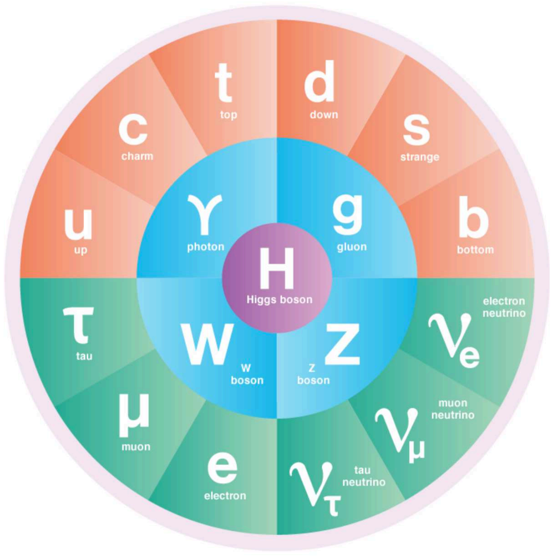
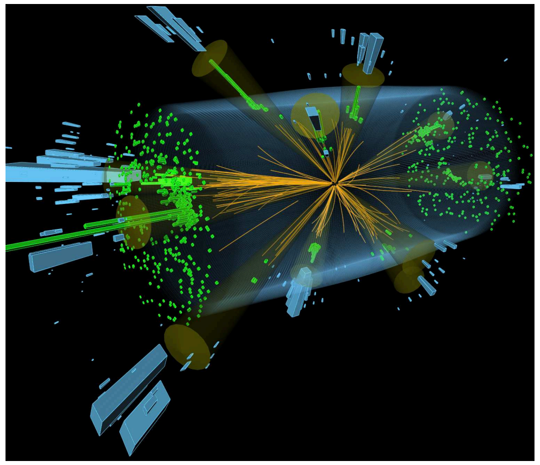
Point Cloud Strategies for Boosted Objects

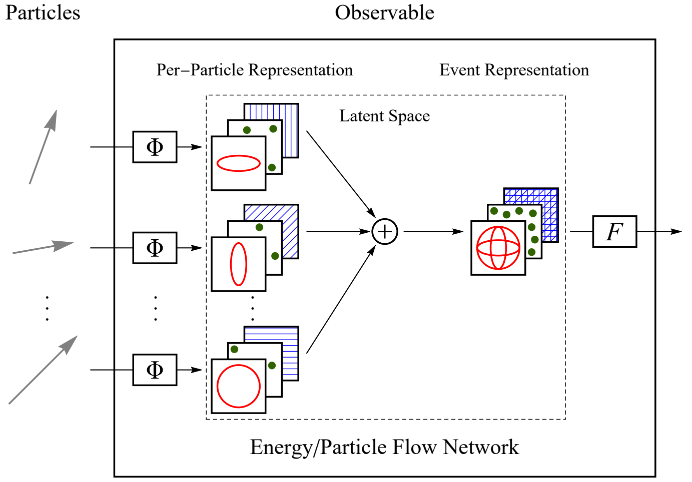
Patrick T. Komiske III
Massachusetts Institute of Technology
Center for Theoretical Physics

with Eric Metodiev and Jesse Thaler, 1810.05165

ML-HEP LBL Meetup
April 17, 2019



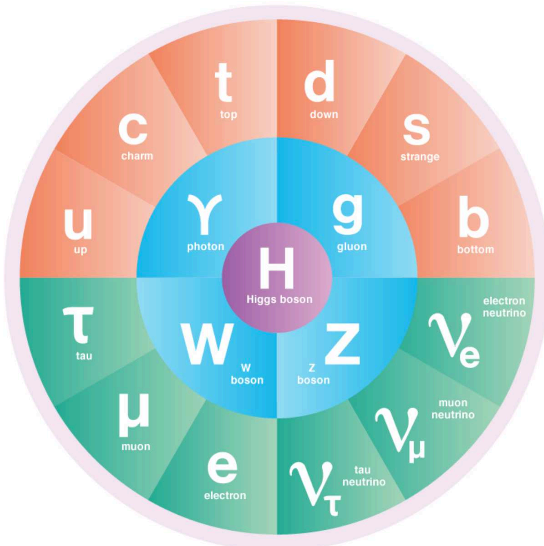




Jets in Particle Physics

Point Clouds

Energy Flow Networks



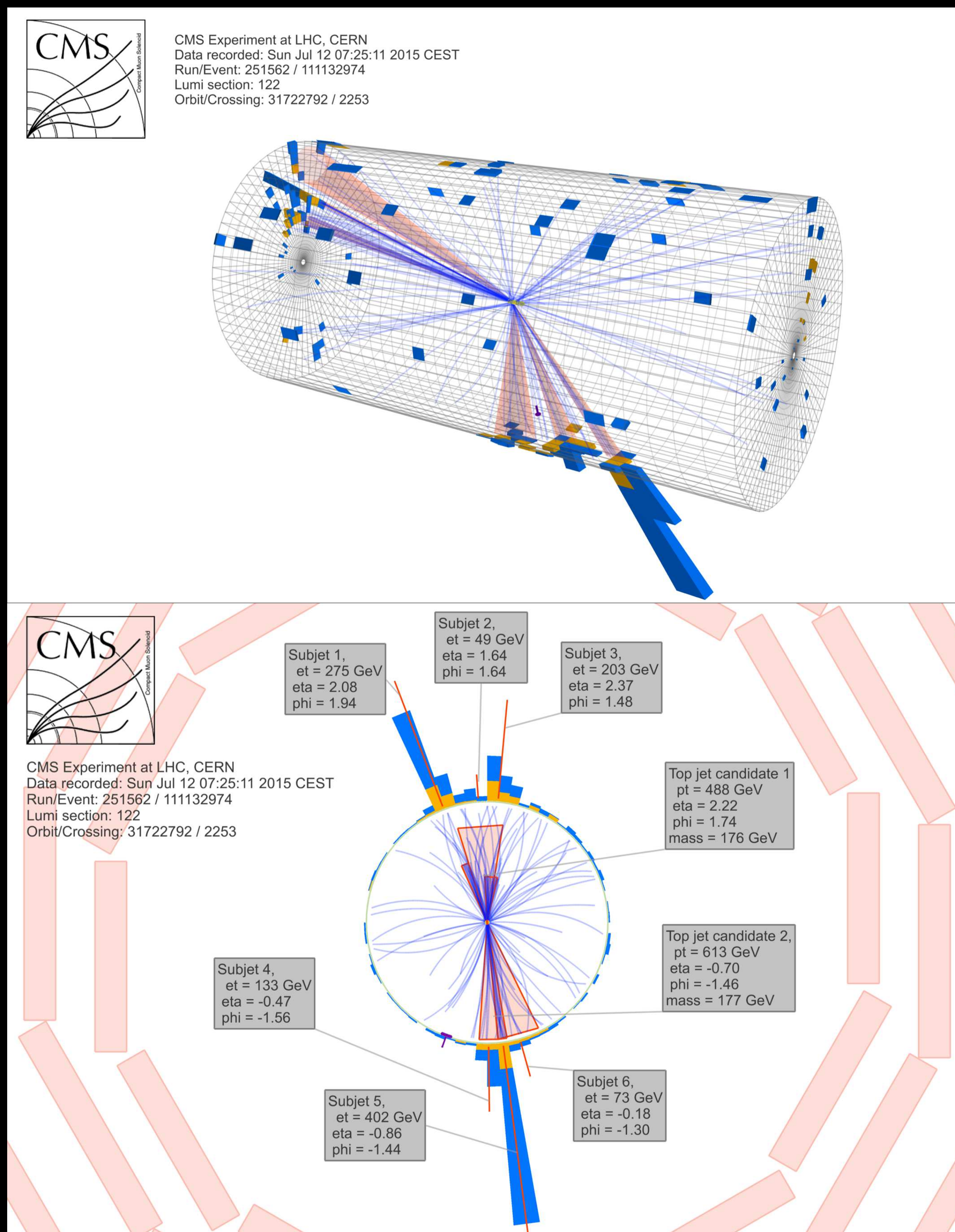
Jets in Particle Physics

Point Clouds

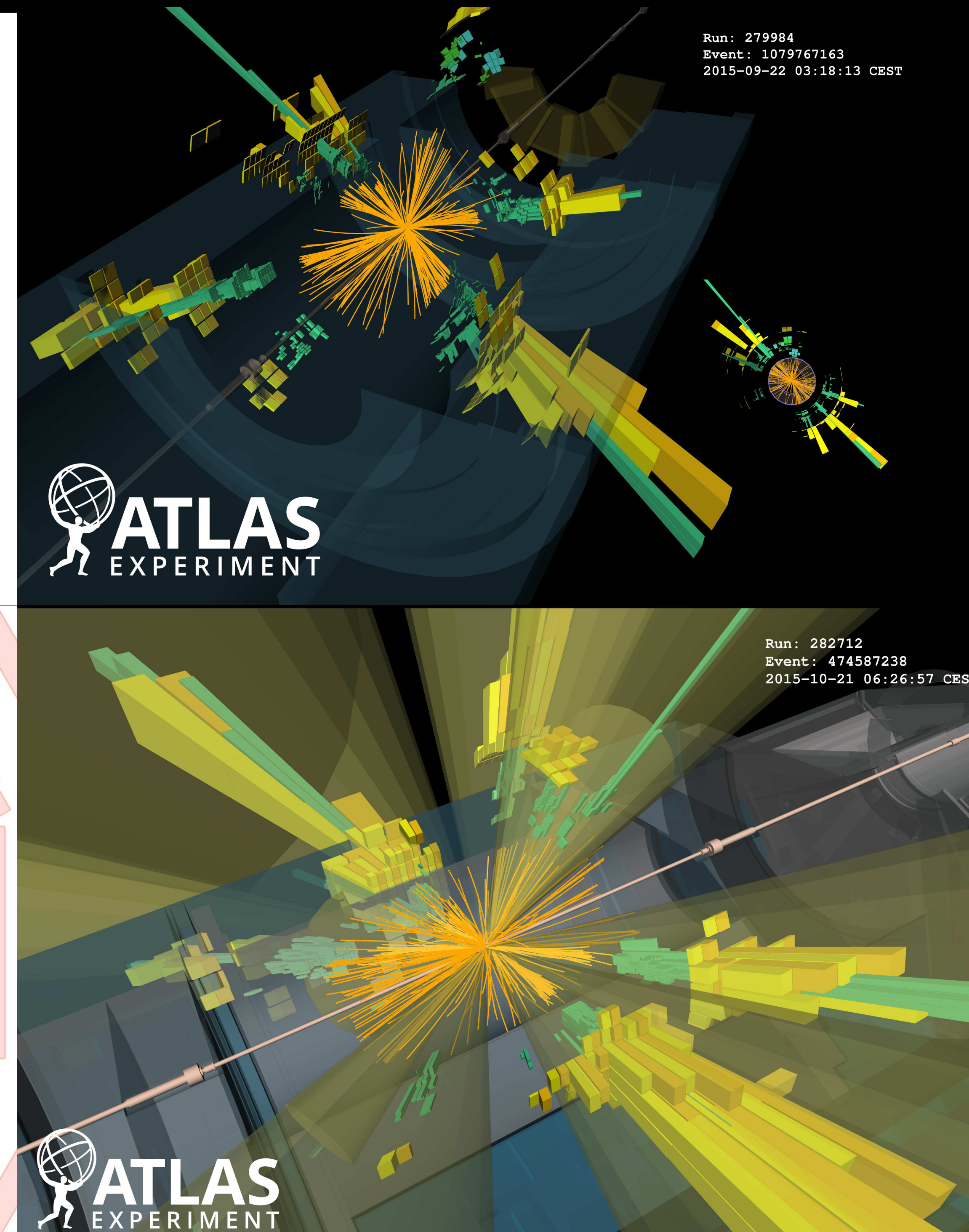
Energy Flow Networks

Jets at the LHC

New physics searches and standard model measurements involve jets (collimated sprays of hadrons)

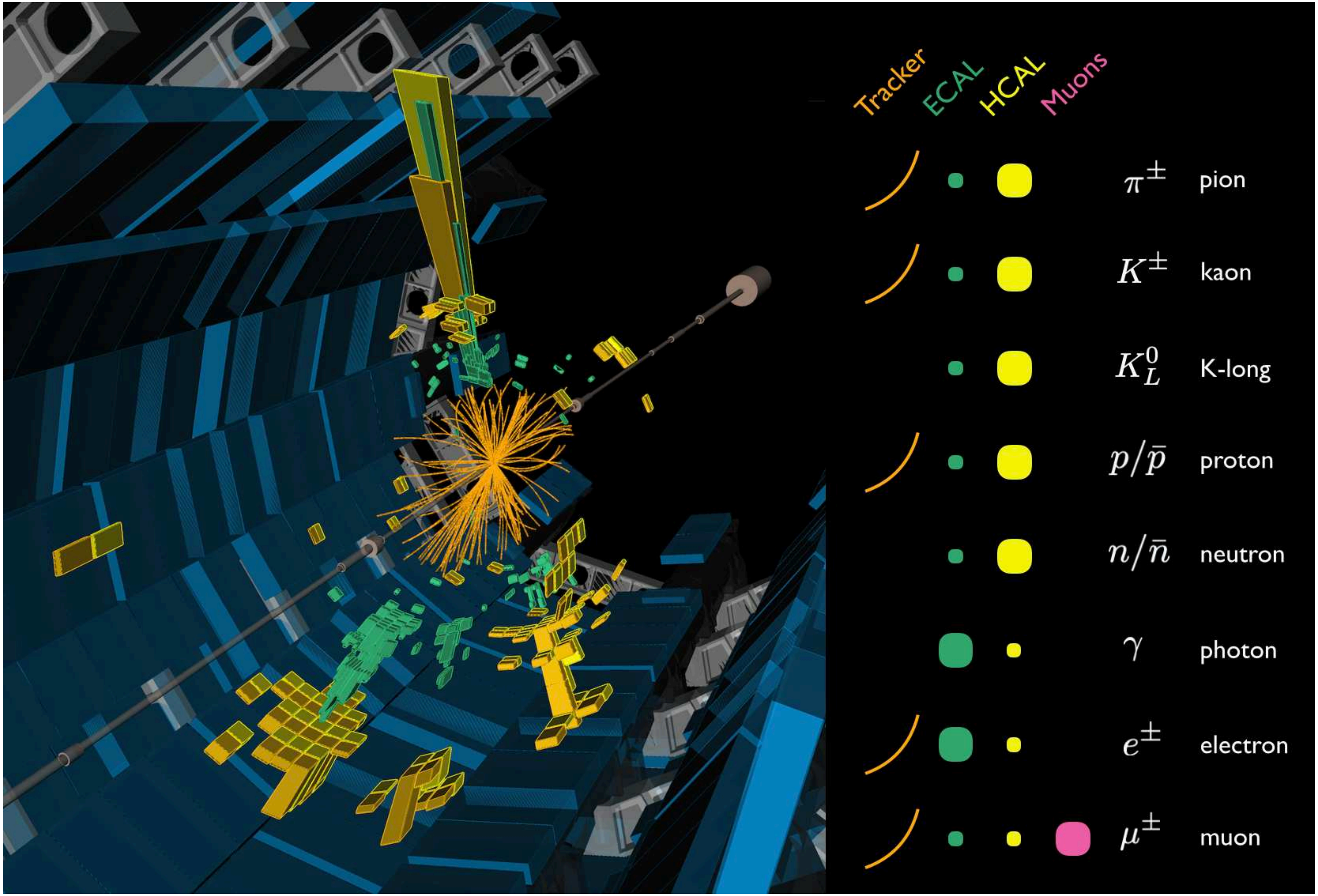


CMS hadronic $t\bar{t}$ event



ATLAS high jet multiplicity events

Jets in Data


Tracker
ECAL
HCAL
Muons
$\pi^{\pm}$ pion
$K^{\pm}$ kaon
K_L^0 K-long
$p/\bar{p}$ proton
$n/\bar{n}$ neutron
γ photon
$e^{\pm}$ electron
$\mu^{\pm}$ muon

Jets in Theory

Hard collision

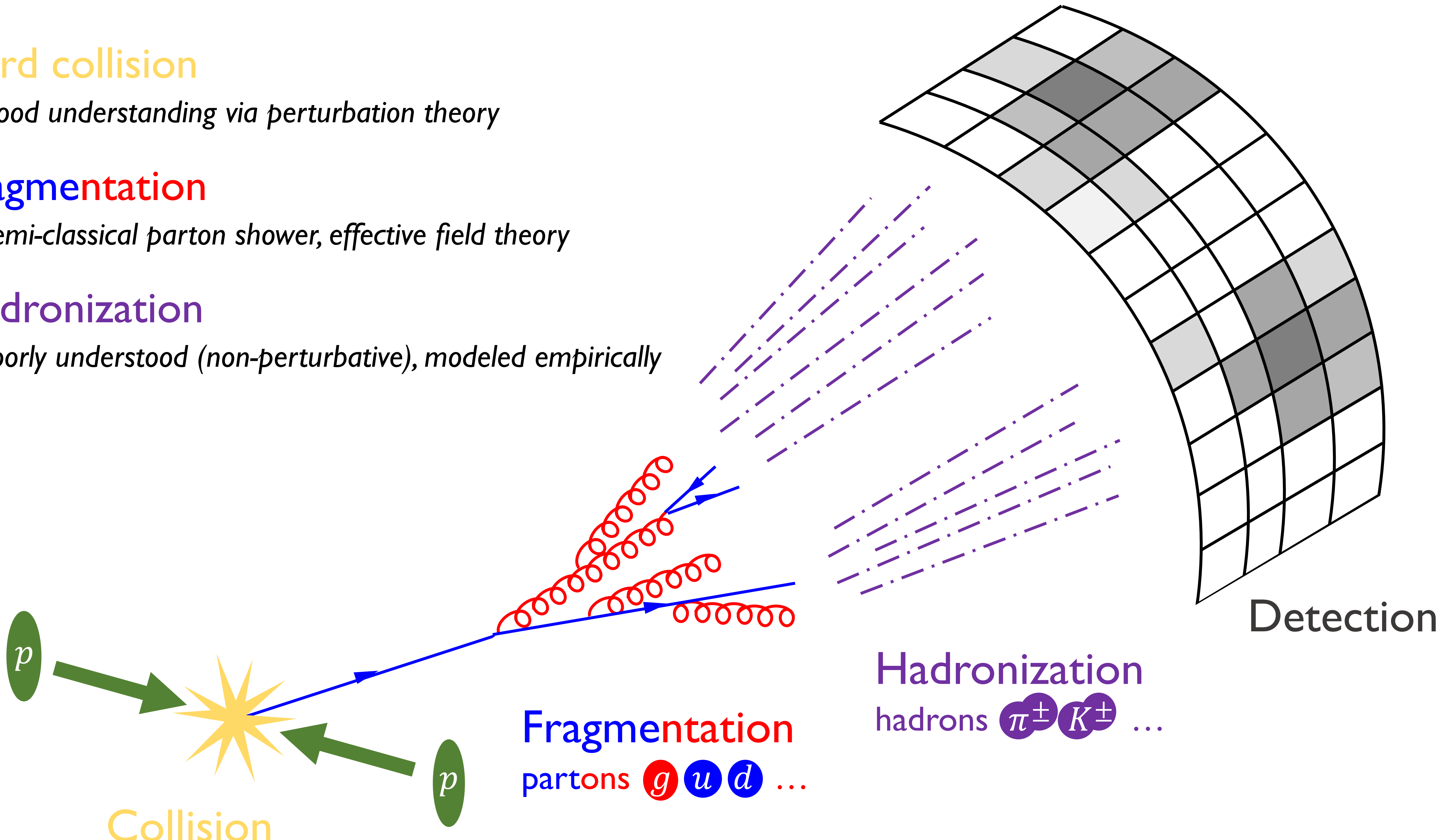
Good understanding via perturbation theory

Fragmentation

Semi-classical parton shower, effective field theory

Hadronization

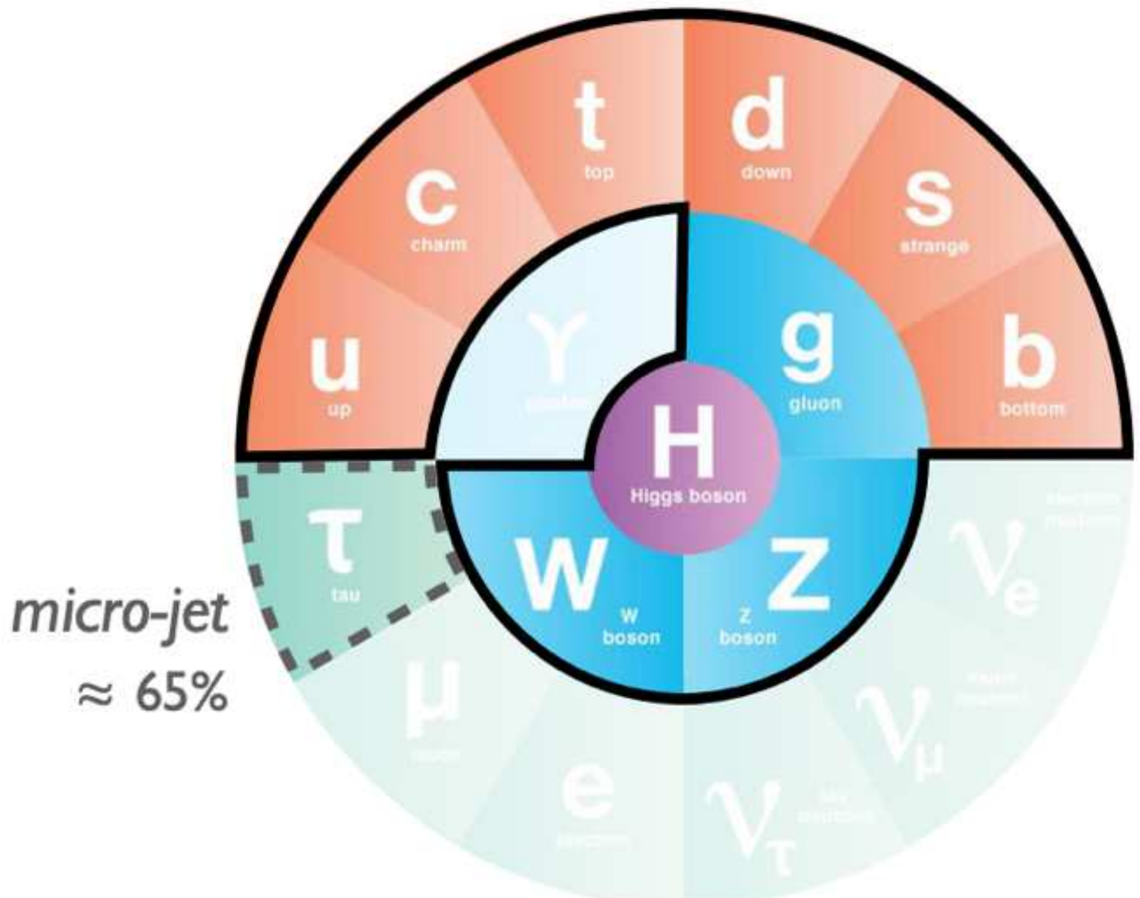
Poorly understood (non-perturbative), modeled empirically



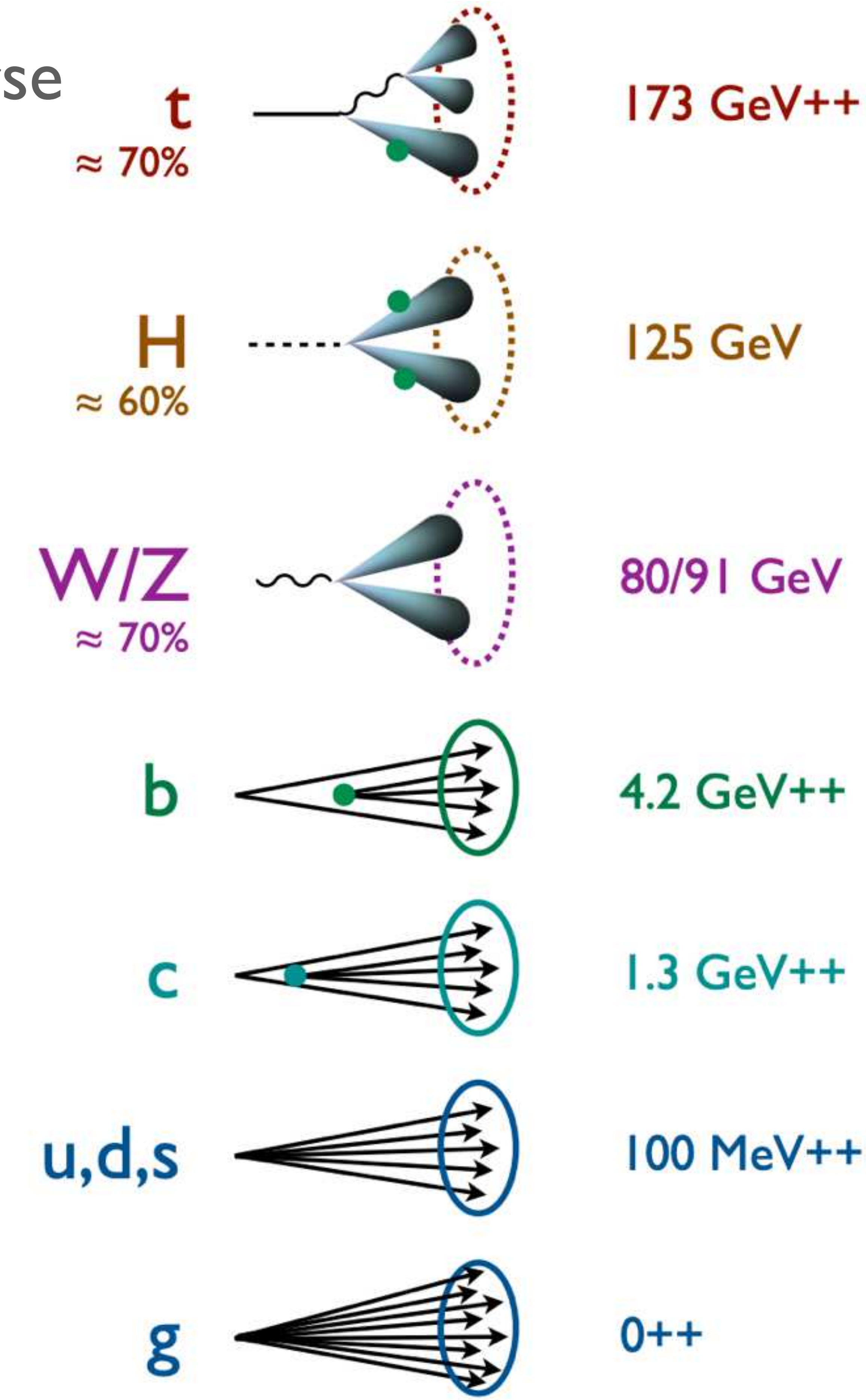
Cartoon of jet formation as a multi-scale process

Diagram by Eric Metodiev

Jet Menu of the (Known) Universe

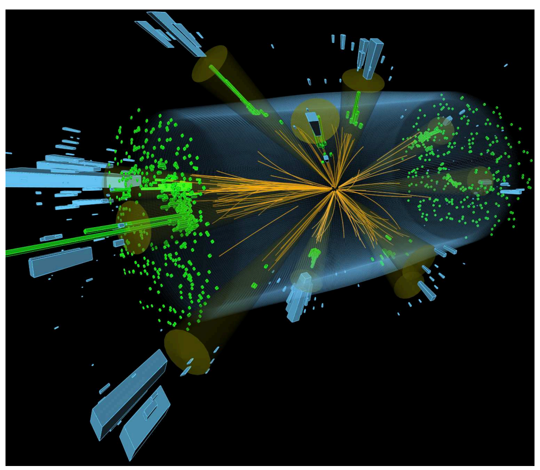


Jets from the Standard Model



++ = Mass from QCD Radiation

Slide by Jesse Thaler



Jets in Particle Physics

Point Clouds

Energy Flow Networks

What is a Jet?

An ***unordered****,* ***variable length*** *collection of particles*

Due to quantum-mechanical indistinguishability
Due to probabilistic nature of jet formation

$$J\left(\{p_1^\mu,\ldots,p_M^\mu\}\right) = J(\{p_{\pi(1)}^\mu,\ldots,p_{\pi(M)}^\mu\}), \qquad \underbrace{M \geq 1}_{\text{Multiplicity}}\ , \qquad \underbrace{\forall\,\pi \in S_M}_{\text{Permutations}}$$

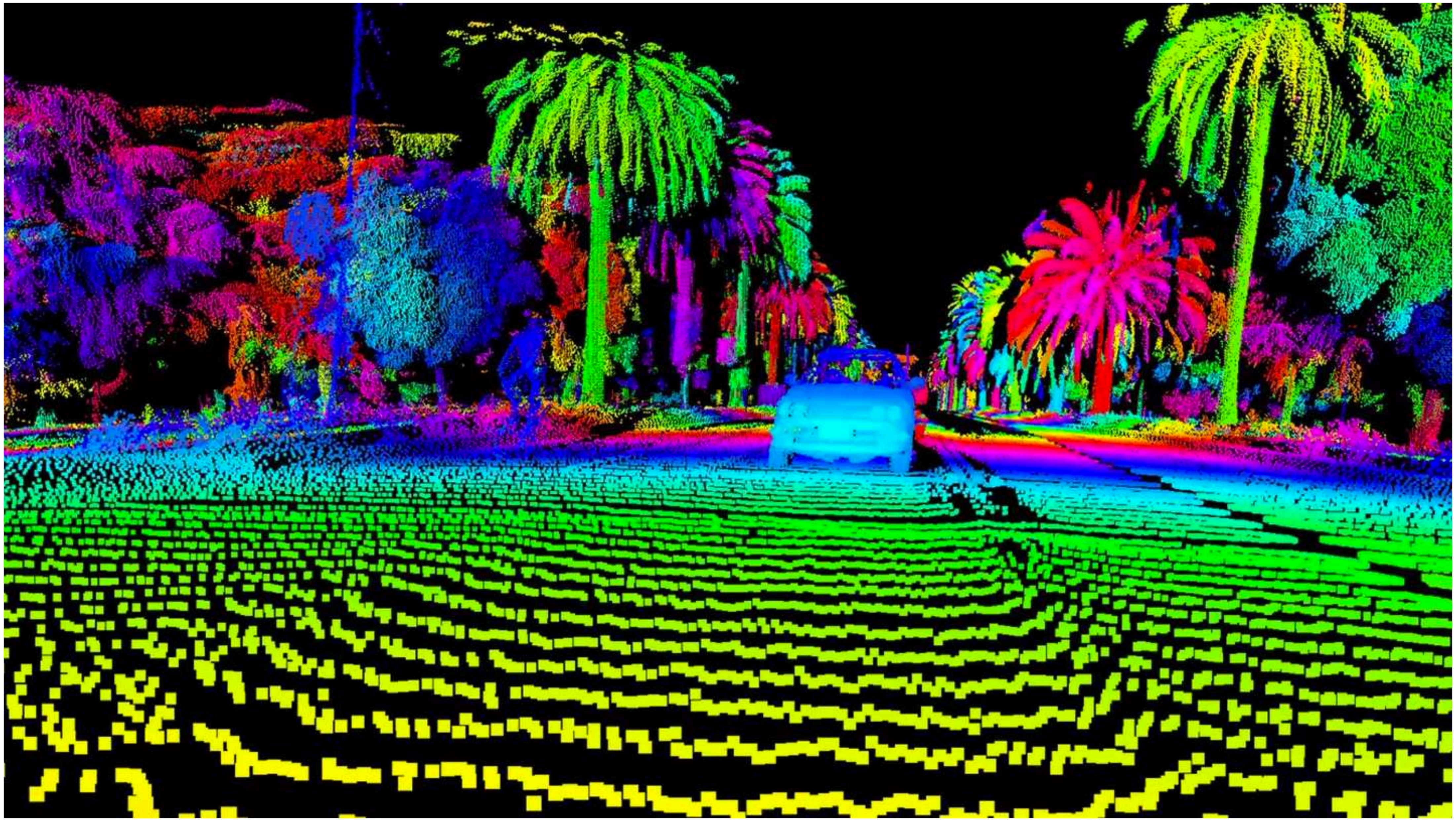
p_i^μ represents *all* the particle properties:

- Four-momentum – $(E, p_x, p_y, p_z)_i^\mu$
- Other quantum numbers (e.g. particle id, charge)
- Experimental information (e.g. vertex info, quality criteria, PUPPI weights)

Point Clouds

Point cloud: "A set of data points in space" –Wikipedia

LIDAR data from self-driving car sensor



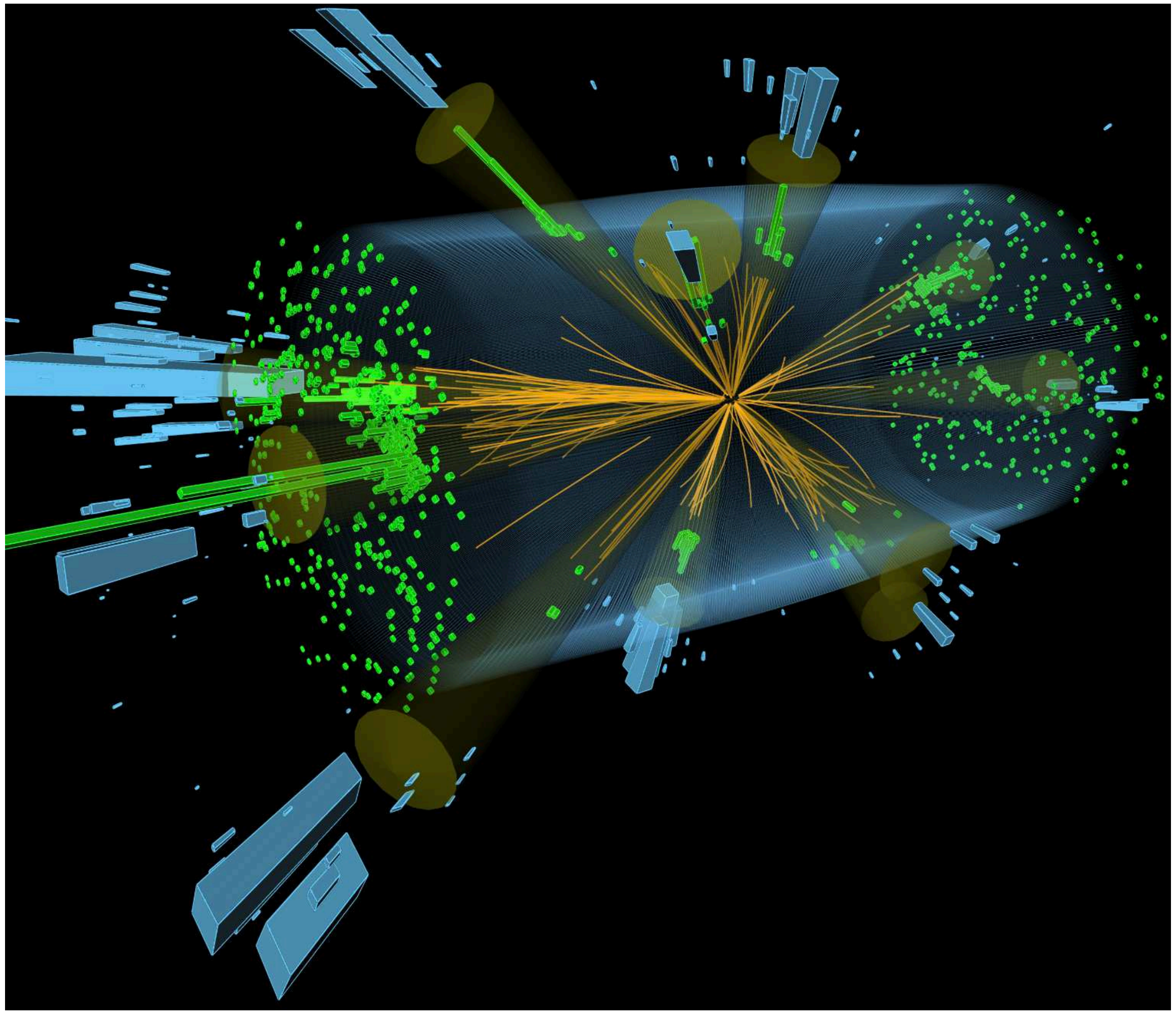
Particle Collision Events as Point Clouds

Point cloud: "A set of data points in space" –Wikipedia

Jet/event — set

Particles — data points

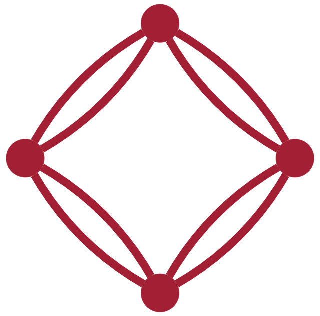
Feature space — space



Multi-jet event at CMS

Processing Point Clouds

Methods for processing point clouds/jets should respect the appropriate symmetries

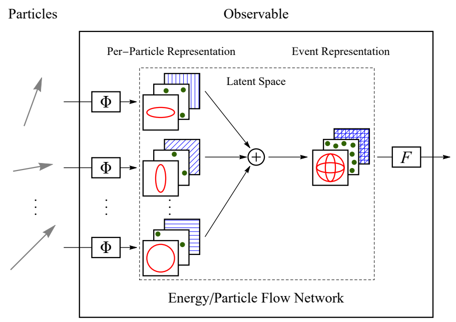


Variable constituent multiplicity requires at least one of:

Preprocessing to another representation (jet images, *N*-subjettiness, etc.)

Truncation to an (arbitrary) fixed size

Recurrent NN structure



Particle permutation symmetry requires:

Permutation symmetric observables

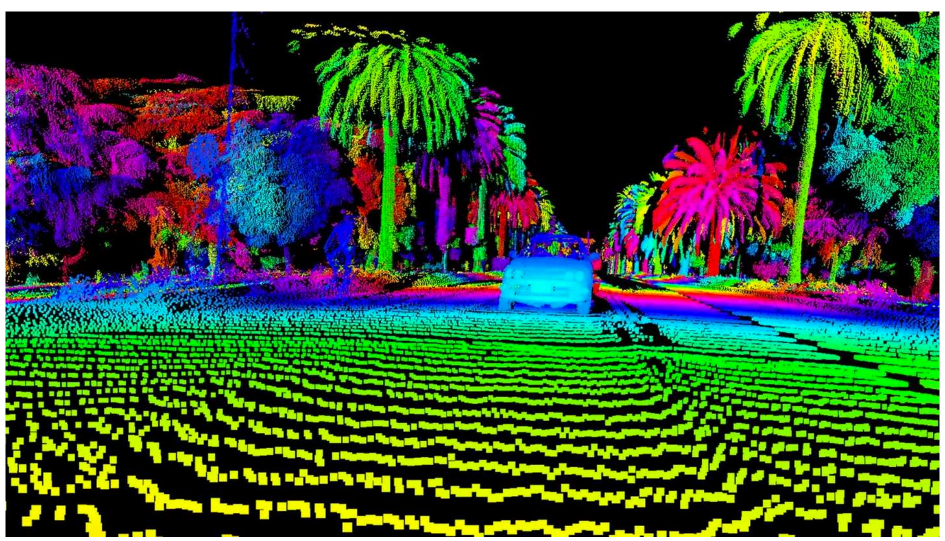
Permutation symmetric architectures

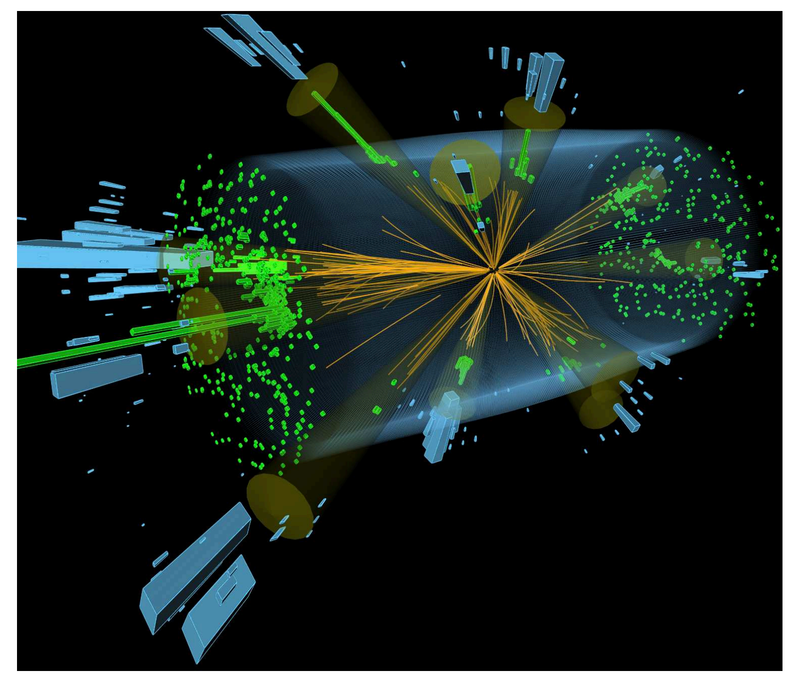
Processing Point Clouds

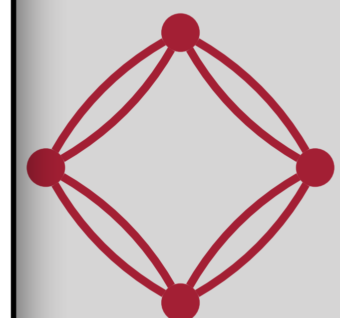
Met

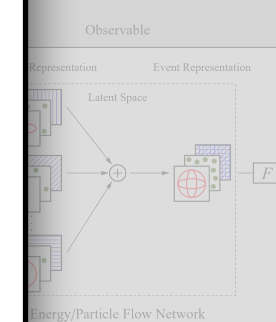
t the

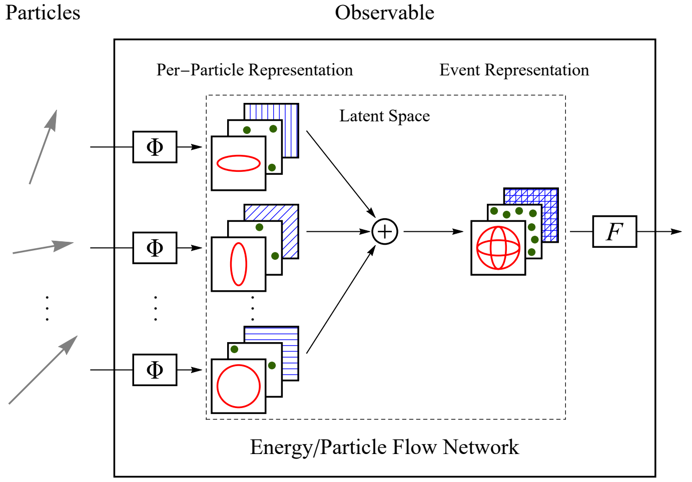
How do we make a machine learning architecture to process point clouds?











Jets in Particle Physics

Point Clouds

Energy Flow Networks

Deep Sets

Namespace for symmetric function parametrization

A general permutation-symmetric function is *additive* in a latent space (piece of math)

Deep Sets

[1703.06114]

Manzil Zaheer[1,2], Satwik Kottur[1], Siamak Ravanbhakhsh[1], Barnabás Póczos[1], Ruslan Salakhutdinov[1], Alexander J Smola[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Deep Sets Theorem [63]. *Let $\mathfrak{X} \subset \mathbb{R}^d$ be compact, $X \subset 2^{\mathfrak{X}}$ be the space of sets with bounded cardinality of elements in $\mathfrak{X}$, and $Y \subset \mathbb{R}$ be a bounded interval. Consider a continuous function $f : X \to Y$ that is invariant under permutations of its inputs, i.e. $f(x_1, \ldots, x_M) = f(x_{\pi(1)}, \ldots, x_{\pi(M)})$ for all $x_i \in \mathfrak{X}$ and $\pi \in S_M$. Then there exists a sufficiently large integer ℓ and continuous functions $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ such that the following holds to an arbitrarily good approximation:*[1]

$$f(\{x_1, \ldots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)$$

Deep Sets

Namespace for symmetric function parametrization

A general permutation-symmetric function is *additive* in a latent space (piece of math)

Deep Sets

[1703.06114]

Manzil Zaheer[1,2], **Satwik Kottur**[1], **Siamak Ravanbhakhsh**[1],
Barnabás Póczos[1], **Ruslan Salakhutdinov**[1], **Alexander J Smola**[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Feature space

Variable length

Deep Sets Theorem [63]. *Let* $\boxed{\mathfrak{X} \subset \mathbb{R}^d}$ *be compact,* $\boxed{X \subset 2^{\mathfrak{X}}$ *be the space of sets*$}$ *with bounded cardinality of elements in* $\mathfrak{X}$, *and* $Y \subset \mathbb{R}$ *be a bounded interval. Consider a continuous function* $f : X \to Y$ *that is invariant under permutations of its inputs, i.e.* $f(x_1, \dots, x_M) = \boxed{f(x_{\pi(1)}, \dots, x_{\pi(M)}) \text{ for all } x_i \in \mathfrak{X} \text{ and } \pi \in S_M}$. *Then there exists a sufficiently large integer* ℓ *and continuous functions* $\Phi : \mathfrak{X} \to \boxed{\mathbb{R}^\ell}$, $F : \boxed{\mathbb{R}^\ell} \to Y$ *such that the following holds to an arbitrarily good approximation:*[1]

Permutation invariance

Latent space

$$\boxed{f(\{x_1, \dots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)}$$

General parametrization for a function of sets

Deep Sets for Particle Jets

[PTK, Metodiev, Thaler, 1810.05165]

Particle Flow Network (PFN)

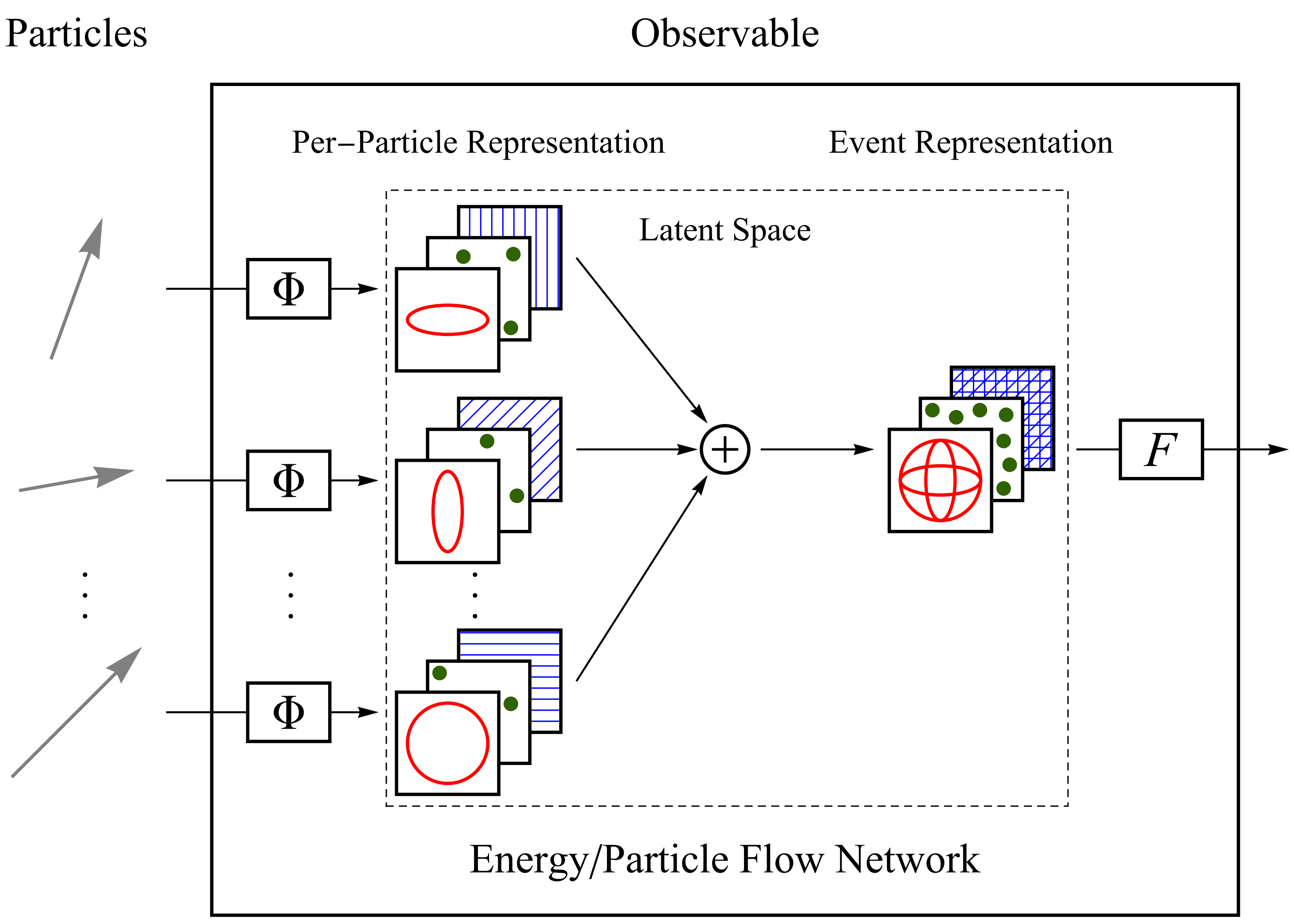
$$\mathrm{PFN}(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M}\Phi(p_i^\mu)\right)$$

Fully general latent space

Energy Flow Network (EFN)

$$\mathrm{EFN}(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M} z_i\,\Phi(\hat{p}_i)\right)$$

Energy-weighted (safe) latent space



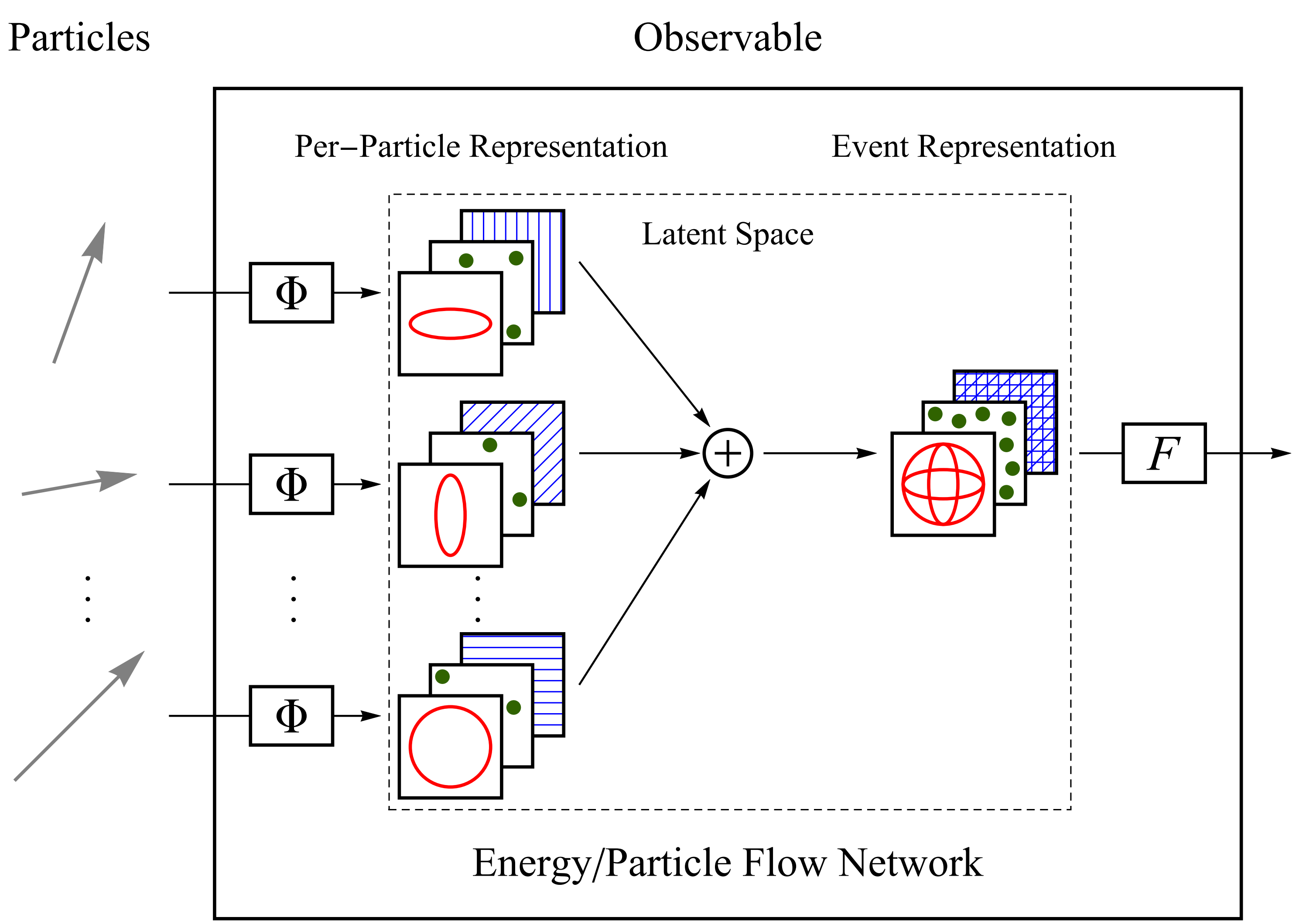
Approximating Φ and *F* with Neural Networks

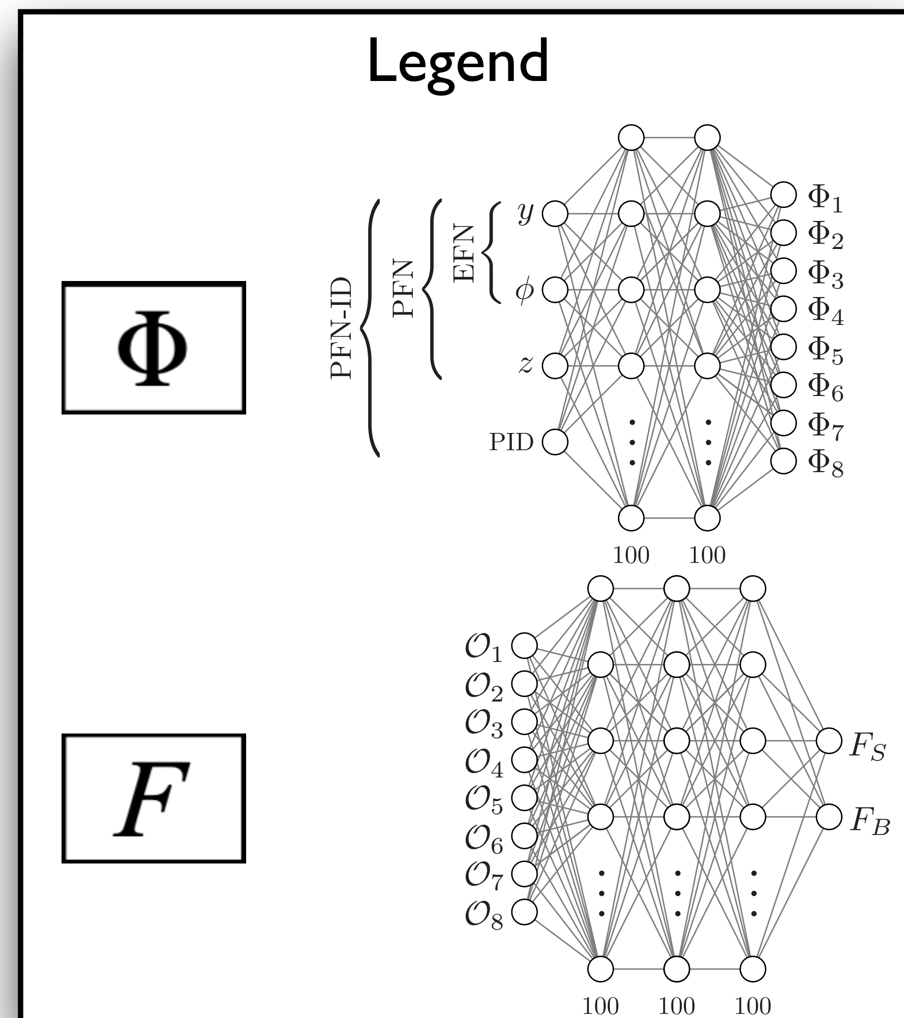
Employ neural networks as arbitrary function approximators

Use fully-connected networks for simplicity

Graph CNNs also interesting (see H. Qu's talk at ML4Jets)

Default sizes – Φ: (100, 100, ℓ), *F*: (100, 100, 100)

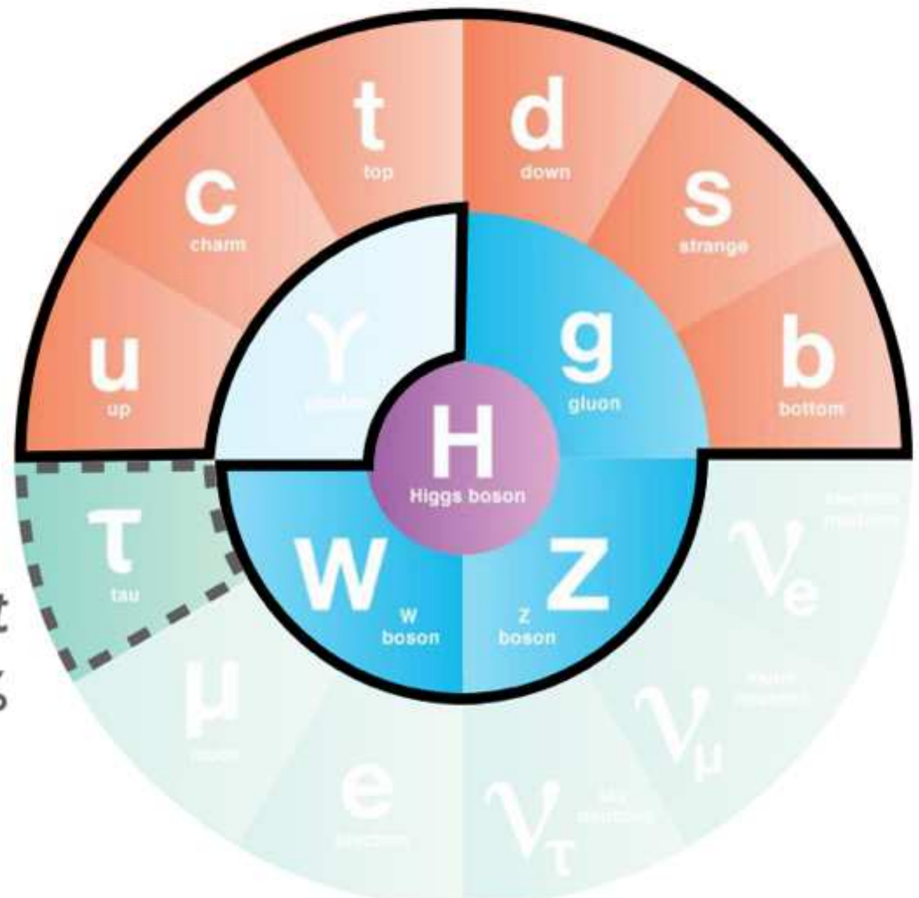




$$\text{EFN}: \quad \mathcal{O}_a = \sum_{i=1}^{M} z_i \Phi_a(y_i, \phi_i)$$

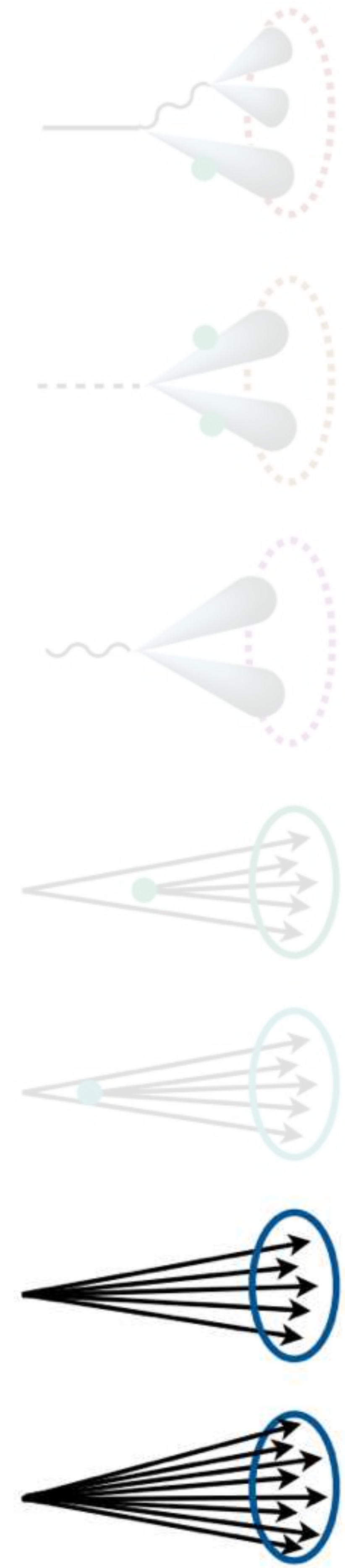
$$\text{PFN}: \quad \mathcal{O}_a = \sum_{i=1}^{M} \Phi_a(z_i, y_i, \phi_i, [\mathrm{PID}_i])$$

Jet Menu of the (Known) Universe



micro-jet
≈ 65%

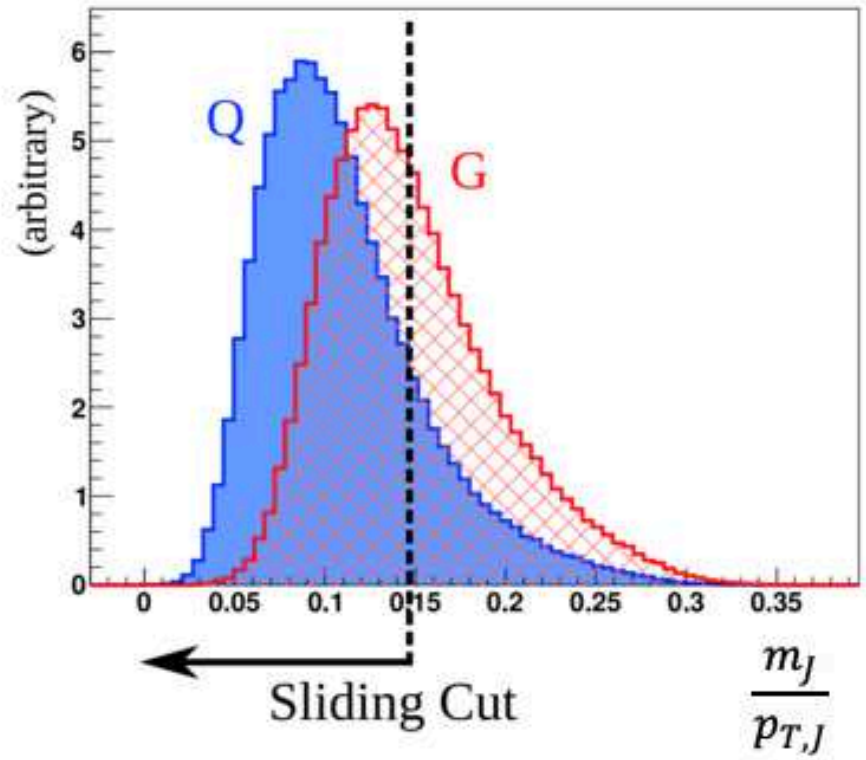
"Hello, world!" of jet classification



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

Quantifying a Classifier

Receiver Operating Characteristic (**ROC**) curve:
True negative rate of the classifier at different true positive rates



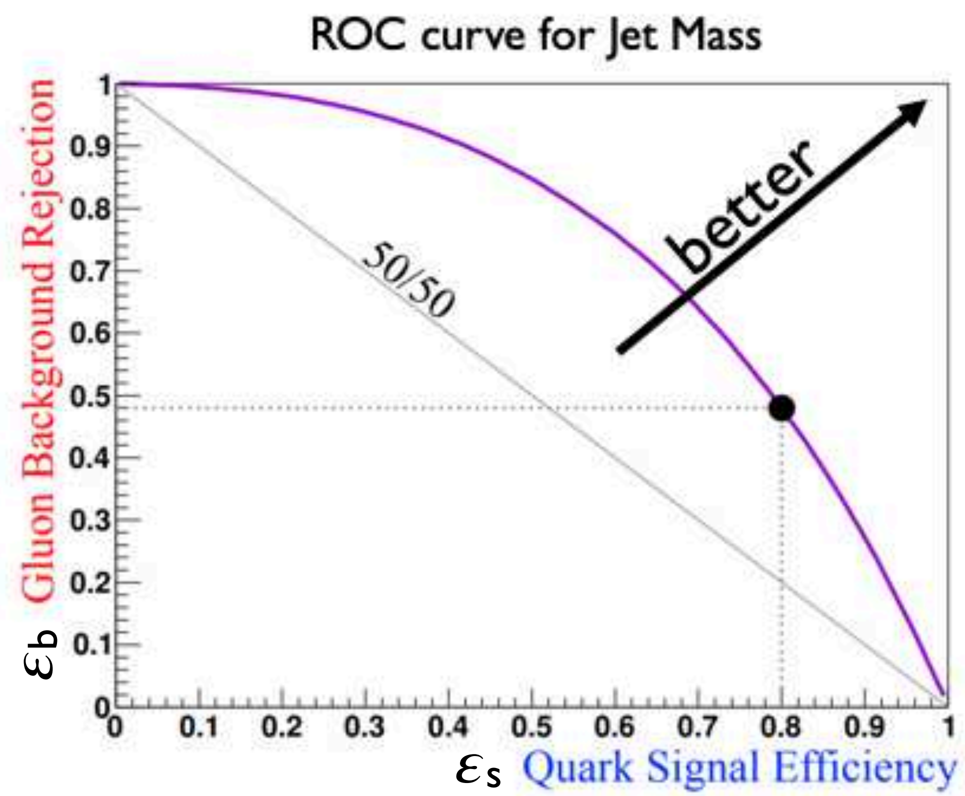


Figure from 1211.7038

Area Under the ROC Curve (**AUC**) captures the classifier performance in a number.

Other formats possible, e.g. $(\varepsilon_s,\, 1/(1-\varepsilon_b)),\ (\varepsilon_s,\, \varepsilon_s/\sqrt{1-\varepsilon_b})$

Quark vs. Gluon: Classification Performance

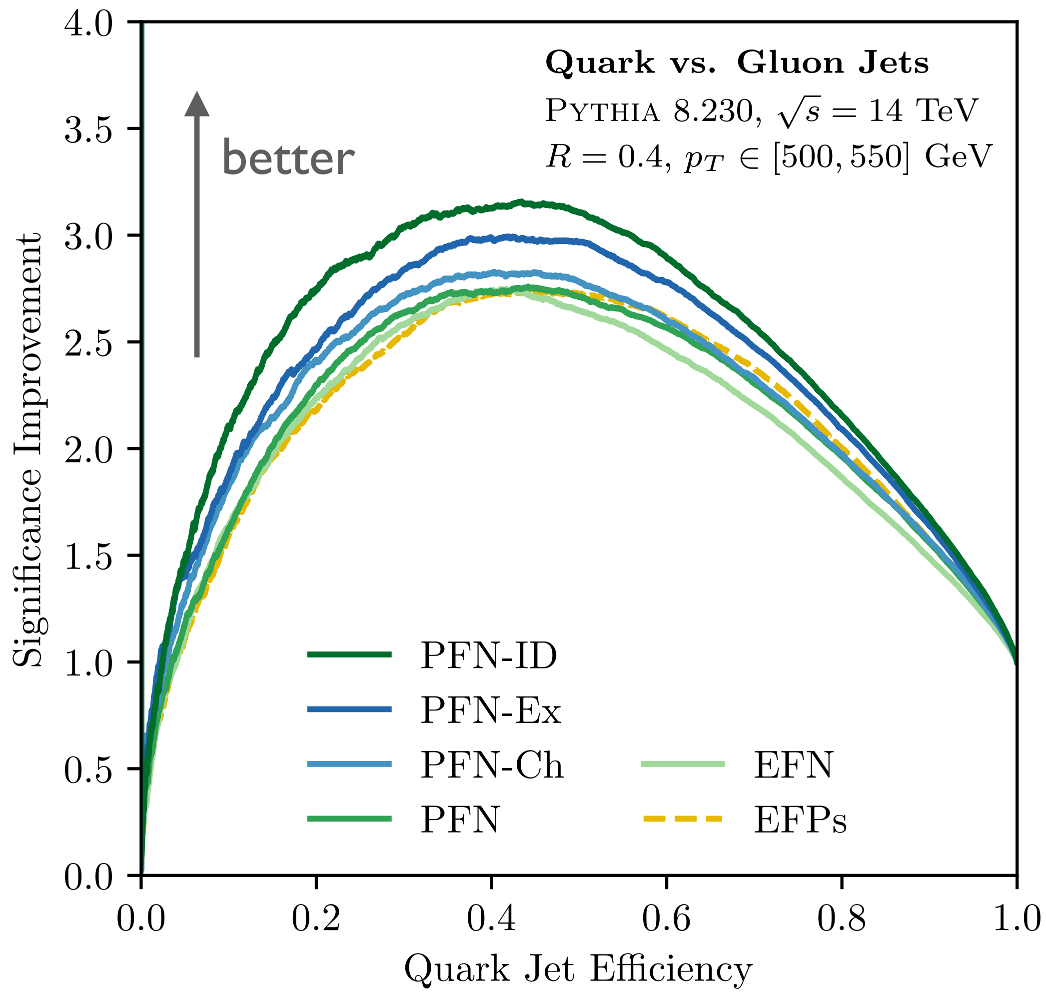
PFN-ID: Full particle flavor info
$(\gamma, \pi^\pm, K^\pm, K_L, p, \bar{p}, n, \bar{n}, e^\pm, \mu^\pm)$

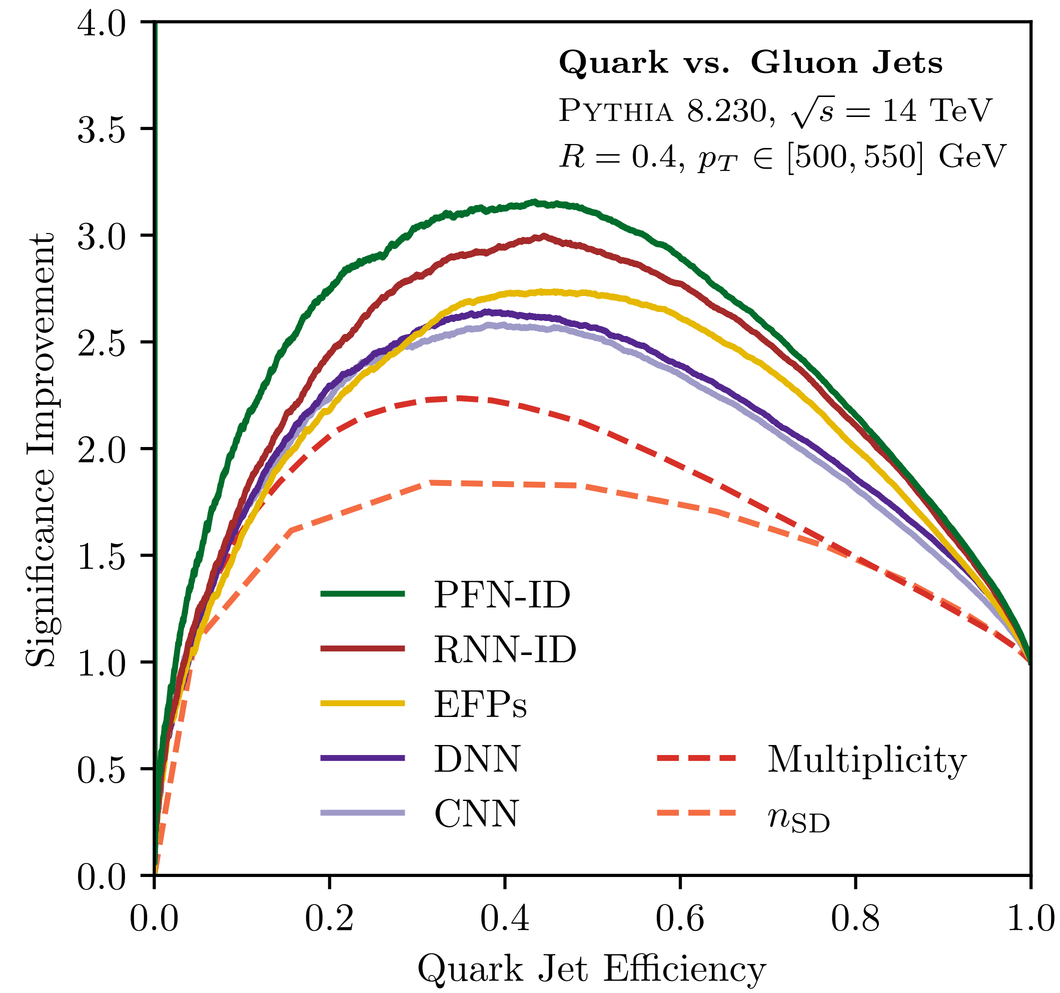
PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^\pm, \mu^\pm)$

PFN-Ch: Particle charge info
(+,0,−)

PFN: No particle type info, arbitrary energy dependence

EFN: Energy-weighted latent space





Latent space dimension $\ell = 256$

PFN-ID slightly better than RNN-ID

EFPs are comparable to EFN

Quark vs. Gluon: EFN Latent Dimension Sweep

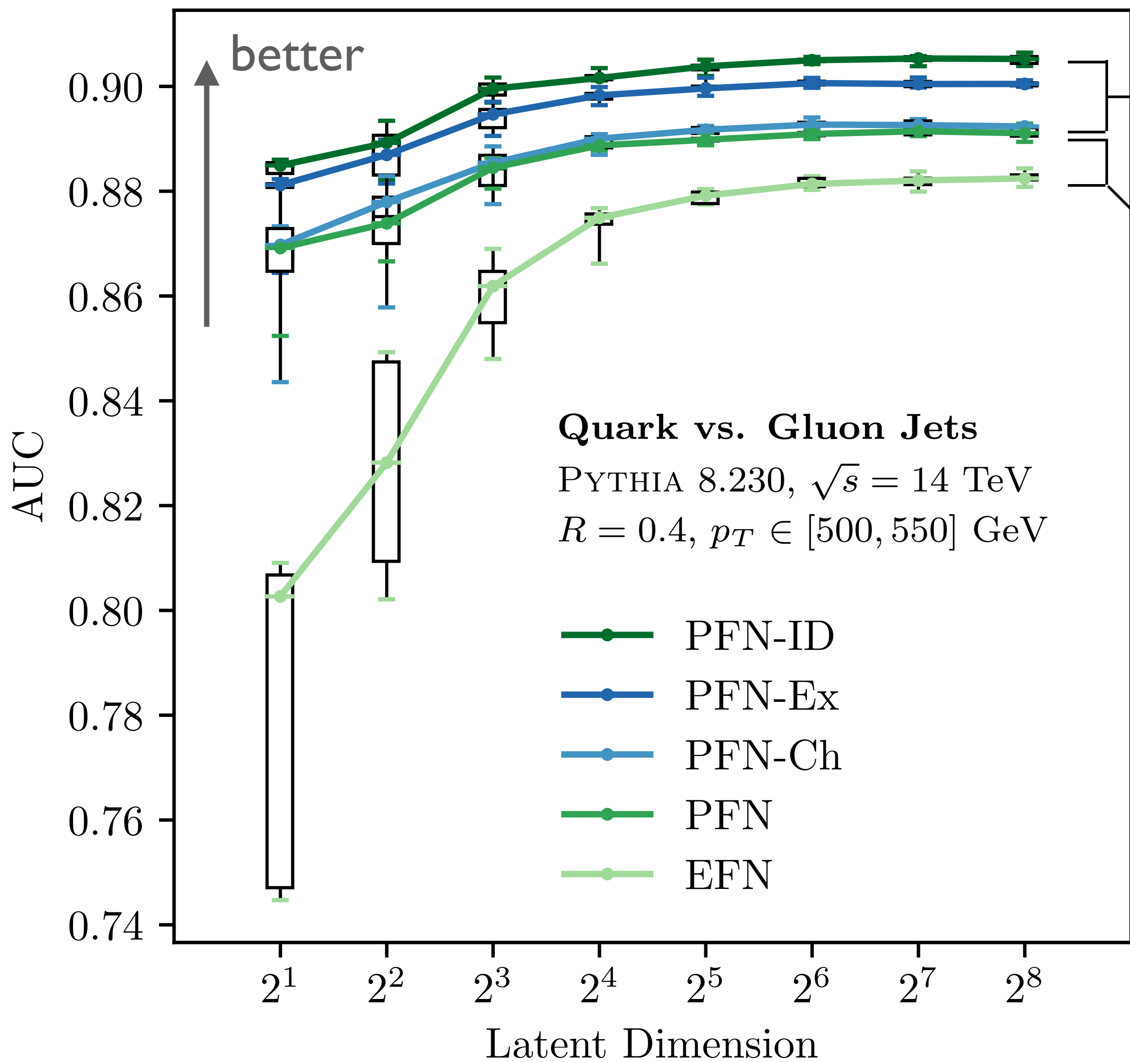
PFN-ID: Full particle flavor info
$(\gamma, \pi^\pm, K^\pm, K_L, p, \bar{p}, n, \bar{n}, e^\pm, \mu^\pm)$

PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^\pm, \mu^\pm)$

PFN-Ch: Particle charge info
$(+,0,-)$

PFN: No particle type info, arbitrary energy dependence

EFN: Energy-weighted latent space

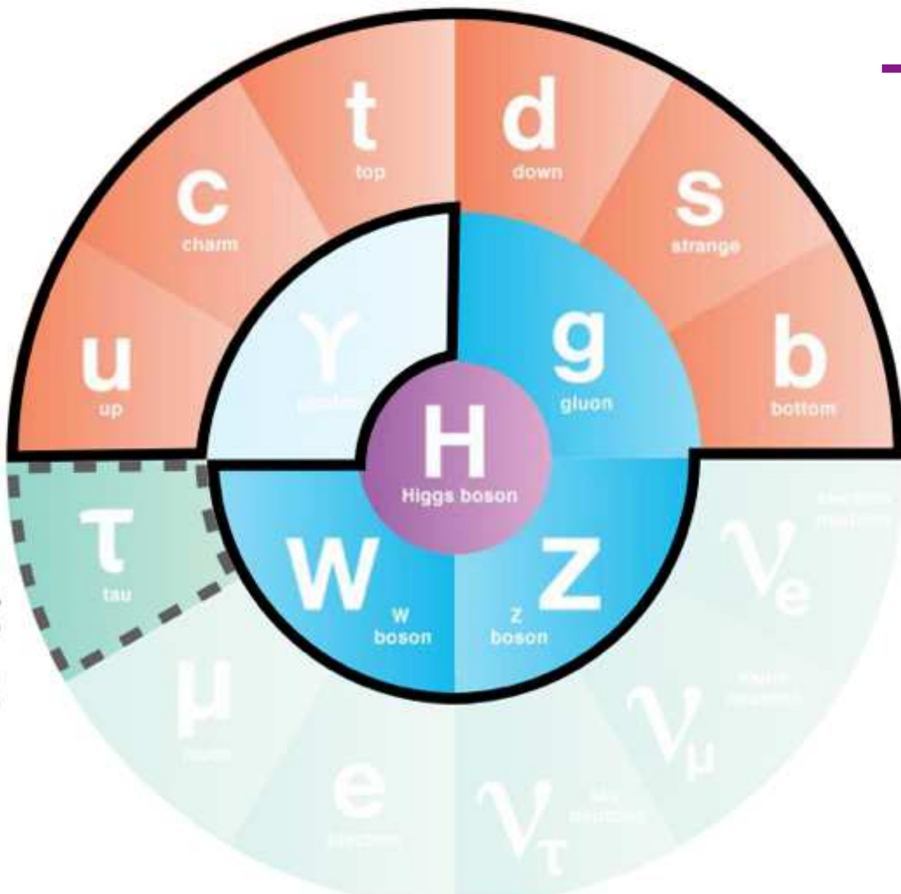


Particle type info helpful

Unsafe information clearly helpful

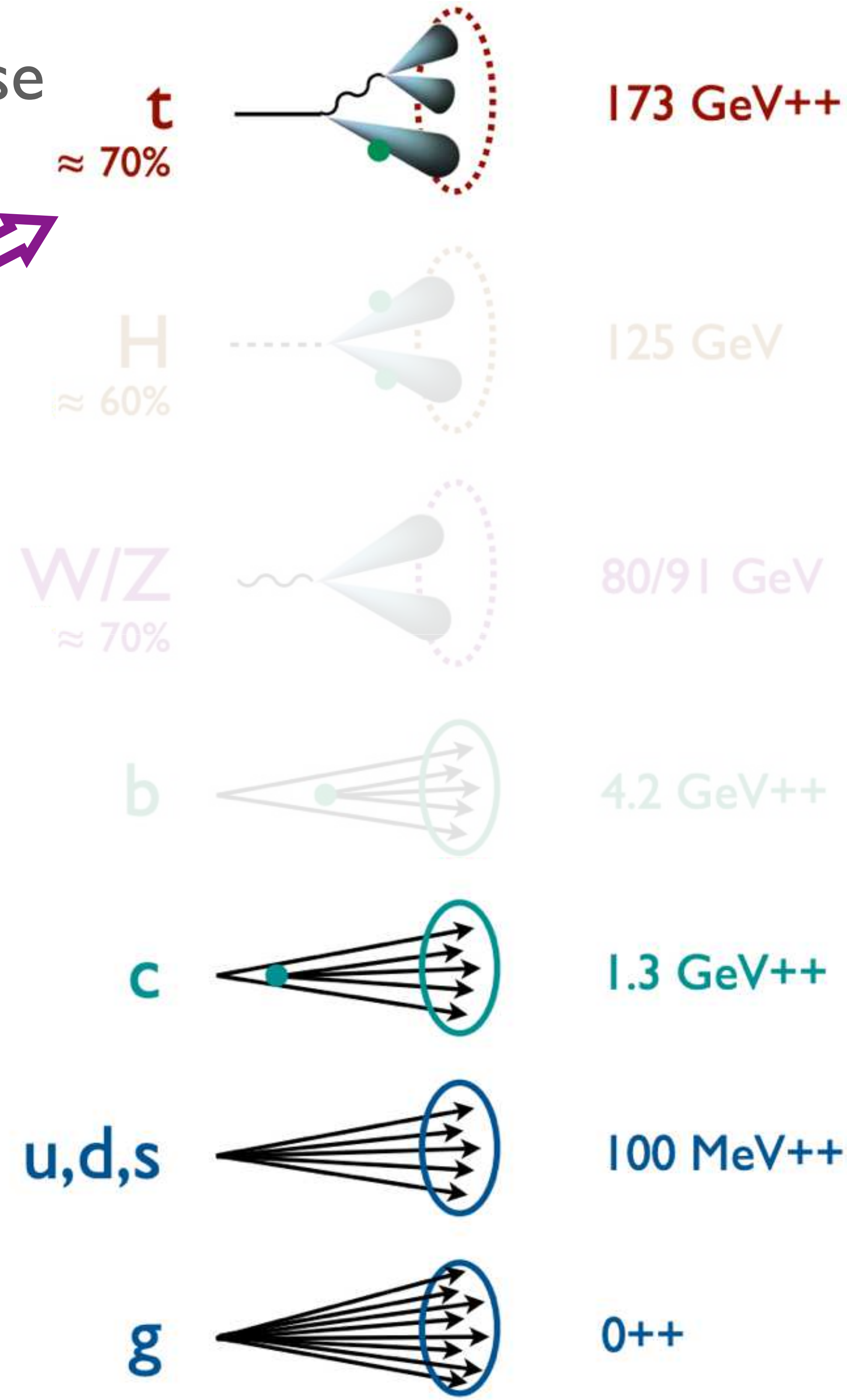
Performance saturates as dimension increases

Jet Menu of the (Known) Universe



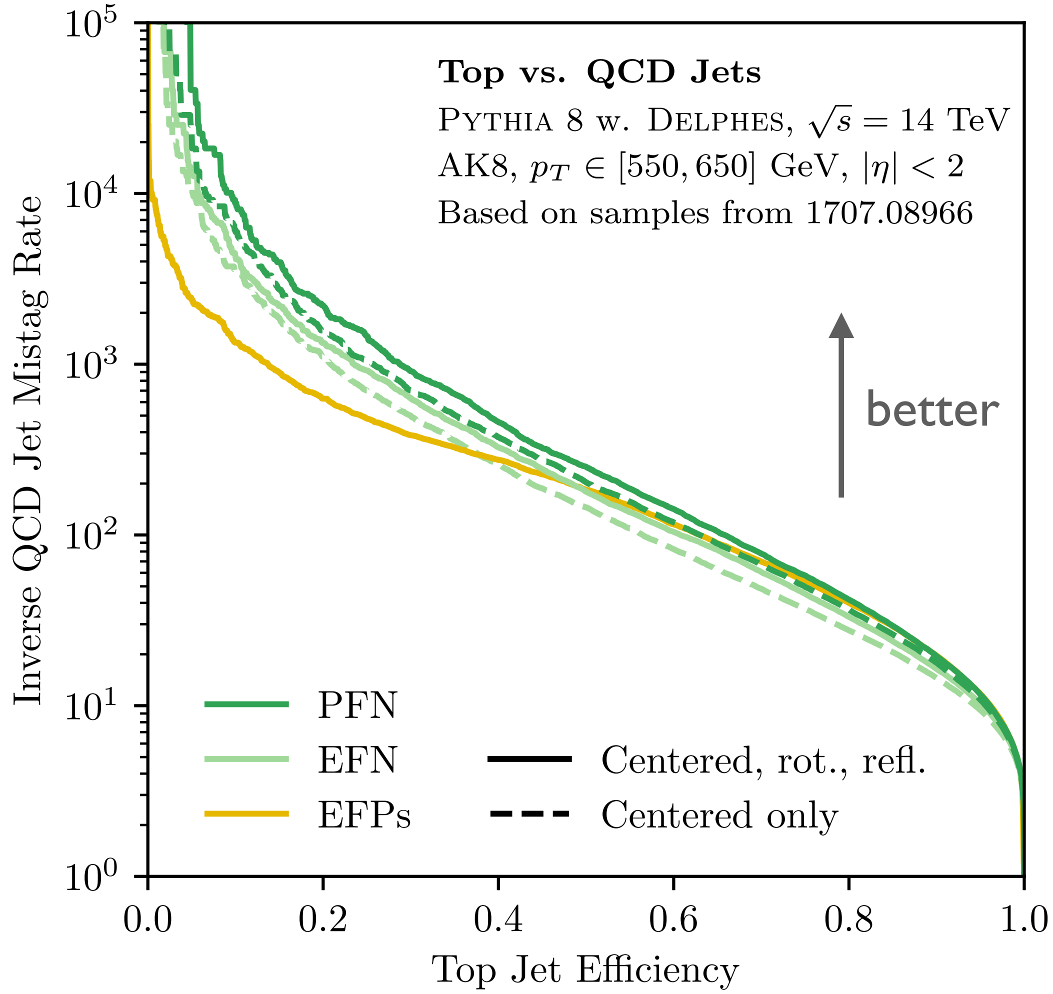
micro-jet
≈ 65%

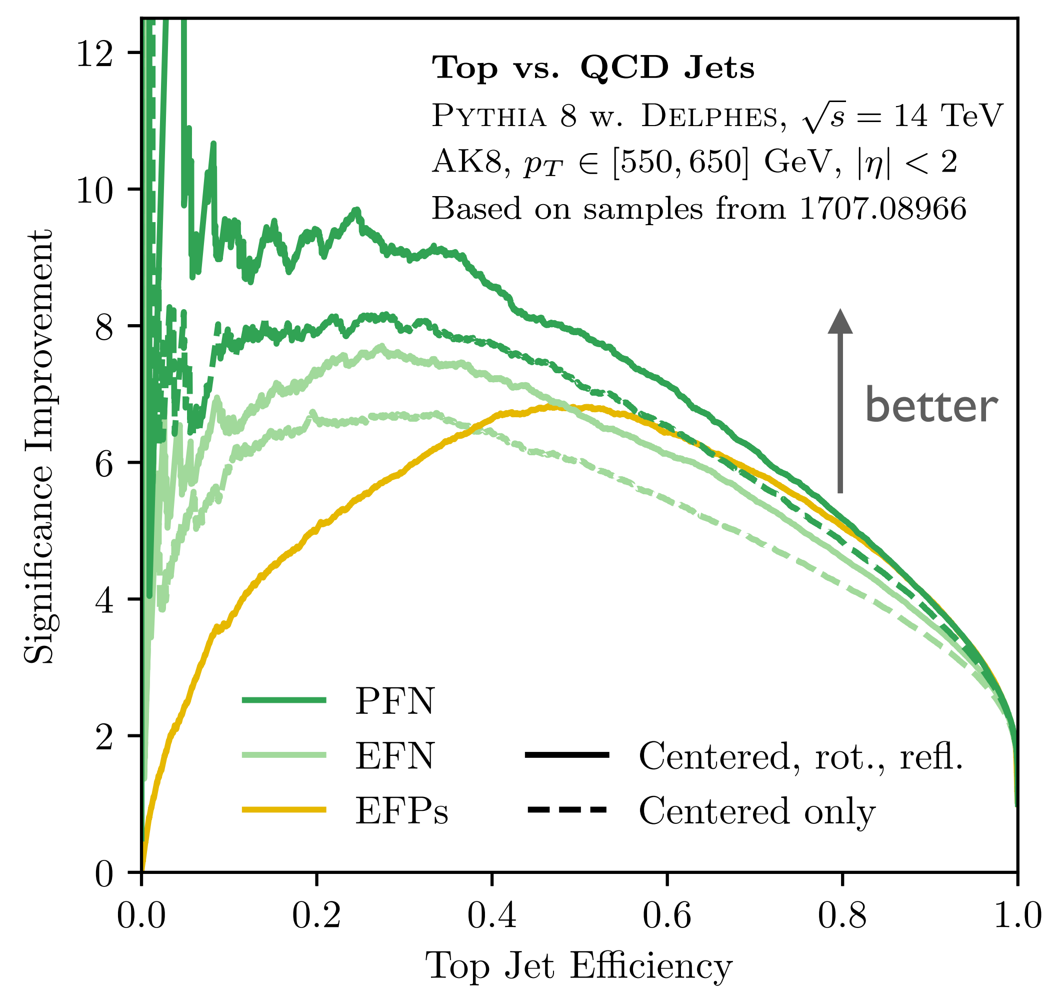
Top



QCD

Boosted Top: Classification Performance



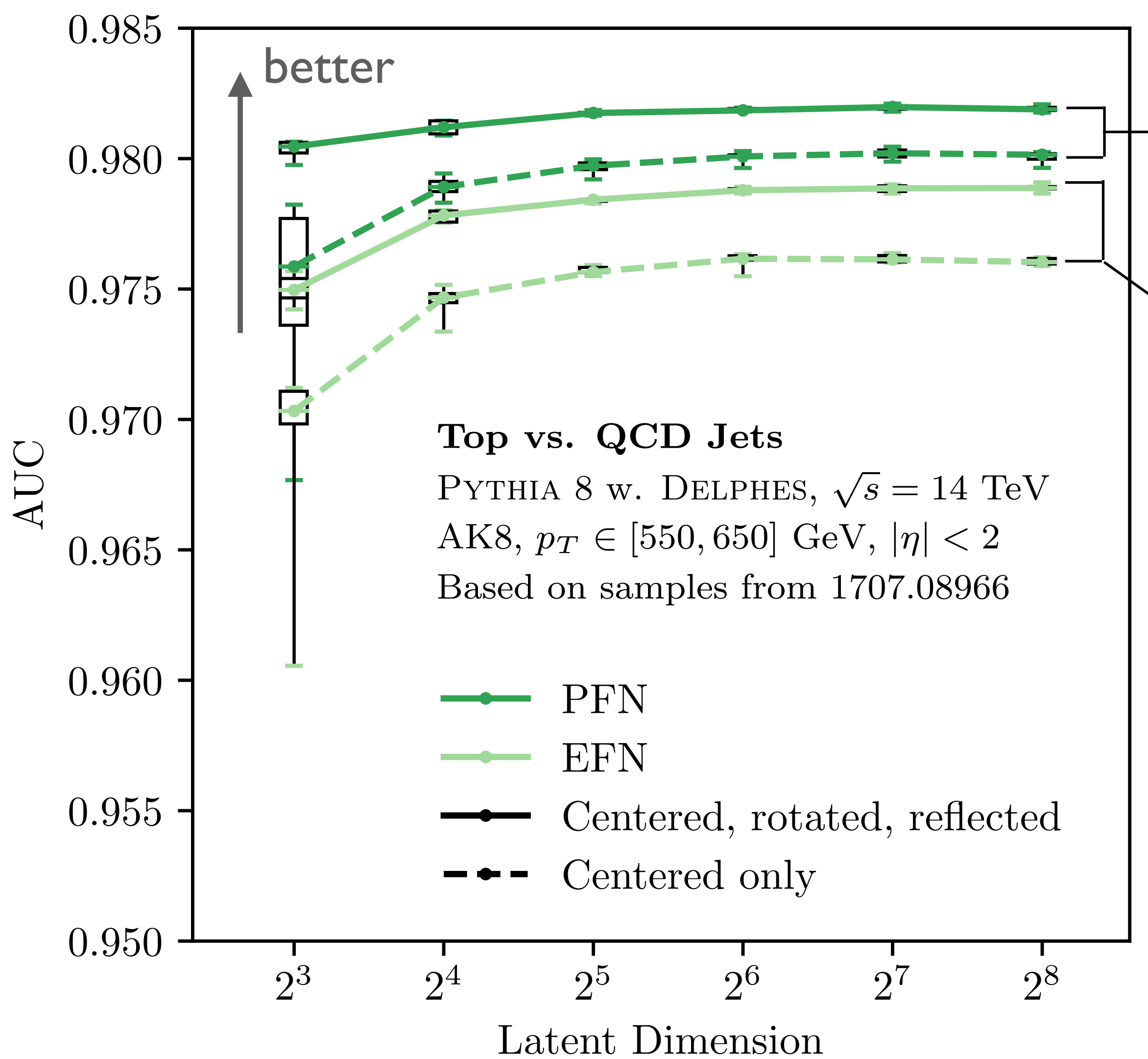


Latent space dimension $\ell = 256$

EFN/PFN rotation and reflection preprocessing helpful

EFPs are comparable to EFN and even better at high signal efficiency

Boosted Top: EFN Latent Dimension Sweep



Preprocessing clearly helpful

Unsafe information clearly helpful

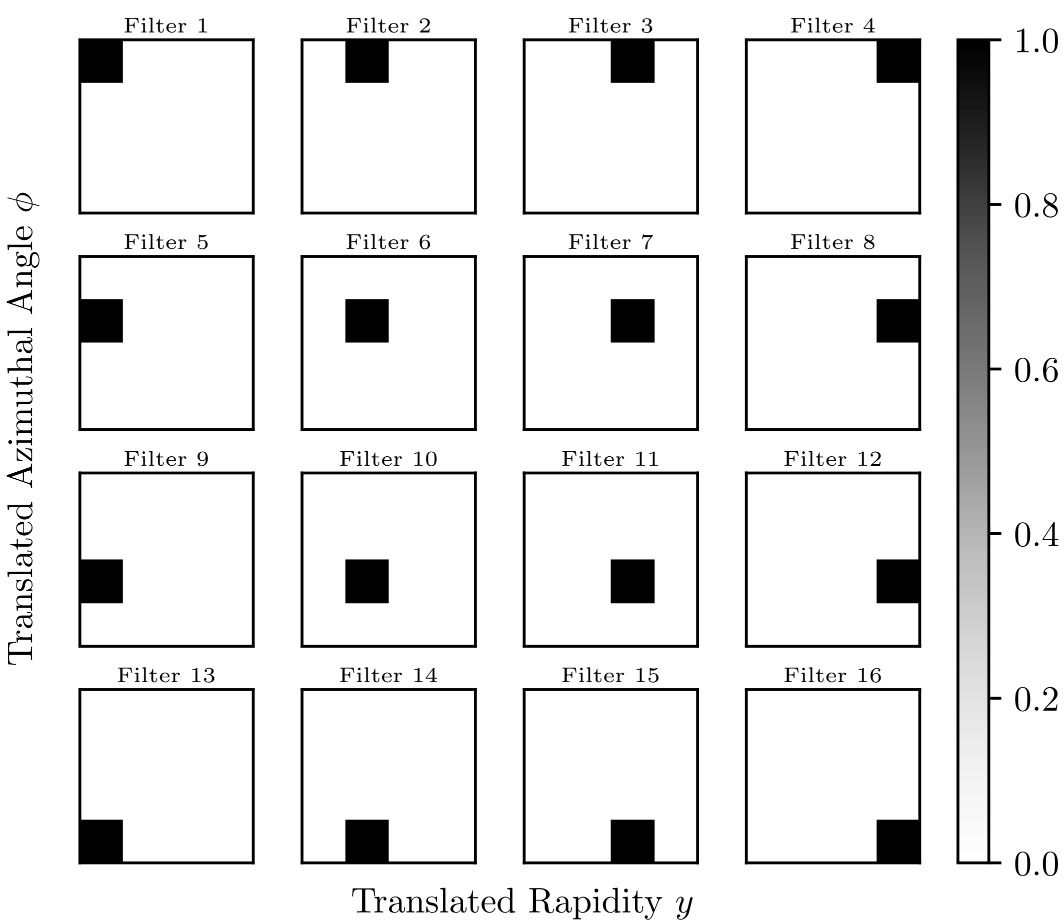
Performance quickly saturates as latent dimension increases

Energy Flow Network Visualization

EFN observables are two-dimensional geometric functions

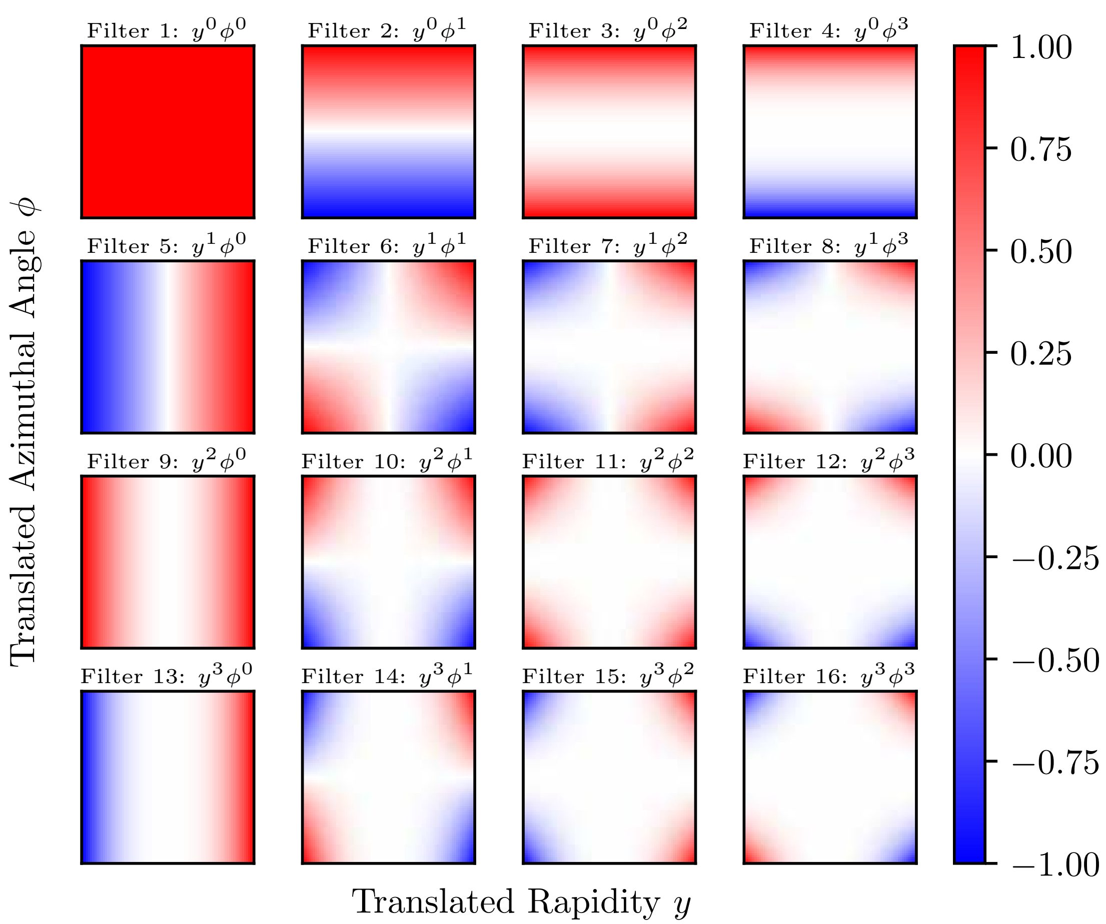
Visualize EFN observables as *filters* in the translated rapidity-azimuth plane

Jet images as EFN filters



[Cogan, Kagan, Strauss, Schwartzman, 2014]
[de Oliviera, Kagan, Mackey, Nachman, Schwartzman, 2015]

Moments as EFN filters



[Donoghue, Low, Pi, 1979]
[Gur-Ari, Papucci, Perez, 2011]

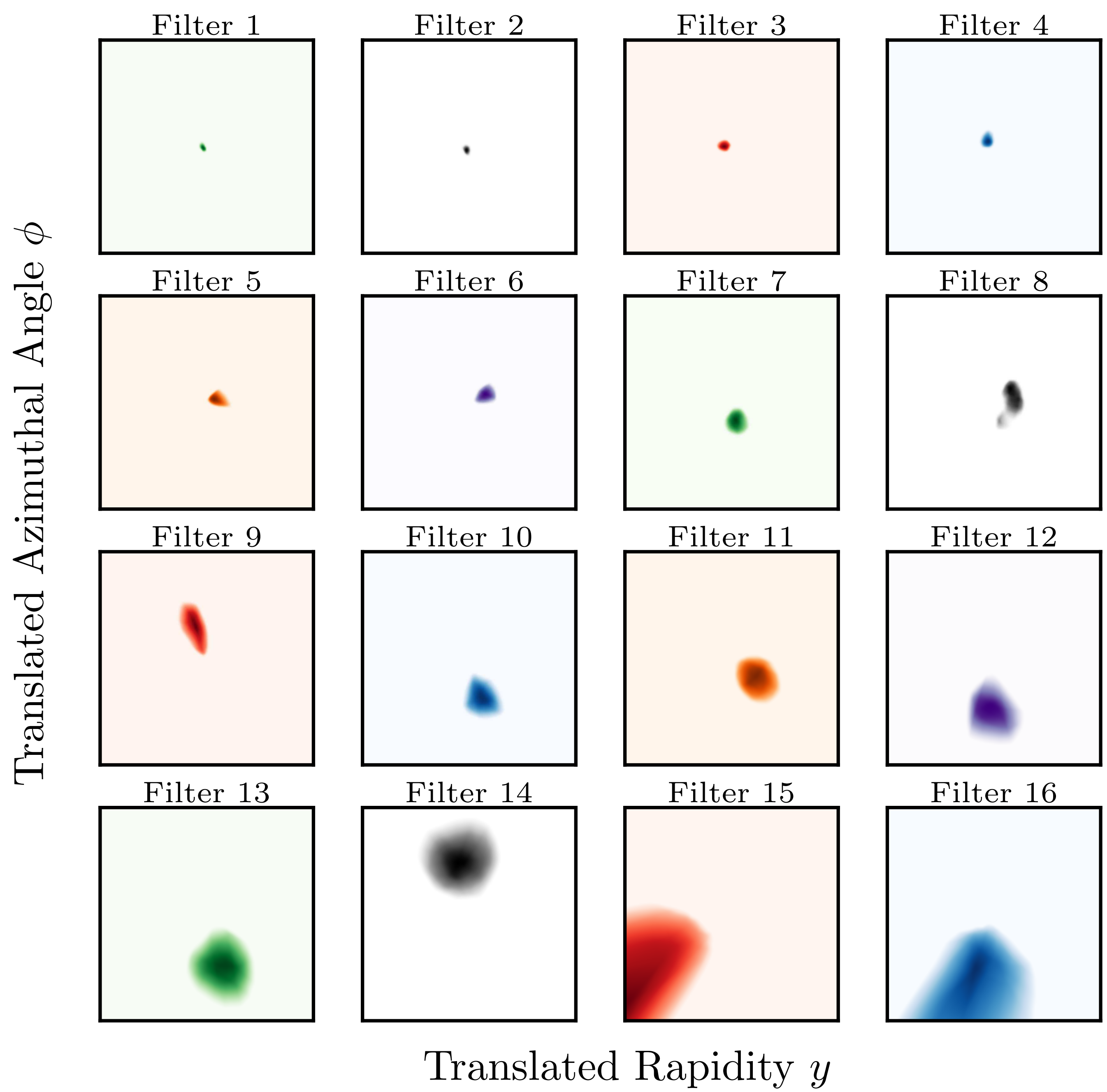
Quark vs. Gluon: Visualizing EFN Filters

Generally see blobs of all scales

Local nature of activated region lends interpretation as "pixels"

EFN seems to have learned a dynamically sized jet image

EFN (ℓ = 256) randomly selected filters, sorted by size



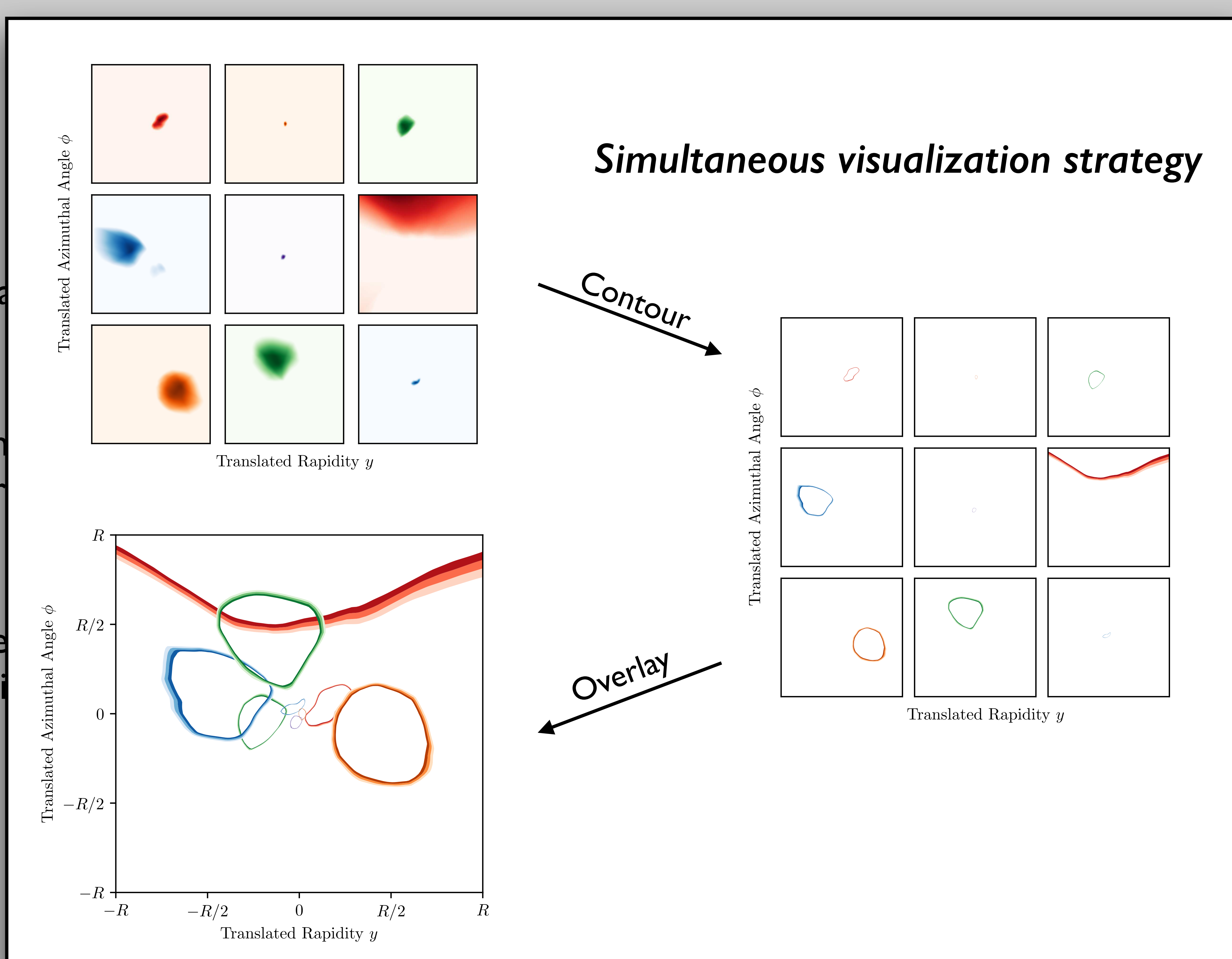
Quark vs. Gluon: Visualizing EFN Filters

by size

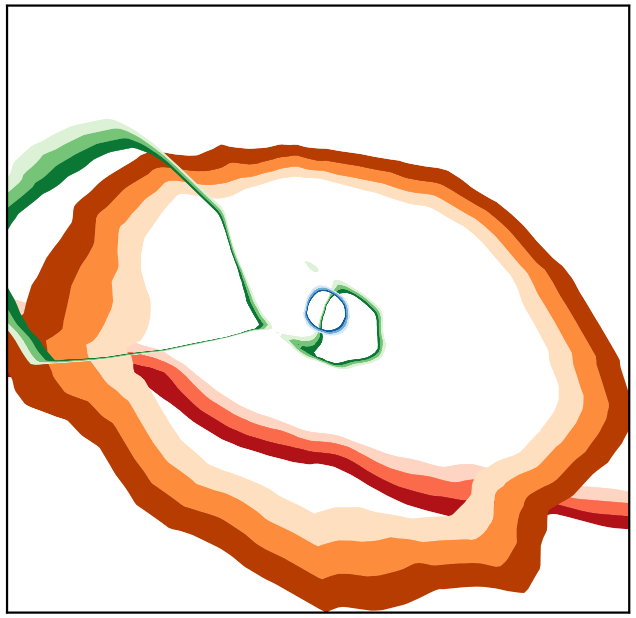
Genera

Local n
interpr

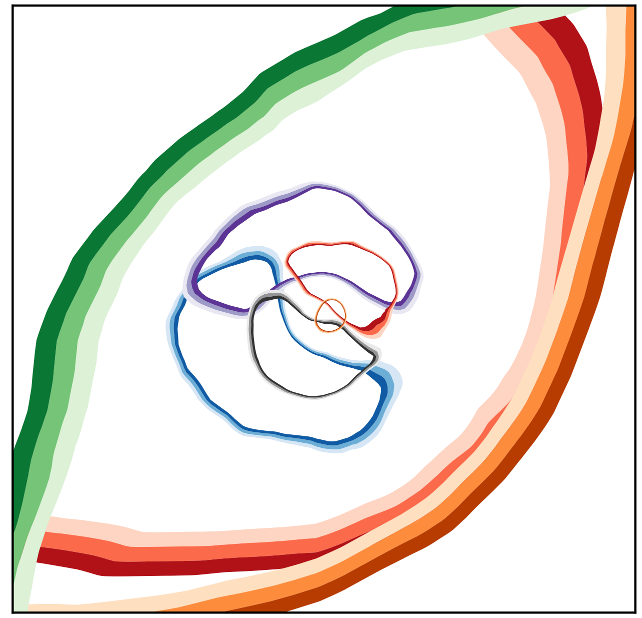
EFN se
dynami



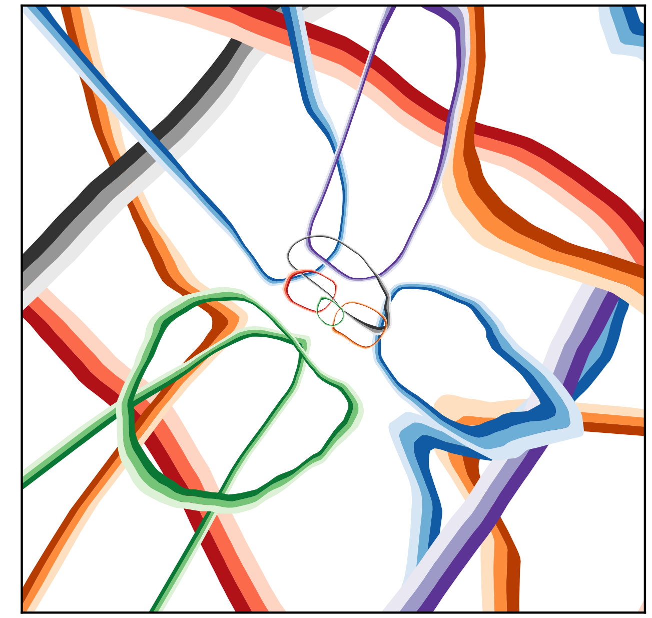
Quark vs. Gluon: Visualizing EFN Filters



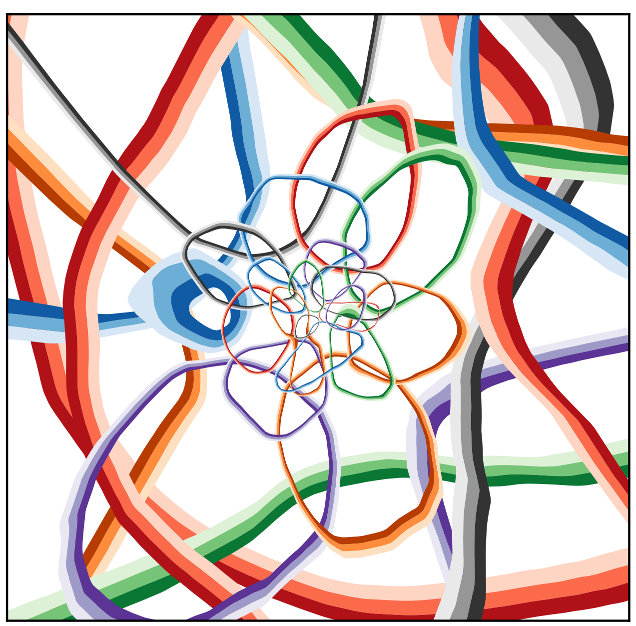
$\ell = 4$



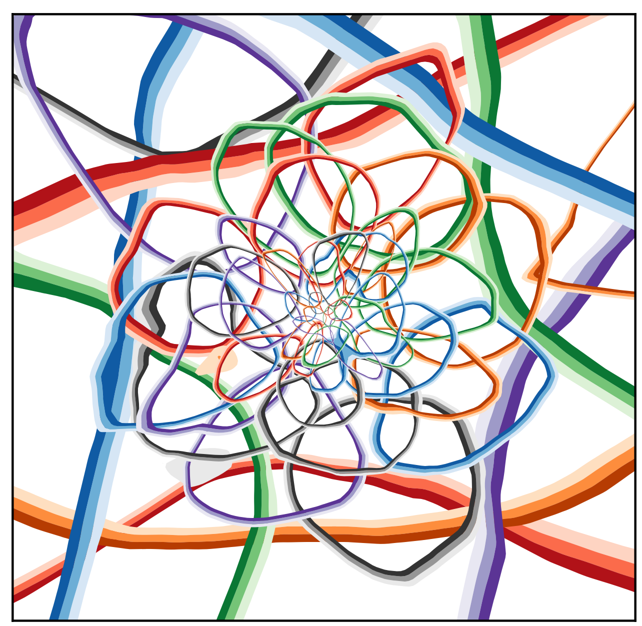
$\ell = 8$



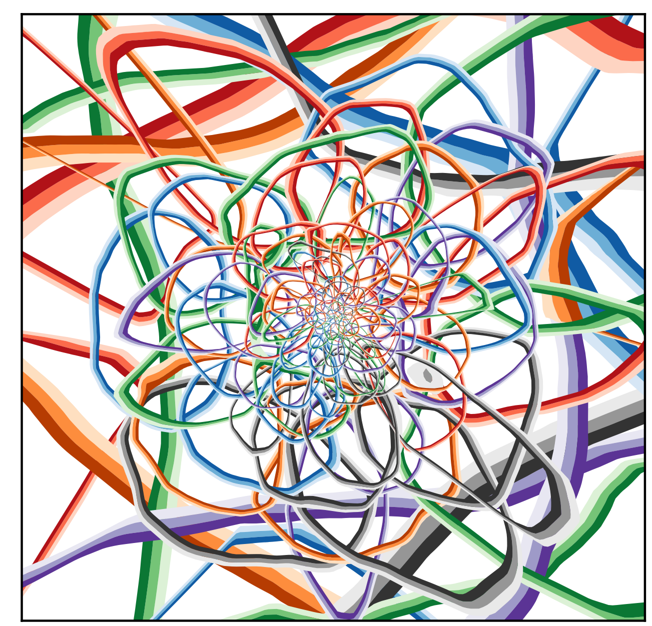
$\ell = 16$



$\ell = 32$

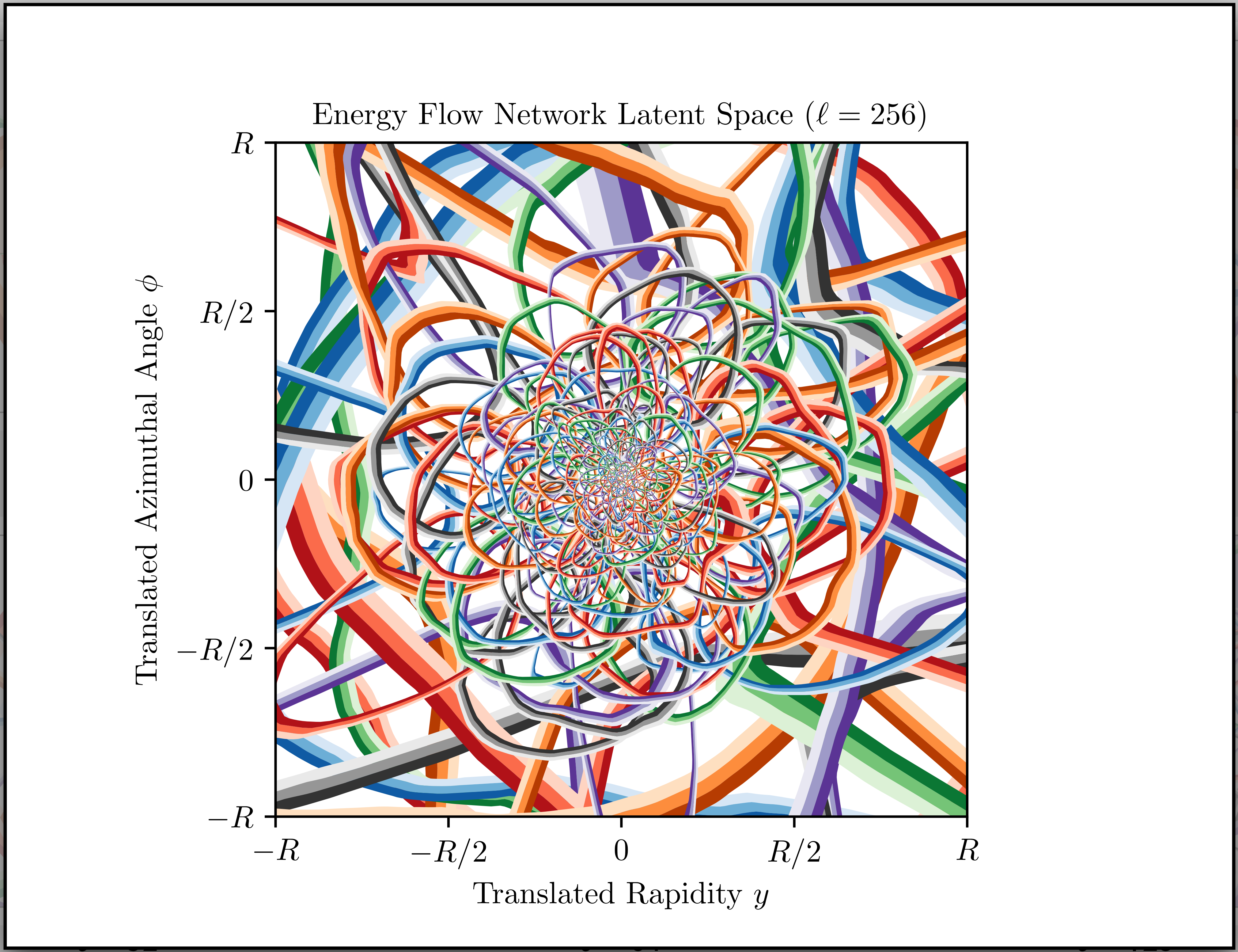


$\ell = 64$

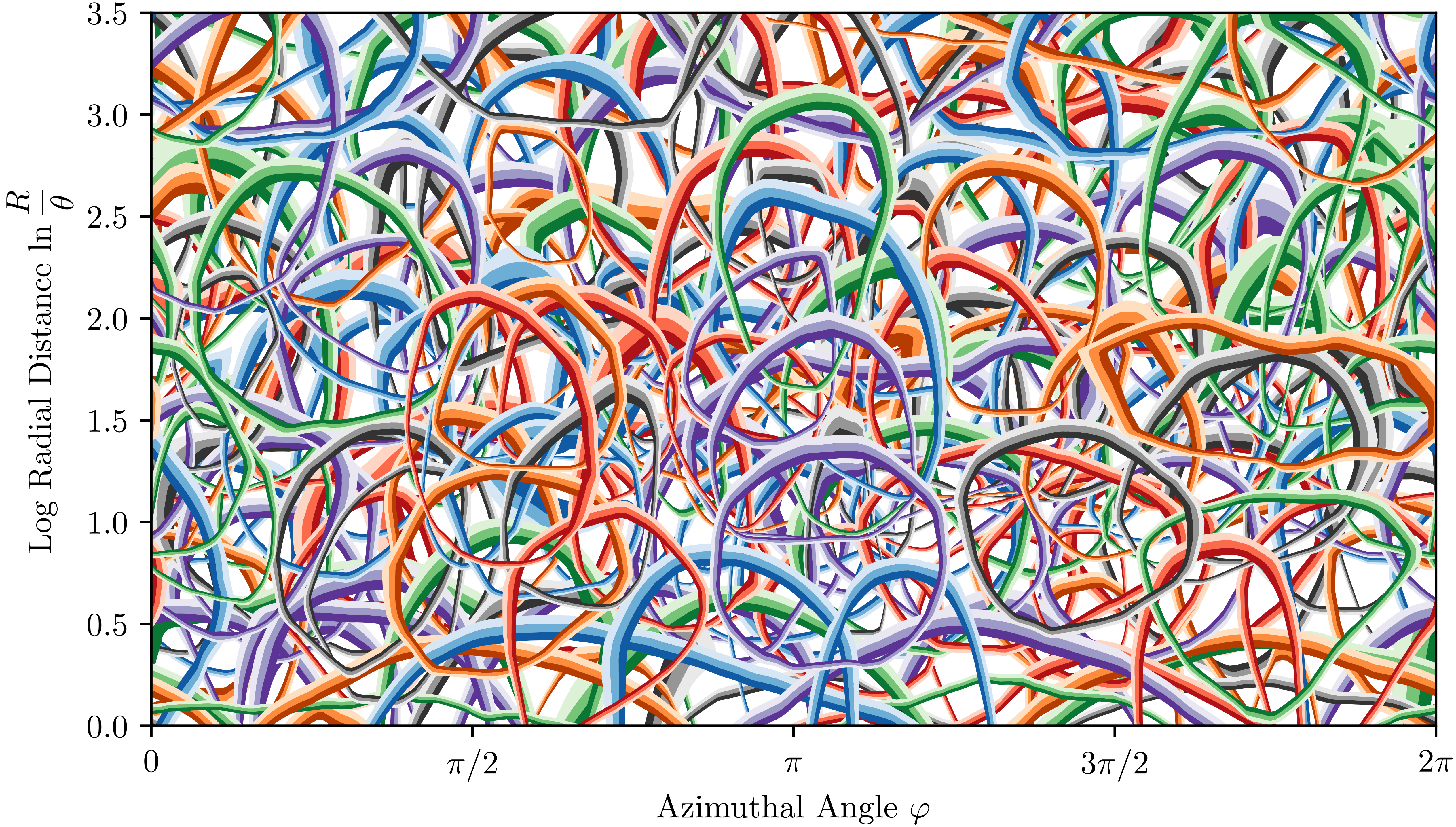


$\ell = 128$

Quark vs. Gluon: Visualizing EFN Filters



Quark vs. Gluon: Visualizing EFN Filters in the Emission Plane



Quark vs. Gluon: Measuring EFN Filters

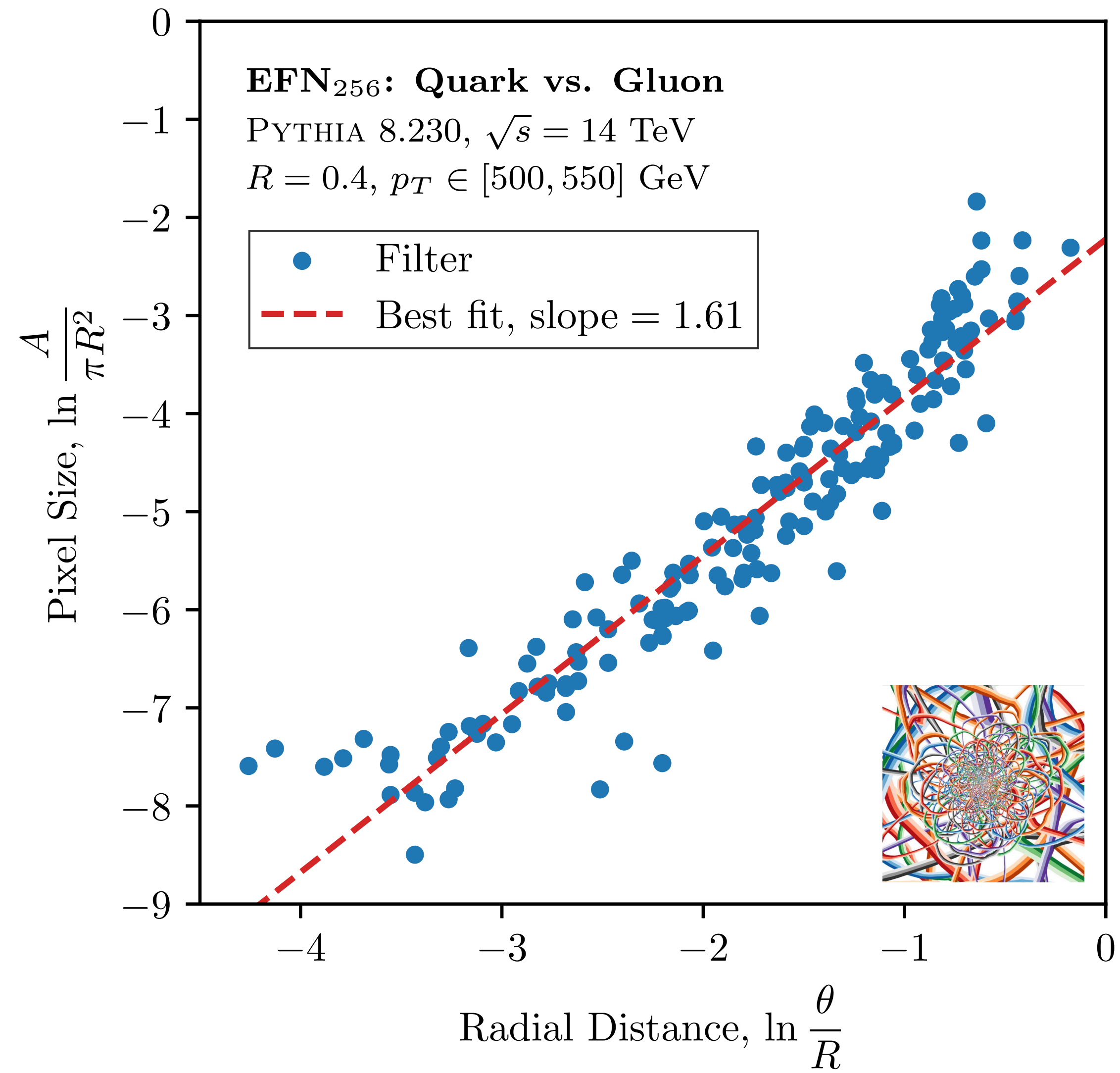
Power-law dependence between filter size and distance from center is observed

Slope of 2 is predicted via a simplified calculation

$$\left|\boxed{\mathrm{d}\ln\frac{\theta}{R}\,\mathrm{d}\varphi}\right| = \theta^2\,\boxed{\mathrm{d}y\,\mathrm{d}\phi}$$

Area element in rap-phi plane

Emission plane area element

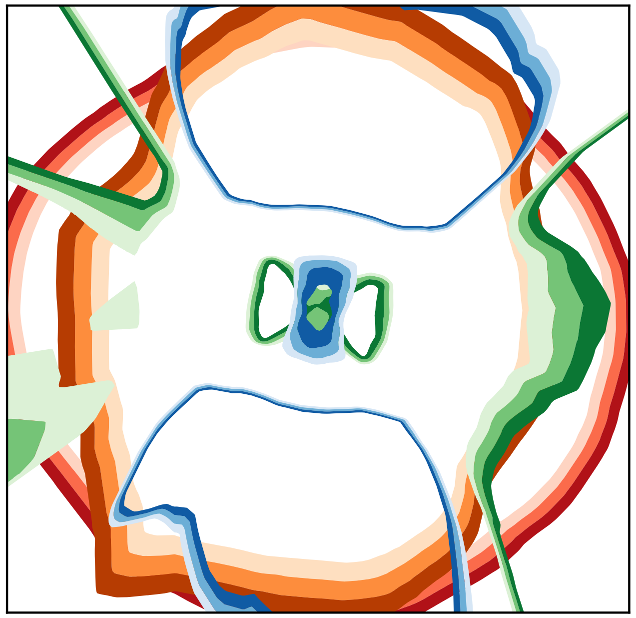


Non-perturbative physics, axis recoil, higher order effects cause deviations from slope of 2

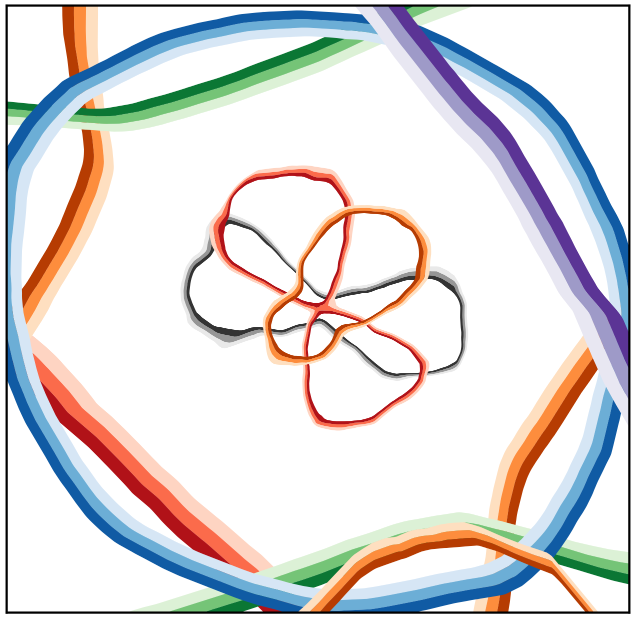
$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta}\frac{dz}{z}$$

Boosted Top: Visualizing EFN Filters

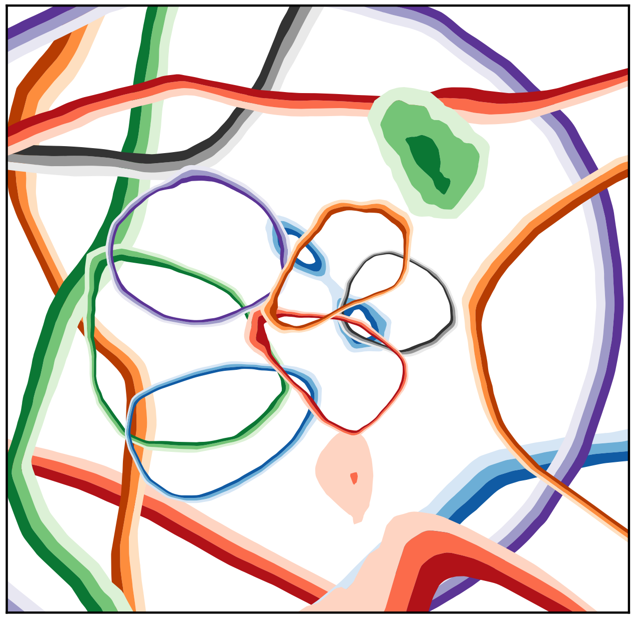
Without rotation/reflection preprocessing



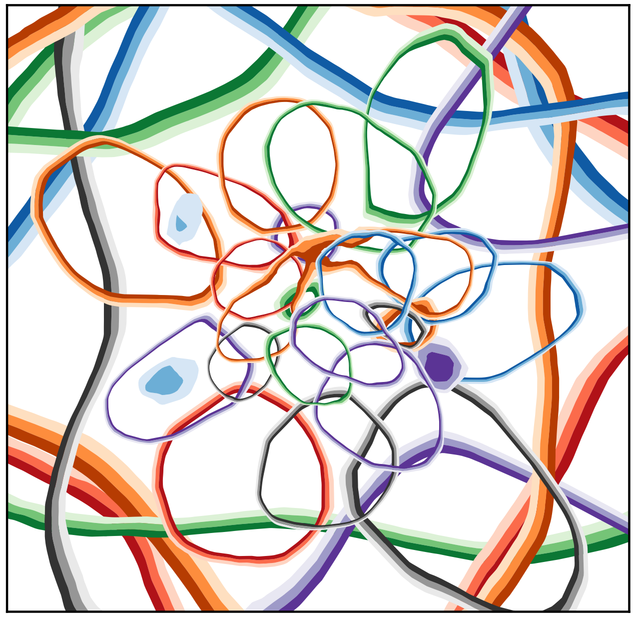
$\ell = 4$



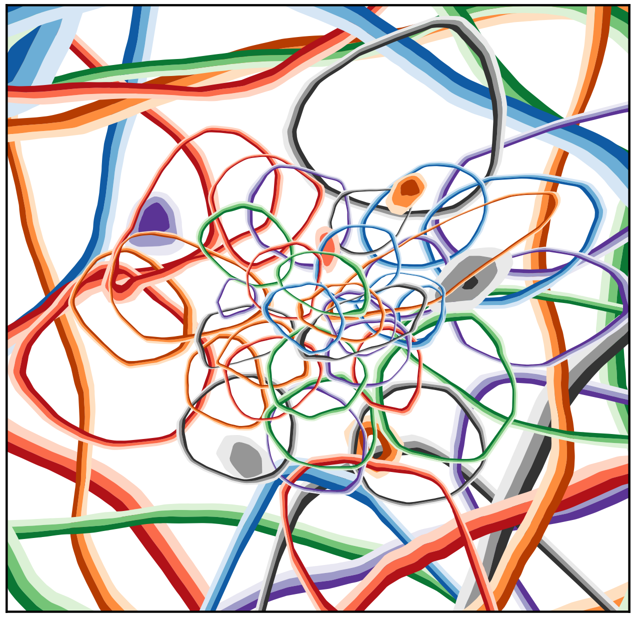
$\ell = 8$



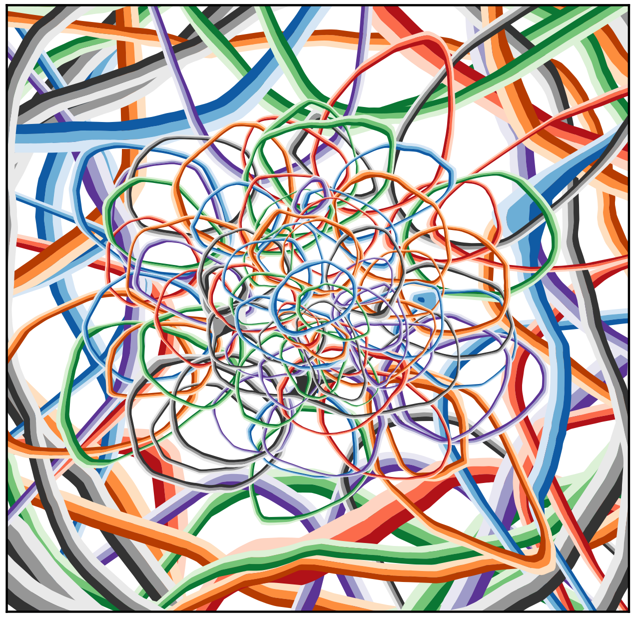
$\ell = 16$



$\ell = 32$



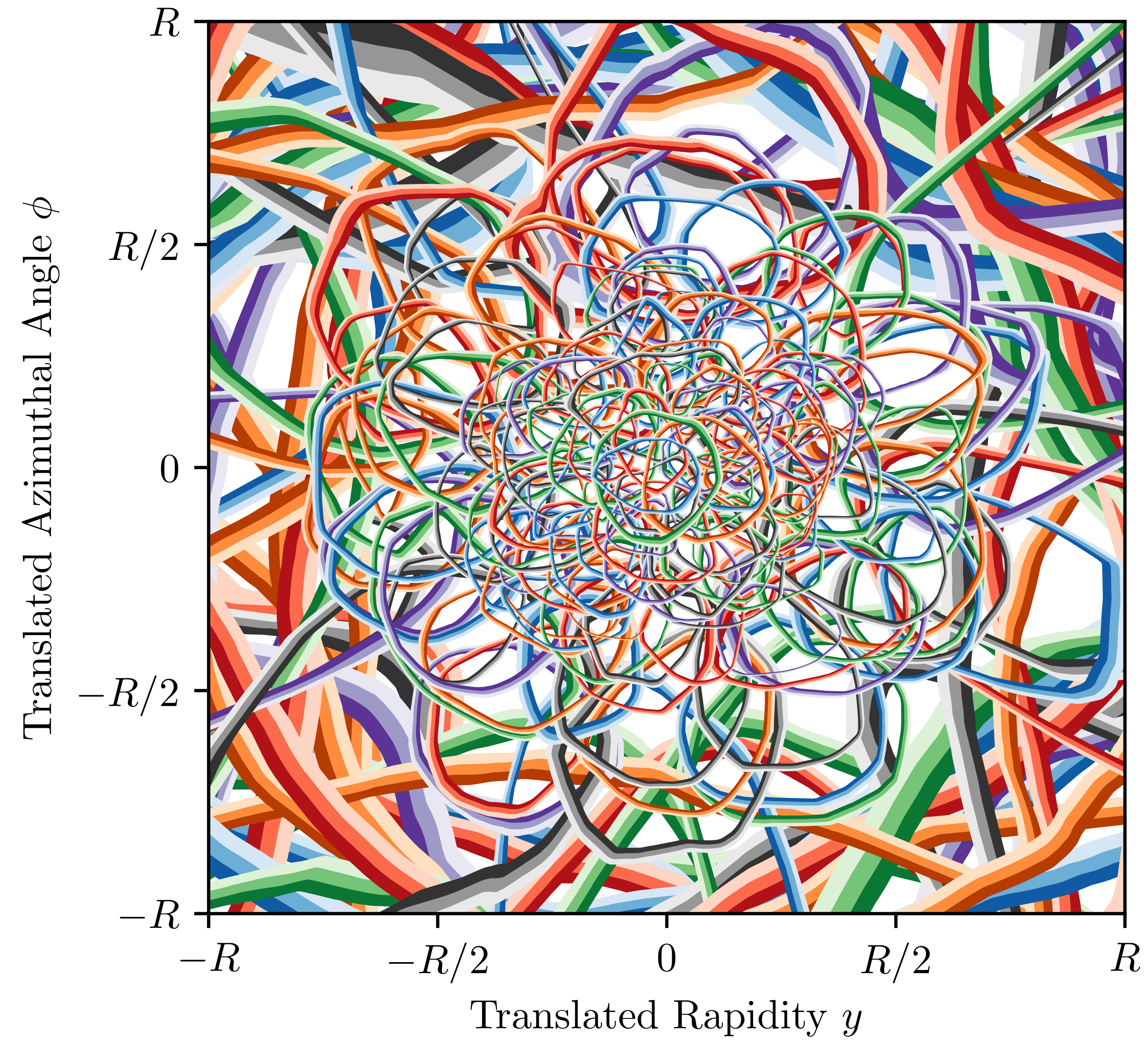
$\ell = 64$



$\ell = 128$

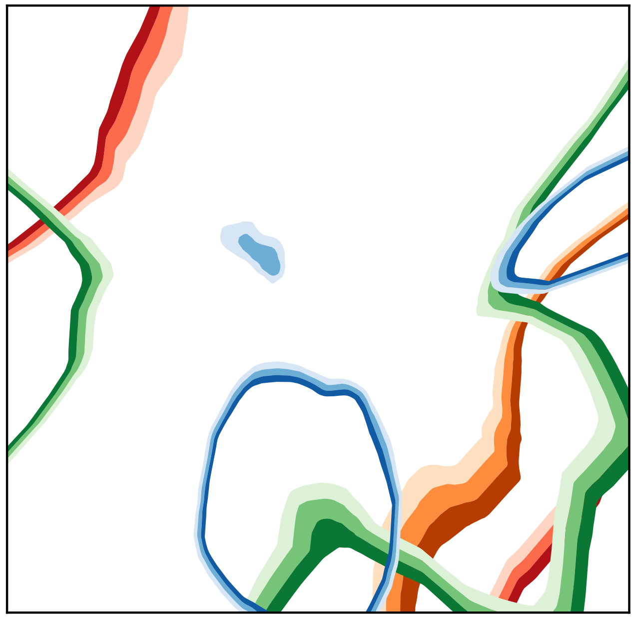
Boosted Top: Visualizing EFN Filters

Without rotation/reflection preprocessing

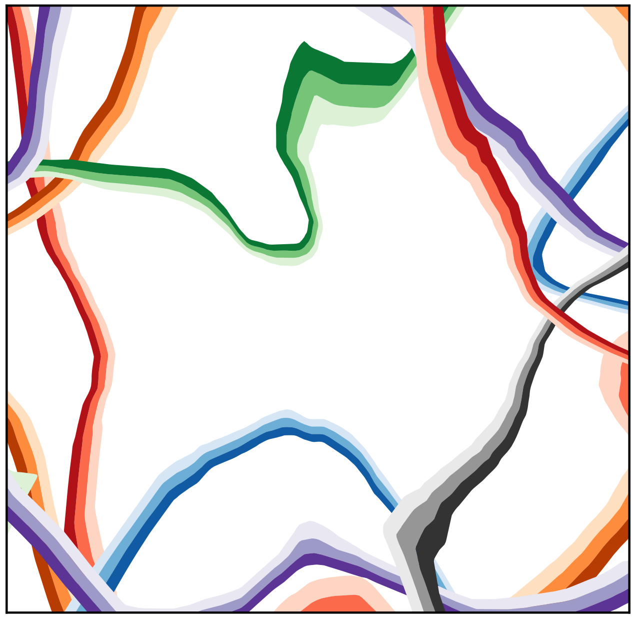


Boosted Top: Visualizing EFN Filters

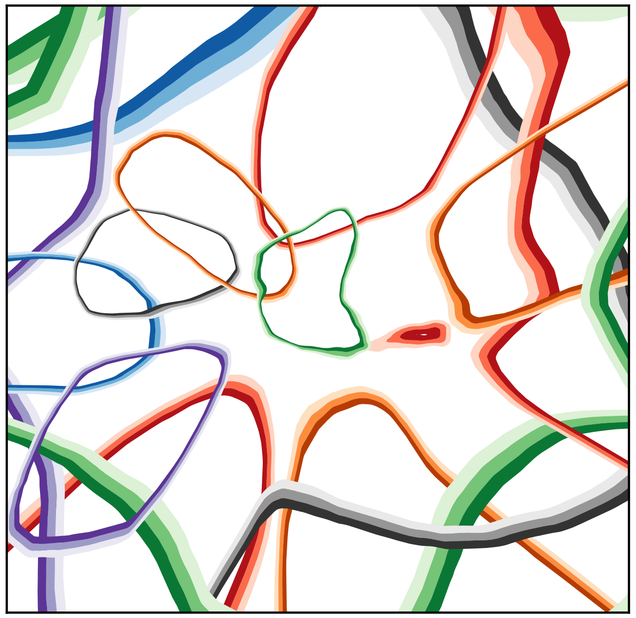
With rotation/reflection preprocessing



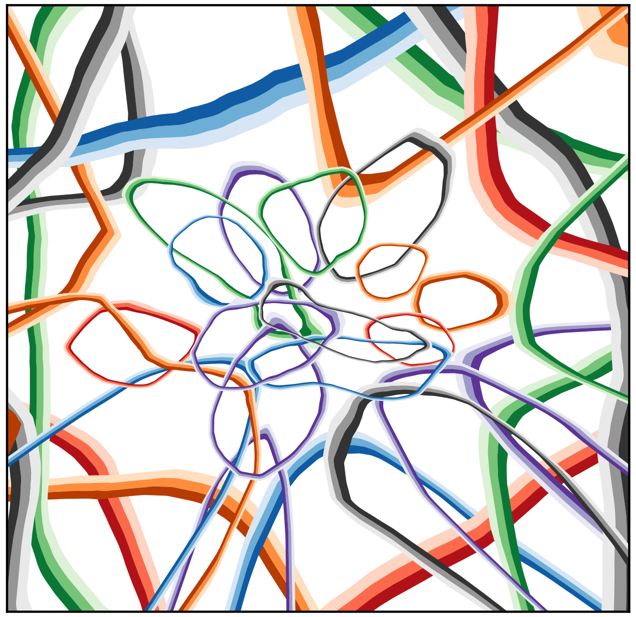
$\ell = 4$



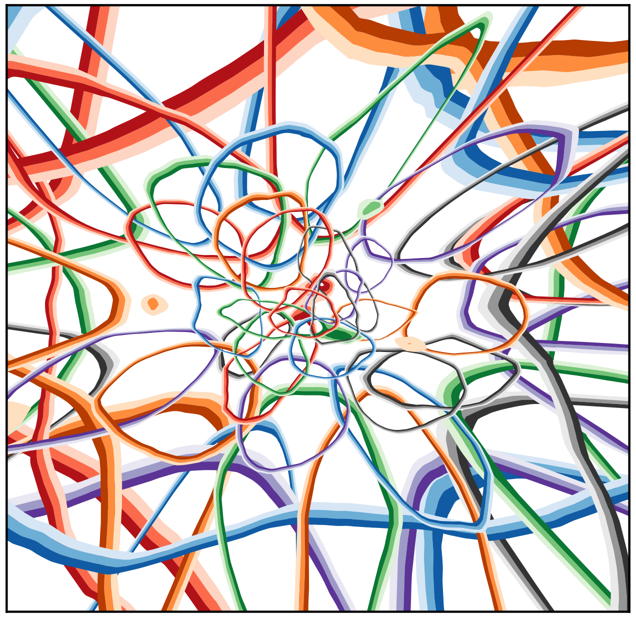
$\ell = 8$



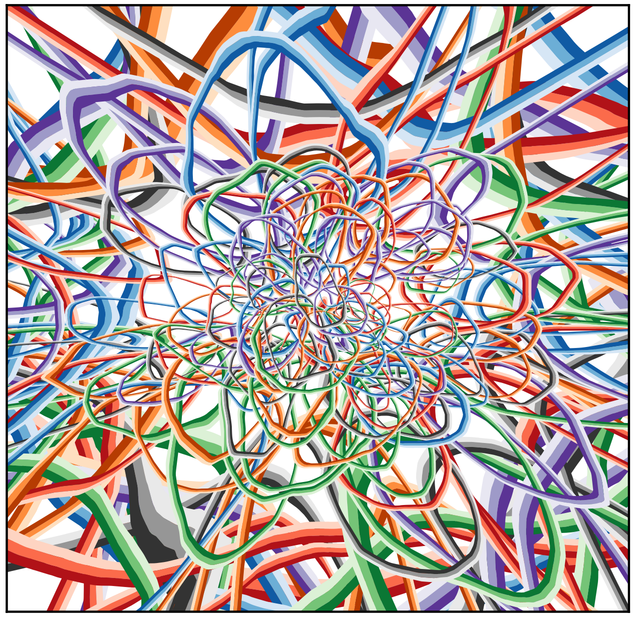
$\ell = 16$



$\ell = 32$



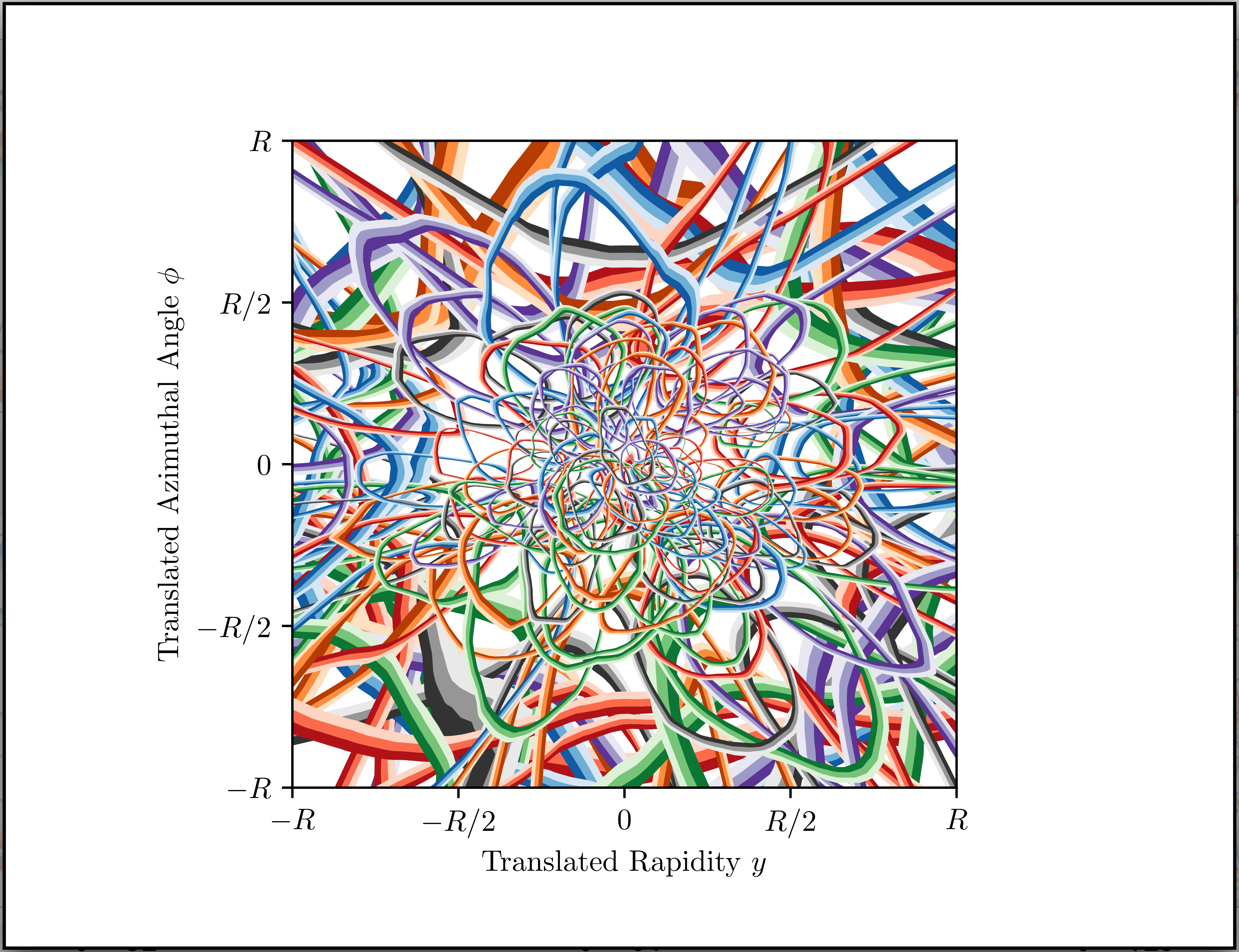
$\ell = 64$



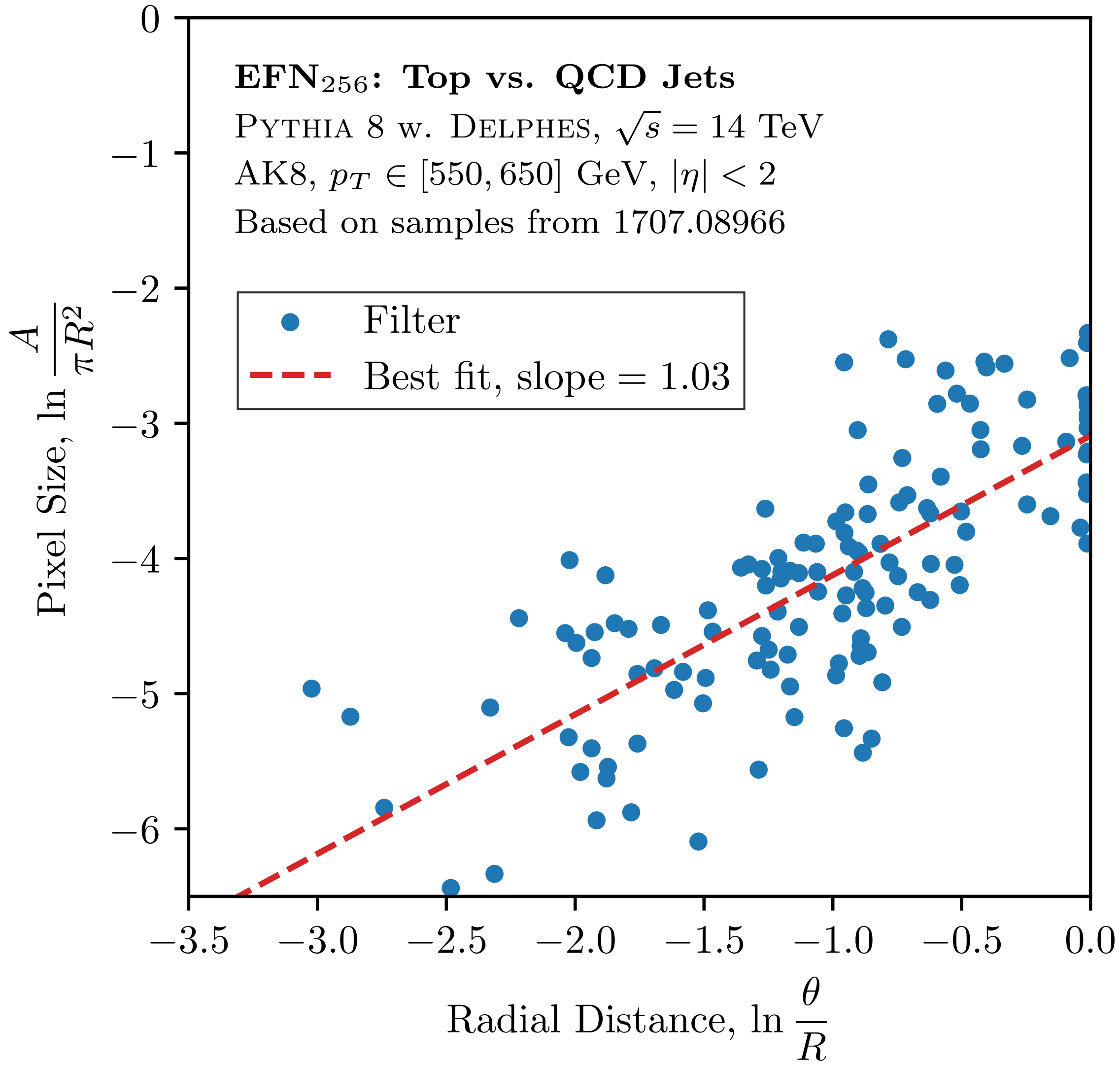
$\ell = 128$

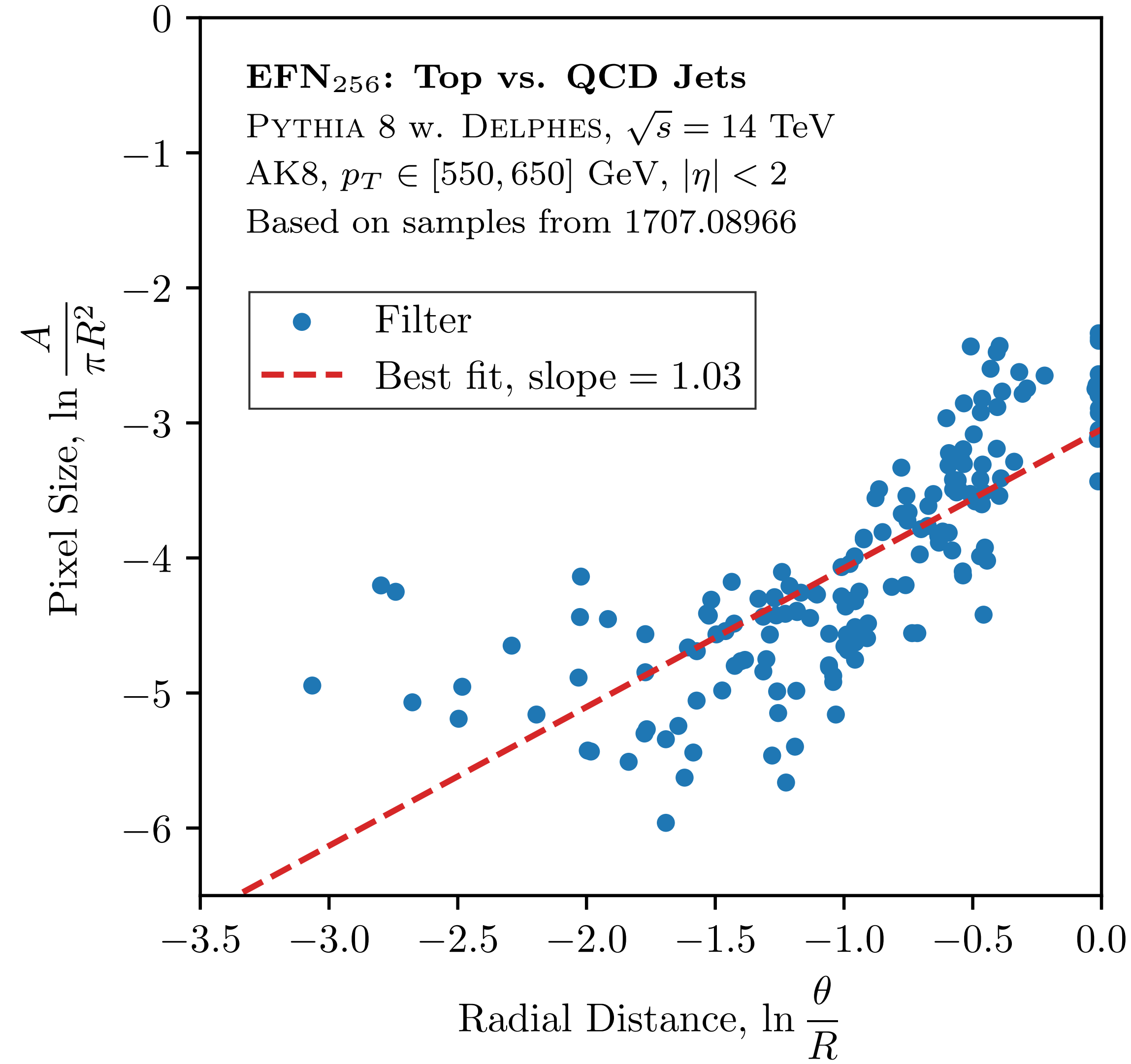
Boosted Top: Visualizing EFN Filters

Without rotation/reflection preprocessing

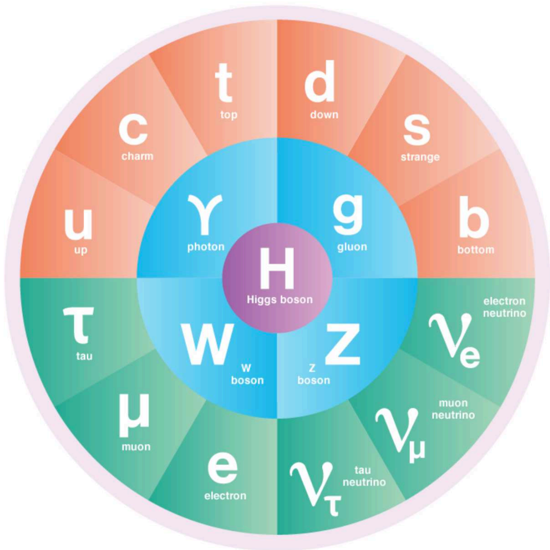


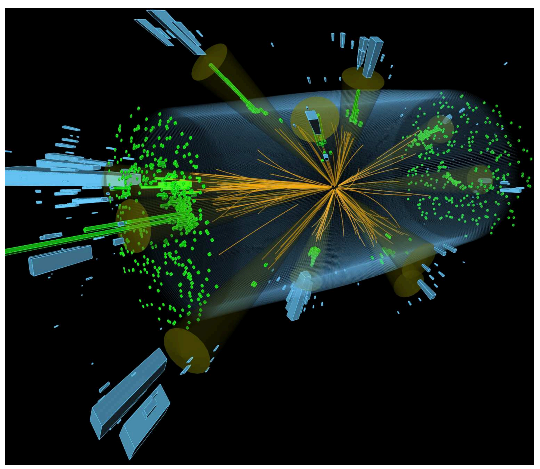
Boosted Top: Measuring EFN Filters

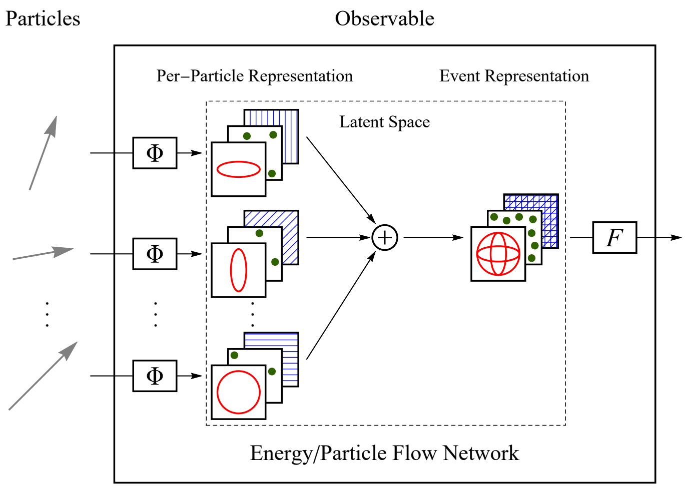




Power law behavior not as clear, slope further away from 2







Jets in Particle Physics

Ubiquitous standard model signature, likely BSM final state

Point Clouds

Same structure as jets/events at colliders

Energy Flow Networks

Jet symmetries, point clouds, Deep Sets, performance, versatility, simplicity, visualization, new analytic observables

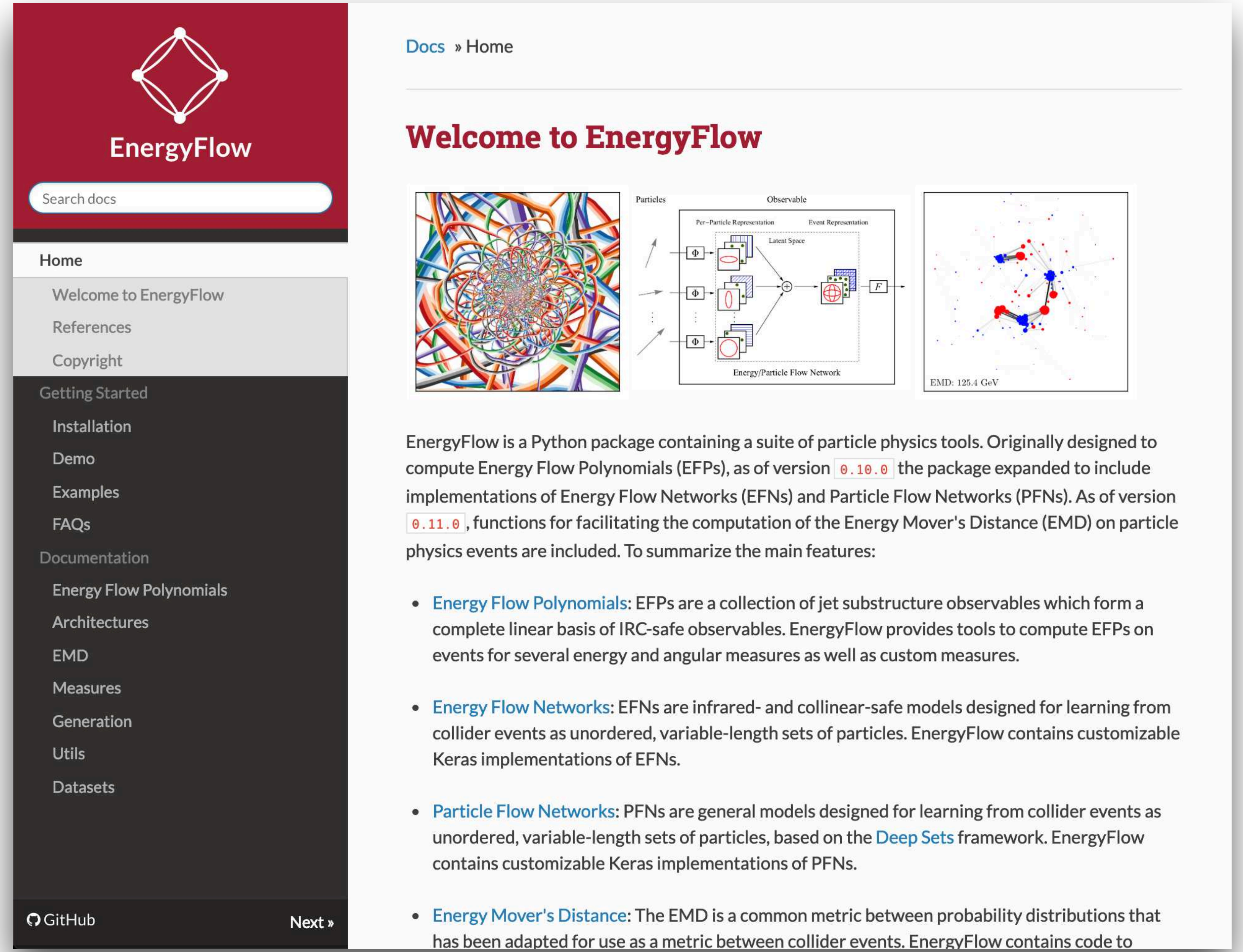
EnergyFlow Python Package

https://energyflow.network

Keras implementations of EFNs, PFNs, DNNs, CNNs, efficient EFP computation

Parallelized EMD calculations via the Python Optimal Transport library

Several detailed examples and demos for common use cases and visualization procedures



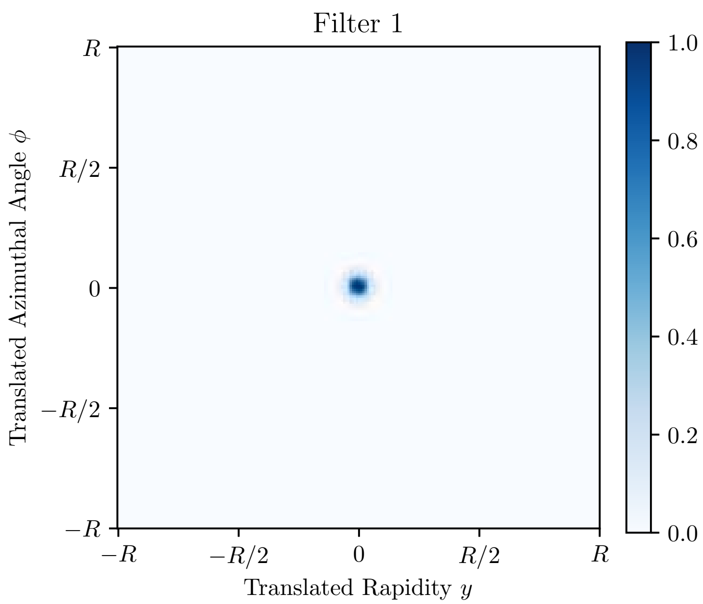
BOSTON 2019

[BOOST 2019, July 22-26, MIT]

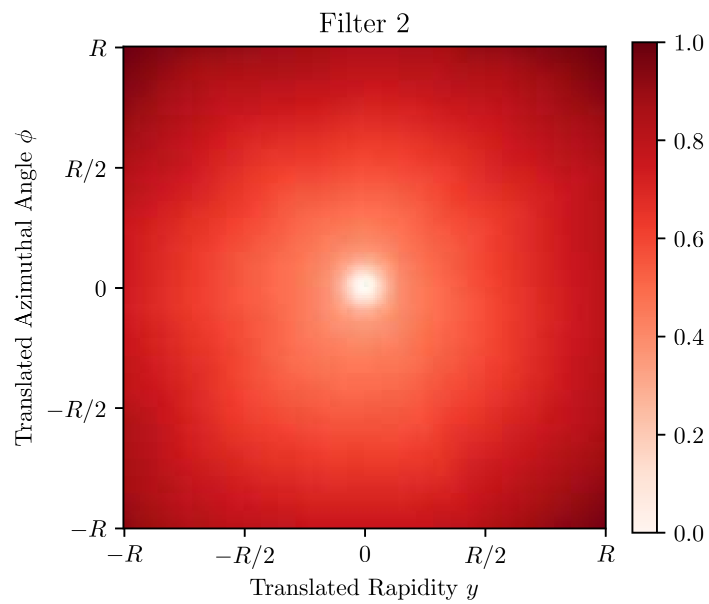
Phenomenology | Reconstruction | Searches | Algorithms | Measurements | Calculations
Modeling | Machine Learning | Pileup Mitigation | Heavy-Ion Collisions | Future Colliders

Additional Slides

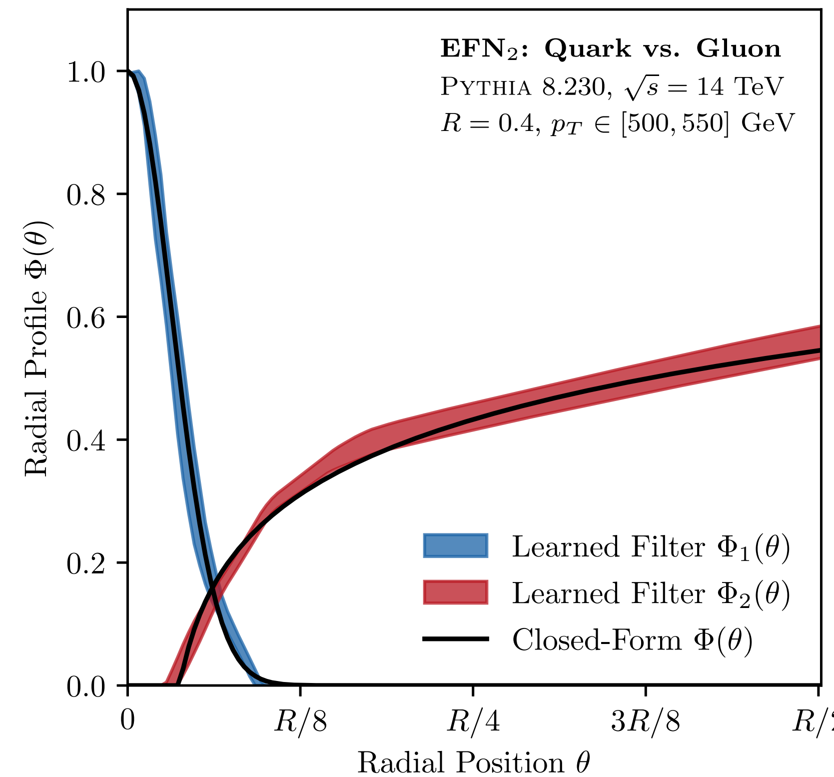
Quark vs. Gluon: Extracting New Analytic Observables



$$\mathcal{O}_1 = \sum_{i=1}^{M} z_i \, \Phi_1(\theta_i)$$



$$\mathcal{O}_2 = \sum_{i=1}^{M} z_i \, \Phi_2(\theta_i)$$



Take radial slices to obtain envelope

EFN (ℓ = 2) has approximately radially symmetric filters

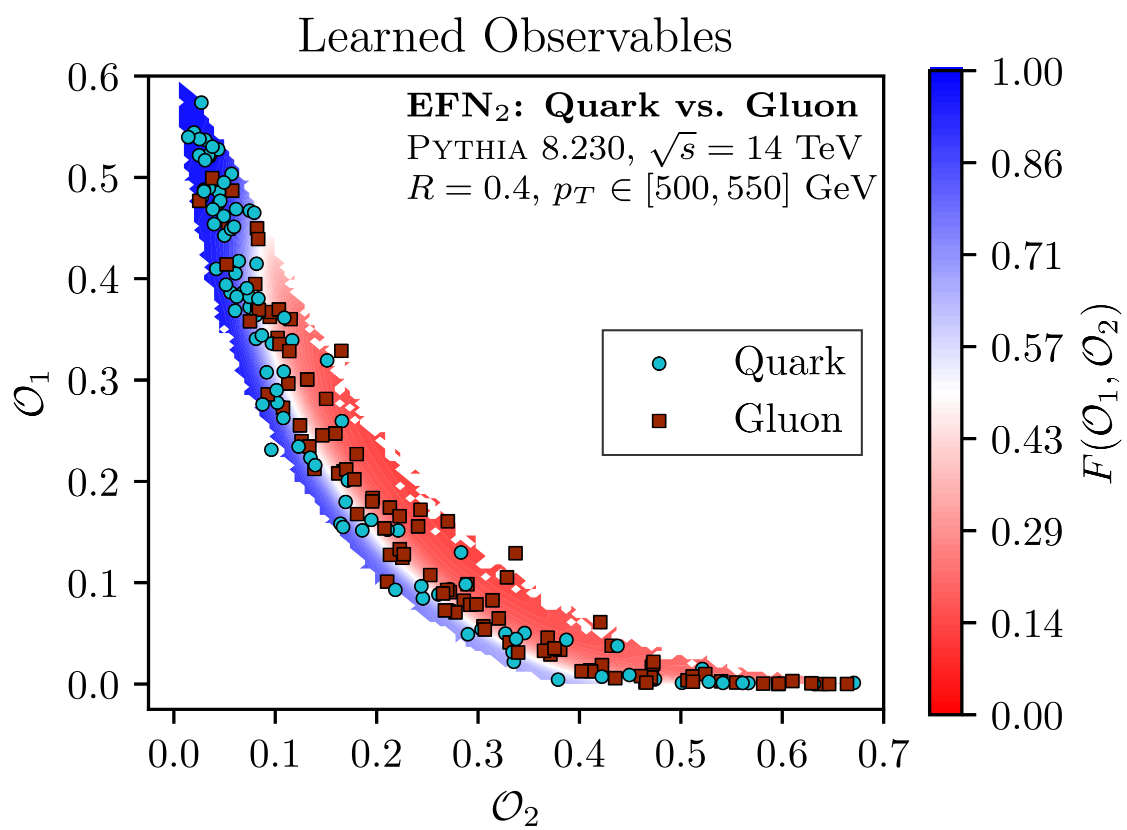
Fit functions of the forms:

$$A_{r_0} = \sum_{i=1}^{M} z_i \, e^{-\theta_i^2/r_0^2}, \qquad B_{r_1,\beta} = \sum_{i=1}^{M} z_i \, \ln(1 + \beta(\theta_i - r_1))\Theta(\theta_i - r_1)$$

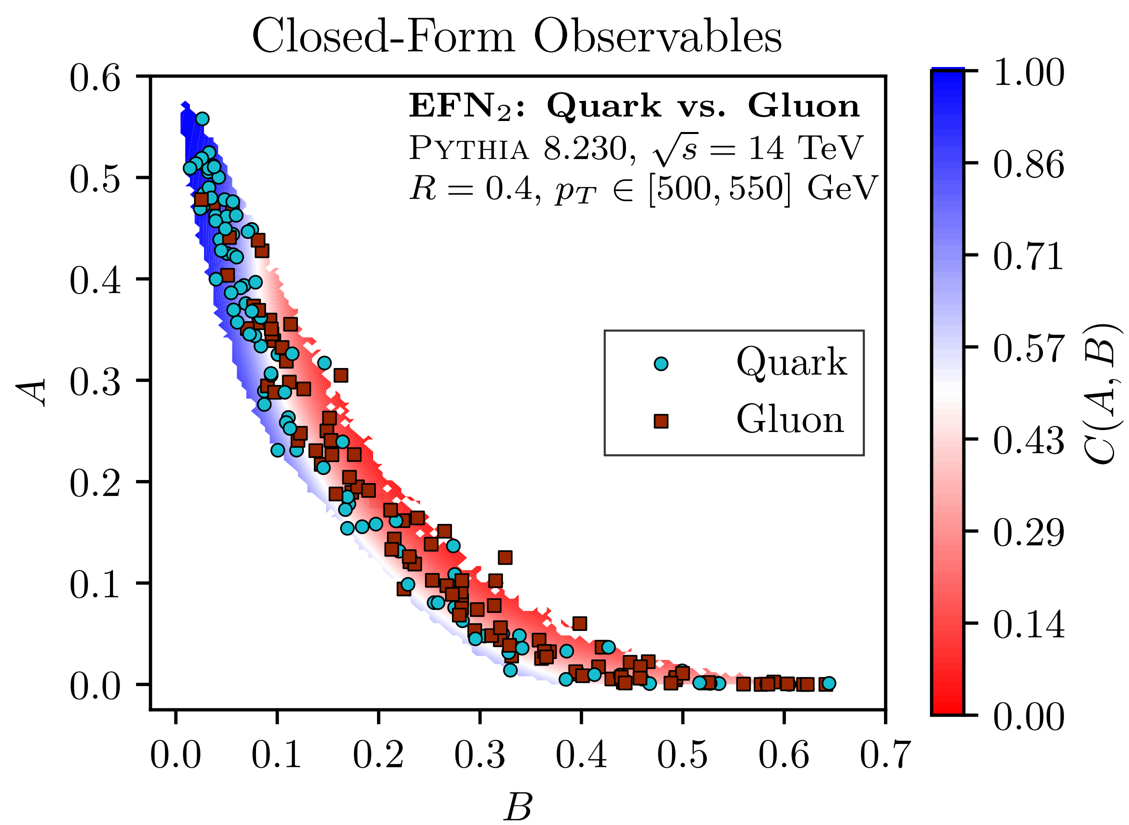
Separate soft and collinear phase space regions

Quark vs. Gluon: Extracting New Analytic Observables

Can visualize *F* in the two dimensional $(\mathcal{O}_1, \mathcal{O}_2)$ phase space



Learned



Extracted

Extract analytic form for *F* as (squared) distance from a point:

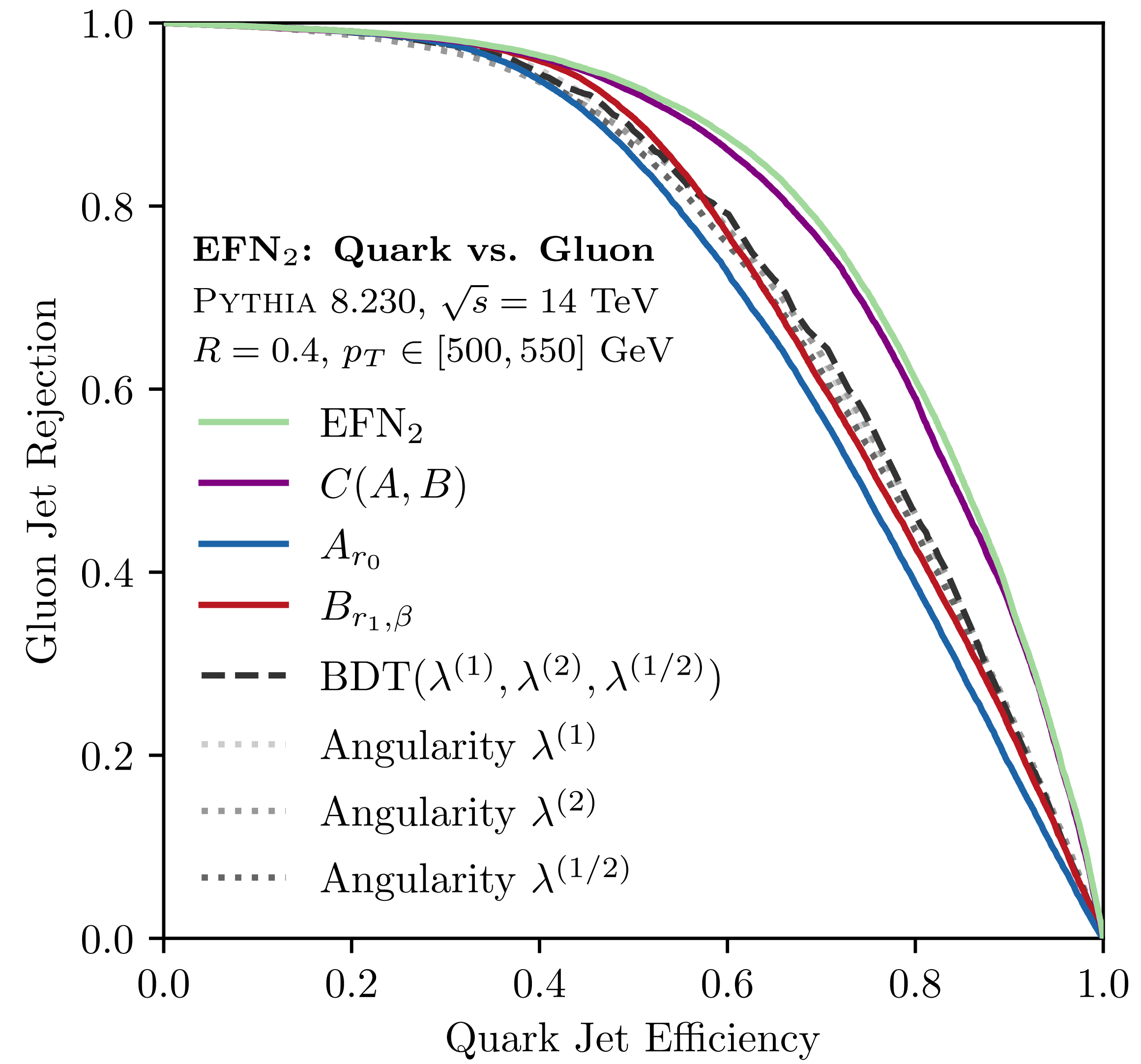
$$C(A, B) = (A - a_0)^2 + (B - b_0)^2$$

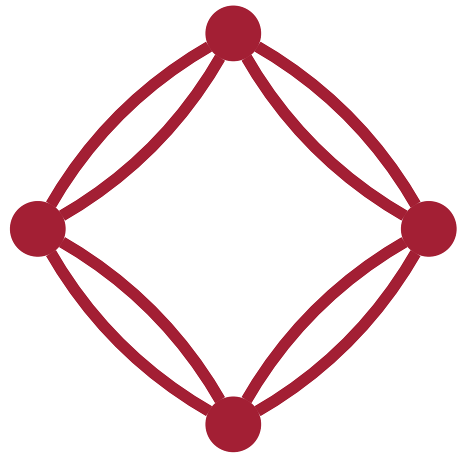
Quark vs. Gluon: Benchmarking New Analytic Observables

Individually, extracted observables are comparable to other angularities

Extracted $C(A, B)$ performs nearly as well as EFN (ℓ = 2)

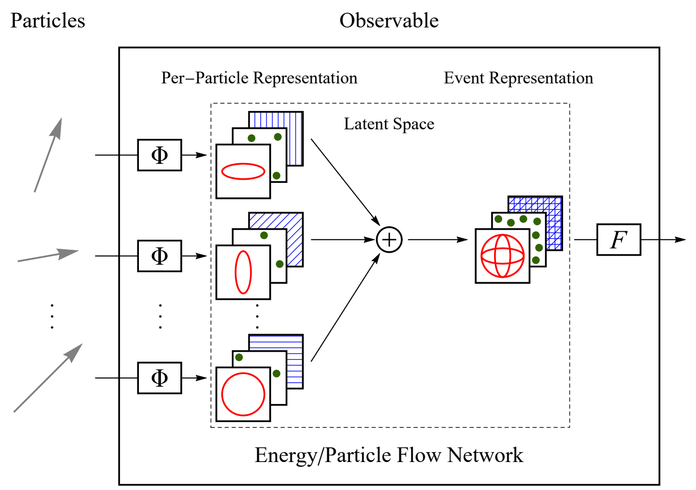
Meanwhile, multivariate combination (BDT) of three other angularities does not show improvement





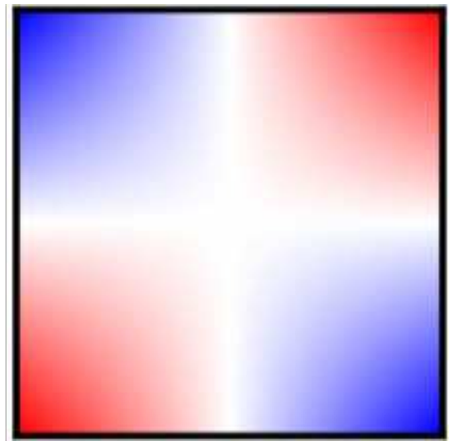
Energy Flow Polynomials

Linear basis of IRC-safe observables, fixed processing of point cloud, identify many common observables as combinations



Energy Flow Networks

Jet symmetries, point clouds, Deep Sets, performance, versatility, simplicity, visualization, new analytic observables



Energy Flow Moments

Connects multiparticle correlators to additive structures, linear in M computation of EFPs, algebraic identities

Infrared and Collinear (IRC) Safety

QCD has soft and collinear divergences associated with gluon radiation

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

$$C_q = C_F = 4/3$$
$$C_g = C_A = 3$$

<u>KLN Theorem</u>: IRC safety of an observable is sufficient to guarantee that soft/collinear divergences cancel at each order in perturbation theory

Infrared (IR) safety – observable is unchanged under addition of a soft particle

$$S(\{p_1^\mu, \ldots, p_M^\mu\}) = S(\{p_1^\mu, \ldots, (1-\lambda)p_M^\mu, \lambda p_M^\mu\}), \quad \forall\, \lambda \in [0,1]$$

Collinear (C) safety – observable is unchanged under a collinear splitting of a particle

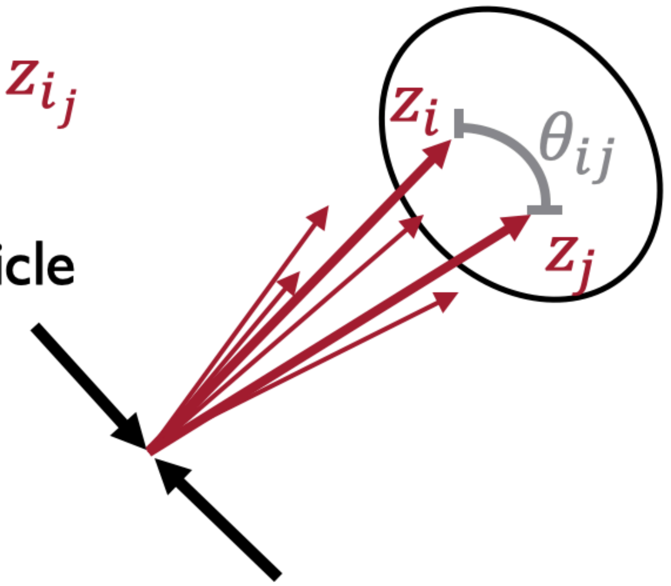
$$S(\{p_1^\mu, \ldots, p_M^\mu\}) = \lim_{\epsilon\to 0} S(\{p_1^\mu, \ldots, p_M^\mu, \epsilon p_{M+1}^\mu\}), \quad \forall\, p_{M+1}^\mu$$

IRC safety is a key theoretical *and experimental* property of observables

Derivation of EFP Linear Spanning Basis

Arbitrary IRC-safe observable: $S(p_1^\mu, \ldots, p_M^\mu)$

- Energy expansion*: Approximate S with polynomials of z_{i_j}
  - IR safety: S is unchanged under addition of soft particle
  - C safety: S is unchanged under collinear splitting of a particle
  - Relabeling symmetry: Particle index is arbitrary



Energy correlator parametrized by angular function f

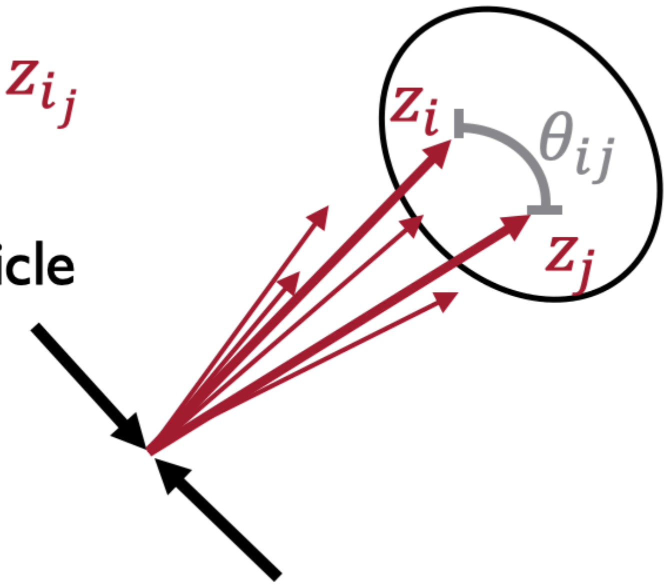
$$\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \ldots z_{i_N} f(\hat{p}_{i_1}, \ldots, \hat{p}_{i_N})$$

[F. Tkachov, hep-ph/9601308]

Derivation of EFP Linear Spanning Basis

Arbitrary IRC-safe observable: $S(p_1^\mu, \ldots, p_M^\mu)$

- Energy expansion*: Approximate S with polynomials of z_{i_j}
  - IR safety: S is unchanged under addition of soft particle
  - C safety: S is unchanged under collinear splitting of a particle
  - Relabeling symmetry: Particle index is arbitrary



[F. Tkachov, hep-ph/9601308]

Energy correlator parametrized by angular function f

$$\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \ldots z_{i_N} f(\hat{p}_{i_1}, \ldots, \hat{p}_{i_N})$$

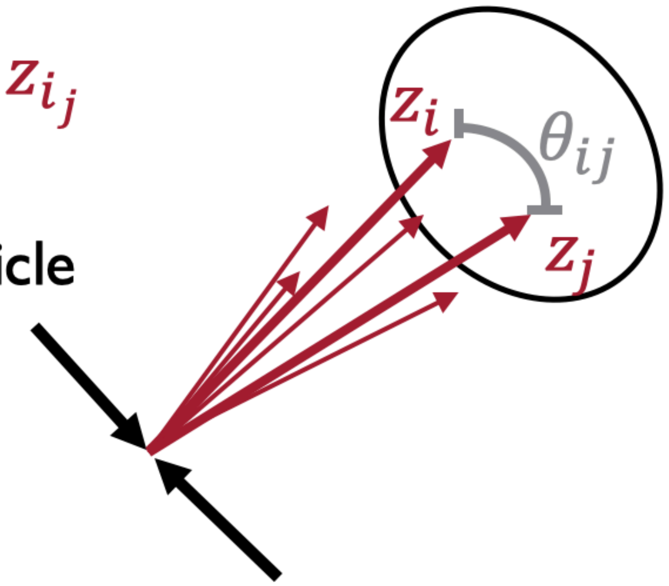
→ Energy correlators linearly span IRC-safe observables

- Angular expansion*: Approximate f with polynomials in θ_{ij}
- Simplify: Identify unique analytic structure that emerge

Derivation of EFP Linear Spanning Basis

Arbitrary IRC-safe observable: $S(p_1^\mu, \dots, p_M^\mu)$

- Energy expansion*: Approximate S with polynomials of z_{i_j}
  - IR safety: S is unchanged under addition of soft particle
  - C safety: S is unchanged under collinear splitting of a particle
  - Relabeling symmetry: Particle index is arbitrary



[F. Tkachov, hep-ph/9601308]

Energy correlator parametrized by angular function f

$$\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \dots z_{i_N} f(\hat{p}_{i_1}, \dots, \hat{p}_{i_N})$$

➡ Energy correlators linearly span IRC-safe observables

- Angular expansion*: Approximate f with polynomials in θ_{ij}
- Simplify: Identify unique analytic structure that emerge

➡ Linear spanning basis in terms of "EFPs" has been found!

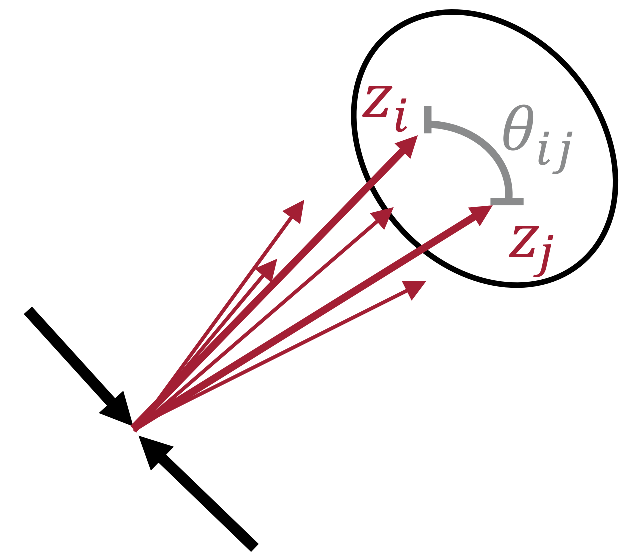
$$S \simeq \sum_{g \in G} s_G \mathrm{EFP}_G\,, \qquad \mathrm{EFP}_G \equiv \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \dots z_{i_N} \prod_{(k,\ell) \in G} \theta_{i_k i_\ell}$$

**Generically, approximations exist by the Stone-Weierstrass theorem

Energy Flow Polynomials (EFPs)

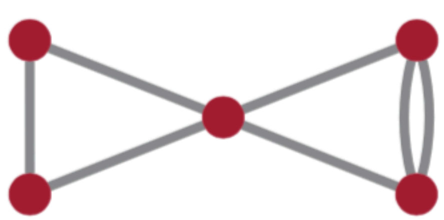
[PTK, Metodiev, Thaler, 1712.07124]

$$\mathrm{EFP}_G = \underbrace{\sum_{i_1=1}^{M}\cdots\sum_{i_N=1}^{M}}_{\text{Correlator}} \;{}_{\text{of}}\; \underbrace{z_{i_1}\cdots z_{i_N}}_{\text{Energies}} \;{}_{\text{and}}\; \underbrace{\prod_{(k,\ell)\in G}\theta_{i_k i_\ell}}_{\text{Angles}}$$



Generalizes many well-known and studied classes of energy correlators observables

A family of energy correlators with angular structures determined by multigraphs



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1}z_{i_2}z_{i_3}z_{i_4}z_{i_5}\theta_{i_1i_2}\theta_{i_2i_3}\theta_{i_1i_3}\theta_{i_1i_4}\theta_{i_1i_5}\theta_{i_4i_5}^2$$

<u>Multigraph correspondence</u>

$\longleftrightarrow z_{i_j}$

$\longleftrightarrow \theta_{i_k i_l}$

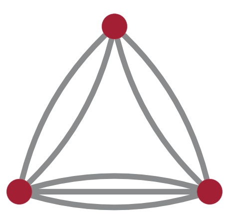
Energy and Angle Measure

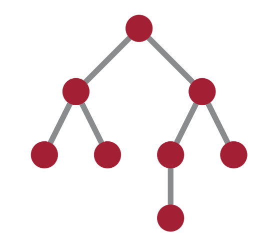
Hadronic : $z_i = \dfrac{p_{Ti}}{\sum_j p_{Tj}}, \quad \theta_{ij} = \left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)^{\beta/2}$

Linear Basis of IRC-Safe Observables

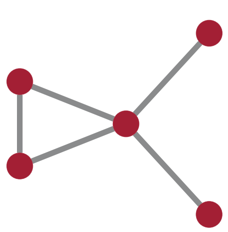
One can show via the Stone-Weierstrass approximation theorem that any IRC-safe observable is a linear combination of EFPs

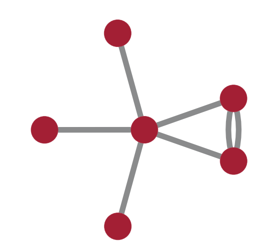
$$\mathcal{S} \simeq \sum_{G \in \mathcal{G}} s_G \mathrm{EFP}_G, \quad \mathcal{G} \text{ a set of multigraphs}$$





Multivariate combinations of EFPs only require <u>linear methods</u> to achieve full generality

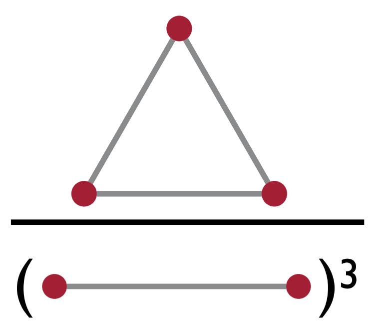




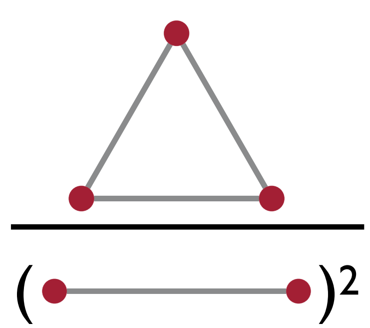
Strategy: Learn coefficients s_G via linear regression or classification

Familiar Observables as EFPs

$m_J^2 =$ 

$D_2 =$ 

[Larkoski, Moult, Neill, 2014]

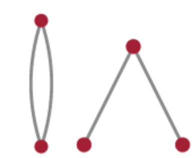
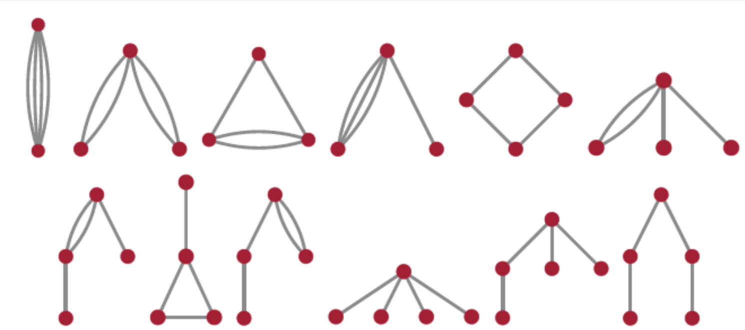
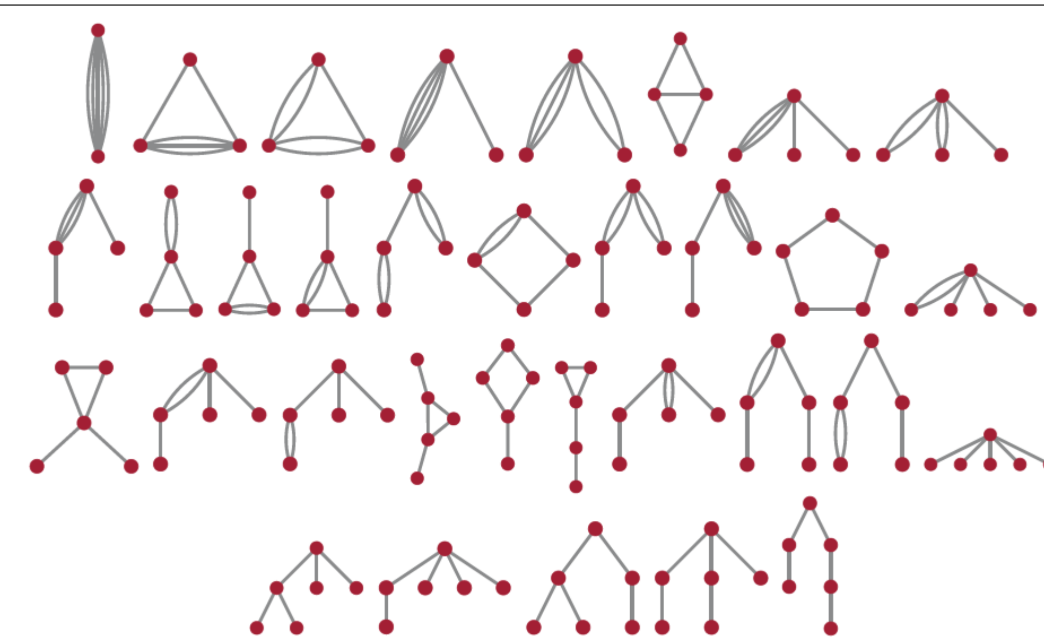
$C_2 =$ 

[Larkoski, Salam, Thaler, 2013]

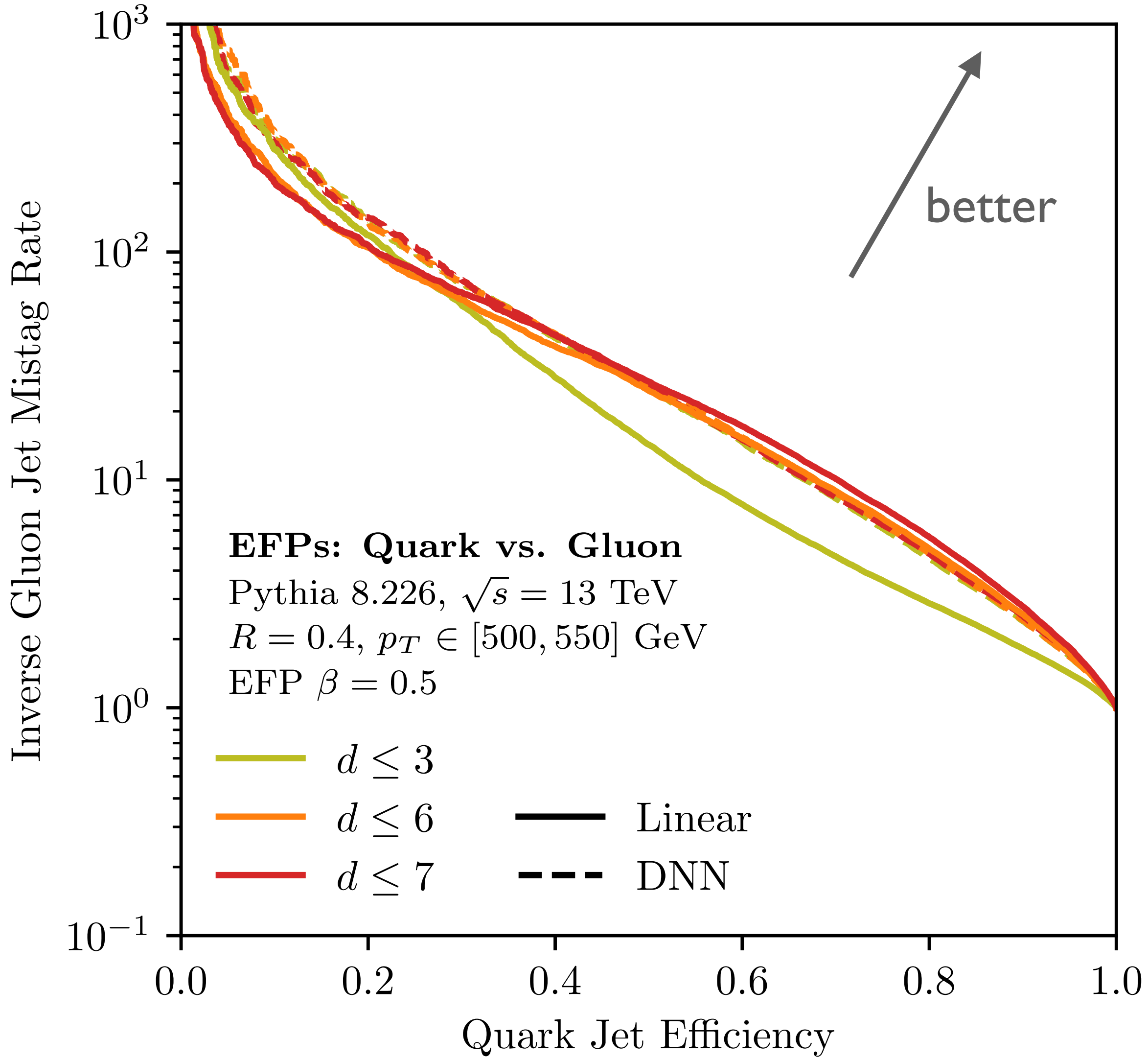
Energy correlation functions are complete graphs

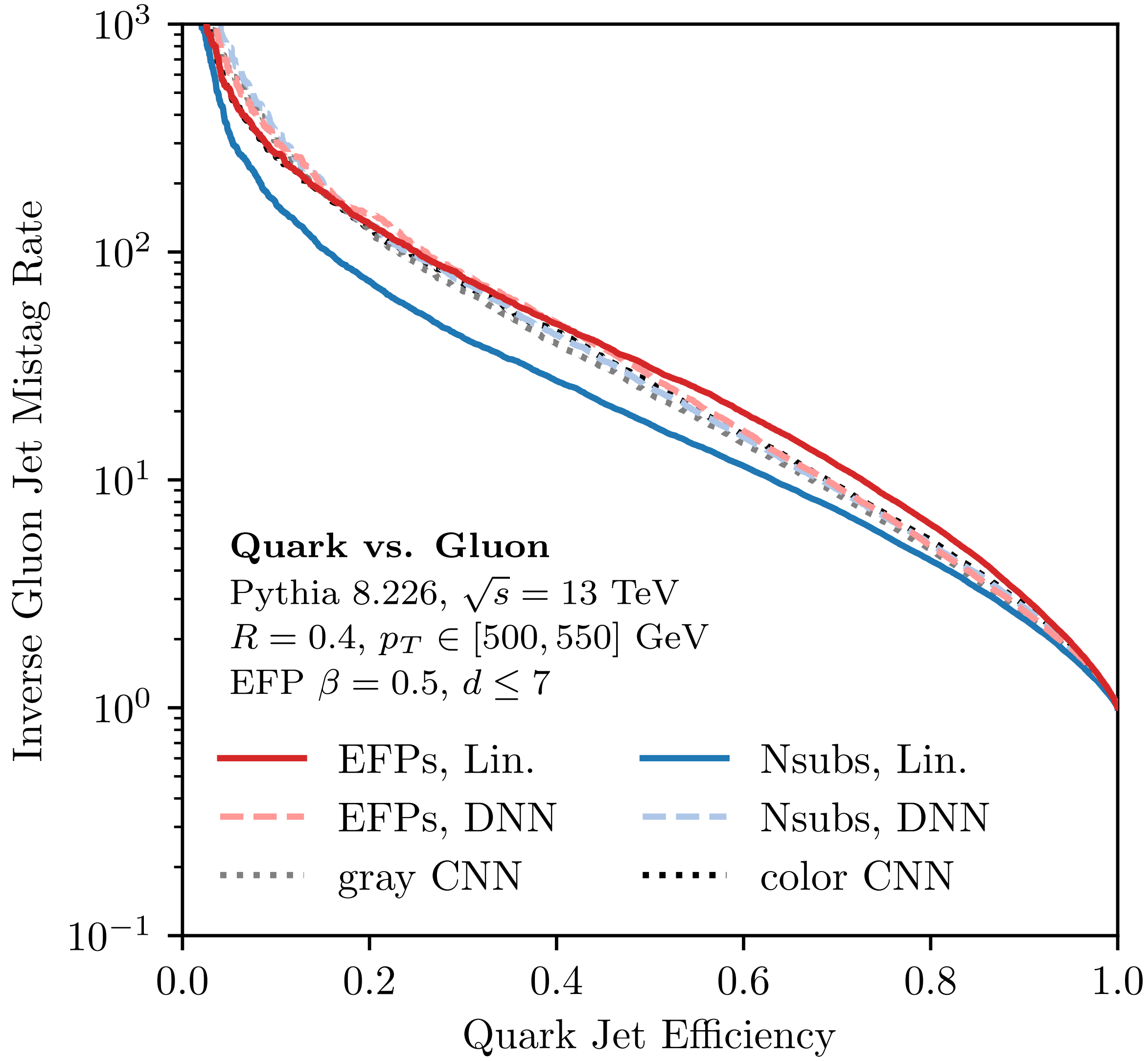
Even angularities are exact linear combinations of EFPs

EFPs organized by degree *d* – number of edges

Degree	Connected Multigraphs
$d = 1$	
$d = 2$	
$d = 3$	
$d = 4$	
$d = 5$	

Quark vs. Gluon: EFP Classification Performance Comparison

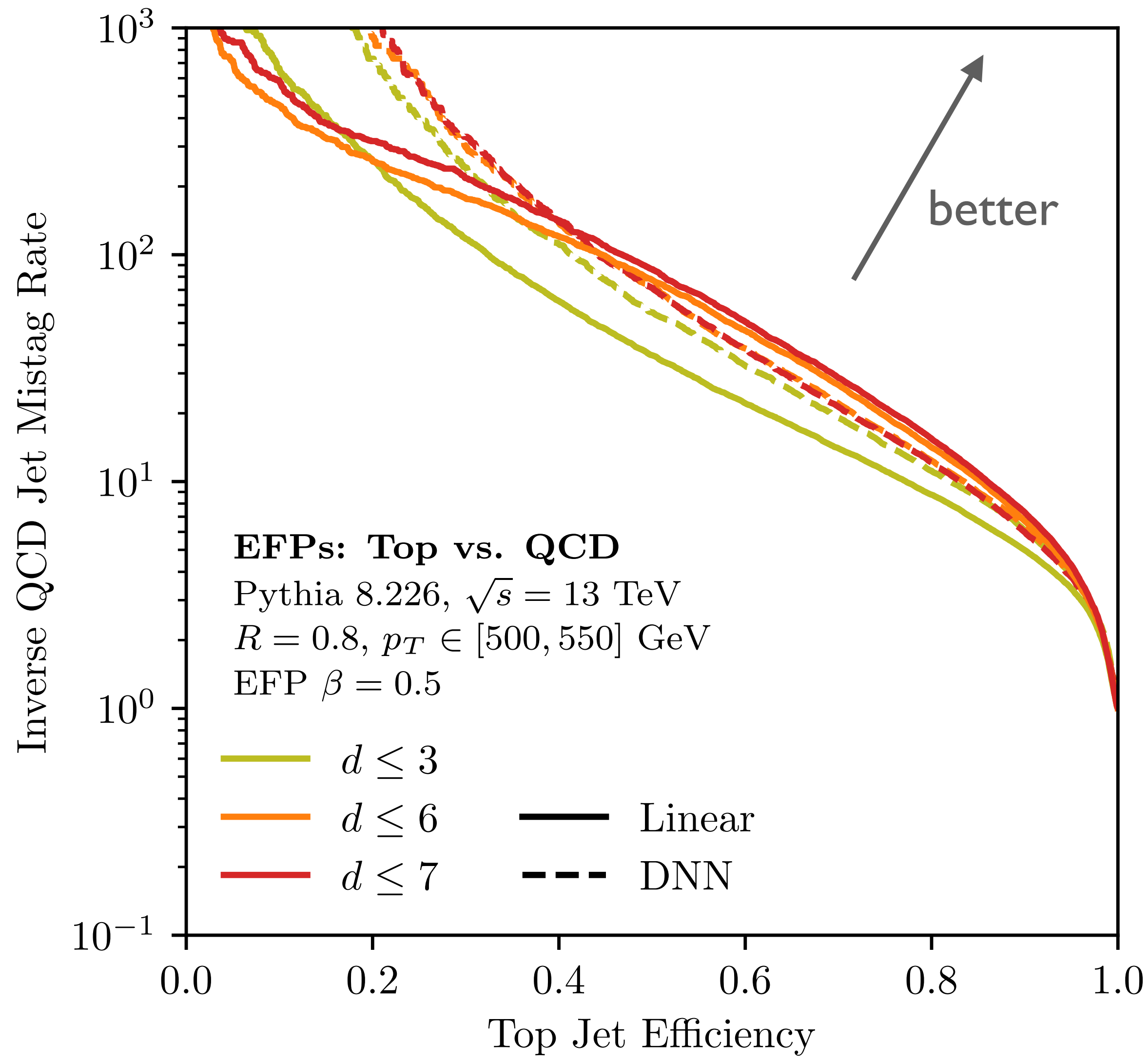


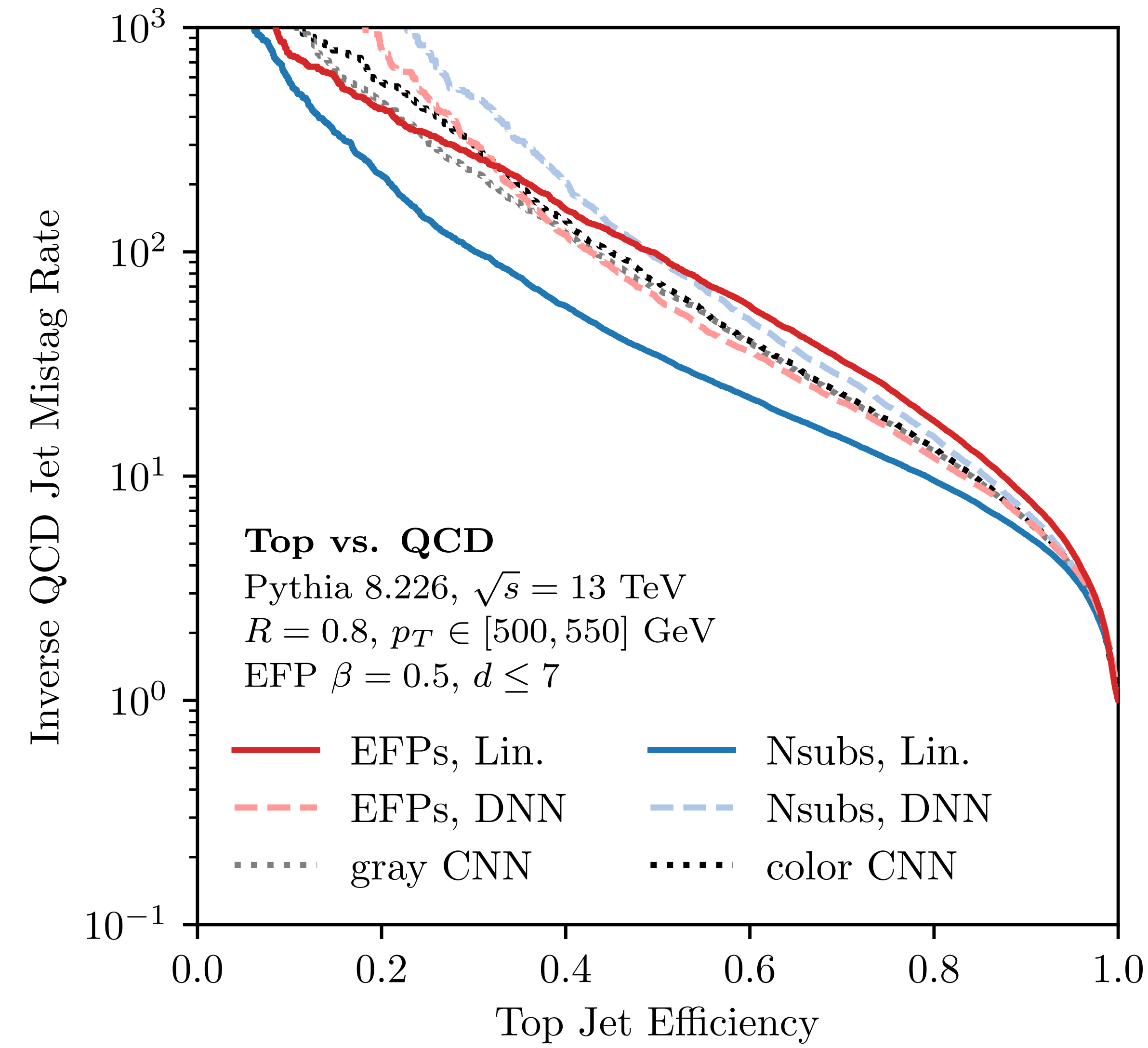


Saturation observed with more EFPs

[de Oliviera, Kagan, Mackey, Nachman, Schwartzman, 2015]
[PTK, Metodiev, Schwartz, 2016]
[Datta, Larkoski, 2017]

Boosted Top: EFP Classification Performance Comparison





Saturation observed with more EFPs

DNN gets there faster but linear suffices

Linear EFPs excel at high efficiency

[de Oliviera, Kagan, Mackey, Nachman, Schwartzman, 2015]
[PTK, Metodiev, Schwartz, 2016]
[Datta, Larkoski, 2017]

Energy Flow Moments

Consider a slightly different hadronic angular measure, $\theta_{ij} = (2\hat{p}_i^{\mu}\hat{p}_{j\mu})^{\frac{\beta}{2}}, \quad \hat{p}_i^{\mu} = \frac{p_i^{\mu}}{p_{Ti}}$

Agrees with previous hadronic measure in the limit of narrow, central jets

When $\beta = 2$, angular measure can be factored, which motivates defining:

Energy Flow Moment (EFM) of valency v: $$\mathcal{I}^{\mu_1\cdots\mu_v} = \sum_{i=1}^{M} z_i \hat{p}_i^{\mu_1}\cdots\hat{p}_i^{\mu_v}$$

$\beta = 2$ EFPs can be rewritten in terms of EFMs, which are $\boxed{\text{linear in M to compute!}}$

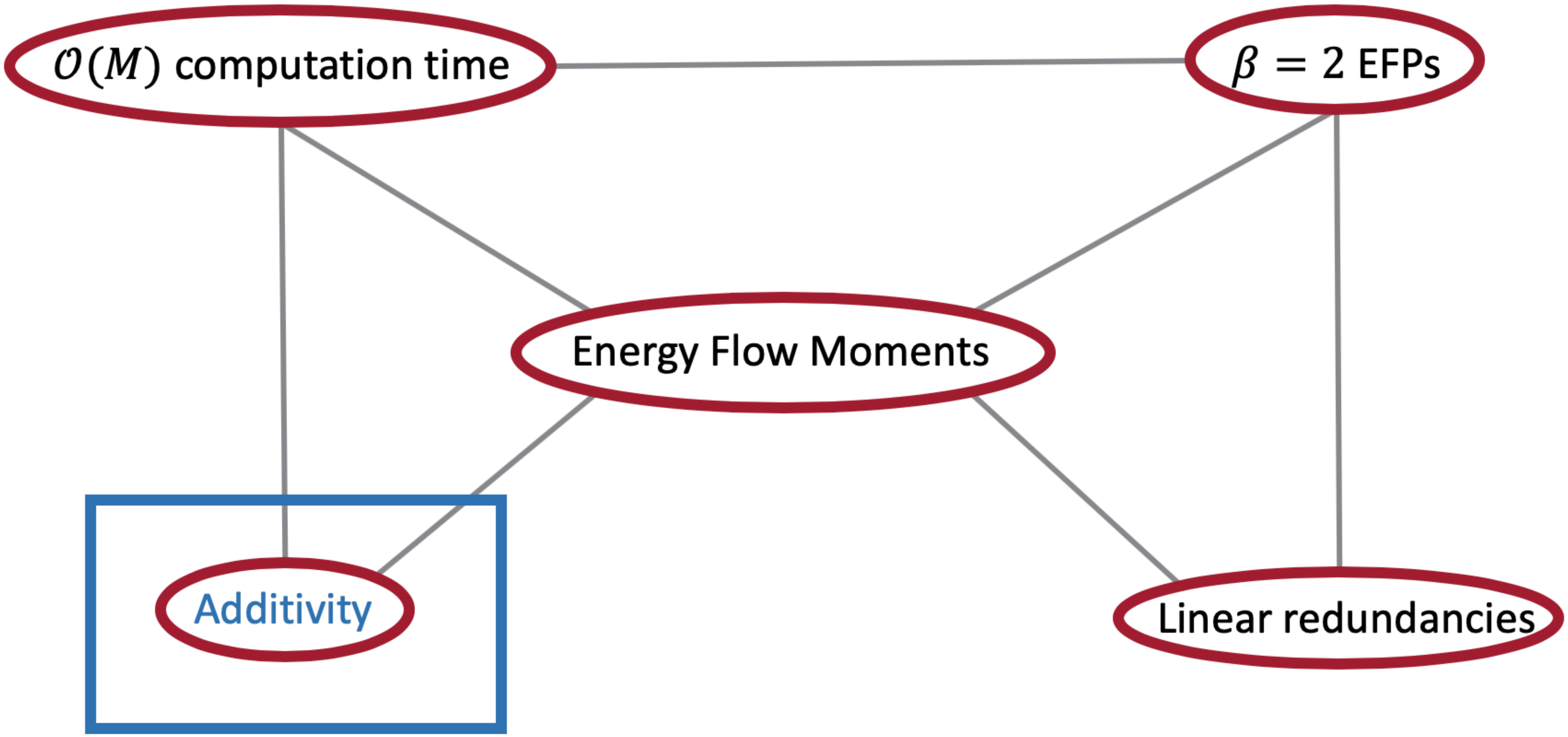
$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1}z_{i_2}z_{i_3}\theta_{i_1i_2}^2\theta_{i_1i_3}^2\theta_{i_2i_3}$$
$$= 2^5 \underbrace{\left(\sum_{i_1=1}^{M} z_{i_1}\hat{p}_{i_1}^{\alpha}\hat{p}_{i_1}^{\beta}\hat{p}_{i_1}^{\gamma}\hat{p}_{i_1}^{\delta}\right)}_{\mathcal{I}^{\alpha\beta\gamma\delta}} \underbrace{\left(\sum_{i_2=1}^{M} z_{i_2}\hat{p}_{i_2\alpha}\hat{p}_{i_2\beta}\hat{p}_{i_2}^{\epsilon}\right)}_{\mathcal{I}_{\alpha\beta}{}^{\epsilon}} \underbrace{\left(\sum_{i_3=1}^{M} z_{i_3}\hat{p}_{i_3\gamma}\hat{p}_{i_3\delta}\hat{p}_{i_3\epsilon}\right)}_{\mathcal{I}_{\gamma\delta\epsilon}}$$

A multigraph correspondence also exists for EFMs:

$\Longleftrightarrow \mathcal{I}^{\mu_k\cdots\mu_\ell}$

$\Longleftrightarrow \eta_{\mu_i\mu_j}$

Energy Flow "Network"



Additivity is the link to the Deep Sets decomposition and EFNs

$5! \times \mathcal{I}^{\mu_2}_{[\mu_1}\mathcal{I}^{\mu_3}_{\mu_2}\mathcal{I}^{\mu_4}_{\mu_3}\mathcal{I}^{\mu_5}_{\mu_4}\mathcal{I}^{\mu_1}_{\mu_5]} = 6 \times$  $- 5 \times$   $= 0$

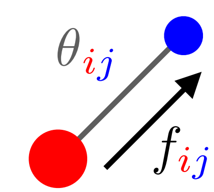
The Energy Mover's Distance

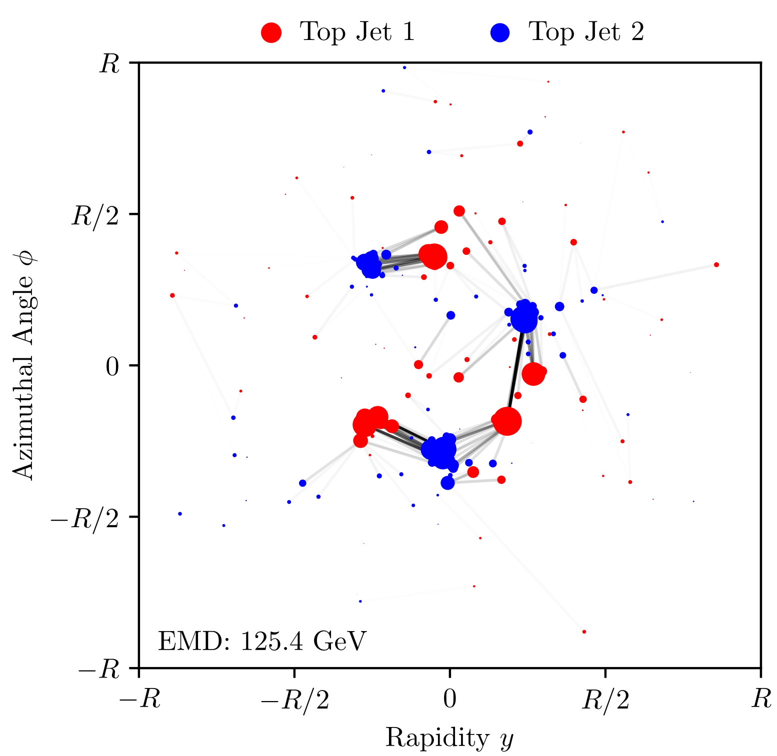
[PTK, Metodiev, Thaler, 1902.02346]

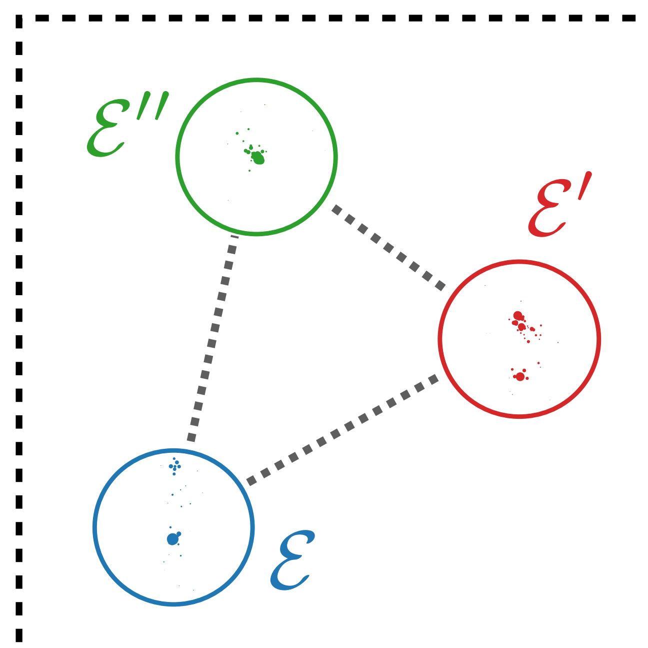
EMD between energy flows defines a metric on the space of events

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \underbrace{\min_{\{f_{ij}\geq 0\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R}}_{\text{Cost of optimal transport}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost of energy creation}}$$

$$\sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)$$







Triangle inequality satisfied for $R \geq d_{\max}/2$

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

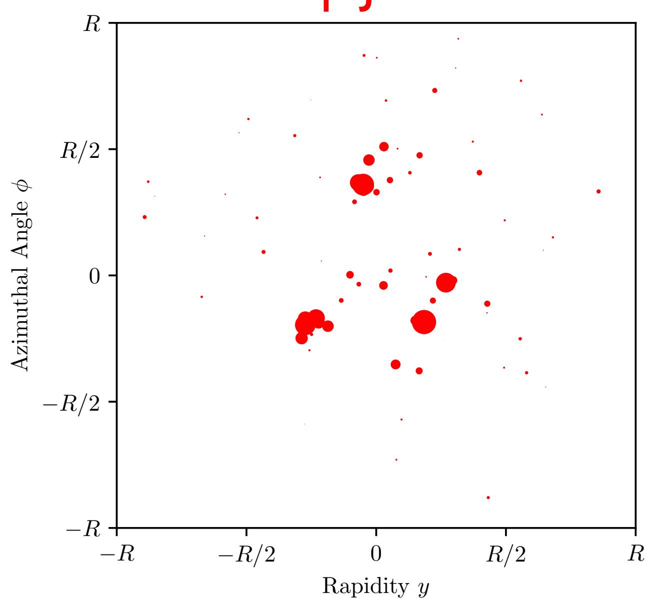
The Earth Mover's Distance

A metric on normalized distributions in a space with a ground distance measure

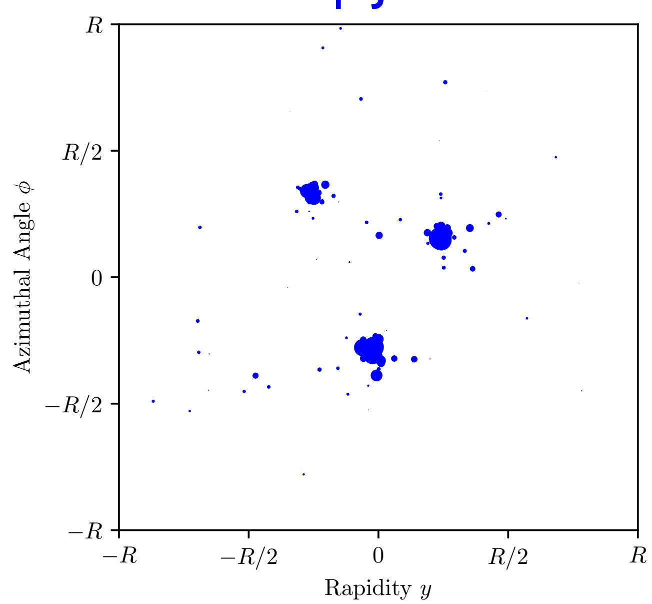
↳ symmetric, non-negative, triangle inequality, zero iff identical

The minimum "work" (stuff x distance) required to transport supply to demand

Top Jet 1



Top Jet 2



Related to *optimal transport* theory – commonly used as a metric on the space of images

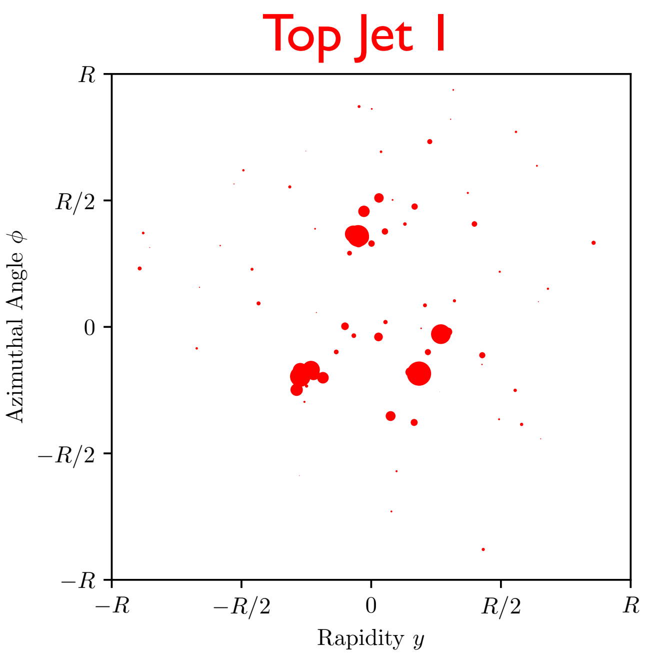
[Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

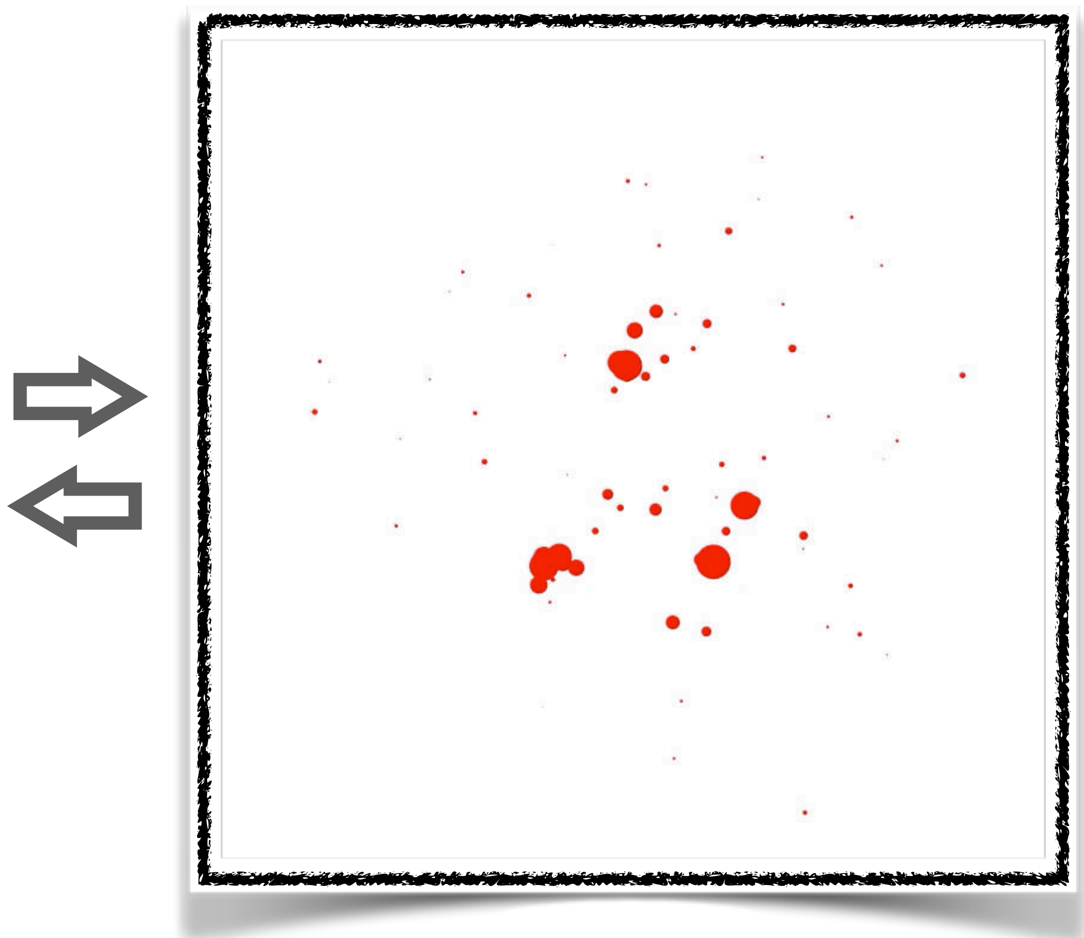
The Earth Mover's Distance

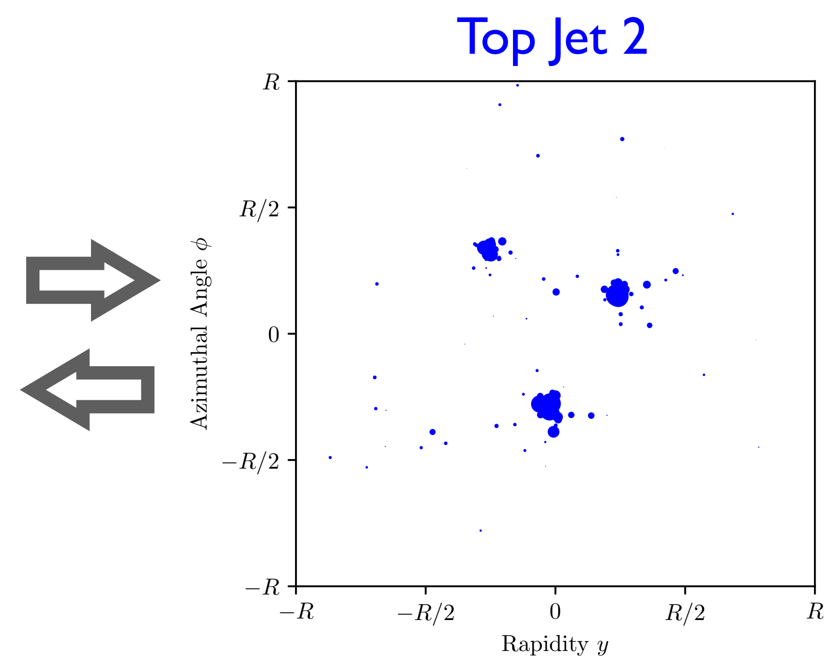
A metric on normalized distributions in a space with a ground distance measure

↳ symmetric, non-negative, triangle inequality, zero iff identical

The minimum "work" (stuff x distance) required to transport supply to demand







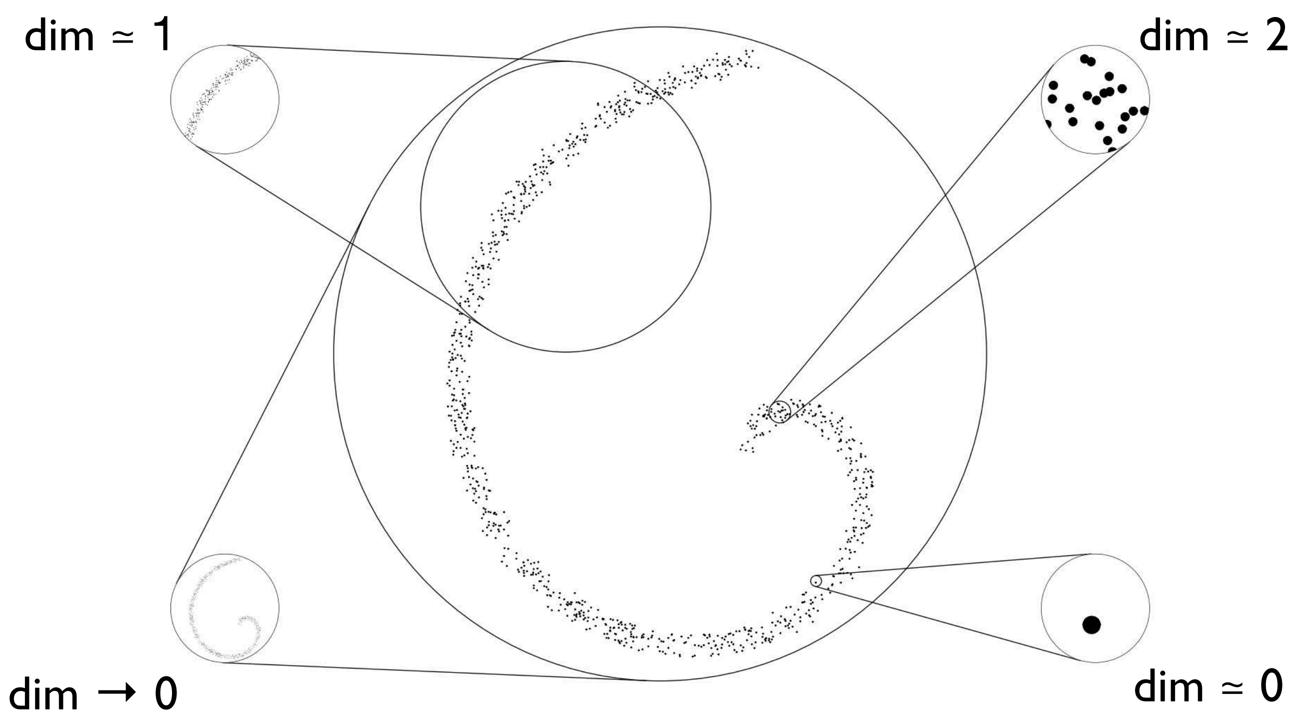
Related to *optimal transport* theory – commonly used as a metric on the space of images

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

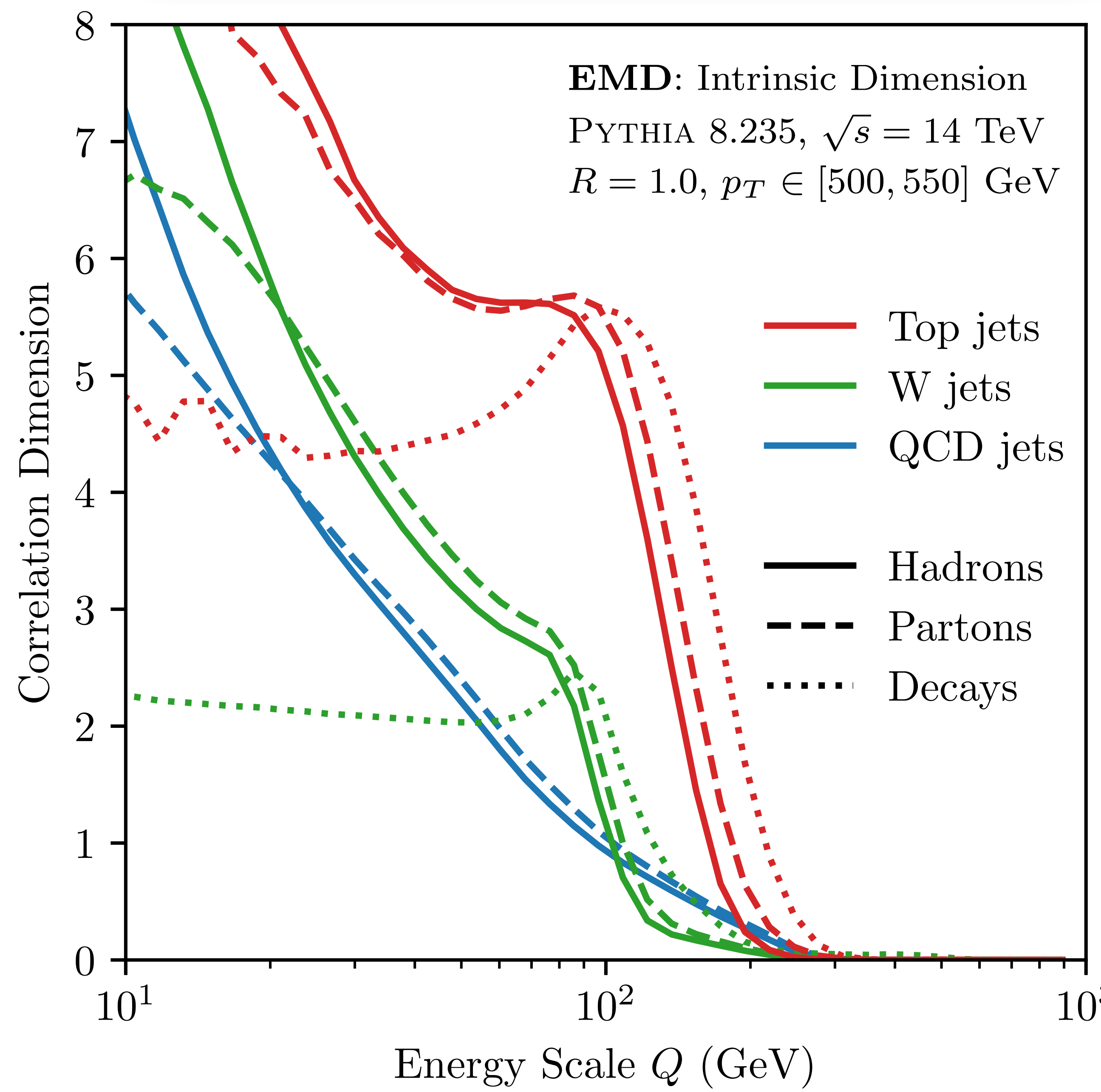
Manifold Dimensions of Event Space

Correlation dimension: how does the # of elements within a ball of size Q change?

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j \Theta(\mathrm{EMD}(\mathcal{E}_i, \mathcal{E}'_j) < Q)$$



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

- Decays are "constant" dim. at low Q
- Complexity hierarchy: QCD < W < Top
- Fragmentation increases dim. at smaller scales
- Hadronization important around 20-30 GeV

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]