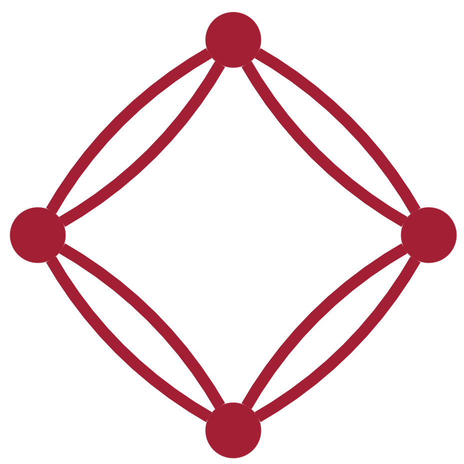
Point Cloud Strategies for Boosted Objects

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

CERN BSM Forum
February 21, 2019

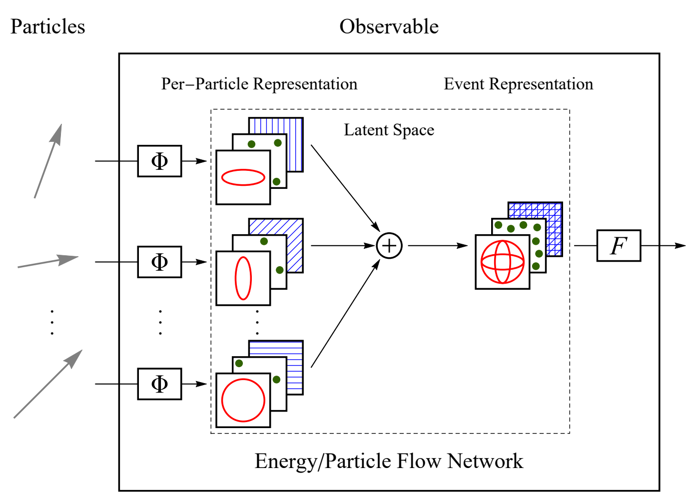
Collaborators: Eric Metodiev and Jesse Thaler

1712.07124 1810.05165



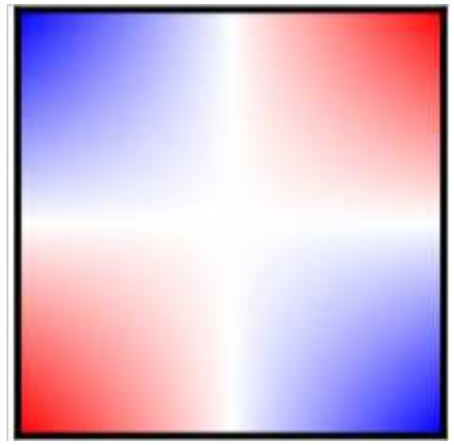
Energy Flow Polynomials

"Understanding the space of IRC-safe observables"



Energy Flow Networks

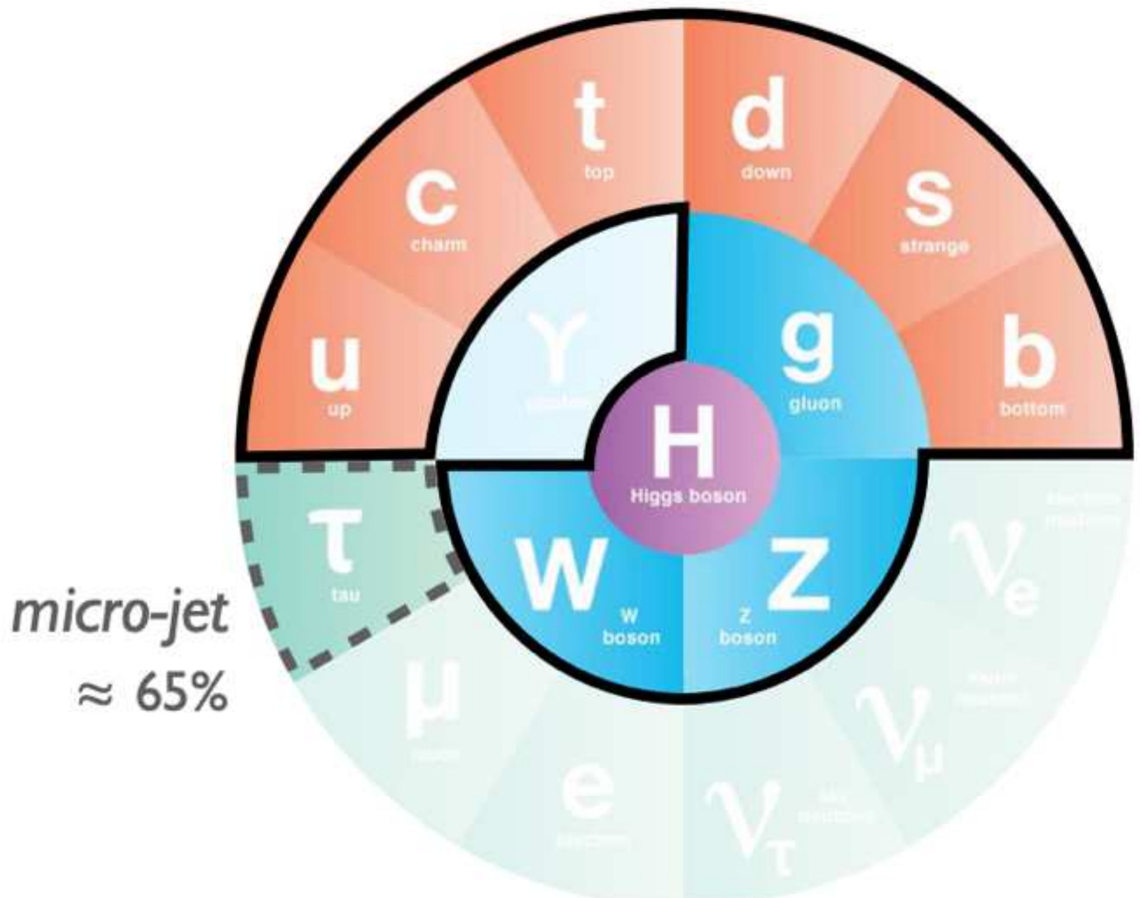
"Power of ML meets IRC-safe physics"



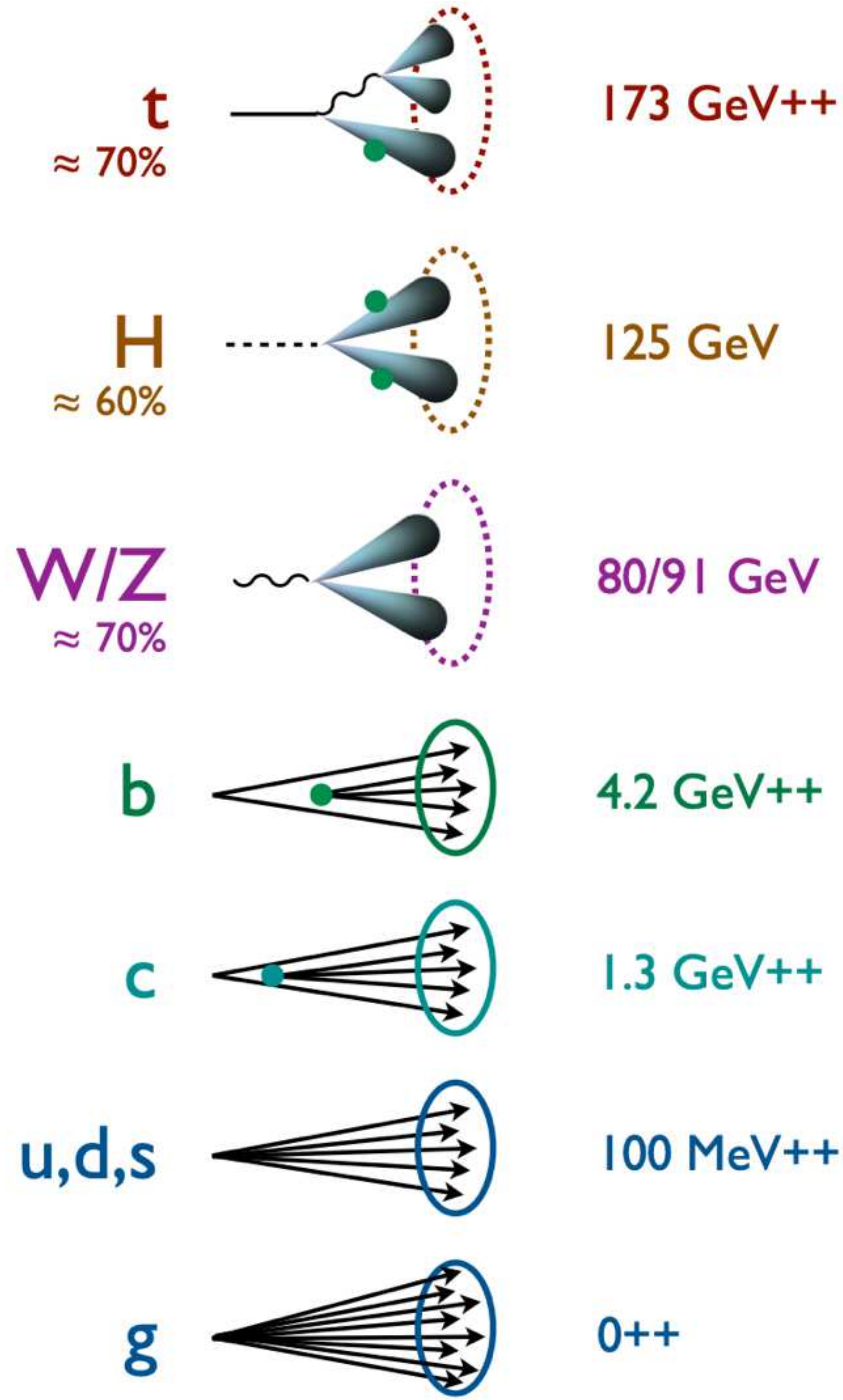
Energy Flow Moments

"A deep connection between EFPs and EFNs"

Jets in Theory

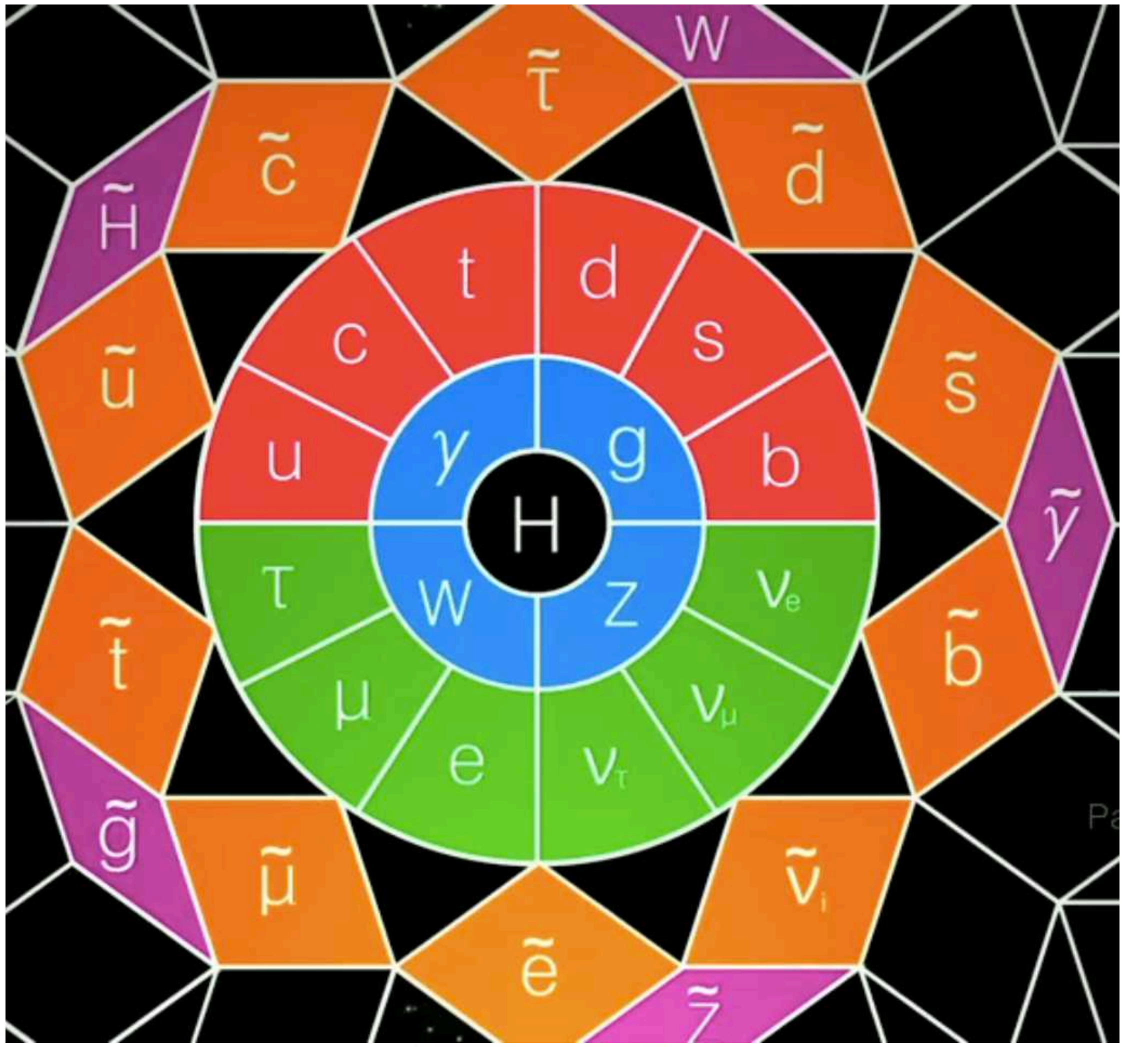


Jets from the Standard Model



++ = Mass from QCD Radiation

Jets in Theory



Jets from Beyond the Standard Model

++ = Mass from QCD Radiation

$\tilde{t} \approx ?$

$\tilde{H} \approx ?$

$\tilde{W}/\tilde{Z} \approx ?$

$\tilde{b} \approx ?$

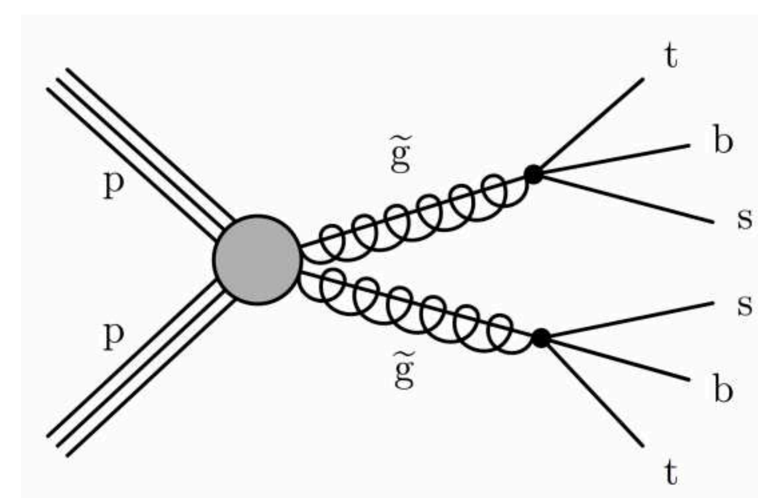
$\tilde{c} \approx ?$

$\tilde{u},\tilde{d},\tilde{s} \approx ?$

$\tilde{g} \approx ?$

Many models of new physics generate boosted Standard Model hadronic final states

e.g. Z' ⟶ $t\bar{t}$, cascade decays, various SUSY scenarios



[CMS-SUS-16-040]

Cartoon of Jet Formation

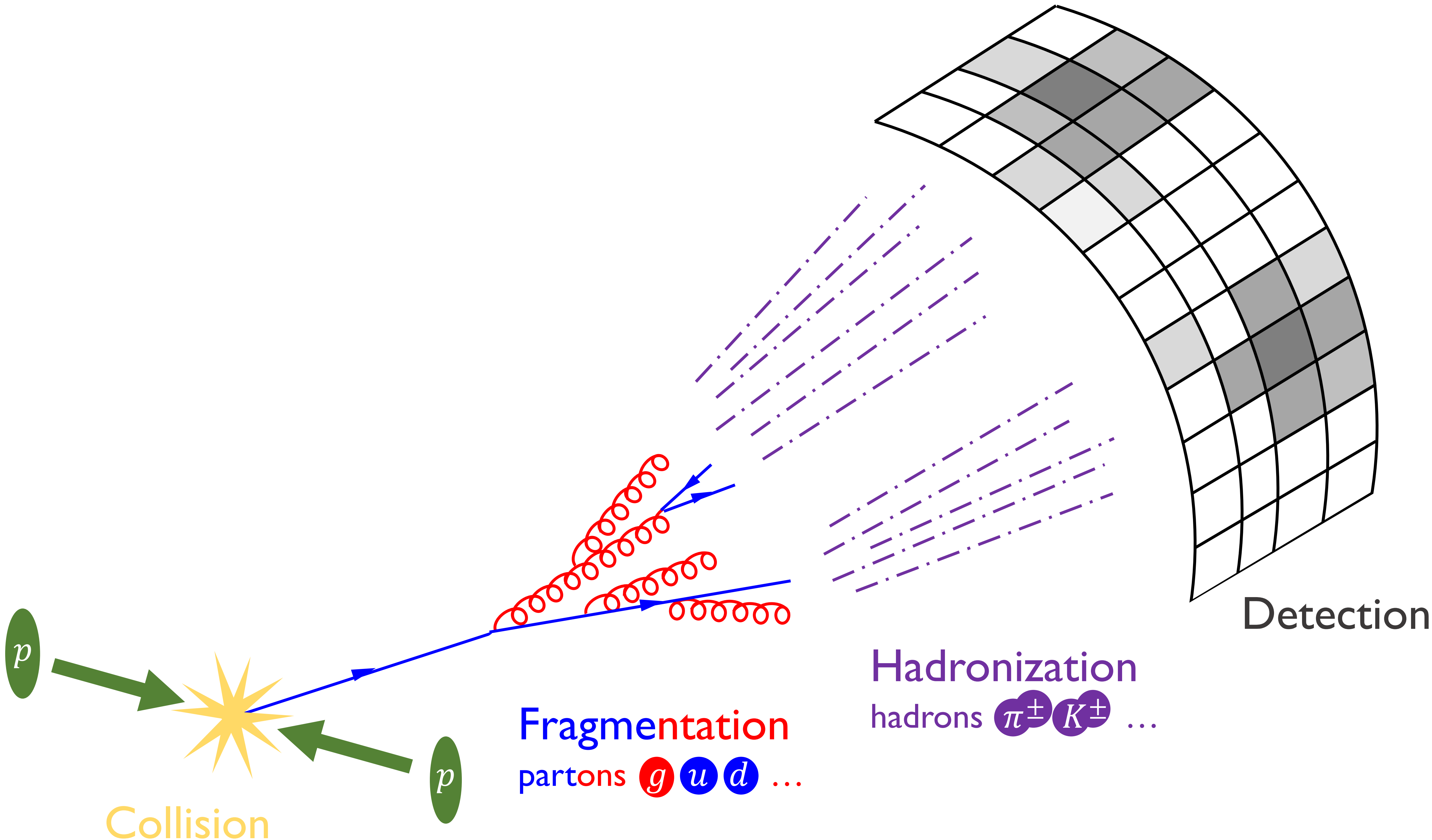
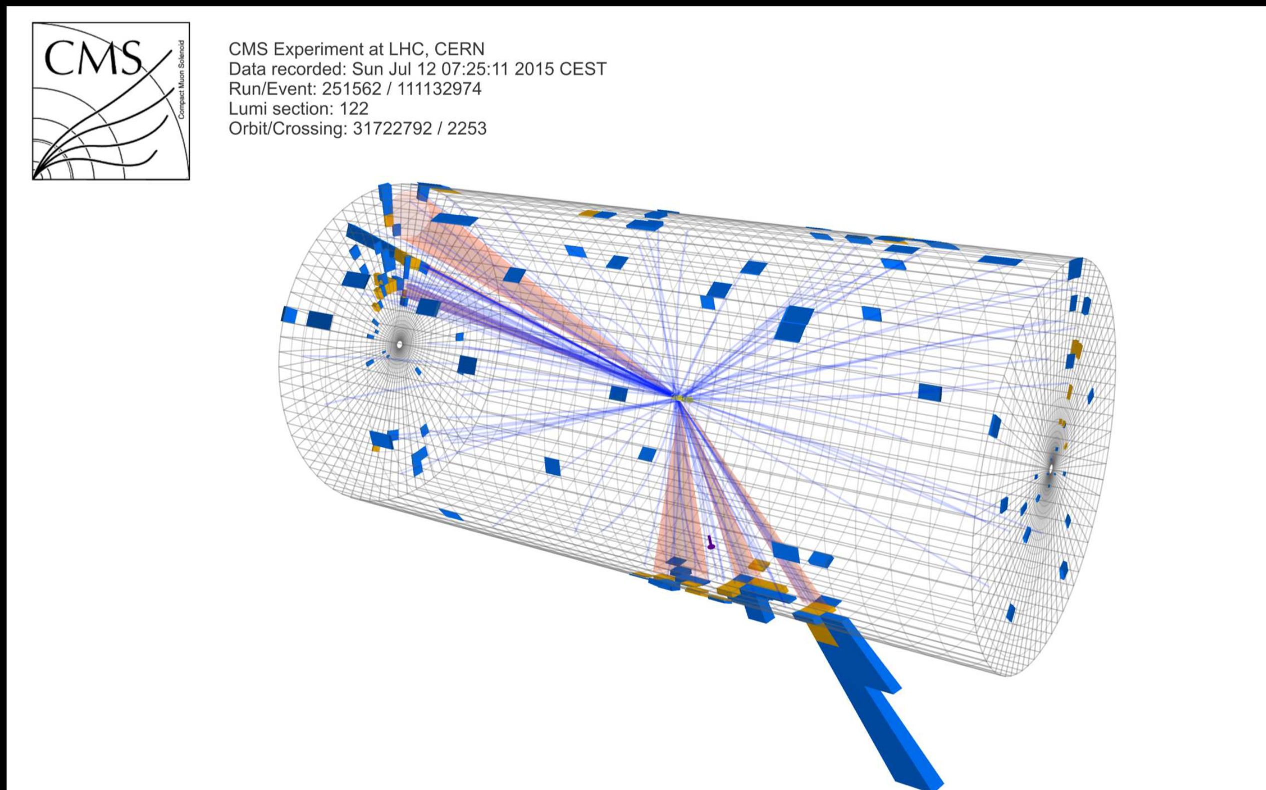
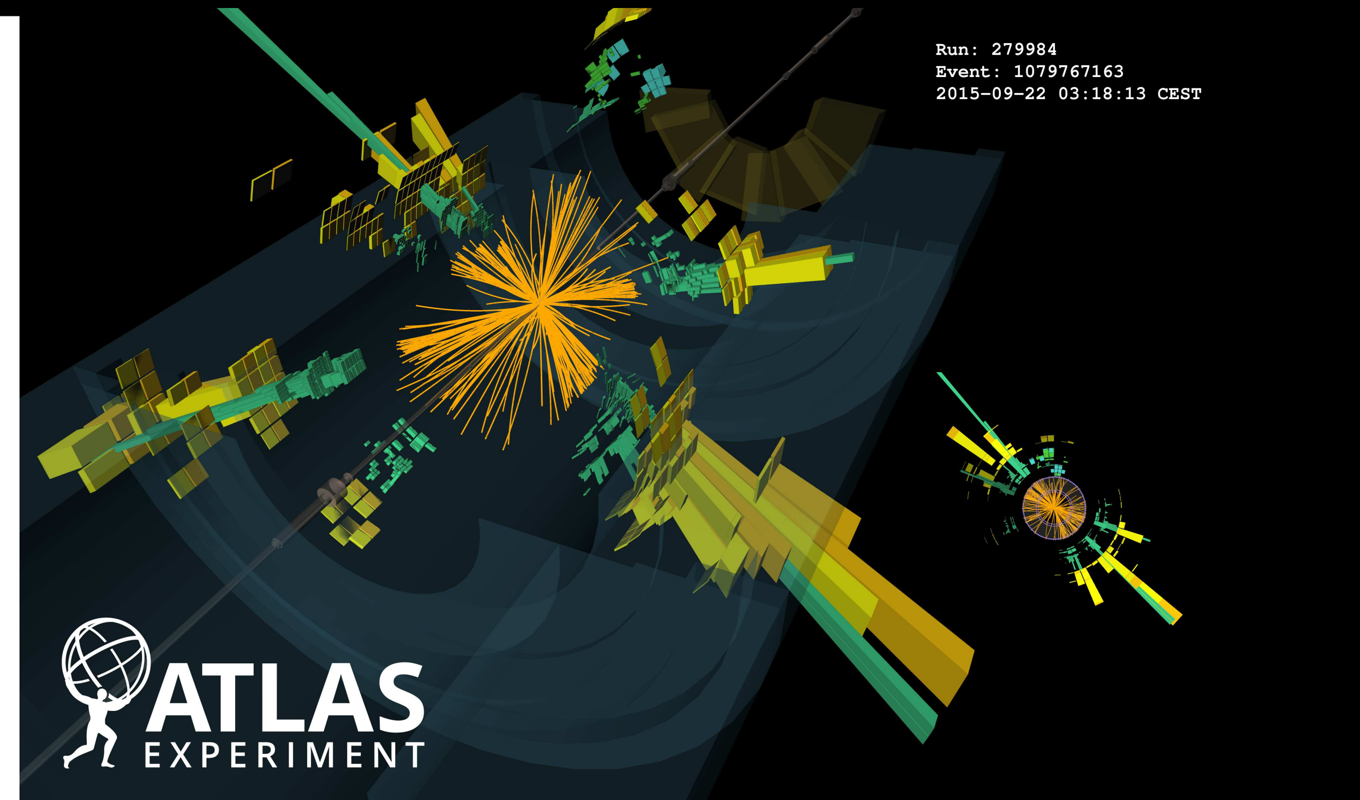
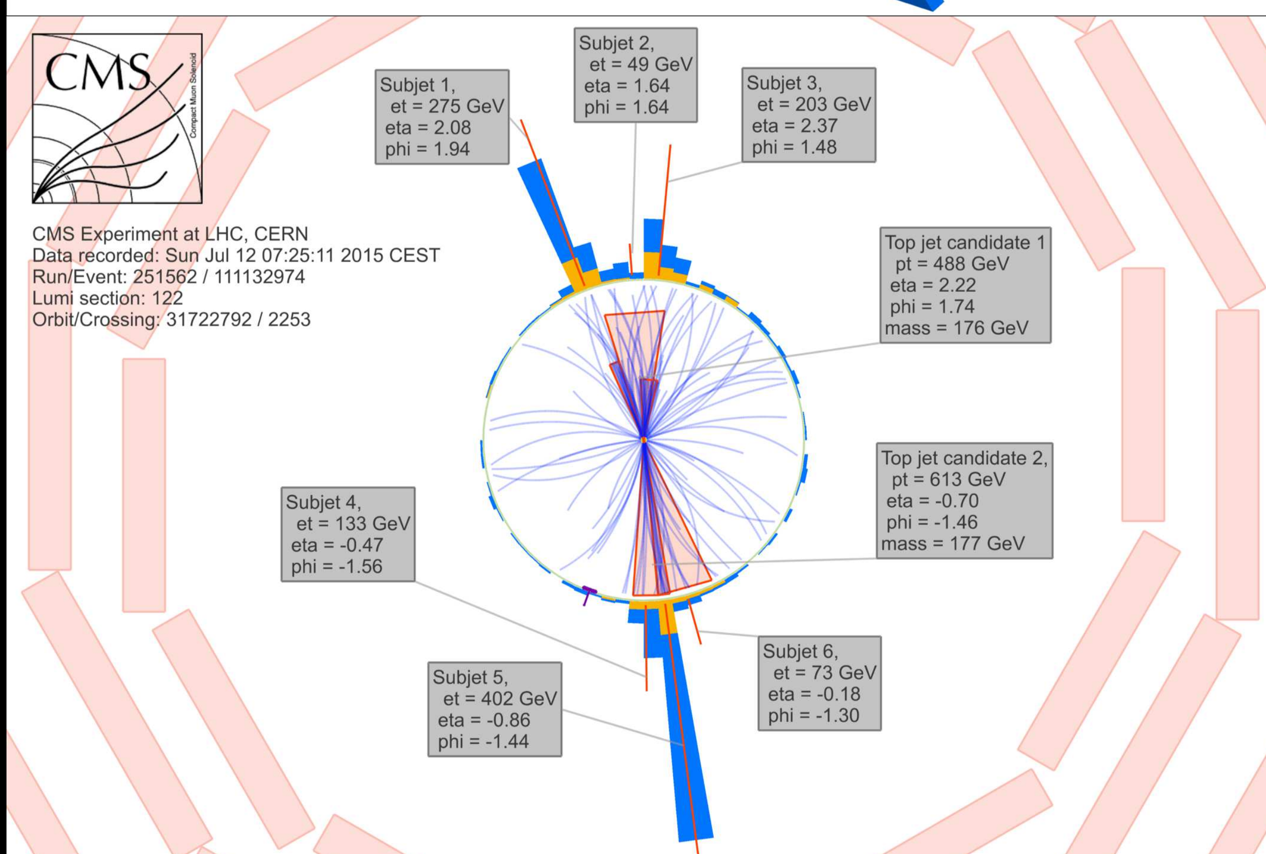


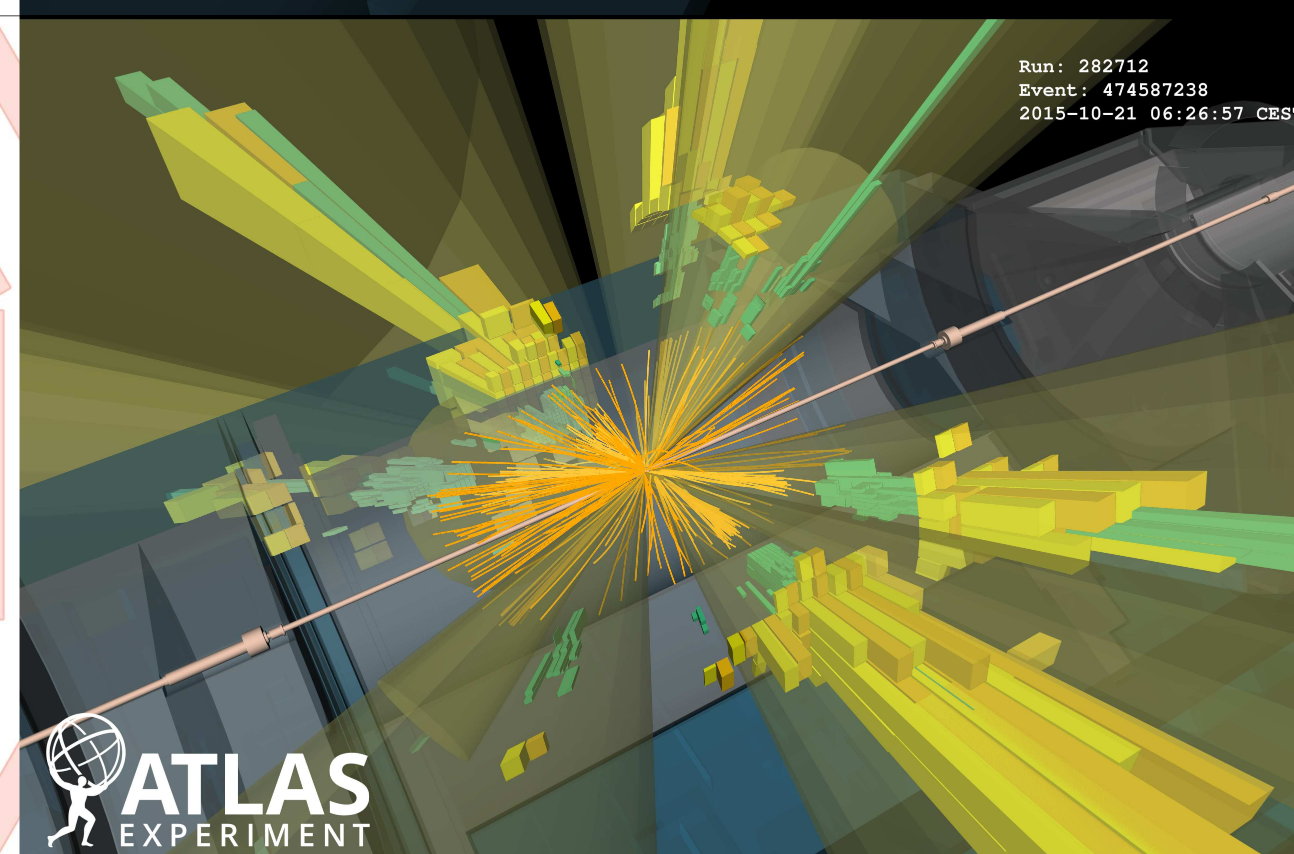
Diagram by Eric Metodiev

Boosted Event Topologies at the LHC









What is a Jet?

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

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

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

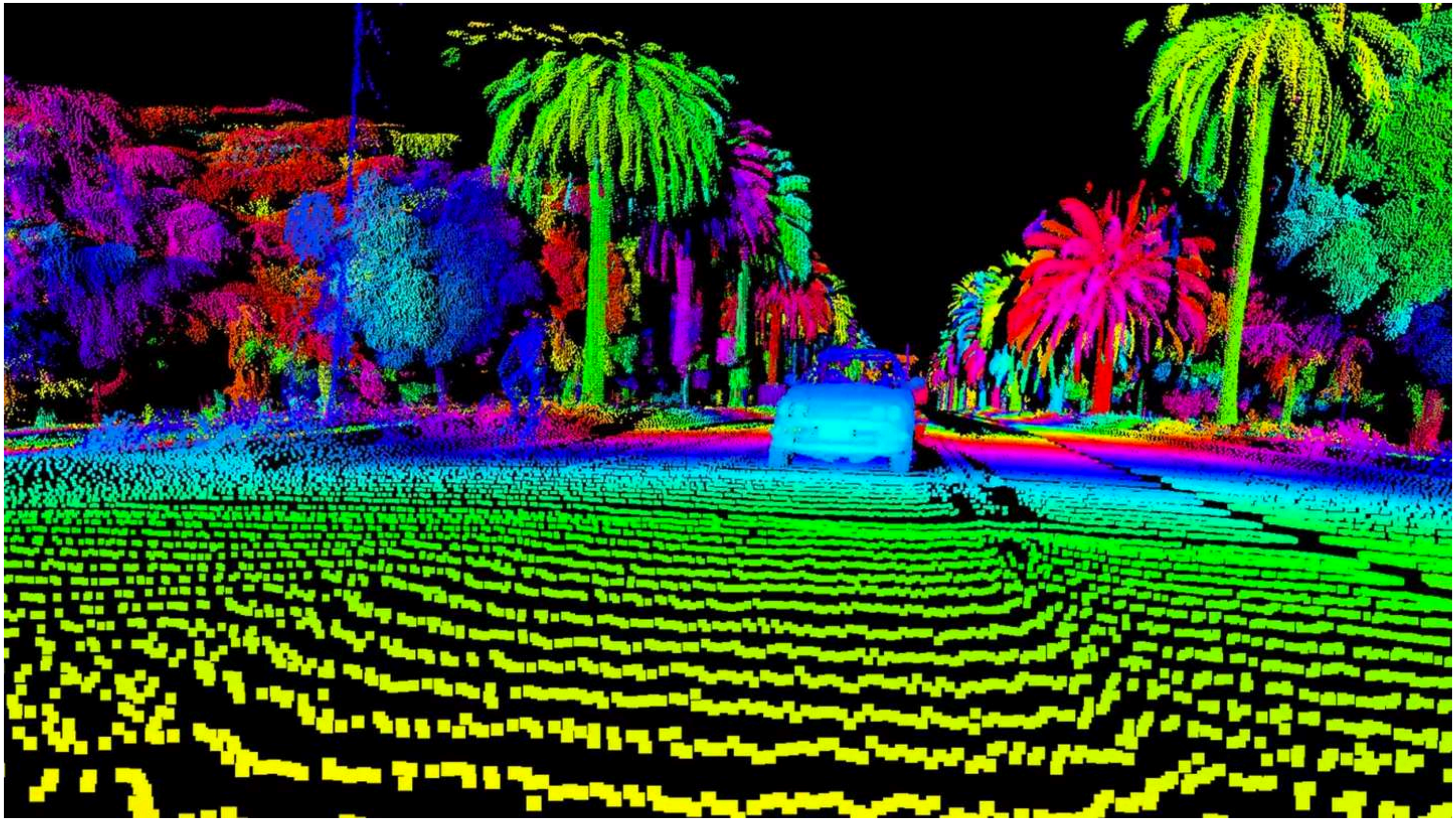
p_i^μ represents *all* the particle properties:

- 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

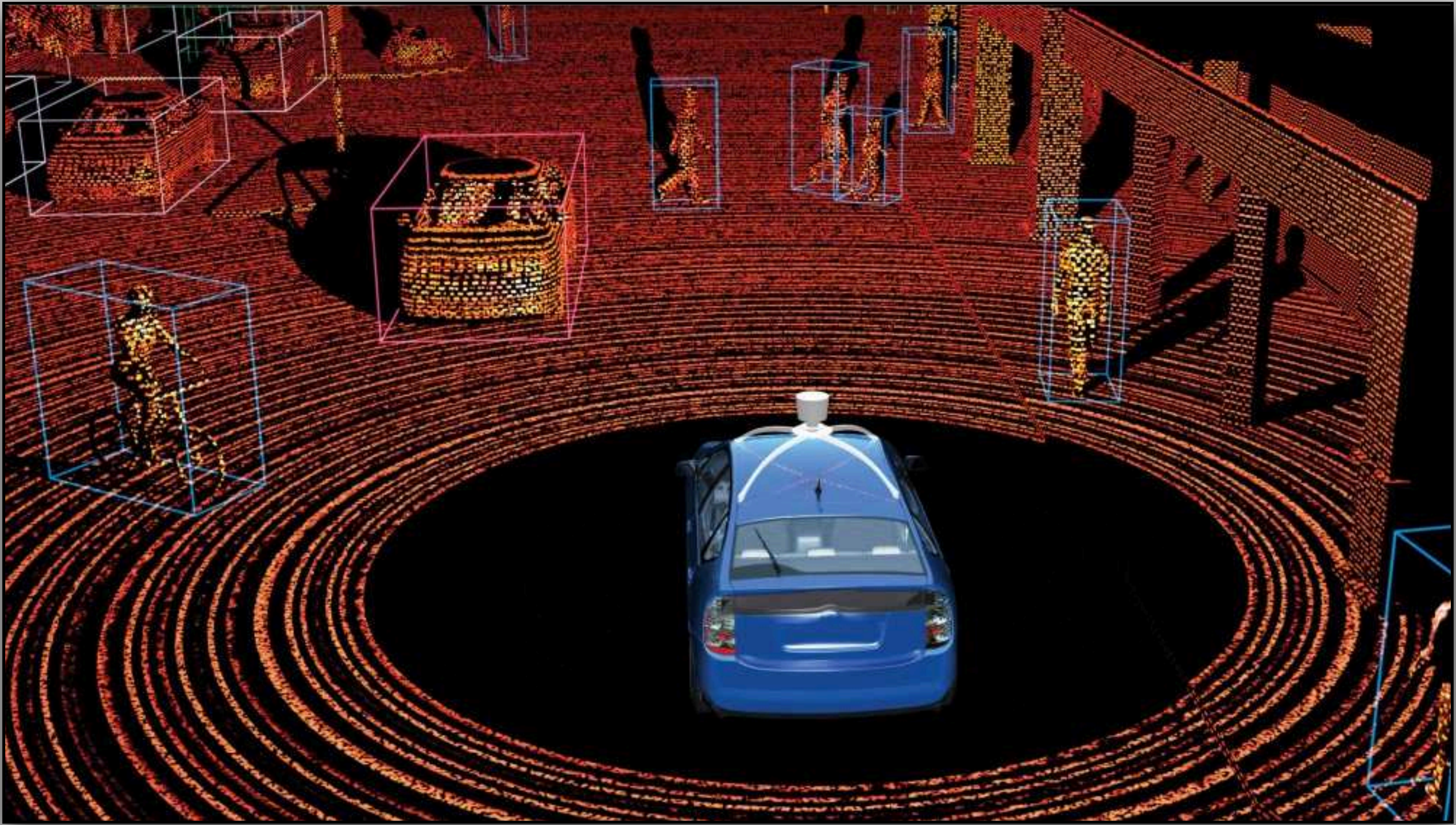
Point cloud: "A set of data points in space" –Wikipedia

LIDAR data from self-driving car sensor



Point Clouds

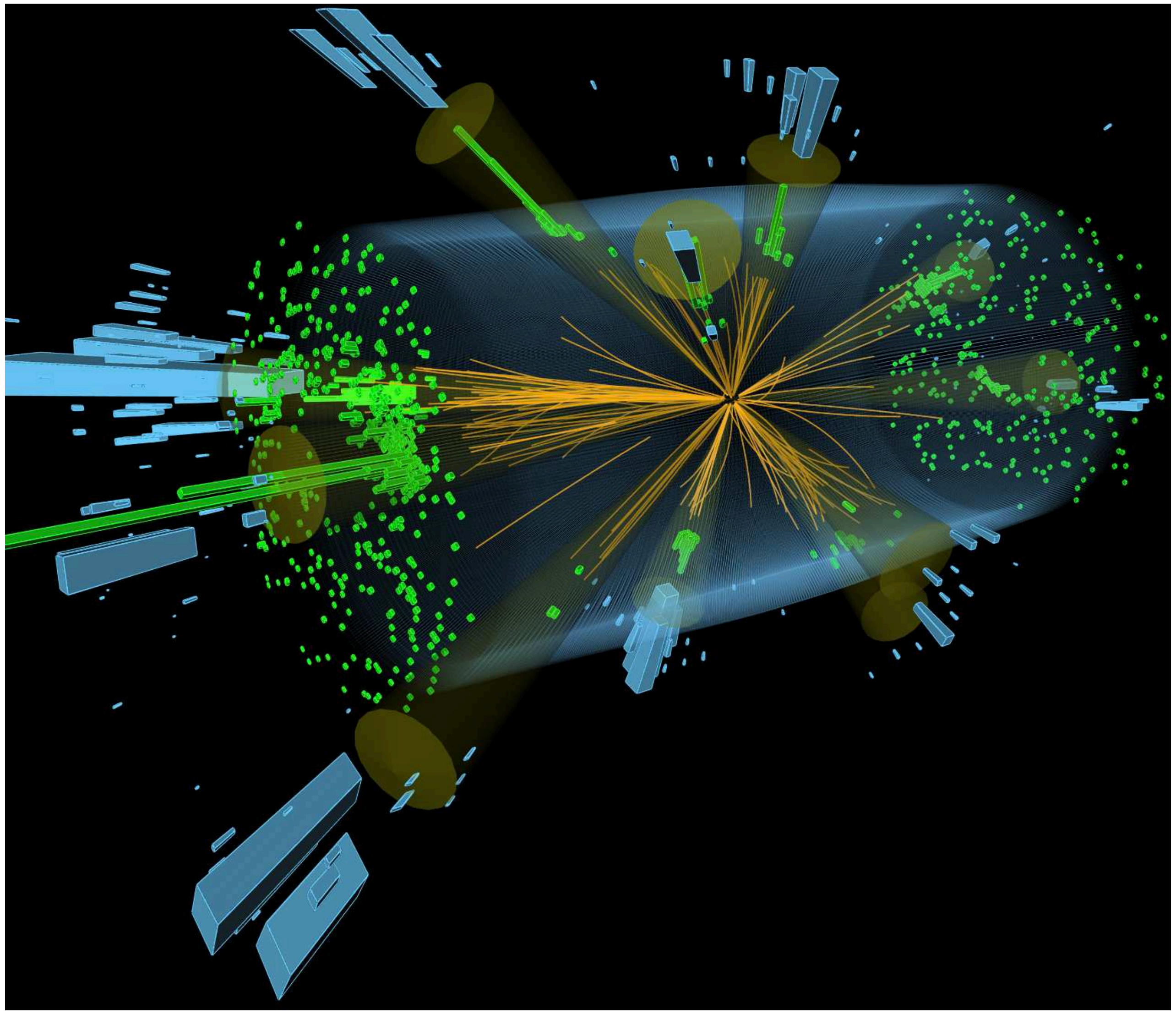
[Popular Science, 2013]



Particle Collision Events as Point Clouds

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

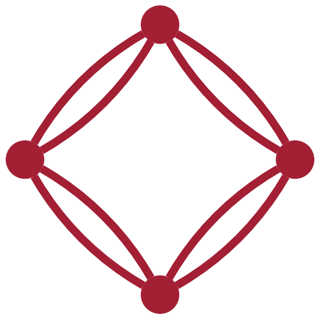
Jet/event — Particles — Feature space



Multi-jet event at CMS

Processing Point Clouds

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

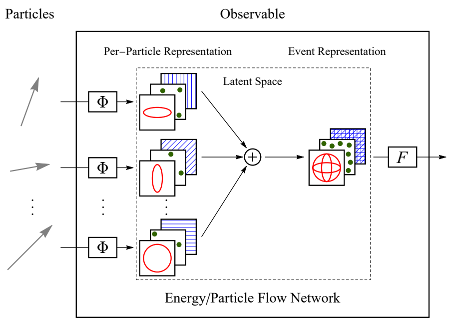


Variable constituent multiplicity requires at least one of:

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

Truncation to an (arbitrary) fixed size

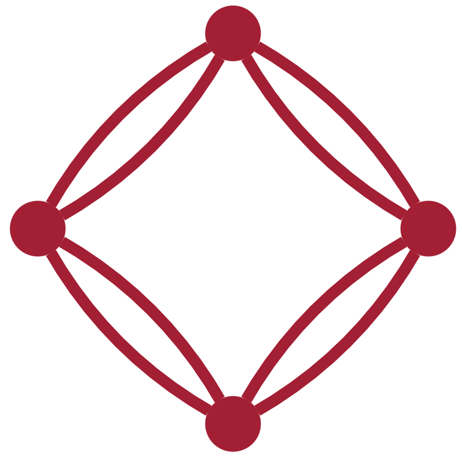
Recurrent NN structure



Particle permutation symmetry requires:

Permutation symmetric observables

Permutation symmetric architectures



Energy Flow Polynomials

Fixed preprocessing of a point cloud

Energy Flow Networks

Energy Flow Moments

Infrared and Collinear (IRC) Safety

QCD has soft and collinear divergences associated with gluon radiation

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

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

KLN Theorem: 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

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

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

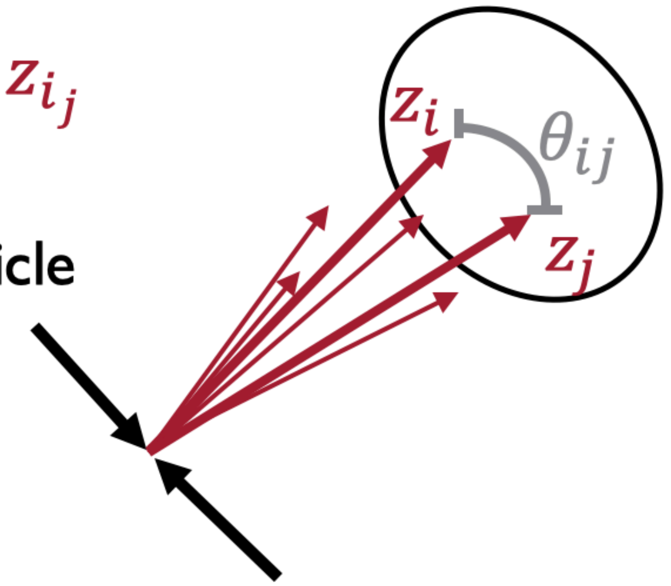
$$S(\{p_1^\mu,\ldots,p_M^\mu\}) = S(\{p_1^\mu,\ldots,(1-\lambda)p_M^\mu,\lambda p_{M+1}^\mu\}), \quad \forall\, \lambda \in [0,1]$$

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



[F. Tkachov, hep-ph/9601308]

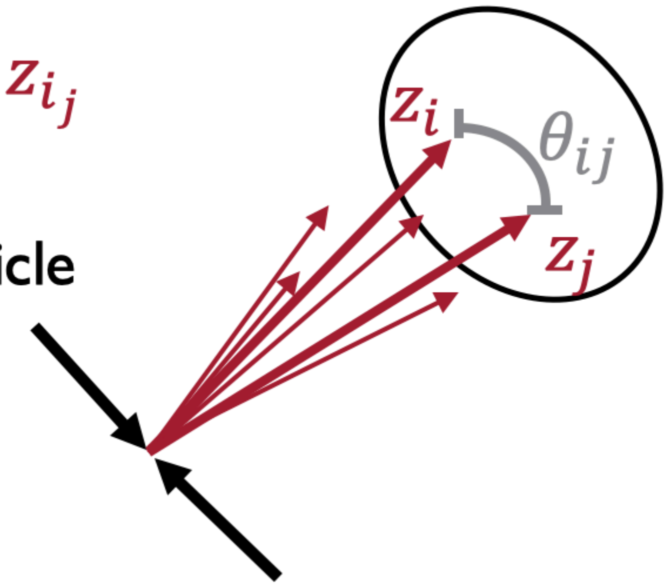
Energy correlator parametrized by angular function f

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

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]

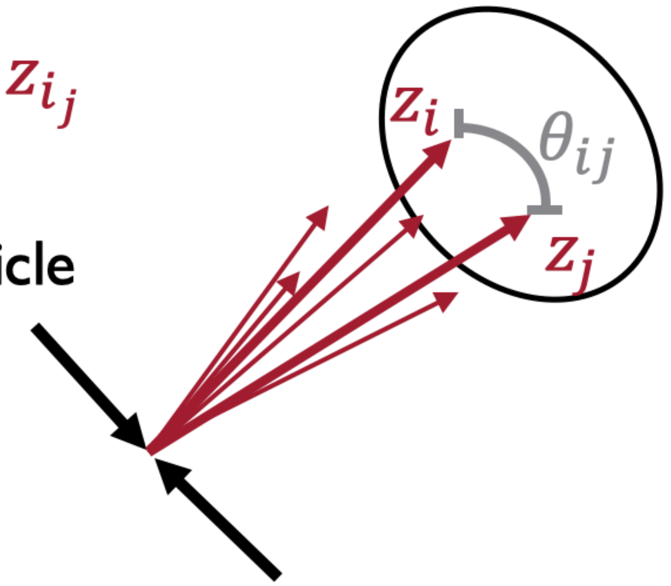
➡ Energy correlators linearly span IRC-safe observables

- 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: $S(p_1^\mu, \ldots, p_M^\mu)$

- 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]

➡ 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!

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

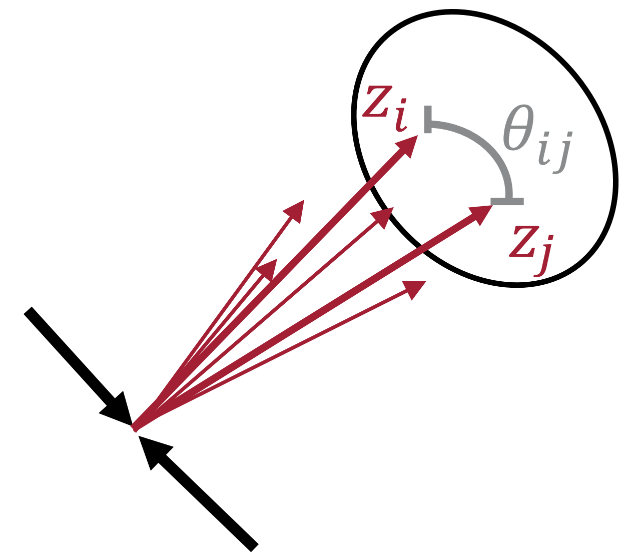
**Generically, approximations exist by the Stone-Weierstrass theorem

Energy Flow Polynomials (EFPs)

[PTK, Metodiev, Thaler, 1712.07124]

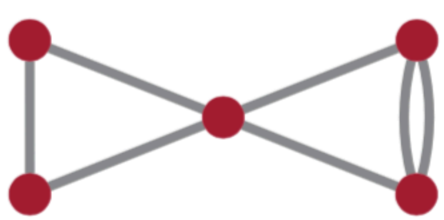
$$\text{EFP}_G = \underbrace{\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \underbrace{z_{i_1} \cdots z_{i_N}}_{\text{Energies}} \underbrace{\prod_{(k,\ell)\in G} \theta_{i_k i_\ell}}_{\text{Angles}}$$

Correlator of Energies and Angles



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

A family of energy correlators with angular structures determined by multigraphs



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

<u>Multigraph correspondence</u>

$\longleftrightarrow$ z_{i_j}

$\longleftrightarrow$ $\theta_{i_k i_l}$

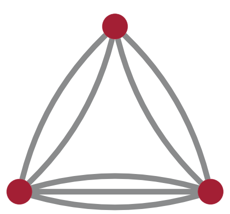
Energy and Angle Measure

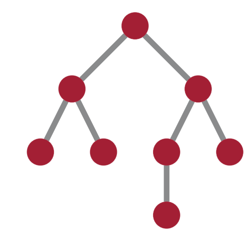
Hadronic : $z_i = \dfrac{p_{Ti}}{\sum_j p_{Tj}}, \quad \theta_{ij} = \left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)^{\beta/2}$

Linear Basis of IRC-Safe Observables

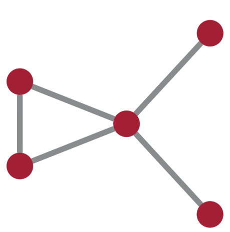
One can show via the Stone-Weierstrass approximation theorem that any IRC-safe observable is a linear combination of EFPs

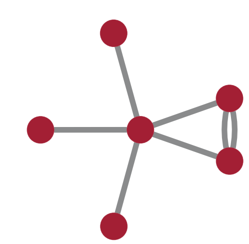
$$\mathcal{S} \simeq \sum_{G\in\mathcal{G}} s_G \, \mathrm{EFP}_G, \quad \mathcal{G} \text{ a set of multigraphs}$$





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

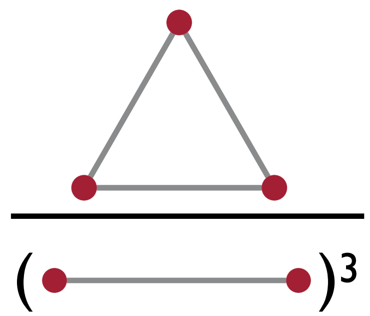




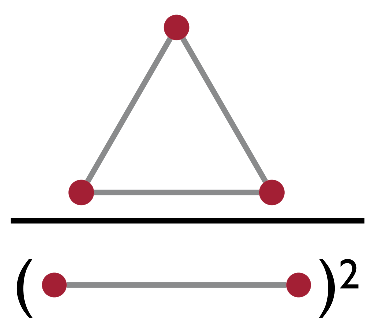
Strategy: Learn coefficients s_G via linear regression or classification

Familiar Observables as EFPs

m_J^2 = 

D_2 = 

[Larkoski, Moult, Neill, 2014]

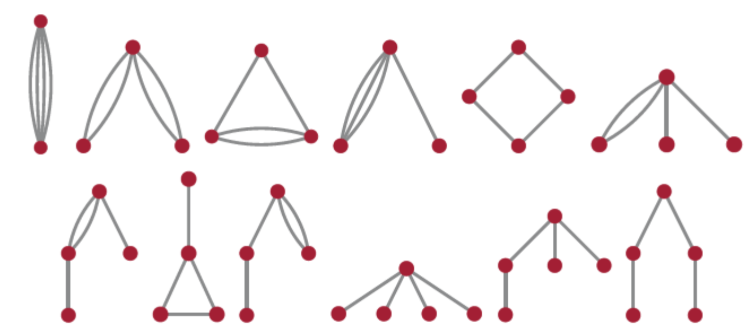
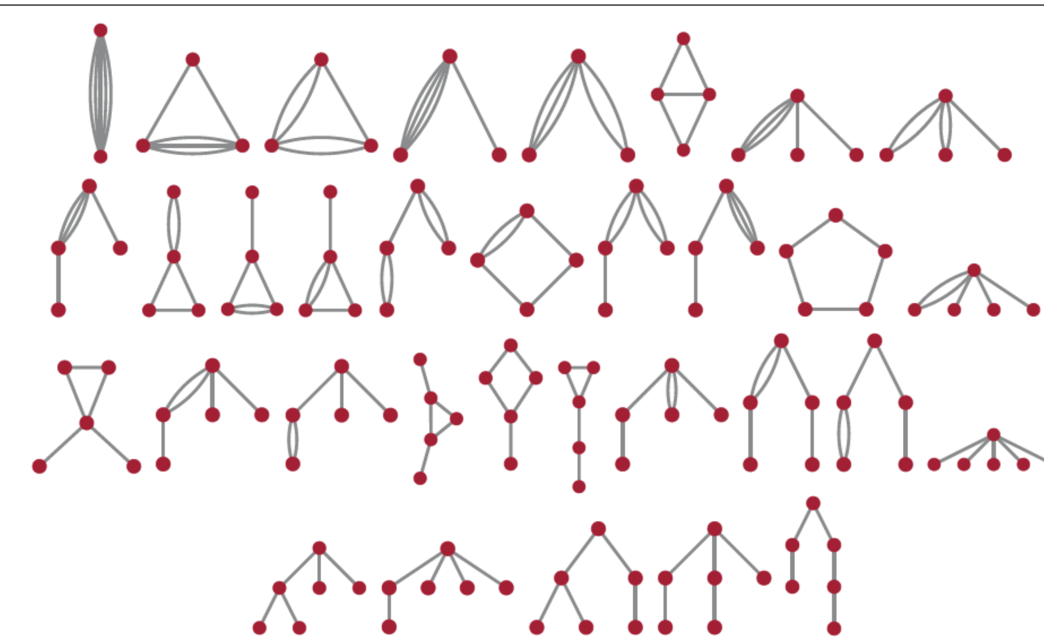
C_2 = 

[Larkoski, Salam, Thaler, 2013]

Energy correlation functions are complete graphs

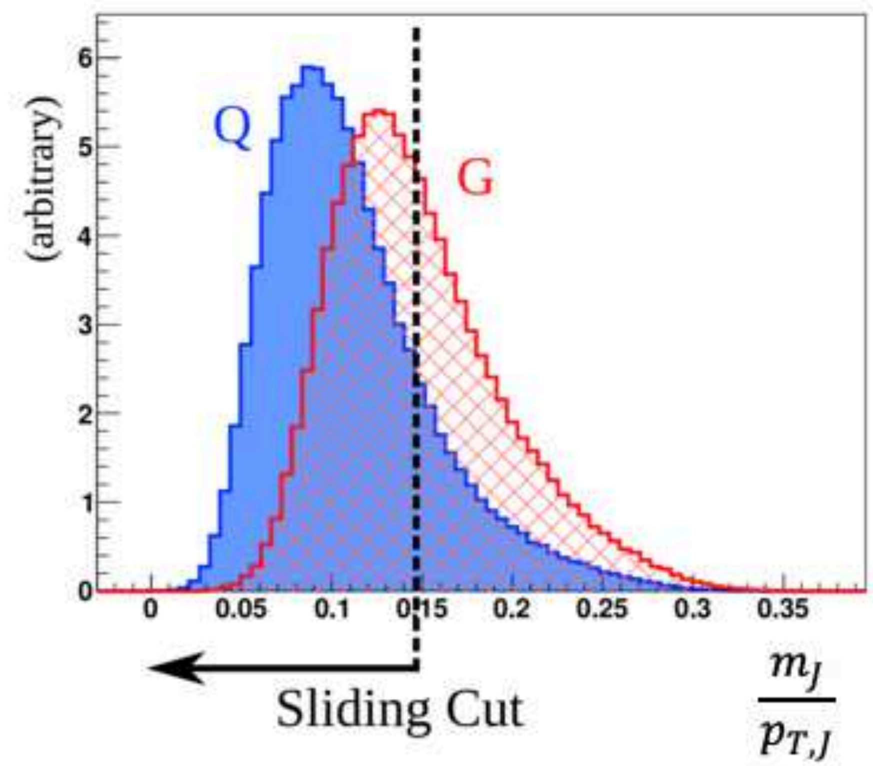
Even angularities are exact linear combinations of EFPs

EFPs organized by degree d – number of edges

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

Quantifying a Classifier

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



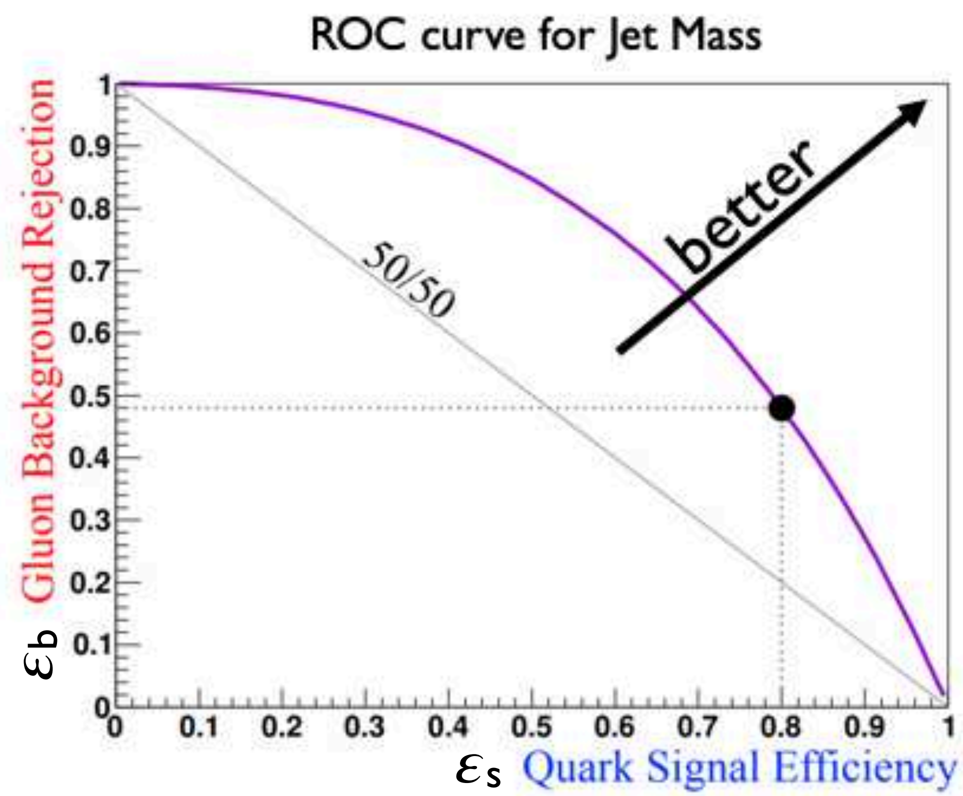
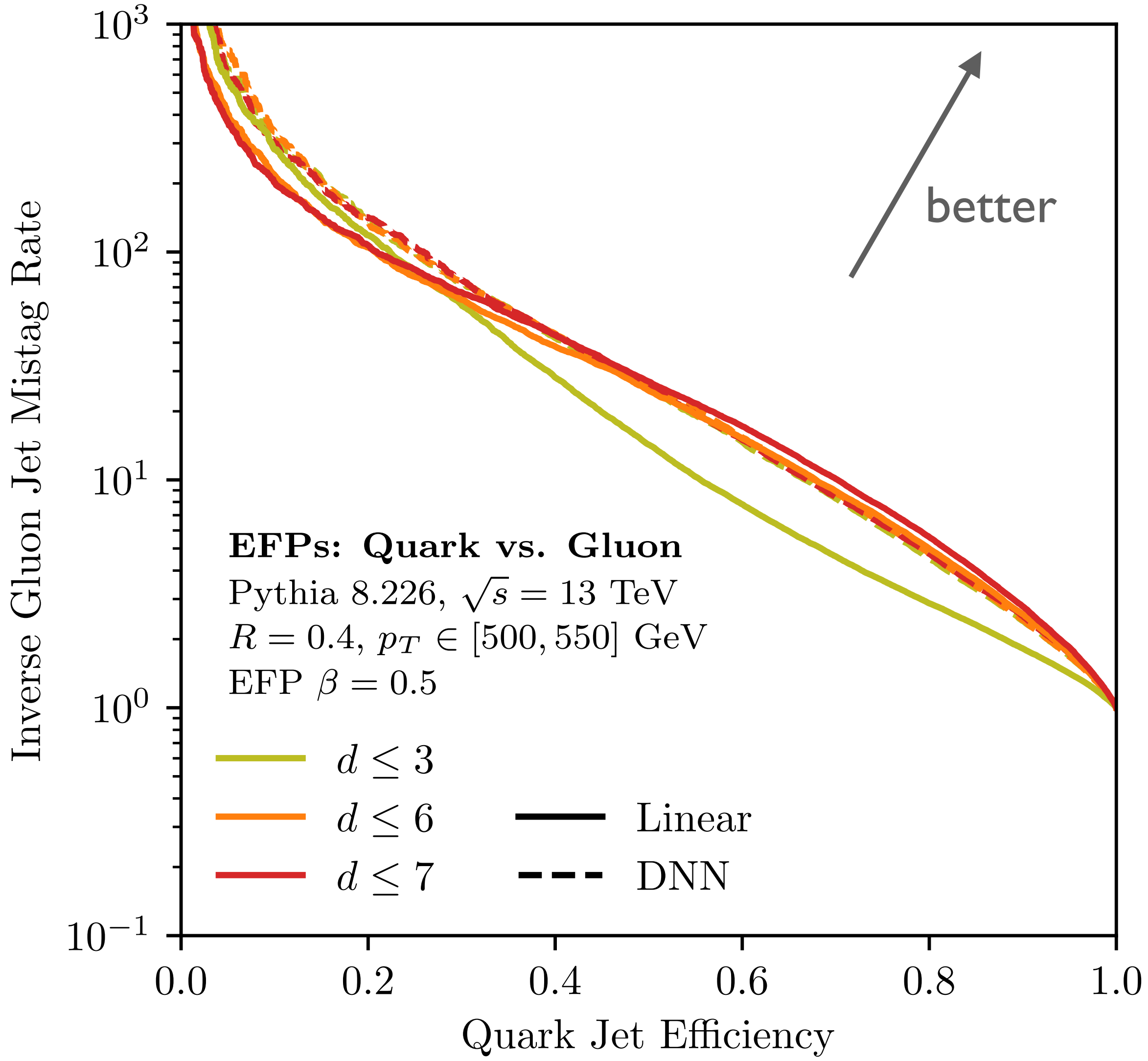


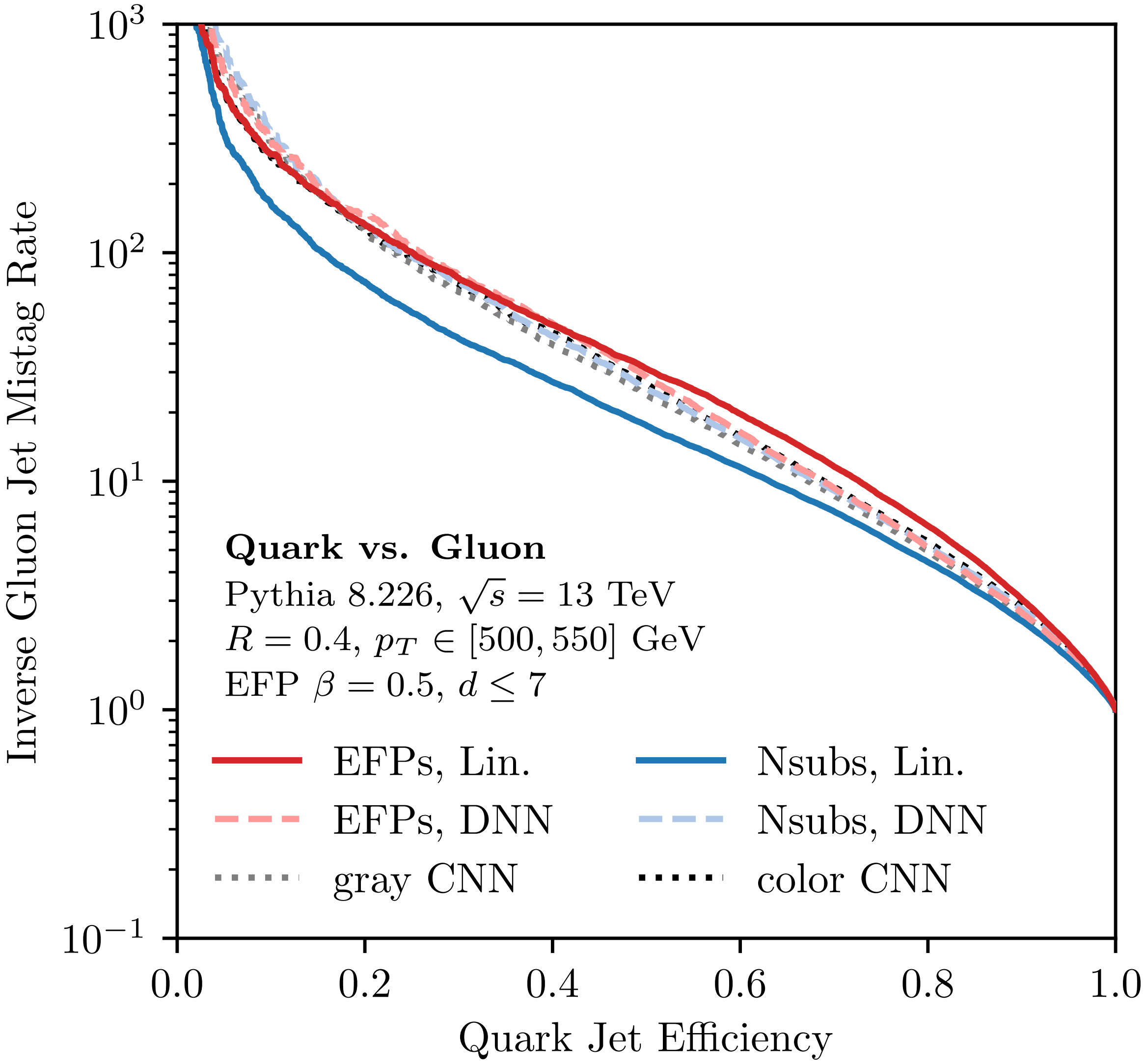
Figure from 1211.7038

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: EFP Classification Performance Comparison

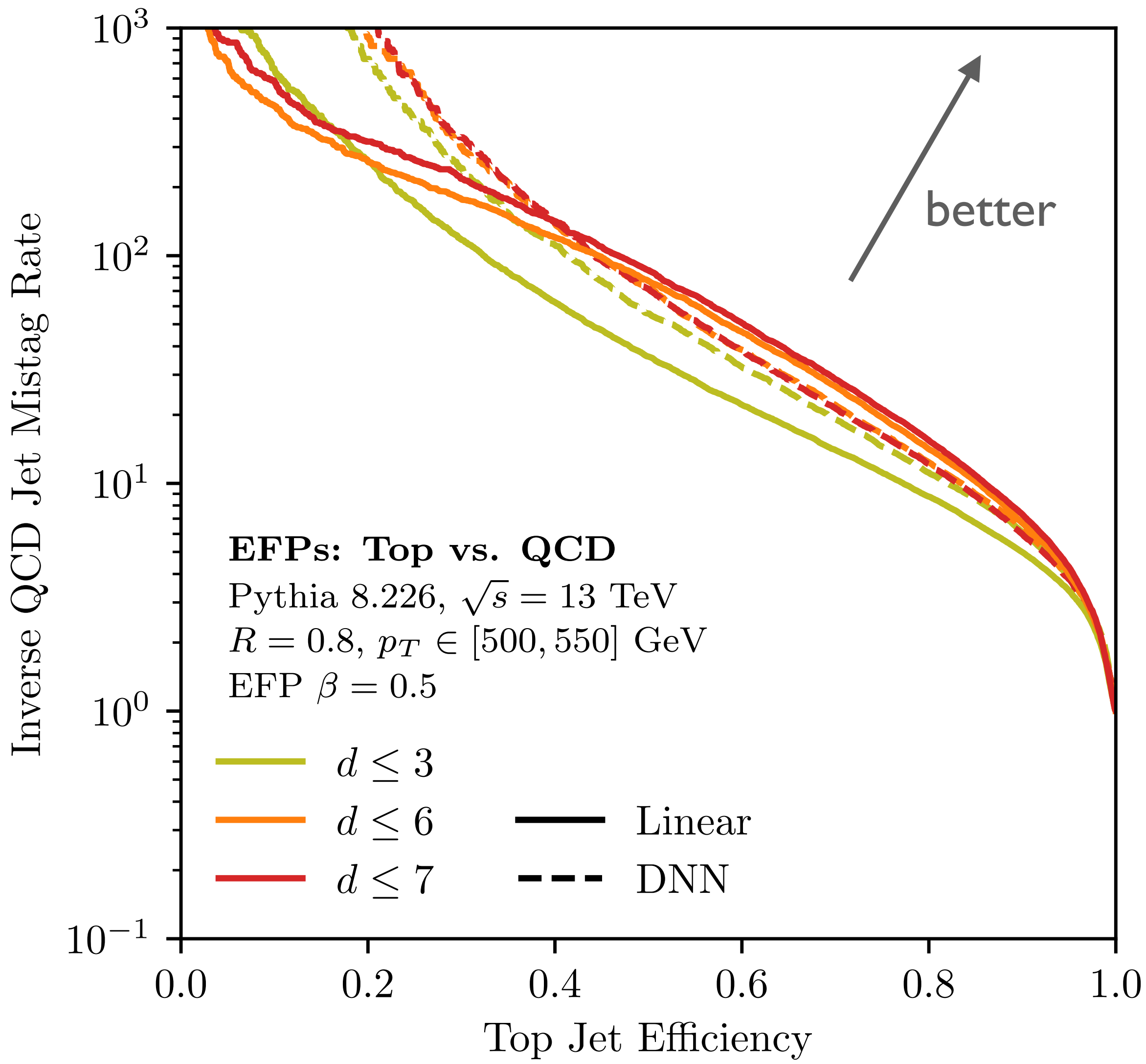


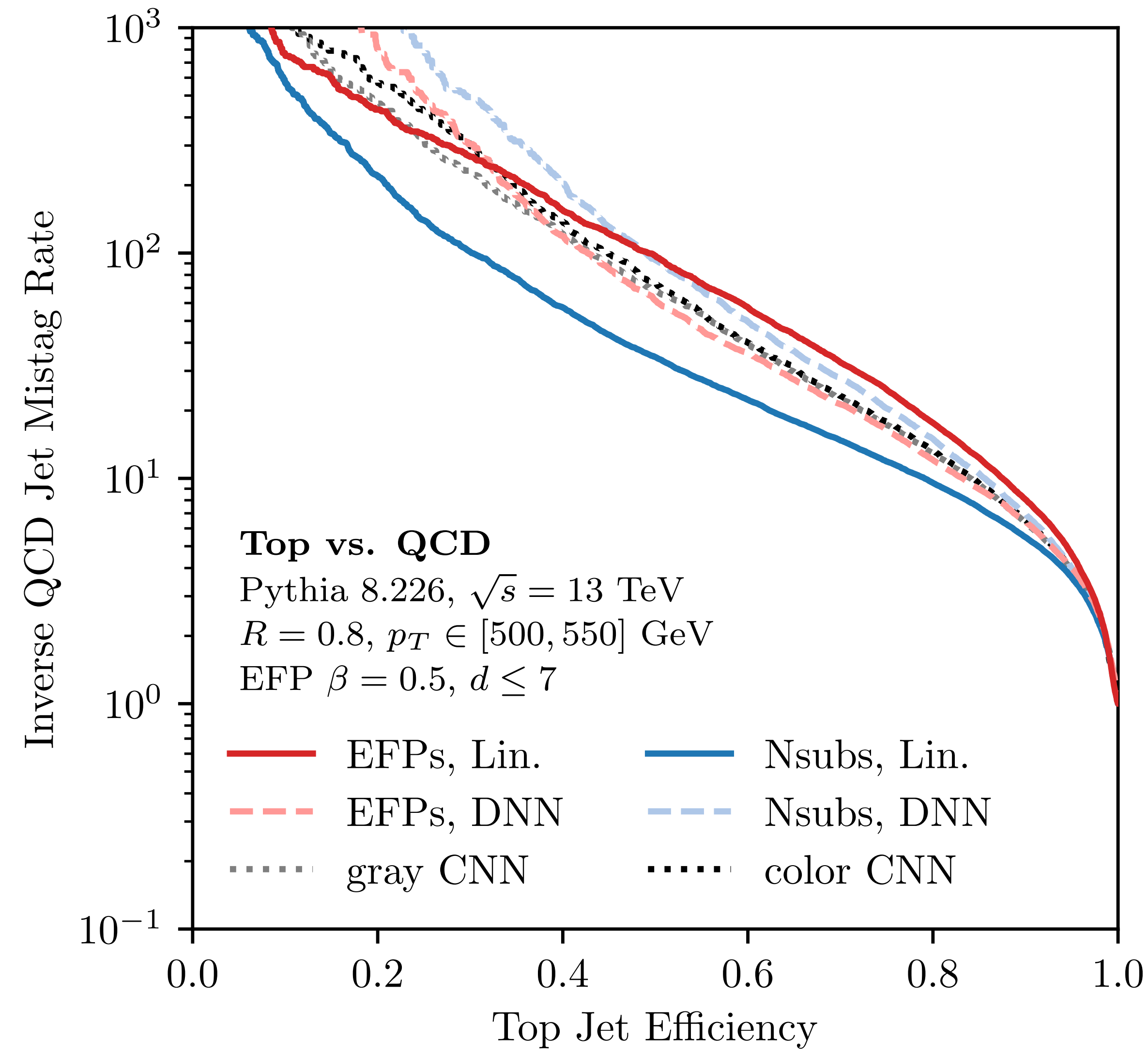


Saturation observed with more EFPs

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

Boosted Top: EFP Classification Performance Comparison



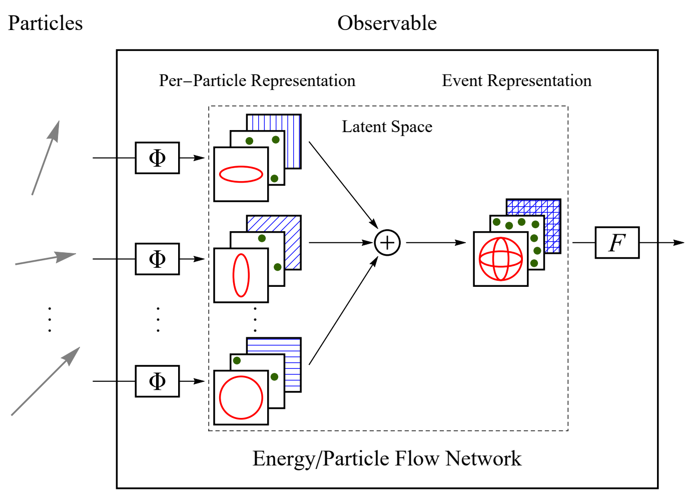


Saturation observed with more EFPs

Linear EFPs excel at high efficiency

DNN gets there faster but linear suffices

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



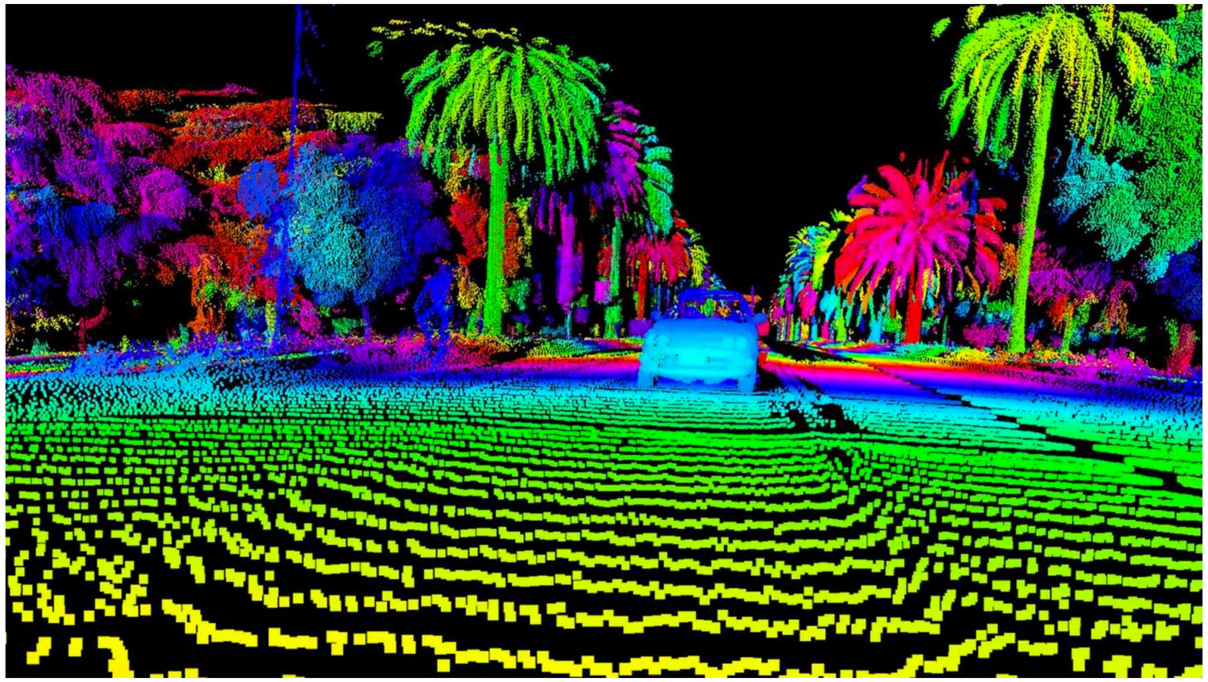
Energy Flow Polynomials

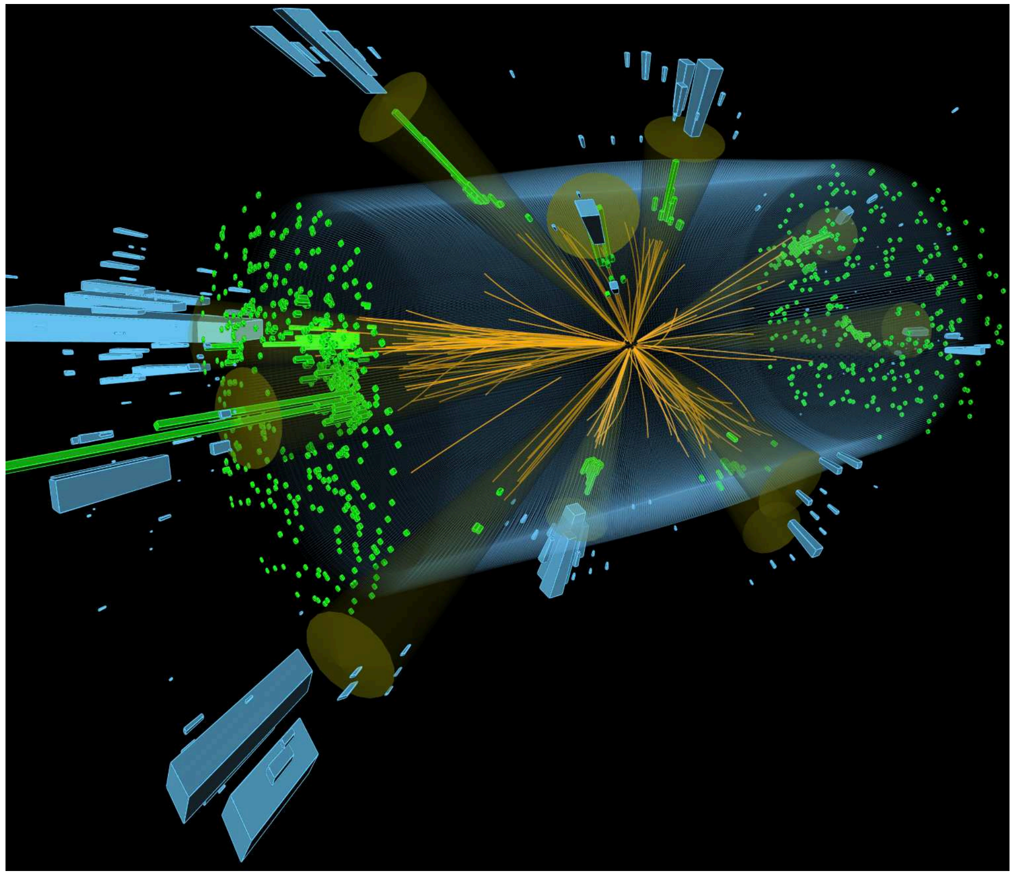
Energy Flow Networks

Learning to process point clouds into observables

Energy Flow Moments

Point Clouds





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

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

Infrared and Collinear (IRC) Safety

QCD h

KLN T ... ar
diverge

Infrare

Colline ... e

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

Theorem: Any IRC-safe observable can be written in the following form:

$$f(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M} z_i \vec{\Phi}(\hat{p}_i)\right), \quad \hat{p}_i = (y_i,\phi_i).$$

Proof: In 1810.05165.

□

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

Deep Sets for Particle Jets

[PTK, Metodiev, Thaler, 1810.05165]

Particle Flow Network (PFN)

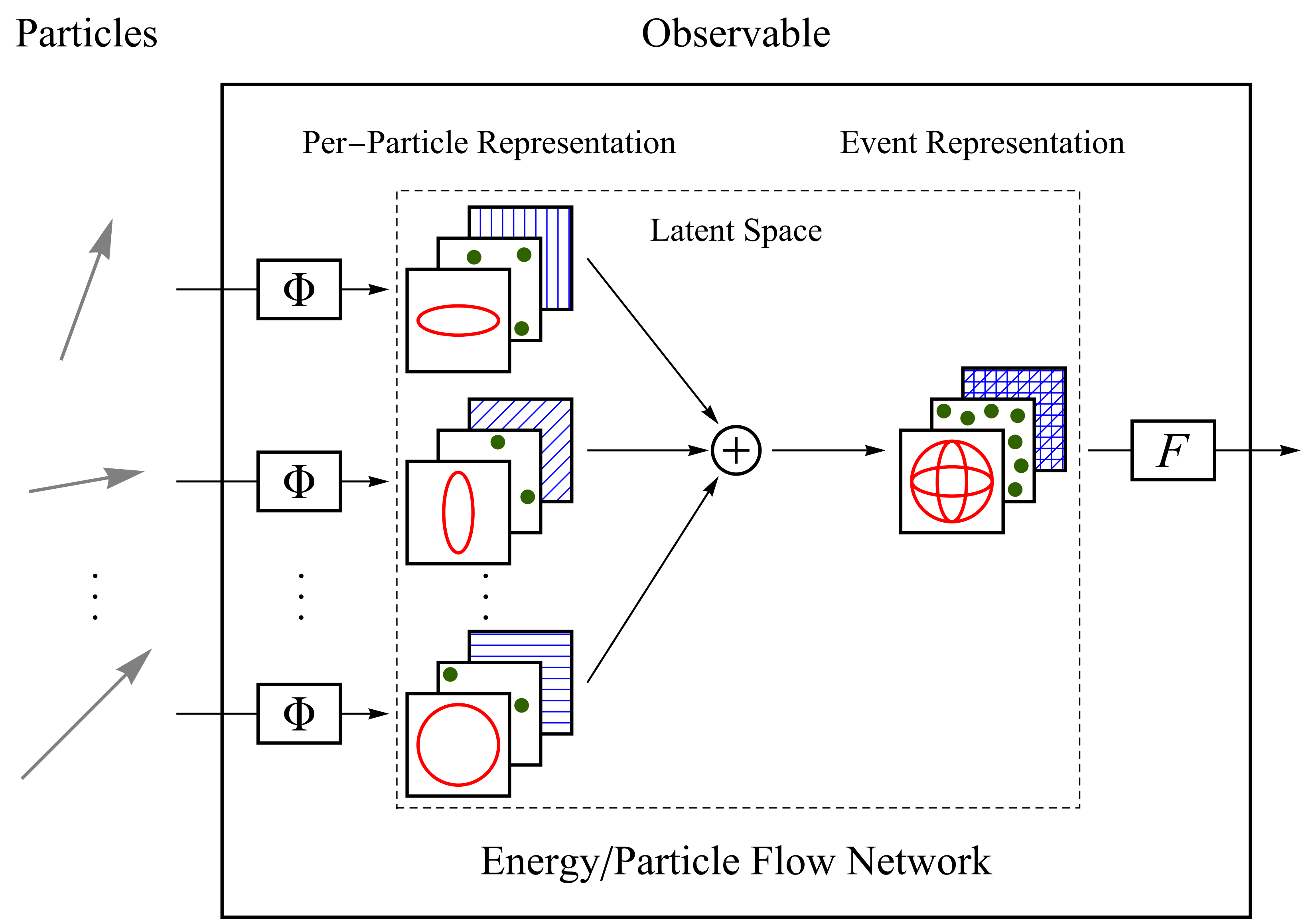
$$\mathrm{PFN}(\{p_1^\mu, \ldots, p_M^\mu\}) = F\left(\sum_{i=1}^{M} \Phi(p_i^\mu)\right)$$

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



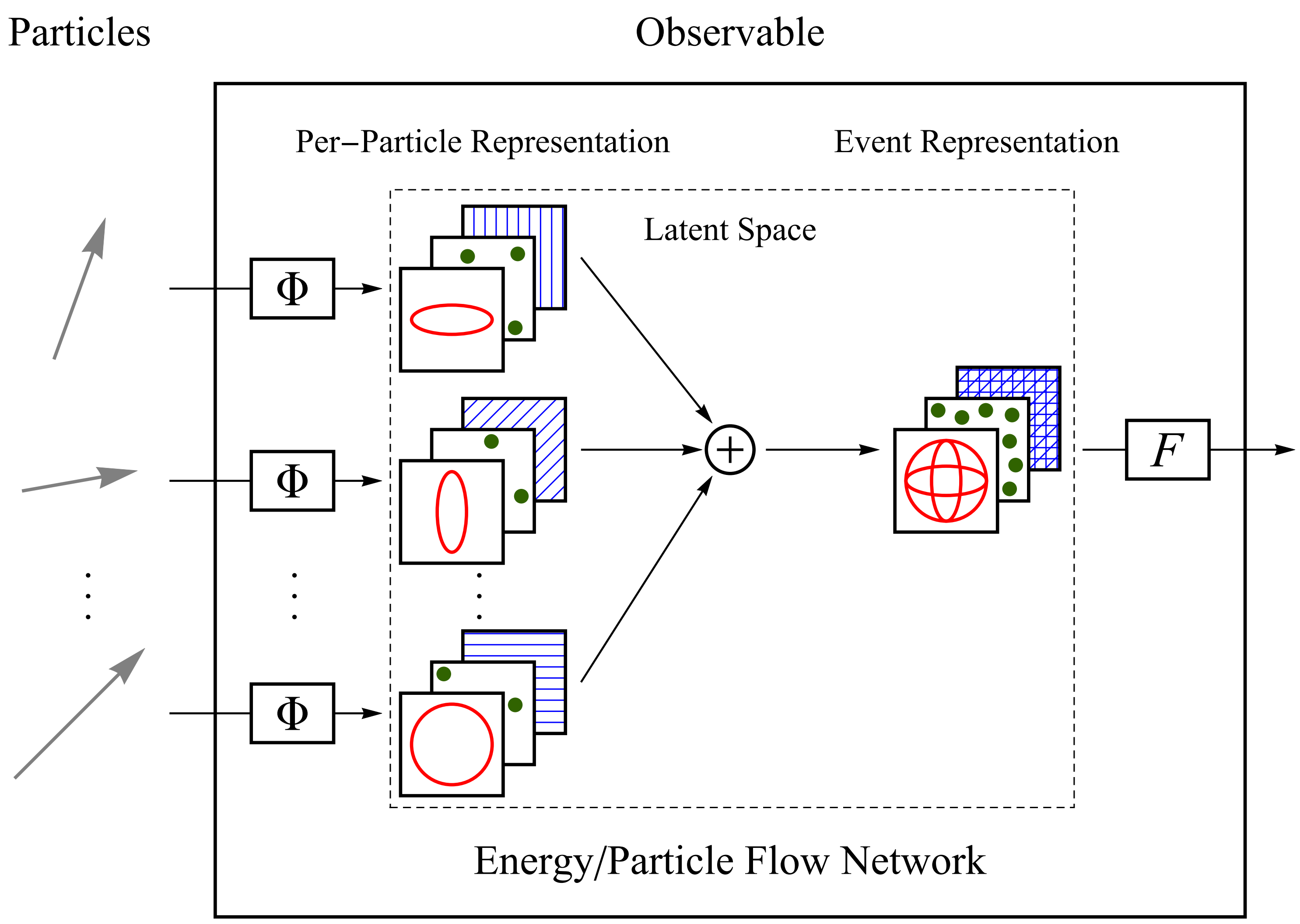
Approximating Φ and *F* with Neural Networks

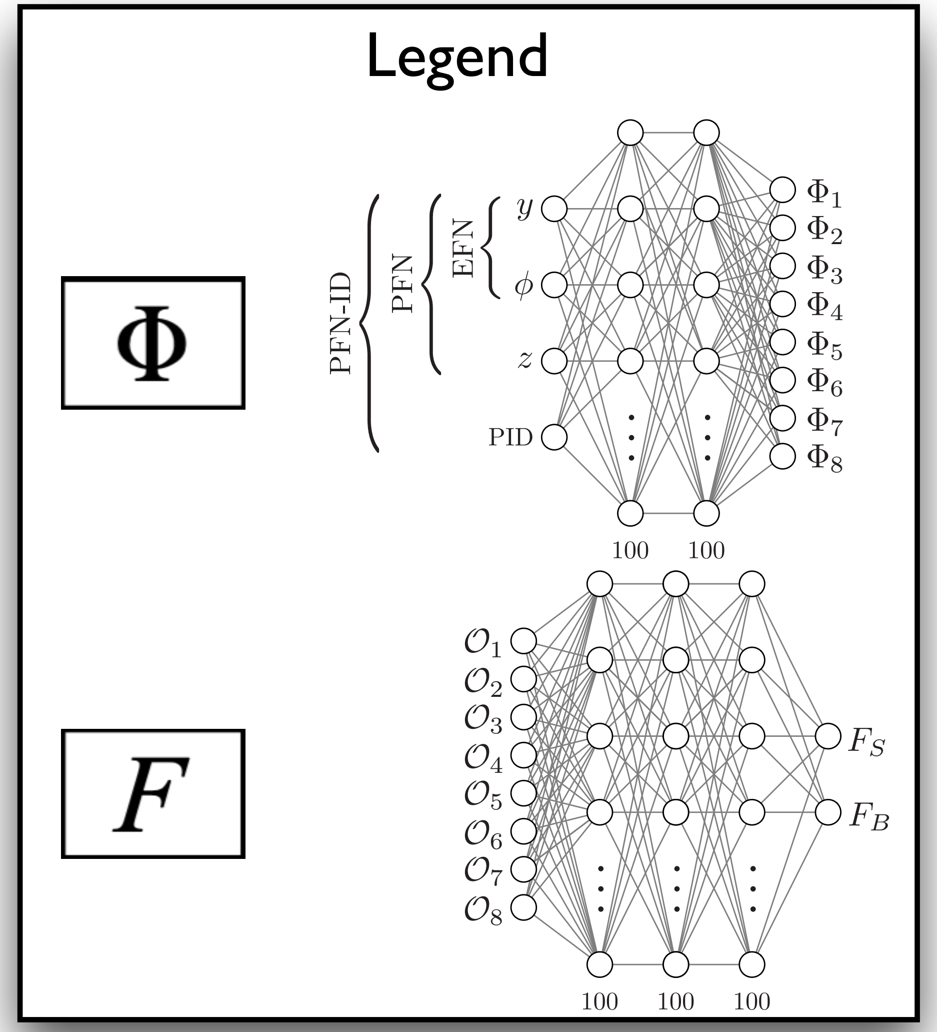
Employ neural networks as arbitrary function approximators

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)





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

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

Quark vs. Gluon: Classification Performance

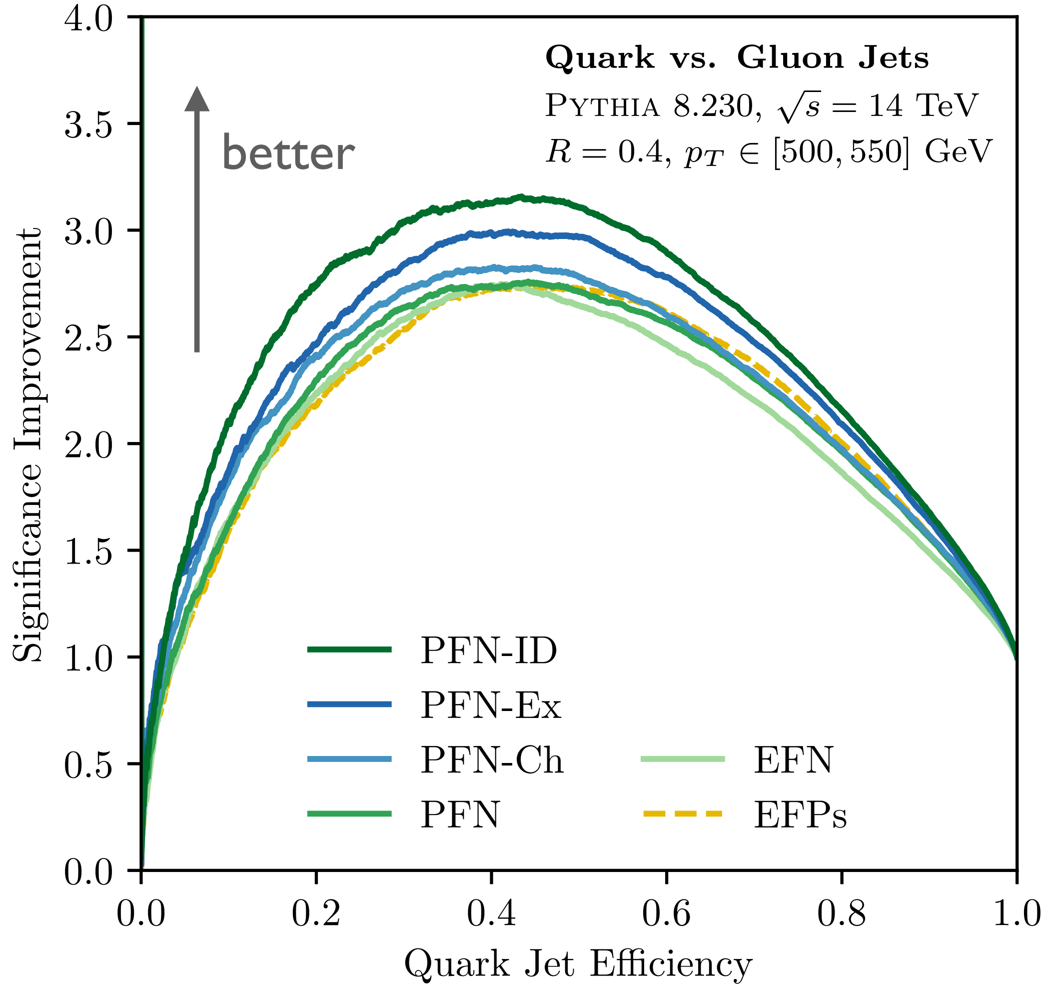
PFN-ID: Full particle flavor info
$(\gamma, \pi^{\pm}, K^{\pm}, K_L, p, \bar{p}, n, \bar{n}, e^{\pm}, \mu^{\pm})$

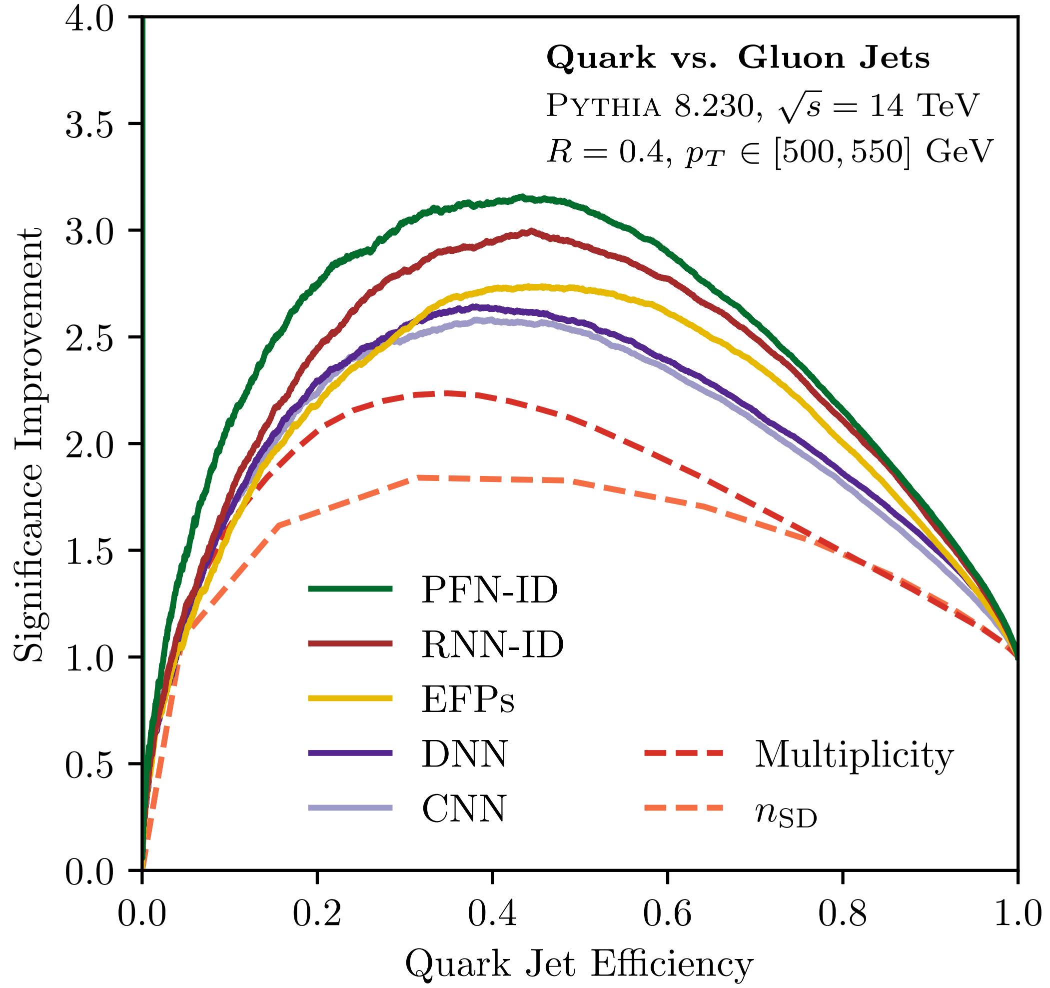
PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^{\pm}, \mu^{\pm})$

PFN-Ch: Particle charge info
(+,0,−)

PFN: No particle type info, arbitrary energy dependence

EFN: IRC-safe 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

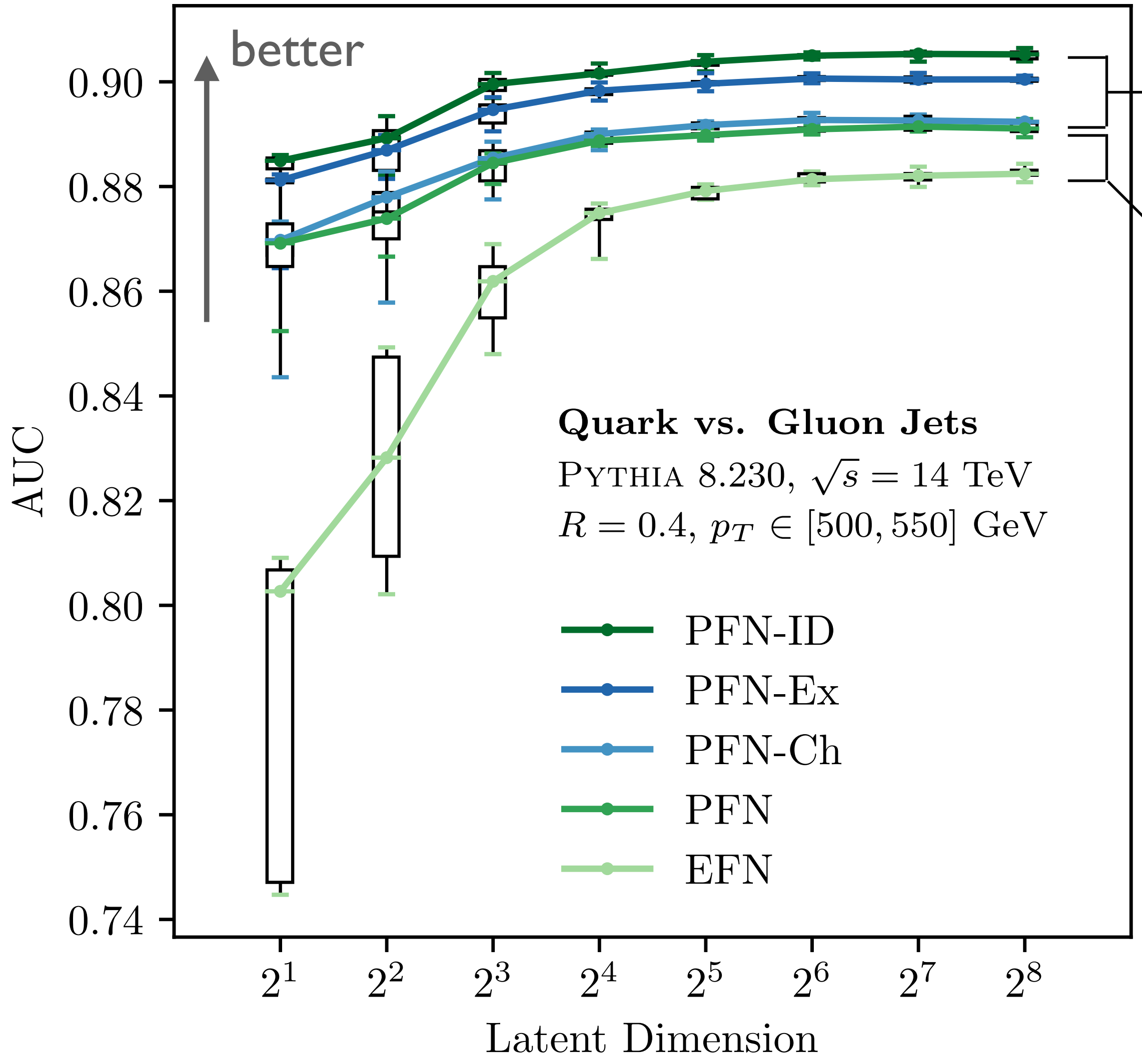
PFN-ID: Full particle flavor info
$(\gamma, \pi^\pm, K^\pm, K_L, p, \bar{p}, n, \bar{n}, e^\pm, \mu^\pm)$

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: IRC-safe latent space

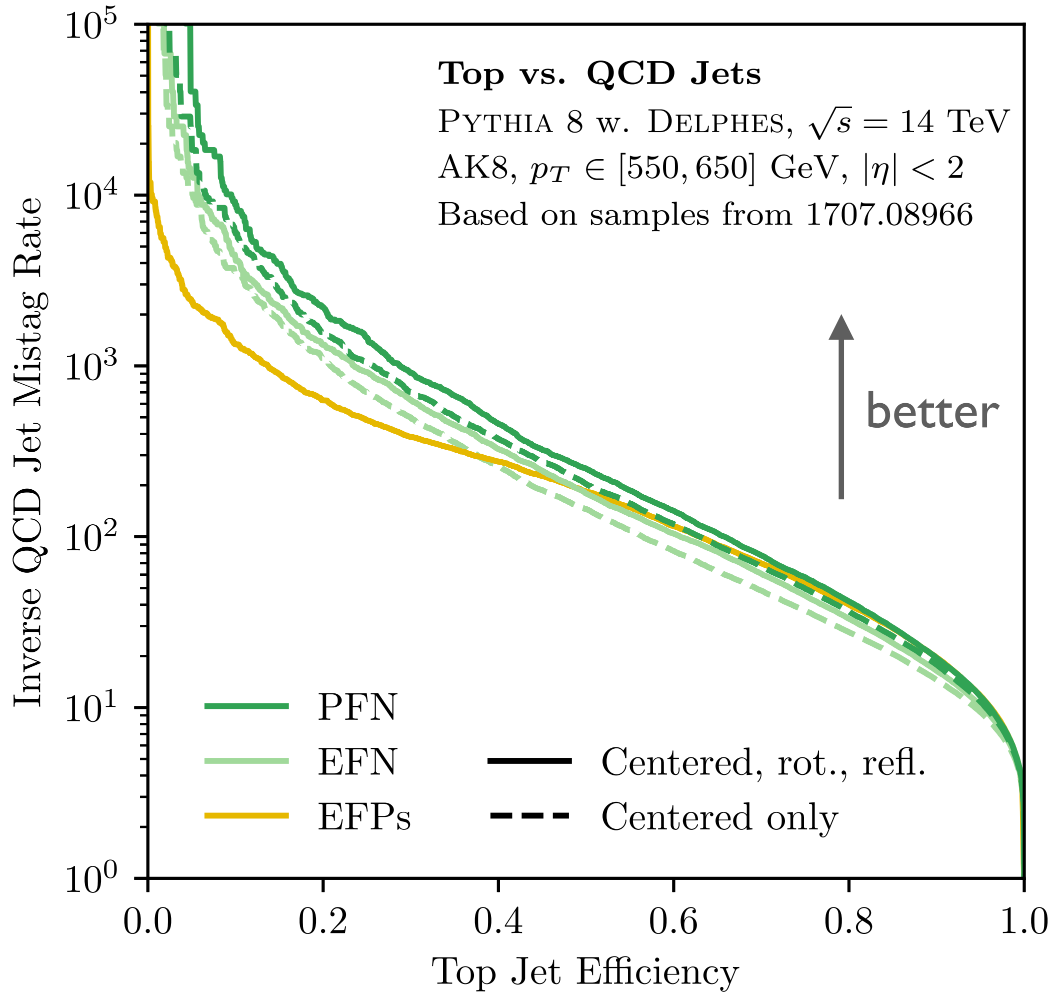


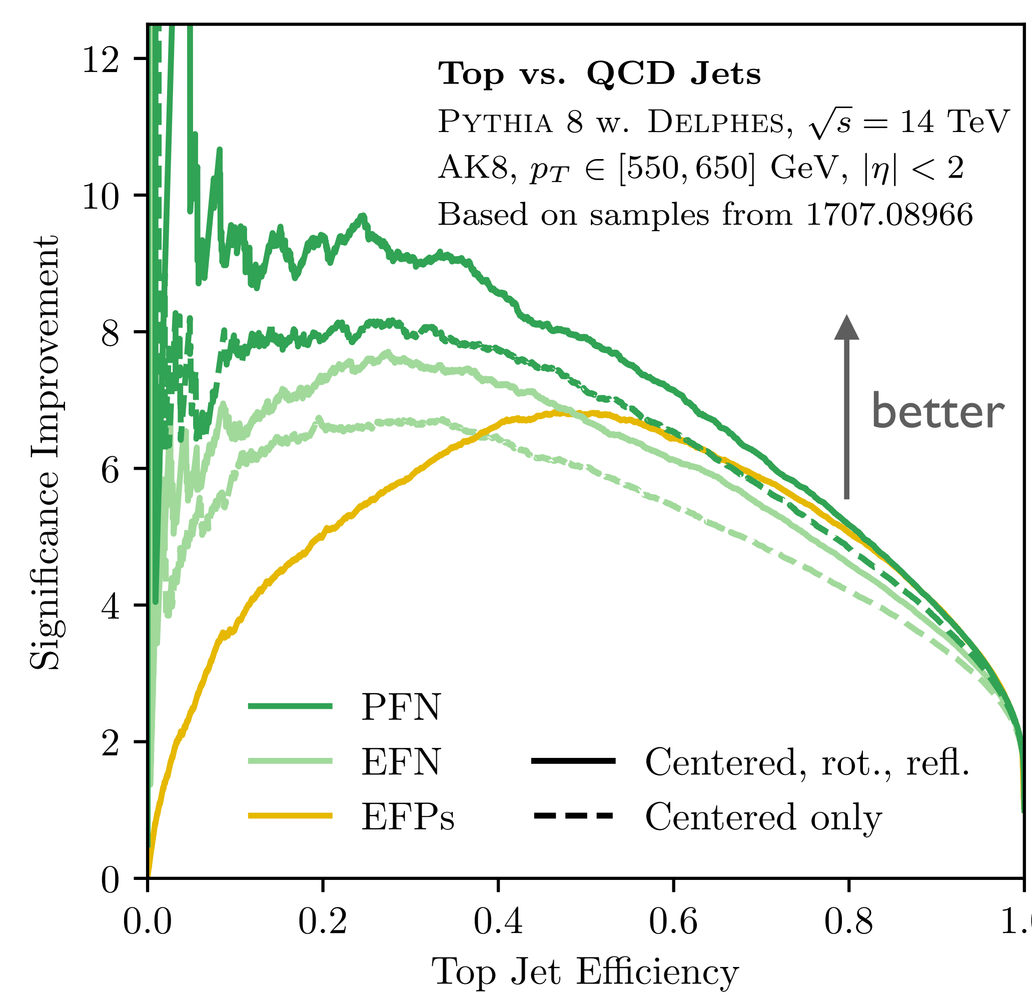
Particle type info helpful

IRC unsafe information clearly helpful

Performance saturates as dimension increases

Boosted Top: Classification Performance



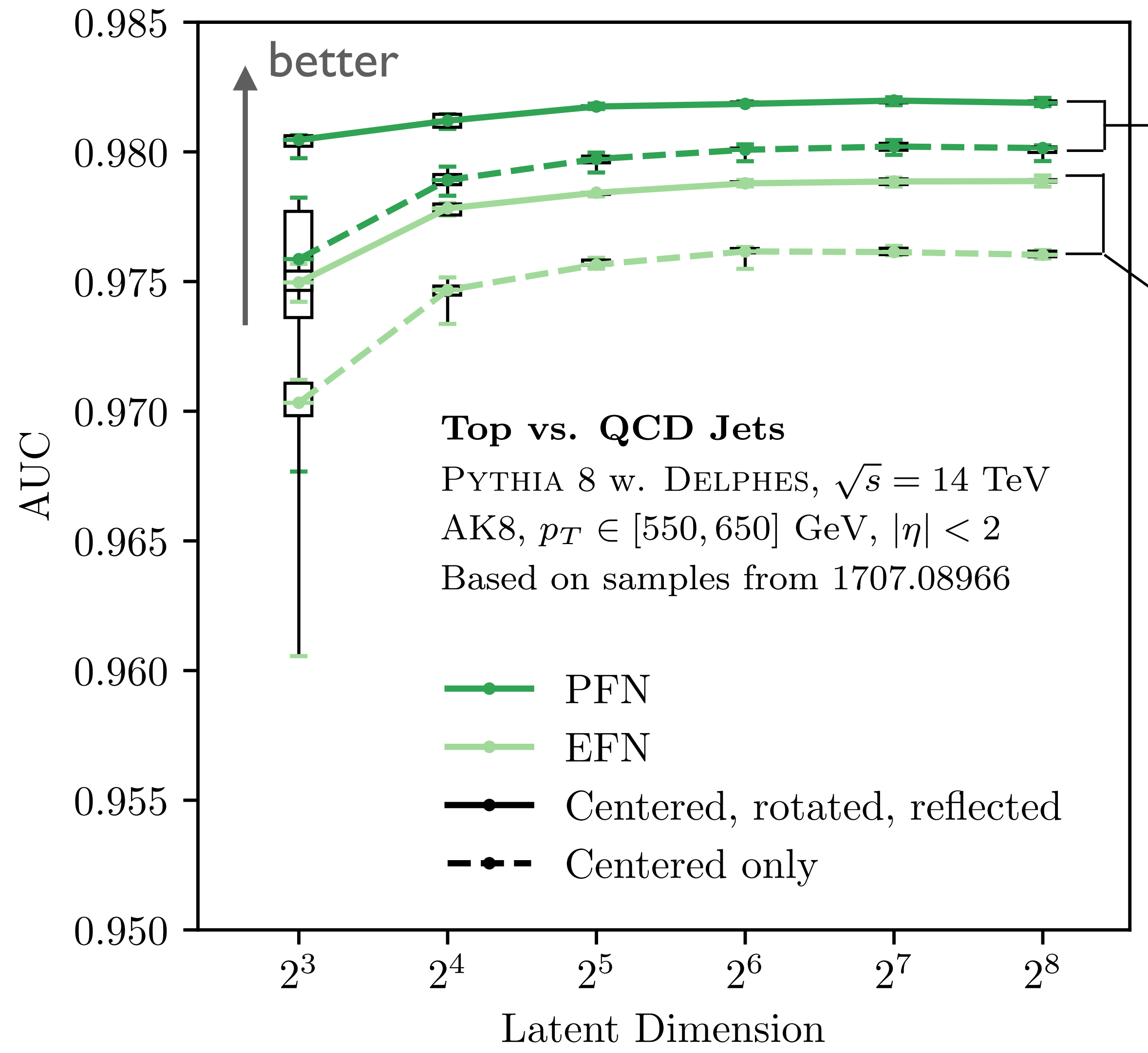


Latent space dimension $\ell = 256$

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

IRC unsafe information clearly helpful

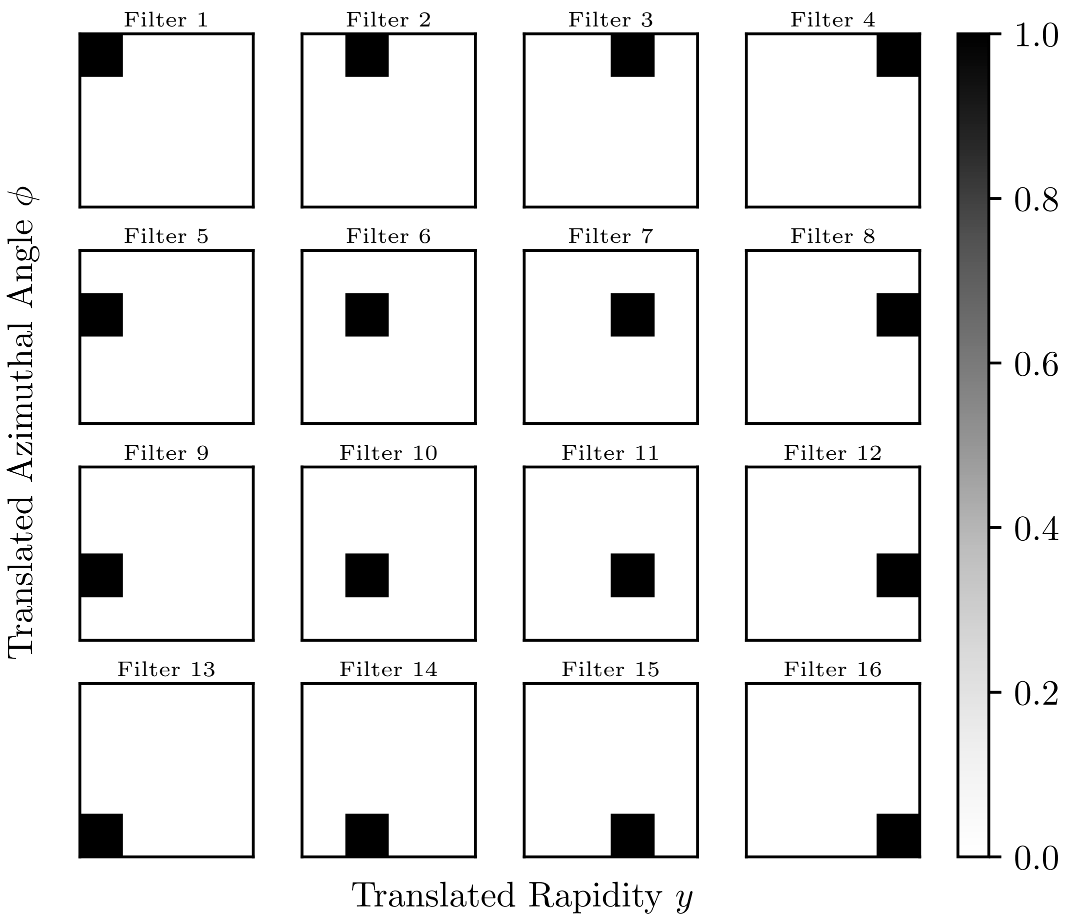
Performance quickly saturates as latent dimension increases

Energy Flow Network Visualization

EFN observables are two-dimensional geometric functions

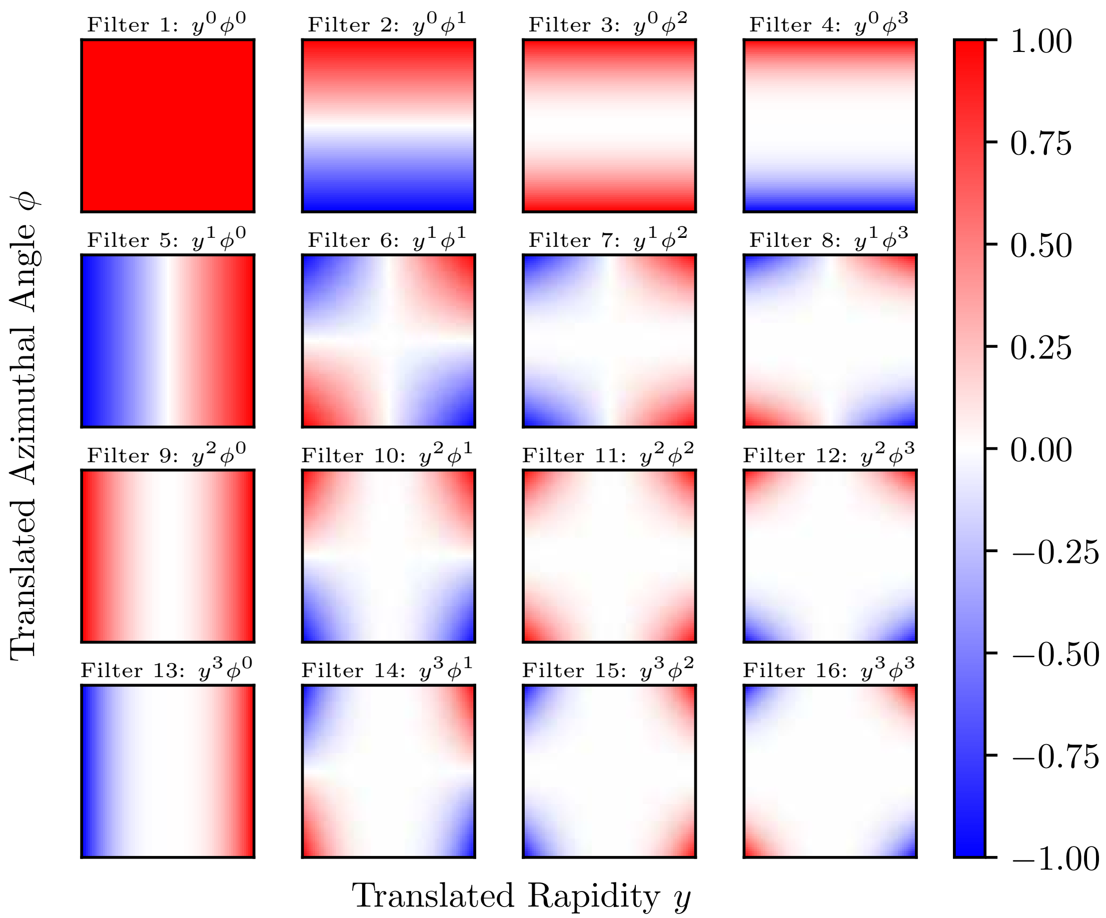
Visualize EFN observables as *filters* in the translated rapidity-azimuth plane

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]

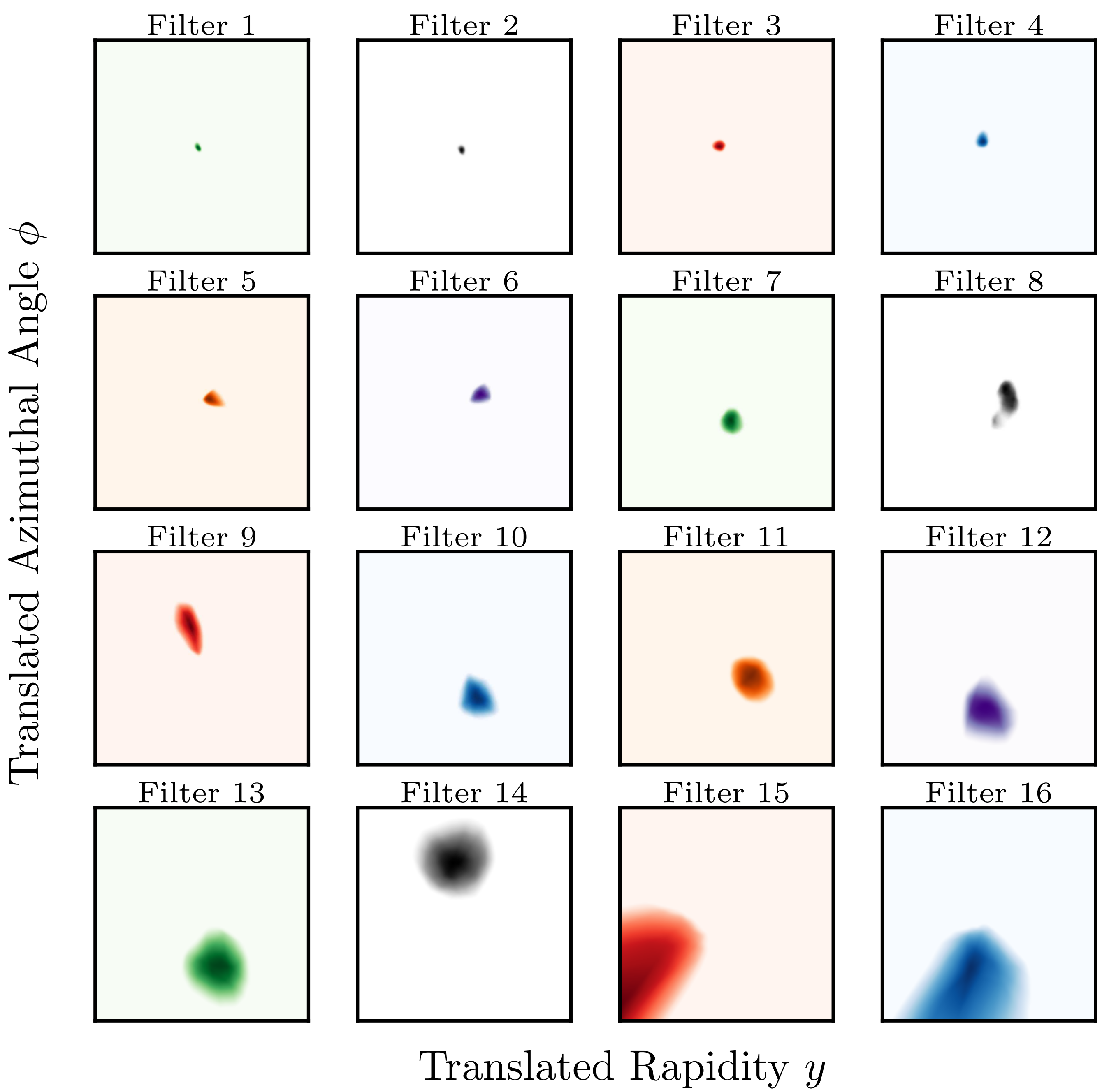
Quark vs. Gluon: Visualizing EFN Filters

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



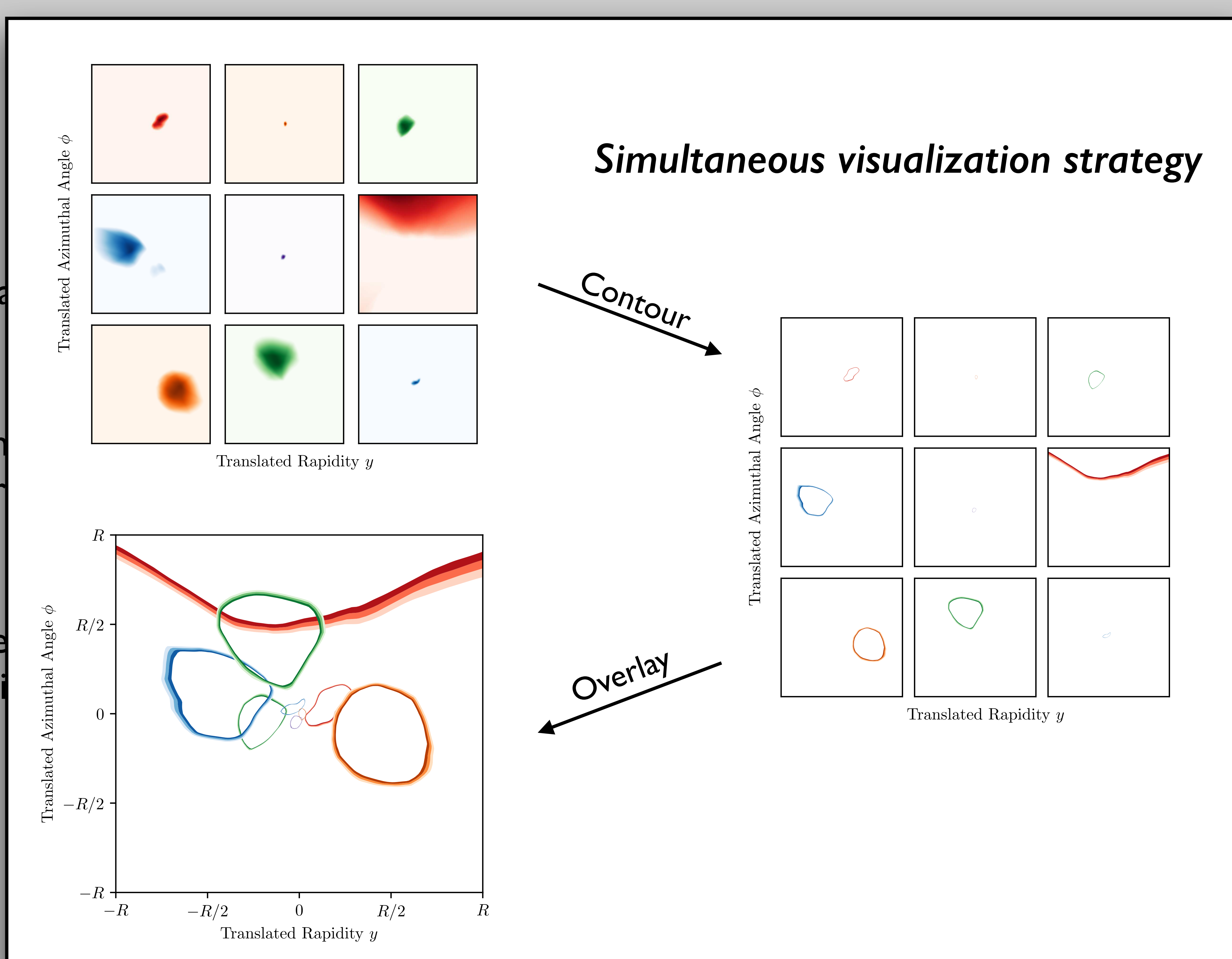
Quark vs. Gluon: Visualizing EFN Filters

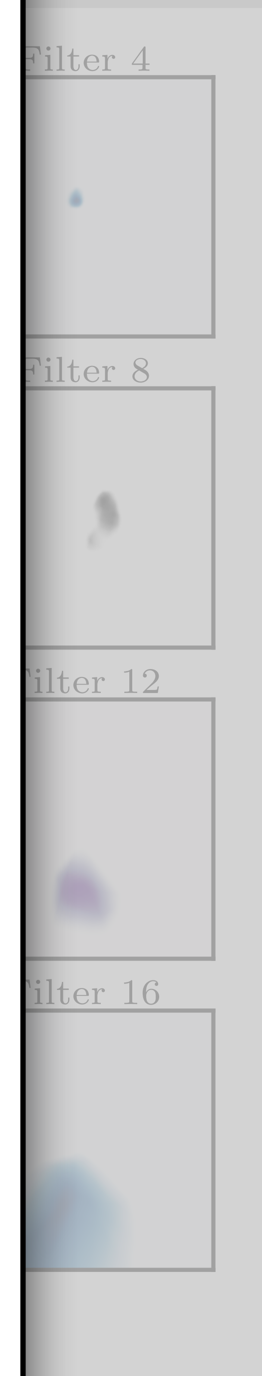
by size

Genera

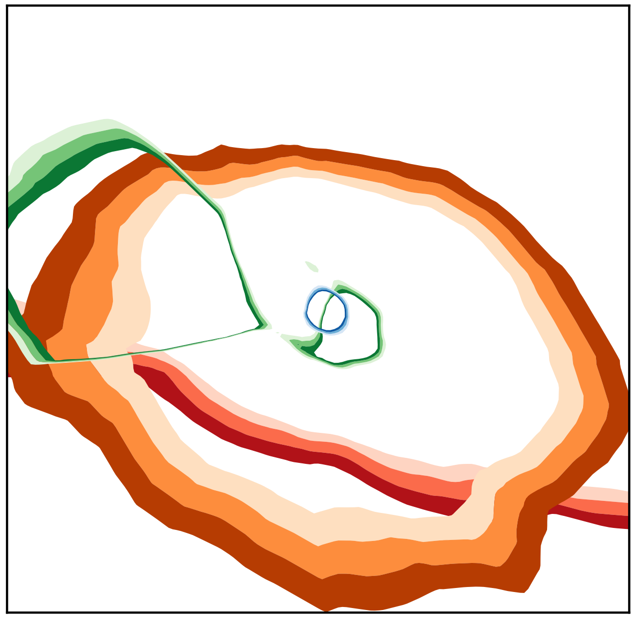
Local n
interpr

EFN se
dynami

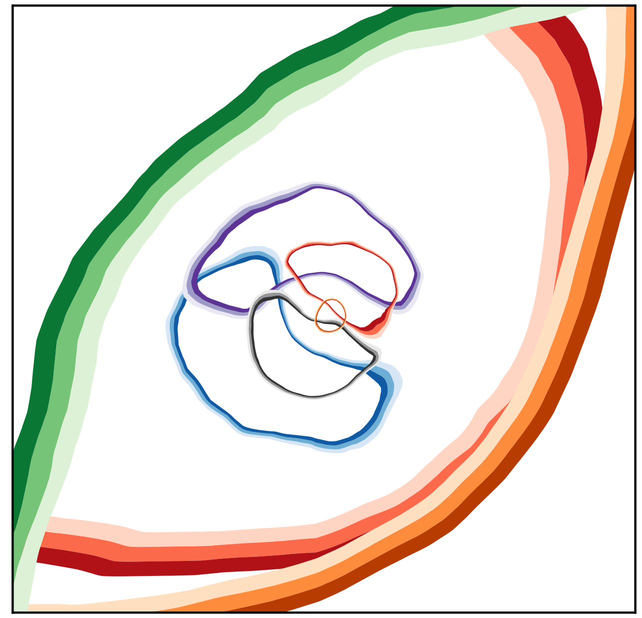




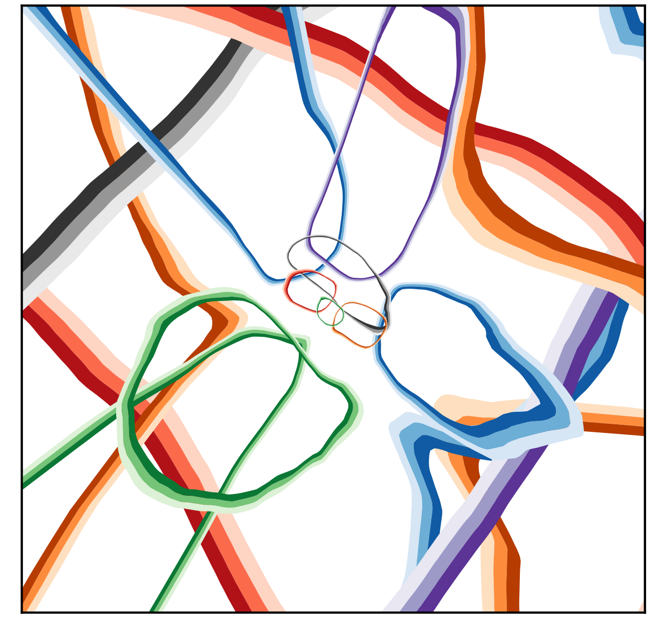
Quark vs. Gluon: Visualizing EFN Filters



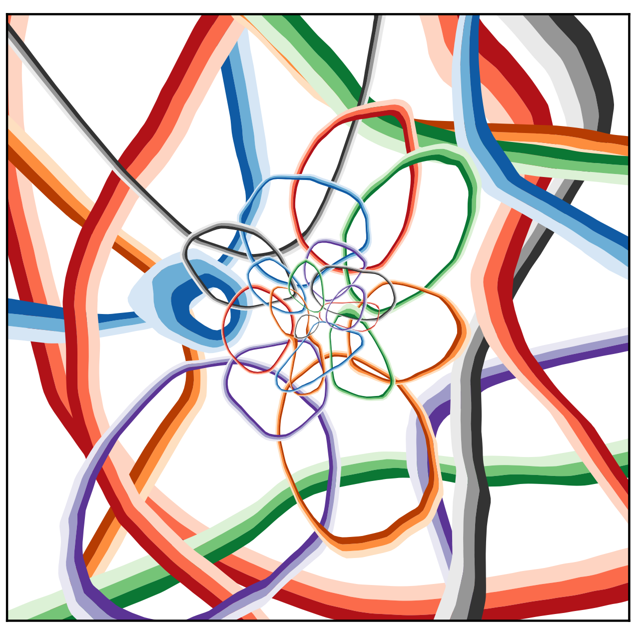
$\ell = 4$



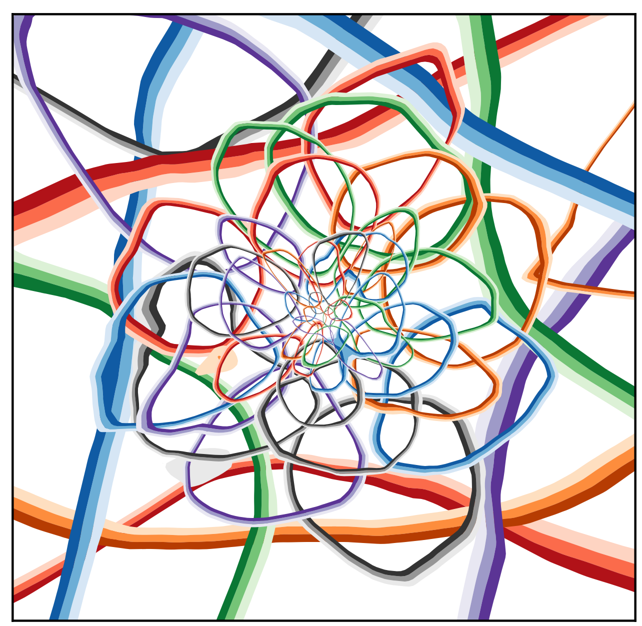
$\ell = 8$



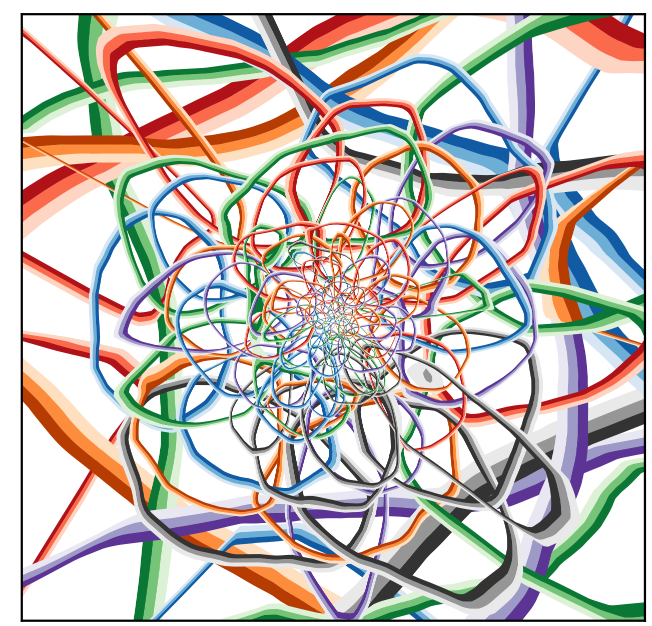
$\ell = 16$



$\ell = 32$

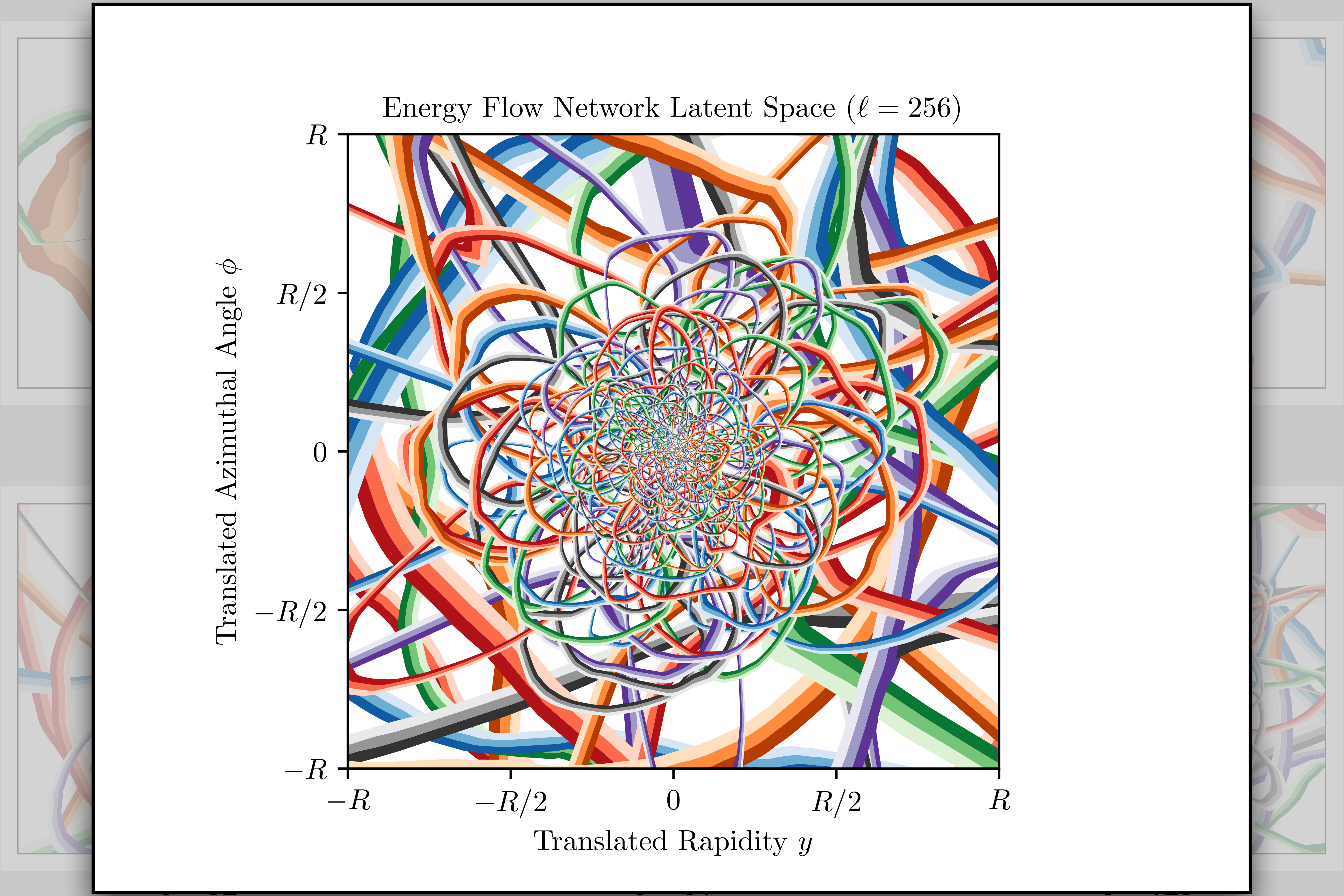


$\ell = 64$

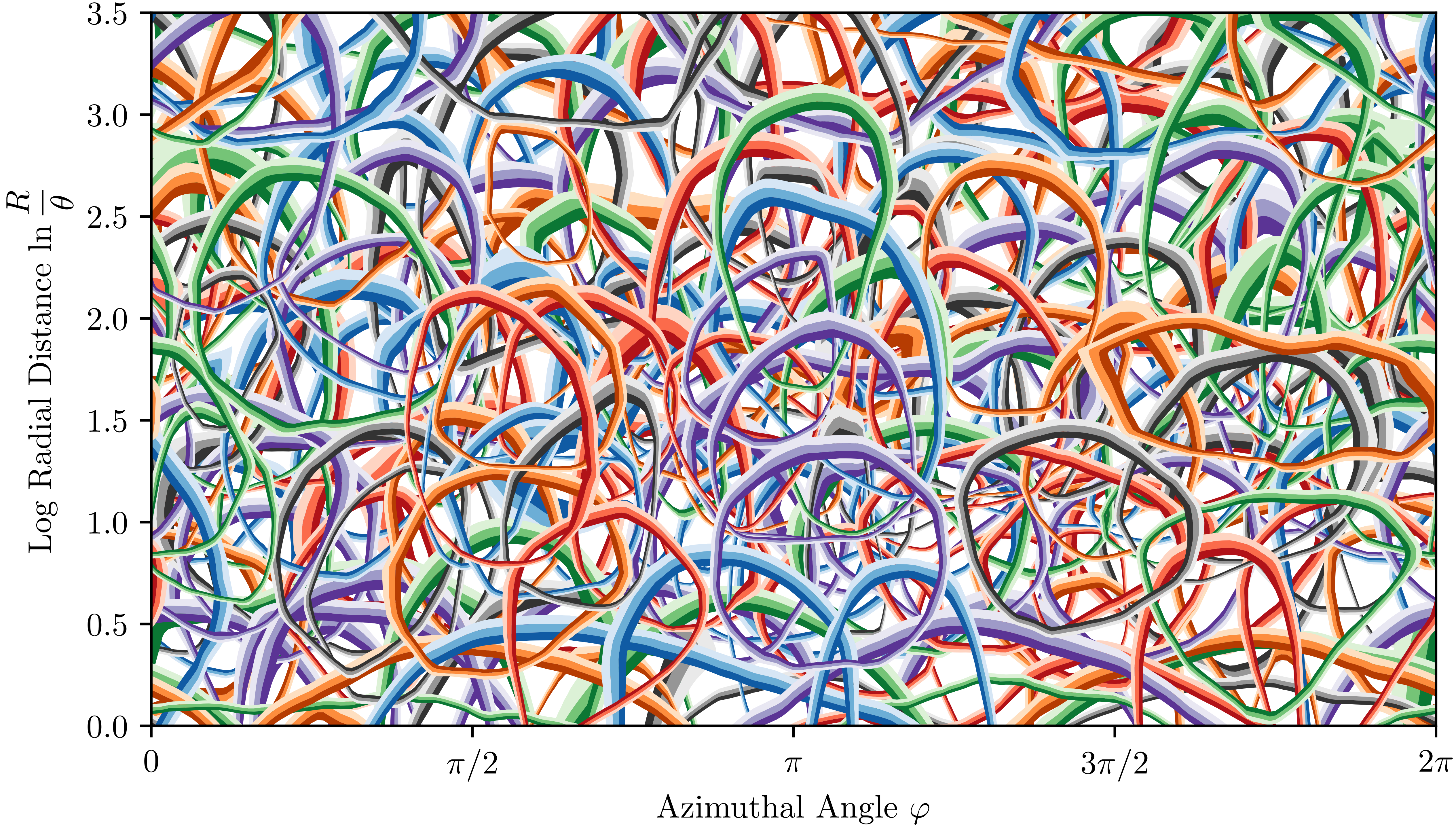


$\ell = 128$

Quark vs. Gluon: Visualizing EFN Filters



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



Quark vs. Gluon: Measuring EFN Filters

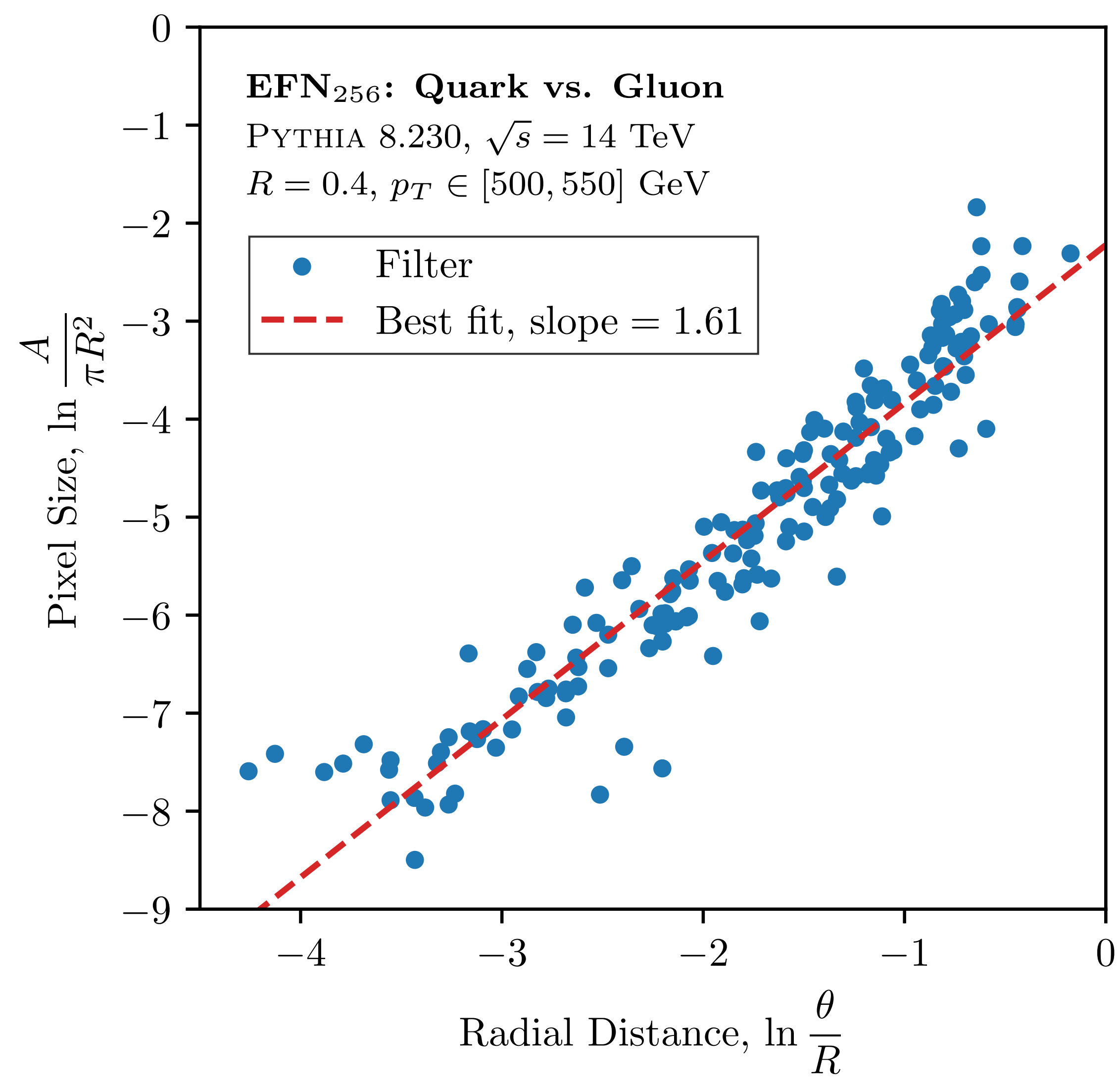
Power-law dependence between filter size and distance from center is observed

Slope of 2 is predicted at leading log

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

Emission plane area element

Area element in rap-phi plane

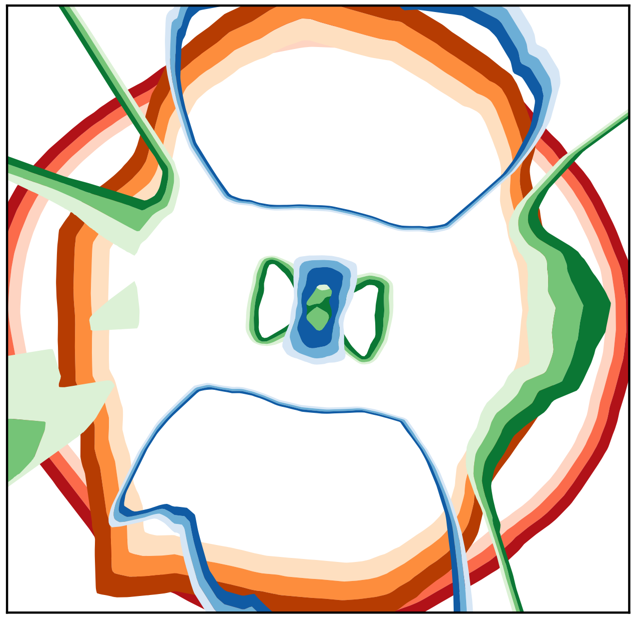


Non-perturbative physics, axis recoil, higher order effects cause deviations from slope of 2

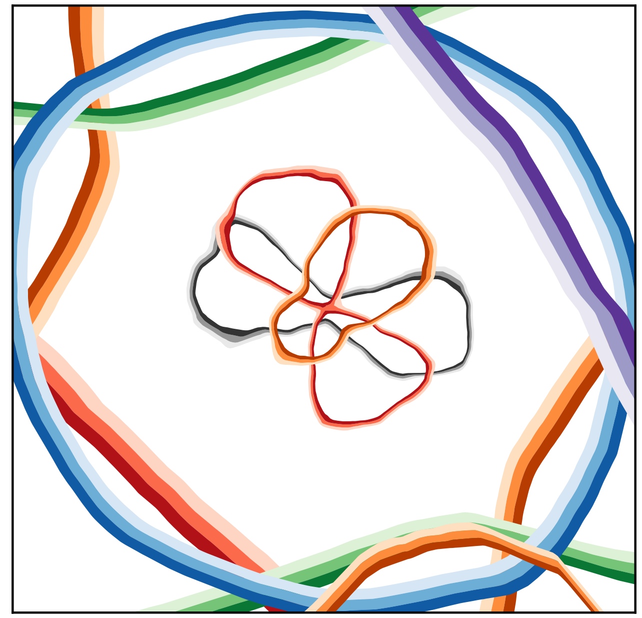
$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta}\frac{dz}{z}$$

Boosted Top: Visualizing EFN Filters

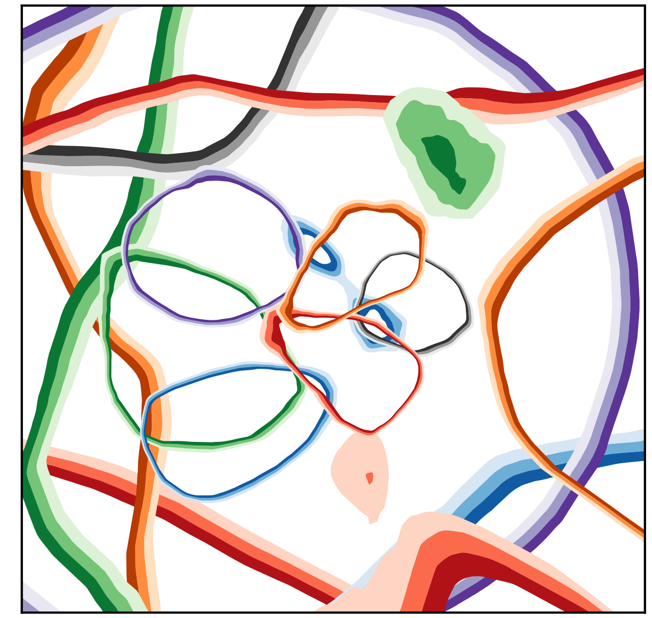
Without rotation/reflection preprocessing



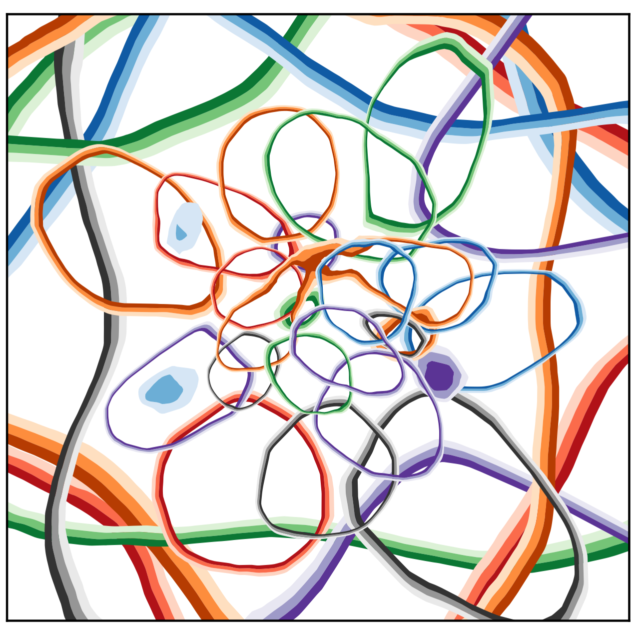
$\ell = 4$



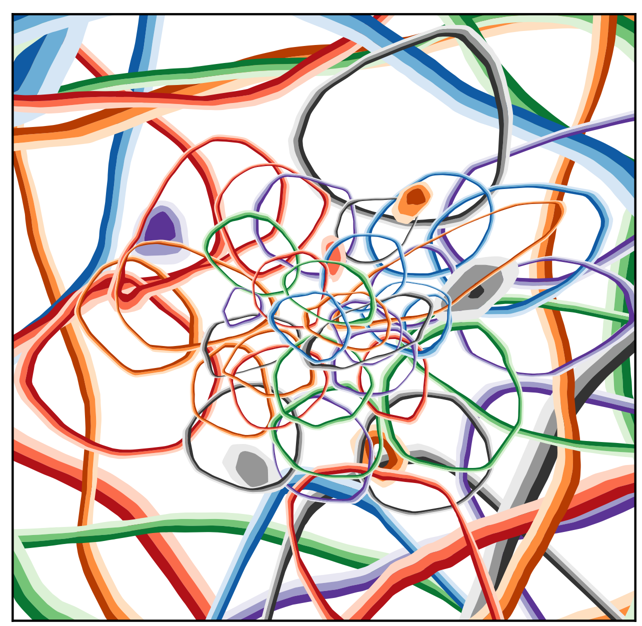
$\ell = 8$



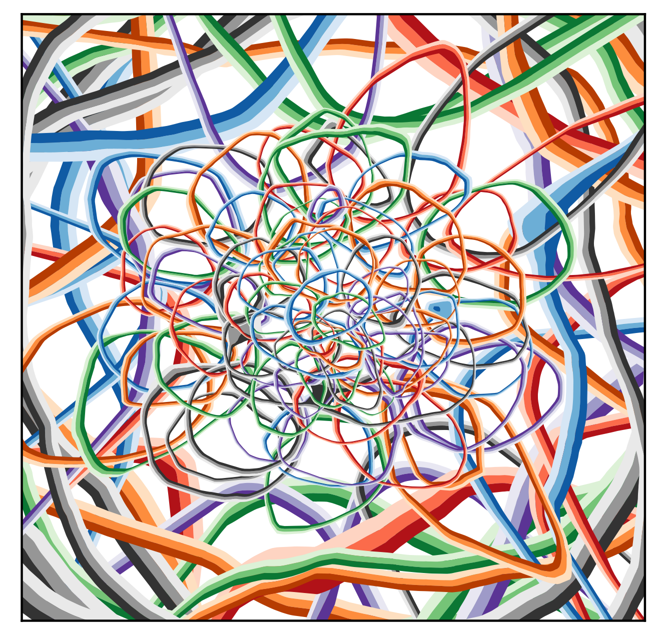
$\ell = 16$



$\ell = 32$



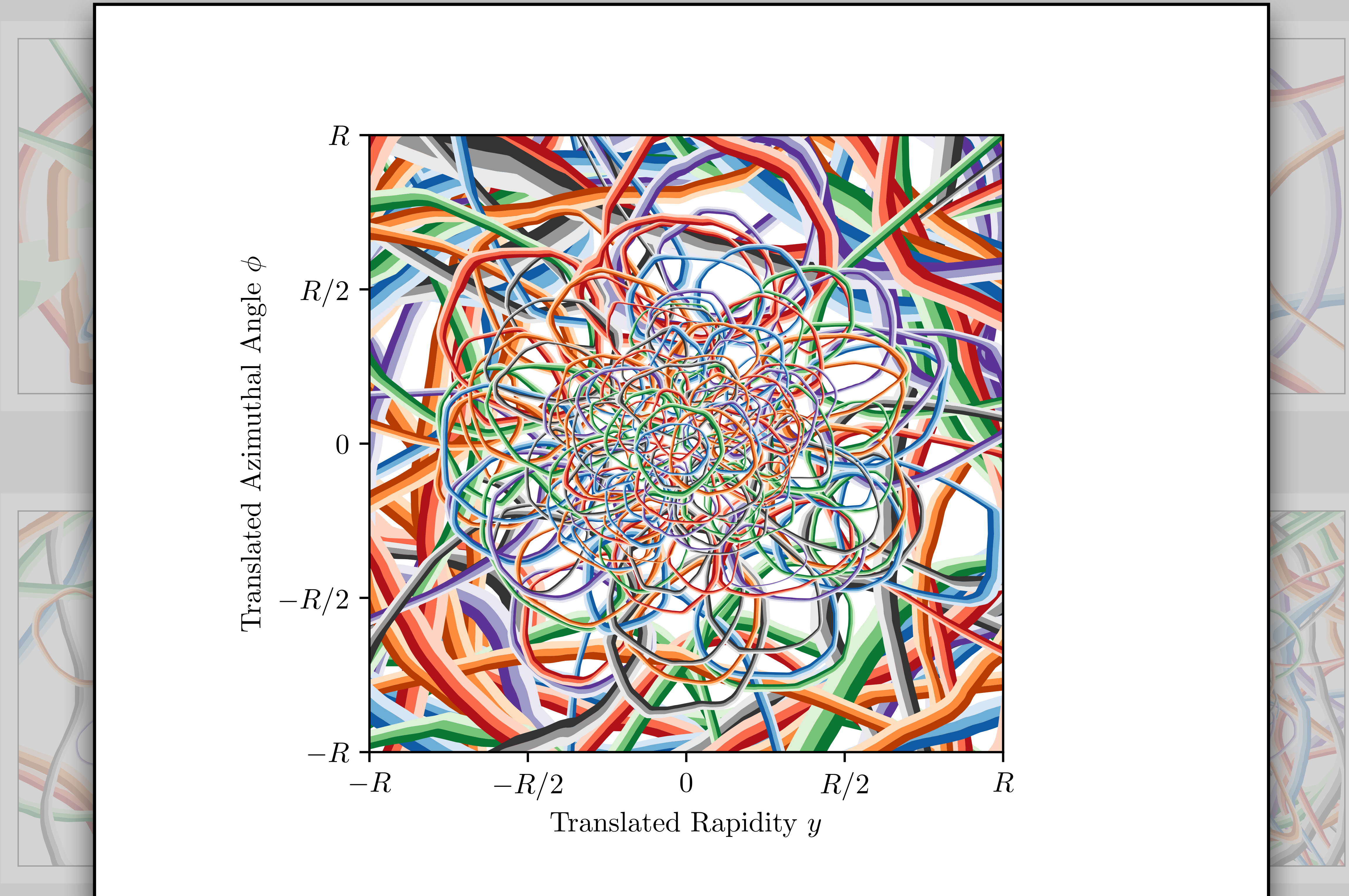
$\ell = 64$



$\ell = 128$

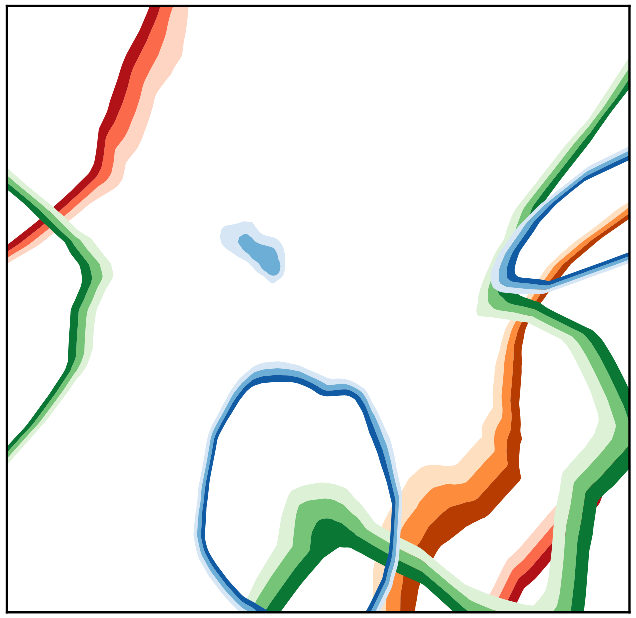
Boosted Top: Visualizing EFN Filters

Without rotation/reflection preprocessing

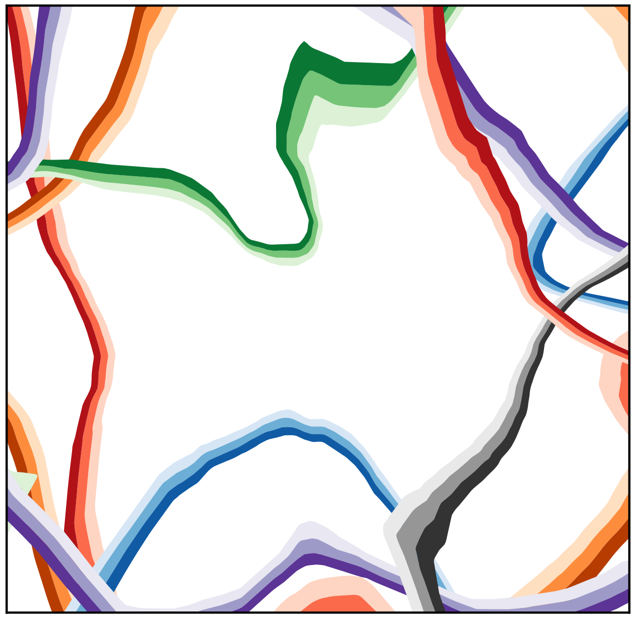


Boosted Top: Visualizing EFN Filters

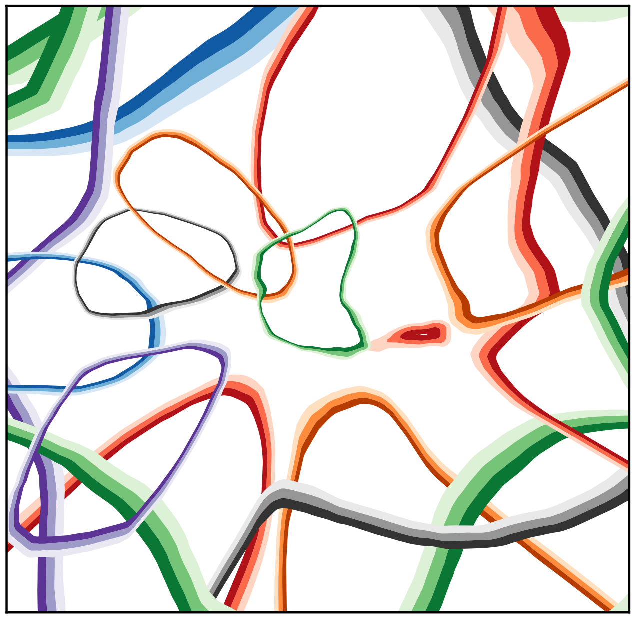
With rotation/reflection preprocessing



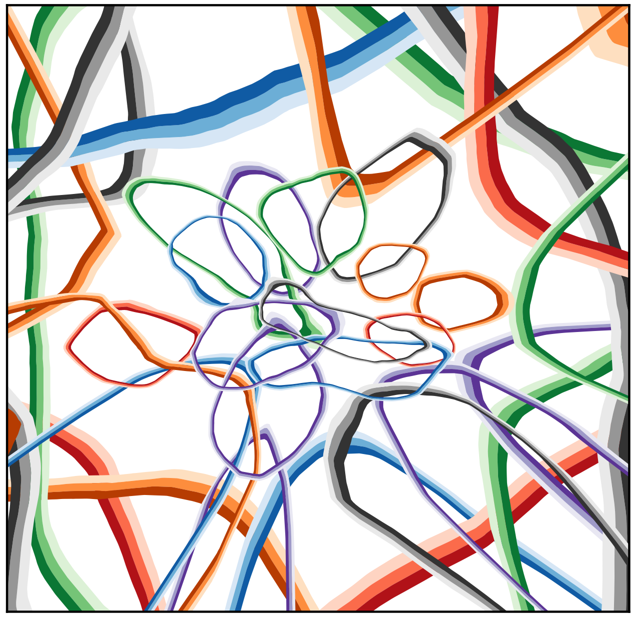
$\ell = 4$



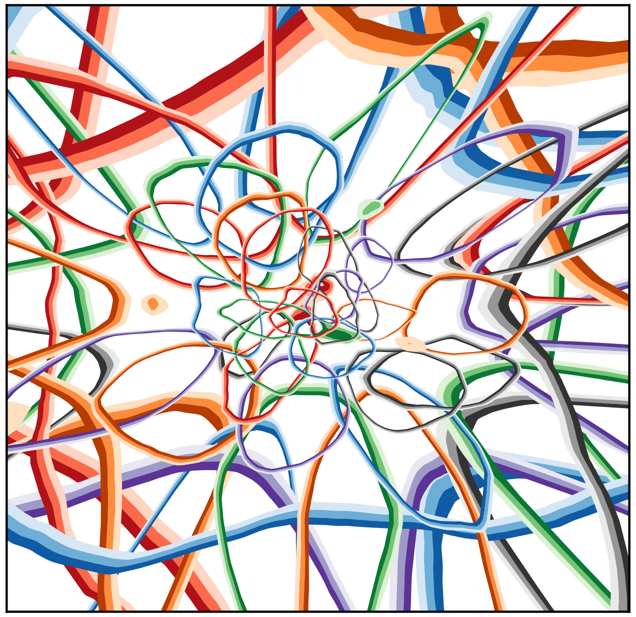
$\ell = 8$



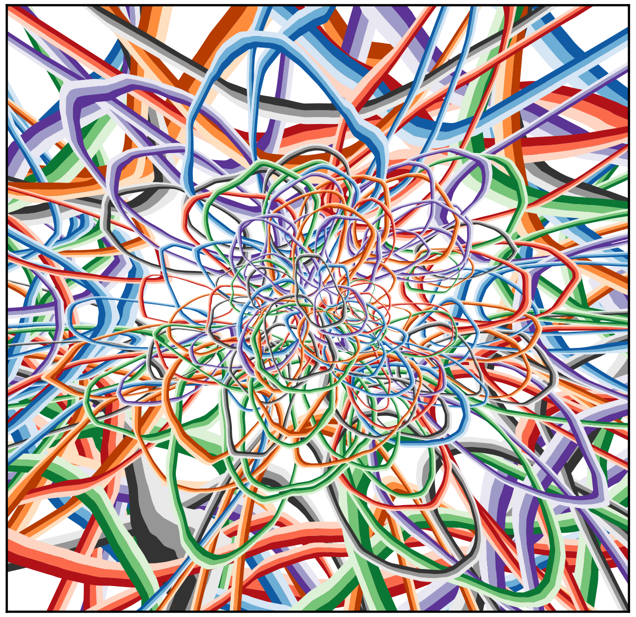
$\ell = 16$



$\ell = 32$



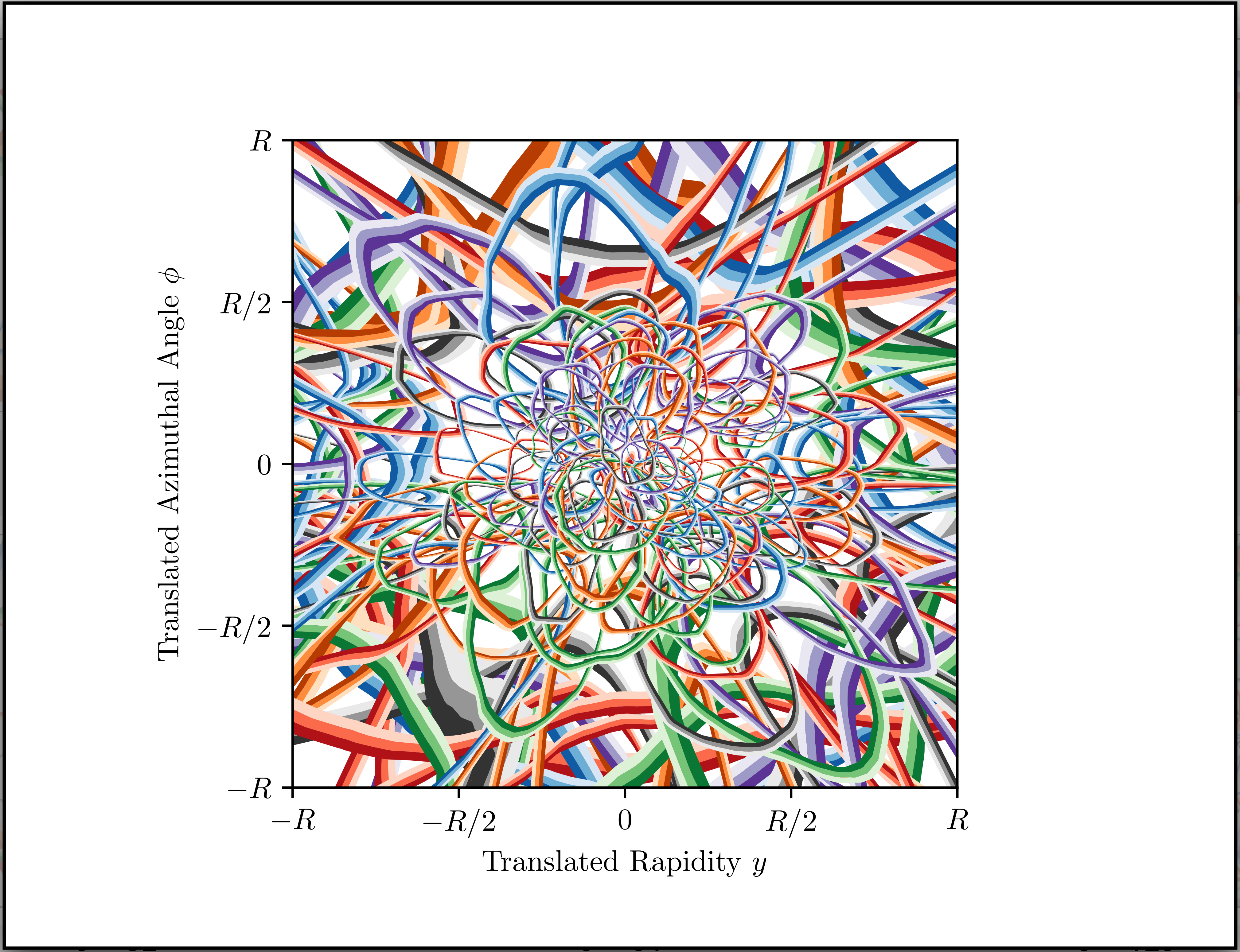
$\ell = 64$



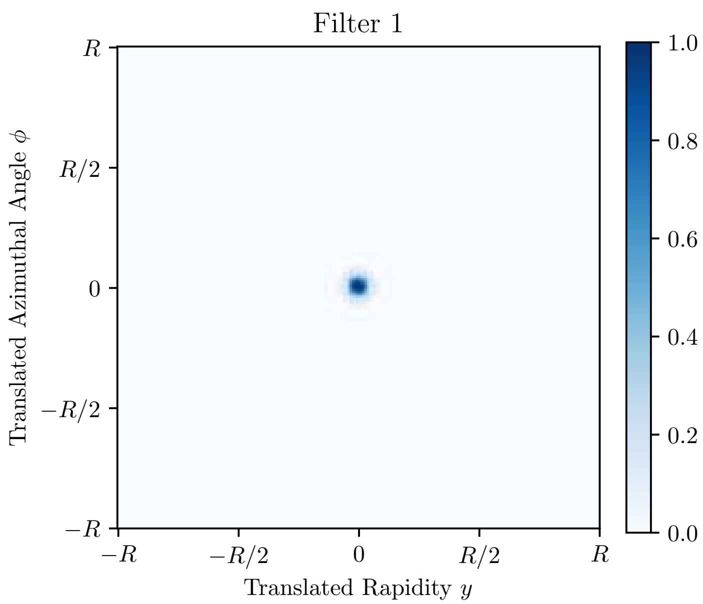
$\ell = 128$

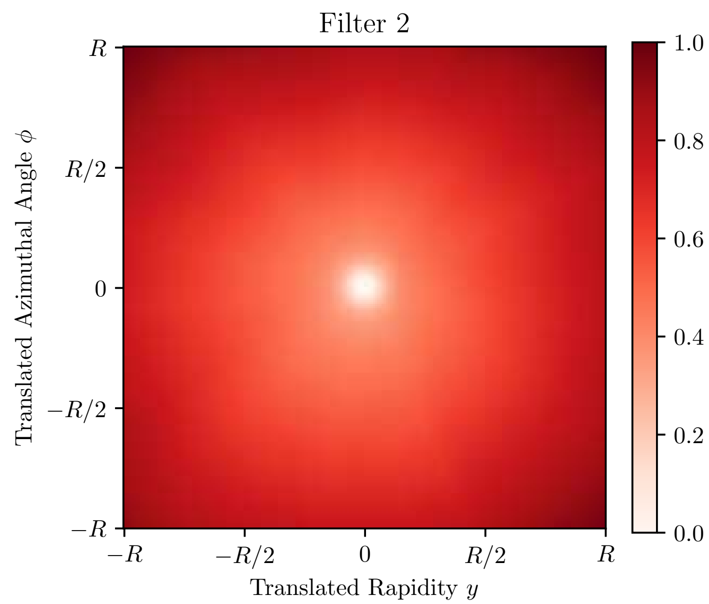
Boosted Top: Visualizing EFN Filters

Without rotation/reflection preprocessing

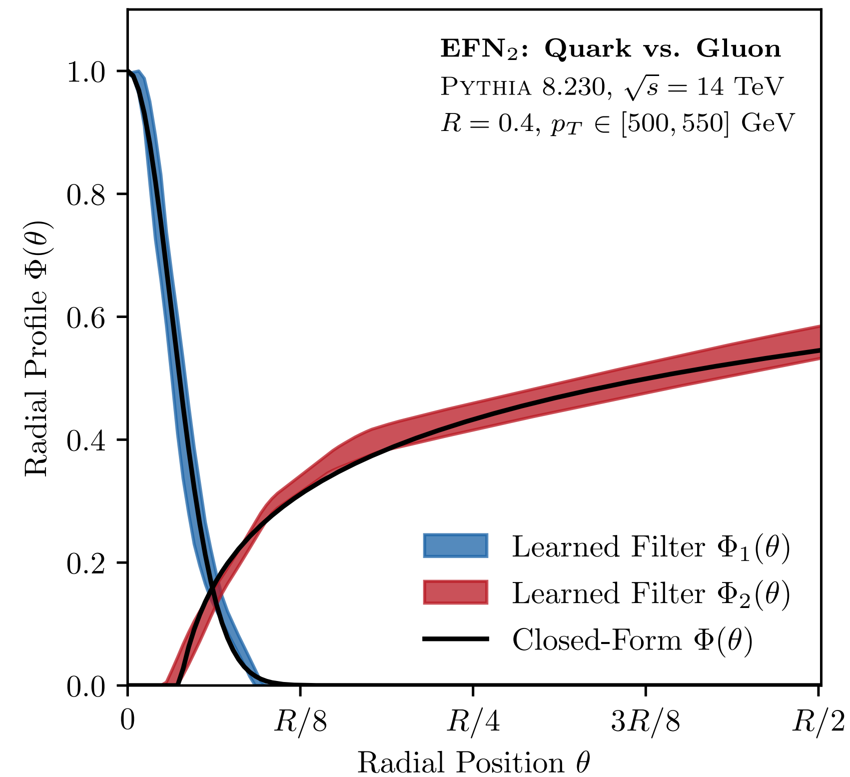


Quark vs. Gluon: Extracting New Analytic Observables





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



Take radial slices to obtain envelope

EFN (ℓ = 2) has approximately radially symmetric filters

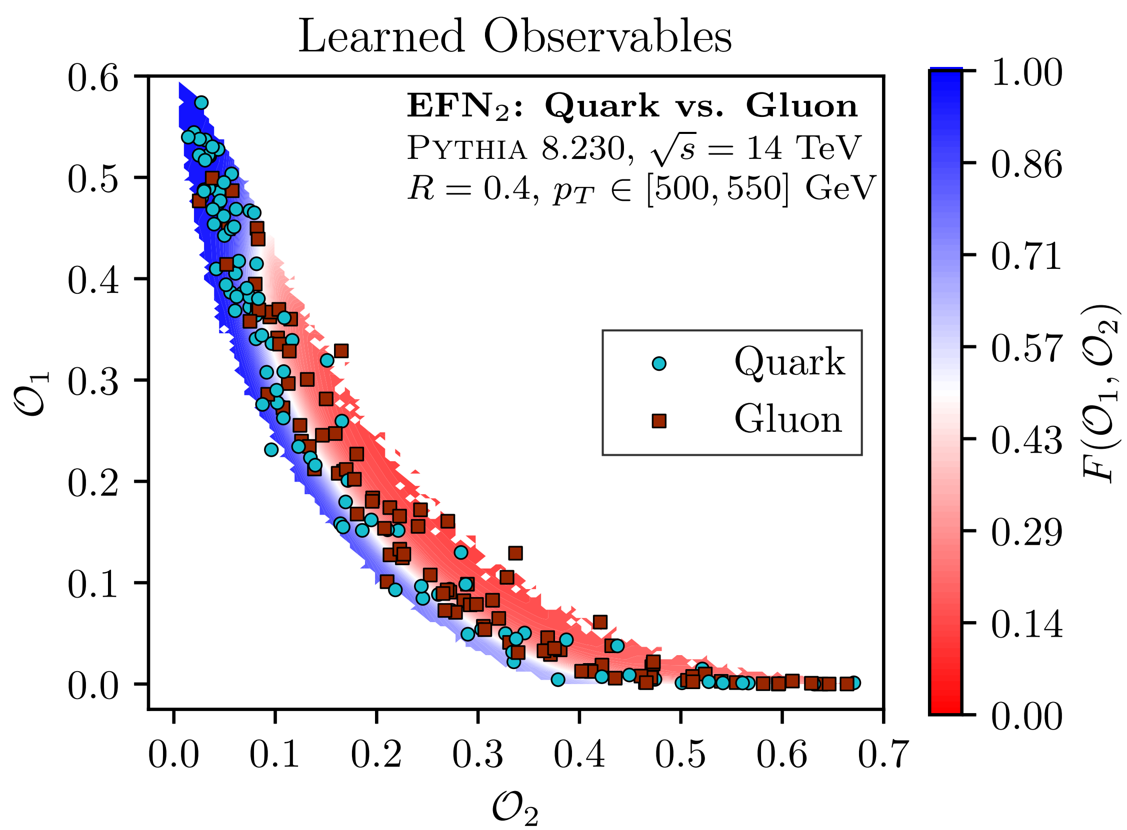
Fit functions of the forms:

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

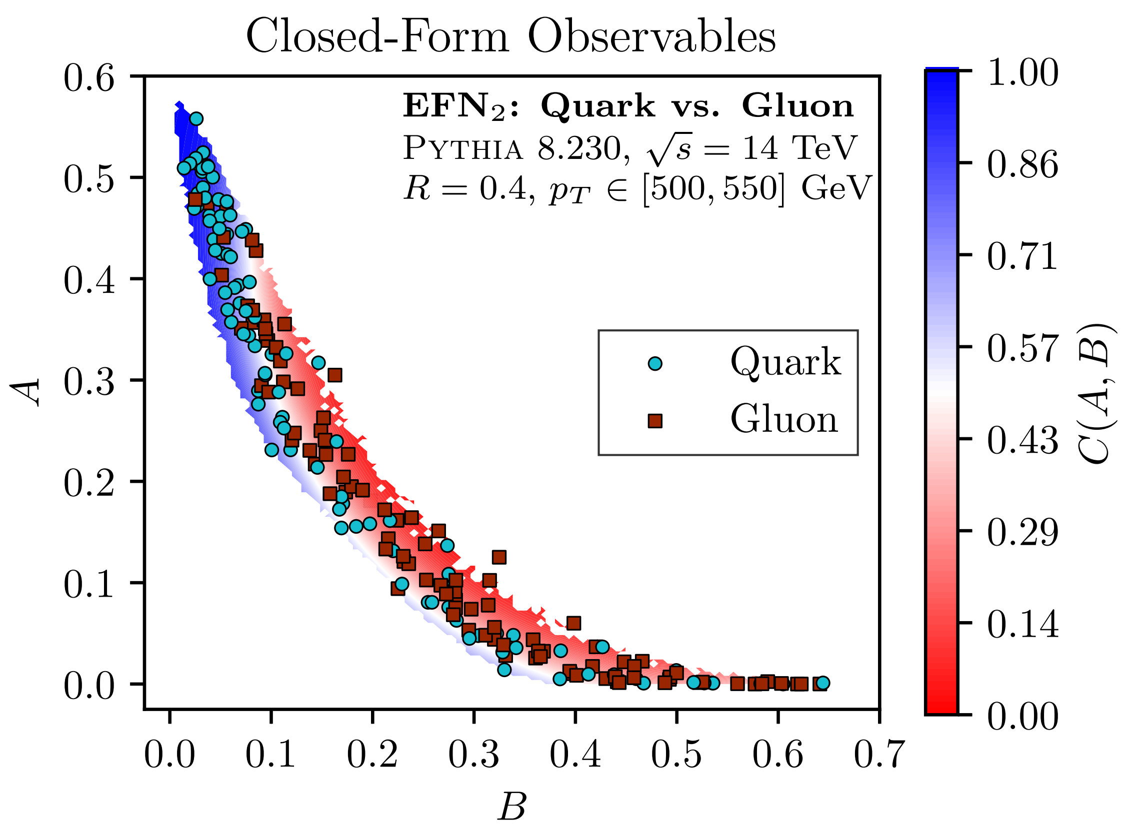
Separate soft and collinear phase space regions

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:

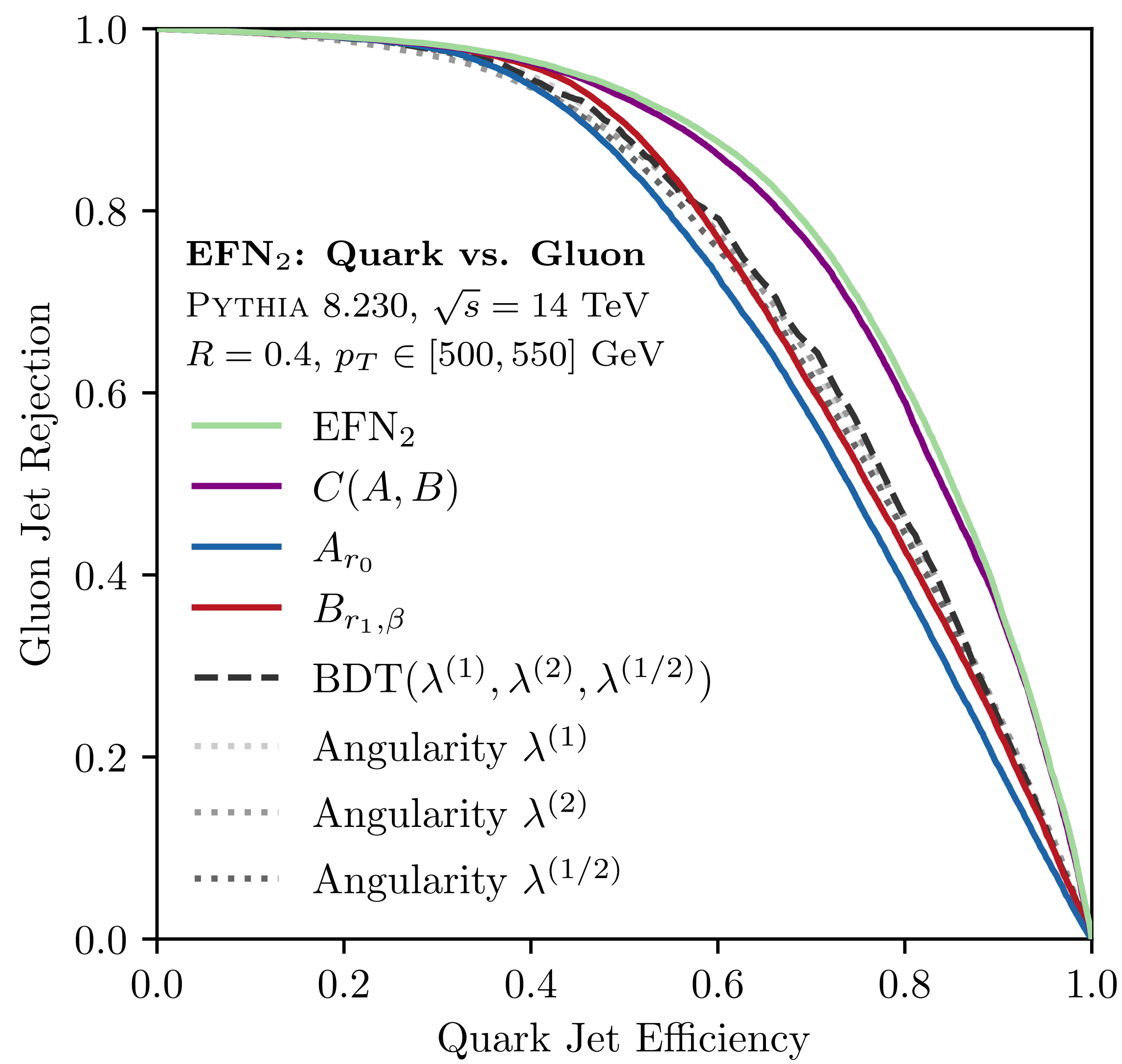
$$C(A,B) = (A - a_0)^2 + (B - b_0)^2$$

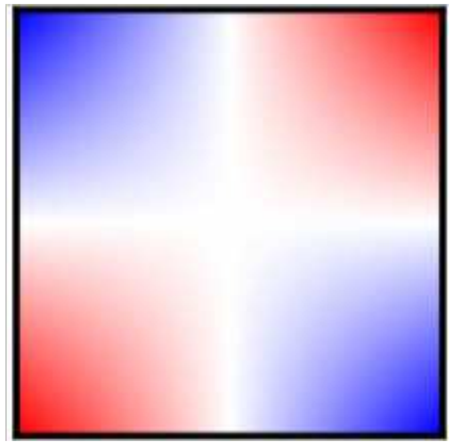
Quark vs. Gluon: Benchmarking New Analytic Observables

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

Energy Flow Networks

Energy Flow Moments

Relating EFPs and EFNs via additivity

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 **linear in M to compute!**

$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1}z_{i_2}z_{i_3}\theta_{i_1i_2}^2\theta_{i_1i_3}^2\theta_{i_2i_3}$$

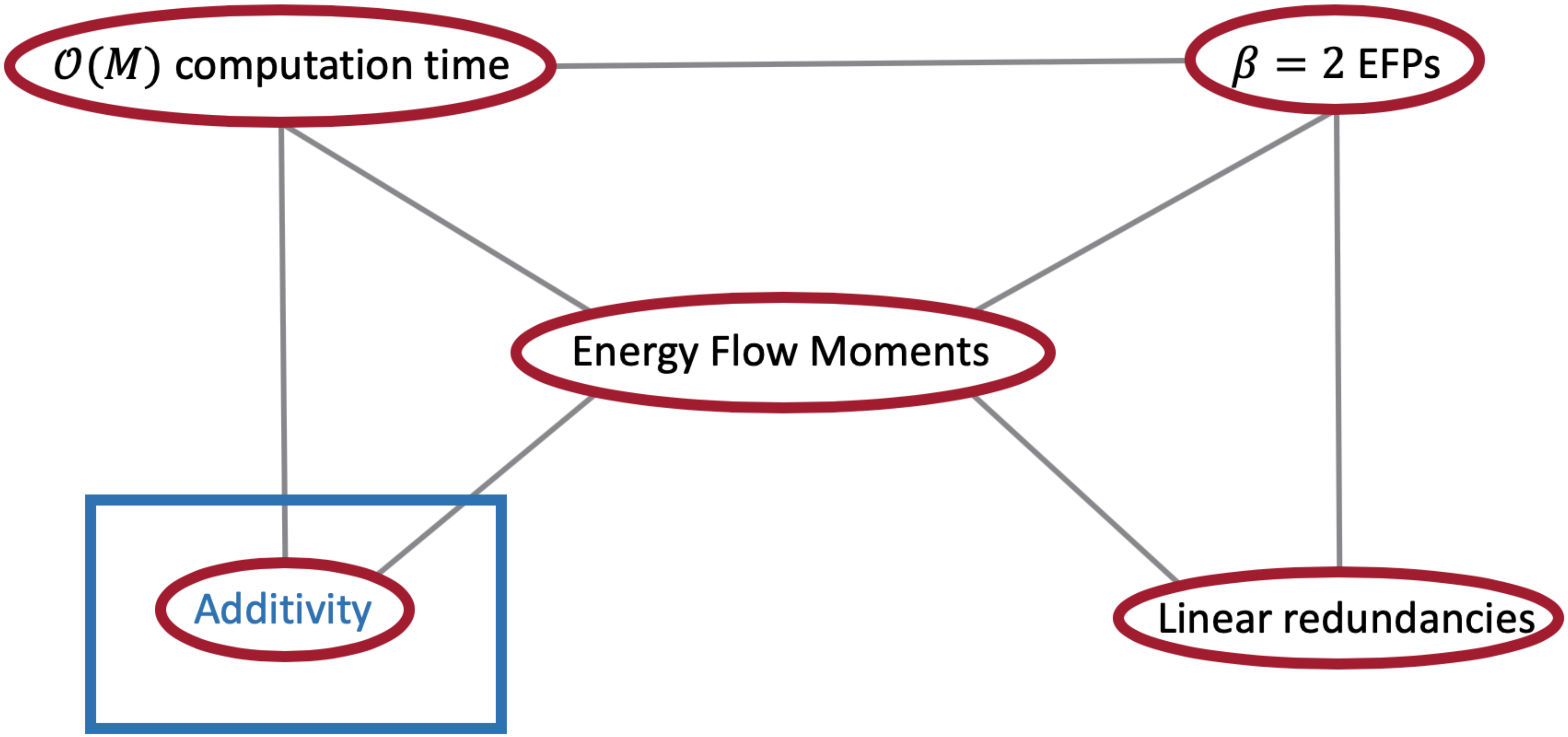
$$= 2^5 \underbrace{\left(\sum_{i_1=1}^{M} z_{i_1}\hat{p}_{i_1}^{\alpha}\hat{p}_{i_1}^{\beta}\hat{p}_{i_1}^{\gamma}\hat{p}_{i_1}^{\delta}\right)}_{\mathcal{I}^{\alpha\beta\gamma\delta}} \underbrace{\left(\sum_{i_2=1}^{M} z_{i_2}\hat{p}_{i_2\alpha}\hat{p}_{i_2\beta}\hat{p}_{i_2}^{\epsilon}\right)}_{\mathcal{I}_{\alpha\beta}{}^{\epsilon}}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

A multigraph correspondence also exists for EFMs:

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

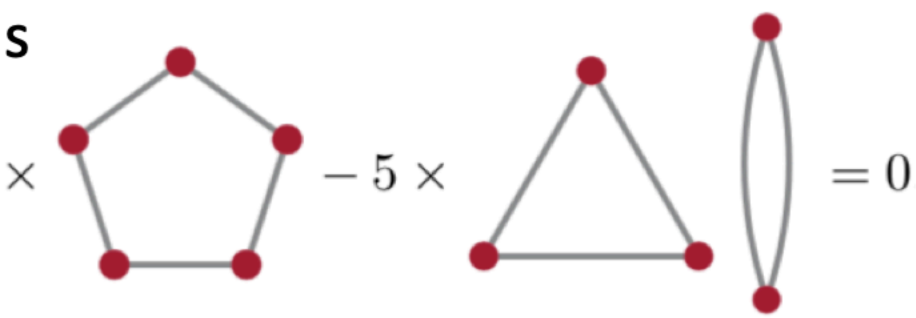
$\Longleftrightarrow \eta_{\mu_i\mu_j}$

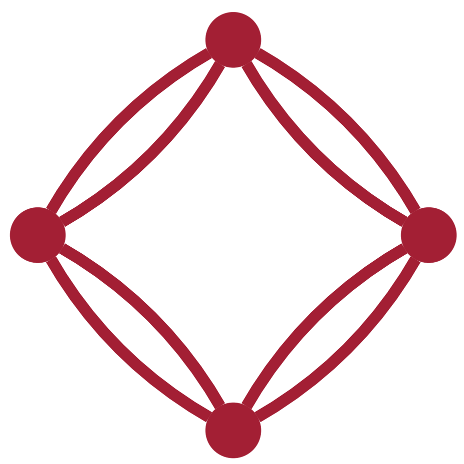
Energy Flow "Network"



Additivity is the link to the Deep Sets decomposition and EFNs

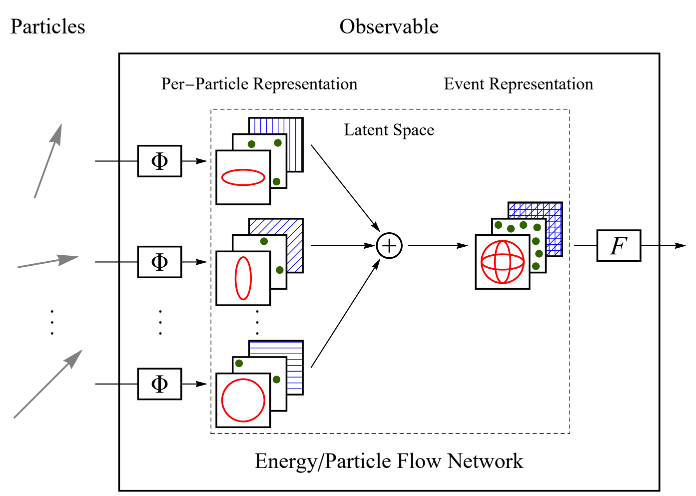
$$5! \times \mathcal{I}^{\mu_2}_{[\mu_1}\mathcal{I}^{\mu_3}_{\mu_2}\mathcal{I}^{\mu_4}_{\mu_3}\mathcal{I}^{\mu_5}_{\mu_4}\mathcal{I}^{\mu_1}_{\mu_5]} = 6 \times$$





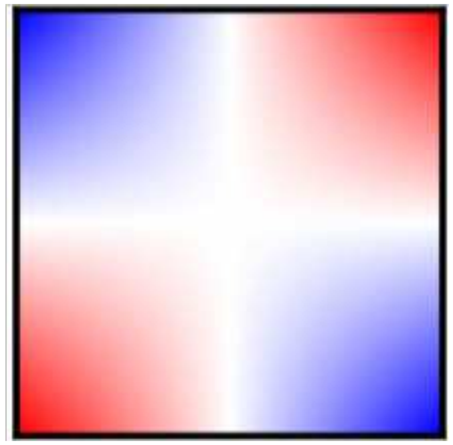
Energy Flow Polynomials

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

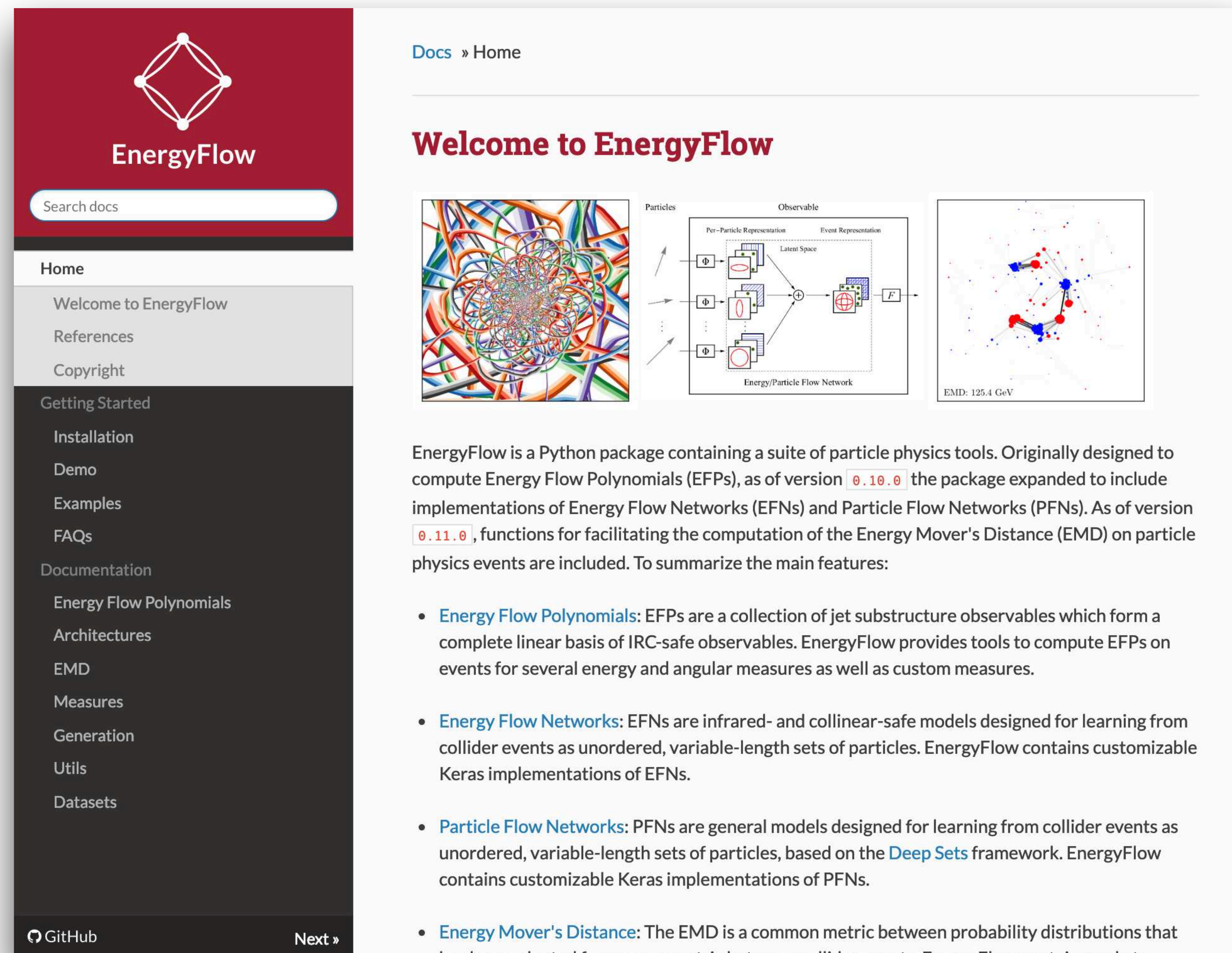
EnergyFlow Python Package

Implements variable elimination for efficient EFP computation

Contains EFN and PFN implementations in Keras

CNN, DNN architectures included for easy model comparison

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

EMD

Measures

Generation

Utils

Datasets

GitHub

Next »

Docs » Home

Welcome to EnergyFlow

EnergyFlow is a Python package containing a suite of particle physics tools. Originally designed to compute Energy Flow Polynomials (EFPs), as of version `0.10.0` the package expanded to include implementations of Energy Flow Networks (EFNs) and Particle Flow Networks (PFNs). As of version `0.11.0`, functions for facilitating the computation of the Energy Mover's Distance (EMD) on particle physics events are included. To summarize the main features:

- 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.
- Energy Mover's Distance: The EMD is a common metric between probability distributions that has been adapted for use as a metric between collider events. EnergyFlow contains code to

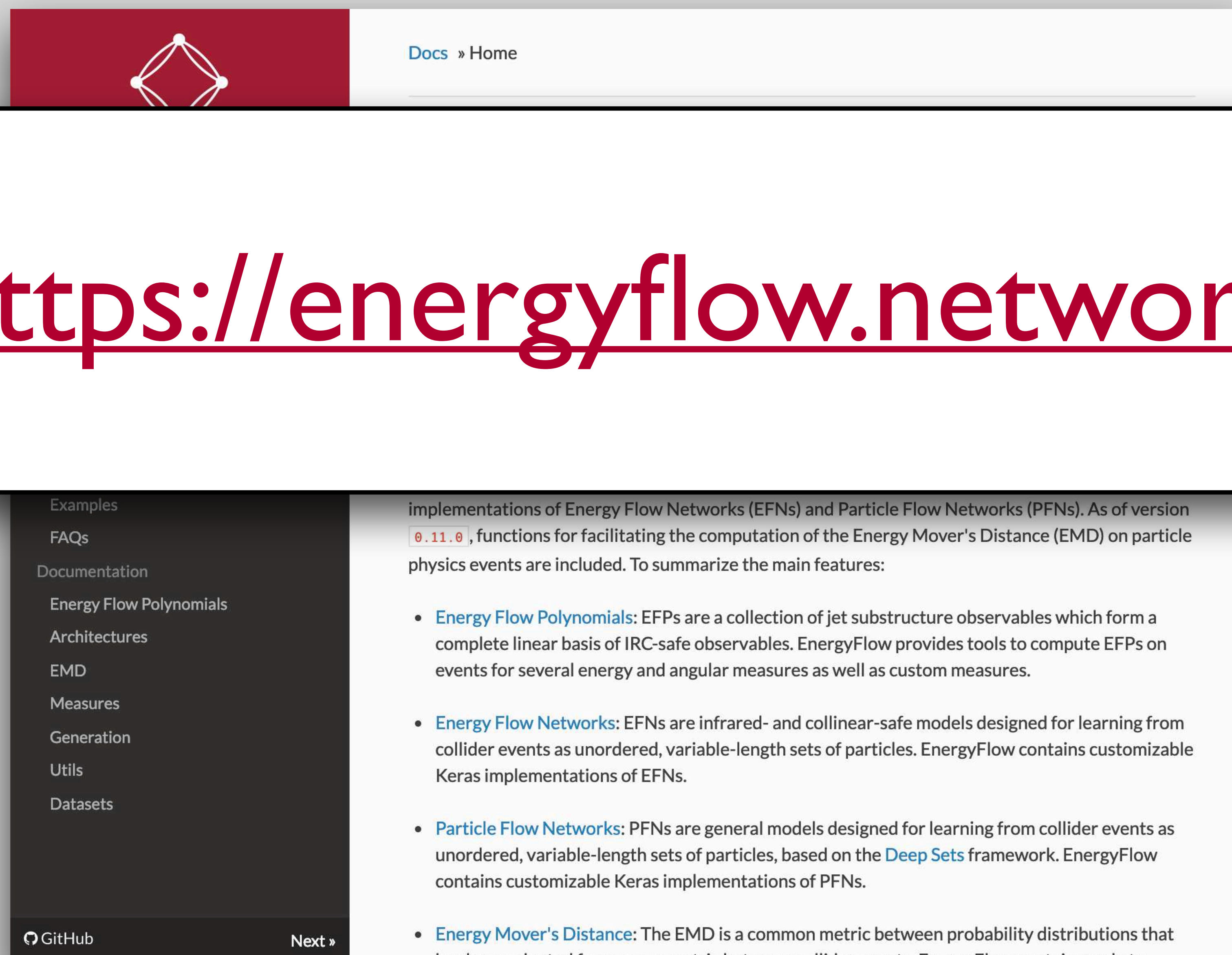
EnergyFlow Python Package

Implements variable elimination for efficient EFP computation

Contains EFN and PFN implementations in Keras

CNN, DNN architectures included for easy model comparison

Several detailed examples demonstrating how to train models and make visualizations



Docs » Home

Examples
FAQs
Documentation
Energy Flow Polynomials
Architectures
EMD
Measures
Generation
Utils
Datasets

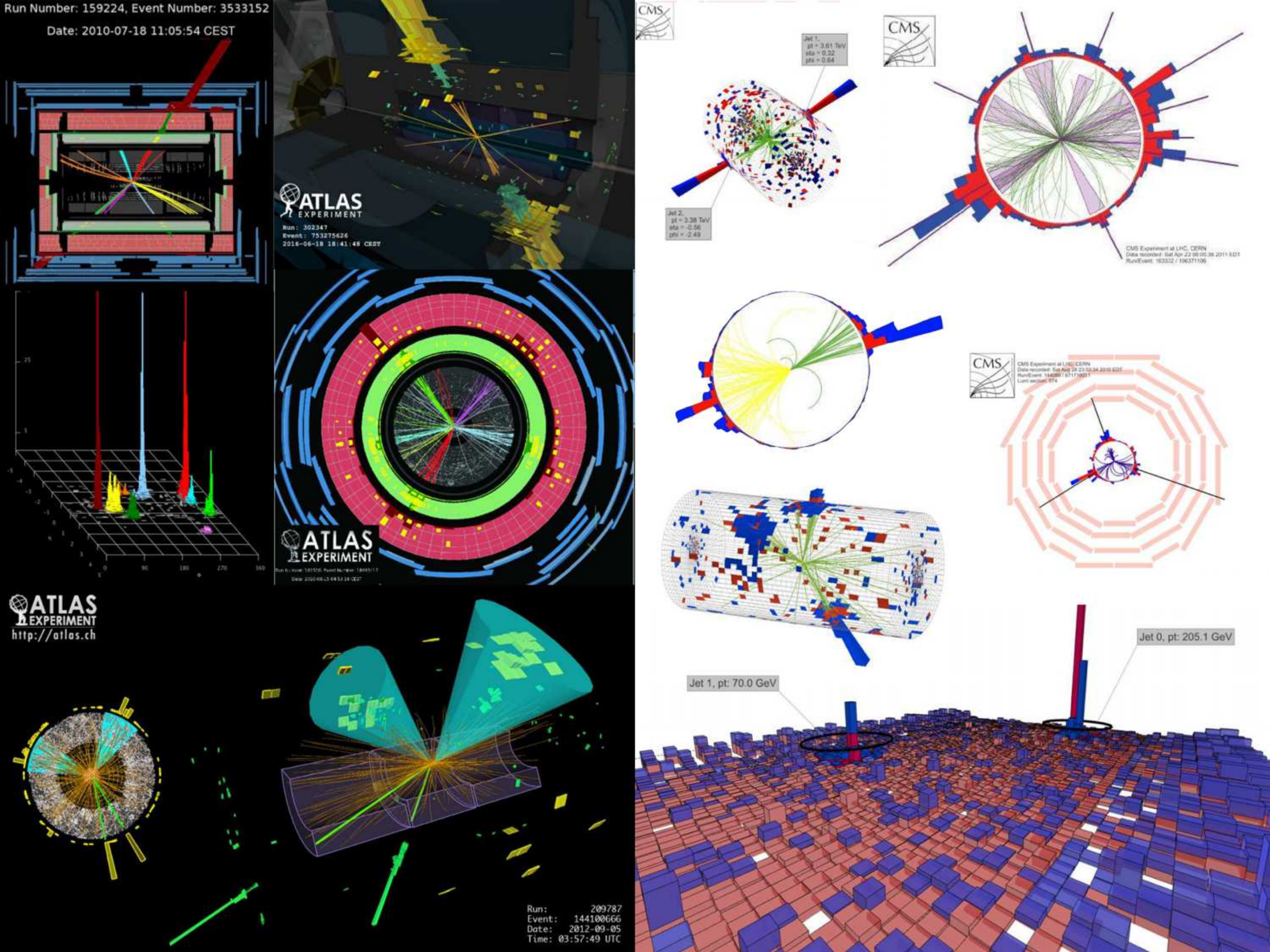
GitHub
Next »

implementations of Energy Flow Networks (EFNs) and Particle Flow Networks (PFNs). As of version `0.11.0`, functions for facilitating the computation of the Energy Mover's Distance (EMD) on particle physics events are included. To summarize the main features:

- 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.
- Energy Mover's Distance: The EMD is a common metric between probability distributions that has been adapted for use as a metric between collider events. EnergyFlow contains code to

https://energyflow.network

Thank You!



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
CMS
ATLAS
EXPERIMENT
CMS
ATLAS
EXPERIMENT
http://atlas.ch
Jet 0, pt: 205.1 GeV
Jet 1, pt: 70.0 GeV
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC

Jet Representations ⟷ Analysis Tools

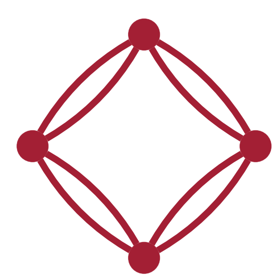
Two key choices when analyzing jets

How to represent the jet

- Single expert observable
- A few expert observables
- Many expert observables
- Jet images
- List of particles
- Clustering tree
- *N*-subjettiness basis
- Energy flow polynomials
- Set of particles

How to analyze that representation

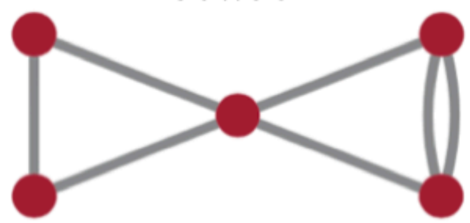
- Threshold cut
- Multidimensional likelihood
- Boosted decision tree (BDT), shallow neural network (NN)
- Convolutional NN (CNN)
- Recurrent/Recursive NN (RNN)
- Fancy RNN
- Dense neural network (DNN)
- Linear classification
- Energy flow network



Multigraph/EFP Correspondence

Multigraph ⟷ **EFP**

"bowtie"



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

⟷ z_{i_j}

⟷ $\theta_{i_k i_l}$

N	Number of vertices	⟷	N-particle correlator
d	Number of edges	⟷	Degree of angular monomial
χ	Treewidth + 1	⟷	Optimal VE Complexity

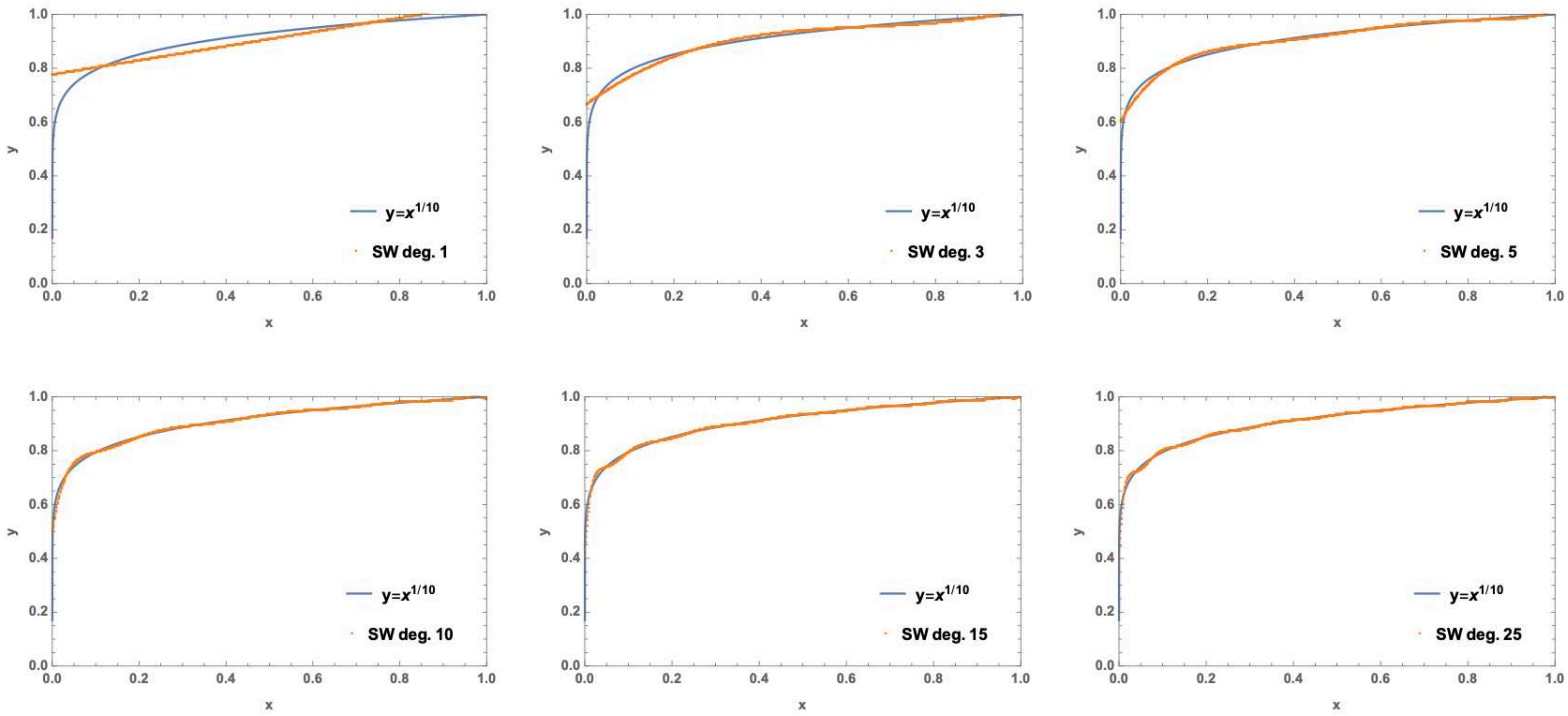
Connected ⟷ Prime

Disconnected ⟷ Composite

⋮

Fun with the Stone-Weierstrass Theorem

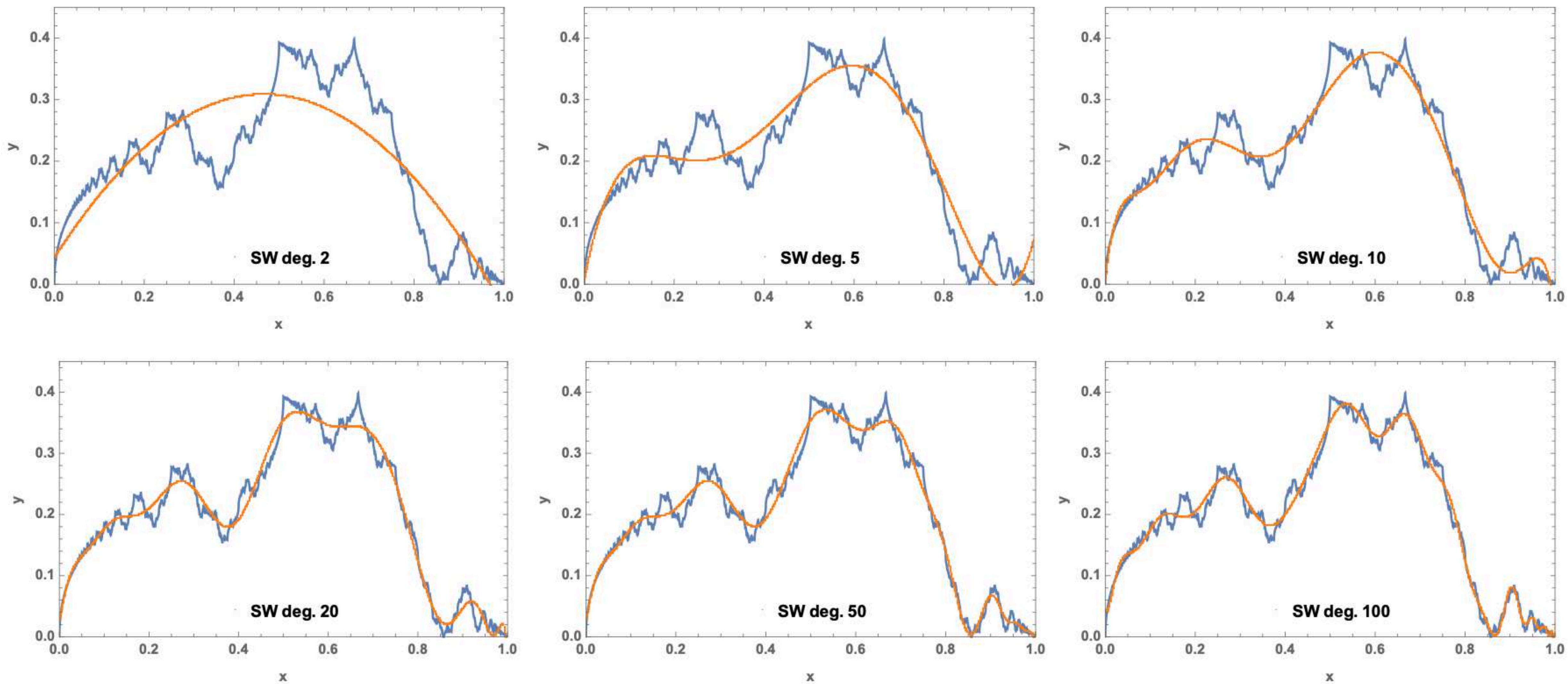
Try to approximate function that has no Taylor expansion around zero



Fun with the Stone-Weierstrass Theorem

$$y = \sum_{k=1}^{\infty} \frac{\sin(\pi k^2 x)}{\pi k^2}$$

That was too easy, try the Weierstrass function (continuous everywhere, differentiable on a measure zero set of points)



Computation Complexity of EFPs – Variable Elimination

Naive computation complexity of an energy correlator is $\mathcal{O}(M^N)$

For ~100 particles this becomes intractable for $N > 4$

Computation Complexity of EFPs – Variable Elimination

Naive computation complexity of an energy correlator is $\mathcal{O}(M^N)$

For ~100 particles this becomes intractable for *N* > 4

⇨ EnergyCorrelator fjcontrib package gives up in this case

```
// if N > 5, then throw error
if (_N > 5) {
    throw Error("EnergyCorrelator is only hard coded for N = 0,1,2,3,4,5");
}
```

Computation Complexity of EFPs – Variable Elimination

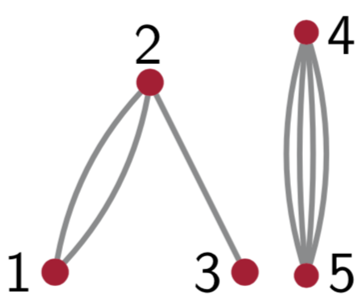
Naive computation complexity of an energy correlator is $\mathcal{O}(M^N)$

For ~100 particles this becomes intractable for *N* > 4

⇨ EnergyCorrelator `fjcontrib` package gives up in this case

```
// if N > 5, then throw error
if (_N > 5) {
    throw Error("EnergyCorrelator is only hard coded for N = 0,1,2,3,4,5");
}
```

Variable elimination (VE) algorithm can speedup EFPs by finding efficient elimination ordering



$$= \left(\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_2 i_3}\right)\left(\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_4} z_{i_5} \theta_{i_4 i_5}^4\right)$$

Disconnected is product of connected

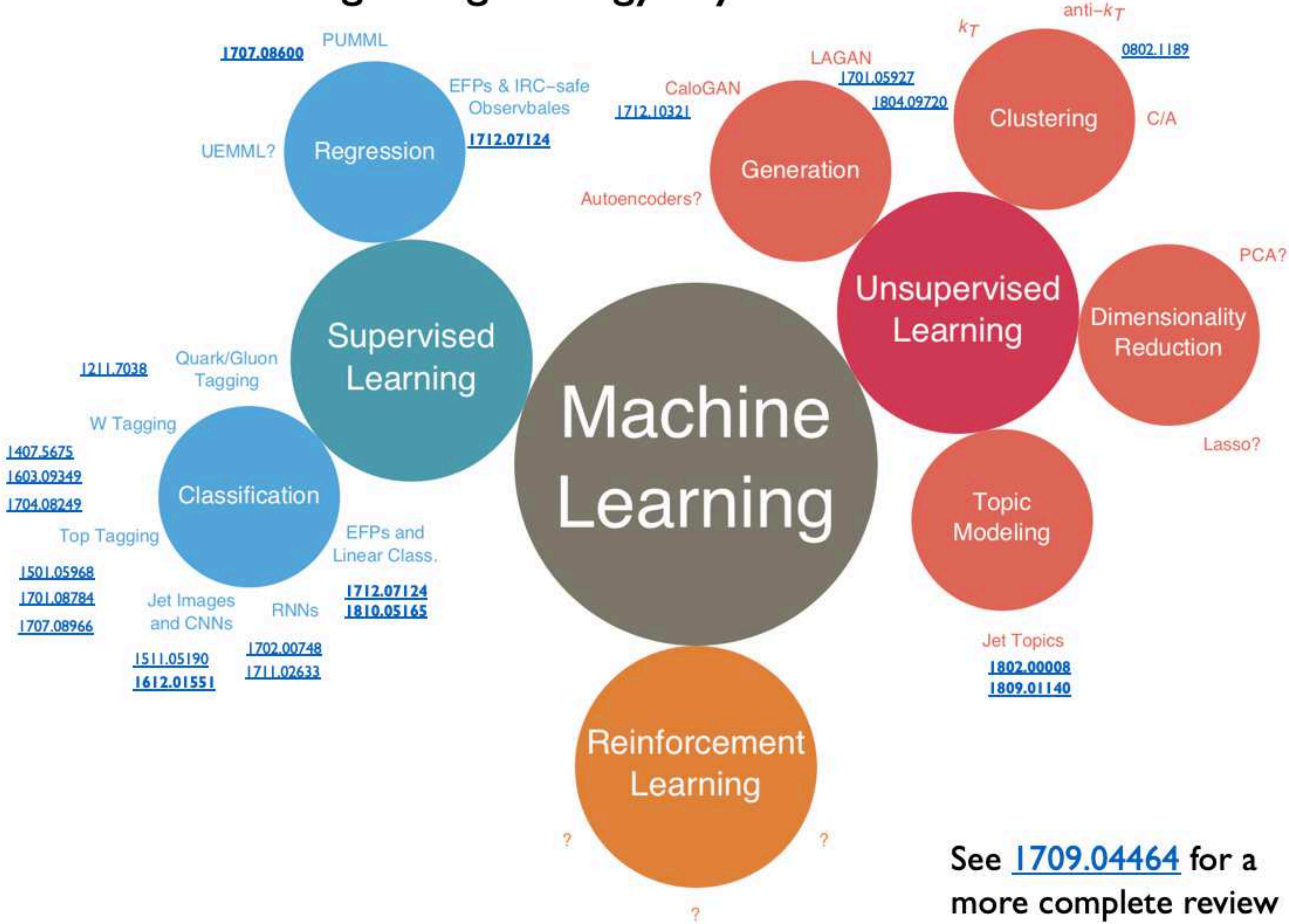
$$= \underbrace{\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M}\sum_{i_6=1}^{M}\sum_{i_7=1}^{M}\sum_{i_8=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} z_{i_6} z_{i_7} z_{i_8} \prod_{j=2}^{7} \theta_{i_1 i_j}}_{\mathcal{O}(M^8)}$$

$$= \underbrace{\sum_{i_1=1}^{M} z_{i_1} \left(\sum_{i_2=1}^{M} z_{i_2} \theta_{i_1 i_2}\right)^7}_{\mathcal{O}(M^2)}$$

Clever parentheses placement corresponds to good elimination ordering

All tree graphs become $\mathcal{O}(M^2)$

Machine Learning in High Energy Physics



See 1709.04464 for a more complete review

Jet Classification Studies

Mix and match

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

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 (and not updating this after July 23, 2018); plus many ATLAS/CMS performance studies]

Top Jet Samples and Other Methods

[Butter, Kasieczka, Plehn, Russell, 2017]

Common top and QCD dijet samples for standardized benchmarking

$p_T \in [550, 650]$ GeV, AK8 jets, fully-merged, Delphes simulation, 2m jets total

Approach	AUC	Acc.	1/eB @ (eS=0.3)	Contact	Comments
LoLa	0.979	0.928		G. Kasieczka S. Leiss	Preliminary number, based on LoLa
LBN	0.981	0.931	863	M. Rieger	Preliminary number
CNN	0.981	0.93	780	D. Shih	Model from *(1803.00107)*
P-CNN (1D CNN)	0.980	0.930	782	H. Qu, L. Gouskos	Preliminary, use kinematic info only
6-body N-subs. (+mass and pT) NN	0.979	0.922	856	K. Nordstrom	Based on 1807.04769
8-body N-subs. (+mass and pT) NN	0.980	0.928	795	K. Nordstrom	Based on 1807.04769
Linear EFPs	0.980	0.932	380	PTK, E. Metodiev	d<= 7, chi <= 3 EFPs with FLD. Based on 1712.07124
Particle Flow Network (PFN)	0.982	0.932	888	PTK, E. Metodiev	Median over ten trainings. Based on Table 5 in 1810.05165
Energy Flow Network (EFN)	0.979	0.927	619	PTK, E. Metodiev	Median over ten trainings. Based on Table 5 in 1810.05165

Performance saturation?

Top Jet Samples and Other Methods

ehn, Russell, 2017]

Comm

$p_T \in [5$

Not yet!

Particle Flow Network (PFN)	0.982	0.932	888	PTK, E. Metodiev	Median over ten trainings. Based on Table 5 in 1810.05165
2D CNN [ResNeXt50]	0.984	0.936	1086	Huilin Qu, Loukas Gouskos	Preliminary from indico.cern.ch/event/745718/contributions/3202526
DGCNN	0.984	0.937	1160	Huilin Qu, Loukas Gouskos	Preliminary from indico.cern.ch/event/745718/contributions/3202526

However, ResNeXt50 has 25m parameters and DGCNN takes 2 days to train

PFN has 40k parameters and takes 1 hour to train