Collision Course

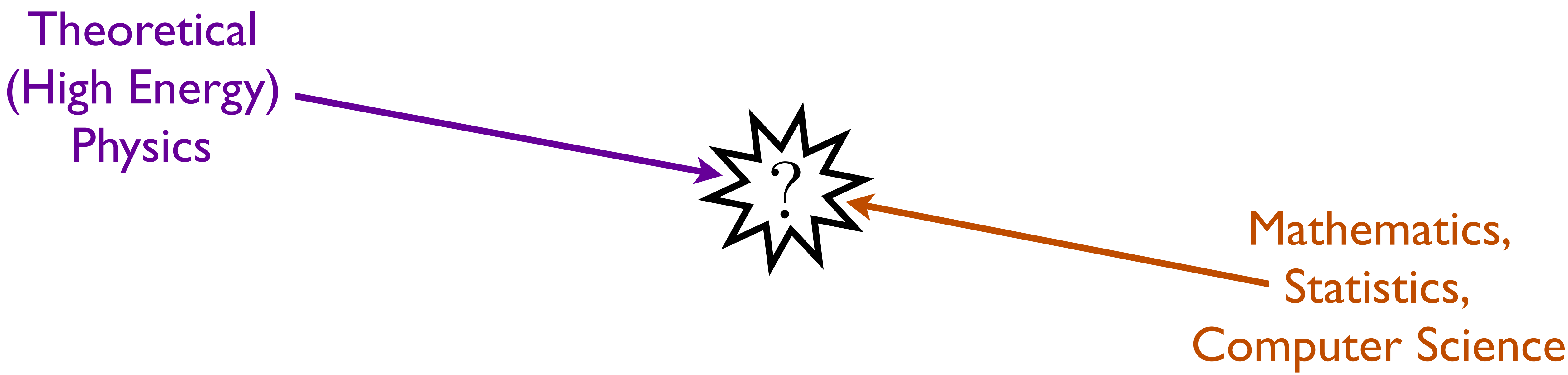
Particle Physics as a Machine-Learning Testbed

Jesse Thaler

Theoretical Physics for Machine Learning, Aspen Center for Physics — January 17, 2019

“Collision Course”

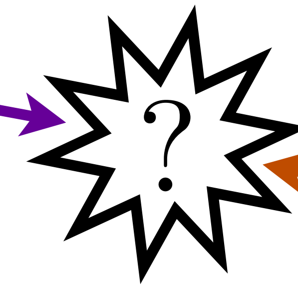
More collider talks on Friday:
Anders Andreassen & David Shih



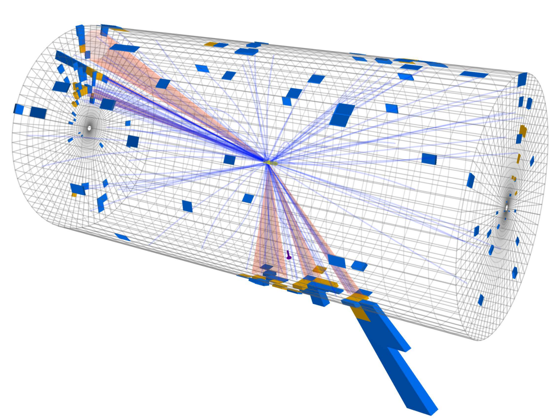
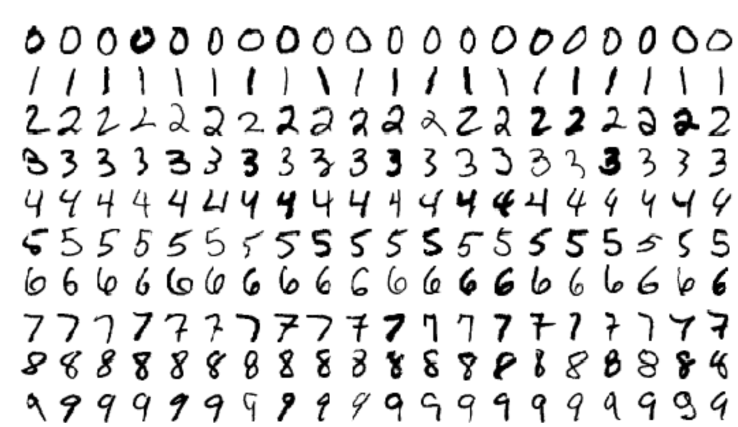
“Collision Course”

More collider talks on Friday:
Anders Andreassen & David Shih

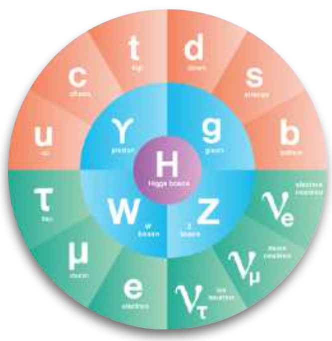
Theoretical
(High Energy)
Physics



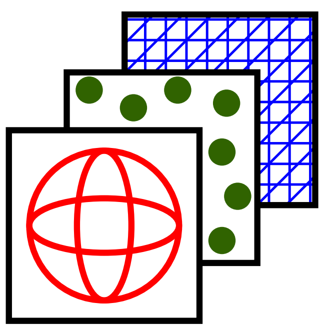
Mathematics,
Statistics,
Computer Science

Could  *be the next*  *?*

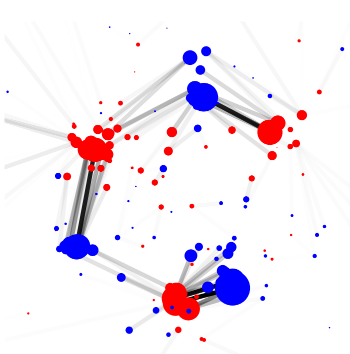
Outline



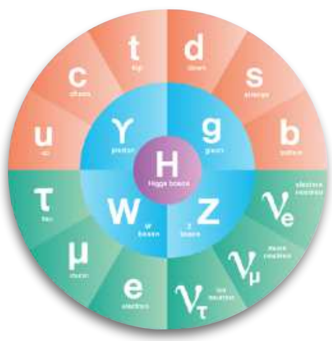
Particle Physics Primer



Point Clouds & Energy Flow Networks



(The Metric Space of Collider Events)

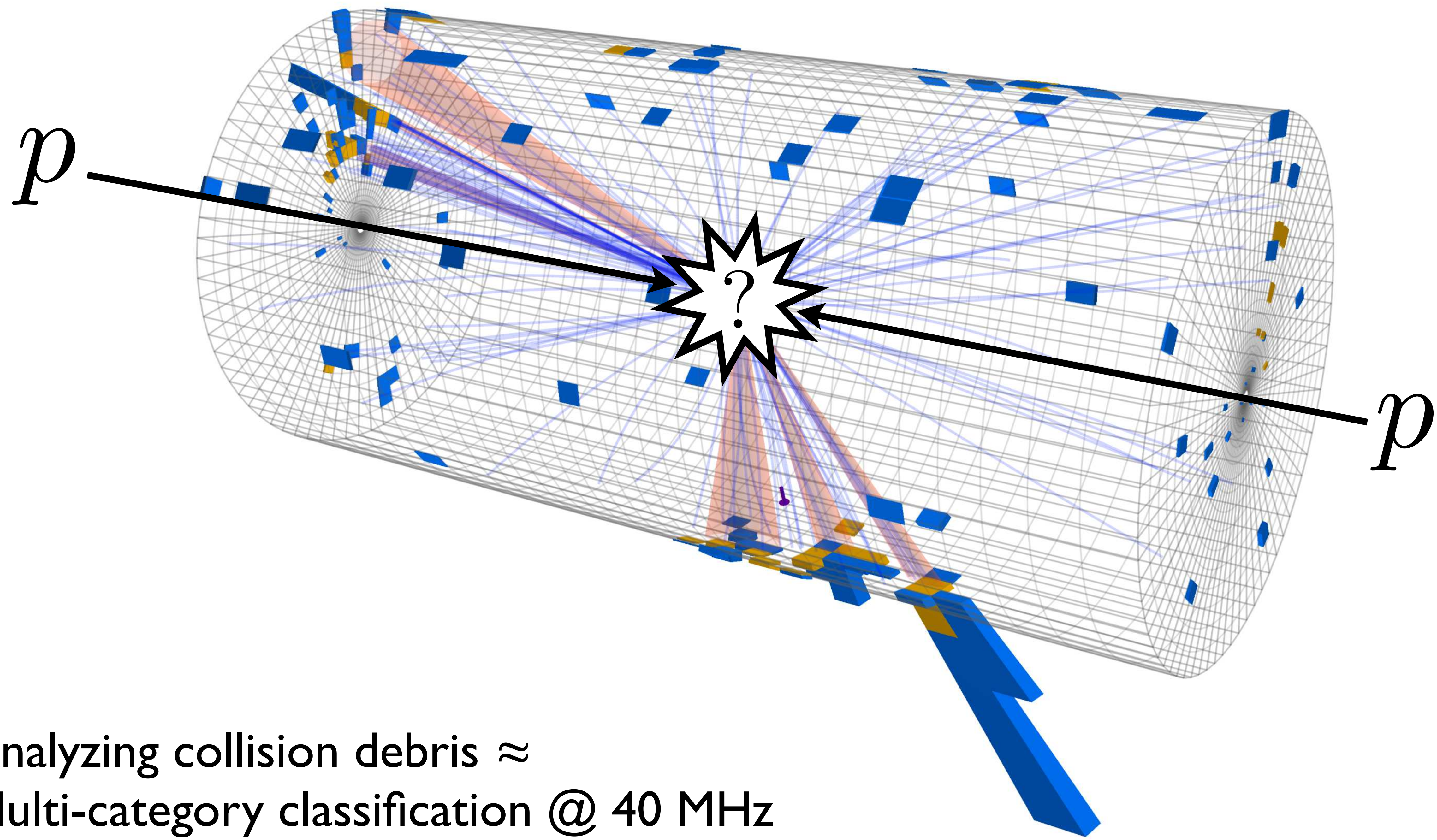


Particle Physics Primer

Point Clouds & Energy Flow Networks

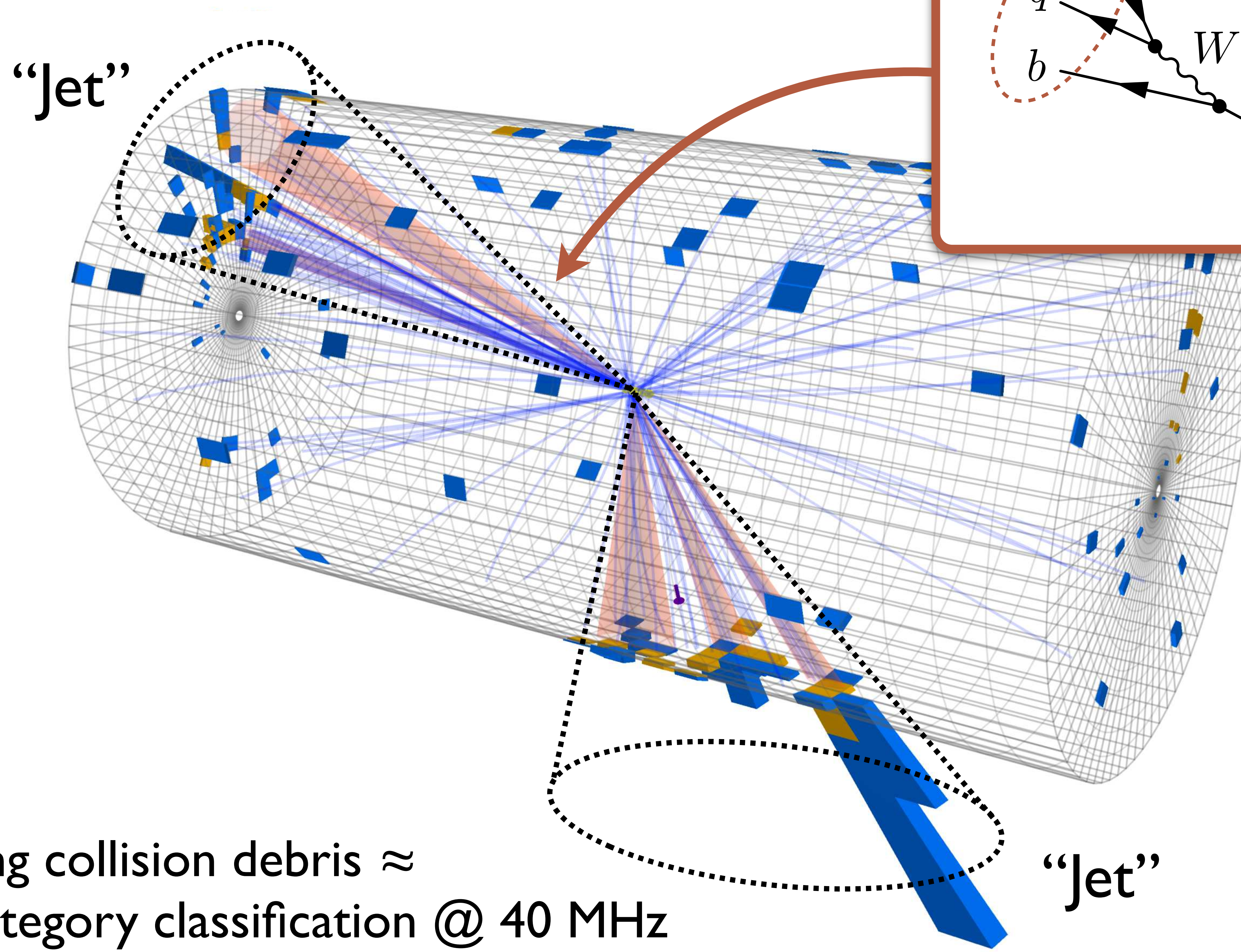
(The Metric Space of Collider Events)

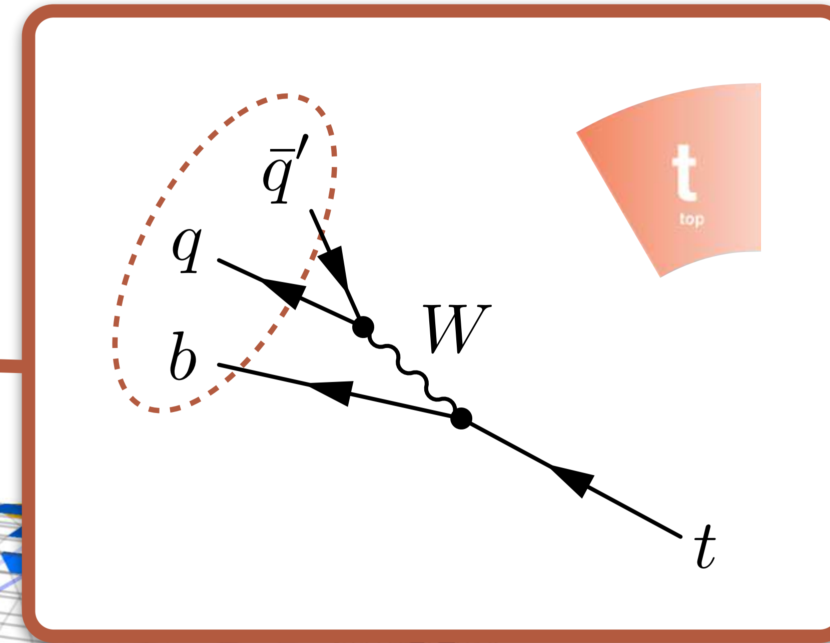
CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253



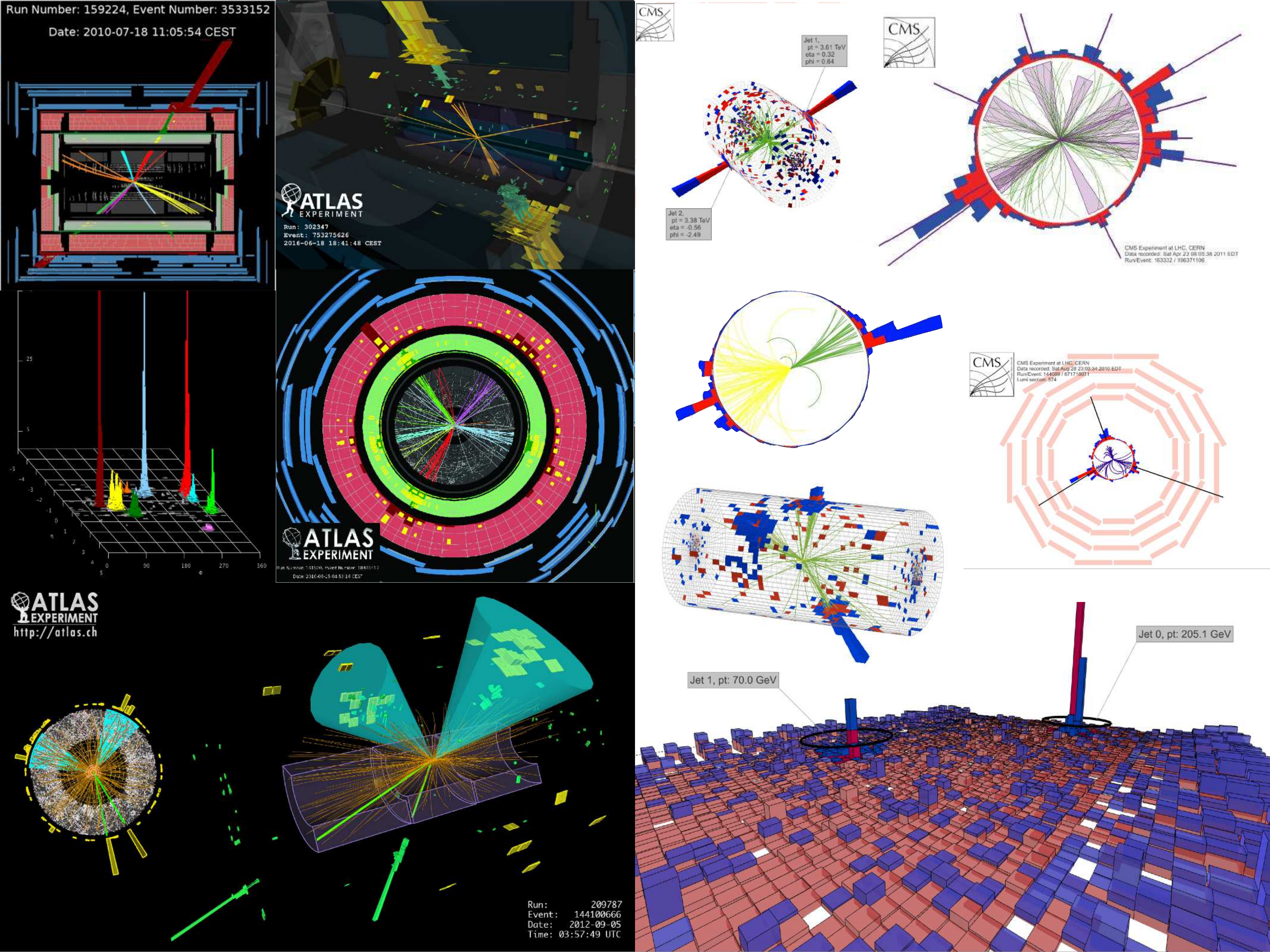
Analyzing collision debris ≈
Multi-category classification @ 40 MHz

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

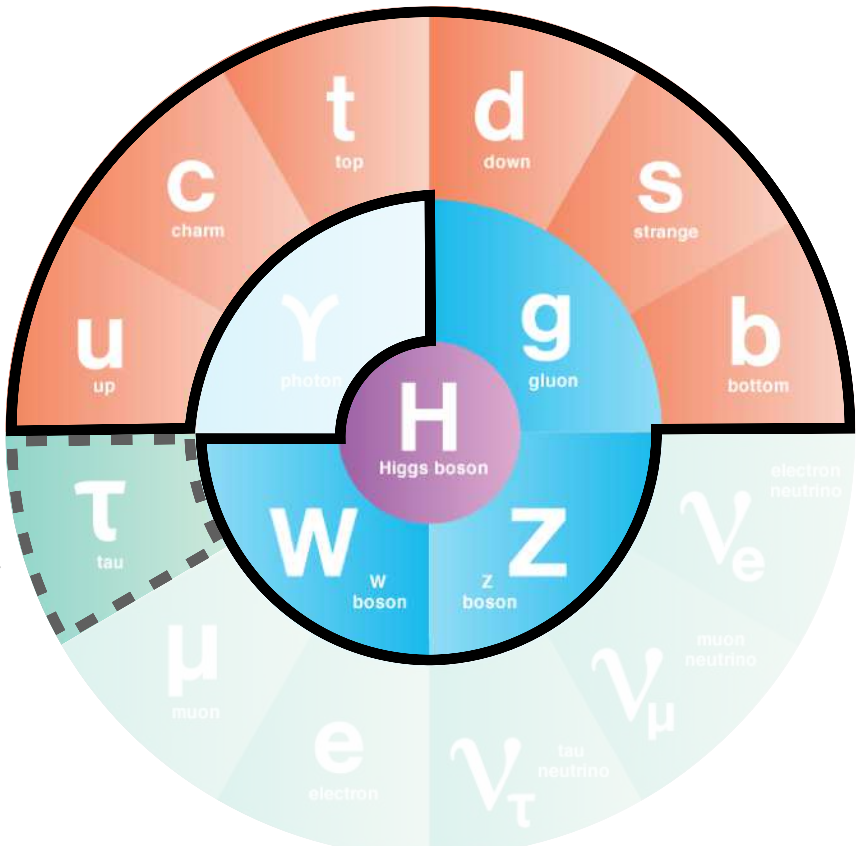




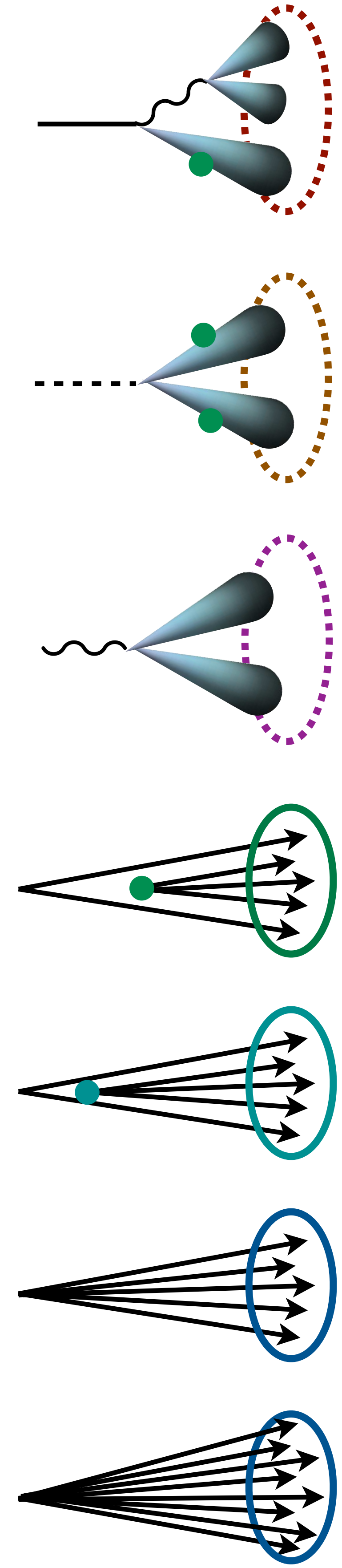
Analyzing collision debris ≈
Multi-category classification @ 40 MHz


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 Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 196371106
ATLAS EXPERIMENT
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

Jets from QCD and the Standard Model



micro-jet
≈ 65%

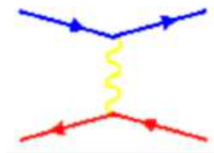


t ≈ 70%	173 GeV++
H ≈ 60%	125 GeV
W/Z ≈ 70%	80/91 GeV
b	4.2 GeV++
c	1.3 GeV++
u,d,s	100 MeV++
g	0++

++ = Mass from QCD Radiation

Particle Physics as ML Testbed

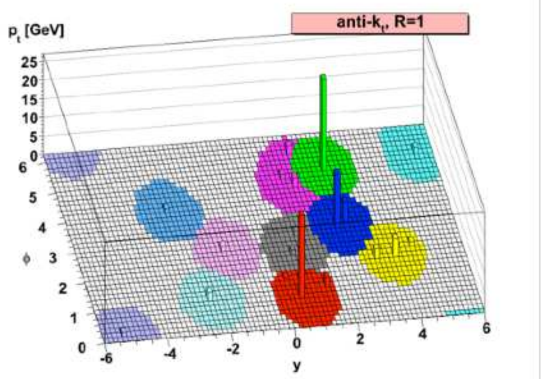
- *Huge datasets with reliable simulations*

e.g. [CMS opendata CERN] with  + [Pythia] + [GEANT4 A SIMULATION TOOLKIT]

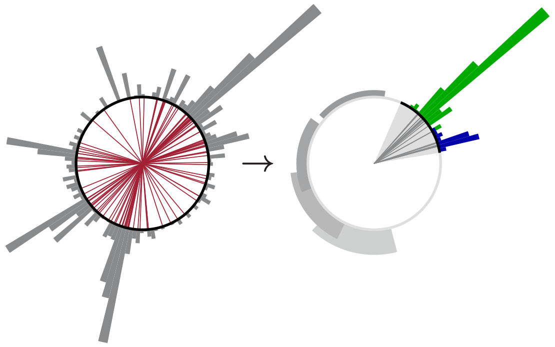
MadGraph   Pythia

- *Broad use of (un)supervised algorithms*

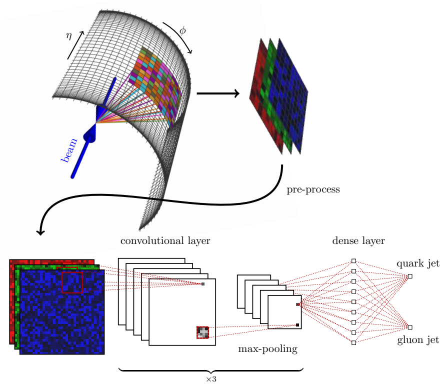
e.g.



Clustering



Noise Mitigation

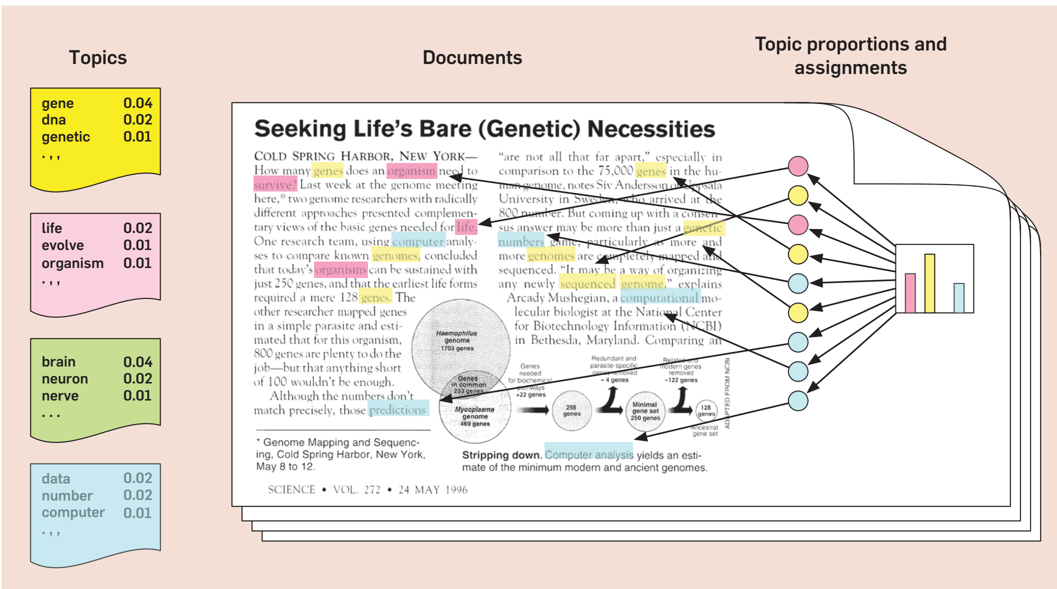


Machine Learning

- *Extensive domain knowledge and strong theory priors*

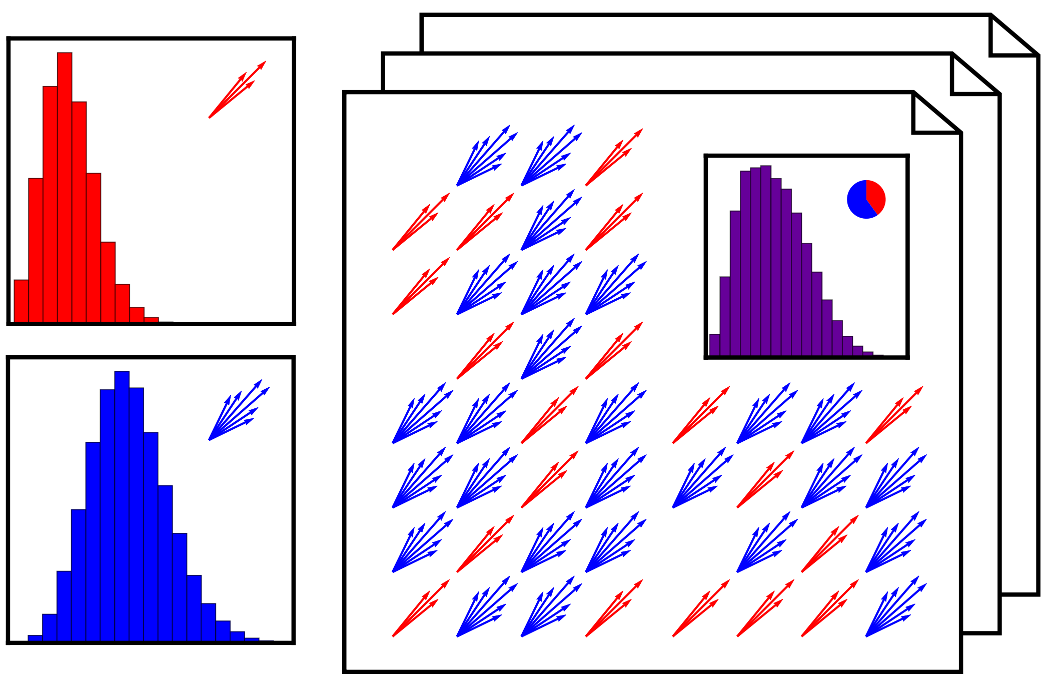
[figures from Cacciari, Salam, Soyez, 0802.1189; Larkoski, Marzani, JDT, Tripathee, Xue, 1704.05066; Komiske, Metodiev, Schwartz, 1612.01551]

E.g. Topic Modeling for Jets



Blind Source Separation:

Documents as bags of words
(which they really aren't)



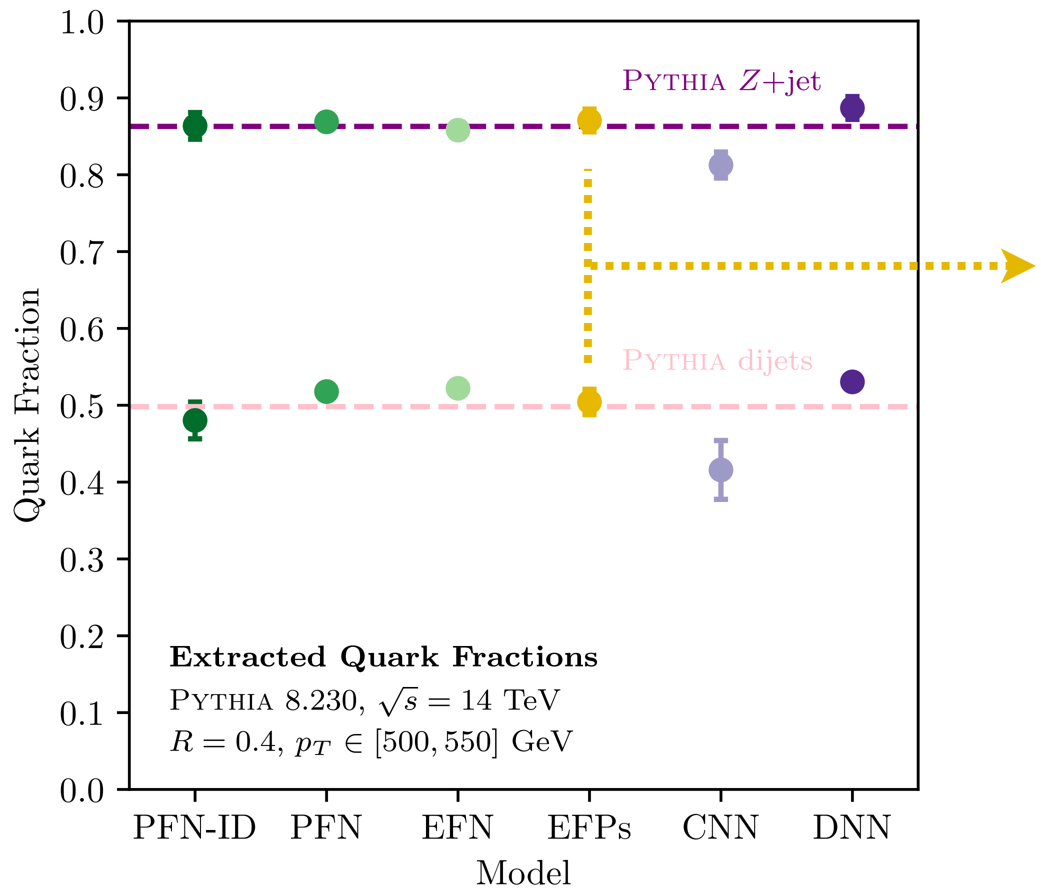
Quark/Gluon Separation:

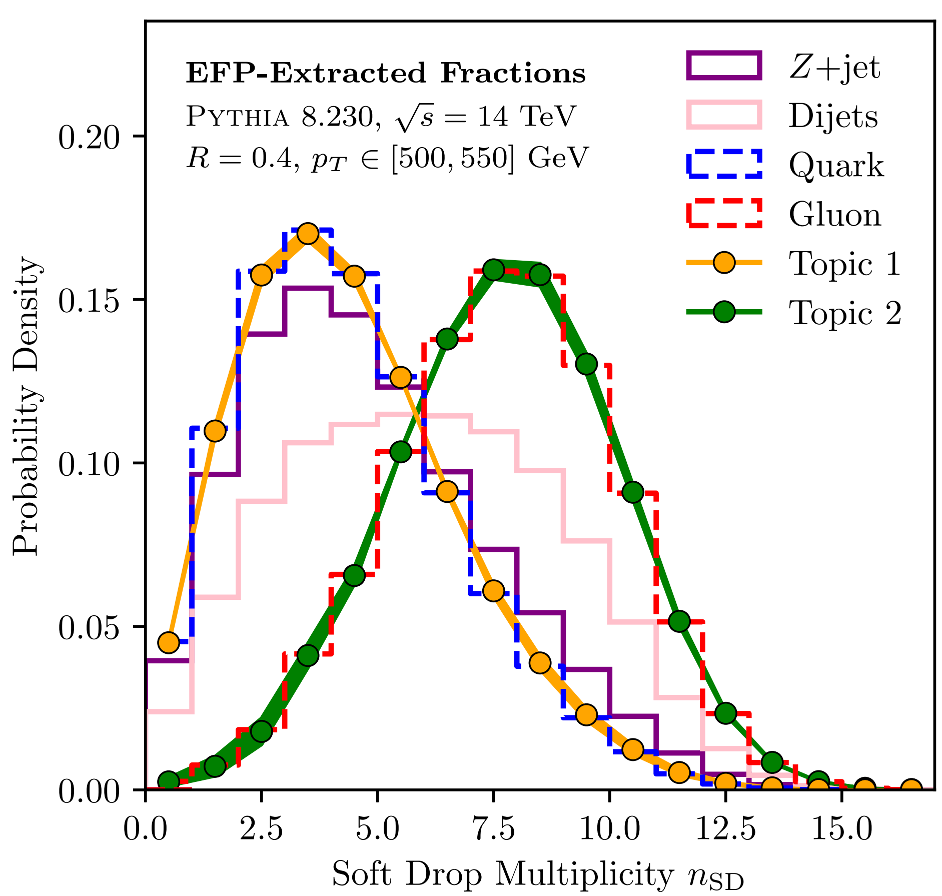
Jet histograms as mixtures of "mutually irreducible" categories
(which they really are)*

[Blei, CACM 2012; Metodiev, JDT, 1802.00008]

E.g. Topic Modeling for Jets

ML validation in complicated (but controlled) environment

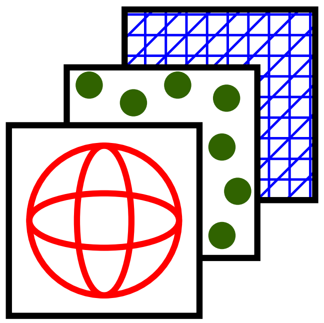




[Komiske, Metodiev, JDT, 1809.01140;
plotting Frye, Larkoski, JDT, Zhou, 1704.06266; Komiske, Metodiev, JDT, 1712.07124;
using Katz-Samuels, Blanchard, Scott, 1710.01167; Metodiev, Nachman, JDT, 1708.02949; Metodiev, JDT, 1802.00008]

[Blei, CACM 2012; Metodiev, JDT, 1802.00008]

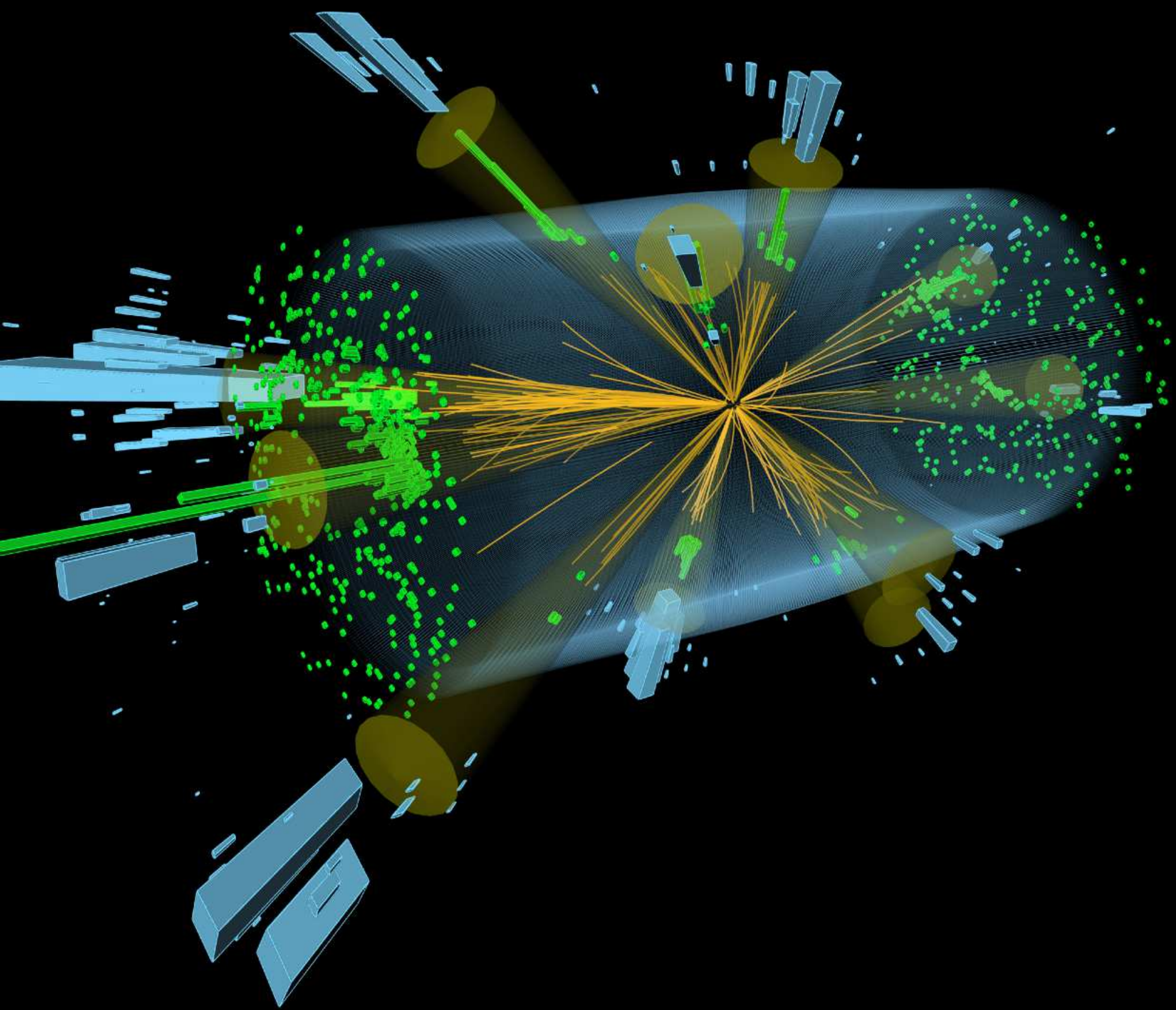
Particle Physics Primer



Point Clouds & Energy Flow Networks

(The Metric Space of Collider Events)

What is a Collision Event?



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

Point Cloud

[Popular Science, 2013]

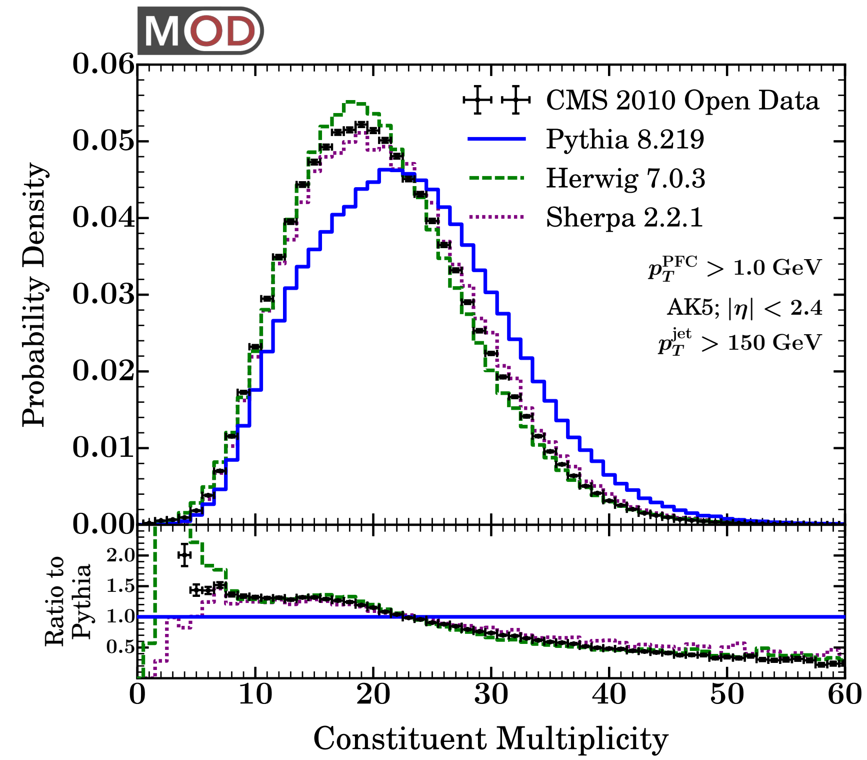
Key Fact #1: Events/Jets are Point Clouds

Jet constituents:

Particle-like objects

Variable-length

Unordered set → QM



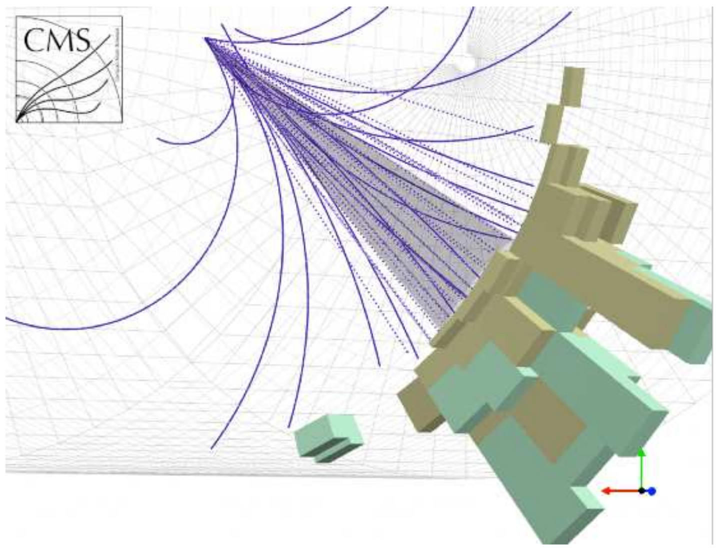
Per particle:

$\{E, p_x, p_y, p_z\}$ *or* $\{p_T, y, \Phi, m\}$

Flavor/charge labels

Vertex information

Quality criteria, etc.



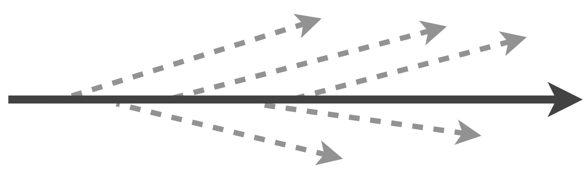
[plot from Tripathee, Xue, Larkoski, Marzani, JDT, 1704.05842]

Key Fact #2: IRC Safety is Important

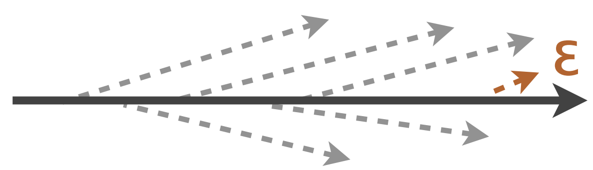
Wide range of interesting observables are "safe"

Interesting ≈ Calculable in fixed-order perturbation theory

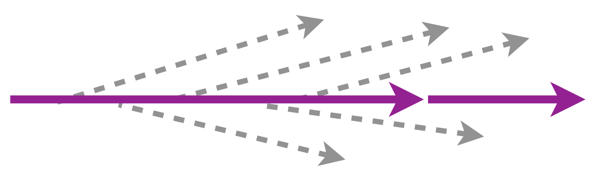
Original Jet



Infrared



Collinear

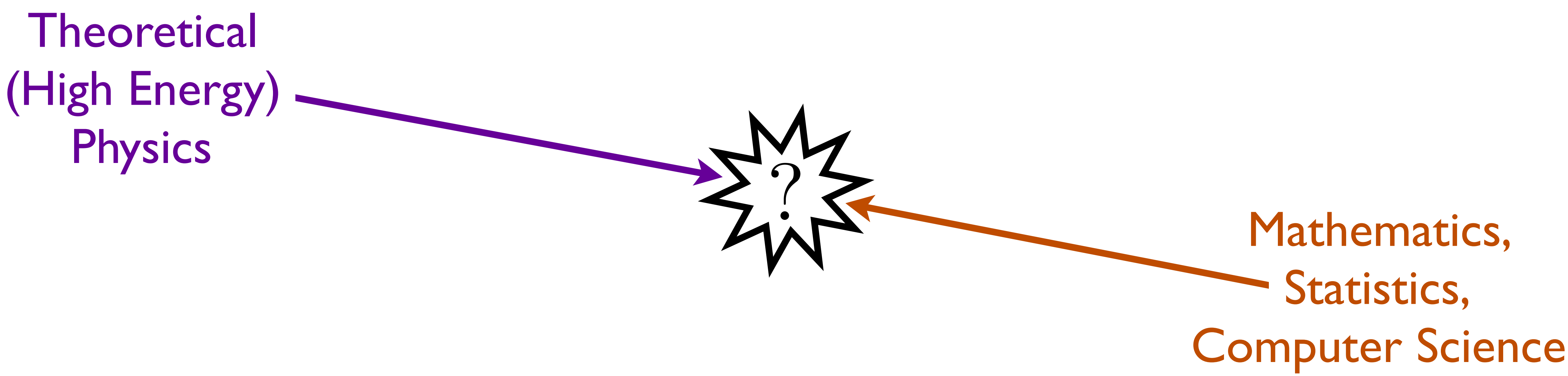


IRC Safe Observable: Insensitive to IR or C emissions

Enforces smooth interpolation between variable-length inputs (i.e. $N \to N{-}1$)

(optionally)

Bottom line: Jets are energy-weighted point clouds


Theoretical
(High Energy)
Physics
?
Mathematics,
Statistics,
Computer Science

Patrick Komiske

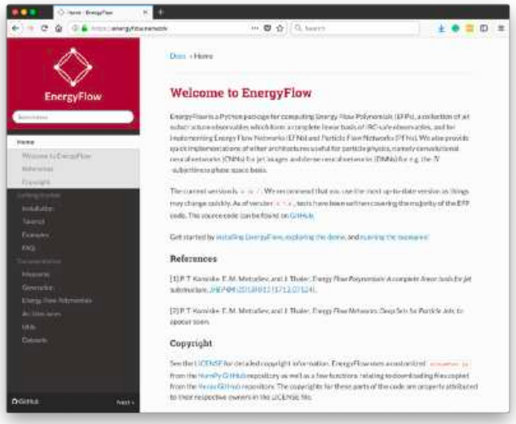
Eric Metodiev

\+ Since Dec 13!

Theoretical (High Energy) Physics

?

Mathematics, Statistics, Computer Science



Energy Flow Networks

https://energyflow.network/

Theory Prior: Dissect Jets with Addition

Additive Observable: $\mathcal{O} = \sum_{i \in \text{jet}} \Phi(E_i, \vec{p}_i, \ldots)$

(relevant for factorization in SCET)

Permutation invariant by construction
Easily adapts to variable-length inputs
Can approximate Φ with neural networks
Can incorporate additional particle properties
Linear runtime in number of particles

Additive Safe Observable: $\mathcal{O} = \sum_{i \in \text{jet}} E_i \, \Phi(\hat{p}_i)$ $\qquad \hat{p}_i = \dfrac{\vec{p}_i}{E_i}$

IRC safety guaranteed by energy weighting

Conjectured Generalization

Arbitrary permutation-symmetric observable?

$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

Latent Additive Observables

$$\mathcal{O}_a = \sum_{i \in \text{jet}} \Phi_a(E_i, \vec{p}_i, \ldots)$$

(see backup for detailed architecture)

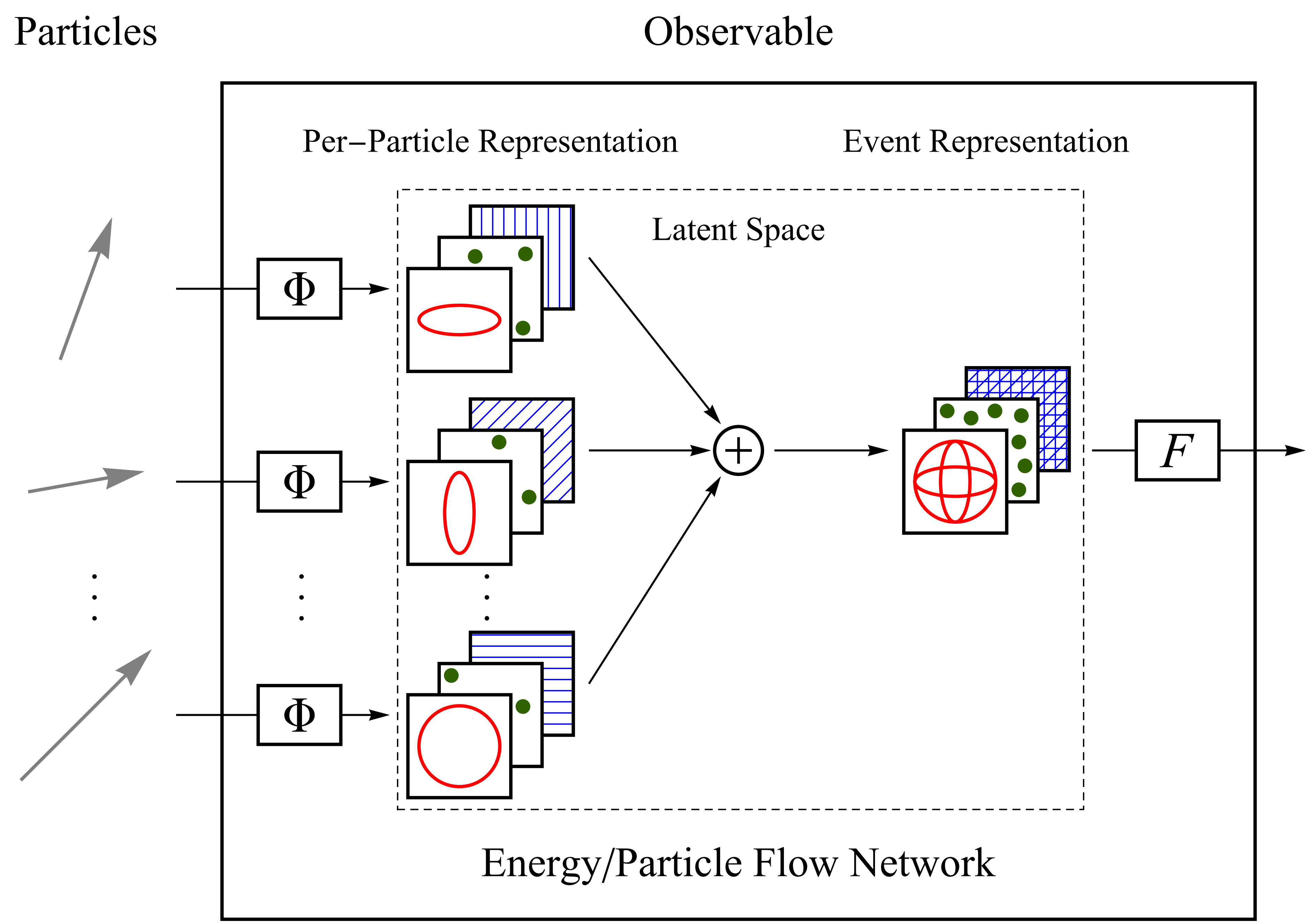
Energy / Particle Flow Networks

IRC-safe Φ

General Φ

[Komiske, Metodiev, JDT, 1810.05165]

Conjectured Generalization

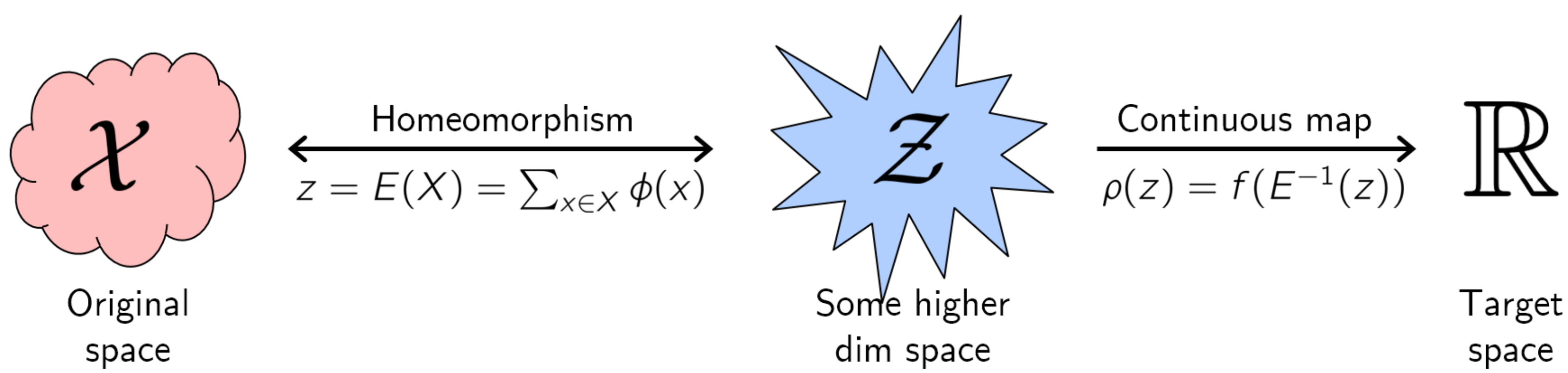


[Komiske, Metodiev, JDT, 1810.05165]

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

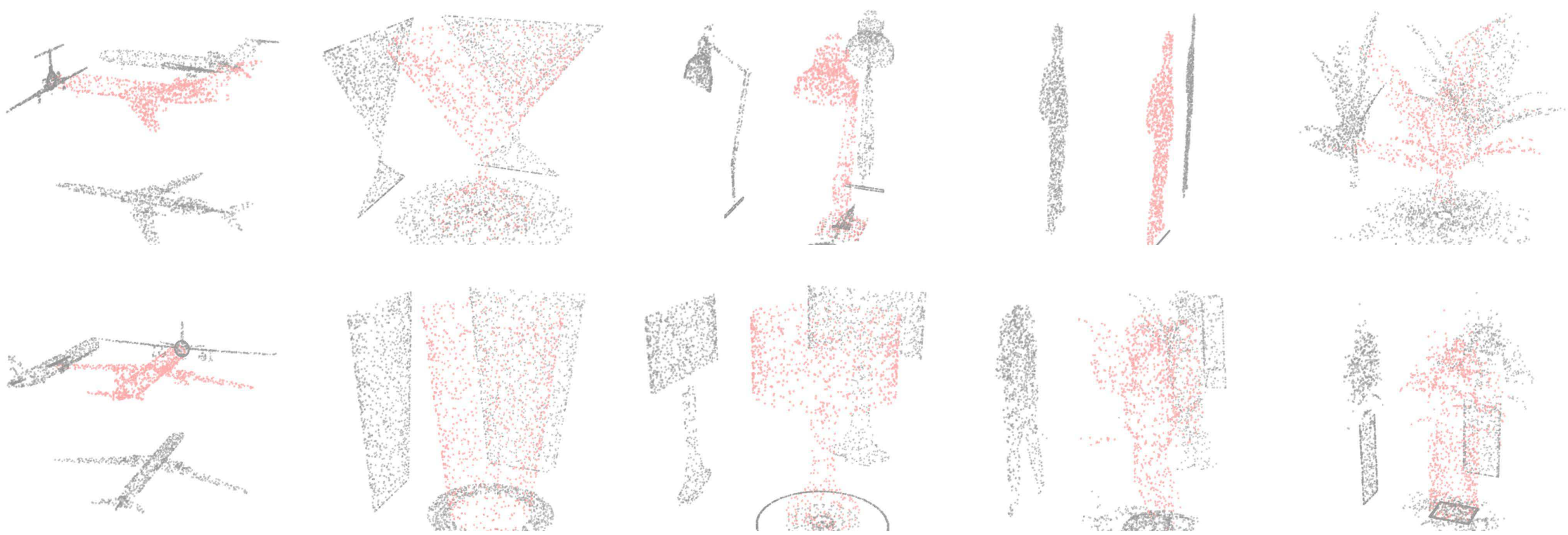
Generic Observable

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

Deep Sets for…

Celebrity Face Anomaly Detection

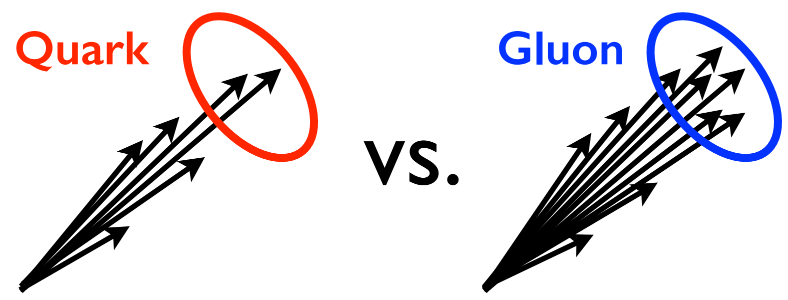
Point Cloud Classification



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

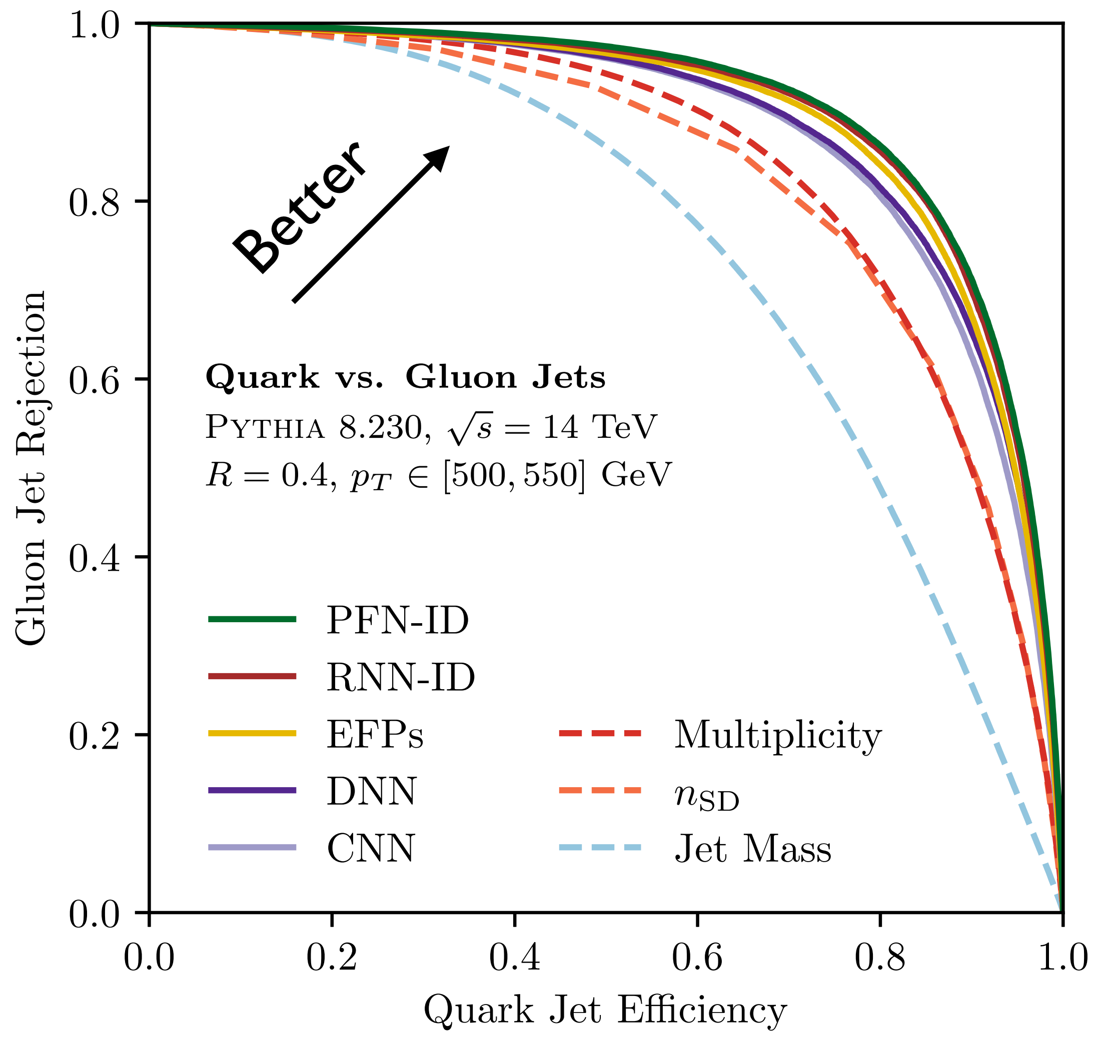
Deep Sets for Particle Jets

Q vs. G: The "Hello, World!" of jet classification



Theory prior:
Network must be exploiting
IRC singularity structure of QCD

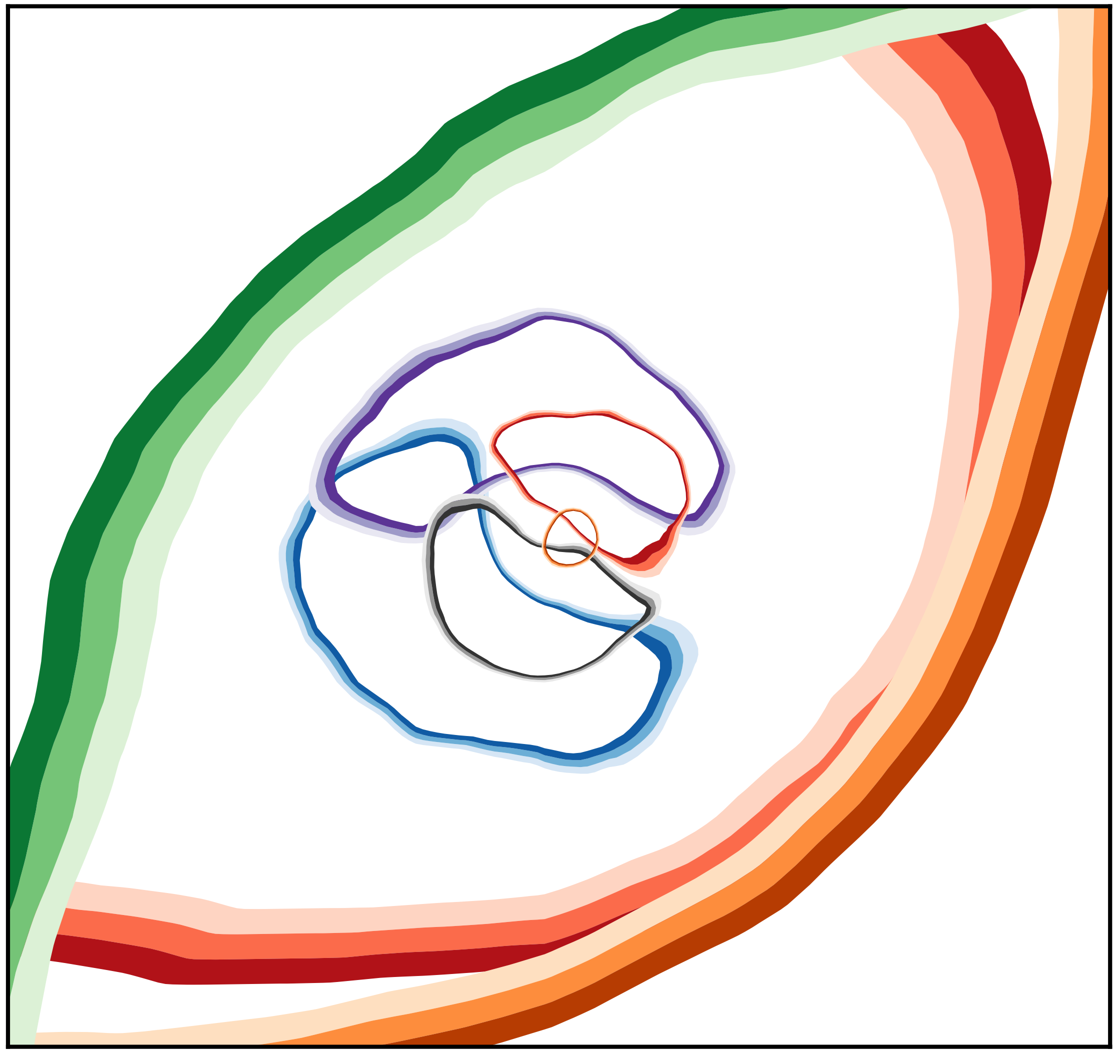
Competitive with previous methods



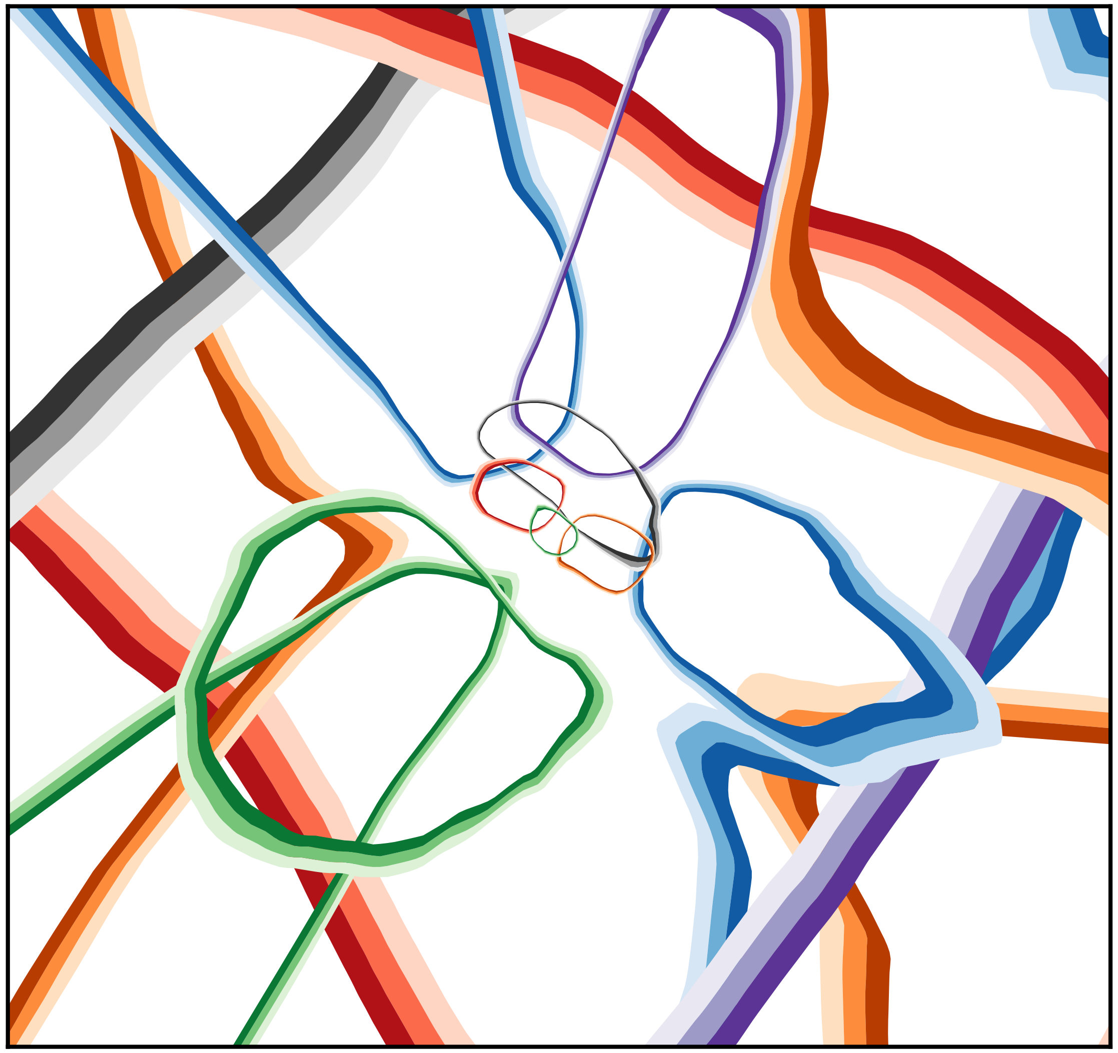
[Komiske, Metodiev, JDT, 1810.05165]

Psychedelic Network Visualization

Latent Dimension 8

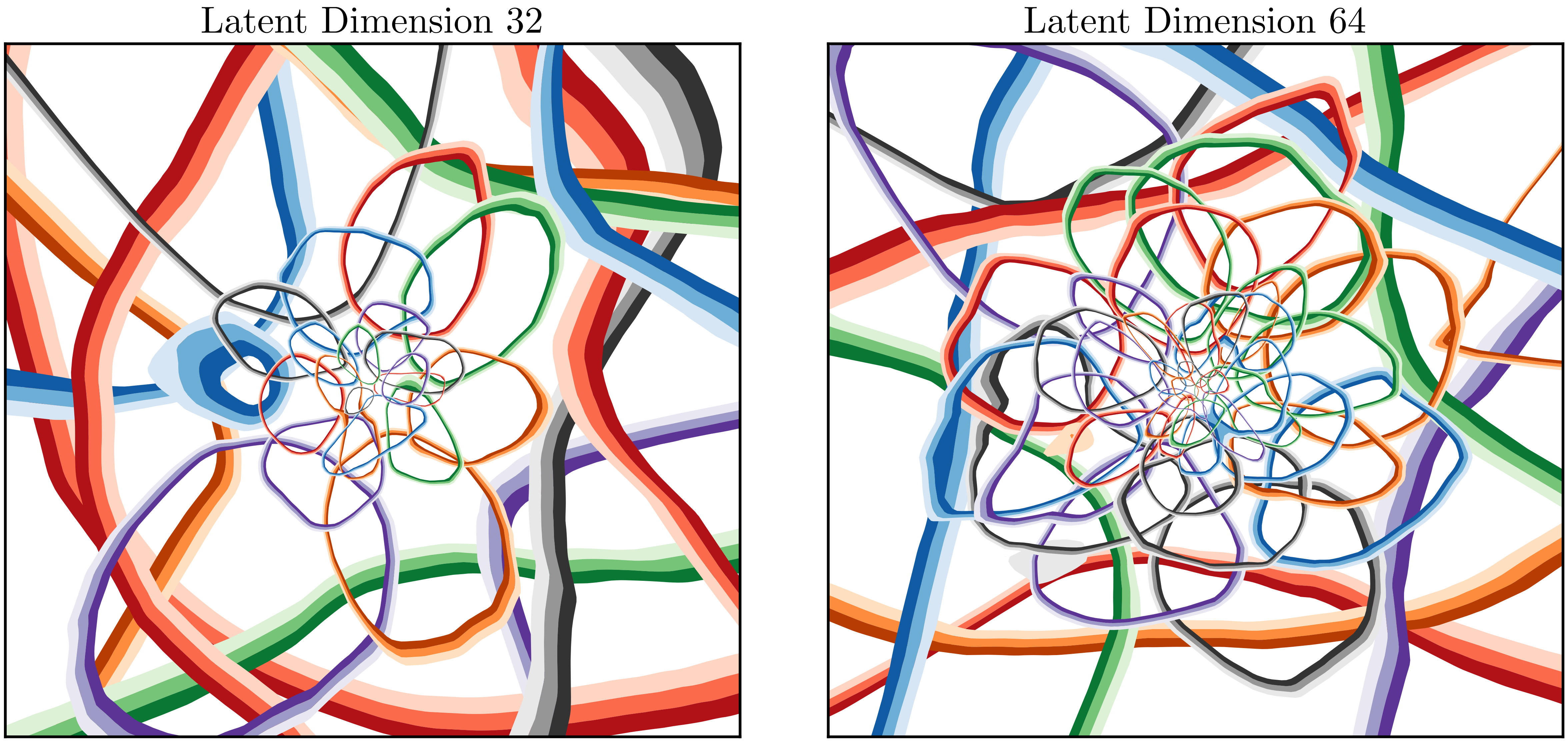


Latent Dimension 16



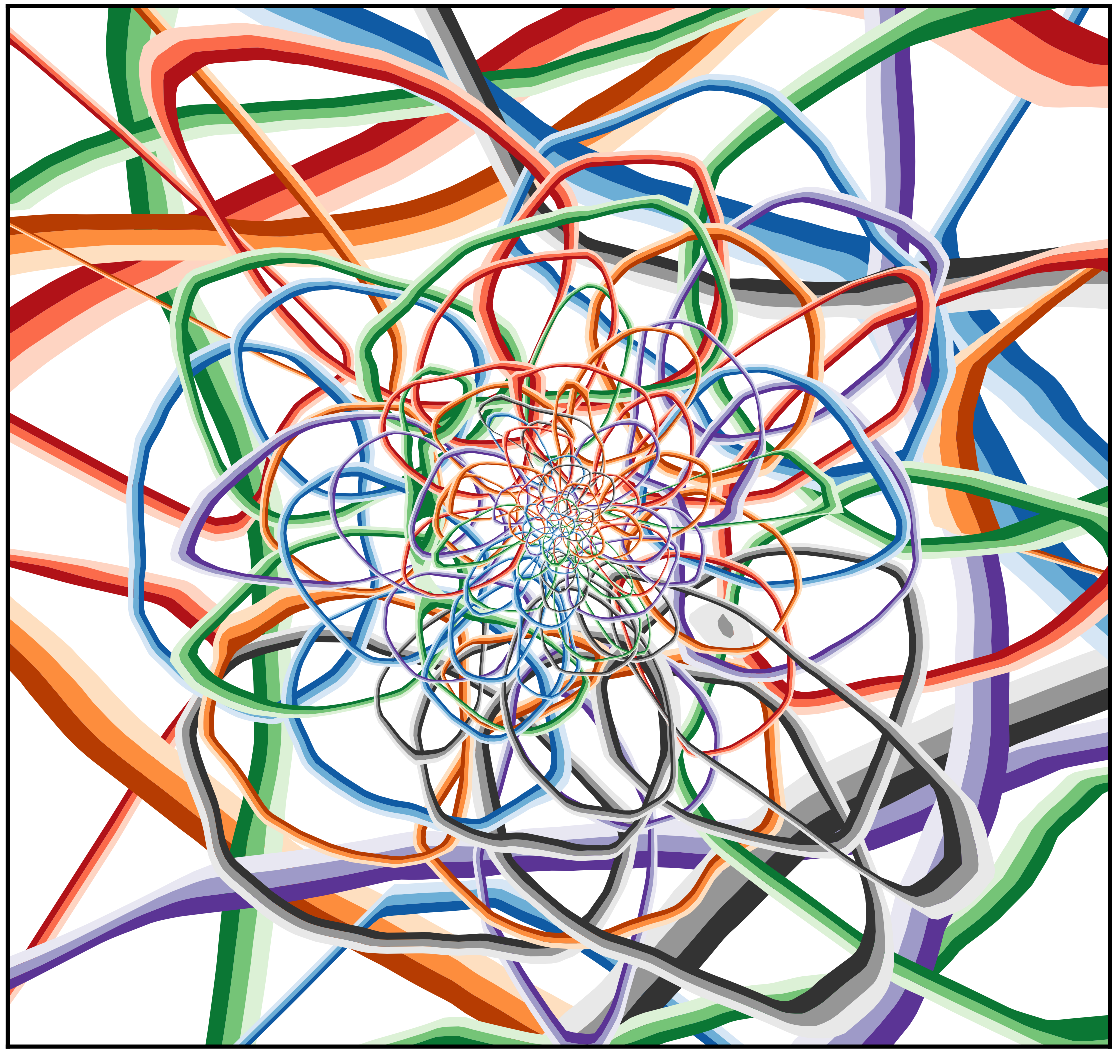
(see backup for how these are made)

Psychedelic Network Visualization

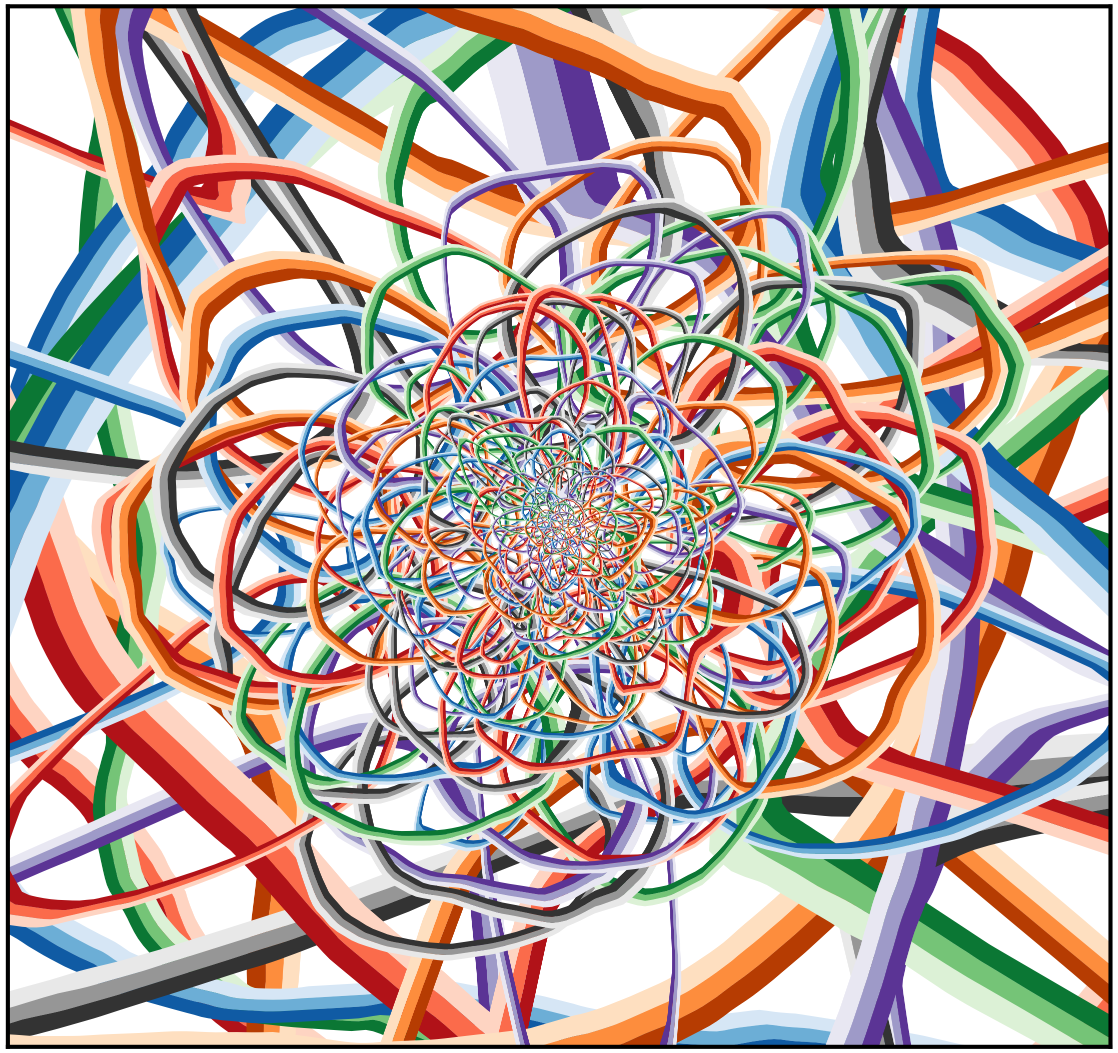


Psychedelic Network Visualization

Latent Dimension 128



Latent Dimension 256



Collinear singularity of QCD!

(more in backup)

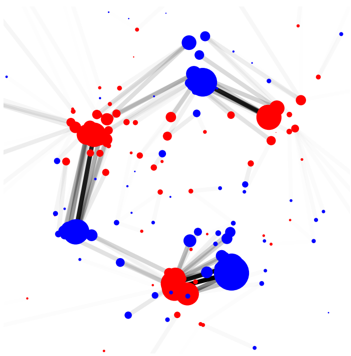
The Broader Lesson

Advances in hep-th, -ph, -ex	$\Leftrightarrow$	*Advances in cs.LG, stat.ML*
Particle Flow Networks	$\longleftarrow$ Point Clouds $\longleftarrow$	Deep Sets
Energy Flow Networks	$\longrightarrow$ Weighted Point Clouds $\longrightarrow$	??

Particle physics is a fascinating domain with rich data sets, established algorithms, and strong theory priors

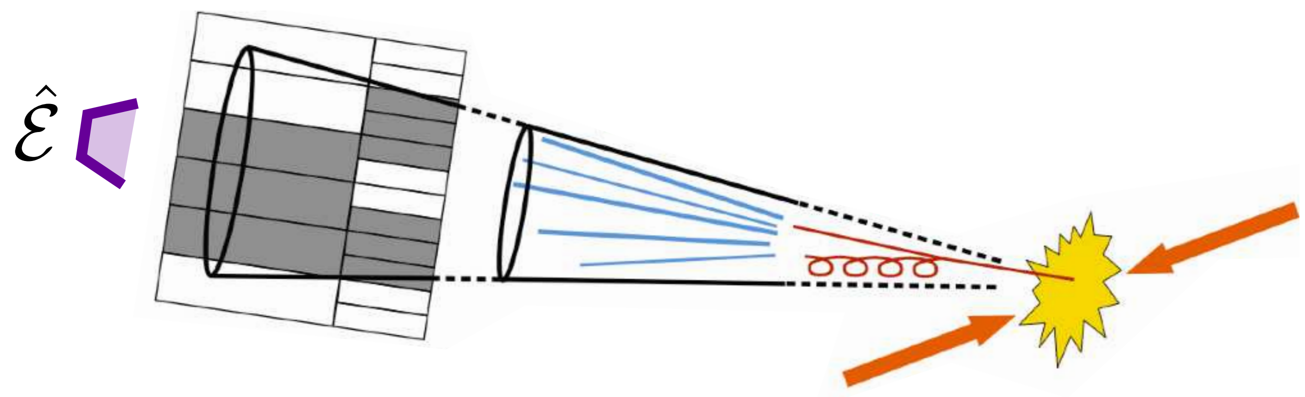
Particle Physics Primer

Point Clouds & Energy Flow Networks

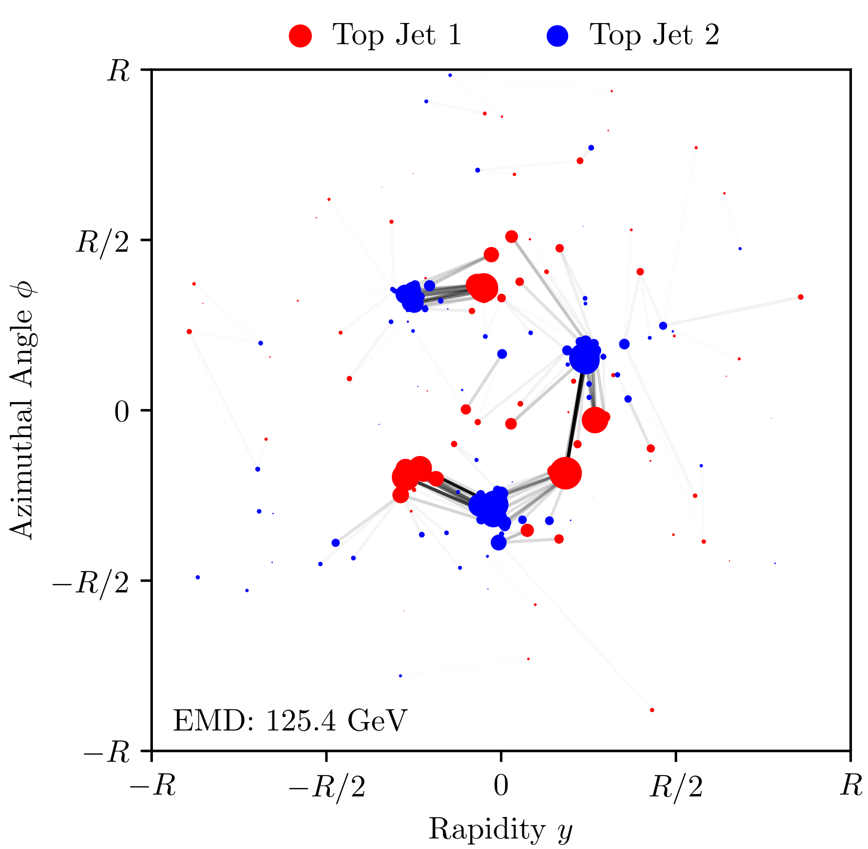


(The Metric Space of Collider Events)

Theory Prior: Keep a "Safe" Distance



Difference in energy flow ↔ *IRC-safe similarity measure*



e.g. Earth Mover's Distance

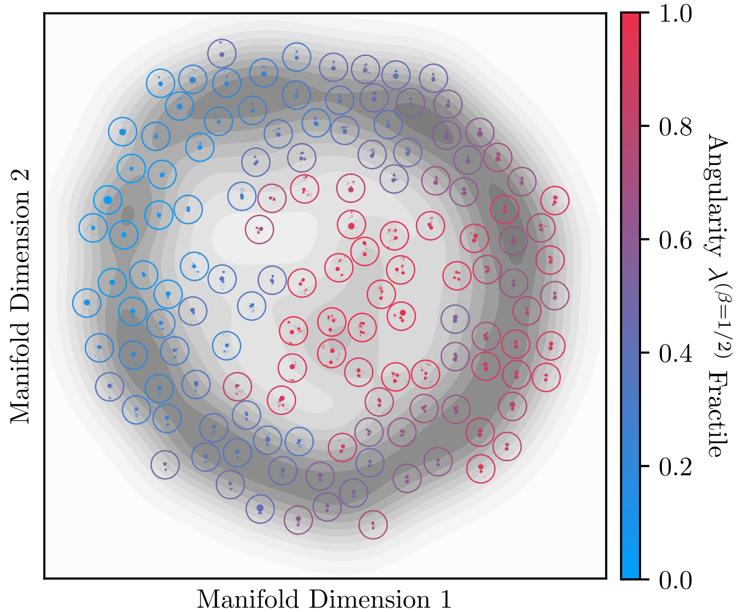
$$\mathrm{EMD}(\mathcal{E}_1, \mathcal{E}_2) = \min_{\{f_{ij}\}} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} f_{ij} \frac{\theta_{ij}}{R} + \left| E_{\mathrm{tot}}^{(1)} - E_{\mathrm{tot}}^{(2)} \right|$$

A metric for weighted point clouds

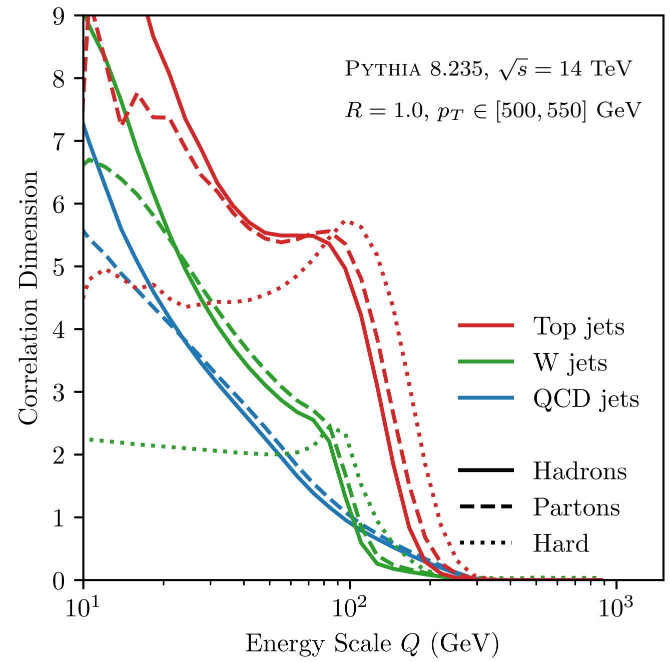
[see Rubner, Tomasi, Guibas, ICJV 2000; Pele, Werman, ECCV 2008]

Rich Opportunities for Data Exploration

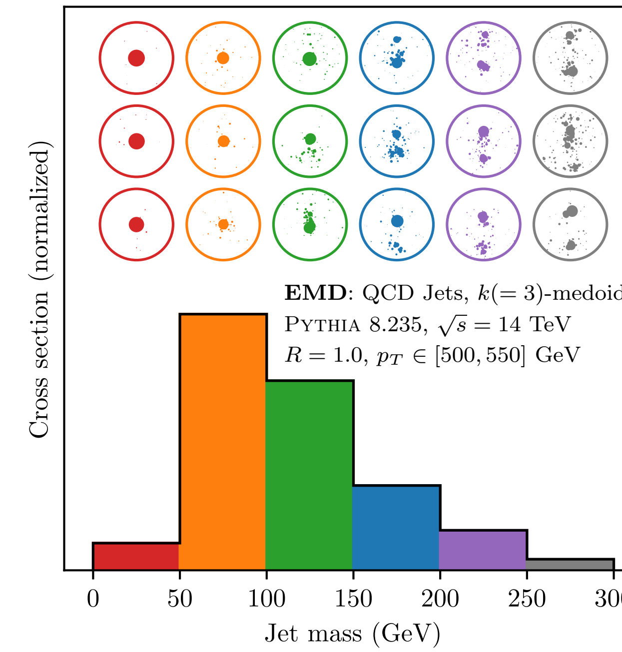
t-SNE



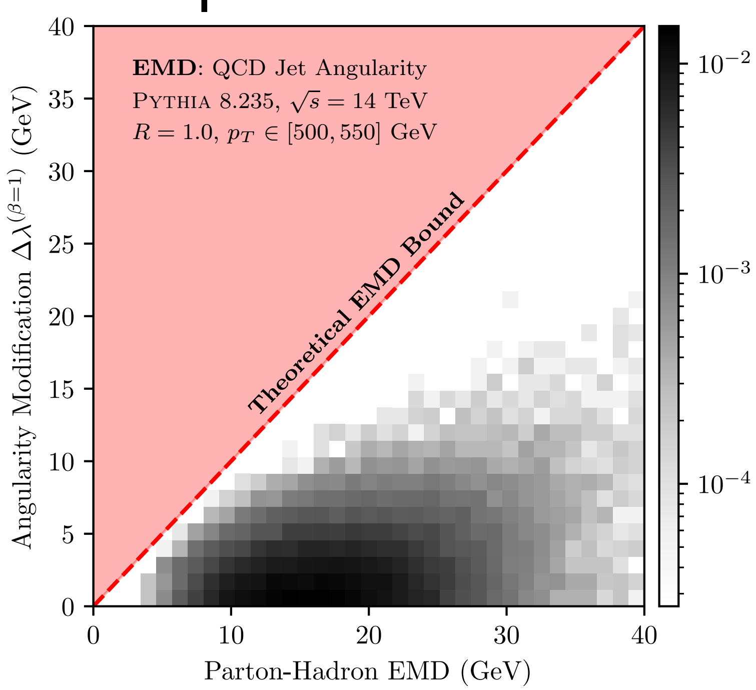
Fractal Dimensions



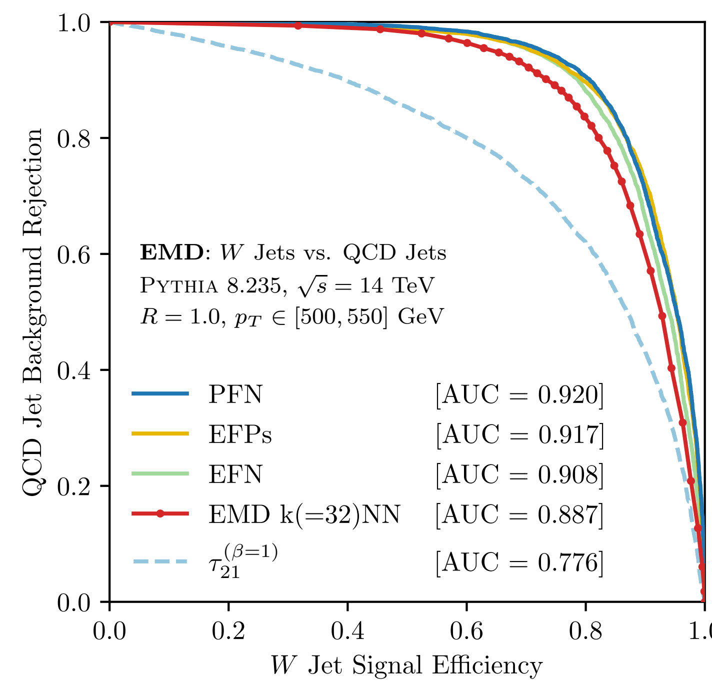
k-medoid Visualization



Lipschitz Bounds



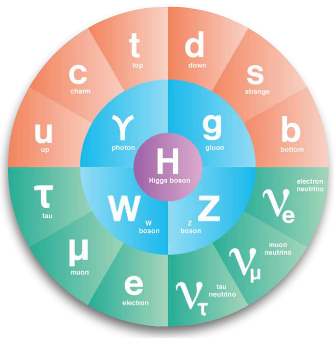
kNN Classification



Plus non-parametric density estimation, anomaly detection, autoencoders, …

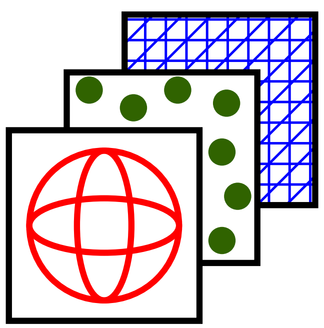
[Komiske, Metodiev, JDT, in progress]

Summary



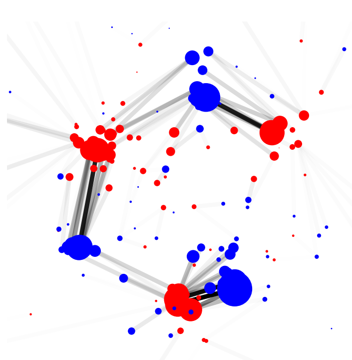
Particle Physics Primer

A rich domain with many machine learning opportunities



Point Clouds & Energy Flow Networks

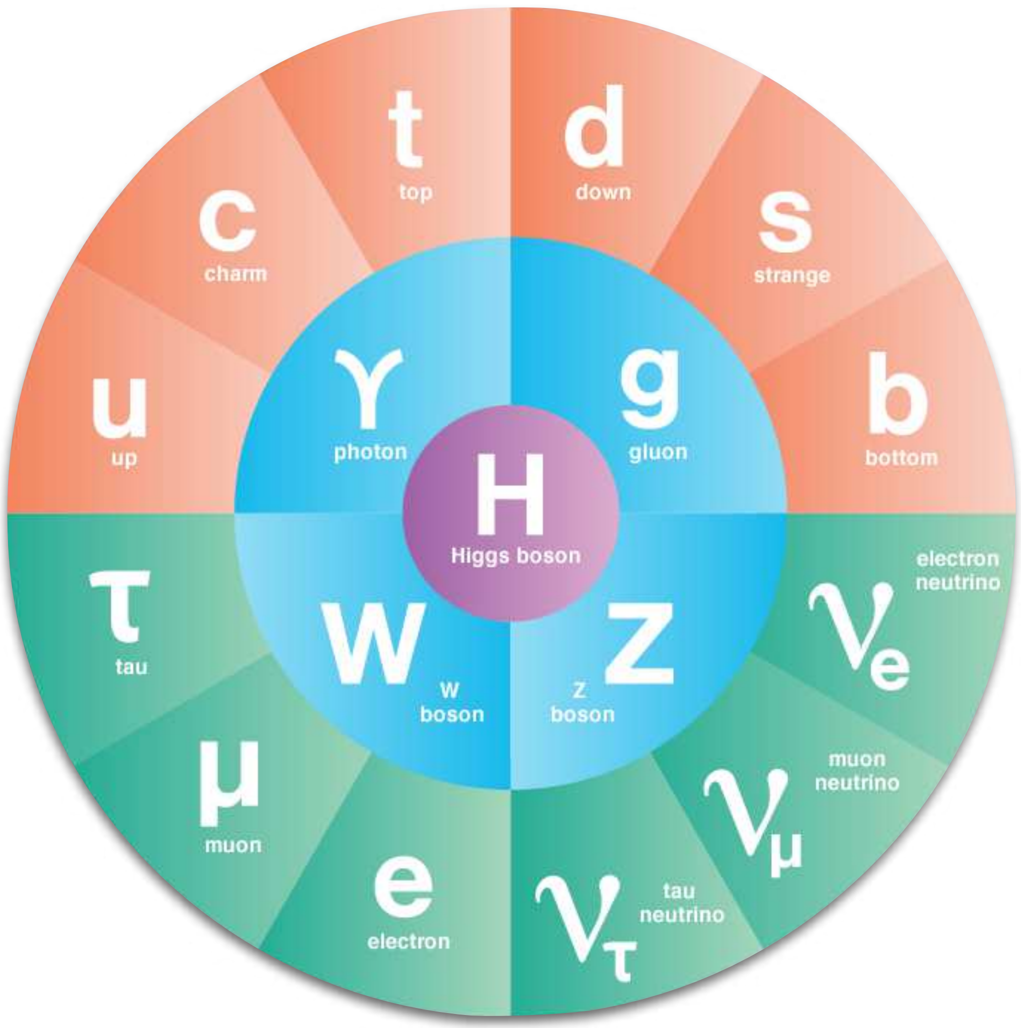
A new architecture for weighted point clouds

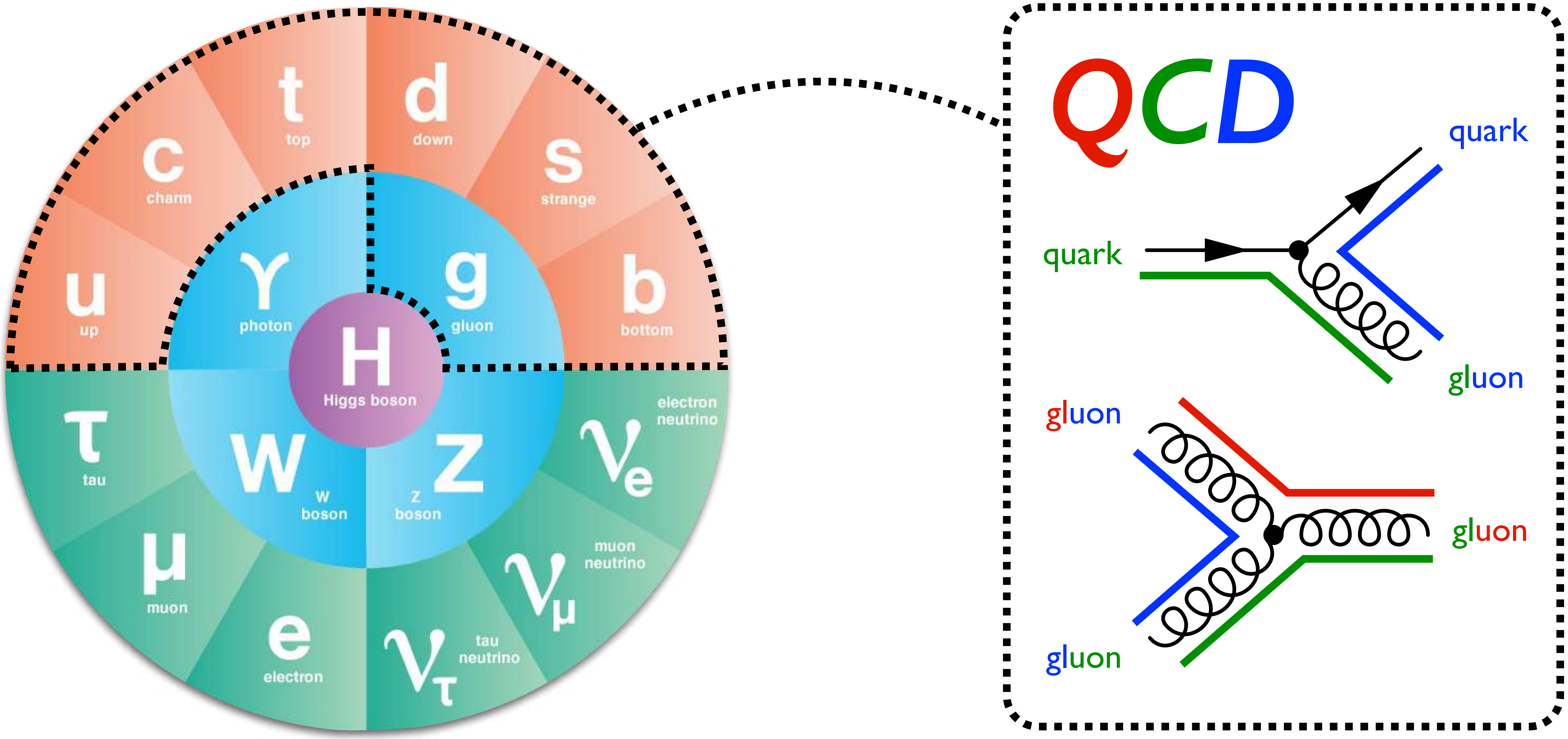


(The Metric Space of Collider Events)

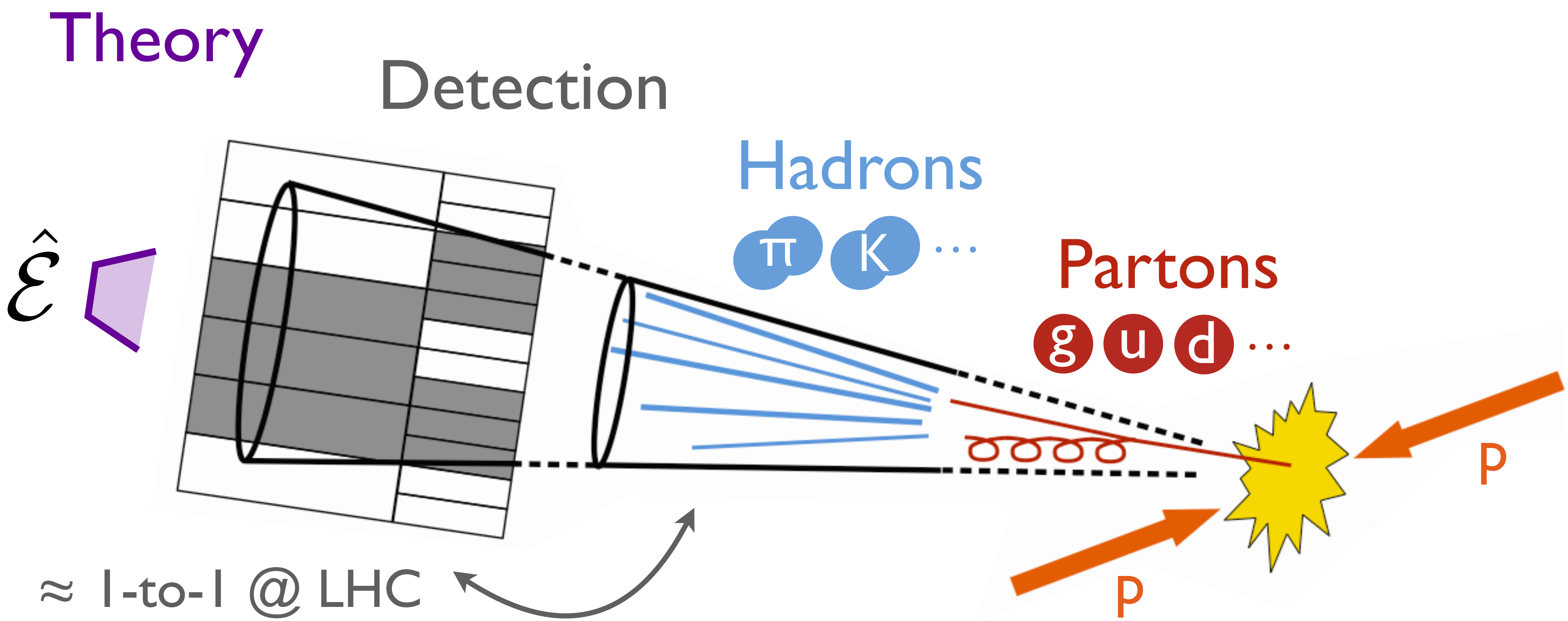
Using geometry for data exploration in particle physics and beyond

Backup Slides


t
top
d
down
c
charm
s
strange
u
up
b
bottom
γ
photon
g
gluon
H
Higgs boson
W
W
boson
Z
Z
boson
τ
tau
electron
neutrino
ν_e
μ
muon
muon
neutrino
ν_μ
e
electron
ν_τ
tau
neutrino



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
QCD
quark
quark
gluon
gluon
gluon
gluon



Stress-Energy Flow Operator:

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

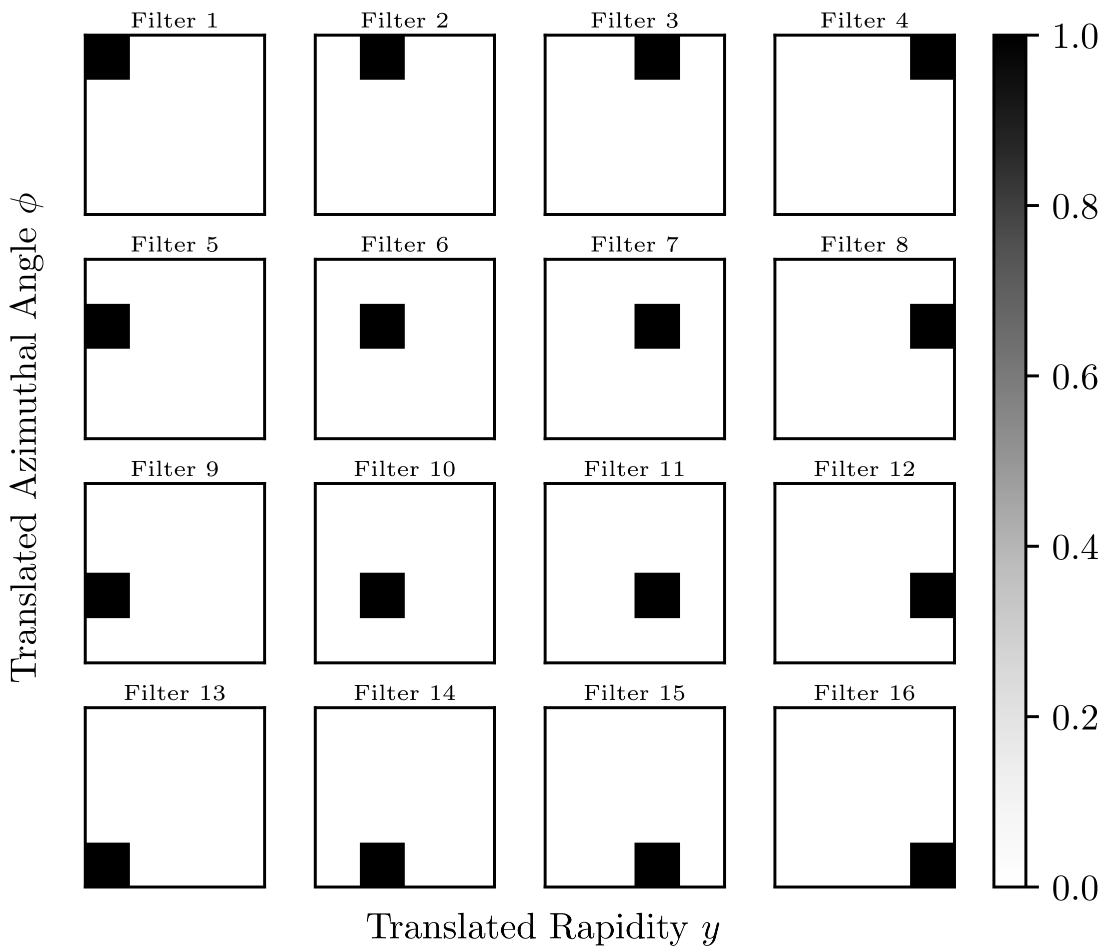
[Sveshnikov, Tkachov, hep-ph/9512370; Hofman, Maldacena, 0803.1467; Mateu, Stewart, JDT, 1209.3781]

Latent Space Visualization

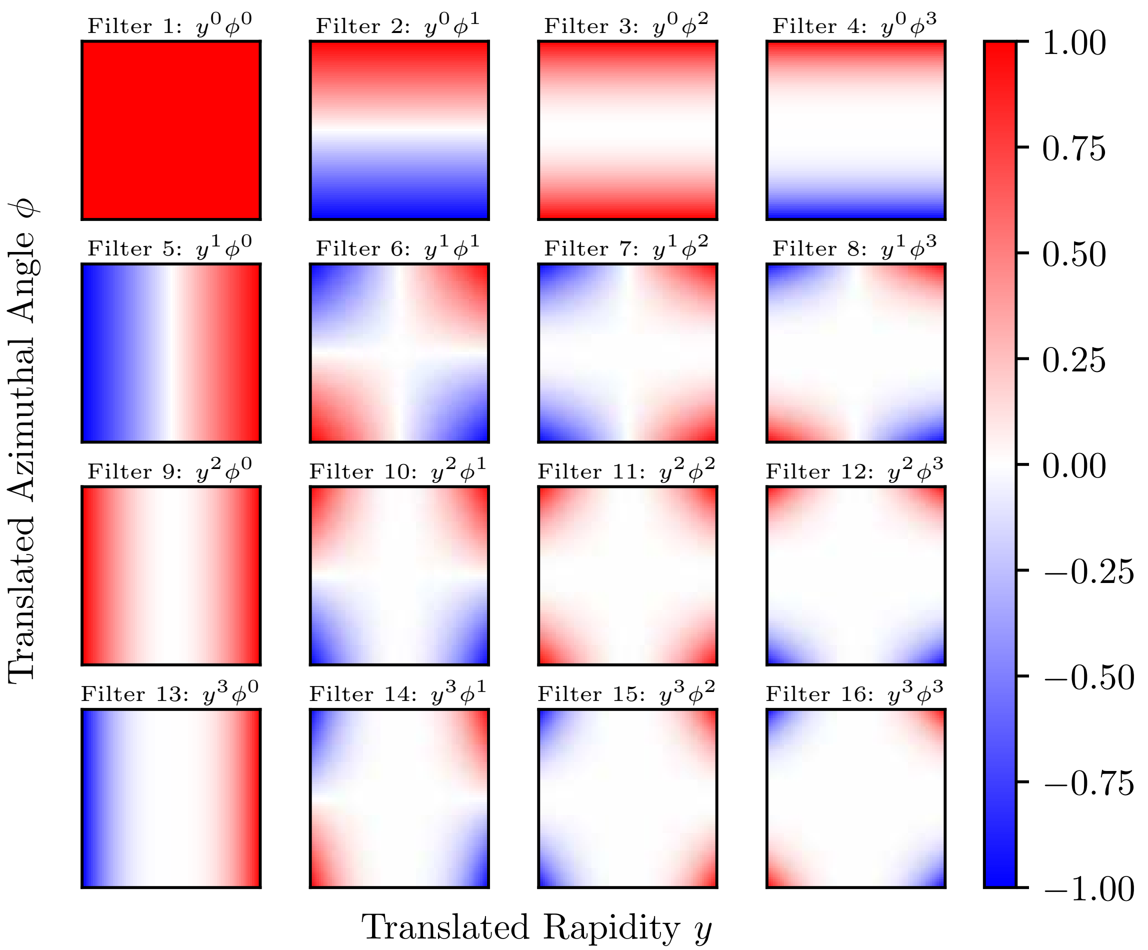
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

IRC-safe: $\mathcal{O}_a = \sum_{i \in \text{jet}} E_i \, \Phi_a(\hat{p}_i)$

Calorimeter Pixels



Radiation Moments

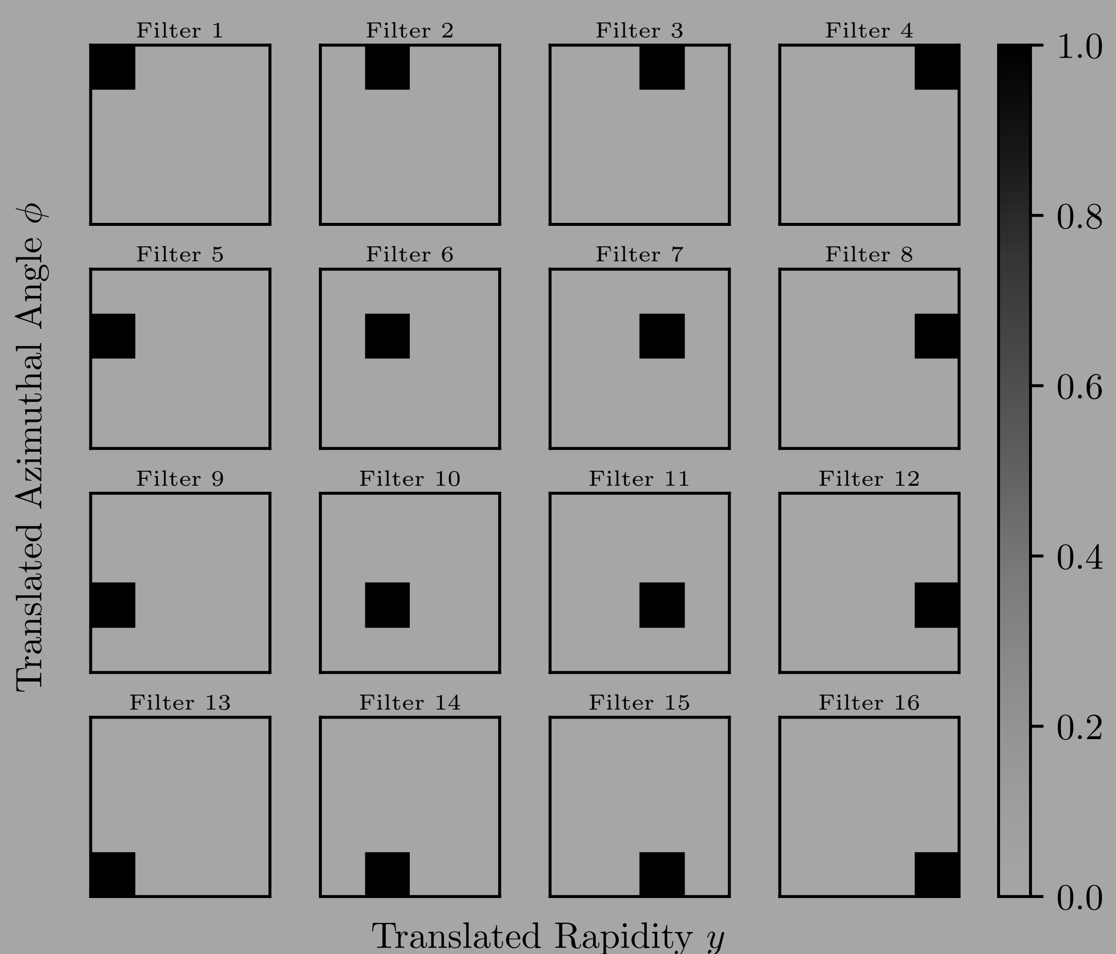


Latent Space Visualization

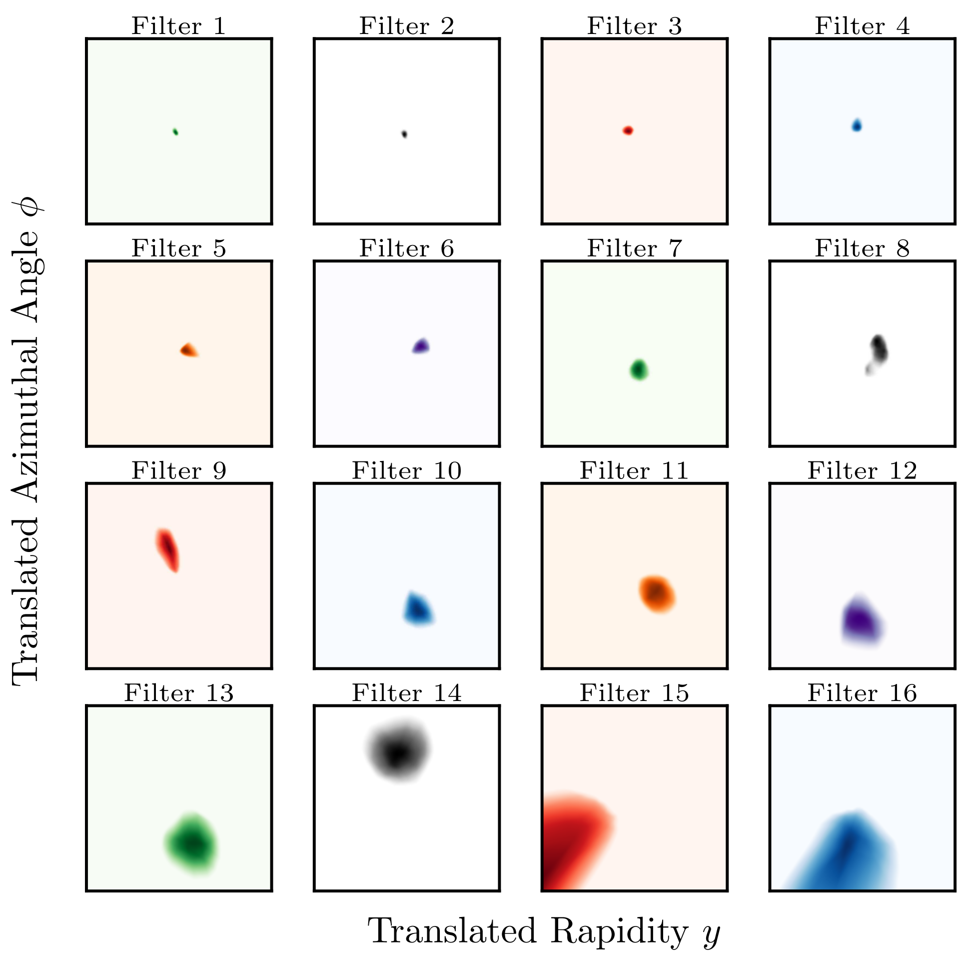
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

IRC-safe: $\mathcal{O}_a = \sum_{i \in \mathrm{jet}} E_i \, \Phi_a(\hat{p}_i)$

Calorimeter Pixels

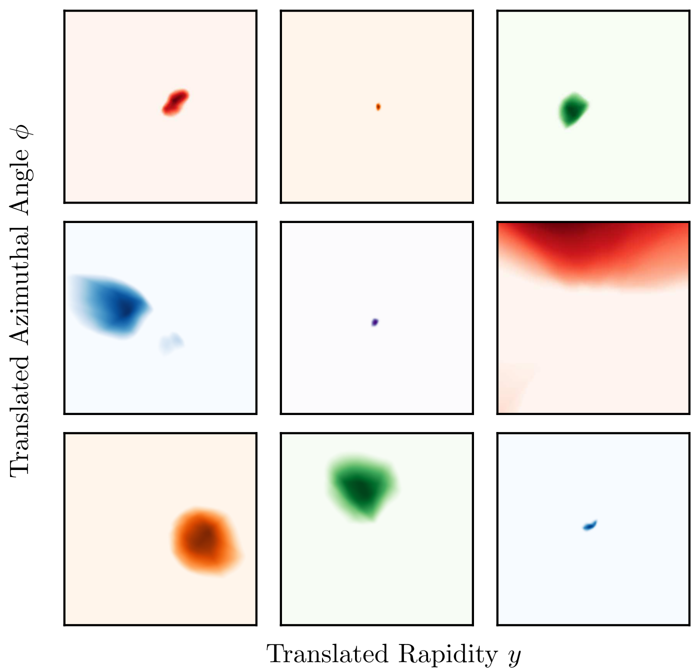


EFNs: Dynamic Pixelation

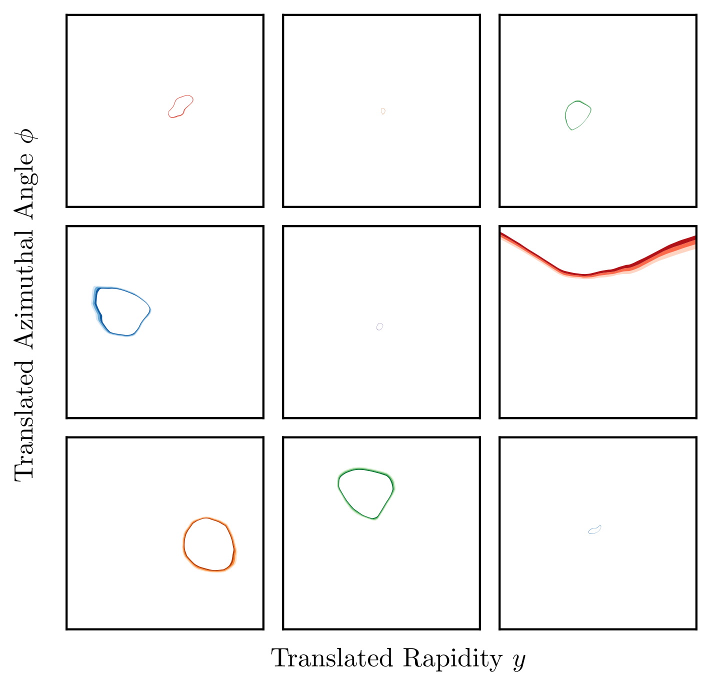


Psychedelic Network Visualization

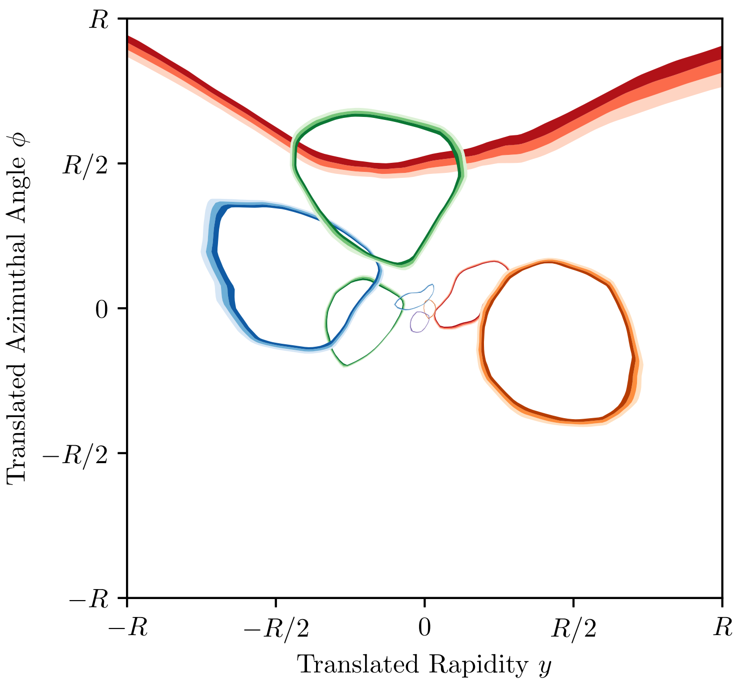
Latent Filters



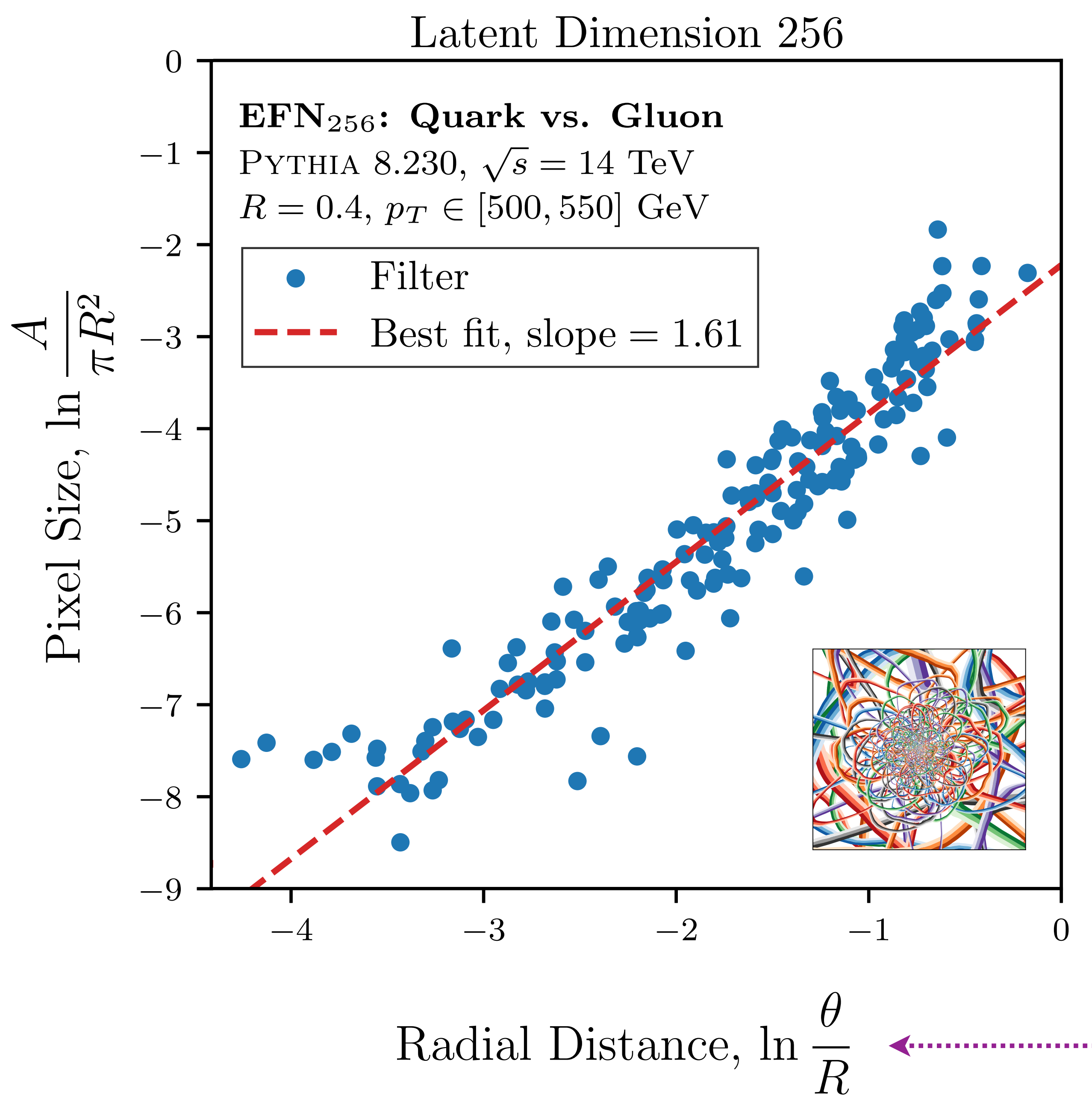
50% Contours

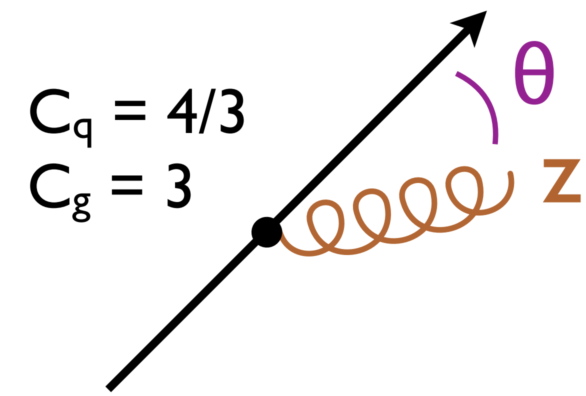


Overlay



Learning the Singularity Structure of QCD

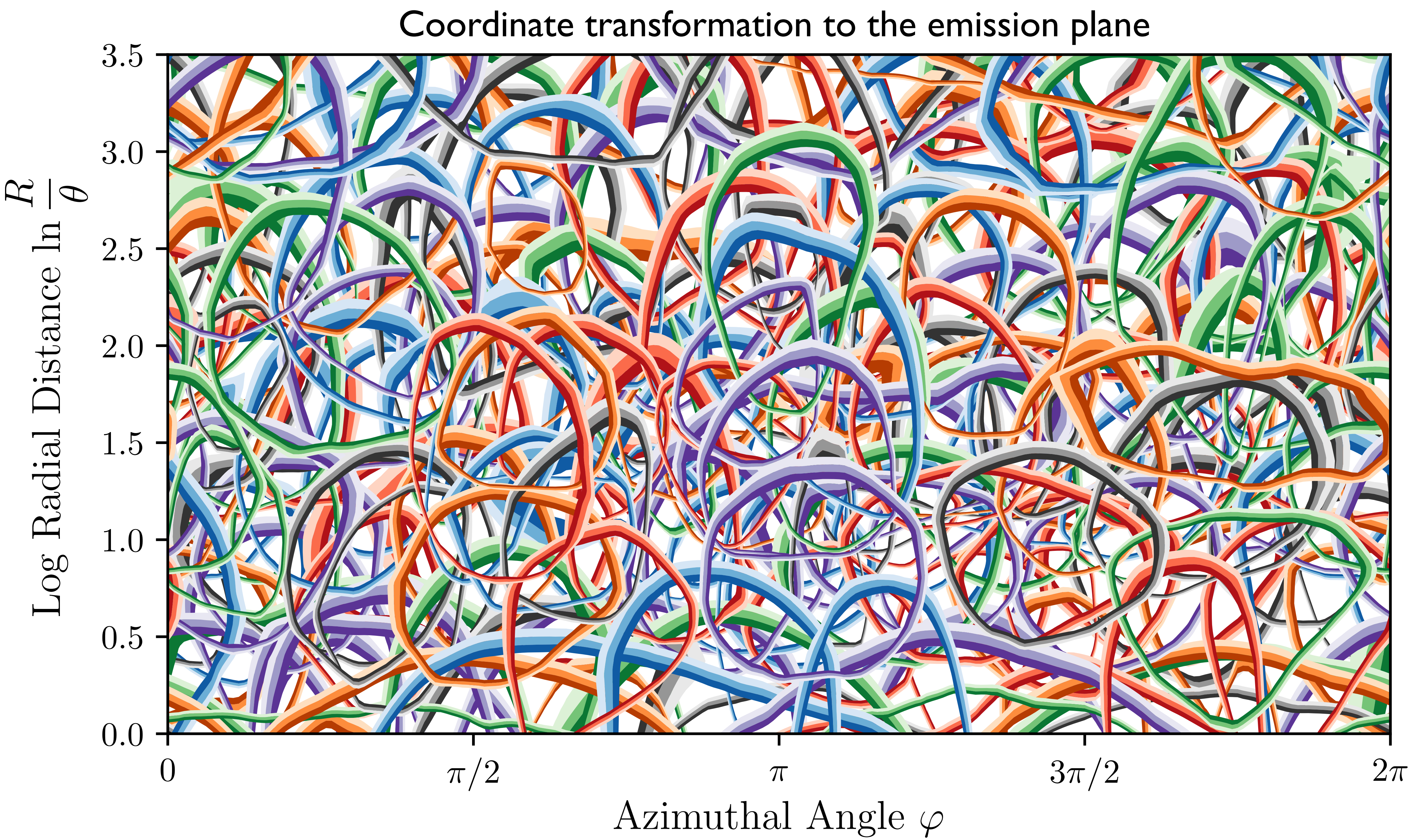




$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{d\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{dz}{z}}_{\text{Soft}}$$

[Komiske, Metodiev, JDT, 1810.05165]

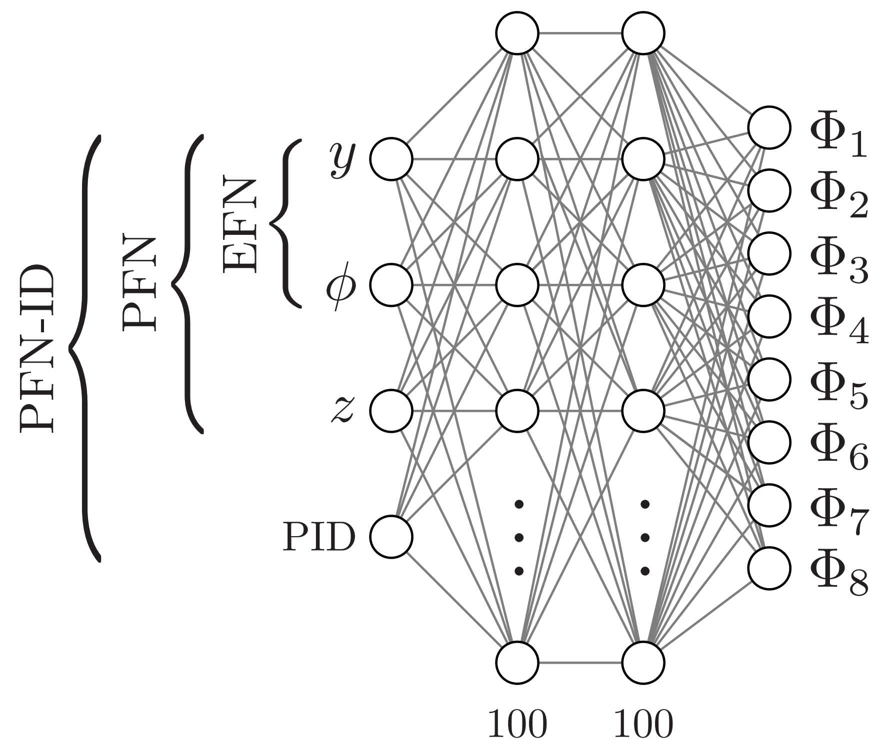
Ready for the MoMA



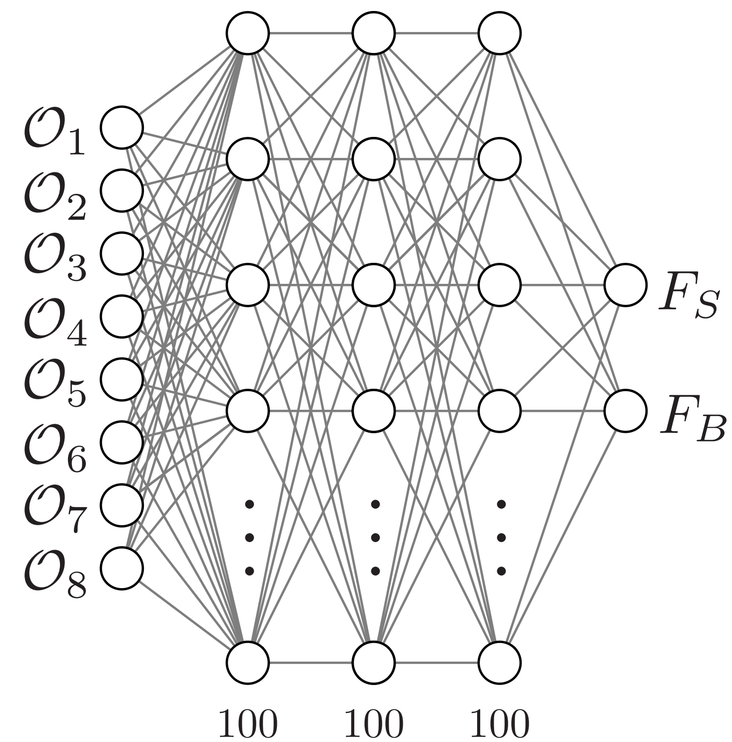
[Komiske, Metodiev, JDT, 1810.05165]

EFN/PFN Architecture Details

Per-Particle:



Latent Combiner: F



Per-Jet (Latent Space):

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

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

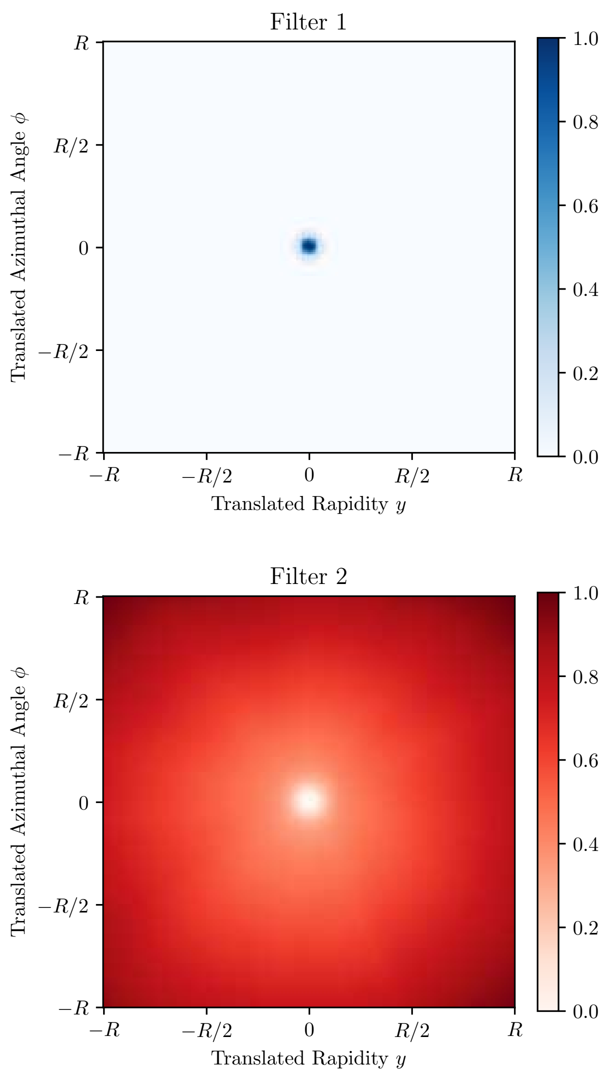
Final Discriminant:

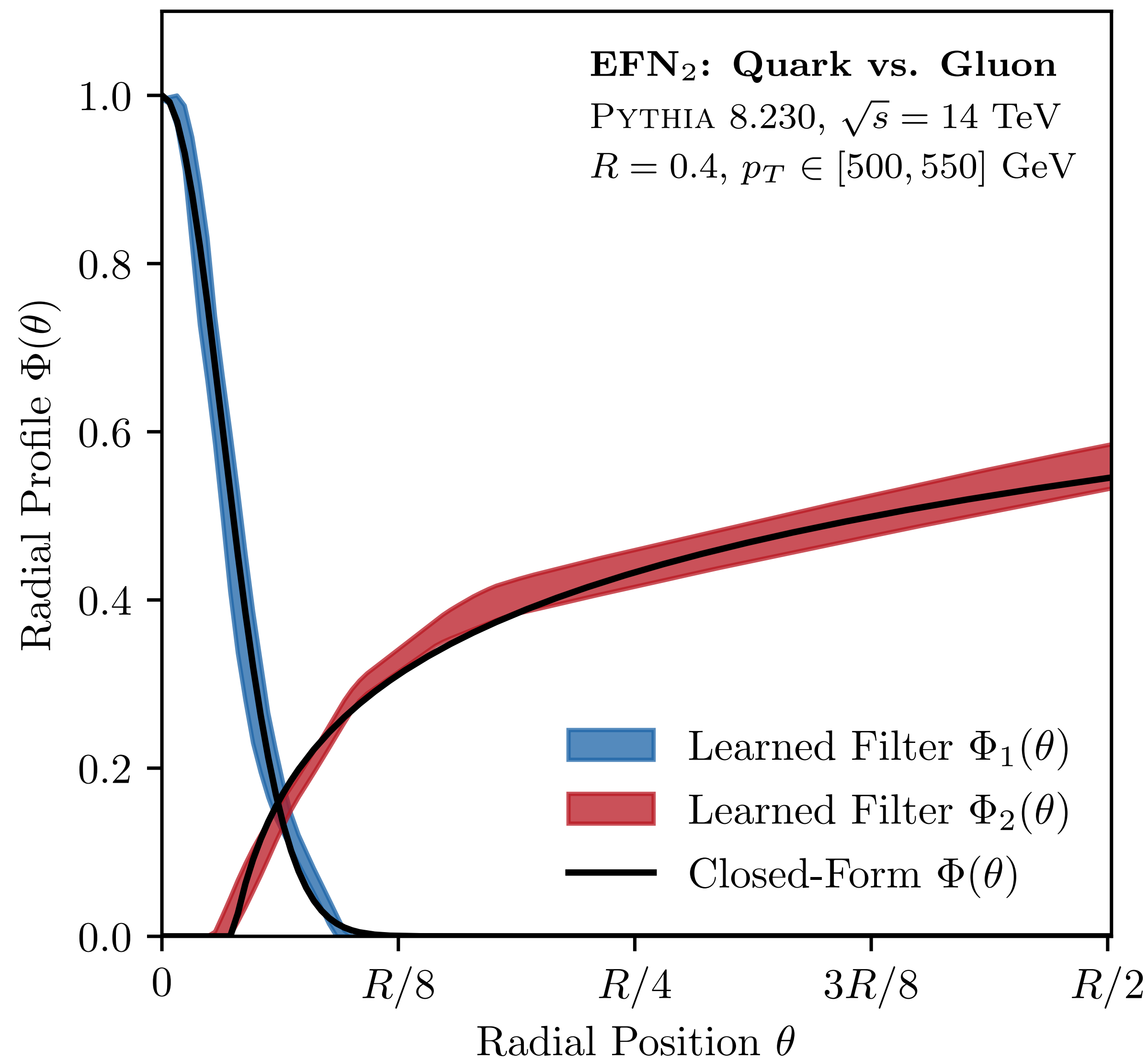
$$\text{softmax}(F_S, F_B)$$

"What is the Machine Learning?"

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

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

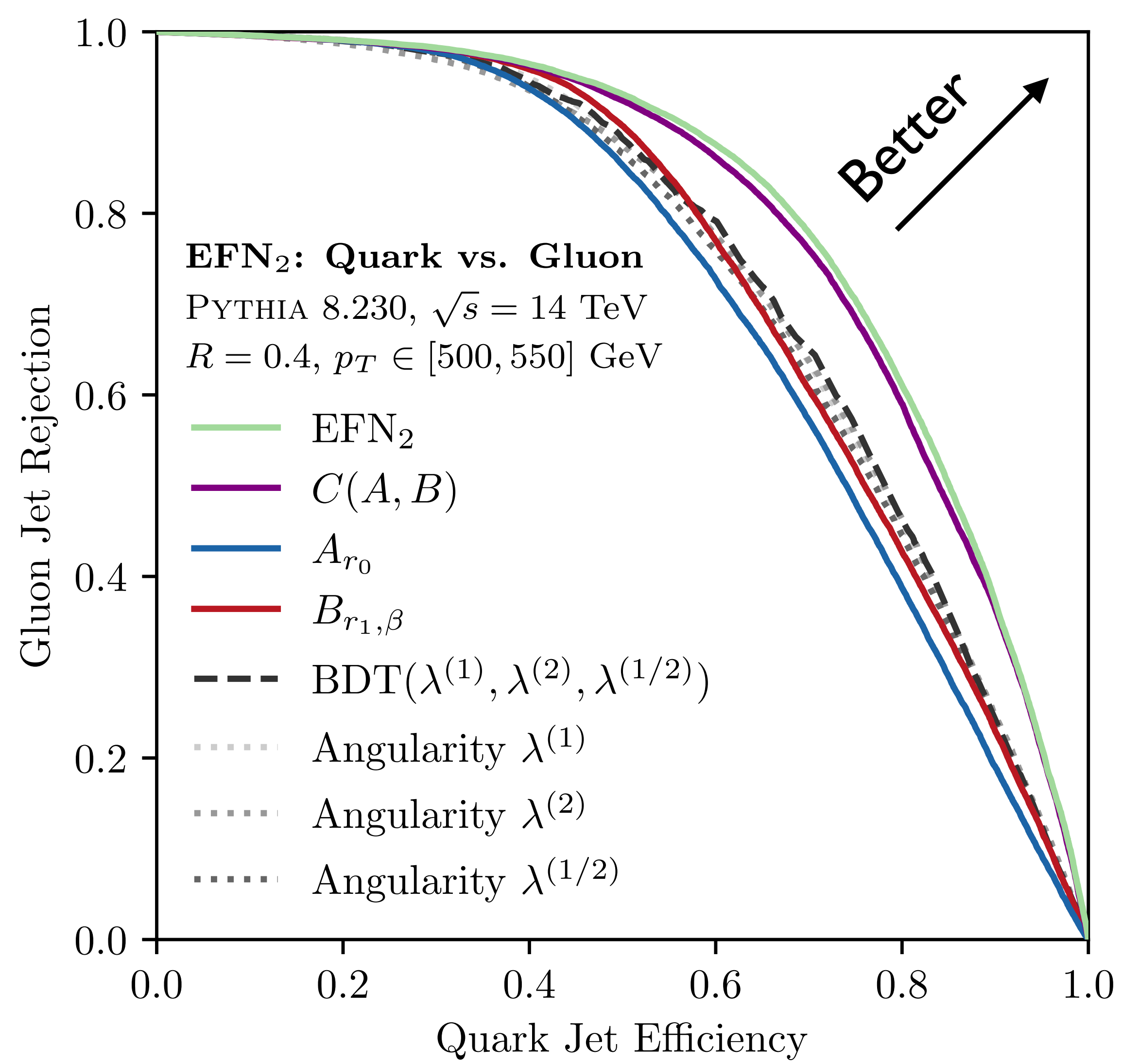




"What is the Machine Learning?"

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

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

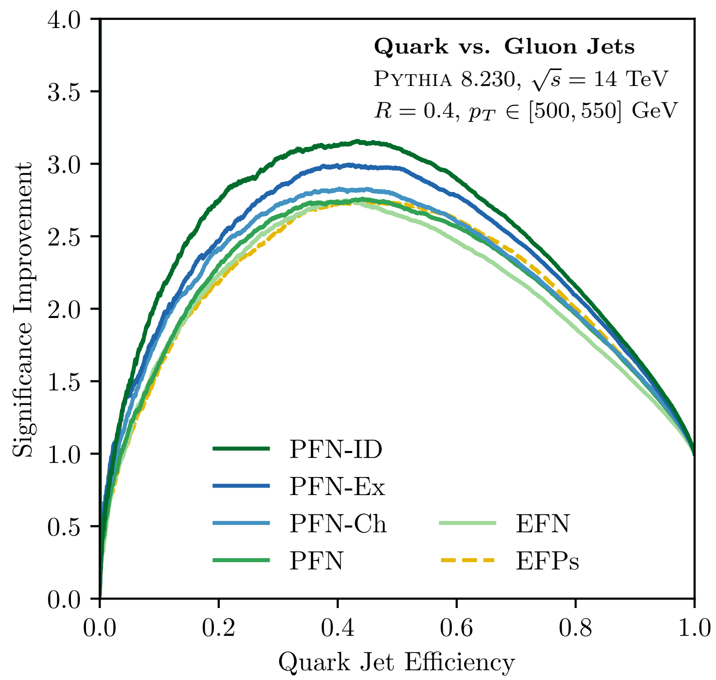


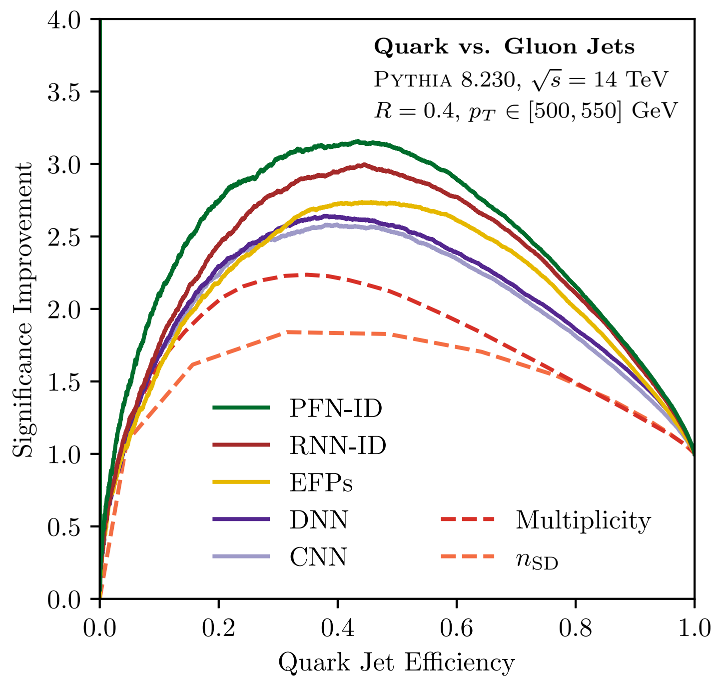
EFN outperformed a domain expert (i.e. me)

But we successfully reverse engineered the machine

More Quark/Gluon Performance

Model	AUC	$1/\varepsilon_g$ at $\varepsilon_q = 50\%$
PFN-ID	$\mathbf{0.9052} \pm 0.0007$	$\mathbf{37.4} \pm 0.7$
PFN-Ex	0.9005 ± 0.0003	34.7 ± 0.4
PFN-Ch	0.8924 ± 0.0001	31.2 ± 0.3
PFN	0.8911 ± 0.0008	30.8 ± 0.4
EFN	0.8824 ± 0.0005	28.6 ± 0.3
RNN-ID	0.9010	34.4
RNN	0.8899	30.5
EFP	0.8919	29.7
DNN	0.8849	26.4
CNN	0.8781	25.5
M	0.8401	19.0
n_{SD}	0.8297	14.2
m	0.7401	7.2



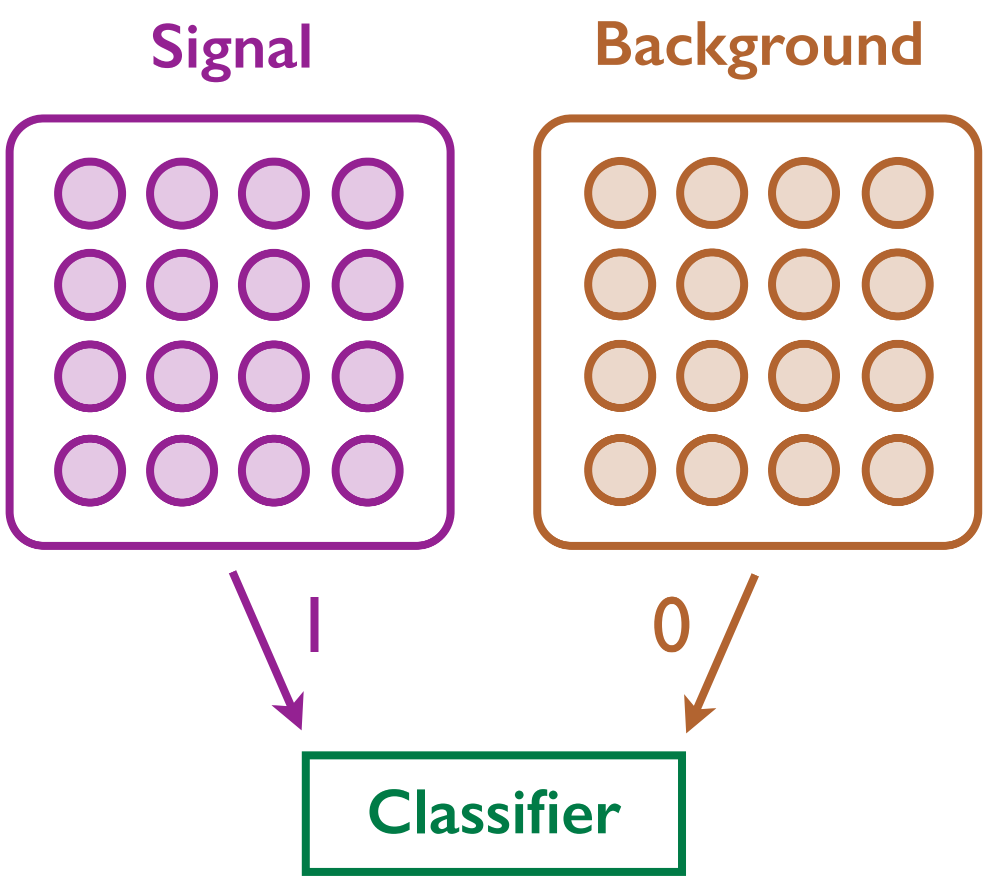


A Cartoon of Machine Learning

For fully-supervised jet classification

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

Classifier Inputs



Minimize Loss Function

(assuming infinite training sets,
and flexible enough functional form)

$$h(\vec{x}) = \frac{p_{\mathrm{sig}}(\vec{x})}{p_{\mathrm{sig}}(\vec{x}) + p_{\mathrm{bkgd}}(\vec{x})}$$

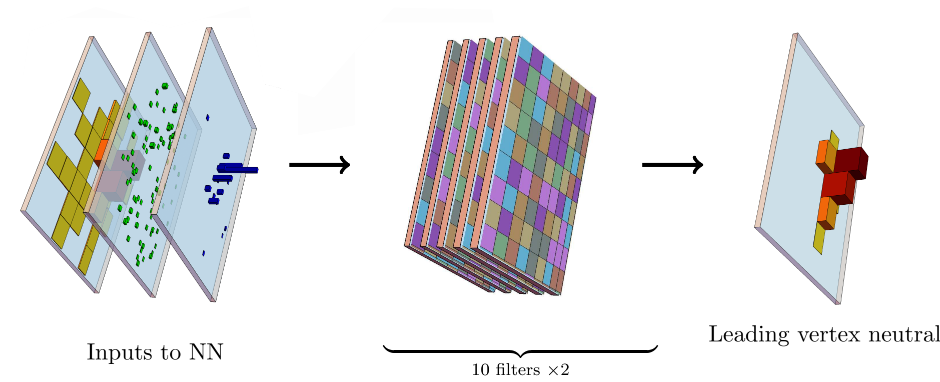
Optimal Classifier (Neyman–Pearson)

Beyond Classification

PUMML

Pileup Mitigation

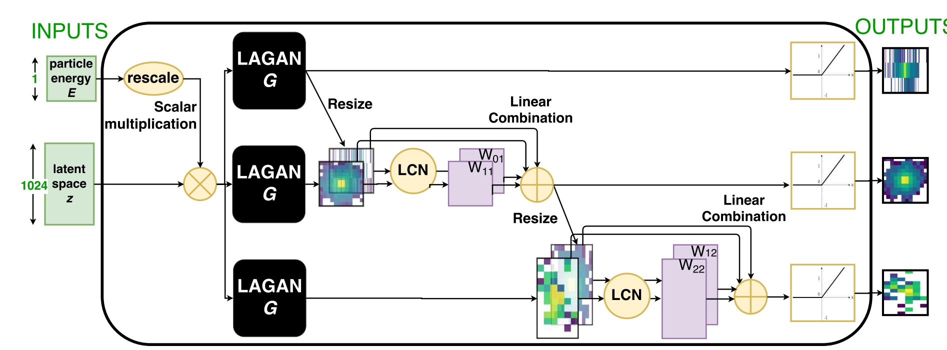
[Komiske, Metodiev, Nachman, Schwartz, 1707.08600]



CaloGAN

Fast Detector Simulation

Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321;
see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



JUNIPR

Probability Modeling

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