Constructing & Using a Quark/Gluon Tagger

How well can we do at the 7 and 8 TeV LHC?

Jason Gallicchio

UC Davis

26 June 2012
1 Big Motivation: Reject Gluey LHC Backgrounds

2 The Tagger: Observables and Performance

3 Verification: Finding Pure Samples of Quark and Gluon Jets

4 ATLAS: Results and Herwig++

5 Theory: Meaningful to What Order?

“Quark and Gluon Tagging at the LHC”  arXiv:1106.3076
“Pure Samples of Quark and Gluon Jets at LHC” arXiv:1104.1175
(with Matt Schwartz at Harvard)

Interactive Plots: http://jets.physics.harvard.edu/qvg/
So chance that all 4 jets $\gtrsim 50$ GeV are quark $\approx (21\%)^4 \approx 1/500$
Most new physics gives quark rather than gluon jets:
Gluon Tagging Motivation

Most *new physics* gives *quark* rather than *gluon* jets:

- 8-jet Gluino event: $pp \rightarrow \tilde{g}\tilde{g}$ and each $\tilde{g}$ decays to 4 quarks:
Most *new physics* gives quark rather than gluon jets:

- 8-jet Gluino event: $pp \rightarrow \tilde{g}\tilde{g}$ and each $\tilde{g}$ decays to 4 quarks:

  ![Diagram of 8-jet Gluino event]

- Higgs $H^+ \rightarrow c\bar{s}$ (for charged Higgs mass between $\tau$ and $t$ mass)
Most *new physics* gives *quark* rather than *gluon* jets:

- 8-jet Gluino event: $pp \rightarrow \tilde{g}\tilde{g}$ and each $\tilde{g}$ decays to 4 *quarks*:

- Higgs $H^+ \rightarrow c\bar{s}$ (for charged Higgs mass between $\tau$ and $t$ mass)
- Measure $Z'$ coupling to hadrons (or find a leptophobic $Z'/W'$)
Gluon Tagging Motivation

Most new physics gives quark rather than gluon jets:

- 8-jet Gluino event: $pp \rightarrow \tilde{g}\tilde{g}$ and each $\tilde{g}$ decays to 4 quarks:

- Higgs $H^+ \rightarrow c\bar{s}$ (for charged Higgs mass between $\tau$ and $t$ mass)
- Measure $Z'$ coupling to hadrons (or find a leptophobic $Z'/W'$)
- For $X \rightarrow$ jets, measure quark/gluon branching ratios.
Motivation

Interesting *standard model physics* also tends to be quark-heavy
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow ud$ or $c\bar{s}$
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \rightarrow b\bar{b} + 0, 2, \text{ or } 4 \text{ light quarks}$)
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \to u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \to b\bar{b} + 0, 2, \text{ or } 4 \text{ light quarks}$)
- Vector Boson Scattering/Fusion (forward ‘tag’ jets are quarks)
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \rightarrow b\bar{b} + 0, 2, \text{or} 4$ light quarks)
- Vector Boson Scattering/Fusion (forward ‘tag’ jets are quarks)

Q vs G is especially useful without $W, Z, \gamma, \ell^\pm, B$-Tags, or $E_T$ (R-Parity violating SUSY easily gives 6 quark jets)
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \rightarrow b\bar{b} \pm 0, 2, \text{ or } 4 \text{ light quarks}$)
- Vector Boson Scattering/Fusion (forward ‘tag’ jets are quarks)

Q vs G is especially useful without $W, Z, \gamma, \ell^\pm, B$-Tags, or $\not{E}_T$ (R-Parity violating SUSY easily gives 6 quark jets)

Some signals consist of *gluon* jets, like Coloron models or Buried Higgs: $h \rightarrow 2a \rightarrow 4g$ ($a$ is CP odd scalar).
Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow u\bar{d}$ or $c\bar{s}$
- Tops ($tt \rightarrow b\bar{b} + 0, 2, \text{or} 4$ light quarks)
- Vector Boson Scattering/Fusion (forward ‘tag’ jets are quarks)

$Q$ vs $G$ is especially useful without $W$, $Z$, $\gamma$, $\ell^\pm$, $B$-Tags, or $E_T$ (R-Parity violating SUSY easily gives 6 quark jets)

Some signals consist of *gluon* jets, like Coloron models or Buried Higgs: $h \rightarrow 2a \rightarrow 4g$ ($a$ is CP odd scalar).

Only model-independent way to measure new particle’s color-charge.
Motivation

Interesting *standard model physics* also tends to be quark-heavy

- $W$’s decaying hadronically (no b’s!): $W^+ \rightarrow u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \rightarrow b\bar{b} + 0, 2, \text{or } 4\text{ light quarks}$)
- Vector Boson Scattering/Fusion (forward ‘tag’ jets are quarks)

Q vs G is especially useful without $W, Z, \gamma, \ell^\pm, B$-Tags, or $E_T$ (R-Parity violating SUSY easily gives 6 quark jets)

Some signals consist of gluon jets, like Coloron models or Buried Higgs: $h \rightarrow 2a \rightarrow 4g$ ($a$ is CP odd scalar).

Only model-independent way to measure new particle’s color-charge.

Must combine Quark/Gluon-Tagging with B-Tagging and $\tau$-Tagging.
- Jet energy scale correction depends on flavor. Can’t calibrate on a quark-heavy sample and blindly apply to a gluon-heavy one.
Jet energy scale correction depends on flavor. Can’t calibrate on a quark-heavy sample and blindly apply to a gluon-heavy one.

Monte Carlo validation and tuning
The Quark/Gluon Tagger
Visual Differences

Same dijet event showered 3 million times. Accumulate $p_T(\eta, \phi)$:

- **Quark Jet**
- **Gluon Jet**

(Same total amount of $p_T$, which is hidden by logarithmic color bands.)
Jet Shape

\[ \Psi(r) \]

Integrated Jet Shape 100 GeV R=0.7

- **Q**
- **G**

\[ r \]

Jason Gallicchio (UC Davis)

Constructing & Using a Quark/Gluon Tagger

26 June 2012 9 / 67
Jet Shape plots are averaged over all events of a particular type.
Integrated Jet Shape out to $r = 0.1$ for 100 GeV

- Distribution is *not* narrow gaussian around average
- Correlations *between* different $r$’s might also be useful
Gluon has a greater effective color charge (squared) than quark:

Gluon adjoint’s $C_A$ vs Quark fundamental’s $C_F$

\[
\frac{C_A}{C_F} = \frac{9}{4}
\]
Gluon has a greater effective color charge (squared) than quark:

Gluon adjoint’s $C_A$ vs Quark fundamental’s $C_F$

$$\frac{C_A}{C_F} = \frac{9}{4}$$

Average Jet Mass in the small angle limit:

$$\langle M^2 \rangle = C \frac{\alpha_s}{\pi} p_T^2 R^2$$

Distribution of Jet Mass....
- Normalizing by $p_T$ (200 GeV in this sample) generalizes better.
- All distributions normalized to equal area.
Evaluating the Observable: Sliding Cut

mass/Pt

![Graph showing the mass/Pt distribution with a sliding cut. The graph has a blue and red distribution for quark (Q) and gluon (G) interactions, respectively. The sliding cut is indicated by a vertical dotted line at mass/Pt = 0.15.](image-url)
ROC Curve for mass/Pt

Quark Signal Efficiency

Gluon Background Rejection
ROC Curve for mass/Pt

Quark Signal Efficiency

Gluon Background Rejection

50/50
Other Jet Sizes and $p_T$s

**Sliding Cut**

**ROC Curve for mass/Pt**

**mass/Pt @ 80% Signal Efficiency**

- Quark Signal Efficiency
- Gluon Background Rejection

**Jet Size**
- 1600 GeV
- 800 GeV
- 400 GeV
- 200 GeV
- 100 GeV
- 50 GeV
Radial Moment – a measure of the “girth” of the jet

Weight $p_T$ deposits by distance from jet center

Radial Moment, or Girth:

$$ g = \frac{1}{p_T} \sum_{i \in \text{jet}} p_T^i |r_i| $$

‘Jet Broadening’ is a similar LEP observable involving $E$ and $\Delta \theta$. 
Gluon adjoint’s $C_A$ vs Quark fundamental’s $C_F$

\[
\frac{C_A}{C_F} = \frac{9}{4}
\]
Gluon adjoint’s $C_A$ vs Quark fundamental’s $C_F$

\[ \frac{C_A}{C_F} = \frac{9}{4} \]

Multiplicity of any particle in a gluon jet should be $C_A/C_F = 9/4$ times greater (confirmed at LEP).

\[ \frac{\langle N_g \rangle}{\langle N_q \rangle} = \frac{C_A}{C_F} \]
Gluon adjoint’s $C_A$ vs Quark fundamental’s $C_F$

$$\frac{C_A}{C_F} = \frac{9}{4}$$

Multiplicity of any particle in a gluon jet should be $C_A/C_F = 9/4$ times greater (confirmed at LEP).

$$\frac{\langle N_g \rangle}{\langle N_q \rangle} = \frac{C_A}{C_F} \quad \frac{\sigma^2_g}{\sigma^2_q} = \frac{C_A}{C_F}.$$ 

(Calculated to $N^3$LO by Capella, et al. hep-ph/9910226)

For this talk, PYTHIA8 will serve as a repository of decades of theoretical and experimental knowledge. (v8.165, default tune.)
No detector simulation, but require charged particles $p_T > 1\text{ GeV}$:

Higher $p_T$ means more tracks and more ‘time’ to establish $C_A/C_F$. 
Types of Variables

The menu, including varying jet size

- Distinguishable particles/tracks/subjets
  - multiplicity, \( \langle p_T \rangle \), \( \sigma_{p_T} \), \( \langle k_T \rangle \),
  - charge-weighted \( p_T \) sum

- Moments
  - mass, girth, jet broadening
  - angularities
  - optimal kernel
  - N-subjettiness
  - 2D: pull, planar flow

- Subjet properties
  - Multiplicity for different algorithms and \( R_{sub} \)
  - First subjet’s \( p_T \), 2nd’s \( p_T \), etc.
  - Ratios of subjet \( p_T \)’s.
  - \( k_T \) splitting scale

- 2-Point Correlators (energy, \( p_T \), possibly times \( r^\# \), etc.)
Combining Variables: Girth and Charged Count

Quark

Gluon

Likelihood: $q/(q + g)$
Best Variables in Each Category for 200 GeV Jets

Gluon Rejection

- best group of 5
- charged mult & girth
- charged mult * girth
- charged mult R=0.5
- subjet mult R_{sub}=0.1
- girth R=0.5
- optimal kernel
- 1st subjet R=0.5
- avg k_T of R_{sub}=0.1
- mass/Pt R=0.3
- decluster k_T R_{sub}=0.1
- jet shape Ψ(0.1)
- |pull| R=0.3
- planar flow R=0.3

Quark Jet Acceptance vs. Gluon Rejection
Can reject 80% of **gluons** while keeping 80% **quarks**.
Can reject 95% of **gluons** while keeping 50% **quarks**. (20x rejection)
Can reject 80% of gluons while keeping 80% quarks. 
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Improve the significance of your quark-heavy signal:

\[ \sigma = \frac{S}{\sqrt{B}} \rightarrow \frac{S\epsilon_s}{\sqrt{B}\epsilon_b} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}} \]
Can reject 80% of gluons while keeping 80% quarks.
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Improve the significance of your quark-heavy signal:

\[ \sigma = \frac{S}{\sqrt{B}} \rightarrow \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}} \]

For \( p_T > 20 \text{ GeV} \), QCD jet background is 83% gluons.
Can reject 80% of gluons while keeping 80% quarks.
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Improve the significance of your quark-heavy signal:

\[
\sigma = \frac{S}{\sqrt{B}} \quad \rightarrow \quad \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}}
\]

For \( p_T > 20 \text{ GeV}, \) QCD jet background is 83% gluons.

Can find operating point where \( \epsilon_s/\sqrt{\epsilon_b} \approx 1.4 \)
Can reject 80% of gluons while keeping 80% quarks.
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Improve the significance of your quark-heavy signal:

$$\sigma = \frac{S}{\sqrt{B}} \rightarrow \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}}$$

For $p_T > 20$ GeV, QCD jet background is 83% gluons.

Can find operating point where $\epsilon_s/\sqrt{\epsilon_b} \approx 1.4$

For signal of 4 quarks $\geq 20$ GeV, significance improvement is $1.4^4 = 3.8$
Can reject 80% of gluons while keeping 80% quarks.
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Improve the significance of your quark-heavy signal:

\[ \sigma = \frac{S}{\sqrt{B}} \quad \rightarrow \quad \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}} \]

For \( p_T > 20 \text{ GeV} \), QCD jet background is 83% gluons.

Can find operating point where \( \epsilon_s/\sqrt{\epsilon_b} \approx 1.4 \)

For signal of 4 quarks \( \geq 20 \text{ GeV} \), significance improvement is \( 1.4^4 = 3.8 \)
For the 6 quark RPV example, significance improvement is 7.5!
Part 3

Finding Pure Samples of Quark and Gluon Jets
Starting Samples

Chance EACH Jet is Quark

$\gamma + 1j$

$Z/W + 1j$

$\gamma + 2j$

$Z/W + 2j$

$2j$

$3j$

$bb + 1j$

$b + 1j$

$p_T$ Cut on All Jets (GeV)
200 GeV Quark Purity

Cross Section in pb

Quark Purity (zoom)
When the softer jet is **quark**, the photon is often radiated off of *it*, rather than the harder jet.
200 GeV Quark Purity

Cross Section in pb

Quark Purity

40% 50% 60% 70% 80% 90% 100%

- 2jet
- 3jet hardest
- γ+2jet harder
- γ+2jet softer
- Z+2jet harder
- Z+2jet softer
- γ+1jet harder
- γ+1jet softer
Cross Section in pb vs Gluon Purity

Best Samples for Gluon Purity

- 50 jj
- 100 jj
- 200 jj
- 400 jj
- 800 jj
- 1600 jj
- 50 jjj
- 100 jjj
- 200 jjj
- 400 jjj
- 800 jjj
- 50 bjj
- 100 bjj
- 200 bjj
- 400 bjj
- 800 bjj
Summary of Finding Samples

- **Quark** samples at 99% purity for $\gamma + 2\text{jet}$
- **Gluon** samples at 90%-95% purity for 3jets
Summary of Finding Samples

- **Quark** samples at 99% purity for $\gamma + 2\text{jet}$
- **Gluon** samples at 90%-95% purity for 3jets

- **Gluon** samples at 95%-99% purity for $b + 2\text{jets}$ with strong B-Tagging and B-Anti-Tagging
ATLAS Results and Herwig++
- Isolated anti-$k_T$ jets with $R = 0.4$
- Only track-based variables to avoid pileup effects
- Charged track $p_T > 1$ GeV
- In MC, jets were matched to highest energy parton within cone
Herwig++ 2.5.2 (darker) as compared to Pythia 8.165 (lighter) for 50 GeV quarks and gluons.

**mass/p_T**

**N-subjettiness $\beta = 1/4$**

**Charged Track Count**

**width (radial-moment)**
Plot the average values, but for different $p_T$ jets. (Note legend)

Charged Track Count differs for gluons

Width (radial-moment) differs at low $p_T$

(from M.Laura Gonzalez Silva’s talk at BOOST2012)
Goal: to measure the quark/gluon shapes from data, dijet (DJ) and photon+jet (γJ) events.

Ideally, solve for q/g (for each bin i) from:

\[
\begin{align*}
    h_i(DJ) &= P_Q(DJ)q_i + P_G(DJ)g_i \\
    h_i(\gamma J) &= P_Q(\gamma J)q_i + P_G(\gamma J)g_i
\end{align*}
\]

- \( P_Q \) = quark percentage, from MC
- \( h \) = histogram value, from data
- \( q/g \) = pure q/g jet distributions (solving for these)

But need to account for \( b \) and \( c \) fractions (taken from MC):

\[
\begin{align*}
    h_i(DJ) &= P_Q(DJ)q_i + P_G(DJ)g_i + P_B(DJ)b_i + P_C(DJ)c_i \\
    h_i(\gamma J) &= P_Q(\gamma J)q_i + P_G(\gamma J)g_i + P_B(\gamma J)b_i + P_C(\gamma J)c_i
\end{align*}
\]
Di-jet data should match linear combination of pure quark + gluon.

from “Jet energy measurement with the ATLAS…” arXiv:1112.6426
The width of the band represents the maximum variation among the Pythia and and the Herwig++ samples.
Charged Track Count differs for gluons

Width (radial-moment) agrees reasonably with Pythia8
Charged Track Count
better in Herwig++
now quarks are off

Width (radial-moment)
worse in Herwig++
Preliminary result shows data not looking as separable.

|η| <0.8, Jet pT ~150 GeV:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Efficiency</th>
<th>Rejection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pythia MC11</td>
<td>50%</td>
<td>8x</td>
</tr>
<tr>
<td>Data 2011</td>
<td>50%</td>
<td>4x</td>
</tr>
</tbody>
</table>

- Purified samples validate these findings.
- Need different variables?
- Need more isolated jets?
QCD Jet Flavor Theory
Example of 2 Quark Jets
Standard Parton Shower
Any flavor tagging is only useful to the extent that there is a correspondence between hard partons and jets.
Any flavor tagging is only useful to the extent that there is a correspondence between hard partons and jets.

This is just as true when using b-tagged jets in kinematic reconstruction (i.e. tops and Higgs)
Any flavor tagging is only useful to the extent that there is a correspondence between hard partons and jets.

This is just as true when using b-tagged jets in kinematic reconstruction (i.e. tops and Higgs)

This is a standard starting point for most searches, and is affected by jet algorithm and event topology.
Any flavor tagging is only useful to the extent that there is a correspondence between hard partons and jets.

This is just as true when using b-tagged jets in kinematic reconstruction (i.e. tops and Higgs)

This is a standard starting point for most searches, and is affected by jet algorithm and event topology.

This is a search-focused rather than precision-QCD-focused view.
Angle of Attack

Any flavor tagging is only useful to the extent that there is a correspondence between hard partons and jets.

This is just as true when using b-tagged jets in kinematic reconstruction (i.e. tops and Higgs)

This is a standard starting point for most searches, and is affected by jet algorithm and event topology.

This is a search-focused rather than precision-QCD-focused view.

Claim: Nothing can go wrong that wouldn’t also destroy the event’s meaning/usefulness/interpretation, and those things are unlikely.
Extra Emissions?

**Gluon emission:**

- If it ends up in same jet (soft), this is exactly what determines the properties of the jet.
- If it creates its own jet (hard), it should have been modeled as a hard emission: ‘matching’
Extra Emissions?

**Gluon** emission:

- If it ends up in the same jet (soft), this is exactly what determines the properties of the jet.
- If it creates its own jet (hard), it should have been modeled as a hard emission: ‘matching’

Gluon splitting to **quarks** (light or b):

- If they end up in the same jet (soft), it’s still a gluon jet.
- If they create their own jets (hard), these are quark (or b) jets.
Flavor is well-defined to all orders in QCD perturbation theory. Ambiguity only when further radiation (hard QCD and soft showering) doesn’t match jet grouping.
Flavor is well-defined to to \textit{all} orders in QCD perturbation theory. Ambiguity only when further radiation (hard QCD and soft showering) doesn’t match jet grouping.

These are described by power corrections that affect \textit{any} collinear and IR safe jet algorithm’s parton correspondence.

They involve $\Lambda_{QCD}/E$, jet size $R$, jet’s mass-to-energy ratio $m/E$, etc.
Flavor is well-defined to to all orders in QCD perturbation theory. Ambiguity only when further radiation (hard QCD and soft showering) doesn’t match jet grouping.

These are described by power corrections that affect any collinear and IR safe jet algorithm’s parton correspondence.

They involve $\Lambda_{QCD}/E$, jet size $R$, jet’s mass-to-energy ratio $m/E$, etc.

So flavor is no more dangerous theoretically than any time jets are used as a proxy for hard partons in kinematic reconstruction.

(All of this is separate from measurement resolution.)
Loops? Same final state. No interference between flavors. Only rates.
Problem Case

Loops? Same final state. No interference between flavors. Only rates. Problem case for unlikely splitting

Given only final state, flavor-blind anti-$k_T$ leaves these ambiguous. Each contribution is not individually gauge invariant and they interfere.
Problem Case

Loops? Same final state. No interference between flavors. Only rates. Problem case for unlikely splitting

Given only final state, flavor-blind \( k_T \) leaves these ambiguous. Each contribution is not individually gauge invariant and they interfere.

For identical final state (same momenta), first amplitude is much larger.

- \( g \to gg \) and \( q \to qg \) (soft \( g \)): both collinear and soft divergences
- \( g \to q\bar{q} \) and \( q \to qg \) (soft \( q \)): only collinear divergence
Another Problem Case

Hard gluon fails to make its own jet

\begin{center}
\begin{tikzpicture}
  \node (q) at (0,0) {$q$};
  \node (qbar) at (1,0) {$\bar{q}$};
  \node (gluon) at (0.5,0) {$\cdots$};
  \node (qprime) at (2,1) {$q$};
  \node (qbarprime) at (2,-1) {$\bar{q}$};
  \draw[->] (q) -- (gluon);
  \draw[->] (gluon) -- (qbar);
  \draw[->] (q) -- (qprime);
  \draw[->] (gluon) -- (qbarprime);
\end{tikzpicture}
\end{center}
Another Problem Case

Hard gluon fails to make its own jet

If the original 2 hard quarks were instead gluons, it wouldn’t make sense to call these ‘quark jets’ either.
Finding a B meson inside a jet makes it a B jet.

This doesn’t really say anything about how well the jet ‘matches the $b$ quark’, i.e. how well two such jets would reconstruct $H \rightarrow b\bar{b}$. 
Finding a B meson inside a jet makes it a B jet.

This doesn’t really say anything about how well the jet ‘matches the $b$ quark’, i.e. how well two such jets would reconstruct $H \rightarrow b\bar{b}$.

“B-Tagging Efficiency” is defined relative to number of tagable jets: ones with a B-hadron with: $\Delta R < 0.4$, $p_T > 1 \text{ GeV}$, and $d_T > 10 \mu m$.

When they operate at 60%, that does not mean, for example, that $60\%^2$ of $t\bar{t}$ events will have 2 B-tags.
Finding a B meson inside a jet makes it a B jet.

This doesn’t really say anything about how well the jet ‘matches the $b$ quark’, i.e. how well two such jets would reconstruct $H \rightarrow b\bar{b}$.

“B-Tagging Efficiency” is defined relative to number of *tagable* jets: ones with a B-hadron with: $\Delta R < 0.4$, $p_T > 1$ GeV, and $d_T > 10\, \mu$m.

When they operate at 60%, that does not mean, for example, that 60%$^2$ of $t\bar{t}$ events will have 2 B-tags.

Ambiguity with $g \rightarrow b\bar{b}$, whether B hadrons end up in the same jet or not. Same fundamental QCD issues we have, but the massive $b$ quark makes problem cases less likely.
What flavor is my Pythia jet!?

“What’s the best way to find the true flavor of a random Pythia jet?”

- Running anti-$k_T$ on the hadrons and assigning flavor based on net baryon number ($N_q - N_{\bar{q}}$) is neither IRC safe nor particularly useful.
What flavor is my Pythia jet!?  

“What’s the best way to find the true flavor of a random Pythia jet?”

- Running anti-$k_T$ on the hadrons and assigning flavor based on net baryon number ($N_q - N_{\bar{q}}$) is neither IRC safe nor particularly useful.
- Matching to ‘hard’ event? (Ignores additional hard radiation.)
- Hardest parton anywhere in event record within jet radius?
- Trace back history of parton branchings? (Pythia randomly assigns parent of soft branching)
“What’s the best way to find the *true* flavor of a random Pythia jet?”

- Running anti-$k_T$ on the hadrons and assigning flavor based on net baryon number ($N_q - N_{\bar{q}}$) is neither IRC safe nor particularly useful.
- Matching to ‘hard’ event? (Ignores additional hard radiation.)
- Hardest parton anywhere in event record within jet radius?
- Trace back history of parton branchings? (Pythia randomly assigns parent of soft branching)

Best way to ‘truth tag’? Maybe:

1. Run experiment’s anti-$k_T$ on hadrons.
2. Run flavor-$k_T$ algorithm on ‘final’ pre-hadronization partons.
3. Assign flavor only if jets overlap sufficiently.
“What’s the best way to find the true flavor of a random Pythia jet?”

- Running anti-$k_T$ on the hadrons and assigning flavor based on net baryon number ($N_q - N_{ar{q}}$) is neither IRC safe nor particularly useful.
- Matching to ‘hard’ event? (Ignores additional hard radiation.)
- Hardest parton anywhere in event record within jet radius?
- Trace back history of parton branchings? (Pythia randomly assigns parent of soft branching)

Best way to ‘truth tag’? Maybe:

1. Run experiment’s anti-$k_T$ on hadrons.
2. Run flavor-$k_T$ algorithm on ‘final’ pre-hadronization partons.
3. Assign flavor only if jets overlap sufficiently.

Whatever is most useful to separate real signals from real backgrounds.
To the extent that things like “fraction of $\gamma + jet$ events we want to call quark-like” is meaningful, measure the width/girth and charged track count distributions for many samples. (ATLAS’s templates)
To the extent that things like “fraction of $\gamma + jet$ events we want to call quark-like” is meaningful, measure the width/girth and charged track count distributions for many samples. (ATLAS’s templates)

Cutting on these observables can make your own sample of jets more like jets that come with $\gamma$’s rather than like QCD multijets.
To the extent that things like “fraction of $\gamma + jet$ events we want to call quark-like” is meaningful, measure the width/girth and charged track count distributions for many samples. (ATLAS’s templates)

Cutting on these observables can make your own sample of jets more like jets that come with $\gamma$’s rather than like QCD multijets.

If a linear combination quark $+$ gluon jet width distributions matches the distribution observed in a sample and it’s the same linear combination predicted by some QCD calculation, it’ll be hard to argue that it’s completely meaningless to talk about jet flavor.
To the extent that things like “fraction of $\gamma + jet$ events we want to call quark-like” is meaningful, measure the width/girth and charged track count distributions for many samples. (ATLAS’s templates)

Cutting on these observables can make your own sample of jets more like jets that come with $\gamma$’s rather than like QCD multijets.

If a linear combination quark + gluon jet width distributions matches the distribution observed in a sample and it’s the same linear combination predicted by some QCD calculation, it’ll be hard to argue that it’s completely meaningless to talk about jet flavor.

Thanks!
Using Flavor Taggers
Cutting gives some signal acceptance and some background acceptance.
Comparison to B-Tagging

Quark vs Gluon with 2 observables

CMS B-Tagging vs charm

CMS B-Tagging vs g, u, d, s

Signal Acceptance

Background Acceptance
A cut on tagger’s score gives

- signal efficiency $\epsilon_s$ (you pick)
- background efficiency $\epsilon_b$ (ROC dictates)
A cut on tagger’s score gives

- signal efficiency $\epsilon_s$ (you pick)
- background efficiency $\epsilon_b$ (ROC dictates)

If you start with $S$ signal events and $B$ background events,

$$\frac{S}{B} \rightarrow \frac{S\epsilon_s}{B\epsilon_b} = \frac{S}{B} \frac{\epsilon_s}{\epsilon_b}$$

Call $\frac{\epsilon_s}{\epsilon_b}$ the “S/B Improvement”
Comparison to B-Tagging

![Comparison to B-Tagging](image-url)

- S/B Improvement vs Signal Acceptance
- Comparison between Quark vs Gluon and B-Tagging vs charm, B-Tagging vs u,d,s
Improvement in statistical significance scales differently

\[ \sigma = \frac{S}{\sqrt{B}} \rightarrow \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma \frac{\epsilon_s}{\sqrt{\epsilon_b}} \]

Call \( \frac{\epsilon_s}{\sqrt{\epsilon_b}} \) the “Significance Improvement”
Comparison to B-Tagging

![Graph showing comparison to B-Tagging]

- B-Tagging vs charm
- B-Tagging vs g, u, d, s
- Quark vs Gluon

Significance Improvement vs Signal Acceptance
But Backgrounds Contain b’s and light quarks!

<table>
<thead>
<tr>
<th></th>
<th>2 Jets</th>
<th>3 Jets</th>
<th>4 Jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>1.9%</td>
<td>1.9%</td>
<td>2.1%</td>
</tr>
<tr>
<td>udsc</td>
<td>15.2%</td>
<td>15.2%</td>
<td>14.3%</td>
</tr>
<tr>
<td>gluons</td>
<td>82.9%</td>
<td>82.9%</td>
<td>83.5%</td>
</tr>
</tbody>
</table>

All Jet $p_T > 20$ GeV
Background Contains 2% ‘Signal’ flavor (B-case)
Background Contains 2% ‘Signal’ flavor (B-case)

![Graph showing CMS B-Tagging vs charm, Quark vs Gluon, and Signal Acceptance.](image)
Background Contains 15% ‘Signal’ flavor (Q-case)

![Graph showing CMS B-Tagging vs g, u, d, s, charm, and Quark vs Gluon with Signal Acceptance and Significance Improvement axes.](image-url)
Operating Points that Maximize Quark Significance

Quark Significance Improvement for Different Contamination Levels

Significance Improvement

Signal Acceptance

0% 10% 20% 30% 40% 50%
Operating Points that Maximize Quark Significance

Quark Significance Improvement vs Contamination Level

15% Quarks for Multijets $\geq 20$ GeV
For signal of 4 quarks $\geq 20 \text{ GeV}$, significance improvement is $1.37^4 = 3.5$
Different Quark Jet Fractions

Chance EACH Jet is Quark

$p_T$ Cut on All Jets (GeV)