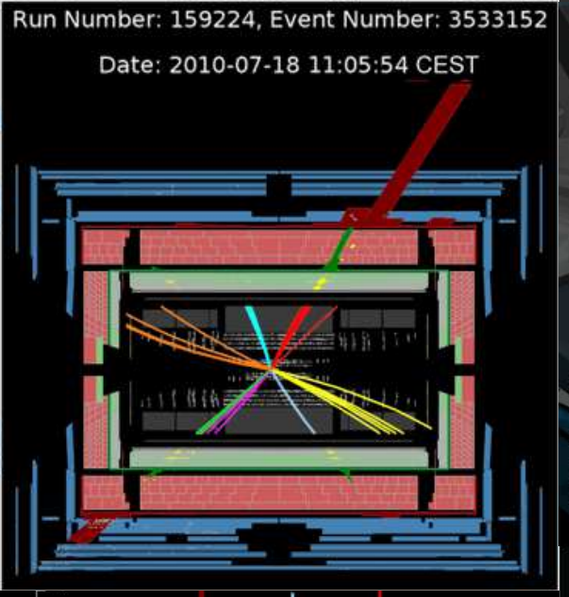
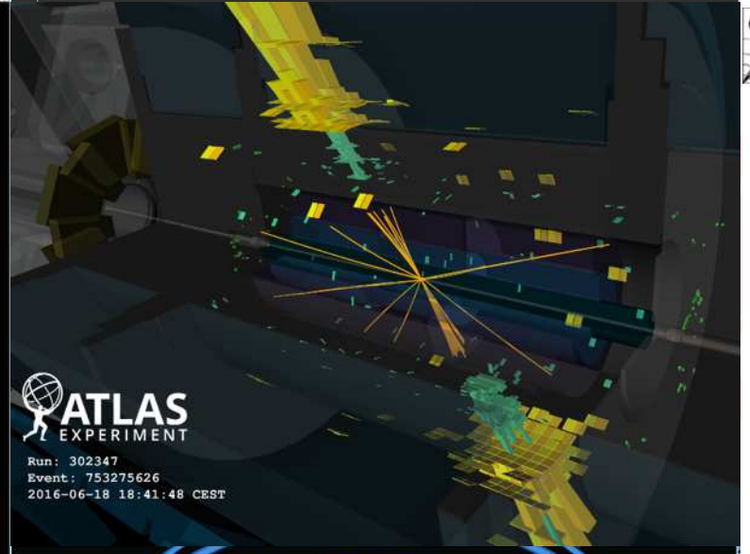
The Energy Flow Basis

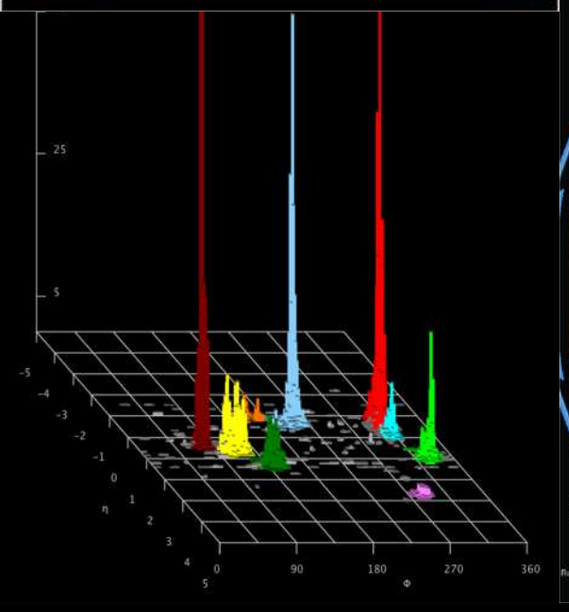
Jesse Thaler

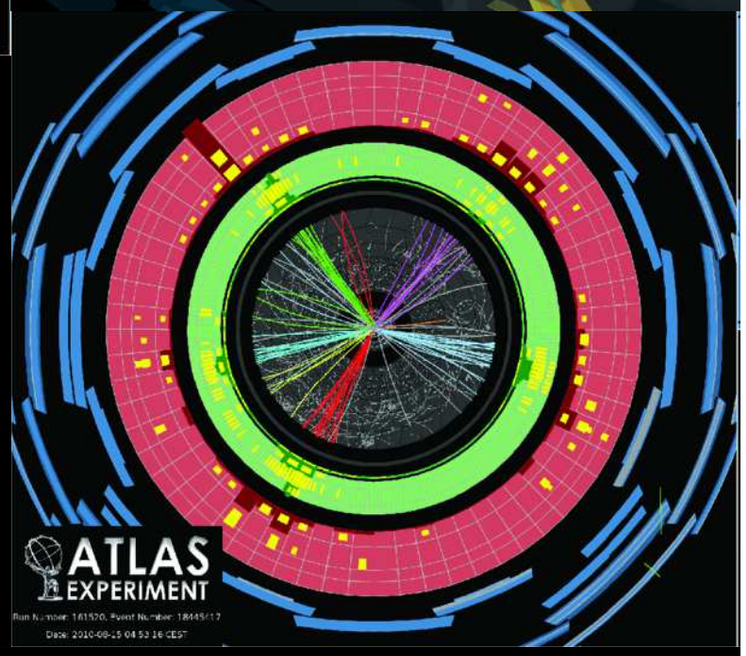
with Patrick Komiske & Eric Metodiev (1712.07124, …)

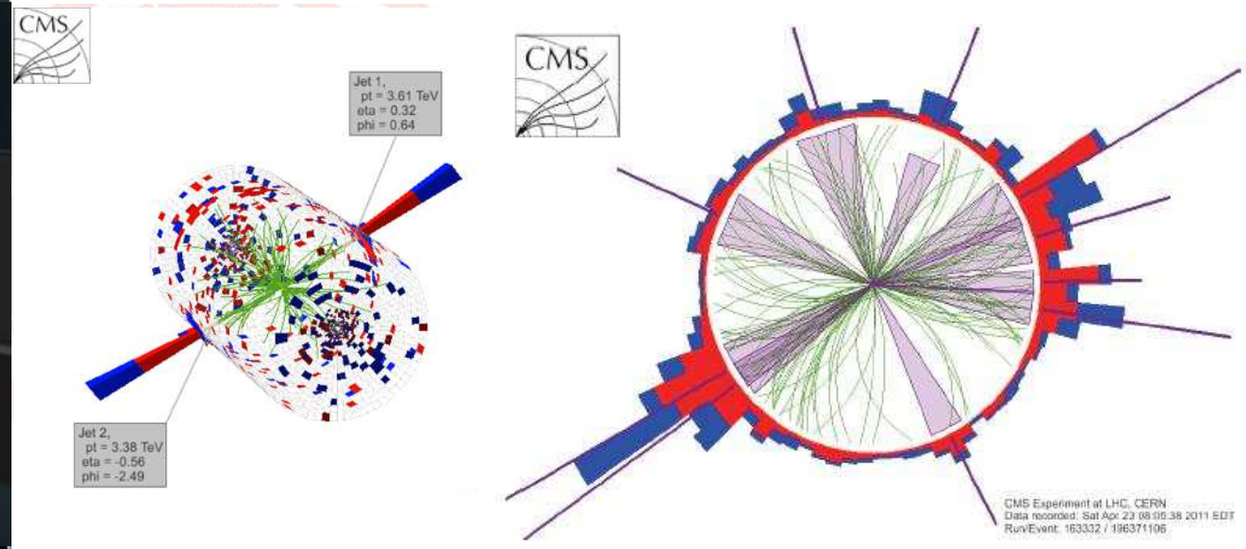
Reconstruction, Trigger, and Machine Learning for the HL-LHC, MIT — April 27, 2018

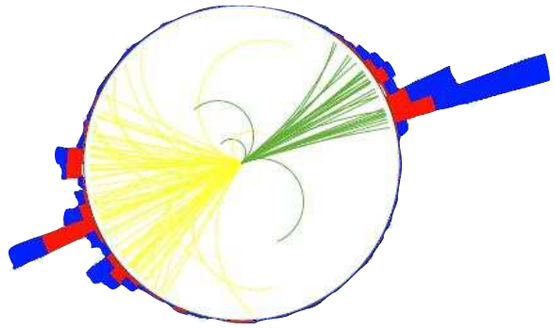

Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST

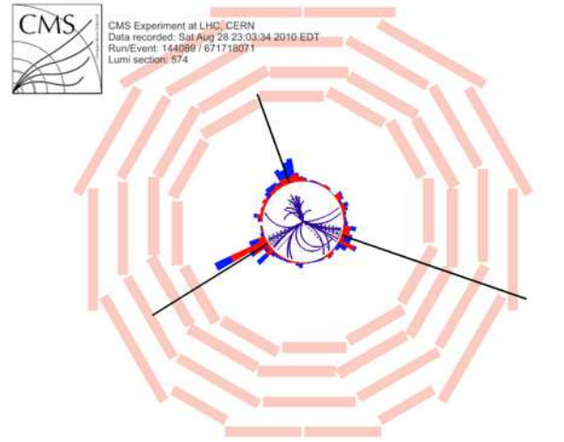

ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

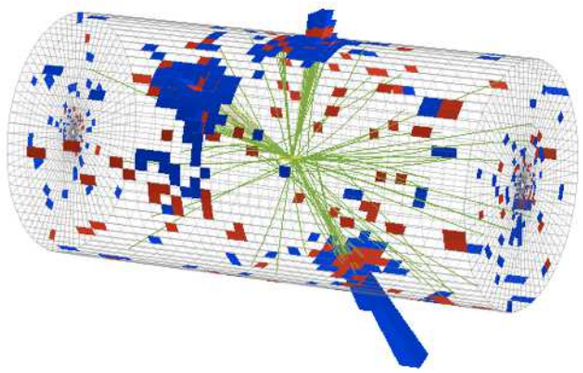

25
5
-5
-4
-3
-2
-1
0
1
2
3
4
5
η
0
90
180
270
360
φ

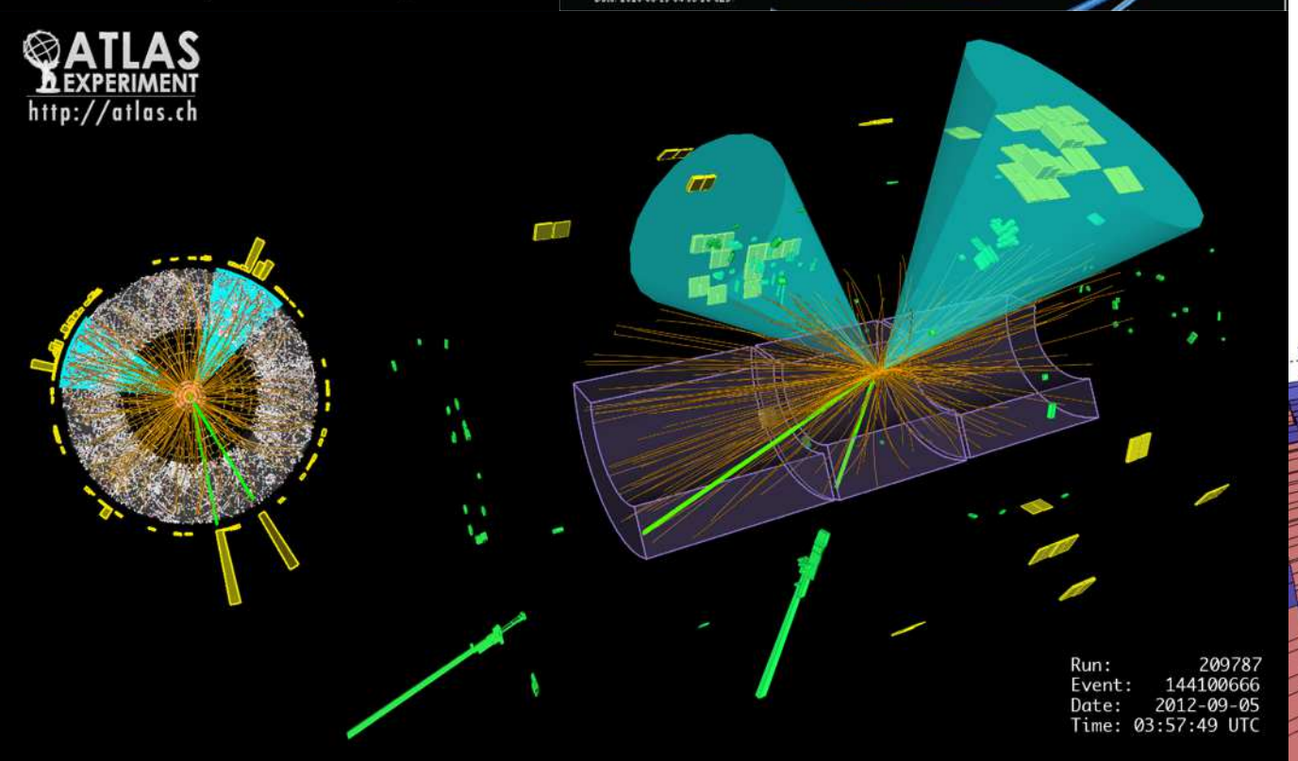

ATLAS
EXPERIMENT

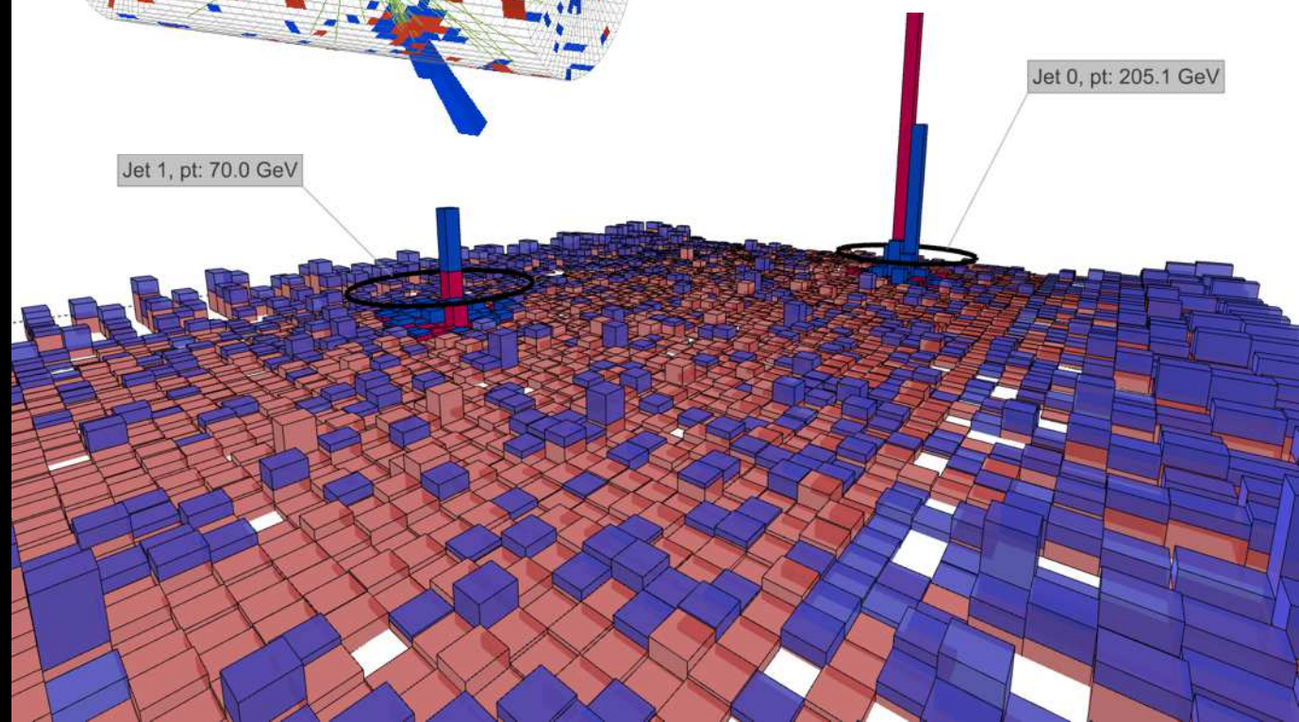

CMS
Jet 1,
pt = 3.61 TeV
eta = 0.32
phi = 0.64
Jet 2,
pt = 3.38 TeV
eta = -0.56
phi = -2.49
CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 196371106



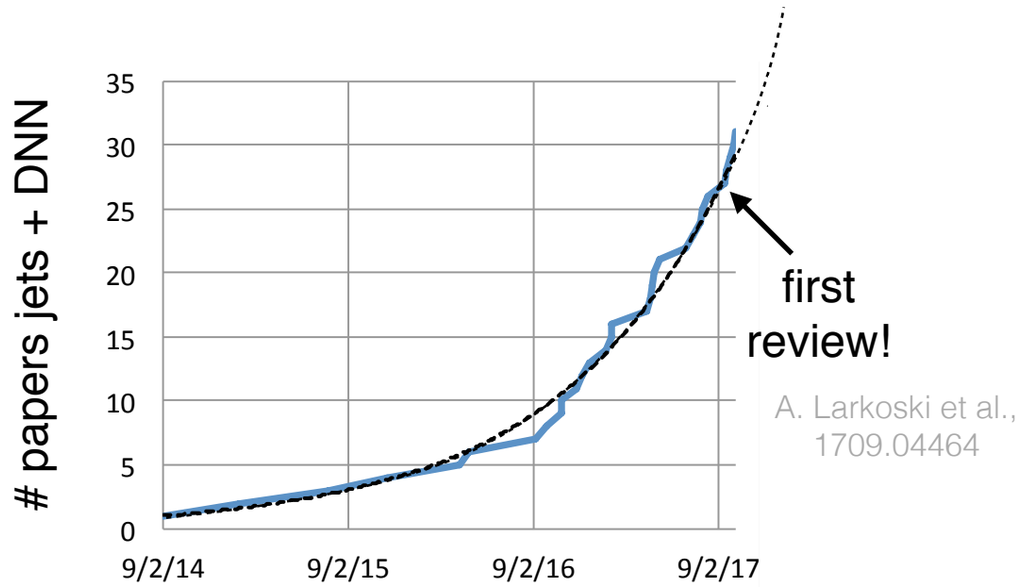

CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Aug 28 23:03:34 2010 EDT
Run/Event: 144089 / 671718071
Lumi section: 574




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

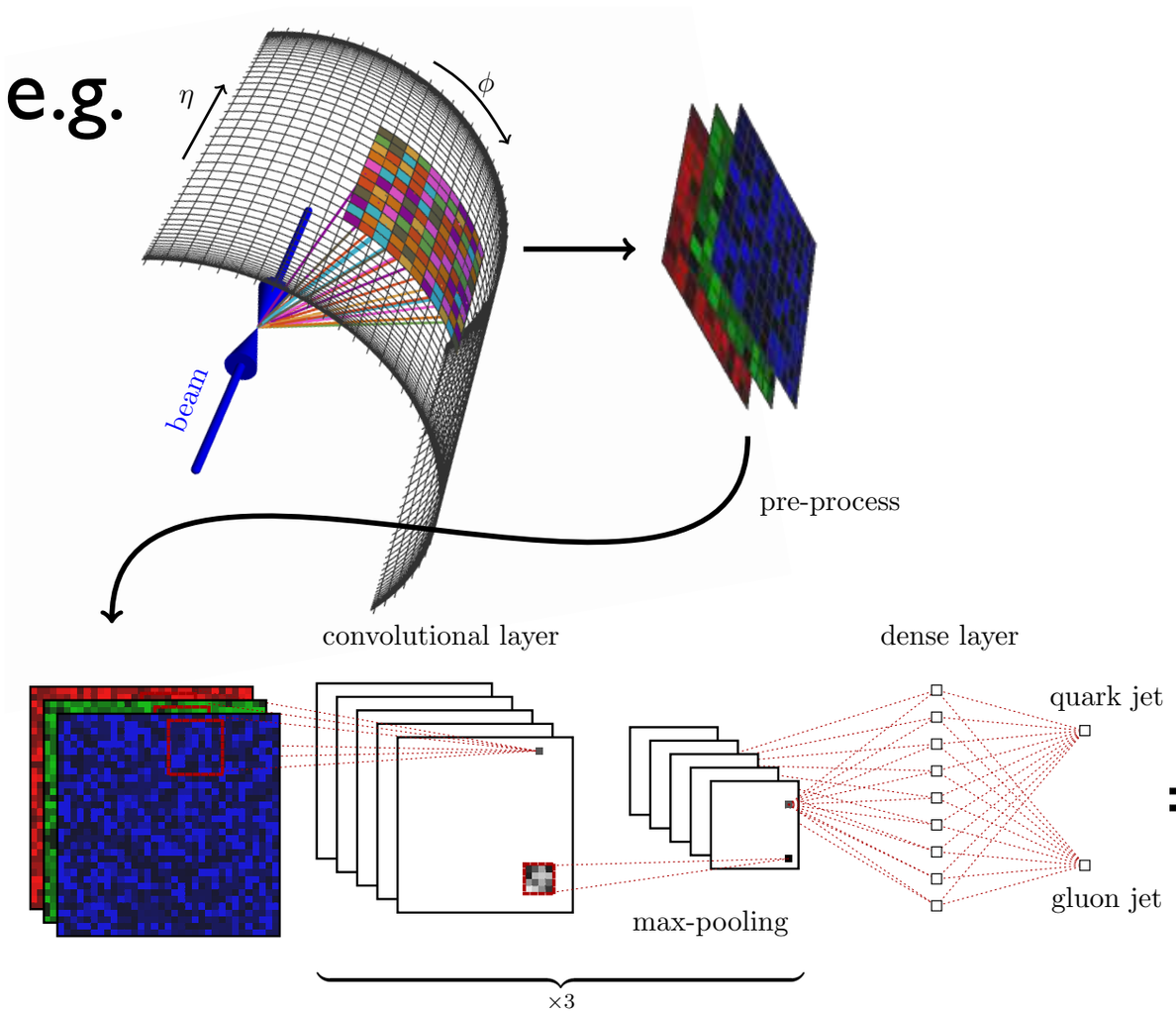

Jet 0, pt: 205.1 GeV
Jet 1, pt: 70.0 GeV

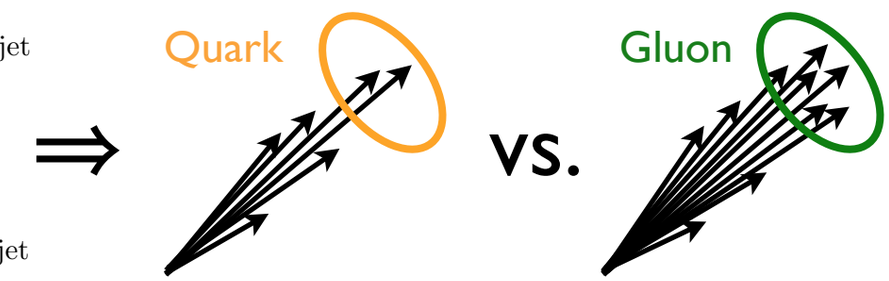
The Rise of Machine Learning for Jets



A. Larkoski et al., 1709.04464

e.g.





[e.g. Komiske, Metodiev, Schwartz, 2016; Nachman, Machine Learning for Jets Workshop, 2017]

My Perspective c. 2016

"Deep Learning" vs. "Deep Thinking"

My Perspective c. 2018

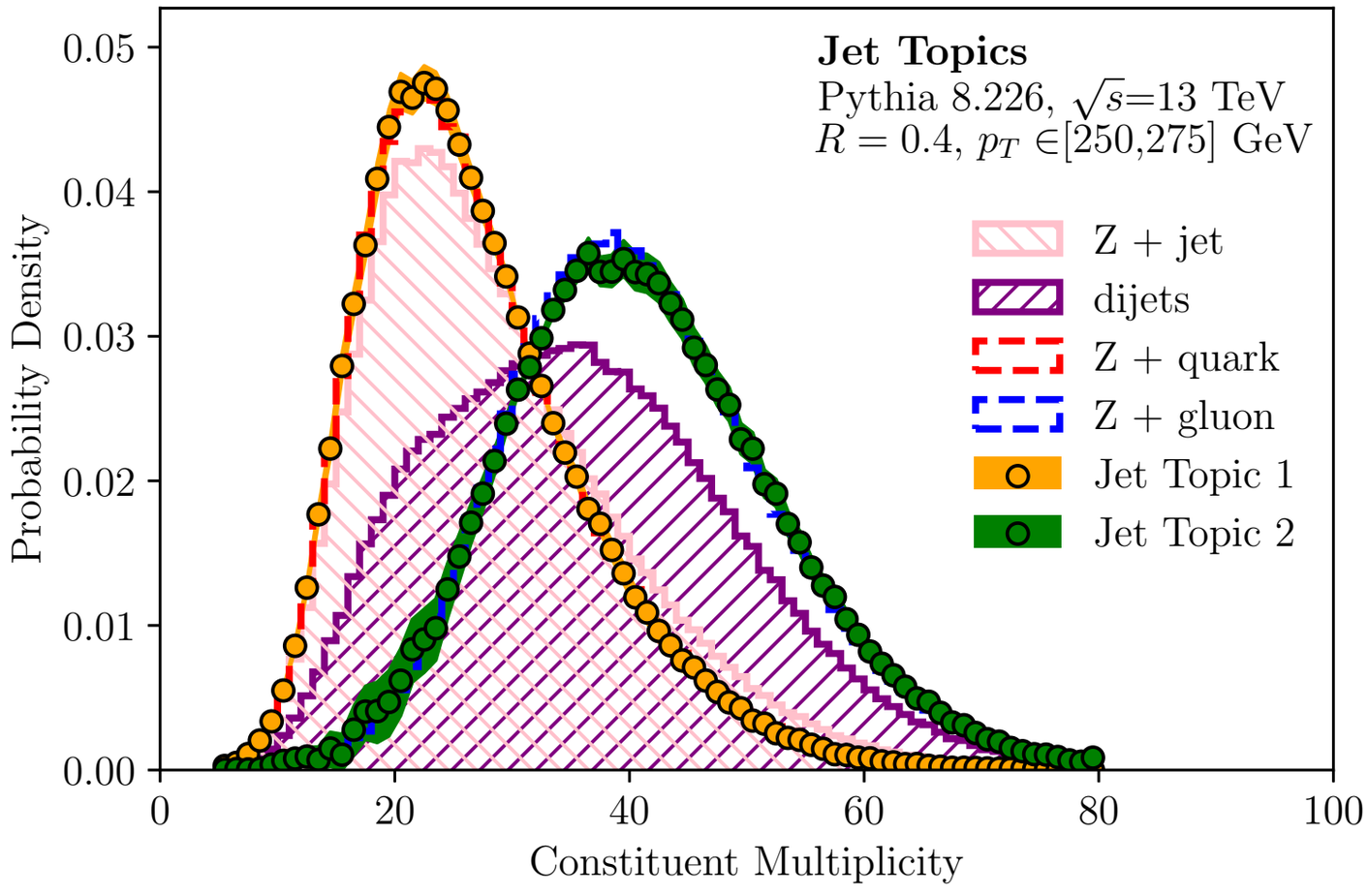
"Deep Learning" & "Deep Thinking"

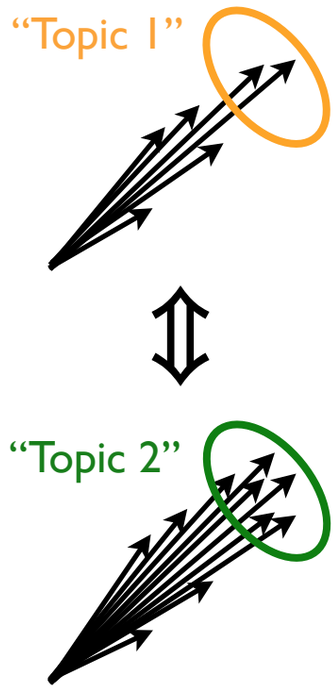
~~vs.~~

New first-principles studies of QCD
facilitated by advances in
statistics, mathematics, and computer science

E.g. Jet Topics for Quark/Gluon Jets

Extract jet categories from data…





…solely* from the assumption they exist

Sample Independence, Different Fractions, Mutual Irreducibility

[Metodiev, JDT, 1802.00008]

Extrapolate Lessons to HL-LHC?

The goal of this workshop is to develop cross-experiment collaborations to work on common problems that will be faced by HL-LHC experiments. Beyond this, we hope to build collaborations between the HEP and the CS communities. **The irregular 3-D geometries of HEP detectors, their heterogeneity, and the extremely small latencies provide unique and interesting data science challenges.** Conversely, we have excellent models of how particles behave in our detectors, and excellent simulations of these detectors, which makes obtaining vast training samples possible. By building HEP-CS collaborations, we hope that HEP data sets can be used for cutting-edge ML research, providing benefits to both communities.

Extrapolate Lessons to HL-LHC?

The goal of this workshop is to develop cross-experiment collaborations to work on common problems that will be faced by HL-LHC experiments. Beyond this, we hope to build collaborations between the HEP and the CS communities. **The irregular 3-D geometries of HEP detectors, their heterogeneity, and the extremely small latencies provide unique and interesting data science challenges.** Conversely, we have excellent models of how particles behave in our detectors, and excellent simulations of these detectors, which makes obtaining vast training samples possible. By building HEP-CS collaborations, we hope that HEP data sets can be used for cutting-edge ML research, providing benefits to both communities.

Jets in Theory-land:

Regular geometry, homogeneous inputs, offline analysis

The Theory-Land Approach

Underlying Physics

"Deep Thinking"

Natural Data Representation

"Deep Learning"

Suitable Algorithm

The Buzzword Approach

Questionable Physics

⇑ "Wishful Thinking"

Unnatural Data Representation

⇑ "Shoehorning"

Cool-Sounding Algorithm

Logic for the Energy Flow Basis

Underlying Physics

Natural Data Representation

Suitable Algorithm

Infrared and Collinear Safety

Energy Flow Polynomials

Linear Regression

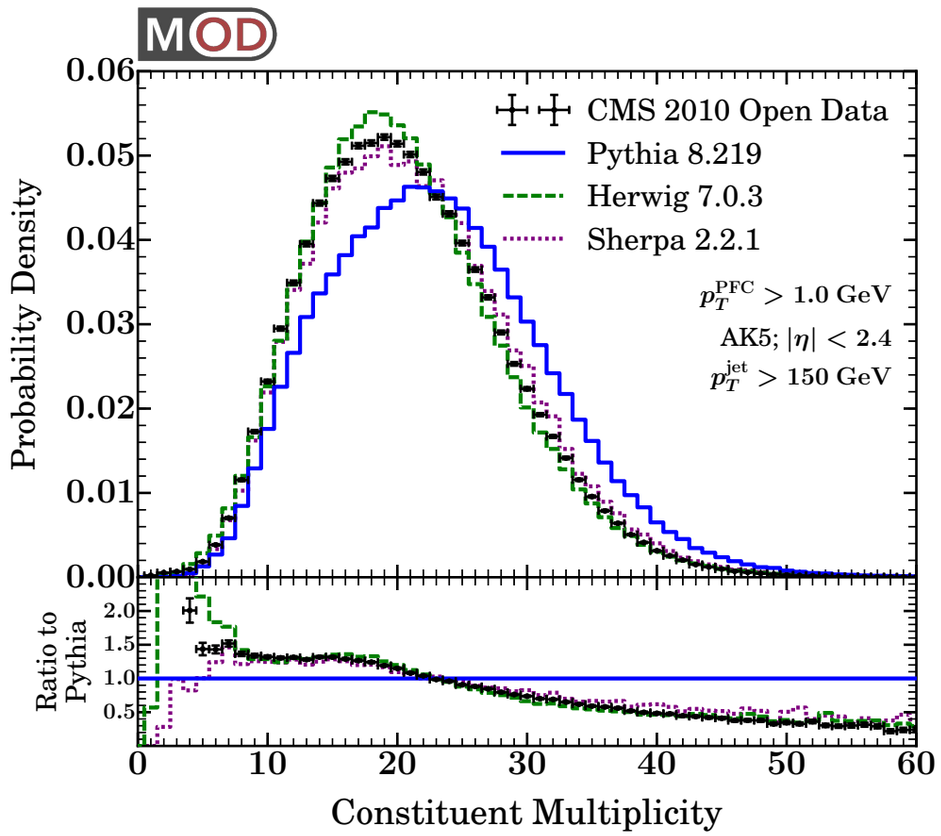
[Komiske, Metodiev, JDT, 1712.07124]

Key Fact #1

→ Underlying Physics

Natural Data Representation

Suitable Algorithm



Jet constituents:

Particle-like objects

Variable-length

Unordered set

Per particle:

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

Flavor/charge labels

Vertex information

Quality criteria, etc.

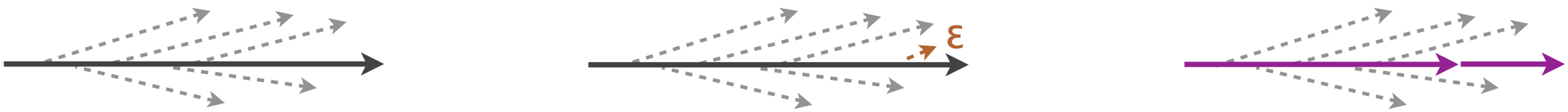
[plot from Tripathee, Xue, Larkoski, Marzani, JDT, 1704.05842]

Key Fact #2

→ Underlying Physics
Natural Data Representation
Suitable Algorithm

Wide range of interesting observables are "safe"

Original Jet | Infrared | Collinear



IRC Safe Observable: Insensitive to IR or C emissions

Enforces smooth interpolation between variable-length inputs (i.e. N → N–1)

Examples from Jet Substructure

→ Underlying Physics
Natural Data Representation
Suitable Algorithm

Jet p_T: $\sum_{i \in \text{jet}} p_{T,i}$ IRC Safe

p_T^D: [CMS HIG-11-027] $\sum_{i \in \text{jet}} \frac{p_{T,i}^2}{p_{T\text{jet}}^2}$ IR Safe, C Unsafe

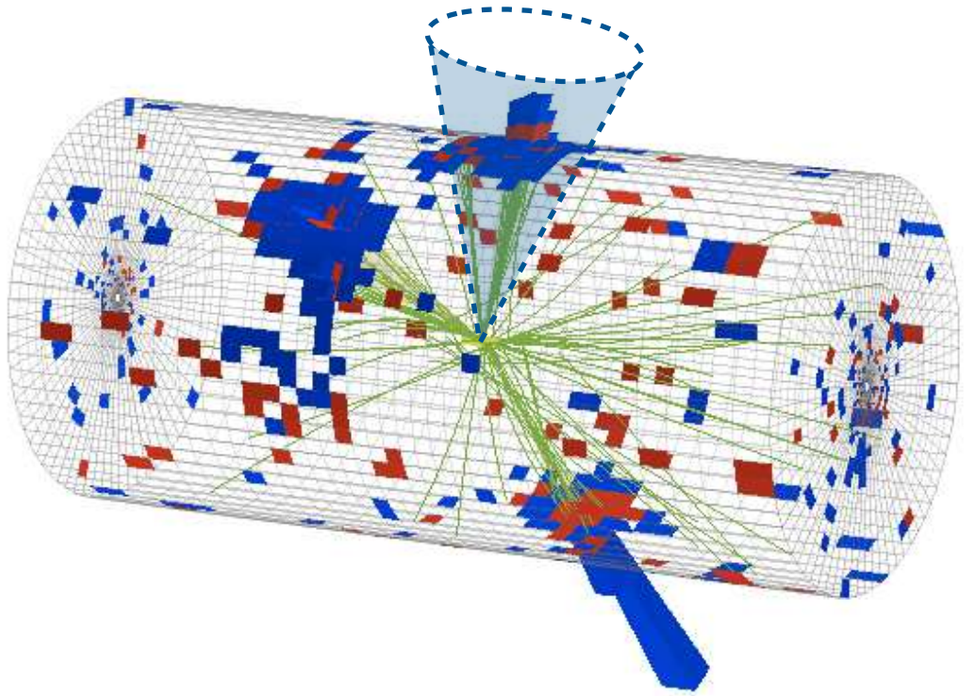
Multiplicity: $\sum_{i \in \text{jet}} 1$ IRC Unsafe

Jet Mass: $\sum_{i,j \in \text{jet}} p_i \cdot p_j$ IRC Safe

N-subjettiness: $\sum_{i \in \text{jet}} p_{T,i} \min\left\{\Delta R_{i,1}, \Delta R_{i,2}, \ldots, \Delta R_{i,N}\right\}^{\beta}$ IRC Safe

[JDT, Van Tilburg, 1011.2268, 1108.2701]

With IRC safe jet algorithm (e.g. anti-k_T)



→ Underlying Physics

Natural Data Representation

Suitable Algorithm

What is the space of *all* IRC-safe observables?

A Systematic Expansion

Underlying Physics
→ Natural Data Representation
Suitable Algorithm

Expand* any IRC safe observable in small energy limit

$$\mathcal{S} = \sum_i E_i \, f_1^{\mathcal{S}}(\hat{n}_i) + \sum_{ij} E_i E_j \, f_2^{\mathcal{S}}(\hat{n}_i, \hat{n}_j) + \sum_{ijk} E_i E_j E_k \, f_3^{\mathcal{S}}(\hat{n}_i, \hat{n}_j, \hat{n}_k) + \dots$$

Form enforced by: Particle Relabeling; Infrared Safety; Collinear Safety

Further expand* each angular function in pairwise angles

$$z_i = \frac{E_i}{E_{\text{jet}}} \qquad \cos\theta_{ij} = \hat{n}_i \cdot \hat{n}_j$$

[Komiske, Metodiev, JDT, 1712.07124; see also Tkachov, hep-ph/9601308]

The Energy Flow Polynomials

Underlying Physics
→ Natural Data Representation
Suitable Algorithm

$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}^{\beta}$$

Multigraph

Angular Scaling

All N-tuples

N Energy Fractions

Polynomial in Pairwise Angles

A *Linear* Basis for Jet Substructure (!)

[Komiske, Metodiev, JDT, 1712.07124]

The Energy Flow Polynomials

Underlying Physics
→ Natural Data Representation
Suitable Algorithm

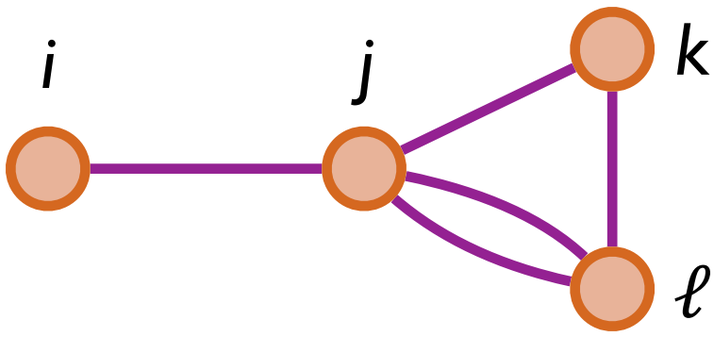
$$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}^{\beta}$$

Multigraph — All N-tuples

N Energy Fractions

Angular Scaling — Polynomial in Pairwise Angles

e.g.



$$= \sum_{ijk\ell} z_i z_j z_k z_\ell \, \theta_{ij}\theta_{jk}\theta_{j\ell}^2\theta_{k\ell}$$

A *Linear* Basis for Jet Substructure (!)

[Komiske, Metodiev, JDT, 1712.07124]

Down the Rabbit Hole

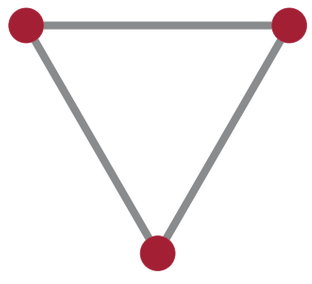
Underlying Physics
→ Natural Data Representation
Suitable Algorithm

Known Structures:

$1e_2$

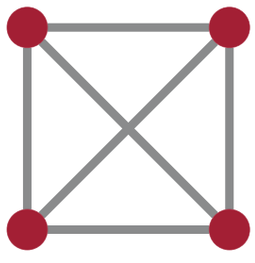
width (β=1), thrust (β=2)

$3e_3$



used in D_2

$6e_4$



used in C_3

Down the Rabbit Hole

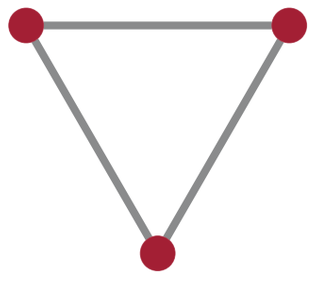
Underlying Physics
→ Natural Data Representation
Suitable Algorithm

Known Structures:

$1e_2$

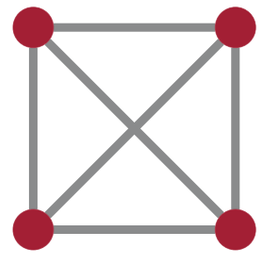
width (β=1), thrust (β=2)

$3e_3$



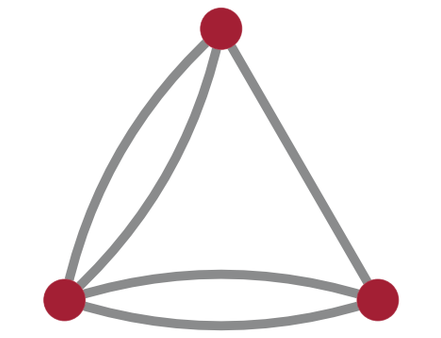
used in D_2

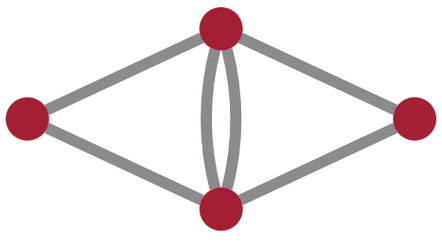
$6e_4$

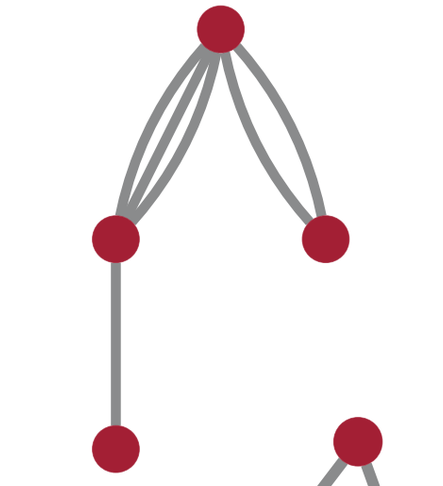


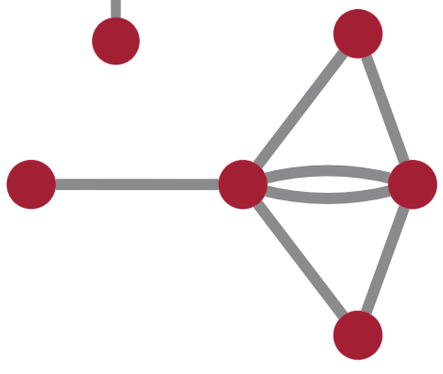
used in C_3

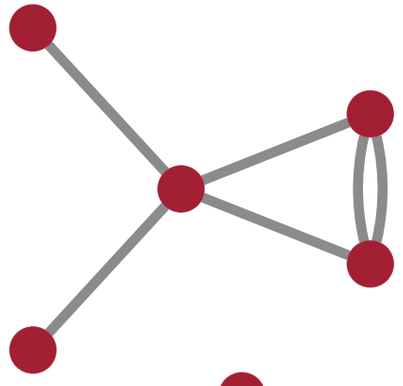
No Idea:

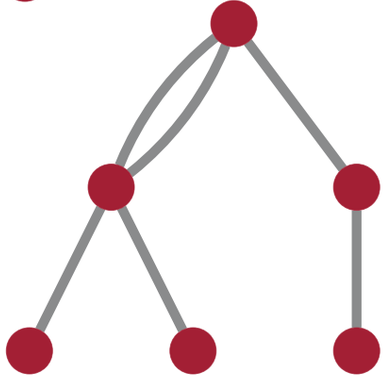












Linear Regression vs. CNN

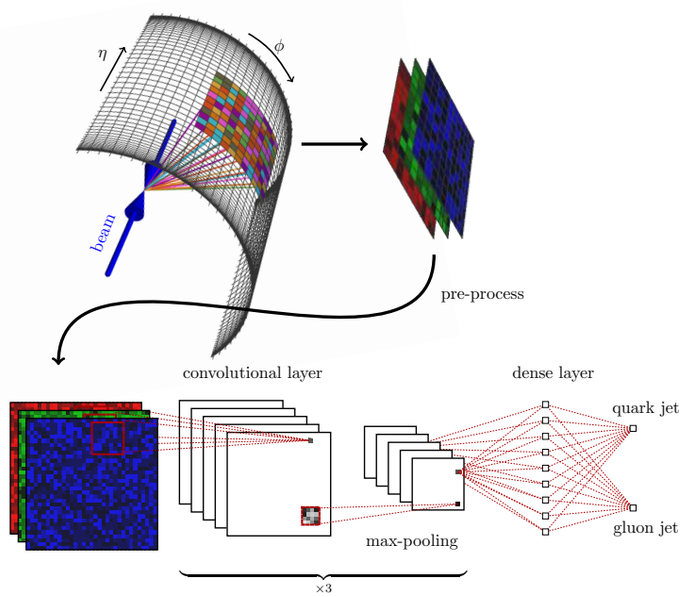
Underlying Physics

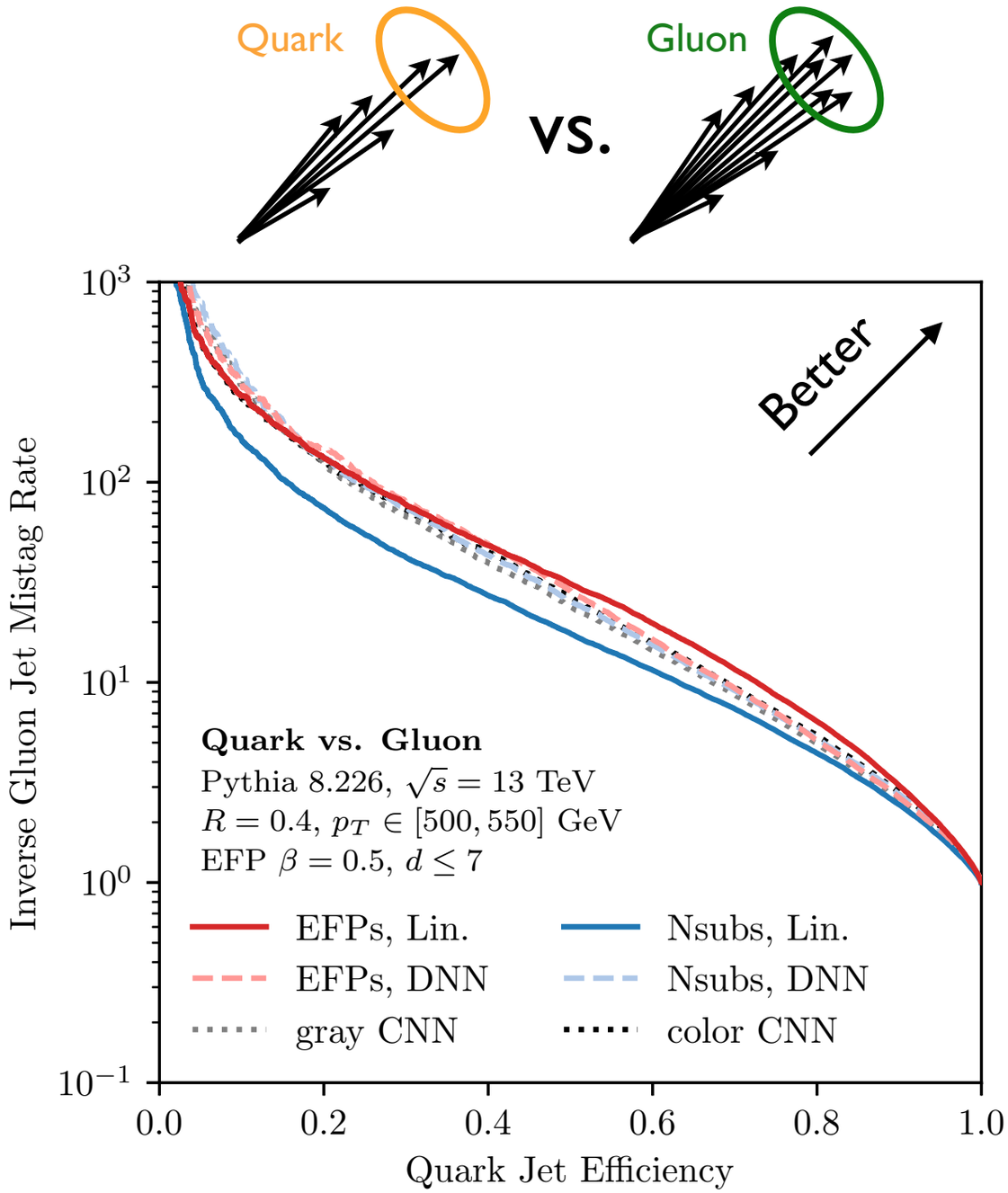
Natural Data Representation

→ Suitable Algorithm

$$\mathcal{S} = \sum_G s_G \, \mathrm{EFP}_G$$

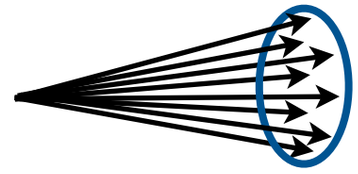
$\approx$





[Komiske, Metodiev, JDT, 1712.07124; Komiske, Metodiev, Schwartz, 1612.01551]

Comparing Data Representations



Original 4-Vectors: $\{p_1^\mu, p_2^\mu, ..., p_N^\mu\}$

Variable-length, unordered set

EFP Basis: $$\mathrm{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$

Automatically permutation invariant

Linear spanning basis (over-complete)

Computational nightmare of $O(M^N)$?

Too naive, can use variable elimination

The Theory-Meets-Computation Approach

Underlying Physics

Natural & Efficient Data Representation

Desired Computational Property

→ Underlying Physics

Natural Data Representation

→ Suitable Algorithm

What is the set of *all*
linearly-computable
permutation-invariant
IRC-safe structures?

The Energy Flow Moments

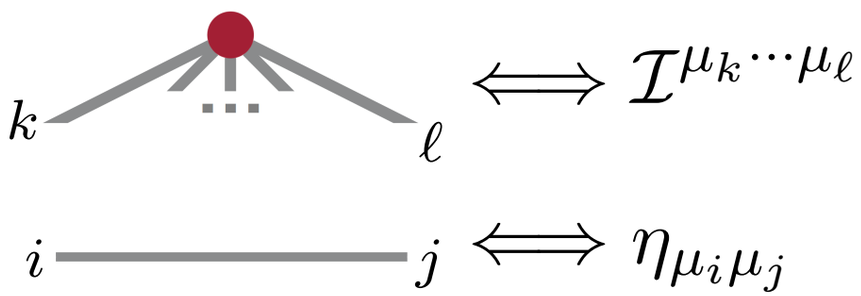
Underlying Physics
→ Natural Data Representation
Suitable Algorithm

$$\mathcal{I}^{\mu_1 \mu_2 \cdots \mu_v} = \sum_{i=1}^{M} E_i \, \hat{p}^{\mu_1} \, \hat{p}^{\mu_2} \cdots \hat{p}^{\mu_v}$$

Particle Relabeling — Infrared Safety

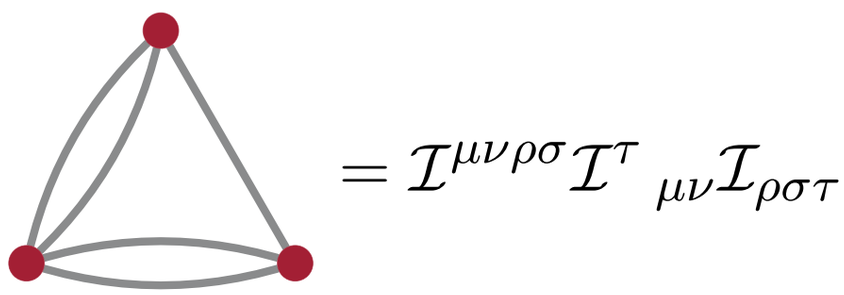
Special Choice of Angle $\theta_{ij} = 2 \, \eta_{\mu\nu} \, \hat{p}_i^{\mu} \hat{p}_j^{\nu}$

EFMs



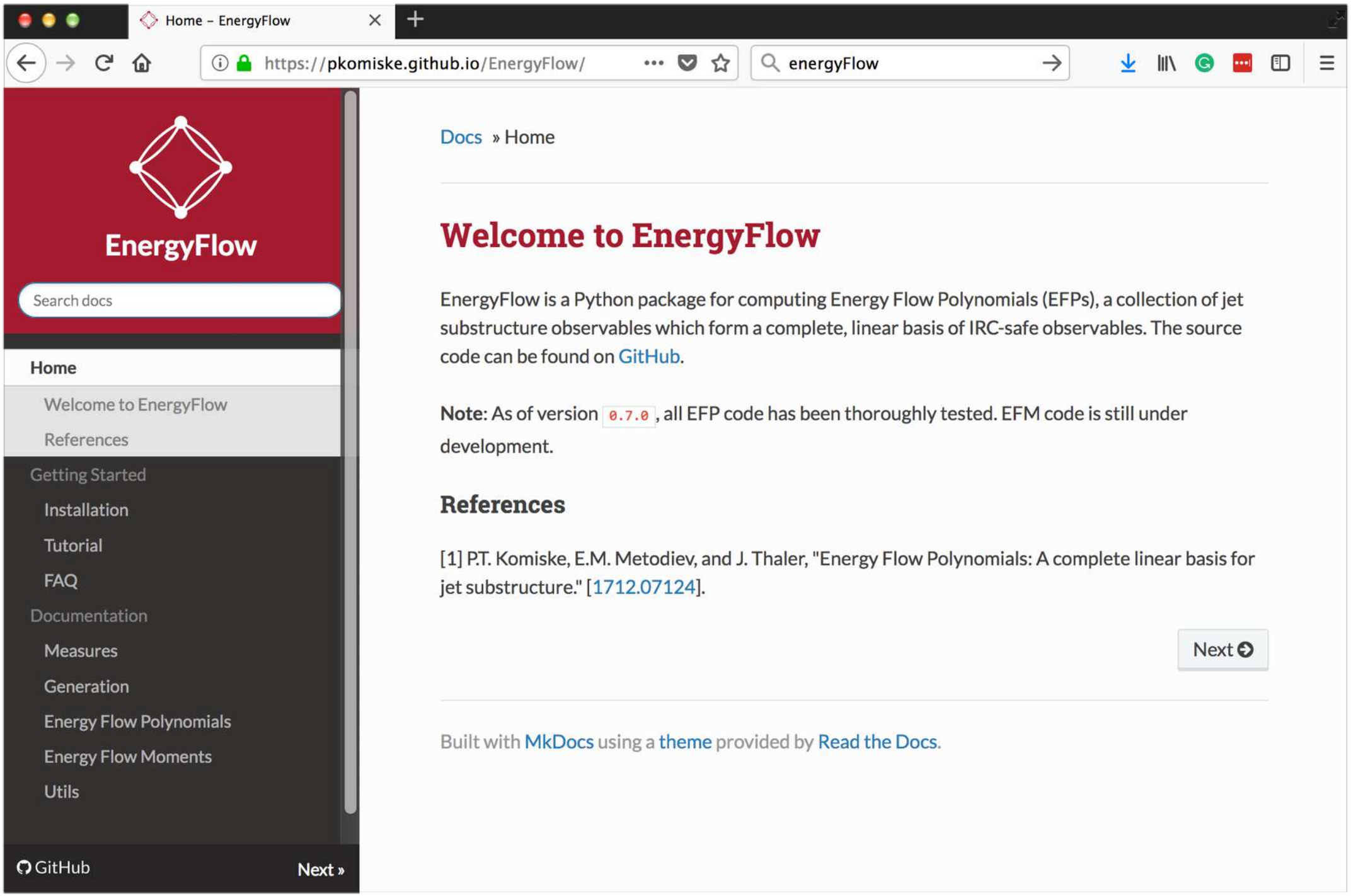
Compute in O(M)!

⇒ EFPs



Also compute in O(M)!

[Komiske, Metodiev, JDT, to appear]


Home – EnergyFlow
https://pkomiske.github.io/EnergyFlow/
energyFlow
EnergyFlow
Search docs
Home
Welcome to EnergyFlow
References
Getting Started
Installation
Tutorial
FAQ
Documentation
Measures
Generation
Energy Flow Polynomials
Energy Flow Moments
Utils
GitHub
Next »
Docs » Home
Welcome to EnergyFlow
EnergyFlow is a Python package for computing Energy Flow Polynomials (EFPs), a collection of jet substructure observables which form a complete, linear basis of IRC-safe observables. The source code can be found on GitHub.
Note: As of version 0.7.0, all EFP code has been thoroughly tested. EFM code is still under development.
References
[1] P.T. Komiske, E.M. Metodiev, and J. Thaler, "Energy Flow Polynomials: A complete linear basis for jet substructure." [1712.07124].
Next
Built with MkDocs using a theme provided by Read the Docs.

Punchlines

Underlying Physics
Natural Data Representation
Suitable Algorithm

Jet Constituents:

Variable-length, unordered set of particle-like objects

Energy Flow Polynomials:

Linear spanning basis for IRC-safe jet observables

Energy Flow Moments:

Linearly computable building blocks for EFPs

Linear Regression:

If it works, use it

Deep Learning or Deep Thinking ?

Both