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

Outline

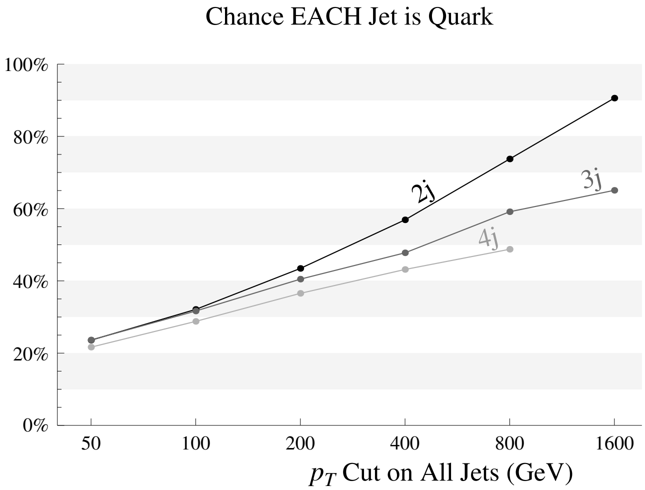
1. **Big Motivation:** Reject Gluey LHC Backgrounds
2. **The Tagger:** Observables and Performance
3. **Verification:** Finding Pure Samples of Quark and Gluon Jets
4. **ATLAS:** Results and Herwig++
5. **Theory:** Meaningful to What Order?

"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/`

There's a Lot of Glue to Get Stuck In (7 TeV LHC)



So chance that all 4 jets $\gtrsim 50\,\text{GeV}$ are quark $\approx (21\%)^4 \approx 1/500$

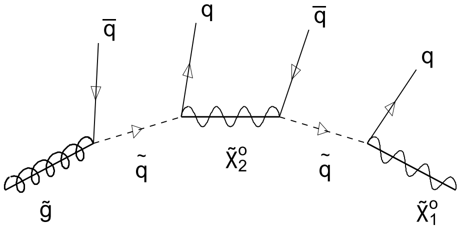
Gluon Tagging Motivation

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

Gluon Tagging Motivation

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

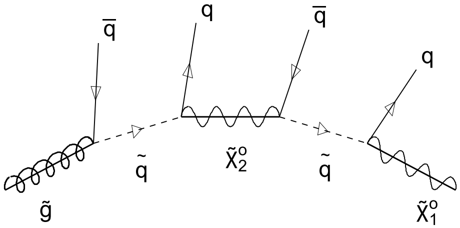
- 8-jet Gluino event: $pp \to \tilde{g}\tilde{g}$ and each $\tilde{g}$ decays to 4 quarks:



Gluon Tagging Motivation

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

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

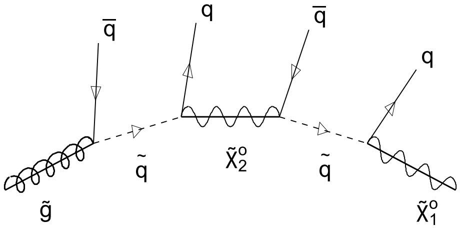


- Higgs $H^+ \to c\bar{s}$ (for charged Higgs mass between τ and t mass)

Gluon Tagging Motivation

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

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

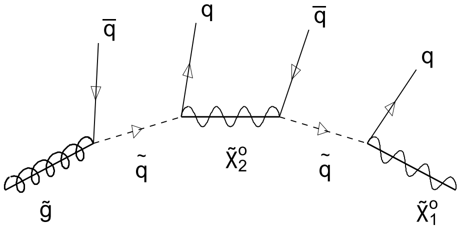


- Higgs $H^+ \to c\bar{s}$ (for charged Higgs mass between τ and t mass)
- Measure Z' coupling to hadrons (or find a leptophobic Z'/W')

Gluon Tagging Motivation

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

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



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

Motivation

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

Motivation

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

- W's decaying hadronically (no b's!): $W^+ \to u\bar{d}$ or $c\bar{s}$

Motivation

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 ($t\bar{t} \to b\bar{b} \quad + \quad$ 0, 2, or 4 light quarks)

Motivation

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 ($t\bar{t} \to b\bar{b} \quad + \quad$ 0, 2, or 4 light quarks)
- Vector Boson Scattering/Fusion (forward 'tag' jets are quarks)

Motivation

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 ($t\bar{t} \to b\bar{b} \quad + \quad 0$, 2, or 4 light quarks)
- Vector Boson Scattering/Fusion (forward 'tag' jets are quarks)

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

Motivation

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 ($t\bar{t} \to b\bar{b} \quad + \quad$ 0, 2, or 4 light quarks)
- Vector Boson Scattering/Fusion (forward 'tag' jets are quarks)

Q vs G is especially useful without W, Z, γ, $\ell^{\pm}$, 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: $h \to 2a \to 4g$ (a is CP odd scalar).

Motivation

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 ($t\bar{t} \to b\bar{b} \quad + \quad 0$, 2, or 4 light quarks)
- Vector Boson Scattering/Fusion (forward 'tag' jets are quarks)

Q vs G is especially useful without W, Z, γ, $\ell^{\pm}$, 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: $h \to 2a \to 4g$ (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!): $W^+ \to u\bar{d}$ or $c\bar{s}$
- Tops ($t\bar{t} \to b\bar{b} \quad + \quad 0, 2,$ or 4 light quarks)
- Vector Boson Scattering/Fusion (forward 'tag' jets are quarks)

Q vs G is especially useful without W, Z, γ, $\ell^{\pm}$, 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: $h \to 2a \to 4g$ (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 τ-Tagging.

Housekeeping Motivation

- Jet energy scale correction depends on flavor. Can't calibrate on a quark-heavy sample and blindly apply to a gluon-heavy one.

Housekeeping Motivation

- Jet energy scale correction depends on flavor. Can't calibrate on a quark-heavy sample and blindly apply to a gluon-heavy one.

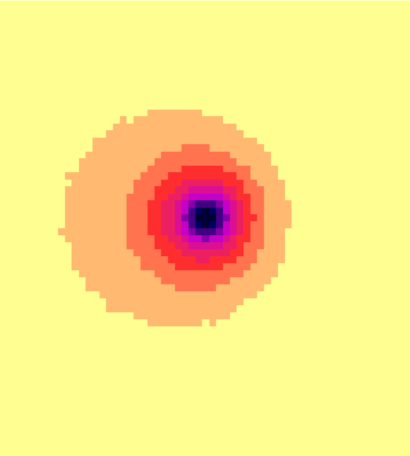
- Monte Carlo validation and tuning

The Quark/Gluon Tagger

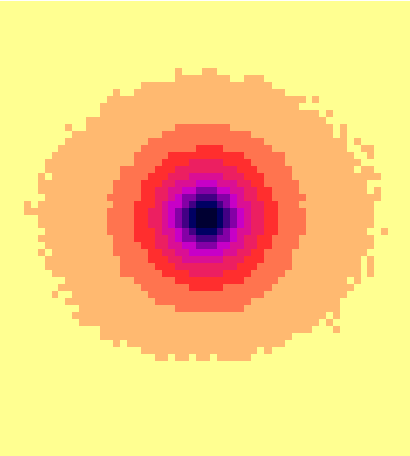
Visual Differences

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

Quark Jet

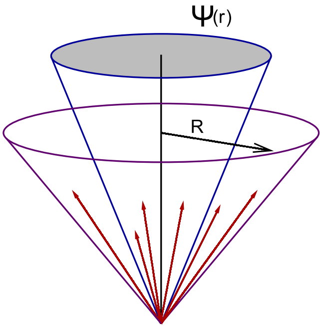


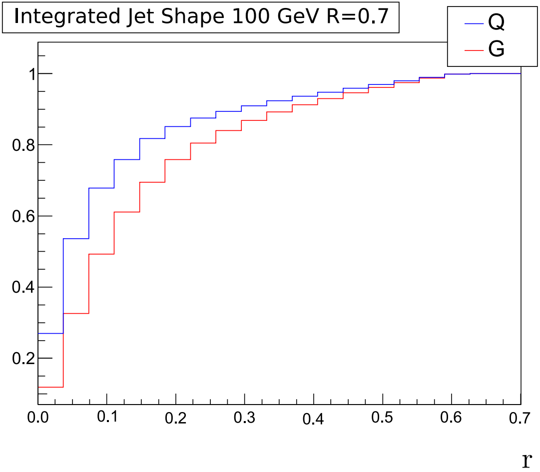
Gluon Jet

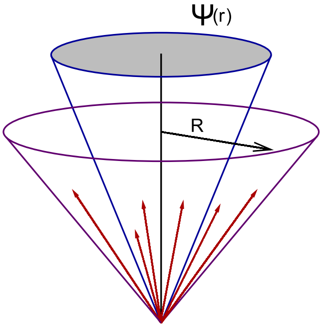


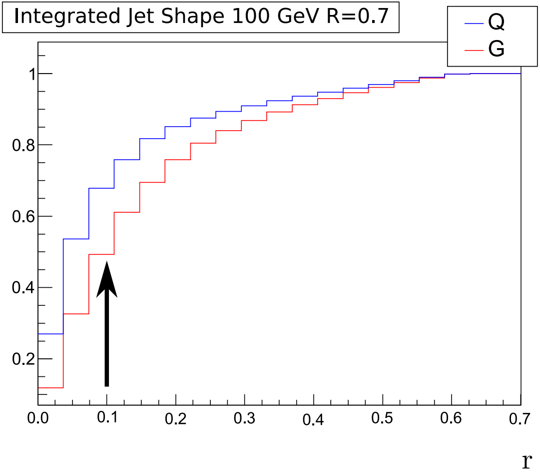
(Same total amount of p_T, which is hidden by logarithmic color bands.)

Jet Shape


Ψ(r)
R

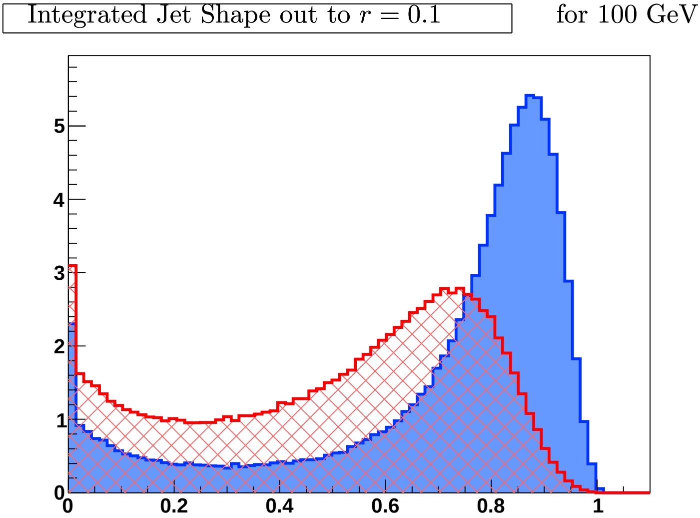

Integrated Jet Shape 100 GeV R=0.7
Q
G
1
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
r





Jet Shape plots are averaged over all events of a particular type.

Jet Shape *Distribution* vs *Average*



- 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

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

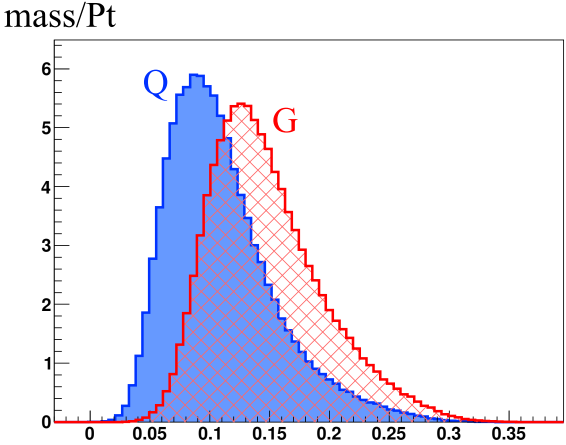
Average Jet Mass in the small angle limit:

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

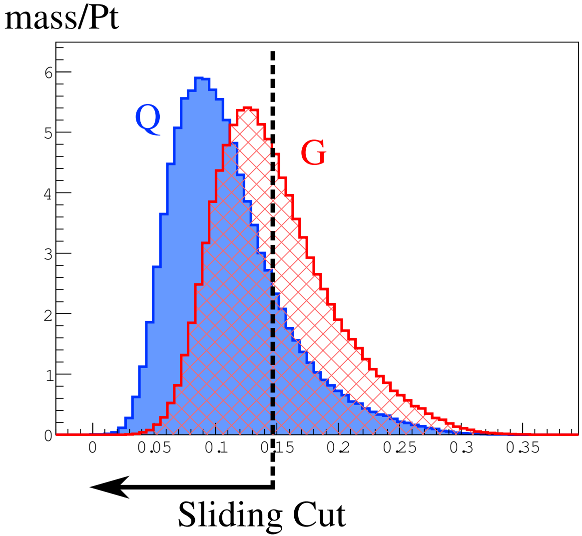
Distribution of Jet Mass....

Jet Mass as an Example Observable

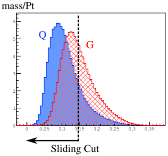
- Normalizing by p_T (200 GeV in this sample) generalizes better.
- All distributions normalized to equal area.

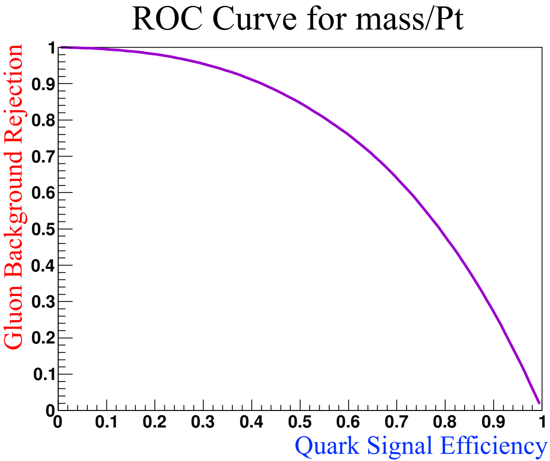


Evaluating the Observable: Sliding Cut

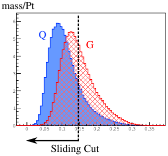


ROC Curve

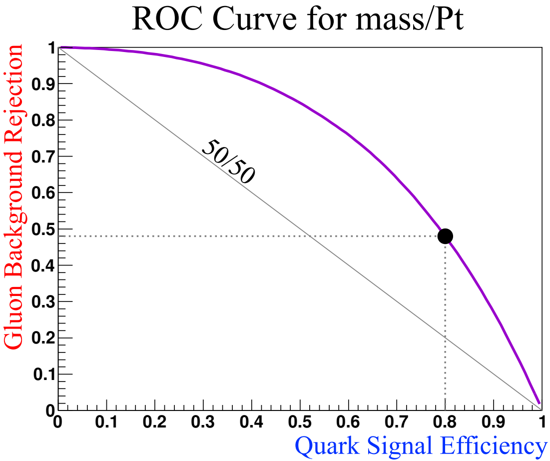

mass/Pt
Q
G
Sliding Cut


ROC Curve for mass/Pt
Gluon Background Rejection
Quark Signal Efficiency
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

ROC Curve

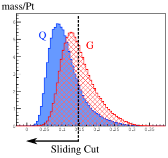


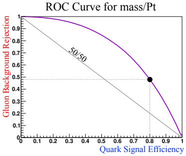
mass/Pt
Q
G
Sliding Cut

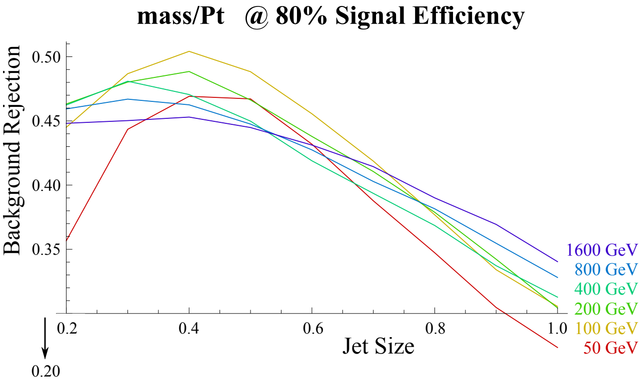


ROC Curve for mass/Pt
Gluon Background Rejection
Quark Signal Efficiency
50/50
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Other Jet Sizes and p_Ts



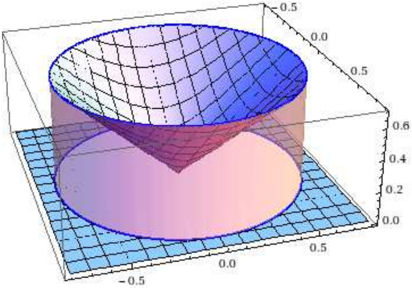


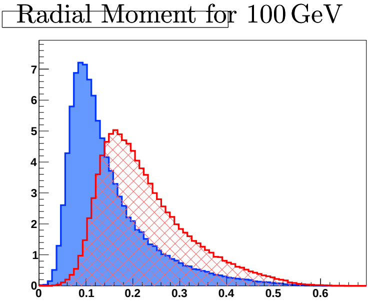


Radial Moment – a measure of the "girth" of the jet

Weight p_T deposits by distance from jet center

Radial Moment, or Girth : $$g = \frac{1}{p_T^{jet}} \sum_{i \in \text{jet}} p_T^i |r_i|$$





'Jet Broadening' is a similar LEP observable involving E and $\Delta\theta$.

Brief Theory II — Particle Count

Gluon adjoint's C_A vs Quark fundamental's C_F

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

Brief Theory II — Particle Count

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).

$$\frac{\langle N_g \rangle}{\langle N_q \rangle} = \frac{C_A}{C_F} \qquad\qquad \frac{\sigma_g^2}{\sigma_q^2} = \frac{C_A}{C_F}.$$

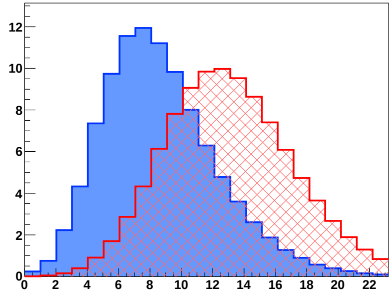
(Calculated to N^3LO by Capella, et al. hep-ph/9910226)

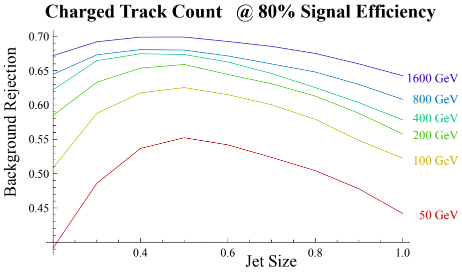
For this talk, PYTHIA8 will serve as a repository of decades of theoretical and experimental knowledge. (v8.165, default tune.)

Charged Particles Count

No detector simulation, but require charged particles $p_T > 1\,\mathrm{GeV}$:

Charged Particle Count 100 GeV





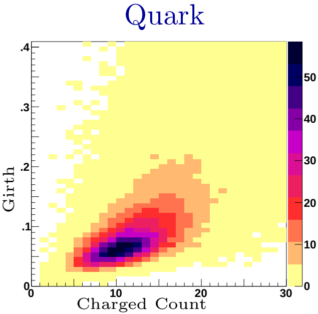
Higher p_T means more tracks and more 'time' to establish C_A/C_F.

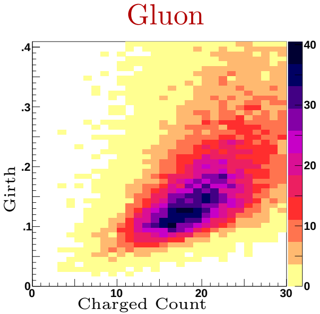
Types of Variables

The menu, including varying jet size

- Distinguishable particles/tracks/subjets
  - multiplicity, $\langle p_T \rangle$, σ_{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.)

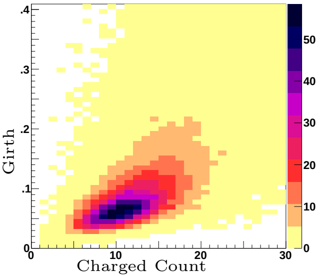
Combining Variables: Girth and Charged Count



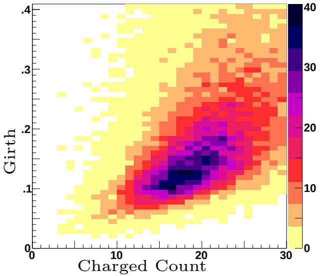


Combining Variables: Girth and Charged Count

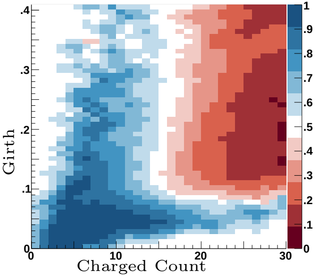
Quark



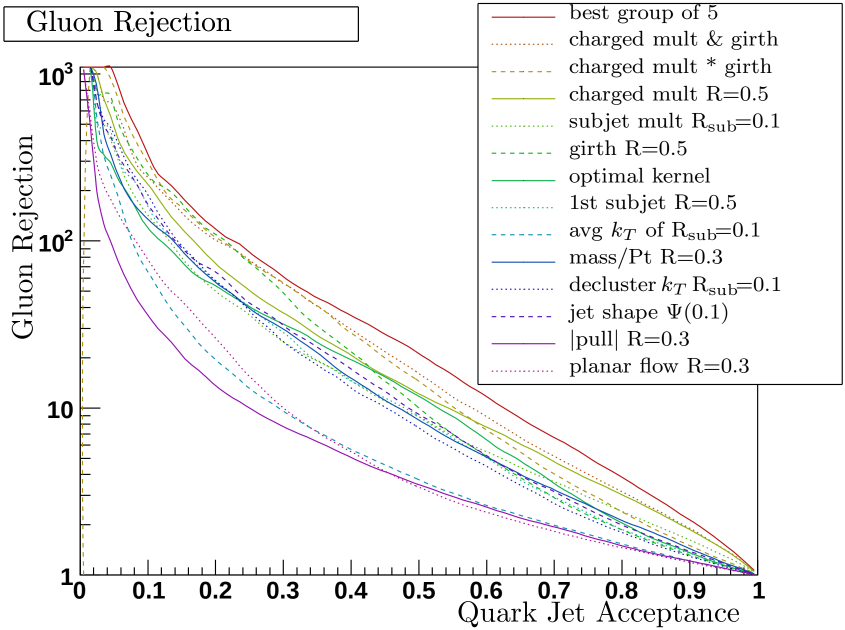
Gluon



Likelihood: $q/(q+g)$



Best Variables in Each Category for 200 GeV Jets



Summary and Use of the Tagger

Can reject 80% of gluons while keeping 80% quarks.
Can reject 95% of gluons while keeping 50% quarks. (20x rejection)

Summary and Use of the Tagger

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:

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

Summary and Use of the Tagger

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 $p_T > 20\,\mathrm{GeV}$, QCD jet background is 83% gluons.

Summary and Use of the Tagger

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 $p_T > 20\,\mathrm{GeV}$, QCD jet background is 83% gluons.

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

Summary and Use of the Tagger

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:

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

For $p_T > 20\,\mathrm{GeV}$, QCD jet background is 83% gluons.

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

For signal of 4 quarks $\geq 20\,\mathrm{GeV}$, significance improvement is $1.4^4 = 3.8$

Summary and Use of the Tagger

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:

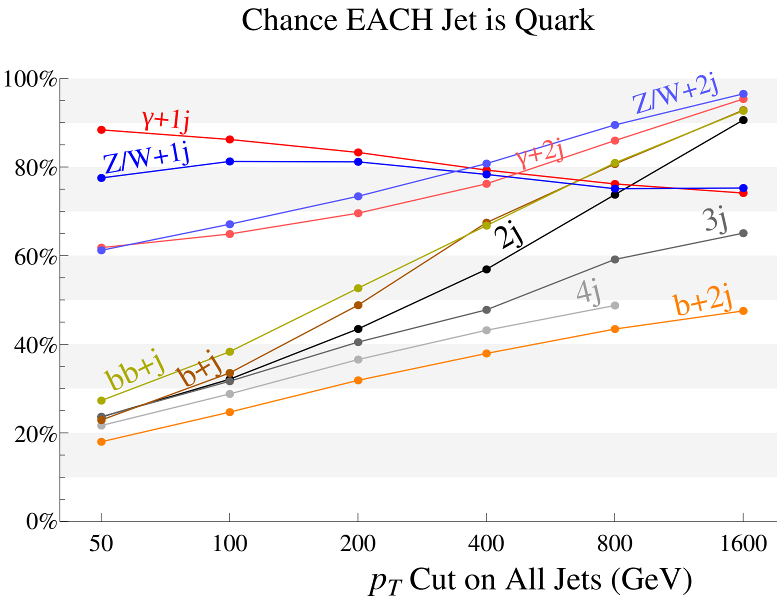
$$\sigma = \frac{S}{\sqrt{B}} \quad \rightarrow \quad \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma\frac{\epsilon_s}{\sqrt{\epsilon_b}}$$

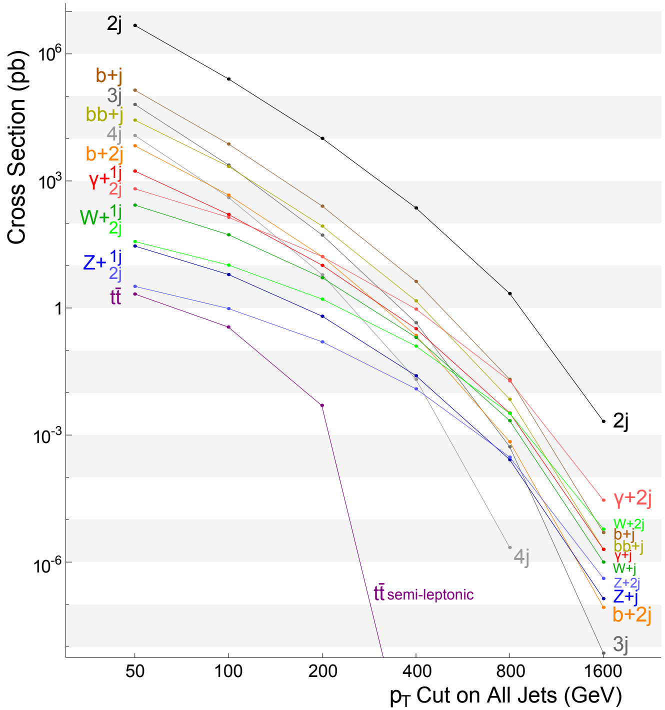
For $p_T > 20\,\mathrm{GeV}$, QCD jet background is 83% gluons.

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

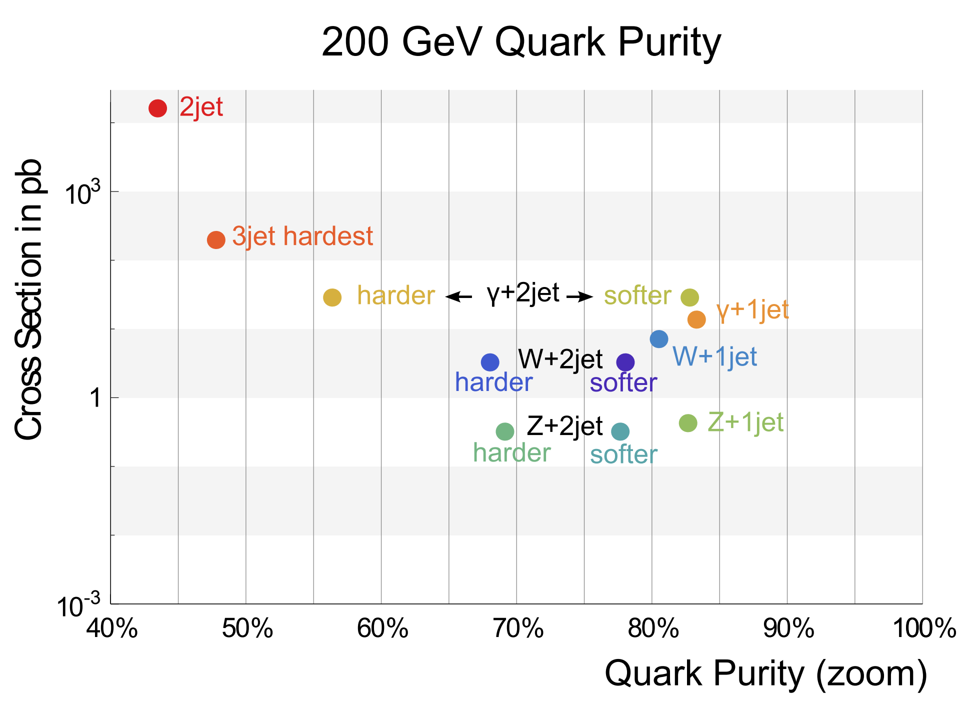
For signal of 4 quarks $\geq 20\,\mathrm{GeV}$, significance improvement is $1.4^4 = 3.8$
For the 6 quark RPV example, significance improvement is 7.5!

Finding Pure Samples of Quark and Gluon Jets

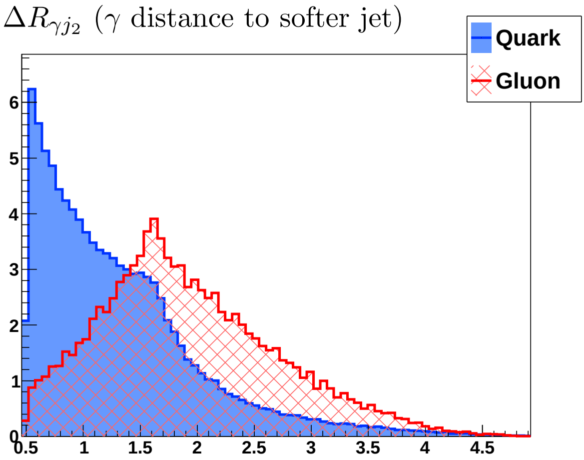

Chance EACH Jet is Quark
100%
80%
60%
40%
20%
0%
50
100
200
400
800
1600
p_T Cut on All Jets (GeV)
γ+1j
Z/W+1j
Z/W+2j
γ+2j
2j
3j
4j
b+2j
bb+j
b+j



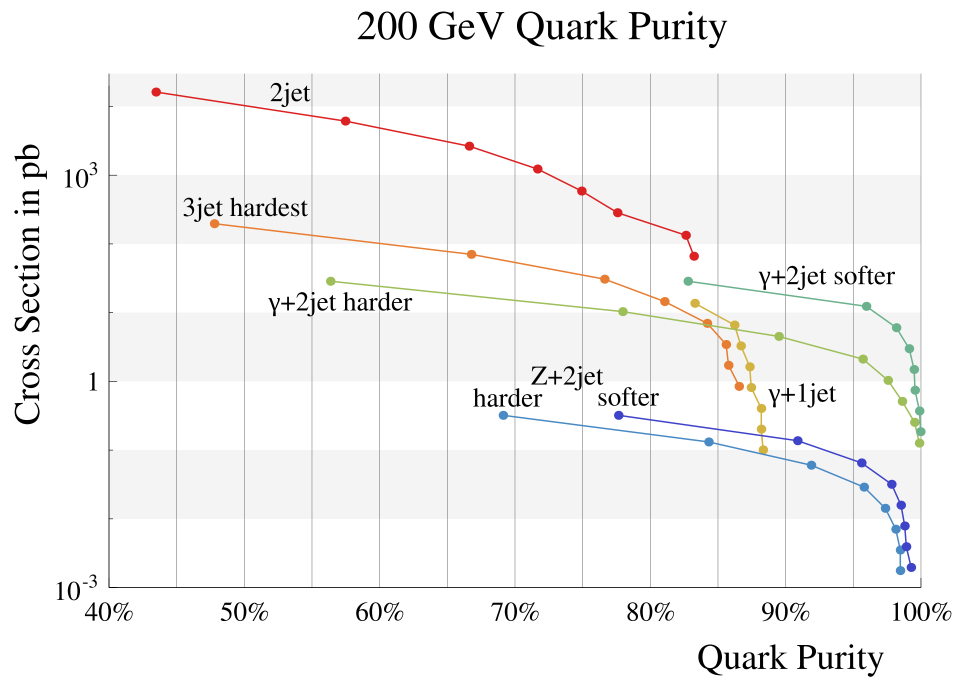
Cross Section (pb)
10^6
10^3
1
10^{-3}
10^{-6}
50
100
200
400
800
1600
p_T Cut on All Jets (GeV)
2j
b+j
3j
bb+j
4j
b+2j
γ+1j
γ+2j
W+1j
W+2j
Z+1j
Z+2j
$t\bar{t}$
$t\bar{t}$ semi-leptonic
γ+2j
W+2j
b+j
bb+j
γ+j
W+j
Z+2j
Z+j
b+2j
3j
4j
2j


200 GeV Quark Purity
Cross Section in pb
10^3
1
10^{-3}
40%
50%
60%
70%
80%
90%
100%
Quark Purity (zoom)
2jet
3jet hardest
harder
γ+2jet
softer
γ+1jet
W+2jet
harder
softer
W+1jet
Z+2jet
harder
softer
Z+1jet

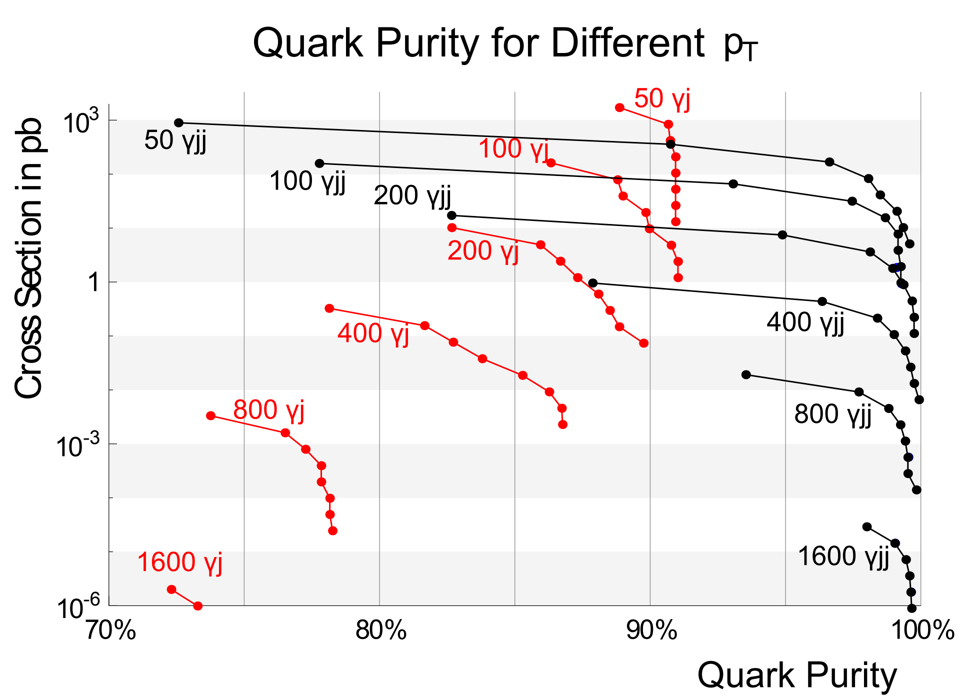
Quark Purification in γ+2jet



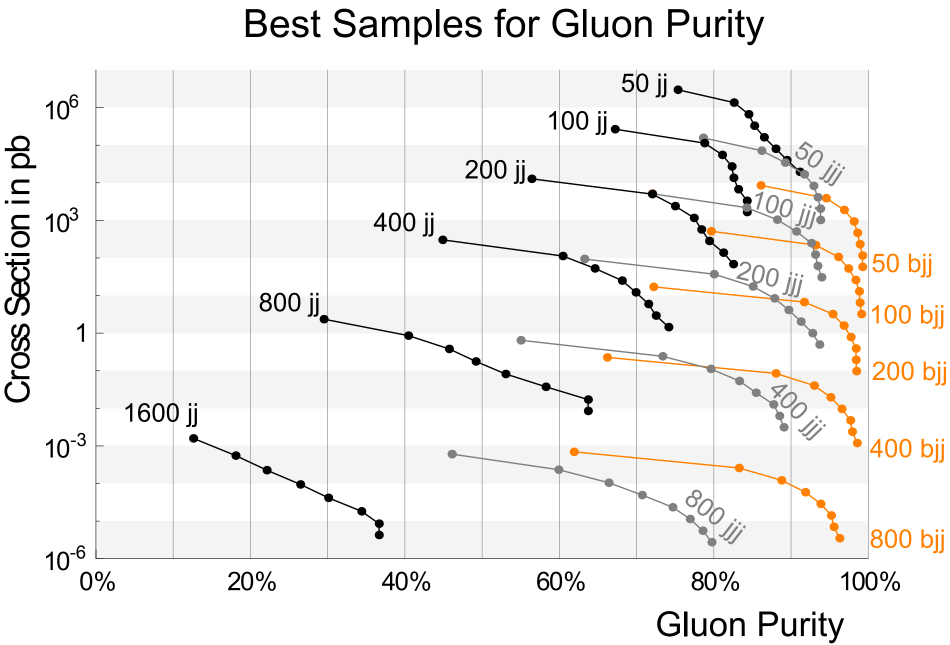
When the softer jet is quark, the photon is often radiated off of *it*, rather than the harder jet.



200 GeV Quark Purity
Cross Section in pb
Quark Purity
10^3
1
10^-3
40%
50%
60%
70%
80%
90%
100%
2jet
3jet hardest
γ+2jet softer
γ+2jet harder
Z+2jet
harder
softer
γ+1jet



Quark Purity for Different p_T
Cross Section in pb
Quark Purity
10^3
1
10^{-3}
10^{-6}
70%
80%
90%
100%
50 γj
100 γj
200 γj
400 γj
800 γj
1600 γj
50 γjj
100 γjj
200 γjj
400 γjj
800 γjj
1600 γjj



Best Samples for Gluon Purity
Cross Section in pb
Gluon Purity
10^6
10^3
1
10^{-3}
10^{-6}
0%
20%
40%
60%
80%
100%
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 γ+2jet
- Gluon samples at 90%-95% purity for 3jets

Summary of Finding Samples

- Quark samples at 99% purity for γ+2jet
- Gluon samples at 90%-95% purity for 3jets

- Gluon samples at 95%-99% purity for b+2jets with strong B-Tagging and B-Anti-Tagging

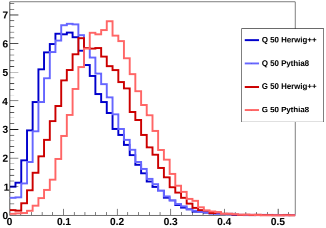
ATLAS Results and Herwig++

ATLAS Measurements

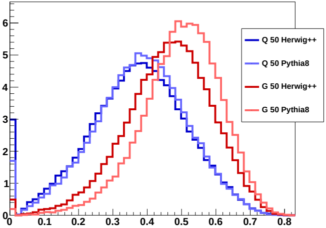
- Isolated anti-k_T jets with $R = 0.4$
- Only track-based variables to avoid pileup effects
- Charged track $p_T > 1\,\text{GeV}$
- In MC, jets were matched to highest energy parton within cone

Herwig++ 2.5.2 (darker) as compared to Pythia 8.165 (lighter) for 50 GeV quarks and gluons.

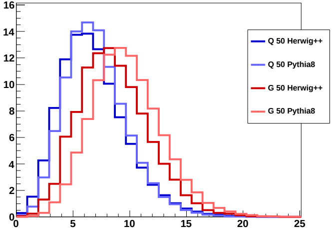
mass/p_T



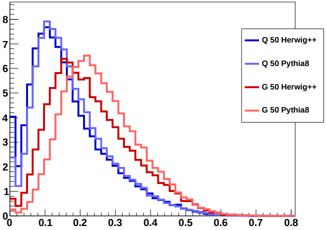
N-subjettiness $\beta = 1/4$



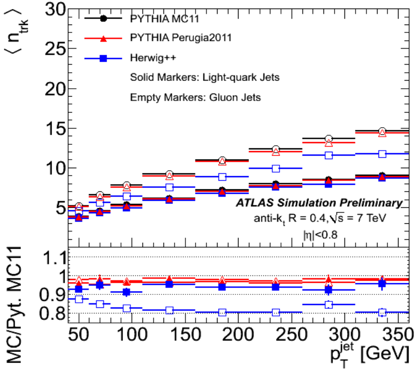
Charged Track Count

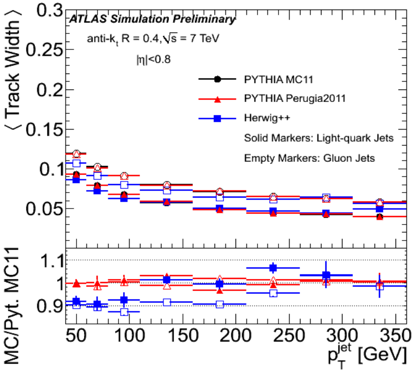


width (radial-moment)



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 Gonzlez Silva's talk at BOOST2012)

ATLAS's Template Method

- Goal: to measure the quark/gluon shapes from data, dijet (*DJ*) and photon+jet *(γ J)* events.
- Ideally, solve for q/g (for each bin i) from:

$$h_i(DJ) = P_Q(DJ)q_i + P_G(DJ)g_i$$
$$h_i(\gamma J) = P_Q(\gamma J)q_i + P_G(\gamma J)g_i$$

P_Q = quark percentage, from MC

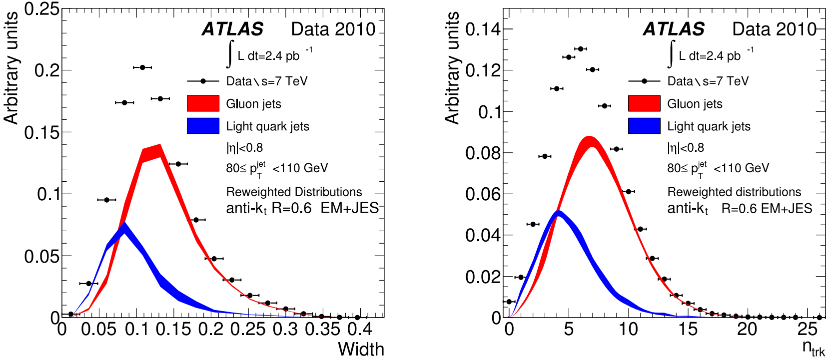
h = histogram value, from data

q/g = pure q/g jet distributions (solving for these)

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

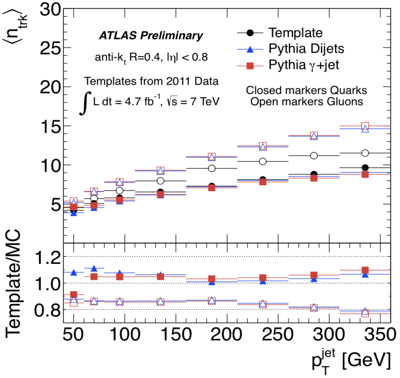
$$h_i(DJ) = P_Q(DJ)q_i + P_G(DJ)g_i + P_B(DJ)b_i + P_C(DJ)c_i$$
$$h_i(\gamma J) = P_Q(\gamma J)q_i + P_G(\gamma J)g_i + P_B(\gamma J)b_i + P_C(\gamma J)c_i$$

Di-jet data should match linear combination of pure quark + gluon.

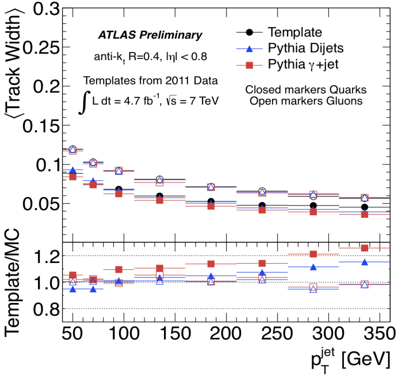


from "Jet energy measurement with the ATLAS..." arXiv:1112.6426
The width of the band represents the maximum variation among the Pythia and and the Herwig++ samples.

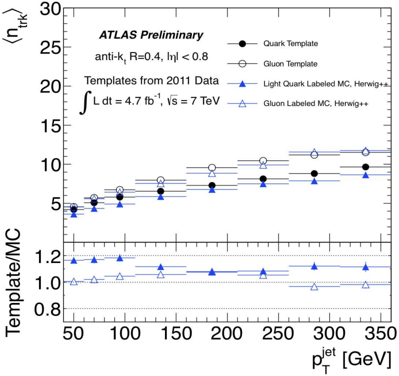
ATLAS Template vs Pythia8



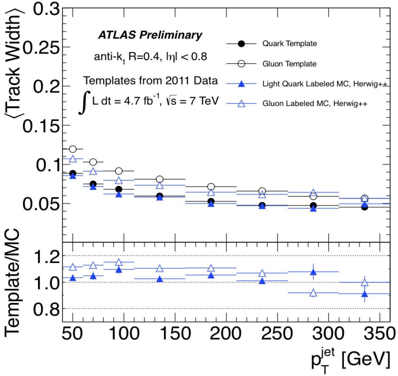
Charged Track Count differs for gluons



Width (radial-moment) agrees reasonably with Pythia8



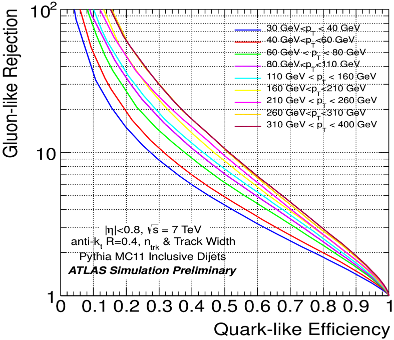
Charged Track Count
better in Herwig++
now quarks are off



Width (radial-moment)
worse in Herwig++

ATLAS ROC Curve for Data

Preliminary result shows data not looking as separable.



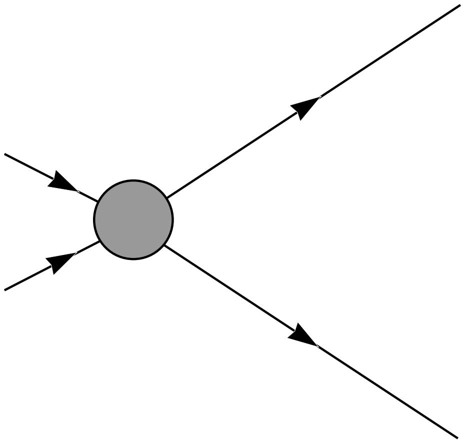
$|\eta|$ <0.8, Jet pT ~150 GeV:

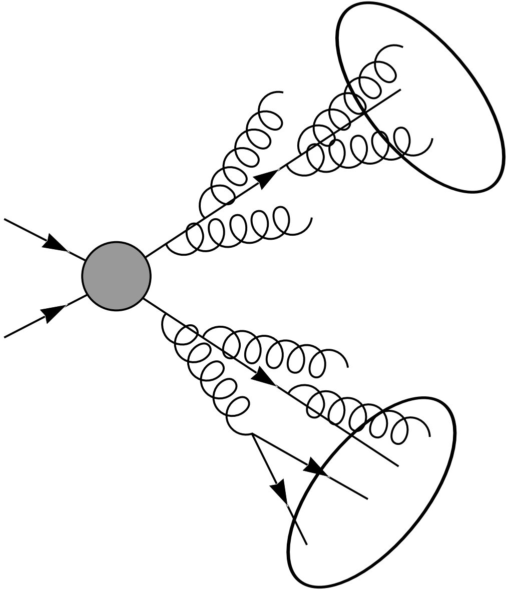
Sample	Efficiency	Rejection
Pythia MC11	50%	8x
Data 2011	50%	4x

- Purified samples validate these findings.
- Need different variables?
- Need more isolated jets?

QCD Jet Flavor Theory

Example of 2 Quark Jets





Angle of Attack

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

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)

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.

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.

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 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.

Is Flavor Meaningful Beyond Leading Order?

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.

Is Flavor Meaningful Beyond Leading Order?

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 Λ_{QCD}/E, jet size R, jet's mass-to-energy ratio m/E, etc.

Is Flavor Meaningful Beyond Leading Order?

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 Λ_{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.

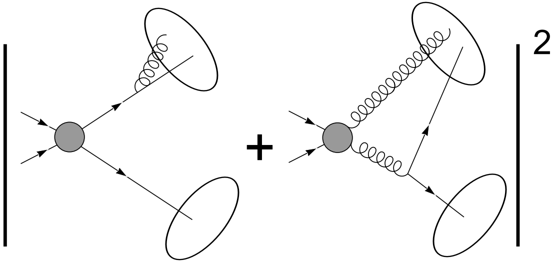
(All of this is sperate from measurement resolution.)

Problem Case

Loops? Same final state. No interference between flavors. Only rates.

Problem Case

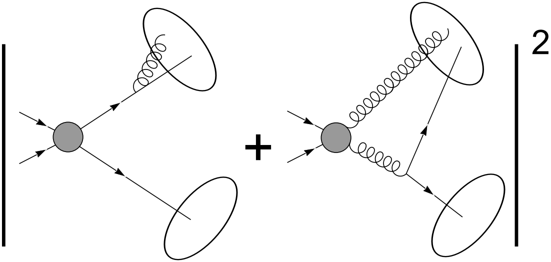
Loops? Same final state. No interference between flavors. Only rates. Problem case for unlikely splitting



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



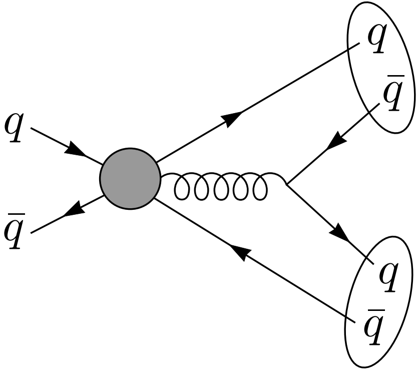
Given only final state, flavor-blind anti-k_T leaves these ambiguous. Each contribution is not individually gauge invariant and they interfere.

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

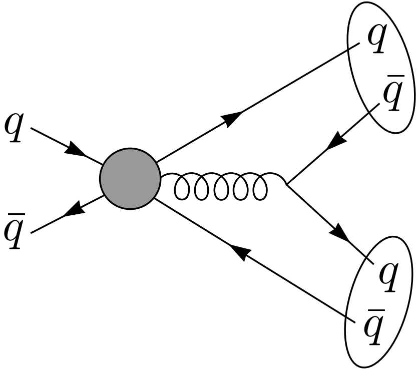
Another Problem Case

Hard gluon fails to make its own jet



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.

B-Tagging Has it Easier

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 $H \to b\bar{b}$.

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 $d_T > 10\,\mu\text{m}$.

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

B-Tagging Has it Easier

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 $H \to b\bar{b}$.

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

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 $g \to b\bar{b}$, 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 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)

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 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)

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.

Data-Driven Definitions

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)

Data-Driven Definitions

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 γ's rather than like QCD multijets.

Data-Driven Definitions

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 γ'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 γ'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.

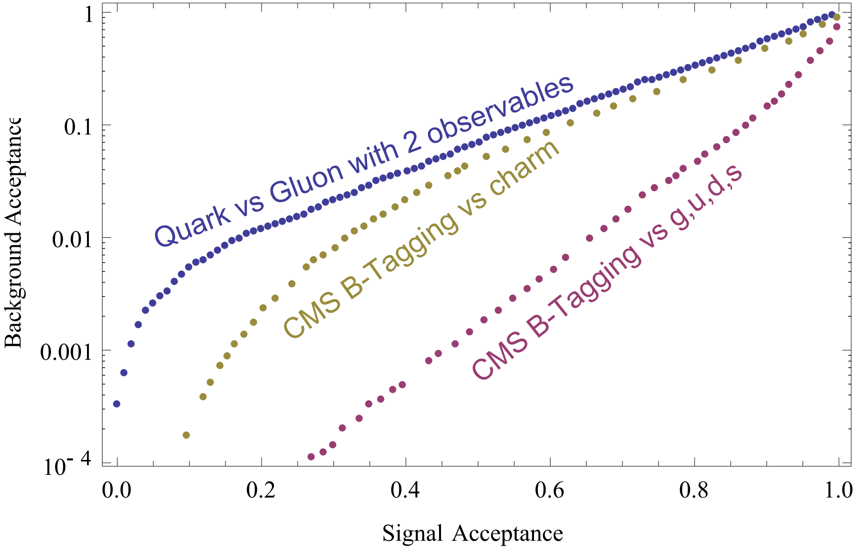
Thanks!

Using Flavor Taggers

Cutting, S/B, and $S/\sqrt{B}$

Cutting gives some signal acceptance and some background acceptance.

Comparison to B-Tagging



Cutting, S/B, and $S/\sqrt{B}$

A cut on tagger's score gives

- signal efficiency ϵ_s (you pick)
- background efficiency ϵ_b (ROC dictates)

Cutting, S/B, and $S/\sqrt{B}$

A cut on tagger's score gives

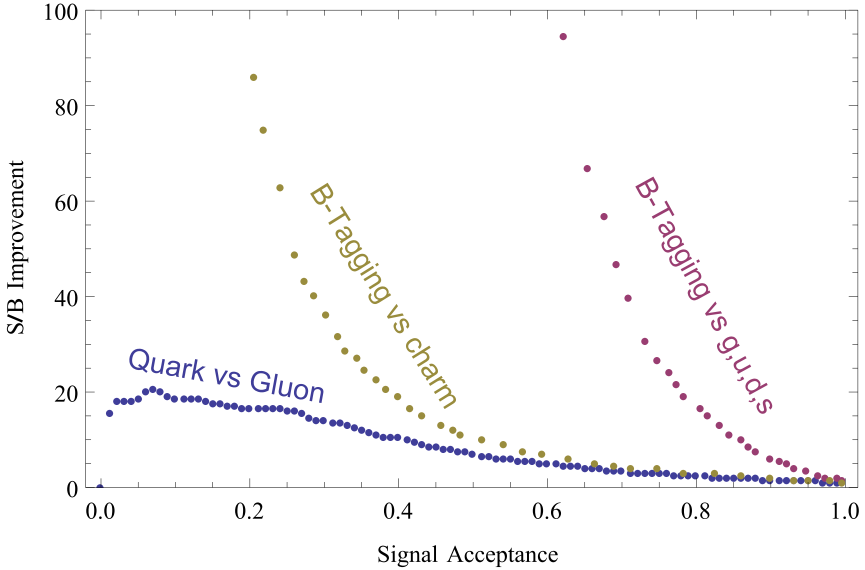
- signal efficiency ϵ_s (you pick)
- background efficiency ϵ_b (ROC dictates)

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

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

Call $\frac{\epsilon_s}{\epsilon_b}$ the "S/B Improvement"

Comparison to B-Tagging



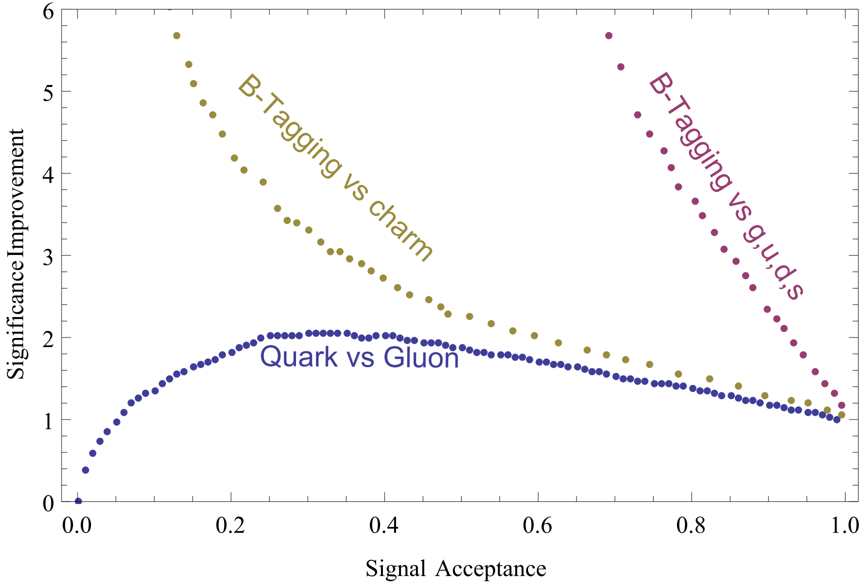
Cutting, S/B, and $S/\sqrt{B}$

Improvement in statistical significance scales differently

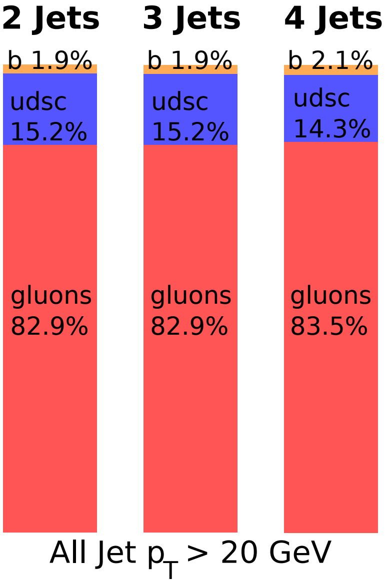
$$\sigma = \frac{S}{\sqrt{B}} \qquad \rightarrow \qquad \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma\frac{\epsilon_s}{\sqrt{\epsilon_b}}$$

Call $\frac{\epsilon_s}{\sqrt{\epsilon_b}}$ the "Significance Improvement"

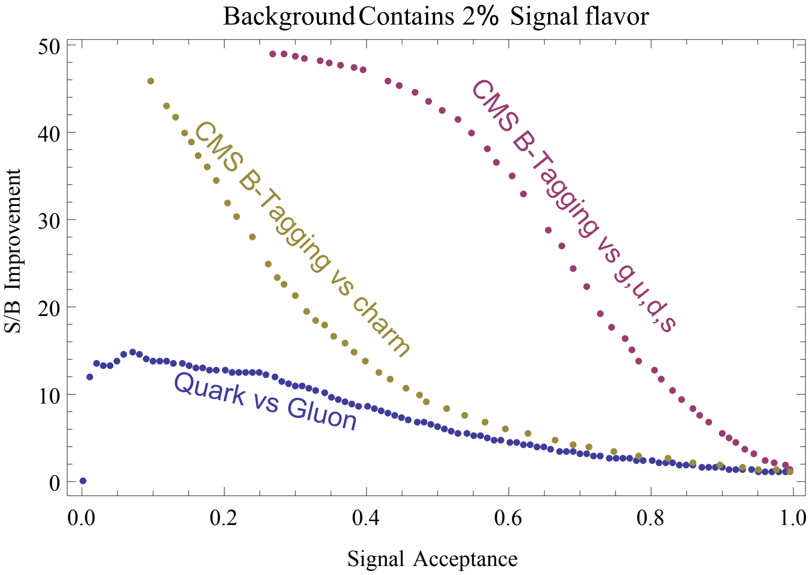
Comparison to B-Tagging



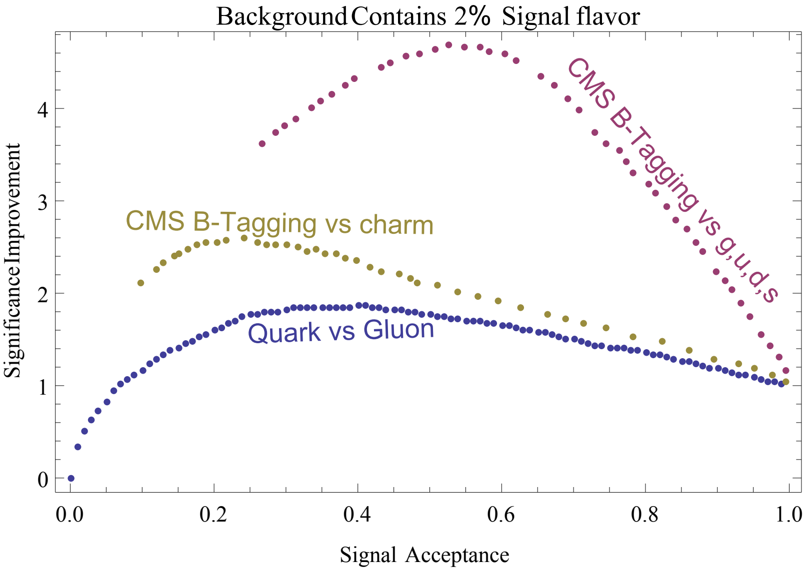
But Backgrounds Contain b's and light quarks!



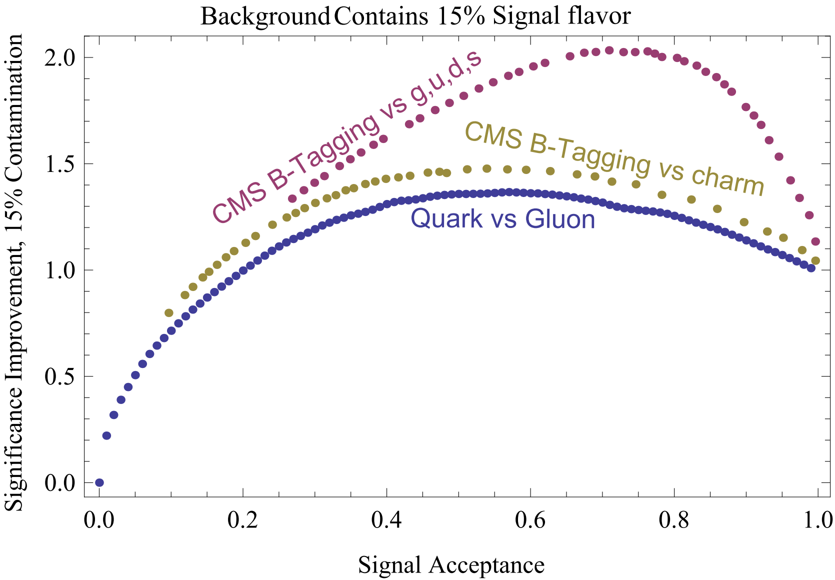
Background Contains 2% ‘Signal’ flavor (B-case)



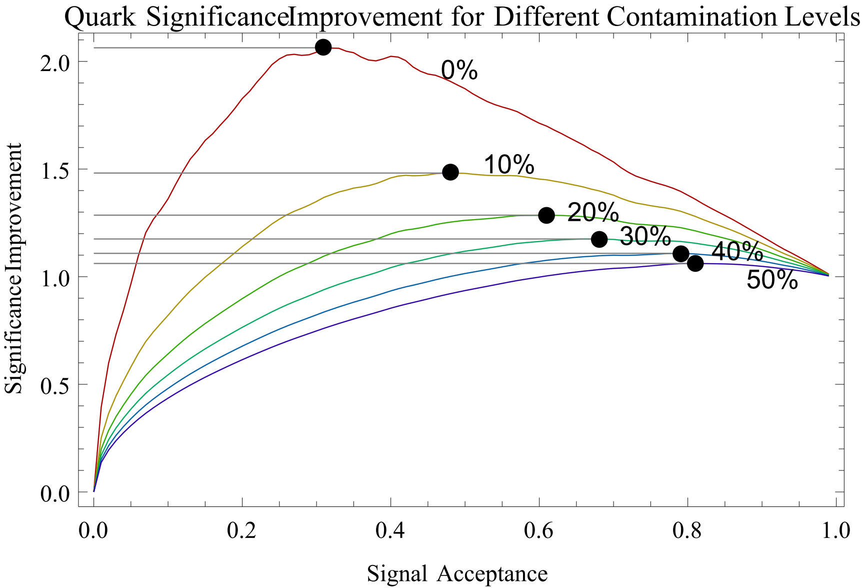
Background Contains 2% ‘Signal’ flavor (B-case)



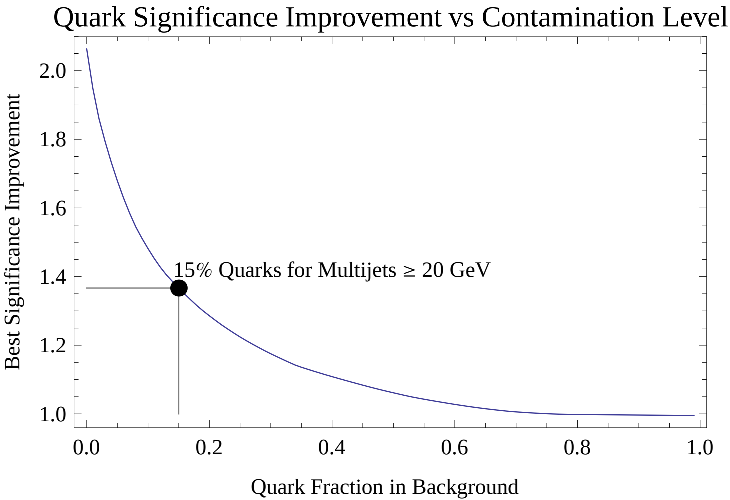
Background Contains 15% ‘Signal’ flavor (Q-case)



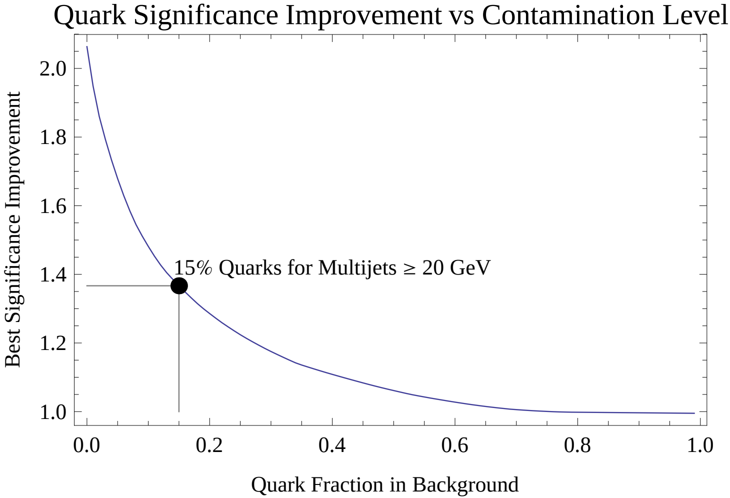
Operating Points that Maximize Quark Significance



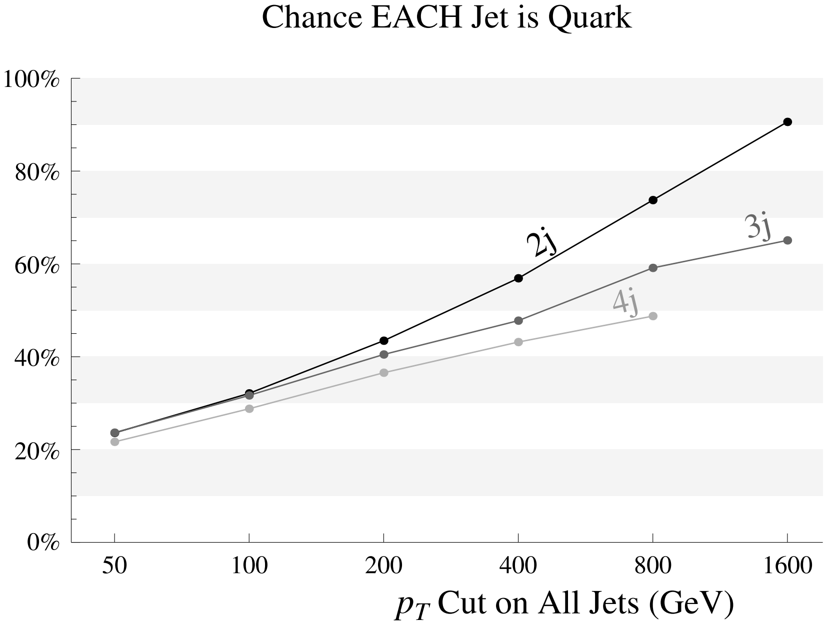
Operating Points that Maximize Quark Significance



Operating Points that Maximize Quark Significance



For signal of 4 quarks $\geq 20\,\mathrm{GeV}$, significance improvement is $1.37^4 = 3.5$


Chance EACH Jet is Quark
100%
80%
60%
40%
20%
0%
2j
3j
4j
50
100
200
400
800
1600
p_T Cut on All Jets (GeV)