Collision Course

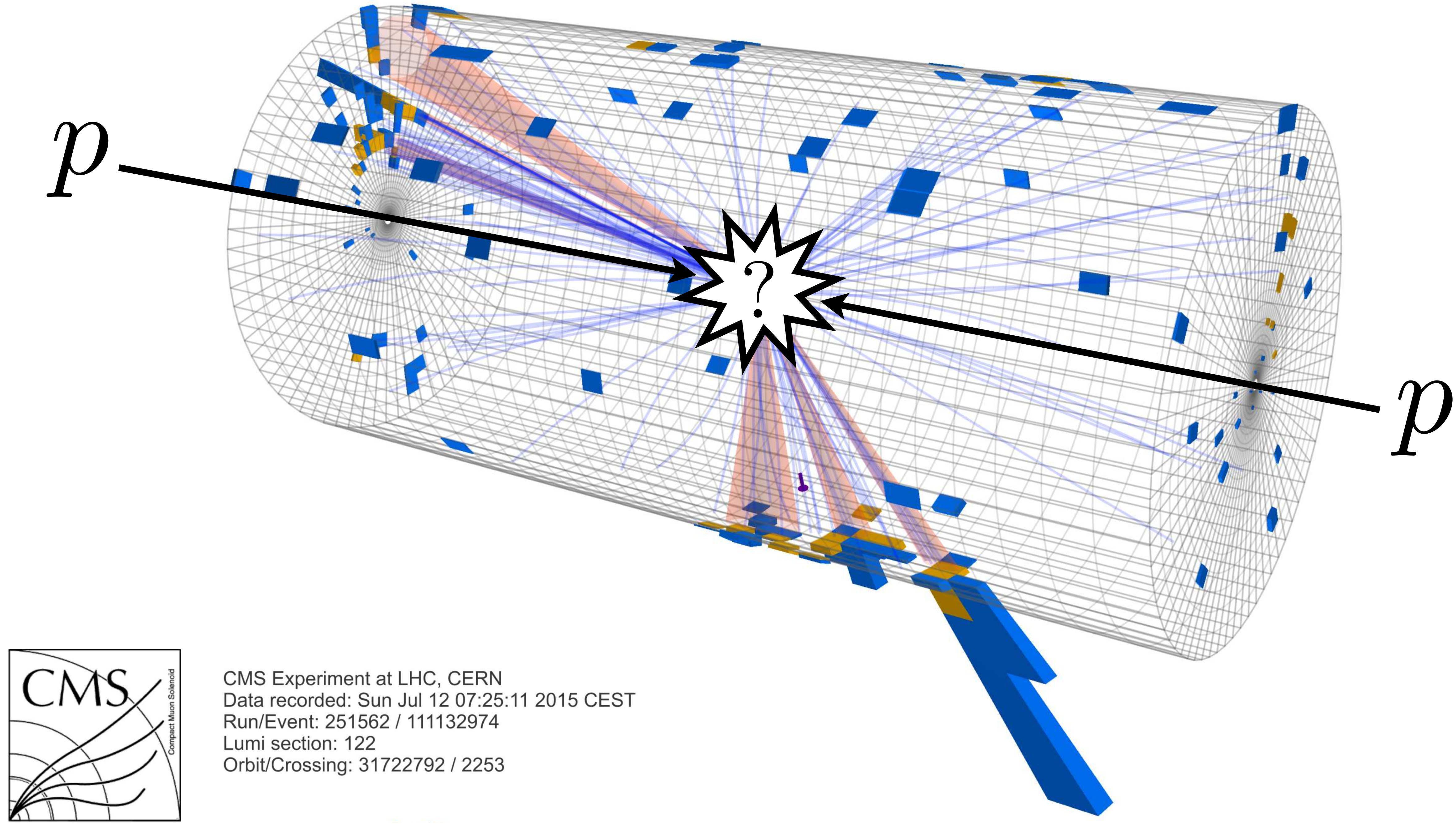
Particle Physics meets Machine Learning

Jesse Thaler

MIT

Oakland University Physics Colloquium — October 3, 2019

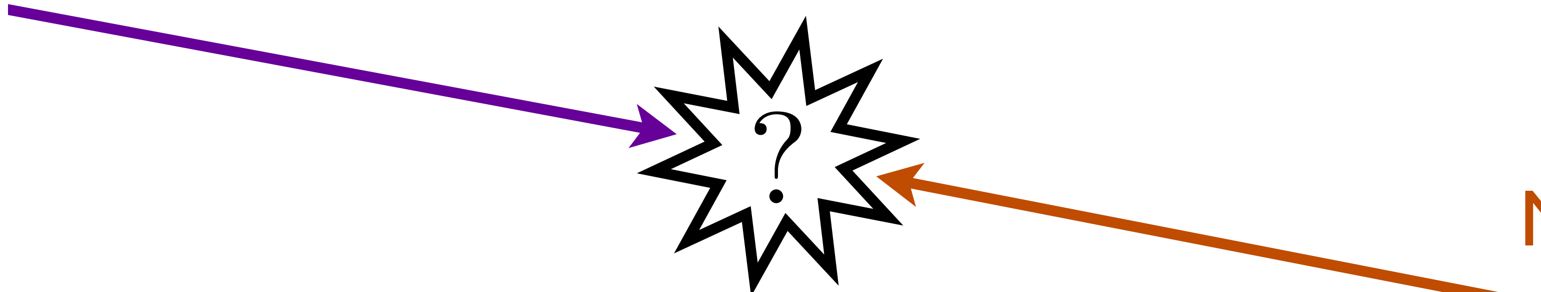
“Collision Course”



“Collision Course”

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

Theoretical
(High Energy)
Physics



Mathematics,
Statistics,
Computer Science

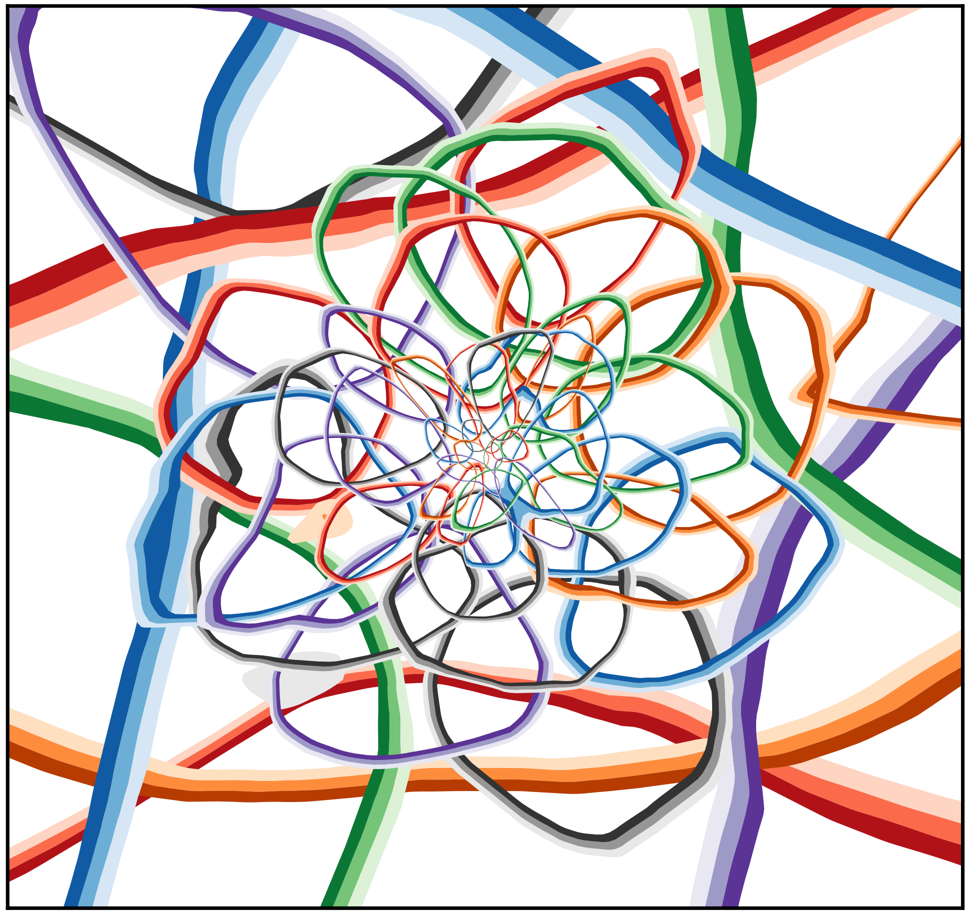
*New insights into particle physics**
*facilitated by advances in machine learning**

(and vice versa?)

Today's Talk: Two Anecdotes

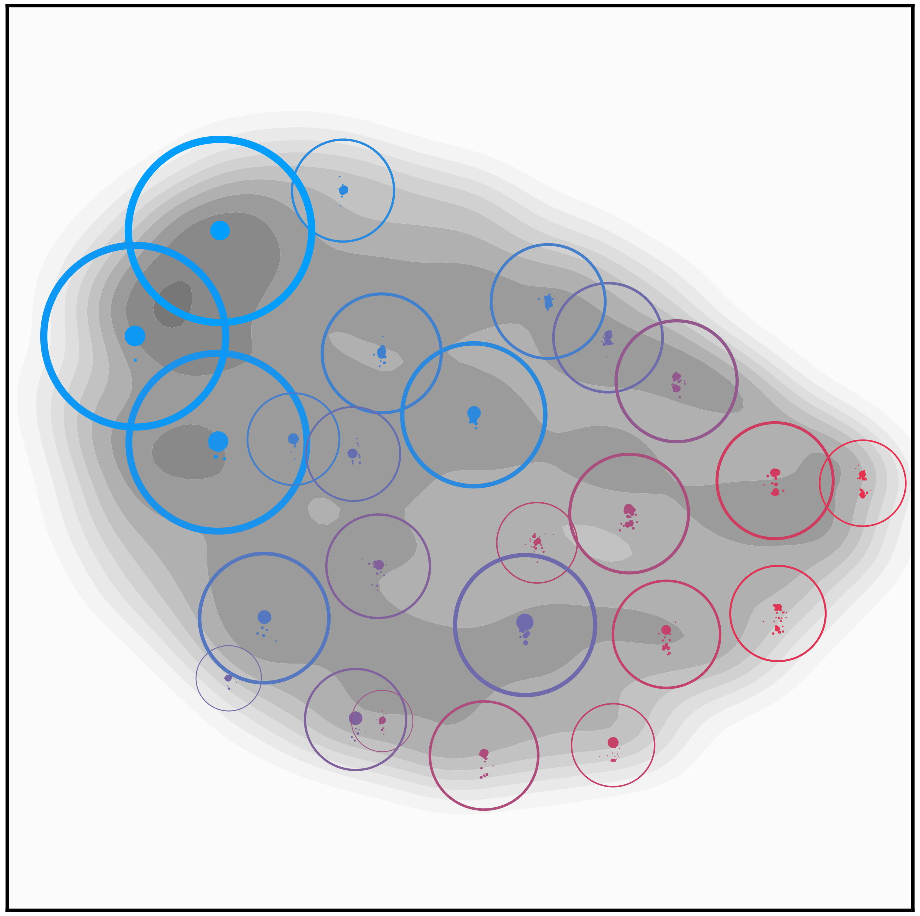



Teaching a Machine to "Think Like a Physicist"



[Komiske, Metodiev, JDT, JHEP 2019]

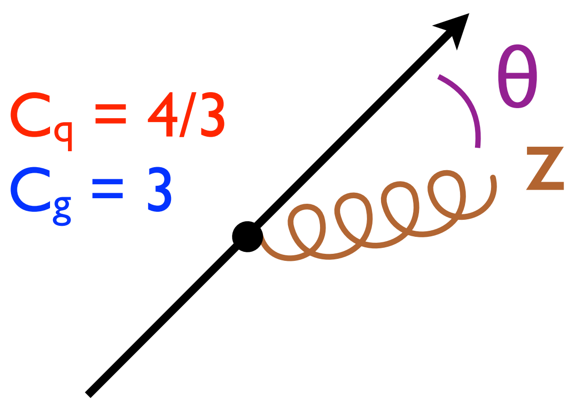
Letting Collider Data Speak for Itself



[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

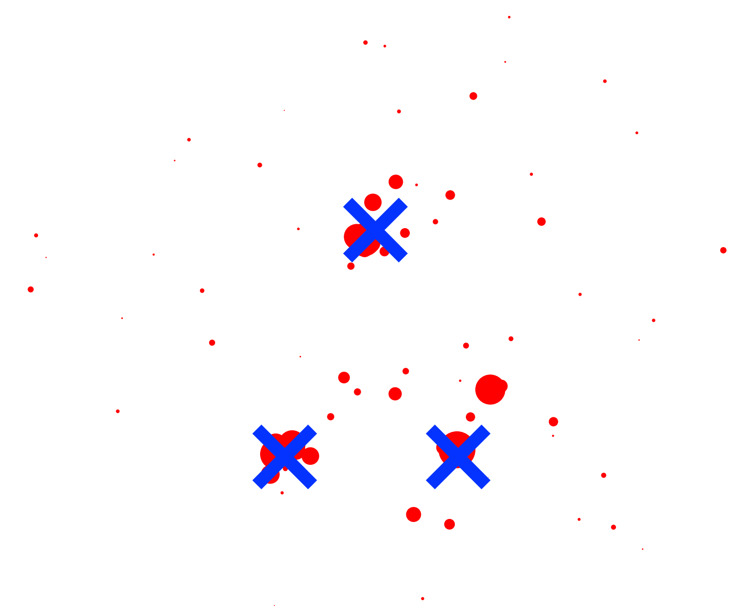
Exploiting a Core Prediction of QCD



$$\mathrm{d}P_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{\mathrm{d}\theta}{\theta}\frac{\mathrm{d}z}{z}$$

[Altarelli, Parisi, NPB 1977]

N-prong Singularities of Gauge Theories

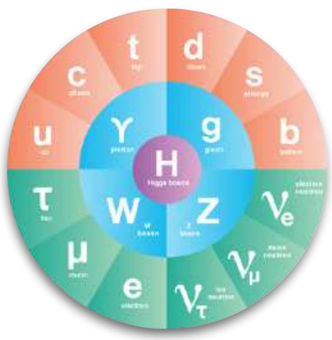


$$\tau_N(\mathcal{E}) = \min_{\text{axes}} \sum_i E_i \min_{a\in[1..N]}\{\theta_{a,i}\}$$

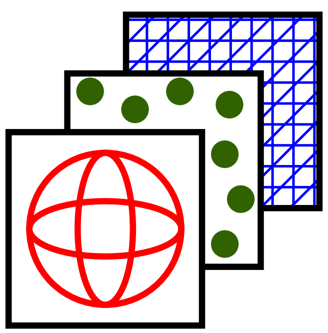
[Stewart, Tackmann, Waalewijn, PRL 2010;
JDT, Van Tilburg, JHEP 2011, JHEP 2012]

New perspectives on *key theoretical concepts*

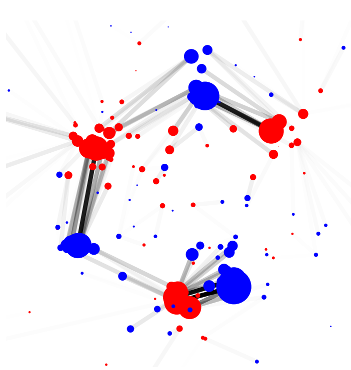
Outline



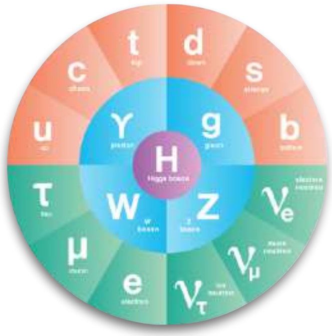
Particle Physics 101



What is a Collider Event?



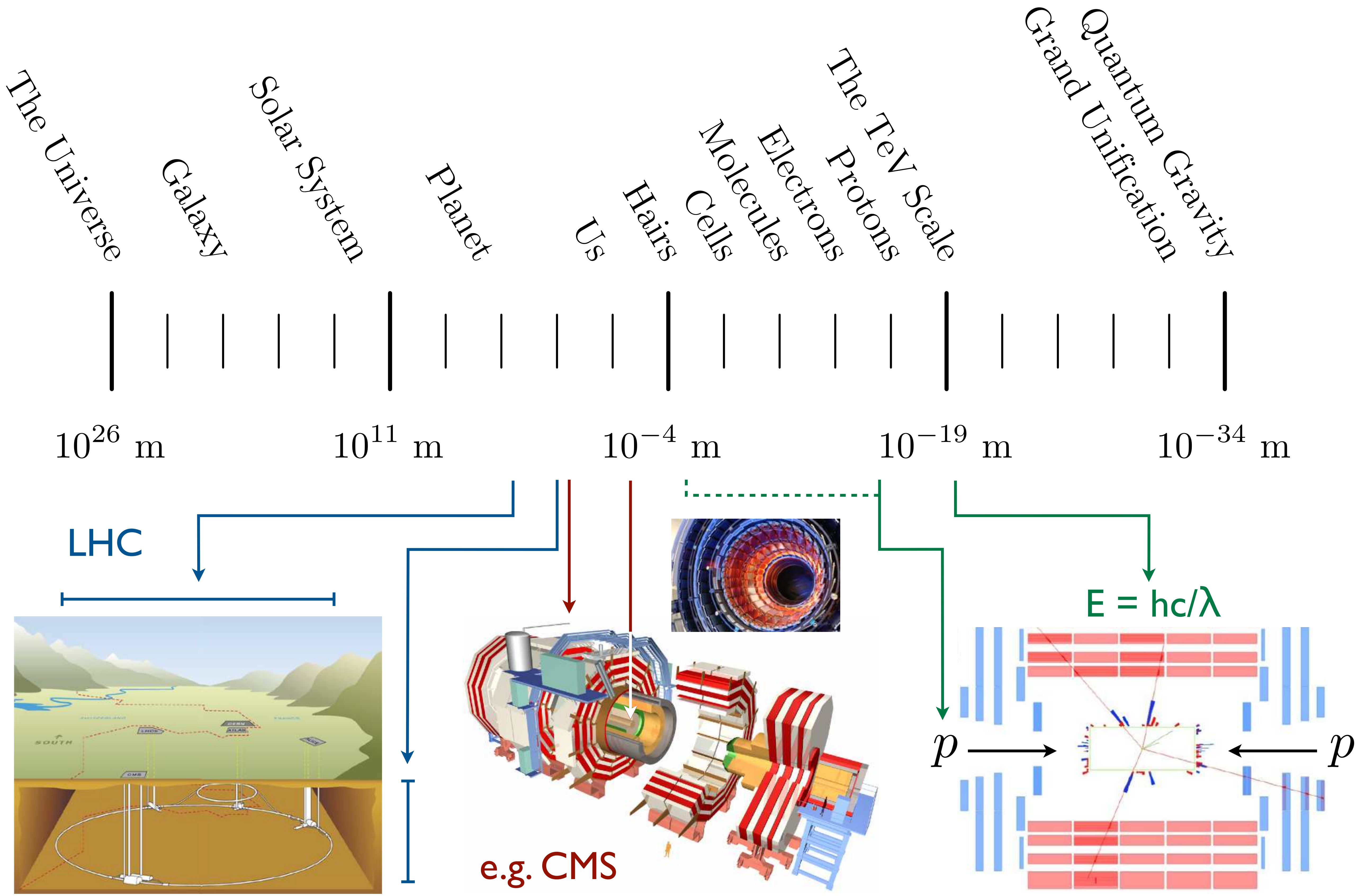
When are Collider Events Similar?



Particle Physics 101

What is a Collider Event?

When are Collider Events Similar?


The Universe
Galaxy
Solar System
Planet
Us
Hairs
Cells
Molecules
Electrons
Protons
The TeV Scale
Grand Unification
Quantum Gravity
10^{26} m
10^{11} m
10^{-4} m
10^{-19} m
10^{-34} m
LHC
e.g. CMS
E = hc/λ
p
p

Mesons

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

$\rho^\pm\,\rho^0\,\omega\,K^{*\pm}\,K^{*0}\,\phi\,D^{*\pm}\,D^{*0}\,D_s^{*\pm}\,J/\psi\,B^{*\pm}\,B^{*0}\,B_s^{*0}\,\Upsilon\,\ldots$

Baryons

$p\,n\,\Lambda^0\,\Sigma^+\,\Sigma^0\,\Sigma^-\,\Xi^0\,\Xi^-\,\ldots$

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

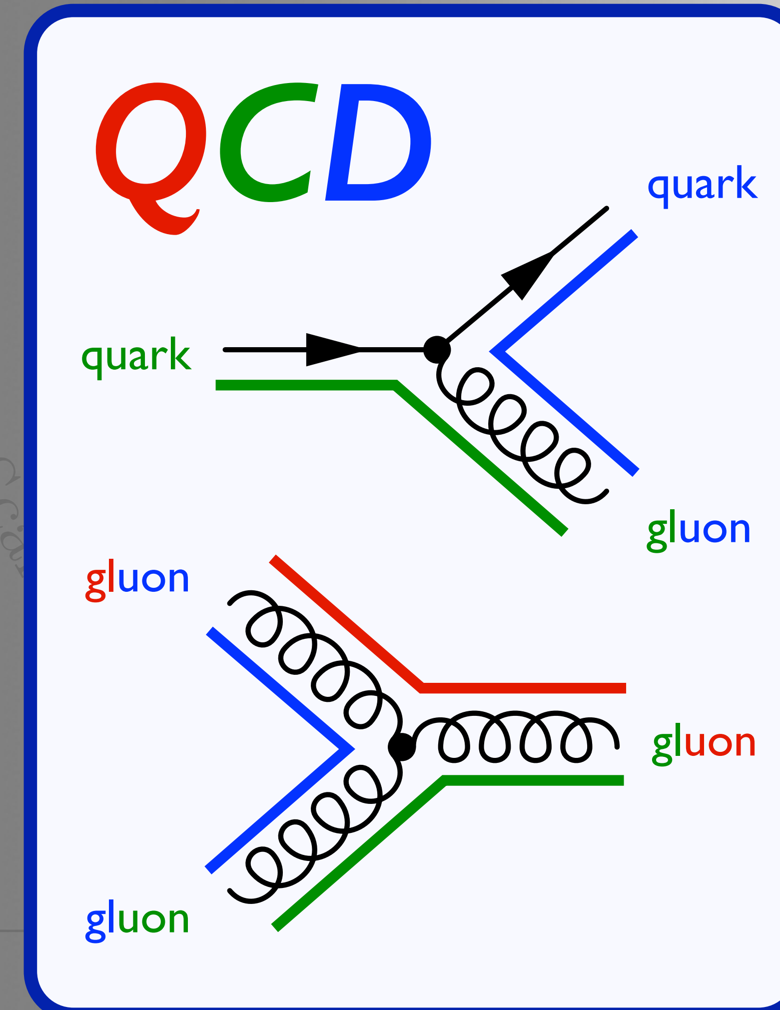
Tetraquarks

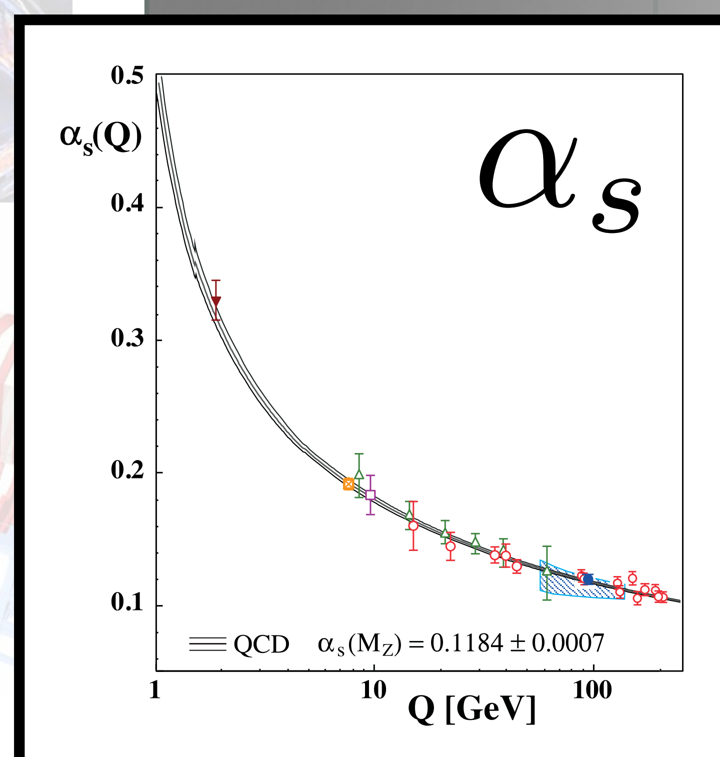
$X(3872)\,Y(4260)\,Z(4430)\,\ldots$

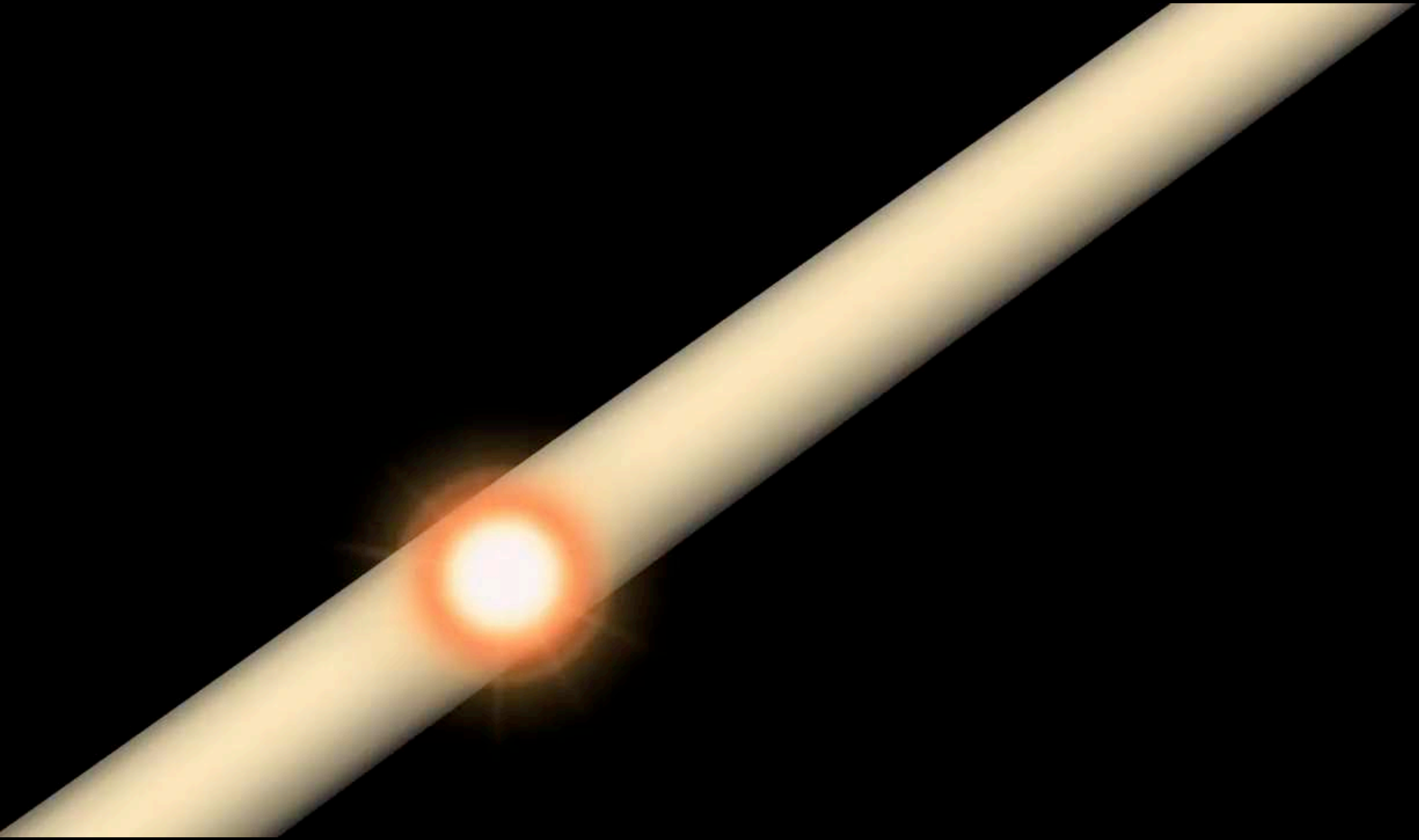
Pentaquarks

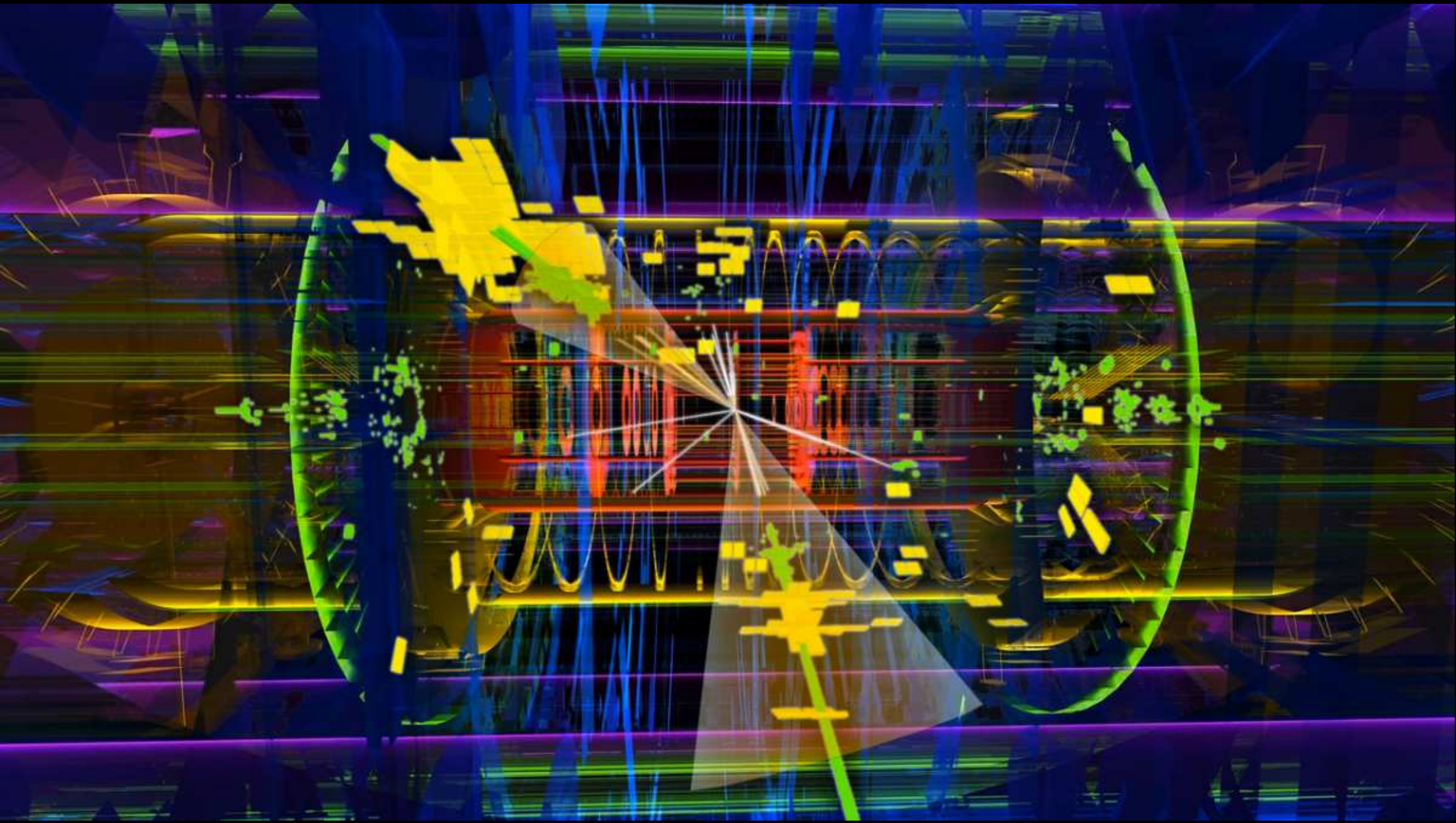
$P_c^+(4450)\,\ldots$

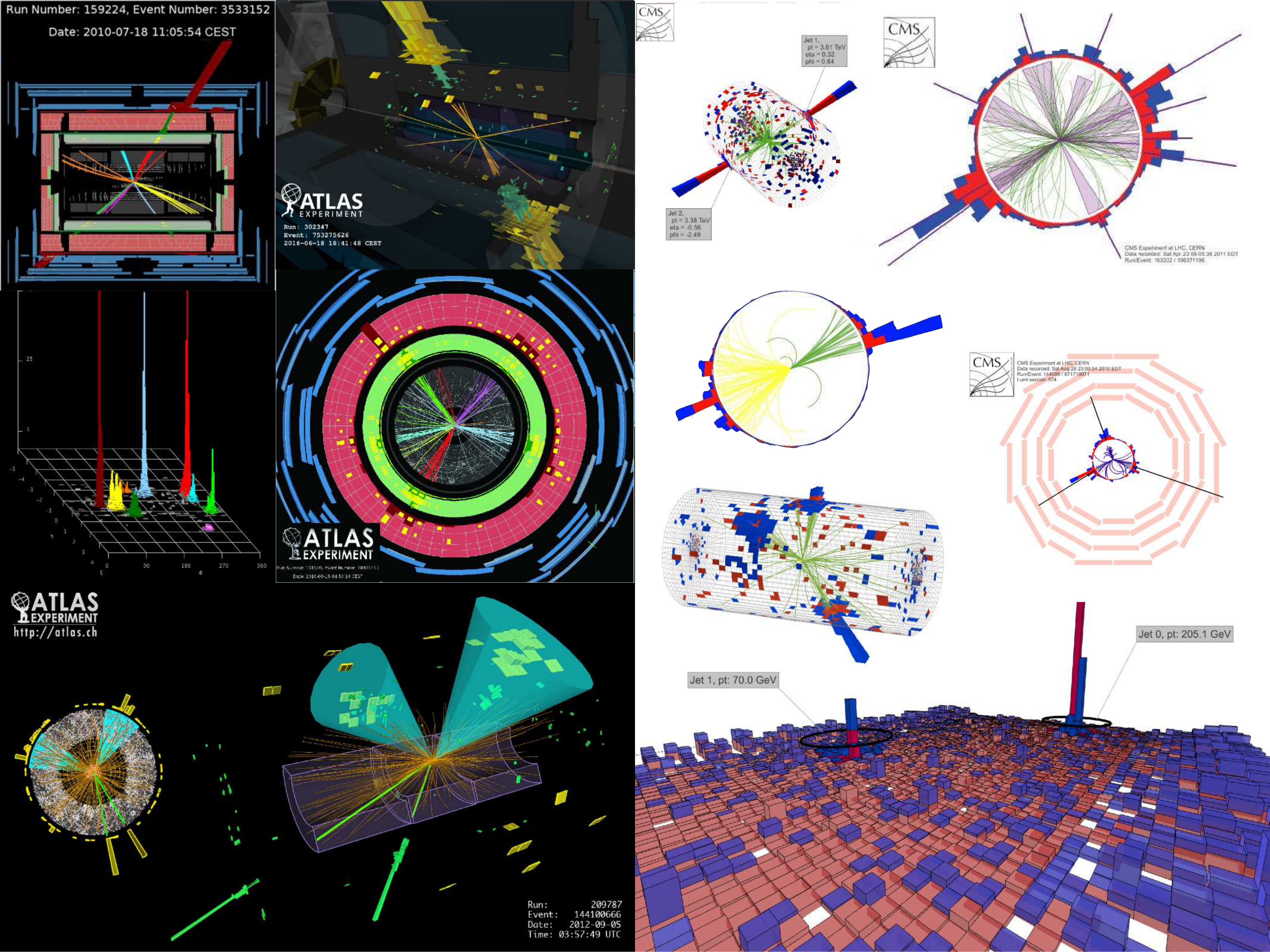
Jets



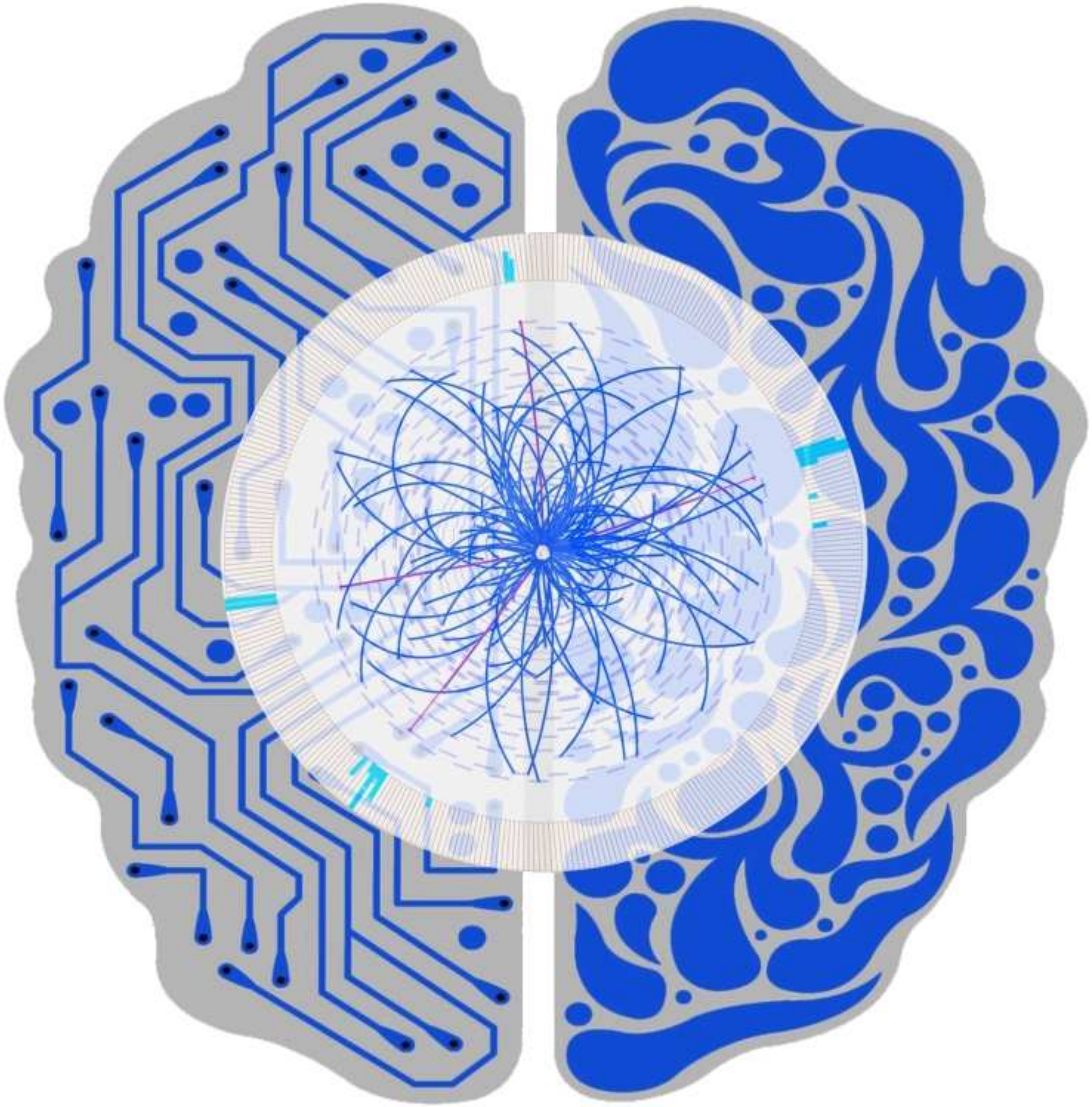






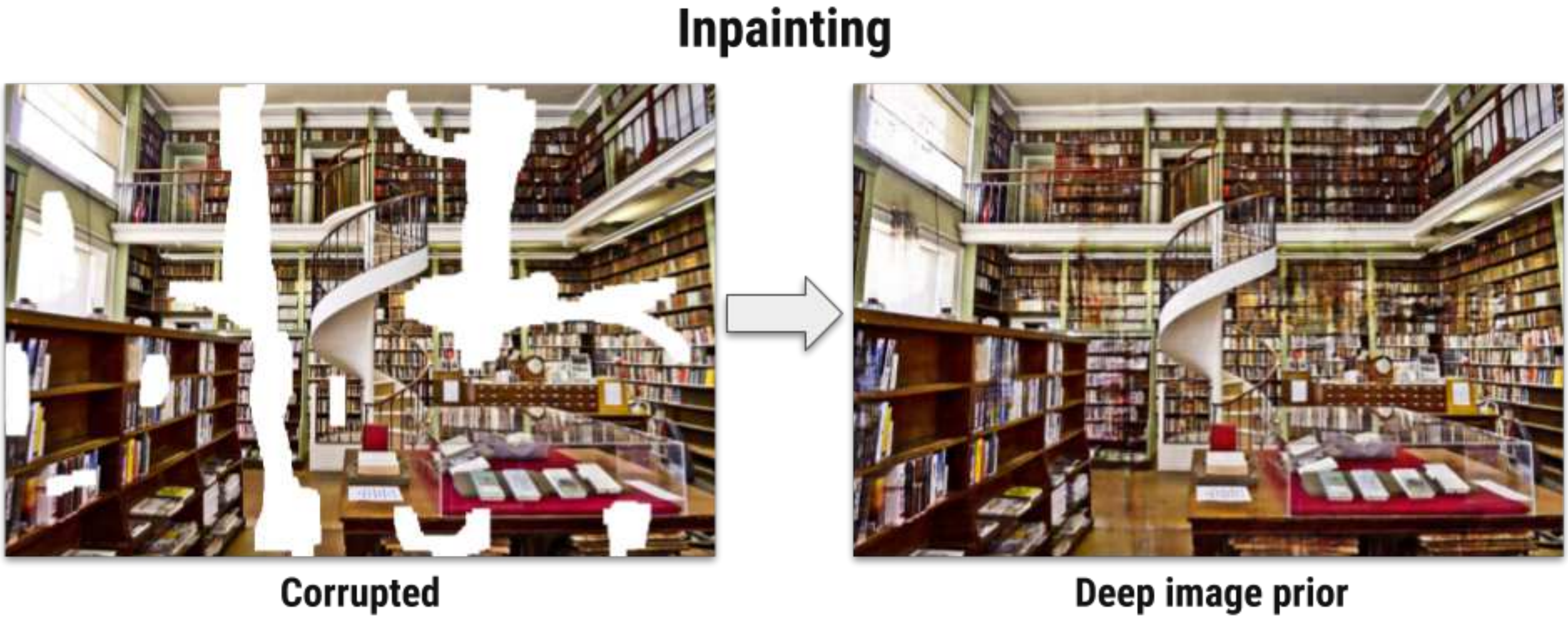


Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST
ATLAS EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST
CMS
Jet 1, pt = 3.61 TeV eta = 0.32 phi = 0.64
Jet 2, pt = 3.38 TeV eta = -0.56 phi = -2.49
CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 196371106
ATLAS EXPERIMENT
CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Aug 28 23:03:34 2010 EDT
Run/Event: 144089 / 671718071
Lumi section: 574
ATLAS EXPERIMENT
http://atlas.ch
Jet 0, pt: 205.1 GeV
Jet 1, pt: 70.0 GeV
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC



The Rise of Machine Learning

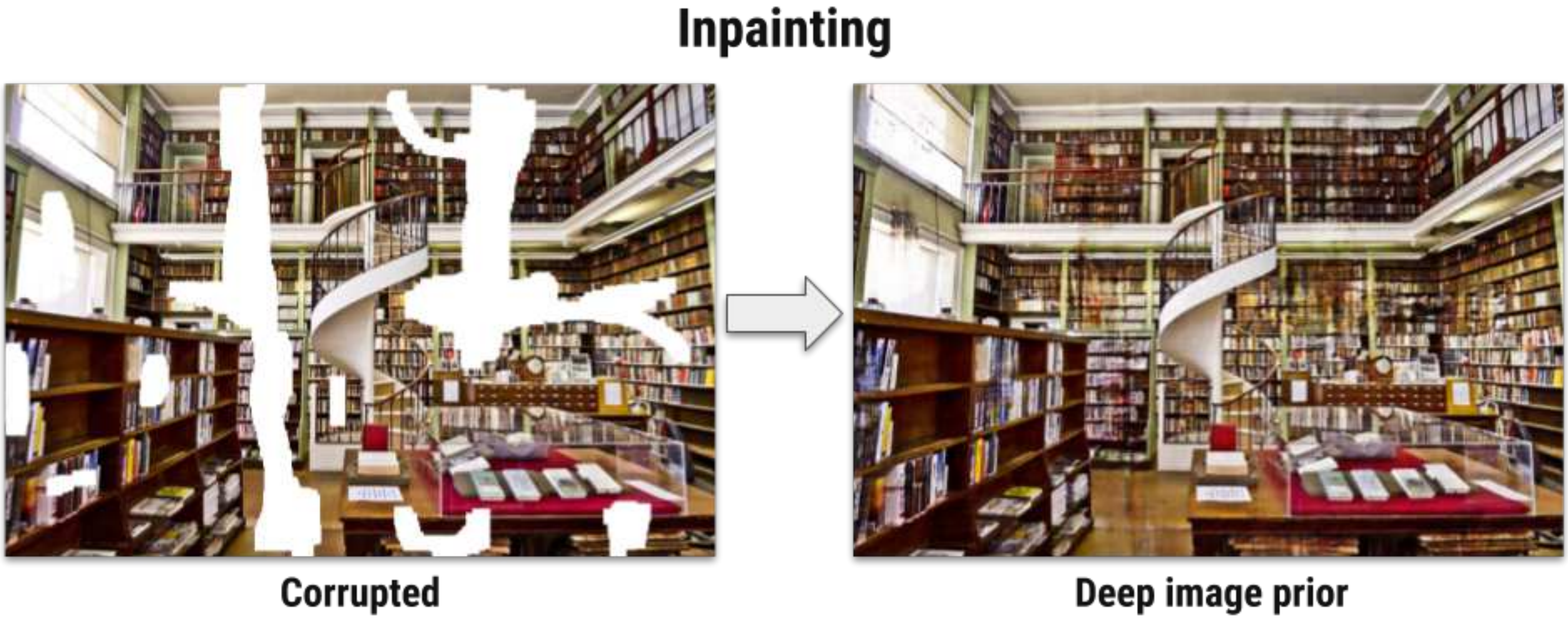
Deep Learning



increased computational power and large data sets

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

Deep Learning (or Deep Thinking?)



Using randomly initialized neural network (!)

Progress made by understanding the structure of problems

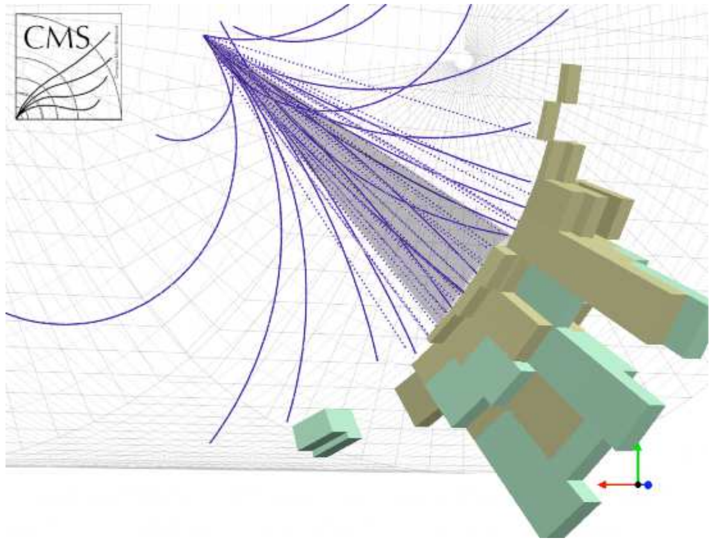
(not just increased computational power and large data sets)

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

Cartoon of Machine Learning

"ML4Jets"
NYU, January 2020

For this talk: $\mathcal{J}$ = jet



Properly specified problem $\Rightarrow$

Many (labeled) examples $\{\mathcal{J}\}$ $\Rightarrow$

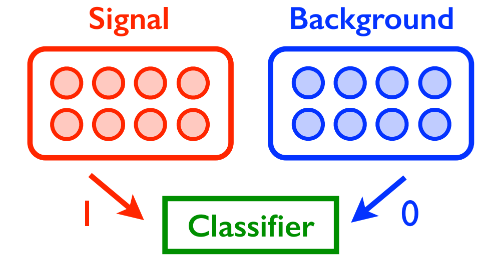
The Machine

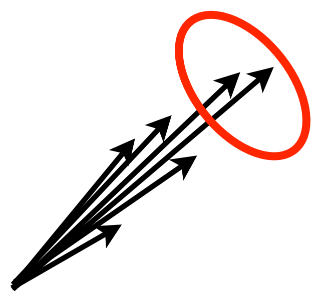
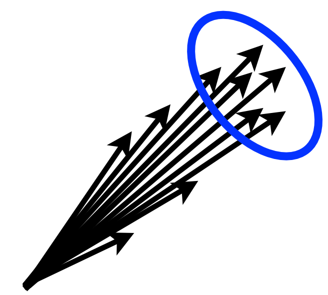
$\Rightarrow$ Solution $S(\mathcal{J})$

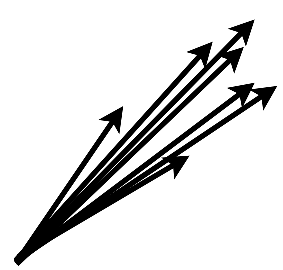
E.g.:	Problem	=	Minimize loss function
	Solution	=	Multi-layer neural network
	Strategy	=	Stochastic gradient descent

E.g. Jet Classification

Key supervised learning task at LHC



e.g.  **Quark** $C_q = 4/3$ vs.  **Gluon** $C_g = 3 = 9/3$

Find $h\Big($  $\Big)$ such that

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

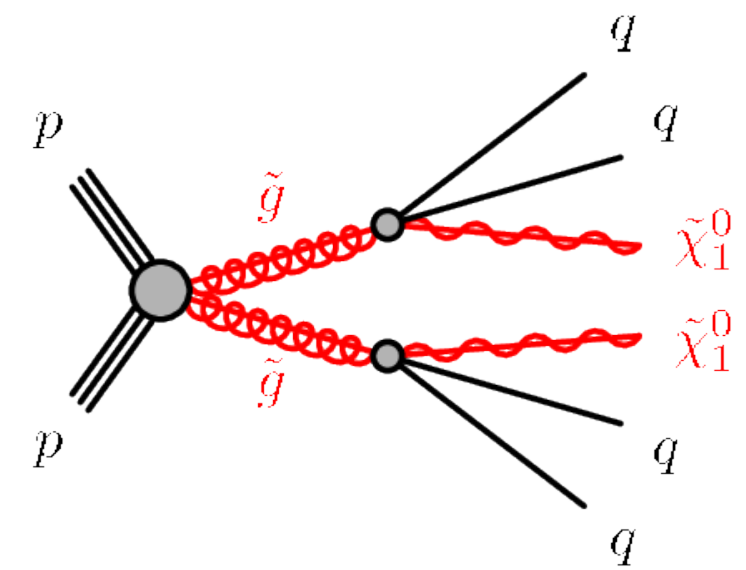
$$h(\text{Gluon}) = 0$$

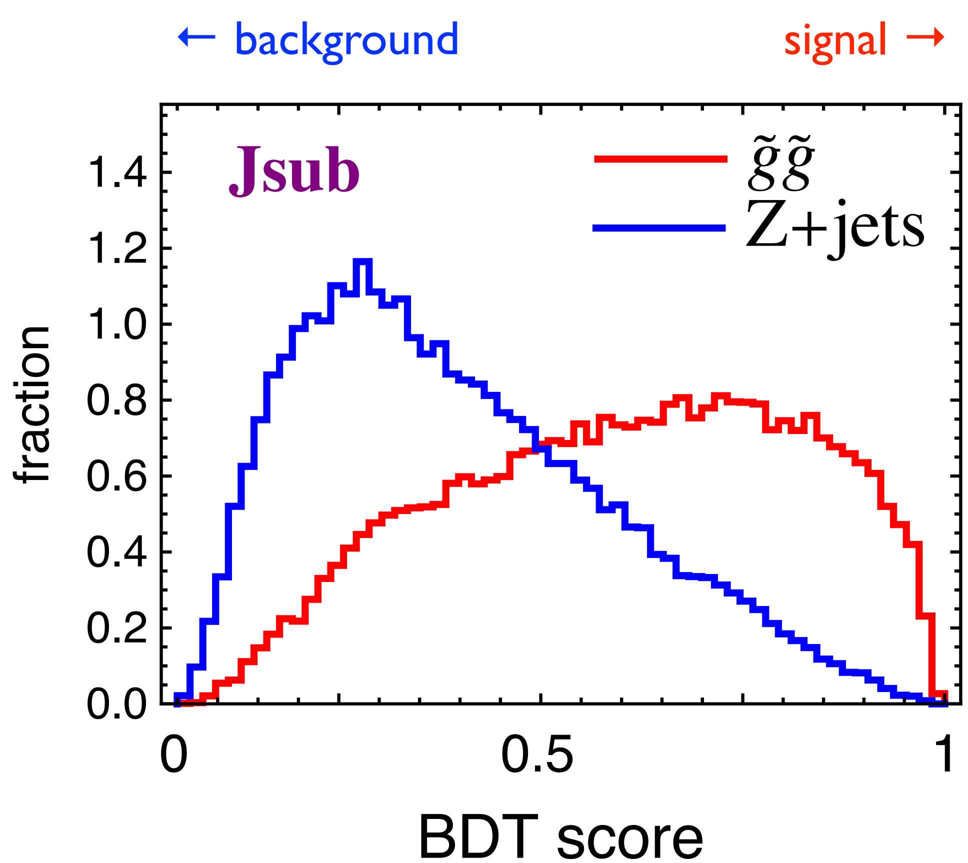
Best you can do: $h(\mathcal{J}) = \dfrac{p(\mathcal{J}|\mathrm{Q})}{p(\mathcal{J}|\mathrm{Q}) + p(\mathcal{J}|\mathrm{G})}$

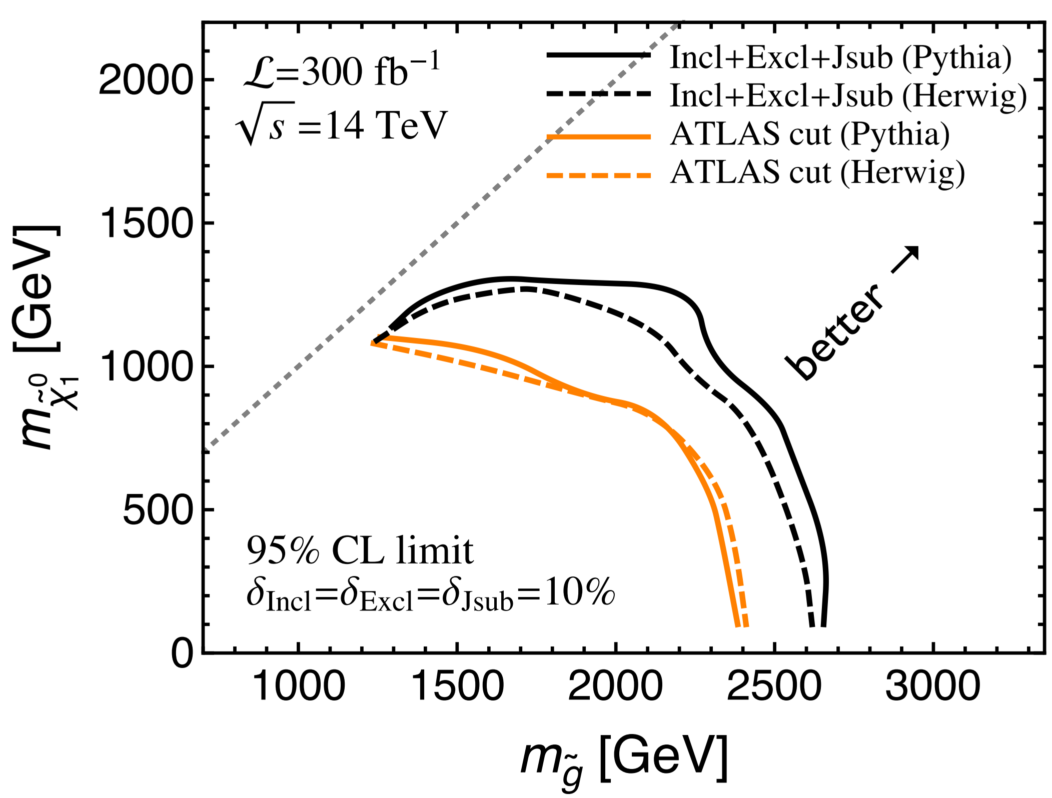
(Neyman-Pearson lemma)

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**

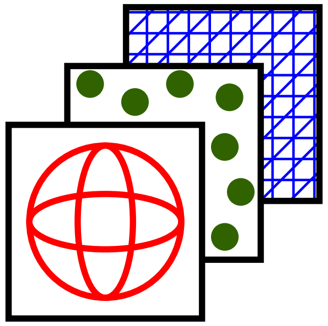






[Bhattacherjee, Mukhopadhyay, Nojiri, Sakakie, Webber, JHEP 2017]

Particle Physics 101

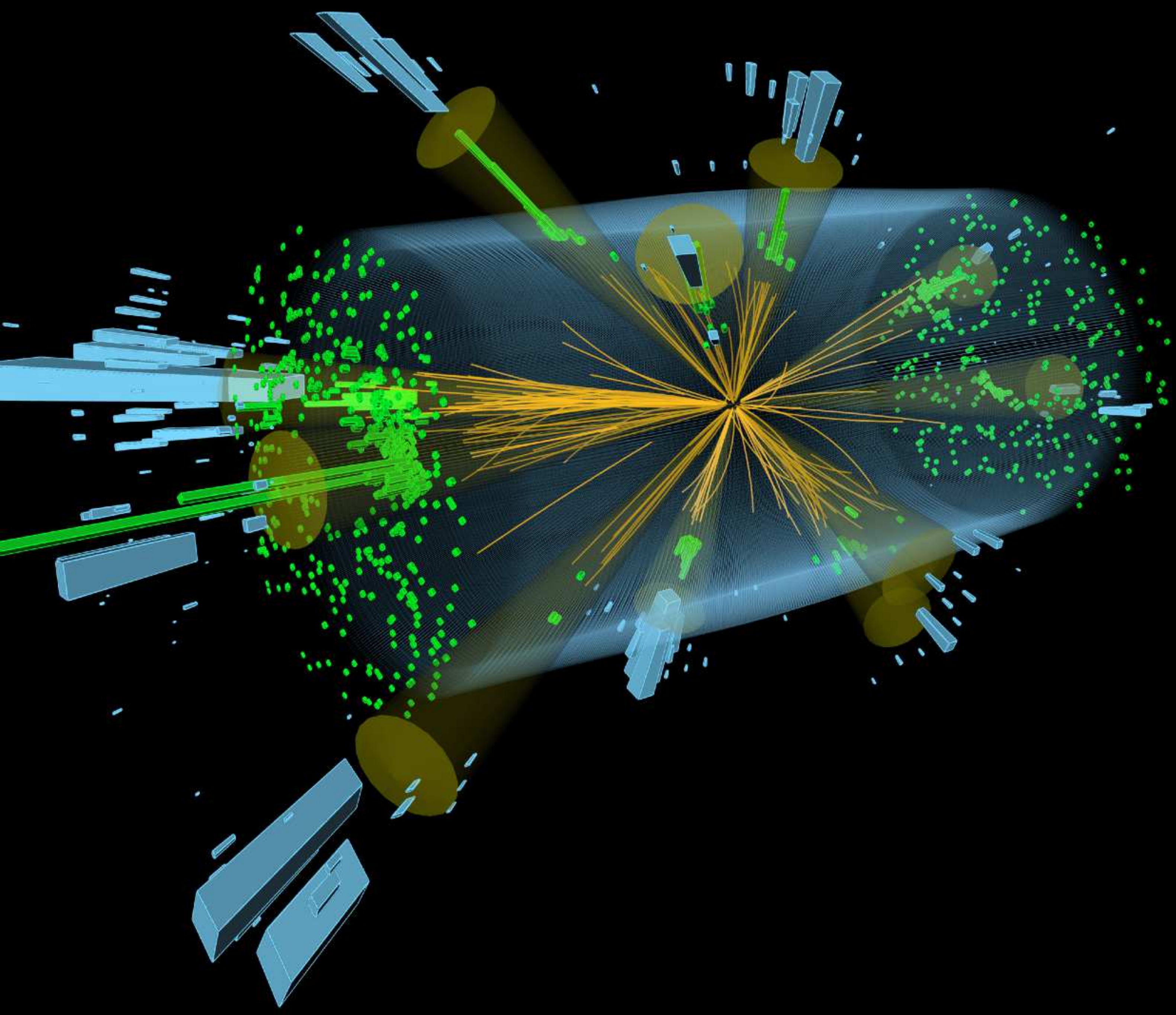


What is a Collider Event?

When are Collider Events Similar?

Collider Event

Collection of points in (momentum) space



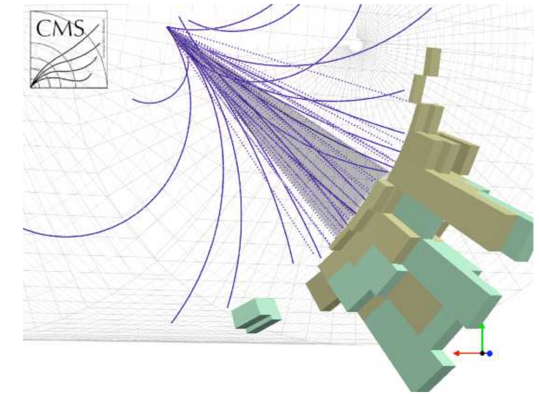
	T	E	H	M		
elementary		●			γ	photon
elementary	●	●			$e^{\pm}$	electron
elementary	●	•	•	●	$\mu^{\pm}$	muon
composite	●	•	●		$\pi^{\pm}$	pion
composite	●	•	●		$K^{\pm}$	kaon
composite		•	●		K_L^0	K-long
composite	●	•	●		$p/\overline{p}$	proton
composite		•	●		$n/\overline{n}$	neutron

Point Cloud

Collection of points in (position) space

[Popular Science, 2013]

Key: Jets are Point Clouds



- Particle: List of properties

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

Energy Momentum Mass, charge, flavor, vertex, quality, …

- Jet: Set of particles

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

Permutation Symmetry

Quantum Mechanics:
"When you've seen one electron, you've seen them all"

Variable Length

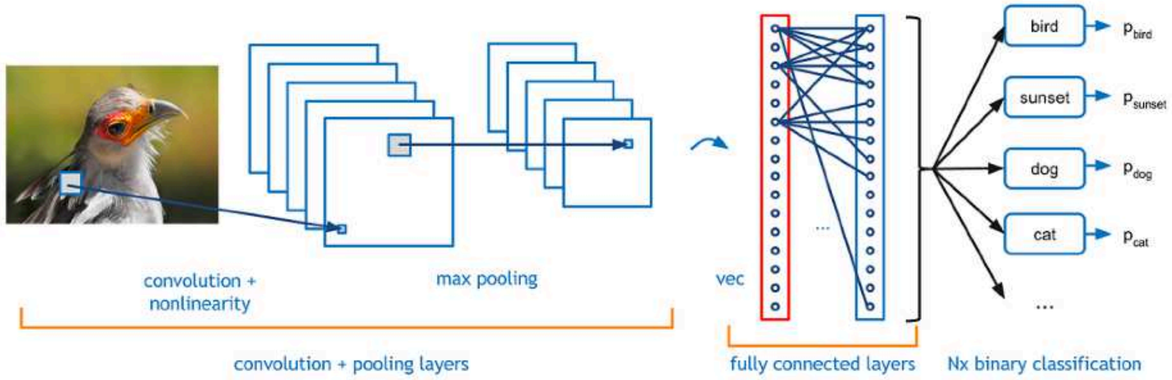
Jet Formation:
Typically 10–50 particles per jet

- Dataset: Set of jets

Off-the-Shelf Machine Learning?

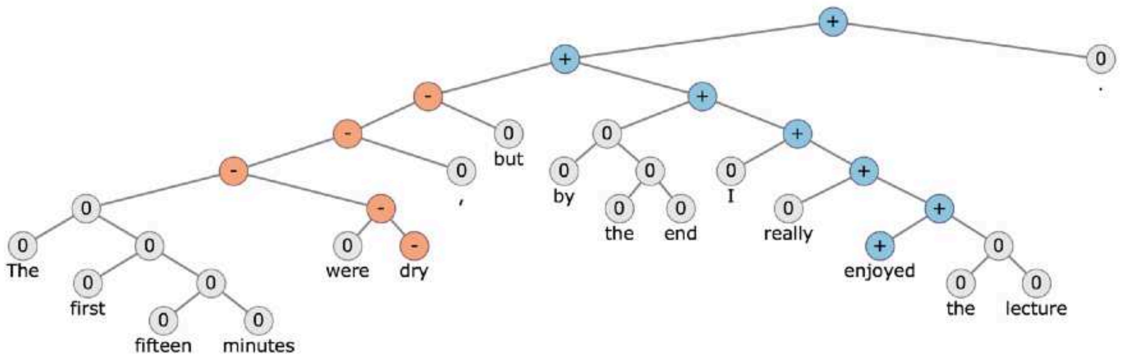
2D Images?

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



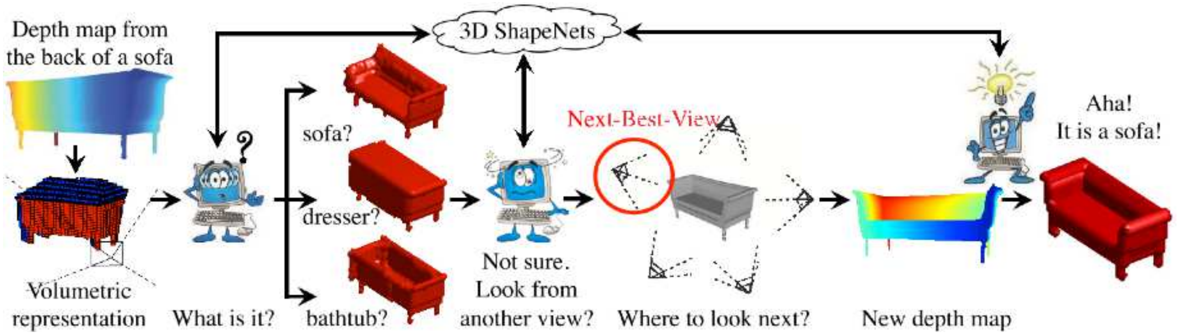
Natural Language?

Clustering can yield "semantic" structure
but inputs are fundamentally symmetric

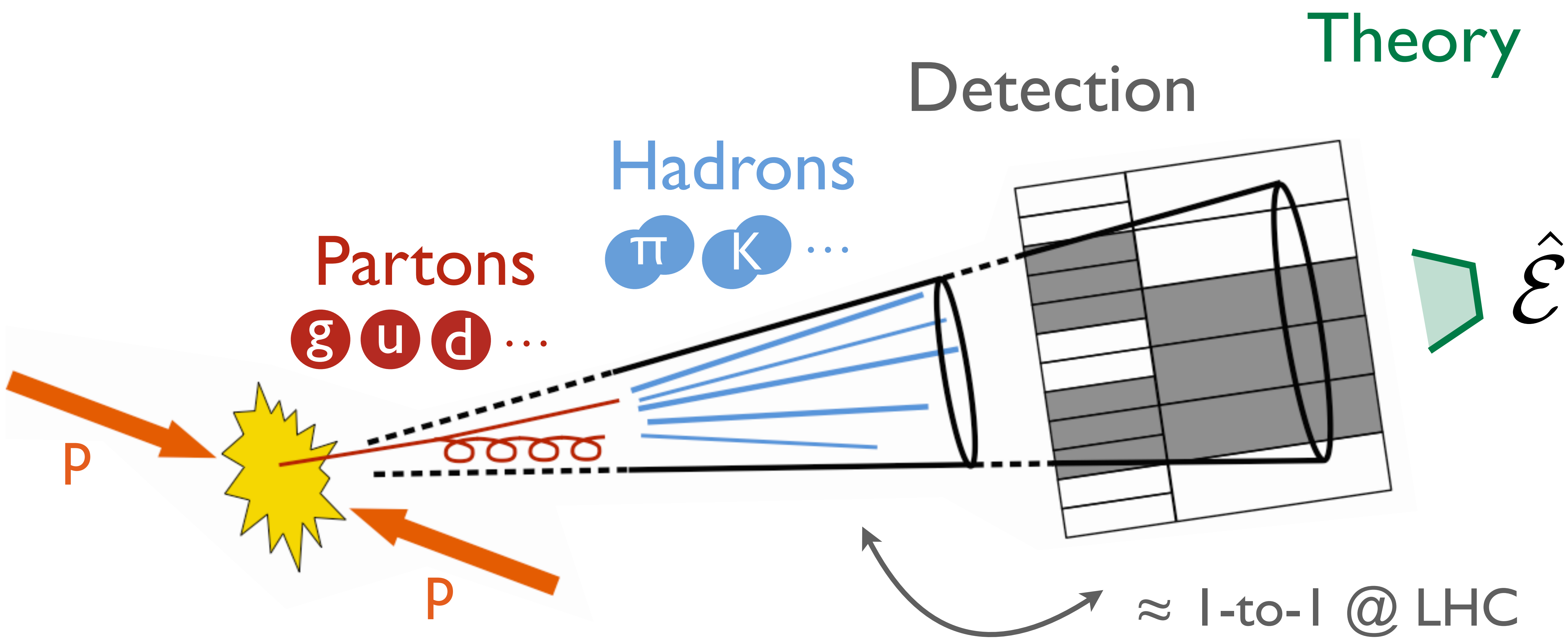


3D Objects?

Much closer to particle physics,
but most architectures are not *"safe"*



Jet Formation Process



Stress-energy flow:

Robust to non-perturbative and detector effects

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

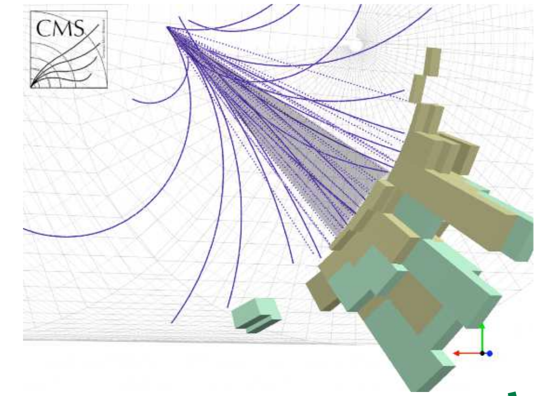
[Sveshnikov, Tkachov, PLB 1996; Hofman, Maldacena, JHEP 2008; Mateu, Stewart, JDT, PRD 2013]

Jets as Weighted Point Clouds

- Energy-weighted directions

$$\vec{p} = \{E, \underbrace{\hat{n}_x, \hat{n}_y, \hat{n}_z}_{\text{Direction}}\}$$

Energy



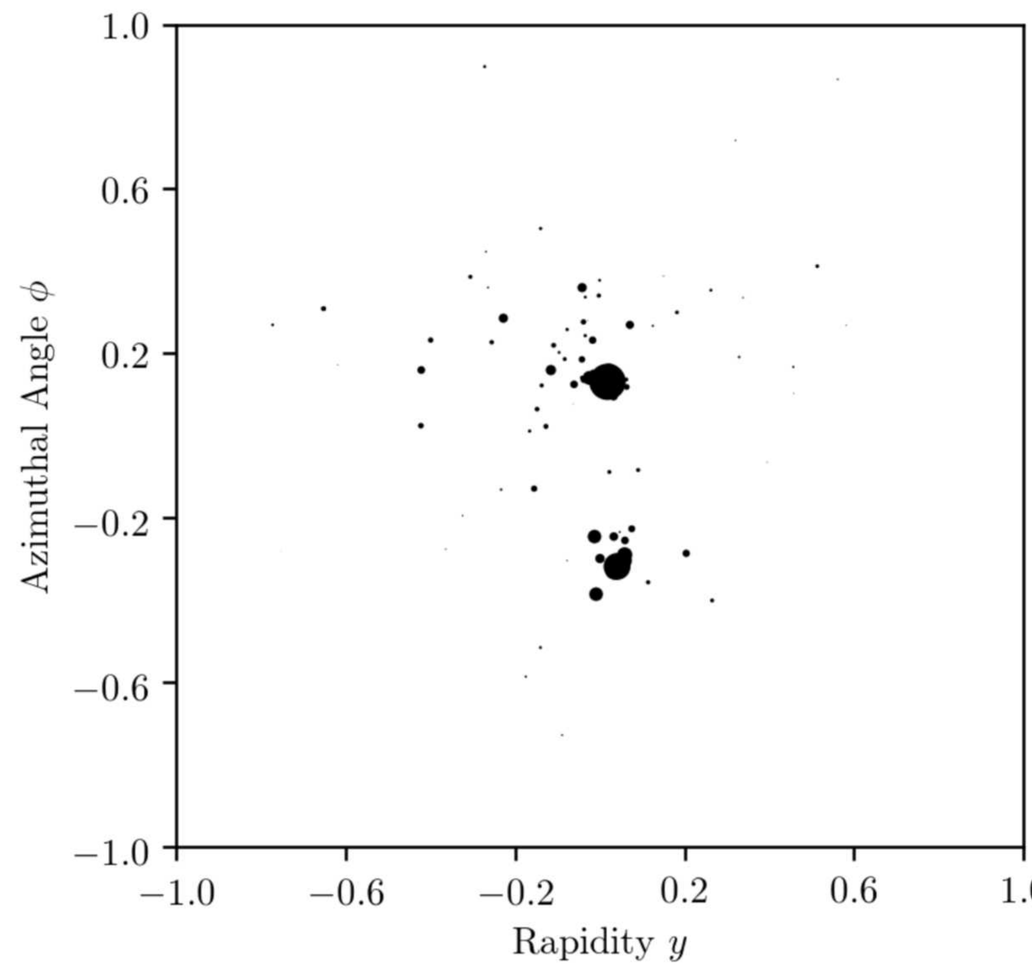
stress-energy flow $\hat{\mathcal{E}}$

- 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

Symmetry: $\mathcal{J} = \{\underbrace{\vec{p}_1, \vec{p}_2, \vec{p}_3, \ldots, \vec{p}_N}_{\text{Unordered, Variable Length Set}}\}$

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

Energy weighting

Properly specified problem ⇨ **The Machine** ⇨ **Solution** $S(\mathcal{J})$

Many (labeled) examples $\{\mathcal{J}\}$ ⇨ **The Machine** ⇨ **Verification** $V(\mathcal{J})$

Check that answer is physically sensible

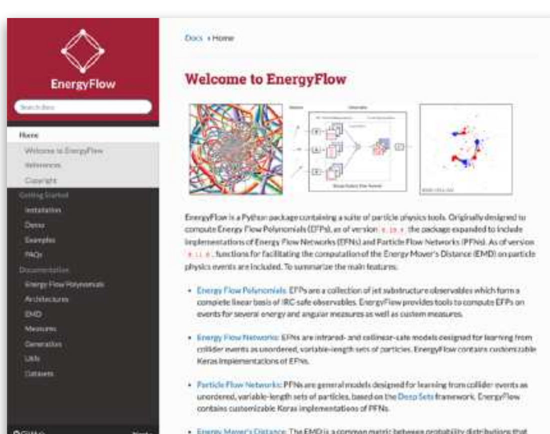
Patrick Komiske

Eric Metodiev

Theoretical (High Energy) Physics

?

Mathematics, Statistics, Computer Science

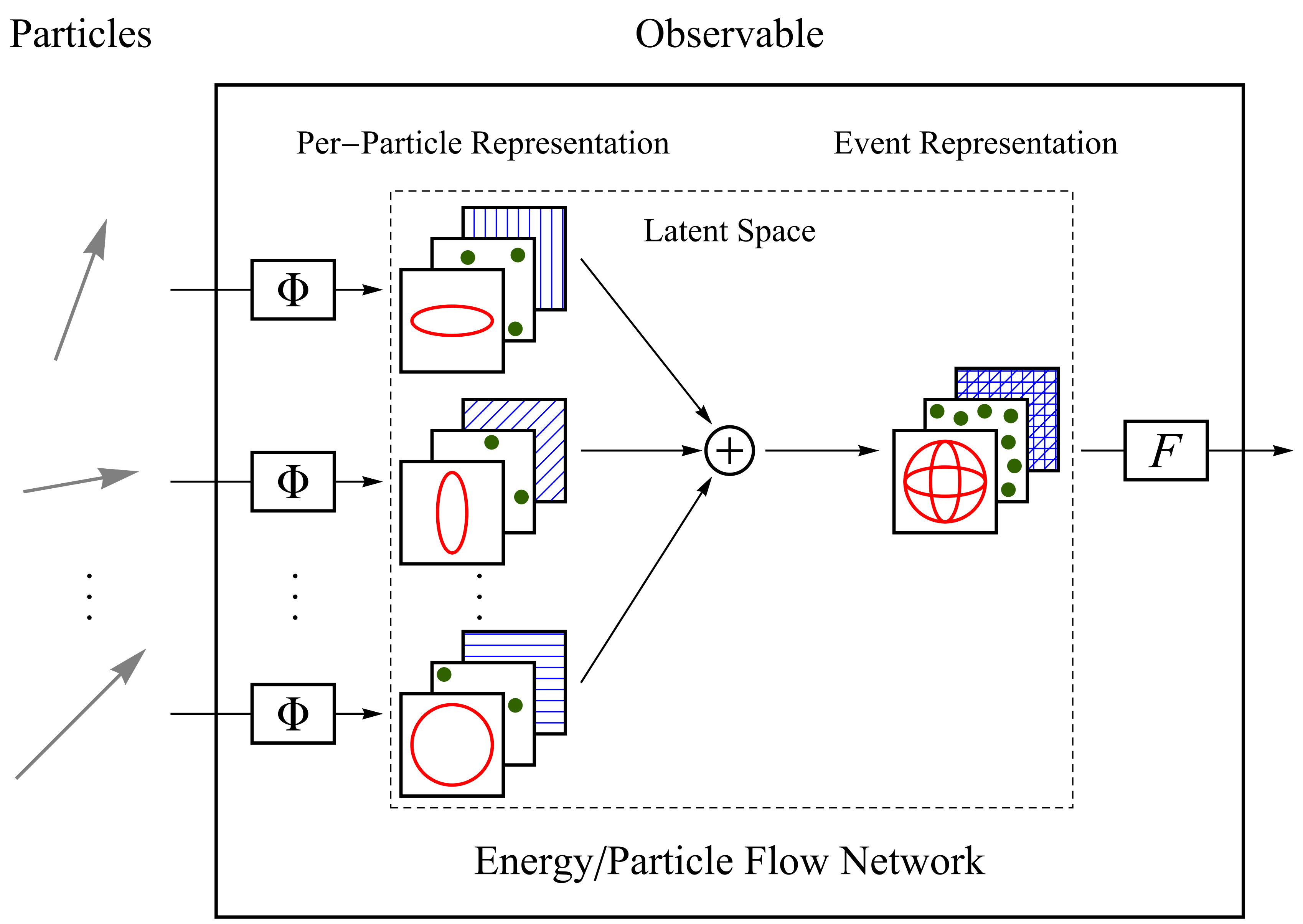


Energy Flow Networks

https://energyflow.network/

Energy Flow Networks

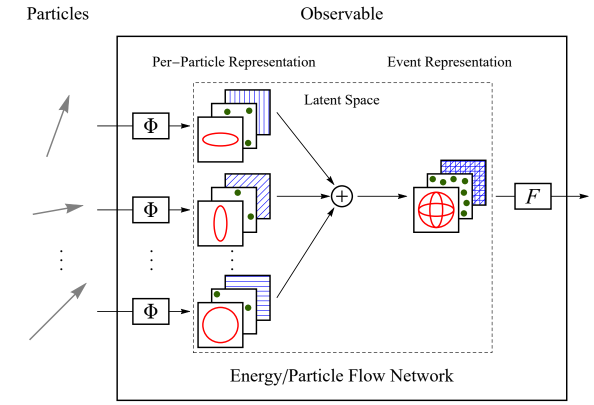
Architecture designed around symmetries and interpretability



[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***



Permutation invariant Linear weights

$$S(\mathcal{J}) = F\big(V_1, V_2, \ldots, V_\ell\big) \qquad V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \, \Phi_a(\hat{n}_i)$$

Parametrized with Neural Networks

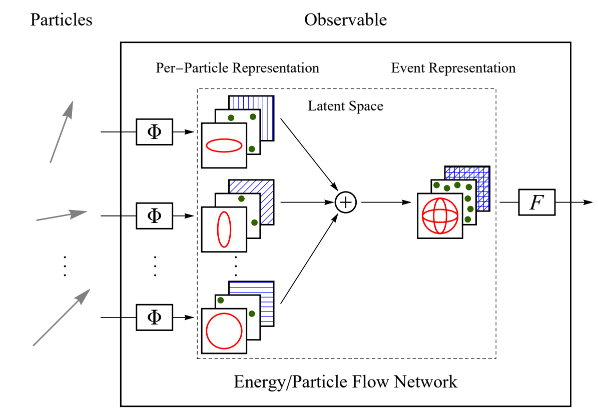
(see backup for details)

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

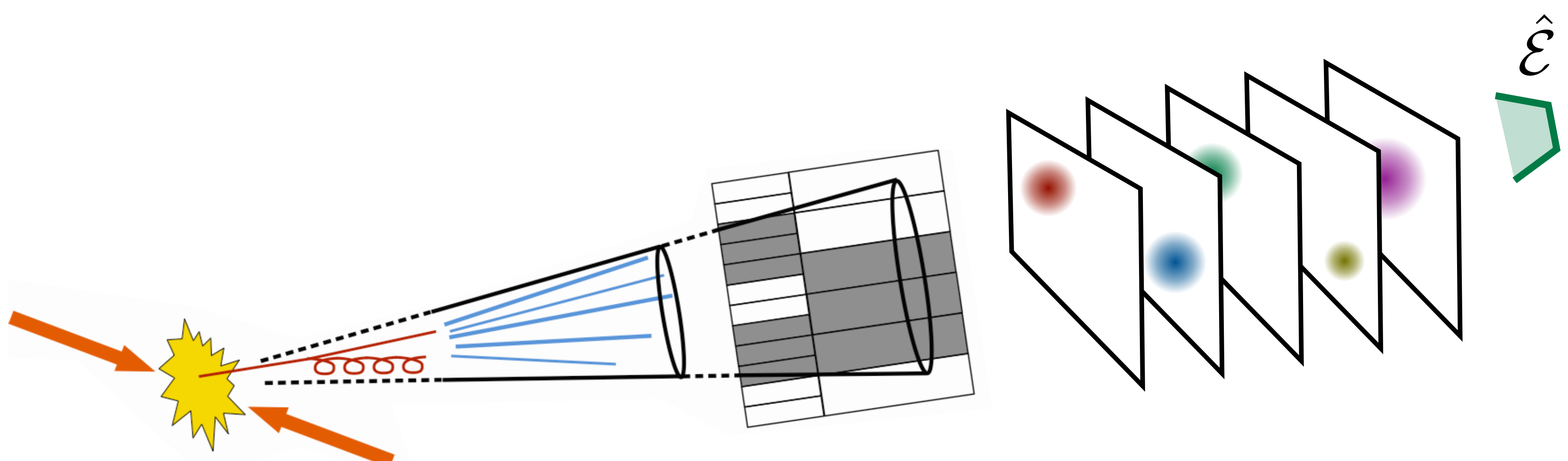


Latent space of dim ℓ

$$S(\mathcal{J}) = F(V_1, V_2, \ldots, V_\ell) \qquad V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \, \Phi_a(\hat{n}_i)$$

Can visualize if ℓ is small

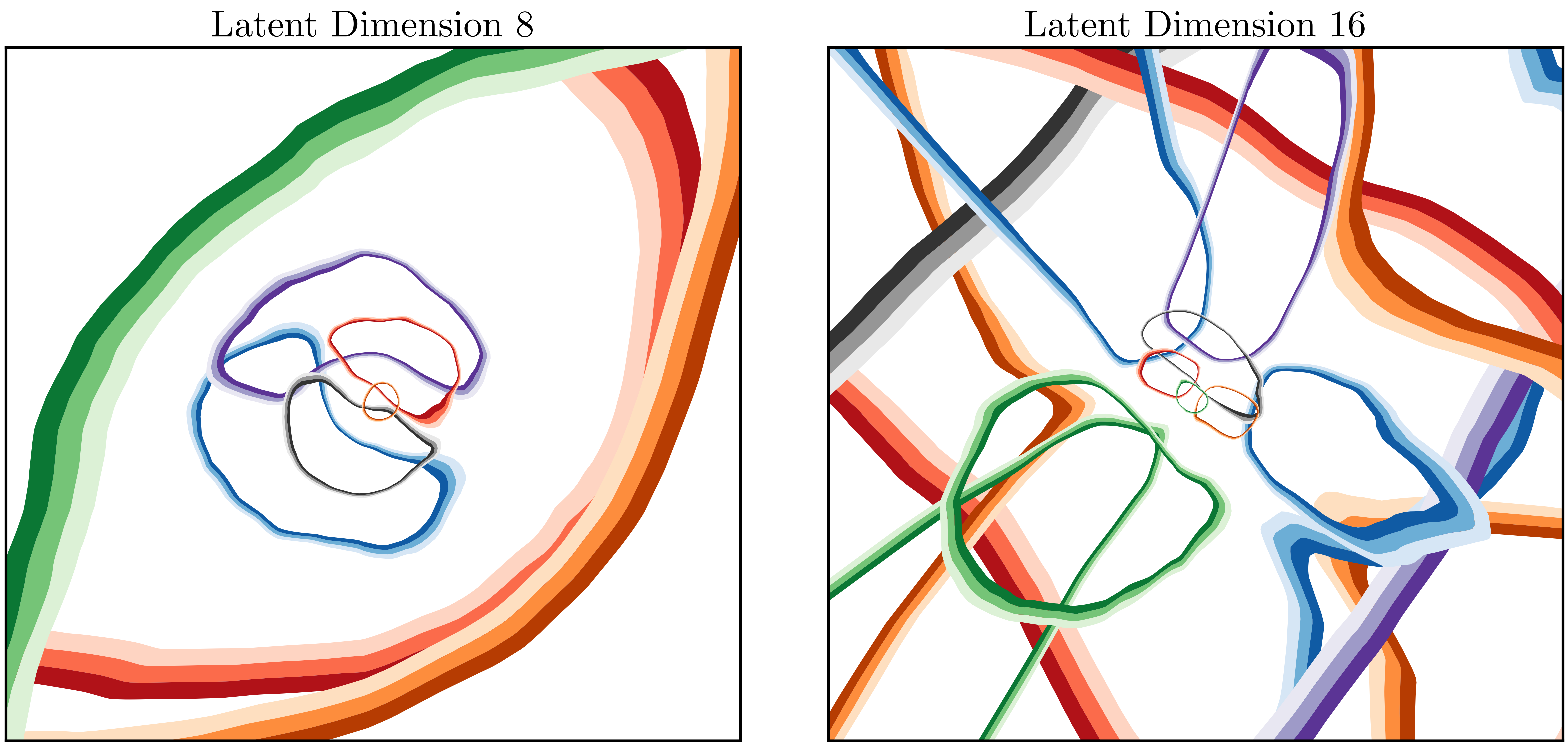
Easy to plot!



(similar to CNN filter activation)

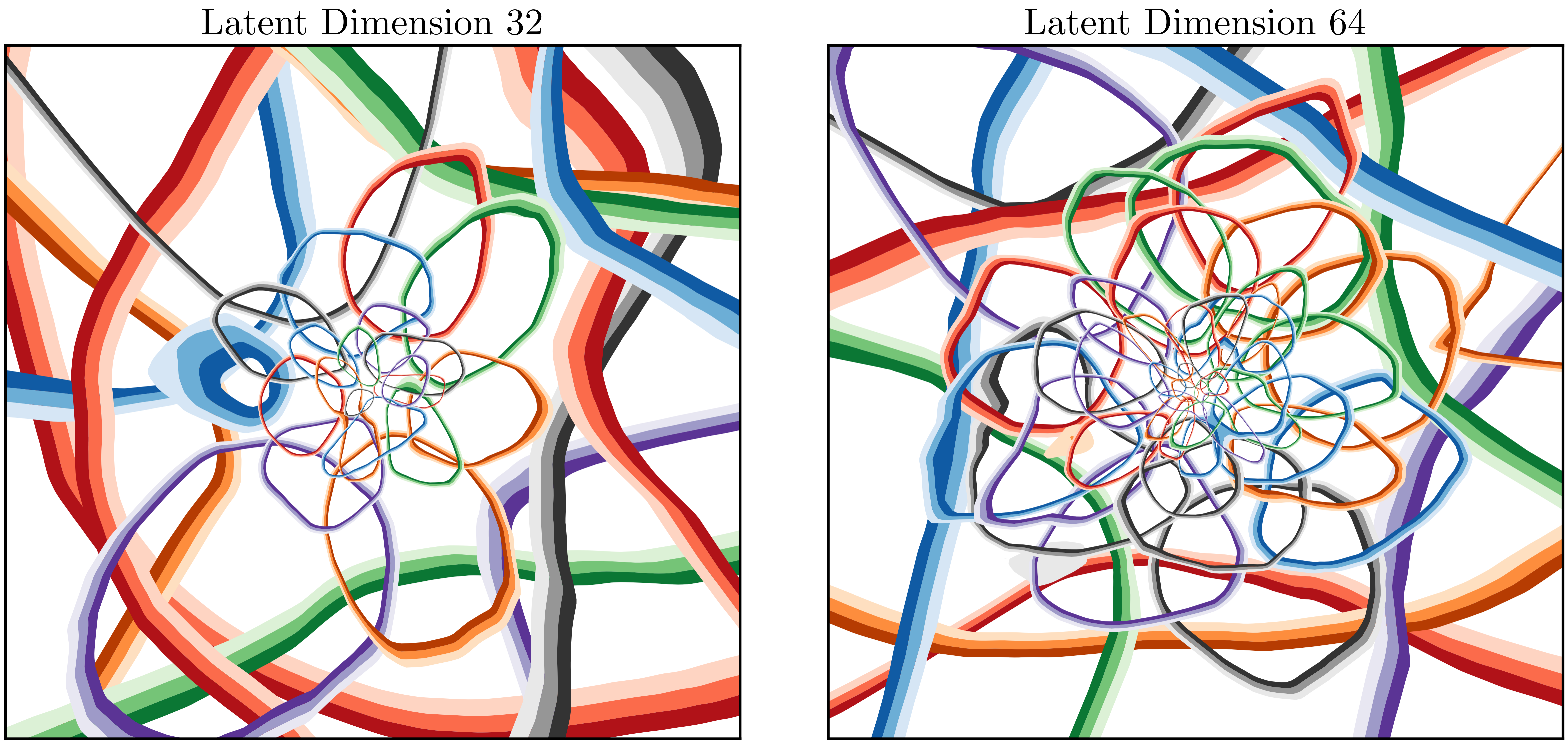
[Komiske, Metodiev, JDT, JHEP 2019; see also Komiske, Metodiev, JDT, JHEP 2018; special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NIPS 2017]

Psychedelic Network Visualization



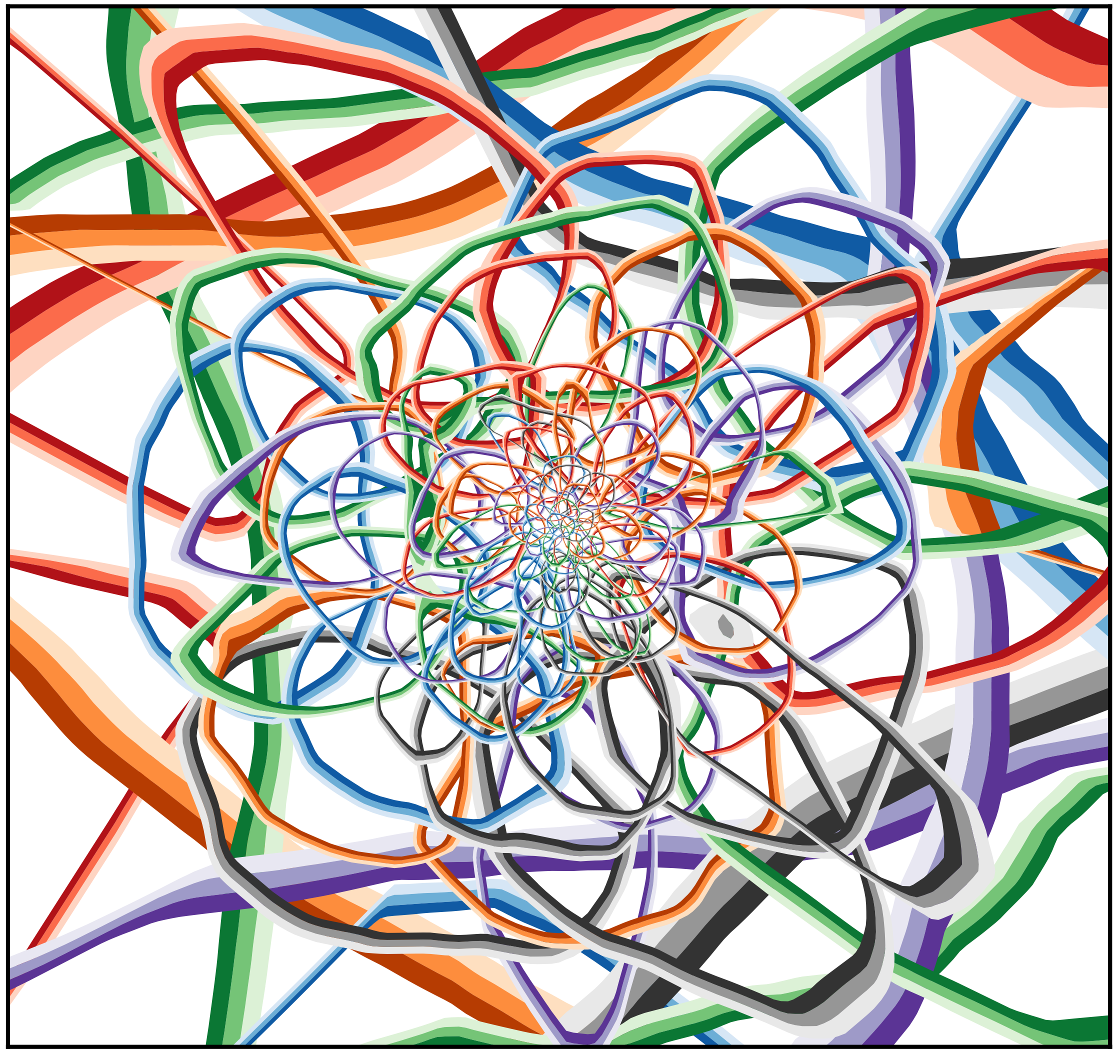
For the case of quark vs. gluon classification

Psychedelic Network Visualization

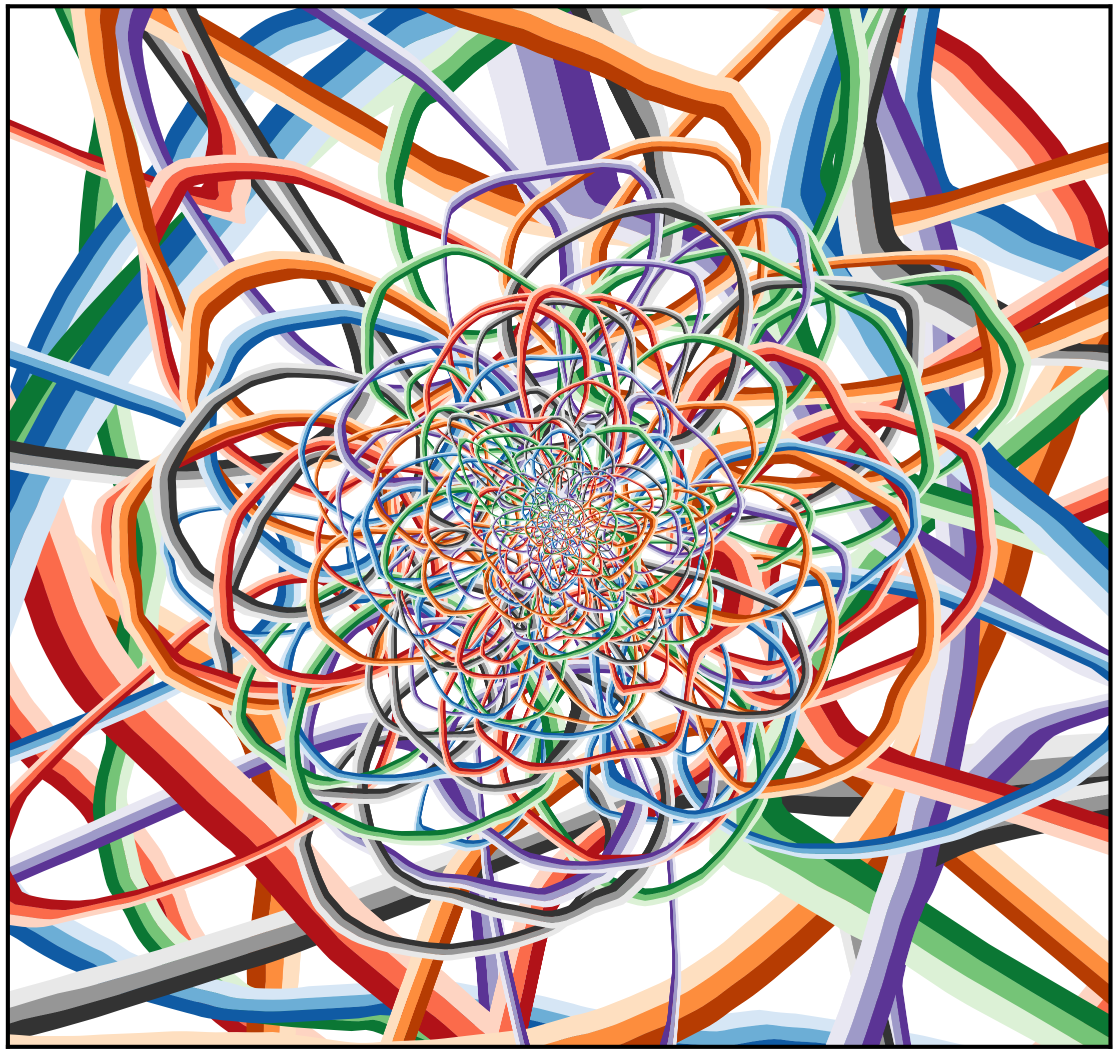


Psychedelic Network Visualization

Latent Dimension 128

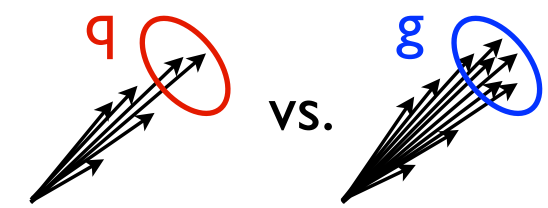


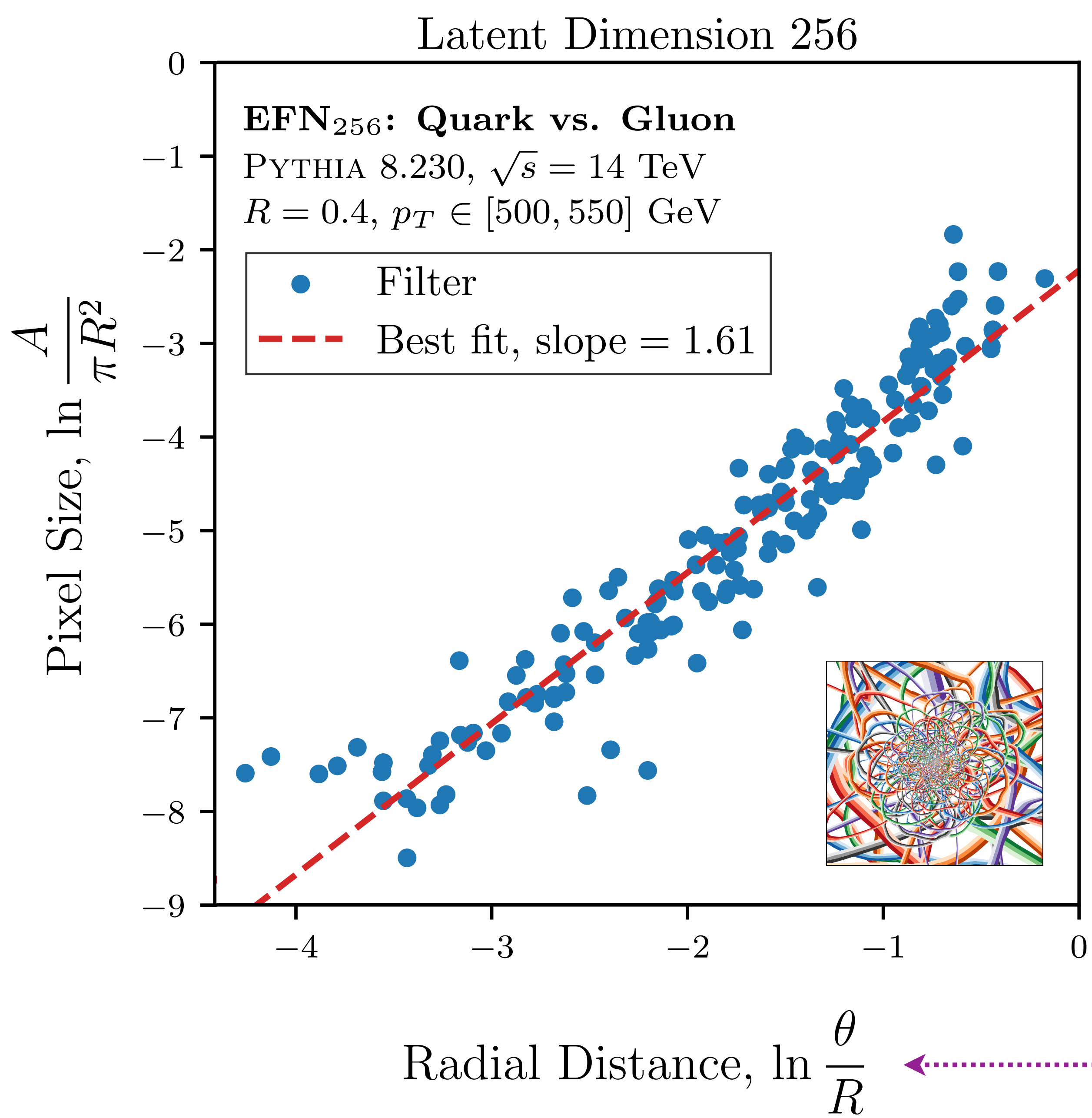
Latent Dimension 256

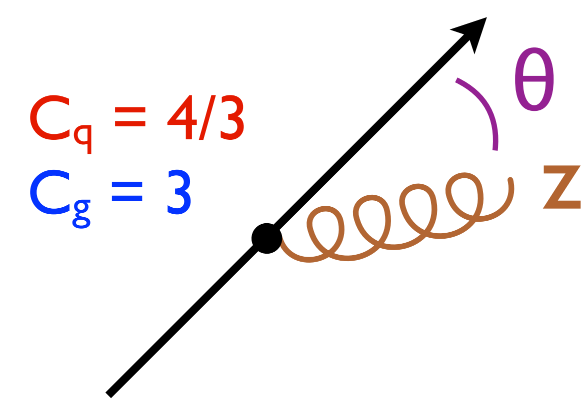


Singularity structure of QCD!

Putting the AI in AItarelli-Parisi



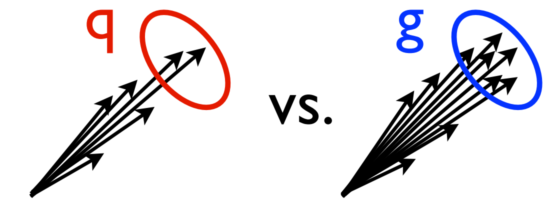




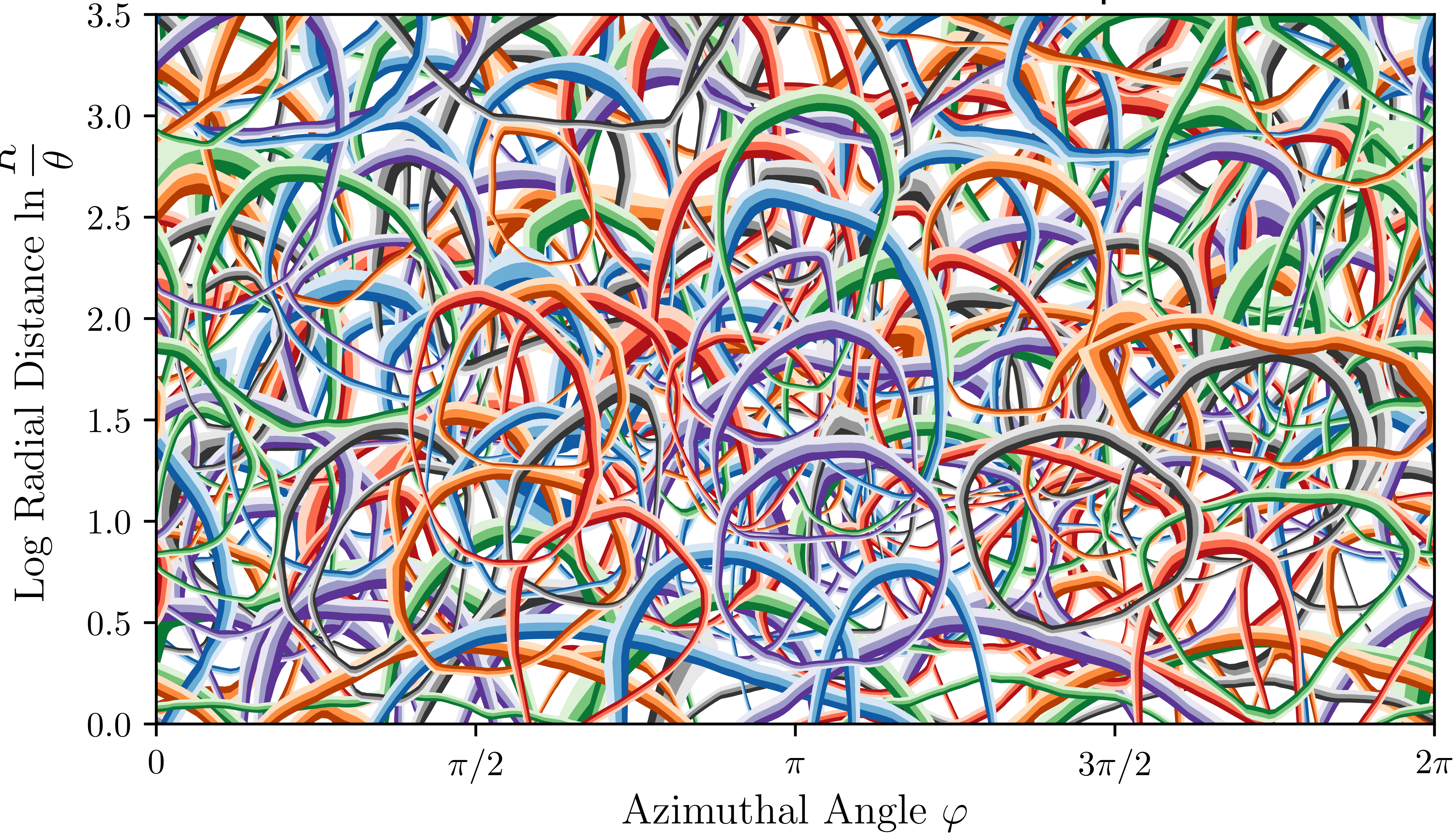
$$dP_{i \to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{d\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{dz}{z}}_{\text{Soft}}$$

[Komiske, Metodiev, JDT, JHEP 2019]

Ready for the MOCAD?



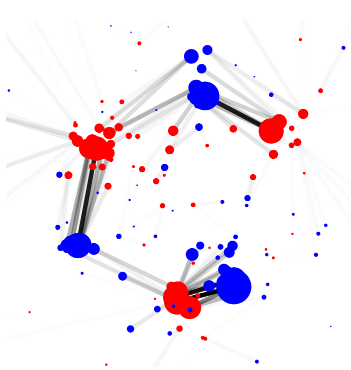
Coordinate transformation to the emission plane



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

Particle Physics 101

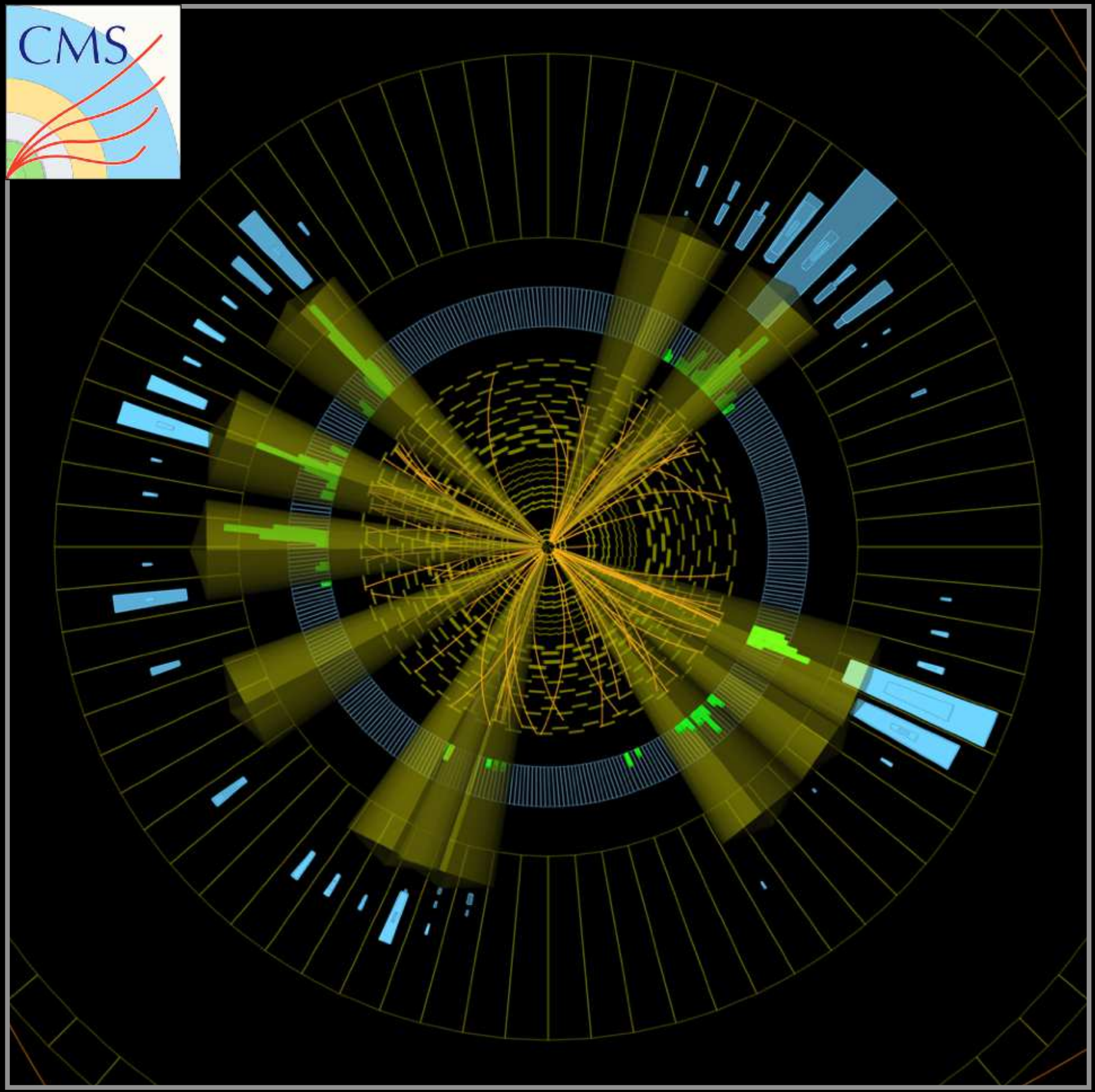
What is a Collider Event?

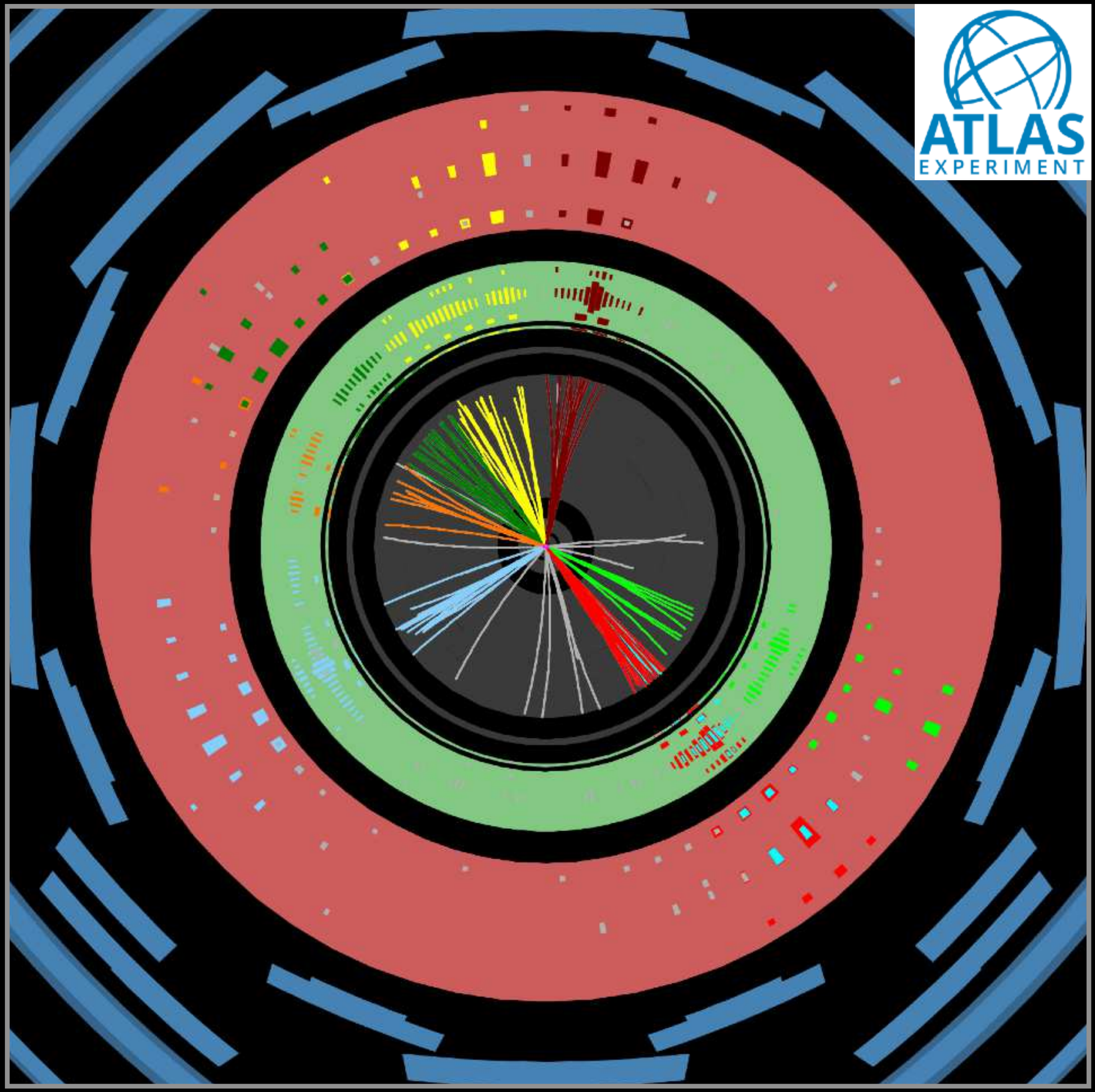


When are Collider Events Similar?

Two Collider Events

Two collections of points in (momentum) space

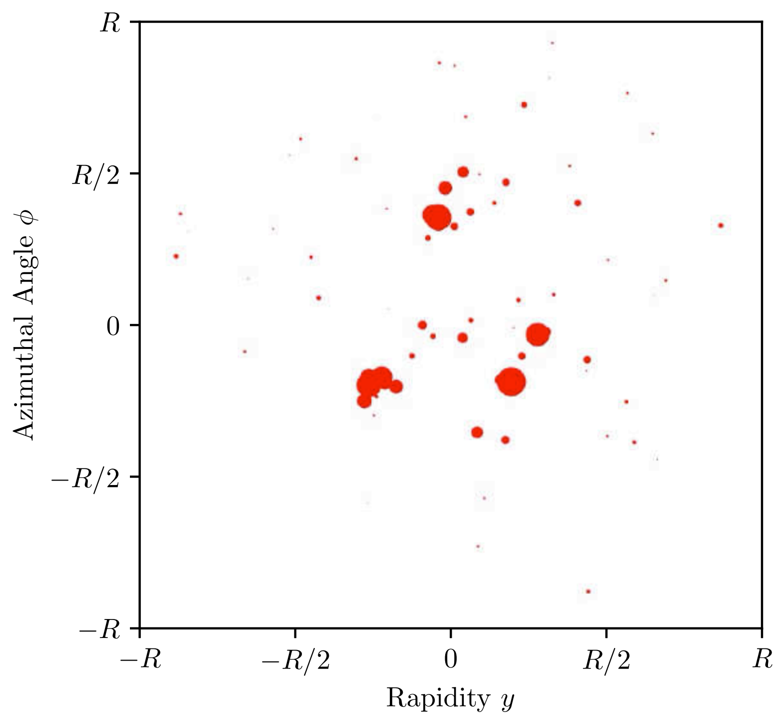




How "close" are these? (8.5 km?)

Similarity of Two Energy Flows?

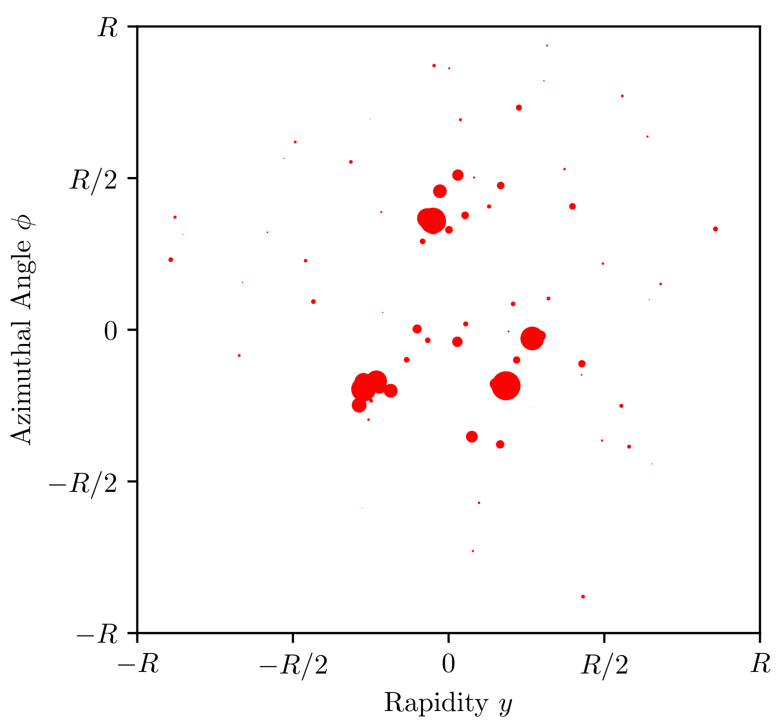
$$\rho(\hat{n}) = \sum_{i \in \mathcal{J}} E_i \, \delta^{(2)}(\hat{n} - \hat{n}_i)$$

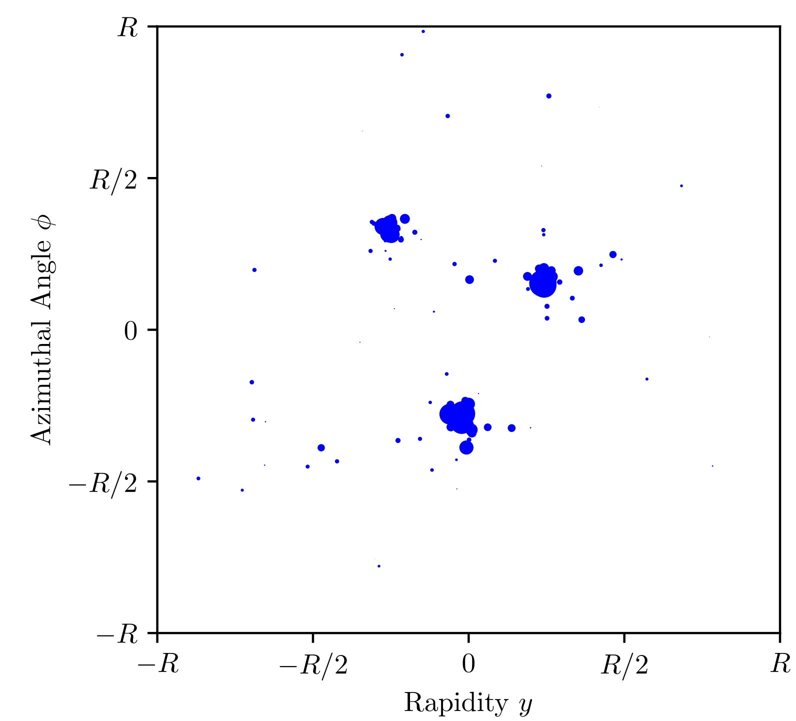


Optimal Transport:

Earth Mover's Distance
a.k.a. 1-Wasserstein metric

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

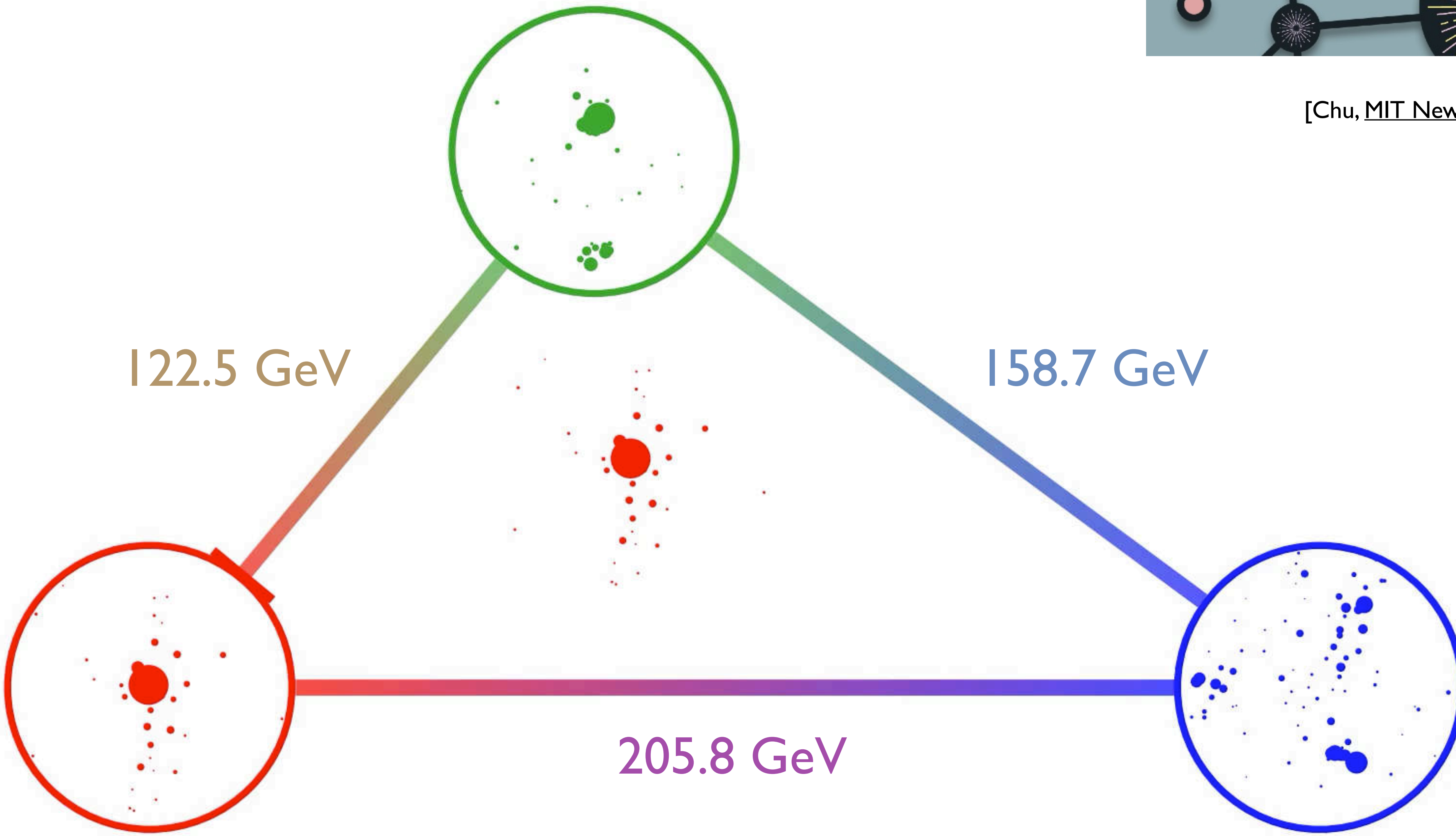




[Komiske, Metodiev, JDT, PRL 2019; code at Komiske, Metodiev, JDT, energyflow.network]

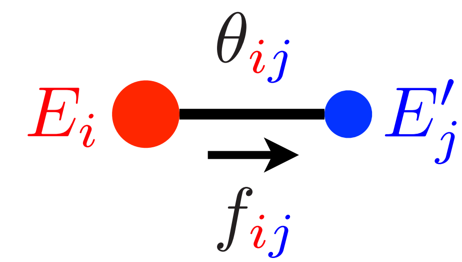
Similarity of Three Energy Flows?

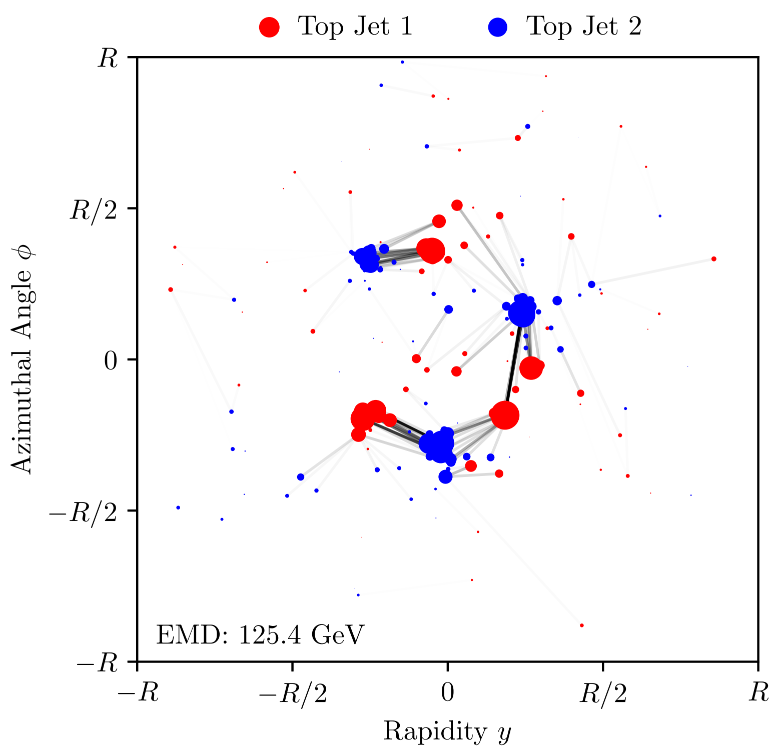
[Chu, MIT News July 2019]



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

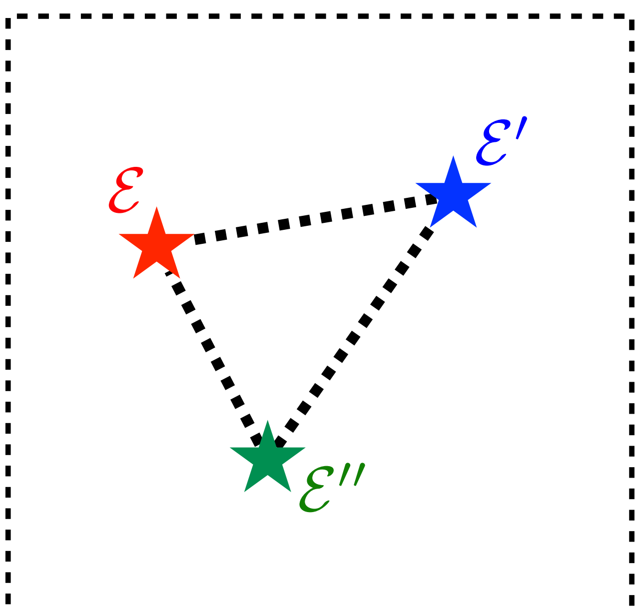
The Energy Mover's Distance





Optimal transport between energy flows…

$$\underset{\text{in GeV}}{\mathrm{EMD}(\mathcal{E},\mathcal{E}')} = \underbrace{\min_{\{f\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R}}_{\text{Cost to move energy}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost to create energy}}$$



…defines a metric on the space of events

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

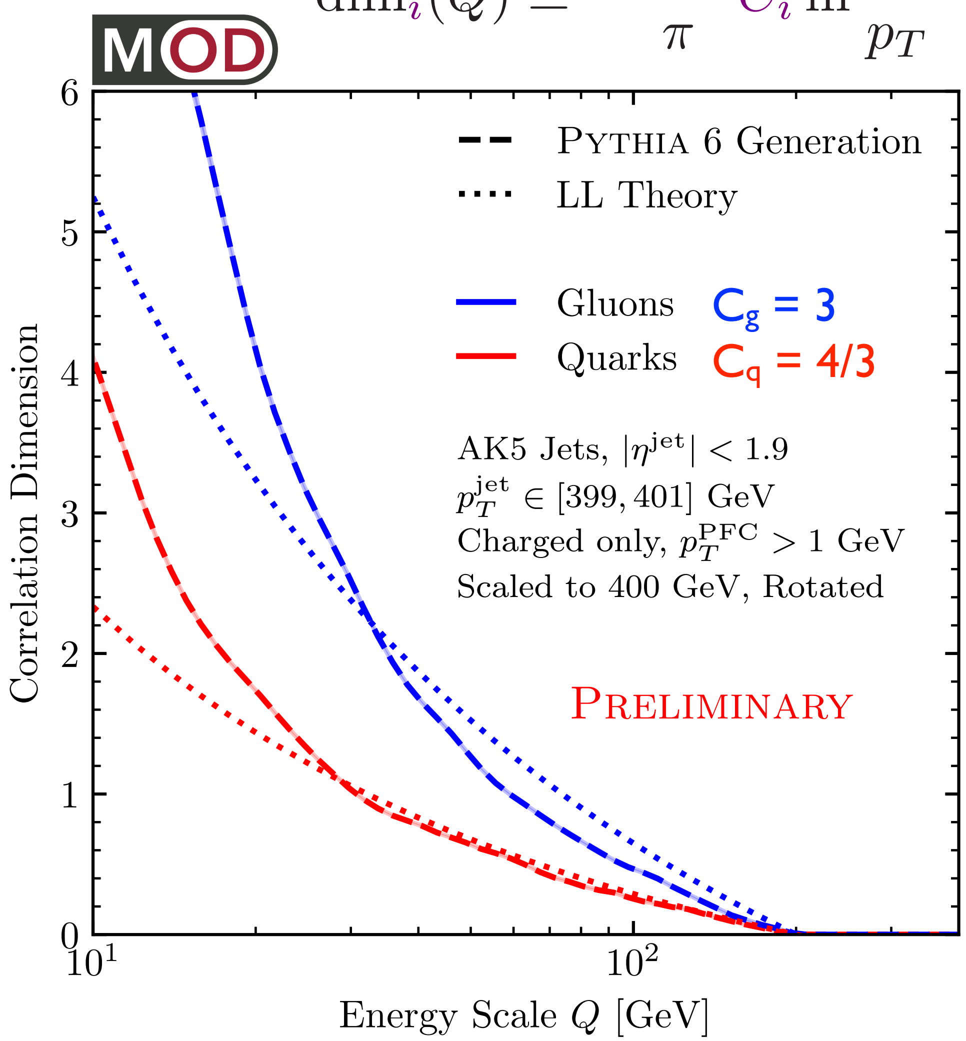
(assuming R ≥ θ_{max}/2, i.e. R ≥ jet radius for conical jets)

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

Dimensionality of Space of Jets

QCD Calculation

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

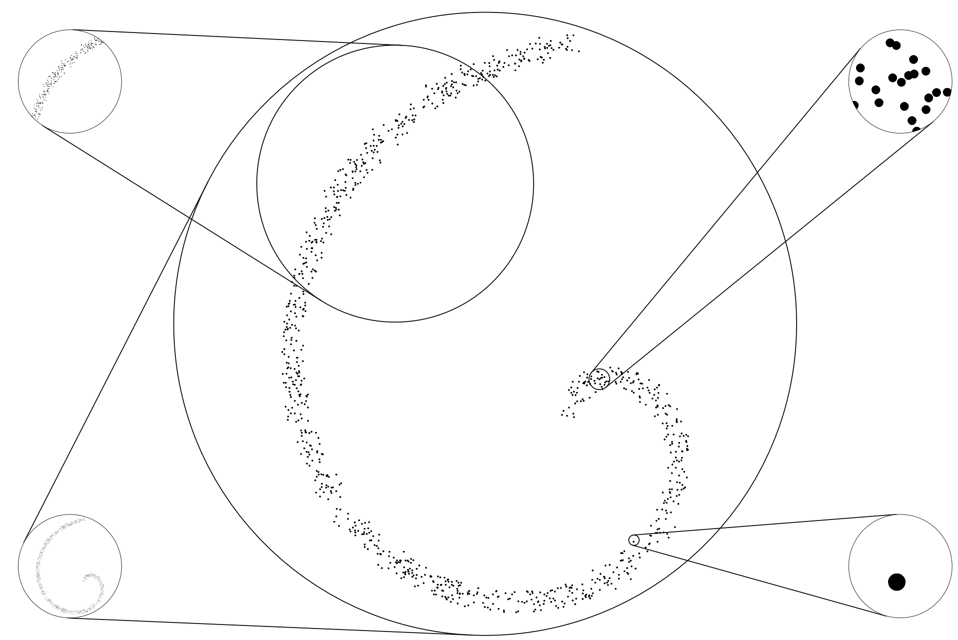


$$N_{\rm neighbors}(r) \sim r^{\dim}$$

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

dim≃1

dim≃2



dim≃2

(eventually 0)

dim≃0

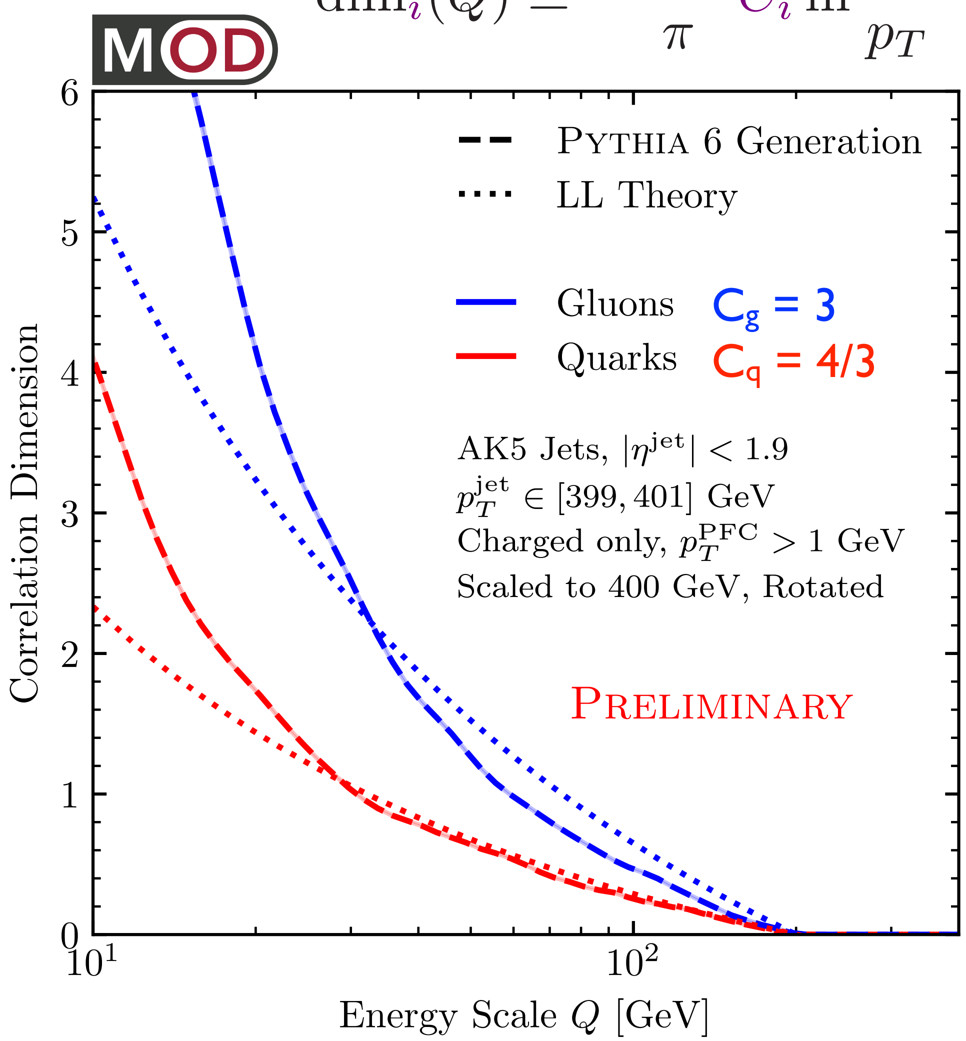
[Grassberger, Procaccia, PRL 1983; Kégl, NIPS 2002]

Dimensionality of Space of Jets

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

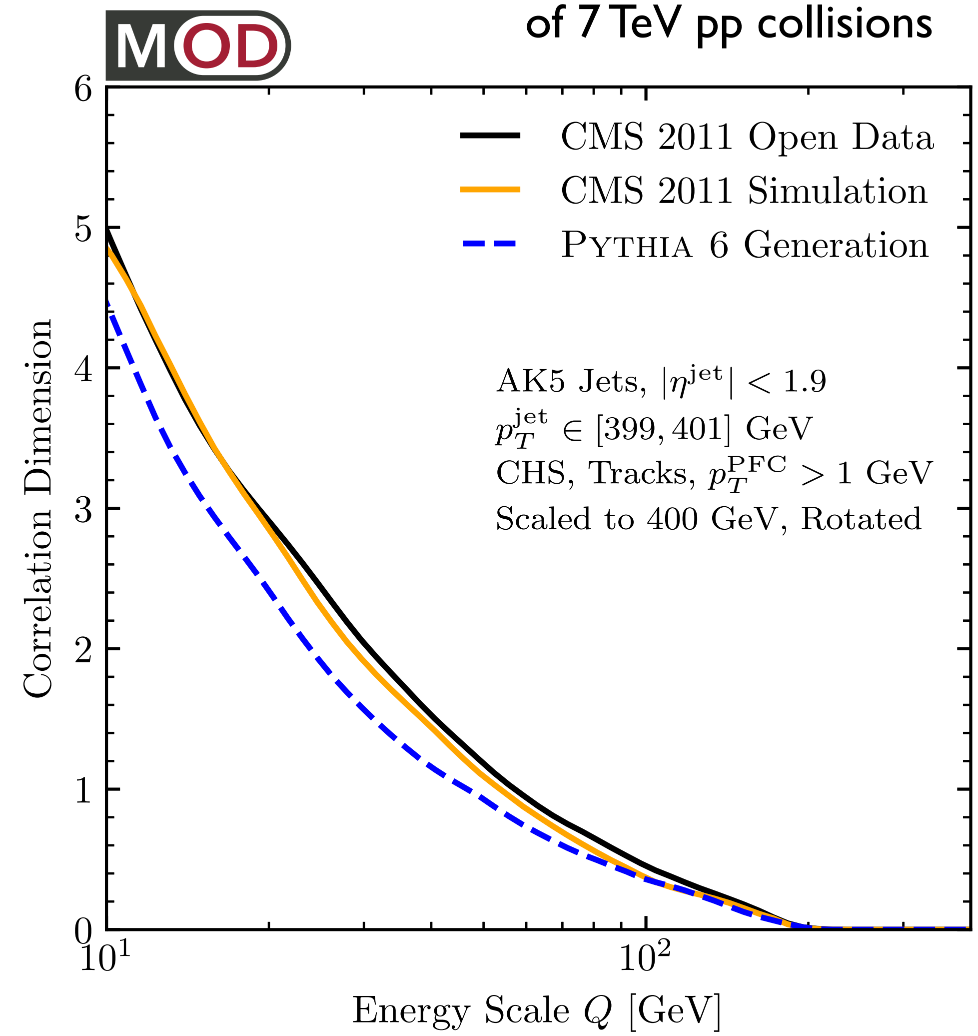
QCD Calculation

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



CMS Open Data

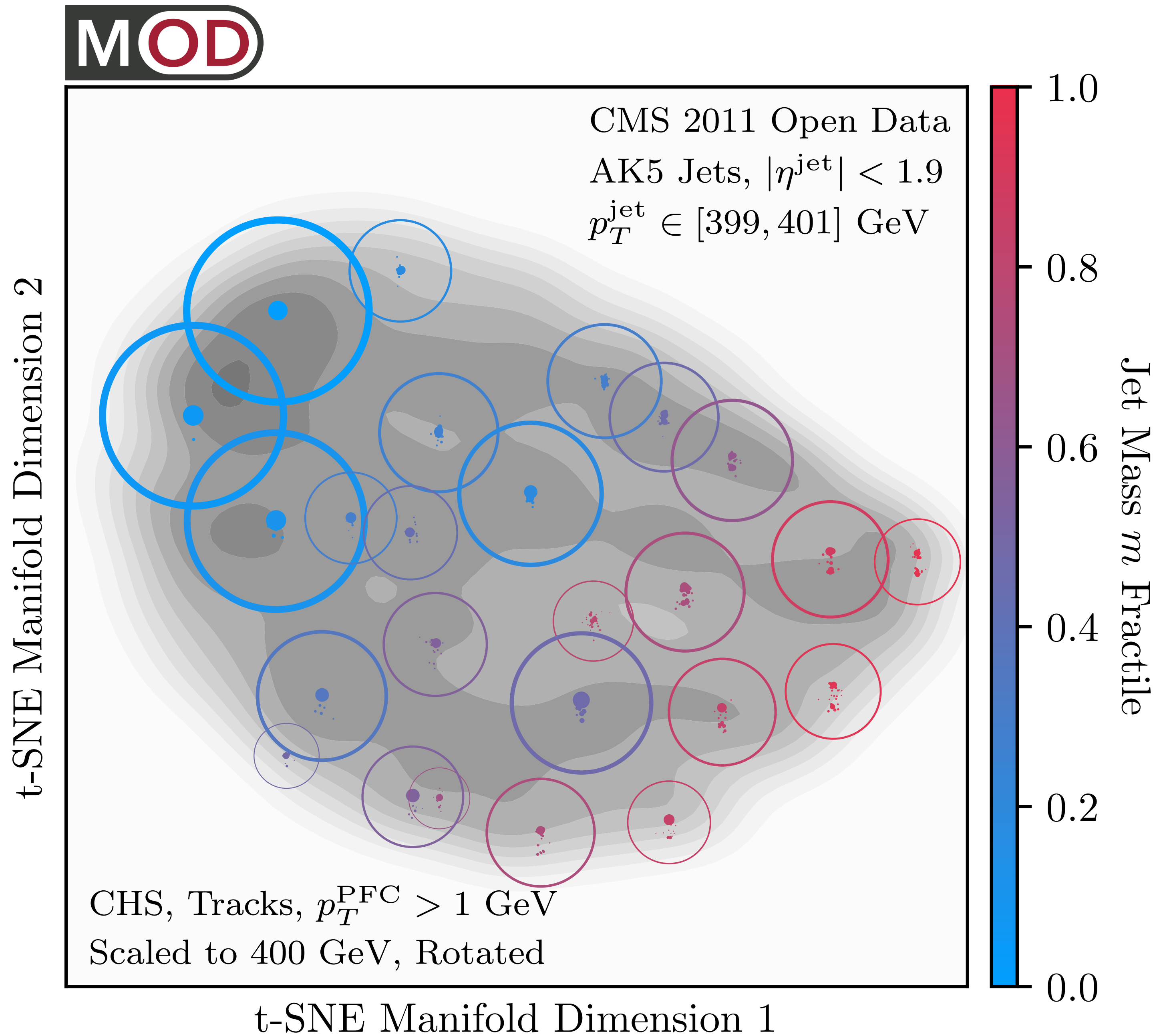
400 GeV jets in 2.3 fb^{-1} of 7 TeV pp collisions



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

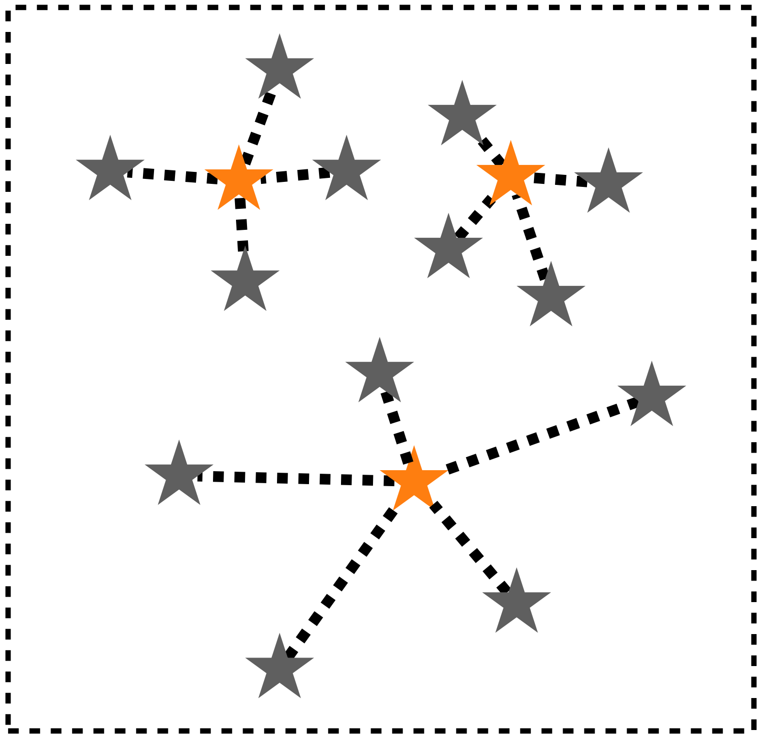
Most Representative Jets

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



k-medoids

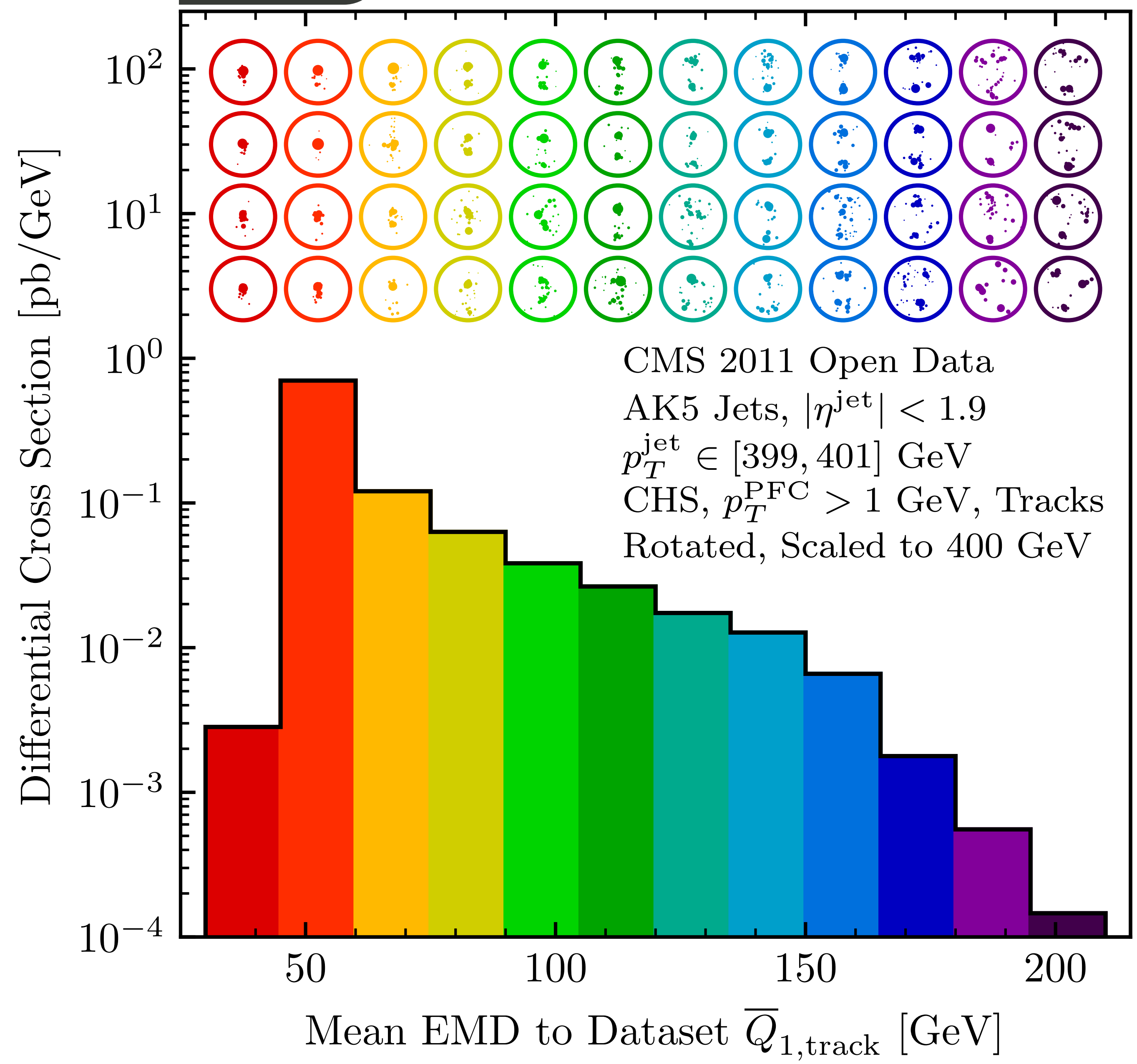
Arranged via t-SNE



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

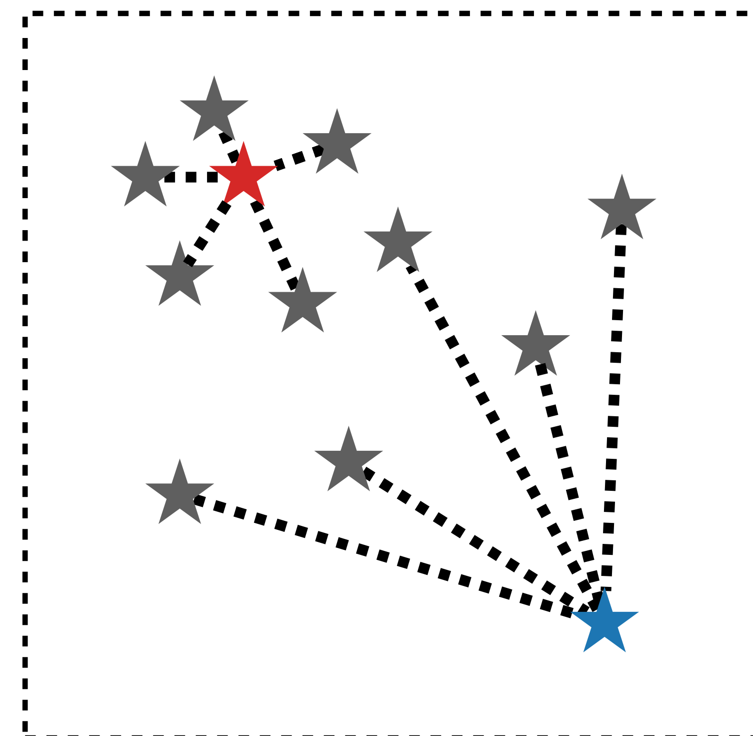
Least Representative Jets

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



New Physics?

Or tails of QCD?



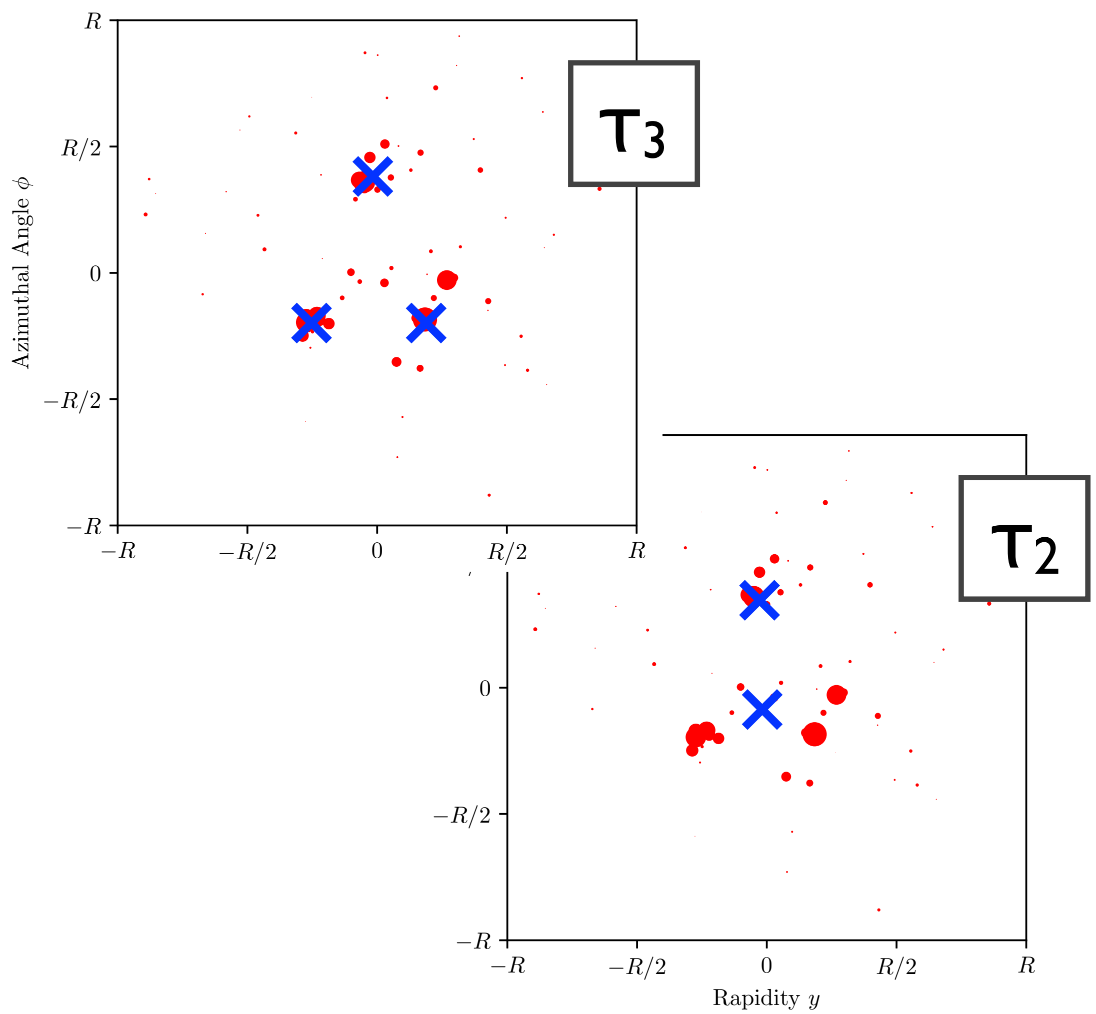
[Komiske, Mastandrea, Metodiev, Naik, JDT, submitted to PRD]

N-subjettiness

Ubiquitous jet substructure observable used for almost a decade…

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

IRC safe



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

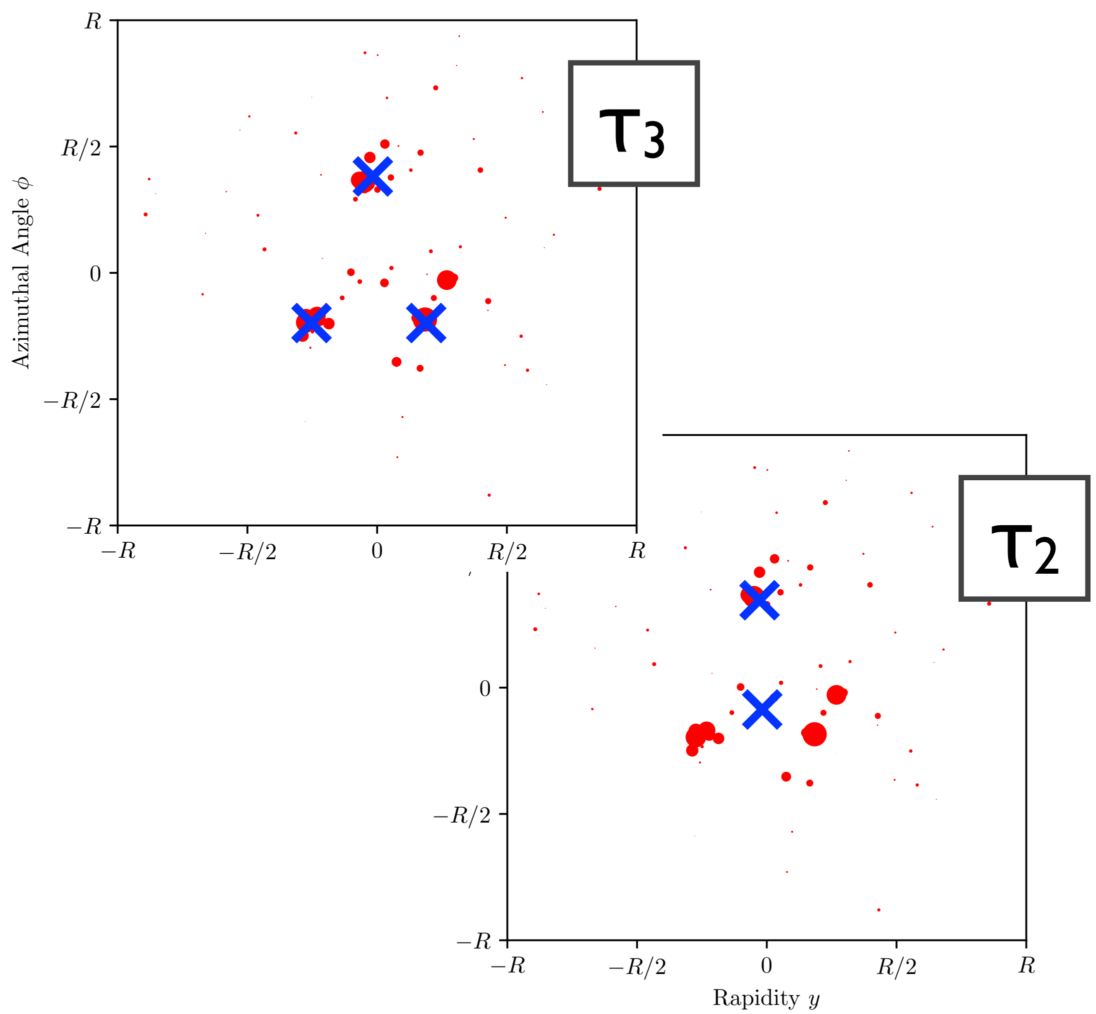
N-subjettiness = Point to Manifold EMD

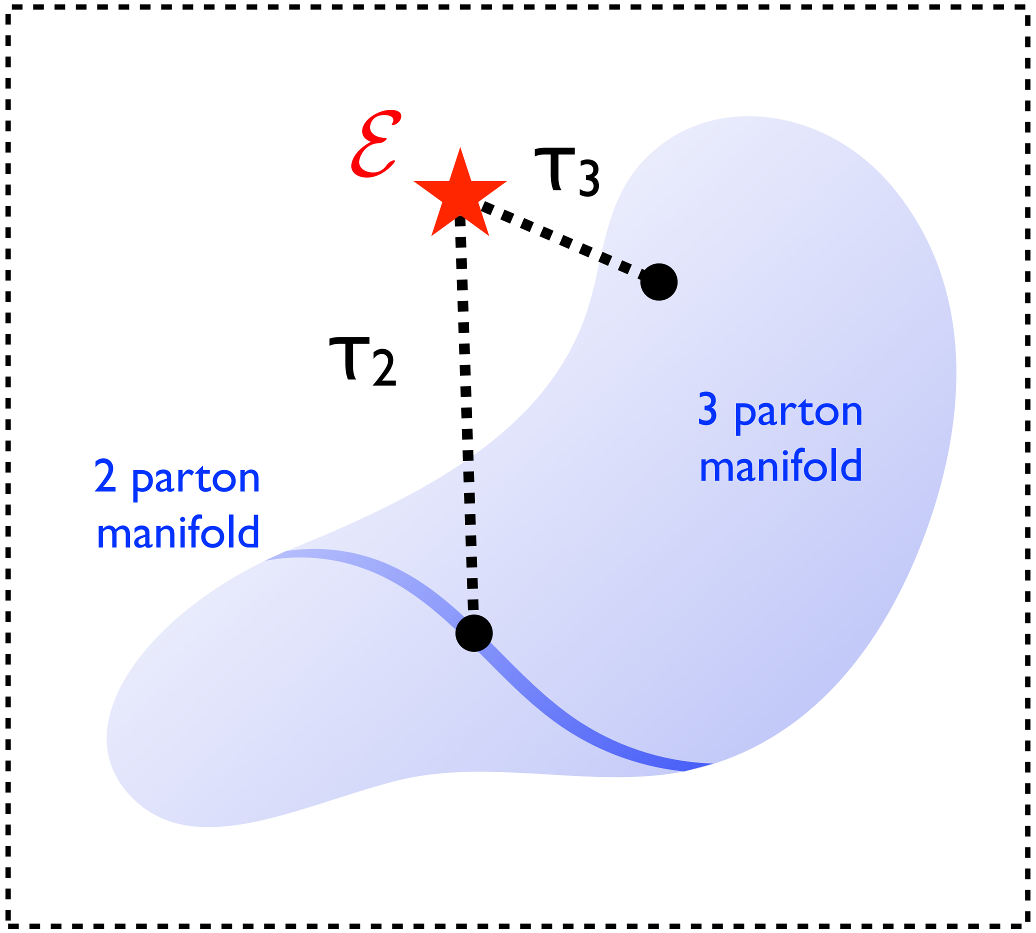
…is secretly an optimal transport problem

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

IRC safe

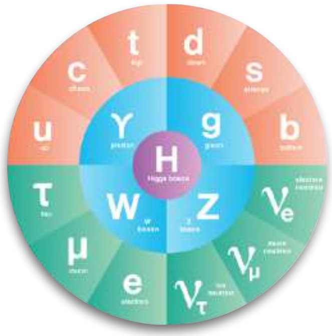
(!)





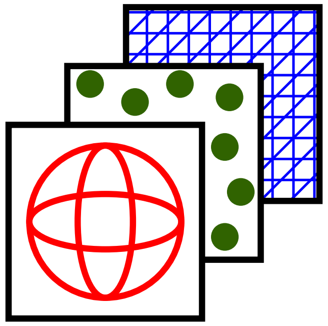
[JDT, Van Tilburg, JHEP 2011, JHEP 2012;
rephrased in the language of Komiske, Metodiev, JDT, PRL 2019]

Summary



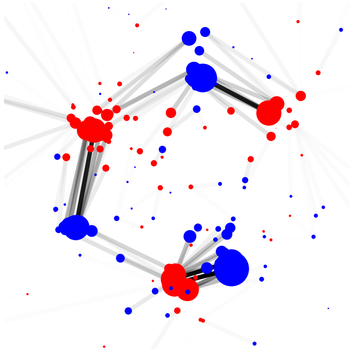
Particle Physics 101

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



What is a Collider Event?

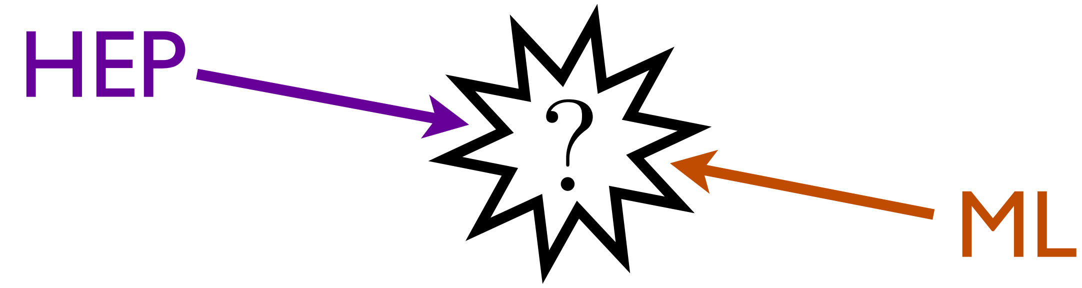
Unordered set of particles describing energy flow of jets
Inspires network architectures designed for symmetry and safety



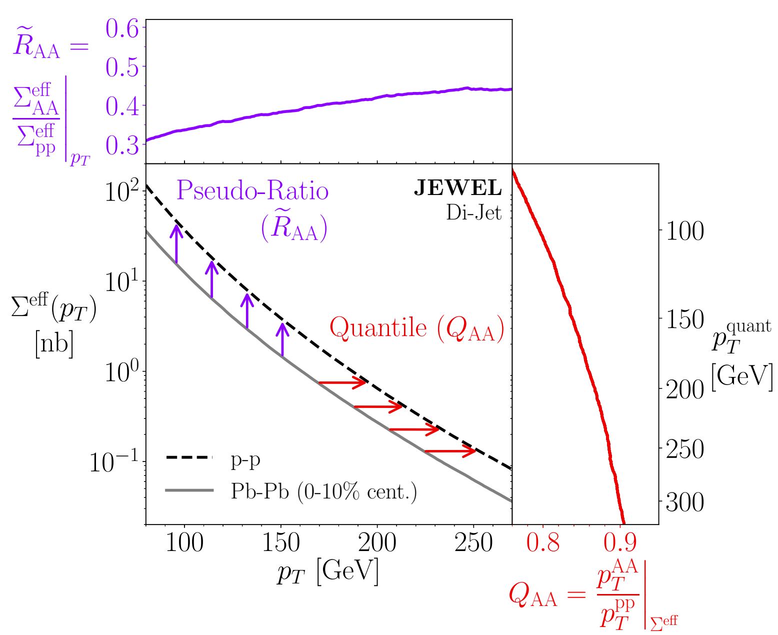
When are Collider Events Similar?

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

More Collisions

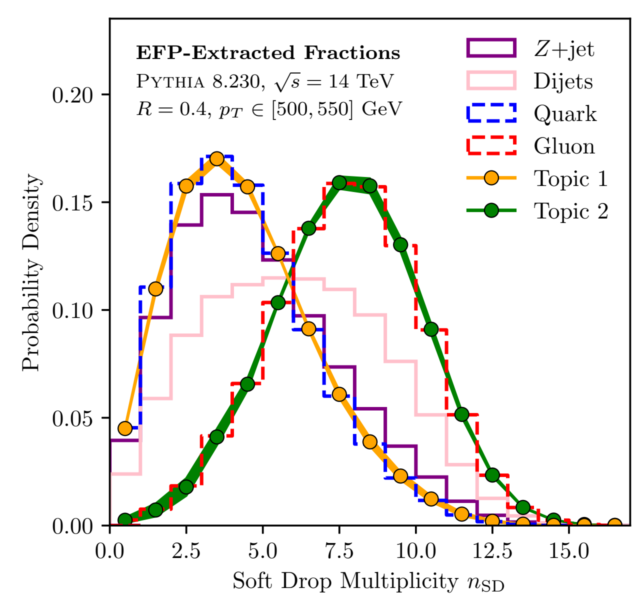


Jet Quenching via Optimal Transport



[Brewer, Milhano, JDT, PRL 2019]

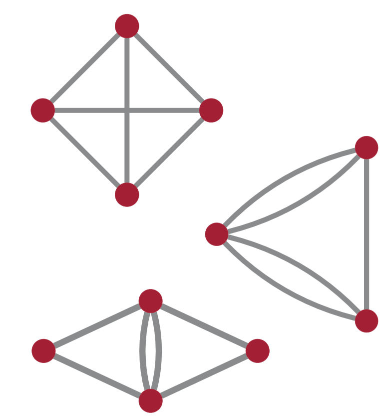
Unsupervised Jet Classification via Blind Source Separation



[Komiske, Metodiev, JDT, JHEP 2018]

Kinematic Decomposition via Graph Theory

	Leafless Multigraphs	
	Connected	All
Edges d	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**

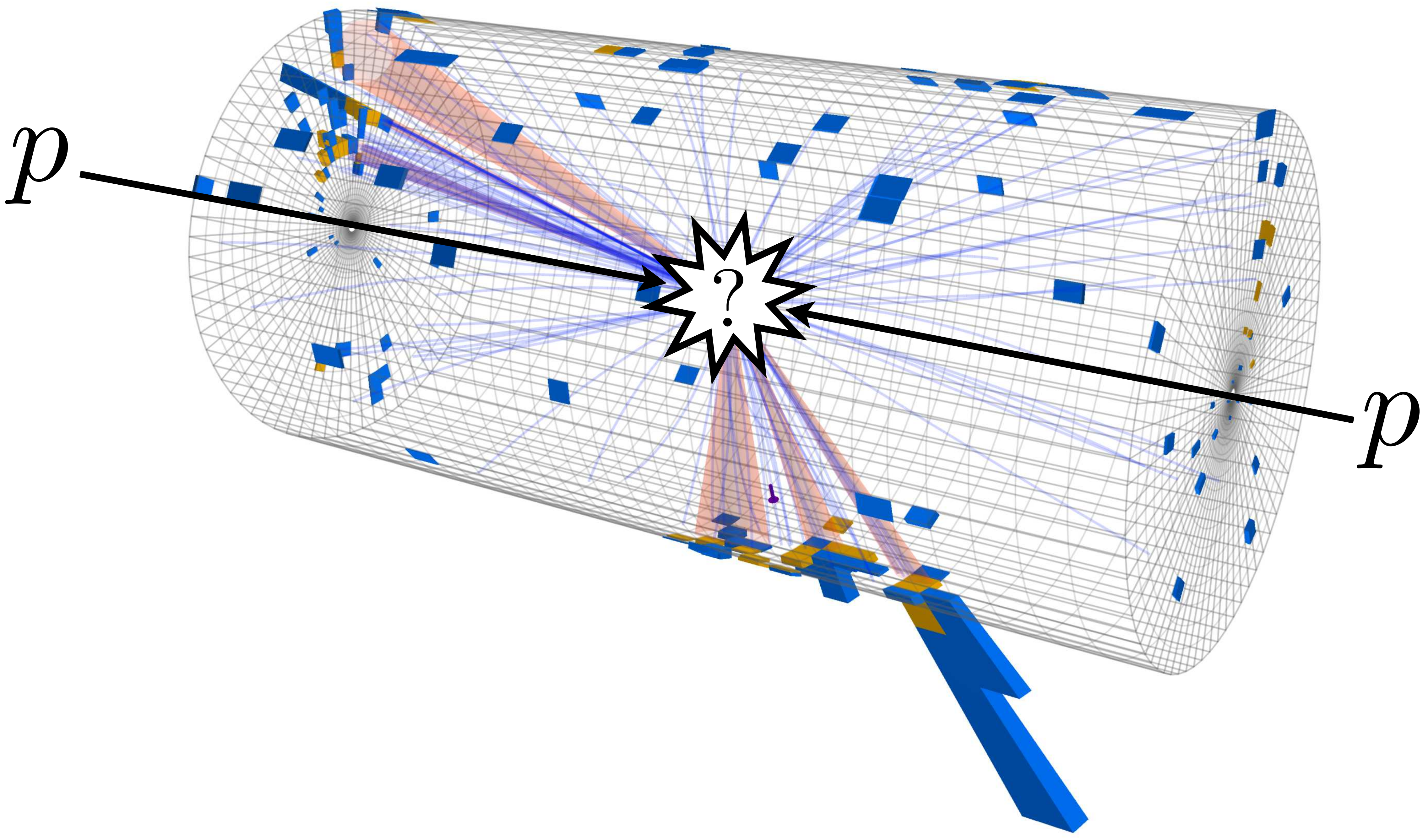


[Komiske, Metodiev, JDT, to appear]

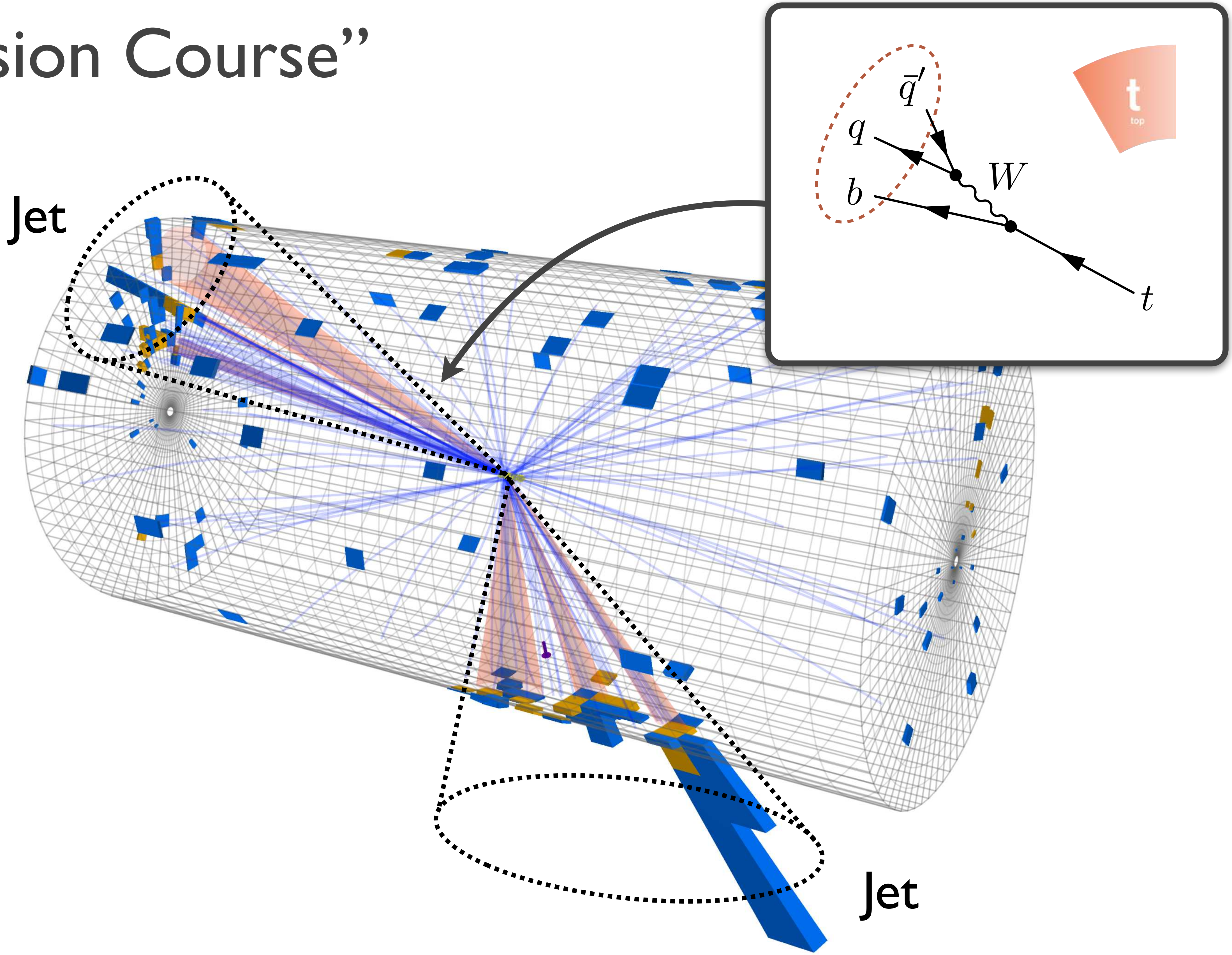
*New insights into particle physics**
*facilitated by advances in machine learning**

Backup Slides

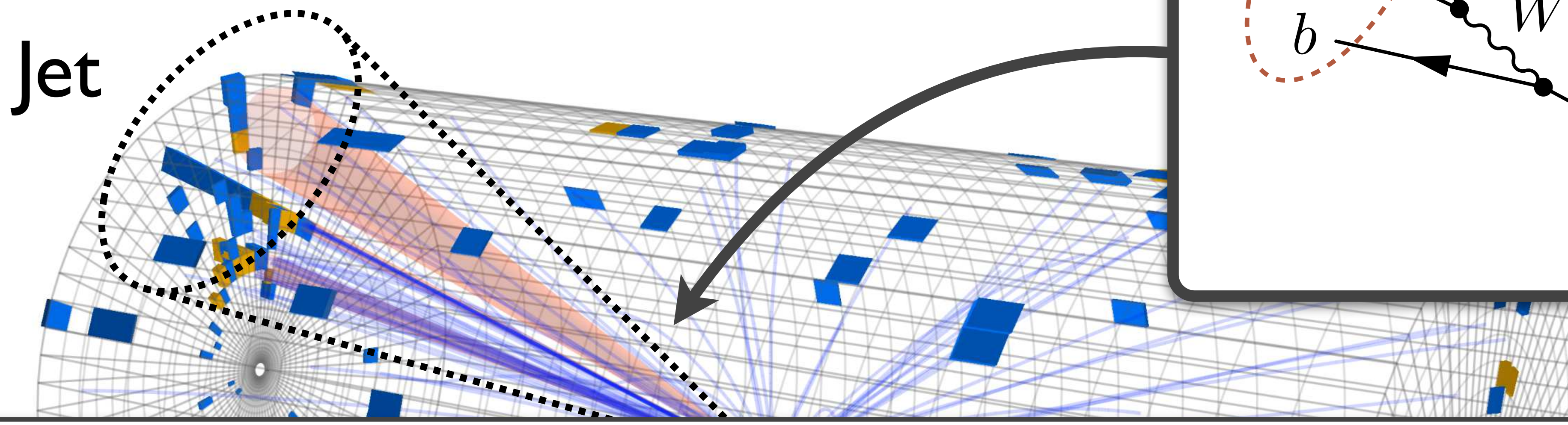
“Collision Course”

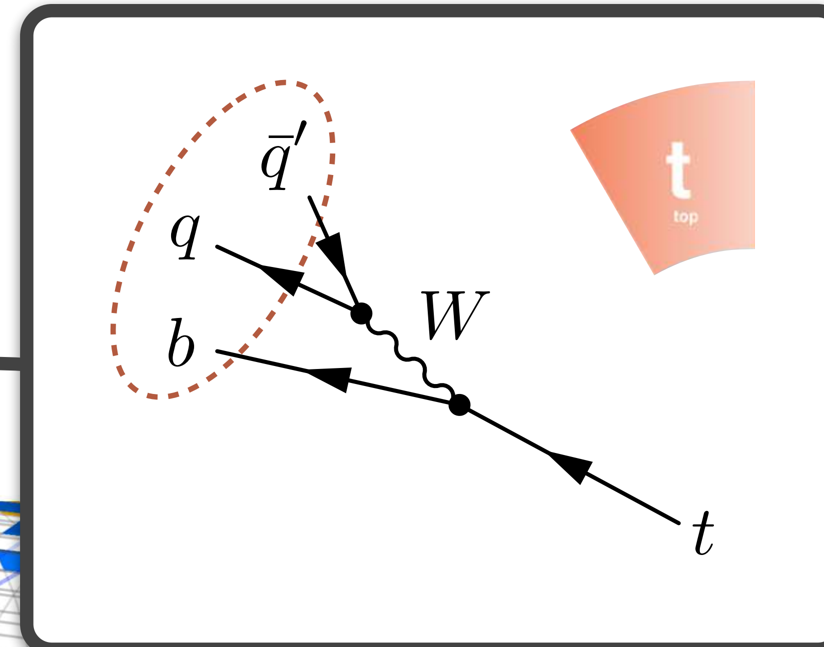


“Collision Course”



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"?

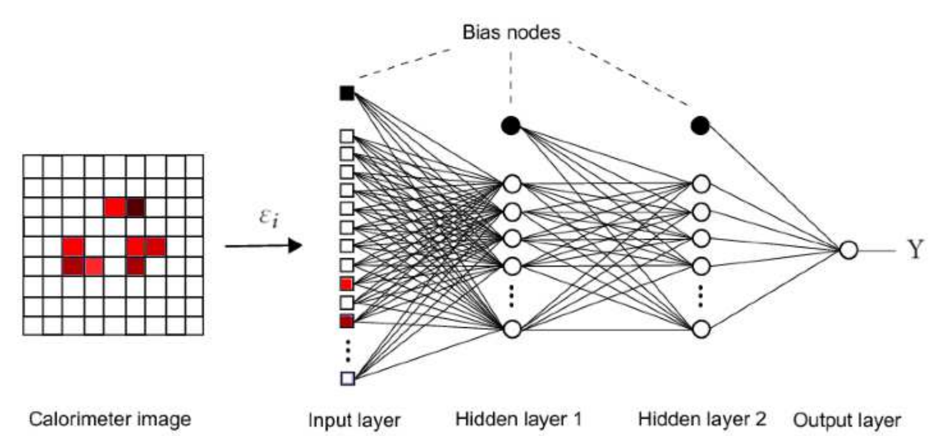
$$\tau_N(\mathcal{J}) = \min_{|\mathcal{J}'|=N} \text{EMD}(\mathcal{J}, \mathcal{J}')$$

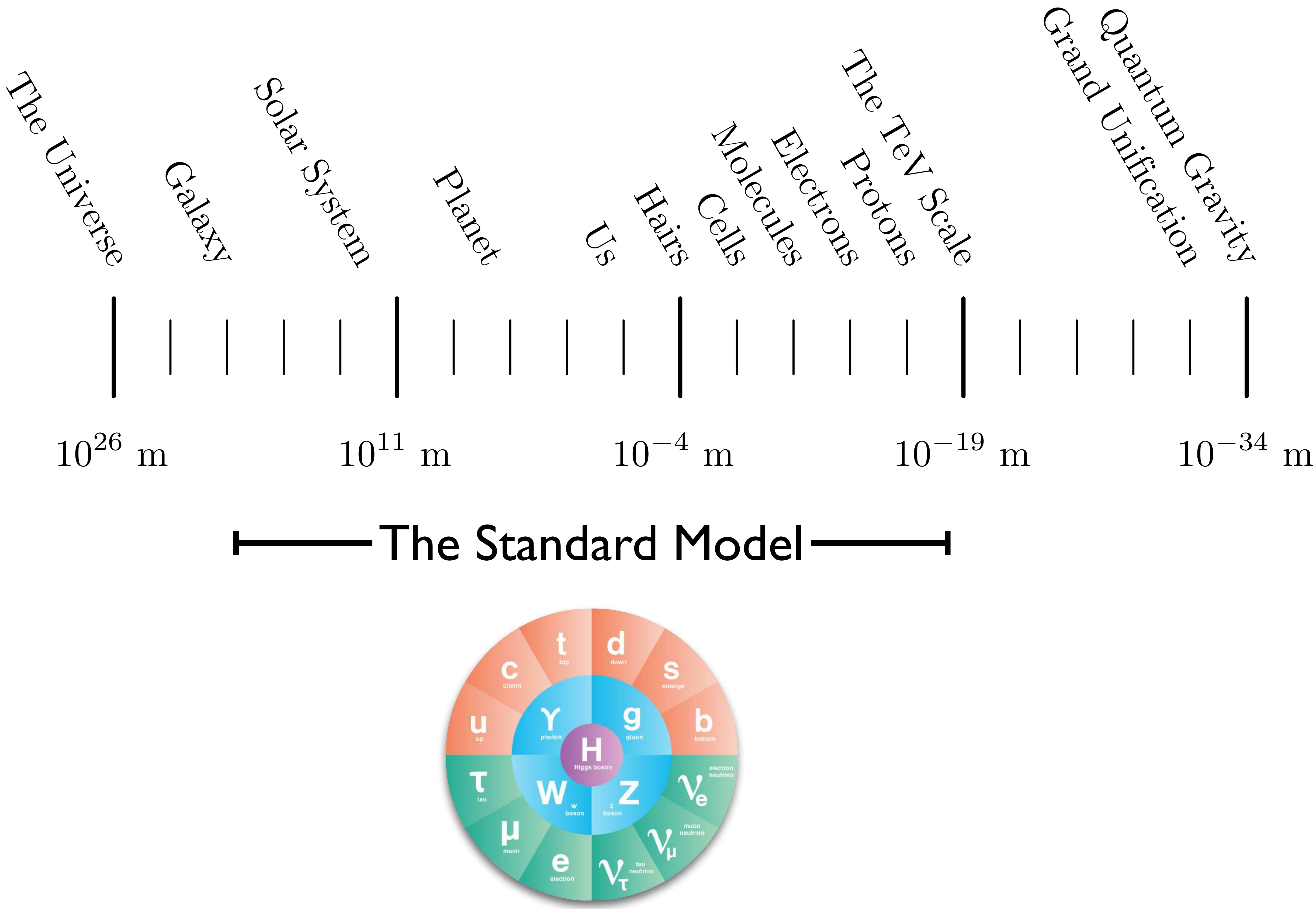
[e.g. JDT, Van Tilburg, 1011.2268, 1108.2701;
rephrased in language of Komiske, Metodiev, JDT, 1902.02346]

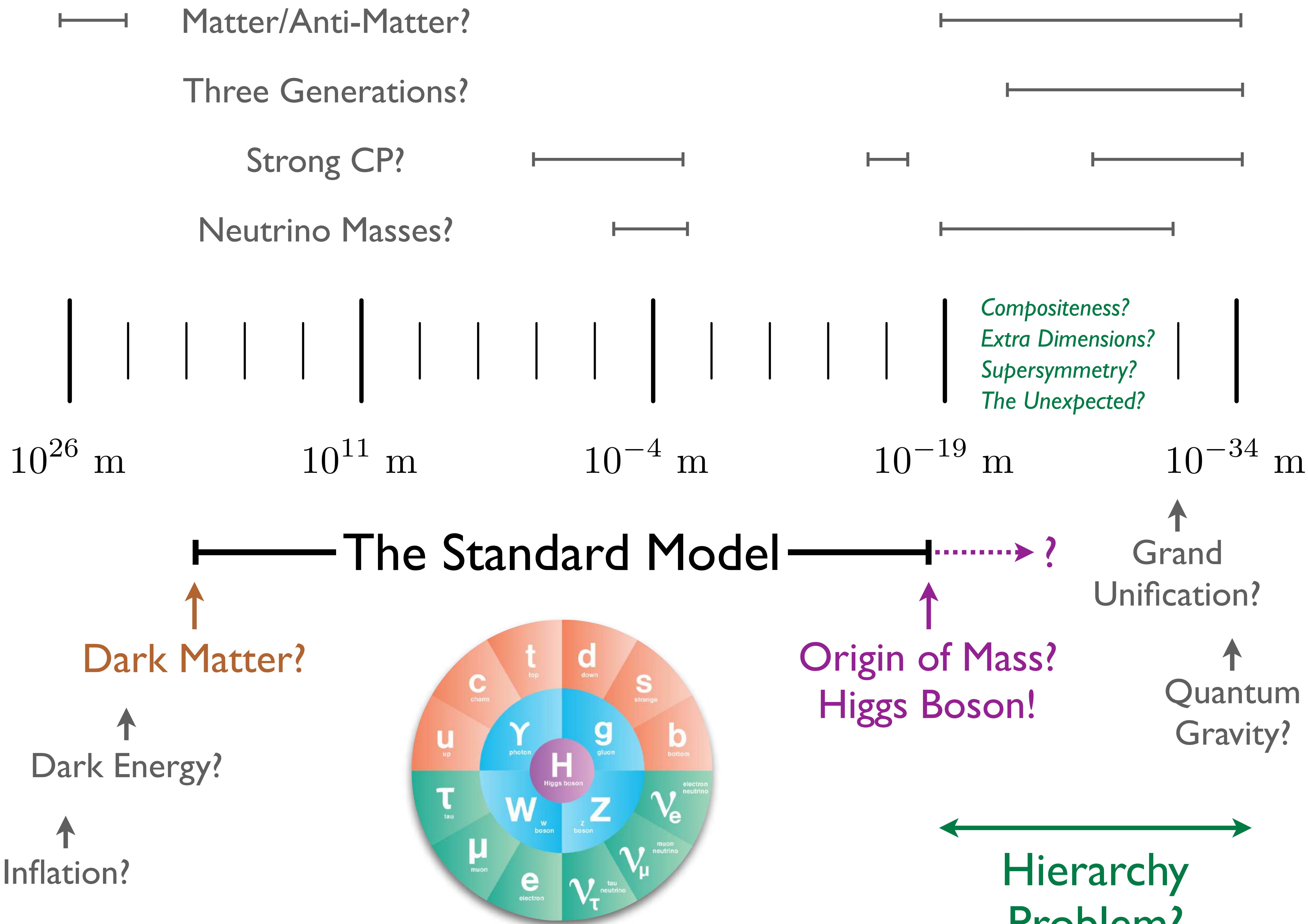
Ask me about
this formula…

"Deep Learning"?

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

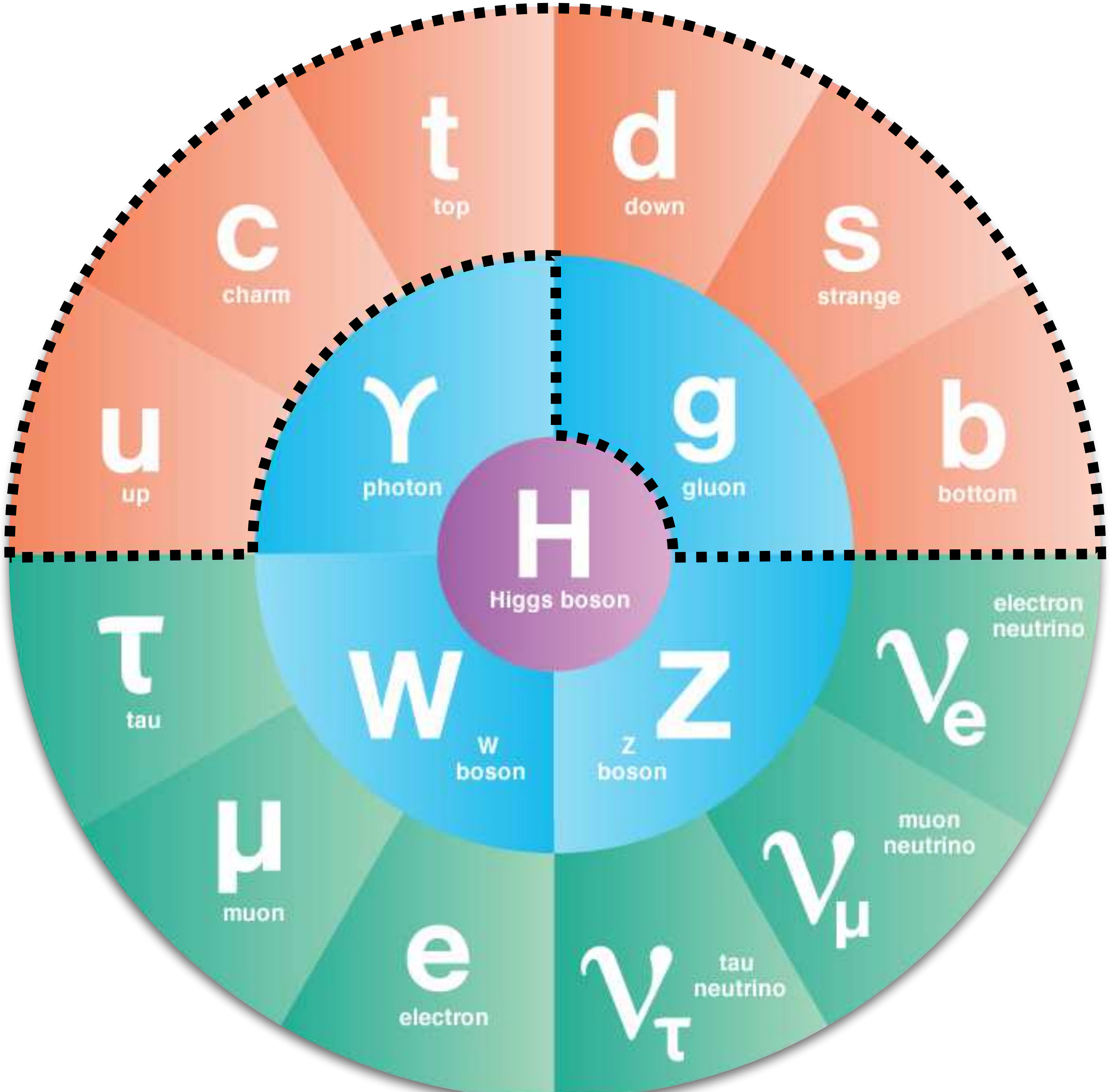


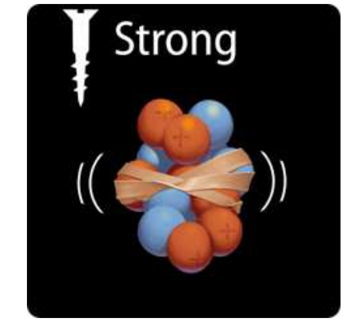

The Universe
Galaxy
Solar System
Planet
Us
Hairs
Cells
Molecules
Electrons
Protons
The TeV Scale
Grand Unification
Quantum Gravity
10^{26} m
10^{11} m
10^{-4} m
10^{-19} m
10^{-34} m
The Standard Model
t top
d down
c charm
s strange
u up
b bottom
γ photon
g gluon
H Higgs boson
τ tau
W w boson
Z z boson
ν_e electron neutrino
μ muon
ν_μ muon neutrino
e electron
ν_τ tau neutrino



Matter/Anti-Matter?
Three Generations?
Strong CP?
Neutrino Masses?
Compositeness?
Extra Dimensions?
Supersymmetry?
The Unexpected?
10^{26} m
10^{11} m
10^{-4} m
10^{-19} m
10^{-34} m
The Standard Model
?
Grand Unification?
Dark Matter?
Dark Energy?
Inflation?
Origin of Mass?
Higgs Boson!
Quantum Gravity?
Hierarchy Problem?
t
top
d
down
c
charm
s
strange
u
up
b
bottom
γ
photon
g
gluon
H
Higgs boson
τ
tau
W
w boson
Z
z boson
ν_e
electron neutrino
μ
muon
ν_μ
muon neutrino
e
electron
ν_τ
tau neutrino

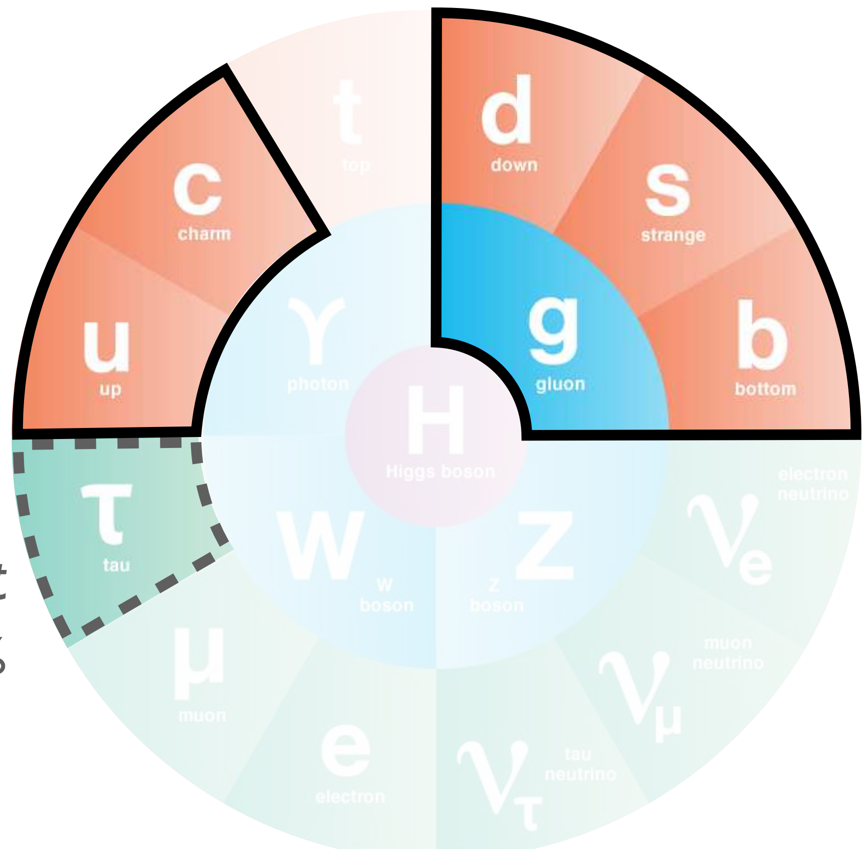
Focus of this talk:



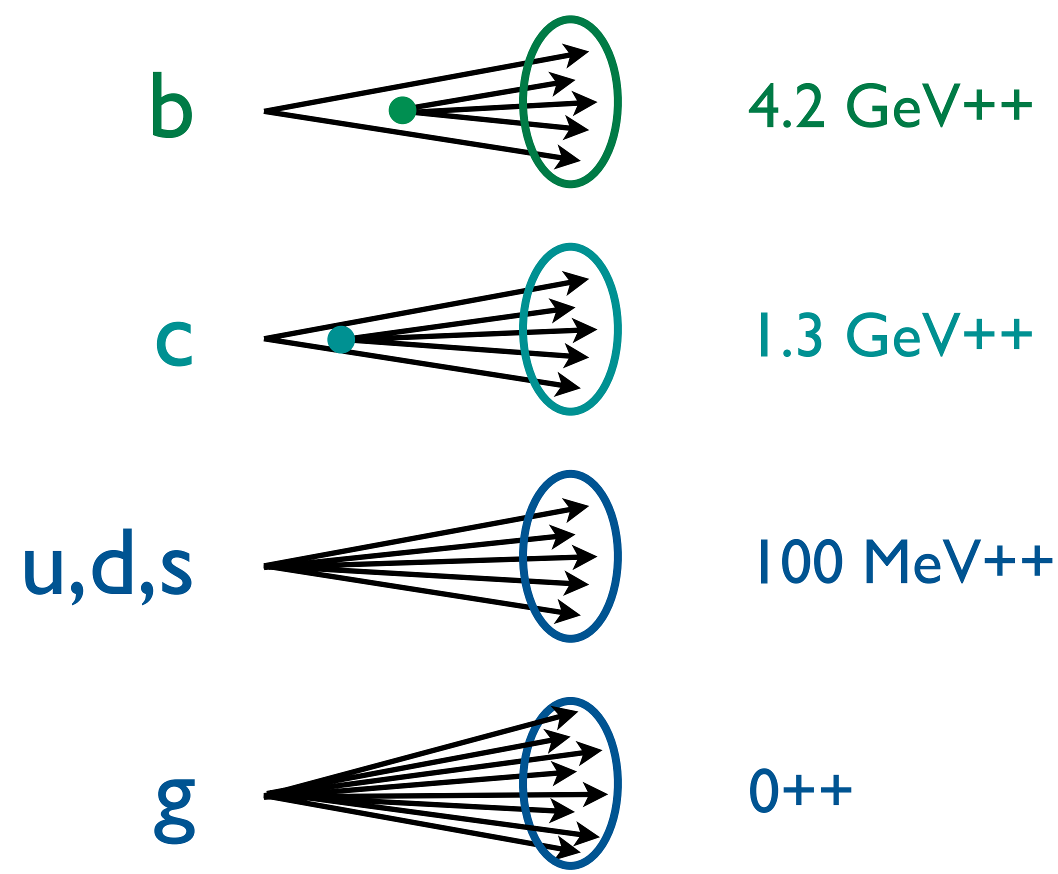


Jet Classification

Key supervised learning task at LHC



micro-jet
≈ 65%

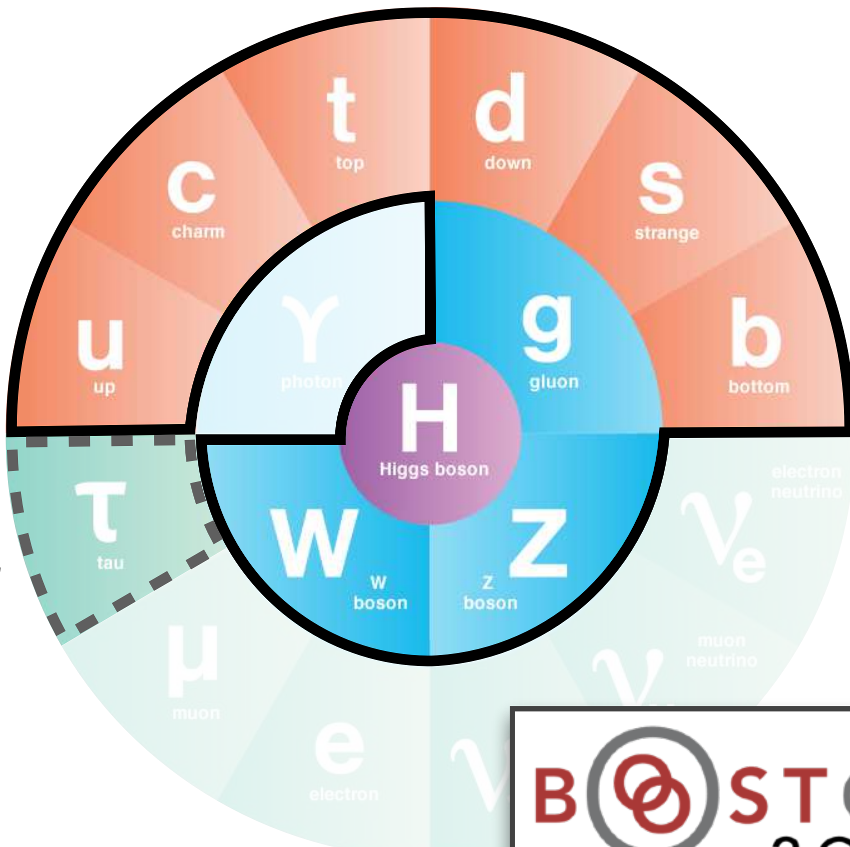


++ = Mass from QCD Radiation

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

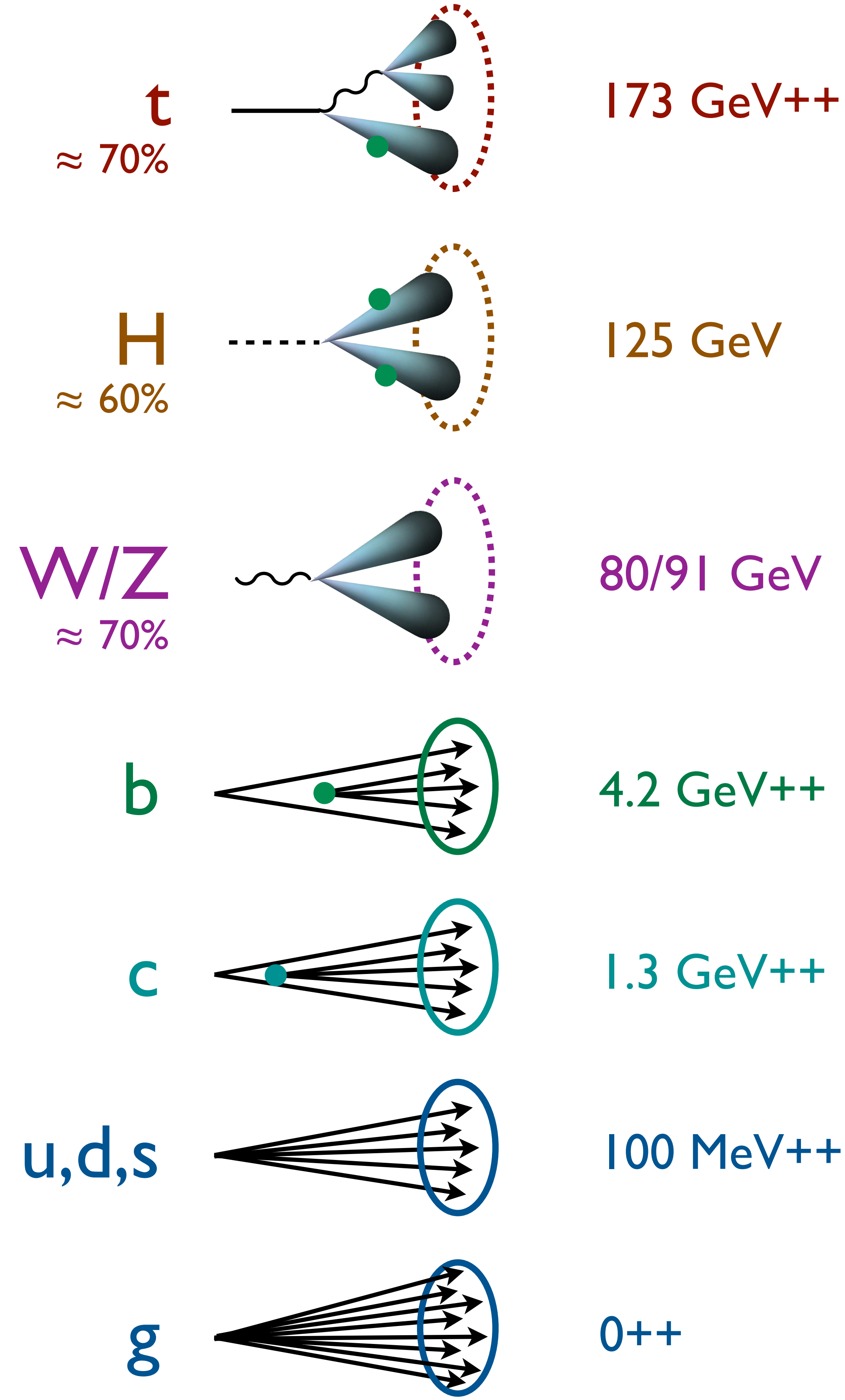
Jet Classification

Key supervised learning task at LHC



micro-jet
≈ 65%

BOOSTON 2019

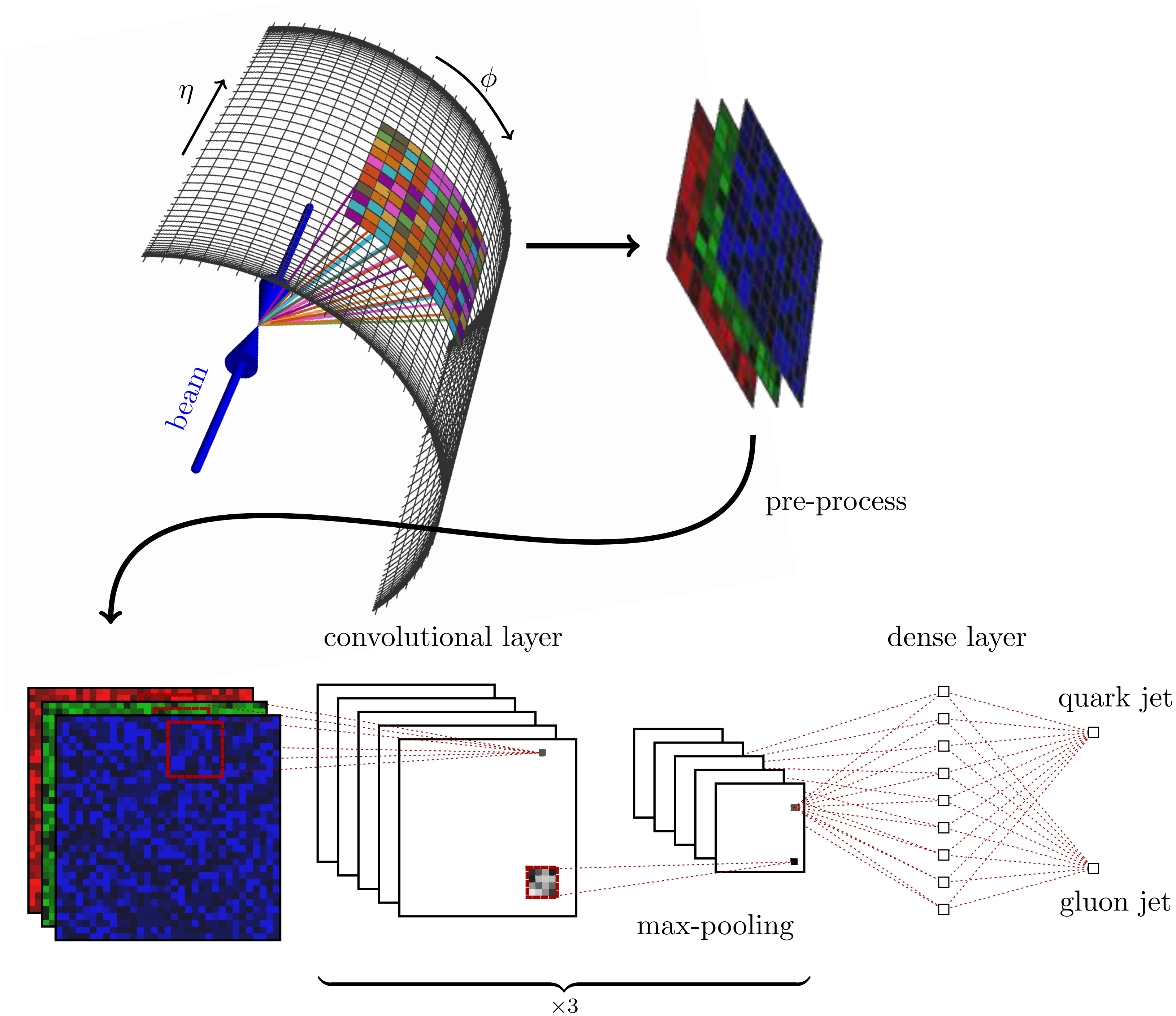


++ = Mass from QCD Radiation

[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



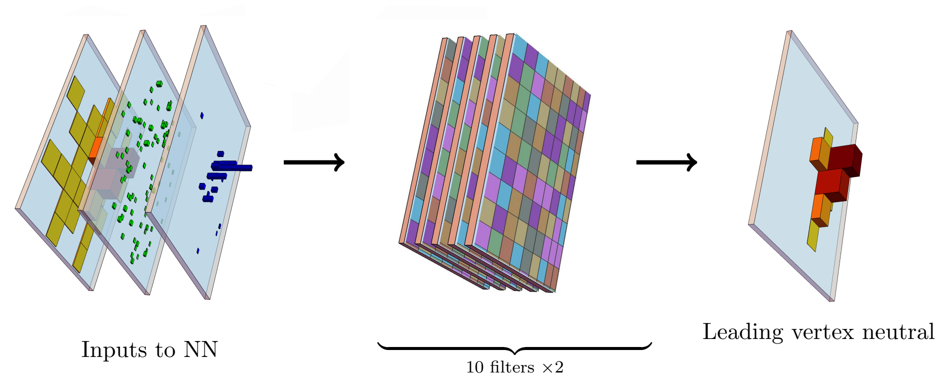
[figure from Komiske, Metodiev, Schwartz, JHEP 2017;
reviews in Larkoski, Moult, Nachman, arXiv 2017; Guest, Cranmer, Whiteson, ARNPS 2018]

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]

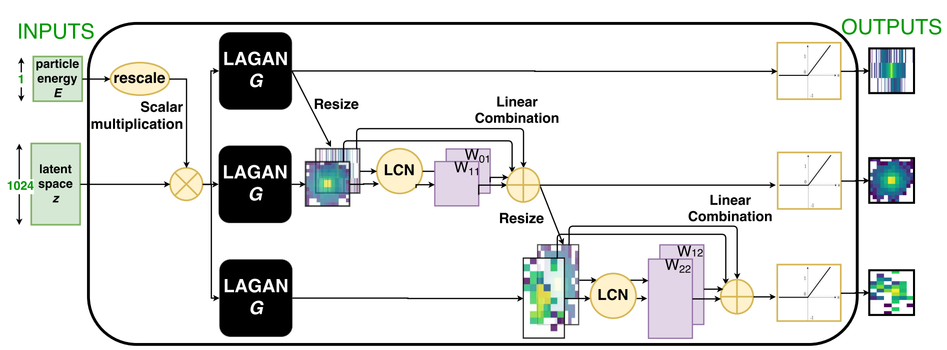


Labeled data: Objects J with property x
Solution: Map from J to x

Generation

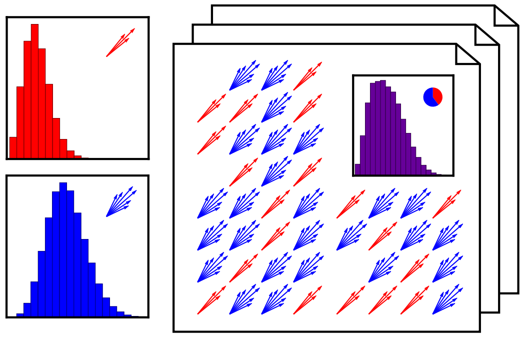
e.g. CaloGAN for fast detector simulation

[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

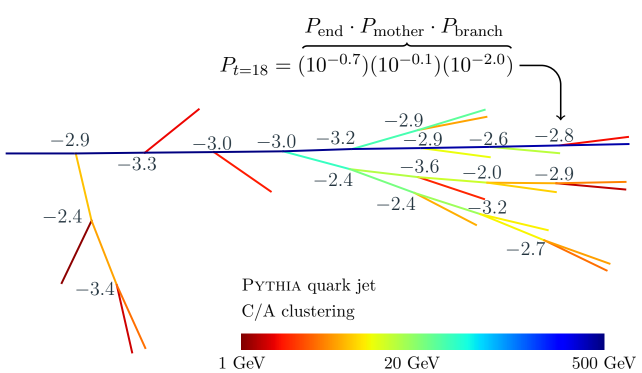
The Rise of Unsupervised Learning



Jet Topics

Blind Source Separation

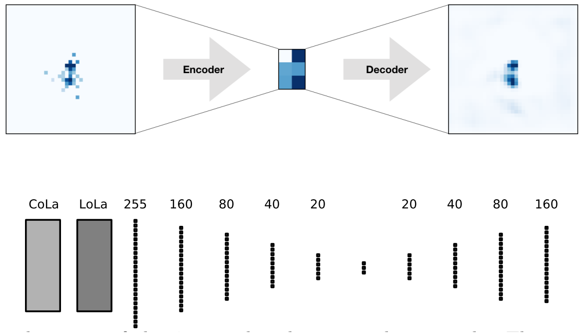
[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140; see also Metodiev, Nachman, JDT, 1708.02949]



JUNIPR

Probability Modeling

[Andreassen, Feige, Frye, Schwartz, 1804.09720; see also Monk, 1807.03685]



Autoencoders

Anomaly Detection

[Hajer, Li, Liu, Wang, 1807.10261; Heimel, Kasieczka, Plehn, Thompson, 1808.08979; Farina, Nakai, Shih, 1808.08992; Cerri, Nguyen, Pierini, Spiropulu, Vlimant, 1811.10276; see also Collins, Howe, Nachman, 1805.02664, 1902.02634; De Simone, Jacques, 1807.06038]

Common theme: Analyze event ensembles, not individual events

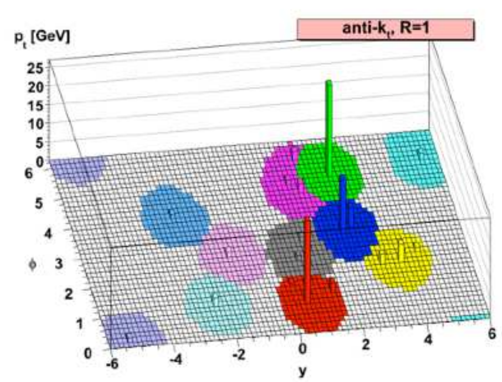
Particle Physics as Machine-Learning Testbed

- *Huge datasets with reliable simulations*

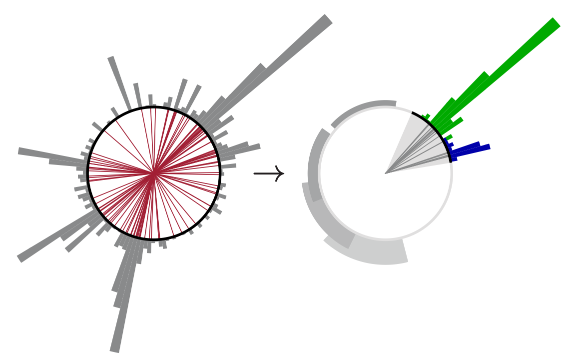
e.g. CMS opendata CERN with MadGraph + Pythia + Geant4 A SIMULATION TOOLKIT

- *Broad use of (un)supervised algorithms*

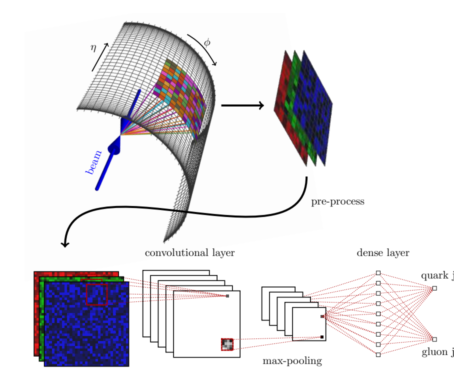
e.g.



Clustering



Noise Mitigation



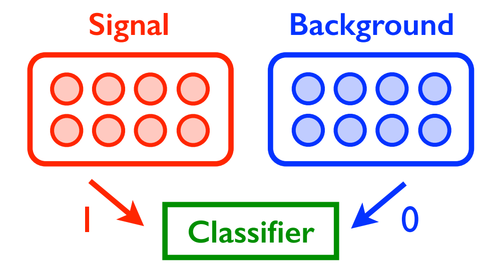
Machine Learning

- *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



$$\ell_{\mathrm{MSE}} = \Big\langle \big(h(\vec{x}) - 1\big)^2 \Big\rangle_{\text{signal}} + \Big\langle \big(h(\vec{x}) - 0\big)^2 \Big\rangle_{\text{background}}$$

Loss Function

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

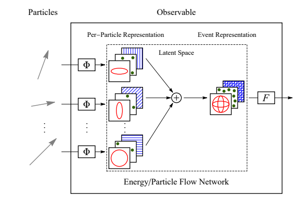
High-Level Features
Basis of High-Level Features
Jet Image
Multi-channel Jet Image
Abstract Jet Image
Sorted Four-Vectors
Clustered Four-Vectors
Lund Plane Emissions
Kitchen Sink
…

Signal vs. Background

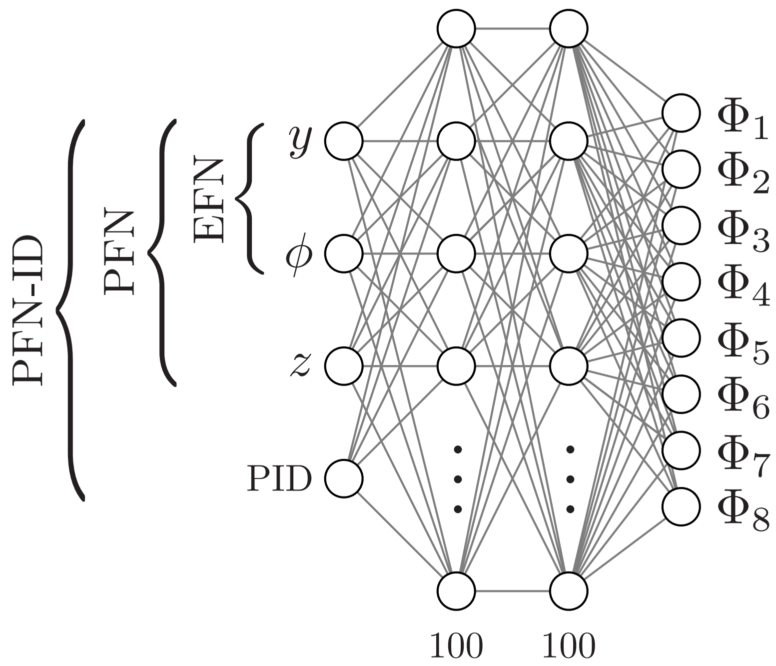
Signal	vs.	Background
Quark Jets	vs.	Gluon Jets
Up-type Quarks	vs.	Down-type Quarks
W/Z Bosons	vs.	QCD Jets
W Bosons	vs.	Z Bosons
Top Quarks	vs.	QCD Jets
Exotic Boosted Objects	vs.	QCD Jets
CMS Open Data Samples	vs.	Each other
…	vs.	…

[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)]

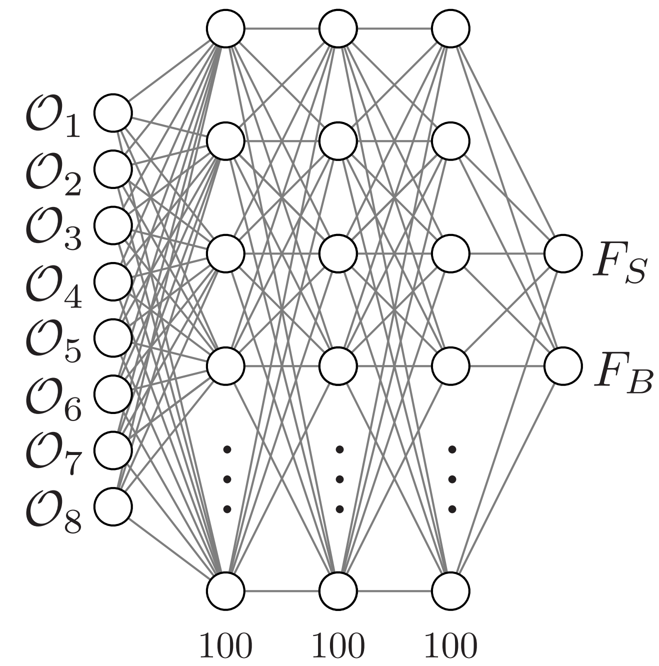
Technical Implementation



Per-Particle Network: Φ



Latent Combiner: F



Per-Jet (Latent Space):

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

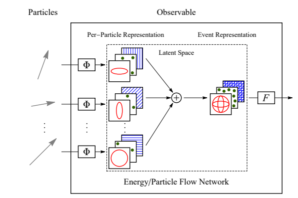
PFN: $\mathcal{O}_a = \sum_{i \in \text{jet}} \Phi_a(y_i, \phi_i, z_i, \text{PID}_i)$

Final Discriminant:

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

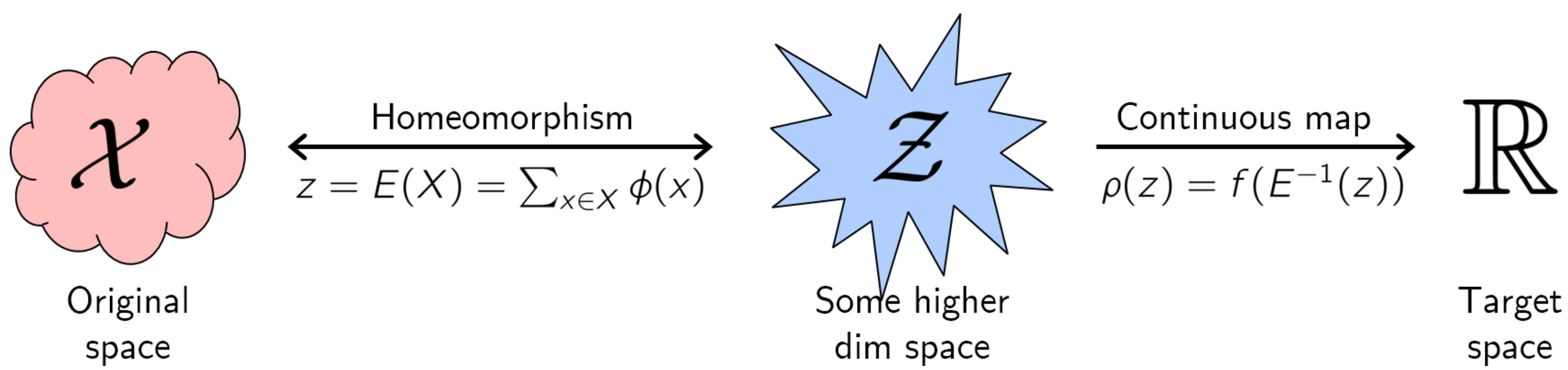
[Komiske, Metodiev, JDT, JHEP 2019]

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 $\rho\left(\sum_{x\in X}\phi(x)\right)$, for suitable transformations ϕ and ρ.*

(!)



Variable-Length Unordered Set of Particles

Additive Latent Space

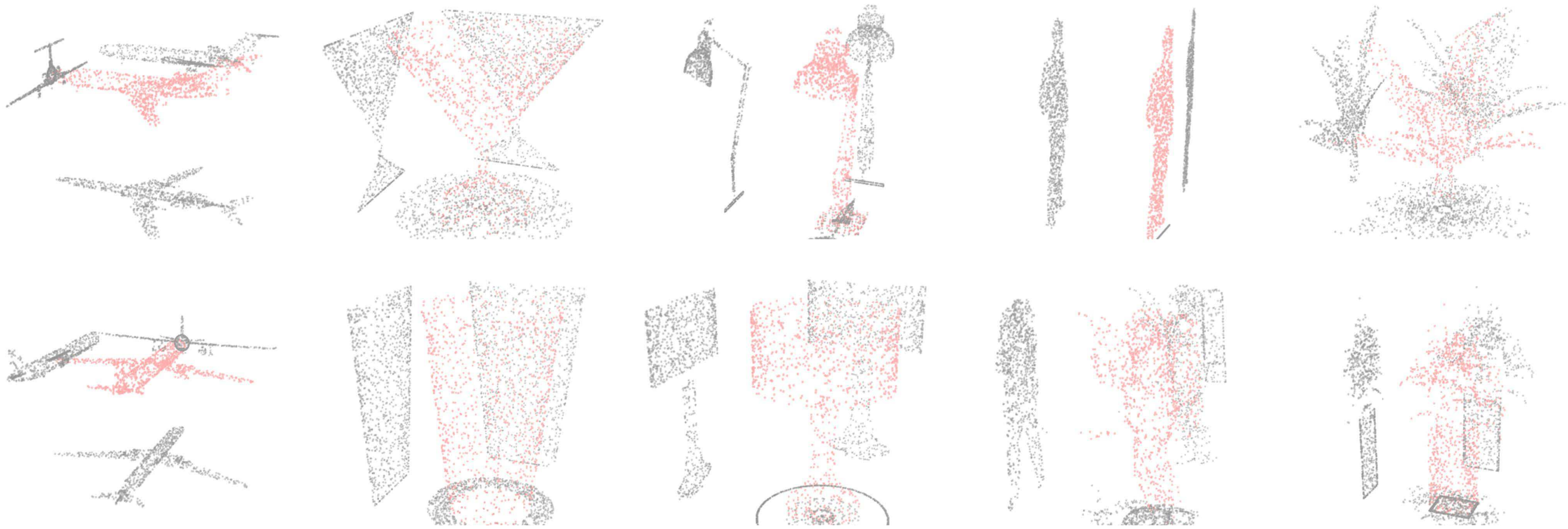
Generic Observable

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

Deep Sets for…

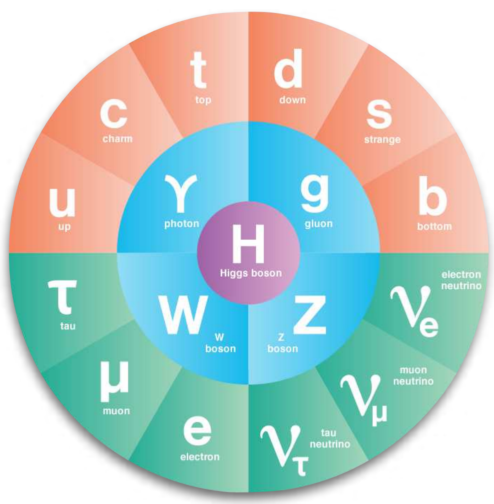
Celebrity Face Anomaly Detection

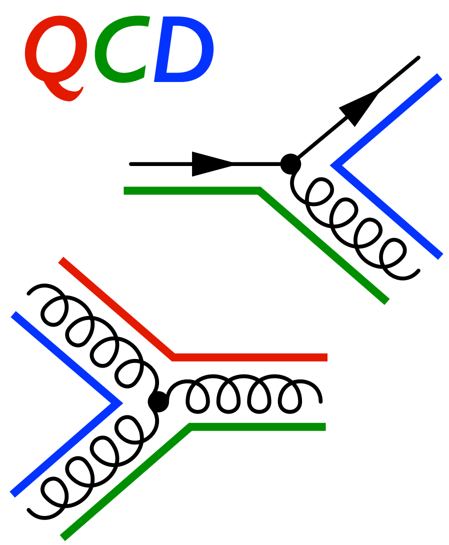
Point Cloud Classification

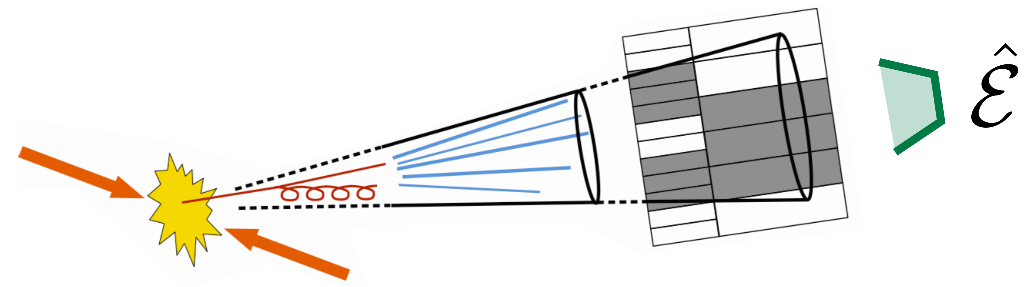


[Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114]

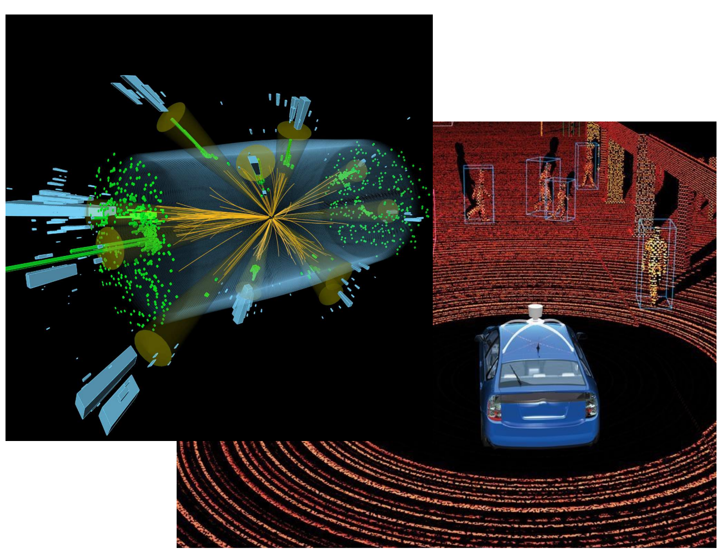
Deep Sets for Particle Jets

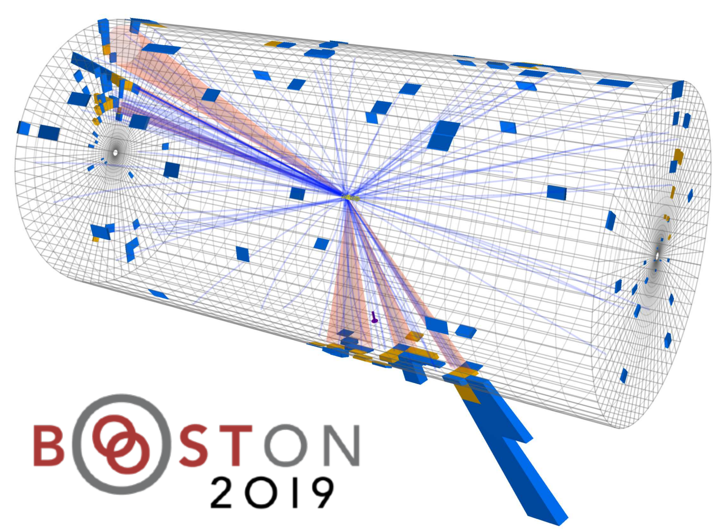


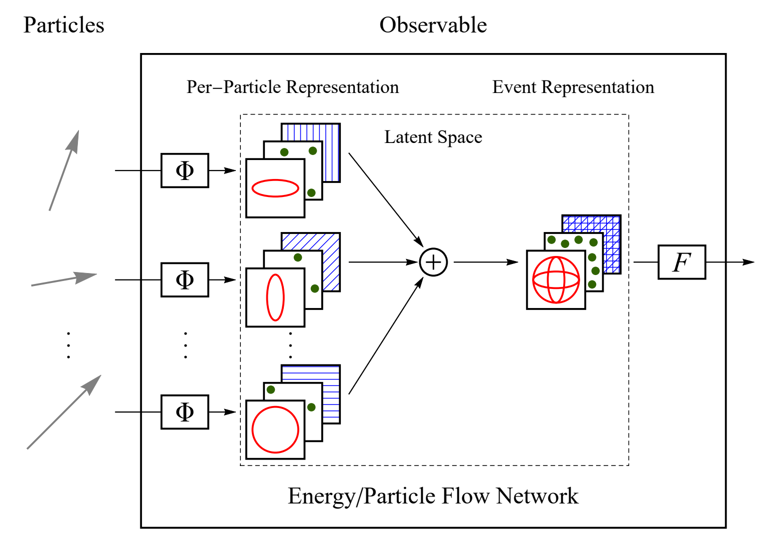




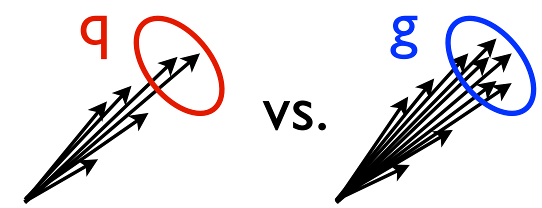
$$\mathrm{d}P_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{\mathrm{d}\theta}{\theta} \frac{\mathrm{d}z}{z}$$

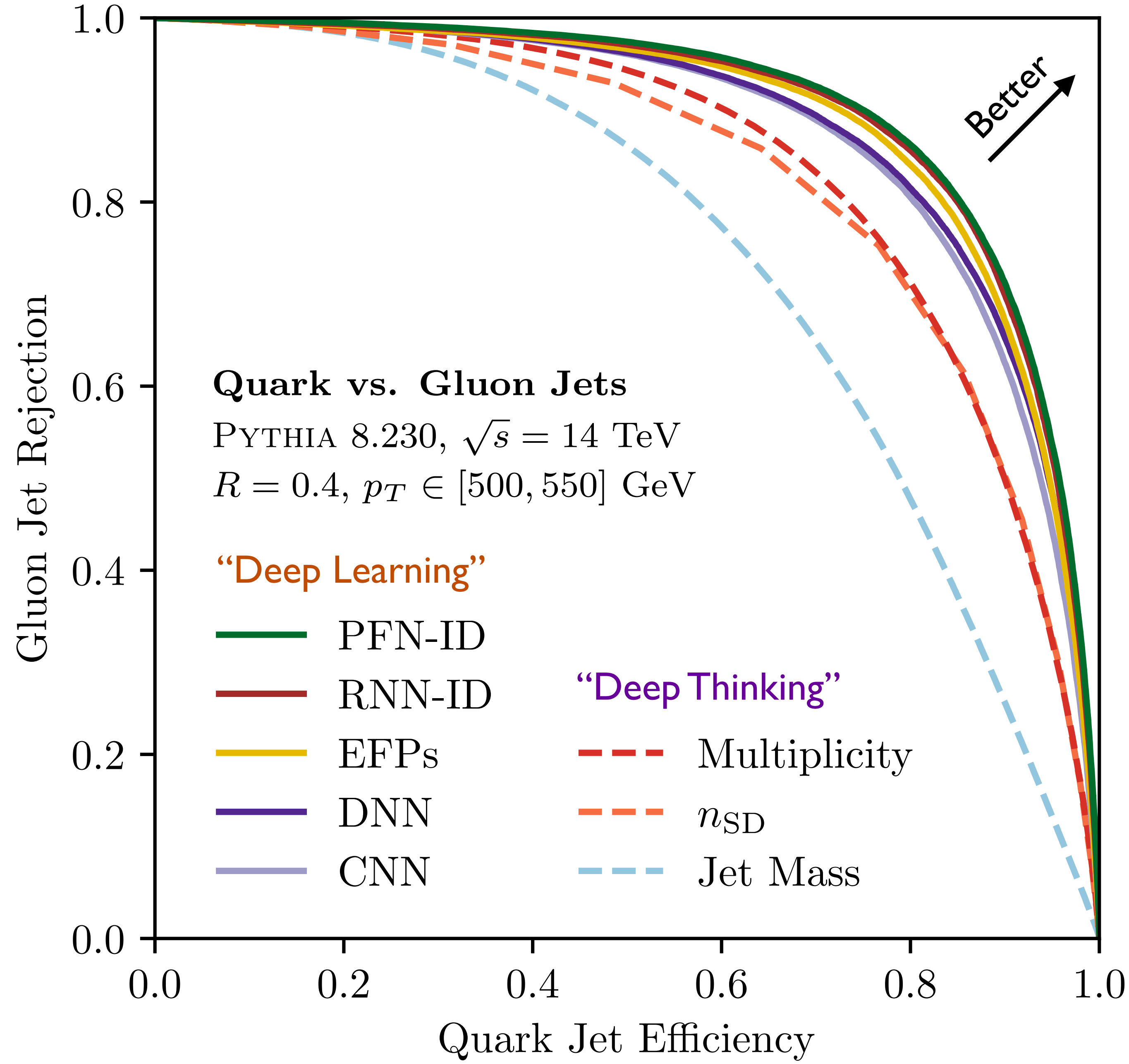






Discrimination Performance



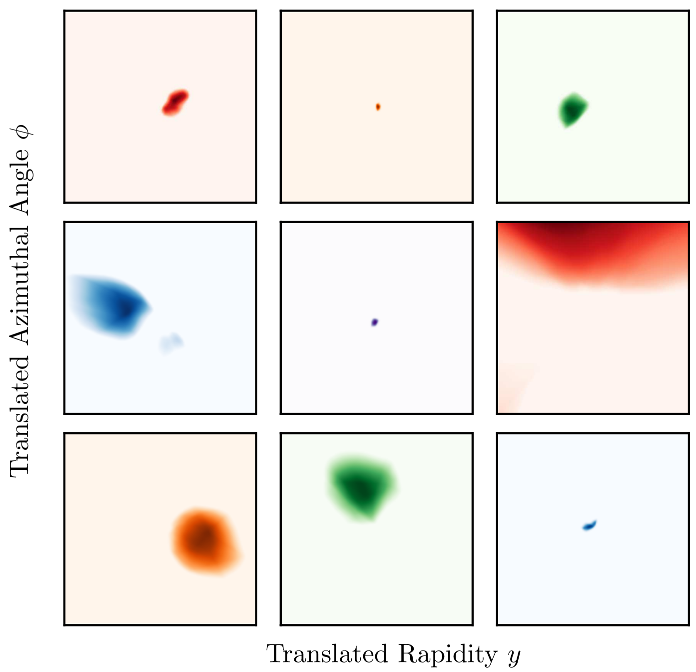


[Komiske, Metodiev, JDT, JHEP 2019]

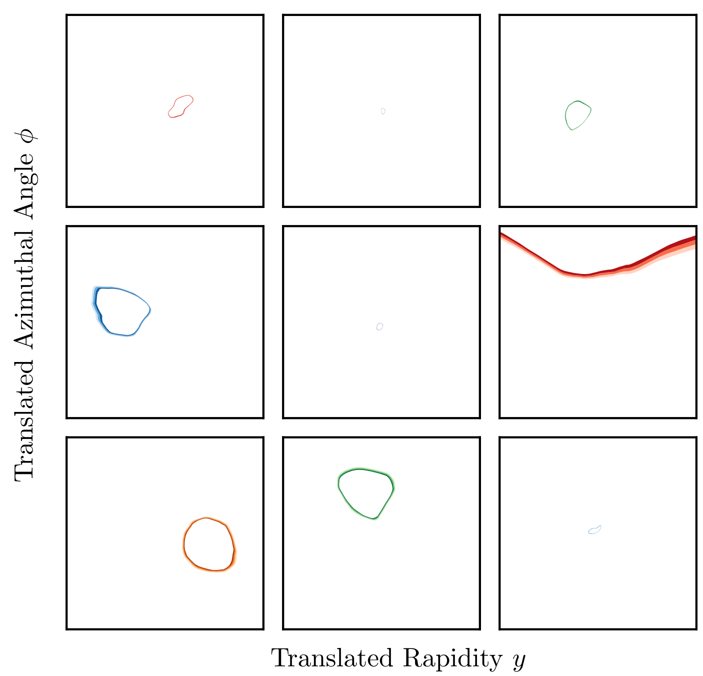
Psychedelic Network Visualization

(go back)

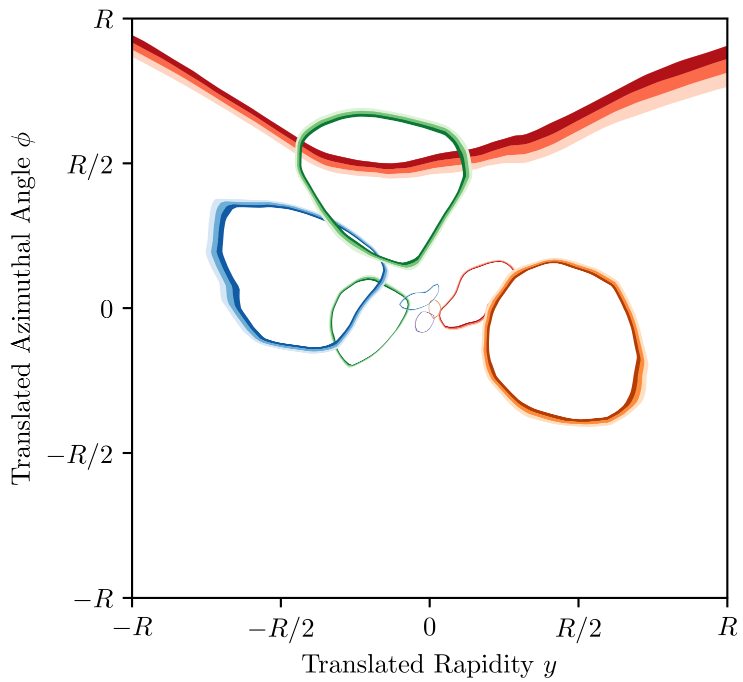
Latent Filters



50% Contours

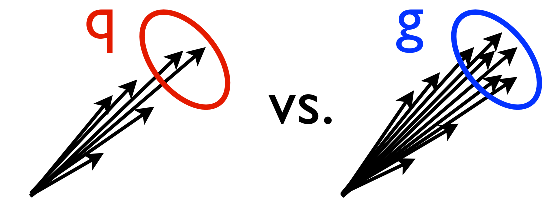


Overlay



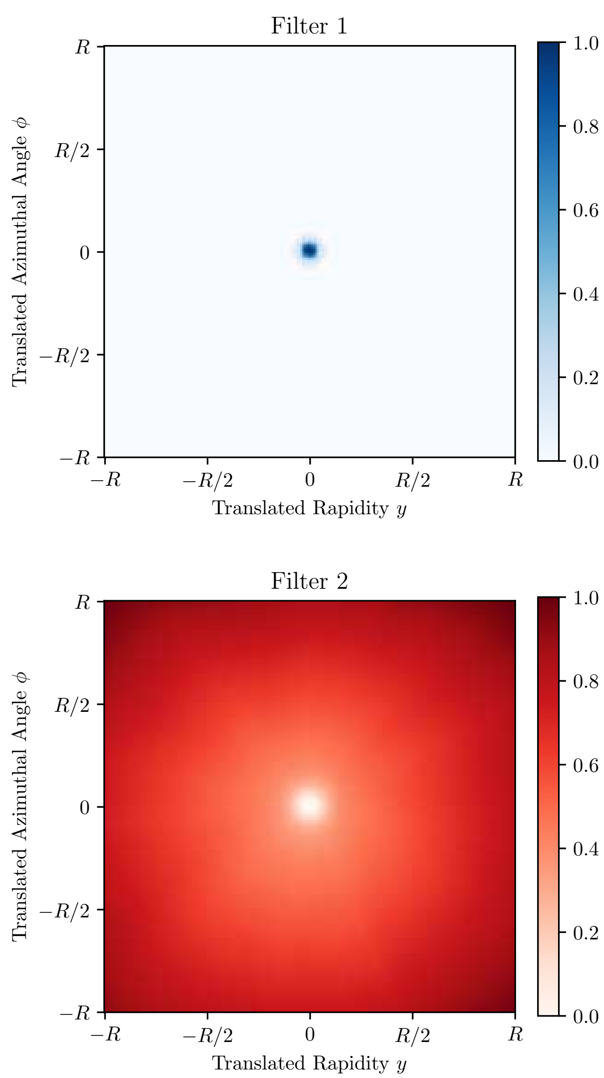
"Ok, but did you really learn something you didn't already know?"

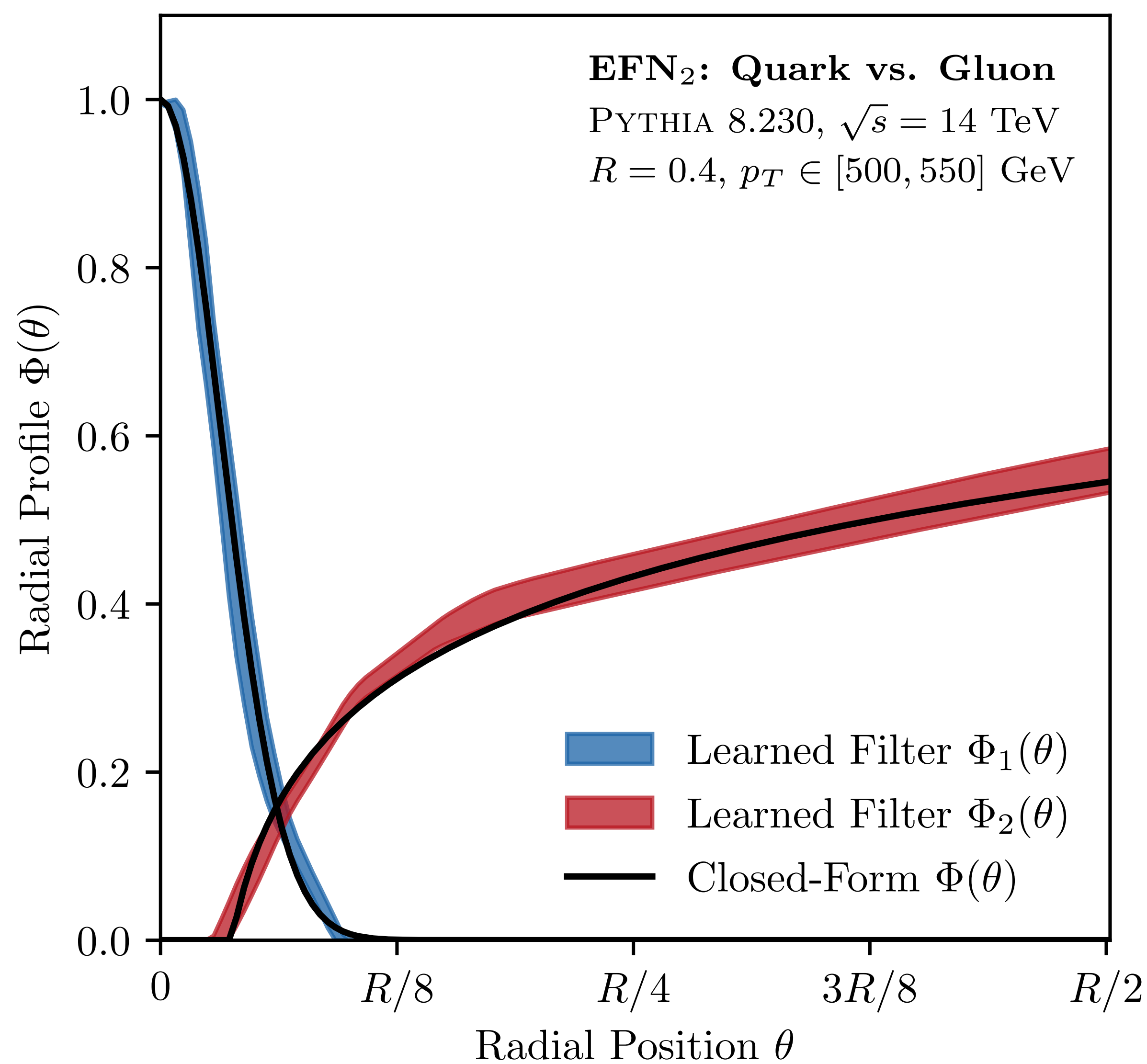
Learning from the Machine



For ℓ = 2 EFN, radial moments: $\sum_{i \in \text{jet}} z_i \, f(\theta_i)$

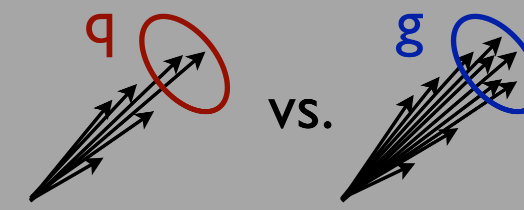
cf. Angularities:
$f(\theta) = \theta^\beta$





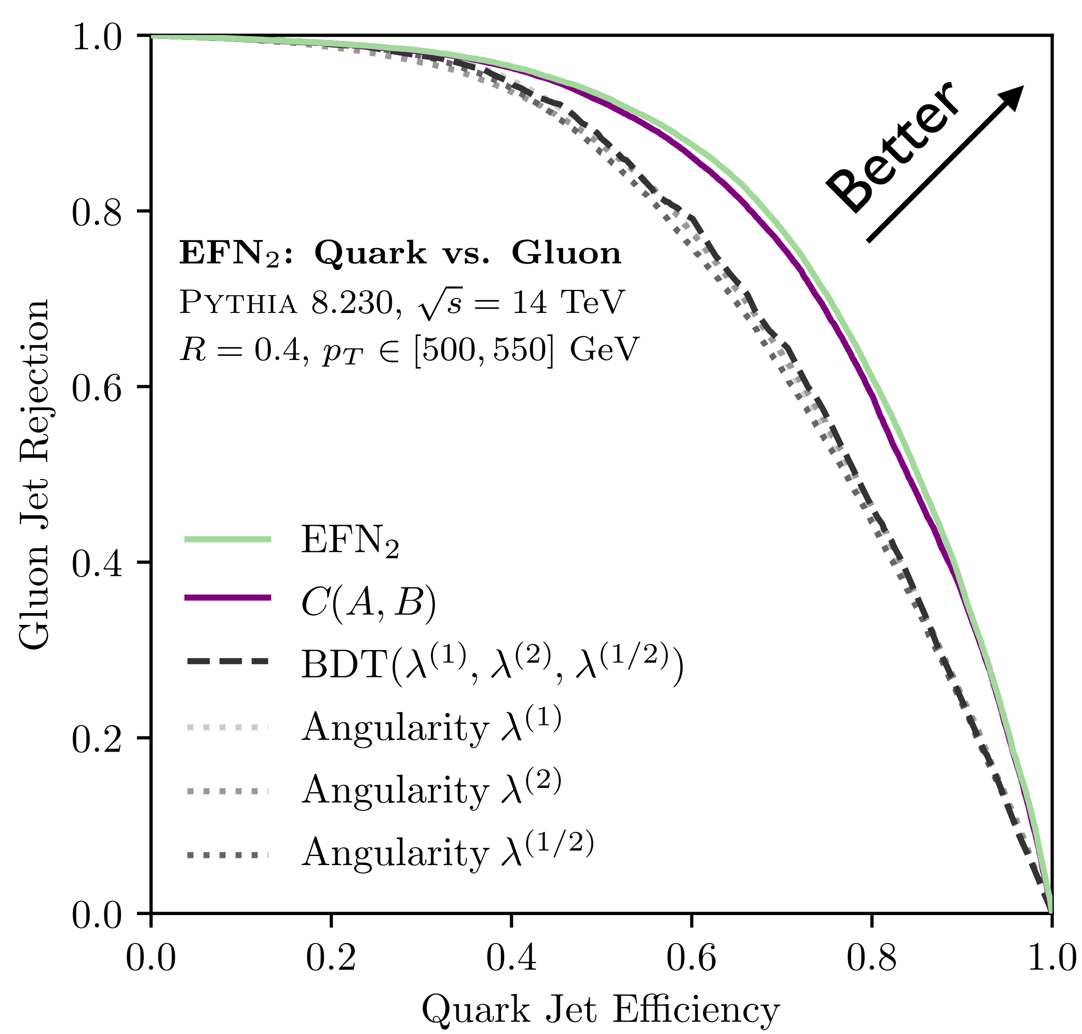
[Komiske, Metodiev, JDT, JHEP 2019;
cf. Larkoski, JDT, Waalewijn, JHEP 2014; using Berger, Kucs, Sterman, PRD 2003; Ellis, Vermilion, Walsh, Hornig, Lee, JHEP 2010]

Learning from the Machine



For ℓ = 2 EFN, radial moments: $$\sum_{i \in \text{jet}} z_i f(\theta_i)$$

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



EFN outperformed a domain expert (i.e. me)

But we reverse engineered the machine (and learned something about QCD)

[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:

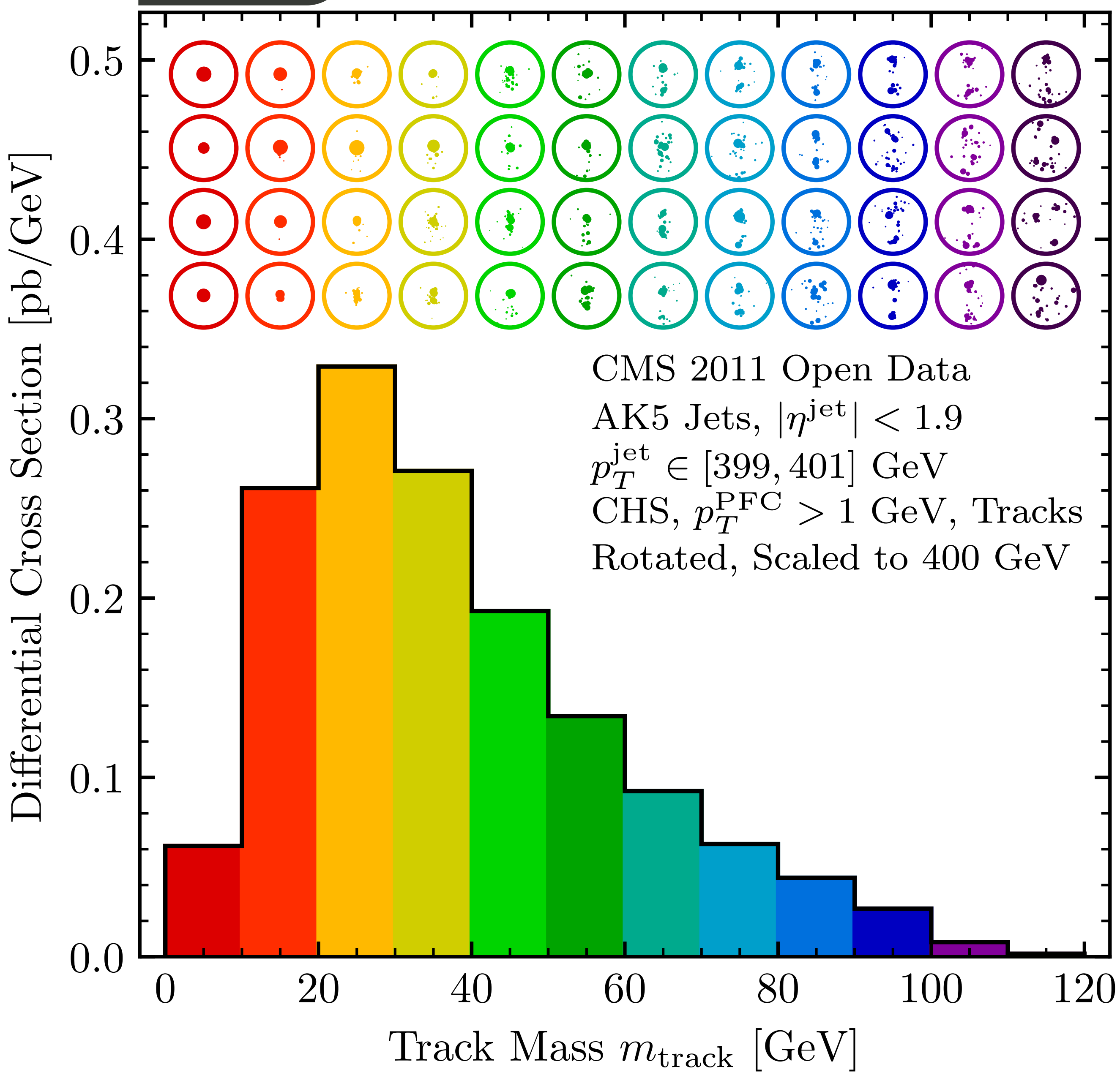
[Peleg, Werman, Rom, IEEE 1989;
Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000;
Pele, Werman, ECCV 2008; Pele Taskar, GSI 2013]

Minimum "work" (stuff x distance) to make one distribution look like another distribution

Equivalent to 1-Wasserstein metric; very popular in ML applications

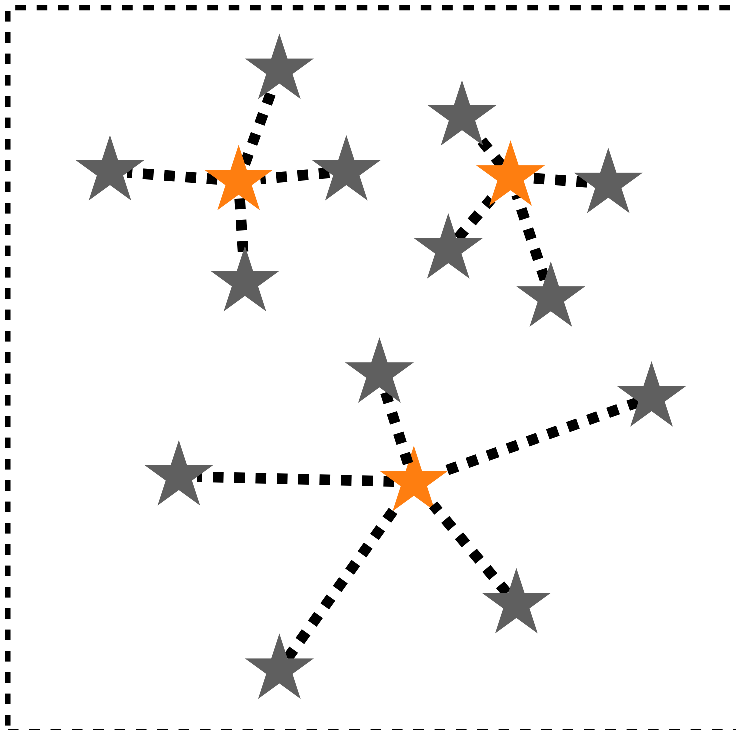
Most Representative Jets

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



k-medoids

Per mass bin



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

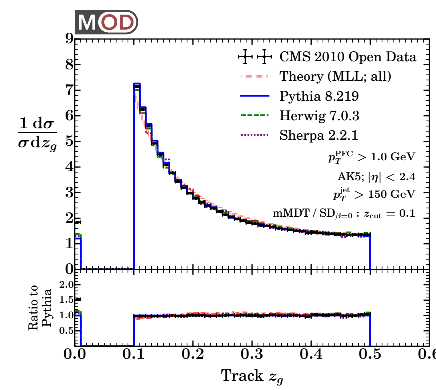
The Rise of Public Collider Data

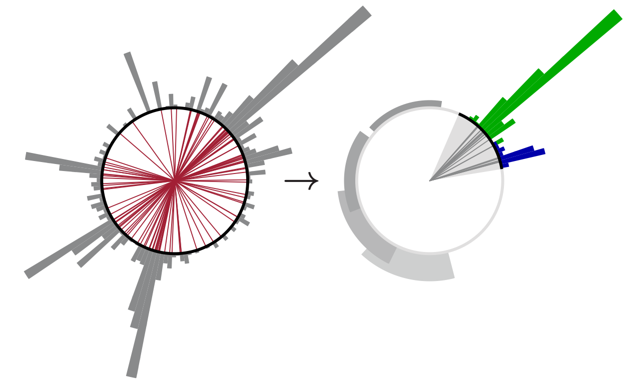
Since 2014: CMS Open Data project

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

Standard Model & QCD

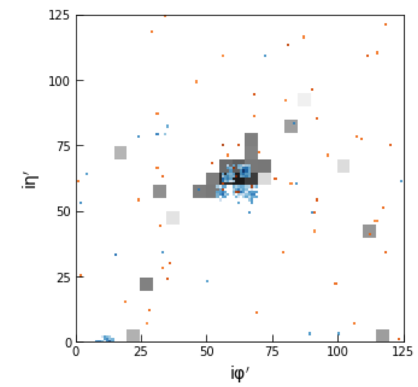
[Larkoski, Marzani, JDT, Tripathee, Xue, PRL 2017, PRD 2017;
Mehdiabadi, Fahim, arXiv 2019;
Apyan, Cuozzo, Klute, Saito, Schott, Sintayehu, arXiv 2019]

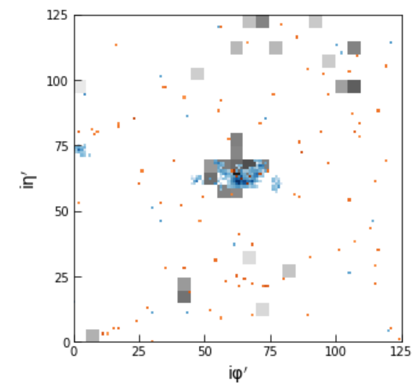




Machine Learning

[Fernndez 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]