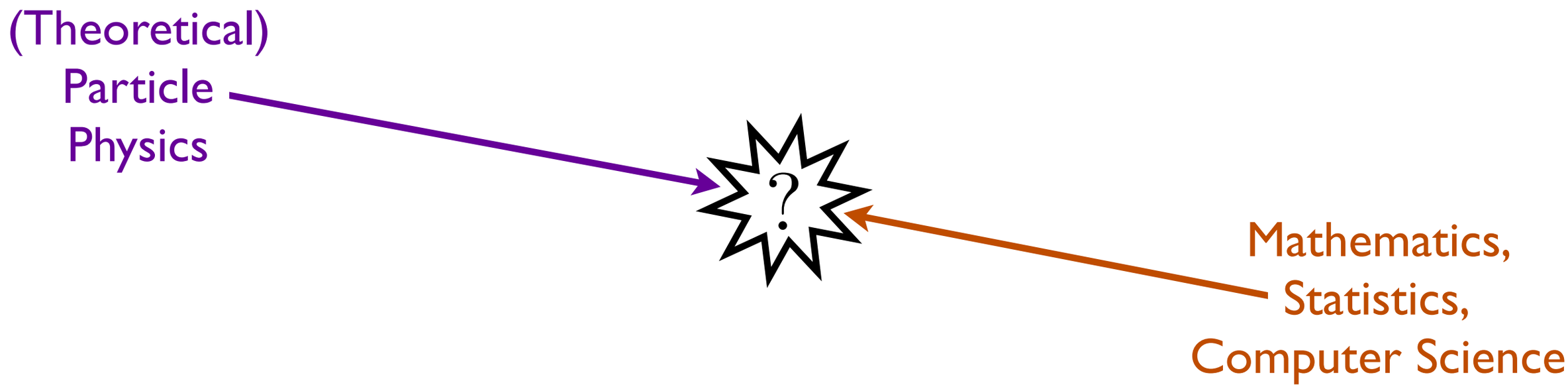
Collision Course

Particle Physics as a Machine-Learning Testbed

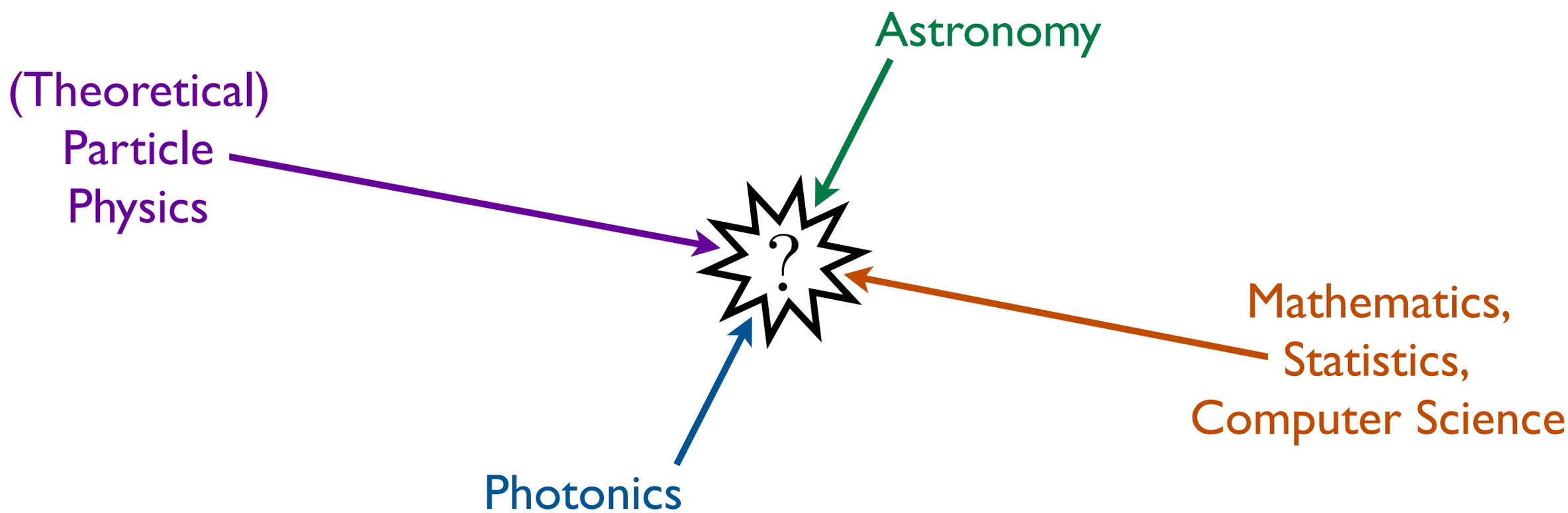
Jesse Thaler

Deep Learning in the Natural Sciences, University of Hamburg — February 28, 2019

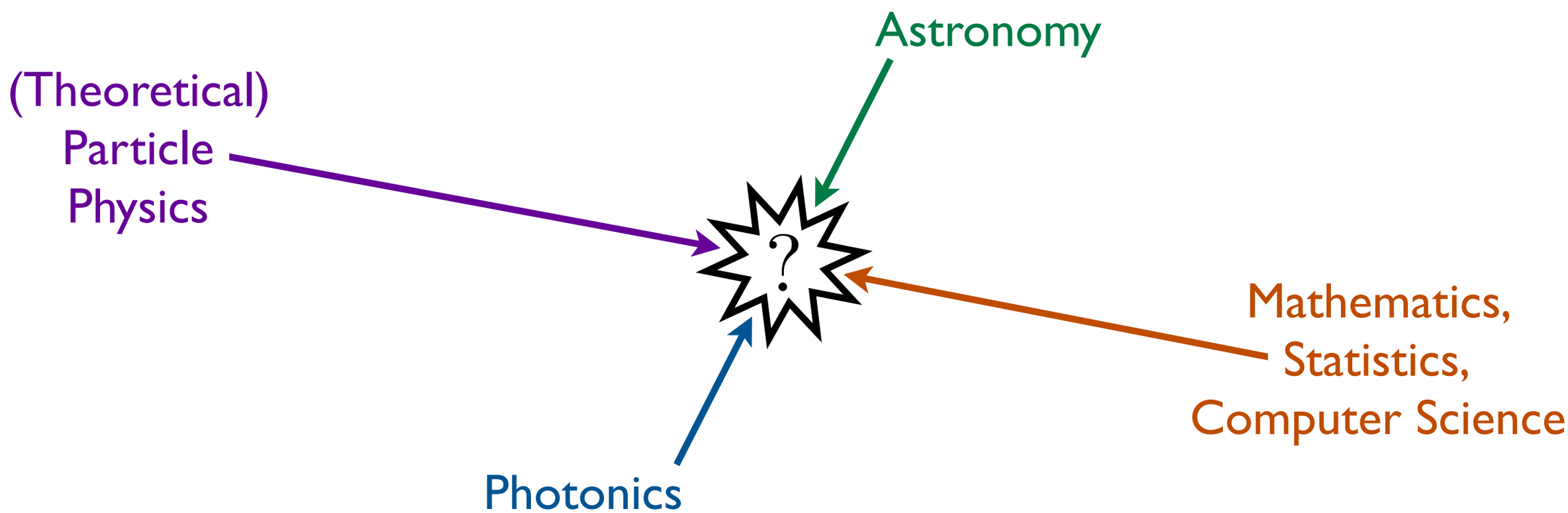
“Collision Course”

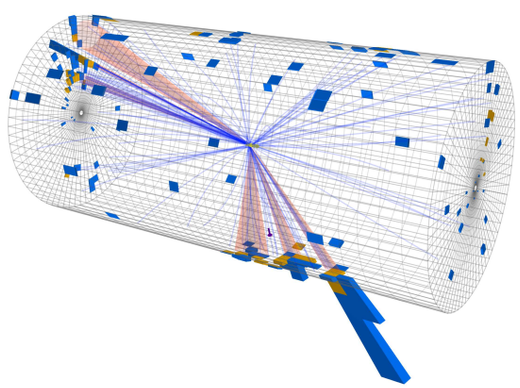
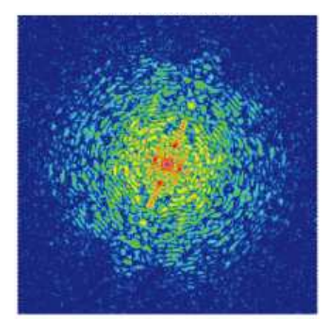


“Collision Course”

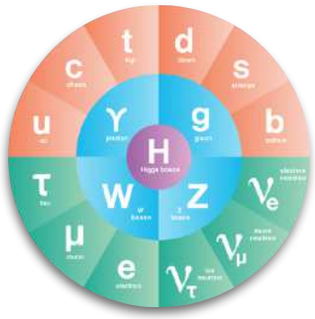


“Collision Course”

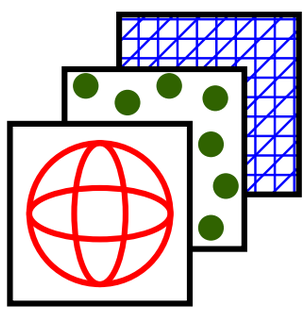


Connecting  ,  &  ?

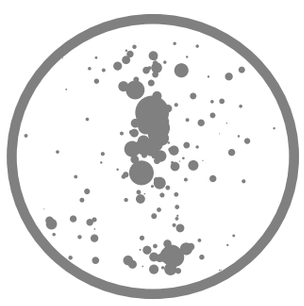
Outline



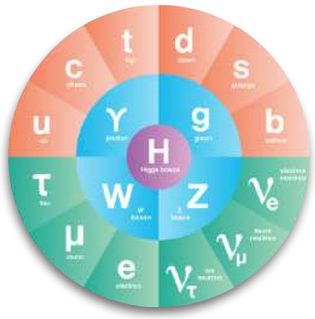
Particle Physics Primer



Point Clouds & Energy Flow Networks



Broader Lessons

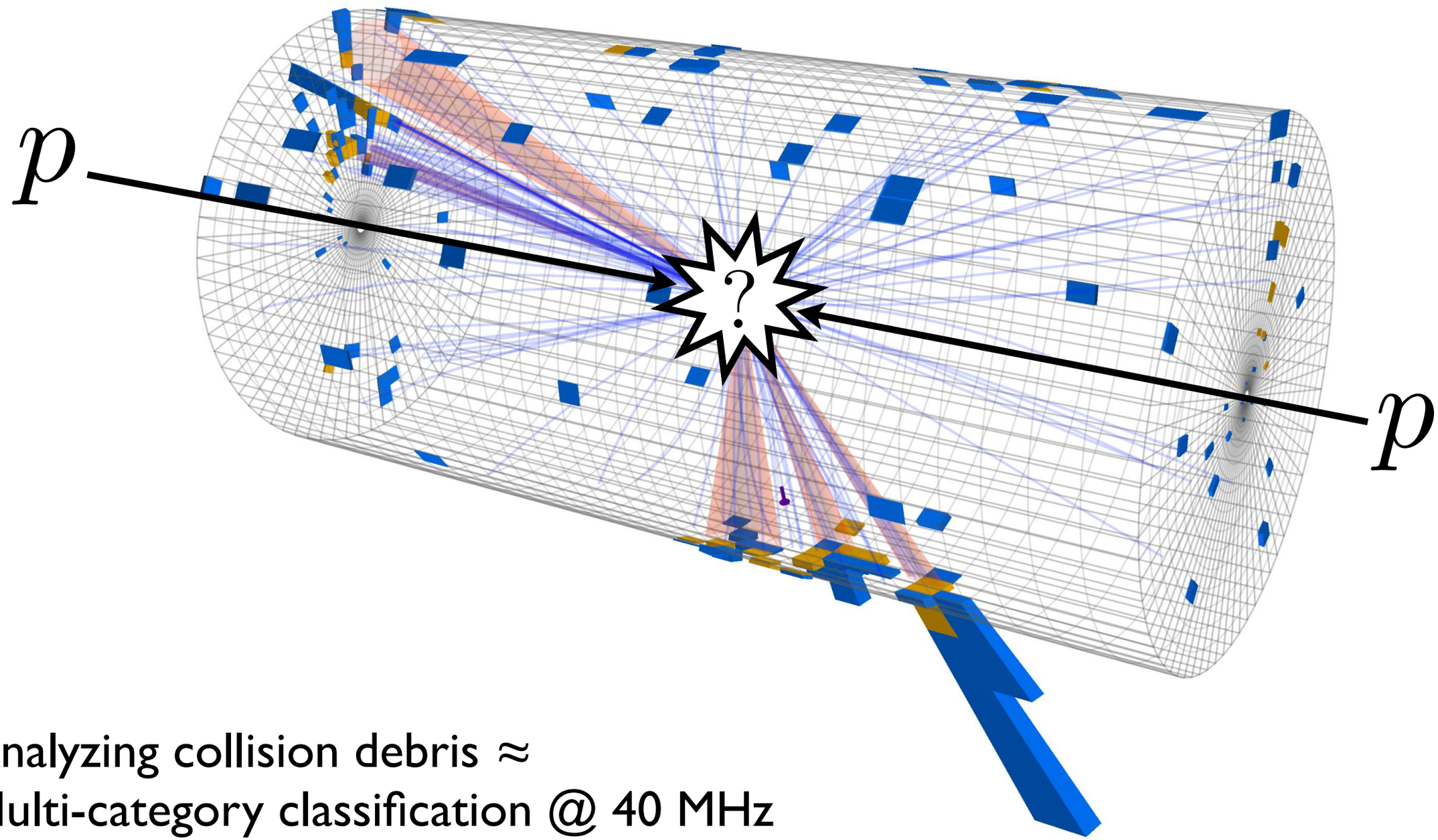


Particle Physics Primer

Point Clouds & Energy Flow Networks

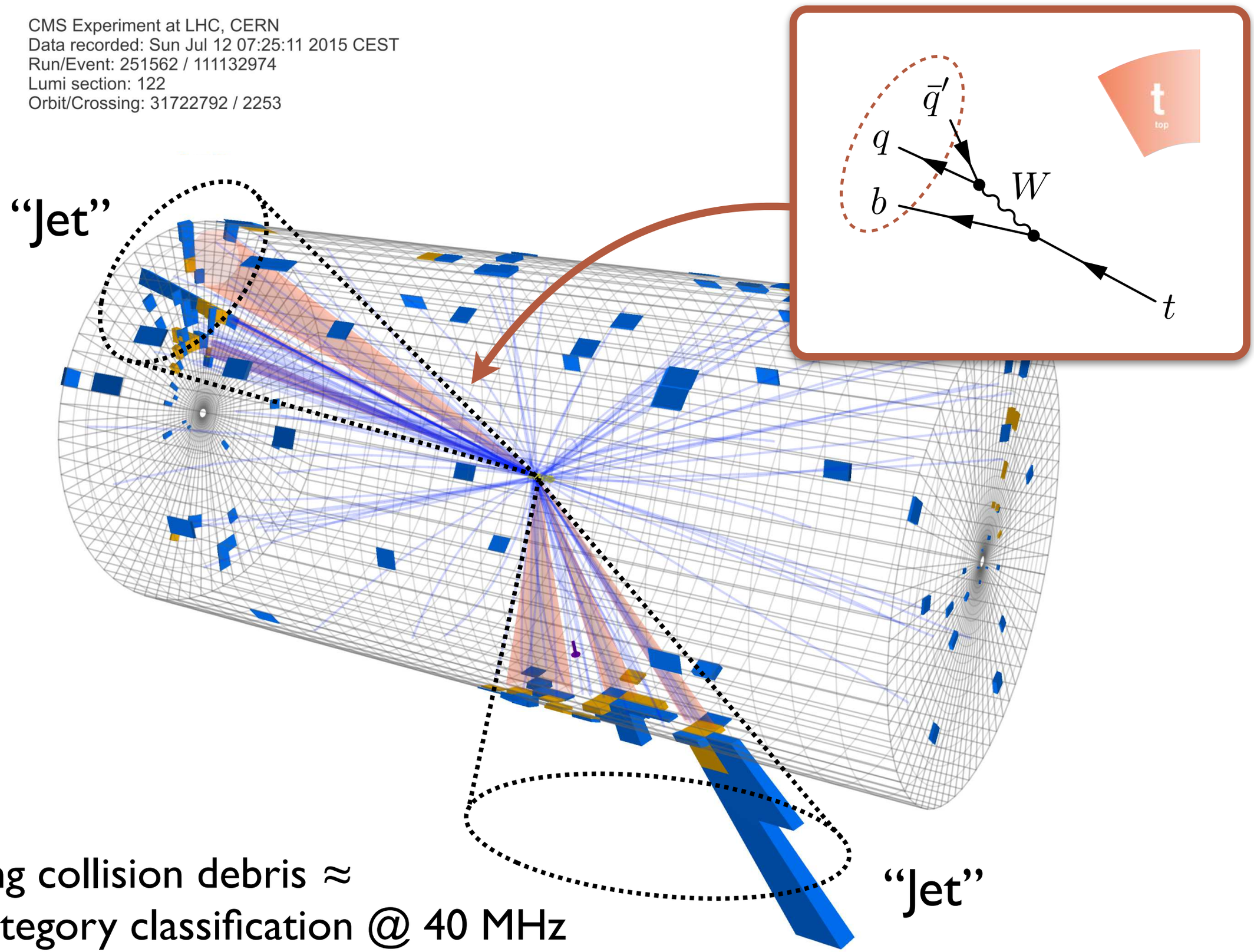
Broader Lessons

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

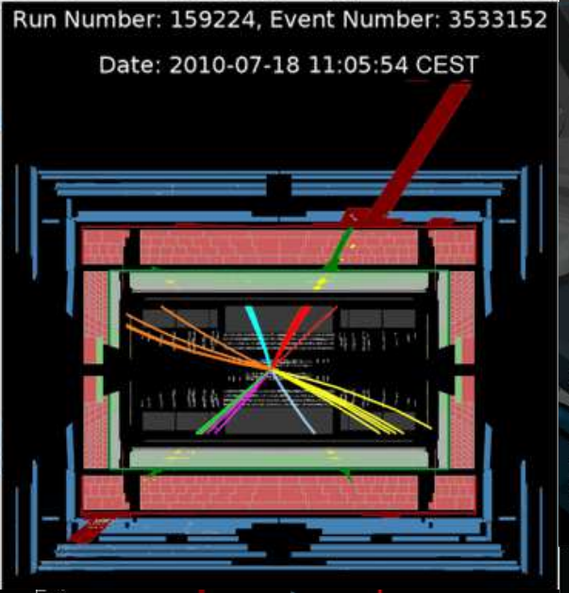


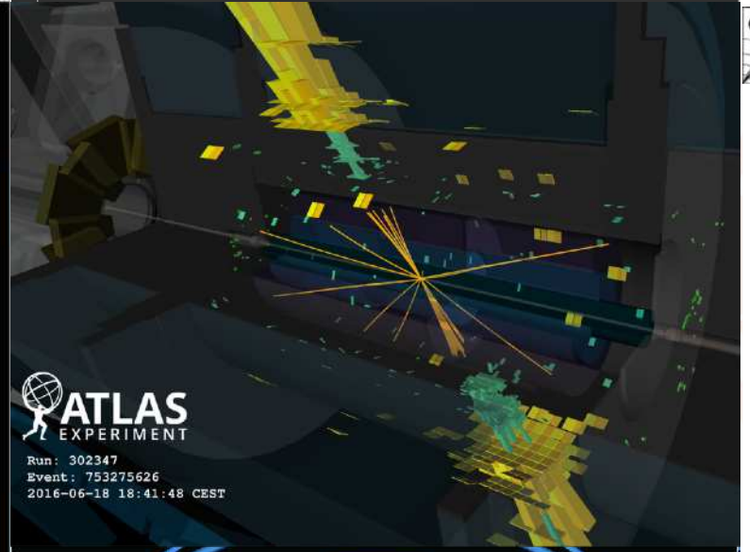
Analyzing collision debris ≈
Multi-category classification @ 40 MHz

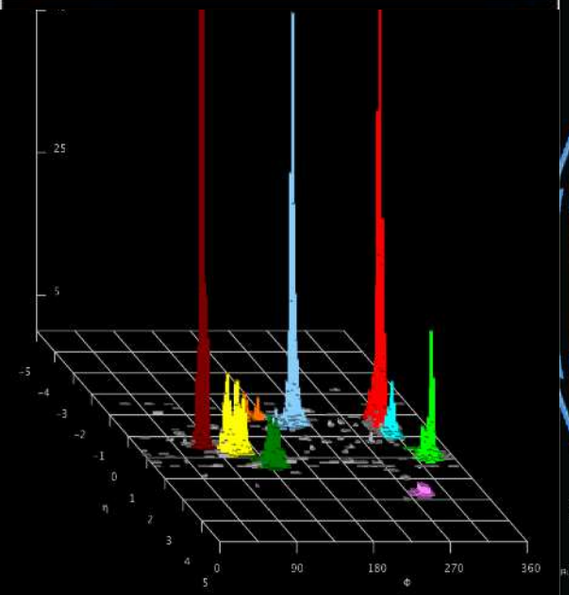
CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253

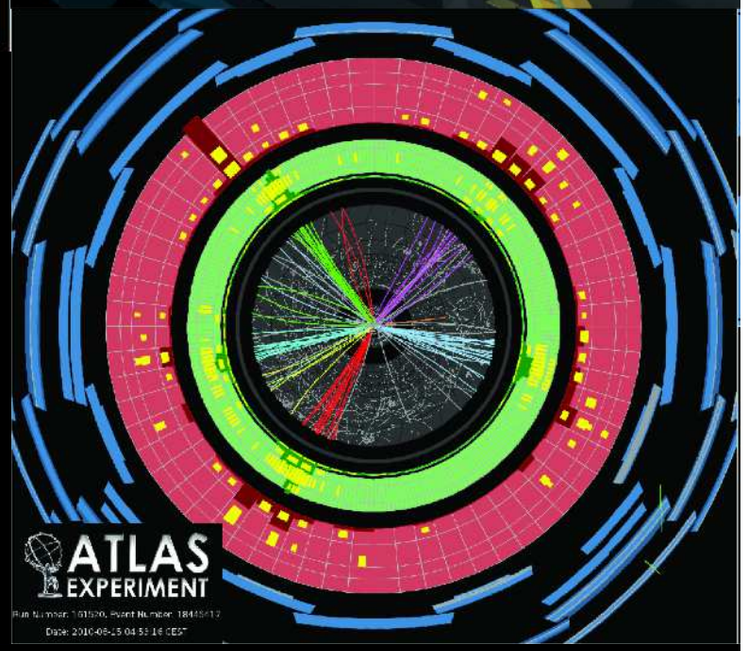


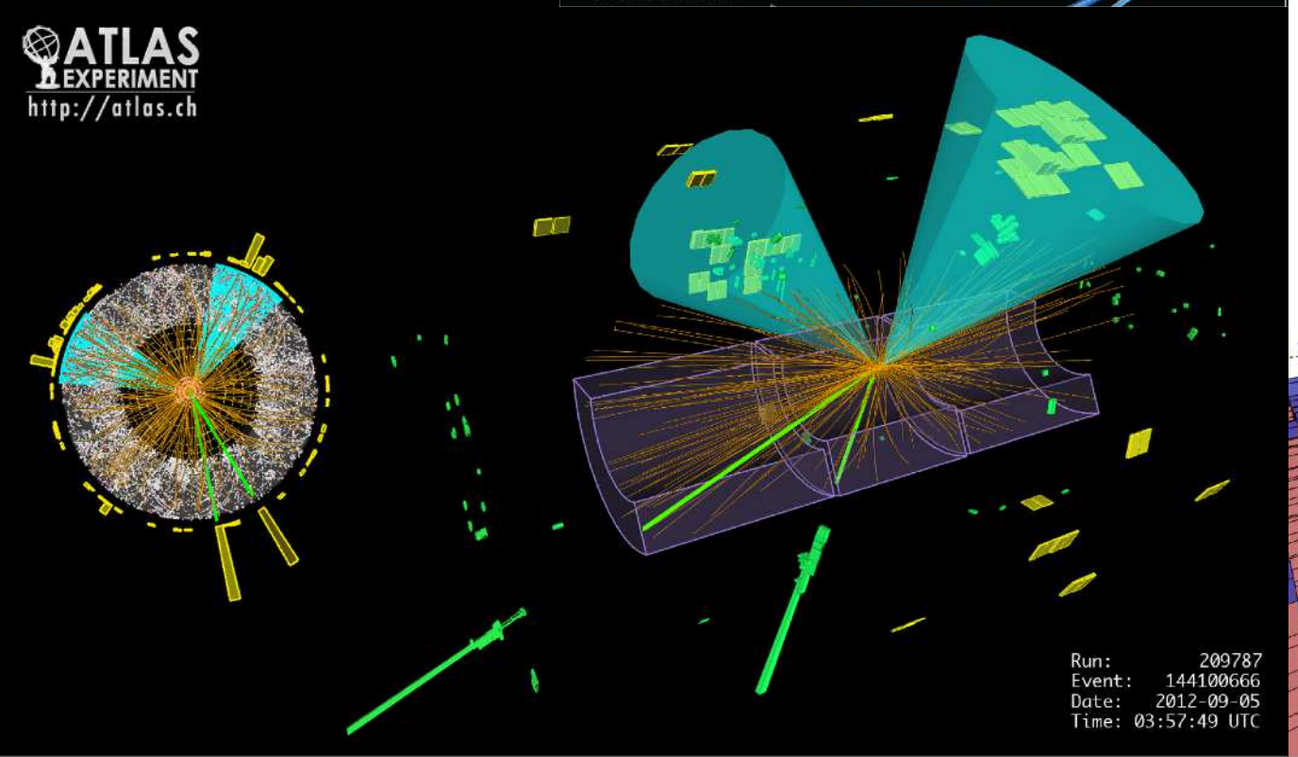
Analyzing collision debris ≈
Multi-category classification @ 40 MHz

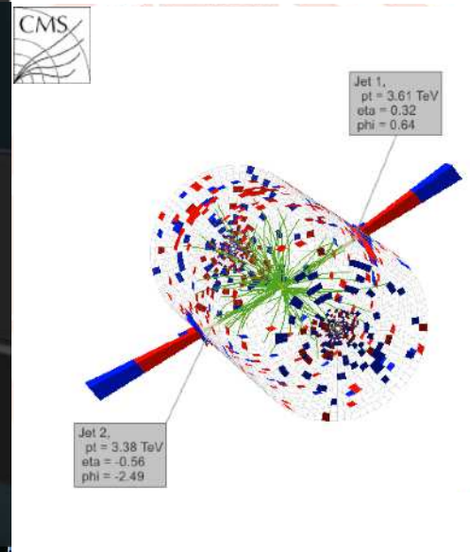

Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST

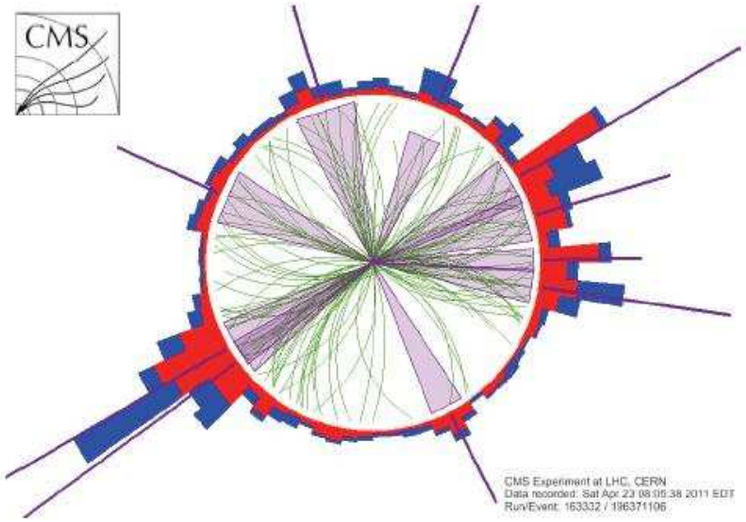

ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

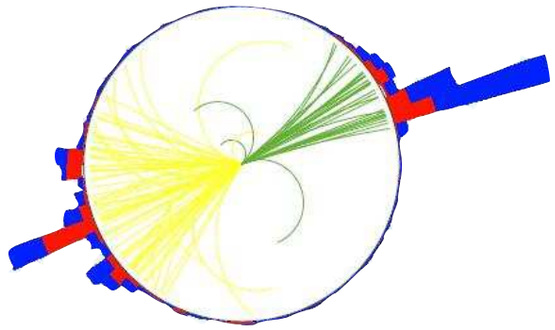


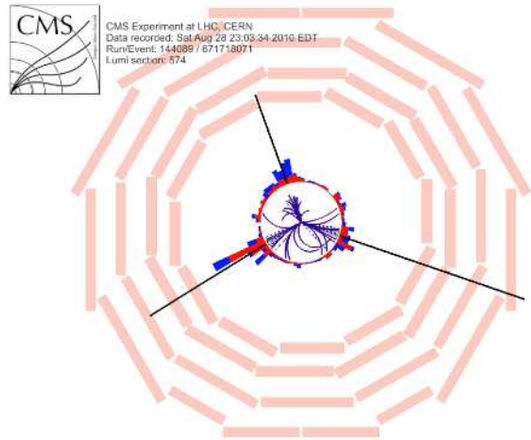

ATLAS
EXPERIMENT

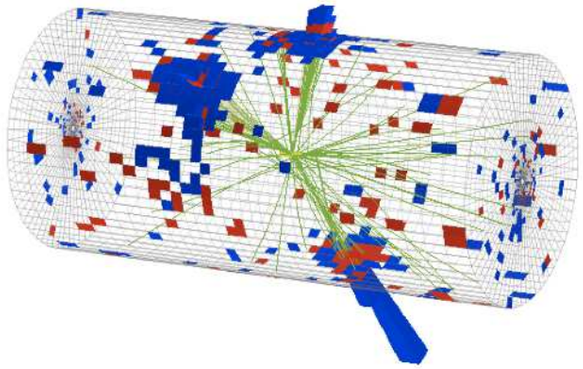

ATLAS
EXPERIMENT
http://atlas.ch
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC

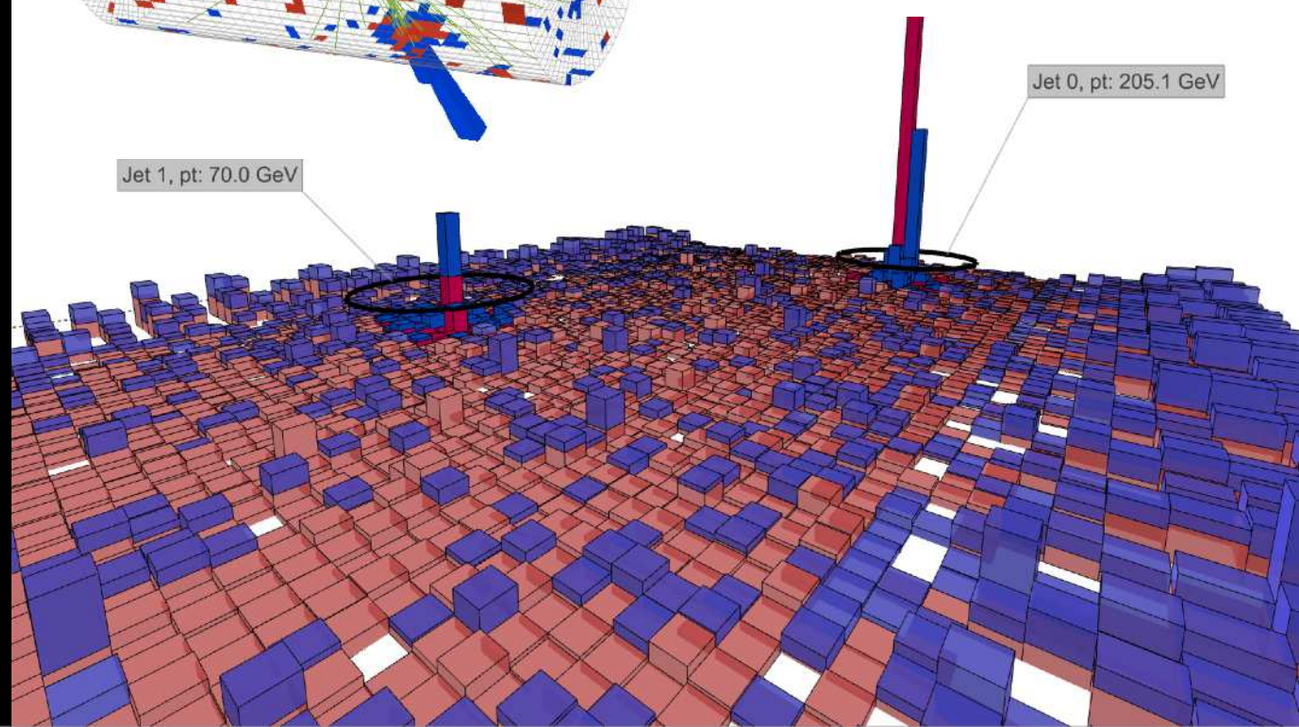

CMS
Jet 1,
pt = 3.61 TeV
eta = 0.32
phi = 0.64
Jet 2,
pt = 3.38 TeV
eta = -0.56
phi = -2.49


CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 186371106

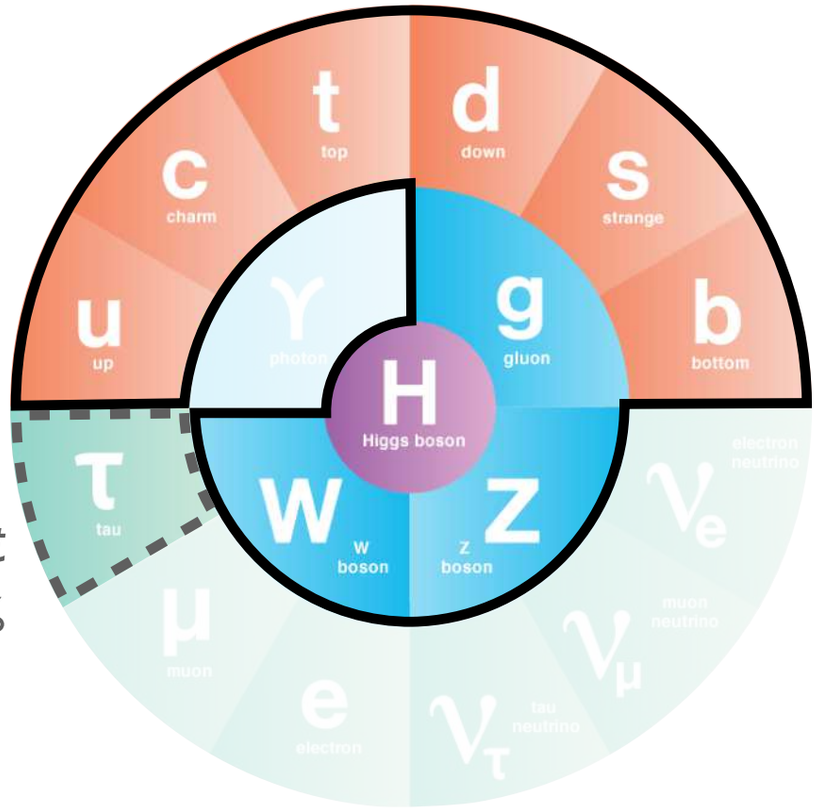



CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Aug 28 23:03:34 2010 EDT
Run/Event: 144089 / 671718071
Lumi section: 574




Jet 0, pt: 205.1 GeV
Jet 1, pt: 70.0 GeV

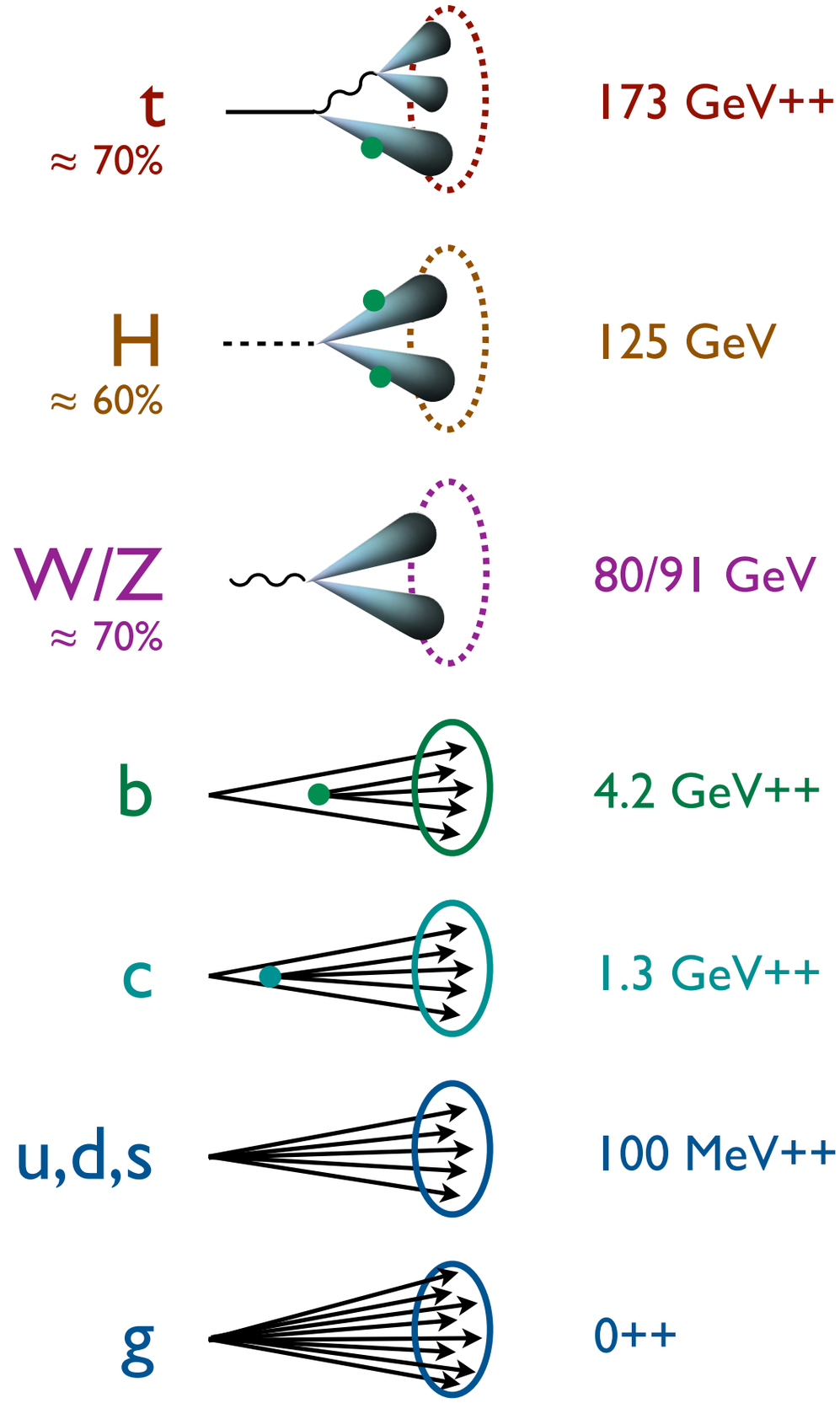
See talk by Tilman Plehn



micro-jet
≈ 65%

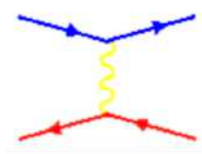
Jets from QCD and the Standard Model

++ = Mass from QCD Radiation



Particle Physics as ML Testbed

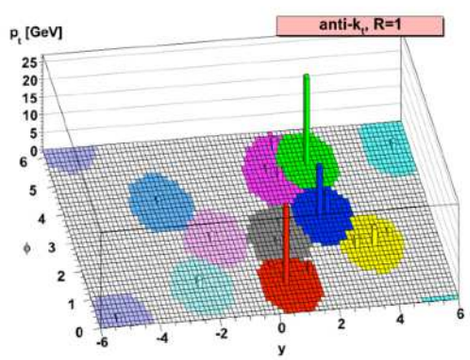
- *Huge datasets with reliable simulations*

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

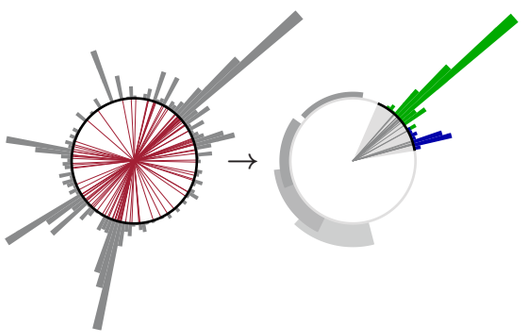
MadGraph Pythia

- *Broad use of (un)supervised algorithms*

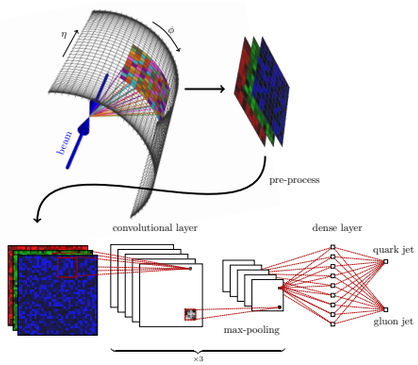
e.g.



Clustering



Noise Mitigation



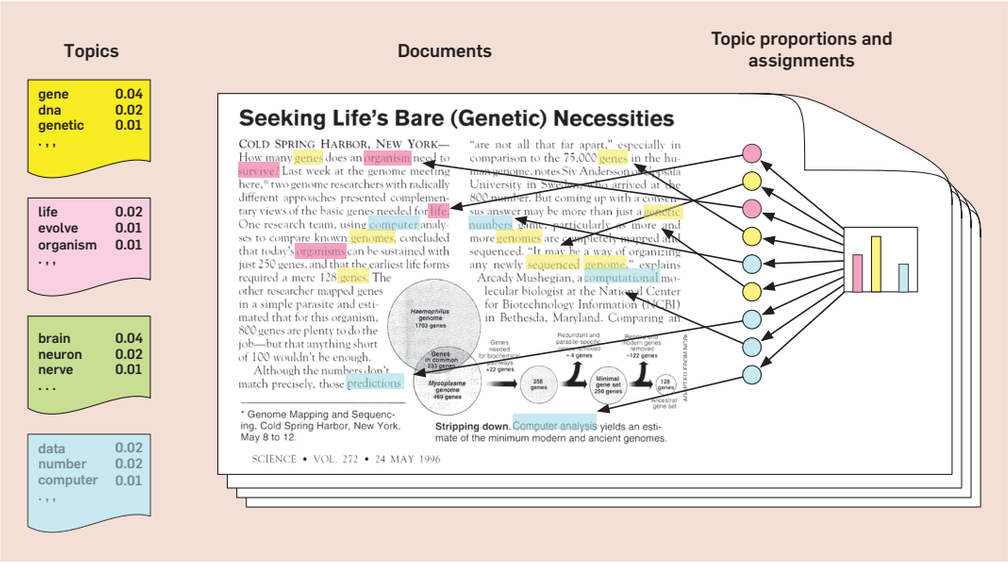
Machine Learning

- *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

Eric Metodiev
Friday, 12:10pm



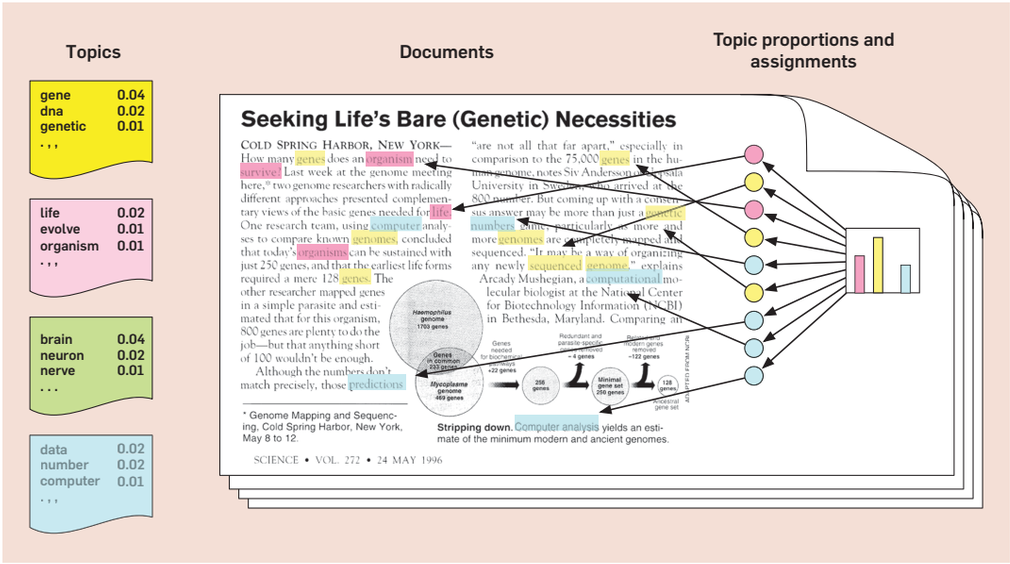
Blind Source Separation:

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

[Blei, CACM 2012; Katz-Samuels, Blanchard, Scott, 1710.01167;
Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140]

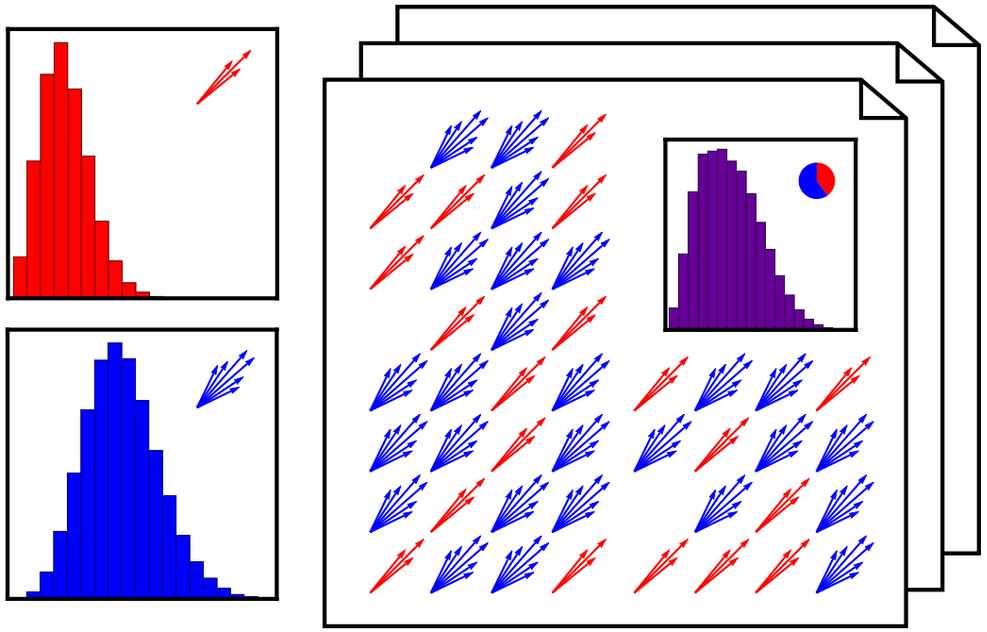
E.g. Topic Modeling for Jets

Eric Metodiev
Friday, 12:10pm



Blind Source Separation:

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



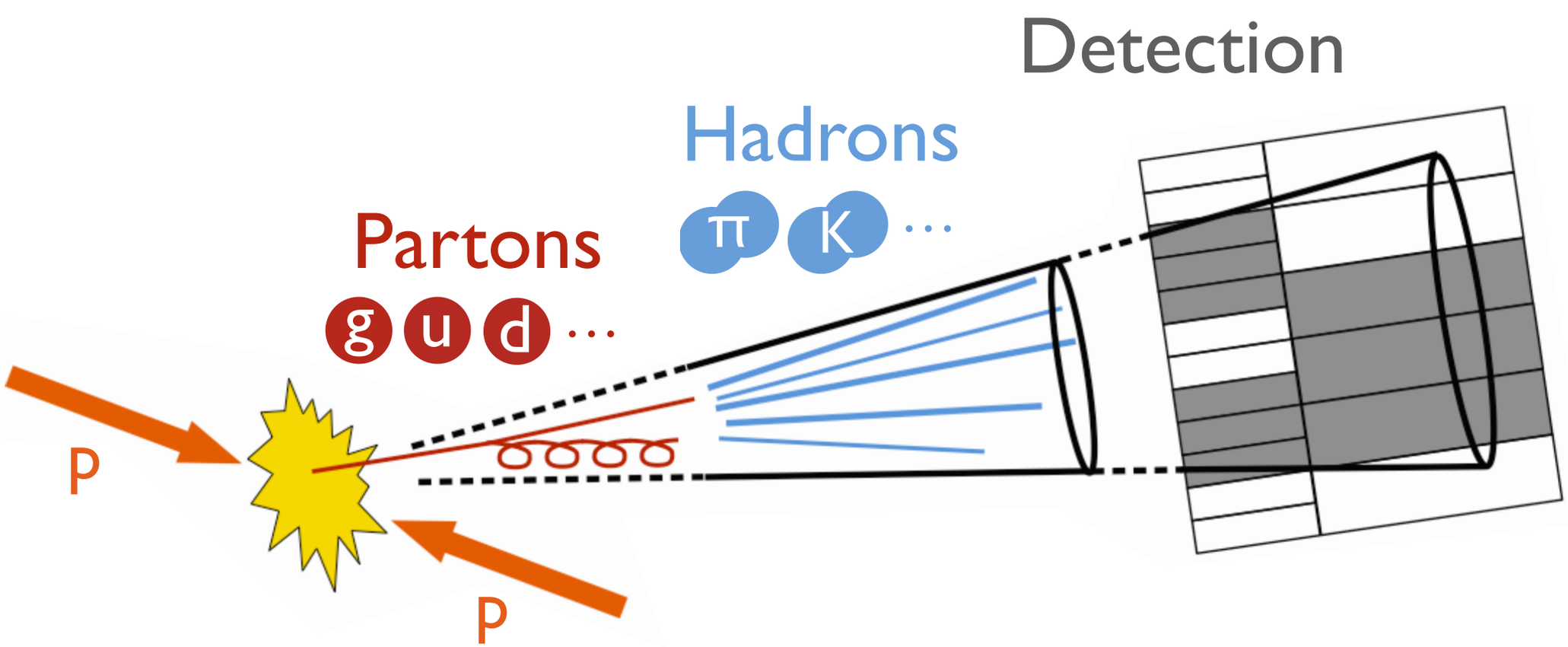
Quark/Gluon Separation:

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

[Blei, CACM 2012; Katz-Samuels, Blanchard, Scott, 1710.01167; Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140]

E.g. A Metric Distance for Jets

Patrick Komiske
Friday, 9:40am

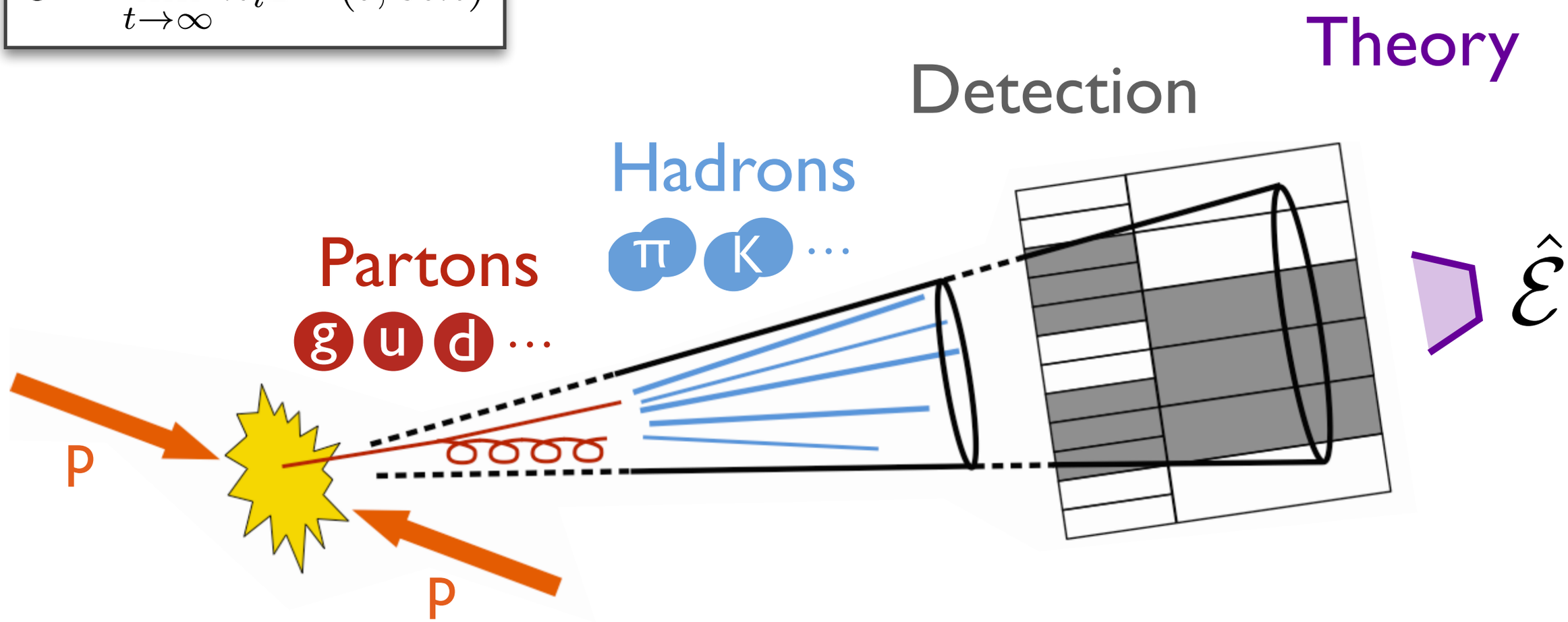


[Sveshnikov, Tkachov, hep-ph/9512370; Hofman, Maldacena, 0803.1467; Mateu, Stewart, JDT, 1209.3781; Komiske, Metodiev, JDT, 1712.07124, 1810.05165, 1902.02346; see also Rubner, Tomasi, Guibas, ICJV 2000; Pele, Werman, ECCV 2008]

E.g. A Metric Distance for Jets

Patrick Komiske
Friday, 9:40am

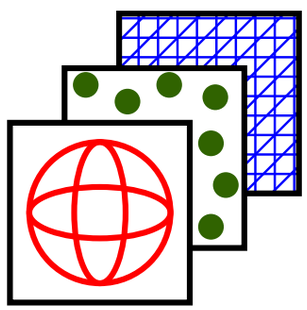
$$\hat{\mathcal{E}} \simeq \lim_{t\to\infty} \hat{n}_i T^{0i}(t, vt\hat{n})$$



Robust measure of jet structure captured by stress-energy flow

[Sveshnikov, Tkachov, hep-ph/9512370; Hofman, Maldacena, 0803.1467; Mateu, Stewart, JDT, 1209.3781; Komiske, Metodiev, JDT, 1712.07124, 1810.05165, 1902.02346; see also Rubner, Tomasi, Guibas, ICJV 2000; Pele, Werman, ECCV 2008]

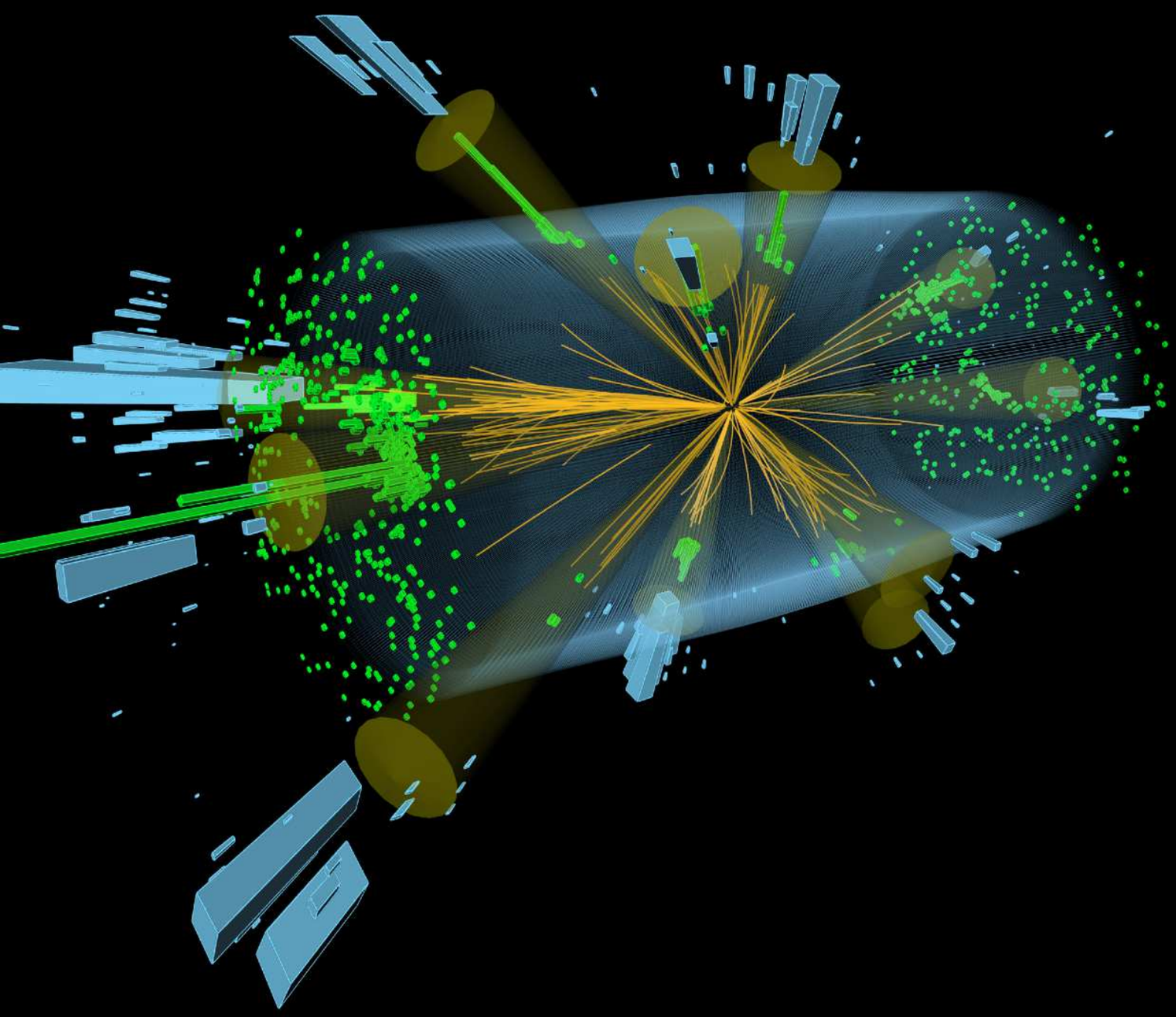
Particle Physics Primer

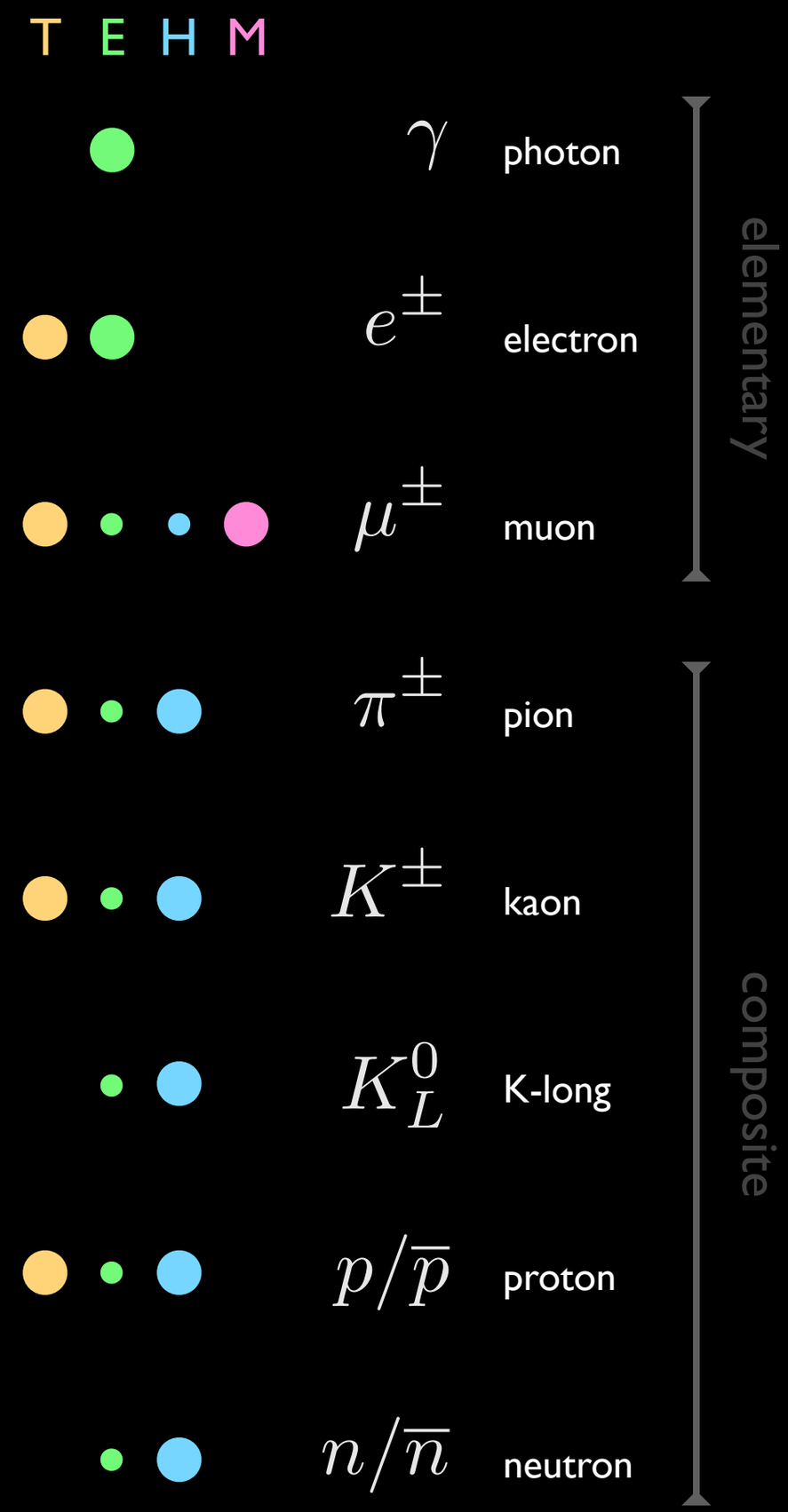


Point Clouds & Energy Flow Networks

Broader Lessons

What is a Collision Event?



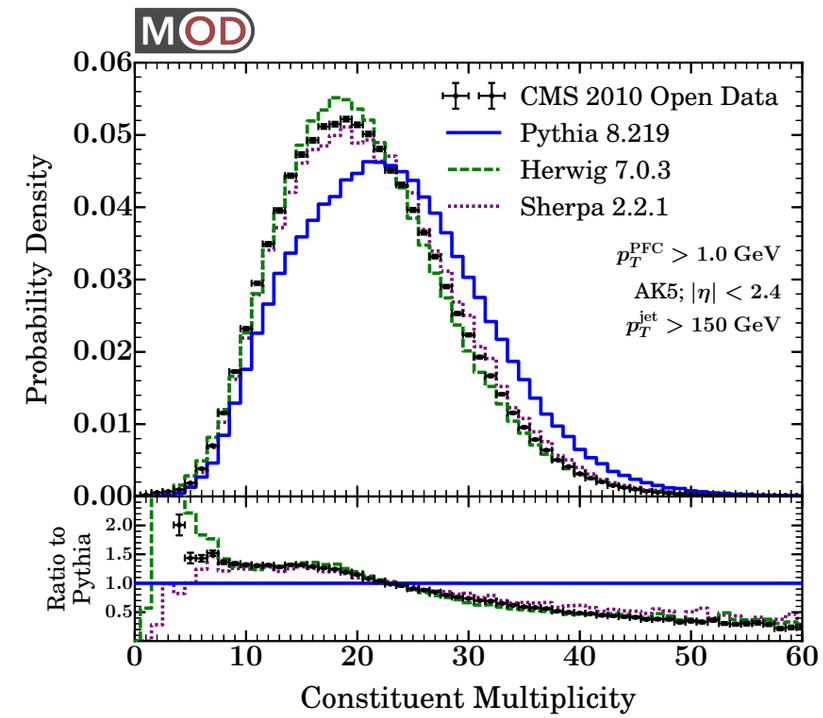


[Popular Science, 2013]

Key Fact #1: Events/Jets are Point Clouds

Jet constituents:

Particle-like objects

Variable-length → 

Unordered set → QM

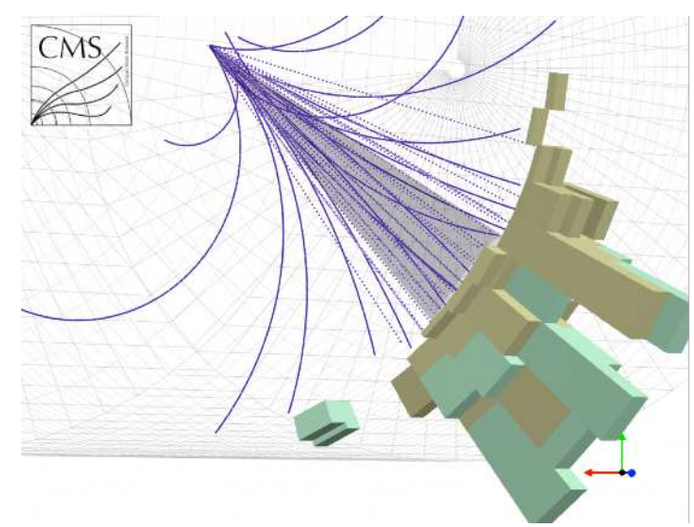
Per particle:

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

Flavor/charge labels

Vertex information

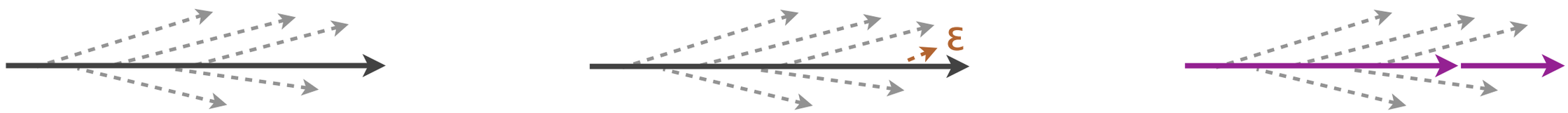
Quality criteria, etc.



[plot from Tripathee, Xue, Larkoski, Marzani, JDT, 1704.05842; see also Qu, Gouskos, 1902.08570]

Key Fact #2: IRC Safety Used for Robustness

Original Jet | Infrared | Collinear

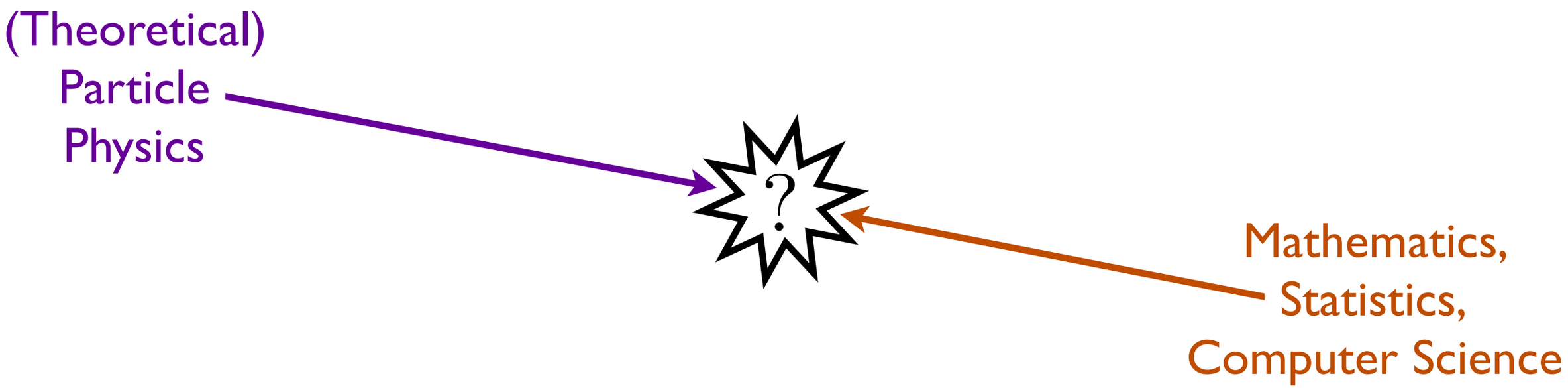


IRC Safe Observable: Insensitive to IR or C emissions

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

Enables calculations in fixed-order perturbation theory

Events/jets are *(optionally)* energy-weighted point clouds

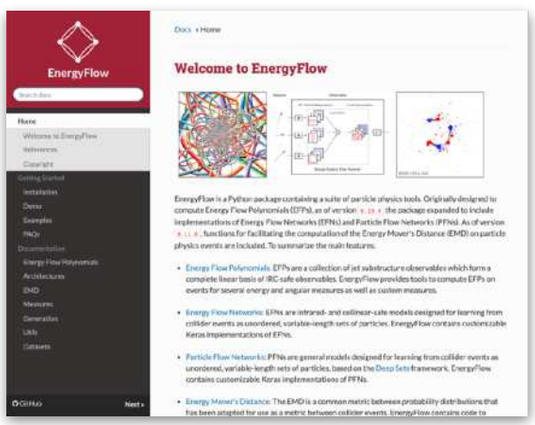

(Theoretical)
Particle
Physics
?
Mathematics,
Statistics,
Computer Science

Patrick Komiske

Eric Metodiev

(Theoretical) Particle 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 effective field theories)

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

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

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

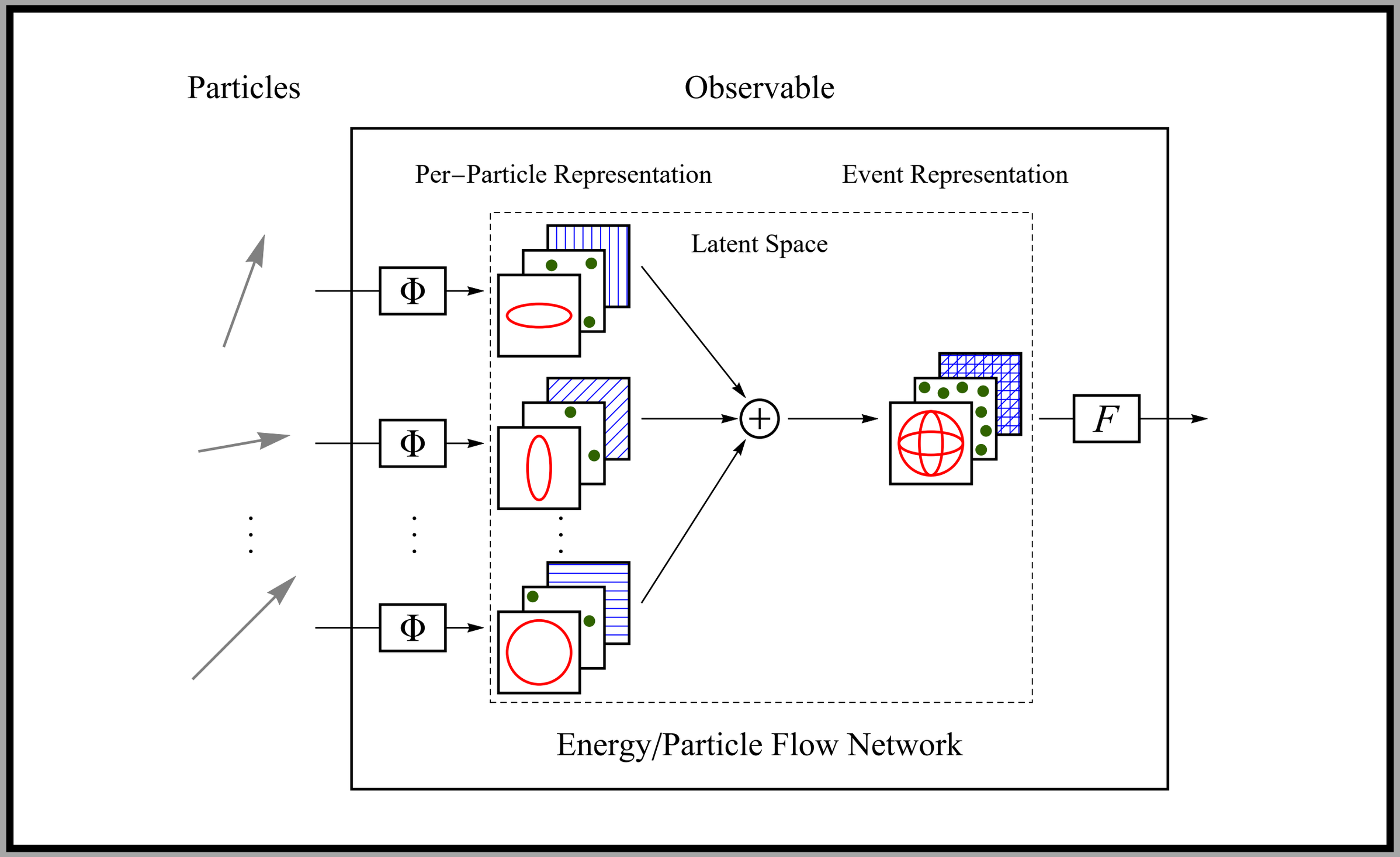
Energy / Particle Flow Networks

IRC-safe Φ

General Φ

[Komiske, Metodiev, JDT, 1810.05165]

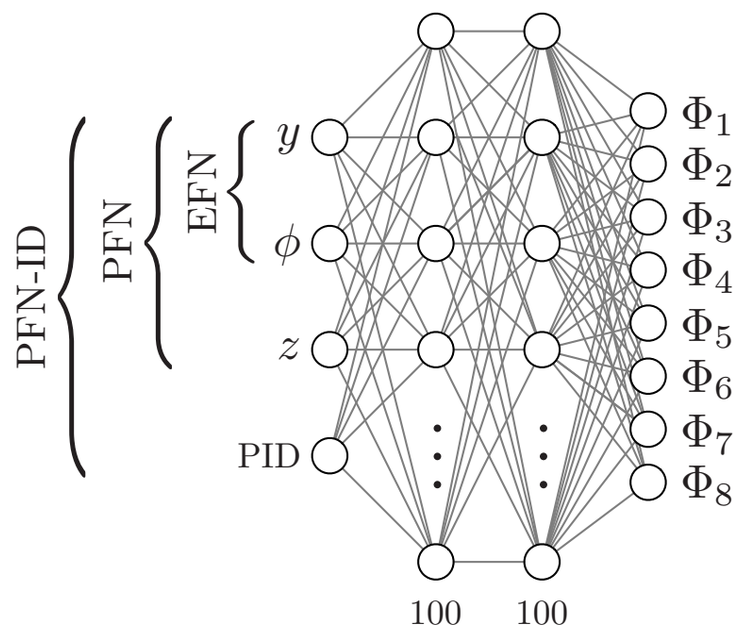
Conjectured Generalization



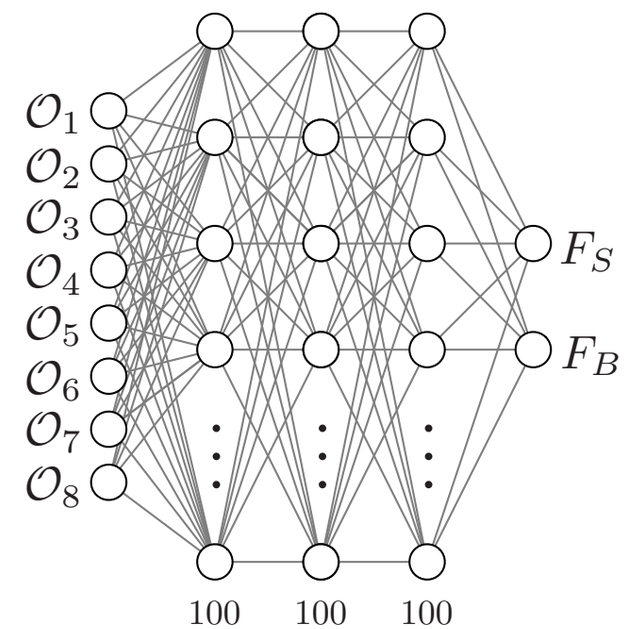
[Komiske, Metodiev, JDT, 1810.05165]

Conjectured Generalization

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, 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.*

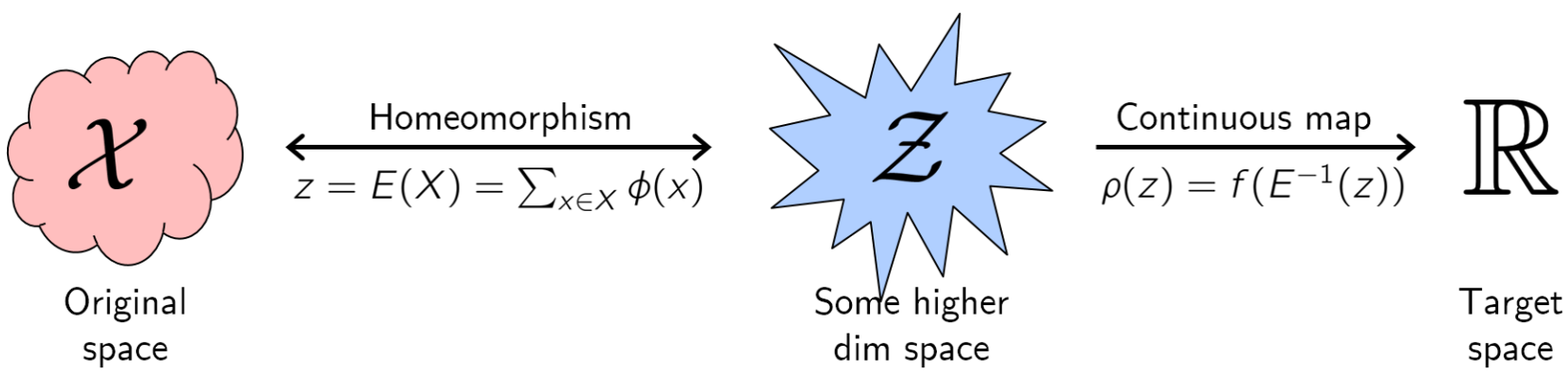
(!)

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

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

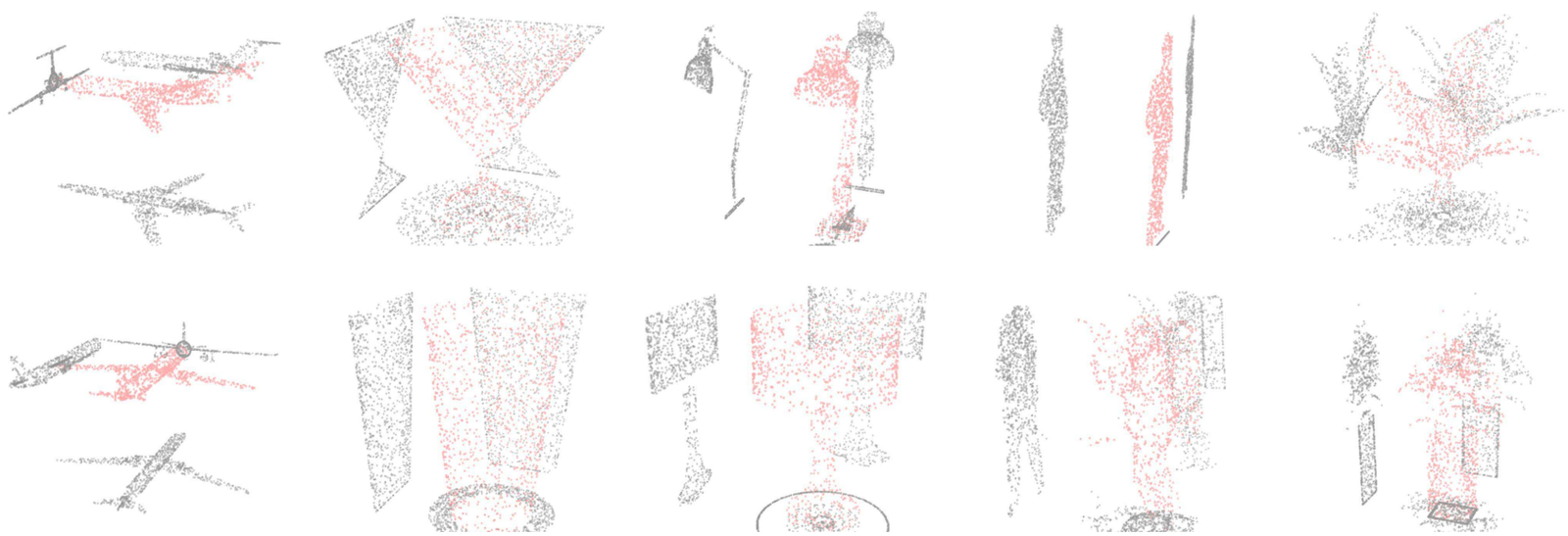
Generic Observable

[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

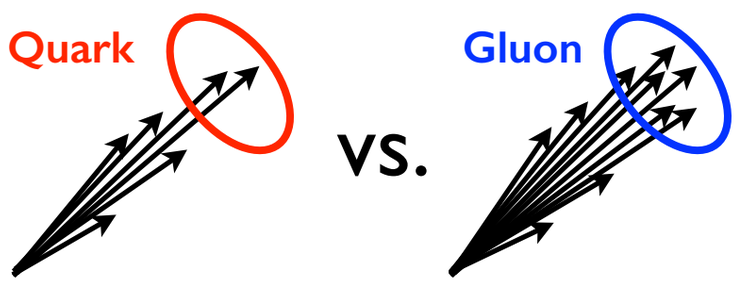
Point Cloud Classification



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

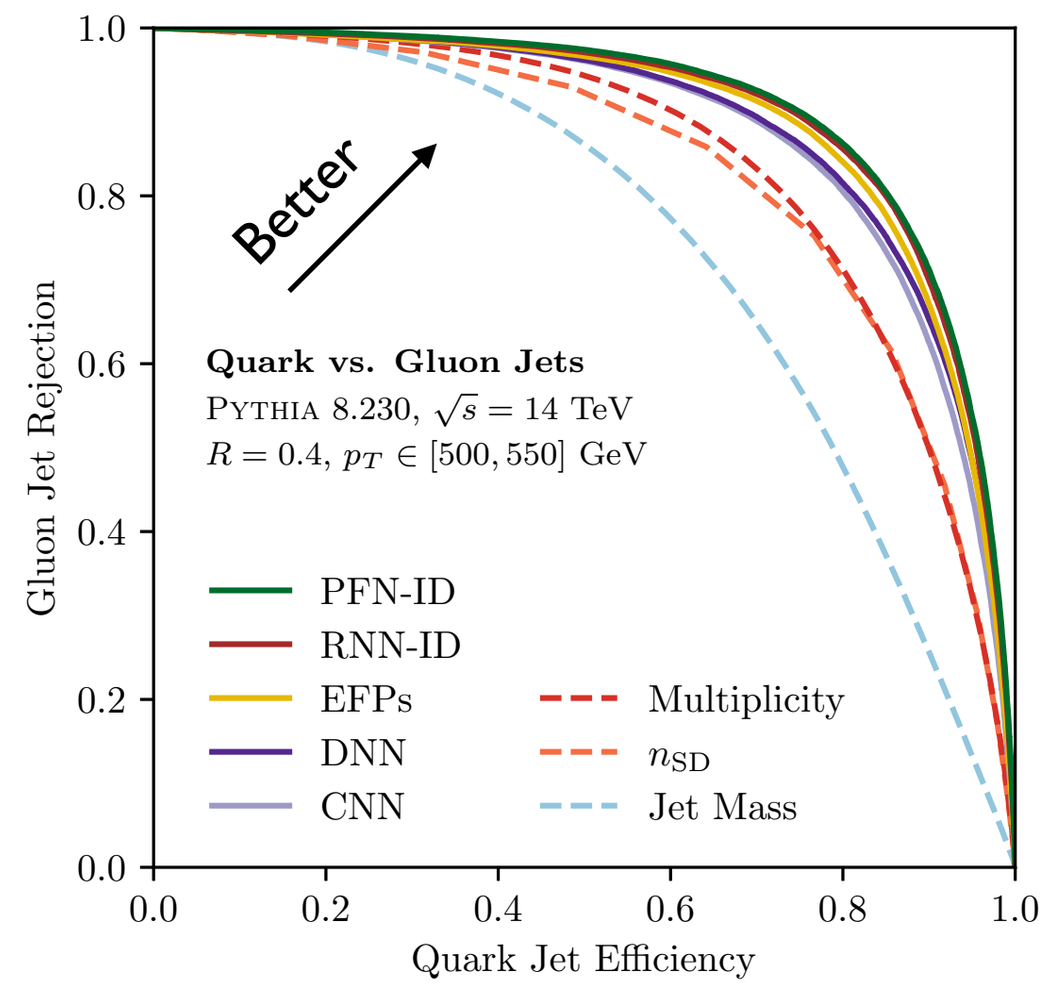
Deep Sets for Particle Jets

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



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

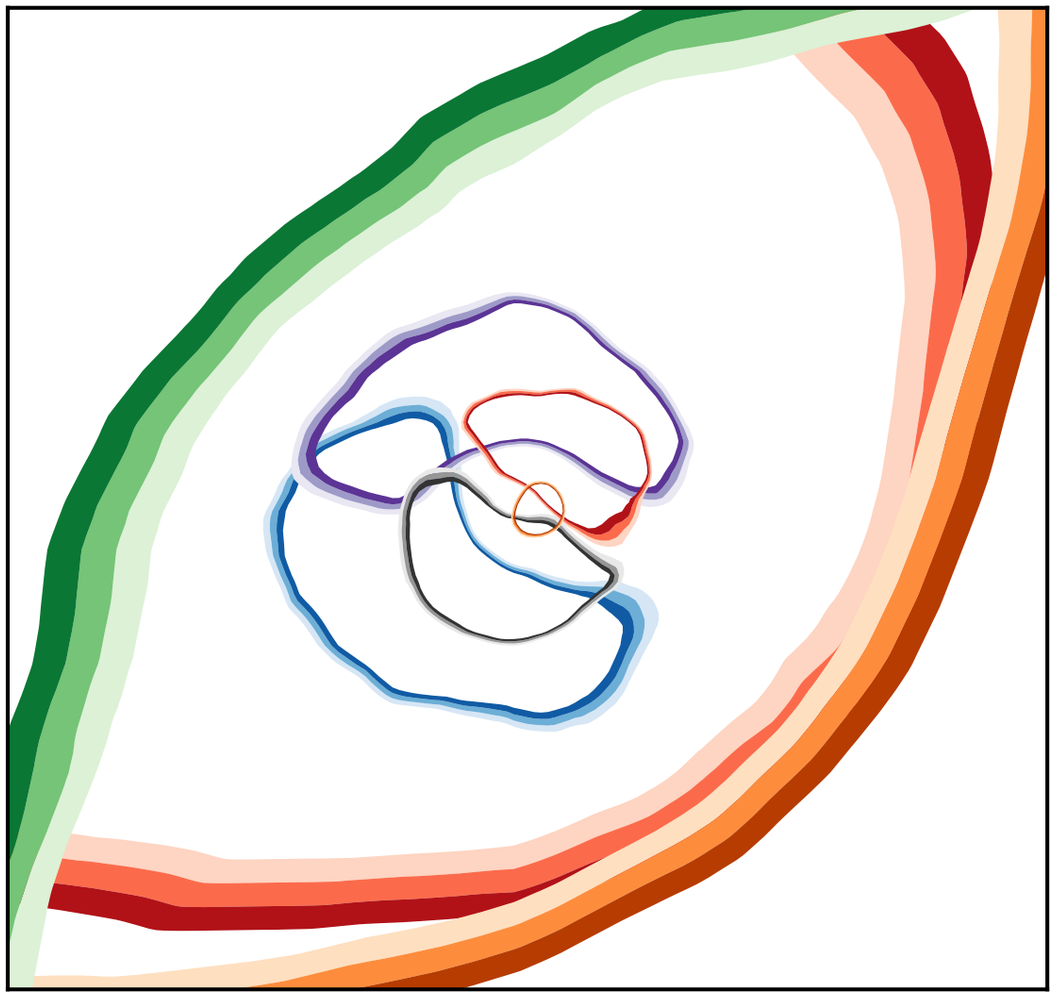
Competitive with previous methods



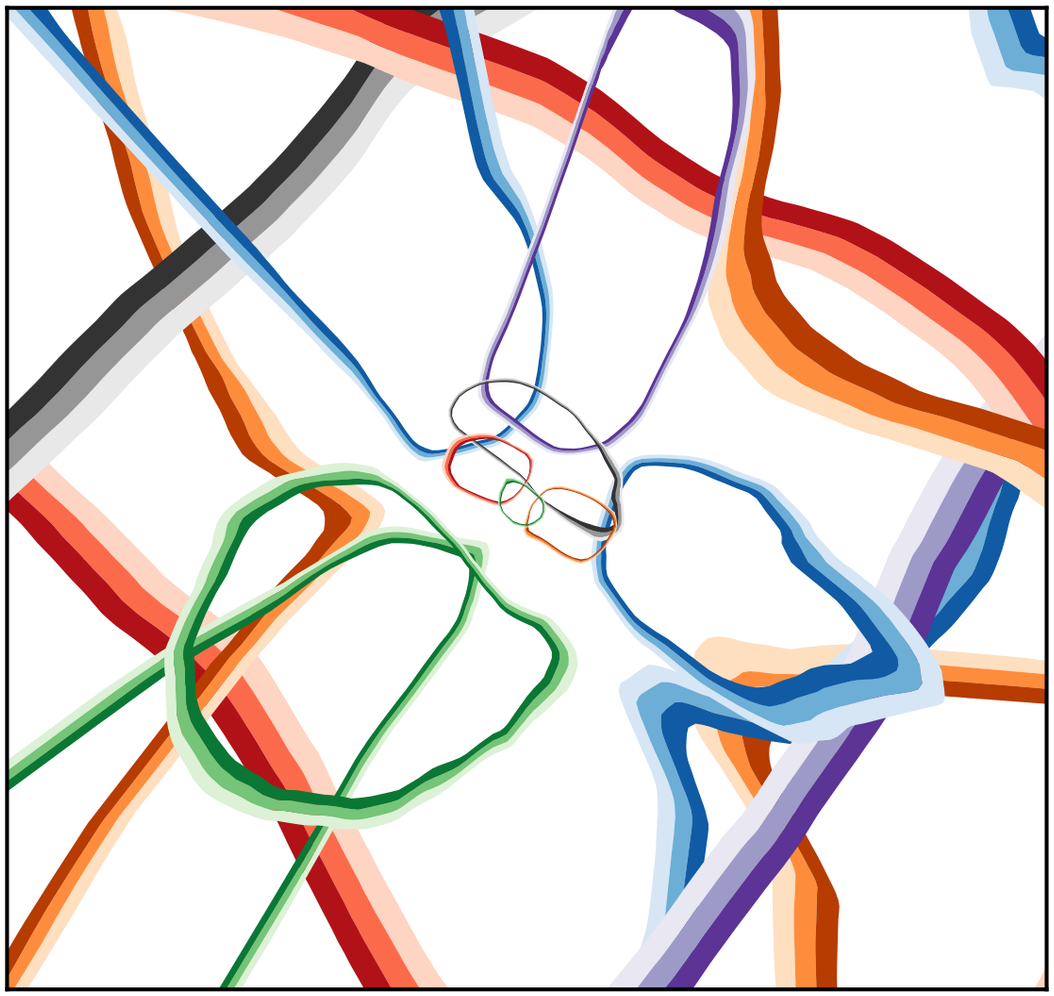
[Komiske, Metodiev, JDT, 1810.05165]

Psychedelic Network Visualization

Latent Dimension 8



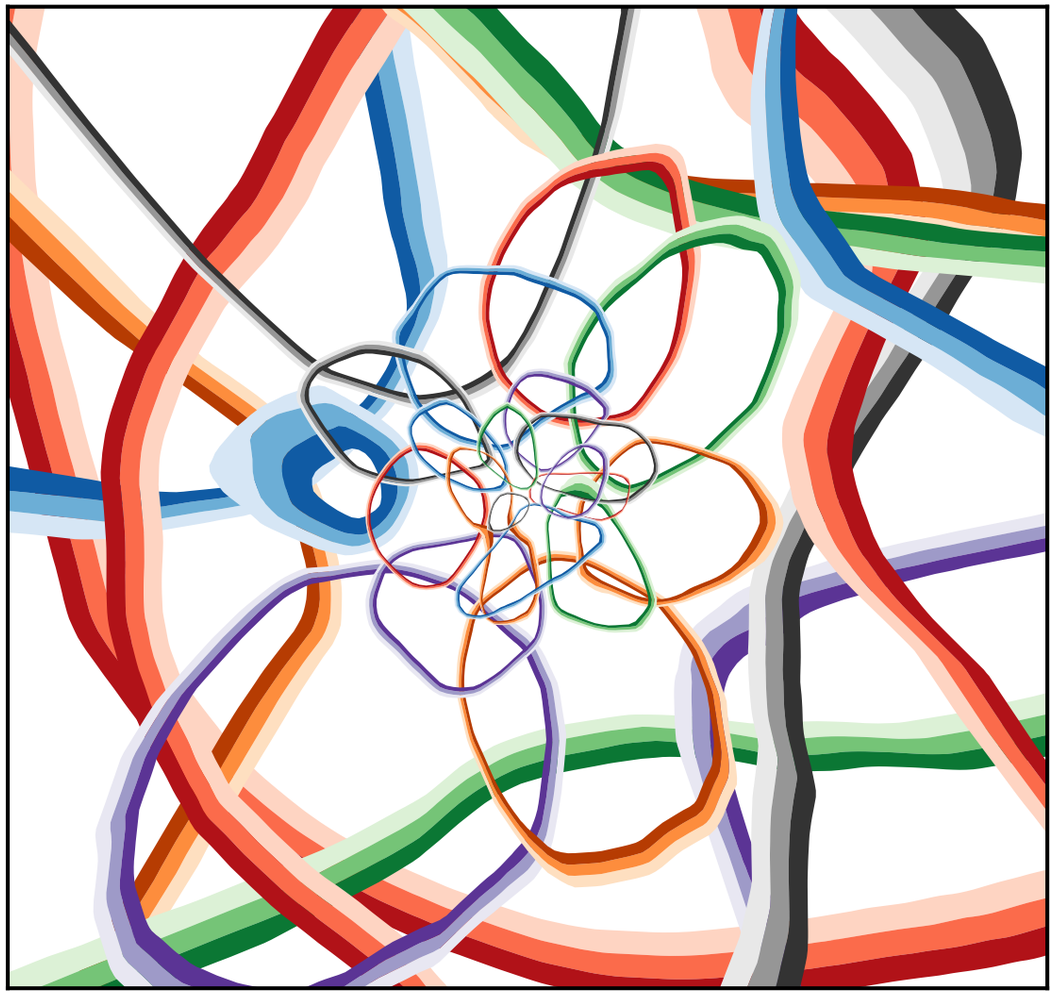
Latent Dimension 16



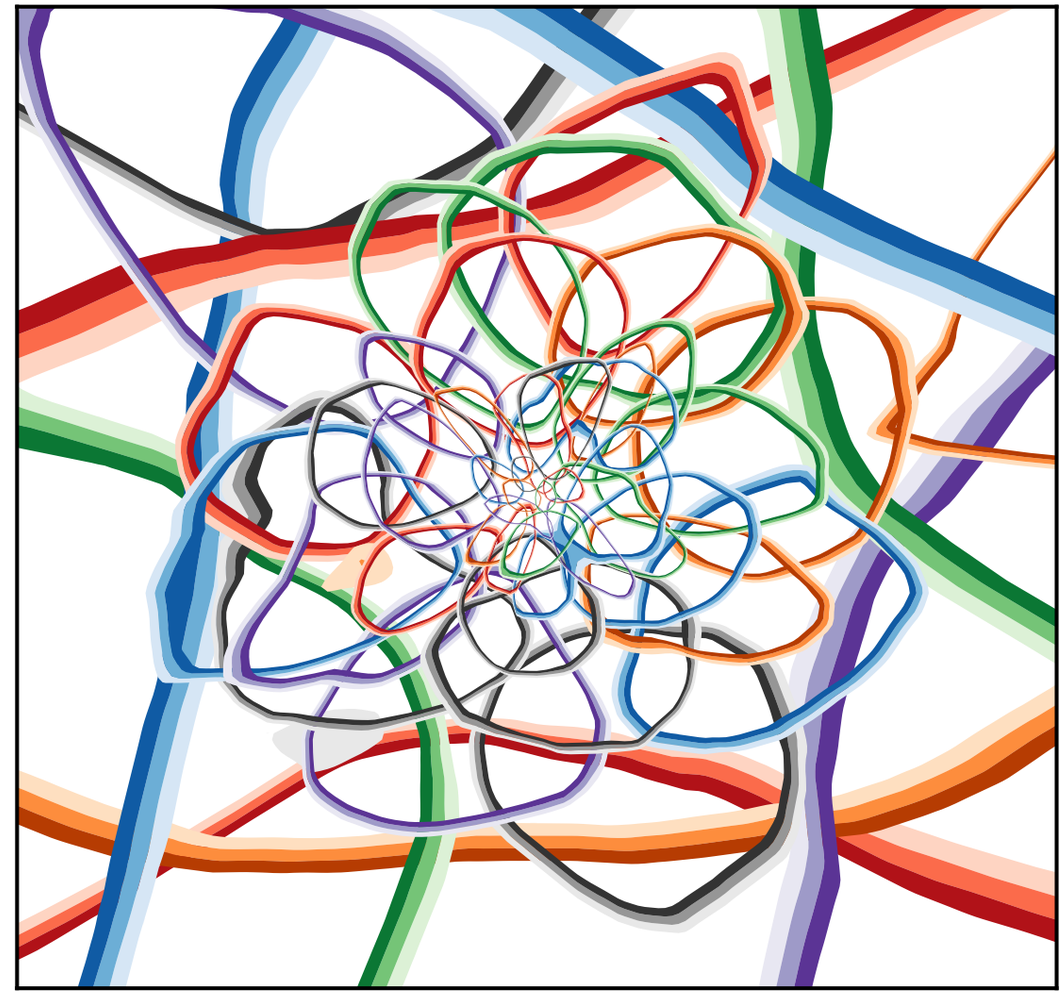
(see backup for how these are made)

Psychedelic Network Visualization

Latent Dimension 32

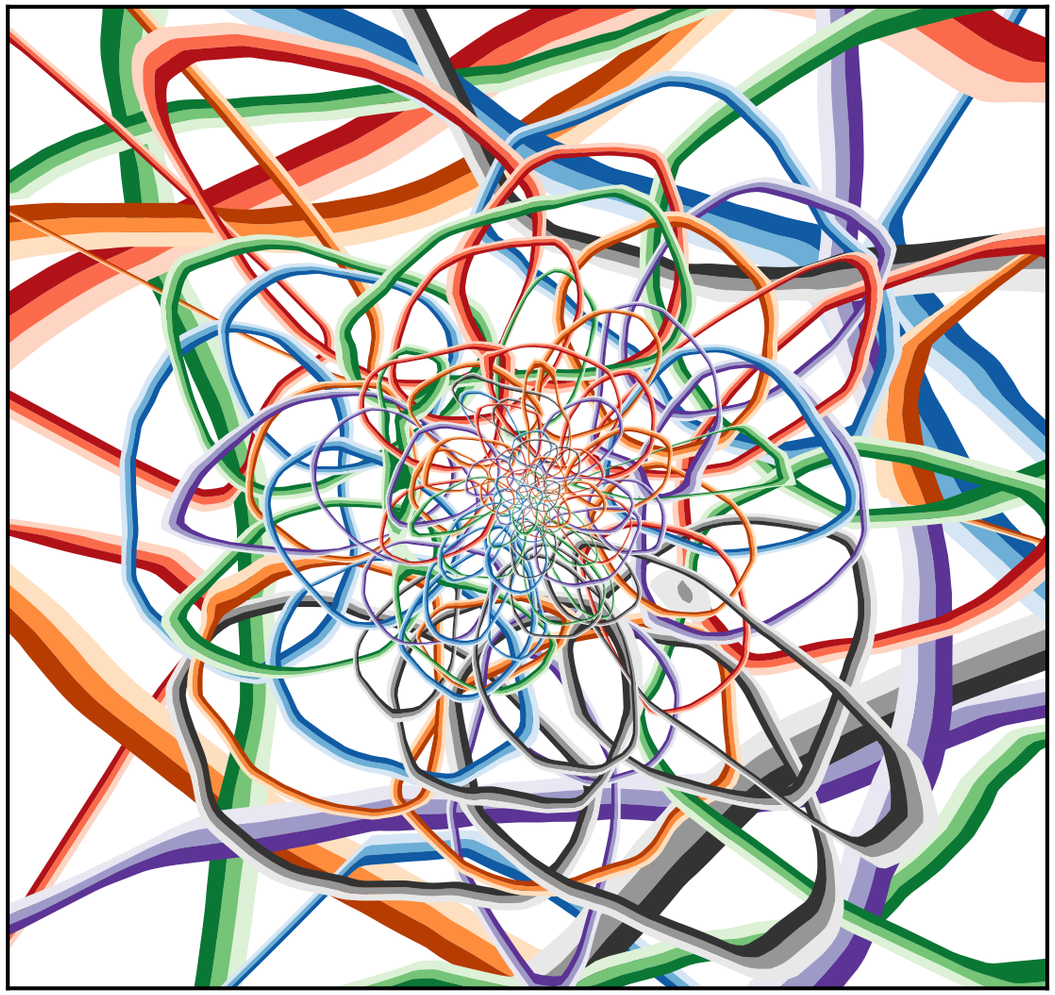


Latent Dimension 64

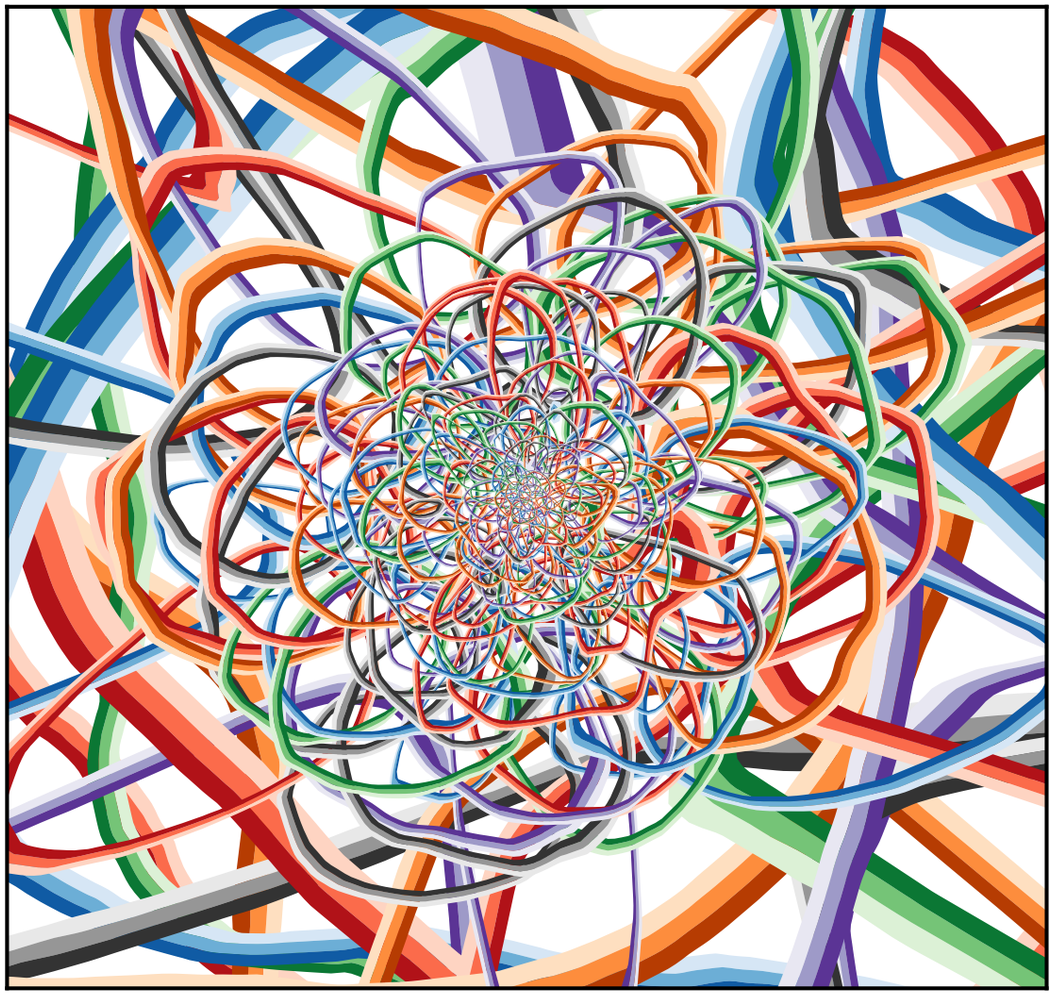


Psychedelic Network Visualization

Latent Dimension 128

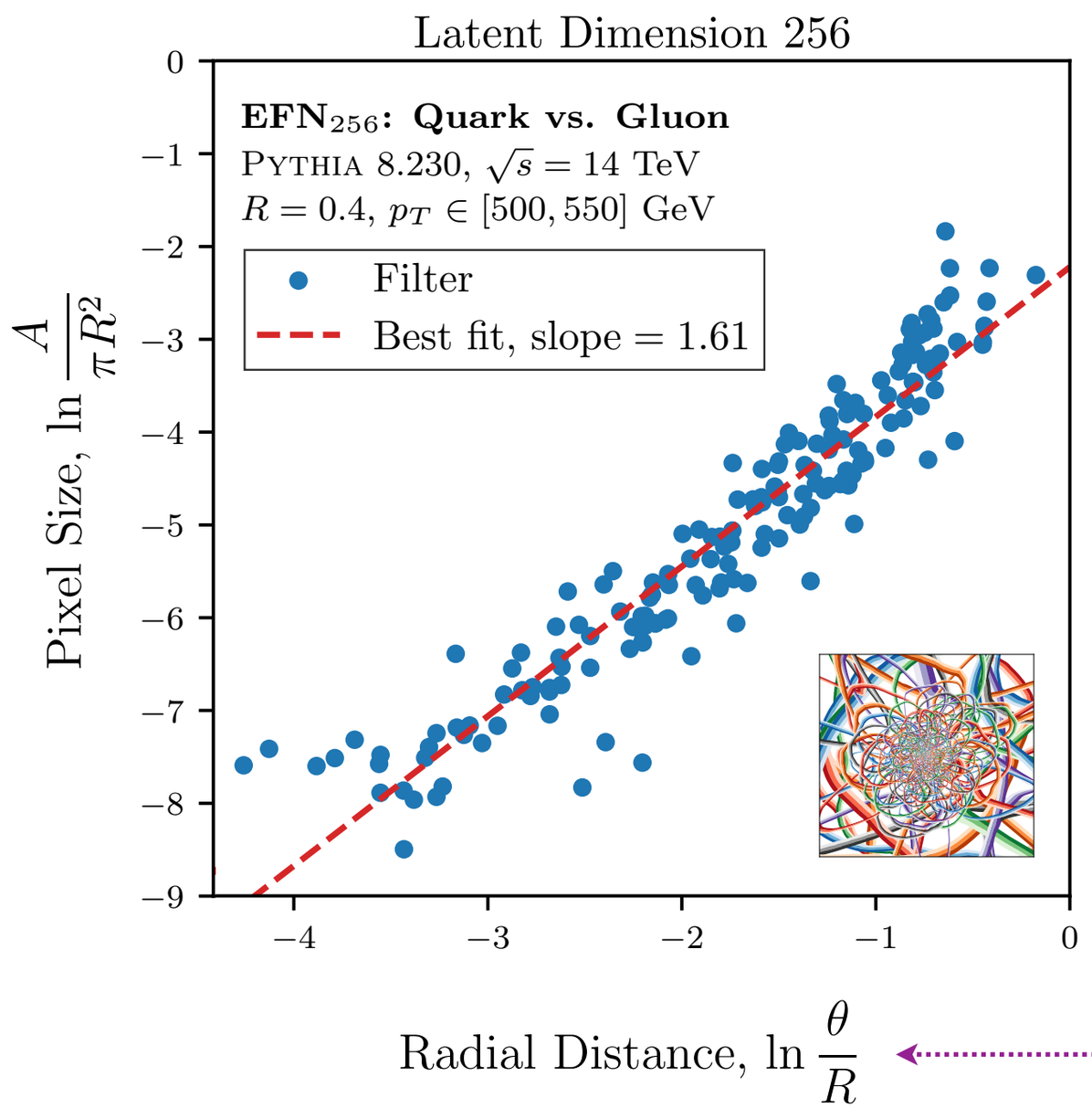


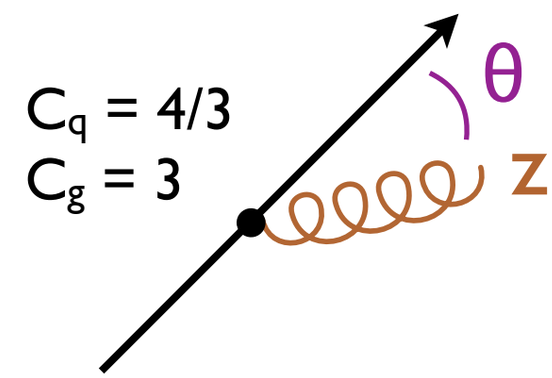
Latent Dimension 256



Collinear singularity of QCD!

Learning the Singularity Structure of QCD

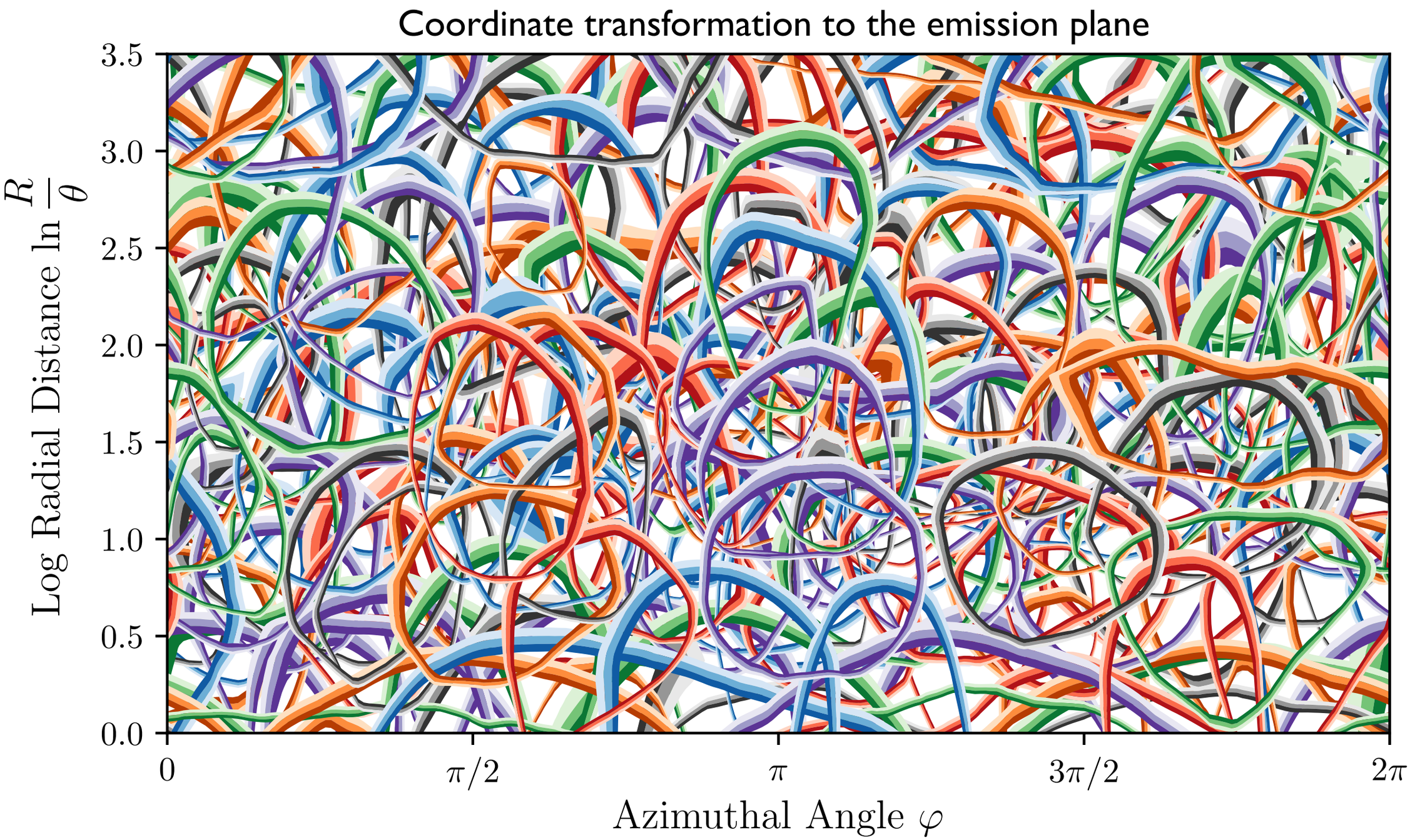




$$\mathrm{d}P_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{\mathrm{d}\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{\mathrm{d}z}{z}}_{\text{Soft}}$$

[Komiske, Metodiev, JDT, 1810.05165]

Ready for the Kunsthalle?



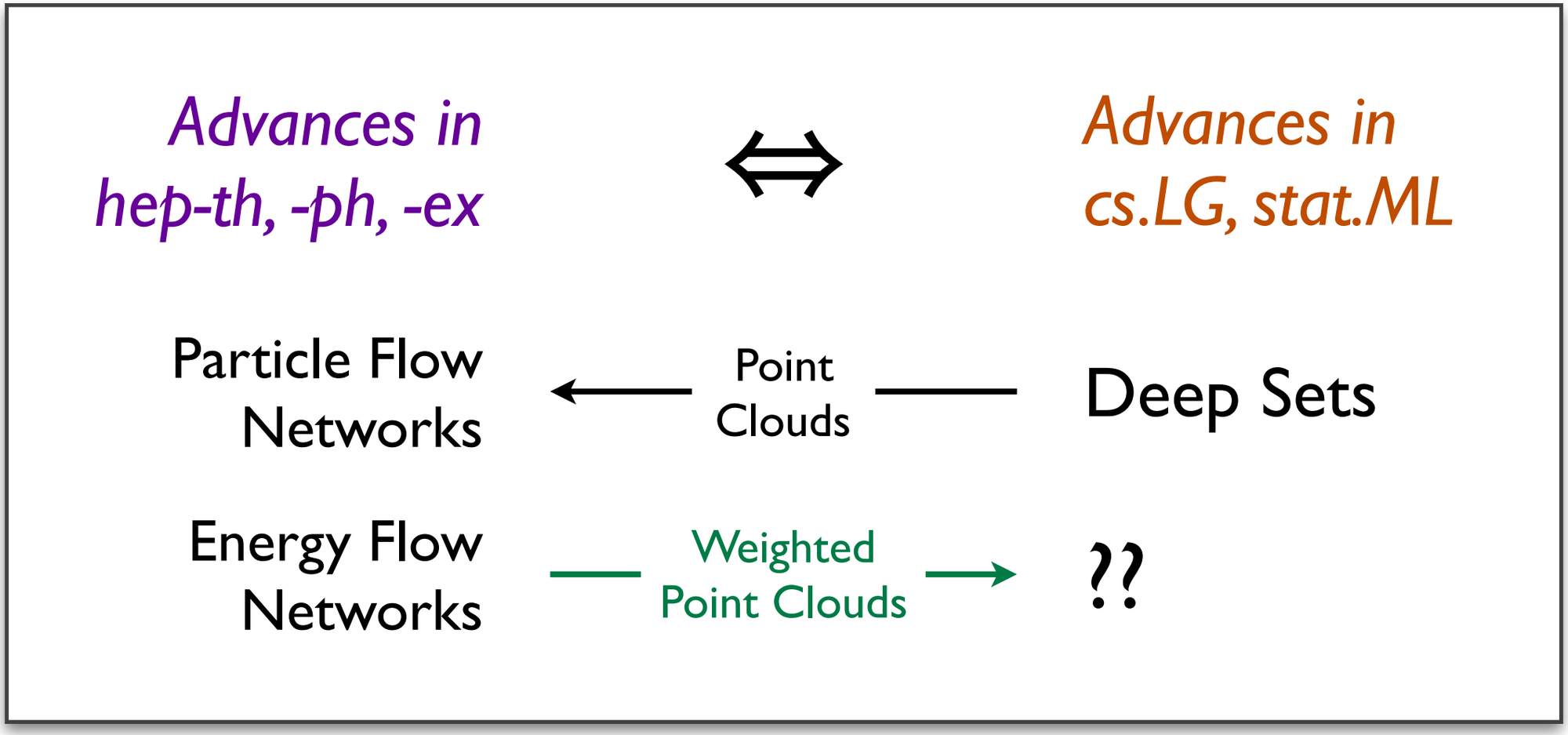
[Komiske, Metodiev, JDT, 1810.05165]

Particle Physics Primer

Point Clouds & Energy Flow Networks

Broader Lessons

“Collision Course”

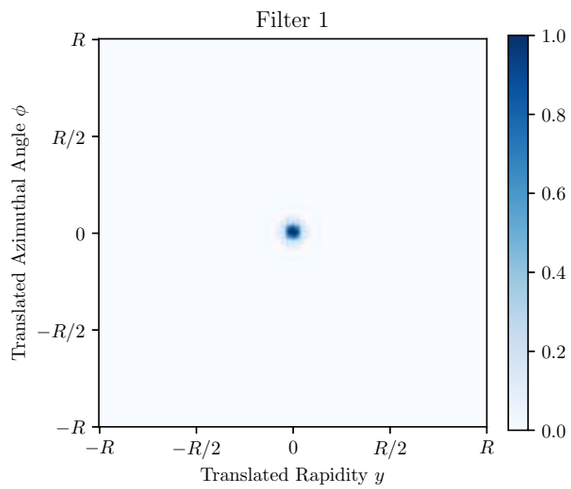


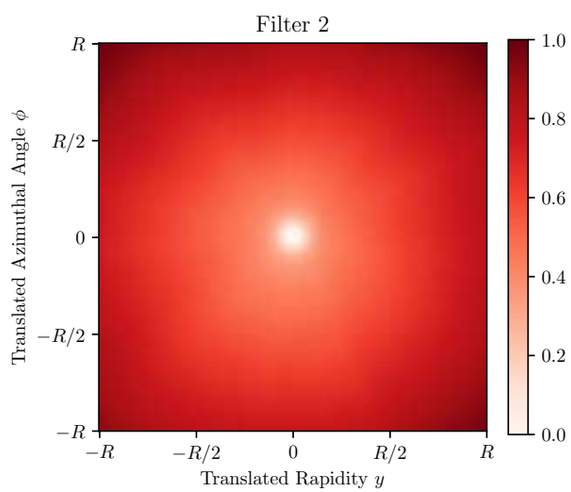
Particle physics is a fascinating domain with rich data sets, established algorithms, and strong theory priors

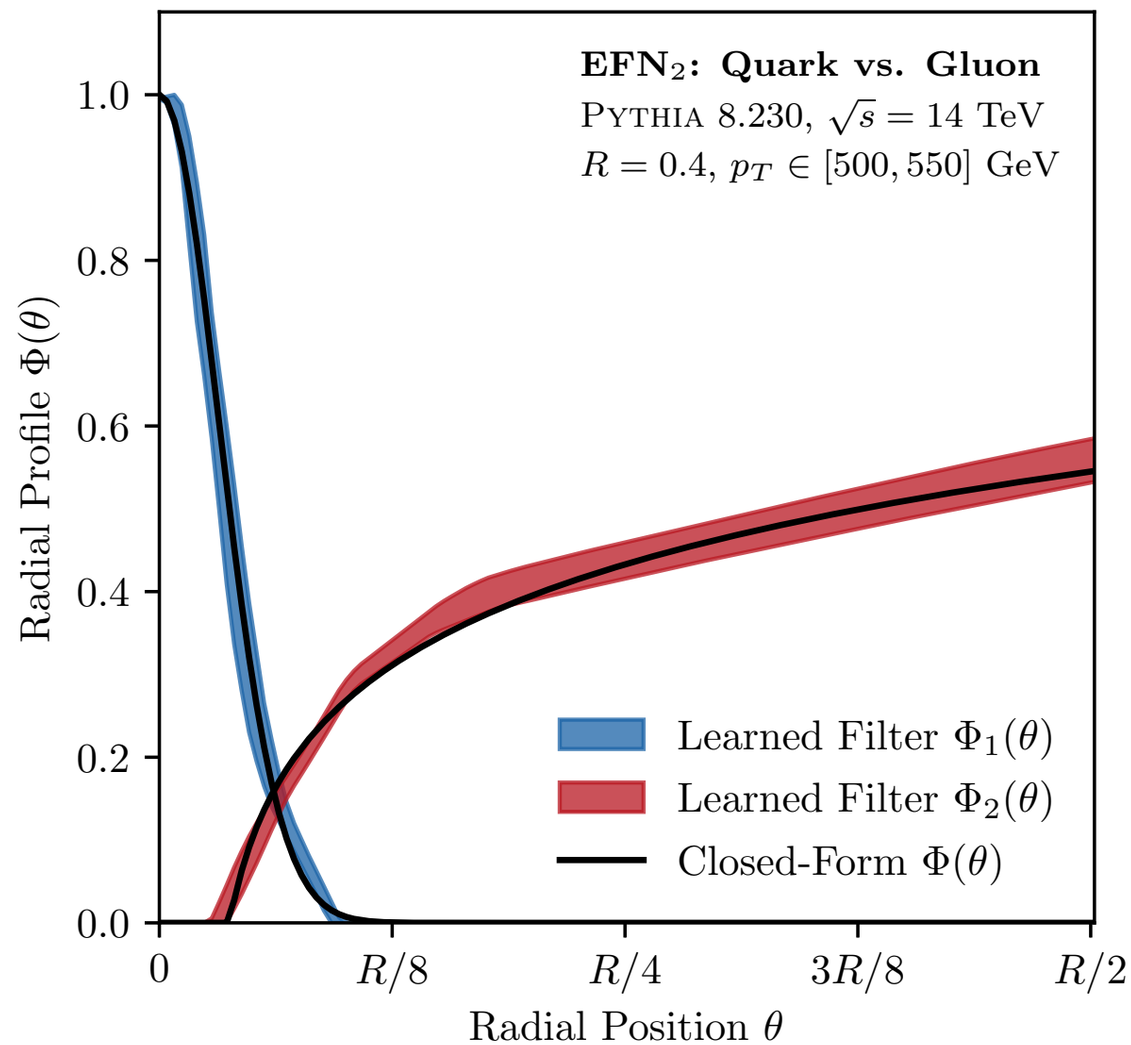
"What is the Machine Learning?"

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

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



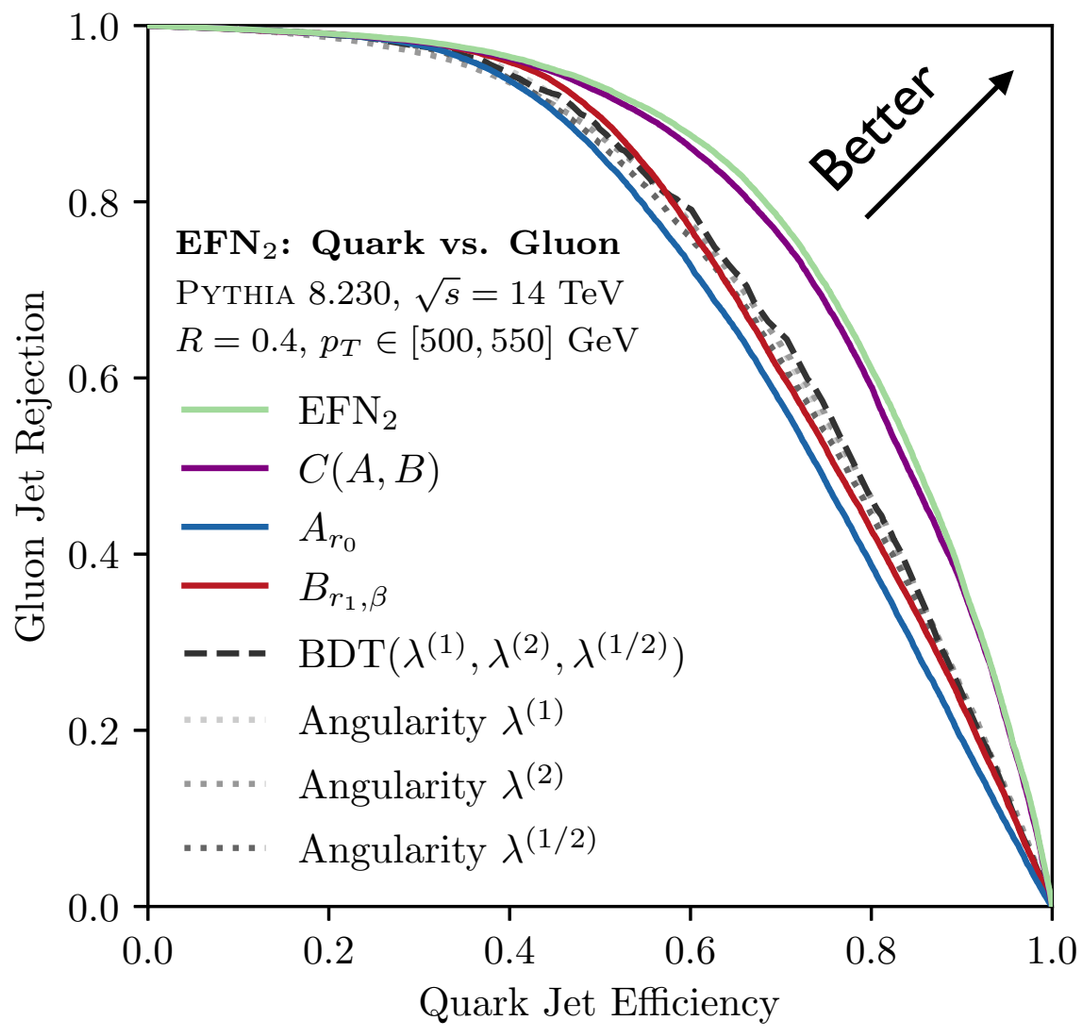




"What is the Machine Learning?"

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

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

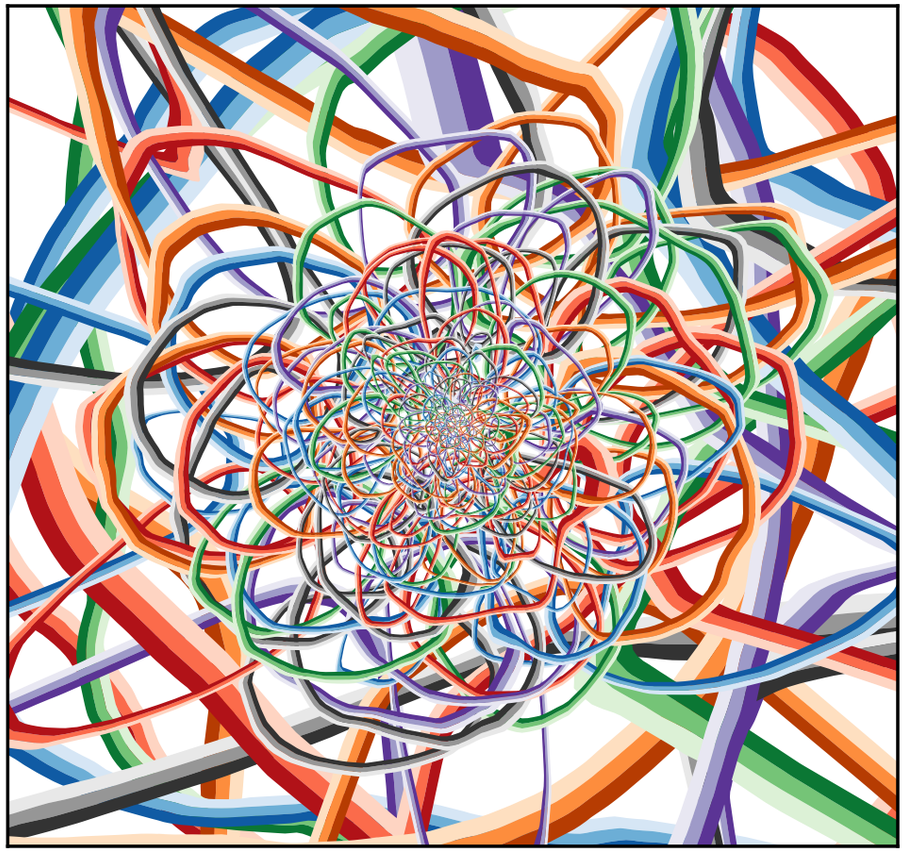


EFN outperformed a domain expert (i.e. me)

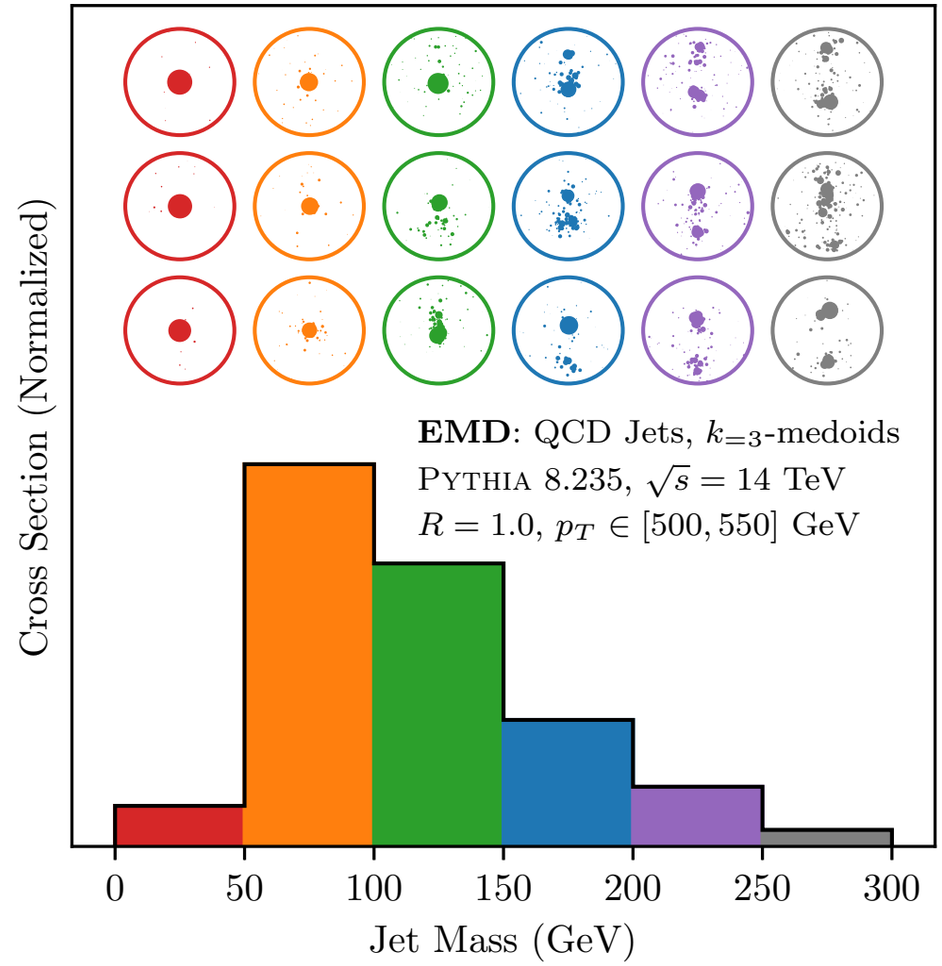
But we reverse engineered the machine (and learned something about QCD)

Opportunities for Network/Data Visualization

Latent Space



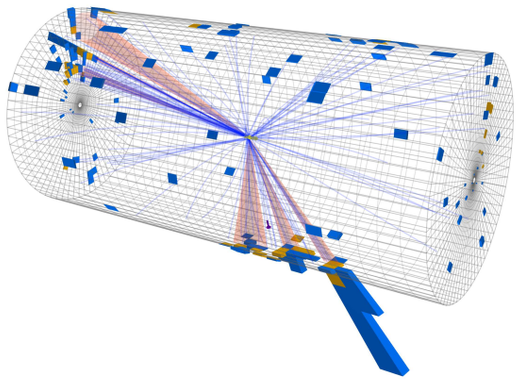
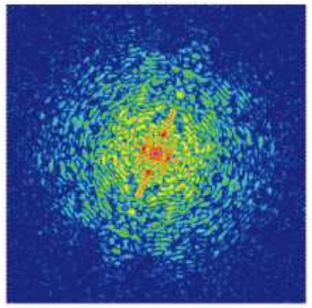
Metric Space



Augment (not replace) our exceptional human/scientific ability to recognize patterns

[Komiske, Metodiev, JDT, 1810.05165; Komiske, Metodiev, JDT, 1902.02346]

Questions for this Workshop

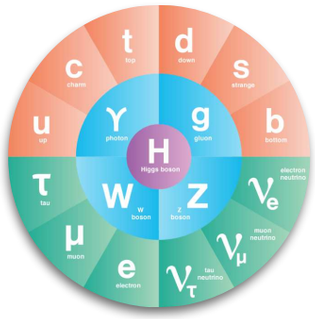
Connecting  ,  &  ?

What are the core principles that drive your approach to ML?

e.g.: Respect the structures & symmetries of datasets
Strive to balance performance, robustness & interpretability

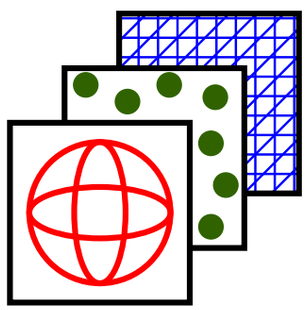
What are the state-of-the-art ML applications/tools in your field?

Summary



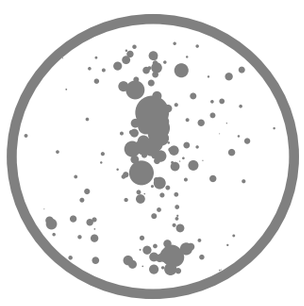
Particle Physics Primer

A rich domain with many machine learning opportunities



Point Clouds & Energy Flow Networks

A new architecture for weighted point clouds



Broader Lessons

Domain-specific knowledge is essential for ML developments

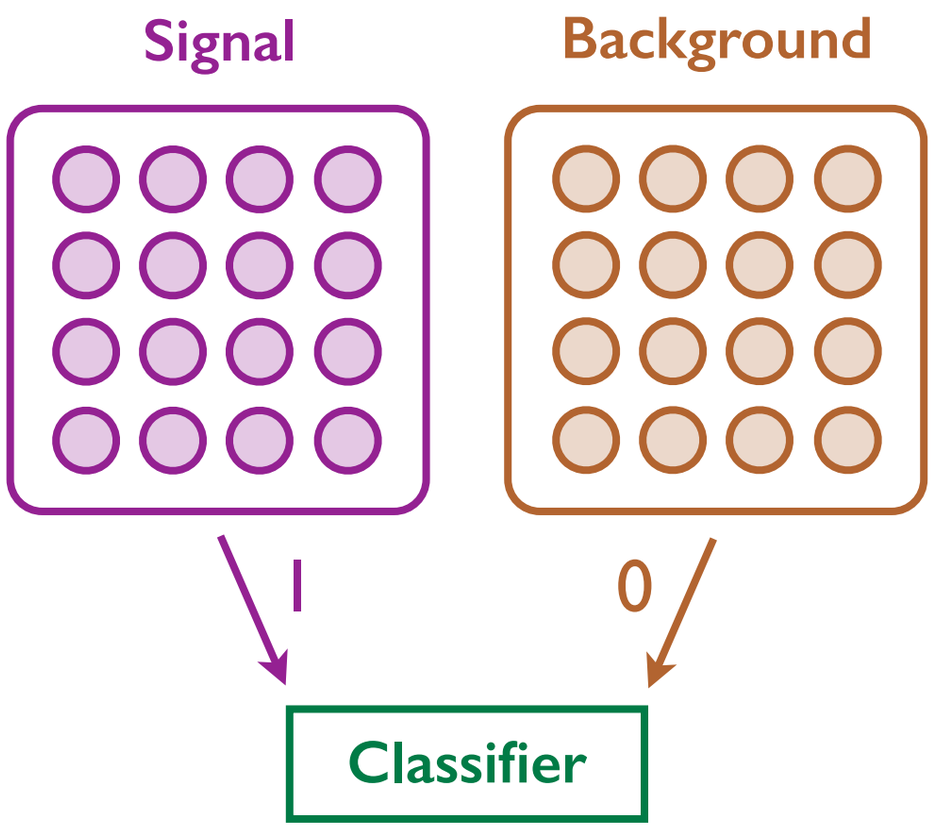
Backup Slides

A Cartoon of Machine Learning

For fully-supervised jet classification

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

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)

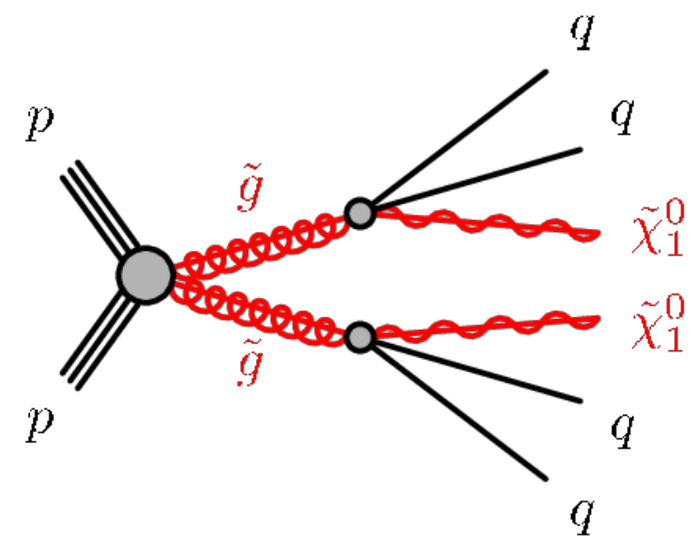
E.g. Search for Gluino Cascade to Dark Matter

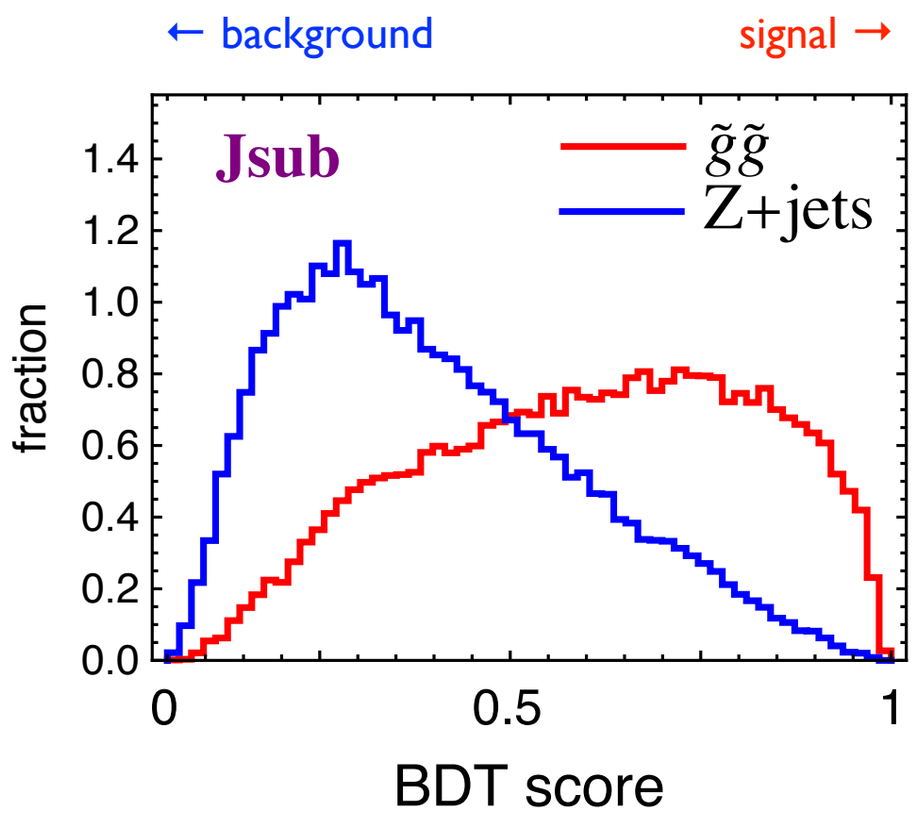
Classifier: Boosted decision tree (for each of 4 jets)

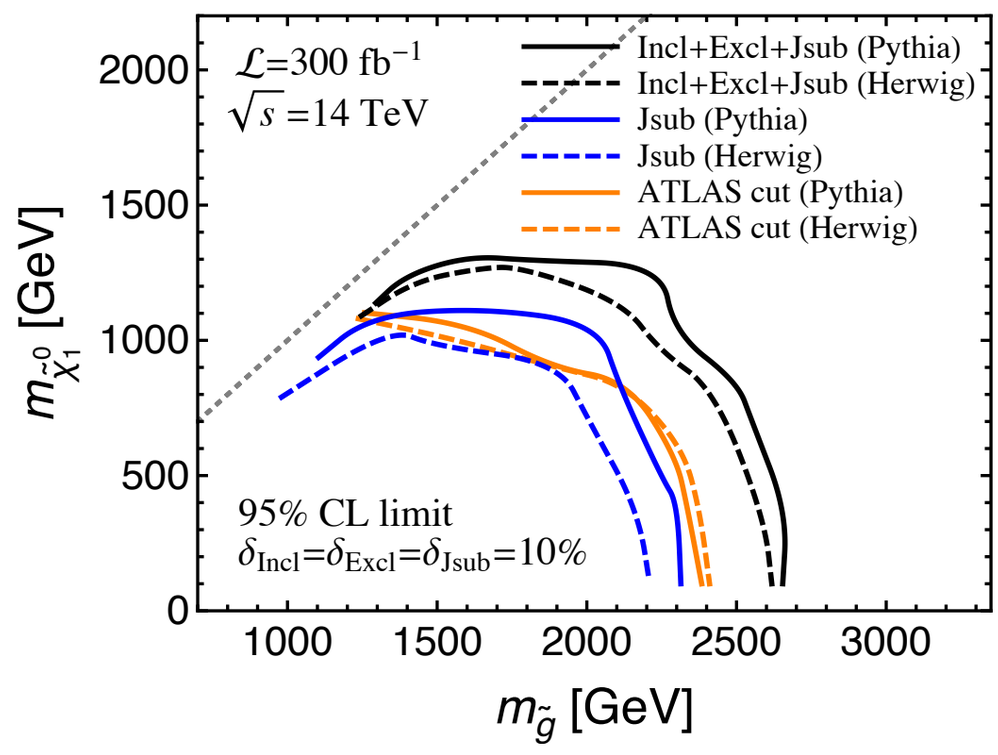
Inputs: Jet mass, width, track multiplicity

Signal: Quark enriched ($C_F = 4/3$)

Background: Gluon enriched ($C_A = 3$)







[Bhattacherjee, Mukhopadhyay, Nojiri, Sakakie, Webber, 1609.08781]

Jet Classification Studies

Mix and match

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

Classifier	Inputs	Signal vs. Background		
Boosted Decision Tree	High-Level Features	Quark Jets	vs.	Gluon Jets
Fisher Linear Discriminant	Basis of High-Level Features	Up-type Quarks	vs.	Down-type Quarks
Shallow Neural Network	Jet Image	W/Z Bosons	vs.	QCD Jets
Deep Neural Network	Multi-channel Jet Image	W Bosons	vs.	Z Bosons
Convolutional Neural Network	Abstract Jet Image	Top Quarks	vs.	QCD Jets
Recurrent Neural Network	Sorted Four-Vectors	Exotic Boosted Objects	vs.	QCD Jets
Recursive Neural Network	Clustered Four-Vectors	CMS Open Data Samples	vs.	Each other
Combination/Lorentz Layers	Lund Plane Emissions	…	vs.	…
…	Kitchen Sink			
	…			

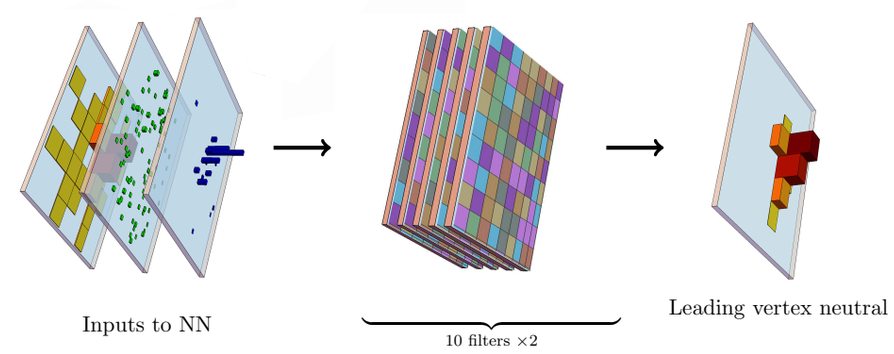
[Lönnblad, Peterson, Rögnvaldsson, 1990, …, Cogan, Kagan, Strauss, Schwartzman, 1407.5675; Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968; de Oliveira, Kagan, Mackey, Nachman, Schwartzman, 1511.05190; Baldi, Bauer, Eng, Sadowski, Whiteson, 1603.09349; Conway, Bhaskar, Erbacher, Pilot, 1606.06859; 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 my friends who will scold me for forgetting their paper (and not updating this after July 23, 2018); plus many ATLAS/CMS performance studies]

Beyond Classification

PUMML

Pileup Mitigation

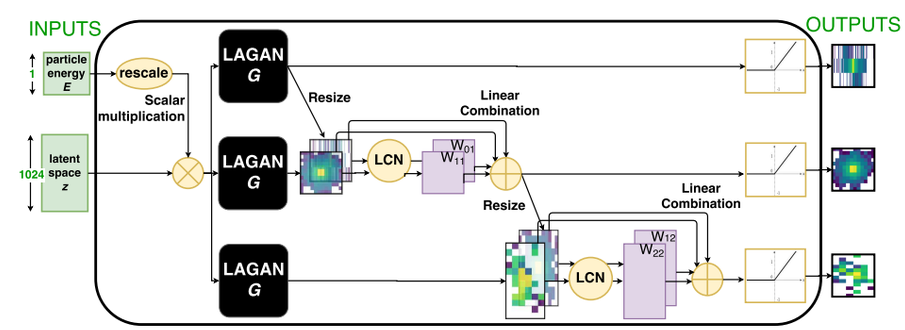
[Komiske, Metodiev, Nachman, Schwartz, 1707.08600]



CaloGAN

Fast Detector Simulation

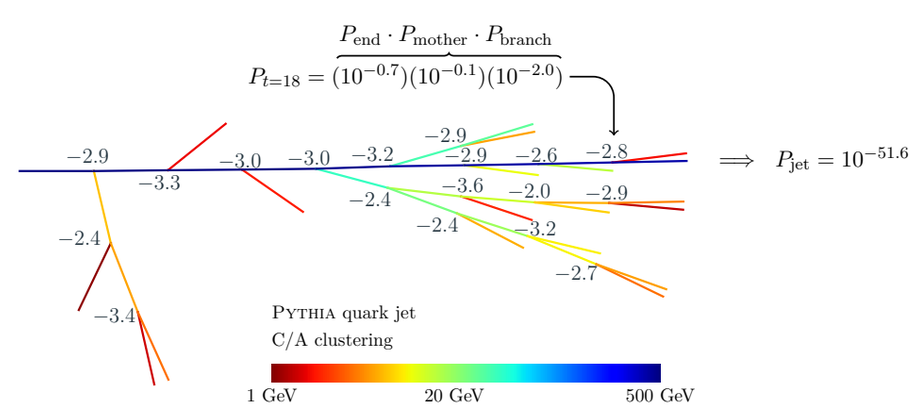
Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321;
see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



JUNIPR

Probability Modeling

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

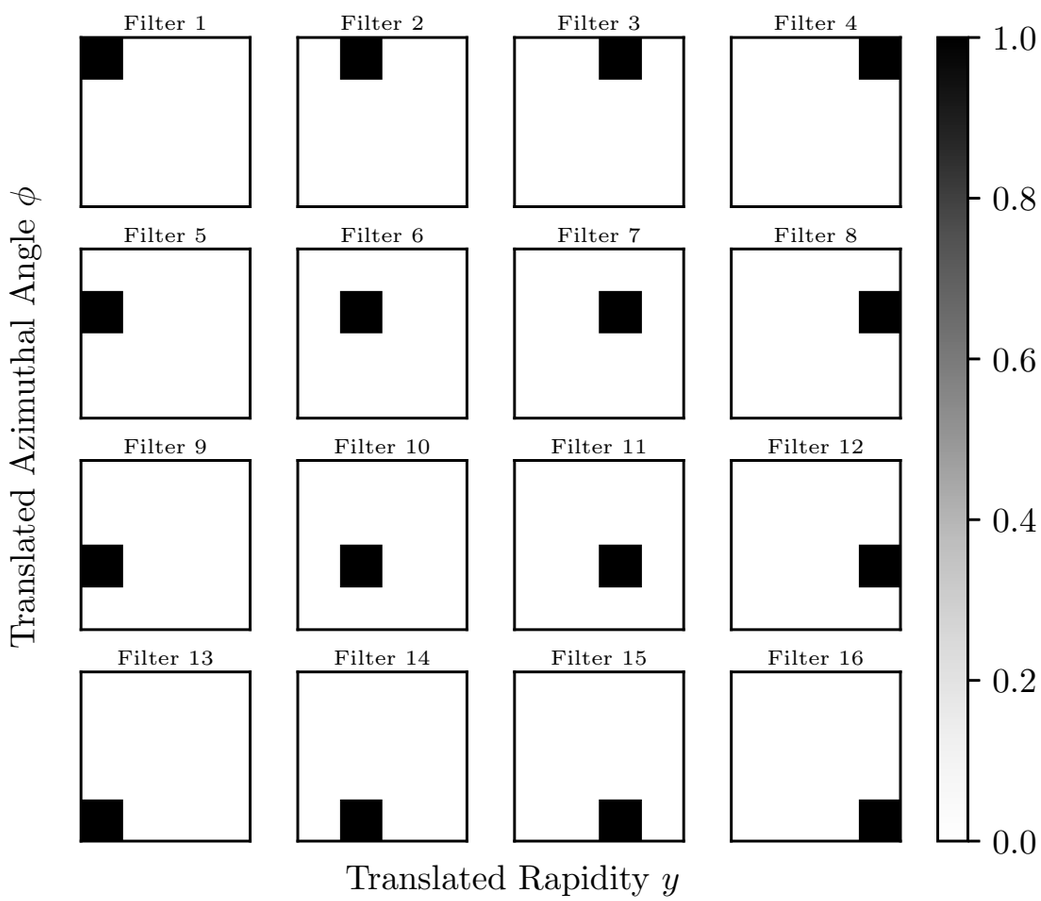


Latent Space Visualization

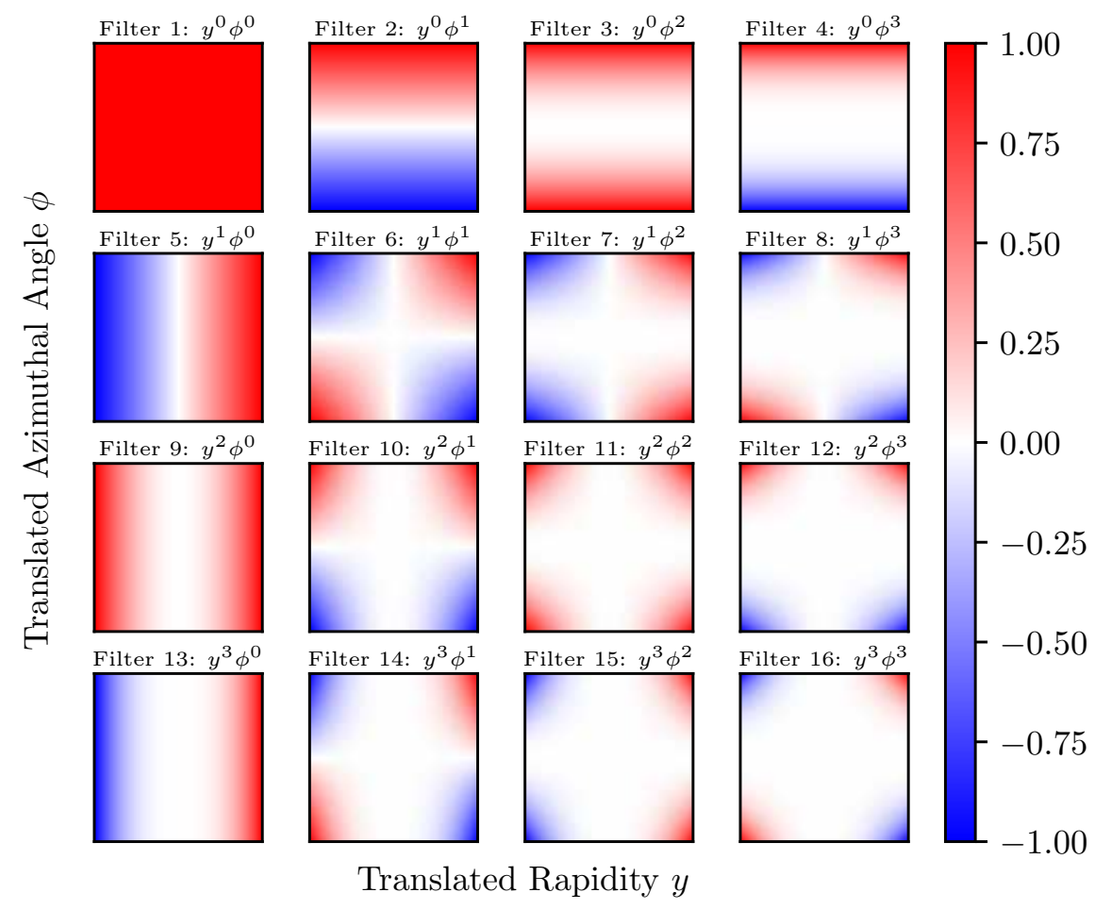
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

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

Calorimeter Pixels



Radiation Moments

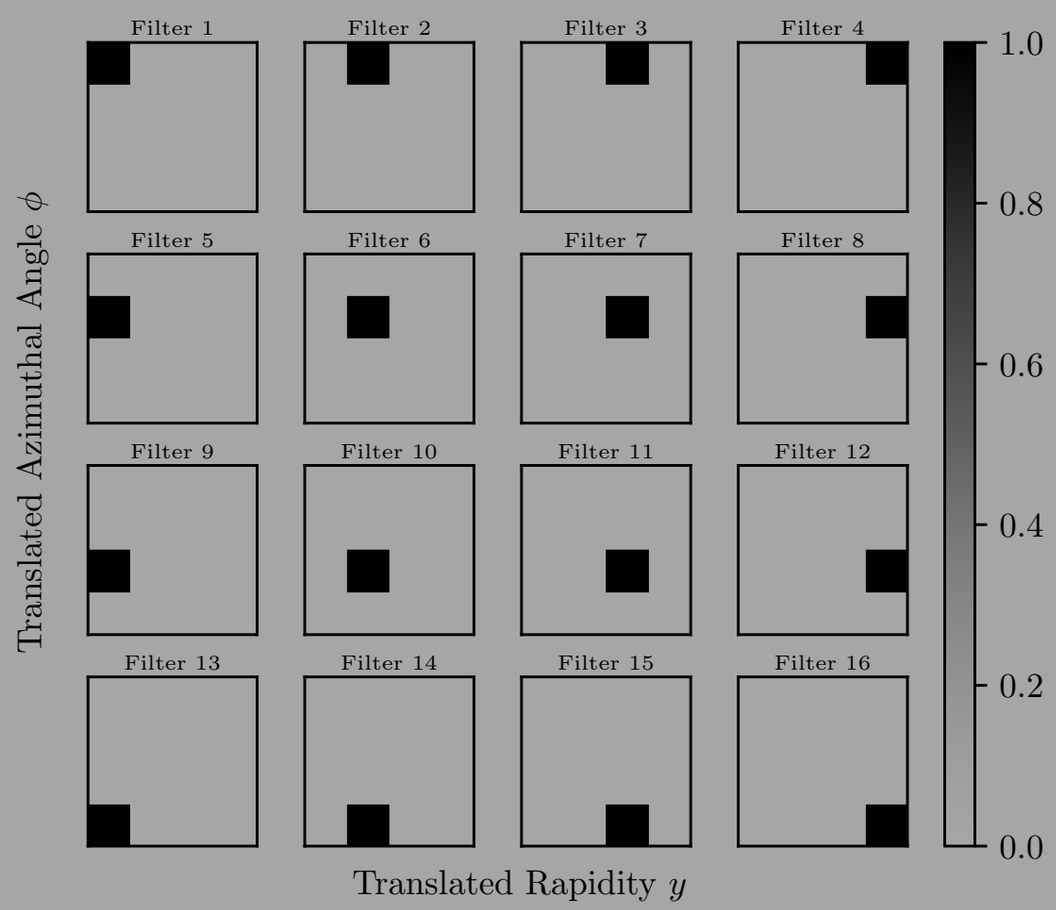


Latent Space Visualization

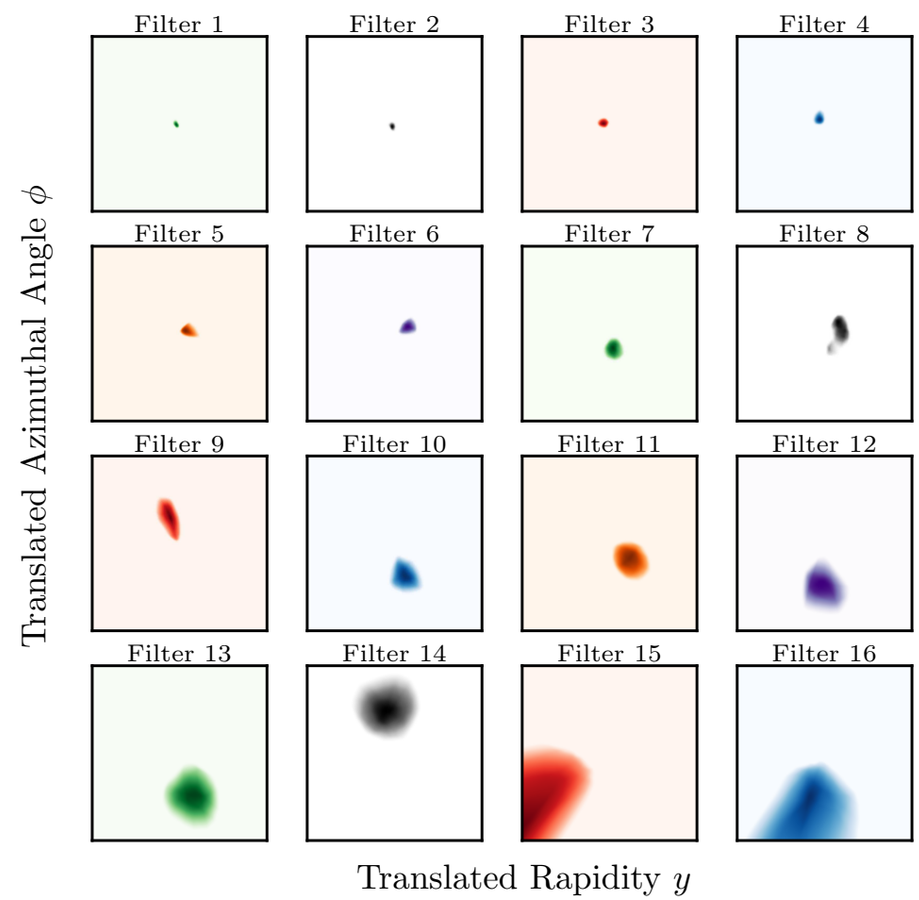
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

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

Calorimeter Pixels

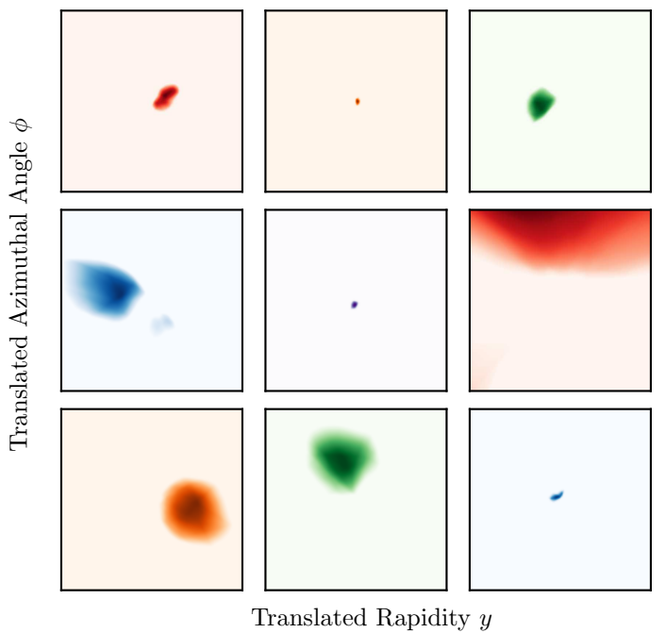


EFNs: Dynamic Pixelation

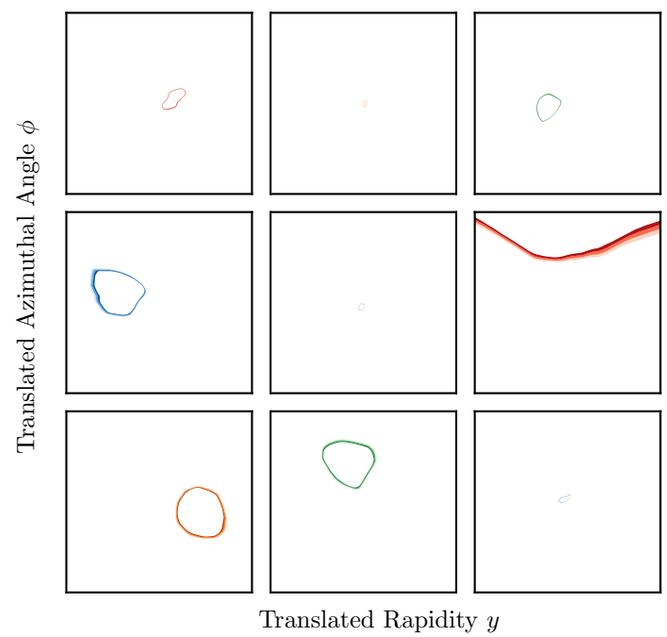


Psychedelic Network Visualization

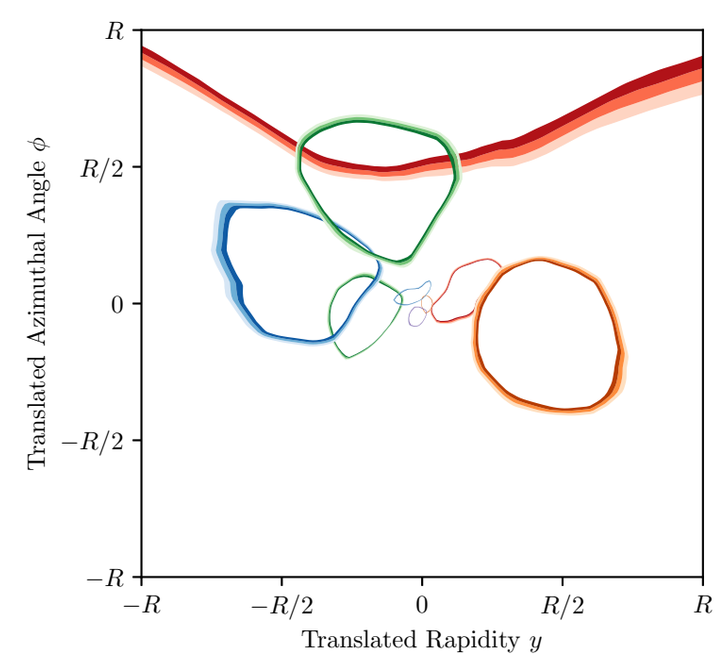
Latent Filters



50% Contours



Overlay



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