New Improvements in Jet Physics

Jesse Thaler

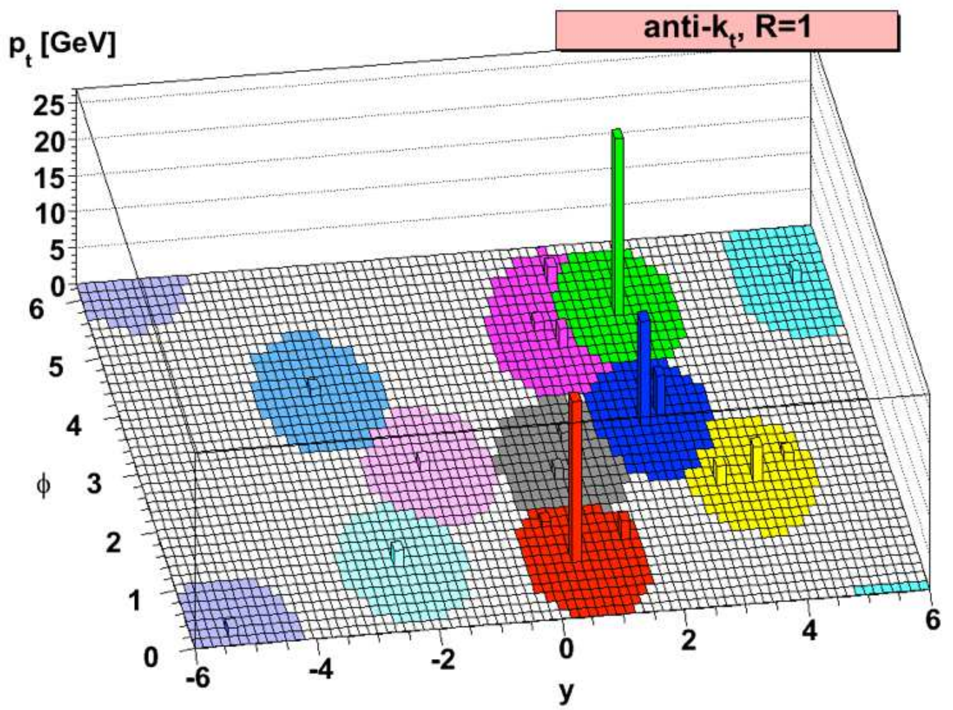
SUSY 2018, Barcelona — July 23, 2018

Last week in Paris…

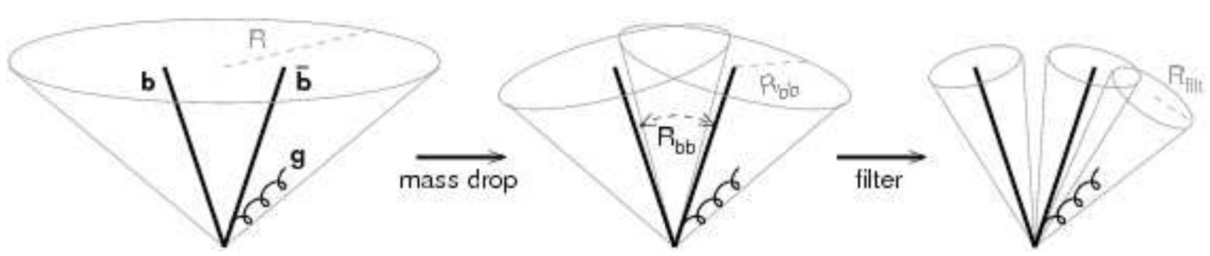
Celebrating 10 Years of Jet Substructure*

Flashback to 2008

Anti-k_t February 8



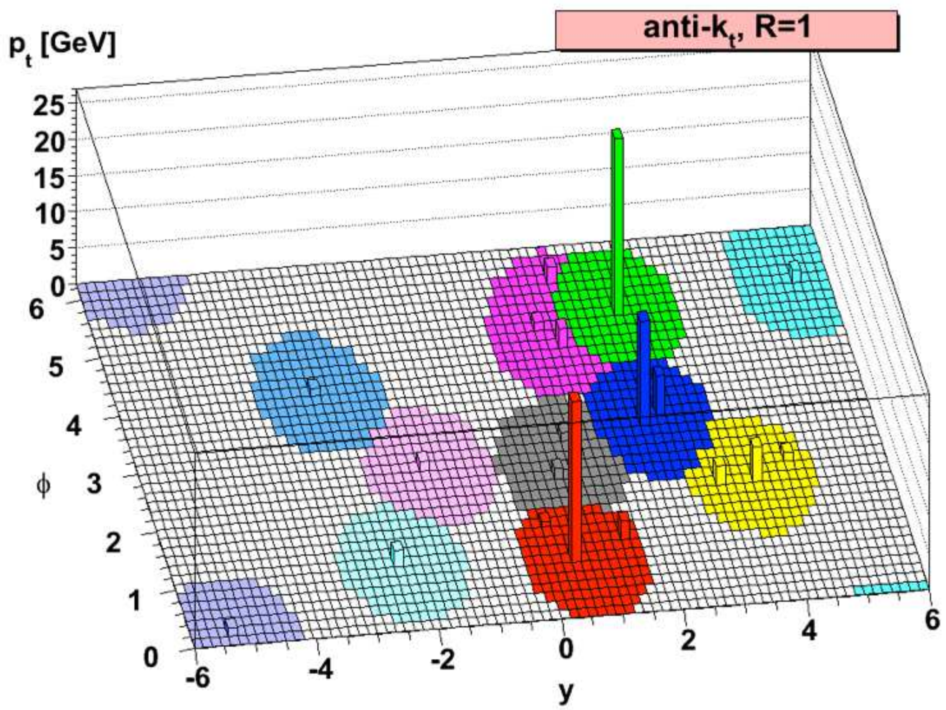
BDRS February 18



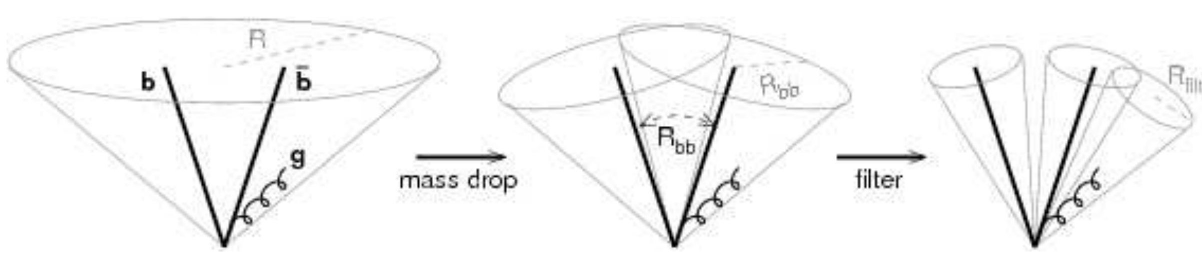
[Cacciari, Salam, Soyez, 0802.1189;
Butterworth, Davison, Rubin, Salam, 0802.2470; see also Seymour, 1991, 1994]

Flashback to 2008

Anti-k_t February 8



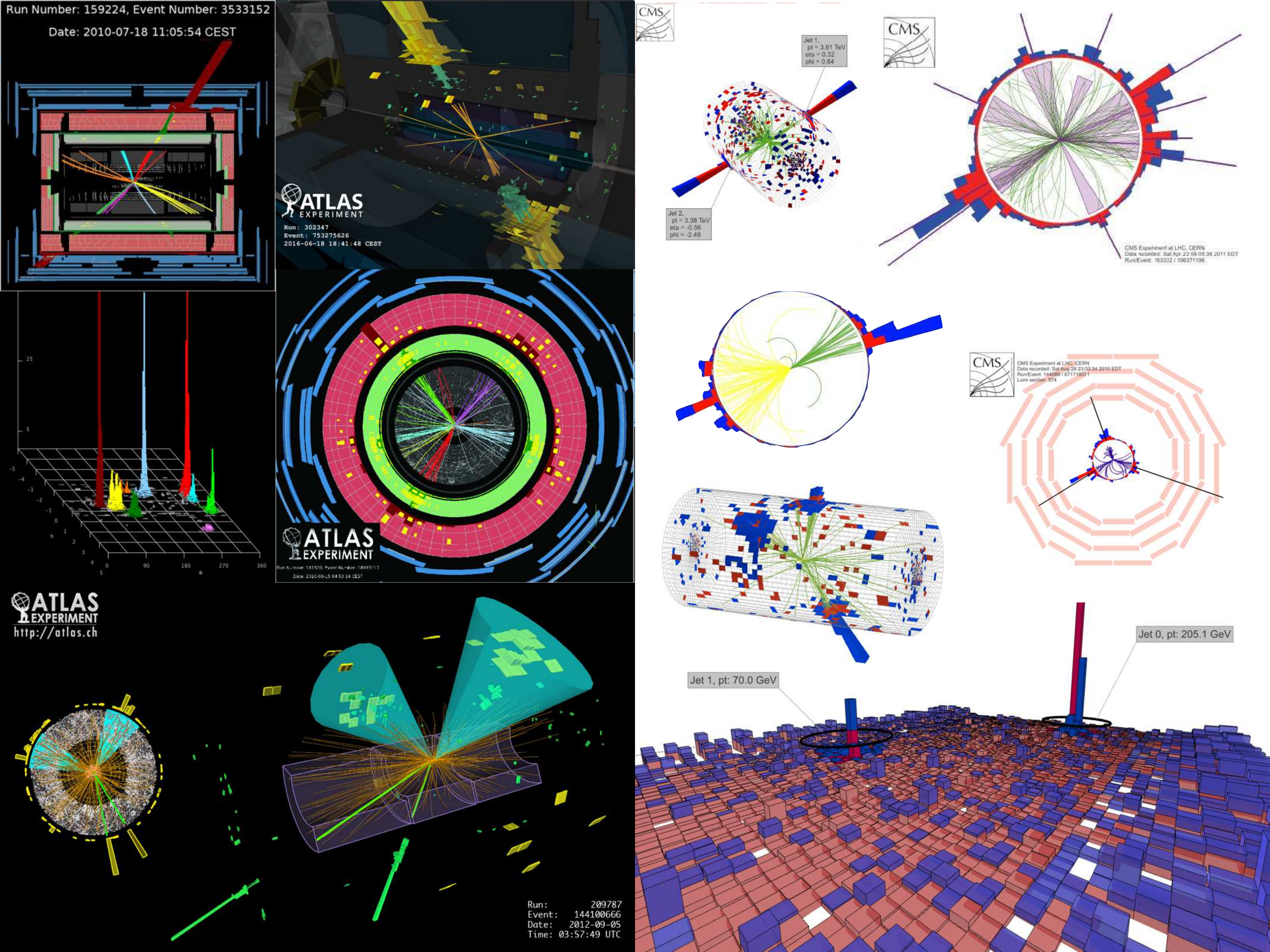
BDRS February 18

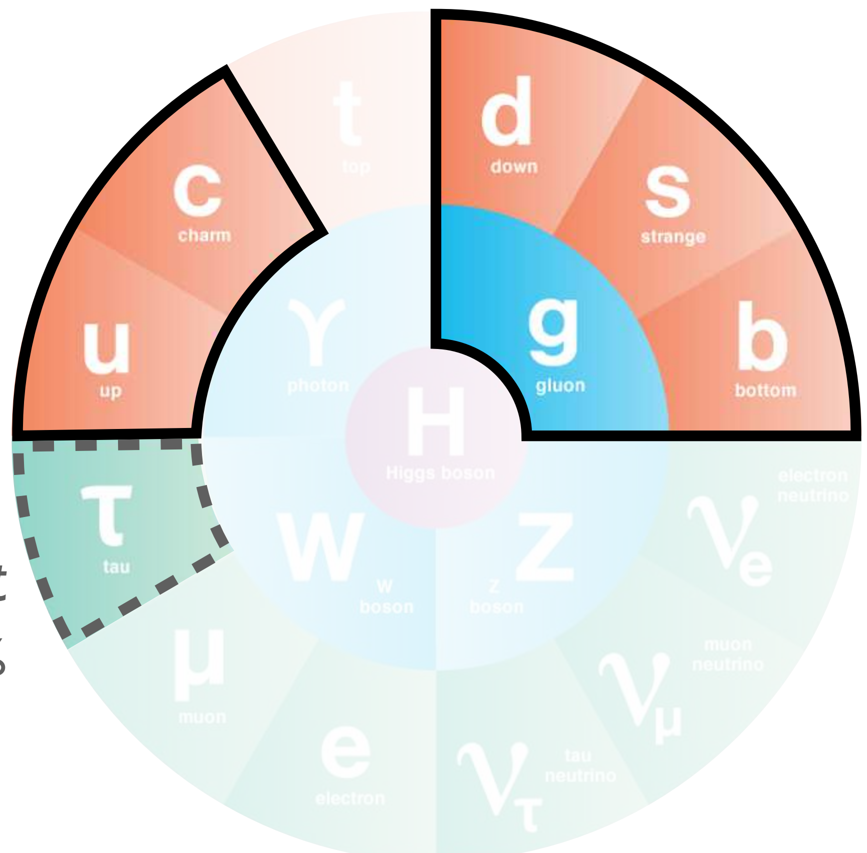


The Incident

September 19

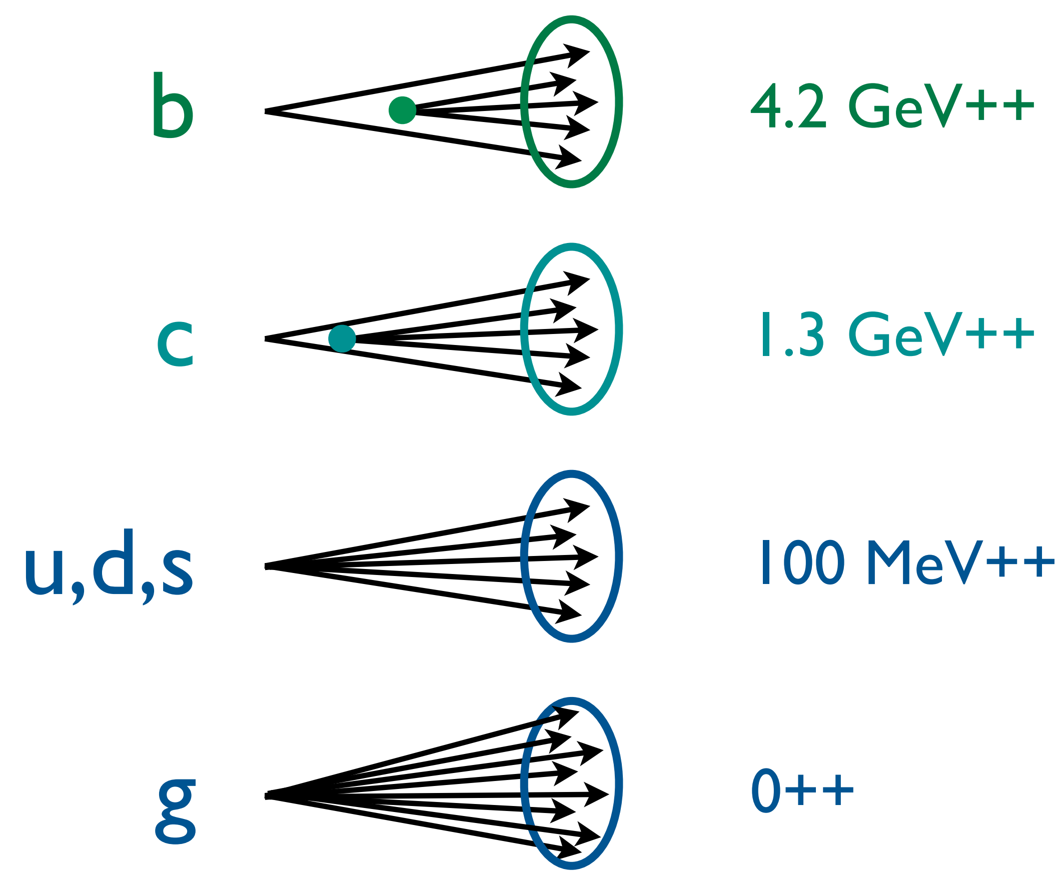
[Cacciari, Salam, Soyez, 0802.1189; Butterworth, Davison, Rubin, Salam, 0802.2470; see also Seymour, 1991, 1994]


Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST
ATLAS EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST
CMS
Jet 1, pt = 3.61 TeV eta = 0.32 phi = 0.64
Jet 2, pt = 3.38 TeV eta = -0.56 phi = -2.49
CMS
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 196371106
ATLAS EXPERIMENT
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
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

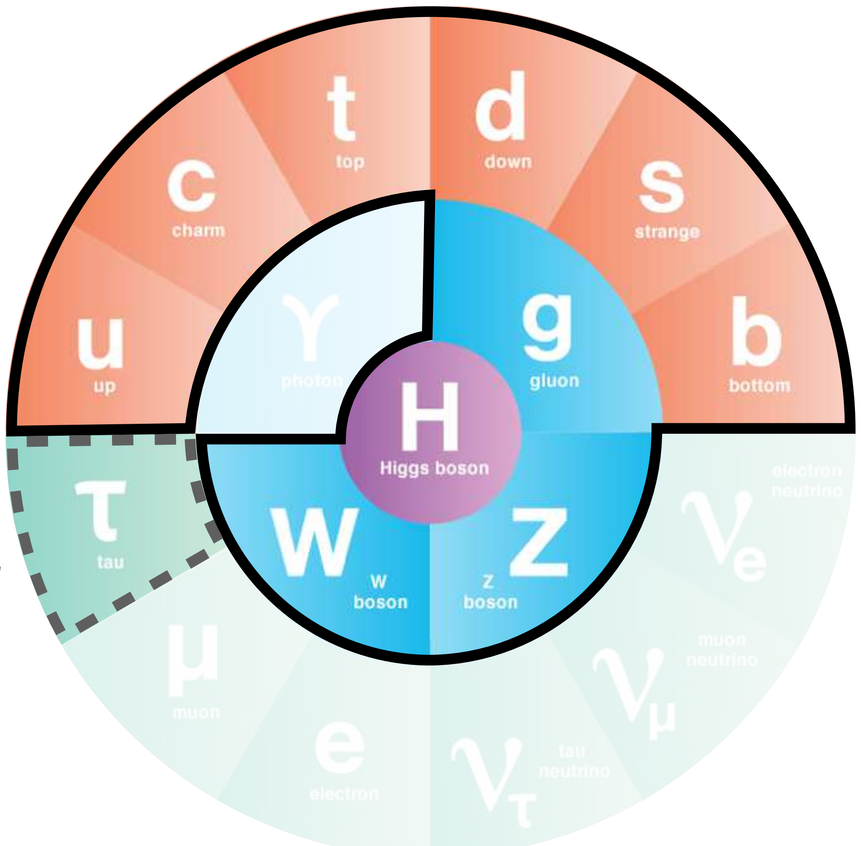


micro-jet
≈ 65%

Jets from the Standard Model



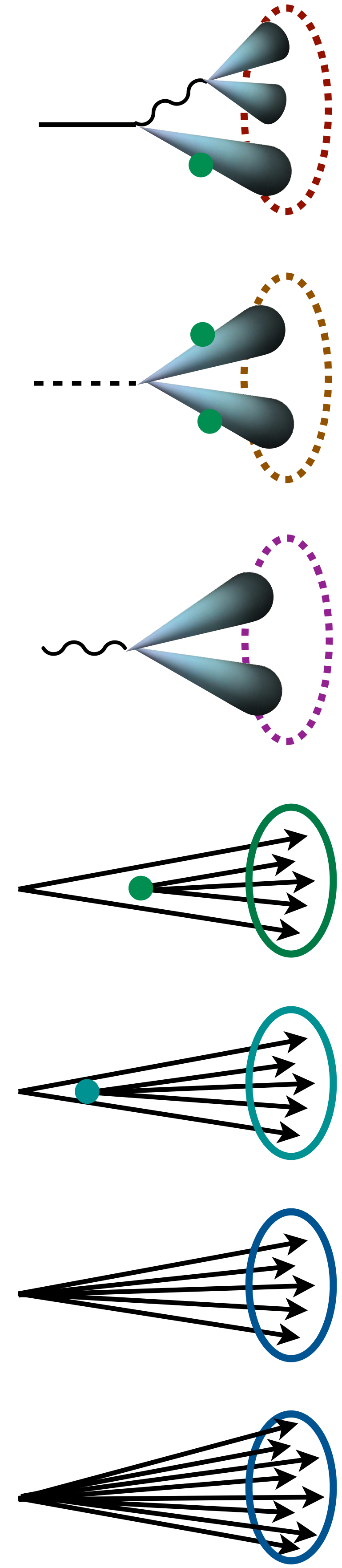
++ = Mass from QCD Radiation



micro-jet
≈ 65%

Jets from the Standard Model

++ = Mass from QCD Radiation



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

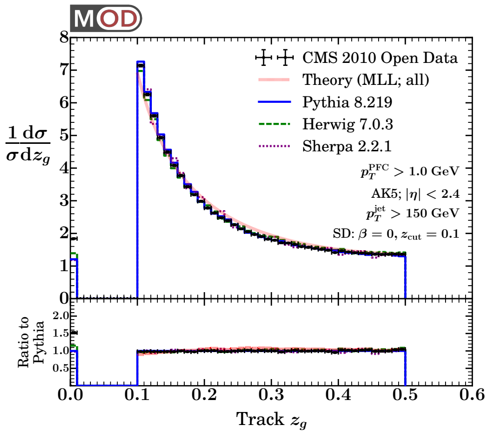
Jet Substructure/Boosted Objects at SUSY 2018

Plenary Talks:	Monday @ 10:00: Seth Zenz (SM Higgs properties measurements) Tuesday @ 09:00: Hannsjorg Weber (SUSY searches - strong production) Tuesday @ 10:00: Roger Wolf (BSM Higgs searches) Wednesday @ 09:00: Xabier Cid Vidal (Exotic searches - prompt signatures)
Supersymmetry: Models, Phenomenology & Experimental Results:	Monday @ 15:00: Louis-Guillaume Gagnon (ATLAS) Monday @ 15:20: Myriam Schoenenberger (CMS) Monday @ 15:40: Jesse Liu (ATLAS) Monday @ 17:50: Nathaniel Pastika (CMS) Wednesday @ 15:50: Alejandro Gomez Espinosa Thursday @ 18:00: Shing Chau Leung (Theory)
Precision Calculations and MC Tools:	Monday @ 17:55: Yasuhito Sakaki (Theory)
Alternatives to Supersymmetry:	Wednesday @ 15:30: Romain Madar (ATLAS) Wednesday @ 15:50: Alberto Iorio (CMS) Wednesday @ 16:10: Robin Erbacher (CMS) Wednesday @ 18:20: Abhishek Iyer (Theory) Thursday @ 14:30: Daniela Schafer (CMS) Thursday @ 14:50: Gabriele Chiodini (ATLAS)
Dark Matter & Astroparticle Physics:	Thursday @ 14:30: Michaela Queitsch-Maitland (ATLAS)
Electroweak, Top & Higgs Physics:	Wednesday @ 17:00: Elisabeth Petit (ATLAS)
Poster Session:	Tuesday @ 18:40: Daniela Salvatore (ATLAS)

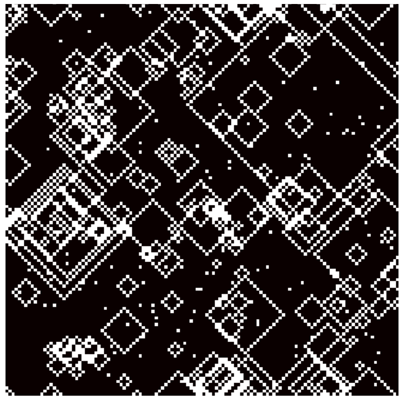
BOOST 2018

Please email me if I missed (or misrepresented) your talk!

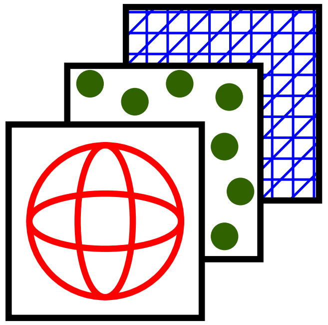
Outline



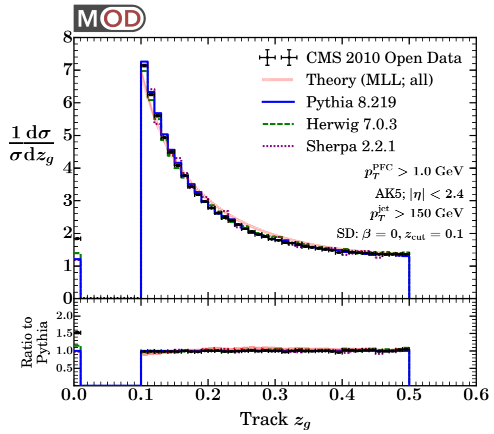
Three Trends in Jet Physics



Into the Network



Data ex Machina



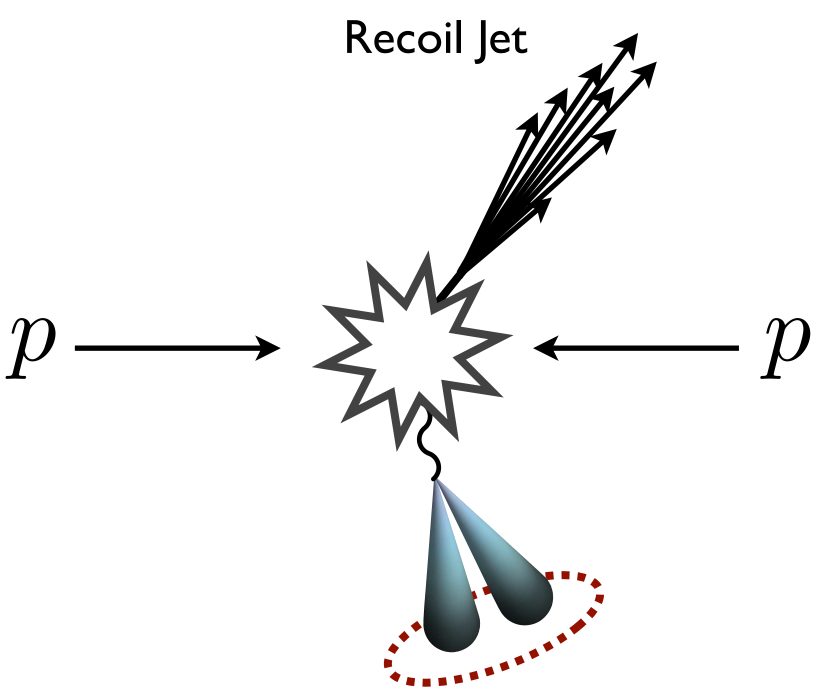
Three Trends in Jet Physics

Into the Network

Data ex Machina

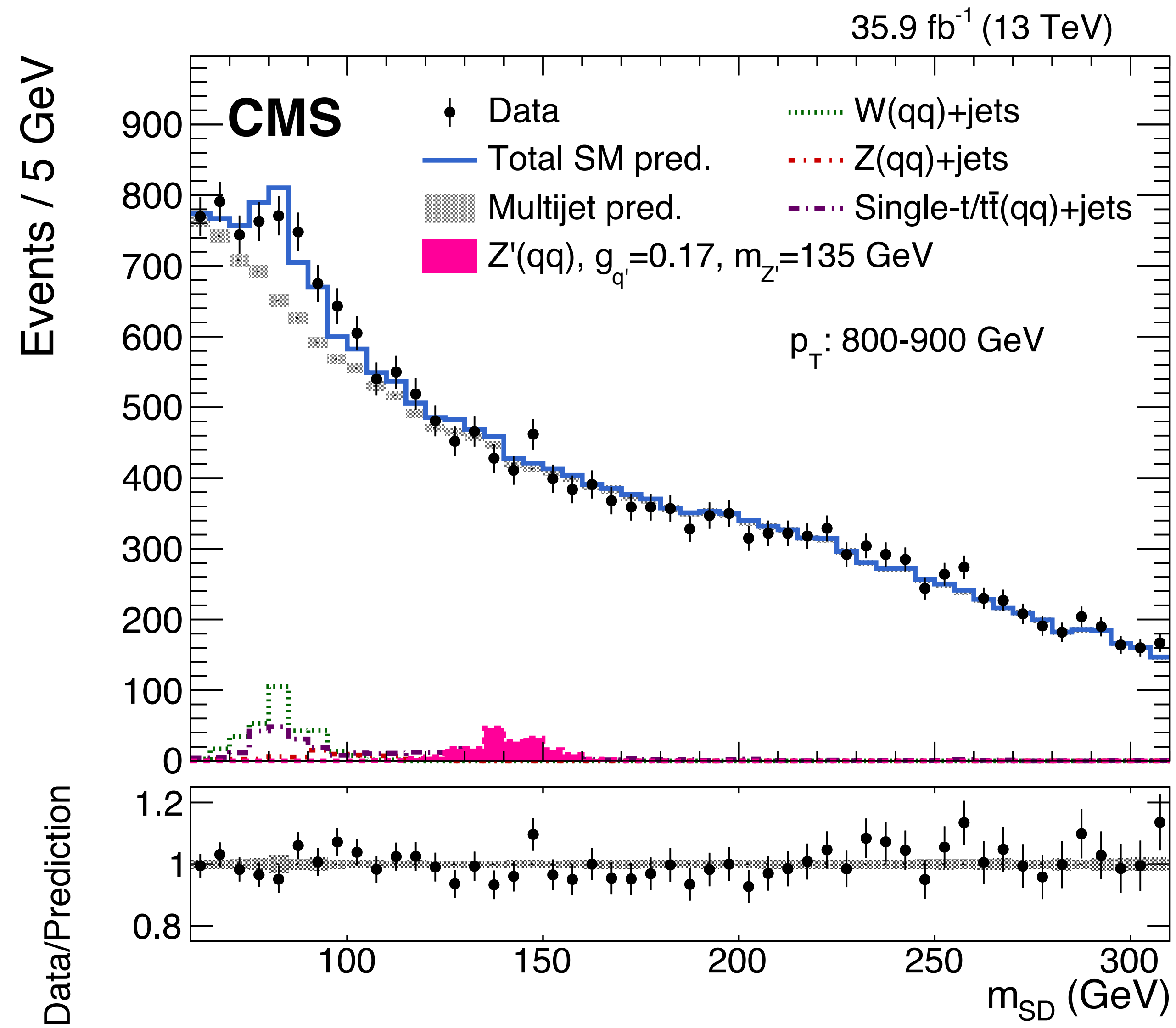
The Rise of Extreme Kinematics

Boosted Hadronic Resonances



High p_T Z or Z′

with PUPPI
+ *Soft Drop*
+ *N_2*
+ *DDT*



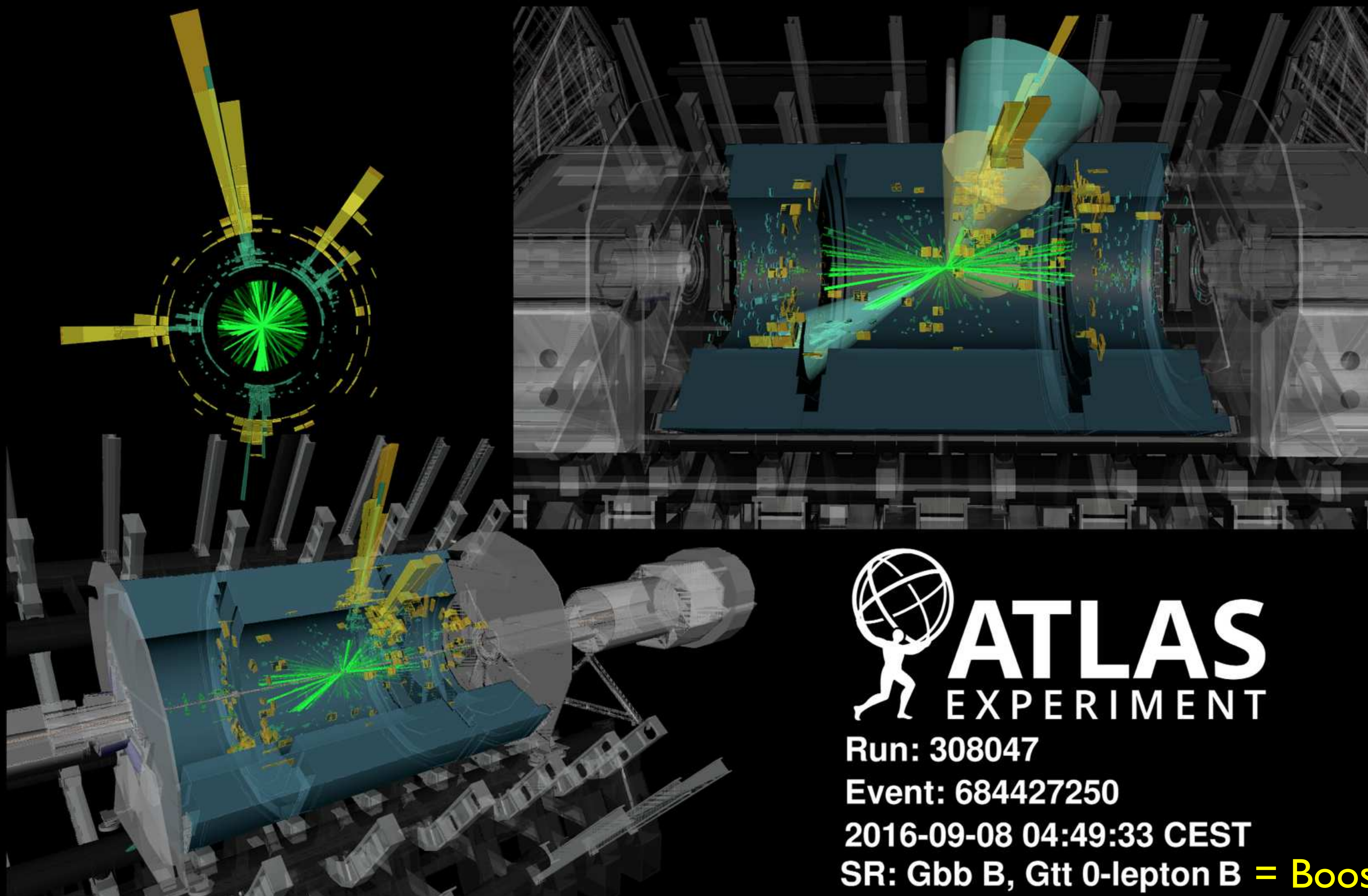
Discovered at SPS

Strongest bounds since SPS (!)

[CMS, 2017; using Bertolini, Harris, Low, Tran, 2014; Larkoski, Marzani, Soyez, JDT, 2014; Moult, Necib, JDT, 2016; Dolen, Harris, Marzani, Rappoccio, Tran, 2016; see boosted H →bb in backup]

The Rise of Extreme Kinematics

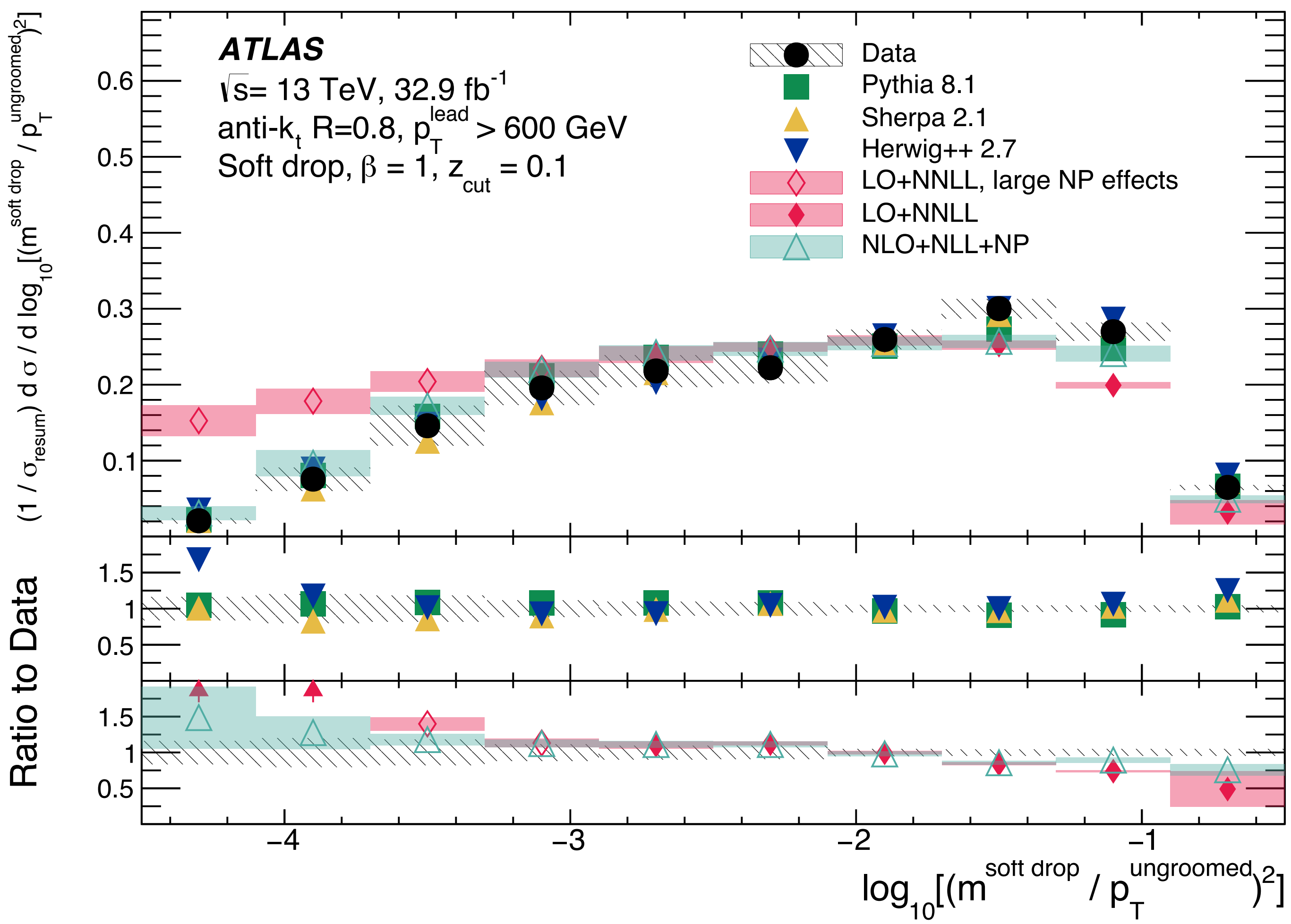
Gluino Cascade Decays



= Boosted

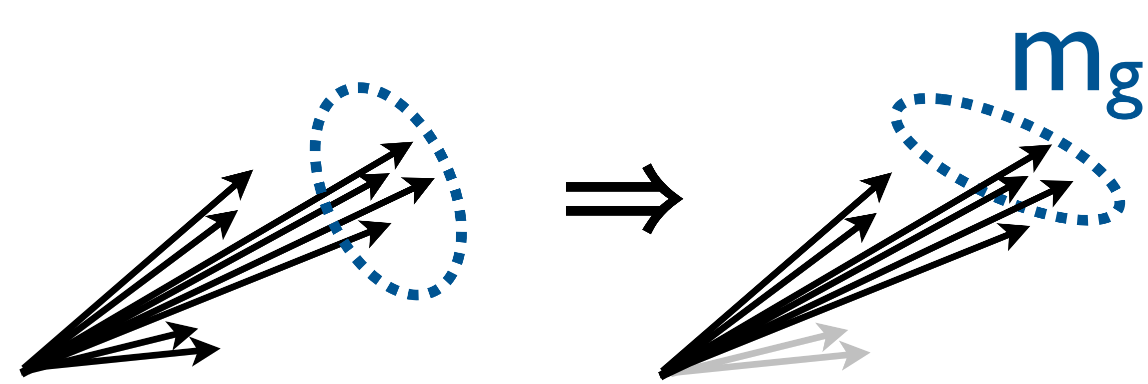
[ATLAS, 1711.01901]

The Rise of Precision Jet Physics



Groomed Jet Mass

Soft Drop (β=1)

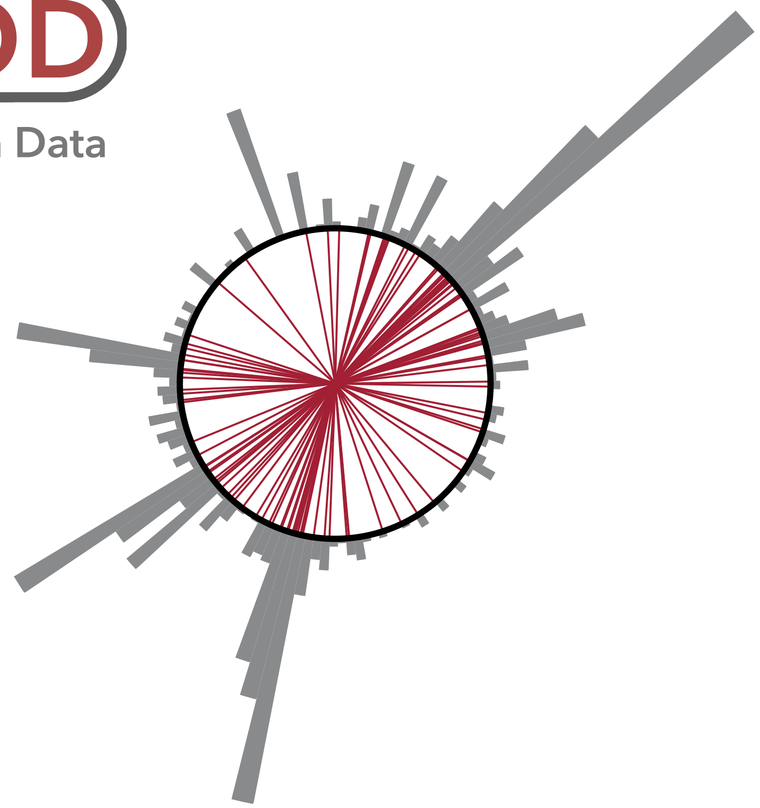


[ATLAS, 1711.08341; see also CMS, 1807.05974 in backup; using Larkoski, Marzani, Soyez, JDT, 1402.2657; compared to Frye, Larkoski, Schwartz, Yan, 1603.06375, 1603.09338; Marzani, Schunk, Soyez, 1704.02210, 1712.05105]

The Rise of Precision Jet Physics

First ever publication on CMS open data (!)

MIT Open Data

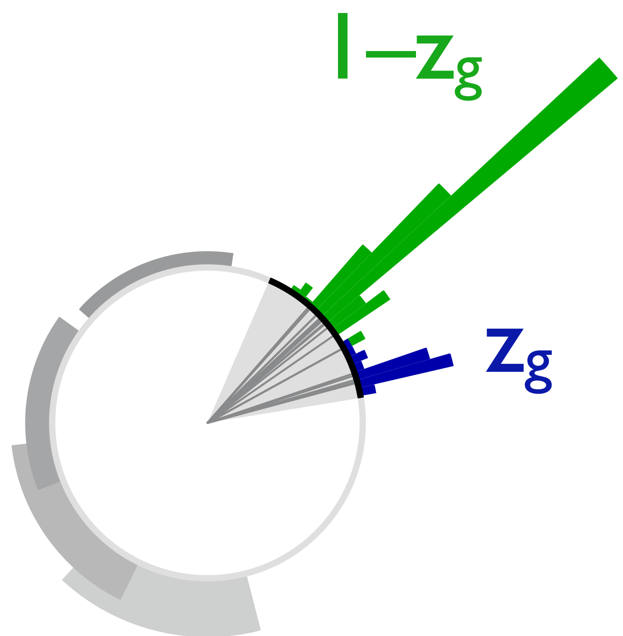


[Larkoski, Marzani, JDT, Tripathee, Xue, 1704.05066, 1704.05842; see also CMS, 1708.09429; using Larkoski, Marzani, Soyez, JDT, 1402.2657; Dasgupta, Fregoso, Marzani, Salam, 1307.0007; compared to Larkoski, Marzani, JDT, 1502.01719; see also Larkoski, JDT, 1307.1699]

The Rise of Precision Jet Physics

First ever publication on CMS open data (!)

MIT Open Data

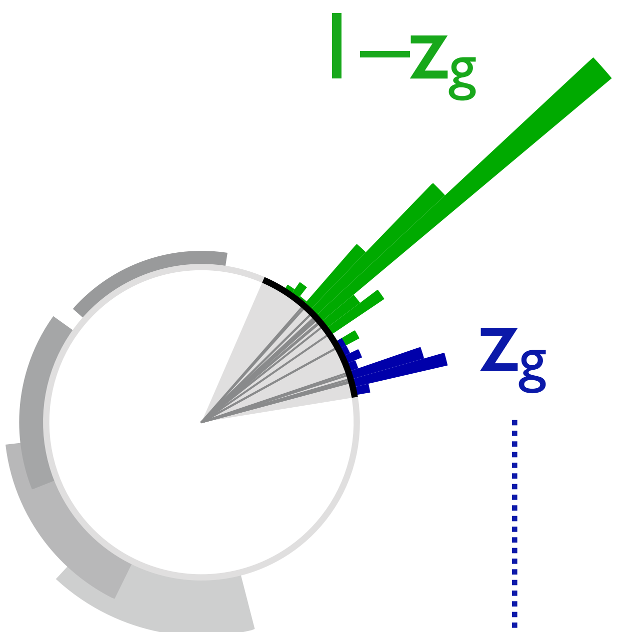


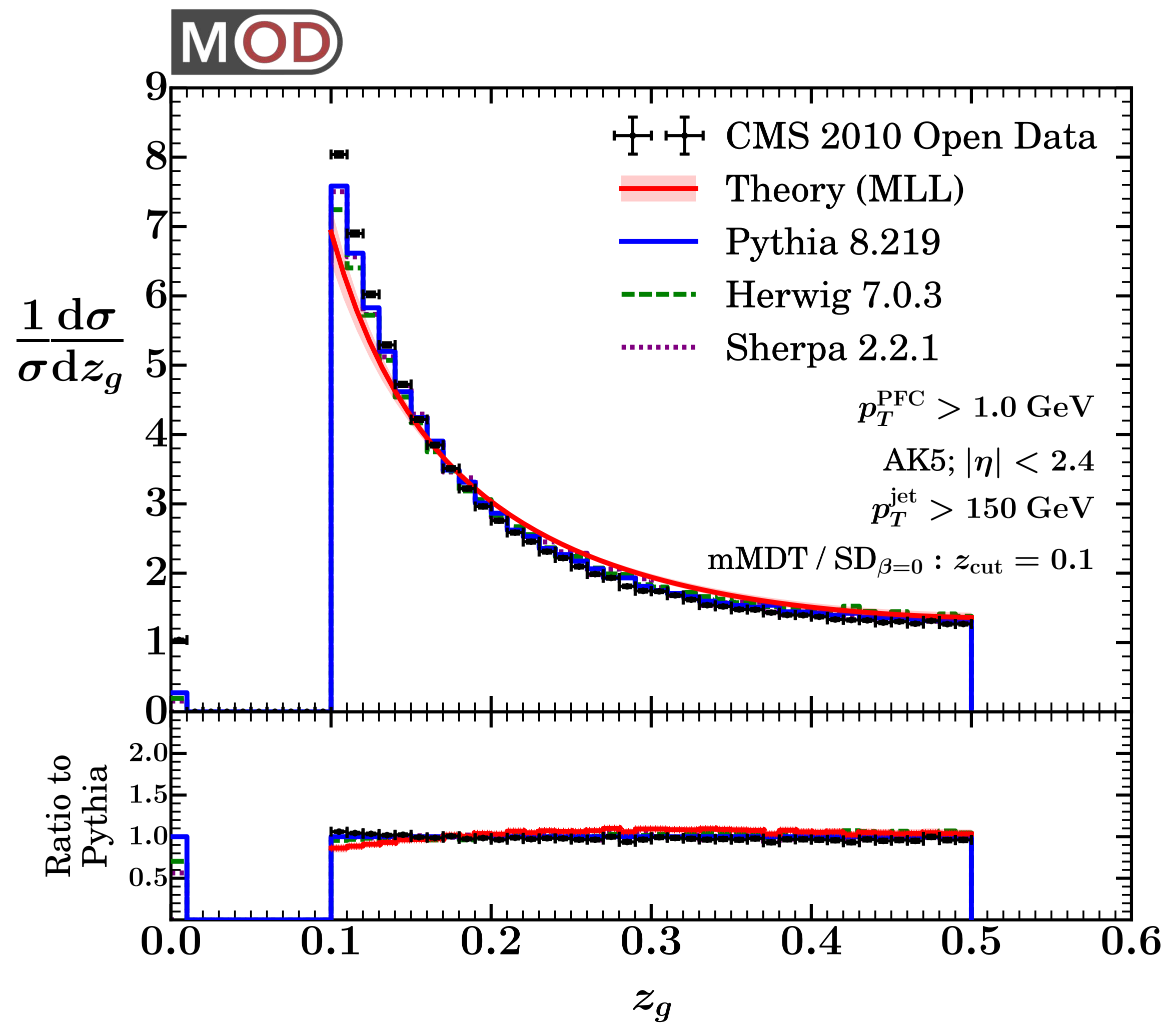
[Larkoski, Marzani, JDT, Tripathee, Xue, 1704.05066, 1704.05842; see also CMS, 1708.09429; using Larkoski, Marzani, Soyez, JDT, 1402.2657; Dasgupta, Fregoso, Marzani, Salam, 1307.0007; compared to Larkoski, Marzani, JDT, 1502.01719; see also Larkoski, JDT, 1307.1699]

The Rise of Precision Jet Physics

First ever publication on CMS open data (!)

MIT Open Data



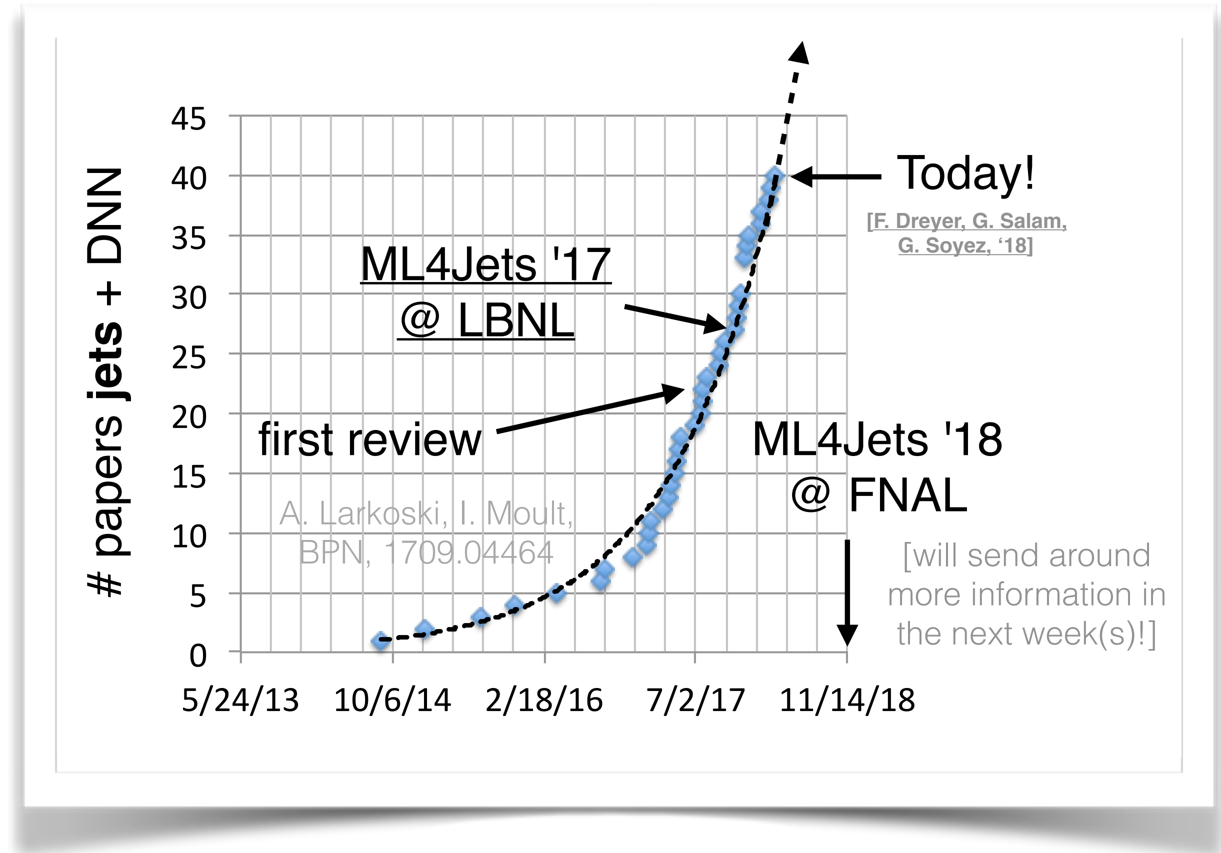


$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{d\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{dz}{z}}_{\text{Soft}}$$

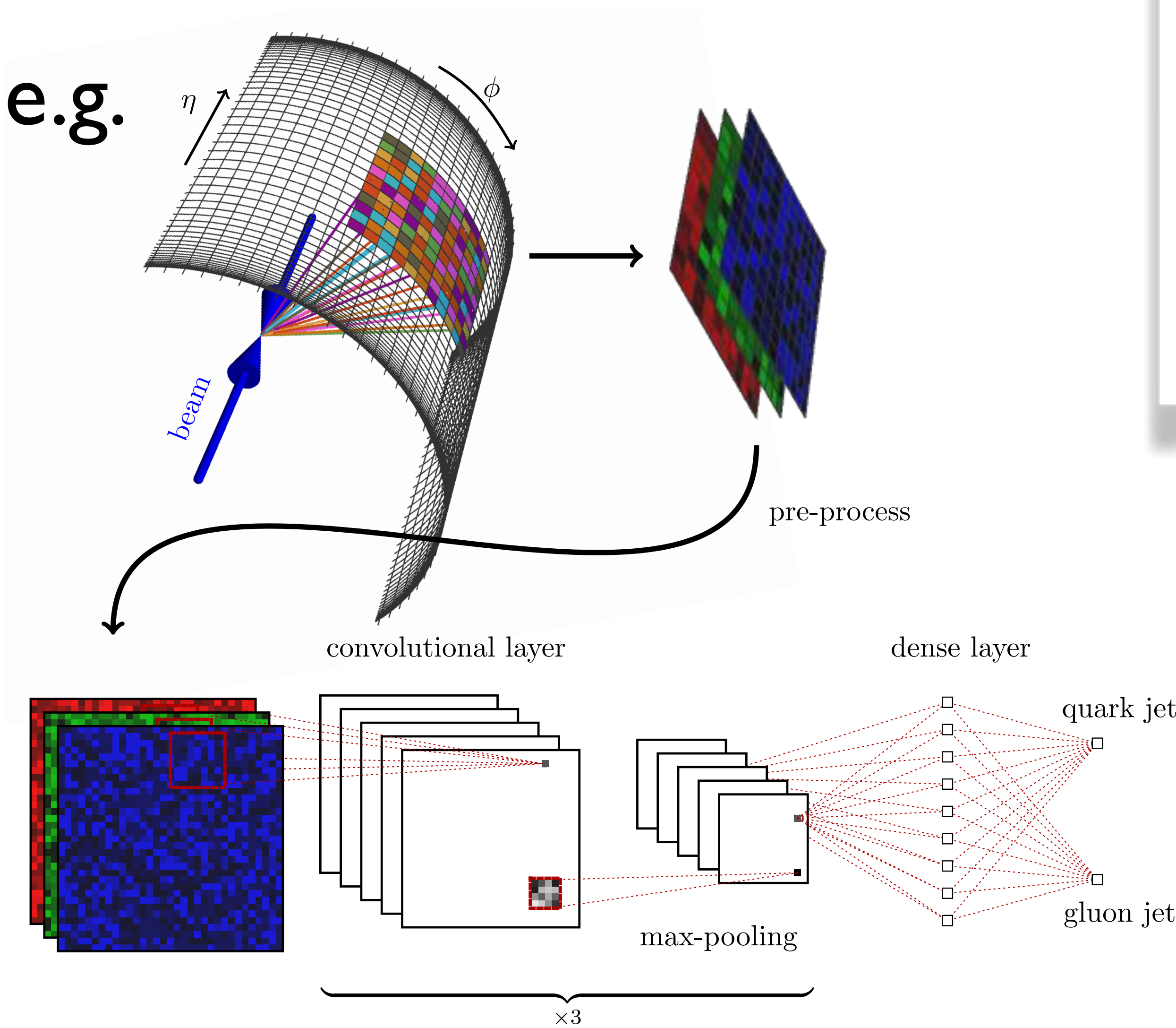
Groomed Momentum Fraction

[Larkoski, Marzani, JDT, Tripathee, Xue, 1704.05066, 1704.05842; see also CMS, 1708.09429; using Larkoski, Marzani, Soyez, JDT, 1402.2657; Dasgupta, Fregoso, Marzani, Salam, 1307.0007; compared to Larkoski, Marzani, JDT, 1502.01719; see also Larkoski, JDT, 1307.1699]

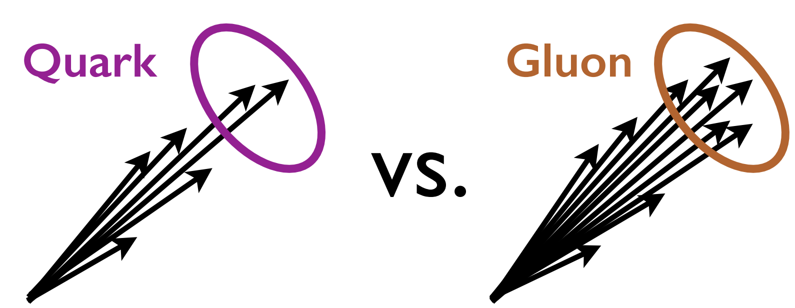
The Rise of Machine Learning for Jets



e.g.

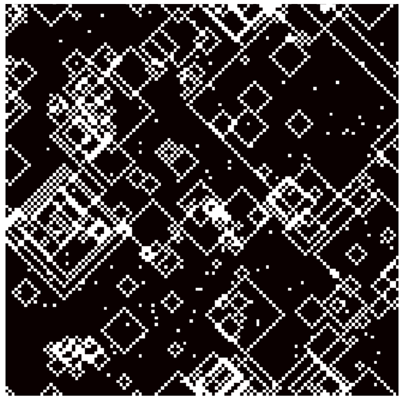


⇒



[e.g. Komiske, Metodiev, Schwartz, 1612.01551; Nachman, Boost 2018 Talk, July 20, 2018; reviews in Larkoski, Moult, Nachman, 1709.04464; Guest, Cranmer, Whiteson, 1806.11484]

Three Trends in Jet Physics



Into the Network

Data ex Machina

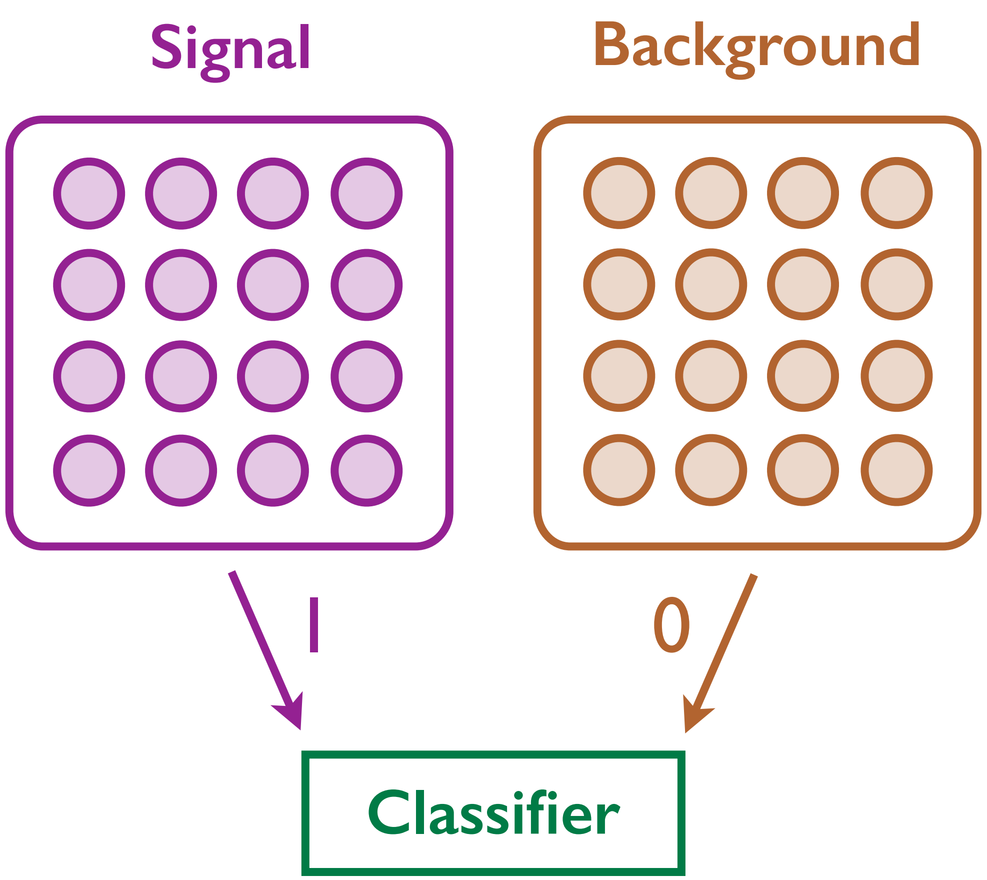
A Cartoon of Machine Learning

For fully-supervised jet classification

(see backup for regression, generation, modeling)

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

Classifier Inputs



Minimize Loss Function

(assuming infinite training sets)

$$h(\vec{x}) = \frac{p_{\mathrm{sig}}(\vec{x})}{p_{\mathrm{sig}}(\vec{x}) + p_{\mathrm{bkgd}}(\vec{x})}$$

Optimal Classifier (Neyman–Pearson)

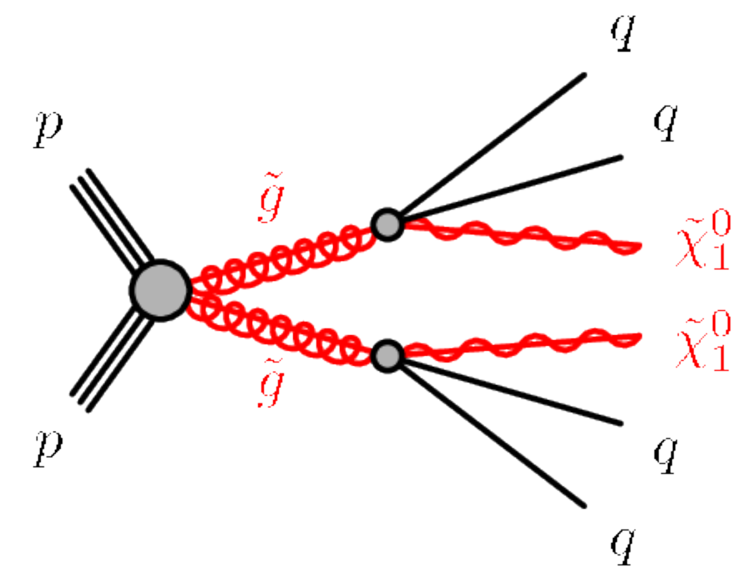
E.g. SUSY Search for Gluino Pairs

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

Inputs: Jet mass, width, track multiplicity

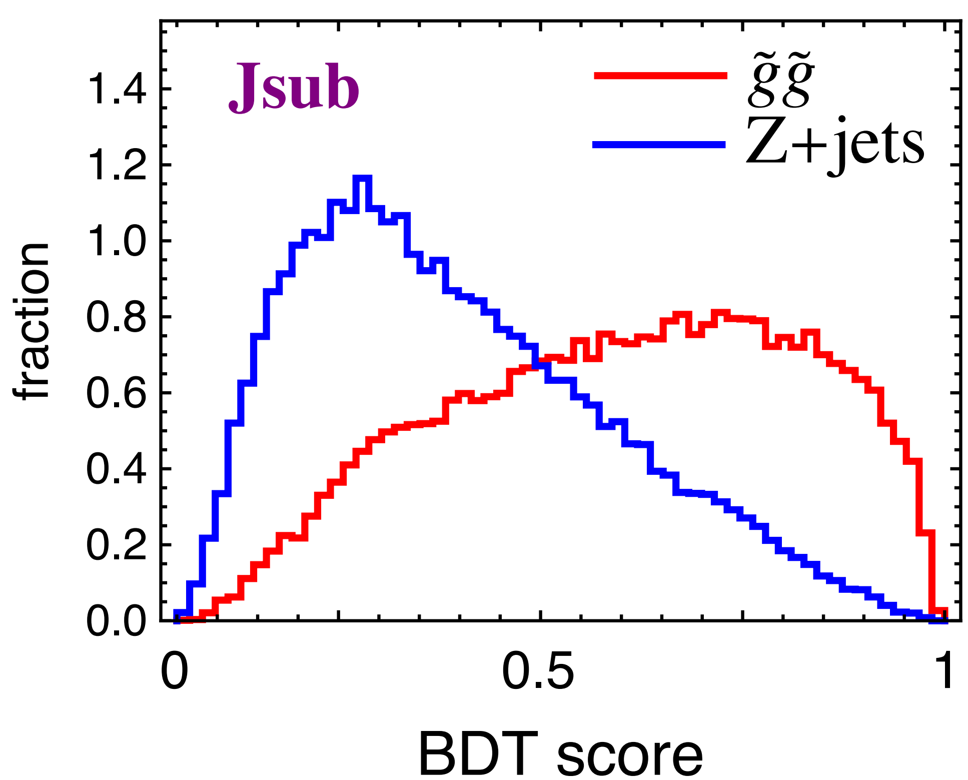
Signal: Quark enriched (C_F = 4/3)

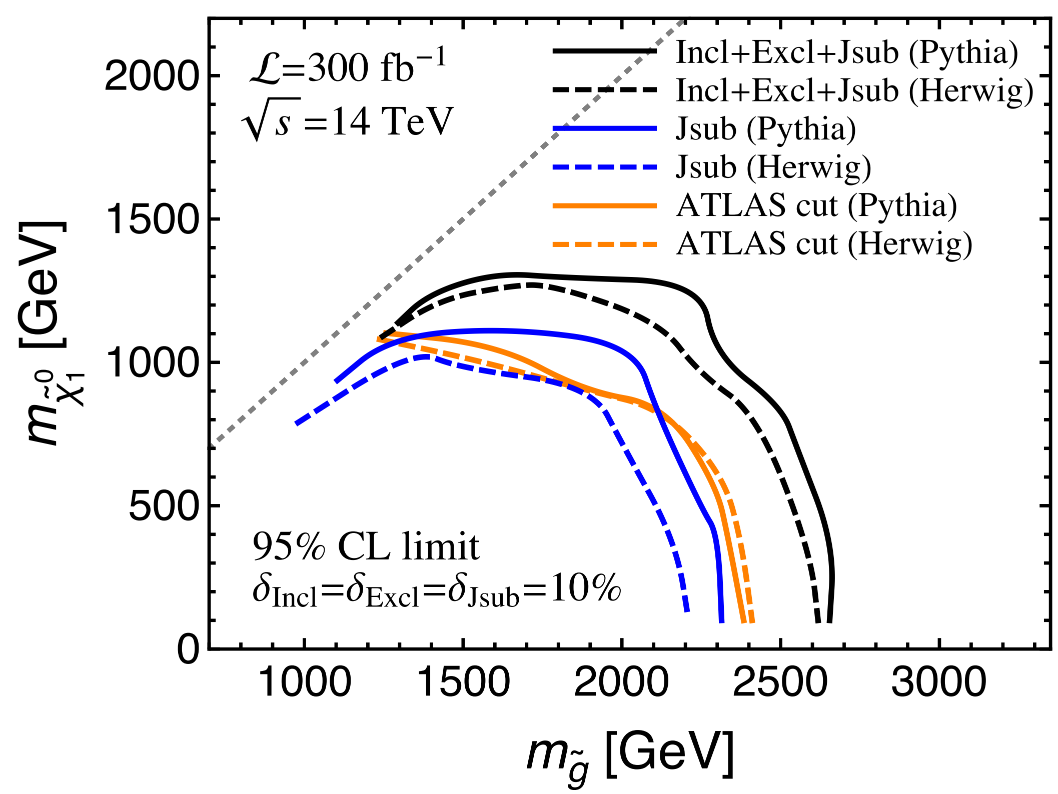
Background: Gluon enriched (C_A = 3)



← background

signal →





[Bhattacherjee, Mukhopadhyay, Nojiri, Sakakie, Webber, 1609.08781]

Jet Classification Studies

Mix and match

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

Classifier

Boosted Decision Tree
Fisher Linear Discriminant
Shallow Neural Network
Deep Neural Network
Convolutional Neural Network
Recurrent Neural Network
Recursive Neural Network
Combination/Lorentz Layers
…

Inputs

High-Level Features
Basis of High-Level Features
Jet Image
Multi-channel Jet Image
Abstract Jet Image
Sorted Four-Vectors
Clustered Four-Vectors
Lund Plane Emissions
Kitchen Sink
…

Signal vs. Background

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

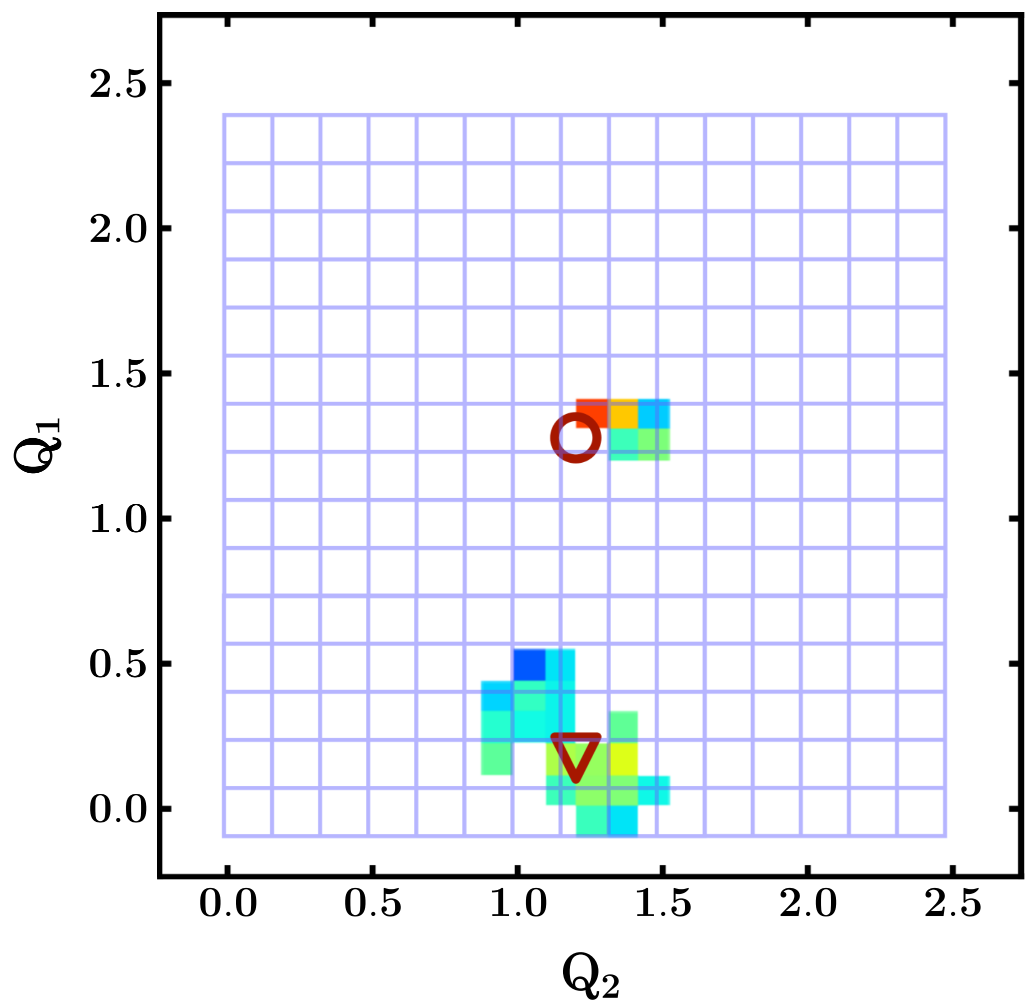
[Lönnblad, Peterson, Rögnvaldsson, 1990, …, Cogan, Kagan, Strauss, Schwartzman, 1407.5675; Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968; de Oliveira, Kagan, Mackey, Nachman, Schwartzman, 1511.05190; Baldi, Bauer, Eng, Sadowski, Whiteson, 1603.09349; Conway, Bhaskar, Erbacher, Pilot, 1606.06859; Guest, Collado, Baldi, Hsu, Urban, Whiteson, 1607.08633; Barnard, Dawe, Dolan, Rajcic, 1609.00607; Komiske, Metodiev, Schwartz, 1612.01551; Kasieczka, Plehn, Russell, Schell, 1701.08784; Louppe, Cho, Becot, Cranmer, 1702.00748; Pearkes, Fedorko, Lister, Gay, 1704.02124; Datta, Larkoski, 1704.08249, 1710.01305; Butter, Kasieczka, Plehn, Russell, 1707.08966; Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, 1708.07034; Aguilar Saavedra, Collin, Mishra, 1709.01087; Cheng, 1711.02633; Luo, Luo, Wang, Xu, Zhu, 1712.03634; Komiske, Metodiev, JDT, 1712.07124; Macaluso, Shih, 1803.00107; Fraser, Schwartz, 1803.08066; Choi, Lee, Perelstein, 1806.01263; Lim, Nojiri, 1807.03312; Dreyer, Salam, Soyez, 1807.04758; Moore, Nordström, Varma, Fairbairn, 1807.04769; plus my friends who will scold me for forgetting their paper; plus many ATLAS/CMS performance studies]

Jet Classification Studies

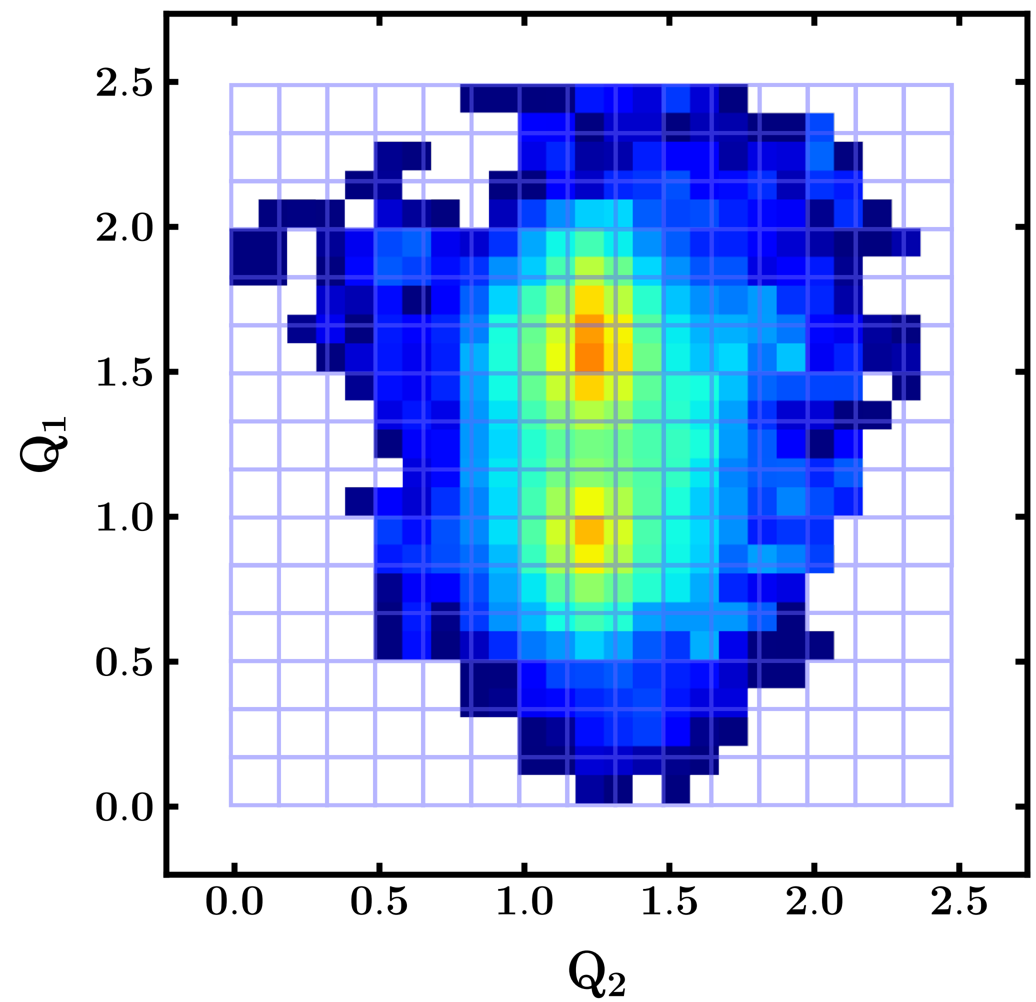
Mix and match

Standard CNN input: Jet images

Individual W jet



Ensemble average

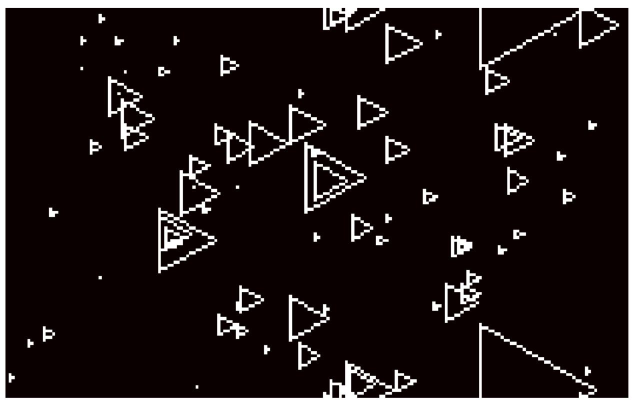


[Cogan, Kagan, Strauss, Schwartzman, 1407.5675]

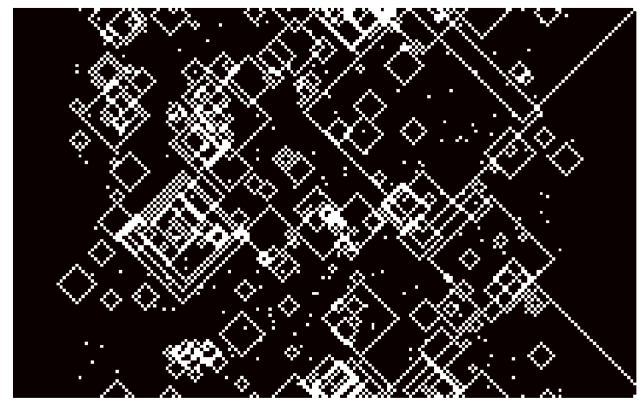
Jet Classification Studies

Mix and match

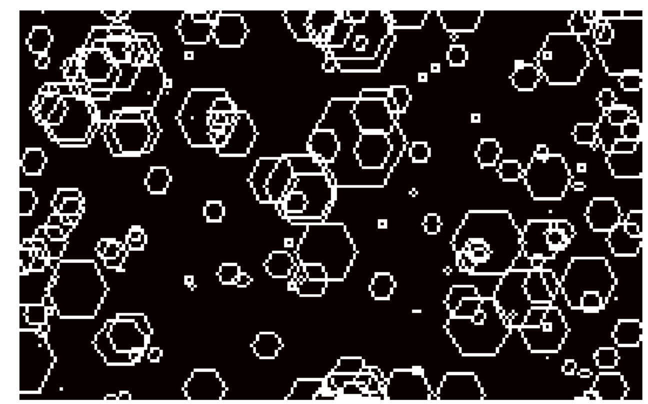
Novel CNN input: Abstract event images



(a) Photons

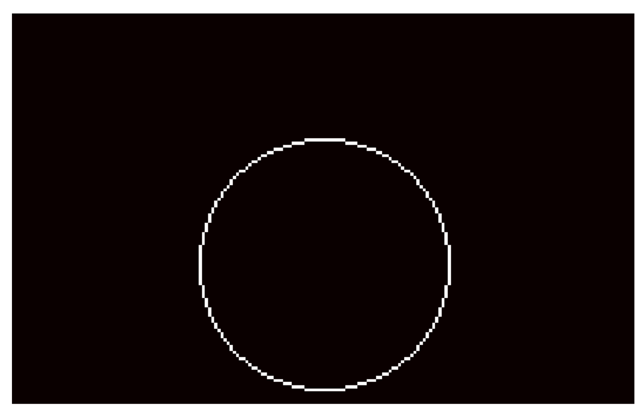


(b) Charged Particles

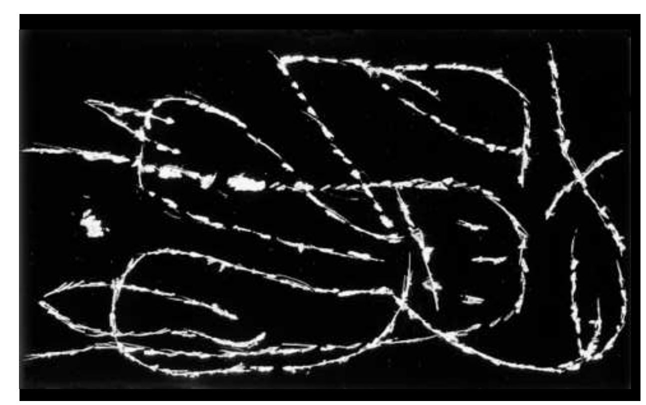


(c) Neutral Hadrons

(d) Lepton



(e) E_T^{miss}



(f) Miro (Untitled, 1977)

Addresses sparsity problem of standard energy-to-intensity mapping

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

Jet Classification Studies

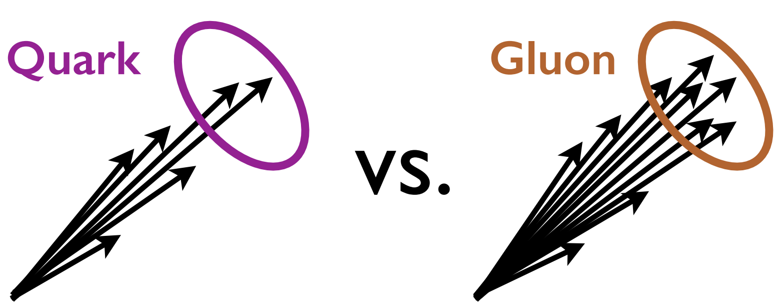
Mix and match

Surprise: Linear Regression ≈ Deep Networks (!)

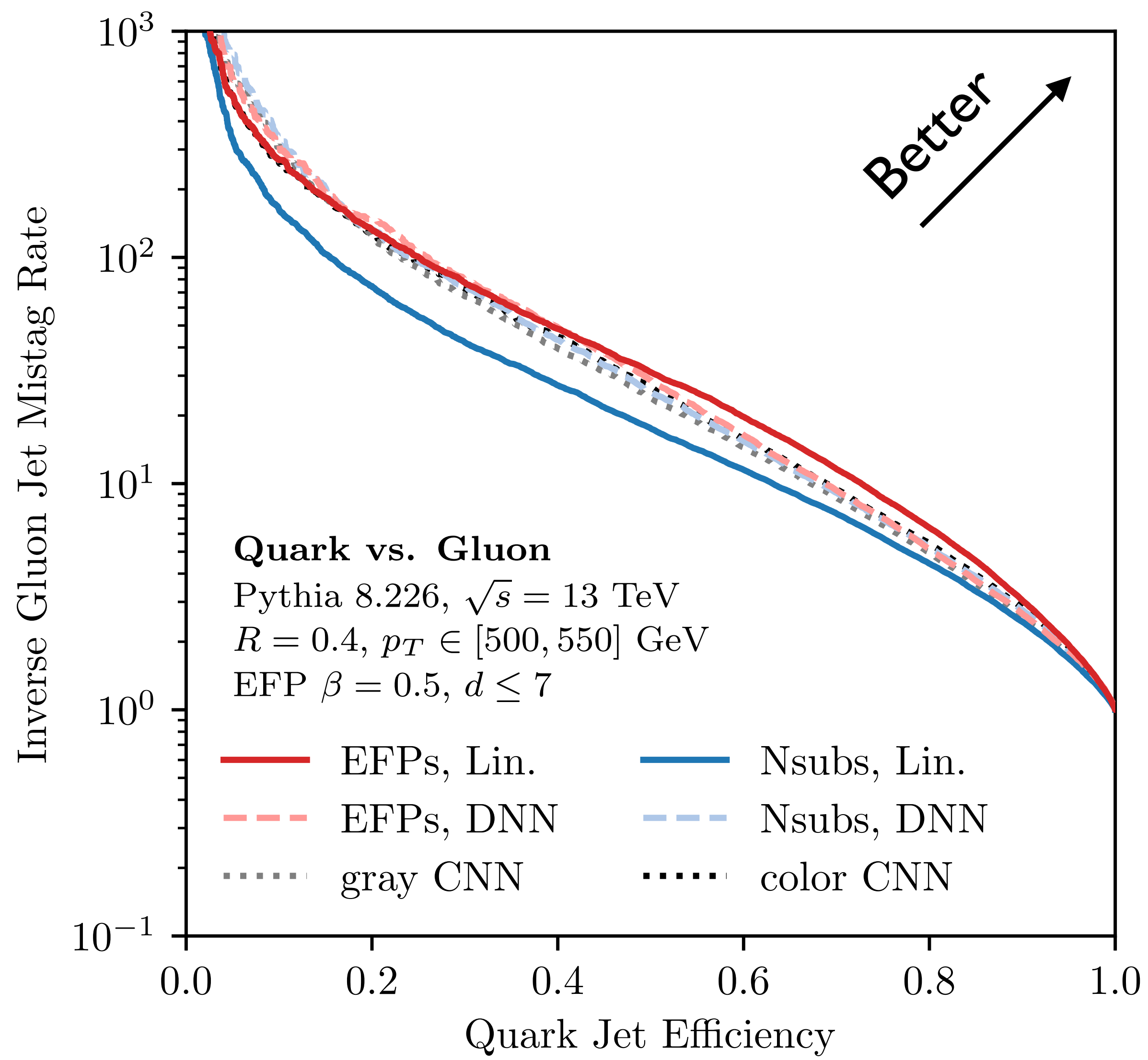
$$\mathcal{S} = \sum_{G} s_G \, \mathrm{EFP}_G$$

Classifier: Fisher Linear Discriminant

Inputs: Energy Flow Polynomials



"Any sufficiently advanced technology is indistinguishable from magic"



[Komiske, Metodiev, JDT, 1712.07124; see also Moore, Nordström, Varma, Fairbairn, 1807.04769; see backup for EFP definition]

My Perspective c. 2016

“Deep Learning”

My Perspective c. 2016

“Deep Learning” vs. “Deep Thinking”

What is the machine learning?
Biases from training on simulations?
Do jet categories even make sense?

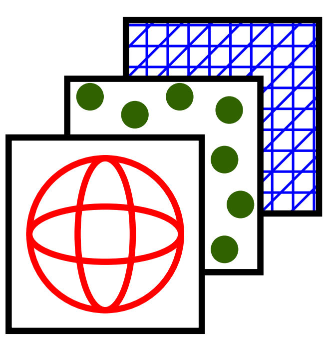
My Perspective c. 2018

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

What is the machine learning? Black box visualization
Biases from training on simulations? Weak supervision
Do jet categories even make sense? Topic modeling

Three Trends in Jet Physics

Into the Network

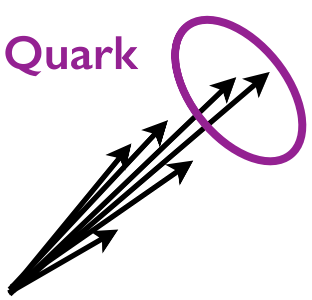


Data ex Machina

Data ex Machina

"A seemingly unsolvable problem is suddenly and abruptly resolved by an unexpected and seemingly unlikely occurrence, typically so much as to seem contrived"

Examples below taken from



vs.

with

Patrick Komiske

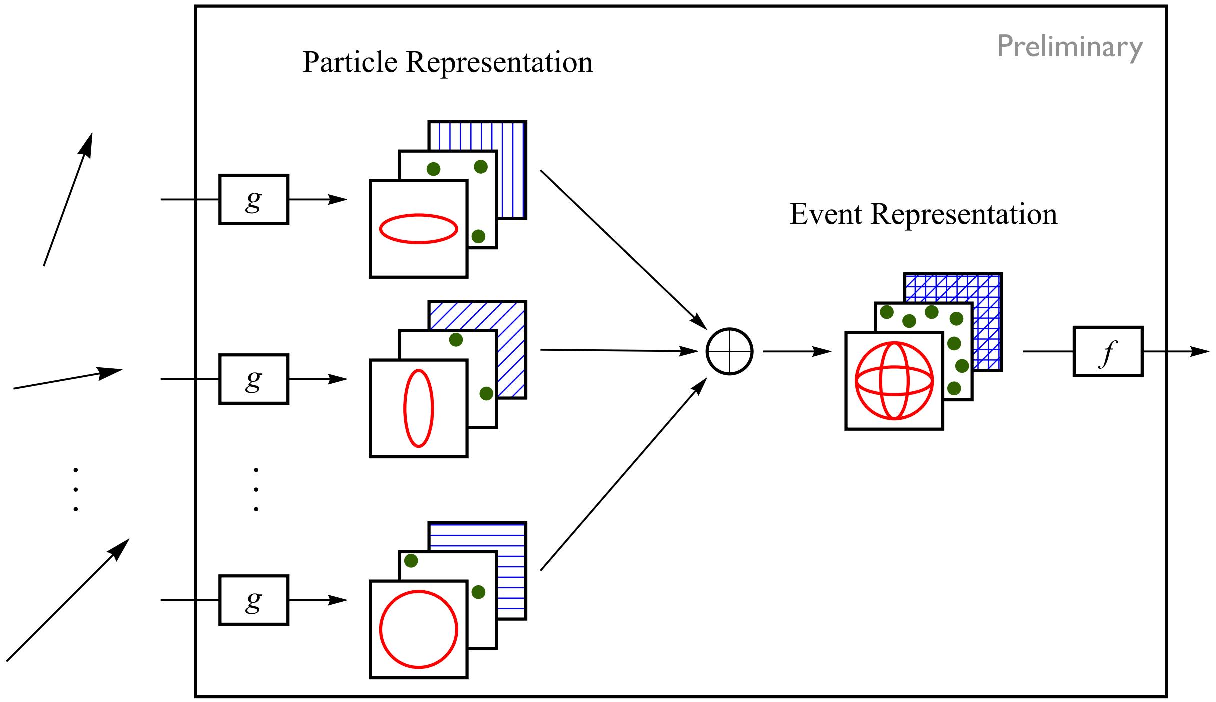
Eric Metodiev

[slogan from Eric Metodiev; quote from Deus ex machina on Wikipedia]

What is the machine learning?

The singularity structure of QCD!

Black box visualization with Energy Flow Networks



$$\underbrace{h}_{\text{EFN}}(\vec{x}) = \underbrace{f}_{\text{DNN}}(\underbrace{\ell_1, \ell_2, \ldots, \ell_n}_{\text{Latent Space}})$$

$$\ell_a = \sum_{i \in \text{jet}} p_{Ti}\, \underbrace{g_a}_{\text{DNN}}(\phi_i, y_i)$$

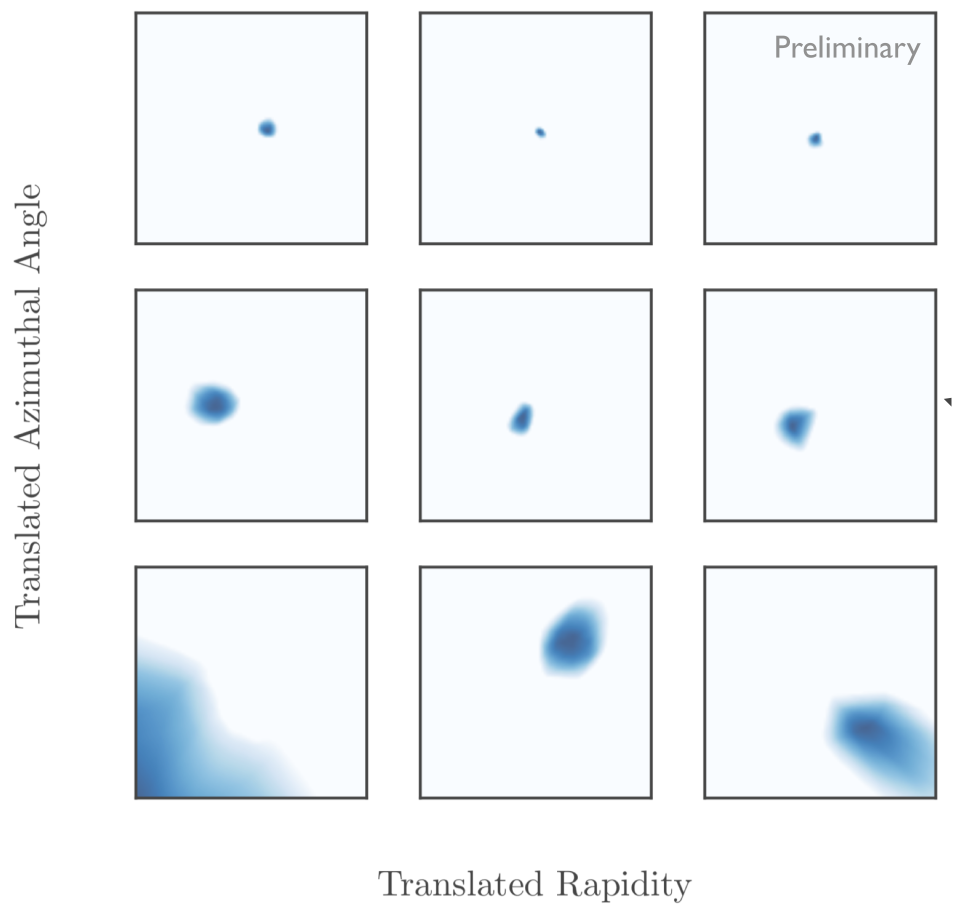
Most general IRC-safe classifier from variable-length, unordered 4-vectors

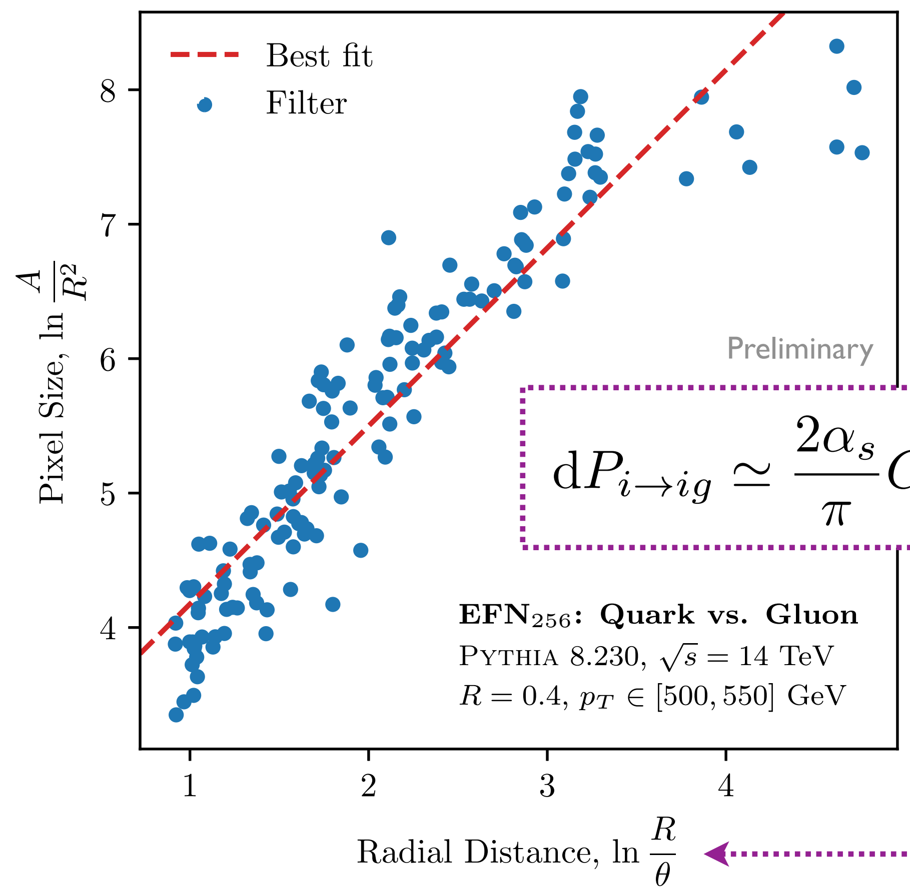
[Komiske, Metodiev, JDT, to appear; see also Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114; alternative perspectives in Chang, Cohen, Ostdiek, 1709.10106; Dreyer, Salam, Soyez, 1807.04758]

What is the machine learning?

The singularity structure of QCD!

Dynamical "pixels" with expected collinear scaling





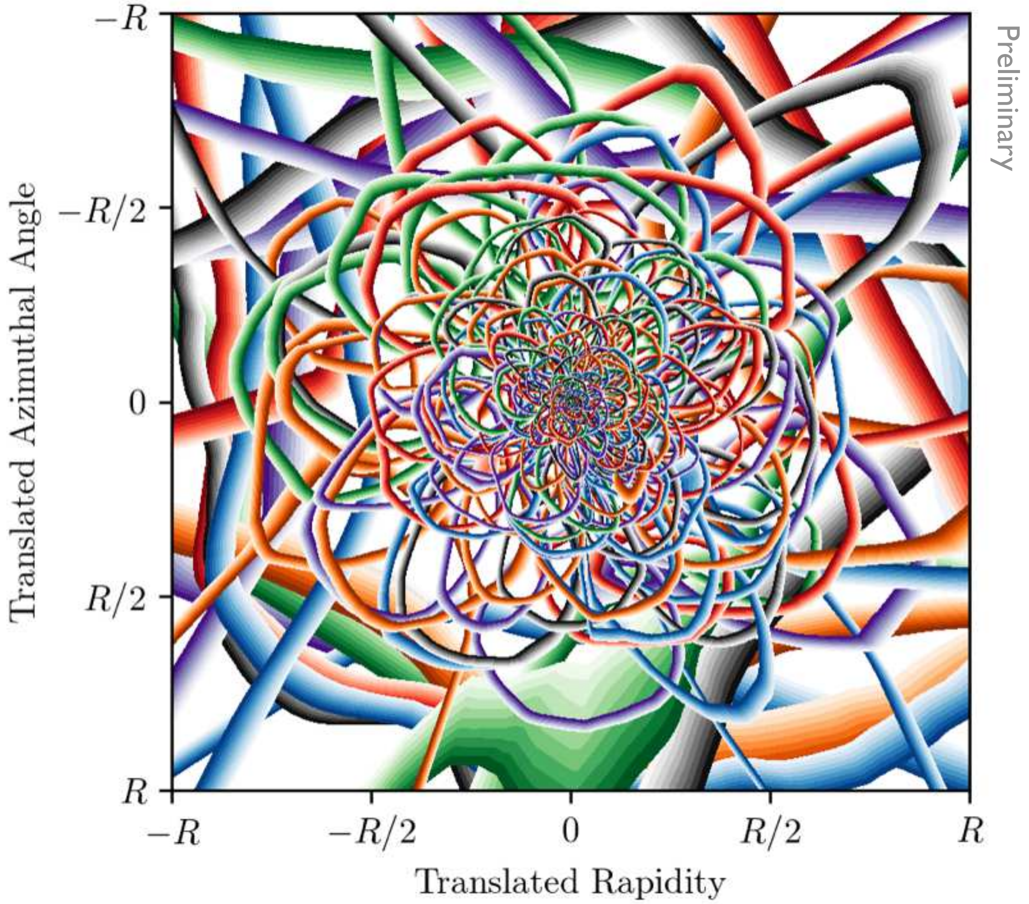
$$dP_{i \to ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{d\theta}{\theta} \frac{dz}{z}$$

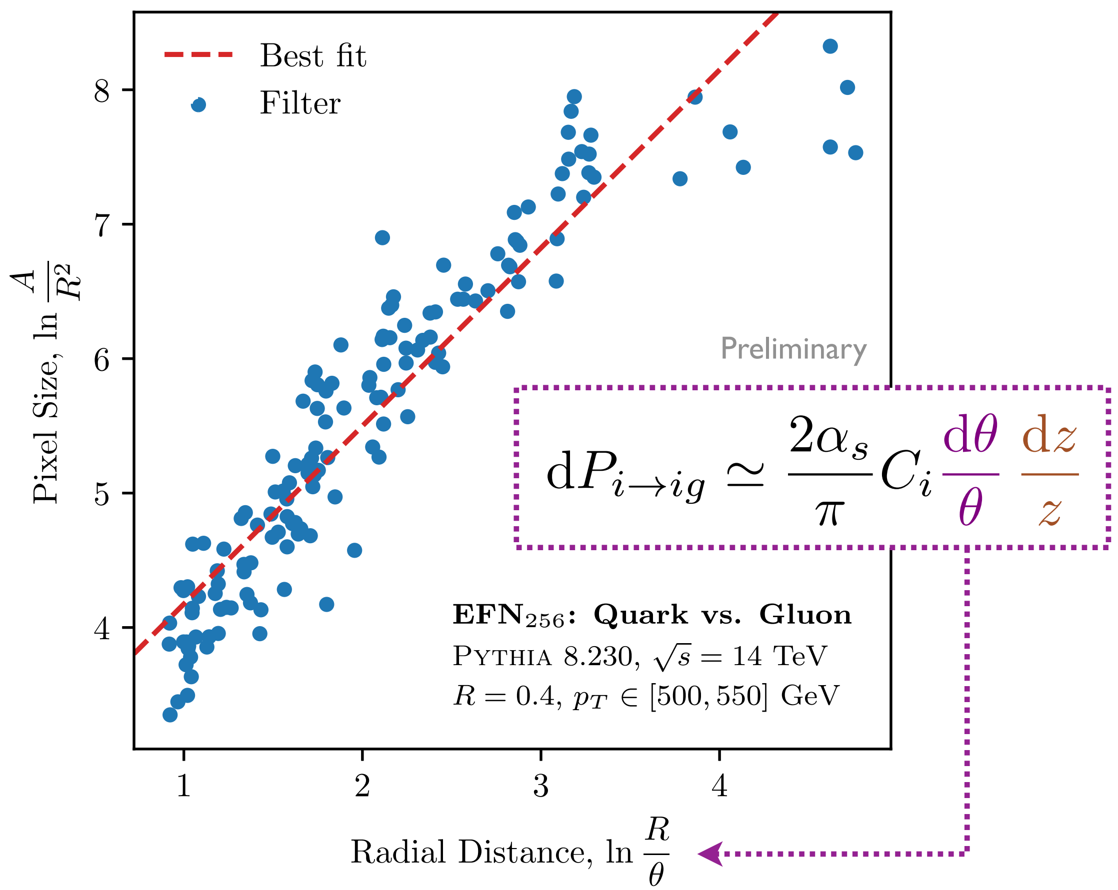
[Komiske, Metodiev, JDT, to appear; see also Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114; alternative perspectives in Chang, Cohen, Ostdiek, 1709.10106; Dreyer, Salam, Soyez, 1807.04758]

What is the machine learning?

The singularity structure of QCD!

Dynamical "pixels" with expected collinear scaling



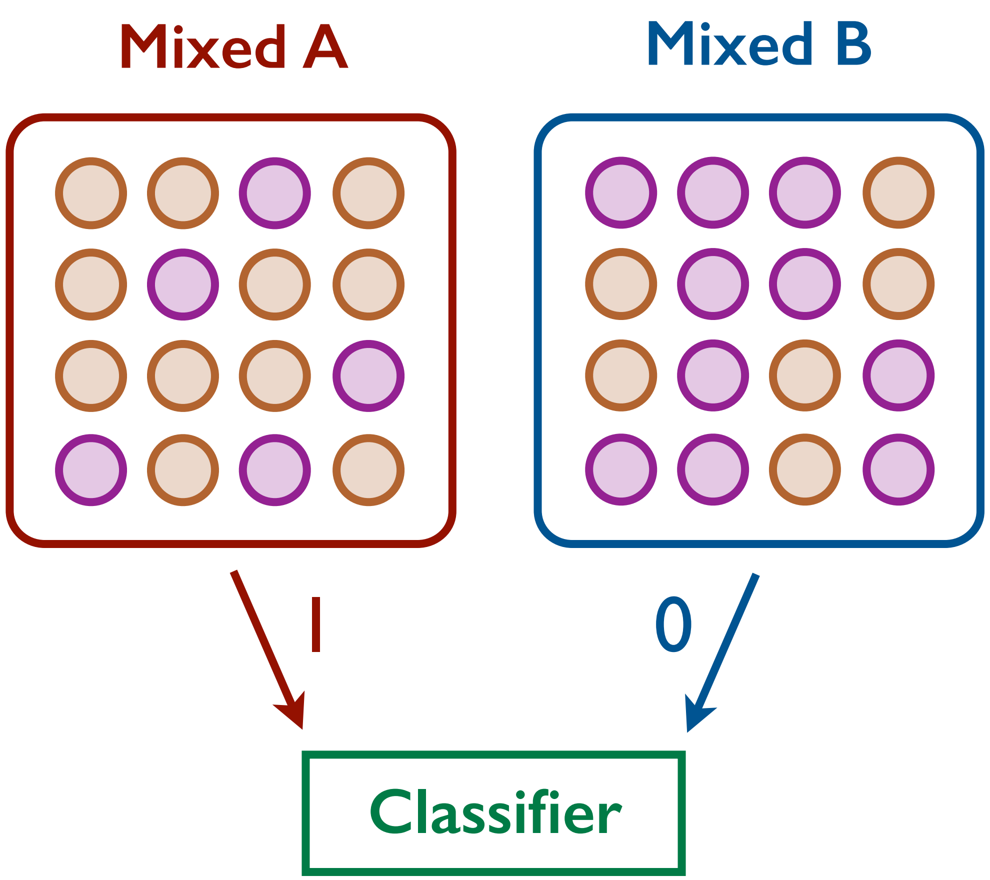


[Komiske, Metodiev, JDT, to appear; see also Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114; alternative perspectives in Chang, Cohen, Ostdiek, 1709.10106; Dreyer, Salam, Soyez, 1807.04758]

Biases from training on simulations?

Train directly on (mixed) data!

$$p_{\mathrm{mixed}}(\vec{x}) = f_q \, p_{\mathrm{quark}}(\vec{x}) + (1 - f_q) \, p_{\mathrm{gluon}}(\vec{x})$$



$$h_{\mathrm{mixed}}(\vec{x}) = \frac{p_A(\vec{x})}{p_A(\vec{x}) + p_B(\vec{x})}$$

$$\neq$$

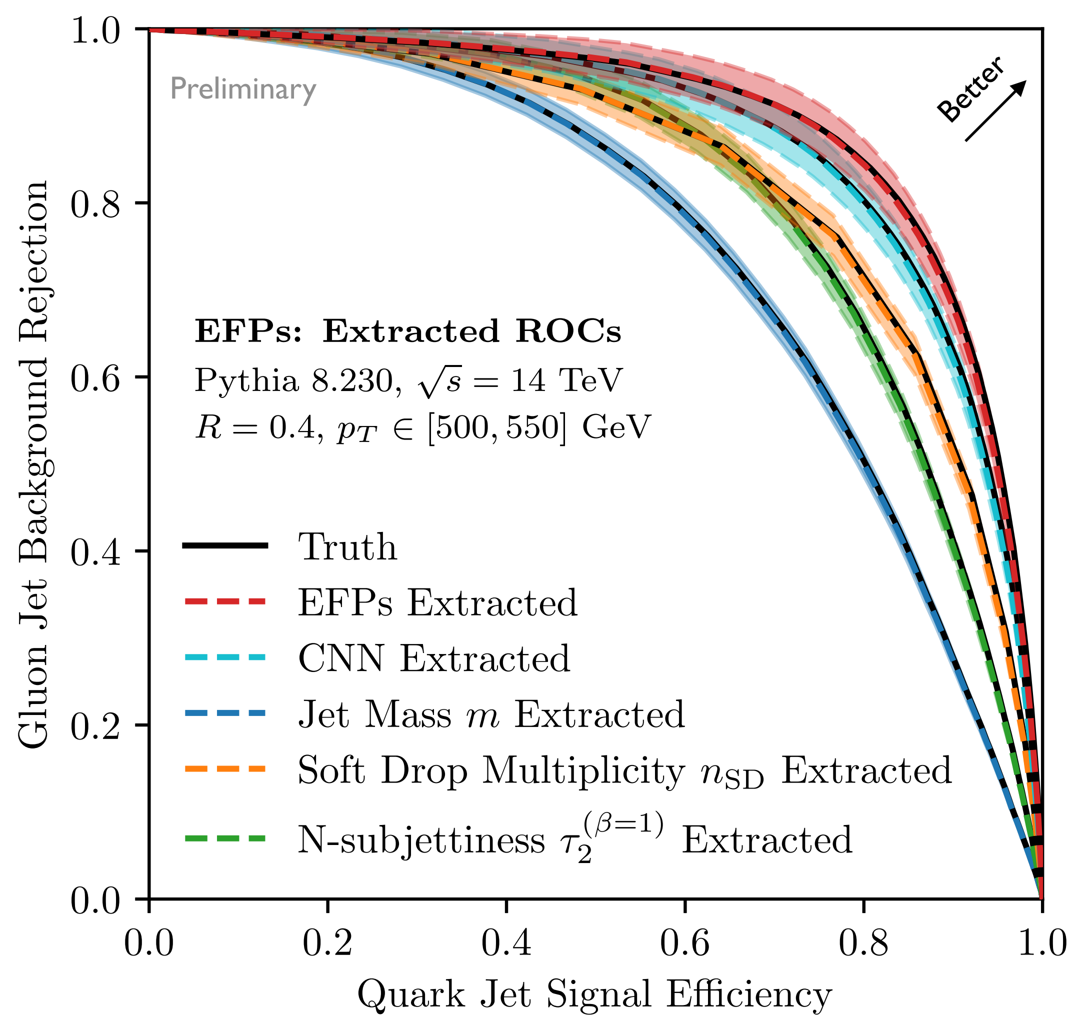
$$h_{\mathrm{pure}}(\vec{x}) = \frac{p_q(\vec{x})}{p_q(\vec{x}) + p_g(\vec{x})}$$

but… $\dfrac{\partial h_{\mathrm{mixed}}(\vec{x})}{\partial h_{\mathrm{pure}}(\vec{x})} > 0$

[Metodiev, Nachman, JDT, 1708.02949;
see also Blanchard, Flaska, Handy, Pozzi, Scott, 1303.1208; Cranmer, Pavez, Louppe, 1506.02169;
Dery, Nachman, Rubbo, Schwartzman, 1702.00414; Cohen, M. Freytsis, and B. Ostdiek, 1706.09451;
Komiske, Metodiev, Nachman, Schwartz, 1801.10158; Collins, Howe, Nachman, 1805.02664]

Biases from training on simulations?

Train directly on (mixed) data!



CWoLa: Classification Without Labels

Robust performance

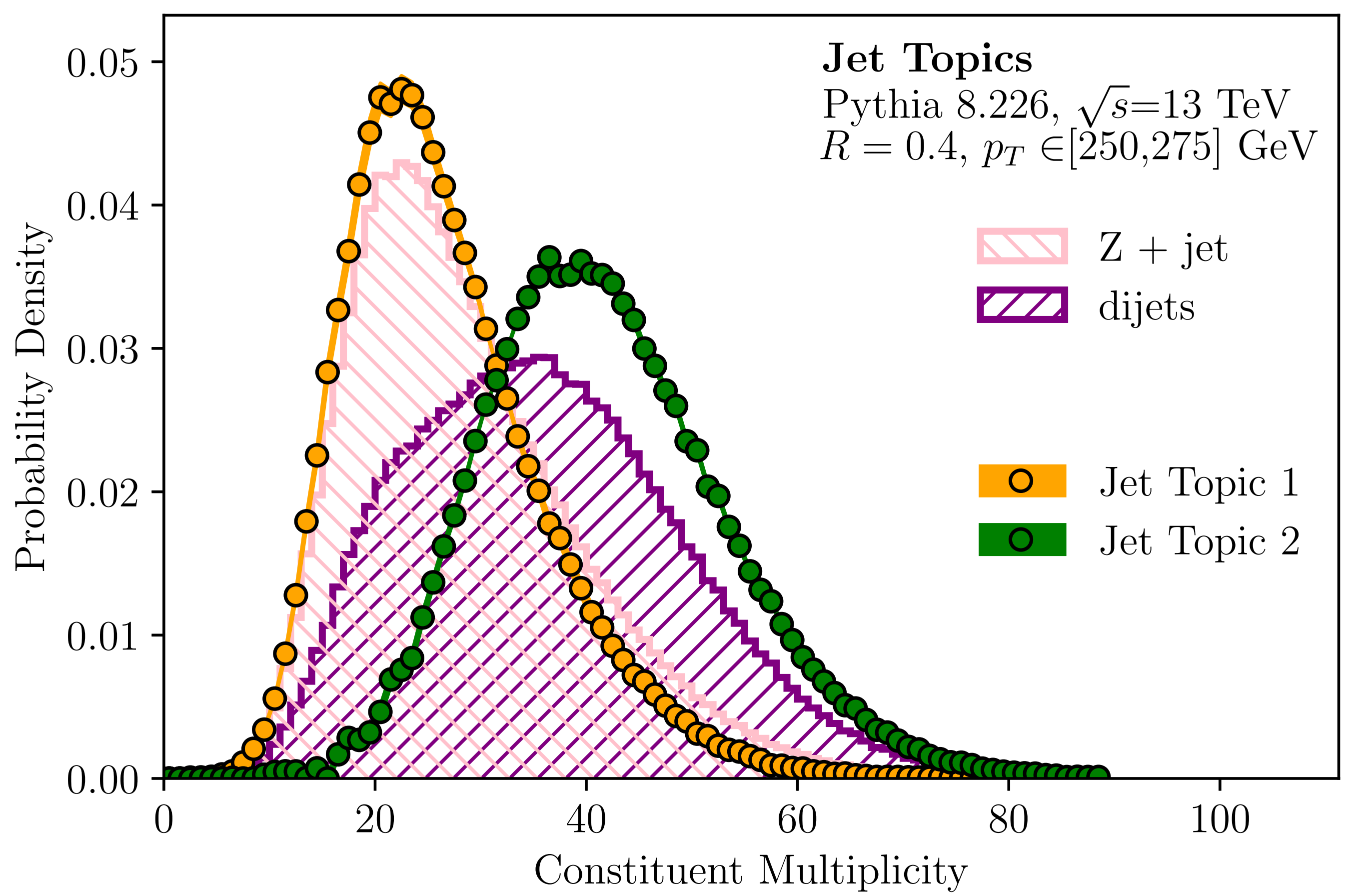
(plus data-driven tricks to calibrate working points and estimate systematic uncertainties)

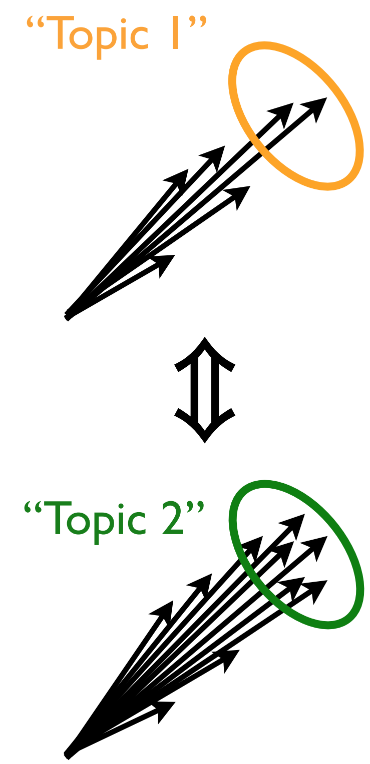
[Komiske, Metodiev, JDT, to appear]

[Metodiev, Nachman, JDT, 1708.02949;
see also Blanchard, Flaska, Handy, Pozzi, Scott, 1303.1208; Cranmer, Pavez, Louppe, 1506.02169;
Dery, Nachman, Rubbo, Schwartzman, 1702.00414; Cohen, M. Freytsis, and B. Ostdiek, 1706.09451;
Komiske, Metodiev, Nachman, Schwartz, 1801.10158; Collins, Howe, Nachman, 1805.02664]

Do jet categories even make sense?

Use classifiers to define categories!





Extract from data, solely* from the assumption they exist

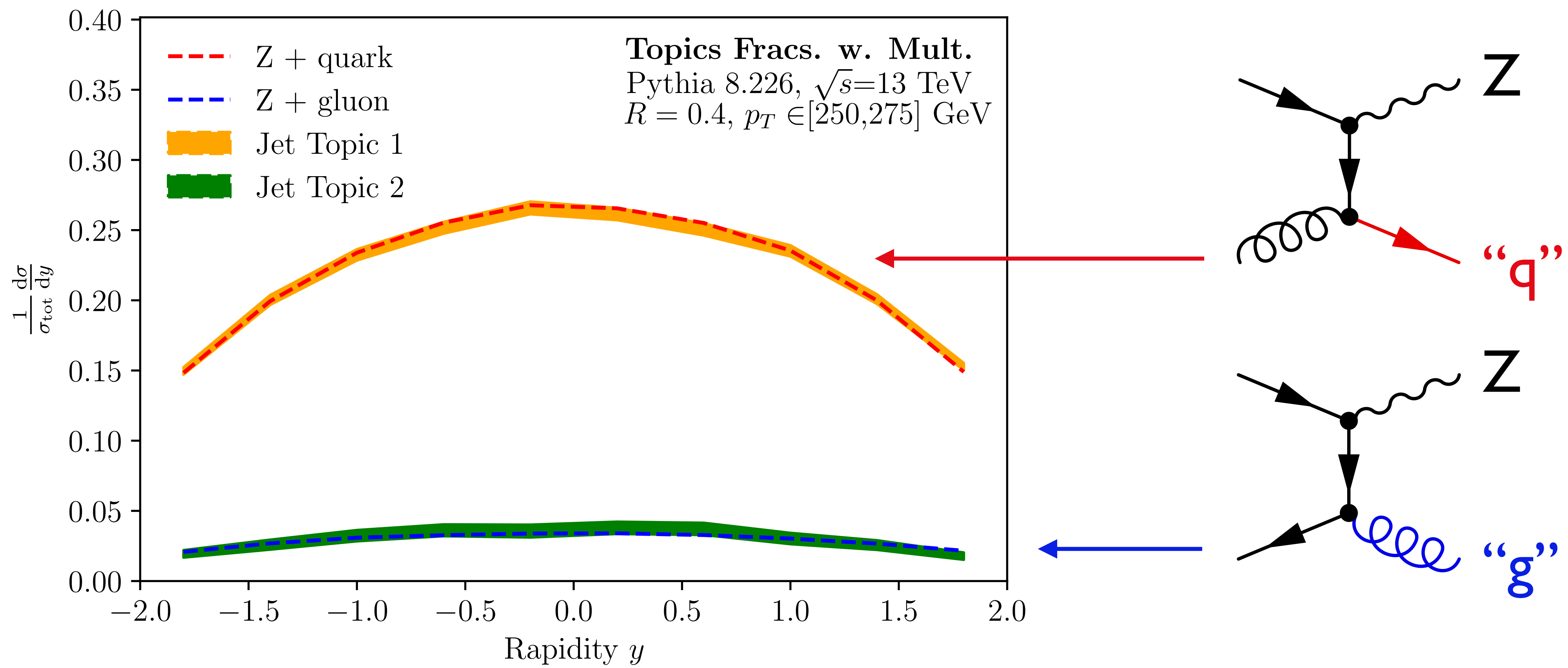
Sample Independence, Different Fractions, Mutual Irreducibility

[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, to appear; using Katz-Samuels, Blanchard, Scott, 1710.01167; alternative perspectives in Gras, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, 1704.03878]

Do jet categories even make sense?

Use classifiers to define categories!

Jet Topics: Data-driven, "parton"-labeled cross sections



[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, to appear; using Katz-Samuels, Blanchard, Scott, 1710.01167; alternative perspectives in Gras, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, 1704.03878]

The Broader Lesson

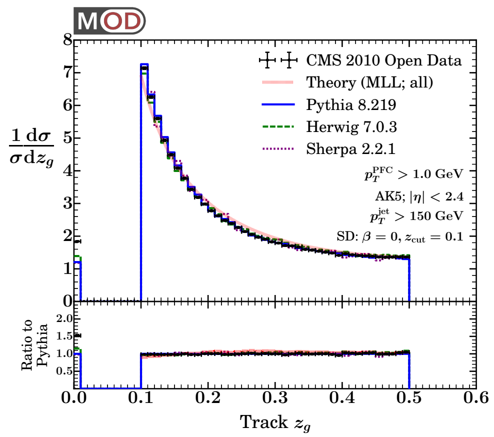
&

"Deep Learning" ~~vs.~~ "Deep Thinking"

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

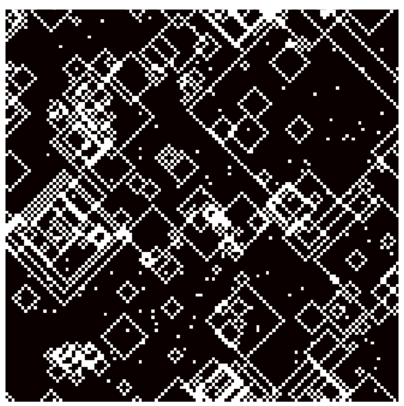
Desired Outcomes $\Leftrightarrow$ Algorithms/Observables

Summary



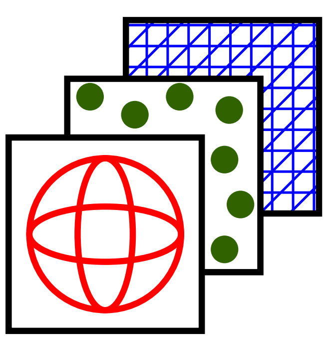
Three Trends in Jet Physics

Extreme kinematics, precision measurements/calculations, and…



Into the Network

…the rise of machine learning



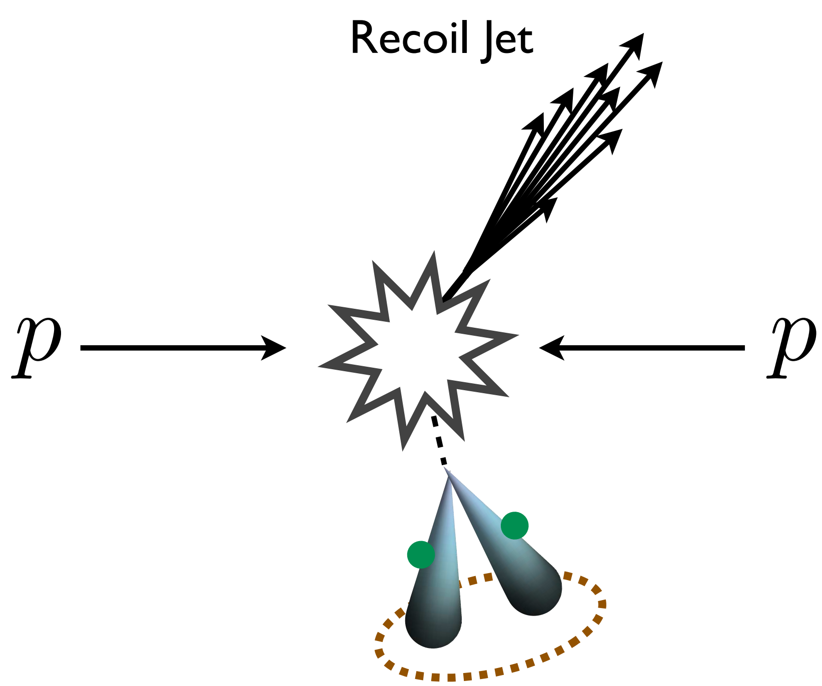
Data ex Machina

Robust data-driven methods for jet physics and beyond

Backup Slides

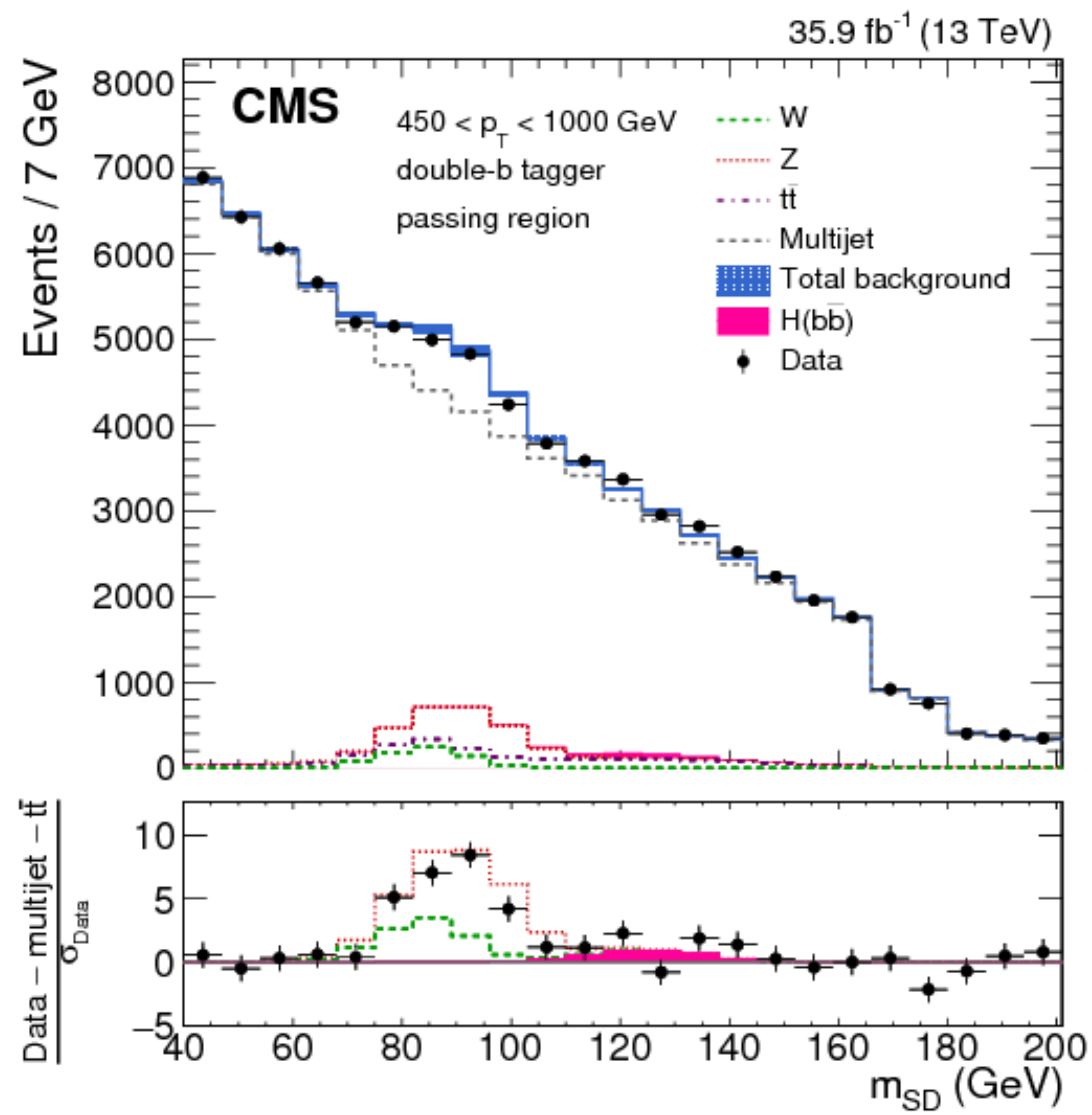
The Rise of Extreme Kinematics

Boosted Higgs



High p_T Higgs

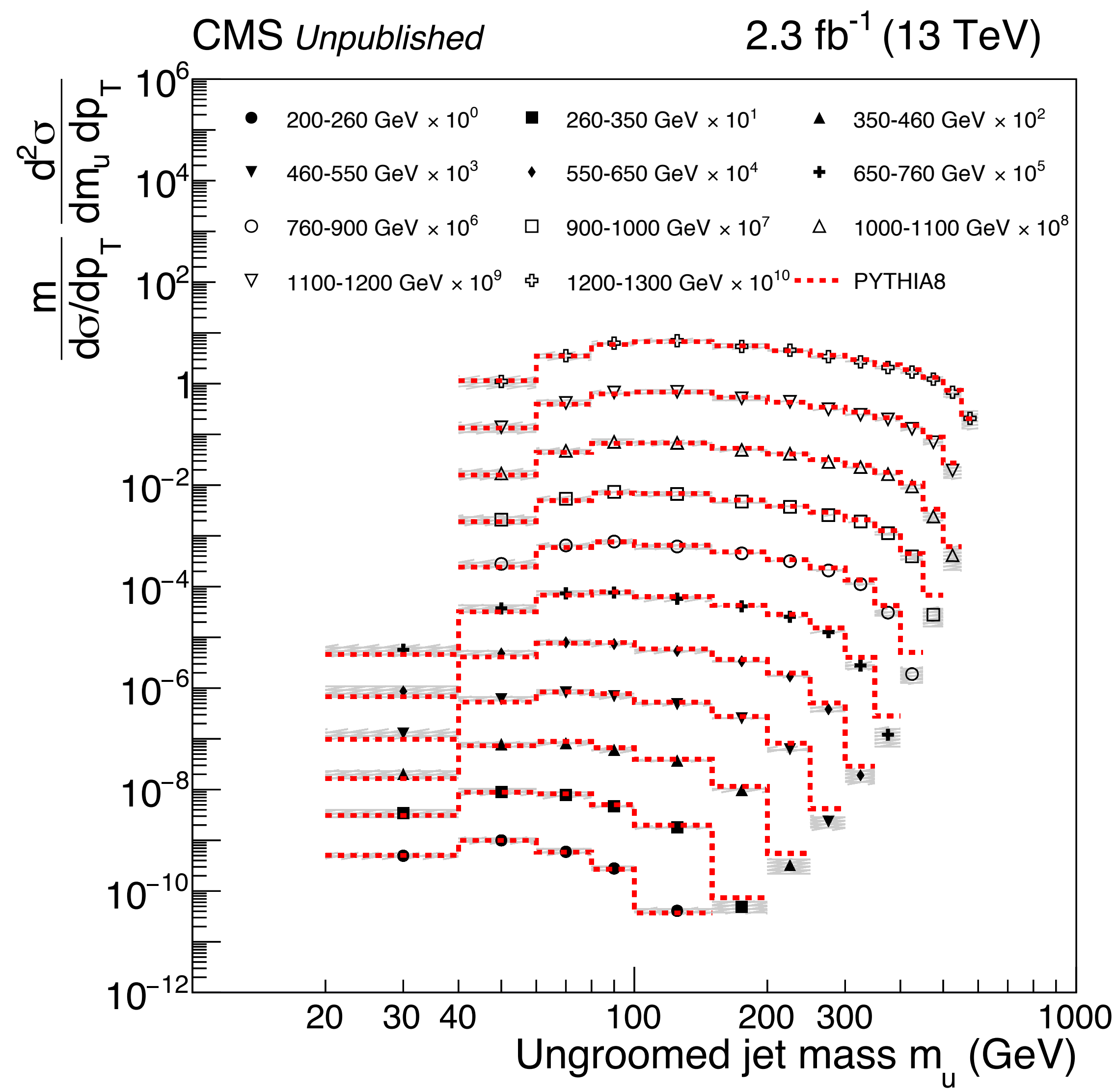
with PUPPI
+ *Soft Drop*
+ *N_2*
+ *B-tagging*
+ *DDT*

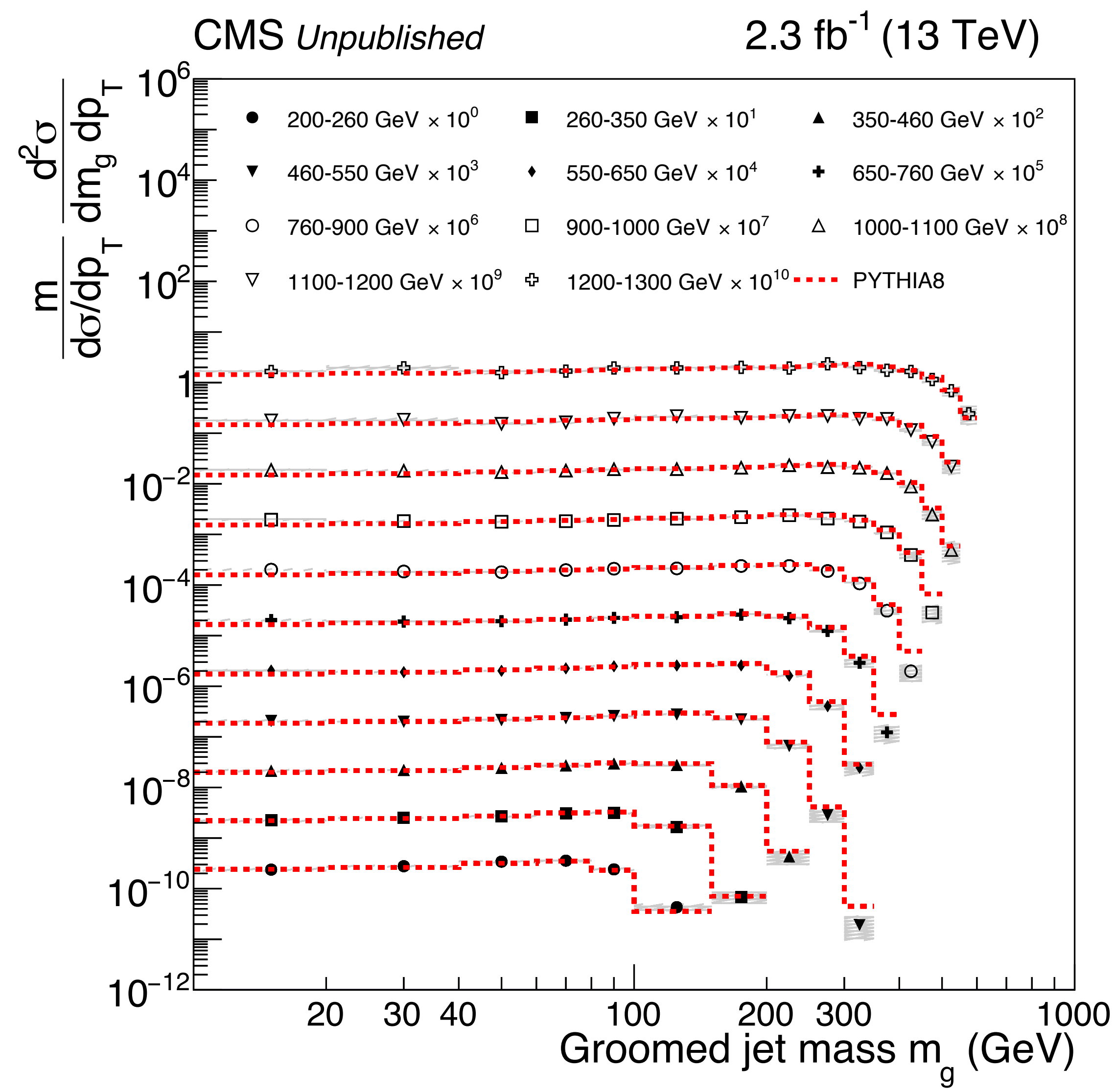


[CMS, 2017; using Bertolini, Harris, Low, Tran, 2014; Larkoski, Marzani, Soyez, JDT, 2014; Moult, Necib, JDT, 2016; CMS, 2015; Dolen, Harris, Marzani, Rappoccio, Tran, 2016]

The Rise of Precision Jet Physics

Groomed Jet Mass from CMS





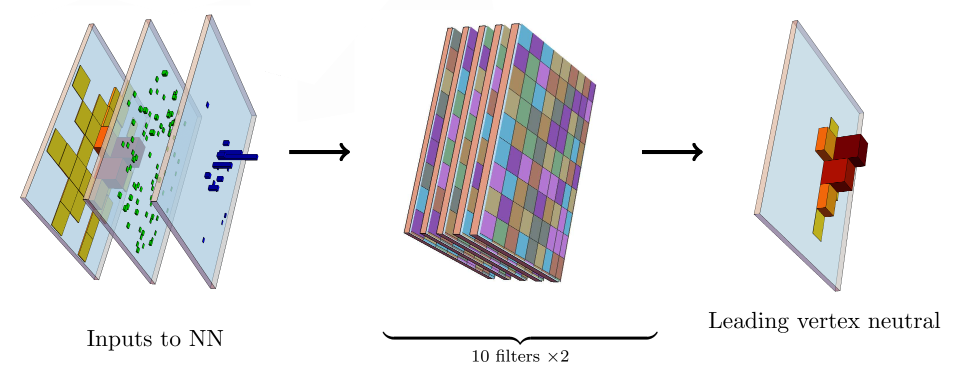
[more intuitive versions of the plots in CMS, 1807.05974]

Beyond Classification

PUMML

Pileup Mitigation

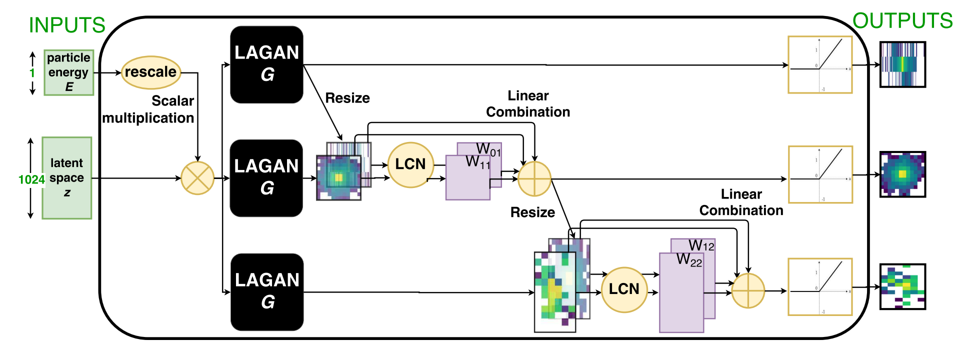
[Komiske, Metodiev, Nachman, Schwartz, 1707.08600]



CaloGAN

Fast Detector Simulation

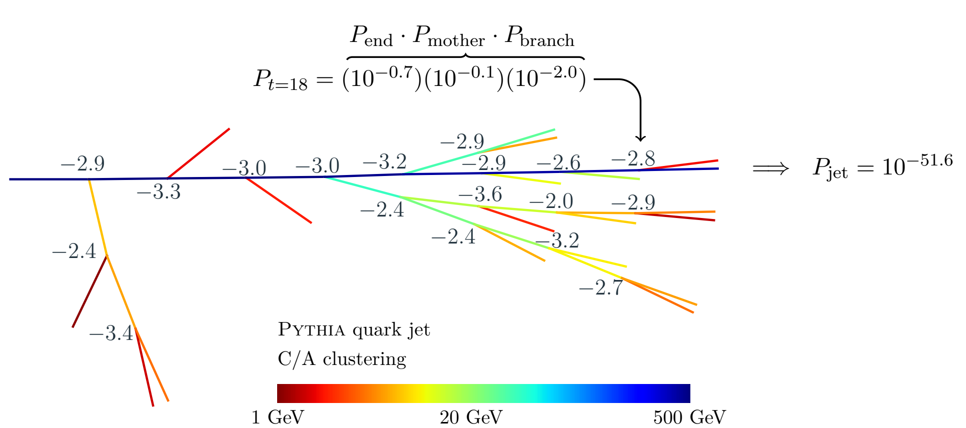
Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321;
see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



JUNIPR

Probability Modeling

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



The Energy Flow Polynomials

Multigraph

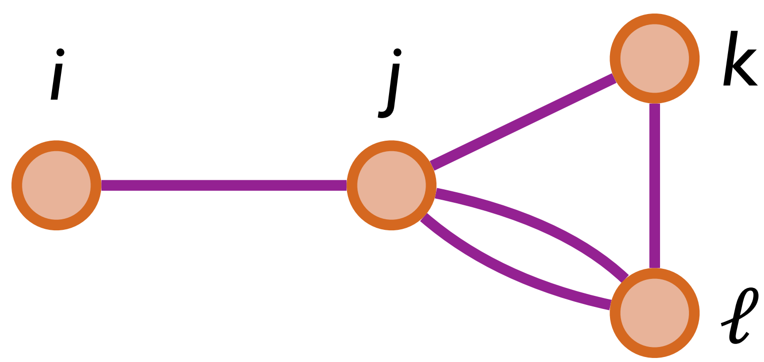
Angular Scaling

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

All N-tuples

N Energy Fractions

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]

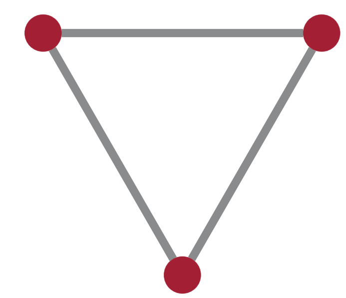
Down the Rabbit Hole

Known Structures:

${}_1e_2$

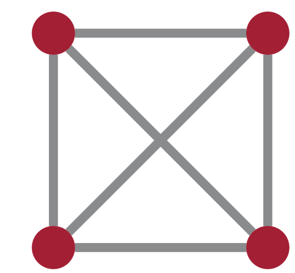
width (β=1), thrust (β=2)

${}_3e_3$



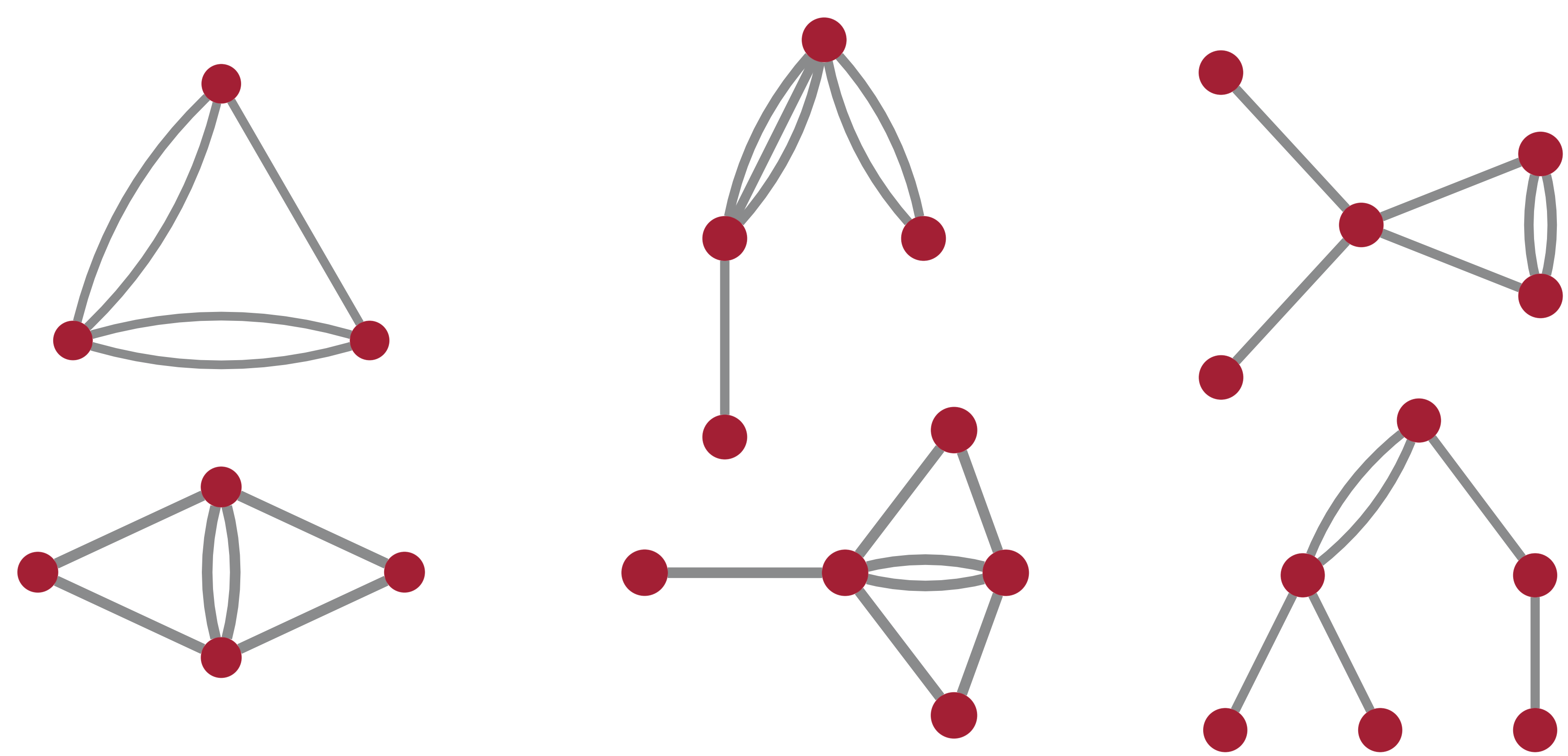
used in D_2

${}_6e_4$



used in C_3

No Idea:



The Energy Flow Moments

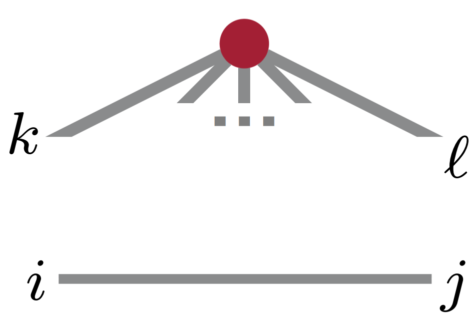
$$\mathcal{I}^{\mu_1\mu_2\cdots\mu_v} = \sum_{i=1}^{M} E_i\ \hat{p}^{\mu_1}\ \hat{p}^{\mu_2} \cdots \hat{p}^{\mu_v}$$

Particle Relabeling

Infrared Safety

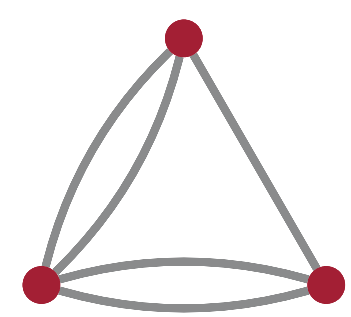
Special Choice of Angle $\theta_{ij} = 2\,\eta_{\mu\nu}\,\hat{p}_i^{\mu}\,\hat{p}_j^{\nu}$

EFMs $\Rightarrow$ EFPs



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

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

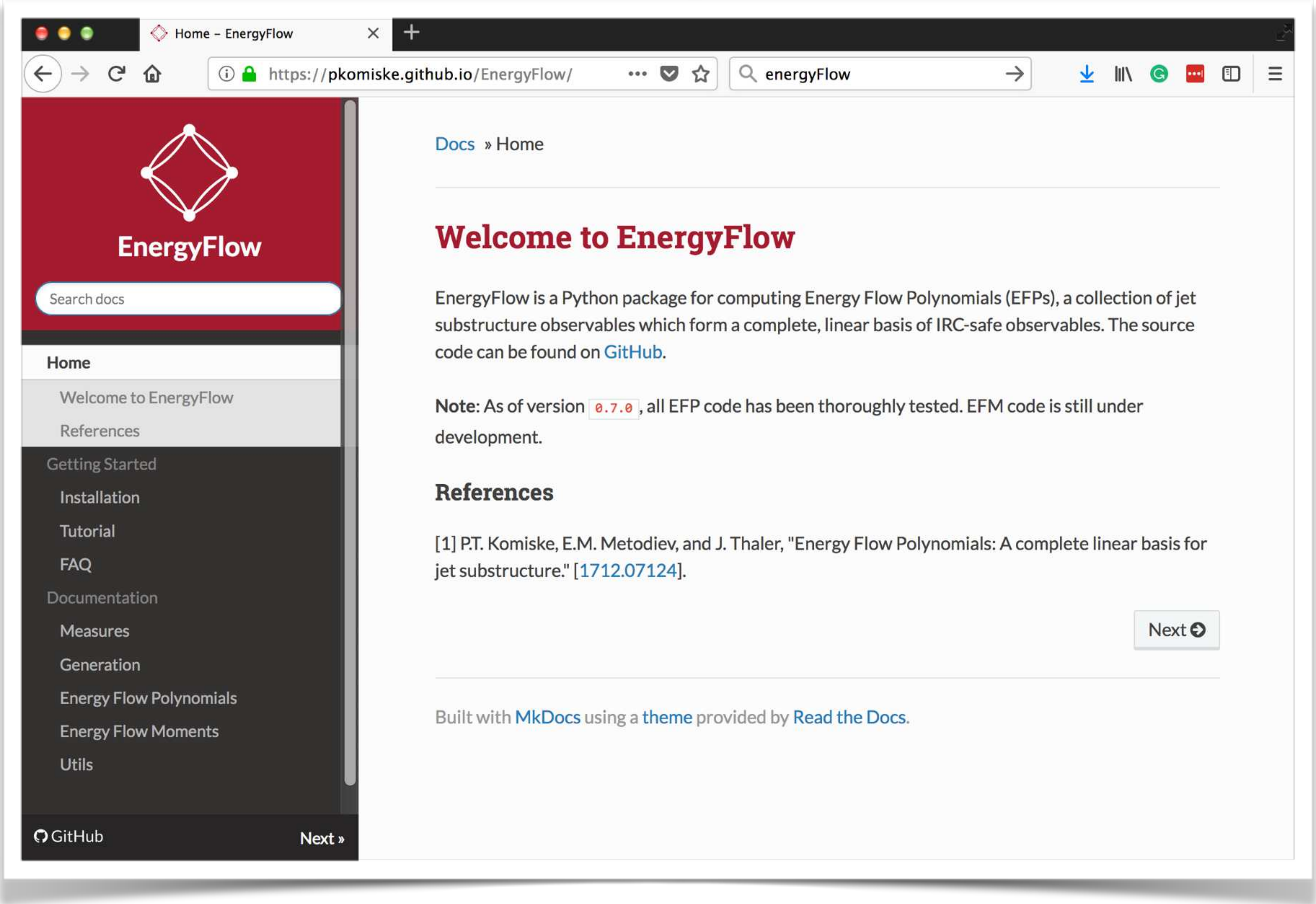


$= \mathcal{I}^{\mu\nu\rho\sigma}\mathcal{I}^{\tau}{}_{\mu\nu}\mathcal{I}_{\rho\sigma\tau}$

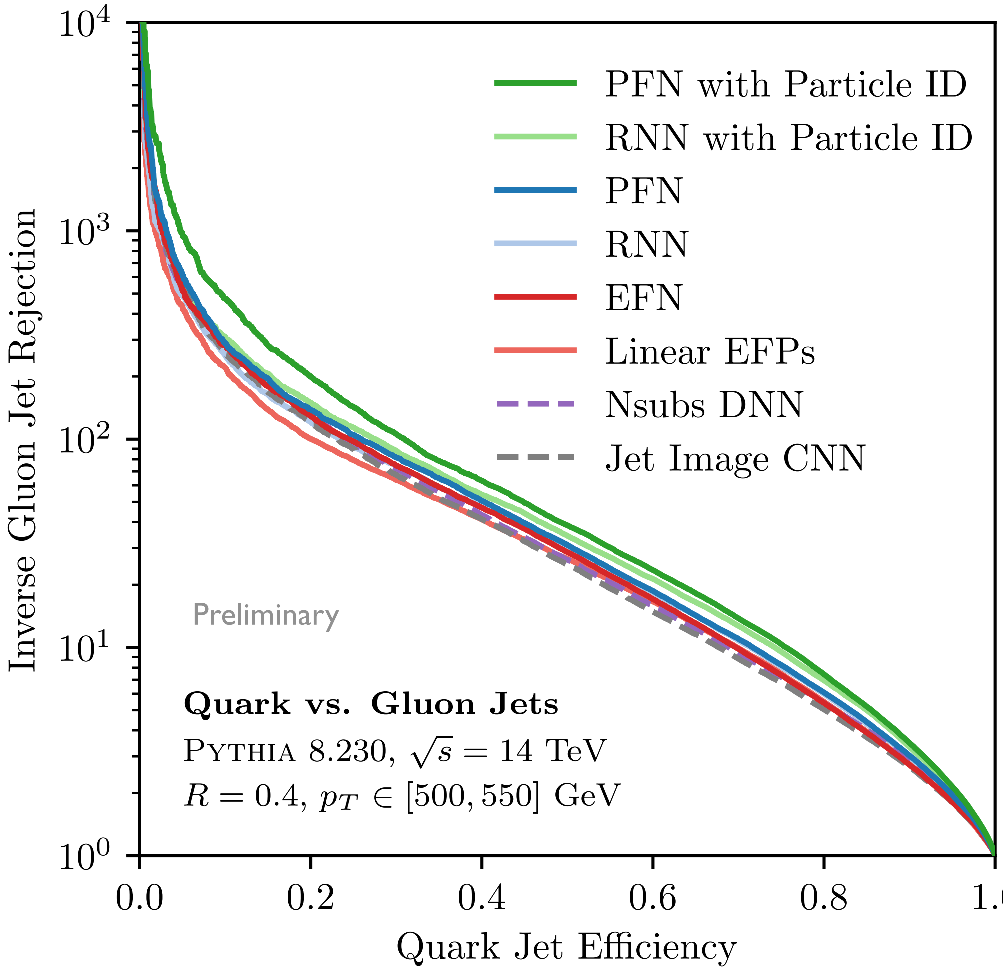
Compute in O(M)!

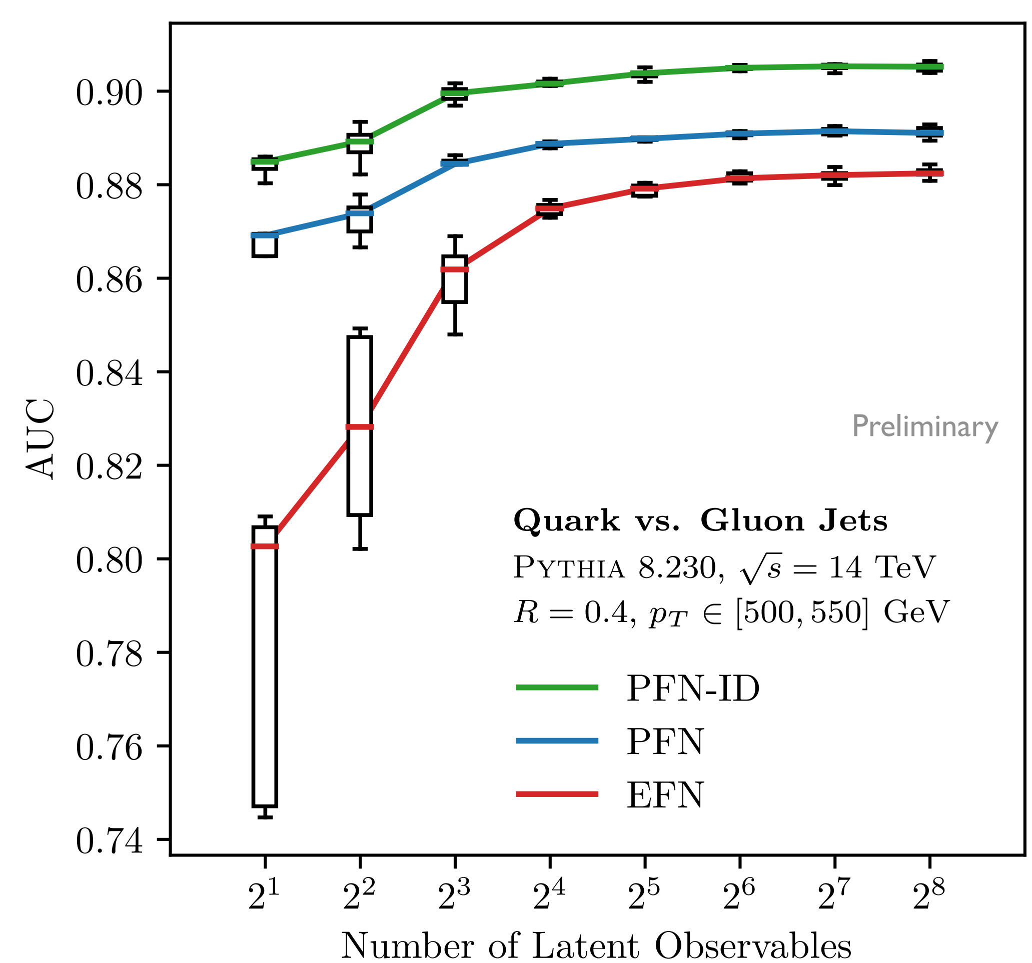
Also compute in O(M)!

[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.

Performance of Energy/Particle Flow Networks





[Komiske, Metodiev, JDT, to appear]

Well-Defined Categories?

Quark (color triplet) vs. Gluon (color octet)?
But jet constituents are color-singlet hadrons!

What is a Quark Jet?
From lunch/dinner discussions

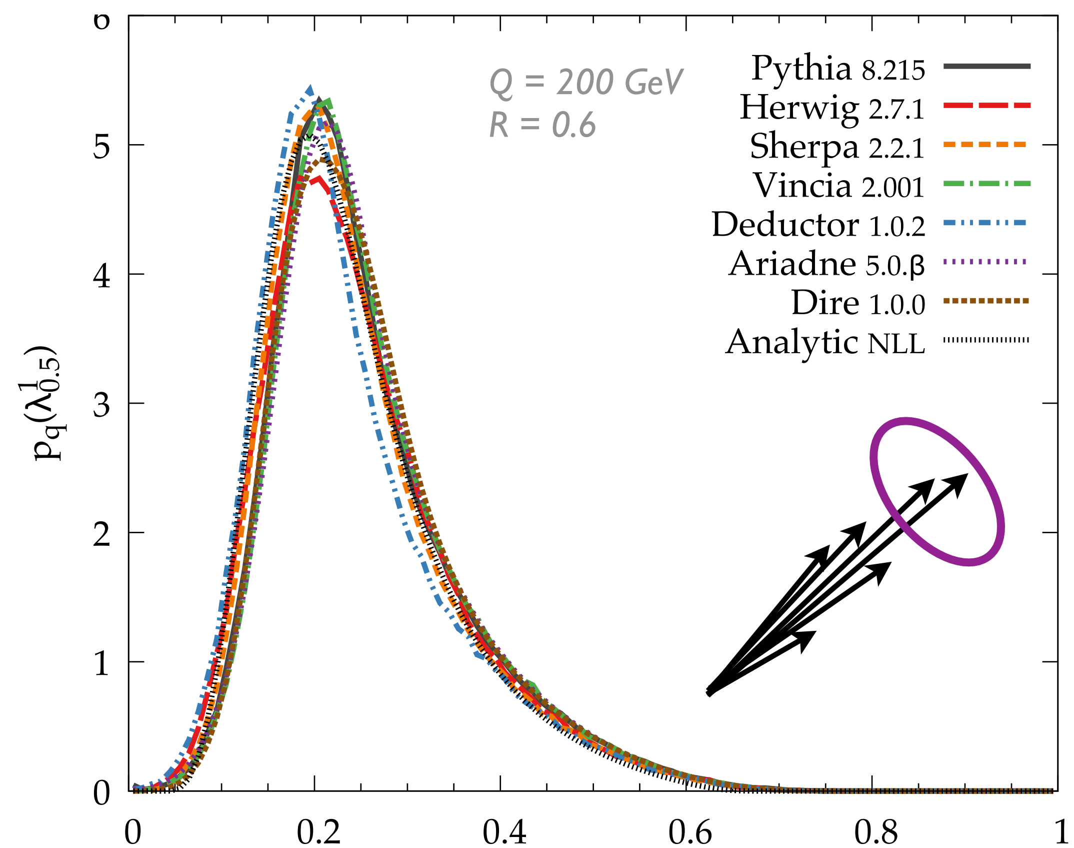
Ill-Defined ↓ Well-Defined	What people sometimes think we mean ↑ What we mean	
		A quark parton
		A Born-level quark parton
		The initiating quark parton in a final state shower
Quark as noun		An eikonal line with baryon number 1/3 and carrying triplet color charge
		A quark operator appearing in a hard matrix element in the context of a factorization theorem
		A parton-level jet object that has been quark-tagged using a soft-safe flavored jet algorithm (automatically collinear safe if you sum constituent flavors)
Quark as adjective		A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

Jesse Thaler — Report of the Les Houches Quark/Gluon Subgroup 3

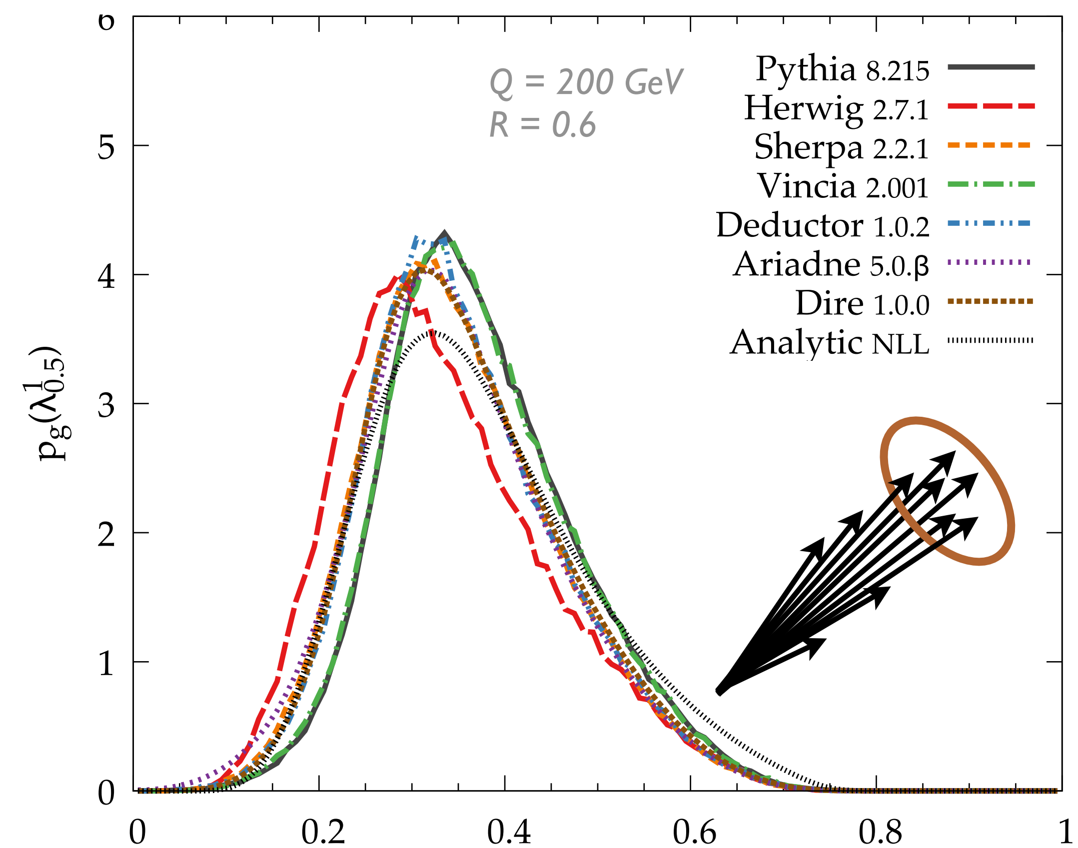
[Gras, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, 1704.03878;
based on Soyez, JDT, Freytsis, Gras, Kar, Lönnblad, Plätzer, Siodmok, Skands, Soper, 1605.04692]

Trustable Training Samples?

$e^+e^- \to$ quarks ($C_F = 4/3$) *vs.* $e^+e^- \to$ gluons ($C_A = 3$)



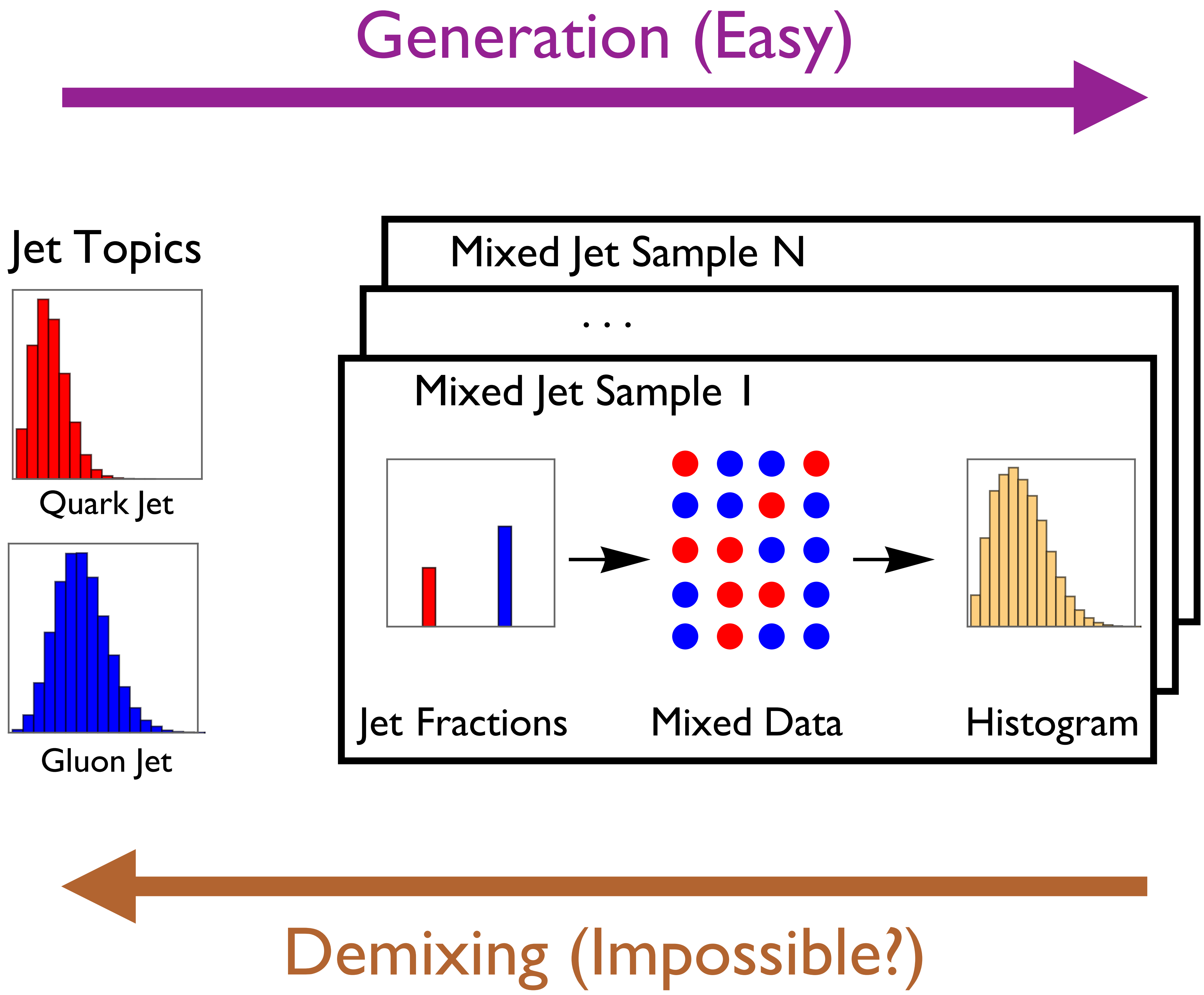
$$\mathrm{LHA} = \sum_i z_i \sqrt{\theta_i}$$



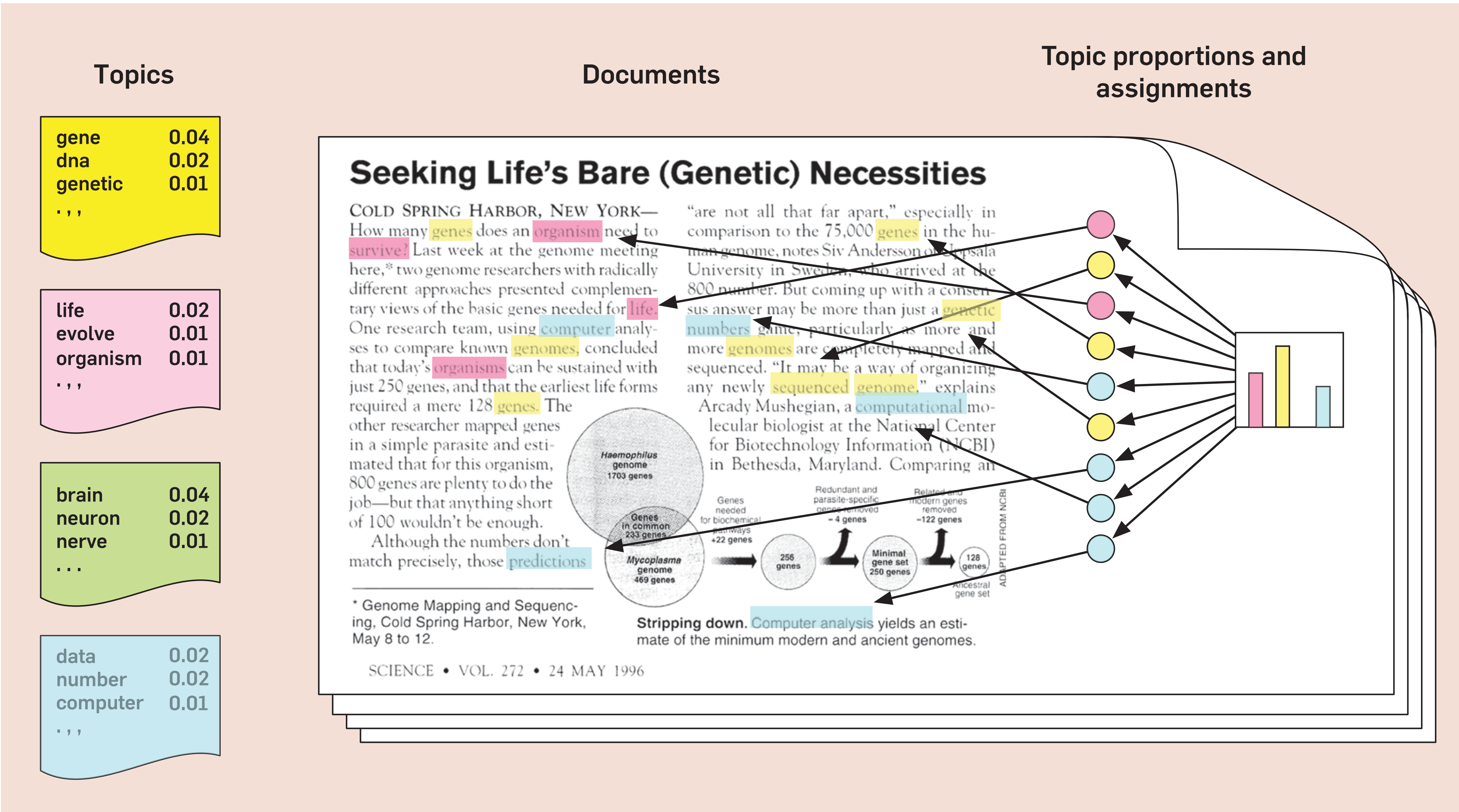
$$\mathrm{LHA} = \sum_i z_i \sqrt{\theta_i}$$

Large shower variations (esp. gluon jets, hard to tune from LEP)

[Gras, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, 1704.03878;
based on Soyez, JDT, Freytsis, Gras, Kar, Lönnblad, Plätzer, Siodmok, Skands, Soper, 1605.04692]

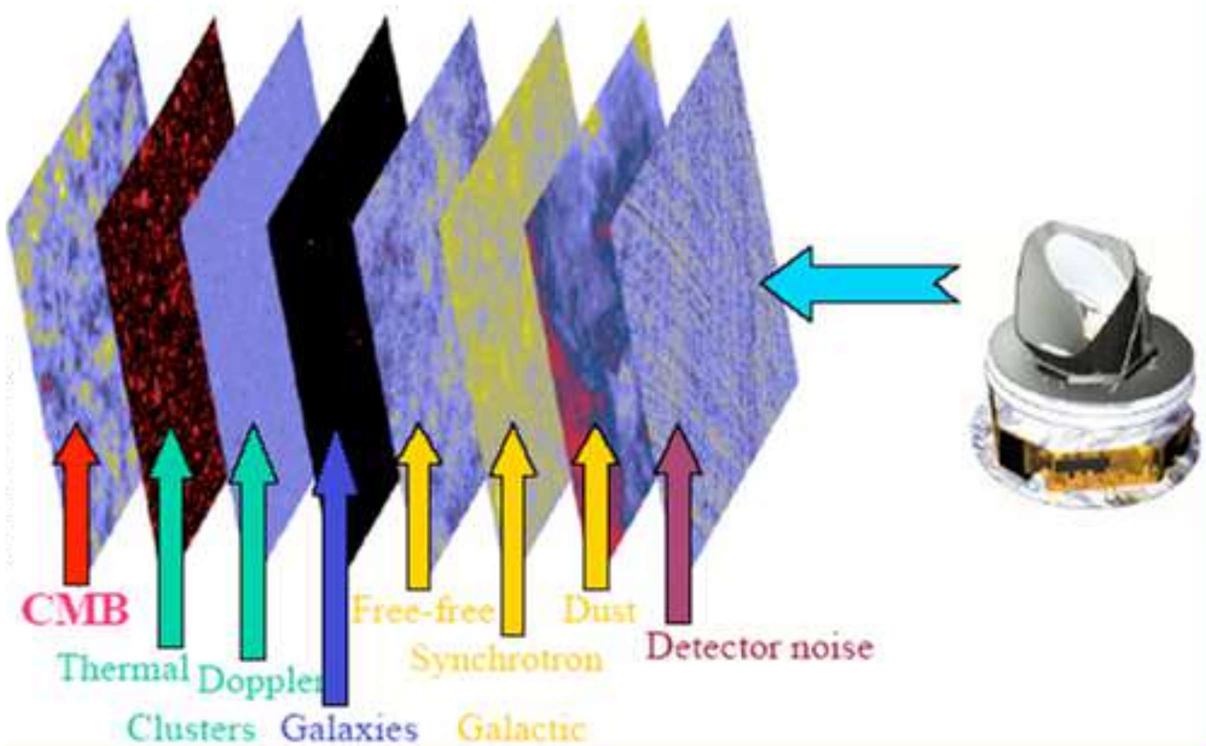

Generation (Easy)
Jet Topics
Quark Jet
Gluon Jet
Mixed Jet Sample N
. . .
Mixed Jet Sample 1
Jet Fractions
Mixed Data
Histogram
Demixing (Impossible?)

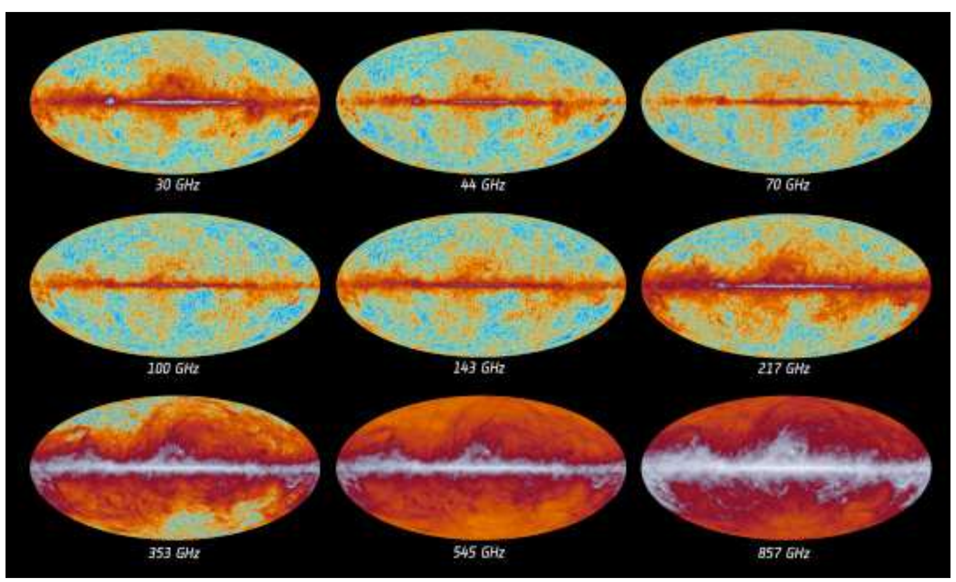
Topic Modeling



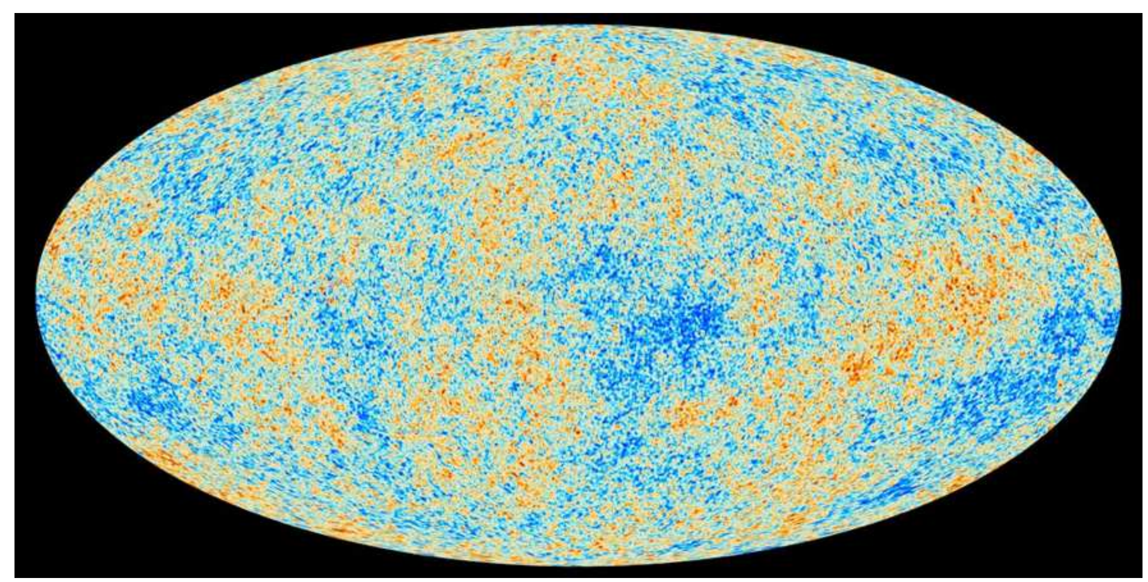
[Blei, 2012]

Related to CMB Foreground Separation





$\Rightarrow$



[Planck Outreach]

The Demix Algorithm

Simplifying to two mixtures of two topics

Just subtract the mixed distributions!

$$p_{\mathrm{T}1}(\vec{x}) = \frac{p_A(\vec{x}) - p_B(\vec{x})\,\kappa_{A|B}}{1 - \kappa_{A|B}}$$

$$p_{\mathrm{T}2}(\vec{x}) = \frac{p_B(\vec{x}) - p_A(\vec{x})\,\kappa_{B|A}}{1 - \kappa_{B|A}}$$

Reducibility Factors

Requires Anchor Bins / "Mutual Irreducibility"

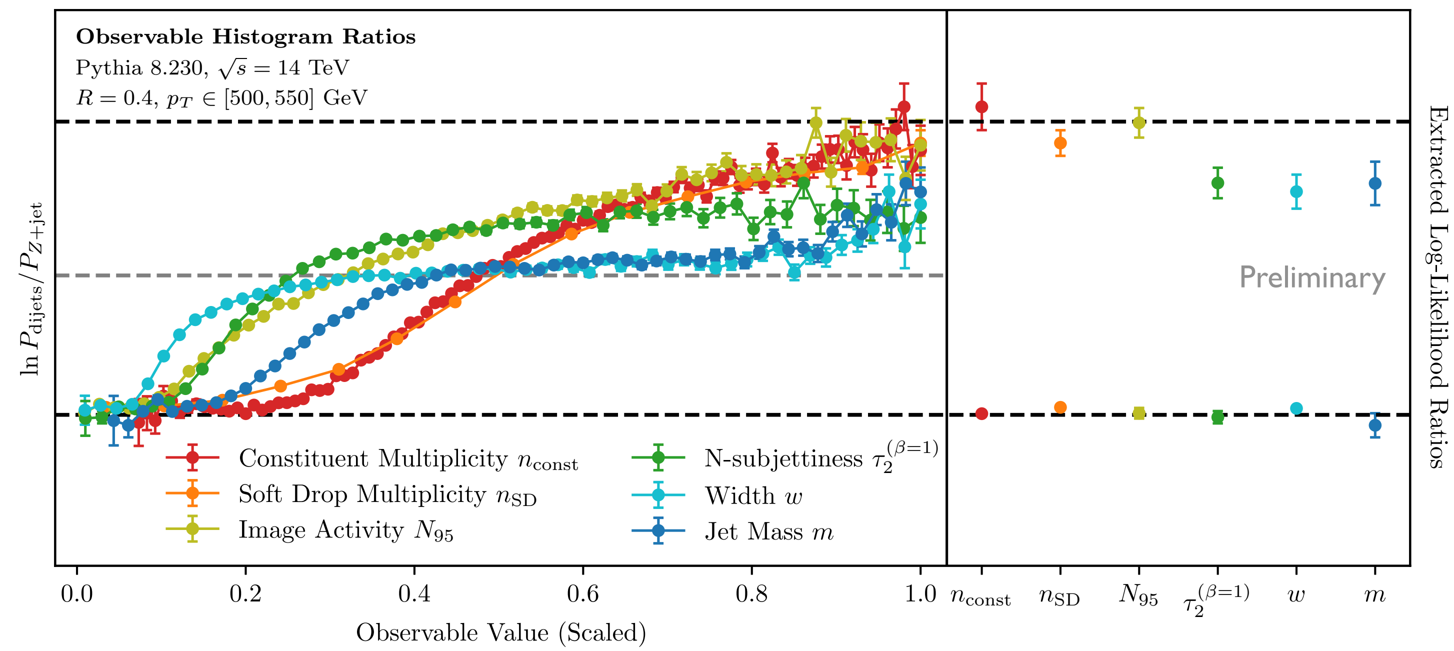
Region of 100% purity for each topic (even if tiny efficiency)
Probabilities are positive, so make κ as large as possible

[Katz-Samuels, Blanchard, Scott, 1710.01167]

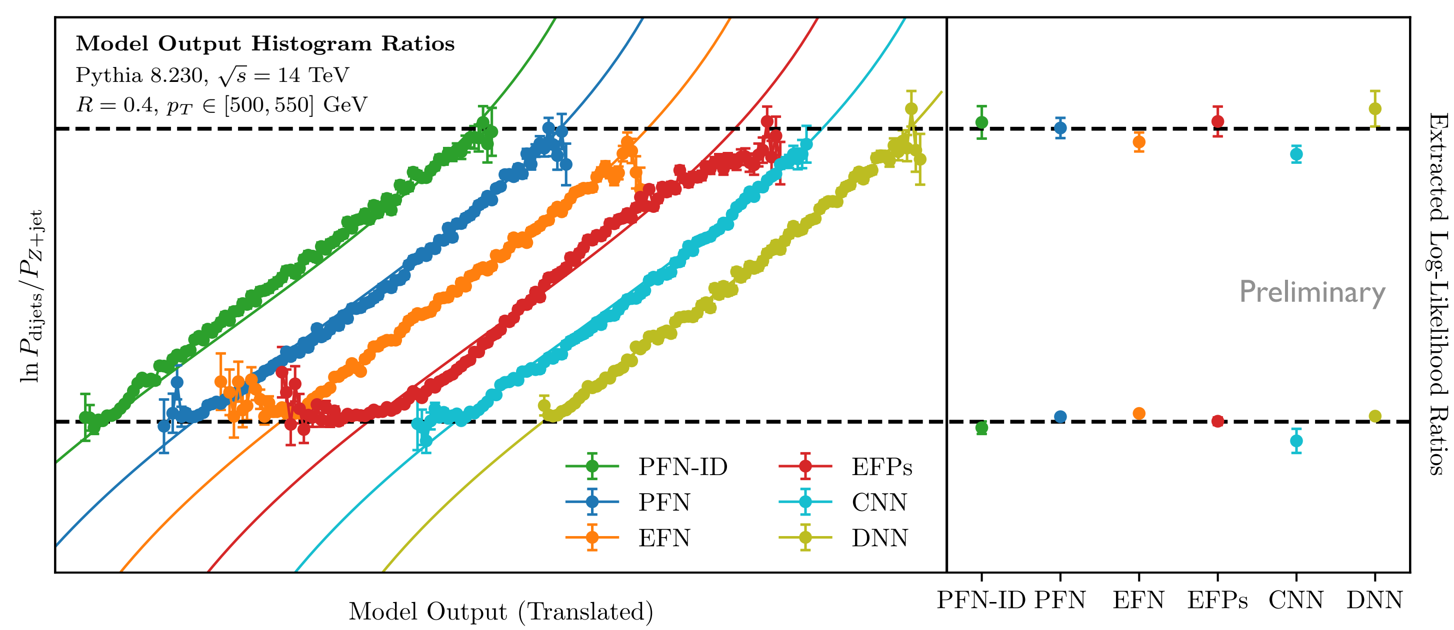
Extracting Quark/Gluon Distributions

Identifying anchor bins

Classic Jet Observables

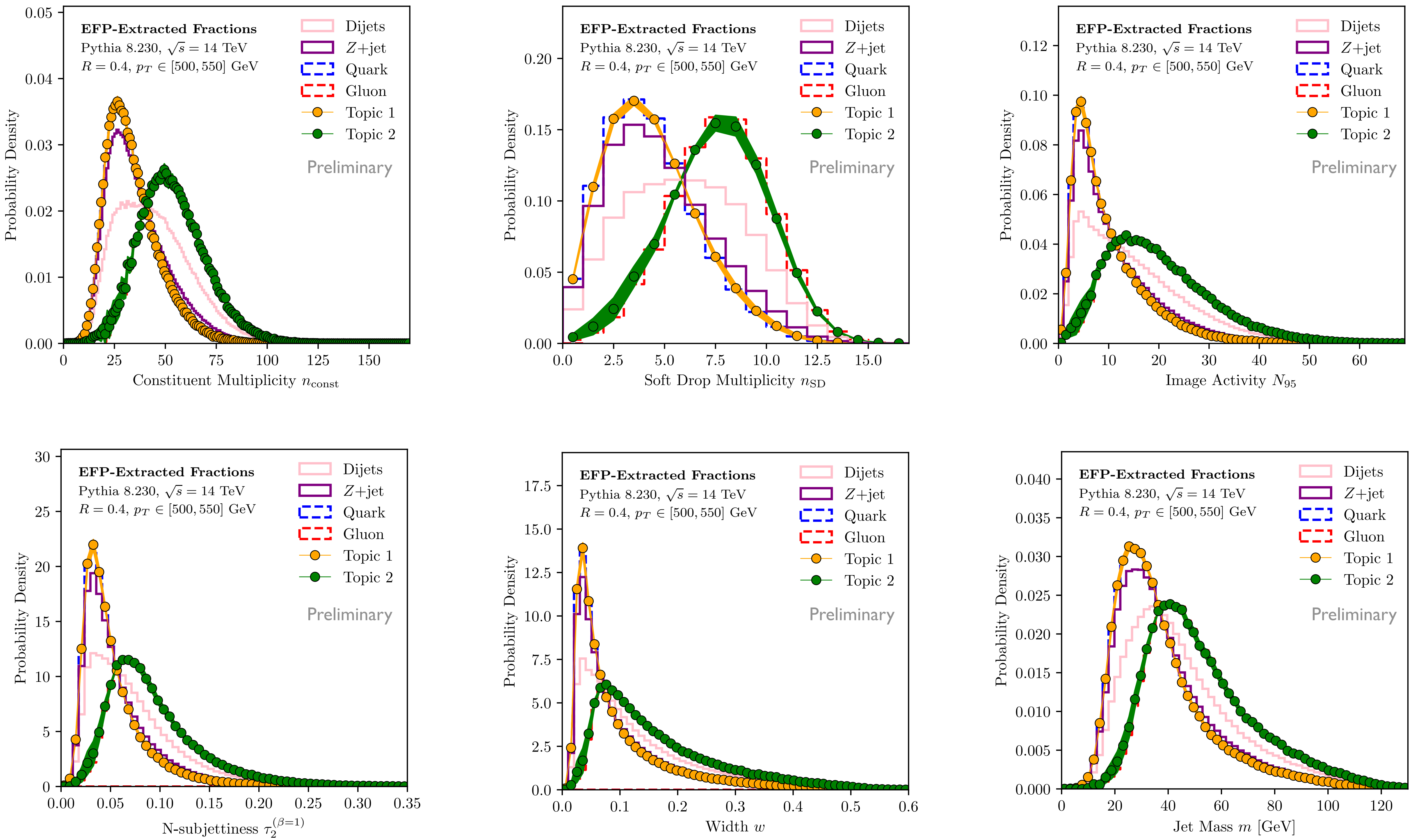


Model Outputs



[Komiske, Metodiev, JDT, to appear]

Extracting Quark/Gluon Distributions

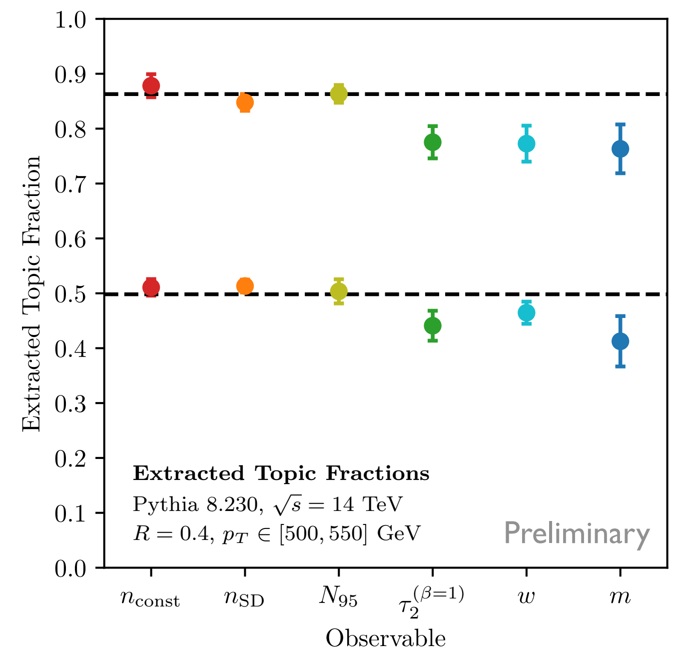


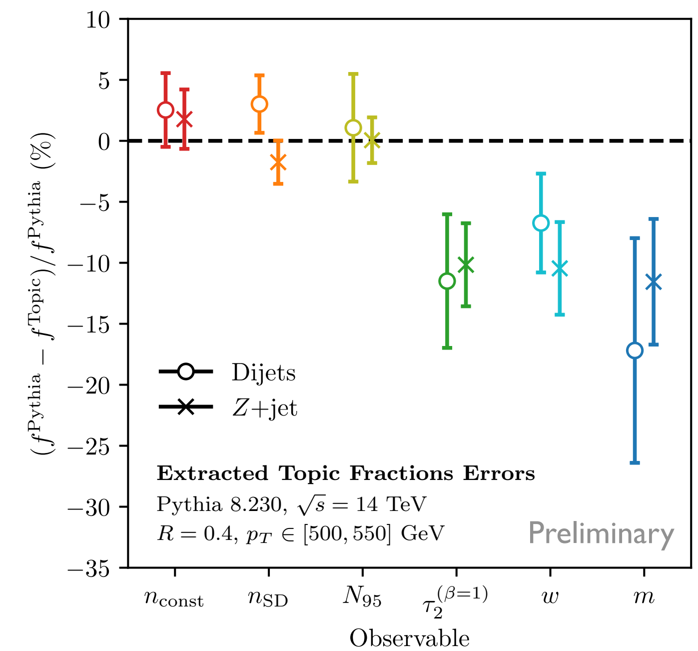
[Komiske, Metodiev, JDT, to appear]

Extracting Quark/Gluon Distributions

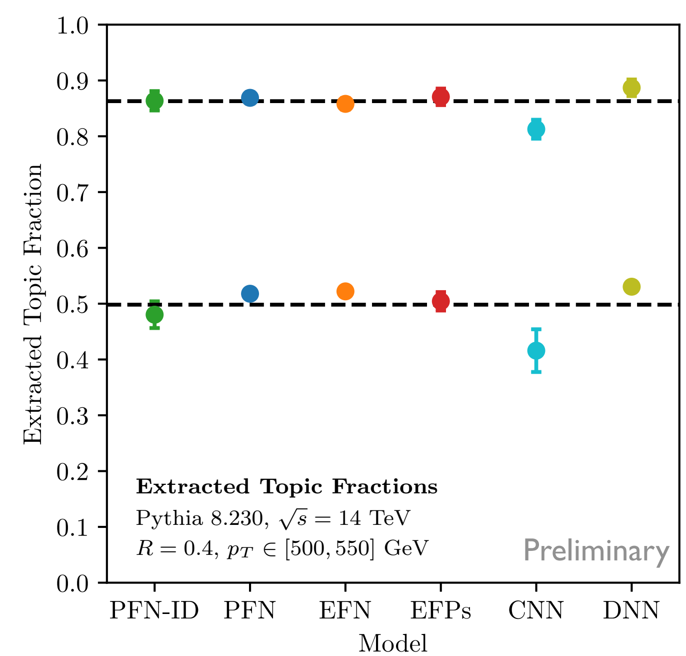
Determining topic fractions and uncertainties

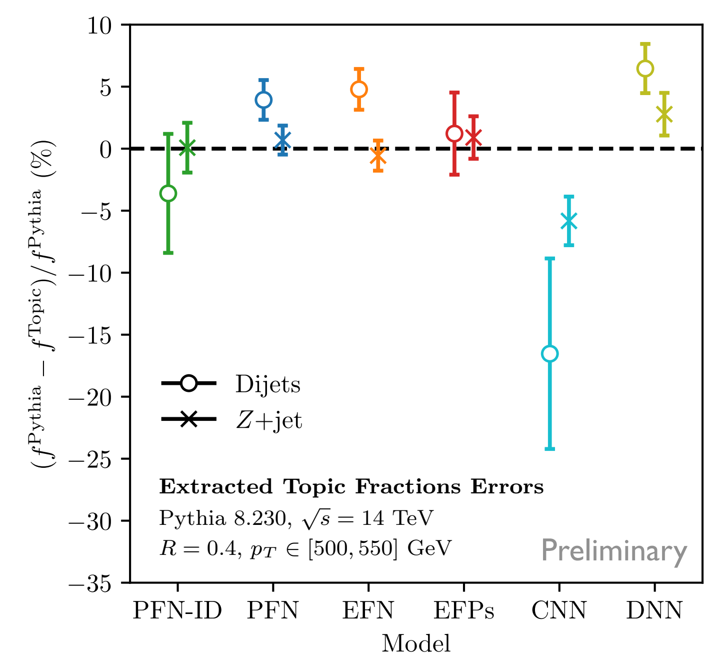
Classic Jet Observables





Model Outputs





[Komiske, Metodiev, JDT, to appear]