

Collision Course

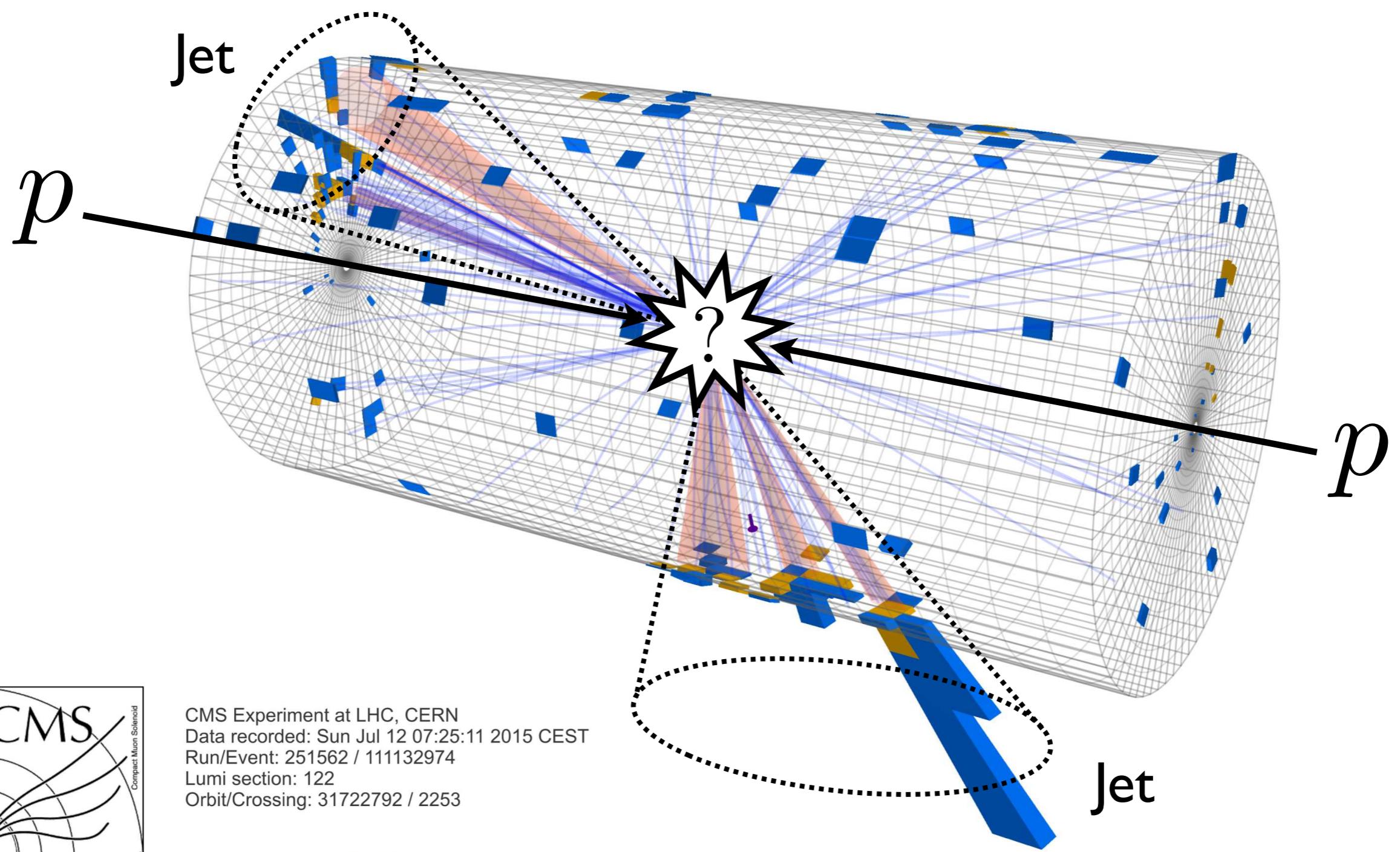
Particle Physics meets Machine Learning

Jesse Thaler



Case Western Reserve Physics Colloquium — November 14, 2019

“Collision Course”

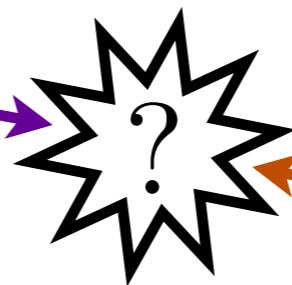


“Collision Course”

“*Theoretical Physics for Machine Learning*”
Aspen Center for Physics, January 2019

Theoretical
(High Energy)
Physics

Mathematics,
Statistics,
Computer Science

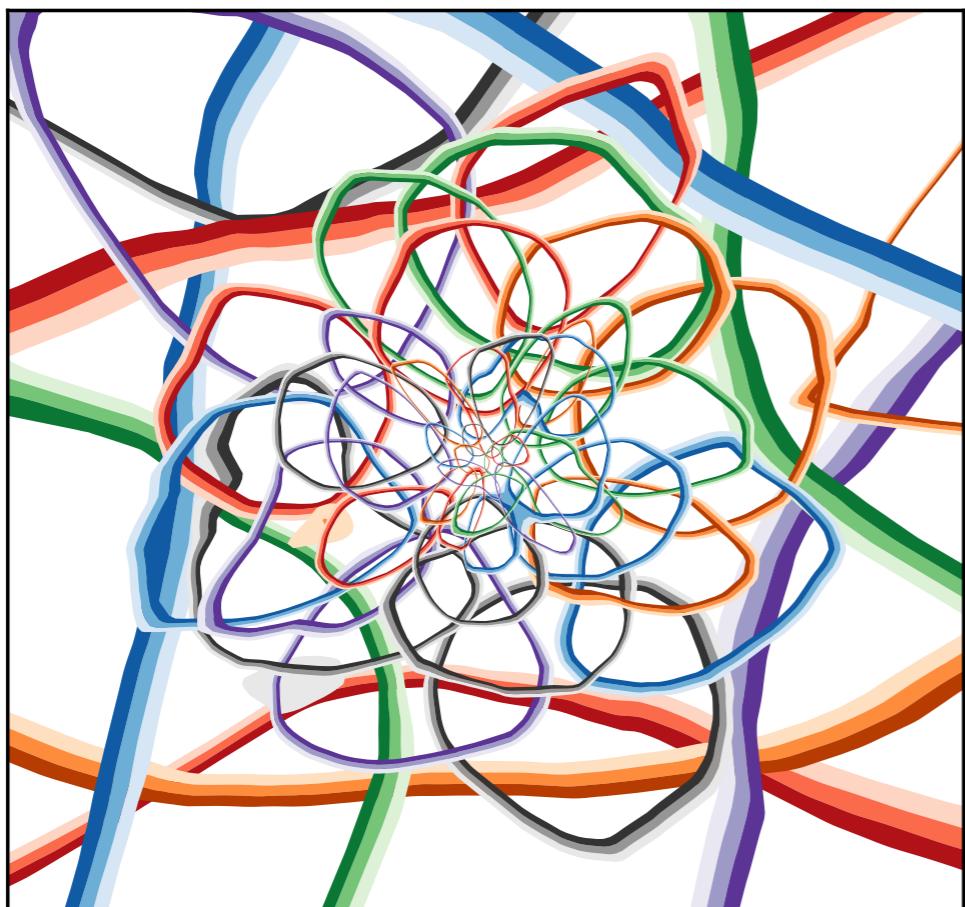


*New insights into particle physics**
*facilitated by advances in machine learning**
(and vice versa?)

Today's Talk: Two Anecdotes

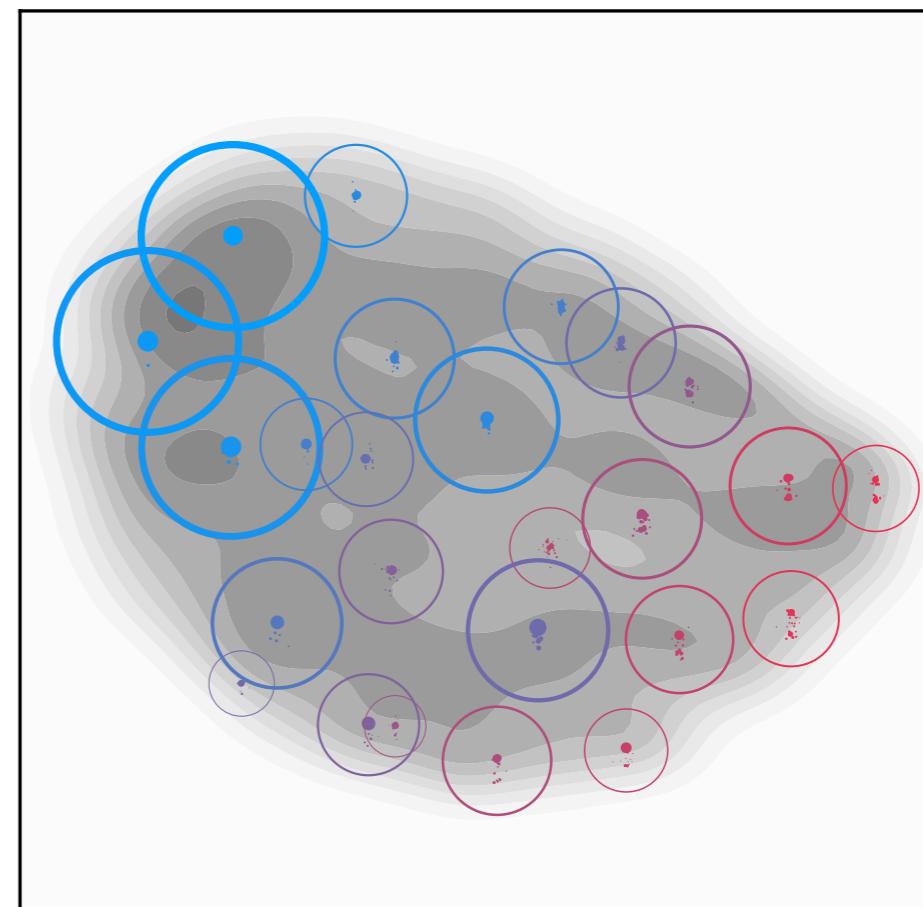


Teaching a Machine to
“Think Like a Physicist”



[Komiske, Metodiev, JDT, JHEP 2019]

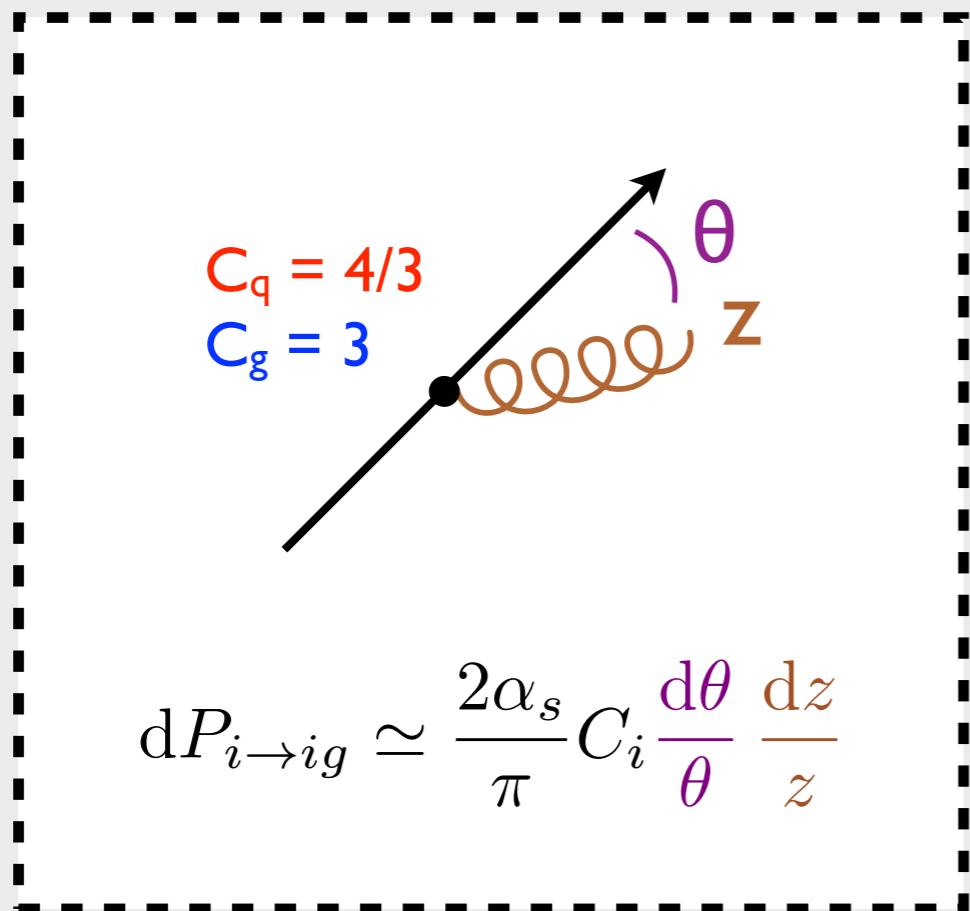
Letting Collider Data
Speak for Itself



[Komiske, Mastandrea, Metodiev, Naik, JDT, arXiv 2019;
based on Komiske, Metodiev, JDT, PRL 2019]

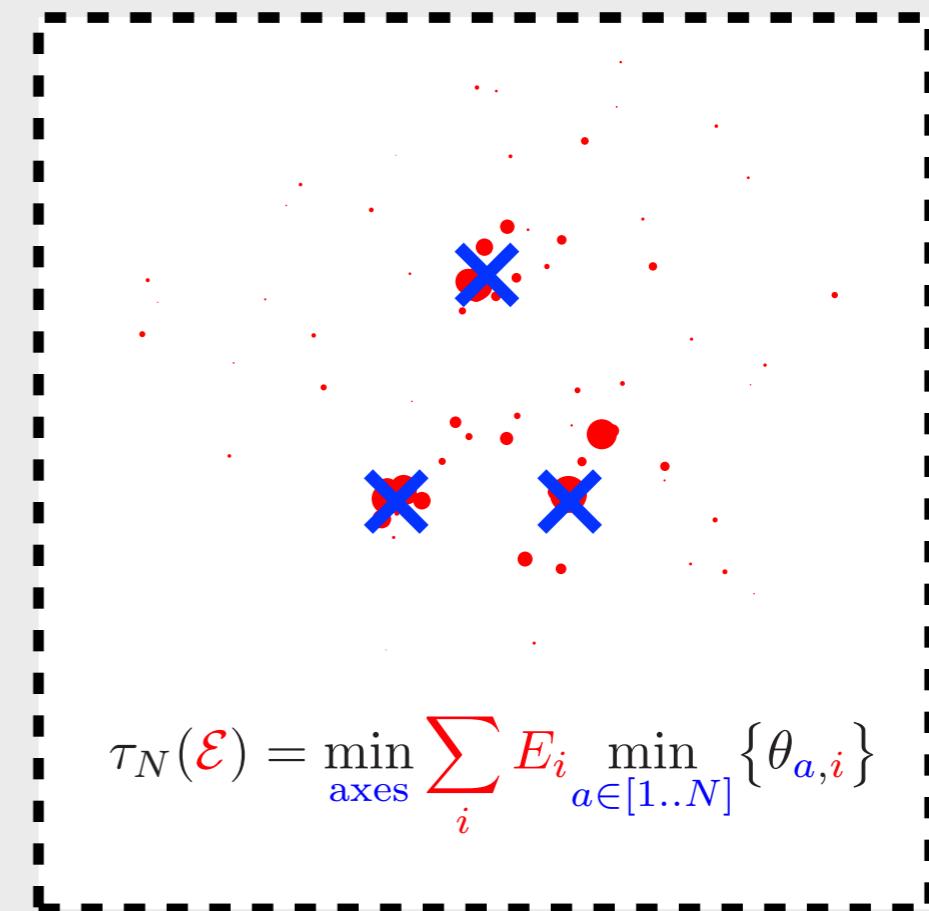
*Data analysis strategies motivated by the
symmetries and structures of particle physics*

Exploiting a Core Prediction of QCD



[Altarelli, Parisi, [NPB 1977](#)]

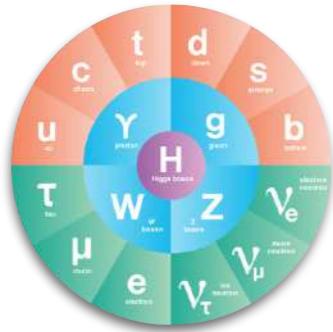
N-prong Singularities of Gauge Theories



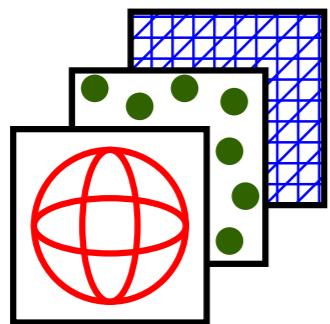
[Stewart, Tackmann, Waalewijn, [PRL 2010](#); JDT, Van Tilburg, [JHEP 2011, JHEP 2012](#)]

New perspectives on key theoretical concepts

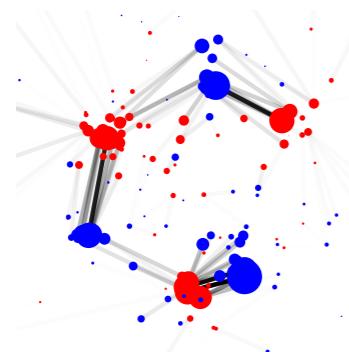
Outline



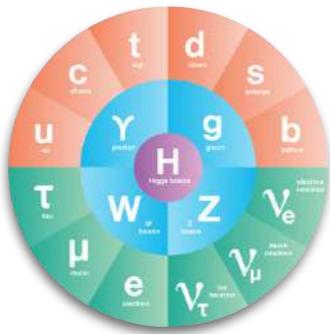
Particle Physics 101



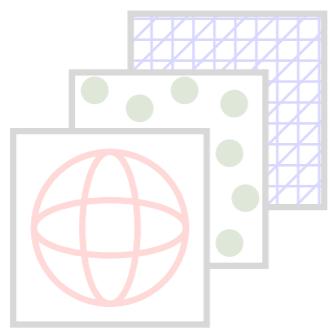
What is a Collider Event?



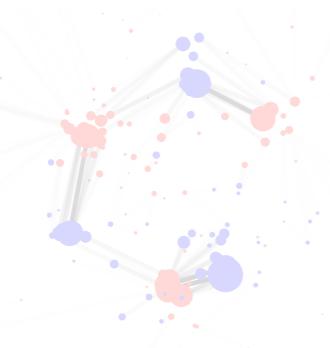
When are Collider Events Similar?



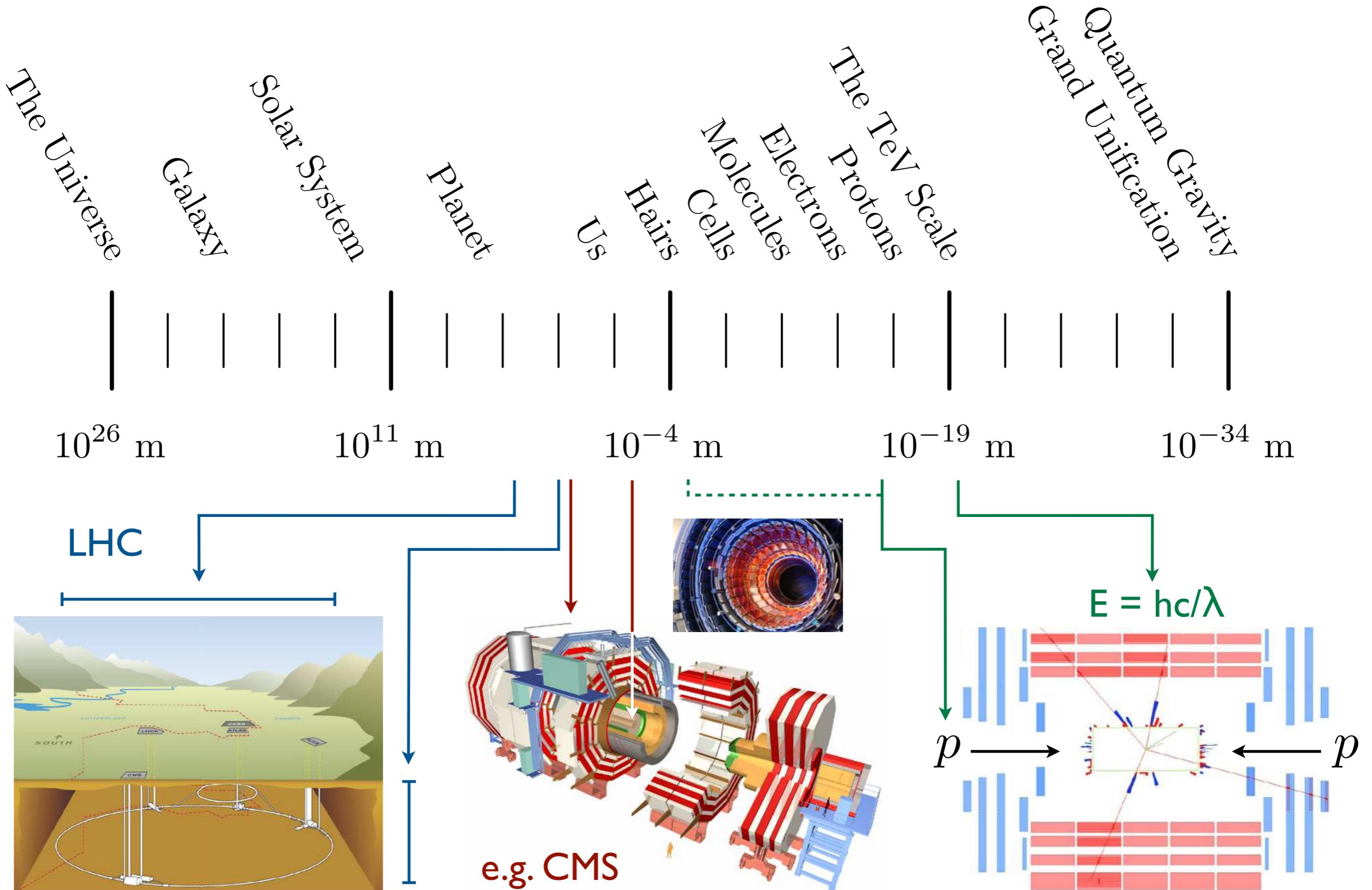
Particle Physics 101



What is a Collider Event?



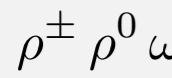
When are Collider Events Similar?





Mesons

$\pi^\pm \pi^0 \eta K^\pm K^0 \eta' D^\pm D^0 D_s^\pm \eta_c B^\pm B^0 B_s^0 \eta_b \dots$



Baryons

$p n \Lambda^0 \Sigma^+ \Sigma^0 \Sigma^- \Xi^0 \Xi^- \dots$

$\Delta^{++} \Delta^+ \Delta^0 \Delta^- \Sigma^{*+} \Sigma^{*0} \Sigma^{*-} \Xi^{*0} \Xi^{*-} \Omega^- \dots$



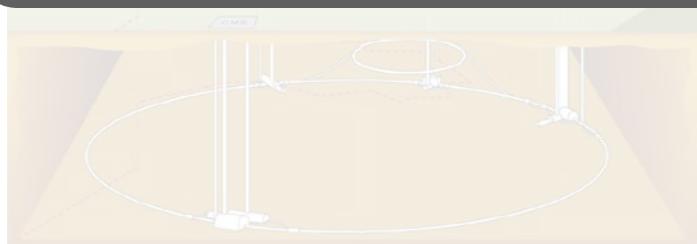
Tetraquarks

$X(3872) Y(4260) Z(4430) \dots$

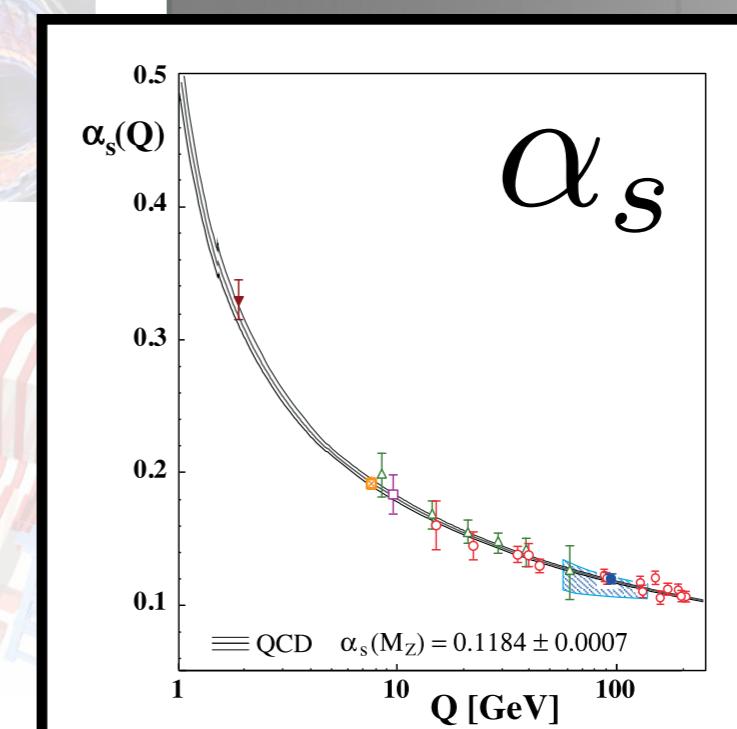
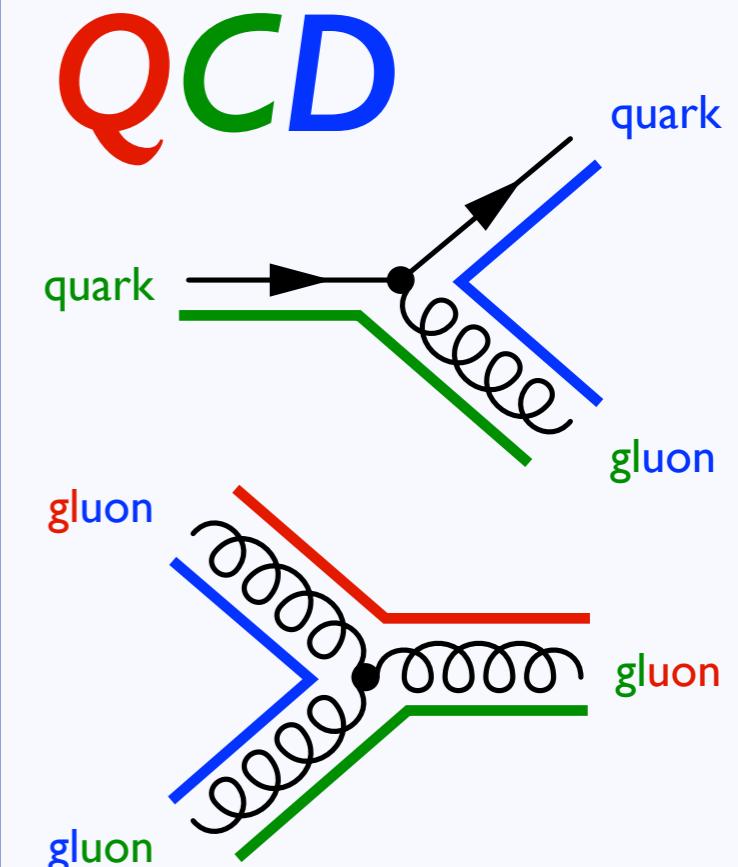
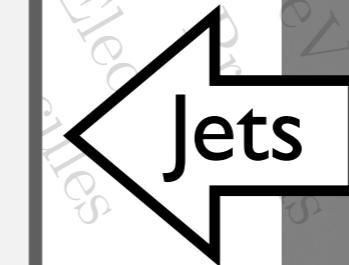


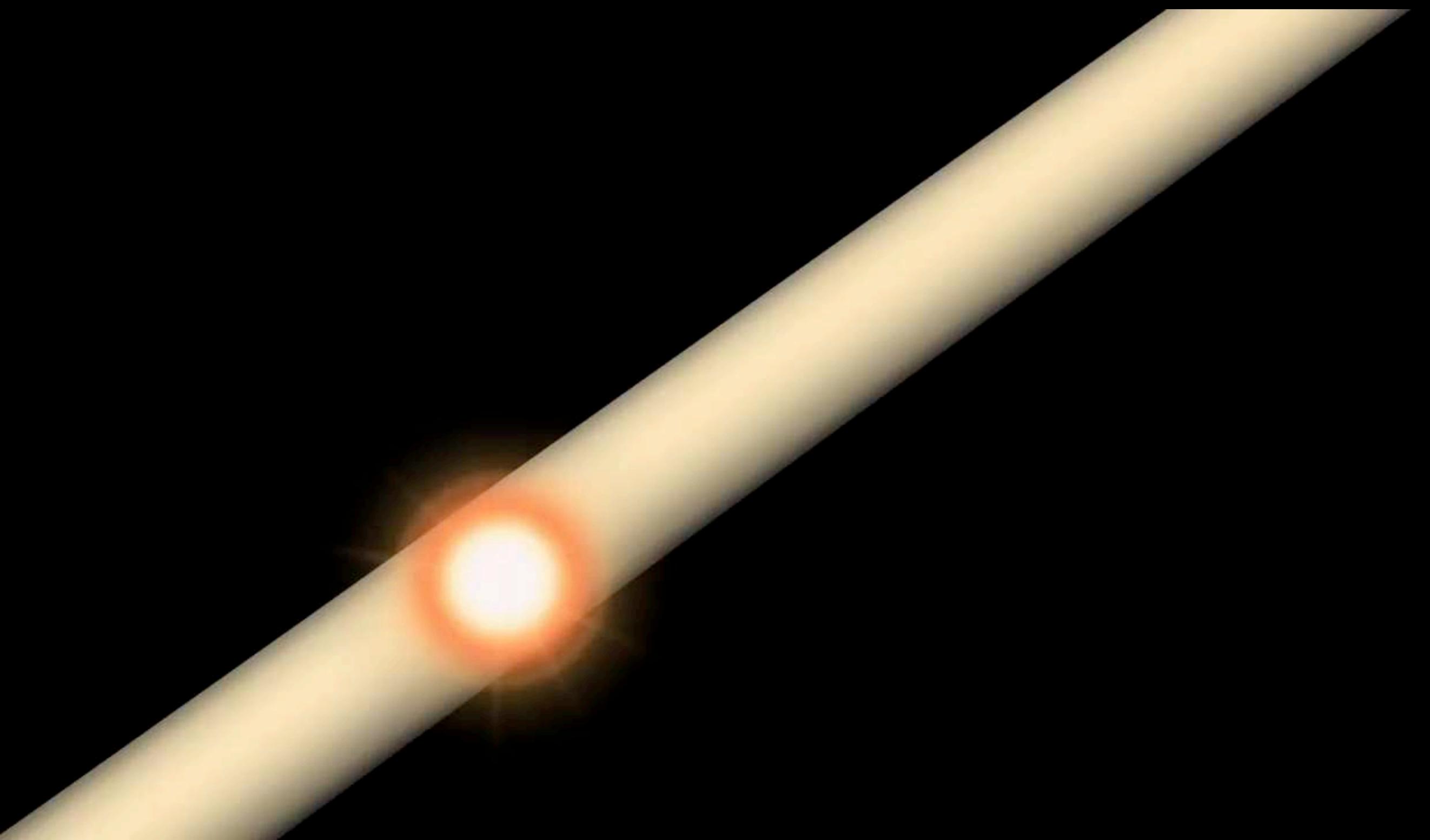
Pentaquarks

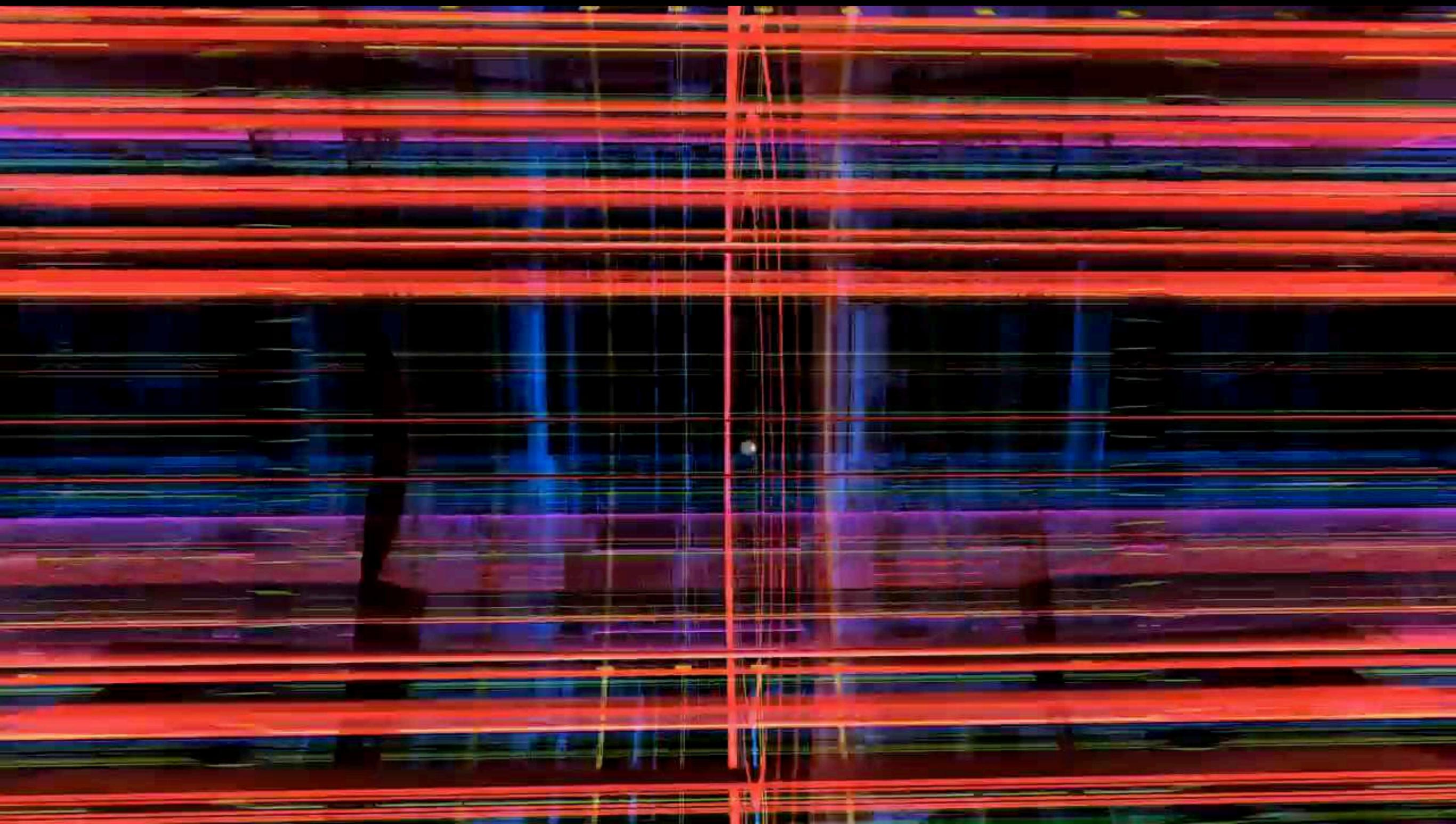
$P_c^+(4450) \dots$

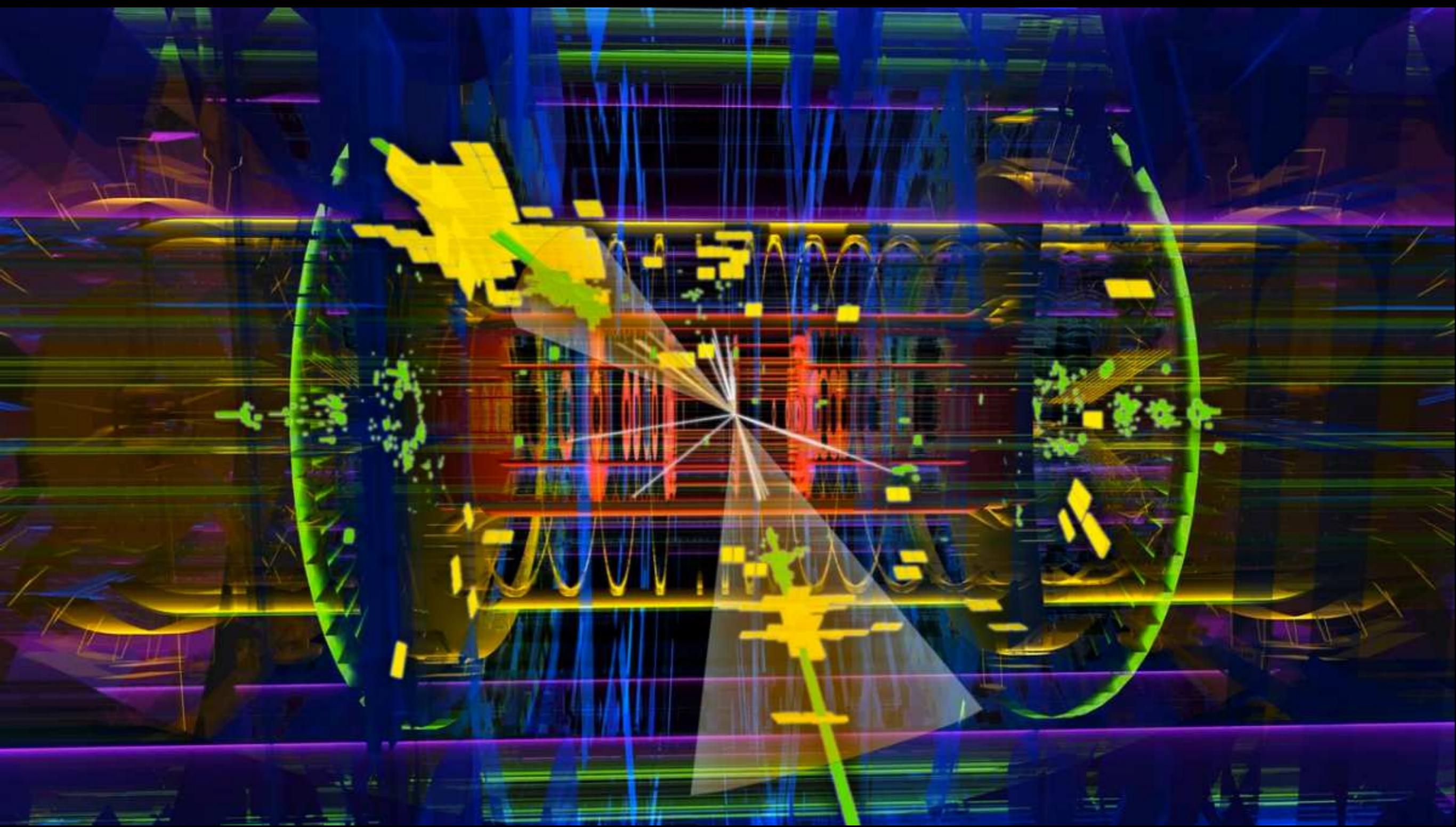


e.g. CMS



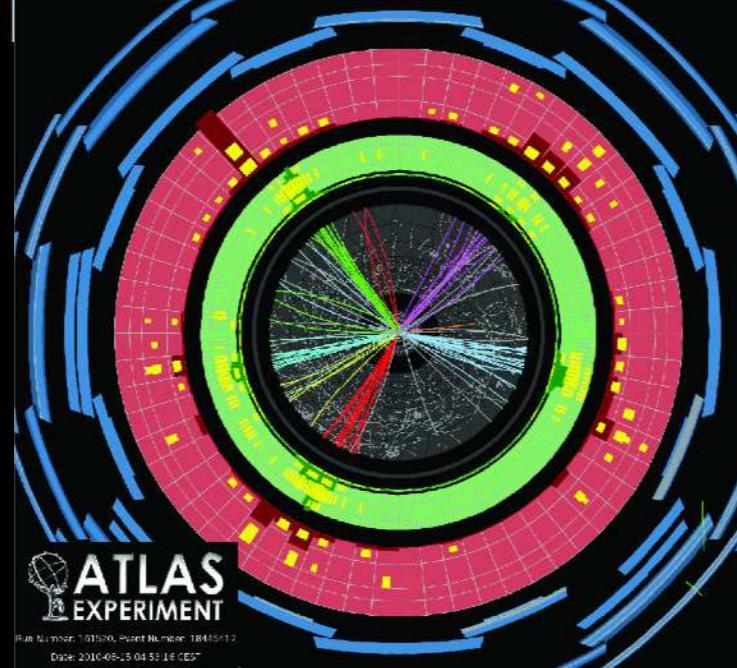
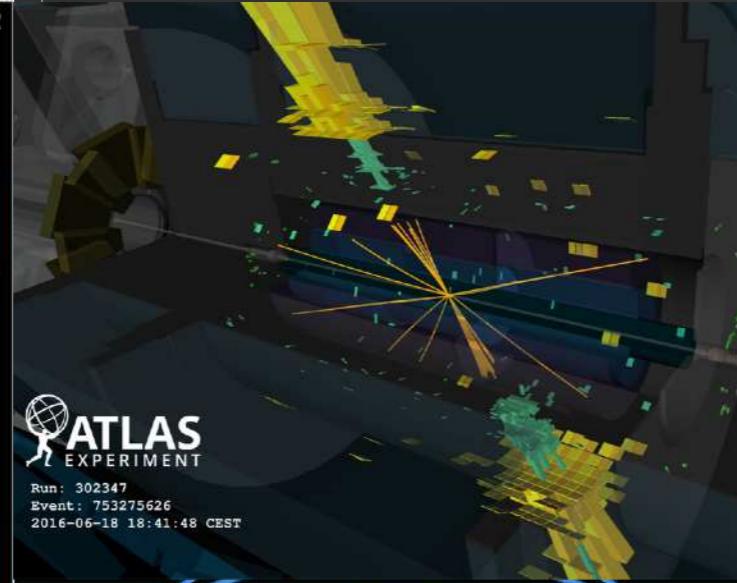
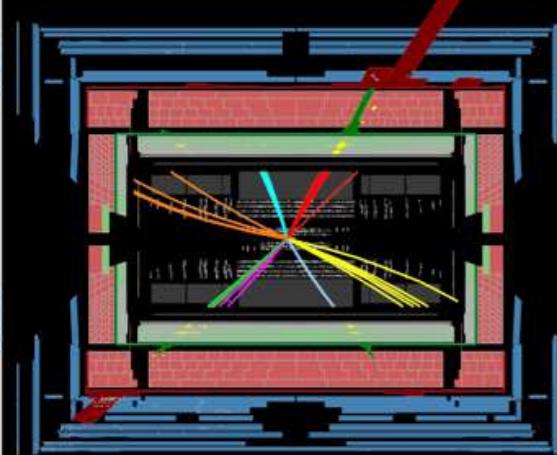




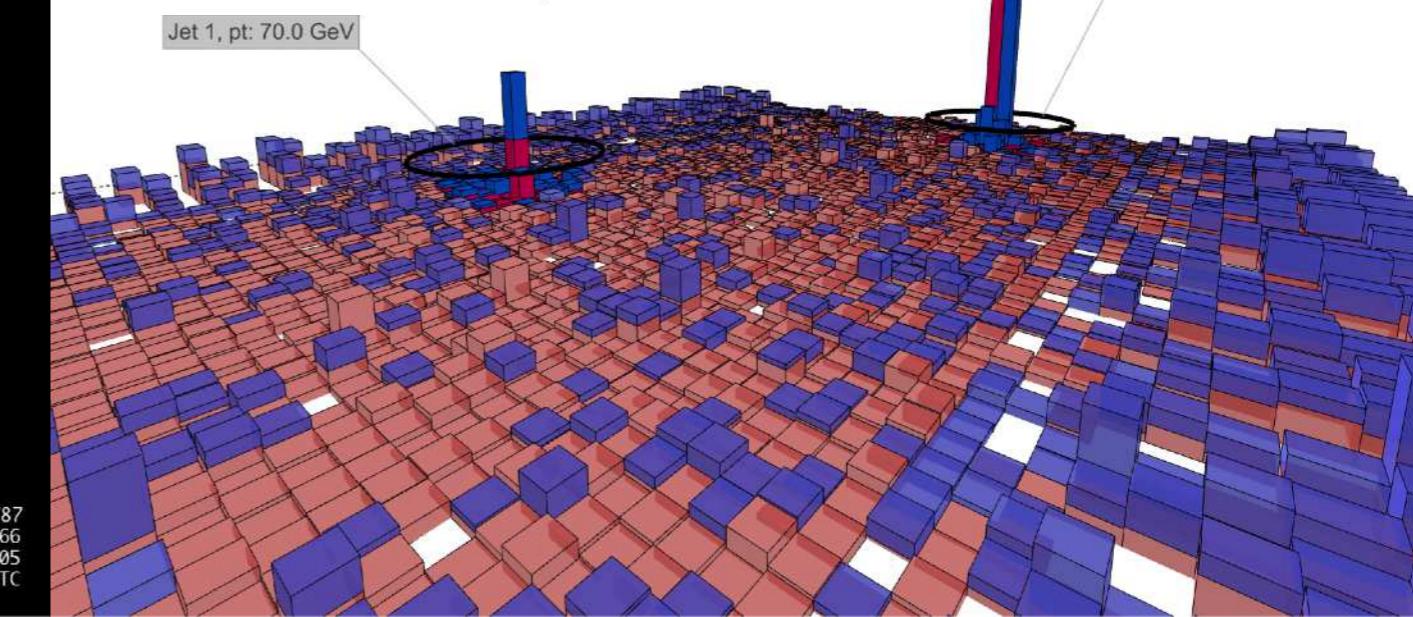
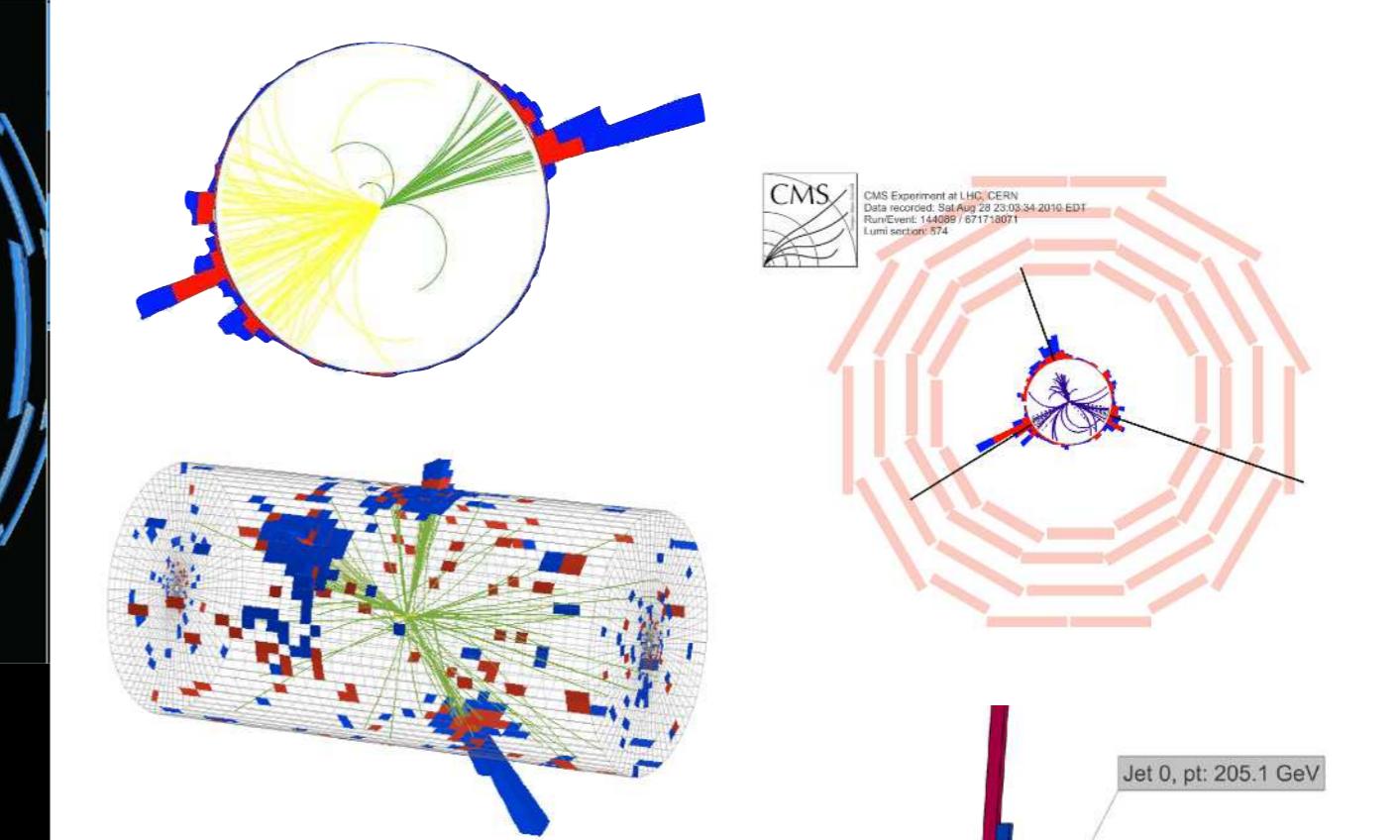
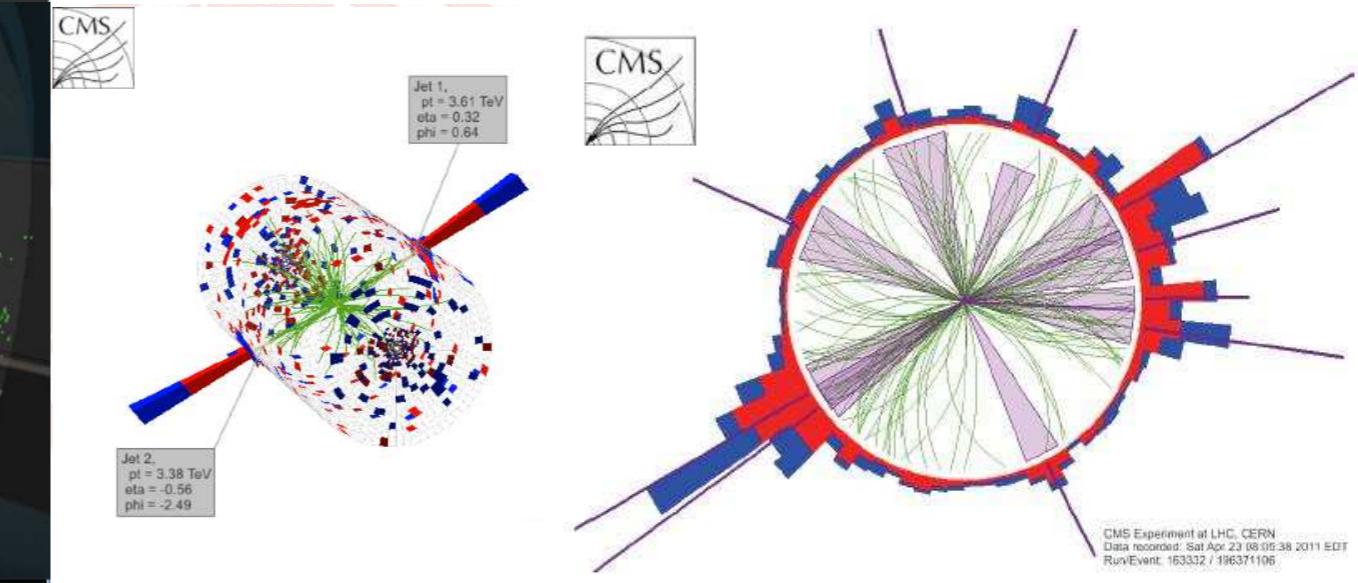
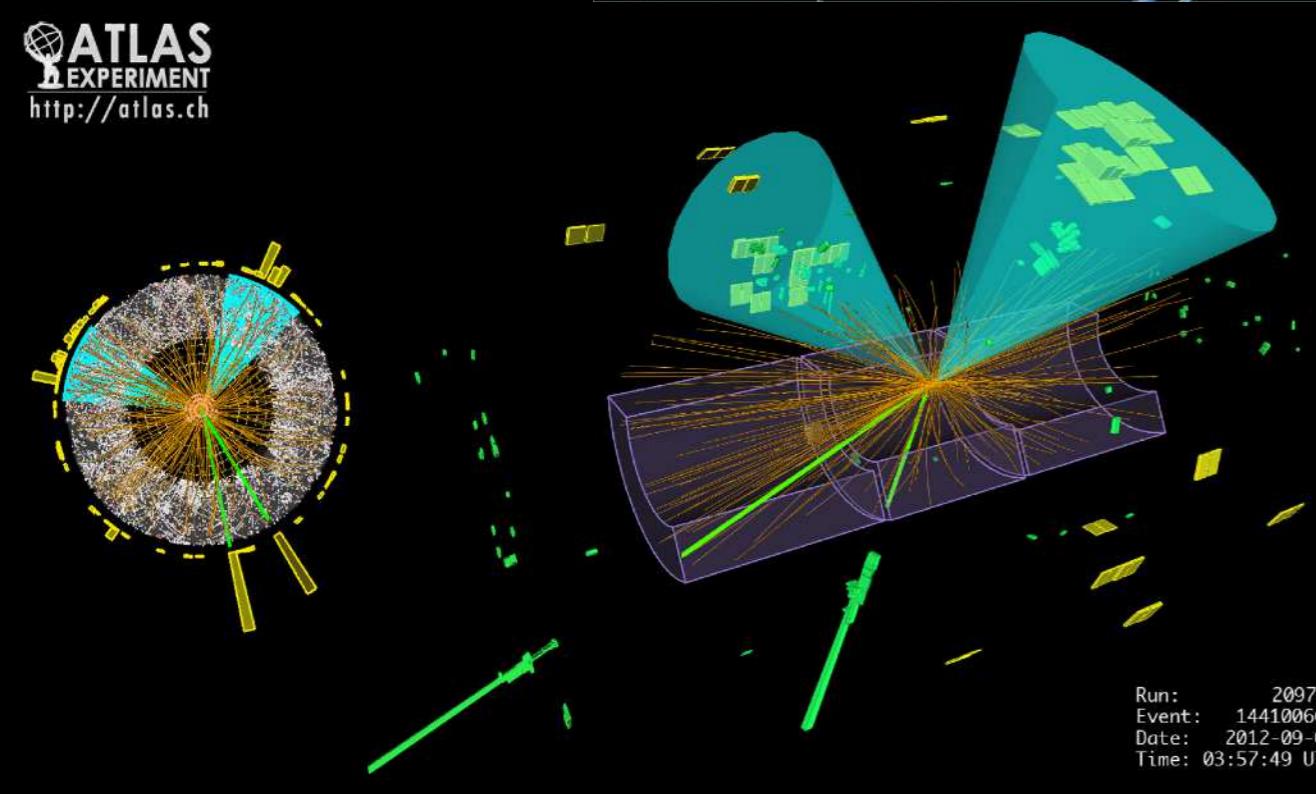


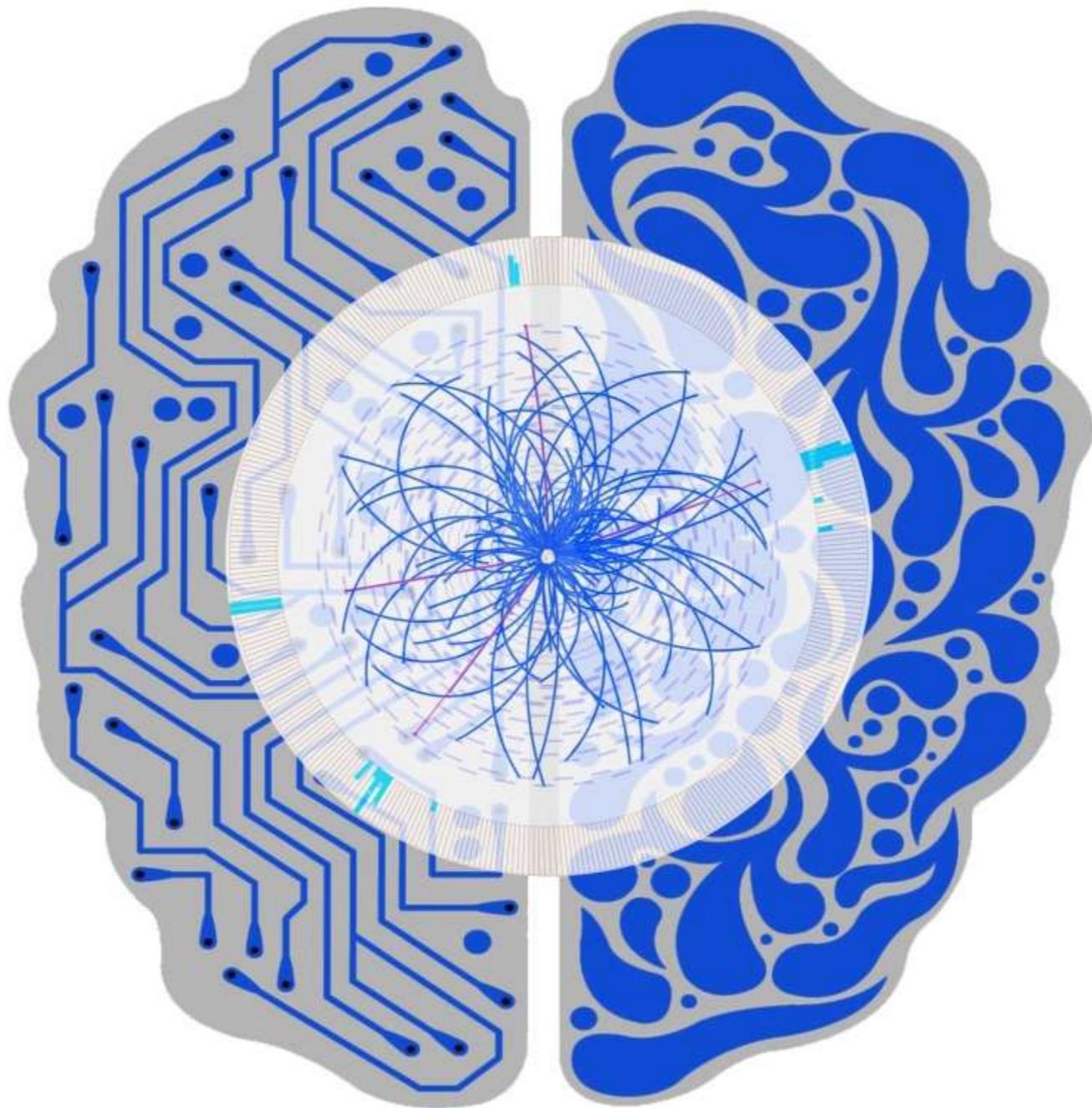
Run Number: 159224, Event Number: 3533152

Date: 2010-07-18 11:05:54 CEST



ATLAS
EXPERIMENT
<http://atlas.ch>





The Rise of Machine Learning

Deep Learning

Inpainting



Corrupted



Deep image prior

increased computational power and large data sets

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

Deep Learning (or Deep Thinking?)

Inpainting



Corrupted



Deep image prior

Using randomly initialized neural network (!)

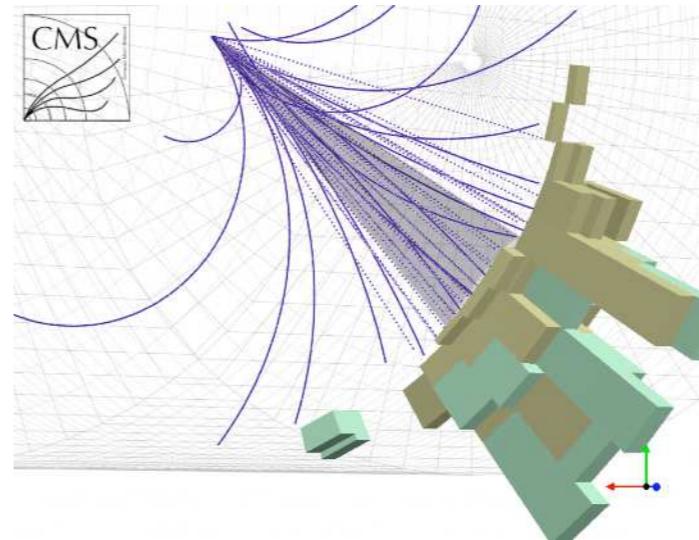
Progress made by understanding the structure of problems
(not just increased computational power and large data sets)

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

Cartoon of Machine Learning

“ML4Jets”
NYU, January 2020

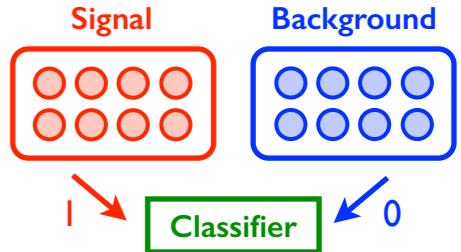
For this talk: \mathcal{J} = jet



E.g.: Problem = Minimize loss function
Solution = Multi-layer neural network
Strategy = Stochastic gradient descent

E.g. Jet Classification

Key supervised learning task at LHC



Find $h\left(\begin{array}{c} \text{jet} \\ \text{fan} \end{array}\right)$ such that

$$h(\text{Quark}) = 1$$
$$h(\text{Gluon}) = 0$$

Best you can do: $h(\mathcal{J}) = \frac{p(\mathcal{J}|Q)}{p(\mathcal{J}|Q) + p(\mathcal{J}|G)}$
(Neyman-Pearson lemma)

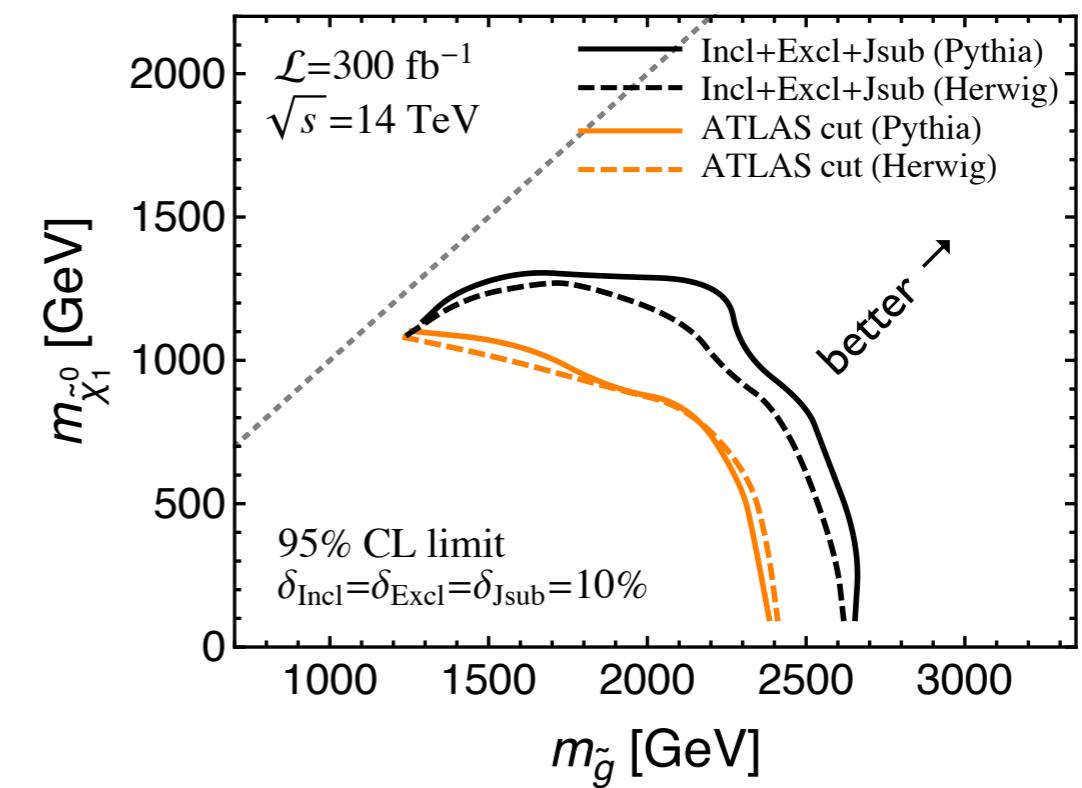
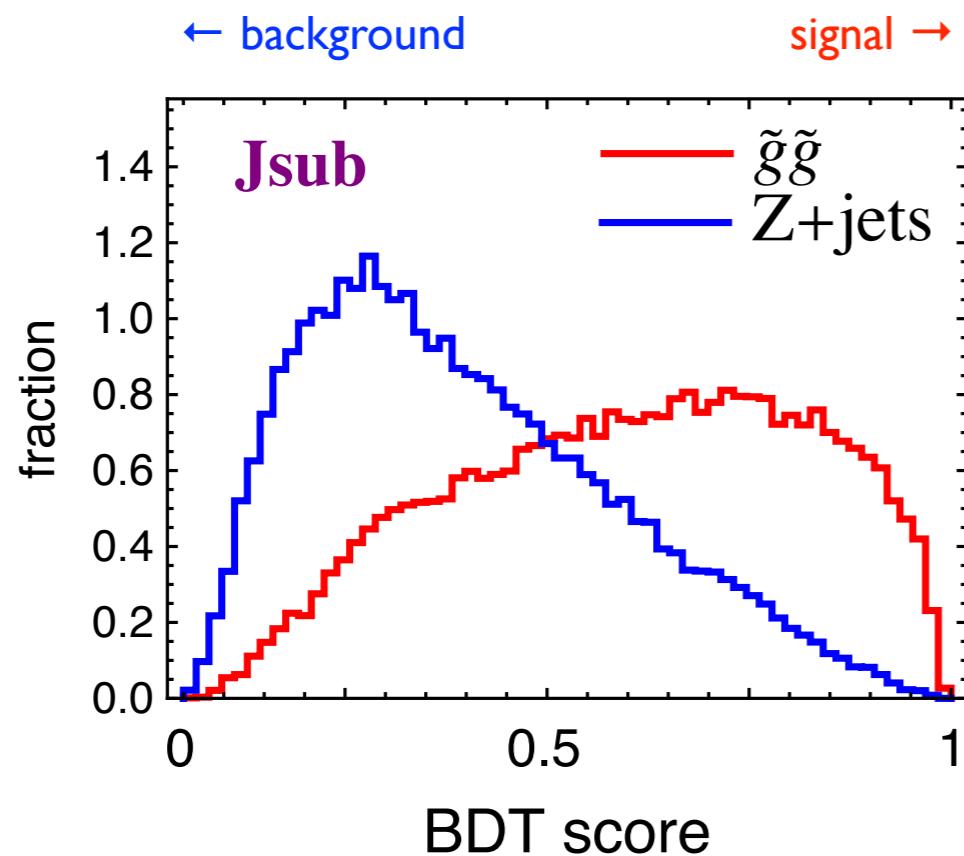
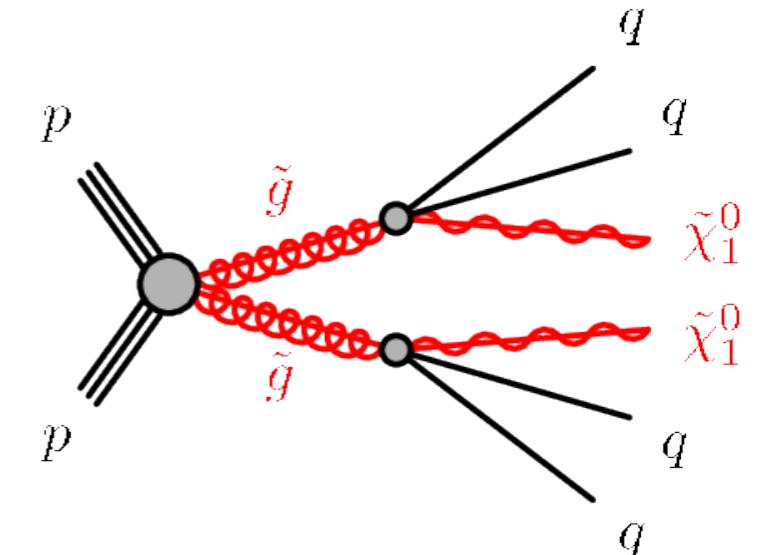
E.g. Search for Supersymmetry

Classifier: Boosted decision tree (for each of 4 jets)

Inputs: Jet mass, width, track multiplicity

Signal: Quark enriched

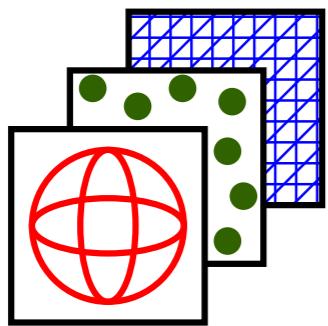
Background: Gluon enriched



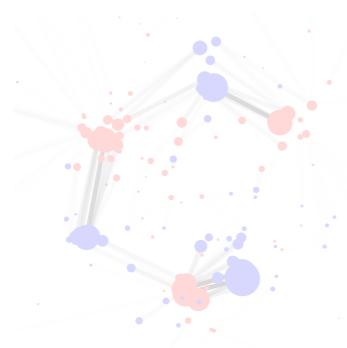
[Bhattacherjee, Mukhopadhyay, Nojiri, Sakakie, Webber, [JHEP 2017](#)]



Particle Physics 101



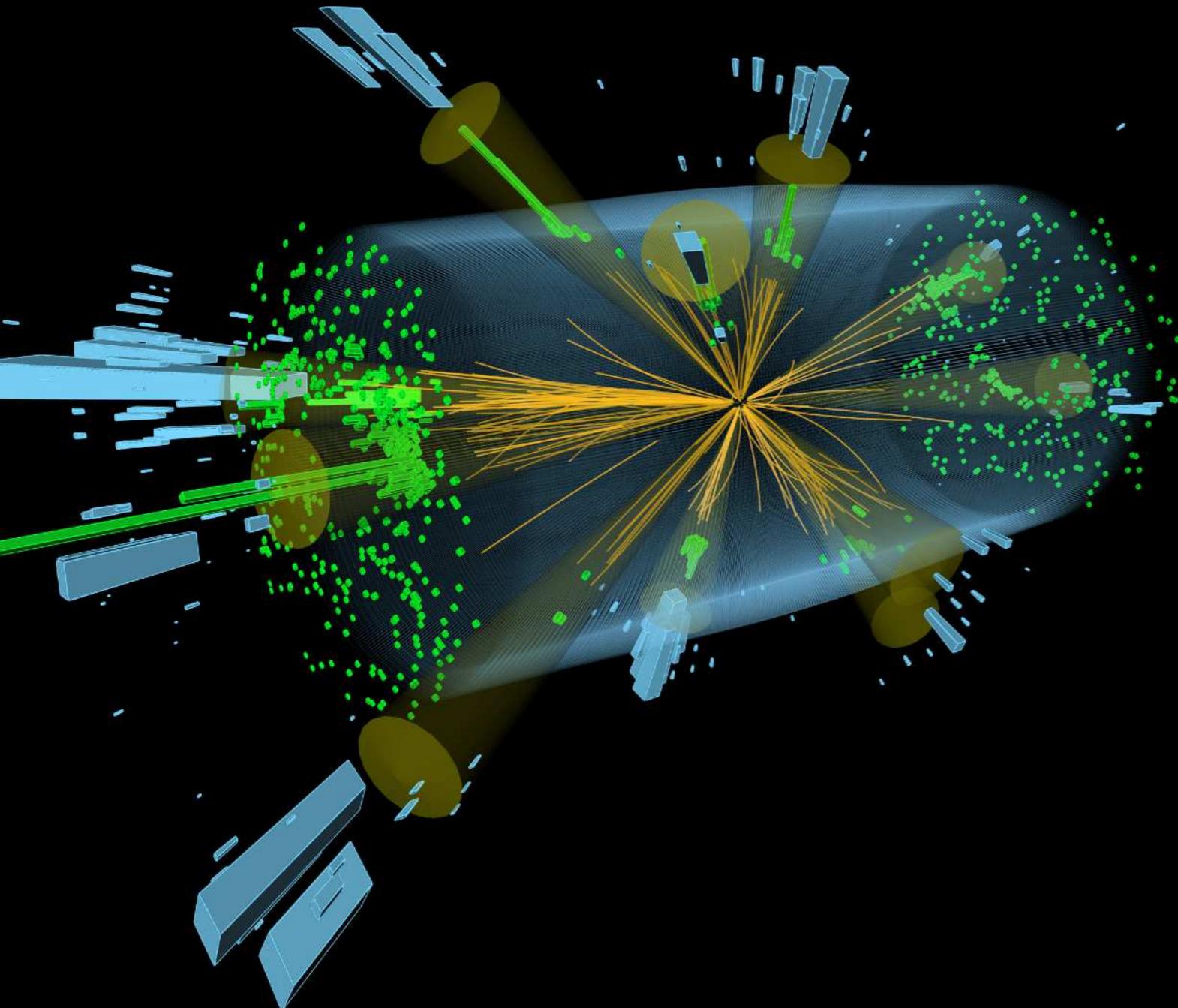
What is a Collider Event?



When are Collider Events Similar?

Collider Event

Collection of points in (momentum) space



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

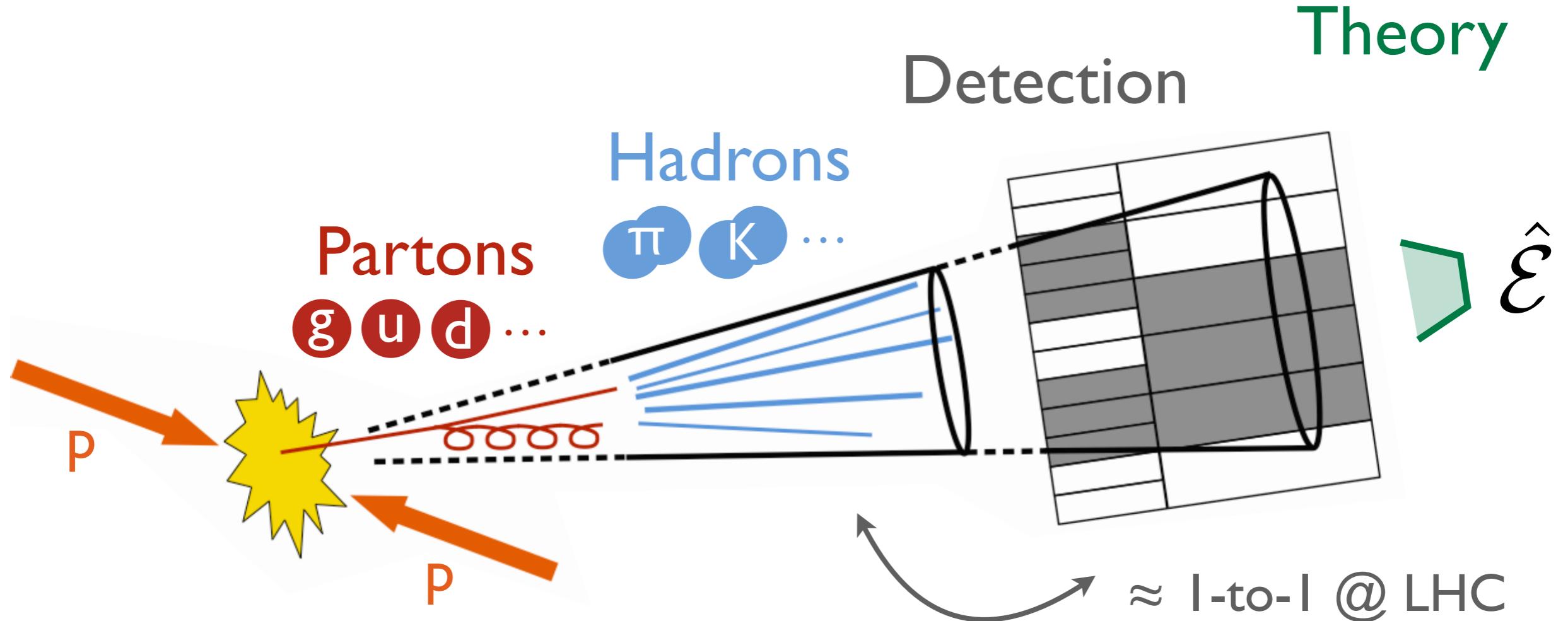
Point Cloud

Collection of points in (position) space



[Popular Science, 2013]

Jet Formation Process



Stress-energy flow:
Robust to non-perturbative and detector effects

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

[Sveshnikov, Tkachov, [PLB 1996](#); Hofman, Maldacena, [JHEP 2008](#); Mateu, Stewart, [JDT, PRD 2013](#)]

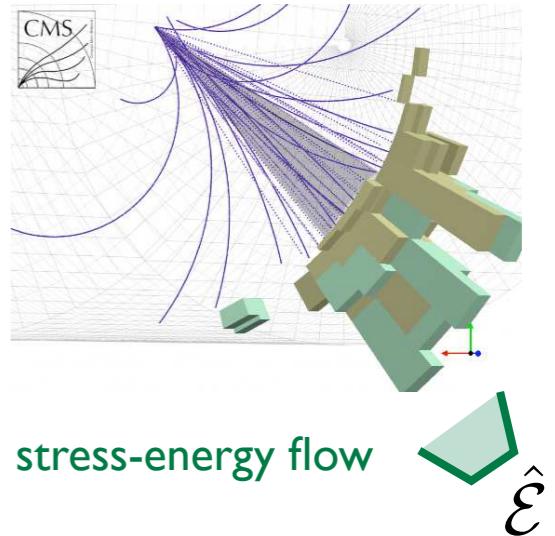
Jets as Weighted Point Clouds

- Energy-weighted directions

$$\vec{p} = \{E, \hat{n}_x, \hat{n}_y, \hat{n}_z\}$$

↑ |
Energy Direction

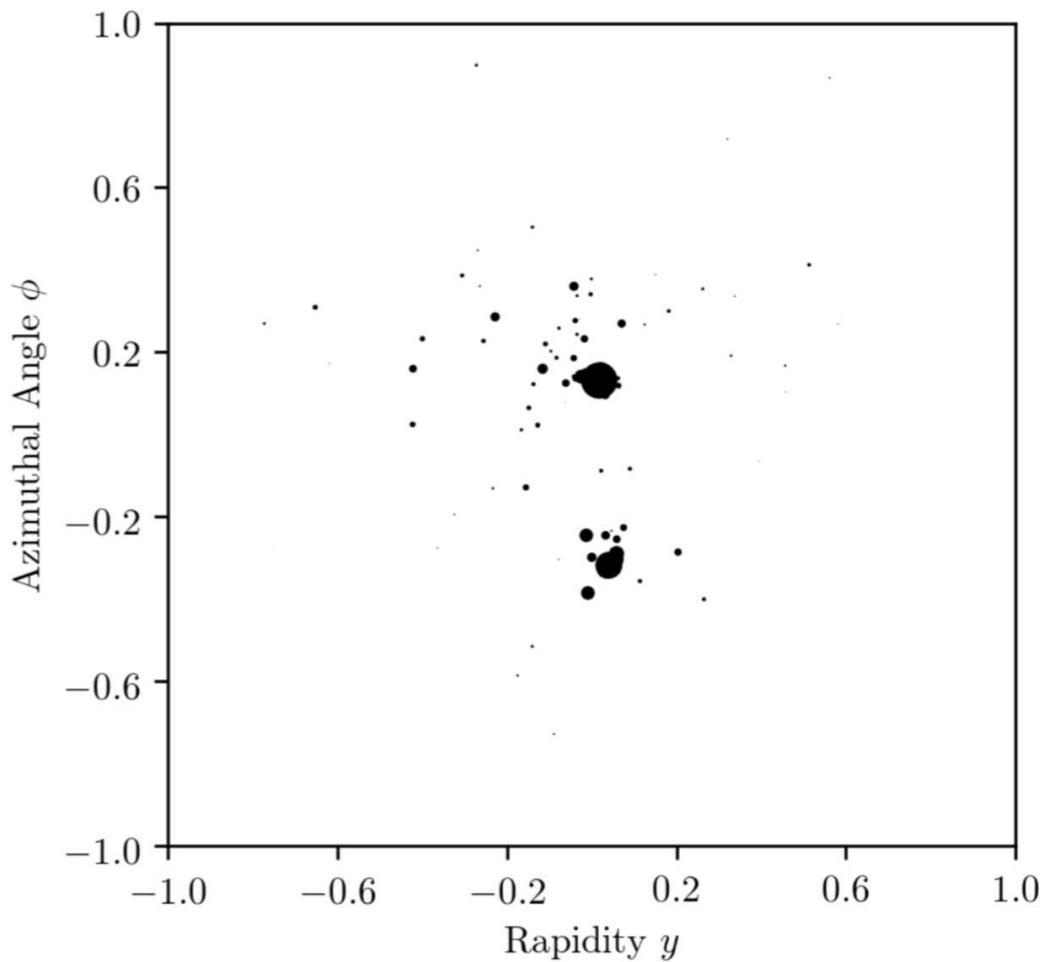
(suppressing “unsafe”
charge/flavor information)



- Visualize as Energy Density

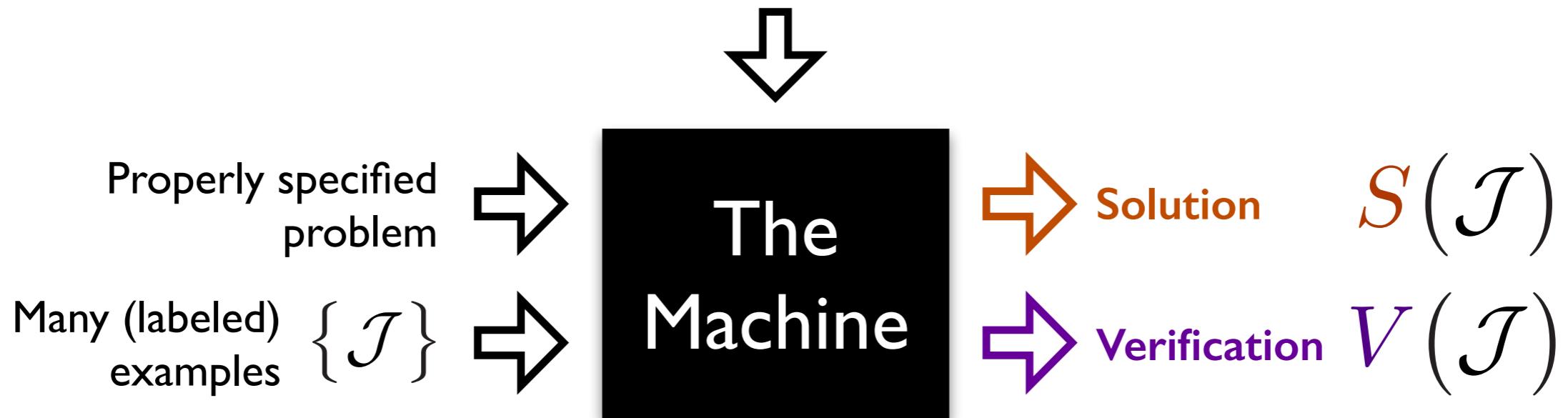
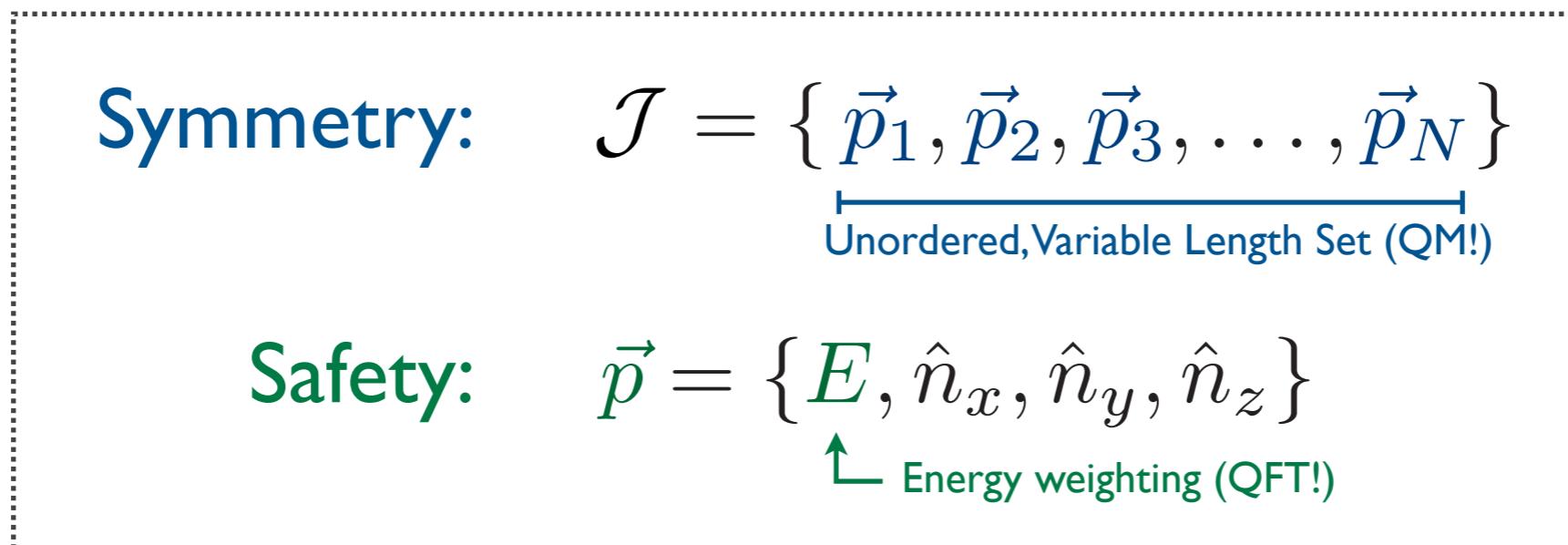
$$\rho(\hat{n}) = \sum_{i \in \mathcal{J}} E_i \delta^{(2)}(\hat{n} - \hat{n}_i)$$

↑ ↑
Energy Direction



- “Safe” to QCD singularities

“Thinking” Like a Physicist



*Check that answer
is physically sensible*

Theoretical (High Energy) Physics



Patrick Komiske



Eric Metodiev



Mathematics,
Statistics,
Computer Science

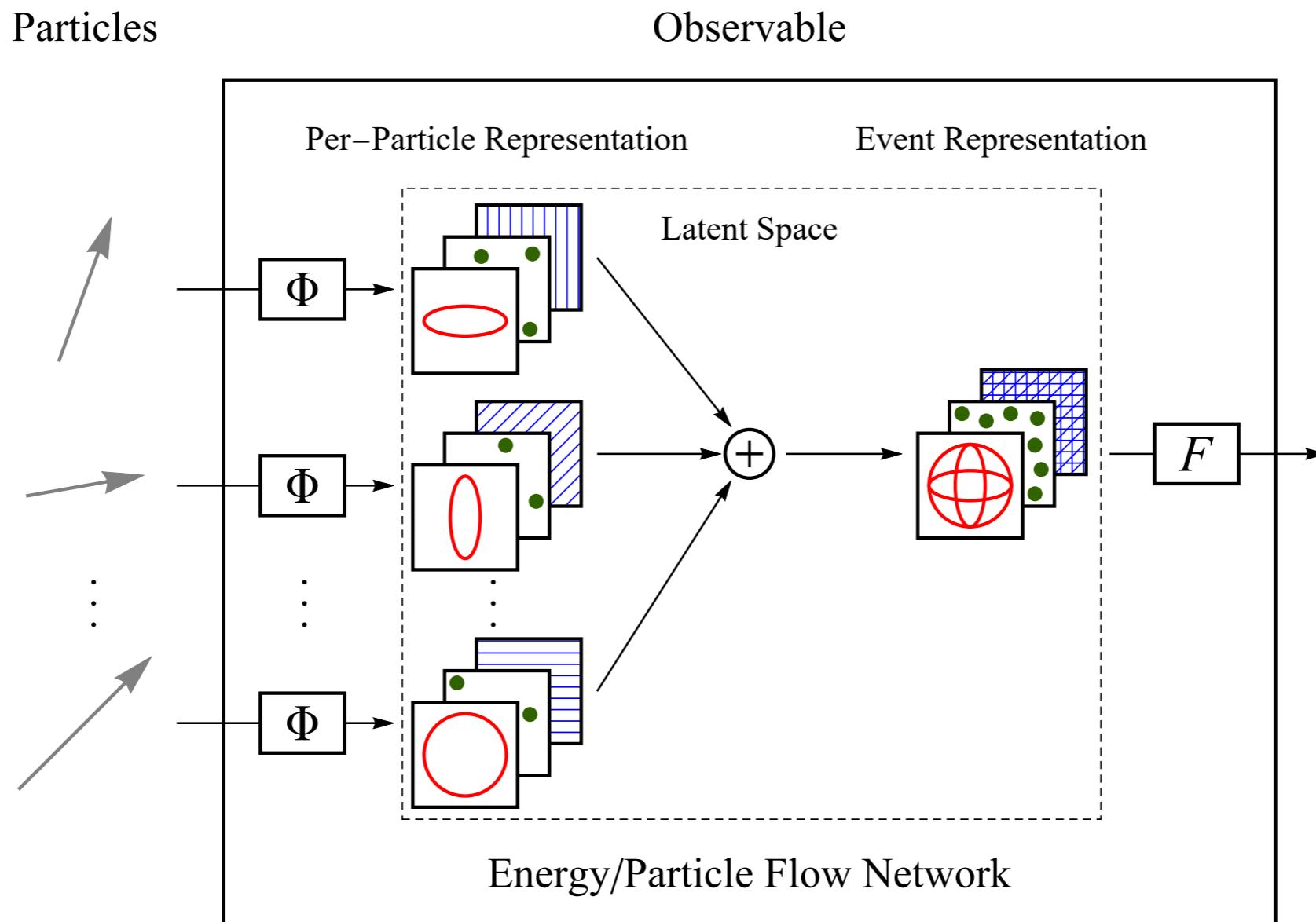
The screenshot shows the 'Welcome to EnergyFlow' page of the EnergyFlow Python package documentation. It features a navigation bar with links like 'Home', 'Welcome', 'Contents', 'Copyright', 'Getting Started', 'Installation', 'Demos', 'Samples', 'FAQs', 'Documentation', 'Energy Flow Polynomials', 'Architecture', 'EMD', 'Measures', 'Generators', 'Utils', and 'Datasets'. The main content area includes a diagram illustrating the flow of energy in particle collisions and a detailed description of the package's features, such as Energy Flow Polynomials (EFPs) and Particle Flow Networks (PFNs).

Energy Flow Networks

<https://energyflow.network/>

Energy Flow Networks

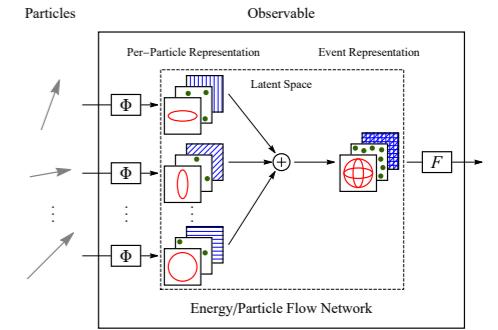
Architecture designed around symmetries and interpretability



[Komiske, Metodiev, JDT, *JHEP* 2019; see also Komiske, Metodiev, JDT, *JHEP* 2018; special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, *NIPS* 2017]

Energy Flow Networks

Architecture designed around symmetries and *interpretability*



$$S(\mathcal{J}) = F(V_1, V_2, \dots, V_\ell)$$

Latent space of dim ℓ

Parametrized with Neural Networks

(see backup for details)

$V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \Phi_a(\hat{n}_i)$

Permutation invariant

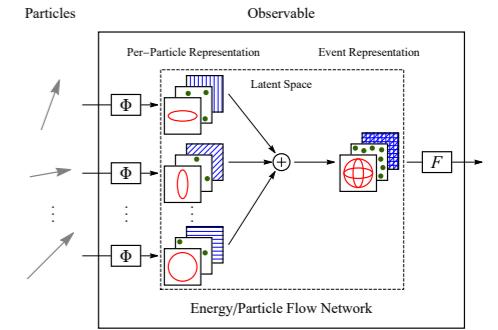
Linear weights
(i.e. safe)

Provably describes any safe observable (!)*
Excellent jet classification performance
Intuitive visualization strategy

[Komiske, Metodiev, JDT, [JHEP 2019](#); see also Komiske, Metodiev, JDT, [JHEP 2018](#);
 special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, [NIPS 2017](#)]

Energy Flow Networks

Architecture designed around symmetries and *interpretability*

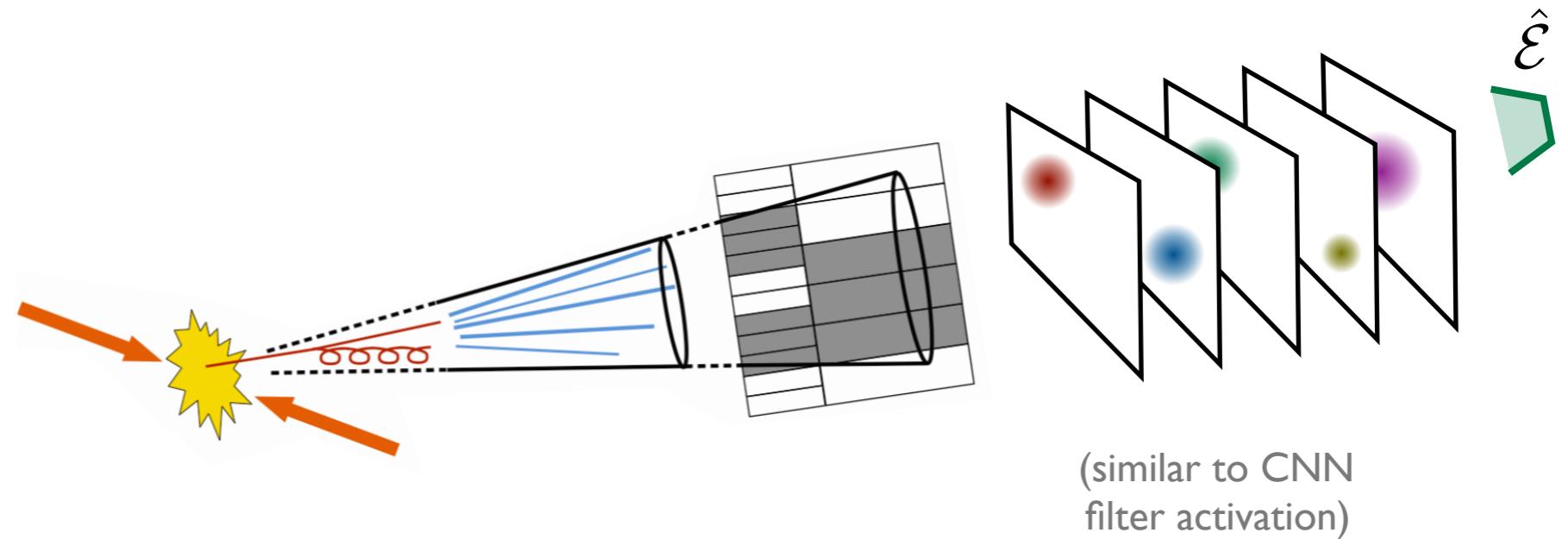


$$S(\mathcal{J}) = F(V_1, V_2, \dots, V_\ell)$$

Latent space of dim ℓ

$V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \Phi_a(\hat{n}_i)$

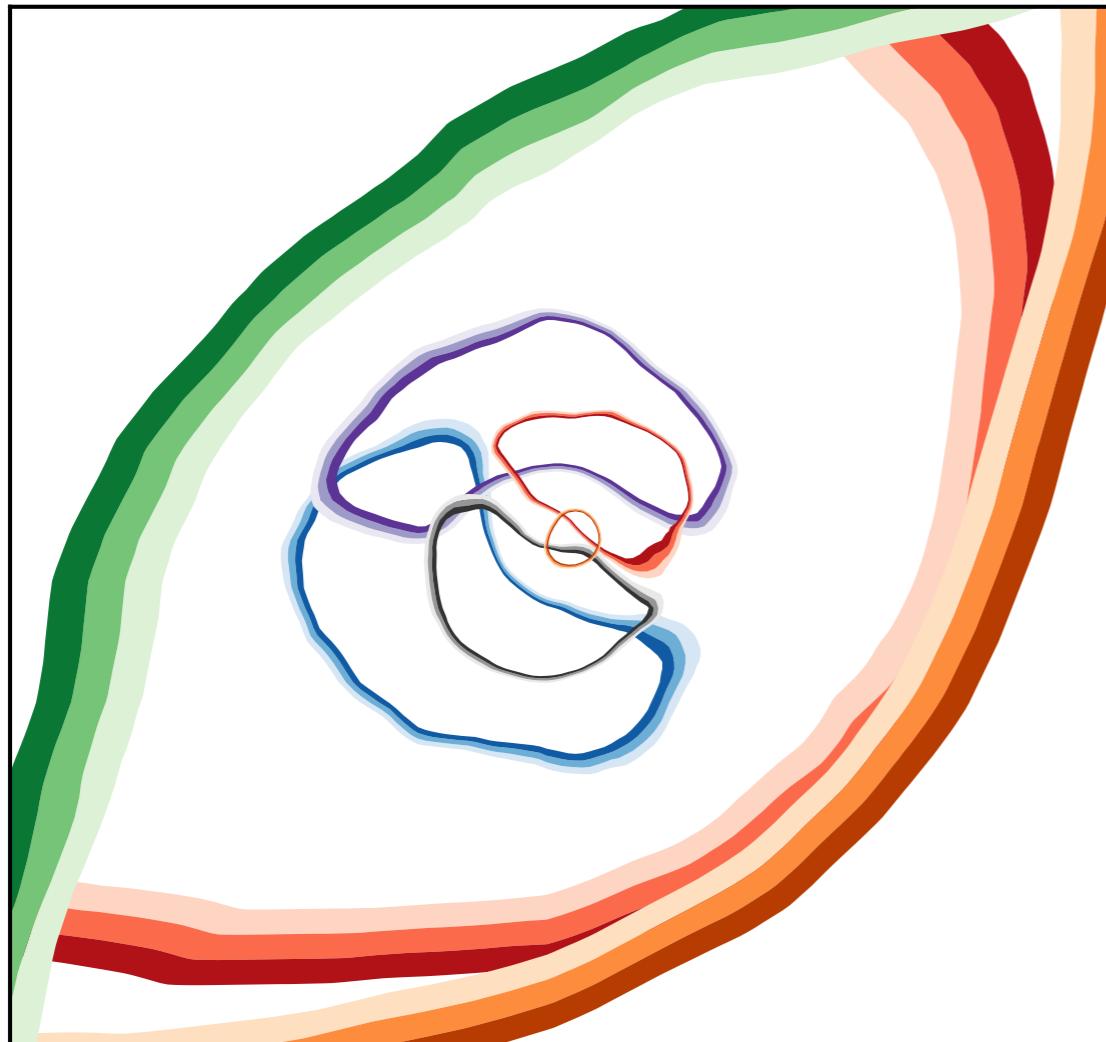
Can visualize if ℓ is small Easy to plot!



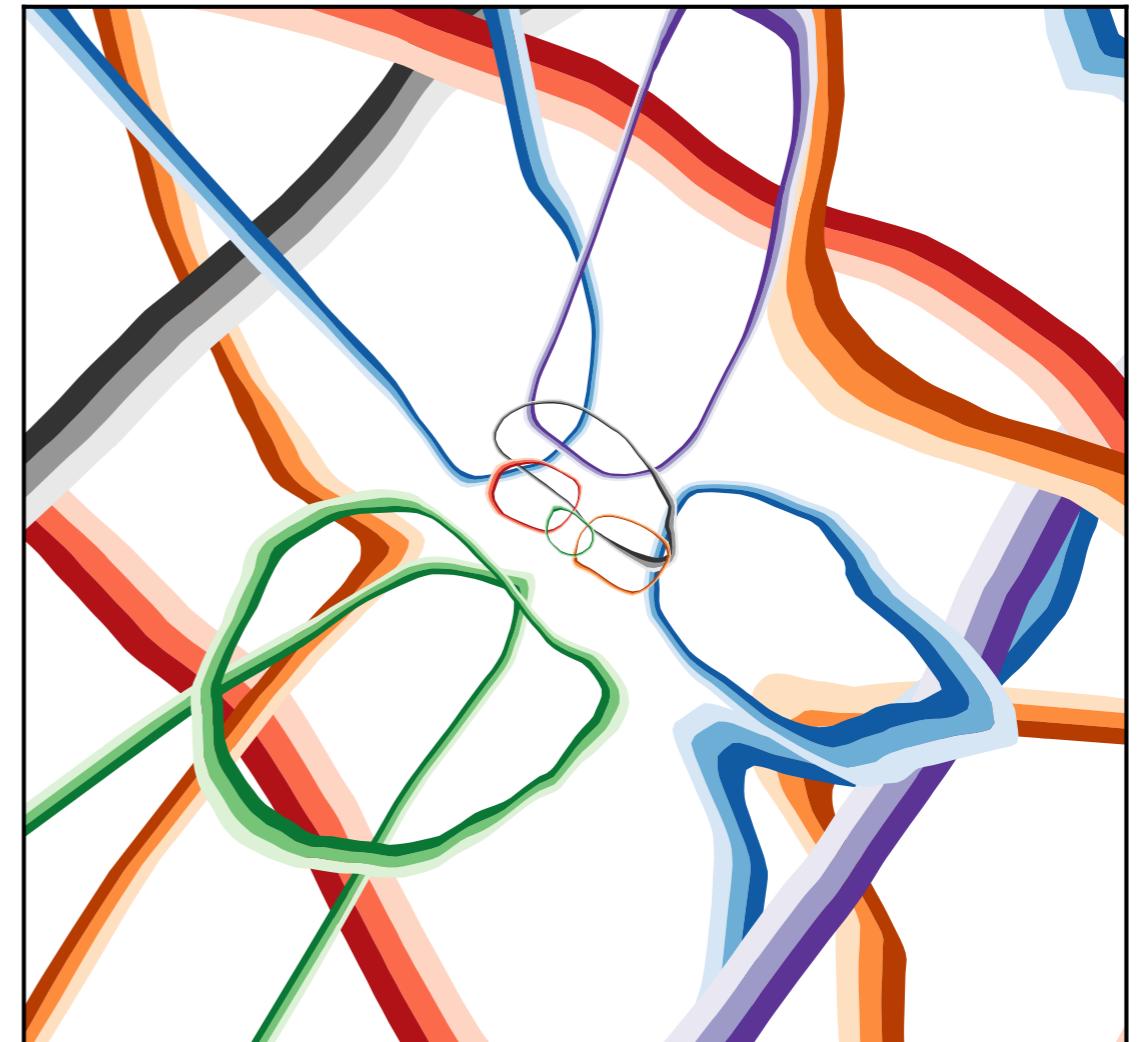
[Komiske, Metodiev, JDT, JHEP 2019; see also Komiske, Metodiev, JDT, JHEP 2018;
special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NIPS 2017]

Psychedelic Network Visualization

Latent Dimension 8



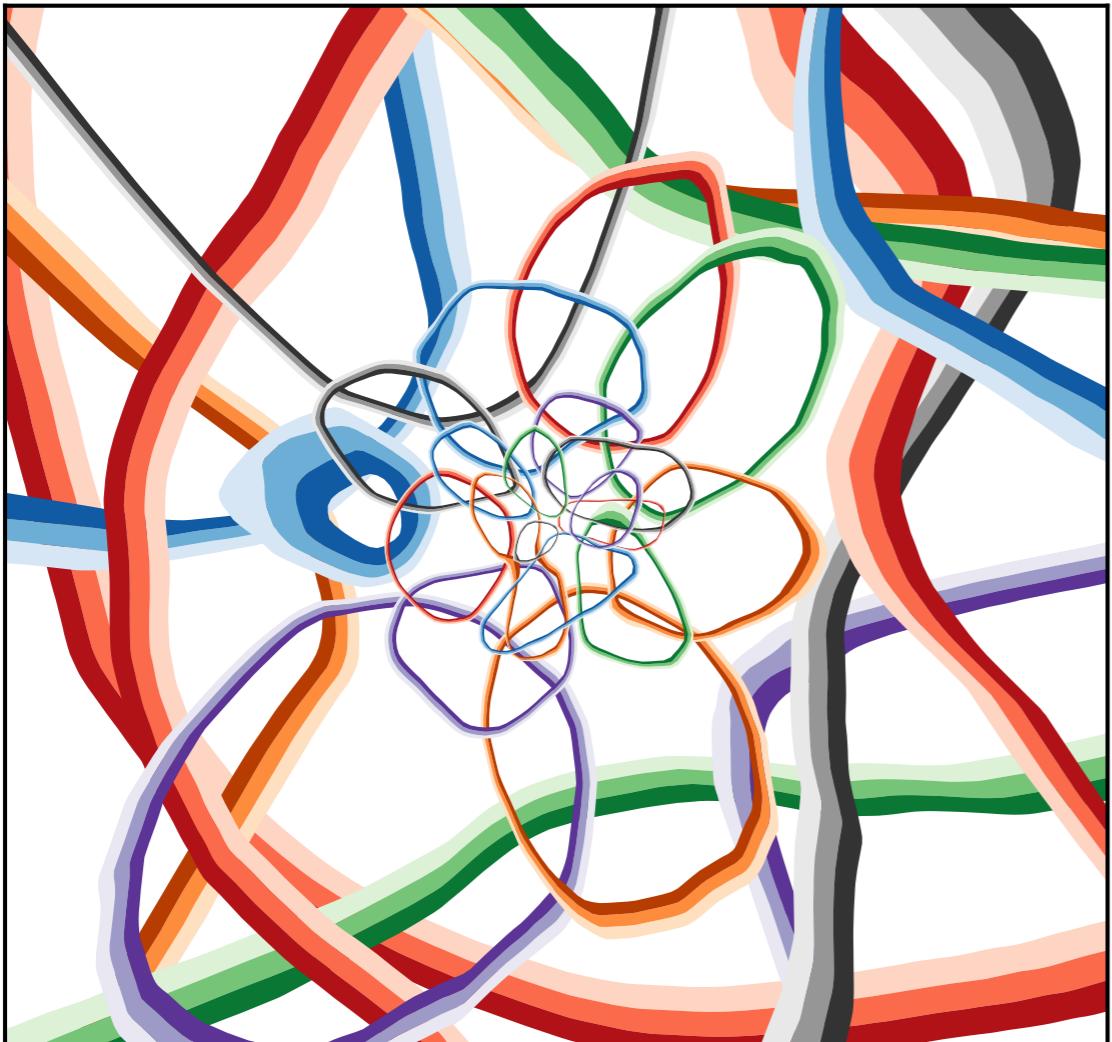
Latent Dimension 16



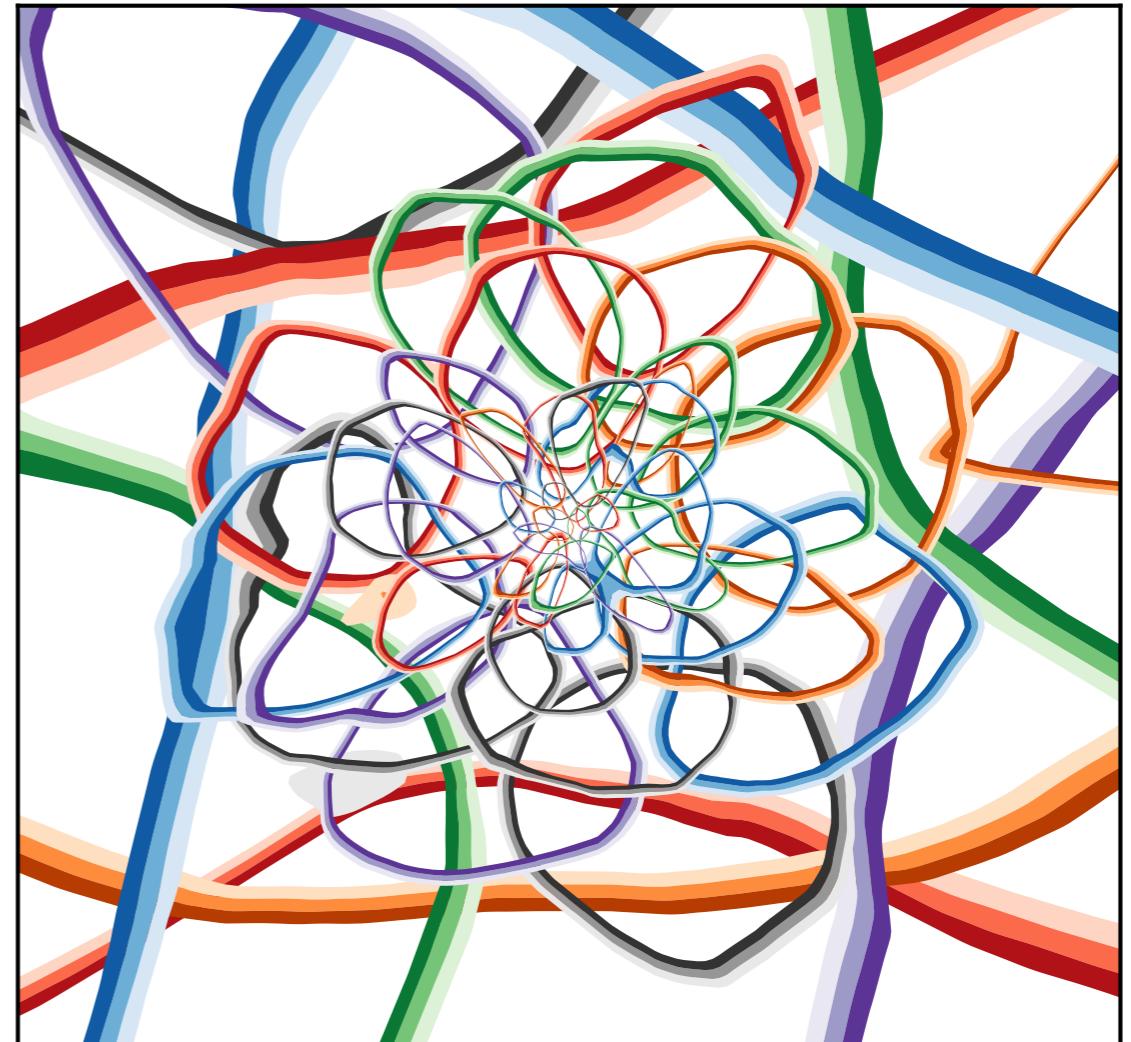
For the case of **quark** vs. **gluon** classification

Psychedelic Network Visualization

Latent Dimension 32

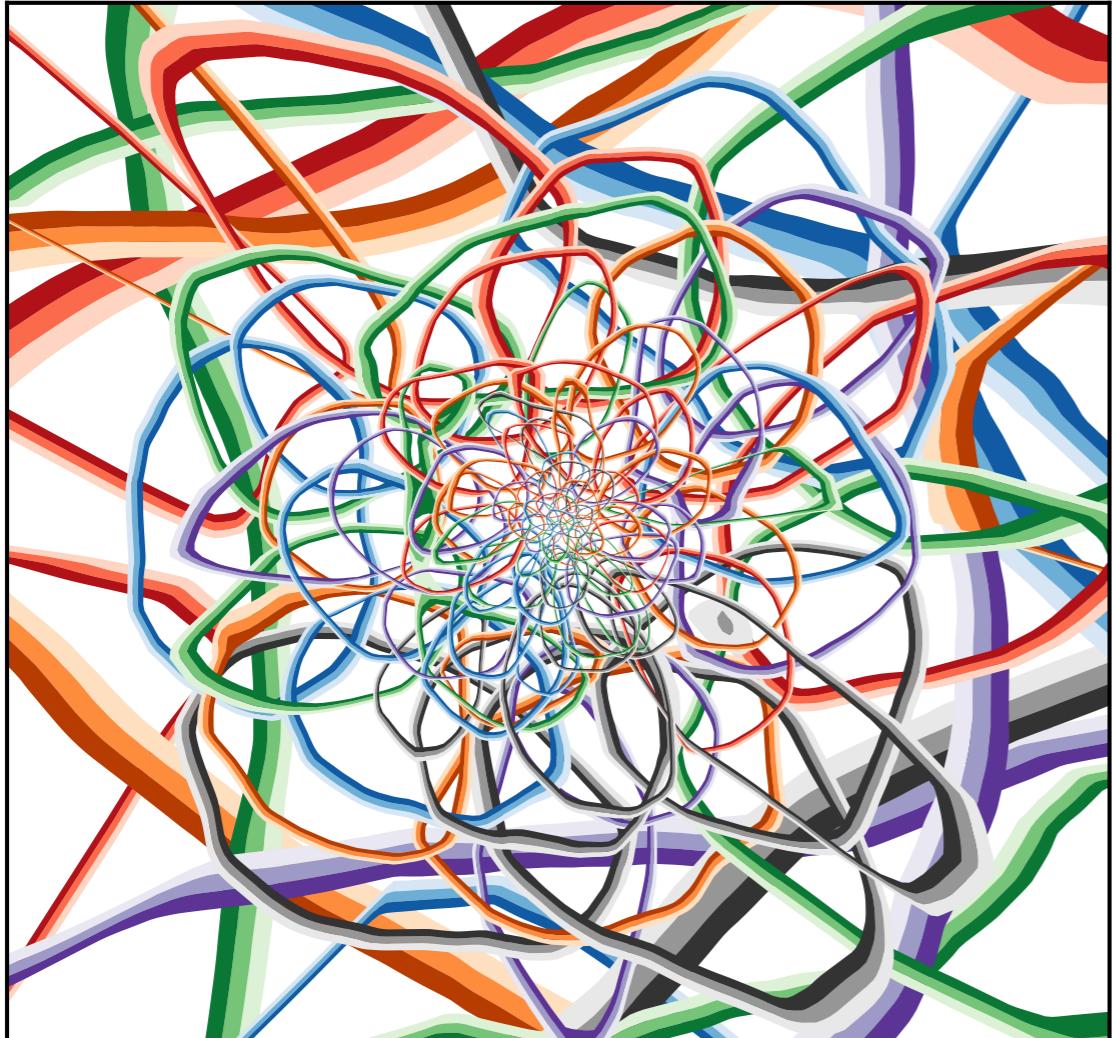


Latent Dimension 64

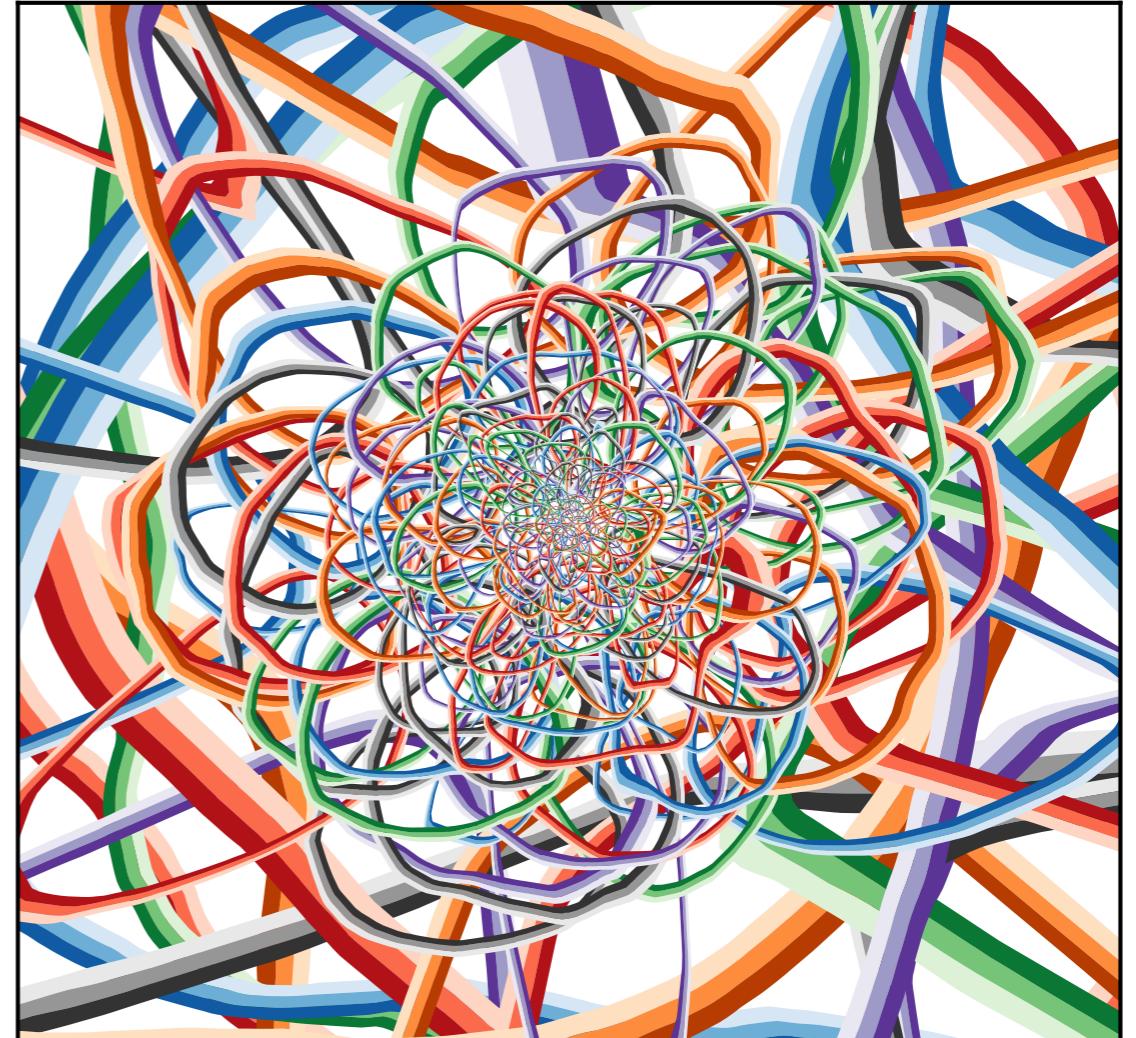


Psychedelic Network Visualization

Latent Dimension 128

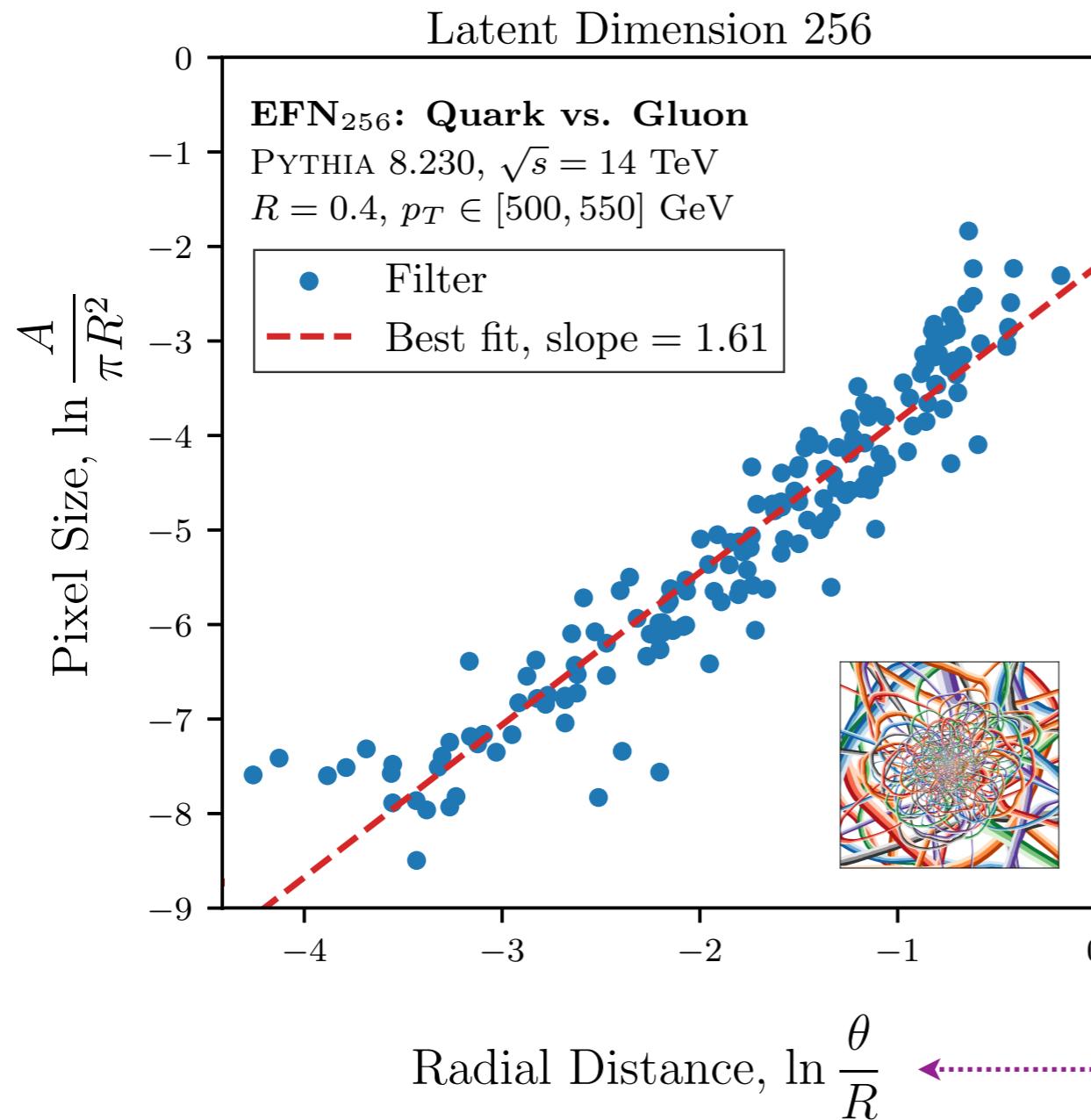


Latent Dimension 256



Singularity structure of QCD!

Putting the AI in Altarelli-Parisi



$$C_q = 4/3$$

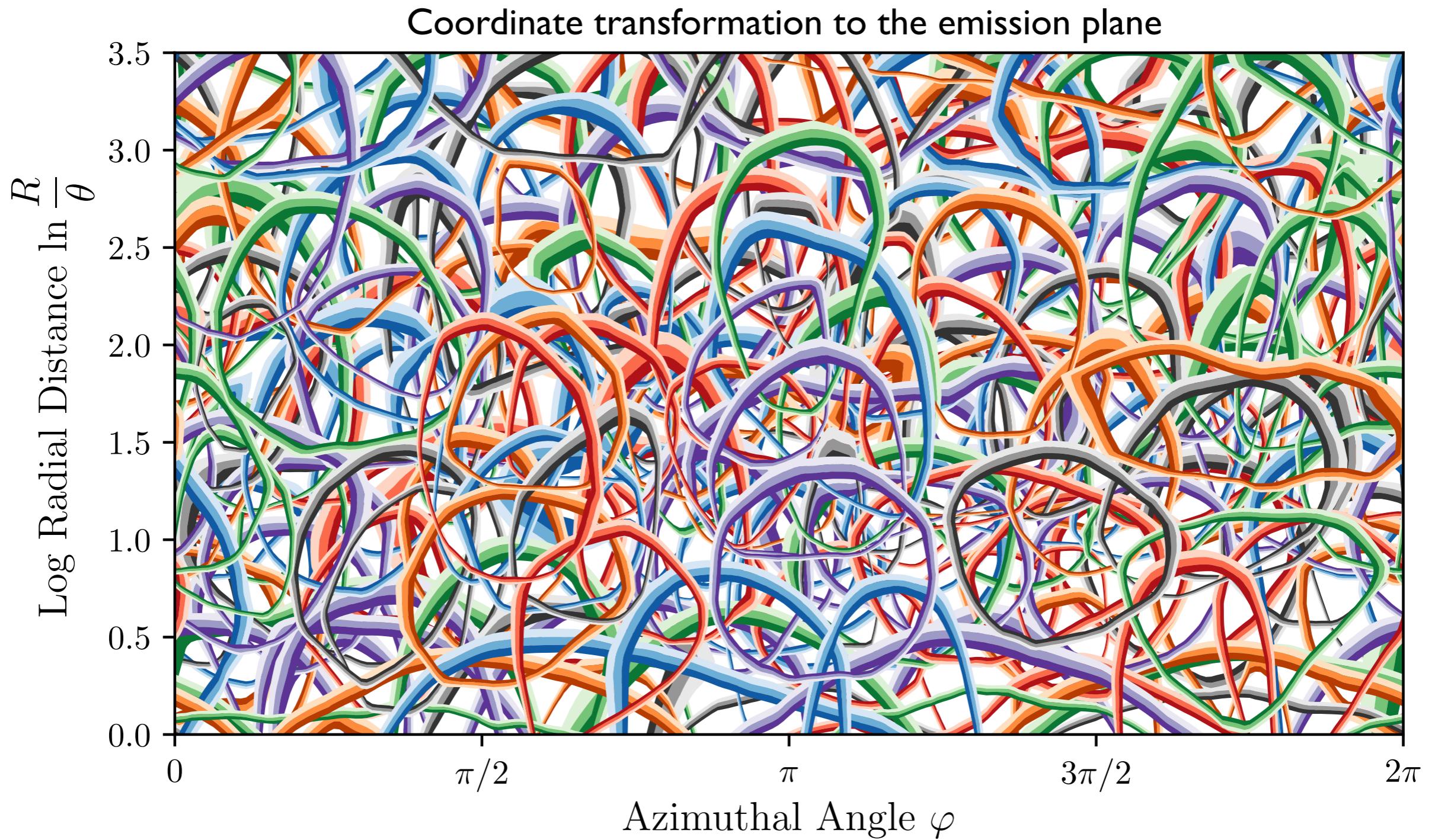
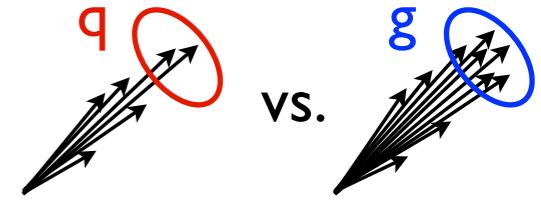
$$C_g = 3$$

$$dP_{i \rightarrow ig} \sim \frac{2\alpha_s}{\pi} C_i \frac{d\theta}{\theta} \frac{dz}{z}$$

Collinear Soft

[Komiske, Metodiev, JDT, JHEP 2019]

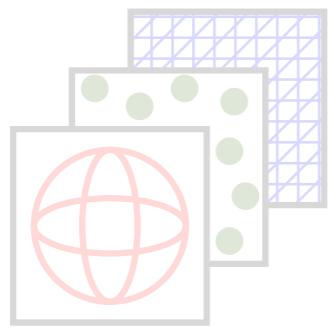
Ready for the moCa?



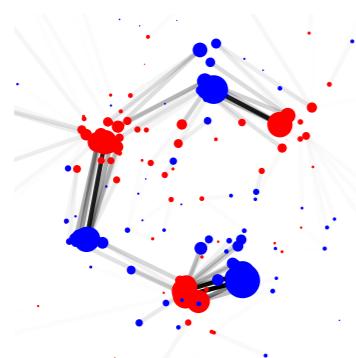
[Komiske, Metodiev, JDT, [JHEP 2019](#); see also Dreyer, Salam, Soyez, [JHEP 2018](#)]



Particle Physics 101



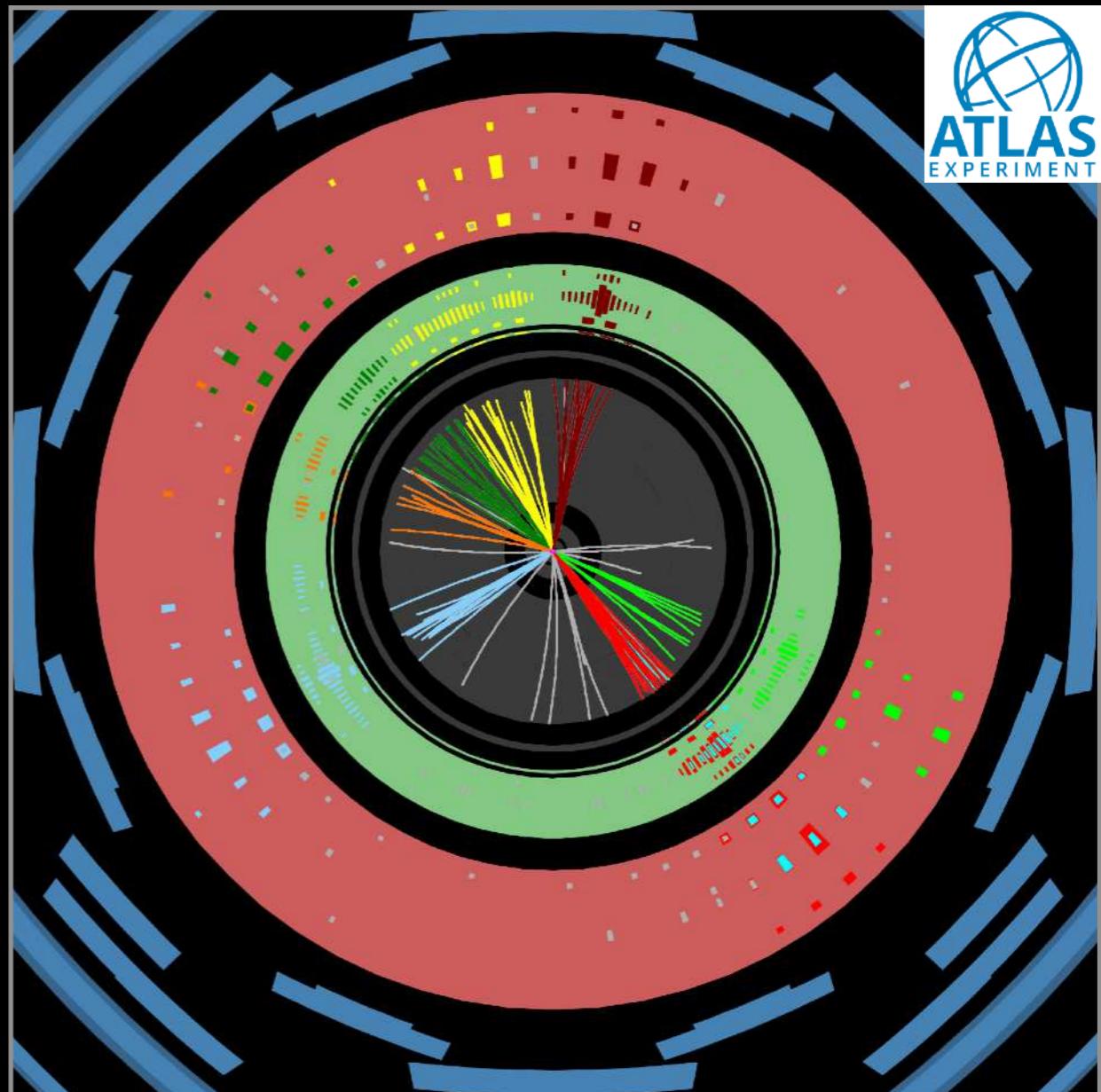
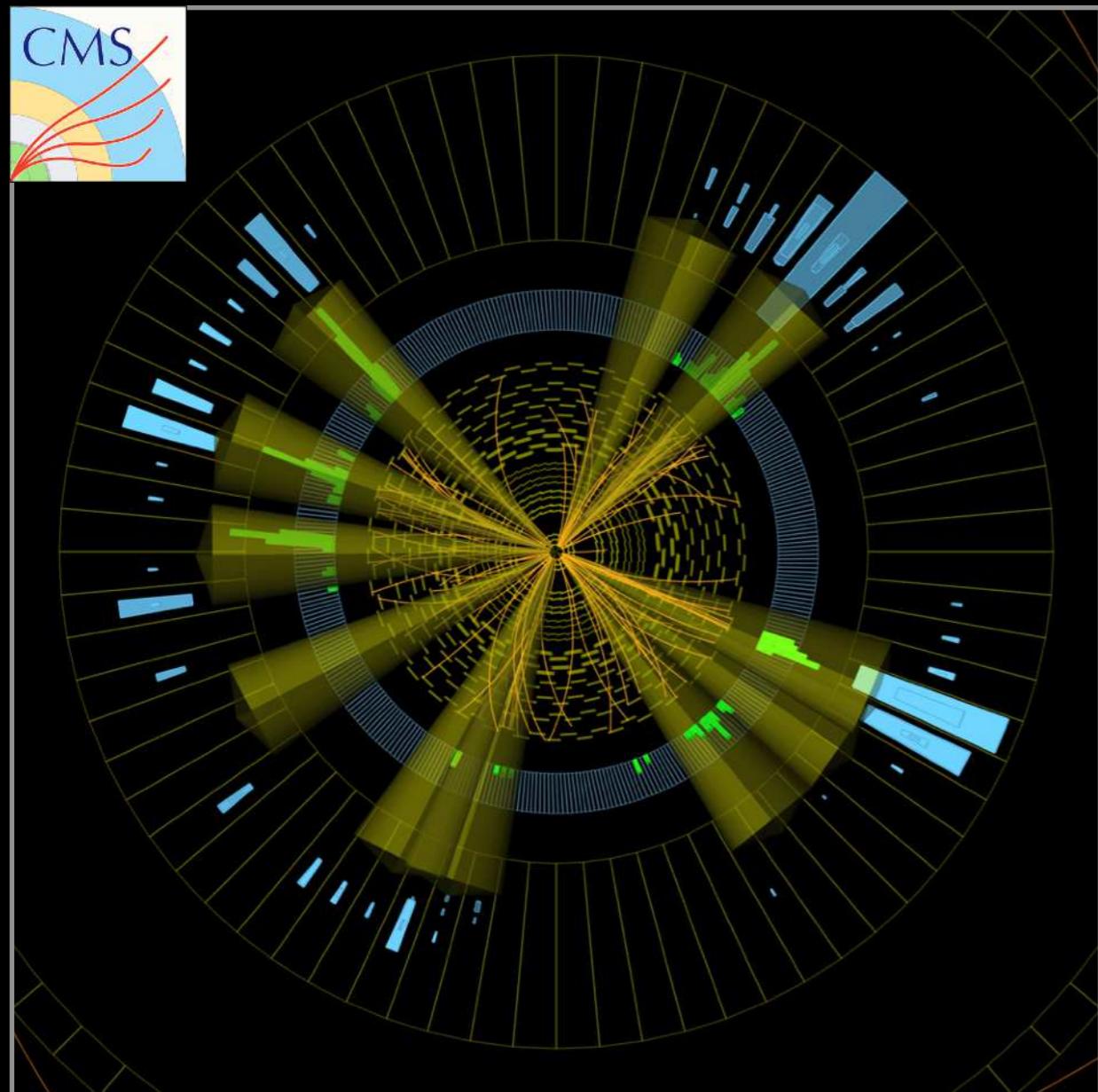
What is a Collider Event?



When are Collider Events Similar?

Two Collider Events

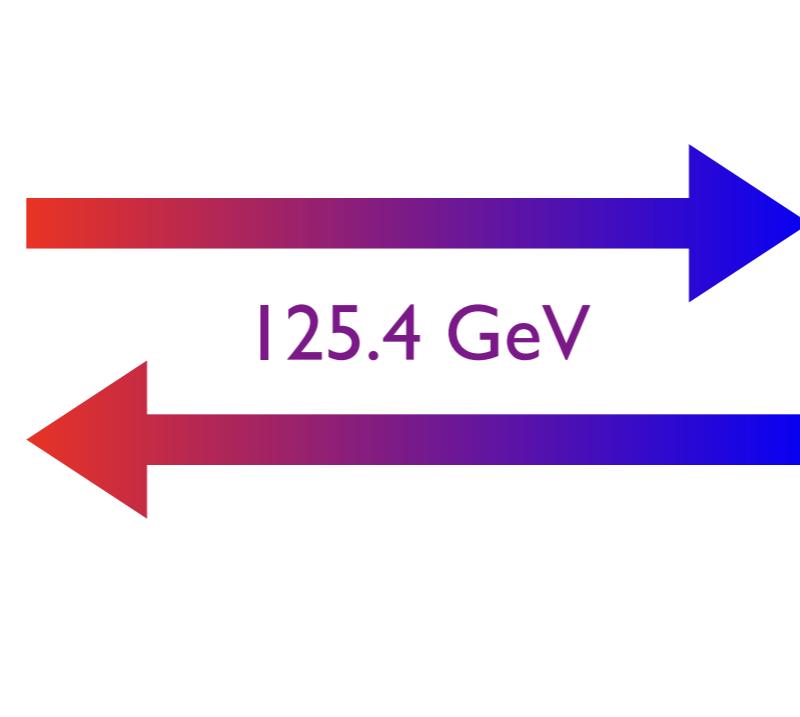
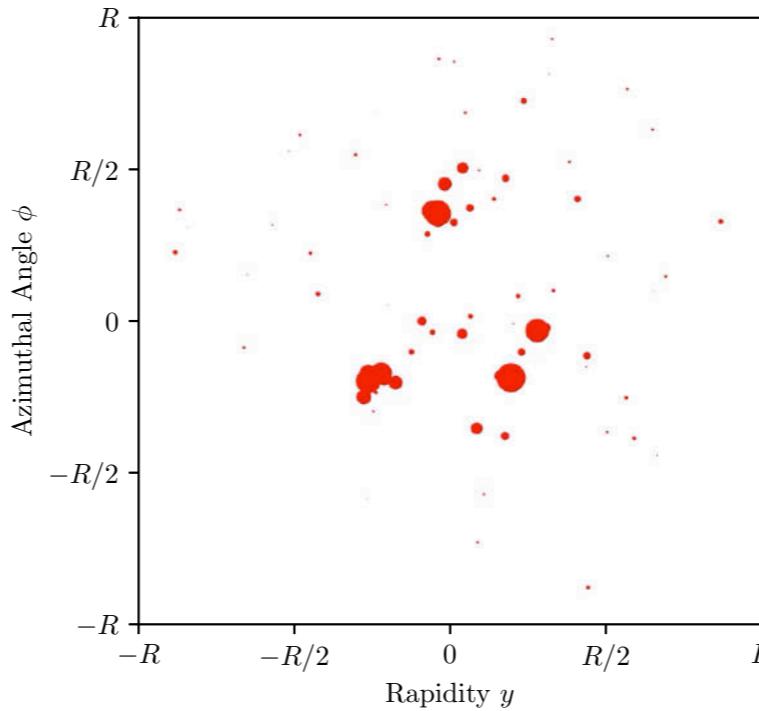
Two collections of points in (momentum) space



How “close” are these? (8.5 km?)

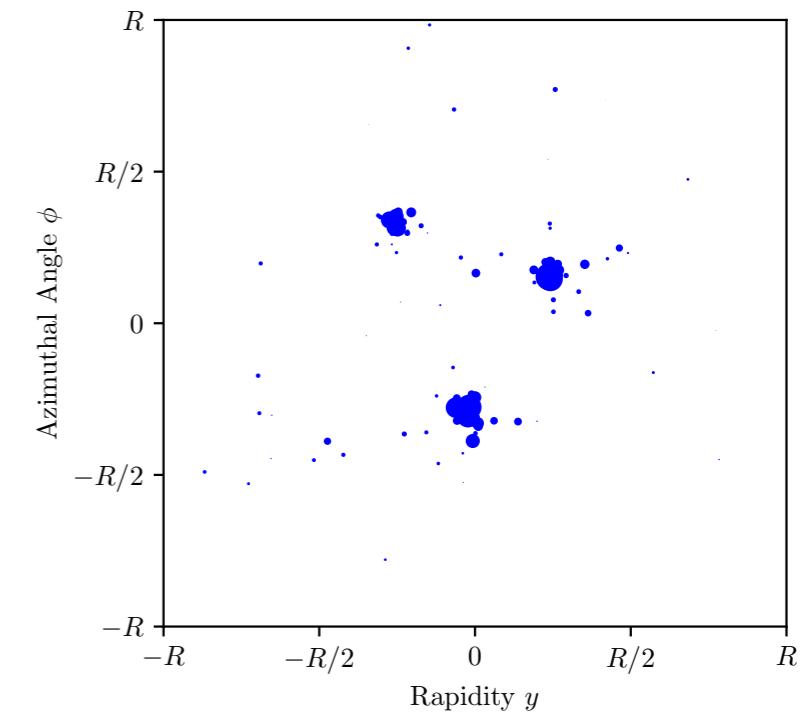
Similarity of Two Energy Flows?

$$\rho(\hat{n}) = \sum_{i \in \mathcal{J}} \textcolor{teal}{E}_i \delta^{(2)}(\hat{n} - \hat{n}_i)$$



Optimal Transport:
Earth Mover's Distance
a.k.a. l -Wasserstein metric

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

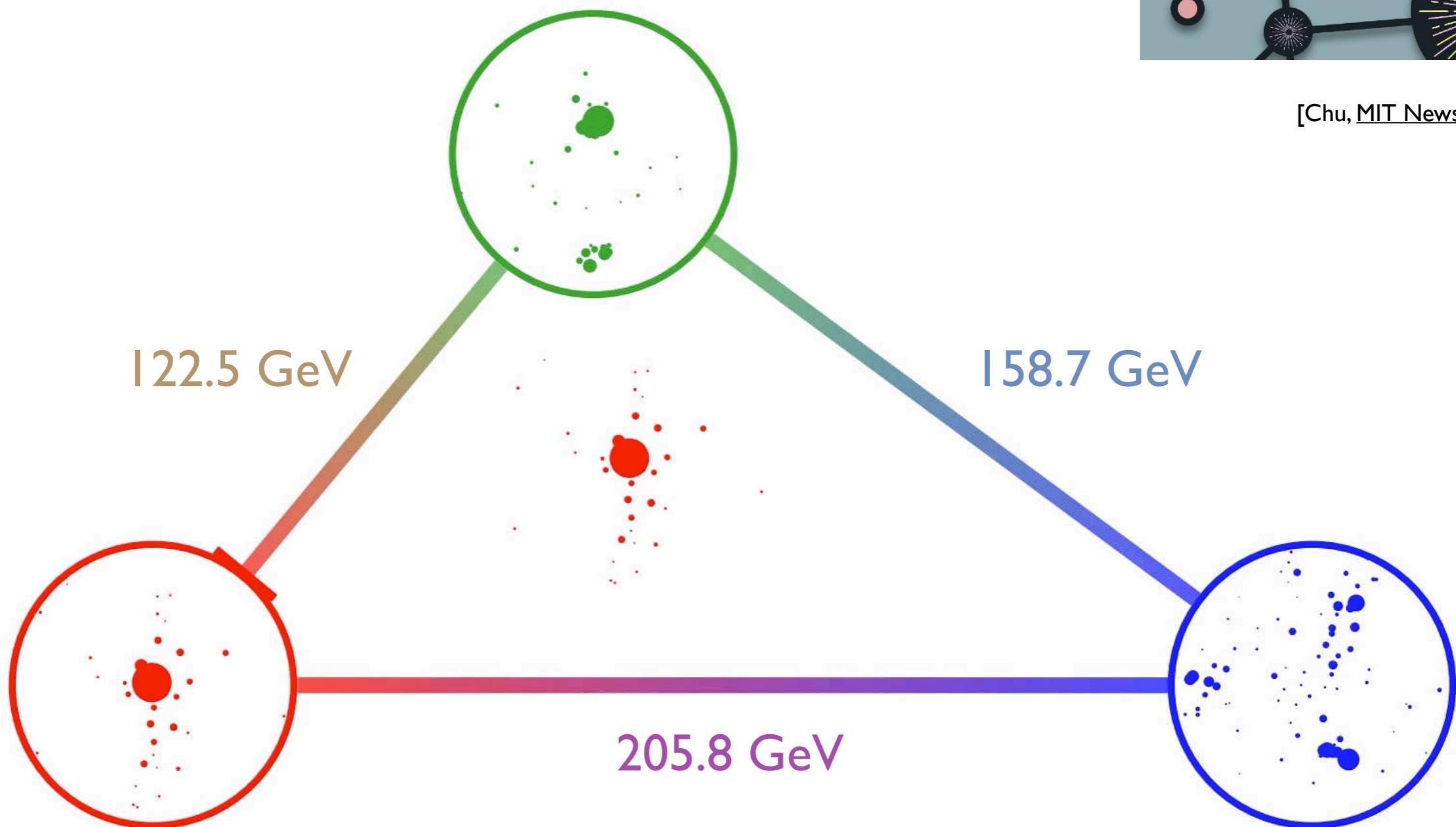


[Komiske, Metodiev, JDT, [PRL 2019](#); code at Komiske, Metodiev, JDT, [energyflow.network](#)]

Similarity of Three Energy Flows?



[Chu, MIT News July 2019]

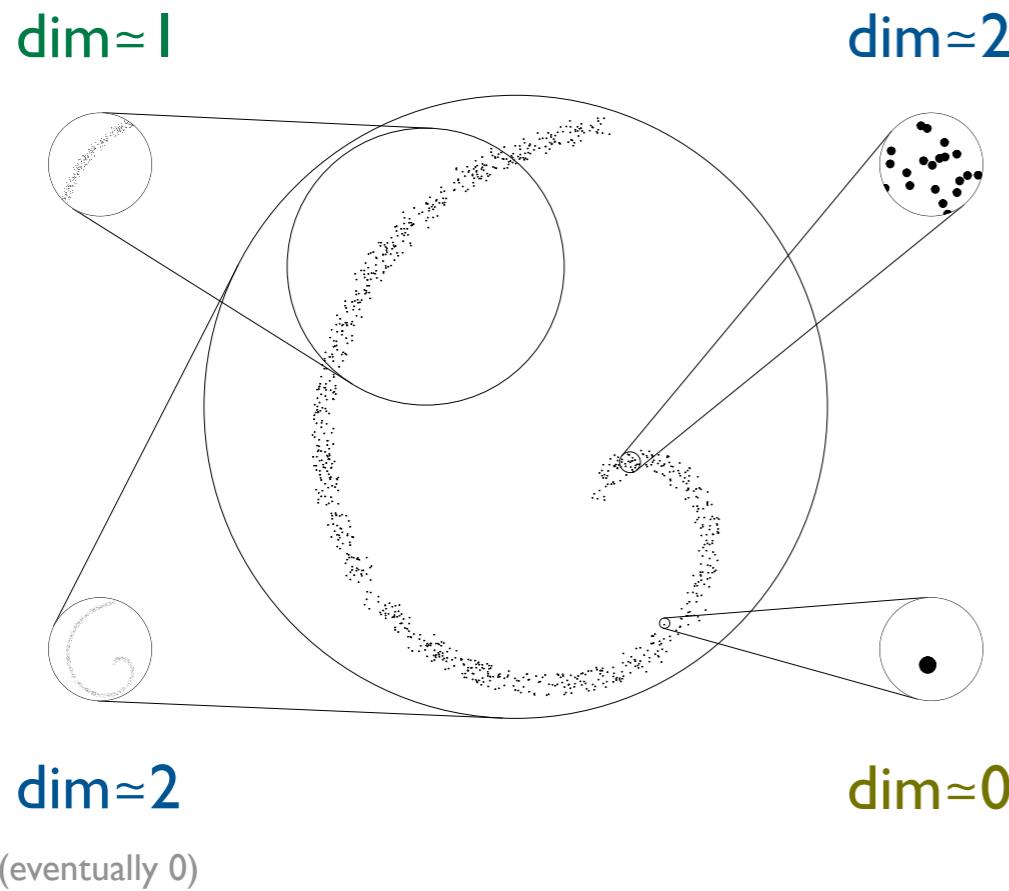


[Komiske, Metodiev, JDT, PRL 2019; code at Komiske, Metodiev, JDT, [energyflow.network](#)]

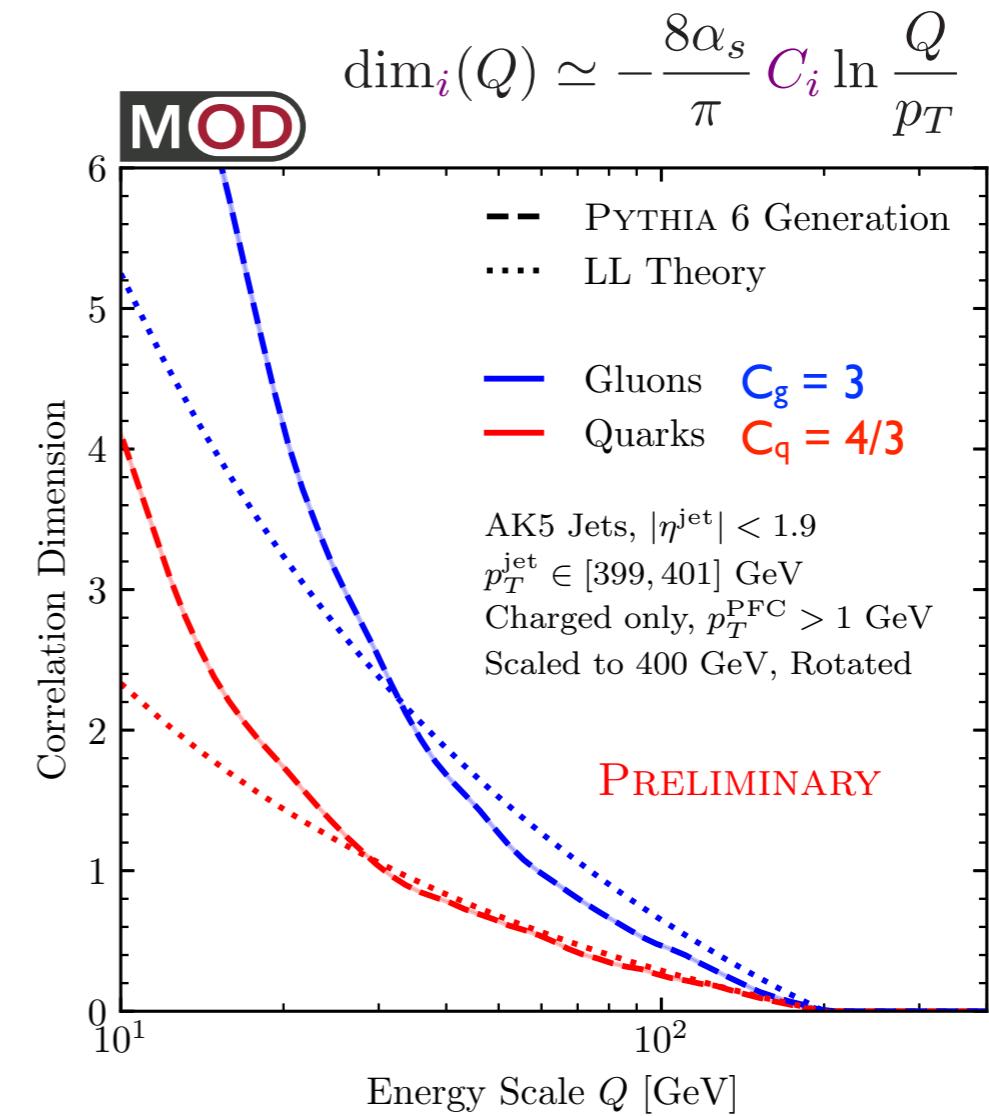
Dimensionality of Space of Jets

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

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



QCD Calculation



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

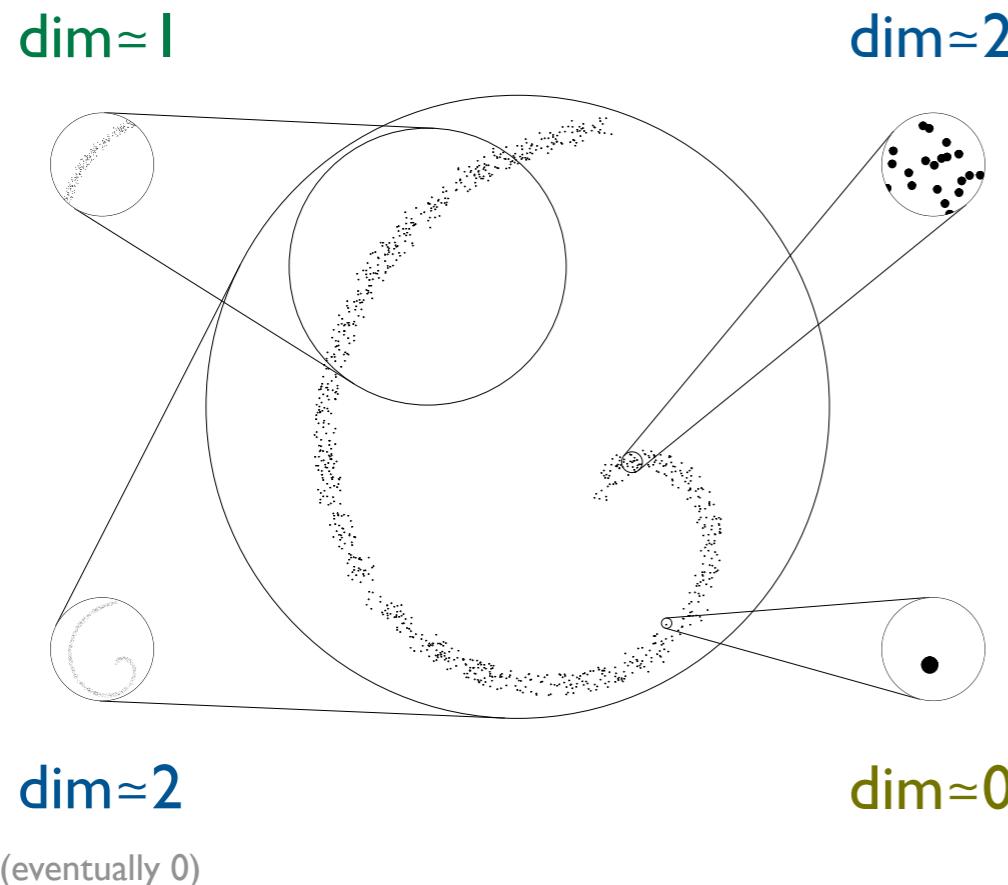
Dimensionality of Space of Jets



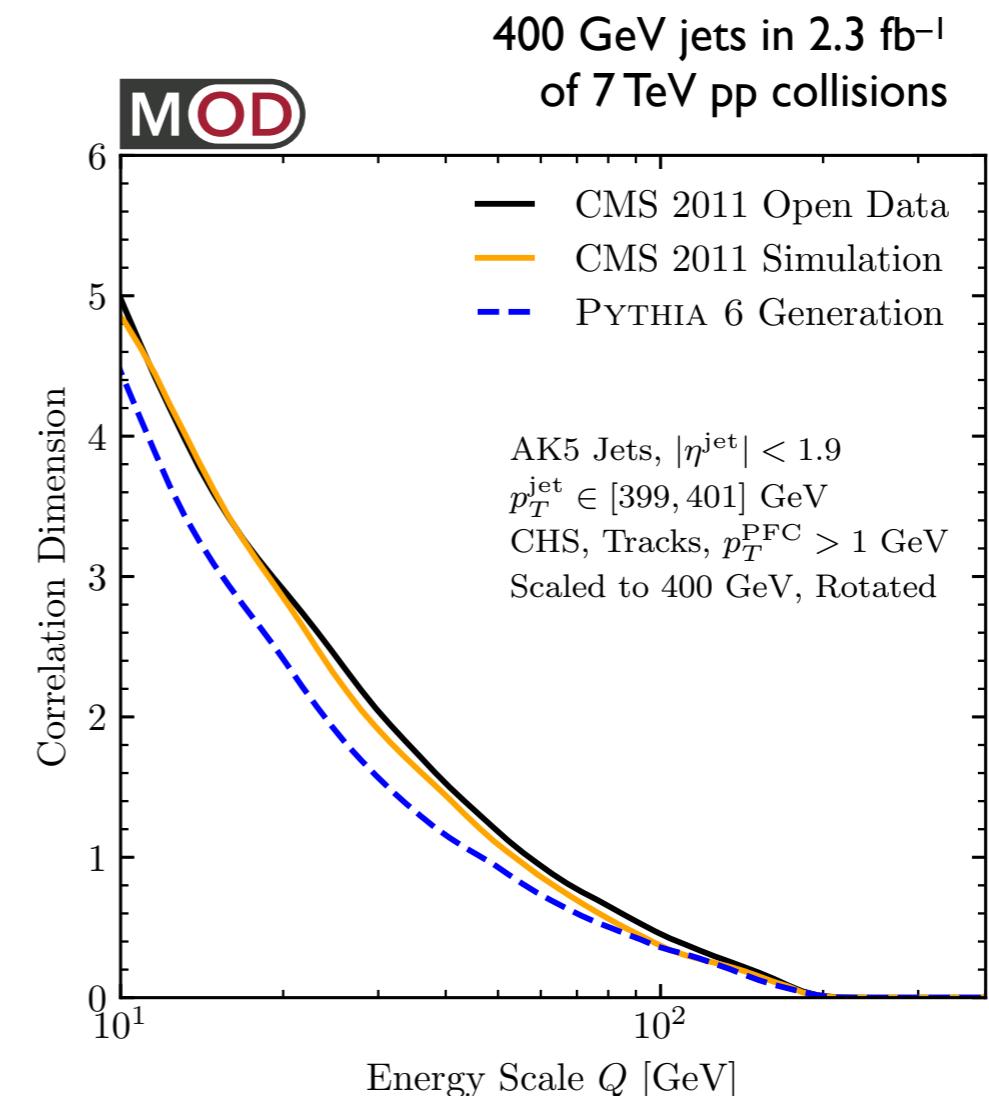
[<http://opendata.cern.ch/>]

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

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



CMS Open Data



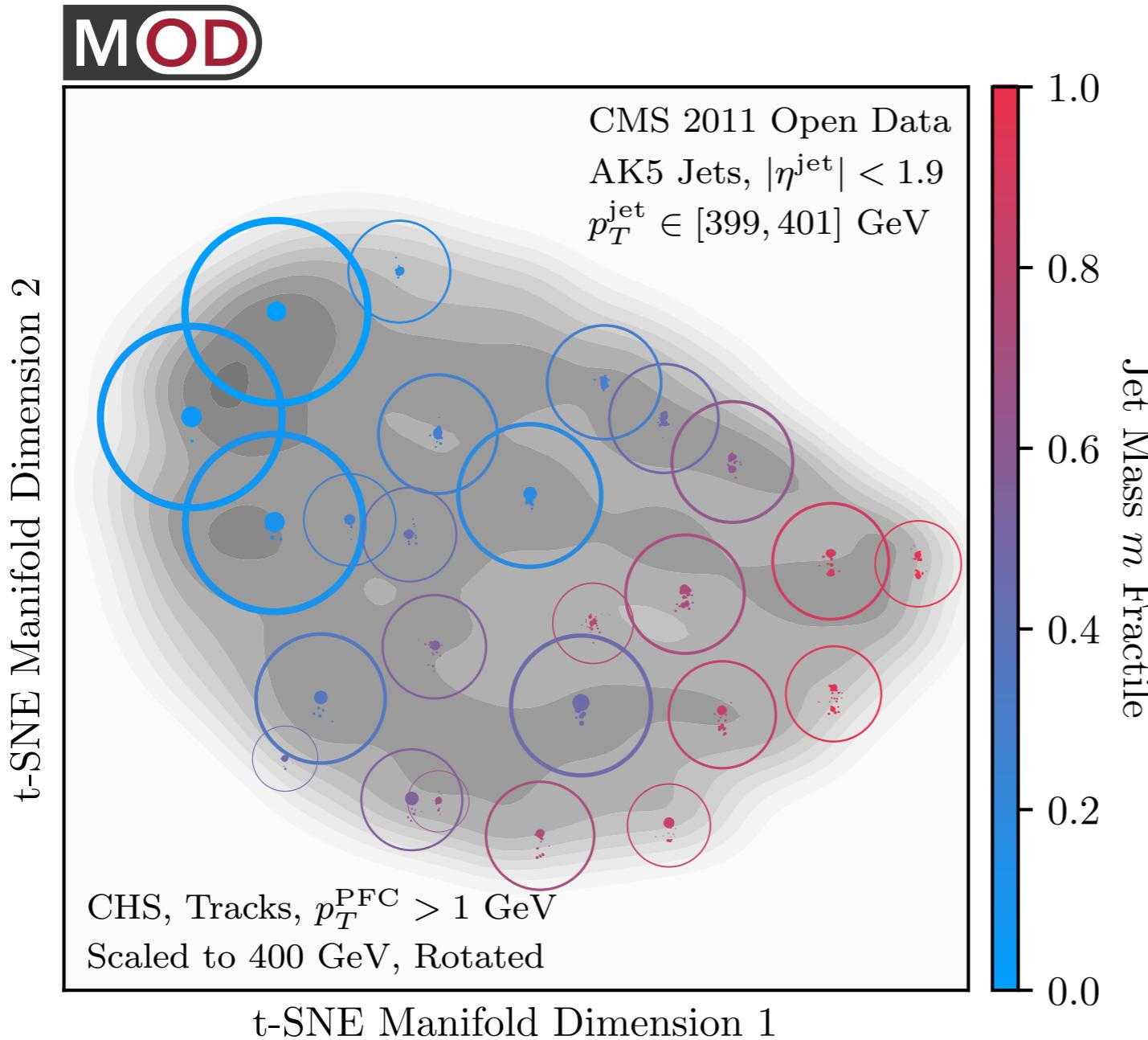
[Komiske, Mastandrea, Metodiev, Naik, JDT, [submitted to PRD](#)]



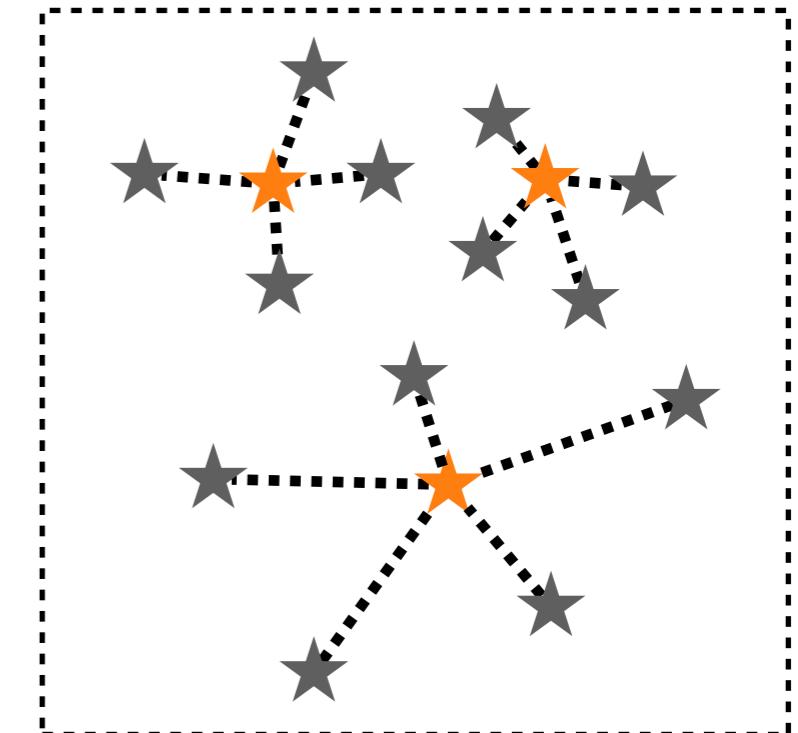
Most Representative Jets



[<http://opendata.cern.ch/>]



k-medoids
Arranged via $t\text{-SNE}$

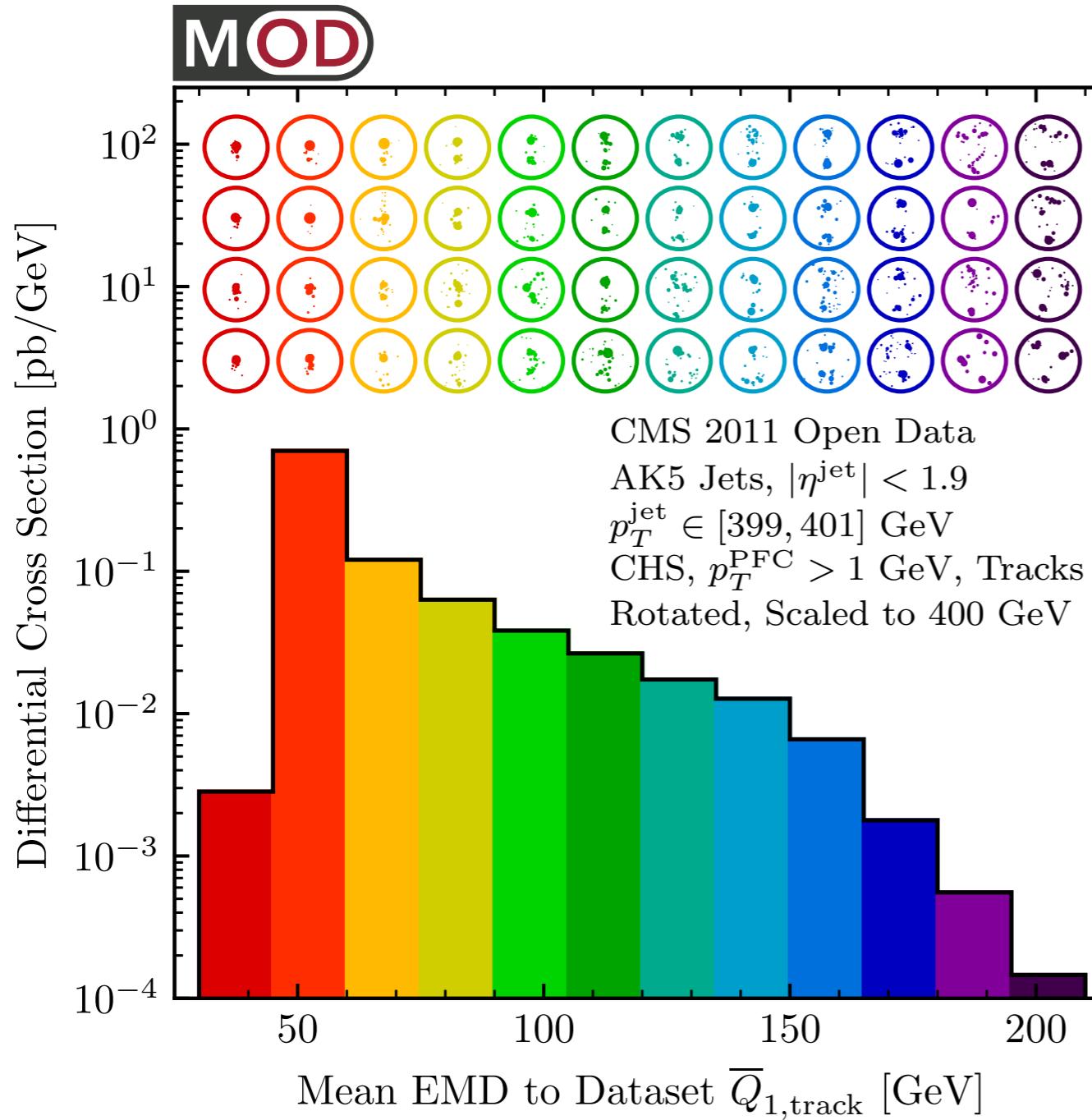


[Komiske, Mastandrea, Metodiev, Naik, JDT, [submitted to PRD](#); using van der Maaten, Hinton, [JMLR 2008](#)]

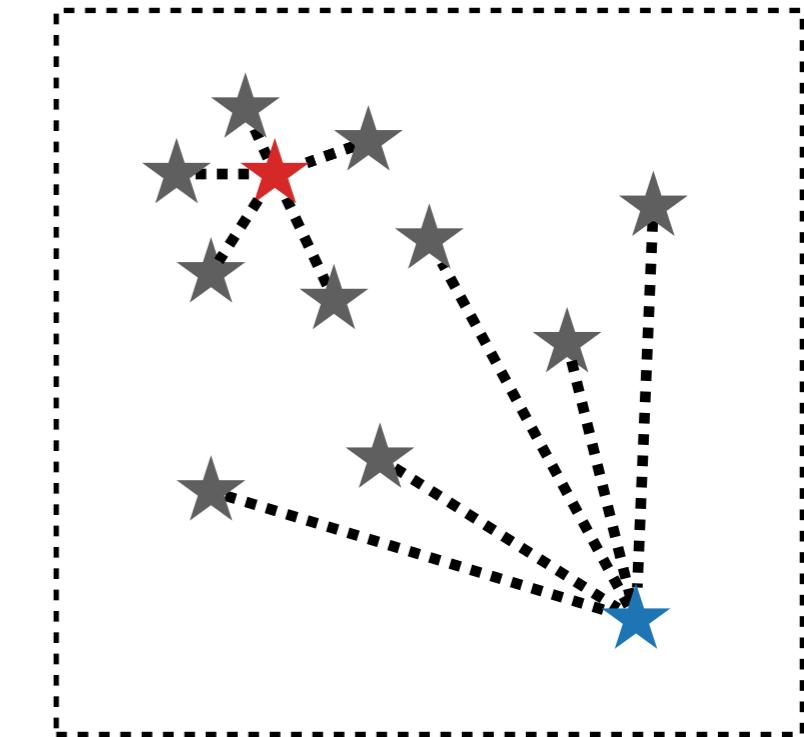
Least Representative Jets



[<http://opendata.cern.ch/>]



New Physics?
Or tails of QCD?



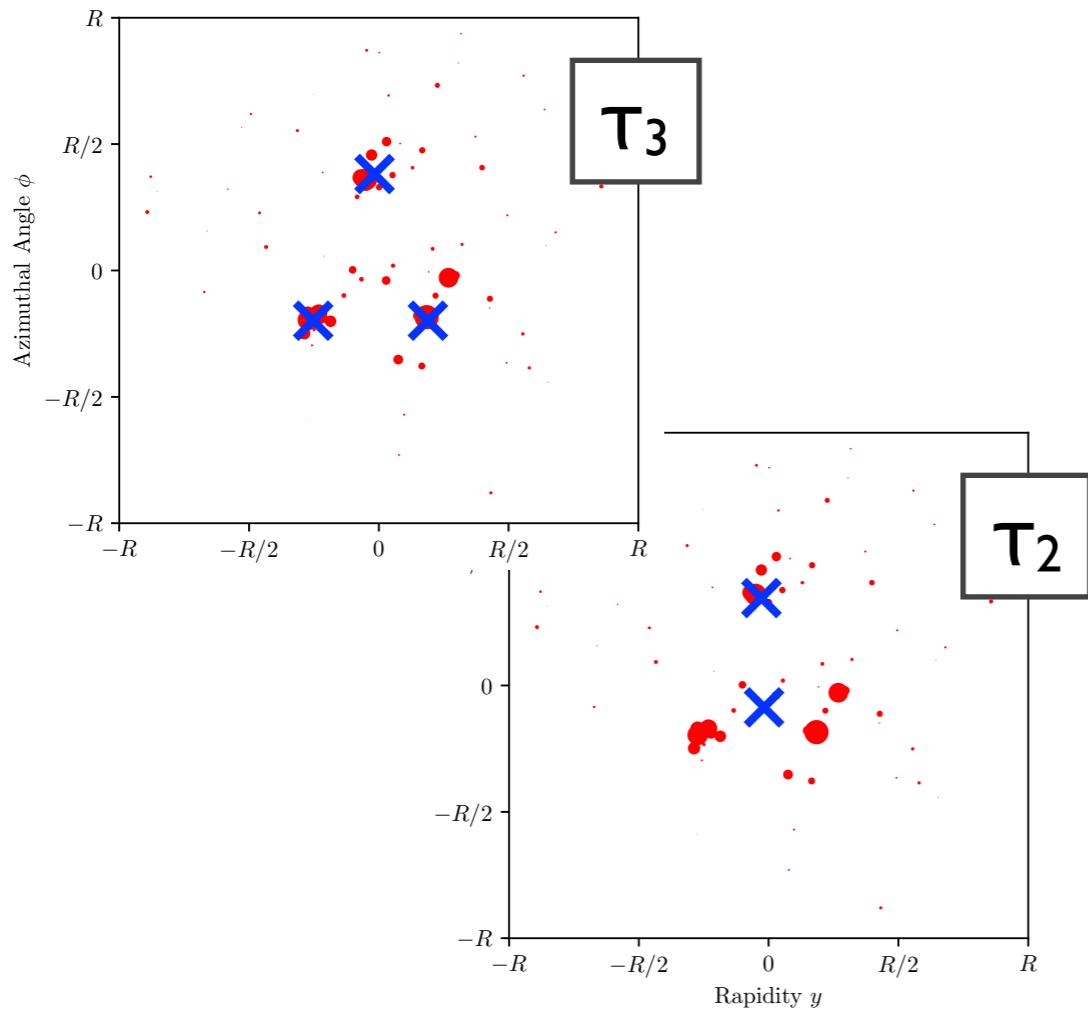
[Komiske, Mastandrea, Metodiev, Naik, JDT, submitted to PRD]

N-subjettiness

Ubiquitous jet substructure observable used for almost a decade...

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

↑ IRC safe



[JDT, Van Tilburg, [JHEP 2011](#), [JHEP 2012](#);
based on Brandt, Dahmen, [ZPC 1979](#); Stewart, Tackmann, Waalewijn, [PRL 2010](#)]

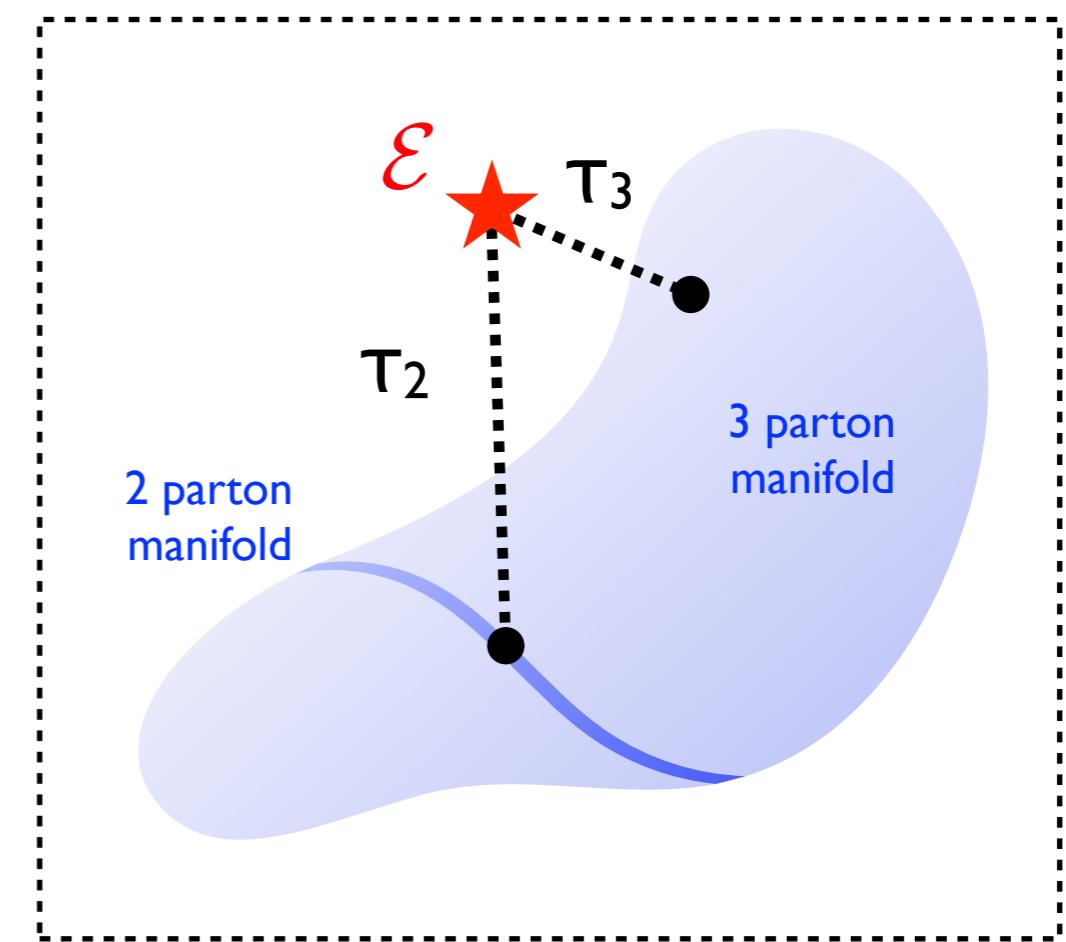
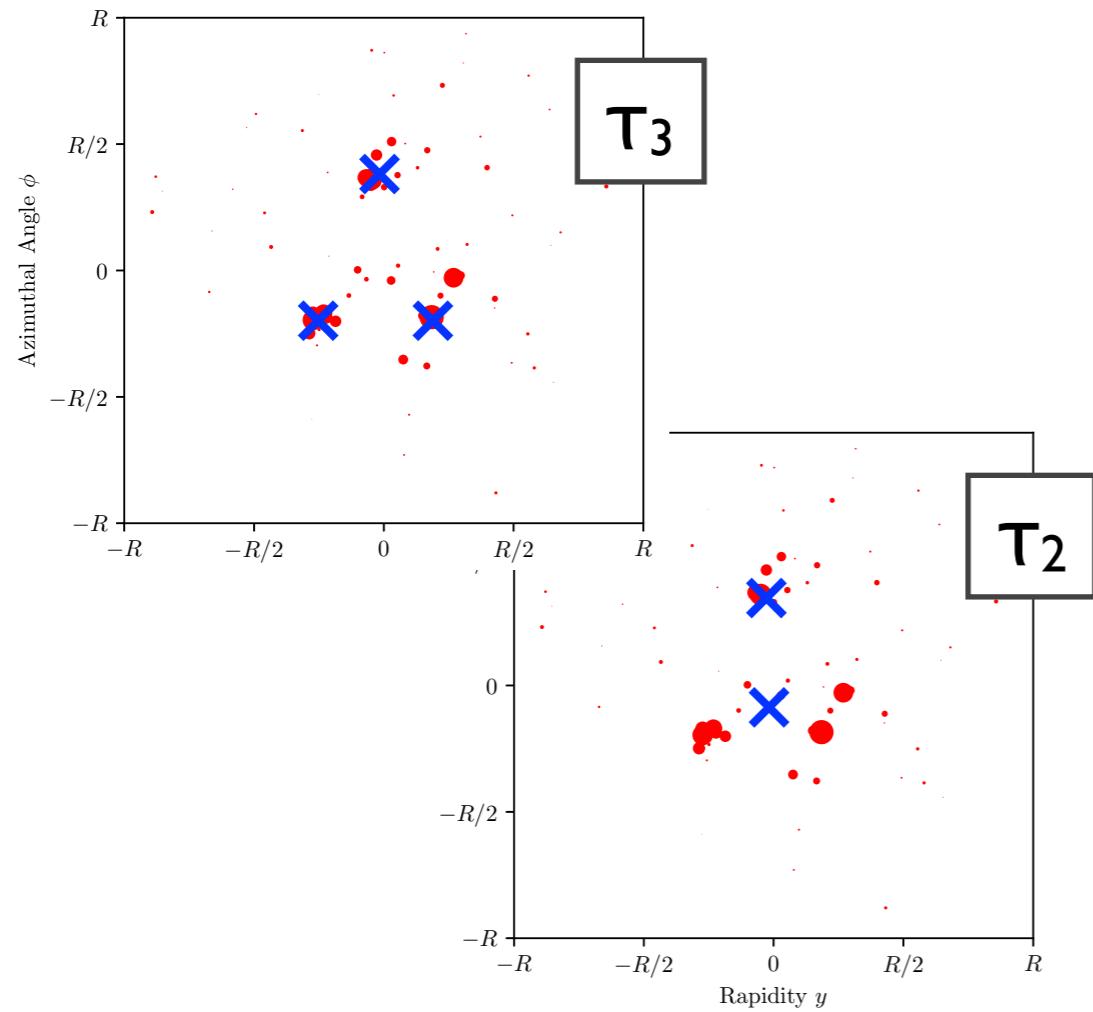
N-subjettiness = Point to Manifold EMD

...is secretly an optimal transport problem

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

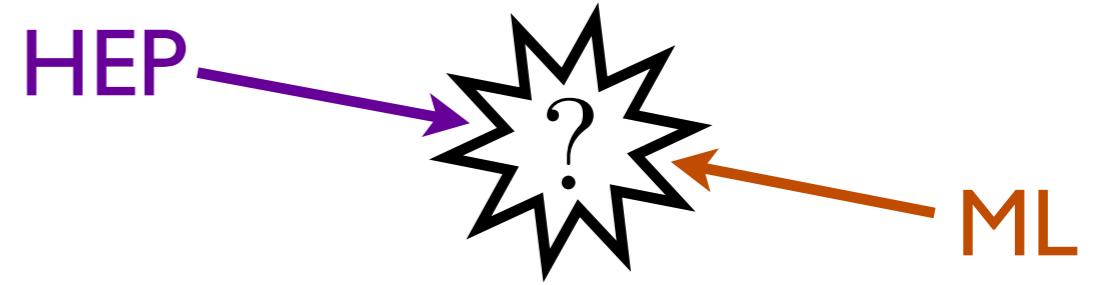
↑ IRC safe

(!)

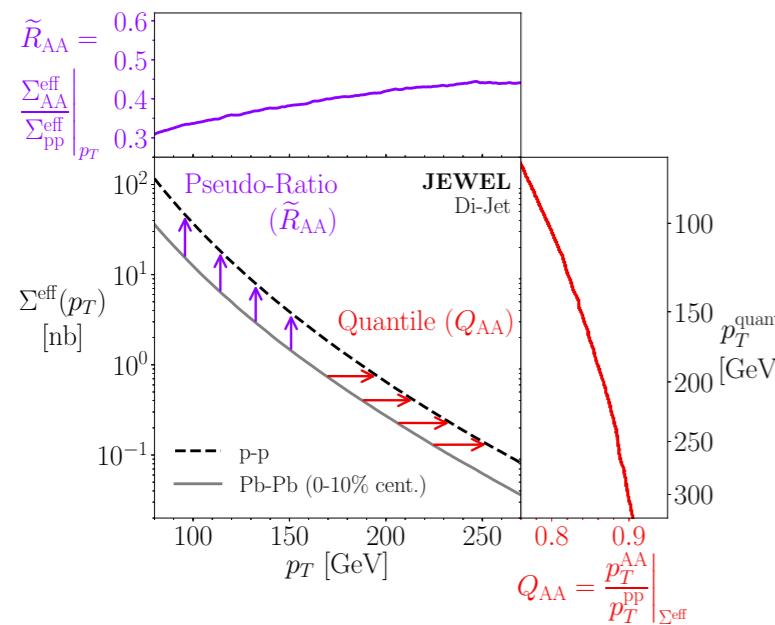


[JDT, Van Tilburg, JHEP 2011, JHEP 2012;
rephrased in the language of Komiske, Metodiev, JDT, PRL 2019]

More Collisions

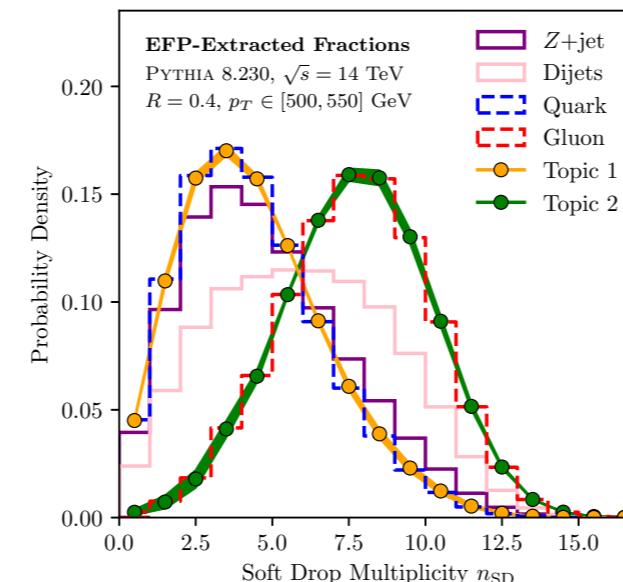


Jet Quenching via Optimal Transport



[Brewer, Milhano, JDT, PRL 2019]

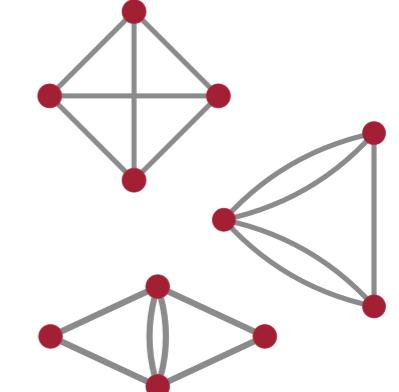
Unsupervised Jet Classification via Blind Source Separation



[Komiske, Metodiev, JDT, JHEP 2018]

Kinematic Decomposition via Graph Theory

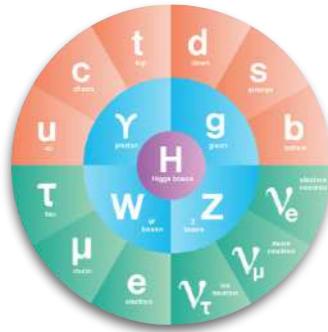
Edges d	Leafless Multigraphs			
	Connected	All	A307317	A307316
1	0	0		
2	1	1		
3	2	2		
4	4	5		
5	9	11		
6	26	34		
7	68	87		
8	217	279		
9	718	897		
10	2 553	3 129		
11	9 574	11 458		
12	38 005	44 576		
13	157 306	181 071		
14	679 682	770 237		
15	3 047 699	3 407 332		
16	14 150 278	15 641 159		



[Komiske, Metodiev, JDT, to appear]

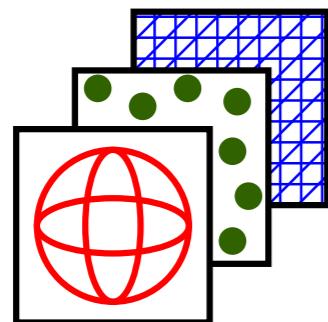
New insights into particle physics*
facilitated by advances in machine learning*

Summary



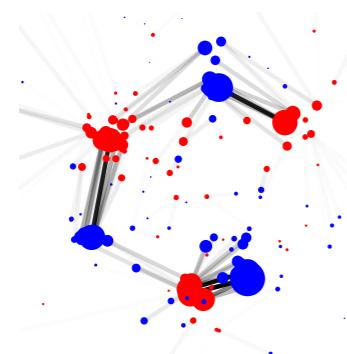
Particle Physics 101

*High-energy collisions can yield insights into fundamental physics
Machine learning offers powerful tools to analyze collision debris*



What is a Collider Event?

*Unordered set of particles describing energy flow of jets
Inspires network architectures designed for symmetry and safety*

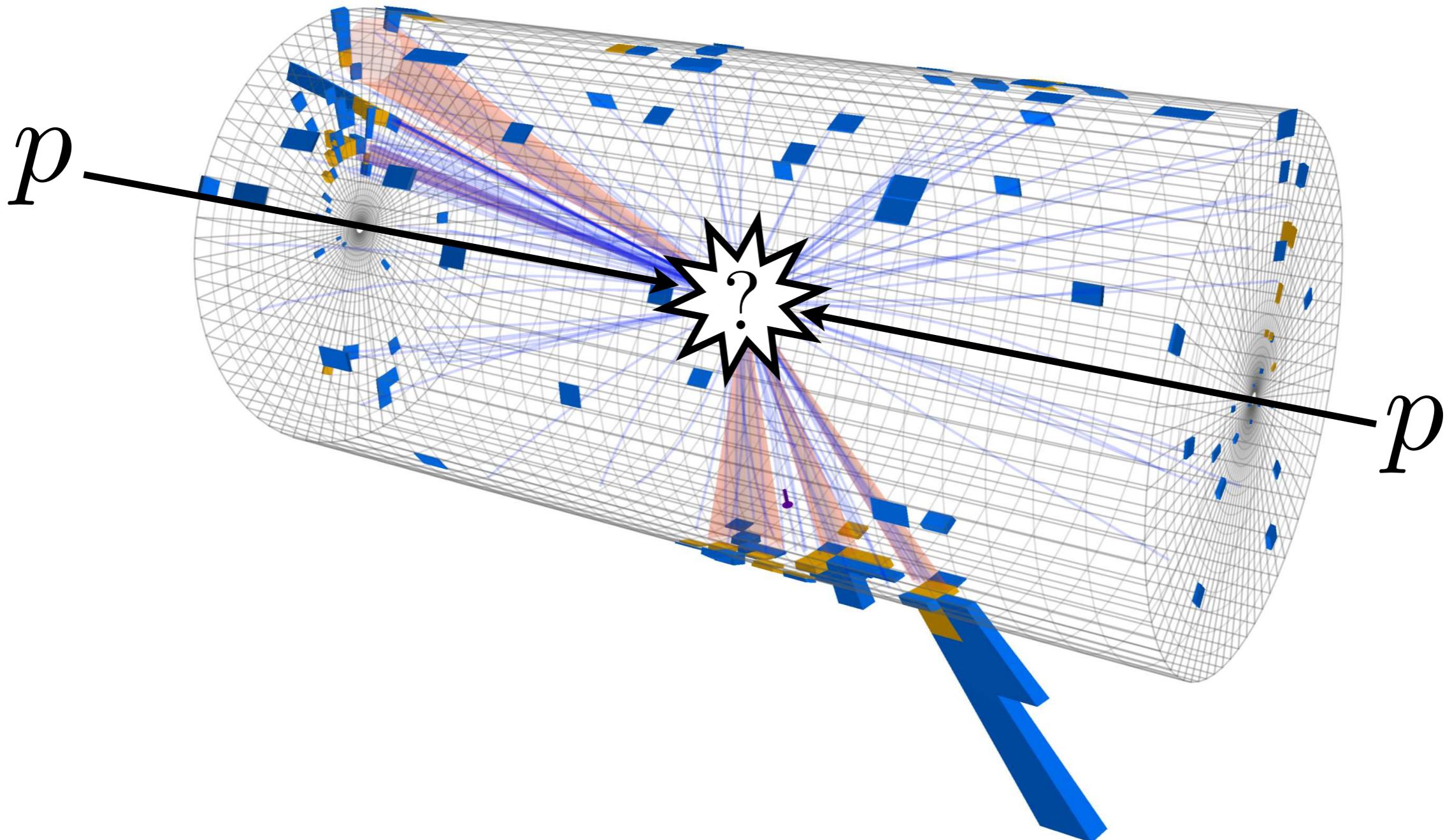


When are Collider Events Similar?

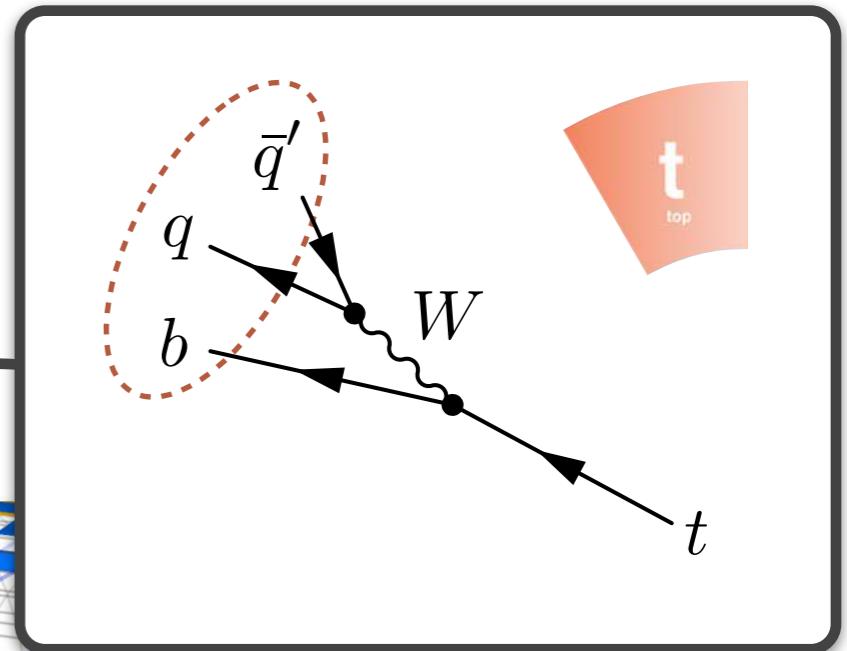
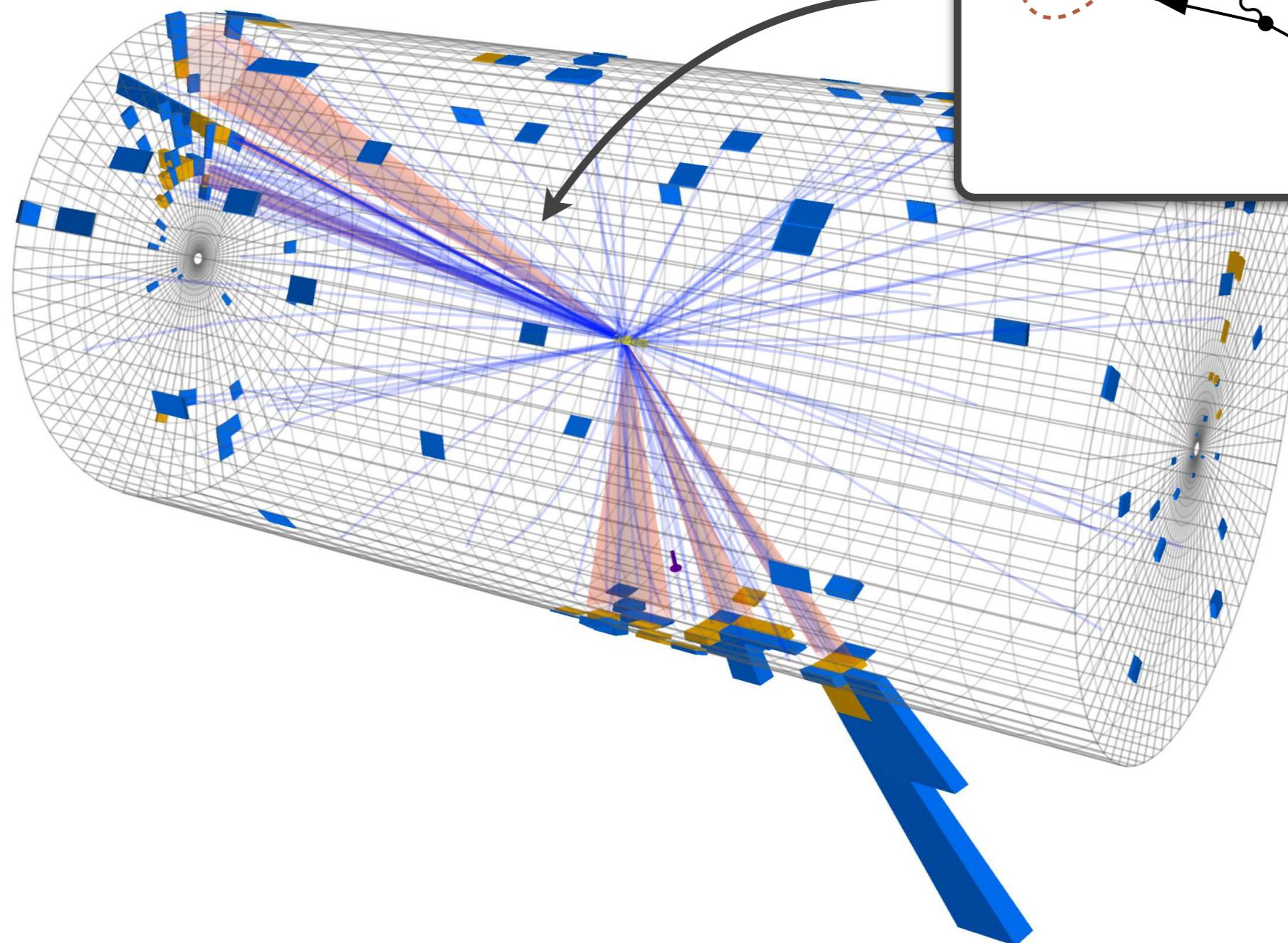
*When their energy flows are similar
Inspires unsupervised learning strategies based on event geometry*

Backup Slides

“Collision Course”

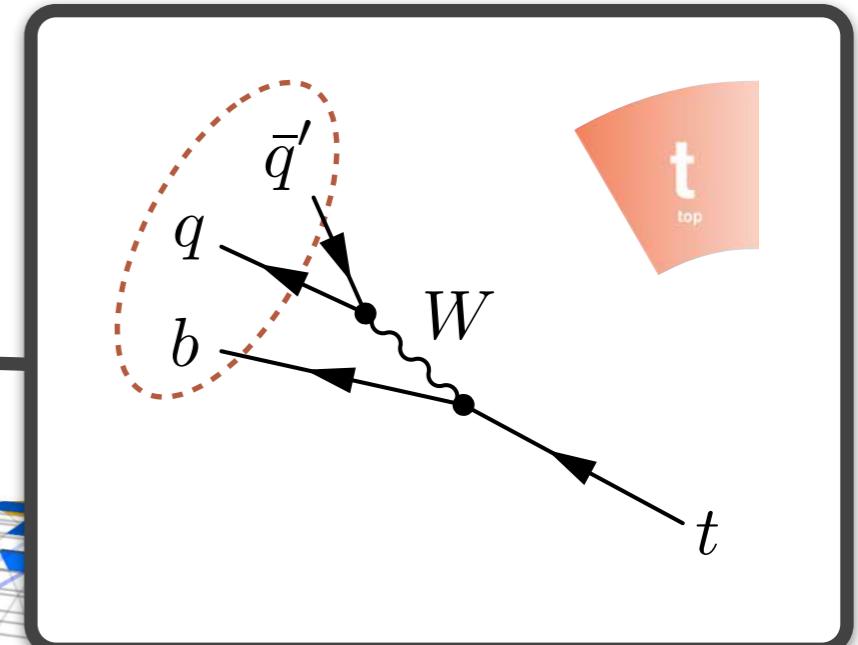
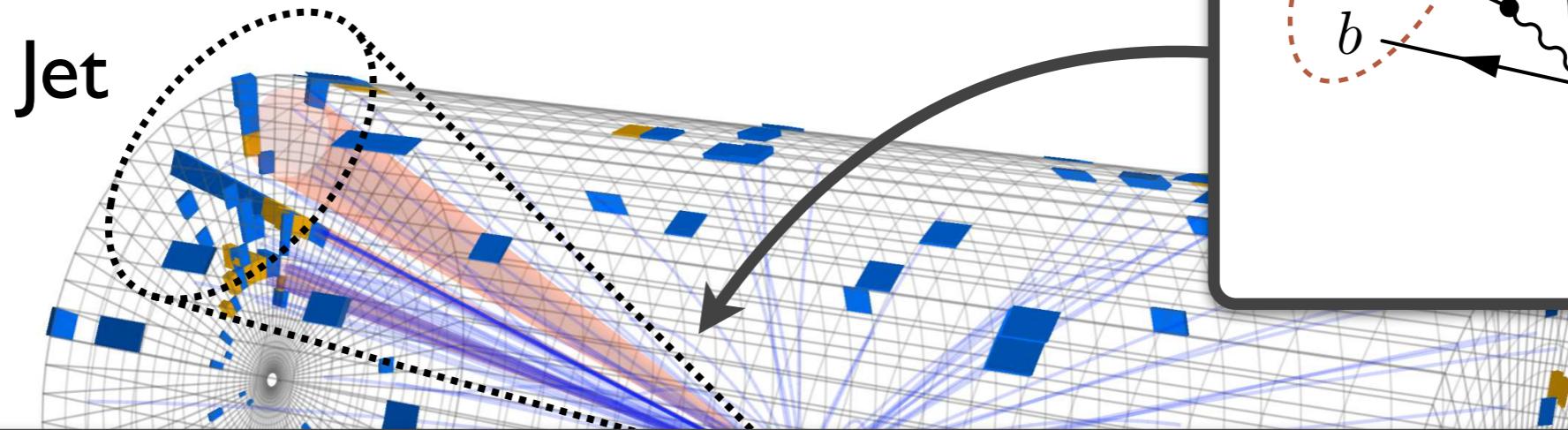


“Collision Course”





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



“Deep Thinking”?

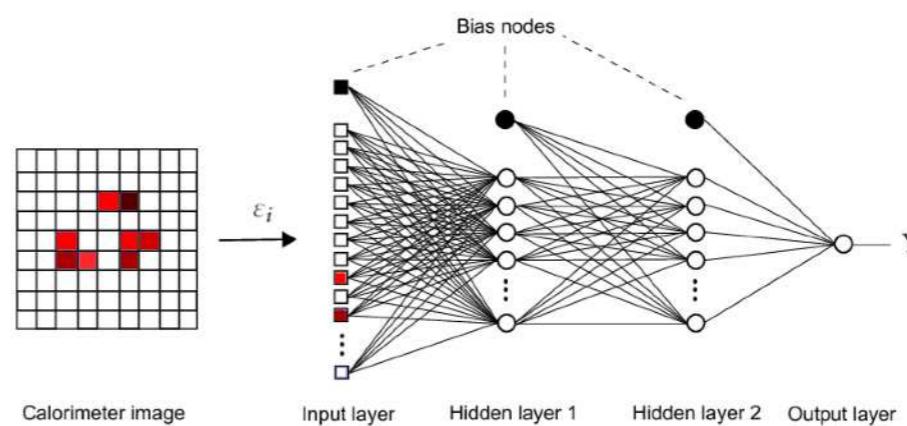
[e.g. JDT, Van Tilburg, [1011.2268](#), [1108.2701](#);
 rephrased in language of Komiske, Metodiev, JDT, [1902.02346](#)]

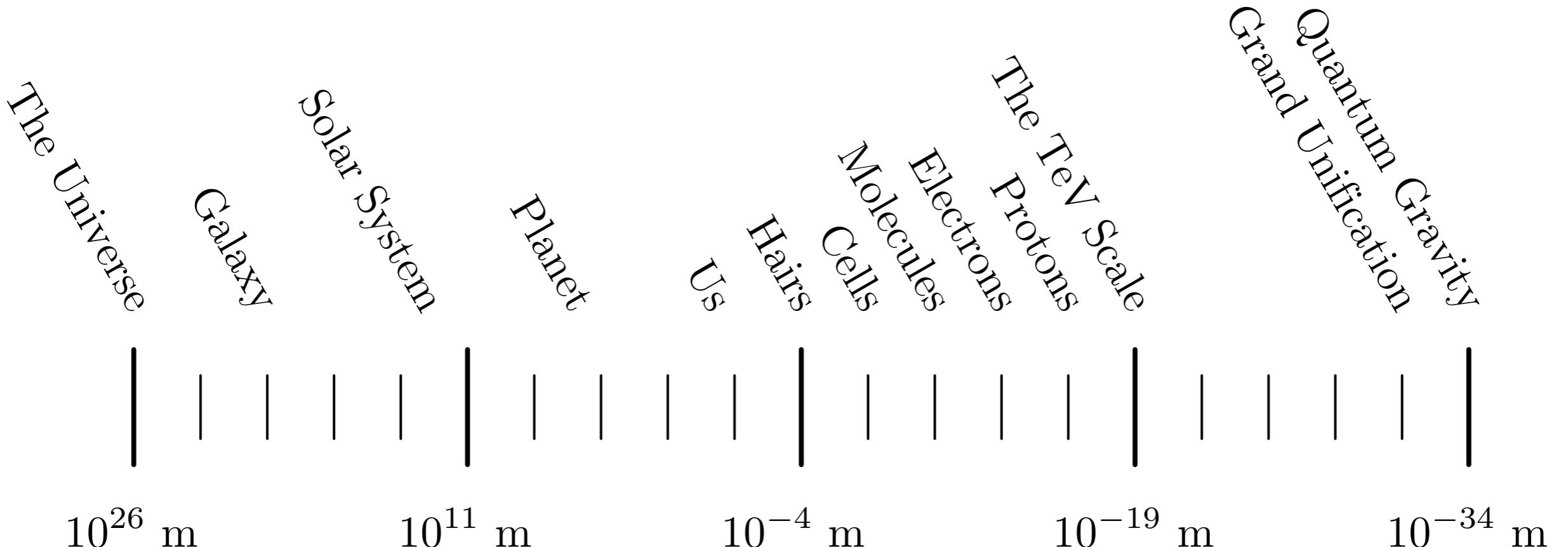
$$\tau_N(\mathcal{J}) = \min_{|\mathcal{J}'|=N} \text{EMD}(\mathcal{J}, \mathcal{J}')$$

Ask me about
 this formula...

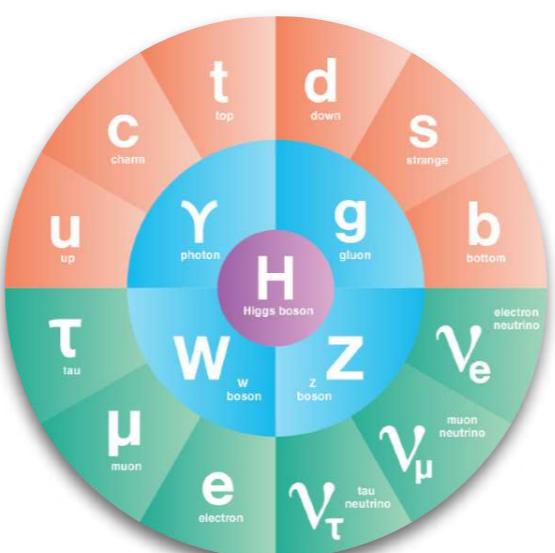
“Deep Learning”?

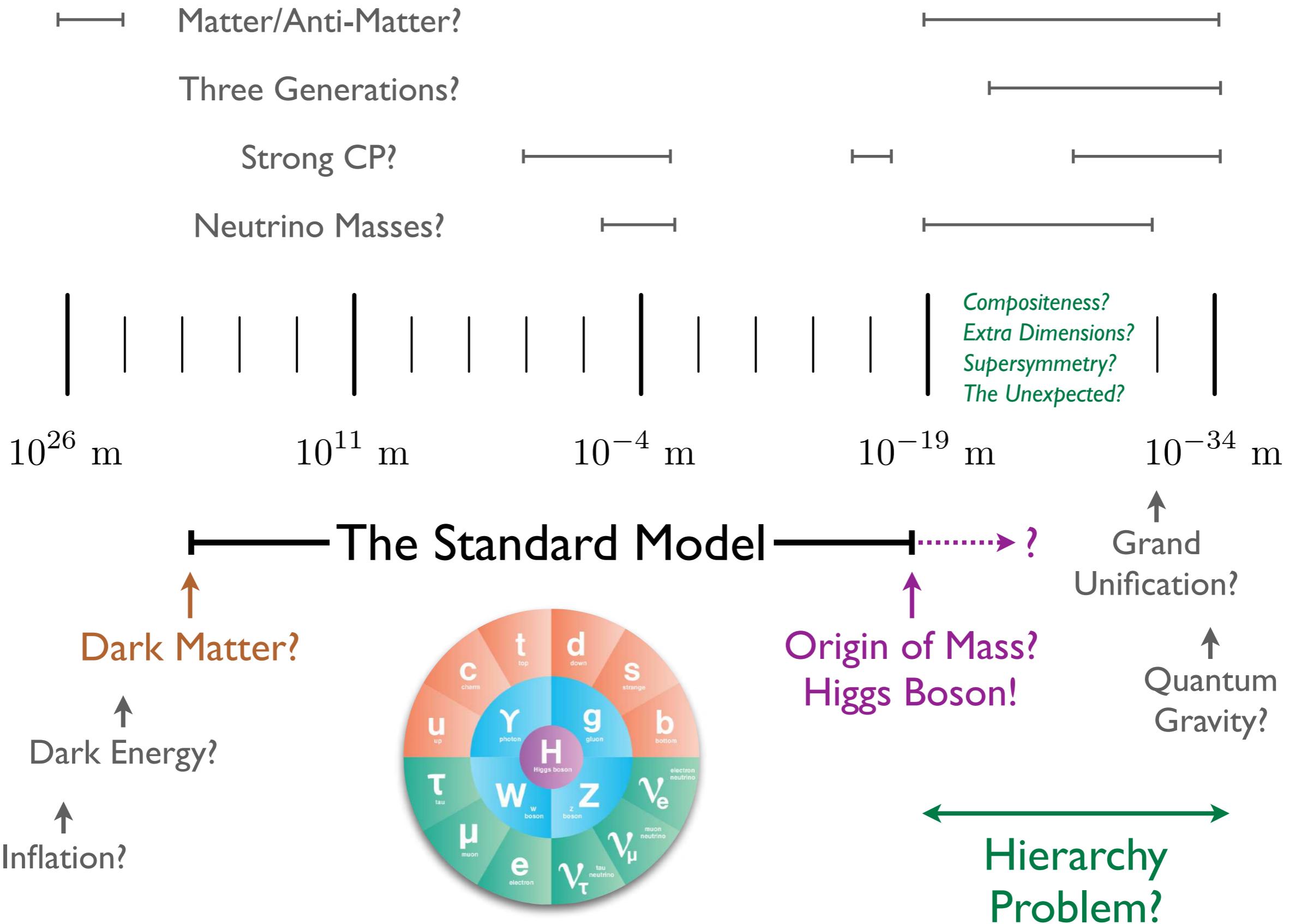
[e.g. Almeida, Backović, Cliche, Lee, Perelstein, [1501.05968](#);
 review in Kasieczka, Plehn, et al., [1902.09914](#)]



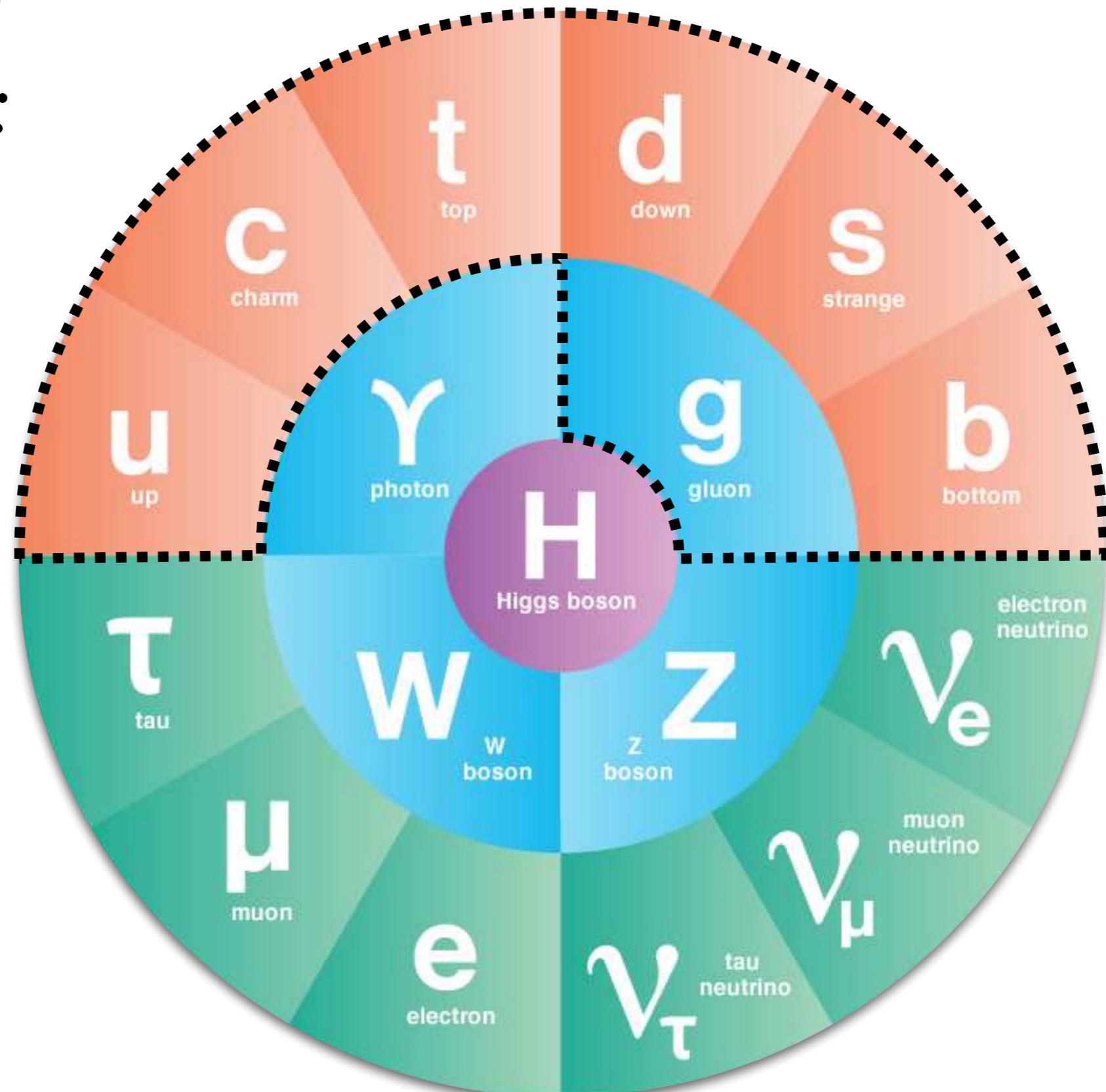
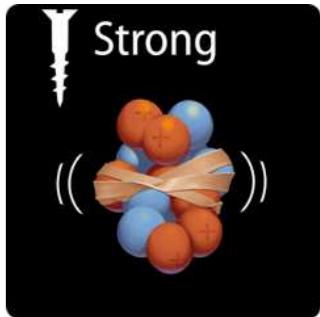


The Standard Model



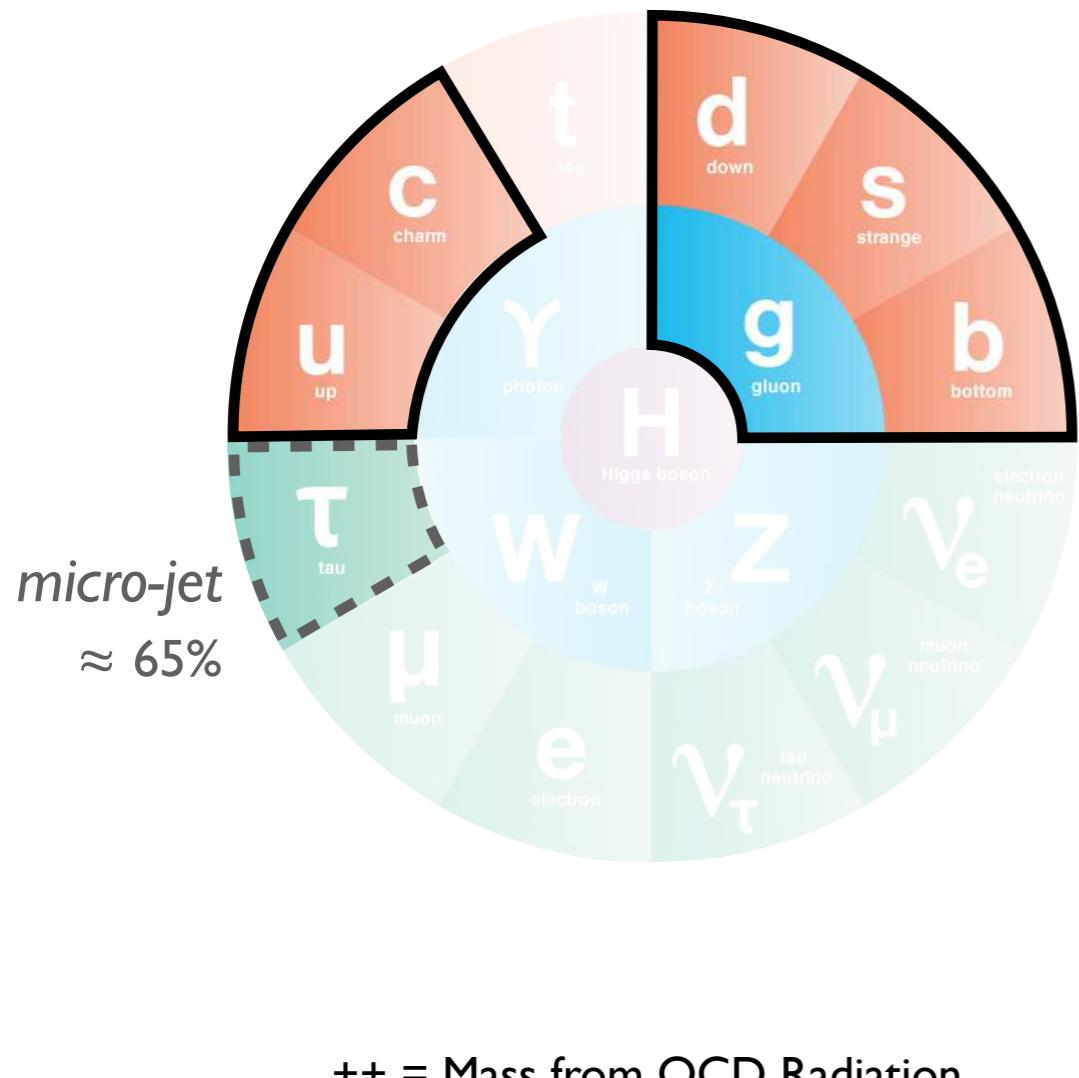


Focus of this talk:

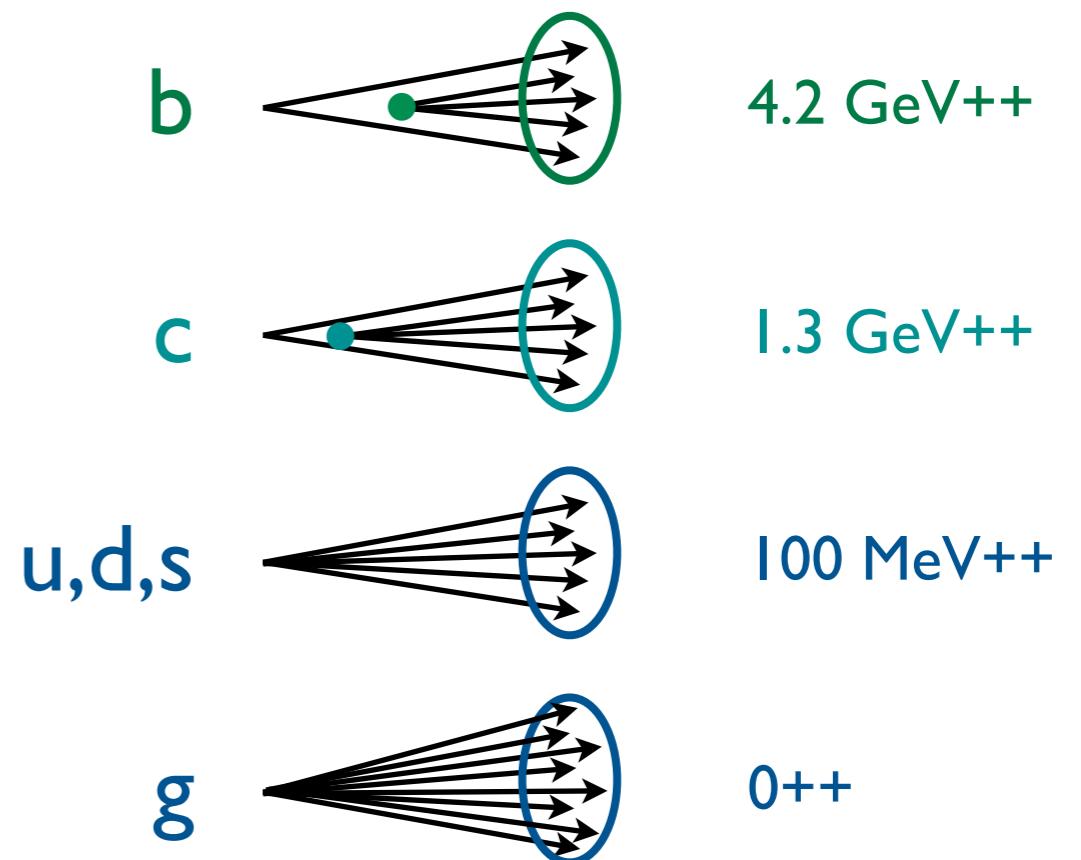


Jet Classification

Key supervised learning task at LHC

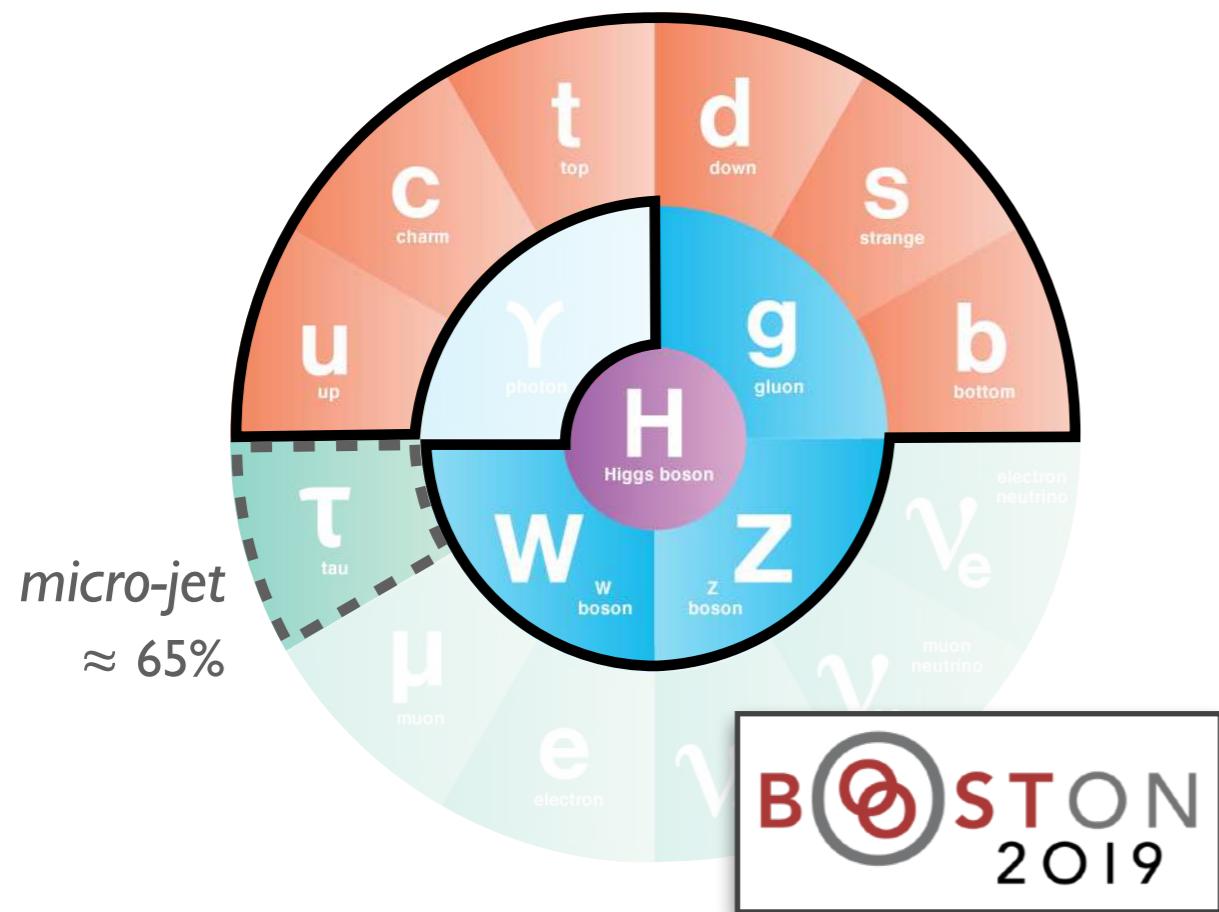


[see reviews in Larkoski, Moult, Nachman, [arXiv 2017](#);
Asquith, Delitzsch, Schmidt, et al., [arXiv 2018](#);
Marzani, Soyez, Spannowsky, [LNP 2019](#)]



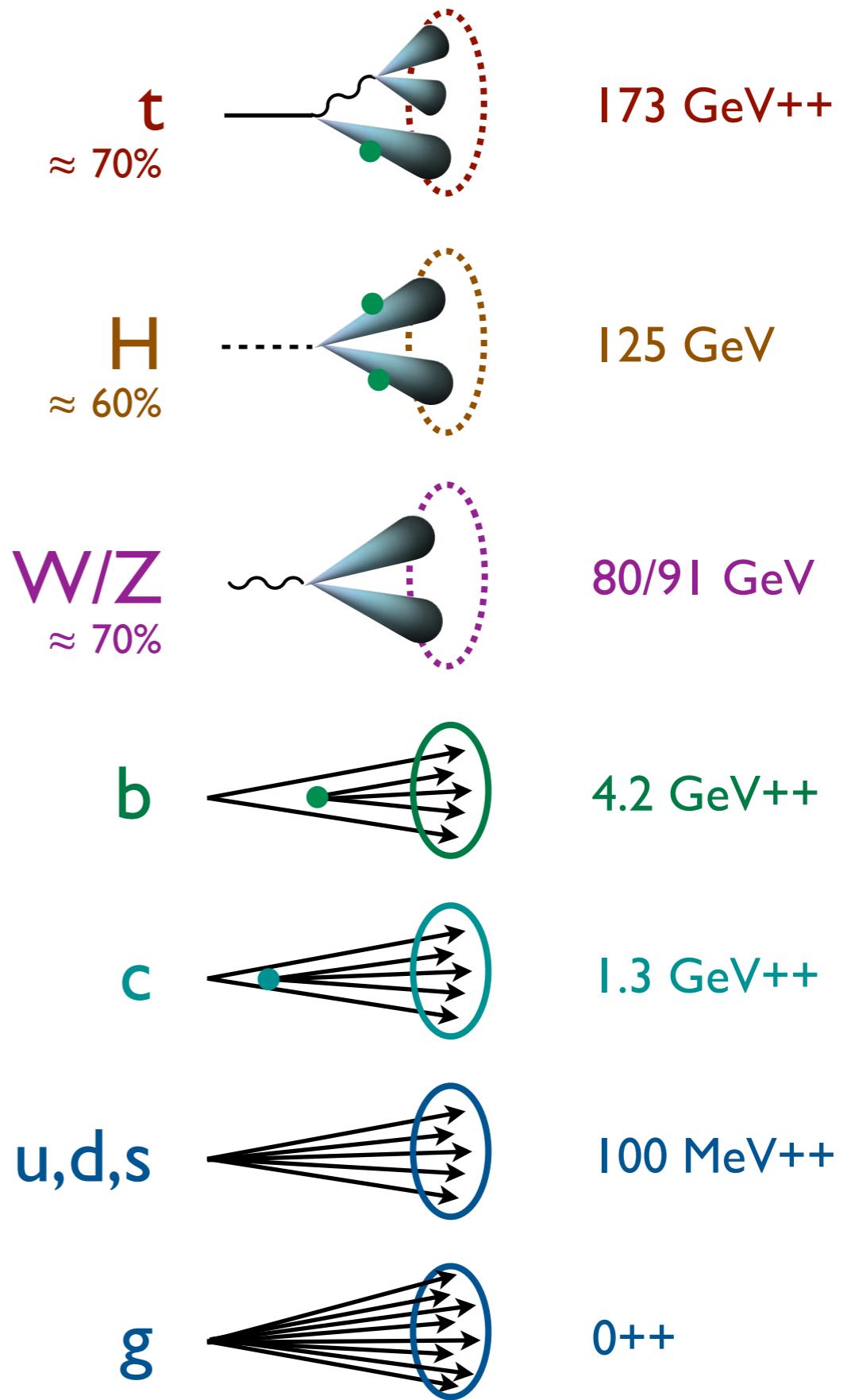
Jet Classification

Key supervised learning task at LHC



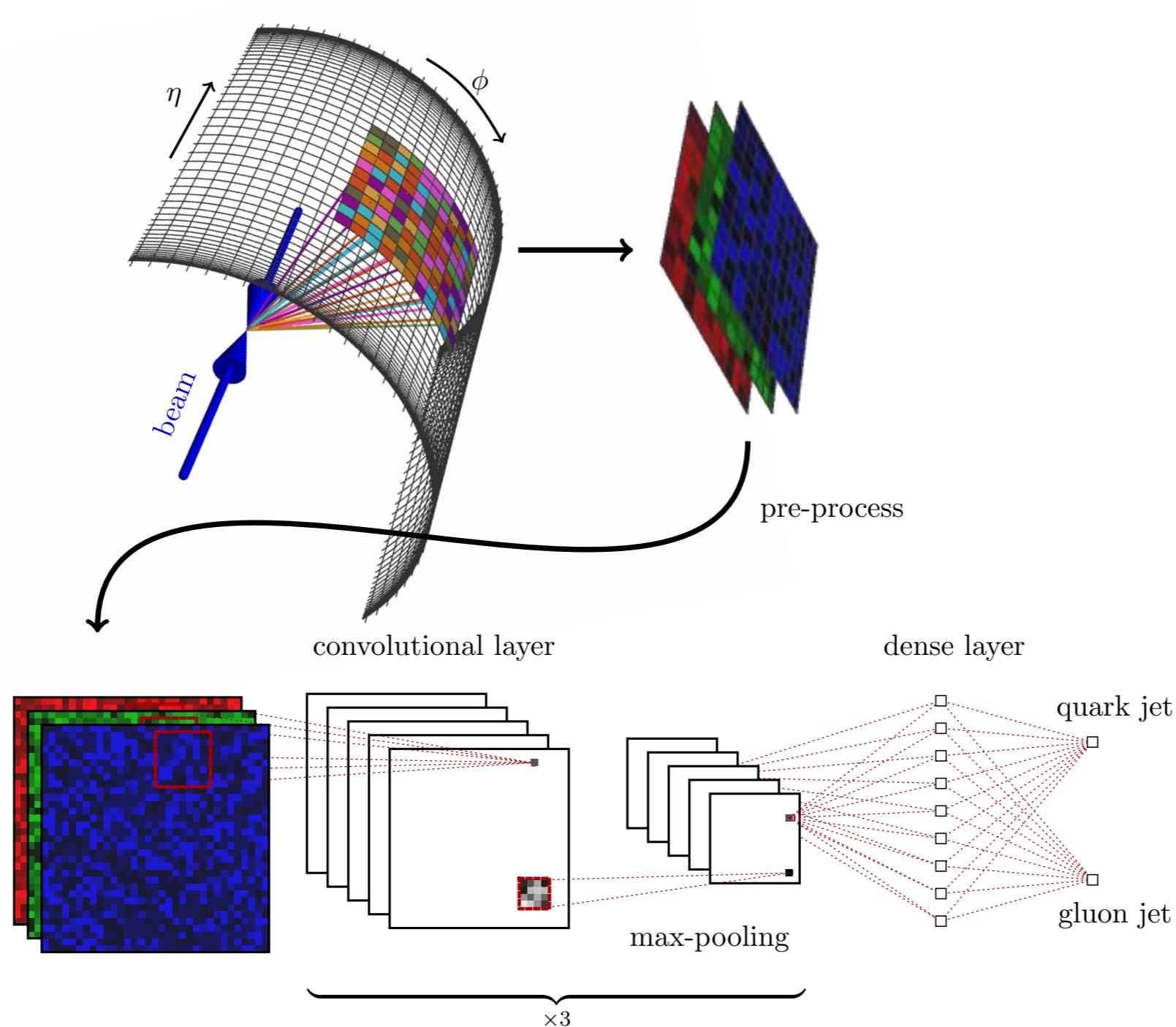
$++$ = Mass from QCD Radiation

[see reviews in Larkoski, Moult, Nachman, [arXiv 2017](#);
Asquith, Delitzsch, Schmidt, et al., [arXiv 2018](#);
Marzani, Soyez, Spannowsky, [LNP 2019](#)]



Machine Learning for Jets

“ML4Jets III”
NYU, January 2020



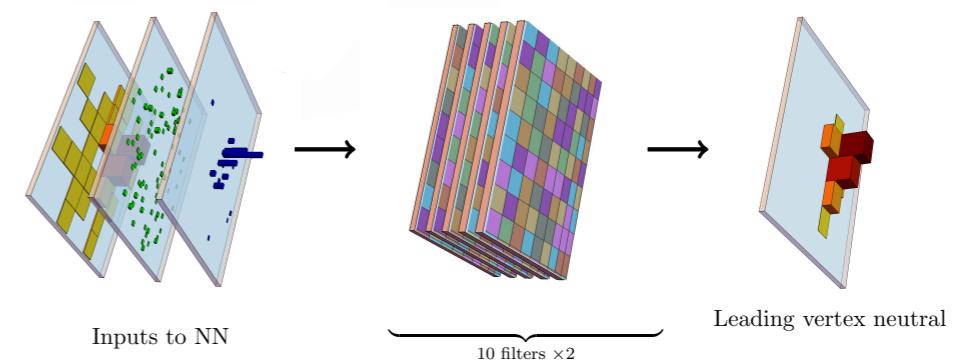
[figure from Komiske, Metodiev, Schwartz, [JHEP 2017](#);
reviews in Larkoski, Moult, Nachman, [arXiv 2017](#); Guest, Cranmer, Whiteson, [ARNPS 2018](#)]

Other Examples of Supervised Learning

Regression

e.g. *PUMML for pileup mitigation*

[Komiske, Metodiev, Nachman, Schwartz, [1707.08600](#);
see also Arjona Martínez, Cerri, Pierini, Spiropulu, Vlimant, [1810.07988](#)]

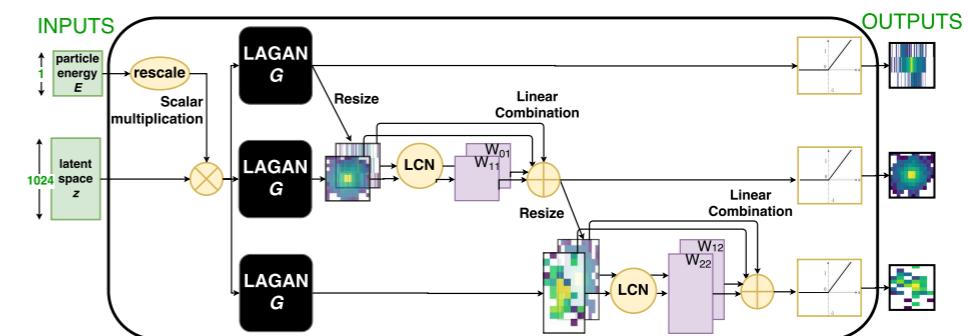


Labeled data: Objects J with property x
Solution: Map from J to x

Generation

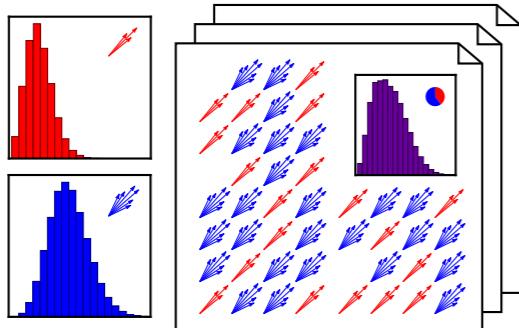
e.g. *CaloGAN for fast detector simulation*

[Paganini, de Oliveira, Nachman, [1705.02355](#), [1712.10321](#);
see also de Oliveira, Michela Paganini, Nachman, [1701.05927](#)]



Labeled data: Objects J with property x
Solution: Map (conditioned on x)
from noise to J

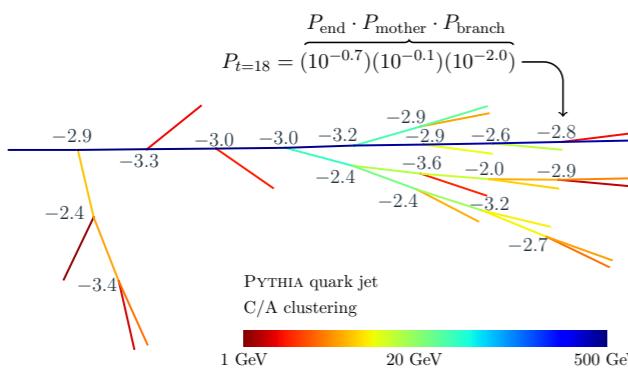
The Rise of Unsupervised Learning



Jet Topics

Blind Source Separation

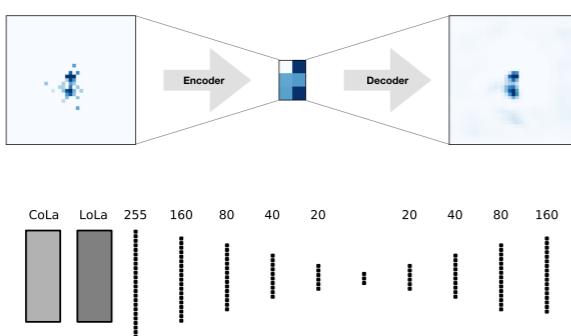
[Metodiev, JDT, [I802.00008](#); Komiske, Metodiev, JDT, [I809.01140](#); see also Metodiev, Nachman, JDT, [I708.02949](#)]



JUNIPR

Probability Modeling

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



Autoencoders

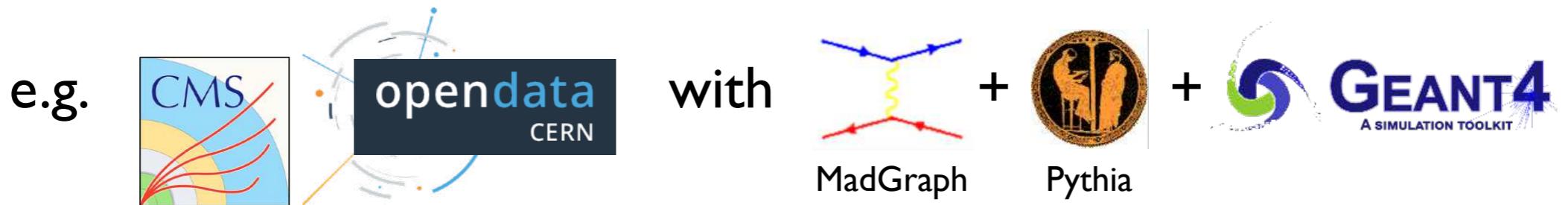
Anomaly Detection

[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](#)]

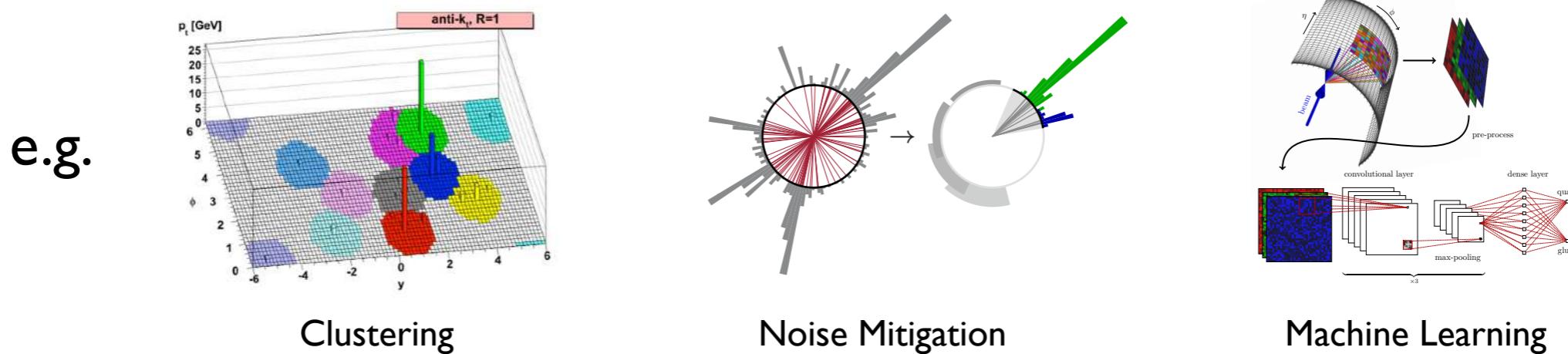
Common theme: Analyze *event ensembles*, not individual events

Particle Physics as Machine-Learning Testbed

- Huge datasets with reliable simulations



- Broad use of (un)supervised algorithms

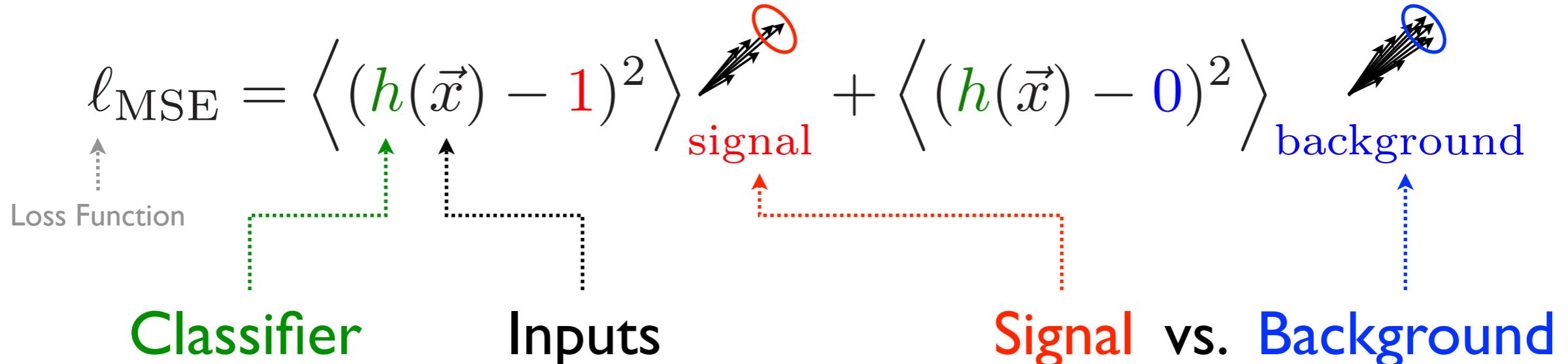
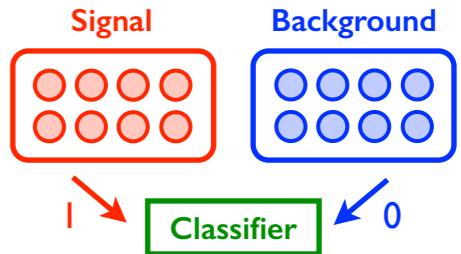


- Extensive domain knowledge and strong theory foundations

[figures from Cacciari, Salam, Soyez, [JHEP 2008](#); Larkoski, Marzani, JDT, Tripathee, Xue, [PRL 2017](#); Komiske, Metodiev, Schwartz, [JHEP 2017](#)]

Jet Classification Studies

Mix and match



Classifier

- Boosted Decision Tree
- Fisher Linear Discriminant
- Shallow Neural Network
- Deep Neural Network
- Convolutional Neural Network
- Recurrent Neural Network
- Recursive Neural Network
- Combination/Lorentz Layers
- ...

Inputs

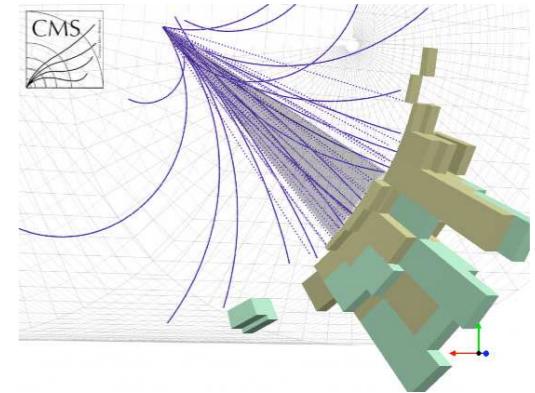
- High-Level Features
- Basis of High-Level Features
- Jet Image
- Multi-channel Jet Image
- Abstract Jet Image
- Sorted Four-Vectors
- Clustered Four-Vectors
- Lund Plane Emissions
- Kitchen Sink
- ...

Signal vs. Background

Quark Jets	vs.	Gluon Jets
Up-type Quarks	vs.	Down-type Quarks
W/Z Bosons	vs.	QCD Jets
W Bosons	vs.	Z Bosons
Top Quarks	vs.	QCD Jets
Exotic Boosted Objects	vs.	QCD Jets
CMS Open Data Samples	vs.	Each other
...	vs.	...

[Lönnblad, Peterson, Rögnvaldsson, [PRL 1990](#), ..., Cogan, Kagan, Strauss, Schwartzman, [1407.5675](#); Almeida, Backović, Cliche, Lee, Perelstein, [1501.05968](#); de Oliveira, Kagan, Mackey, Nachman, Schwartzman, [1511.05190](#); Baldi, Bauer, Eng, Sadowski, Whiteson, [1603.09349](#); Conway, Bhaskar, Erbacher, Pilot, [1606.06859](#); Guest, Collado, Baldi, Hsu, Urban, Whiteson, [1607.08633](#); Barnard, Dawe, Dolan, Rajcic, [1609.00607](#); Komiske, Metodiev, Schwartz, [1612.01551](#); Kasieczka, Plehn, Russell, Schell, [1701.08784](#); Louppe, Cho, Becot, Cranmer, [1702.00748](#); Pearkes, Fedorko, Lister, Gay, [1704.02124](#); Datta, Larkoski, [1704.08249](#), [1710.01305](#); Butter, Kasieczka, Plehn, Russell, [1707.08966](#); Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, [1708.07034](#); Aguilar Saavedra, Collin, Mishra, [1709.01087](#); Cheng, [1711.02633](#); Luo, Luo, Wang, Xu, Zhu, [1712.03634](#); Komiske, Metodiev, JDT, [1712.07124](#); Macaluso, Shih, [1803.00107](#); Fraser, Schwartz, [1803.08066](#); Choi, Lee, Perelstein, [1806.01263](#); Lim, Nojiri, [1807.03312](#); Dreyer, Salam, Soyez, [1807.04758](#); Moore, Nordström, Varma, Fairbairn, [1807.04769](#); plus many ATLAS/CMS performance studies; plus my friends who will scold me for forgetting their paper (and not updating this after July 23, 2018)]

Key: Jets are Point Clouds



- **Particle:** List of properties

$$\vec{p} = \{E, p_x, p_y, p_z, \dots\}$$

↑ ↑
Energy Momentum Mass, charge, flavor, vertex, quality, ...

- **Jet:** Set of particles

$$\mathcal{J} = \{\vec{p}_1, \vec{p}_2, \vec{p}_3, \dots, \vec{p}_N\}$$

—————
Permutation Symmetry **Variable Length**

Quantum Mechanics:
“When you’ve seen one electron, you’ve seen them all”

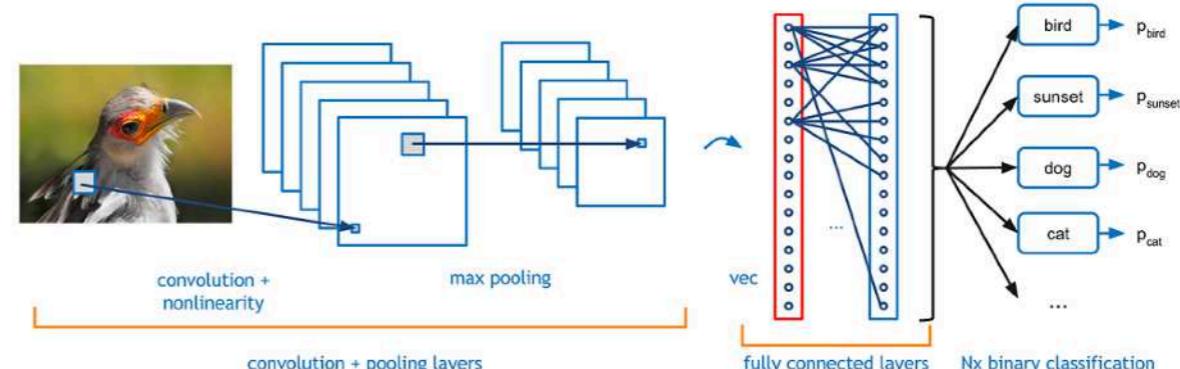
Jet Formation:
Typically 10–50 particles per jet

- **Dataset:** Set of jets

Off-the-Shelf Machine Learning?

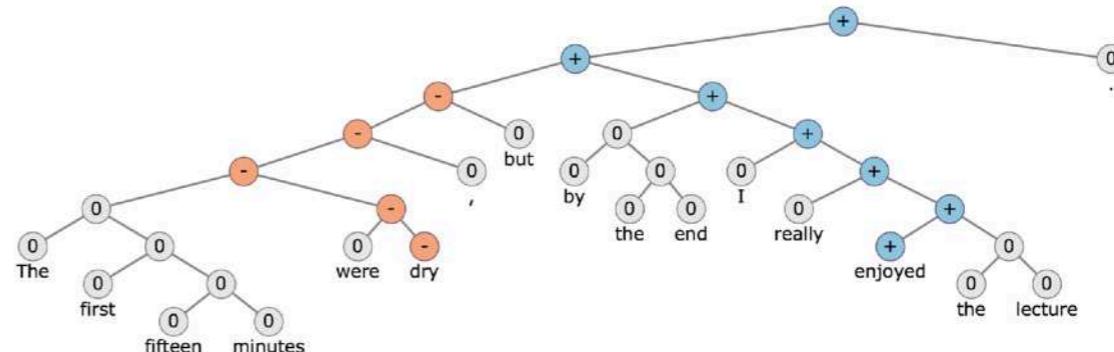
2D Images?

Appropriate for fixed-grid calorimeters,
but less ideal for tracking detectors



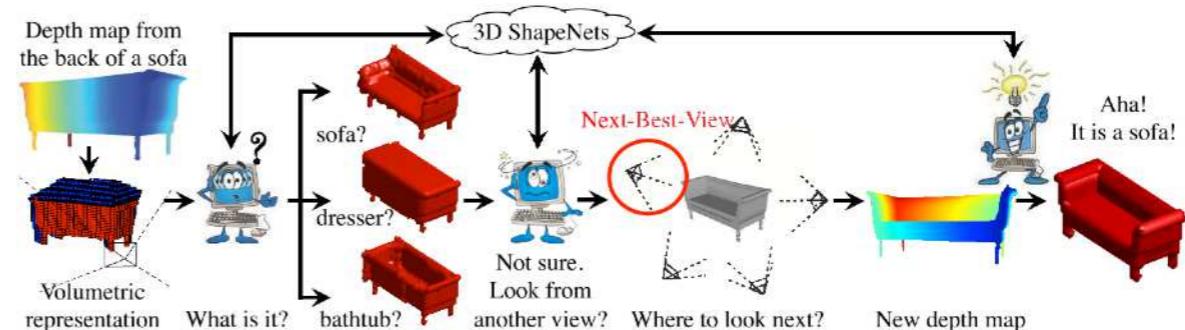
Natural Language?

Clustering can yield “semantic” structure
but inputs are fundamentally symmetric

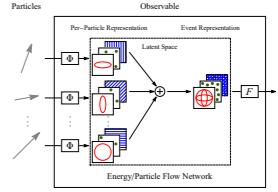


3D Objects?

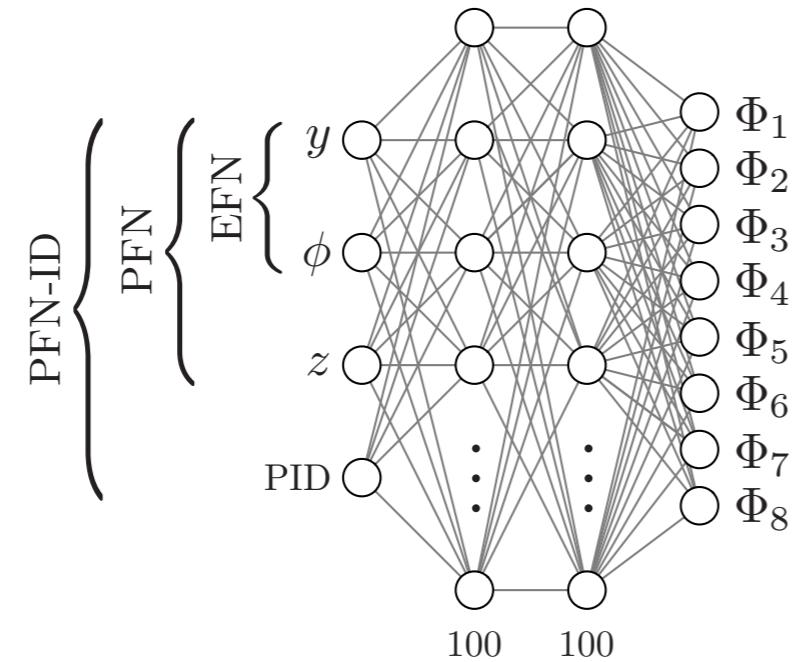
Much closer to particle physics,
but most architectures are not “safe”



Technical Implementation



Per-Particle Network: Φ

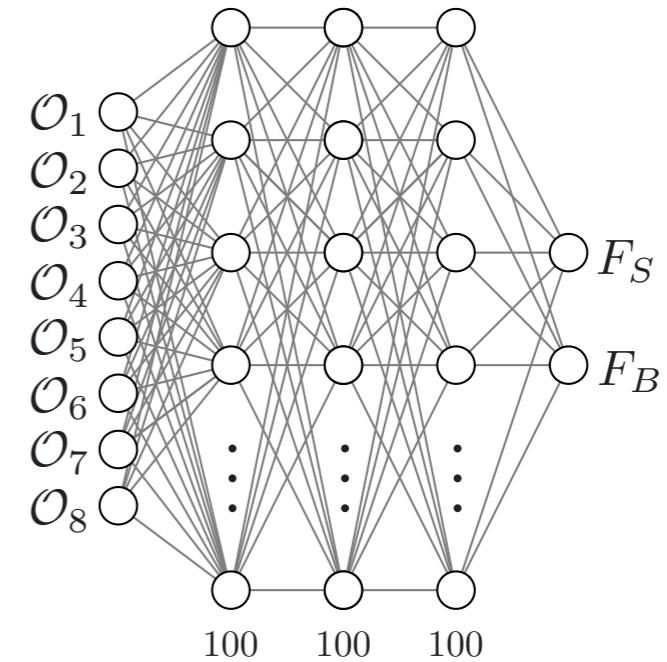


Per-Jet (Latent Space):

$$\text{EFN: } \mathcal{O}_a = \sum_{i \in \text{jet}} z_i \Phi_a(y_i, \phi_i) \quad z_i = \frac{p_{Ti}}{\sum_j p_{Tj}}$$

$$\text{PFN: } \mathcal{O}_a = \sum_{i \in \text{jet}} \Phi_a(y_i, \phi_i, z_i, \text{PID}_i)$$

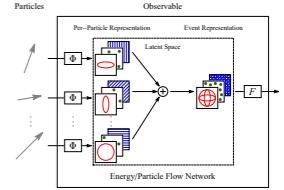
Latent Combiner: F



Final Discriminant:

$$\text{softmax}(F_S, F_B)$$

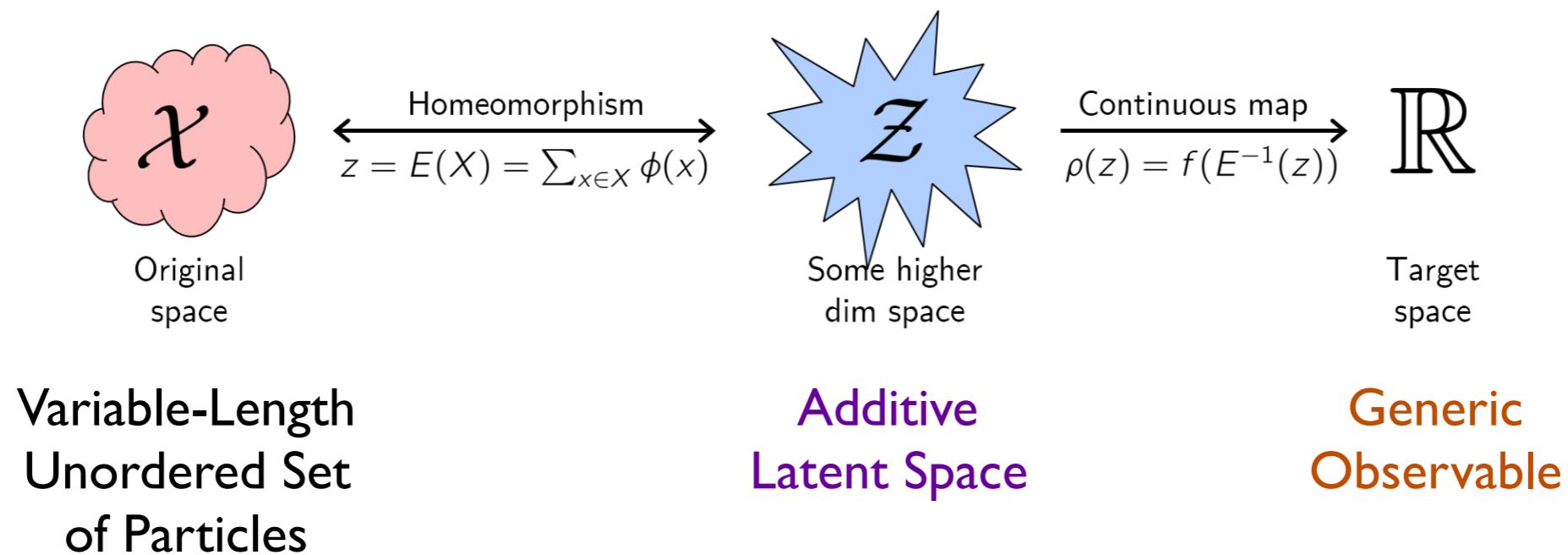
[Komiske, Metodiev, JDT, [JHEP 2019](#)]



Meanwhile in ML-Land: Deep Sets

Theorem 2 A function $f(X)$ operating on a set X having elements from a countable universe, is a valid set function, i.e., **invariant** to the permutation of instances in X , iff it can be decomposed in the form $\rho(\sum_{x \in X} \phi(x))$, for suitable transformations ϕ and ρ .

↑
(!)



[Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, [1703.06114](#); see also Vinyals, Bengio, Kudlur, [1511.06391](#); Qi, Su, Mo, Guibas, [1612.00593](#)]

Deep Sets for...

Celebrity Face Anomaly Detection

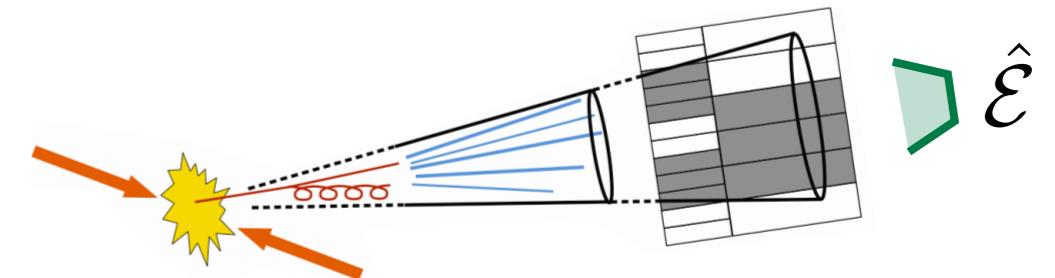
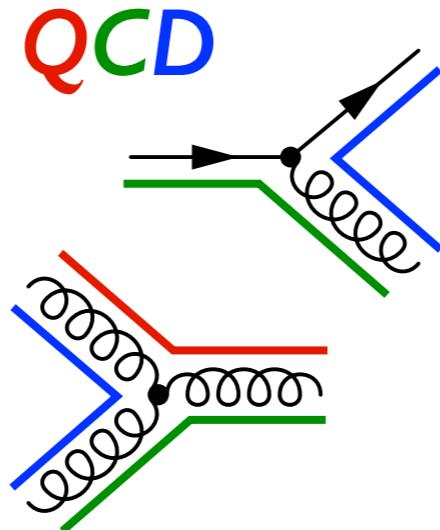
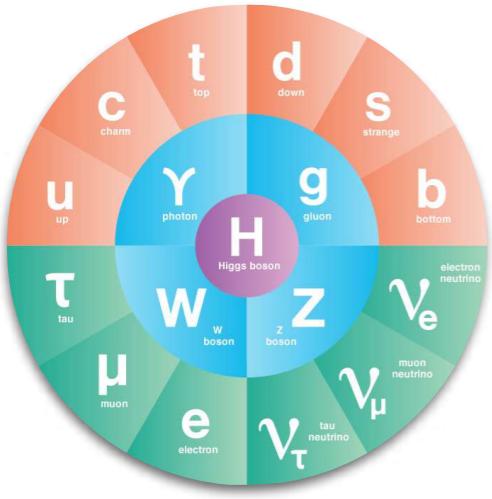


Point Cloud Classification

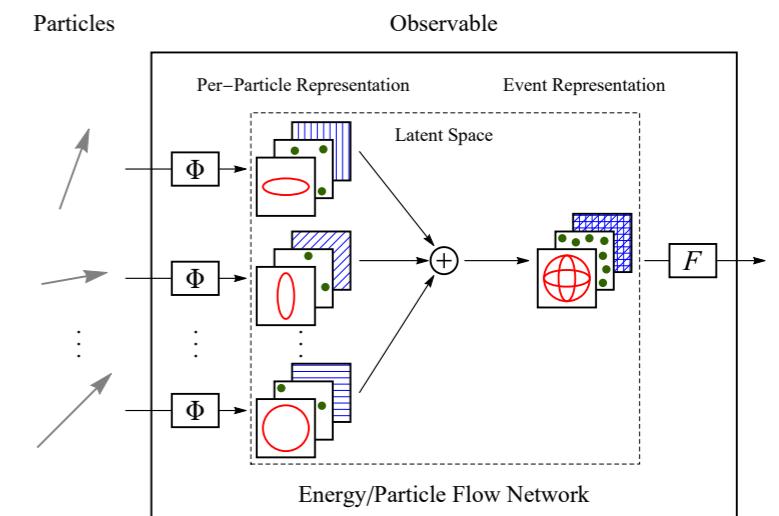
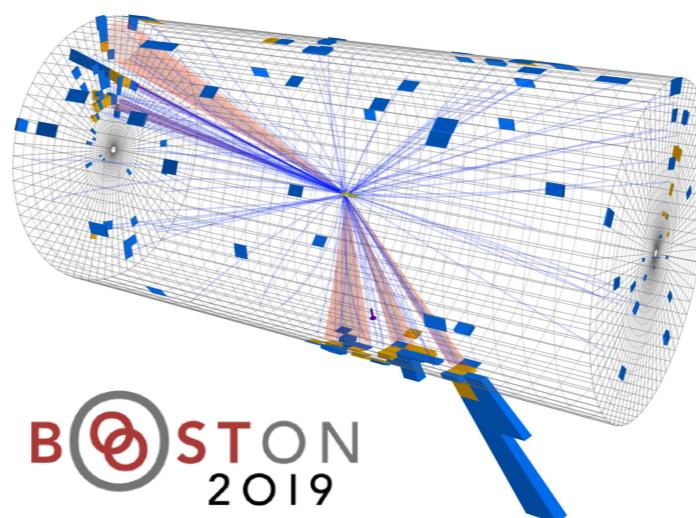
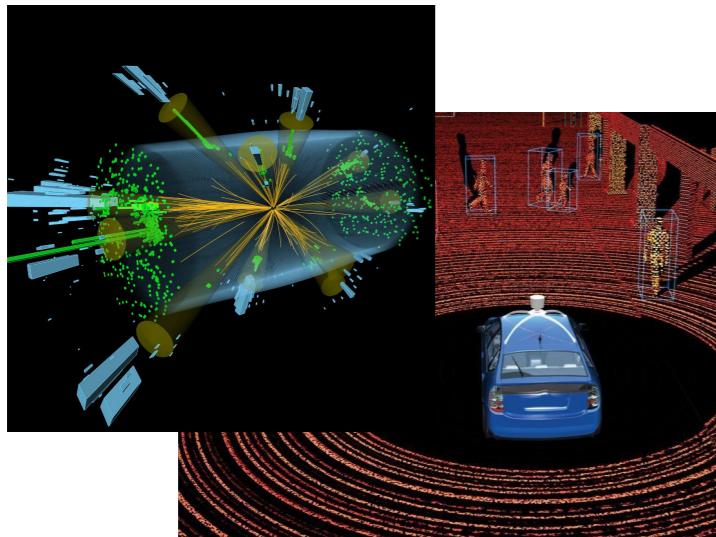


[Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, [1703.06114](#)]

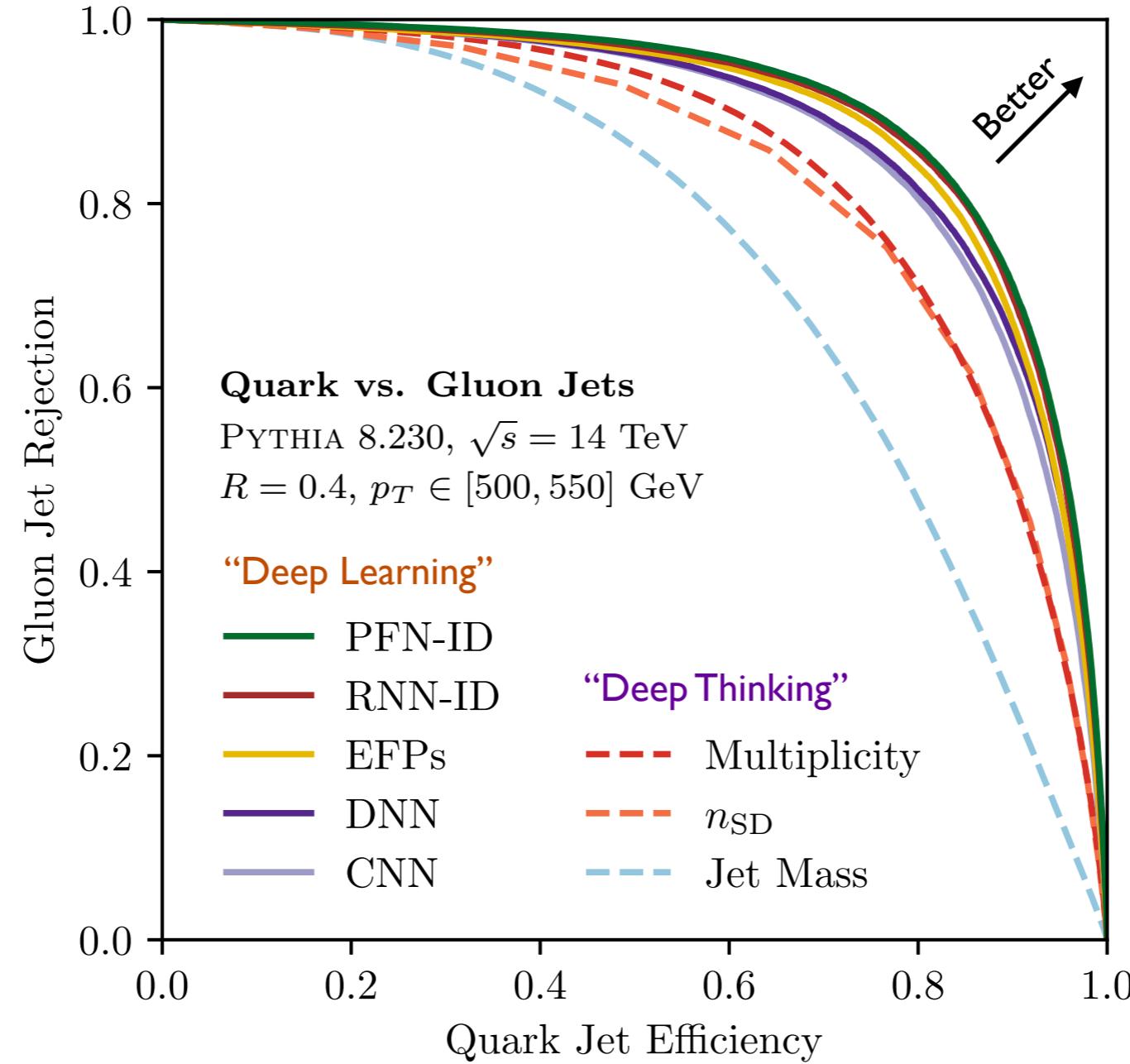
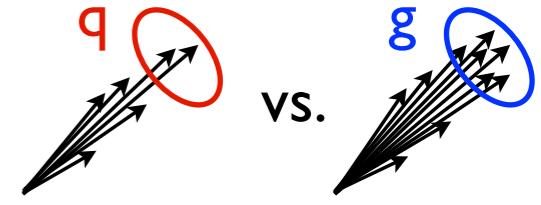
Deep Sets for Particle Jets



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



Discrimination Performance

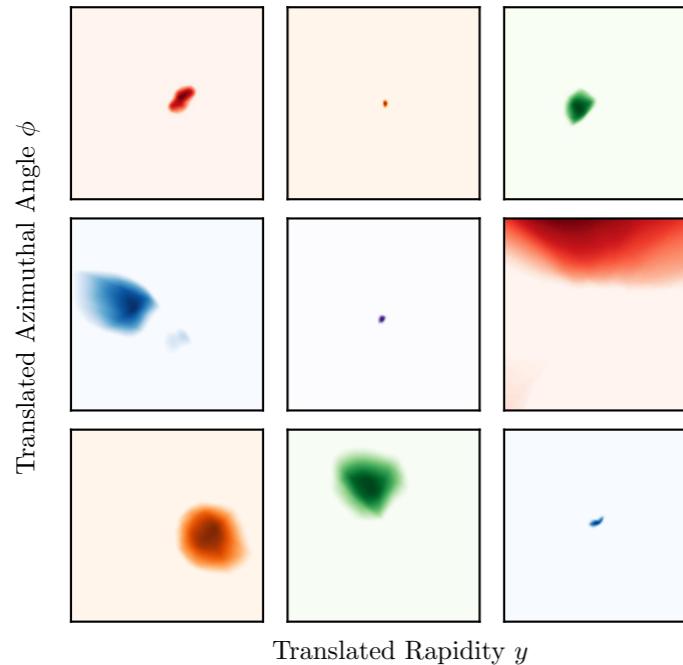


[Komiske, Metodiev, JDT, JHEP 2019]

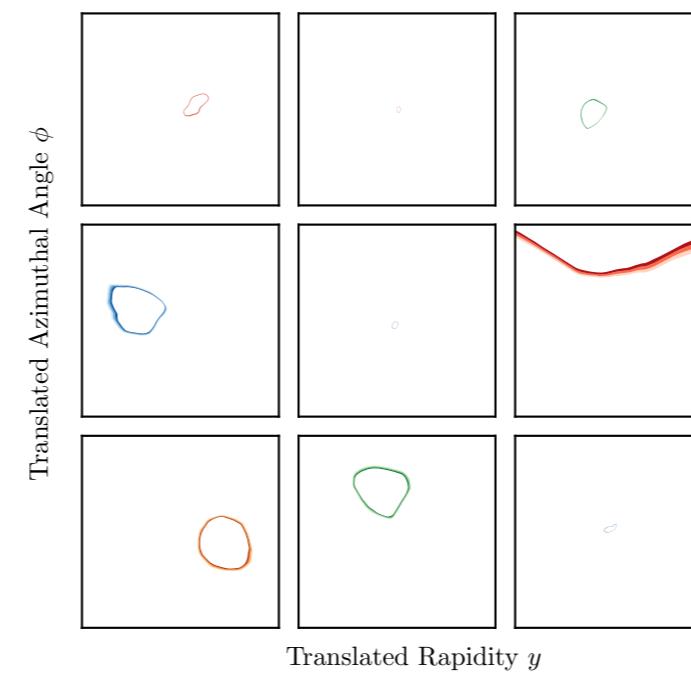
Psychedelic Network Visualization

(go back)

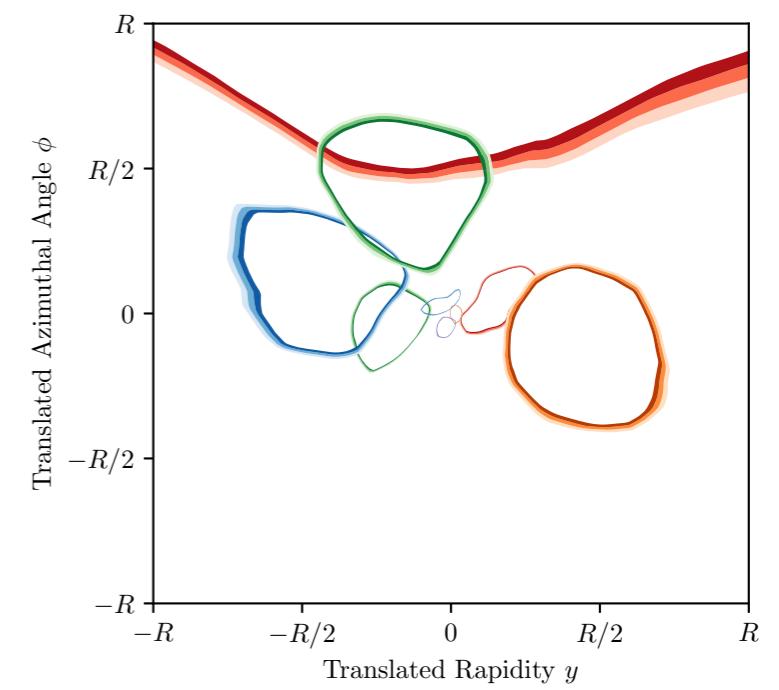
Latent Filters



50% Contours

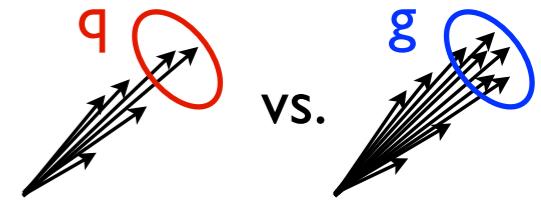


Overlay

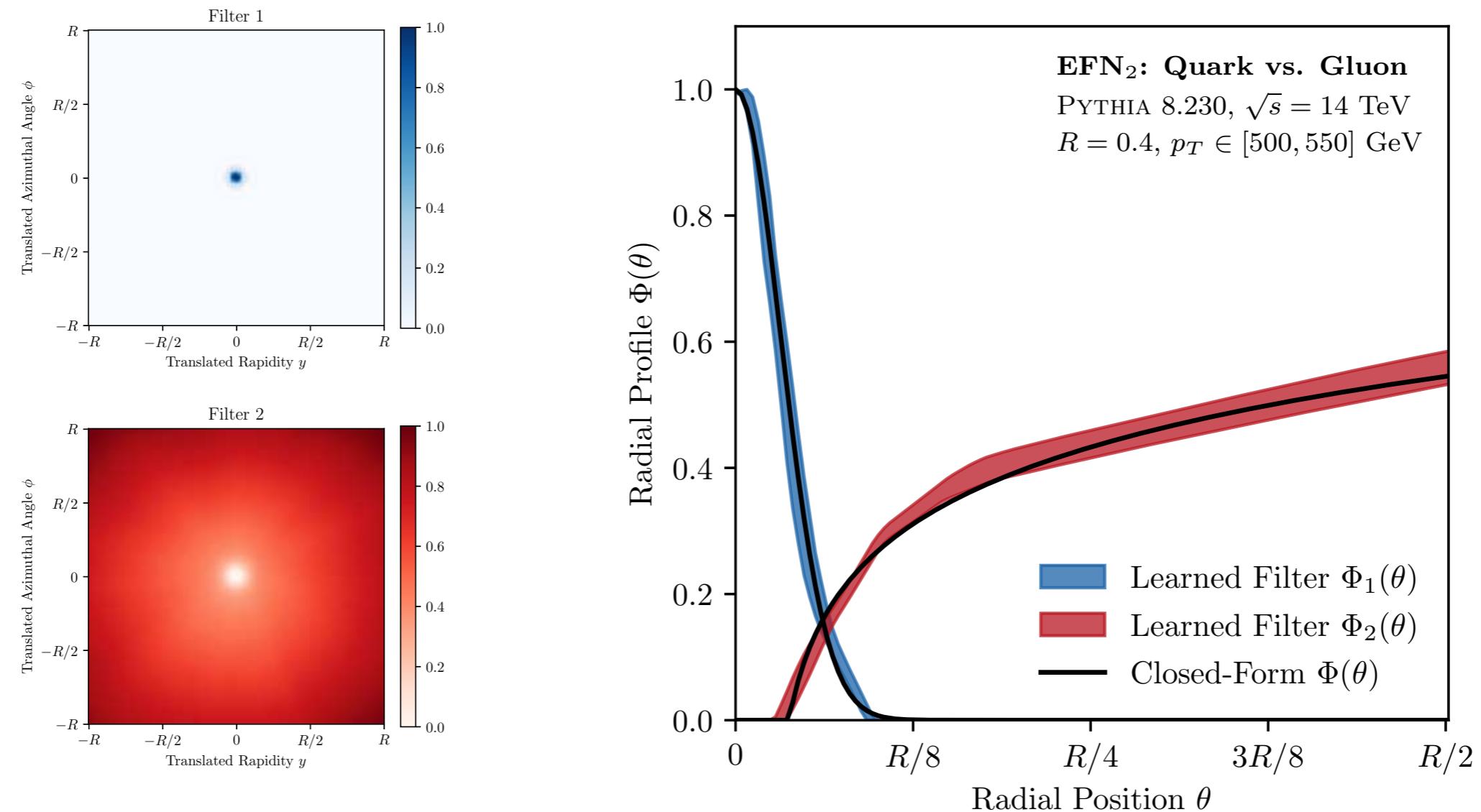


*“Ok, but did you really learn something
you didn’t already know?”*

Learning from the Machine



For $\ell = 2$ EFN, radial moments: $\sum_{i \in \text{jet}} z_i f(\theta_i)$ cf. Angularities:
 $f(\theta) = \theta^\beta$



[Komiske, Metodiev, JDT, JHEP 2019;
cf. Larkoski, JDT, Waalewijn, JHEP 2014; using Berger, Kucs, Sterman, PRD 2003; Ellis, Vermilion, Walsh, Hornig, Lee, JHEP 2010]

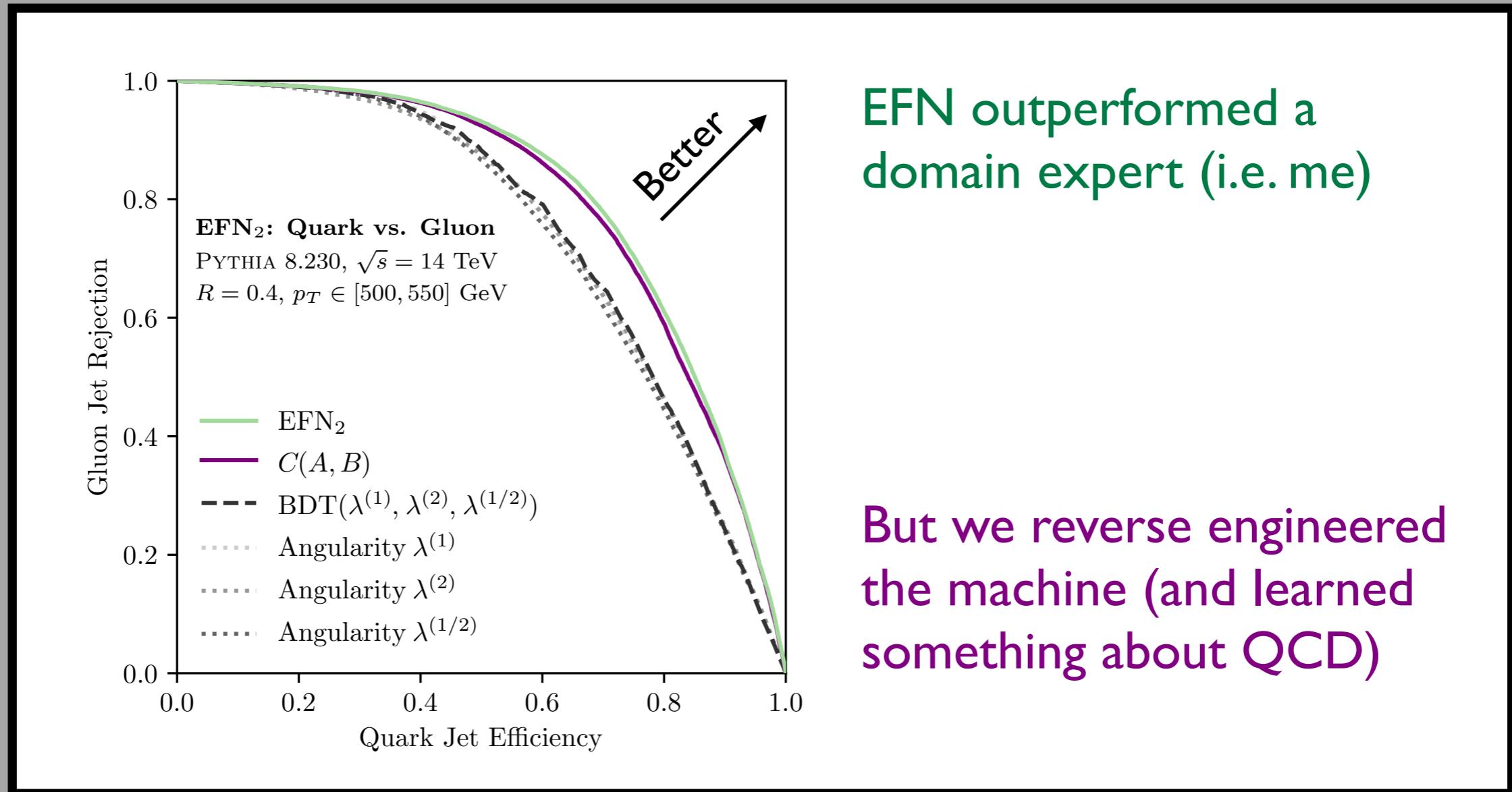
Learning from the Machine



For $\ell = 2$ EFN, radial moments:

$$\sum_{i \in \text{jet}} z_i f(\theta_i)$$

cf. Angularities:
 $f(\theta) = \theta^\beta$



[Komiske, Metodiev, JDT, [JHEP 2019](#);
cf. Larkoski, JDT, Waalewijn, [JHEP 2014](#); using Berger, Kucs, Sterman, [PRD 2003](#); Ellis, Vermilion, Walsh, Hornig, Lee, [JHEP 2010](#)]

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:

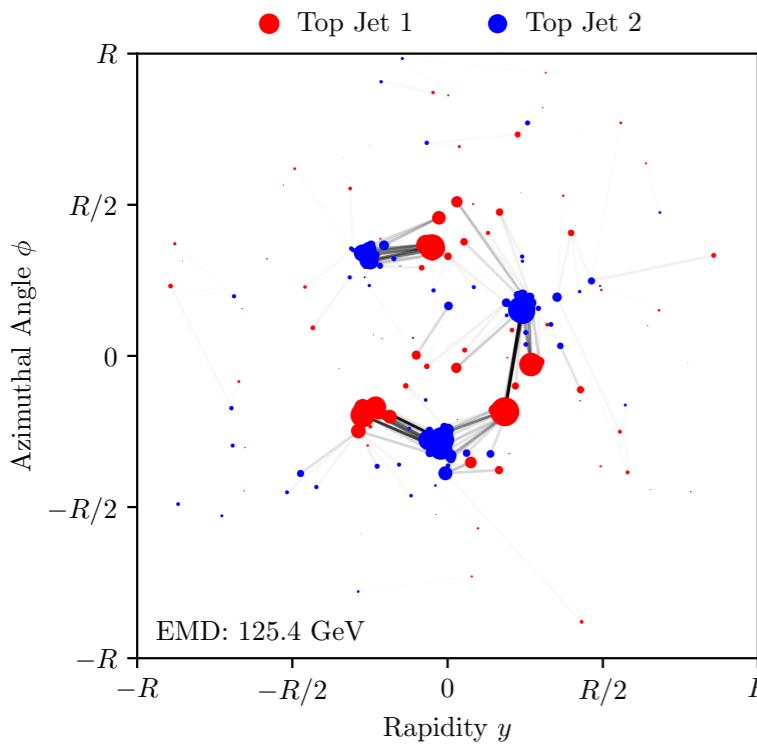
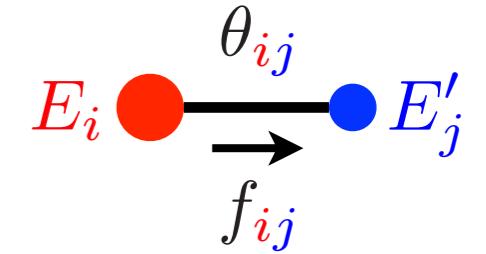
[Peleg, Werman, Rom, [IEEE 1989](#);
Rubner, Tomasi, Guibas, [ICCV 1998](#), [ICJV 2000](#);
Pele, Werman, [ECCV 2008](#); Pele Taskar, [GSI 2013](#)]

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



Equivalent to l -Wasserstein metric; very popular in ML applications

The Energy Mover's Distance

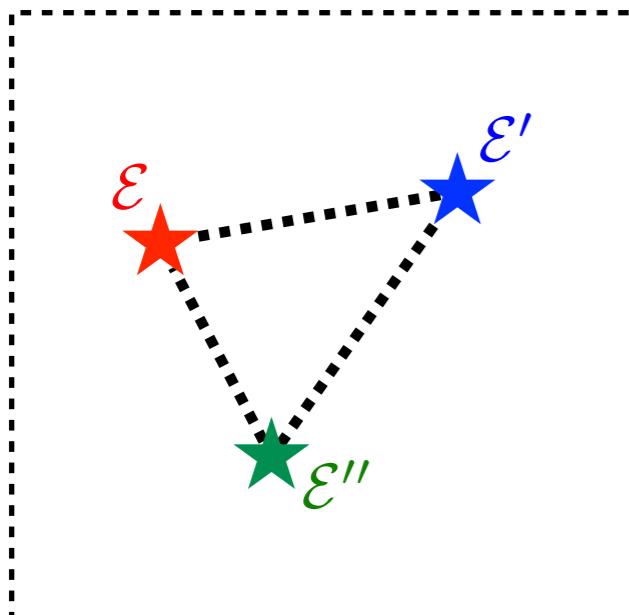


Optimal transport between energy flows...

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

↑
in GeV

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



...defines a metric on the space of events

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

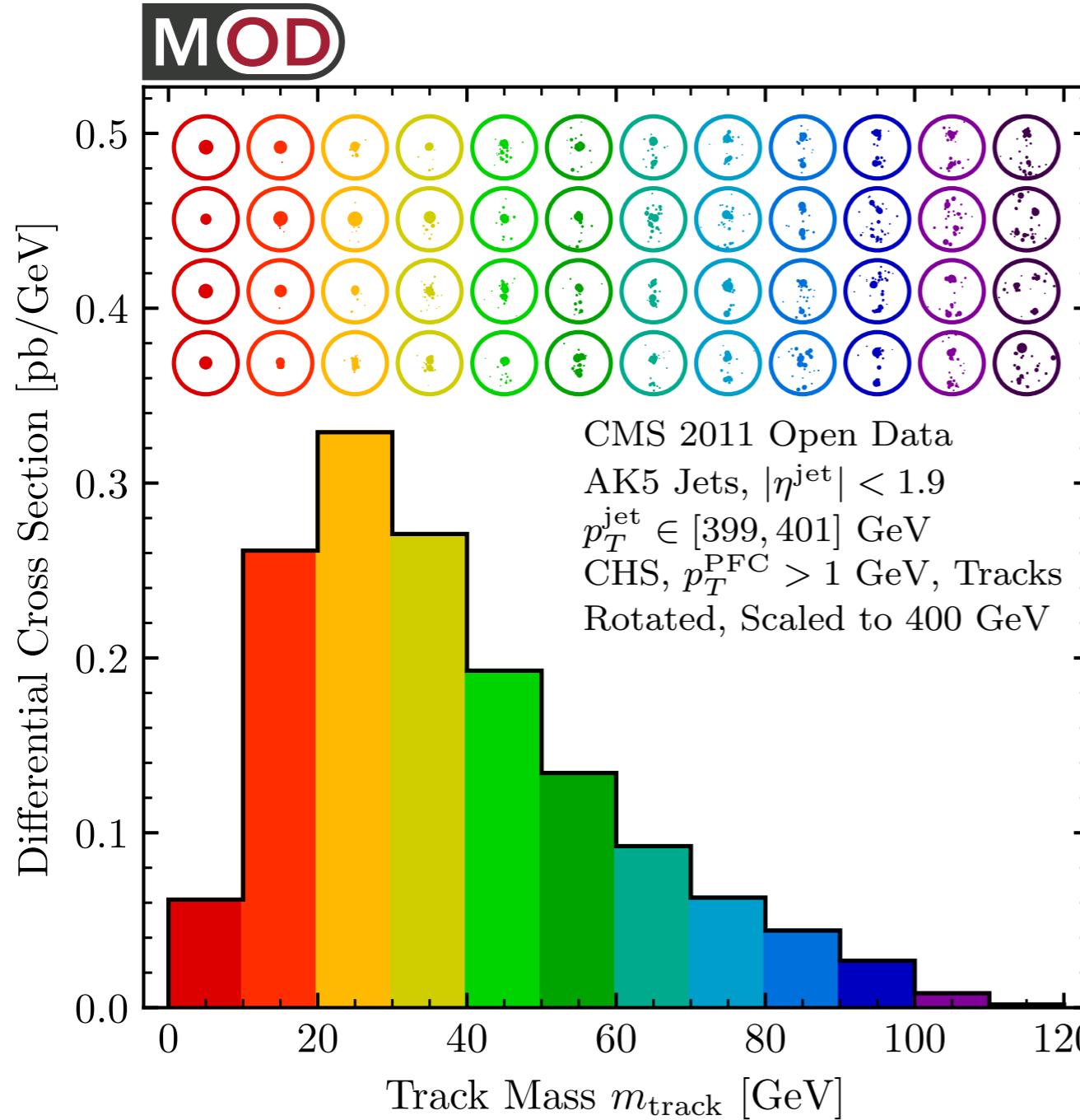
(assuming $R \geq \theta_{\max}/2$, i.e. $R \geq$ jet radius for conical jets)

[Komiske, Metodiev, JDT, PRL 2019;
see also Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

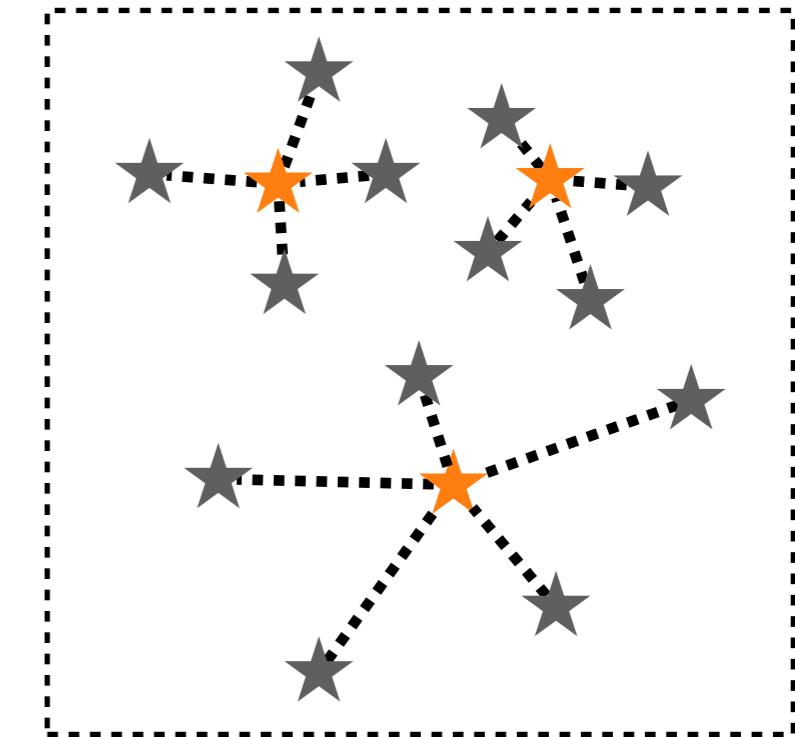
Most Representative Jets



[<http://opendata.cern.ch/>]



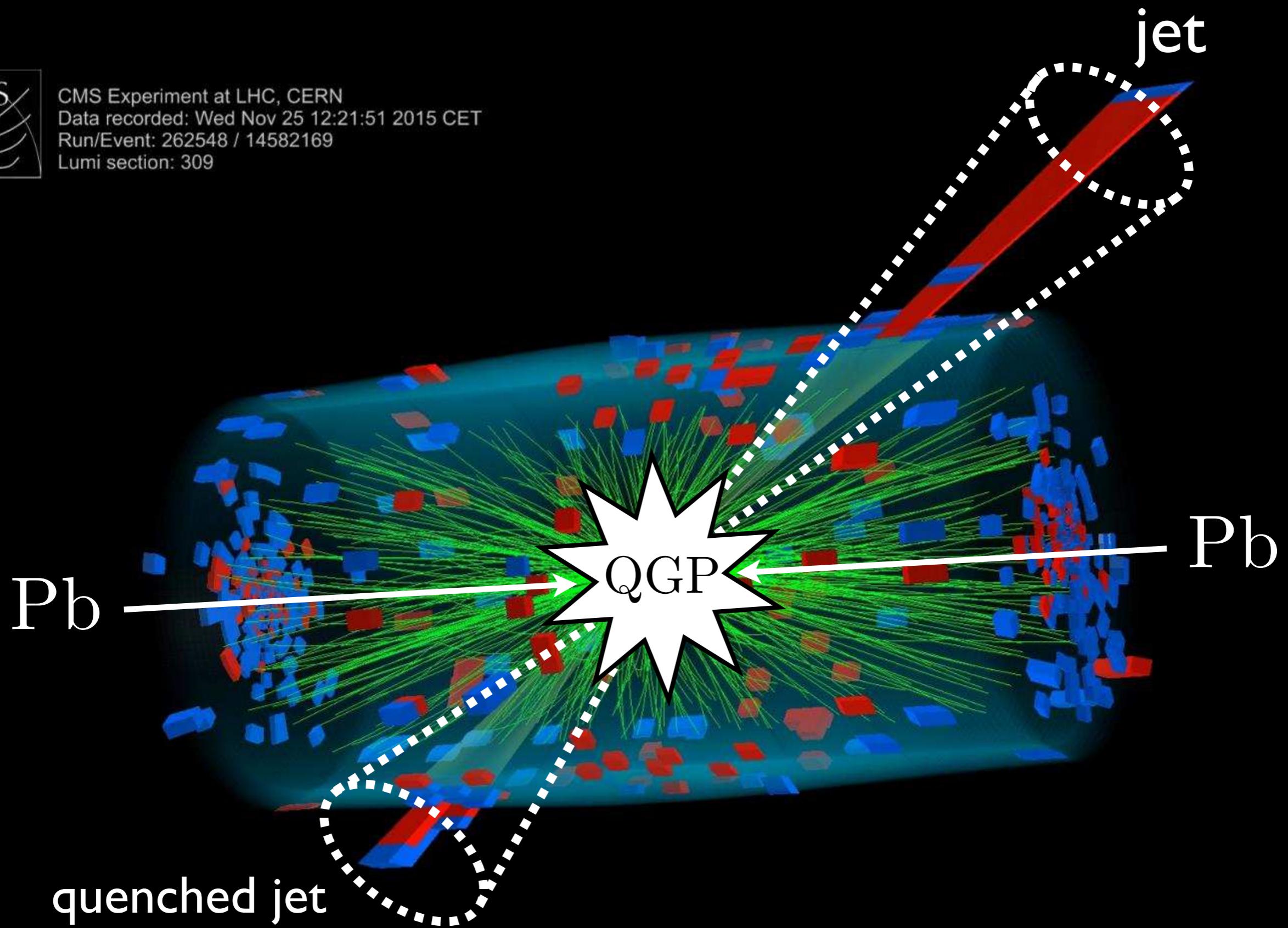
k-medoids
Per mass bin



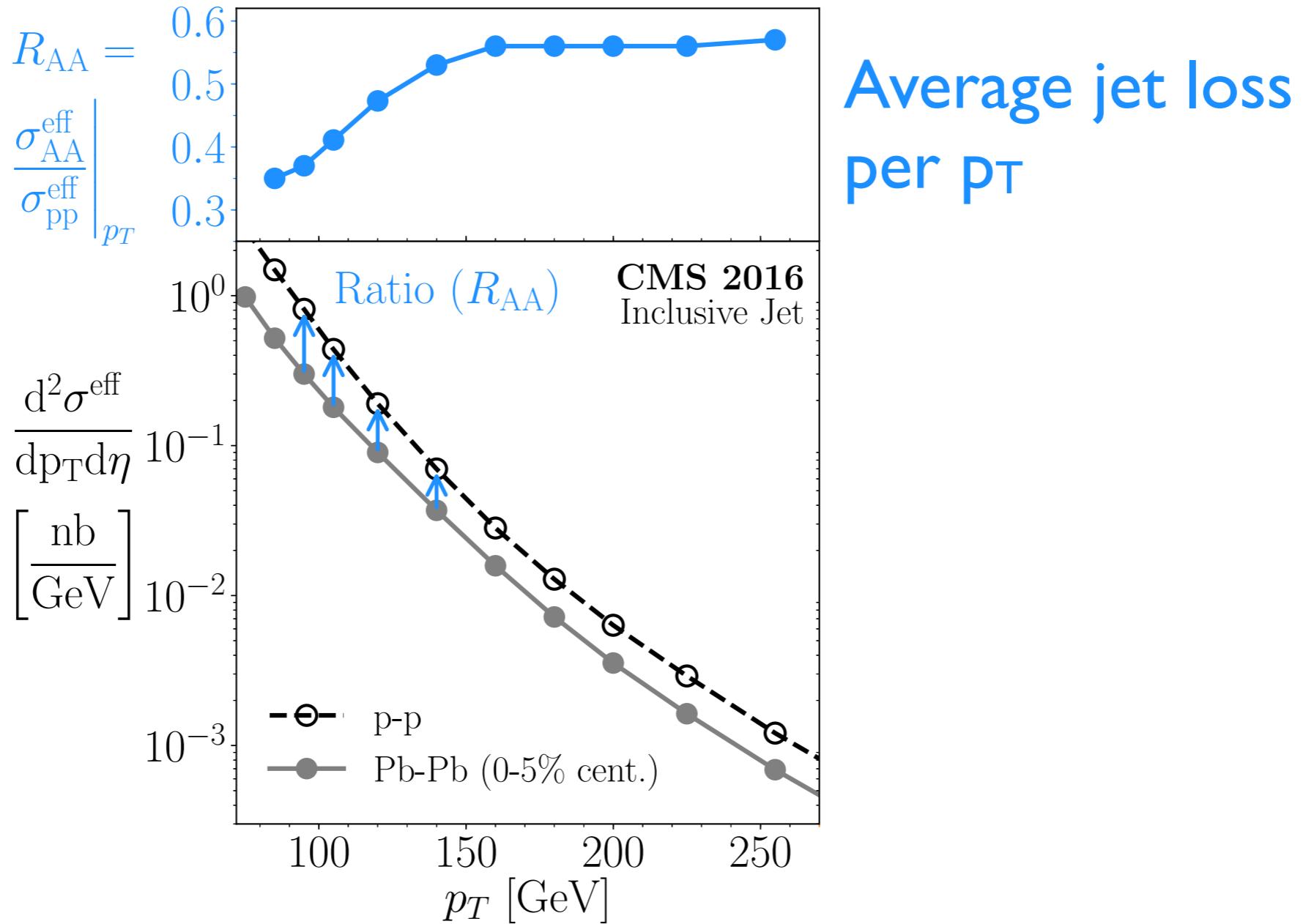
[Komiske, Mastandrea, Metodiev, Naik, JDT, arXiv 2019]



CMS Experiment at LHC, CERN
Data recorded: Wed Nov 25 12:21:51 2015 CET
Run/Event: 262548 / 14582169
Lumi section: 309



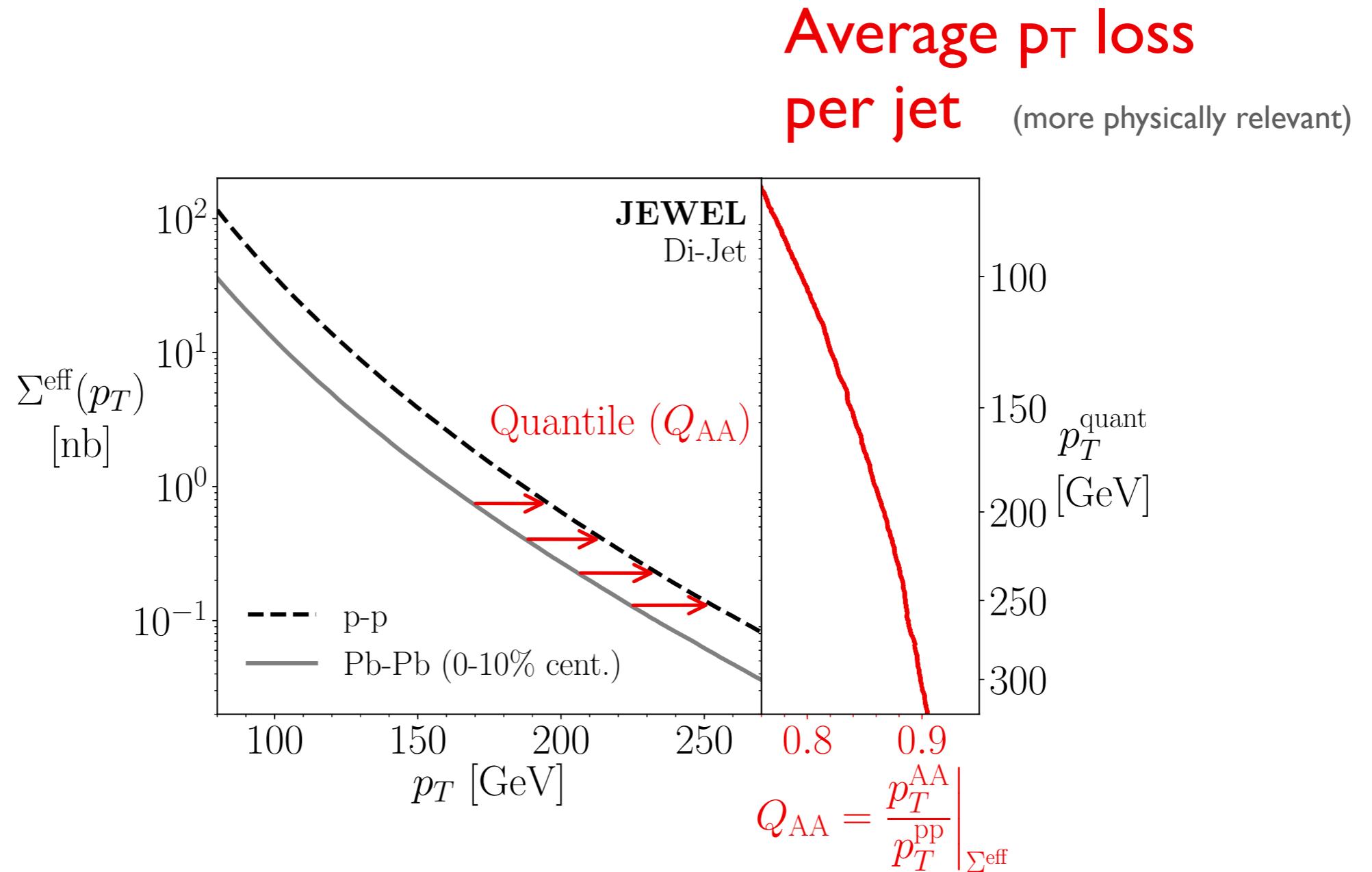
Jet Quenching via Cross Section Ratios



[CMS, PRC 2017]

Jet Quenching via Quantile Matching

Equivalent to one-dimensional optimal transport (!)



[Brewer, Milhano, JDT, PRL 2019]

The Rise of Public Collider Data

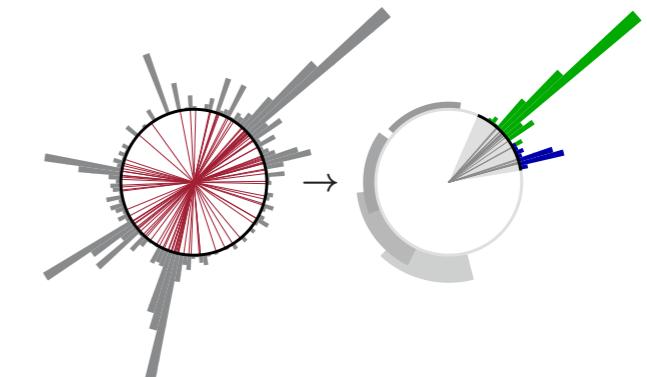
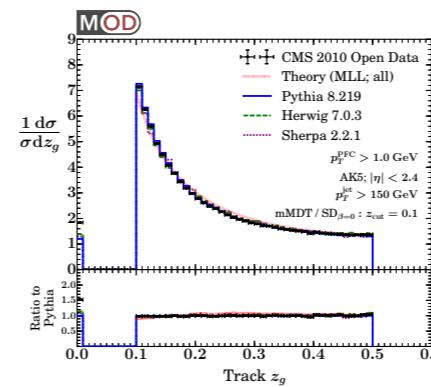
Since 2014: CMS Open Data project



[<http://opendata.cern.ch/>]

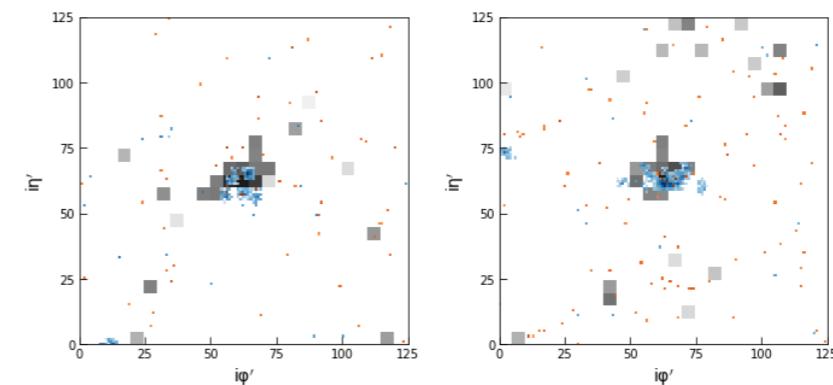
Standard Model & QCD

[Larkoski, Marzani, JDT, Tripathi, Xue, [PRL 2017, PRD 2017](#);
Mehdiabadi, Fahim, [arXiv 2019](#);
Apyan, Cuozzo, Klute, Saito, Schott, Sintayehu, [arXiv 2019](#)]



Machine Learning

[Fernández Madrazo, Heredia Cacha, Lloret Iglesias, de Lucas, [arXiv 2017](#);
Andrews, Paulini, Gleyzer, Poczos, [arXiv 2018](#);
APGP + Alison, An, Bryant, Burkle, Narain, Usai, [arXiv 2019](#);
Komiske, Mastandrea, Metodiev, Naik, JDT, [arXiv 2019](#)]



New Physics Searches

[Cesarotti, Soreq, Strassler, JDT, Xue, [PRD 2019](#);
Lester, Schott, [arXiv 2019](#)]

