Energy Flow Networks: Deep Sets for Particle Jets

Patrick T. Komiske III

Massachusetts Institute of Technology
Center for Theoretical Physics

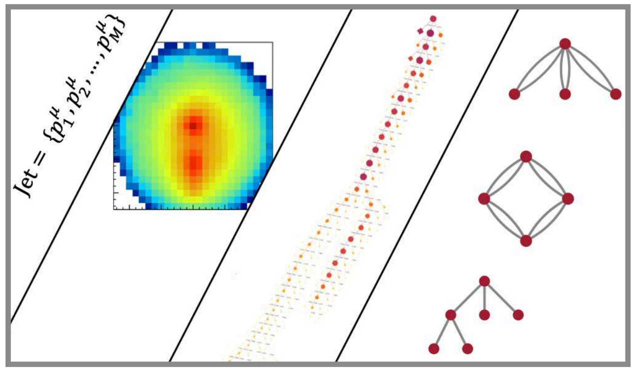
Machine Learning for Jet Physics Workshop

Fermilab, Illinois – 11/15/2018

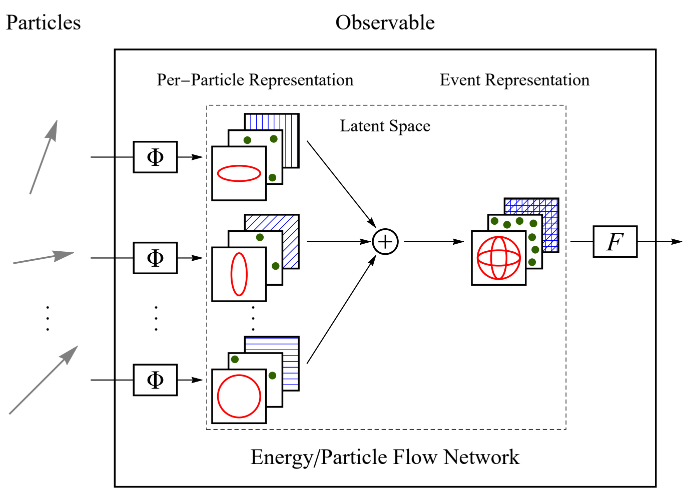
Based on work with Eric Metodiev and Jesse Thaler

1810.05165

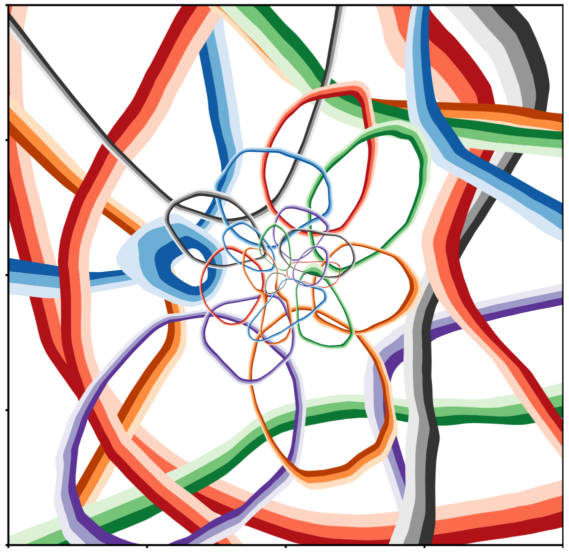
https://energyflow.network



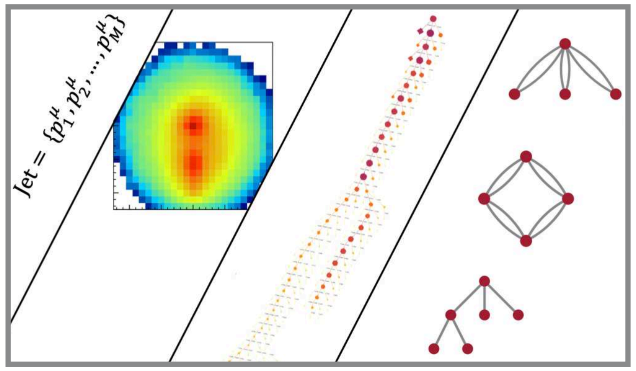
Jets as Point Clouds



Energy Flow Networks



Quark vs. Gluon Tagging



Jets as Point Clouds

Energy Flow Networks

Quark vs. Gluon Tagging

What is a Jet?

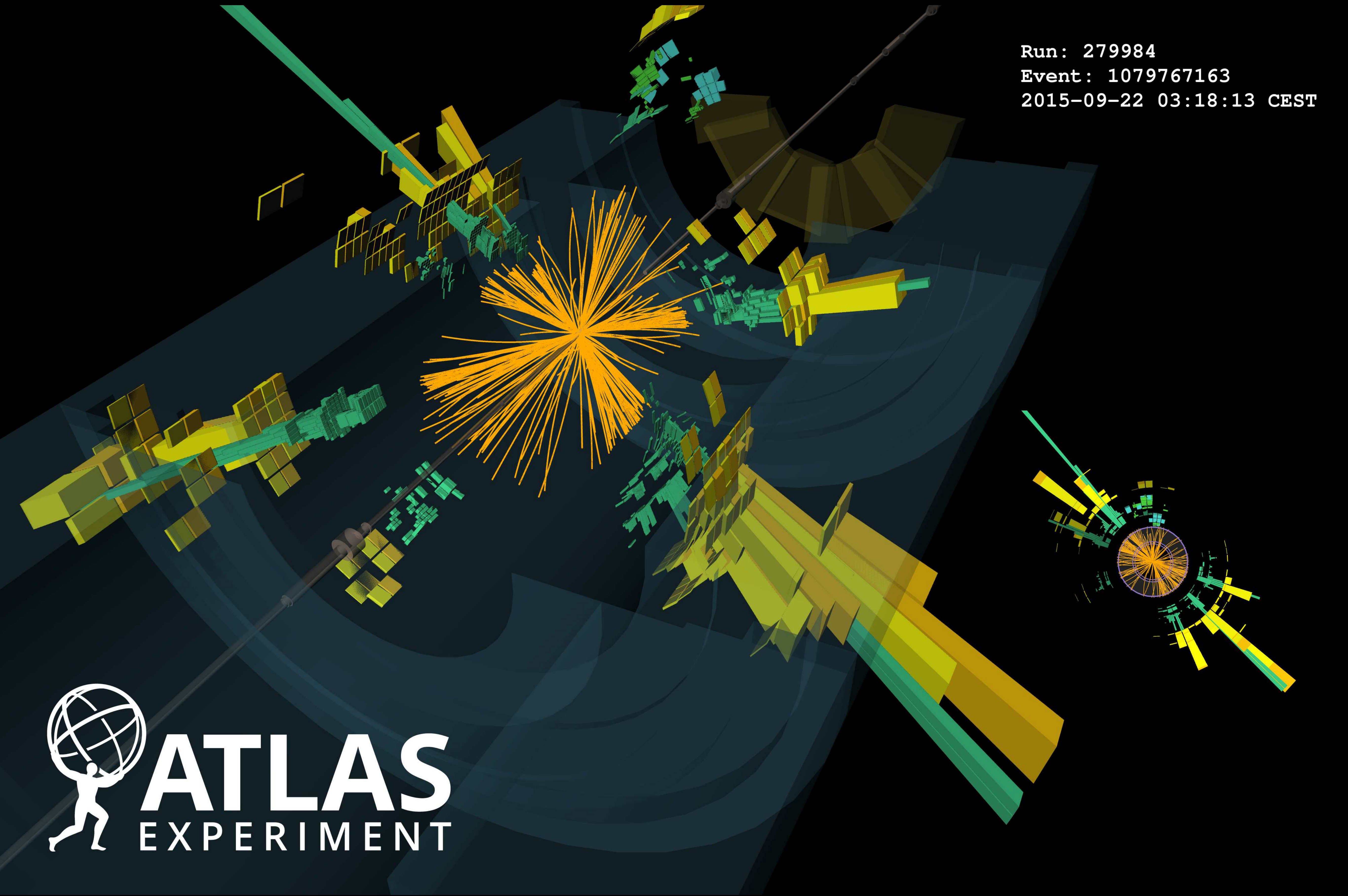
*An **unordered**, **variable length** collection of particles*

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

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

p_i^μ represents *all* the particle properties:

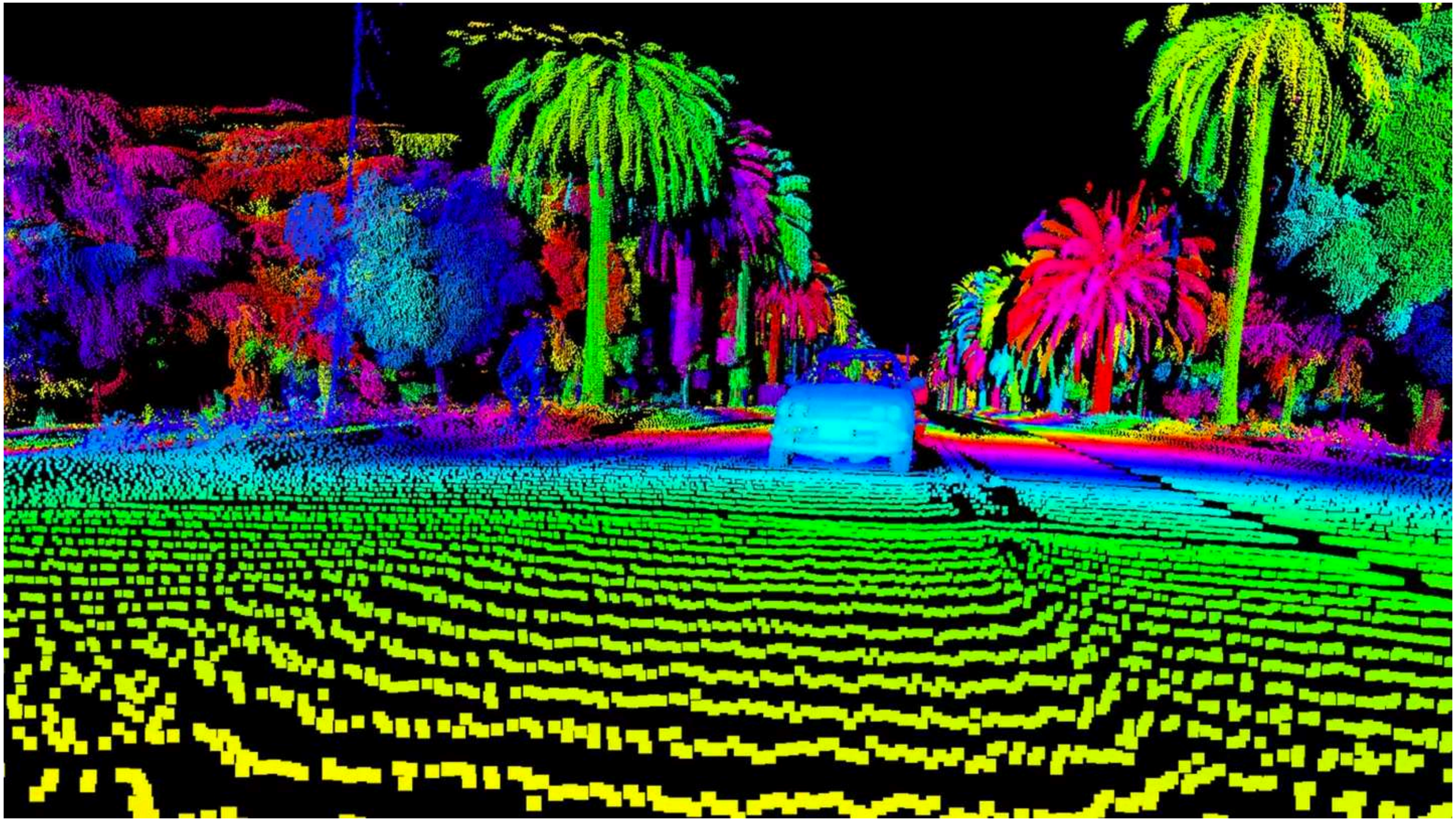
- Four-momentum – $(E, p_x, p_y, p_z)_i^\mu$
- Other quantum numbers (e.g. particle id, charge)
- Experimental information (e.g. vertex info, quality criteria, PUPPI weights)


Run: 279984
Event: 1079767163
2015-09-22 03:18:13 CEST
ATLAS
EXPERIMENT

Point Clouds

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

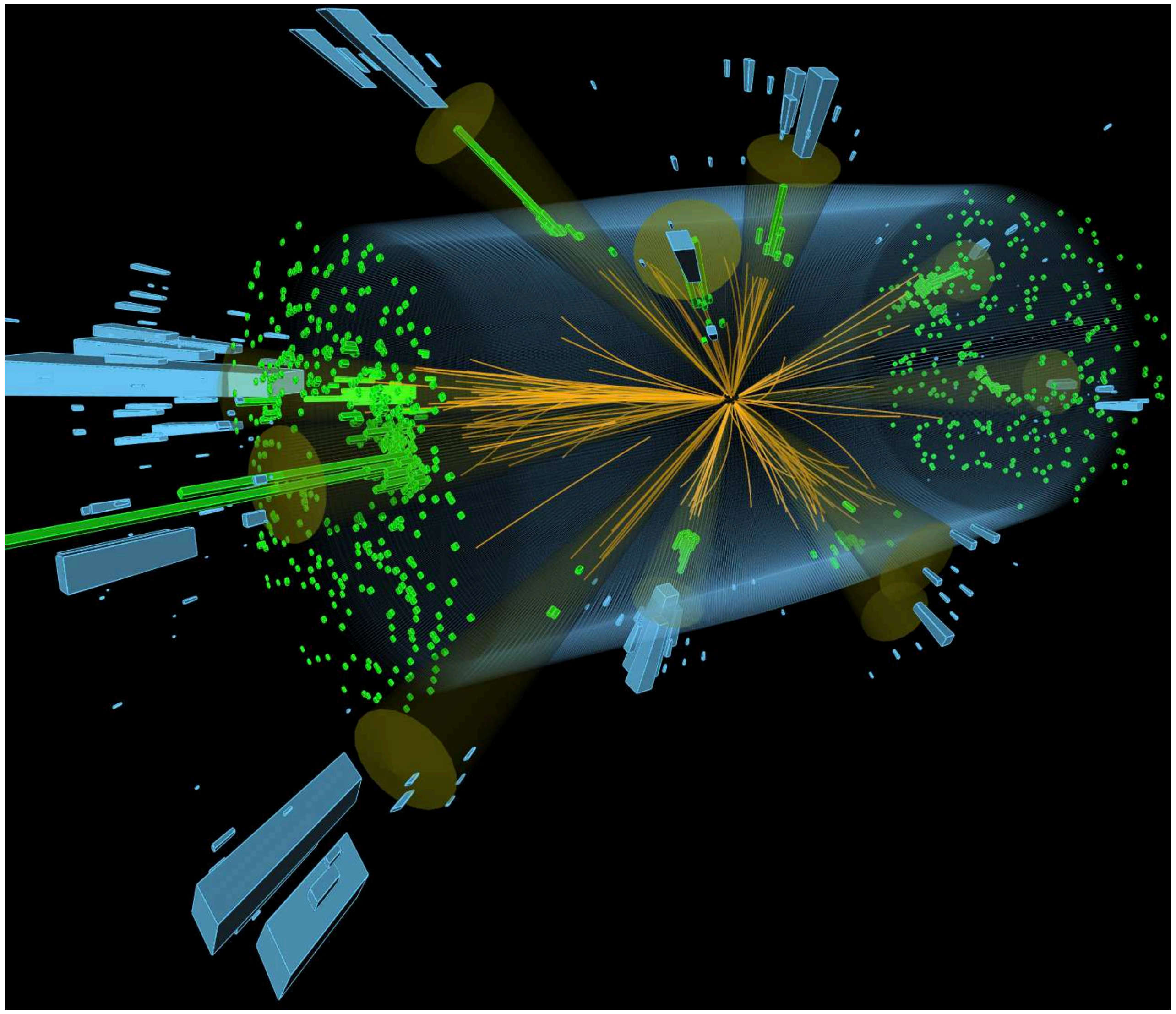
LIDAR data from self-driving car sensor



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

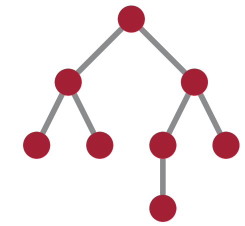
Permutation symmetric architectures

Jet Representations ⟷ Analysis Tools

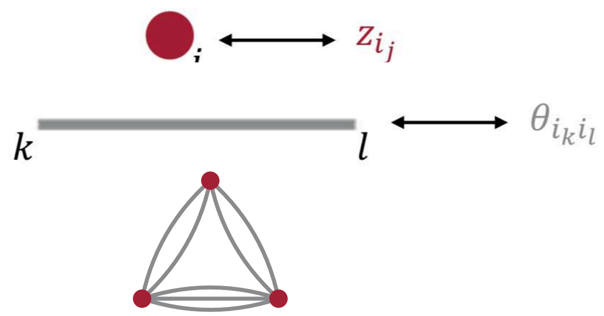
Two key choices when analyzing jets

How to represent the jet		How to analyze that representation
• Single expert observable		• Threshold cut
• A few expert observables		• Multidimensional likelihood
• Many expert observables		• Boosted decision tree (BDT), shallow neural network (NN)
• Jet images		• Convolutional NN (CNN)
• List of particles		• Recurrent/Recursive NN (RNN)
• Clustering tree		• Fancy RNN
• *N*-subjettiness basis		• Dense neural network (DNN)
• Energy flow polynomials	PTK ML4Jets 2017	• Linear classification
• Set of particles	PTK ML4Jets 2018	• Energy flow network

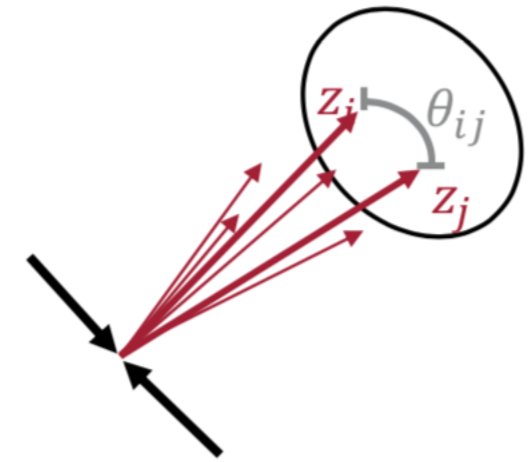
Energy Flow Polynomials (EFPs)

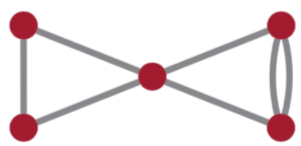


[PTK, Metodiev, Thaler, 1712.07124]



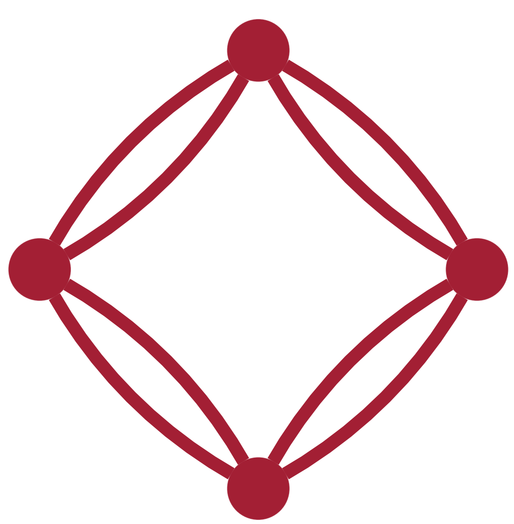
$$\text{EFP}_G = \underbrace{\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \text{ of } \underbrace{z_{i_1} \cdots z_{i_N}}_{\text{Energies}} \text{ and } \underbrace{\prod_{(k,\ell)\in G} \theta_{i_k i_\ell}}_{\text{Angles}}$$

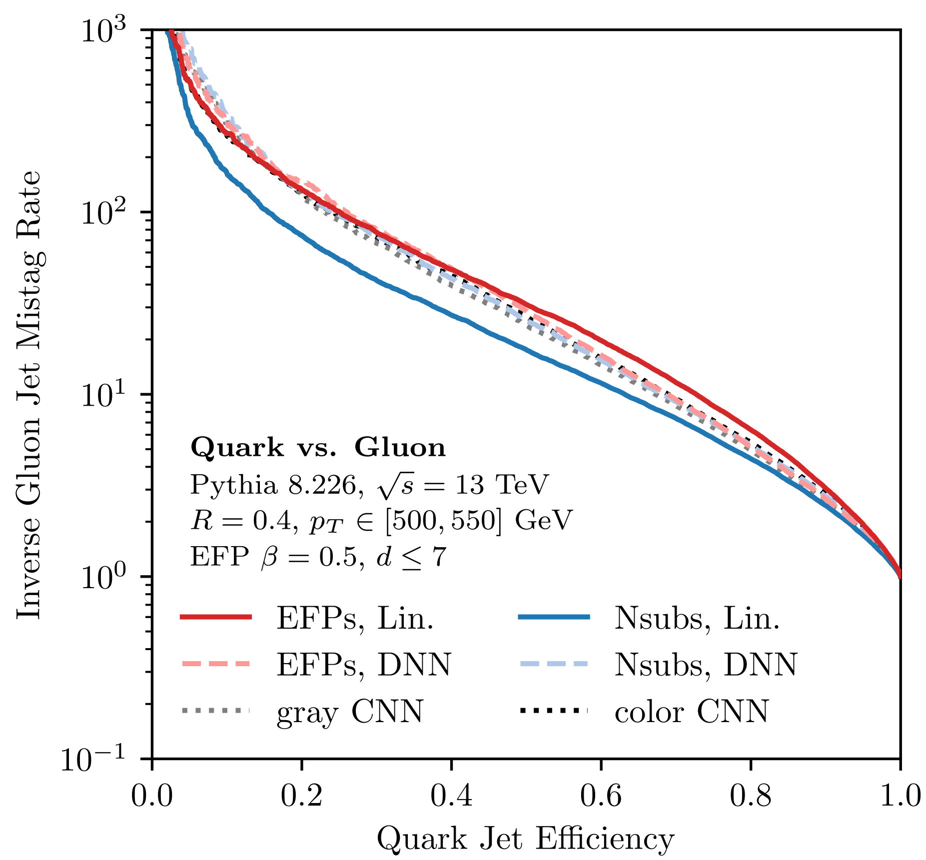


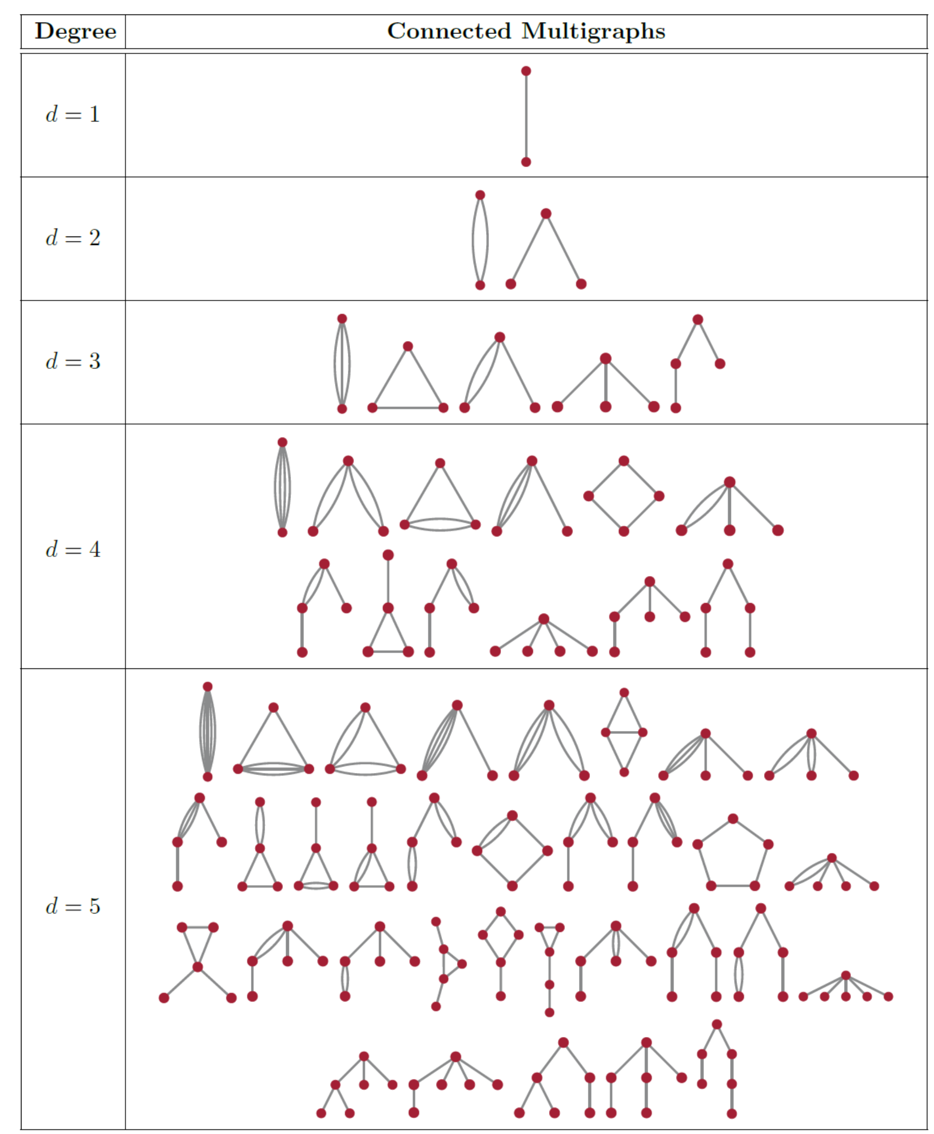


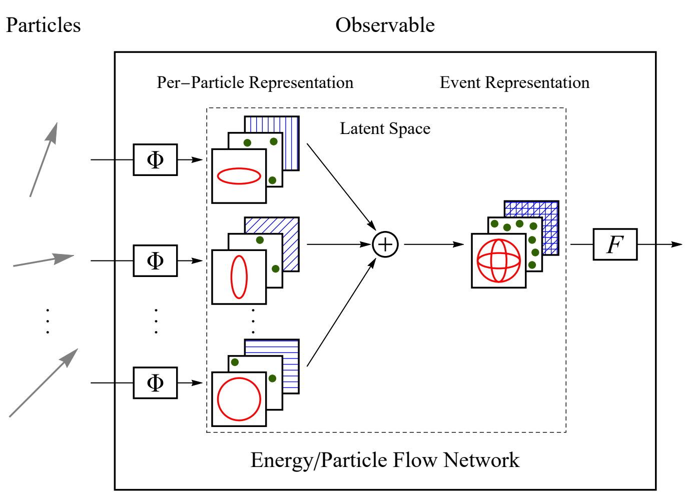
$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_1 i_3} \theta_{i_1 i_4} \theta_{i_1 i_5} \theta_{i_4 i_5}^2$$

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









Jets as Point Clouds

Energy Flow Networks

Quark vs. Gluon Tagging

Deep Sets

Namespace for symmetric function parametrization

A general permutation-symmetric function is *additive* in a latent space

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

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 $\mathfrak{X} \subset \mathbb{R}^d$ 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]

Permutation invariance

Latent space

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

General parametrization for a function of sets

Deep Sets for Particle Jets

[PTK, Metodiev, Thaler, 1810.05165]

Particle Flow Network (PFN)

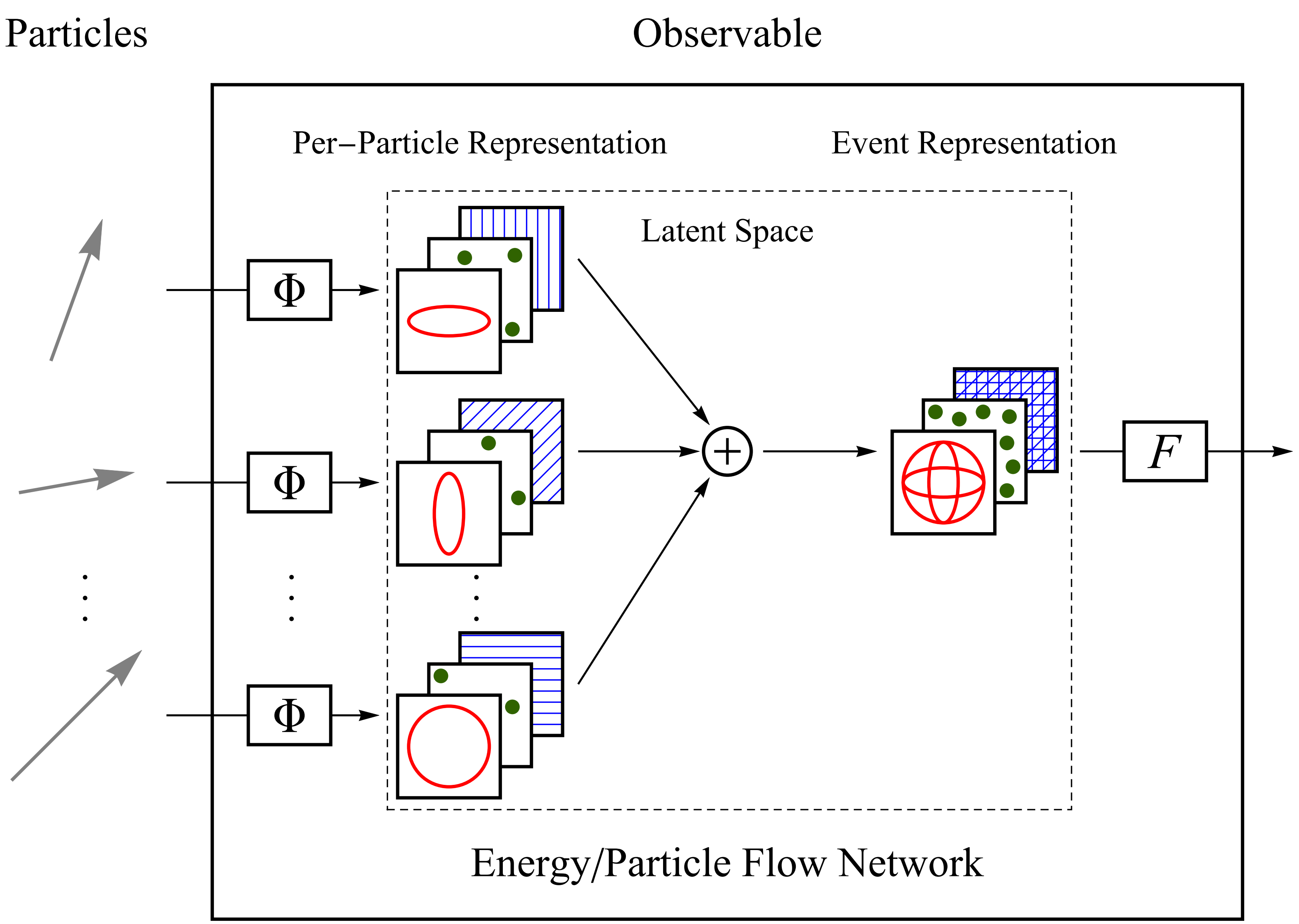
$$\mathrm{PFN}(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M}\Phi(p_i^\mu)\right)$$

Fully general latent space

Energy Flow Network (EFN)

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

IRC-safe latent space



Latent Space IRC Safety

Latent space defines new physics observables

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

QCD has soft and collinear divergences associated with gluon radiation

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

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

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

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

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

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

Latent Space IRC Safety

Latent

IRC saf

QCD h

IRC safety is a statement of *linearity* in energy and *continuity* in geometry

Theorem: A generic function of four-momenta can be made IRC safe via the following replacement:

$$\sum_{i=1}^{M} f(p_i^\mu) \longrightarrow \sum_{i=1}^{M} z_i f(\hat{p}_i).$$

Proof: In 1810.05165. $\square$

Infrare

Colline

e

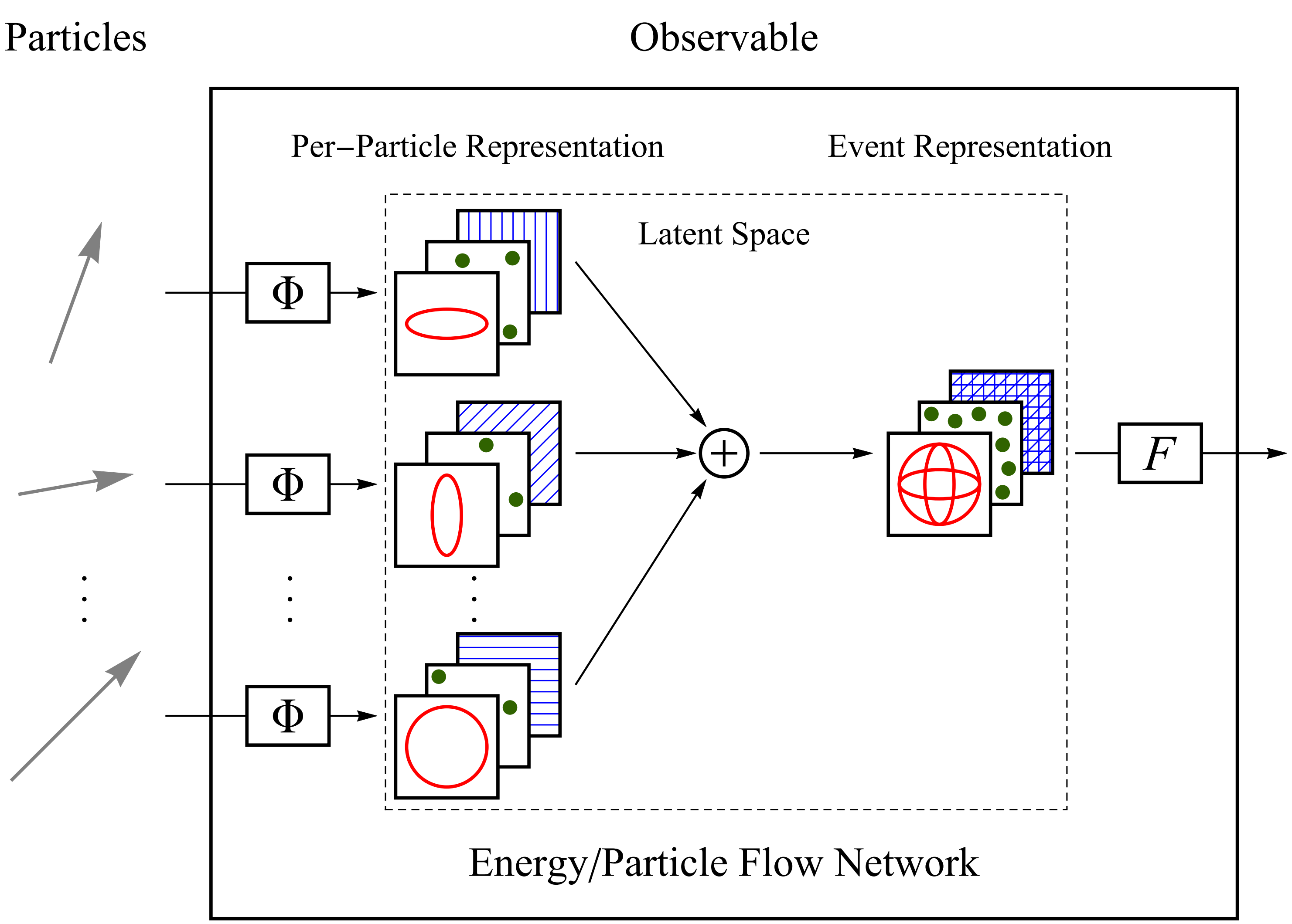
$$S(\{p_1^\mu,\ldots,p_M^\mu\}) = S(\{p_1^\mu,\ldots,(1-\lambda)p_M^\mu,\lambda p_{M+1}^\mu\}), \quad \forall\,\lambda\in[0,1]$$

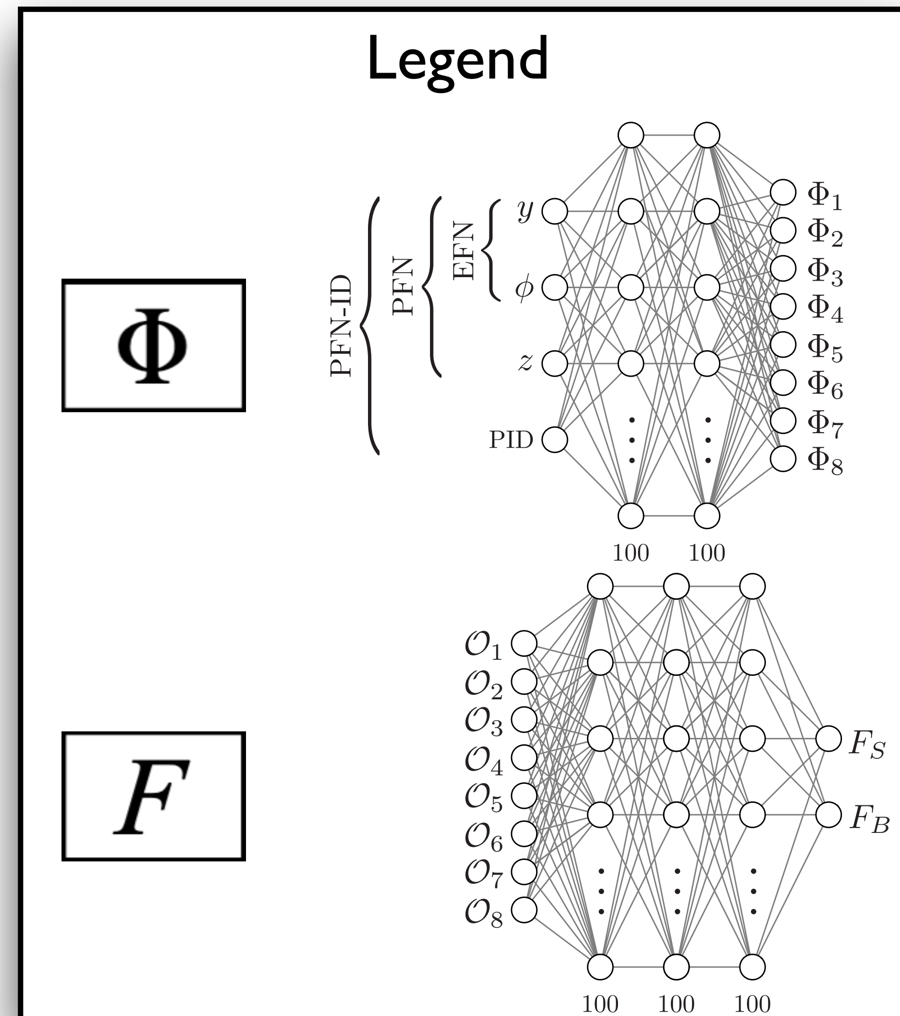
Approximating Φ and *F* with Neural Networks

Employ neural networks as arbitrary function approximators

Use fully-connected networks for simplicity

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



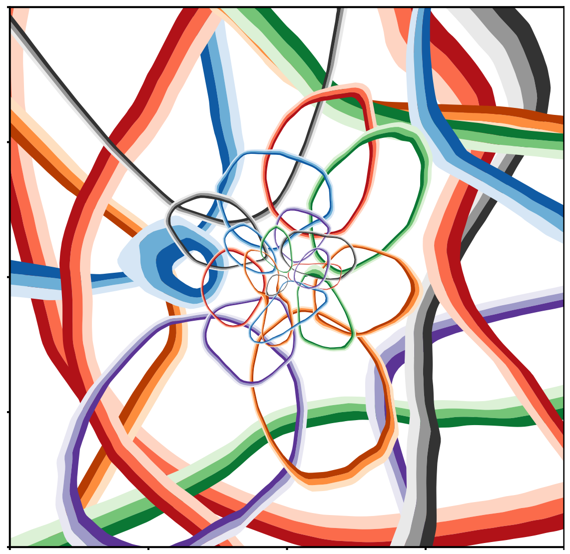


$$\text{EFN}: \quad \mathcal{O}_a = \sum_{i=1}^{M} z_i \Phi_a(y_i, \phi_i)$$

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

Jets as Point Clouds

Energy Flow Networks



Quark vs. Gluon Tagging

Classification Performance

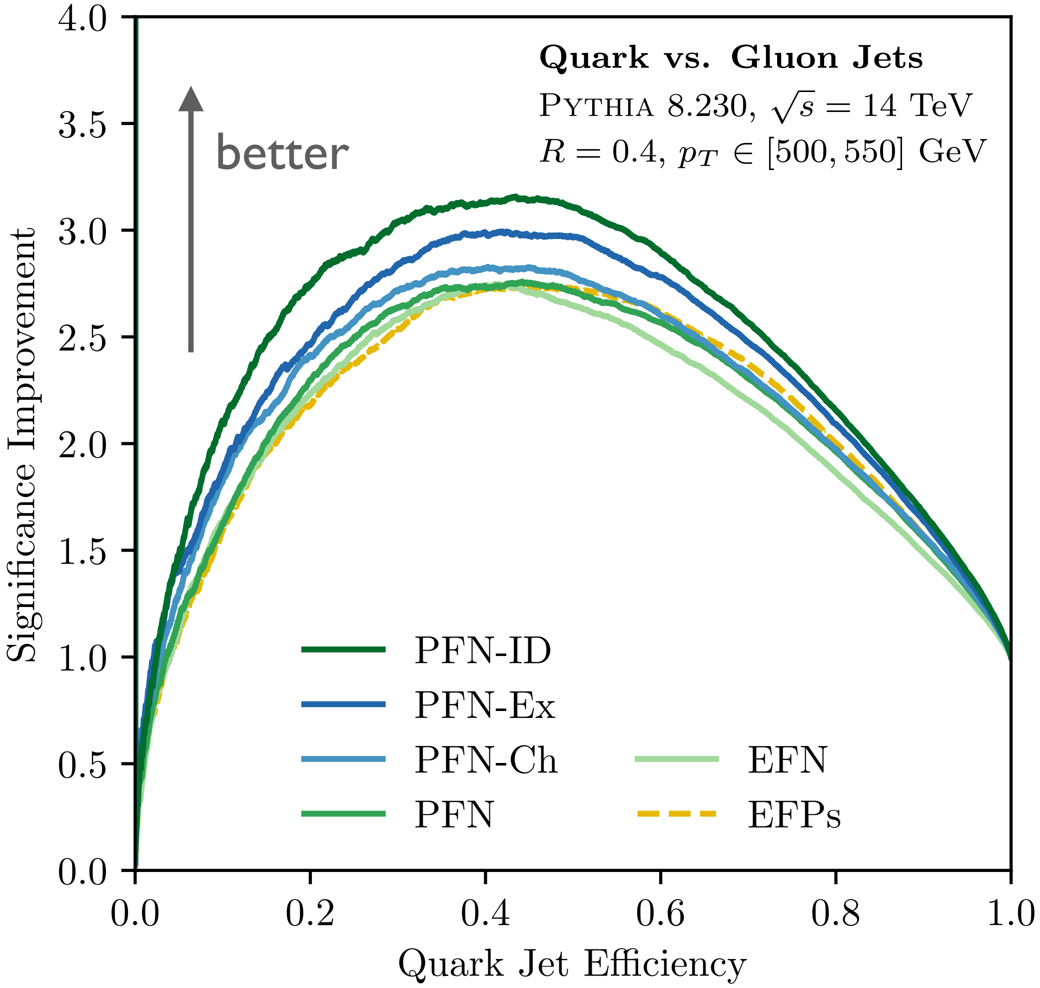
PFN-ID: Full particle flavor info
$(\gamma, \pi^{\pm}, K^{\pm}, K_L, p, \bar{p}, n, \bar{n}, e^{\pm}, \mu^{\pm})$

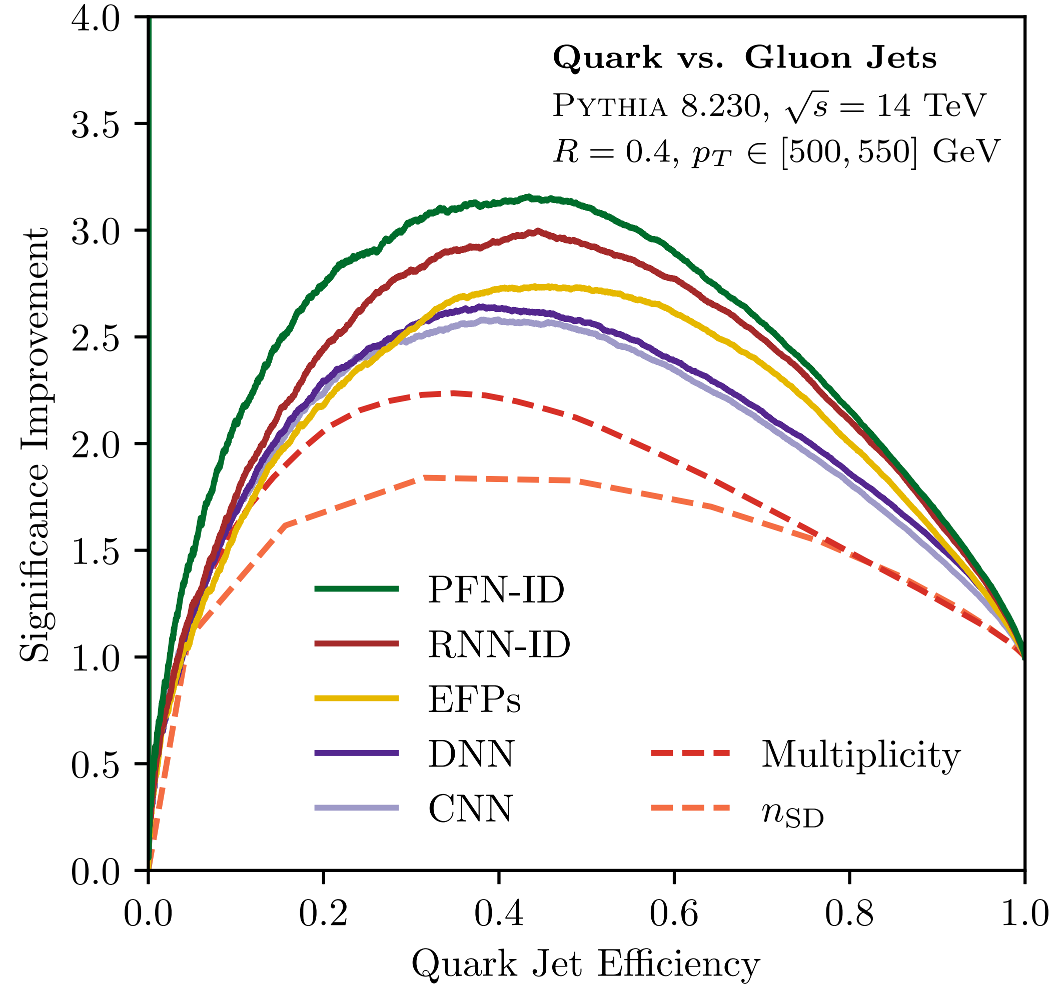
PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^{\pm}, \mu^{\pm})$

PFN-Ch: Particle charge info
(+,0,–)

PFN: No particle type info, arbitrary energy dependence

EFN: IRC-safe latent space





Latent space dimension ℓ = 256

EFPs are comparable to EFN

PFN-ID slightly better than RNN-ID

EFN Latent Dimension Sweep

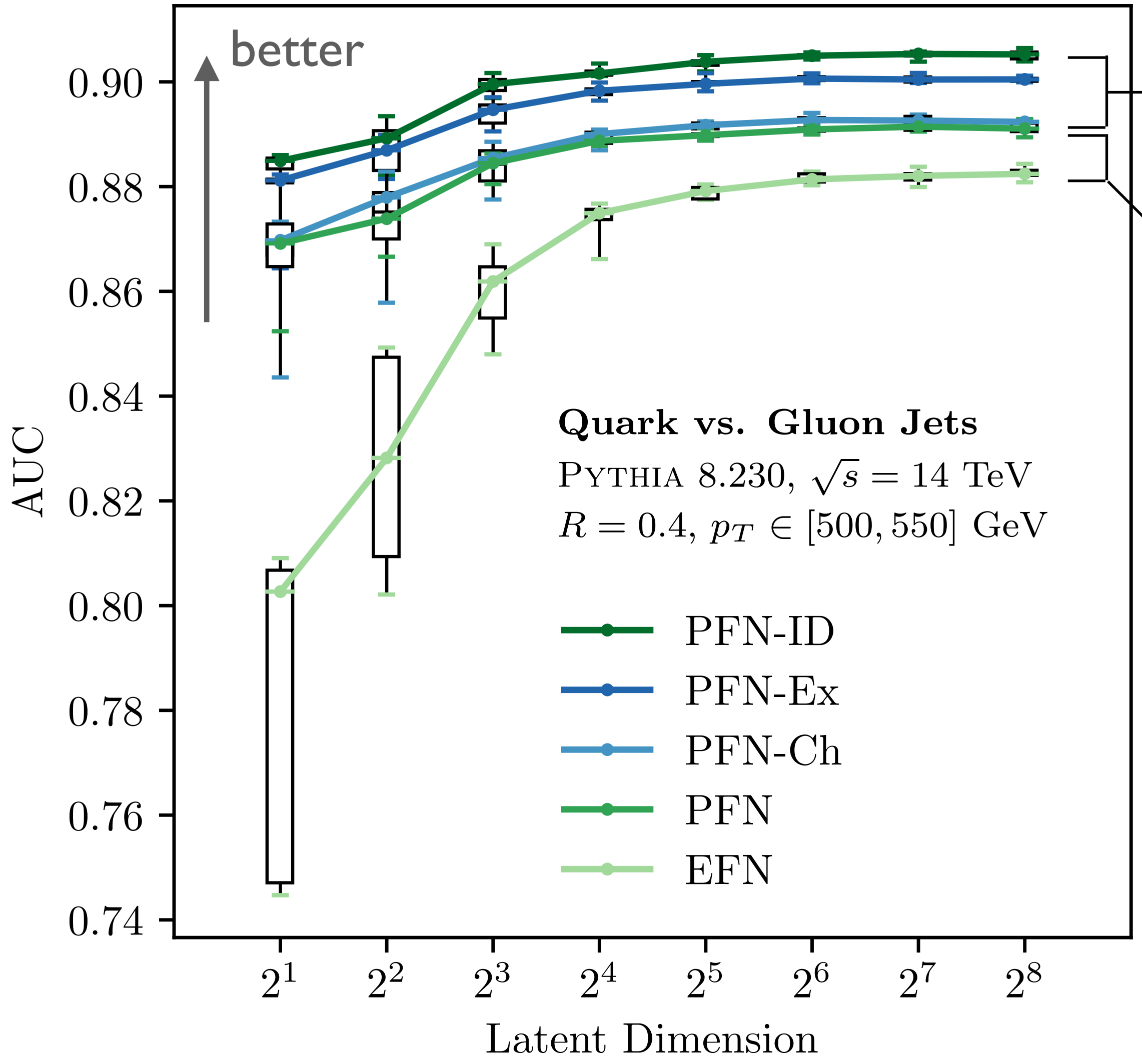
PFN-ID: Full particle flavor info
$(\gamma, \pi^{\pm}, K^{\pm}, K_L, p, \bar{p}, n, \bar{n}, e^{\pm}, \mu^{\pm})$

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: IRC-safe latent space



Particle type info helpful

IRC unsafe information clearly helpful

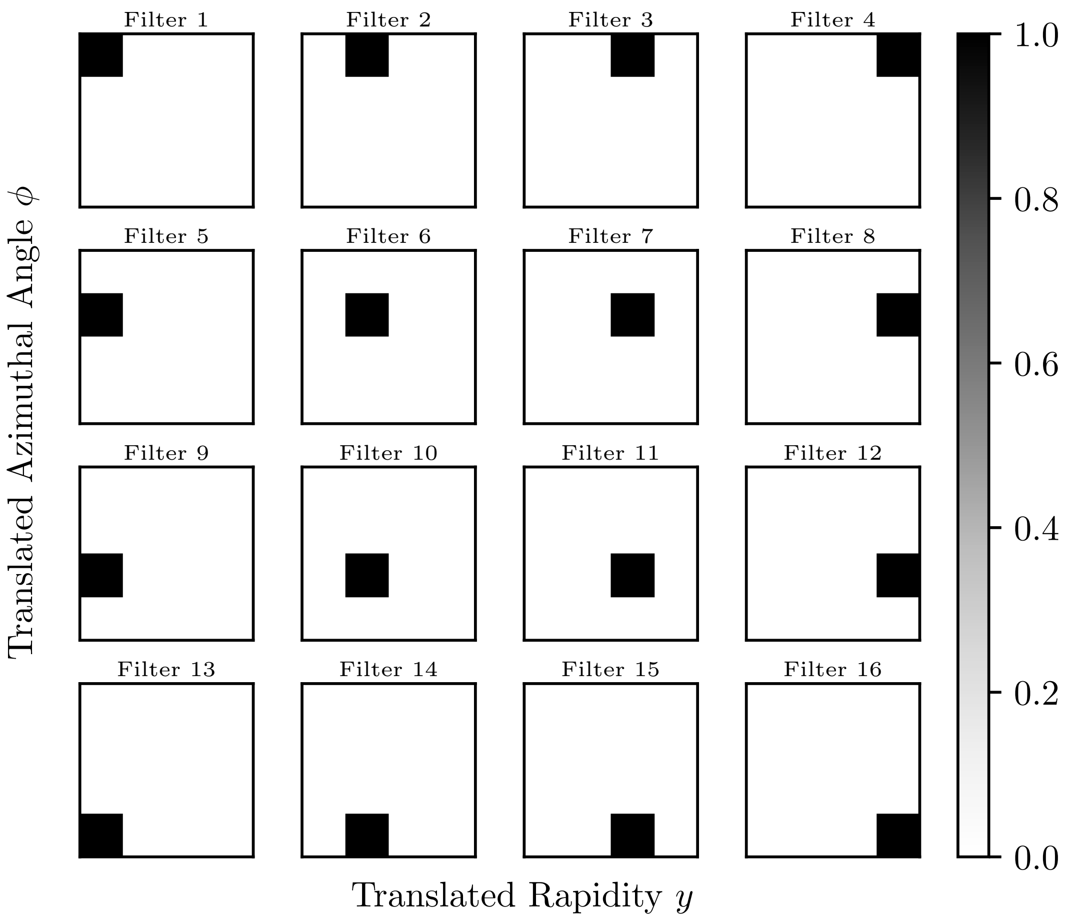
Performance saturates as dimension increases

Energy Flow Network Visualization

EFN observables are two-dimensional geometric functions

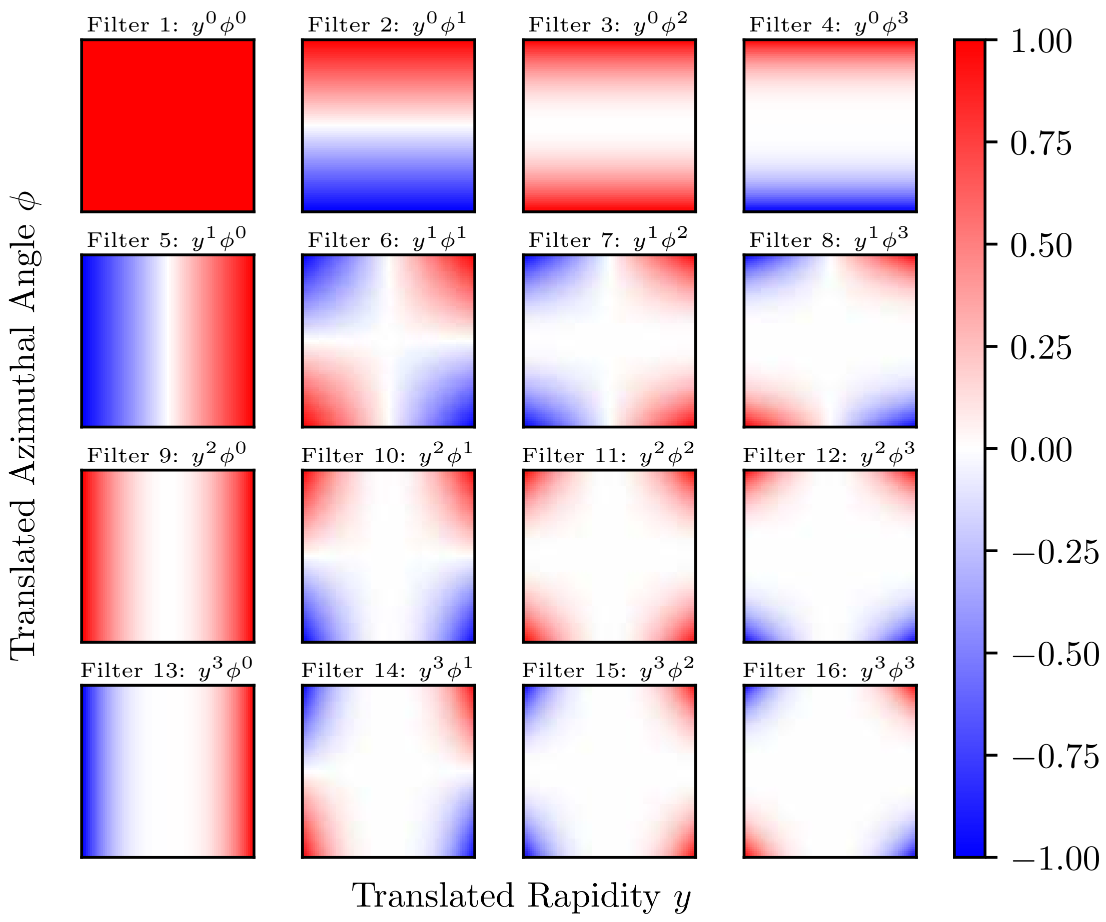
Visualize EFN observables as *filters* in the translated rapidity-azimuth plane

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]

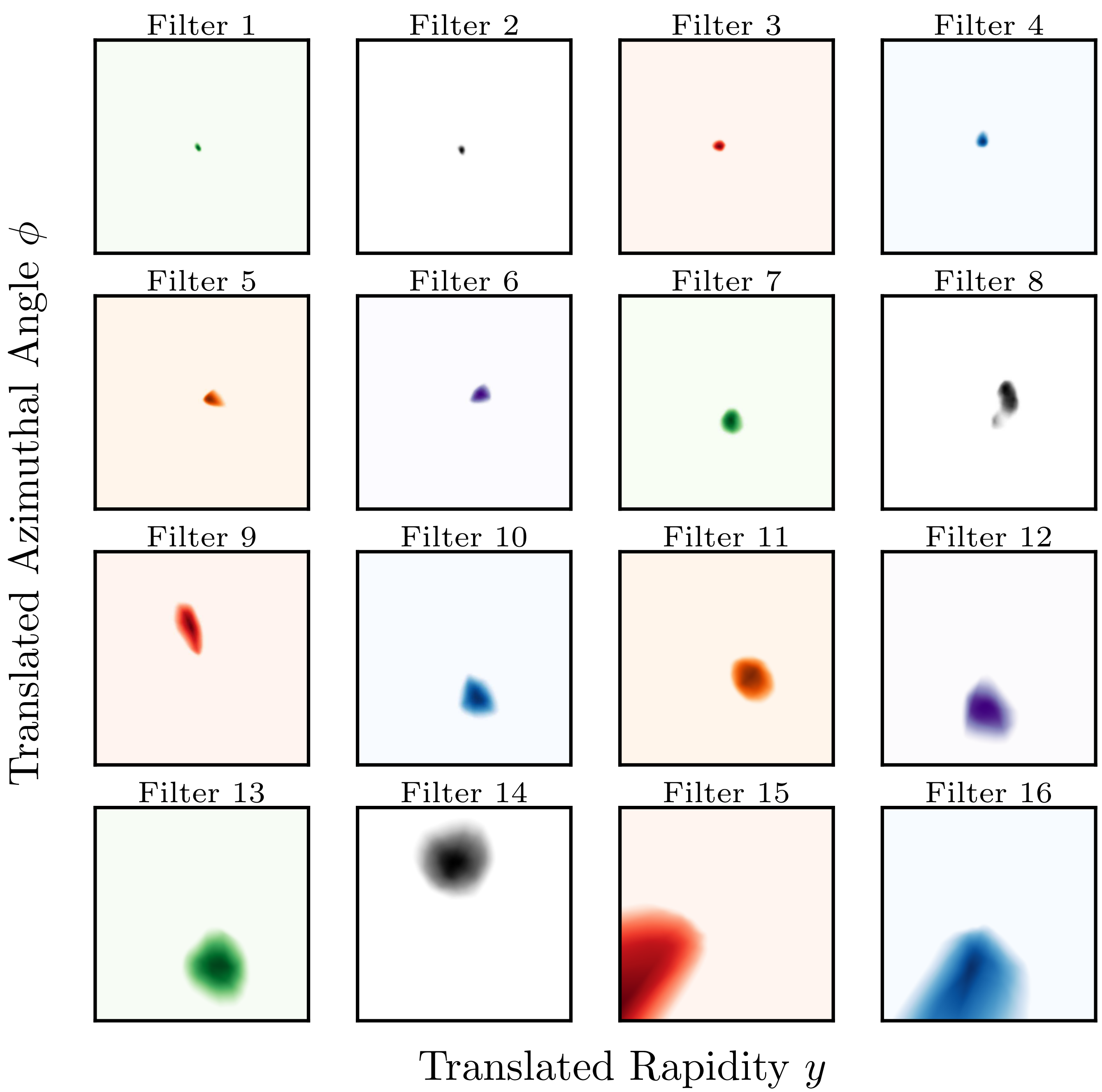
Visualizing Q/G EFN Filters

Generally see blobs of all scales

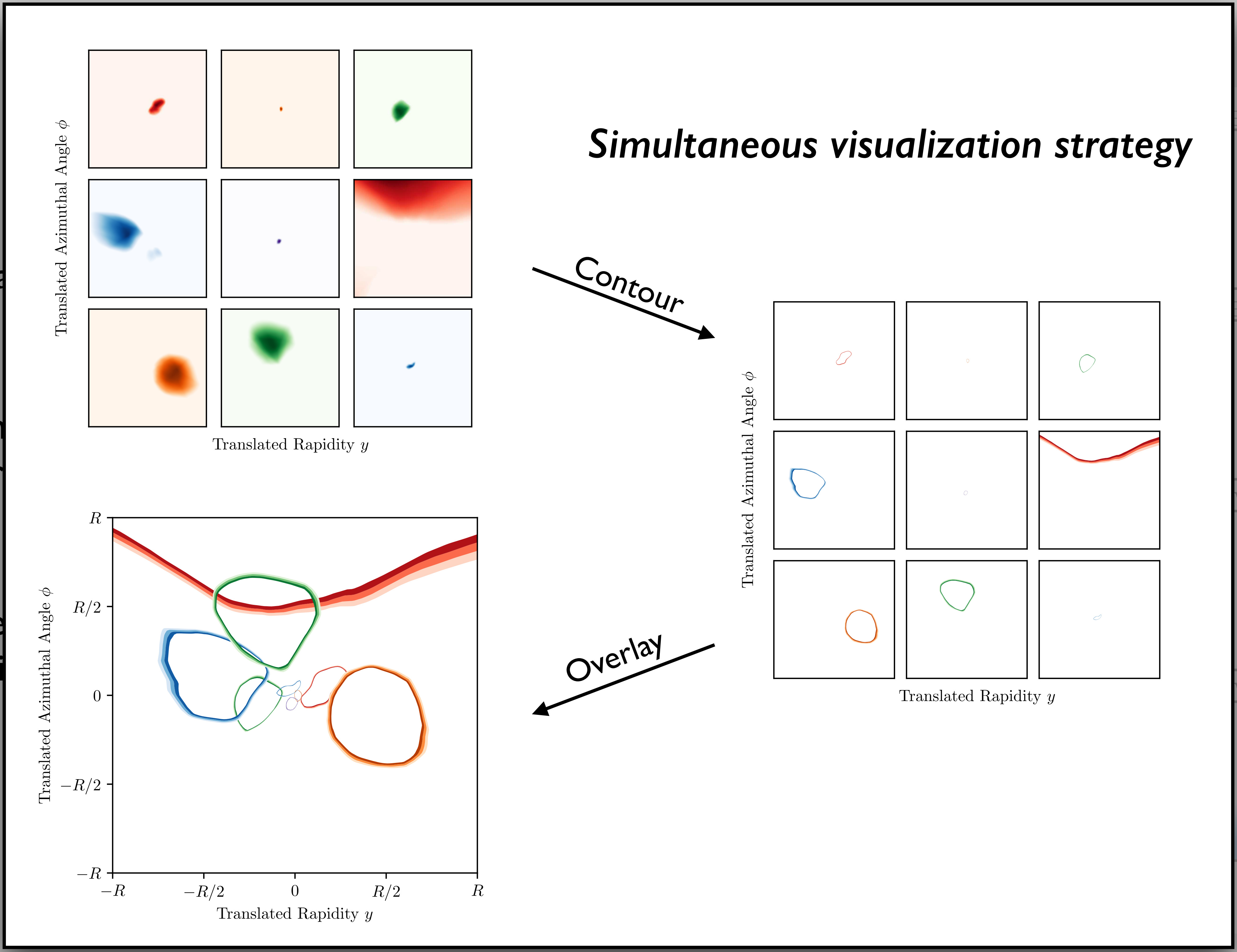
Local nature of activated region lends interpretation as "pixels"

EFN seems to have learned a dynamically sized jet image

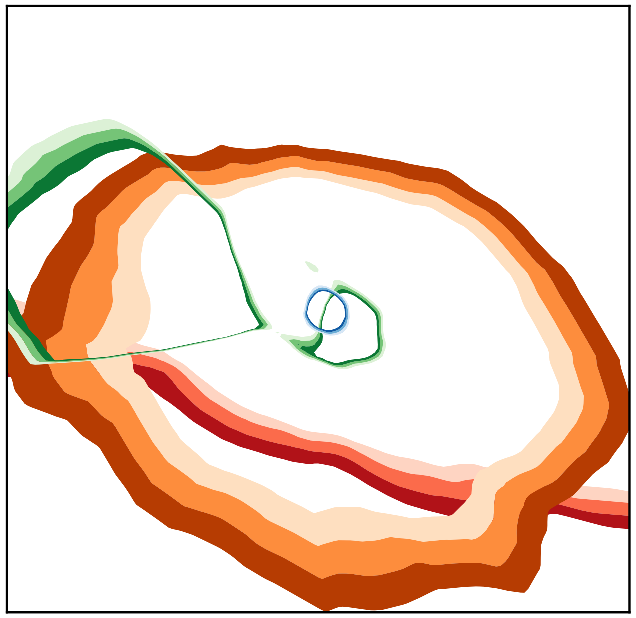
EFN (ℓ = 256) randomly selected filters, sorted by size



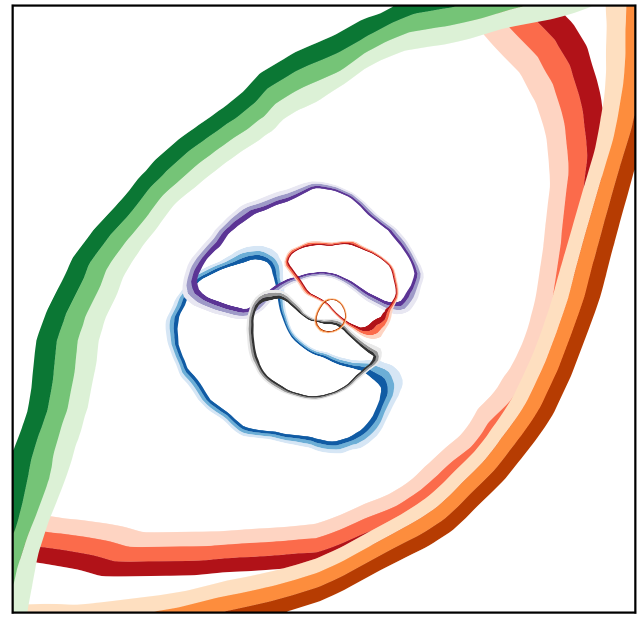
Visualizing Q/G EFN Filters



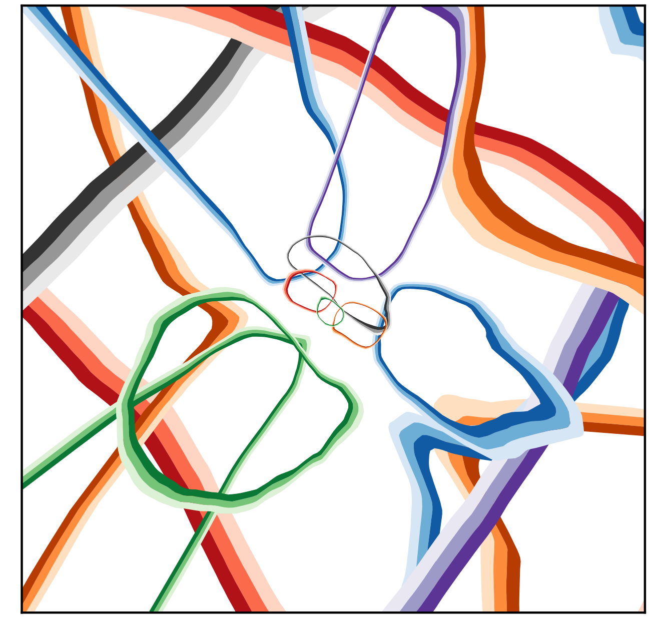
Visualizing Q/G EFN Filters



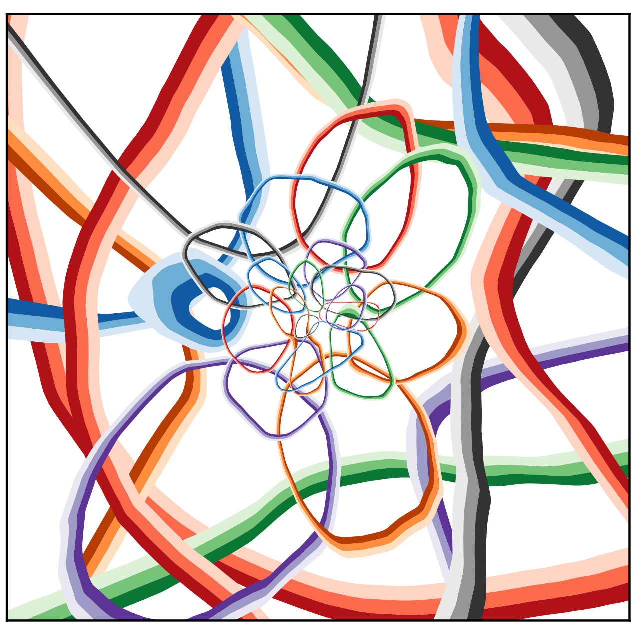
$\ell = 4$



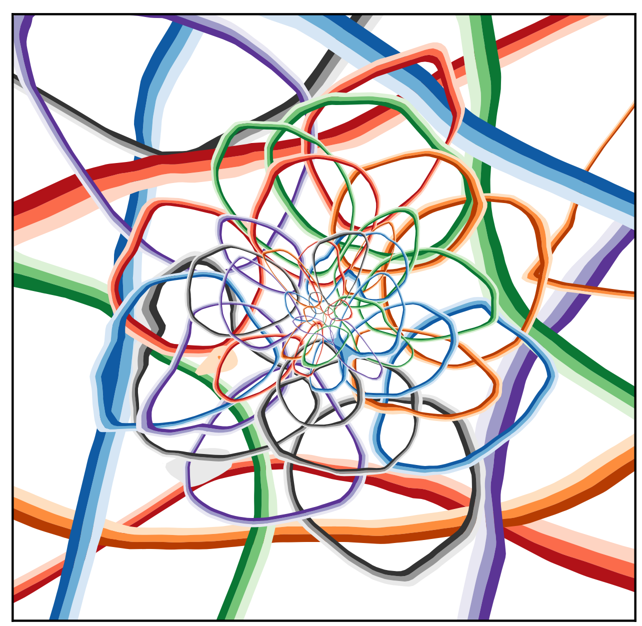
$\ell = 8$



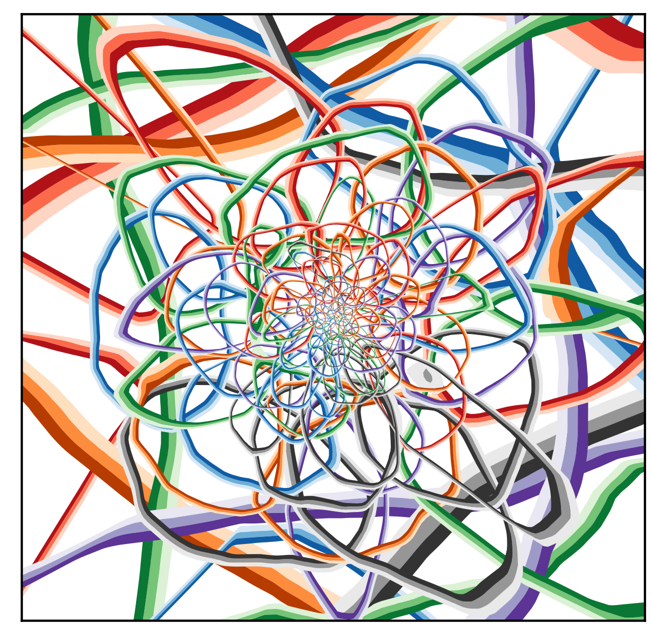
$\ell = 16$



$\ell = 32$

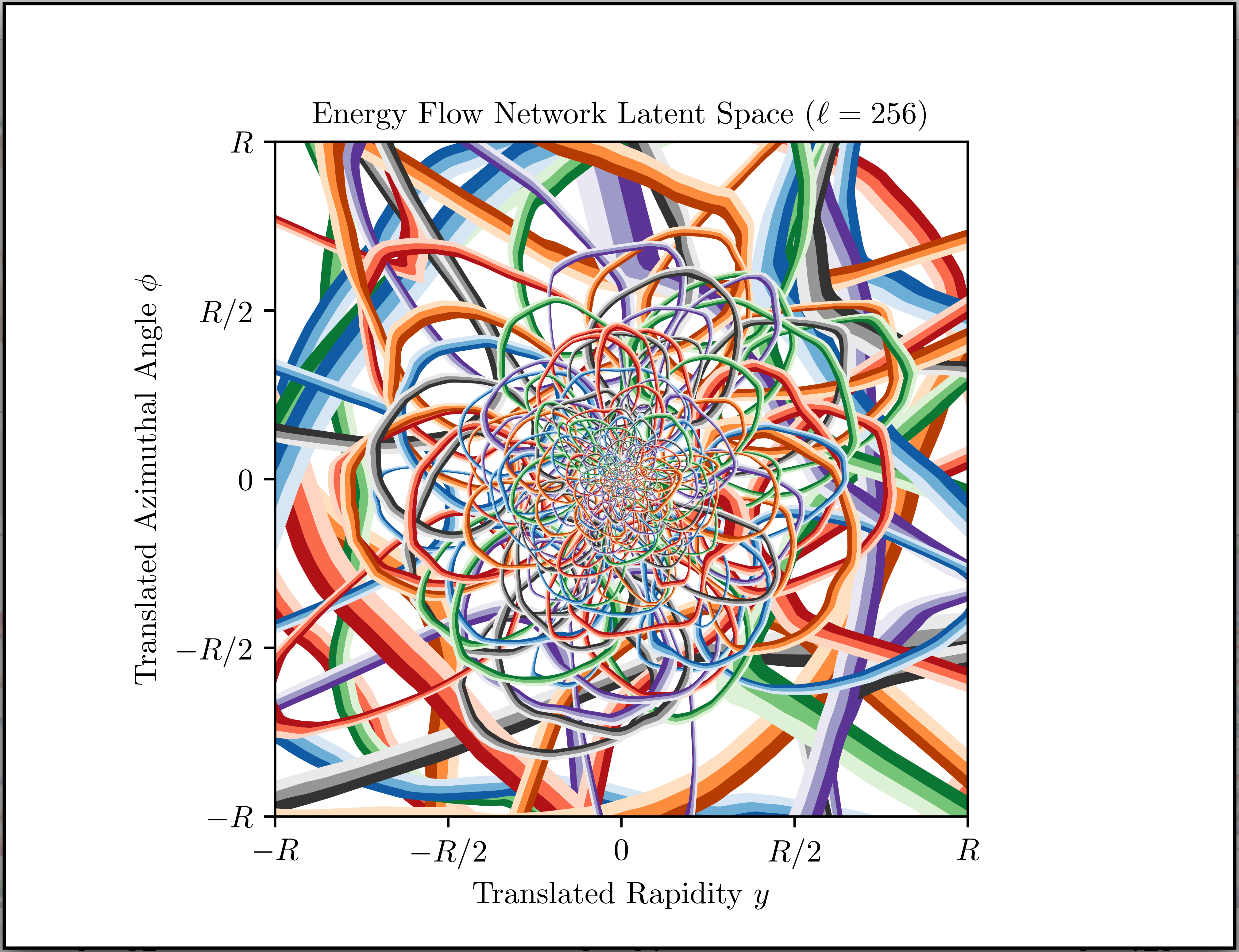


$\ell = 64$



$\ell = 128$

Visualizing Q/G EFN Filters



Measuring Q/G EFN Filters

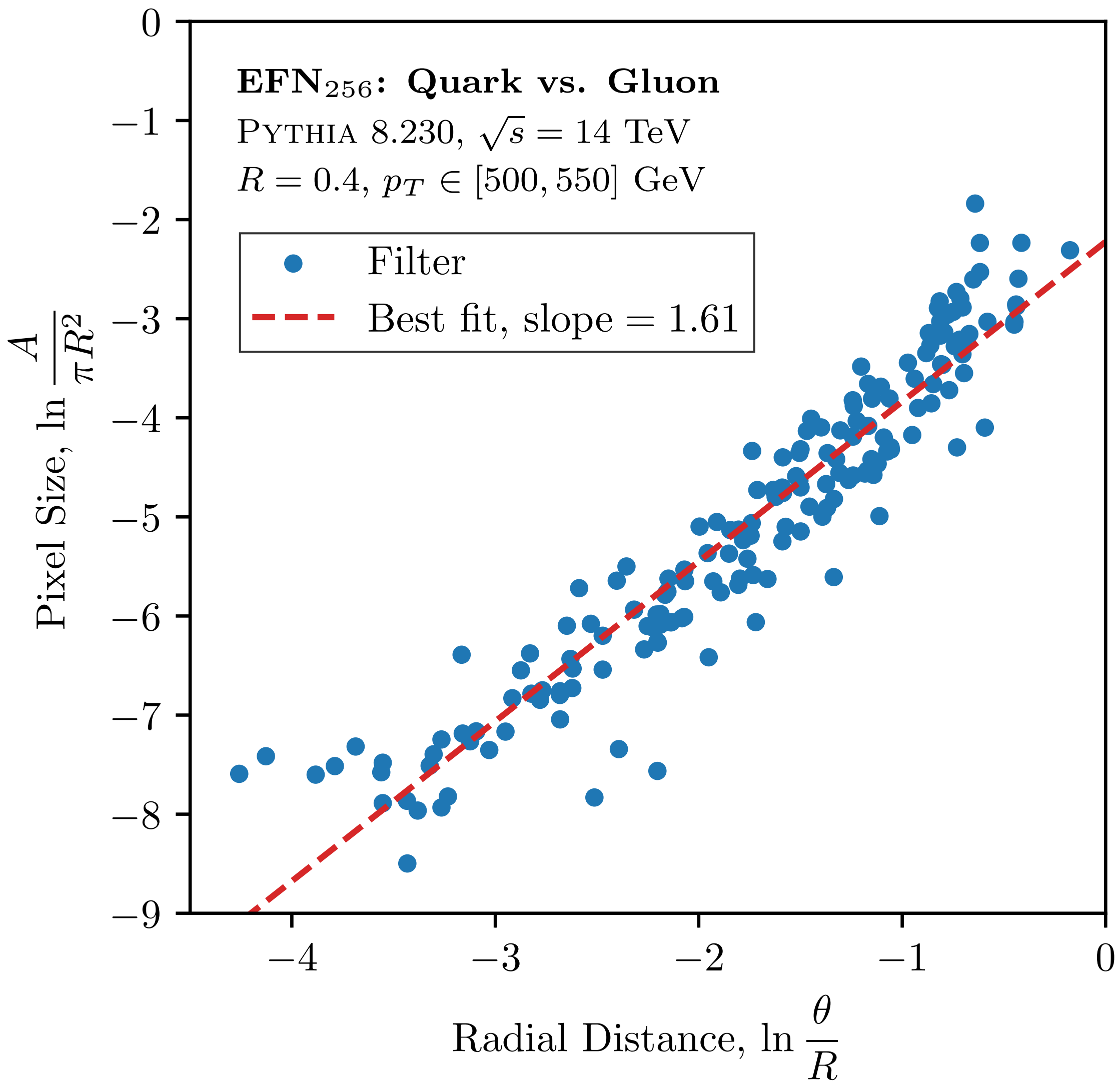
Power-law dependence between filter size and distance from center is observed

Slope of 2 is predicted at leading log

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

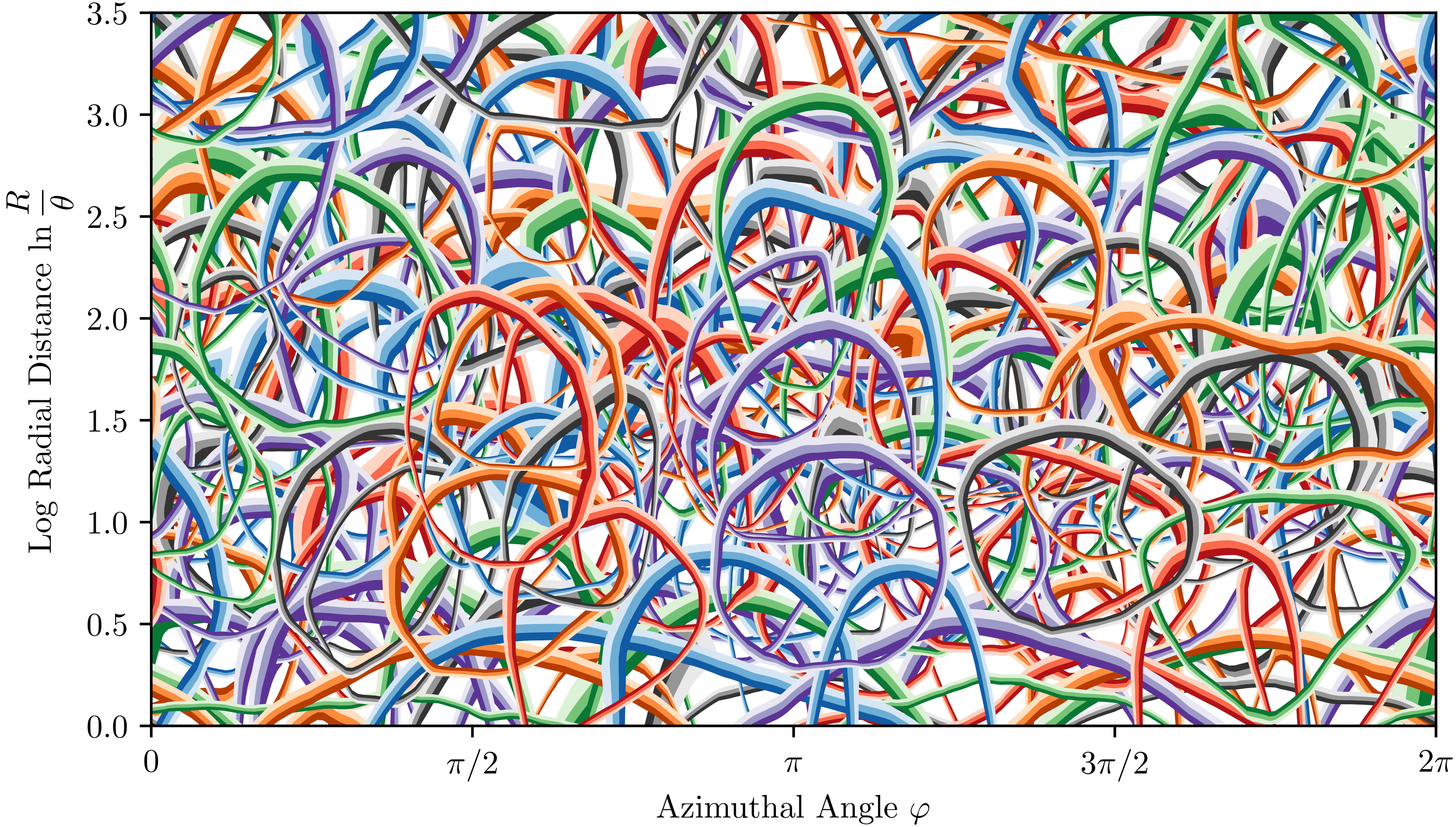
Area element in rap-phi plane

Emission plane area element

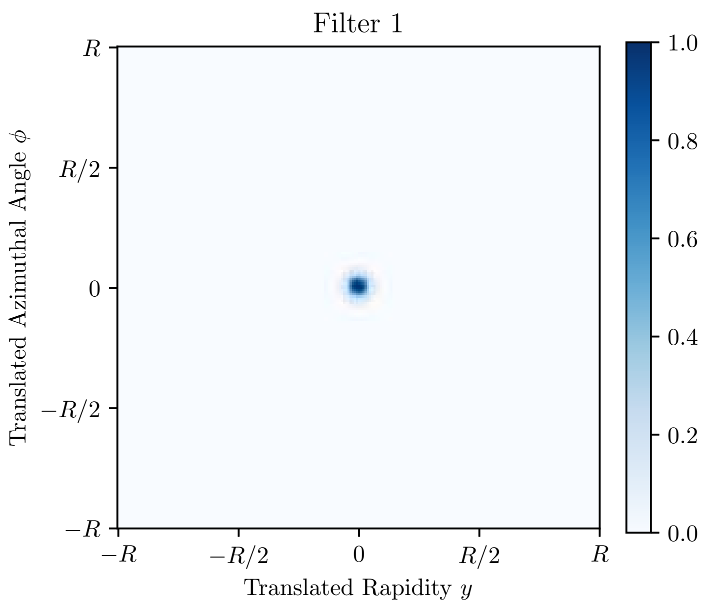


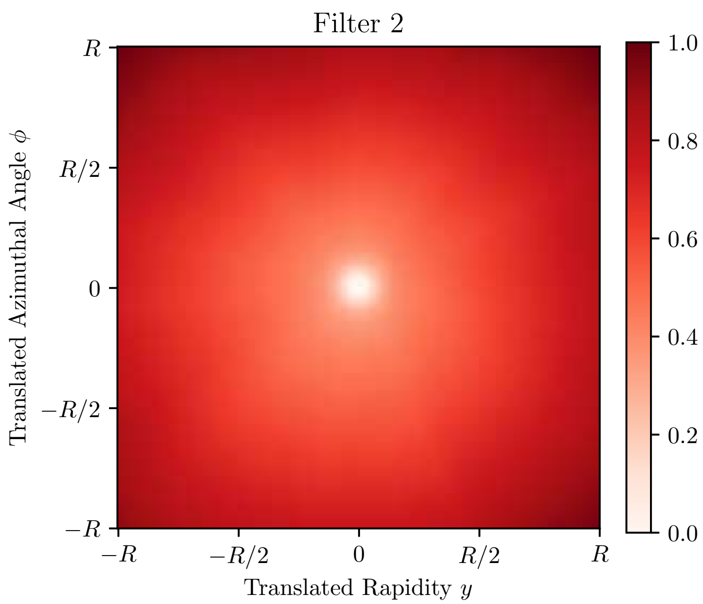
Non-perturbative physics, axis recoil, higher order effects cause deviations from slope of 2

Visualizing Q/G EFN Filters in the Emission Plane

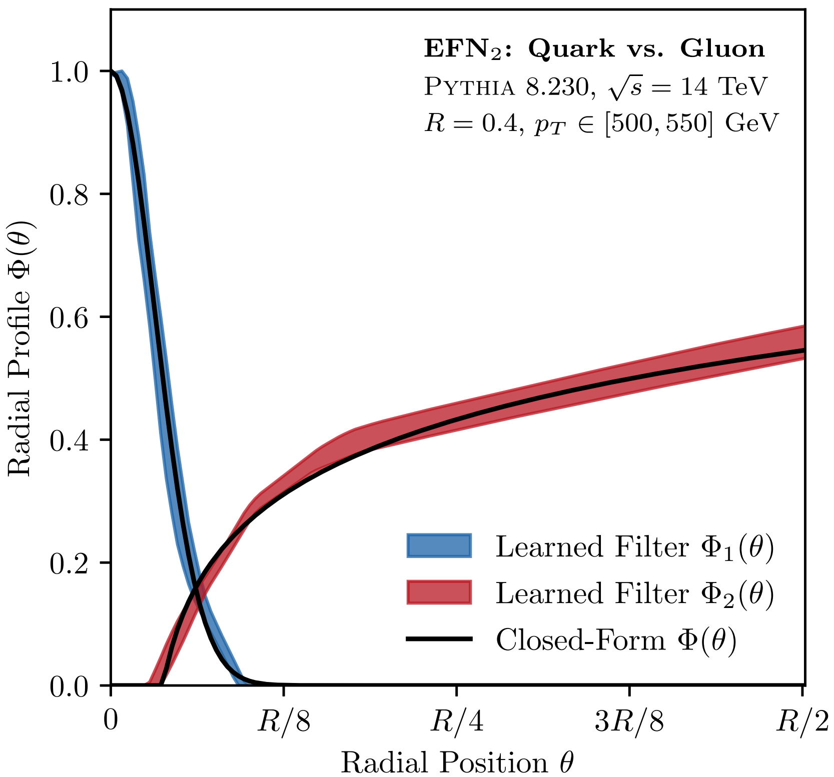


Extracting New Analytic Observables





$$\mathcal{O}_1 = \sum_{i=1}^{M} z_i \, \Phi_1(\theta_i) \qquad \mathcal{O}_2 = \sum_{i=1}^{M} z_i \, \Phi_2(\theta_i)$$



Take radial slices to obtain envelope

EFN (ℓ = 2) has approximately radially symmetric filters

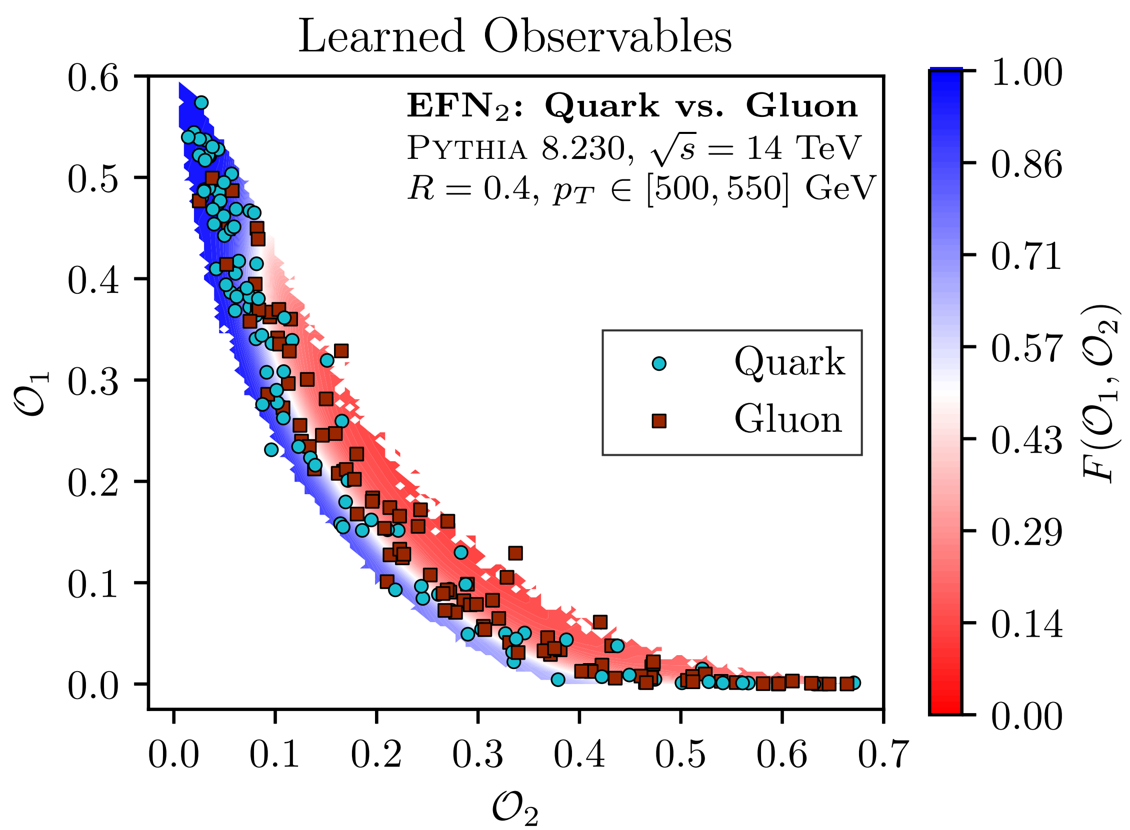
Fit functions of the forms:

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

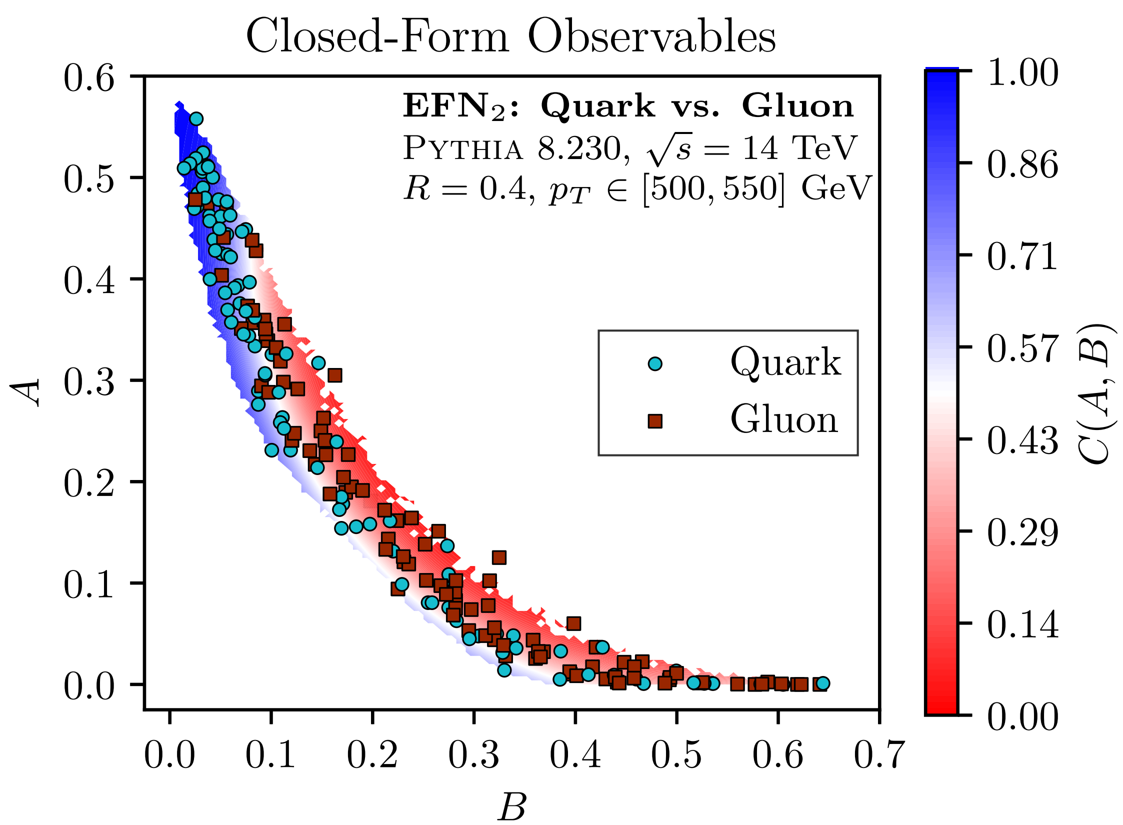
Separate soft and collinear phase space regions

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:

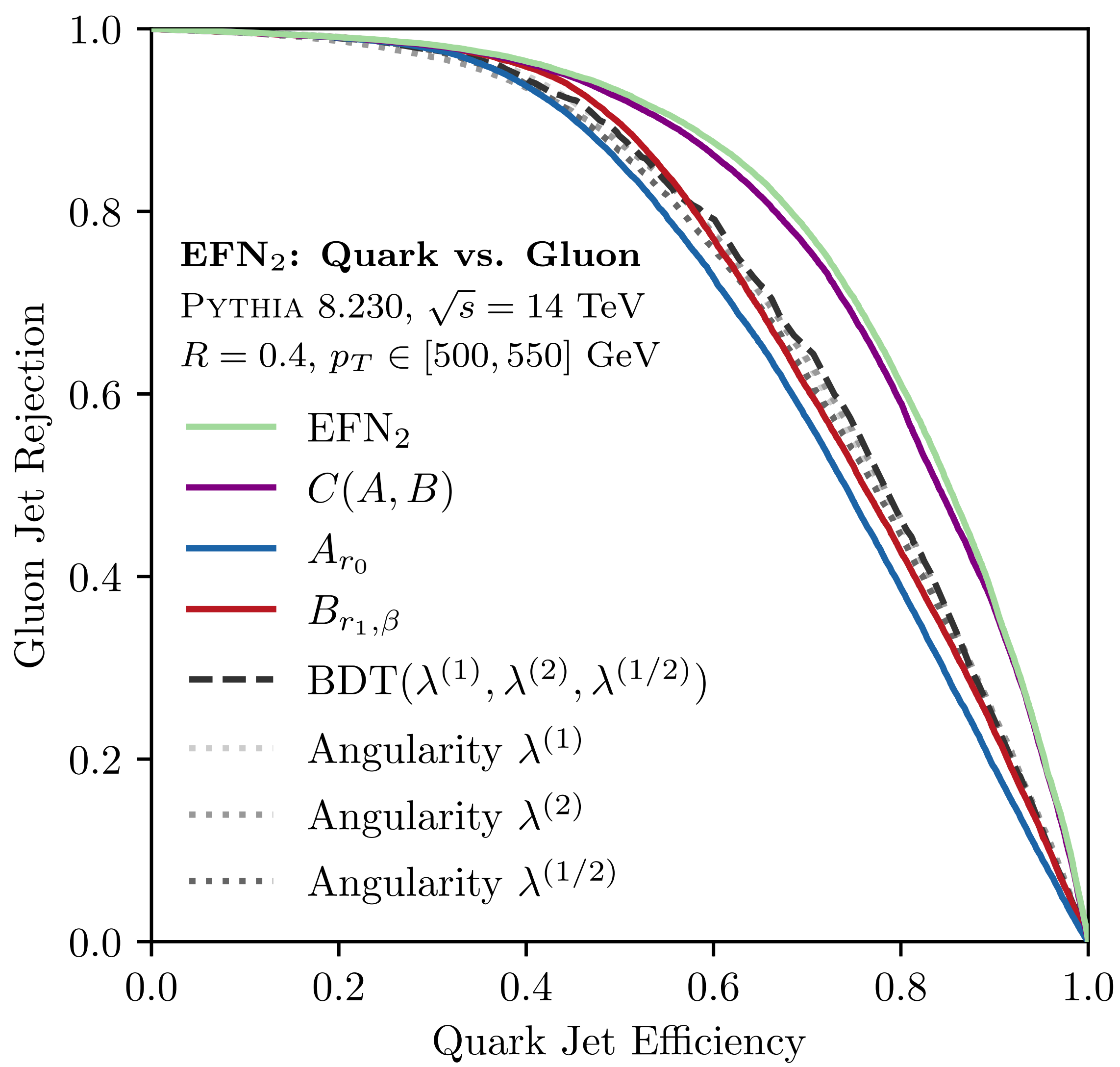
$$C(A, B) = (A - a_0)^2 + (B - b_0)^2$$

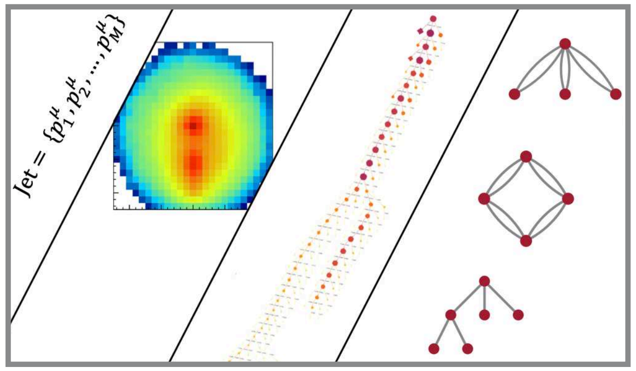
Benchmarking New Analytic Observables

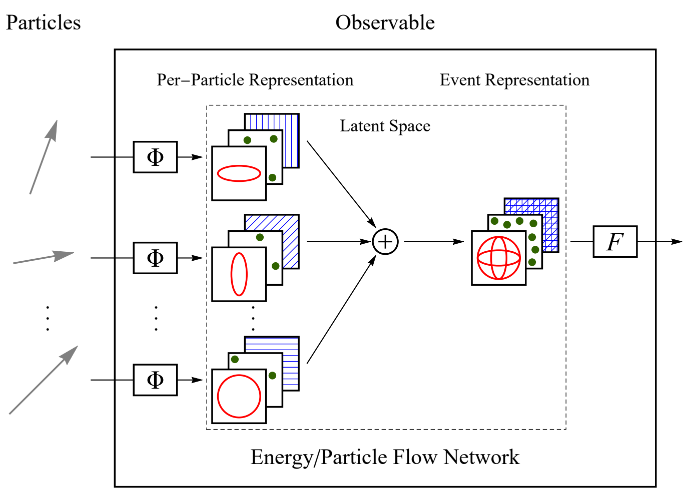
Individually, extracted observables are comparable to other angularities

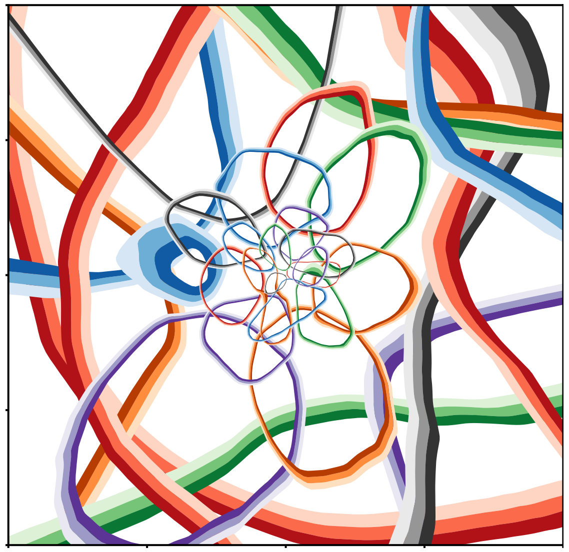
Extracted $C(A, B)$ performs nearly as well as EFN (ℓ = 2)

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









Jets as Point Clouds

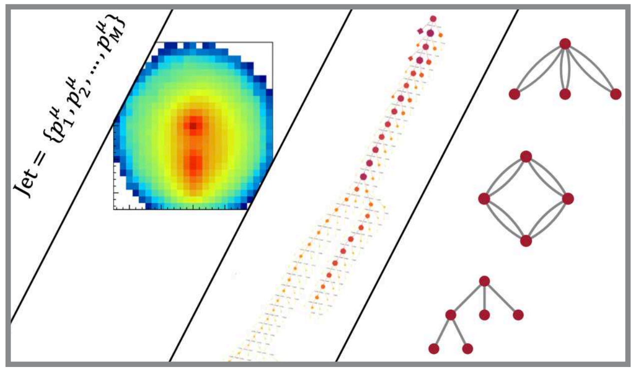
Point clouds have variable size and permutation symmetry

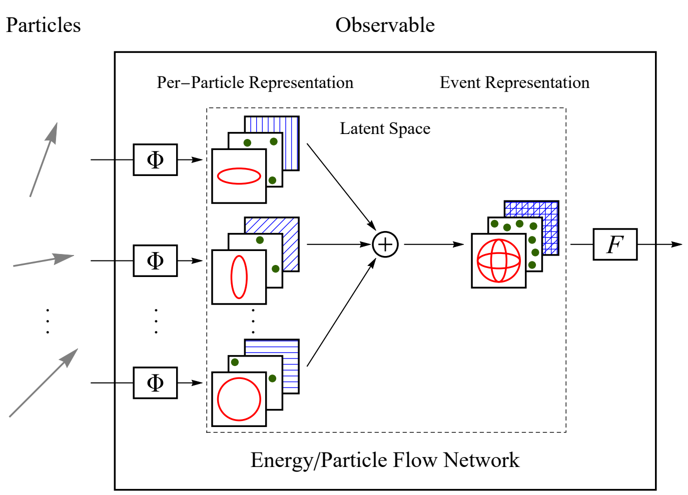
Energy Flow Networks

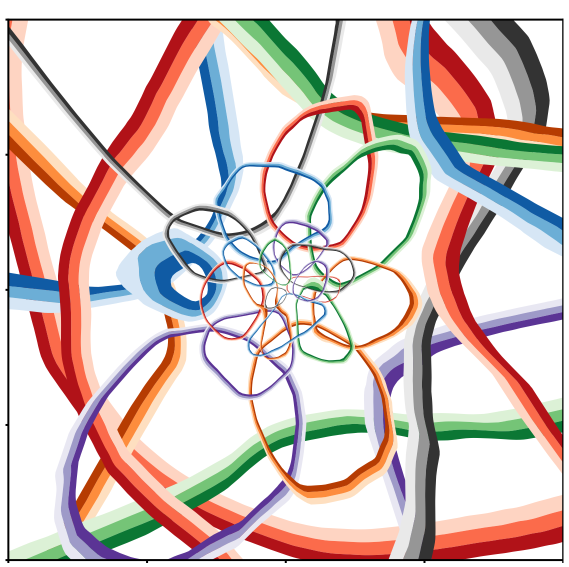
Deep Sets architecture, IRC-safe latent space

Quark vs. Gluon Tagging

Performance, visualization, new analytic observables







Jets as Point Clouds

Point clouds have variable size and permutation symmetry

Energy Flow Networks

Deep Sets architecture, IRC-safe latent space

Quark vs. Gluon Tagging

Performance, visualization, new analytic observables

Versatility? Transparency? Verifiability? Robustness? Deployment?

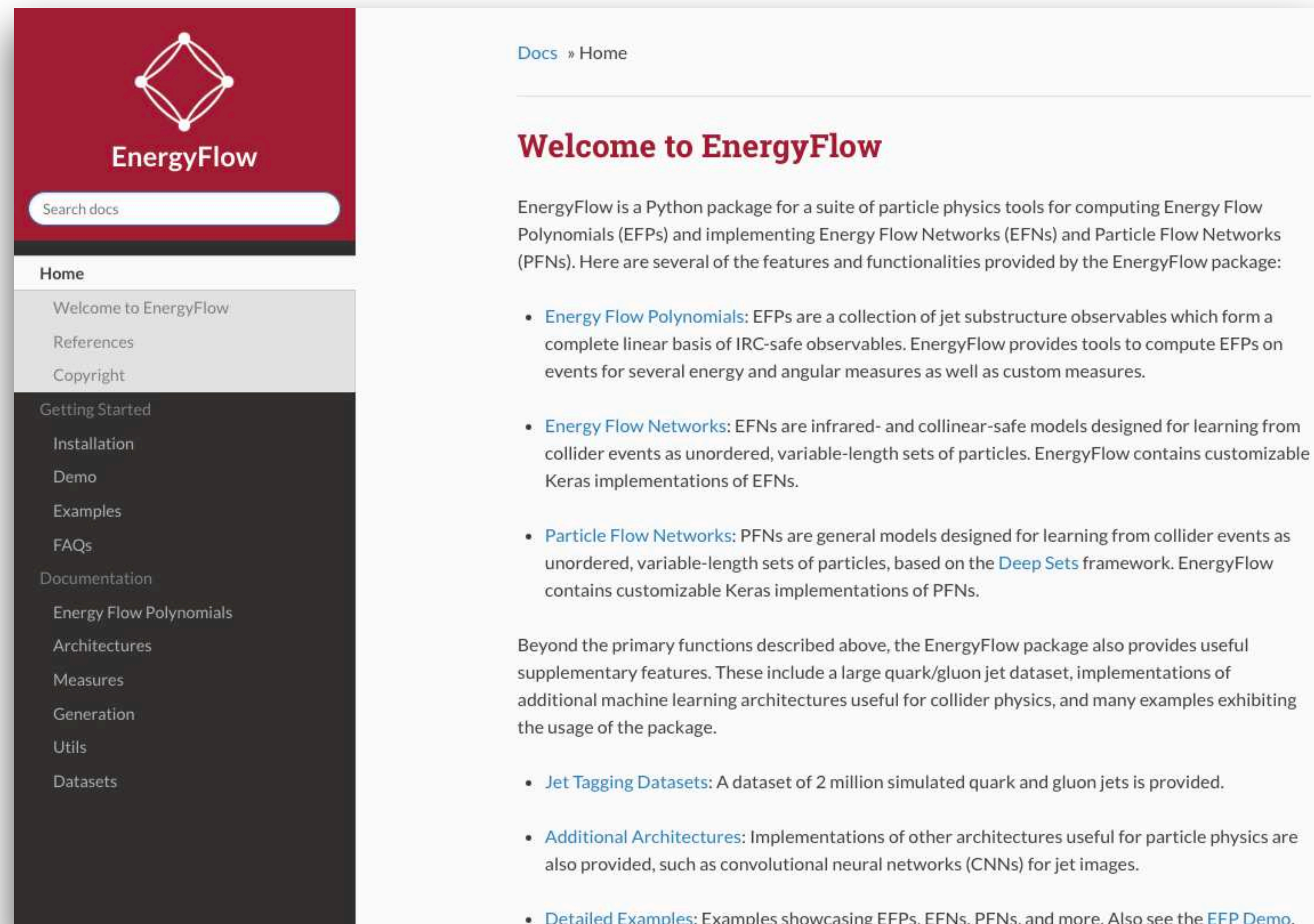
EnergyFlow Python Package

Contains EFN and PFN implementations in Keras

CNN, DNN architectures included for easy model comparison

Includes quark/gluon jet samples used in [1810.01565]

Several detailed examples demonstrating how to train models and make visualizations



EnergyFlow

Search docs

Home
- Welcome to EnergyFlow
- References
- Copyright

Getting Started
- Installation
- Demo
- Examples
- FAQs

Documentation
- Energy Flow Polynomials
- Architectures
- Measures
- Generation
- Utils
- Datasets

Docs » Home

Welcome to EnergyFlow

EnergyFlow is a Python package for a suite of particle physics tools for computing Energy Flow Polynomials (EFPs) and implementing Energy Flow Networks (EFNs) and Particle Flow Networks (PFNs). Here are several of the features and functionalities provided by the EnergyFlow package:

- Energy Flow Polynomials: EFPs are a collection of jet substructure observables which form a complete linear basis of IRC-safe observables. EnergyFlow provides tools to compute EFPs on events for several energy and angular measures as well as custom measures.
- Energy Flow Networks: EFNs are infrared- and collinear-safe models designed for learning from collider events as unordered, variable-length sets of particles. EnergyFlow contains customizable Keras implementations of EFNs.
- Particle Flow Networks: PFNs are general models designed for learning from collider events as unordered, variable-length sets of particles, based on the Deep Sets framework. EnergyFlow contains customizable Keras implementations of PFNs.

Beyond the primary functions described above, the EnergyFlow package also provides useful supplementary features. These include a large quark/gluon jet dataset, implementations of additional machine learning architectures useful for collider physics, and many examples exhibiting the usage of the package.

- Jet Tagging Datasets: A dataset of 2 million simulated quark and gluon jets is provided.
- Additional Architectures: Implementations of other architectures useful for particle physics are also provided, such as convolutional neural networks (CNNs) for jet images.
- Detailed Examples: Examples showcasing EFPs, EFNs, PFNs, and more. Also see the EFP Demo.

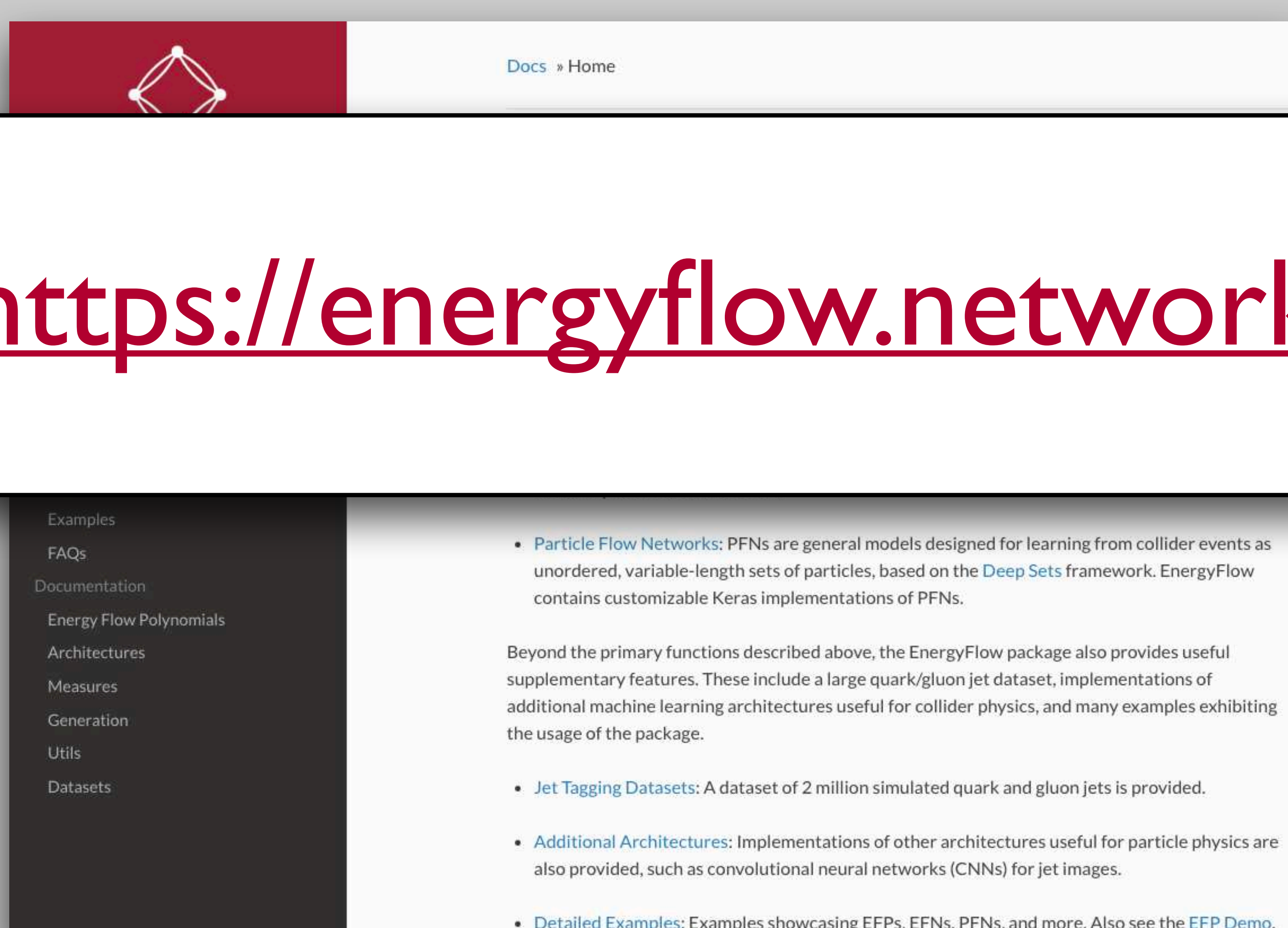
EnergyFlow Python Package

Contains EFN and PFN implementations in Keras

CNN, DNN architectures included for easy model comparison

Includes quark/gluon jet samples used in [1810.01565]

Several detailed examples demonstrating how to train models and make visualizations



Docs » Home

Examples
FAQs
Documentation
Energy Flow Polynomials
Architectures
Measures
Generation
Utils
Datasets

- Particle Flow Networks: PFNs are general models designed for learning from collider events as unordered, variable-length sets of particles, based on the Deep Sets framework. EnergyFlow contains customizable Keras implementations of PFNs.

Beyond the primary functions described above, the EnergyFlow package also provides useful supplementary features. These include a large quark/gluon jet dataset, implementations of additional machine learning architectures useful for collider physics, and many examples exhibiting the usage of the package.

- Jet Tagging Datasets: A dataset of 2 million simulated quark and gluon jets is provided.
- Additional Architectures: Implementations of other architectures useful for particle physics are also provided, such as convolutional neural networks (CNNs) for jet images.
- Detailed Examples: Examples showcasing EFPs, EFNs, PFNs, and more. Also see the EFP Demo.

https://energyflow.network

Thank You!