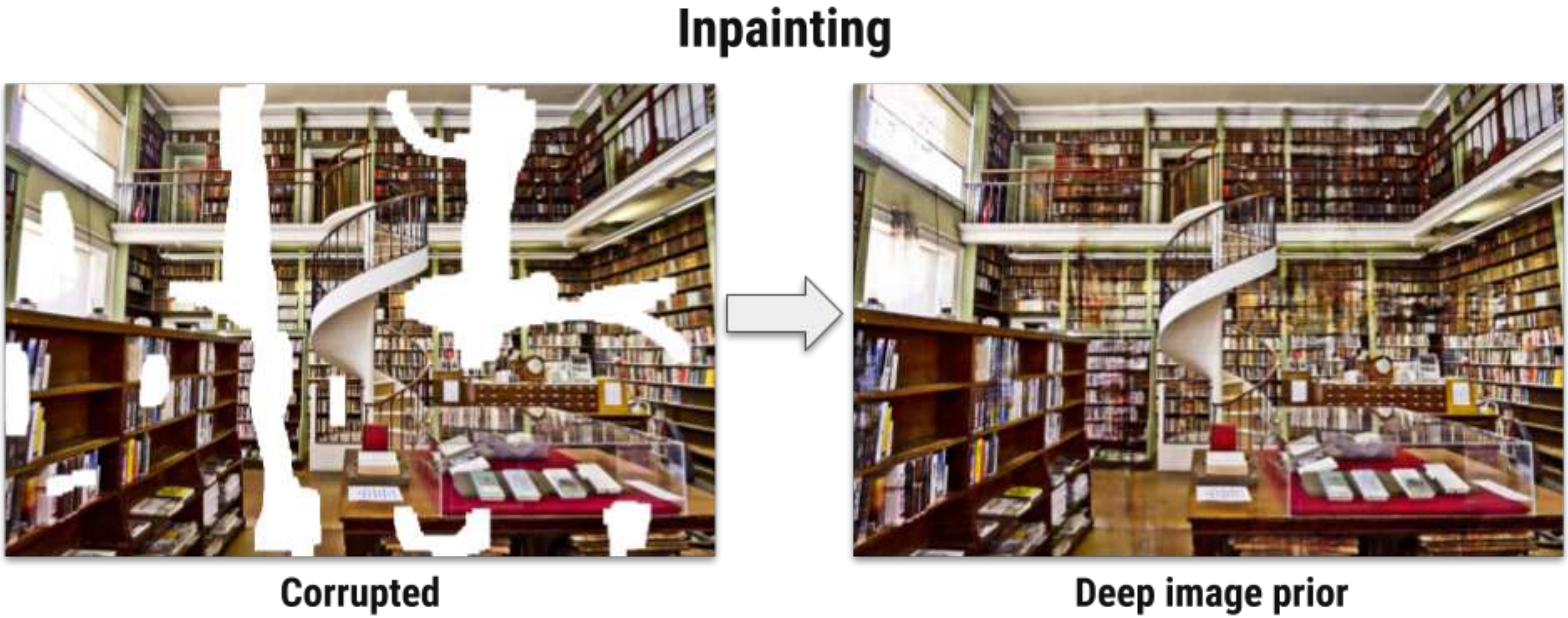
Deep Learning (and Deep Thinking) for QCD

Jesse Thaler

QCD@LHC 2019, University at Buffalo — July 15, 2019

Deep Learning

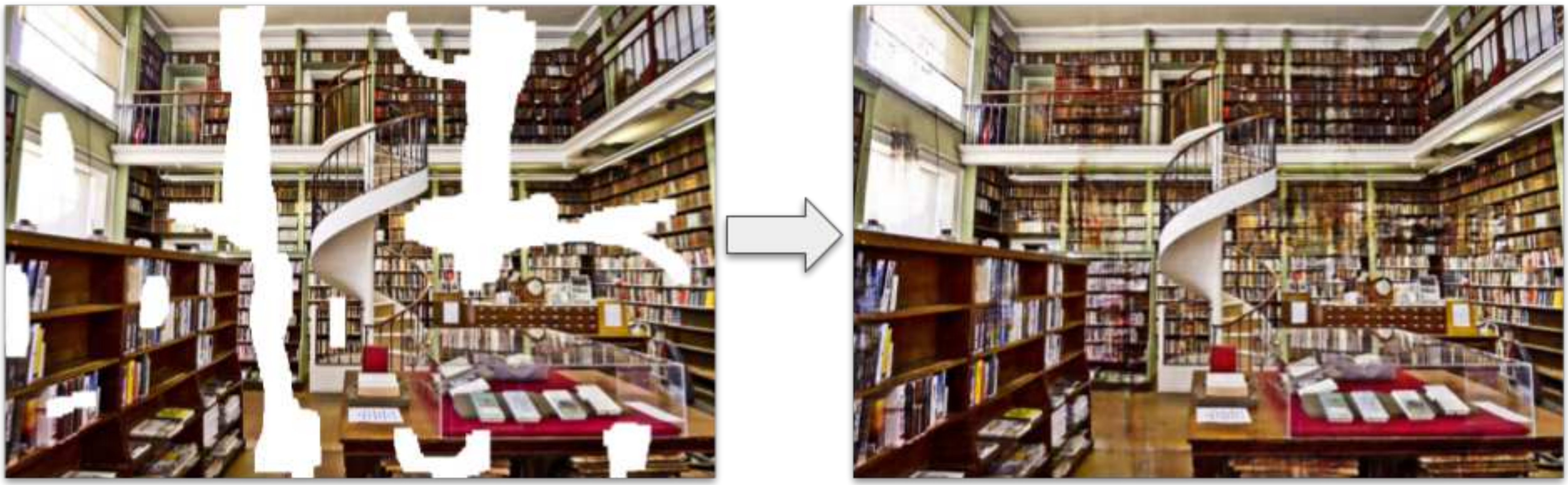


increased computational power and large data sets

[Ulyanov, Vedaldi, Lempitsky, 1711.10925]

Deep Learning (or Deep Thinking?)

Inpainting



Corrupted | **Deep image prior**

Using randomly initialized neural network (!)

Progress made by understanding the structure of problems

(not just increased computational power and large data sets)

[Ulyanov, Vedaldi, Lempitsky, 1711.10925]

Deep Learning for QCD?

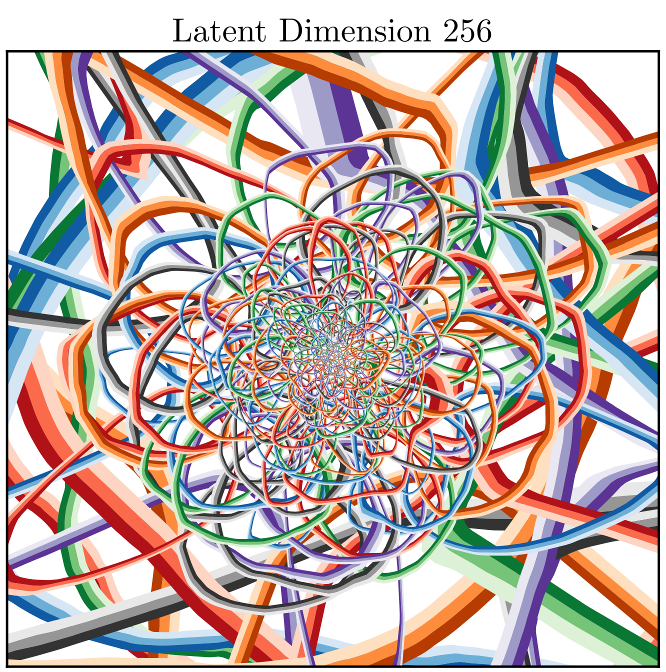
& other advanced data analysis strategies

New insights into structure of jets?
Robust handles on hadronic final states?

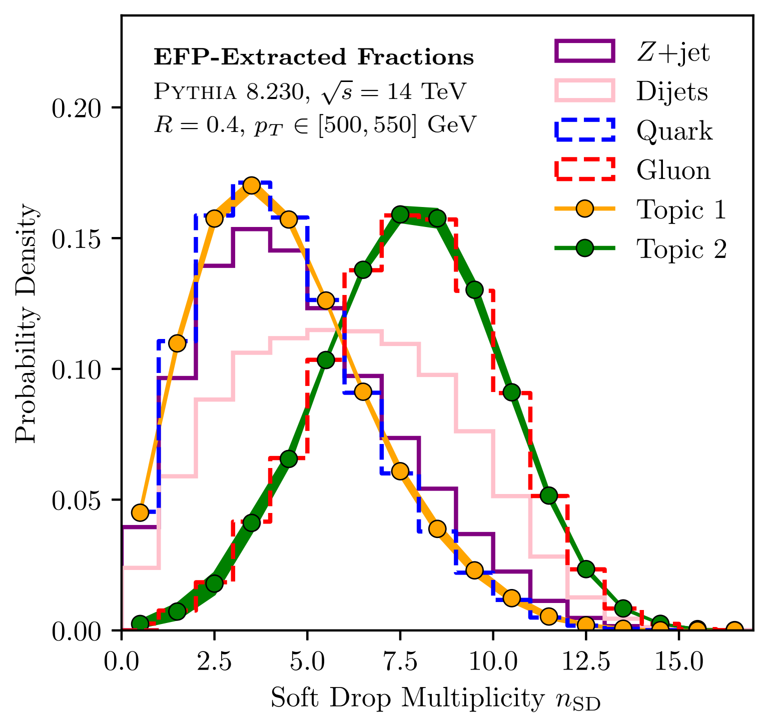
Three case studies from my research group

See also broader ML4Jets community

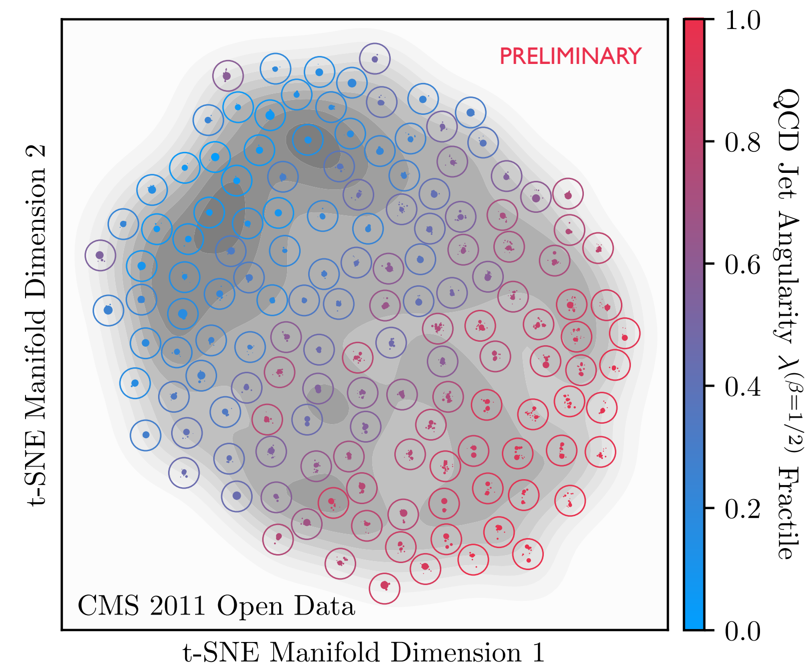
Energy Flow Networks



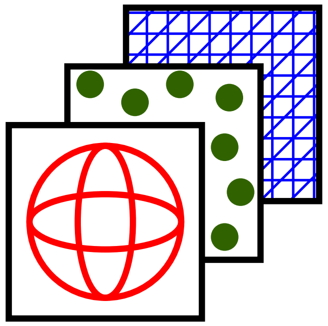
Jet Topics



(Energy Mover's Distance)



[Komiske, Metodiev, JDT, 1810.05165; Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140;
Komiske, Metodiev, JDT, 1902.02346; Komiske, Mastandrea, Metodiev, Naik, JDT, in preparation]



Into the Network

Data Ex Machina

(The Space of Jets)

Cartoon of Machine Learning

Properly specified problem $\Rightarrow$

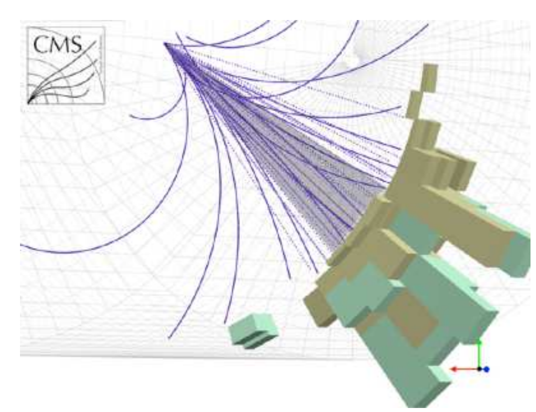
Many (labeled) examples $\{\mathcal{J}\}$ $\Rightarrow$

The Machine

$\Rightarrow$ Solution $S(\mathcal{J})$

E.g.: Problem = Minimize loss function
Solution = Multi-layer neural network
Strategy = Stochastic gradient descent

For most of this talk: $\mathcal{J}$ = "jet"

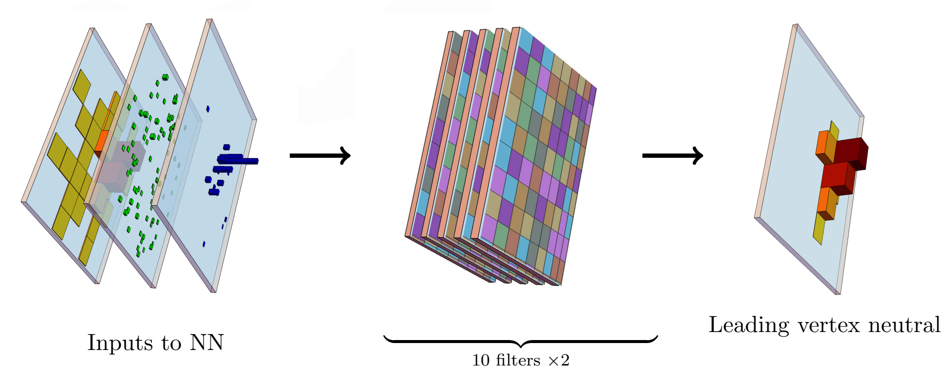


Examples of Supervised Learning

Regression

e.g. PUMML for pileup mitigation

[Komiske, Metodiev, Nachman, Schwartz, 1707.08600; see also Arjona Martínez, Cerri, Pierini, Spiropulu, Vlimant, 1810.07988]

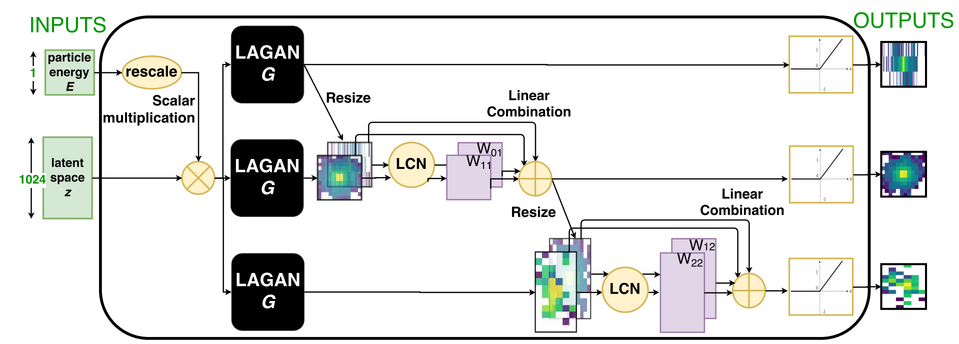


Labeled data: Objects J with property x
Solution: Map from J to x

Generation

e.g. CaloGAN for fast detector simulation

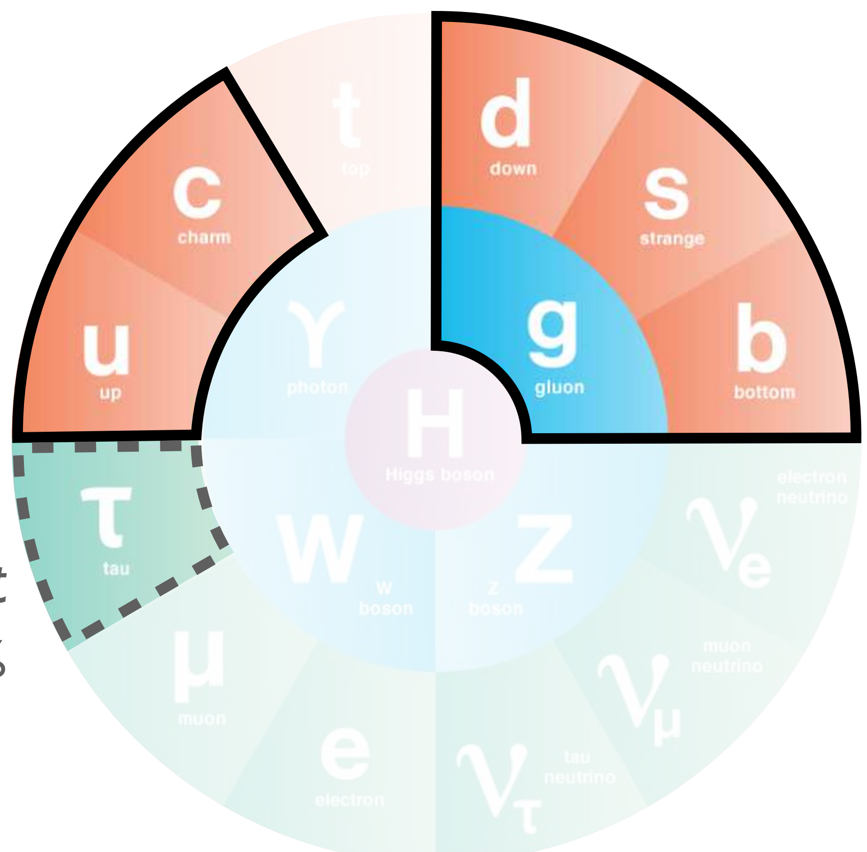
[Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321; see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



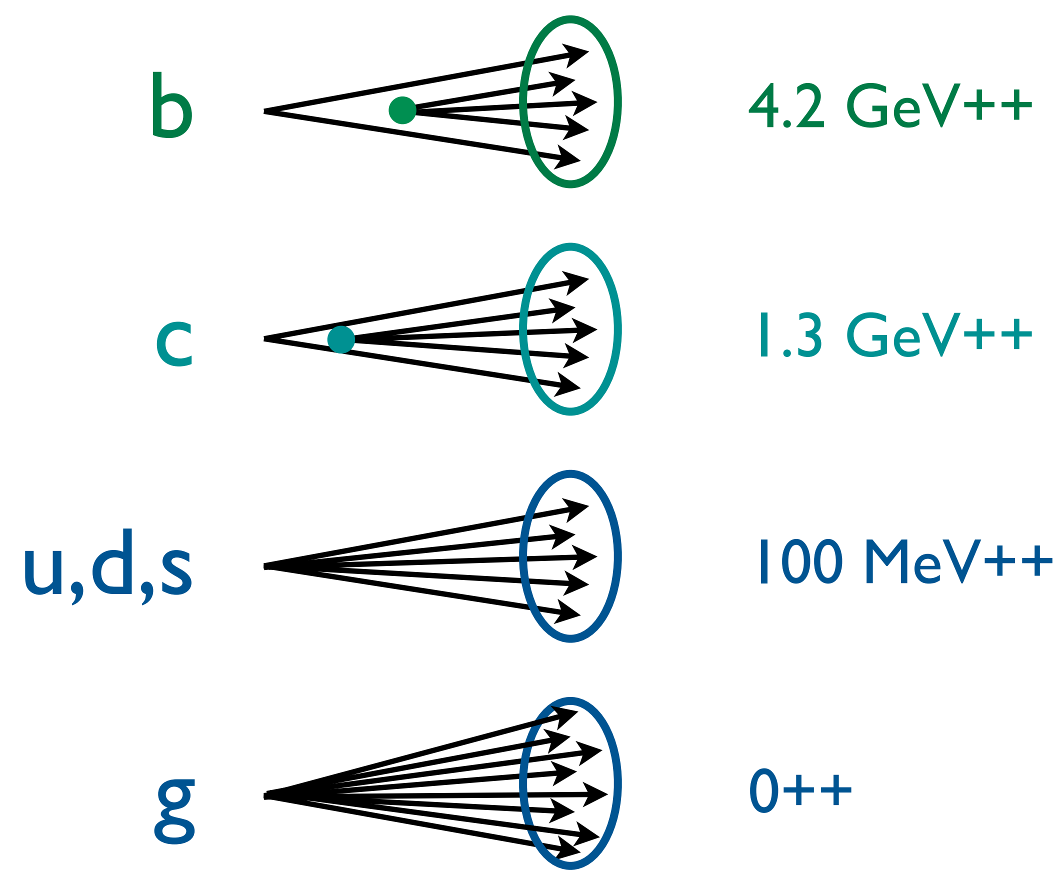
Labeled data: Objects J with property x
Solution: Map (conditioned on x) from noise to J

Jet Classification

Key supervised learning task at LHC



micro-jet
≈ 65%

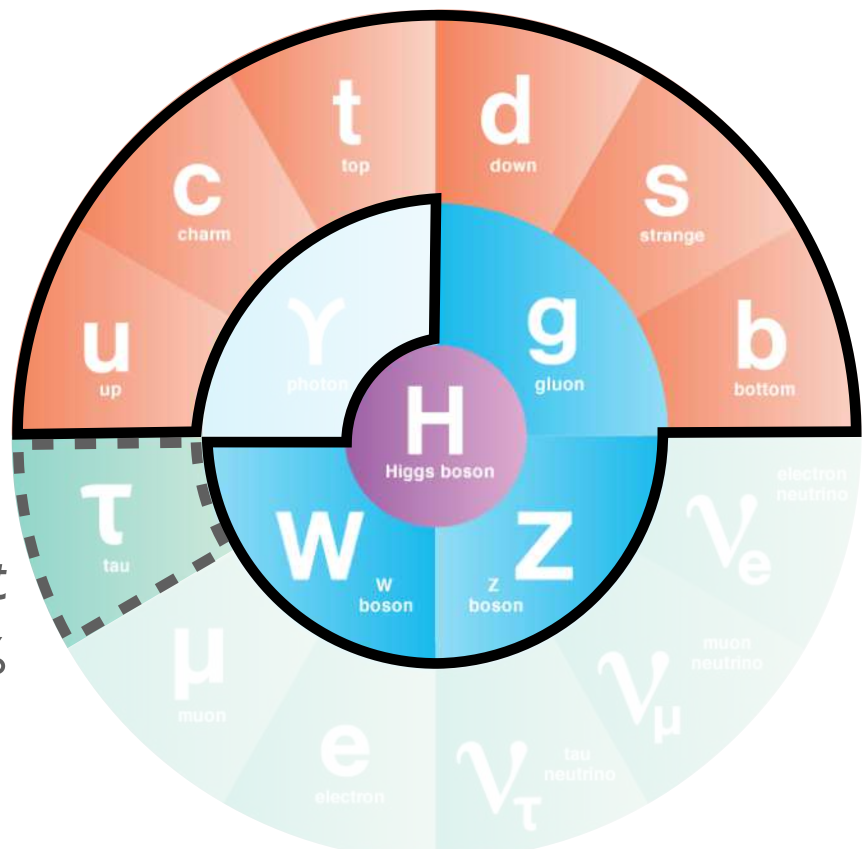


++ = Mass from QCD Radiation

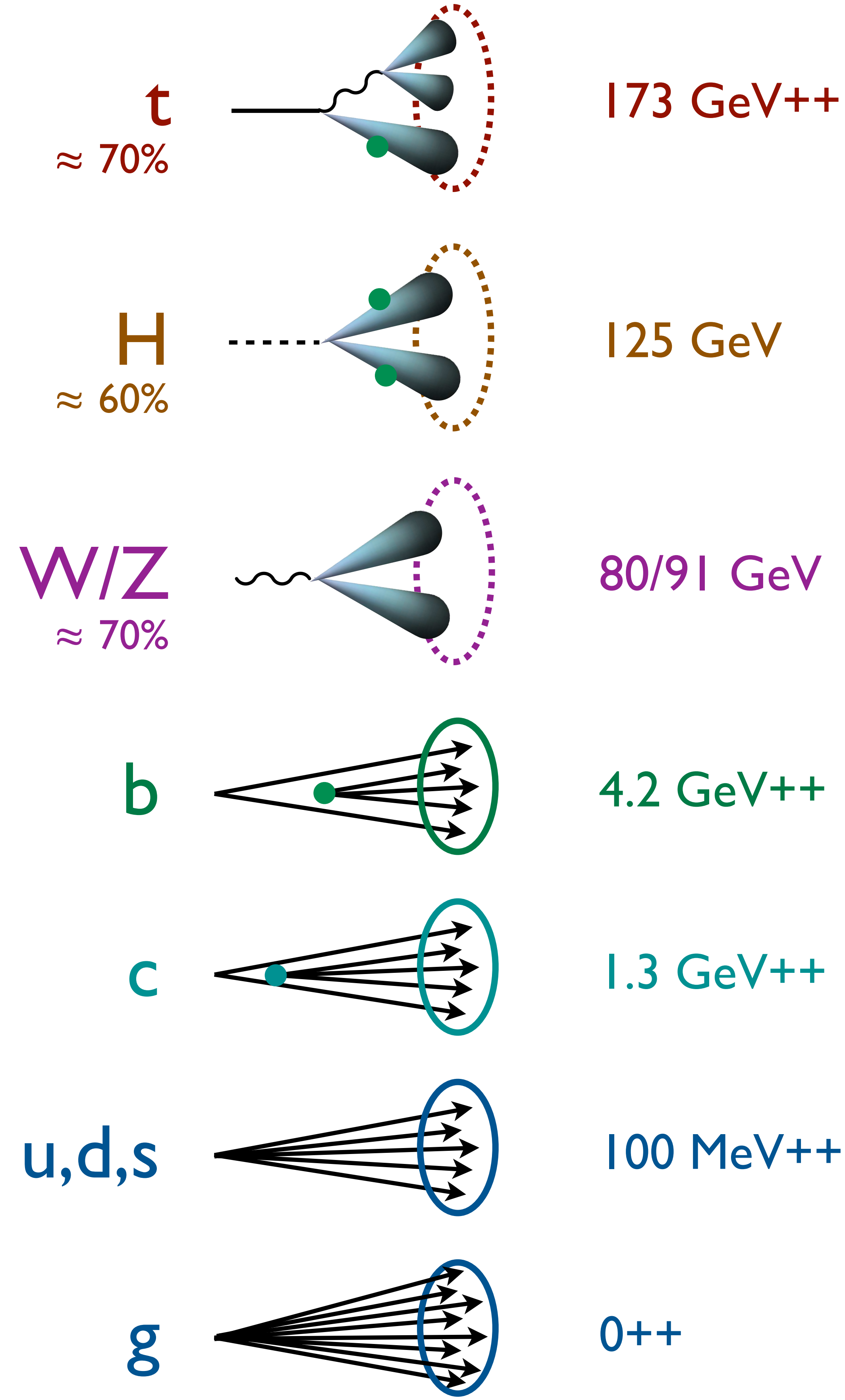
[see review in Larkoski, Moult, Nachman, 1709.04464]

Jet Classification

Key supervised learning task at LHC



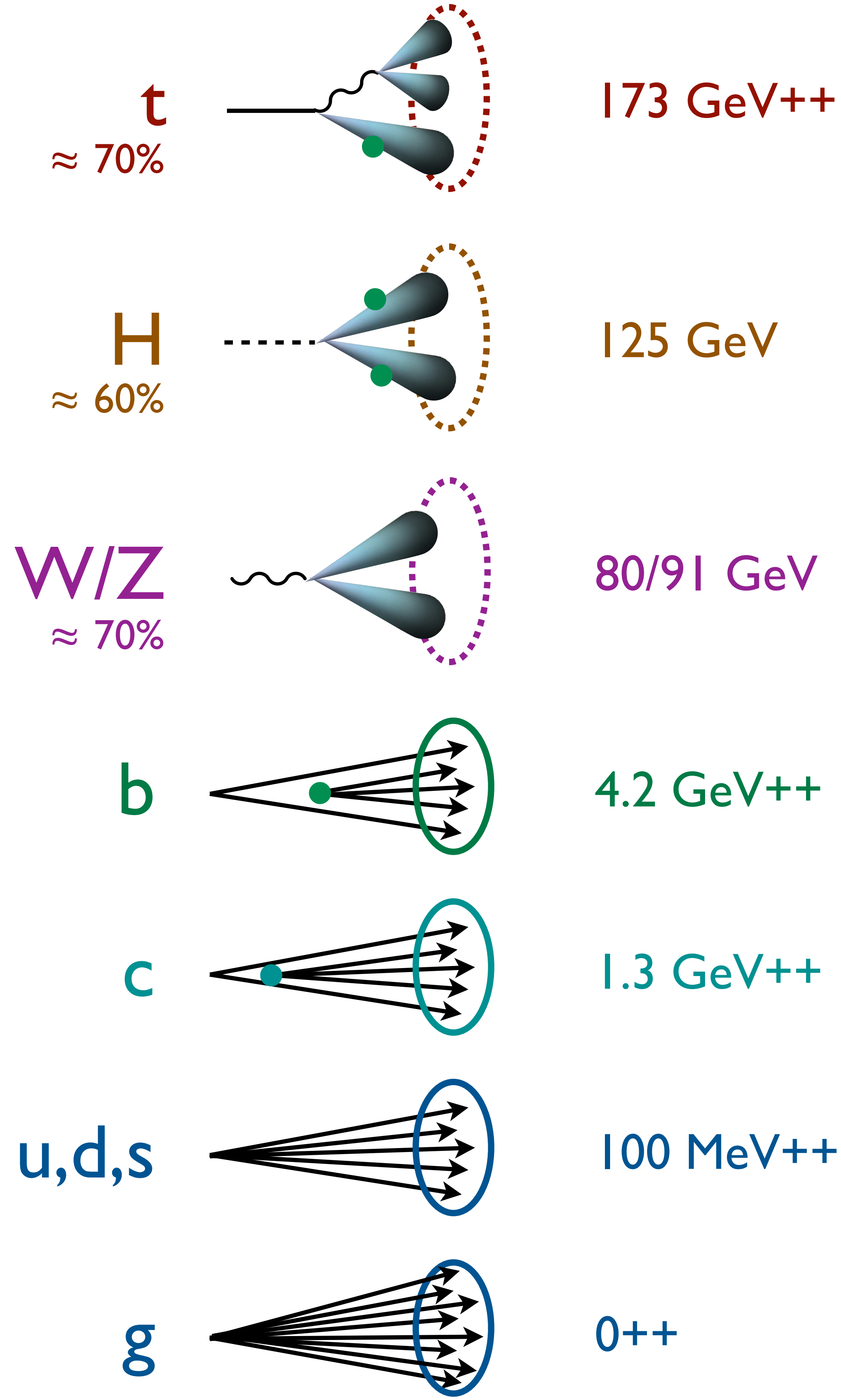
micro-jet
≈ 65%



++ = Mass from QCD Radiation

[see review in Larkoski, Moult, Nachman, 1709.04464]

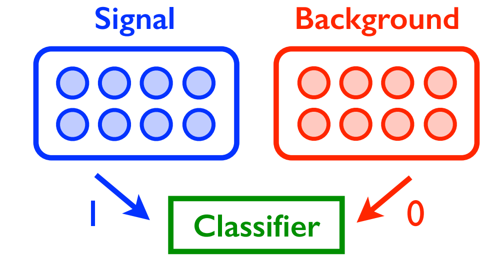
11th International Workshop on Boosted Objects
BOOSTON
2019
Phenomenology | Reconstruction | Searches | Algorithms | Measurements | Calculations
Modeling | Machine Learning | Pileup Mitigation | Heavy-Ion Collisions | Future Colliders
Local Organizing Committee:
Zeynep Demiragli (BU)
Philip Harris (MIT)
Yen-Jie Lee (MIT)
Matthew Schwartz (Harvard)
Jesse Thaler (MIT)
International Advisory Committee:
Ayana Arce (Duke)
Lily Asquith (Sussex)
Jon Butterworth (UCL)
Reina Camacho Toro (CNRS)
Mrinal Dasgupta (Manchester)
Robin Erbacher (UC Davis)
Gregor Kasieczka (Hamburg)
Andrew Larkoski (Reed)
Peter Loch (Arizona)
David Miller (Chicago)
Mihoko M. Nojiri (KEK)
Sal Rappoccio (Buffalo)
Gavin Salam (Oxford)
Alexander Schmidt (Aachen)
Ariel Schwartzman (SLAC)
Gregory Soyez (CNRS)
Marcel Vos (Valencia)
July 22-26, 2019
Stata Center, MIT
https://indico.cern.ch/e/boost2019

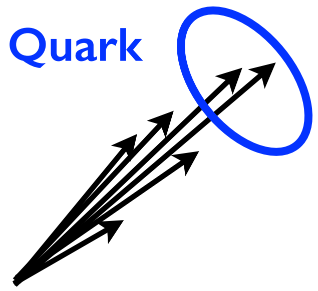


t
≈ 70%
173 GeV++
H
≈ 60%
125 GeV
W/Z
≈ 70%
80/91 GeV
b
4.2 GeV++
c
1.3 GeV++
u,d,s
100 MeV++
g
0++

Binary Classification

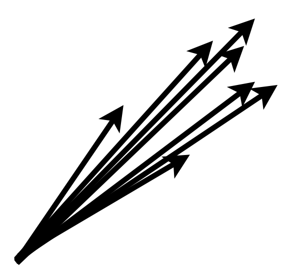
Much more in backup



e.g. 

vs. 

assuming trustable training data (more later…)

Find h () such that

$$h\big(\textbf{Quark}\big) = 1$$

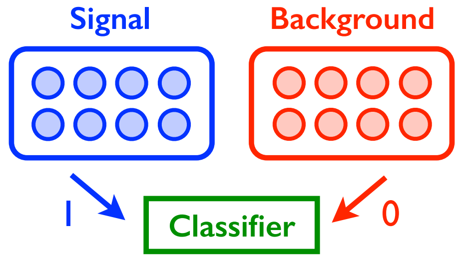
$$h\big(\textbf{Gluon}\big) = 0$$

Best you can do: $h(\mathcal{J}) = \dfrac{p(\mathcal{J}|\mathrm{Q})}{p(\mathcal{J}|\mathrm{Q}) + p(\mathcal{J}|\mathrm{G})}$

(Neyman-Pearson lemma)

Jet Classification Studies

Mix and match



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

Loss Function

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
- …

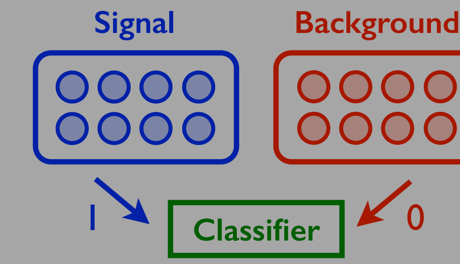
Signal vs. Background

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.	…

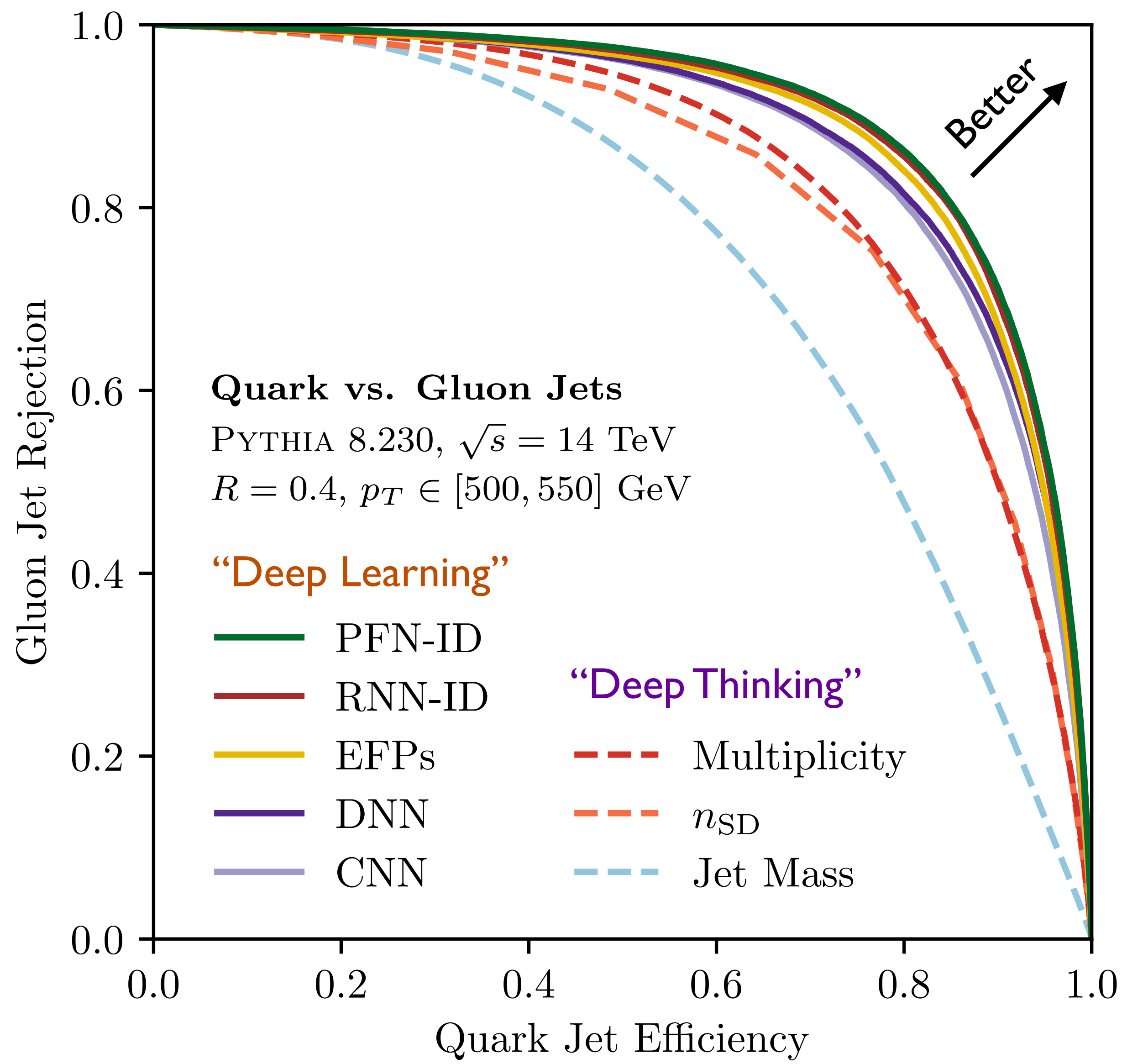
[Lönnblad, Peterson, Rögnvaldsson, PRL 1990, …, Cogan, Kagan, Strauss, Schwartzman, 1407.5675; Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968; de Oliveira, Kagan, Mackey, Nachman, Schwartzman, 1511.05190; Baldi, Bauer, Eng, Sadowski, Whiteson, 1603.09349; Conway, Bhaskar, Erbacher, Pilot, 1606.06859; 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 many ATLAS/CMS performance studies; plus my friends who will scold me for forgetting their paper (and not updating this after July 23, 2018)]

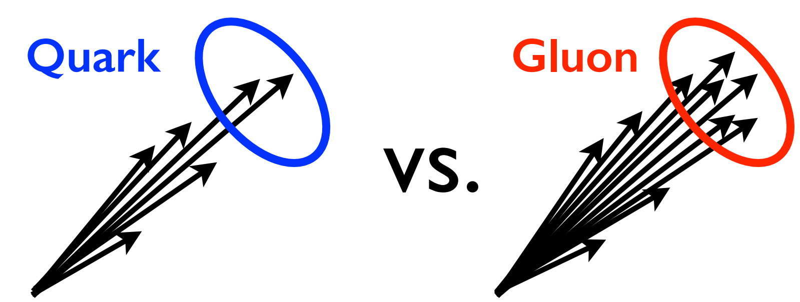
Jet Classification Studies

Mix and match



The “Hello, World!” of Jet Classification





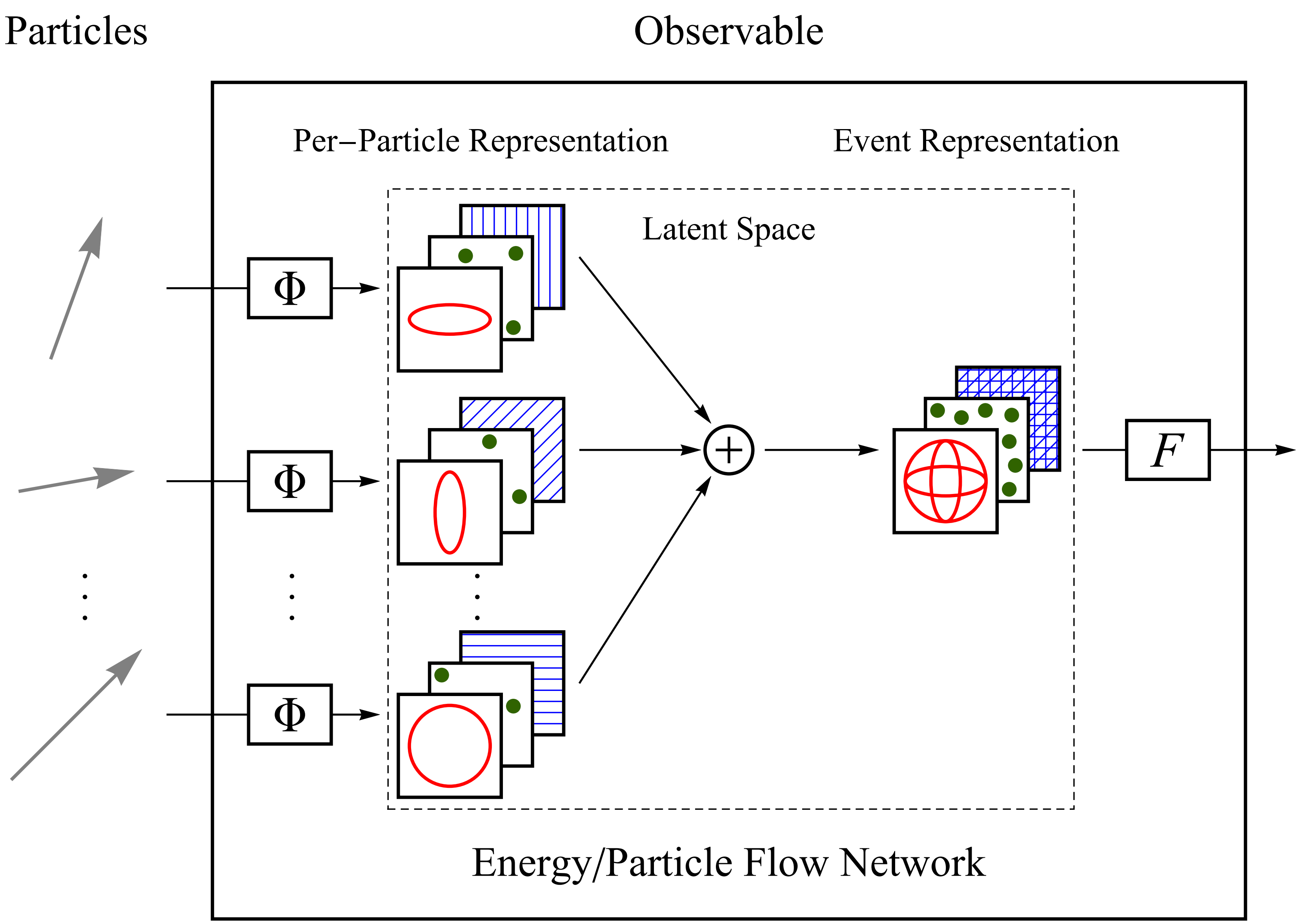
Substantial gains from deep learning… but why?

[Komiske, Metodiev, JDT, 1810.05165]

Introducing Energy Flow Networks

(see backup for detailed architecture)

An architecture designed for interpretability



[Komiske, Metodiev, JDT, 1810.05165;
special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114]

Introducing Energy Flow Networks

(see backup for detailed architecture)

An architecture designed for interpretability

Latent space of dim ℓ

Linear weights

$$S(\mathcal{J}) = F(V_1, V_2, \ldots, V_\ell) \qquad V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} p_{Ti}\, \Phi_a(y_i, \phi_i)$$

Parametrized with Neural Networks

Flexible enough to describe any* IRC-safe observable

(assuming large enough ℓ)

Generalization: Particle Flow Networks (aka "Deep Sets")

[Komiske, Metodiev, JDT, 1810.05165;
special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114]

Introducing Energy Flow Networks

An architecture designed for interpretability

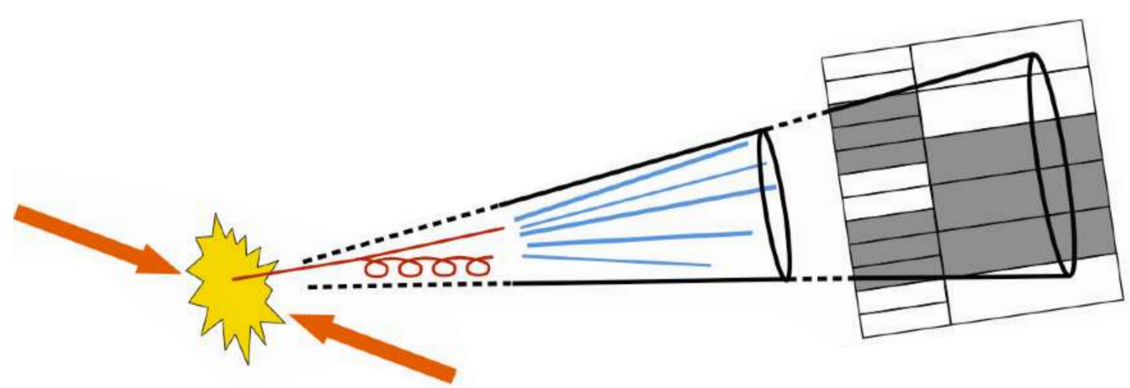
(see backup for detailed architecture)

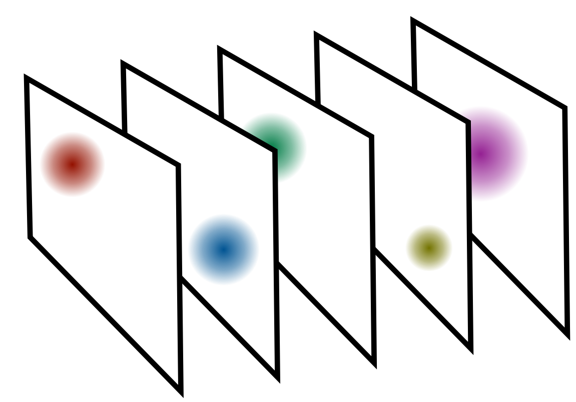
Visualization Strategy

$$S(\mathcal{J}) = F(V_1, V_2, \ldots, V_\ell) \qquad V_a(\mathcal{J}) = \sum_{i\in\mathcal{J}} p_{Ti}\, \Phi_a(y_i, \phi_i)$$

Difficult to visualize (unless ℓ is small)

Easy to plot these!



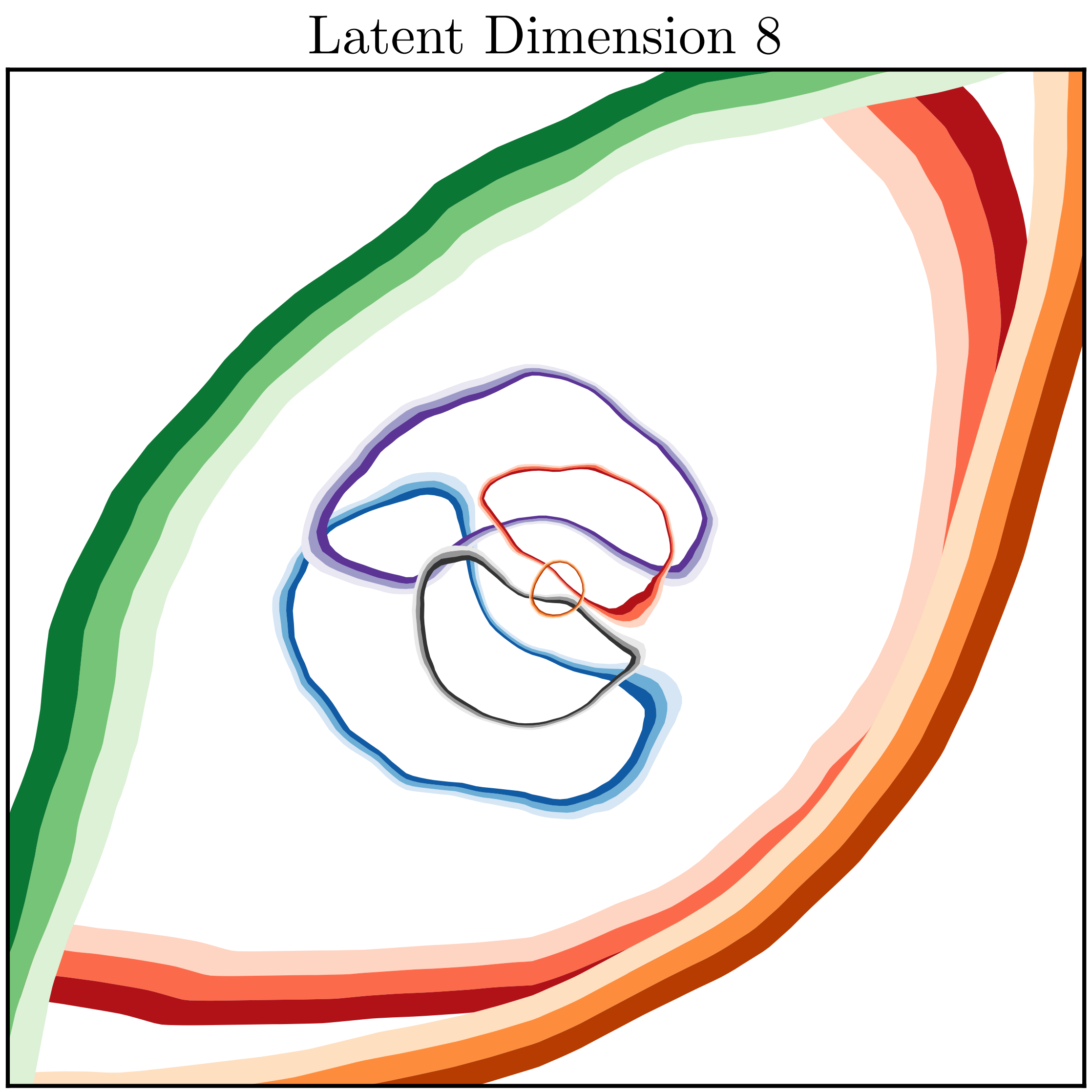


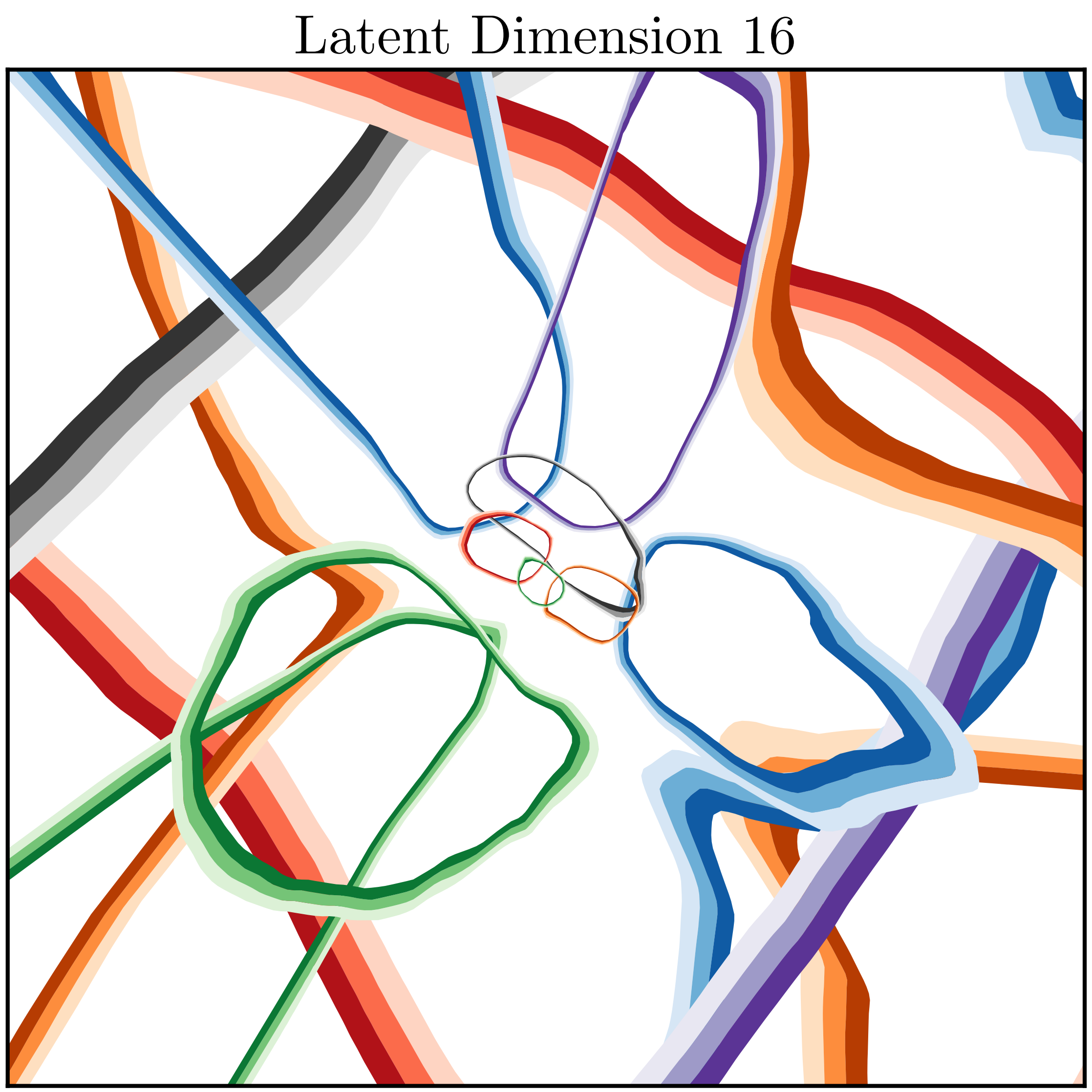
(similar to CNN filter activation)

[Komiske, Metodiev, JDT, 1810.05165; special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114]

Psychedelic Network Visualization

(see backup for how these are made)

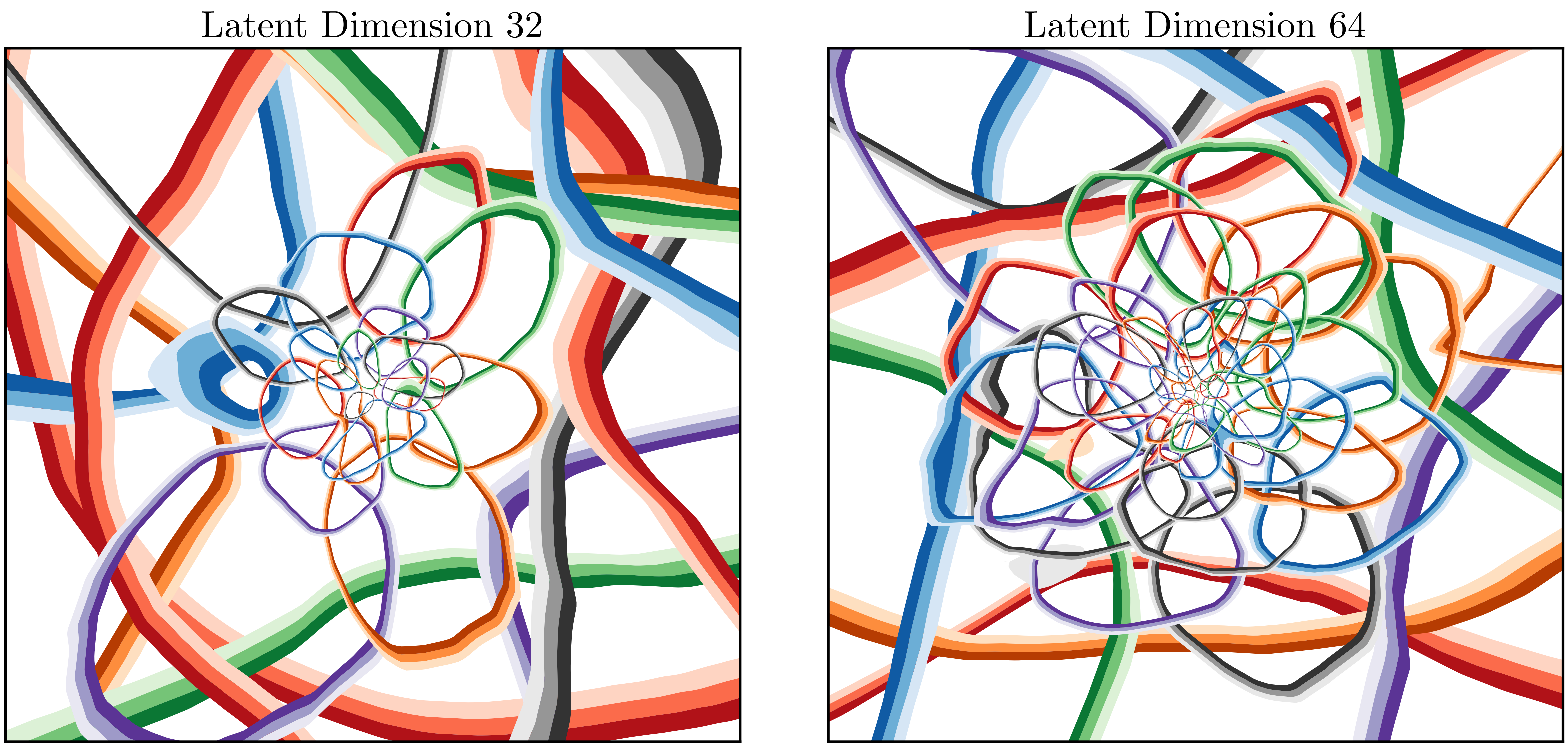




For the case of quark vs. gluon classification

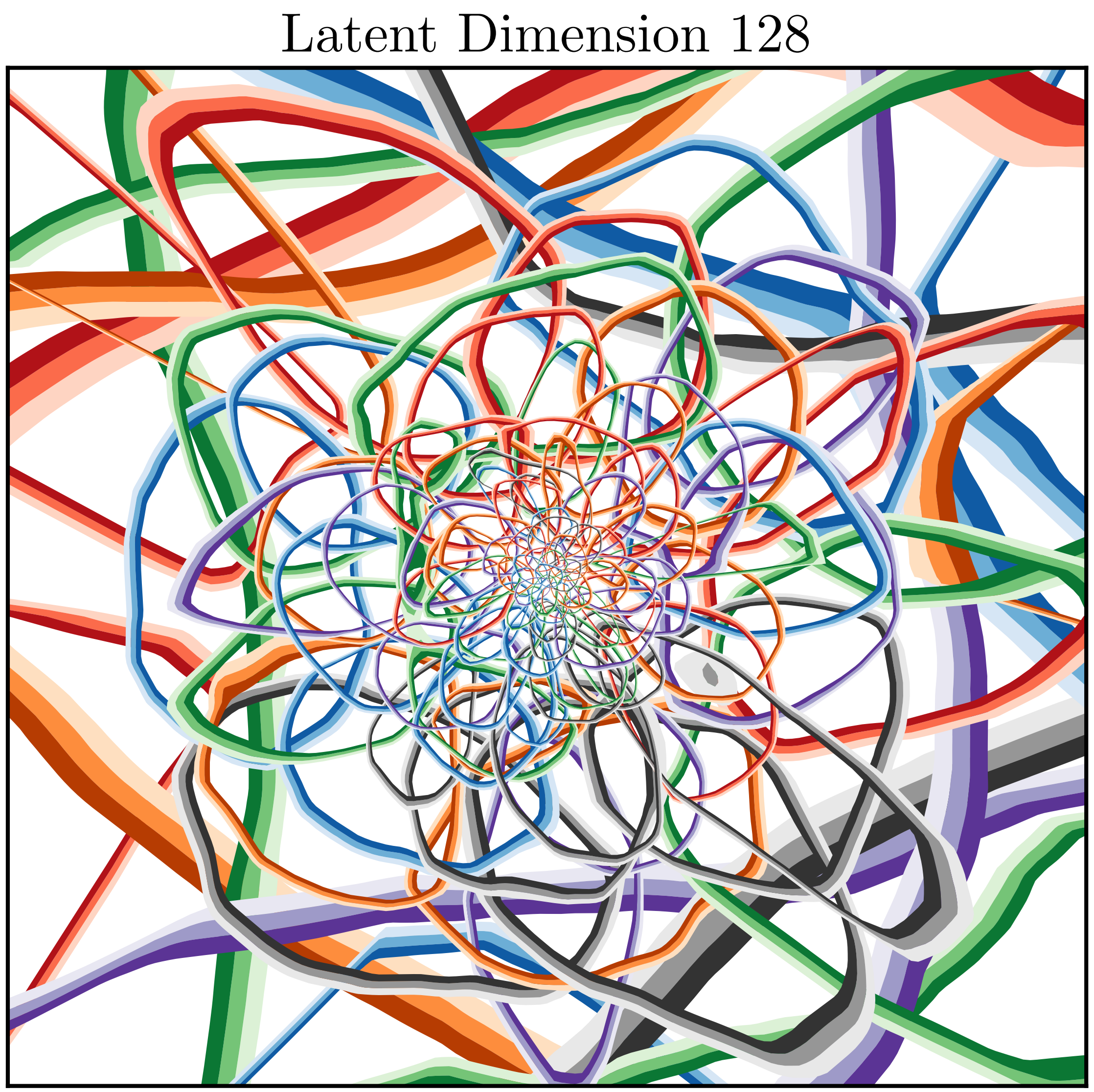
Psychedelic Network Visualization

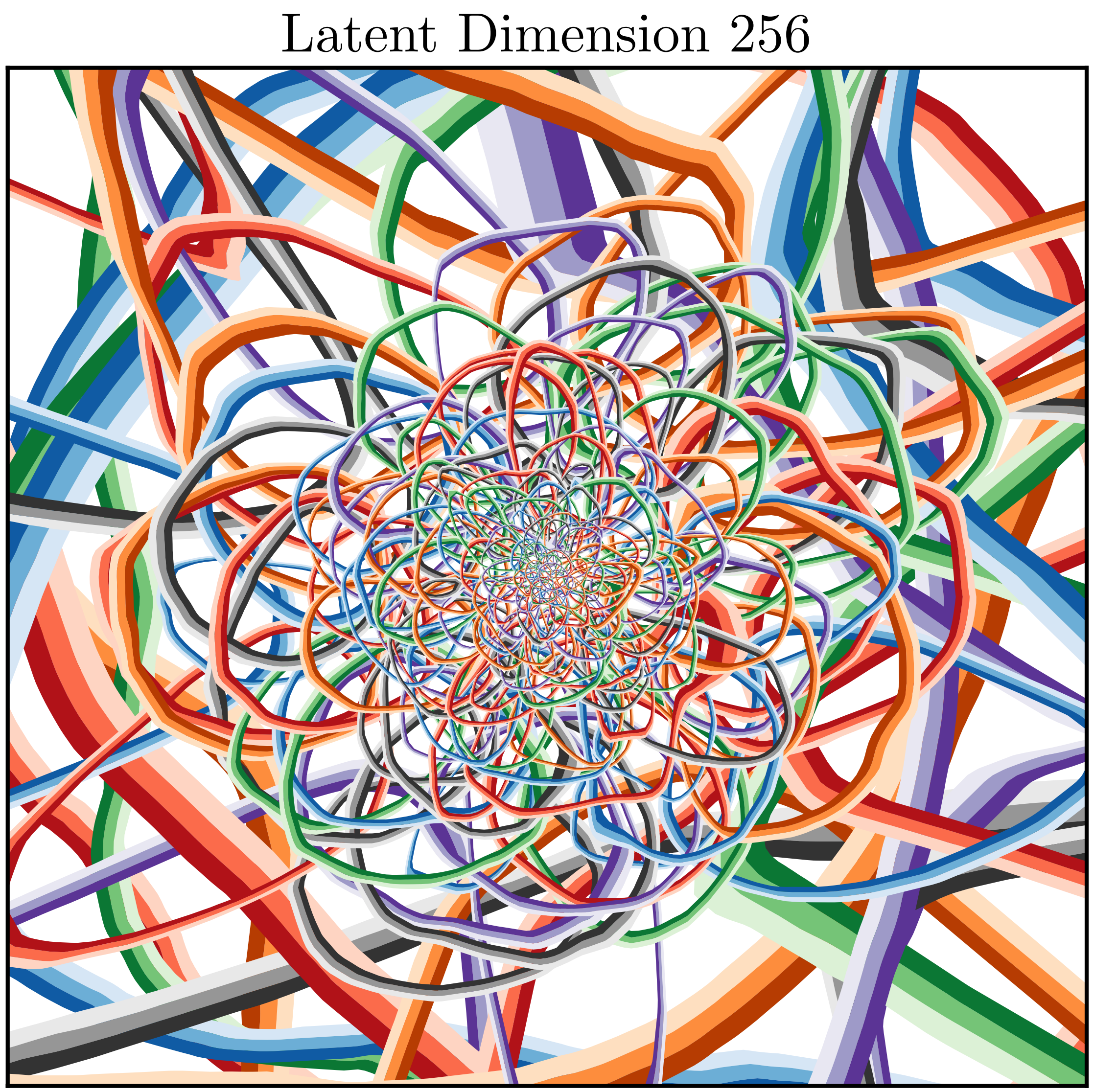
(see backup for
how these are made)



Psychedelic Network Visualization

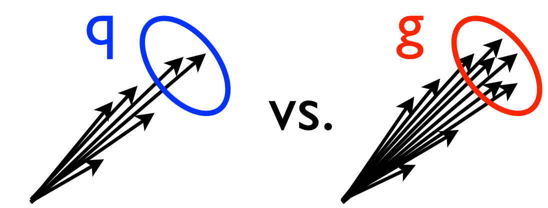
(see backup for how these are made)

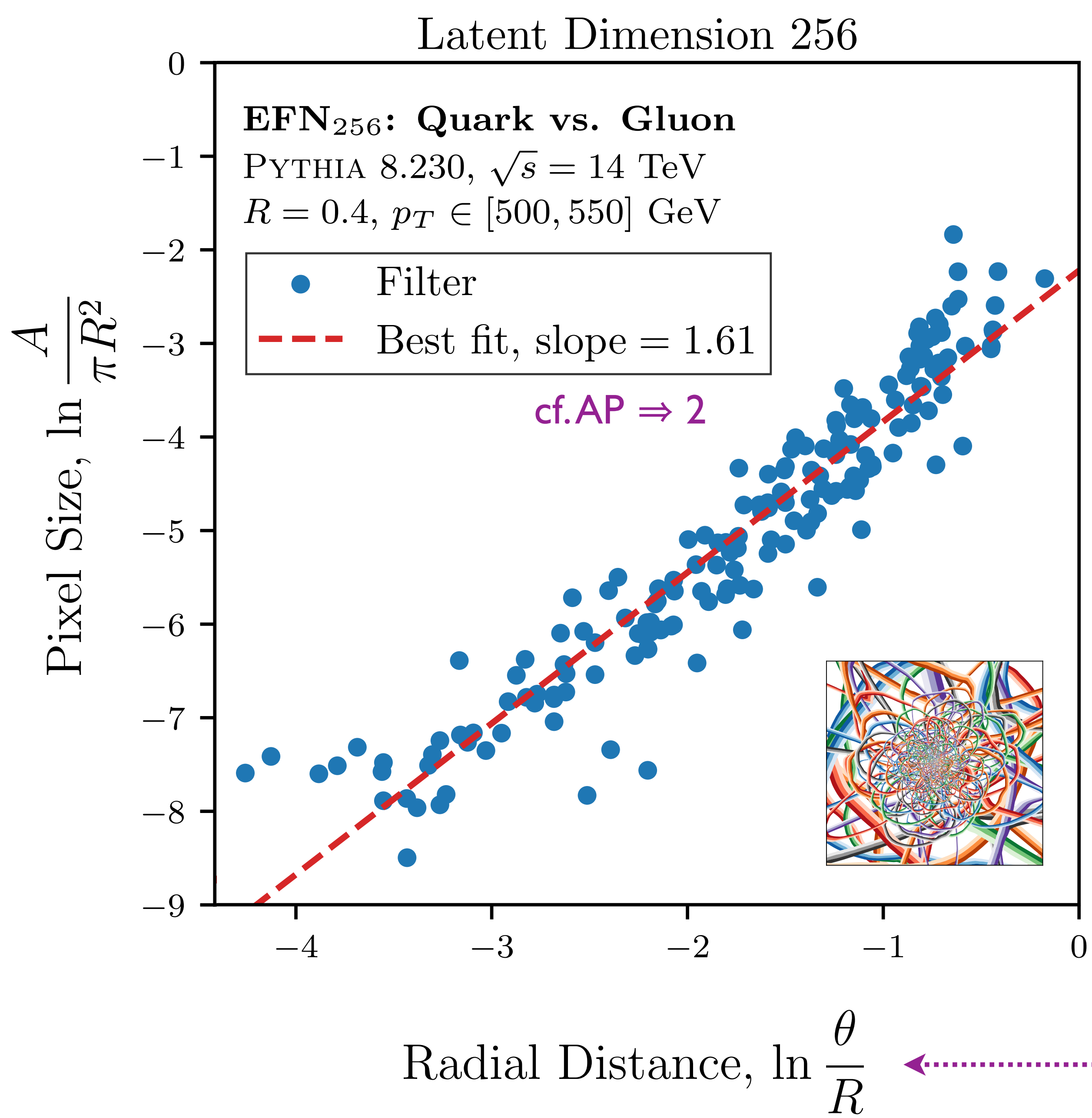


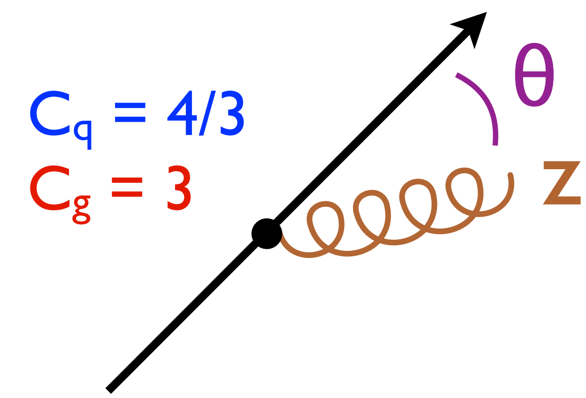


Singularity structure of QCD!

Putting the AI in AItarelli-Parisi



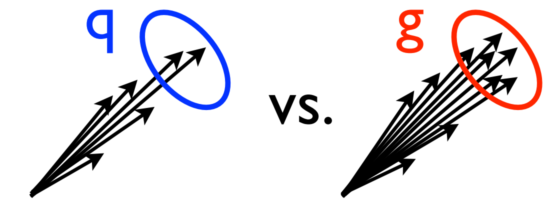




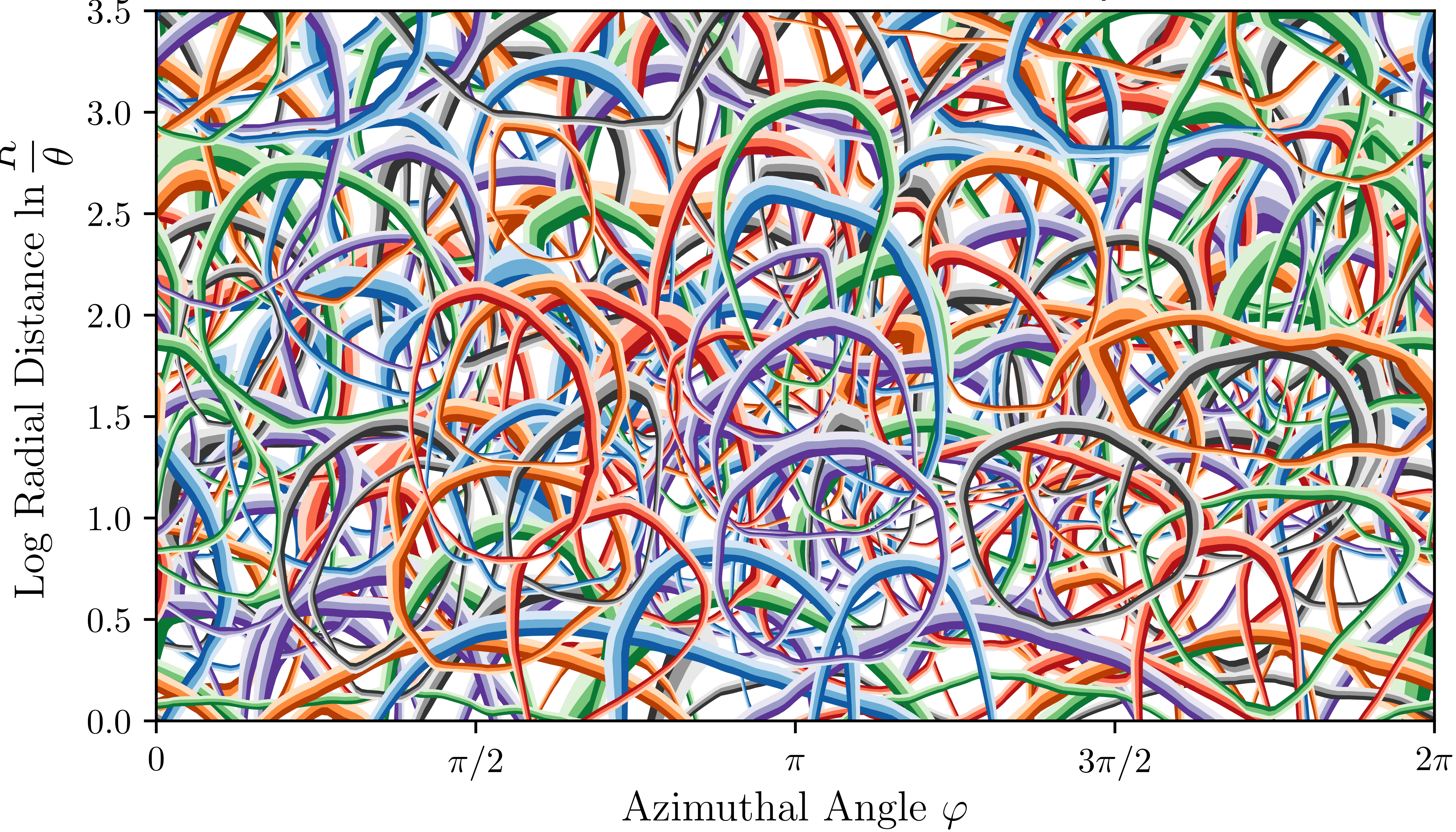
$$\mathrm{d}P_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{\mathrm{d}\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{\mathrm{d}z}{z}}_{\text{Soft}}$$

[Komiske, Metodiev, JDT, 1810.05165]

Suitable for Framing



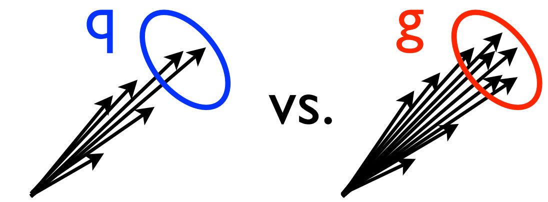
Coordinate transformation to the emission plane



[Komiske, Metodiev, JDT, 1810.05165; see also Dreyer, Salam, Soyez, 1807.04758]

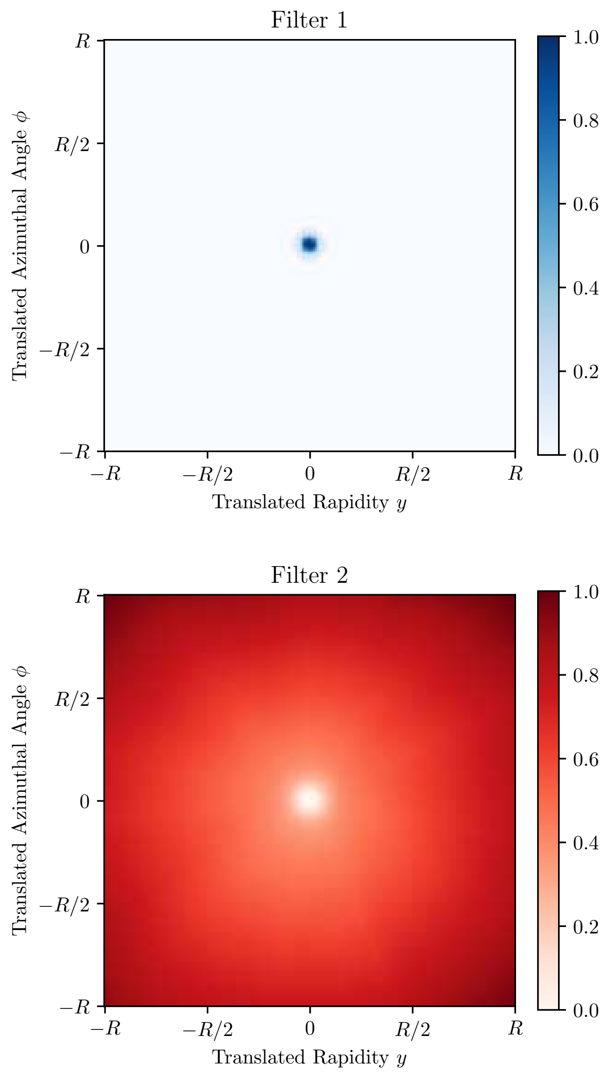
“Ok, but did you really learn something you didn’t already know?”

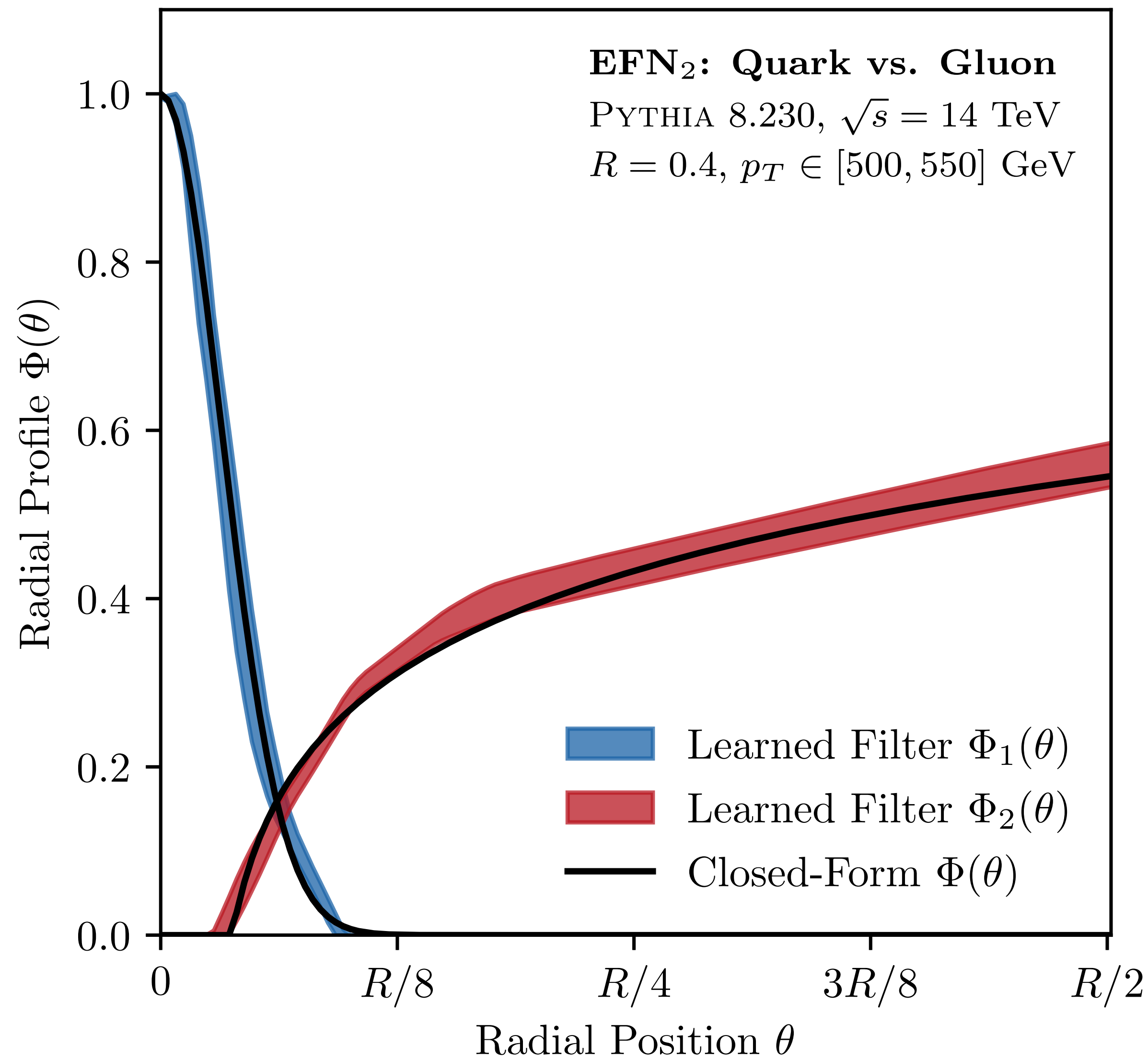
Learning from the Machine



For ℓ = 2 EFN, radial moments: $\sum_{i \in \text{jet}} z_i \, f(\theta_i)$

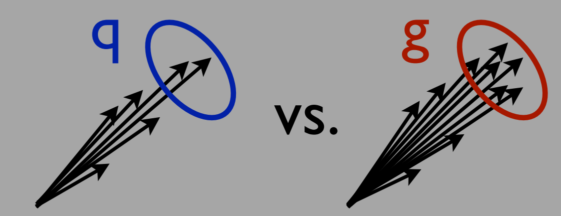
cf. Angularities:
$f(\theta) = \theta^\beta$





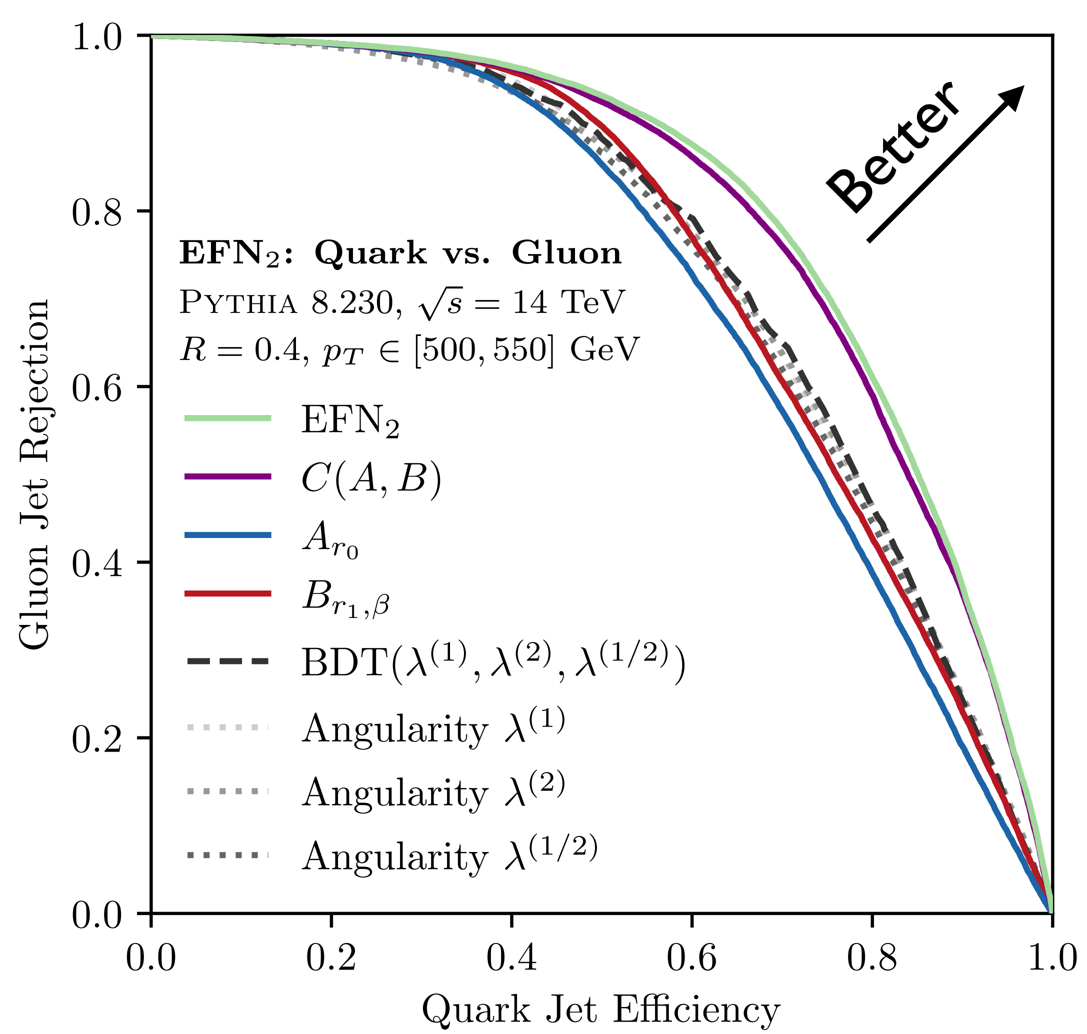
[Komiske, Metodiev, JDT, 1810.05165;
cf. Larkoski, JDT, Waalewijn, 1408.3122; using Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014]

Learning from the Machine



For ℓ = 2 EFN, radial moments: $\sum_{i \in \text{jet}} z_i f(\theta_i)$

cf. Angularities:
f(θ) = θ^β

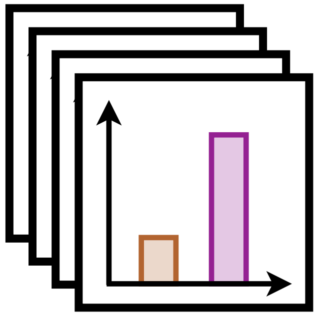


EFN outperformed a domain expert (i.e. me)

But we reverse engineered the machine (and learned something about QCD)

[Komiske, Metodiev, JDT, 1810.05165;
cf. Larkoski, JDT, Waalewijn, 1408.3122; using Berger, Kucs, Sterman, hep-ph/0303051; Ellis, Vermilion, Walsh, Hornig, Lee, 1001.0014]

Into the Network



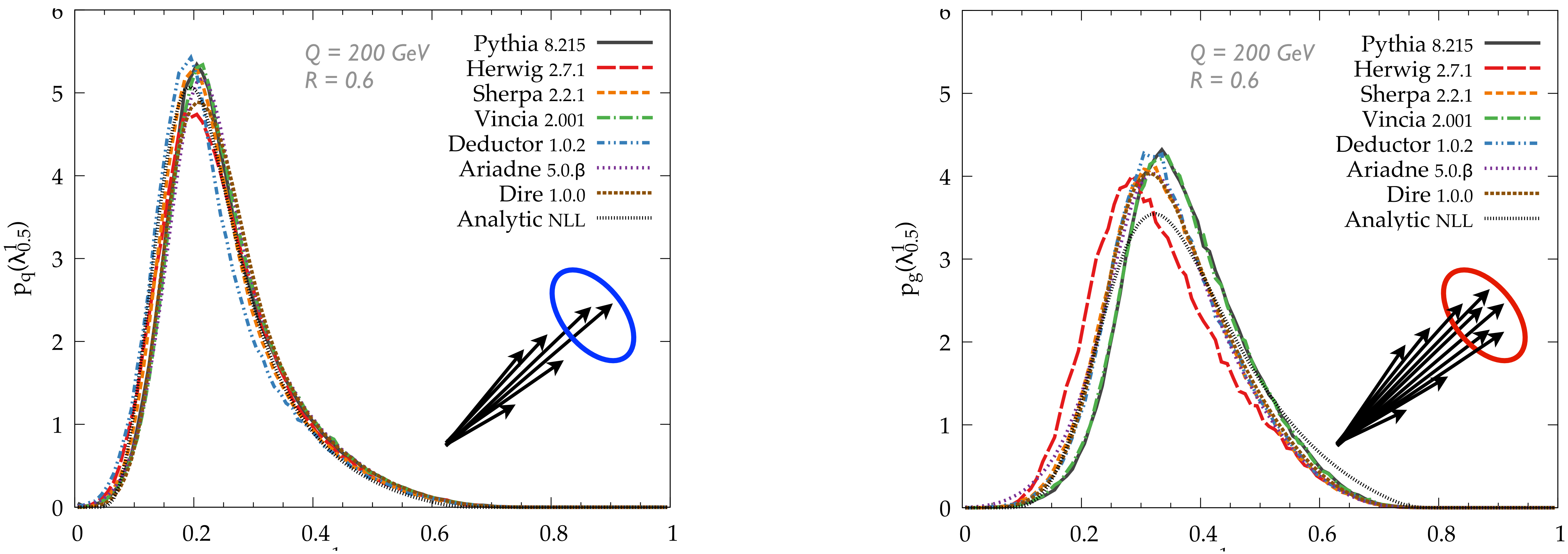
Data Ex Machina

(The Space of Jets)

“Ok, but isn’t supervised learning only as reliable as your training samples?”

Uncertainties in Monte Carlo 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;
see progress in Reichelt, Richardson, Siódmok, 1708.01491]

What are "Quarks" and "Gluons" anyways?

Color triplet vs. Color octet?

But jet constituents are color-singlet hadrons!

What is a Quark Jet?

From lunch/dinner discussions

Les Houches 2015
(& 2017 & 2019 & …)

Ill-Defined	What people sometimes think 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)
Well-Defined	What we mean	

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

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

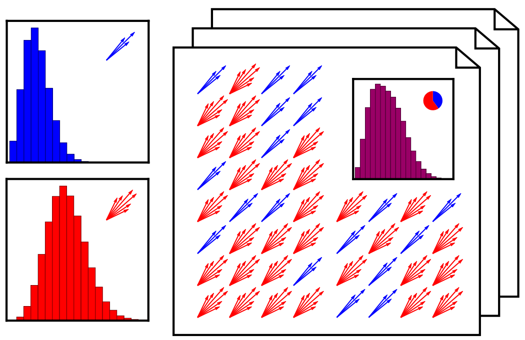
Data ex Machina

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

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

Enter Unsupervised Learning

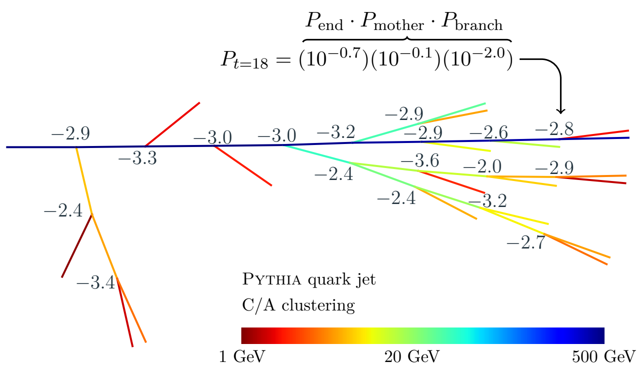
Learning from unlabeled (or barely labeled) data



Jet Topics

Blind Source Separation

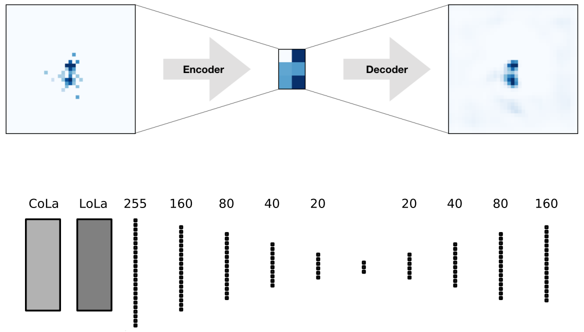
[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140;
see also Metodiev, Nachman, JDT, 1708.02949; Dillon, Faroughy, Kamenik, 1904.04200]



JUNIPR

Probability Modeling

[Andreassen, Feige, Frye, Schwartz, 1804.09720, 1906.10137;
see also Monk, 1807.03685]



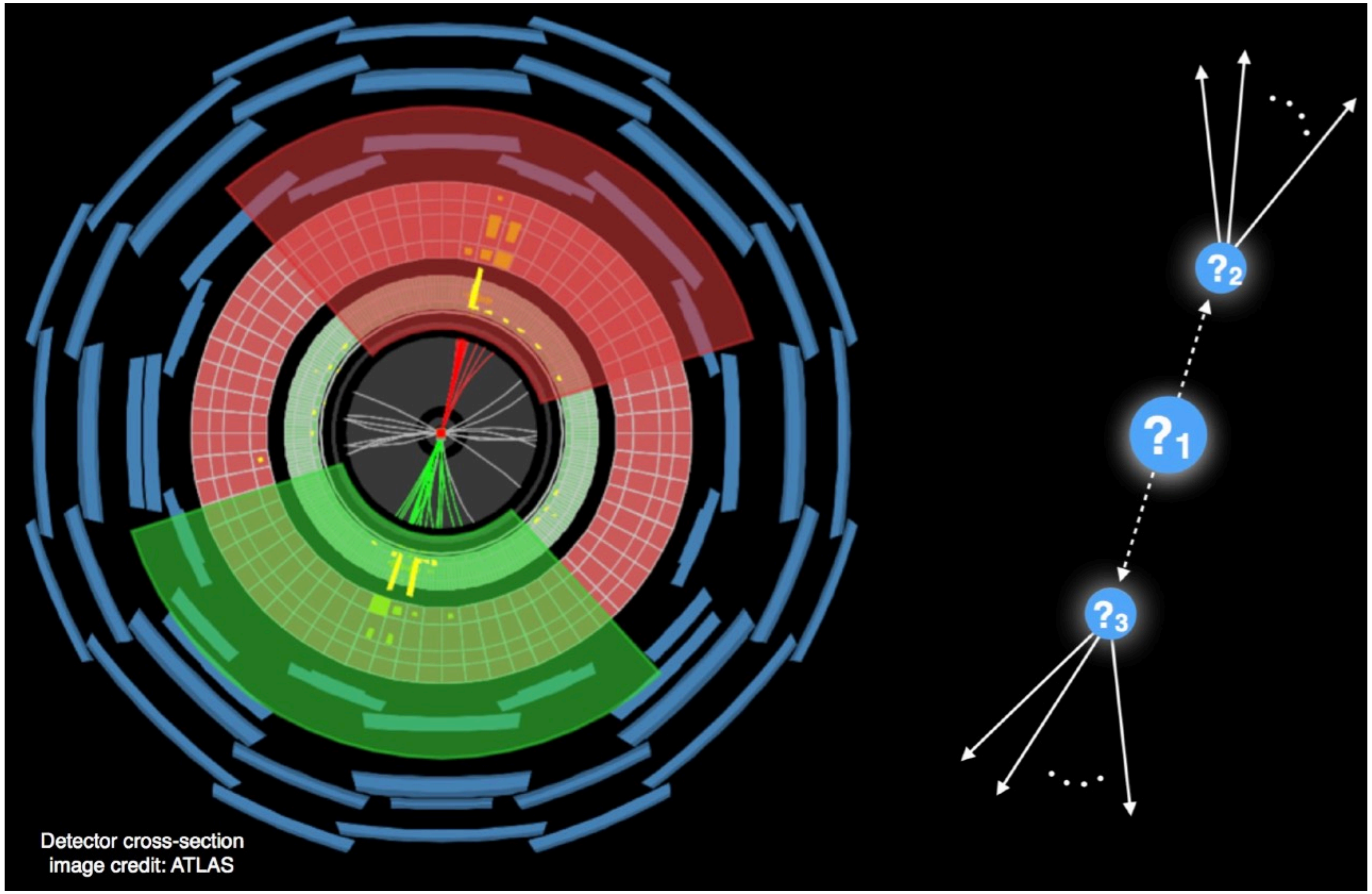
Autoencoders

Anomaly Detection

[Hajer, Li, Liu, Wang, 1807.10261; Heimel, Kasieczka, Plehn, Thompson, 1808.08979;
Farina, Nakai, Shih, 1808.08992; Cerri, Nguyen, Pierini, Spiropulu, Vlimant, 1811.10276;
see also Collins, Howe, Nachman, 1805.02664, 1902.02634; De Simone, Jacques, 1807.06038]

LHC Olympics 2020

@ ML4Jets, NYU, January 15-17

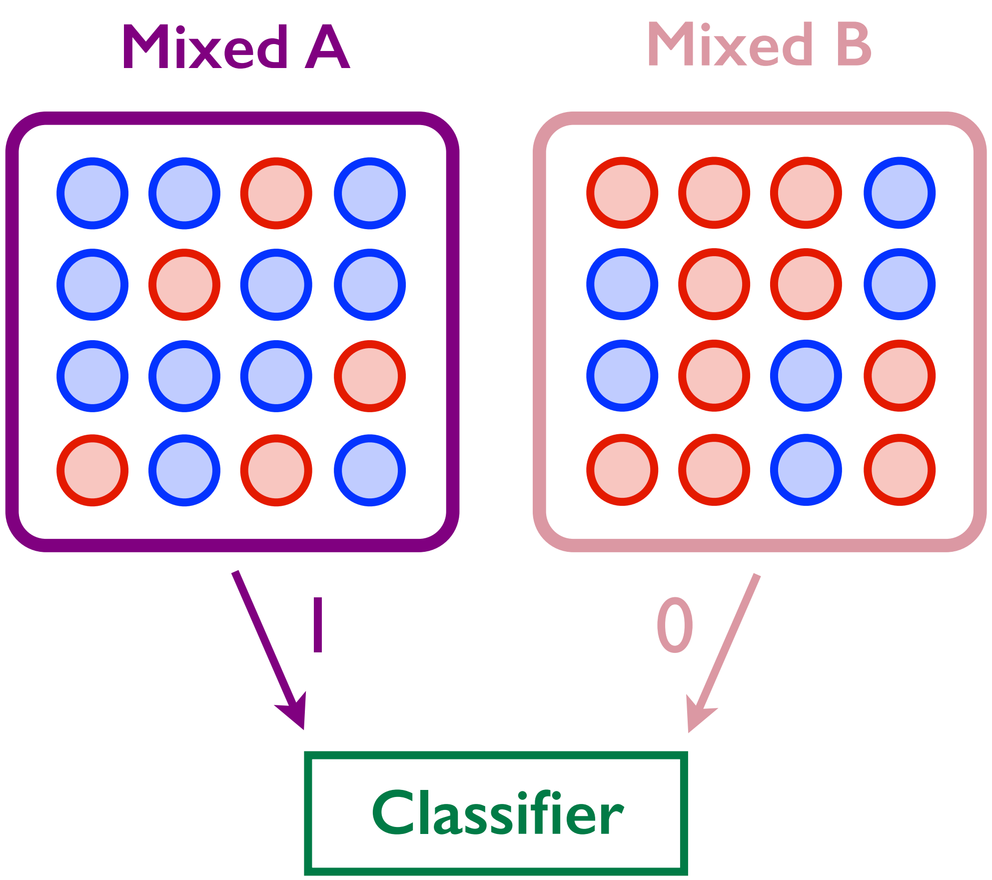


An opportunity to stress test new anomaly detection strategies

Biases from Training on Simulation?

Train directly on mixed data!

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



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

$$\neq$$

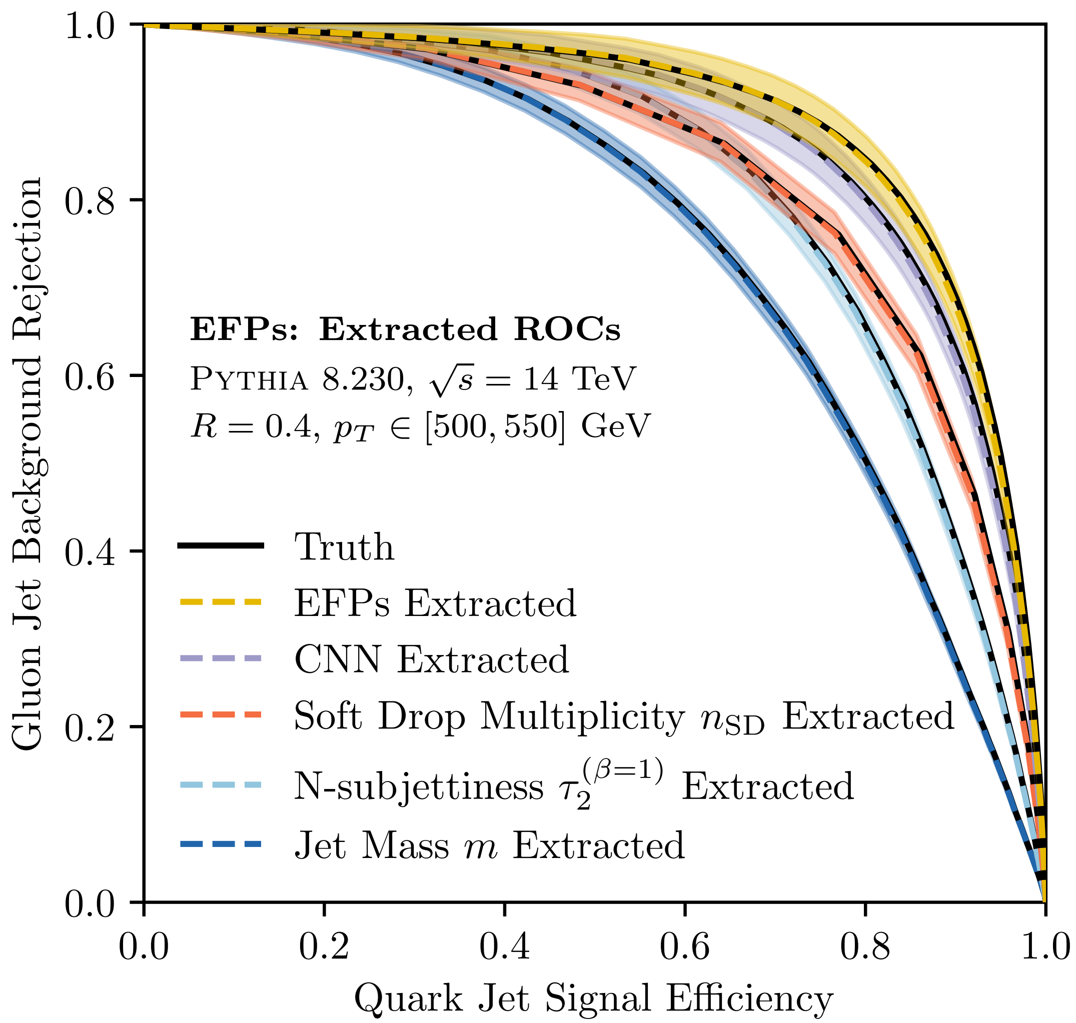
$$h_{\mathrm{pure}}(\vec{x}) = \frac{p_q(\vec{x})}{p_q(\vec{x}) + p_g(\vec{x})}$$

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, Freytsis, Ostdiek, 1706.09451; Komiske, Metodiev, Nachman, Schwartz, 1801.10158; Collins, Howe, Nachman, 1805.02664, 1902.02634]

Biases from Training on Simulation?

Train directly on mixed data!



CWoLa:
Classification Without Labels

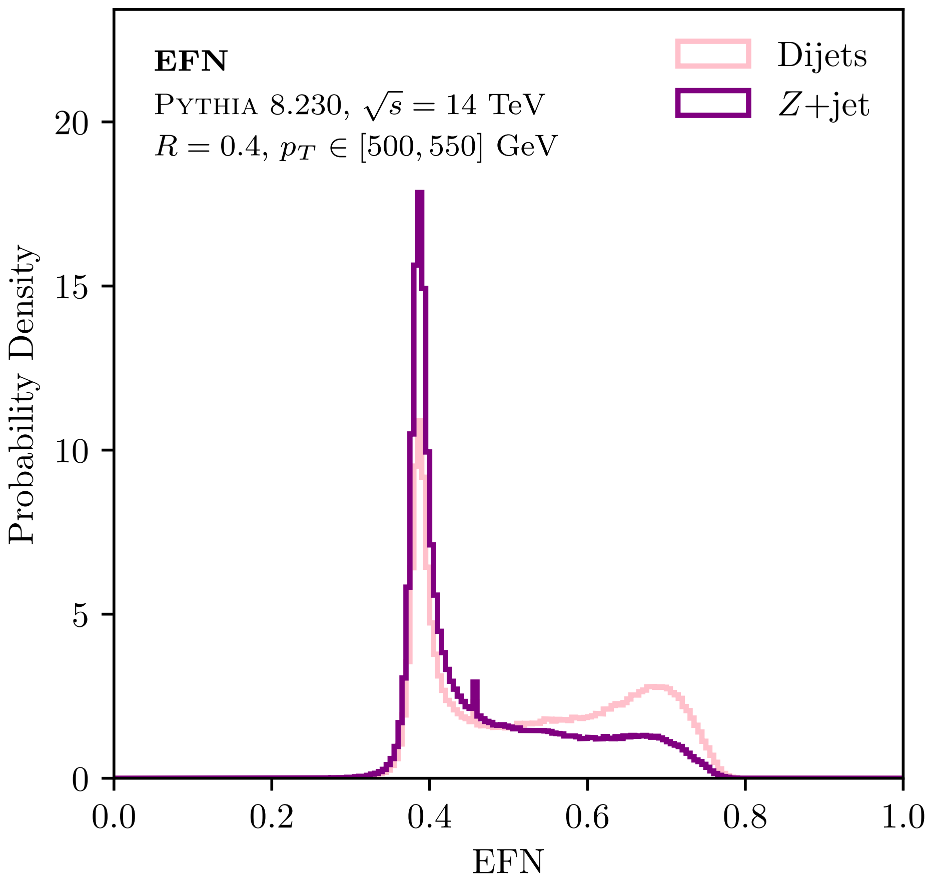
Robust tagging performance
(plus data-driven tricks to calibrate working points and estimate systematic uncertainties)

[Komiske, Metodiev, JDT, 1809.01140]

[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, Freytsis, Ostdiek, 1706.09451; Komiske, Metodiev, Nachman, Schwartz, 1801.10158; Collins, Howe, Nachman, 1805.02664, 1902.02634]

Ambiguous Definition of Jet Categories?

Use classifiers to define categories!



Extract optimal jet categories from data, solely from assumption they exist (!)*

Mutual Irreducibility

"Anchor bins":
Pure representatives exist for each category (even if very rare)

Sample Independence

Mixed samples have different category fractions but same category properties

$$h_{\text{pure}} \in [0, 1]$$

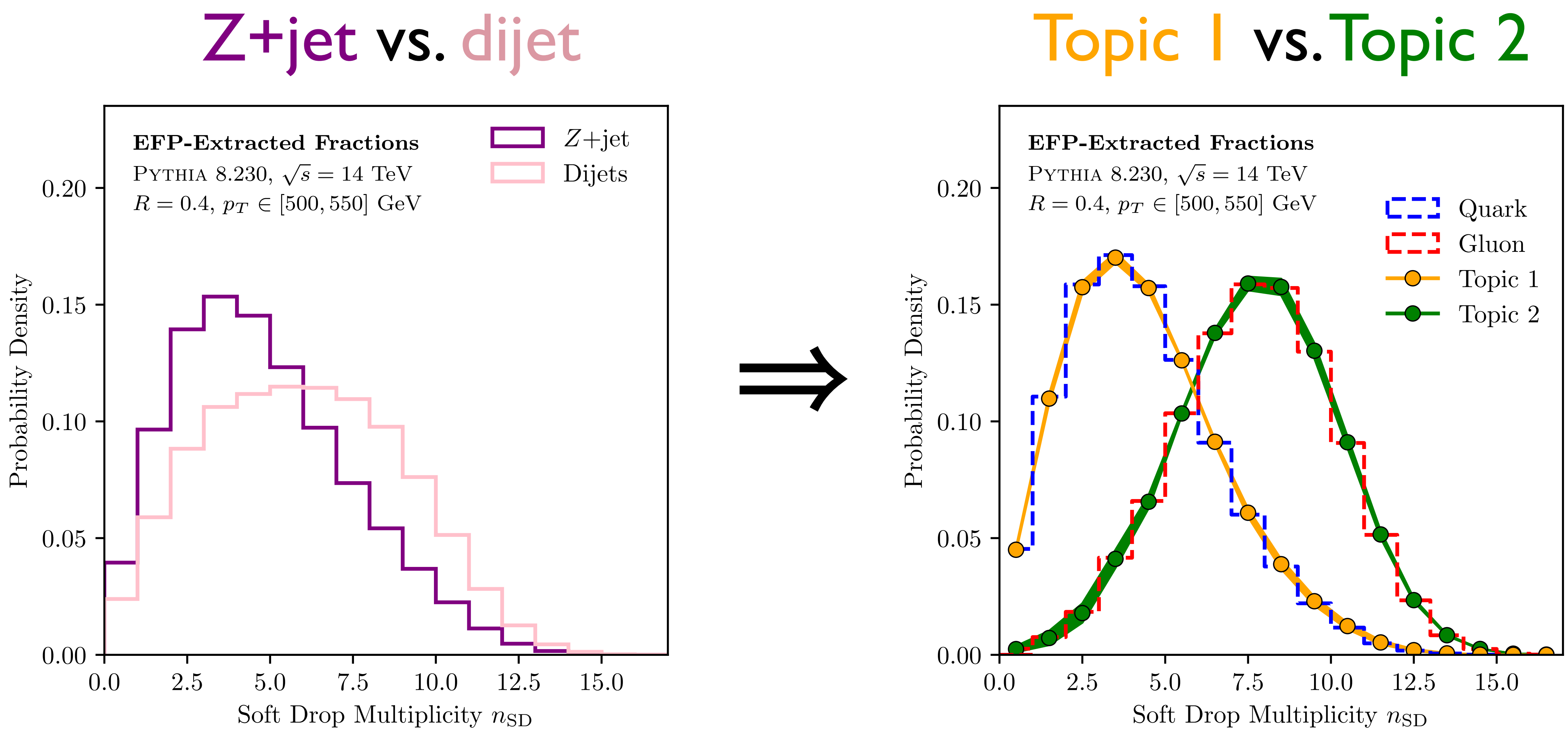
$$h_{\text{mixed}} \in \left[\frac{f_g^A}{f_g^A + f_g^B}, \frac{f_q^A}{f_q^A + f_q^B}\right]$$

[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140; using Katz-Samuels, Blanchard, Scott, 1710.01167]

Ambiguous Definition of Jet Categories?

Use classifiers to define categories!

Z+jet vs. dijet → Topic 1 vs. Topic 2



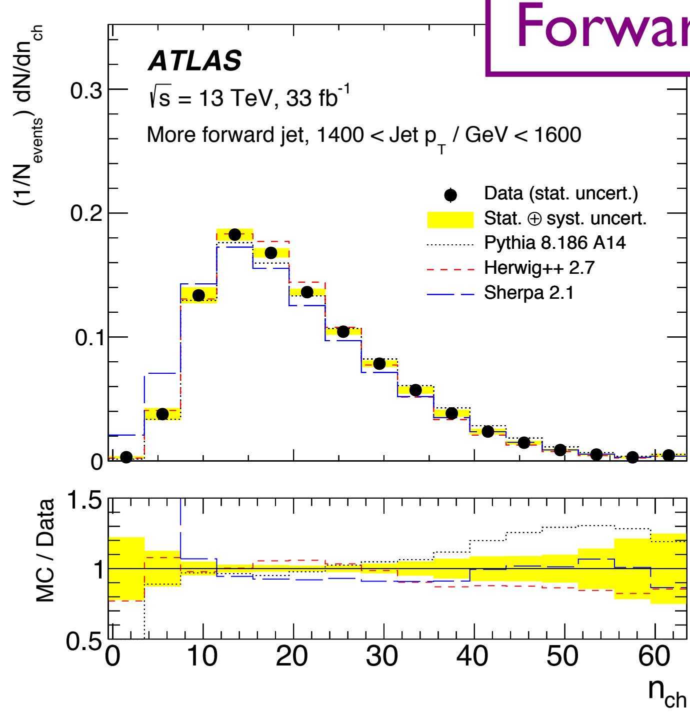
⇒ Operational Definition of "Quark" vs. "Gluon"

[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140; using Katz-Samuels, Blanchard, Scott, 1710.01167]

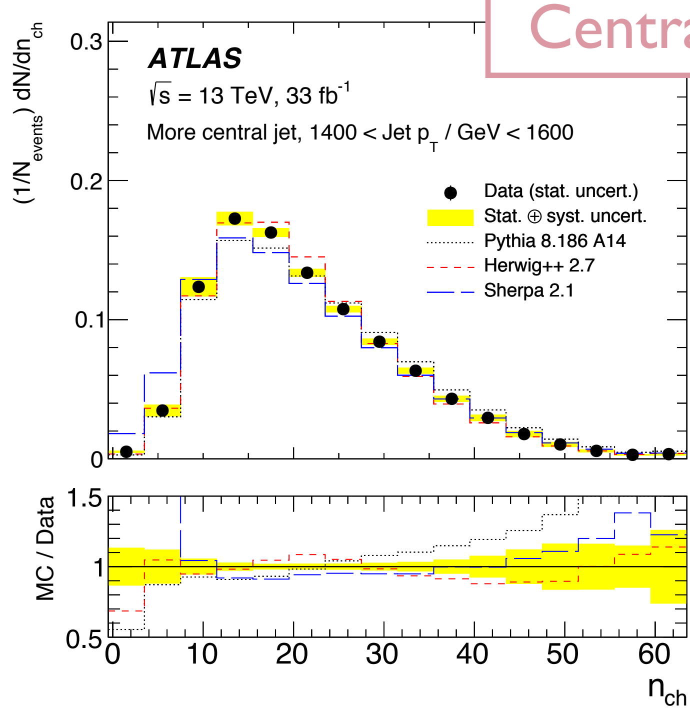
"Ok, but do you really think these techniques will ever be applied to real LHC data?"

First Jet Topics Result from ATLAS

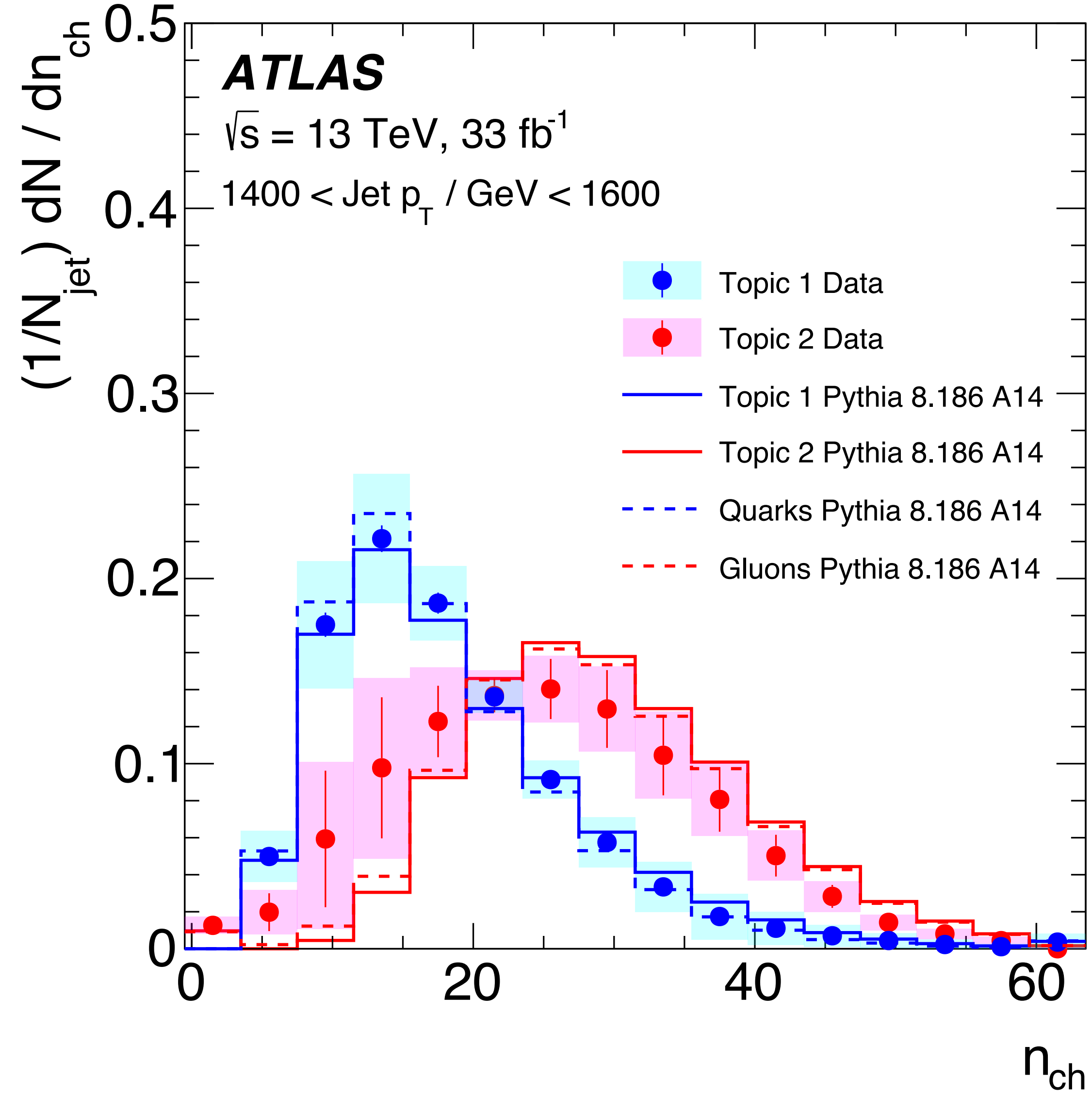
Forward



Central



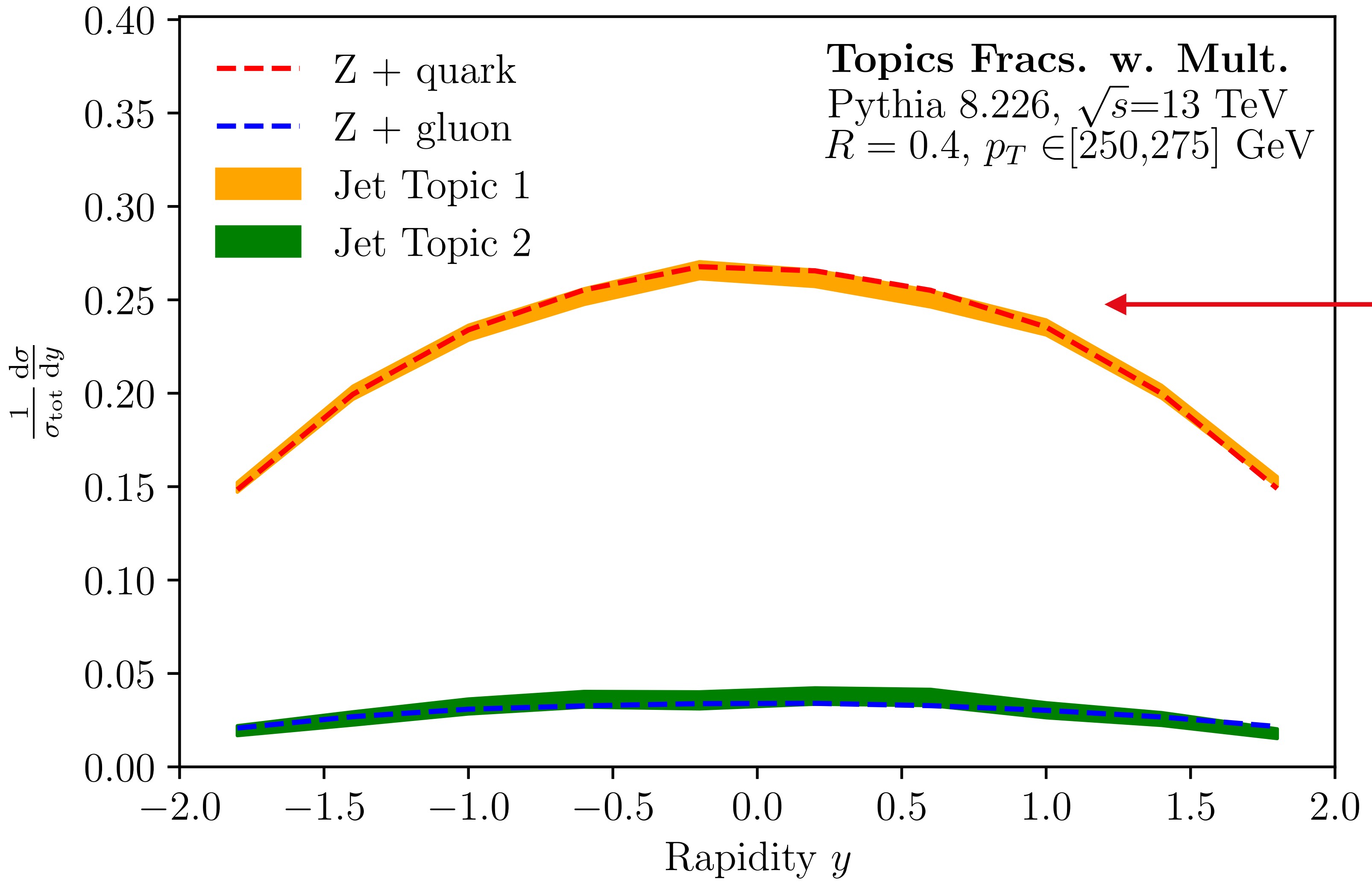
Track multiplicity for "Topic 1" and "Topic 2"

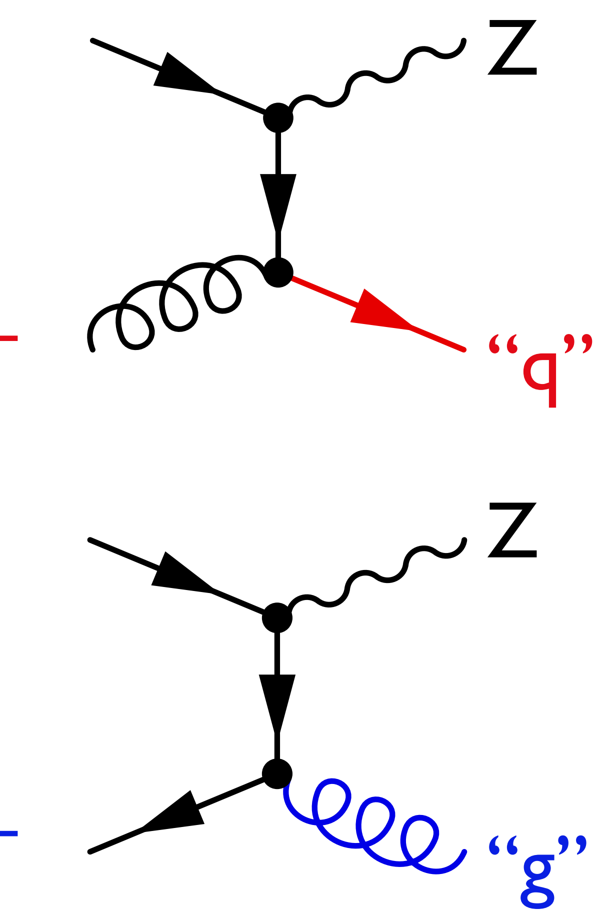


[ATLAS, 1906.09254]

“Parton”-Labeled Cross Sections?

Potential boon for PDF extraction at colliders



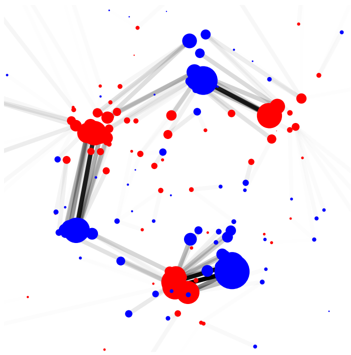


Key Challenges:
- Sample dependence from color coherence
- Limited statistics in anchor bin region
- Defining jet topics at fixed order

[Metodiev, JDT, 1802.00008]

Into the Network

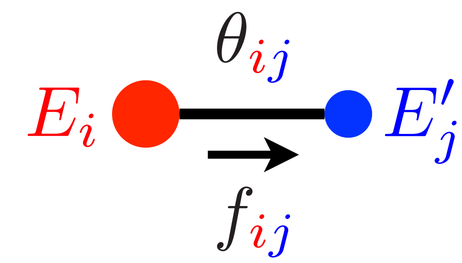
Data Ex Machina

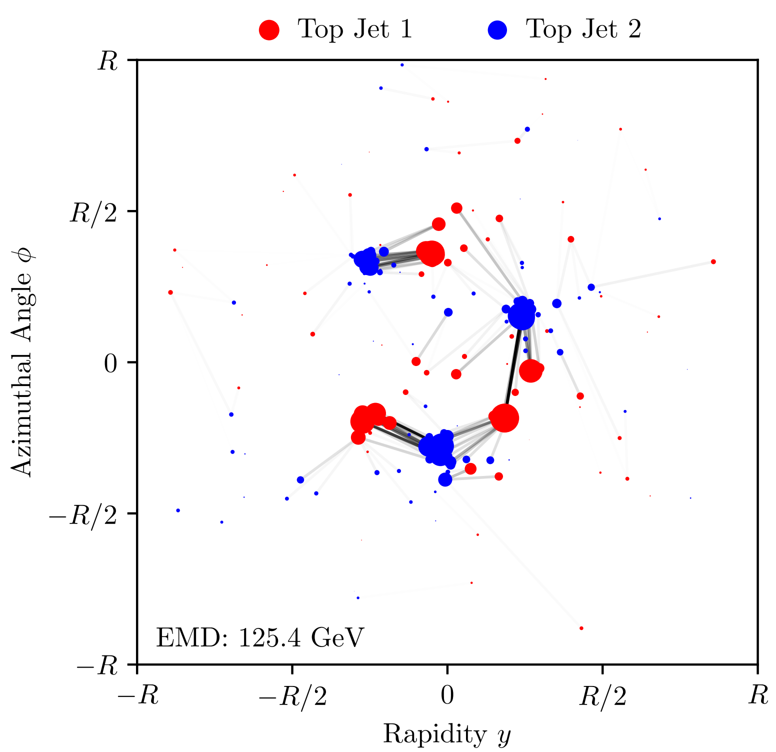


(The Space of Jets)

The Energy Mover's Distance

Closely related to 1-Wasserstein metric





Optimal transport between energy flows…

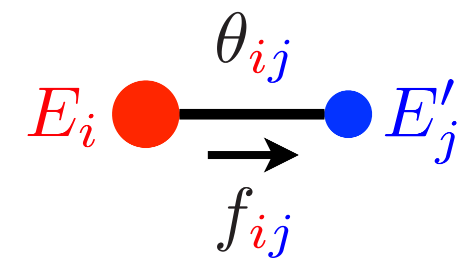
$$\mathrm{EMD}(\mathcal{E}, \mathcal{E}') = \underbrace{\min_{\{f\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R}}_{\text{Cost to move energy}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost to create energy}}$$

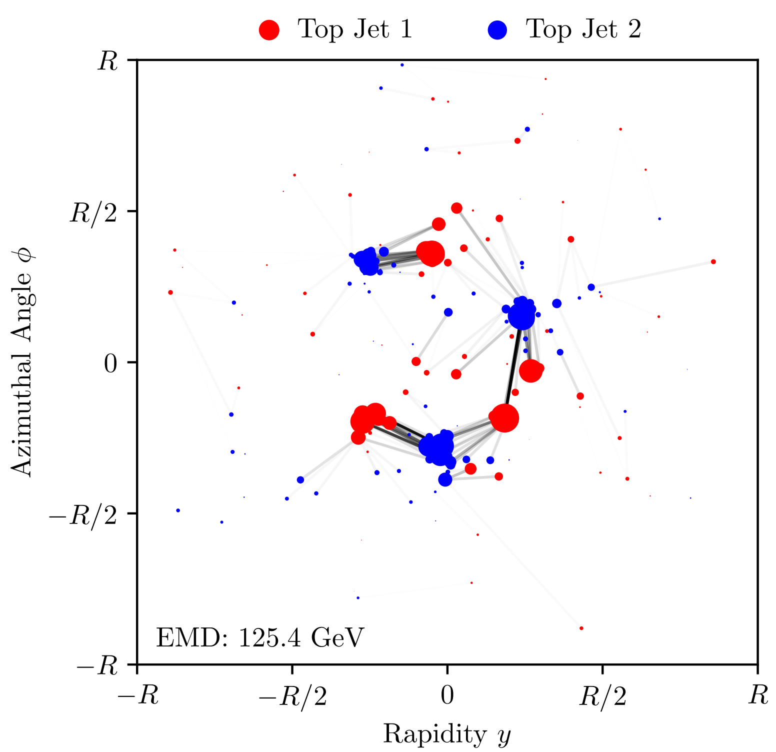
in GeV

[Komiske, Metodiev, JDT, 1902.02346;
see also Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele Taskar, GSI 2013]

The Energy Mover's Distance

Closely related to 1-Wasserstein metric





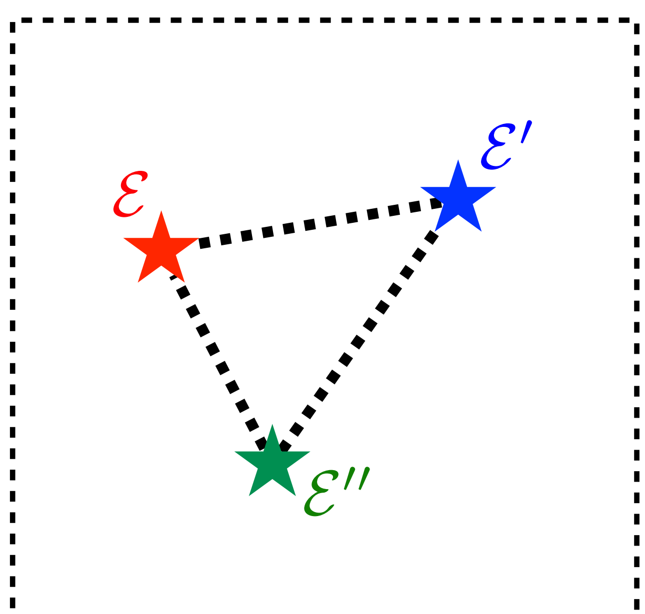
Optimal transport between energy flows…

$$\mathrm{EMD}(\mathcal{E},\mathcal{E}') = \min_{\{f\}} \underbrace{\sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R}}_{\text{Cost to move energy}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost to create energy}}$$

in GeV

Cost to move energy

Cost to create energy



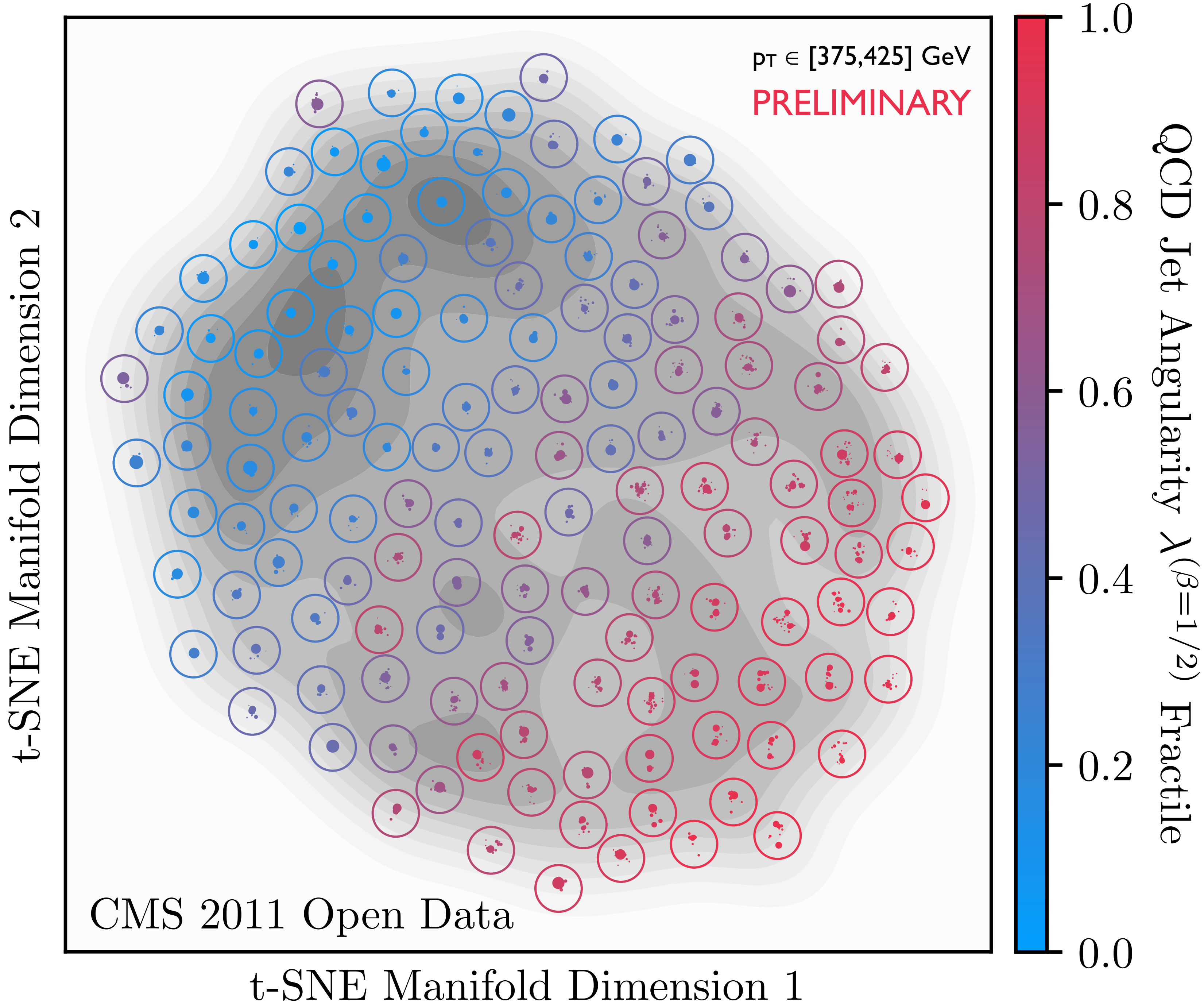
…defines a metric on the space of events

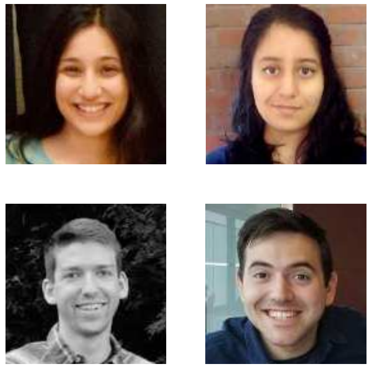
$$0 \le \mathrm{EMD}(\mathcal{E},\mathcal{E}') \le \mathrm{EMD}(\mathcal{E},\mathcal{E}'') + \mathrm{EMD}(\mathcal{E}',\mathcal{E}'')$$

(assuming R ≥ $\theta_{max}/2$, i.e. R ≥ jet radius for conical jets)

[Komiske, Metodiev, JDT, 1902.02346;
see also Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele Taskar, GSI 2013]

The Space of Quark/Gluon Jets

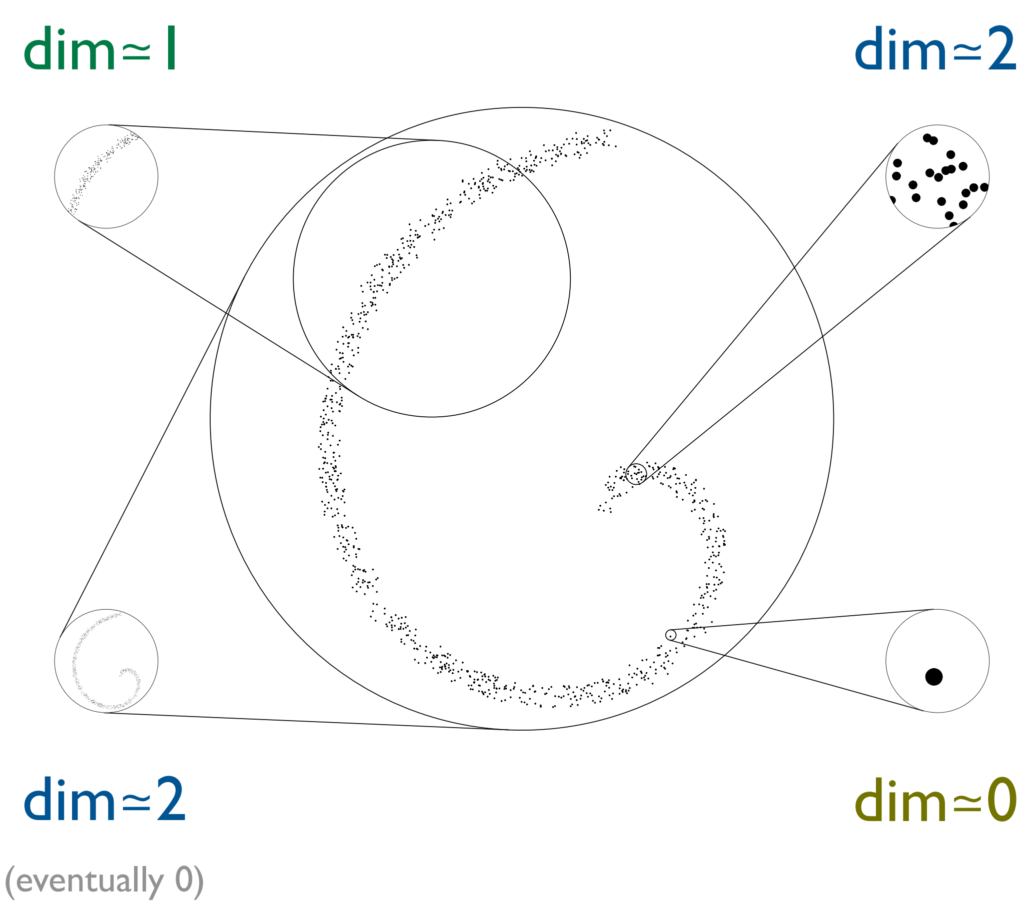




[Komiske, Mastandrea, Metodiev, Naik, JDT, in preparation]

Quantifying Dimensionality

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



$$N_{\text{neighbors}}(r) \sim r^{\dim}$$

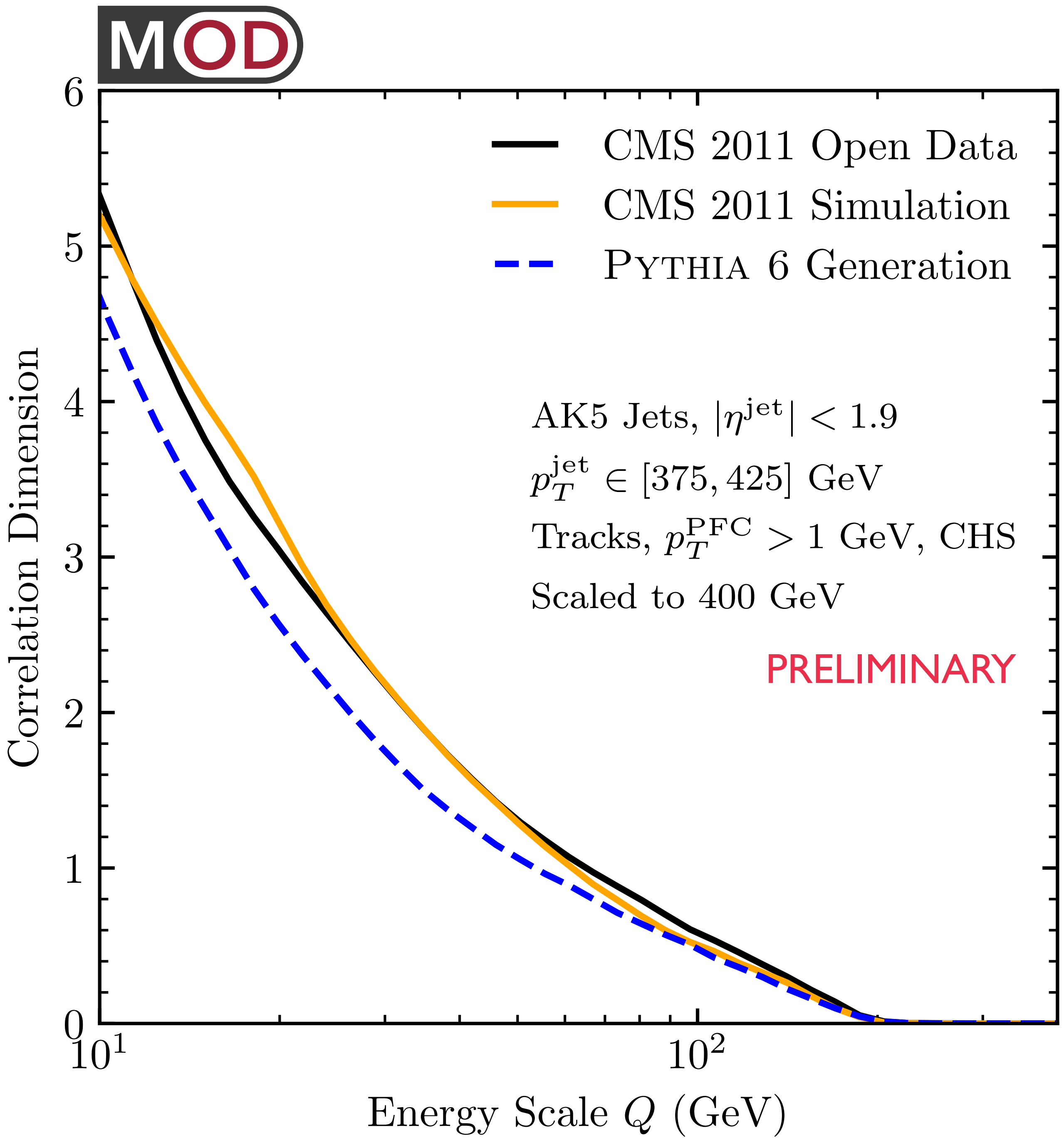
$$\Downarrow$$

$$\dim(r) \sim r \frac{\partial}{\partial r} \ln N_{\text{neighbors}}(r)$$

[Grassberger, Procaccia, PRL 1983; Kégl, NIPS 2002]

The Dimension of Quark/Gluon Jets

Fascinating impact of detector effects






[Komiske, Mastandrea, Metodiev, Naik, JDT, in preparation]

Preliminary Calculation

Leading Log:

(single log, since dim has derivative)

$$\dim_i(Q) \simeq -\frac{8\alpha_s}{\pi} C_i \ln\frac{Q}{p_T}$$

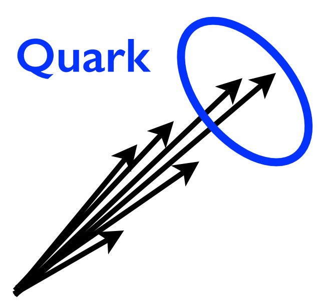
Color Factor

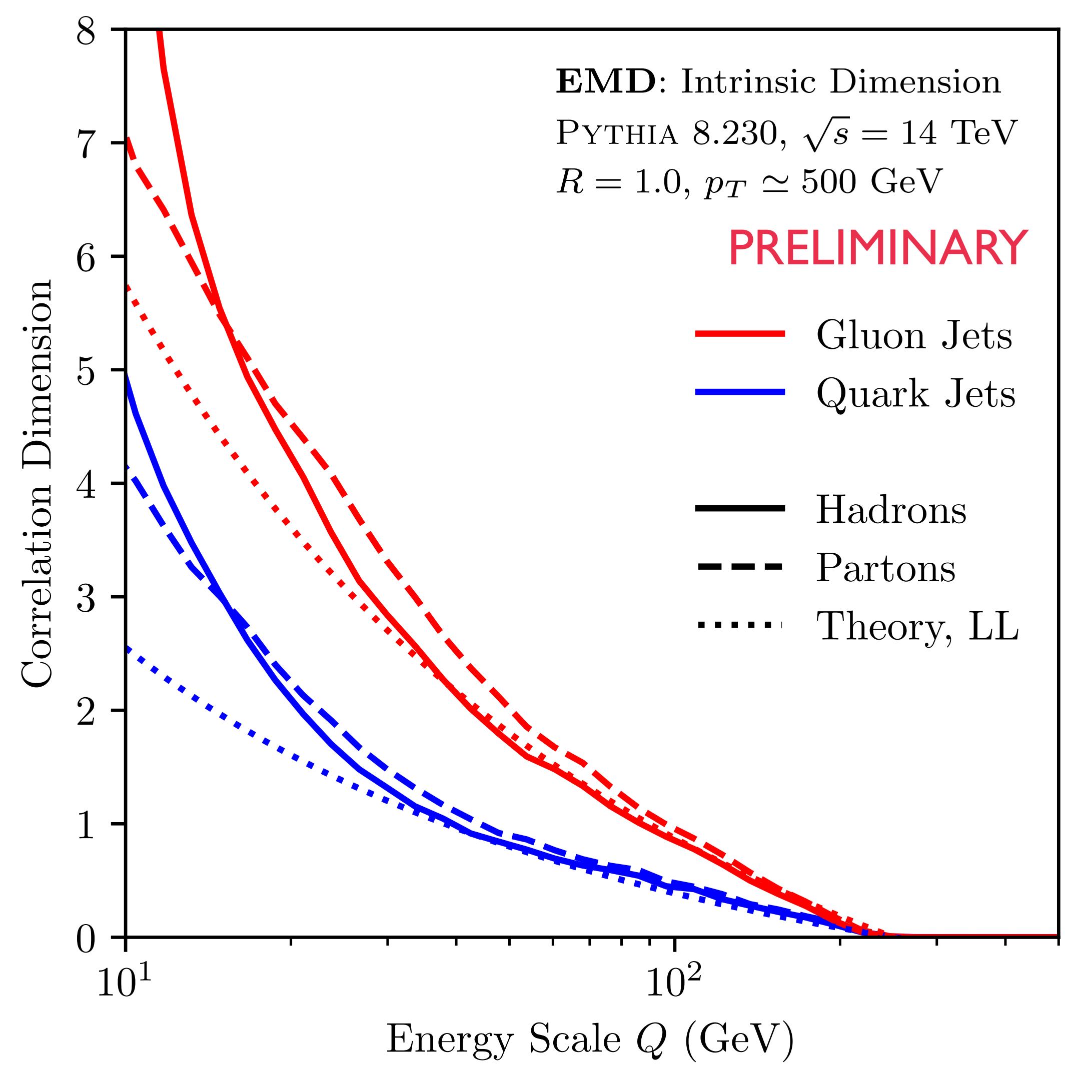
$C_A = 3$

Gluon

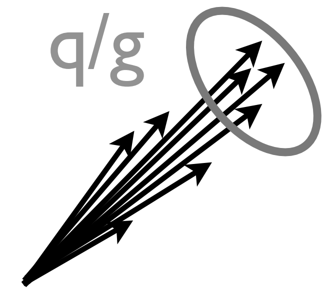
$C_F = 4/3$

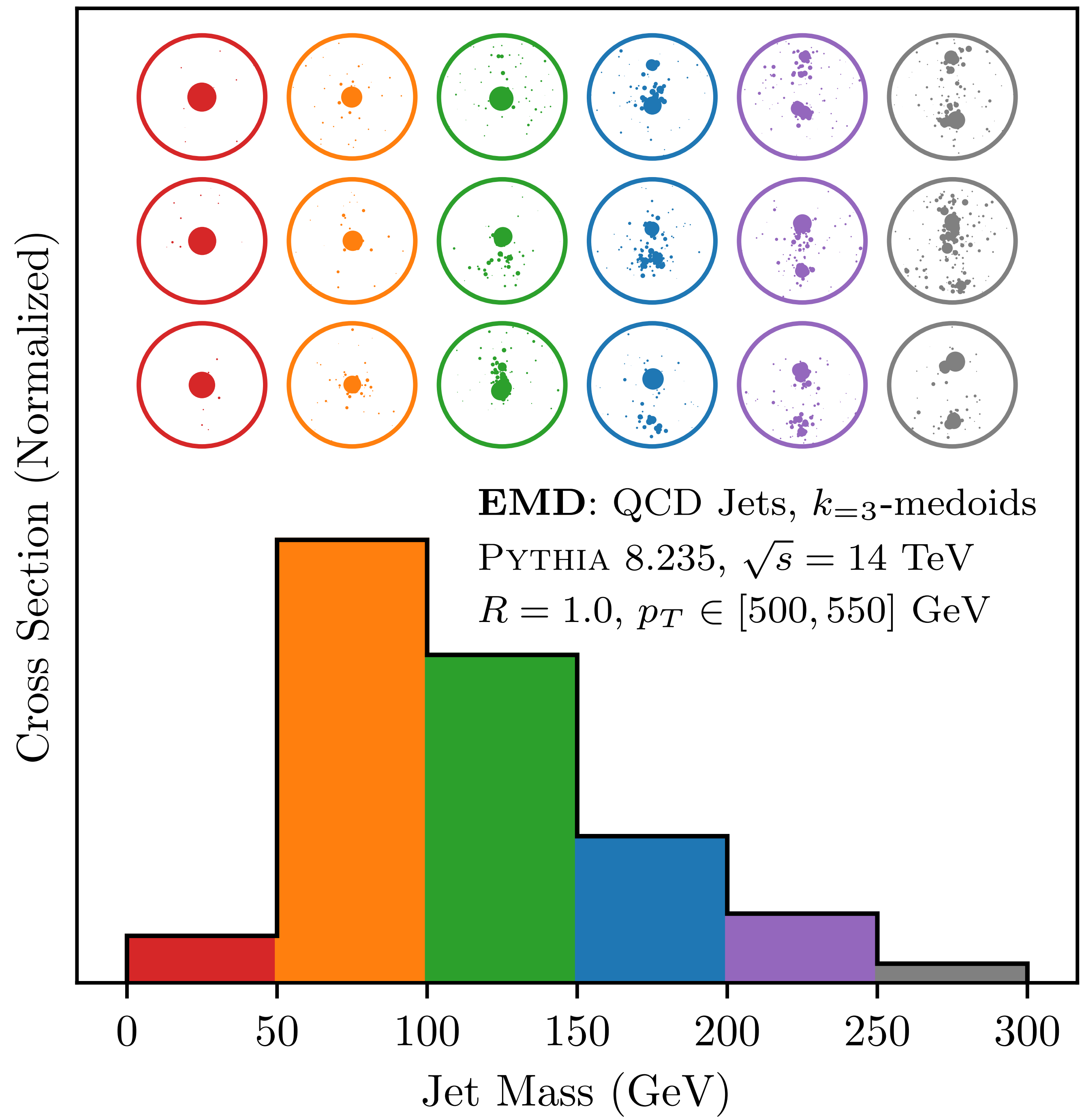
Quark



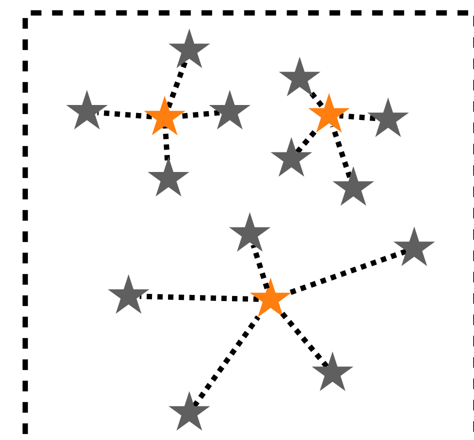


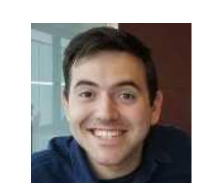
Histograms meet Event Displays





3-medoid: Three most representative jets in each bin



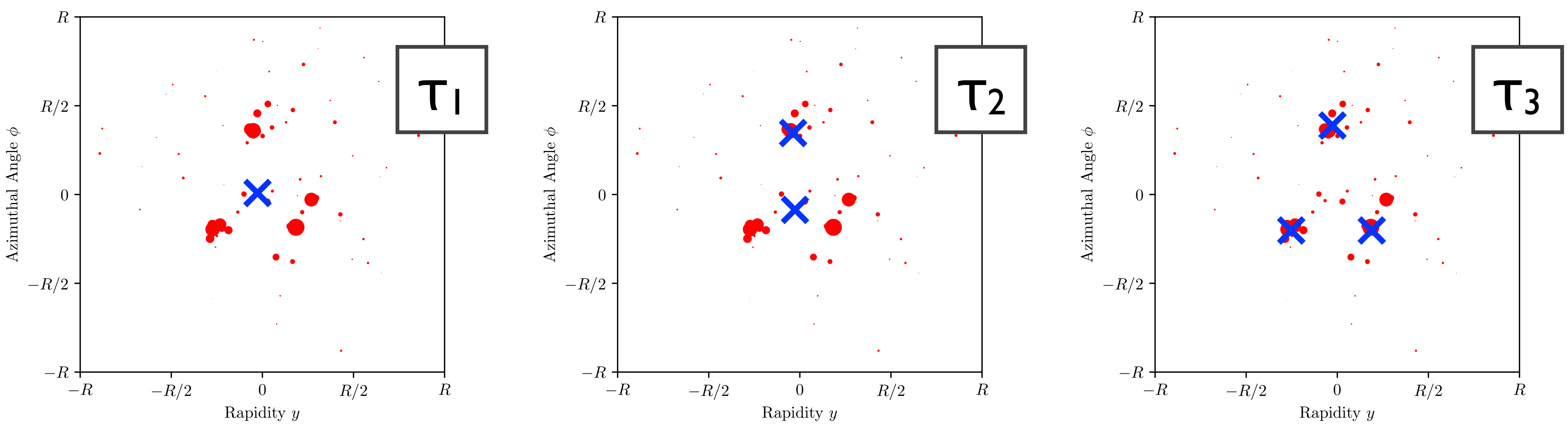
[Komiske, Metodiev, JDT, 1902.02346]

Insight into N-subjettiness

$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N \text{ axes}} \sum_i E_i \min\left\{\theta_{1,i}^\beta, \theta_{2,i}^\beta, \ldots, \theta_{N,i}^\beta\right\}$$

IRC safe

kind of arbitrary



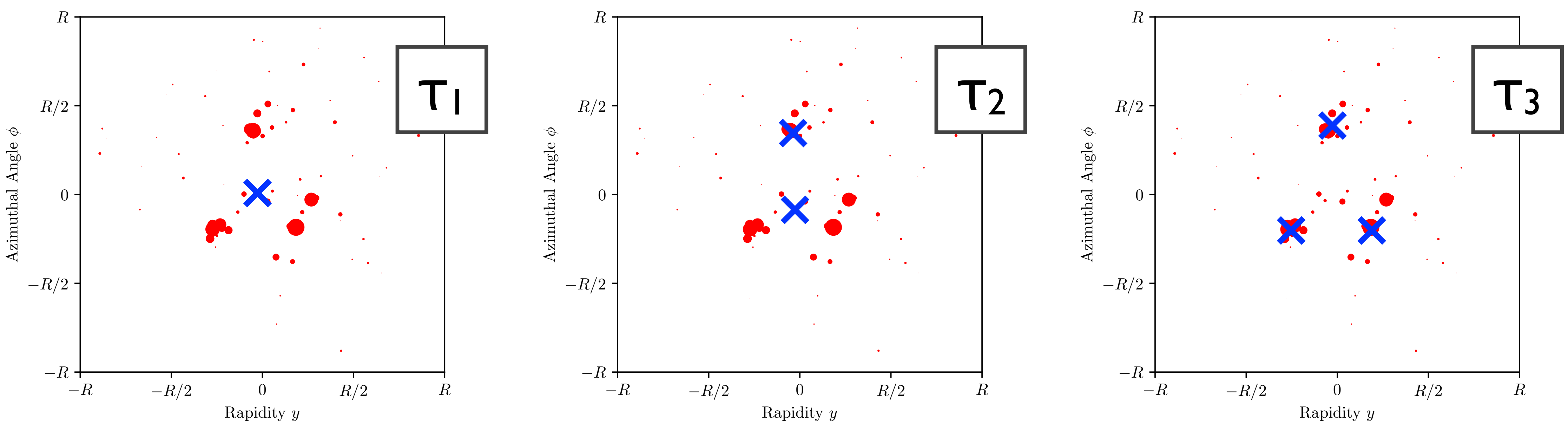
[JDT, Van Tilburg, 1011.2268, 1108.2701;
based on Brandt, Dahmen, ZPC 1979; Stewart, Tackmann, Waalewijn, 1004.2489]

Insight into N-subjettiness

$$\tau_N^{(\beta)}(\mathcal{E}) = \min_{N \text{ axes}} \sum_i E_i \min\left\{\theta_{1,i}^{\beta}, \theta_{2,i}^{\beta}, \ldots, \theta_{N,i}^{\beta}\right\}$$

IRC safe

kind of arbitrary



$$\tau_N(\mathcal{E}) = \min_{|\mathcal{E}'|=N} \mathrm{EMD}(\mathcal{E}, \mathcal{E}')$$

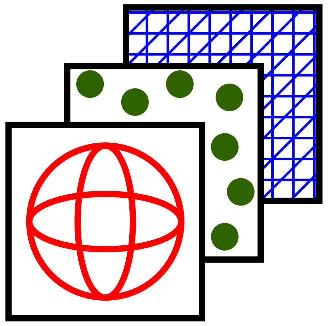
for β = 1

Related to p-Wasserstein metric for p = β > 1

very satisfying

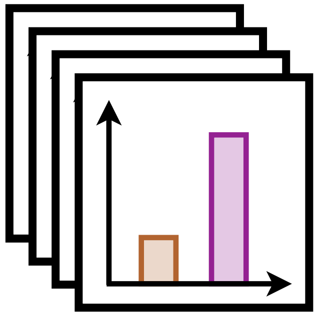
[JDT, Van Tilburg, 1011.2268, 1108.2701; based on Brandt, Dahmen, ZPC 1979; Stewart, Tackmann, Waalewijn, 1004.2489]

Summary



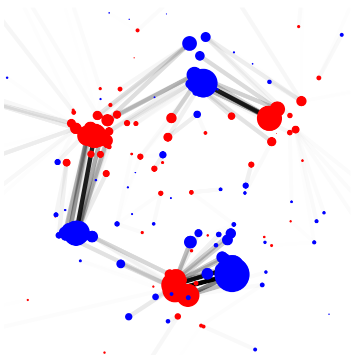
Into the Network

Designing architectures around symmetries and interpretability



Data Ex Machina

Unsupervised learning to interpret hadronic final states



(The Space of Jets)

Computational geometry as a new collider data analysis strategy

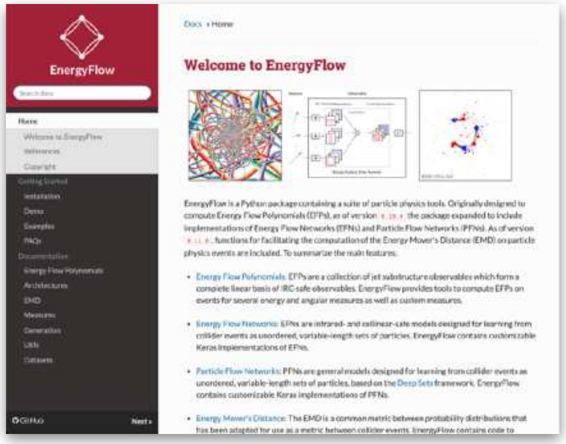
Patrick Komiske

Eric Metodiev

(Theoretical) High Energy Physics

?

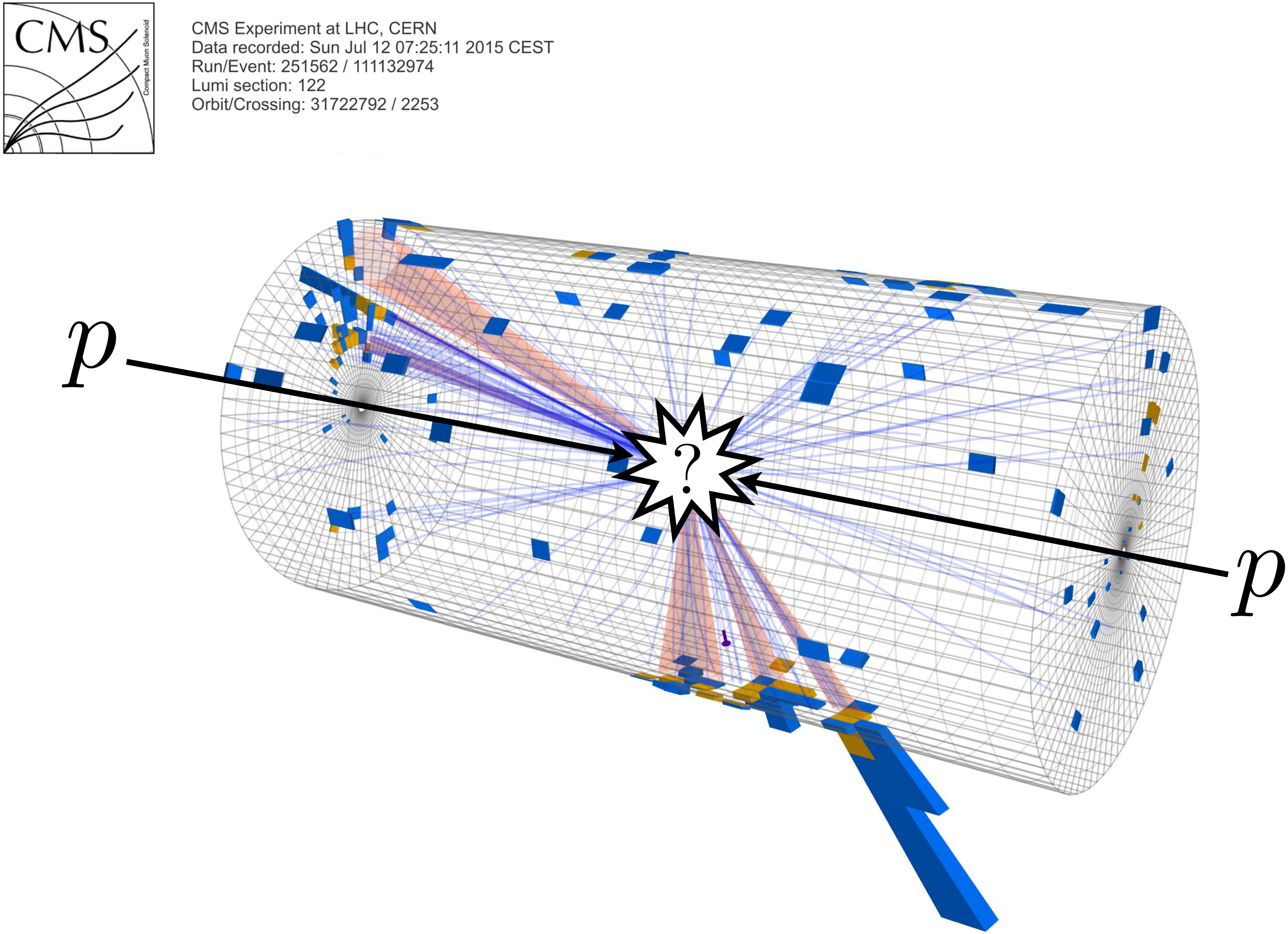
Mathematics, Statistics, Computer Science



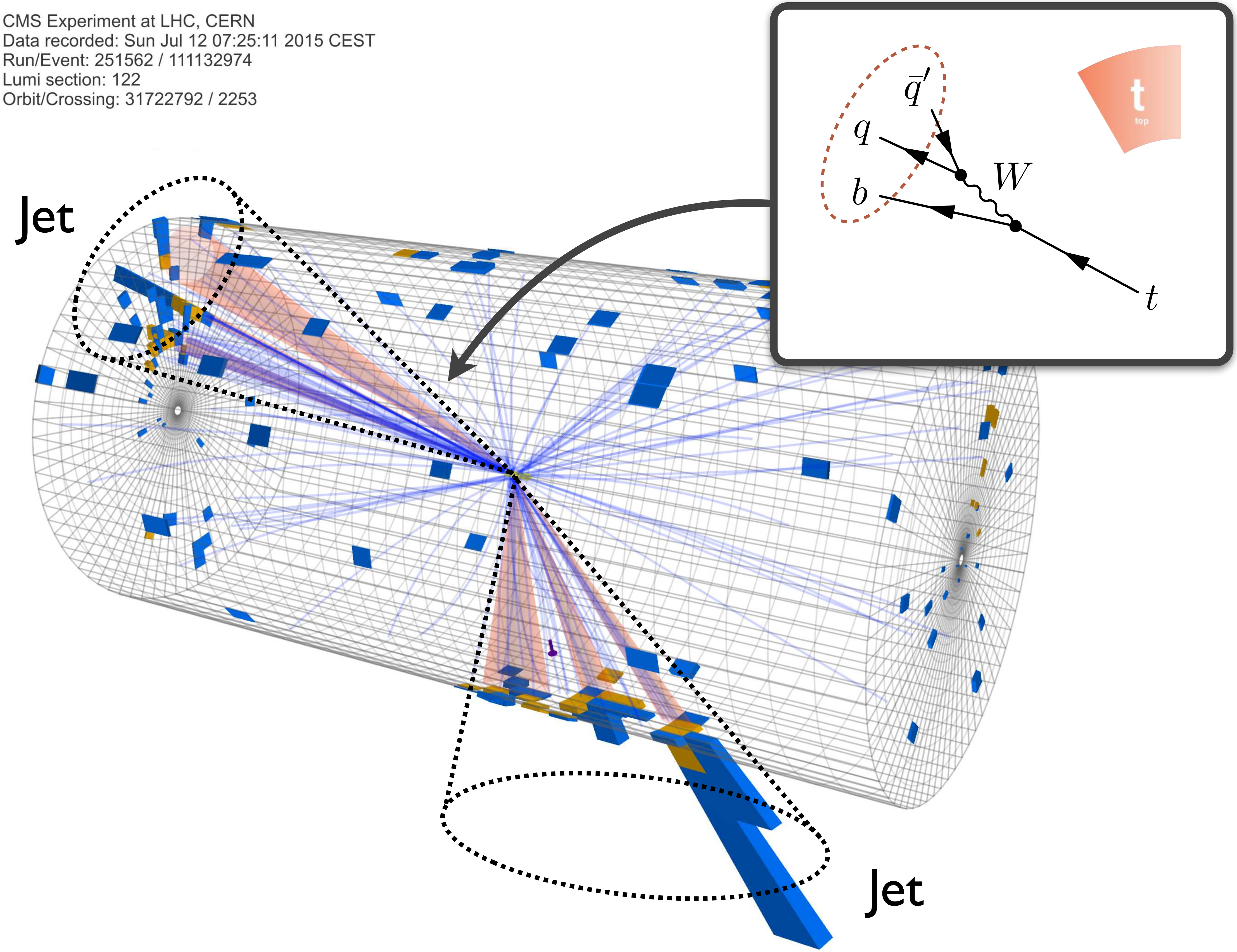
Energy Flow Package

https://energyflow.network/

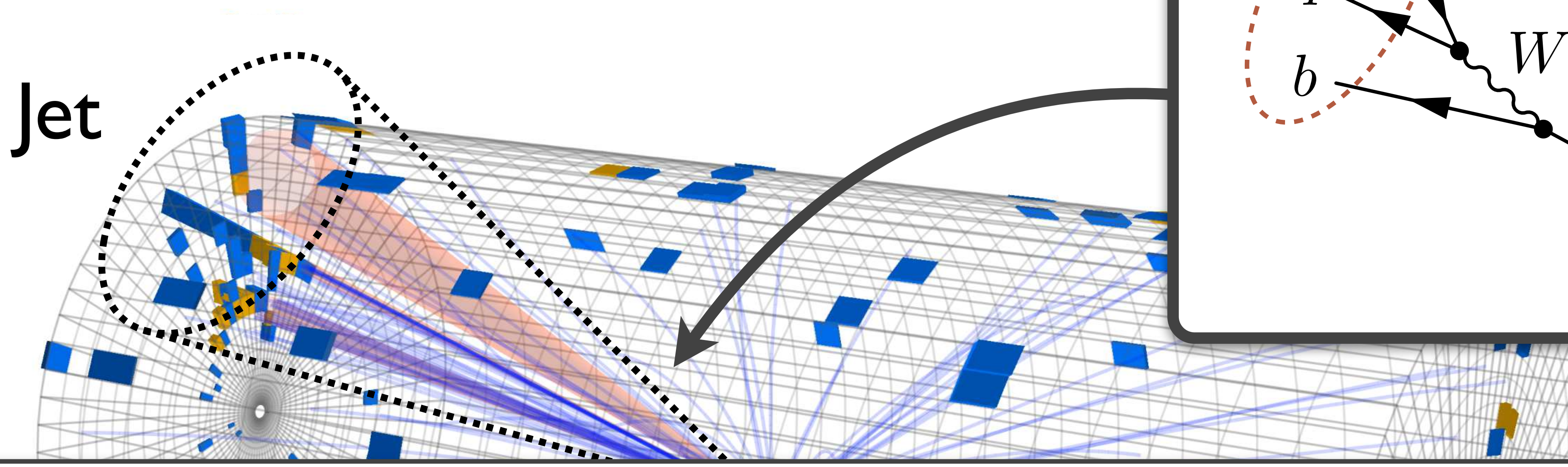
Backup Slides

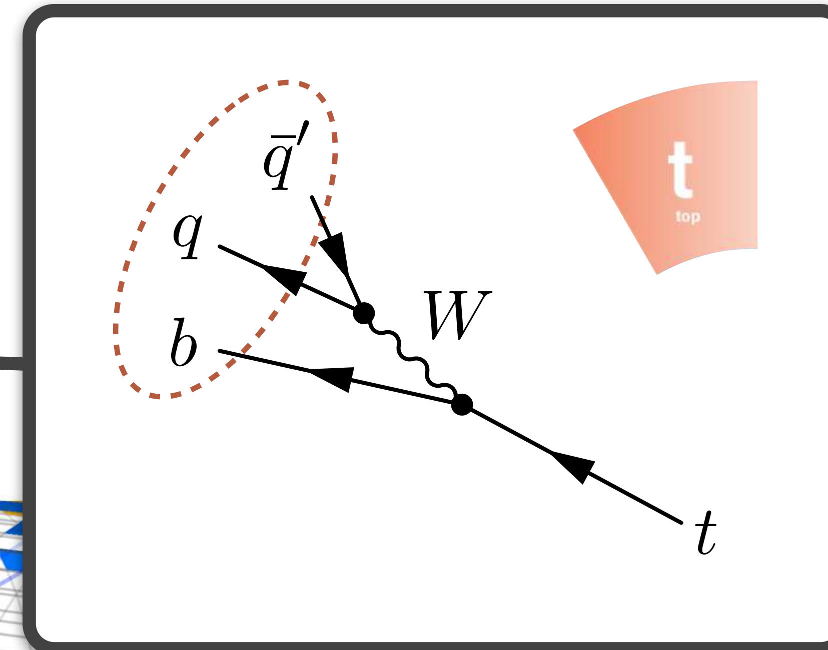

CMS
CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253
p
?
p

CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253



CMS Experiment at LHC, CERN
Data recorded: Sun Jul 12 07:25:11 2015 CEST
Run/Event: 251562 / 111132974
Lumi section: 122
Orbit/Crossing: 31722792 / 2253



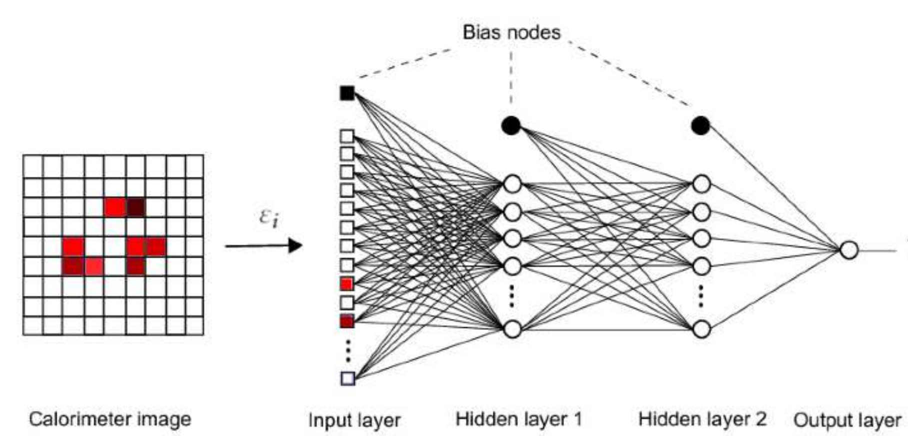


“Deep Thinking”?

$$\tau_N(\mathcal{J}) = \min_{|\mathcal{J}'|=N} \mathrm{EMD}(\mathcal{J}, \mathcal{J}')$$

[e.g. JDT, Van Tilburg, 1011.2268, 1108.2701;
rephrased in language of Komiske, Metodiev, JDT, 1902.02346]

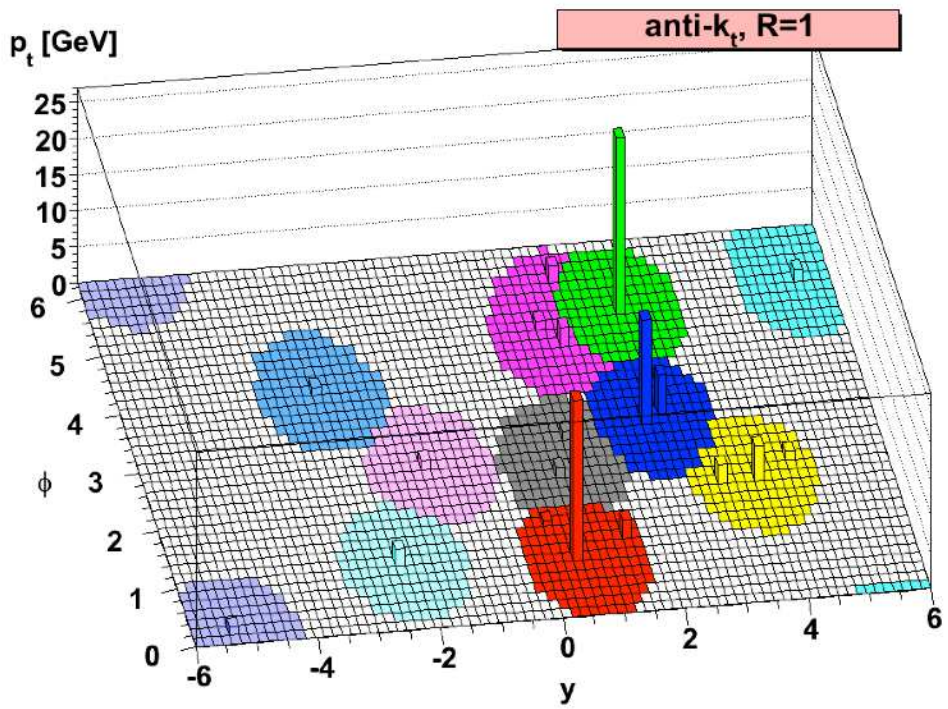
“Deep Learning”?



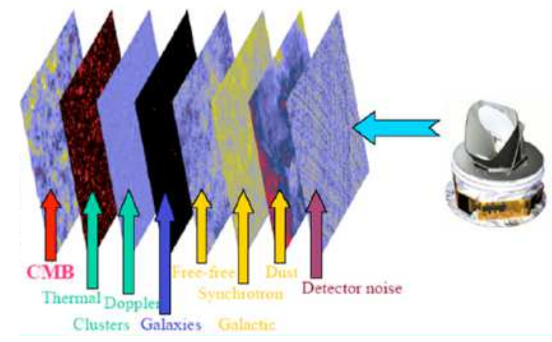
[e.g. Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968;
review in Kasieczka, Plehn, et al., 1902.09914]

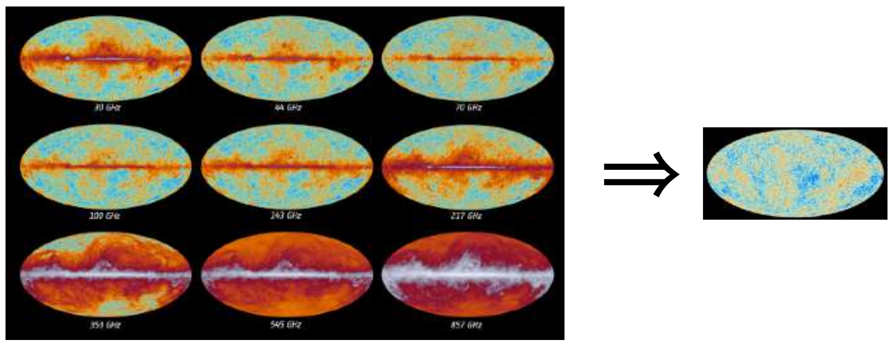
Examples of Unsupervised Learning

Clustering



Topic Modeling



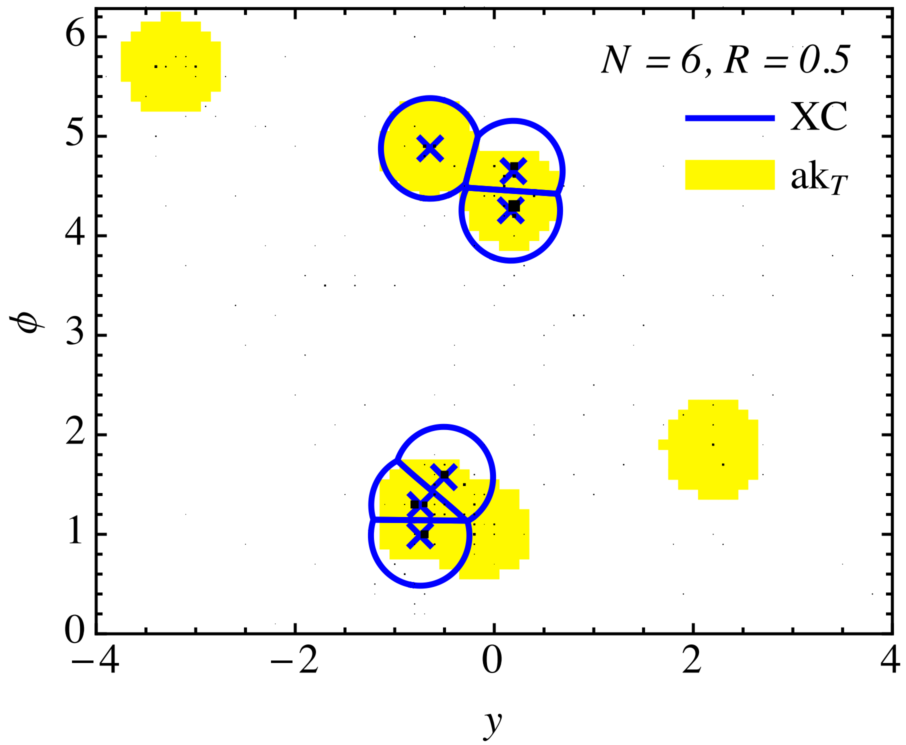


(Approximate) solutions to properly specified problems

[figures from Cacciari, Salam, Soyez, 0802.1189; Planck Outreach]

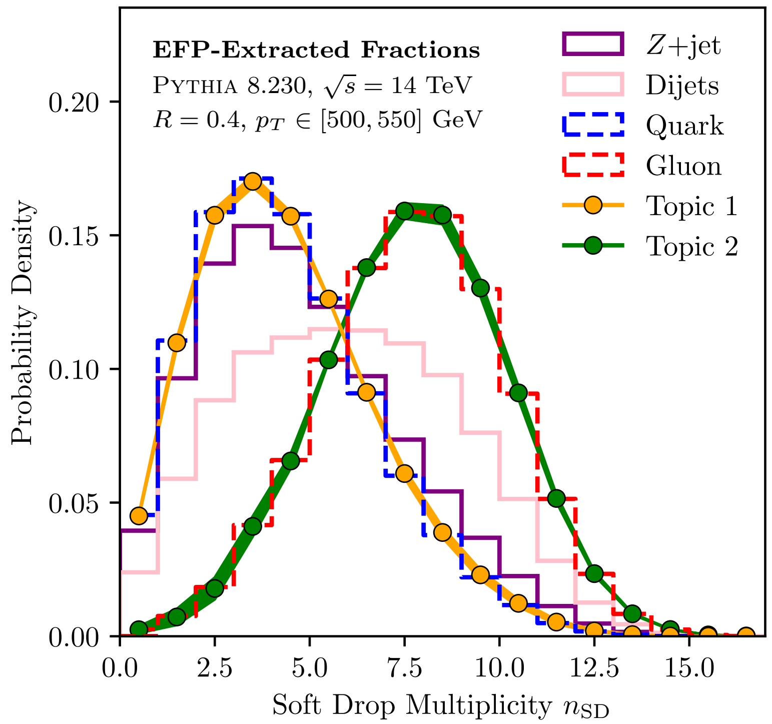
Examples of Unsupervised Learning

XCone Jet Finding



"Find N axes that minimize N-jettiness"

Jet Topics



"Find two mutually irreducible distributions"

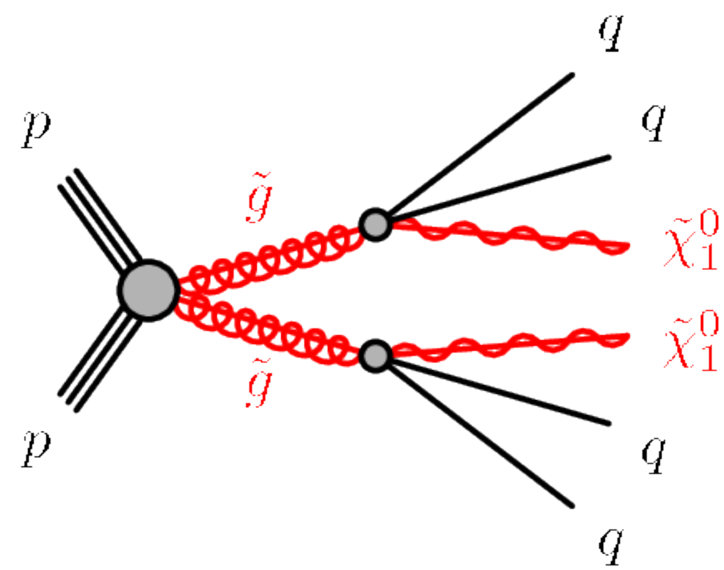
[Stewart, Tackmann, JDT, Vermilion, Wilkason, 1508.01516; based on Stewart, Tackmann, Waalewijn, 1004.2489]
[Metodiev, JDT, 1802.00008; Komiske, Metodiev, JDT, 1809.01140; see also Dillon, Faroughy, Kamenik, 1904.04200]

(Approximate) solutions to properly specified problems

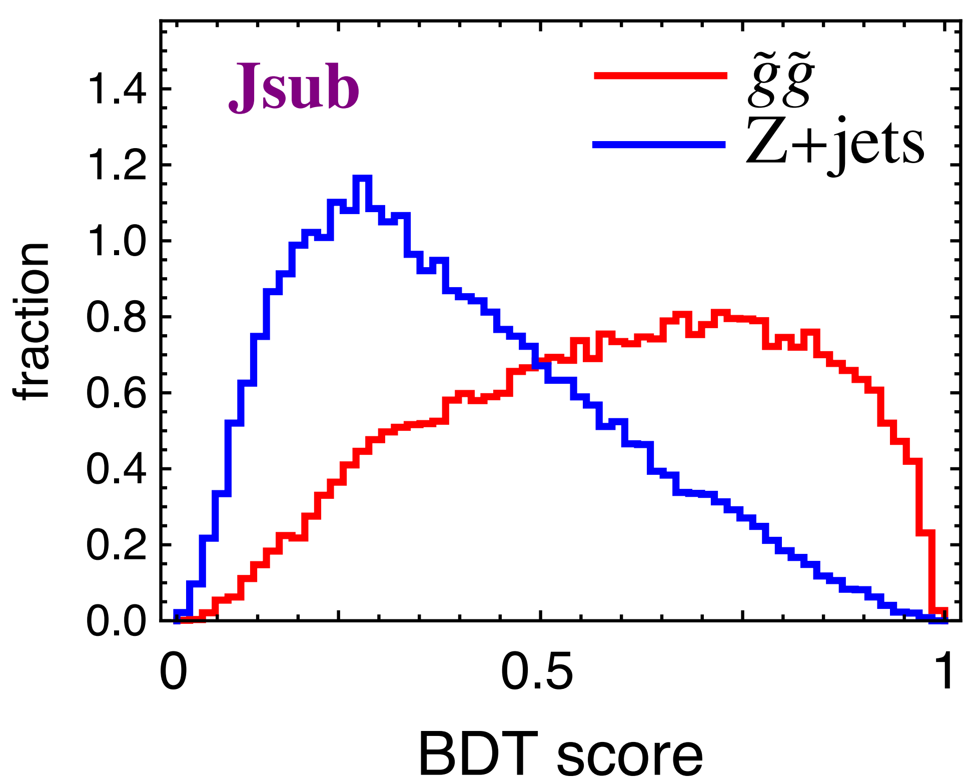
[figures from Cacciari, Salam, Soyez, 0802.1189; Planck Outreach]

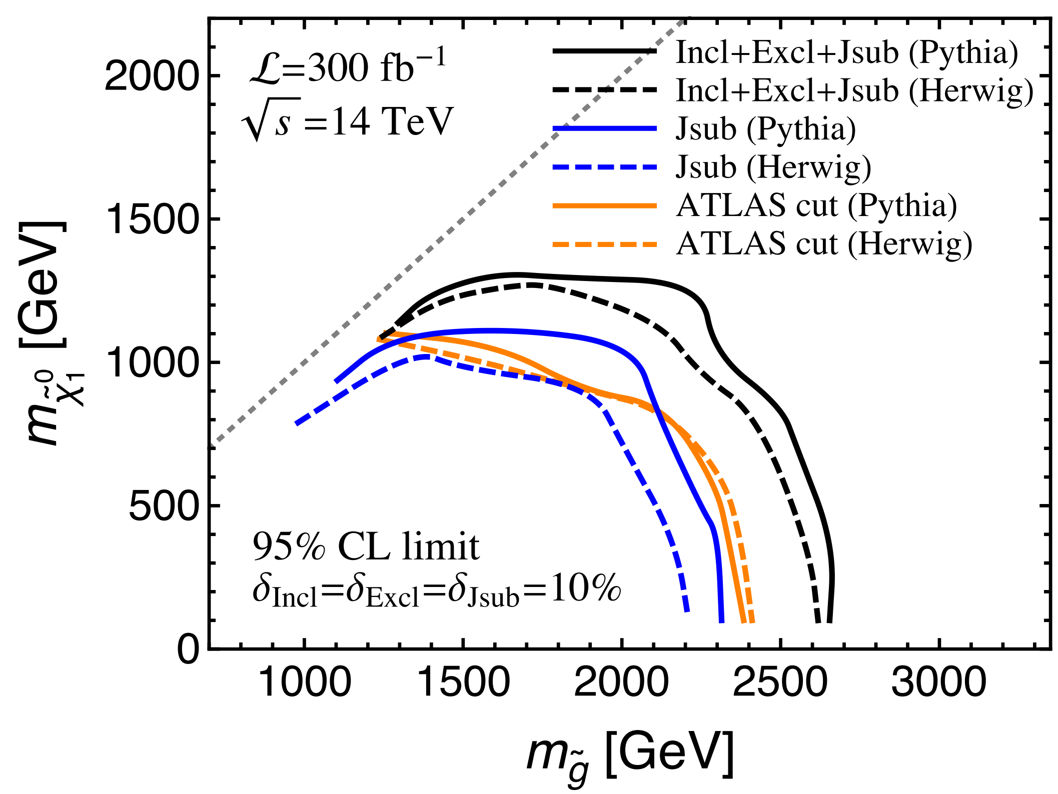
E.g. SUSY Search for Gluino Pairs

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 →

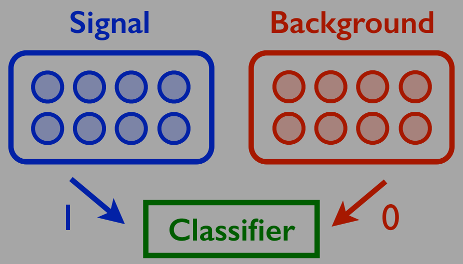




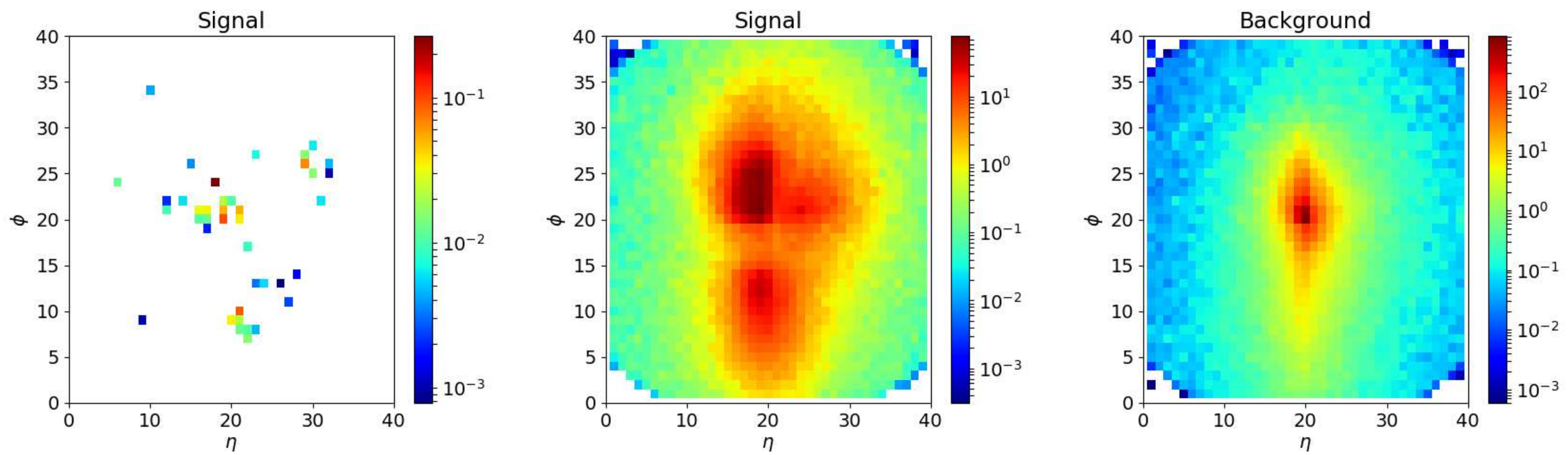
[Bhattacherjee, Mukhopadhyay, Nojiri, Sakakie, Webber, 1609.08781]

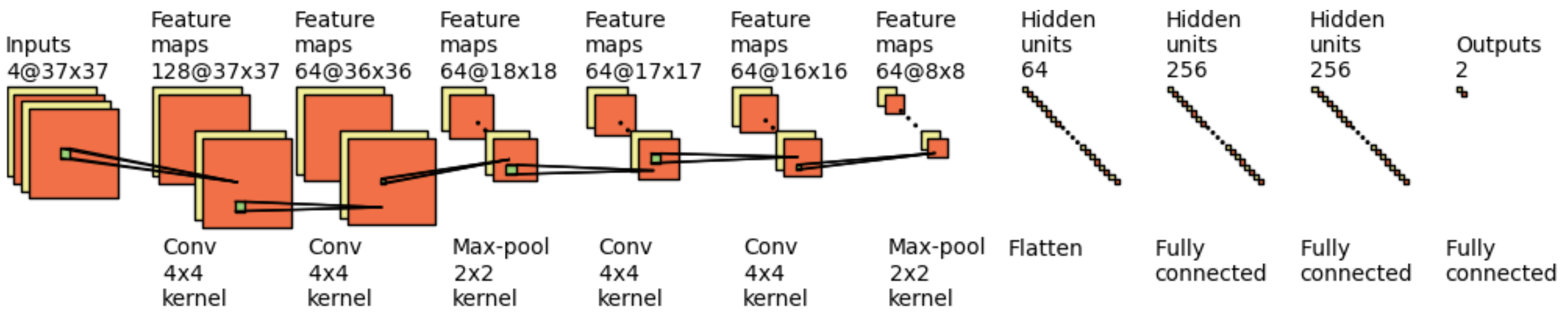
Jet Classification Studies

Mix and match



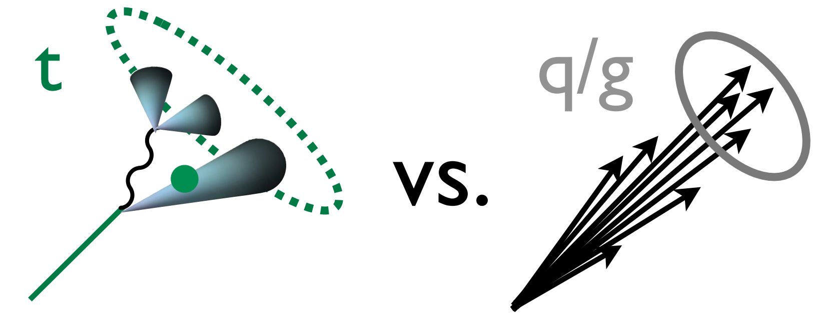
Deep Learning: Jet Image Strategy with CNNs

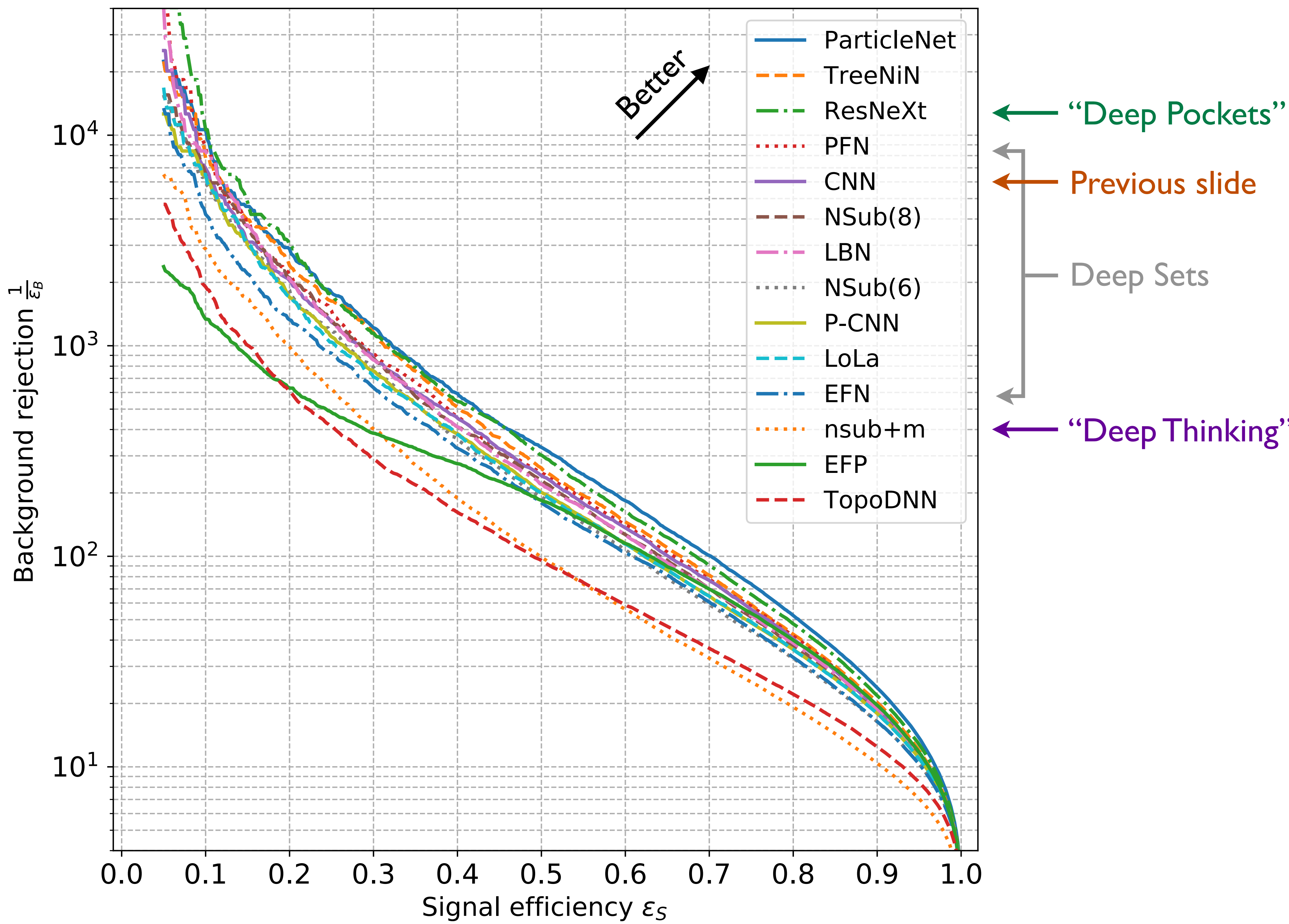




[Macaluso, Shih 1803.00107; building off Kasieczka, Plehn, Russell, Schell, 1701.08784; based on Cogan, Kagan, Strauss, Schwartzman, 1407.5675; de Oliveira, Kagan, Mackey, Nachman, Schwartzman, 1511.05190]

Throwing Down the Gauntlet

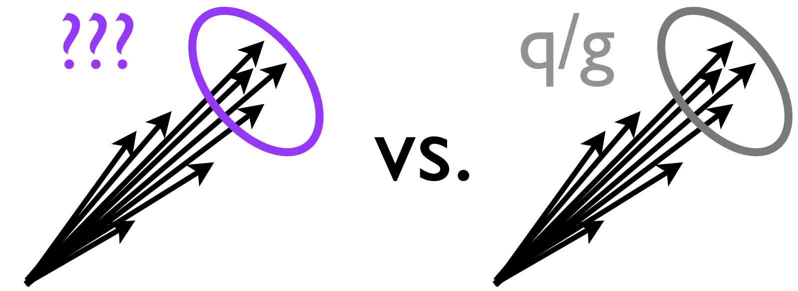


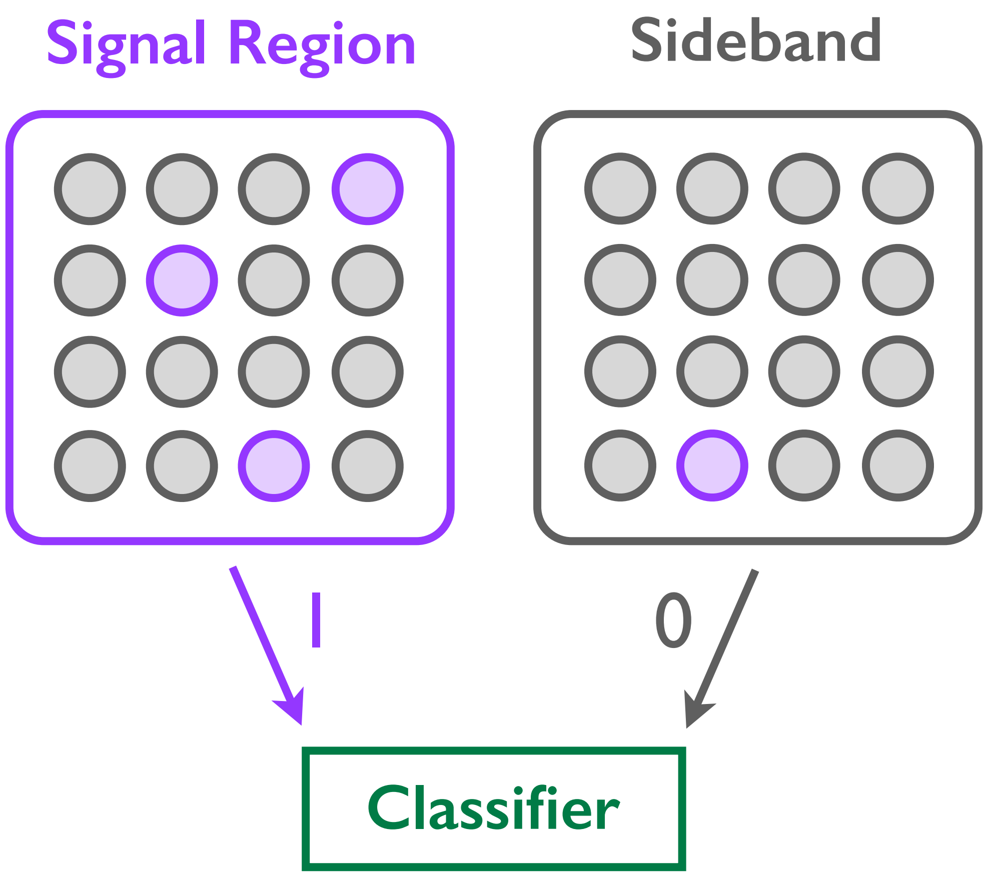


[Kasieczka, Plehn, et al., 1902.09914;
comparison of JDT, Van Tilburg, 1011.2268, 1108.2701; Xie, Girshick, Dollár, Tu, He, 1611.05431; CMS-DP-2017-049; Pearkes, Fedorko, Lister, Gay, 1704.02124;
Butter, Kasieczka, Plehn, Russell, 1707.08966; Komiske, Metodiev, JDT, 1712.07124; Macaluso, Shih 1803.00107; Moore, Nordström, Varma, Fairbairn 1807.04769;
Komiske, Metodiev, JDT, 1810.05165; Erdmann, Geiser, Rath, Rieger, 1812.09722; Qu, Gouskos, 1902.08570; Macaluso, Cranmer, to appear]

CWoLa Hunting

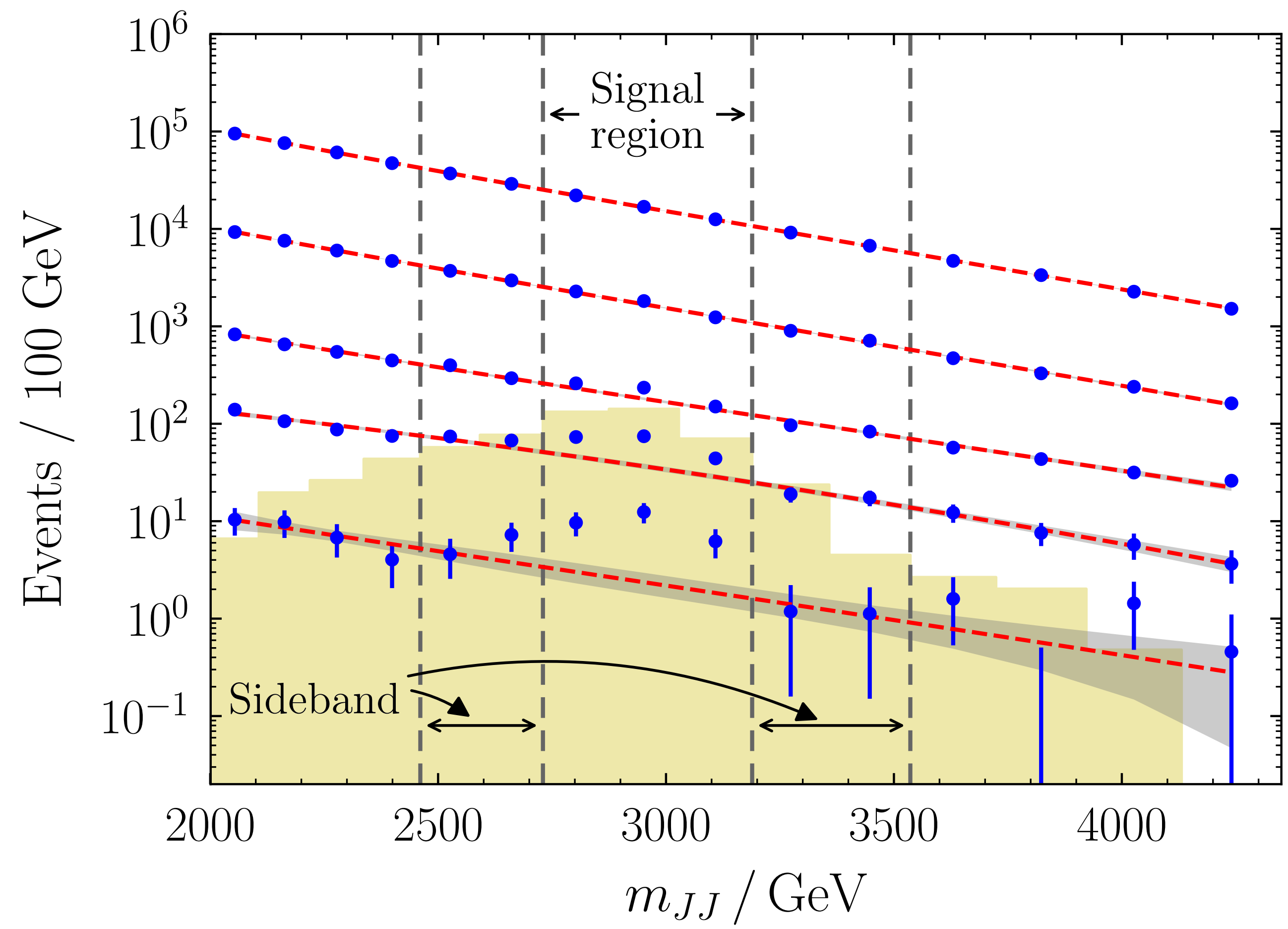
Using "Classification Without Labels"





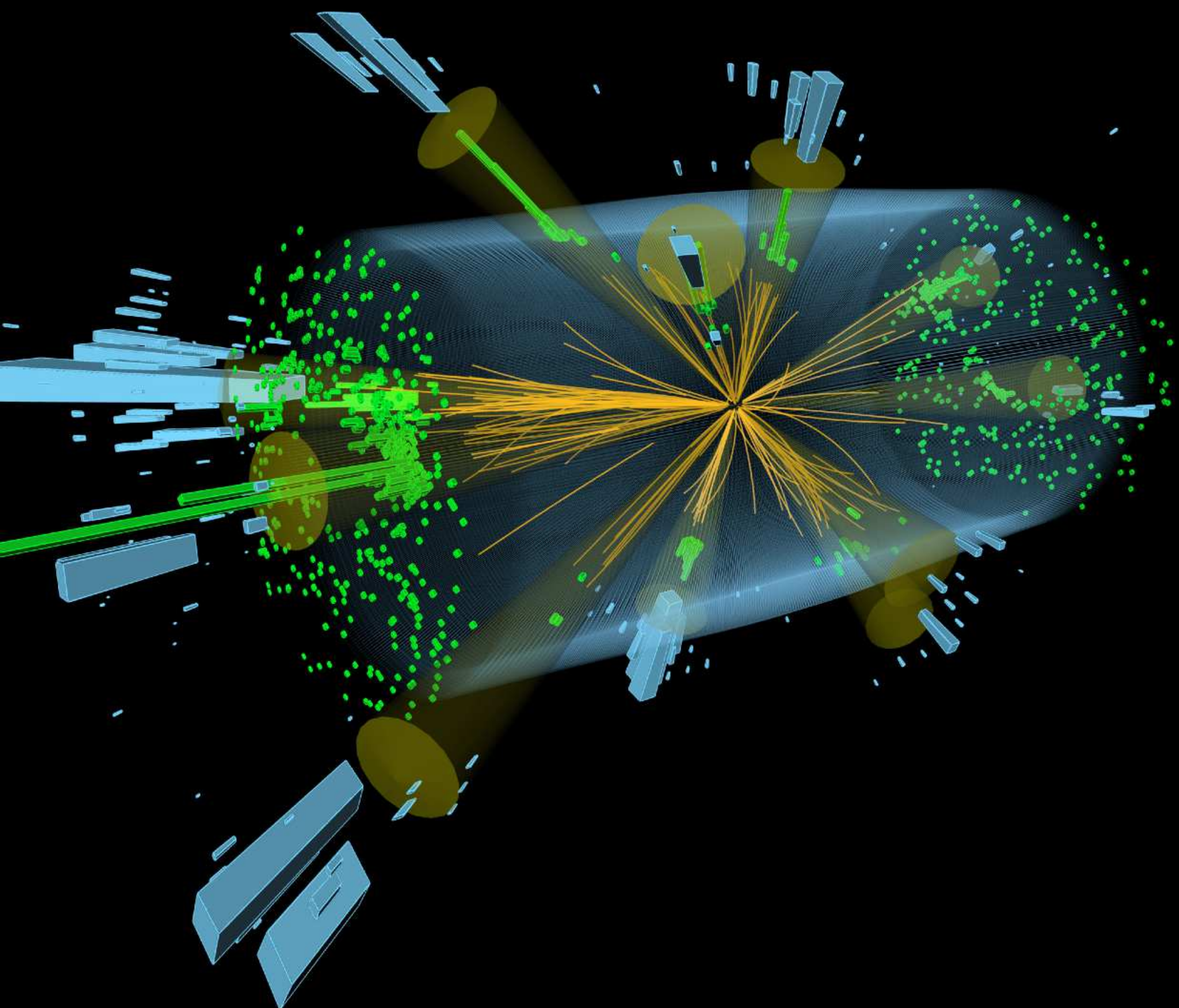
With enough data, monotonic w.r.t. optimal classifier (!)

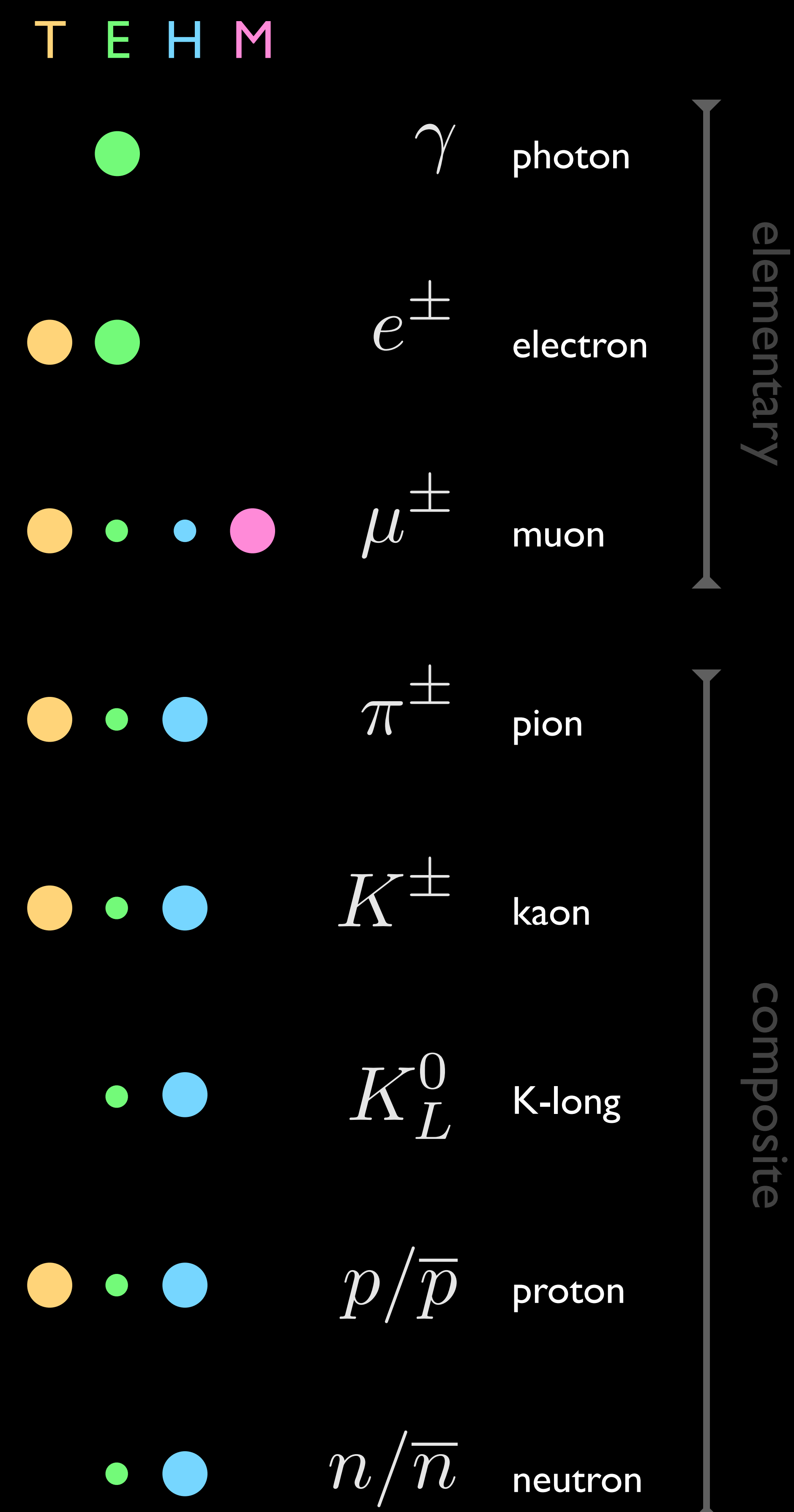
Model-Agnostic Bump Hunt



[Collins, Howe, Nachman, 1805.02664, 1902.02634; using Metodiev, Nachman, JDT, 1708.02949; see also Blanchard, Flaska, Handy, Pozzi, Scott, 1303.1208; Cranmer, Pavez, Louppe, 1506.02169]

What is a Collision Event?





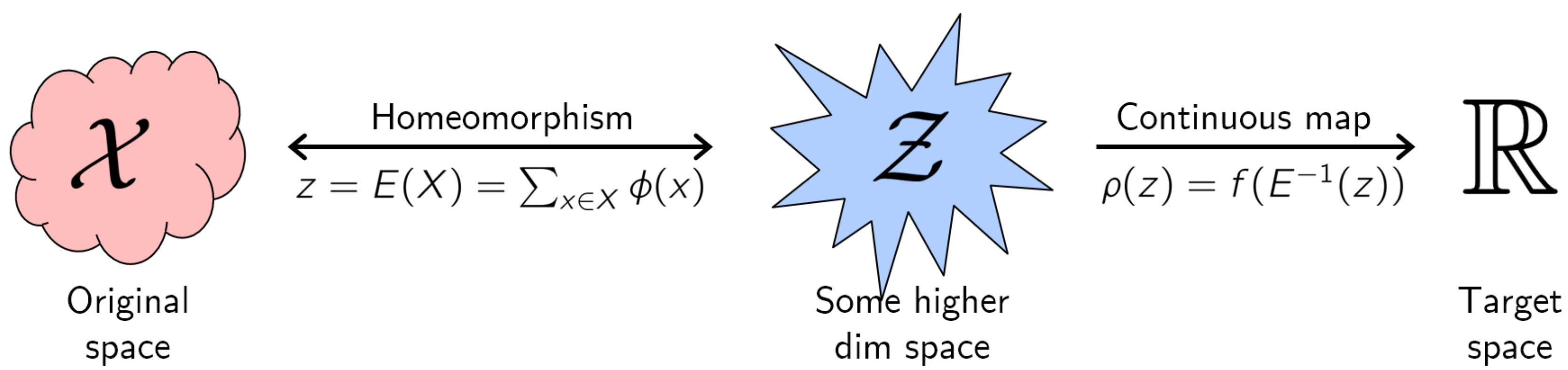
Point Cloud

[Popular Science, 2013]

Meanwhile in ML-Land: Deep Sets

Theorem 2 *A function $f(X)$ operating on a set X having elements from a countable universe, is a valid set function,* i.e., ***invariant*** *to the permutation of instances in X, iff it can be decomposed in the form $\rho\left(\sum_{x \in X} \phi(x)\right)$, for suitable transformations ϕ and ρ.*

(!)



Variable-Length Unordered Set of Particles

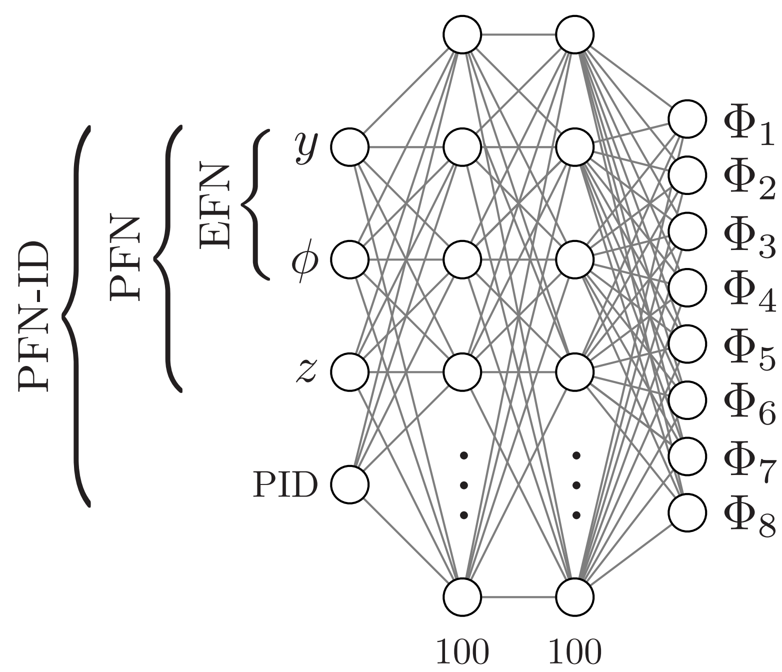
Additive Latent Space

Generic Observable

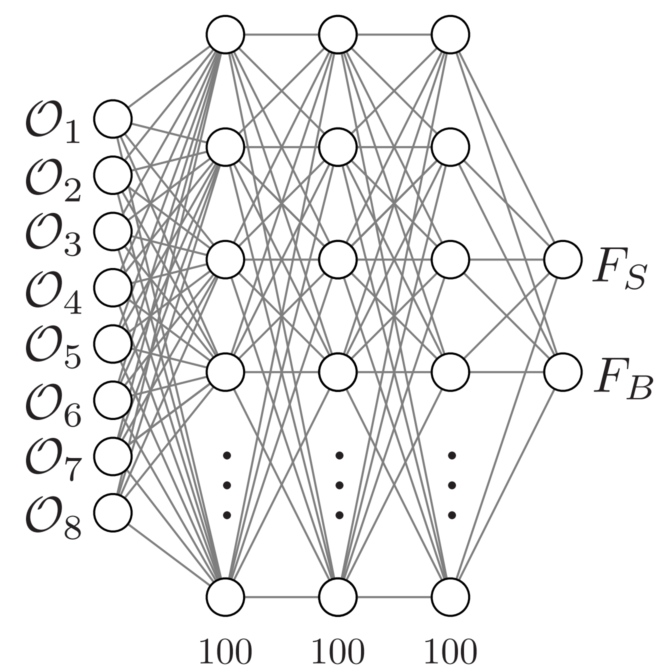
[Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, 1703.06114; see also Vinyals, Bengio, Kudlur, 1511.06391; Qi, Su, Mo, Guibas, 1612.00593]

Technical Implementation

Per-Particle Network: Φ



Latent Combiner: F



Per-Jet (Latent Space):

EFN: $\mathcal{O}_a = \sum_{i \in \text{jet}} z_i \, \Phi_a(y_i, \phi_i) \qquad z_i = \frac{p_{Ti}}{\sum_j p_{Tj}}$

PFN: $\mathcal{O}_a = \sum_{i \in \text{jet}} \Phi_a(y_i, \phi_i, z_i, \text{PID}_i)$

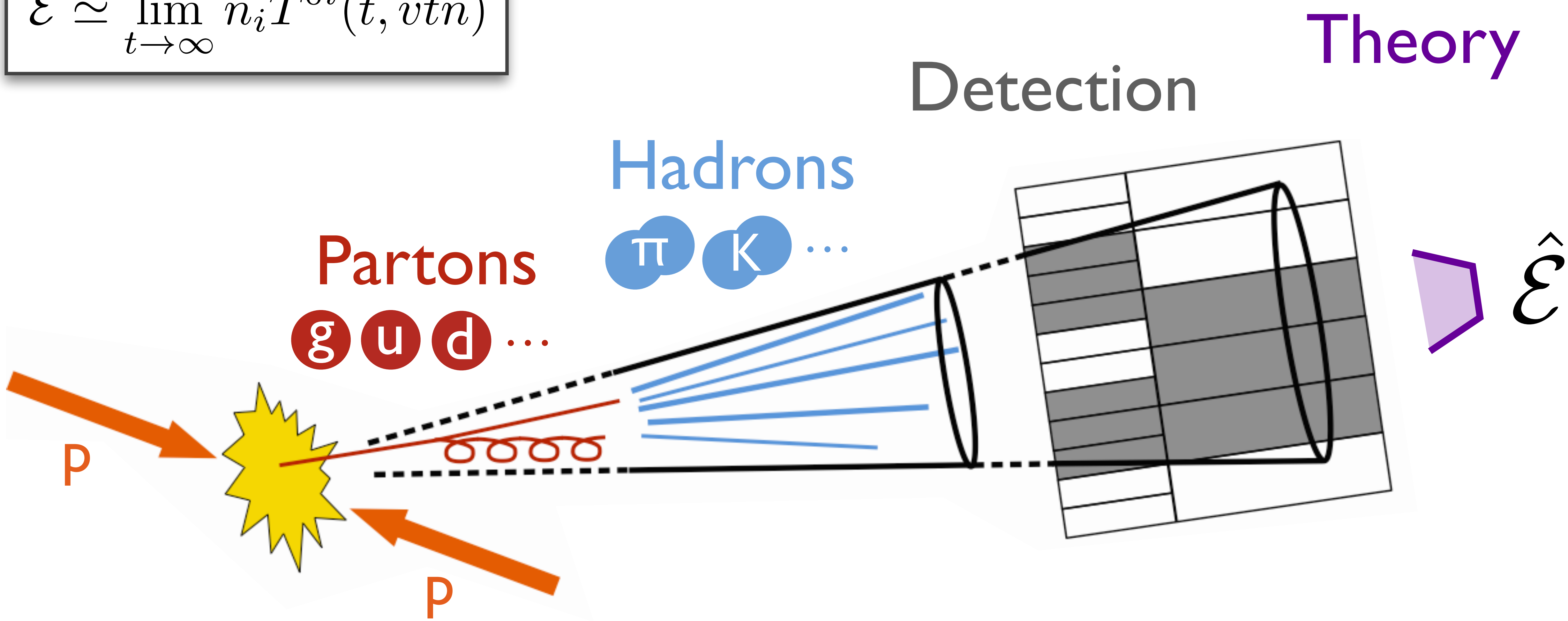
Final Discriminant:

$$\text{softmax}(F_S, F_B)$$

[Komiske, Metodiev, JDT, 1810.05165]

Focus on Energy Flow

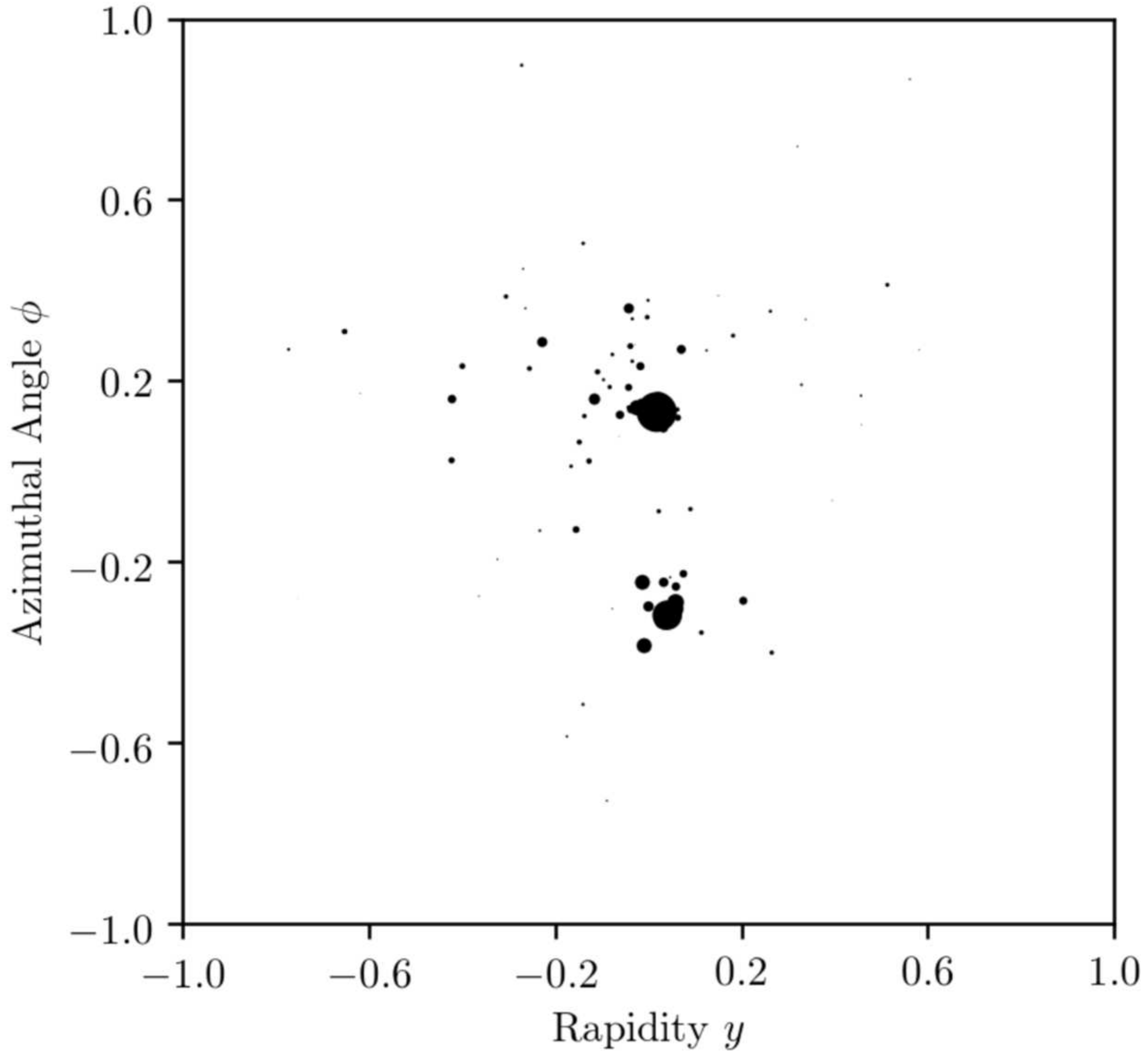
$$\boxed{\hat{\mathcal{E}} \simeq \lim_{t\to\infty} \hat{n}_i T^{0i}(t, vt\hat{n})}$$



Stress-energy flow: Measure of event/jet structure robust to non-perturbative and detector effects (i.e. IRC safe)

[Sveshnikov, Tkachov, hep-ph/9512370; Hofman, Maldacena, 0803.1467; Mateu, Stewart, JDT, 1209.3781; Komiske, Metodiev, JDT, 1712.07124, 1810.05165]

Focus on Energy Flow



Represent jet as:

$$\rho(\hat{p}) = \sum_{i \in \text{jet}} E_i \, \delta(\hat{p} - \hat{p}_i)$$

Energy (p_T)

Direction (y,φ)

Safe to infrared & collinear splittings
No flavor/charge information
No pixelation needed

Stress-energy flow: Measure of event/jet structure robust to non-perturbative and detector effects (i.e. IRC safe)

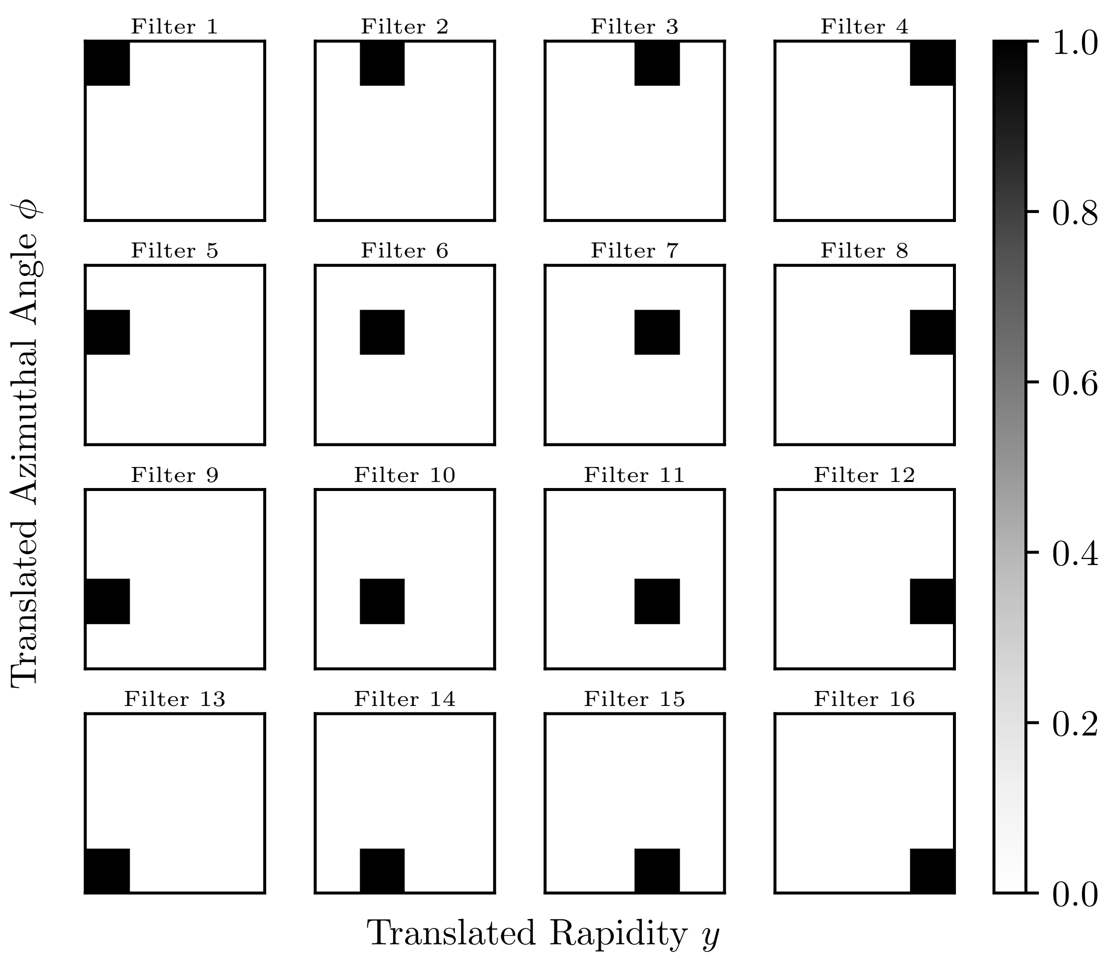
[Sveshnikov, Tkachov, hep-ph/9512370; Hofman, Maldacena, 0803.1467; Mateu, Stewart, JDT, 1209.3781; Komiske, Metodiev, JDT, 1712.07124, 1810.05165]

Latent Space Visualization

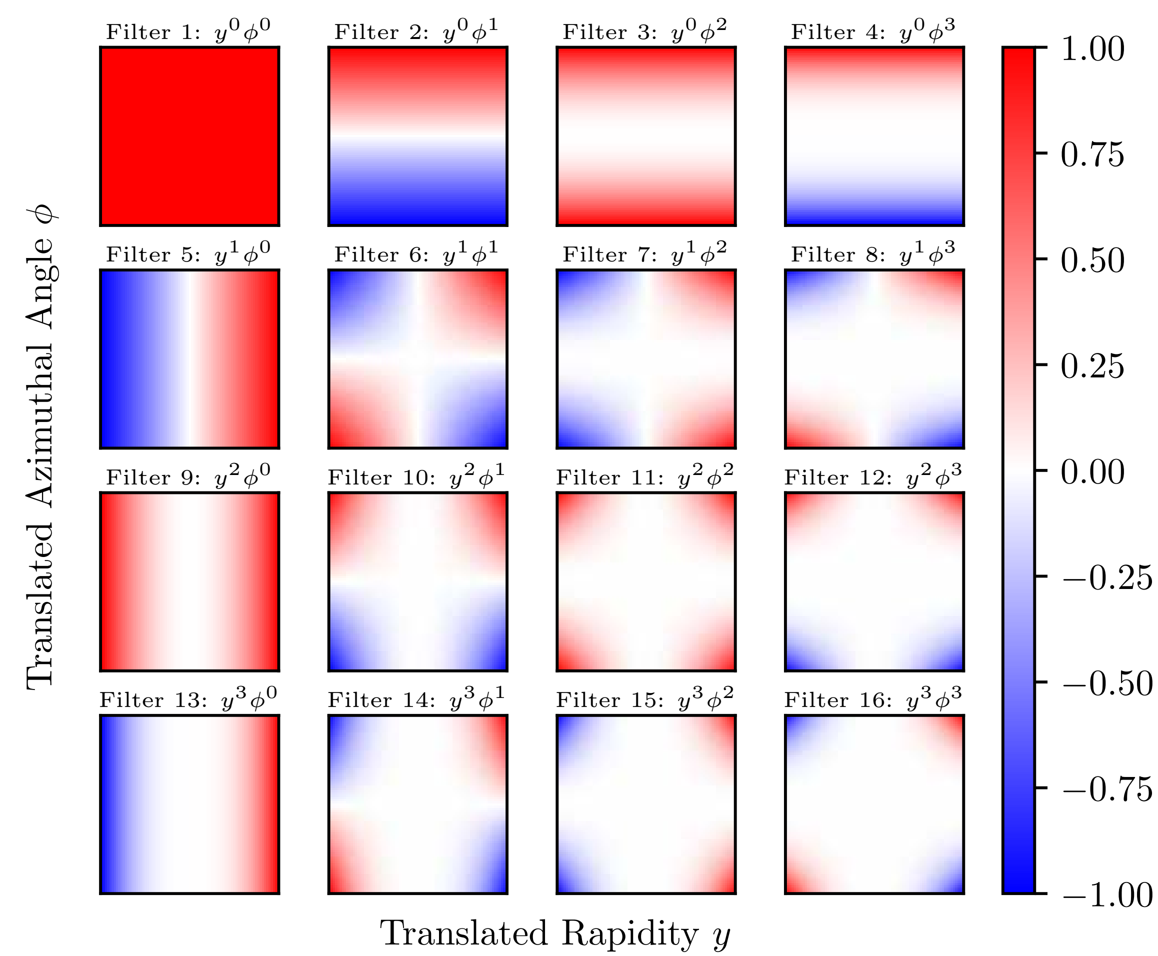
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

IRC-safe: $\mathcal{O}_a = \sum_{i \in \text{jet}} E_i \, \Phi_a(\hat{p}_i)$

Calorimeter Pixels



Radiation Moments

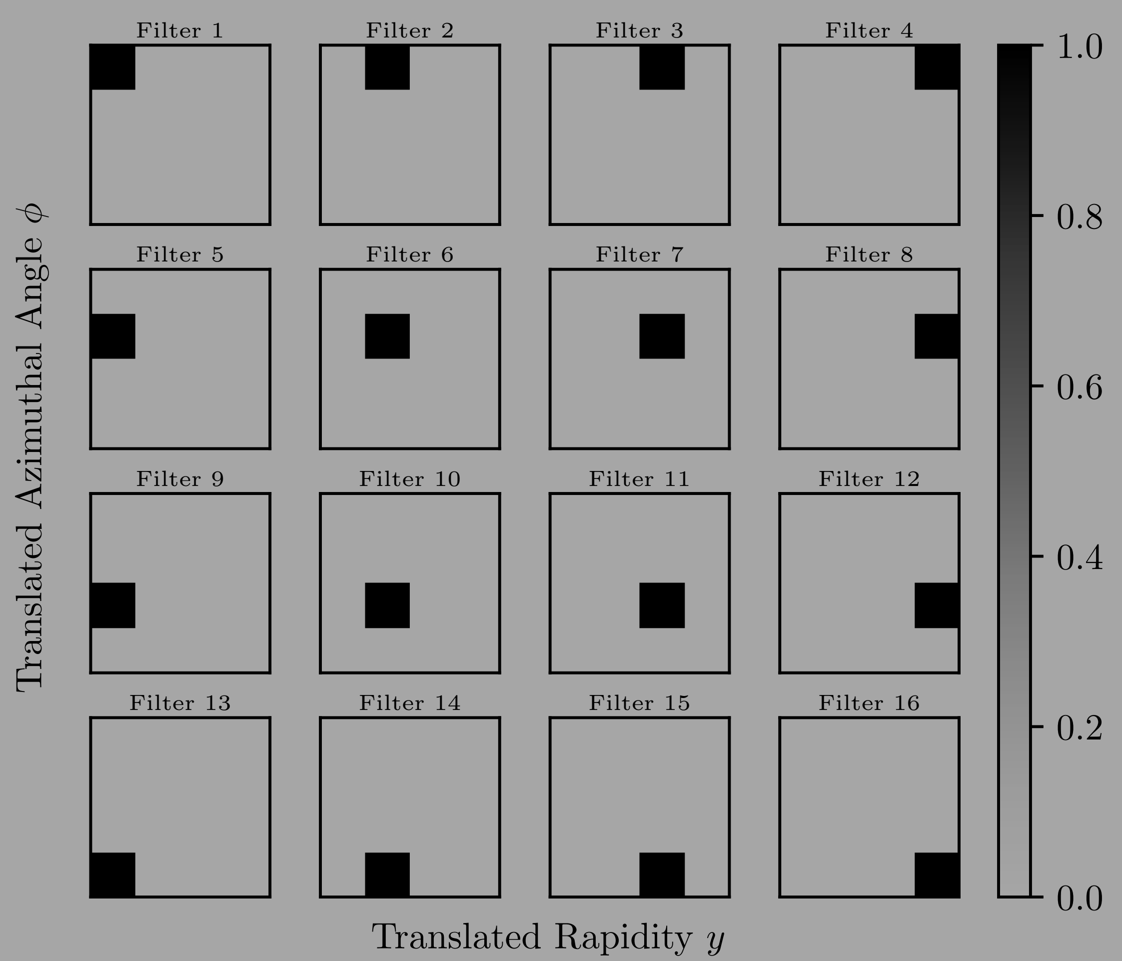


Latent Space Visualization

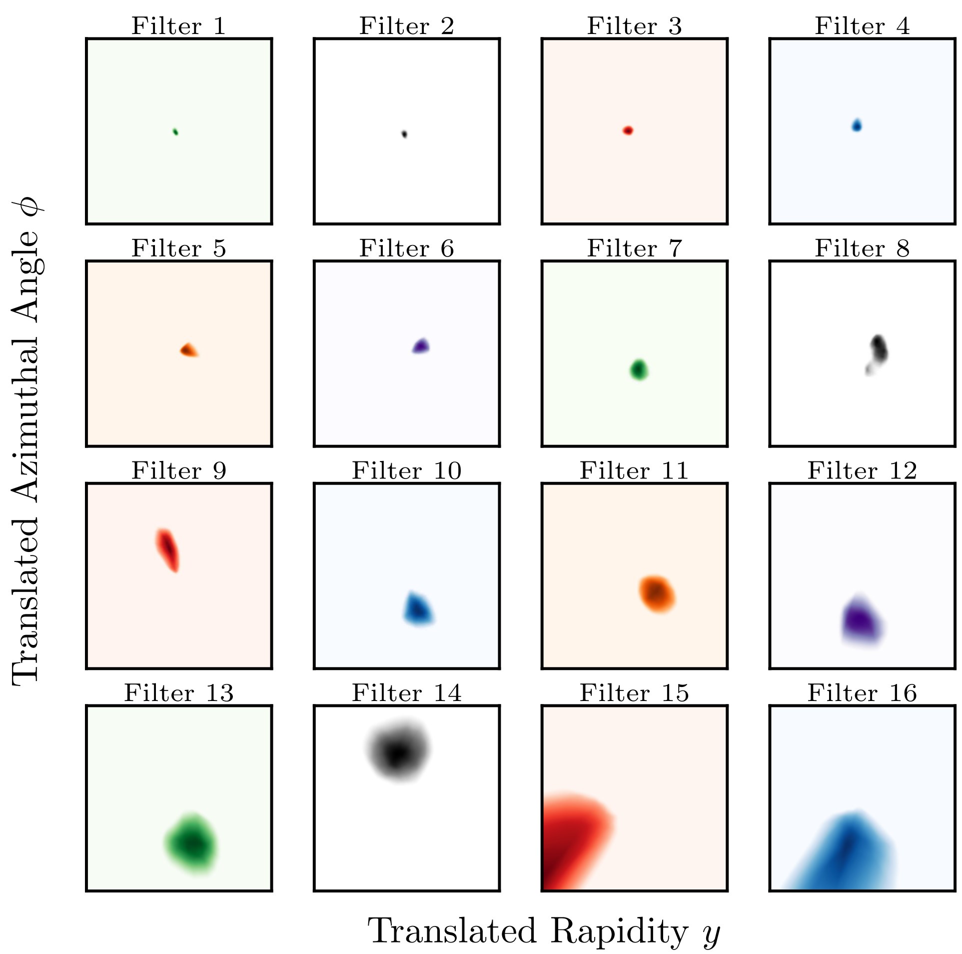
$$\mathcal{O}(\{p_i^\mu\}) = F(\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_\ell)$$

IRC-safe: $\mathcal{O}_a = \sum_{i \in \text{jet}} E_i \, \Phi_a(\hat{p}_i)$

Calorimeter Pixels

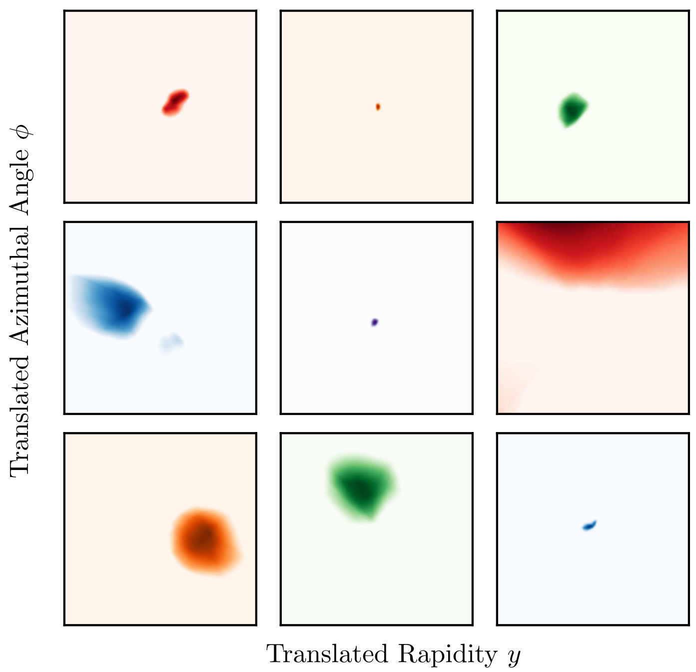


EFNs: Dynamic Pixelation

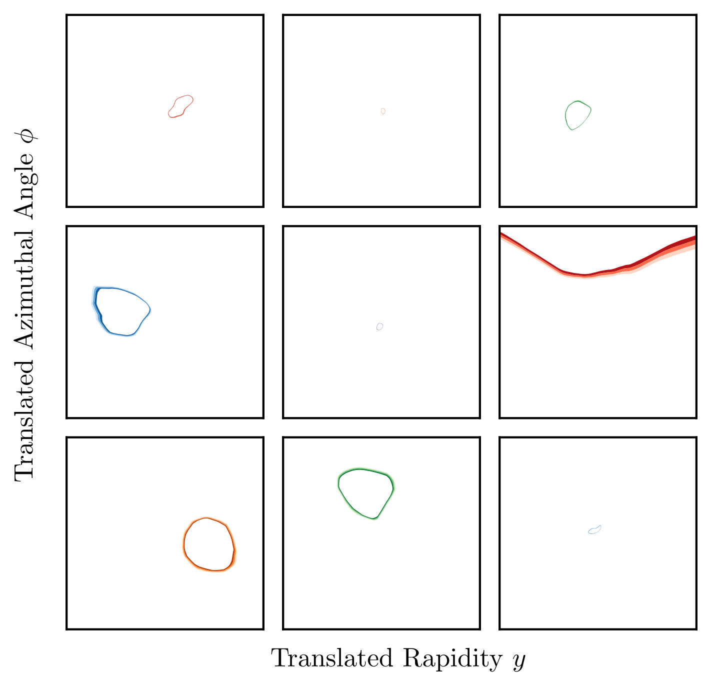


Psychedelic Network Visualization

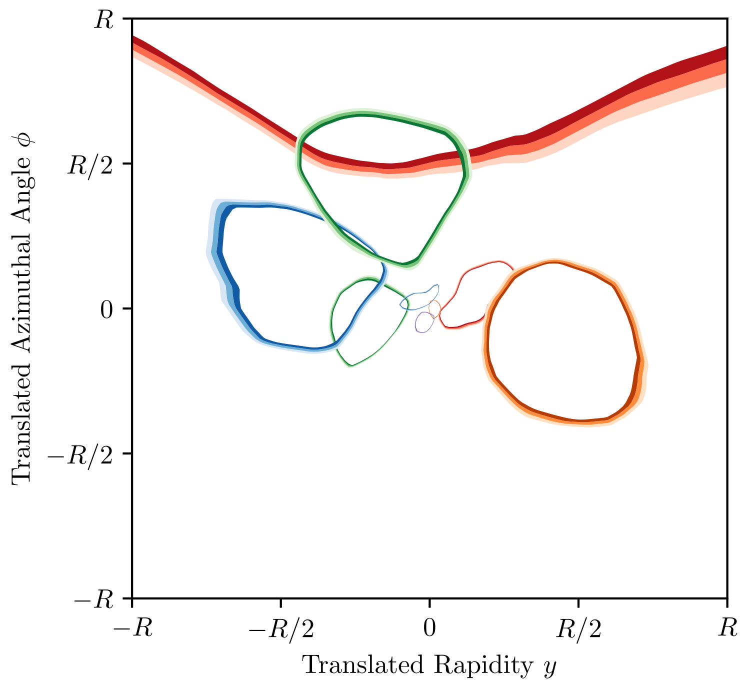
Latent Filters

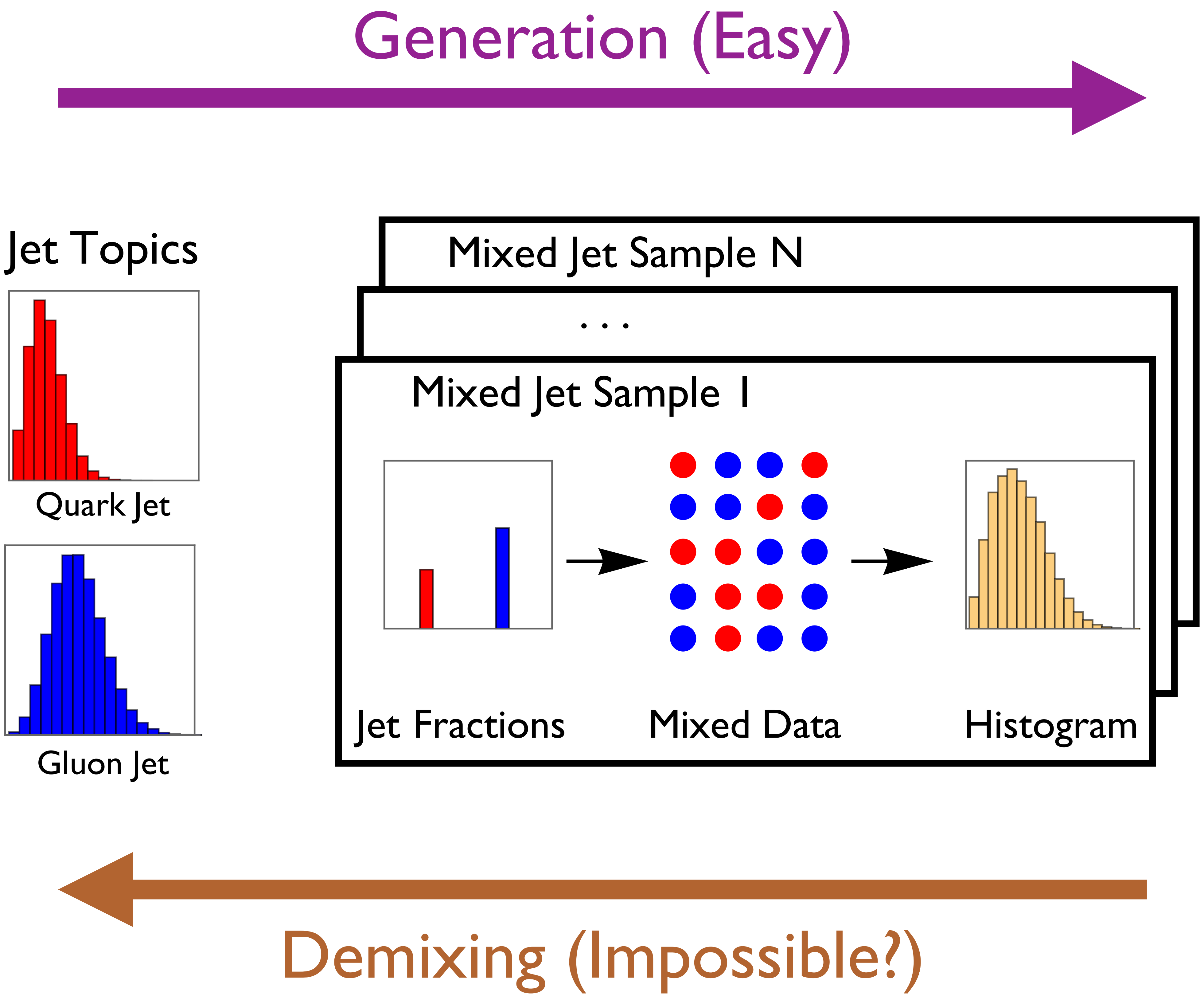


50% Contours

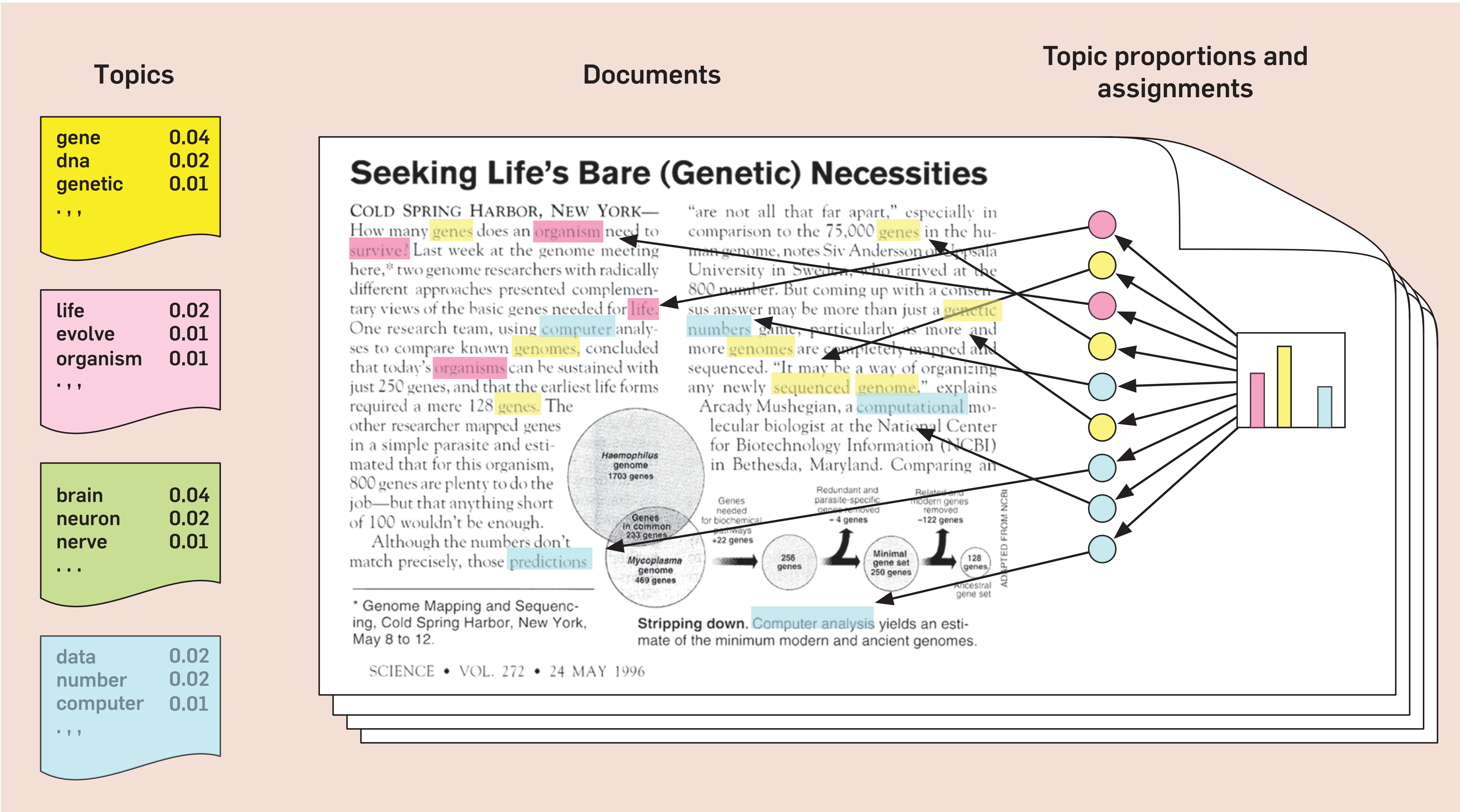


Overlay



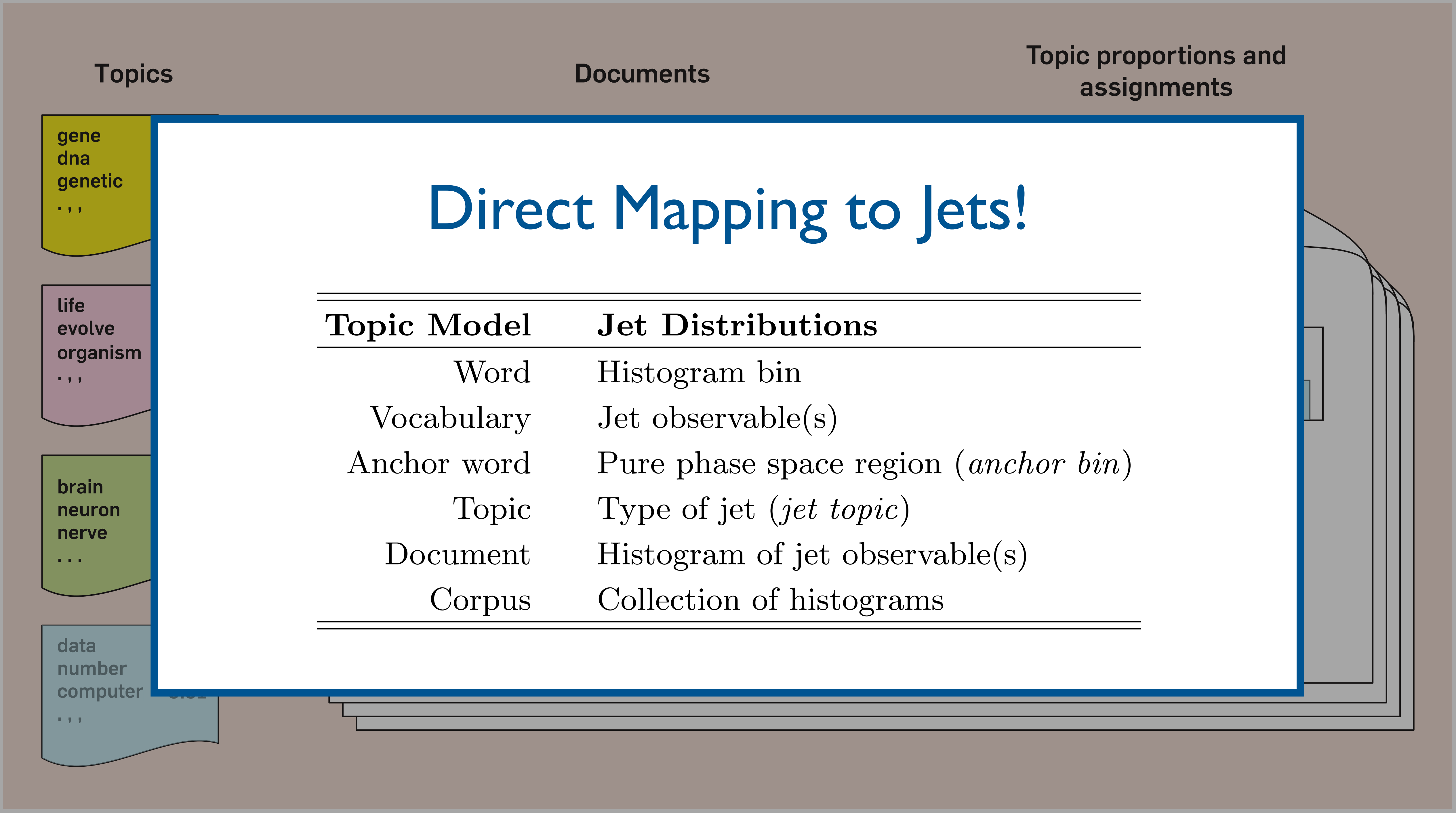

Generation (Easy)
Jet Topics
Quark Jet
Gluon Jet
Mixed Jet Sample N
. . .
Mixed Jet Sample 1
Jet Fractions
Mixed Data
Histogram
Demixing (Impossible?)

Topic Modeling



[Blei, 2012]

Topic Modeling

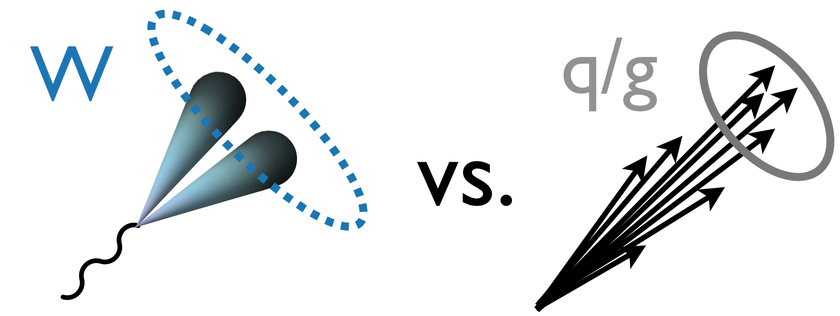


Direct Mapping to Jets!

Topic Model	**Jet Distributions**
Word	Histogram bin
Vocabulary	Jet observable(s)
Anchor word	Pure phase space region (*anchor bin*)
Topic	Type of jet (*jet topic*)
Document	Histogram of jet observable(s)
Corpus	Collection of histograms

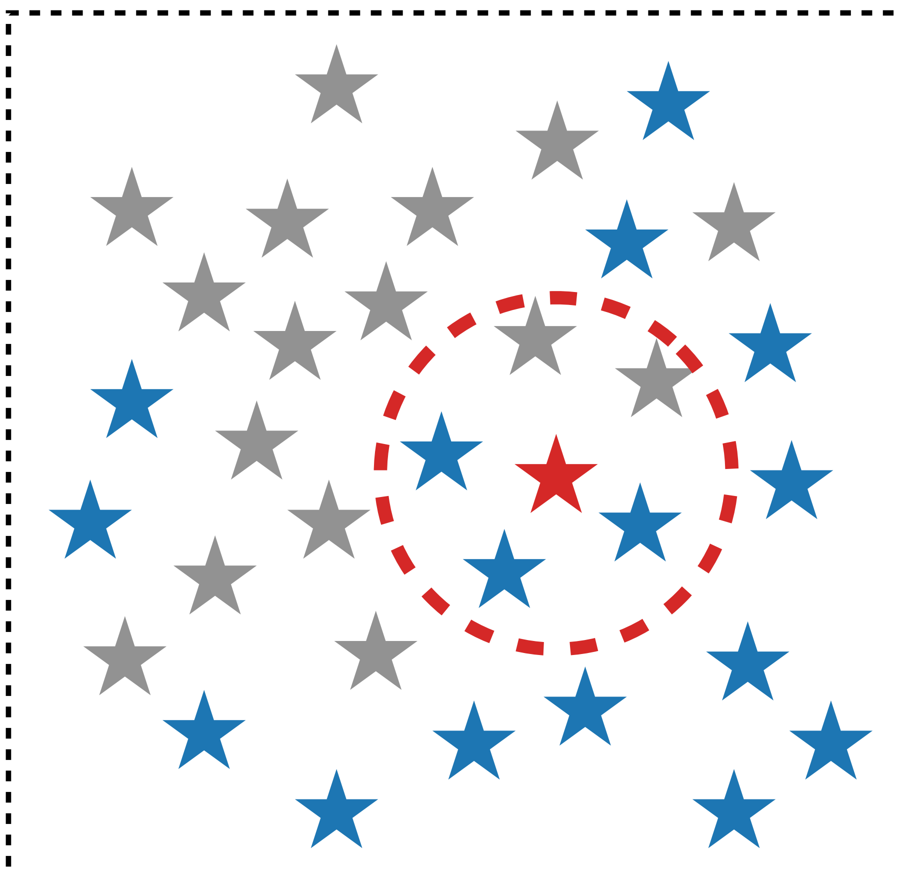
[Blei, 2012]

Revisiting Jet Classification

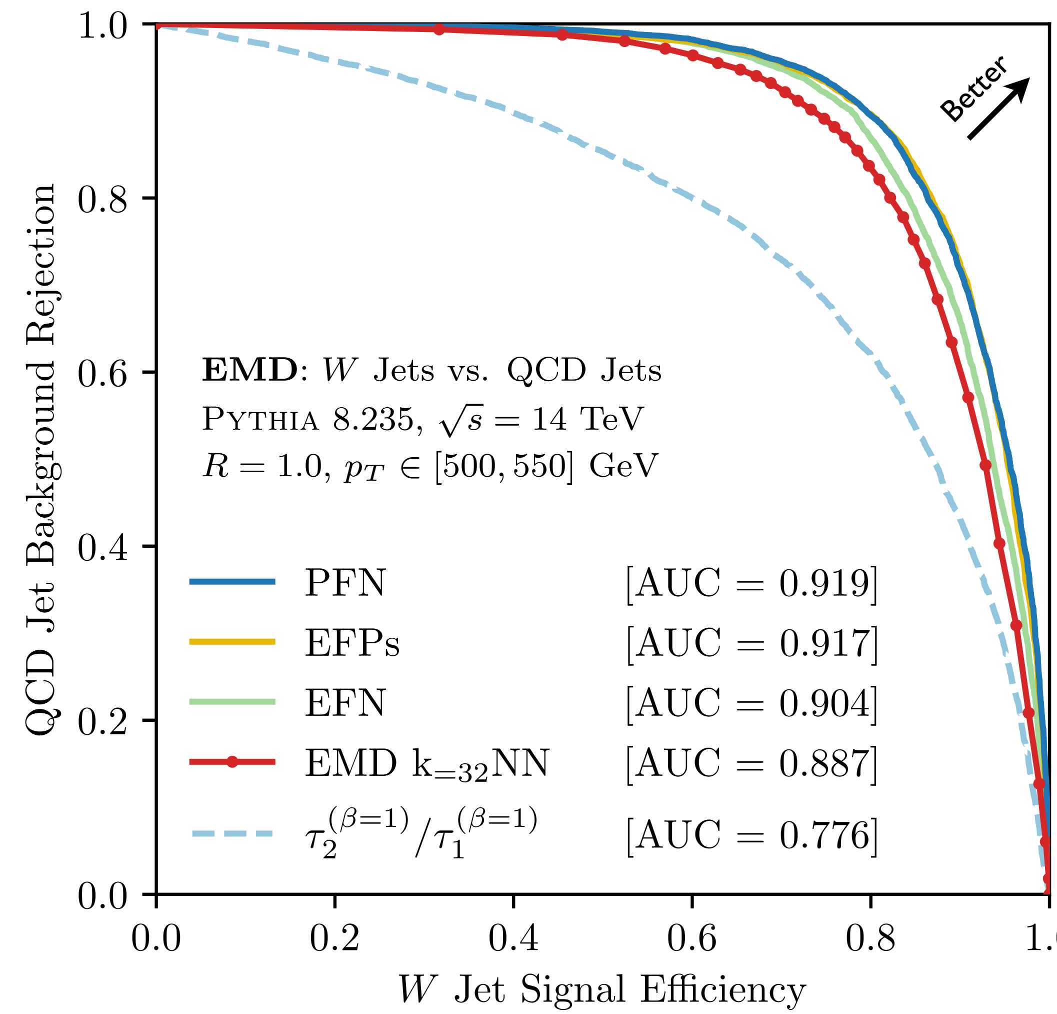


Estimate jet label by k nearest neighbors in training data

Approaches performance of modern machine learning

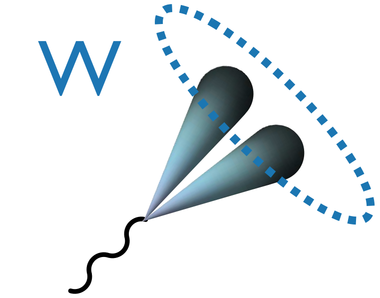


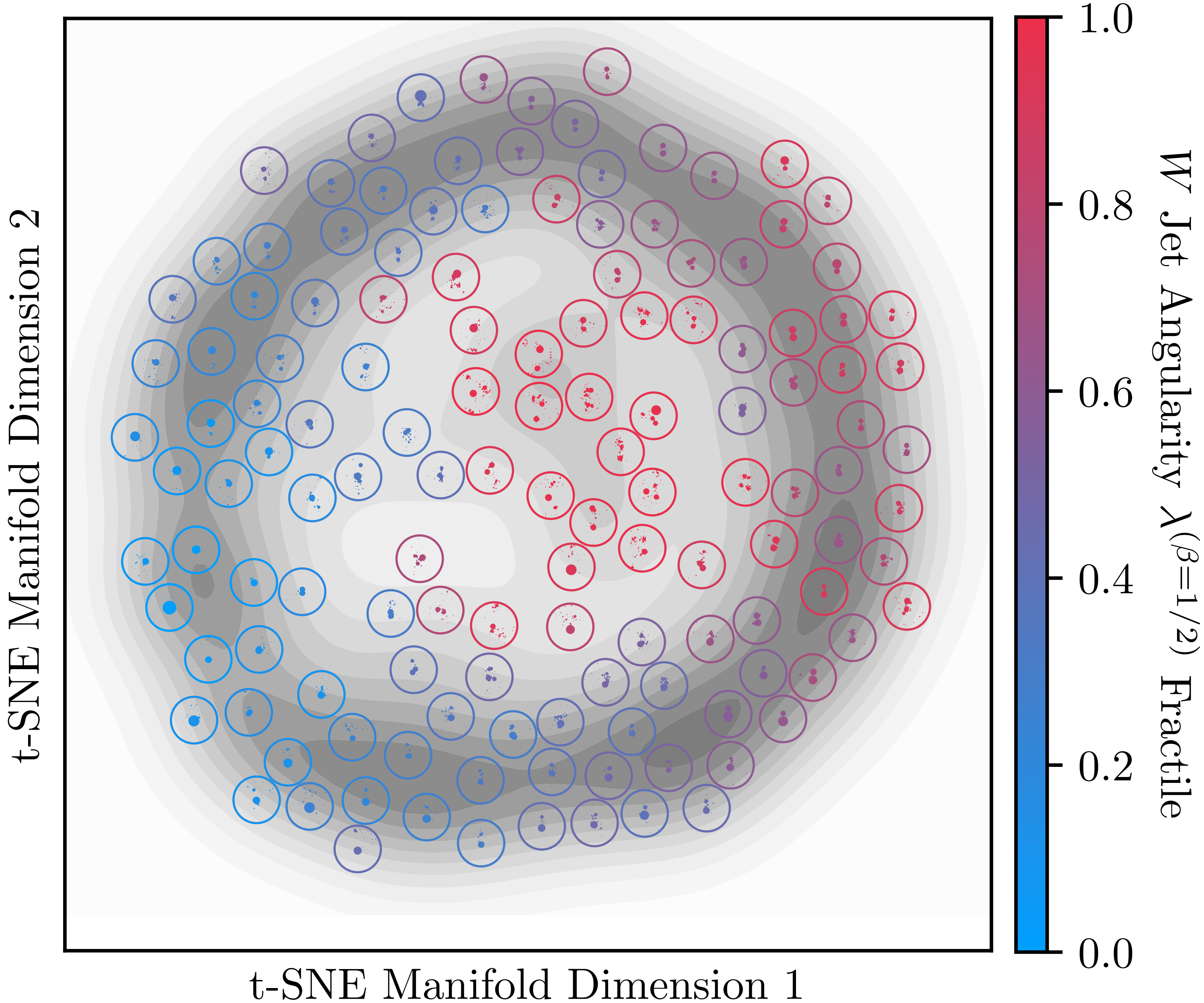
5-NN



[Komiske, Metodiev, JDT, 1902.02346;
comparison to JDT, Van Tilburg, 1011.2268, 1108.2701; Komiske, Metodiev, JDT, 1712.07124, 1810.05165]

The Space of Boosted W Bosons





[Komiske, Metodiev, JDT, 1902.02346]