Hbb meeting, 20.01.2016, CERN

Jet substructure in VHbb analysis

Felix Müller (Post-Doc)
Ilona Weimer (Master student)
Sandra Kortner (Senior scientist)

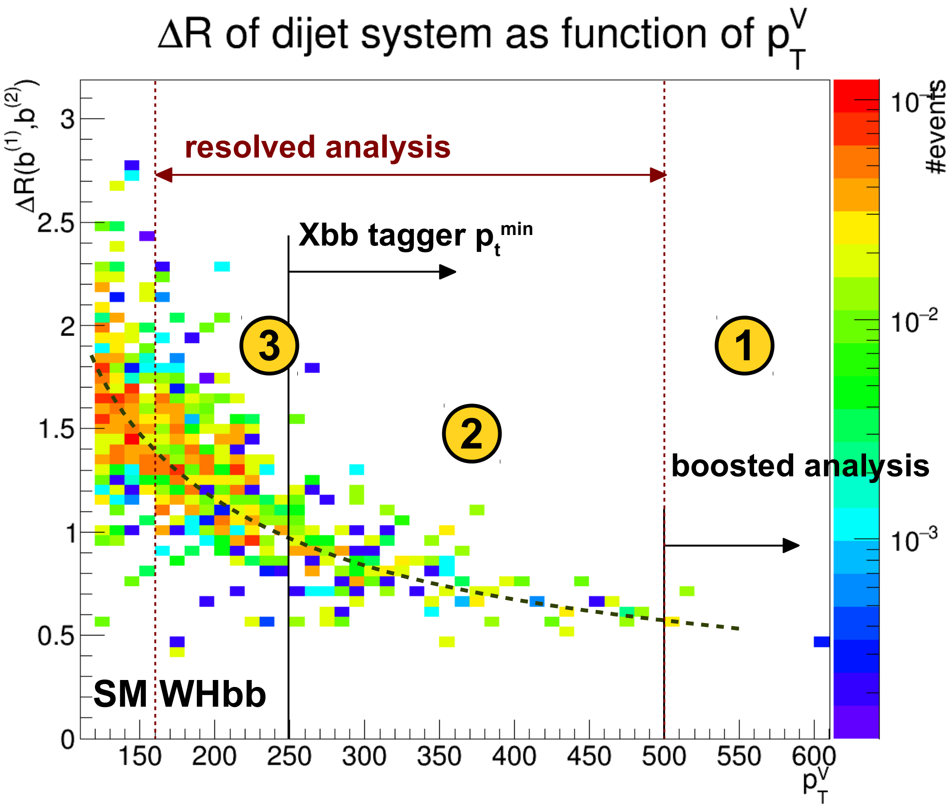
$\Delta p \cdot \Delta q \geq \frac{1}{2}\hbar$

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Introduction

Idea

- Exploit additional information from substructure inside jets
- Boosted topologies where jets cannot be resolved anymore
- Separation power of substructure in H → bb vs. ttbar, W+jets, ...



Different regimes

(1) Boosted analysis in VH resonance
Test Xbb tagger with different substructure variables

(2) Boosted analysis in SM VHbb
Benefit from Xbb tagger in SM analysis?
Which substructure variables to use?

(3) Resolved analysis in SM VHbb
Can we profit from substructure information?

In this talk: (1) + (3)

- Substructure in HVT
  → New result
- Substructure in resolved VHbb
  → Update from brainstorming session

Book keeping

Samples

- CxAOD 00-18-01
  Private production to add substructure variables
- MC15a, no pile-up reweighting
  Match PU conditions in full dataset 2016

Event selection

- 1-lep selection

Resolved analysis	Boosted analysis
$E_T^{miss} > 30$ GeV	$E_T^{miss} > 30$ GeV
$m_T^W > 20$ GeV	$m_T^W > 20$ GeV
$p_T^W > 120$ GeV	$p_T^W > 120$ GeV
$d\phi$ (E_T^{miss}, jets) > 1.0	$d\phi$ (E_T^{miss}, jets) > 1.0
$p_T^V < 500$ GeV	$p_T^{fat} > 500$ GeV
2 b-tags (leading jets, 70% WP)	Xbb tagger (medium WP)
95 GeV < m_{bb} < 140 GeV	93 GeV < m_{bb} < 134 GeV

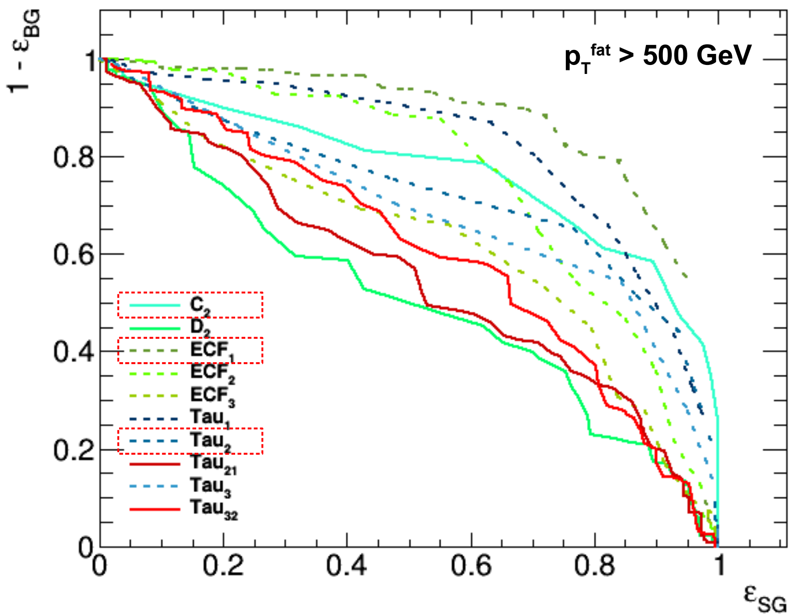
(Boosted analysis, first four cuts:) shoud be $E_T^{miss} > 120$ GeV

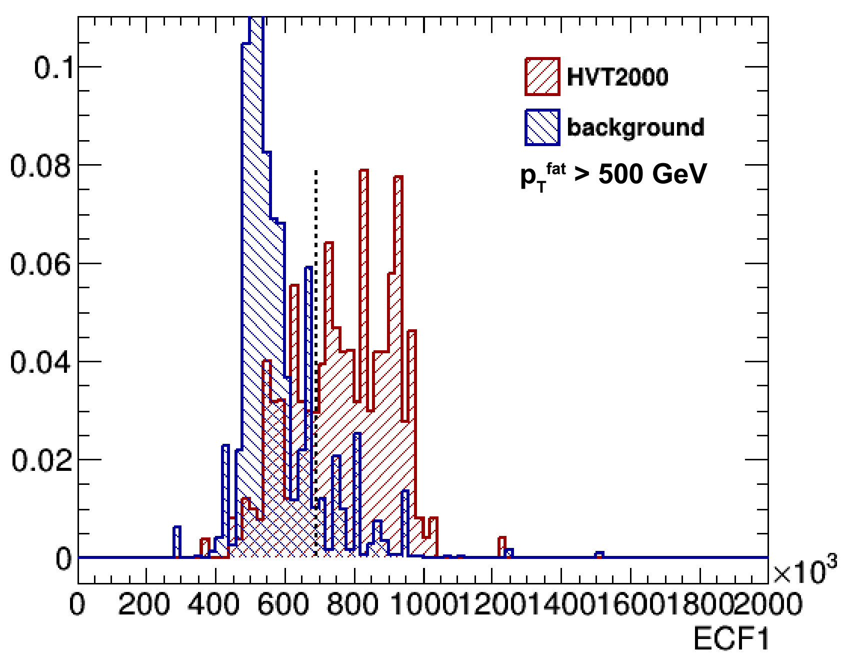
(Boosted analysis, p_T^{fat}:) shoud be p_T^V

- Muon-in-jet correction applied [OneMu]
- Simplified overlap removal

1 Boosted analysis for VHbb resonances

Boosted VHbb resonance





Analysis

- Signal HVT model m=2000 GeV
- 1-lep, anti-QCD cuts
- p_T^{fat} > 500 GeV
- Xbb tagger medium WP

Study substructure variables

- Using Xbb tagged jet
- Study energy correlation functions (ECF) and n-subjettiness (Tau) and ratios thereof (D2, Tau21, ...)
- Simple one-sided cut on distribution of substructure variable

ROC curves

- No separation power from D2
- ECF1, C2, Tau1 seem promising

ECF1 distribution

- Clear separation of signal and background
- Some statistical fluctuations, but result should hold

Substructure variable in boosted VHbb resonance

by mistake

$95 < m_{bb} < 133$ GeV, 3.209 fb^{-1}, $p_T^{fat} > 500$ GeV

	Xbb medium + jss			
	cut	S	B	Z
Xbb medium	-	15.9 (15.7)	3.49 *(3.90)*	**5.88** ***(5.64)***
Xbb tight	D2(p_T)	*(14.0)*	*(2.25)*	***(6.02)***
Tau21	0.965	15.9	3.49	**5.89**
Tau32	0.175	15.9	3.49	**5.89**
D2	2.3125	15.1	3.05	**5.89**
Tau21_wta	0.695	15.8	3.36	**5.93**
ECF3	4.65E+014	15.8	3.34	**5.94**
Tau32_wta	0.835	15.9	3.43	**5.92**
ECF2	41E+009	8.8	0.44	**6.21**
Tau3_wta	0.035	14.6	2.24	**6.23**
Tau3	0.035	13.4	1.60	**6.36**
Tau1_wta	0.145	14.3	1.80	**6.48**
Tau2	0.065	14.8	1.94	**6.53**
Tau2_wta	0.045	13.5	1.42	**6.58**
Tau1	0.115	10.0	0.46	**6.75**
C2	0.125	14.2	1.45	**6.79**
ECF1	690000	11.5	0.40	**7.57**

using $p_T^V > 500$ GeV

Quantitative approach

- Significance in m_{bb} window to estimate performance of tagger

$$\sigma = \sqrt{2 \cdot \{ (s+b) \cdot \ln(1+(s/b)) - s \}}$$

- Look for relative improvements
  Numbers don't indicate improvements on m_{VH} or limits

Results

- Tight WP better than Medium in 1-lep
  p_T dependent D2 cut
- Simple D2 cut does not improve
- **ECF1 gives best result (+28%)**

Feedback from Xbb group

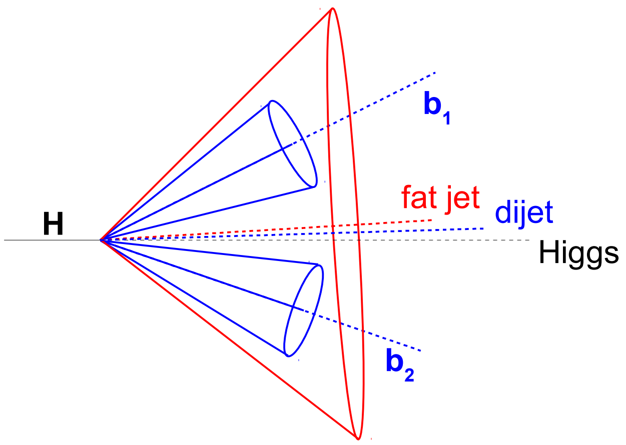
- Studied QCD, ttbar rejection
  W+jet might be different
- Studied only C2, D2
  No improvements seen on ttbar from C2
- C2 uncertainty could be derived on short time scale
- Cross-checks and deeper understanding required

3 Resolved analysis in SM VHbb

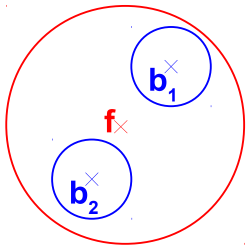
3. Substructure variable in resolved VHbb SM

Low-p_T Xbb tagger

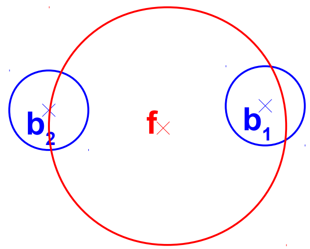
- Prerequisites
  - Large radius parameter
    → include full substructure information
  - Avoid calibration of Large-R jet
    → using only substructure information, not m_{bb}
    → small dependency on calibration
- Simplistic approach
  - Baseline is resolved analysis
  - Use m_{bb} from b-jets
  - Match Large-R jet to dijet system
    → look at substructure variables only
- Matching
  - Substructure from both b-jets must be included
  - dR(fat jet, b-jet$_{1,2}$) < 1.0
  - Not many events for p_T^{fat} < 250 GeV;
    need larger radius parameter



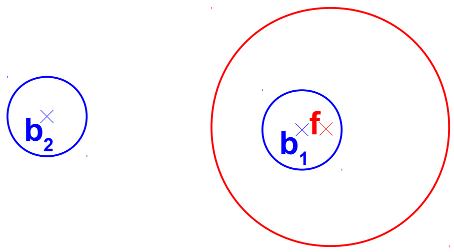
good



ugly



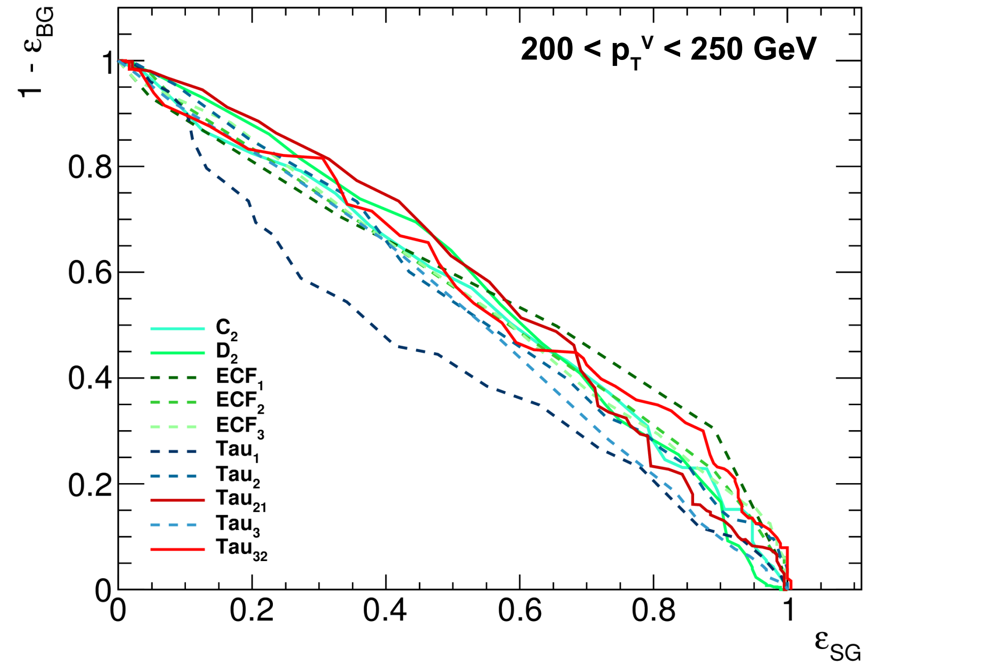
bad

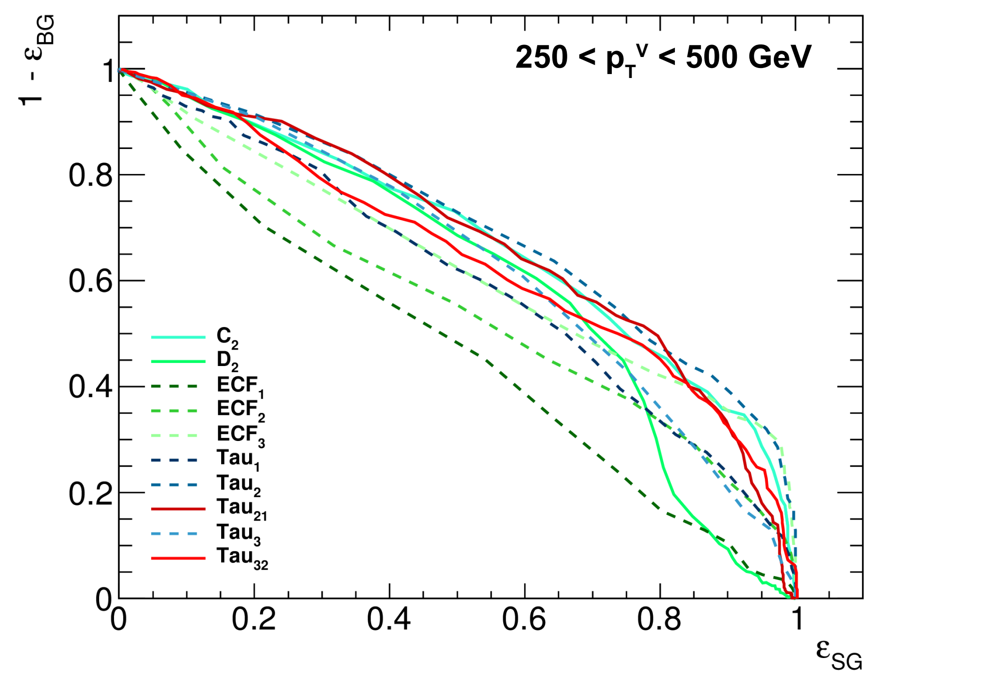


3. ROC curves for resolved VHbb SM

Analysis

- Signal: SM VHbb
- 1-lep, anti-QCD cuts
- Two regions: 200 GeV < p_T^V < 250 GeV
  250 GeV < p_T^V < 500 GeV
- Mass window: 95 < m_{bb} < 140 GeV
- Two event categories
  - Matched category
    Require small-R b-jets to be inside large-R jet
  - Resolved catogry
    Standard analysis





Observation

- Not much potential for 200 < p_T^V < 250 GeV
  → Matching efficiency < 20%
  → Need larger jet radius to capture full substructure information
- Improvements for p_T^V > 250 GeV
  - Candidates: C2, Tau2, ECF3

3. Substructure variable in resolved VHbb SM

$95 < m_{bb} < 140$ GeV, 3.209 fb^{-1}, 200 GeV $< p_T^V <$ 250 GeV

	Matched			Resolved			Com
	S	B	Z	S	B	Z	Z
Resolved only				1.49	68.1	0.179	**0.179**
No substructure	0.31	8.50	**0.107**	1.18	59.6	0.152	**0.186**
D2	0.31	8.50	**0.107**	1.18	59.6	0.152	**0.186**
C2	0.30	7.37	**0.109**	1.19	60.7	0.152	**0.187**
ECF1	0.28	5.91	**0.114**	1.21	62.2	0.153	**0.191**
ECF3	0.31	7.44	**0.111**	1.24	60.7	0.159	**0.194**
Tau2	0.31	7.68	**0.110**	1.24	60.4	0.159	**0.193**

- Significance improvement due to separate categories
  (dR requirement for matching; will be exploited by MVA)
- **7% significance improvement in matched category using ECF1**
  - 2-sided cut might improve further
  - No improvement from D2, C2
- **2% improvement in total**

$95 < m_{bb} < 140$ GeV, 3.209 fb^{-1}, 250 GeV $< p_T^V <$ 500 GeV

	Matched			Resolved			Com
	S	B	Z	S	B	Z	Z
Resolved only				1.53	38.4	0.245	**0.245**
No substructure	1.18	21.9	**0.250**	0.34	16.4	0.084	**0.264**
D2	0.86	11.4	**0.252**	0.67	27.0	0.128	**0.283**
C2	1.09	14.3	**0.286**	0.43	24.1	0.089	**0.300**
ECF1	1.17	21.4	**0.251**	0.35	17.0	0.085	**0.265**
ECF3	1.15	15.3	**0.290**	0.38	23.1	0.079	**0.301**
Tau2	1.14	14.9	**0.290**	0.39	23.5	0.080	**0.301**

- Small significance improvement due to separate categories
  (almost all events in matched category)
- **16% significance improvement in matched category using ECF3**
  - C2, ECF3 and Tau2 similar
  - Small improvement from D2
- **14% improvement in total**

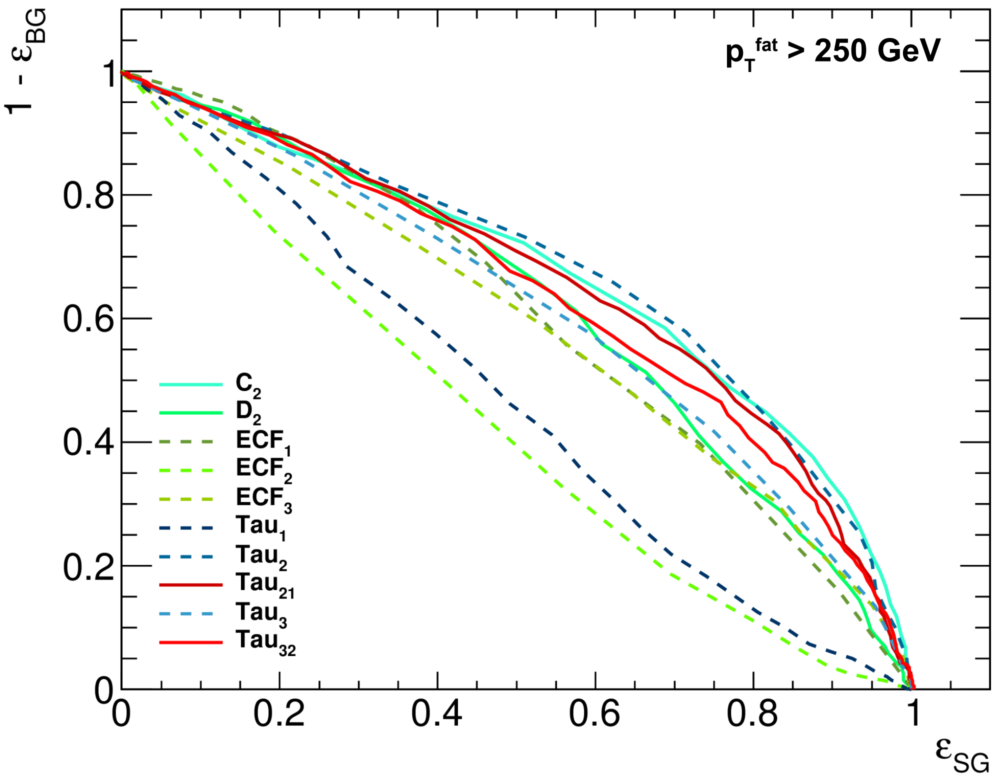
Summary and outlook

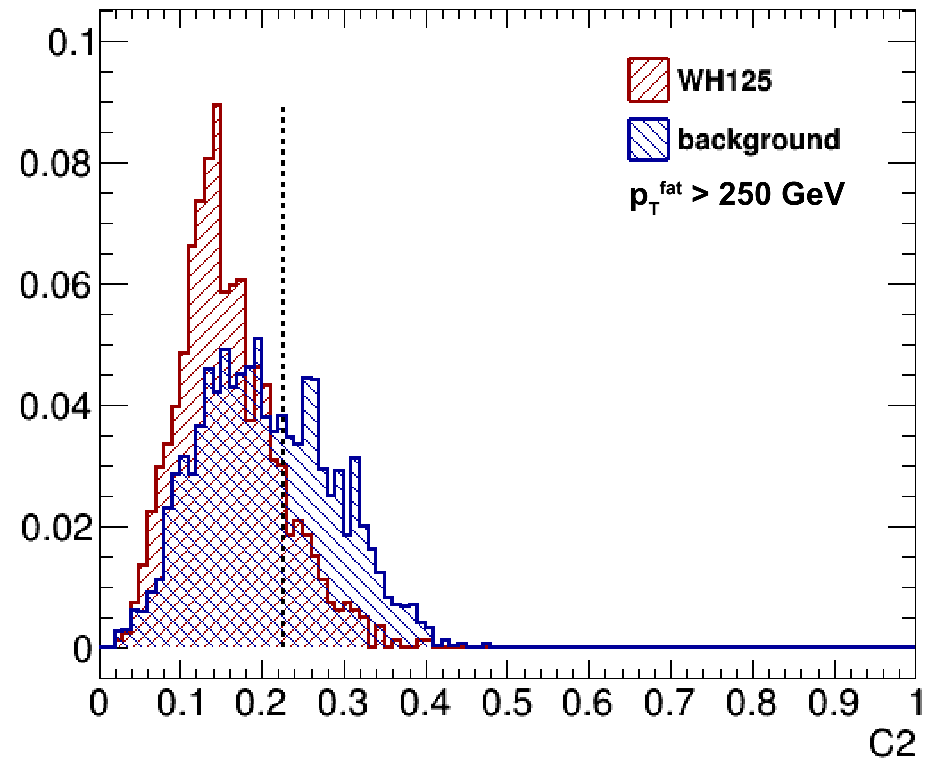
- **Study of substructure variables in VHbb 1-lep channel**
- **Boosted analysis in VHbb resonance search**
  - Much potential in variables other than D2 (up to 28% improvement in tagging)
    - Candidates: ECF1, C2, Tau1
    - Investigate why these variables work
  - Cross-checks with Xbb studies needed
    - Too late for VH resonance, analysis fixed
    - Eligible for next update of analysis
- **Resolved analysis in SM VHbb**
  - Promising results by adding substructure from large-R jet (14% in significance)
    - Candidates: ECF3, C2, Tau2; MVA analysis to combine
    - Investigate why these variable work
  - Larger radius required for low p_T
    - Study different algorithms, improve matching
  - Positive feedback from Xbb tagging group
    - Possibly large-R jet calibration / uncertainty can be avoided
    - Alternatively could try jet reclustering approach (Large-R jets built from Small-R jets)

Backup

② Boosted analysis in SM VHbb

ROC curves for boosted SM VHbb





- Event selection
  - 1-lep selection, QCD cuts
  - Xbb tagger with medium WP
- Simple one-sided cut in jss distribution, no p_T dependence
  - Variables: Tau21 Tau32 Tau1 Tau2 Tau3 [+ wta, not shown] ECF1 ECF2 ECF3 C2 D2
  - Promising candidates: C2, Tau2

Substructure variable in boosted VHbb SM

$93 < m_{bb} < 134$ GeV, 3.209 fb^{-1}, $p_T^{fat} > 250$GeV for Xbb, $p_T > 120$ GeV for resolved

	Xbb medium			Resolved			Combined
	S	B	Z	S	B	Z	Z
				8.0	900	0.266	**0.266**
Xbb medium	1.3	108	**0.126**	6.7	793	0.237	**0.268**
Xbb tight D2 (p_T)	1.0	68.0	**0.121**	7.0	832	0.242	**0.271**
Xbb medium + fixed D2 cut	1.2	93.1	**0.127**	6.8	807	0.239	**0.271**
Xbb medium + fixed C2 cut	1.2	68.0	**0.140**	6.8		0.235	**0.274**

Strategy

- Use Xbb tagger when possible
- Use resolved analysis when not Xbb tagged

Observations

- No improvement from tight WP
  → large penalty from signal loss; p_T dependent cut not ideal
- 11% significance improvement in Xbb category using C2
- 3% improvement in total from using Xbb + C2
- Only small improvement from boosted category
  → SM just not boosted enough
  → Xbb tagger not optimized / optimal for SM VHbb

Ideas for improvements

- Optimize jet substructure: choice of variable, p_T dependency, …
- Potential improvement for 1-tag + substructure category?