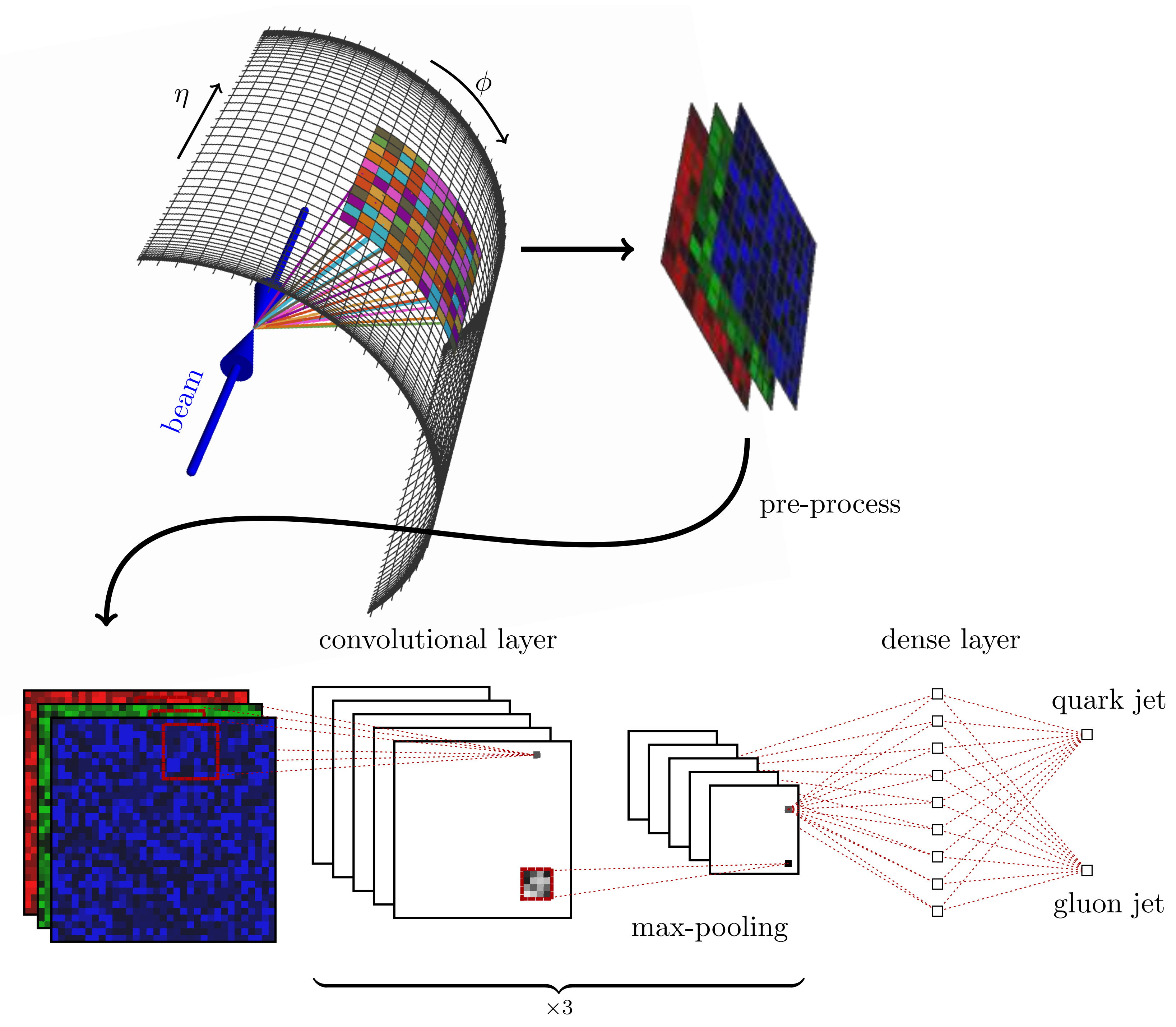
Collision Course

Particle Physics as a Machine-Learning Testbed

Jesse Thaler

PDT Partners — October 16, 2018

The Rise of Machine Learning for Colliders



[e.g. Komiske, Metodiev, Schwartz, 1612.01551]

My Perspective c. 2016

“Deep Learning” vs. “Deep Thinking”

My Perspective c. 2018

“Deep Learning” ~~vs.~~ & “Deep Thinking”

Advances in physics facilitated by advances in mathematics, statistics, and computer science

And vice versa?

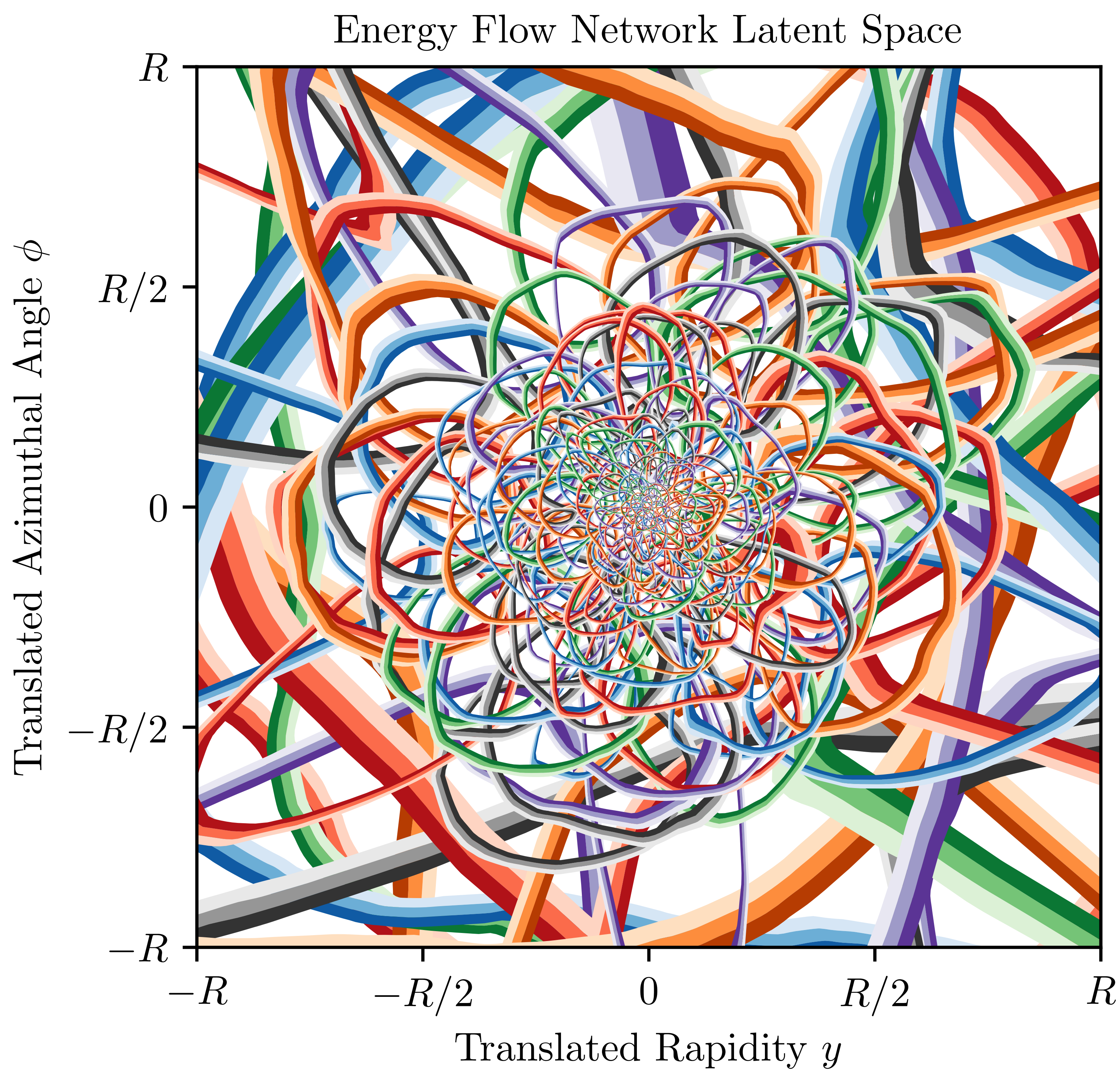
Proximate Reasons for My Conversion

Patrick Komiske

Eric Metodiev

plus Ben Nachman, Kyle Cranmer, Daniel Whiteson, Mike Williams, Matt Schwartz, Dan Roberts, Phiala Shanahan, …

Physics Reasons for My Conversion



Today's Talk

(see backup for topic modeling)

Collider physics meets neural networks with same underlying symmetries as…

[Komiske, Metodiev, JDT, 1810.05165; see also Komiske, Metodiev, JDT, 1712.07124]

Point Clouds

[Popular Science, 2013]

Particle Physics as ML Testbed

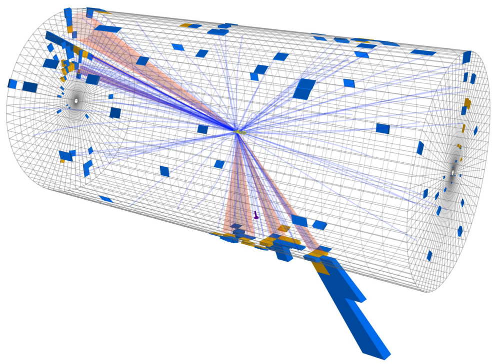
&

“Deep Learning” ~~vs.~~ “Deep Thinking”

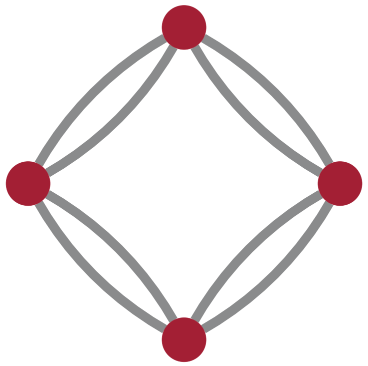
Particle physics is a fascinating domain with rich data sets, established algorithms, and first-principles calculations

Today: designing a machine to “think like a physicist”

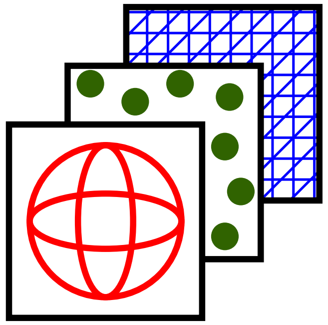
Outline



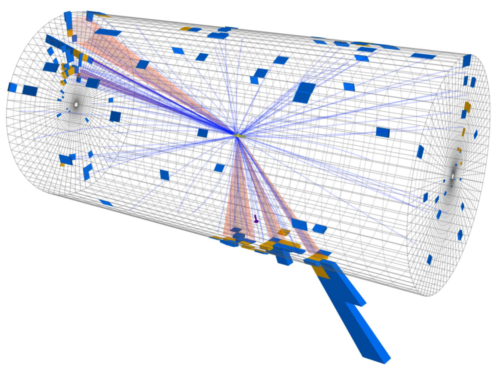
Jets at the LHC



The Importance of Symmetries



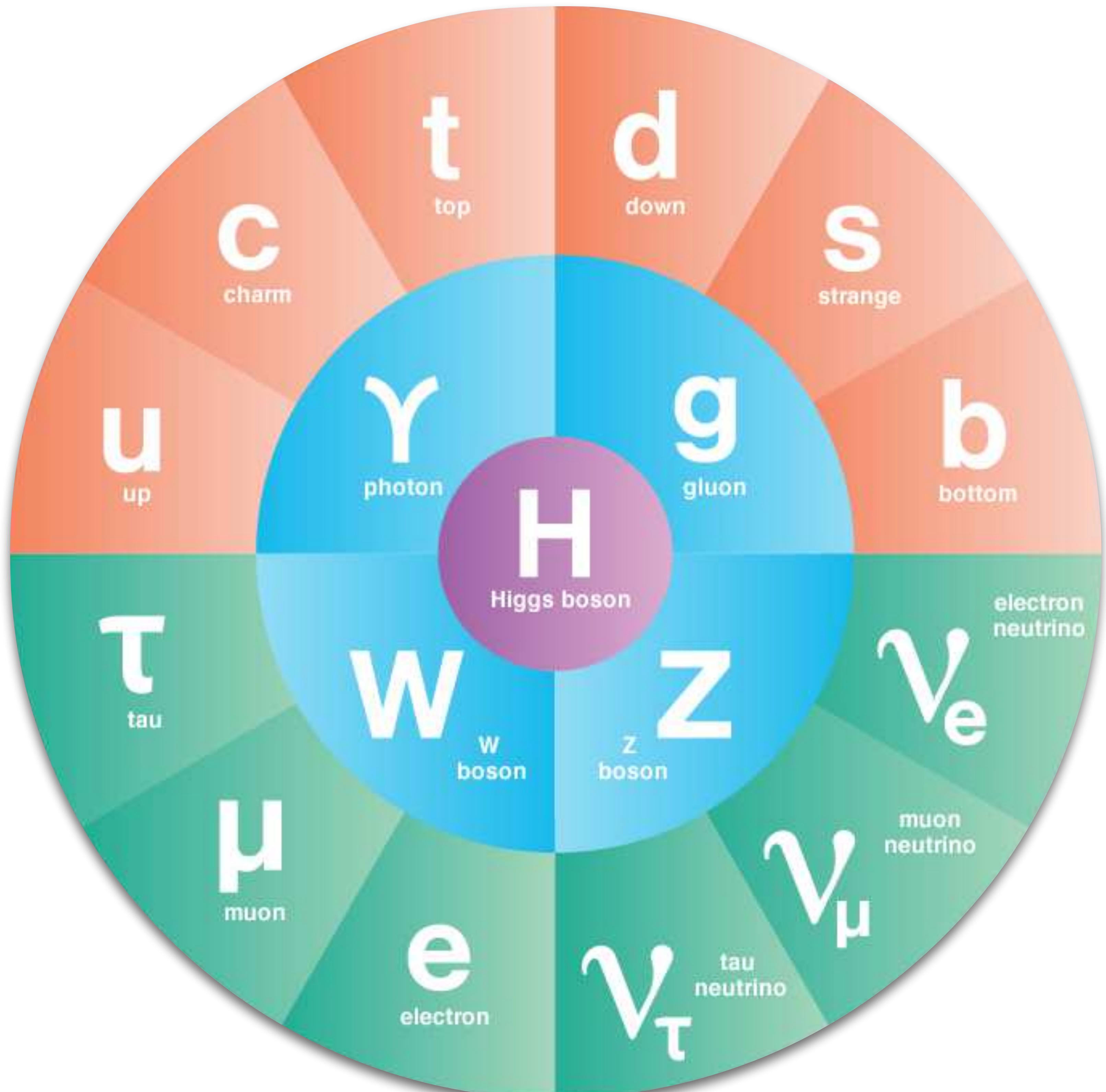
Energy Flow Networks

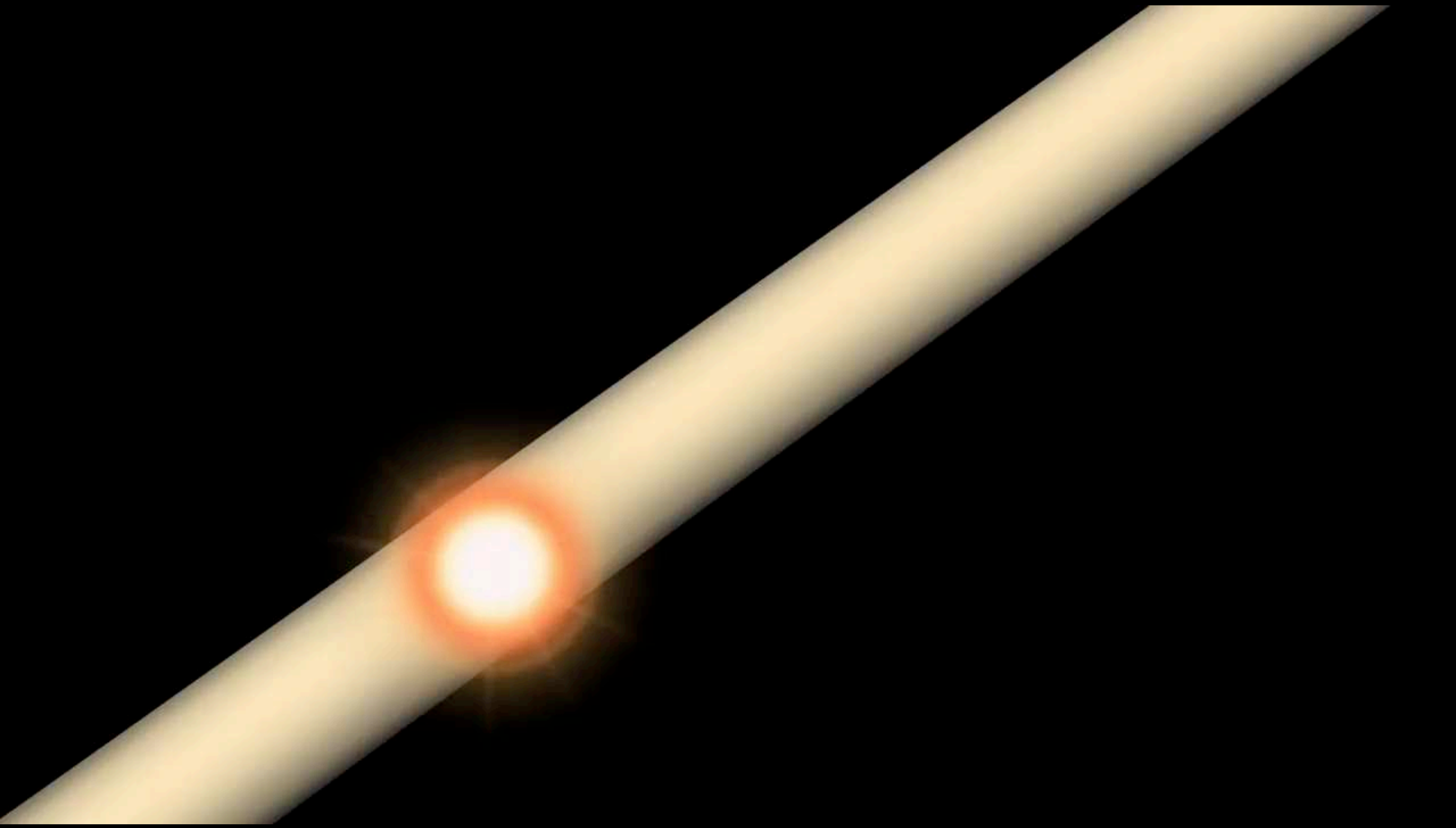


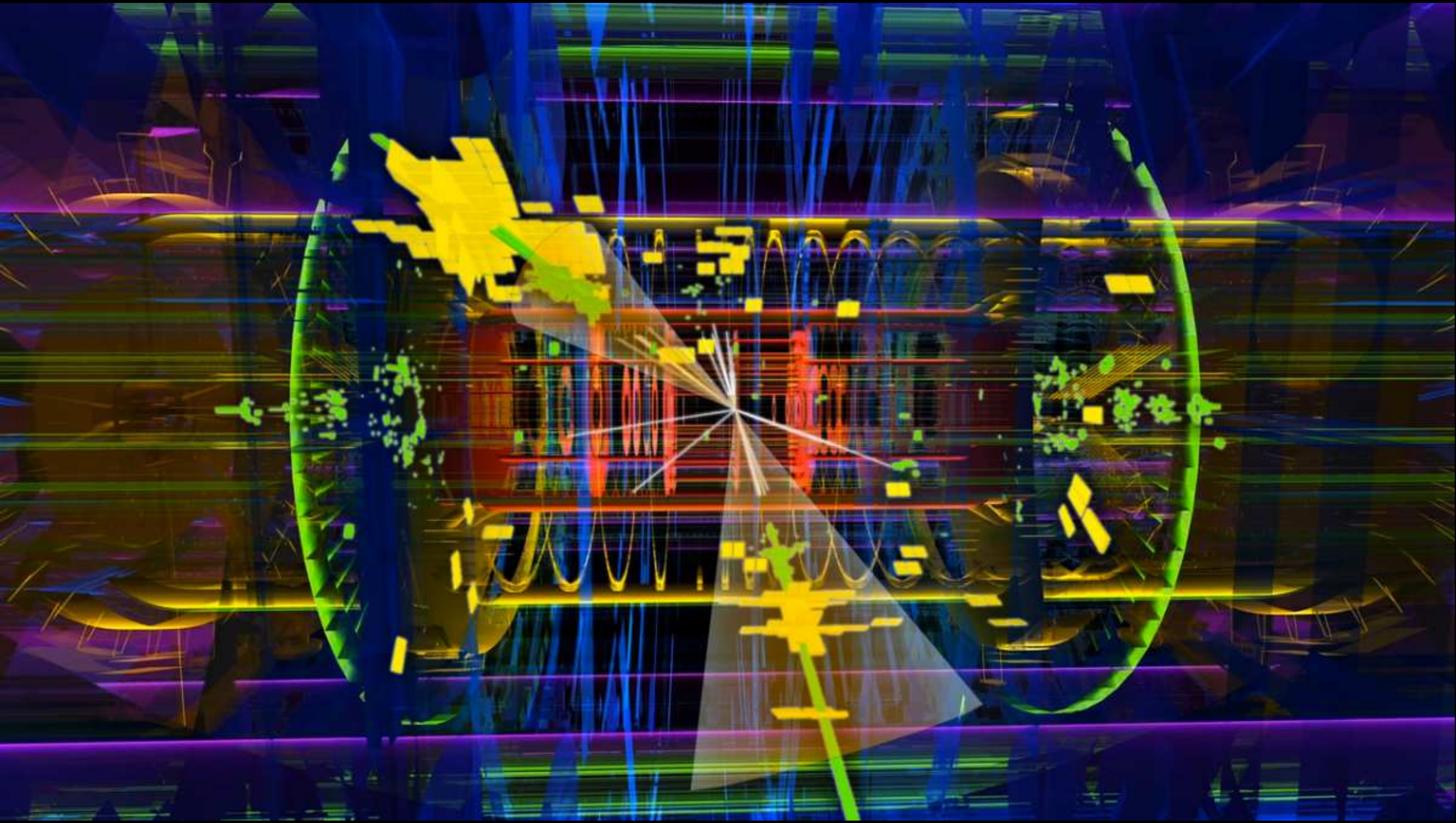
Jets at the LHC

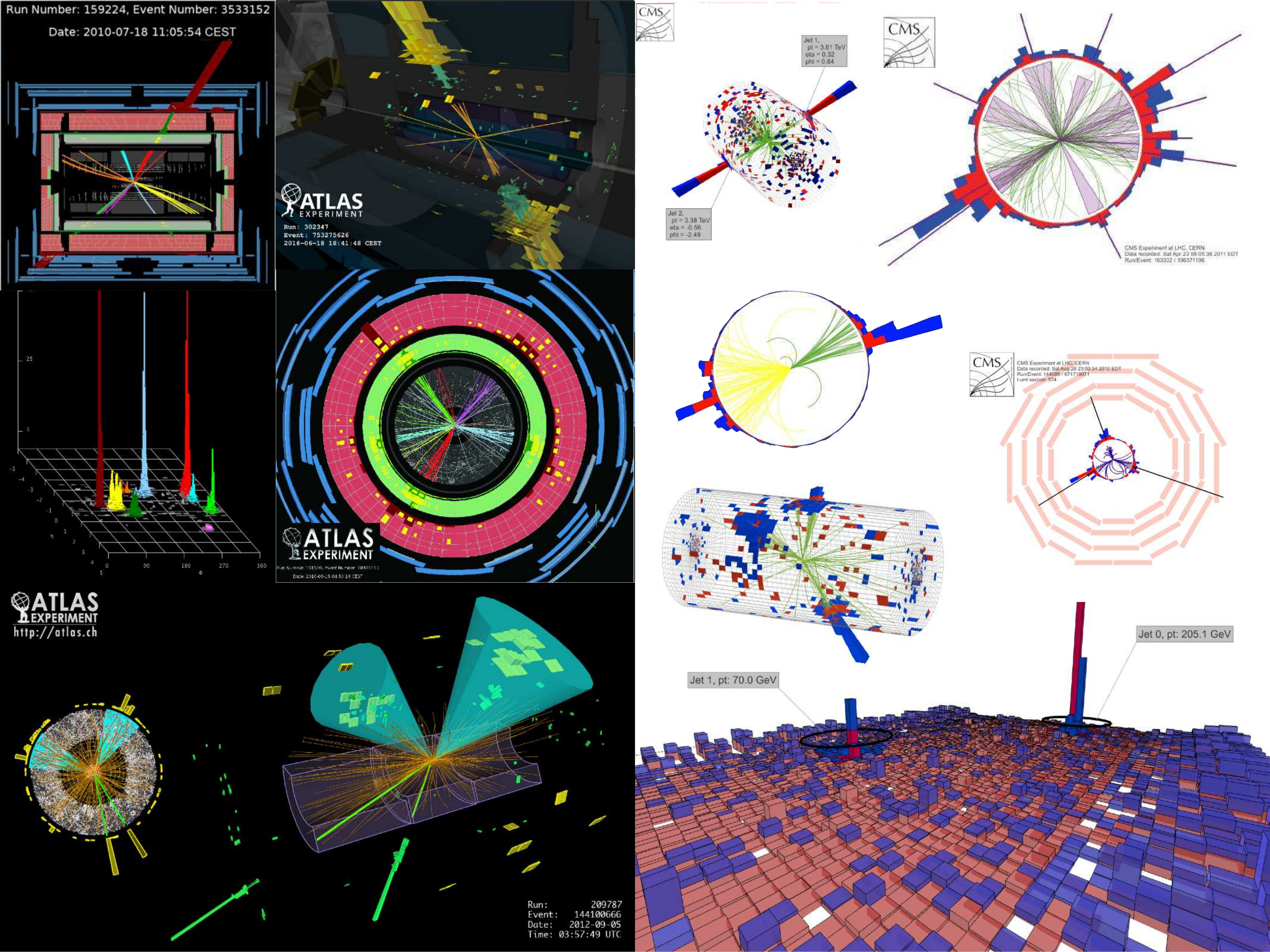
The Importance of Symmetries

Energy Flow Networks

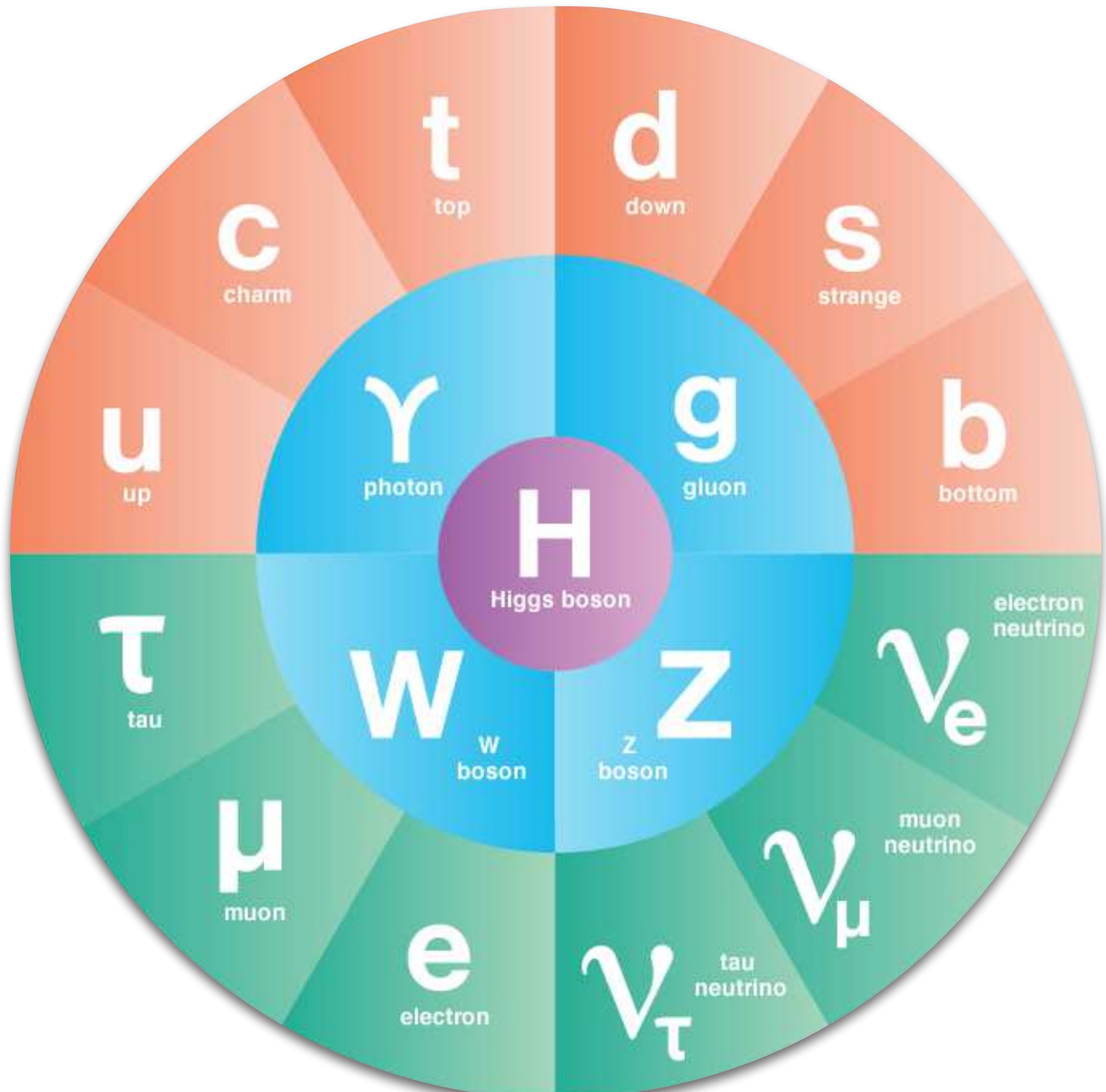

t
top
d
down
c
charm
s
strange
u
up
γ
photon
g
gluon
b
bottom
H
Higgs boson
τ
tau
W
W
boson
Z
Z
boson
ν_e
electron
neutrino
μ
muon
ν_μ
muon
neutrino
e
electron
ν_τ
tau
neutrino

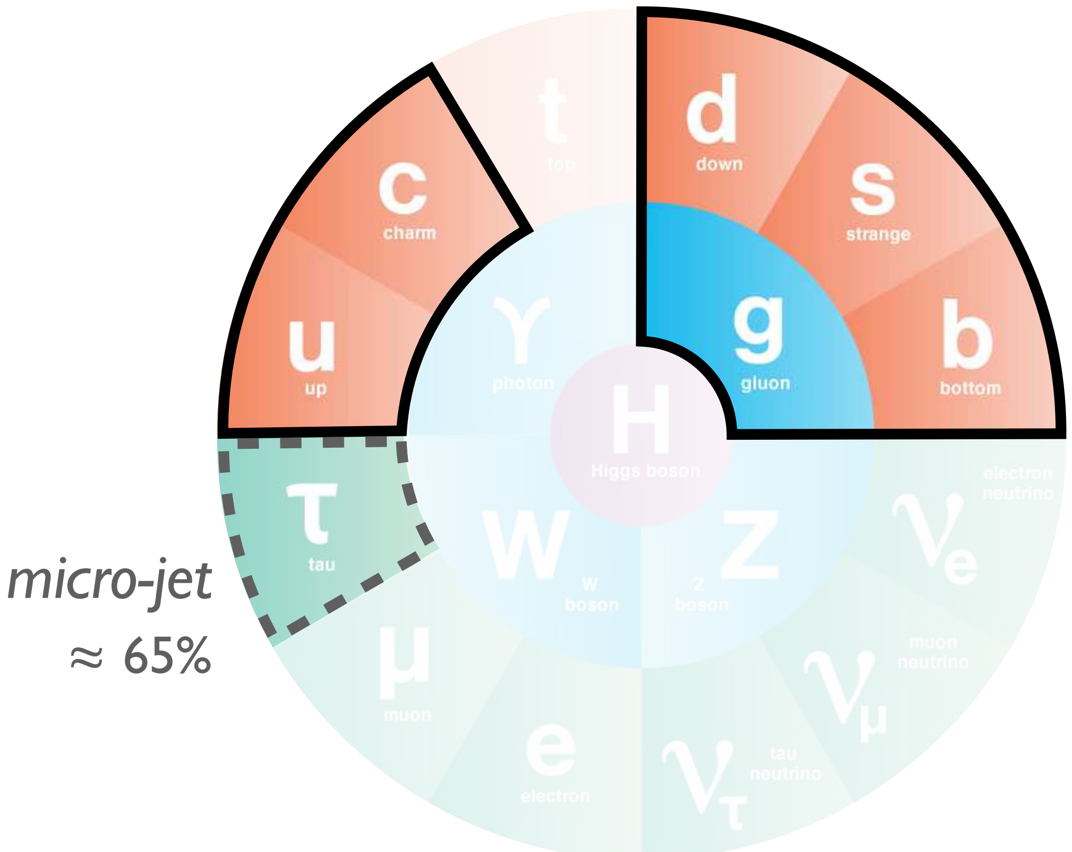






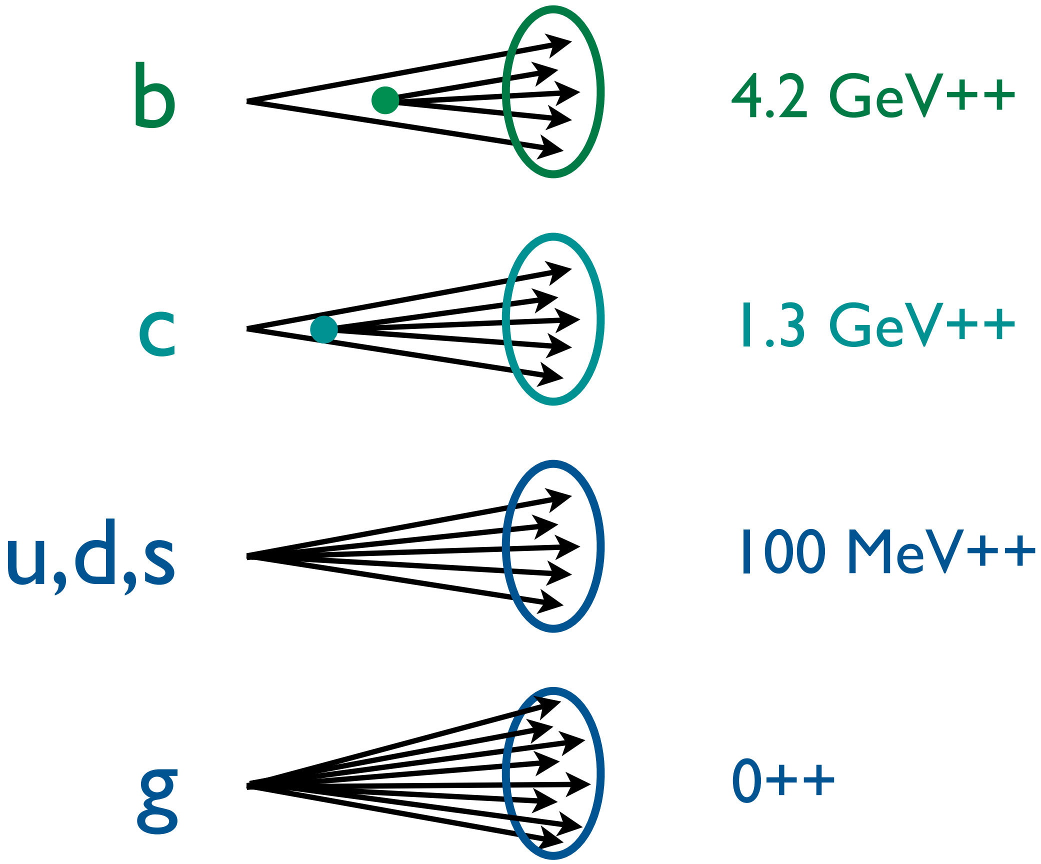
Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST
CMS
Jet 1,
pt = 3.61 TeV
eta = 0.32
phi = 0.64
Jet 2,
pt = 3.38 TeV
eta = -0.56
phi = -2.49
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011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

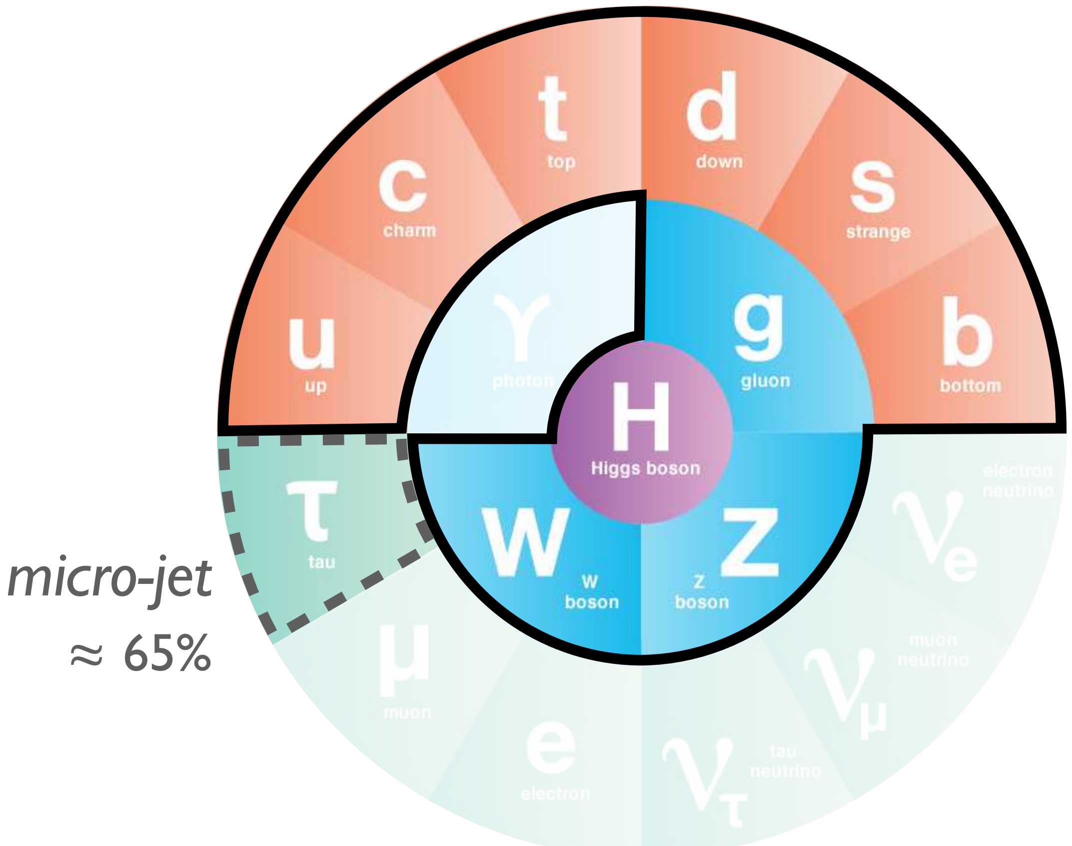

t
top
d
down
c
charm
s
strange
u
up
γ
photon
H
Higgs boson
g
gluon
b
bottom
τ
tau
W
W
boson
Z
Z
boson
ν_e
electron
neutrino
μ
muon
ν_μ
muon
neutrino
e
electron
ν_τ
tau
neutrino



Jets from the Standard Model

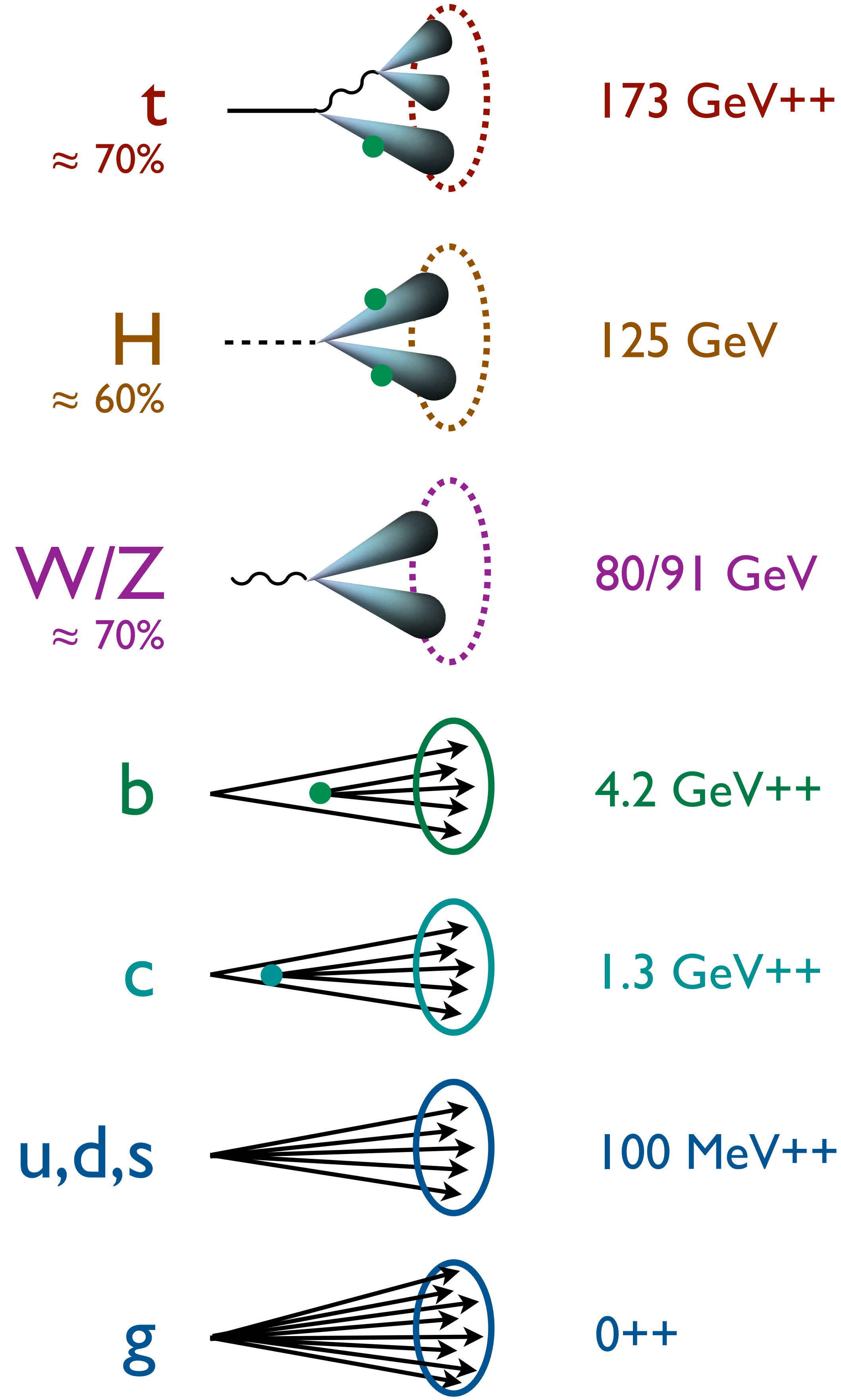
++ = Mass from QCD Radiation

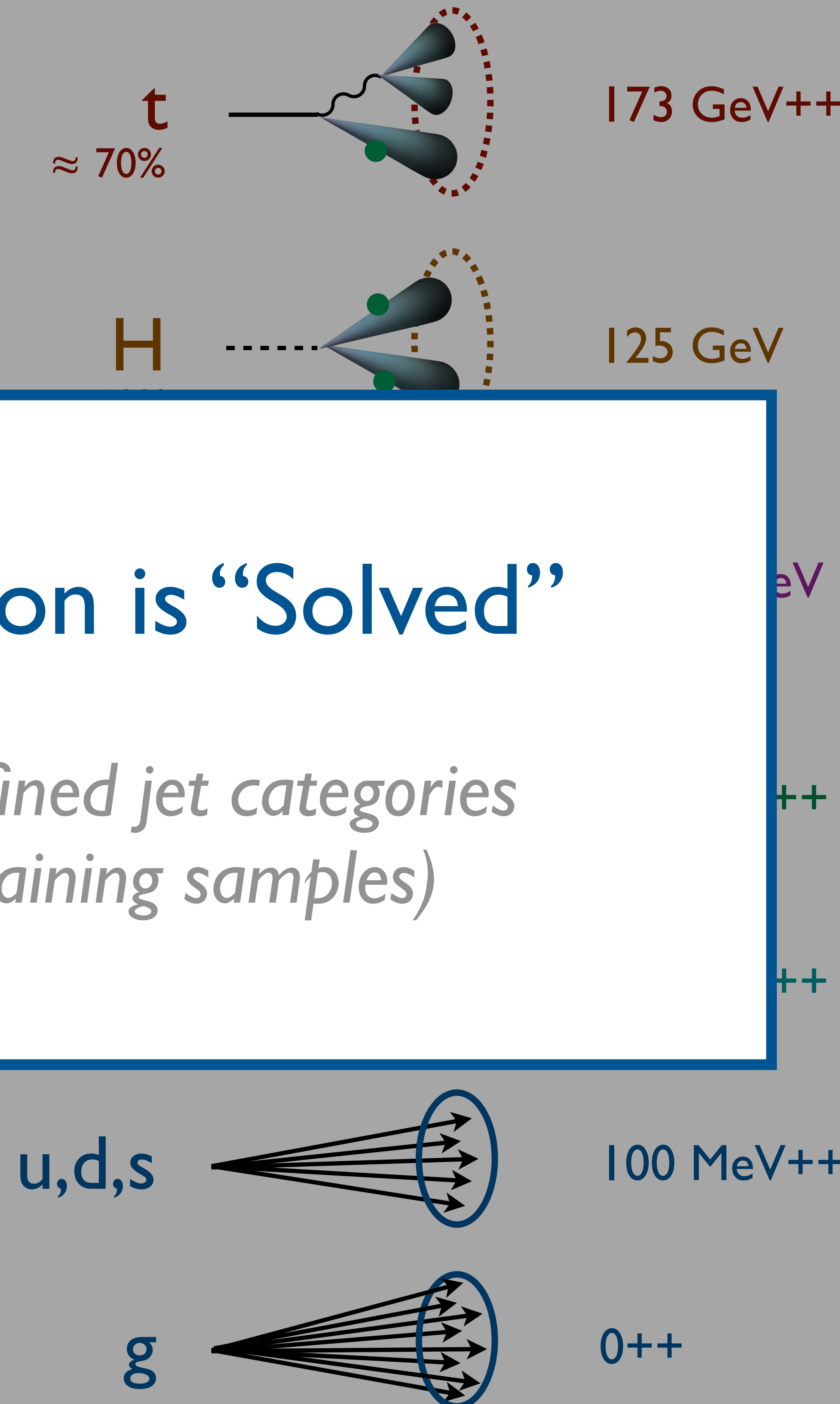


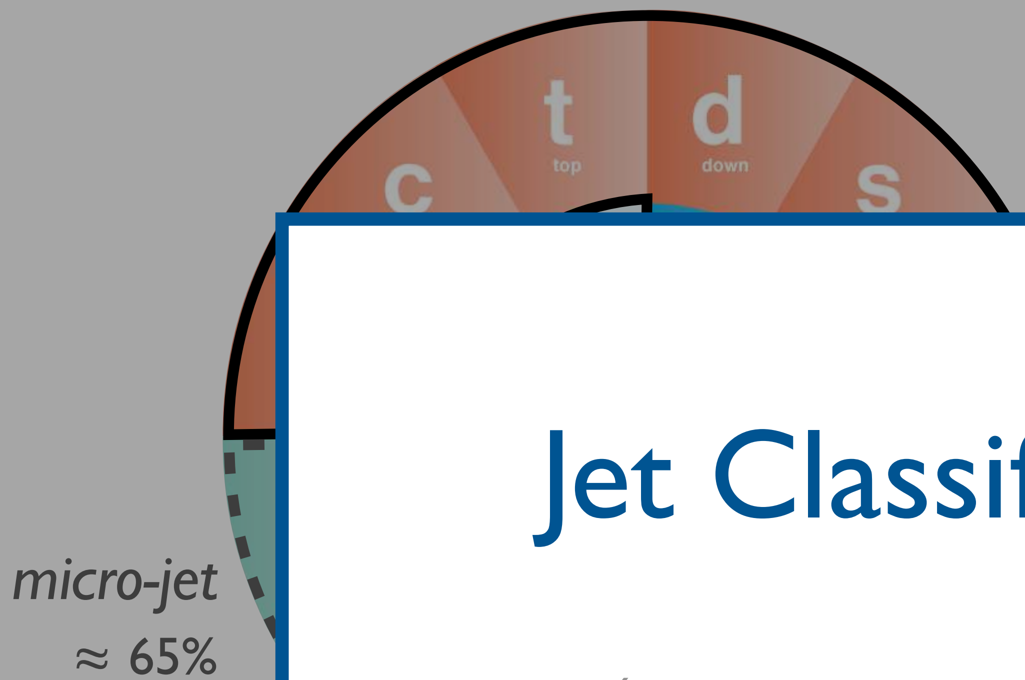


Jets from the Standard Model

++ = Mass from QCD Radiation





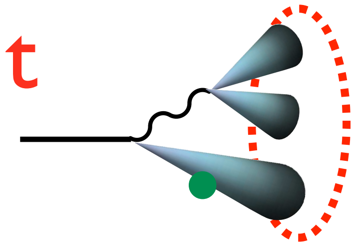


Jet Classification is “Solved”

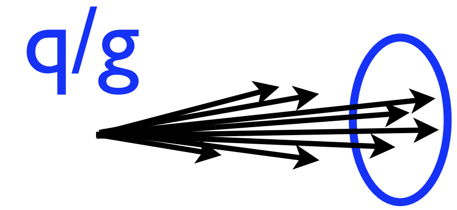
(assuming well-defined jet categories and trustable training samples)

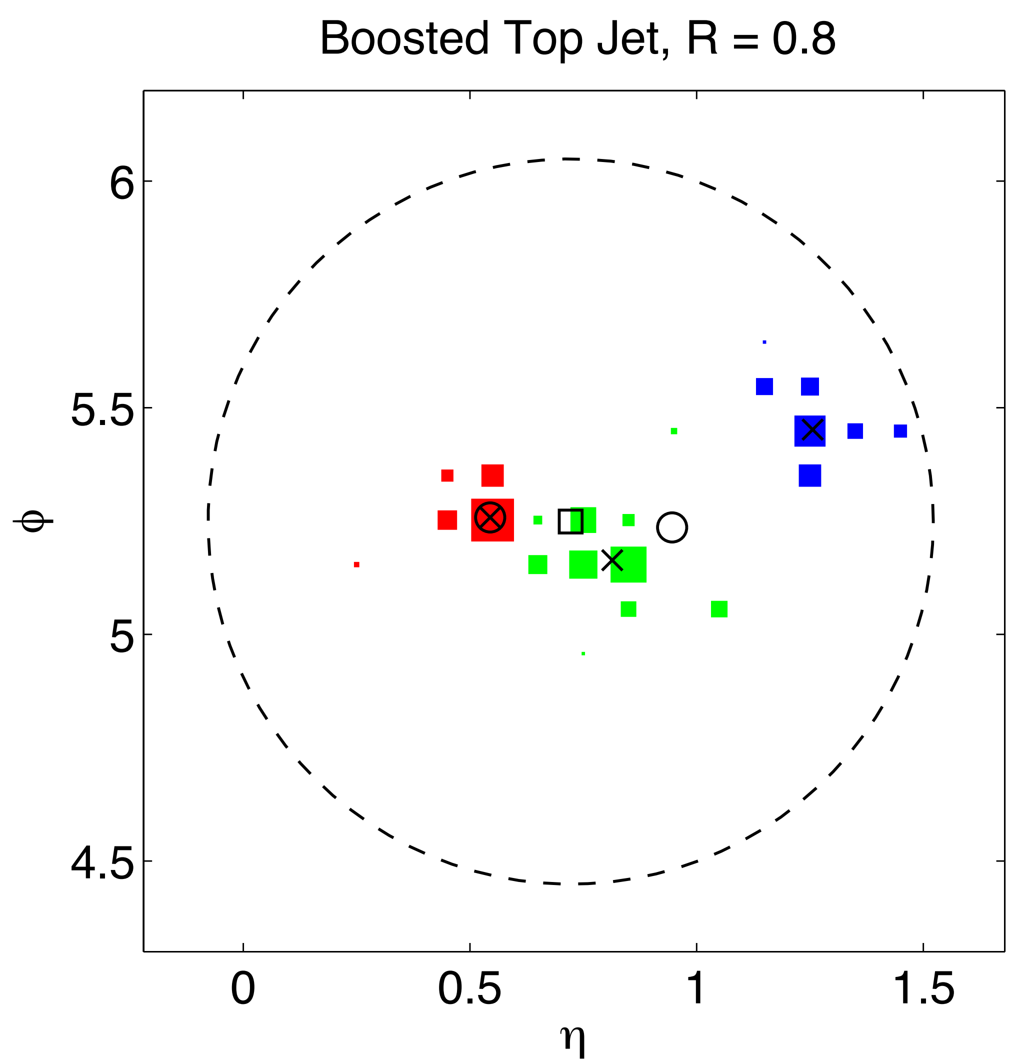
Standard Model

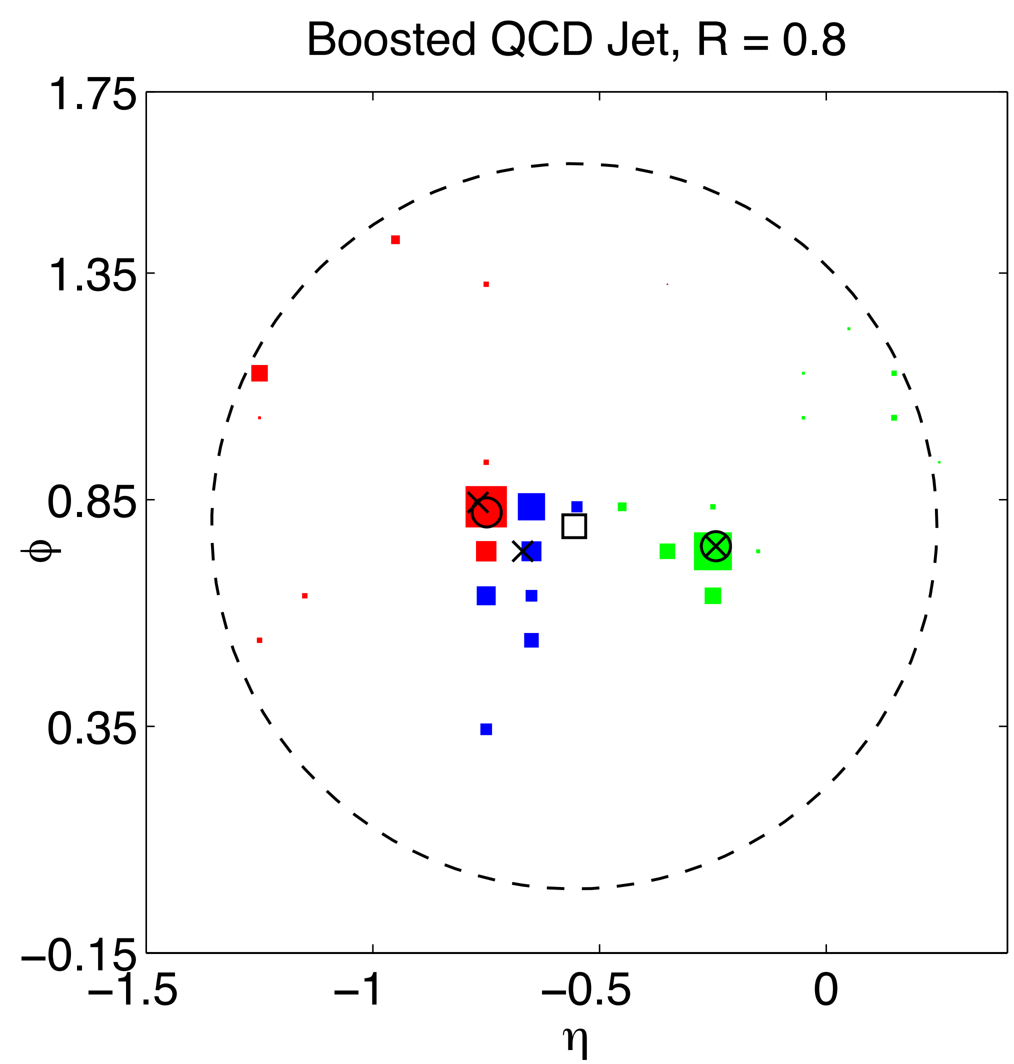
++ = Mass from QCD Radiation



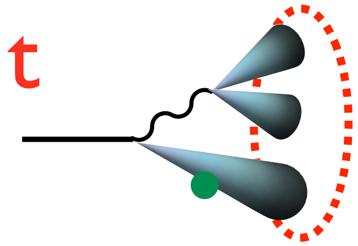
3-Prong vs. 1-Prong



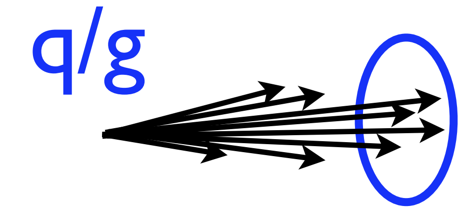


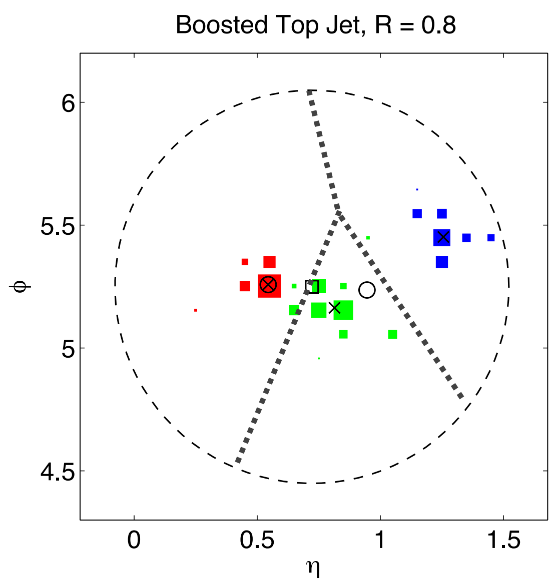


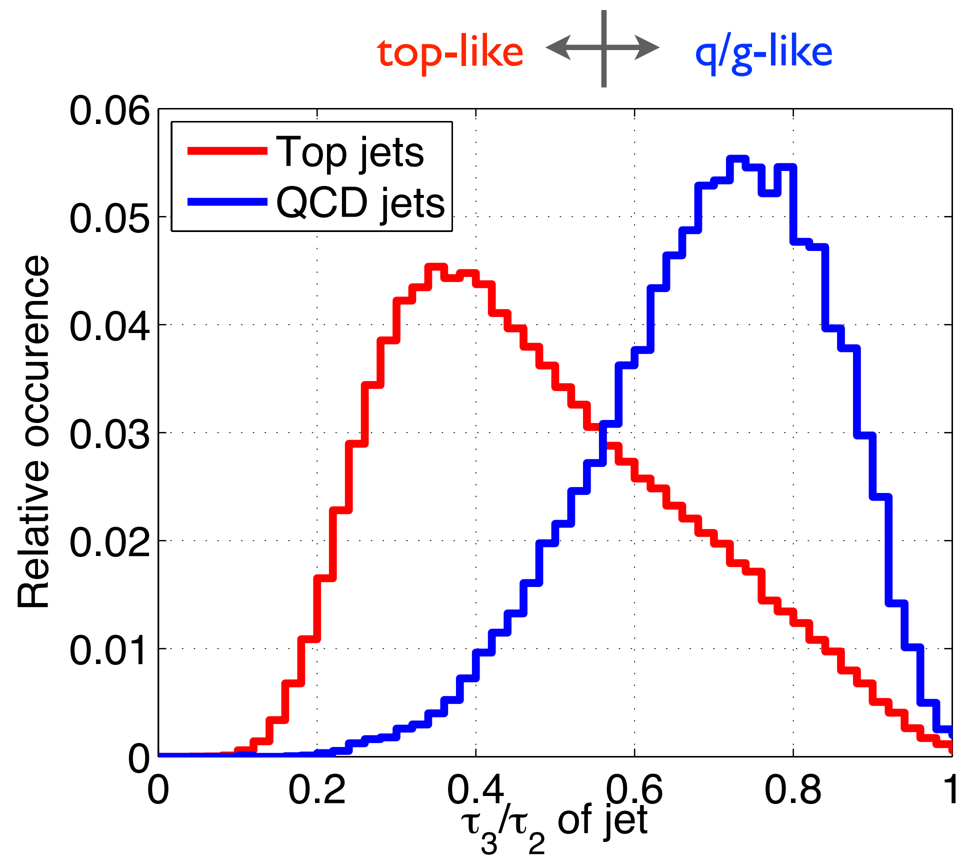
If your eyes can do it…

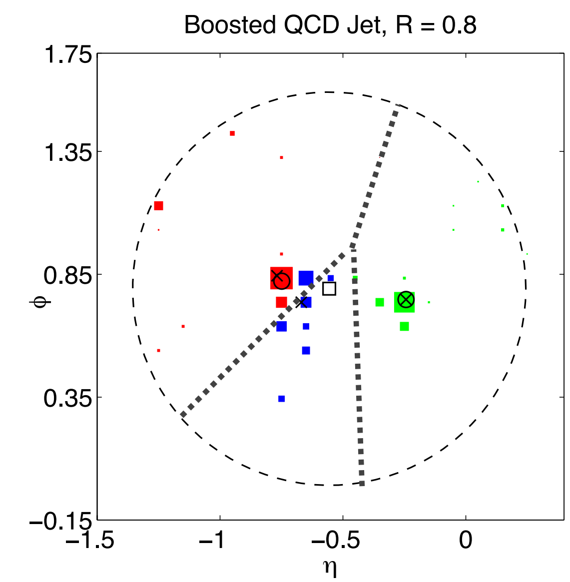


3-Prong vs. 1-Prong





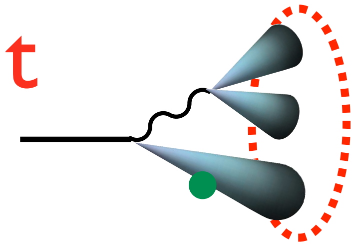




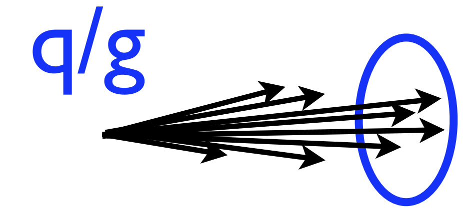
N-subjettiness

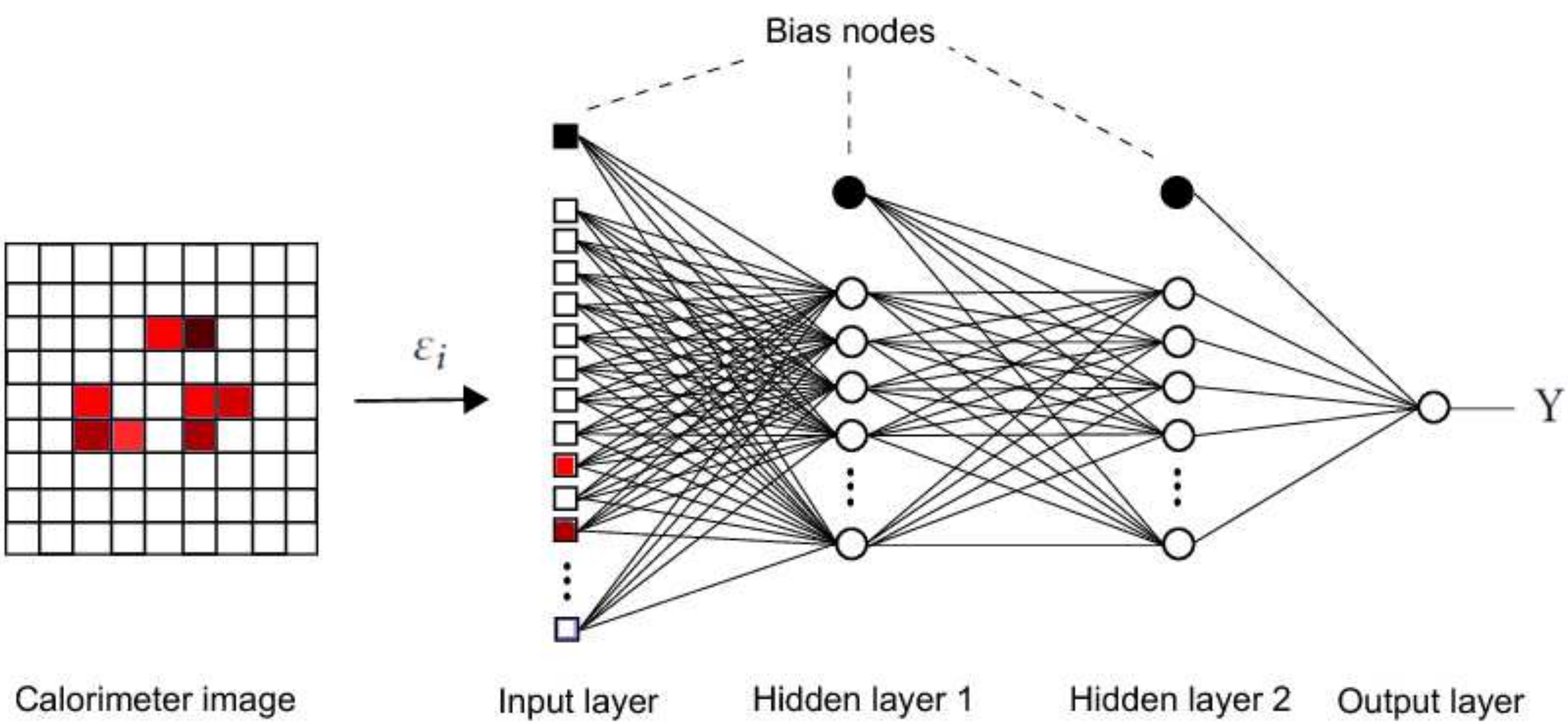
"Deep Thinking": $$\tau_N = \sum_k p_{T,k} \min\left\{\Delta R_{k,1}, \Delta R_{k,2}, \ldots, \Delta R_{k,N}\right\}$$

[JDT, Van Tilburg, 1011.2268, 1108.2701]



3-Prong vs. 1-Prong



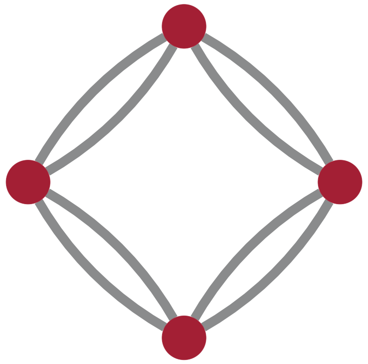


“Deep Learning”: BDTs, FLDs, DNNs, CNNs, RNNs, …

(More discussion in backup)

[Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968]

Jets at the LHC



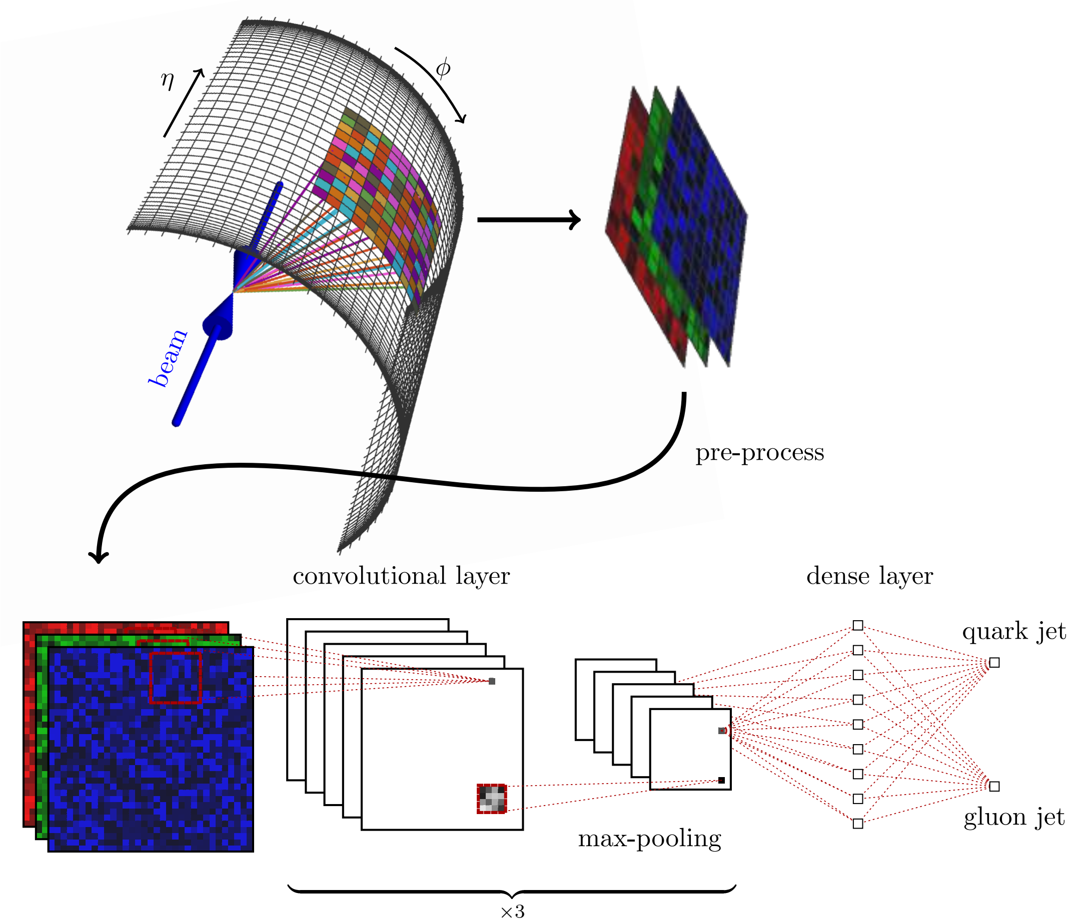
The Importance of Symmetries

Energy Flow Networks

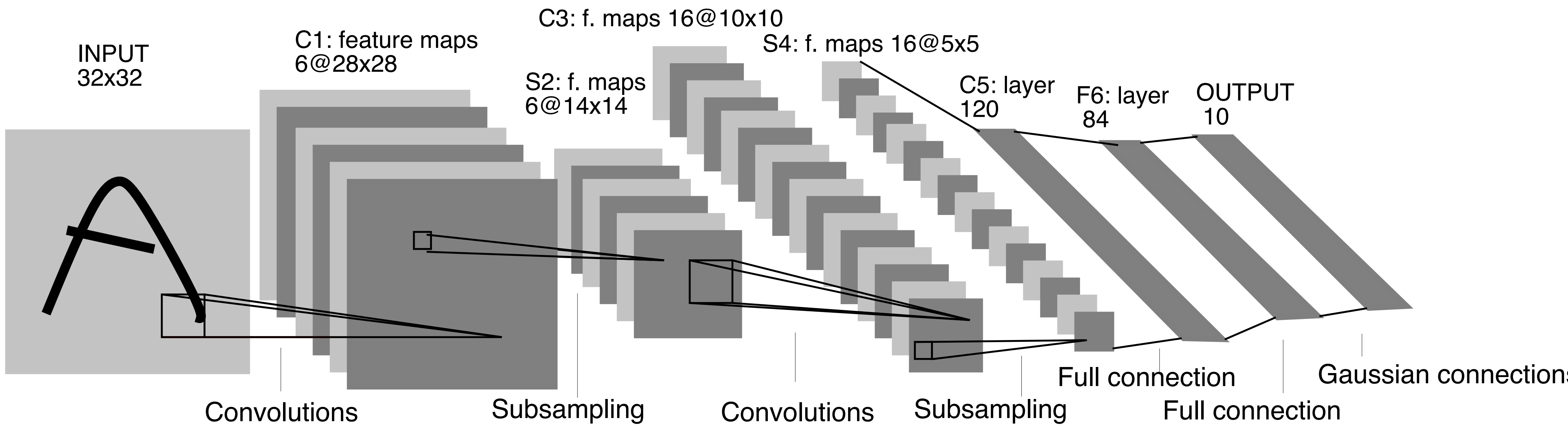
Patrick Komiske

Eric Metodiev

Two grad students walk into my office with their CNN…

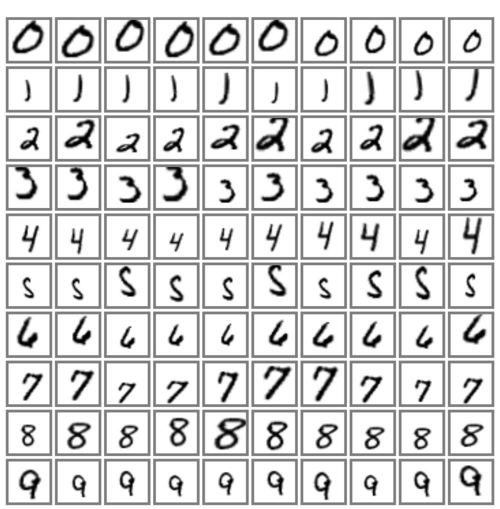


Symmetries of a CNN



Explores different length scales

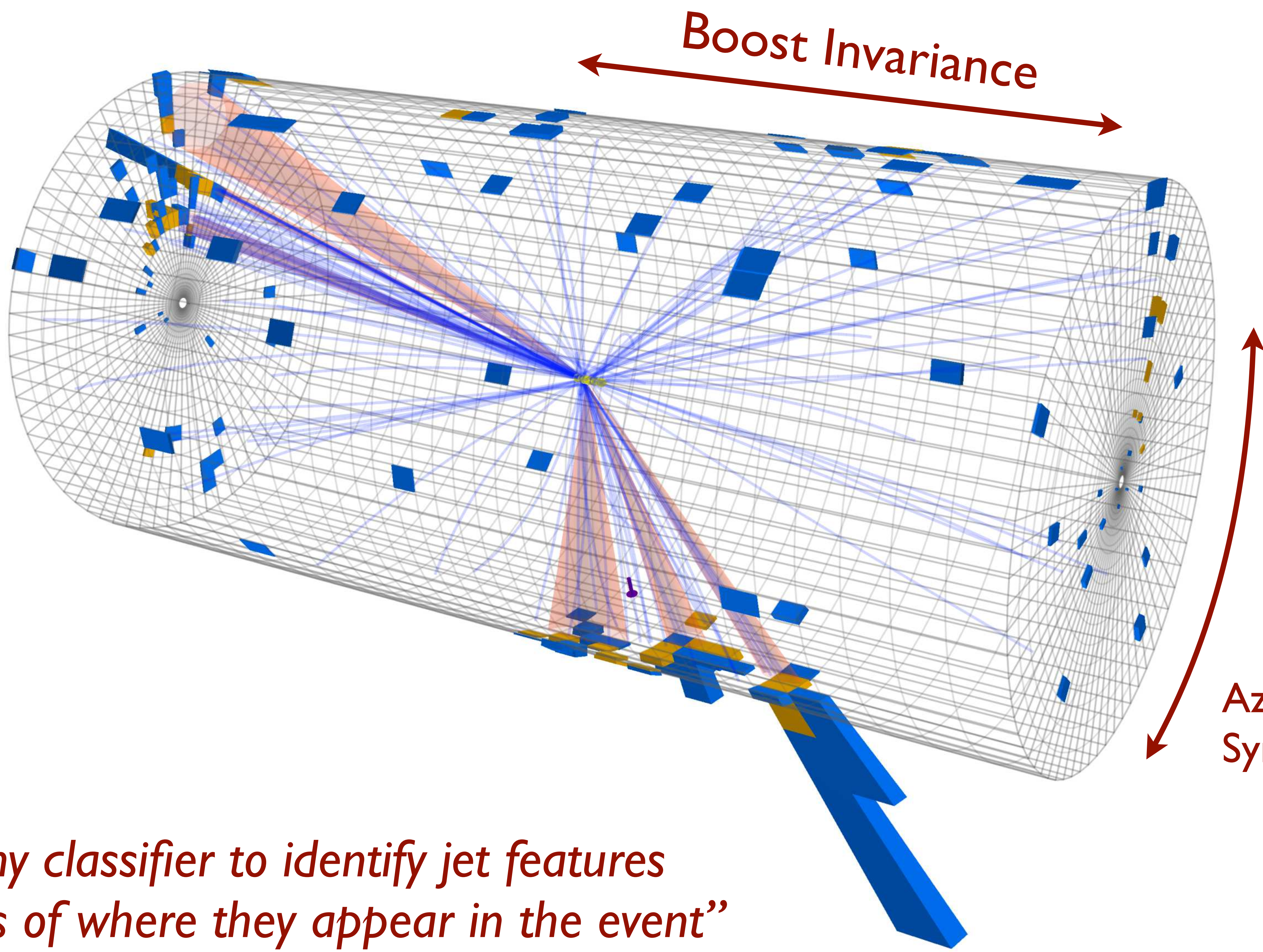
Enforces Translation Invariance



"I want my classifier to identify localized features regardless of where they appear in the image"

[image from LeCun, Bottou, Bengio, Haffner, 1998]

Symmetries of LHC Events



"I want my classifier to identify jet features regardless of where they appear in the event"

[image from CMS, 2015]

The Physics-First Approach

Underlying Physics

$\Downarrow$ "Deep Thinking"

Natural Data Representation

$\Downarrow$ "Deep Learning"

Suitable Algorithm

The Buzzword-First Approach

Questionable Physics

$\Uparrow$ "Wishful Thinking"

Unnatural Data Representation

$\Uparrow$ "Shoehorning"

Cool-Sounding Algorithm

Why CNNs Aren't Ideal

Underlying Physics
Natural Data Representation
Suitable Algorithm

Translations/Boost Invariance

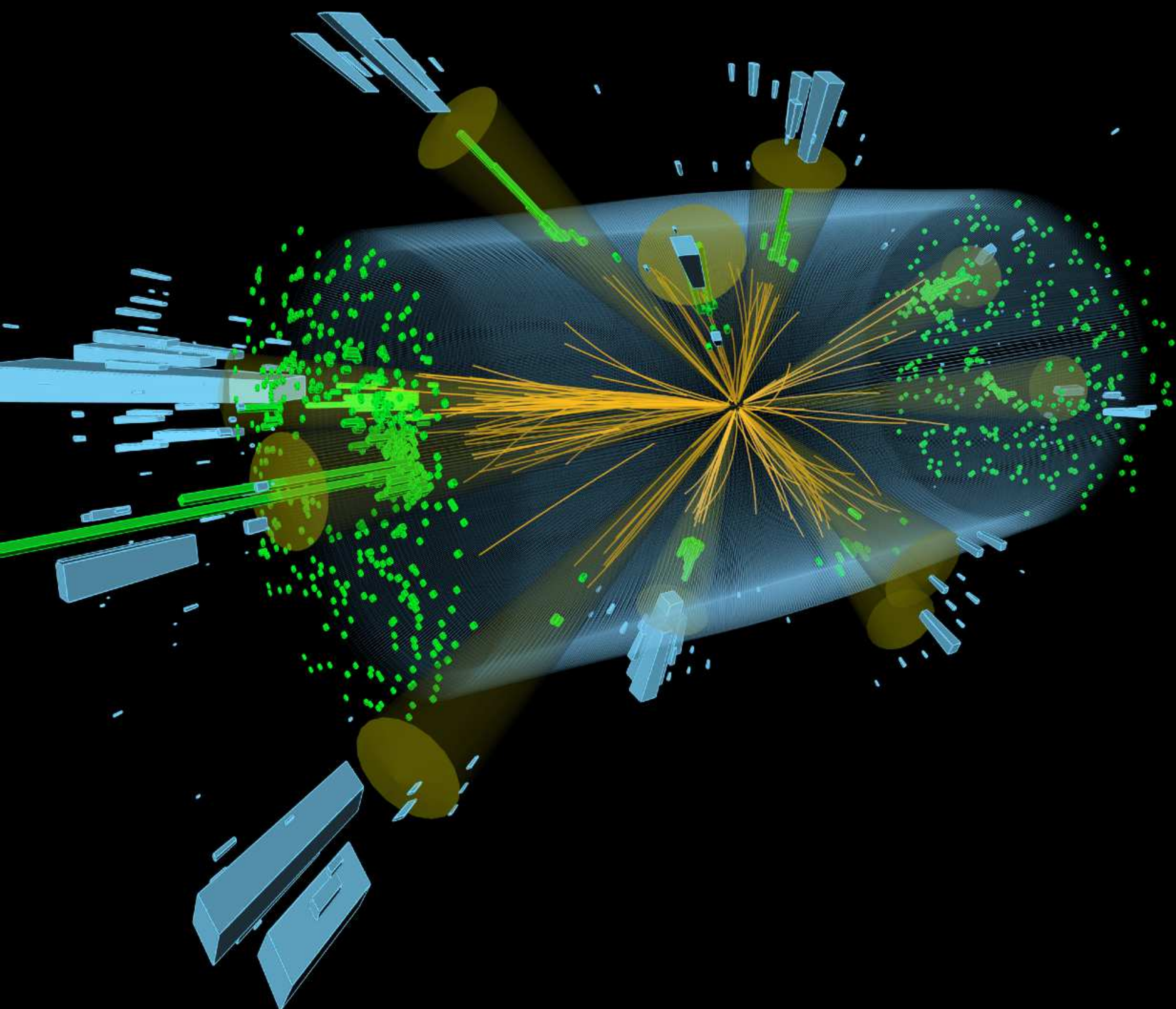
$\Downarrow$ "Deep Thinking"

Pixelated Image

$\Uparrow$ "Shoehorning"

Convolution Neural Networks

Debris Taxonomy

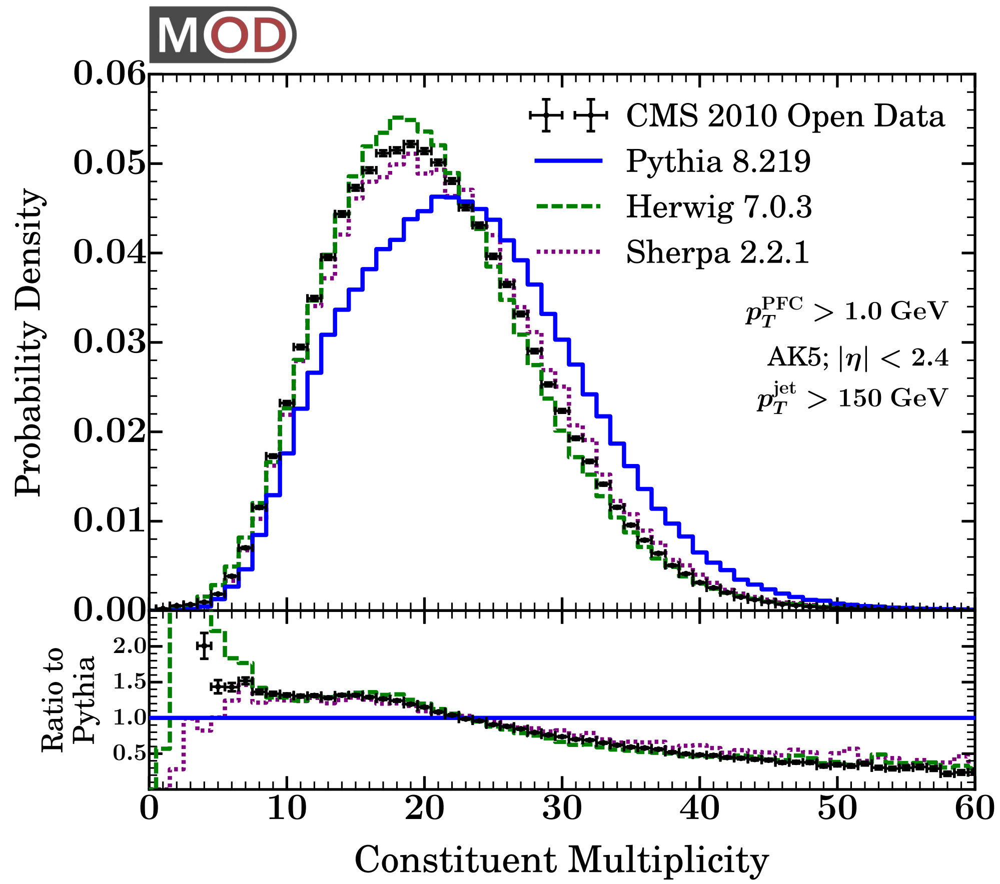


T	E	H	M			
	●			γ	photon	elementary
●	●			$e^{\pm}$	electron	elementary
●	•	•	●	$\mu^{\pm}$	muon	elementary
●	•	●		$\pi^{\pm}$	pion	composite
●	•	●		$K^{\pm}$	kaon	composite
	•	●		K_L^0	K-long	composite
●	•	●		$p/\overline{p}$	proton	composite
	•	●		$n/\overline{n}$	neutron	composite

Point Clouds

[Popular Science, 2013]

Key Fact



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

Jet constituents:

Variable-length

Unordered set

Each particle:

$\{E, p_x, p_y, p_z\}$

Optional labels

Key Jargon

"Infrared/Collinear Safe" ≈ Weighted by Energy

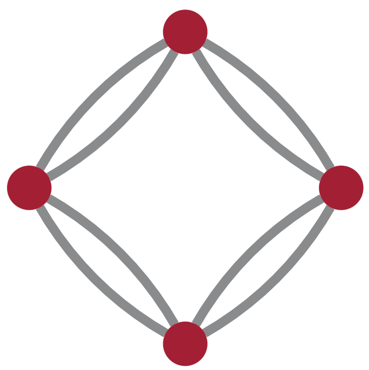
Our Understanding c. 2017

Underlying Physics
Natural Data Representation
Suitable Algorithm

Infrared/Collinear Safety

$\Downarrow$

Energy Flow Polynomials



$\Downarrow$

Linear Regression

[Komiske, Metodiev, JDT, 1712.07124;
https://energyflow.network]

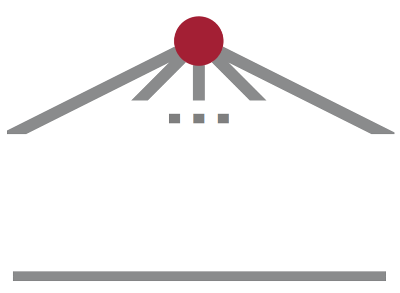
Our Understanding c. 2019

Underlying Physics
Natural Data Representation
Suitable Algorithm

Infrared/Collinear Safety

$\Downarrow$

Energy Flow Moments



$\Uparrow$

Linear Regression + Linear Runtime

[Komiske, Metodiev, JDT, we've been promising this paper for 9 months]

Today's Talk (2018)

Underlying Physics
Natural Data Representation
Suitable Algorithm

Quantum-Mechanical Indistinguishability

$\Downarrow$

Variable-Length Unordered Sets of Particles

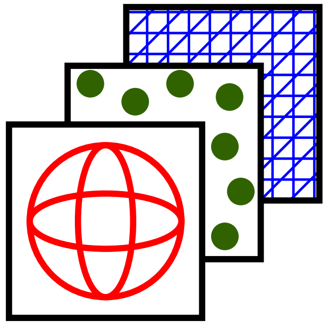
$\Downarrow$

???

[Komiske, Metodiev, JDT, 1810.05165]

Jets at the LHC

The Importance of Symmetries



Energy Flow Networks

The Power of Addition

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

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

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

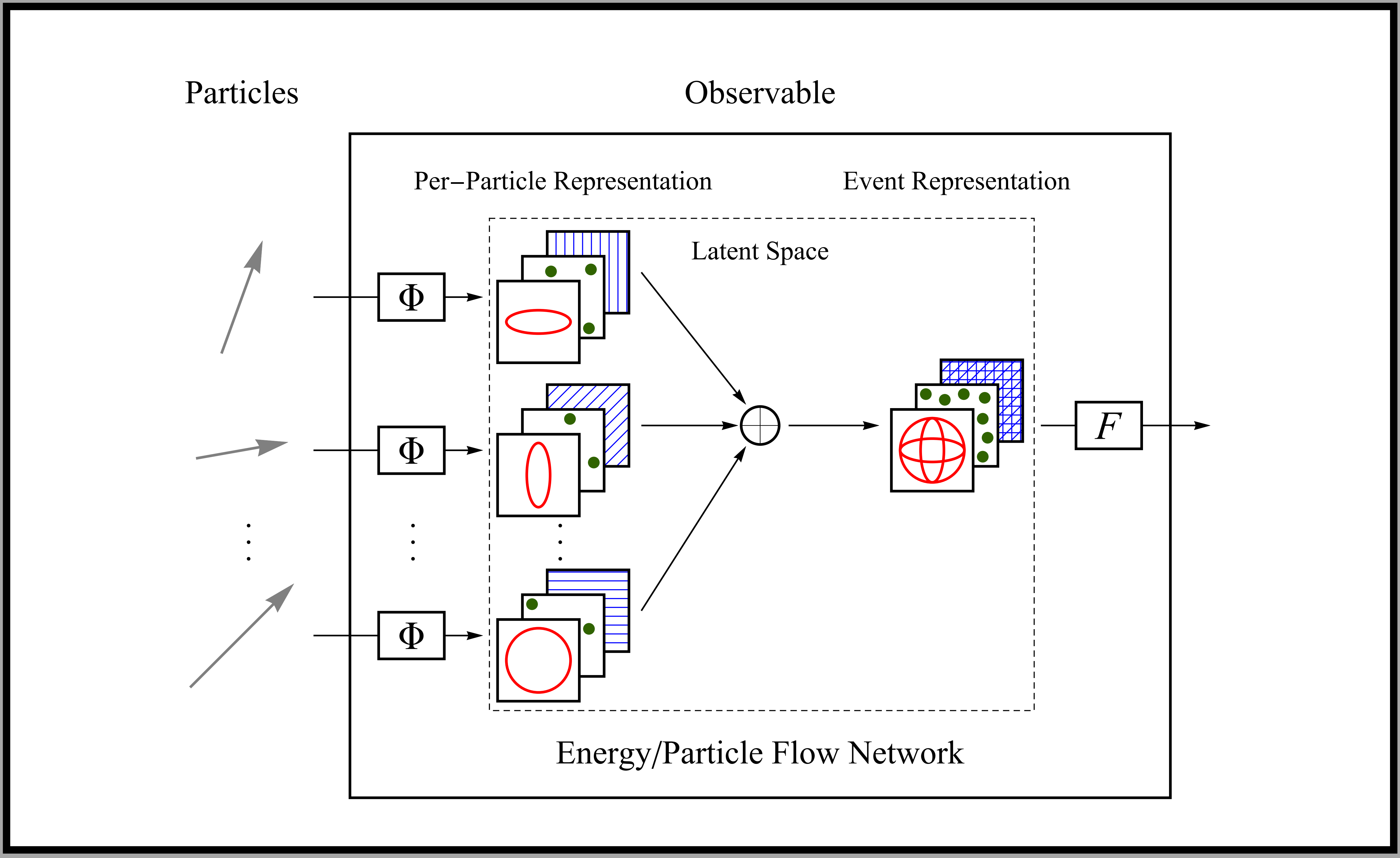
Energy / Particle Flow Networks

IRC-safe Φ

General Φ

[Komiske, Metodiev, JDT, 1810.05165]

Conjectured Generalization

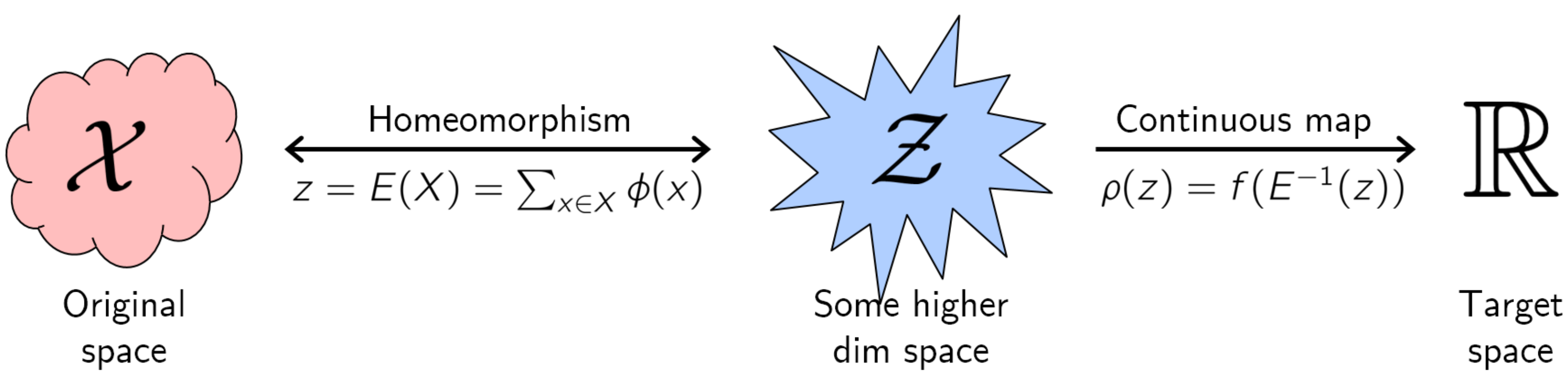


[Komiske, Metodiev, JDT, 1810.05165]

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

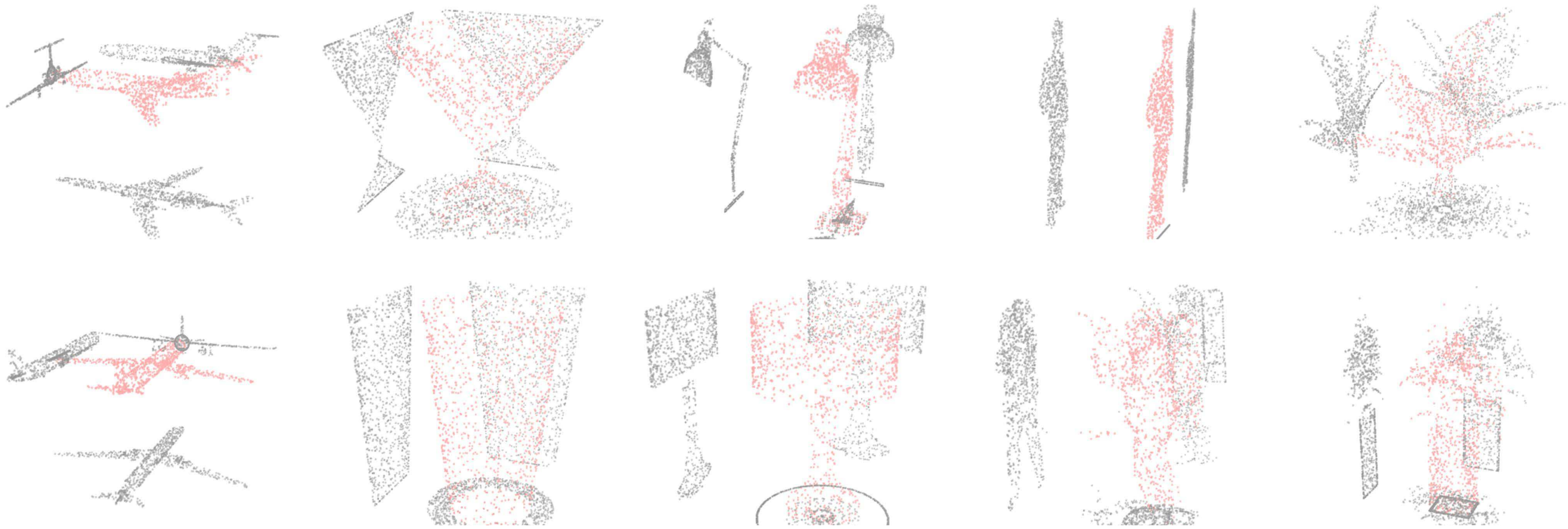
Generic Observable

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

Deep Sets for…

Celebrity Face Anomaly Detection

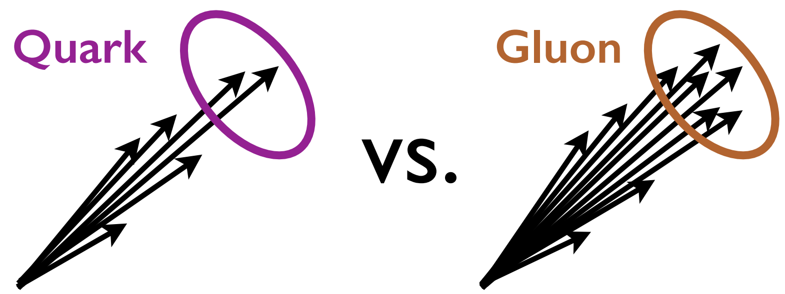
Point Cloud Classification



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

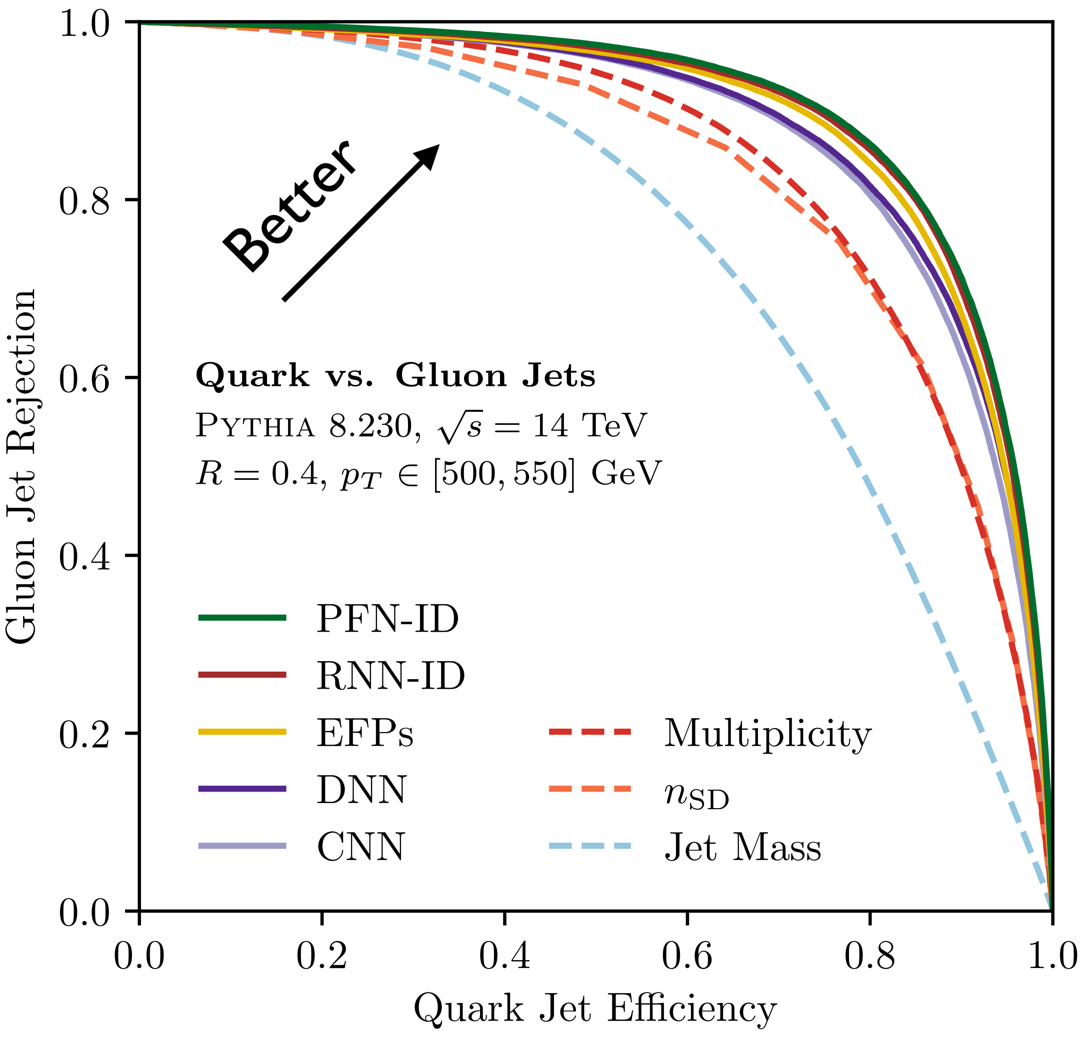
Deep Sets for Quark/Gluon Discrimination

The "Hello, World!" of jet classification



As performance saturates, next frontiers are symmetries & visualization

Competitive with previous methods

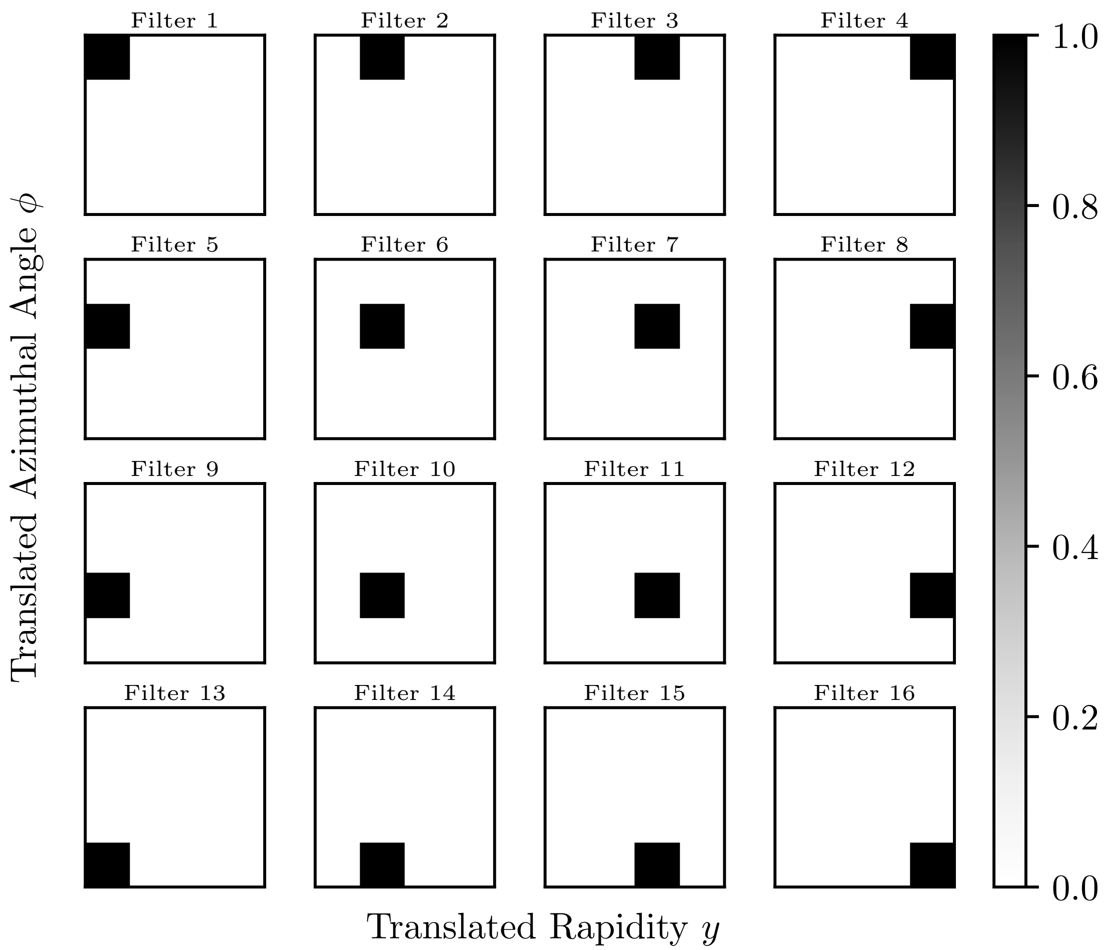


Latent Space Visualization

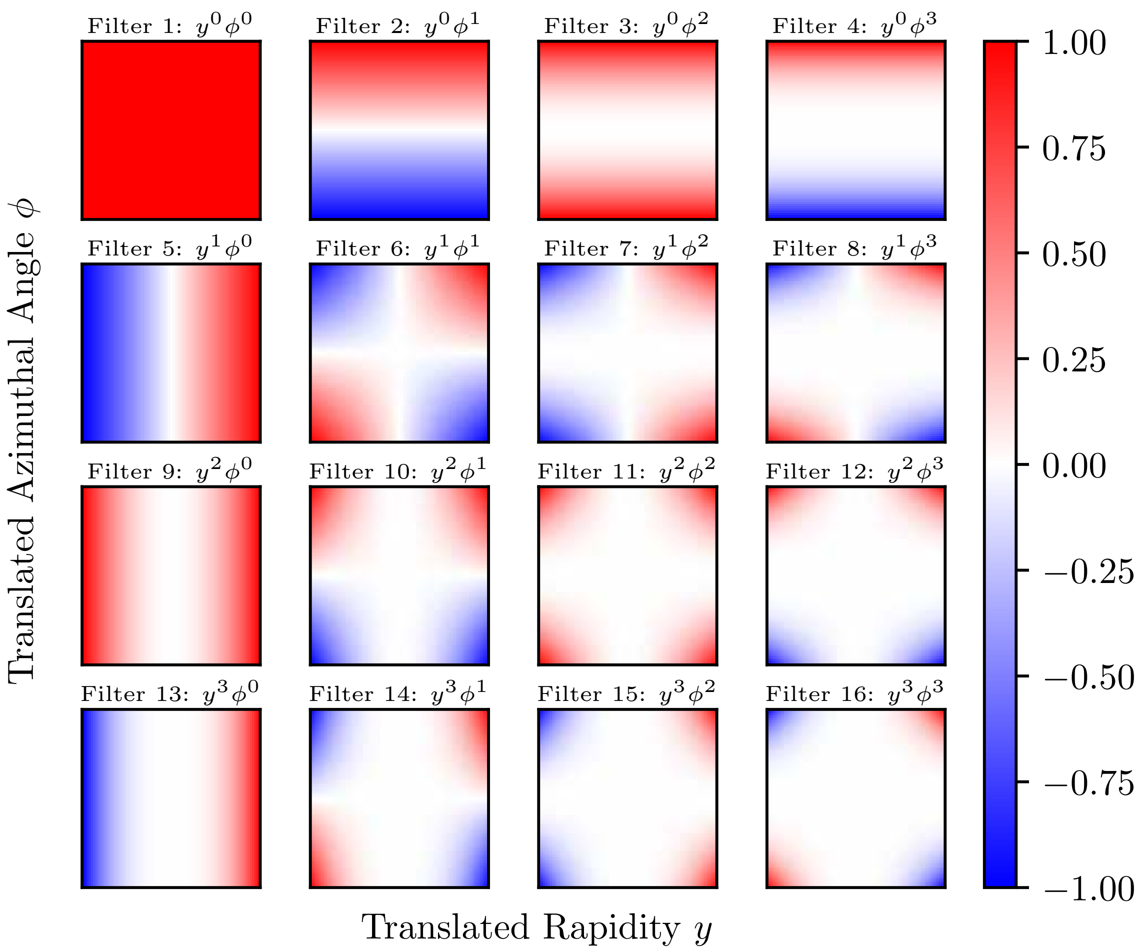
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

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

Calorimeter Pixels



Radiation Moments

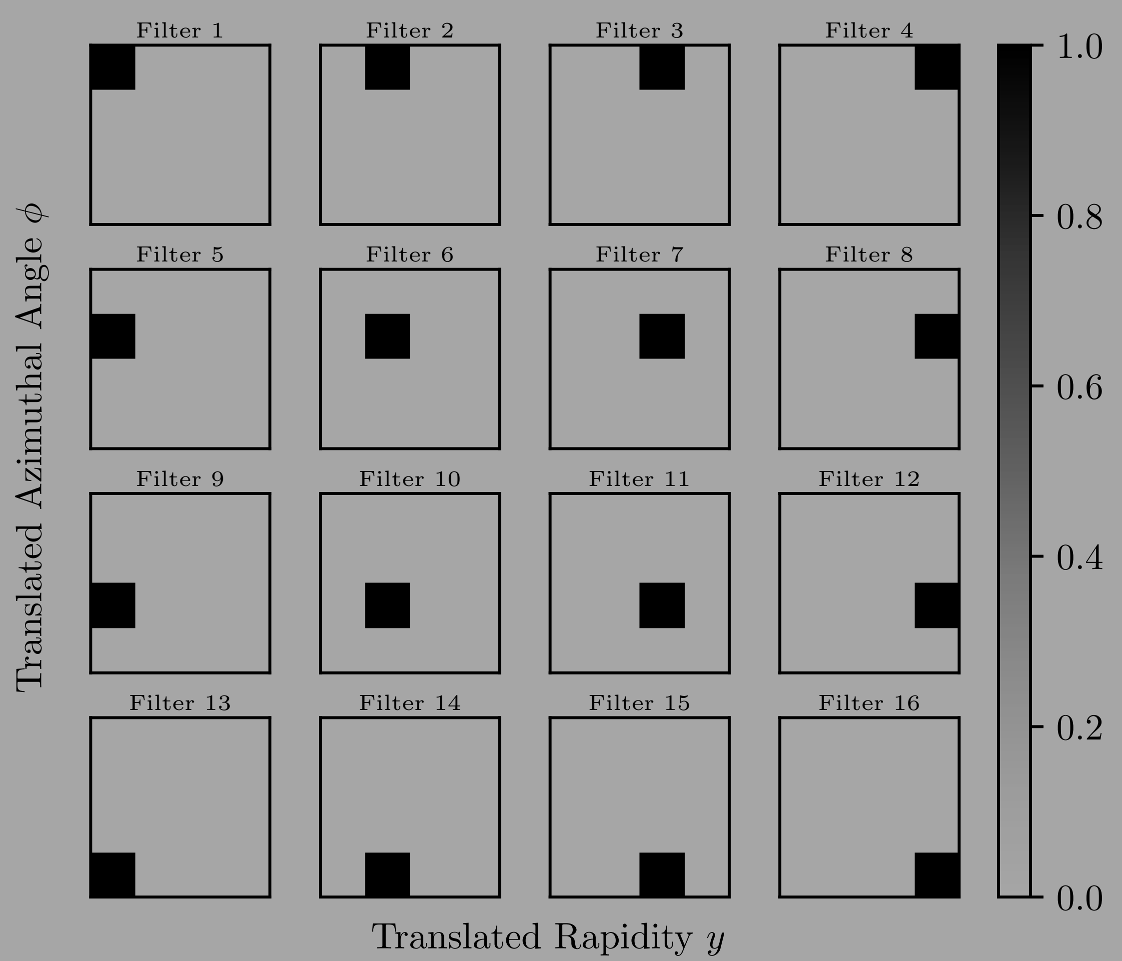


Latent Space Visualization

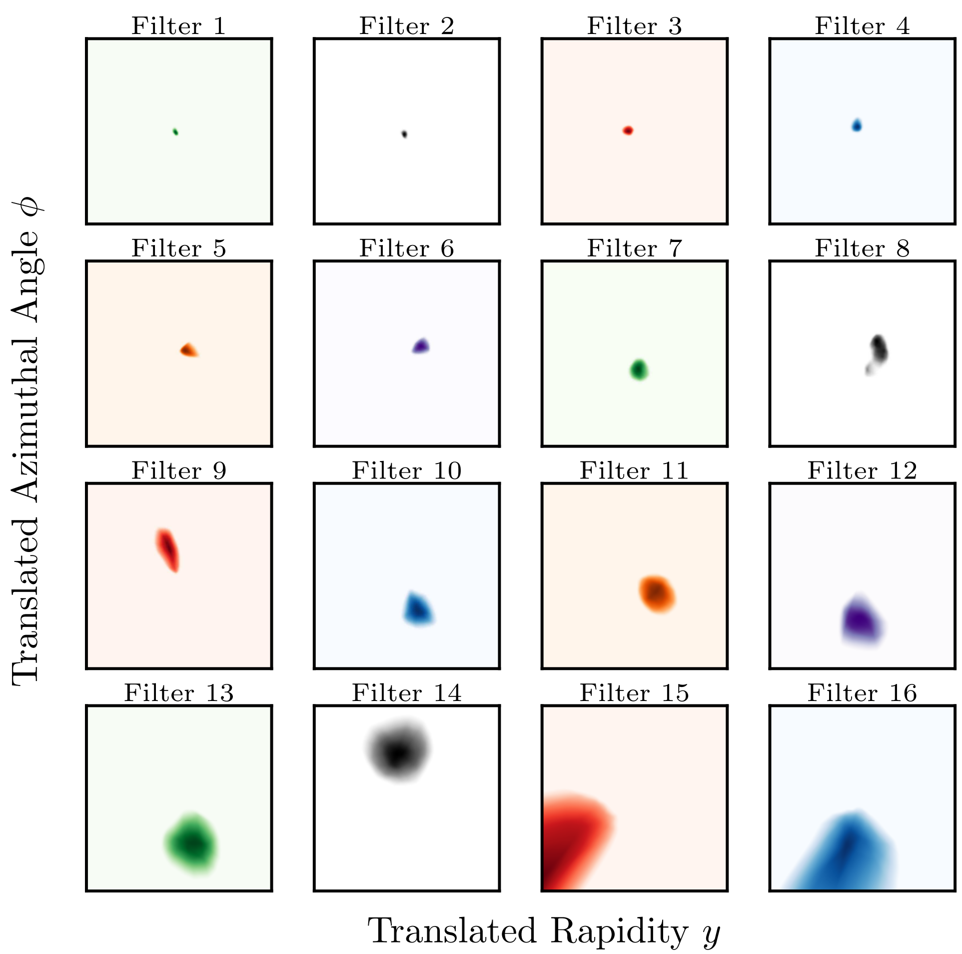
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

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

Calorimeter Pixels

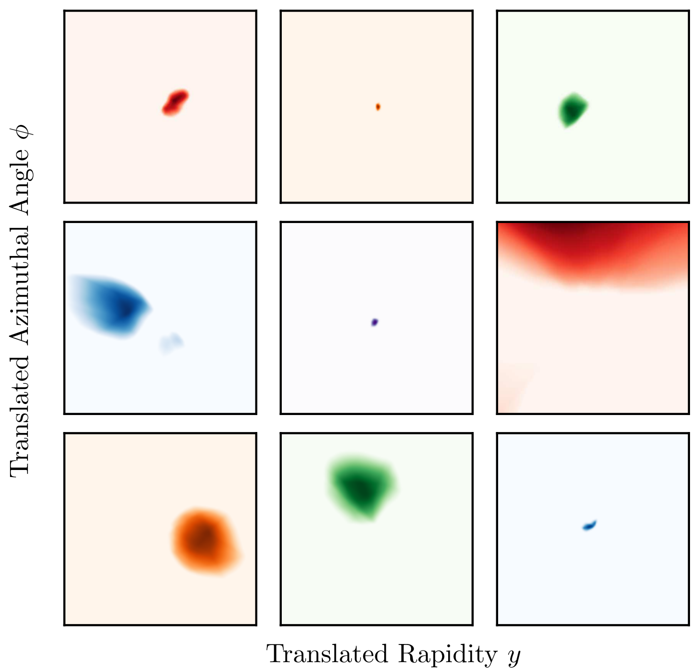


EFNs: Dynamic Pixelation

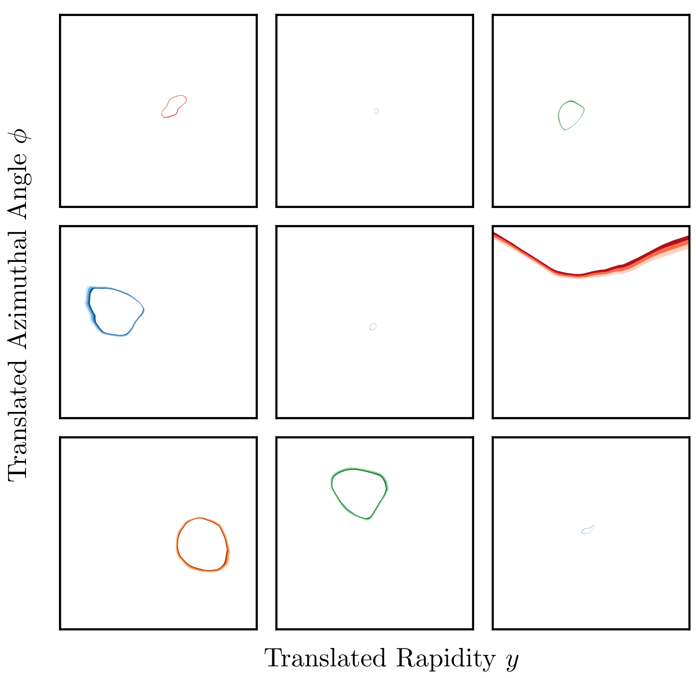


Psychedelic Visualization

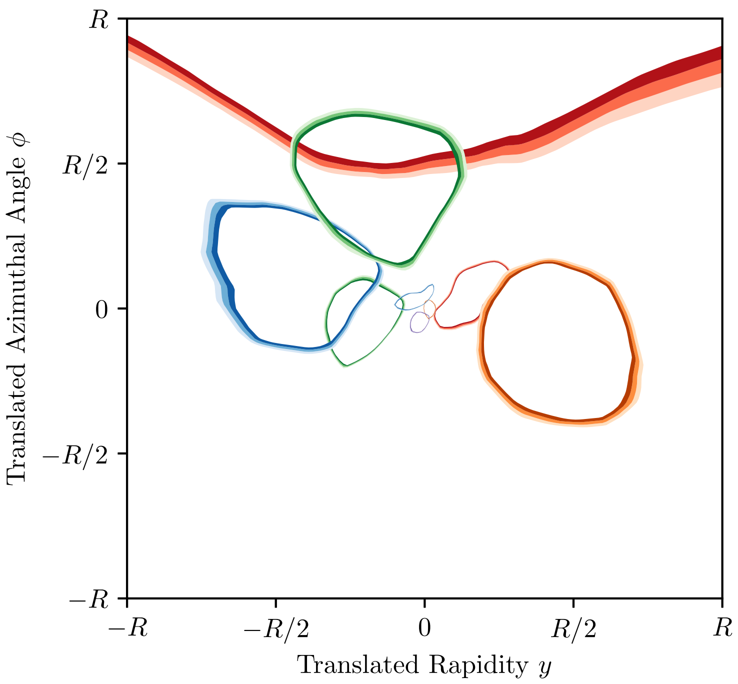
Latent Filters



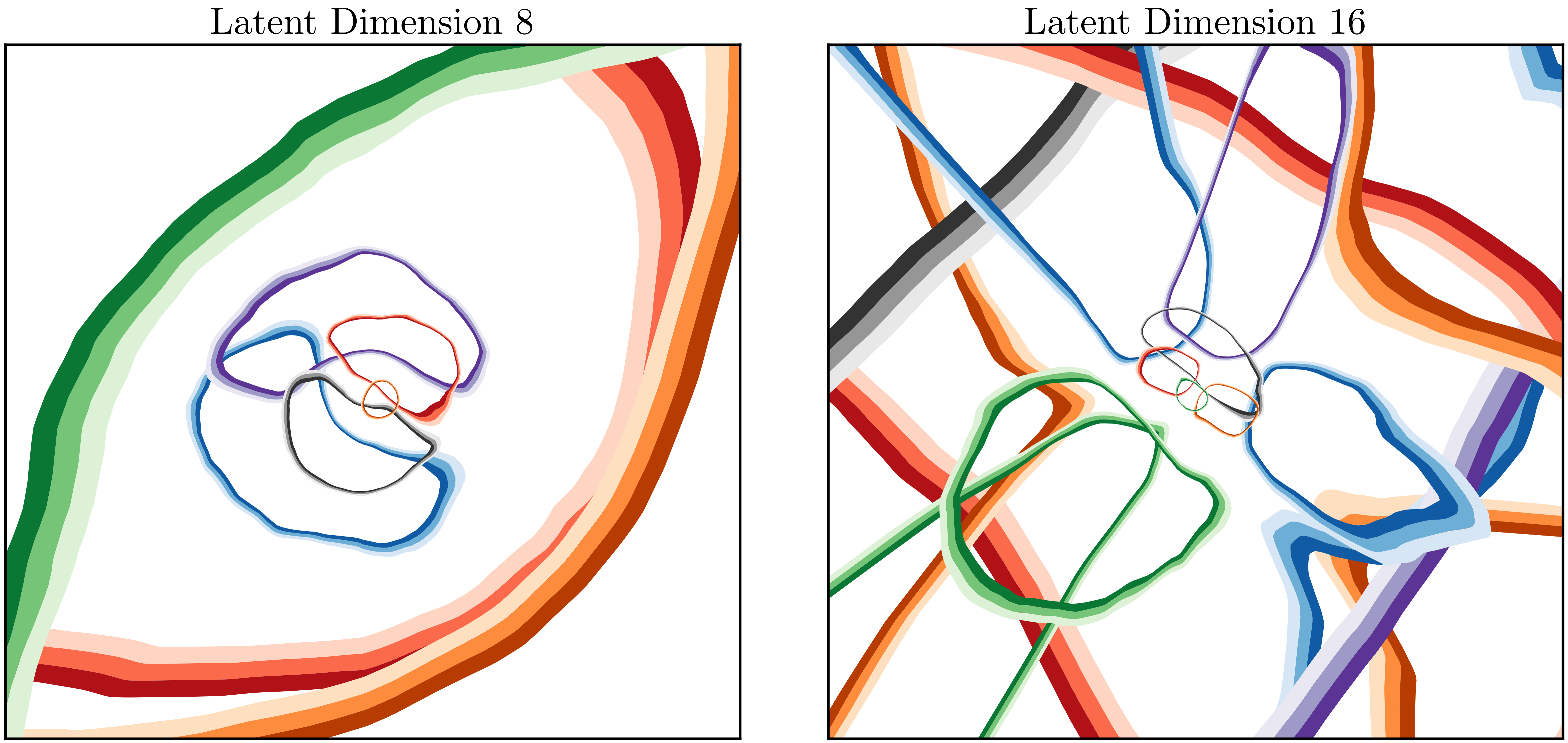
50% Contours



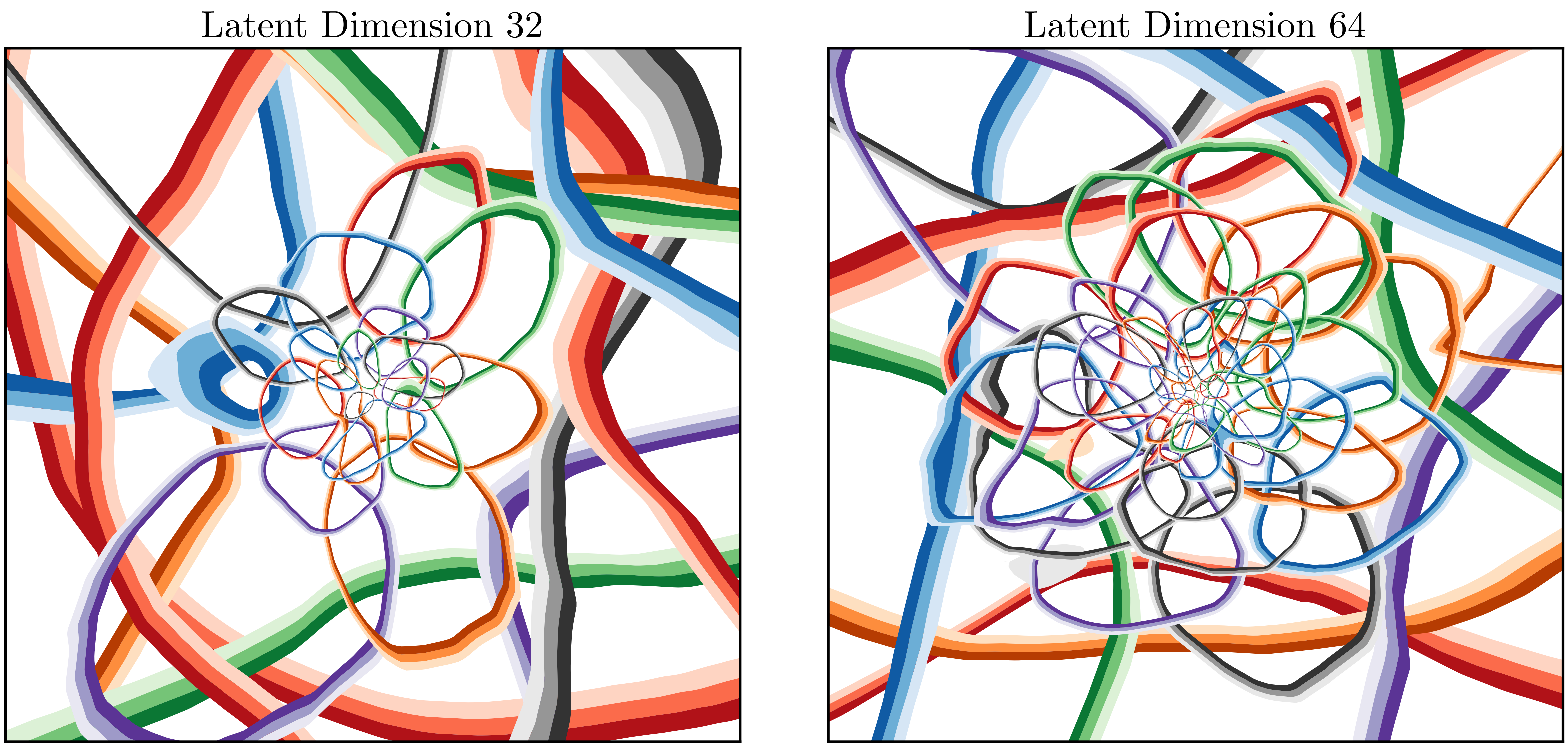
Overlay



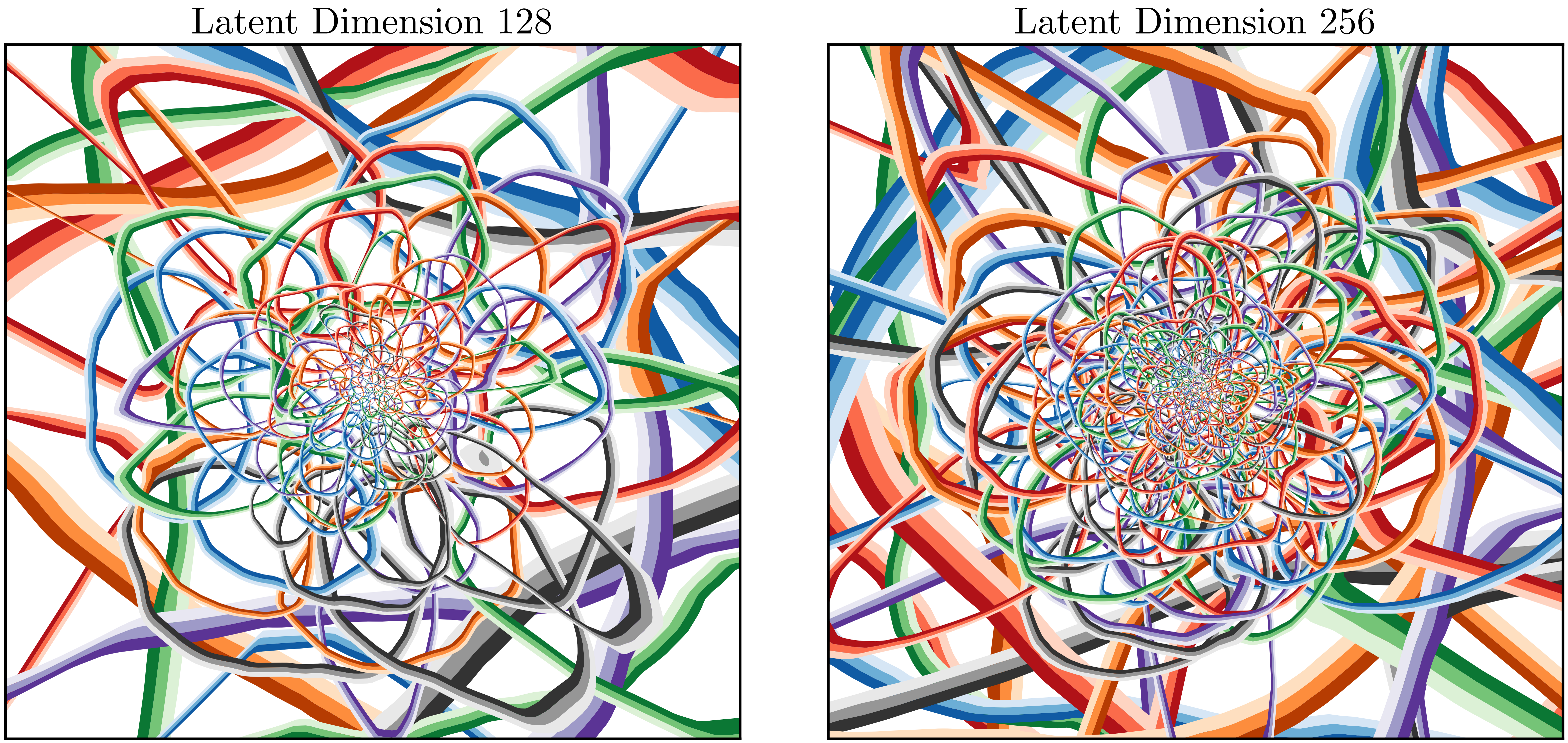
Psychedelic Visualization



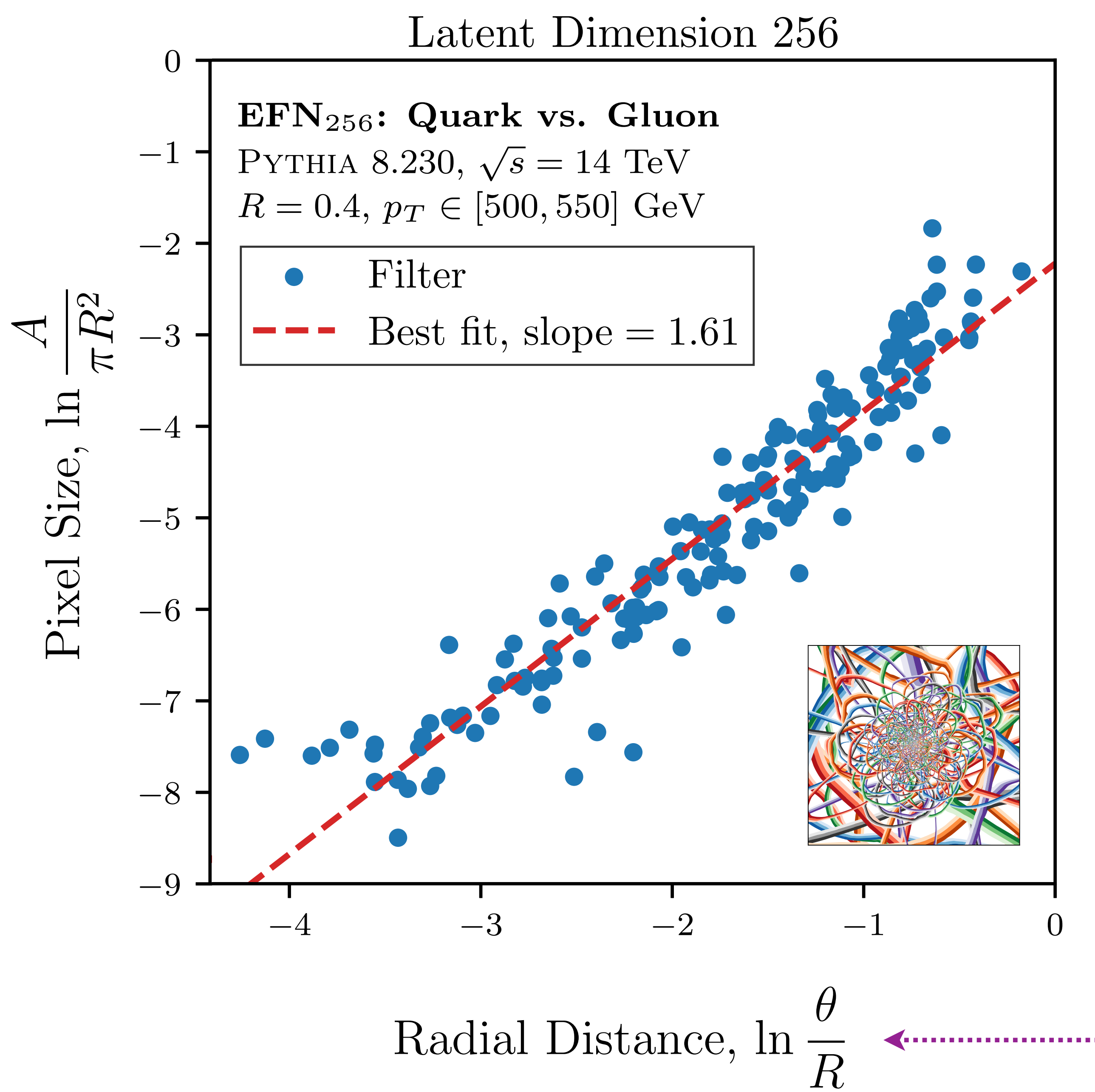
Psychedelic Visualization

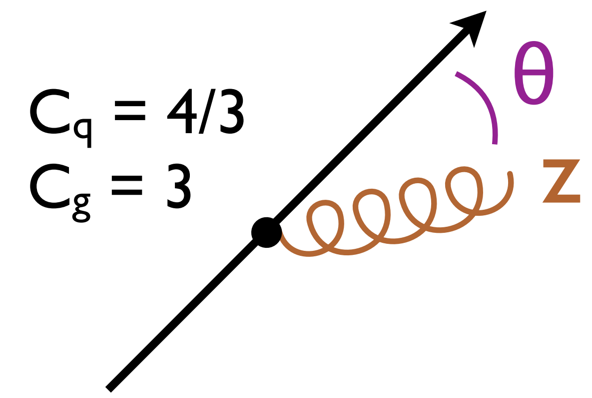


Psychedelic Visualization



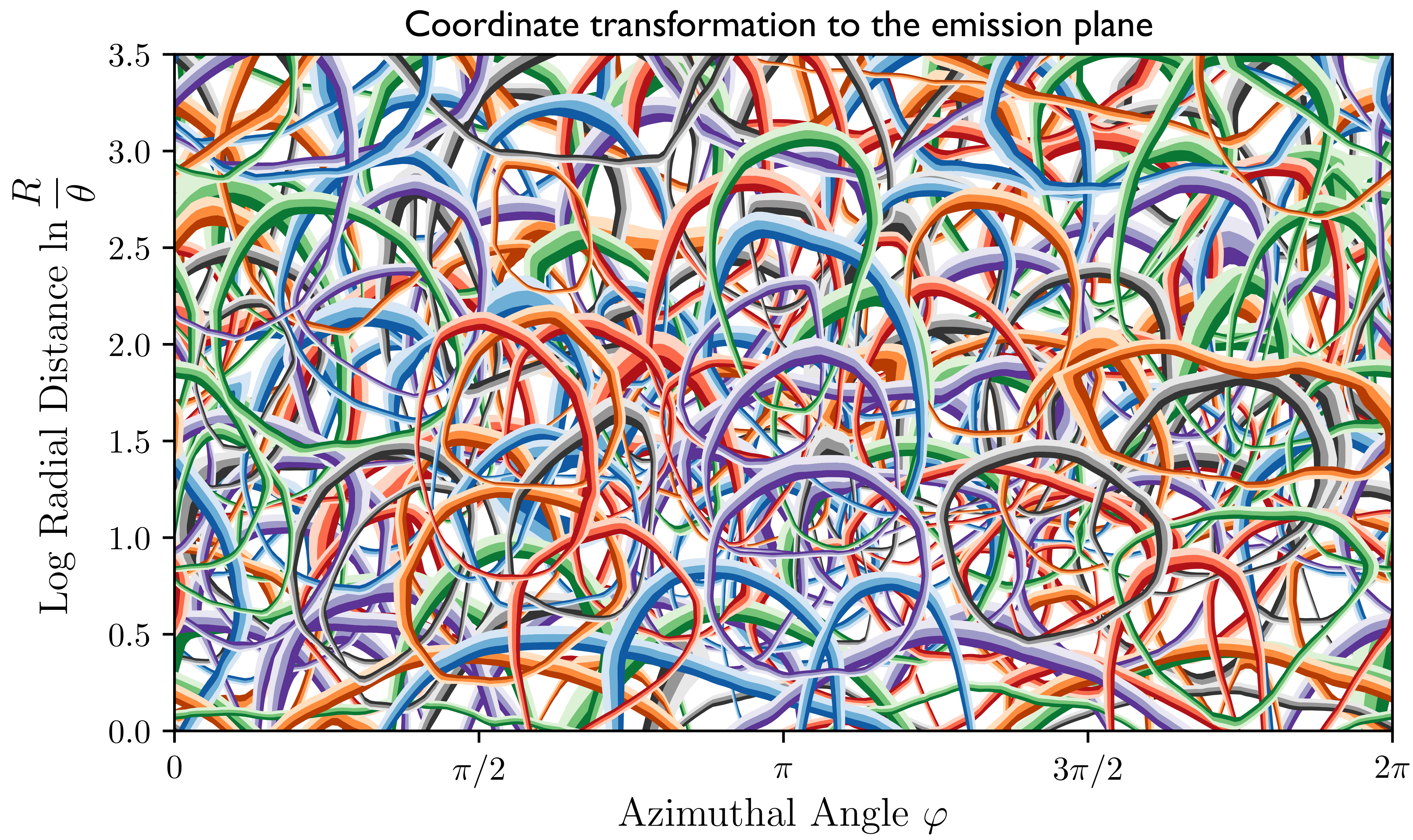
Learning the Singularity Structure of QCD





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

Ready for the MoMA



The Broader Lesson

“Deep Learning” ~~vs.~~ & “Deep Thinking”

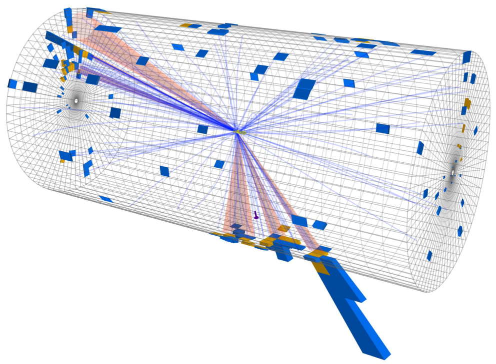
Advances in mathematics and computer science (Deep Sets)

$\Downarrow \quad \Uparrow$

Advances in collider physics (EFN/PFN)

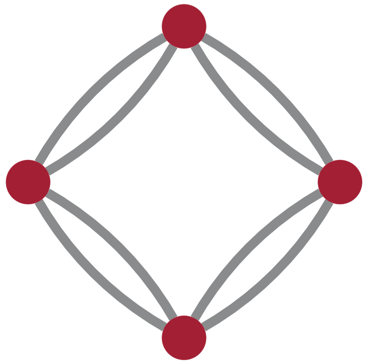
Weighted Point Sets $\Leftrightarrow$ *IRC Safety*

Summary



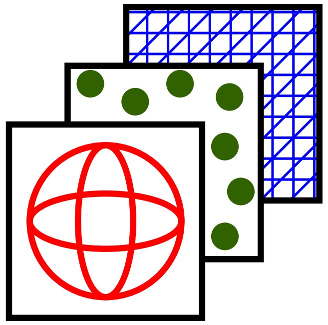
Jets at the LHC

The rise of machine learning for collider physics



The Importance of Symmetries

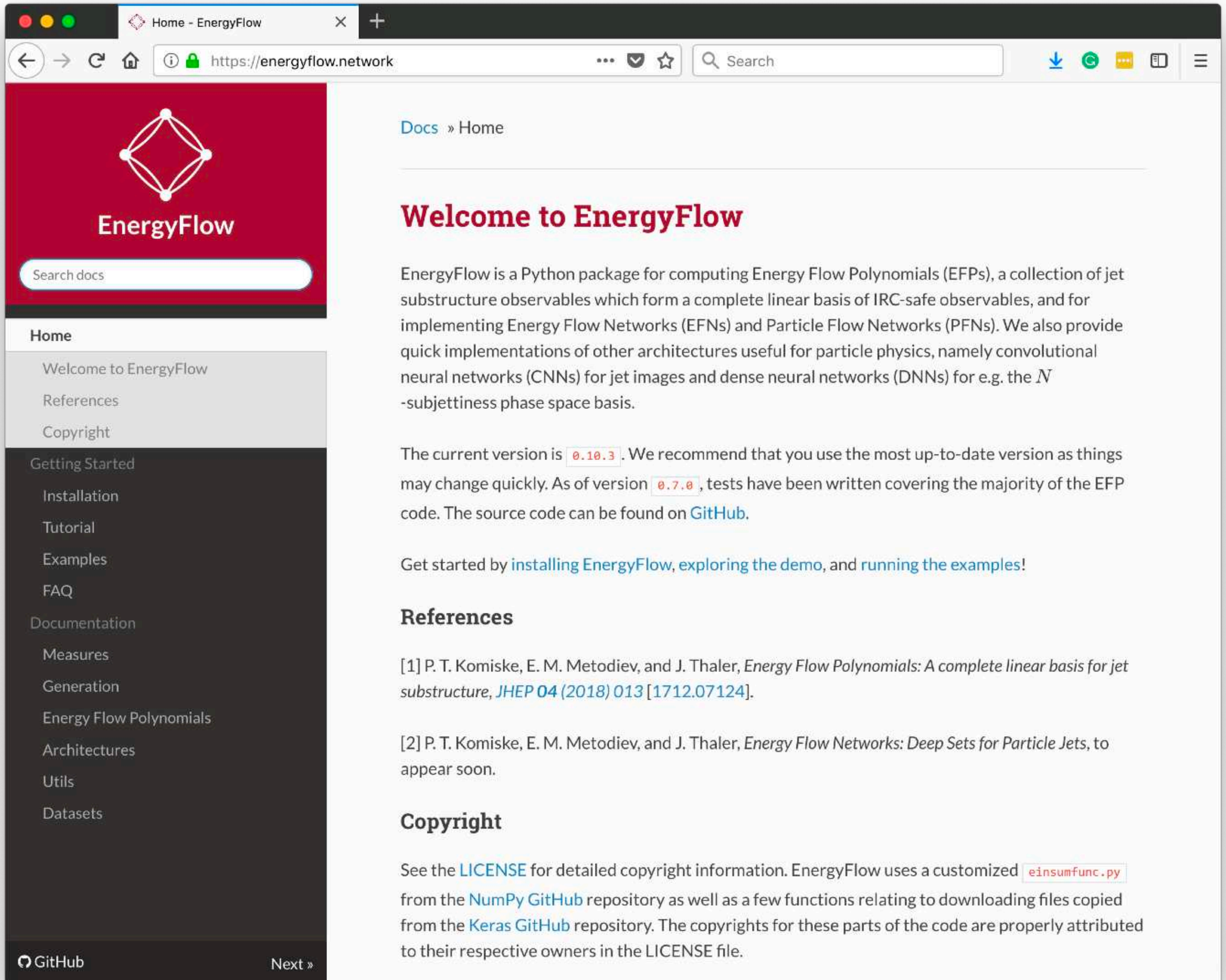
Essential tools to "think like a physicist"



Energy Flow Networks

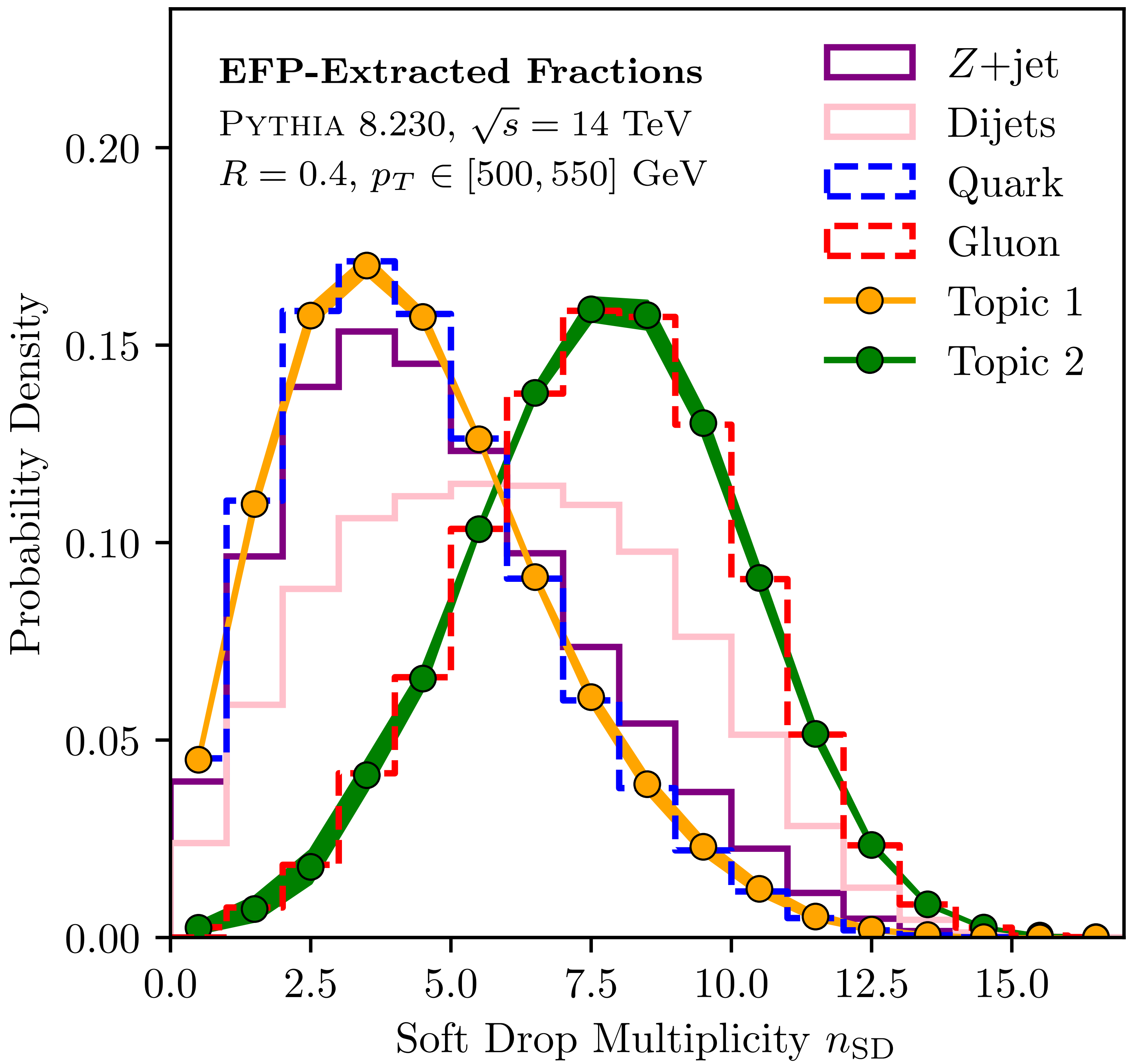
A bridge between collider physics and machine learning

energyflow.network



Backup Slides

Physics Reasons for My Conversion

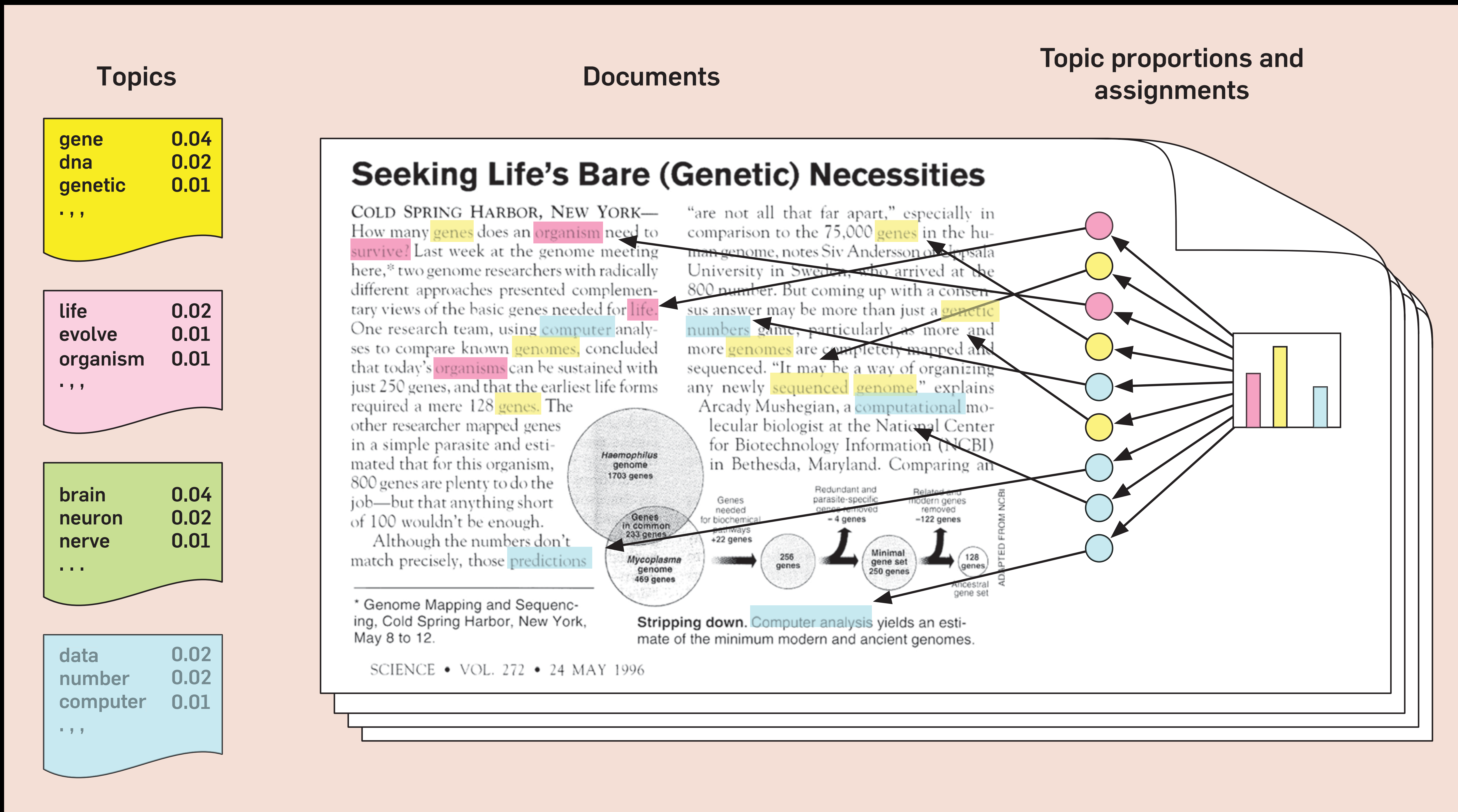


For Offline Discussions

Collider physics meets blind source separation with same underlying structure as…

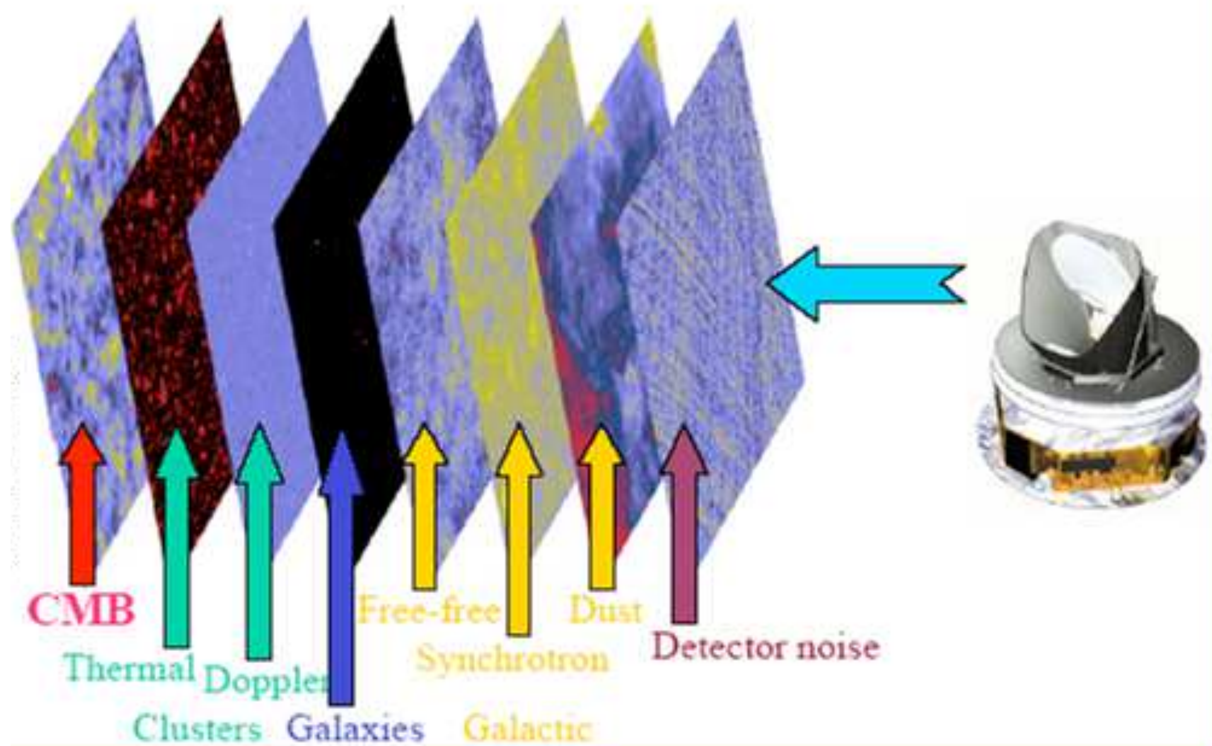
[Komiske, Metodiev, JDT, 1809.01140;
see also Metodiev, JDT, 1802.00008; Metodiev, Nachman, JDT, 1708.02949]

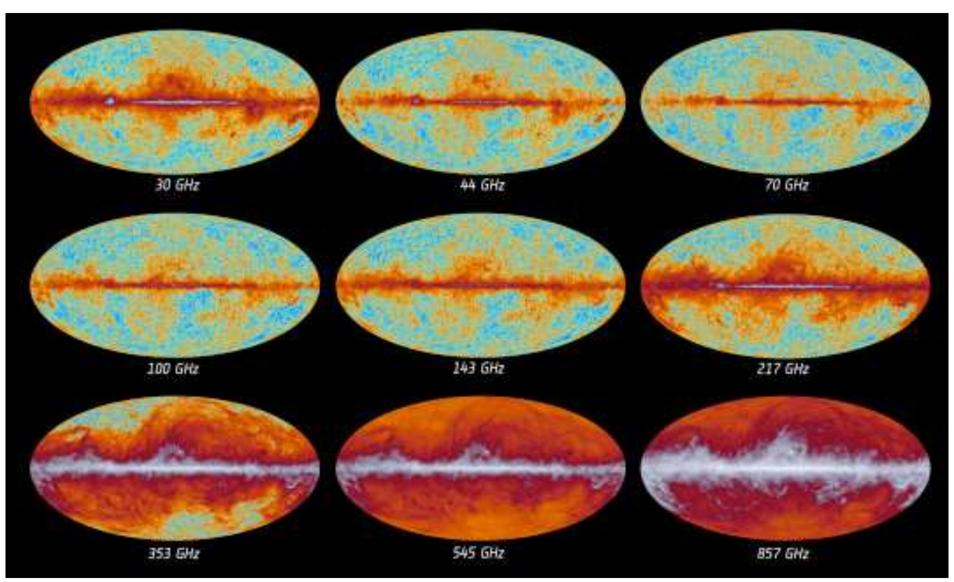
Topic Modeling



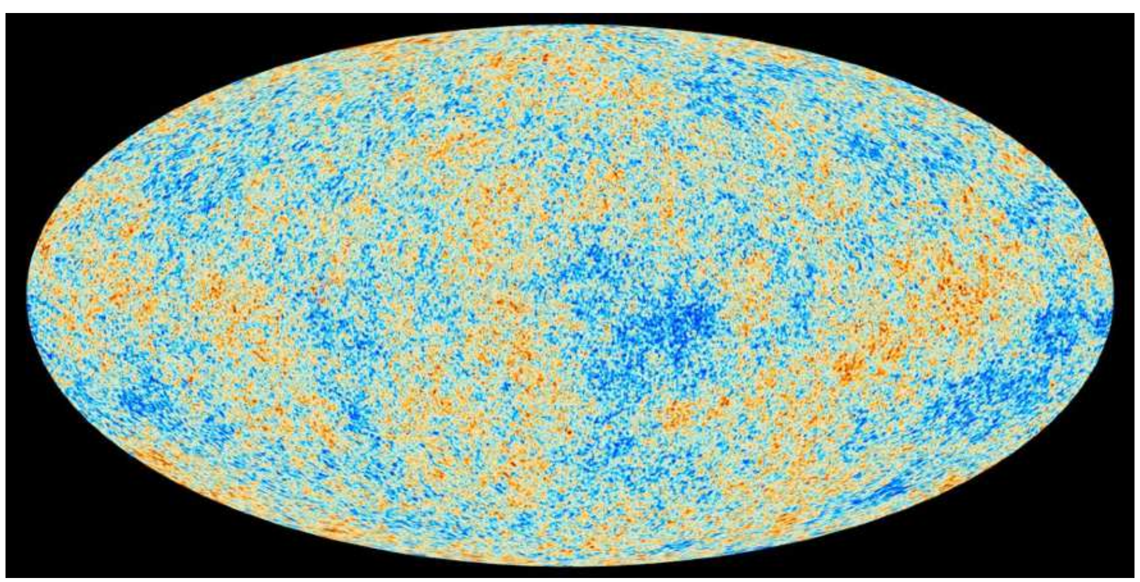
[Blei, 2012]

Related to CMB Foreground Separation

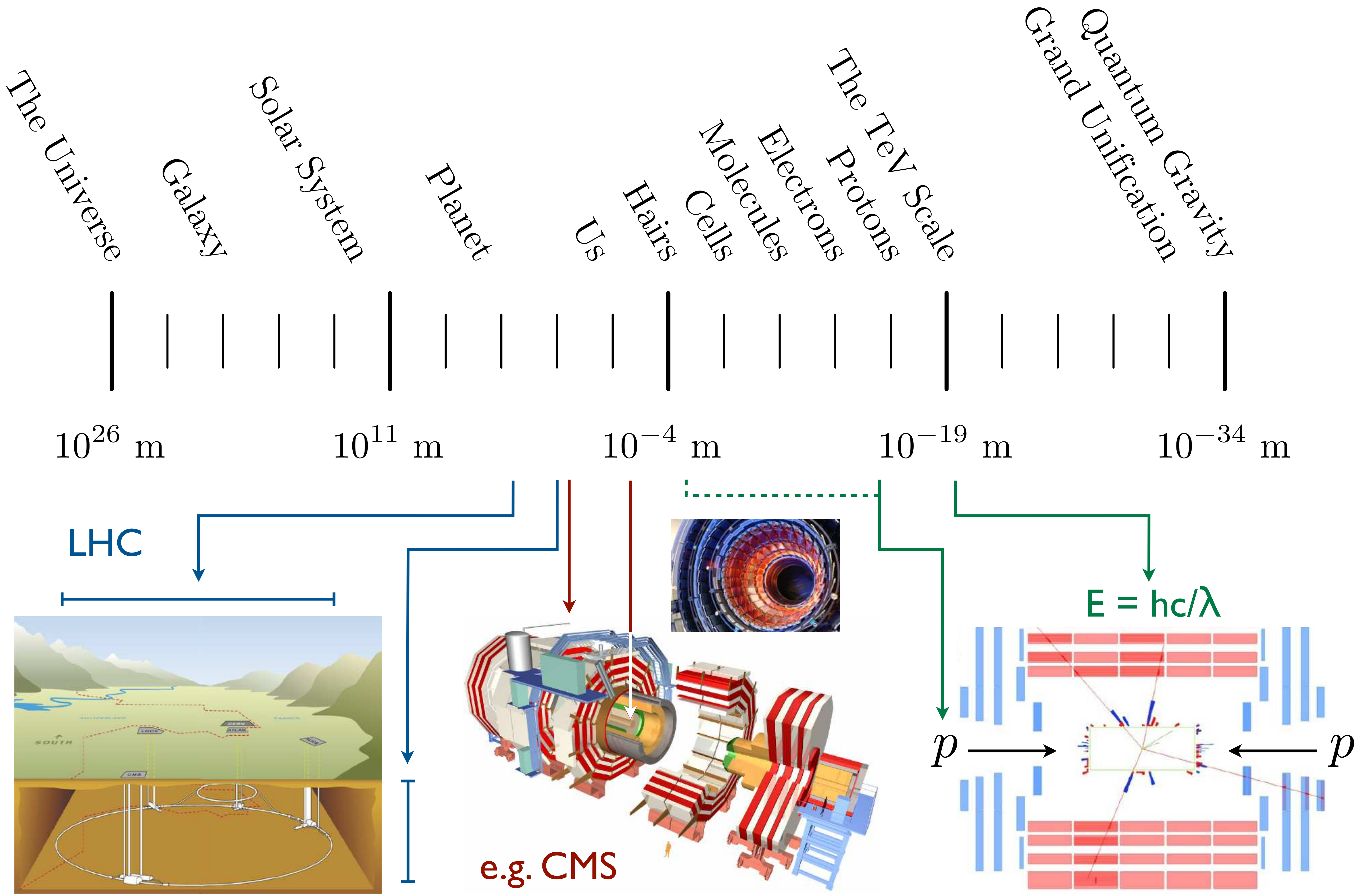




$\Rightarrow$



[Planck Outreach]


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

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

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

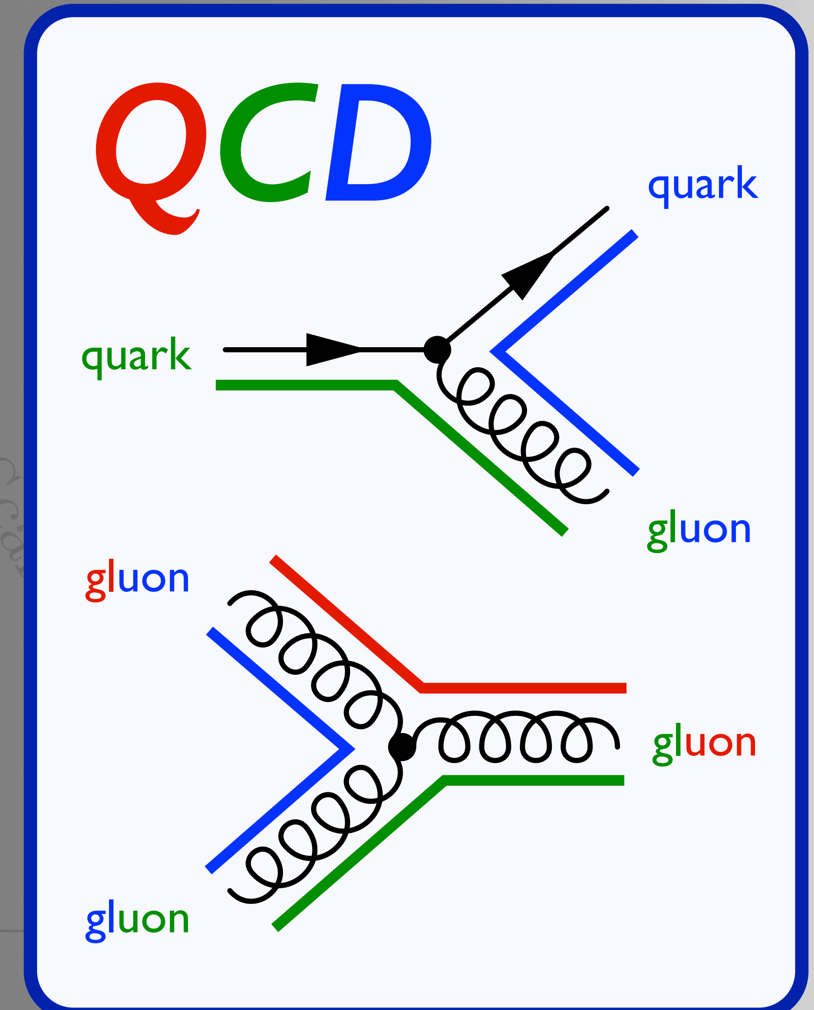
Tetraquarks (?)

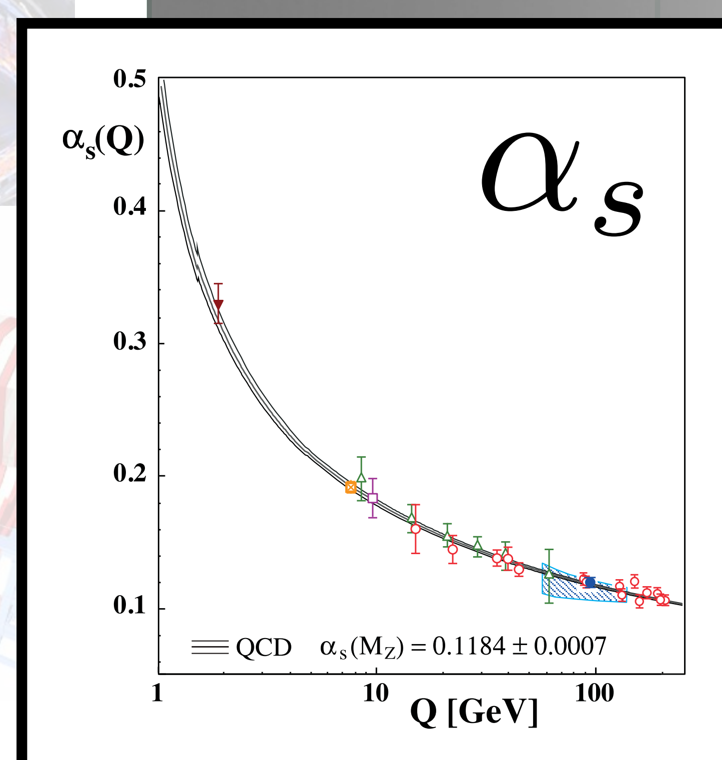
$X(3872)\,Y(4260)\,Z(4430)\,\ldots$

Pentaquarks (?)

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

Jets





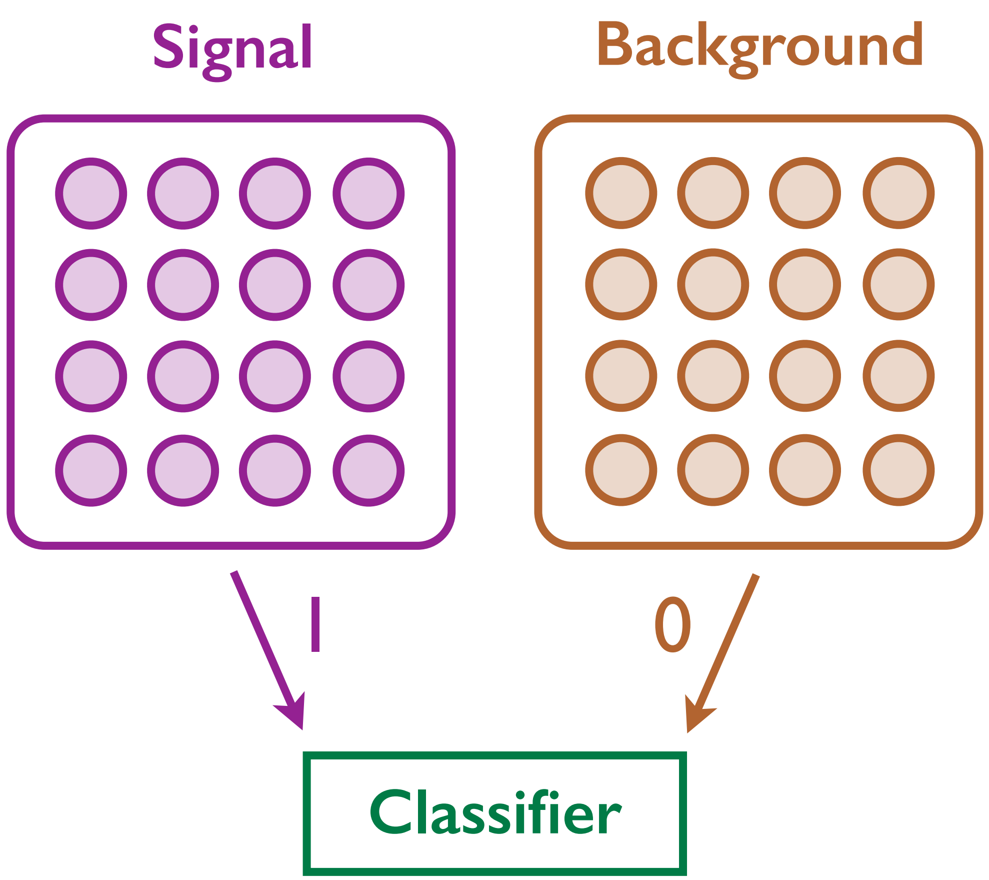
A Cartoon of Machine Learning

For fully-supervised jet classification

(see backup for regression, generation, modeling)

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

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)

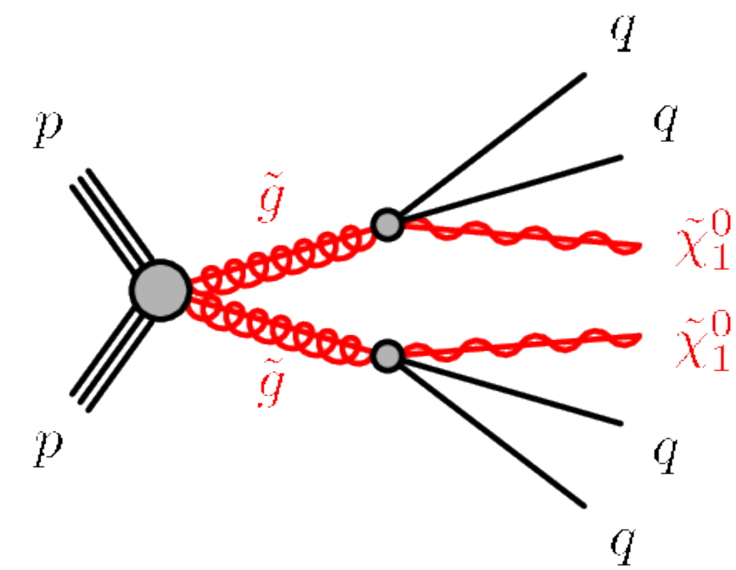
E.g. SUSY Search for Gluino Pairs

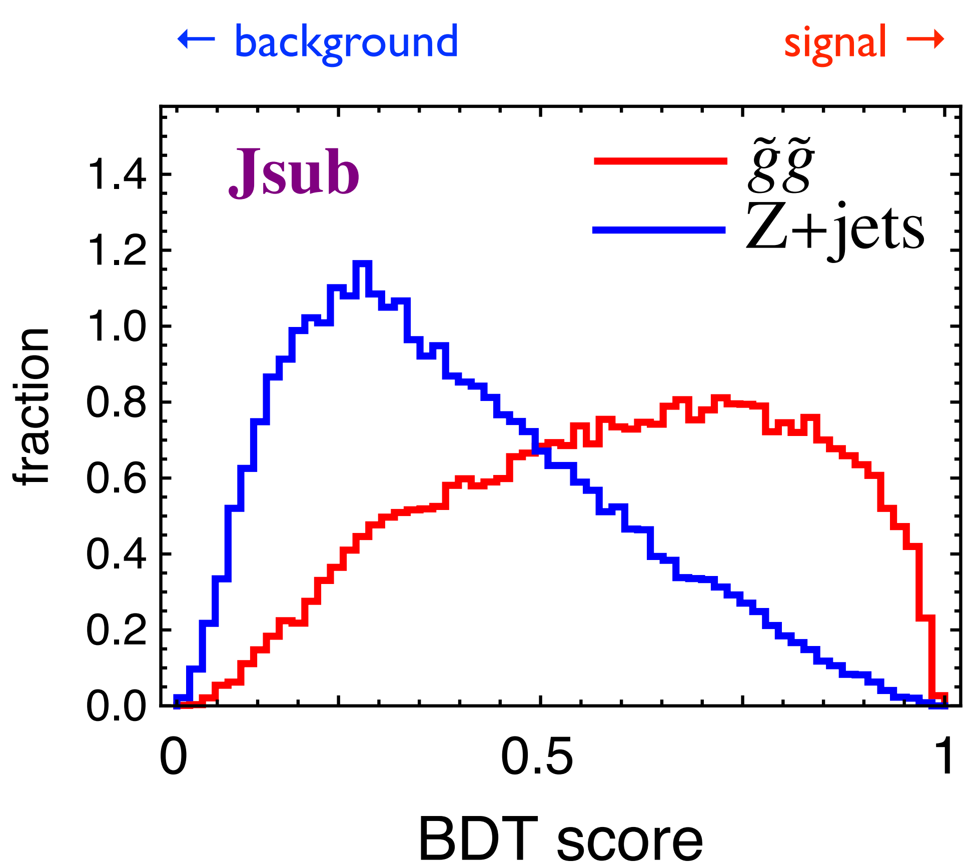
Classifier: Boosted decision tree (for each of 4 jets)

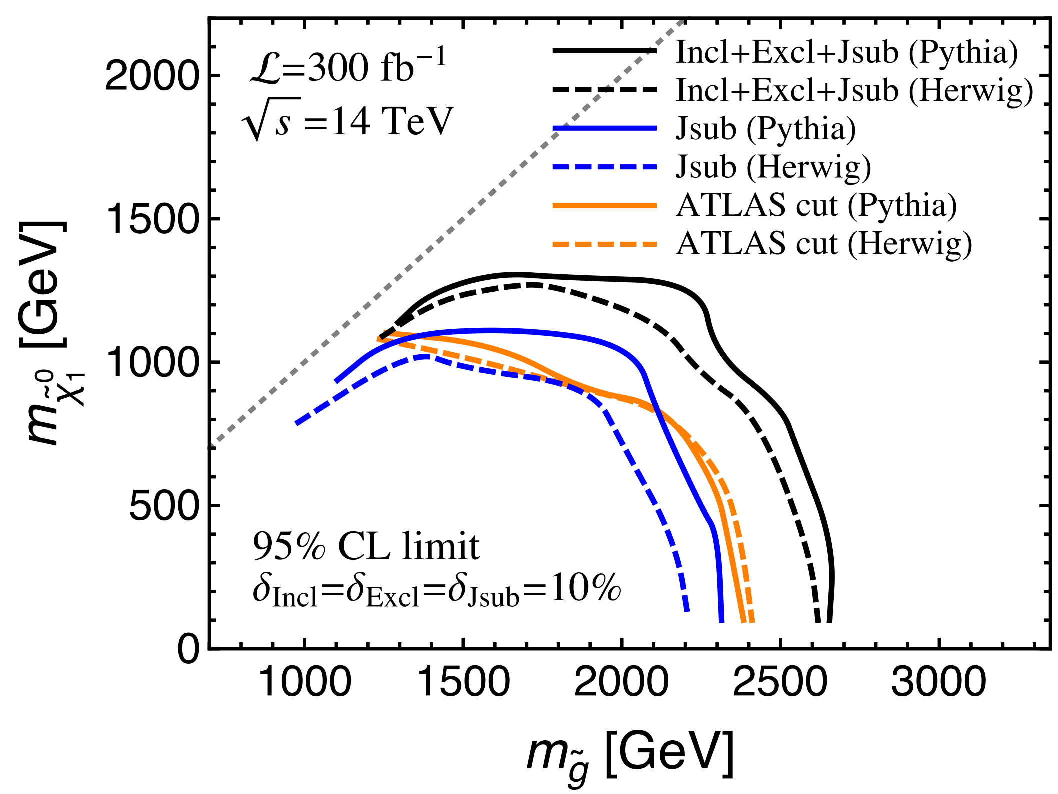
Inputs: Jet mass, width, track multiplicity

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

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

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

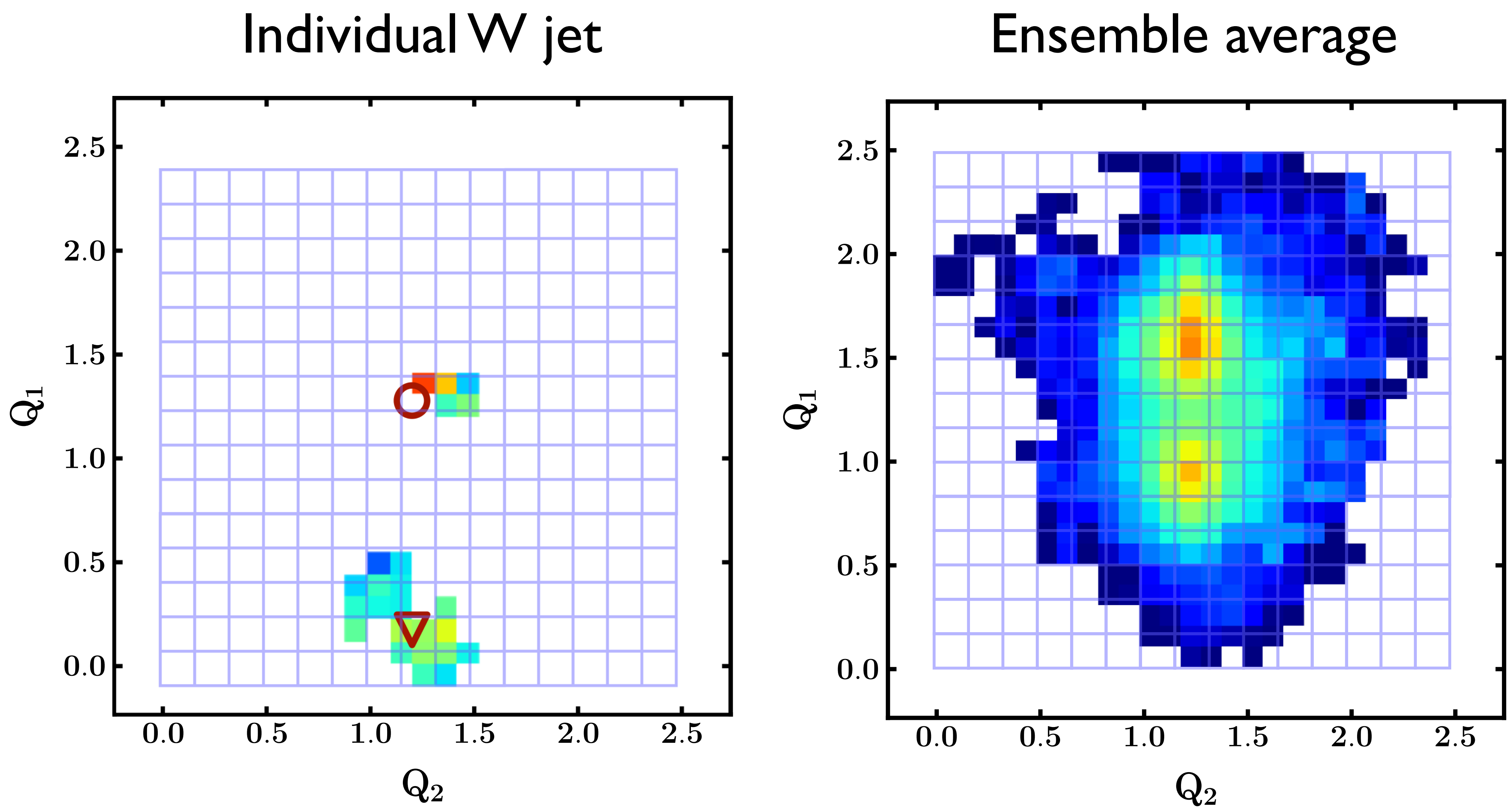
Signal	vs.	Background
Quark Jets	vs.	Gluon Jets
Up-type Quarks	vs.	Down-type Quarks
W/Z Bosons	vs.	QCD Jets
W Bosons	vs.	Z Bosons
Top Quarks	vs.	QCD Jets
Exotic Boosted Objects	vs.	QCD Jets
CMS Open Data Samples	vs.	Each other
…	vs.	…

[Lönnblad, Peterson, Rögnvaldsson, 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; plus many ATLAS/CMS performance studies]

Jet Classification Studies

Mix and match

Standard CNN input: Jet images

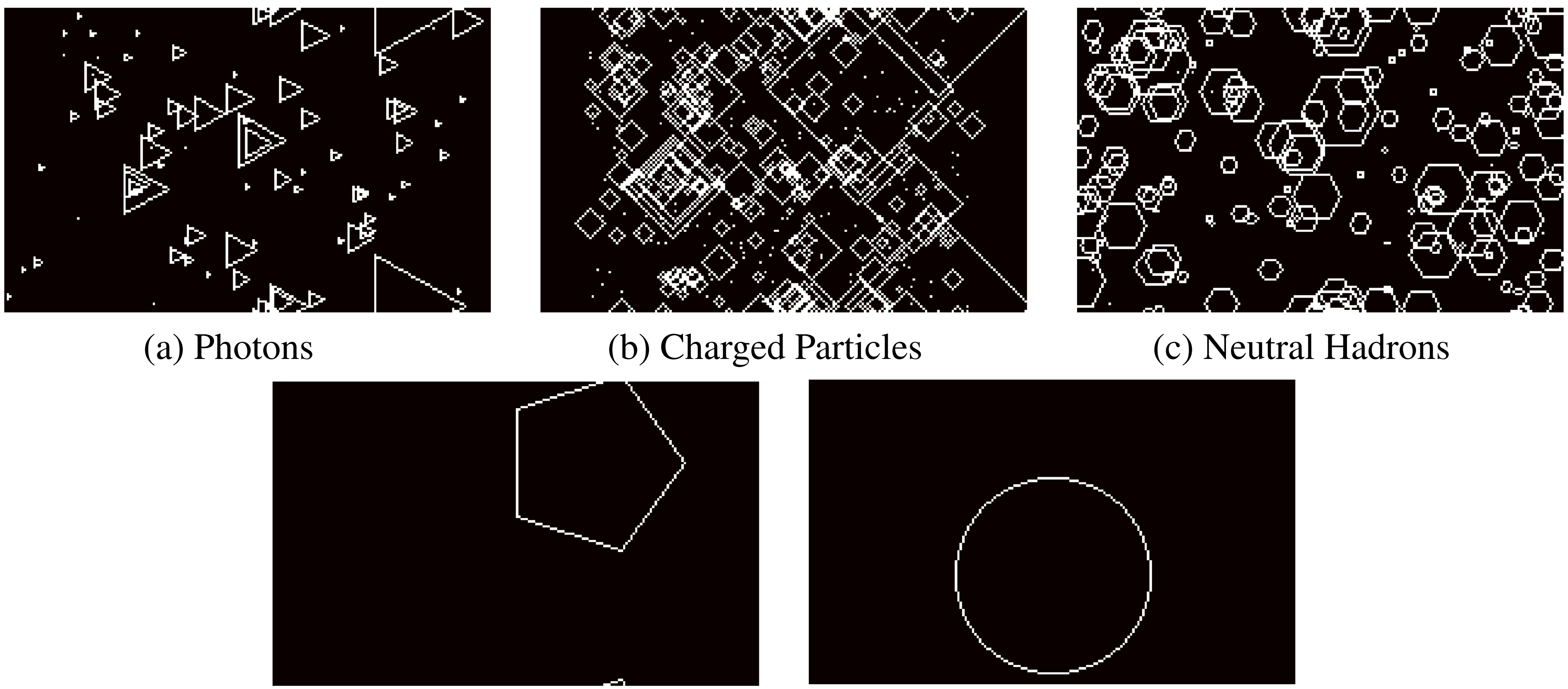


[Cogan, Kagan, Strauss, Schwartzman, 1407.5675]

Jet Classification Studies

Mix and match

Novel CNN input: Abstract event images



(a) Photons (b) Charged Particles (c) Neutral Hadrons

(d) Lepton (e) E_T^{miss}

Addresses sparsity problem of standard energy-to-intensity mapping

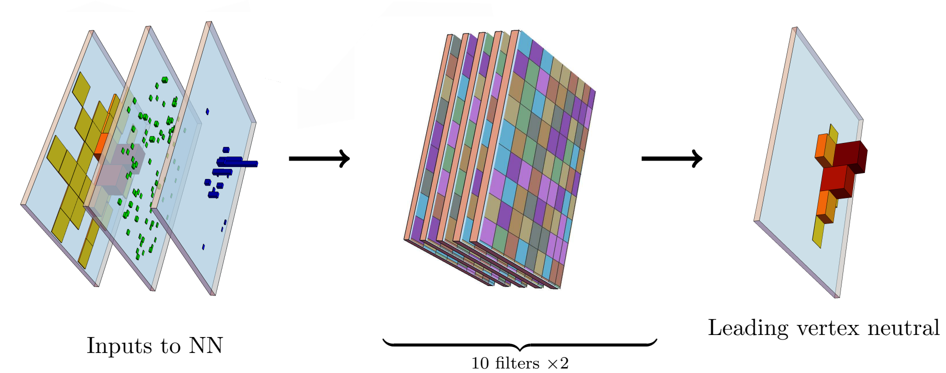
[Nguyen, Weitekamp, Anderson, Castello, Cerri, Pierini, Spiropulu, Vlimant, 1807.00083; using Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, 1708.07034]

Beyond Classification

PUMML

Pileup Mitigation

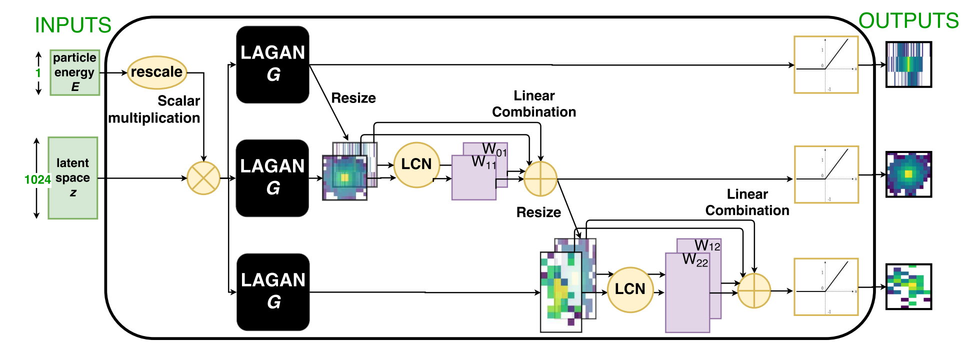
[Komiske, Metodiev, Nachman, Schwartz, 1707.08600]



CaloGAN

Fast Detector Simulation

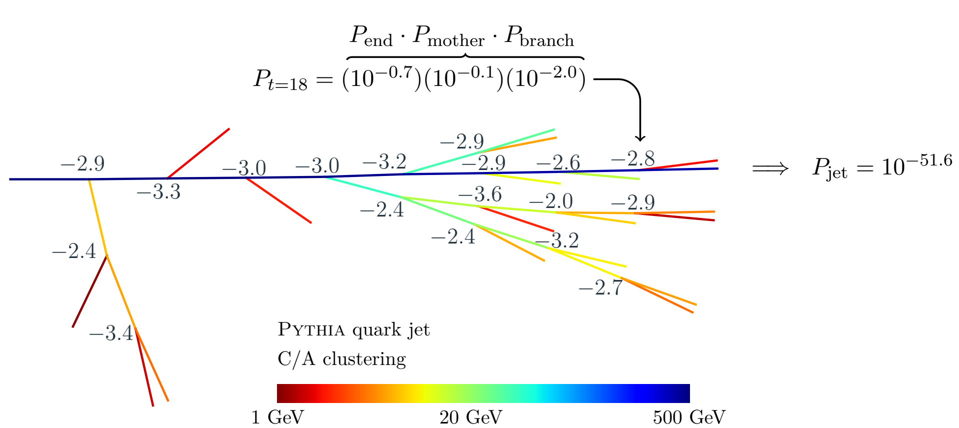
Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321;
see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



JUNIPR

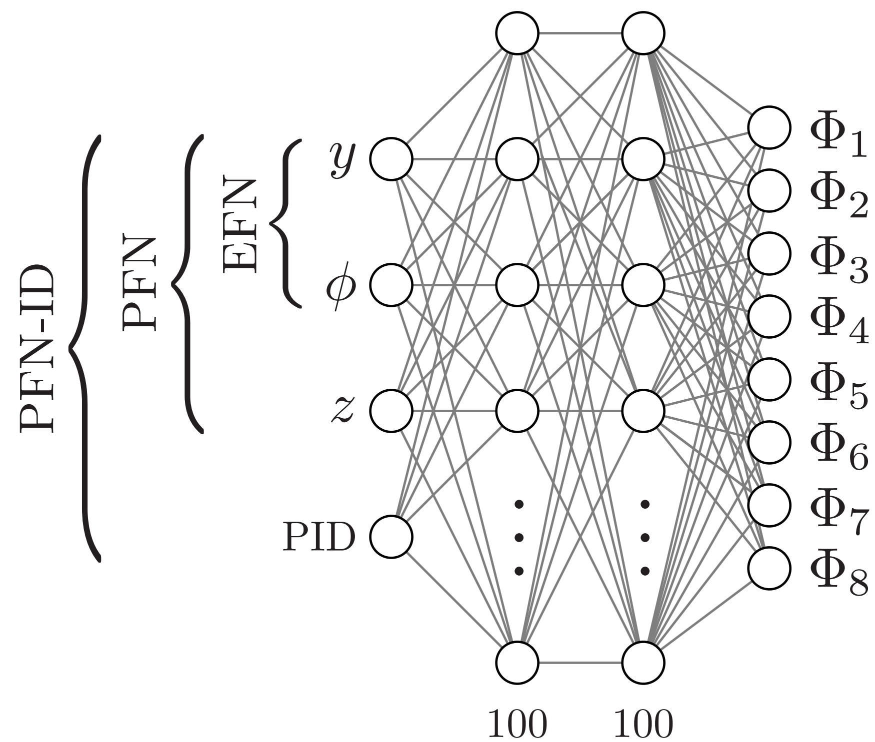
Probability Modeling

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

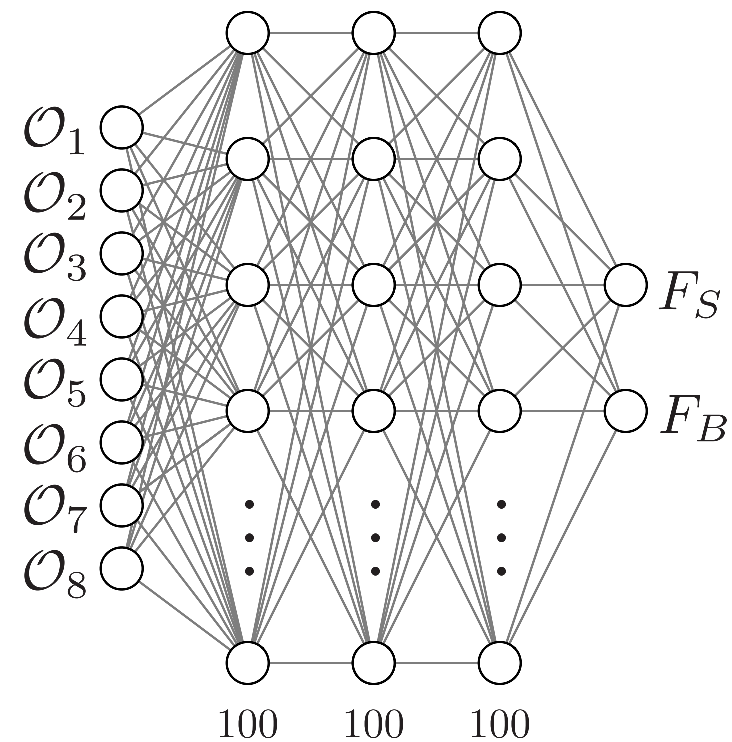


Architecture Details

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

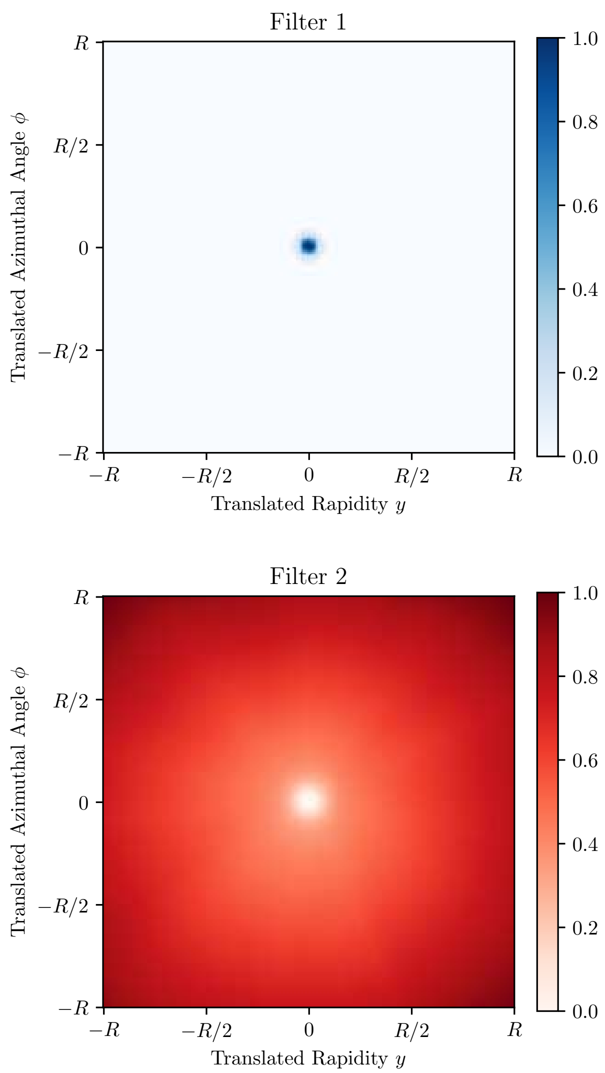
Final Discriminant:

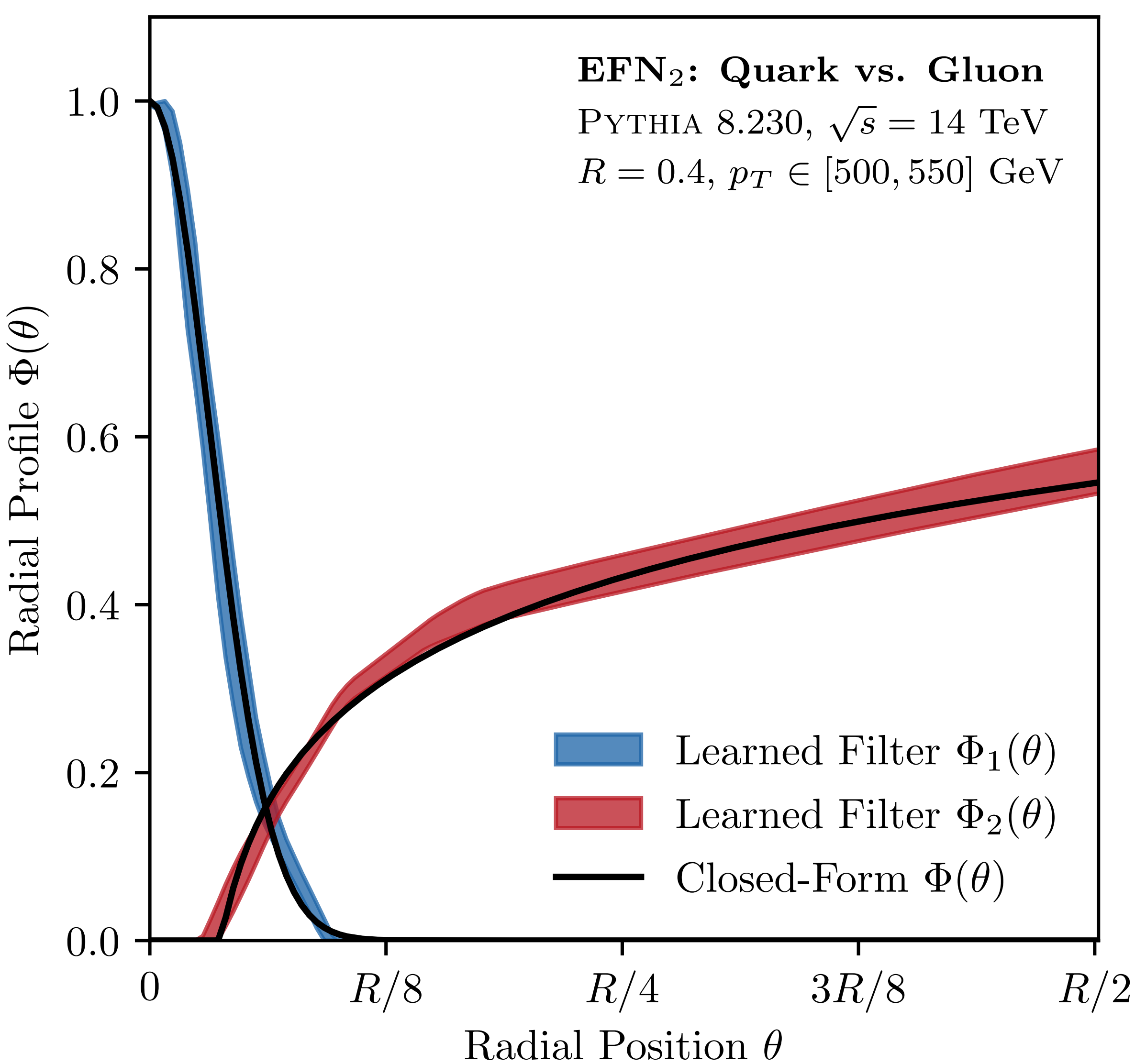
$$\text{softmax}(F_S, F_B)$$

“What is the Machine Learning?”

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

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

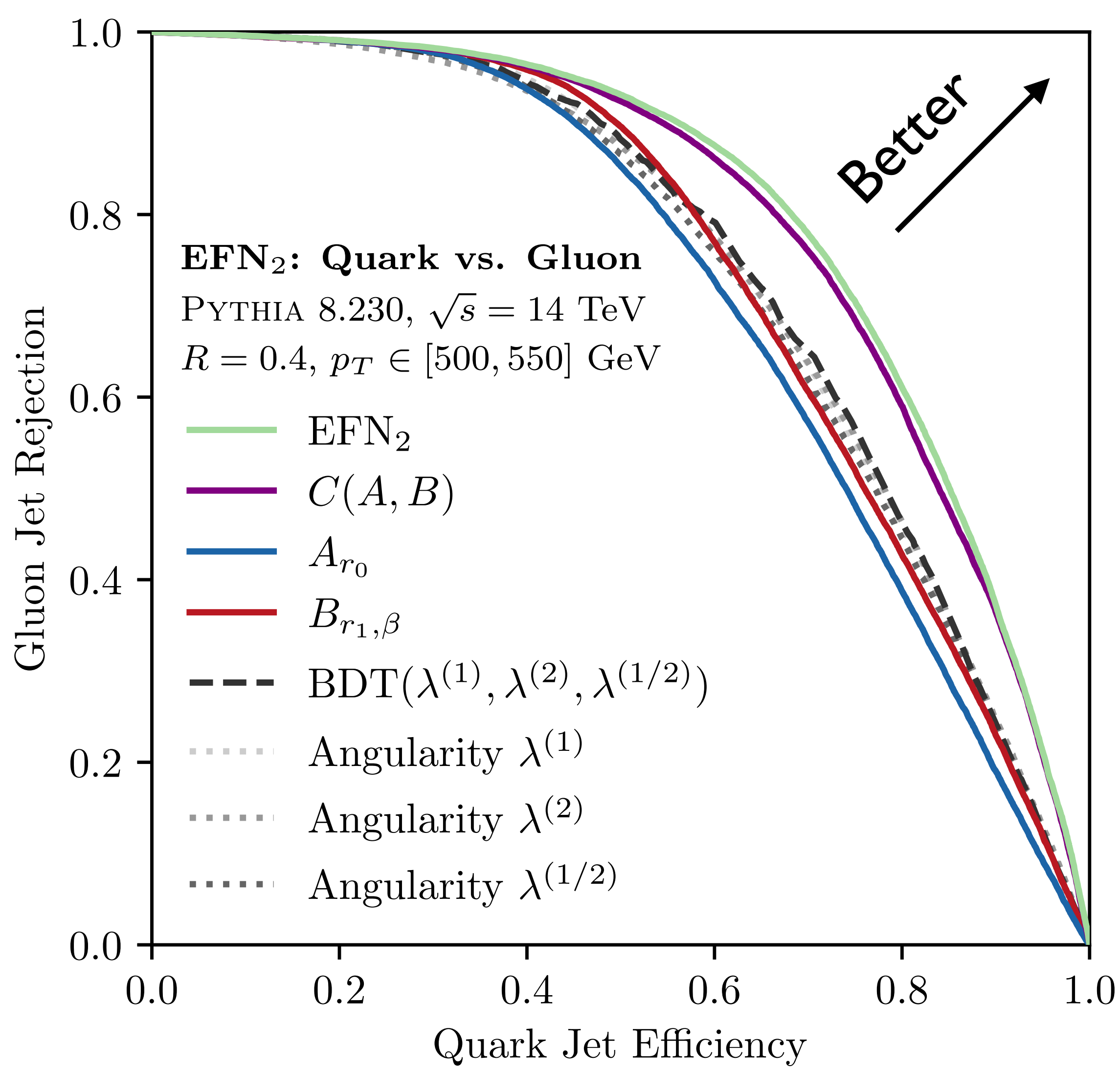




"What is the Machine Learning?"

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