Towards Data-Driven Particle Physics Classifiers

Deep Learning in the Natural Sciences

University of Hamburg

Eric M. Metodiev

Center for Theoretical Physics

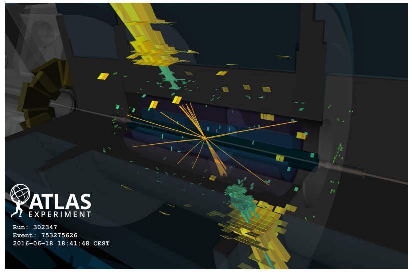
Massachusetts Institute of Technology

Based on work with Patrick Komiske, Benjamin Nachman, Matthew Schwartz, and Jesse Thaler

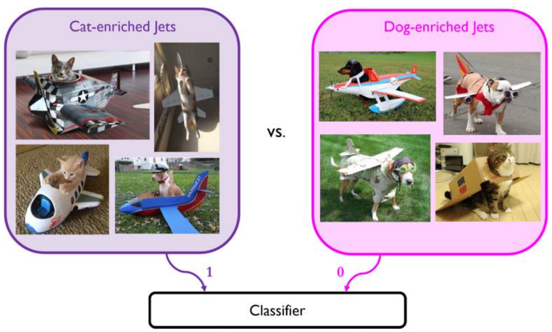
[1708.02949] [1801.10158] [1802.00008] [1809.01140]

March 1, 2019

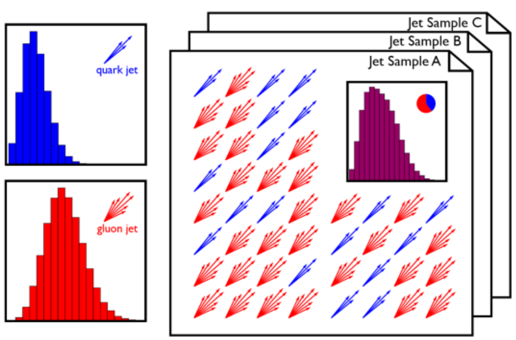
Outline



Classification at Colliders

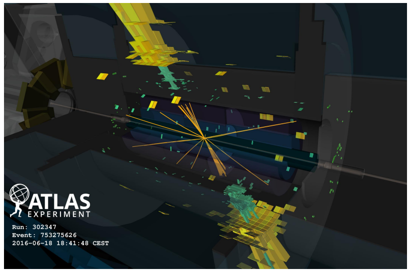


Training on Data



Disentangling Categories

Outline

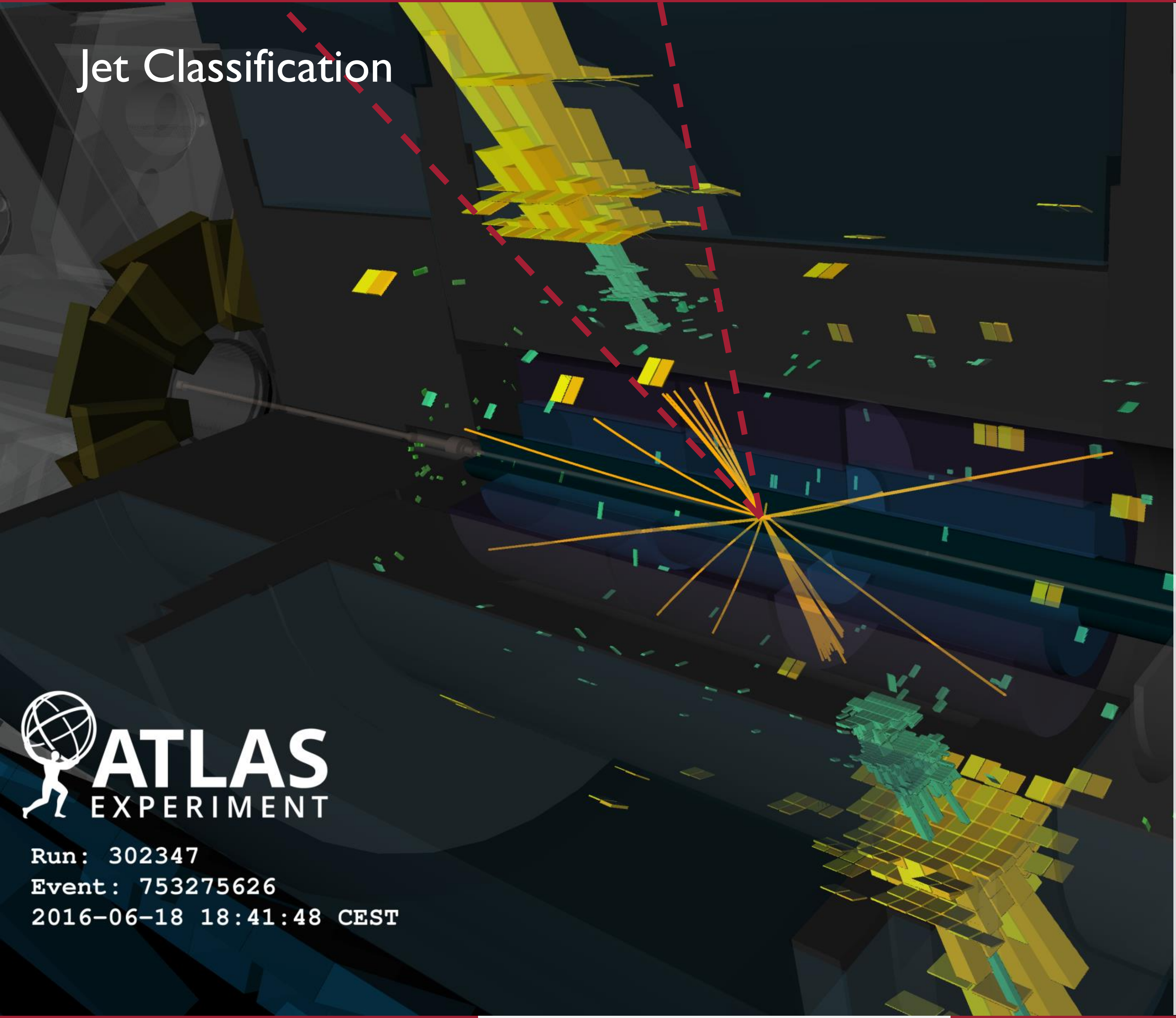


Classification at Colliders

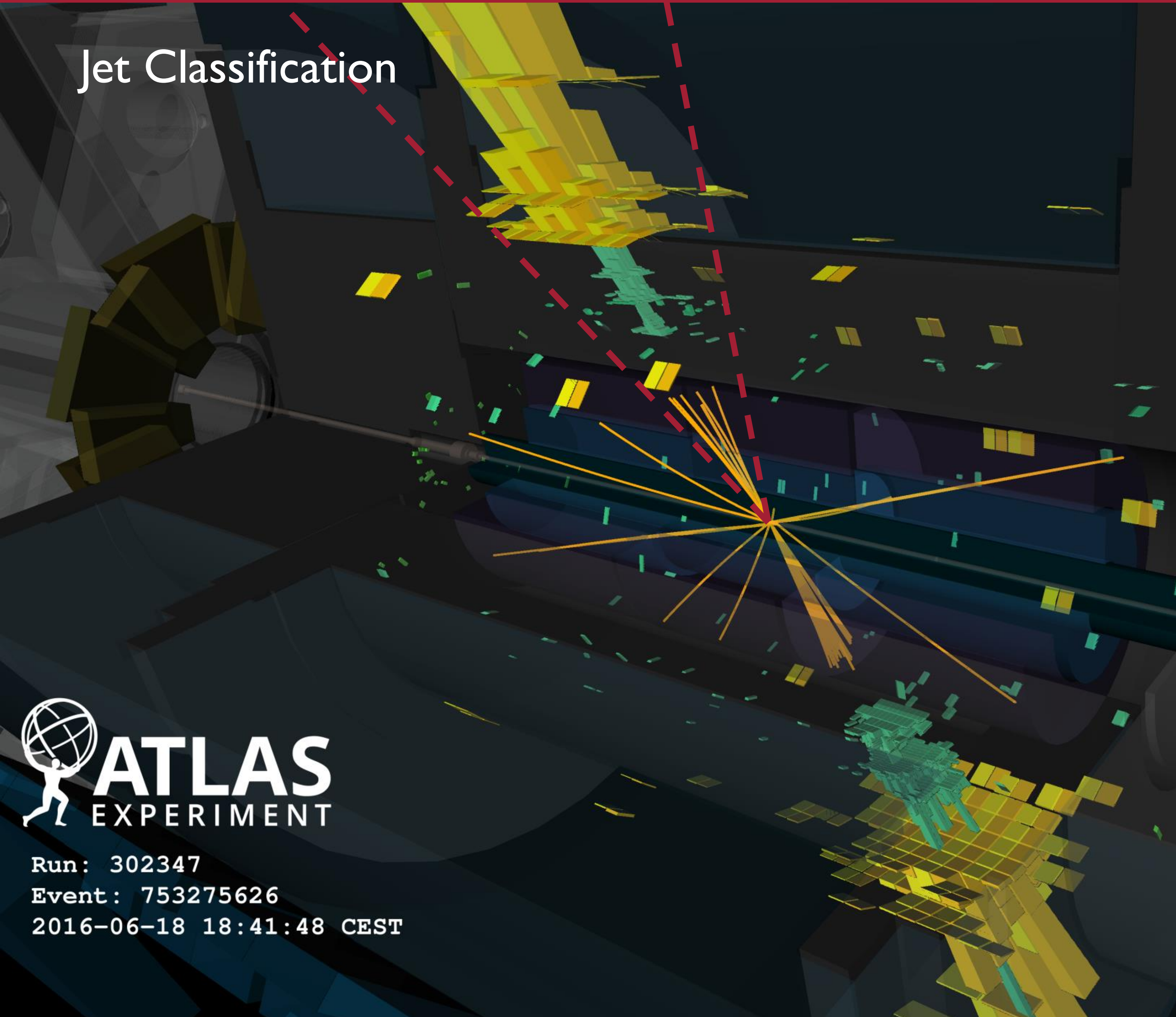
Training on Data

Disentangling Categories

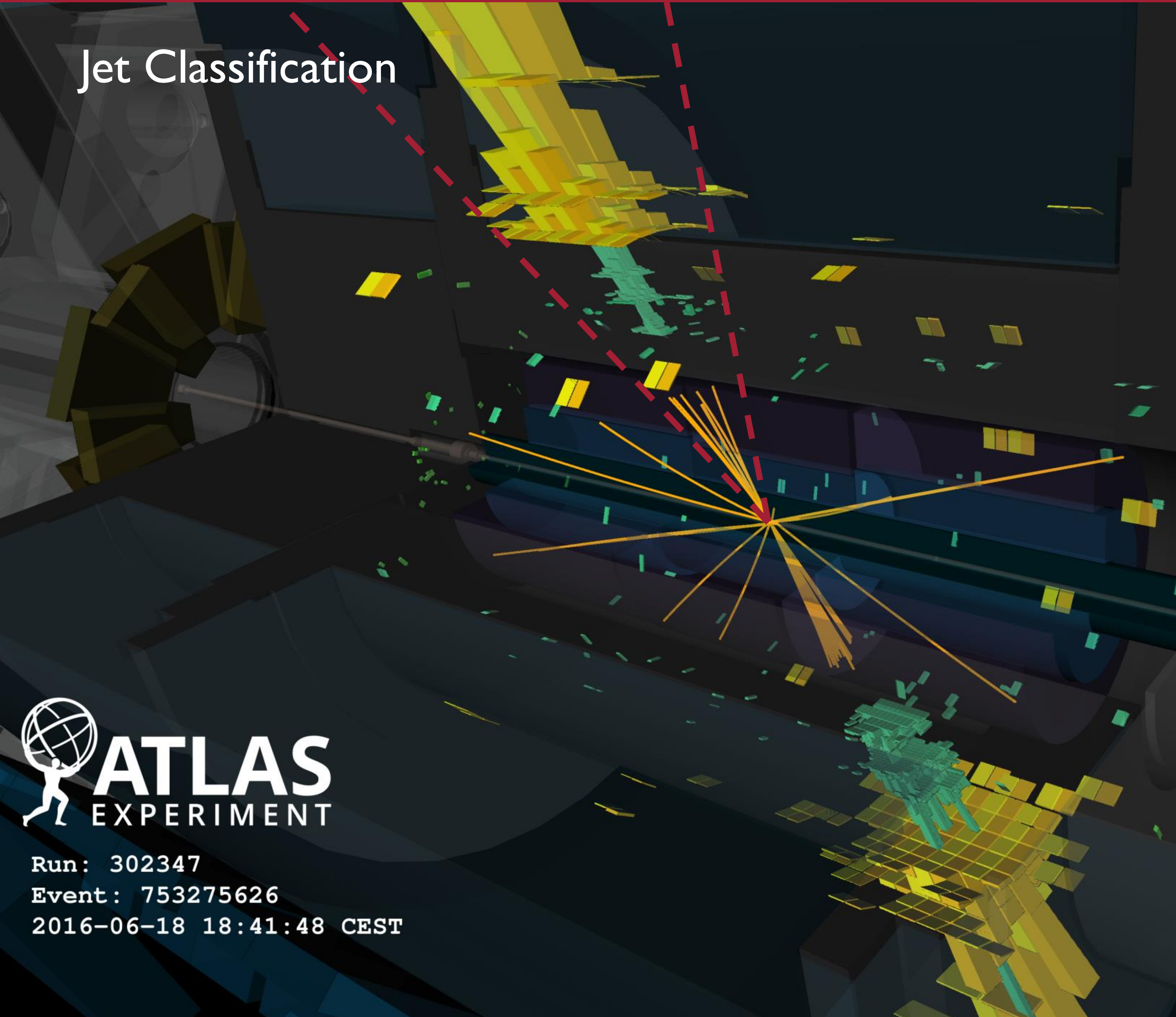
Jet Classification



Jet Classification



Jet Classification



quark

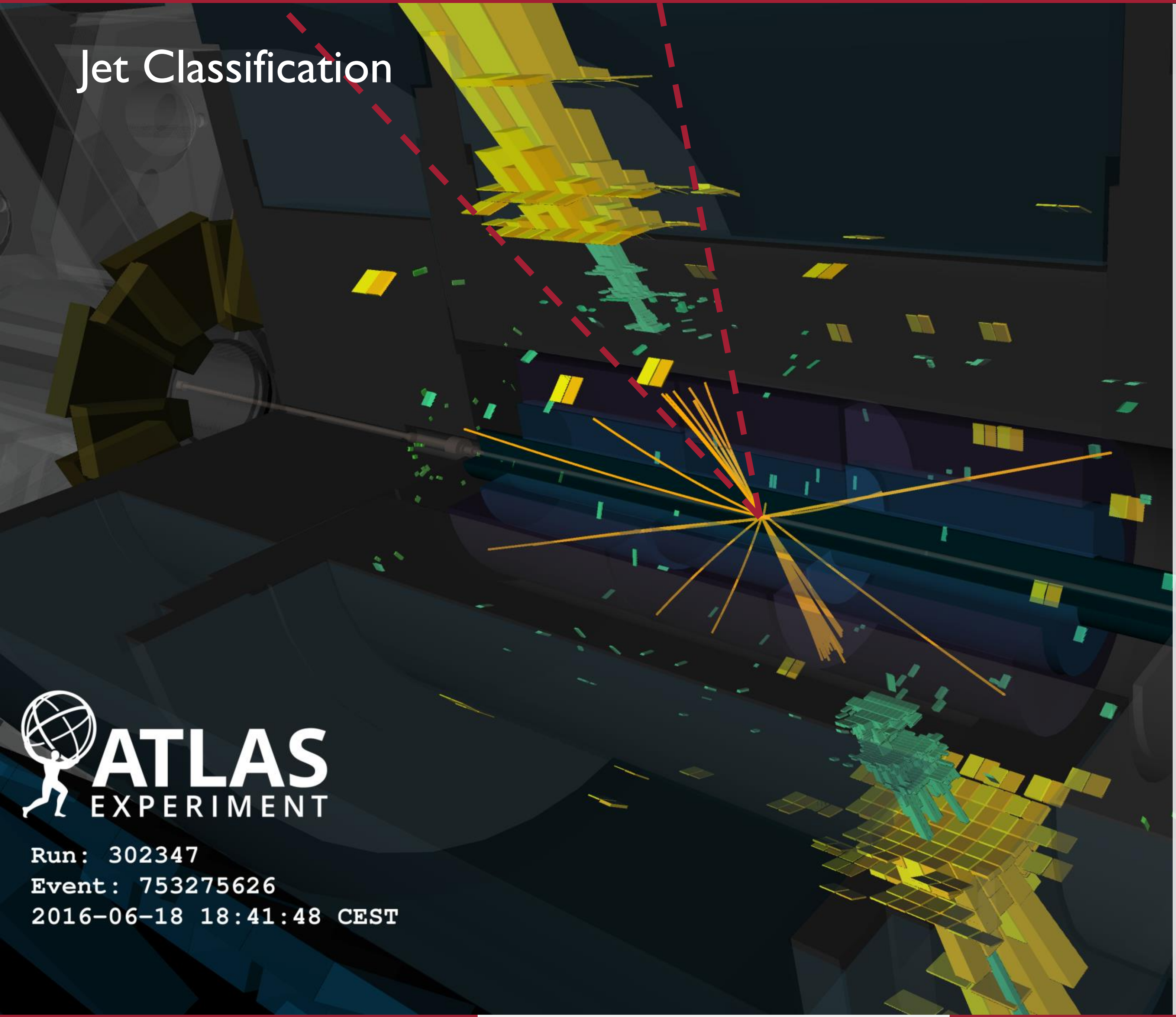
gluon

§ ——— §

New Physics
signal
⬇
quark jets

QCD
background
⬇
gluon jets

Jet Classification



quark

gluon

New Physics **signal**

⬇

quark jets

QCD **background**

⬇

gluon jets

$C_q = 4/3$

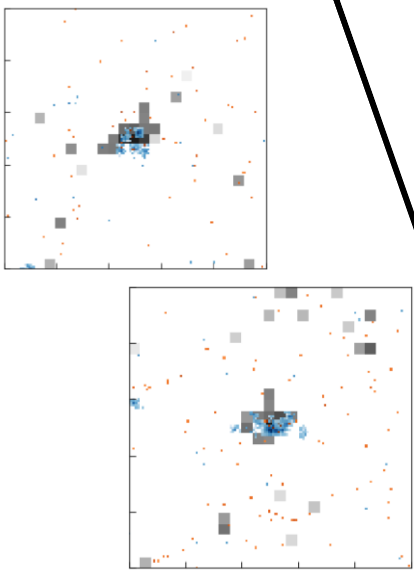
$C_g = 3$

⬇

gluon jets are "twice as wide" as quark jets

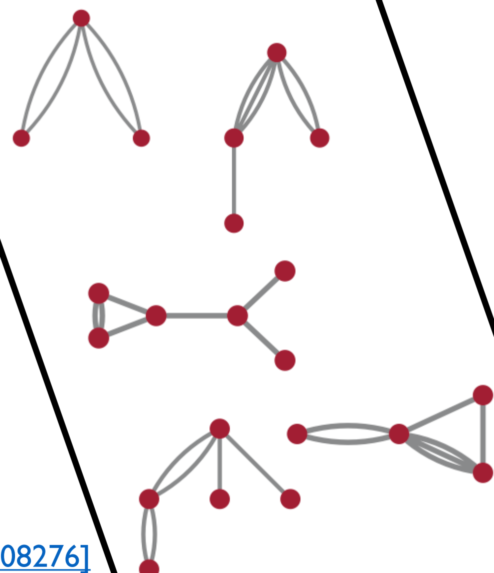
Machine Learning with Jets

Images



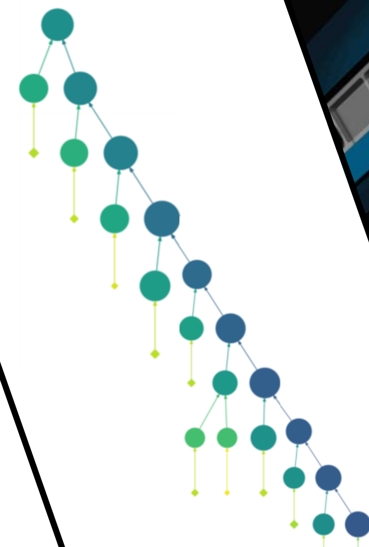
e.g. [M. Andrews, *et al.*, 1902.08276]
[L. de Oliveira, et al., 1511.05190]
[P.T. Komiske, EMM, M.D. Schwartz, 1612.01551]

Observables



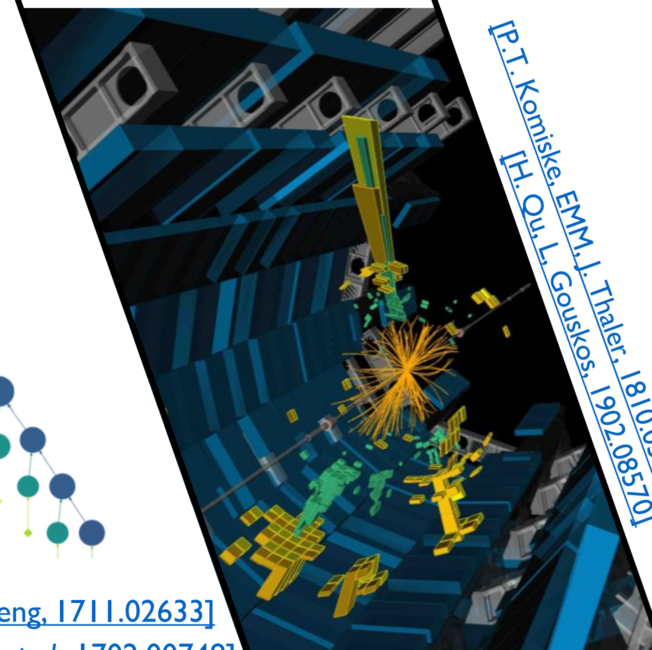
e.g.
[K. Datta, A. Larkoski, 1704.08249]
[P.T. Komiske, EMM, J. Thaler, 1712.07124]

Sequences



e.g. [T. Cheng, 1711.02633]
[G. Louppe, *et al.*, 1702.07748]
[G. Kasieczka, N. Kiefer, T. Plehn, J. Thompson, 1812.09223]

Point Clouds



…

[P.T. Komiske, EMM, J. Thaler, 1810.05165]
[H. Qu, L. Gouskos, 1902.08570]

All supervised classification methods require training data.

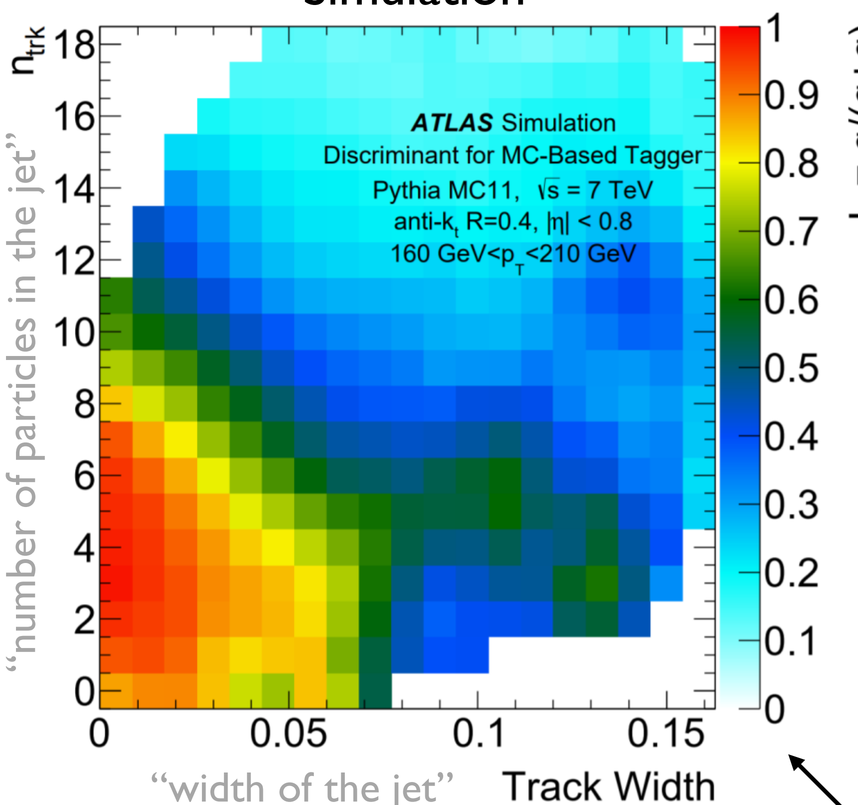
Impossible to isolate pure samples of quark jets and gluon jets.

Often rely on simulation, which is sensitive to mismodeling.

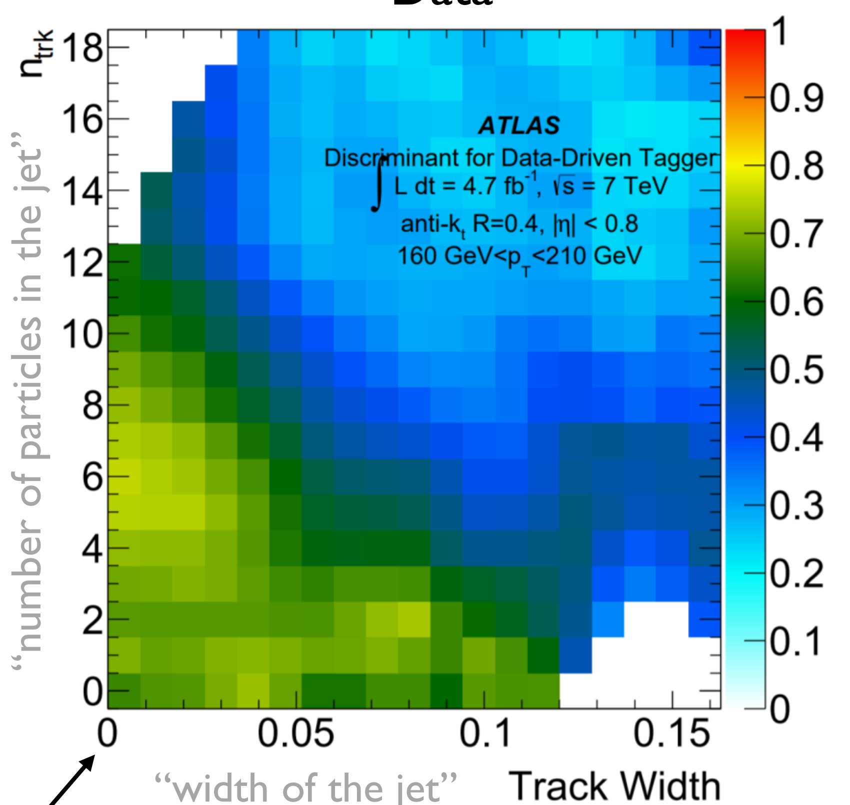
Simulation vs. Data

Simple two-feature quark vs. gluon jet classifier using **simulation** and **data**.

Simulation



Data

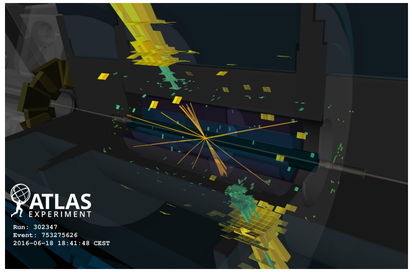


Very different!

[ATLAS Collaboration, 1405.6583]

Is it possible to train classifiers on data?

Outline



Classification at Colliders

Classifying jets based on their originating particles.

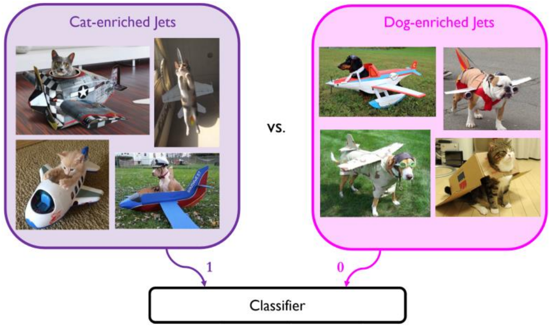
Training on Data

Disentangling Categories

Outline

Classification at Colliders

Classifying jets based on their originating particles.

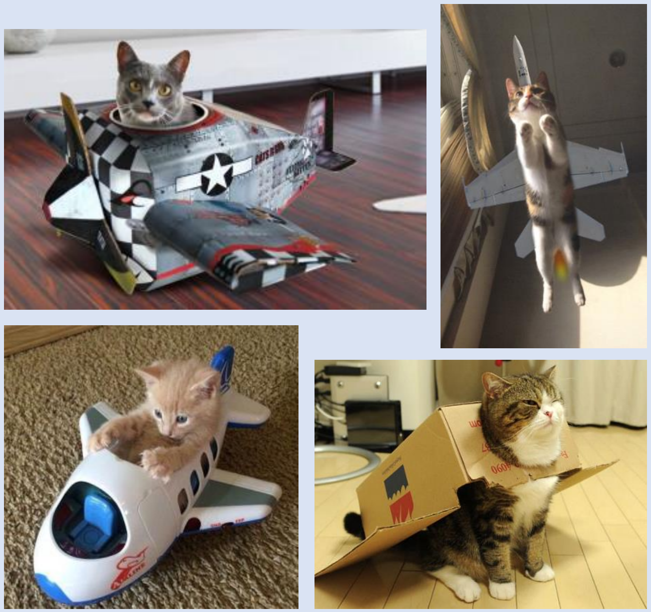


Training on Data

Disentangling Categories

Training on pure samples: Cat vs. Dog jets

Cat Jets



vs.

Dog Jets

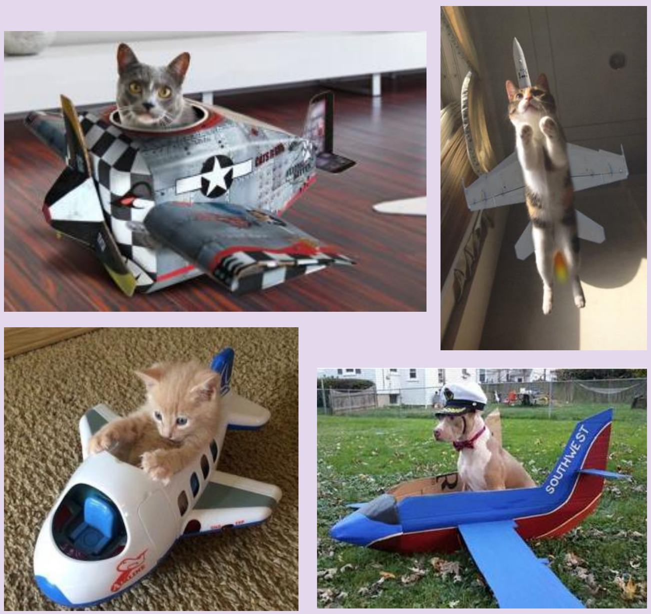
1

0

Classifier

Training on mixed samples: Cat vs. Dog jets

Cat-enriched Jets



vs.

Dog-enriched Jets

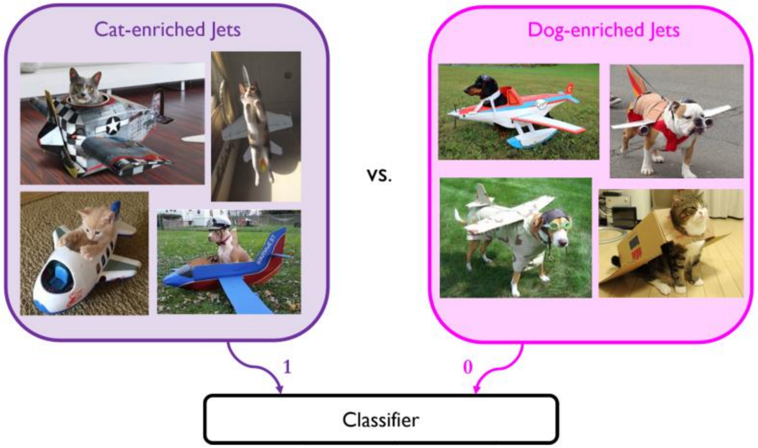
1

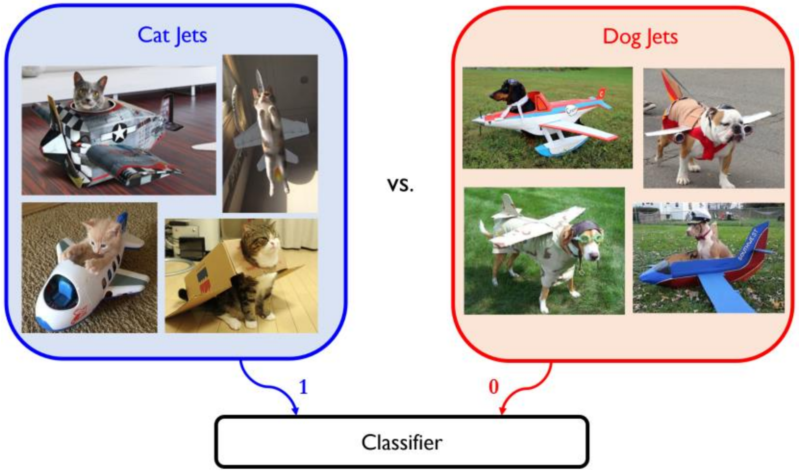
0

Classifier

This defines an equivalent classifier to the pure case!

Classification without labels (CWoLa)





$$\underbrace{L_{\frac{M_1}{M_2}}(x)} = \frac{p_{M_1}(\boldsymbol{x})}{p_{M_2}(\boldsymbol{x})} = \frac{f_1^{\mathrm{cat}}\, p_{\mathrm{cat}}(\boldsymbol{x}) + (1 - f_1^{\mathrm{cat}})\, p_{\mathrm{dog}}(\boldsymbol{x})}{f_2^{\mathrm{cat}}\, p_{\mathrm{cat}}(\boldsymbol{x}) + (1 - f_2^{\mathrm{cat}})\, p_{\mathrm{dog}}(\boldsymbol{x})} = \frac{f_1^{\mathrm{cat}}\, L_{\frac{\mathrm{cat}}{\mathrm{dog}}}(\boldsymbol{x}) + (1 - f_1^{\mathrm{cat}})}{f_2^{\mathrm{cat}}\, \underbrace{L_{\frac{\mathrm{cat}}{\mathrm{dog}}}}(\boldsymbol{x}) + (1 - f_2^{\mathrm{cat}})}$$

Optimal mixed sample classifier

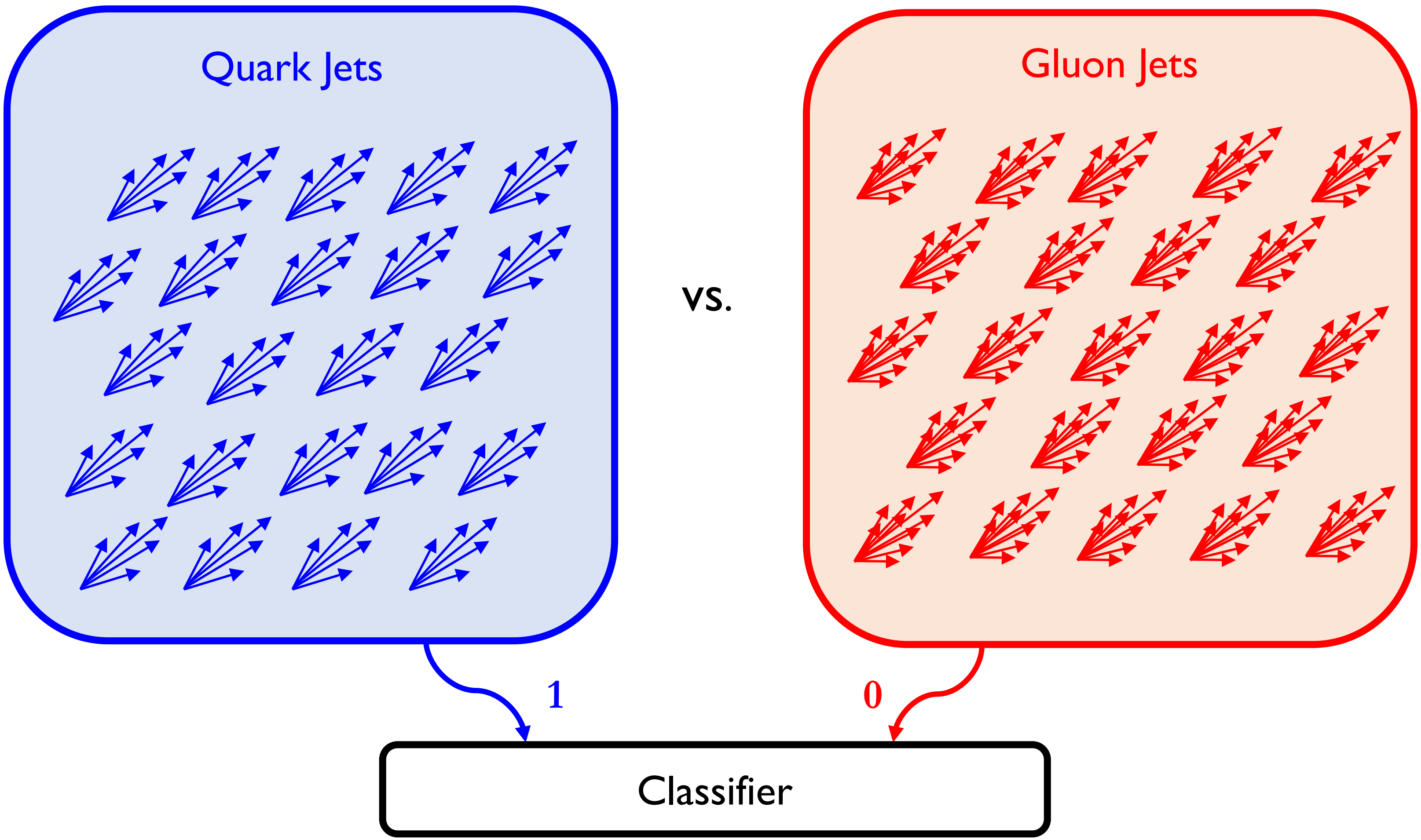
is a monotonic rescaling of

Optimal cat vs. dog classifier

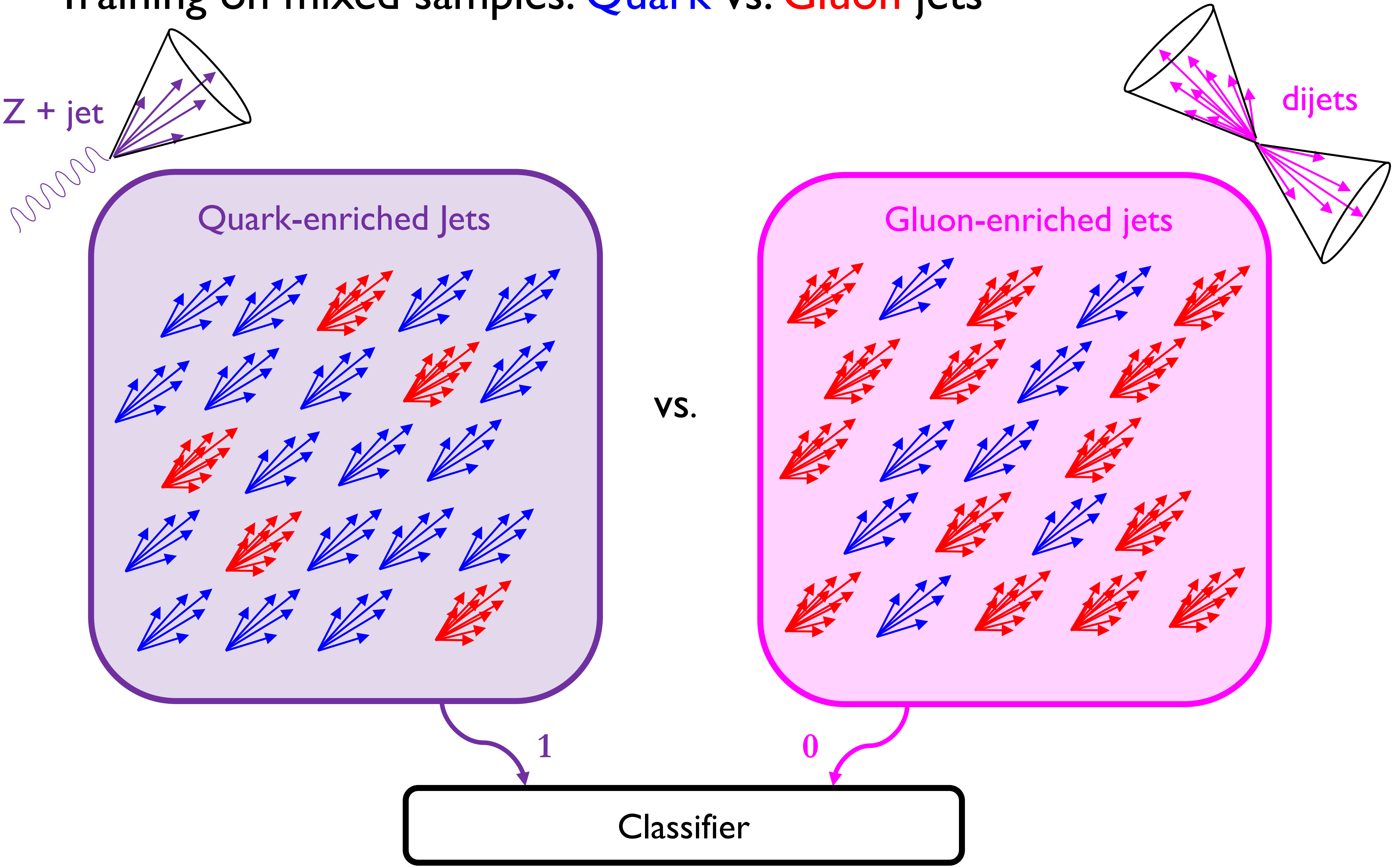
Hence they define equivalent classifiers. □

[EMM, B. Nachman, J. Thaler, 1708.02949] [P.T. Komiske, EMM, B. Nachman, M.D. Schwartz, 1801.10158]
see also [L. Dery, B. Nachman, F. Rubbo, A. Schwartzman, 1702.00414] [T. Cohen, M. Freytsis, B. Ostdiek, 1706.09451]

Training on pure samples: Quark vs. Gluon jets



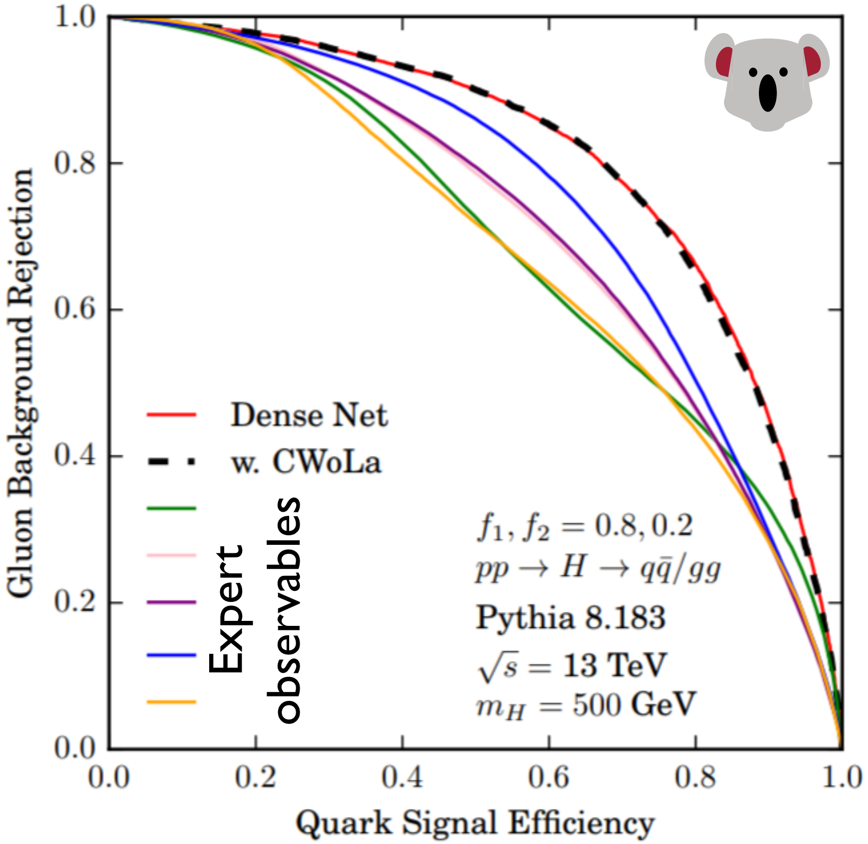
Training on mixed samples: Quark vs. Gluon jets



This defines an equivalent classifier to the pure case!

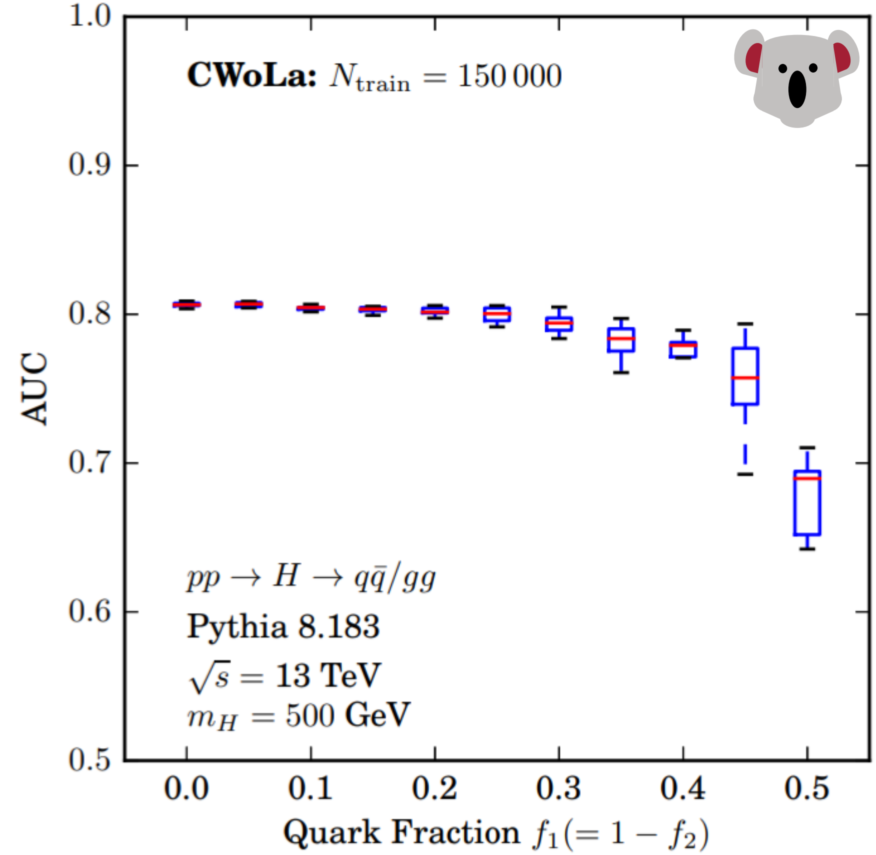
Performance

Compare 80-20% mixtures to pure samples



Can train on mixed samples!

Vary the mixture purity



Works for very impure mixtures!

[EMM, B. Nachman, J. Thaler, 1708.02949]

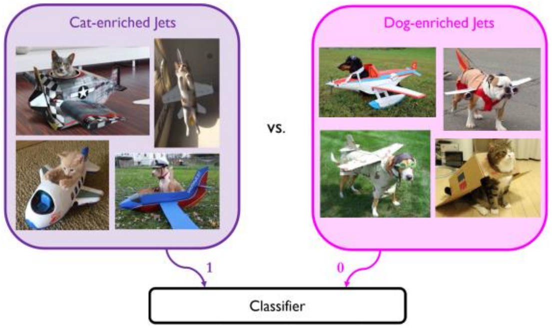
Also works for convolutional neural networks and jet images.

[P.T. Komiske, EMM, B. Nachman, M.D. Schwartz, 1801.10158]

Outline

Classification at Colliders

Classifying jets based on their originating particles.



Training on Data

Weak supervision with mixed jet samples.

Disentangling Categories

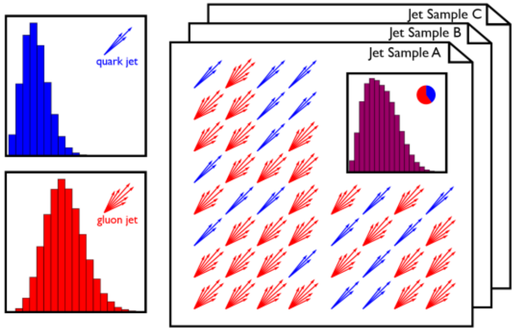
Outline

Classification at Colliders

Classifying jets based on their originating particles.

Training on Data

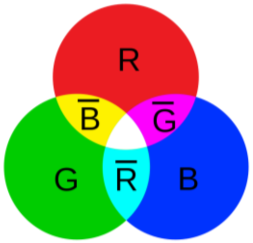
Weak supervision with mixed jet samples.



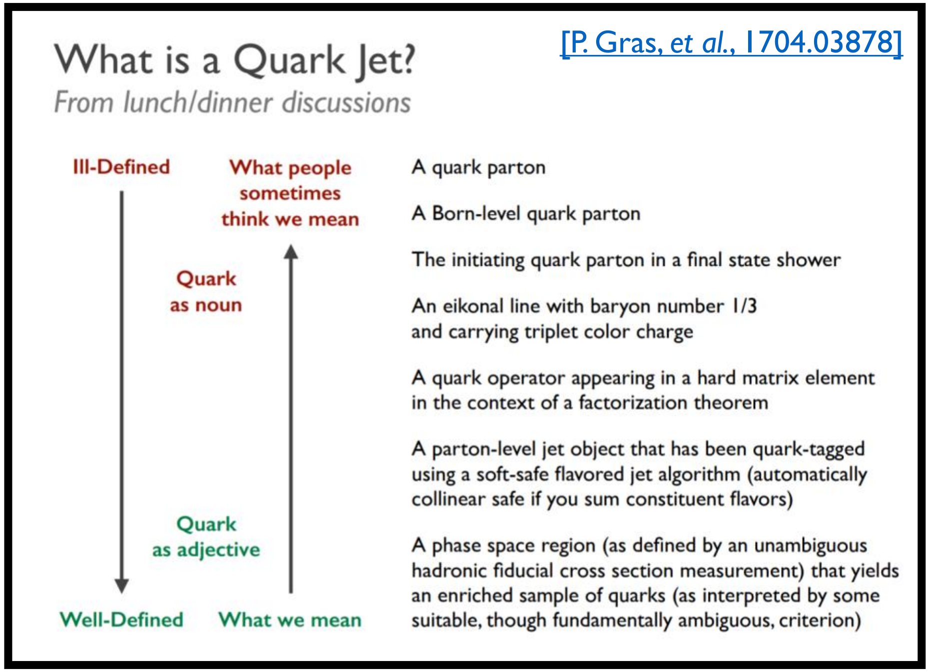
Disentangling Categories

What do we even mean by **quark** and **gluon** jets?

Quarks are color triplets. Gluons are color octets. Hadrons in jets are color singlets.



No unambiguous definition of quark and gluon jets.


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

What is a Quark Jet?

From lunch/dinner discussions

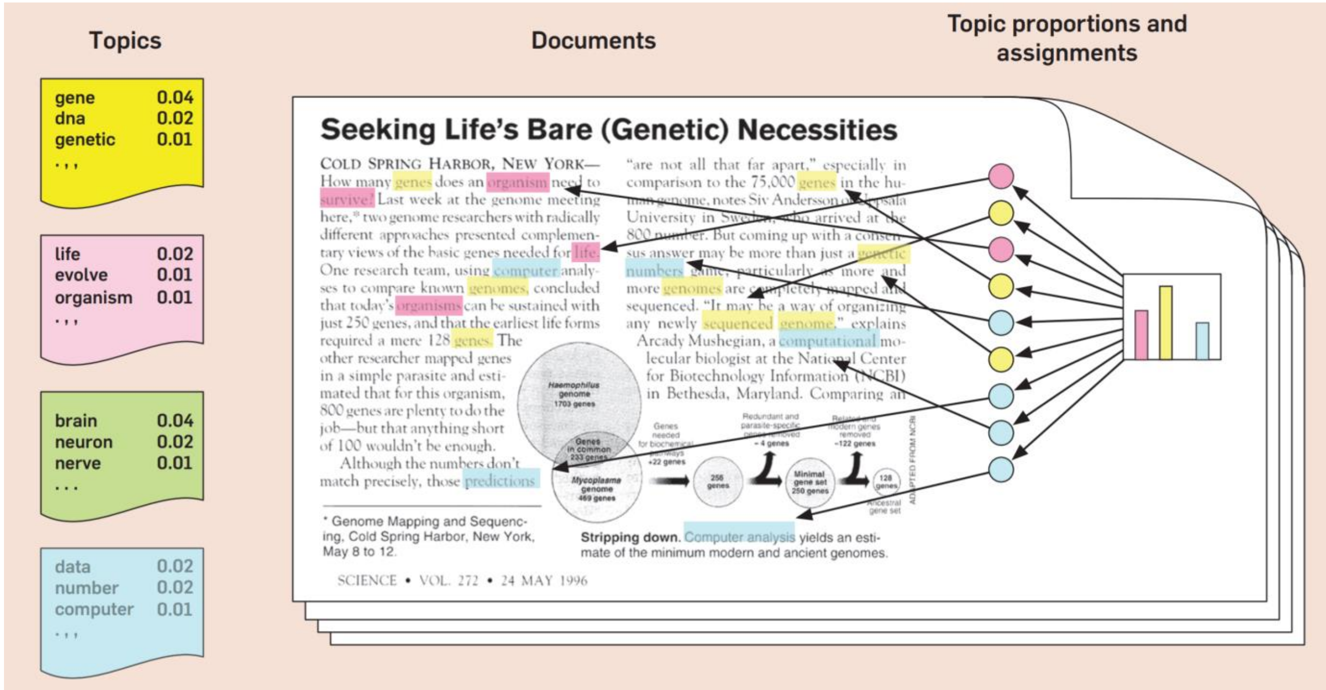
Ill-Defined	What people sometimes think we mean	
Quark as noun		A quark parton
		A Born-level quark parton
		The initiating quark parton in a final state shower
		An eikonal line with baryon number 1/3 and carrying triplet color charge
		A quark operator appearing in a hard matrix element in the context of a factorization theorem
		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)
Quark as adjective		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	What we mean	

Various definitions of increasing verbosity

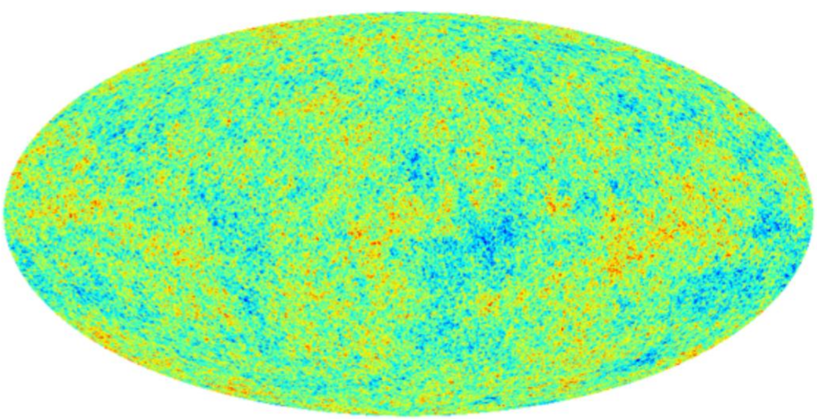
We obtained a quark vs. gluon jet classifier without a definition…

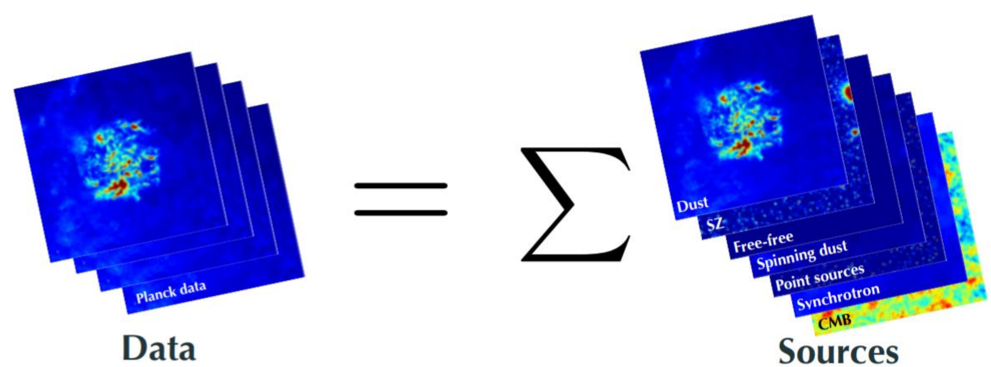
Operational data-driven definition of quark and gluon jets

Topic Modeling and Blind Source Separation



[Image: D. Blei]



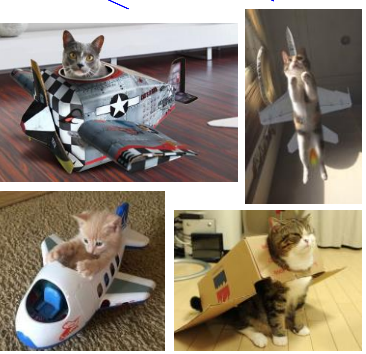


[Image: J. Bobin]

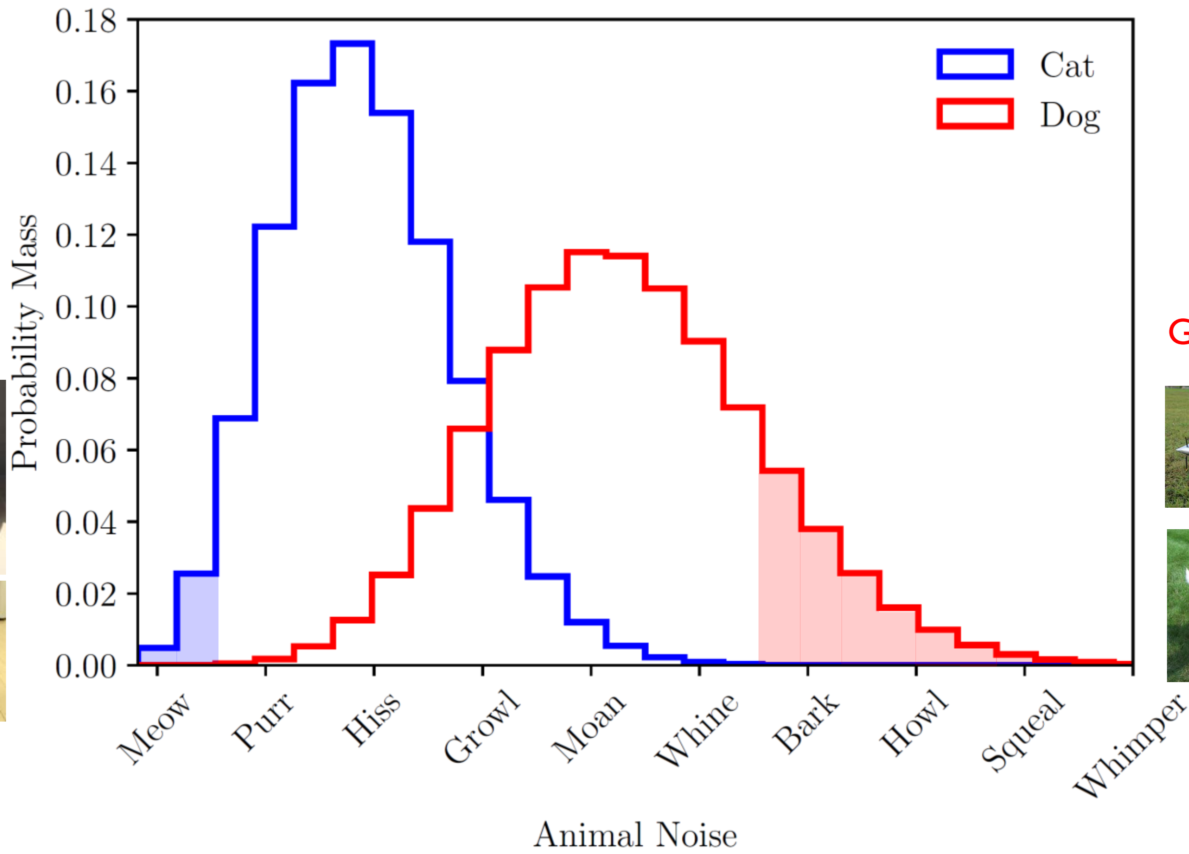
Disentangling Categories

Let's model cats and dogs as random animal noise producers.

Meow Growl



Purr



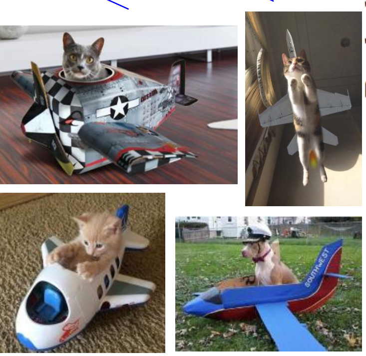
Growl Woof

Howl

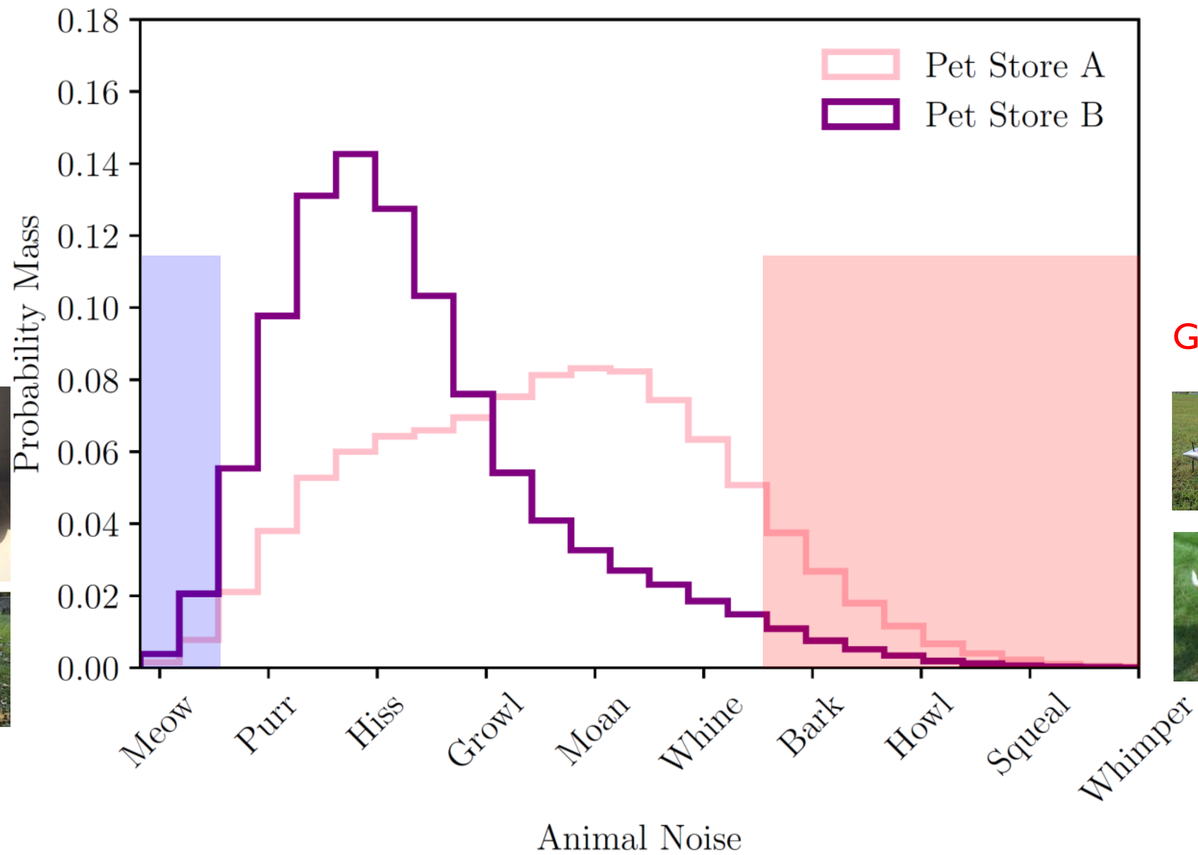
Disentangling Categories

Listen to the animal noises from two different pet stores.

Meow Growl



Howl



Growl Woof

Purr

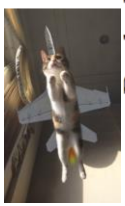
$$\frac{N^{\text{Store }A}_{\text{"Meow"}}}{N^{\text{Store }B}_{\text{"Meow"}}} = \frac{f^{\text{Store }A}_{\text{Cat}}}{f^{\text{Store }B}_{\text{Cat}}}$$

$$\frac{N^{\text{Store }A}_{\text{"Bark"}}}{N^{\text{Store }B}_{\text{"Bark"}}} = \frac{1 - f^{\text{Store }A}_{\text{Cat}}}{1 - f^{\text{Store }B}_{\text{Cat}}}$$

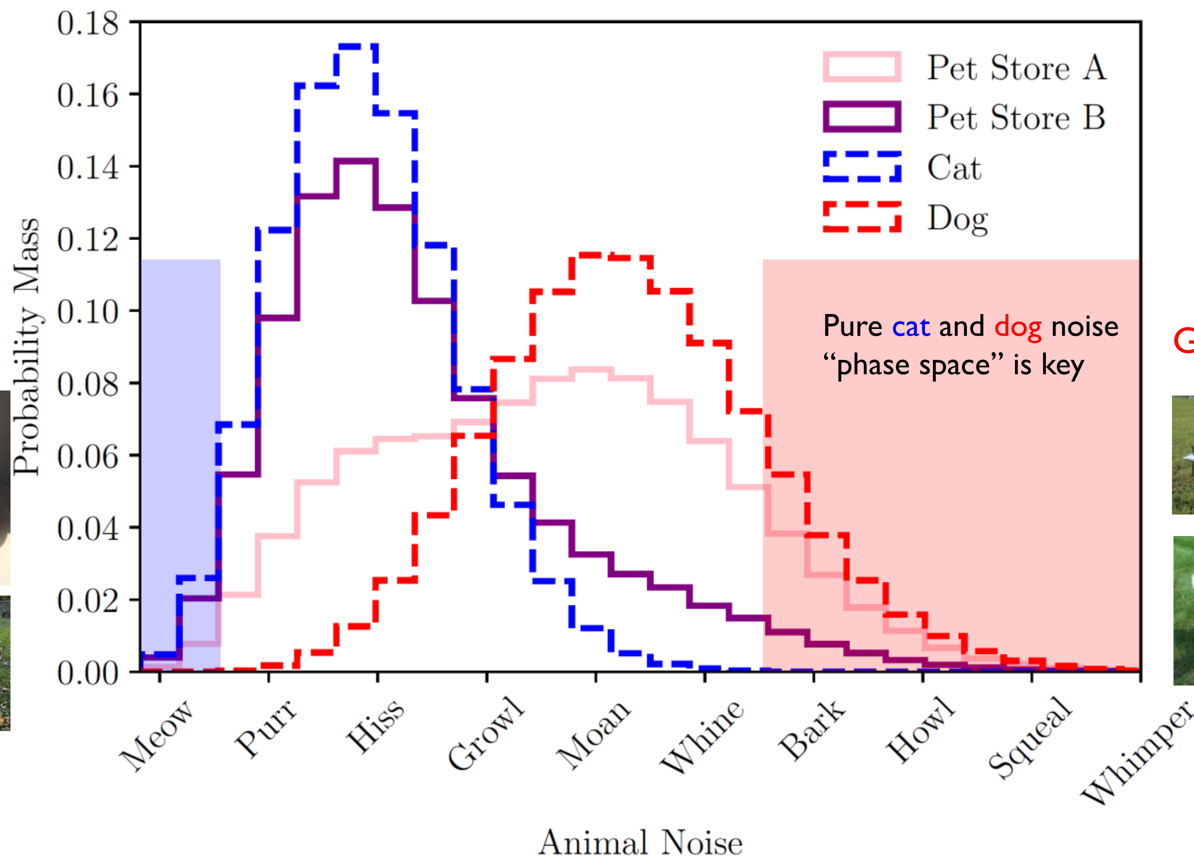
Disentangling Categories

Disentangle cat and dog vocabularies from the animal noises at pet stores.

Meow Growl



Howl



Growl Woof

Purr

$$\frac{N_{\text{"Meow"}}^{\text{Store }A}}{N_{\text{"Meow"}}^{\text{Store }B}} = \frac{f_{\text{Cat}}^{\text{Store }A}}{f_{\text{Cat}}^{\text{Store }B}}$$

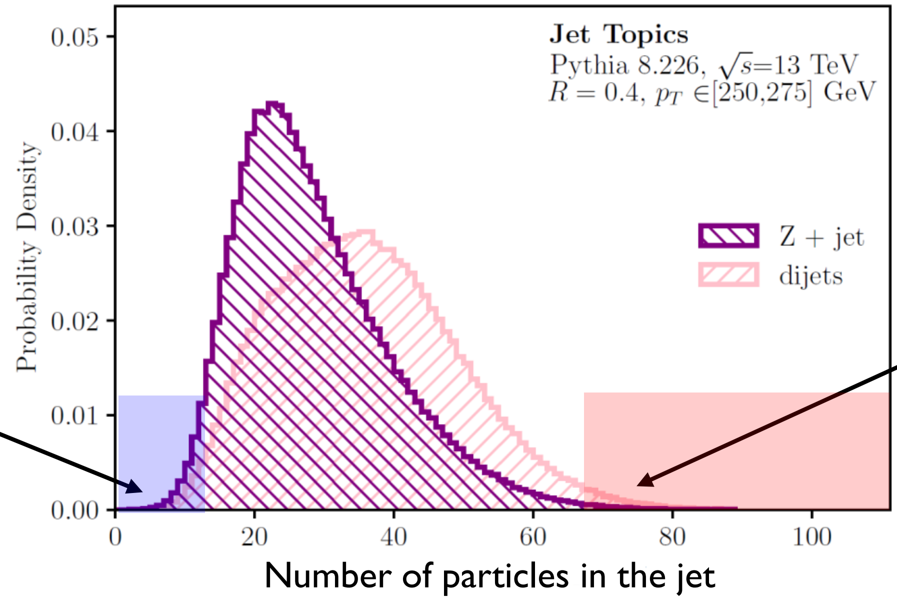
$$\frac{N_{\text{"Bark"}}^{\text{Store }A}}{N_{\text{"Bark"}}^{\text{Store }B}} = \frac{1 - f_{\text{Cat}}^{\text{Store }A}}{1 - f_{\text{Cat}}^{\text{Store }B}}$$

Disentangling Categories

An operational definition of quark and gluon jets.

[EMM, J. Thaler, 1802.00008]
[P.T. Komiske, EMM, J. Thaler, 1809.01140]

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



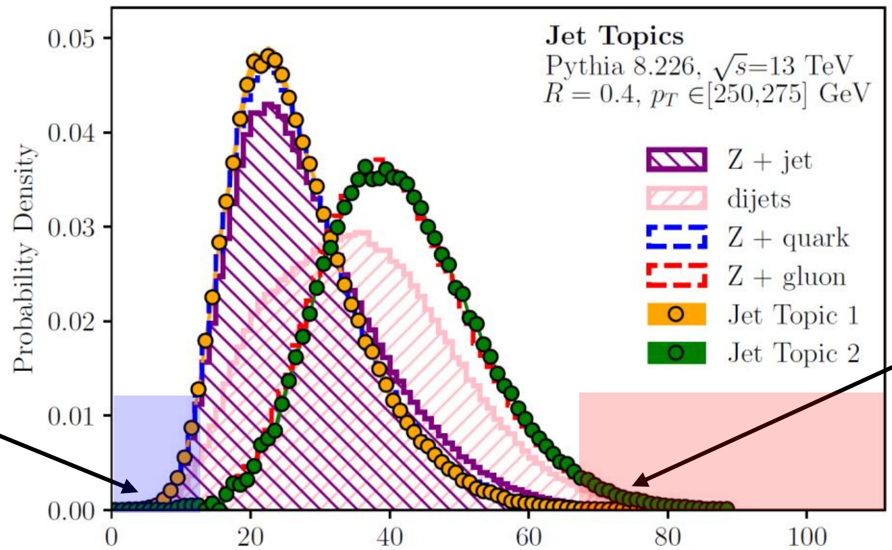
$$\kappa_{\mathrm{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

Disentangling Categories

An operational definition of quark and gluon jets.

[EMM, J. Thaler, 1802.00008]
[P.T. Komiske, EMM, J. Thaler, 1809.01140]

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



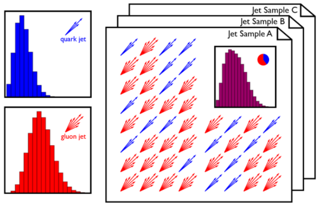
$$\kappa_{\mathrm{AB}} \equiv \min_{\boldsymbol{x}} \frac{p_A(\boldsymbol{x})}{p_B(\boldsymbol{x})} = \frac{1-f_A^q}{1-f_B^q}$$

With **reducibility factors** κ_{AB} and κ_{BA}, solve for the quark and gluon distributions:

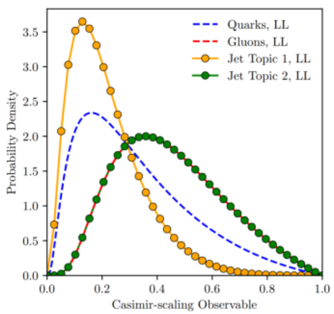
$$p_{\mathrm{quark}}(\boldsymbol{x}) = \frac{p_A(\boldsymbol{x}) - \kappa_{\mathrm{AB}}\, p_B(\boldsymbol{x})}{1-\kappa_{\mathrm{AB}}} \qquad p_{\mathrm{gluon}}(\boldsymbol{x}) = \frac{p_B(\boldsymbol{x}) - \kappa_{\mathrm{BA}}\, p_A(\boldsymbol{x})}{1-\kappa_{\mathrm{BA}}}$$

Can also use machine learning to determine the feature space.

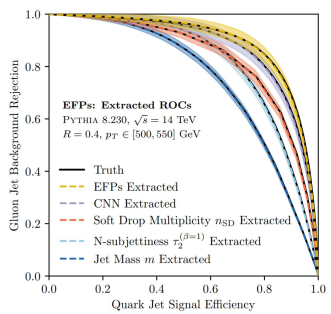
Collider data as mixtures of jet types



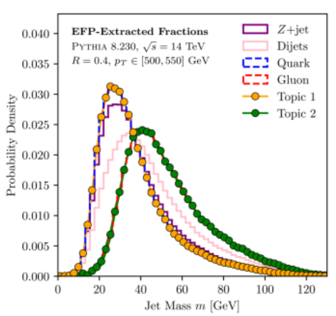
Theoretical and experimental definition of jet categories.



Theoretically tractable: calculate reducibility factors from perturbative QCD for certain observables.



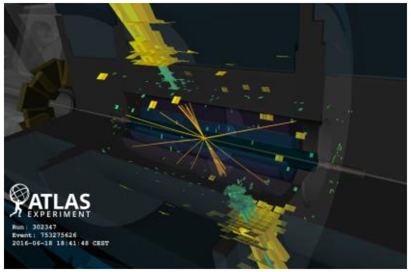
Can use the fractions to calibrate ROC curves.



Allows for any observable distributions to be extracted for quark and gluon jets separately.

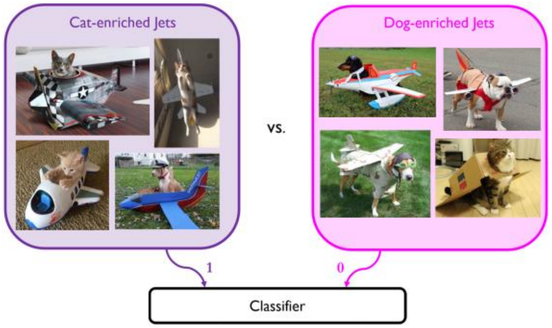
See extra slides for more.

Summary



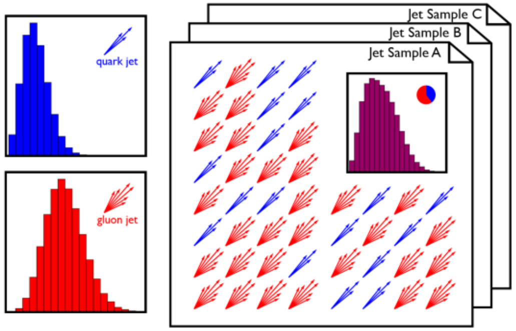
Classification at Colliders

Classifying jets based on their originating particles.



Training on Data

Weak supervision with mixed jet samples.

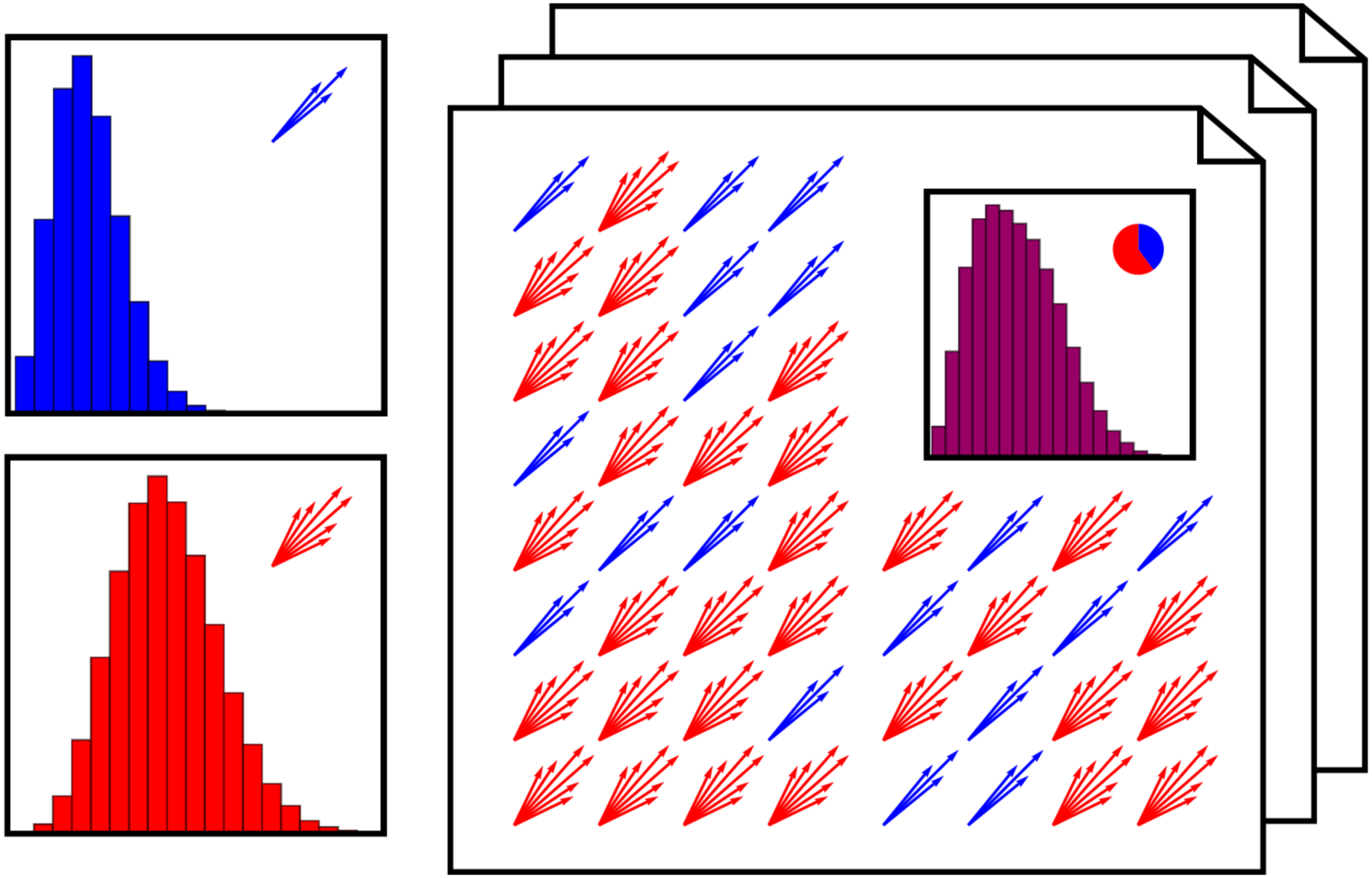


Disentangling Categories

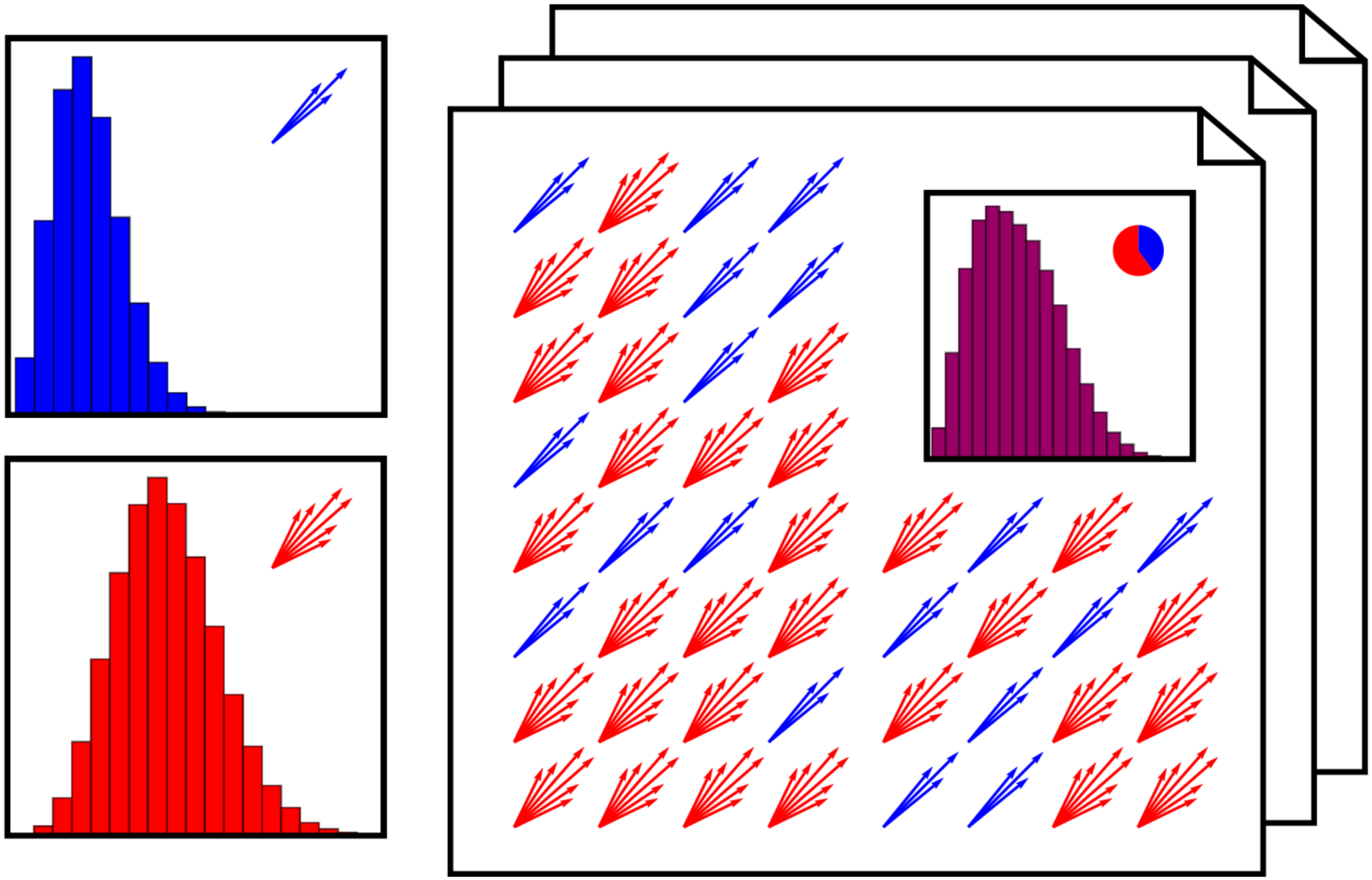
Topic modeling to define data-driven jet categories.

The End

Thank you!



Extra Slides

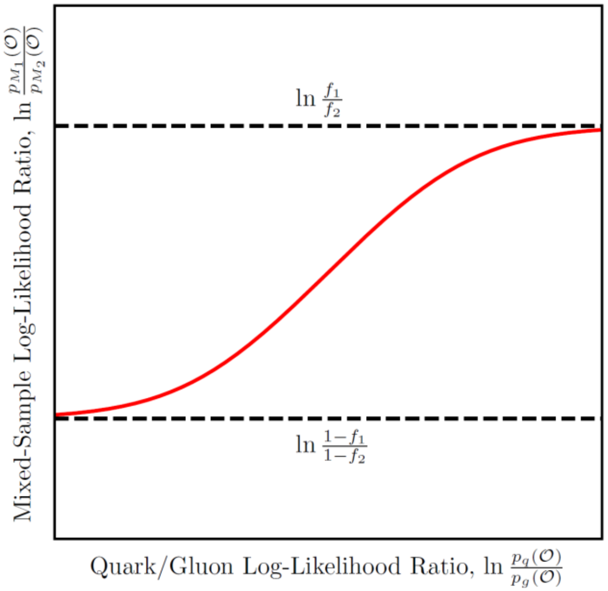


A/B Likelihood Ratio

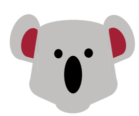
$$p_{\text{sample }A}(\boldsymbol{x}) = f_A^q\, p_{\text{quark}}(\boldsymbol{x}) + \left(1 - f_A^q\right) p_{\text{gluon}}(\boldsymbol{x})$$

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

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



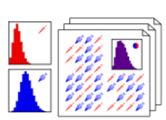
The A/B and quark/gluon likelihood ratios are monotonic!



Classification without labels (CWoLa)
- Optimal A/B classifier is the optimal quark/gluon classifier.
- Use machine learning to approximate A/B likelihood ratio.

[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]

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:

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

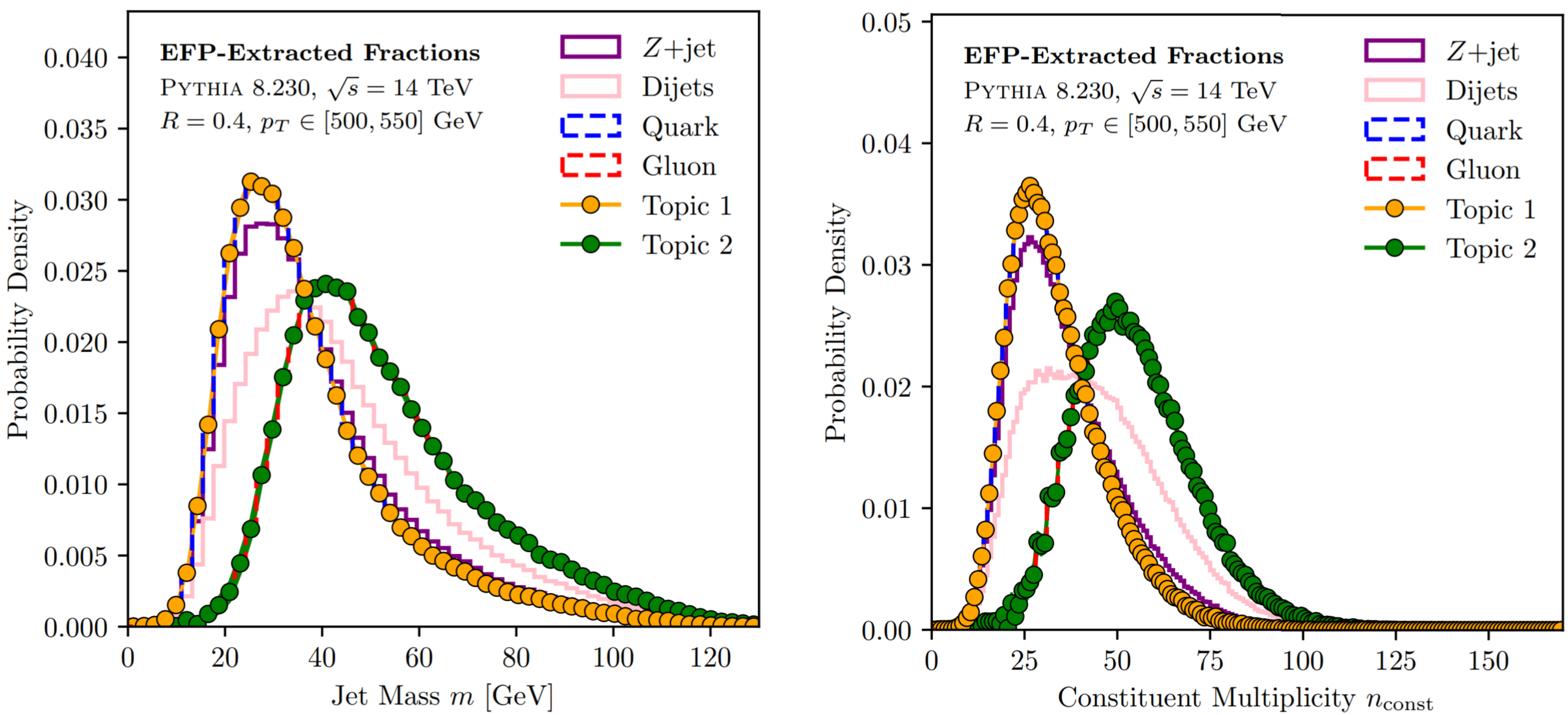
Well-defined and operational statement in terms of hadronic cross sections.

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

Jets themselves are operationally defined.

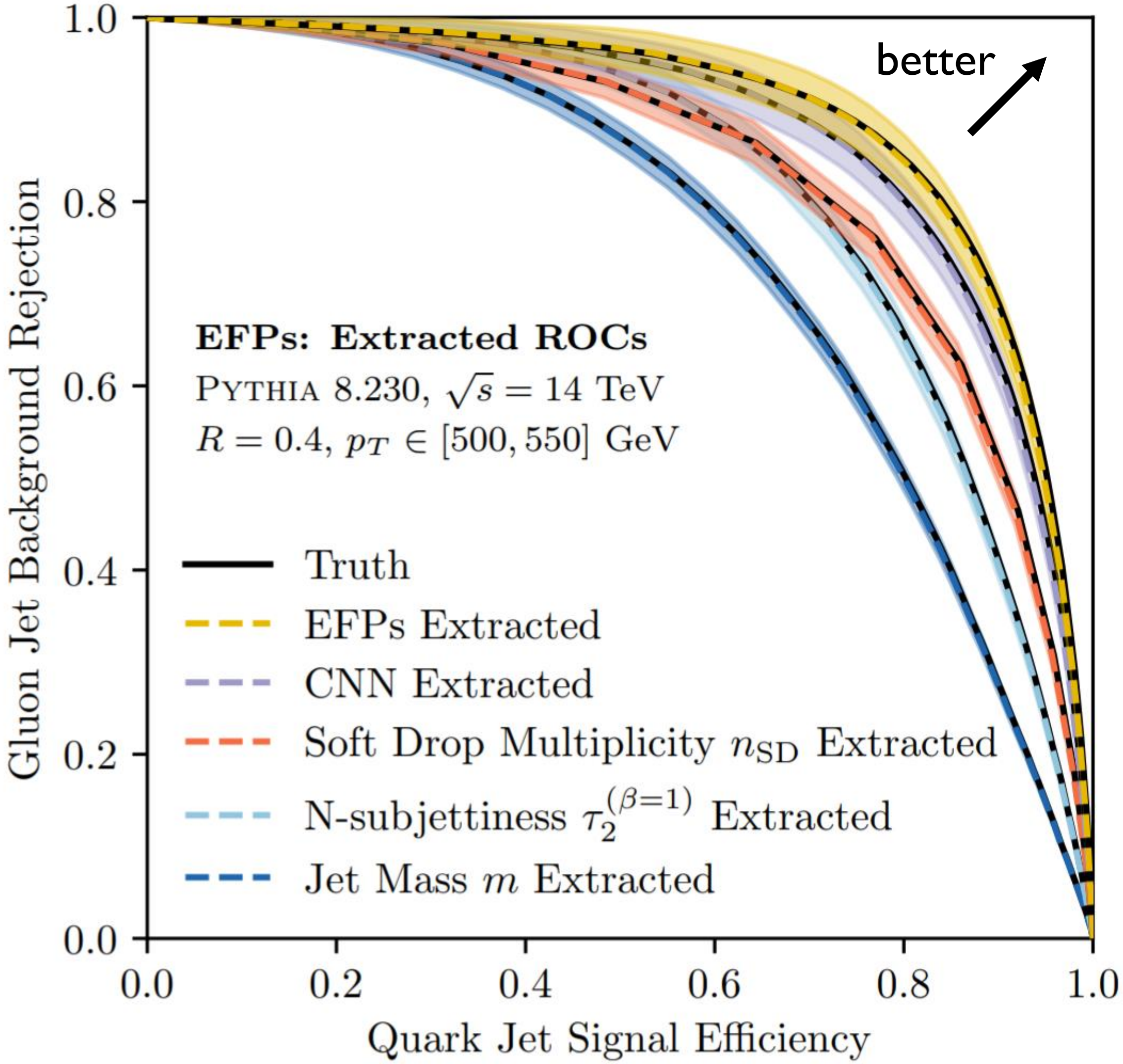
Extracting quark and gluon distributions

The extracted quark and gluon fractions can be used to obtain any quark/gluon distributions.



(Self-)calibrating quark and gluon classifiers

The extracted quark and gluon fractions can calibrate any data-driven quark/gluon classifiers.



Sample dependence?

How does "sample dependence" manifest in this language?

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

Comparing definitions from different pairs of 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:

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

Thus topic modeling techniques could be an interesting avenue to explore issues of sample dependence directly in data.

Jet topics from QCD: Casimir scaling

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

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

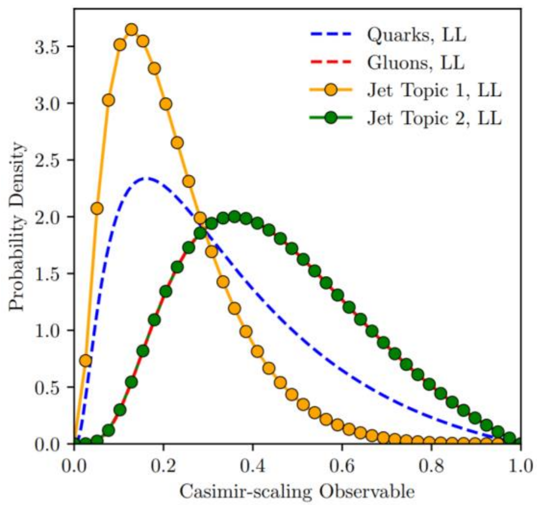
$C_F = \frac{4}{3}$ for quarks

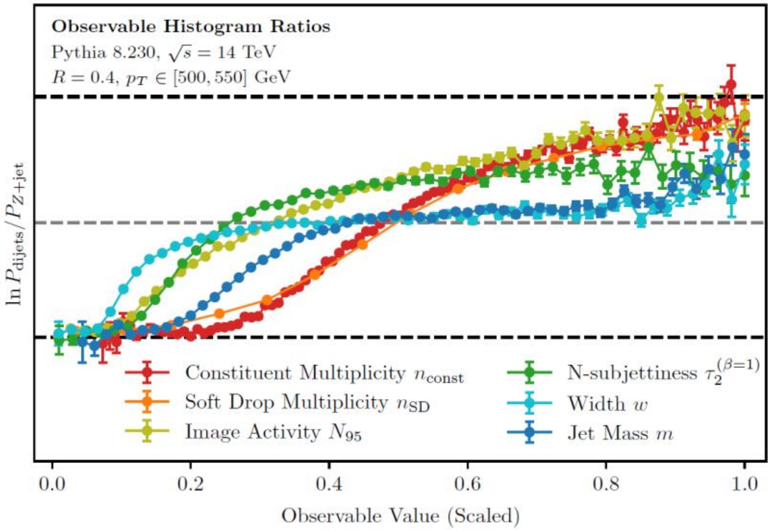
$C_A = 3$ for gluons

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:

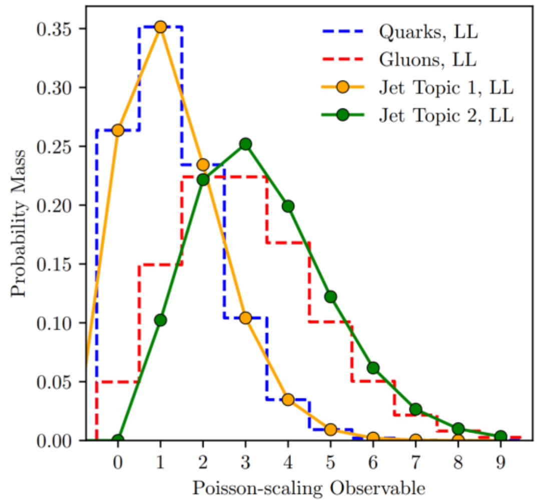
$C_F = \frac{4}{3}$ for quarks

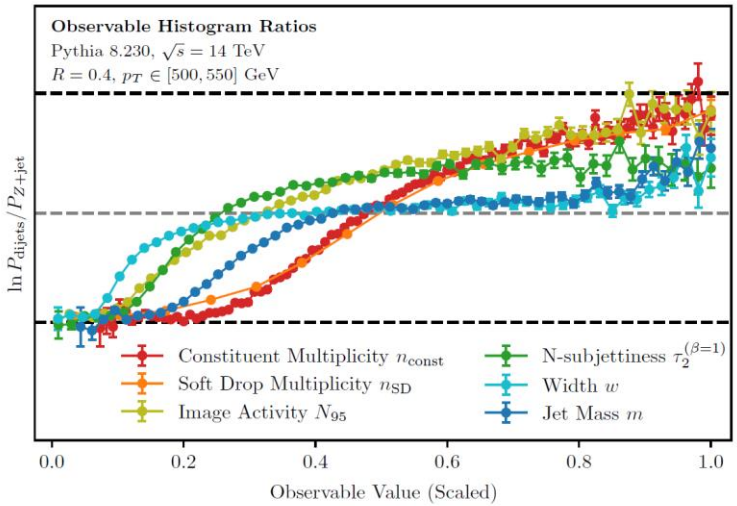
$C_A = 3$ for gluons

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

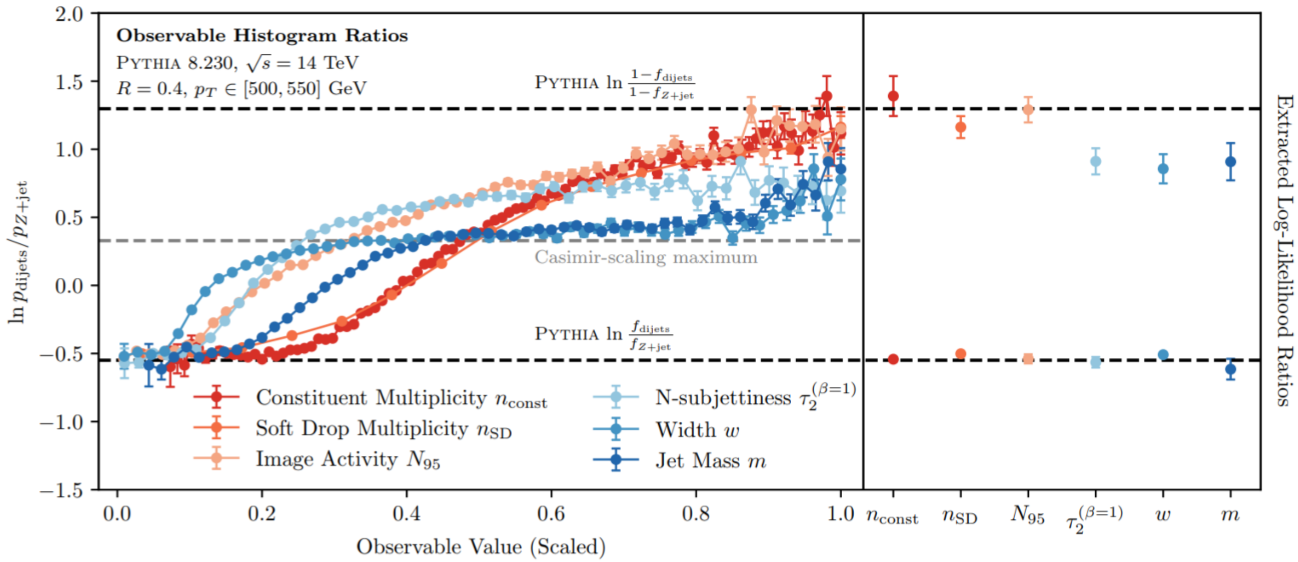
The quark/gluon reducibility factors at LL for any Poisson scaling observable are:

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





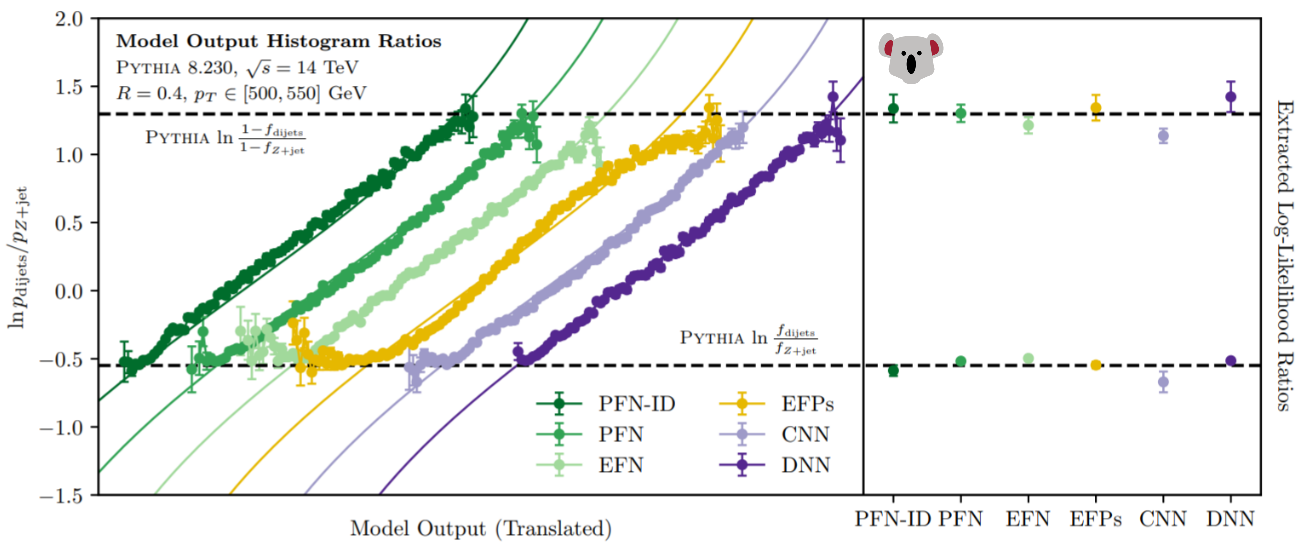
Exploring substructure feature spaces



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

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

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.

Parton-labeled sample dependence in Pythia