Boosted Heavy Particles and Jet Substructure with the CMS Detector
Introduction

Boosted regime: **classical selection methods fail.**

- multiple, **well separated jets** from W, Higgs or top hadronic decays
- typical jet collection: **AK5** (anti-$k_T$, R=0.5)
- decay products from heavy particles merged into large fat-jets
- jet substructure provides fundamental selection tools (top-tagging, W-tagging...)
- typically, **larger jet collections**: **CA8** (Cambridge-Aachen, R=0.8) or **CA15**
Substructure Techniques
W-Tagging
[CMS-PAS-JME-13-006]

Based on jet mass pruning (Ellis, Vermillion, Walsh [arXiv:0903.5081], [CMS-PAS-SMP-12-019]).

Starting with CA8 jets.

Re-cluster jet and apply requirements when merging clusters \(i\) and \(j\) into cluster \(p\).

Veto soft and large angle re-combinations, removing softer component if:

\[
\begin{align*}
\text{(1)} & \quad \min(p_T^i, p_T^j)/p_T^p < 0.1 \\
\text{(2)} & \quad \Delta R_{ij} > 0.5 \frac{m_{\text{orig}}}{p_T^{\text{orig}}}
\end{align*}
\]

W-tagging:
- 2 pruned subjets
- pruned jet mass \([60, 100] \text{ GeV}\)

Signal: resonance (600 GeV) → WW
QCD background
Pruning can be combined with additional observables:

- mass-drop $\mu$
- N-subjettiness $\tau_N$: $\tau_2/\tau_1$ used for W-tagging
- also examined: Qjet volatility $\Gamma_{QJet}$
- generalized energy correlation function $C_2^\beta$

N-subjettiness shows the best single discriminating power.

Observables are correlated: moderate improvement with multivariate combination using TMVA.

Efficiency: $H \rightarrow WW$, $m_H = 600$ GeV

Mistag: QCD
W-Tagging: MC vs Data
[CMS-PAS-JME-13-006]

- Detailed data/MC comparisons for all substructure observables
- Different topologies and generators considered

**ttbar**
- **powheg**
- **mc@nlo**

Sensitive to: efficiency

**leptonic W+jets**
- **Pythia6**
- **Herwig**

Mistag
- pT [250, 350] GeV

**di-jets**
- **MG+Pythia6**
- **Herwig**
- **Pythia8**

Mistag
- pT [400, 600] GeV

general good agreement, more observables in the backup
Scale factors (SF) to correct for residual discrepancies.

Extract:

- **W-jet mass scale (peak position):**
  - Data: 84.5±0.4 GeV
  - MC: 83.4±0.4 GeV

- **W-jet mass resolution:**
  - Data: 8.7±0.6 GeV
  - MC: 7.5±0.4 GeV

- **data/MC correction for W-tagging efficiency (SF):**
  - 0.905 ± 0.08
  (operating point: \(m_{\text{pruned}}\) cut + \(\tau_2/\tau_1 < 0.5\))
Based on JHU top-tagger (Kaplan et al [PRL 101 (2008) 142001]):

- start with CA8 jets
- reverse clustering sequence and examine clusters pairwise
- clusters are split if:
  \[ \Delta R > 0.4 - 0.0004 \ p_T^C \]
  \( p_T^C \) is the parent cluster \( p_T \)
- low \( p_T \) clusters removed if:
  \[ p_T < 0.05 \ p_T^{jet} \]

Top-tagger requirements:

- \( 140 < m_{jet} < 250 \ \text{GeV} \)
- \( N_{subjets} \geq 3 \)
- Min pairwise mass > 50 GeV

ROC Curves

top-mistag (QCD) vs efficiency in simulation
Performance
[CMS-PAS-B2G-12-005]

**μ+jets: semileptonic ttbar**

![Histogram of Top Candidate Jet Mass](image1)

- **Data**
- t\(\bar{t}\)
- W+jets
- Z+jets
- Single Top

**QCD**

**Mistag** rate can be measured from data, using **anti-tag method**:
- two high-\(p_T\) jets, \(p_T > 400\) GeV
- anti-tag one jet, inverting min pairwise mass requirement
- top-tag of other jet is a mistag

![Histogram of Minimum Pairwise Mass](image2)

- **Data**
- t\(\bar{t}\)
- W+jets
- Z+jets
- Single Top

**top-tagging data/MC scale factor derived from selection efficiency of hadronic top candidate: 0.93±0.04**
B-Tagging in Boosted Topologies
[CMS-PAS-BTV-13-001]

- B-tagging at CMS traditionally developed on isolated AK5 jets, mostly suitable for the non-boosted regime.

- First study at LHC dedicated to b-tagging in the boosted regime. Benchmark topologies:
  - Boosted top, hadronic decay: selected using HEPTopTagger [JHEP 1010 (2010) 078], CA15 jet collection
  - Boosted Higgs → bb: studies based on pruned CA8 jets

- CSV developed on AK5 jets: currently no dedicated re-training for the boosted regime.
Two scenarios considered:

- **subjet CSV:**
  - CSV b-tagger applied to subjets (2 b-tags for Higgs-tagging, \( \geq 1 \) for top-tagging)

- **fat-jet CSV:**
  - CSV b-tagger applied to the Higgs/top candidate fat-jet

**Subjet b-tagging** generally performs better: chosen as **default technique**

**Fat-jet b-tagging** suitable at **very high** \( p_T \)

where subjets start to merge
Control samples
Boosted top:

- $\mu+\text{jets}$, semileptonic $\text{t}\overline{\text{t}}\text{bar}$

**Boosted Higgs**: challenging definition of the control sample

- similar topology: **gluon splitting jets**, two closeby b's

---

**e.g. Top channel, HEPTopTagger subjets**

- Good data/MC agreement for b-tagging observables.
- All observables cross-checked (backup).

---

**e.g. subjets of gluon splitting CA8 fat-jets**

- SF~1, compatibly with SF for standard b-tagging in the non-boosted regime, for both channels.
- Nothing pathological in the boosted regime.
Pile-Up Jet-ID  
[CMS-PAS-JME-13-005]

- Traditional PU subtraction: subtract charged particles not pointing to the primary vertex.

- PU Jet-ID:
  - exploit also **non-tracking quantities** (jet shape) to extend PU rejection outside of the tracking acceptance
  - **multivariate discriminant**

### Z ($\rightarrow\mu\mu$) + jets events

<table>
<thead>
<tr>
<th>$\mu\mu$ + jets events</th>
<th><strong>PU jets</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt;</td>
<td>\eta</td>
</tr>
</tbody>
</table>

\[
\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R^2_{iT} p^2_{Ti}}{\sum_i p^2_{Ti}}
\]

- **Example non-tracking observable:** radial distribution of Particle-Flow jet-constituents

- Known residual discrepancy due to out-of-time pile-up simulation
Performance:
- tag-and-probe method from $Z \rightarrow \mu\mu$ + jets events, where probe is jet recoiling against $Z$
- data/MC agreement within 10%, corrected using SF

Several applications:
- e.g. extensions of jet vetos to low $p_T$ (Higgs searches)
Quark-Gluon Discrimination
[CMS-PAS-JME-13-002]

Quark/gluon discrimination: similarly to PU Jet-ID, combine discriminating variables in likelihood

Variables:

- **multiplicity:**
  - charged, neutral, total

- **spread:**
  - $\eta-\phi$ spread
  - major $\eta-\phi$ matrix axes $\sigma_1$
  - minor $\eta-\phi$ matrix axes $\sigma_2$

- **energy sharing:**
  - hardest candidate off-centering/energy

\[ p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \]

combined in likelihood

Quark and gluon have different colour interaction:

- quark jet
- gluon jet

+ multiplicity
+ width
more homogeneous energy sharing
Quark-Gluon Discrimination
[CMS-PAS-JME-13-002]

- Quark/gluon discrimination: similarly to PU Jet-ID, combine discriminating variables in likelihood

- Quark and gluon have different colour interaction:

  - Quark jet
  - Gluon jet

  + multiplicity
  + width
  more homogeneous energy sharing

Single-variable and combined likelihood discrimination power
Quark-Gluon Discrimination
[CMS-PAS-JME-13-002]

- Validation in two different samples:
  - **Z+jets**: quark enriched
  - **di-jets**: gluon enriched

- Overall good data/MC agreement. Some discrepancy at low $p_T$ in di-jets, probably due to gluon fragmentation mismodeling. Covered by systematics.

- Useful tool for several searches:
  - many channels with jets are **flavor specific**
  - pioneer analyses at CMS:
    - Higgs→ZZ→2l2q
      [JHEP 04 (2012) 036]
    - VBF Higgs→bb
      [CMS-PAS-HIG-13-011]
Searches Using Substructure
Resonances $\rightarrow$ ttbar All-Hadronic Final State


Flagship for boosted searches for new physics.
Sensitive to several models. Considered:
- extra dimensions, RS gluon
- extended gauge, $Z'$
  - narrow $\Gamma/m=0.01$
  - broad $\Gamma/m=0.1$

Selection:
- 2 back-to-back high $p_T$ jets
- both top-tagged

exclusion limits from combination with semi-leptonic channel
exclusion up to 2.7 TeV depending on the channel

High-purity version of the analysis underway:
reduce QCD with combination top-tagging+subjet b-tagging
Vector-like heavy quarks predicted by several theories:
- little/composite Higgs models
- extra dimensions
Solution to the hierarchy problem.

Signal:
- pair-produced $B'$ with charge $-1/3$
- decay modes: $B'\rightarrow tW, bZ, bH$
- all branching fractions

Selection:
- single muon or electron
- substructure used in event categories based on number of V-tags ($V=W/Z/H$):
  - CA8 jet, $p_T \geq 200$ GeV
  - mass drop $\mu < 0.4$
  - 2 pruned subjets
  - $m_{\text{pruned}} [50,150]$ GeV

final limit up to 732 GeV
Top Partners
[CMS-PAS-B2G-12-012 and 015]

- [CMS-PAS-B2G-12-012] Signal:
  - pair-produced \( T' \) with charge 5/3
  - BR 100% \( T' \rightarrow tW \)

- Selection:
  - two same sign leptons
  - top-tagging
  - \( W \)-tagging \( (m_{\text{pruned}} \ [60,130] \text{ GeV}) \)

- [CMS-PAS-B2G-12-015] Signal:
  - pair-produced \( T' \) with charge 2/3
  - decay modes: \( T' \rightarrow tH, tZ, bW \)
  - all branching fractions

- Two final states:
  - multilepton: counting experiment, no substructure
  - single lepton: multivariate analysis, number of \( W \)- and top-tags enter the BDT discriminant.

observed limit 770 GeV

final combined limit up to 782 GeV
High Mass Dibosons  
[CMS-PAS-EXO-12-021/024]

Predicted by several models. Here considered:

- **bulk graviton** production: $G_{\text{bulk}} \rightarrow WW \rightarrow l + \text{jet} + \text{MET}$

- **RS graviton**, $W$ heavy partner $W'$:
  $G_{\text{RS}} \rightarrow WW/ZZ$, $W' \rightarrow WZ$
  all-hadronic decay

W-tagging:
- $m_{\text{pruned}} [65,105] \text{GeV}$
- N-subjettiness:
  - high-purity $\tau_2/\tau_1 < 0.5$
  - low-purity $0.5 < \tau_2/\tau_1 < 0.75$

- double $W/Z$-tagging:
  $m_{\text{pruned}} [70,100] \text{GeV}$, same $\tau_2/\tau_1$ cuts as above

- $G_{\text{bulk}} \sigma \times \text{BR}_{WW}$ limits between 70 fb and 3 fb
- $G_{\text{RS}} \rightarrow WW$ exclusion: [1.00, 1.59] TeV
- $G_{\text{RS}} \rightarrow ZZ$ exclusion: [1.00, 1.17] TeV
- $W' \rightarrow WZ$ exclusion: up to 1.73 TeV
Substructure techniques

- major developments recently: subjet b-tagging, W-tagging, pile-up jet-ID, gluon/quark discriminator, ...
- **new results on top-tagging expected soon**
- extensive data/MC comparisons: generally good agreement

Searches:

- increased number of analyses using substructure, beyond typical ttbar resonance searches
- **searches exploiting powerful new tools** (subjet b-tagging, new top-taggers, ...) **expected before the end of the year**
Substructure techniques

- major developments recently: subjet b-tagging, W-tagging, pile-up jet-ID, gluon/quark discriminator, …
- new results on top-tagging expected soon
- extensive data/MC comparisons: generally good agreement

Searches:

- increased number of analyses using substructure, beyond typical ttbar resonance searches
- searches exploiting powerful new tools (subjet b-tagging, new top-taggers, …) expected before the end of the year
Additional Slides
W-Tagging: Additional Observables

Pruning can be combined with additional observables:

- mass-drop
Pruning can be combined with additional observables:

- **mass-drop**

  \[
  \mu = \frac{m_1}{m_{\text{jet}}}
  \]

  \(m_1\) is the highest mass pruned subjet
Pruning can be combined with additional observables:

- **mass-drop**
- N-subjettiness $\tau_N : \tau_2 / \tau_1$ used for W-tagging

**probability that jet is composed by N subjets**

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min \{ \Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k} \}$$

$$d_0 = \sum_k p_{T,k} R_0, \text{ and } R_0 \text{ is the original jet radius}$$
Pruning can be combined with additional observables:

- **mass-drop**
- N-subjettiness $\tau_N: \tau_2/\tau_1$ used for W-tagging
- also examined: Qjet volatility $\Gamma_{QJet}$, generalized energy correlation function $C_2^\beta$
Pruning can be combined with additional observables:

- **mass-drop**
- **N-subjettiness** $\tau_N : \tau_2/\tau_1$ used for W-tagging
- **Qjet volatility** $\Gamma_{QJet}$

$$\text{RMS (mass jet trees)} / m_{\text{jet}}$$

where a jet is interpreted as a distribution of trees based on its clustering sequence
Pruning can be combined with additional observables:

- **mass-drop**
- N-subjettiness $\tau_N = \tau_2 / \tau_1$ used for W-tagging
- Qjet volatility $\Gamma_{QJet}$
- generalized energy correlation function $C_2^\beta$

$$C_2^\beta = \frac{\sum_{i,j,k} p_T^i p_T^j p_T^k (R_{ij} R_{ik} R_{jk})^\beta \sum_i p_T^i}{(\sum_{i,j} p_T^i p_T^j (R_{ij})^\beta)^2}$$

based on momentum and pair-wise angles of particles within the jet
B-Quark Signatures

Life-time b-hadron $\rightarrow$ jets with:
- secondary vertex
- tracks with large impact parameter

Large mass, $\sim 5$ GeV

Fragmentation function:
- high $p_T$ of the b-hadron relatively to jet $p_T$

B-decay produces often leptons: soft muon or electron within jet

Several taggers implemented at CMS. Boosted studies based on the Combined Secondary Vertex CSV tagger:
- likelihood ratio combination of secondary vertex + single track information;
- currently the best tagger in CMS, improvements ongoing.

b-tagging algorithms ROC curves
[JINST 8 (2013) P04013]
B-Tagging at CMS

- **jet-tracks association**: static cone
  \[ \Delta R(\text{tracks}, \text{jet}) < 0.3 \]

- **OBSERVABLES**
  - apply tight selection on tracks, mainly for pile-up rejection
  - determine b-tagging observables

- **DISCRIMINATORS**
  - calculate b-tagging discriminators
  - several operating points defined for taggers, selecting different regions of purity/efficiency:
    - loose \( L \); 10% misidentification from light quarks/gluons
    - medium \( M \); 1% misidentification from light quarks/gluons
    - tight \( T \); 0.1% misidentification from light quarks/gluons
B-Tagging Algorithms

- Boosted studies based on the Combined Secondary Vertex CSV tagger:
  - likelihood ratio combination of secondary vertex + single track information;
  - currently the best tagger in CMS, improvements ongoing.

- For performance measurements used also Jet-Probability JP tagger:
  - likelihood estimate of the probability that the jet-tracks come from the PV, based on the IP significance of all jet-tracks;
  - calibrated on data from tracks with negative IP.
Based on **CA8 jet collection**: boosted regime for $p_T > 300$ GeV.

- Signal: $B' \rightarrow bH$ pair production. B-tagging studied on $H \rightarrow bb$.

- Inclusive **mistag** from QCD and mistags from hadronically-decaying $W/Z/top$.

- Subjet $b$-tagging based on pruned subjets:
  - cut on **pruned jet mass** can be combined with $b$-tagging requirement (see next slides).
Top Channel

- Based on CA15 collection, default for HEPTopTagger.

- Large cone-size allows to reach lower $p_T$'s (~200GeV) without switching from merged-top to unmerged top selection.

- Signal: $T' \rightarrow tH$ pair production. Consistency of the results checked also on SM ttbar production.

- Inclusive mistag from QCD.

- HEPTopTagger forces 3 filtered subjets: used for subjet b-tagging.
B-Tagging Performance

Higgs channel

**Subjet b-tagging** performs better

Fat-jet **b-tagging** suitable at very high $p_T$

Top channel

Overall **subjet b-tagging** performs better
Tagging Performance

Higgs channel

Higgs-tagging = double b-tagging + 75 < m_{jet} < 135 GeV

Top channel

QCD mistag rate reduced up to a factor 10 with minor loss of efficiency
Challenging definition of the control sample. Similar topology: **gluon splitting jets**, two closeby b's clustered in the same fat-jet.

Event selection:
- 1 CA8 jet, $p_T > 400$ GeV, $|\eta|<2.4$;
- $\Delta R(\text{subjets}) > m_{\text{jet}}/p_T$: remove infrared unsafe configurations;
- MC samples: inclusive and muon-enriched QCD, $tt, Z\to qq$.

**Muon-tag** to b-enrich subjets sample: require muon with $p_T > 5$ GeV within subjet cone.

Sample of CA8 fat-jets enriched in gluon splitting, requiring both subjets to be muon-tagged: **Higgs-like sample**.
ttbar semi-leptonic decays.

- Leptonic decay:
  - isolated muon;
  - 1 standard b-tag.

- Hadronic decay selected using HEPTopTagger.

- MC samples: ttbar + all SM backgrounds (single-top, Z/W+jets).
Lifetime Tagger Method

Method based on Jet-Probability b-tagger. Advantage:
- JP discriminant can be defined for most jets (>90%);
- calibrated on data.

Template fit to JP discriminant, before and after applying CSV. Discriminant shape from MC, while relative flavor fractions are free parameters.

Tagging efficiency in data given by ($C_b$ is fraction of jets for which JP computable):

$$\epsilon_b^{\text{tag}} = \frac{C_b \cdot f_b^{\text{tag}} \cdot N_{\text{data}}^{\text{tag}}}{f_b^{\text{before tag}} \cdot N_{\text{data}}^{\text{before tag}}}$$
LT method applied to individual muon-tagged subjets of CA8 fat jets (w/ and w/o the companion subjet b-tagged).

Very good agreement with the standard scale factors.

Results for the loose operating point of CSV.
Mistag Scale Factor

- Measurement of \( \text{mistag rate \ SF} \) for CA8 subjets based on negative taggers, which use tracks with negative impact parameter.
- Very good agreement with the standard scale factors.
Flavor Tag Consistency Method

- Method based on distribution of number of b-tags for the 3 subjets of CA15 HEPTopTagged fat-jet: expected distribution fitted to data, with scale factors as free parameters.

- Expected number $n$ of tags for ttbar signal can be expressed as:

$$\langle N_n \rangle = \mathcal{L} \cdot \sigma_{t\bar{t}} \cdot \varepsilon \cdot \sum_{i,j,k} F_{ijk} \sum_{i'+j'+k' = n} [C_{i'}^{i'} \varepsilon_b^{i'} (1 - \varepsilon_b)^{(i-i')} C_{j'}^{j'} \varepsilon_c^{j'} (1 - \varepsilon_c)^{(j-j')} C_{k'}^{k'} \varepsilon_l^{k'} (1 - \varepsilon_l)^{(k-k')}]$$

- $\varepsilon_b$, $\varepsilon_c$, $\varepsilon_l$ are the tagging efficiencies;
- $C^a_b$ are the binomial coefficients;
- $F_{ijk}$ are the fractions of events with $i$ b-subjets, $j$ c-subjets and $k$ light-subjets: taken from MC.
- backgrounds included in the fit.
Fit Modalities

2 parameters fit:
- $\sigma_{tt}$, SF$_{b}$ are free parameters.
- Fixed SF$_{c} = $ SF$_{b}$ and fixed SF$_{light}$ to SF$_{light}$ for standard b-tagging on AK5 jets.

3 parameters fit:
- $\sigma_{tt}$, SF$_{b}$ and SF$_{light}$ are free parameters. Fixed SF$_{c} = $ SF$_{b}$.

Excellent data/MC agreement after fit of subjet b-tag multiplicity.

Post-fit distribution
Measured $S_{b}$ for boosted top subjets are in agreement with standard $S_{b}$ for AK5 jets.

No significant deviation at high top-$p_T$ of the measured $S_{b}$.

Mistag $S_{\text{light}}$ are in agreement with standard $S_{\text{light}}$ for AK5 jets.

<table>
<thead>
<tr>
<th>SF$_b$ for non-boosted jets</th>
<th>CSVL</th>
<th>CSVM</th>
<th>CSVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{b}$ for HEPTopTagger subjets</td>
<td>1.010±0.013</td>
<td>0.970±0.013</td>
<td>0.950±0.015</td>
</tr>
<tr>
<td>$S_{b}$ for HEPTopTagger subjets</td>
<td>1.003±0.026</td>
<td>0.979±0.023</td>
<td>0.960±0.036</td>
</tr>
<tr>
<td>$p_T \geq 350$ GeV/$c$</td>
<td>—</td>
<td>0.978$^{+0.023}_{-0.023}$</td>
<td>—</td>
</tr>
<tr>
<td>$p_T \geq 450$ GeV/$c$</td>
<td>—</td>
<td>0.993$^{+0.034}_{-0.034}$</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SF$_{\text{light}}$ for non-boosted jets</th>
<th>CSVL</th>
<th>CSVM</th>
<th>CSVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{\text{light}}$ for HEPTopTagger subjets</td>
<td>1.080$^{+0.063}_{-0.072}$</td>
<td>1.136$^{+0.090}_{-0.110}$</td>
<td>1.088$^{+0.039}_{-0.086}$</td>
</tr>
<tr>
<td>$S_{\text{light}}$ for HEPTopTagger subjets</td>
<td>1.185±0.080</td>
<td>1.580±0.47</td>
<td>—</td>
</tr>
</tbody>
</table>

pT dependence
Track Sharing

- Cross-check of sharing of tracks selected for b-tagging between subjets.
- Consider tracks in a cone of $\Delta R < 0.3$ around subjet axis (as used by CSV).

- Track-sharing increases with $p_T$ of the fat-jet. At very high boost, the level of track sharing becomes significantly large. One solution is to switch to fat-jet b tagging.
Mistag SF

- Use tracks with negative IP or SV with negative decay length to define a negative tagger for each tagger.
- Scale factor for mistag obtained according to:

\[ SF_{\text{mistag}} = SF_{\text{neg tag}} \cdot \frac{R_{\text{data}}}{R_{\text{MC}}} \]

given by:
B-tagging Observables

- Checking data/Monte Carlo agreement for b-tagging quantities.
  
  Presentation ordering:

  - Top channel validation: HEPTopTagger Subjets
  - Higgs channel validation: Multijet sample (CA8 jets)
  - Higgs channel validation: Multijet sample (CA8 muon-tagged subjets)
  - Higgs channel validation: Multijet sample (double muon-tagged CA8 jets)
3D Impact Parameter
Secondary Vertex Multiplicity

CMS Preliminary, 19.8 fb$^{-1}$ at $\sqrt{s} = 8$ TeV

**μ+jets sample (HEPTopTagger subjets)**

- Data
- b quark
- c quark
- uds quark or gluon

**Multijet sample (CA8 jets)**

- Data
- b quark
- c quark
- uds quark or gluon
- t\bar{t}
- Z→q\bar{q}

**Multijet sample (Muon-tagged CA8 subjets)**

- Data
- b quark
- c quark
- uds quark or gluon
- t\bar{t}
- Z→q\bar{q}

**Multijet sample (Double-muon-tagged CA8 jets)**

- Data
- b quark
- c quark
- uds quark or gluon
- t\bar{t}
- Z→q\bar{q}
SV Flight Distance Significance
Secondary Vertex Mass
Secondary Vertex Mass

Overall good data/Monte Carlo agreement, at the same level as standard b-tagging
Vector-like heavy quarks are part of several theories:

- little/composite Higgs models
- extra dimensions

Solution to the hierarchy problem.

Signal:

- pair-produced T' with charge 5/3
- BR 100% $T' \rightarrow tW$

Selection:

- two same sign leptons
- top-tagging
- W-tagging ($m_{\text{pruned}} [60,130] \text{ GeV}$)

Limits from event yields.

reconstruction of T' mass from all channels
Vector-like heavy quarks are part of several theories:
- little/composite Higgs models
- extra dimensions

Solution to the hierarchy problem.

Signal:
- pair-produced \( T' \) with charge \( 5/3 \)
- BR 100%

Selection:
- two same sign leptons
- top-tagging
- \( W \)-tagging

Limits from event yields.
- observed limit 770 GeV

reconstruction of \( T' \) mass from all channels
T'\(2/3\) Top Partners [CMS-PAS-B2G-12-015]

Signal:
- pair-produced T' with charge 2/3
- decay modes: T'\(\rightarrow tH, tZ, bW\)
- all possible branching fractions

Combination of two analysis strategies:

- **Multivariate analysis, single lepton:**
  - two event categories: with or without \(W\)-tag
  - top-tagging applied
  - relevant observables combined in BDT:
    - multiplicity/\(p_T\) of reconstructed objects (lepton, jets, tagged jets...)
    - \(N\) of \(b\)-, \(W\)- and top-tags

- **Counting experiment, multilepton channel. No substructure.**

BDT discriminant, single \(\mu\) channel

final combined limit up to 782 GeV
**B' -1/3 Bottom Partners**

**[CMS-PAS-B2G-12-019]**

**Signal:**
- pair-produced $B'$ with charge $-1/3$
- decay modes: $B' \rightarrow tW, bZ, bH$
- all possible branching fractions

**Selection:**
- single muon or electron
- $\geq 4$ AK5 jets, $\geq 1$ b-tagged
- event categories based on number of V-tags ($V=W/Z/H$):
  - CA8 jet, $p_T \geq 200$ GeV
  - mass drop $\mu < 0.4$
  - 2 pruned subjets
  - $m_{\text{pruned}} [50,150]$ GeV

**Limits based on $S_T$ distribution:**

$$S_T = p_T^{\text{lept}} + p_T^{\text{miss}} + \sum p_T^{\text{jets}}$$

**$S_T$ distribution, for 1 V-tag category**
Predicted by several models. Here considered:

- **bulk graviton production:** $G_{\text{bulk}} \rightarrow WW \rightarrow l + \text{jet} + \text{MET}$

- **RS graviton, W heavy partner $W'$**
  
  $G_{\text{RS}} \rightarrow WW/ZZ, W' \rightarrow WZ$

  all-hadronic decay

- **W-tagging:**
  - $m_{\text{pruned}} [65,105] \text{GeV}$
  - N-subjettiness:
    - high-purity $\tau_2/\tau_1 < 0.5$
    - low-purity $0.5 < \tau_2/\tau_1 < 0.75$

- **double W/Z-tagging:**
  - $m_{\text{pruned}} [70,100] \text{GeV}$, same $\tau_2/\tau_1$ cuts as above
High Mass Dibosons
[CMS-PAS-EXO-12-021/024]

- $G_{\text{bulk}} \times \text{BR}_{WW}$ limits between 70fb and 3fb
- $G_{RS} \rightarrow WW$ excluded between [1.00, 1.59] TeV
- $G_{RS} \rightarrow ZZ$ excluded between [1.00, 1.17] TeV
- $W' \rightarrow WZ$ excluded up to 1.73 TeV
Performance:
- tag-and-probe method from 
  $Z (\rightarrow \mu\mu)$ + jets events, where 
  probe is jet recoiling against $Z$
- data/MC agreement within 
  10%, corrected using SF

Several applications:
- e.g.: extensions of jet vetos 
  to low $p_T$ (Higgs searches)
W-Tagging
[CMS-PAS-JME-13-006]

Event topologies considered

- **Dijet**
  - two hard jets
  - $p_T = 400-600$ GeV
  - accesses high $p_T$ region
  - QCD-jet dominated
  - used to study fake rate

- **W+jets**
  - leptonic $W +$ jet
  - $p_T = 250-350$ GeV
  - accesses low $p_T$ region
  - QCD-jet dominated
  - presence of non-dominant background ($t\bar{t}$, single top)
  - used to study fake rate

- **$t\bar{t}$**
  - leptonic top decay + hadronic top
  - highly pure sample of $W$-jets
  - used to study efficiency

Benchmark signal: $X \rightarrow W_L W_L$, $M_X = 600$ GeV, 1 TeV
W-Tagging: MC vs Data
[CMS-PAS-JME-13-006]

- Detailed data/MC comparisons for all substructure observables
- Different topologies and generators considered

- leptonlic $W+\text{jets}$
  - $\text{Pythia6}$
  - $\text{Herwig}$

- mistag
  - pT [250, 350] GeV

- di-jets
  - $\text{MG+Pythia6}$
  - $\text{Herwig}$
  - $\text{Pythia8}$

- mistag
  - pT [400, 600] GeV
Detailed **data/MC comparisons** for all substructure observables

**Different topologies and generators considered**

- ttbar
- powheg
- mc@nlo
- *efficiency*

- leptonic W+jets
- Pythia6
- Herwig

- di-jets
- MG+Pythia6
- Herwig
- Pythia8

- mistag
- pT [250, 350] GeV

- mistag
- pT [400, 600] GeV
MVA correlations

**Signal**

**Background (W+jets)**
W-Tagging

[CMS-PAS-JME-13-006]

Substructure variables: mass drop, \( \mu \)

\[ \rho_T = 250 - 350 \text{ GeV} \]

(W+jet) - no pruned mass cut

\[ \rho_T = 250 - 350 \text{ GeV} \]

(W+jet) - pruned mass cut

Good discrimination power

Discrimination power reduced: correlation with mass cut

Emanuele Usai BOOST13
**W-Tagging**
[CMS-PAS-JME-13-006]

Substructure variables: N-subjettiness

Three variants considered:
- $\tau_2/\tau_1$: one step optimization of the $k_T$ subjet axes
- $\tau_2/\tau_1$ $k_T$ axes: no optimization
- pruned $\tau_2/\tau_1$: uses only pruned constituents + one pass optimization.

$p_T = 400 - 600$ GeV

(dijet) - no pruned mass cut

$p_T = 400 - 600$ GeV

(dijet) - pruned mass cut

Emanuele Usai BOOST13
W-Tagging
[CMS-PAS-JME-13-006]

Performance in function of $p_T$

Performance studied for: $60 < m_{jet} < 100$ GeV + $\tau_2/\tau_1 < 0.5$

Efficiency vs $p_T$ (W+jets topology)

- low $p_T$: W decay products begin to be reconstructed inside CA8 jets
- high $p_T$: detector resolution for jet substructures degrades, pruning remove too much of the mass of the W

Fake rate vs $p_T$ (dijet topology)

- drops at high $p_T$ similarly to efficiency

Emanuele Usai BOOST13
W-Tagging
[CMS-PAS-JME-13-006]

Performance in function of number of vertices

Efficiency vs Nvtx (W+jets topology)
- slight degrade of performance
- jet pruning fails to remove all soft contributions

Fake rate vs Nvtx (dijet topology)
- constant behavior with respect to Nvtx

Emanuele Usai BOOST13
W-Tagging
[CMS-PAS-JME-13-006]

Quark and Gluons w/Substructure

- Quark/gluon separation vs W same after cuts
- Mass cut more effective on quark separation
- N-subjettiness more effective on gluon separation
  - Once mass the cut is applied
Jet charge, $Q^\kappa$

$$Q^\kappa = \frac{\sum_i q_i (p_T^i)^\kappa}{(\sum_i p_T^i)^\kappa}$$

Used to discriminate between $W^+$ and $W^-$

Right plot, note: $\langle \text{jet charge} \rangle \neq 0$
Jet charge distribution

$\tau \bar{\tau}$ sample for $W^+$ and $W^-$ jets in simulation and data. Simulated distributions are a sum of all processes.

$\tau \bar{\tau}$ semileptonic selection

By selecting on the lepton charge, we can isolate $W^+$ from $W^-$ jets.
W-Tagging
[CMS-PAS-JME-13-006]

Polarization studies

- Polarization can affect substructure distribution
- Sample used: scalar $\chi \to W_{\text{lept}} L W_{\text{had}}^L$ and $\chi \to W_{\text{lept}} U W_{\text{had}}^T$

![Diagram](image)

- Parton level helicity angle for hadronic $W$
- Observable helicity angle from subjets

Emanuele Usai BOOST13
W-Tagging
[CMS-PAS-JME-13-006]

Polarization studies - $\tau_2/\tau_1$

- pruned jet mass acceptance different for $W_L$ and $W_T$
- $\Delta R$ between partons smaller on average for $W_L$
- $W_L$ more likely to be accepted by CA8 jet
- in $W_T$ topology $p_T$ of the subjets is more asymmetric, thus more QCD-like
Pile-Up Jet-ID
[CMS-PAS-JME-13-005]

Pileup Jet Id Algorithm: Tracking

- 13 variables for the full discrimination
- 4 Vertexing related variables (2 most imp! shown):
  \[ \beta = \frac{\sum_{i \in PV} p_{Ti}}{\sum_{i} p_{Ti}} \]
  \[ \beta^* = \frac{\sum_{i \in otherPV} p_{Ti}}{\sum_{i} p_{Ti}} \]

Pileup tends to degrade performance of these variables
Pile-Up Jet-ID
[CMS-PAS-JME-13-005]

Pileup Jet Id Algorithm: Cones

- Additional shape variables: $\Delta R$ annuli

$10^3$ CMS Preliminary, $\sqrt{s} = 8$ TeV $L=20$ fb$^{-1}$

$Z \rightarrow \mu \mu$

$|\eta| < 2.5$ Jet $p_T > 25$ GeV

Pileup peaks at in $0.1 < \Delta R < 0.2$

$\Delta R 0.1 => 0.5$
# Pile-Up Jet-ID

[CMS-PAS-JME-13-005]

## Algorithm Construction

- **Construct a Boosted decision tree real vs PU Jets**

  - Train in four separate regions of $\eta$

| $|\eta| < 2.5$ tracking | $2.5 < |\eta| < 2.75$ Weak tracking (tracking ends at 2.5) |
|------------------------|--------------------------------------------------------|
| Shape variables        | Shape variables                                       |

| $2.75 < |\eta| < 3.0$ Shape variables | $3.0 < |\eta| < 5.0$ Forward HCAL Shape variables |

Construct a Boosted decision tree (trained on $Z+$jets for each)
Pile-Up Jet-ID
[ CMS-PAS-JME-13-005 ]

08/13/13

Philip Harris BOOST

Pileup Jet Id in Data

- Fraction of pileup grows with higher $|\eta|$
Quark-Gluon Discrimination
[CMS-PAS-JME-13-002]

Single-variable ROCs and likelihood combination
Quark-Gluon Discrimination

[CMS-PAS-JME-13-002]

Single variables discrimination power
Quark-Gluon Discrimination

[CMS-PAS-JME-13-002]

**di-jets: derived corrections**

**Z+jets: applied corrections. Very good closure**
Quark-Gluon Discrimination
[CMS-PAS-JME-13-002]

Discrimination power slightly decreases after smearing
Quark-Gluon Discrimination

**[CMS-PAS-JME-13-002]**

- Quark/gluon discrimination: similarly to PU Jet-ID, combine discriminating variables in likelihood

- Quark and gluon have different colour interaction:
  - Quark jet
  - Gluon jet
  - + multiplicity
  - + width
  - more homogeneous energy sharing

**good discrimination power in different $\eta$, $p_T$ ranges**
Quark-Gluon Discrimination

QG Performance + Usage
- QG discrimination used in VBF selection
  - Reduces the QCD/Pileup bkgs for forward jets
- QG discrimination used in Z boson tagging
- Reduction of 60% gluon for 80% quark eff

VBF $H \rightarrow bb$ Search

H → ZZ → 2l2q Search

Old version of QG Likelihood