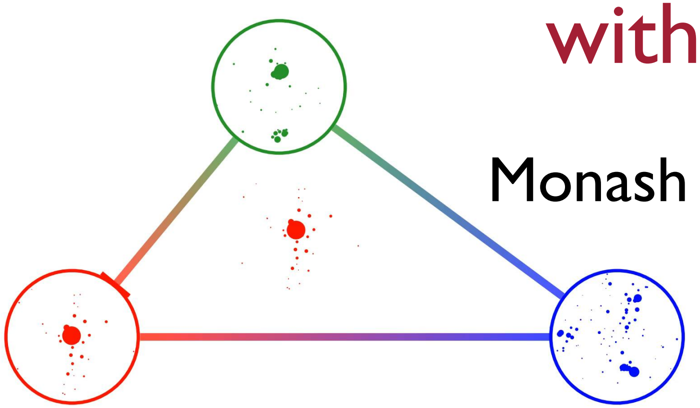
Exploring the (Metric) Space of Collider Events with CMS Open Data



Monash University Virtual Seminar

Eric M. Metodiev

Center for Theoretical Physics

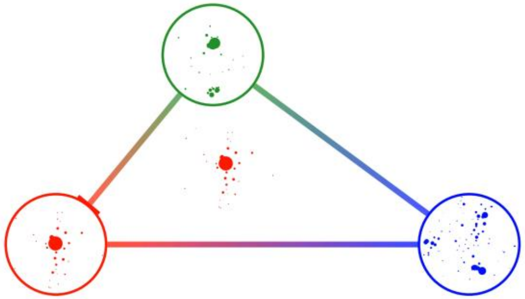
Massachusetts Institute of Technology

Joint work with Patrick Komiske and Jesse Thaler [1902.02346]

CMS Open Data also with Radha Mastandrea and Preksha Naik [1908.08542]

November 19, 2019

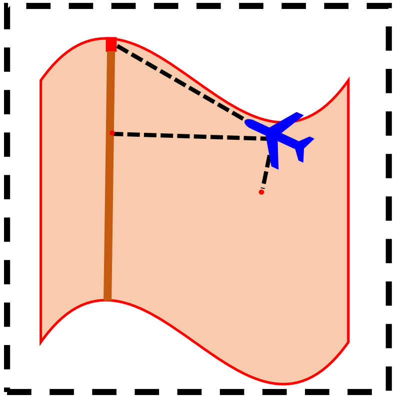
Outline



The Metric Space of Collider Events

When are two events similar?

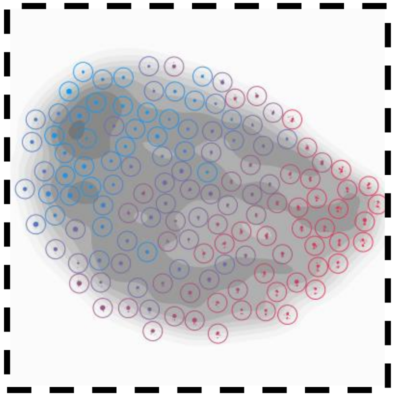
The Energy Mover's Distance



A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization

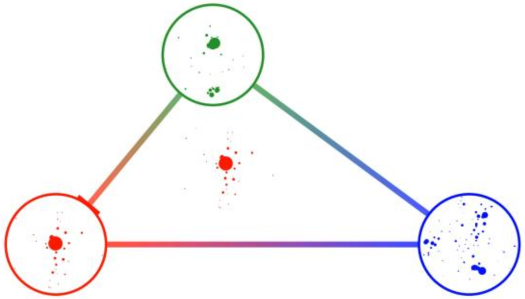


Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

Outline



The Metric Space of Collider Events

When are two events similar?

The Energy Mover's Distance

A Geometric Language for Observables

Old Observables in a New Light

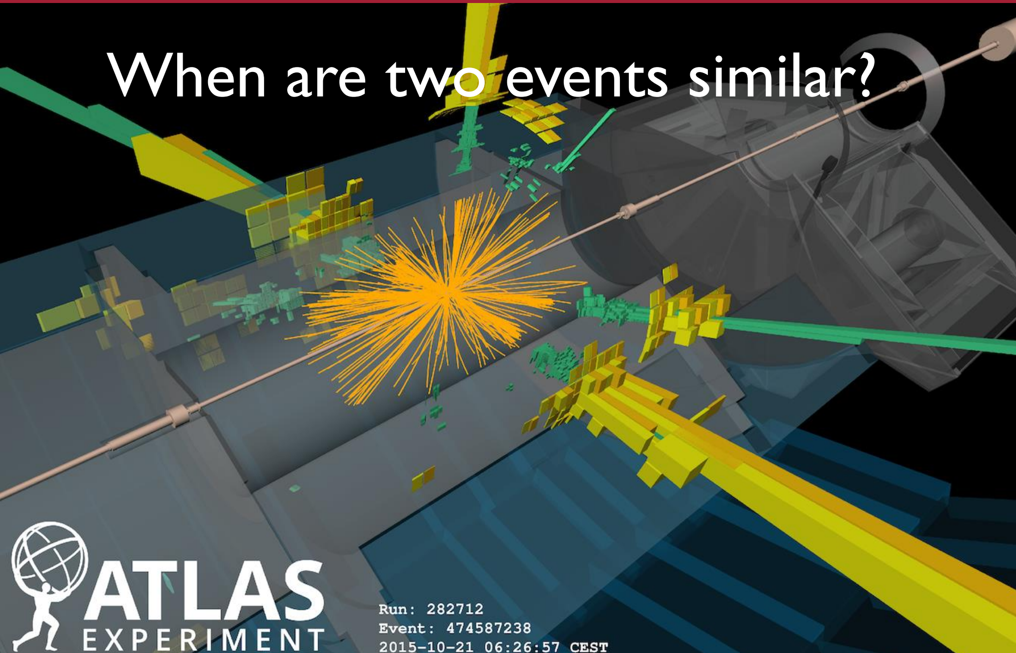
Quantifying Hadronization

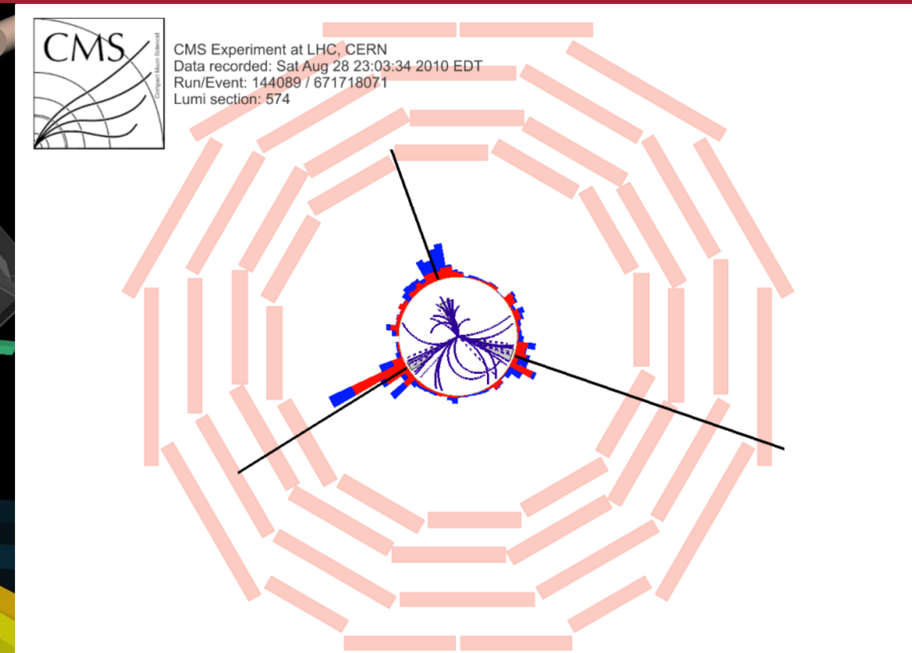
Exploring the Space of Jets with CMS Open Data

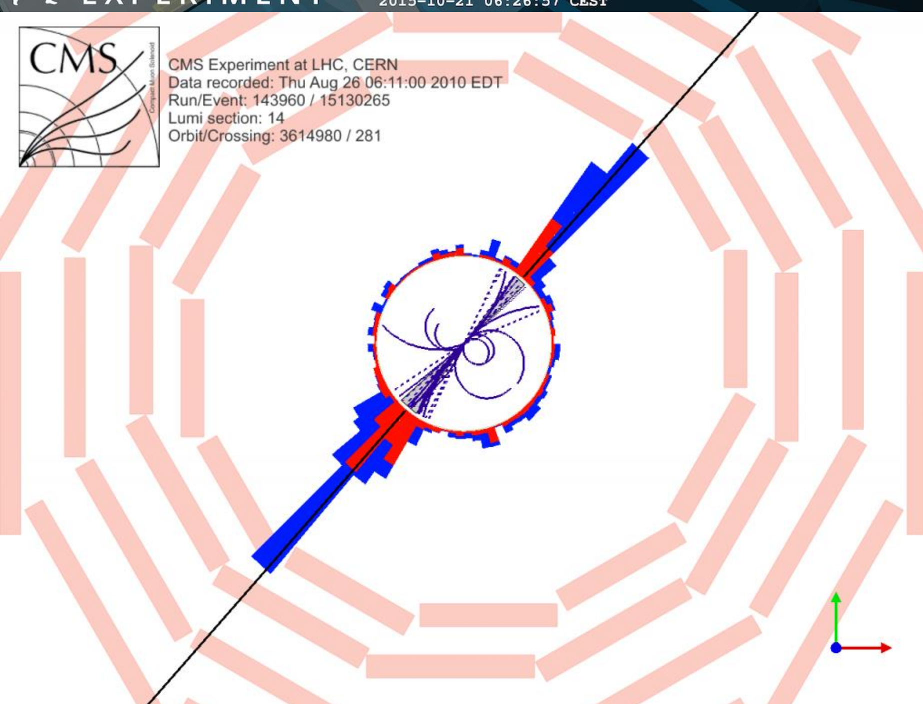
Most Representative and Anomalous Jets

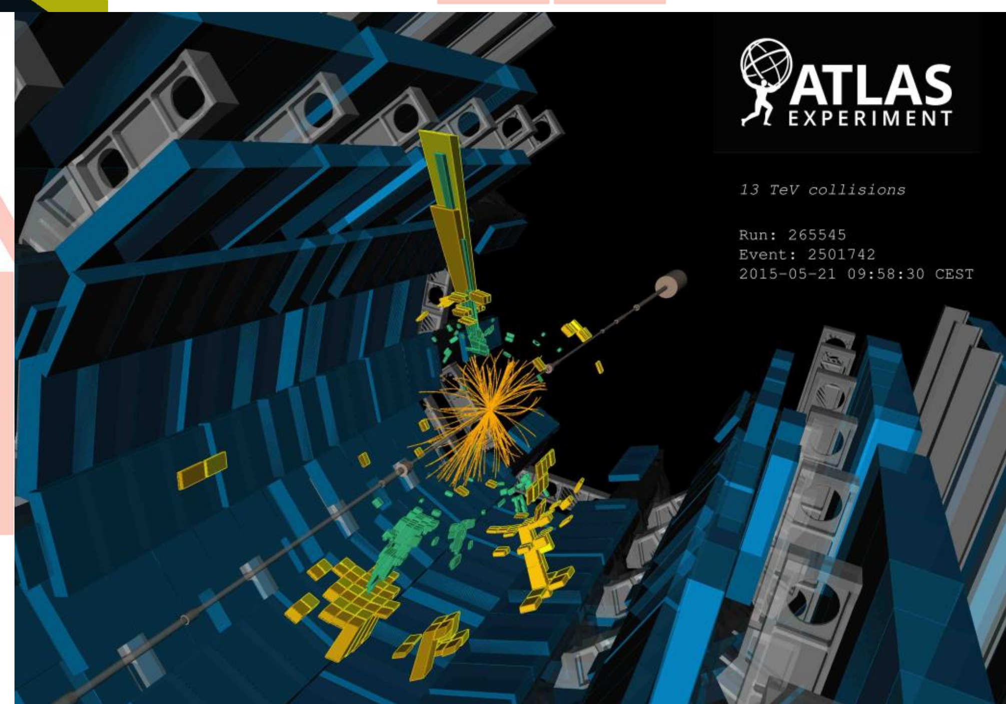
Visualizing the Space and its (fractal) dimension

When are two events similar?





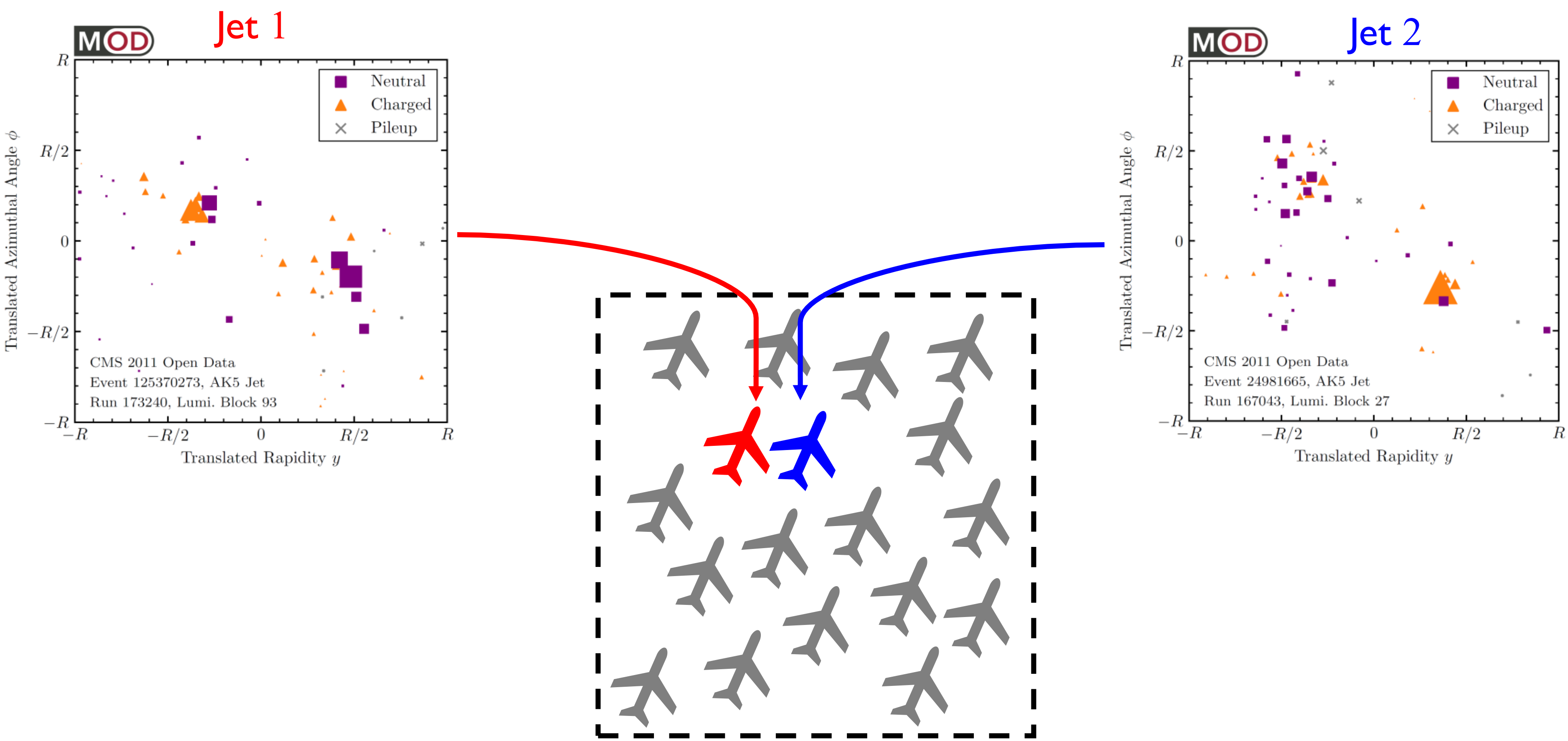




When are two events similar?

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

How do we quantify this?

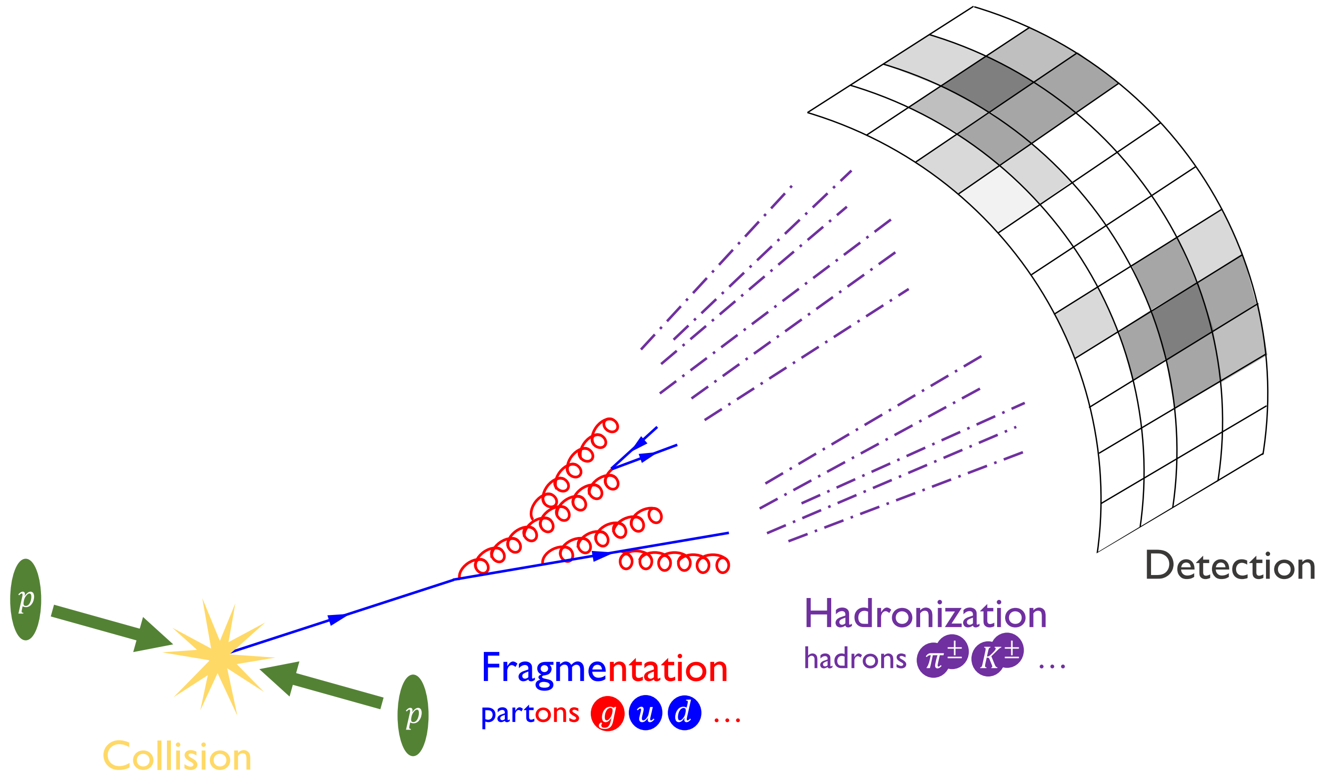


"Space of Jets"

400 GeV $R = 0.5$ anti- k_T Jets from CMS Open Data

When are two events similar?

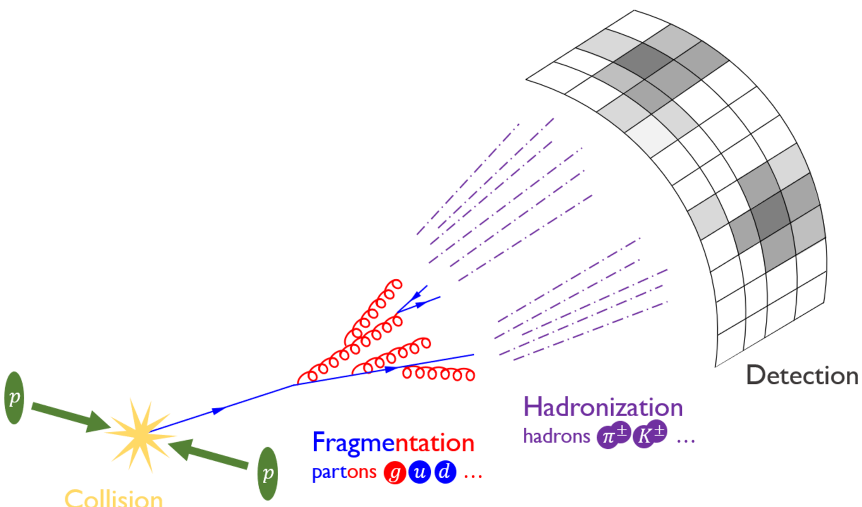
How a jet gets its shape



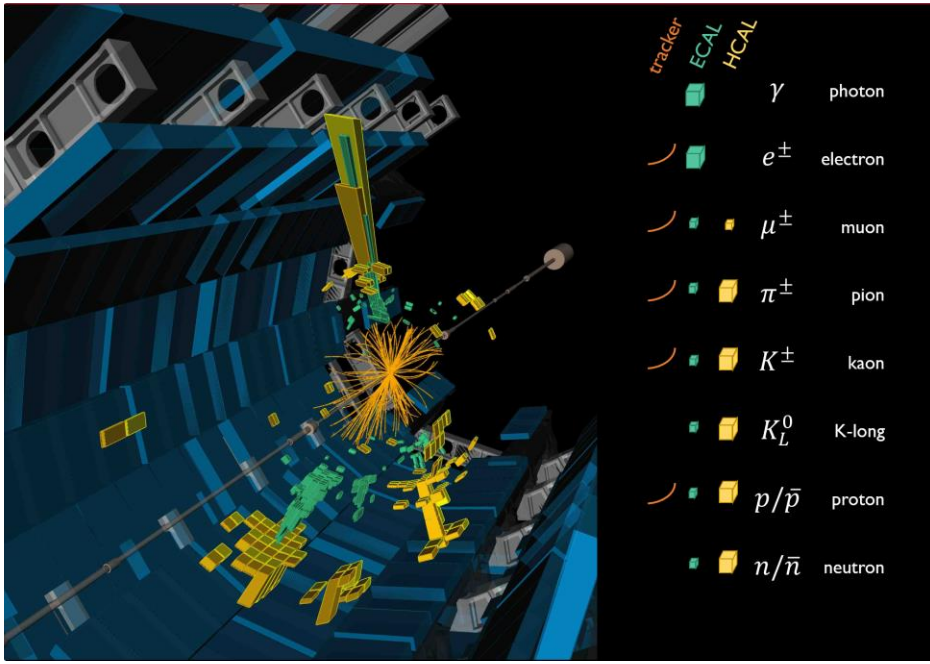
When are two events similar?

An event is…

Theoretically: very complicated



Experimentally: very complicated



However:

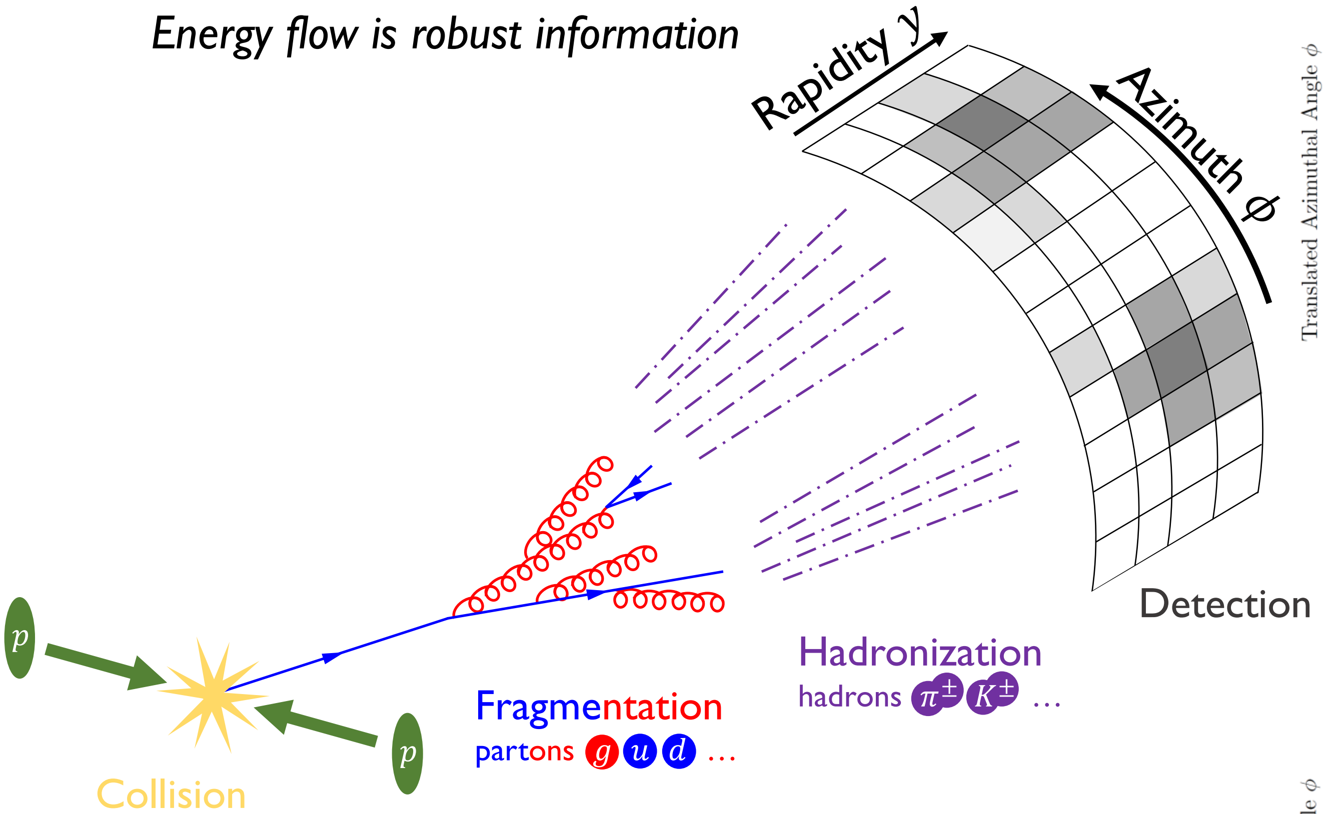
The *energy flow* (distribution of energy) is the information that is robust to:
fragmentation, hadronization, detector effects, …

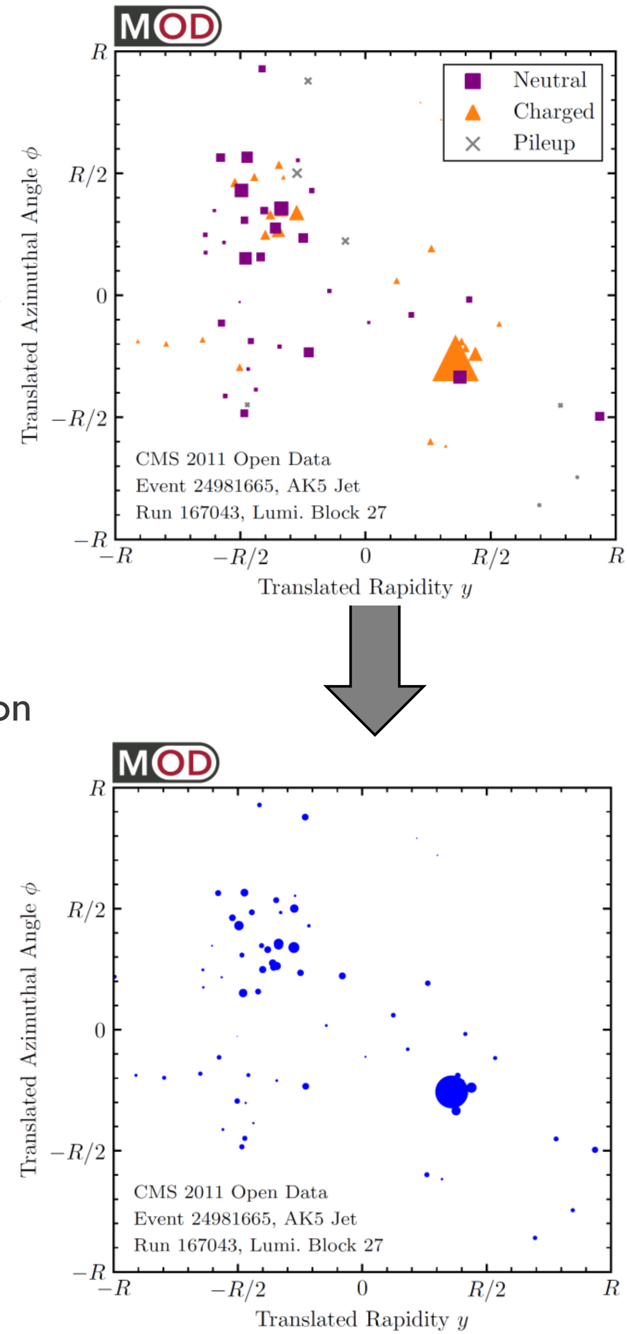
[N.A. Sveshnikov, F.V. Tkachov, 9512370]
[F.V. Tkachov, 9601308]
[P.S. Cherzor, N.A. Sveshnikov, 9710349]

Energy flow ⇔ Infrared and Collinear (IRC) Safe information

When are two jets similar?

Energy flow is robust information



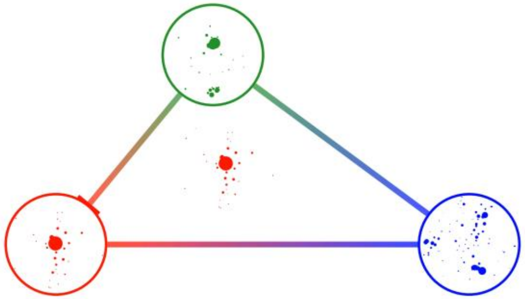


Treat events as distributions of energy: $\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} E_i \, \delta(\hat{n} - \hat{n}_i)$

energy direction

Ignoring particle flavor, charge…

Outline



The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

A Geometric Language for Observables

Old Observables in a New Light

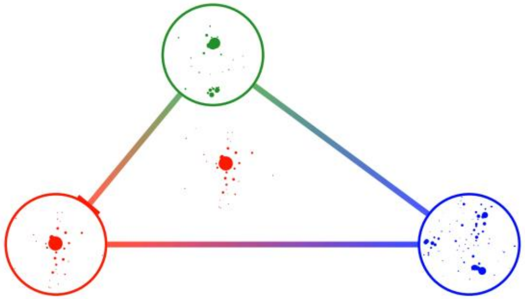
Quantifying Hadronization

Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

Outline



The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization

Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

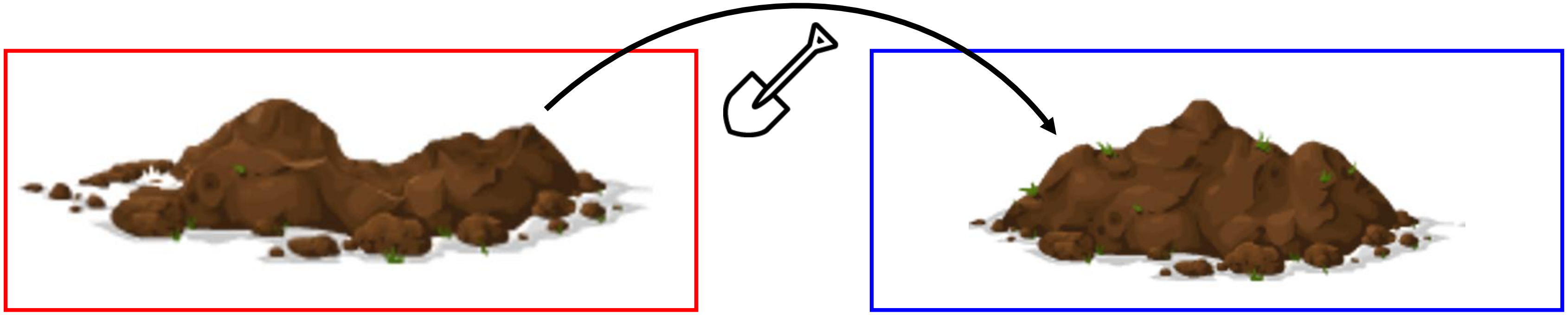
Visualizing the Space and its (fractal) dimension

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*

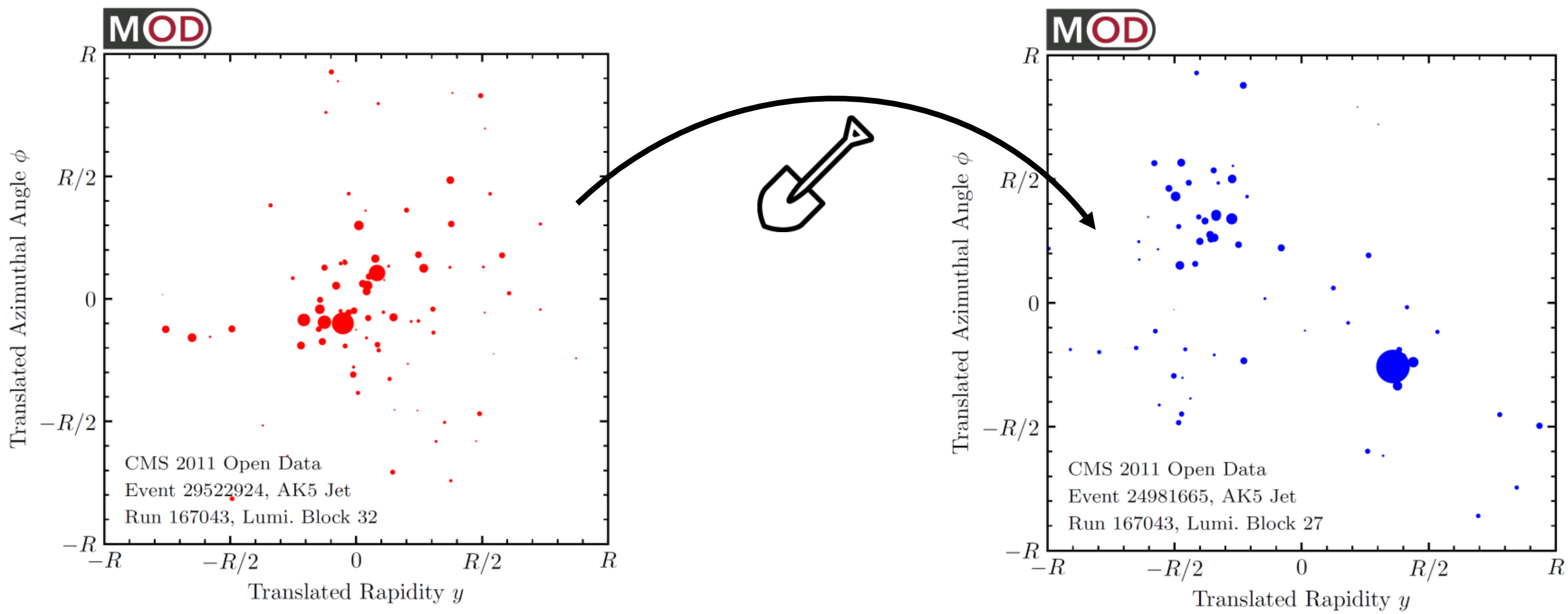
Distributions are close in EMD ⇔ their expectation values are close.

Also known as the 1-Wasserstein metric.

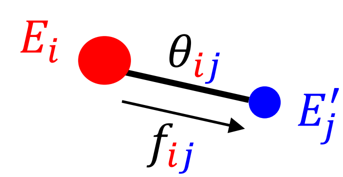
The Energy Mover's Distance

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]



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



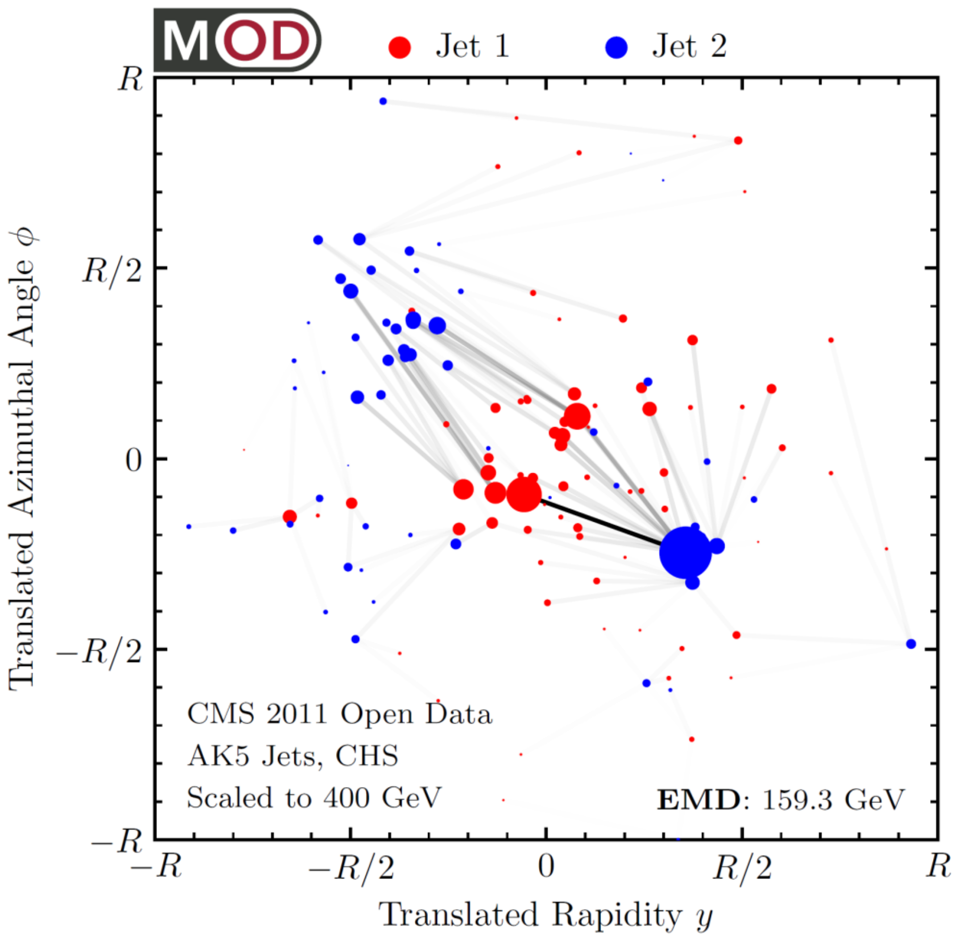
Difference in radiation pattern

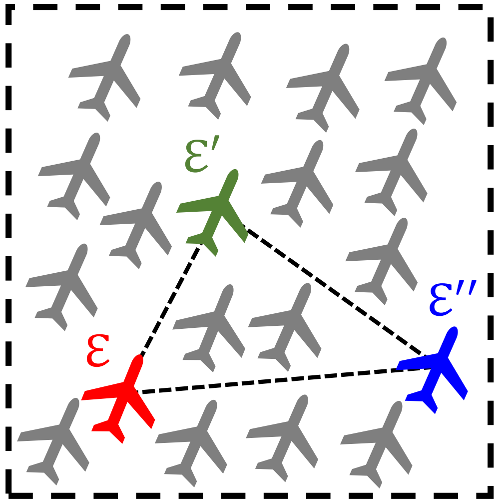
Difference in total energy

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]





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

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

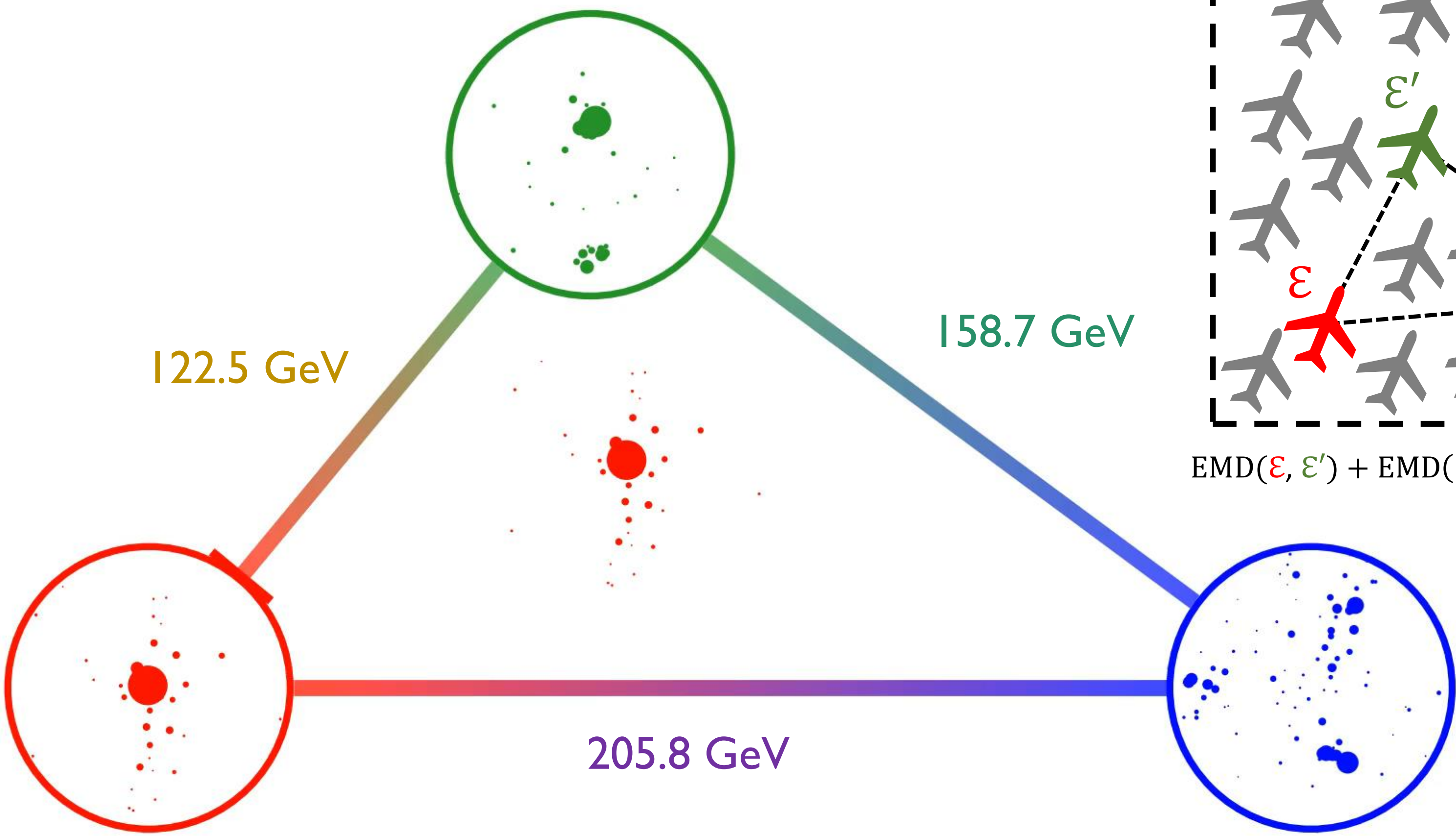
Solvable via Optimal Transport problem.

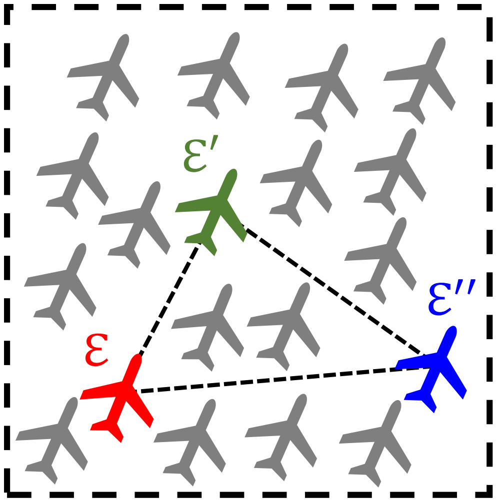
~1ms to compute EMD for two jets with 100 particles.

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]

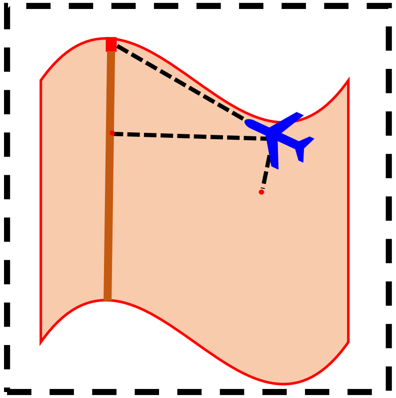




$$\mathrm{EMD}(\mathcal{E}, \mathcal{E}') + \mathrm{EMD}(\mathcal{E}', \mathcal{E}'') \geq \mathrm{EMD}(\mathcal{E}, \mathcal{E}'')$$

https://energyflow.network

Outline



The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

The "work" to rearrange one event into another

A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization

Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

Outline

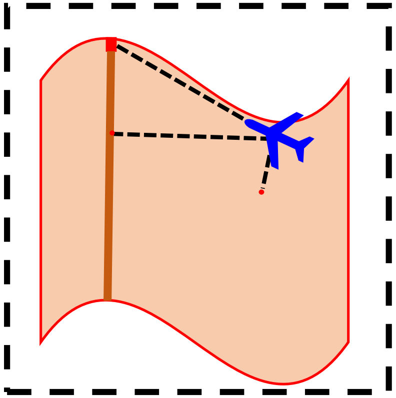
The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

The "work" to rearrange one event into another



A Geometric Language for Observables

Old Observables in a New Light

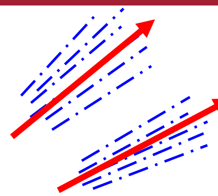
Quantifying Hadronization

Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

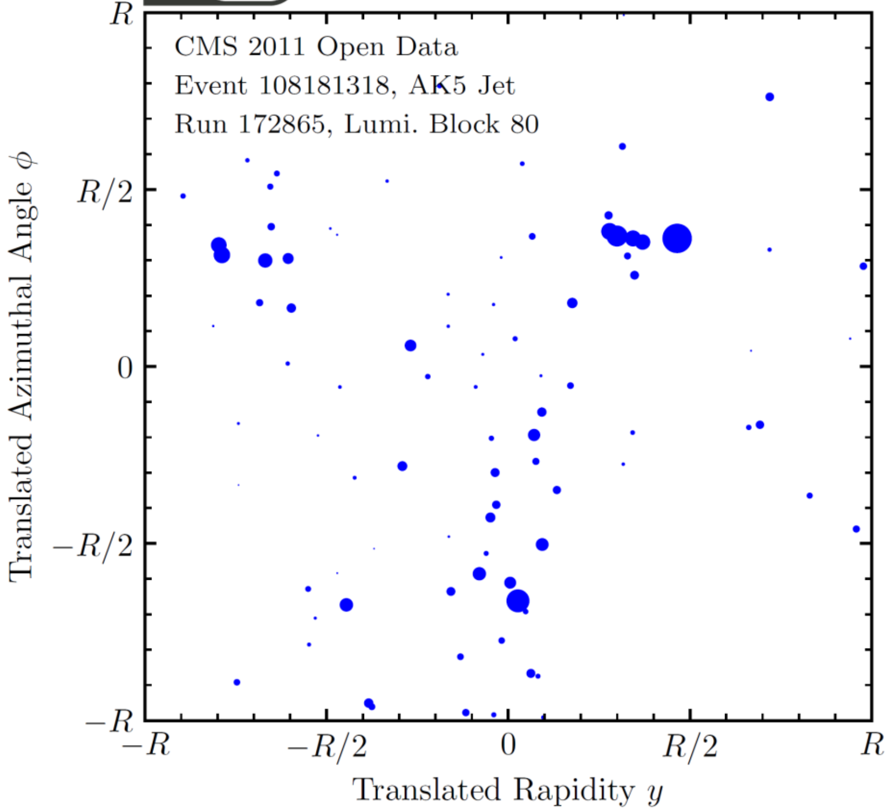
A Geometric Language for Observables



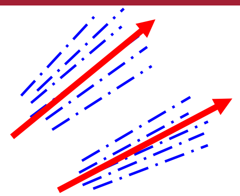
***N*-(sub)jettiness** is a ubiquitous "N-prong" observable used at the LHC

[Thaler, Van Tilburg, 1011.2268]

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



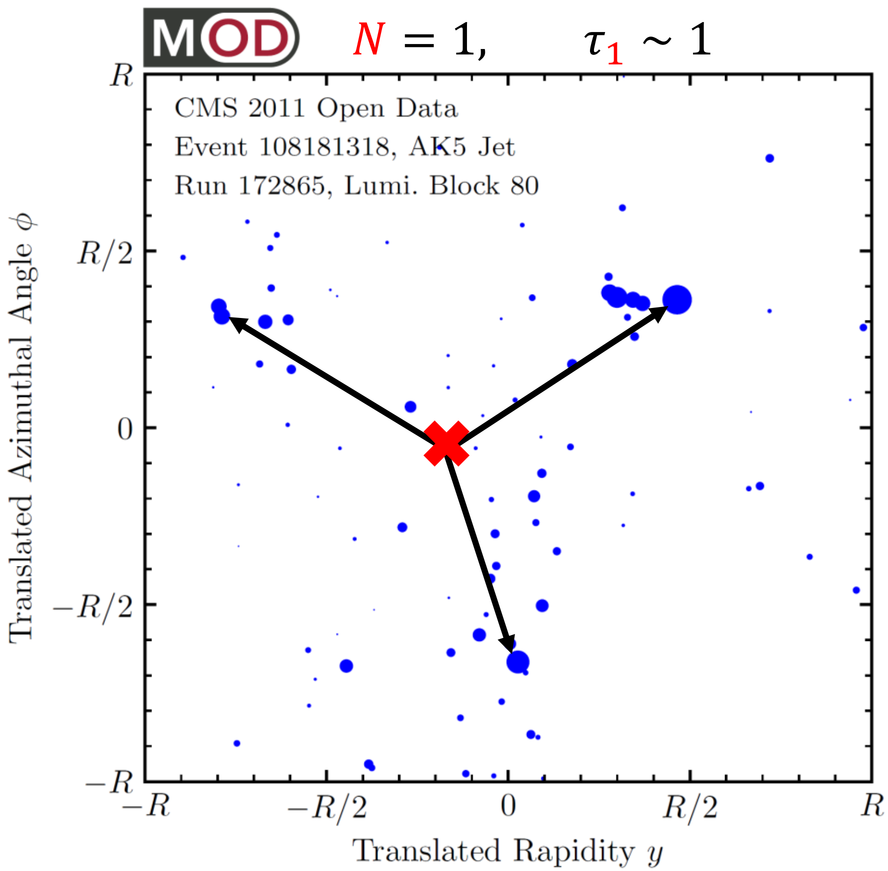
A Geometric Language for Observables



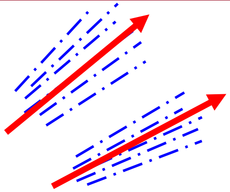
***N*-(sub)jettiness** is a ubiquitous "N-prong" observable used at the LHC

[Thaler, Van Tilburg, 1011.2268]

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



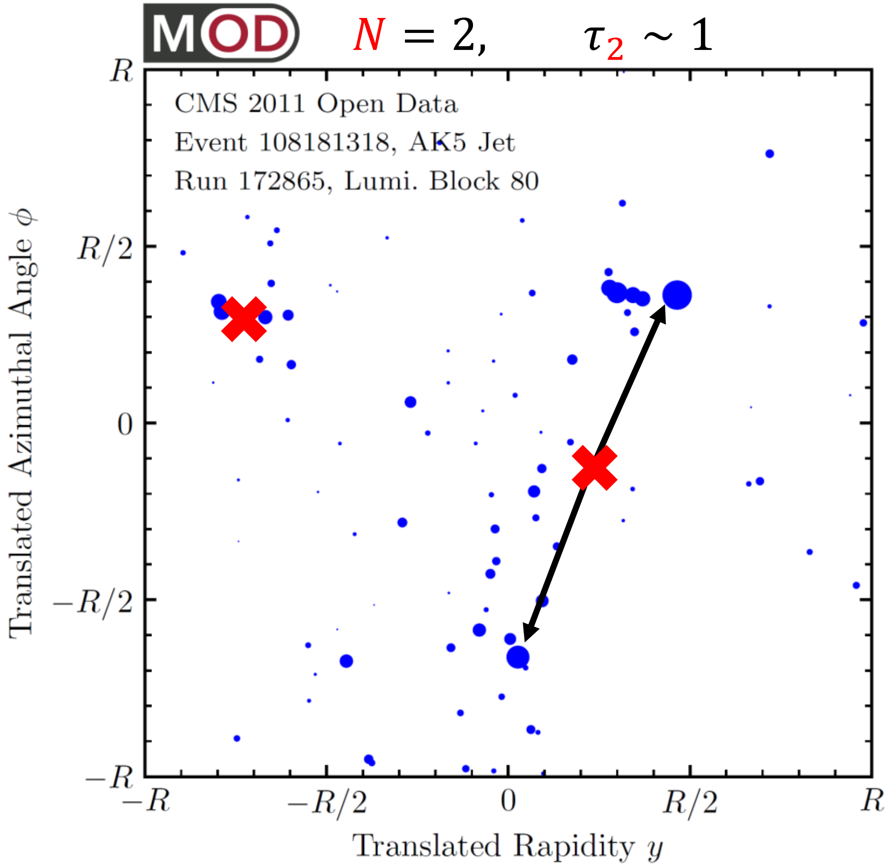
A Geometric Language for Observables



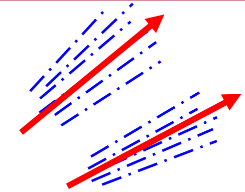
***N*-(sub)jettiness** is a ubiquitous "N-prong" observable used at the LHC

[Thaler, Van Tilburg, 1011.2268]

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



A Geometric Language for Observables

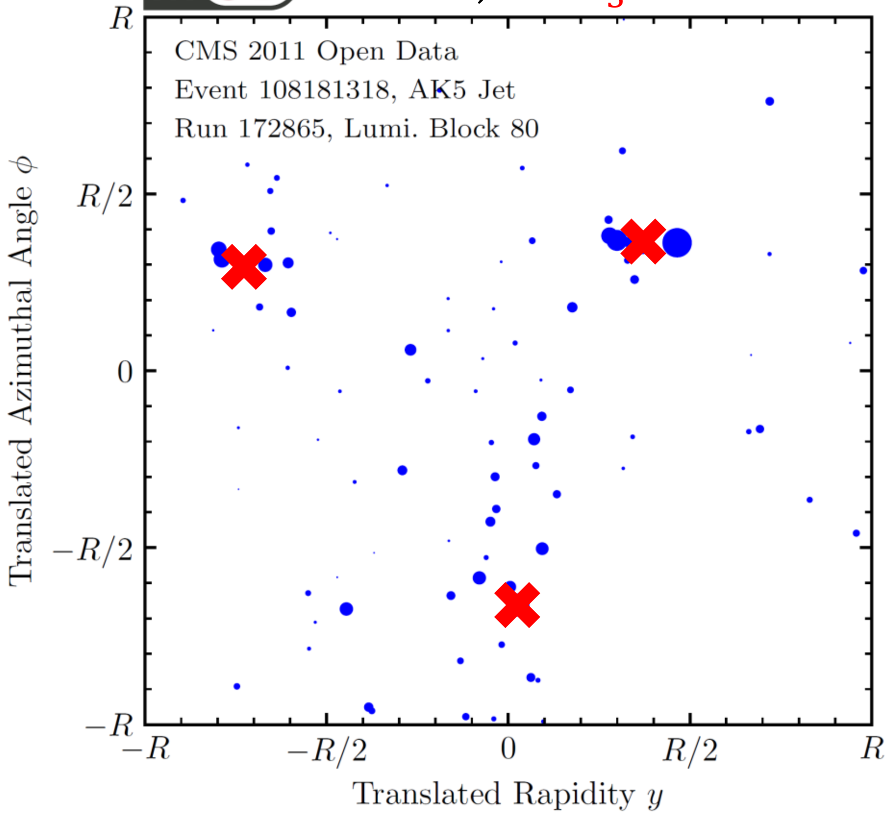


***N*-(sub)jettiness** is a ubiquitous "N-prong" observable used at the LHC

[Thaler, Van Tilburg, 1011.2268]

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

$N = 3, \qquad \tau_3 \ll 1$



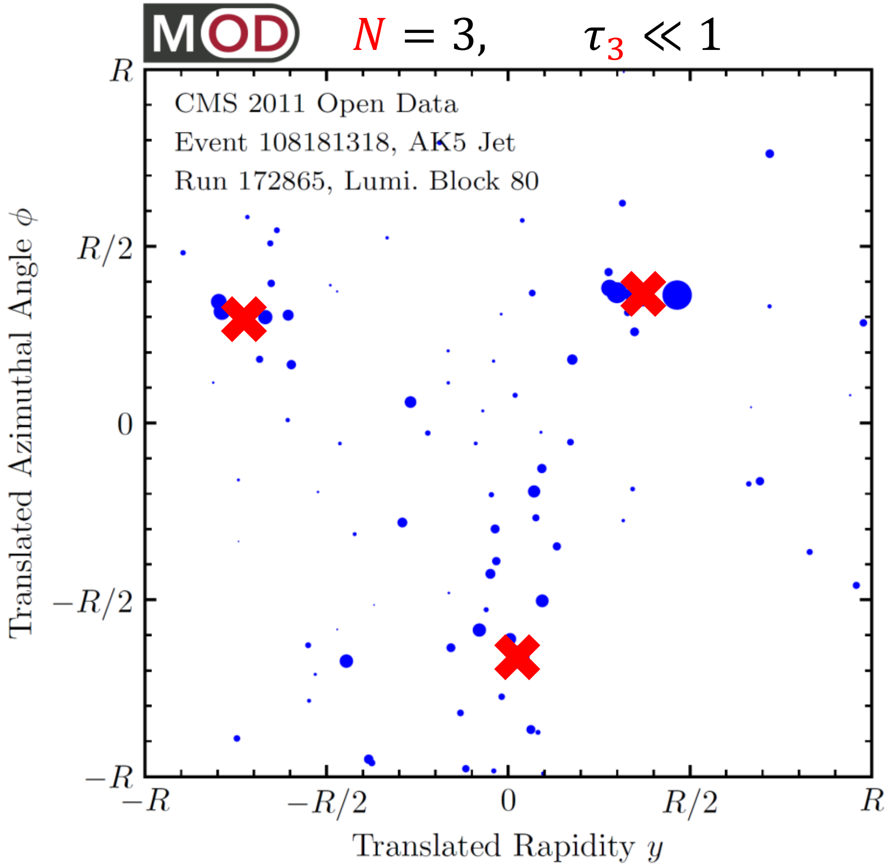
A Geometric Language for Observables

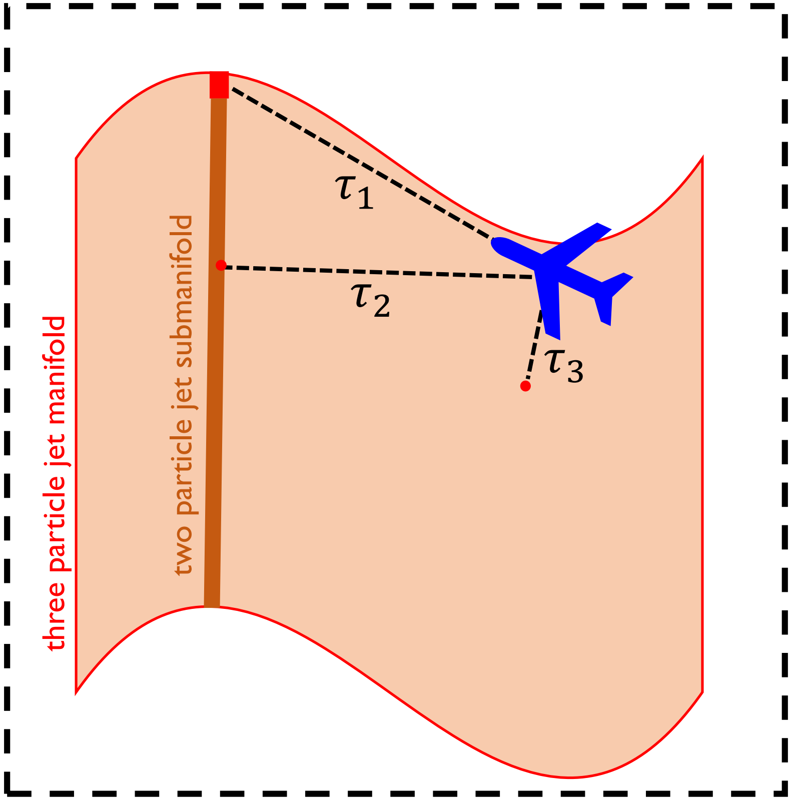
N-(sub)jettiness is the EMD between the event and the closest N-particle event.

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

β-Wasserstein distance

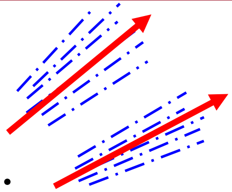
$N = 3, \quad \tau_3 \ll 1$





Geometry in the space of events

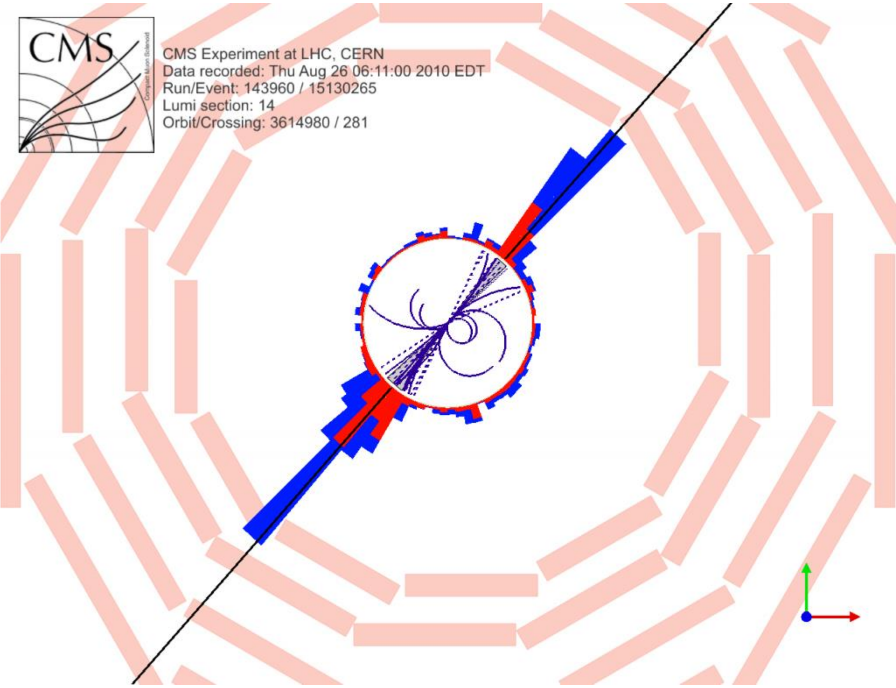
A Geometric Language for Observables



Thrust is a classic event shape that measures how "pencil-like" an event is.

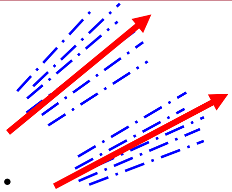
[Farhi, PRL 1977]

$$t(\mathcal{E}) = E - \max_{\hat{n}} \sum_i |\vec{p}_i \cdot \hat{n}|$$



A Geometric Language for Observables

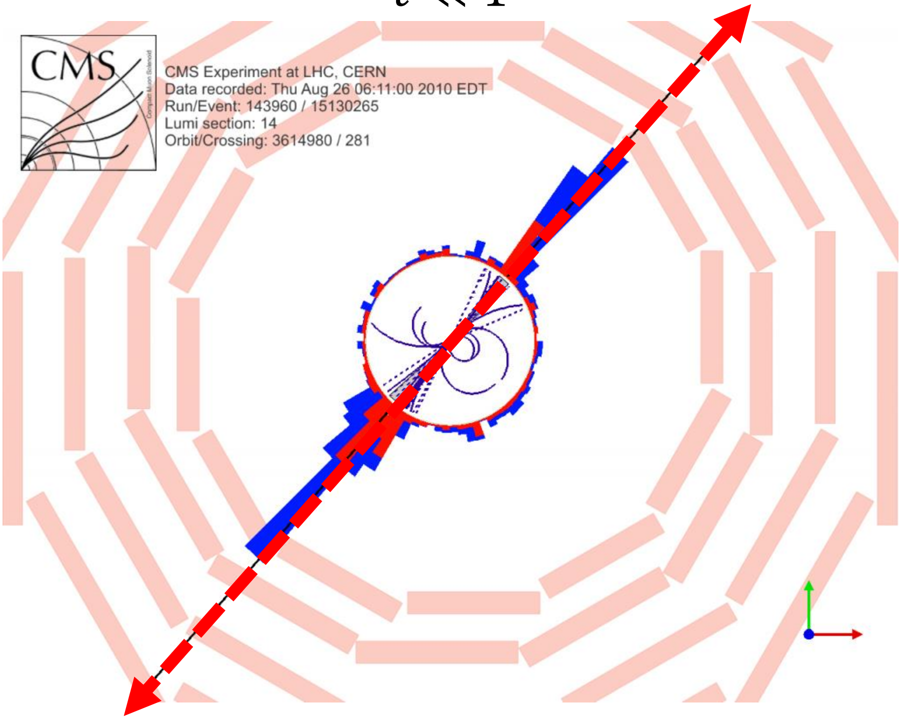
Thrust is a classic event shape that measures how "pencil-like" an event is.



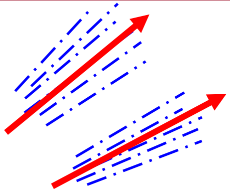
[Farhi, PRL 1977]

$$t(\mathcal{E}) = E - \max_{\hat{n}} \sum_i |\vec{p}_i \cdot \hat{n}|$$

$t \ll 1$



A Geometric Language for Observables

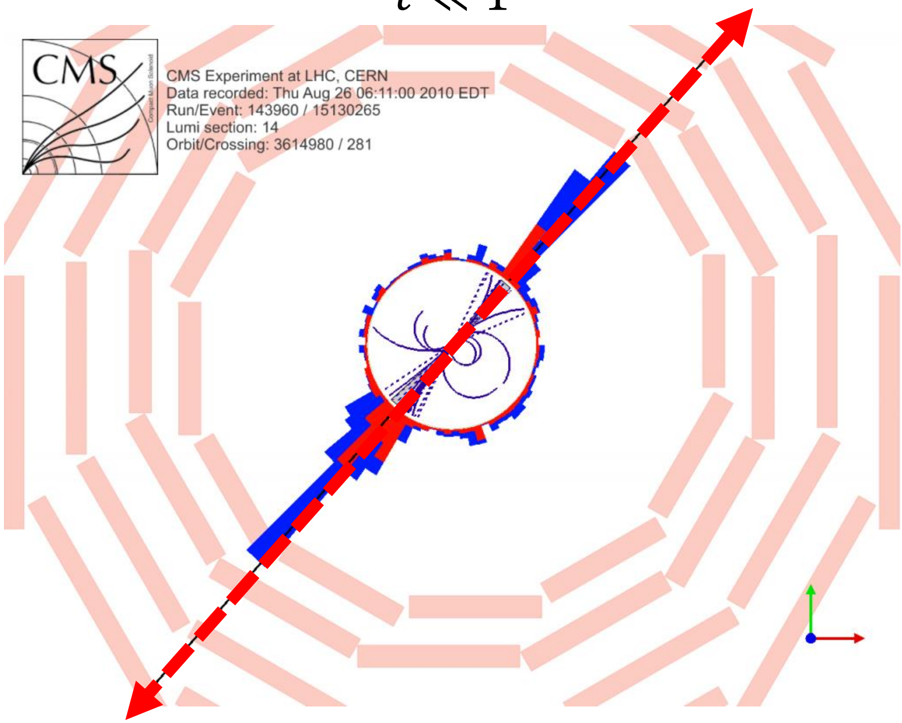


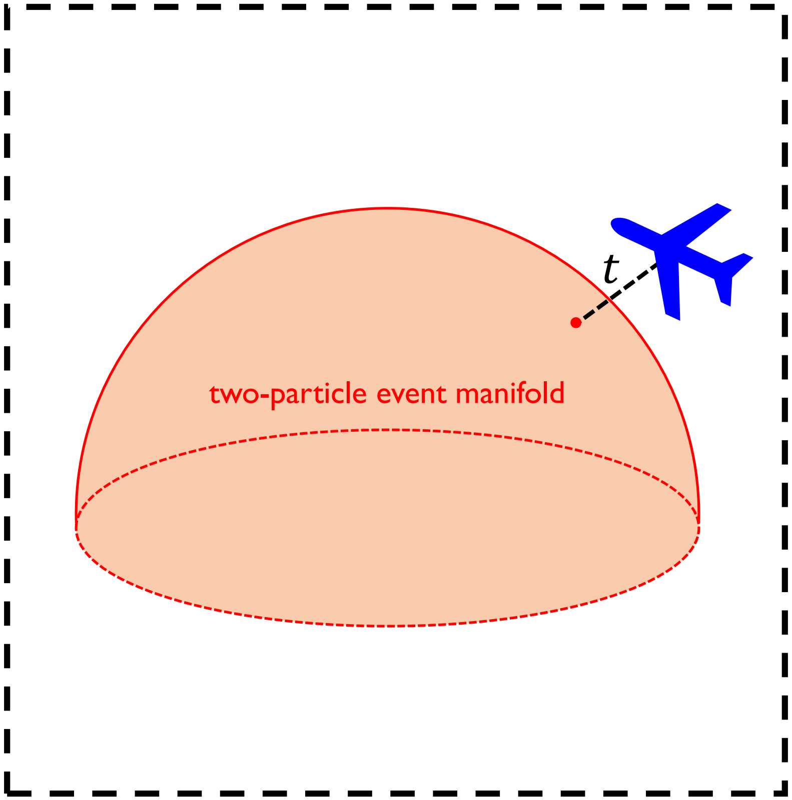
Thrust is the EMD between the event and the closest two-particle event.

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

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

$t \ll 1$





Geometry in the space of events

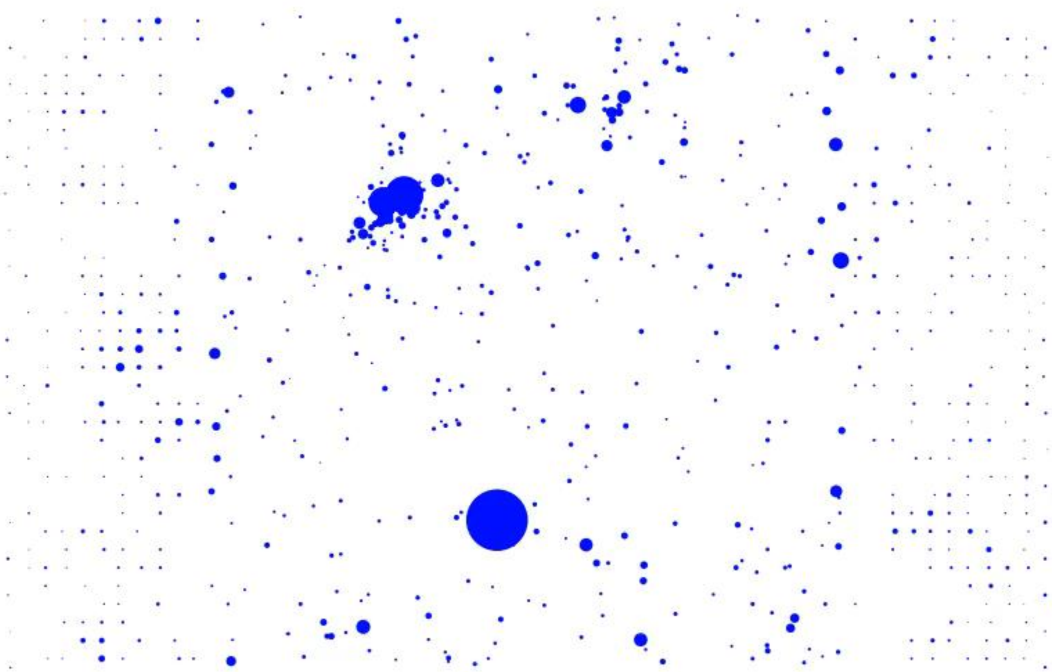
A Geometric Language for Observables

Isotropy is a new observable to probe how "uniform" an event is.

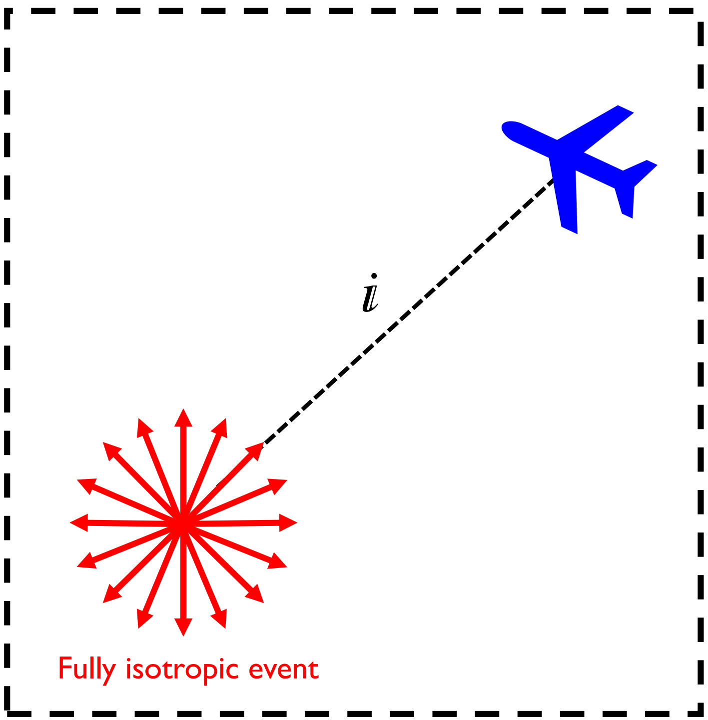
It is sensitive to very different new physics signals than existing event shapes.

e.g. uniform radiation from micro black holes [Cari Cesarotti and Jesse Thaler, coming soon!]

$$\imath(\mathcal{E}) = \mathrm{EMD}(\mathcal{E}, \mathcal{E}_{\mathrm{iso}})$$ where $\mathcal{E}_{\mathrm{iso}}$ is a fully isotropic event



dijet event from CMS Open Data



Outline

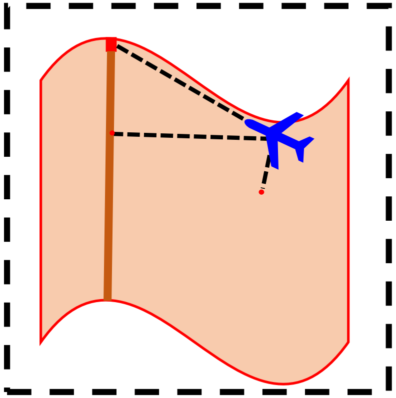
The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

The "work" to rearrange one event into another



A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization

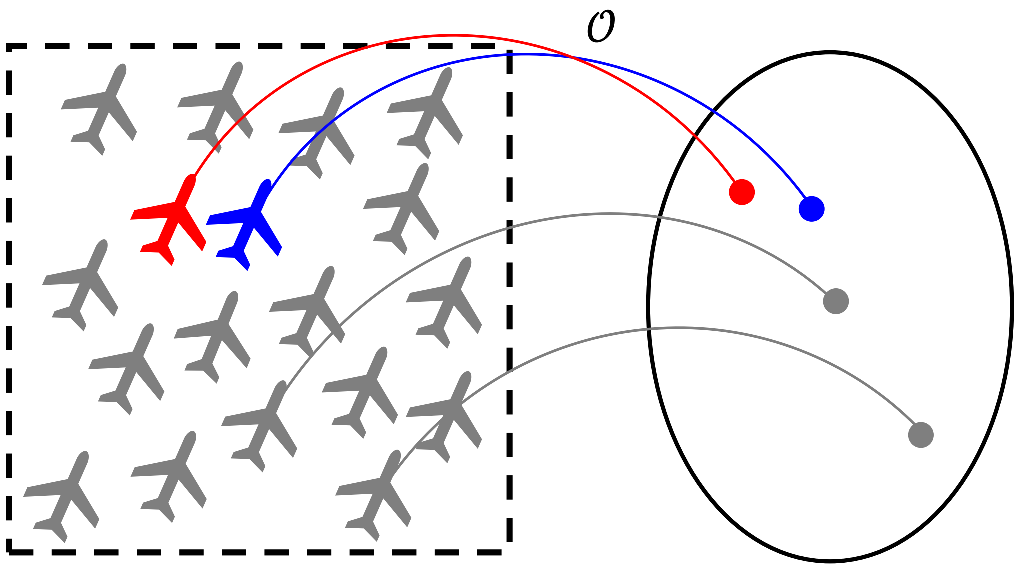
Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

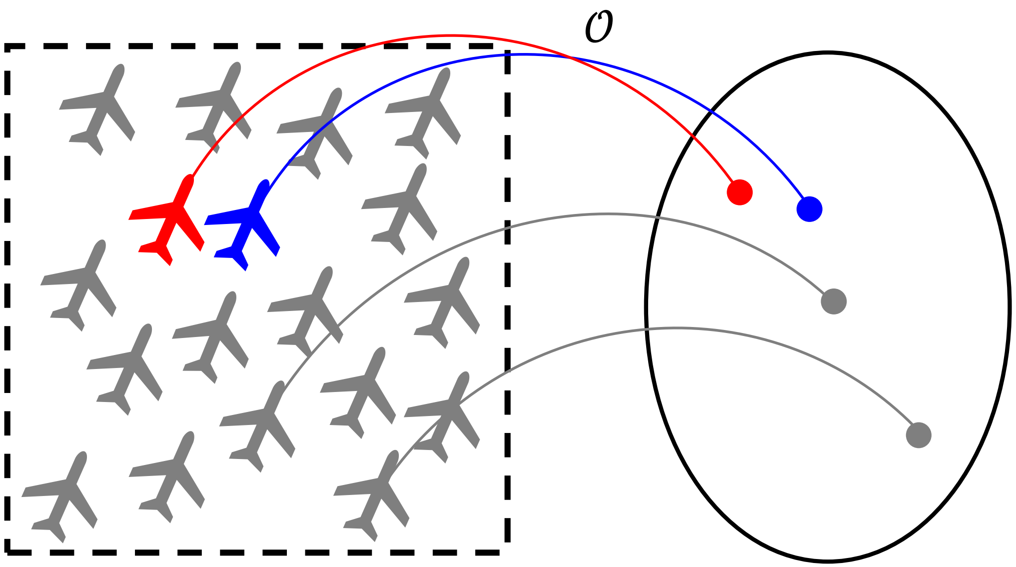
A Geometric Language for Observables

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



A Geometric Language for Observables

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



Additive IRC-safe observables: $\mathcal{O}(\mathcal{E}) = \sum_{i=1}^{M} E_i \, \Phi(\hat{n}_i)$

Energy Mover's Distance

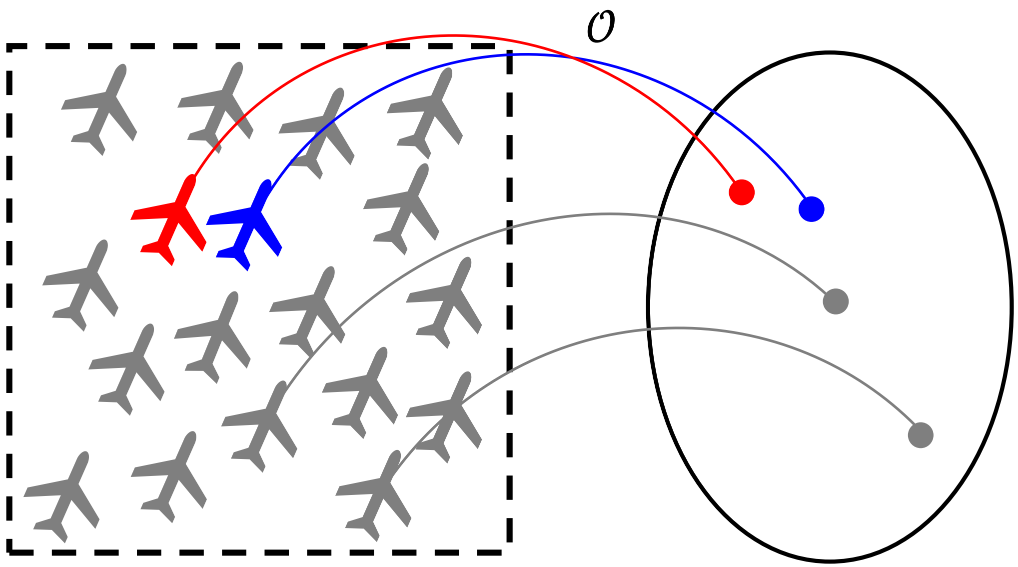
$$\boxed{\mathrm{EMD}(\mathcal{E}, \mathcal{E}') \geq \frac{1}{RL} \left|\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')\right|}$$

Difference in observable values

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

A Geometric Language for Observables

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



Jet angularities with $\beta \geq 1$:

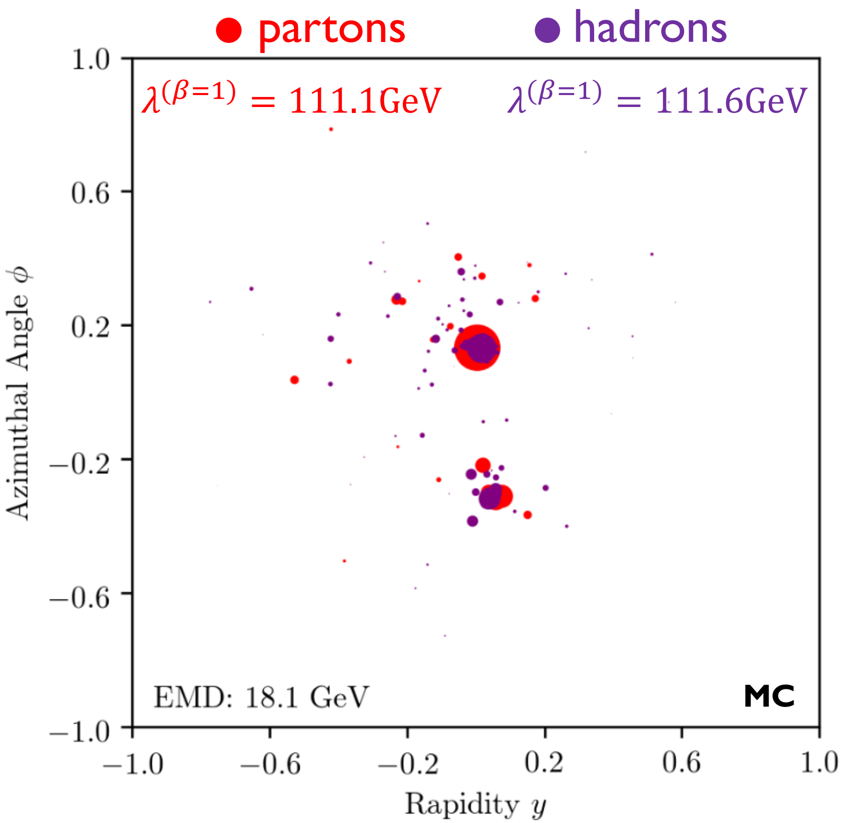
[C. Berger, T. Kucs, and G. Sterman, 0303051]
[A. Larkoski, J. Thaler, and W. Waalewijn, 1408.3122]

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

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

A Geometric Language for Observables

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



$$\left|\lambda^{(\beta=1)}(\mathcal{E}) - \lambda^{(\beta=1)}(\mathcal{E}')\right| = 0.5 \text{ GeV}$$

$$\text{EMD}(\mathcal{E}, \mathcal{E}') = 18.1 \text{ GeV}$$

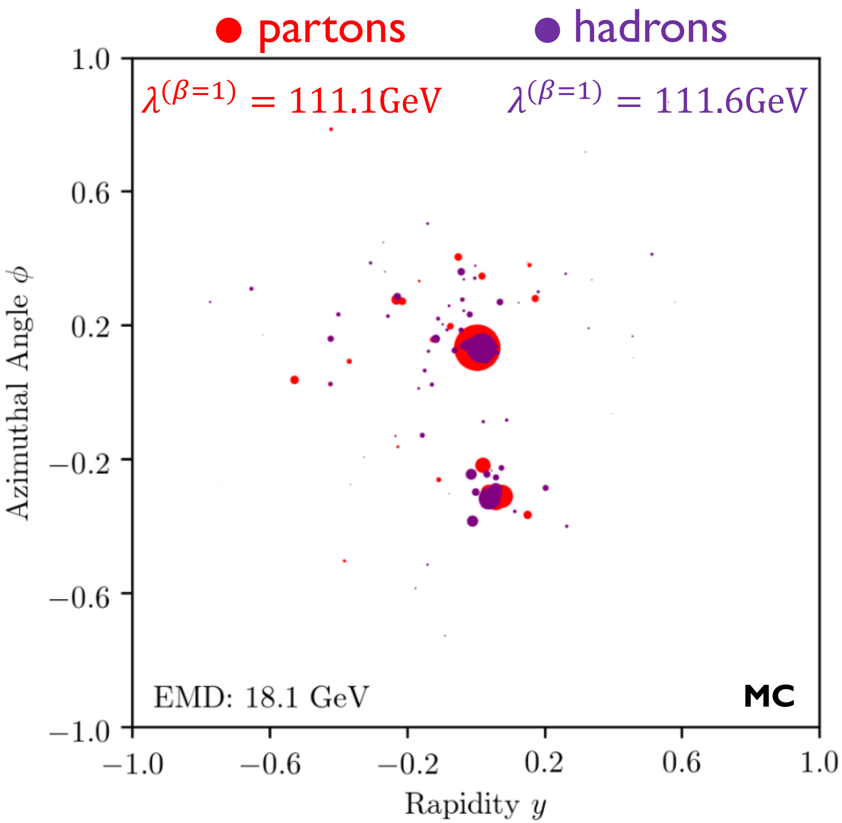
$$\mathcal{E} = \mathcal{E}_{\text{partons}}$$

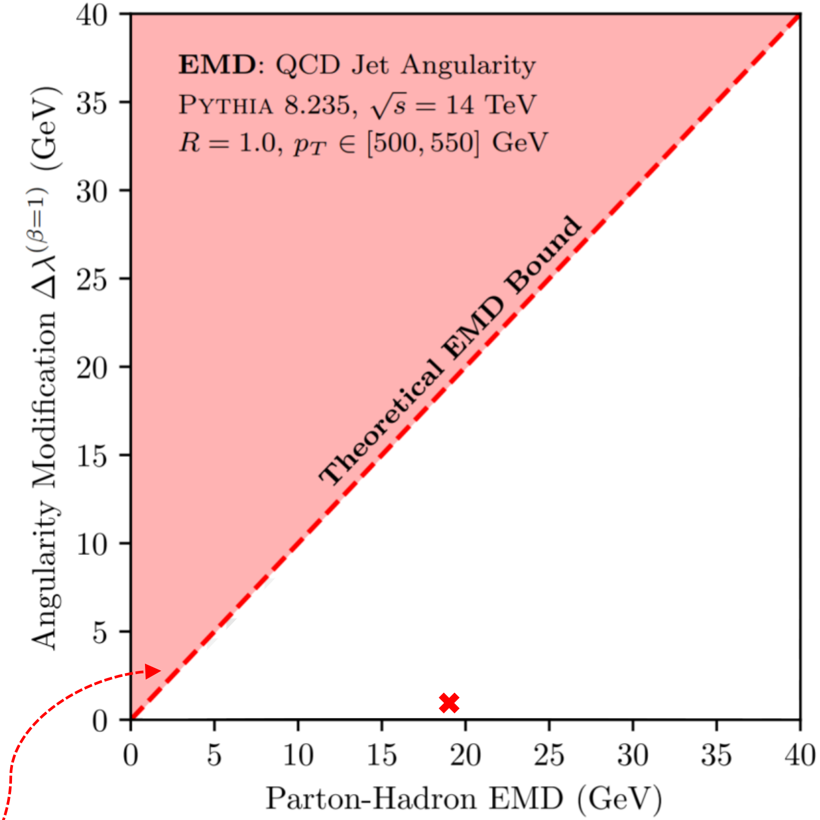
$$\mathcal{E}' = \mathcal{E}_{\text{hadrons}}$$

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

A Geometric Language for Observables

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





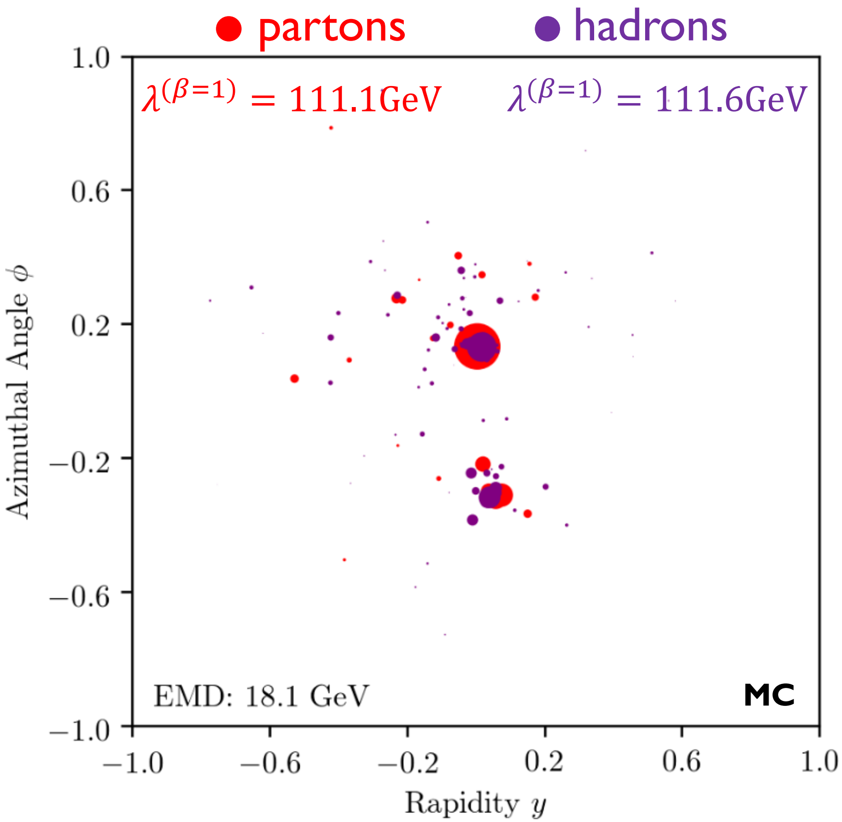
$\mathcal{E} = \mathcal{E}_{\text{partons}}$

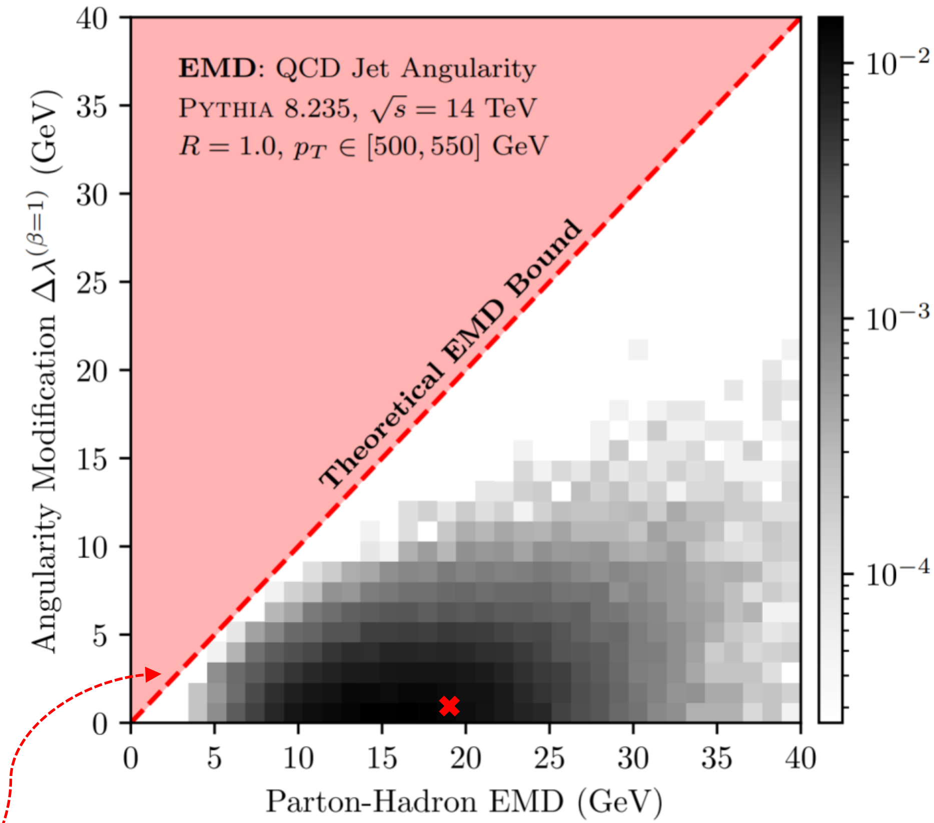
$\mathcal{E}' = \mathcal{E}_{\text{hadrons}}$

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

A Geometric Language for Observables

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





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

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

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

Can similarly bound modifications due to detector effects and pileup. Shown in Extra Slides.

Outline

The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

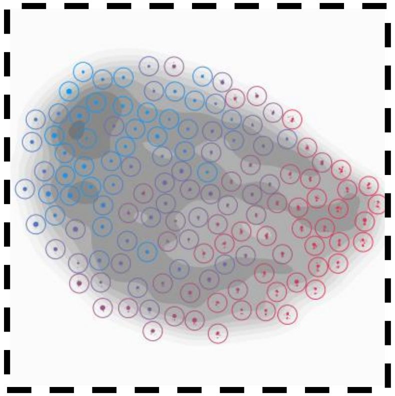
The Energy Mover’s Distance

The “work” to rearrange one event into another

A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization



Exploring the Space of Jets with CMS Open Data

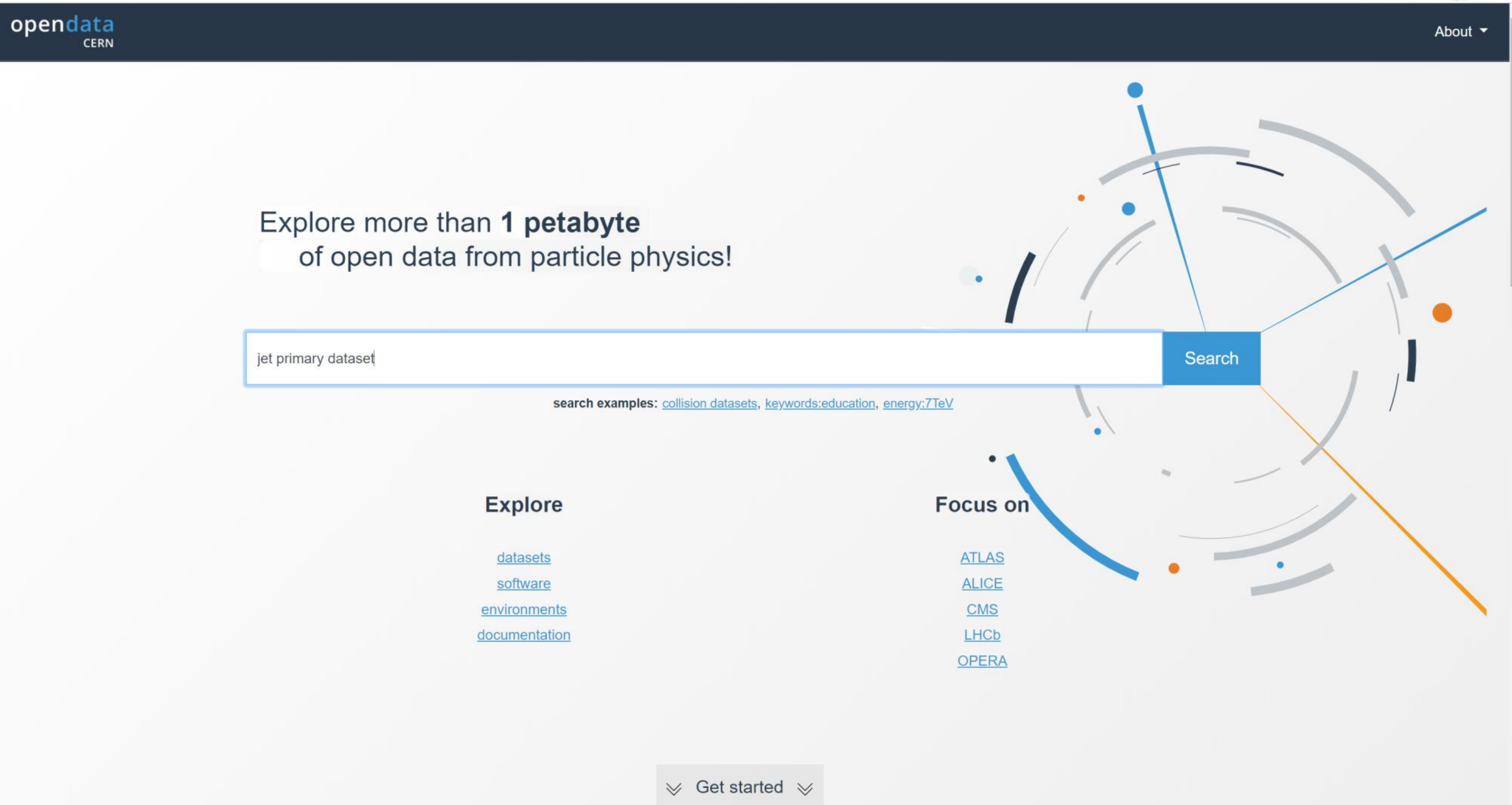
Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

CMS Open Data

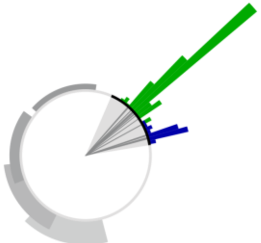
An amazing resource for physics exploration and exploratory studies.

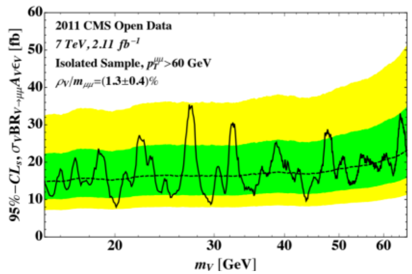
opendata.cern.ch

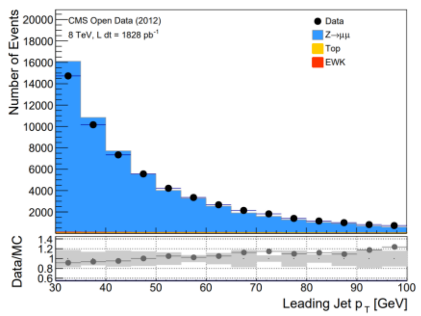


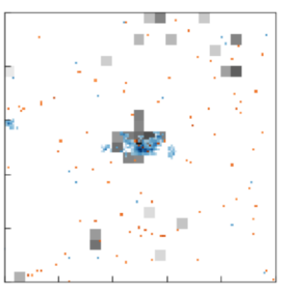
CMS Open Data

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]

Searching for new physics

[Cesarotti, Soreq, Strassler, Thaler, Xue, 1902.04222]
[Lester, Schott, 1904.11195]

Understanding the Standard Model

[A. Apyan, et al., 1907.08197]

[S.P. Mehdiabadi, A. Fahim, et al., 1907.08842]

Analyzing collision data with deep learning techniques

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

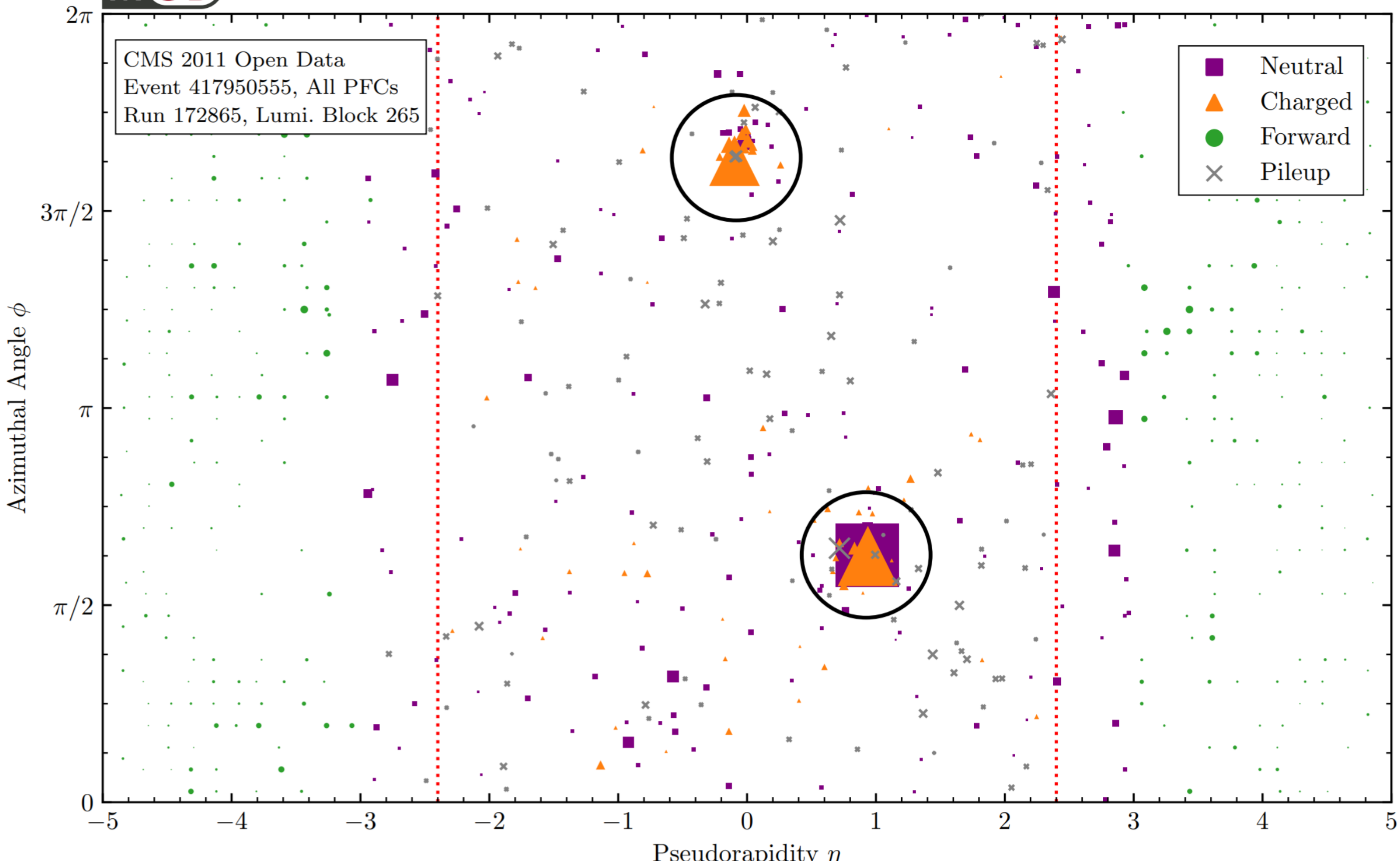
[Andrews, Paulini, Gleyzer, Poczos, 1807.11916]

[Andrews, et al., 1902.08276]

CMS 2011A Jet Primary Dataset (+ Simulation)

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

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



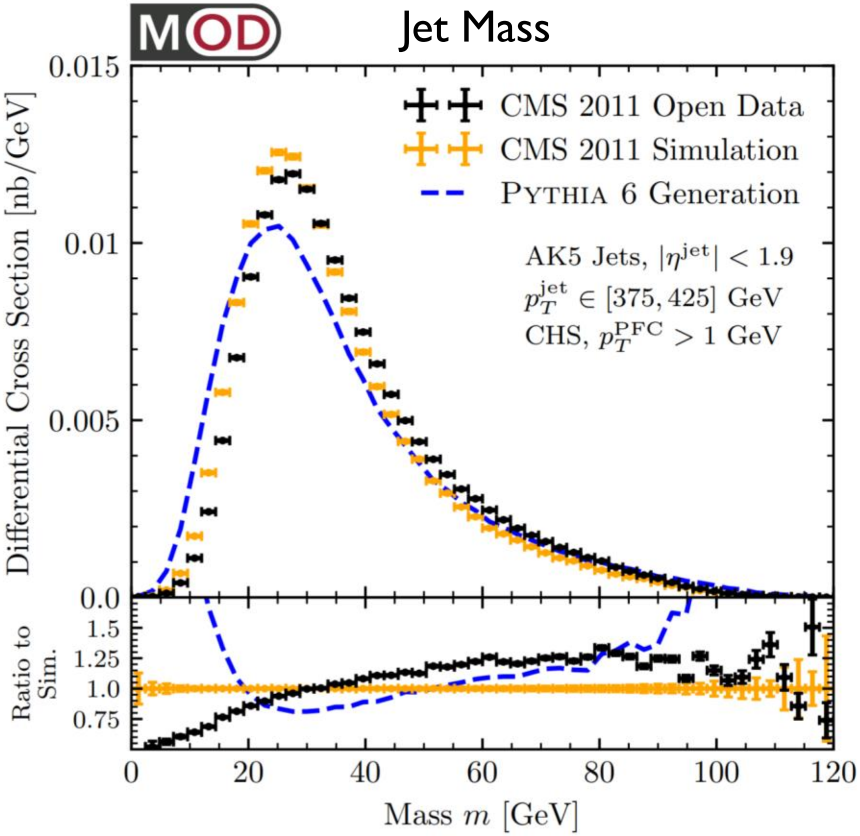
[Komiske, Mastandrea, **EMM**, Naik, Thaler, 1908.08542]

Jet Substructure Observables

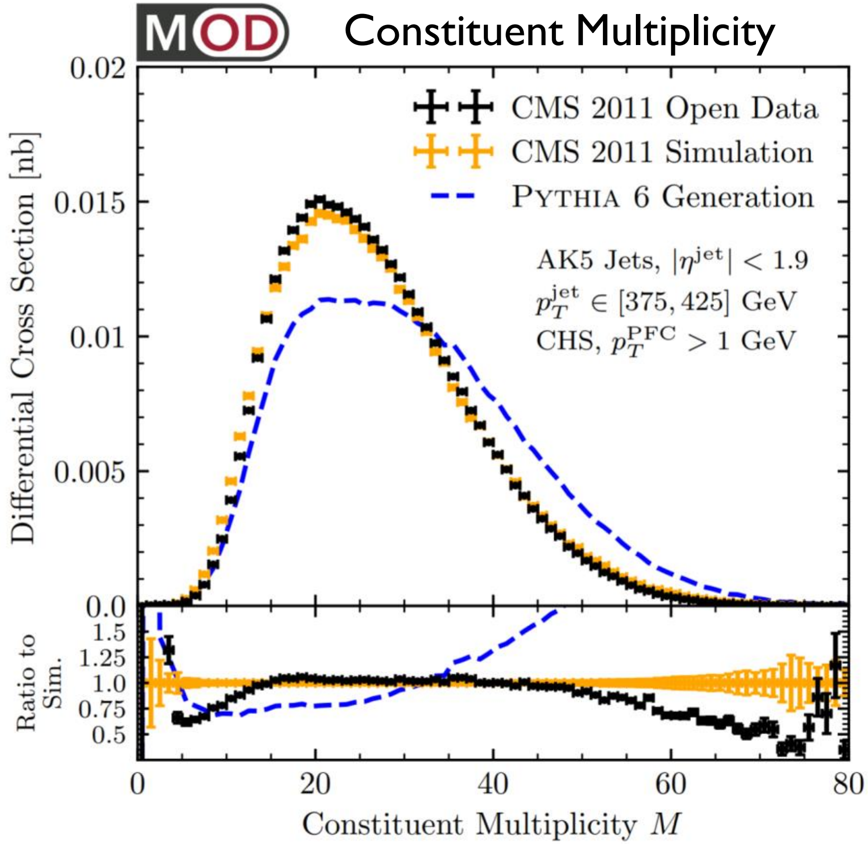
Study the substructure of jets at detector-level and particle-level.

$$m^2 = \left(\sum_{i\in\text{Jet}} p_i^\mu\right)^2 \qquad\qquad M = \sum_{i\in\text{Jet}} 1$$

Jet Mass

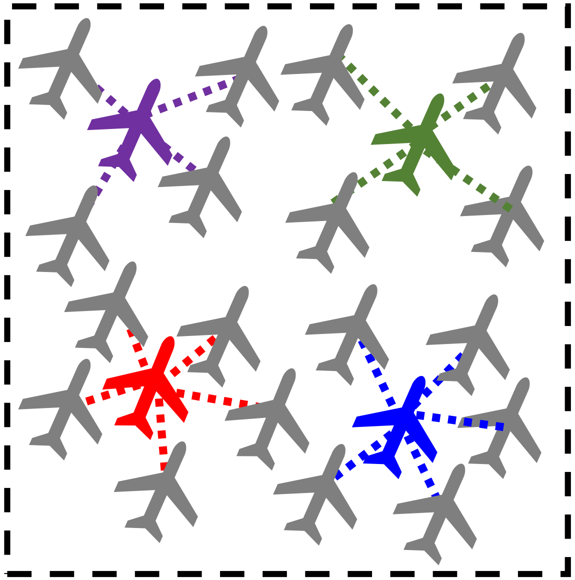


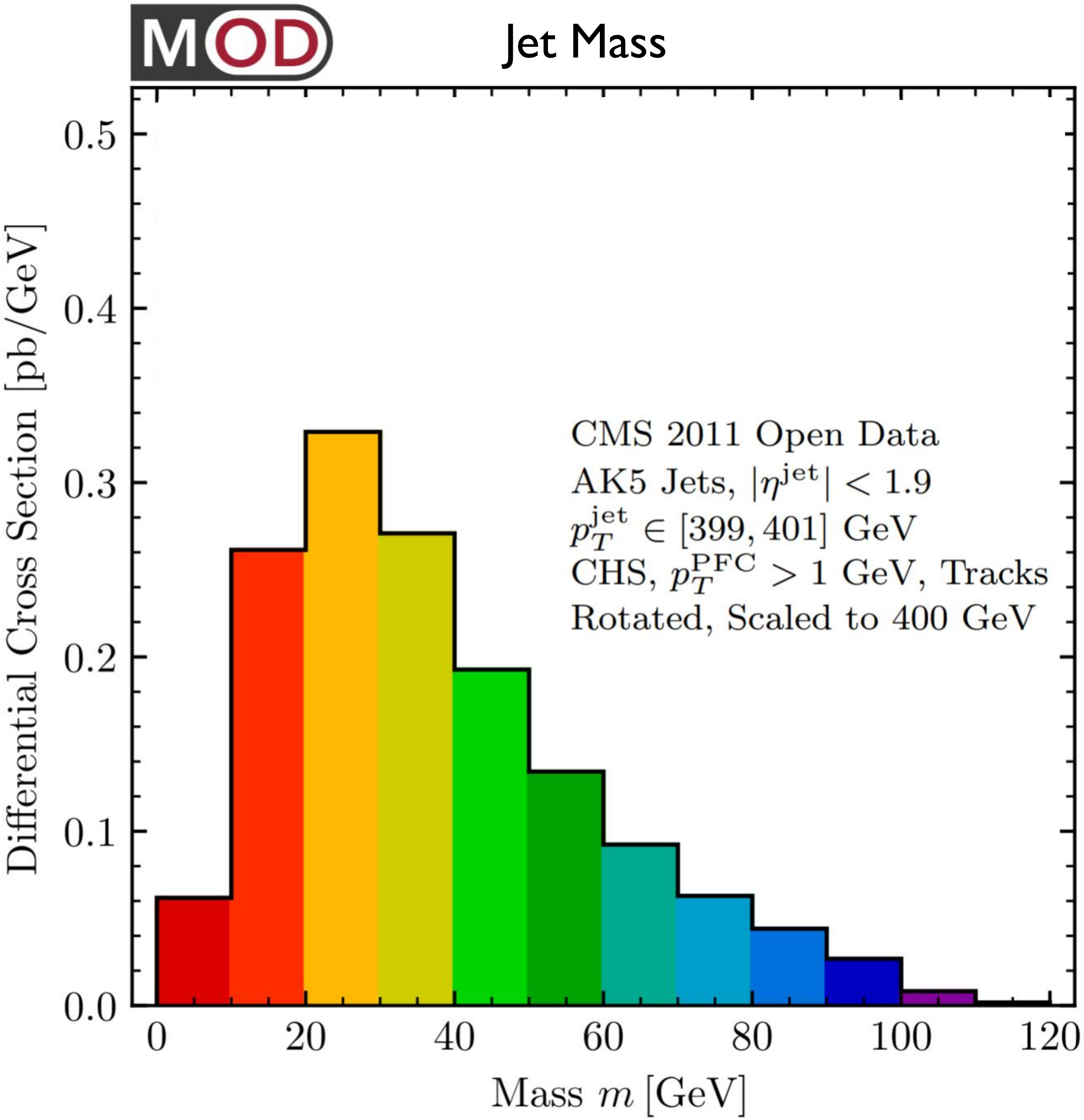
Constituent Multiplicity



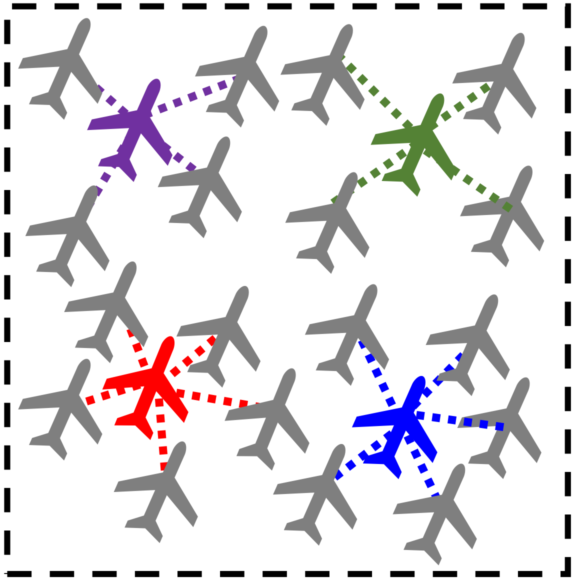
[Komiske, Mastandrea, **EMM**, Naik, Thaler, 1908.08542] similar to [Tripathee, Xue, Larkoski, Marzani, Thaler, 1704.05842]

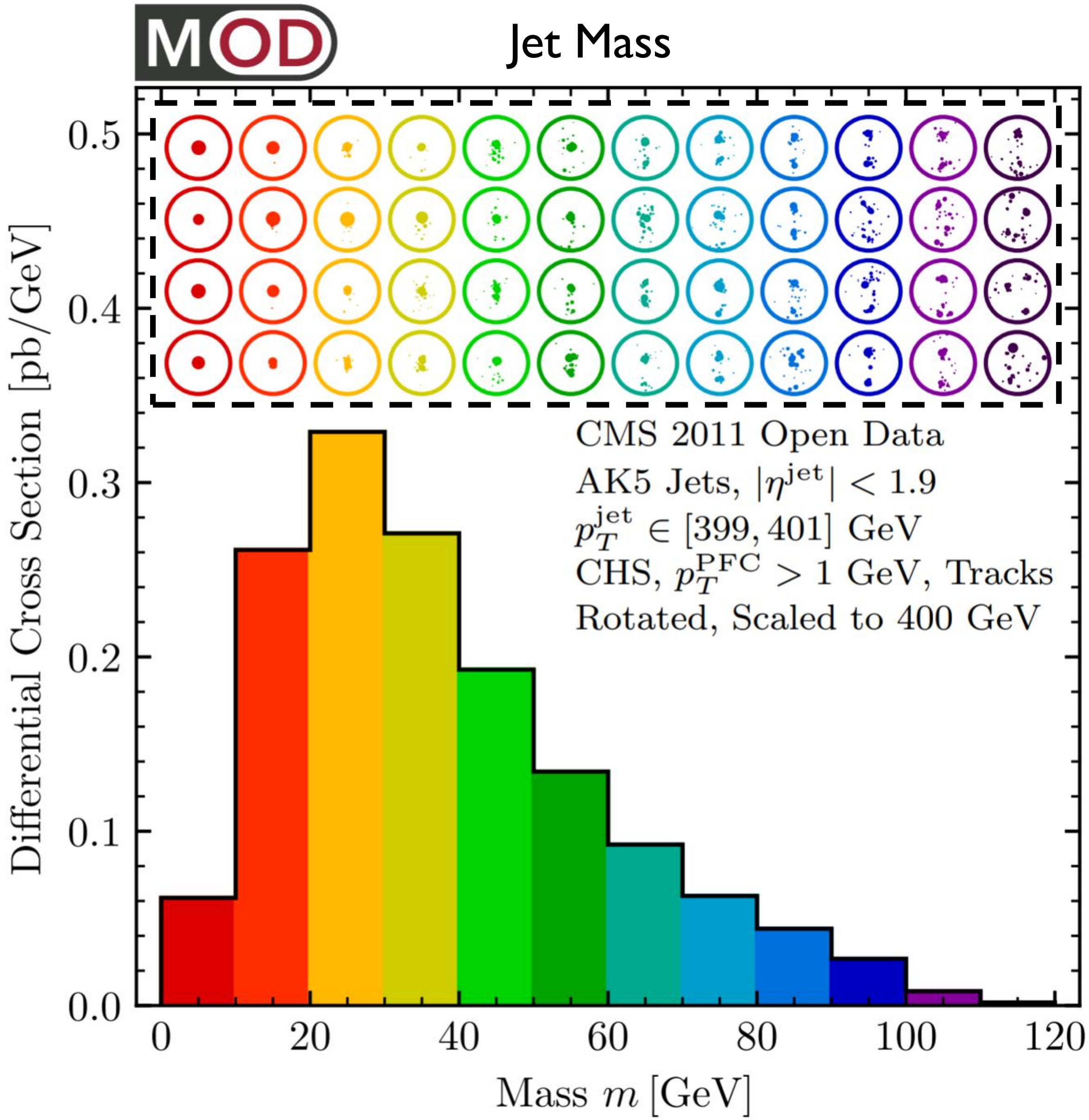
Most Representative Jets: K-medoids



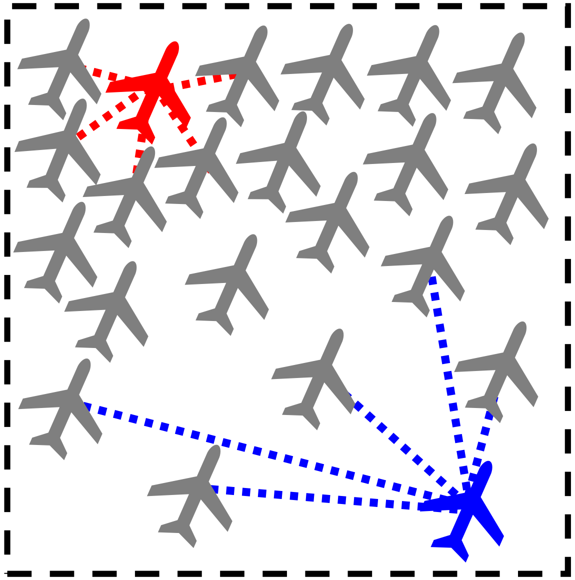


Most Representative Jets: K-medoids

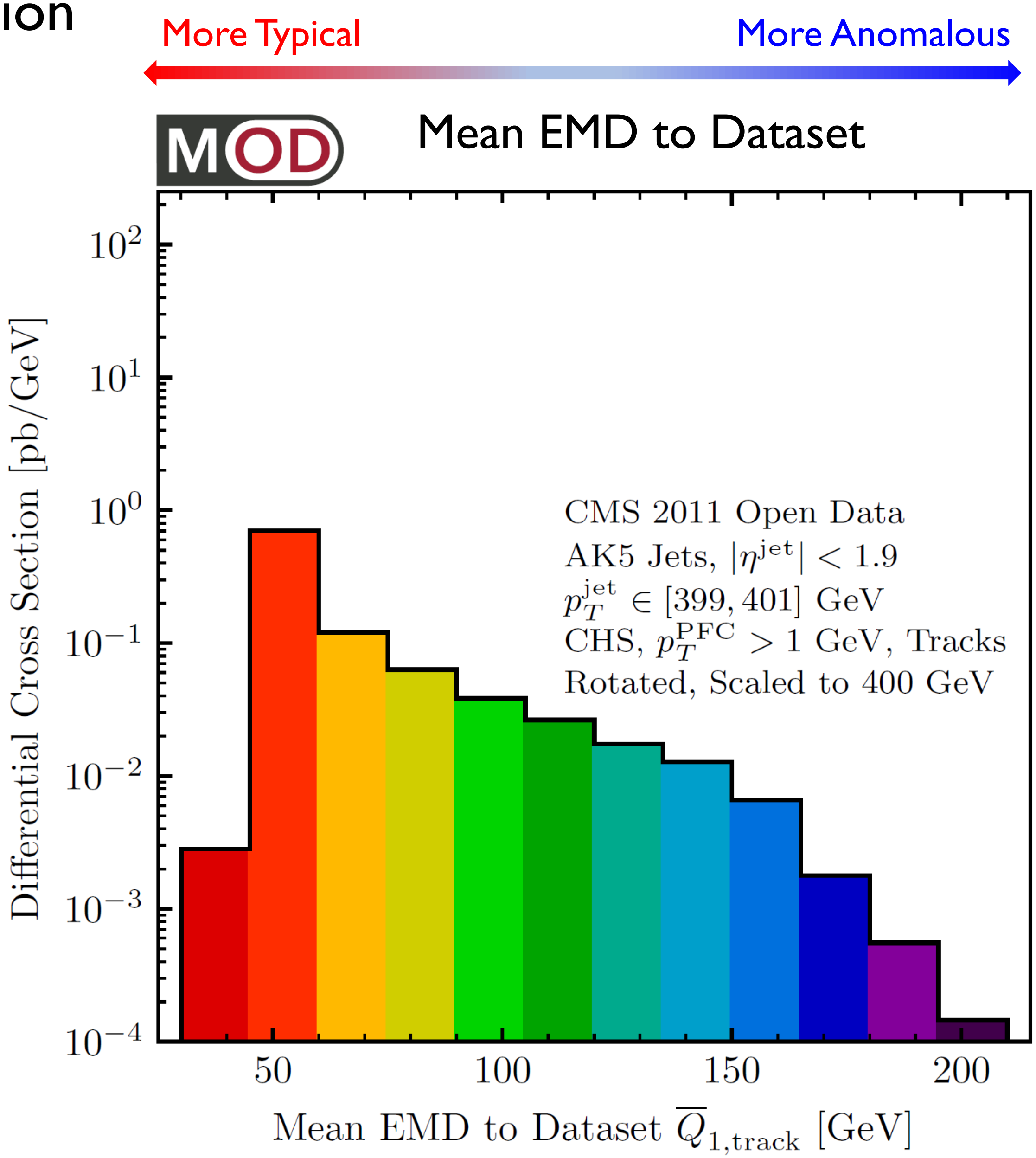




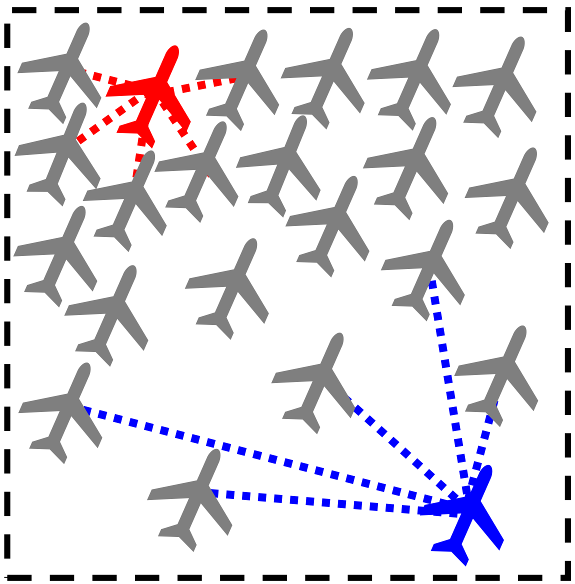
Towards Anomaly Detection



$$\bar{Q}(\mathcal{E}) = \sum_{i=1}^{N} \mathrm{EMD}\,(\mathcal{E}, \mathcal{E}_i)$$



Towards Anomaly Detection

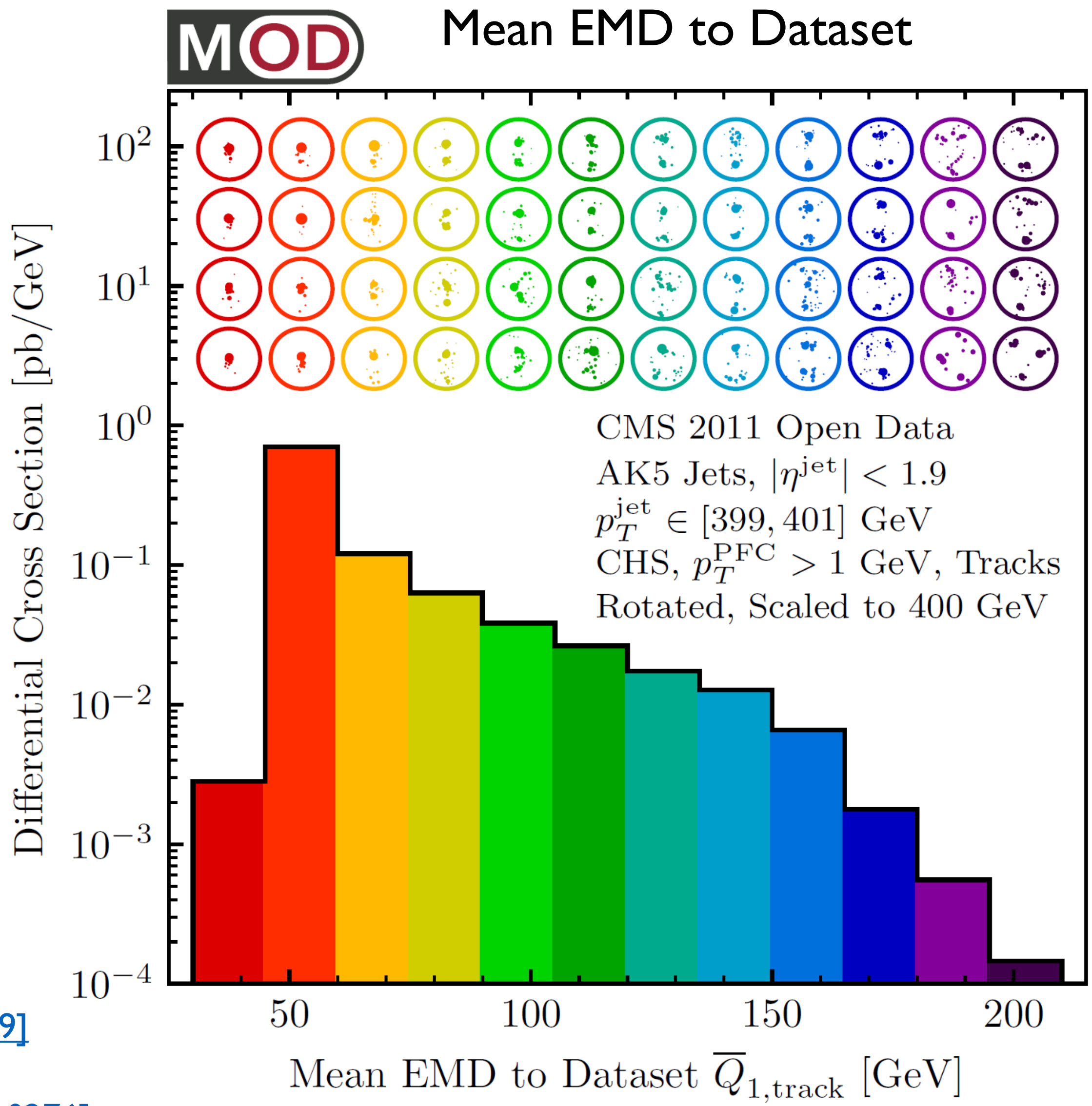


$$\bar{Q}(\mathcal{E}) = \sum_{i=1}^{N} \text{EMD}\,(\mathcal{E}, \mathcal{E}_i)$$

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]

More Typical — More Anomalous



Outline

The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

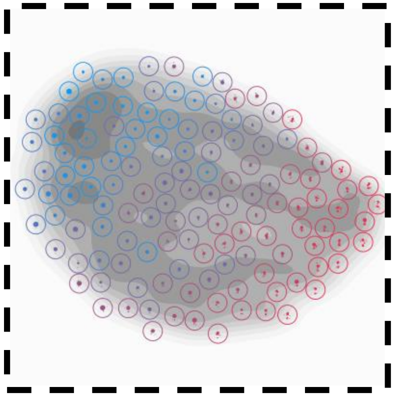
The Energy Mover's Distance

The "work" to rearrange one event into another

A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization



Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

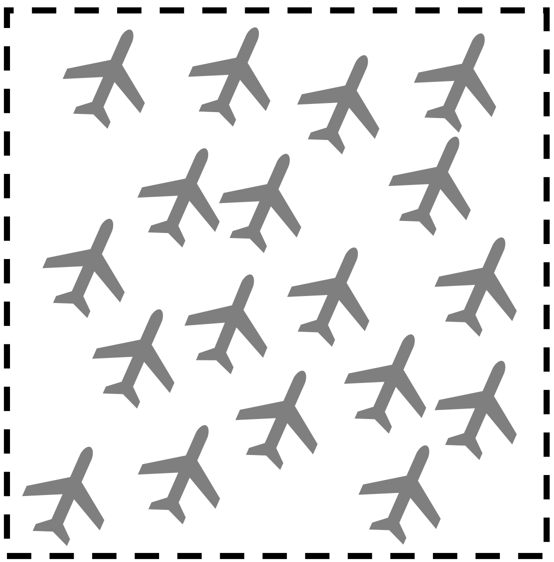
Visualizing the Space and its (fractal) dimension

Exploring the Space of Jets: Visualizing the Manifold

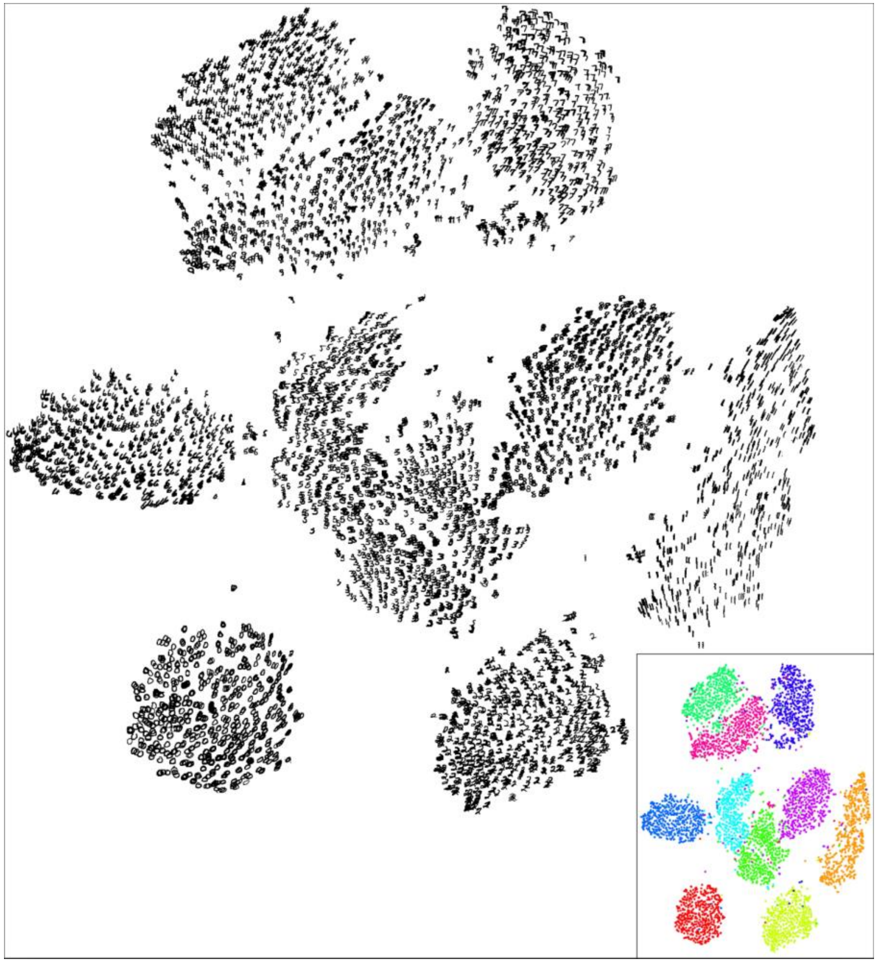
Visualize the space of events with t-Distributed Stochastic Neighbor Embedding (t-SNE).

[L. van der Maaten, G. Hinton]

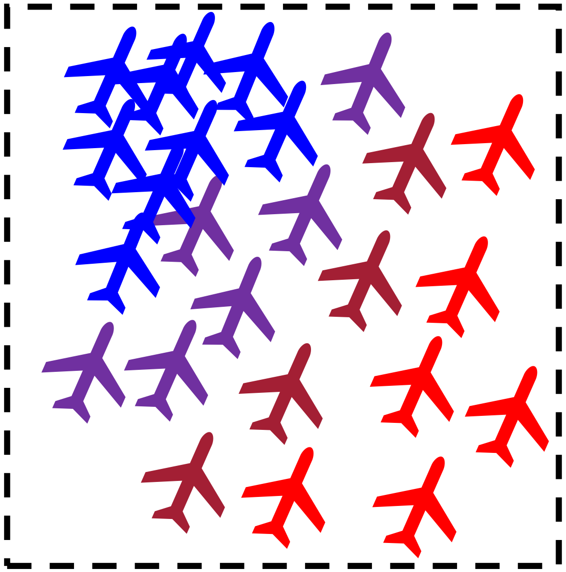
Finds an embedding into a low-dimensional manifold that respects distances.



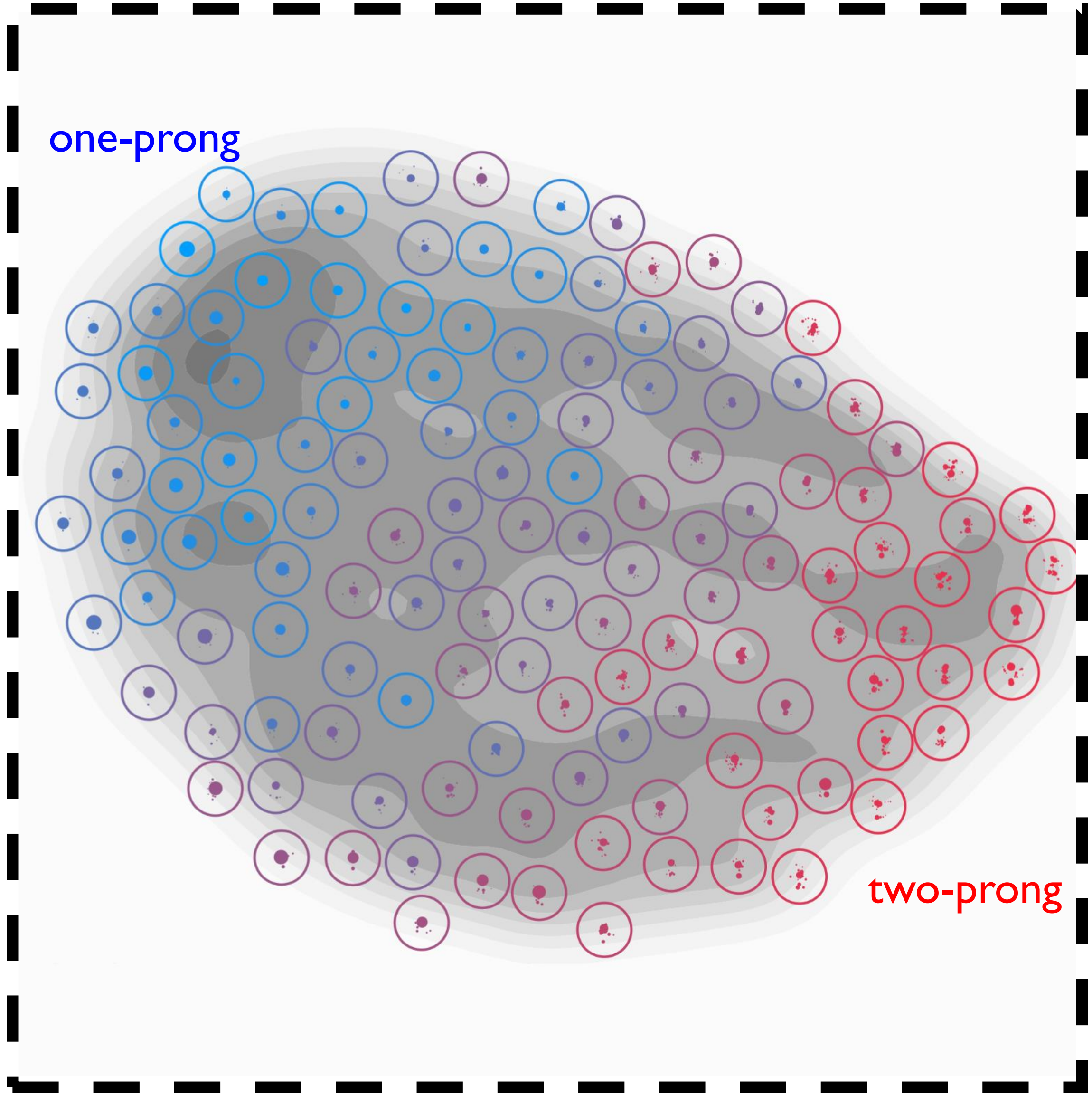
What does the space of jets look like?



Exploring the Space of Jets: Visualizing the Manifold

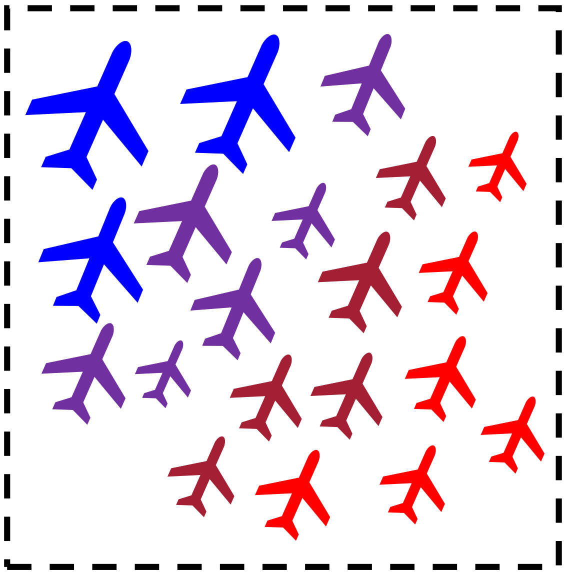


What does the space of jets look like?

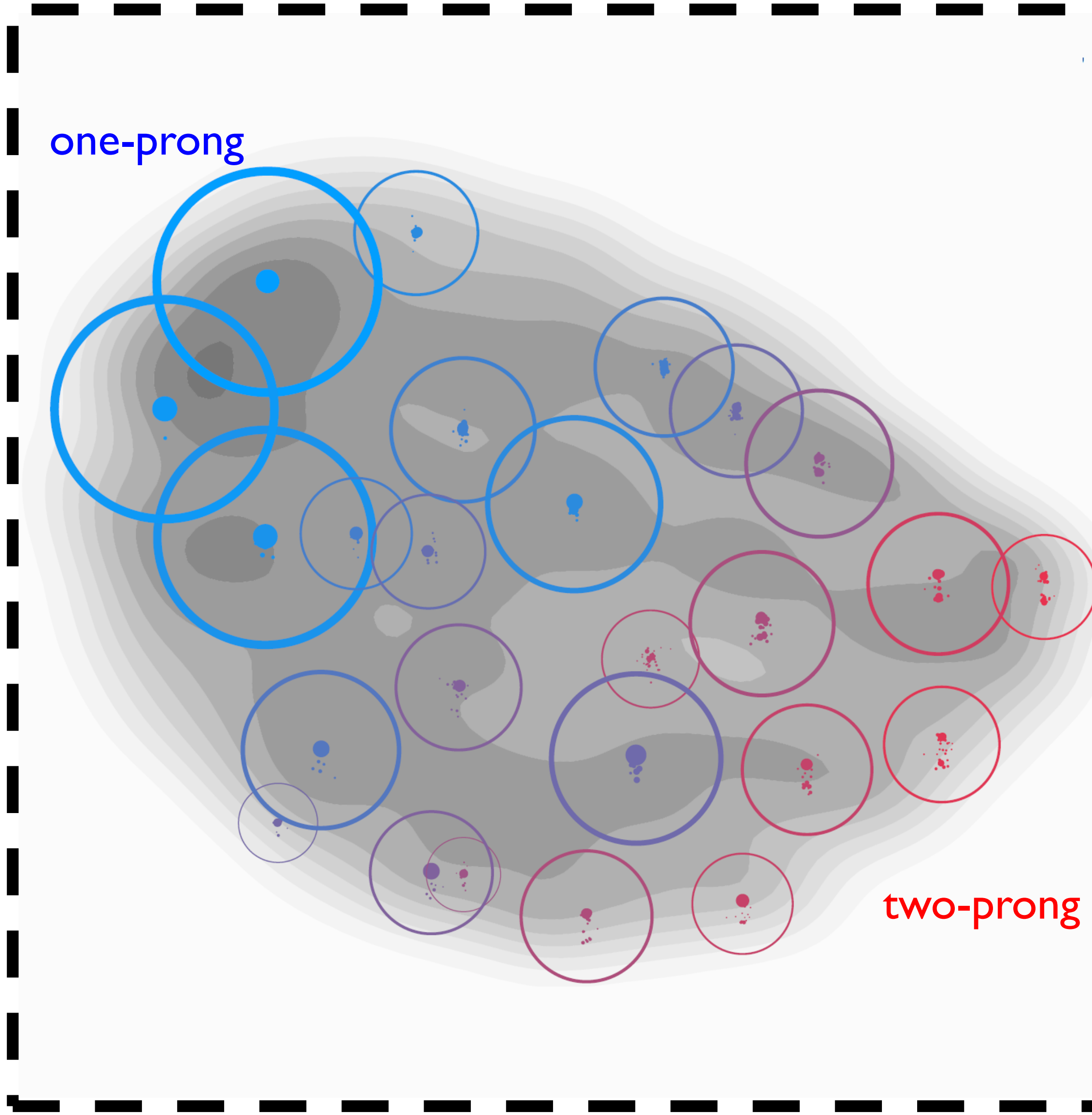


Uniformly distributed example jets

Exploring the Space of Jets: Visualizing the Manifold



What does the space of jets look like?



25 most representative jets

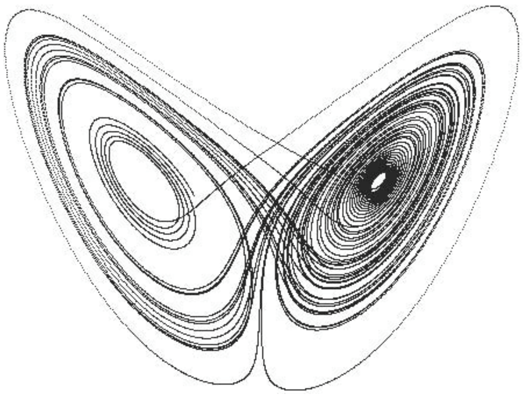
Exploring the Space of Jets: Correlation Dimension

VOLUME 50, NUMBER 5 PHYSICAL REVIEW LETTERS 31 JANUARY 1983

Characterization of Strange Attractors

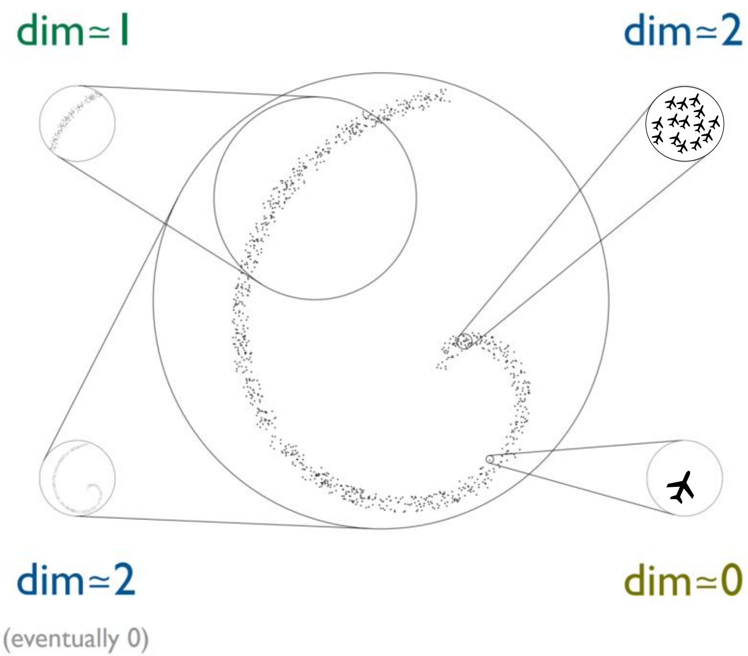
Peter Grassberger[a] and Itamar Procaccia
Chemical Physics Department, Weizmann Institute of Science, Rehovot 76100, Israel
(Received 7 September 1982)

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:

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



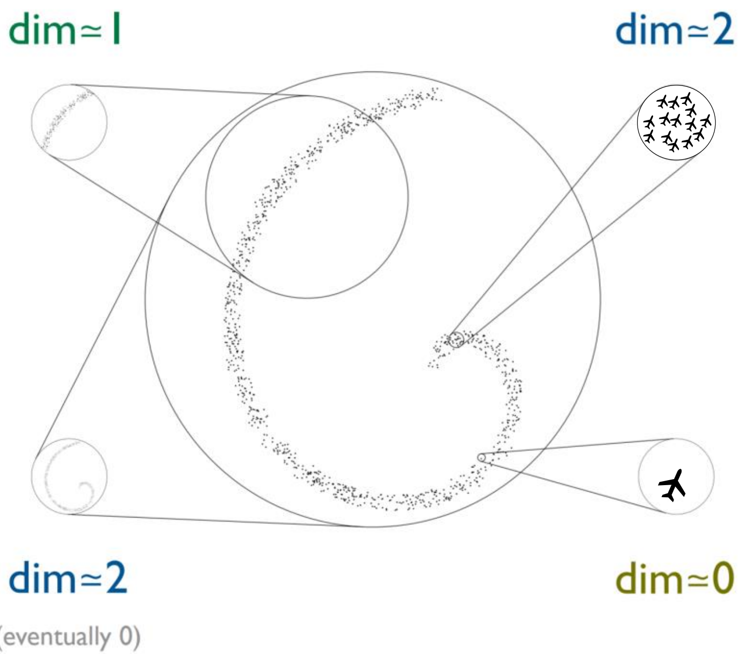
Correlation dimension:

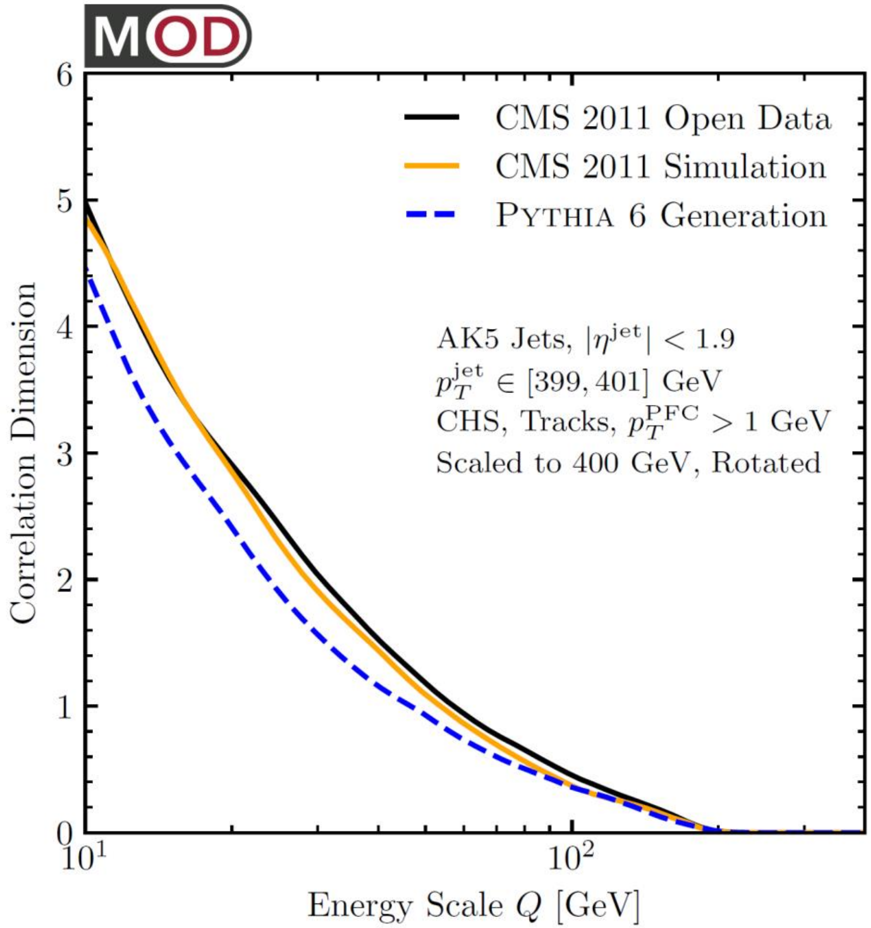
$$\boxed{\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_{i=1}^{N}\sum_{j=1}^{N}\Theta[\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}_j) < Q]}$$

Energy scale Q dependence

Count neighbors in ball of radius Q

Exploring the Space of Jets: Correlation Dimension



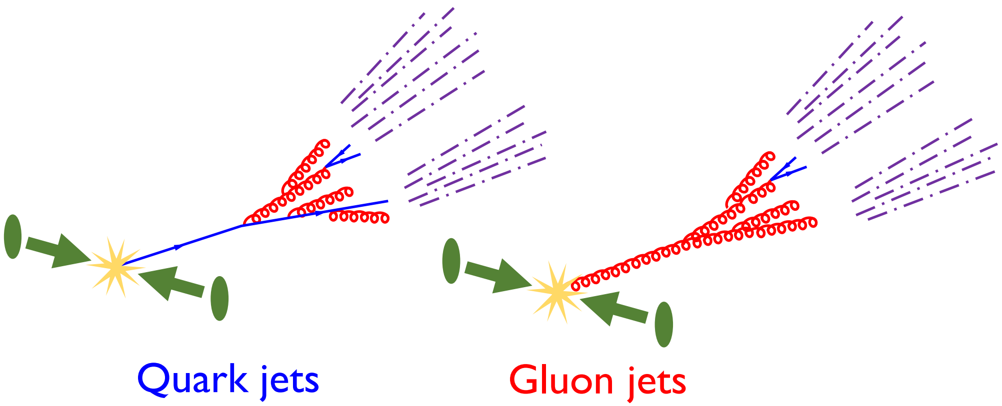


Dimension blows up at low energies.

Jets are "more than fractal"

Can we understand this analytically?

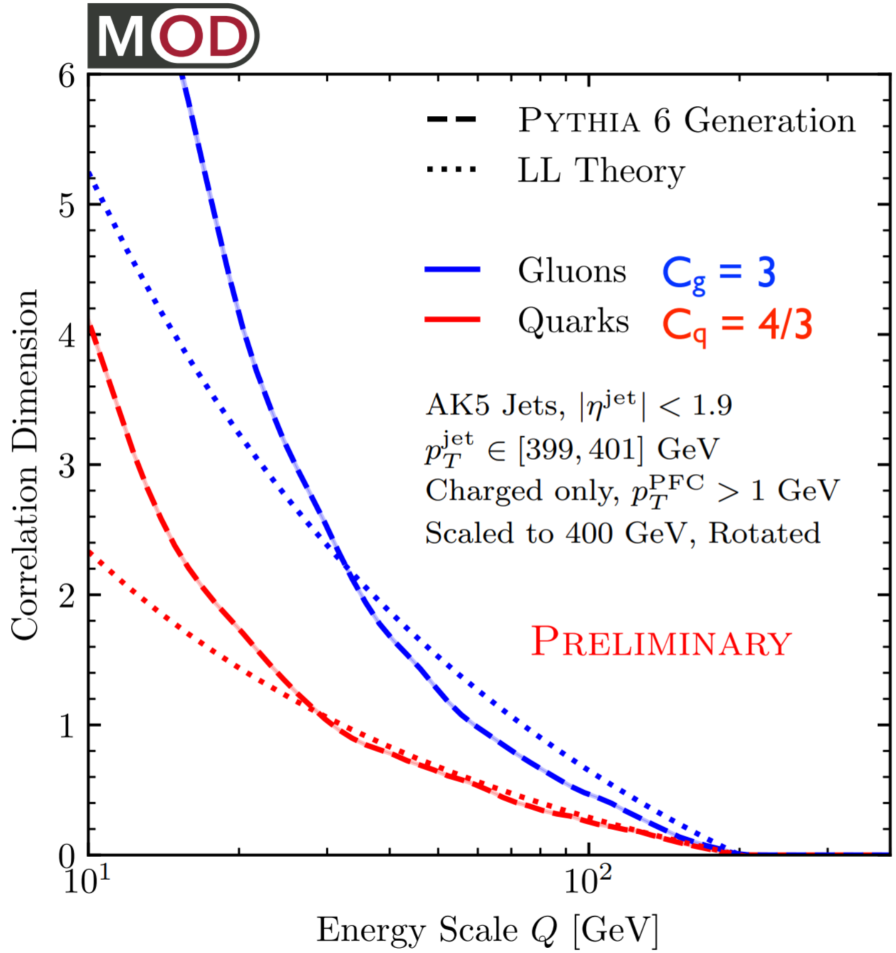
Exploring the Space of Jets: Correlation Dimension



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

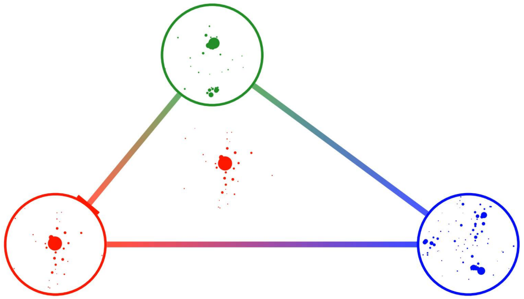
+ 1-loop running of α_s

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



See extra slides for sketch of calculation.

Outline



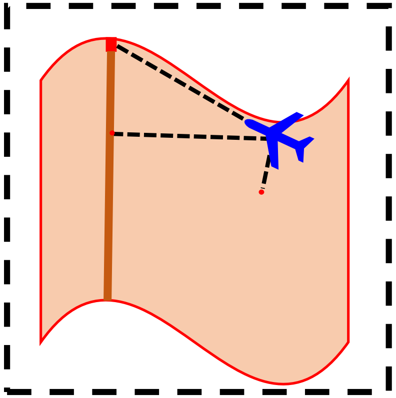
The Metric Space of Collider Events

When are two collider events similar?

When they have similar energy distributions

The Energy Mover's Distance

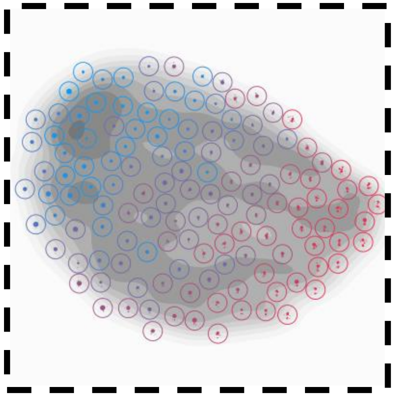
The "work" to rearrange one event into another



A Geometric Language for Observables

Old Observables in a New Light

Quantifying Hadronization

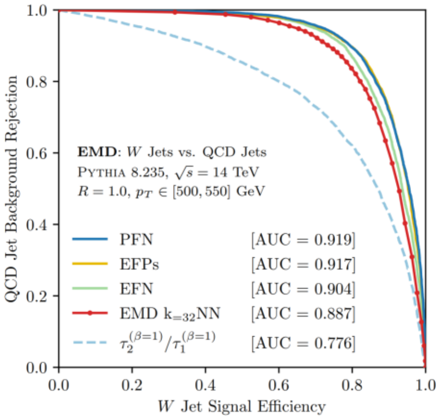


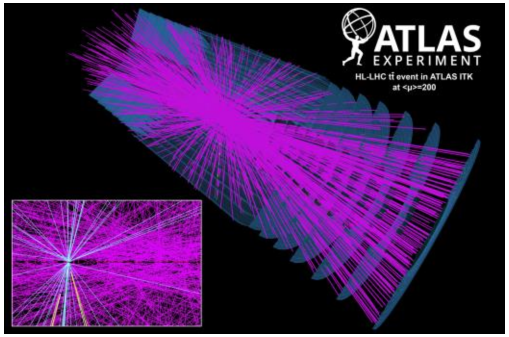
Exploring the Space of Jets with CMS Open Data

Most Representative and Anomalous Jets

Visualizing the Space and its (fractal) dimension

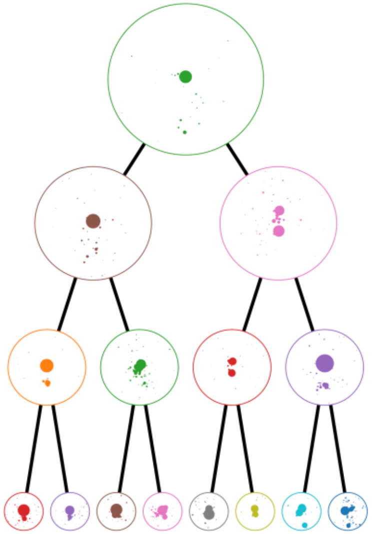
Going Beyond





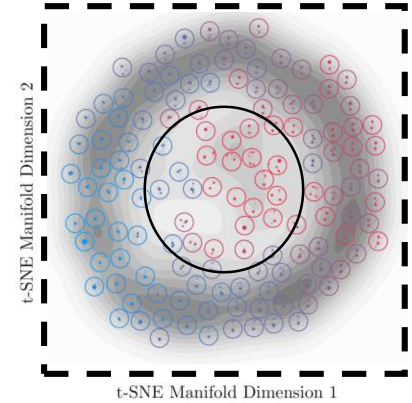
Classification with EMD

Clustering sets of events



Quantifying pileup and detector effects

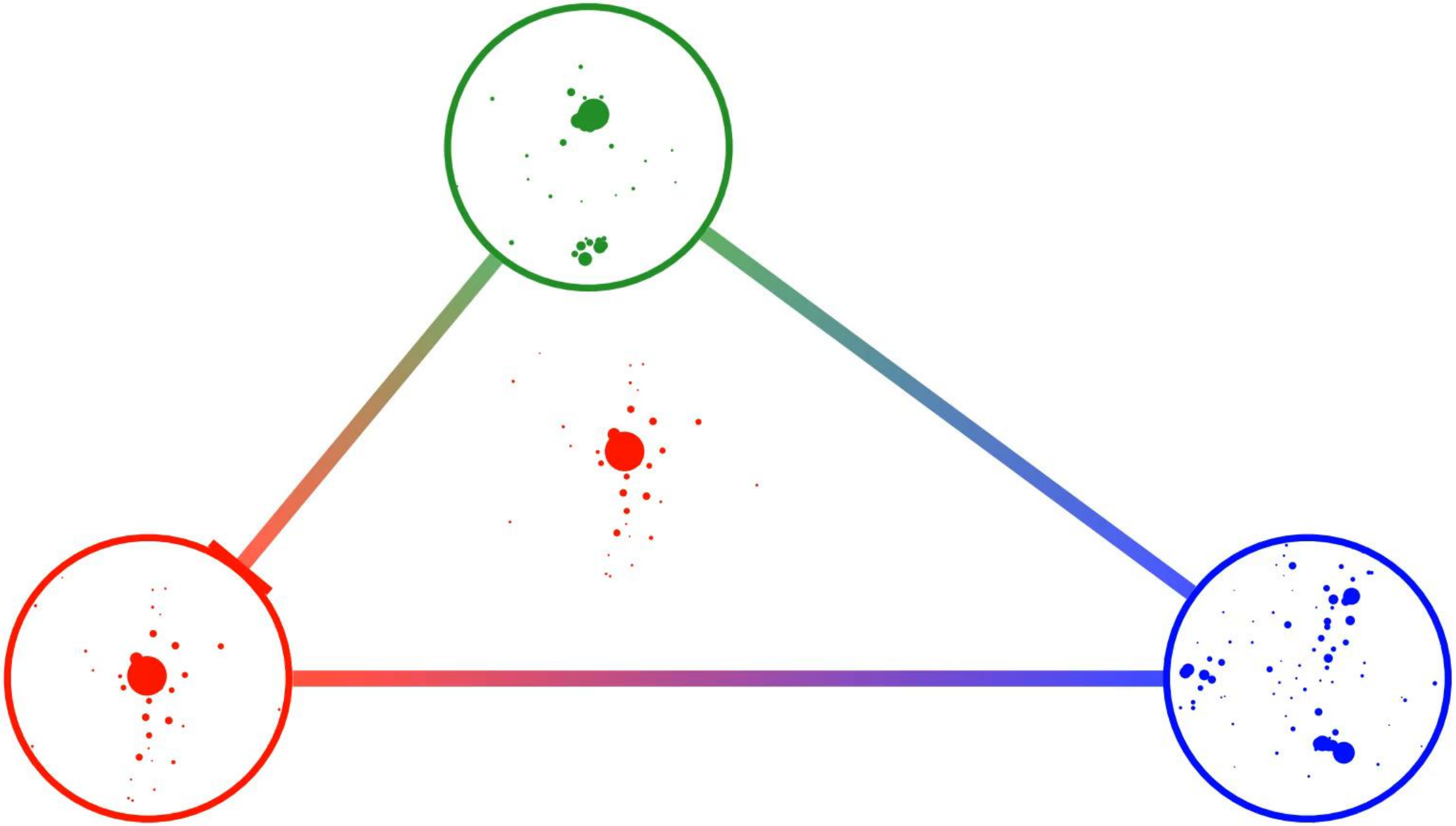
“Event” mover’s distance between ensembles?



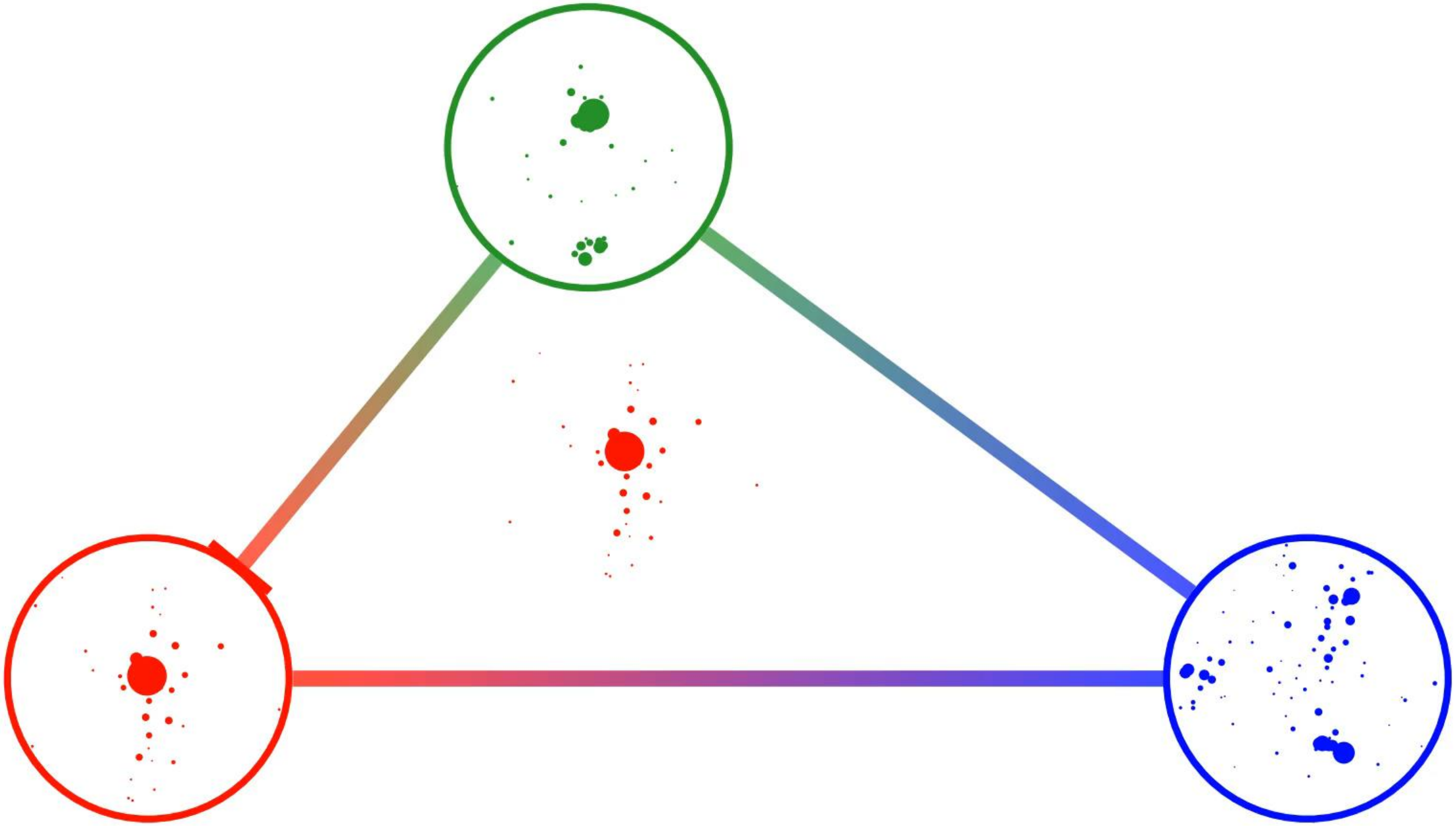
Include flavor information?

The End

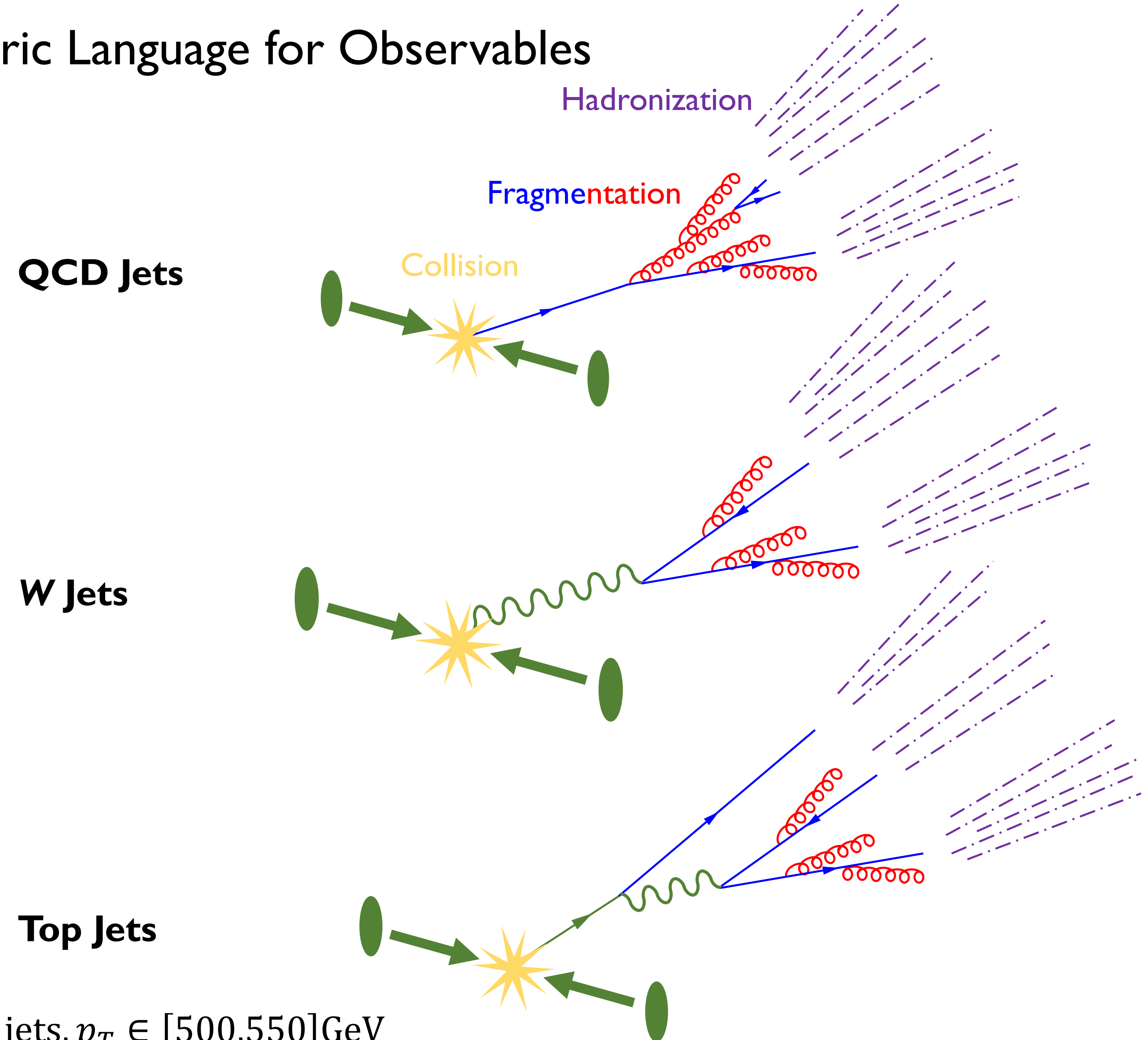
Thank you!



Extra Slides

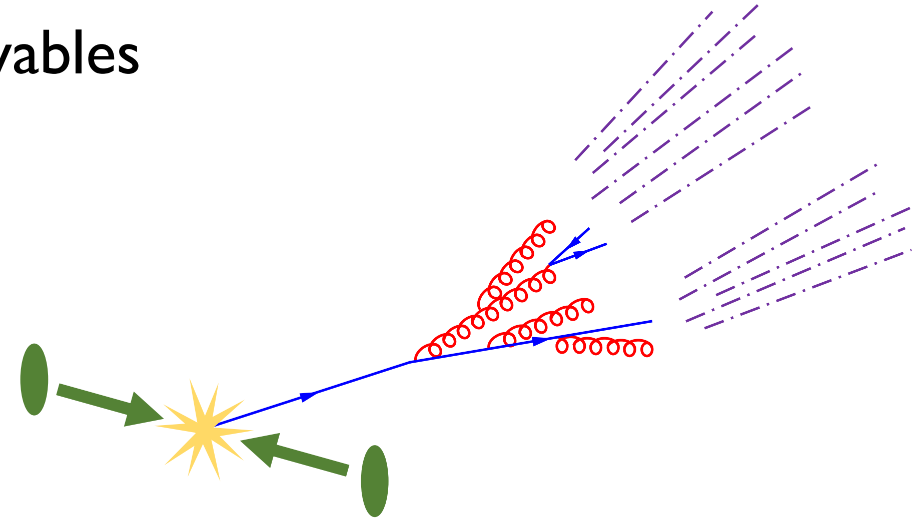


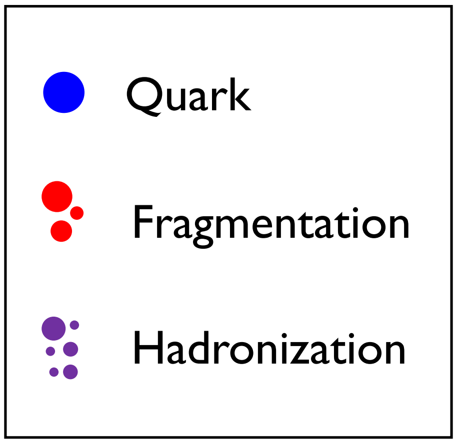
A Geometric Language for Observables

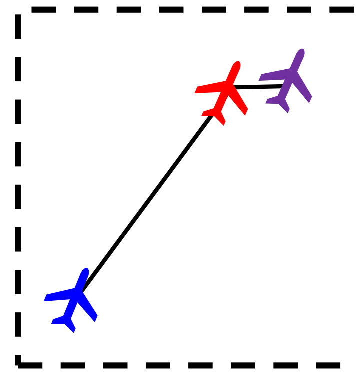


Pythia 8, $R = 1.0$ jets, $p_T \in [500{,}550]\,\mathrm{GeV}$

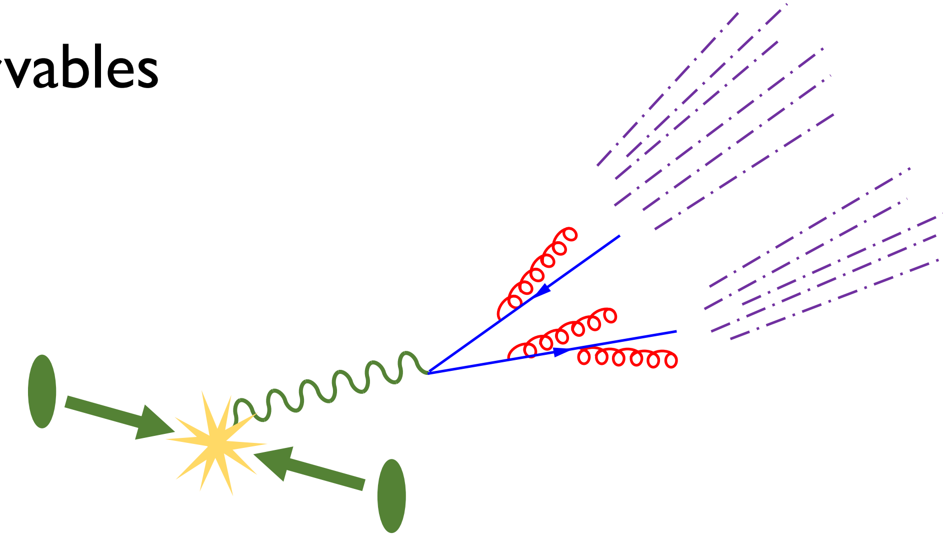
A Geometric Language for Observables

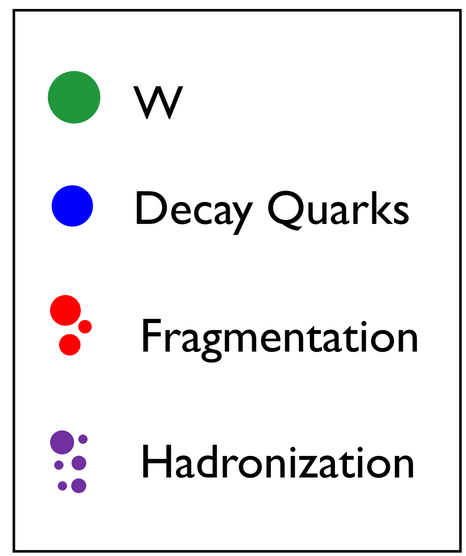


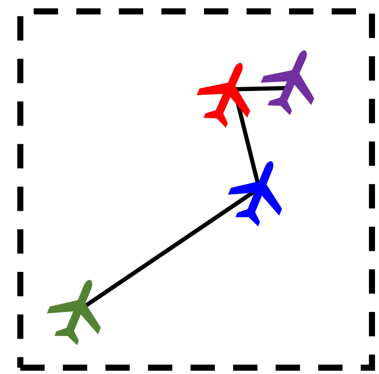




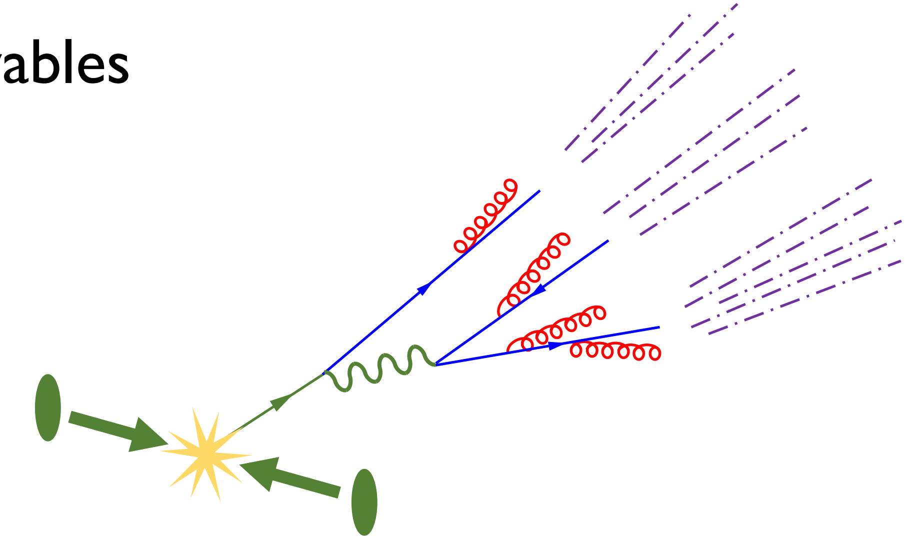
A Geometric Language for Observables

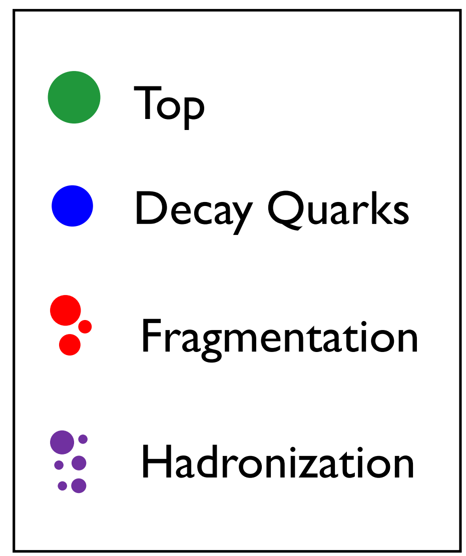


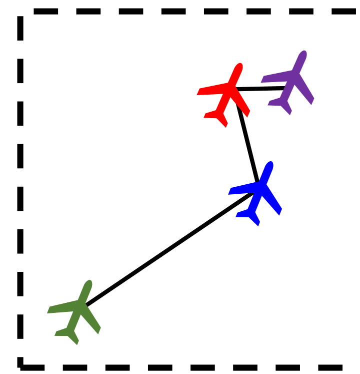




A Geometric Language for Observables

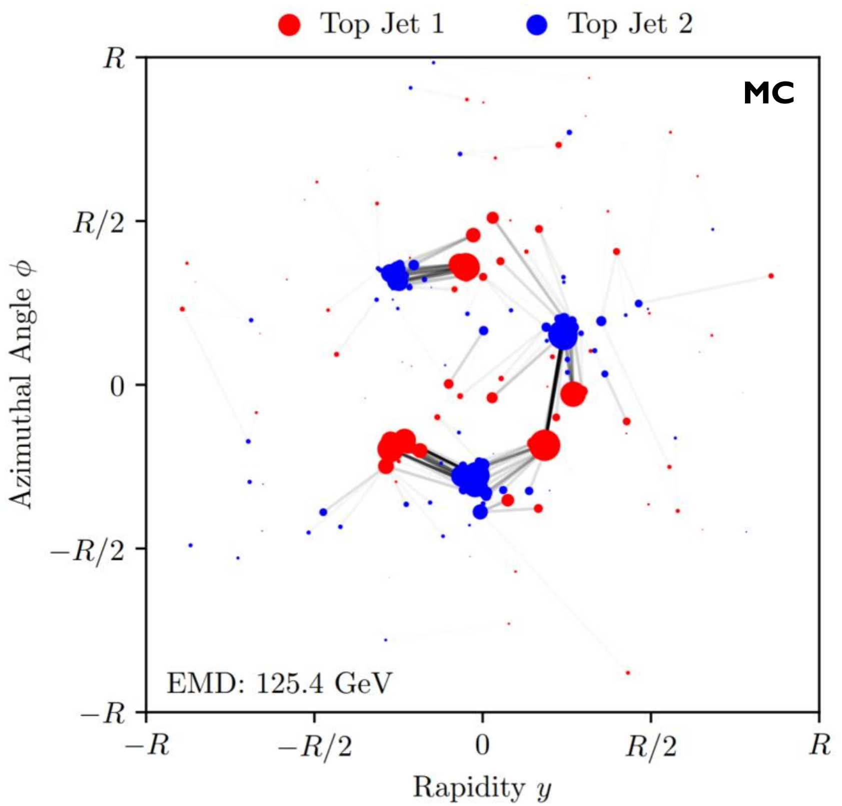


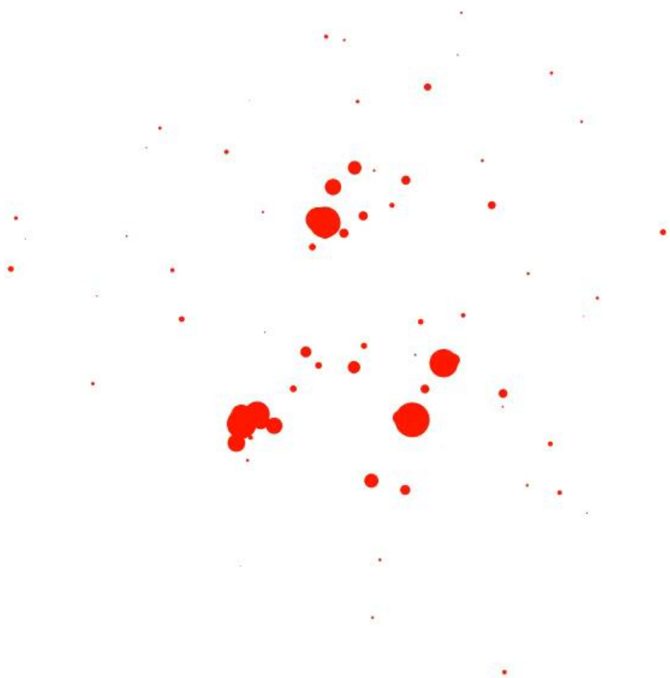


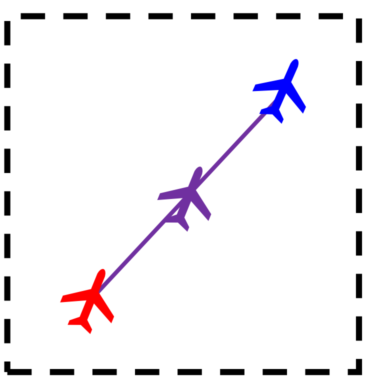


Movie Time: Visualizing the EMD

Taking a walk in the space of events





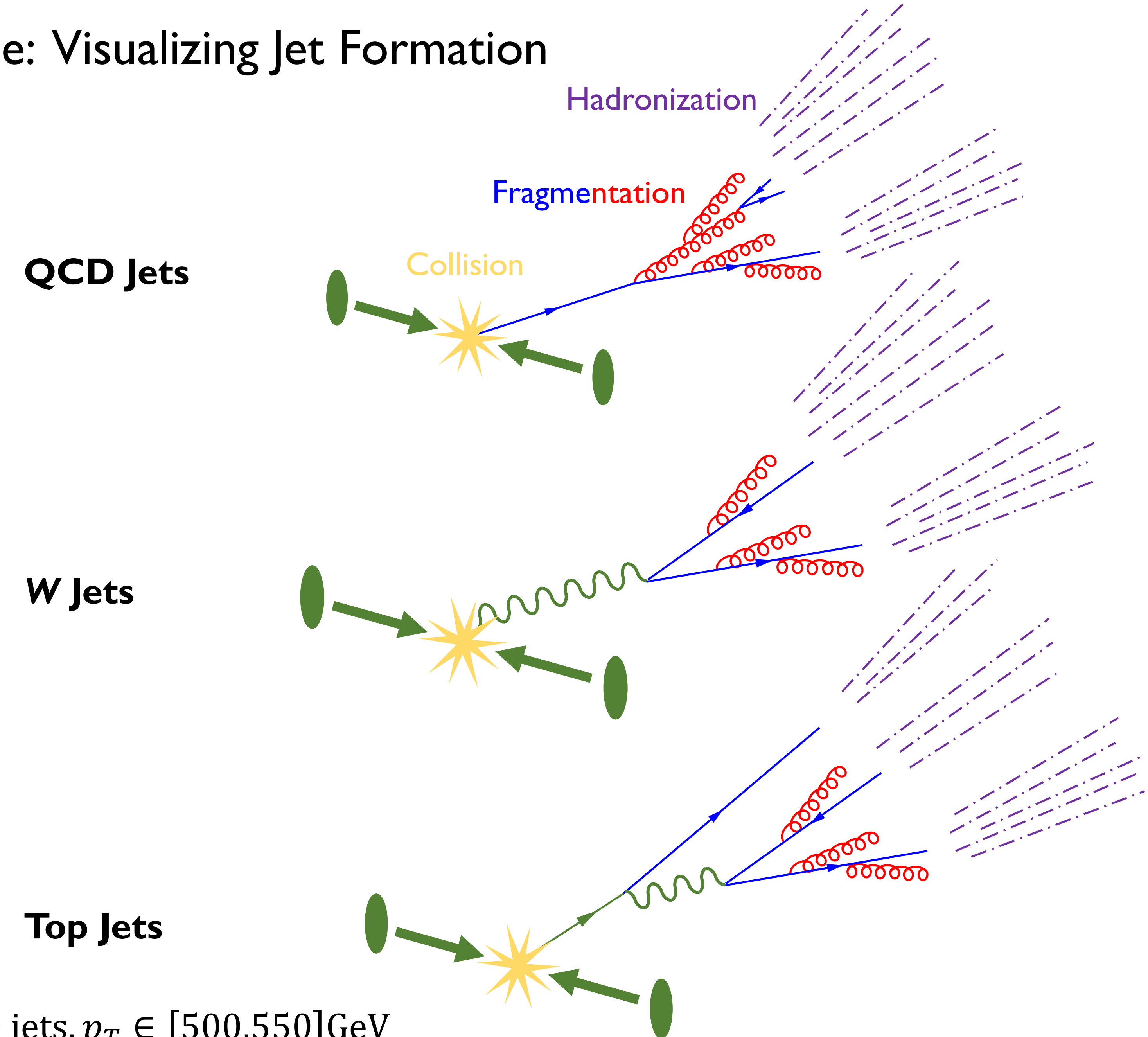


EMD is the cost of an optimal transport problem.

We also get the *shortest path* between the events.

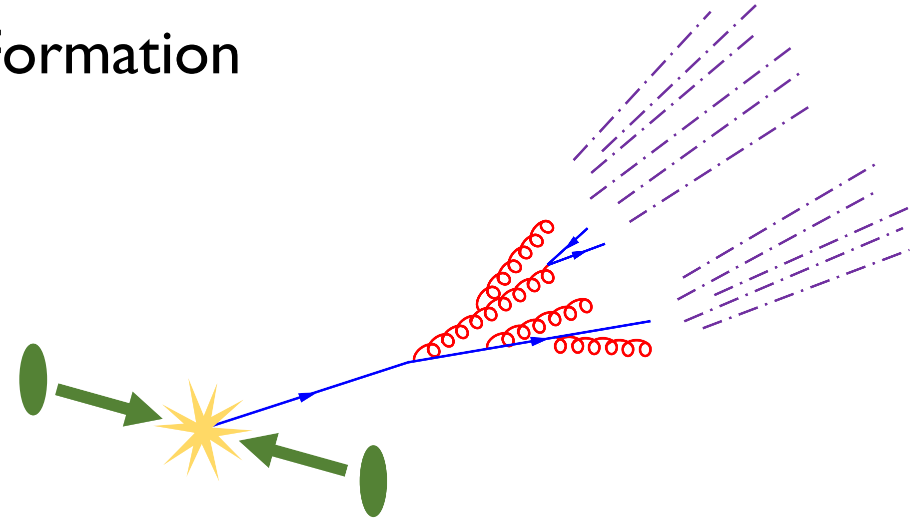
Interpolate along path to visualize!

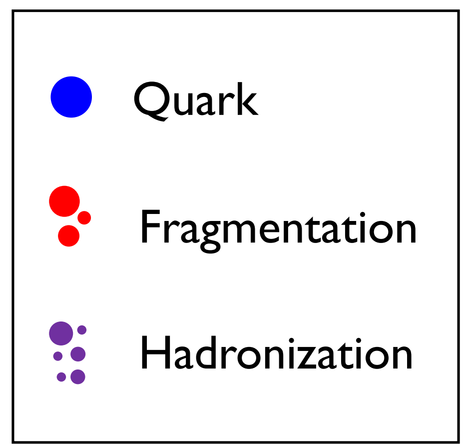
Movie Time: Visualizing Jet Formation

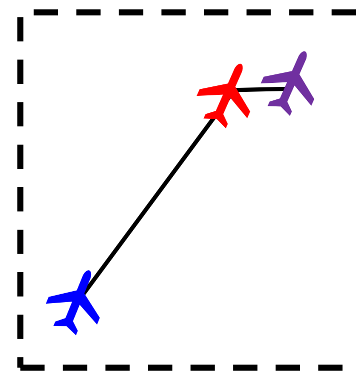


Pythia 8, $R = 1.0$ jets, $p_T \in [500{,}550]\mathrm{GeV}$

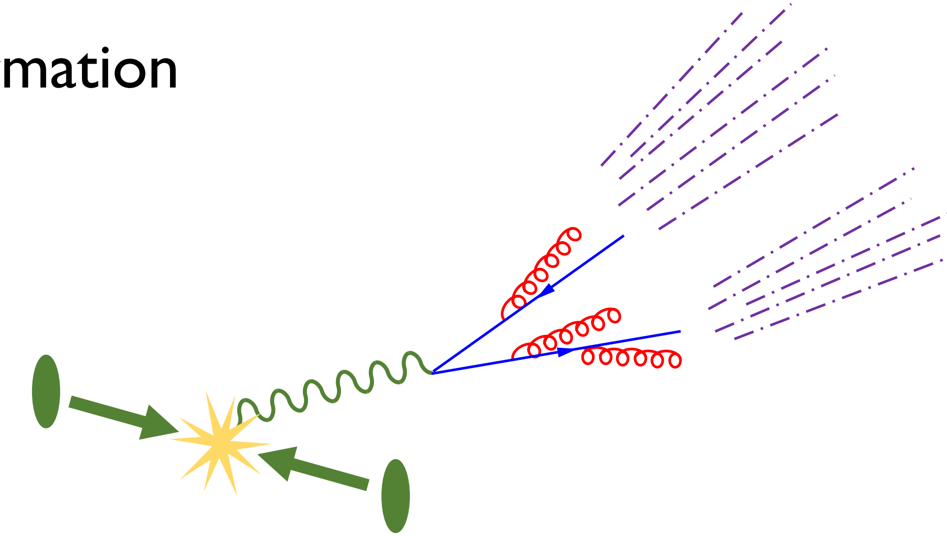
Movie Time: Visualizing QCD Jet Formation

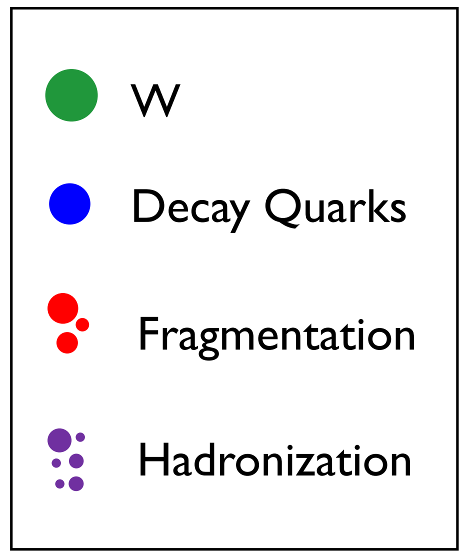


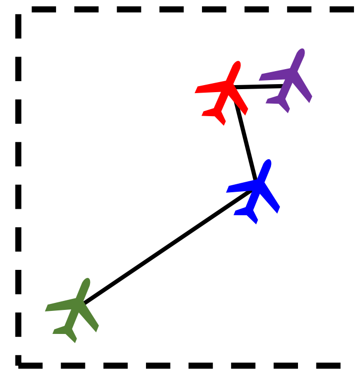




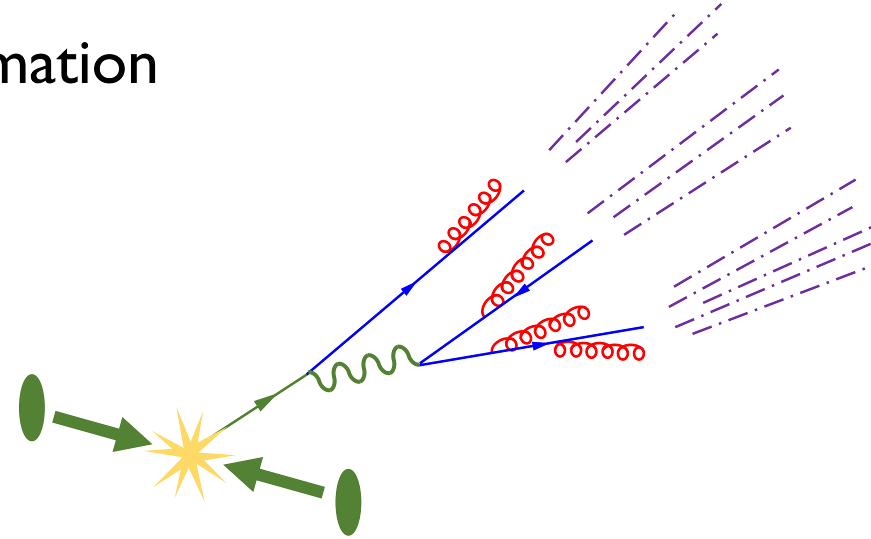
Movie Time: Visualizing W Jet Formation

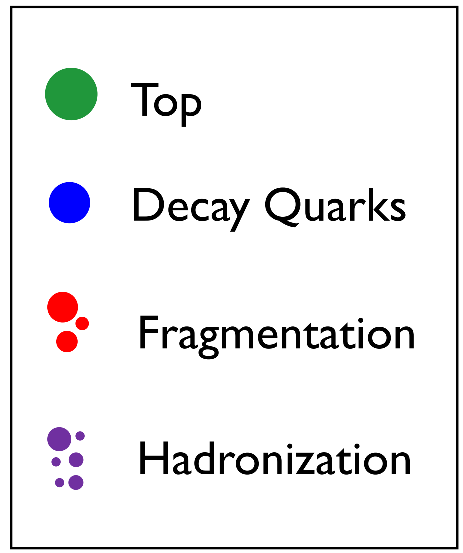


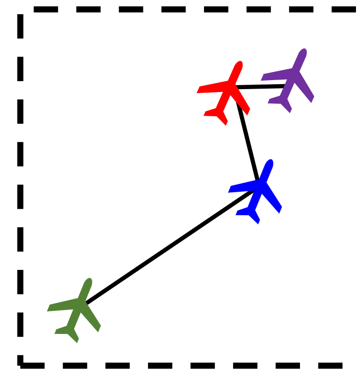




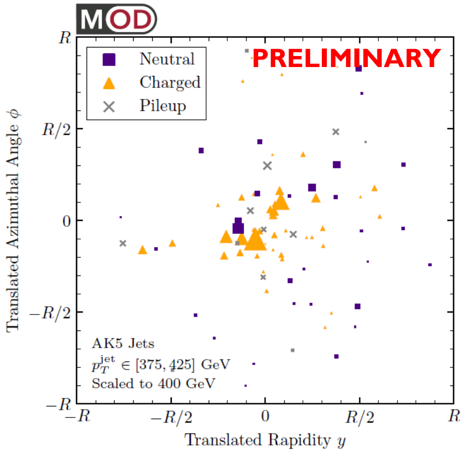
Movie Time: Visualizing Top Jet Formation

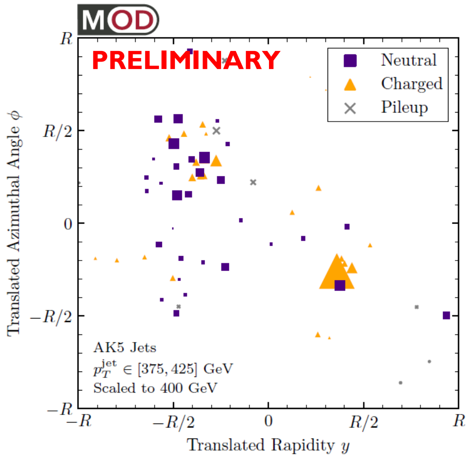


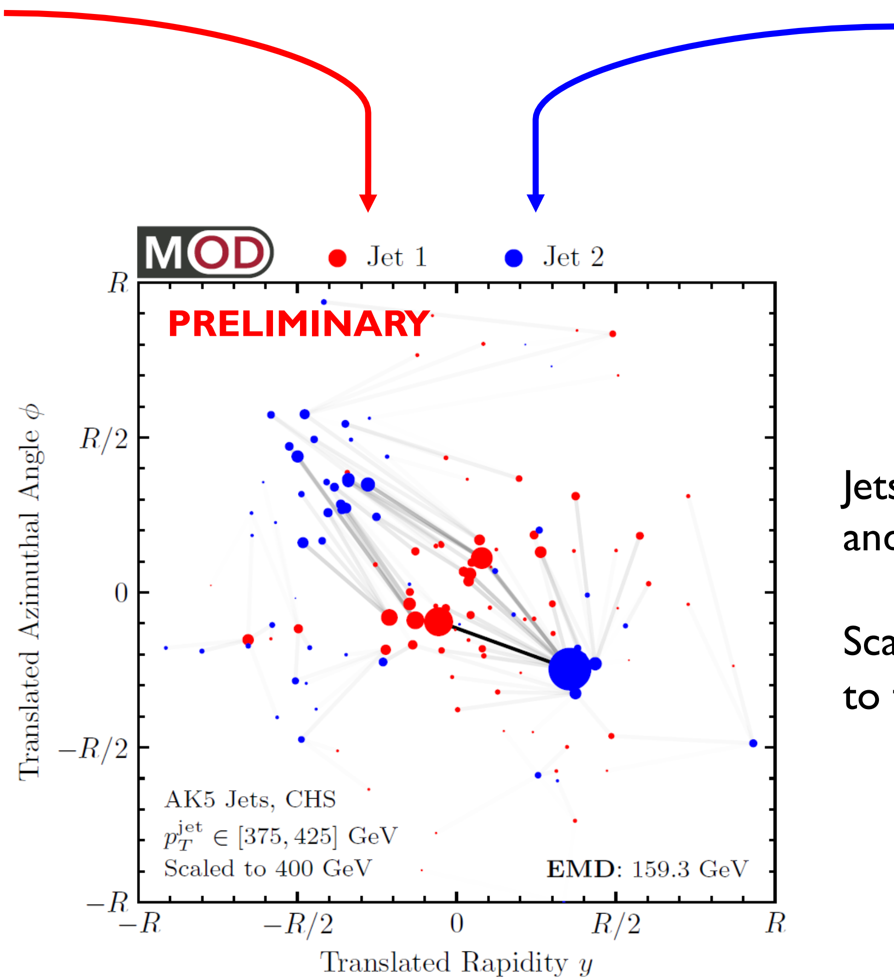




Comparing Jets in CMS Open Data



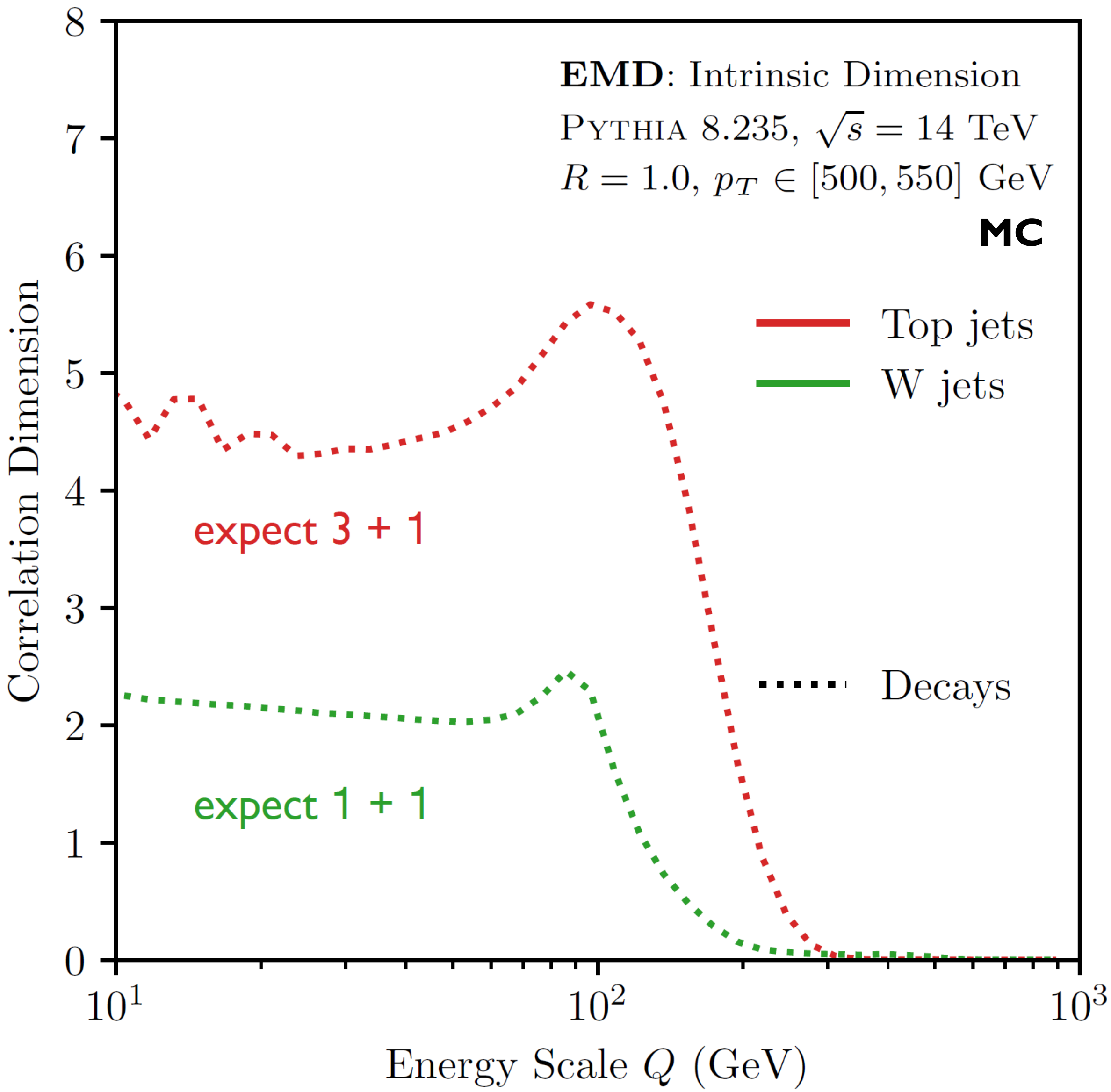


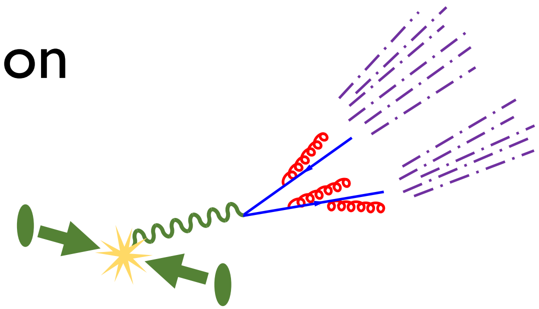


Jets are longitudinally boosted and rotated to $(y, \phi) = (0,0)$.

Scaling $\sum_{i=1}^{M} p_{T,i} \rightarrow 400$ GeV to focus on substructure.

Exploring the Space of Jets: Correlation Dimension





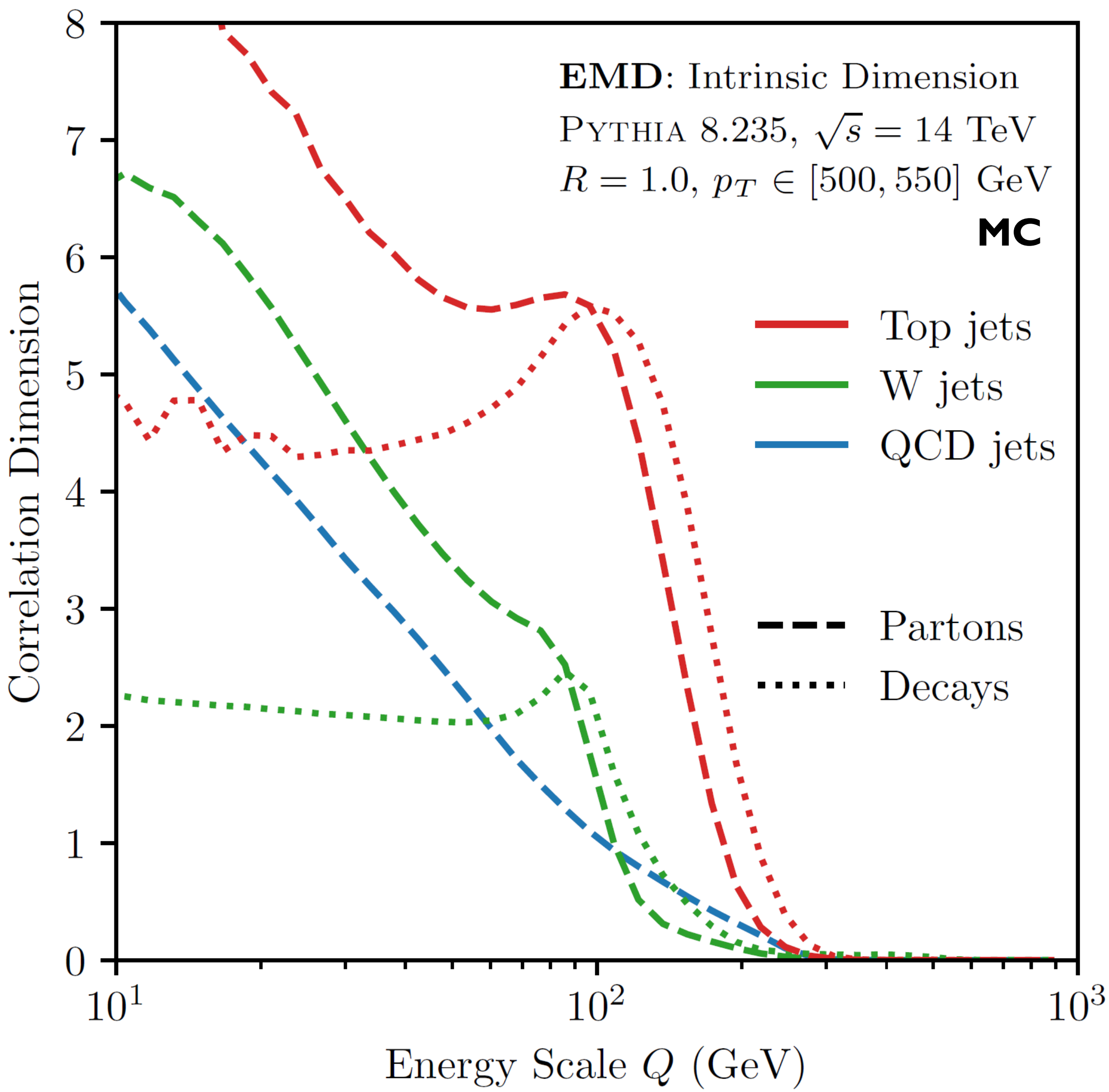
QCD jets are simplest.

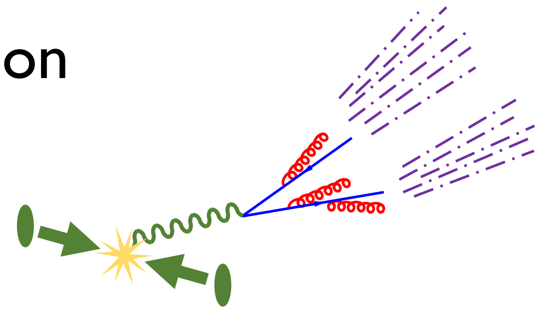
W jets are more complicated.

Top jets are most complex.

"Decays" have ~constant dimension.

Exploring the Space of Jets: Correlation Dimension





QCD jets are simplest.

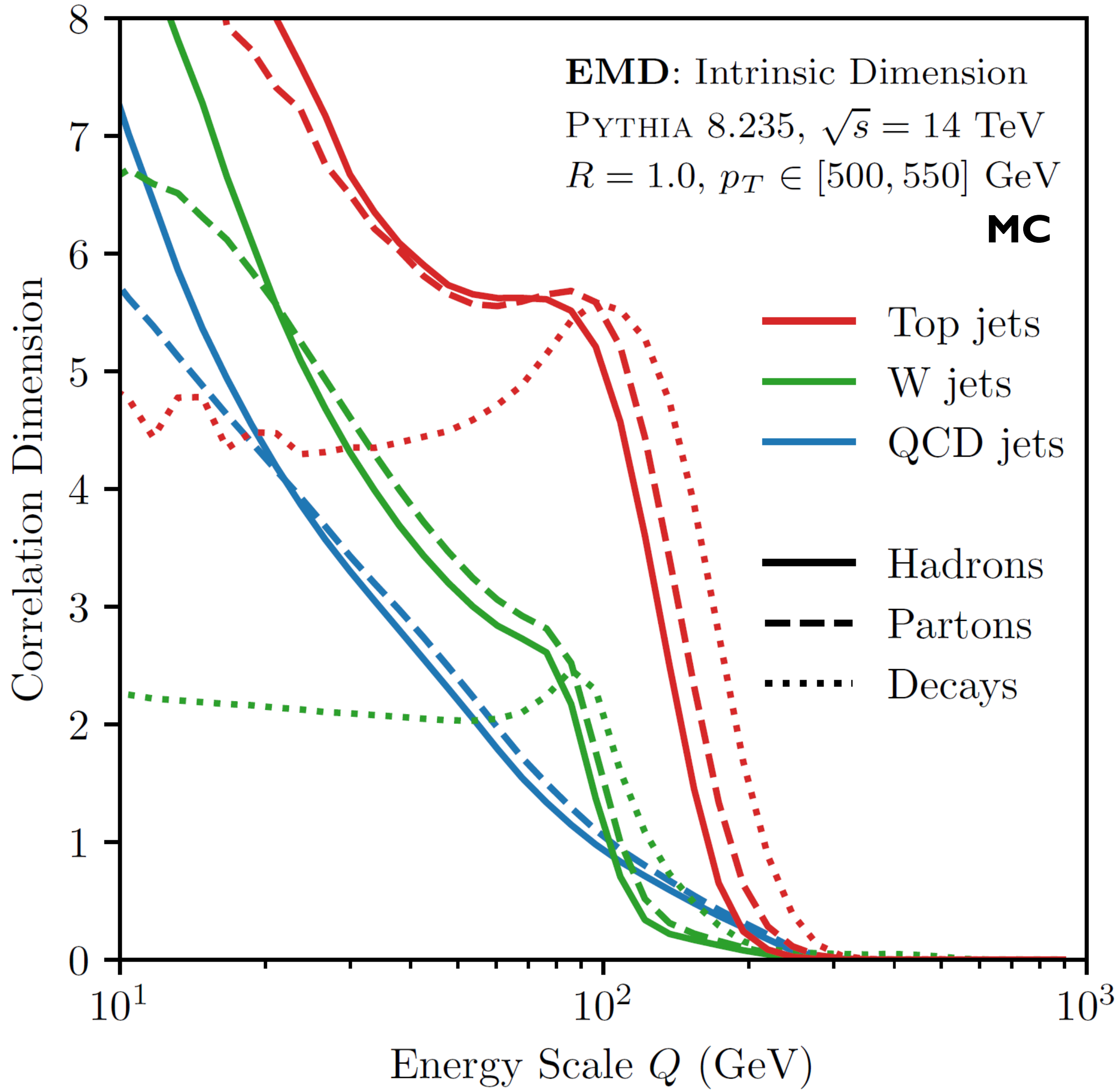
W jets are more complicated.

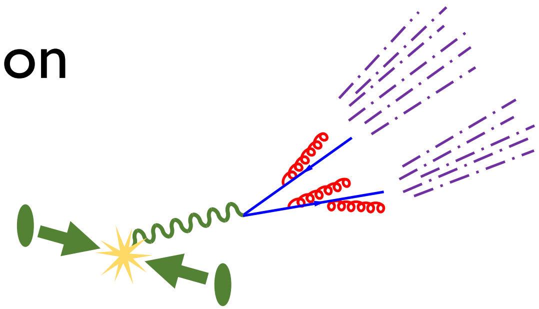
Top jets are most complex.

"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.

EMD and IRC-Safe Observables

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

Additive IRC-safe observables: $\mathcal{O}(\mathcal{E}) = \sum_{i=1}^{M} E_i \, \Phi(\hat{n}_i)$

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

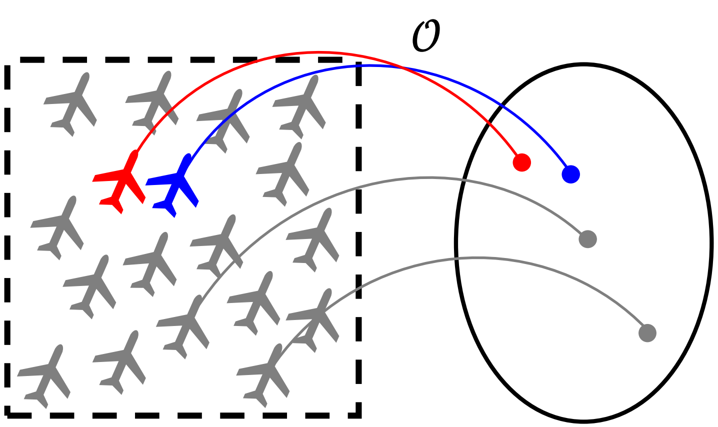
"Lipschitz constant" of Φ
i.e. bound on its derivative

e.g. jet angularities:
[C. Berger, T. Kucs, and G. Sterman, 0303051]
[A. Larkoski, J. Thaler, and W. Waalewijn, 1408.3122]

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

For $\beta \geq 1$ jet angularities:

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

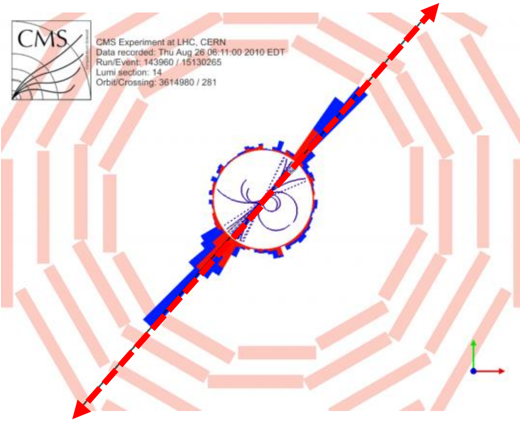


Old Observables in a New Language

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

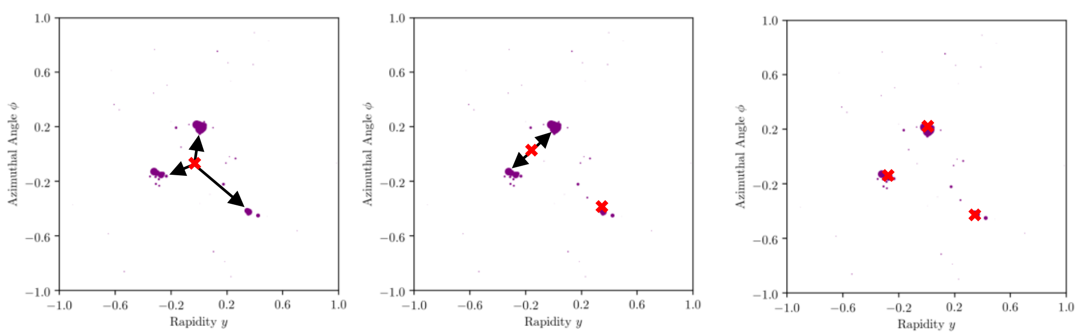
$$t(\mathcal{E}) = E - \max_{\hat{n}} \sum_i |\vec{p}_i \cdot \hat{n}| \longrightarrow t(\mathcal{E}) = \min_{|\mathcal{E}'|=2} \text{EMD}(\mathcal{E}, \mathcal{E}')$$

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

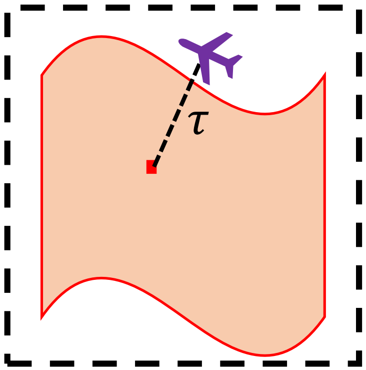


***N*-(sub)jettiness** is the EMD between the event and the closest N-particle event.

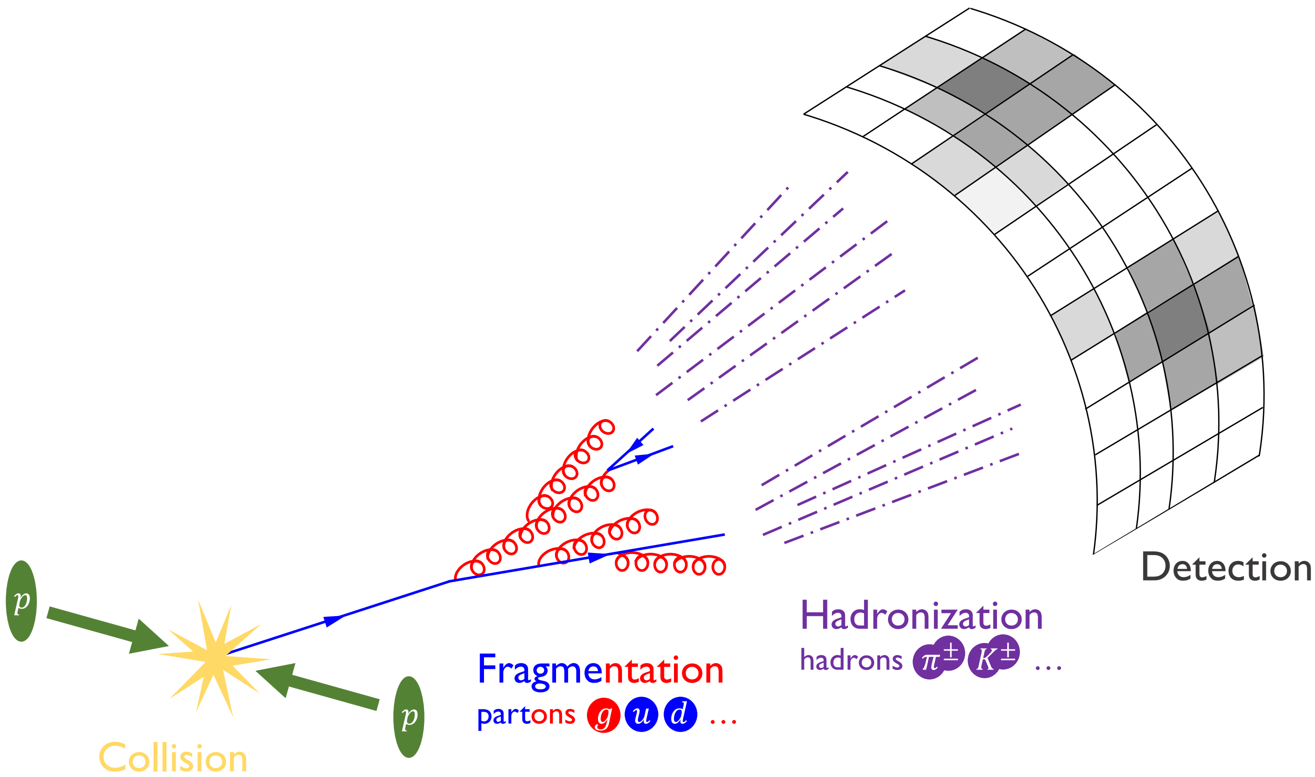
$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\text{ axes}} \sum_{i=1}^{M} E_i \min_k \{\theta_{1,k}^\beta, \theta_{2,k}^\beta, \ldots, \theta_{N,k}^\beta\} \longrightarrow \tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \text{EMD}(\mathcal{E}, \mathcal{E}').$$



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

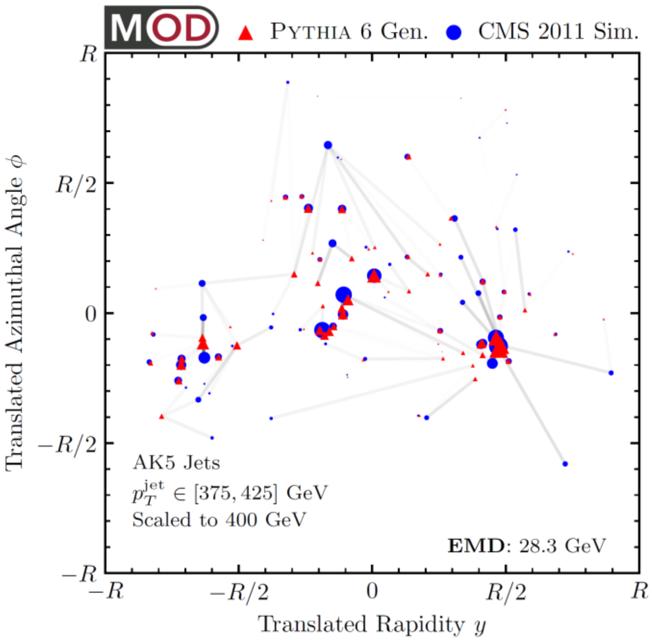


Quantifying Detector Effects

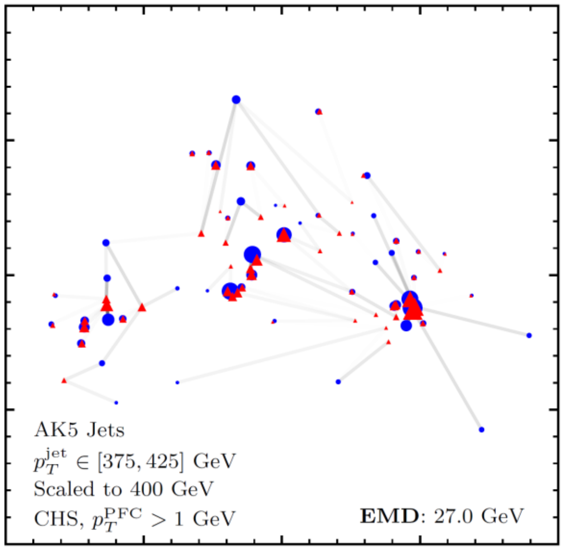


Quantifying Detector Effects with EMD

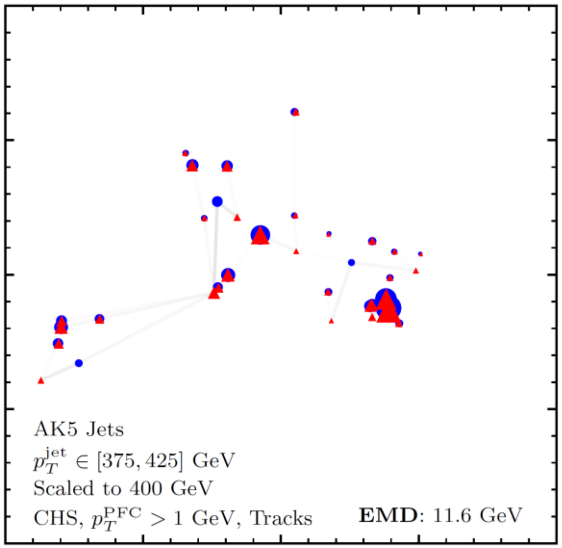
Gen./Sim. EMD: 28.3 GeV



Gen./Sim. EMD: 27.0 GeV



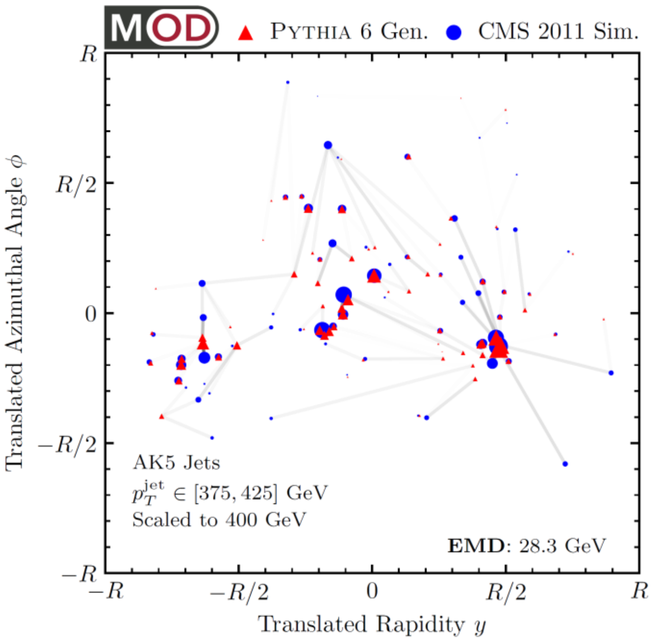
Gen./Sim. EMD: 11.6 GeV

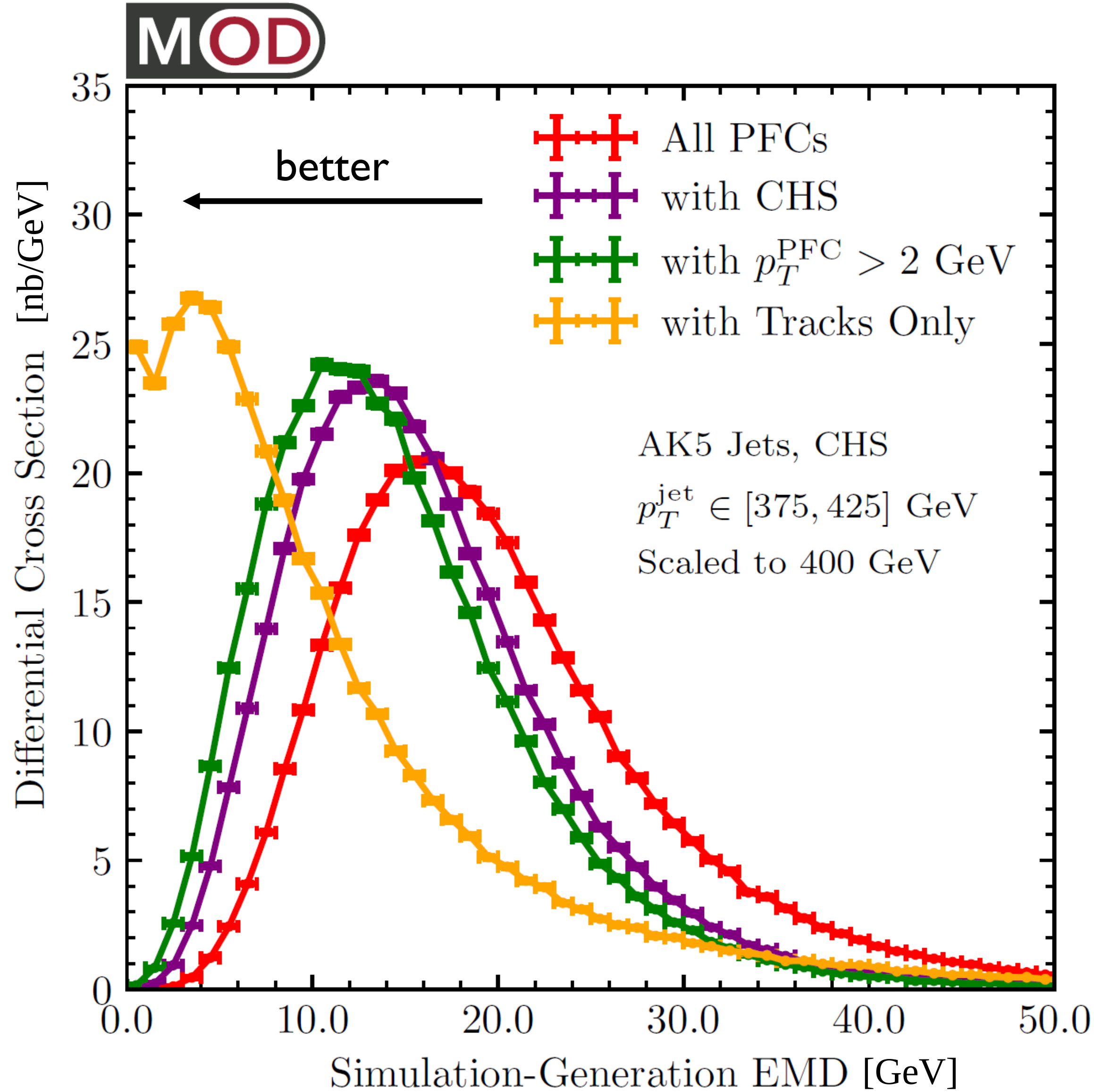


+ charged hadron subtraction
+ $p_T^{\rm PFC} > 1$ GeV cut

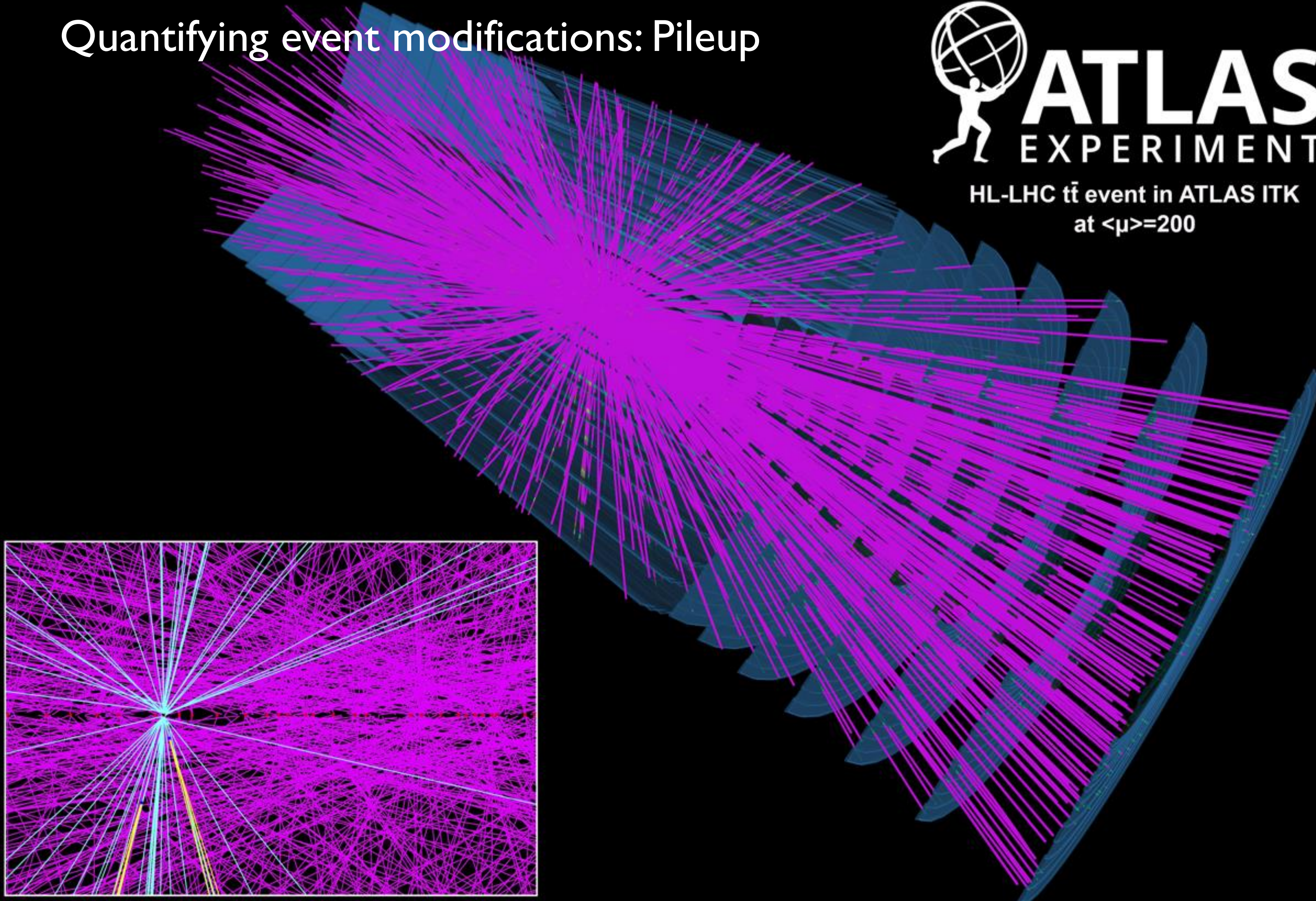
Tracks only

Quantifying Detector Effects with EMD





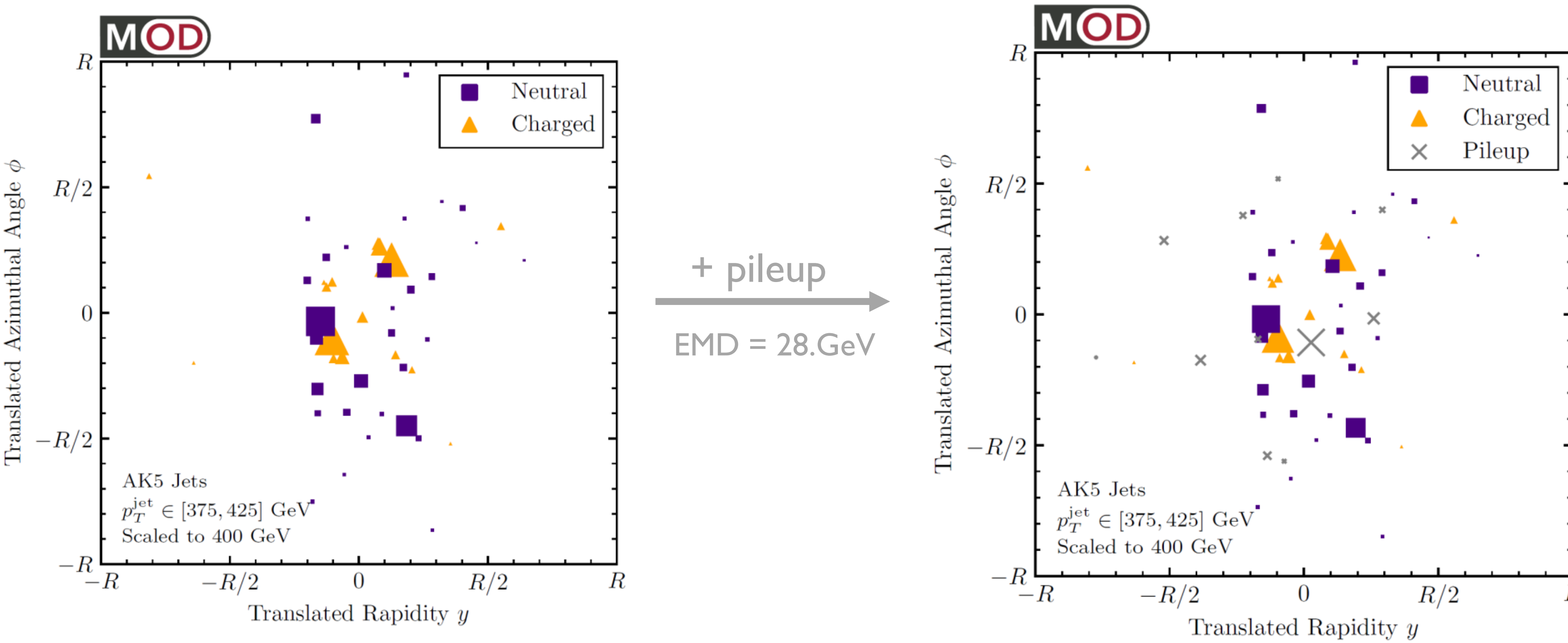
Quantifying event modifications: Pileup



Pileup Mitigation

Can use vertex information in CMS Open Data to find charged pileup particles.

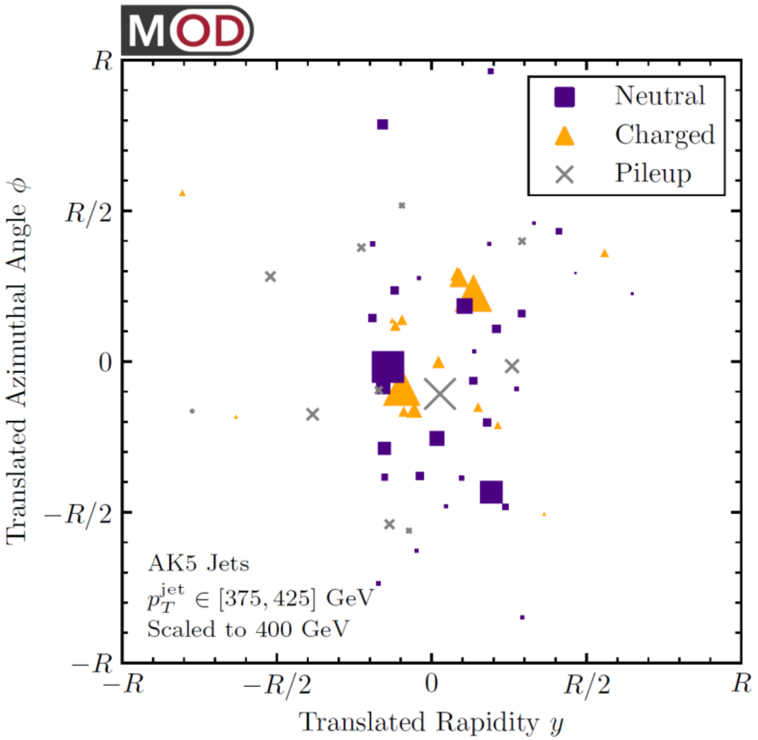
Allows us to study the effect of pileup on radiation pattern and observables.

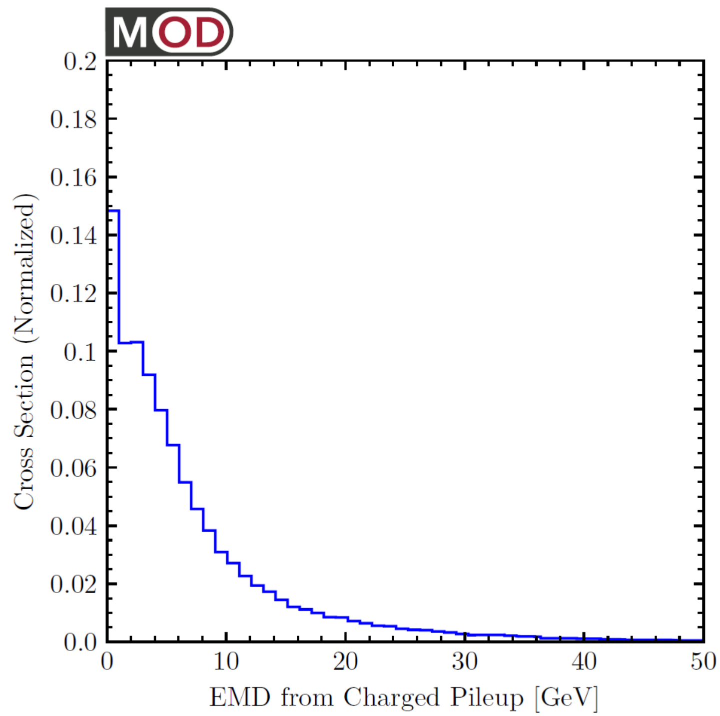


Pileup Mitigation

Can use vertex information in CMS Open Data to find charged pileup particles.

Allows us to study the effect of pileup on radiation pattern and observables.





Quantify pileup mitigation performance with EMD (more in backup).

Optimize machine learning-based pileup mitigation methods to minimize EMD?

[Komiske, EMM, Nachman, Schwartz, 1707.08600]

[Martinez, Cerri, Pierini, Spiropulu, Vlimant, 1810.07988]

EnergyFlow

https://energyflow.network

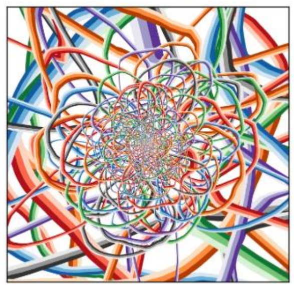
pip install energyflow

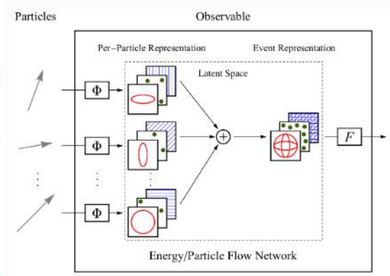
EnergyFlow

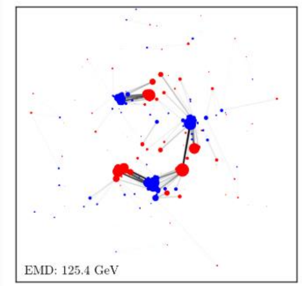
Search docs

Docs » Home

Welcome to EnergyFlow



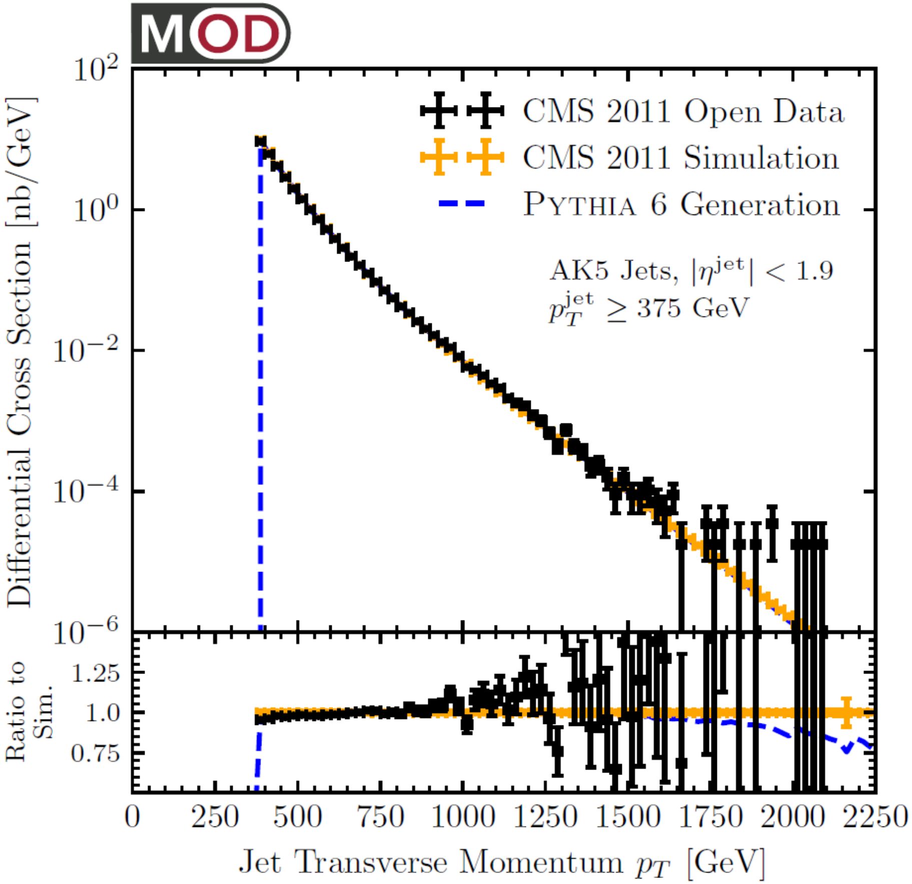


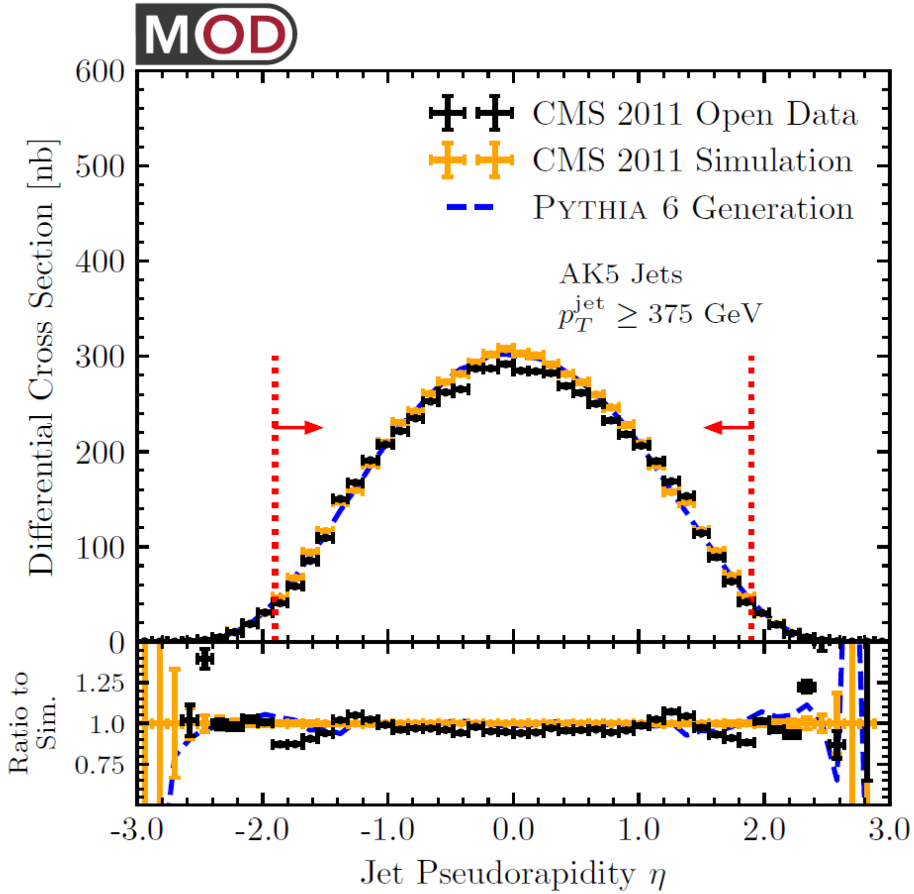


EnergyFlow is a Python package containing a suite of particle physics tools. Originally designed to compute Energy Flow Polynomials (EFPs), as of version `0.10.0` the package expanded to include implementations of Energy Flow Networks (EFNs) and Particle Flow Networks (PFNs). As of version `0.11.0`, 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

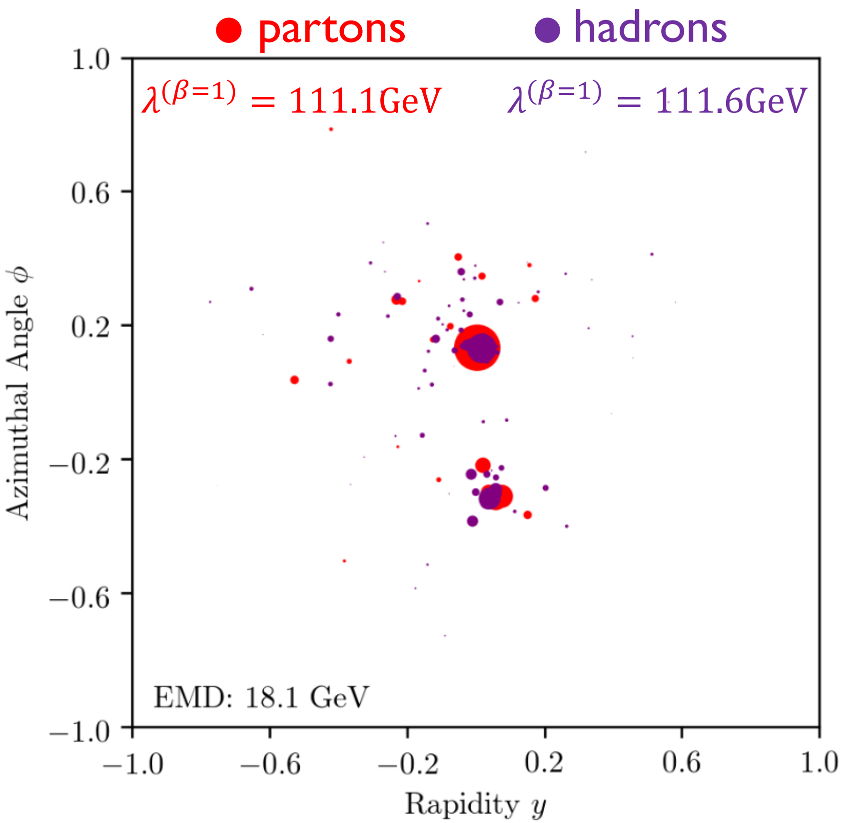
Jet Kinematic Distributions

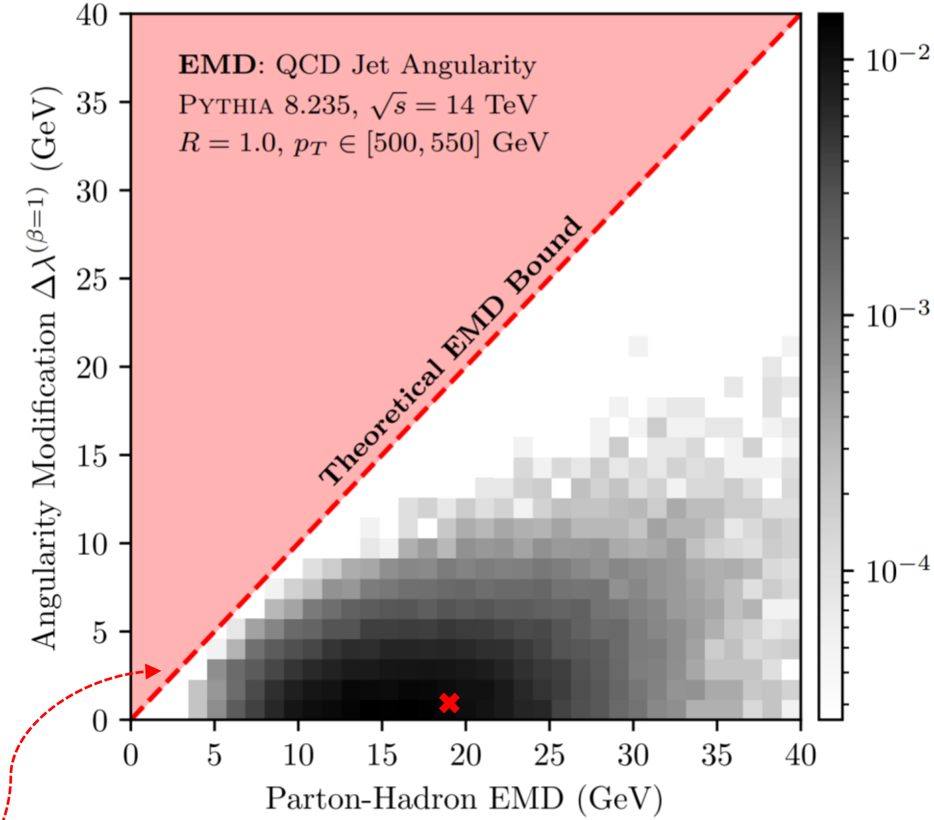




Quantifying event modifications: Hadronization

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



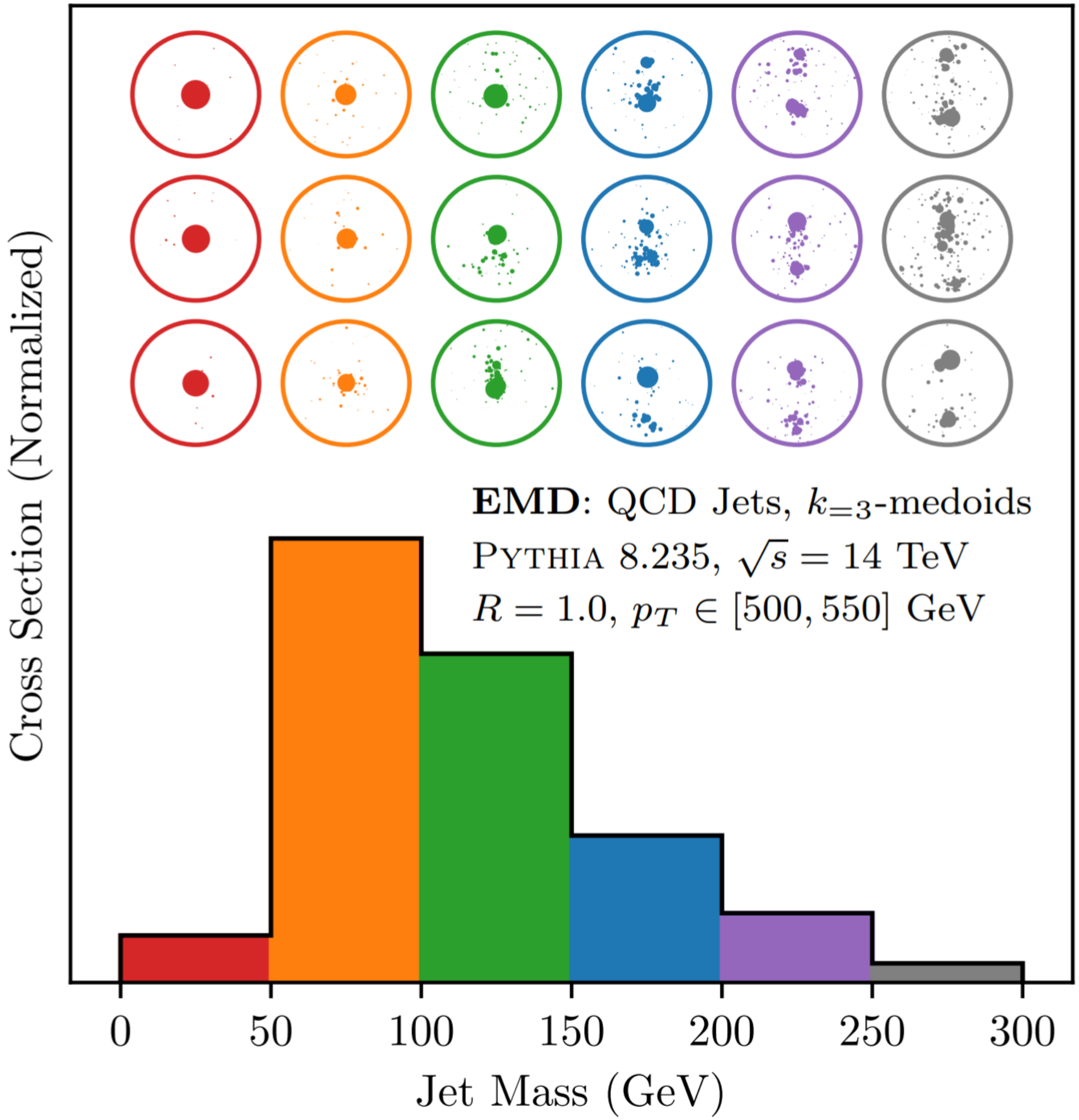


$\mathcal{E} = \mathcal{E}_{\text{partons}}$

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

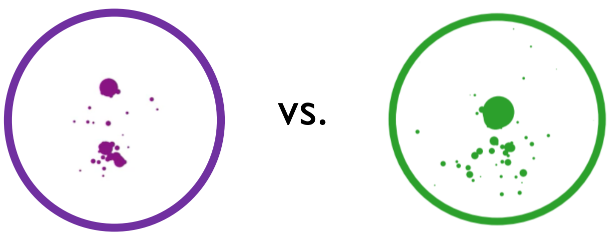
$$\boxed{\left|\lambda^{(\beta=1)}(\mathcal{E}) - \lambda^{(\beta=1)}(\mathcal{E}')\right| \leq \mathrm{EMD}(\mathcal{E}, \mathcal{E}')}$$

Exploring the Space of Events: *k*-medoids



Exploring the Space of Events: Jet Classification

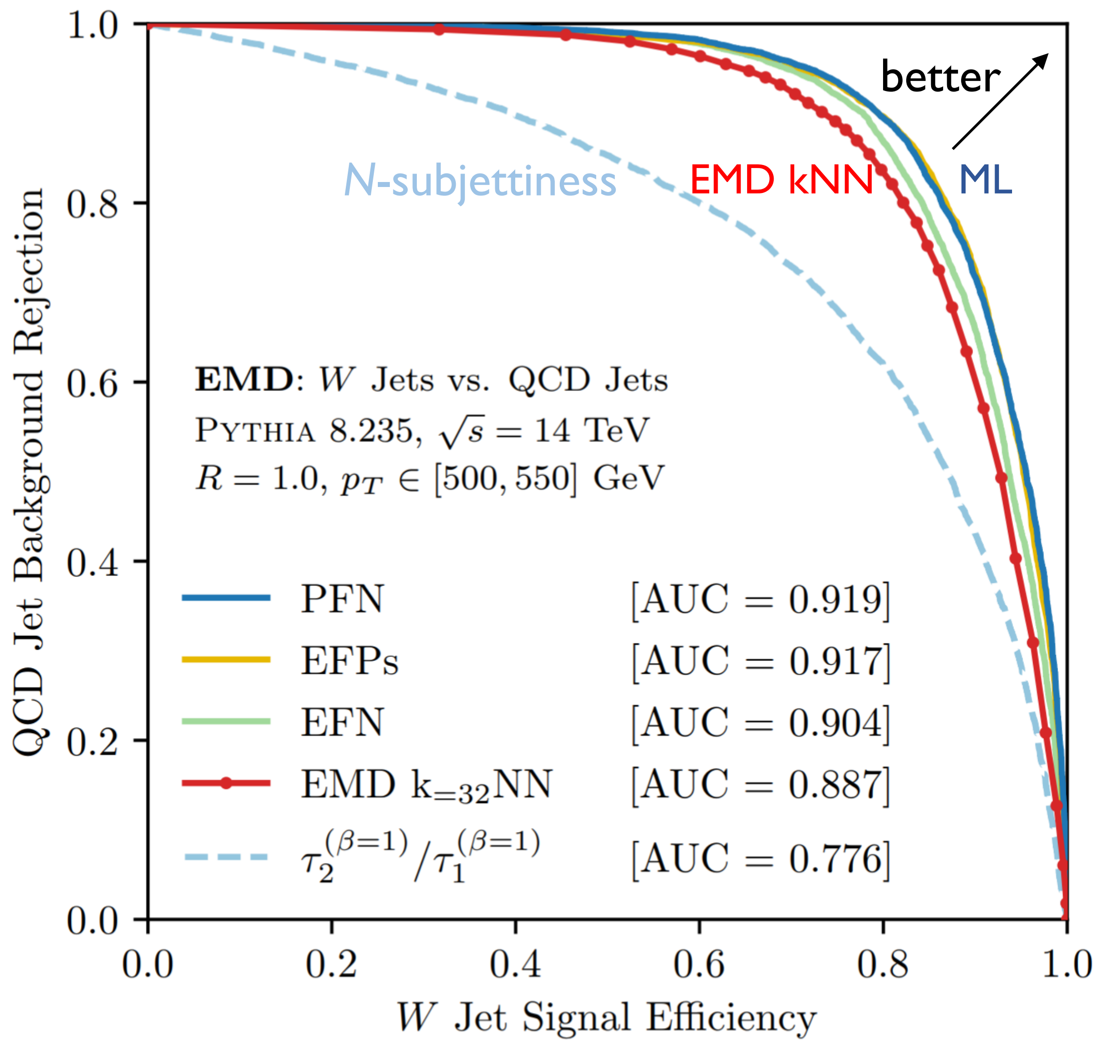
Classify *W* jets vs. QCD jets



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

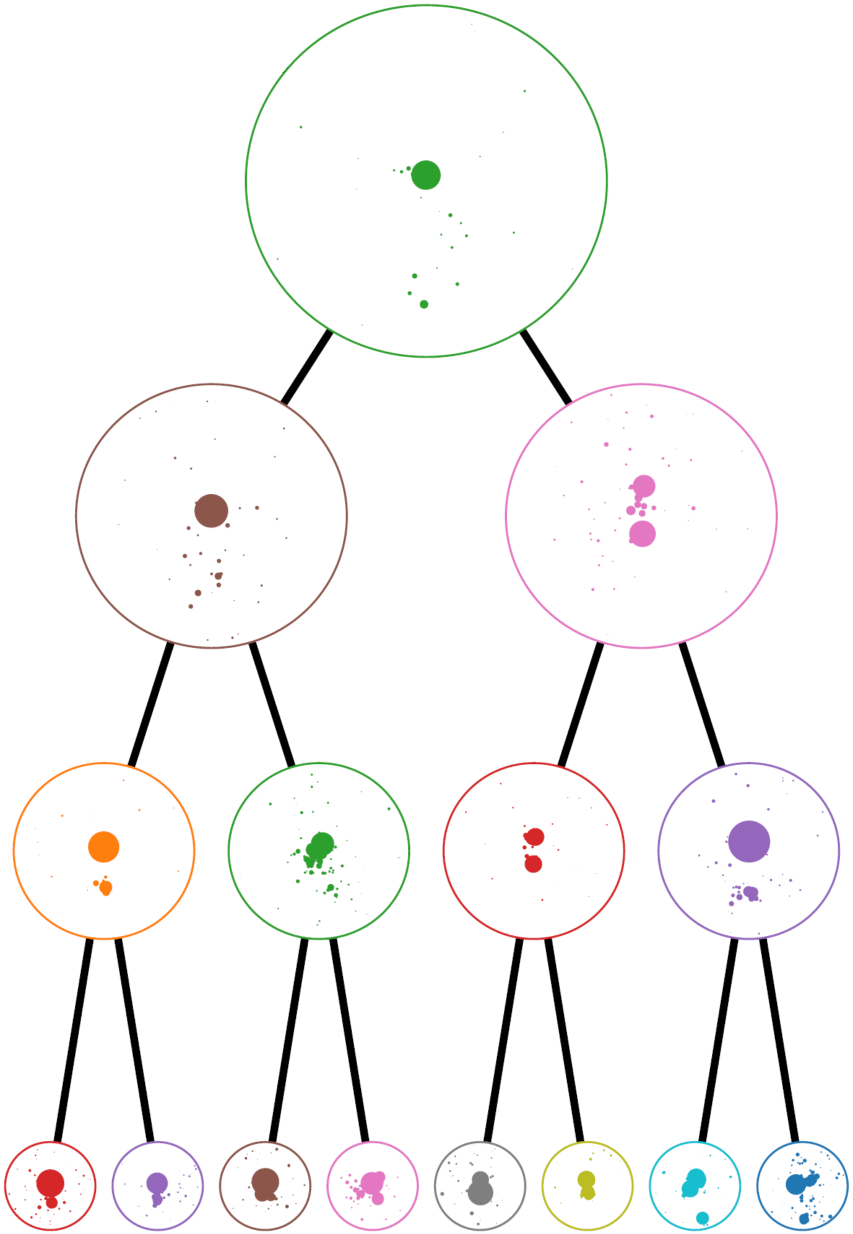
Unsupervised clustering algorithms can be used to cluster events

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

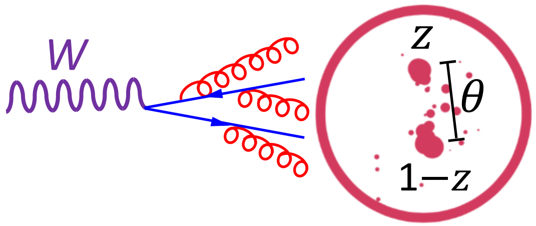
VP Tree: O(log(N)) neighbor query time

Much more to explore.

Vantage Point (VP) Tree



Exploring the Space of Events: *W* jets



W jets are 2-pronged:

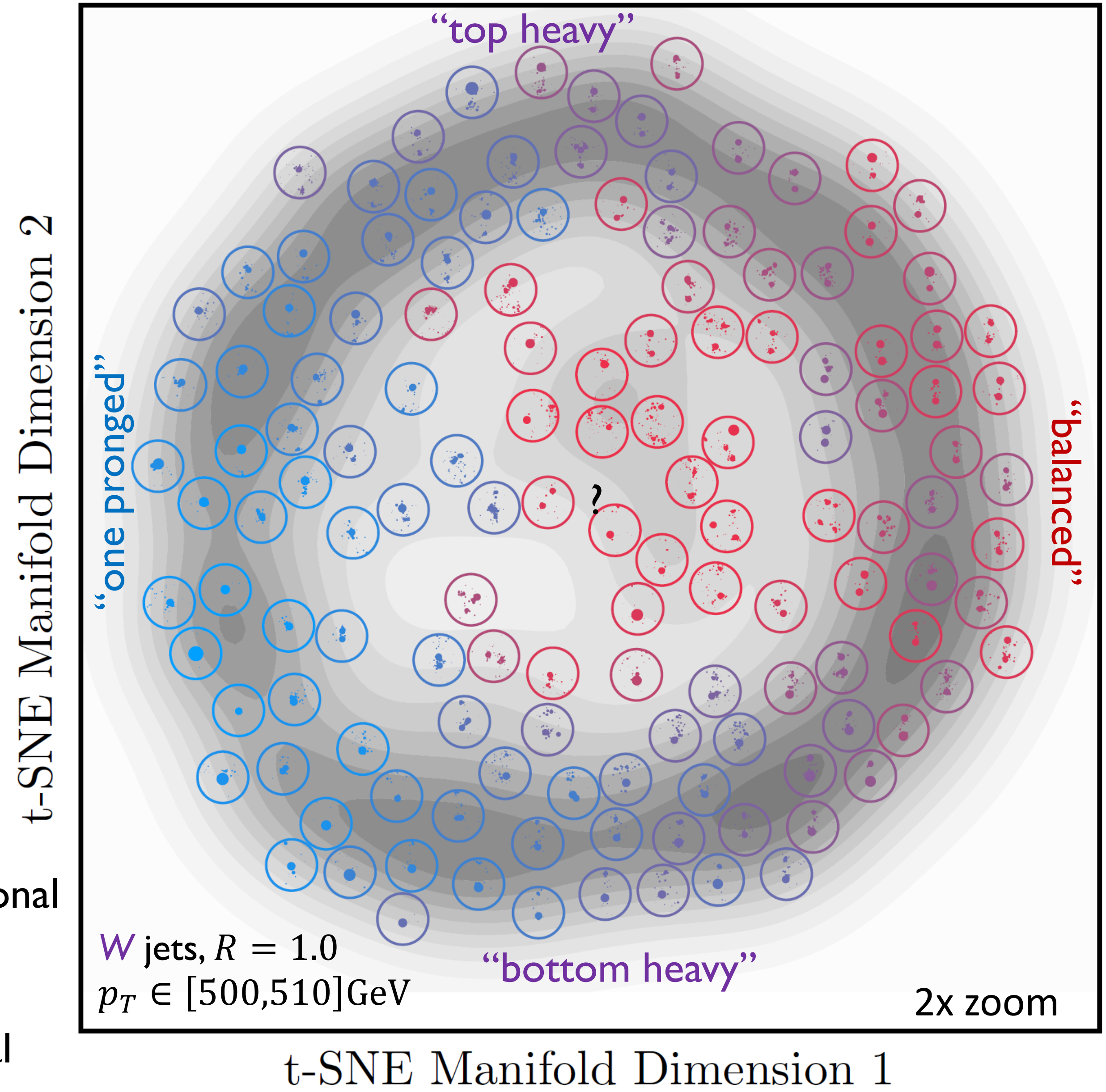
z: Energy Sharing of Prongs
θ: Angle between Prongs
φ: Azimuthal orientation

Constrained by W mass:

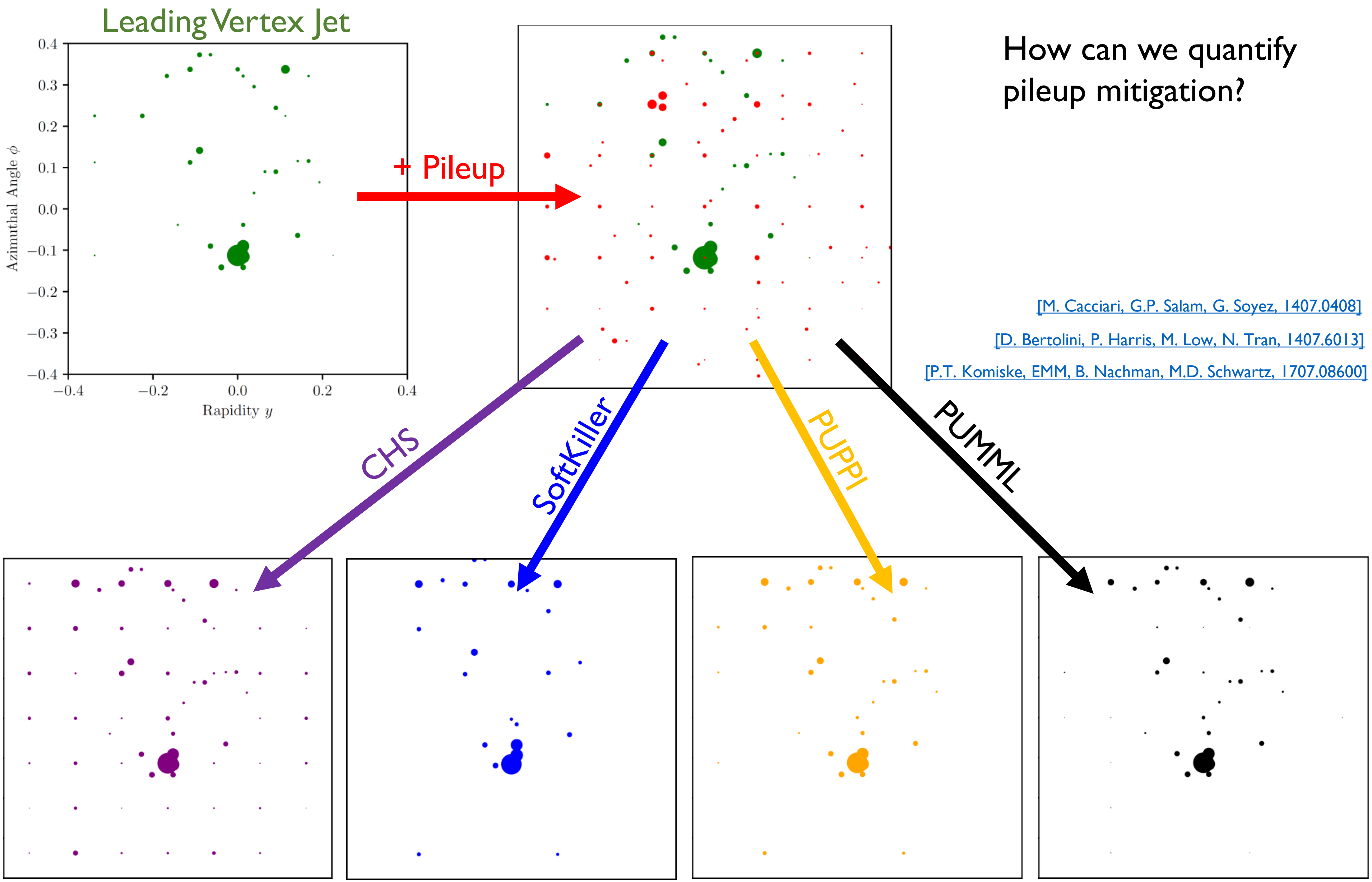
$$z(1-z)\theta^2 = \frac{p_{\mu J}^2}{p_T^2} = \frac{m_W^2}{p_T^2}$$

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

After φ rotation: **one**-dimensional



Quantifying event modifications: Pileup



How can we quantify pileup mitigation?

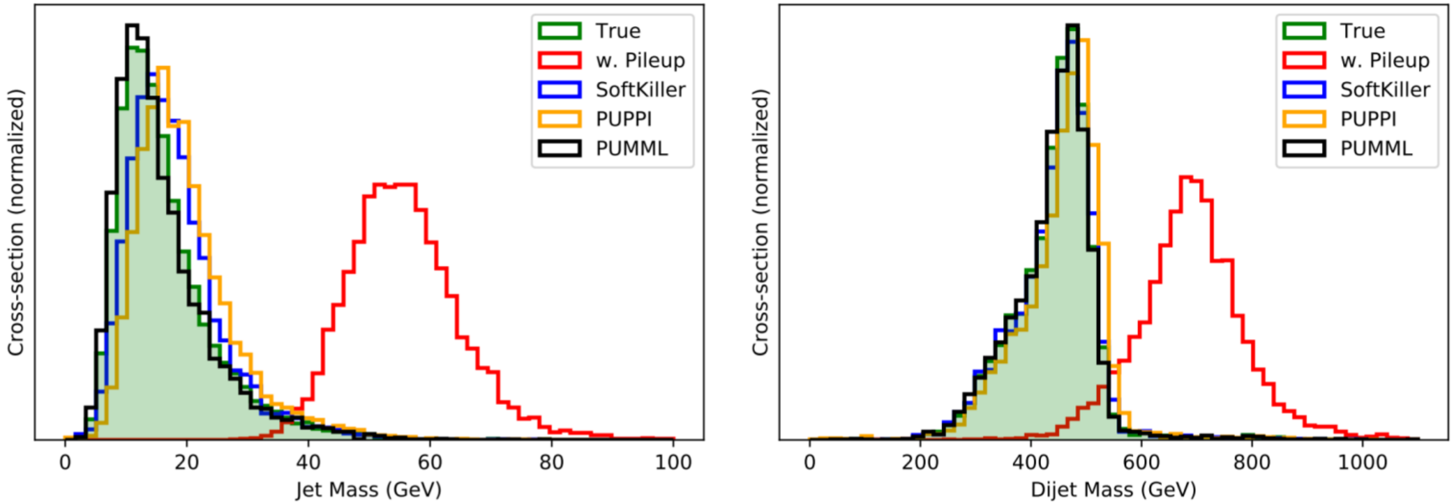
[M. Cacciari, G.P. Salam, G. Soyez, 1407.0408]

[D. Bertolini, P. Harris, M. Low, N. Tran, 1407.6013]

[P.T. Komiske, EMM, B. Nachman, M.D. Schwartz, 1707.08600]

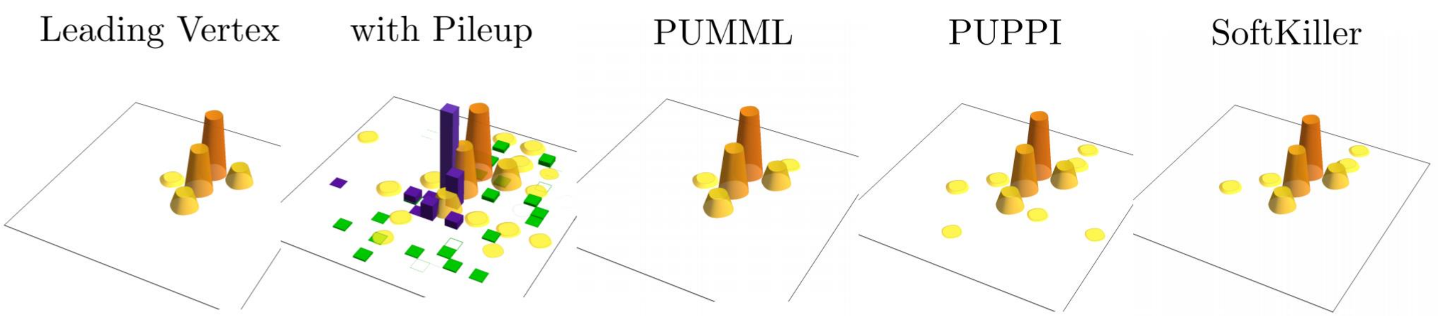
Quantifying event modifications: Pileup

Compare on a collection of observables?



Requires ad hoc choices of observables.

Compare calorimeter images pixel by pixel?

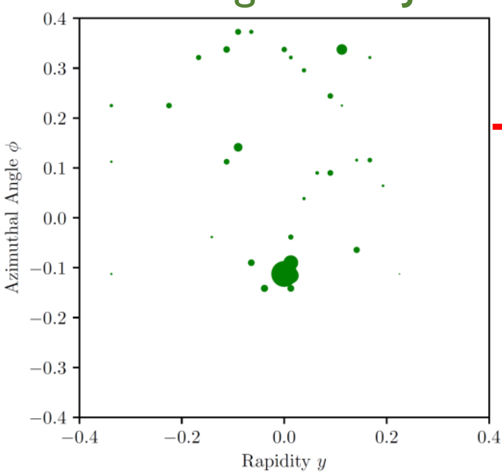


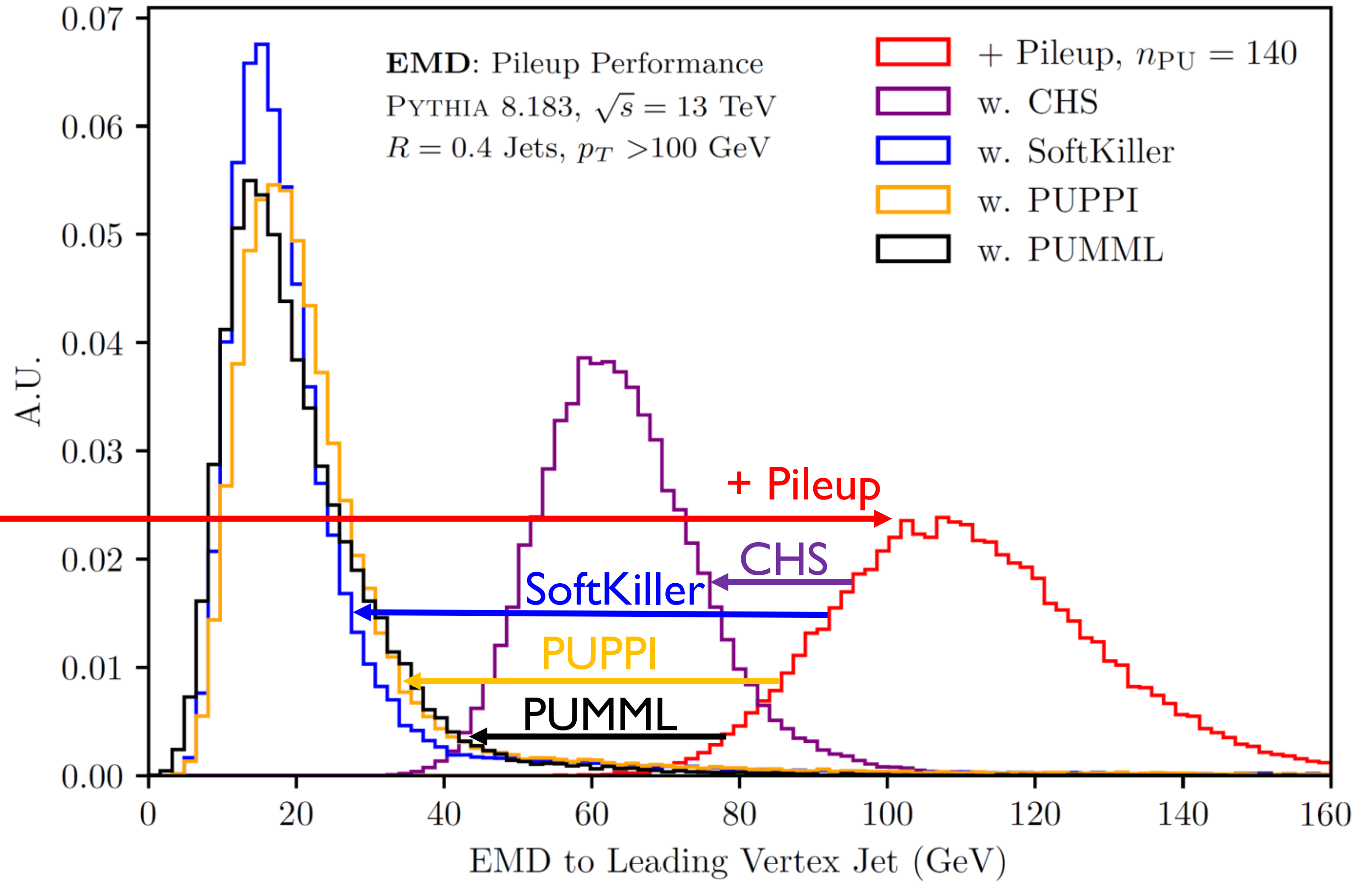
Discontinuous under physically-sensible single-pixel perturbations.
Undesirable behavior with increasing resolution.

Quantifying event modifications: Pileup

Measure pileup mitigation performance with EMD!

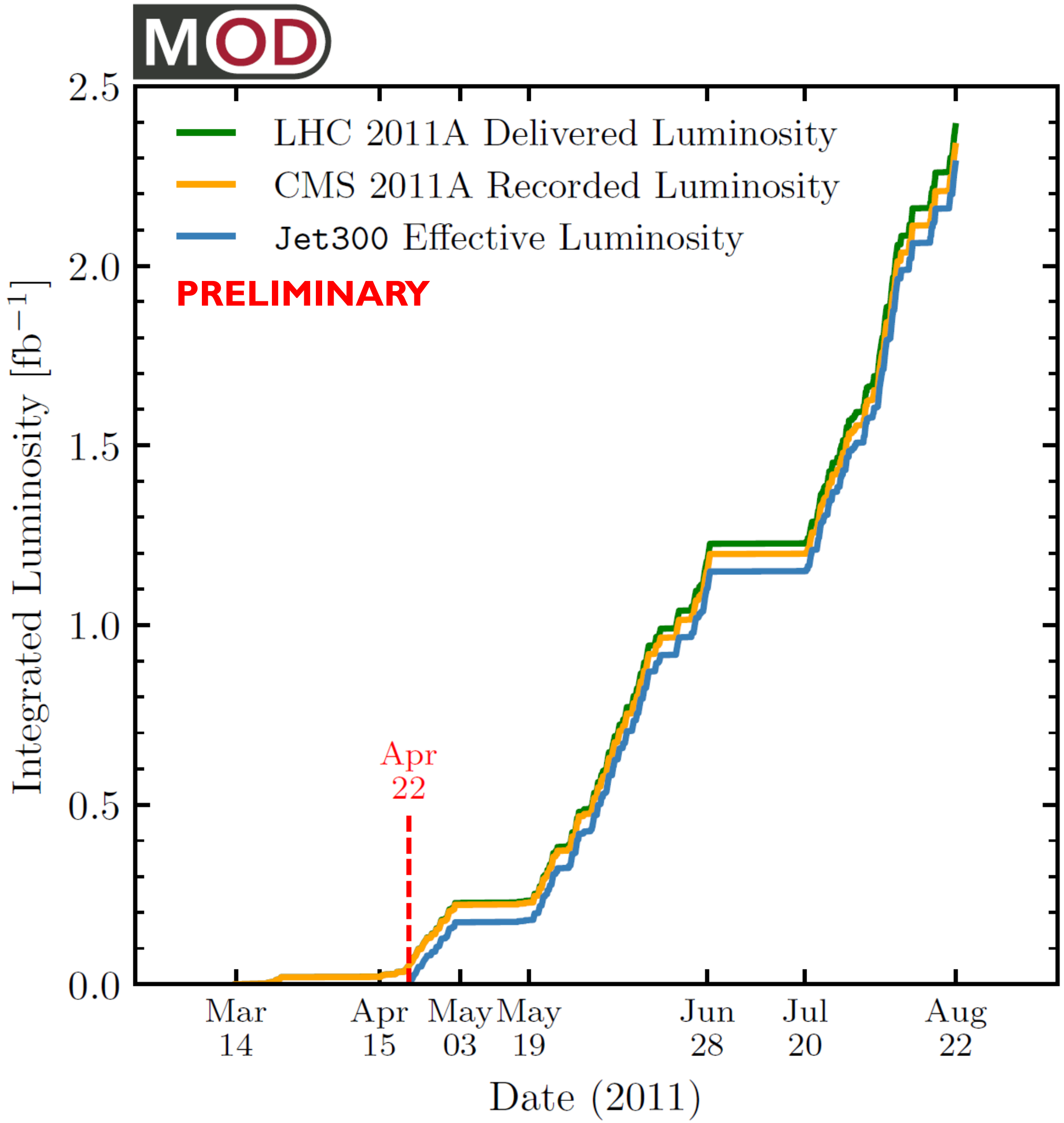
Leading Vertex Jet





Guarantees performance on IRC safe observables.
Stable under physically-sensible perturbations.
Train to optimize EMD with machine learning?

Integrated Luminosity



Exploring the Space of Jets: Correlation Dimension

Sketch of leading log (one emission) calculation:

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

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

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

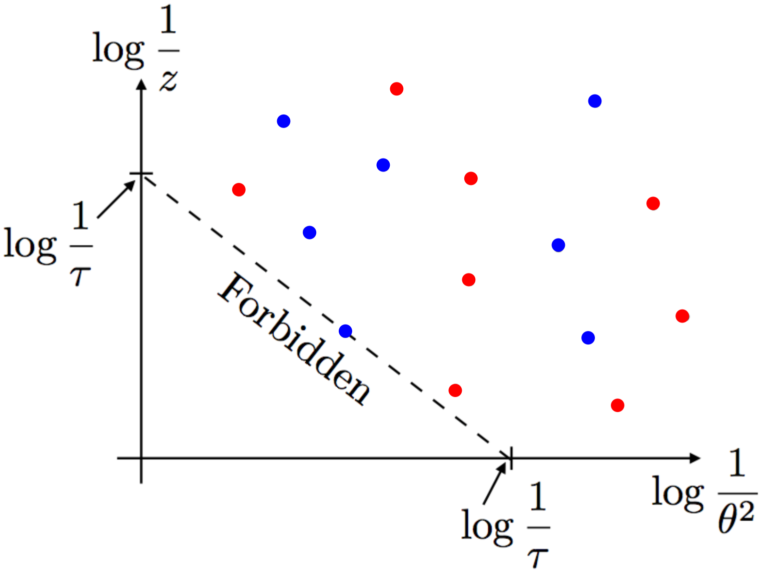
$$= Q\frac{\partial}{\partial Q}\ln\exp\left(-\frac{4\alpha_s C_{q/g}}{\pi}\ln^2\frac{Q}{p_T/2}\right)$$

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

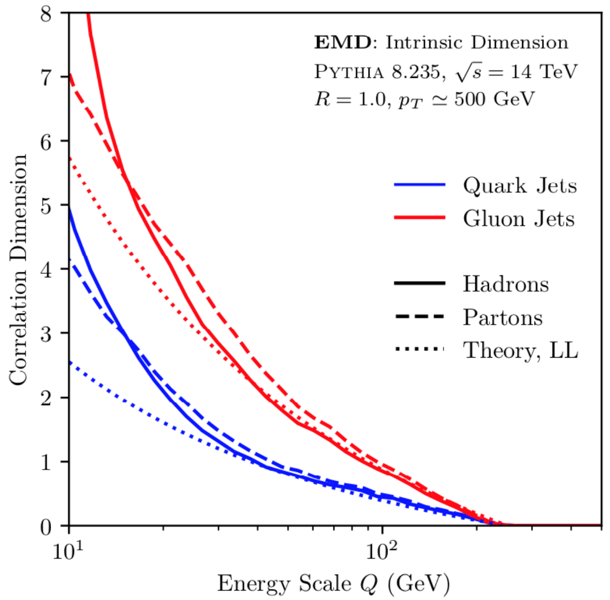
+ 1-loop running of α_s

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

$$C_g = C_A = 3$$

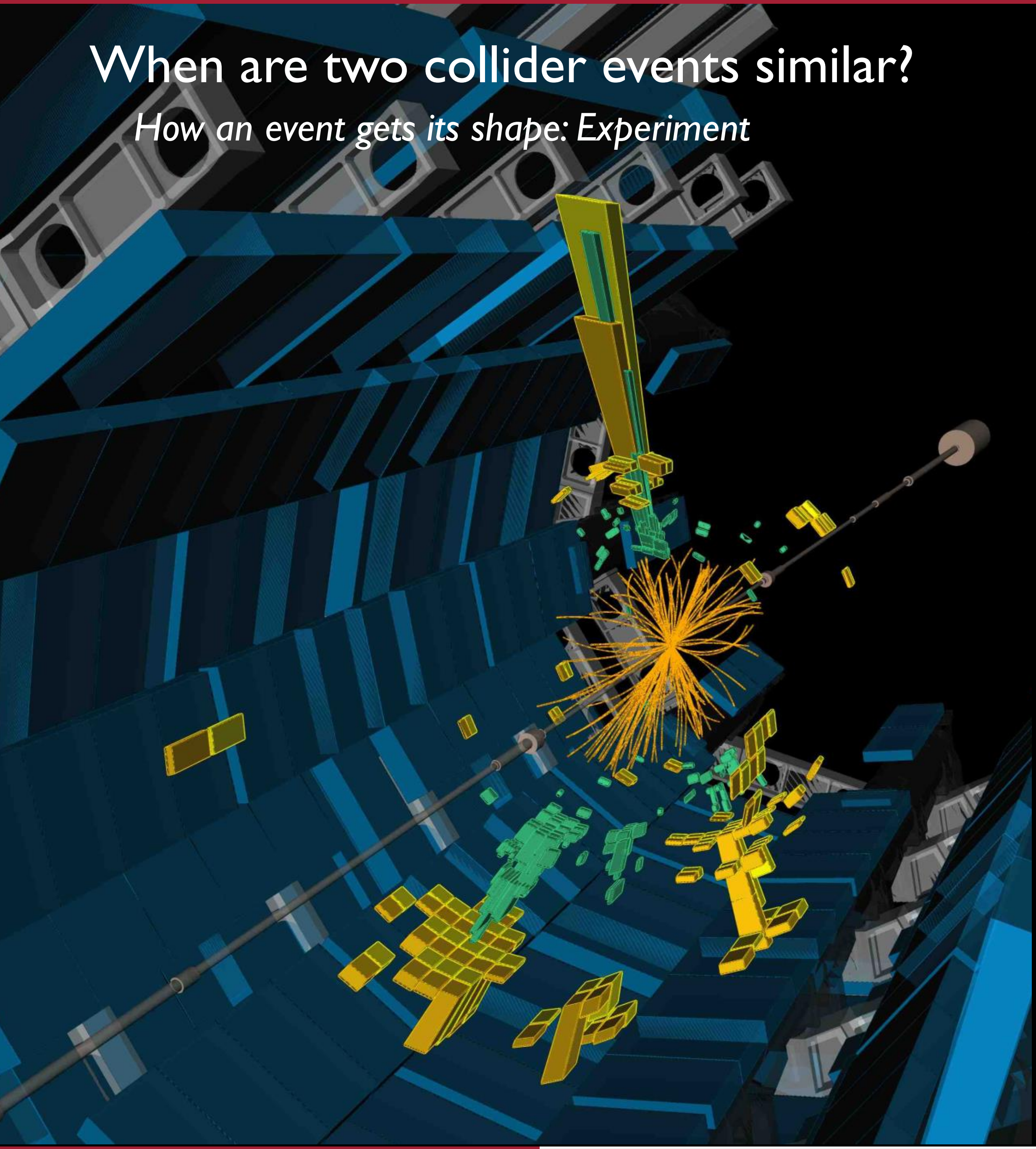


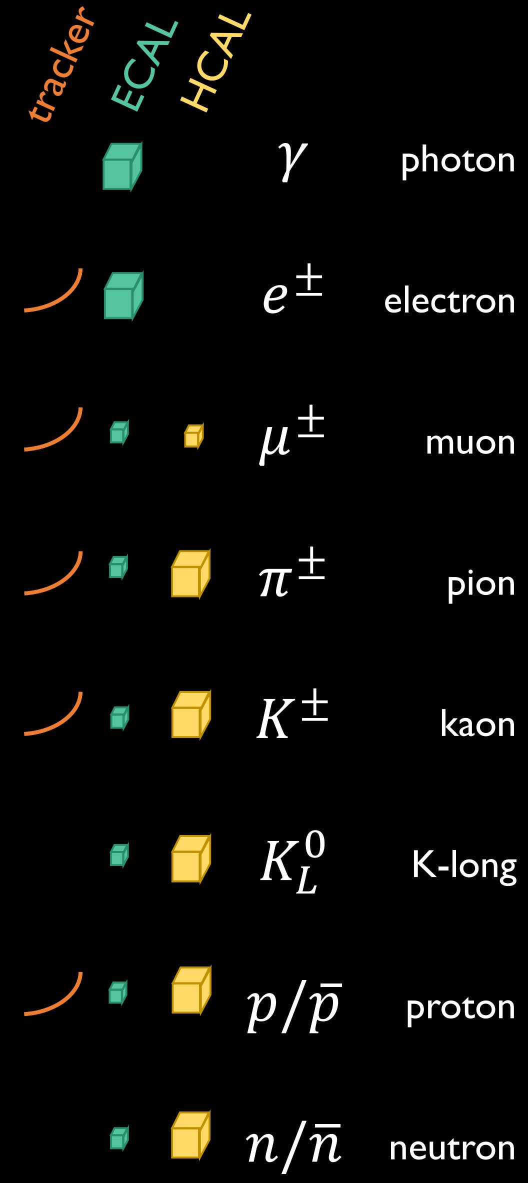
[A. Larkoski, 1709.06195]



When are two collider events similar?

How an event gets its shape: Experiment





Pileup Mitigation with PUMML