Search for Emerging Jets: Experimental Perspective

Kevin Pedro (FNAL)
on behalf of the CMS Collaboration
October 23, 2018
CMS-PAS-EXO-18-001
• Dark QCD force $SU_d(N_C)$; $N_C = 3 \rightarrow 8$ dark gluons

• $N_F = 7$ flavors of fermionic dark quarks $Q_d$ (charged under $SU_d(3)$)

• Complex scalar mediator $X_d$
  (charged under $SU_d(3)$, $SU(3)$ triplet, $Q_{EM} = \pm1/3$ or $\pm2/3$)

• Dark quarks form dark mesons & baryons, decay to SM hadrons w/ non-negligible lifetime → emerging jets

• Lightest dark baryon = dark matter candidate
Varied (independent) parameters:

- Mediator mass: \( m_{X_d} \)
- Dark pion mass: \( m_{\pi_d} \)
- Dark pion decay length: \( c\tau_{\pi_d} \)

Fixed or dependent parameters:

- Mediator cross section: \( \sigma_{X_d} = 3\sigma_t \)
  (same quantum numbers, 3 colors)
- Mediator width: \( \Gamma_{X_d} = 10 \text{ GeV} \)
- Dark quark mass: \( m_{Q_d} = 2m_{\pi_d} \)
- Dark rho mass: \( m_{\rho_d} = 4m_{\pi_d} \)
- Dark QCD scale: \( \Lambda_d = m_{Q_d} \)

Model points:

- 336 total
Kinematic Variables

Varying mass ($c\tau = 25 \text{ mm}$)

$m_{X_d} = 1 \text{ TeV}$

Varying lifetime ($m_{\pi_d} = 2 \text{ GeV}$)
Kinematic Selections

**Jets**
- Anti-\(k_T\), \(R = 0.4\)
- \(|\eta| < 2.0\)
- \(n_{\text{trk}} \geq 1, \frac{p_T(\text{trk})}{p_T(j)} < 0.6\)
- \(f_{\text{charged EM}} < 0.9\)
- \(f_{\text{neutral EM}} < 0.9\)

**Tracks**
- \(p_T > 1\) GeV
- High-purity quality
- Match to jets: \(\Delta R(j,\text{trk}) < 0.4\)

**Baseline**
- \(N_{\text{jet}} \geq 4\)
- \(p_T(j_1) > 225\) GeV
- \(p_T(j_{2,3,4}) > 100\) GeV
- \(H_T = \sum p_T(j_{1,2,3,4}) > 900\) GeV
- \(N_{\text{EMJ}} \geq 2, \text{ OR}\)
- \(N_{\text{EMJ}} \geq 1 \text{ and } E_T^{\text{miss}} > 200\) GeV

**Signal topology**
- Trigger
- Large \(c\tau\)
Emerging Jets

• Emerging jets contain *multiple* displaced vertices

• Tracks from displaced vertices have large *impact parameters*

• Vertex reconstruction difficult for $c\tau > 30$ cm
  → use impact parameters to discriminate
  (more robust calculation)

Per-track quantities:

• $\Delta z_{PU} = |z_{PV} - z_{trk}|$ → reject PU

• $D_N = \sqrt{(\Delta z_{PU}/0.01 \text{ cm})^2 + (S_{IP})^2}$,
  $S_{IP} = d_0^{(2D)}/\sigma d_0^{(2D)}$ → 3D IP significance (reject prompt)

Per-jet quantities:

• $\langle d_0^{(2D)} \rangle = \text{median}[d_0^{(2D)}(trk)]$ → decay length

• $\alpha_{3D} = \frac{\sum p_T(\text{trk w/ } D_N < D_{\text{cut}})}{\sum p_T(\text{trk})}$ → fraction of $p_T$ from prompt
Emerging Jet Variables

Varying mass
\( (c\tau = 25 \text{ mm}) \)

\( m_{\chi_d} = 1 \text{ TeV} \)

Varying lifetime
\( (m_{\pi_d} = 5 \text{ GeV}) \)

---

**CMS Simulation**

- QCD light jets
- Dark pion mass 1 GeV
- Dark pion mass 2 GeV
- Dark pion mass 5 GeV
- Dark pion mass 10 GeV

**Fraction of Jets / 0.2**

\( \log(\langle d_0^{(2D)} \rangle / 1\text{ cm}) \)

---

**CMS Simulation**

- QCD light jets
- \( c\tau_{\pi_{dK}} = 1 \text{ mm} \)
- \( c\tau_{\pi_{dK}} = 5 \text{ mm} \)
- \( c\tau_{\pi_{dK}} = 25 \text{ mm} \)
- \( c\tau_{\pi_{dK}} = 60 \text{ mm} \)
- \( c\tau_{\pi_{dK}} = 100 \text{ mm} \)
- \( c\tau_{\pi_{dK}} = 300 \text{ mm} \)

**Fraction of Jets / 0.05**

\( \alpha_{3D} \)
Emerging Jets Tagging

- Combine cuts on 4 variables ($\Delta z_{PU}$, $D_N$, $\langle d_0^{(2D)} \rangle$, $\alpha_{3D}$) to tag emerging jets
- Different cut sets optimized for different models (lifetime, mass, etc.)
  - Maximize significance: $S/\sqrt{(S + B + (0.1B)^2)}$

<table>
<thead>
<tr>
<th>Criteria group</th>
<th>$\Delta z_{PU}$ (&lt;) [cm]</th>
<th>$D_N$ (&lt;)</th>
<th>$\langle d_0^{(2D)} \rangle$ (&gt;) [cm]</th>
<th>$\alpha_{3D}$ (&lt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMJ-1</td>
<td>2.5</td>
<td>4</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-2</td>
<td>4.0</td>
<td>4</td>
<td>0.10</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-3</td>
<td>4.0</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-4</td>
<td>2.5</td>
<td>4</td>
<td>0.10</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-5</td>
<td>2.5</td>
<td>20</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-6</td>
<td>2.5</td>
<td>10</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>EMJ-7</td>
<td>2.5</td>
<td>4</td>
<td>0.05</td>
<td>0.40</td>
</tr>
<tr>
<td>EMJ-8</td>
<td>4.0</td>
<td>20</td>
<td>0.10</td>
<td>0.50</td>
</tr>
</tbody>
</table>

QCD control selections
Selection Sets

• Combine emerging jet tagging criteria and basic kinematic selections

• Optimize for different models

• For interpretations, use selection set with best expected sensitivity

<table>
<thead>
<tr>
<th>Set number</th>
<th>$H_T$</th>
<th>$p_{T,1}$</th>
<th>$p_{T,2}$</th>
<th>$p_{T,3}$</th>
<th>$p_{T,4}$</th>
<th>$p_T^{\text{miss}}$</th>
<th>$n_{\text{EMJ}}(\geq)$</th>
<th>EMJ group</th>
<th>no. models</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>900</td>
<td>225</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>900</td>
<td>225</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>900</td>
<td>225</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>1</td>
<td>3</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>1100</td>
<td>275</td>
<td>250</td>
<td>150</td>
<td>150</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>250</td>
<td>150</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>1000</td>
<td>250</td>
<td>150</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>1200</td>
<td>300</td>
<td>250</td>
<td>200</td>
<td>150</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>103</td>
</tr>
<tr>
<td>8</td>
<td>900</td>
<td>225</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>SM QCD-enhanced</td>
</tr>
<tr>
<td>9</td>
<td>900</td>
<td>225</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>1</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>
Signal Efficiency

CMS Simulation Preliminary (13 TeV)

Dark pion lifetime [mm]

Dark mediator mass [GeV]

(Each bin labeled with number of optimal selection set)
Candidate Event

Tracks

Primary vertex

Secondary vertices

Pixel layer 1

Passes selection sets 1, 5

Tracks

ECAL

HCAL
QCD Background Estimation

- Tight selection removes most SM backgrounds
- Remaining events arise from QCD multijets
  - Jets from SM quarks or gluons misidentified as emerging jets
- Measure misidentification rate in photon CR
- Apply to signal-depleted QCD CRs to estimate yield in SRs

<table>
<thead>
<tr>
<th>Photon CRs</th>
<th>QCD CRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_\gamma = 1, p_T &gt; 175 \text{ GeV, }</td>
<td>\eta</td>
</tr>
<tr>
<td>( N_{\text{SMJ}} \geq 1, p_T &gt; 50 \text{ GeV, } ) CSVv2 &gt; 0.8 OR CSVv2 &lt; 0.2</td>
<td>( N_{\text{EMJ}} = 1, \text{ no } E_T^{\text{miss}} ) req.</td>
</tr>
</tbody>
</table>
Misidentification Rate

- Misid. rate higher for b quarks vs. light flavor
- Split photon CR into low and high CSVv2 subregions
  - Determine flavor composition by fitting CSVv2 distribution w/ MC templates
- Similarly, determine flavor composition of QCD CR (right), then apply misid. rate

- Can’t use single weighted-avg fake rate because b quarks come in pairs (gluon splitting)
Background Prediction & Validation

- Closure test: apply misid. rate to QCD MC in CR, compare to QCD MC in SR
- Validation: QCD-enhanced selection sets 8, 9

<table>
<thead>
<tr>
<th></th>
<th>Set 8</th>
<th>Set 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Closure</td>
<td>Pred.</td>
<td>207 ± 30</td>
</tr>
<tr>
<td>test (MC)</td>
<td>Obs.</td>
<td>231 ± 18</td>
</tr>
<tr>
<td>Data</td>
<td>Pred.</td>
<td>317 ± 35</td>
</tr>
<tr>
<td>validation</td>
<td>Obs.</td>
<td>279</td>
</tr>
</tbody>
</table>

Systematic uncertainties:
- b quark fraction: difference between CSVv2 fit and MC truth → 0.6–5%
- non-b quark composition: difference between photon and QCD CRs → 1.4–6.3% (28.3% for set 3 w/ $E_T^{\text{miss}} > 200$ GeV)
Results

- Data: 16.1 fb⁻¹, √s = 13 TeV, 2016, CMS

- Observed data agree with background predictions within uncertainties (statistical, systematic)

- Signal yields in table shown for largest m_{X_d} excluded by each selection set

- Limits do not depend strongly on m_{π_d}

- Exclude m_{X_d} between 400 and 1250 GeV for cτ_{π_d} between 5 and 225 mm

<table>
<thead>
<tr>
<th>Set number</th>
<th>Expected</th>
<th>Observed</th>
<th>Signal</th>
<th>Model parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m_{X_d} [GeV]</td>
<td>m_{π_d} [GeV]</td>
<td>cτ_{π_d} [mm]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>168 ± 15 ± 5</td>
<td>131</td>
<td>36.7 ± 4.0</td>
<td>600 5 1</td>
</tr>
<tr>
<td>2</td>
<td>31.8 ± 5.0 ± 1.4</td>
<td>47</td>
<td>(14.6 ± 2.6) × 10²</td>
<td>400 1 60</td>
</tr>
<tr>
<td>3</td>
<td>19.4 ± 7.0 ± 5.5</td>
<td>20</td>
<td>15.6 ± 1.6</td>
<td>1250 1 150</td>
</tr>
<tr>
<td>4</td>
<td>22.5 ± 2.5 ± 1.5</td>
<td>16</td>
<td>15.1 ± 2.0</td>
<td>1000 1 2</td>
</tr>
<tr>
<td>5</td>
<td>13.9 ± 1.9 ± 0.6</td>
<td>14</td>
<td>35.3 ± 4.0</td>
<td>1000 2 150</td>
</tr>
<tr>
<td>6</td>
<td>9.4 ± 2.0 ± 0.3</td>
<td>11</td>
<td>20.7 ± 2.5</td>
<td>1000 10 300</td>
</tr>
<tr>
<td>7</td>
<td>4.40 ± 0.84 ± 0.28</td>
<td>2</td>
<td>5.61 ± 0.64</td>
<td>1250 5 225</td>
</tr>
</tbody>
</table>
Experimental Outlook

- **Strong first results with limited dataset**

  - **Uncertainty reduction:**
    - More data to improve statistical uncertainty in misid. rate (photon CR)
    - Or find other control regions with more events
      - + more similar flavor composition to QCD in SR
  
  - **Develop trigger strategy for lower-mass mediators** ($H_T$ threshold too high)

  - **Improve primary and secondary vertex reconstruction** to handle long-lived, as well as track reconstruction (for very long-lived)

  - **Improve 3D impact parameter significance calculation** (understanding of resolutions)

  - **Discriminate between b quarks** (background) and emerging jets (signal)
Other Channels, Parameters

- Direct $t$-channel production of emerging jets
  - Fewer final state objects $\rightarrow$ more background?

- Resonant production of emerging jets
  - Can we reconstruct the masses of these objects?

- Vary dark force parameters to see effects on showers and sensitivity, e.g.:
  - Different $N_C$, $N_F$
  - Dark pion multiplicity enhanced for $\Lambda_d \sim m_{\pi_d}$
  - More fragmentation for $\Lambda_d > m_{\pi_d}$

- Model-independent interpretations?
Other Flavors

- Consider couplings to heavy flavor (top and bottom), not just light flavor (down – this analysis)

- Secondary vertices from SM hadrons as well as dark pions
  - Need to understand how b-tagging algorithms perform on emerging jets

- Also understand bounds from FCNC, etc. (and implications or evasions)

- Multiple flavors in dark sector: different decay lengths for dark mesons, different SM hadrons produced at different lengths w/in emerging jets
  - See arXiv:1803.08080
Other Dark Sector Models

- Prompt decays of $\pi_d$:
  - missing energy aligned with jets $\rightarrow$ semi-visible jets
- Both resonant ($Z'$) and non-resonant ($X_d$) production
  - Search for resonant production ongoing in CMS
- Future work: explore mixture of semi-visible and emerging jet techniques, for moderate $c\tau_{\pi_d}$

![Graphs showing distributions of $E_T$ and $\Delta \phi$ for QCD, semi-visible jet, and WIMP signals.](arXiv:1503.00009)
Backup
References

Data Quality

- Data analyzed: 16.1 fb\(^{-1}\), \(\sqrt{s} = 13\) TeV, 2016, CMS
- First 19.8 fb\(^{-1}\) from 2016 excluded:
  - Suffered from saturation dead time in strip tracker
  - Luminosity-dependent inefficiencies in track reconstruction
  - Unreliable modeling of \(d_0^{(2D)} > 10\)mm
- Primary vertex (PV) chosen as highest \(\sum p_T^2\), considering jets + \(E_T^{\text{miss}}\)
  - Require PV to have largest \(\sum p_T\) from all associated tracks
  - Require \(\sum p_T(\Delta z_{\text{PU}} < 0.1\)mm)/\(\sum p_T > 10\%
  - Position resolution 10–12 \(\mu\text{m}\) in x, y, z
Track Resolutions

(1 < p_T < 10 GeV, |η| < 1.4)
- p_T: 1.5%
- transverse: 25–90 μm
- longitudinal: 45–150 μm

arXiv:1405.6569
Misidentification Rates (MC)

![Graphs showing misidentification rates for QCD MC, b jet, and light jet for different jet momenta and pseudorapidity.](image)

EMJ-1
Misidentification Rates (subregions)

Selection set 8
Background Validation (Set 9)

Selection set 9

Contact information:
Kevin Pedro
26
LLP2018
QCD Background Cross-check

• Alternative method:
  unfold $N_{t,i}$ (# b jets from MC) vs. $N_{m,i}$ (# jets passing CSVv2 loose WP)

• $A_{i,j}$ combines b-tag efficiency and misidentification probability from CSVv2

• Invert $A$ matrix to predict true # b jets, then apply misidentification rate from photon CR

\[
\begin{pmatrix}
N_{m,0} \\
N_{m,1} \\
N_{m,2} \\
N_{m,3} \\
N_{m,4}
\end{pmatrix}
= \begin{pmatrix}
A_{0,0} & A_{0,1} & A_{0,2} & A_{0,3} & A_{0,4} \\
A_{1,0} & A_{1,1} & A_{1,2} & A_{1,3} & A_{1,4} \\
A_{2,0} & A_{2,1} & A_{2,2} & A_{2,3} & A_{2,4} \\
A_{3,0} & A_{3,1} & A_{3,2} & A_{3,3} & A_{3,4} \\
A_{4,0} & A_{4,1} & A_{4,2} & A_{4,3} & A_{4,4}
\end{pmatrix}
\begin{pmatrix}
N_{t,0} \\
N_{t,1} \\
N_{t,2} \\
N_{t,3} \\
N_{t,4}
\end{pmatrix}
\]
## Background Cross-check Validation

<table>
<thead>
<tr>
<th></th>
<th>Set 8</th>
<th>Set 9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Closure test (MC)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pred.</td>
<td>207 ± 30</td>
<td>52.8 ± 9.2</td>
</tr>
<tr>
<td>Pred. (2)</td>
<td>209.2 ± 1.3</td>
<td>53.1 ± 1.2</td>
</tr>
<tr>
<td>Obs.</td>
<td>231 ± 18</td>
<td>52.1 ± 6.2</td>
</tr>
<tr>
<td><strong>Data validation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pred.</td>
<td>317 ± 35</td>
<td>115 ± 28</td>
</tr>
<tr>
<td>Pred. (2)</td>
<td>312.2 ± 2.0</td>
<td>112.0 ± 1.6</td>
</tr>
<tr>
<td>Obs.</td>
<td>279</td>
<td>98</td>
</tr>
</tbody>
</table>

- Closure test: apply misid. rate to QCD MC in CR, compare to QCD MC in SR
- Pred. (2): from cross-check (statistical uncertainty only)
Systematic Uncertainty Tables

Background:

<table>
<thead>
<tr>
<th>Set number</th>
<th>Source of uncertainty (%)</th>
<th>b quark fraction</th>
<th>non-b quark composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>2.8</td>
<td>1.4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.6</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2.9</td>
<td>28.3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5.0</td>
<td>4.4</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.9</td>
<td>4.0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1.0</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Signal:

- Track modeling: smear $\Delta z_{PU}$ and $d_0^{(2D)}$ in MC to match data
- Trigger: modeling in MC vs. data
Limits for Other $m_{\pi_d}$