Search for strong production of vector-like quarks at CMS

Sadia Khalil
University of Kansas
On behalf of CMS Collaboration

29th International Symposium on Lepton Photon Interactions at High Energies
Toronto, Canada
Aug 5-10, 2019
Introduction

• Solution to naturalness invokes the existence of top quark partners, that has sizable mixing with third generation of SM quarks.

• Top partners can be boson (stop) or fermion (VLQ). If VLQ exists, can
  • Induce electroweak breaking and explain the observed lightness of the Higgs boson
  • Emerge as fermion resonances in partial-compositeness of theory of flavor
  • Introduce new source of CP violation
Vector-like Quarks

- They are spin 1/2, non-chiral particles with its own mass: $\mathcal{L}_M = - M \bar{\psi} \gamma^\mu \psi$
- Left-and right-handed chiralities transforms under the same representation of the SM gauge symmetry: $SU(3)_C \times SU(2)_L \times U(1)_Y$

  - SM Chiral Quarks: $J^{\mu^+} = J^{\mu^+}_L = V - A$ ;  
  - VLQs: $J^{\mu^+} = J^{\mu^+}_L + J^{\mu^+}_R = V$

- VLQ Production:

- Decays and allowed combinations

<table>
<thead>
<tr>
<th>Q Ele. Charge</th>
<th>Decays</th>
</tr>
</thead>
<tbody>
<tr>
<td>T^{2/3}</td>
<td>bW^+, tH, tZ</td>
</tr>
<tr>
<td>B^{-1/3}</td>
<td>tW^-, bH, bZ</td>
</tr>
<tr>
<td>X^{5/3}</td>
<td>tW^+</td>
</tr>
<tr>
<td>Y^{-4/3}</td>
<td>bW^-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$T$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SU(2)_L multiplet</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Charge</td>
<td>2/3</td>
<td>-1/3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$(X/T)$</th>
<th>$(T/B)$</th>
<th>$(B/Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(5/3, 2/3)$</td>
<td>$(2/3, -1/3)$</td>
<td>$(-1/3, -4/3)$</td>
<td></td>
</tr>
<tr>
<td>$(5/3, 2/3)$</td>
<td>$(2/3, -1/3)$</td>
<td>$(-1/3, -4/3)$</td>
<td></td>
</tr>
</tbody>
</table>
• Experimentally
  • VLQ are investigated at large mass, $M(\text{VLQ}) \geq 700 \text{ GeV}$
  • $W$, $Z$, $H$ and top, when decaying hadronically can be reconstructed as a single jet
    • Investigation of boosted topologies
• Interpretation
  • Pair production \(\rightarrow\) not only set limits on a benchmark but also scan all possible BR

![Branching Ratio Graph](image1)

---

**Exclude Triangles not Points**

- M. E. Peskin
- SEARCH workshop
- U of Maryland
- March 2012
Boosted Jets

- The decay products such as W/Z/H bosons and top quark decaying hadronically, receive more boost from a heavy particle.

Jet Grooming: remove soft and wide-angle radiation from within the jet.

Dramatically improves the separation of QCD and top quark jets.

Unmerged hadronic top quark in lab frame (resolved jets)

Fully merged hadronic top quark in lab frame (boosted jets)
Boosted Jets Techniques

- Algorithms to tag **boosted** W/Z/H bosons and top quark decaying hadronically against QCD jets
- Use **wide jets** after grooming
- Use **jet mass** and **jet substructure** to discriminate between 0/2/3 subjets inside the wide jet → N-subjettiness

![Graphs showing tagging efficiency and jet mass distributions](image1.jpg)

- More taggers in backup!
Boosted Event Shape Tagger (BEST) algorithm

- A multi-object classification machine learning algorithm that allows to search in all VLQ decay channels at once rather than single final states

- Transform the assumed particle from laboratory frame to the rest frame

- Use event angles & kinematics to build a neural network discriminant, and classify a wide jet as: W, Z, H, top, b or light quark jet

\[ A_L = \frac{\sum_{\text{jet}} p_{\text{jet}}^L}{\sum_{\text{jet}} p_{\text{jet}}} \]

Fox-Wolfram moments

Longitudinal jet asymmetry

NN output differences for a four-output NN, trained to discriminate jets from \( t \to q q' b, H \to bb, Z \to q q, W \to q q' \), b

\[ E_T > 500 \text{ GeV} \]
VLQ pair search with BEST

**Left:** Kinematic distributions, such as $m_{SD}$ from training samples (boosted SM particles from high mass resonances: $X \rightarrow tt$, HH etc)

**Below:** Output probability for $H$ classification after evaluating each sample ($t$, $W$, $Z$, $H$, $b$, light jet) with BEST procedure.
**VLQ pair search with BEST**

- **Signal**: Using exactly 4 AK8 jets, count multiplicities of objects according to BEST classification:
  - For any $N_i \leq 4; N_t + N_H + N_W + N_Z + N_b + N_j = 4 \Rightarrow 126 \text{ categories}$
  - Discriminant: $H_T > 1600 \text{ GeV} \ (\text{by construction})$

- **QCD Backgrounds**: In each BEST category $N_X$, using 3 AK8 jets in data, measure tagging rates
  \[ \epsilon_X(p_T) = \frac{N_X}{N_{jets}} \]
  - Apply rates to events in inclusive 4-jet DATA sample $\notin$ any 126 signal category: e.g; [t, W, b, j]
  - Permute over all BEST labels in an event to obtain QCD shape and yield
  \[ R = \sum_{\text{events}} \left[ \sum_{\text{perm}} \left( \prod_{i=1}^{4} \epsilon_X(p_T) \right) \right] \]
VLQ pair search using CutBased

- Analysis targets $T \rightarrow bW$ decay mode, by deploying $W$- and $b$-tagging
  - At least 4 jets: 2 AK8 jets + 2 AK4 jets
  - Background: control regions from data used to measure QCD multijet background yields and shapes

![Diagram](image)

- $M_1 - M_2 = \frac{\Delta M}{\text{avg}(M_1 + M_2)}$
  - $H_T > 1200$ & $\Delta M > 0.1$

- C - Signal
- D - QCD Prediction
  - Same shape as C, normalize by A/B.
  - Corrections applied from fitting C/D ratio

Additional SF from 0b sideband region are applied to 1b and 2b signal region

CMS-PAS-B2G-18-005 (all-hadronic)
- **T search**: cut-based most sensitive to $bW$ decay, NN to other decays
- **B search**: cut-based most sensitive to $bZ$ decay, NN to $tW$
B2G-17-012: TT/BB → di-leptons

- Final states with **opposite sign di-lepton from a Z decay**, b jets, and hadronically decaying W, Z, H boson and top quarks

- Events are categorized in **mutually exclusive groups/categories** according to number of $N_b$, $N_W$, $N_Z$, $N_H$, $N_t$ counted for both the boosted or resolved jets

- **Discriminant**: $S_T \geq 1000$ GeV, where it is define as sum $p_T$ of all final state particles (see next slides)
**Backgrounds:**

- All background events are estimated from MC (shape and rate)
- **DY+jets:** Correction to shape is applied with SFs defined as data/MC events as a function of jet multiplies in a CR0b control region
- Background modeling is validated in ST control regions

<table>
<thead>
<tr>
<th>Selection</th>
<th>CR0b</th>
<th>CR1b+ST(low)</th>
<th>CR0b+ST(high)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N(bjets)</td>
<td>0</td>
<td>≥ 1</td>
<td>0</td>
</tr>
<tr>
<td>ST</td>
<td>≥ 0</td>
<td>≤ 1000 GeV</td>
<td>≥ 1000 GeV</td>
</tr>
</tbody>
</table>

![Graphs showing data, DY+jets, tt, tZ, and other backgrounds in CR1b+ST(low) and CR0b+ST(high) control regions.](image)
**B2G-17-012: TT/BB → di-leptons**

- **T search**: sensitive to final states with $T \rightarrow tZ$ decay
- **B search**: sensitive to final states with $B \rightarrow bZ$ decay

• **Search for TT or BB production**

• **Three different channel**
  
  • **single lepton**, optimized for signal events with $T \rightarrow bW$ or $T \rightarrow tH$, with a lepton from top decay and boosted W or H tagged jet

  • **same sign leptons**, most sensitive to signal events with $T \rightarrow tH$, with the Higgs decaying to WW and one W decaying leptonically

  • **trilepton**, highly sensitive to VLQ pair production with at least one $T \rightarrow tZ$, $B \rightarrow bZ$, or $B \rightarrow tW$ decay
B2G-17-011: TT/BB → leptons

- Each channel sensitive to different VLQ decay modes → strongest sensitivity to TT/BB achieved by combining the three channels:
  - **single-lepton** channel most sensitive to at least one $T \rightarrow bW$
  - the **SS dilepton** channel most sensitive to at least one $B \rightarrow tW$
  - the **trilepton** channel most sensitive to at least one $T \rightarrow tZ$

- This search excludes $T$ ($B$) quark masses below 1140-1300 GeV (910-1240) GeV, depending on the BR.

![Graph showing CMS results for TT and BB channels](image-url)
Search for TT or YY → bWbW → blvbqq

Final states with a single isolated $\mu$ or $e$, missing transverse momentum, and at least four jets with high transverse momenta.

W-tagging is used to improve sensitivity at higher masses.

Performed a constrained kinematic fit with 4 resolved jets or W-tag subjets + 2 jets.

Background is estimated from simulation with corrections from data.

- tt+jets (W+jets): Event weights are applied to MC based on generator-level $p_T (H_T)$ distribution vs reconstruction.
• **Search for:** TT or YY → bWbW → blνbqq

• Exclusion limits on the TT/YY cross section, considering $\text{BR}(T/Y \rightarrow bW) = 100\%$
  
  • Only categories with W-tagged jets are used when inferring signal or setting limits

  • T/Y quark masses are excluded below 1295 GeV
Backup
Jet Reconstruction

- CMS uses the Particle Flow algorithm to reconstruct all particles produced in a collision event.
- Charged and neutral hadrons are associated to initial quarks and gluons by a sequential recombination jet algorithm.

- Clustering algorithm used with size parameter $R$ (“jet cone size”)
- If $d_{ij} < d_{ii}$, combine constituents

\[
d_{ii} = p_{T,i}^{2p}
\]

\[
d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}
\]

- $p = -1, 0, 1$ for anti-kt (AK), Cambridge Aachen (CA), kt (KT) algorithms.
W-tagging

Pruning

Redo clustering
remove soft
large single
constituents

More pure

Less pure

S.Rappoccio, ICHEP 2016
JME-13-006,JME-16-003,SMP-12-019
Reconstruct the two B hadrons from the $b$ and $\bar{b}$ within the same fat jet

Double tagger
\[
\Delta R < 0.8
\]

Mistag efficiency

BTV-13-001, BTV-15-002
Event Pre-selection

<table>
<thead>
<tr>
<th>Variable</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z \rightarrow \ell\ell$ candidate multiplicity</td>
<td>$= 1$</td>
</tr>
<tr>
<td>$p_T(Z)$</td>
<td>$&gt; 100$ GeV</td>
</tr>
<tr>
<td>AK4 jet multiplicity</td>
<td>$\geq 3$</td>
</tr>
<tr>
<td>$H_T$</td>
<td>$&gt; 200$ GeV</td>
</tr>
<tr>
<td>$p_T$ of leading AK4 jet</td>
<td>$&gt; 100$ GeV</td>
</tr>
<tr>
<td>$p_T$ of subleading AK4 jet</td>
<td>$&gt; 50$ GeV</td>
</tr>
<tr>
<td>b-tagged AK4 jet multiplicity</td>
<td>$\geq 1$</td>
</tr>
<tr>
<td>$p_T$ of b jet</td>
<td>$&gt; 50$ GeV</td>
</tr>
<tr>
<td>$S_T$</td>
<td>$&gt; 1000$ GeV</td>
</tr>
</tbody>
</table>
### Systematic Uncertainties

<table>
<thead>
<tr>
<th>Source</th>
<th>Shape</th>
<th>Uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>TT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background yield</td>
</tr>
<tr>
<td>t(\bar{t})+jets rate</td>
<td>15</td>
<td>–</td>
</tr>
<tr>
<td>DY+jets rate</td>
<td>15</td>
<td>–</td>
</tr>
<tr>
<td>Diboson rate</td>
<td>15</td>
<td>–</td>
</tr>
<tr>
<td>Integrated luminosity</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Lepton identification</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Trigger efficiency</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PDF</td>
<td>✓</td>
<td>4.8–6.6</td>
</tr>
<tr>
<td>(\mu_f) and (\mu_r)</td>
<td>✓</td>
<td>12.9–25.8</td>
</tr>
<tr>
<td>Pileup</td>
<td>✓</td>
<td>3.5–5.0</td>
</tr>
<tr>
<td>DY+Jets correction factor</td>
<td>✓</td>
<td>4.2–11.4</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>✓</td>
<td>5.4–8.2</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>✓</td>
<td>2.0–3.8</td>
</tr>
<tr>
<td>V and H tagging</td>
<td>✓</td>
<td>1.5–2.5</td>
</tr>
<tr>
<td>t tagging</td>
<td>✓</td>
<td>0.5–3.0</td>
</tr>
<tr>
<td>misidentification of V</td>
<td>✓</td>
<td>0.6–2.3</td>
</tr>
<tr>
<td>misidentification of H</td>
<td>✓</td>
<td>0.0–0.7</td>
</tr>
<tr>
<td>misidentification of t</td>
<td>✓</td>
<td>1.0–2.3</td>
</tr>
<tr>
<td>b tagging</td>
<td>✓</td>
<td>4.1–6.2</td>
</tr>
</tbody>
</table>