quark/gluon discrimination

theory and data/MC interplay

Marat Freytsis

University of Oregon

SM@LHC — University of Pittsburgh, May 3–6, 2016

The white whale of jet substructure

- any measurement must involve treatment of jets
  - no uncolored channels *cf.* boosted tops/W/Z/Higgs/$\cdots$
- q/g distinction intrinsically from structure of jets
- no parametric separation
  - C_F vs. C_A, spin-1 vs. spin-$\frac{1}{2}$
- ubiquitous q/g differences in S vs. B at LHC
  - signals almost always "quark-rich"

Plan

- What is a quark jet?
- Calculable discrimination
  - Universality of LL behavior
  - Beyond leading log
  - “Marginally calculable” discriminants
- Tracking non-perturbative effects in MC and data
  - What do shower/hadronization tunes constrain?
  - Improving shape tunes at the LHC

What is a quark jet?

ill-defined

A quark parton

A Born-level quark parton

The initiating quark parton in a final state shower

An eikonal line with baryon number $\frac{1}{3}$ and carrying triplet color charge

A parton-level jet object that has been quark-tagged using an IRC-safe flavored jet algorithm

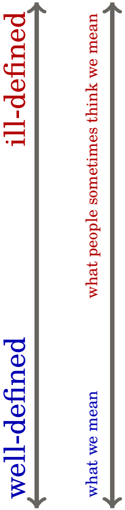
A quark operator appearing in a hard matrix element in the context of a factorization theorem

A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

well-defined

[J. Thaler, Les Houches summary talk]

What is a quark jet?



A quark parton

A Born-level quark parton

The initiating quark parton in a final state shower

An eikonal line with baryon number $\frac{1}{3}$ and carrying triplet color charge

A parton-level jet object that has been quark-tagged using an IRC-safe flavored jet algorithm

A quark operator appearing in a hard matrix element in the context of a factorization theorem

A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

[J. Thaler, Les Houches summary talk]

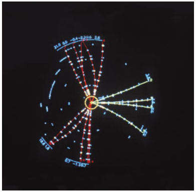
A rebranding, perhaps?

q/g ~~discrimination~~ enrichment

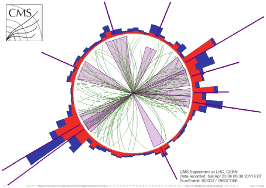
The goal is to enhance S over B using beyond-fixed-order information

…and it so happens

(perturbatively IRC-safe) ways of talking about parton flavor are useful for this



$\Longrightarrow$

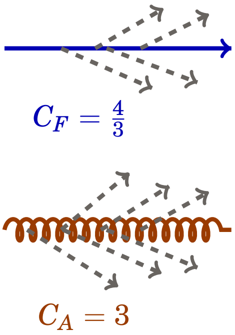


Plan

- What is a quark jet?
- Calculable discrimination
  - Universality of LL behavior
  - Beyond leading log
  - "Marginally calculable" discriminants
- Tracking non-perturbative effects in MC and data
  - What do shower/hadronization tunes constrain?
  - Improving shape tunes at the LHC

Semi-classical radiation

A cartoon:



- gluons carry larger color charge
  - radiate more, radiate wider
- shower is an iterative process
  - effect will exponentiate

Expectation:

$$1 - \epsilon_g \sim \epsilon_q^{C_A/C_F} = \epsilon_q^{9/4}$$

Universality of leading logarithm

let's see how this happens more carefully for jet angularities

$$e_\beta = \sum_{i \in \text{jet}} z_i \theta_i^\beta, \quad z_i = E_i / E_{\text{jet}}$$

integrated probability up to e_β:

$$\Sigma(e_\beta) = \exp\left(-\frac{\alpha}{\pi}\frac{C_{F,A}}{\beta}\ln^2\frac{R^\beta}{e_b}\right)$$

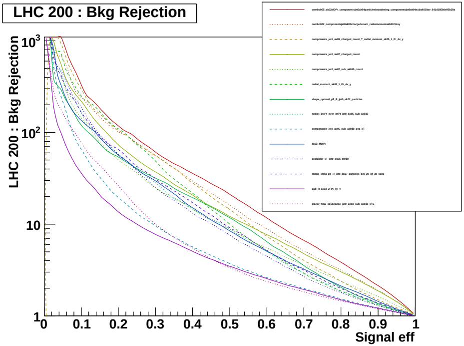
cut on e_β as expected, $\Sigma_g(e_\beta) = \Sigma_q(e_\beta)^{C_F/C_A}$

[Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014]

[Larkoski, Salam, Thaler, 1305.0007]

Multivariate combinations

can do better by combining multiple variables



[Gallicchio, Schwartz, 1106.3076]

$O(10000)$ variable combinations considered!
some improvement, but not much — LL behavior identical

NLL corrections

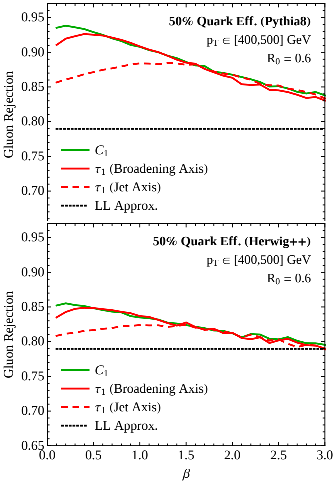
angularities simple, corrections to LL tractable
NLL corrections analytically known

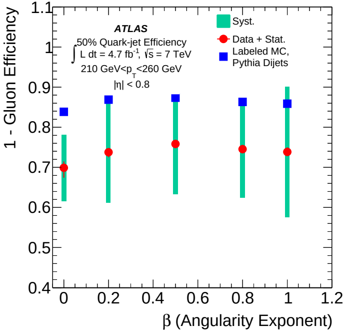
$$\ln \Sigma_g \simeq \frac{C_A}{C_F}\left(1 + \frac{n_F - C_A}{3C_A}\sqrt{-\frac{\alpha_s C_F}{\pi\beta \ln \Sigma_q}} + \frac{n_F - C_A}{C_A}\frac{\alpha_s}{36\pi}\frac{b_0}{\beta}(2-\beta) + \frac{\alpha_s \pi}{3}\frac{C_A - C_F}{\beta} + \frac{17}{36}\frac{\alpha_s}{\pi}\frac{C_F}{C_A}\frac{n_f - C_A}{\beta \ln \Sigma_q}\right)\ln \Sigma_q$$

Lower values of β preferred
$\implies$ increased sensitivity to wide-angle emission

[Larkoski, Salam, Thaler, 1305.0007]

NLL predictions vs. data





[ATLAS, 1405.6583]

(maybe) Pythia is too optimistic, Herwig++ too pessimistic
but why is data below LL prediction?

Broaden scope

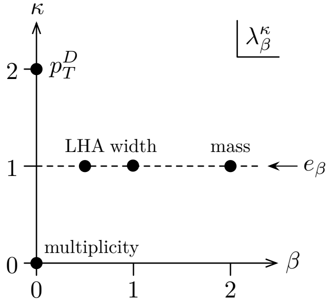
Generalized angularities:

$$\lambda_\beta^\kappa = \sum_{\text{jet}} z_i^\kappa \theta_i^\beta$$

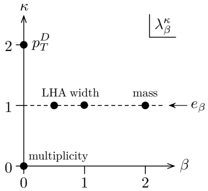
parametrizes a family of IRC safe and unsafe variables

many already used for q/g discrimination

[Larkoski, Thaler, Waalewijn, 1408.3122]



Beyond IRC safety



- $\kappa = 1$ IRC safe
  - breaks as $\beta \to 0$
- $\beta = 0$ is IR safe
  - breaks as $\kappa \to 0$
  - non-perturbative collinear function
  - perturbative evolution

$$\mu \frac{\partial}{\partial \mu} \boldsymbol{F}_\kappa^i \simeq \frac{\alpha_s}{\pi} \boldsymbol{P}_{i \to jk} \otimes \boldsymbol{F}_\kappa^j \otimes \boldsymbol{F}_\kappa^k$$

[Larkoski, Thaler, Waalewijn, 1408.3122]

- region around $\kappa \approx 1$ well described by 1 NP parameter

$$\lambda_\beta^\kappa \approx (e_{\beta/\kappa})^\kappa e^{\langle \ln z^\kappa \rangle_g}$$

Diagnosing MC behavior with λ_β^κ

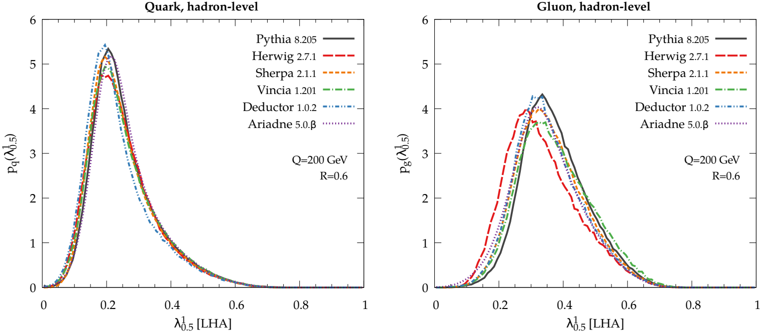
λ_β^κ includes many commonly used IRC un/safe variables
looking at behavior in MCs checks many expectations

- IRC safe variables insensive to hadronization details
- Scale dependence of IR safe variables predicted
- Jet parameters indirectly depend on scale of $\alpha_s(\mu)$

Plan

- What is a quark jet?
- Calculable discrimination
  - Universality of LL behavior
  - Beyond leading log
  - “Marginally calculable” discriminants
- Tracking non-perturbative effects in MC and data
  - What do shower/hadronization tunes constrain?
  - Improving shape tunes at the LHC

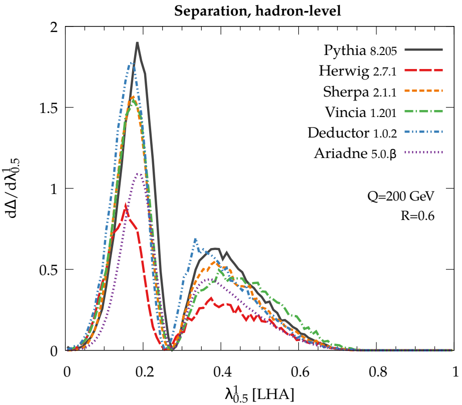
LHA



Agreement on q, variation in performance from g modeling

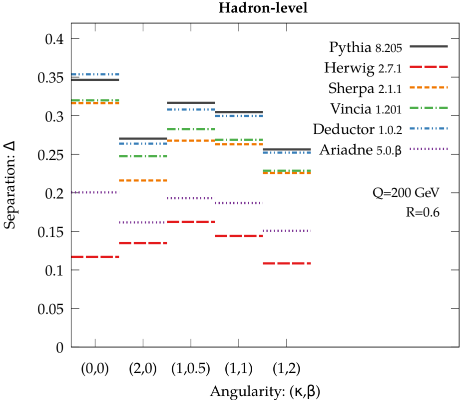
All results here with $e^+e^- \to Z/\gamma^* \to u\bar{u}$, $e^+e^- \to h \to gg$

LHA



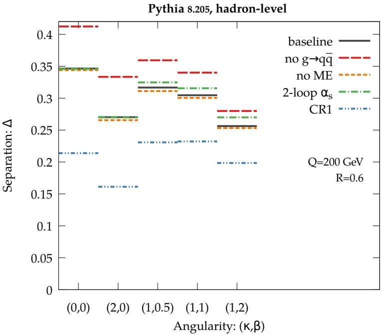
$$\Delta = \frac{1}{2}\int d\lambda \frac{(p_q(\lambda) - p_q(\lambda)^2}{p_q(\lambda) + p_q(\lambda}$$

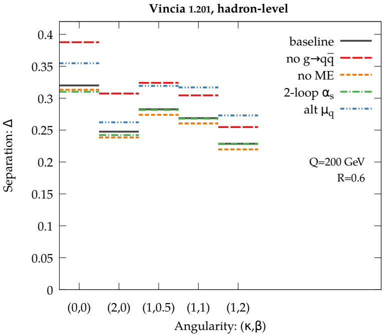
Classifier separation



Shower options

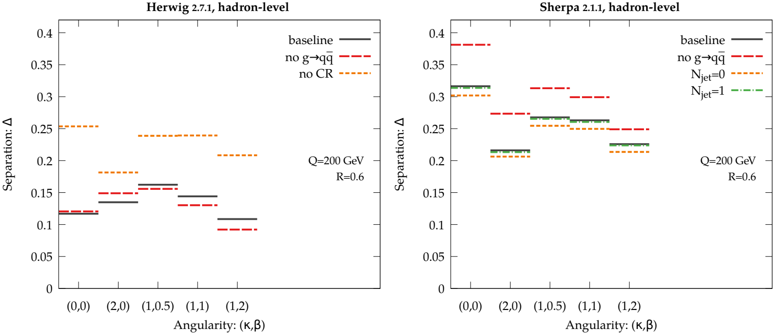
String hadronization



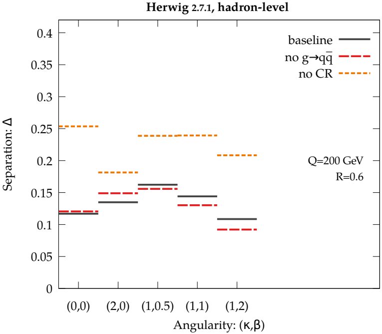


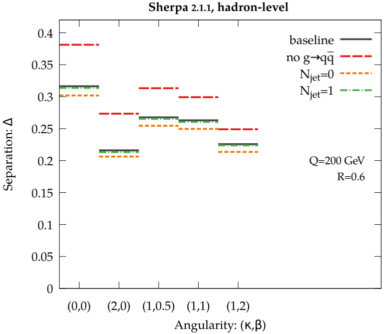
Shower options

Cluster hadronization



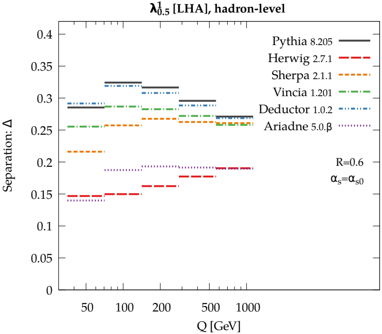
Shower options
Cluster hadronization

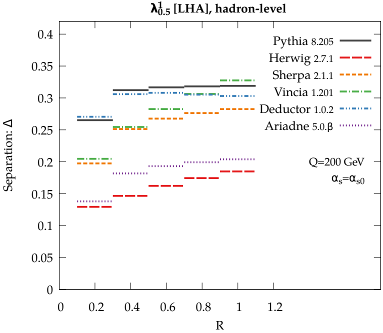




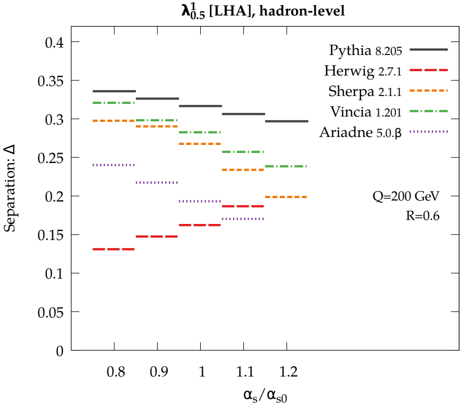
color reconnection and $g \to q\bar{q}$ splitting just as important for IRC safe observables!

Parameter variation





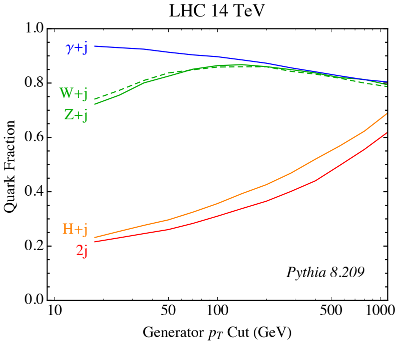
Parameter variation



what's happening with Herwig here?

A: nonperturbative physics

Towards the LHC



can we tune some of this information from LHC data?

- goal not necessarily pure samples
  - See Gallicchio, Schwartz, 1104.1175 for that
- find processes with significant gluon fraction + robust control over event/jet shape uncertainties

Conclusions

- known knowns
  - good understanding of IRC observables up to NLL
  - single log corrections improve performance at low β
  - for most unsafe variables, still some calculable control
    - reduction in nonpert. sensitivity or pert. running
- known unknowns
  - gluon jets unconstrained by LEP event shape data
  - higher-order effects important even for IRC observables
    - color reconnection at e^+e^-, $g \to q\bar{q}$
- unknowns unknowns
  - best way to constrain gluon jet shapes at LHC
  - other handles lying beyond NLL

Thank you!