# DATA-DRIVEN EXTRACTION OF QUARK AND GLUON JET SUBSTRUCTURE


Yueyang (Kylie) Ying

Relativistic Heavy Ion Group

Massachusetts Institute of Technology


Work in collaboration with Jasmine Brewer, Yi Chen, and Yen-Jie Lee

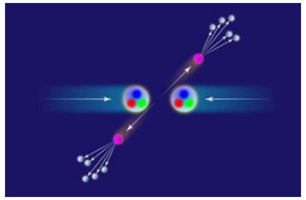
April 7, 2022

Massachusetts Institute of Technology

# What happens during a particle collision?

## Particle collisions

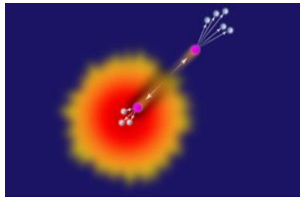
- Collision "kicks out" a high-energy quark or gluon from the nuclei



- Quark/gluon then fragments and hadronizes into sprays of particles

---

- In heavy-ion collisions, high energy can create the quark-gluon plasma

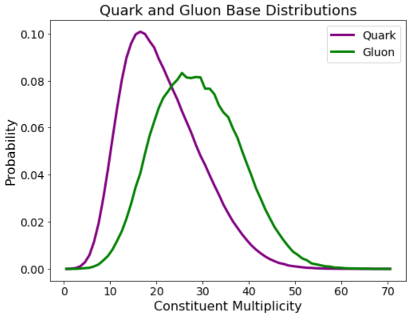


[Manuel, APS]

- Formation of QGP leads to jet quenching

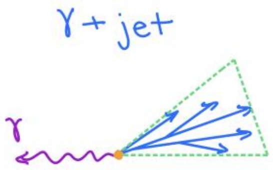
## Our goal

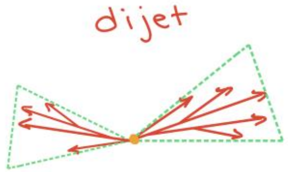
- We want to understand how quark and gluon jets interact with the QGP in a data-driven way
- We use a method known as **topic modeling** in order to extract the quark and gluon base distributions:



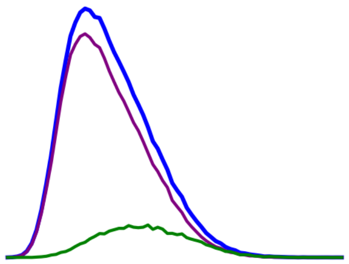

- We apply the results to obtain quark/gluon jet substructure and observe the modification in the QGP

## Two types of jet samples:

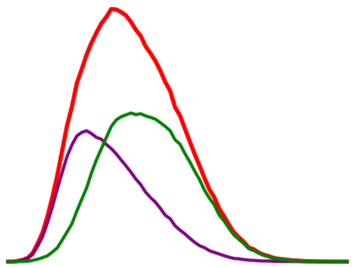


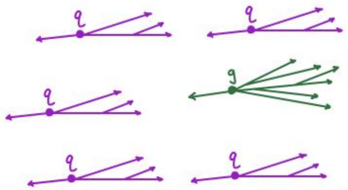

## Both are mixtures of quark and gluon jets:



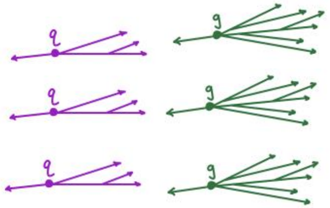
Constituent Multiplicity



Constituent Multiplicity



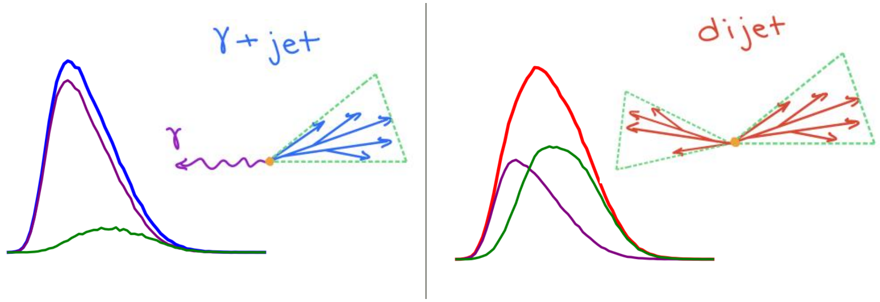

At LHC energies, $\gamma$+jet and dijets have different quark/gluon contributions




Note: The proton-proton and heavy-ion events are simulated using the PYQUEN generator, with $\widehat{p_T} = 80$ GeV. We apply $80 < p_T < 100$ GeV and $|\eta_{jet}| < 1$ cuts to the events.

# Distinguishing quark and gluon jets

Given the two input distributions:




We can express these as mixtures of the base distributions:

$$p_{\gamma+jet}(x) = f_A p_1(x) + (1 - f_A)p_2(x)$$
$$p_{dijets}(x) = f_B p_1(x) + (1 - f_B)p_2(x)$$

BUT! There are infinitely many ways to define the fractions and distributions here!

Thus, we impose a condition on the base distributions, known as **mutual irreducibility**

[Brewer, *et al.*, 2008.08596]

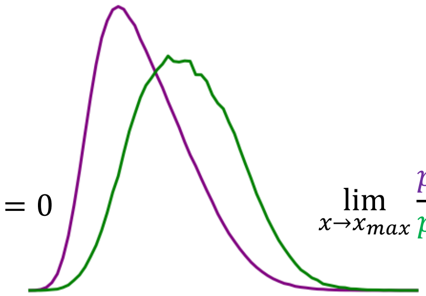
## Mutual irreducibility

- We cannot write $p_1(x) = c\, p_2(x) + (1 - c)F$, where $F$ is some probability distribution and $0 < c \leq 1$
- At the limits, the samples are "pure" topic 1 or topic 2

Let us define:

$$\kappa_{mn} = \inf_x \frac{p_m(x)}{p_n(x)}$$

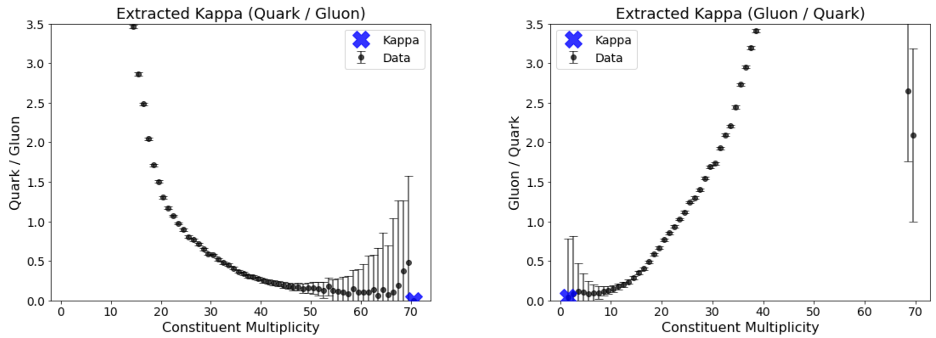
$$\lim_{x \to x_{min}} \frac{p_1(x)}{p_2(x)} = 0$$



$$\lim_{x \to x_{max}} \frac{p_2(x)}{p_1(x)} = 0$$

Mutual irreducibility means $\kappa = 0$!

- The mutual irreducibility requirement is approximately satisfied for quark and gluon jet base distributions (using constituent multiplicity)!




[Komiske, *et al.*, 1809.01140]

# Topic modeling procedure and results

If we can find:

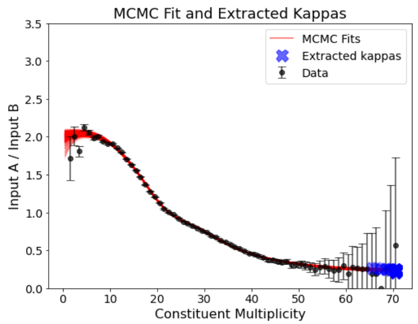
$$\kappa_{AB} = \inf \frac{p_{\gamma+jet}(x)}{p_{dijets}(x)}$$

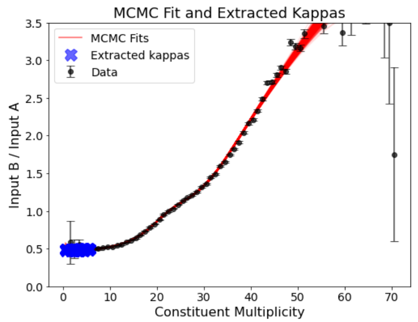
$$\kappa_{BA} = \inf \frac{p_{dijets}(x)}{p_{\gamma+jet}(x)}$$

Then, we can compute the base distributions:

$$p_1(x) = \frac{p_{\gamma+jet}(x) - \kappa_{AB} p_{dijets}(x)}{1 - \kappa_{AB}}$$

$$p_2(x) = \frac{p_{dijets}(x) - \kappa_{BA} p_{\gamma+jet}(x)}{1 - \kappa_{BA}}$$



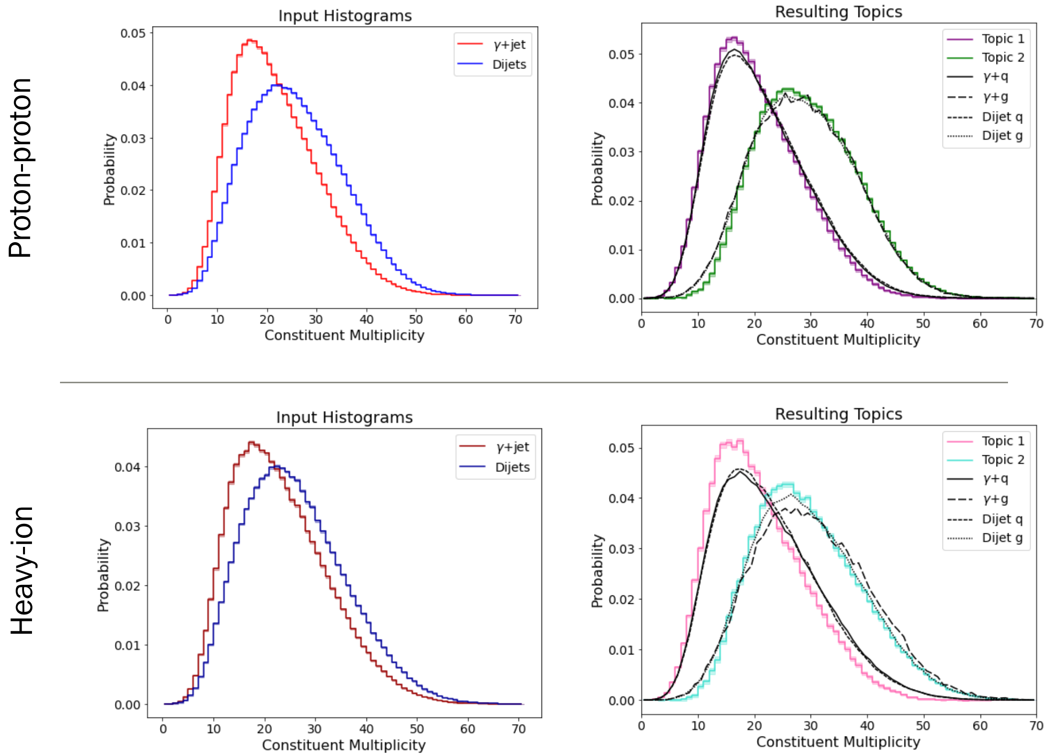




## We can sample for $\kappa_{AB}$ and $\kappa_{BA}$

- Poor statistics at tails of distribution, where $\kappa$ would be, lead to large uncertainties
- Thus, fit a model (sum of 4 skew-normal distributions) to the data, leveraging the rest of the distribution
- Use a sampling algorithm known as **Markov Chain Monte Carlo (MCMC)** to sample parameters that fit the data
- Sample for $\kappa$ by finding the minimum of each MCMC fit ratio (shown on left)

[Brewer, *et al.*, 2008.08596]

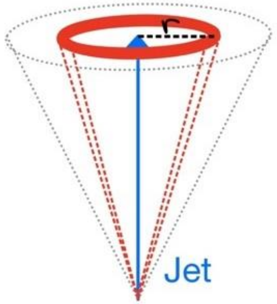
## Topic modeling results




Note: The "truth" labels are assigned by matching the jet to the nearest outgoing matrix element (in terms of angular distance) and represent an approximation of the quark/gluon truth.

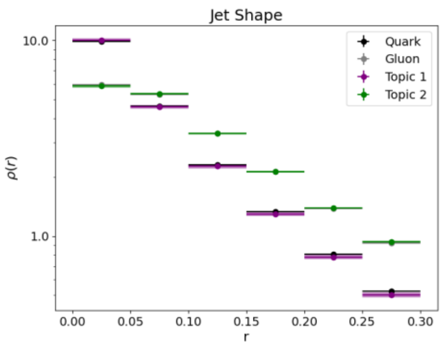
# Extracting jet substructures and modification

## Jet Shape

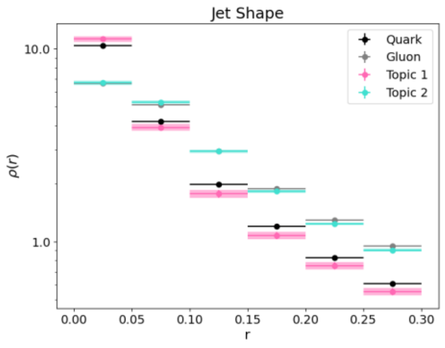



Transverse momentum as a function of radial distance from the jet axis

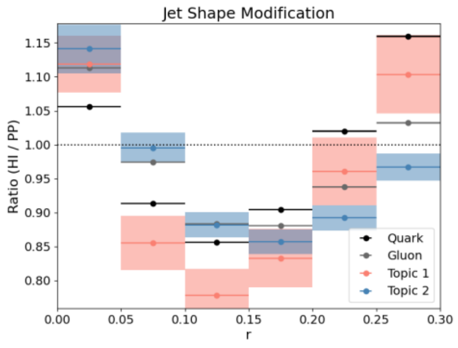
$$\rho(r) = \frac{1}{\delta r}\frac{1}{N_{jet}}\sum_{jets}\frac{\sum_{tracks\in[r_a,r_b)} p_T^{track}}{p_T^{jet}}$$




Proton-proton

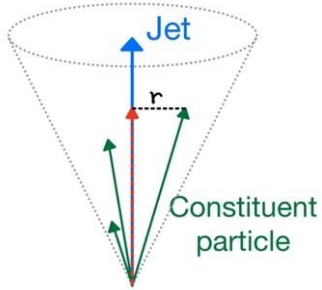



Heavy-ion




Modification (HI/PP)

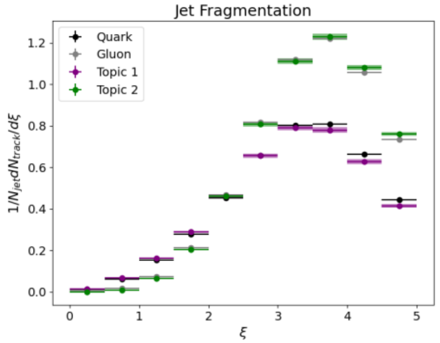
## Jet Fragmentation



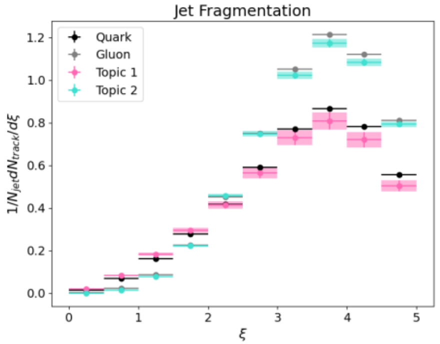

Longitudinal energy distribution of tracks within a jet

$$\epsilon = \ln(1/z)$$

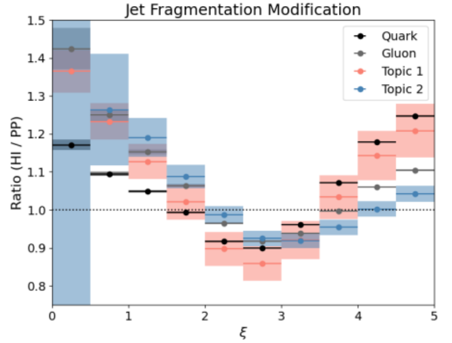
$$z = \frac{p_T}{p_T^{jet}}\cos\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$




Proton-proton



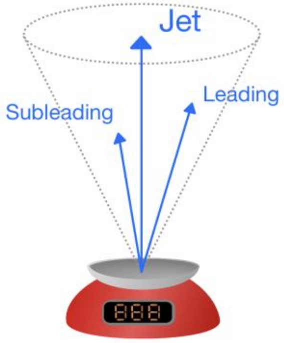

Heavy-ion




Modification (HI/PP)

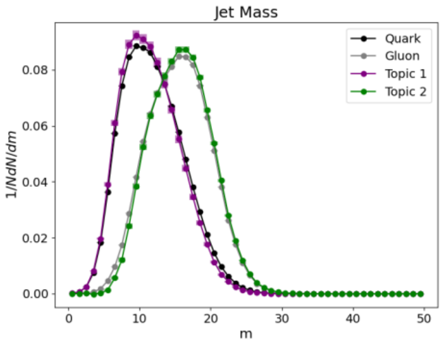
# Extracting jet substructures and modification

## Jet Mass

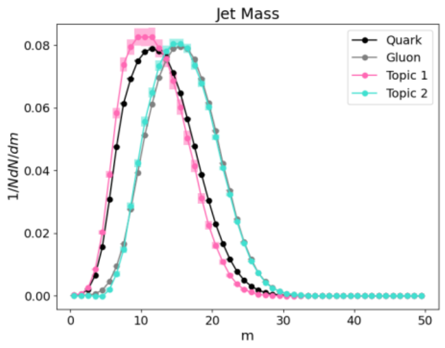



Invariant mass of the jet, calculated from the jet four-momentum

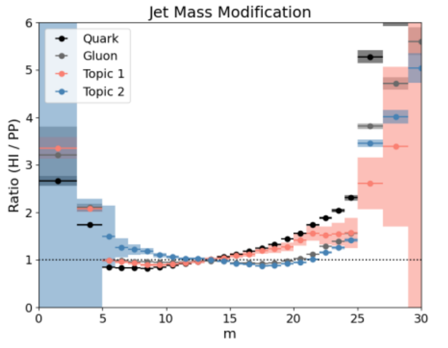
$$m = \sqrt{E^2 - |\vec{p}|^2}$$




Proton-proton

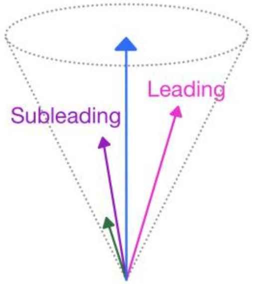



Heavy-ion



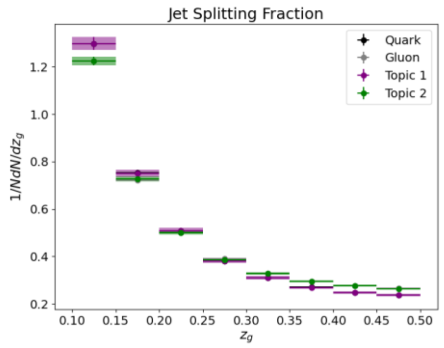

Modification (HI/PP)

## Jet Splitting Fraction

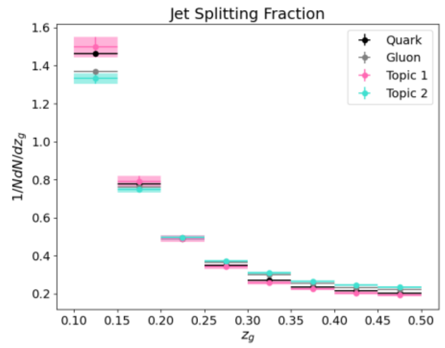



Momentum ratio of the two leading subjets within the jet

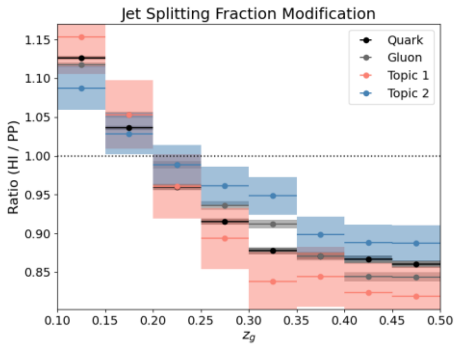
$$z_g = \frac{p_{T,subleading}}{p_{T,leading} + p_{T,subleading}}$$




Proton-proton




Heavy-ion




Modification (HI/PP)

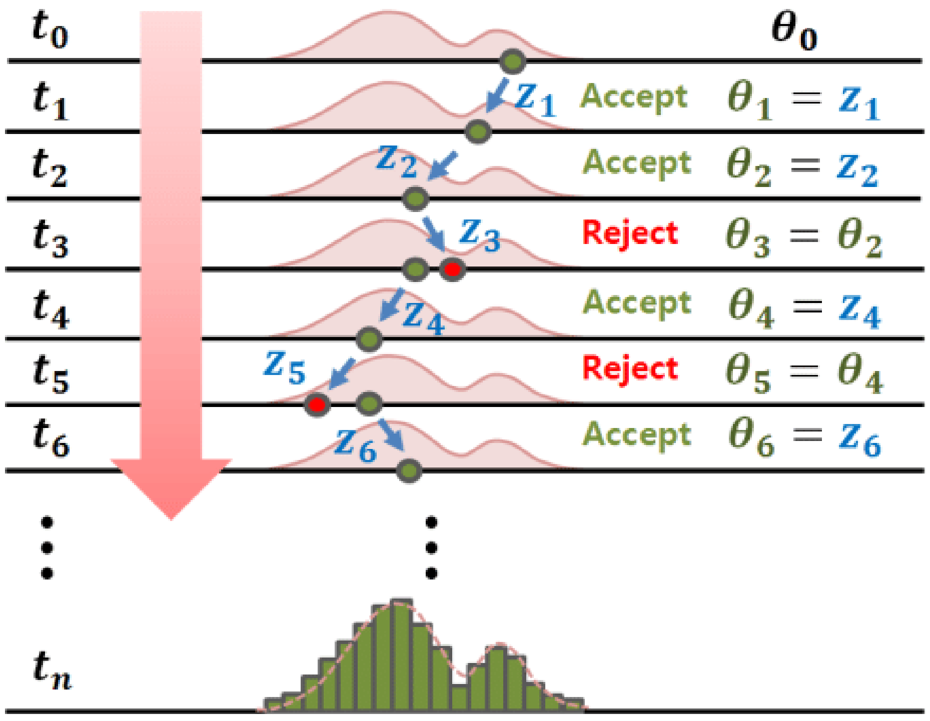
# THANK YOU

This work is supported by US DOE-NP

# BACKUP SLIDES

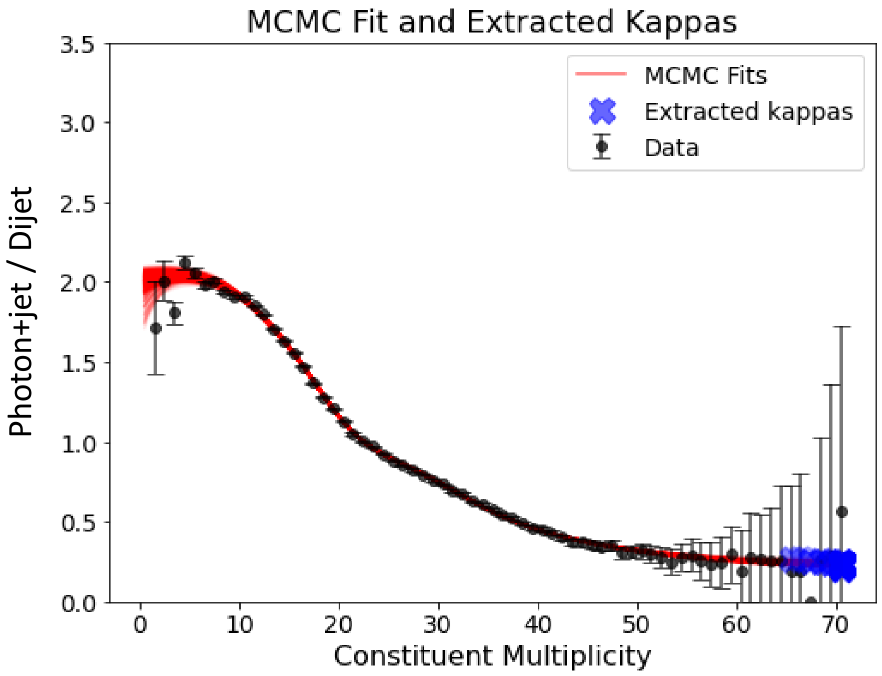
# Markov Chain Monte Carlo

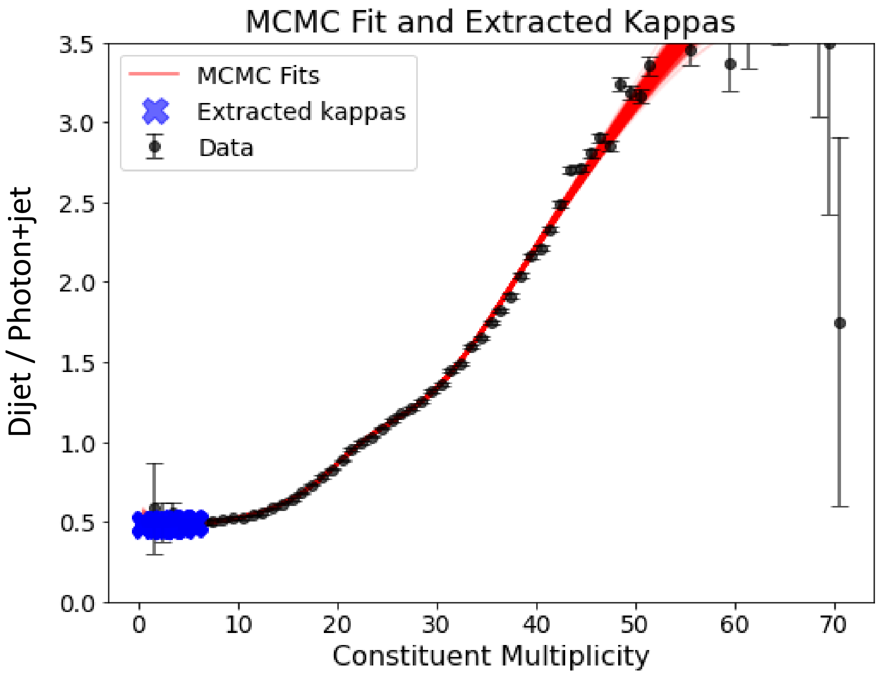
- Each of parameters in our model has a **probability distribution** of its value
  - *MCMC attempts to find this through sampling!*
- Algorithm:
  - *Initialize* $\theta_0$
  - *For t = 0, 1, 2, …, n:*
    - Draw a tentative sample $z_t$ from $Q(\theta|\theta_t)$
    - Accept new $z_t$ with probability A: $$A = \min(1, \frac{P(z_t|D)Q(\theta_t|z_t)}{P(\theta_t|D)Q(z_t|\theta_t)})$$
    - If $z_t$ is accepted, then set $\theta_{t+1} \leftarrow z_t$
    - Else set $\theta_{t+1} \leftarrow \theta_t$




[Jin, *et al.*, 10.1080]

# Extracting Kappas

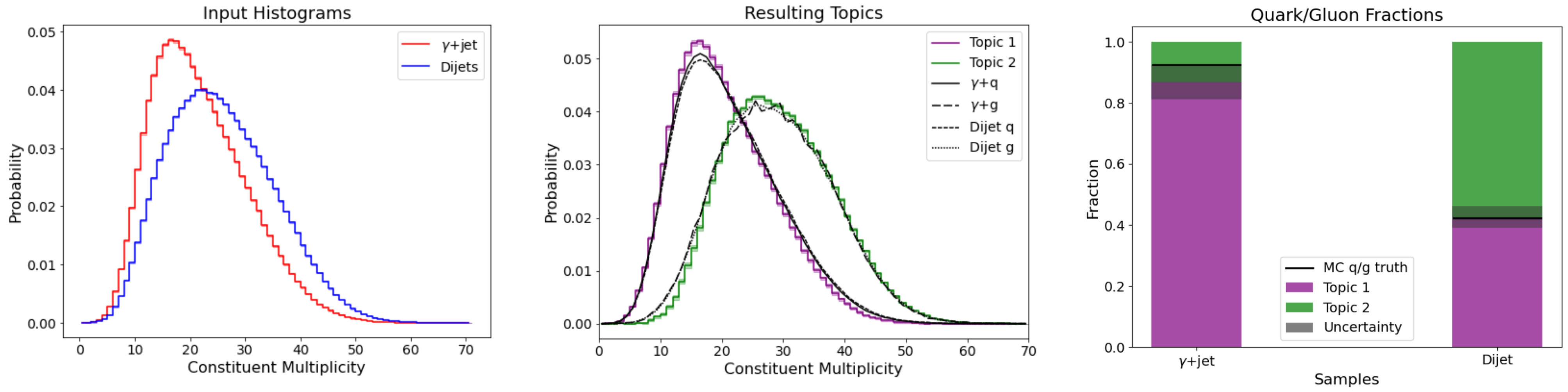



Red lines represent each MCMC fit ratio for the histograms
Black points represent the collected data ratios for the histograms
Blue crosses represent the extracted kappa values

# Topic Modeling Results

Proton-proton




Heavy-ion