On the Topic of Jets

BOOST 2018

Eric M. Metodiev
Center for Theoretical Physics
Massachusetts Institute of Technology

Joint work with Patrick T. Komiske and Jesse Thaler

July 19, 2018

Quark and Gluon Jets

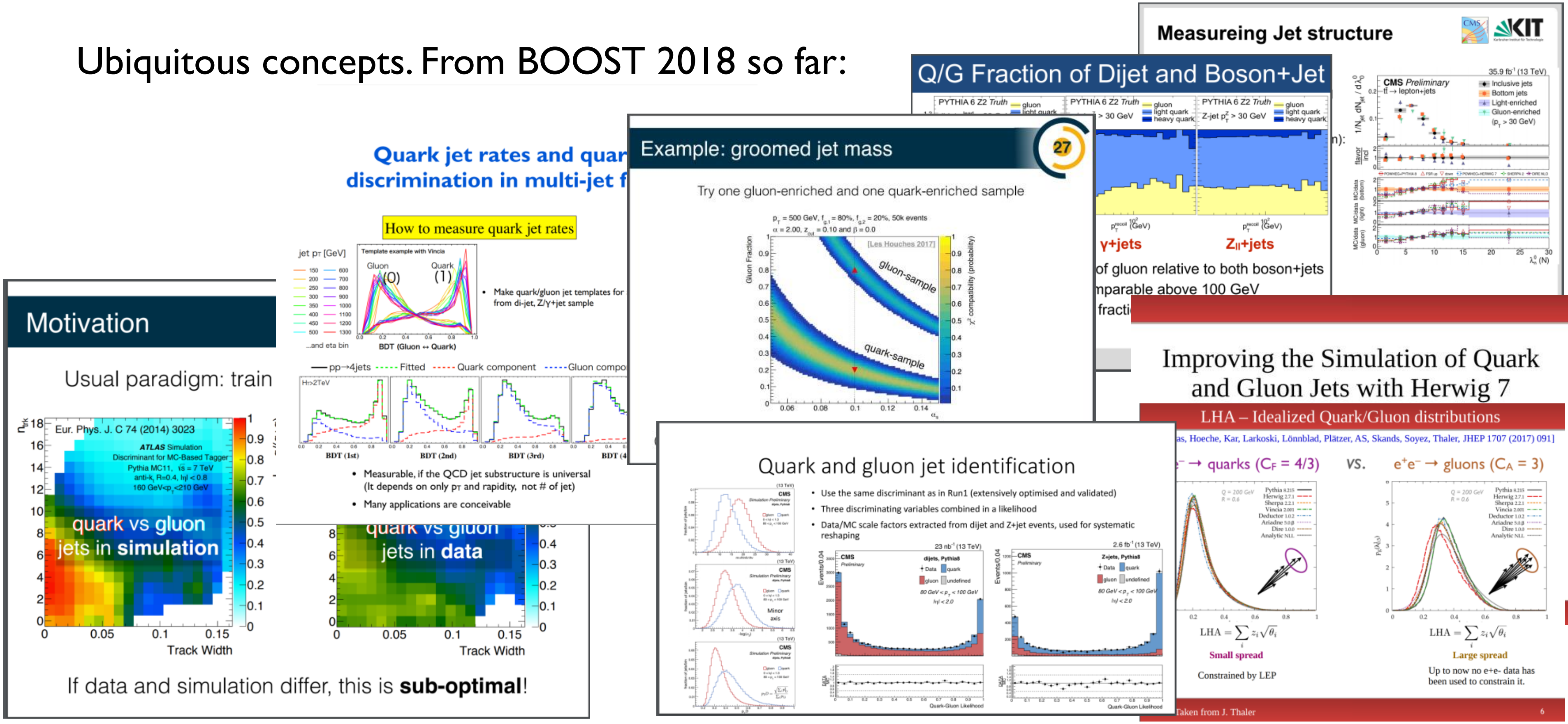
Quarks are color triplets and Gluons are color octets. We observe color-singlet hadrons.

No unambiguous hadron-level definition of jet flavor.

We often rely on unphysical notions such as parton shower event records to define jet flavor in practice.

Can quark and gluon be made well-defined nonetheless? Similar to defining jets themselves.

Ubiquitous concepts. From BOOST 2018 so far:



Motivation

Usual paradigm: train

If data and simulation differ, this is **sub-optimal**!

Quark jet rates and quar

discrimination in multi-jet f

How to measure quark jet rates

- Make quark/gluon jet templates for di-jet, Z/γ+jet sample
- Measurable, if the QCD jet substructure is universal (It depends on only pT and rapidity, not # of jet)
- Many applications are conceivable

Example: groomed jet mass

Try one gluon-enriched and one quark-enriched sample

Q/G Fraction of Dijet and Boson+Jet

of gluon relative to both boson+jets

mparable above 100 GeV

fracti

Measureing Jet structure

Quark and gluon jet identification

- Use the same discriminant as in Run1 (extensively optimised and validated)
- Three discriminating variables combined in a likelihood
- Data/MC scale factors extracted from dijet and Z+jet events, used for systematic reshaping

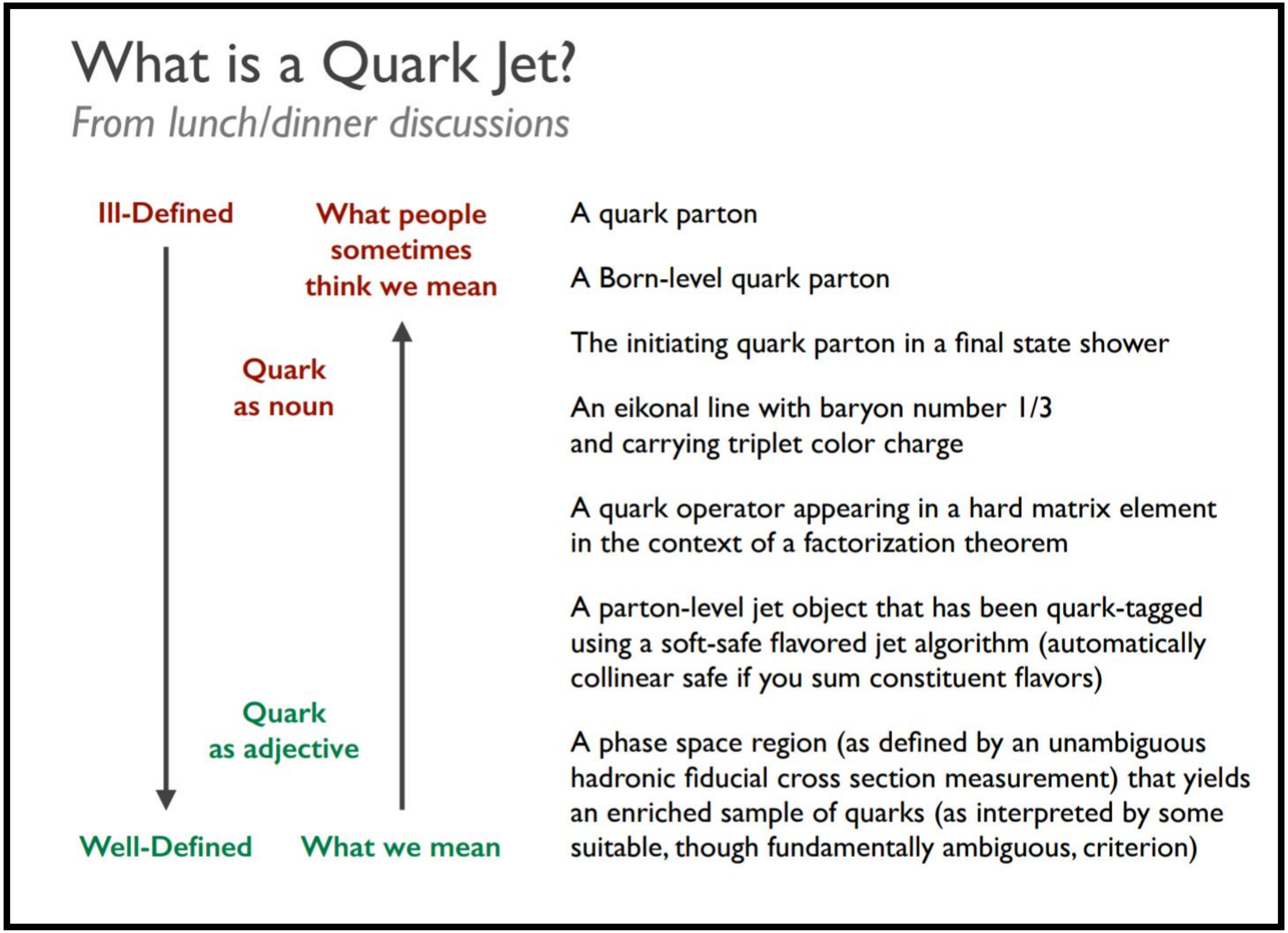
Improving the Simulation of Quark and Gluon Jets with Herwig 7

LHA – Idealized Quark/Gluon distributions

as, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, AS, Skands, Soyez, Thaler, JHEP 1707 (2017) 091]

Taken from J. Thaler

What are “Quark” and “Gluon” Jets?



What is a Quark Jet?

From lunch/dinner discussions

Ill-Defined

Well-Defined

What people sometimes think we mean

Quark as noun

Quark as adjective

What we mean

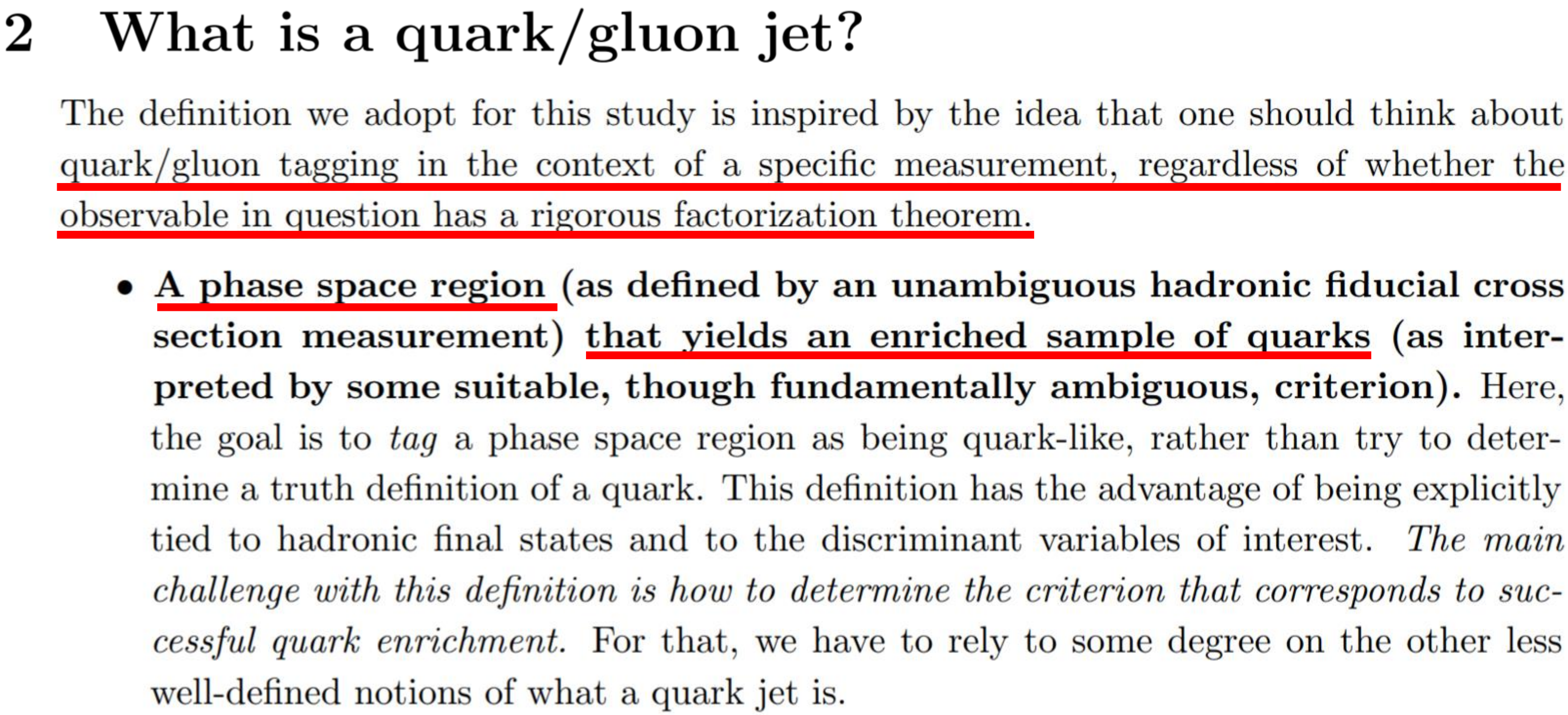
	Word Count
A quark parton	3
A Born-level quark parton	4
The initiating quark parton in a final state shower	9
An eikonal line with baryon number 1/3 and carrying triplet color charge	12
A quark operator appearing in a hard matrix element in the context of a factorization theorem	16
A parton-level jet object that has been quark-tagged using a soft-safe flavored jet algorithm (automatically collinear safe if you sum constituent flavors)	22
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)	30

[Les Houches 2015 Report]
[P. Gras, *et al.*, 1704.03878]

[P. Gras, *et al.*, 1704.03878]

Systematics of quark/gluon tagging

Philippe Gras,[a] **Stefan Höche,**[b] **Deepak Kar,**[c] **Andrew Larkoski,**[d]
Leif Lönnblad,[e] **Simon Plätzer,**[f,g] **Andrzej Siódmok,**[h,i] **Peter Skands,**[j]
Gregory Soyez,[k,†] **and Jesse Thaler**[l,†]



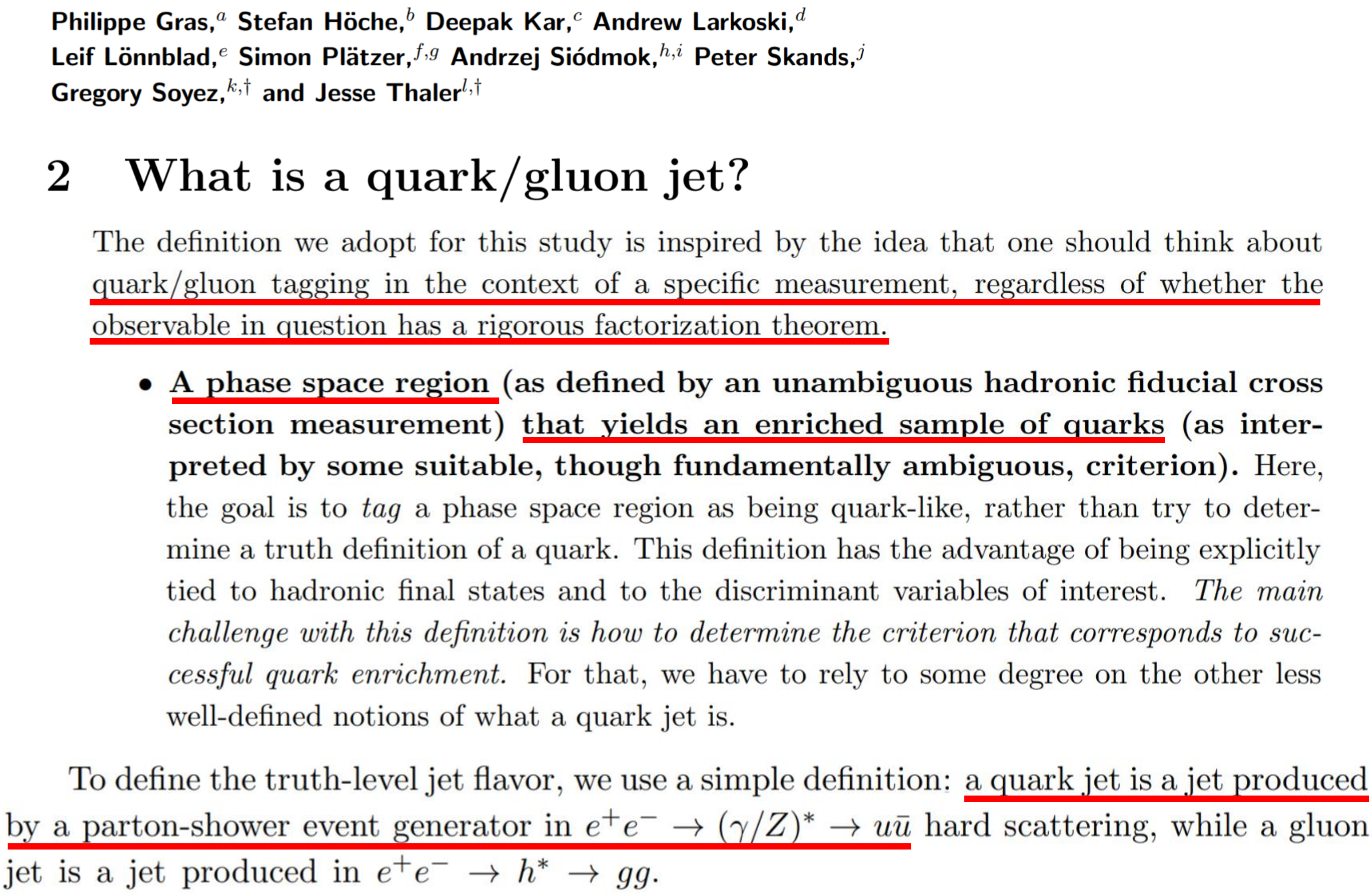
2 What is a quark/gluon jet?

The definition we adopt for this study is inspired by the idea that one should think about quark/gluon tagging in the context of a specific measurement, regardless of whether the observable in question has a rigorous 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).** Here, the goal is to *tag* a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. *The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment.* For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is.

Systematics of quark/gluon tagging

[P. Gras, *et al.*, 1704.03878]



Philippe Gras,[a] Stefan Höche,[b] Deepak Kar,[c] Andrew Larkoski,[d]
Leif Lönnblad,[e] Simon Plätzer,[f,g] Andrzej Siódmok,[h,i] Peter Skands,[j]
Gregory Soyez,[k,†] and Jesse Thaler[l,†]

2 What is a quark/gluon jet?

The definition we adopt for this study is inspired by the idea that one should think about quark/gluon tagging in the context of a specific measurement, regardless of whether the observable in question has a rigorous factorization theorem.

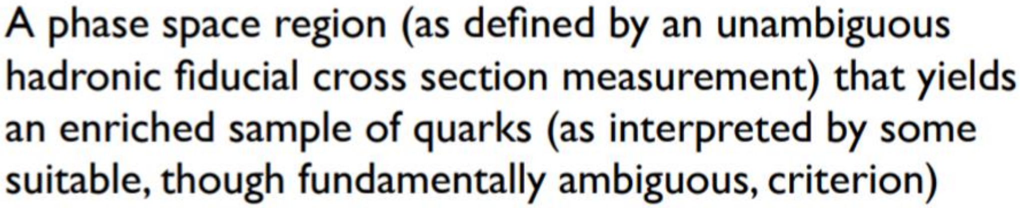
- **A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion).** Here, the goal is to *tag* a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. *The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment.* For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is.

To define the truth-level jet flavor, we use a simple definition: a quark jet is a jet produced by a parton-shower event generator in $e^+e^- \to (\gamma/Z)^* \to u\bar{u}$ hard scattering, while a gluon jet is a jet produced in $e^+e^- \to h^* \to gg$.

Our Plan: An operational definition of quark and gluon jets

That definition:

[A quark jet is defined by:]



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)

This talk: Translating those 30 words to these 2 equations:

$$p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1 - \kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1 - \kappa_{\text{BA}}}$$

A picture of quark and gluon jets

1. Take your favorite jet algorithm

2. Consider two jet samples A and B of QCD jets

3. Choose a jet substructure observable $\boldsymbol{x}$

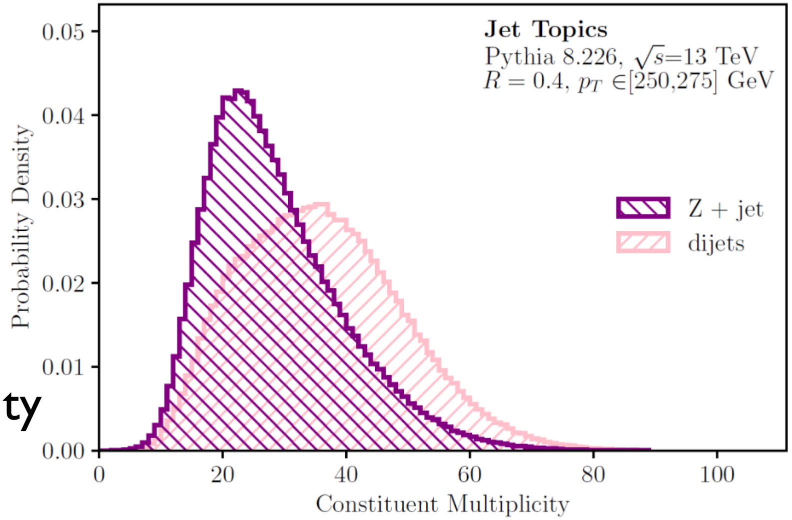
4. "Assume" that "quark" and "gluon" jets exist

5. "Assume" "quark/gluon" jet mutual irreducibility

Anti-kT R=0.4 jets

Z+jet and Dijets

Consistuent Multiplicity



The samples A and B are statistical mixtures of quark and gluon:

$$p_{\text{sample }A}(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + f_A^g\, p_{\text{gluon}}(\boldsymbol{x}), \qquad f_A^g = 1 - f_A^q$$

$$p_{\text{sample }B}(\boldsymbol{x}) = f_B^q\, p_{\text{quark}}(\boldsymbol{x}) + f_B^g\, p_{\text{gluon}}(\boldsymbol{x}), \qquad f_B^g = 1 - f_A^q$$

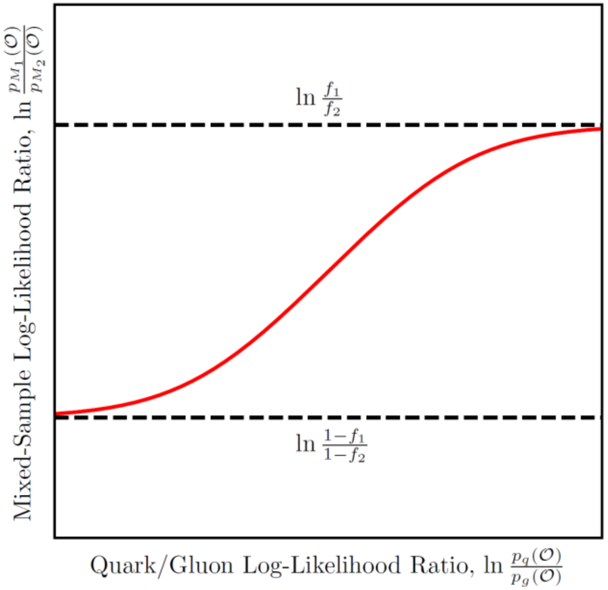
Similar picture to template- and fraction-based methods.

A/B Likelihood Ratio

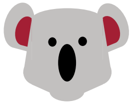
$$p_{\text{sample }A}(\boldsymbol{x}) = f_A^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_{\text{sample }B}(\boldsymbol{x}) = f_B^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$\boxed{L_{\frac{\text{A}}{\text{B}}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{f_A^q \, L_{\frac{\text{quark}}{\text{gluon}}}(\boldsymbol{x}) + \left(1 - f_A^q\right)}{f_B^q \, L_{\frac{\text{quark}}{\text{gluon}}}(\boldsymbol{x}) + \left(1 - f_B^q\right)}}$$



The A/B and quark/gluon likelihood ratios are monotonic!

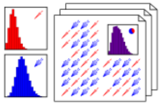


Classification without labels (CWoLa)

- Optimal A/B classifier is the optimal quark/gluon classifier.
- Use machine learning to approximate A/B likelihood ratio.

See Ben's talk! [EMM, B. Nachman, J. Thaler, 1708.02949]

The A/B likelihood ratio is bounded between $\frac{f_A^q}{f_B^q}$ and $\frac{1-f_A^q}{1-f_B^q}$!



Jet Topics

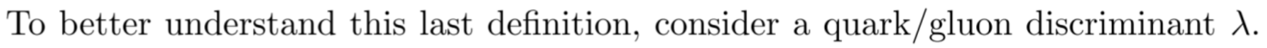
- "Mutually irreducibility" means the bounds saturate
- Obtain the maxima and minima of the A/B likelihood ratio.
- Solve for the quark/gluon fractions and distributions.

[EMM, J. Thaler, 1802.00008]

[P. Gras, *et al.*, 1704.03878]

Systematics of quark/gluon tagging

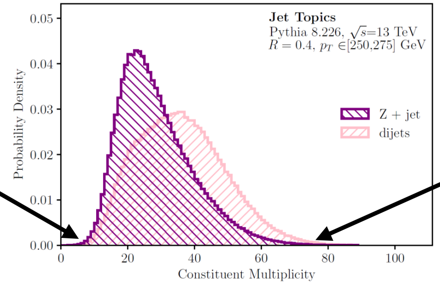
Philippe Gras,[a] Stefan Höche,[b] Deepak Kar,[c] Andrew Larkoski,[d] Leif Lönnblad,[e] Simon Plätzer,[f,g] Andrzej Siódmok,[h,i] Peter Skands,[j] Gregory Soyez,[k,†] and Jesse Thaler[l,†]



To better understand this last definition, consider a quark/gluon discriminant λ.

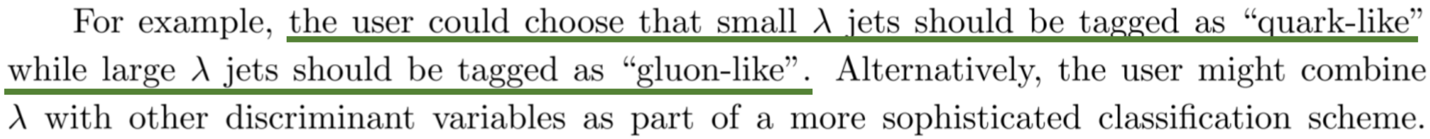
“quarks”

$$L_{\frac{\text{quark}}{\text{gluon}}}(x) \to \infty$$



“gluons”

$$L_{\frac{\text{quark}}{\text{gluon}}}(x) \to 0$$



For example, the user could choose that small λ jets should be tagged as “quark-like” while large λ jets should be tagged as “gluon-like”. Alternatively, the user might combine λ with other discriminant variables as part of a more sophisticated classification scheme.

Mutual irreducibility!

These concepts are not new in physics, and have been around for a while.

Quark/gluon mutual irreducibility: There are some substructure phase space regions where quark and gluon jets are pure.

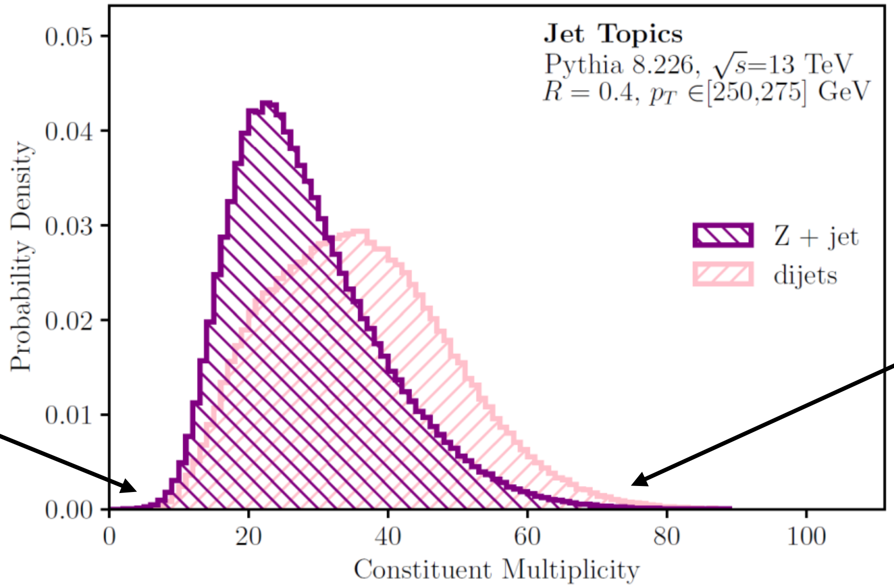
$$\min_{x} \frac{p_B(x)}{p_A(x)} = \frac{f_B^q}{f_A^q} \qquad \min_{x} \frac{p_A(x)}{p_B(x)} = \frac{1 - f_A^q}{1 - f_B^q}$$

Demixing the mixtures

$$p_A(\boldsymbol{x}) = f_A^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_B(\boldsymbol{x}) = f_B^q \, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$\kappa_{\text{BA}} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})} = \frac{f_B^q}{f_A^q}$$



$$\kappa_{\text{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

Solve for the quark and gluon distributions and fractions:

$$f_A^q = \frac{1 - \kappa_{\text{AB}}}{1 - \kappa_{\text{AB}}\kappa_{\text{BA}}} \qquad f_{\text{B}}^q = \frac{\kappa_{\text{BA}}(1 - \kappa_{\text{AB}})}{1 - \kappa_{\text{AB}}\kappa_{\text{BA}}}$$

$$p_{\text{quark}}(\boldsymbol{x}) = \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}} \, p_B(\boldsymbol{x})}{1 - \kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) = \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}} \, p_A(\boldsymbol{x})}{1 - \kappa_{\text{BA}}}$$

Demixing the mixtures

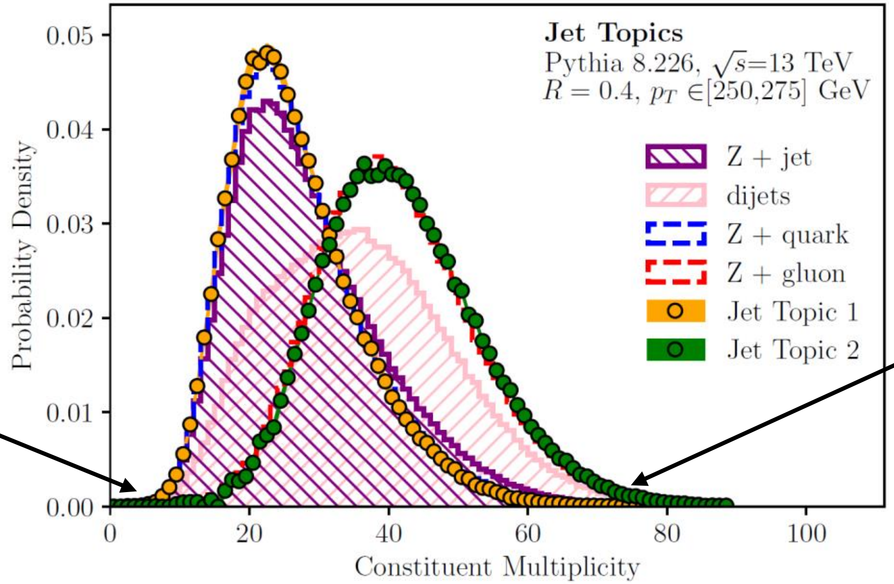
Defined from data

$$p_A(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

$$p_B(\boldsymbol{x}) = f_B^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_B^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

Ambiguous?

$$\kappa_{\text{BA}} \equiv \min_{\boldsymbol{x}} \frac{p_B(\boldsymbol{x})}{p_A(\boldsymbol{x})} = \frac{f_B^q}{f_A^q}$$



$$\kappa_{\text{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

Solve for the quark and gluon distributions and fractions:

$$f_A^q = \frac{1 - \kappa_{\text{AB}}}{1 - \kappa_{\text{AB}}\kappa_{\text{BA}}} \qquad f_{\text{B}}^q = \frac{\kappa_{\text{BA}}(1 - \kappa_{\text{AB}})}{1 - \kappa_{\text{AB}}\kappa_{\text{BA}}}$$

$$p_{\text{quark}}(\boldsymbol{x}) = \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1 - \kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) = \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1 - \kappa_{\text{BA}}}$$

An operational definition of quark and gluon jets

Quark and Gluon Jet Definition (Operational): Given two samples A and B of QCD jets at a fixed p_T obtained by a suitable jet-finding procedure, taking A to be "quark-enriched" compared to B, and a jet substructure feature space $\boldsymbol{x}$, quark and gluon jet distributions are defined to be:

$$\boxed{p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1-\kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1-\kappa_{\text{BA}}}}$$

Well-defined and operational statement in terms of hadronic cross sections.

Not a per-jet flavor label, but rather an aggregate distribution label.

Defined in the context of a specific pair of samples A and B, regardless of whether the observable in question has a rigorous factorization theorem.

Additional jet processing (e.g. grooming) can be folded into definition of A and B.

Extracting topics well is fundamentally easier than tagging well.

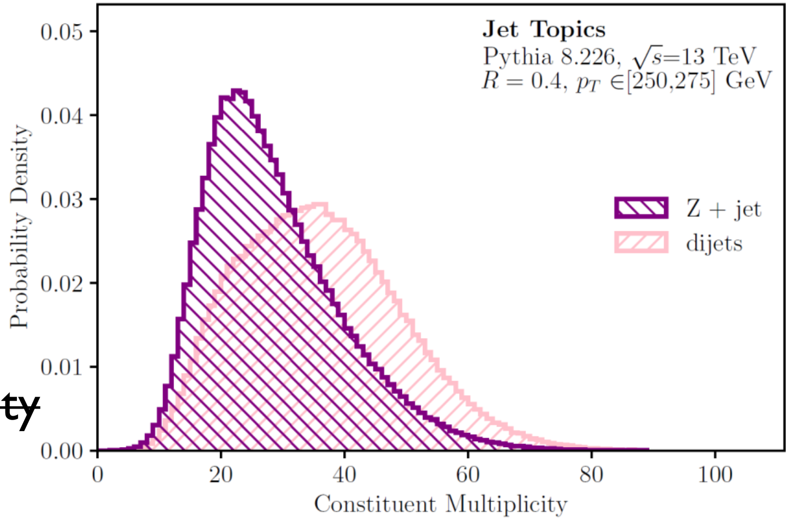
A picture of quark and gluon jets

1. Take your favorite jet algorithm
2. Consider two jet samples A and B of QCD jets
3. Choose a jet substructure observable $\boldsymbol{x}$
4. ~~"Assume" that "quark" and "gluon" jets exist~~
5. ~~"Assume" "quark/gluon" jet mutual irreducibility~~

Anti-kT R=0.4 jets

Z+jet and Dijets

Consistuent Multiplicity



The samples A and B are statistical mixtures of quark and gluon:

$$p_{\text{sample }A}(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + f_A^g\, p_{\text{gluon}}(\boldsymbol{x}), \qquad f_A^g = 1 - f_A^q$$

$$p_{\text{sample }B}(\boldsymbol{x}) = f_B^q\, p_{\text{quark}}(\boldsymbol{x}) + f_B^g\, p_{\text{gluon}}(\boldsymbol{x}), \qquad f_B^g = 1 - f_A^q$$

Firm foundation for data-driven methods.

Exploring substructure feature spaces

Why restrict ourselves to multiplicity? It works, but we can explore this choice.

We can also use a *trained model* (with CWoLa) as an observable in its own right.

Observables

- Multiplicity n_{const}
  Number of particles in the jet
- Soft Drop Multiplicity n_{SD}
  Probes number of perturbative emissions
- Image Activity N_{95}
  Number of pixels with 95% of jet p_T
- N-subjettiness $\tau_2^{(\beta=1)}$
  Probes how multi-pronged the jet is
- Jet Mass m
  Mass of the total jet four-vector
- Width w
  Probes the girth of the jet

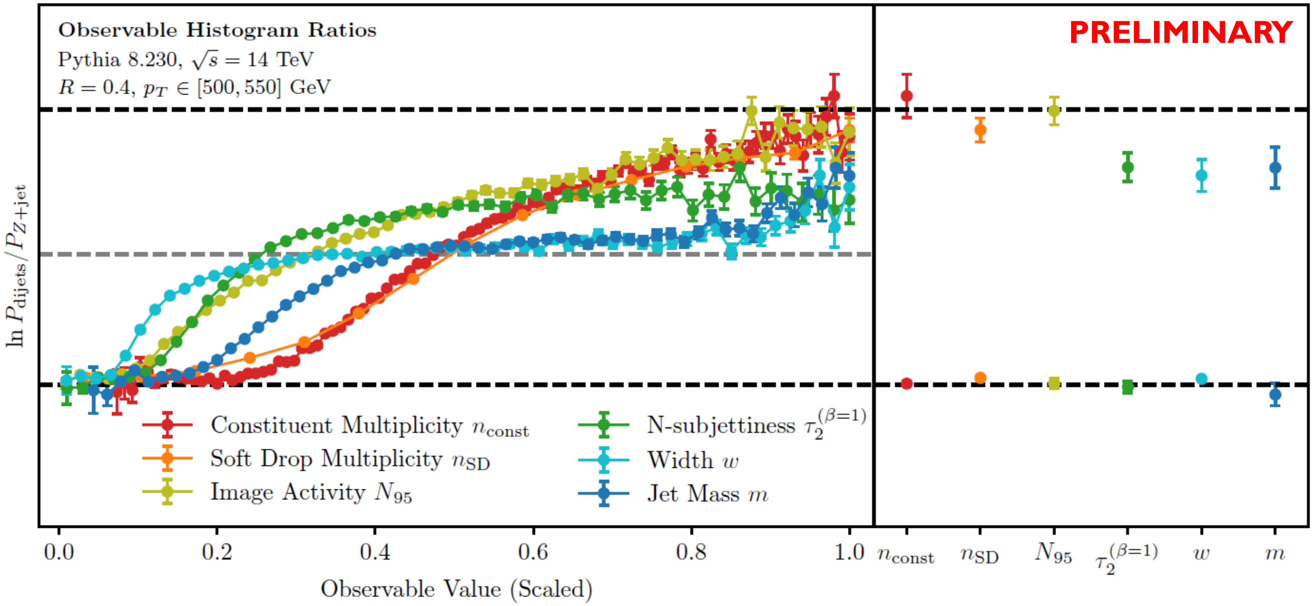
Models

- PFN-ID
  Full particle-level information
- PFN
  Full four-momentum information
- EFN
  Full IRC-safe information
- EFPs
  Full IRC-safe information, linearly

See Patrick's talk!

- CNN
  Trained on two-channel jet images
- DNN
  Trained on an N-subjettiness basis

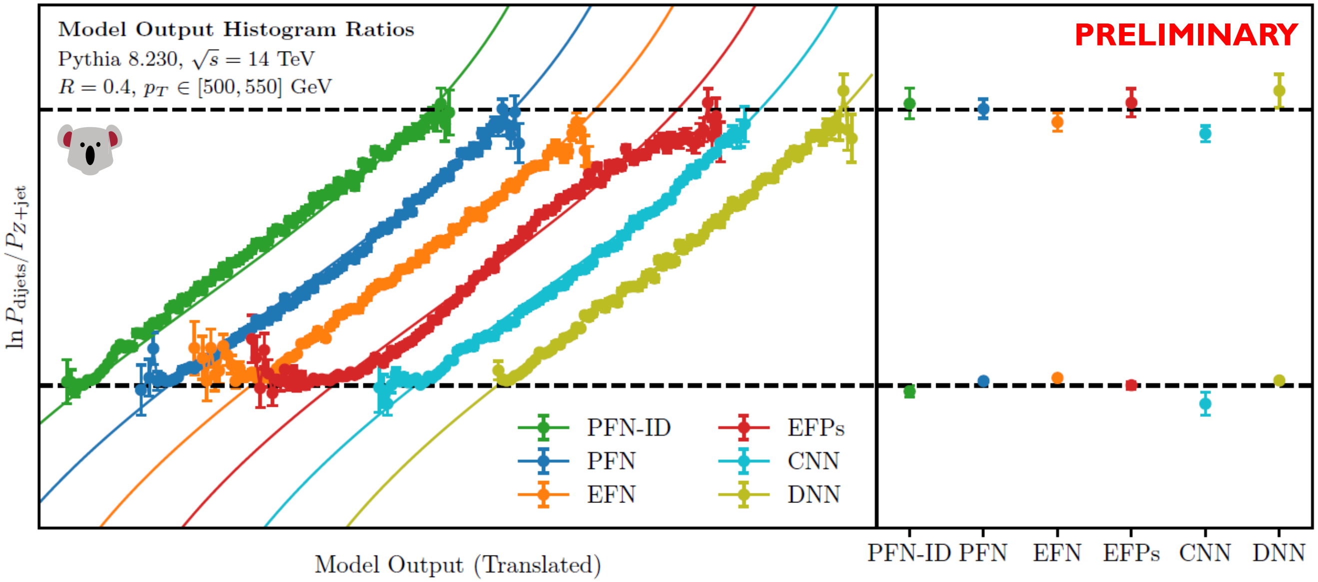
Exploring substructure feature spaces



Casimir scaling of mass and width is observed (gray).

Count observables come closer to saturating the bounds (black) than shape observables.

Lower bound easier to extract than upper. (i.e. Gluons are easy!)



Models CWoLa-trained. Fully data-driven.

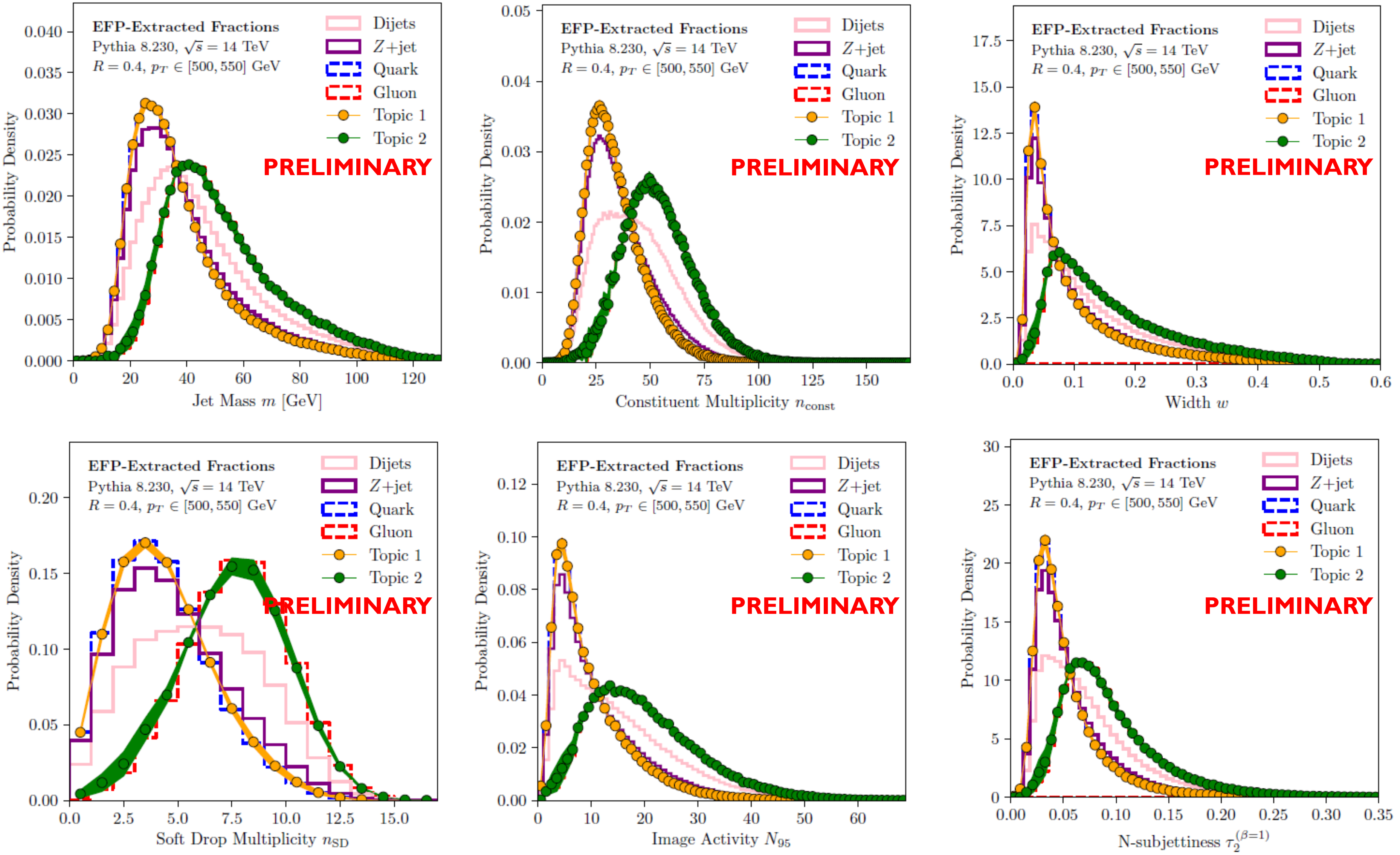
Well-behaved likelihoods close to S/(S+B) expectation.

All different models manifest the same bounds.

Insensitive to the model details.

[P.T. Komiske, EMM, J. Thaler, Upcoming.]

Extracting quark and gluon distributions

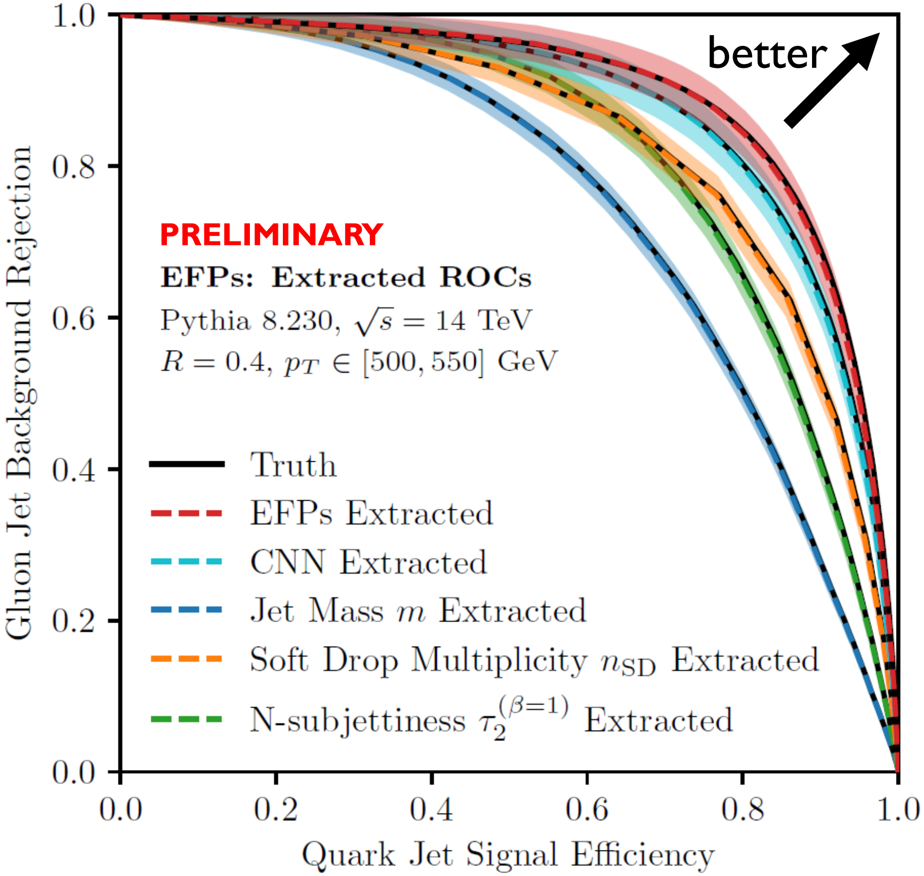


(Self-)calibrating quark and gluon classifiers

The extracted quark and gluon fractions can calibrate quark/gluon classifiers and evaluate tagging performance.

Even the classifier that was used to extract the fractions in the first place!

Note: To compare classifiers, one can just use the performance on A vs B directly.



Looking ahead

How does "sample dependence" manifest in this language?

Pairs of samples define quark and gluon. Different pairs of samples yield different flavor definitions.

Comparing definitions from different samples (dijets, Z+jet, gamma+jet, …) in data could probe how universal quark and gluon are. Can grooming improve this?

There are ways to quantify how "explainable" a new sample C is by quark and gluon:

$$\max(f^q + f^g) \quad \text{s.t.} \quad p_C(\boldsymbol{x}) = f^q\, p_q(\boldsymbol{x}) + f^g\, p_g(\boldsymbol{x}) + (1 - f^q - f^g)\, p_{\text{other}}(\boldsymbol{x})$$

Beyond quarks and gluons?

Multi-sample & multi-category generalizations of these ideas exist (though become more complicated).

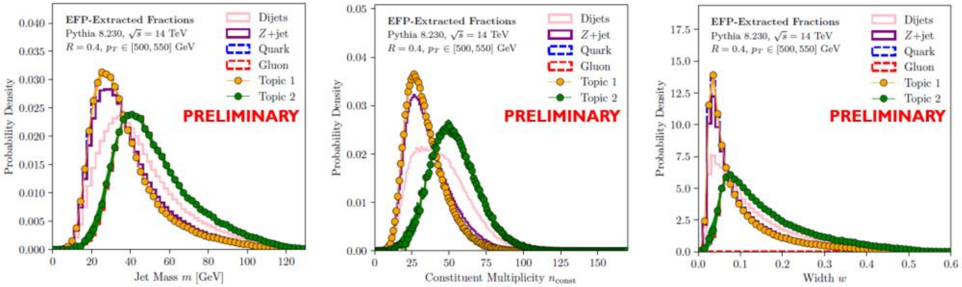
These ideas may be useful for other boosted hadronic objects as well.

Summary

- An operational definition of quark and gluon jets defined directly in terms of hadronic cross sections:

$$\boxed{p_{\text{quark}}(\boldsymbol{x}) \equiv \frac{p_A(\boldsymbol{x}) - \kappa_{\text{AB}}\, p_B(\boldsymbol{x})}{1-\kappa_{\text{AB}}} \qquad p_{\text{gluon}}(\boldsymbol{x}) \equiv \frac{p_B(\boldsymbol{x}) - \kappa_{\text{BA}}\, p_A(\boldsymbol{x})}{1-\kappa_{\text{BA}}}}$$

- Allows quark and gluon jet distributions to be measured separately without fraction or template inputs:

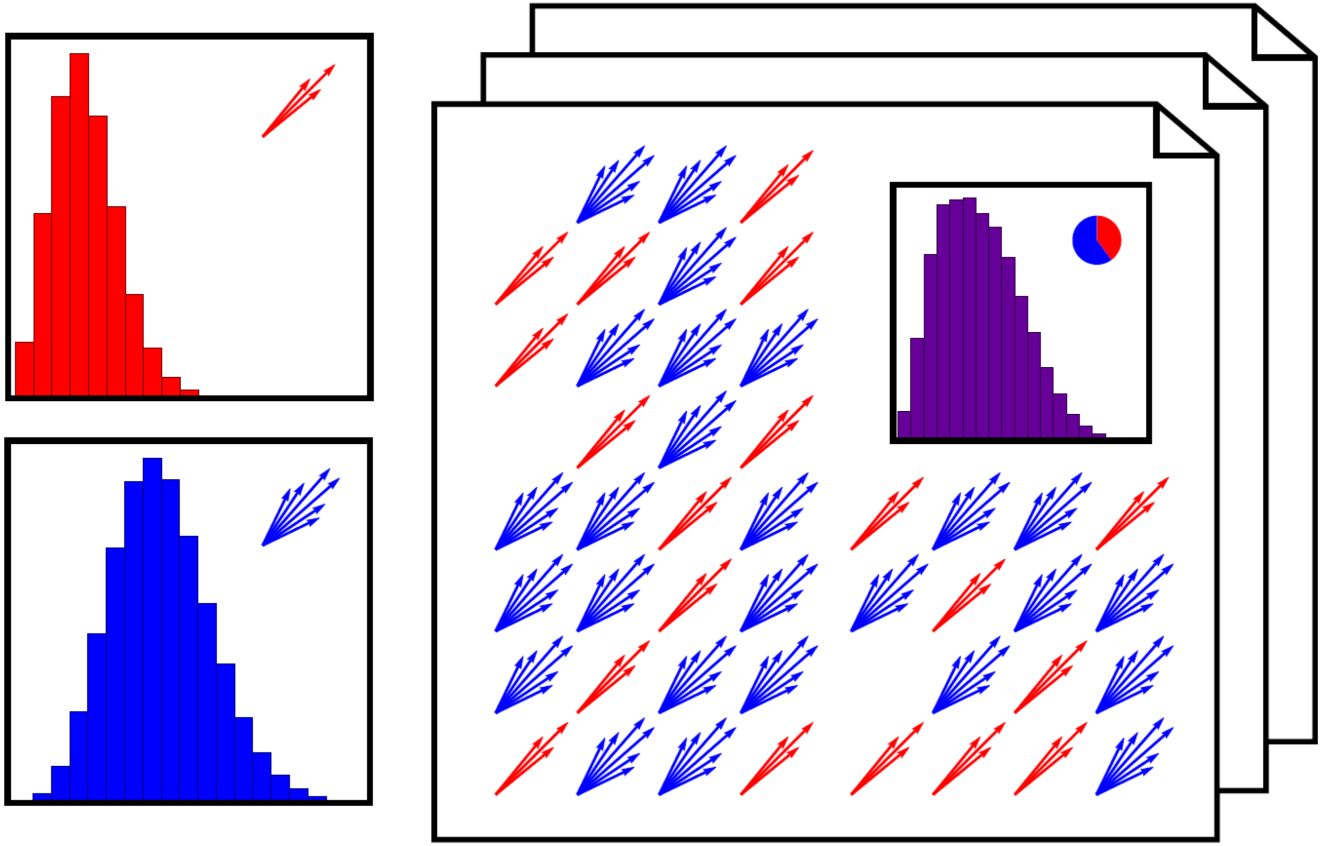


- Provide a firm foundation for data-driven techniques
  - Template methods, classification without labels, etc.

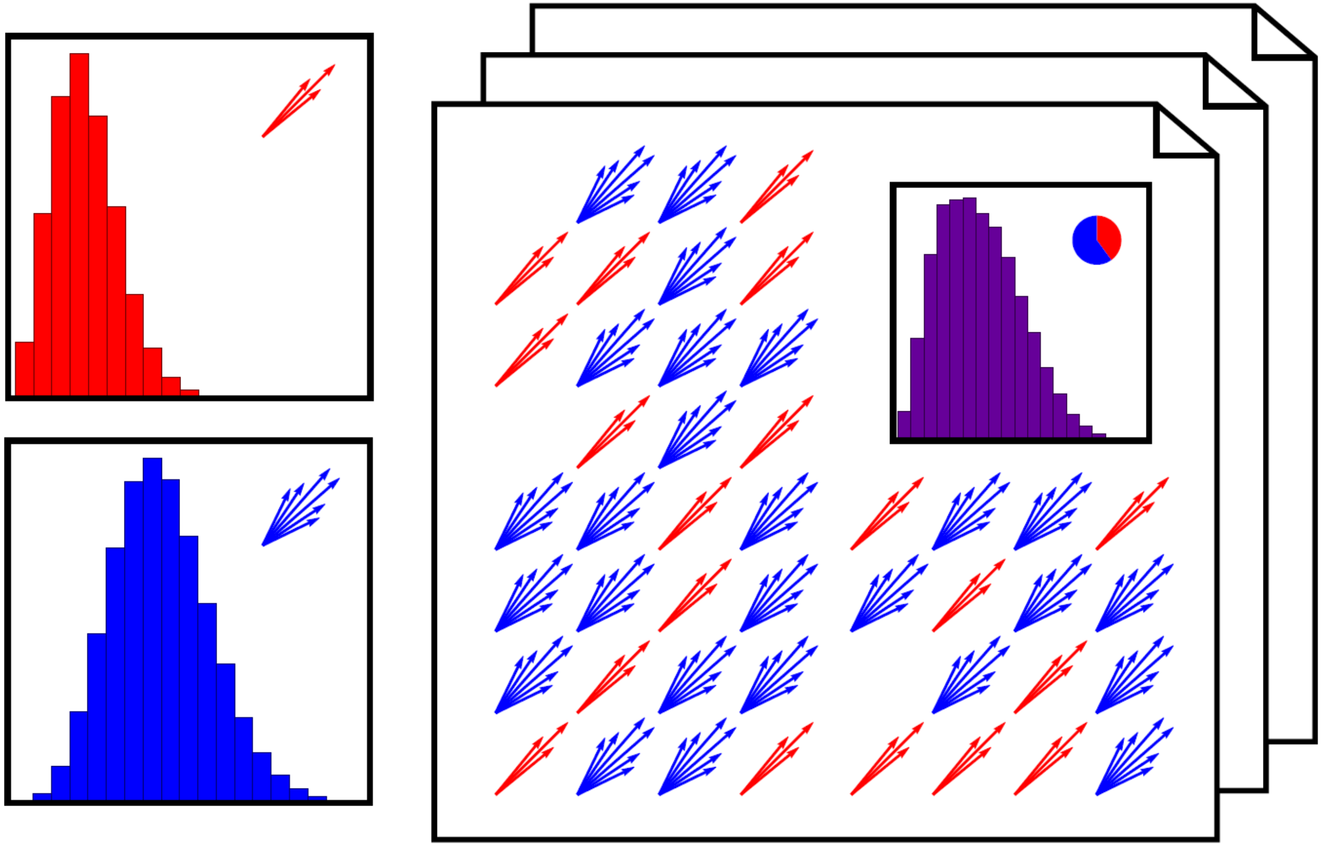
- Potential to probe questions of sample dependence in data

The End

Thank you!



Extra Slides



Jet topics from QCD: Casimir scaling

Jet mass (and many substructure observables) exhibits Casimir scaling at Leading Logarithmic accuracy:

$C_F = \frac{4}{3}$ for quarks

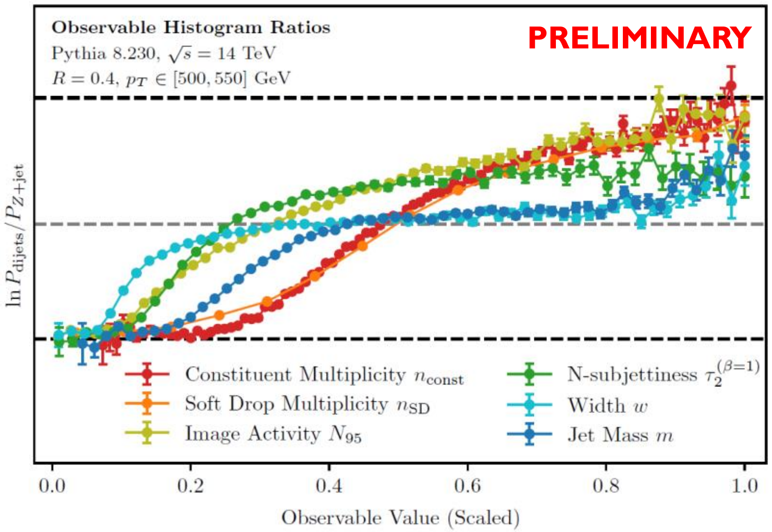
$C_A = 3$ for gluons

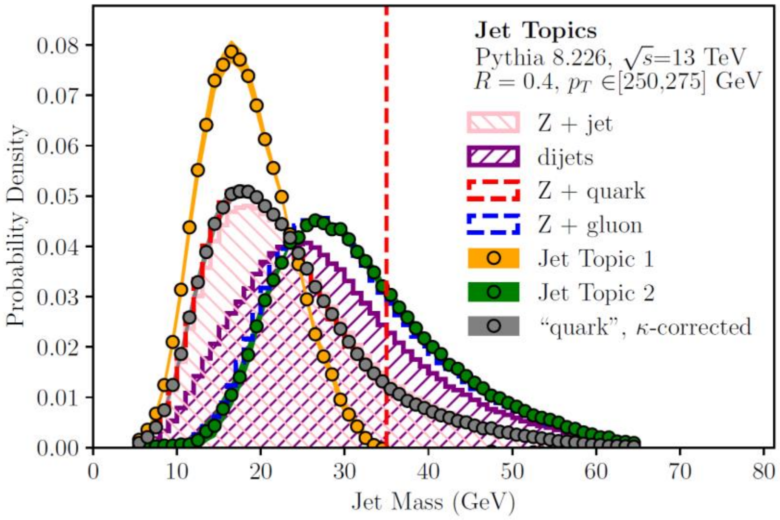
$$\Sigma_g(m) = \Sigma_q(m)^{\frac{C_A}{C_F}}$$

The quark/gluon reducibility factors at LL for any Casimir scaling observable are:

$$\kappa_{gq} = \min_m \frac{p_g(m)}{p_q(m)} = \min_m \frac{\Sigma_g{}'(m)}{\Sigma_q{}'(m)} = \frac{C_A}{C_F} \min_m \Sigma_q'(m)^{\frac{C_A}{C_F}-1} = 0$$

$$\kappa_{qg} = \min_m \frac{p_q(m)}{p_g(m)} = \min_m \frac{\Sigma_q{}'(m)}{\Sigma_g{}'(m)} = \frac{C_F}{C_A} \min_m \Sigma_q'(m)^{1-\frac{C_A}{C_F}} = \frac{C_F}{C_A} = \frac{4}{9}$$





Jet topics from QCD: Poisson scaling

Soft Drop Multiplicity (and other count observables) exhibits Poisson scaling at Leading Logarithmic accuracy:

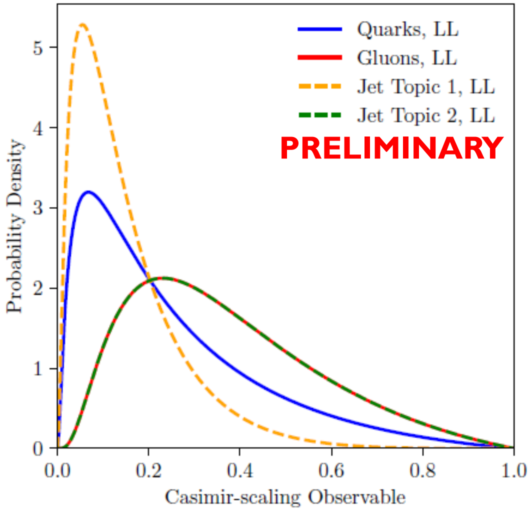
$C_F = \frac{4}{3}$ for quarks

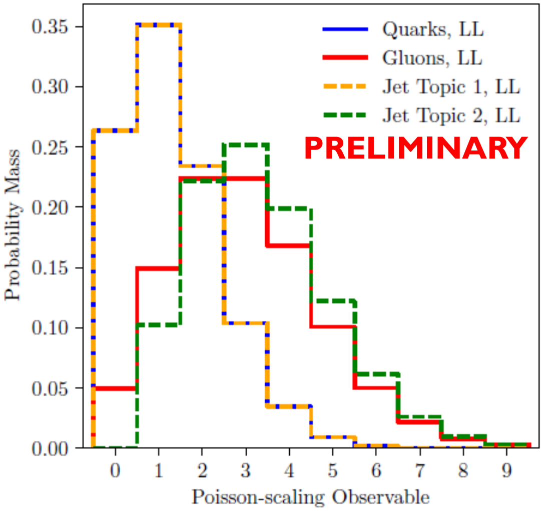
$C_A = 3$ for gluons

$$p_q(n) = \mathrm{Pois}(n; C_F\lambda), \qquad p_g(n) = \mathrm{Pois}(n; C_A\lambda).$$

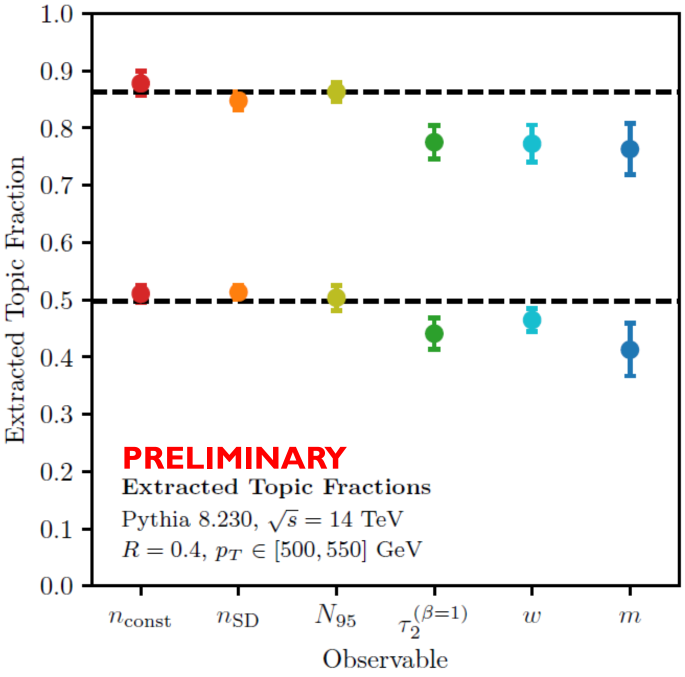
The quark/gluon reducibility factors at LL for any Poisson scaling observable are:

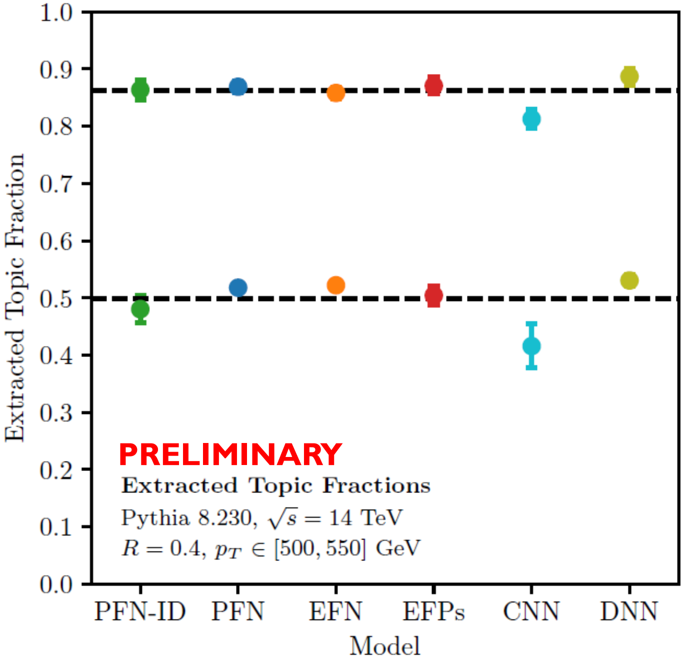
$$\boxed{\begin{aligned}
\kappa_{gq} &= \min_n \frac{p_g(n)}{p_q(n)} = \min_n \frac{(C_A\lambda)^n \; e^{-C_A\lambda}}{(C_F\lambda)^n \; e^{-C_F\lambda}} = e^{\lambda(C_F - C_A)} \min_n \left(\frac{C_A}{C_F}\right)^n = e^{\lambda(C_F - C_A)} \\
\kappa_{qg} &= \min_n \frac{p_q(n)}{p_g(n)} = \min_n \frac{(C_F\lambda)^n \; e^{-C_F\lambda}}{(C_A\lambda)^n \; e^{-C_A\lambda}} = e^{\lambda(C_A - C_F)} \min_n \left(\frac{C_F}{C_A}\right)^n = 0
\end{aligned}}$$





Extracting quark and gluon fractions





From the reducibility factors, the quark and gluon fractions of the samples can be obtained.

MC-labeled sample dependence in Pythia