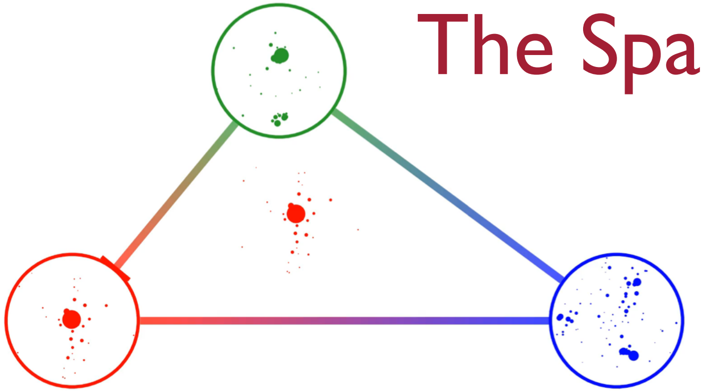


The Space of Collider Events

BOOST 2019

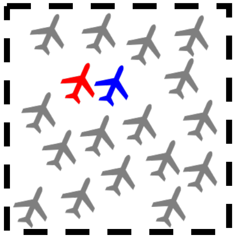
Eric M. Metodiev
Center for Theoretical Physics
Massachusetts Institute of Technology

Joint work with Patrick Komiske, Radha Mastandrea, Preksha Naik, and Jesse Thaler

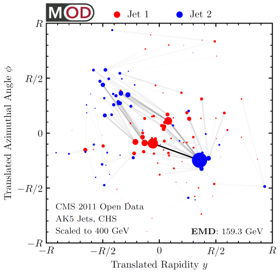
[1902.02346], to appear in PRL [19xx.xxxxx]

July 22, 2019

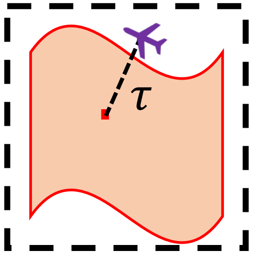
Outline



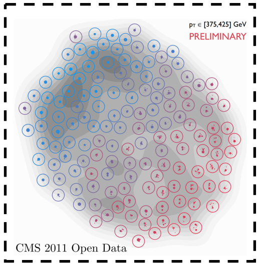
When are two jets similar?



Energy Mover's Distance

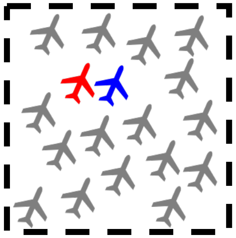


Quantifying Jet Similarity

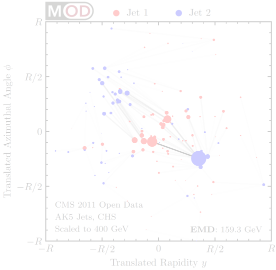


Exploring the Space of Jets

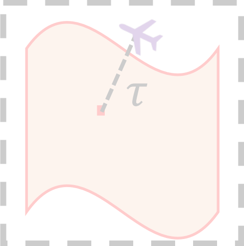
Outline



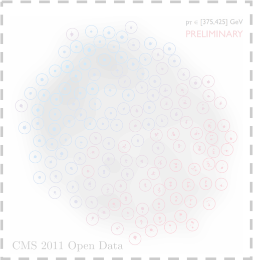
When are two jets similar?



Energy Mover's Distance



Quantifying Jet Similarity

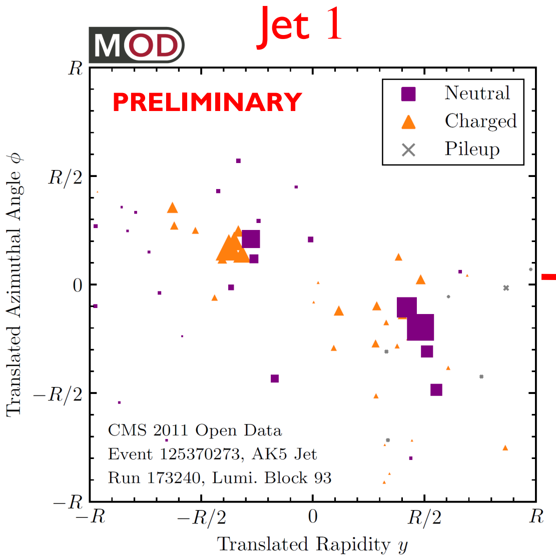


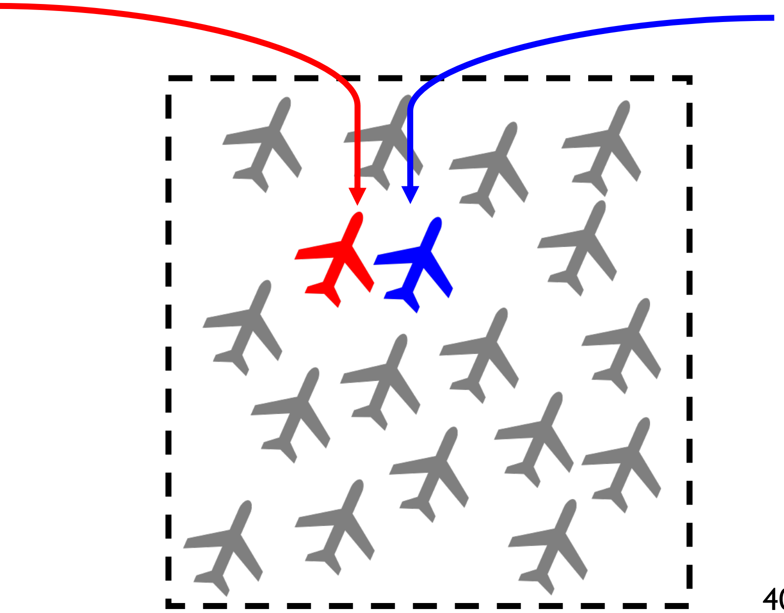
Exploring the Space of Jets

When are two jets similar?

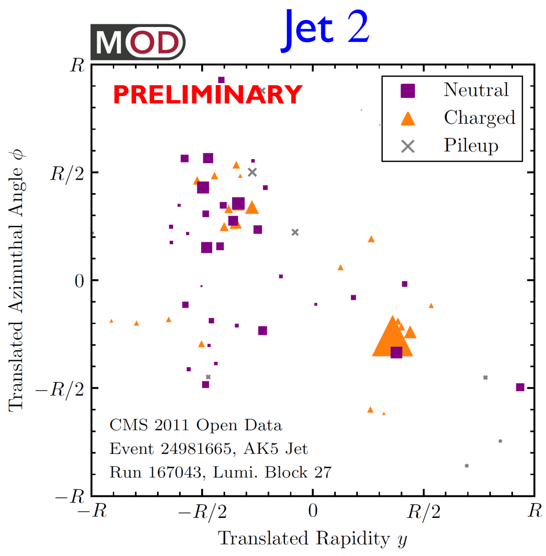
These two jets "look" similar, but have different numbers of particles, flavors, and locations.

How do we quantify this?





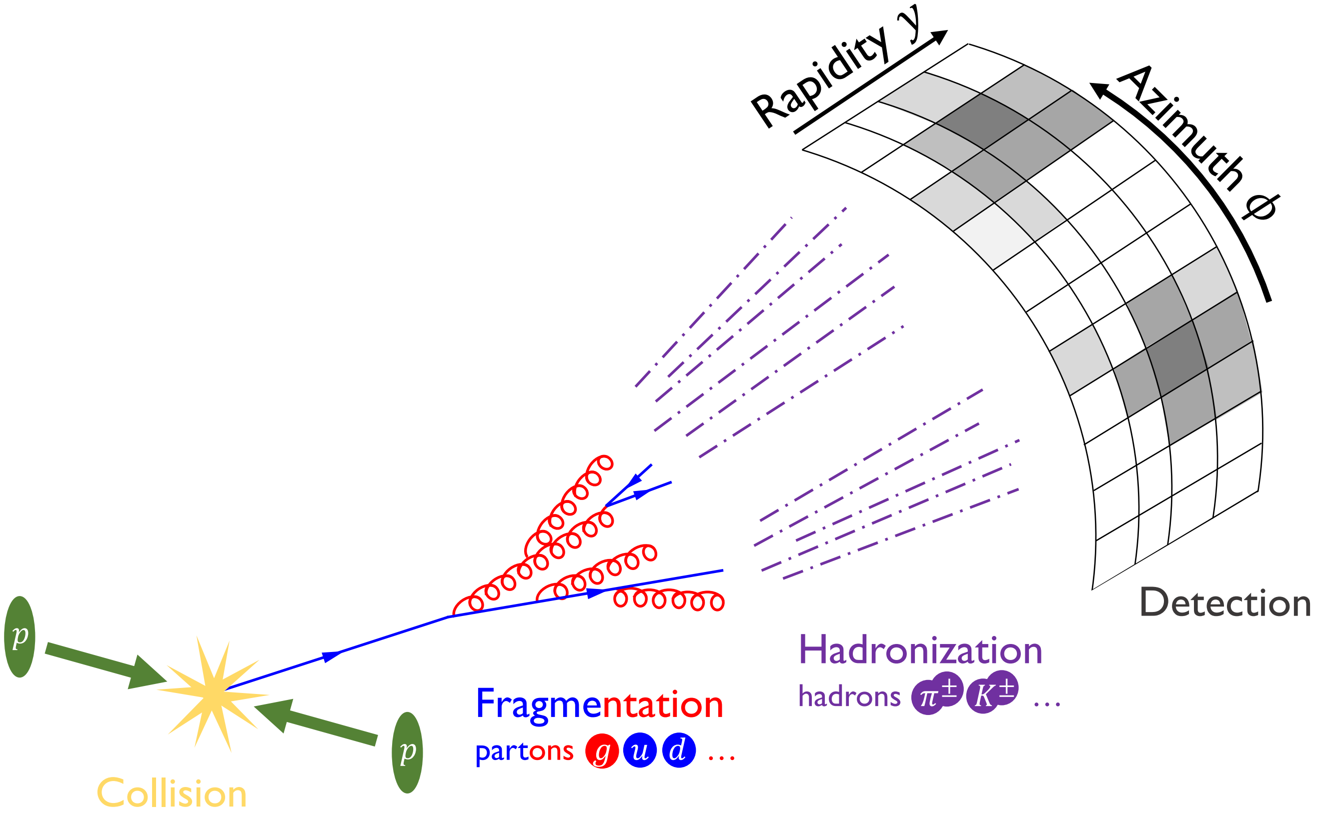
"Space of Jets"



400 GeV AK5 Jets from CMS Open Data

See Radha's talk on Thursday for more!

When are two jets similar?

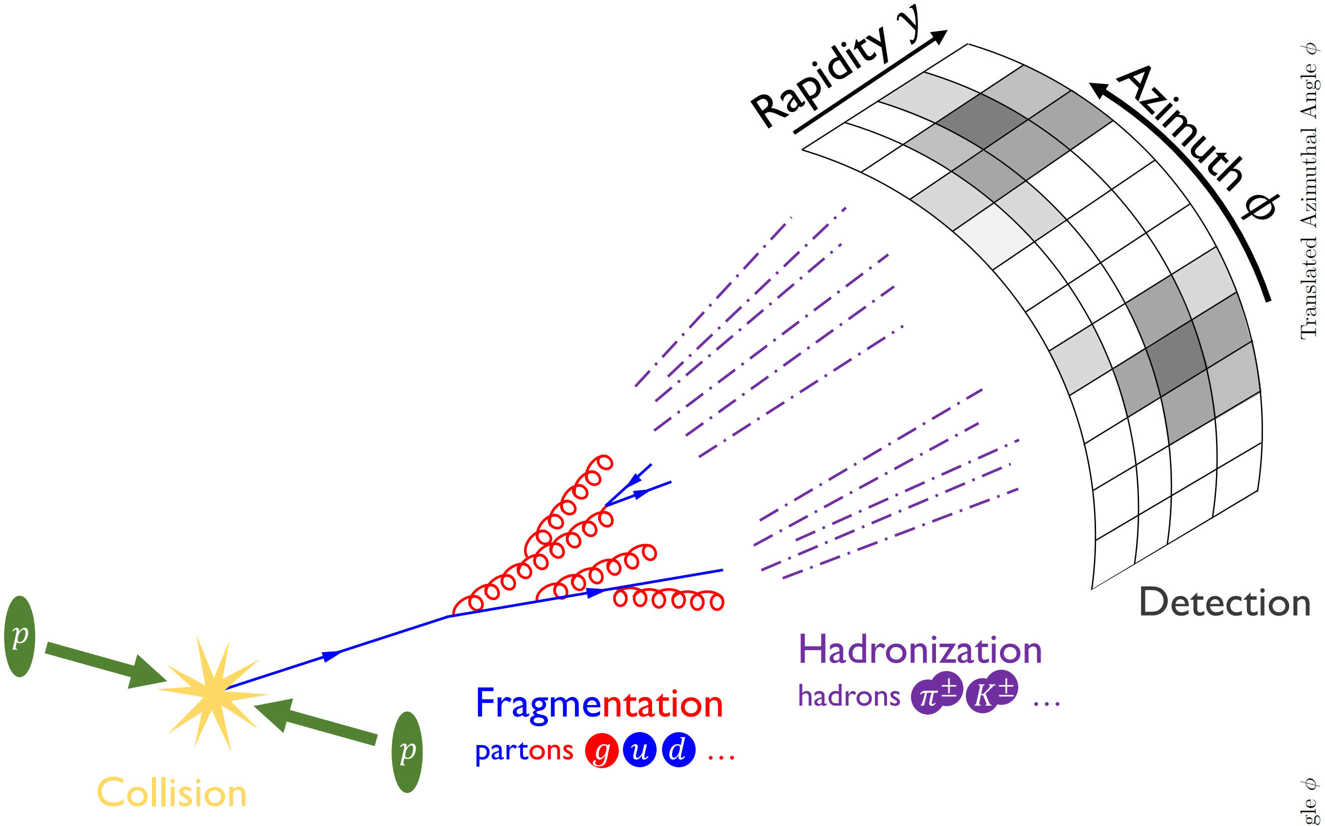


The *energy flow* (distribution of energy) is the information that is robust to: fragmentation, hadronization, detector effects, …

[N.A. Sveshnikov, F.V. Tkachov, 9512370]
[F.V. Tkachov, 9601308]
[P.S. Cherzor, N.A. Sveshnikov, 9710349]

Energy flow ⇔ Infrared and Collinear (IRC) Safe information

When are two jets similar?



Detection

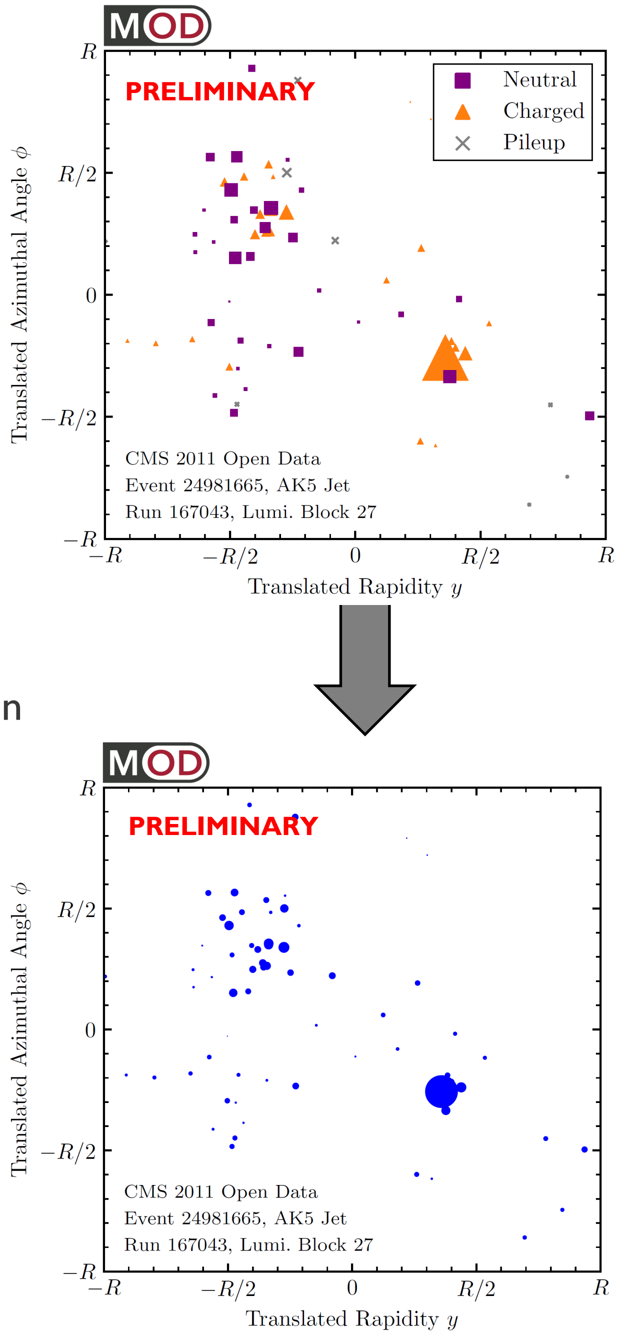
Hadronization

hadrons $\pi^\pm$ $K^\pm$ …

Fragmentation

partons g u d …

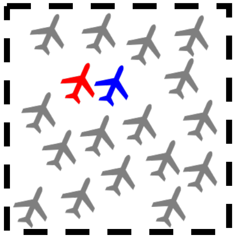
Collision



Treat jets as distributions of energy: $\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} \underset{\substack{\uparrow \\ \text{energy}}}{E_i}\, \delta(\hat{n} - \underset{\substack{\uparrow \\ \text{direction}}}{\hat{n}_i})$

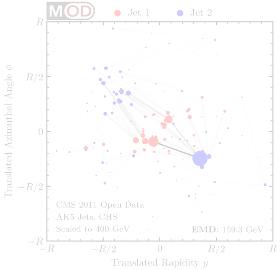
Ignoring particle flavor, charge…

Outline

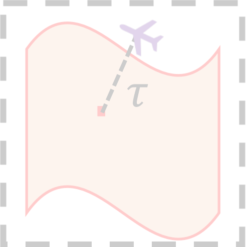


When are two jets similar?

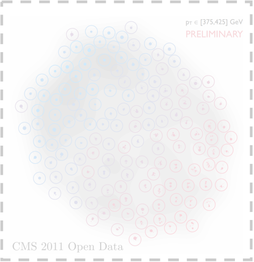
When they have similar distributions of energy



Energy Mover's Distance



Quantifying Jet Similarity

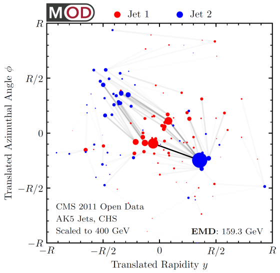


Exploring the Space of Jets

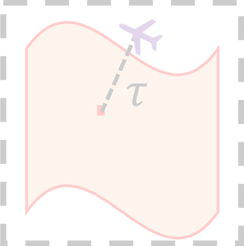
Outline

When are two jets similar?

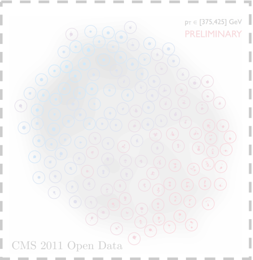
When they have similar distributions of energy



Energy Mover's Distance



Quantifying Jet Similarity



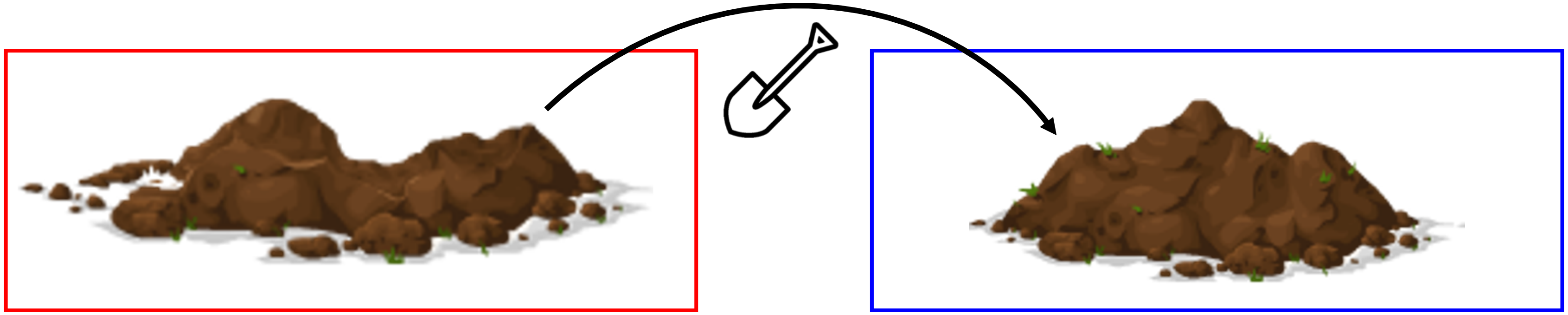
Exploring the Space of Jets

The Energy Mover’s Distance

Review: The Earth Mover’s Distance

Earth Mover’s Distance: the minimum “work” (stuff x distance) to rearrange one pile of dirt into another

[Peleg, Werman, Rom]
[Rubner, Tomasi, Guibas]



Metric on the space of (normalized) distributions: *symmetric, non-negative, triangle inequality*

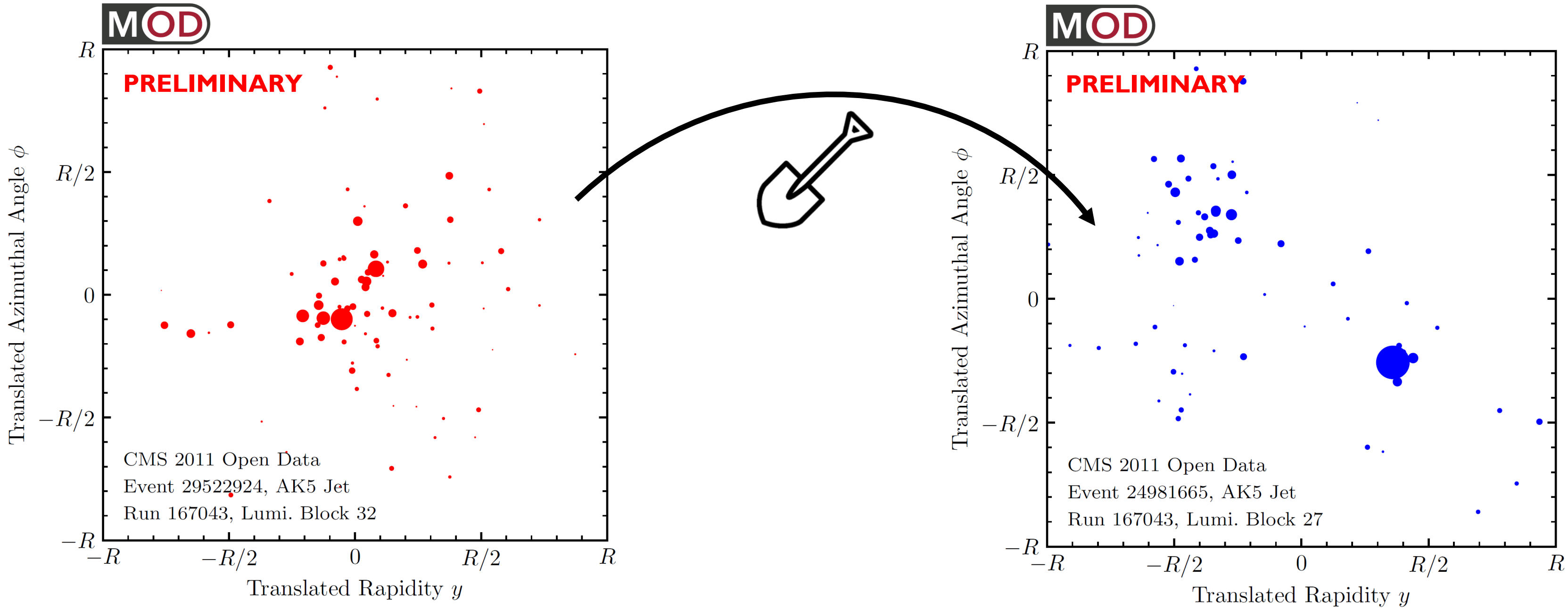
Distributions are close in EMD ⇔ their expectation values are close.

Also known as the 1-Wasserstein metric.

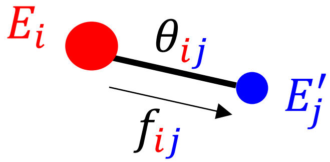
The Energy Mover's Distance

From Earth to Energy

Energy Mover's Distance: the minimum "work" (energy x angle) to rearrange one *jet* (pile of energy) into another [Komiske, EMM, Thaler, 1902.02346]



$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \min_{\{f\}} \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij}\frac{\theta_{ij}}{R} + \left|\sum_{i=1}^{M}E_i - \sum_{j=1}^{M'}E'_j\right|$$



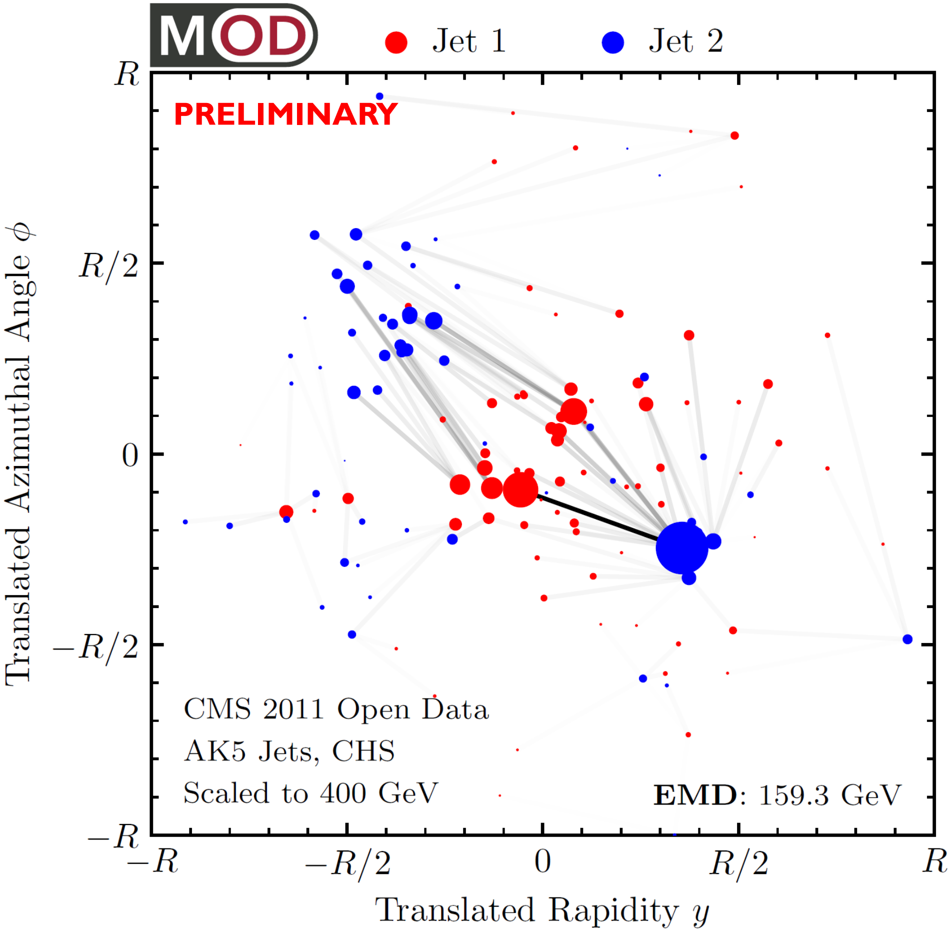
Difference in radiation pattern

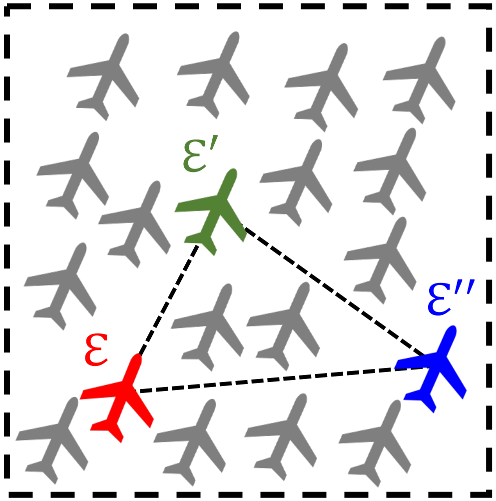
Difference in total energy

The Energy Mover's Distance

From Earth to Energy

Energy Mover's Distance: the minimum "work" (energy x angle) to rearrange one *event* (pile of energy) into another [Komiske, EMM, Thaler, 1902.02346]





$$\text{EMD}(\mathcal{E}, \mathcal{E}') + \text{EMD}(\mathcal{E}', \mathcal{E}'') \geq \text{EMD}(\mathcal{E}, \mathcal{E}'')$$

EMD has dimensions of energy

True metric as long as $R \geq \frac{1}{2}\theta_{\max}$

$R \geq$ the jet radius, for conical jets

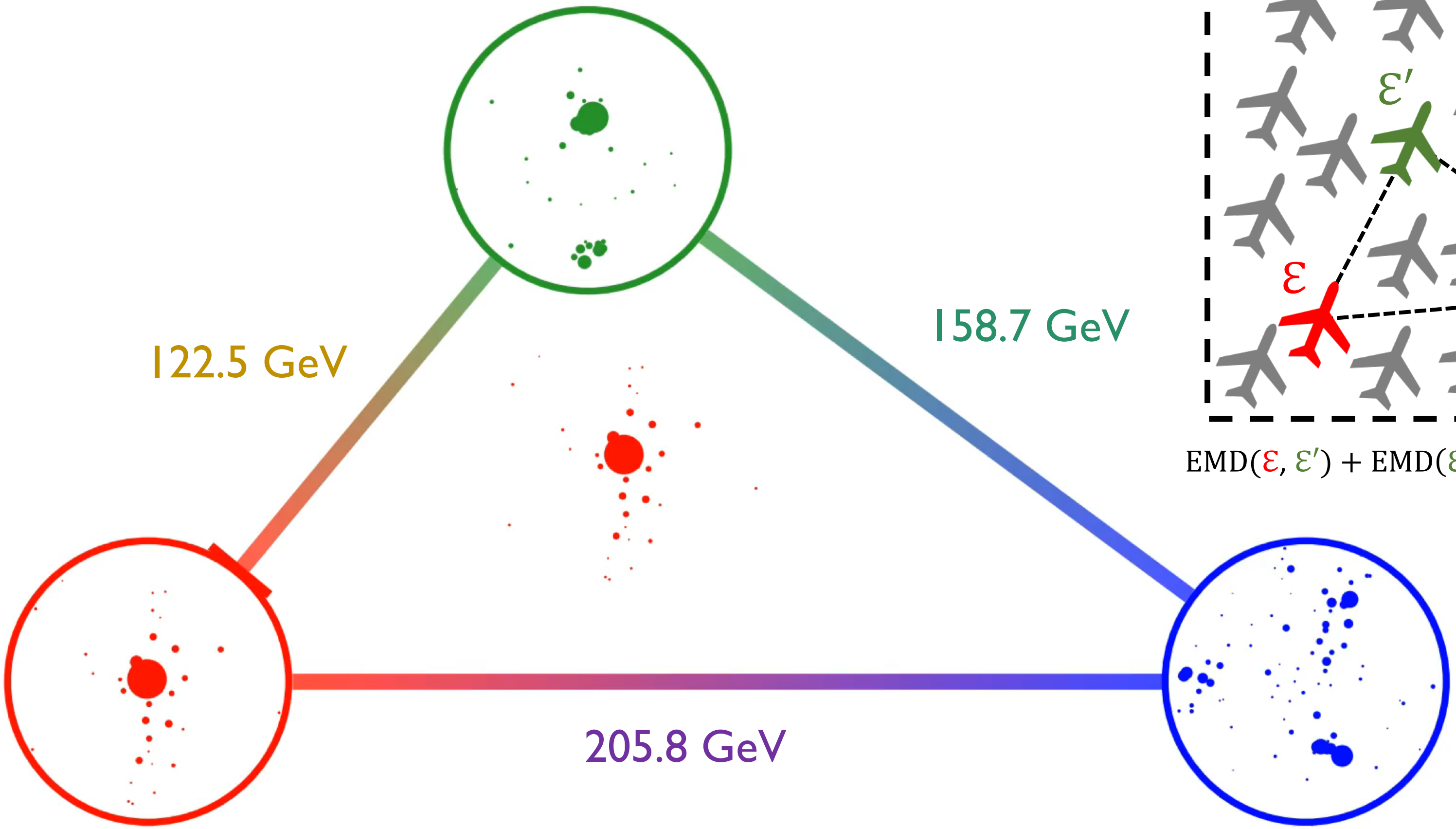
Solvable via Optimal Transport problem.

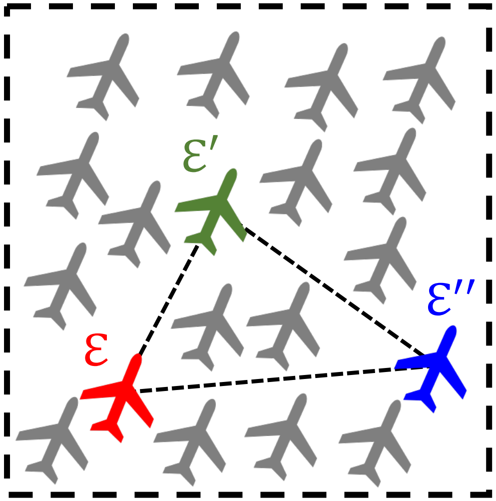
~1ms to compute EMD for two jets with 100 particles.

The Energy Mover's Distance

From Earth to Energy

Energy Mover's Distance: the minimum "work" (energy x angle) to rearrange one *event* (pile of energy) into another [Komiske, EMM, Thaler, 1902.02346]





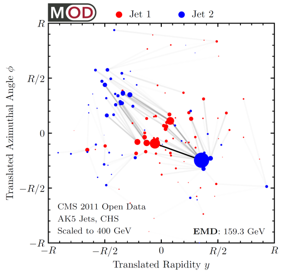
$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') + \mathrm{EMD}(\mathcal{E}',\mathcal{E}'') \geq \mathrm{EMD}(\mathcal{E},\mathcal{E}'')$$

https://energyflow.network

Outline

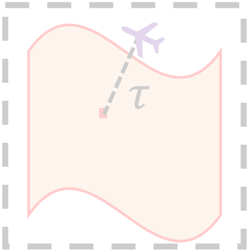
When are two jets similar?

When they have similar distributions of energy

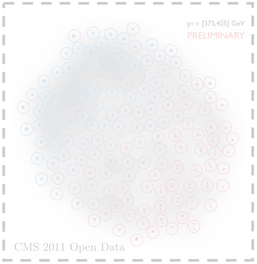


Energy Mover's Distance

The "work" to rearrange one jet into another



Quantifying Jet Similarity

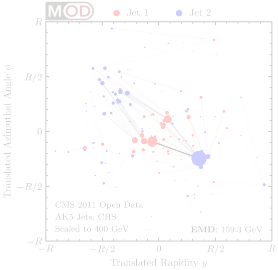


Exploring the Space of Jets

Outline

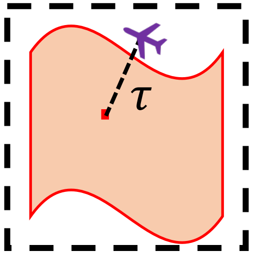
When are two jets similar?

When they have similar distributions of energy

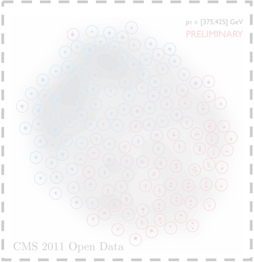


Energy Mover’s Distance

The “work” to rearrange one jet into another



Quantifying Jet Similarity



Exploring the Space of Jets

Energy Moving and IRC Safety

Events close in EMD are close in any infrared and collinear safe observable!

Additive IRC-safe observables:

Energy Mover's Distance

$$\boxed{\mathrm{EMD}(\mathcal{E},\mathcal{E}') \geq \frac{1}{RL}\left|\mathcal{O}(\mathcal{E})-\mathcal{O}(\mathcal{E}')\right|}$$

Difference in observable values

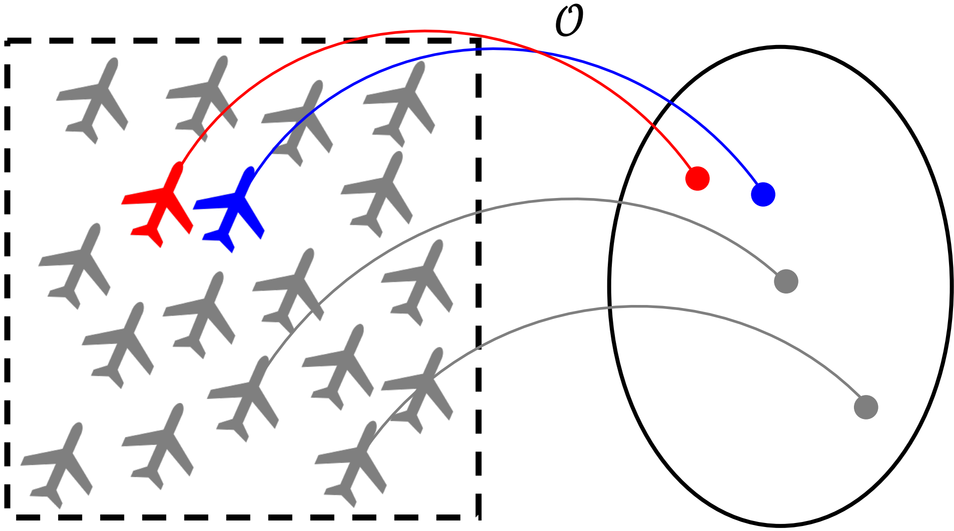
"Lipschitz constant" of Φ
i.e. bound on its derivative

$$\mathcal{O}(\mathcal{E})=\sum_{i=1}^{M} E_i\,\Phi(\hat{n}_i)$$

e.g. $\beta \geq 1$ jet angularities:

[Berger, Kucs, Sterman, 0303051]
[Larkoski, Thaler, Waalewijn, 1408.3122]

$$\left|\lambda^{(\beta)}(\mathcal{E})-\lambda^{(\beta)}(\mathcal{E}')\right| \leq \beta\,\mathrm{EMD}(\mathcal{E},\mathcal{E}')$$

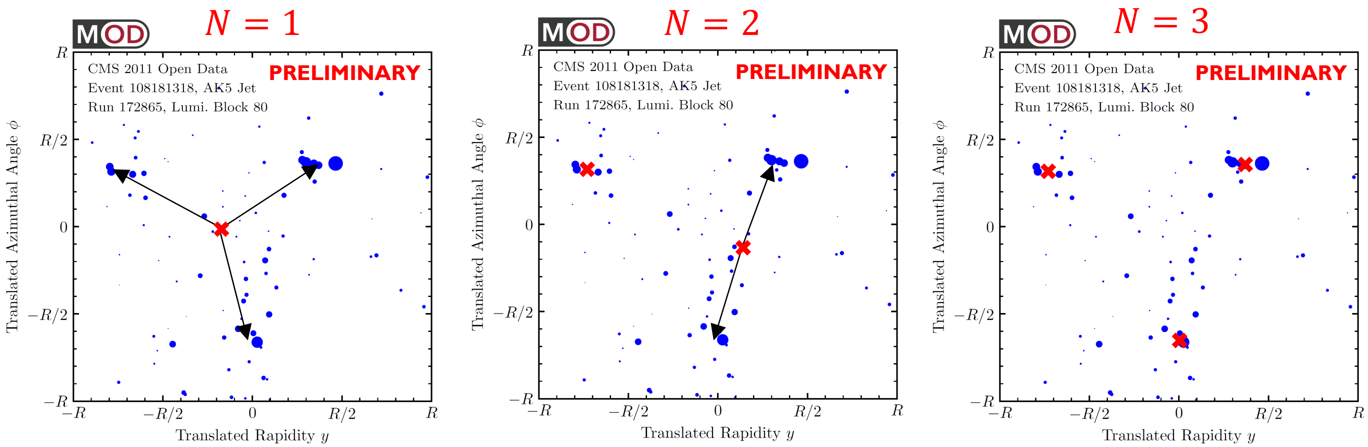


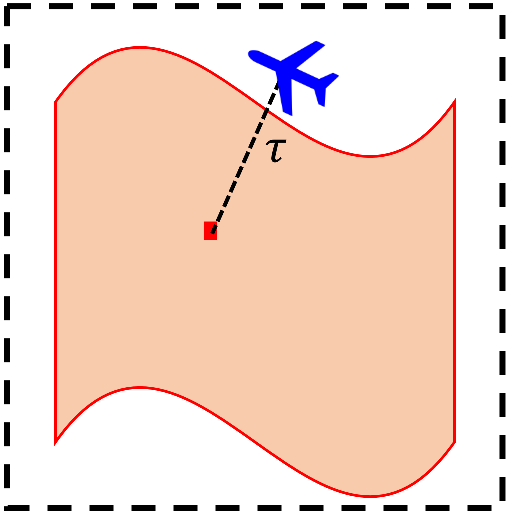
Old Observables in a New Language

***N*-subjettiness** is the EMD between the event and the closest N-particle event.

$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\text{ axes}} \sum_{i=1}^{M} E_i \min\{\theta_{1,i}^{\beta}, \theta_{2,i}^{\beta}, \dots, \theta_{N,i}^{\beta}\} \longrightarrow \tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \text{EMD}(\mathcal{E},\mathcal{E}').$$

$\beta \geq 1$ is p-Wasserstein distance with $\text{p} = \beta$.





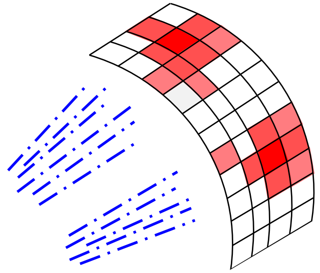
Geometry in the space of events

Thrust is the EMD between the event and two back-to-back particles.

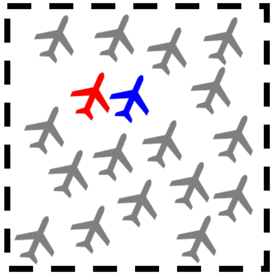
$$t(\mathcal{E}) = E - \max_{\hat{n}} \sum_{i} |\vec{p}_i \cdot \hat{n}| \longrightarrow t(\mathcal{E}) = \min_{|\mathcal{E}'|=2} \text{EMD}(\mathcal{E},\mathcal{E}')$$

with $\theta_{ij} = \hat{n}_i \cdot \hat{n}_j, \;\; \hat{n} = \vec{p}/E$

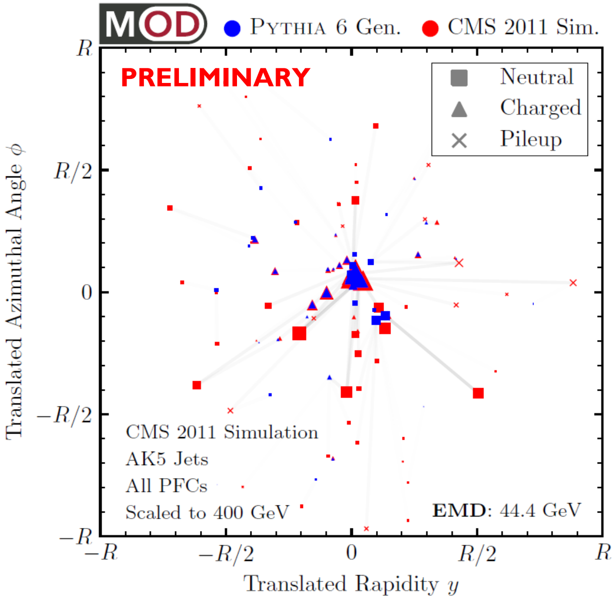
Quantifying Pileup and Detector Effects with EMD



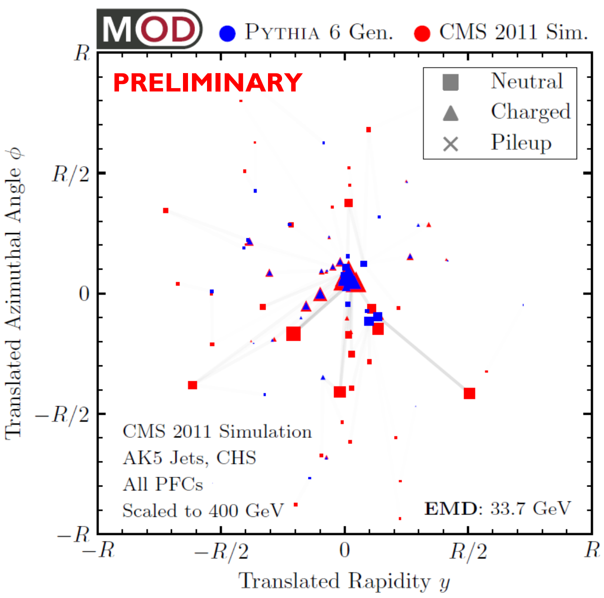
EMD universally quantifies pileup and detector effects.



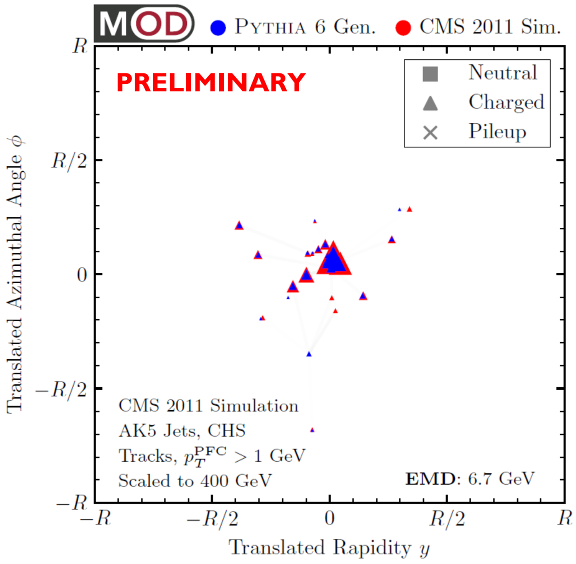
Gen./Sim. EMD: 44.4 GeV



Gen./Sim. EMD: 33.7 GeV



Gen./Sim. EMD: 6.7 GeV



+ charged hadron subtraction

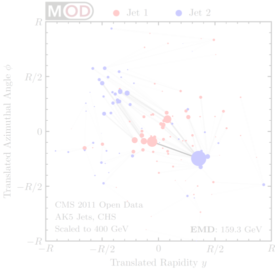
+ Tracks only, $p_T^{\mathrm{PFC}} > 1$ GeV cut

See extra slides for histograms. Can also quantify hadronization effects this way.

Outline

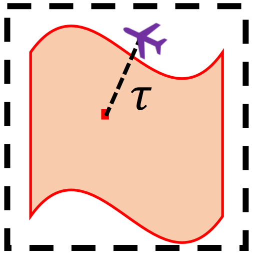
When are two jets similar?

When they have similar distributions of energy



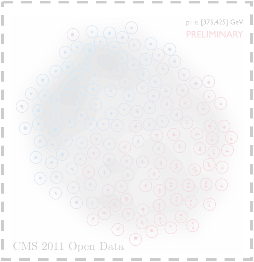
Energy Mover's Distance

The "work" to rearrange one jet into another



Quantifying Jet Similarity

Geometry of the space of jets. Bounds for pileup, detector effects

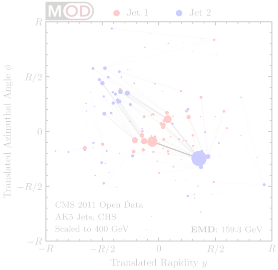


Exploring the Space of Jets

Outline

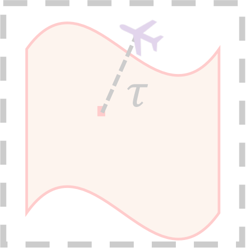
When are two jets similar?

When they have similar distributions of energy



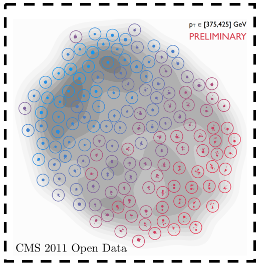
Energy Mover's Distance

The "work" to rearrange one jet into another



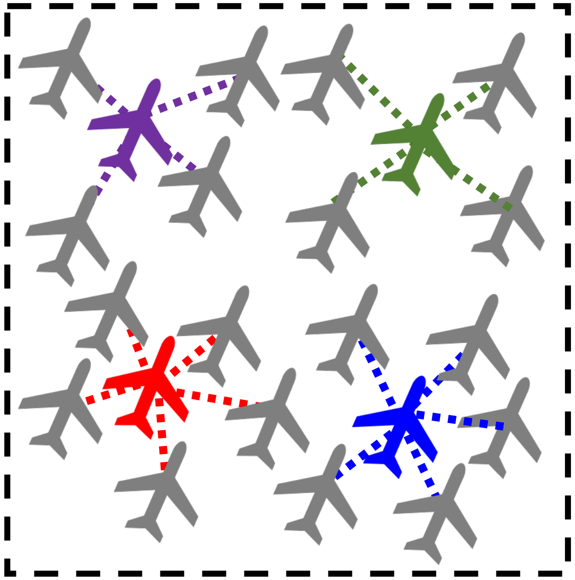
Quantifying Jet Similarity

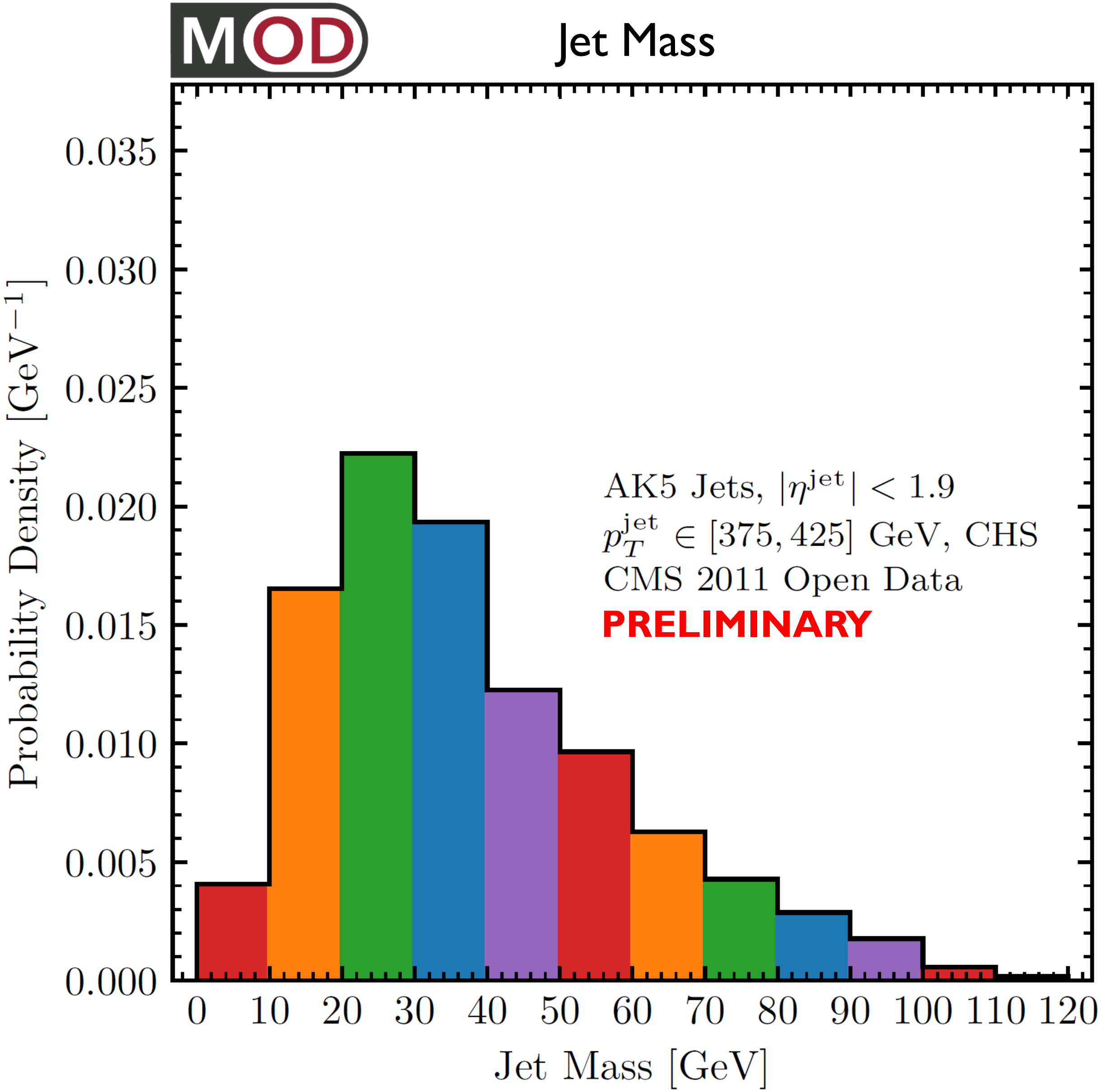
Geometry of the space of jets. Bounds for pileup, detector effects



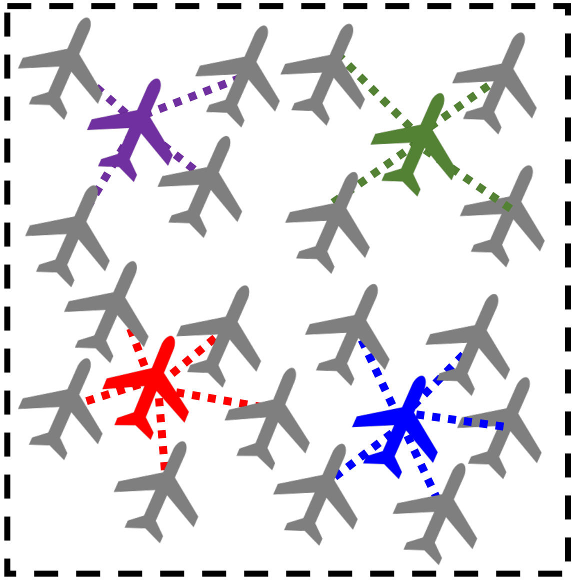
Exploring the Space of Jets

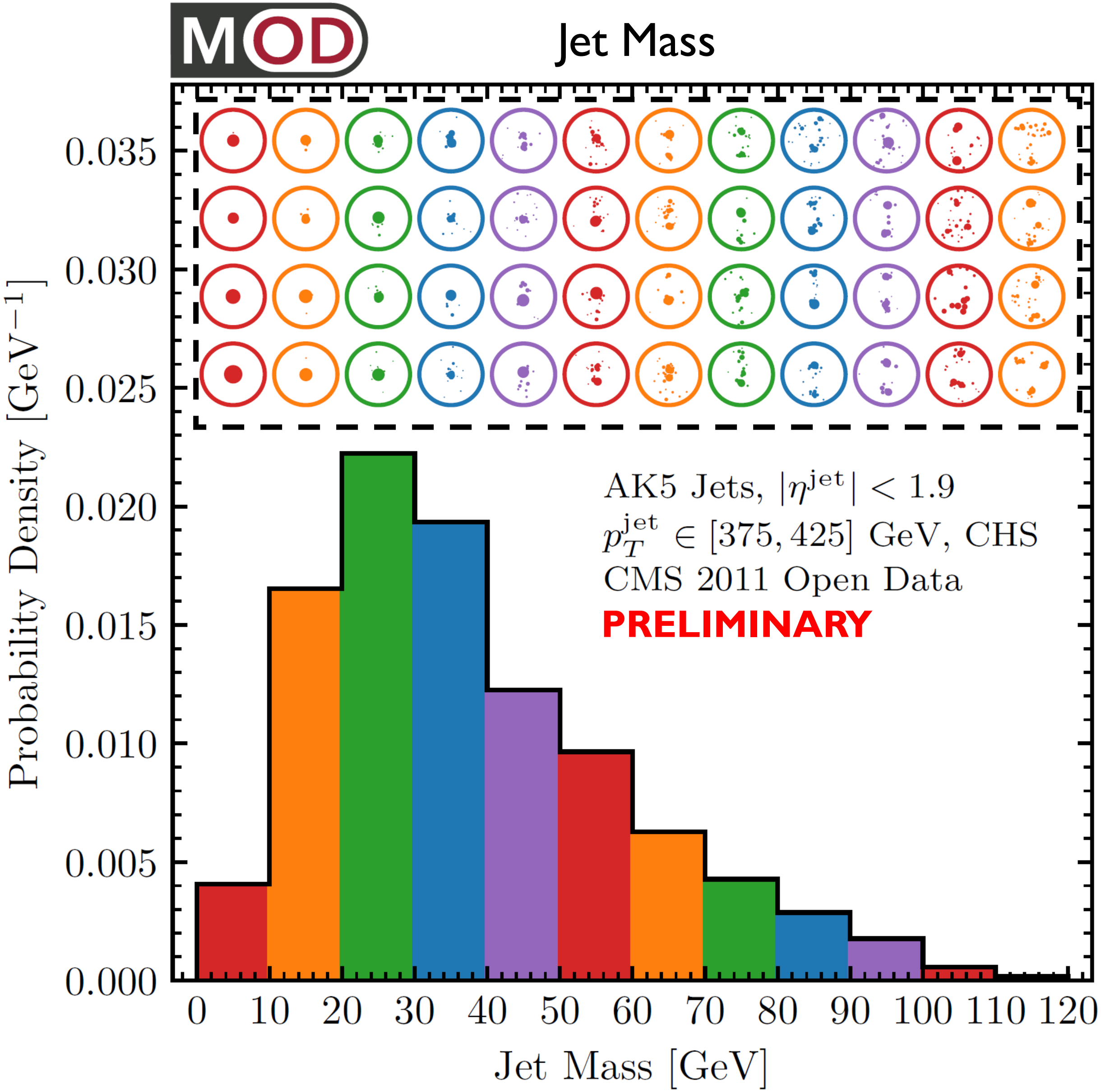
Most Representative Jets: K-medoids



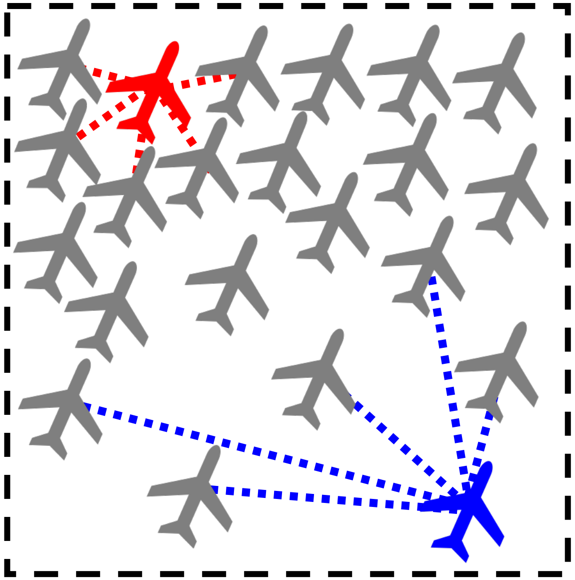


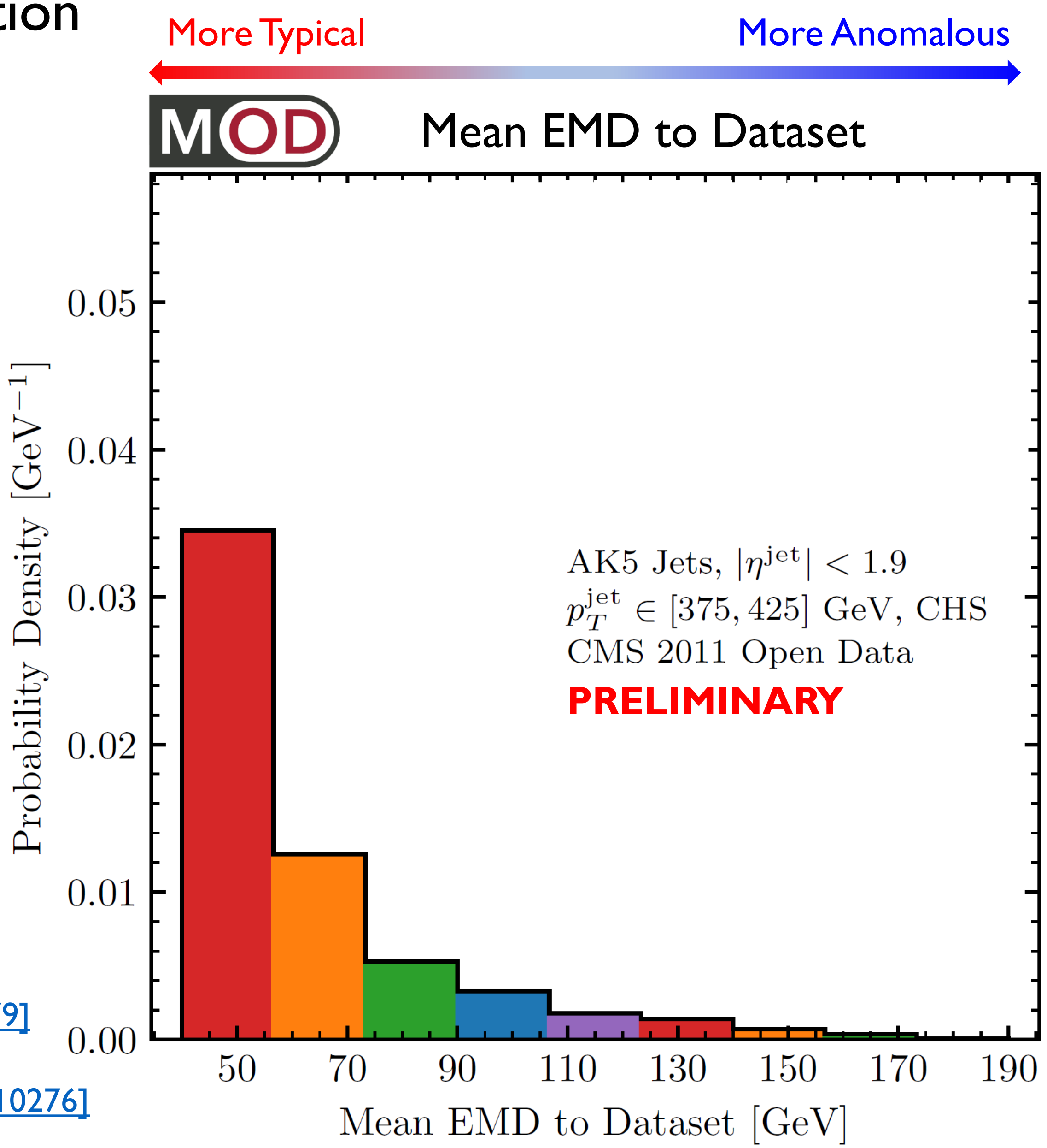
Most Representative Jets: K-medoids





Towards Anomaly Detection

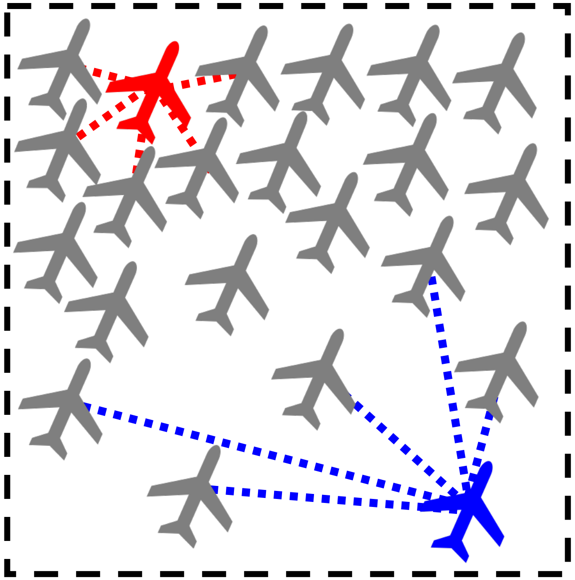


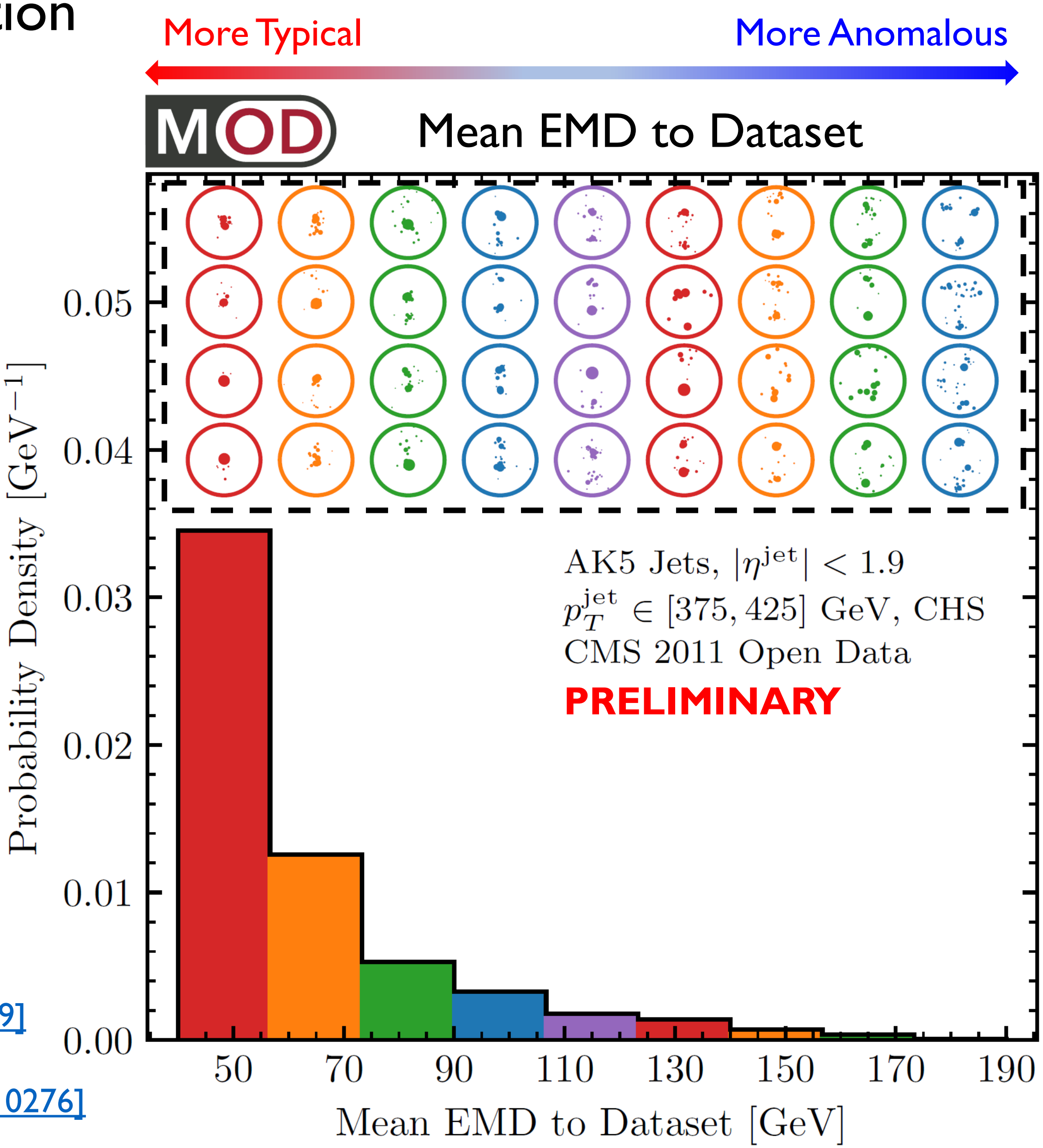


Complements recent developments in anomaly detection for collider physics.

[Collins, Howe, Nachman, 1805.02664]
[Heimel, Kasieczka, Plehn, Thompson, 1808.08979]
[Farina, Nakai, Shih, 1808.08992]
[Cerri, Nguyen, Pierini, Spiropulu, Vlimant, 1811.10276]

Towards Anomaly Detection





Complements recent developments in anomaly detection for collider physics.

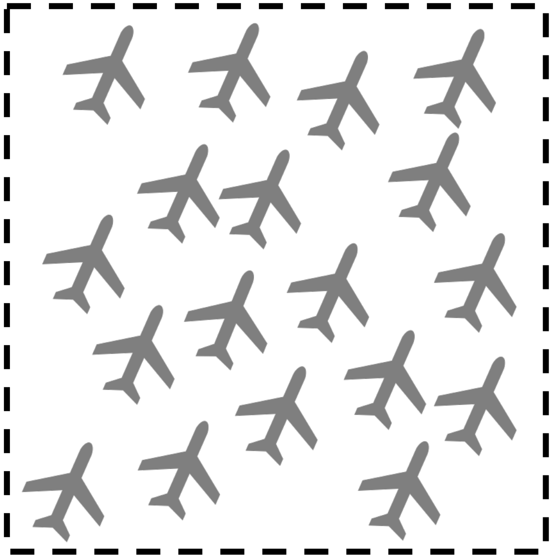
[Collins, Howe, Nachman, 1805.02664]
[Heimel, Kasieczka, Plehn, Thompson, 1808.08979]
[Farina, Nakai, Shih, 1808.08992]
[Cerri, Nguyen, Pierini, Spiropulu, Vlimant, 1811.10276]

Exploring the Space of Jets: Visualizing the Manifold

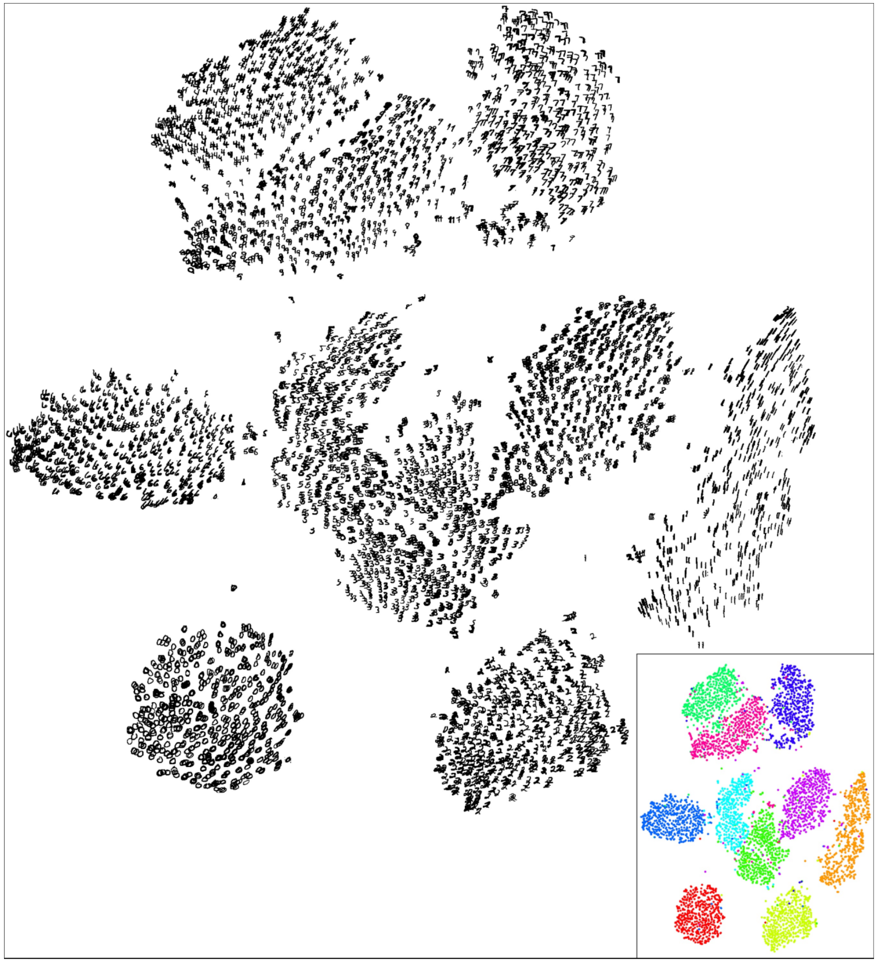
Visualize the space of events with t-Distributed Stochastic Neighbor Embedding (t-SNE).

[L. van der Maaten, G. Hinton]

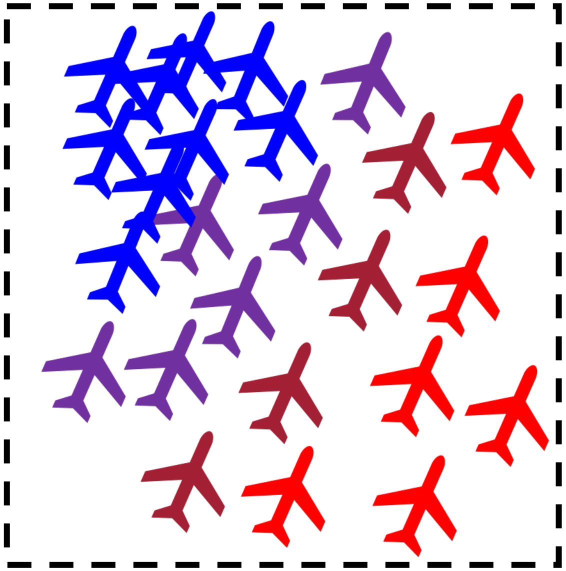
Finds an embedding into a low-dimensional manifold that respects distances.



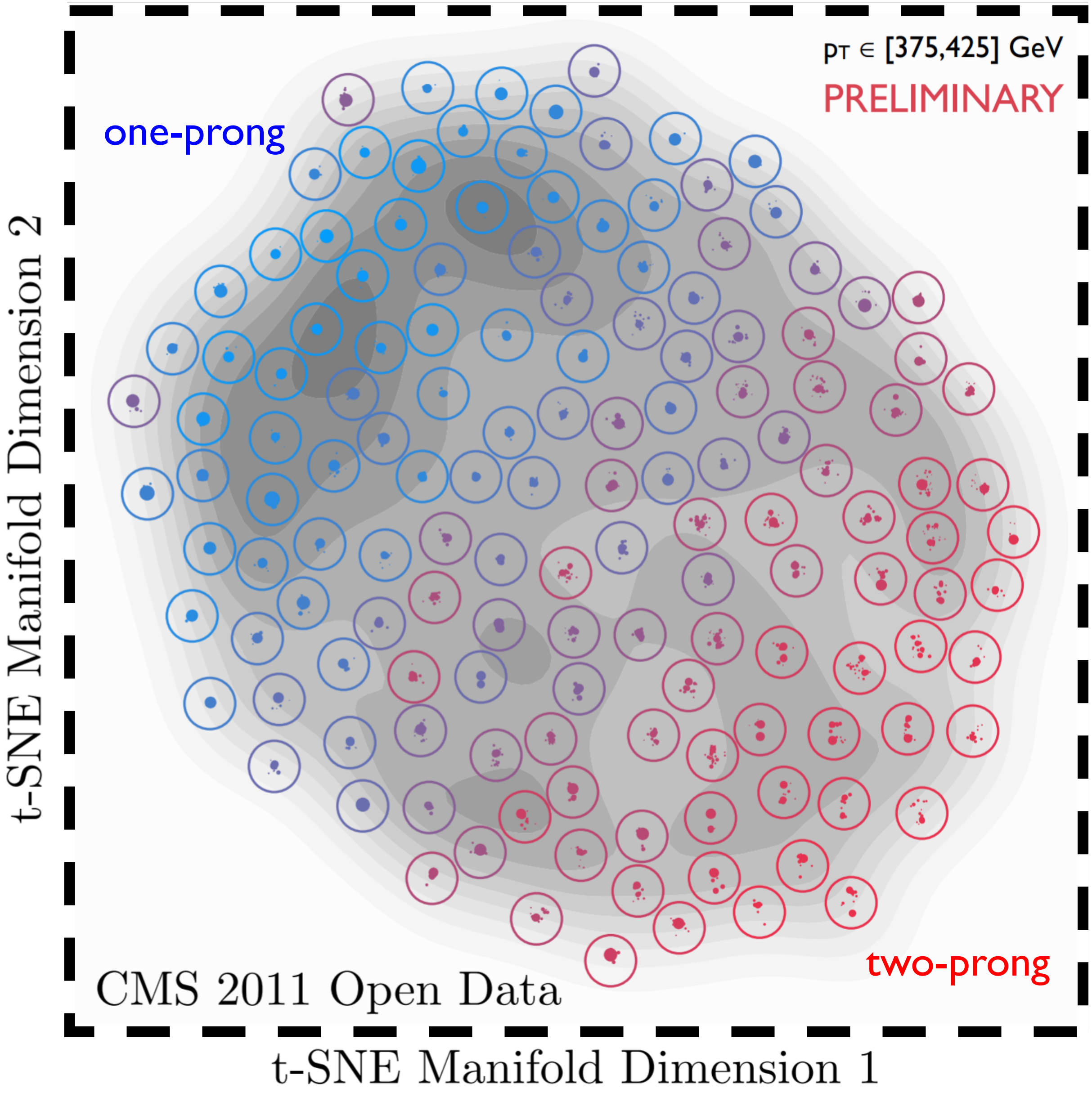
What does the space of jets look like?



Exploring the Space of Jets: Visualizing the Manifold



What does the space of jets look like?



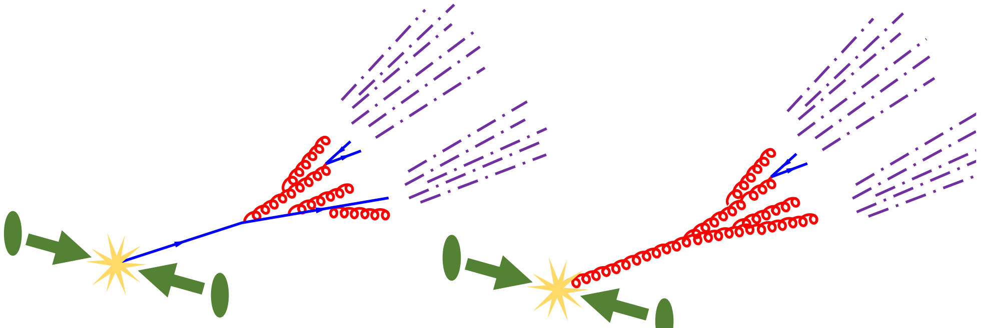
Exploring the Space of Jets: Correlation Dimension

Correlation dimension:

$$\boxed{\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_{i=1}^{N}\sum_{j=1}^{N}\Theta[\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}_j) < Q]}$$

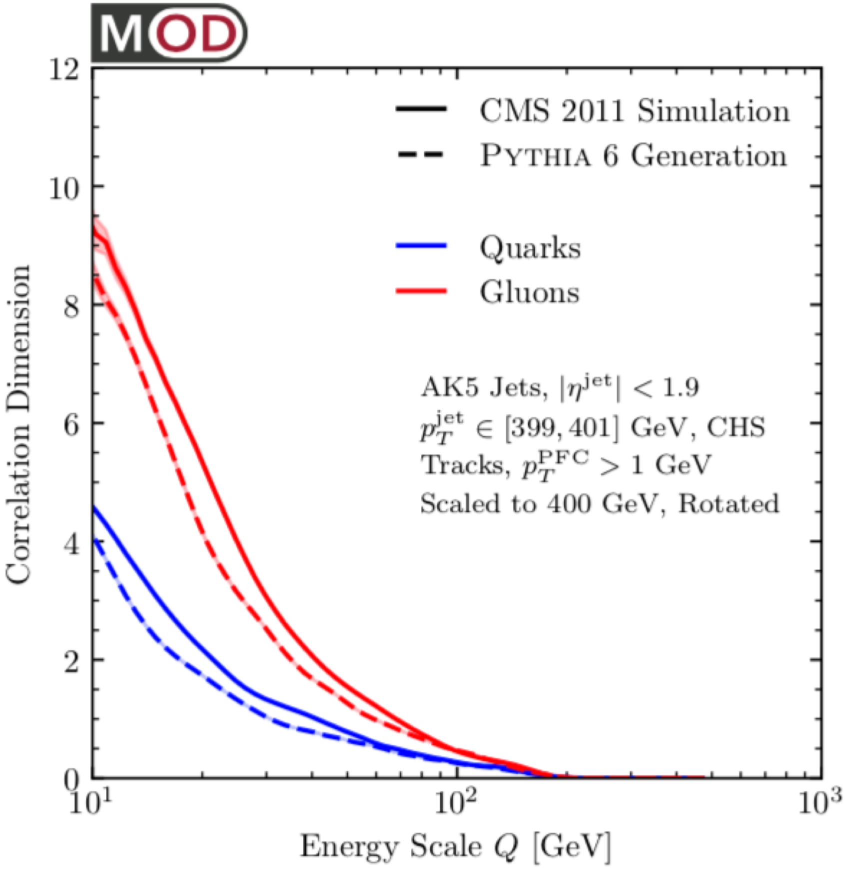
Energy scale Q dependence

Count neighbors in ball of radius Q



Quark jets

Gluon jets

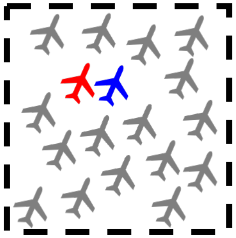


$$\boxed{\text{At LL:}\ \dim_{q/g}(Q) = -\frac{8\alpha_s C_{q/g}}{\pi}\ln\frac{Q}{p_T/2}}$$

Dimension blows up at low energies.

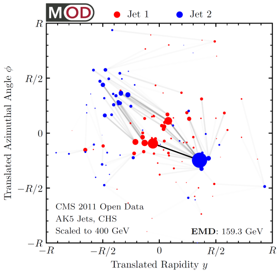
Jets are "more than fractal"

Outline



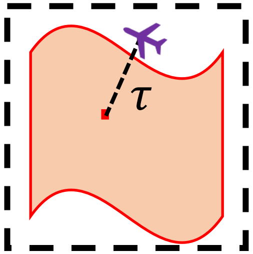
When are two jets similar?

When they have similar distributions of energy



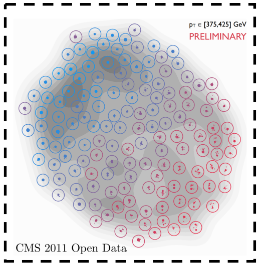
Energy Mover's Distance

The "work" to rearrange one jet into another



Quantifying Jet Similarity

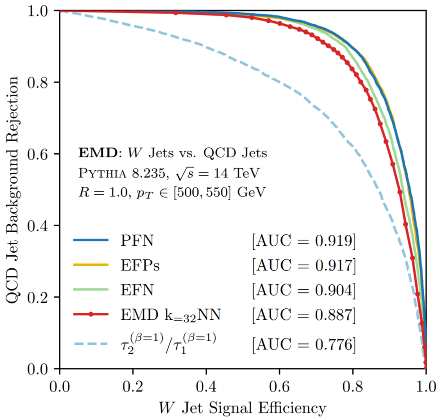
Geometry of the space of jets. Bounds for pileup, detector effects

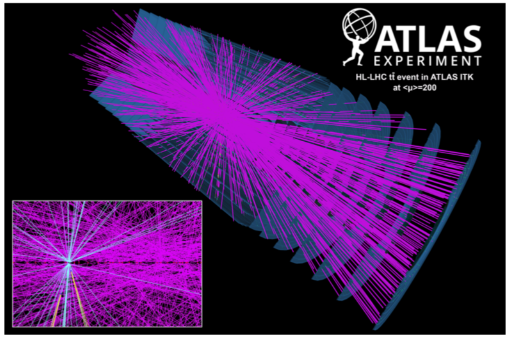


Exploring the Space of Jets

New ways to visualize and probe jet data

Going Beyond





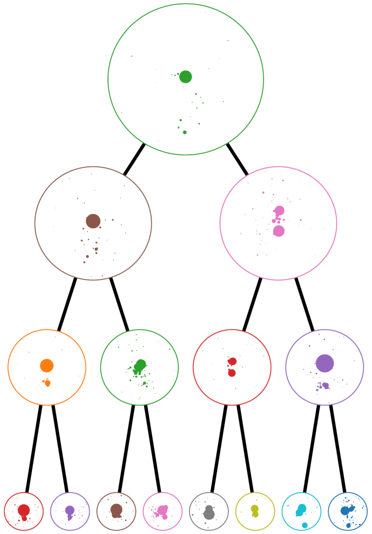
Classification with EMD

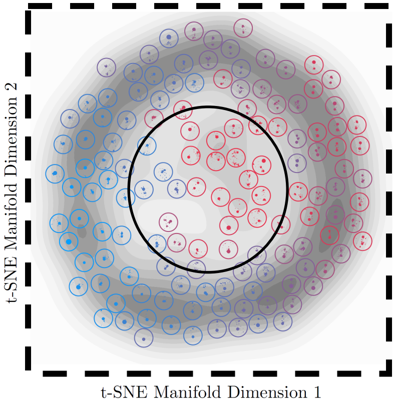
Clustering sets of events

New observables through EMD geometry?

“Event” mover’s distance between ensembles?

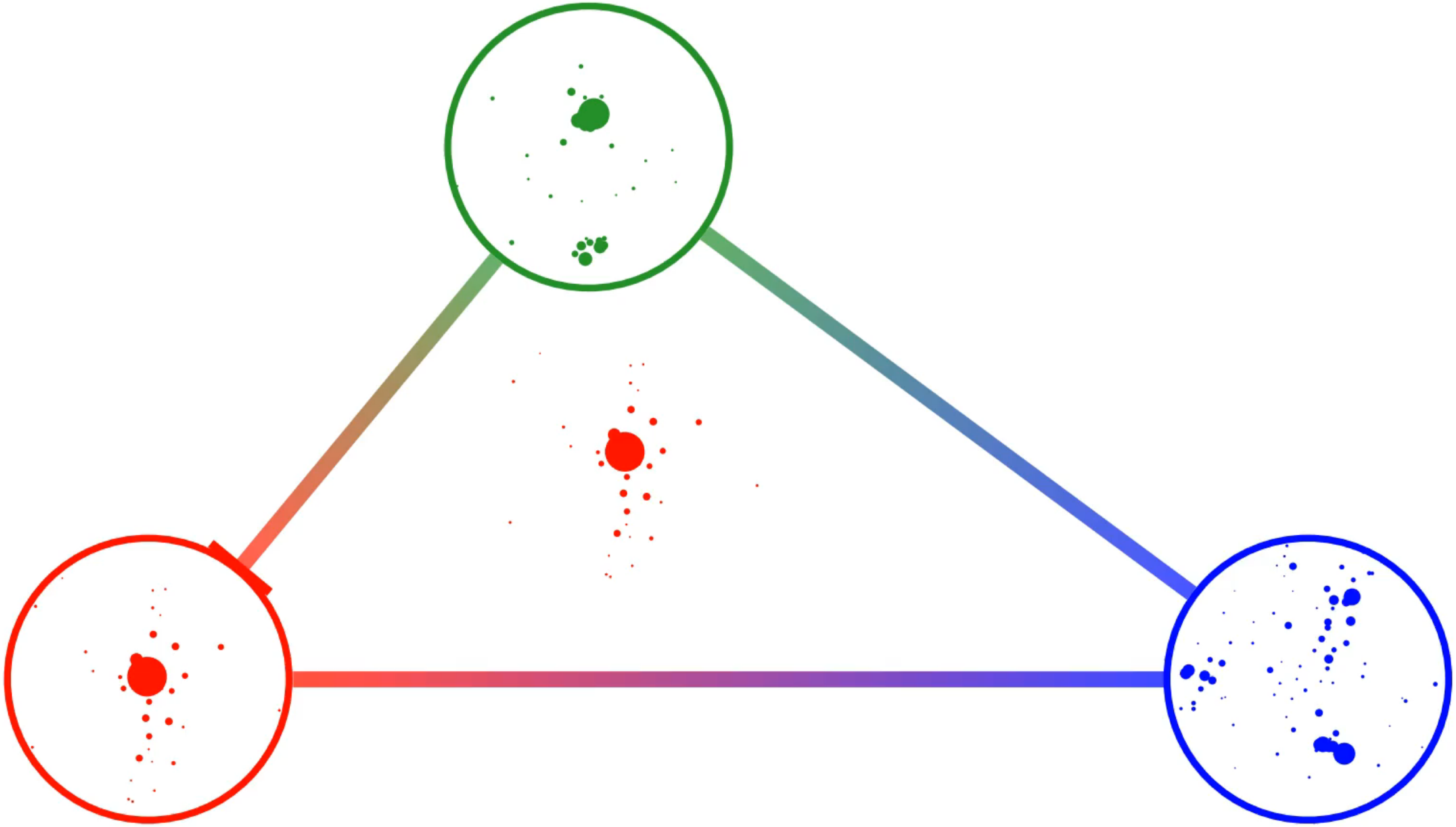
Include flavor information?



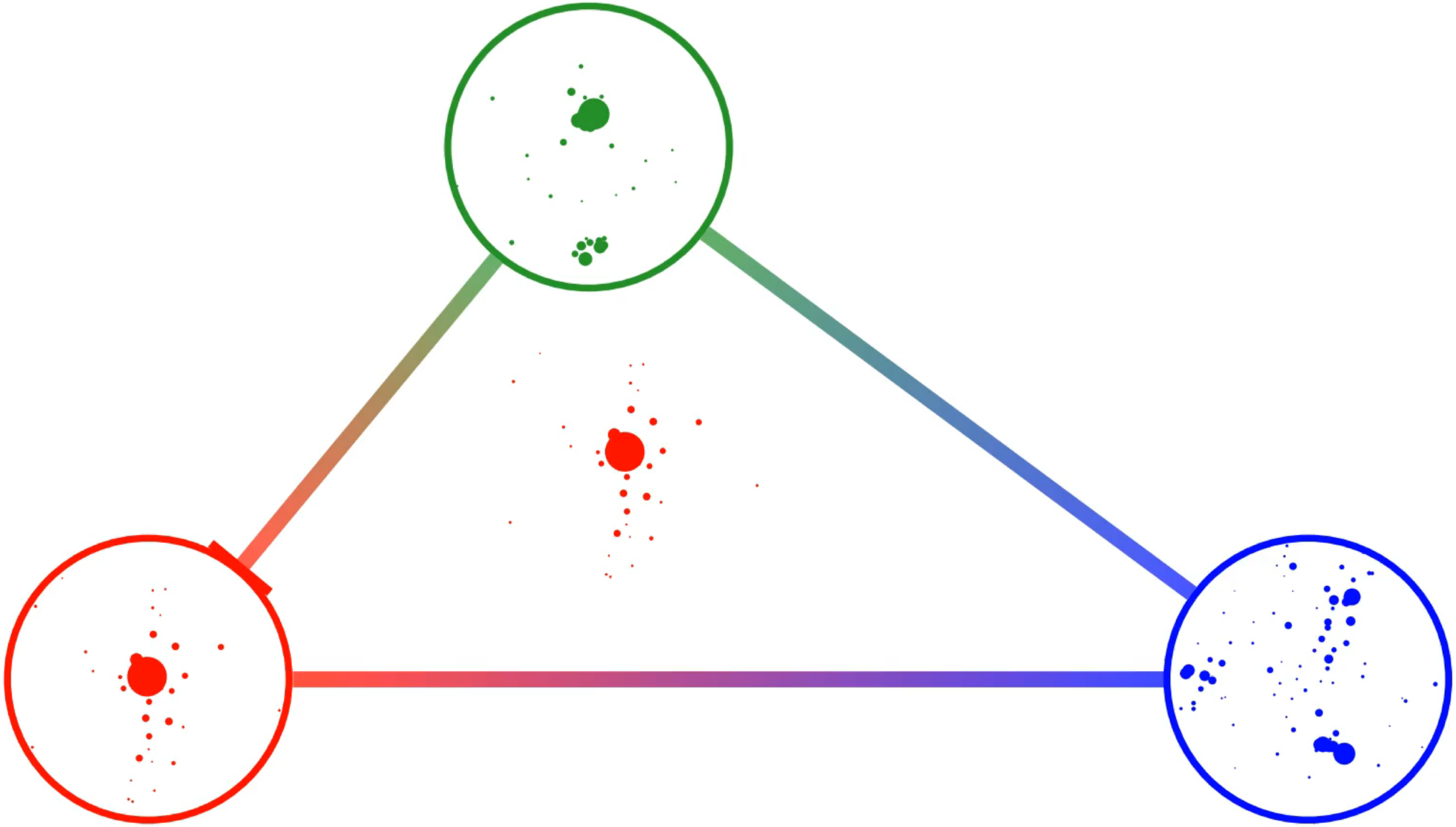


The End

Thank you!



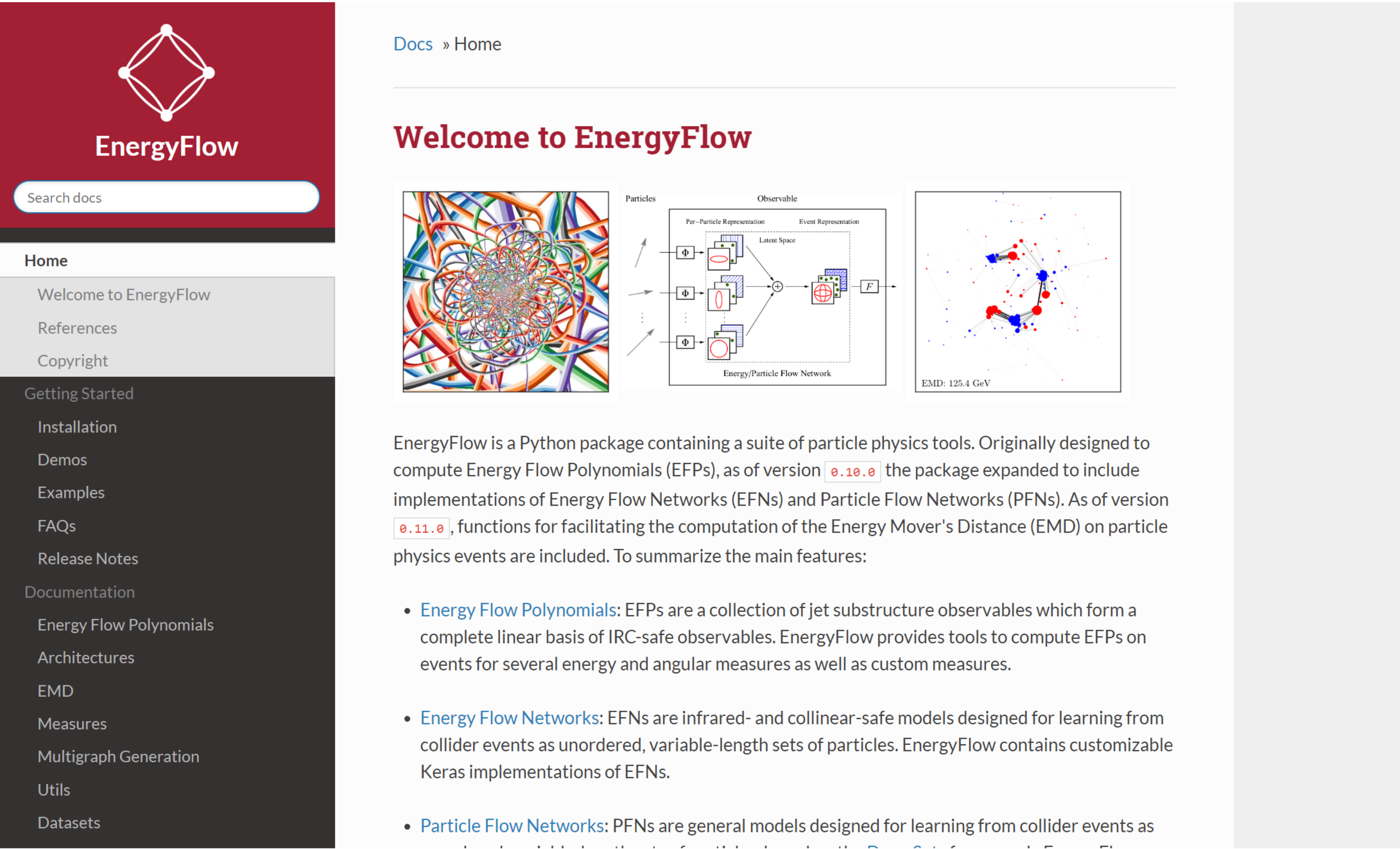
Extra Slides



EnergyFlow

https://energyflow.network

```
pip install energyflow
```



EnergyFlow

Search docs

- Home
  - Welcome to EnergyFlow
  - References
  - Copyright
- Getting Started
  - Installation
  - Demos
  - Examples
  - FAQs
  - Release Notes
- Documentation
  - Energy Flow Polynomials
  - Architectures
  - EMD
  - Measures
  - Multigraph Generation
  - Utils
  - Datasets

Docs » Home

Welcome to EnergyFlow

EnergyFlow is a Python package containing a suite of particle physics tools. Originally designed to compute Energy Flow Polynomials (EFPs), as of version `0.10.0` the package expanded to include implementations of Energy Flow Networks (EFNs) and Particle Flow Networks (PFNs). As of version `0.11.0`, functions for facilitating the computation of the Energy Mover's Distance (EMD) on particle physics events are included. To summarize the main features:

- Energy Flow Polynomials: EFPs are a collection of jet substructure observables which form a complete linear basis of IRC-safe observables. EnergyFlow provides tools to compute EFPs on events for several energy and angular measures as well as custom measures.

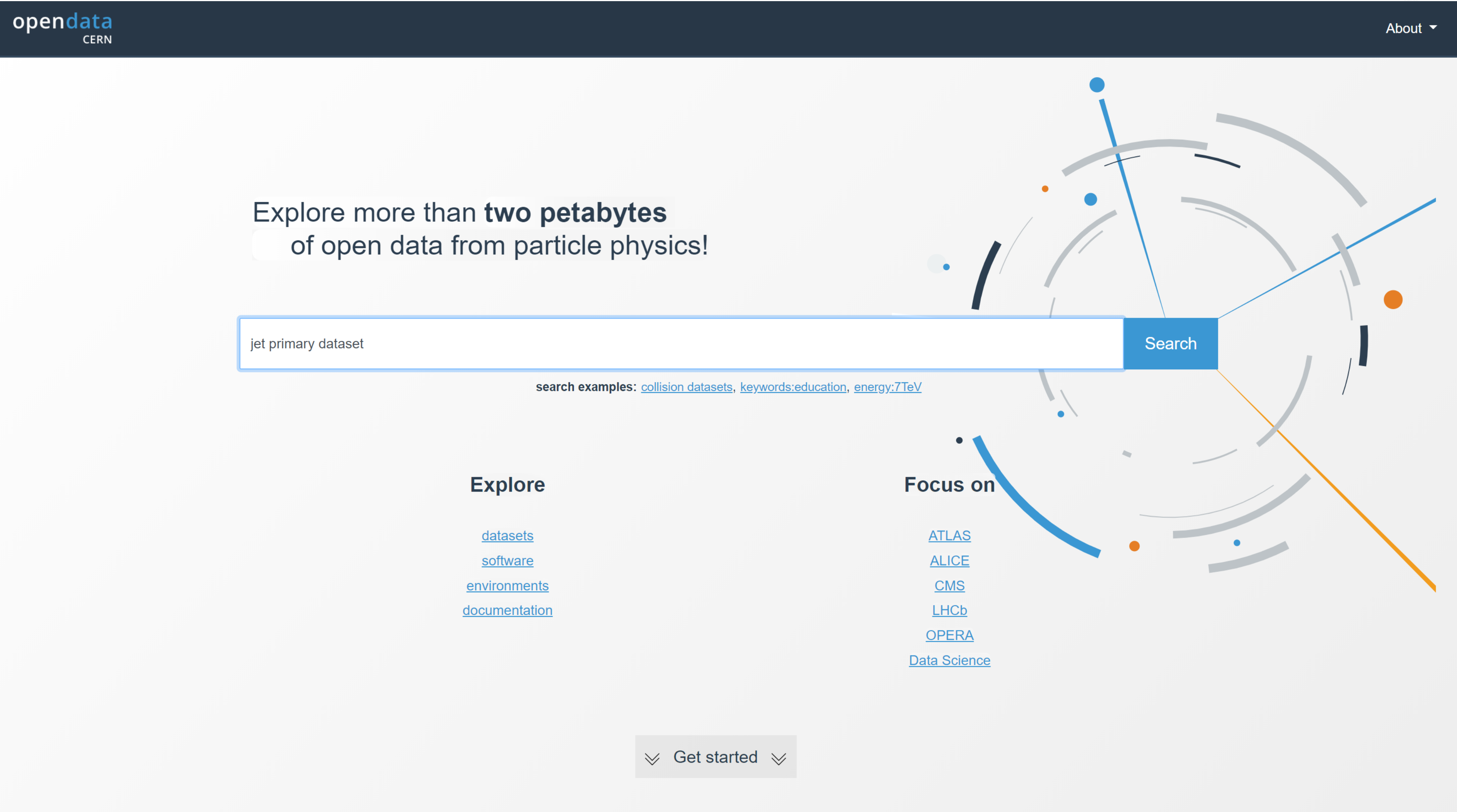
- Energy Flow Networks: EFNs are infrared- and collinear-safe models designed for learning from collider events as unordered, variable-length sets of particles. EnergyFlow contains customizable Keras implementations of EFNs.

- Particle Flow Networks: PFNs are general models designed for learning from collider events as

CMS Open Data

An amazing resource for physics exploration and proof-of-principle studies.

opendata.cern.ch



opendata CERN

About

Explore more than **two petabytes** of open data from particle physics!

jet primary dataset

Search

search examples: collision datasets, keywords:education, energy:7TeV

Explore

- datasets
- software
- environments
- documentation

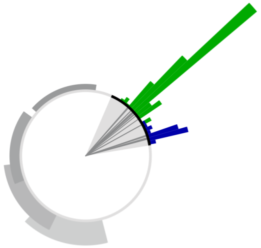
Focus on

- ATLAS
- ALICE
- CMS
- LHCb
- OPERA
- Data Science

Get started

CMS Open Data

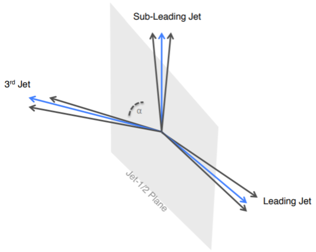
Many exciting physics applications with the CMS Open Data already.



Exposing the QCD splitting function

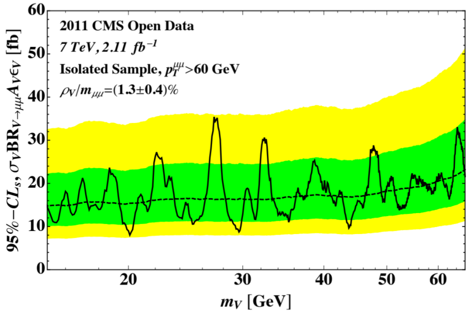
[Tripathee, Xue, Larkoski, Marzani, Thaler, 1704.05842]

[Larkoski, Marzani, Thaler, Tripathee, Xue, 1704.05066]



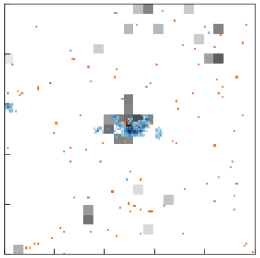
Looking for parity violation in jets

[Lester, Schott, 1904.11195]



Searching for dimuon resonances

[Cesarotti, Soreq, Strassler, Thaler, Xue, 1902.04222]



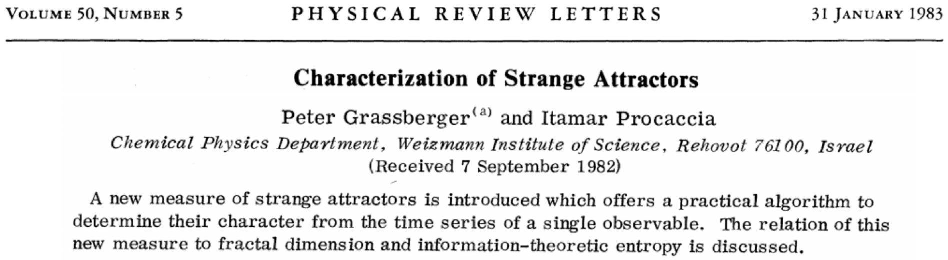
Analyzing collision data with deep learning techniques

[Madrazo, Cacha, Iglesias, de Lucas, 1708.07034]

[Andrews, Paulini, Gleyzer, Poczos, 1807.11916]

[Andrews, et al., 1902.08276]

Exploring the Space of Jets: Correlation Dimension



VOLUME 50, NUMBER 5 PHYSICAL REVIEW LETTERS 31 JANUARY 1983

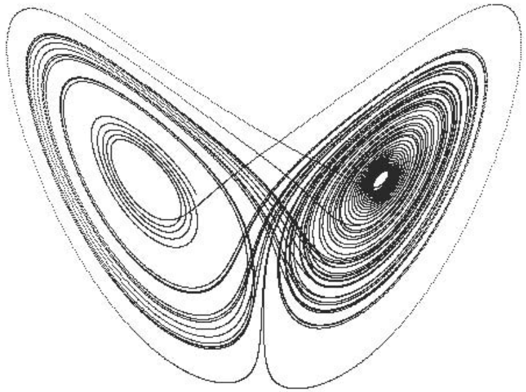
Characterization of Strange Attractors

Peter Grassberger[(a)] and Itamar Procaccia

Chemical Physics Department, Weizmann Institute of Science, Rehovot 76100, Israel

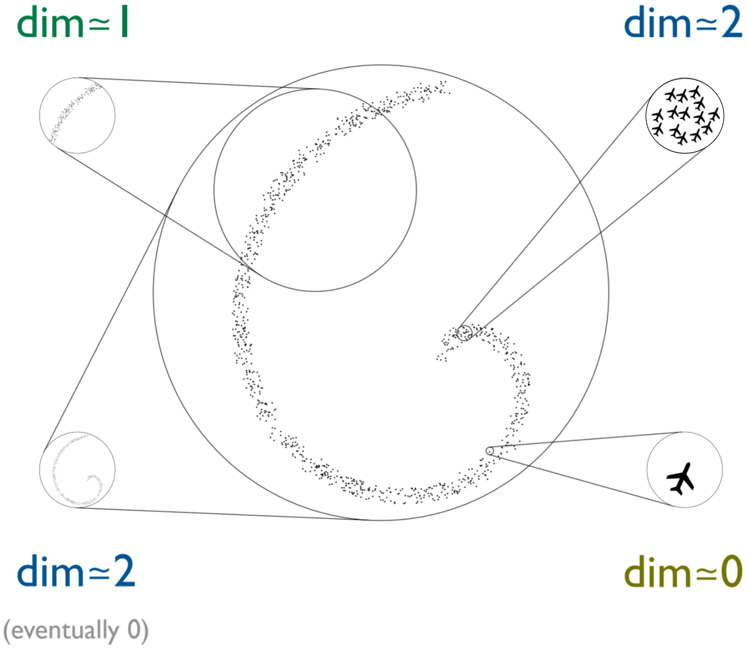
(Received 7 September 1982)

A new measure of strange attractors is introduced which offers a practical algorithm to determine their character from the time series of a single observable. The relation of this new measure to fractal dimension and information-theoretic entropy is discussed.



Intuition:

$$N_{\substack{\text{neighboring}\\ \text{points}}}(r) \propto r^{\text{dim}} \quad \Longrightarrow \quad \text{dim}(r) = \text{r}\frac{\partial}{\partial r}\ln N_{\text{neighbors}}(r)$$



Correlation dimension:

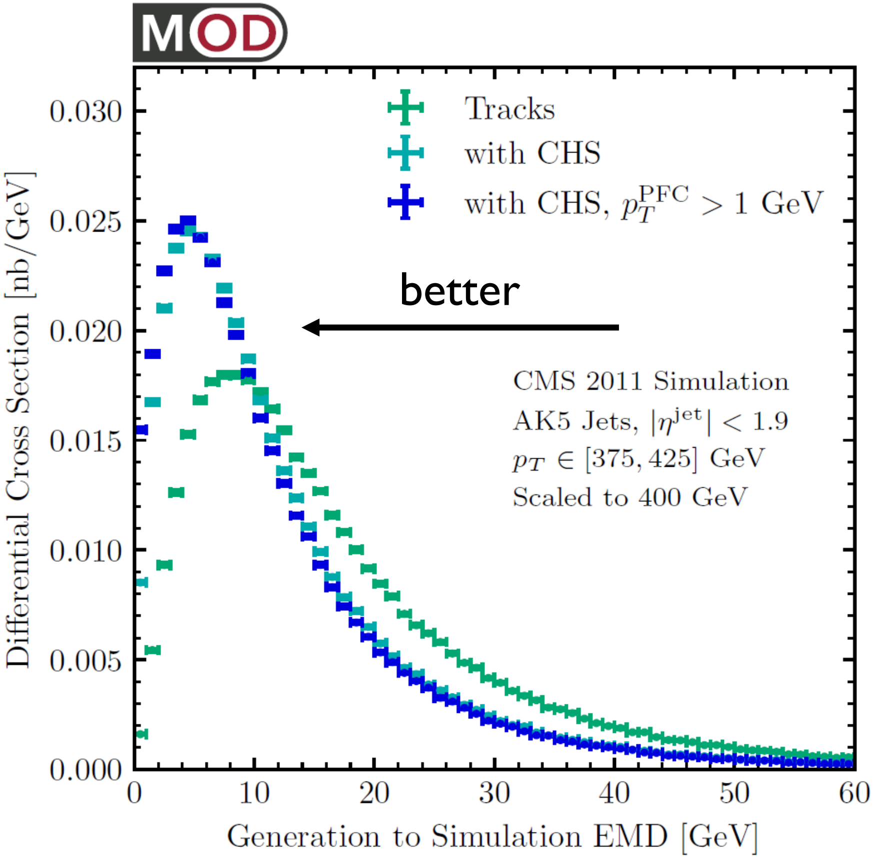
$$\boxed{\text{dim}(Q) = Q\frac{\partial}{\partial Q}\ln\sum_{i=1}^{N}\sum_{j=1}^{N}\Theta[\text{EMD}(\mathcal{E}_i,\mathcal{E}_j) < Q]}$$

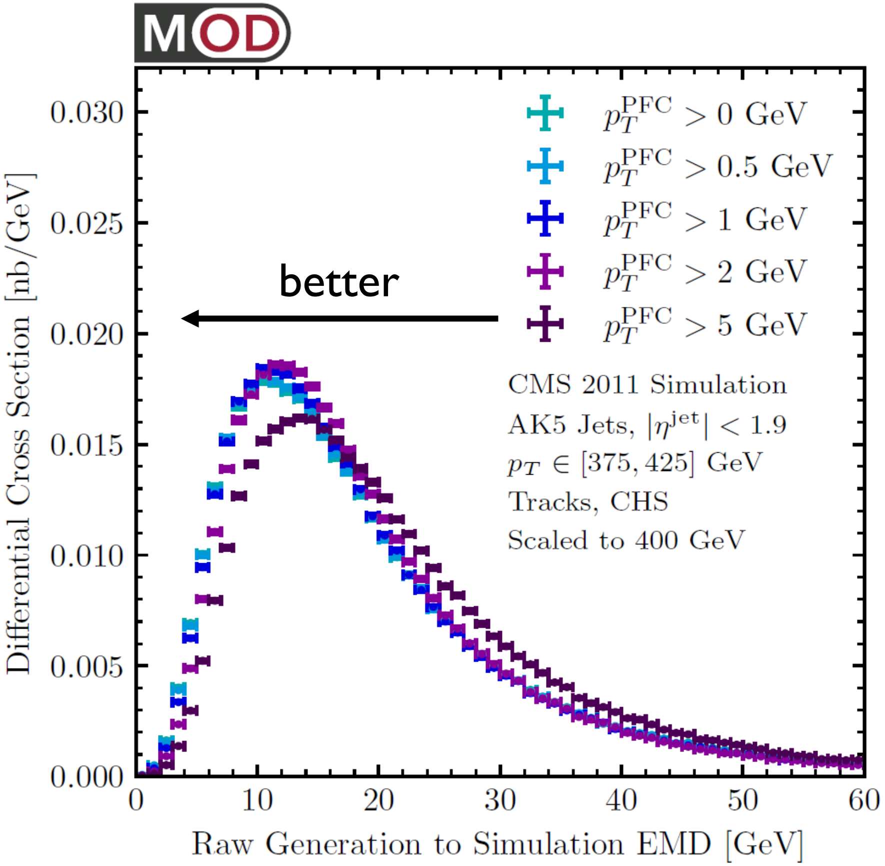
Energy scale Q dependence

Count neighbors in ball of radius Q

Quantifying Pileup and Detector Effects with EMD

Gen./Sim. EMD universally quantifies pileup mitigation and detector effects.

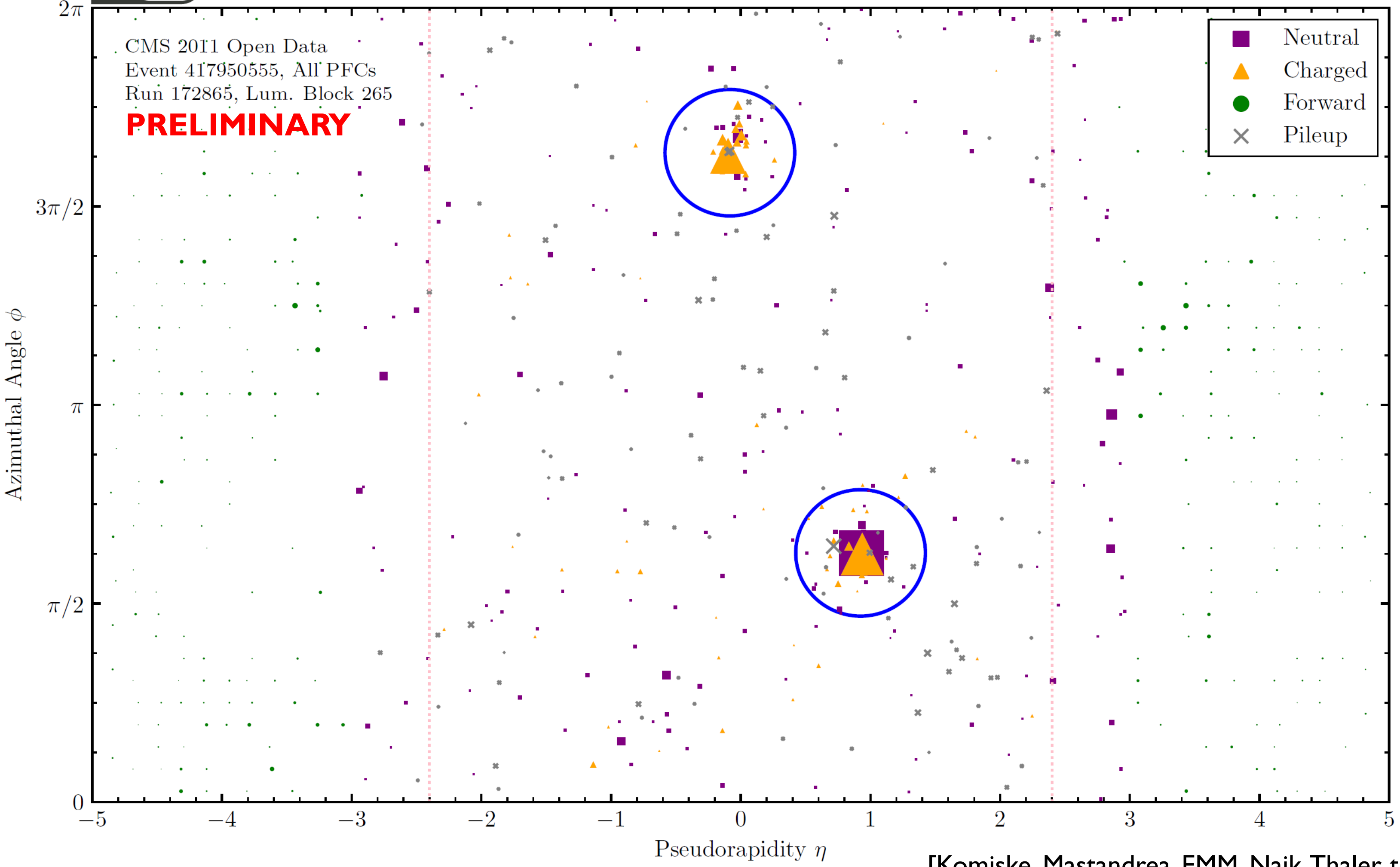




CMS 2011A Jet Primary Dataset (+ Simulation)

$2.3\ \mathrm{fb}^{-1}$ of 7 TeV proton-proton collision data. [link]

~ 1 million $R = 0.5$ jets with $p_T \in [375{,}425]\mathrm{GeV}, |\eta| < 1.9$



[Komiske, Mastandrea, EMM, Naik, Thaler, to appear]

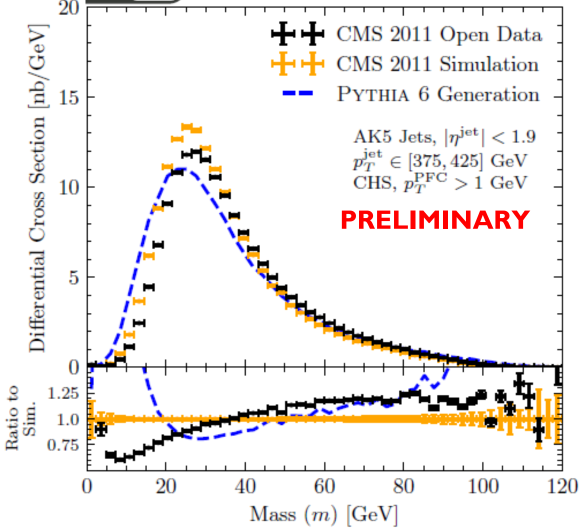
Jet Substructure Observables

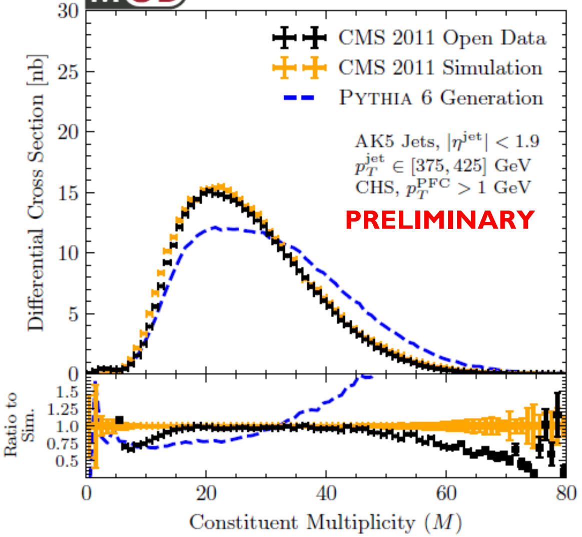
Study jet substructure at truth and detector level.

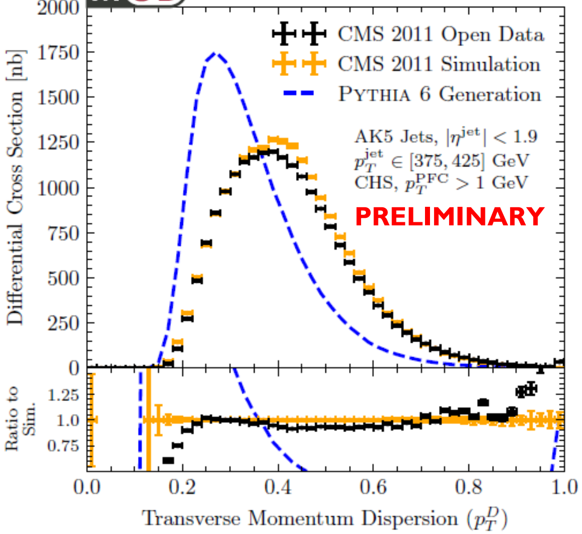
$$m^2 = \left(\sum_{i\in\text{Jet}} p_i^\mu\right)^2$$

$$M = \sum_{i\in\text{Jet}} 1$$

$$p_T^D = \frac{\sum_{i\in\text{Jet}} p_{T,i}^2}{\left(\sum_{i\in\text{Jet}} p_{T,i}\right)^2}$$

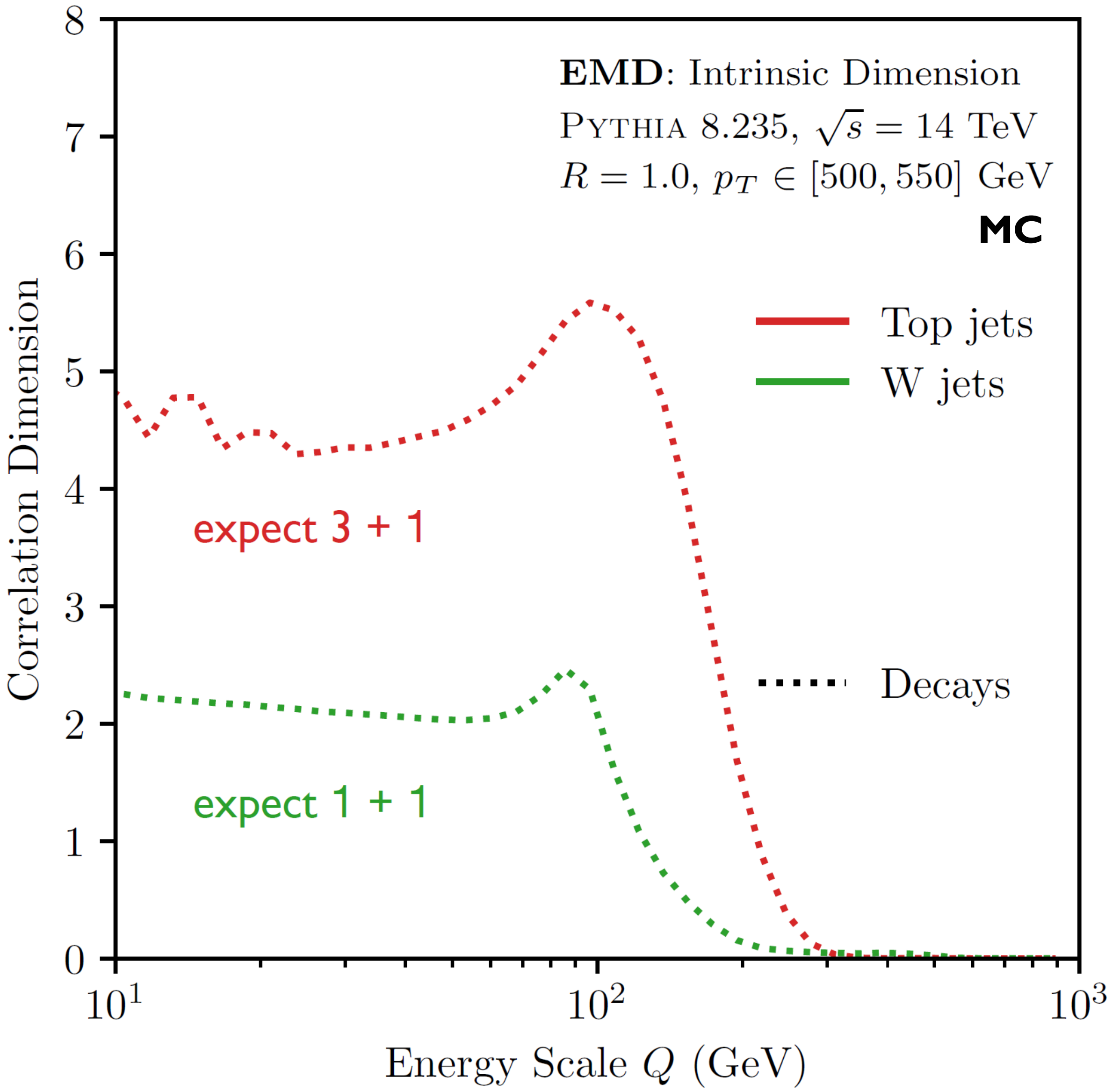

Jet Mass

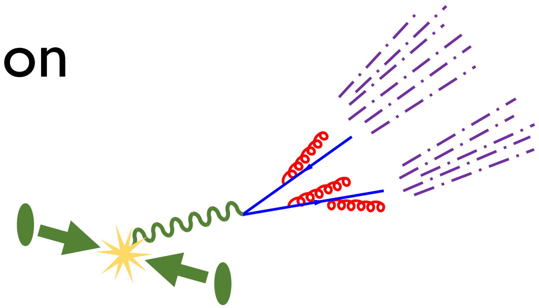

Constituent Multiplicity


Momentum Dispersion

Similar to: [Larkoski, Marzani, Thaler, Tripathee, Xue, 1704.05066]

Exploring the Space of Jets: Correlation Dimension





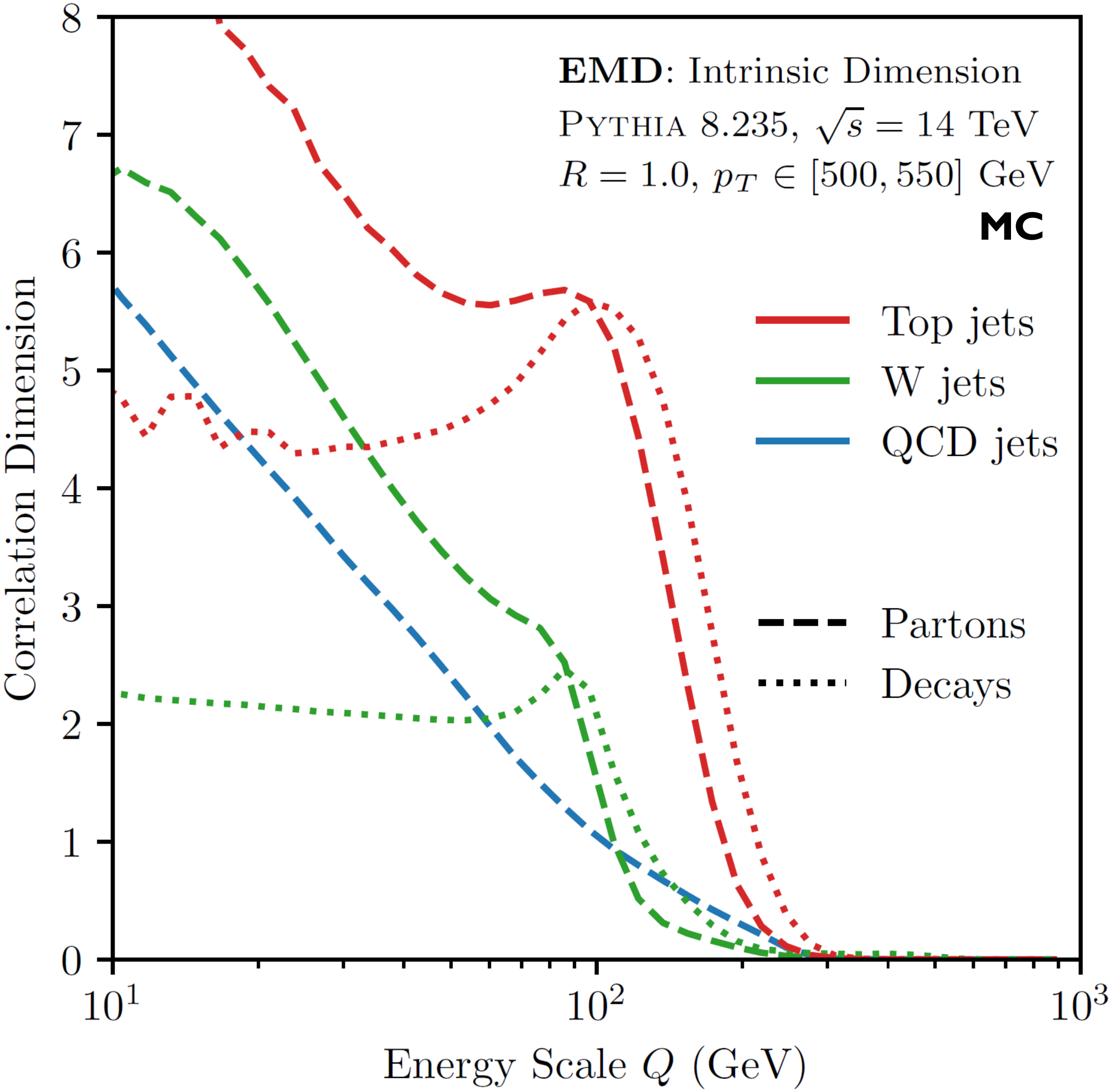
QCD jets are simplest.

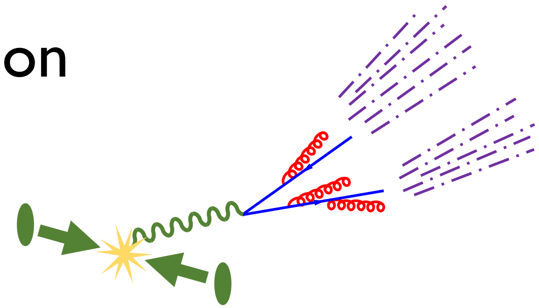
W jets are more complicated.

Top jets are most complex.

“Decays” have ~constant dimension.

Exploring the Space of Jets: Correlation Dimension





QCD jets are simplest.

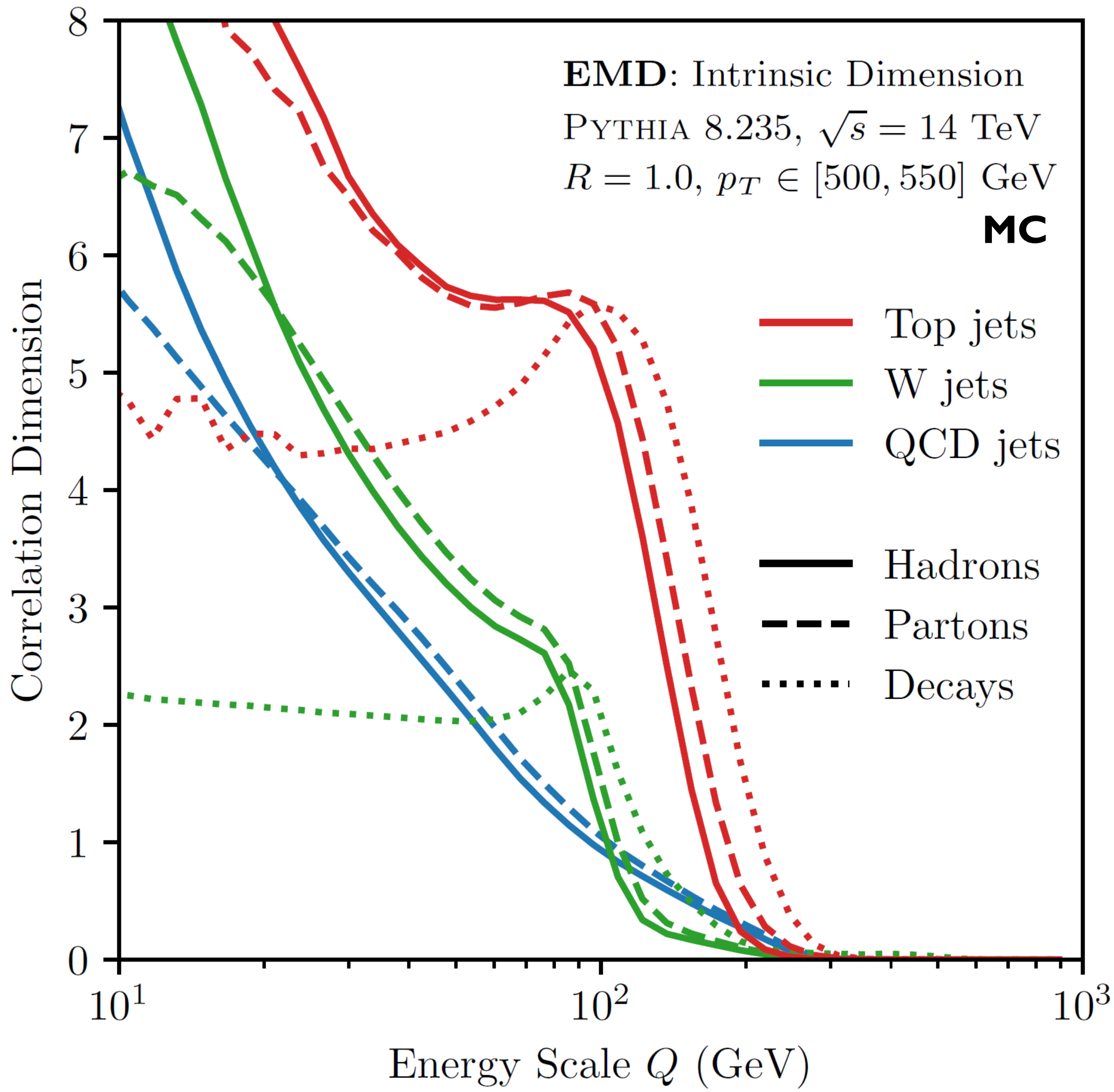
W jets are more complicated.

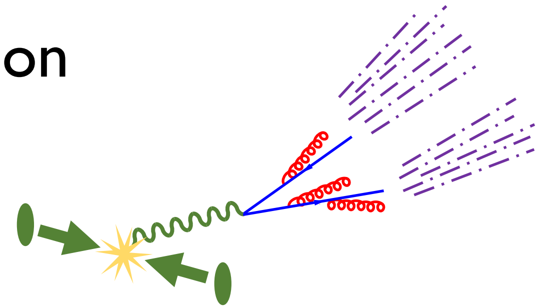
Top jets are most complex.

“Decays” have ~constant dimension.

Fragmentation becomes more complex at lower energy scales.

Exploring the Space of Jets: Correlation Dimension





QCD jets are simplest.

W jets are more complicated.

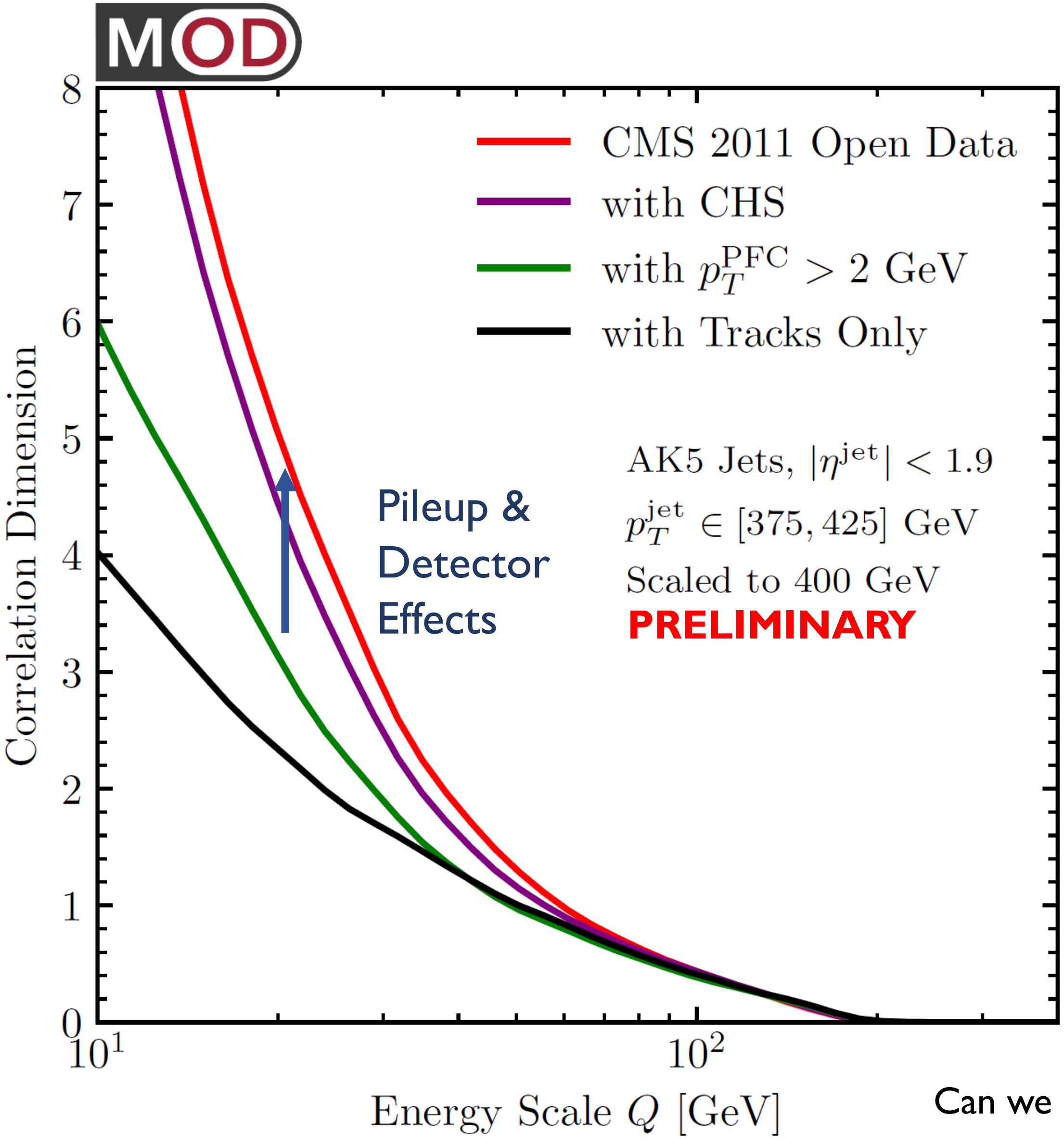
Top jets are most complex.

"Decays" have ~constant dimension.

Fragmentation becomes more complex at lower energy scales.

Hadronization becomes relevant at scales around 20 GeV.

Exploring the Space of Jets: Correlation Dimension

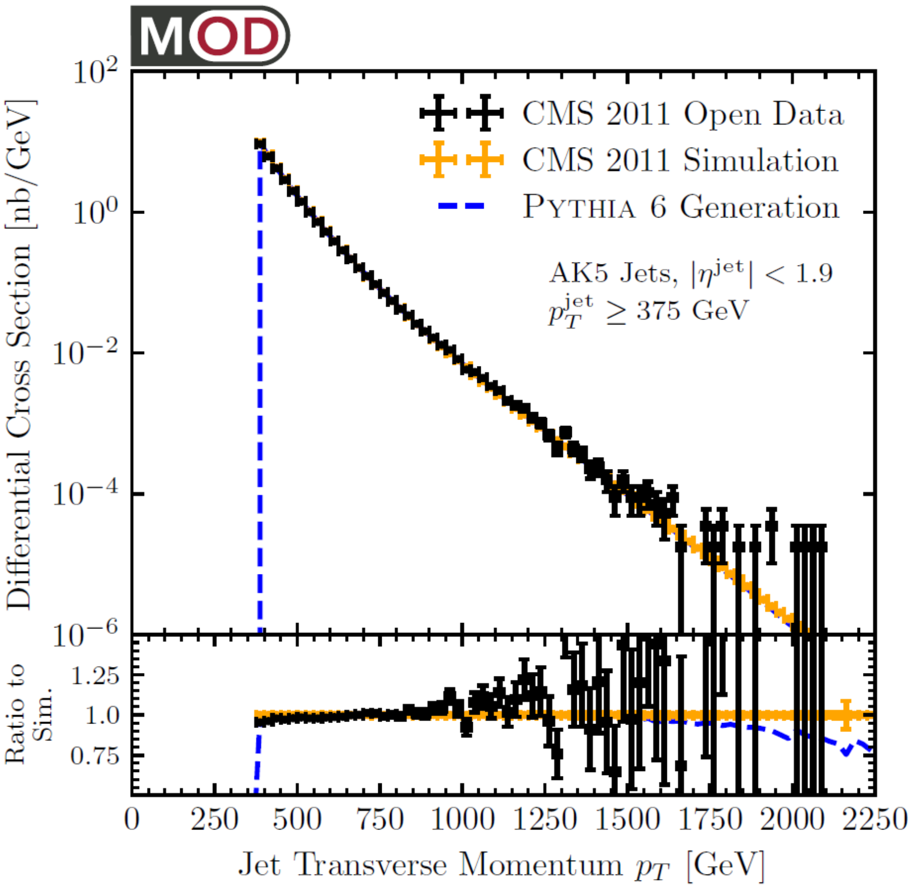


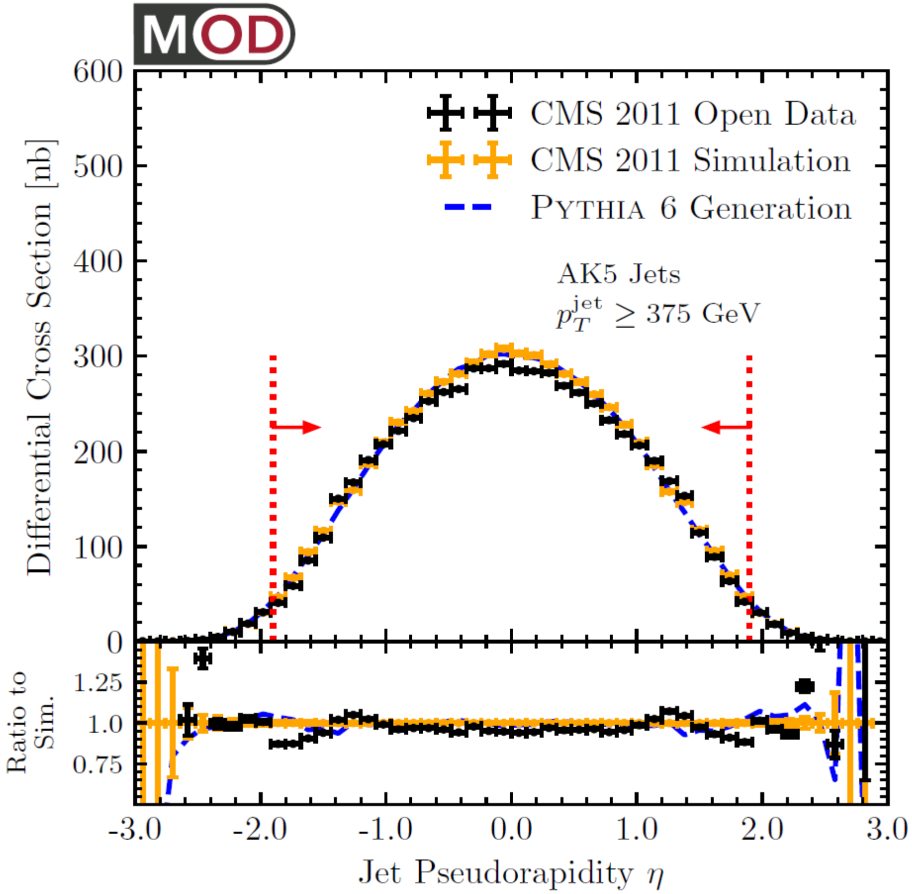
Dimension blows up at low energies.

Jets are "more than fractal"

Can we understand this analytically?

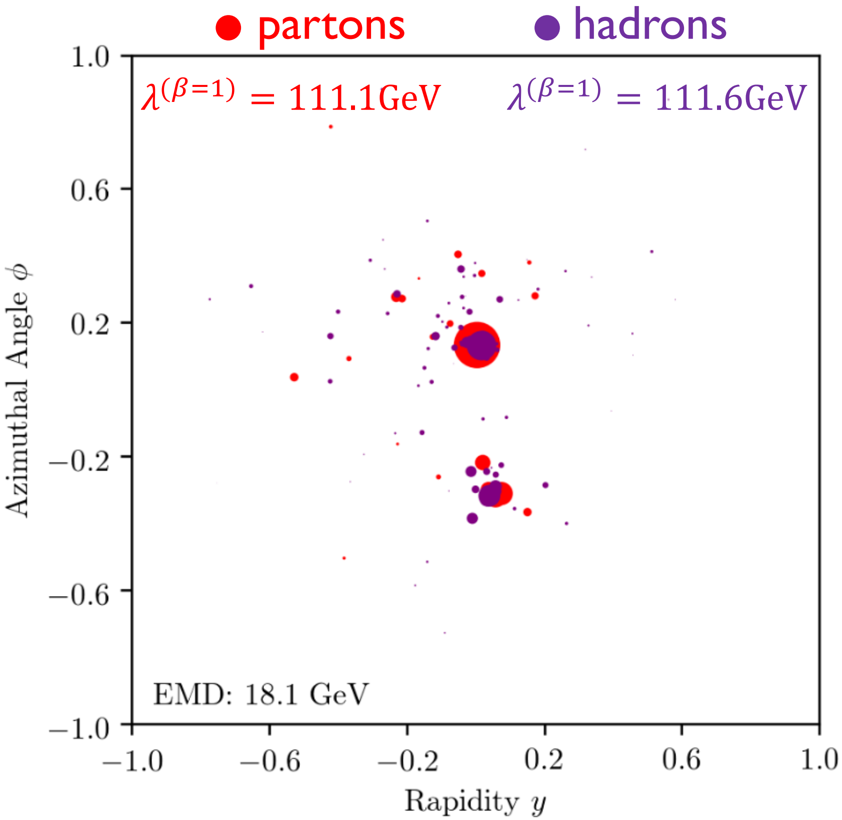
Jet Kinematic Distributions

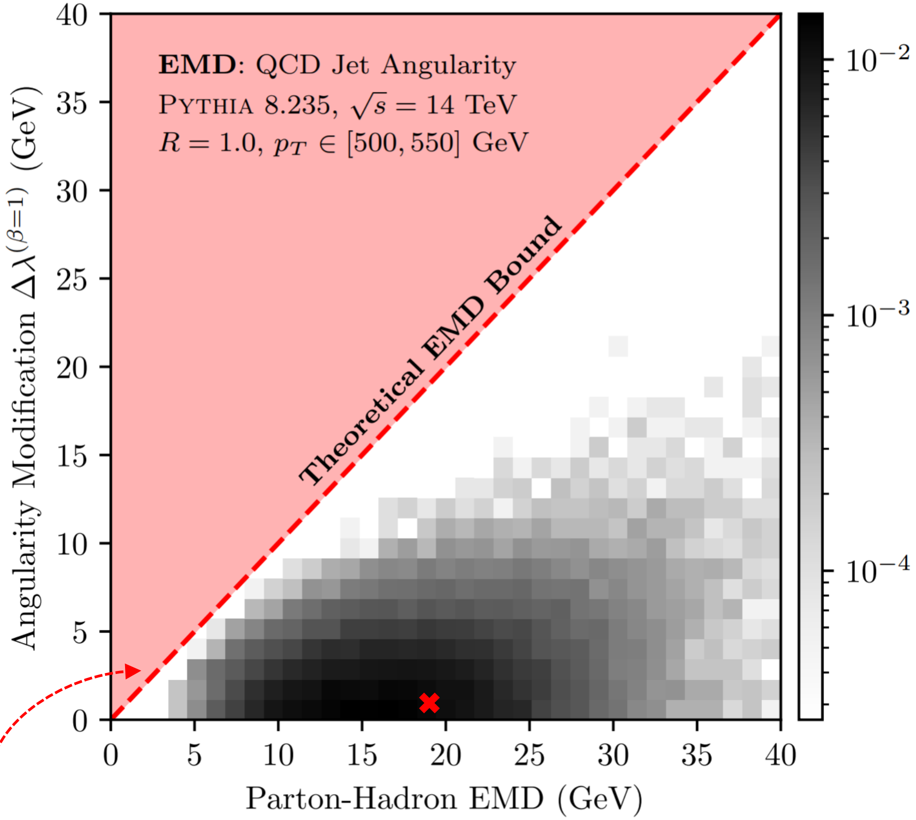




Quantifying event modifications: Hadronization

$$\lambda^{(\beta=1)} = \sum_{i=1}^{M} E_i\, \theta_i$$





$$\mathcal{E} = \mathcal{E}_{\text{partons}}$$

$$\mathcal{E}' = \mathcal{E}_{\text{hadrons}}$$

$$\boxed{\left|\lambda^{(\beta=1)}(\mathcal{E}) - \lambda^{(\beta=1)}(\mathcal{E}')\right| \leq \text{EMD}(\mathcal{E}, \mathcal{E}')}$$

Exploring the Space of Events: Jet Classification

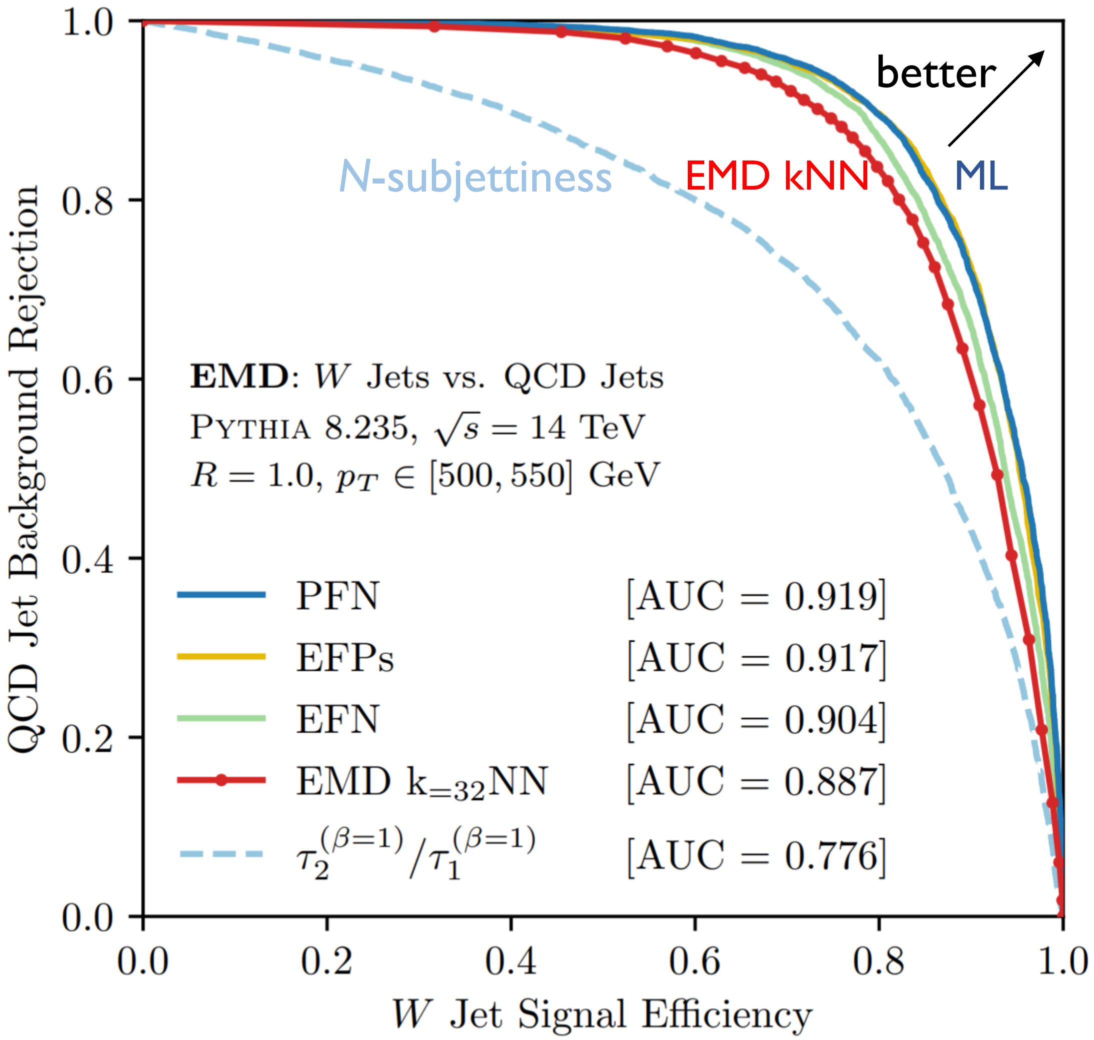
Classify *W* jets vs. QCD jets

vs.

Look at a jet's nearest neighbors (kNN) to predict its class.

Optimal IRC-safe classifier with enough data.

Nearing performance of ML.



Exploring the Space of Events

Use EMD as a measure of event similarity

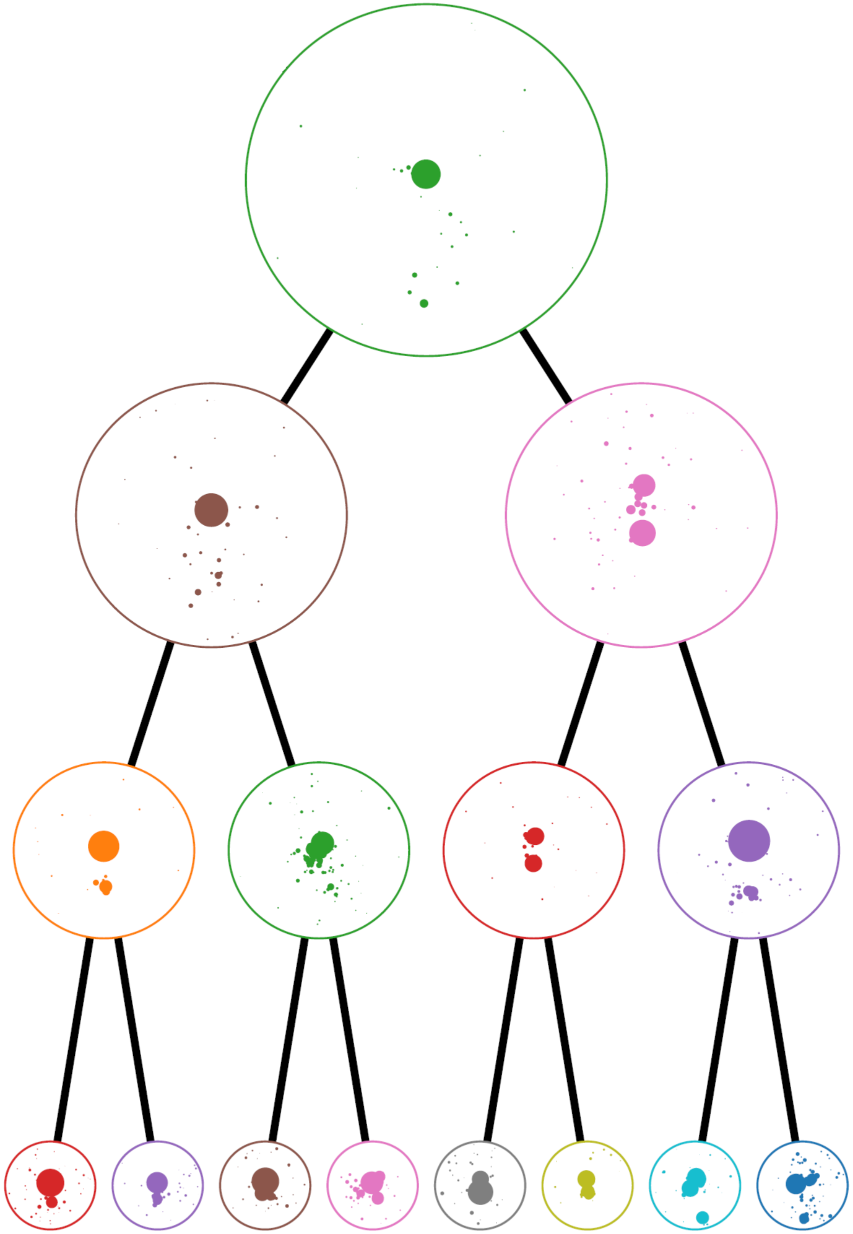
Unsupervised clustering algorithms can be used to cluster events

Jets are clusters of particles
???? are clusters of jets

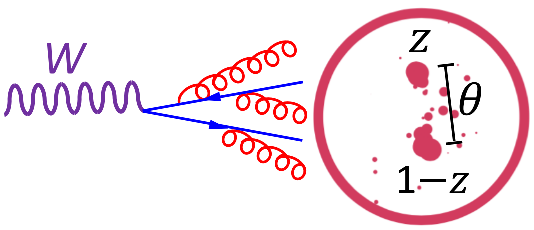
VP Tree: O(log(N)) neighbor query time

Much more to explore.

Vantage Point (VP) Tree



Exploring the Space of Events: *W* jets



W jets are 2-pronged:

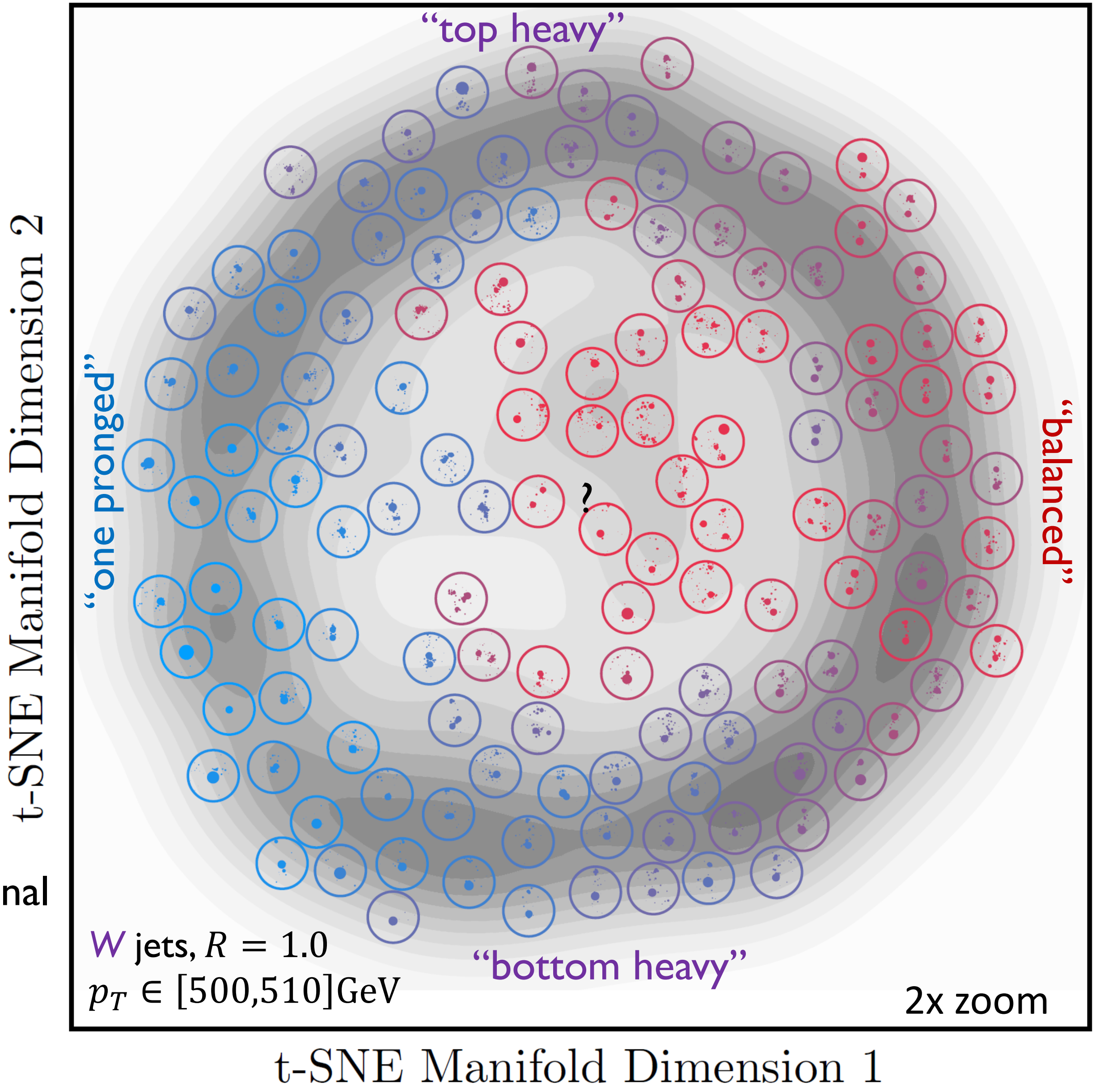
z: Energy Sharing of Prongs
θ: Angle between Prongs
φ: Azimuthal orientation

Constrained by *W* mass:

$$z(1-z)\theta^2 = \frac{p_{\mu J}^2}{p_T^2} = \frac{m_W^2}{p_T^2}$$

Hence we expect a **two**-dimensional space of *W* jets.

After φ rotation: **one**-dimensional



Exploring the Space of Jets: Correlation Dimension

Sketch of leading log (one emission) calculation:

$$\dim_{q/g}(Q) = Q\frac{\partial}{\partial Q}\ln \sum_{i=1}^{N}\sum_{j=1}^{N}\Theta[\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}_j) < Q]$$

$$= Q\frac{\partial}{\partial Q}\ln \Pr\left[\mathrm{EMD} < Q\right]$$

$$= Q\frac{\partial}{\partial Q}\ln \Pr\left[\lambda^{(\beta=1)} < Q; C_{q/g} \to 2\,C_{q/g}\right]$$

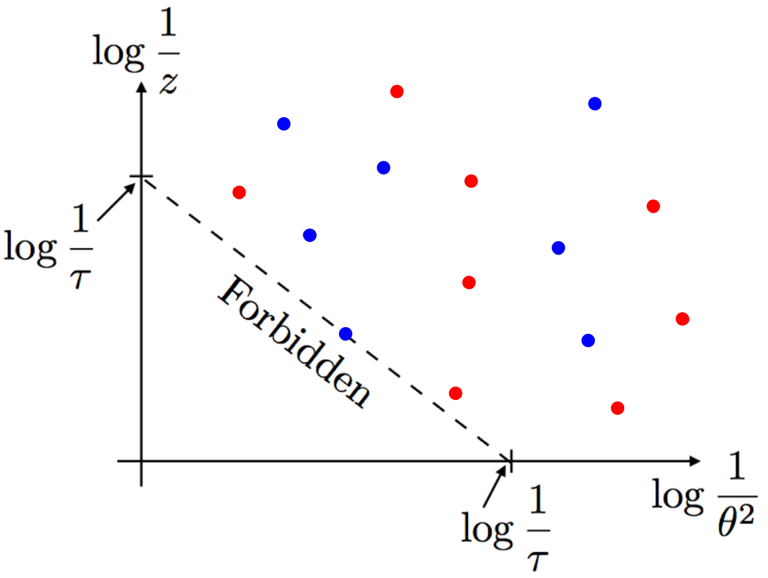
$$= Q\frac{\partial}{\partial Q}\ln \exp\left(-\frac{4\alpha_s C_{q/g}}{\pi}\ln^2\frac{Q}{p_T/2}\right)$$

$$= -\frac{8\alpha_s C_{q/g}}{\pi}\ln\frac{Q}{p_T/2}$$

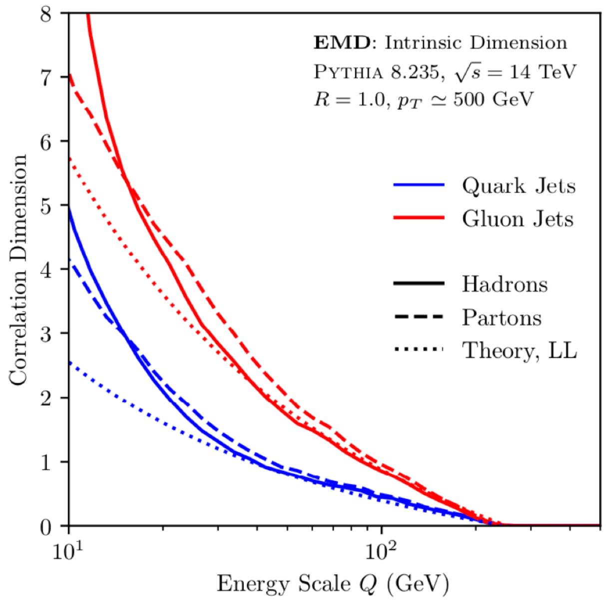
+ 1-loop running of α_s

$$C_q = C_F = \frac{4}{3}$$

$$C_g = C_A = 3$$

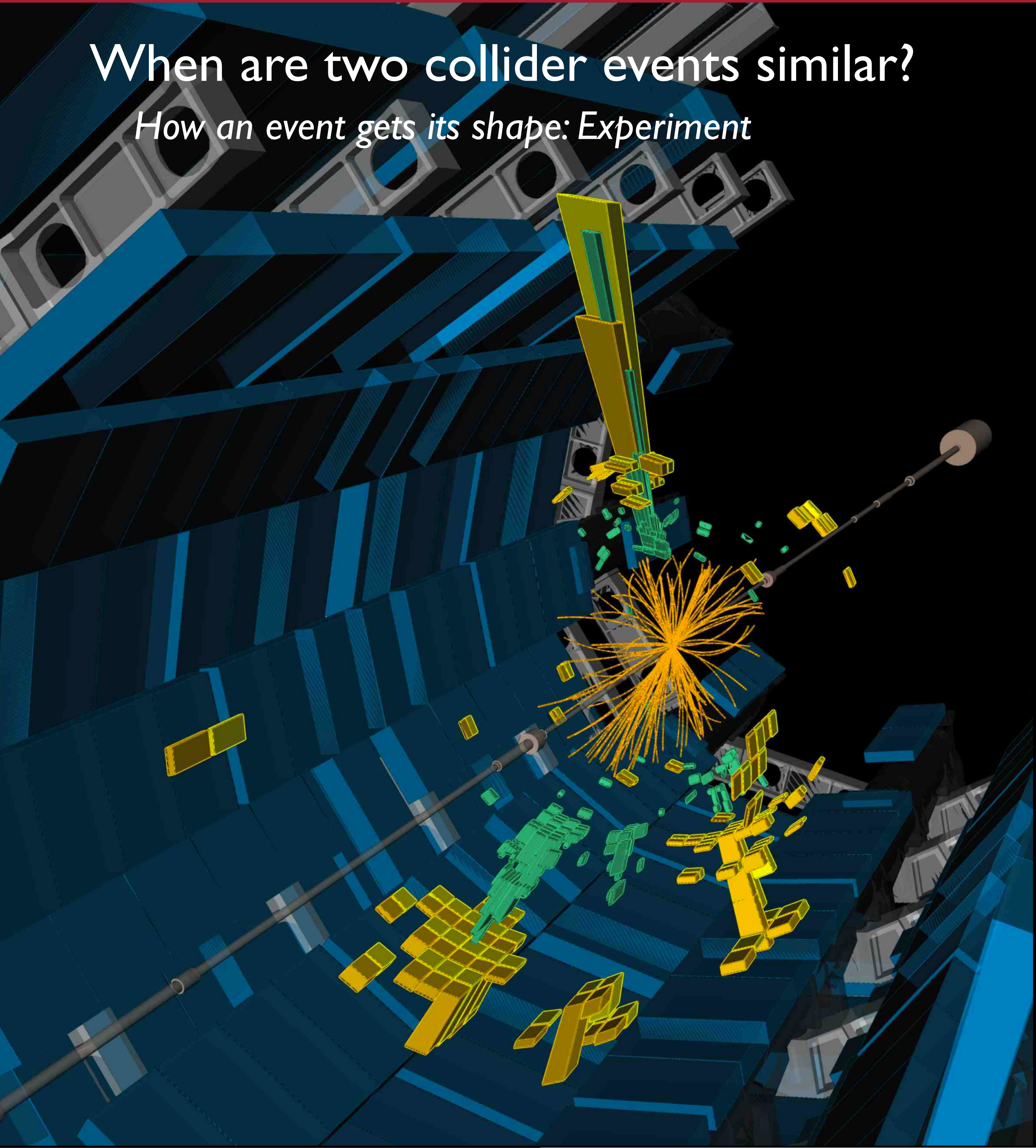


[A. Larkoski, 1709.06195]



When are two collider events similar?

How an event gets its shape: Experiment



tracker	ECAL	HCAL		
	🟩		γ	photon
⌣	🟩		$e^{\pm}$	electron
⌣	▪	▪	$\mu^{\pm}$	muon
⌣	▪	🟨	$\pi^{\pm}$	pion
⌣	▪	🟨	$K^{\pm}$	kaon
	▪	🟨	K_L^0	K-long
⌣	▪	🟨	$p/\bar{p}$	proton
	▪	🟨	$n/\bar{n}$	neutron

Pileup Mitigation with PUMML