Boosted jets and substructures for Higgs physics

Nick Smith, on behalf of ATLAS+CMS
Higgs 2022
9 November, 2022
Outline

• History: boosted Higgs as a path to H(bb) observation?
• Developments in jet substructure
  - Mass grooming & flavor tagging
• Enabling a diverse set of Higgs and di-Higgs analyses
  - VH(bb)
  - ggH(bb)
  - VH(cc)
  - ggH(cc)
  - HH(bb bb)
  - Multi-boson resonances
Popular for some time

- Inaugural BOOST 2009

**Combined particle-level result**

- Note excellent Z peak for calibration
- 5.9 $\sigma$; potentially very competitive
- $b\bar{b}$ branching information critical for extracting Higgs properties
- Studies within ATLAS are promising and nearly public.

https://indico.cern.ch/event/56985/#3-why-subject-structure-works
The original plan

- Trigger on leptonic V decay
- Cluster into Cambridge-Aachen R=1.2 jet
  - Higgs $p_T > 200$ GeV
- Decluster until mass drop
  - If subjets are b-tagged, call it a Higgs
- $6 \rightarrow 3\sigma$ in ATLAS full-sim follow up (cds:1201444)
  - For 30/fb @ 14 TeV
A decade later

- VH(bb) observation 2018
  - ATLAS arxiv:1808.08238 CMS arxiv:1808.08242
  - Neither used a declustering technique, rather Anti-$k_T$ R=0.4 (AK4) dijet system
    - Though $p_T^V > 150$ GeV*
    - CMS: gluon radiation recovery

* Z(II) channel has additional 50(75)-150 GeV bin
Meanwhile

- 14 years of development in jet substructure
  - Latest usually found at BOOST
- Main products for our purposes:
  - Jet mass grooming
  - Jet flavor tagging
Jet mass

- ATLAS & CMS converging on approach, mostly
  - Anti-$k_T$, radius 0.8-1.5
  - Pileup suppression: CS+SK vs. Puppi
  - Grooming: soft-drop $\beta=0$ (CMS) or 1 (ATLAS Run 3), $z=0.1$

\[ \frac{\min(p_{T,j_1}, p_{T,j_2})}{p_{T,j_1} + p_{T,j_2}} > z_{cut} \left( \frac{\Delta R_{12}}{R} \right)^\beta \]

\[ j \]

\[ j_1 \]

\[ j_2 \]

\[ \text{arxiv:2009.04986} \]
Jet mass

• ATLAS & CMS converging on approach, mostly
  - Anti-\(k_T\), radius 0.8-1.5
  - Pileup suppression: CS+SK vs. Puppi
  - Grooming: soft-drop \(\beta=0\) (CMS) or 1 (ATLAS Run 3), \(z=0.1\)
    • Can send mass to zero

\[ \frac{\min(p_{T,j_1}, p_{T,j_2})}{p_{T,j_1} + p_{T,j_2}} > z_{\text{cut}} \left( \frac{\Delta R_{12}}{R} \right)^{\beta} \]

\[ m_{\text{SD}} [\text{GeV}] \]

\[ \text{A.U.} \]

\( \text{CMS} \)

\( \text{Simulation} \)

AK8

500 < \(p_T^{\text{jet}}\) < 1000 GeV, \(|\eta^{\text{jet}}| < 2.4\)

QCD multijet

W boson

Z boson

Higgs boson

\text{arxiv:2004.08262}
Jet mass

- ATLAS & CMS converging on approach, mostly
  - Anti-$k_T$, radius 0.8-1.5
  - Pileup suppression: CS+SK vs. Puppi
  - Grooming: soft-drop $\beta=0$ (CMS) or 1 (ATLAS Run 3), $z=0.1$
    - Can send mass to zero, recover with ParticleNet regression

\[ \frac{\min(p_{T,j_1}, p_{T,j_2})}{p_{T,j_1} + p_{T,j_2}} > z_{\text{cut}} \left( \frac{\Delta R_{12}}{R} \right)^\beta \]

CMS Simulation Preliminary

\[ M_{\text{target}} = \begin{cases} M_{\text{SD}}^{\text{gen}} & \text{if jet is QCD} \\ m_H \in [15, 250] & \text{otherwise} \end{cases} \]
Jet mass

- ATLAS & CMS converging on approach, mostly
  - Anti-$k_T$, radius 0.8-1.5
  - Pileup suppression: CS+SK vs. Puppi
  - Grooming: soft-drop $\beta=0$ (CMS) or 1 (ATLAS Run 3), $z=0.1$
    - Can send mass to zero, recover with ParticleNet regression
    - Well-modeled observable? Yes except at very high boost

- ATLAS [arxiv:1912.09837]
  - Soft-drop (1402.2657) observables
  - “All of the parton shower … calculations provide an excellent description of the data in most regions of phase space.”

\begin{center}
\begin{tabular}{l}
\textbf{ATLAS} \\
$\sqrt{s}=13$ TeV, 32.9 fb$^{-1}$ \\
Track-based, anti-$k_T$, $R=0.8$ \\
Soft Drop, $z_{\text{cut}}=0.1$, $\beta=0$ \\
$p_T^{\text{lead}}>300$ GeV
\end{tabular}
\end{center}
Jet flavor

- Whole large-R jet tagger wins over sub-jet tag
  - Tracks, secondary vertices along subjet axis

[Diagram and graph showing tagging efficiency and misidentification probability for different jet substructures]

arxiv:1712.07158
Jet flavor

- Whole large-R jet tagger wins over sub-jet tag
  - Tracks, secondary vertices along subjet axis

  ![CMS Simulation](image)

- Cannot calibrate using top tag & probe
  - Calibration through gluon splitting proxy jet
    - Tag via soft muons from weak B hadron decay

  ![Diagram](image)

**arxiv:1712.07158**
Jet flavor

• Gluon splitting modeling ok except for
  - Low $\rho$
    • High boost, important for EFT
  - High momentum asymmetry
    • Soft-drop: $z > 0.1$
• Common theme: use $Z$ as validation (later)

arxiv:1812.09283
Jet flavor

- Big gains with more complex DNN

arXiv:2006.13251

Latest flavor tagging: Nilotpal Kakati's talk
Jet flavor

- Big gains with more complex DNN
  - SOTA graph networks

![Graph Network Diagram](image)

**CMS Simulation**
- $450 < p_T < 1200$ GeV
- $40 < m_{b\bar{b}} < 200$ GeV
- DBT, AUC = 93.0%
- DDBT, AUC = 97.3%

**DBT, AUC = 93.0%**
**DDBT, AUC = 97.3%**

**ARXIV:2006.13251**

**Latest flavor tagging:** Nilotpal Kakati's talk

**ARXIV:1902.08570**

---

9 Nov. 2022  Nick Smith I Boosted jets and substructures for Higgs physics
Jet flavor

- Big gains with more complex DNN
  - SOTA graph networks

Latest flavor tagging: Nilotpal Kakati's talk

arXiv:2006.13251

arxiv:1902.08570

arXiv:2205.05550
Jet flavor

- Big gains with more complex DNN
  - SOTA graph networks transformers

---

alxiv:2202.03772

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>AUC</th>
<th>$H \rightarrow b\bar{b}$ Rej$_{50%}$</th>
<th>$H \rightarrow c\bar{c}$ Rej$_{50%}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFN</td>
<td>0.772</td>
<td>0.9714</td>
<td>2924</td>
<td>841</td>
</tr>
<tr>
<td>P-CNN</td>
<td>0.809</td>
<td>0.9789</td>
<td>4890</td>
<td>1276</td>
</tr>
<tr>
<td>ParticleNet</td>
<td>0.844</td>
<td>0.9849</td>
<td>7634</td>
<td>2475</td>
</tr>
<tr>
<td>ParT</td>
<td>0.861</td>
<td>0.9877</td>
<td>10638</td>
<td>4149</td>
</tr>
<tr>
<td>ParT (plain)</td>
<td>0.849</td>
<td>0.9859</td>
<td>9569</td>
<td>2911</td>
</tr>
</tbody>
</table>

 latest flavor tagging: Nilotpal Kakati's talk

---

Model complexity

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th># params</th>
<th>FLOPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFN</td>
<td>0.772</td>
<td>86.1 k</td>
<td>4.62 M</td>
</tr>
<tr>
<td>P-CNN</td>
<td>0.809</td>
<td>354 k</td>
<td>15.5 M</td>
</tr>
<tr>
<td>ParticleNet</td>
<td>0.844</td>
<td>370 k</td>
<td>540 M</td>
</tr>
<tr>
<td>ParT</td>
<td>0.861</td>
<td>2.14 M</td>
<td>340 M</td>
</tr>
<tr>
<td>ParT (plain)</td>
<td>0.849</td>
<td>2.13 M</td>
<td>260 M</td>
</tr>
</tbody>
</table>

Computation cost still under control.
Jet flavor

- Per-particle information beyond kinematics crucial for performance
  - Particle-flow ID (μ / e / charged had. / γ / neutral had.), charge
  - Secondary vertices

![Diagram of jet flavor with displaced tracks, charged lepton, heavy-flavour jet, PV, and SV.]

---

**CMS**

Simulation

$Z \rightarrow bb$ vs. QCD multijet

- $1000 < p_{T}^{gen} < 1500$ GeV, $|\eta^{gen}| < 2.4$
- $65 < m_{SD}^{AK8} < 105$ GeV

![Graph showing signal and background efficiency.]

---

Enabling boosted VH(bb)

- ATLAS boosted VH(bb) result
  - Following the 2008 proposal in spirit (with 5x data)
  - 2.1 (2.7 exp.) $\sigma$ excess

![Graph showing event distribution](image)

\[ \mu_{VH}^{bb} = 0.72^{+0.39}_{-0.36} = 0.72^{+0.29}_{-0.28} \text{(stat.)}^{+0.26}_{-0.22} \text{(syst.)} \]

\[ \mu_{VZ}^{bb} = 0.91^{+0.29}_{-0.23} = 0.91 \pm 0.15 \text{(stat.)}^{+0.24}_{-0.17} \text{(syst.)} \]

[arxiv:2008.02508]
Enabling boosted ggH(bb)

- CMS boosted ggH result
  - Anti-\(k_T\), R=0.8, PUPPI softdrop mass
  - 2.5 (0.7 exp.) \(\sigma\) excess over background

\[
\mu_H = 3.7 \pm 1.2 \text{ (stat)}^{+0.8}_{-0.7} \text{ (syst)}^{+0.8}_{-0.5} \text{ (theo)} \\
\mu_Z = 1.01 \pm 0.05 \text{ (stat)}^{+0.20}_{-0.15} \text{ (syst)}^{+0.13}_{-0.09} \text{ (theo)}
\]

[Link to arXiv:2006.13251]

High-\(p_T\) Z(bb): tagging uncertainty larger than theory
Enabling boosted ggH(bb)

- ATLAS boosted ggH result
  - Anti-$k_T$, $R=1$, trimmed mass
  - $\mu_H = 0.8\pm3.2$

$arxiv:2111.08340$
Improving STXS tails

- ATLAS, CMS: full Run-2 VH(bb) STXS
  - Both include boosted categories
  - Merged jet: CMS: AK8, ATLAS: AK10
- Strong constraint at \( p_T(V) > 400 \)

Further details on CMS VH(bb): Nicholas Haubrich’s talk

**HIG-20-001**

**ATLAS-CONF-2021-051**
Improving EFT constraints

- ATLAS result: SMEFT interpretation
  - Boosted/resolved crossover ~ 400 GeV
- Combination resolves degeneracies
  - Complementary kinematic regimes

\[ \begin{align*}
  &\text{Resolved only} \\
  &\text{Boosted only} \\
  &\text{Overlap}
\end{align*} \]
Enabling boosted VH(cc)

- CMS full Run II VH(cc)
  - Combines resolved and merged-jet channels
    - Merged: one AK15 ParticleNet cc-tagged jet
- Validated using VZ(cc)
  - Excess significance 5.7 (5.9) σ
  - First observation of Z(cc) at a hadron collider
- VH(cc) upper limit: 14x SM

Further details on this result: Marko Stamenkovic's talk

**Higgs physics**

**Higgs boson signal**

- **95% CL limit on $\mu_{\text{VH} \to cc}$**
  - Combined Expected 7.60
  - Observed 14.4
  - Median expected
  - 68% expected
  - 95% expected

**Background efficiency**

- **CMS Simulation**
  - DeepAK15
  - ParticleNet

**Signal efficiency**

- **CMS**
  - $p_T > 300$ GeV, $|\eta| < 2.4$
  - $H\to c\bar{c}$ vs. $H\to b\bar{b}$
  - $H\to c\bar{c}$ vs. $V+\text{jets}$

**Fermilab**

**arxiv:2205.05550**
Enabling boosted ggH(cc)

- CMS full Run II ggH(cc)
  - Similar technique to ggH(bb): one AK8 DeepDoubleX cc-tagged jet
  - QCD estimation via CR + sideband constraint ("differential alphabet")

- Validated using Z(cc)
  - Observation of Z(cc)+jets

- ggH(cc) upper limit: 45x SM

Further details on this result: Marko Stamenkovic's talk

HIG-21-012
Towards HH

- CMS: merged jet channel for HH(4b)
  - R=1.5 jet improves low-\(p_T\) acceptance
  - Leading performance for cross section limit

* bb\(\tau\tau\) also has a boosted region, details in Jaime Holgado's talk

Further details on CMS combination: Jin Wang's talk

arxiv:2205.06667

arxiv:2207.00043, CMSPublic twiki: SummaryResultsHIG
Towards HH

- CMS: merged jet channel for HH(4b)
  - Quartic coupling $\kappa_{2V} = 0$ excluded at 6.3 $\sigma$
  - For $\kappa_\lambda$, slightly less sensitive

**CMS**

$$\kappa_i = \kappa_V = 1$$

+ Observed

- SM Higgs boson

**138 fb^{-1} (13 TeV)**

arxiv:2205.06667
CMS Experiment at the LHC, CERN
Data recorded: 2017-Aug-07 19:13:22.727552 GMT
Run / Event / LS: 300633 / 525384863 / 347
Resonant HH

- Hadronic Higgs at very high boost ➔ HH resonance search
  - And YH resonance in NMSSM model

arxiv:2204.12413

arxiv:2202.07288
Resonant multiboson

- ATLAS: MSSM A(Zh) search
  - Merged-jet category extends m_A reach

$\sqrt{s} = 13$ TeV, 139 fb$^{-1}$

0L channel, $gg \rightarrow b\bar{b}A, A \rightarrow Zh$

ATLAS Simulation

$\sqrt{s} = 13$ TeV, 139 fb$^{-1}$

All signal regions

- 1 b-tag, resolved
- 2 b-tags, resolved
- 3+ b-tags, resolved
- 1 b-tag, 0 add, merged
- 2 b-tags, 0 add, merged
- 2 b-tags, 1+ add, merged

Acceptance $\times$ efficiency

ATLAS

$95\%$ CL limit

Observed limit

Expected limit

Expected $\pm 1$ s.d.

Expected $\pm 2$ s.d.

Expected limit (0L)

Expected limit (2L)

Also relevant: CMS all-hadronic VH resonant: arxiv:2210.00043

arxiv:2207.00230
Conclusion

• Jet substructure methods are delivering results
• Boosted channels add value to LHC Higgs program
  - Across a diverse set of analyses
• Outlook: more final states!
  - Boosted $H(\tau\tau)$, $H(WW)$ on the horizon
  - $H$(gluons) ?
Backup
**UFO jets**

**Calorimeter only:**
- **LCTopo**: Topological calorimeter clusters

**Combined with tracking:**
- **PFlow**: Particle Flow Objects
  - Low $p_T$: Use *track* 4-vector for charged particles, subtract energy from *cluster* 4-vectors
  - High $p_T$: Use *cluster* 4-vectors, ignore *tracks*
- **TCC**: Track Calo Clusters
  - Low $p_T$: Use *cluster* 4-vectors, ignore *tracks*
  - High $p_T$: Split *clusters* using *tracks*, get energy from *clusters* but angles from *tracks*

**Combining PFlow and TCC:**
- **UFO** combines **TCC** and **PFlow** to achieve optimal performance over a broad kinematic ($p_T$) range

T. Fitschen
Softdrop observables ($\beta=0,1$)

(a) $\beta = 0$, calorimeter-based

(b) $\beta = 0$, track-based

(c) $\beta = 1$, calorimeter-based

(d) $\beta = 1$, track-based
Mass decorrelation

- Always one “impossible” background
  - Too expensive and unreliable to model from sim
  - For boosted H(bb), QCD. For VH(bb), V+HF
- Use data in mass sideband with “alphabet method”
- Required assumption:
  - Minimal correlation between tagger and mass in background
- Many ways to accomplish decorrelation in design
  - DDT 1603.00027, DisCo 2001.05310, adversarial training…
  - Engineer synthetic signals at many mass points

\[
\begin{align*}
D & \quad B \\
C & \quad A = C \cdot B / D
\end{align*}
\]
Boosted $Z(bb)$

CMS w.r.t. NLO (Z+1,2j FxFx) QCD + NLO EW

CMS

\[
\begin{align*}
\text{[800, 1200]} \text{ GeV} & \quad \mu_Z = 0.61^{+0.25}_{-0.21} \\
\text{[675, 800]} \text{ GeV} & \quad \mu_Z = 0.78^{+0.24}_{-0.20} \\
\text{[600, 675]} \text{ GeV} & \quad \mu_Z = 0.87^{+0.24}_{-0.20} \\
\text{[550, 600]} \text{ GeV} & \quad \mu_Z = 0.93^{+0.25}_{-0.21} \\
\text{[500, 550]} \text{ GeV} & \quad \mu_Z = 0.85^{+0.22}_{-0.18} \\
\text{[450, 500]} \text{ GeV} & \quad \mu_Z = 1.24^{+0.30}_{-0.24}
\end{align*}
\]

\[137 \text{ fb}^{-1} (13 \text{ TeV})\]

ATLAS

\[\sqrt{s} = 13 \text{ TeV}, 136 \text{ fb}^{-1}\]

Data SRL+SRS

K-Factor

- Sherpa 2.2.8 Z with EW NLO
- QCD NNLO (NNLOjets)
- QCD NNLO + EW NLO

\[\begin{align*}
\text{Ratio to QCD NLO + EW LO} \\
\text{p_T} [\text{GeV}] \\
\end{align*}\]

\[\text{arxiv:2006.13251}\]

\[\text{arxiv:2111.08340}\]
Boosted ggH(bb) unfolded

**ggH only (others fixed)**

**SM = HJ-MiNLO (ggH) + others**

**ATLAS**
\[\sqrt{s} = 13 \text{ TeV}, 136 \text{ fb}^{-1}\]

95% C.L. Upper Limit

Fit

**arxiv:2006.13251**

**arxiv:2111.08340**
ATLAS resolved+boosted VH(bb)

**Expected**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_{Hq}^{(3)}$</td>
<td>$2.0 \times 10.0$</td>
</tr>
<tr>
<td>$c_{Hu}$</td>
<td>$5.0$</td>
</tr>
<tr>
<td>$c_{HW}$</td>
<td>$2.0$</td>
</tr>
</tbody>
</table>

**Observed**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_{Hq}^{(3)}$</td>
<td>$2.0 \times 10.0$</td>
</tr>
<tr>
<td>$c_{Hu}$</td>
<td>$5.0$</td>
</tr>
<tr>
<td>$c_{HW}$</td>
<td>$2.0$</td>
</tr>
</tbody>
</table>

**ATLAS Preliminary**

- $\sqrt{s} = 13$ TeV, 139 fb$^{-1}$
- 68% CL ... 95% CL
- VH, $H \to b\bar{b}$, $\Lambda = 1$ TeV

- Best-fit, expected
- Resolved (EPJC 81 178)
- Boosted (PLB 816 136204)
- Combination

- 68% CL ... 95% CL
- VH, $H \to b\bar{b}$, $\Lambda = 1$ TeV
- Best-fit, observed
- Resolved (EPJC 81 178)
- Boosted (PLB 816 136204)
- Combination

ATLAS-CONF-2021-051
ggHH differential XS

LHC 14 TeV
PDF4LHC15
NLO, $\mu = m_{hh}/2$

$\kappa_\lambda = 1.0$  red
$\kappa_\lambda = -1.0$  green
$\kappa_\lambda = 3.0$  blue
$\kappa_\lambda = 5.0$  yellow

$\frac{d\sigma}{dm_{hh}}$ [pb/GeV]

LHC 14 TeV
PDF4LHC15
NLO, $\mu = m_{hh}/2$

$\kappa_\lambda = 1.0$  red
$\kappa_\lambda = -1.0$  green
$\kappa_\lambda = 3.0$  blue
$\kappa_\lambda = 5.0$  yellow

$\frac{d\sigma}{dp_T^h}$ [pb/GeV]

arxiv:1903.08137