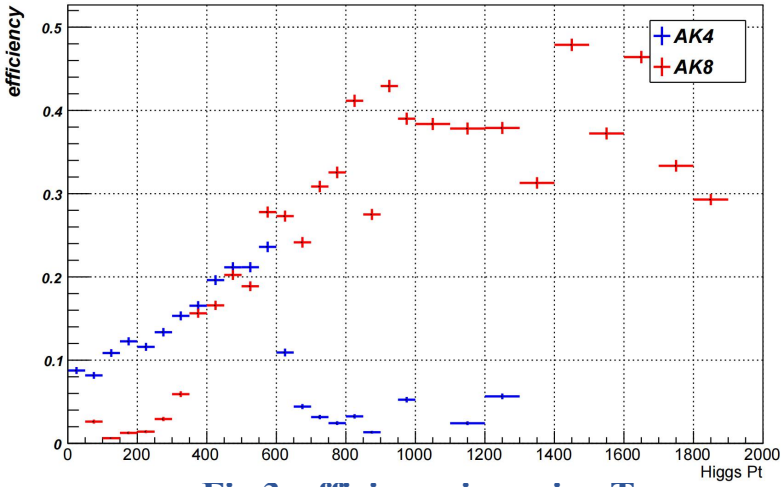


Fig 3. efficiency in region T

Note: This is the comparision for AK4 channel and AK8 channel, C2V_4p5 used. The cross is around 300GeV in region L, around 400GeV in region M, and around 500GeV in region T.

The efficiency drop at high energy in AK8 channel turns out to be sample issue. After reskimming, the curve increases with energy.

Higgs Pt in T region

- The left plot below is the signal strength plot. ttW, TTJets_DiLept, TTJets_SingleLeptFromT/Tbar backgrounds are considered. While the red line is our C2V_4p5 signal. AK8 channel.
- We have illustrated at high energy, double b tagging shows high signal efficiency. The below right plot also shows that events we selected out have high purity. We match the generation level jets with the reco level Higgs. If both gen level jets are within the cone of reco level fatjet, we tag it as “2pass”. That means, the fatjet we are tagging has very good accuracy of being the boosted Higgs instead of misidentification from elsewhere.

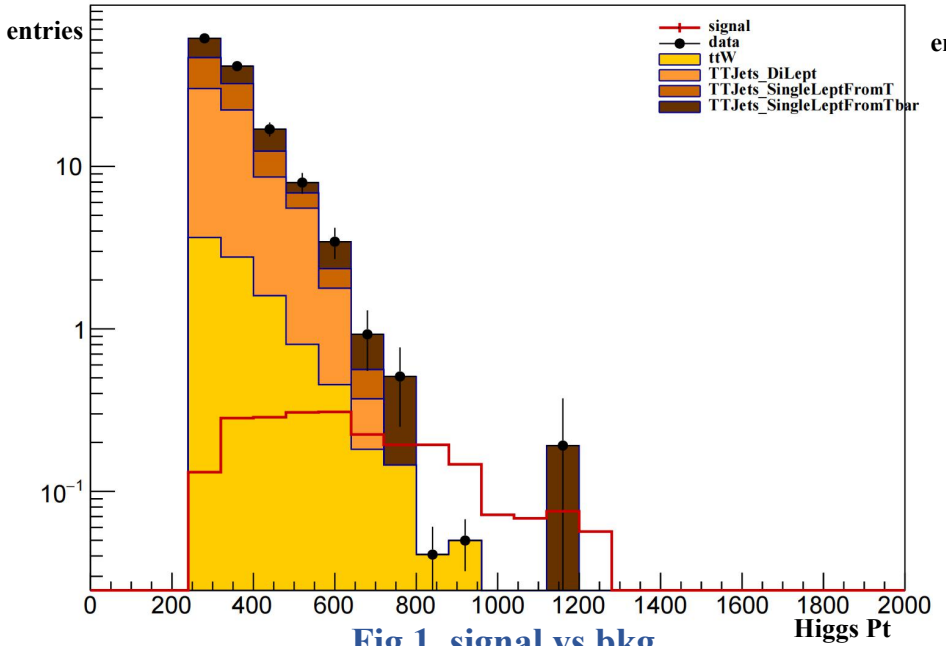


Fig 1. signal vs bkg

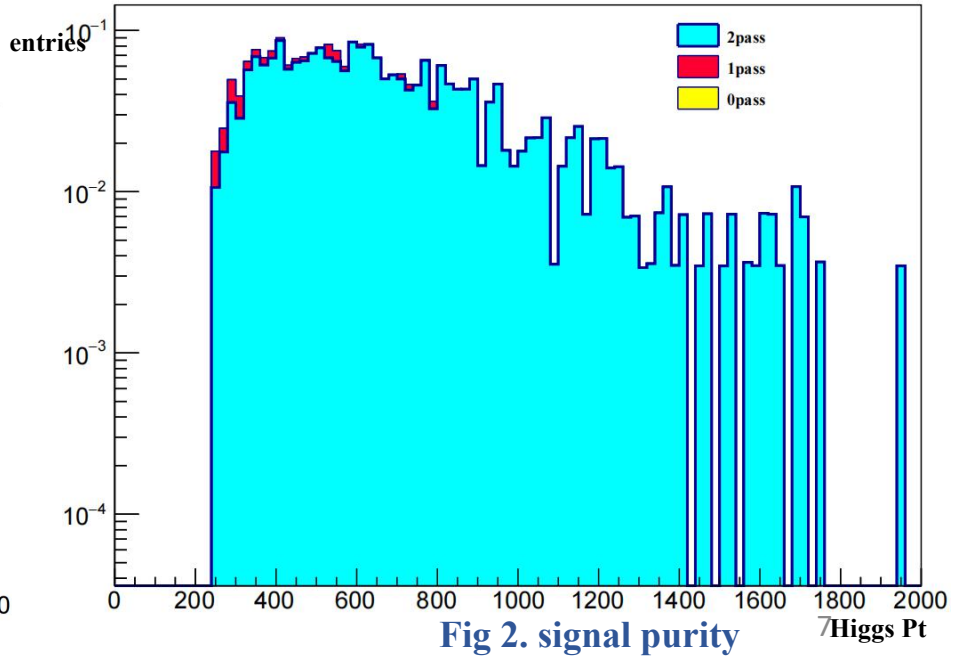


Fig 2. signal purity

New skimmed samples

We are doing another skim requiring only one lepton, as we are studying the single lepton + boosted Higgs.

One lepton can reduce some backgrounds, to make the bkgs are not too large, but on the other hand, it's not a specific requirement as for our signals. That may be useful for other signals in general. Leptons does not include tau.

- **Pre-selections on the skim**
- at least one looseID e or μ

  check tightID for leptons in loop.

- **Event selection cuts** **AK8 channel**
- at least one goodTightID lepton (e or μ)
- kinematic cuts on fatjets:

  $|\eta_{fatjet}| < 2.5$

  fatjet $P_T > 250GeV$

  fatjet $M > 50GeV$

  Fatjet_msoftdrop > 40GeV

  FatJet_Filter == true
- take the leading jet as fatjet
- FatJet_btagDDBvL: L,M,T

Entries comparison for S and B

- **Entries comparison**

	C2V_4p5	TTJets_Dilept	TTJets_SingleLeptFromT	TTJets_SingleLeptFromTbar
Raw events	347.69	5426222	10904416	10904416
Pre-selection	166.70	4681060	7016042	7015312
LeptonTightID	110.44	2621117	3057609	3058120
Acceptance cut	72.66	116208	260609	260473
Loose b tag	23.89	60329.9	101647	101457
Medium b tag	11.94	26370.6	29947.4	29884.4
Tight b tag	6.97	14499.9	13628.4	13543.9

Efficiency comparison for S and B

- **Efficiency comparison**

	C2V_4p5	TTJets_Dilept	TTJets_SingleLeptFromT	TTJets_SingleLeptFromTbar
Pre-selection	1	1	1	1
LeptonTightID	0.66	0.560	0.436	0.436
Acceptance cut	0.44	0.025	0.037	0.037
Loose b tag	0.14	0.013	0.014	0.014
Medium b tag	0.07	0.006	0.004	0.004
Tight b tag	0.04	0.003	0.002	0.002