Hadron-level quark and gluon jet discrimination

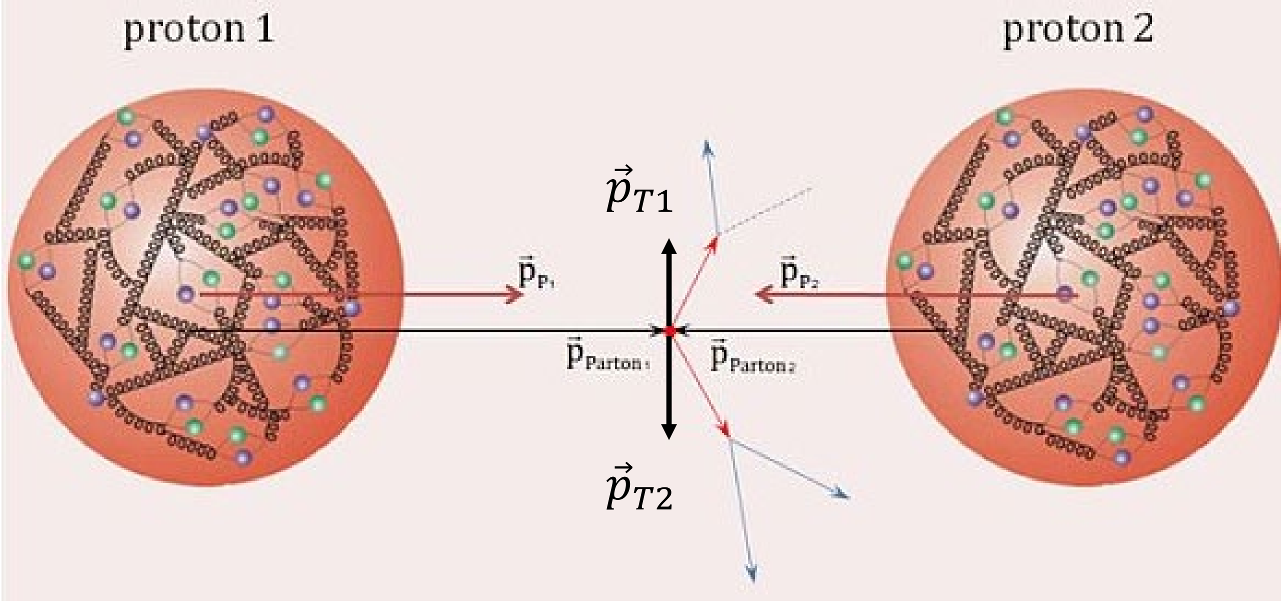
Kristopher Samant

in collaboration with Dag Gillberg and Dylan Pizzi

May 27, 2021

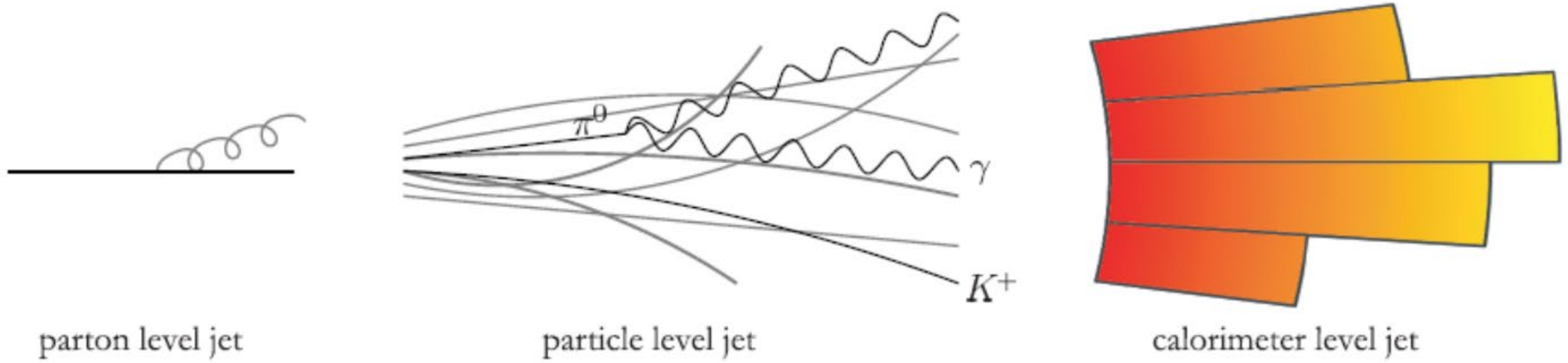
What is a jet?

- At the LHC two protons collide in a **hard scatter** (high p_T)
  - Constituent partons (quarks/gluons) interact
- Partons are emitted outward
  - Creates parton jets and showers
- Partons are "unphysical"
  - Cannot be directly measured



What is a jet?

- Partons in jet **hadronize** and form observable particles
  - E.g. Protons, neutrons, pion, kaons
- Hadrons in a “narrow cone” called a hadron jet (particle jet, truth jet)
- Hadron jets measured in detectors
  - Measurable properties: momentum, mass, number of particles, etc



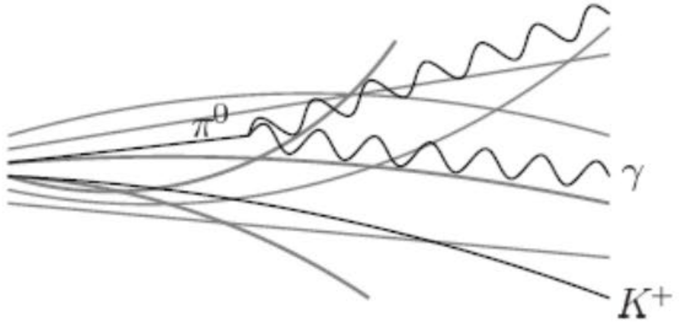
= hadron jet

What is a jet?

- Partons in jet **hadronize** and form observable particles
  - E.g. Protons, neutrons, pion, kaons
- Hadrons in a “narrow cone” called a hadron jet (particle jet, truth jet)
- Hadron jets measured in detectors
  - Measurable properties: momentum, mass, number of particles, etc

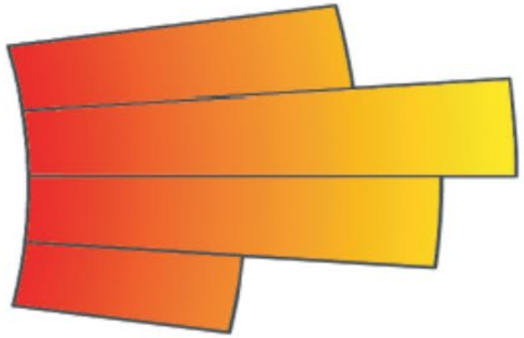
Focus of my project

parton level jet



particle level jet

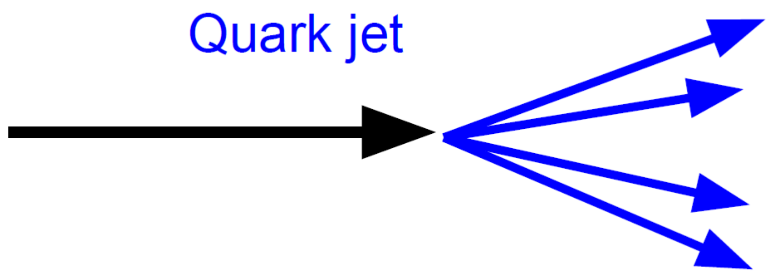
= hadron jet

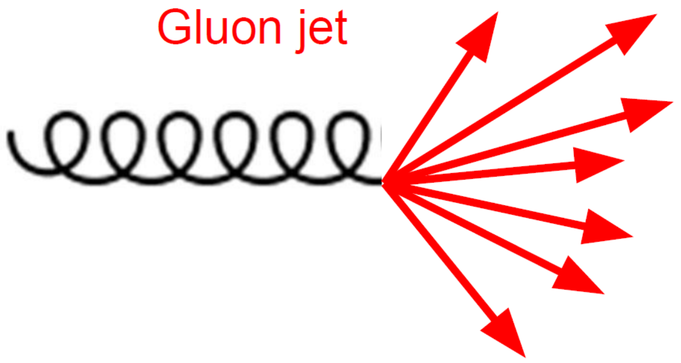


calorimeter level jet

Quark vs gluons jets

- A jet is characterized by the initial parton that produces it
- Quarks and gluons produce jets with different properties after hadronization
- On average, for jets with the same p_T:
  - Quarks $\Longrightarrow$ Fewer hadrons produced, narrower spread, lesser total mass
  - Gluons $\Longrightarrow$ More hadrons produced, wider spread, greater total mass



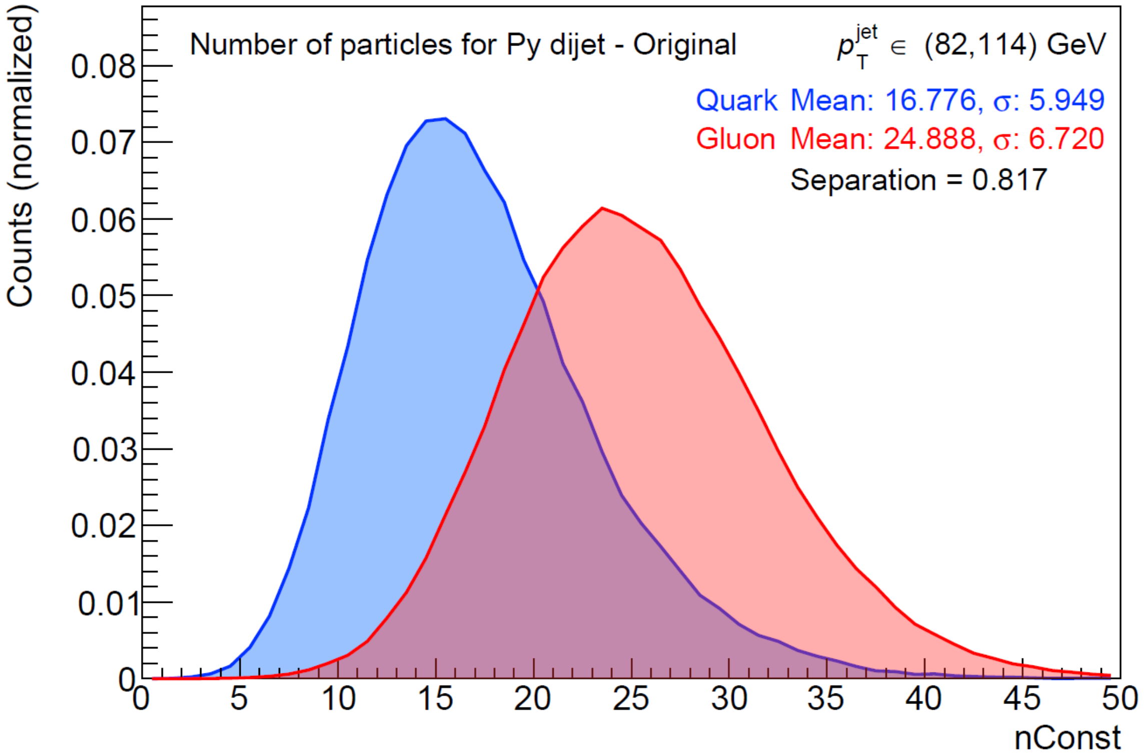


Summer Project Goal

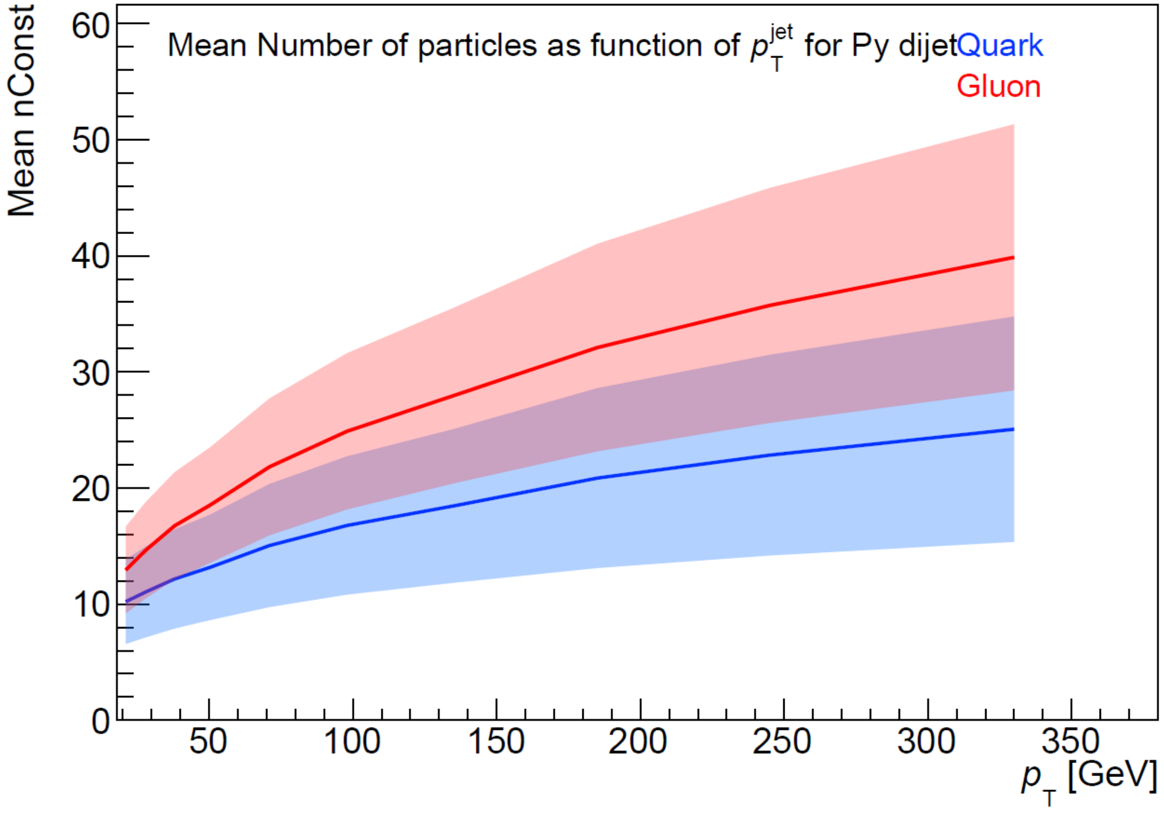
- Create a simple, robust definition of quark and gluon jets
  - Based on hadron-level properties
  - Long desired, but never been done
  - Important for reducing jet flavour uncertainty
- Use simulated data from Monte Carlo generators (Dylan’s work)
- Compare properties of quark and gluon jets
- Use multivariate analysis to classify quark and gluon jets
- Continuation of work by Gareth Smith during summer 2018
  - Internal ATLAS note and Summer student note

Analysis and Separation

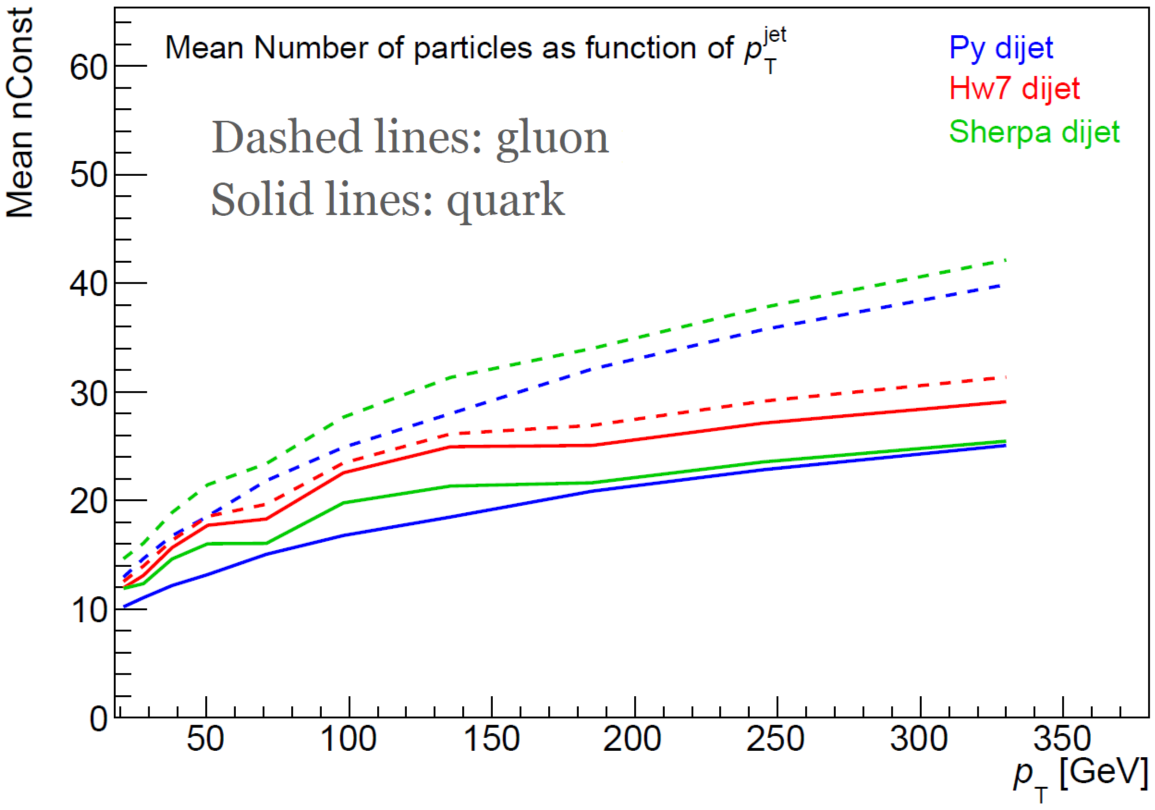
- From MC generator, we choose a narrow p_T range and classify jets as "quark jets" or "gluon jets" (described later)
- Create distribution for a specific hadron-level property
- Define the separation between quark and gluon jets
  - Separation $= \frac{(\mu_q - \mu_g)^2}{\sigma_q^2 + \sigma_g^2}$
  - Quantifies ability of property to distinguish between quark and gluon jets



Example: number of constituents



There is a clear p_T dependence of many hadron-level properties



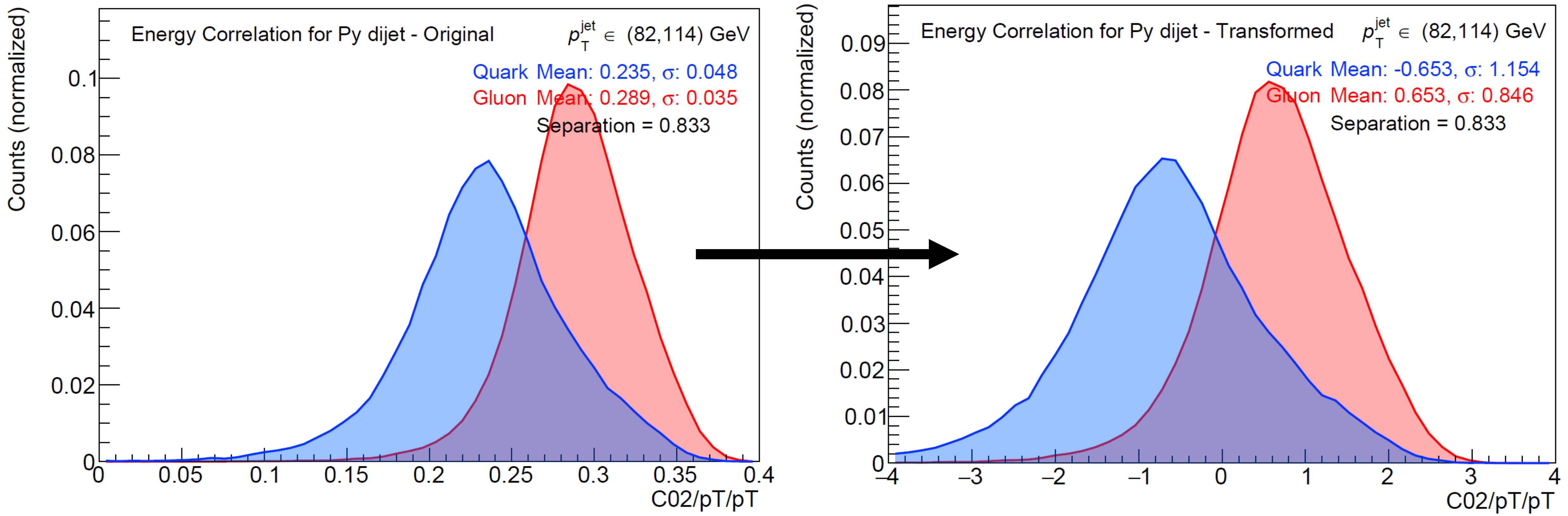
Generators use different modeling techniques and assumptions

Multivariate Analysis with Fisher discriminant

- Takes two classes of data and finds best linear combination of properties to discriminate between them
  - E.g. quark jet vs gluon jet
- Provides a tool for studying many hadron-level properties simultaneously
- Takes into consideration correlation between different properties
  - The same property twice doesn't help!
- Found to be equally as capable as more advanced statistical methods (boosted decision trees) by Gareth Smith
- Described in 4.4.1 of Statistical Data Analysis by Cowan
  - PHYS 4807/PHYS 5002 textbook

Transformed distributions

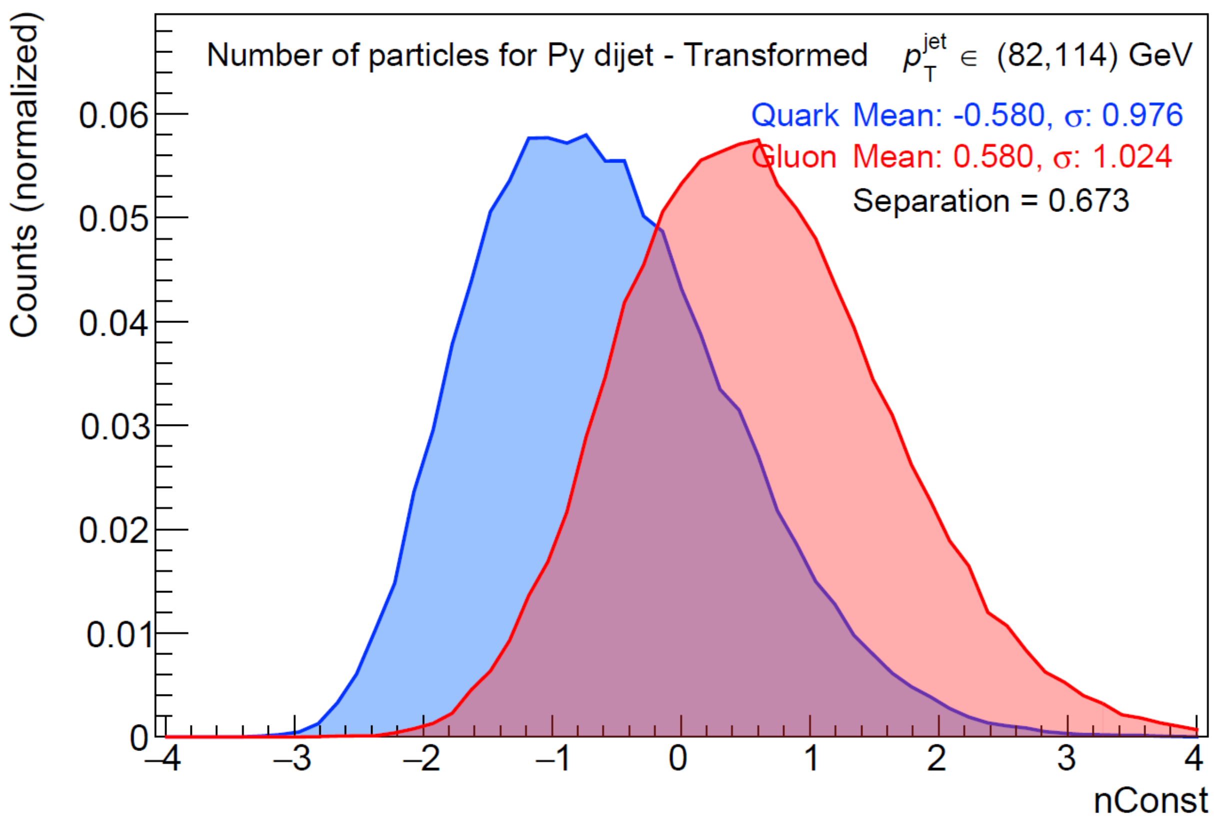
- Distributions are transformed to obtain $\frac{\mu_q+\mu_g}{2} = 0$ and $\frac{\sigma_q+\sigma_g}{2} = 1$
- Puts all properties on an equal footing

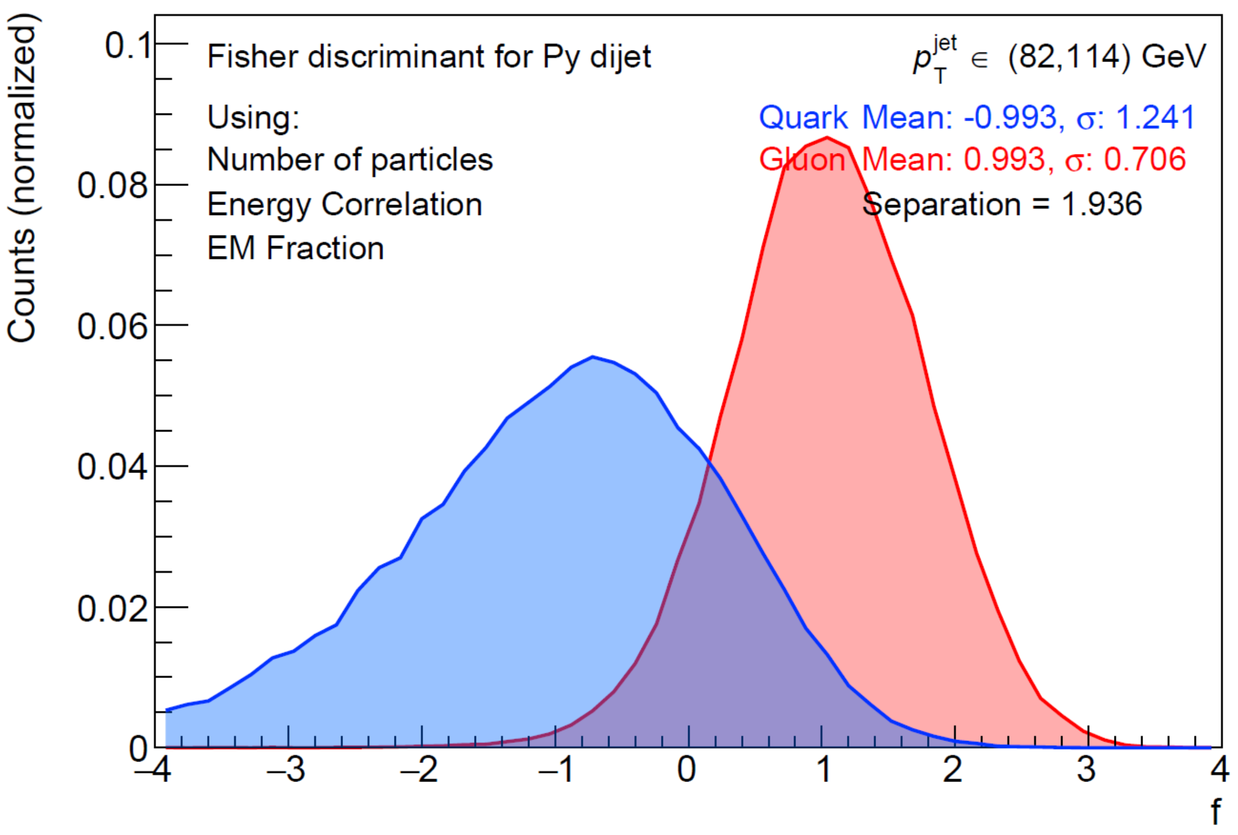


Fisher discriminant

- Assume we have a n hardon-level properties
- Let $\mathrm{x}_1^{jet}, \mathrm{x}_2^{jet}, \dots, \mathrm{x}_\mathrm{n}^{jet}$ be the transformed quantity for a jet
- Define a new quantity $f_{jet} = a_1 x_1^{jet} + a_2 x_2^{jet} + \cdots + a_n x_n^{jet}$
  - $a_1, a_2, \dots, a_n$ are arbitrary coefficients
  - Coefficients same for all jets, both quarks and gluons
- Provides an algebraic method to find coefficients which maximize the separation between the distributions for f_{quark} and f_{gluon}
- Magnitude of coefficient corresponds to importance of a property
  - Tells us which ones to discard in the future

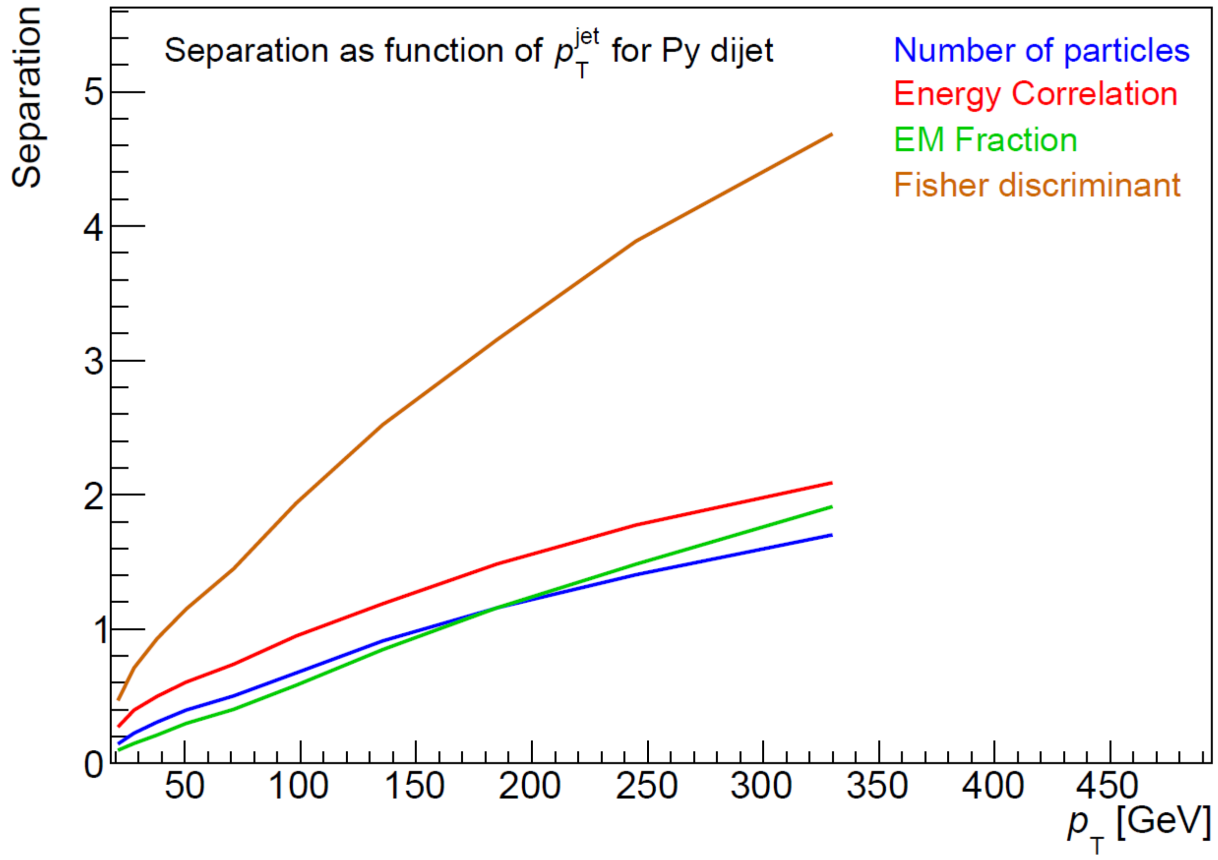
Individual properties vs Fisher discriminant





Distribution for the Fisher discriminant shows greater separation than any of the individual properties

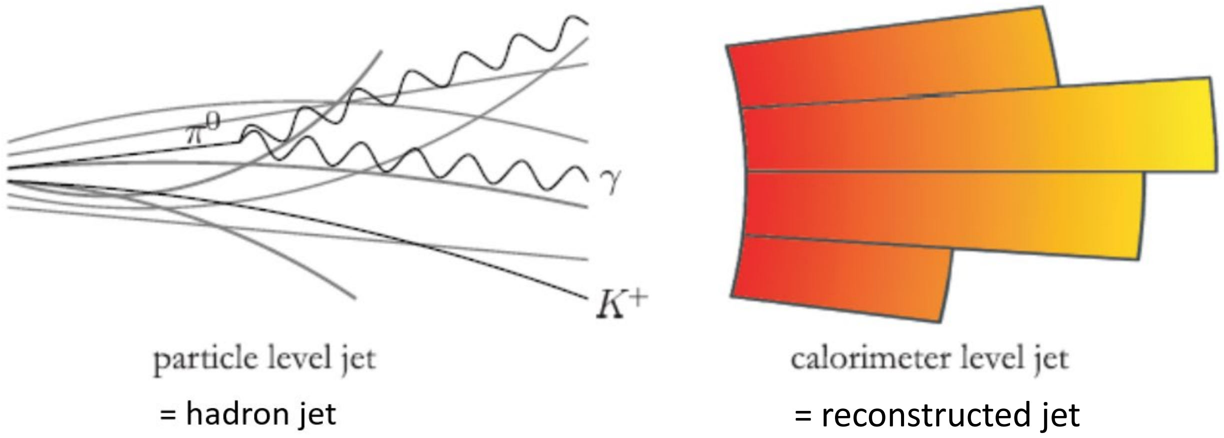
Individual properties vs Fisher discriminant



Separation curves show the power of Fisher discriminant. Highly dependent on p_T.

Labeling jets from MC generators

- Multivariate techniques need two classes of data to compare
- There are several options for classifying jets
  - Parton label – classifies jet based on most energetic parton at shower-level
  - Dijets vs Z boson + jet
    - Dijets – 70% gluon jets
    - Z boson + jet – 70% quark jets
  - Detector response – $R = \frac{p_T^{reco}}{p_T^{hadron}}$
    - Quark – More energetic hadrons, narrower jet ⟹ R > 1
    - Gluon – Less energetic hadrons, wider jet ⟹ R < 1



- Ideally want hadron-level properties that works well with all

What’s next?

- Examine classification using dijets vs Z boson + jet
- Find hadron-level properties that best discriminate
  - Compare across generators, labeling methods and p_T ranges
  - Hopefully 2 or 3 properties
- Develop method of assigning a continuous score between 0 to 1 for a parton jet using these properties
  - $0 \implies$ gluon jet
  - $1 \implies$ quark jet
- Write ATLAS PUBnote and convince other to adopt this definition
  - Based on previous talk by Dag Gillberg, many current physicists have strong feelings on the matter

Thank you for listening.
Any questions?