

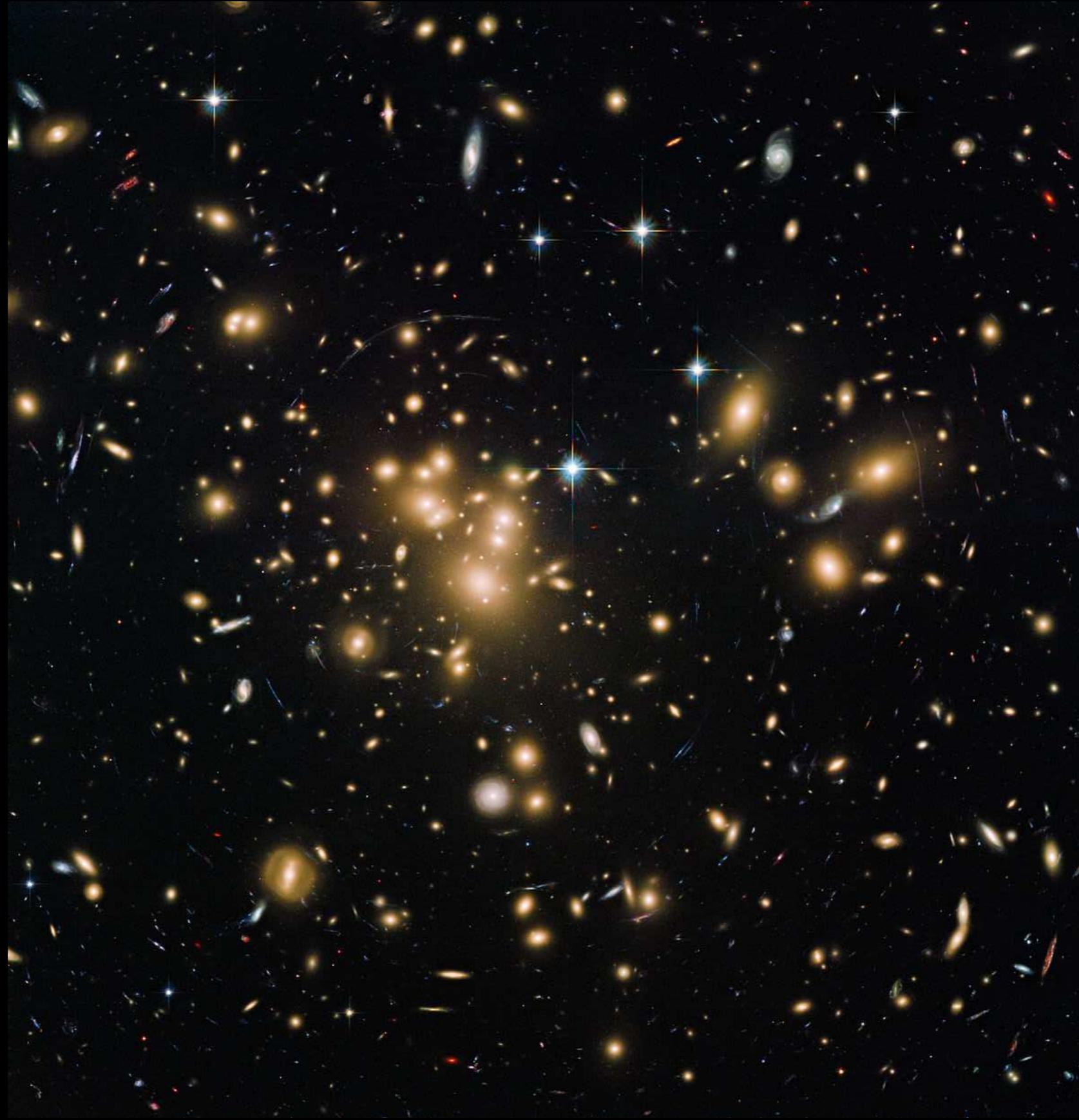
The Space of Collider Events

Jesse Thaler

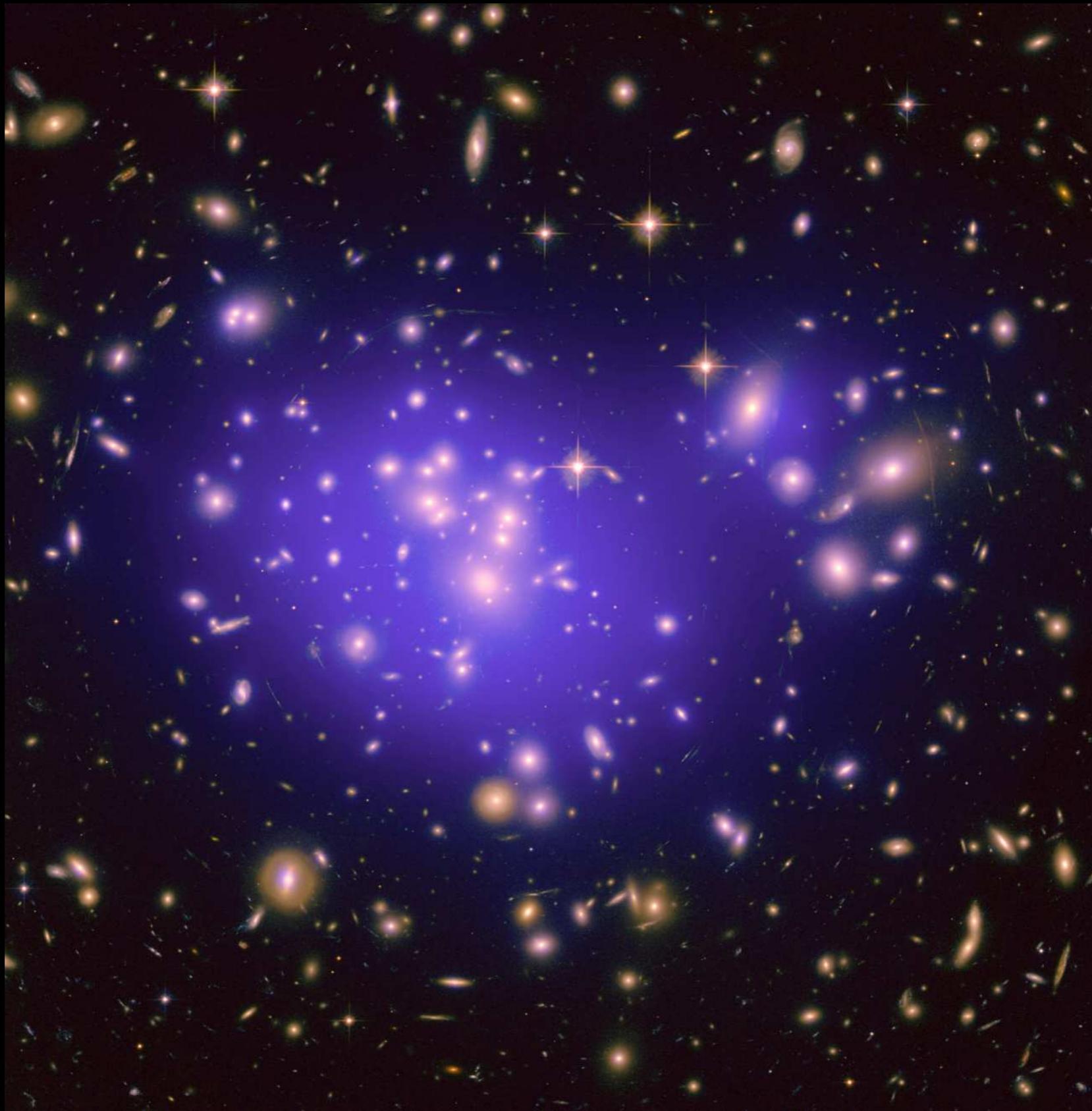


with Patrick Komiske & Eric Metodiev, [1902.02346](#)

Special SCIPP Friday Seminar, UC Santa Cruz — April 19, 2019

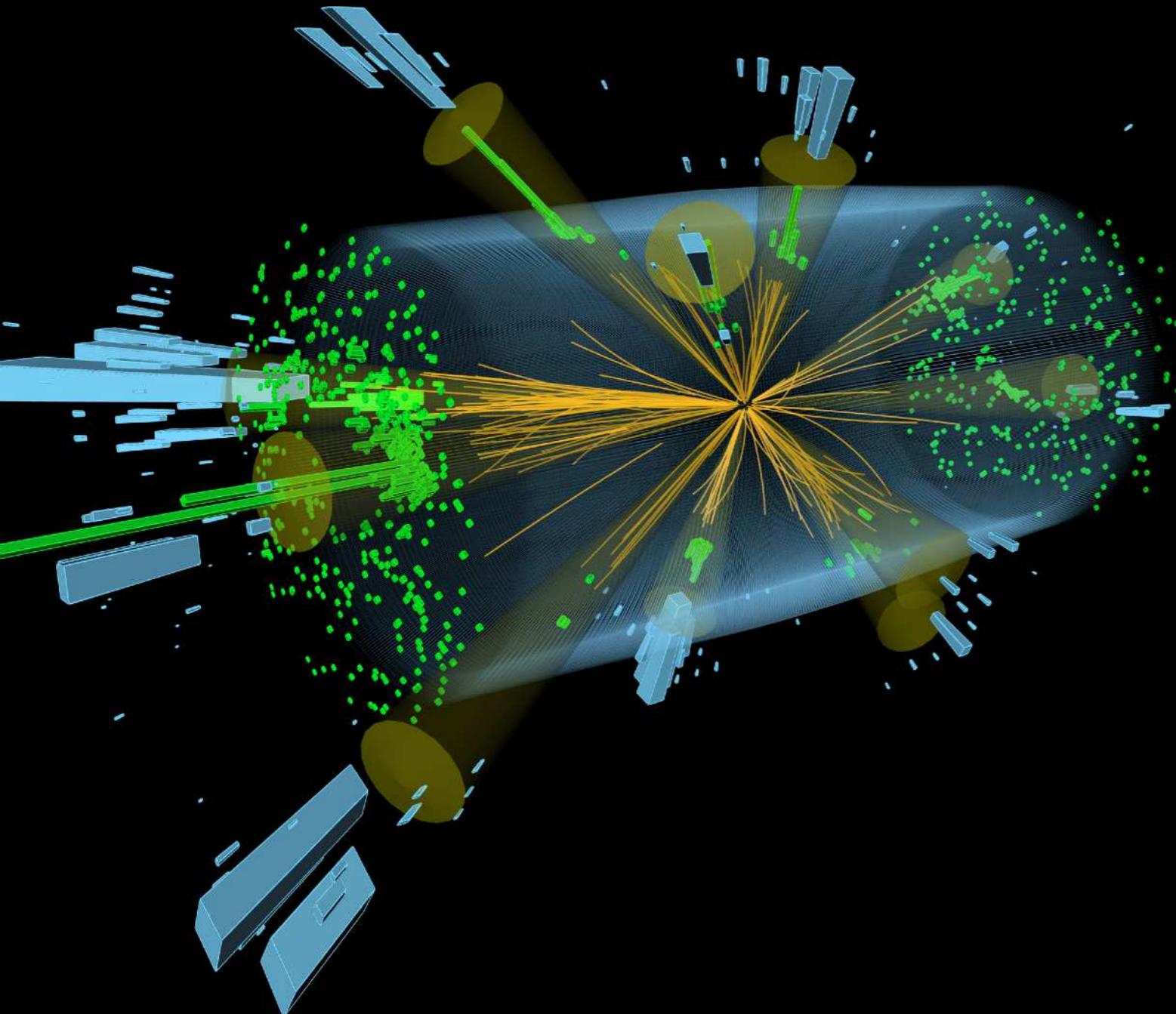


ESA/Hubble: Abell 1689



ESA/Hubble: Abell 1689

Event Displays ...



T E H M

 γ

photon

 e^+

electron

 μ^+

muon

 π^+

pion

 K^+

kaon

 K_L^0

K-long

 p/\bar{p}

proton

 n/\bar{n}

neutron

elementary

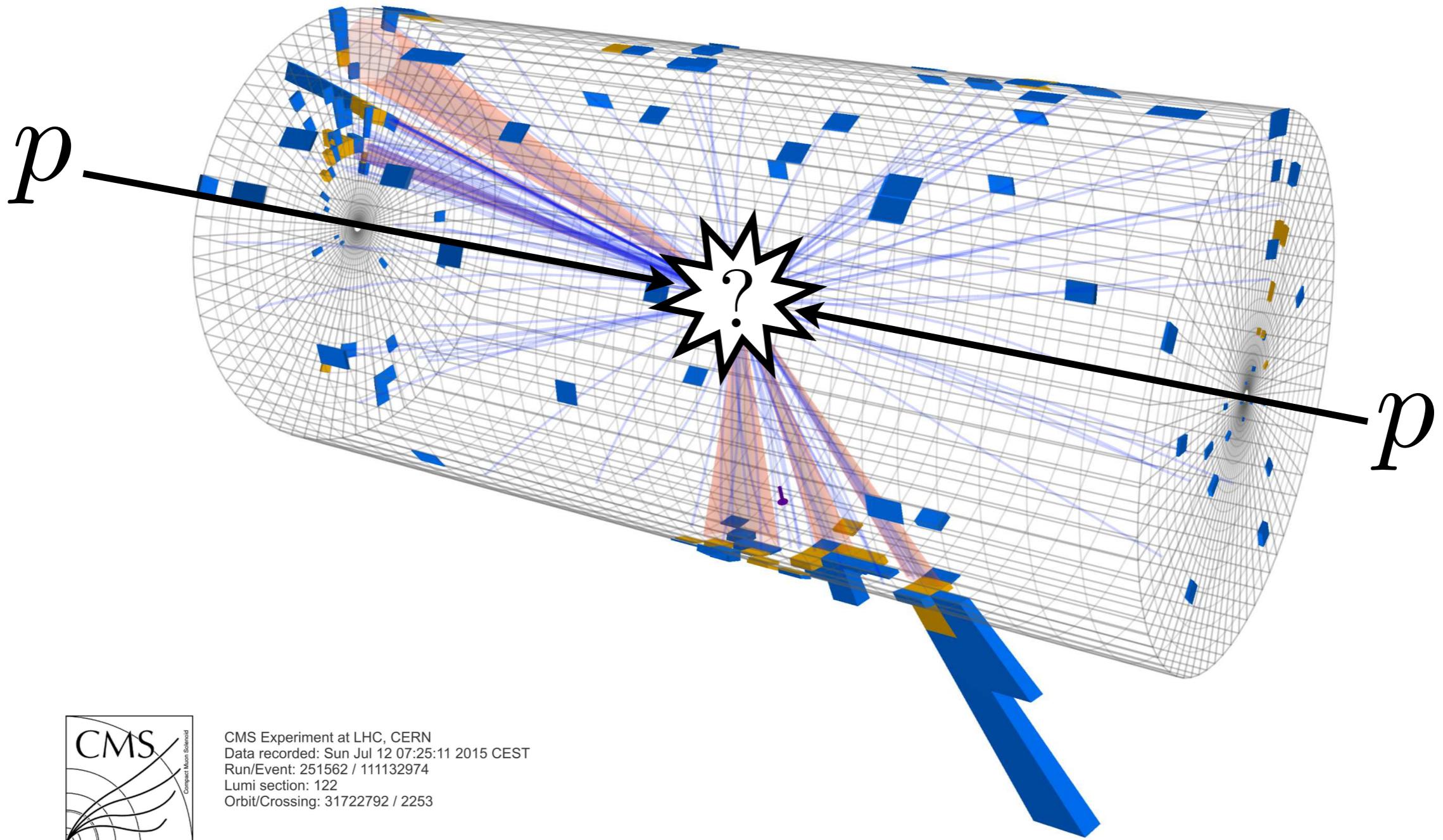
composite

Event Displays (for an Event)

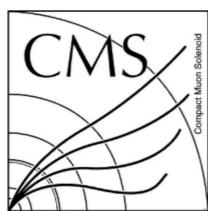
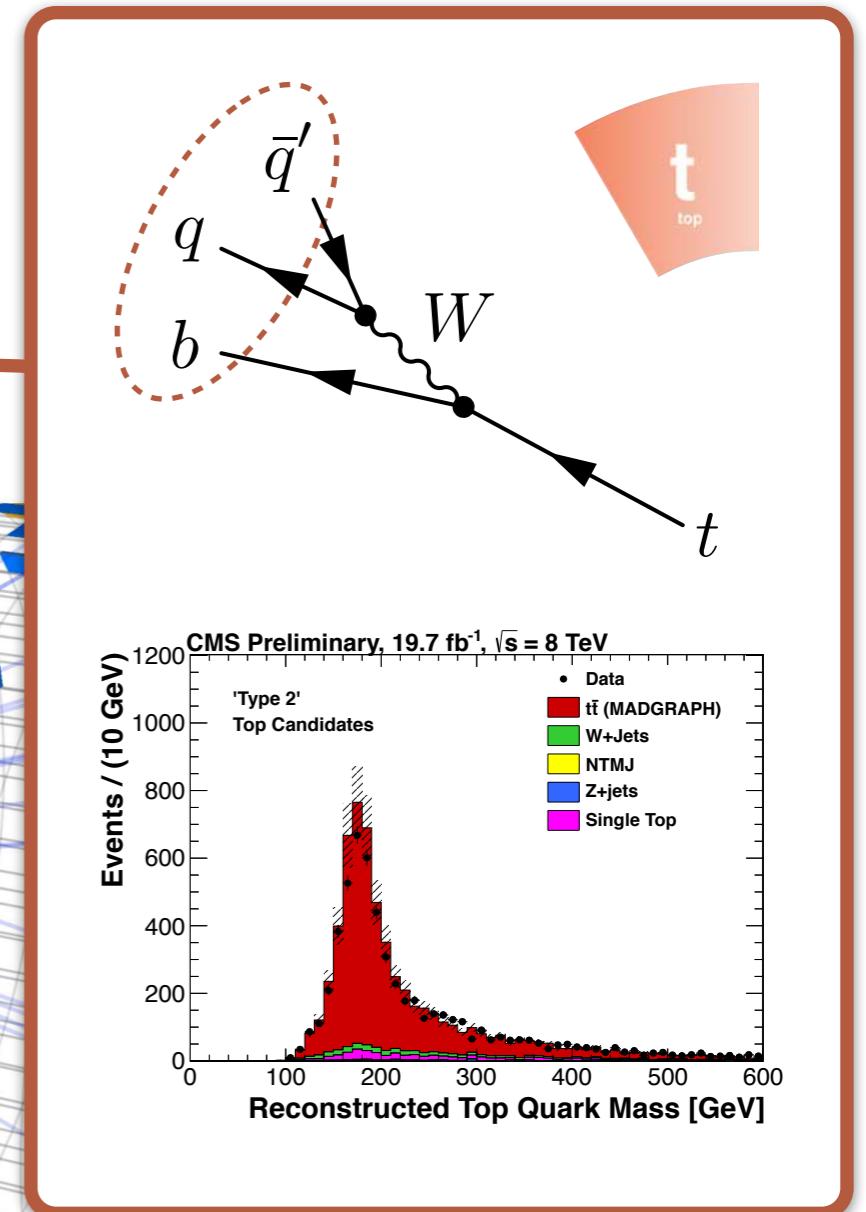
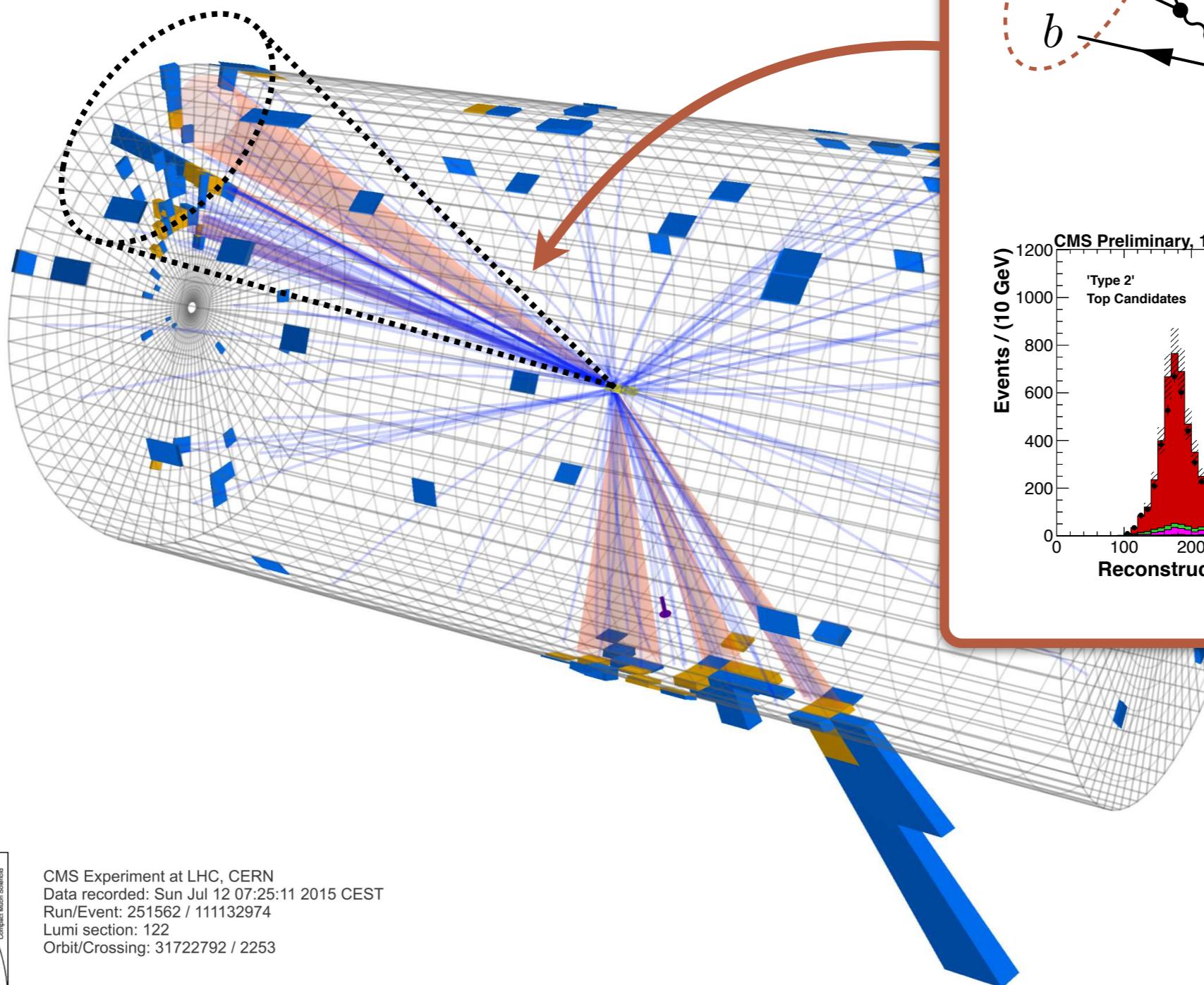
BOSTON
2019

[BOOST 2019: July 22-26, 2019, MIT]

Science in the Visualization?

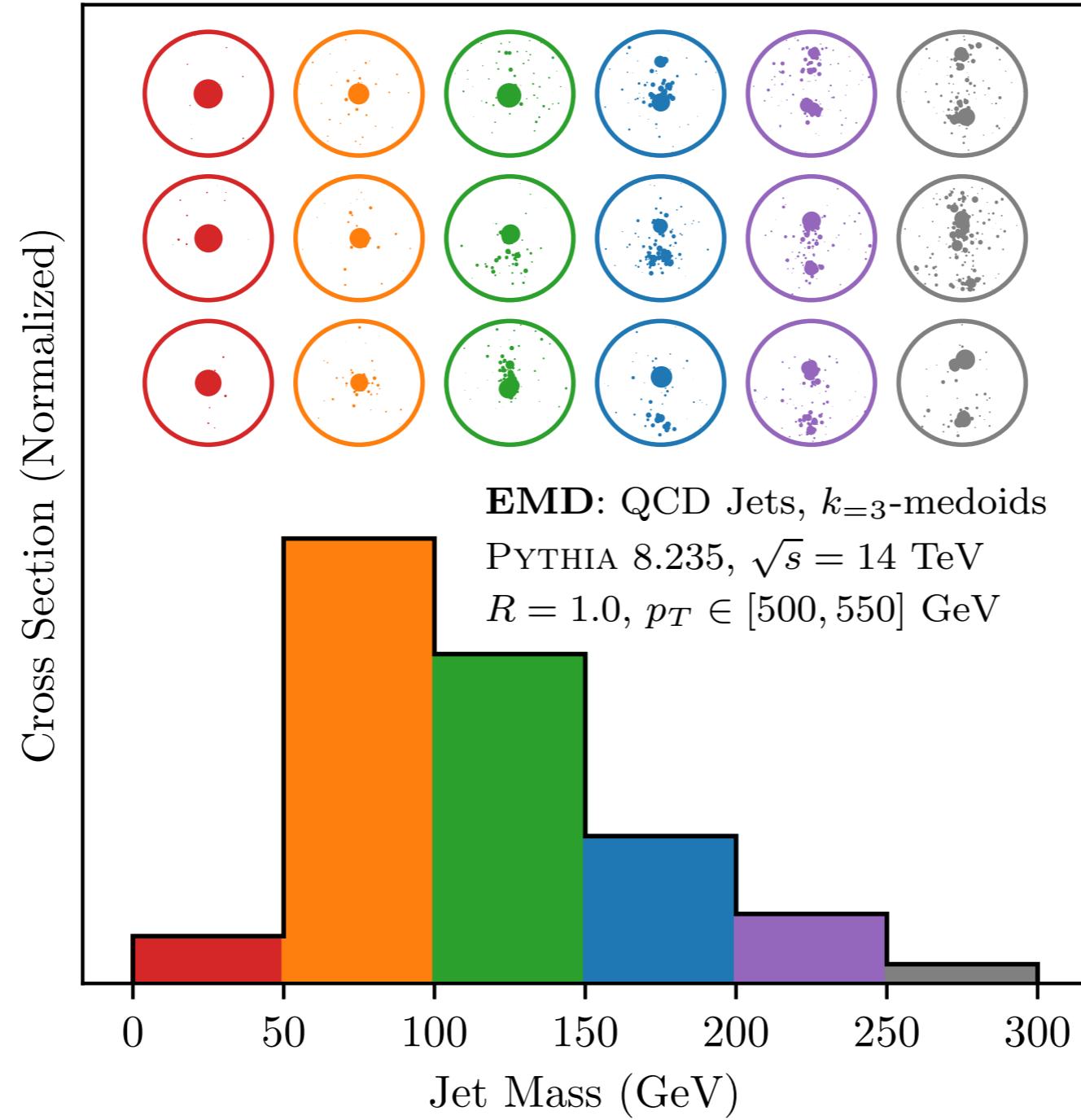


Science in the Visualization?



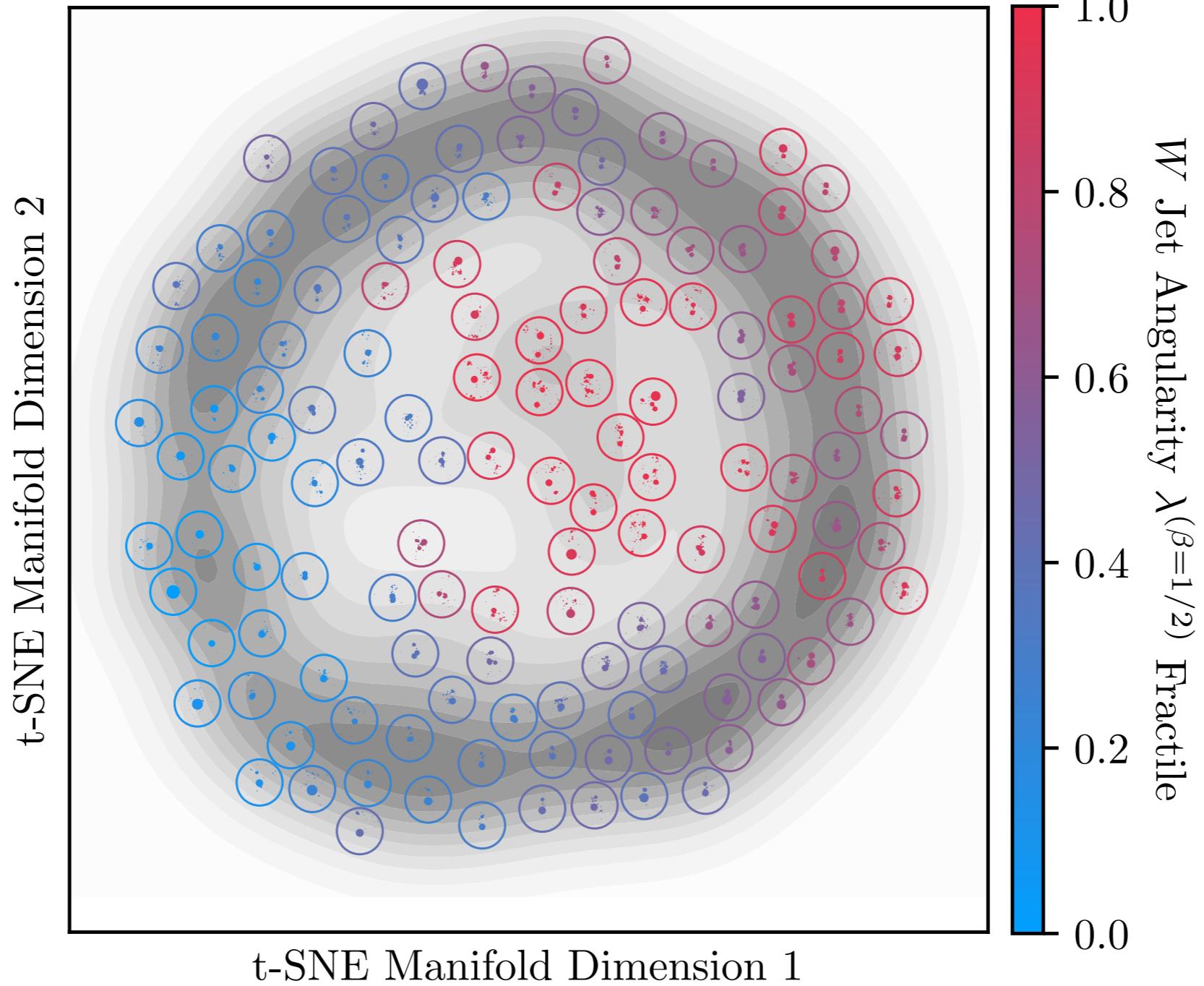
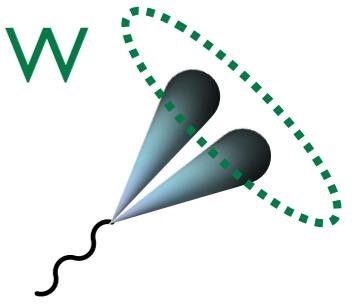
CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253

My New Favorite Histogram



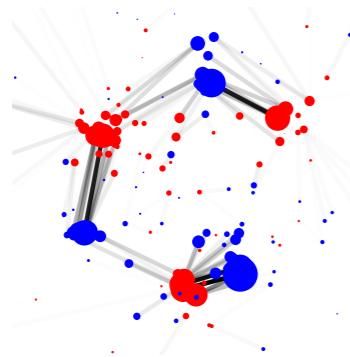
[Komiske, Metodiev, JDT, [1902.02346](#)]

The Space of Boosted W Bosons

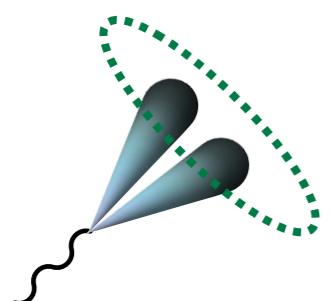


[Komiske, Metodiev, JDT, [1902.02346](#)]

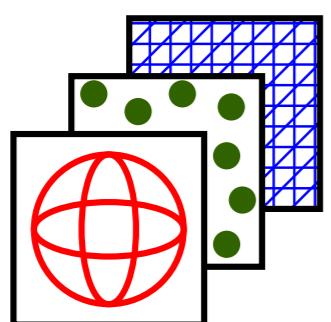
Outline



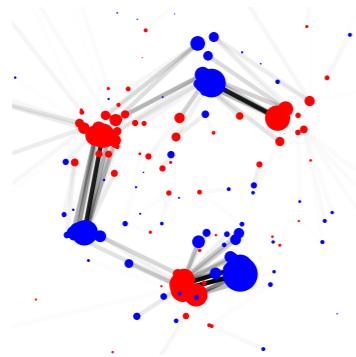
The Energy Mover's Distance



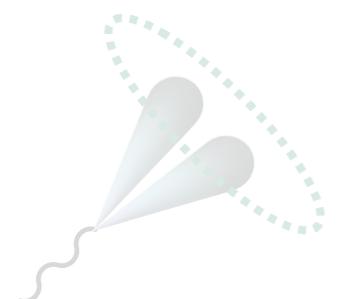
Initial Physics Studies



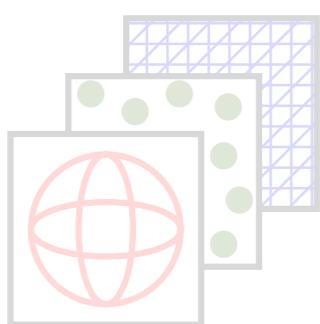
(Broader Comments)



The Energy Mover's Distance

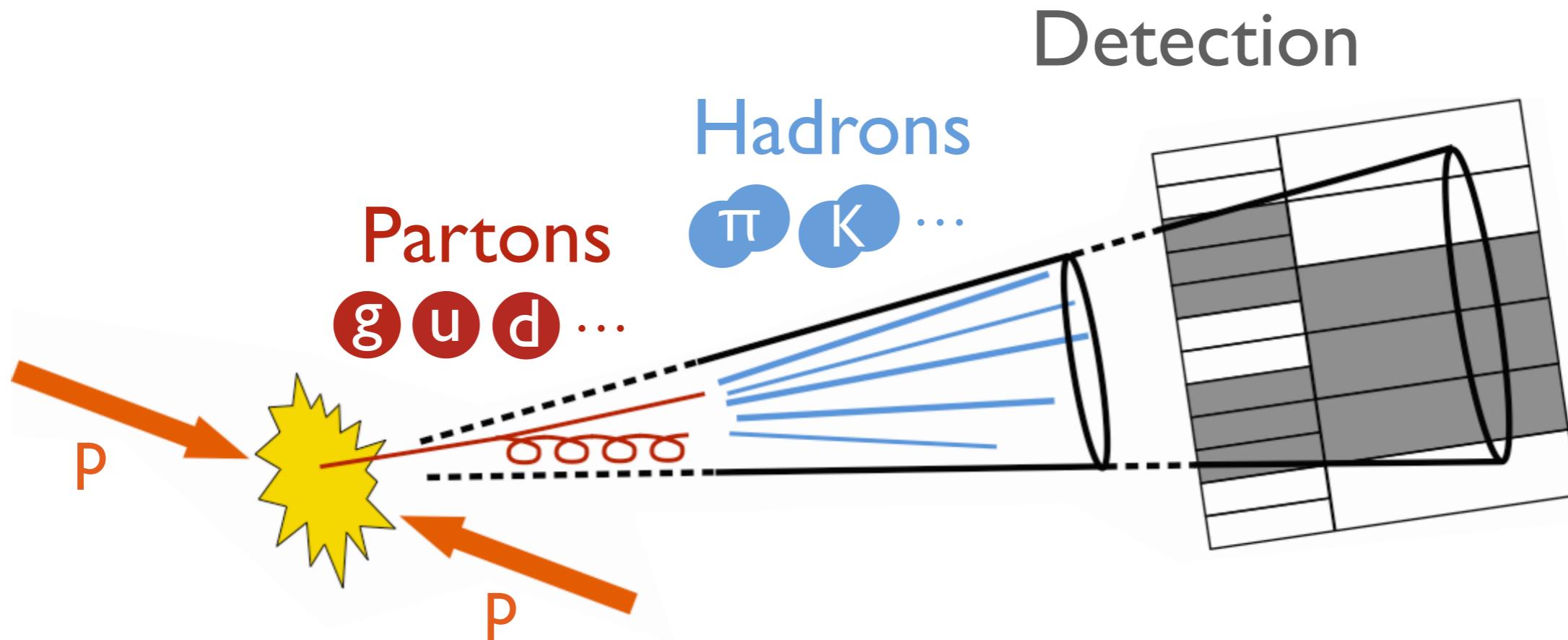


Initial Physics Studies



(Broader Comments)

Focus on Energy Flow



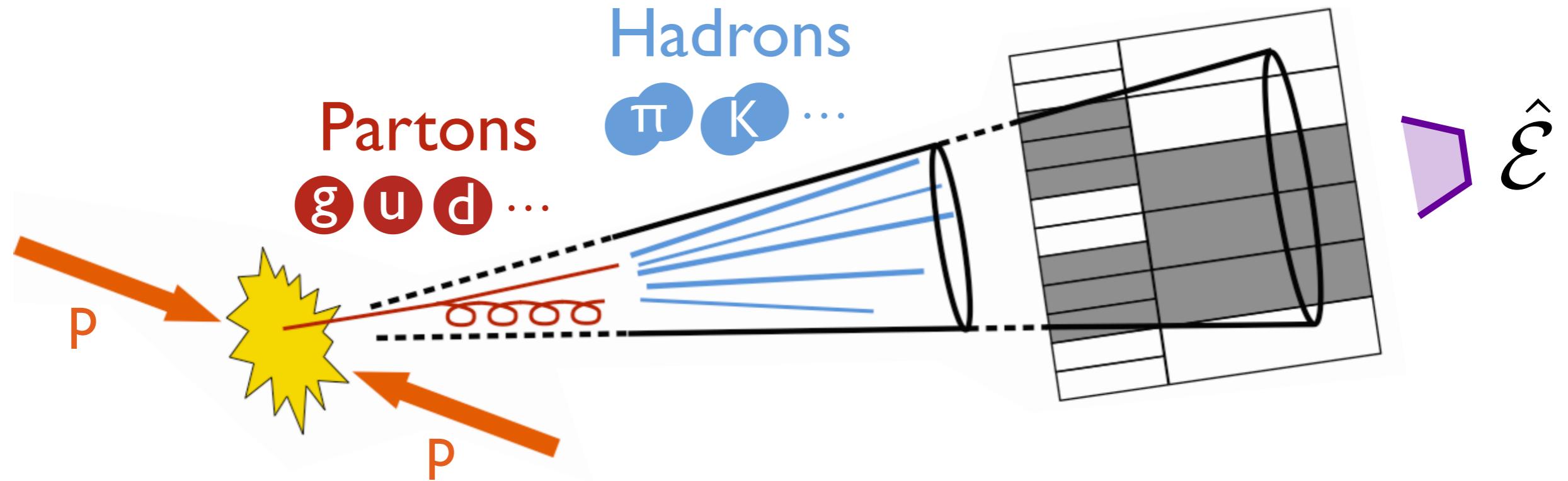
[Sveshnikov, Tkachov, [hep-ph/9512370](#); Hofman, Maldacena, [0803.1467](#); Mateu, Stewart, JDT, [1209.3781](#); Komiske, Metodiev, JDT, [1712.07124](#), [1810.05165](#)]

Focus on Energy Flow

$$\hat{\mathcal{E}} \simeq \lim_{t \rightarrow \infty} \hat{n}_i T^{0i}(t, vt\hat{n})$$

Theory

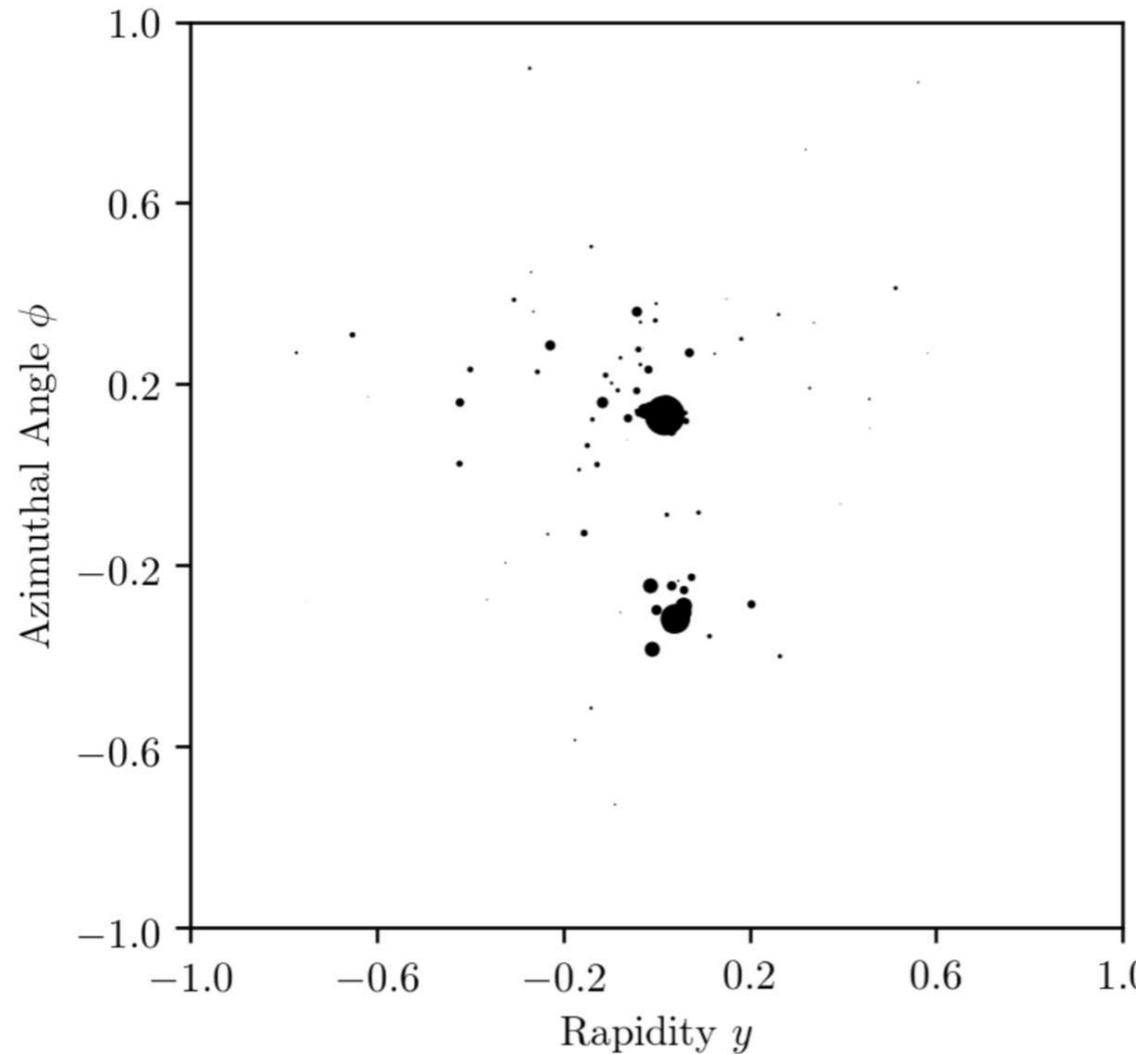
Detection



*Stress-energy flow: Measure of event/jet structure
robust to non-perturbative and detector effects (i.e. **IRC safe**)*

[Sveshnikov, Tkachov, [hep-ph/9512370](#); Hofman, Maldacena, [0803.1467](#); Mateu, Stewart, JDT, [I209.3781](#); Komiske, Metodiev, JDT, [I712.07124](#), [I810.05165](#)]

Focus on Energy Flow



Represent jet as:

$$\rho(\hat{p}) = \sum_{i \in \text{jet}} E_i \delta(\hat{p} - \hat{p}_i)$$

Energy
(p_T)

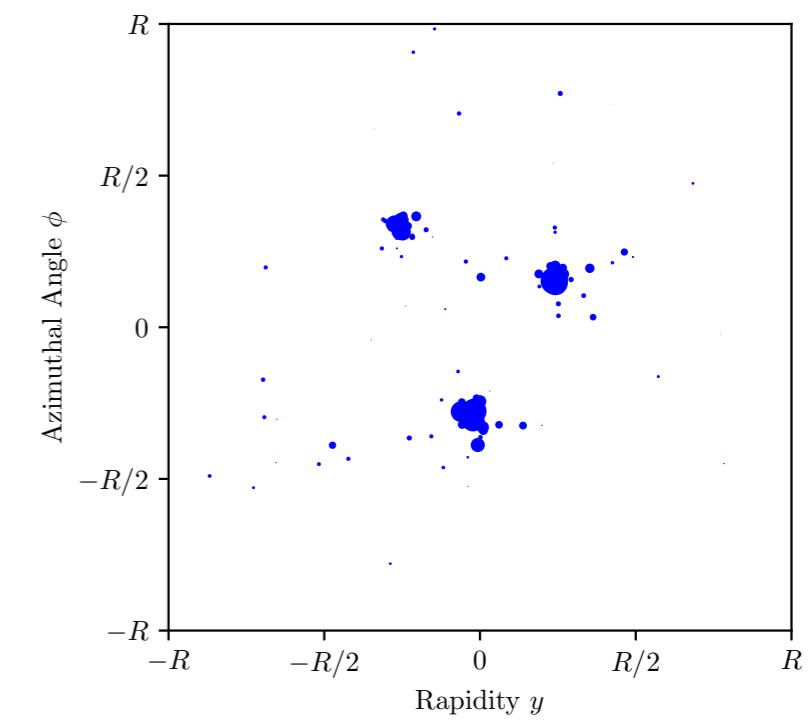
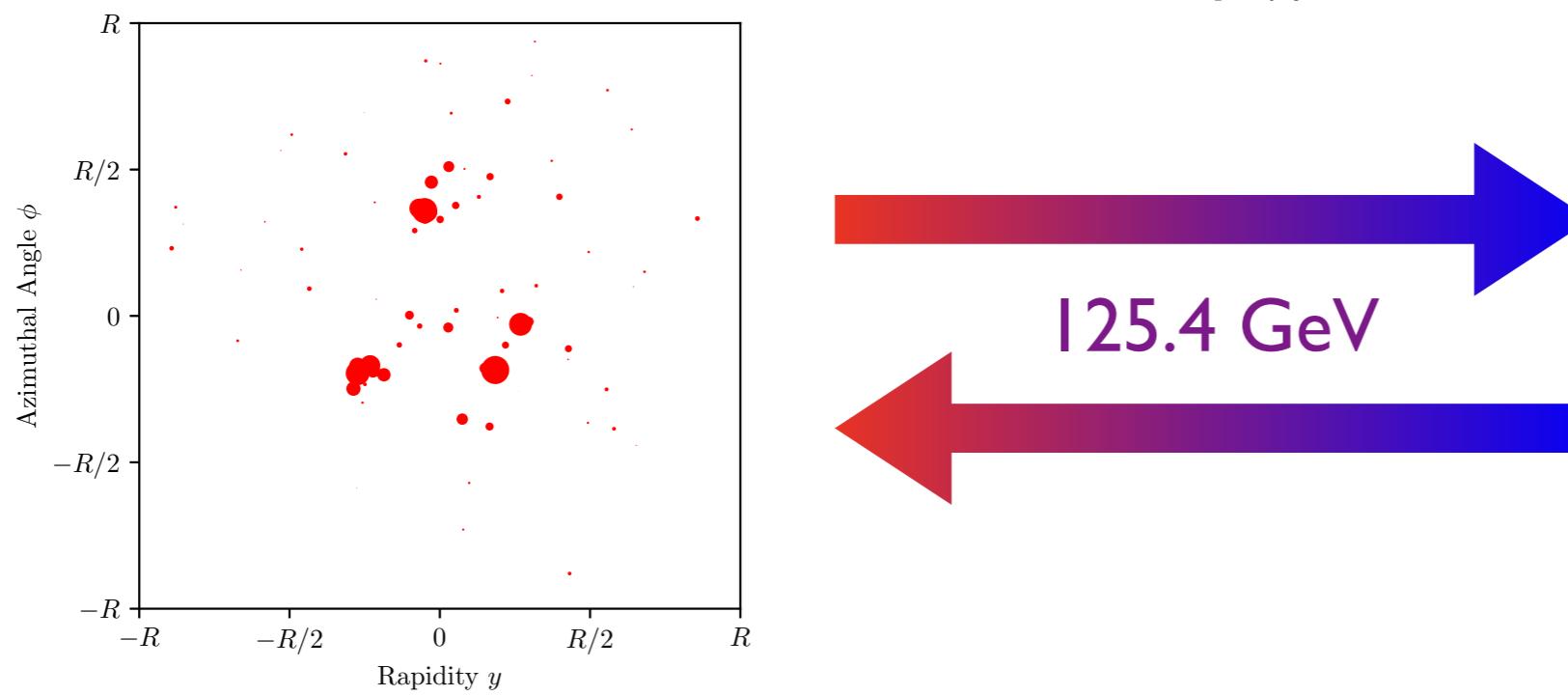
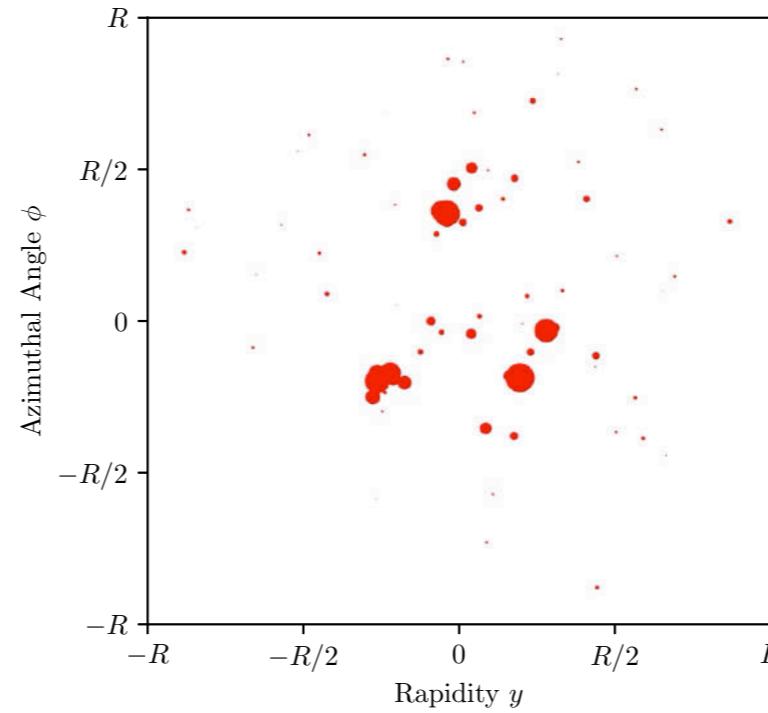
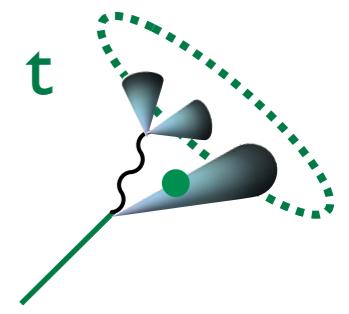
Direction
(y, ϕ)

Safe to infrared & collinear splittings
No flavor/charge information
No pixelation needed

*Stress-energy flow: Measure of event/jet structure
robust to non-perturbative and detector effects (i.e. **IRC safe**)*

[Sveshnikov, Tkachov, [hep-ph/9512370](#); Hofman, Maldacena, [0803.1467](#); Mateu, Stewart, JDT, [1209.3781](#); Komiske, Metodiev, JDT, [1712.07124](#), [1810.05165](#)]

Similarity of Two Energy Flows?



[h/t [@pkomiske](#) & [@EricMetodiev](#) for animations]

The Earth Mover's Distance

Optimal Transport:

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

Minimum “work” (**stuff x distance**) to make
one distribution ...



The Earth Mover's Distance

Optimal Transport:

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

Minimum “work” (**stuff** × **distance**) to make
one distribution look like **another distribution**



The Earth Mover's Distance

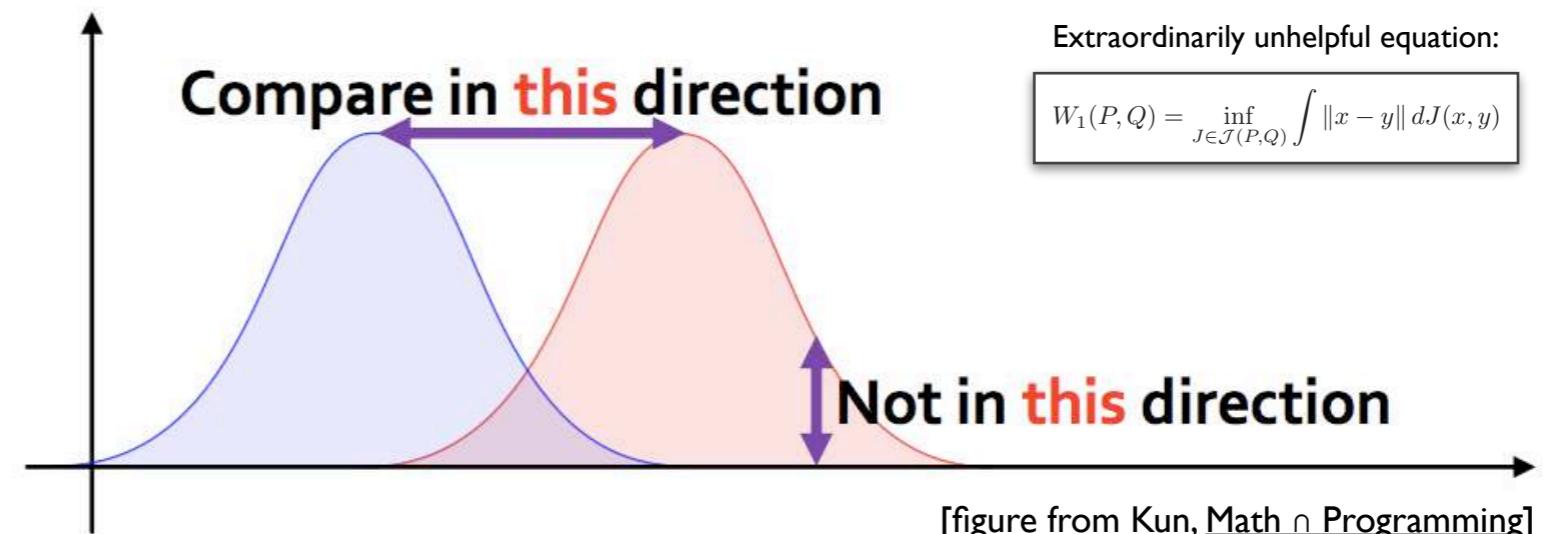
Optimal Transport:

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

Minimum “work” (**stuff** × **distance**) to make one distribution look like another distribution

Equivalent to \mathbb{L} -Wasserstein:

Metric on normalized distributions
(symmetric, non-negative,
triangle inequality, zero iff equal)



Very popular in ML applications:

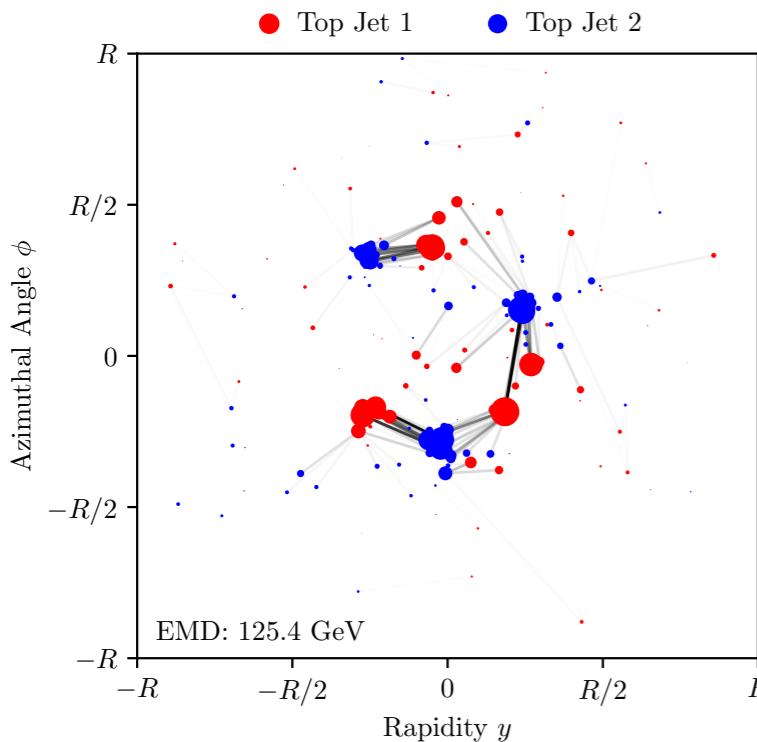
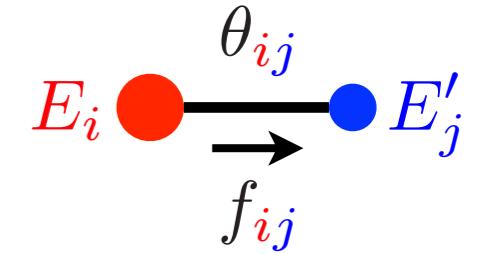
e.g. *Wasserstein Generative Adversarial Networks*

[Arjovsky, Chintala, Bottou, [1701.07875](#);
see also Erdmann, Geiger, Glombitza, Schmidt, [1802.03325](#); Erdmann, Glombitza, Quast, [1807.01954](#)]

e.g. *Wasserstein(-Wasserstein) Autoencoders*

[Tolstikhin, Bousquet, Gelly, Schoelkopf, [1711.01558](#); Zhang, Gao, Jiao, Liu, Wang, Yang, [1902.09323](#)]

The Energy Mover's Distance

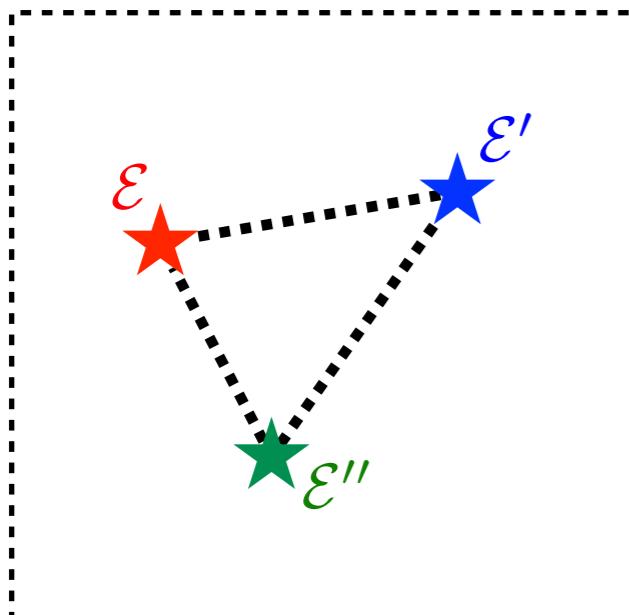


Optimal transport between energy flows...

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

↑
in GeV

Cost to move energy **Cost to create energy**

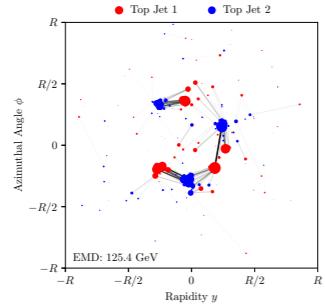


...defines a metric on the space of events

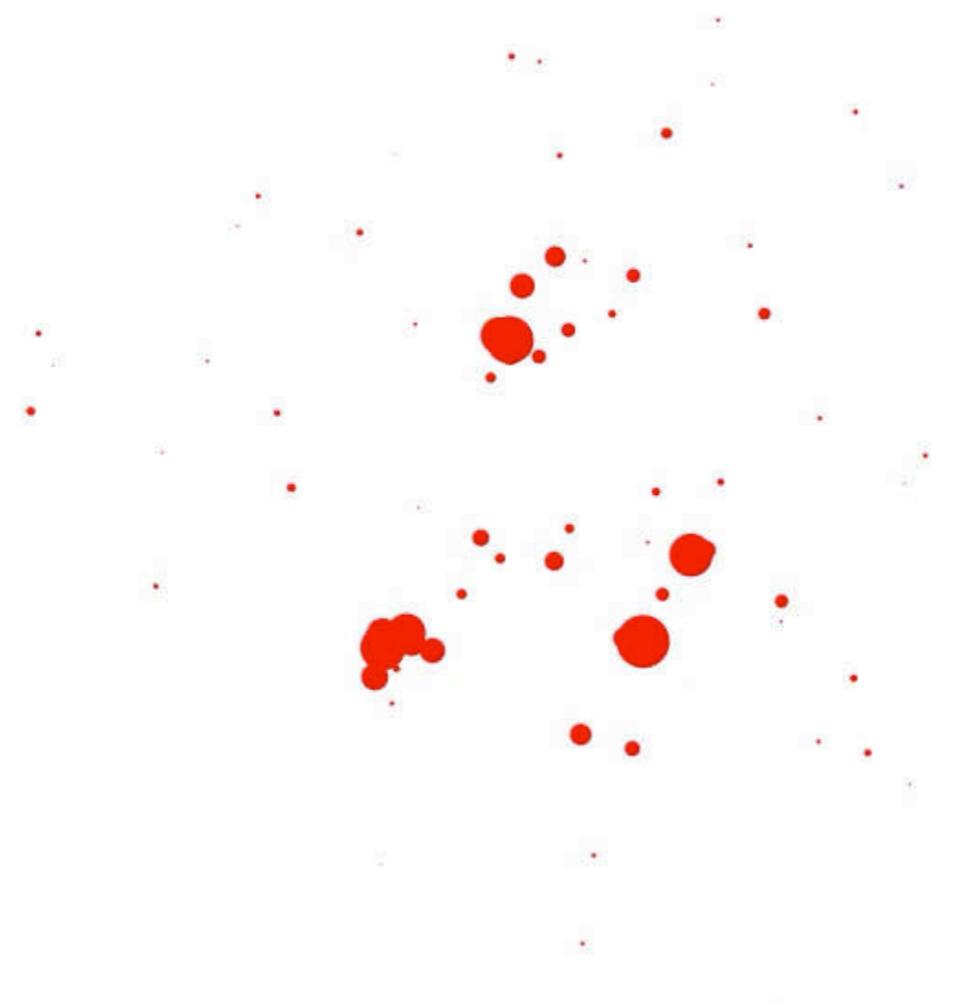
$$0 \leq \text{EMD}(\mathcal{E}, \mathcal{E}') \leq \text{EMD}(\mathcal{E}, \mathcal{E}'') + \text{EMD}(\mathcal{E}', \mathcal{E}'')$$

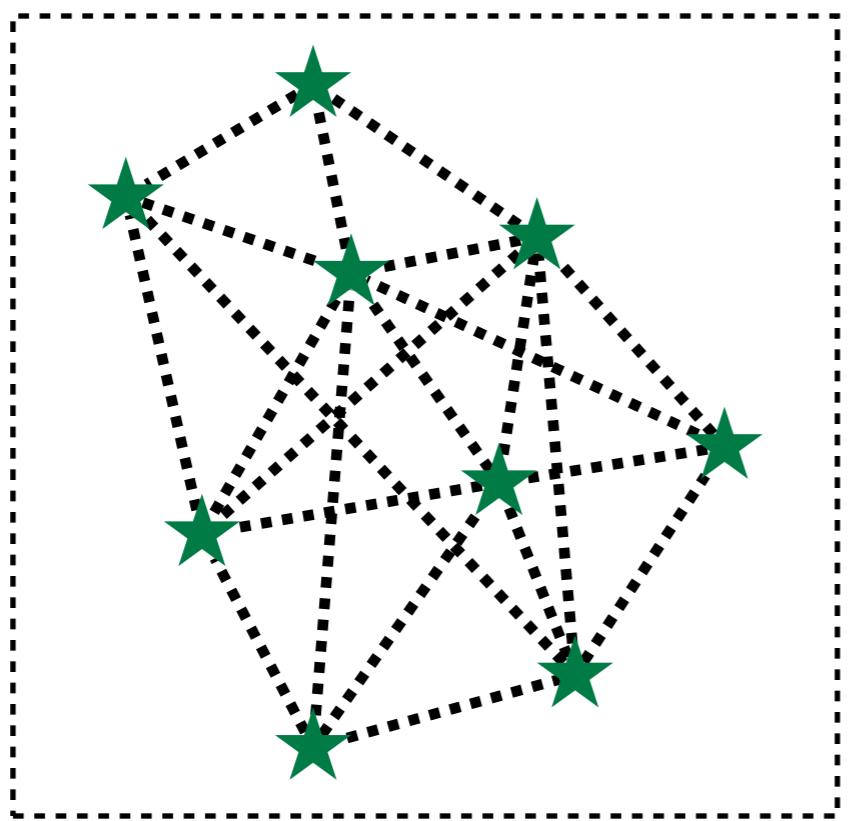
(assuming $R \geq \theta_{\max}/2$, i.e. $R \geq$ jet radius for conical jets)

[Komiske, Metodiev, JDT, [1902.02346](#);
see also Pele, Werman, [ECCV 2008](#); Pele Taskar, [GSI 2013](#)]



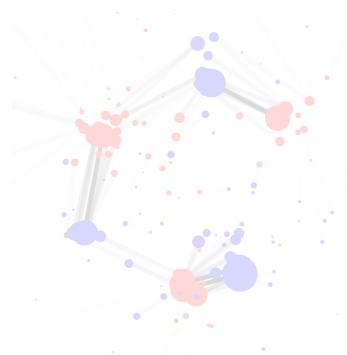
$$\text{EMD}(\mathcal{E}, \mathcal{E}') = \min_{\{f\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R} + \left| \sum_i E_i - \sum_j E'_j \right|$$



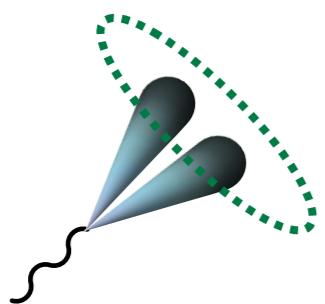


We defined a (metric)
space for collider events

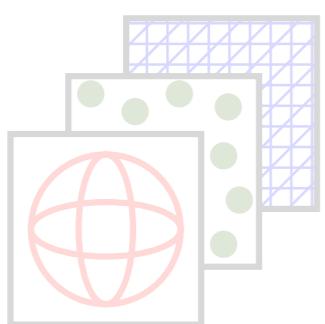
Now what?



The Energy Mover's Distance



Initial Physics Studies



(Broader Comments)

Visualizing Jet Evolution

500 GeV

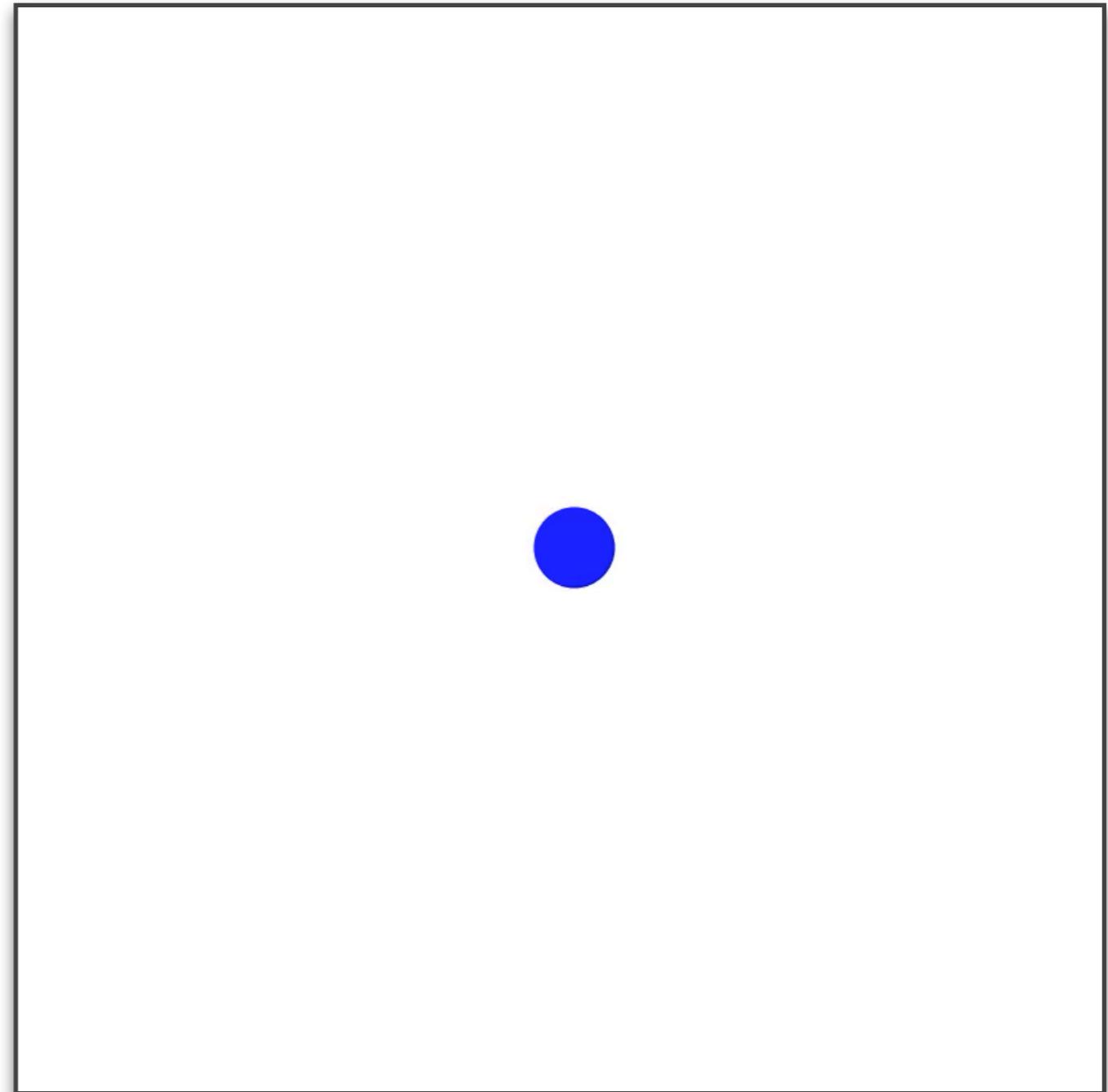
Hard Quark ●

Showering
EMD: 111.6 GeV

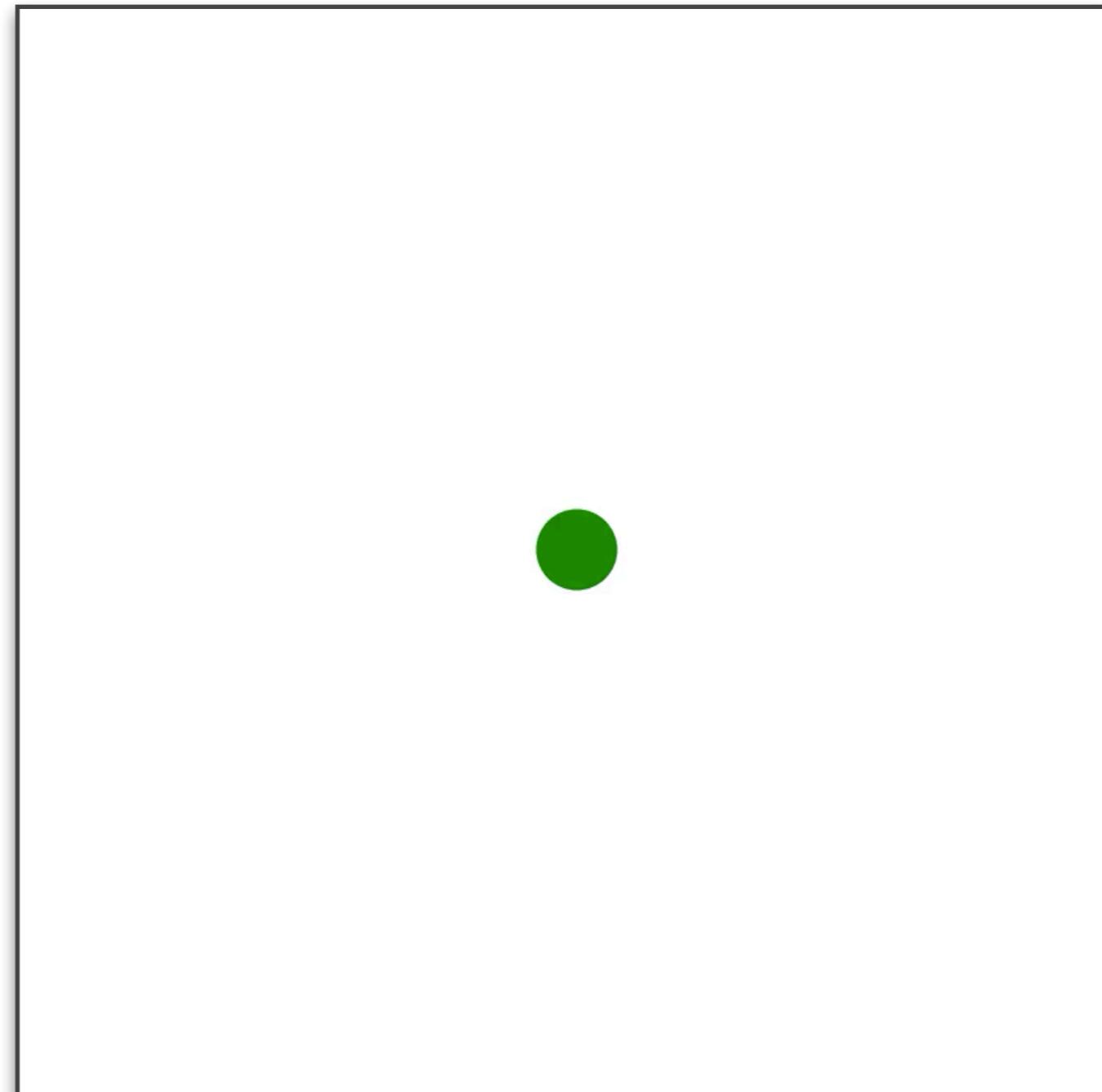
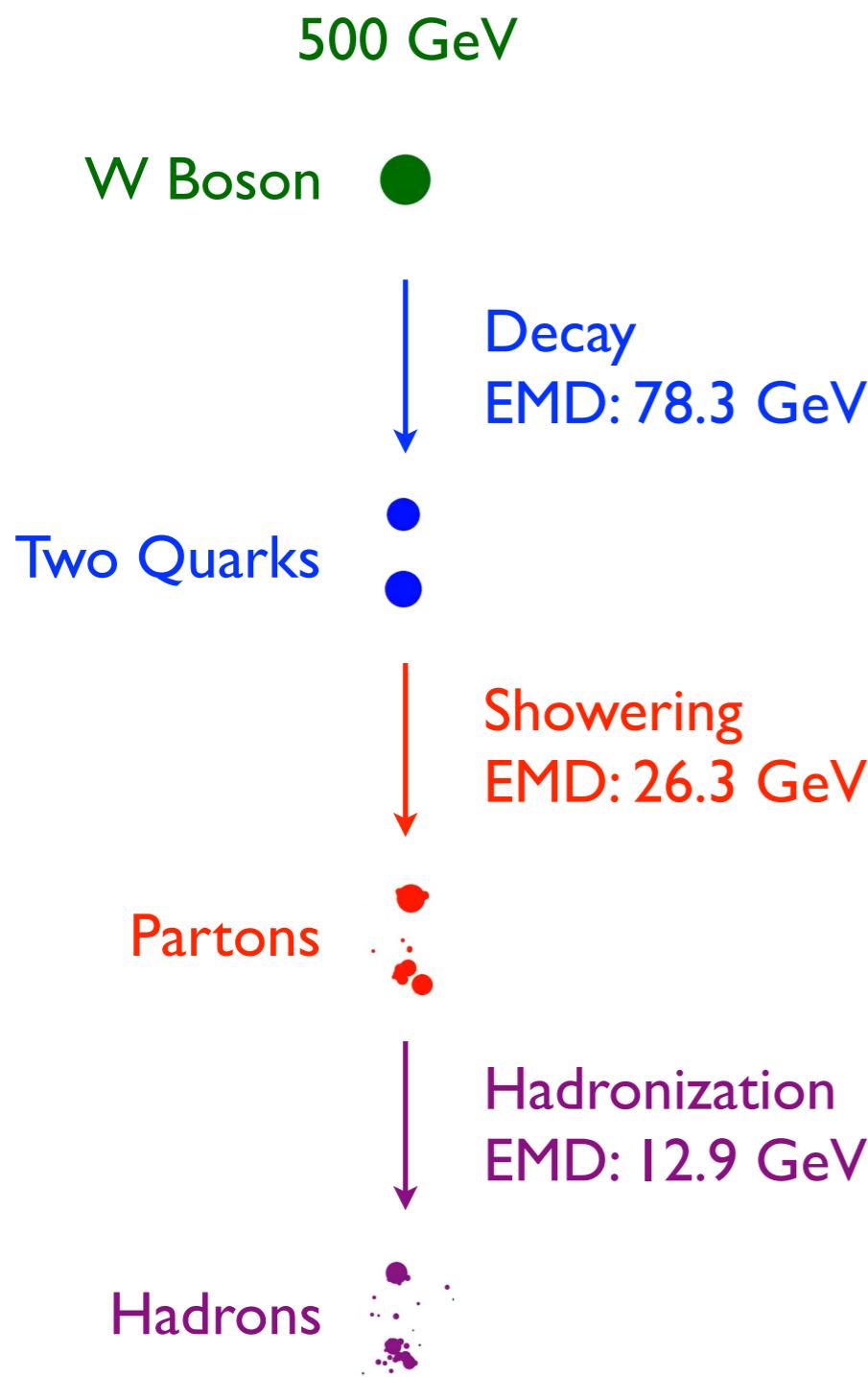
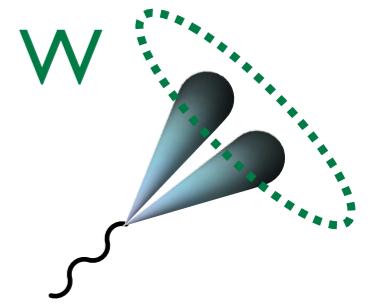
Partons

Hadronization
EMD: 18.1 GeV

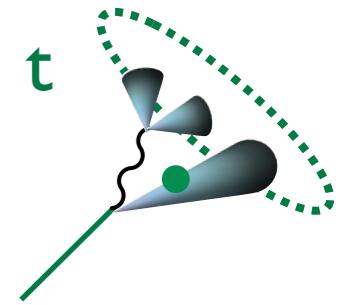
Hadrons



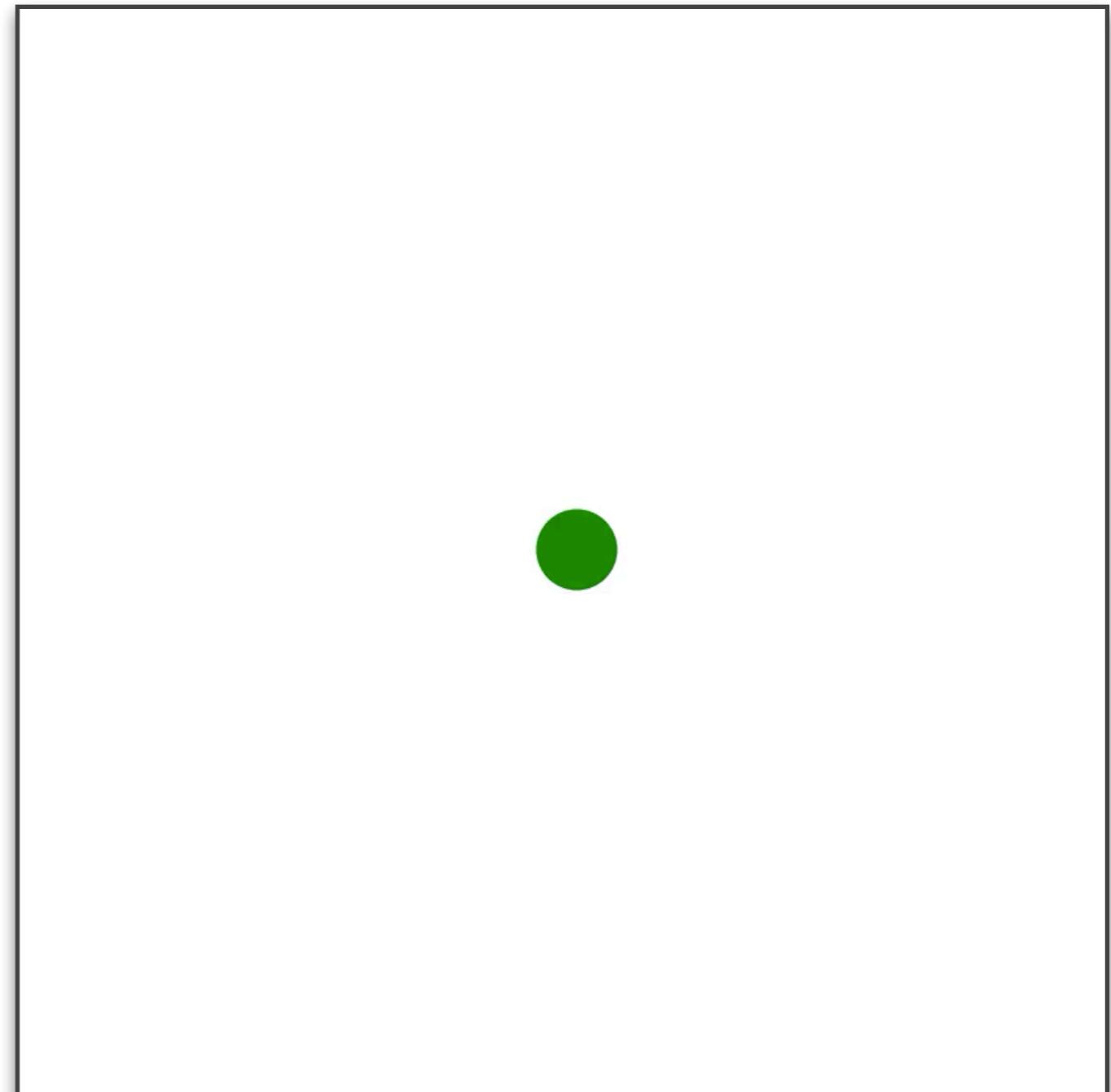
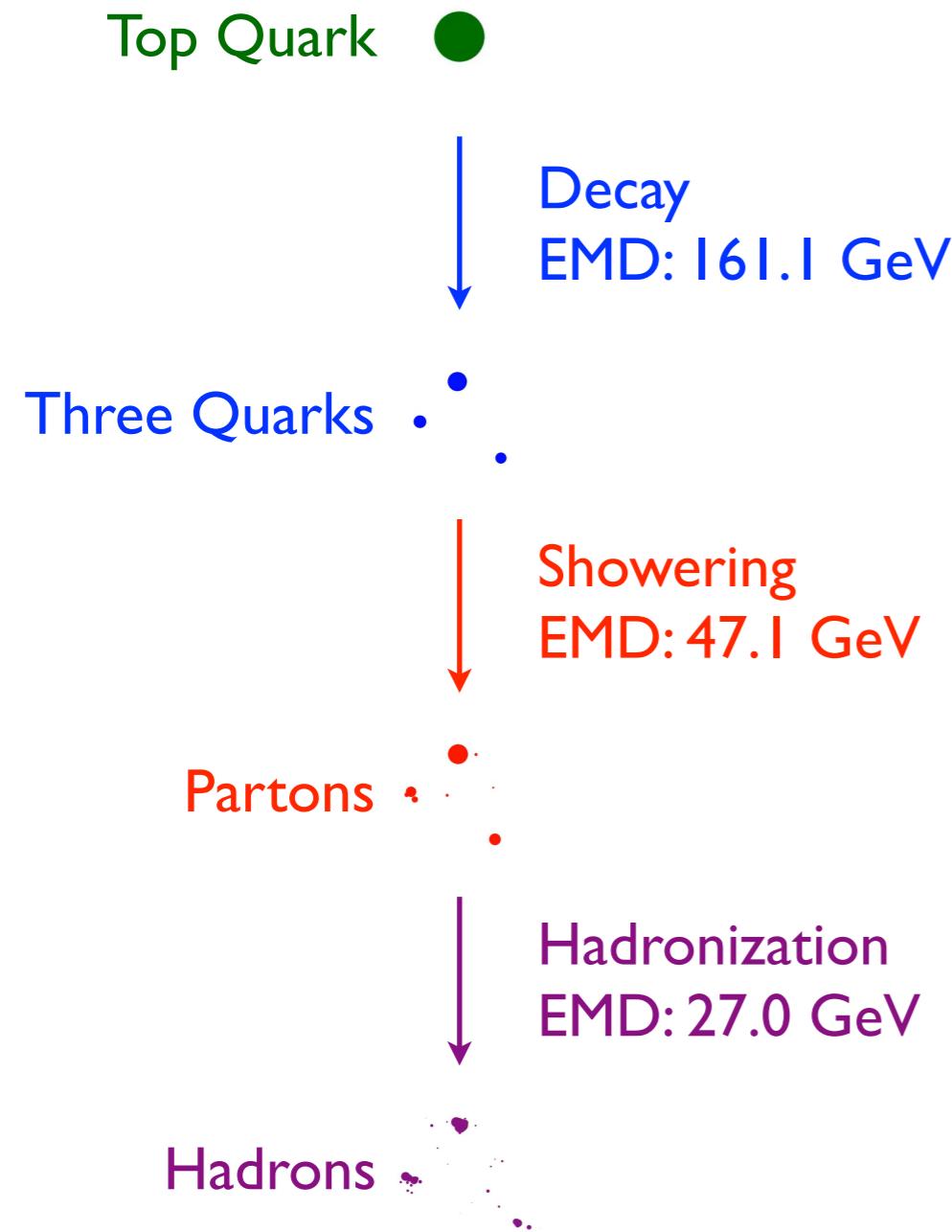
Visualizing W Boson Evolution



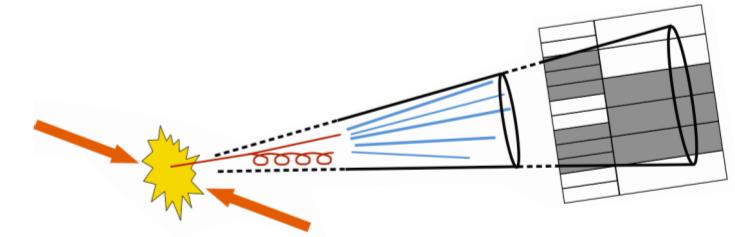
Visualizing Top Quark Evolution



500 GeV



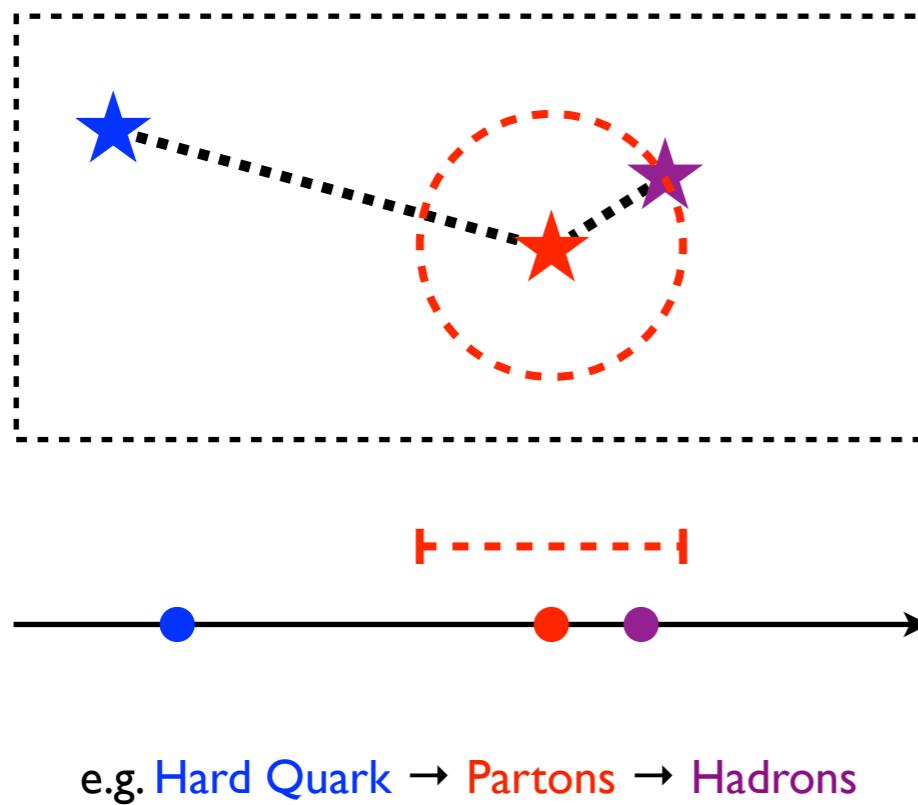
Quantifying Jet Evolution



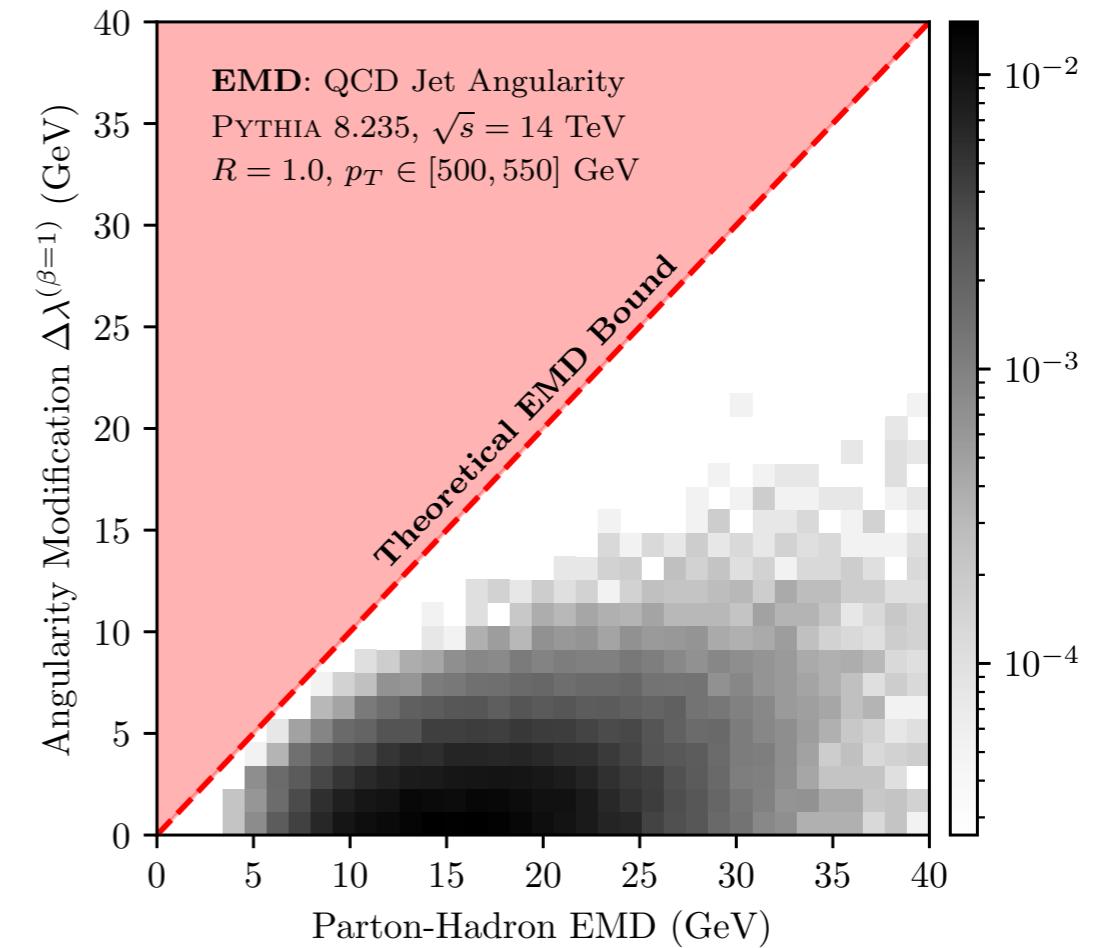
Events **close in EMD** are close in any* **IRC-safe** observable

Using Kantorovich-Rubinstein duality

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

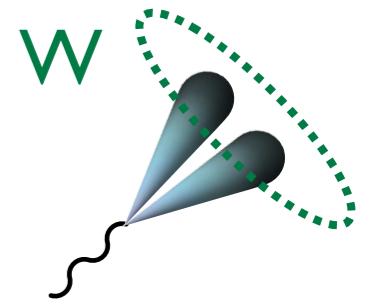


e.g. Angularities ($\beta \geq 1$) $\lambda^{(\beta)}(\mathcal{E}) = \sum_i E_i \theta_i^\beta$

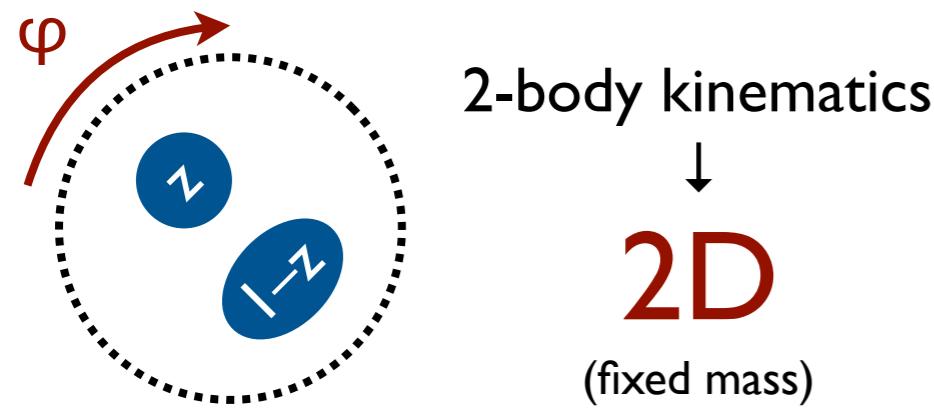


[Komiske, Metodiev, JDT, [1902.02346](#); using Berger, Kucs, Sterman, [hep-ph/0303051](#); Ellis, Vermilion, Walsh, Hornig, Lee, [1001.0014](#); Larkoski, JDT, Waalewijn, [1408.3122](#)]

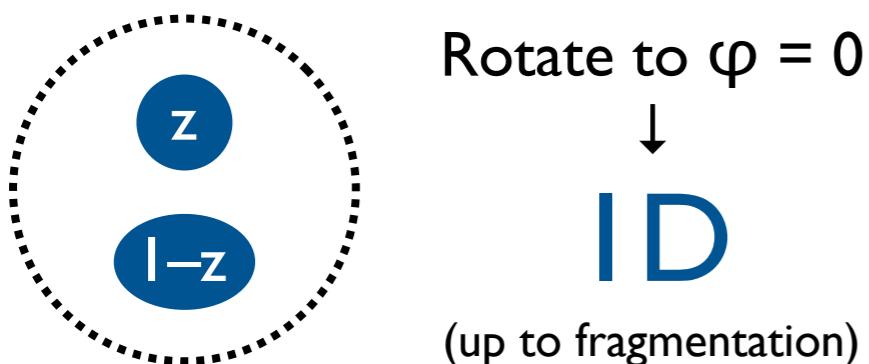
The Space of Boosted W Bosons



*Identify low-D manifold
in high-D space?*

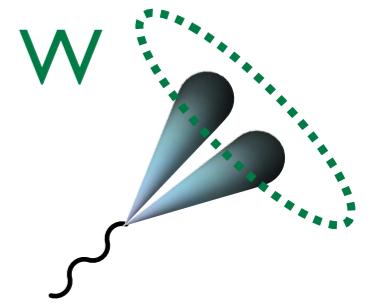


2-body kinematics
↓
2D
(fixed mass)



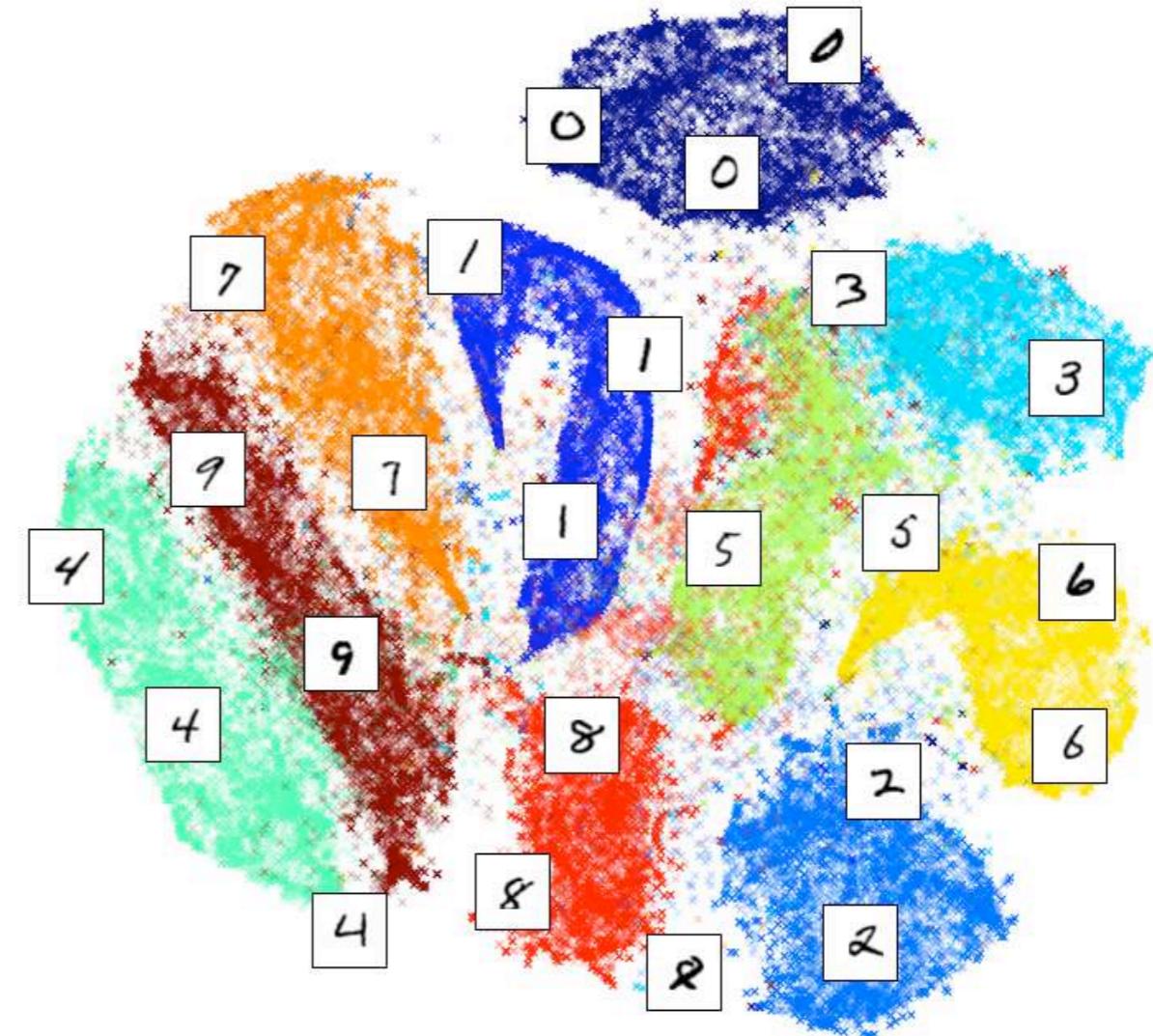
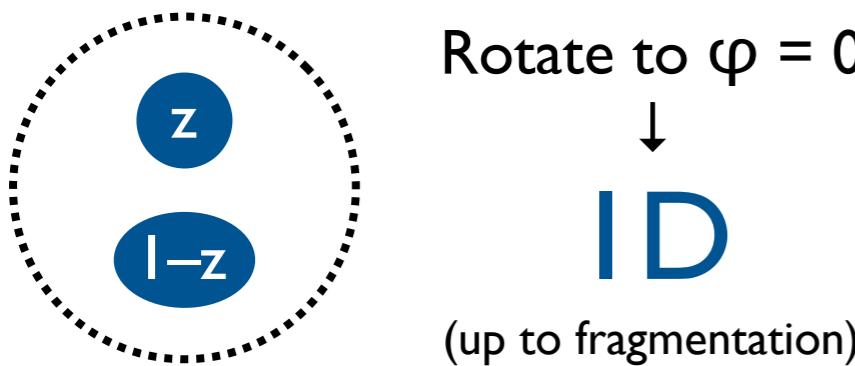
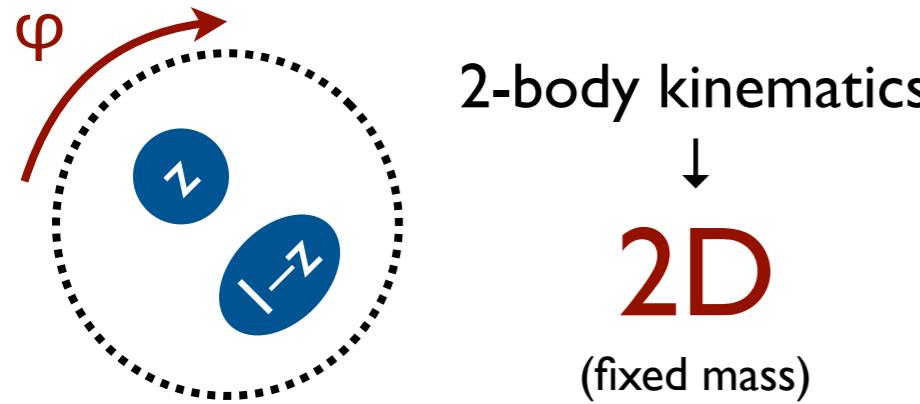
Rotate to $\varphi = 0$
↓
1D
(up to fragmentation)

The Space of Boosted W Bosons



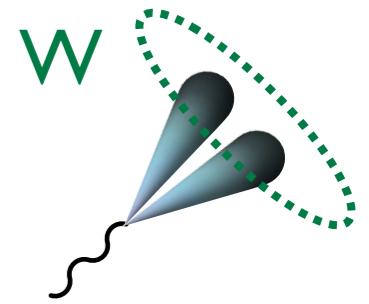
*Identify low-D manifold
in high-D space?*

e.g. t-SNE
(t-distributed Stochastic Neighbor Embedding)

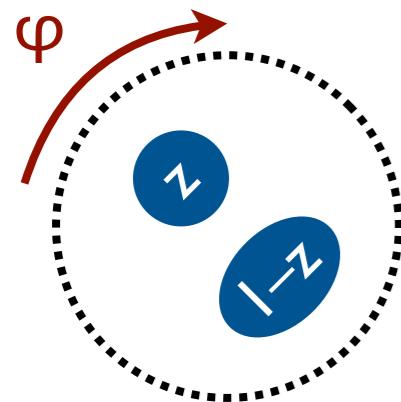


[van der Maaten, Hinton, *JMLR* 2008; figure from [BigSnarf](#)]

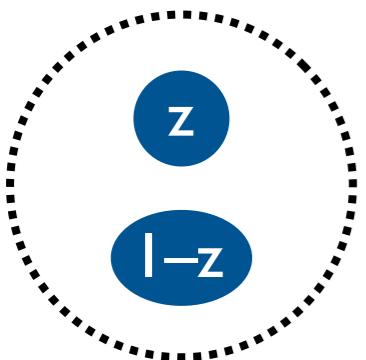
The Space of Boosted W Bosons



*Identify low-D manifold
in high-D space?*

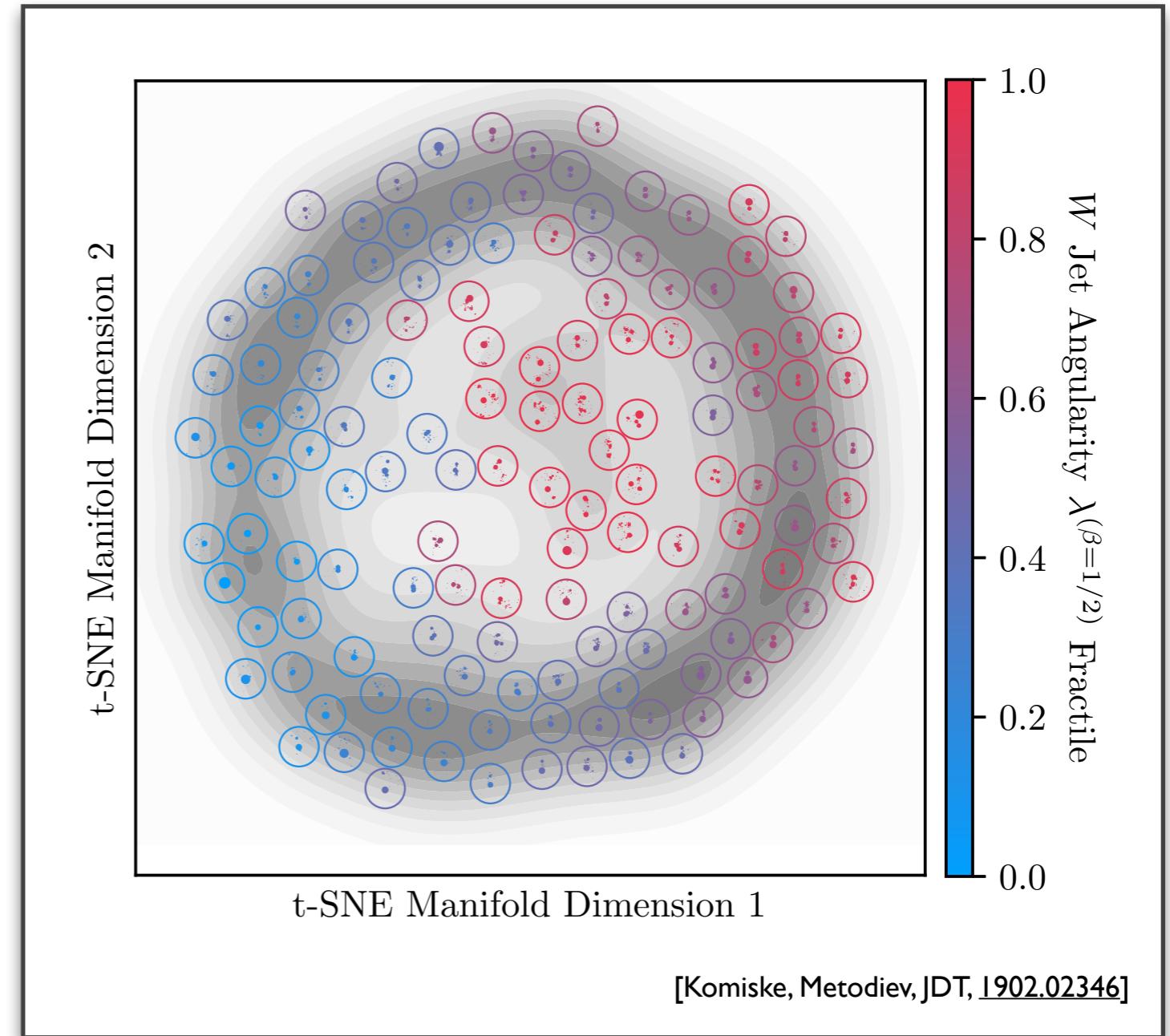


2-body kinematics
↓
2D
(fixed mass)



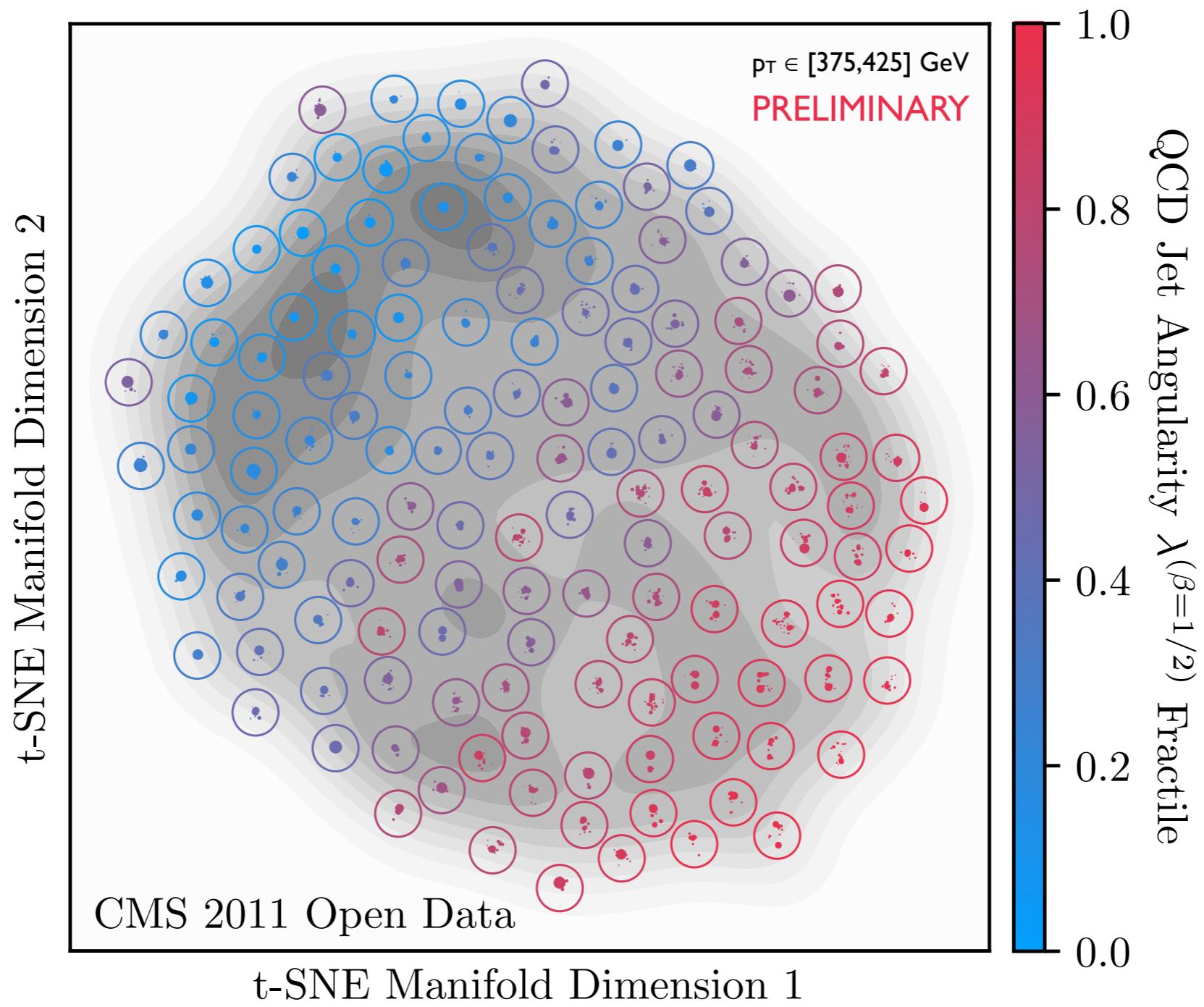
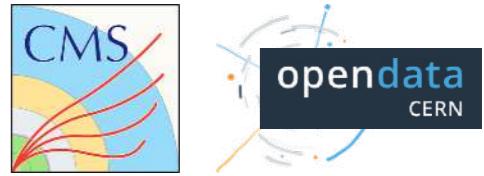
Rotate to $\varphi = 0$
↓
1D
(up to fragmentation)

e.g. t-SNE
(t-distributed Stochastic Neighbor Embedding)

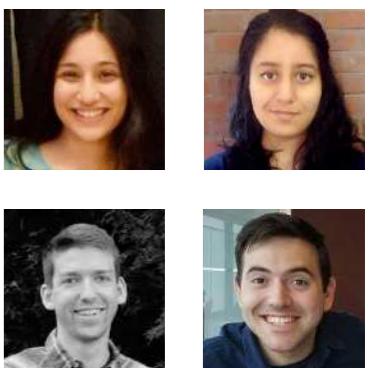


[van der Maaten, Hinton, [JMLR 2008](#); figure from [BigSnarf](#)]

Hot off the Press: The Space of Jets

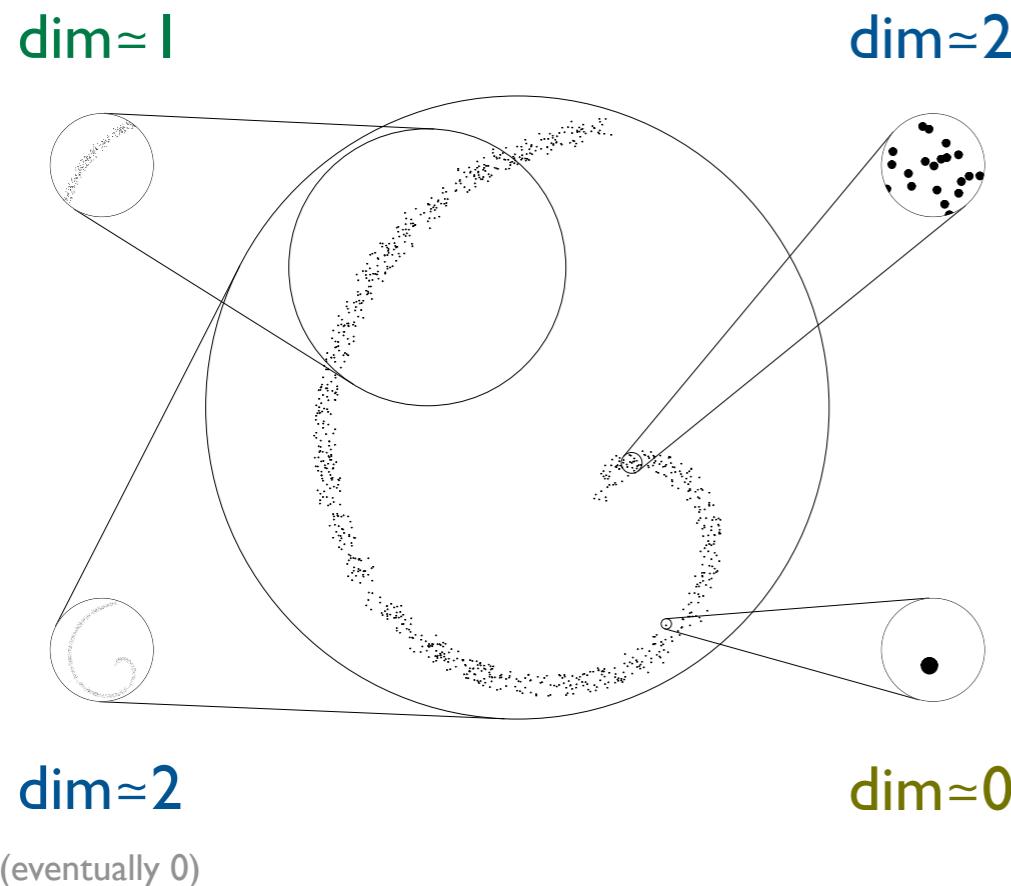


[Mastandrea, Naik, Komiske, Metodiev, JDT, in preparation]



Quantifying Dimensionality

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



$$N_{\text{neighbors}}(r) \sim r^{\dim}$$



$$\dim(r) \sim r \frac{\partial}{\partial r} \ln N_{\text{neighbors}}(r)$$

[Grassberger, Procaccia, [PRL 1983](#); Kégl, [NIPS 2002](#)]

Decay-Level Dimension

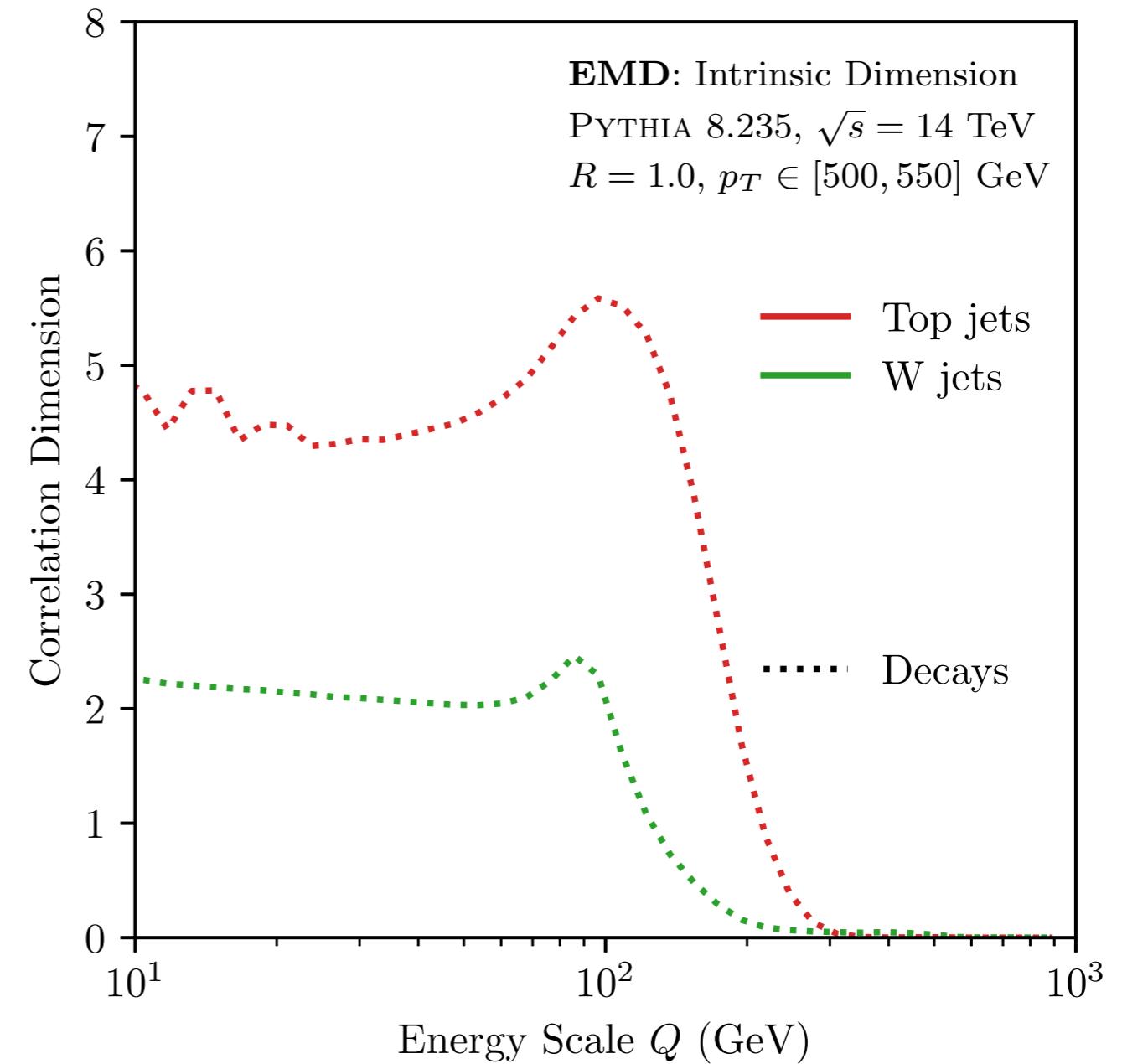
$$\dim(Q) = Q \frac{\partial}{\partial Q} \ln \sum_i \sum_j \Theta(\text{EMD}(\mathcal{E}_i, \mathcal{E}_j) < Q)$$

3-body phase space —

with W mass constraint

2-body phase space —

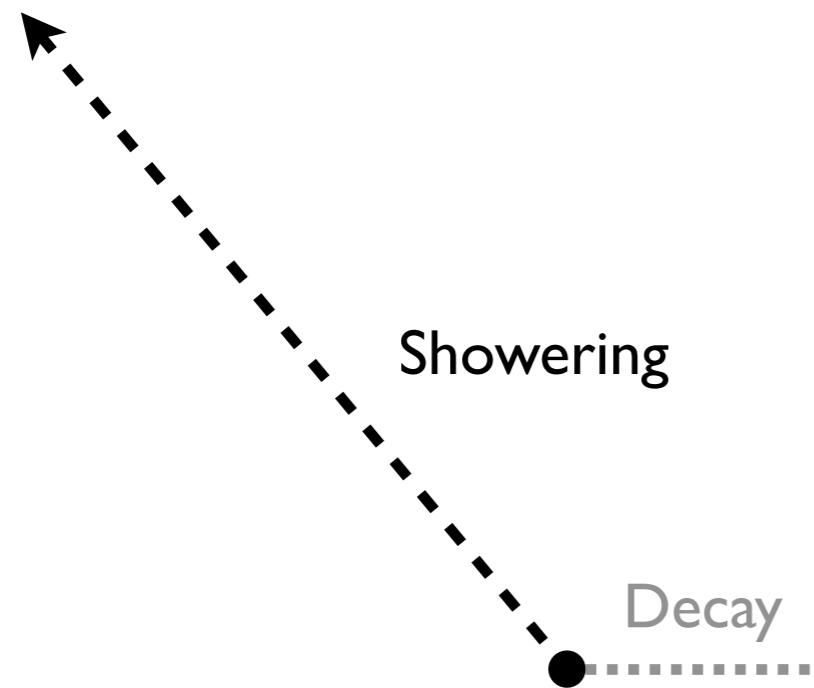
Increasing complexity: multi-body phase space



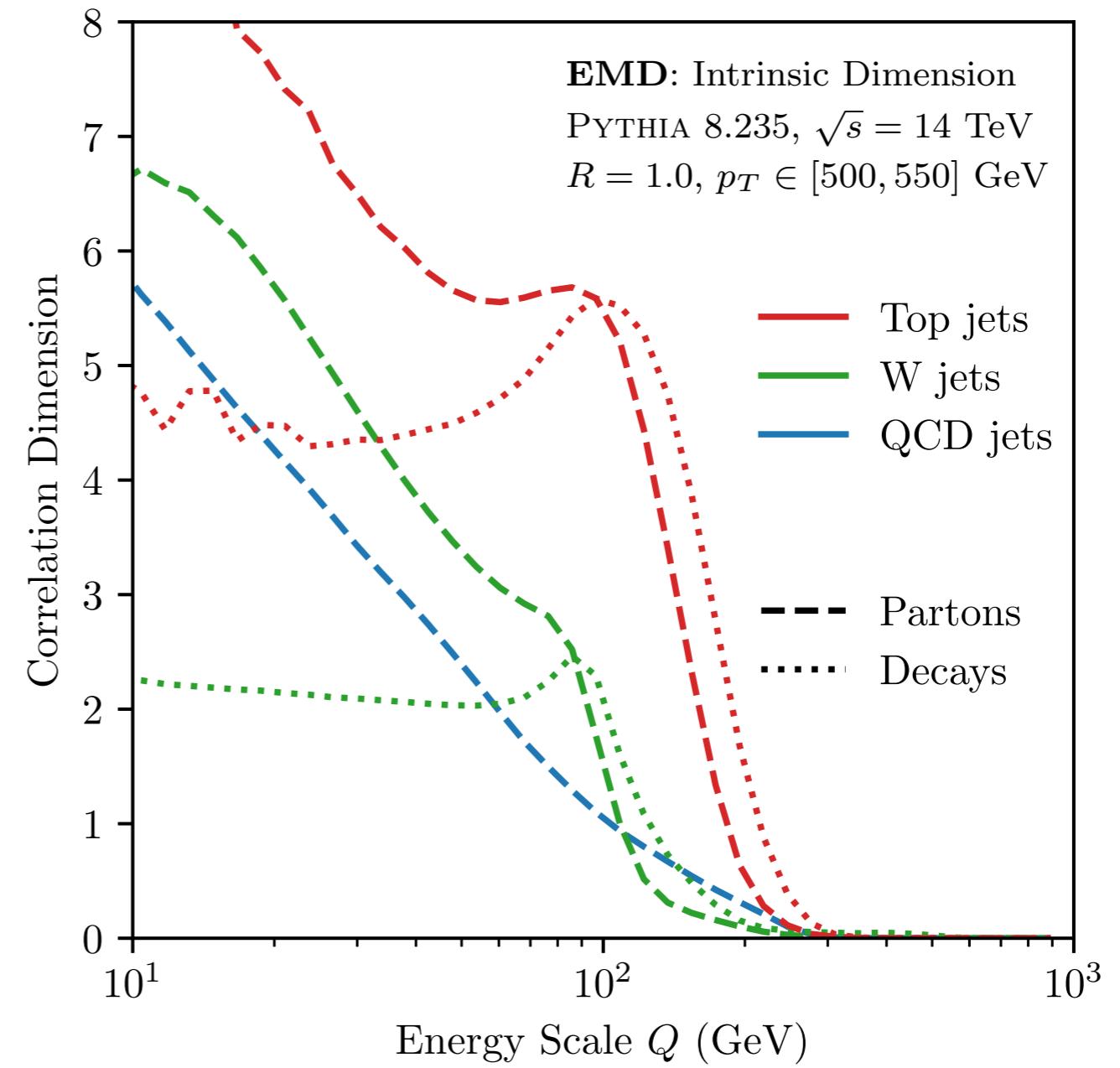
[Komiske, Metodiev, JDT, 1902.02346]

Parton-Level Dimension

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



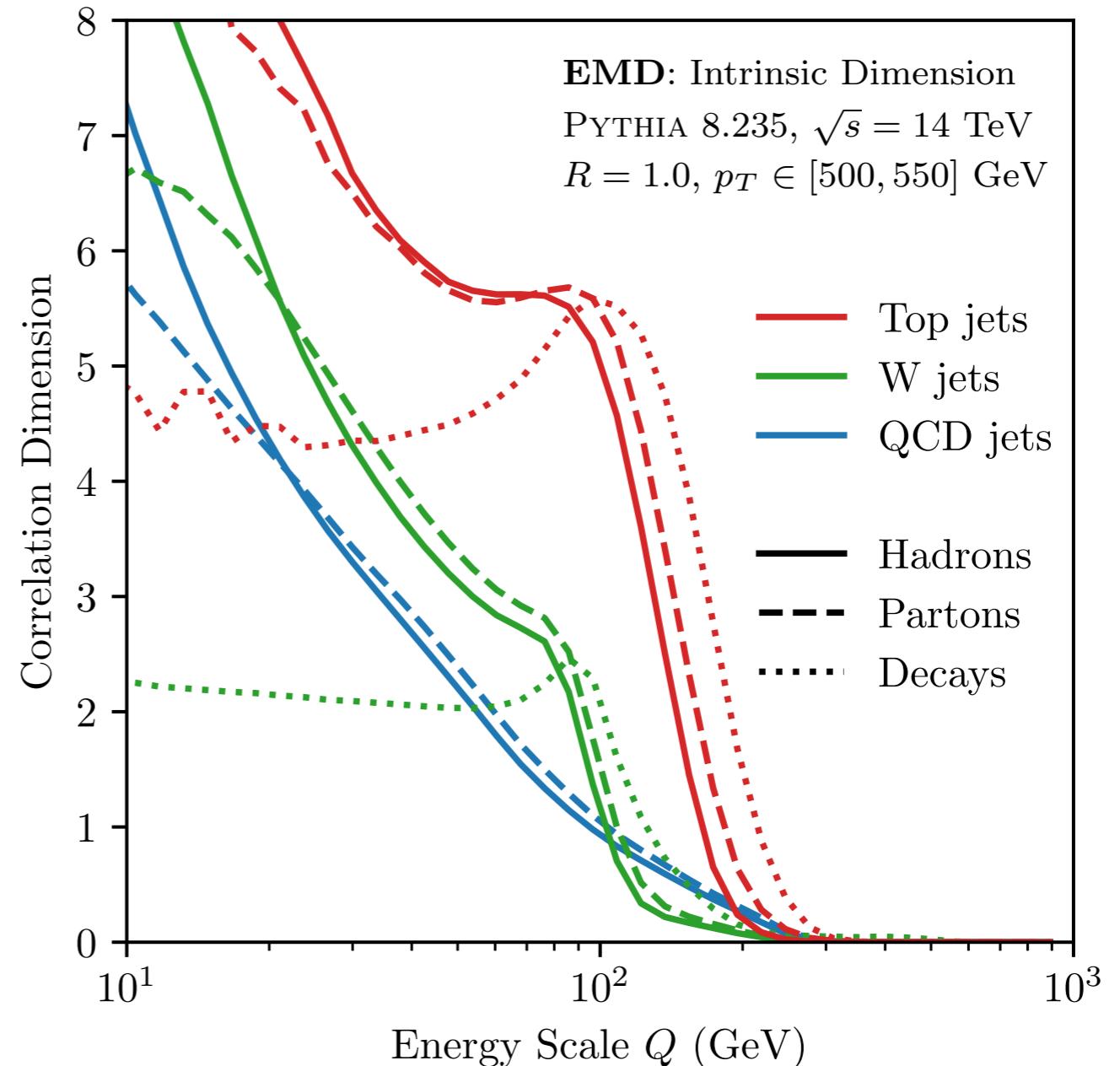
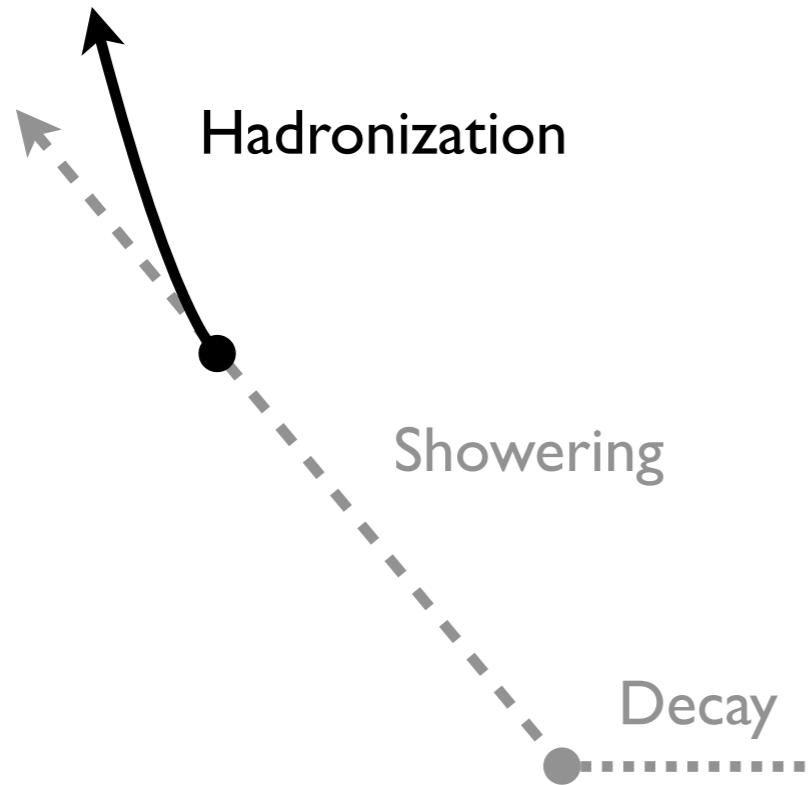
Increasing complexity: multi-body phase space
perturbative emissions



[Komiske, Metodiev, JDT, 1902.02346]

Hadron-Level Dimension

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



Increasing complexity:

- multi-body phase space
- perturbative emissions
- non-perturbative dynamics

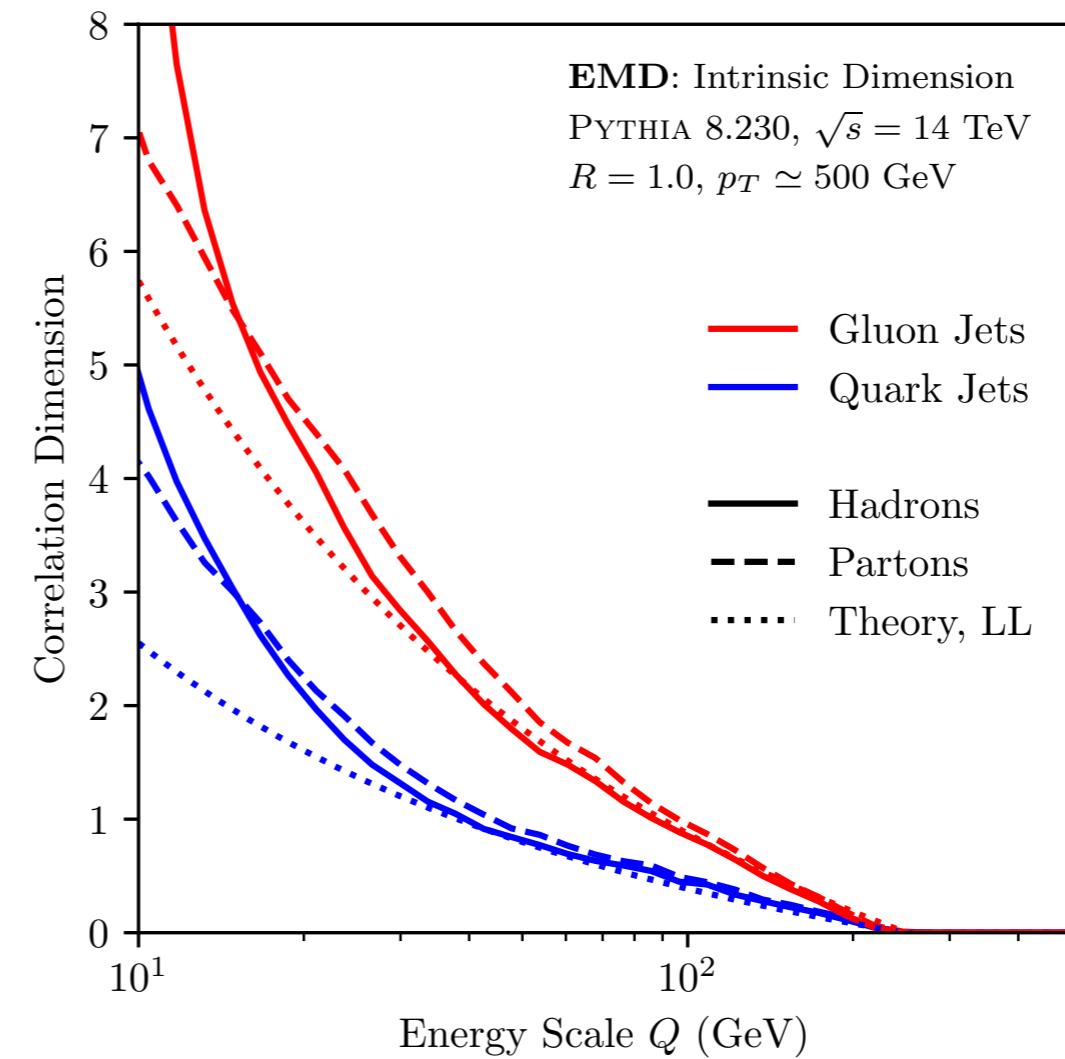
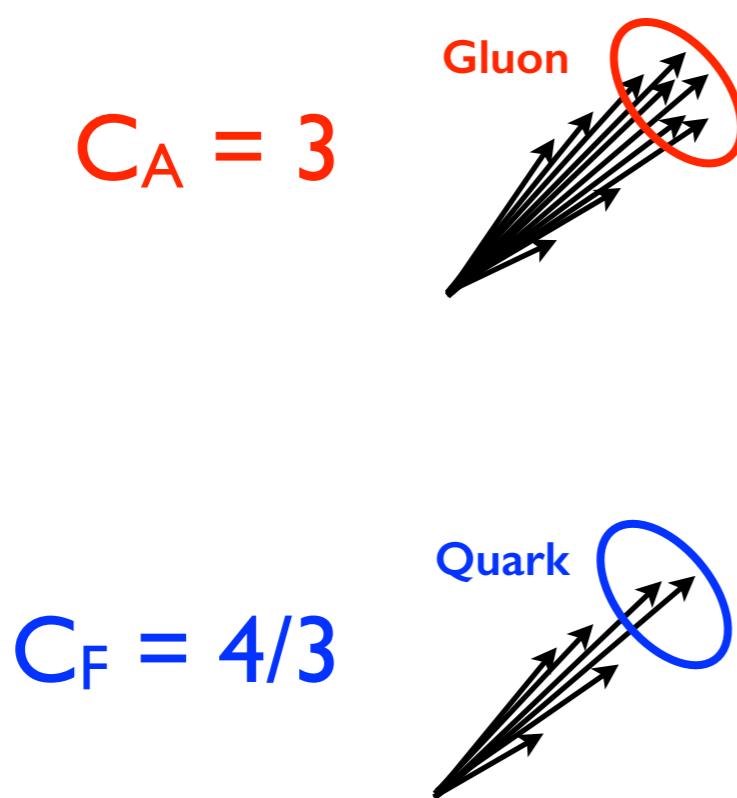
[Komiske, Metodiev, JDT, 1902.02346]

Preliminary Calculation

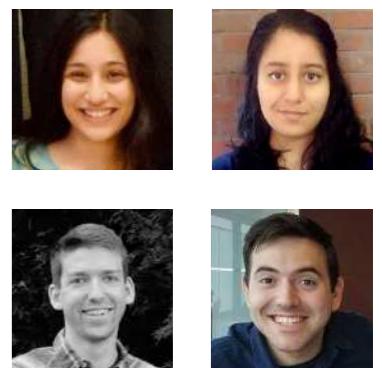
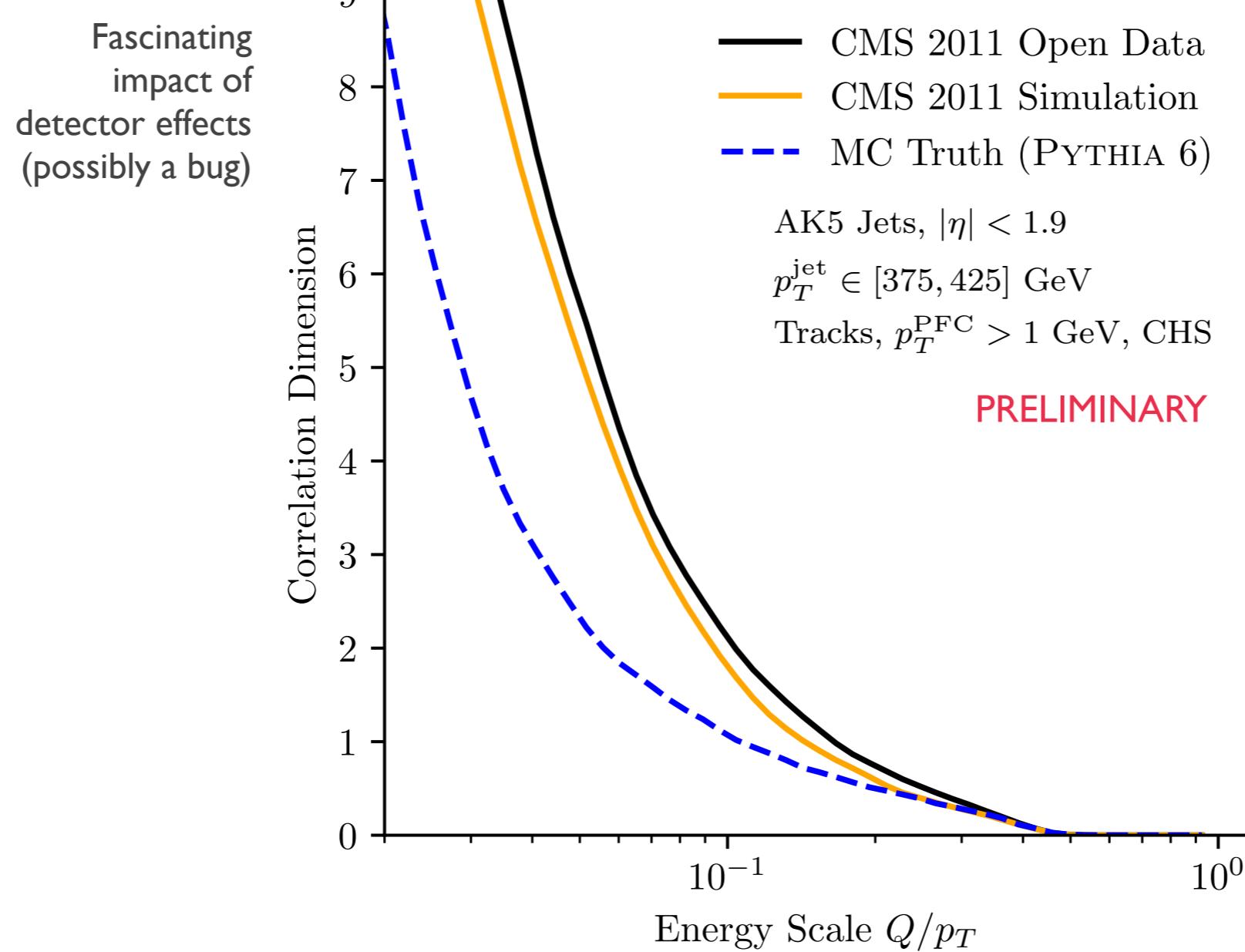
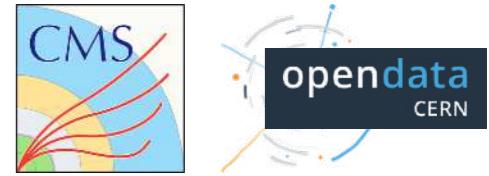
Leading Log:
(single log, since dim has derivative)

$$\dim_i(Q) \sim -\frac{8\alpha_s}{\pi} C_i \ln \frac{Q}{p_T}$$

↑
Color Factor

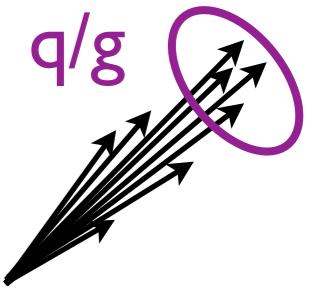


Hot off the Press: Dimension of Jets



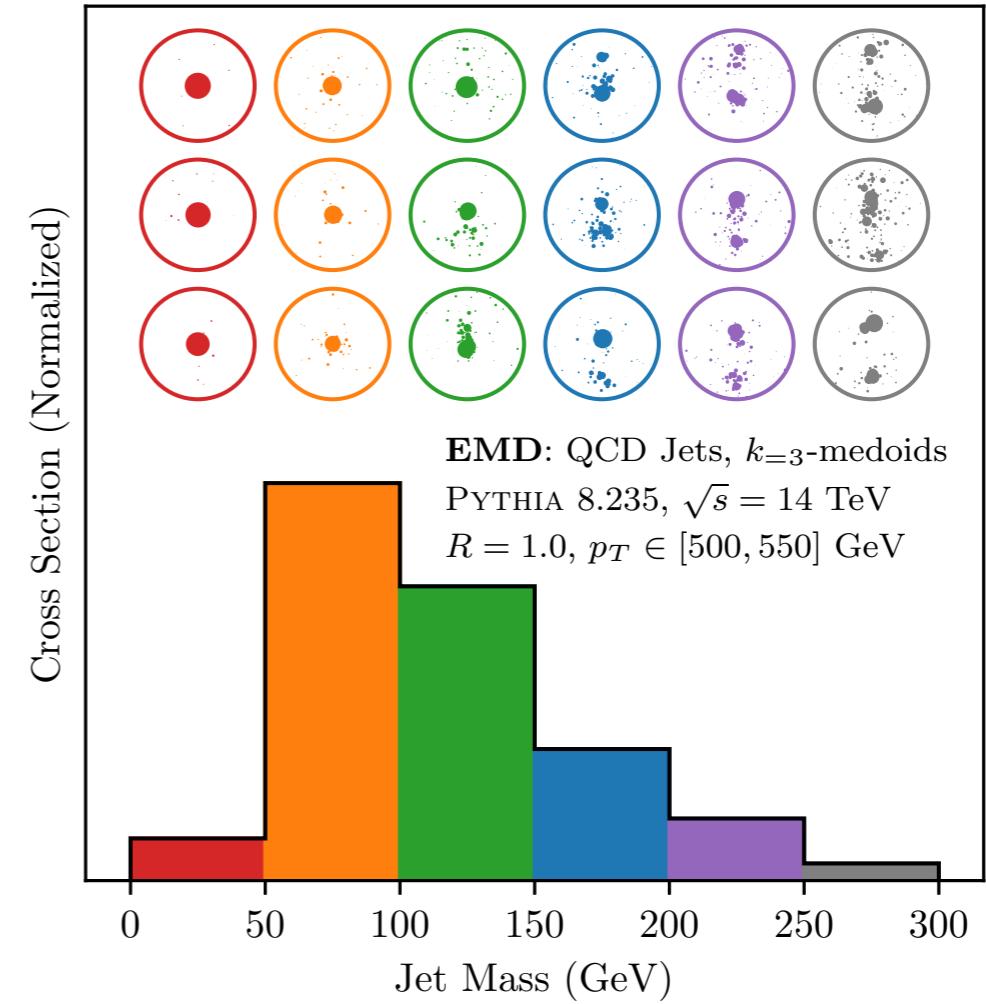
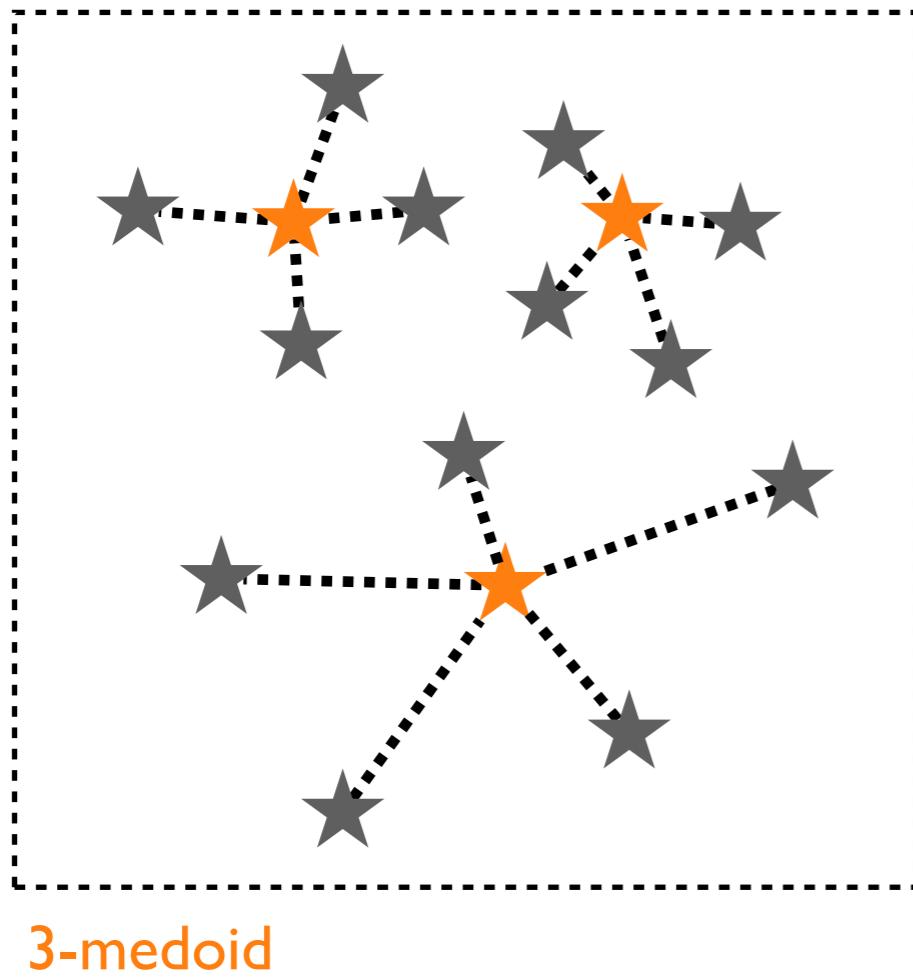
[Mastandrea, Naik, Komiske, Metodiev, JDT, in preparation]

Identifying Representative Jets



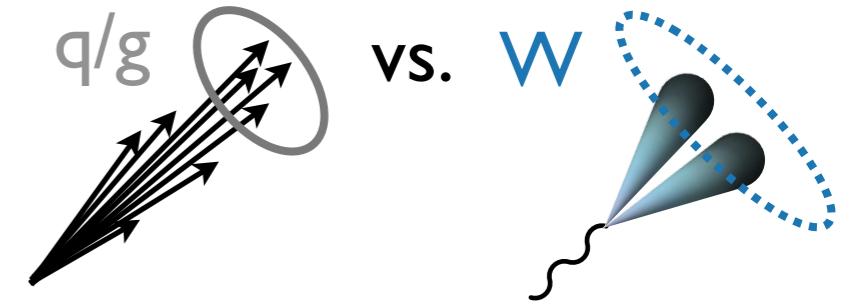
medoid: jet that minimizes total EMD to other jets

k-medoids: k clusters to minimize total medoid distance



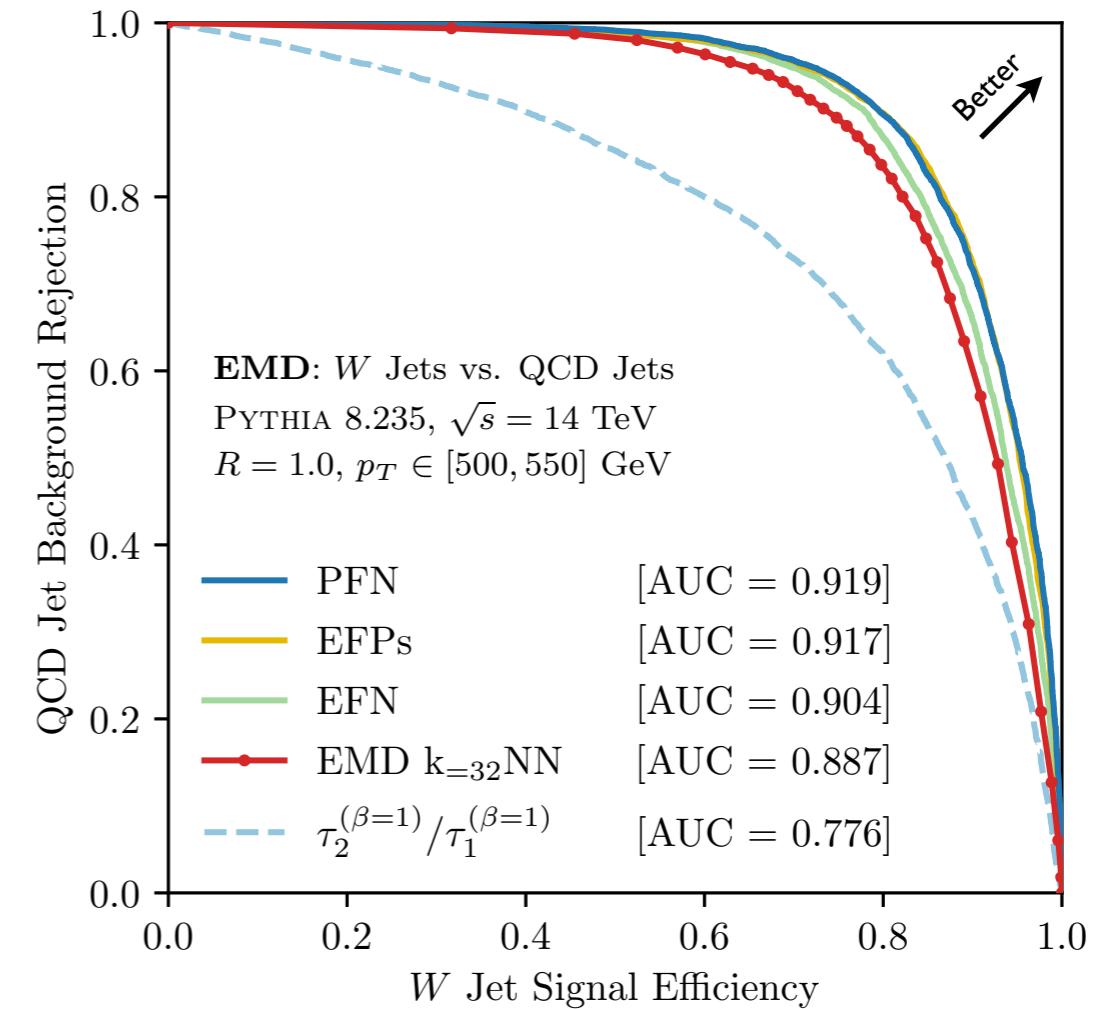
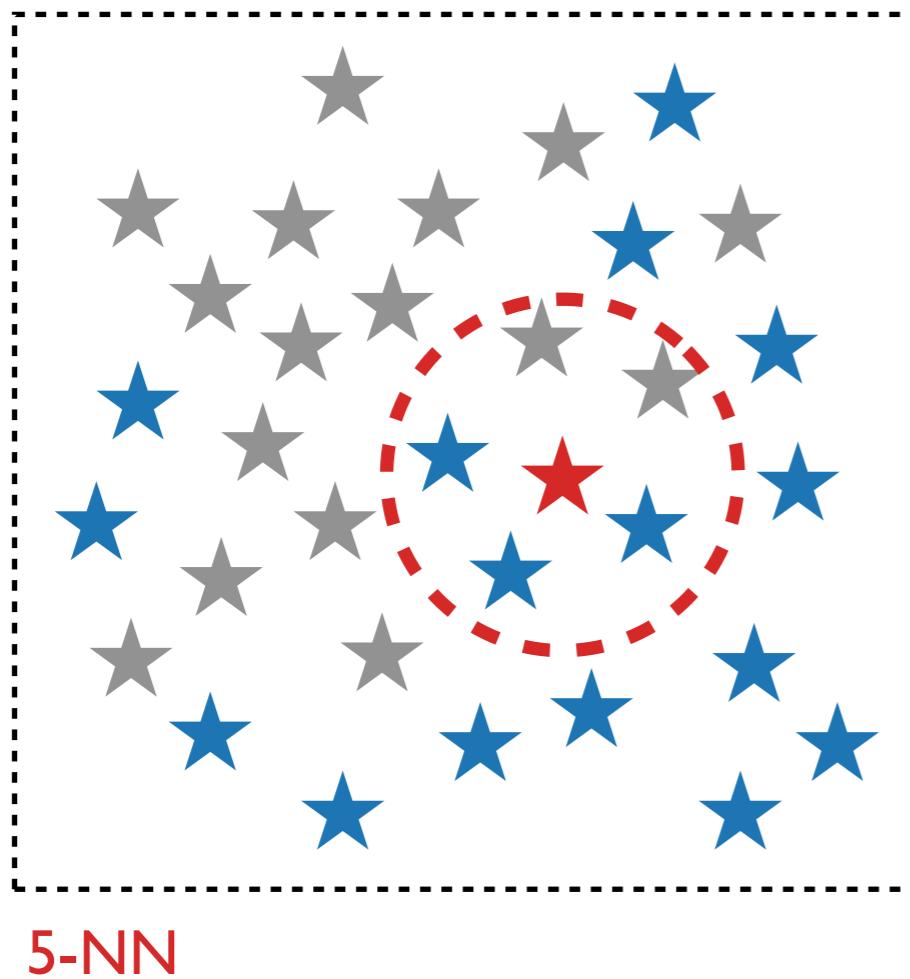
[Komiske, Metodiev, JDT, 1902.02346]

Jet Classification



Estimate jet label by **k nearest neighbors** in training data

Approaches performance of **modern machine learning**



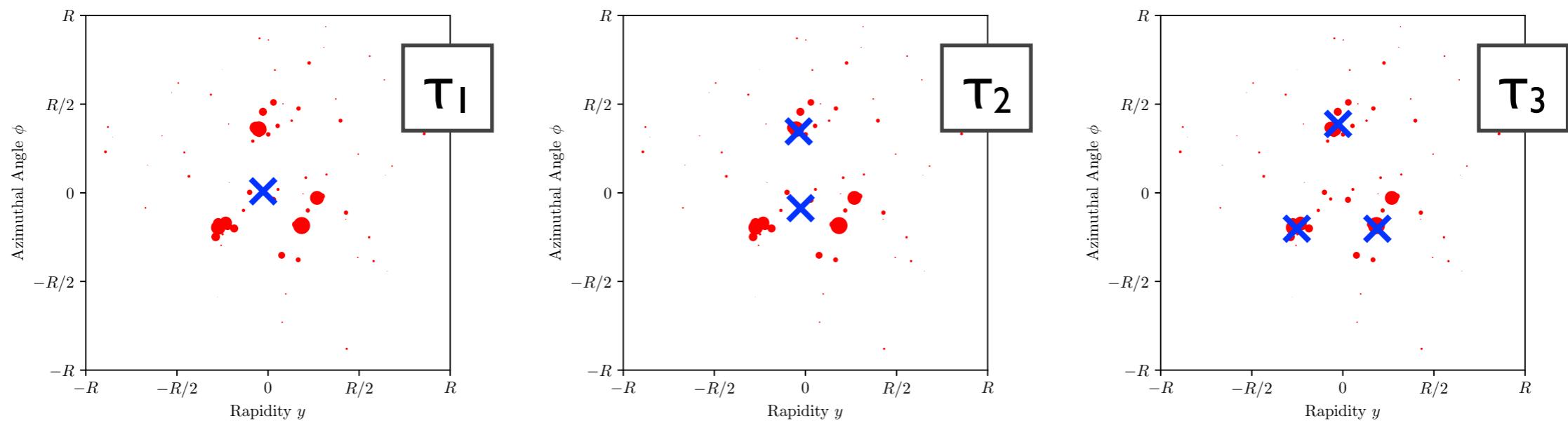
[Komiske, Metodiev, JDT, [1902.02346](#);
comparison to JDT, Van Tilburg, [1011.2268](#), [1108.2701](#); Komiske, Metodiev, JDT, [1712.07124](#), [1810.05165](#)]

Insight into N-subjettiness

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

↑ kind of arbitrary

↑ IRC safe



$$\tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \text{EMD}(\mathcal{E}, \mathcal{E}') \quad \text{for } \beta = 1$$

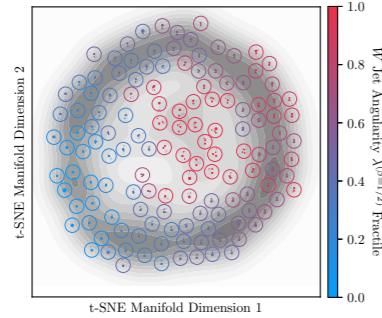
↑ very satisfying

Related to p-Wasserstein metric for $p = \beta > 1$

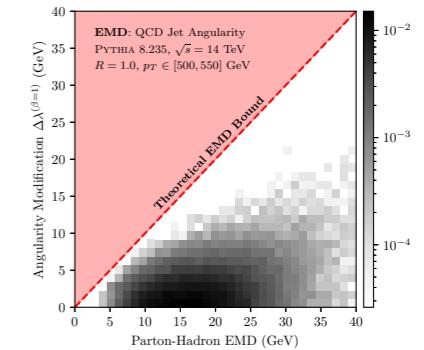
JDT, Van Tilburg, [1011.2268](#), [1108.2701](#);
based on Brandt, Dahmen, [ZPC 1979](#); Stewart, Tackmann, Waalewijn, [1004.2489](#)

Future Directions

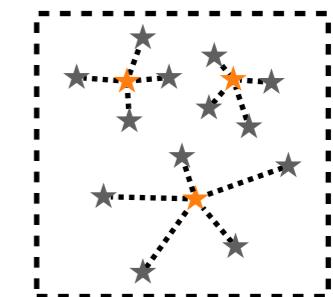
Quantify (and mitigate?) pileup/detector effects



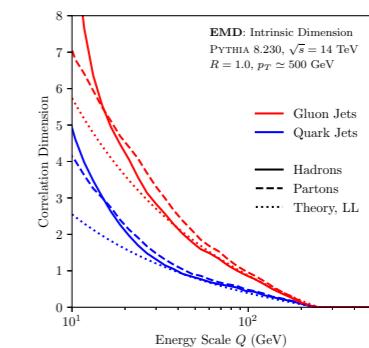
Non-parametric density estimates (& unfolding?)



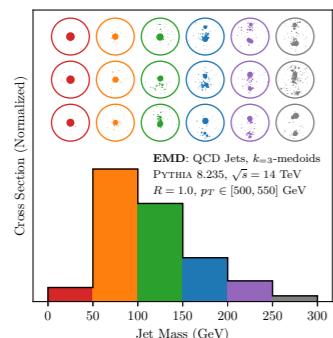
Automated data compression (& triggering?)



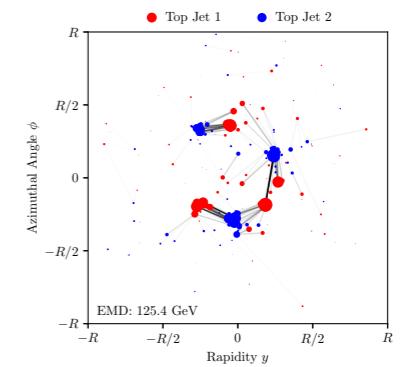
Define new observables through EMD geometry



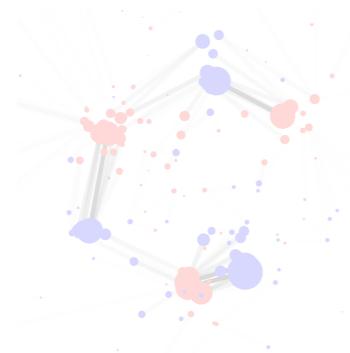
Precision QCD calculations of pairwise metrics



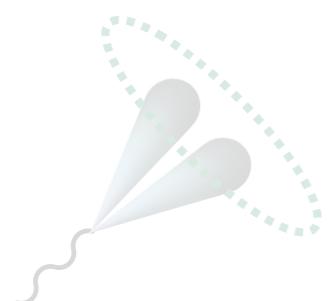
Novel analyses with and beyond histograms



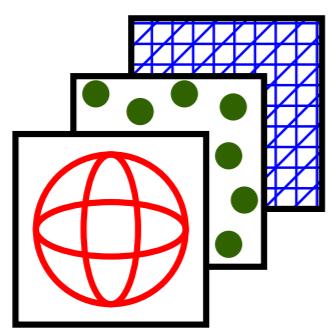
...



The Energy Mover's Distance



Initial Physics Studies



(Broader Comments)

High-Energy
Particle
Physics

Mathematics,
Statistics,
Computer Science



High-Energy Particle Physics



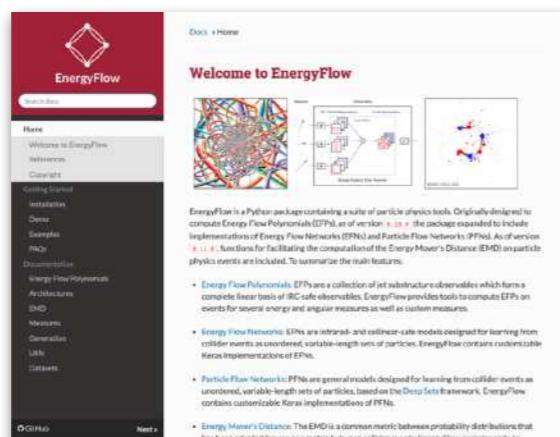
Patrick Komiske



Eric Metodiev



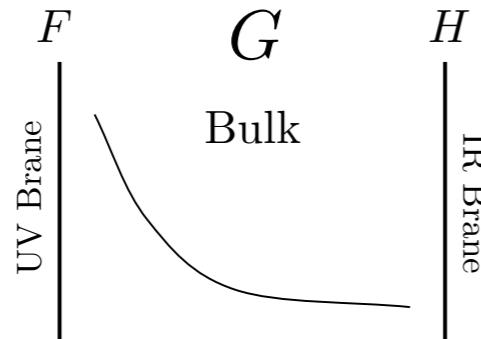
Mathematics,
Statistics,
Computer Science



Energy Flow Package

<https://energyflow.network/>

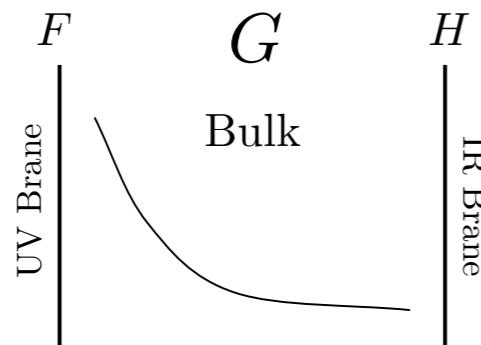
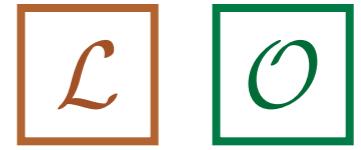
Evolution of a “Model Builder”



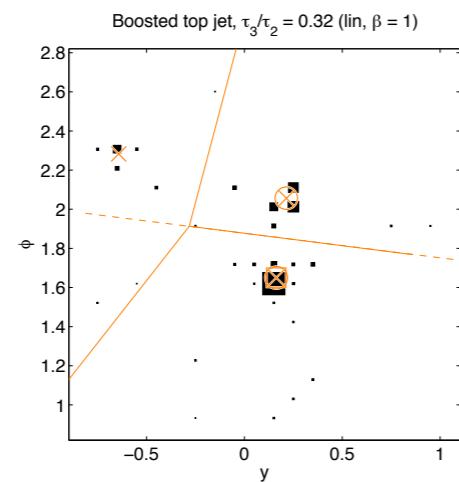
Build Lagrangians & Paradigms

[images from JDT, [hep-ph/0502175](#); JDT, Van Tilburg, [1108.2701](#); Komiske, Metodiev, JDT, [1810.05165](#)]

Evolution of a “Model Builder”



Build Lagrangians & Paradigms

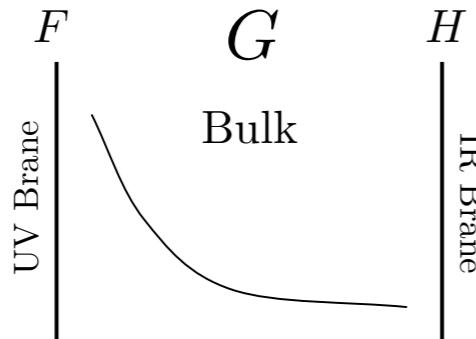


Build Observables & Algorithms

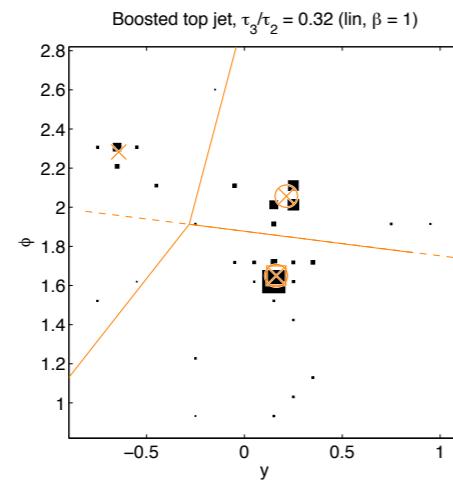
[images from JDT, [hep-ph/0502175](#); JDT, Van Tilburg, [1108.2701](#); Komiske, Metodiev, JDT, [1810.05165](#)]

Evolution of a “Model Builder”

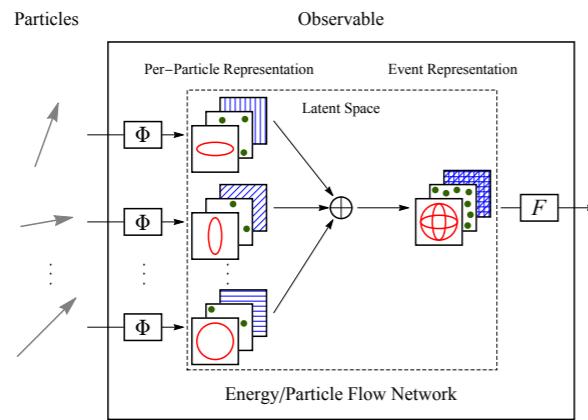
\mathcal{L} \mathcal{O} \mathcal{M}



Build Lagrangians & Paradigms



Build Observables & Algorithms



Build Models & Loss Functions

[images from JDT, [hep-ph/0502175](#); JDT, Van Tilburg, [1108.2701](#); Komiske, Metodiev, JDT, [1810.05165](#)]

Evolution of a “Model Builder”



Given current status of the LHC, which strategy:

*Makes maximal **verifiable** use of collider data?*

Can scale up to the challenges of HL-LHC?

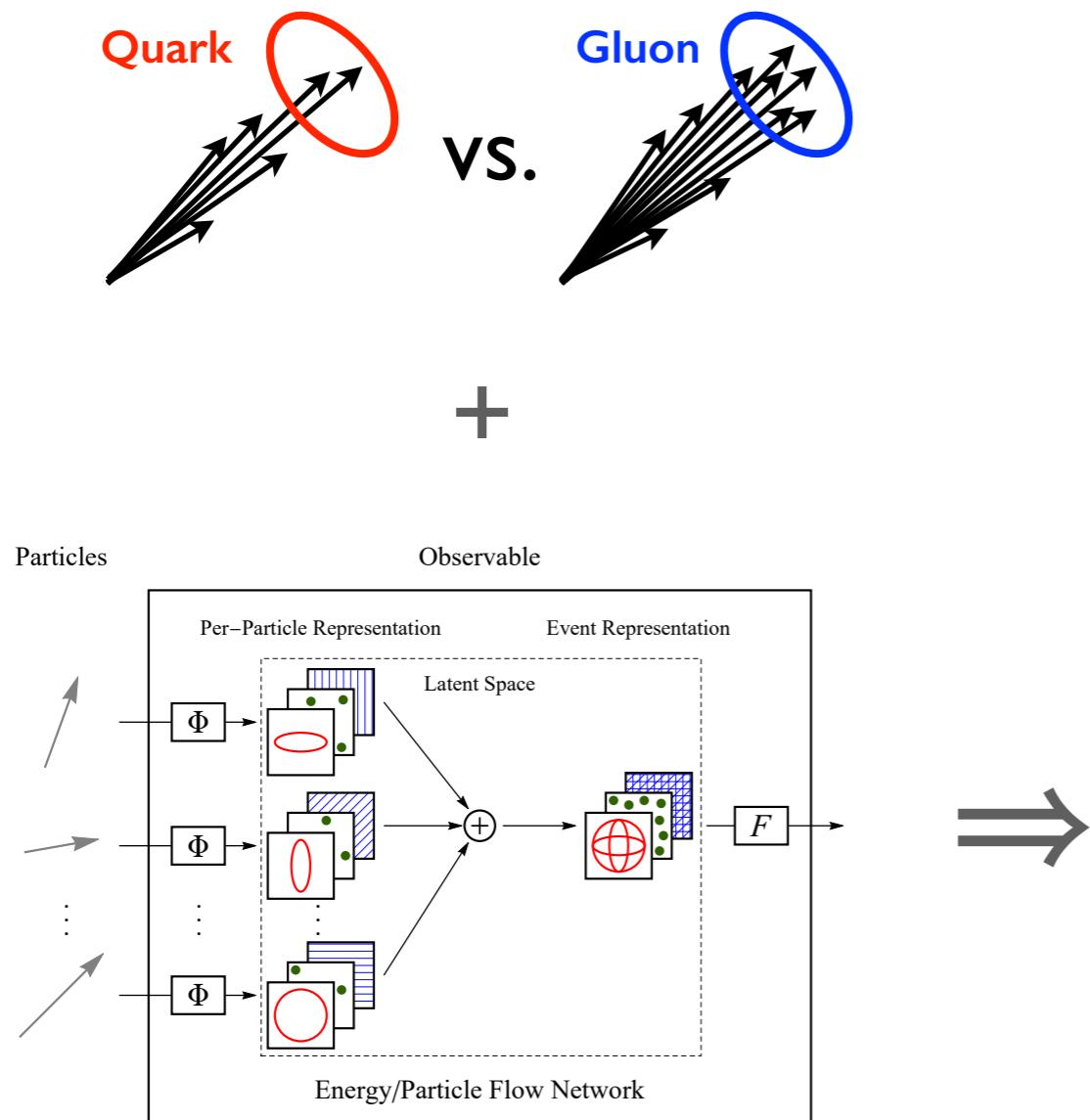
*Has the greatest **discovery potential**?*

*Offers new **insights** into fundamental physics?*

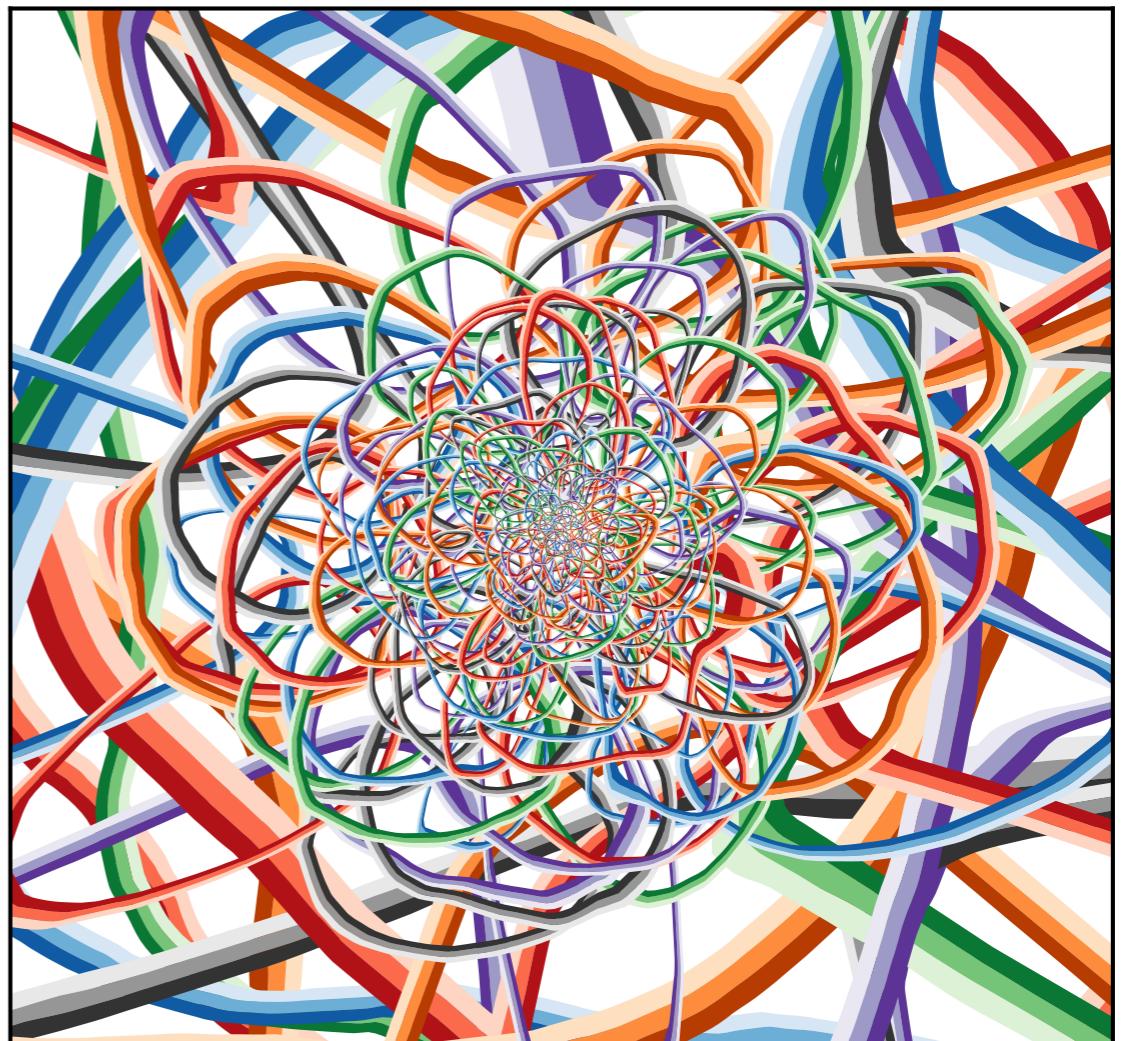
[images from JDT, [hep-ph/0502175](#); JDT, Van Tilburg, [1108.2701](#); Komiske, Metodiev, JDT, [1810.05165](#)]

Opportunities for Visualization

\mathcal{L} \mathcal{O} \mathcal{M}

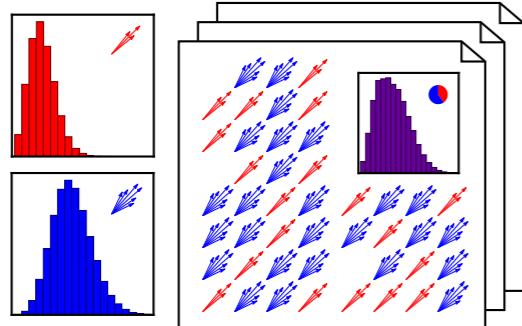


“Seeing” the collinear singularity of QCD



Augmenting our exceptional human/scientific ability to recognize patterns

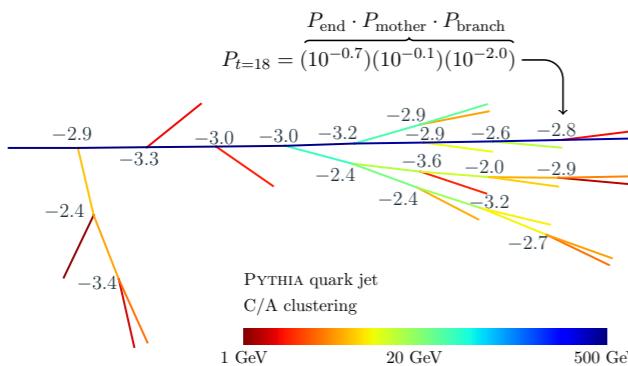
The Rise of Unsupervised Learning



Jet Topics

Blind Source Separation

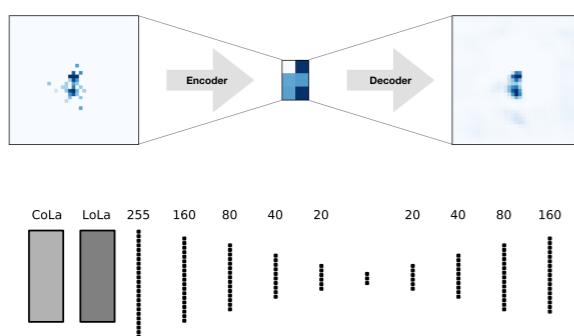
[Metodiev, JDT, [I802.00008](#); Komiske, Metodiev, JDT, [I809.01140](#); see also Metodiev, Nachman, JDT, [I708.02949](#)]



JUNIPR

Probability Modeling

[Andreassen, Feige, Frye, Schwartz, [I804.09720](#); see also Monk, [I807.03685](#)]



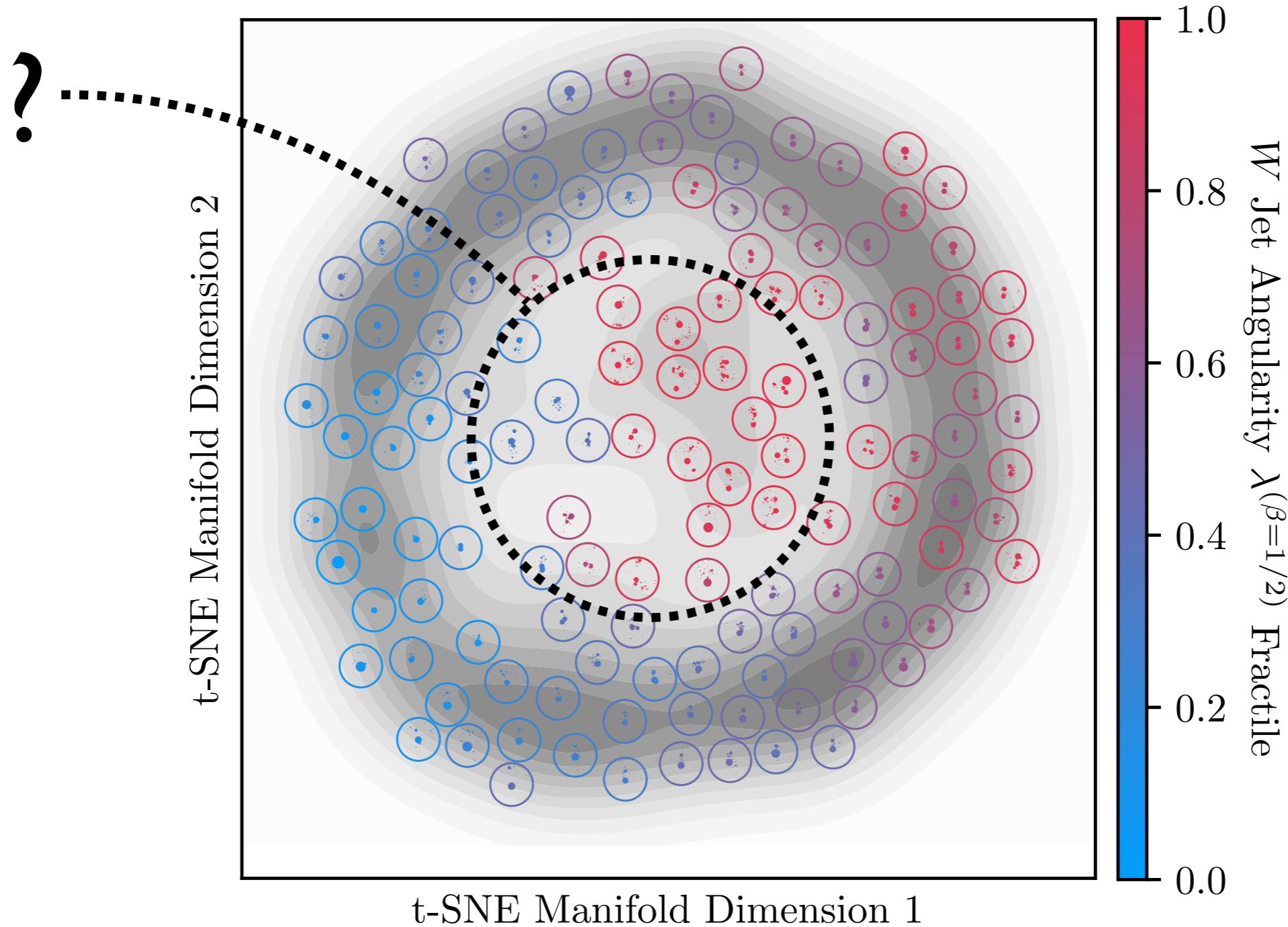
Autoencoders

Anomaly Detection

[Hajer, Li, Liu, Wang, [I807.10261](#); Heimel, Kasieczka, Plehn, Thompson, [I808.08979](#); Farina, Nakai, Shih, [I808.08992](#); Cerri, Nguyen, Pierini, Spiropulu, Vlimant, [I811.10276](#); see also Collins, Howe, Nachman, [I805.02664](#), [I902.02634](#); De Simone, Jacques, [I807.06038](#)]

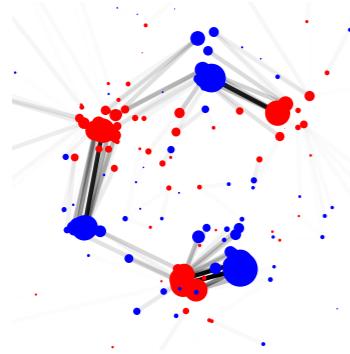
Common theme: Analyze *event ensembles*, not individual events

Visualization meets Anomaly Detection

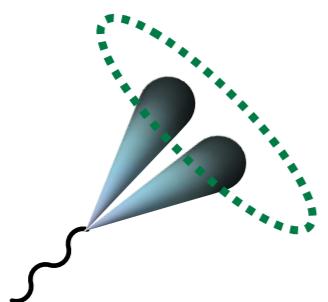


[Komiske, Metodiev, JDT, 1902.02346]

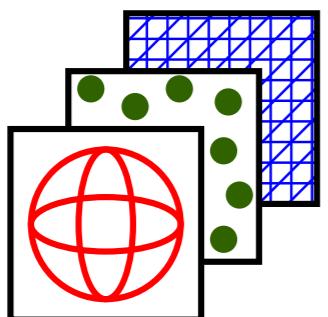
Summary



The Energy Mover's Distance
A metric on the space of IRC-safe energy flows



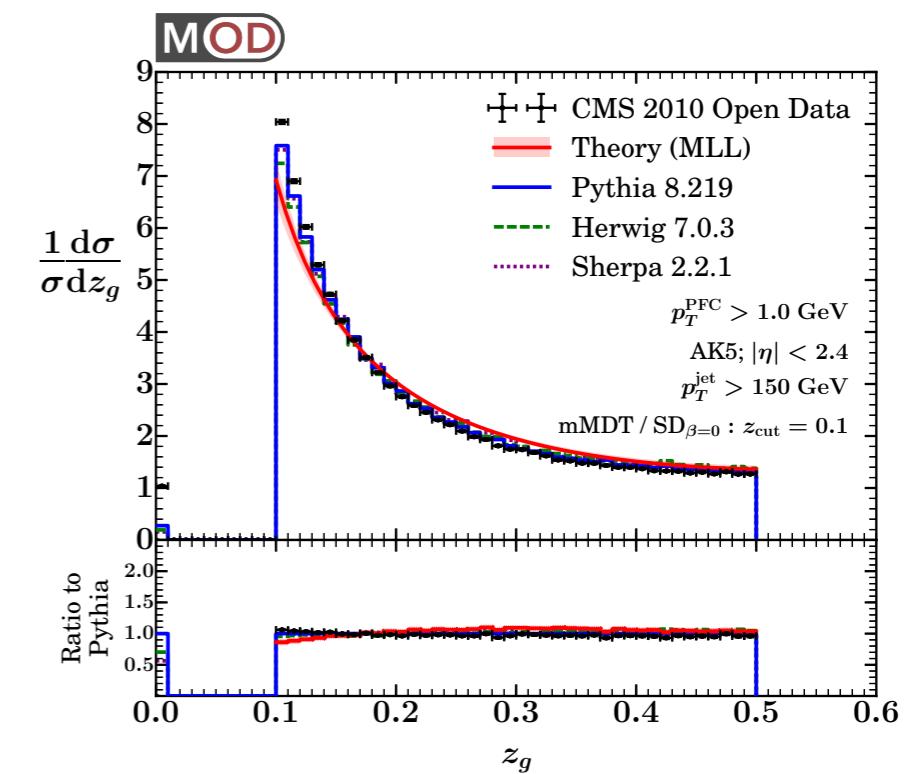
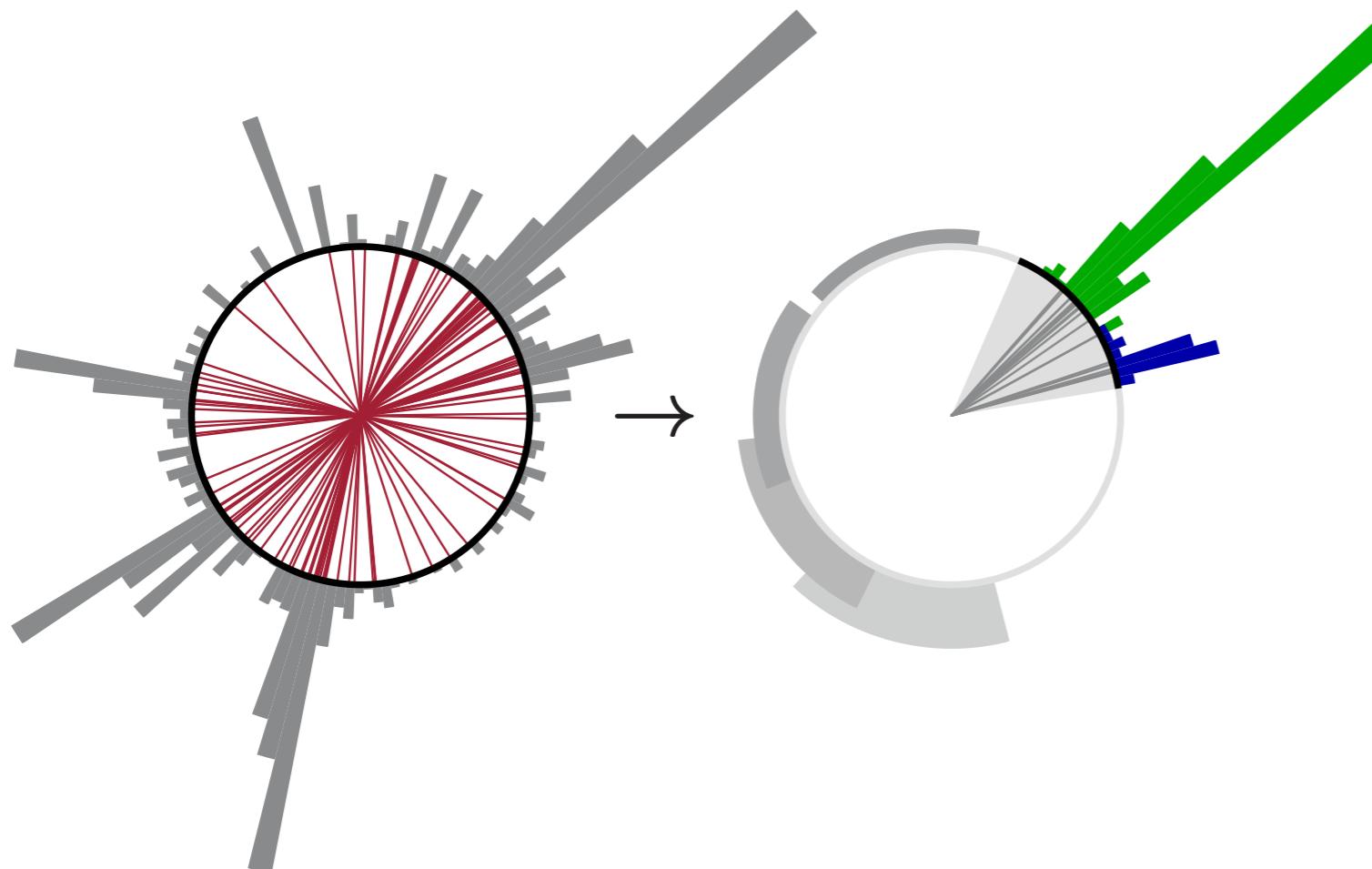
Initial Physics Studies
A geometric language for collider data exploration



(Broader Comments)
Machine learning as a pathway to LHC discoveries

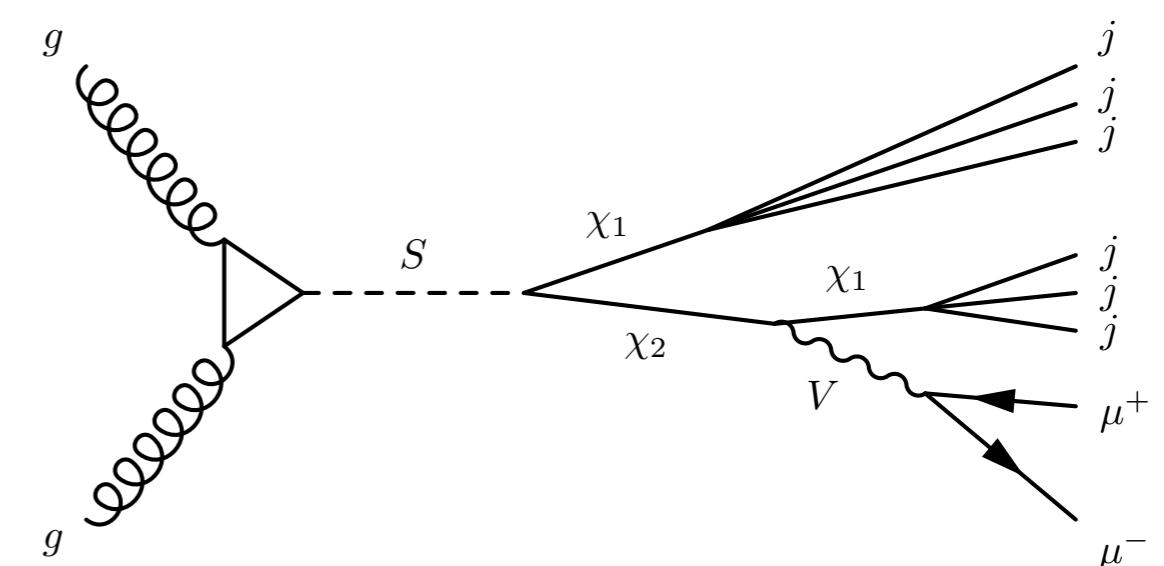
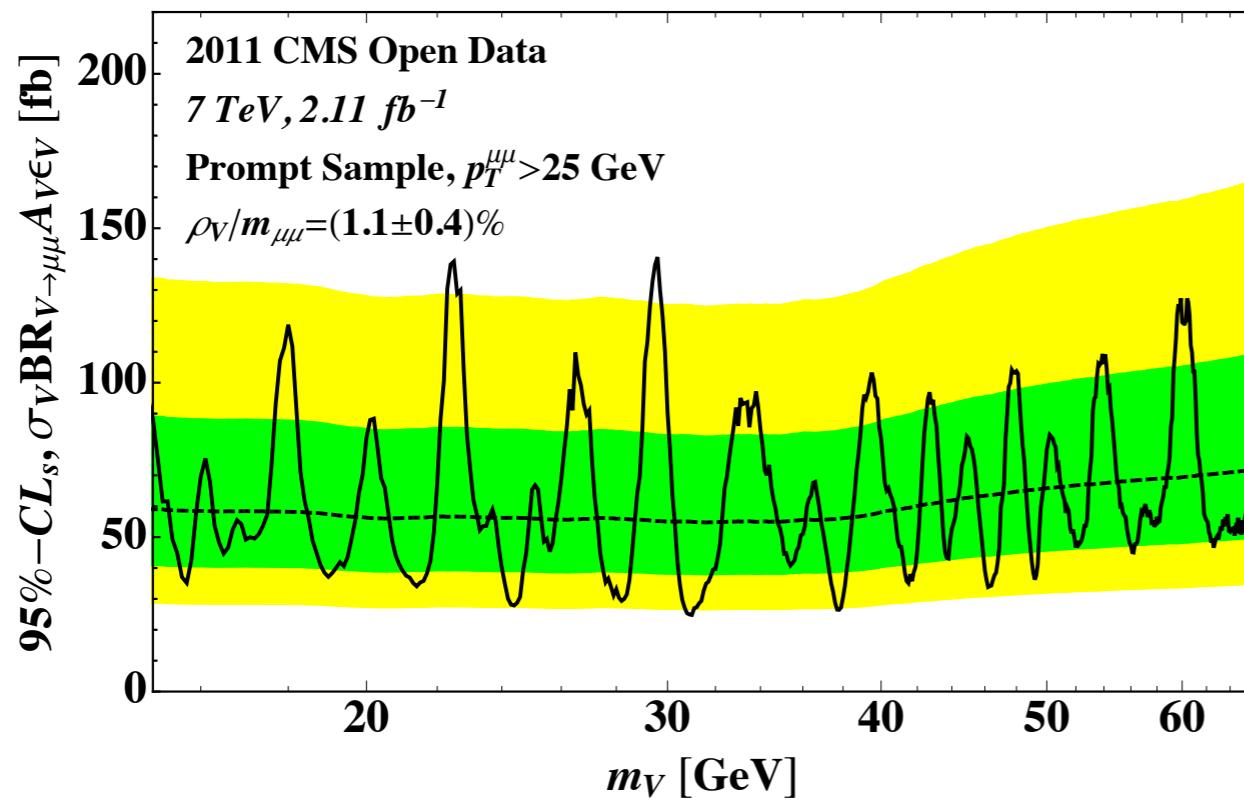
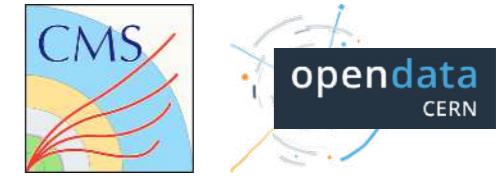
Backup Slides

Event Displays vs. Histograms ...



[Larkoski, Marzani, JDT, Tripathee, Xue, [I704.05066](#)]

... vs. Brazil Plots vs. Cartoons



[Cesarotti, Soreq, Strassler, JDT, Xue, [1902.04222](#)]