Opportunities for Hadronization Measurements with Heavy-Flavor-Tagged Jets at LHCb

Kara Mattioli
University of Michigan

On behalf of the LHCb Collaboration

Santa Fe Jets and Heavy Flavor Workshop
January 28 - 30, 2019
Heavy-Flavor-Tagged Jets: Ideal Systems for Hadronization Studies

- Flavor tagging provides information about 
  both the initial and final states of the hadronization process - information we have never had experimental access to in jets before!

- Unprecedented opportunities for hadronization measurements exist at LHCb, which has excellent heavy-flavor jet tagging and particle ID capabilities

In this talk:
- LHCb’s advantage for hadronization studies
- Possible measurements
- Implications for our understanding of hadronization

Kara Mattioli, University of Michigan
The Large Hadron Collider beauty (LHCb) Experiment

- Forward detector ($2 < \eta < 5$) designed to study decays of beauty and charm hadrons
LHCb Tracking

- Precision tracking achieved with silicon microstrip and straw tube detectors.
LHCb Tracking

- $\delta p/p = 0.5\%$ at low momentum to $1\%$ at 200 GeV/c
LHCb Particle Identification

- Excellent Particle ID achieved with two Ring Imaging Cherenkov (RICH) detectors

Particle ID

- RICH 1
  - $2 < \eta < 4.4$
  - 2 - 60 GeV/c

- RICH 2
  - $2.8 < \eta < 5$
  - 15 - 100 GeV/c
LHCb Particle Identification

- Particle ID performance remains excellent even in high-multiplicity events.

Particle ID

RICH 1
2 < η < 4.4
2 - 60 GeV/c

RICH 2
2.8 < η < 5
15 - 100 GeV/c

Particle ID performance remains excellent even in high-multiplicity events.

Heavy Flavor Jet Tagging at LHCb

- The Secondary Vertex tagging (SV-tagger) algorithm finds secondary vertices within a jet cone and uses two boosted decision trees (BDTs) for flavor discrimination.

- BDT(bc|udsg) discriminates between heavy and light flavor jets, while BDT(b|c) discriminates between beauty and charm jets:

  - Simulation shows clearly distinguishable BDT distributions for each jet tag.
  - Tagging efficiencies measured in data: for jets with $p_T > 20$ GeV and $2.2 < \eta < 4.2$, $\varepsilon_{b\text{-jet}}$ is $\sim 65\%$ and $\varepsilon_{c\text{-jet}}$ is $\sim 25\%$, with a light-parton jet misidentification probability of $0.3\%$.

JINST 10, P06013 (2015)
Heavy Flavor Jet Tagging at LHCb: An Example in Data

BDT distribution from data for LHCb B + jet events

Projected 1D BDT distributions from 2D fit to data

Obtain yields of flavor-tagged jets
Selected Heavy-Flavor-Tagged Jet Results from LHCb

**Z+b-jet cross section**

\[ \sigma_{Z+b}(p_T > 10 \text{ GeV}) \text{ [fb]} \]

\[ \sigma_{Z+b}(p_T > 20 \text{ GeV}) \text{ [fb]} \]

\[ \text{LHCb, } \sqrt{s} = 7 \text{ TeV} \]

- MCFM MSTW08 massive LO
- MCFM MSTW08 massless LO
- MCFM MSTW08 massless NLO
- Data_{stat}
- Data_{tot}

**Simultaneous extraction of W+b-jet and W+c-jet yields**

\[ \text{Candidates/0.1} \]

\[ \text{BDT(bcludsg)} \]

**First observation of forward Z->bb production in pp collisions at } \sqrt{s} = 8 \text{ TeV**}

\[ \text{LHCb, } \sqrt{s} = 13 \text{ TeV} \]

- data
- \( t \bar{t} \)
- Wt
- Z+jets
- Lepton Mis-ID

**Forward top pair production in the dilepton channel in pp collisions at } \sqrt{s} = 13 \text{ TeV**}

\[ \text{Candidates/0.3 GeV} \]

\[ \text{Dijet Mass [GeV]} \]

\[ \text{LHCb 8 TeV Signal Region} \]

- Data
- Z
- Total uncertainty

\[ \text{LHCb 8 TeV Control Region} \]

- Data
- Z
- Total uncertainty

\[ \text{PLB 776, 430-439 (2018)} \]

JHEP 08, 174 (2018)

JHEP 01, 064 (2015)

PRD 92, 052001 (2015)
Selected Heavy-Flavor-Tagged Jet Results from LHCb

Simultaneous extraction of W+b-jet and W+c-jet yields Z+b-jet cross section

Forward top pair production in the dilepton channel in pp collisions at $\sqrt{s} = 13$ TeV

First observation of forward $Z\rightarrow bb$ production in pp collisions at $\sqrt{s} = 8$ TeV

Still many avenues to explore with heavy-flavor-tagged jets at LHCb, including hadronization!
Heavy Flavor Jet Tagging at the LHC

LHC Run 1 ($\sqrt{s} = 7, 8$ TeV) Heavy-Flavor Jet Tagging for Jets with $p_T > 20$ GeV/c

<table>
<thead>
<tr>
<th></th>
<th>ATLAS\textsuperscript{1}</th>
<th>CMS\textsuperscript{2}</th>
<th>ALICE\textsuperscript{3}</th>
<th>LHCb\textsuperscript{4}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_{b\text{-jet}}$</td>
<td>70%</td>
<td>70%</td>
<td>None</td>
<td>65%</td>
</tr>
<tr>
<td>$\varepsilon_{c\text{-jet}}$</td>
<td>25-40%</td>
<td>None</td>
<td>None</td>
<td>25%</td>
</tr>
<tr>
<td>Mis-ID as light parton</td>
<td>1%</td>
<td>1.5%</td>
<td>N/A</td>
<td>0.3%</td>
</tr>
<tr>
<td>Hadron PID</td>
<td>None</td>
<td>None</td>
<td>$</td>
<td>\eta</td>
</tr>
</tbody>
</table>

LHCb is the only LHC experiment that has both b- and c-jet tagging capabilities \textit{and} hadron PID over a large range in both pseudorapidity and momentum.

\textsuperscript{1} JINST 11 P04008 (2016)
\textsuperscript{2} JINST 8 P04013 (2013)
\textsuperscript{3} JINST 3 S08002 (2008)
\textsuperscript{4} JINST 10 P06013 (2015)
LHCb’s Advantage for Hadronization Studies

- Excellent tracking and particle ID detectors allow for the identification of final-state hadrons in a fully reconstructed jet
- Forward acceptance gives access to a mix of quark and gluon jets
- Established techniques exist for tagging heavy-flavor jets with high purity
- Emphasis in studying the production and decays of heavy flavor hadrons makes it an ideal experiment to study how beauty and charm quarks hadronize
- Hadronization studies in heavy-flavor-tagged jets at LHCb would expand on recent work done at LHCb to study J/Ψ production in jets:

A heavy-flavor-tagged jet hadronization study would be complementary to the LHCb Z + jet hadronization study currently in progress, which preferentially selects light-quark jets

PRL 118, 192001 (2017)
Hadronization Measurements Accessible at LHCb

- Longitudinal momentum distribution of hadrons in a heavy-flavor-tagged jet

\[ z = \frac{p_{jet} \cdot p_h}{|p_{jet}|^2} \]

- Transverse momentum distribution of hadrons in a heavy-flavor-tagged jet

\[ j_T = \frac{|p_h \times p_{jet}|}{|p_{jet}|} \]

- Radial profile of hadrons in a heavy-flavor-tagged jet

\[ r = \sqrt{(\phi_h - \phi_{jet})^2 + (y_h - y_{jet})^2} \]

- Number of heavy-flavor and light-flavor hadrons in the jet and their flavor composition

- Number of baryons and mesons in the jet

- Comparison of these observables between beauty and charm jets

\[ \text{Suggestions for more observables are very welcome!} \]
• LHCb has collected 9.23 fb$^{-1}$ of data since 2010:
  
  • Subsets of Run 1 and Run 2 data indicate a lower bound of several million for the heavy-flavor-tagged jet yield
    - Yields likely to increase with improved tagging techniques

• Range of jet $p_T$ for multidifferential studies:

LHCb Cumulative Integrated Recorded Luminosity in pp, 2010-2018

- 2018 (6.5 TeV): 2.19 /fb
- 2017 (6.5+2.51 TeV): 1.71 /fb + 0.10 /fb
- 2016 (6.5 TeV): 1.67 /fb
- 2015 (6.5 TeV): 0.33 /fb
- 2012 (4.0 TeV): 2.06 /fb
- 2011 (2.5 TeV): 1.11 /fb
- 2010 (3.5 TeV): 0.04 /fb

**LHCb Z+jet cross section**

\[ \text{JHEP 01, 033 (2014)} \]

\[ p_T^{\text{jet}} > 10 \text{ GeV} \]

\[ \sqrt{s} = 7 \text{ TeV Data} \]

LHCb

Data (stat.)

Data (tot.)

POWHEG + PYTHIA:

- MSTW08, $O(\alpha_s)$
- MSTW08, $O(\alpha_s^2)$
- CTEQ10, $O(\alpha_s^2)$
- NNPDF 2.3, $O(\alpha_s^2)$

**LHCb Z+jet cross section**

\[ \frac{1}{\sigma} \frac{d\sigma}{dp_T^{\text{jet}}} \]

\[ \text{[1/GeV]} \]

\[ 10^{-2} \]

\[ 10^{-1} \]

\[ 10^{-3} \]

\[ 10^{-4} \]

\[ 20 \]

\[ 40 \]

\[ 80 \]

\[ 100 \]

\[ 120 \]

\[ 140 \]

\[ p_T^{\text{jet}} \text{ [GeV]} \]
Implications of Hadronization Studies with Heavy-Flavor-Tagged Jets

- Identifying final-state hadrons in fully-reconstructed, flavor-tagged jets will offer new insights into mechanisms of color neutralization in hadronization.

- All of the proposed studies in this talk can be done at LHCb with existing data within the next few years.

- We hope that the capability to measure identified final-state hadrons in a fully-reconstructed, flavor-tagged jet will encourage the theoretical community to calculate distributions for multiple hadrons in a jet.
  
  - Wealth of 1D and multi-dimensional projections possible to facilitate theoretical comparisons.
  
  - Suggestions for additional observables within b- and c-tagged jets are very welcome!

- Hadronization will be a major component of the Electron-Ion Collider physics program. Learning more about hadronization now will help the community refine its goals for hadronization measurements at a future Electron-Ion Collider.
Implications of Hadronization Studies with Heavy-Flavor-Tagged Jets

- Identifying final-state hadrons in fully-reconstructed, flavor-tagged jets will offer new insights into mechanisms of color neutralization in hadronization.

- All of the proposed studies in this talk can be done at LHCb with existing data within the next few years.

- We hope that the capability to measure identified final-state hadrons in a fully-reconstructed, flavor-tagged jet will encourage the theoretical community to calculate distributions for multiple hadrons in a jet.
  - Wealth of 1D and multi-dimensional projections possible to facilitate theoretical comparisons.
  - **Suggestions for additional observables within b- and c-tagged jets are very welcome!**

- Hadronization will be a major component of the Electron-Ion Collider physics program. Learning more about hadronization now will help the community refine its goals for hadronization measurements at a future Electron-Ion Collider.
Backup
LHCb Calorimetry

- Calorimeter system includes a Scintillating Pad Detector, a Preshower Detector, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL)

**Calorimetry**

\[
\frac{\sigma(E)}{E} = \frac{69\%}{\sqrt{E}} \pm 9\%
\]

**HCAL:**

\[
\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \pm 1\%
\]

**ECAL:**

**Scintillating Pad & Preshower Detectors**

**Shashlik scintillator/lead**

JINST 3, S08005 (2008)
LHCb Muon System

- Muon ID achieved with five muon stations consisting of Gas Electron Multiplier (GEM) and 1380 Multi-Wire Proportional Chambers (MWPC) interleaved with iron absorbers.
Light-parton Jet Mis-ID with the SV-Tagger Algorithm

- Light-parton mis-ID probability studied in simulation and measured in data

- Mis-identification of a heavy-flavor jet as a light-parton jet is predominantly due to prompt tracks mis-reconstructed as displaced tracks to form a fake secondary vertex
Heavy Flavor Jet Tagging with the Topological Trigger

- Software trigger designed to identify decays of beauty hadrons
- Has been used as a b-jet tagger by requiring the tagged b-hadron to be inside a jet
- Jet tagging performance studied in both data and simulation:

JINST 10, P06013 (2015)

Kara Mattioli, University of Michigan
Calculation of SV-Tagger Efficiency

- Efficiency is calculated as the number of SV-tagged beauty and charm jets divided by the total number of beauty and charm jets.

- Number of SV-tagged beauty and charm jets is determined by fitting the 2D BDT distribution from data (see slide 8).

- Total number of beauty and charm jets is determined by fitting $\chi^2_{IP}$: the difference in $\chi^2$ of a primary interaction vertex reconstructed with and without the highest-pT track in the jet.

- The highest-pT track in light-parton jets usually originates from the primary vertex, while the highest-pT track in heavy-flavor jets usually originates from the beauty or charm hadron decay.

- Fit templates obtained from simulation and calibrated by comparison to data from $W +$ jet events (primarily light-parton jet dominated) to determine detector resolution effects.
The Secondary Vertex (SV)-Tagger Algorithm

- Requires that either a high-$p_T$ muon or a beauty or charm hadron passes trigger requirement:
  - Muon candidate must have $p_T > 10$ GeV
  - Beauty or charm hadron candidate must have $p_T > 1.7$ GeV, and the difference in $\chi^2$ of a primary vertex reconstructed with and without the considered track greater than 16

- For events with a candidate passing the trigger requirement, jets are clustered with the anti-$k_T$ algorithm

- Tracks for secondary vertex reconstruction within the jet are required to have $p_T > 0.5$ GeV and a difference in $\chi^2$ of a primary vertex reconstructed with and without the considered track greater than 16
  - No hadron PID is used - all particles are assigned a pion mass
  - Tracks are not required to be in the jet cone

- All possible two-track secondary vertices are reconstructed subject to the following cuts:
  - Distance of closest approach between tracks is less than 2mm
  - $\chi^2$ of vertex fit < 10
  - Two-body mass is in the range $0.4$ GeV < $M < M(B)$, where $M(B)$ is the nominal $B^0$ mass

JINST 10, P06013 (2015)
The Secondary Vertex (SV)-Tagger Algorithm

- All two-track secondary vertices with $\Delta R < 0.5$ relative to the jet axis are merged until none of the secondary vertices share tracks.

- The weighted average of the two-body secondary vertices is calculated with the inverse of the vertex $\chi^2$ values as the weights. The weighted average is taken to be the position of the secondary vertex in the jet.

- The merged and weighted secondary vertices are required to satisfy the following:
  - $p_T > 2$ GeV
  - Significant spatial separation from primary vertex
  - Contain at most one track with $\Delta R > 0.5$ relative to jet axis
  - Pass quality cuts to suppress strange-hadron decays

- Information about secondary vertices is passed to two BDTs for flavor discrimination: BDT $(bc|udsg)$ and BDT $(b|c)$
Input to SV-Tagger BDTs

- Secondary vertex mass
- Secondary vertex corrected mass:  \[ M_{cor} = \sqrt{M^2 + p^2 \sin^2 \theta + p \sin \theta} \]
  - \( M \) is the invariant mass of the particles that form the secondary vertex, \( p \) is the momentum of the particles that form the secondary vertex, \( \theta \) is the angle between the momentum and direction of flight of the secondary vertex
- Transverse flight distance of the two-track secondary vertex closest to the primary vertex
- Fraction of the jet \( p_T \) carried by the secondary vertex
- \( \Delta R \) between the secondary vertex flight direction and the jet
- Number of tracks in the secondary vertex
- Number of secondary vertex tracks with \( \Delta R < 0.5 \) relative to the jet axis
- Net charge of the tracks that form the secondary vertex
- Flight distance \( \chi^2 \)
- Sum of all secondary vertex track \( (\chi^2_P)^2 \): the difference in \( \chi^2 \) of a primary vertex reconstructed with and without a considered track
Additional Heavy-Flavor-Tagged Jet Results from LHCb

First observation of top quark production in the forward region

PRL 115, 112001 (2015)

First measurement of the charge asymmetry in $b\bar{b}$ production

$A_{C}^{bb} \equiv \frac{N(\Delta y > 0) - N(\Delta y < 0)}{N(\Delta y > 0) + N(\Delta y < 0)}$

PRL 113, 082003 (2014)
Advances in Theoretical Approaches to Hadronization

Early 1970s
- Fragmentation functions introduced to describe the probability for a parton to fragment into a specific hadron as a function of $z$
  
  - PRD 4, 3388 (1971)

Late 1970s
- Dihadron fragmentation functions introduced to describe the probability of a single parton fragmenting into two specific hadrons
  

Early 1980s
- Transverse-momentum-dependent fragmentation functions introduced, adding a dependence on the transverse momentum of the produced hadron in addition to $z$
  

1990s
- Dihadron interference fragmentation function introduced
  

2010s
- Theoretical predictions for distributions of hadron-in-jet observables introduced
  
  - PRD 81, 074009 (2010), JHEP 1404, 147 (2014)