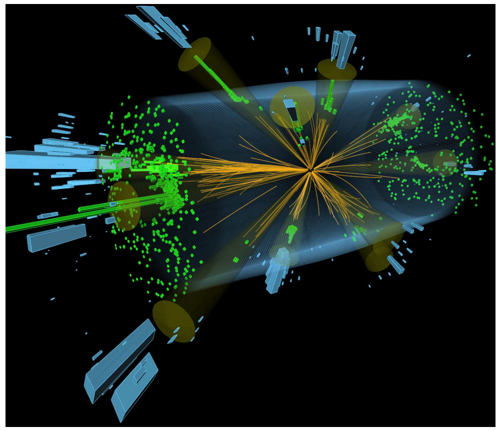
The Metric Space of Collider Events

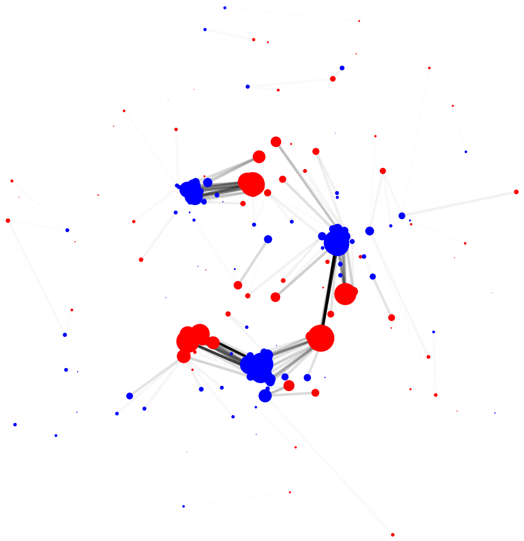
Patrick T. Komiske III
Massachusetts Institute of Technology
Center for Theoretical Physics

with Eric Metodiev and Jesse Thaler, 1902.02346, to appear in PRL

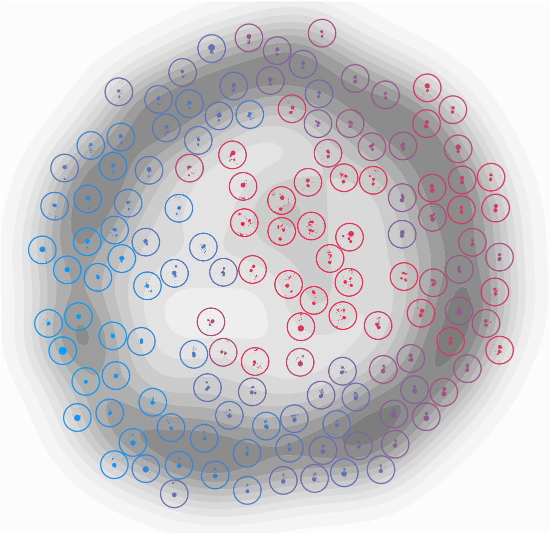
Particle Physics Seminar – University of Chicago
May 29, 2019



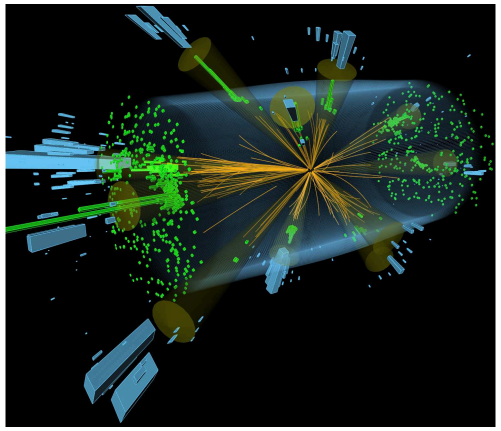
When are two events similar?

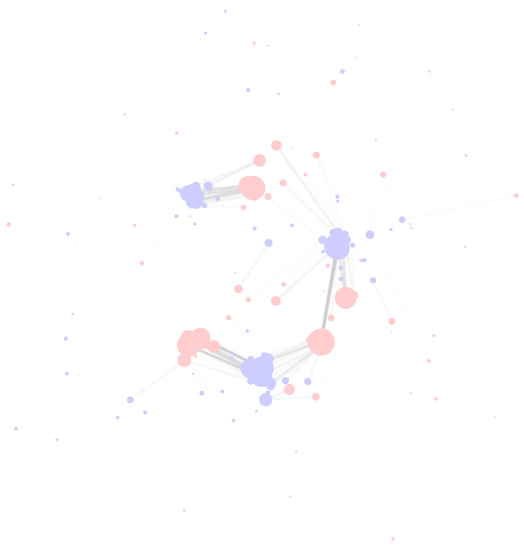


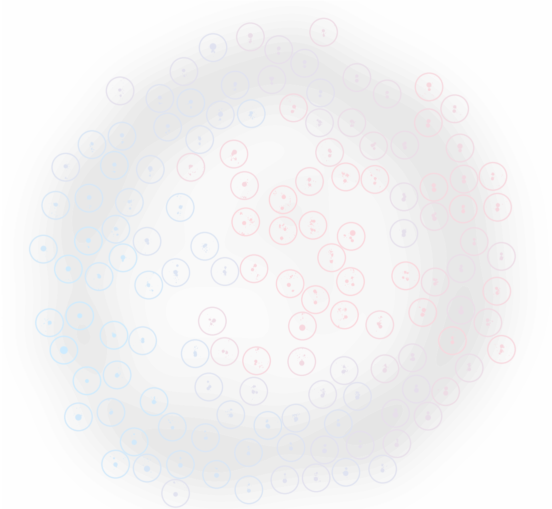
The Energy Mover’s Distance



Particle Physics Applications







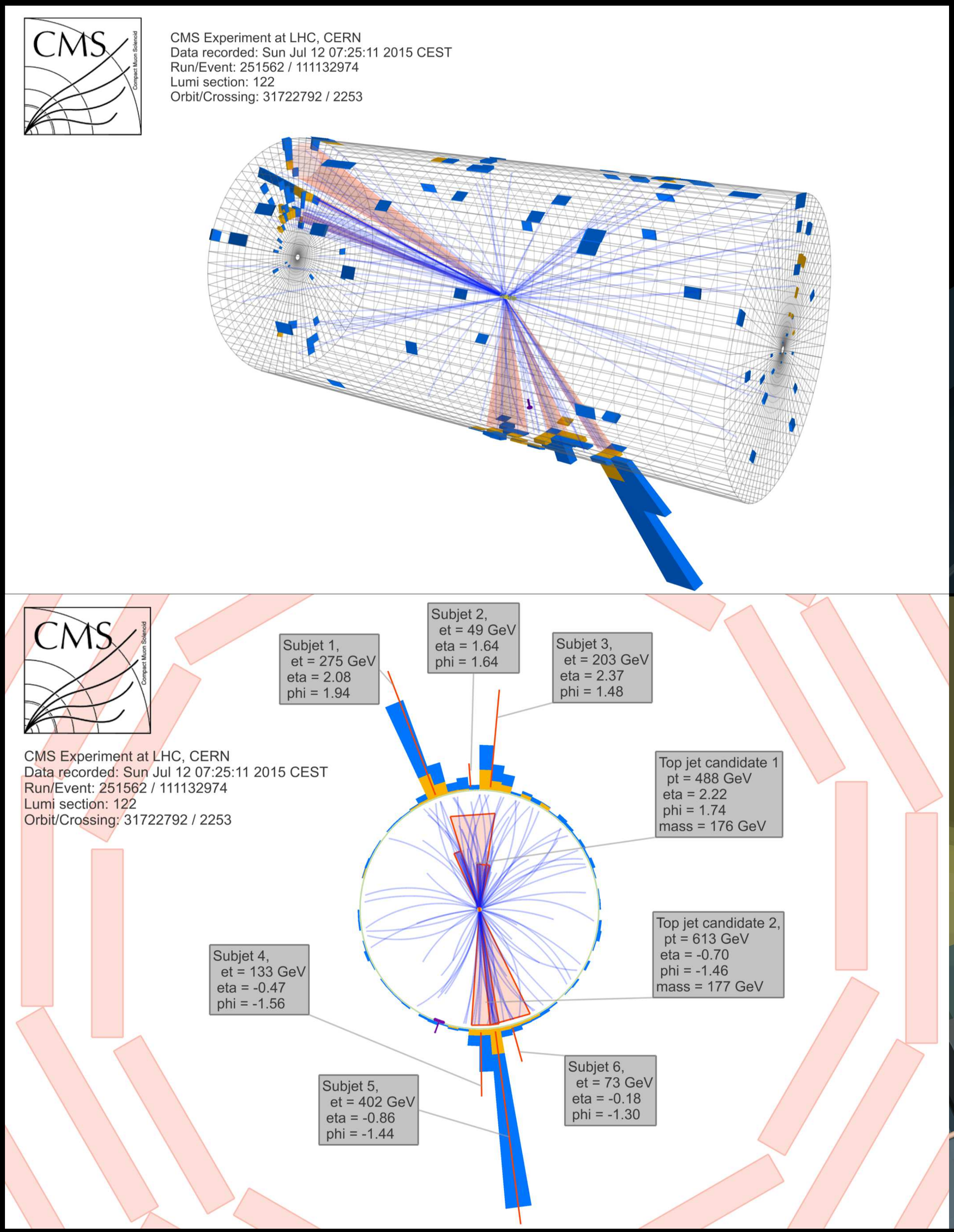
When are two events similar?

The Energy Mover's Distance

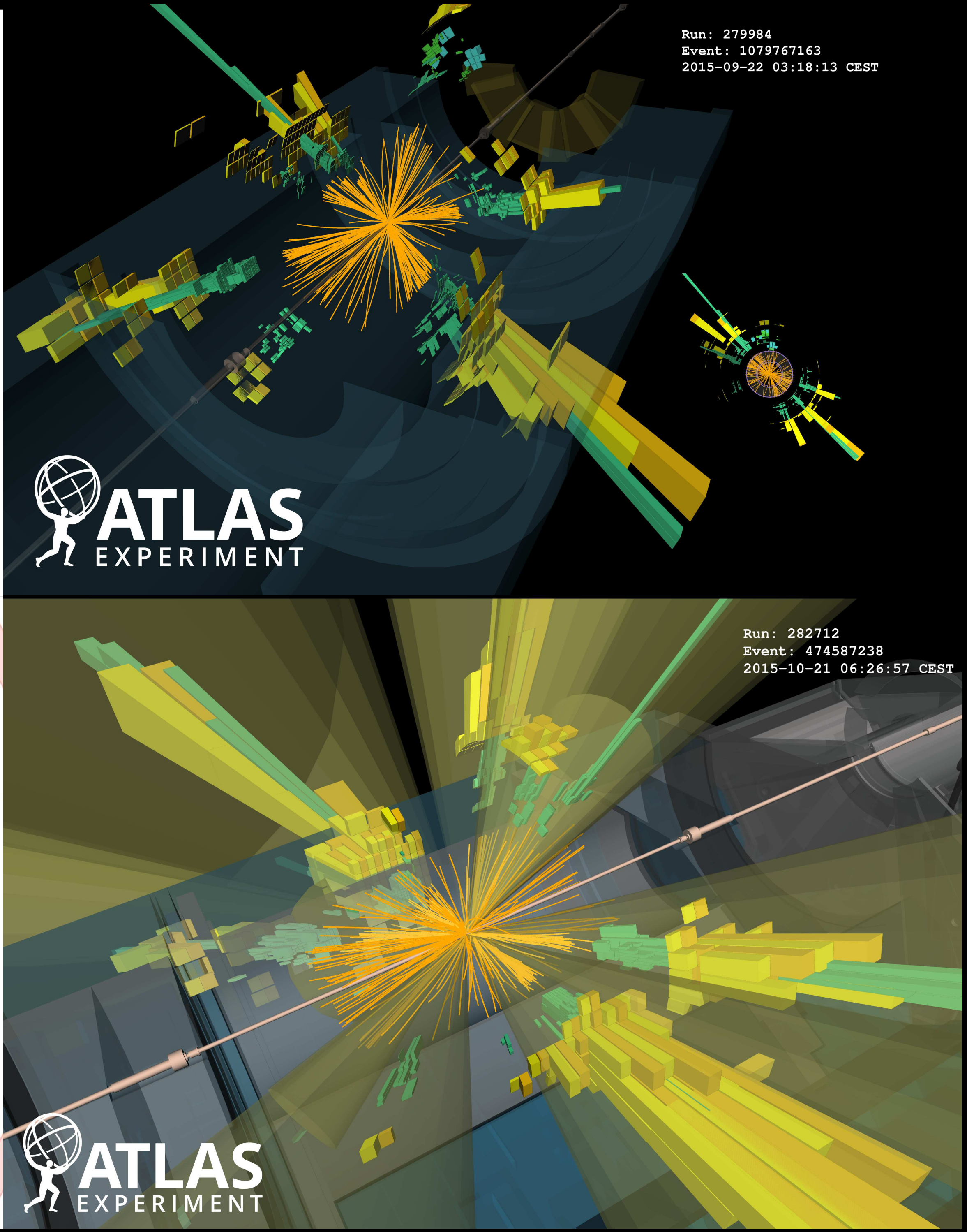
Particle Physics Applications

Events at the Large Hadron Collider

Jets (collimated sprays of color-neutral particles) are ubiquitous at high-energy colliders



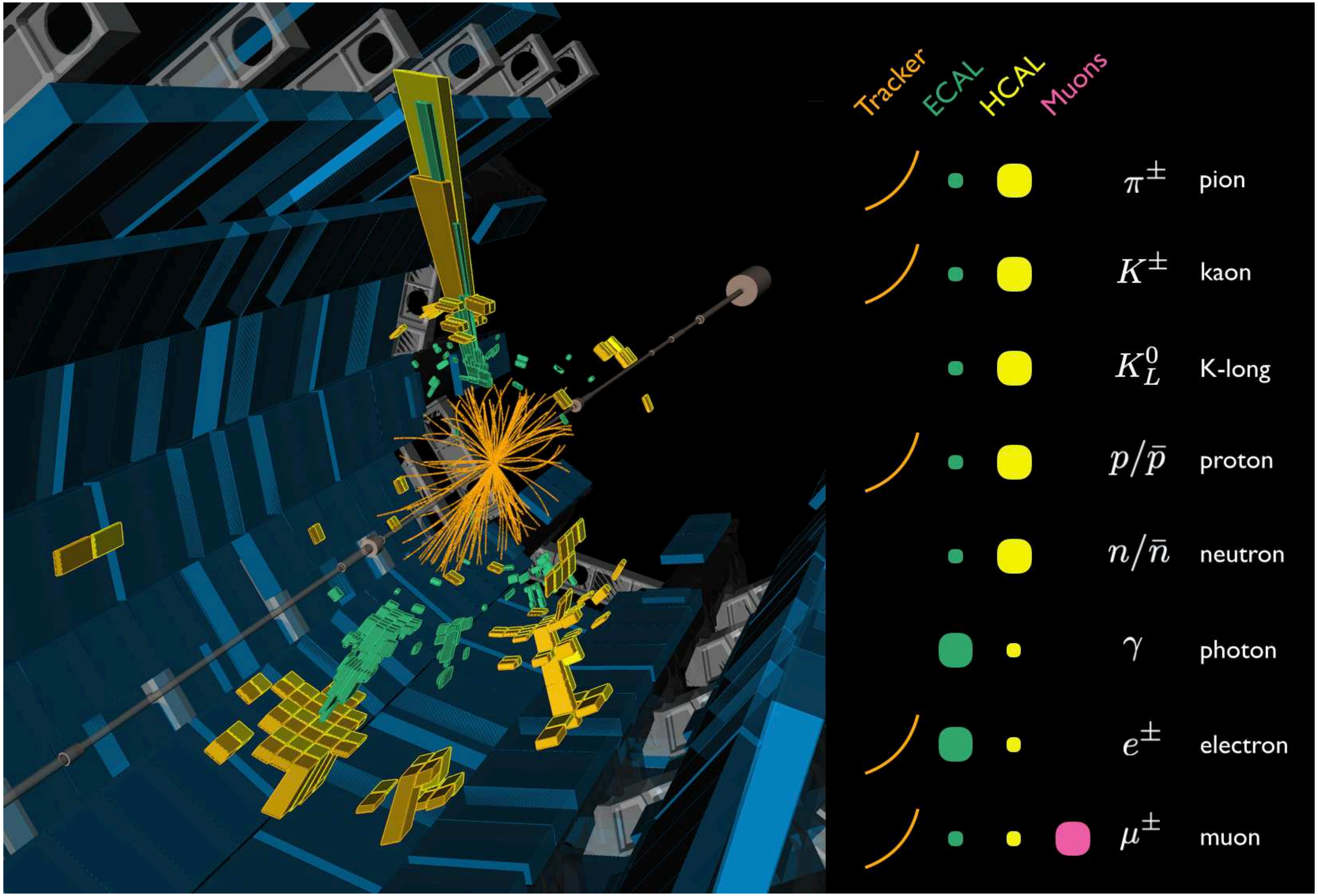
CMS hadronic $t\bar{t}$ event



ATLAS high jet multiplicity events

Events in Detectors

Information synthesized from numerous detector subsystems each with different resolutions and idiosyncrasies



Event Formation in Theory

Hard collision

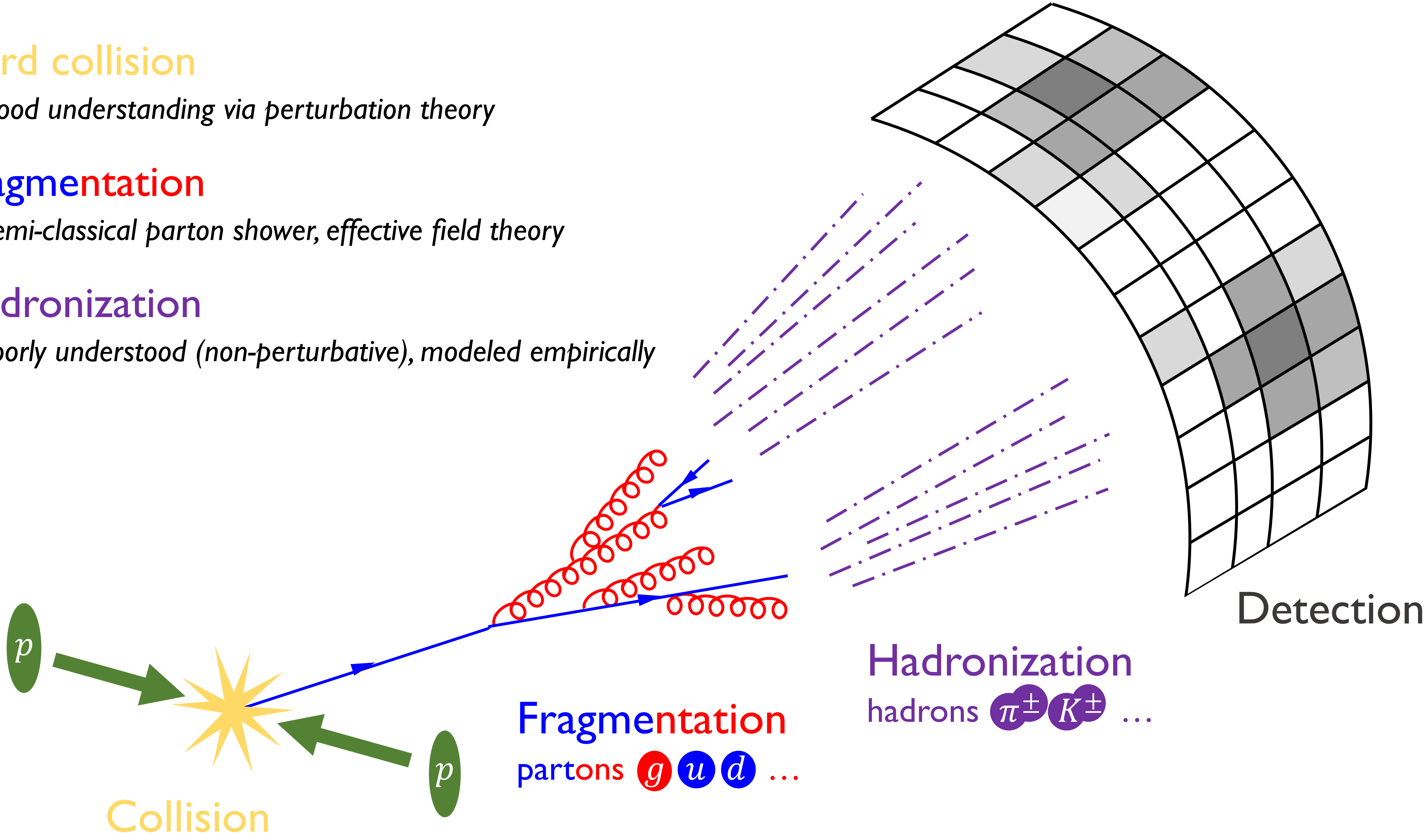
Good understanding via perturbation theory

Fragmentation

Semi-classical parton shower, effective field theory

Hadronization

Poorly understood (non-perturbative), modeled empirically

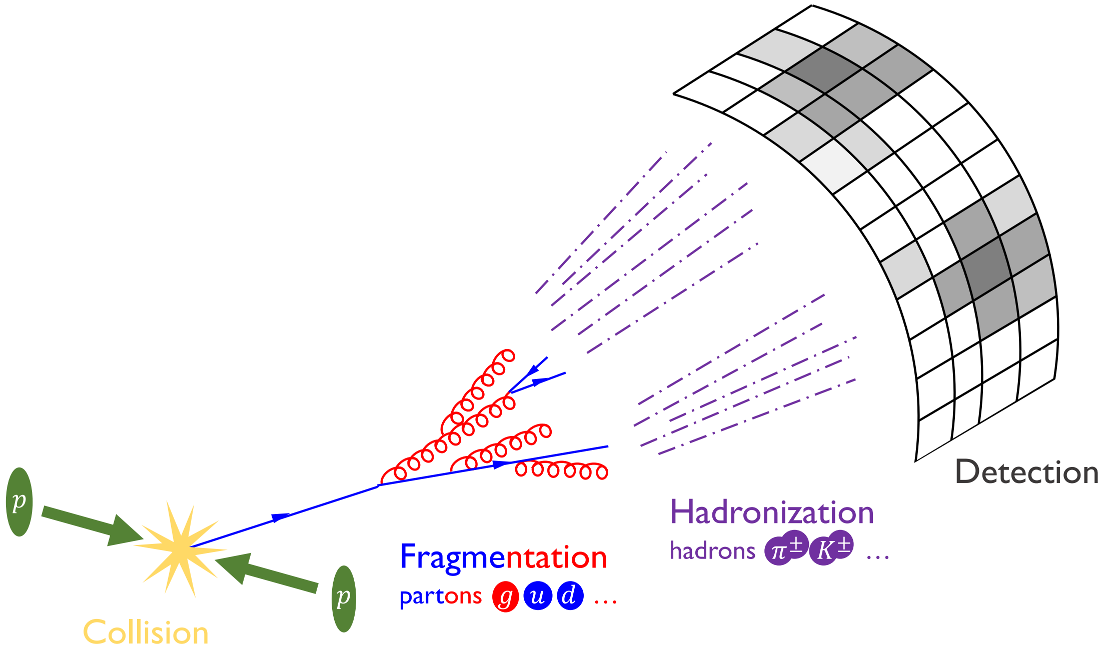


Cartoon of jet formation as a multi-scale process

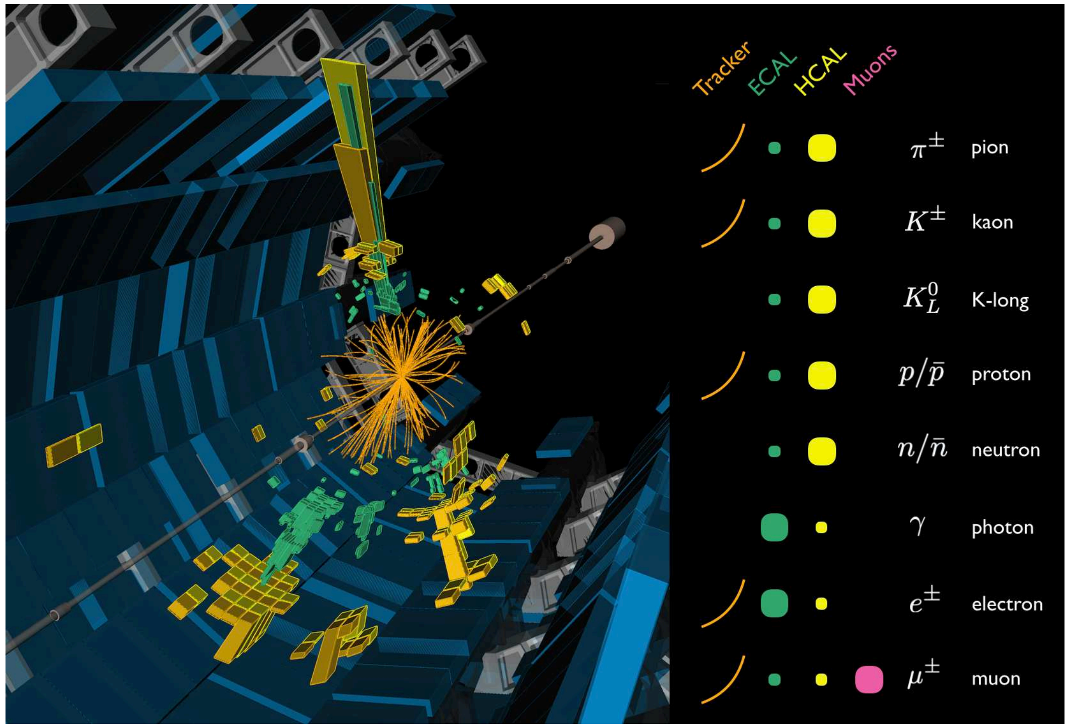
Diagram by Eric Metodiev

Events in Theory vs. Experiment

Theory



Experiment

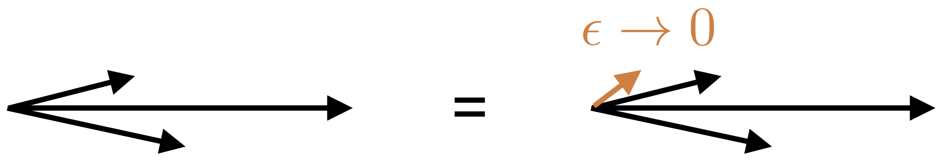


What information is both theoretically and experimentally robust?

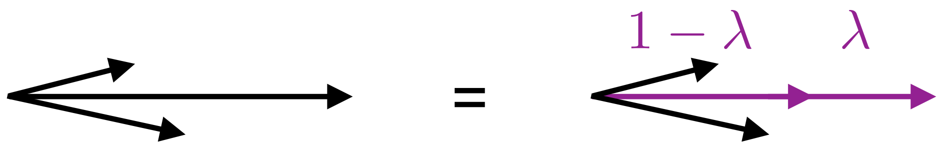
Theoretically and Experimentally Robust Information

Infrared and Collinear Safe Information

Infrared (IR) safety – observable is unchanged under addition of a soft particle



Collinear (C) safety – observable is unchanged under a collinear splitting of a particle



Theoretically

QCD has soft and collinear divergences associated with gluon radiation

$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta} \frac{dz}{z}$$

$$C_q = C_F = 4/3$$
$$C_g = C_A = 3$$

Experimentally

IRC safety is a statement of *linearity* in energy and *continuity* in geometry

[Sveshnikov, Tkachov, hep-ph/9512370; Tkachov, hep-ph/9601308;
Cherzor, Sveshnikov, hep-ph/9710349]

Events as Distributions of Energy

Energy flow distribution fully captures IRC-safe information

[PTK, Metodiev, Thaler, 1810.05165;
PTK, Metodiev, Thaler, *to appear soon*]

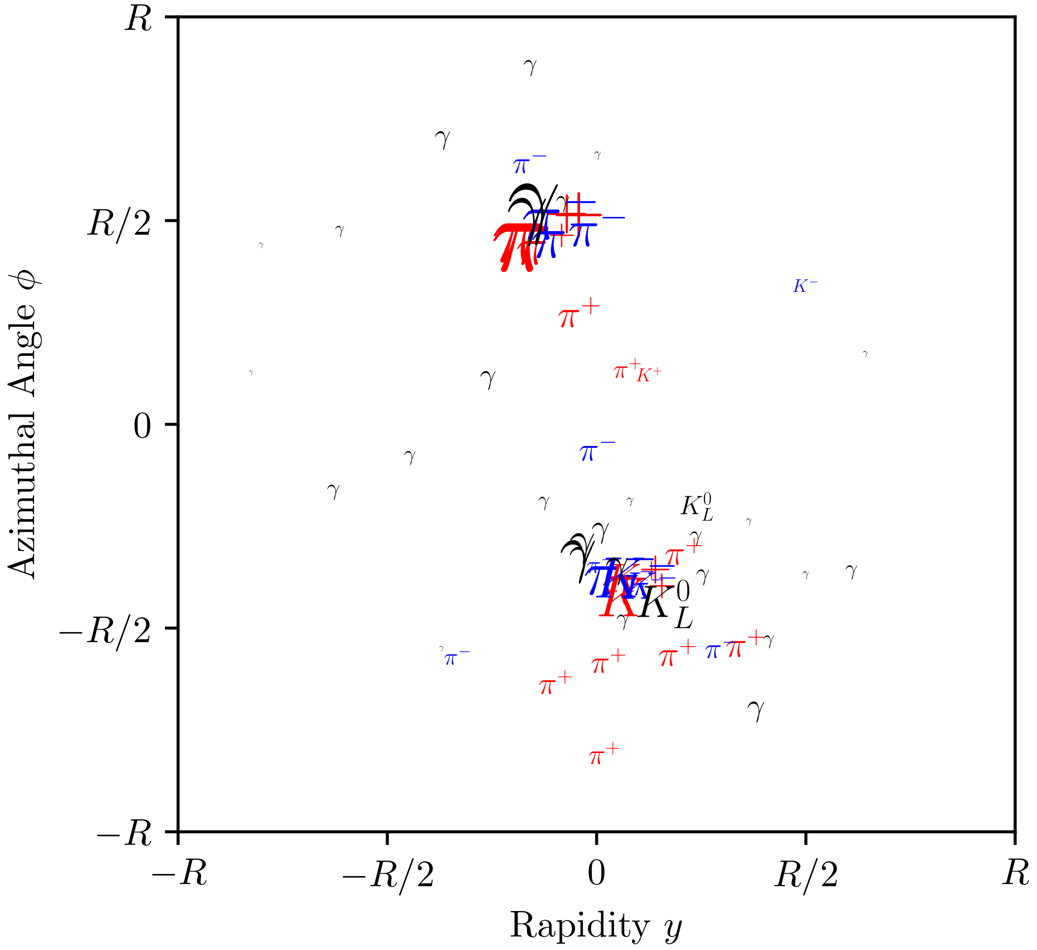
$$\rho(\hat{n}) = \sum_{i=1}^{M} E_i \delta(\hat{n} - \hat{n}_i)$$

Energy Flow Distribution

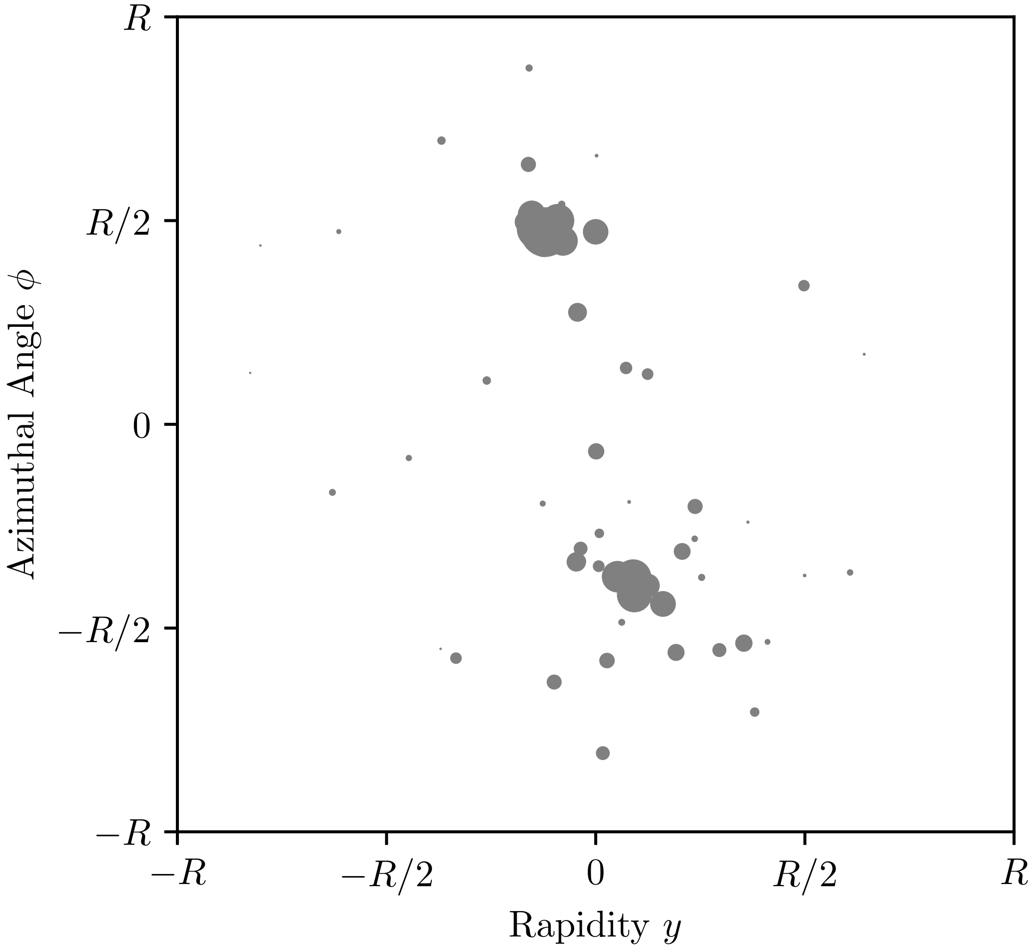
Energy (p_T)

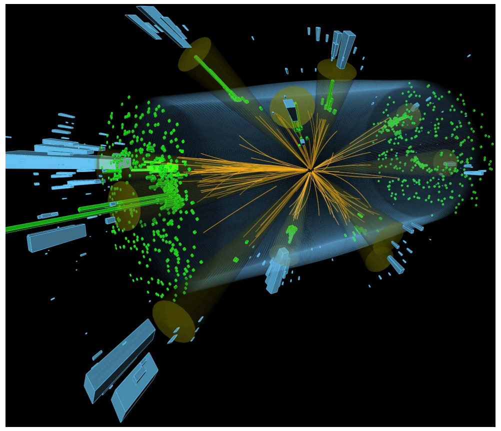
Direction (y,φ)

Full event is a set of particles having momentum and charge/flavor



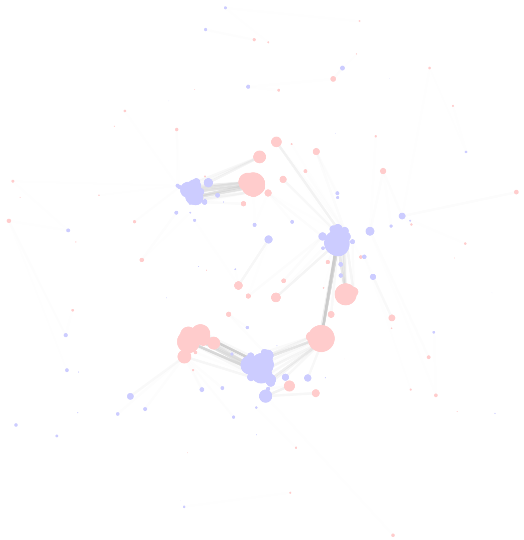
The energy flow is unpixelized and ignores charge/flavor information



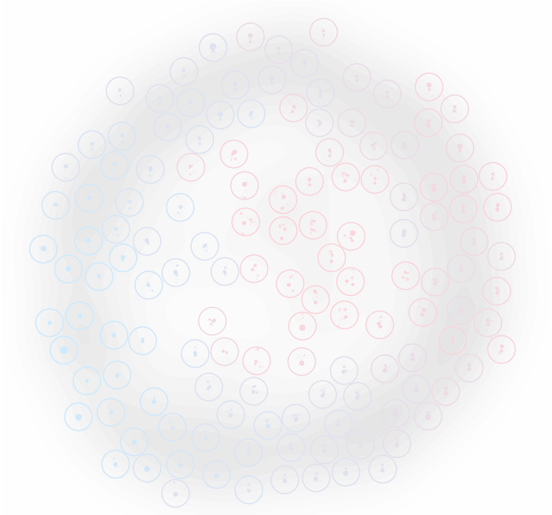


When are two events similar?

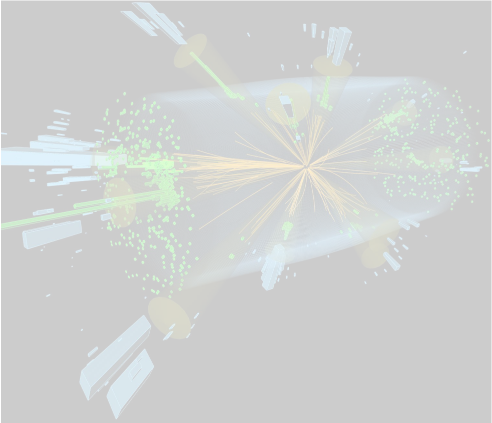
IRC-safe energy flow is theoretically and experimentally robust



The Energy Mover's Distance

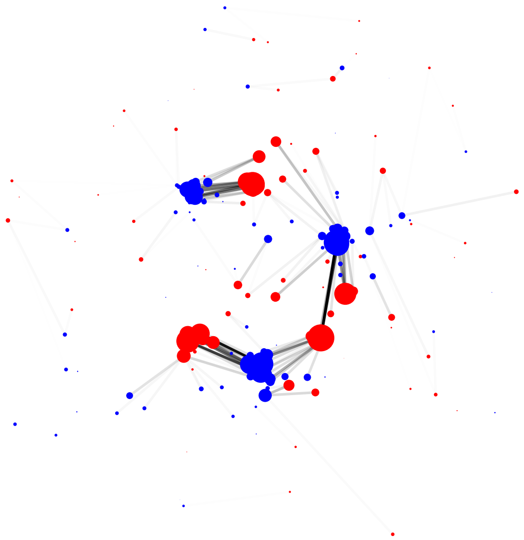


Particle Physics Applications

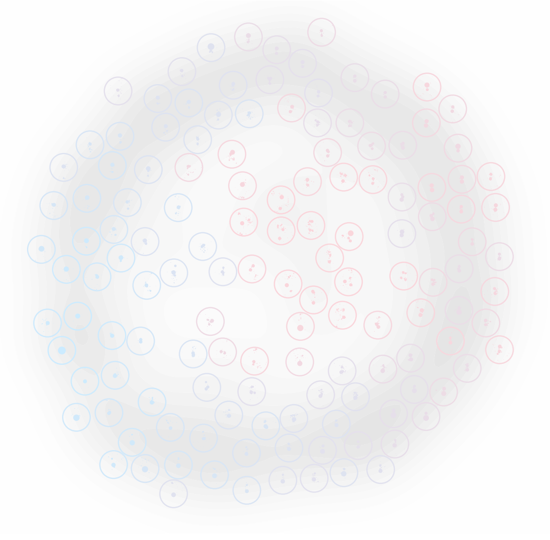


When are two events similar?

IRC-safe energy flow is theoretically and experimentally robust



The Energy Mover’s Distance



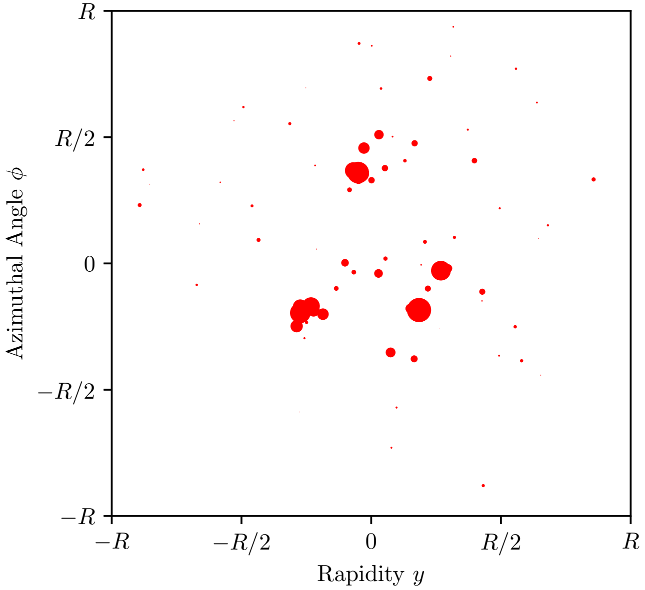
Particle Physics Applications

The Earth Mover's Distance (EMD)

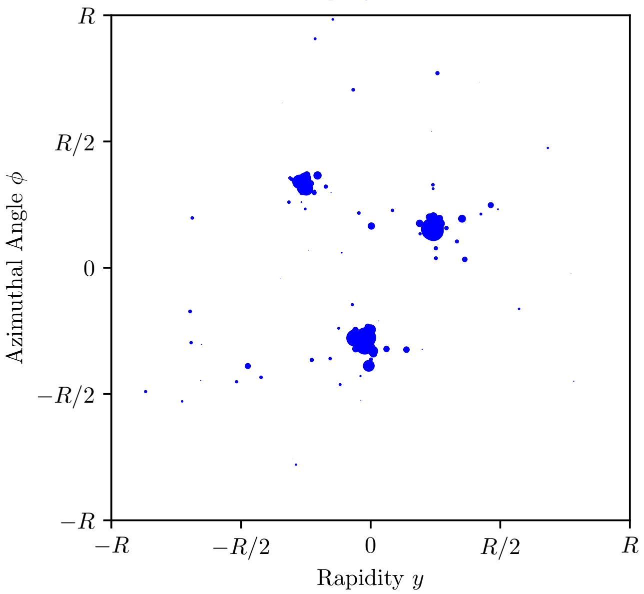
A metric on normalized distributions in a space with a ground distance measure

↳ symmetric, non-negative, triangle inequality, zero iff identical

The minimum "work" (stuff x distance) required to transport supply to demand



Top Jet 1



Top Jet 2

Related to *optimal transport* theory – commonly used as a metric on the space of images

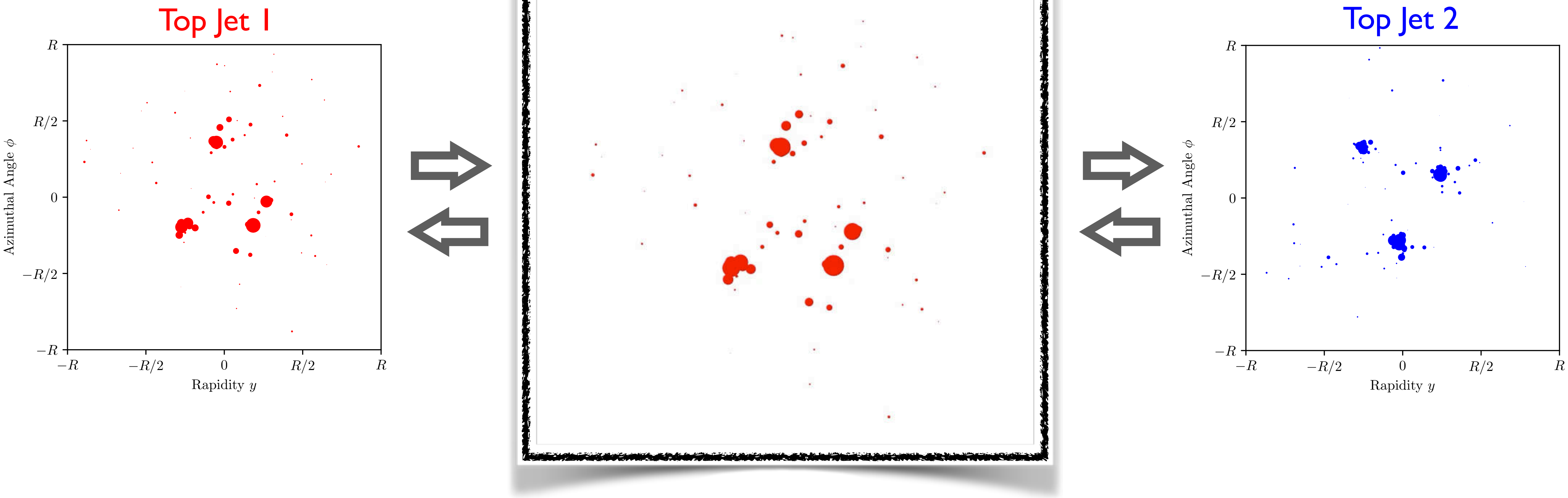
[Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

The Earth Mover's Distance (EMD)

A metric on normalized distributions in a space with a ground distance measure

↳ symmetric, non-negative, triangle inequality, zero iff identical

The minimum "work" (stuff x distance) required to transport supply to demand



Related to *optimal transport* theory – commonly used as a metric on the space of images

[Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

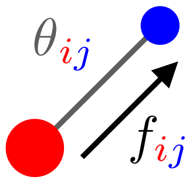
The Energy Mover's Distance (EMD)

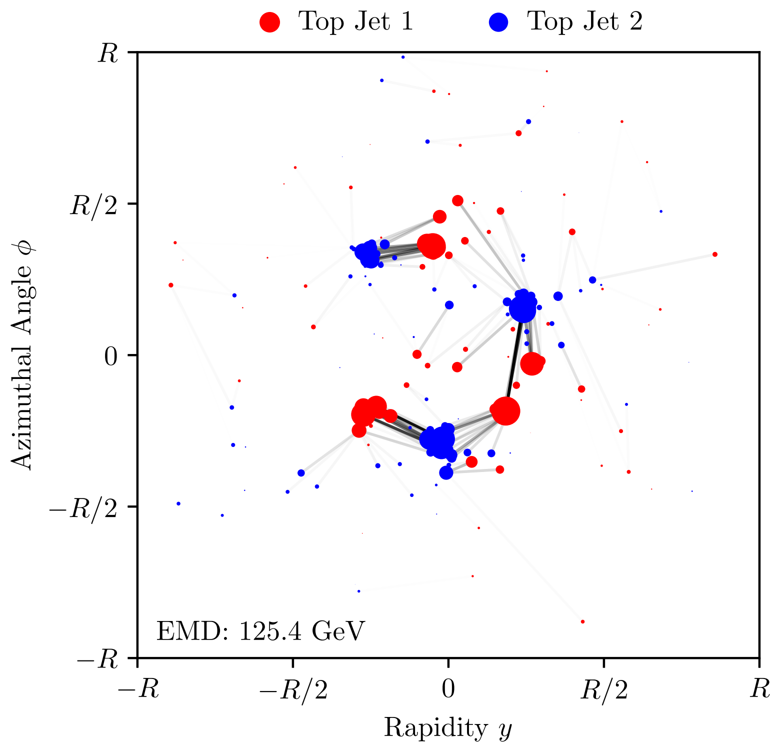
[PTK, Metodiev, Thaler, 1902.02346]

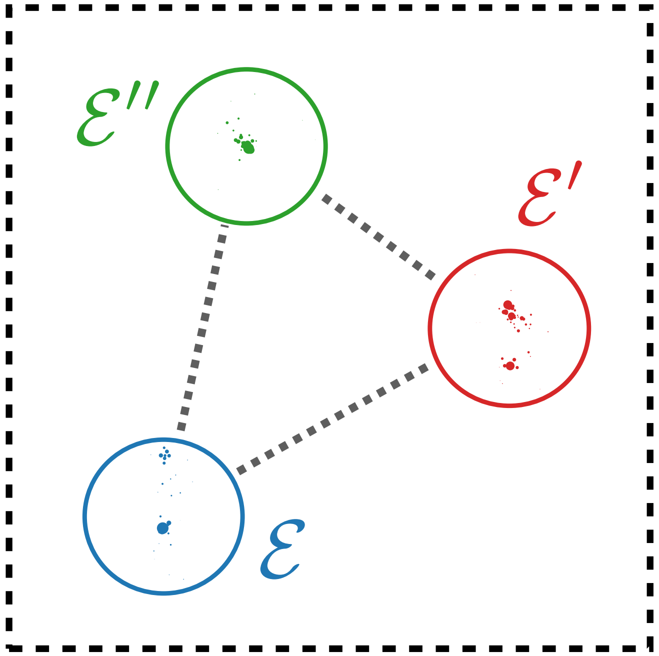
EMD between energy flows defines a metric on the space of events

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \underbrace{\min_{\{f_{ij}\geq 0\}} \sum_i \sum_j f_{ij}\frac{\theta_{ij}}{R}}_{\text{Cost of optimal transport}} + \underbrace{\left|\sum_i E_i - \sum_j E'_j\right|}_{\text{Cost of energy creation}}$$

$$\sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)$$



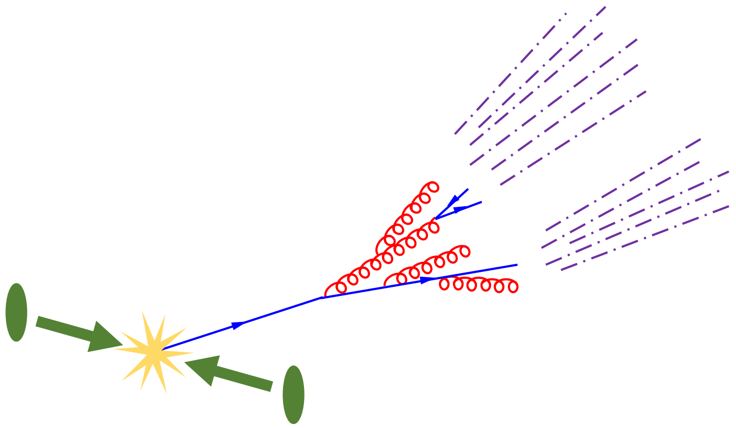


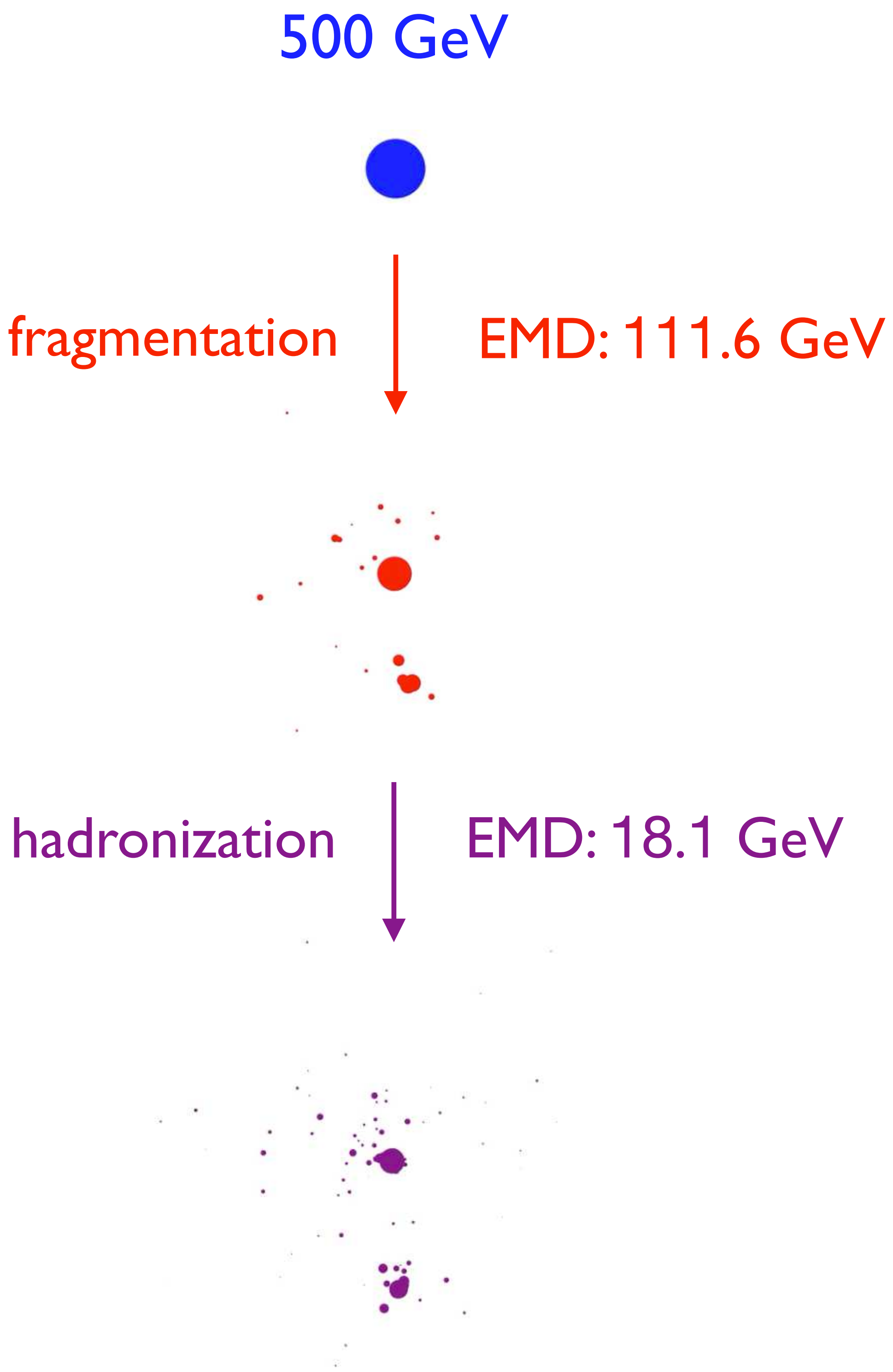


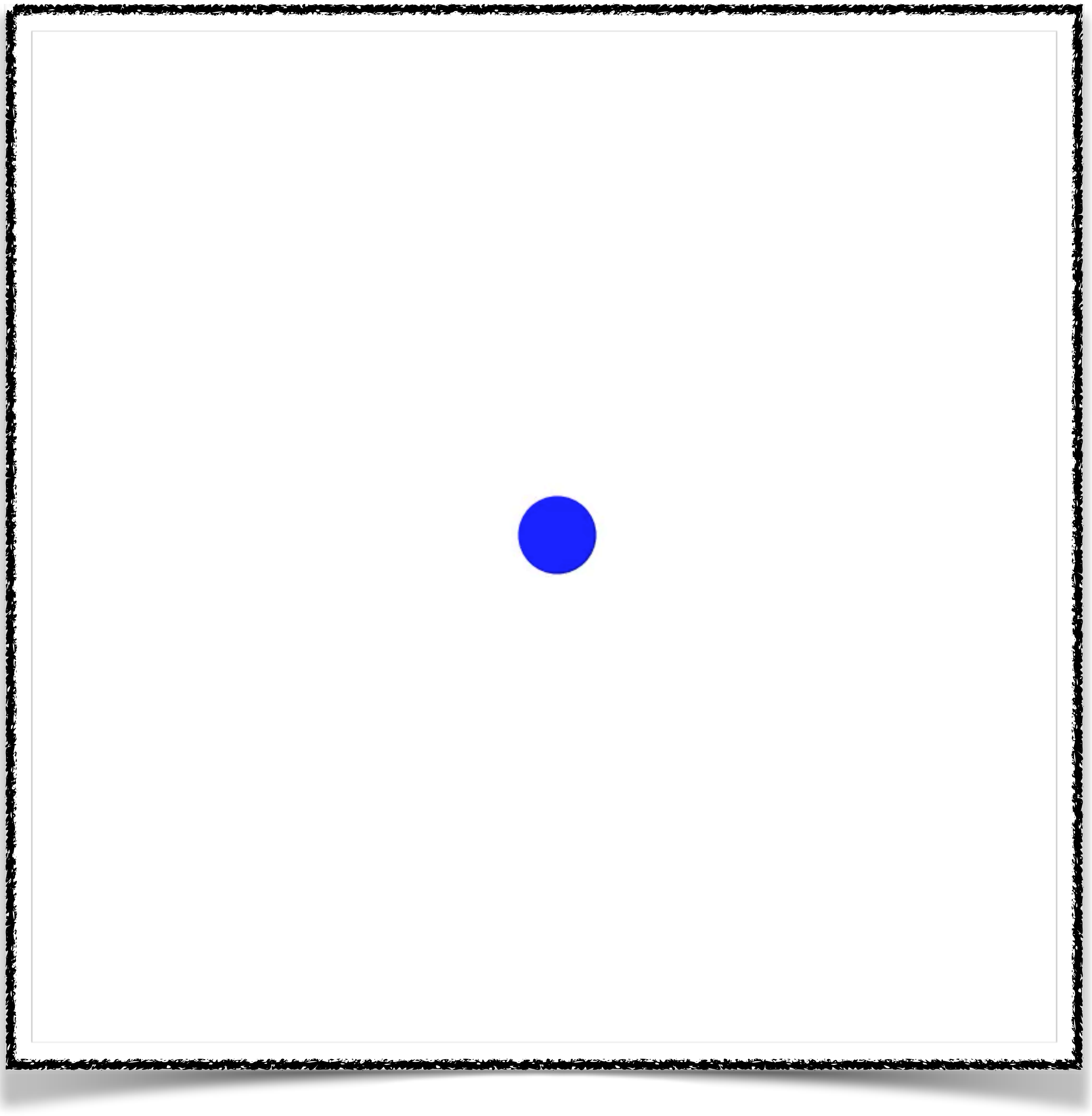
Triangle inequality satisfied for $R \geq d_{\max}/2$

$$0 \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}') \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}'') + \mathrm{EMD}(\mathcal{E}'',\mathcal{E}')$$

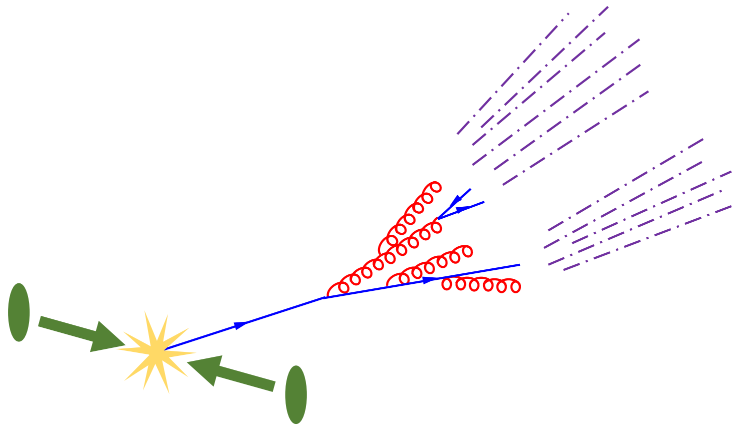
Visualizing Jet Formation – QCD Jets

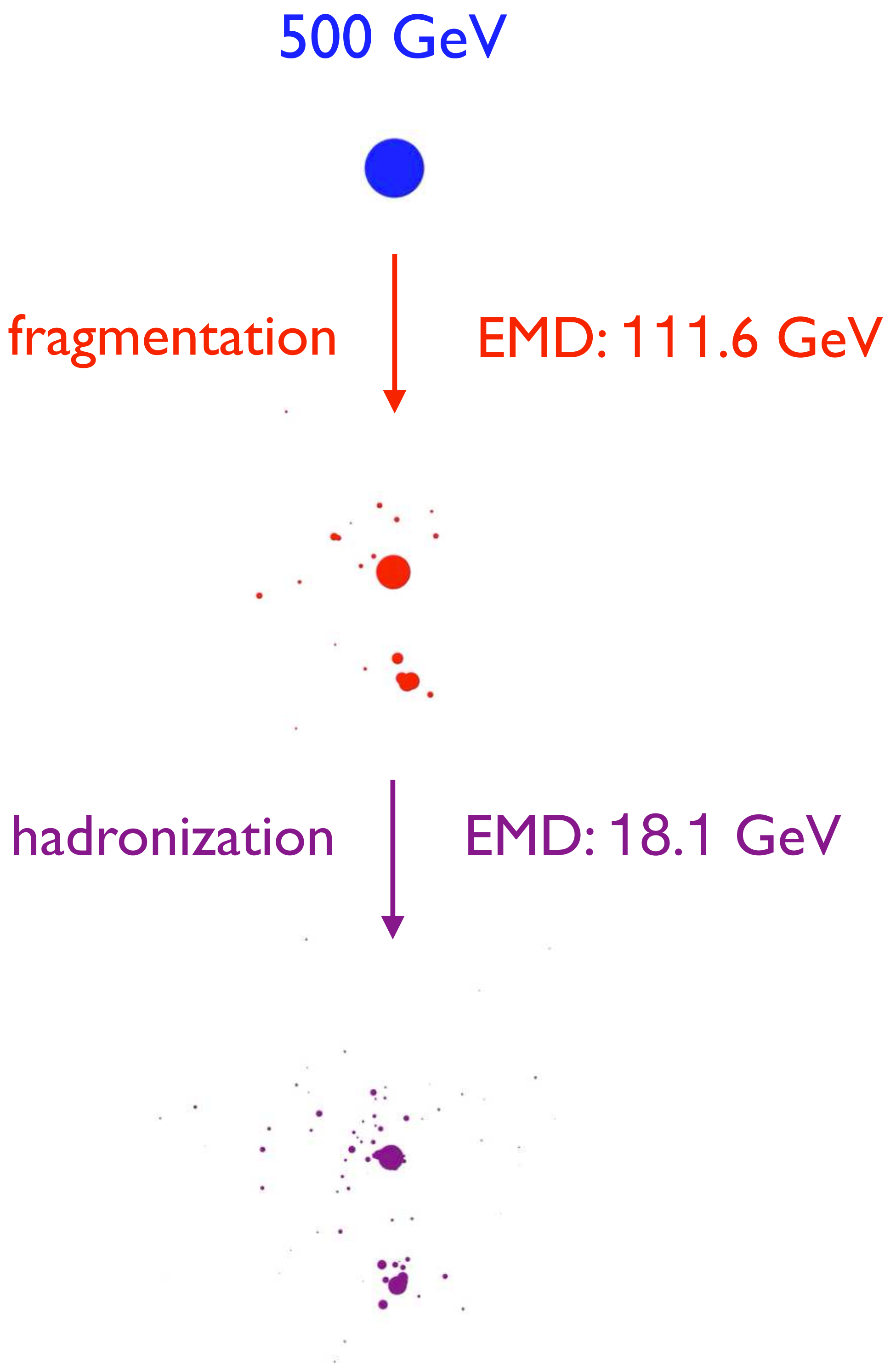






Visualizing Jet Formation – QCD Jets





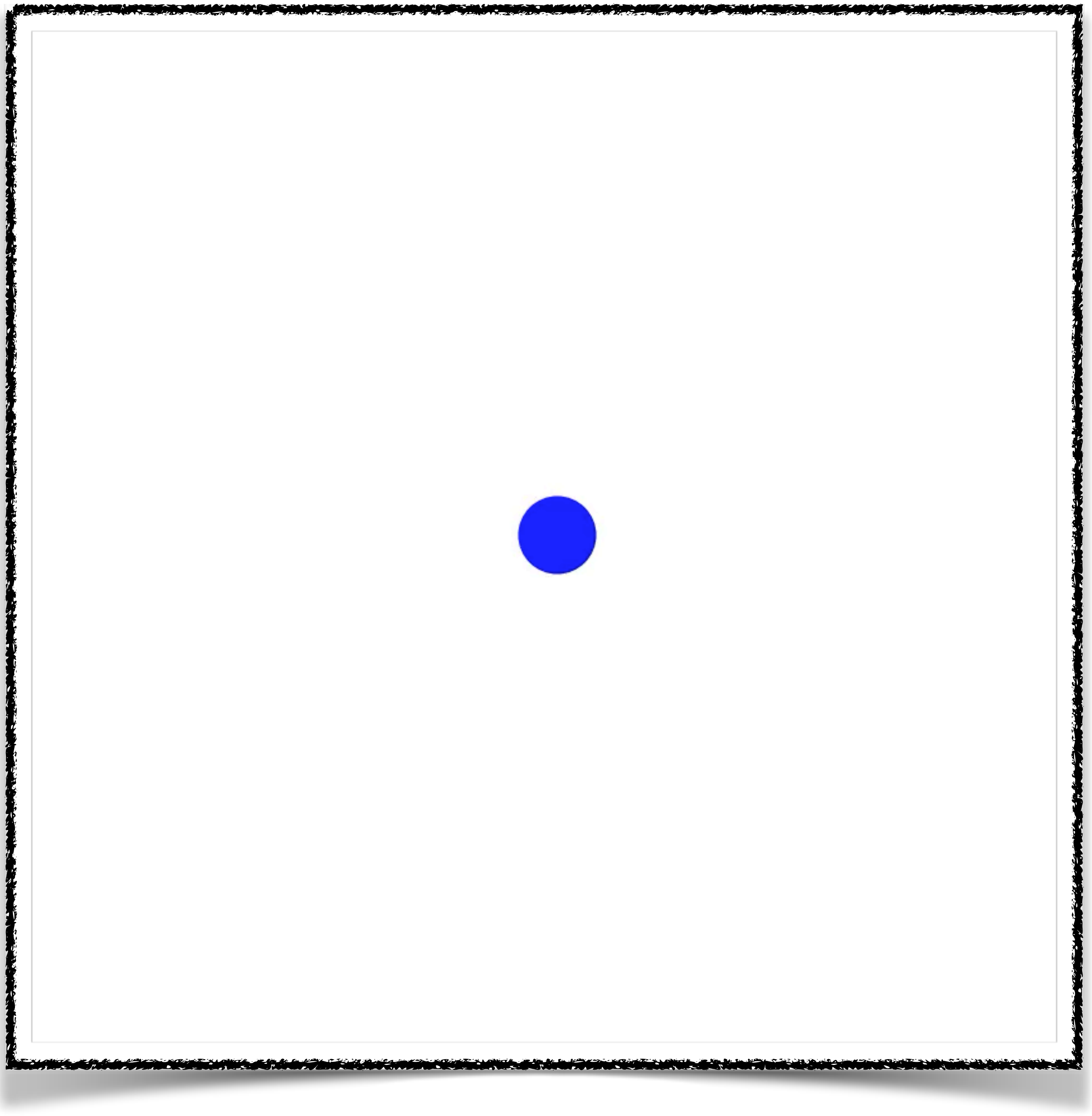
500 GeV

fragmentation

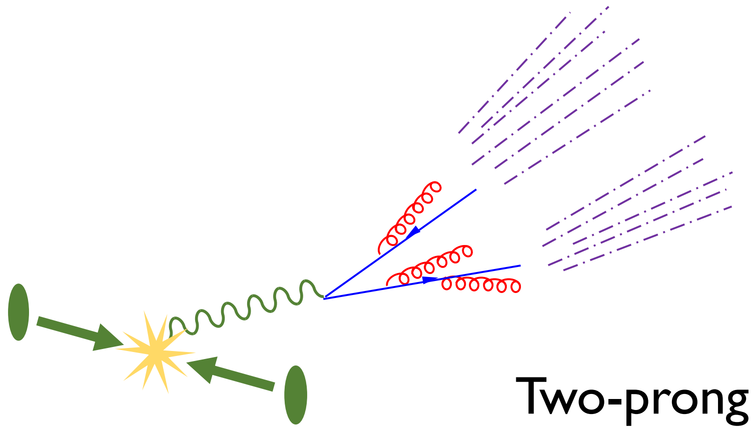
EMD: 111.6 GeV

hadronization

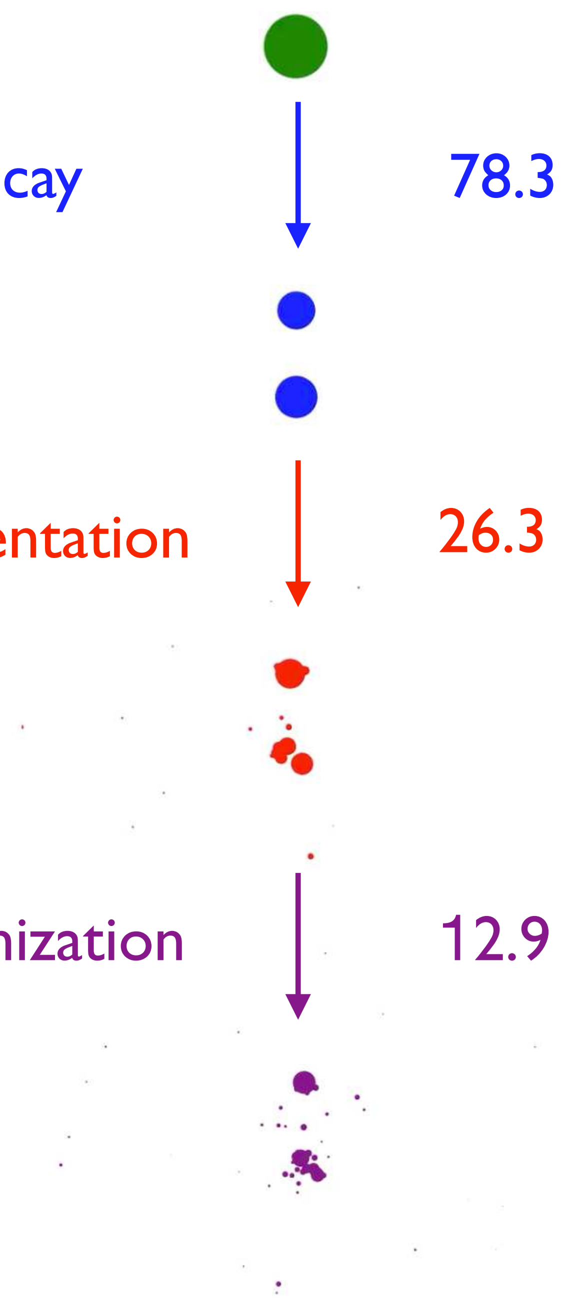
EMD: 18.1 GeV



Visualizing Jet Formation – W Jets



Two-prong decay



500 GeV

decay

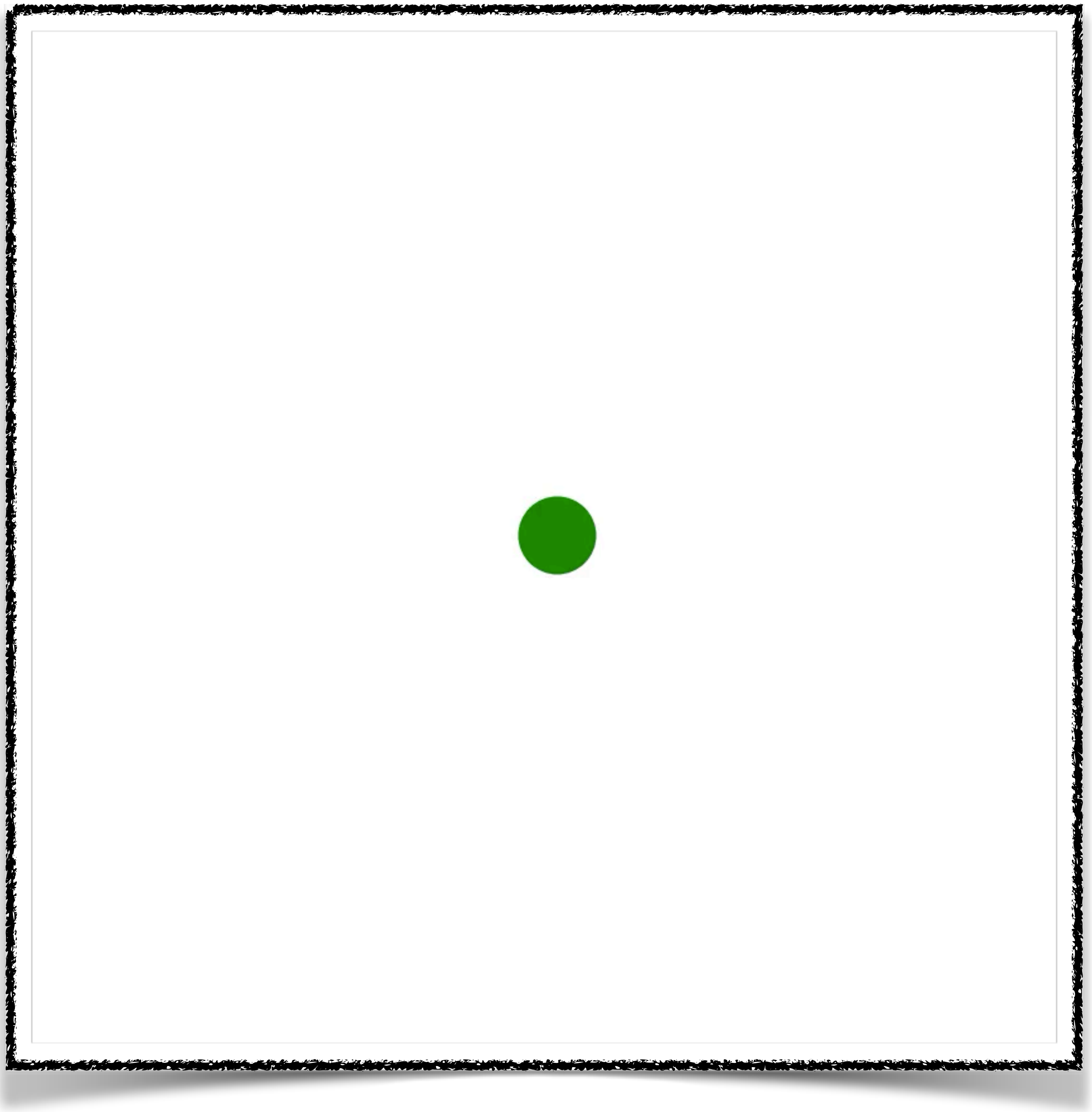
78.3 GeV

fragmentation

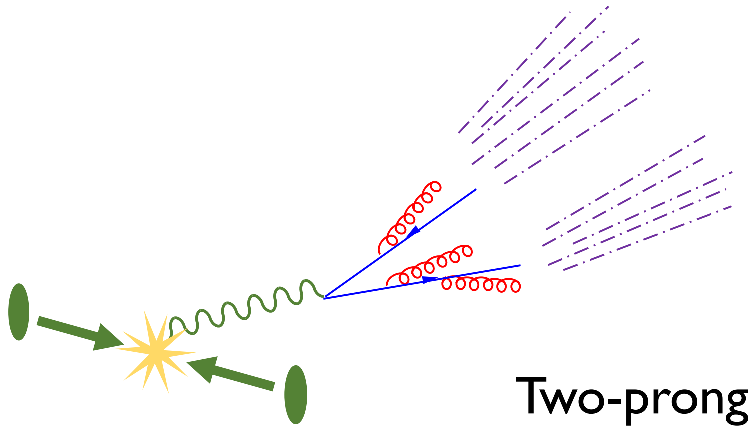
26.3 GeV

hadronization

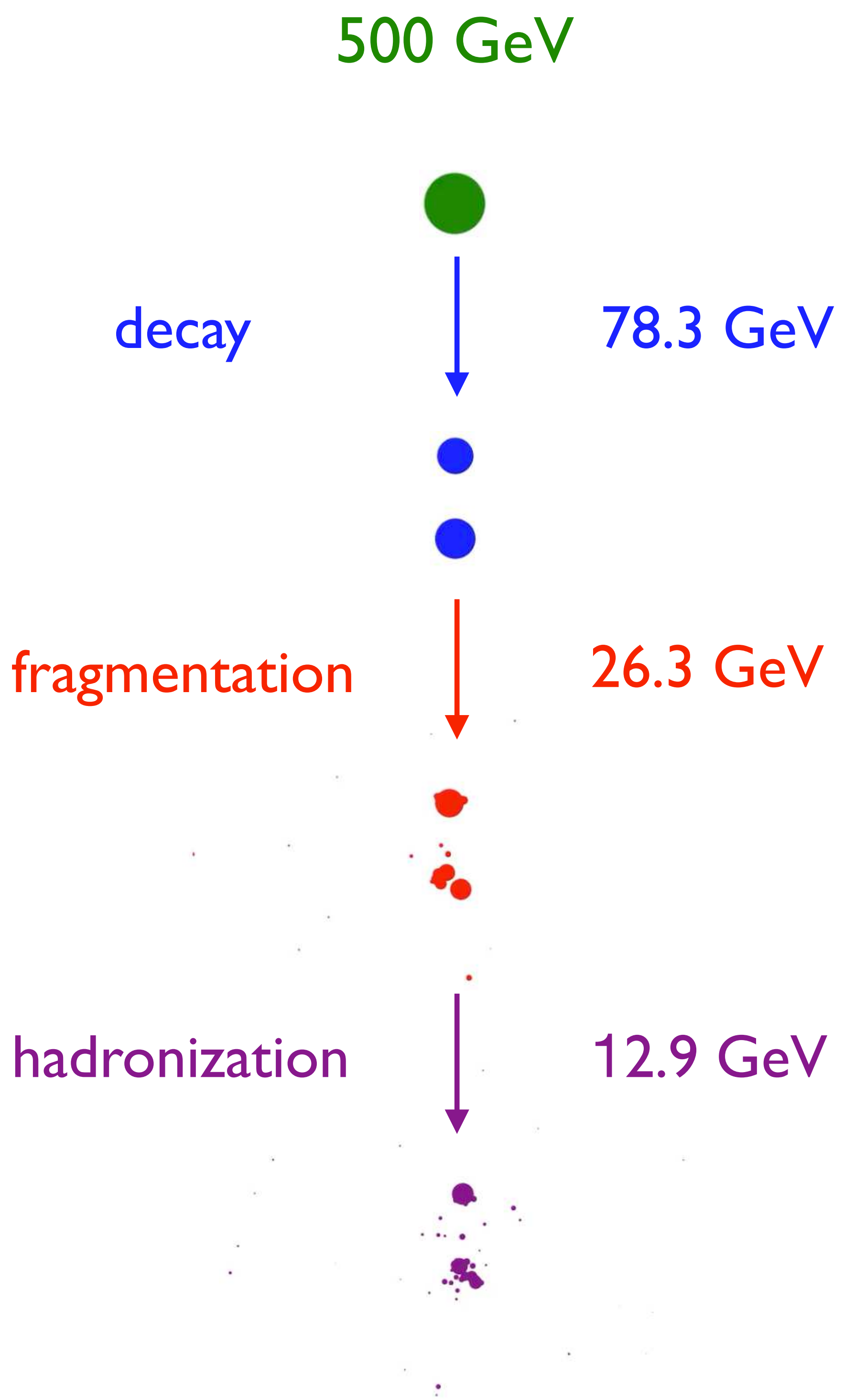
12.9 GeV



Visualizing Jet Formation – W Jets



Two-prong decay



500 GeV

decay

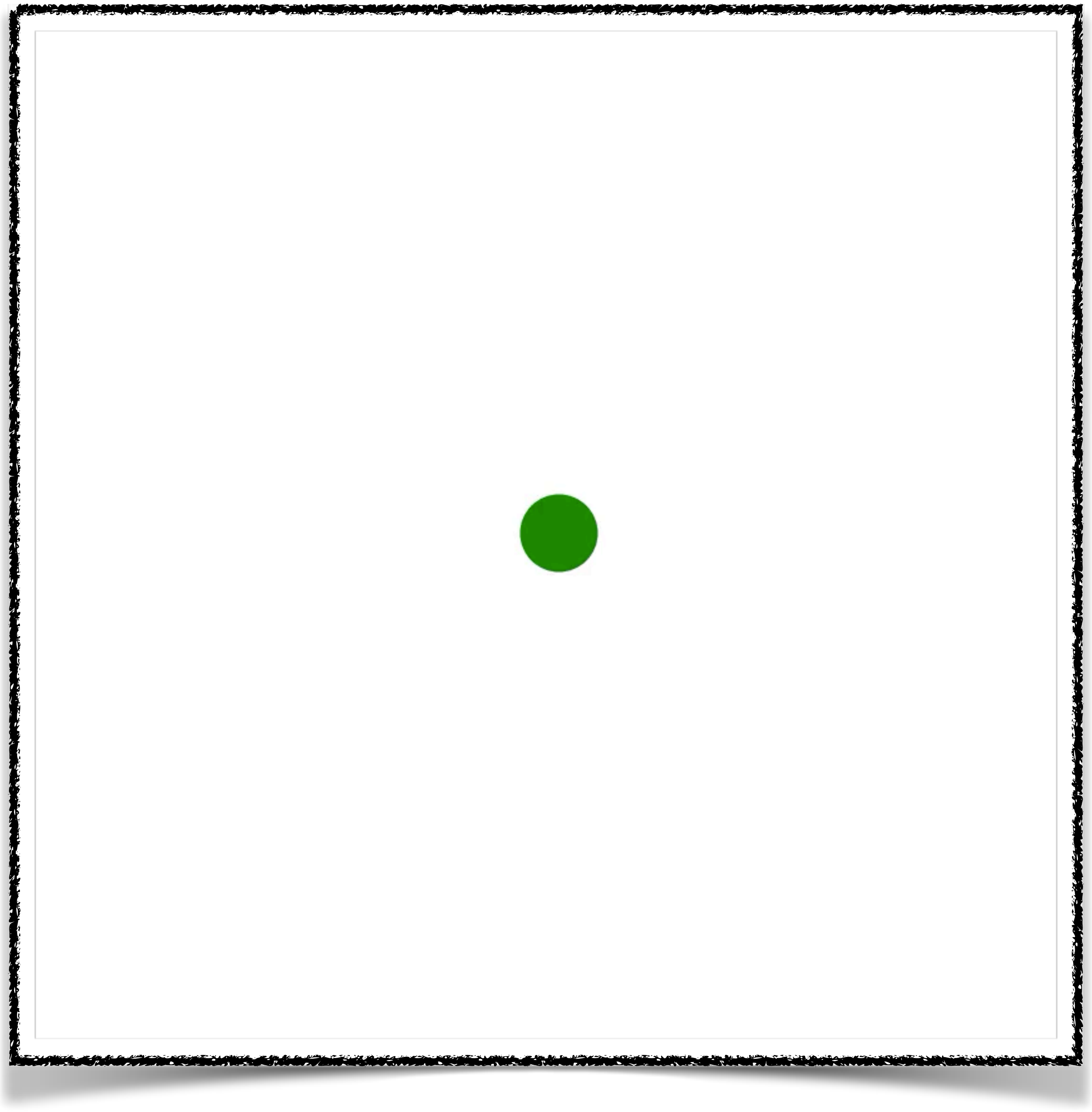
78.3 GeV

fragmentation

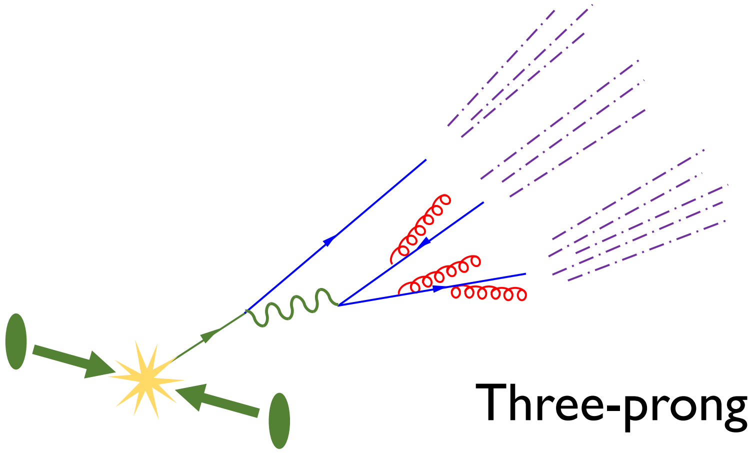
26.3 GeV

hadronization

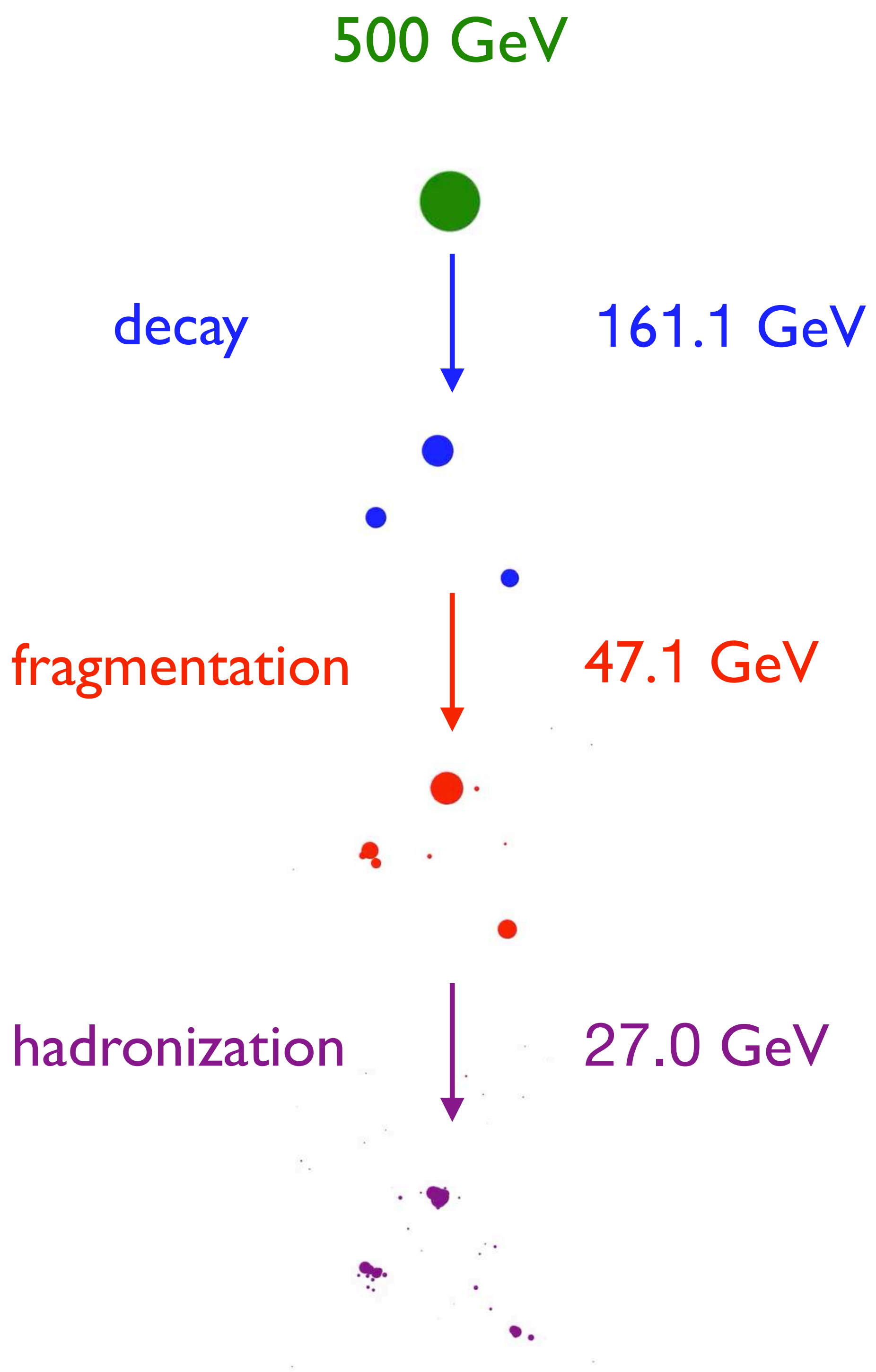
12.9 GeV

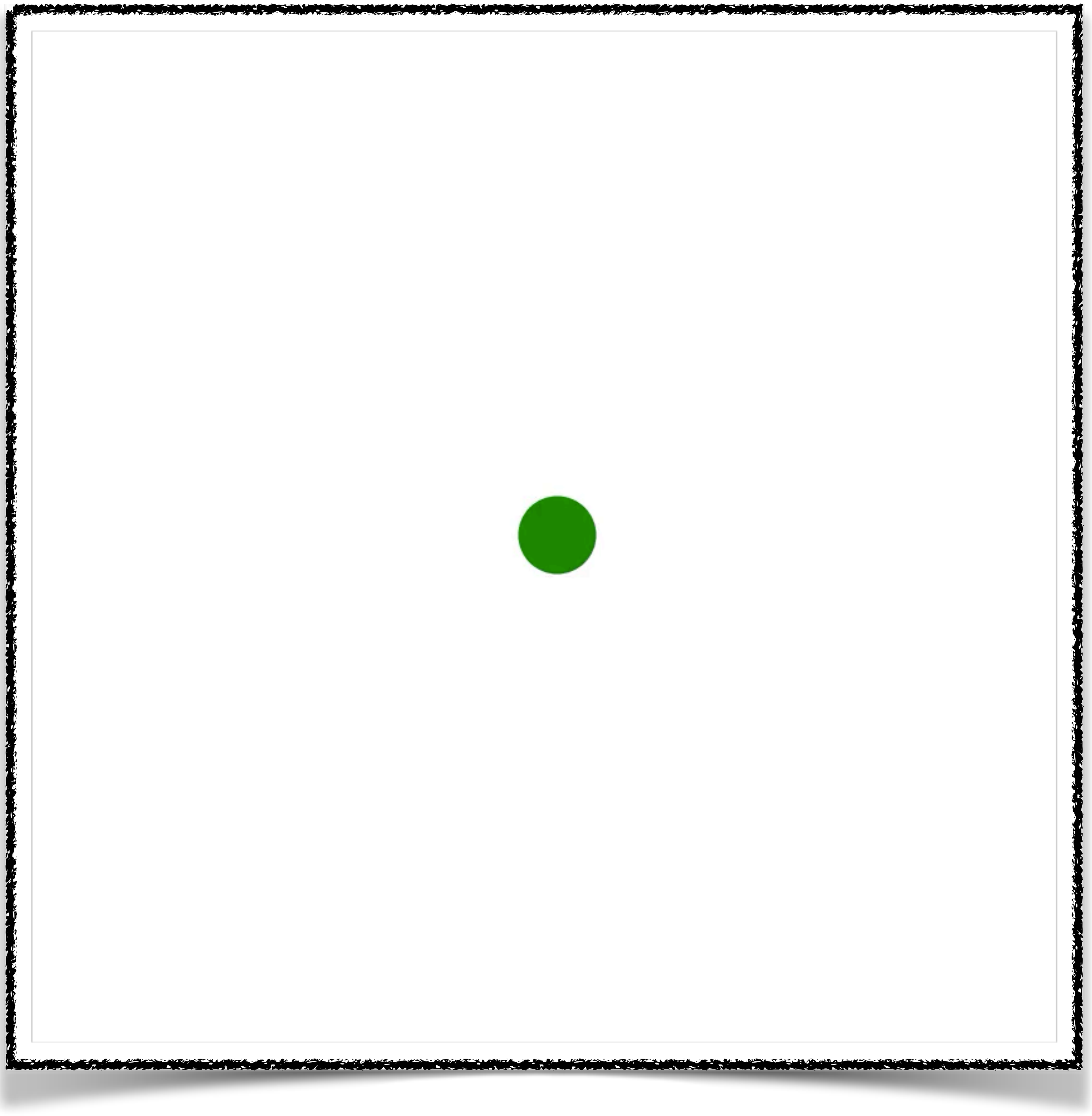


Visualizing Jet Formation – Top Jets

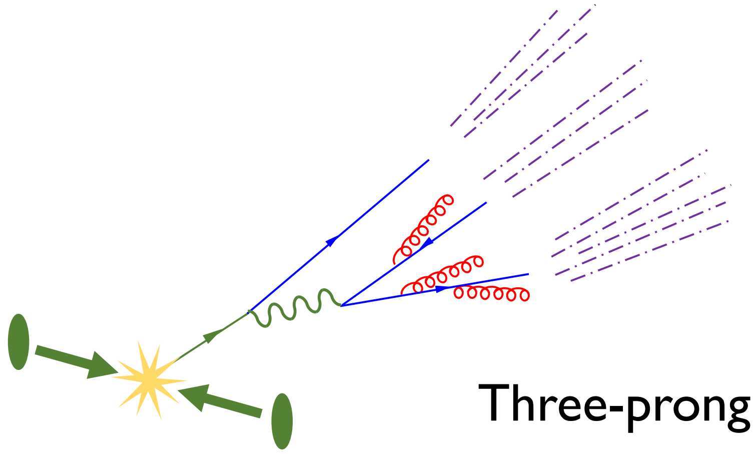


Three-prong decay

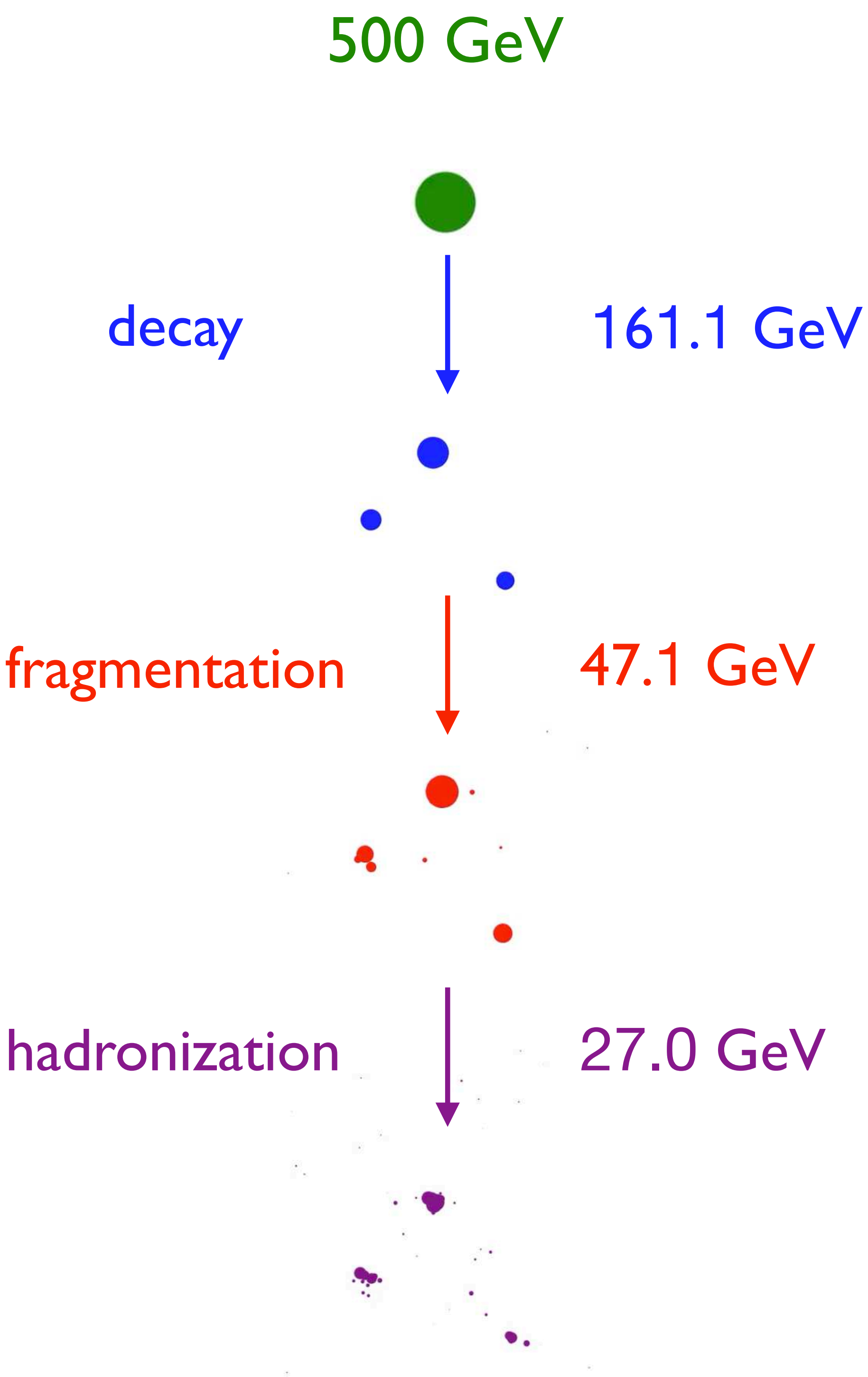




Visualizing Jet Formation – Top Jets



Three-prong decay



500 GeV

decay

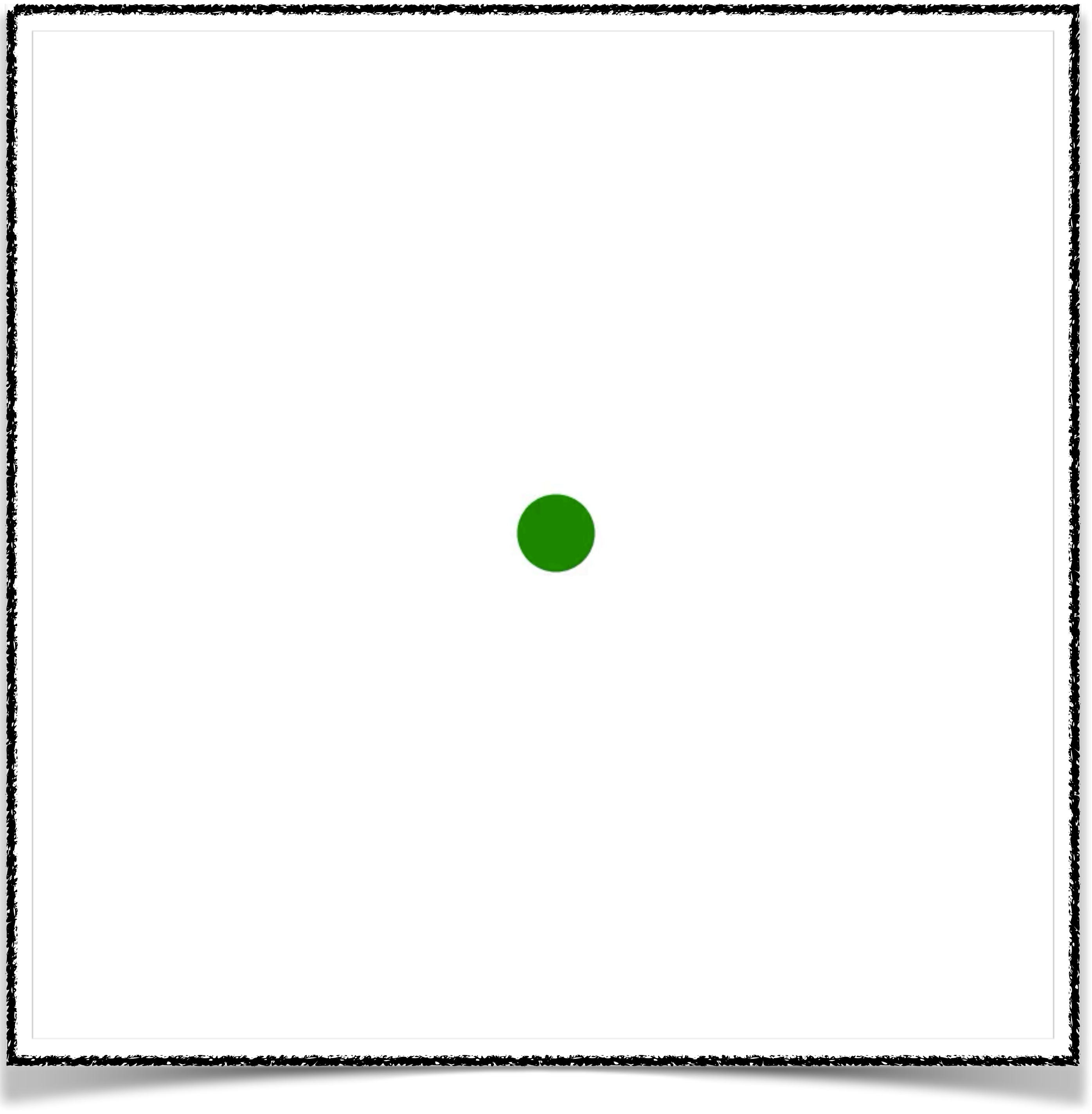
161.1 GeV

fragmentation

47.1 GeV

hadronization

27.0 GeV



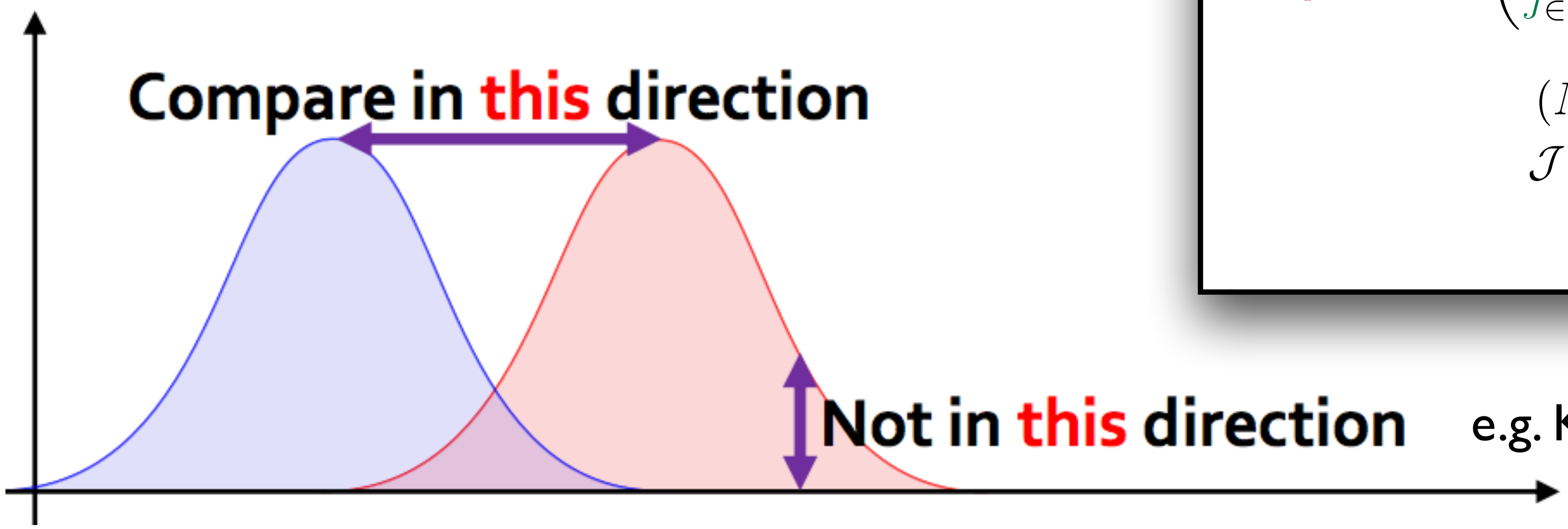
Earth Mover's Distance for the Connoisseur

p-Wasserstein distance is a metric on probability distributions

$$W_p(\mu,\nu) \equiv \left(\inf_{J\in\mathcal{J}(\mu,\nu)} \int_{M\times M} d(x,y)^p \, \mathrm{d}J(x,y)\right)^{1/p}$$

$(M,\,d)$, metric space

$\mathcal{J}(\mu,\nu)$, space of joint distributions
with marginals $\mu,\,\nu$



e.g. K-L divergence compares vertically

[figure from Kun, Math ∩ Programming]

Earth mover's distance is 1-Wasserstein
metric on discrete distributions

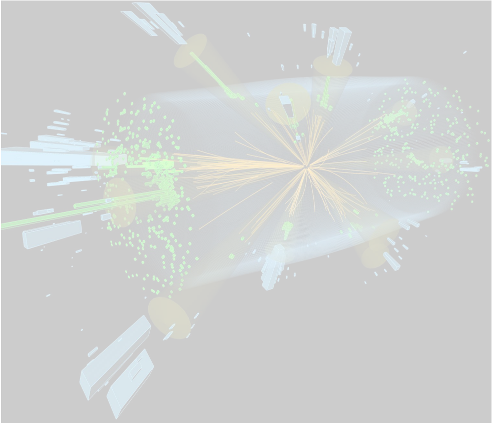
Recent use in Machine Learning

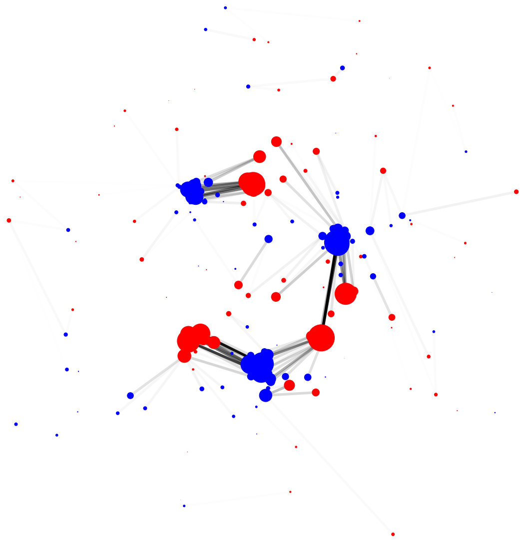
Wasserstein Generative Adversarial Networks

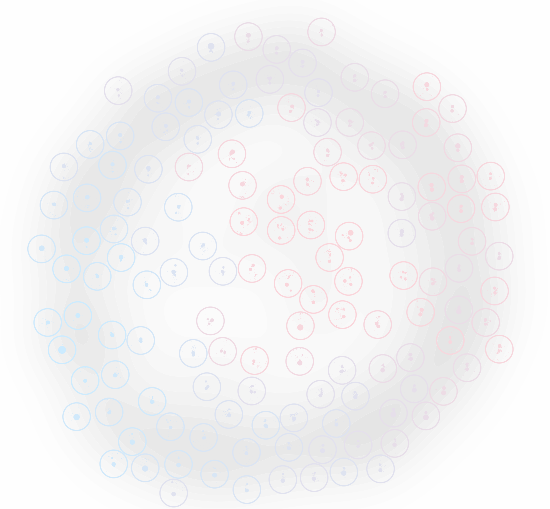
[Arjovsky, Chintala, Bottou, 1701.07875;
in particle physics:
- Erdmann, Geiger, Glombitza, Schmidt, 1802.03325
- Erdmann, Glombitza, Quast, 1807.01954]

Wasserstein(-Wasserstein) Autoencoders

[Tolstikhin, Bousquet, Gelly, Shoelkopf, 1711.01558]
[Zhang, Gao, Jiao, Liu, Wang, Yang, 1902.09323]







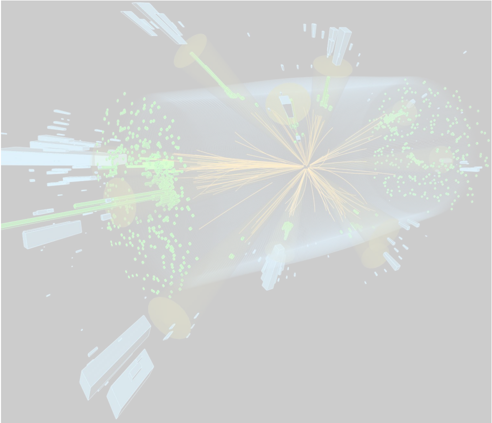
When are two events similar?

IRC-safe energy flow is theoretically and experimentally robust

The Energy Mover's Distance

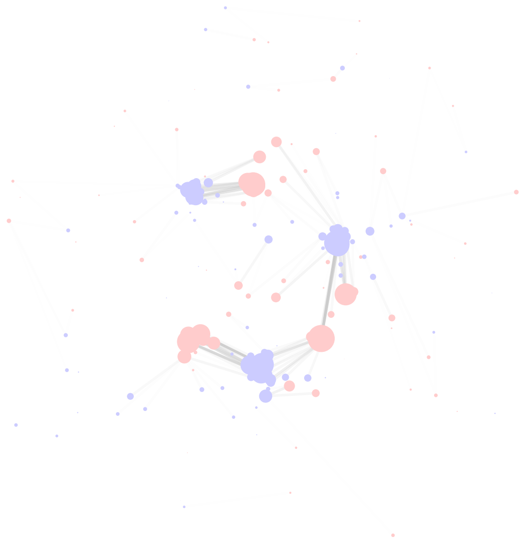
Quantifies the difference in energy flow between events

Particle Physics Applications



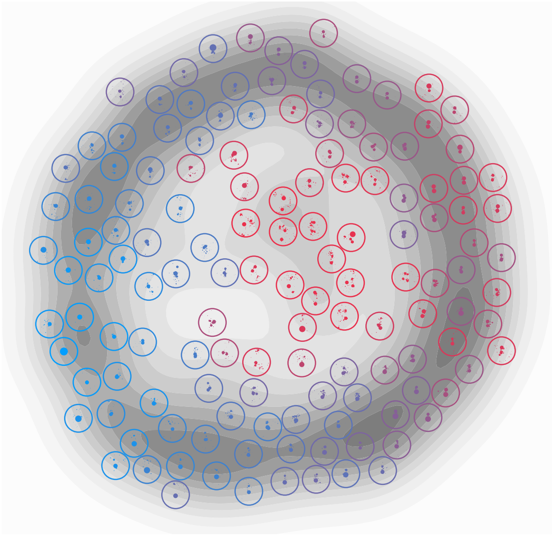
When are two events similar?

IRC-safe energy flow is theoretically and experimentally robust



The Energy Mover's Distance

Quantifies the difference in energy flow between events



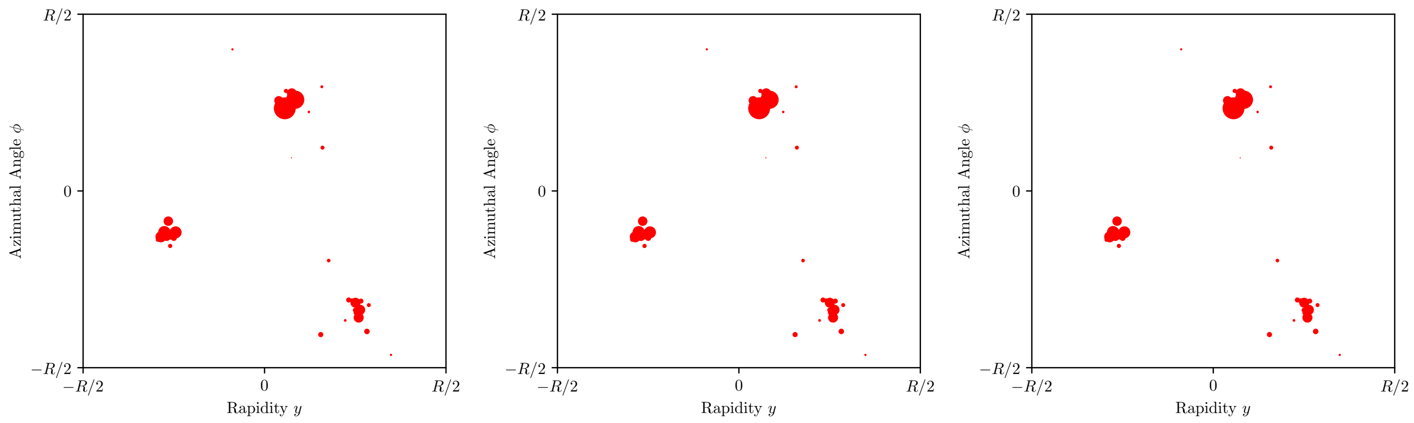
Particle Physics Applications

Old – Geometric Interpretation of Familiar Observables

N-subjettiness

[Thaler, Van Tilburg, 1011.2268, 1108.2701]

$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\ \text{axes}} \sum_i E_i \min\left(\theta_{i1}^\beta, \theta_{i2}^\beta, \ldots, \theta_{iN}^\beta\right)$$



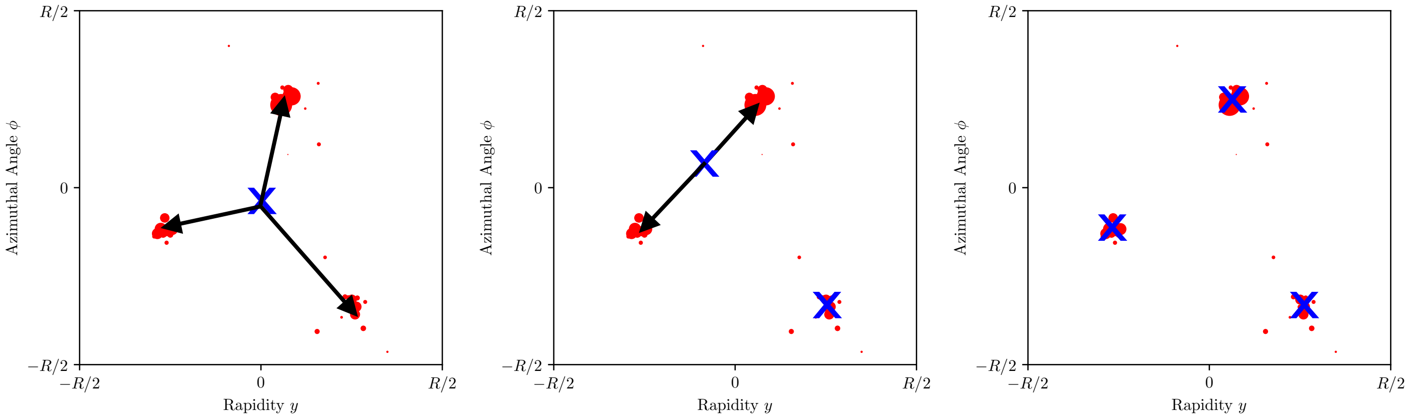
Old – Geometric Interpretation of Familiar Observables

N-subjettiness

[Thaler, Van Tilburg, 1011.2268, 1108.2701]

$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\ \text{axes}} \sum_i E_i \min\left(\theta_{i1}^{\beta}, \theta_{i2}^{\beta}, \ldots, \theta_{iN}^{\beta}\right)$$

$$\tau_1(\mathcal{E}) \quad \gg \quad \tau_2(\mathcal{E}) \quad \gg \quad \tau_3(\mathcal{E}) \simeq 0$$



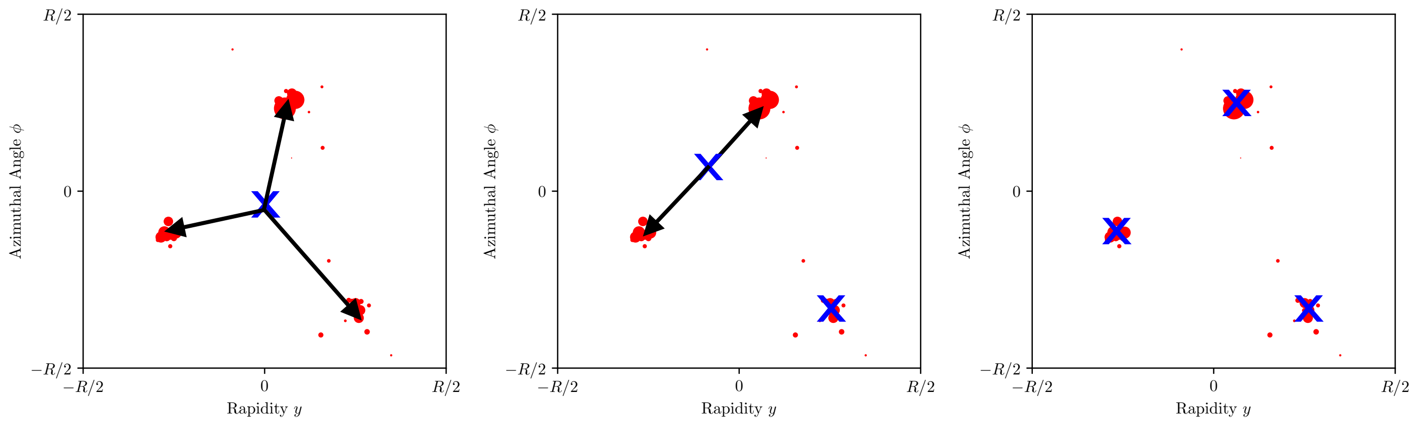
Old – Geometric Interpretation of Familiar Observables

N-subjettiness

[Thaler, Van Tilburg, 1011.2268, 1108.2701]

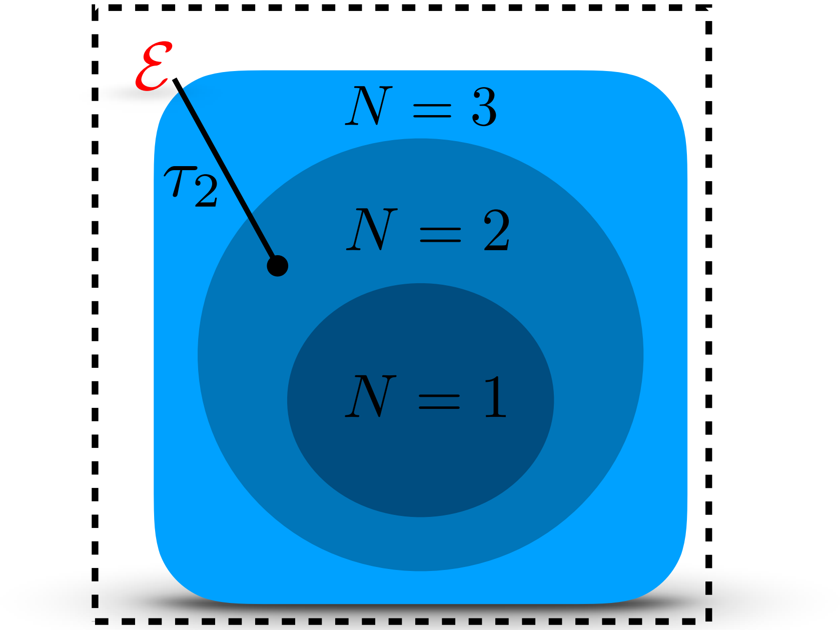
$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N \text{ axes}} \sum_i E_i \min\left(\theta_{i1}^\beta, \theta_{i2}^\beta, \ldots, \theta_{iN}^\beta\right)$$

$$\tau_1(\mathcal{E}) \quad \gg \quad \tau_2(\mathcal{E}) \quad \gg \quad \tau_3(\mathcal{E}) \simeq 0$$



β≠1 is p-Wasserstein distance with p=β

$$\tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \mathrm{EMD}(\mathcal{E}, \mathcal{E}')$$



N parton manifolds in event space

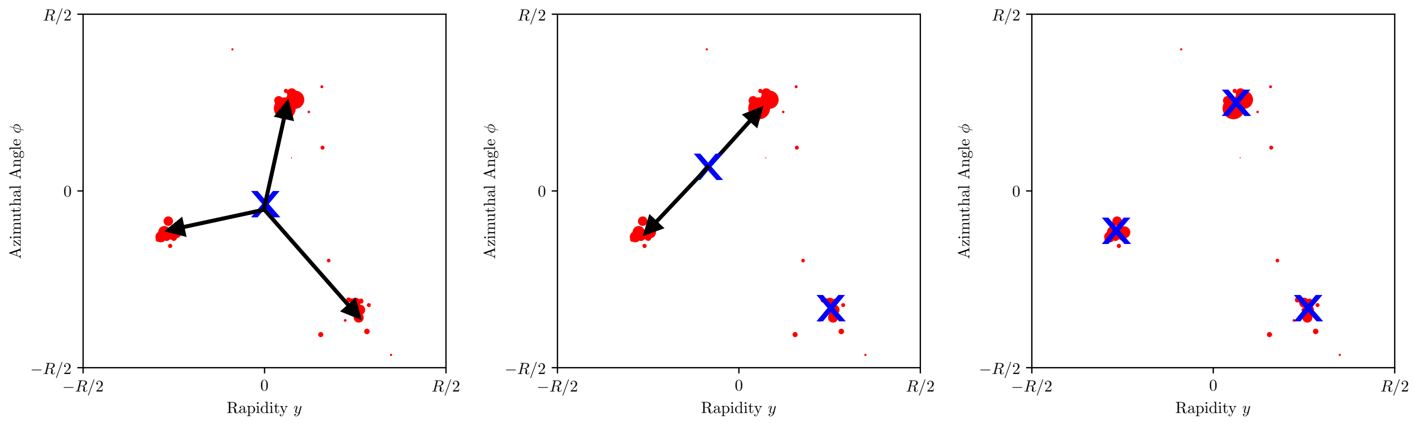
Old – Geometric Interpretation of Familiar Observables

N-subjettiness

[Thaler, Van Tilburg, 1011.2268, 1108.2701]

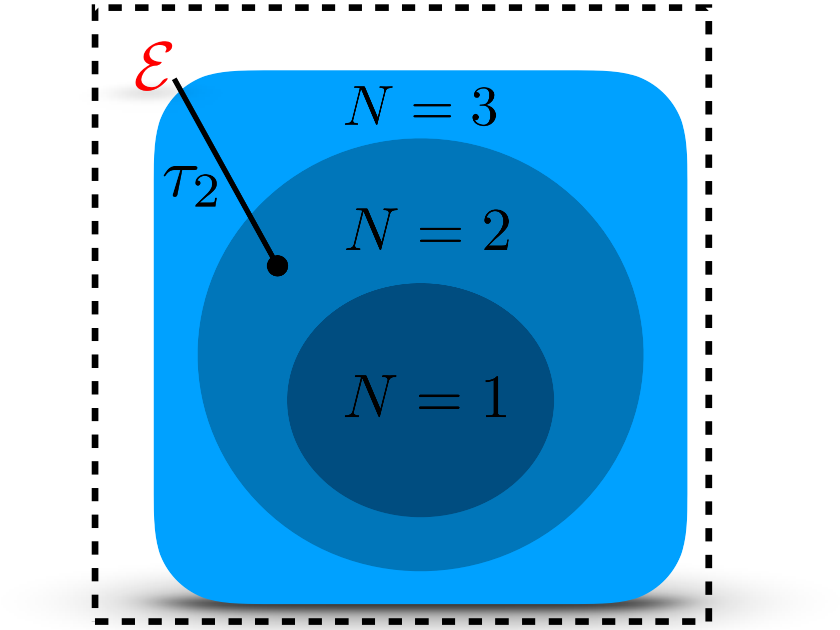
$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\ \text{axes}} \sum_i E_i \min\left(\theta_{i1}^{\beta}, \theta_{i2}^{\beta}, \dots, \theta_{iN}^{\beta}\right)$$

$$\tau_1(\mathcal{E}) \quad \gg \quad \tau_2(\mathcal{E}) \quad \gg \quad \tau_3(\mathcal{E}) \simeq 0$$



β≠1 is p-Wasserstein distance with p=β

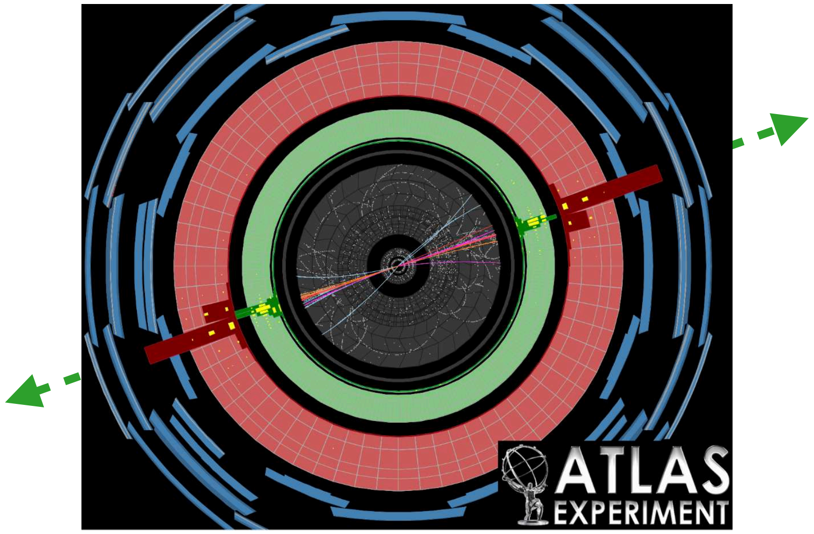
$$\tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \text{EMD}(\mathcal{E},\mathcal{E}')$$



N parton manifolds in event space

Thrust

$$t(\mathcal{E}) = 1 - \max_{\hat{n}} \sum_i |\hat{p}_i \cdot \hat{n}|, \quad \hat{p}_i = \vec{p}_i / E_i$$



[Farhi, PRL 1977]

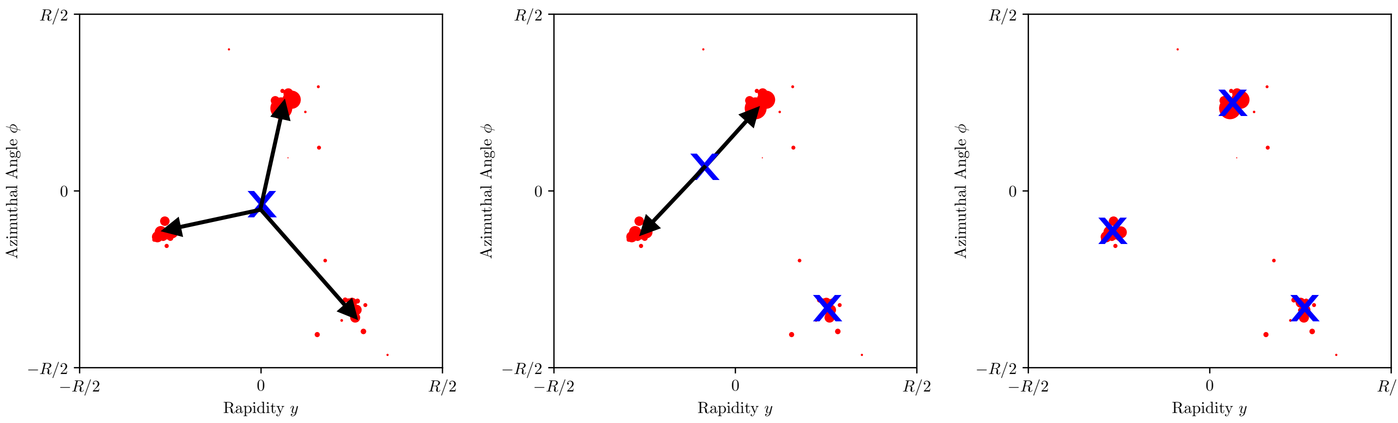
Old – Geometric Interpretation of Familiar Observables

N-subjettiness

[Thaler, Van Tilburg, 1011.2268, 1108.2701]

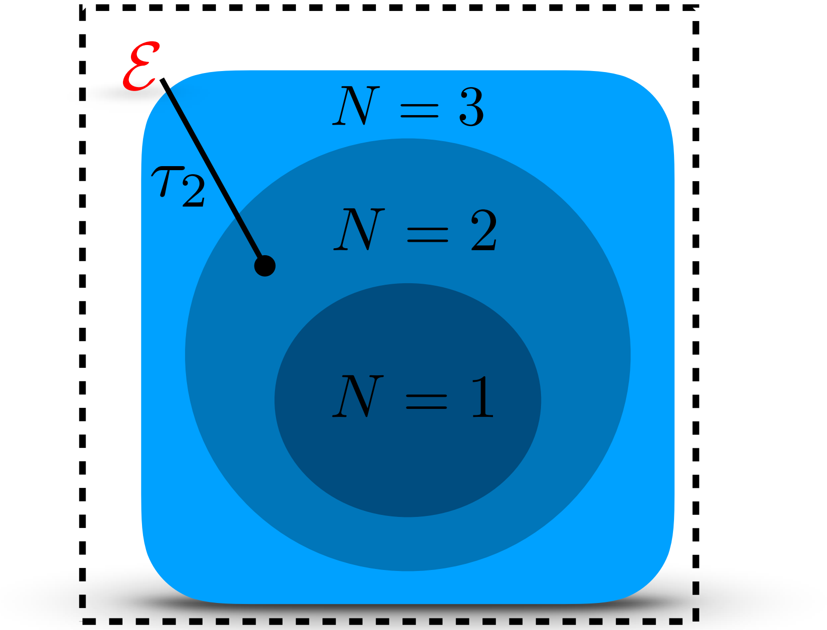
$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N\ \text{axes}} \sum_i E_i \min\left(\theta_{i1}^{\beta}, \theta_{i2}^{\beta}, \ldots, \theta_{iN}^{\beta}\right)$$

$$\tau_1(\mathcal{E}) \quad \gg \quad \tau_2(\mathcal{E}) \quad \gg \quad \tau_3(\mathcal{E}) \simeq 0$$



β≠1 is p-Wasserstein distance with p=β

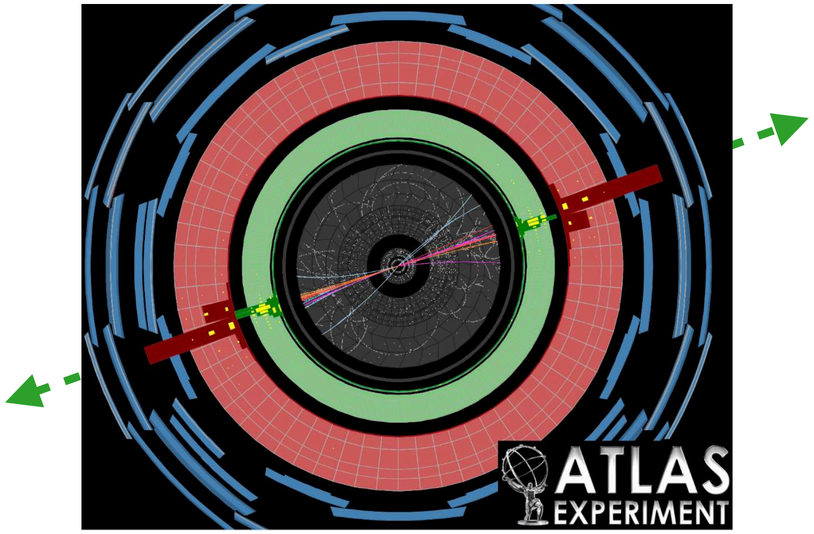
$$\tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \text{EMD}(\mathcal{E}, \mathcal{E}')$$



N parton manifolds in event space

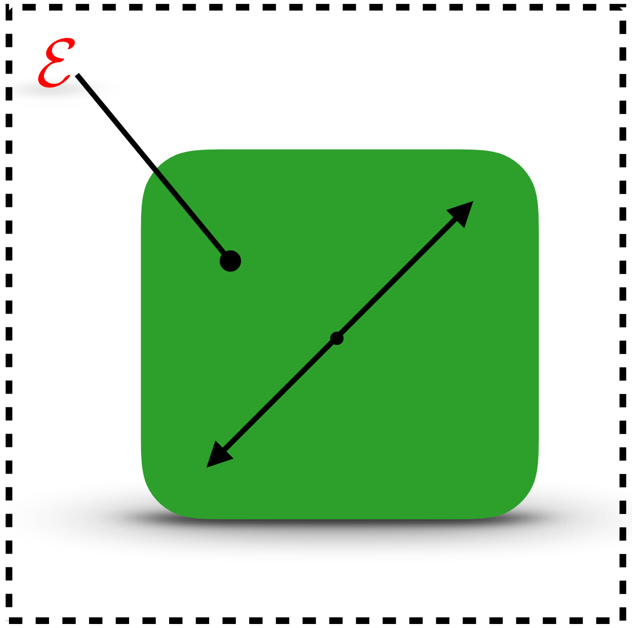
Thrust

$$t(\mathcal{E}) = 1 - \max_{\hat{n}} \sum_i |\hat{p}_i \cdot \hat{n}|\,, \qquad \hat{p}_i = \vec{p}_i / E_i$$



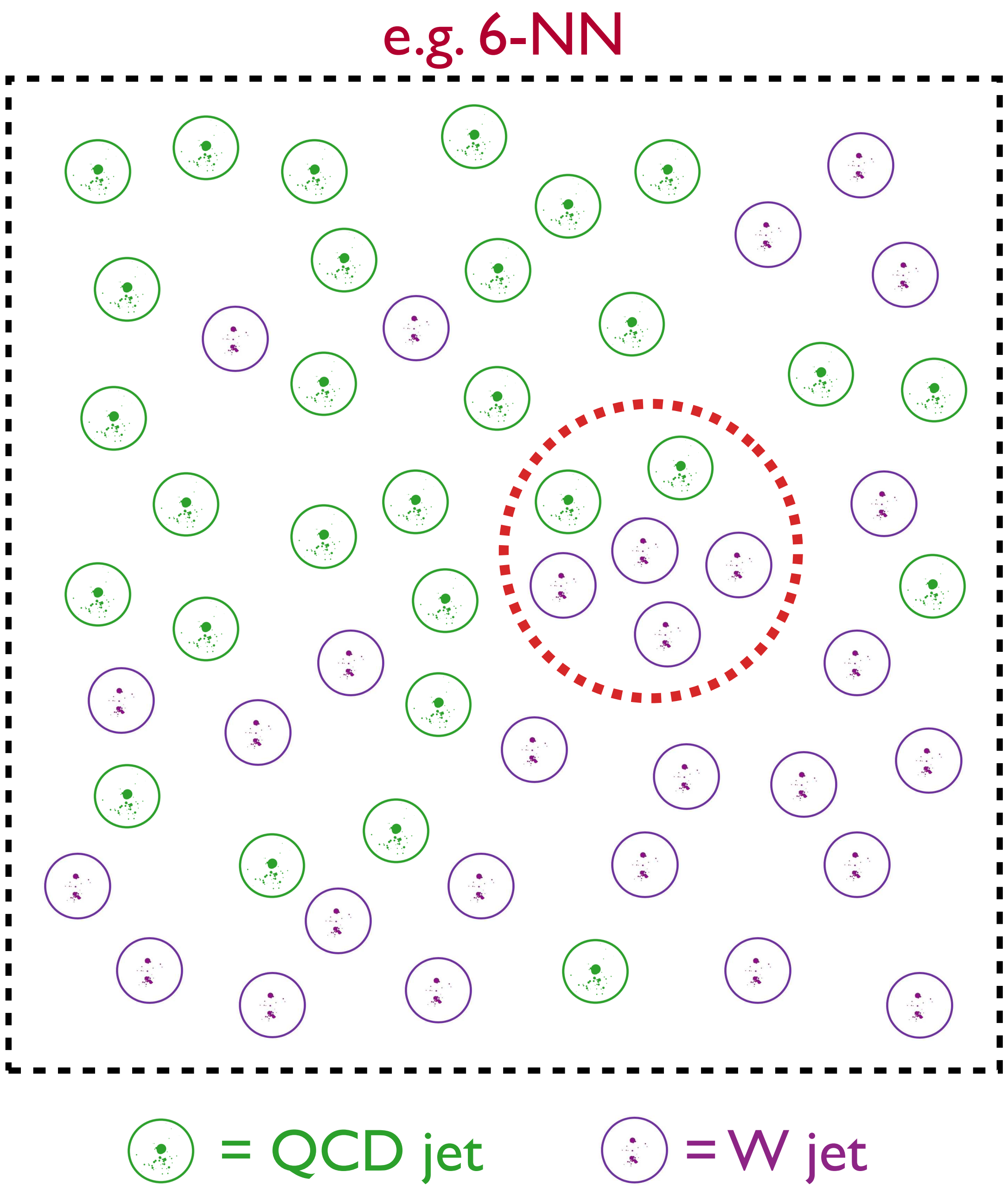
[Farhi, PRL 1977]

$$t(\mathcal{E}) = \min_{\mathcal{E}'=\updownarrow} \text{EMD}(\mathcal{E}, \mathcal{E}')$$

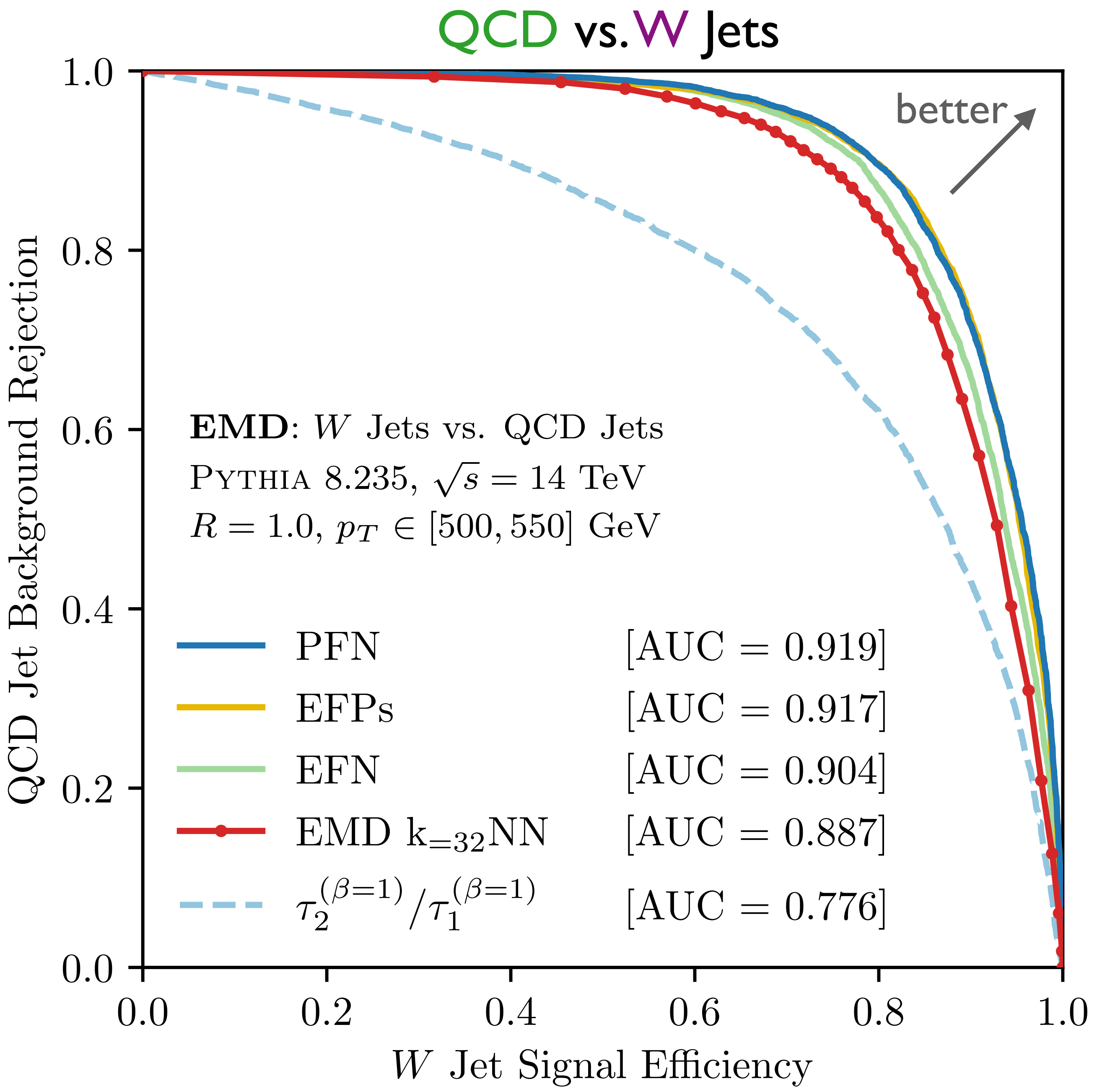


space of back-to-back configurations

Current – Jet Classification by Nearest-Neighbor Density Estimation



Abstract space of events



Approaches performance of modern machine learning

[PTK, Metodiev, Thaler, 1902.02346;
comparison to Thaler, Van Tilburg, 1011.2268, 1108.2701; PTK, Metodiev, Thaler, 1712.07124, 1810.05165;]

New – Quantifying Event Modifications

[PTK, Metodiev, Thaler, 1902.02346]

Mathematics

1-Wasserstein metric bounds the difference in expectation values between distributions

⇨

Physics

Events close in EMD are close according to IRC-safe observables

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') \geq \frac{1}{RL}\left|\sum_i E_i\Phi(\hat{p}_i) - \sum_j E'_j\Phi(\hat{p}'_j)\right| = \frac{1}{RL}\left|\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')\right|$$

↑ via Kantorovich-Rubinstein duality

↑ ↑ Additive IRC-safe observable

New – Quantifying Event Modifications

[PTK, Metodiev, Thaler, 1902.02346]

Mathematics

1-Wasserstein metric bounds the difference in expectation values between distributions

⇨

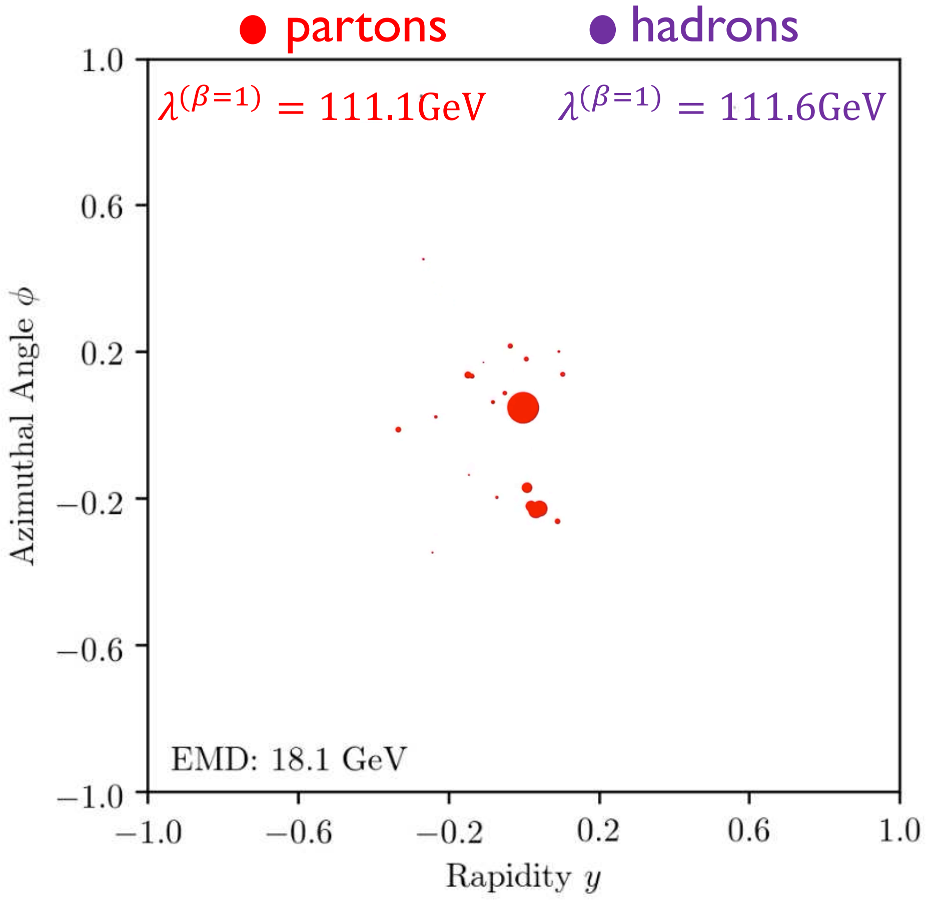
Physics

Events close in EMD are close according to IRC-safe observables

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') \underset{\uparrow}{\geq} \frac{1}{RL}\left|\sum_i E_i\Phi(\hat{p}_i) - \sum_j E'_j\Phi(\hat{p}'_j)\right| = \frac{1}{RL}\,|\underset{\uparrow}{\mathcal{O}}(\mathcal{E}) - \underset{\uparrow}{\mathcal{O}}(\mathcal{E}')|$$

via Kantorovich-Rubinstein duality

Additive IRC-safe observable

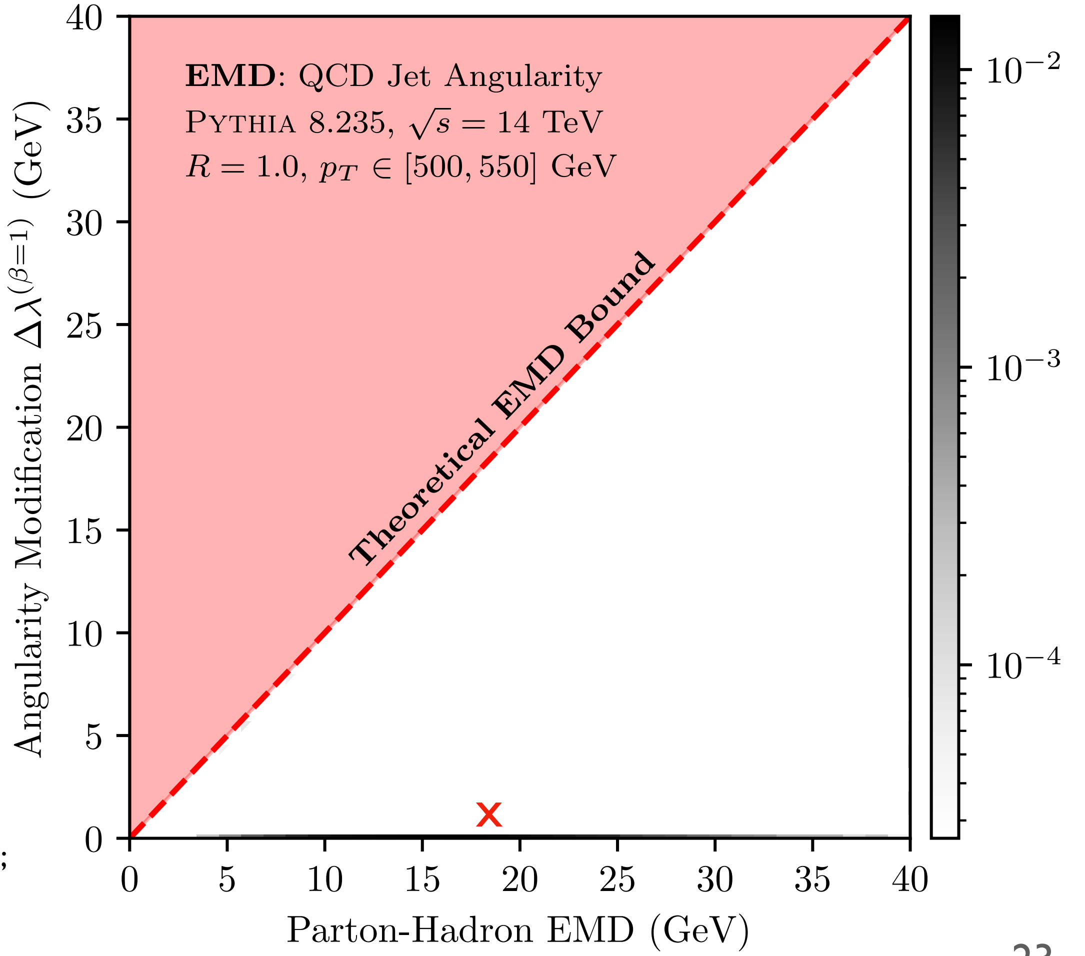


e.g. bounding IRC-safe angularities

$$\lambda^{(\beta)}(\mathcal{E}) = \sum_i E_i\theta_i^\beta$$

Can do the same for pileup, detector effects

[Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014; Larkoski, Thaler, Waalewijn, 1408.3122]



New – Quantifying Event Modifications

[PTK, Metodiev, Thaler, 1902.02346]

Mathematics

1-Wasserstein metric bounds the difference in expectation values between distributions

⇒

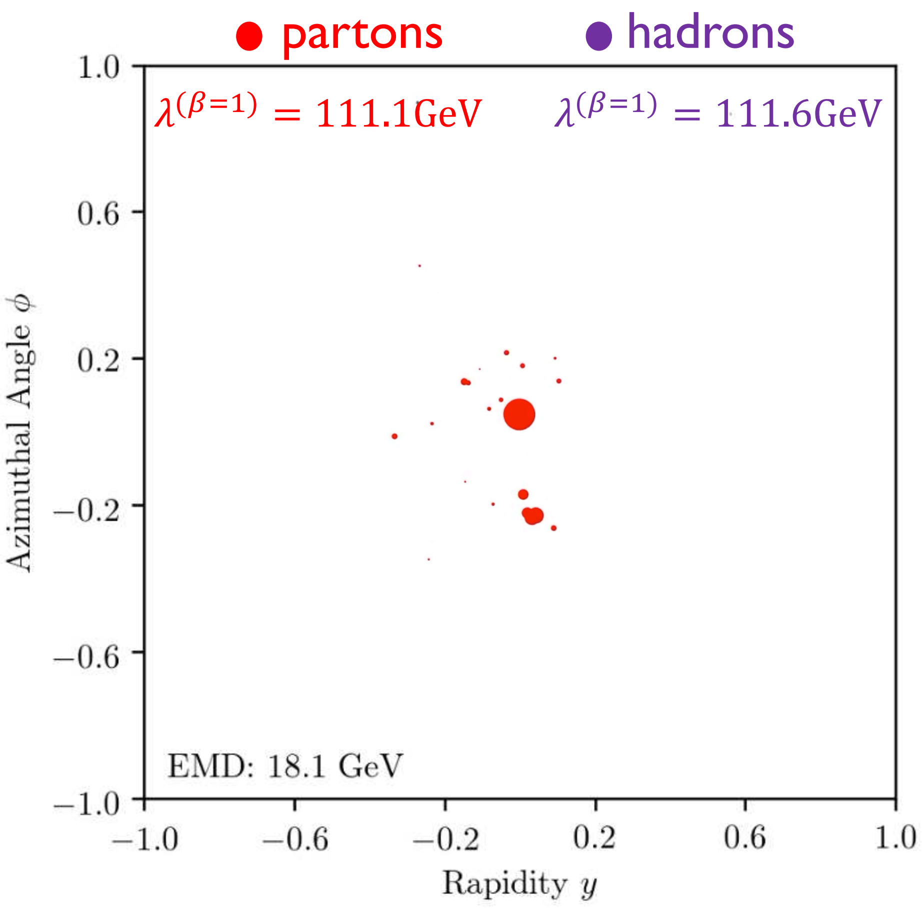
Physics

Events close in EMD are close according to IRC-safe observables

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') \underset{\uparrow}{\geq} \frac{1}{RL}\left|\sum_i E_i\Phi(\hat{p}_i) - \sum_j E'_j\Phi(\hat{p}'_j)\right| = \frac{1}{RL}\,|\underset{\uparrow}{\mathcal{O}}(\mathcal{E}) - \underset{\uparrow}{\mathcal{O}}(\mathcal{E}')|$$

via Kantorovich-Rubinstein duality

Additive IRC-safe observable

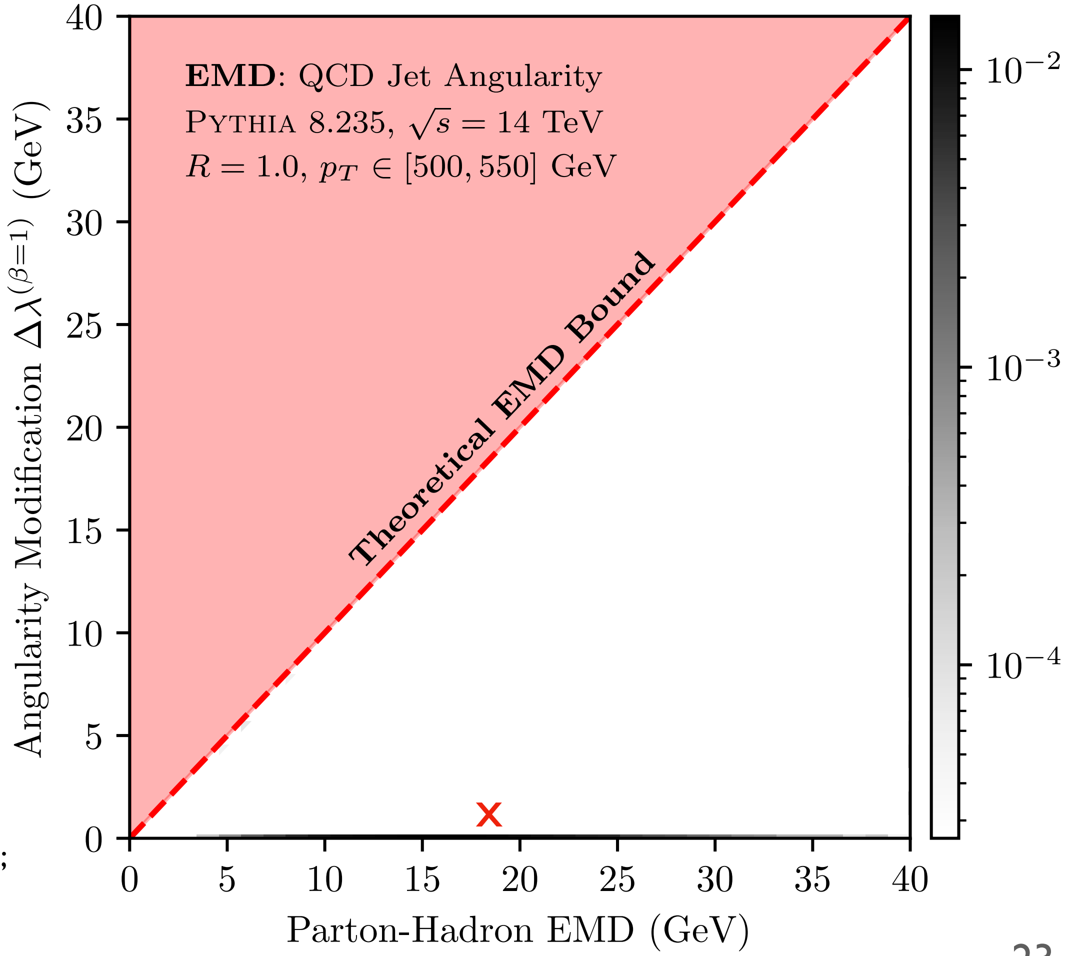


e.g. bounding IRC-safe angularities

$$\lambda^{(\beta)}(\mathcal{E}) = \sum_i E_i\theta_i^\beta$$

Can do the same for pileup, detector effects

[Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014; Larkoski, Thaler, Waalewijn, 1408.3122]



New – Quantifying Event Modifications

[PTK, Metodiev, Thaler, 1902.02346]

Mathematics

1-Wasserstein metric bounds the difference in expectation values between distributions

⇨

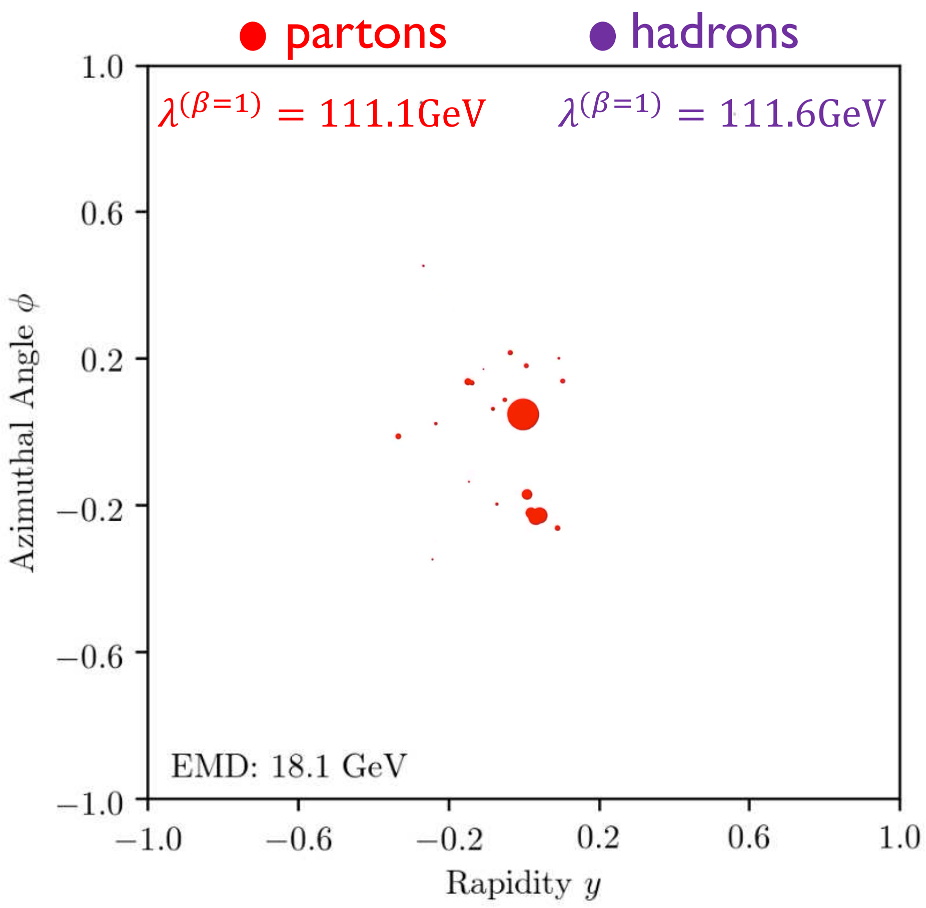
Physics

Events close in EMD are close according to IRC-safe observables

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') \underset{\uparrow}{\geq} \frac{1}{RL}\left|\sum_i E_i\Phi(\hat{p}_i) - \sum_j E'_j\Phi(\hat{p}'_j)\right| = \frac{1}{RL}\left|\underset{\uparrow}{\mathcal{O}}(\mathcal{E}) - \underset{\uparrow}{\mathcal{O}}(\mathcal{E}')\right|$$

via Kantorovich-Rubinstein duality

Additive IRC-safe observable

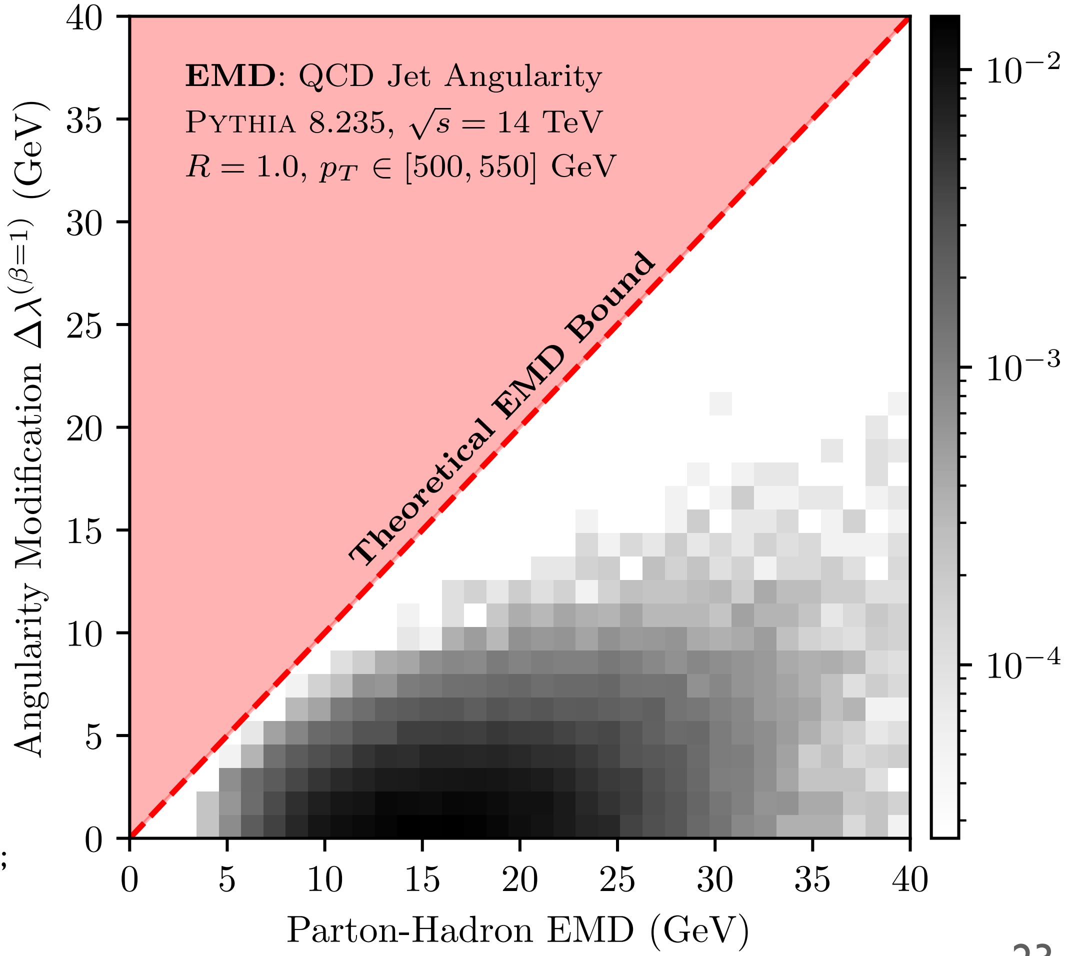


e.g. bounding IRC-safe angularities

$$\lambda^{(\beta)}(\mathcal{E}) = \sum_i E_i \theta_i^\beta$$

Can do the same for pileup, detector effects

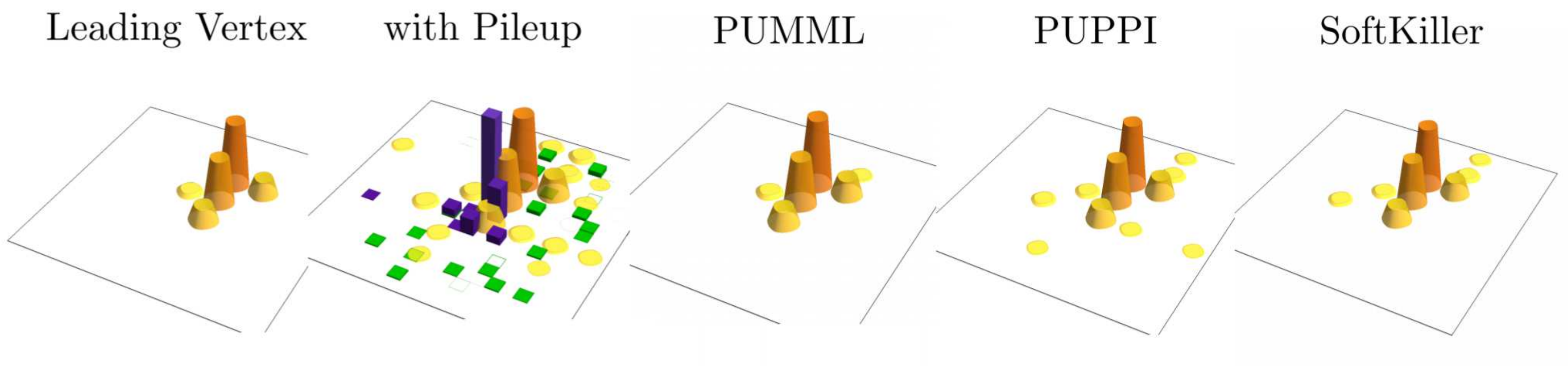
[Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014; Larkoski, Thaler, Waalewijn, 1408.3122]



Future – Optimizing Pileup Removal

[PTK, Metodiev, Nachman, Schwartz, 1707.08600]

PileUp Mitigation with Machine Learning (PUMML)

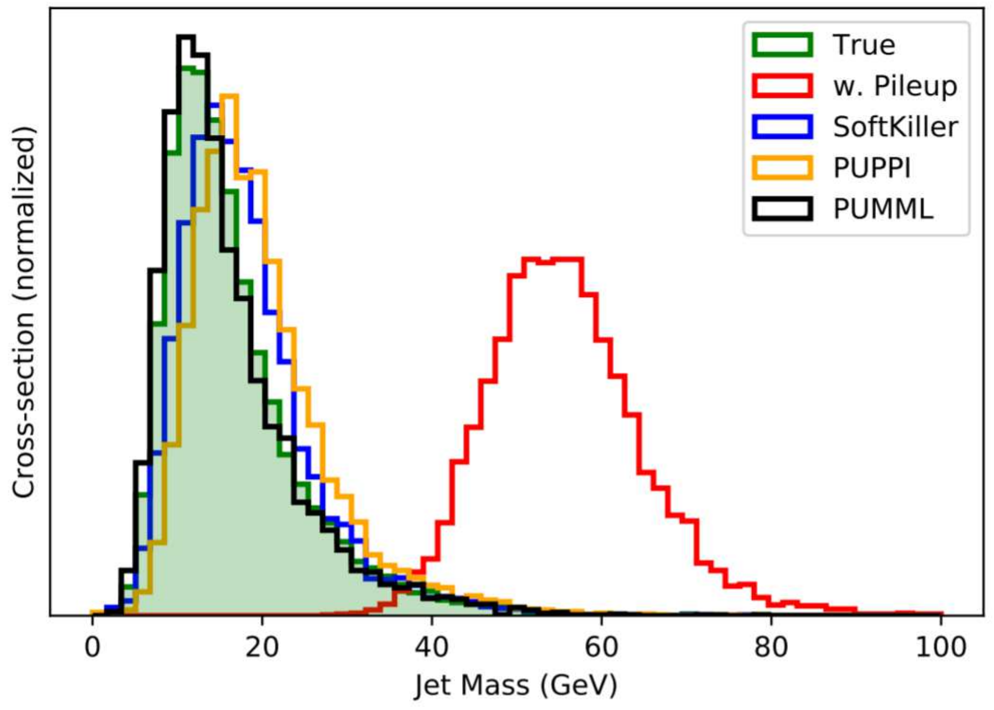


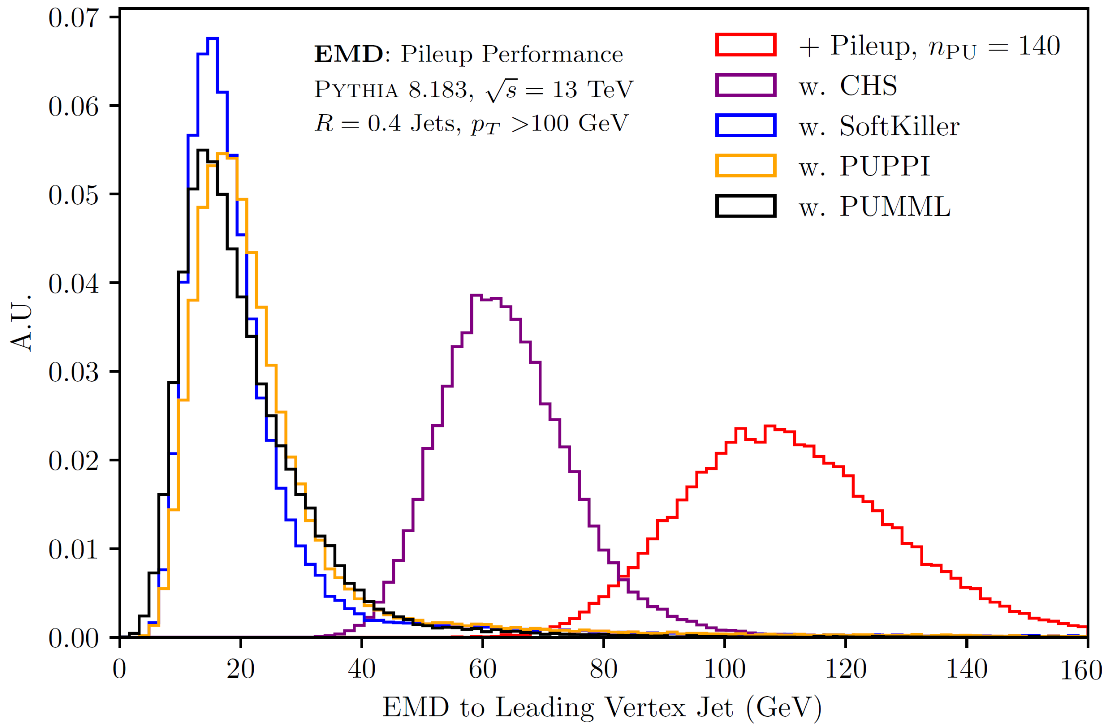
PUMML with jet images

- *pixel-based custom loss function*
- *compared specific IRC-safe observables*

PUMML with EMD?

- *no pixelation, automatic loss function*
- *related to all IRC-safe observables*



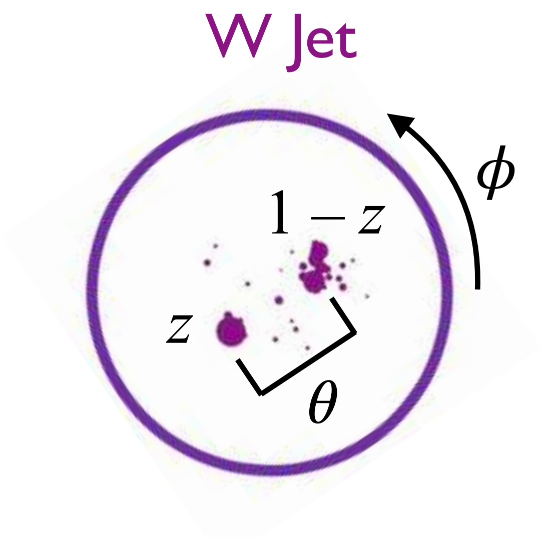


(Event) Space Exploration

[Khanna, phys.org]

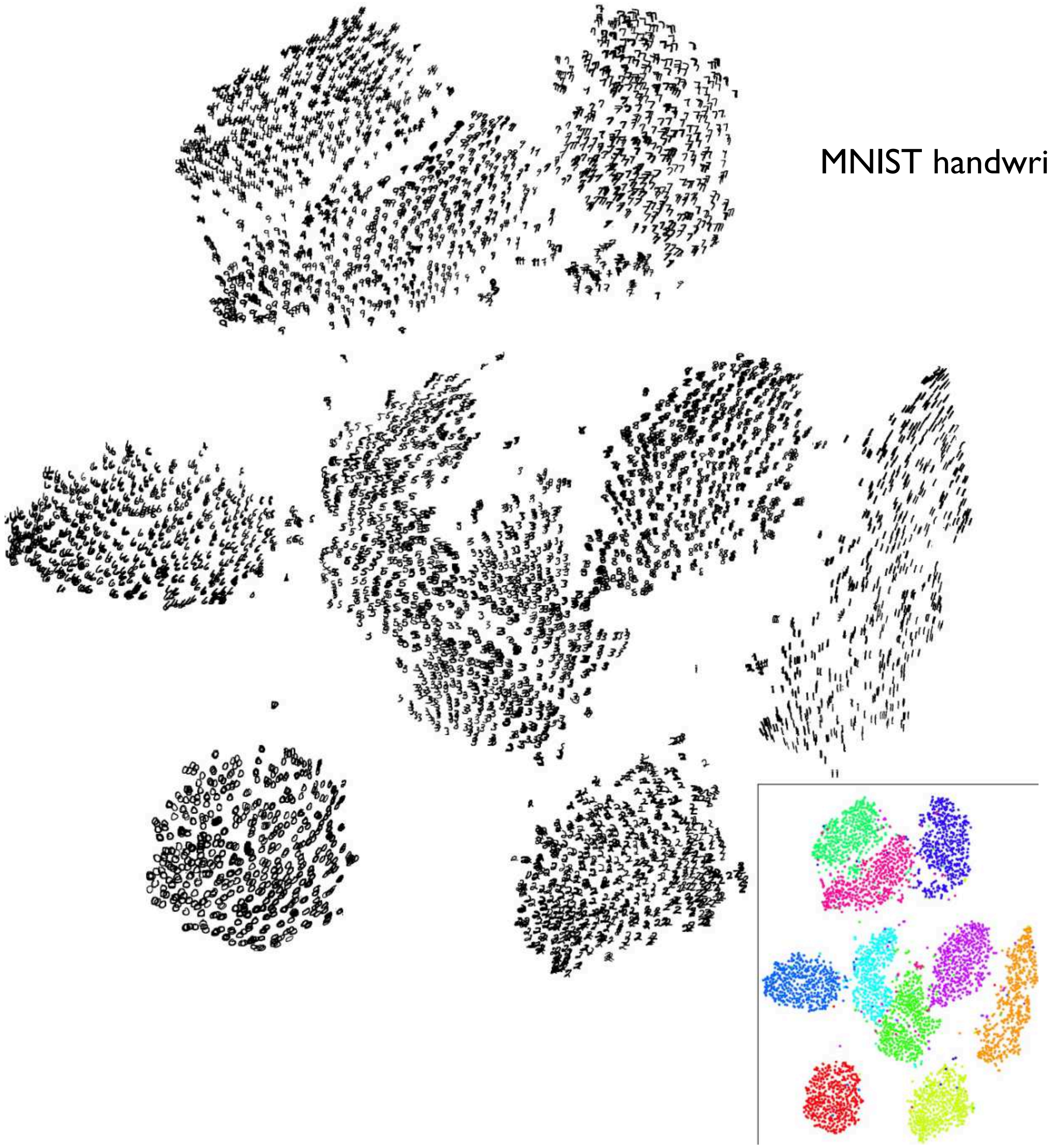
Visualizing the Metric Space of *W* Jets

Embed high-dimension manifold in low-dimensional space?



Constraints: *W* Mass and ϕ = 0 preprocessing

t-Distributed Stochastic Neighbor Embedding

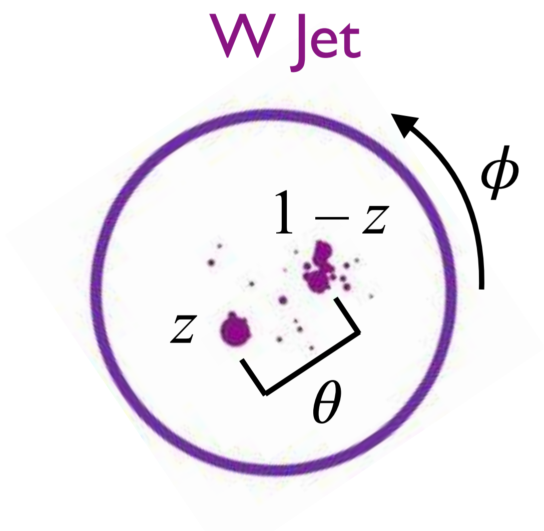


MNIST handwritten digits

[L. van der Maaten, G. Hinton, JMLR 2008]

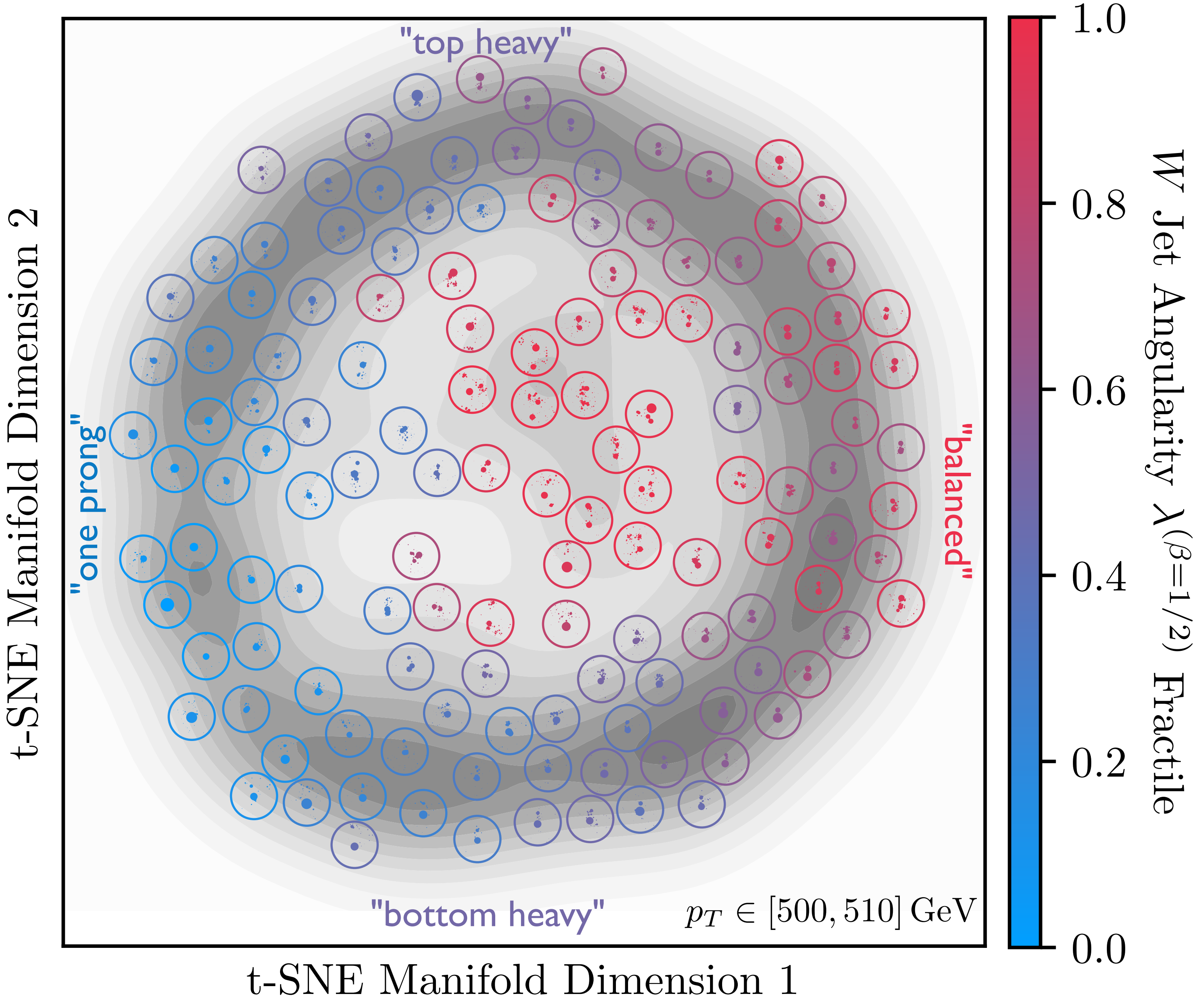
Visualizing the Metric Space of W Jets

Embed high-dimension manifold in low-dimensional space?



Constraints: W Mass and ϕ = 0 preprocessing

t-Distributed Stochastic Neighbor Embedding



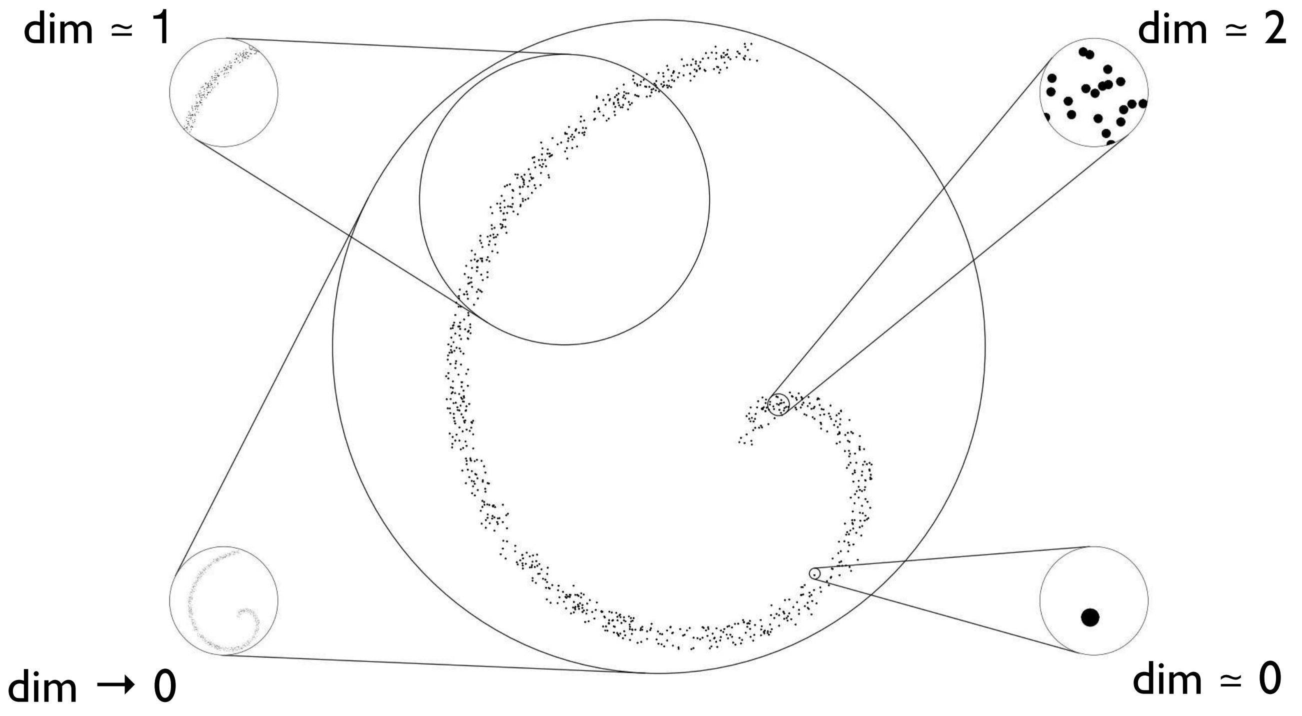
Gray contours represent the density of jets
Each circle is a particular W jet

[PTK, Metodiev, Thaler, 1902.02346]

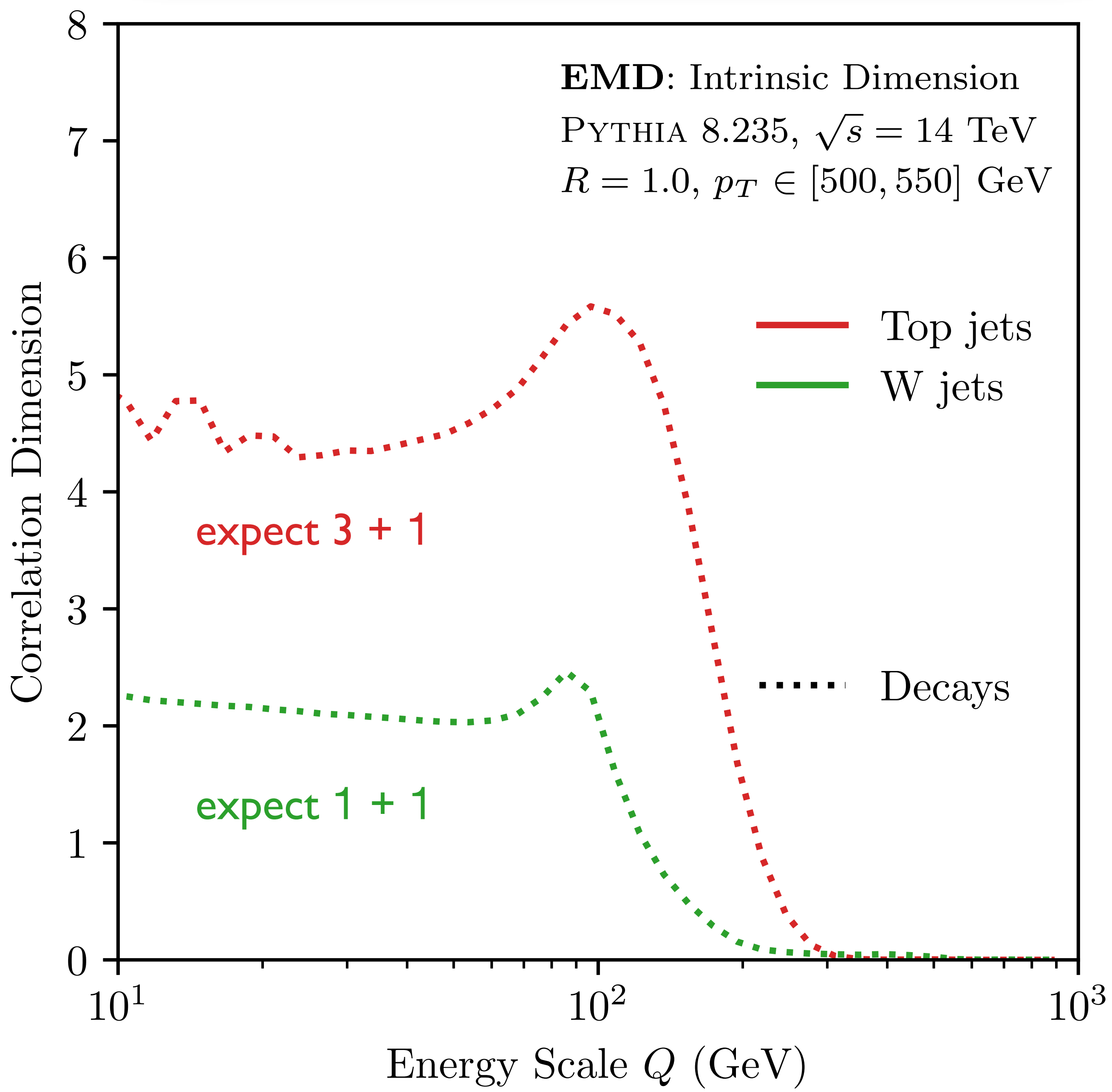
Manifold Dimensions of Event Space

Correlation dimension: *how does the # of elements within a ball of size Q change?*

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j\Theta(\text{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

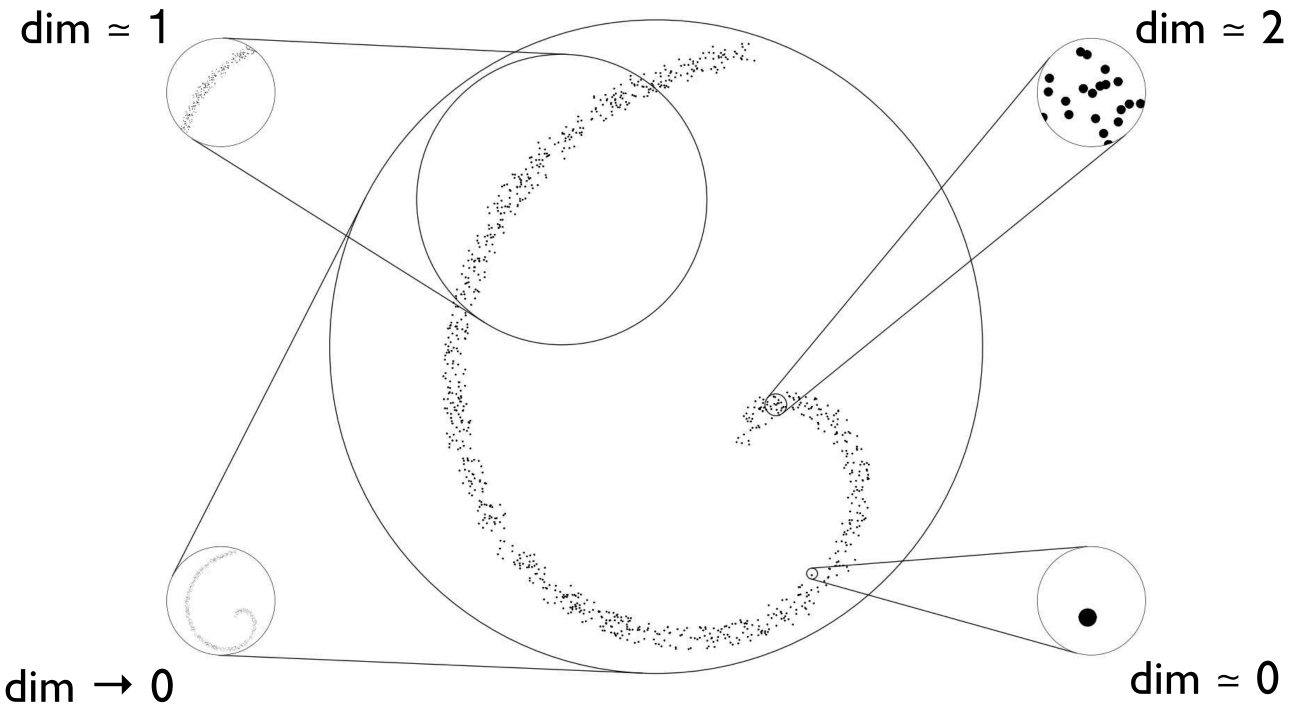
Decays are "constant" dim. at low *Q*

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

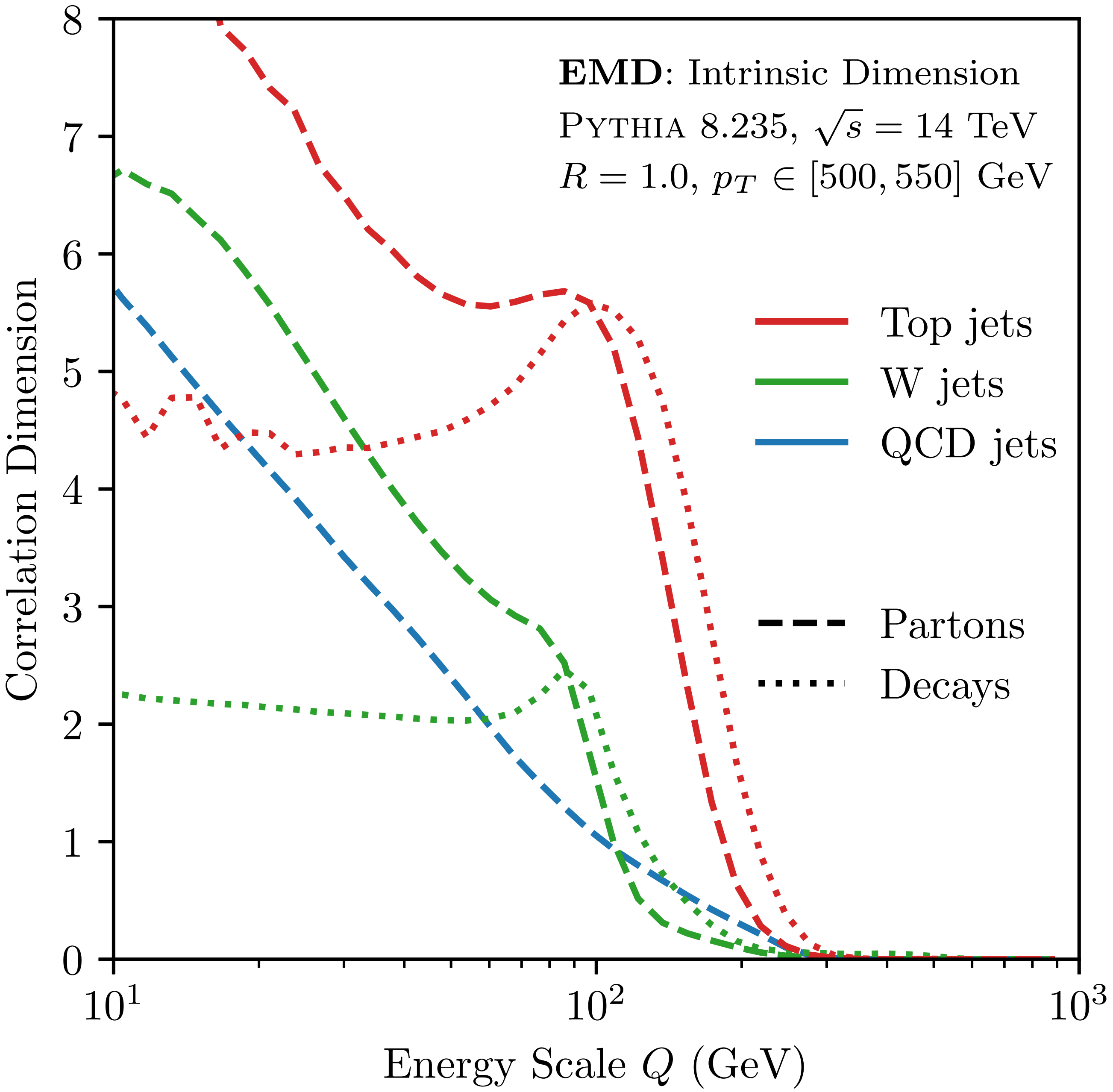
Manifold Dimensions of Event Space

Correlation dimension: *how does the # of elements within a ball of size Q change?*

$$\boxed{\dim(Q) = Q\frac{\partial}{\partial Q}\ln \sum_i \sum_j \Theta(\text{EMD}(\mathcal{E}_i, \mathcal{E}'_j) < Q)}$$



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

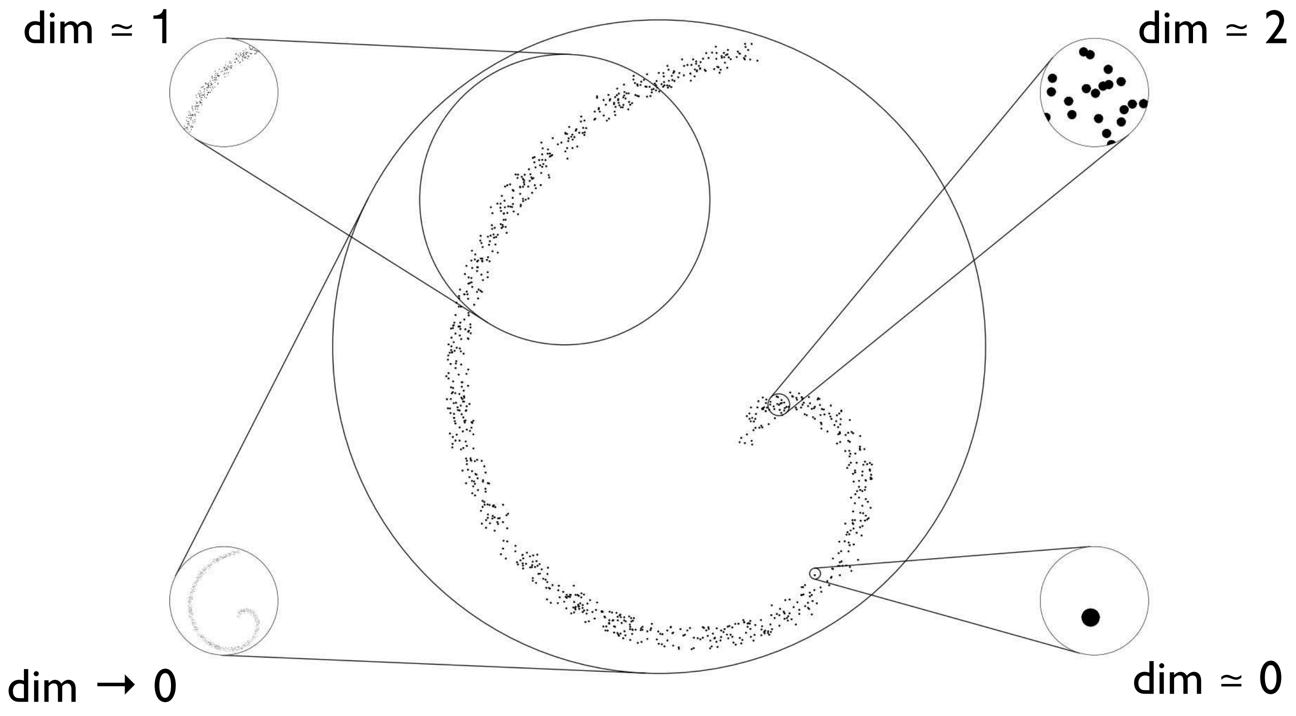
Decays are "constant" dim. at low Q
Complexity hierarchy: QCD < W < Top
Fragmentation increases dim. at smaller scales

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

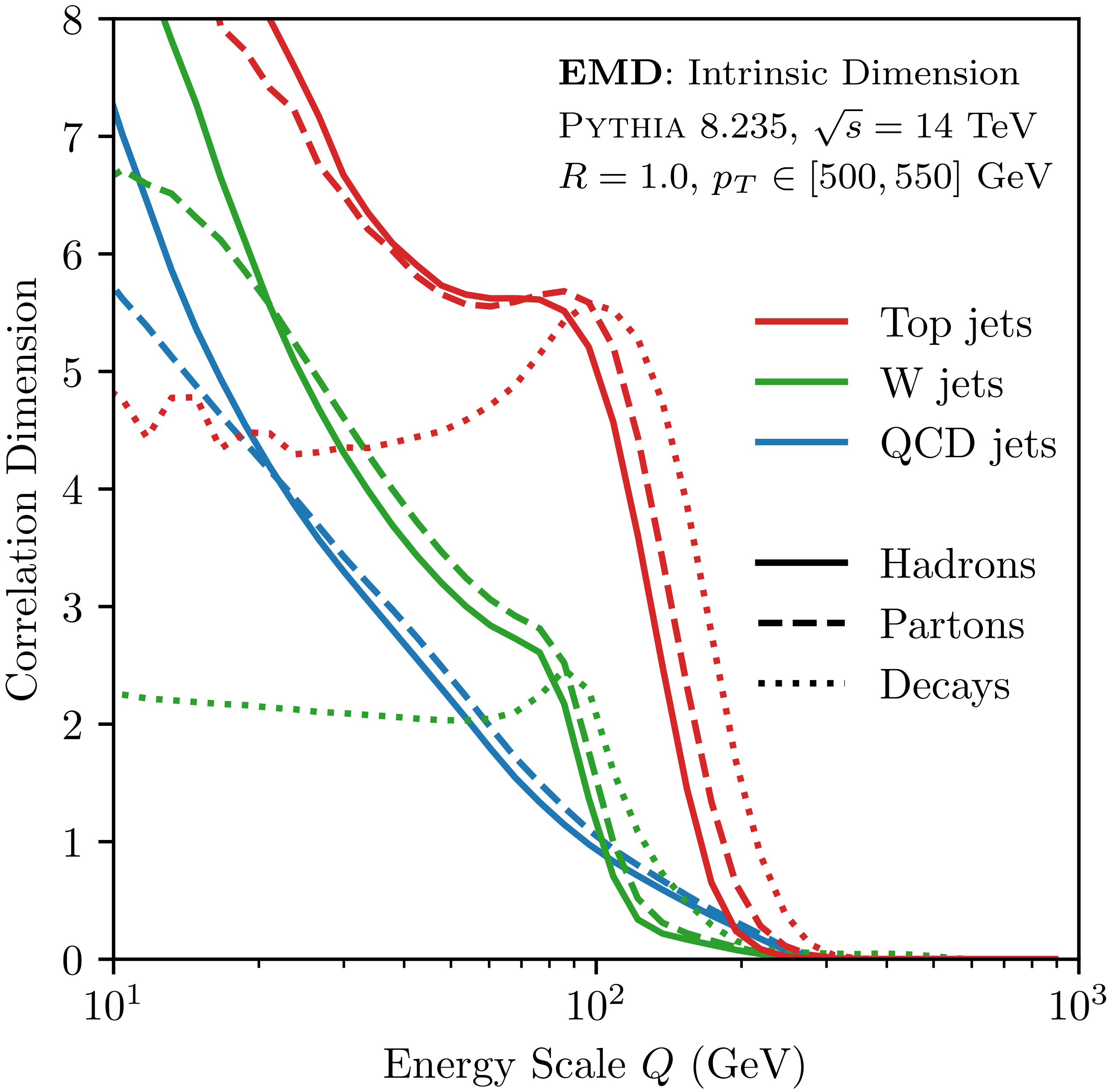
Manifold Dimensions of Event Space

Correlation dimension: *how does the # of elements within a ball of size Q change?*

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

- Decays are "constant" dim. at low Q
- Complexity hierarchy: QCD < W < Top
- Fragmentation increases dim. at smaller scales
- Hadronization important around 20-30 GeV

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

Quark and Gluon Correlation Dimensions

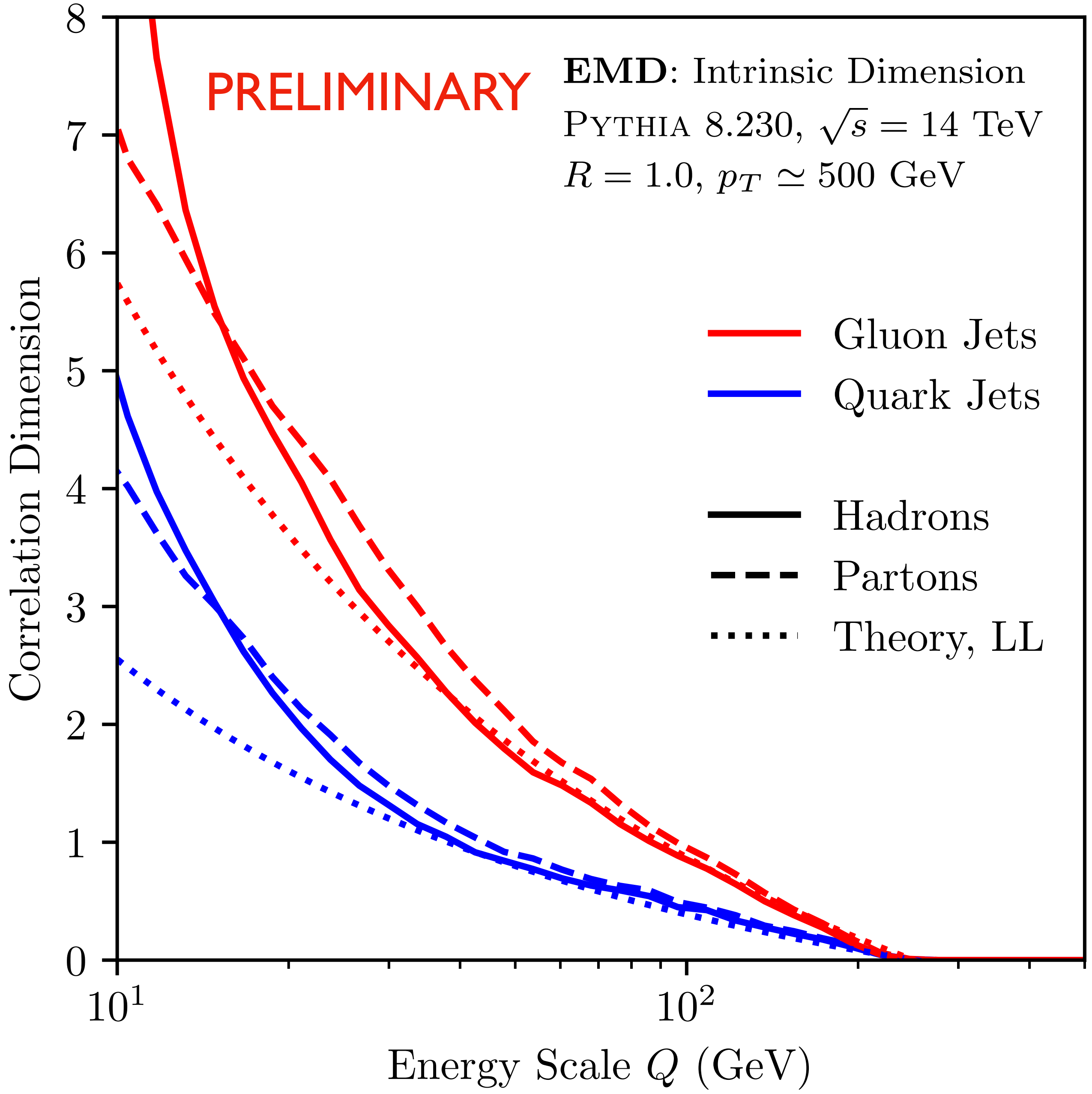
$$\boxed{\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_{i}\sum_{j}\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)}$$

Leading log (single emission) calculation:

$$\dim_i(Q) \simeq -\frac{8\alpha_s}{\pi}\underset{\substack{\uparrow \\ \text{color factor}}}{C_i}\ln\frac{Q}{p_T/2}$$

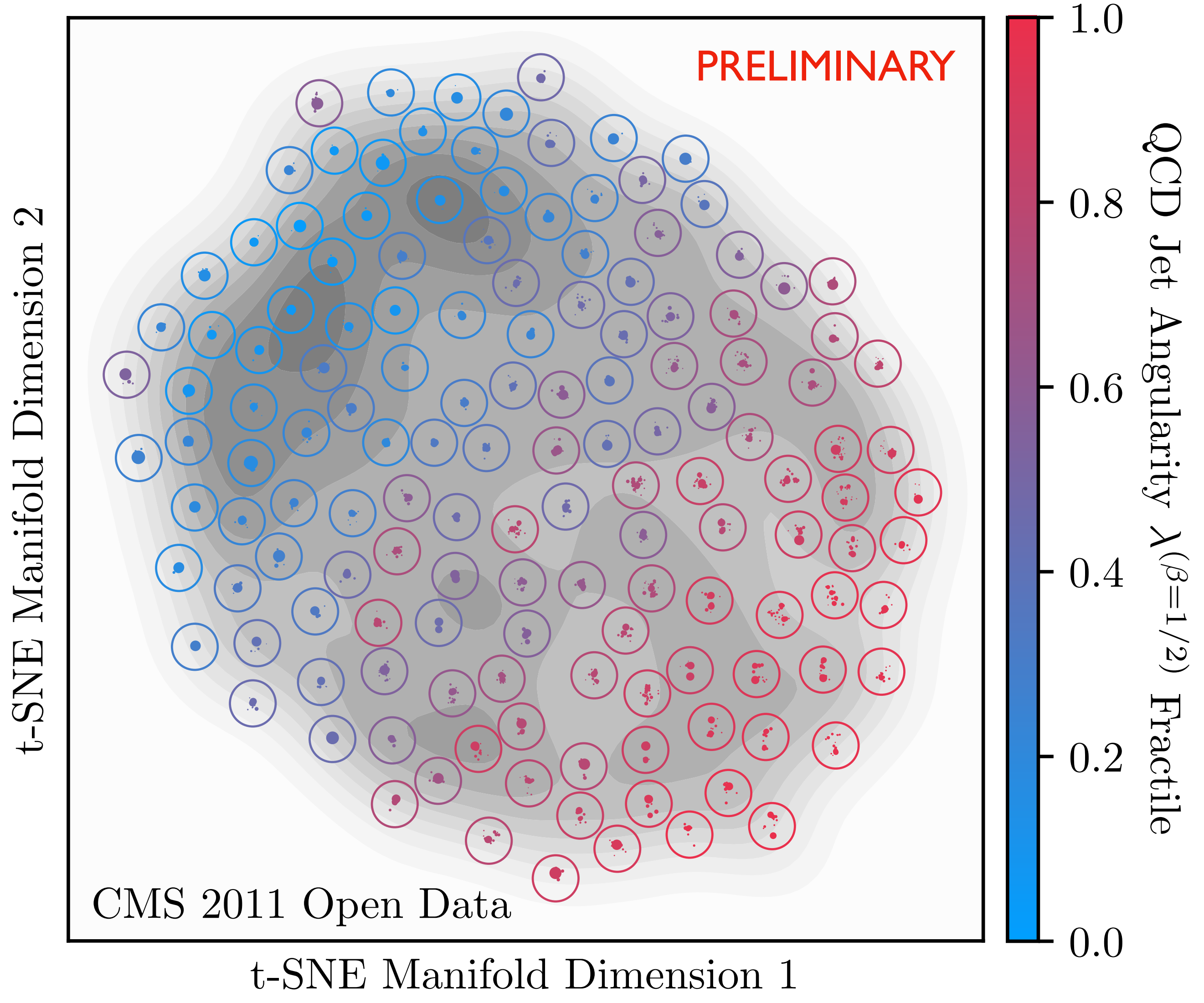
C_F = 4/3

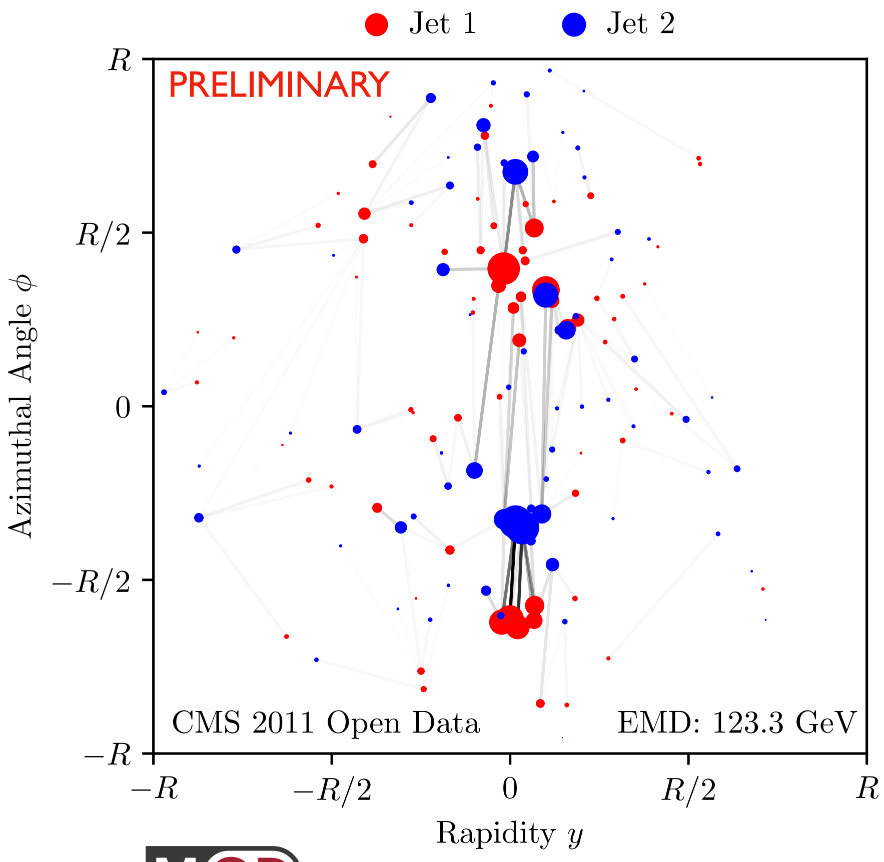
C_A = 3

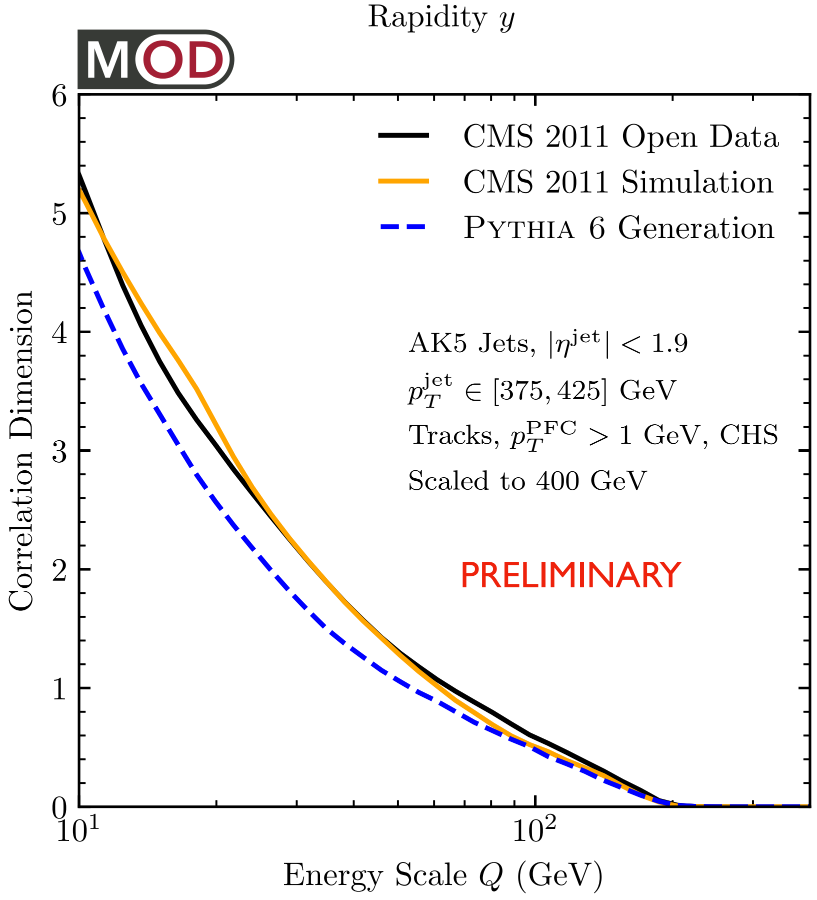


Visualizing Jets with CMS Open Data

Application of EMD techniques to the jet primary dataset in CMS 2011 Open Data





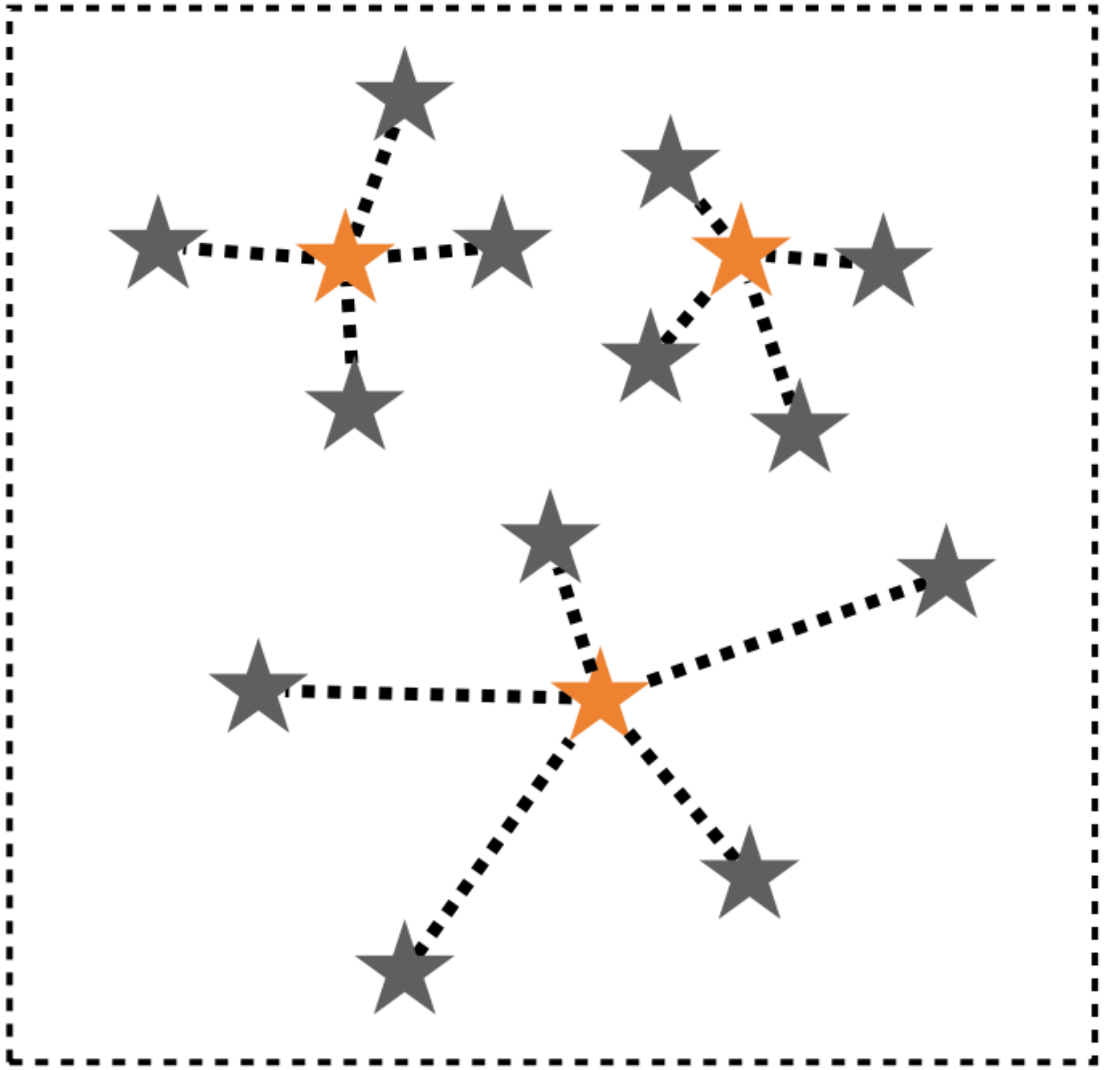


[Mastandrea, Naik, PTK, Metodiev, Thaler, *in progress*]

Identifying Representative Jets

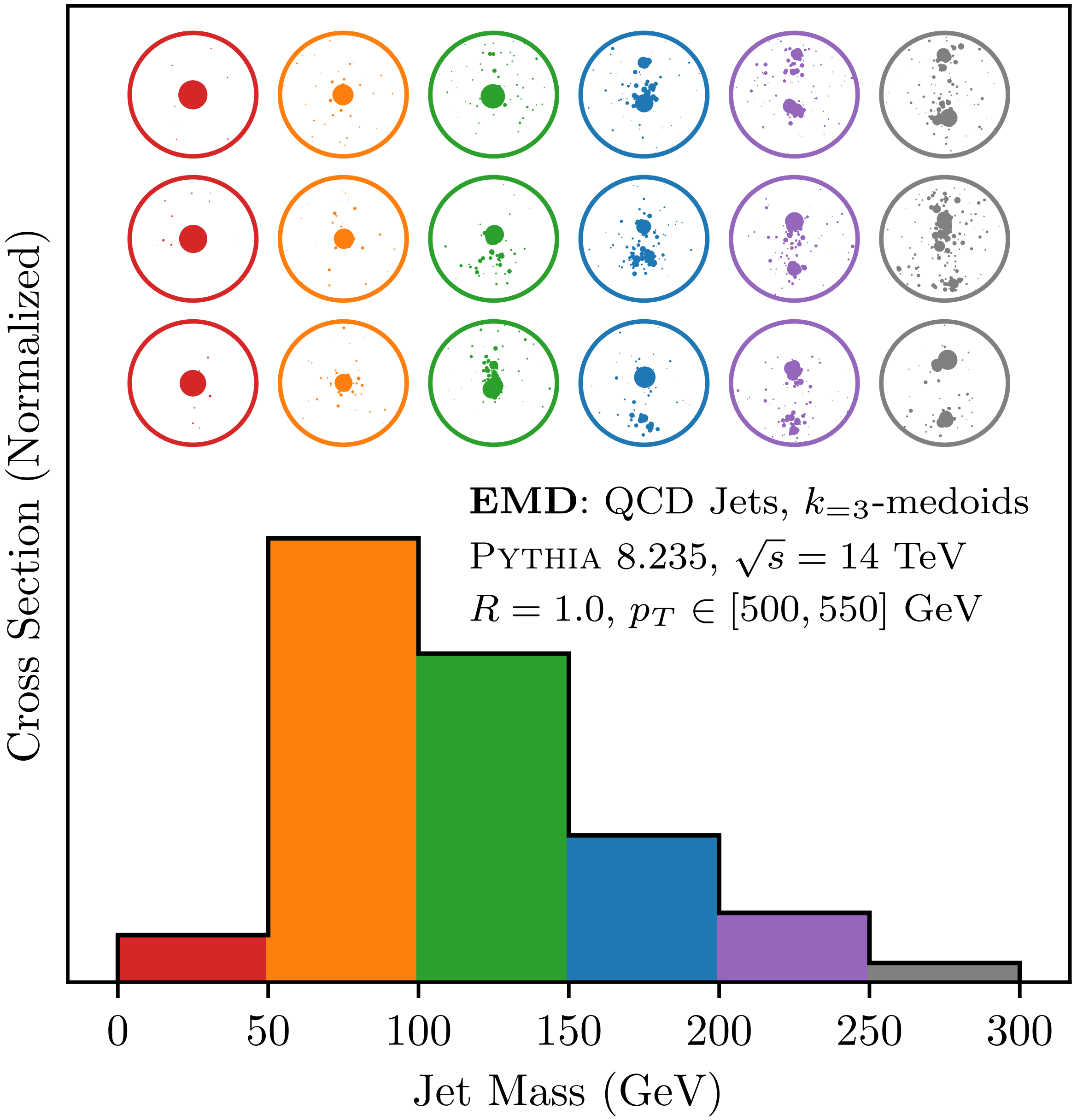
medoid: element selected to best represent a set of elements

k-medoids: k clusters to minimize total distance of points to medoids



3-medoid

[diagram by Jesse Thaler]

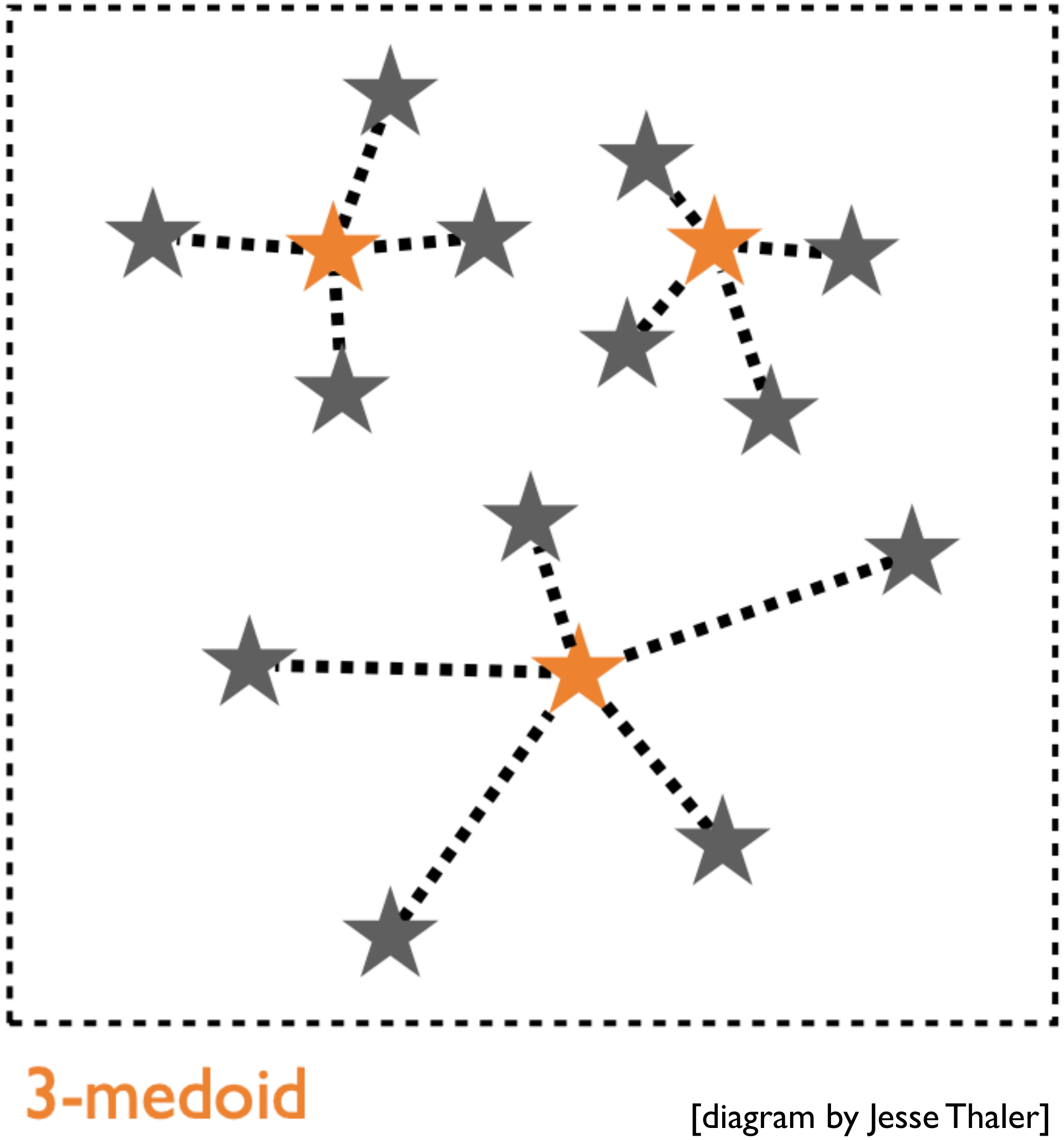


[PTK, Metodiev, Thaler, 1902.02346]

Identifying Representative Jets

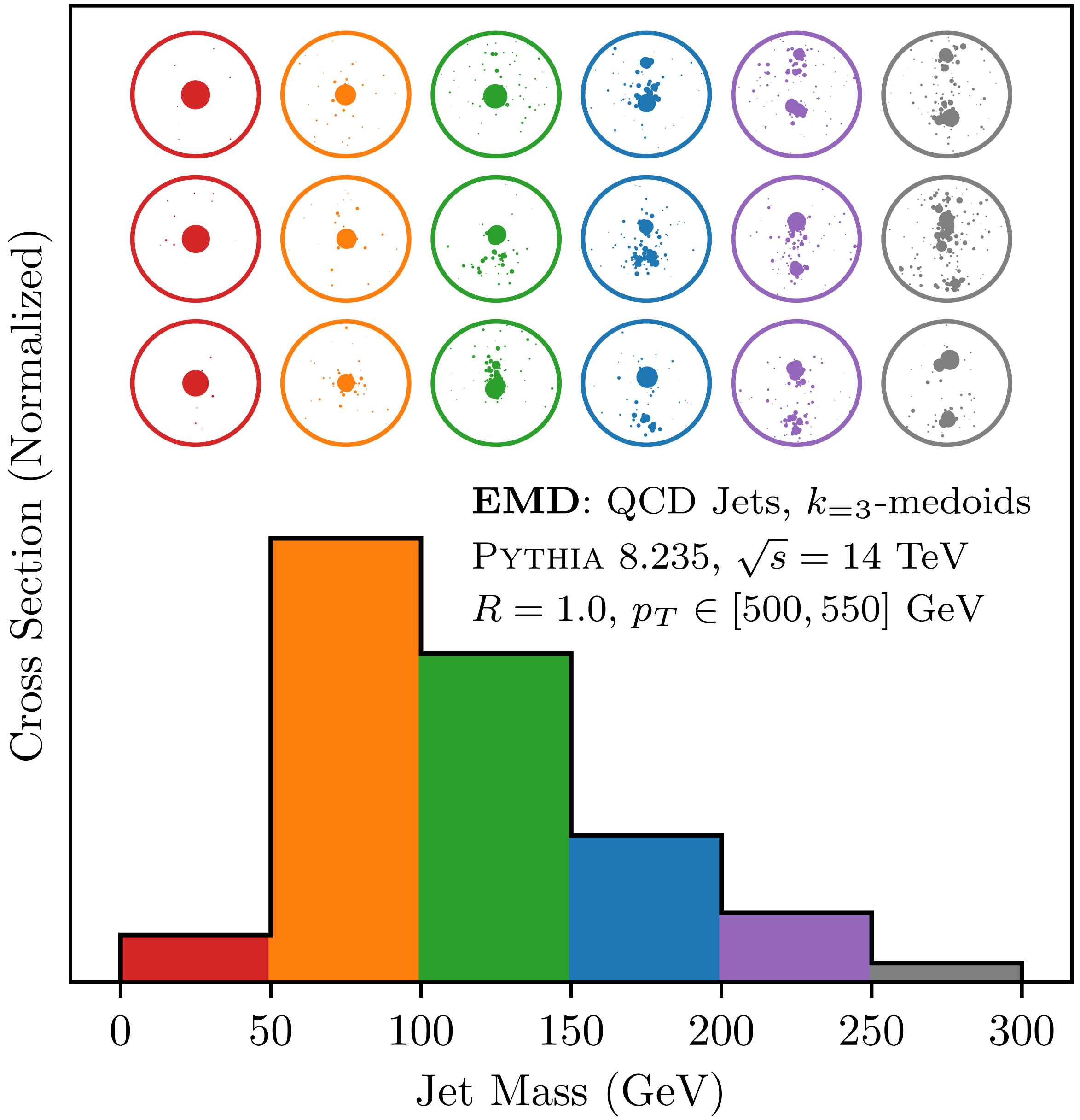
medoid: element selected to best represent a set of elements

k-medoids: k clusters to minimize total distance of points to medoids

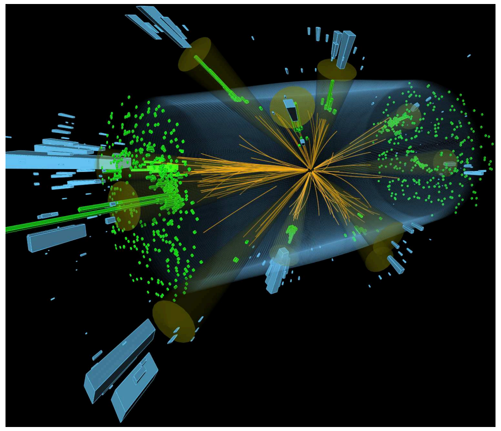


[diagram by Jesse Thaler]

Three "most" representative jets in each bin

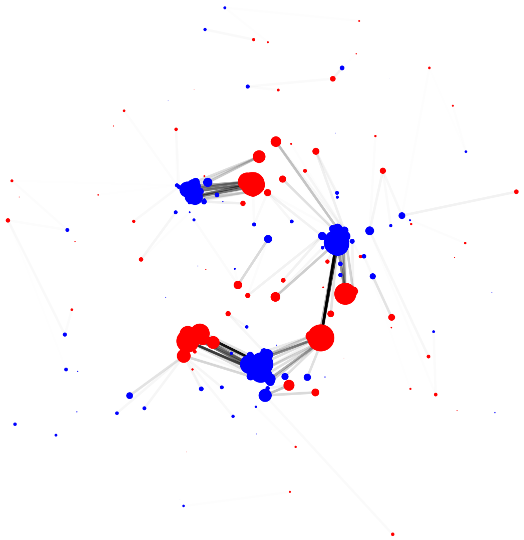


[PTK, Metodiev, Thaler, 1902.02346]



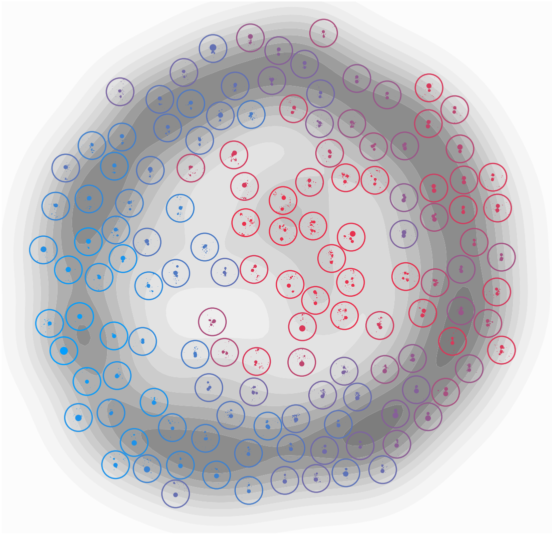
When are two events similar?

IRC-safe energy flow is theoretically and experimentally robust



The Energy Mover's Distance

Quantifies the difference in energy flow between events



Particle Physics Applications

Classification, quantifying modifications, understanding observables, exploring and visualizing event space

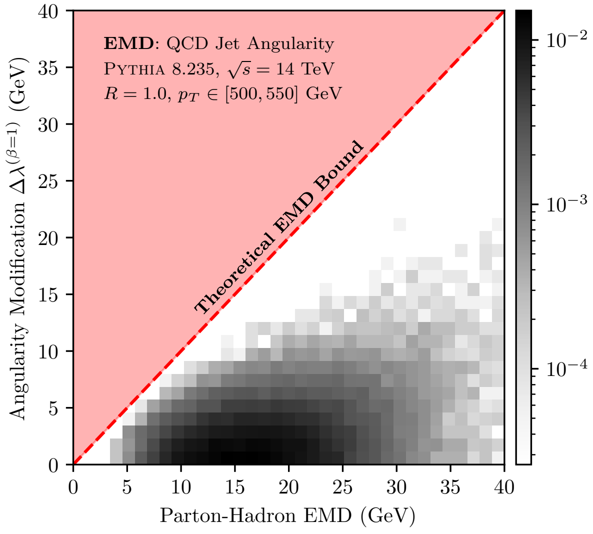
Further Directions

Experimental

Quantify (or even mitigate?) pileup/detector effects

Non-parametric density estimates (unfolding?)

Automated data compression (triggering?)

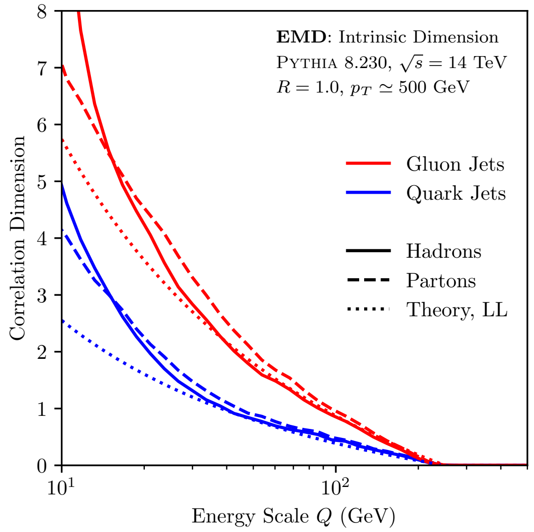


Theoretical

Define new observables with EMD?

Precision QCD calculations of event space geometry?

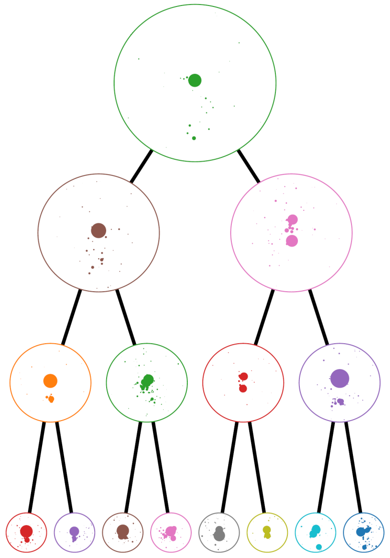
Event Mover's Distance between ensembles?



Algorithmic

Loss function for modern ML in particle physics?

Metric trees to turn $O(N^2)$ into $O(N \log N)$?



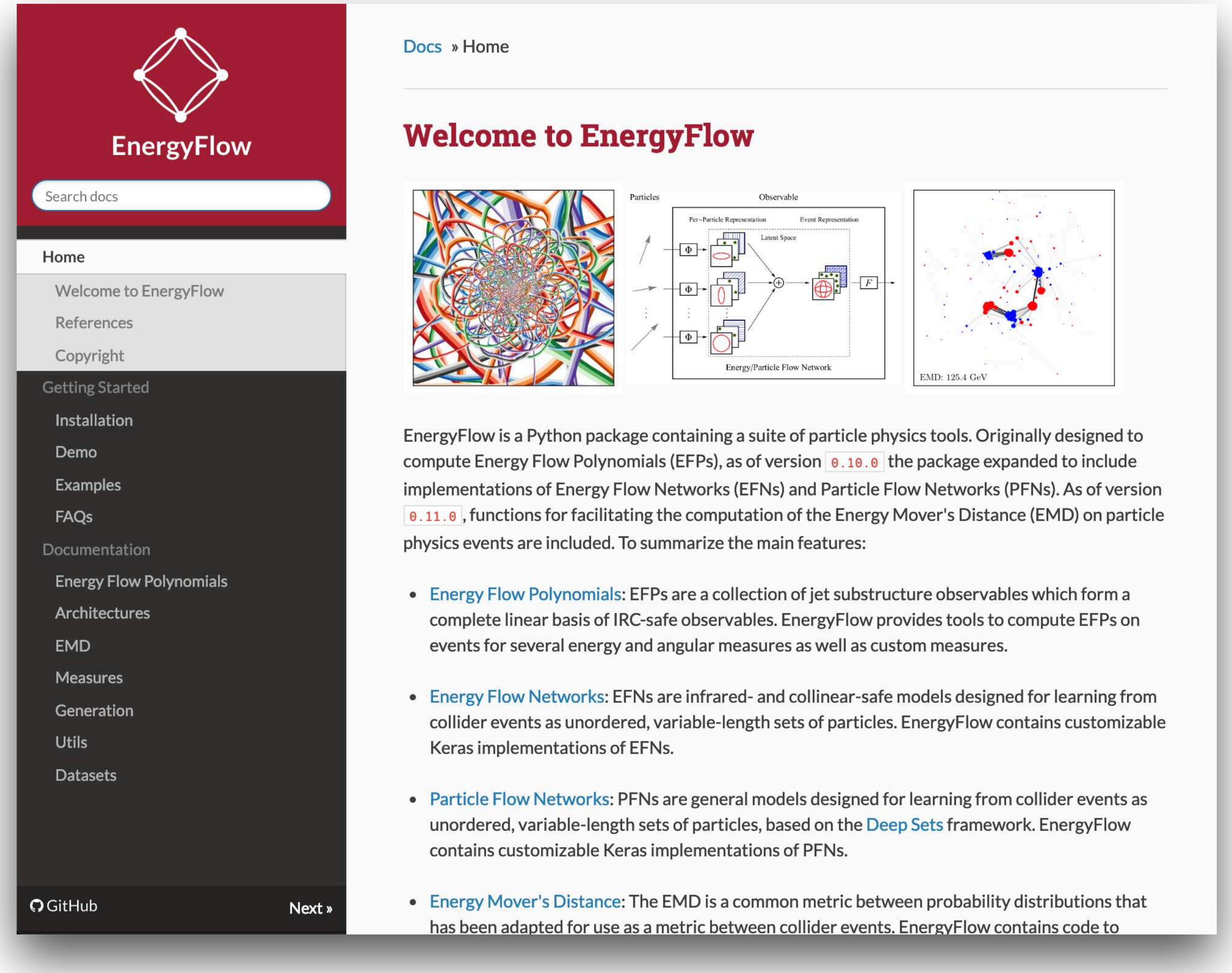
EnergyFlow Python Package

https://energyflow.network

Parallelized EMD calculations via the Python Optimal Transport library

Keras implementations of EFNs, PFNs, DNNs, CNNs, efficient EFP computation

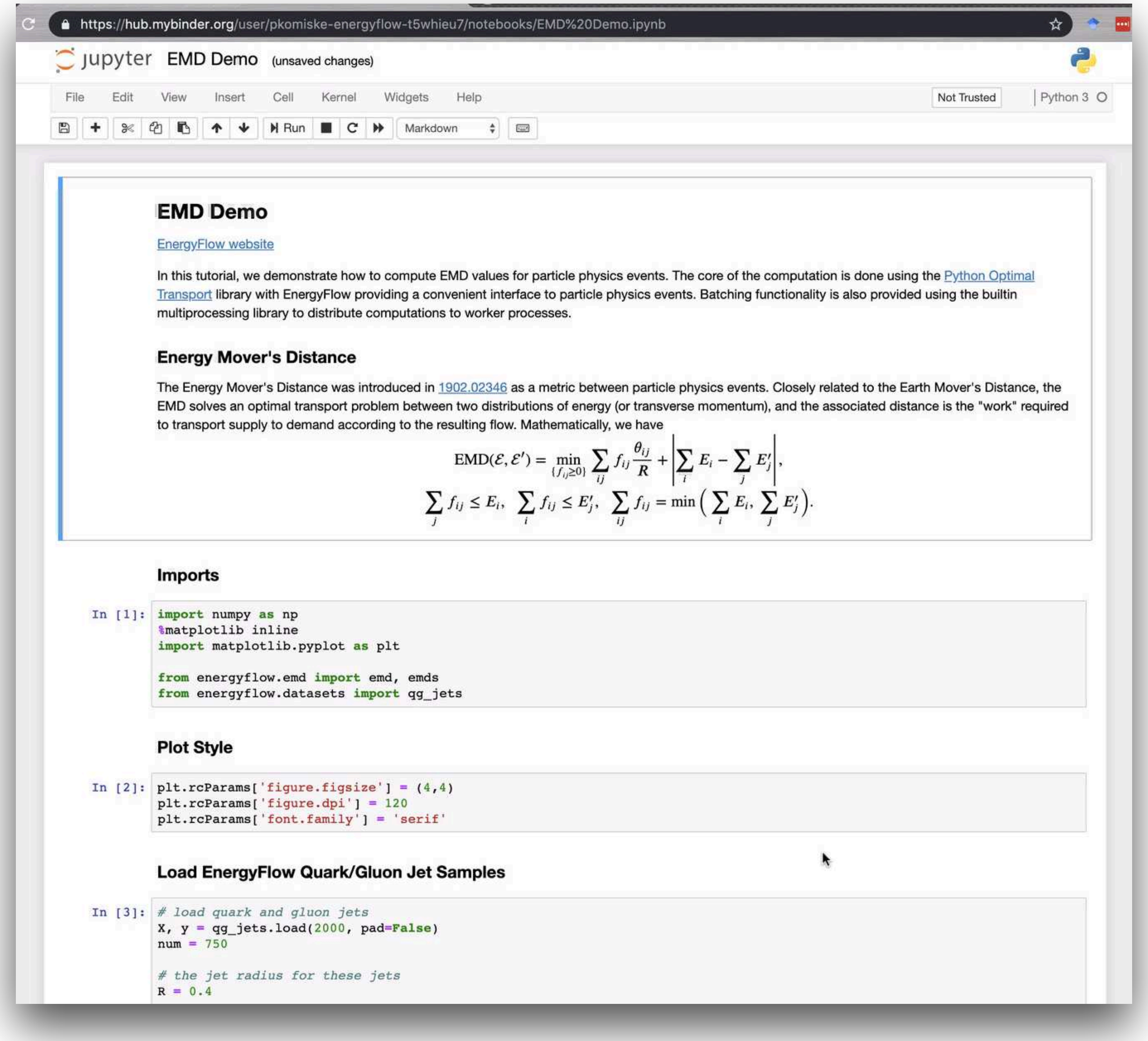
Several detailed examples and demos for common use cases and visualization procedures



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 unordered, variable-length sets of particles, based on the Deep Sets framework. EnergyFlow contains customizable Keras implementations of PFNs.
- Energy Mover's Distance: The EMD is a common metric between probability distributions that has been adapted for use as a metric between collider events. EnergyFlow contains code to



EMD Demo

EnergyFlow website

In this tutorial, we demonstrate how to compute EMD values for particle physics events. The core of the computation is done using the Python Optimal Transport library with EnergyFlow providing a convenient interface to particle physics events. Batching functionality is also provided using the builtin multiprocessing library to distribute computations to worker processes.

Energy Mover's Distance

The Energy Mover's Distance was introduced in 1902.02346 as a metric between particle physics events. Closely related to the Earth Mover's Distance, the EMD solves an optimal transport problem between two distributions of energy (or transverse momentum), and the associated distance is the "work" required to transport supply to demand according to the resulting flow. Mathematically, we have

$$\text{EMD}(\mathcal{E},\mathcal{E}') = \min_{\{f_{ij}\geq 0\}} \sum_{ij} f_{ij}\frac{\theta_{ij}}{R} + \left|\sum_i E_i - \sum_j E'_j\right|,$$

$$\sum_j f_{ij} \leq E_i,\ \ \sum_i f_{ij} \leq E'_j,\ \ \sum_{ij} f_{ij} = \min\Big(\sum_i E_i, \sum_j E'_j\Big).$$

Imports

In [1]:

```
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt

from energyflow.emd import emd, emds
from energyflow.datasets import qg_jets
```

Plot Style

In [2]:

```
plt.rcParams['figure.figsize'] = (4,4)
plt.rcParams['figure.dpi'] = 120
plt.rcParams['font.family'] = 'serif'
```

Load EnergyFlow Quark/Gluon Jet Samples

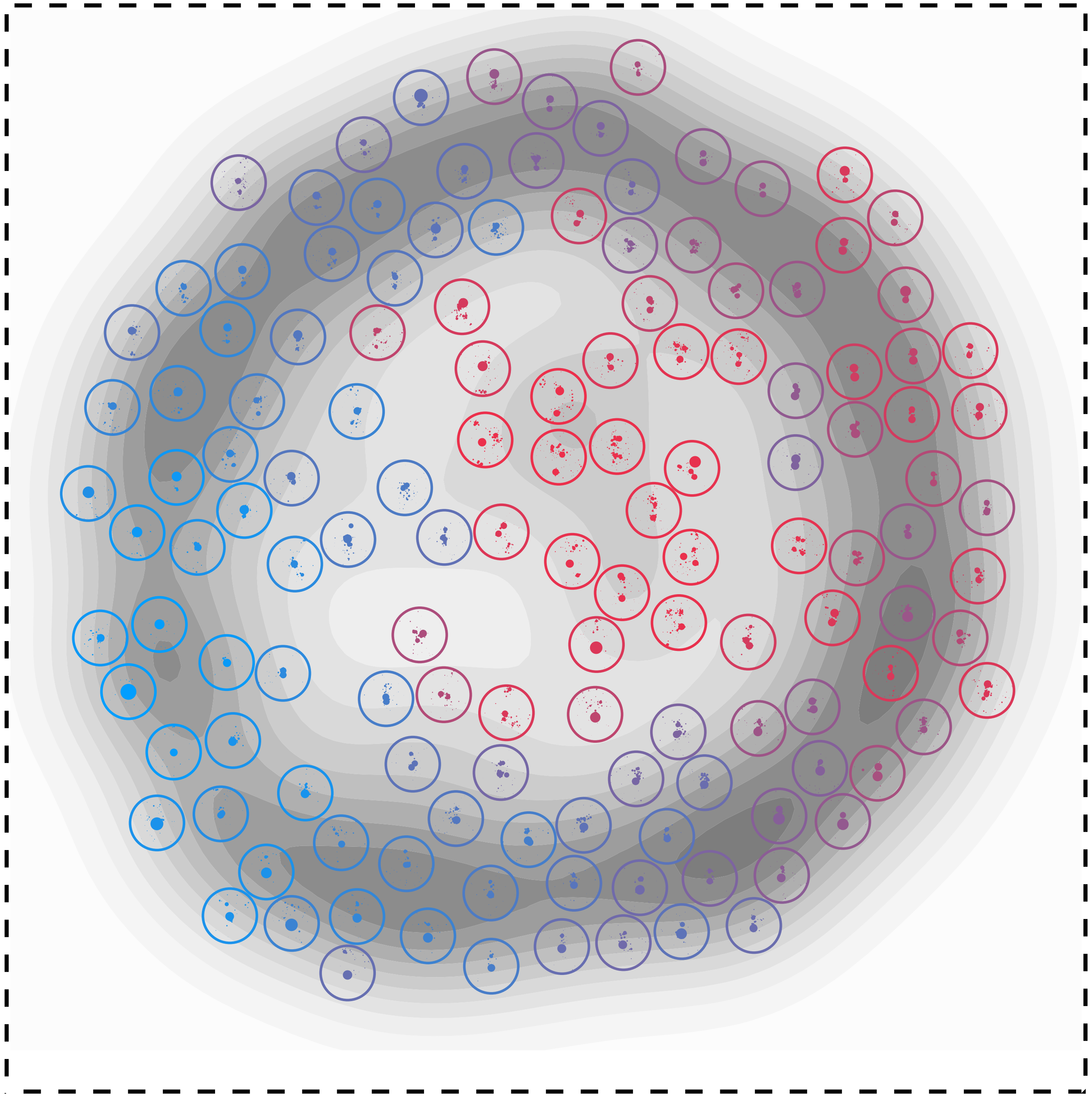
In [3]:

```
# load quark and gluon jets
X, y = qg_jets.load(2000, pad=False)
num = 750

# the jet radius for these jets
R = 0.4
```

Additional Slides

Boosted *W* Jets

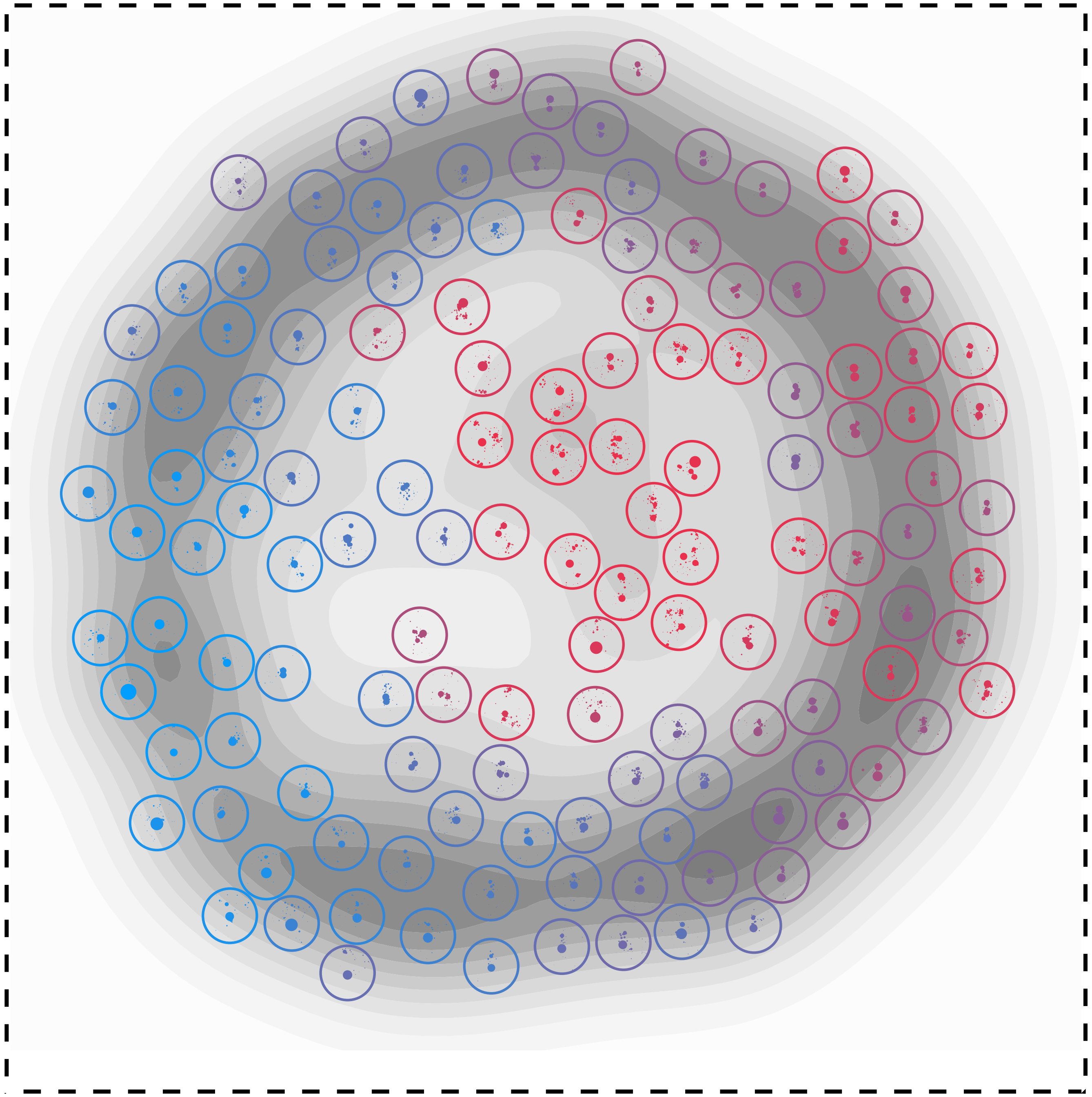


Abstract space of W jets

[PTK, Metodiev, Thaler, 1902.02346]

Boosted *W* Jets

Gray contours represent the density of jets



Each circle is a particular *W* jet

Abstract space of W jets

[PTK, Metodiev, Thaler, 1902.02346]

BOOSTON

2019

[BOOST 2019, July 22-26, MIT]

Phenomenology | Reconstruction | Searches | Algorithms | Measurements | Calculations
Modeling | Machine Learning | Pileup Mitigation | Heavy-Ion Collisions | Future Colliders