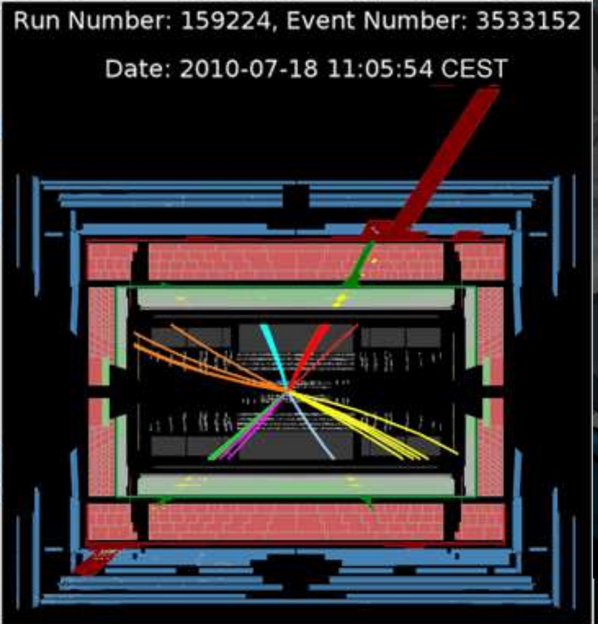
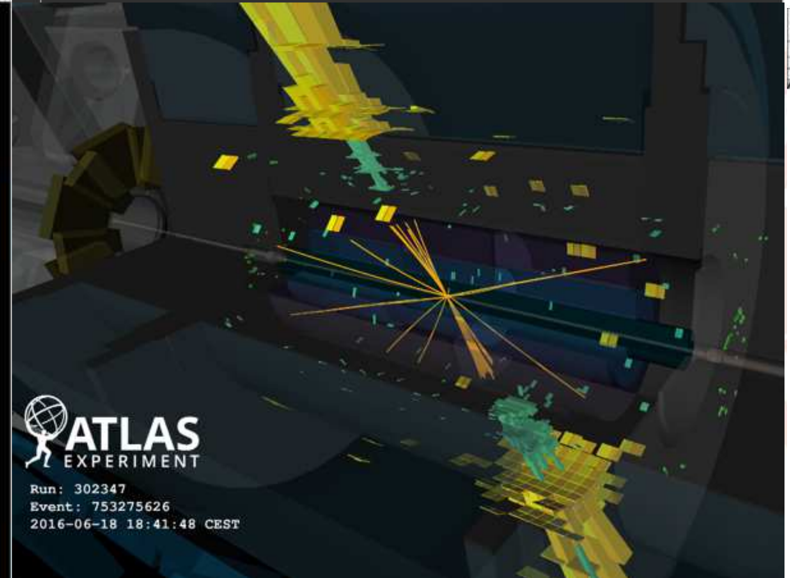
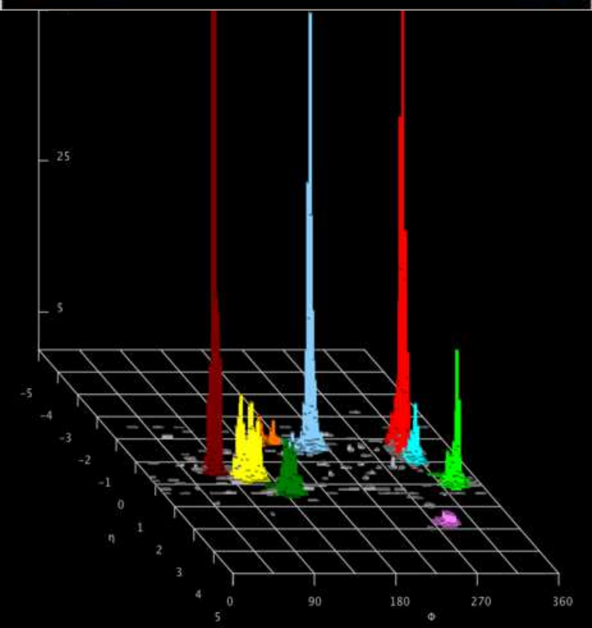
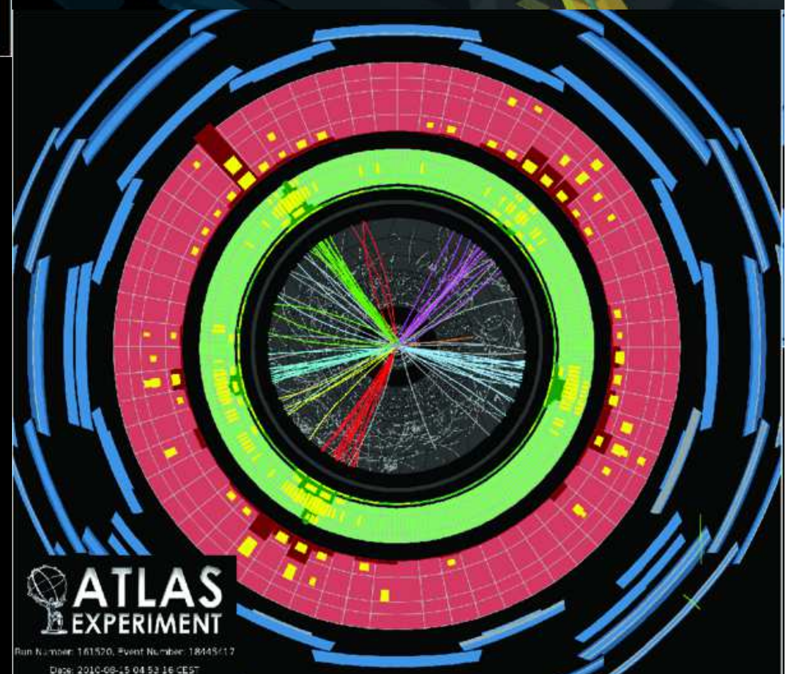
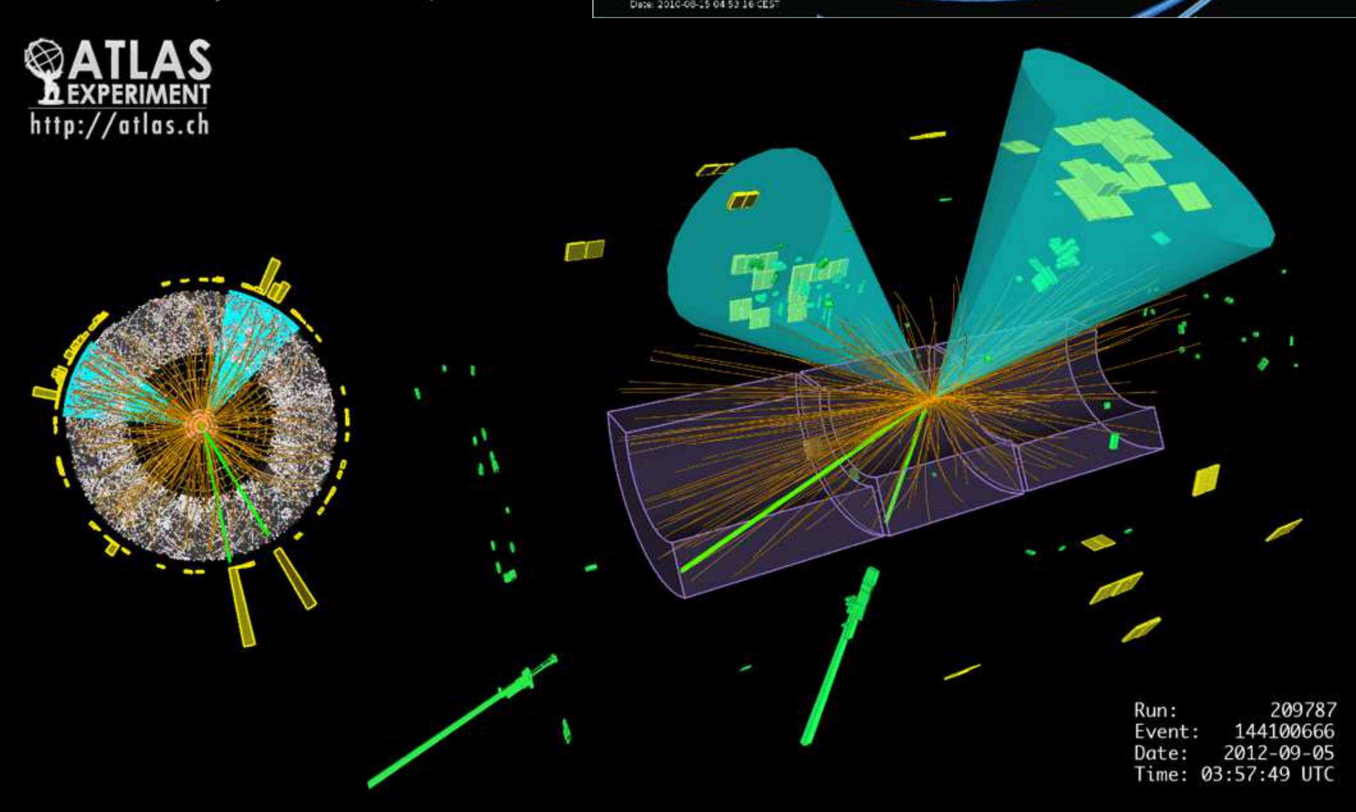
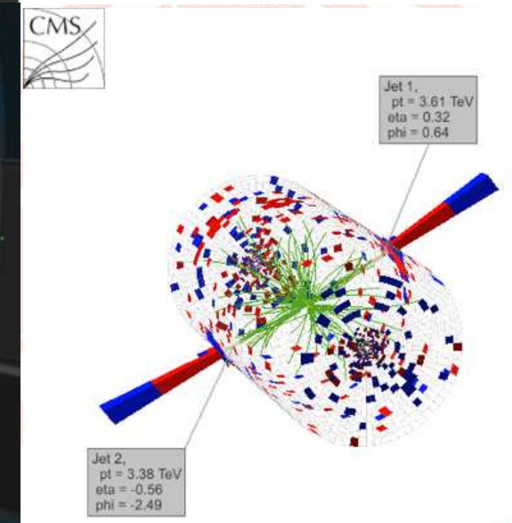
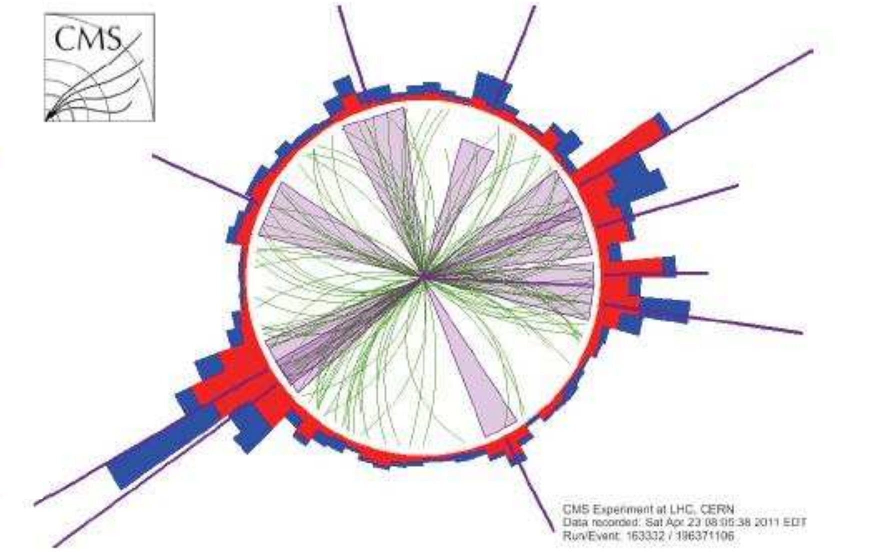
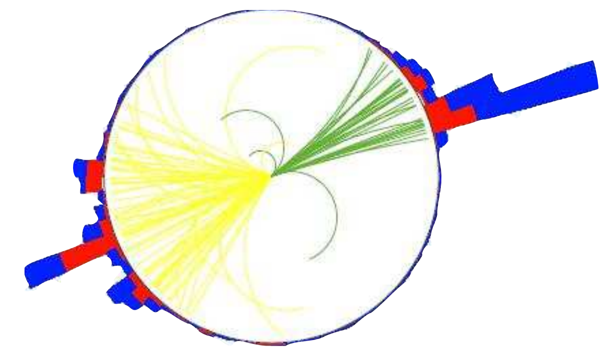
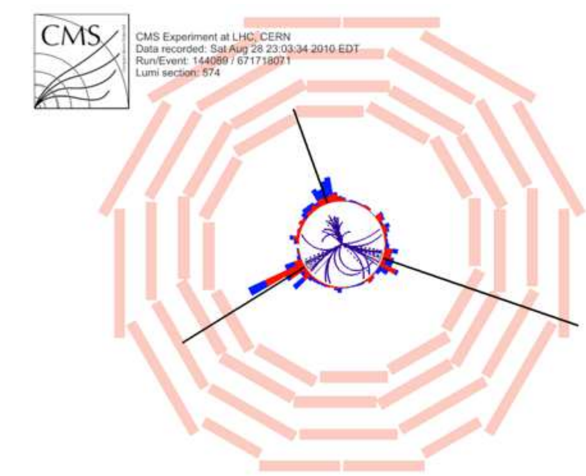
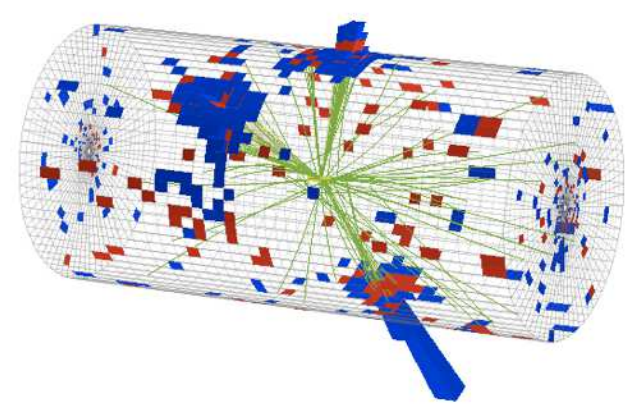
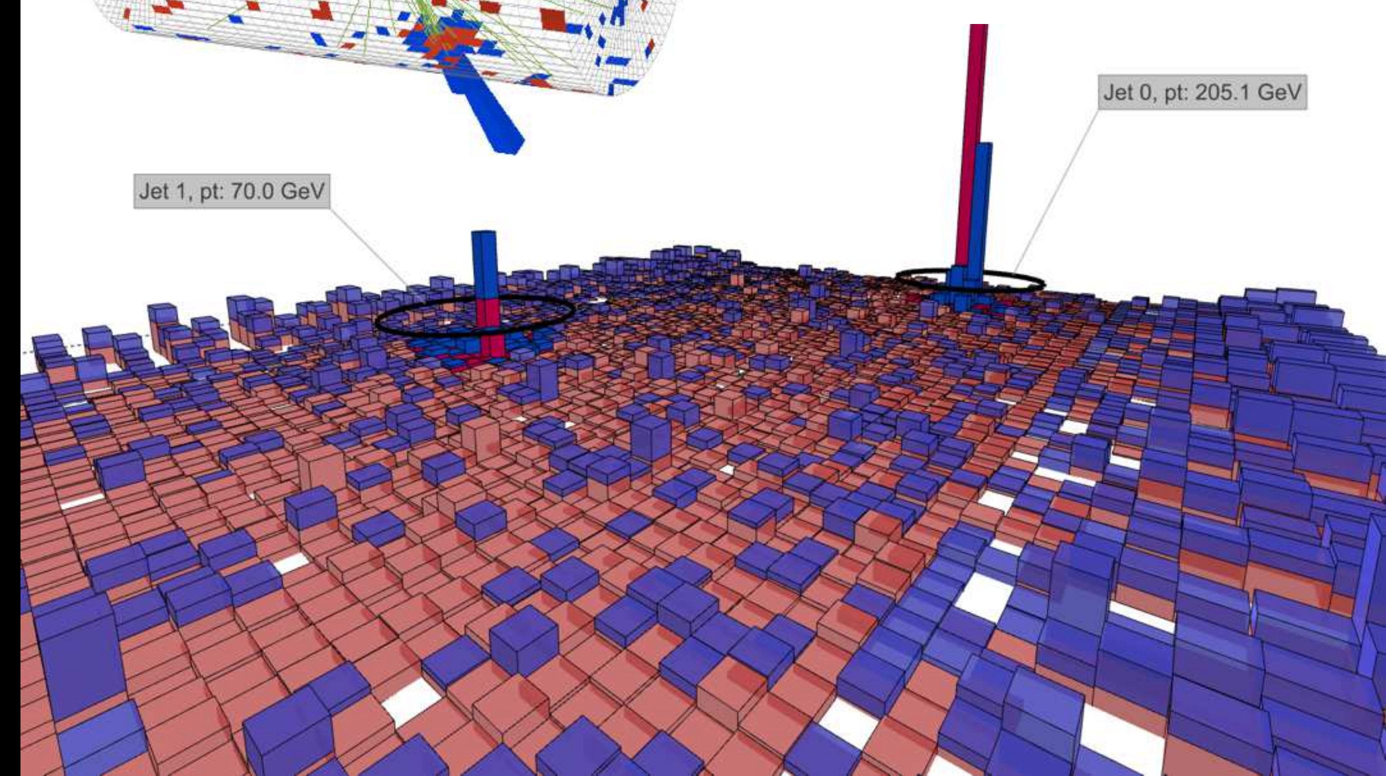
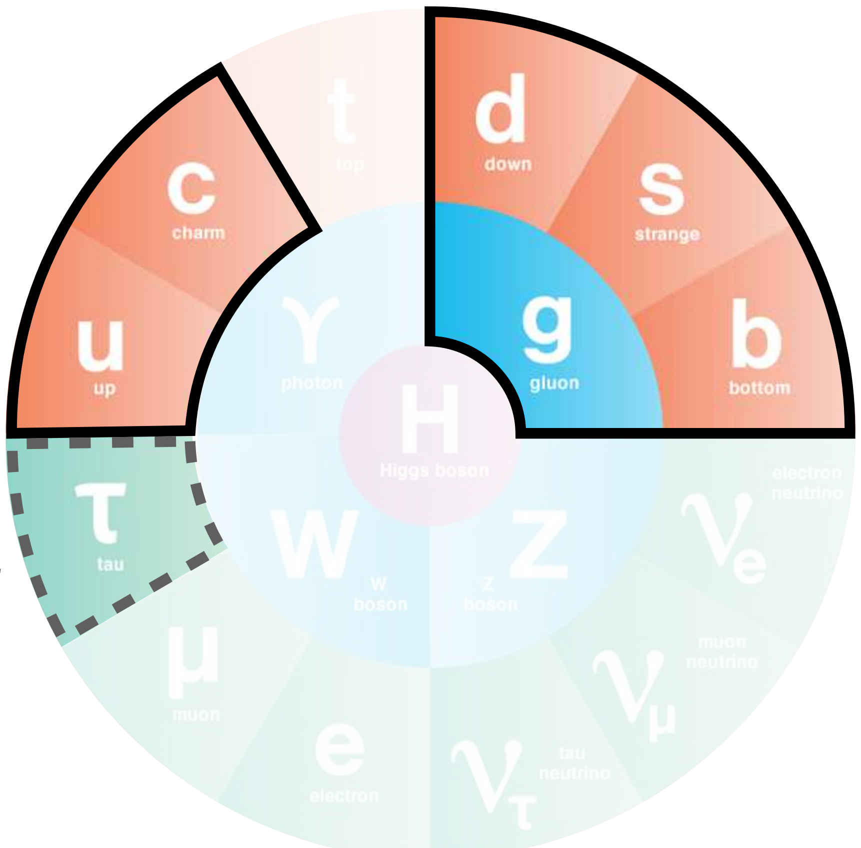
On the Topic of Jets

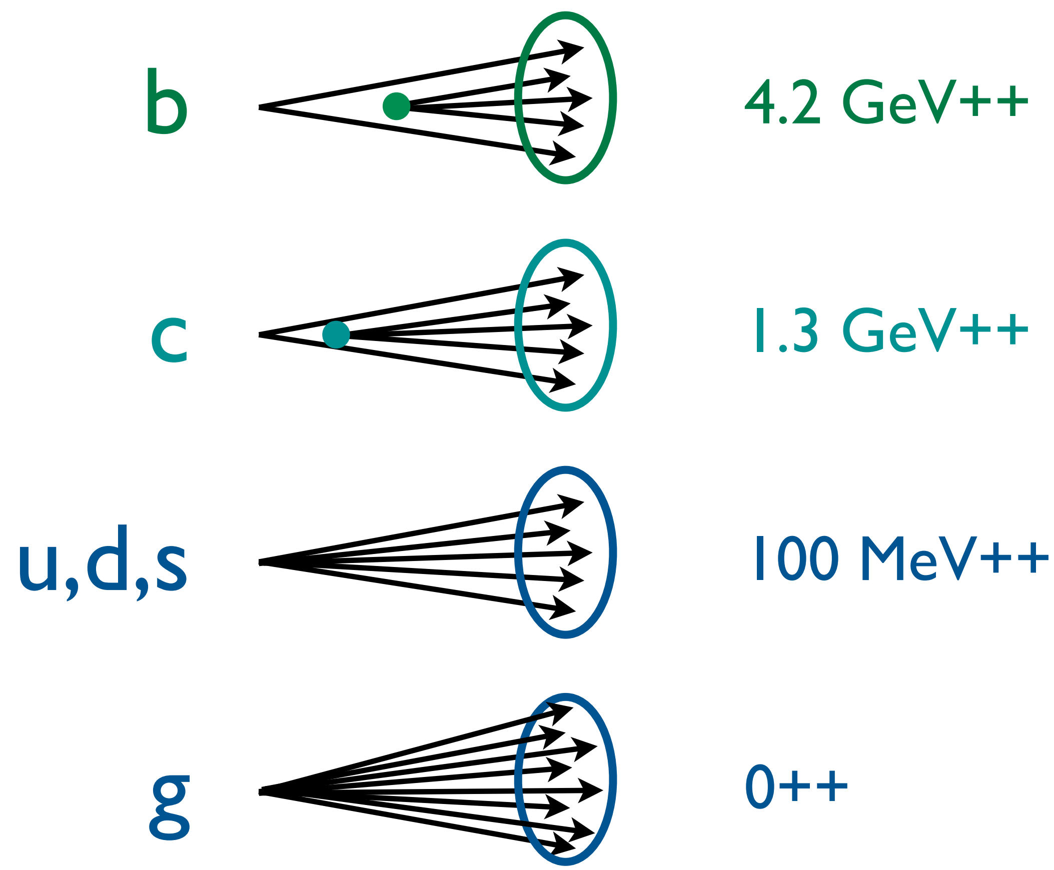
Jesse Thaler

Self-Invited Seminar, DAMTP Cambridge — May 1, 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


ATLAS
EXPERIMENT

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

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


CMS
CMS Experiment at LHC, CERN


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

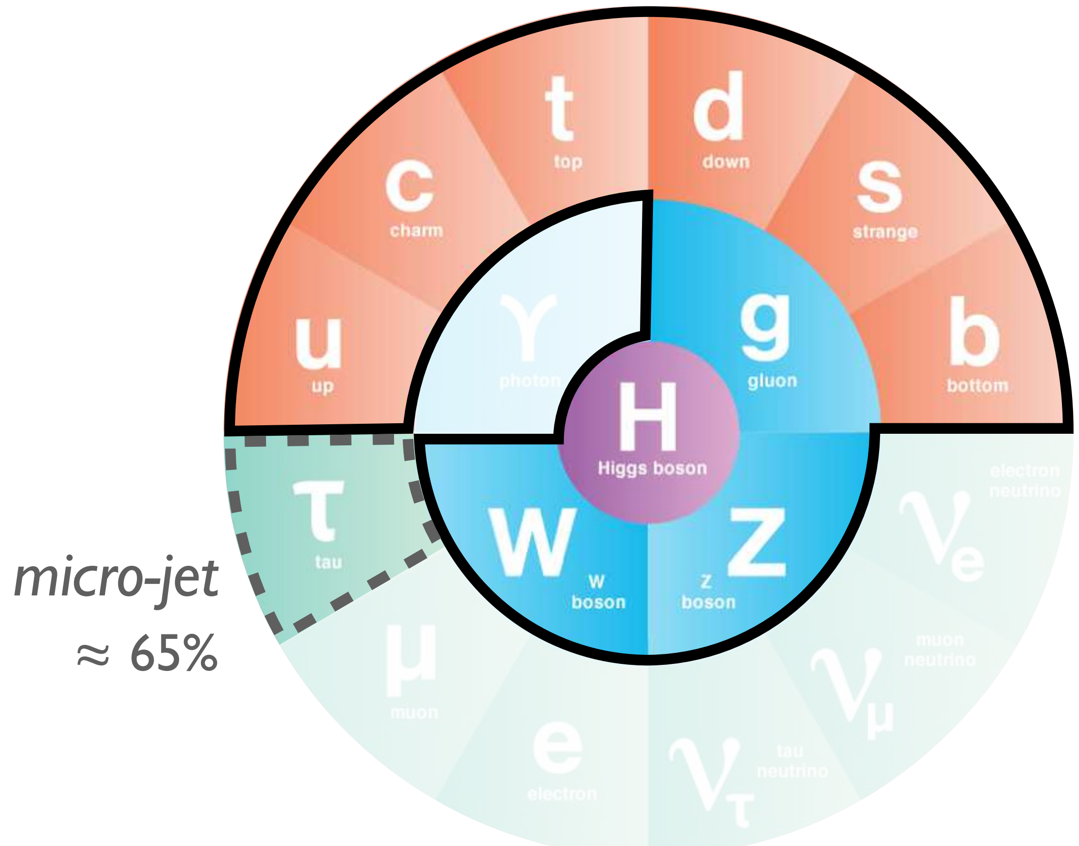


micro-jet
≈ 65%



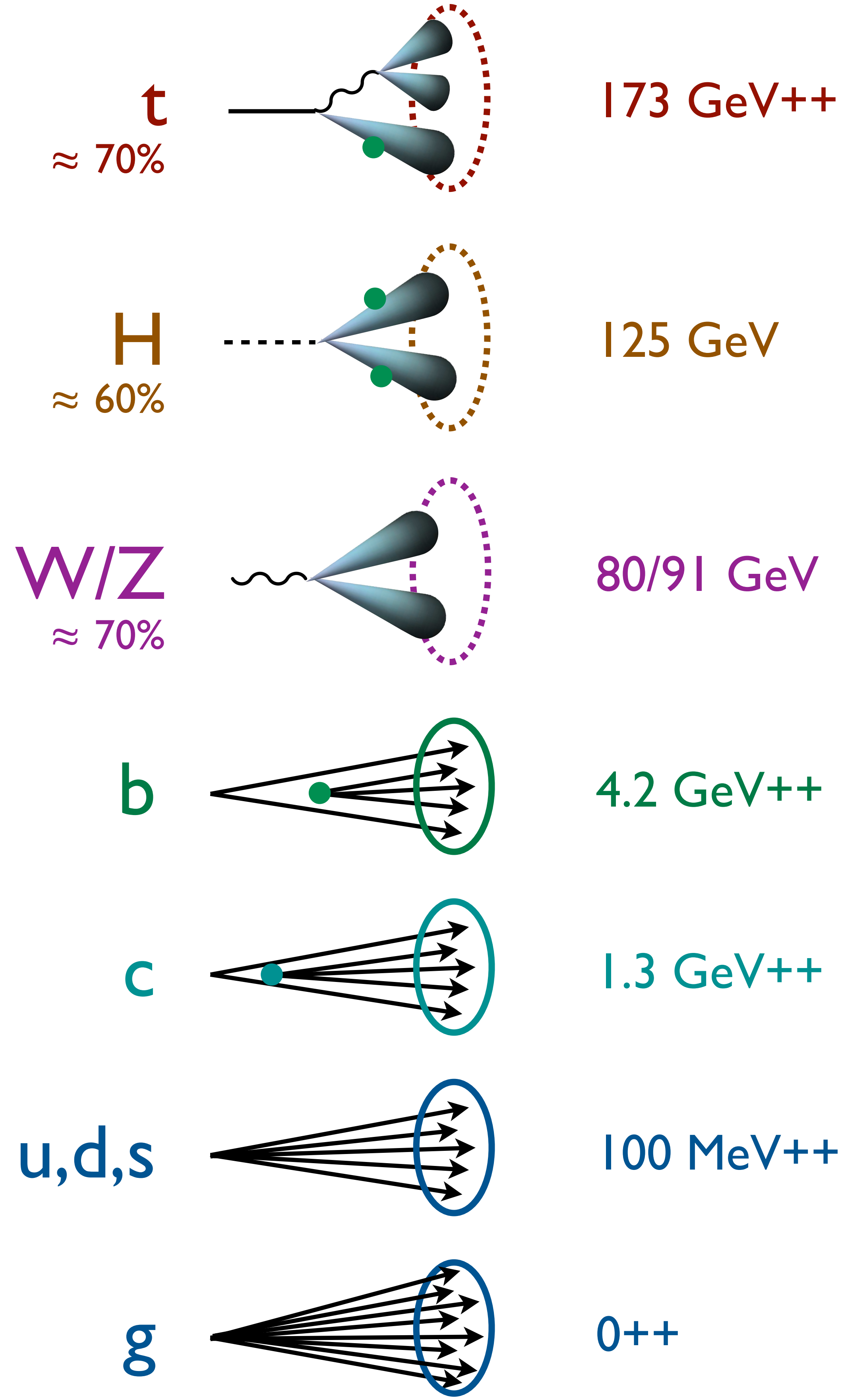
Jets from the Standard Model

++ = Mass from QCD Radiation



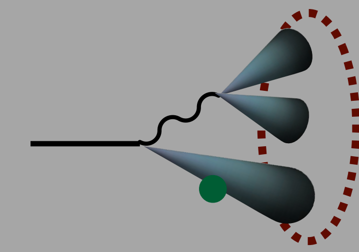
Jets from the Standard Model

++ = Mass from QCD Radiation



t

≈ 70%



173 GeV++

micro-j

≈ 65

Bottom Line:
Jet Classification is "Solved"

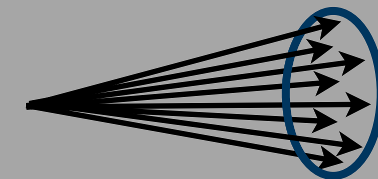
Trustable training samples?

Well-defined categories?

Controlled systematics?

++ = Mass from QCD Radiation

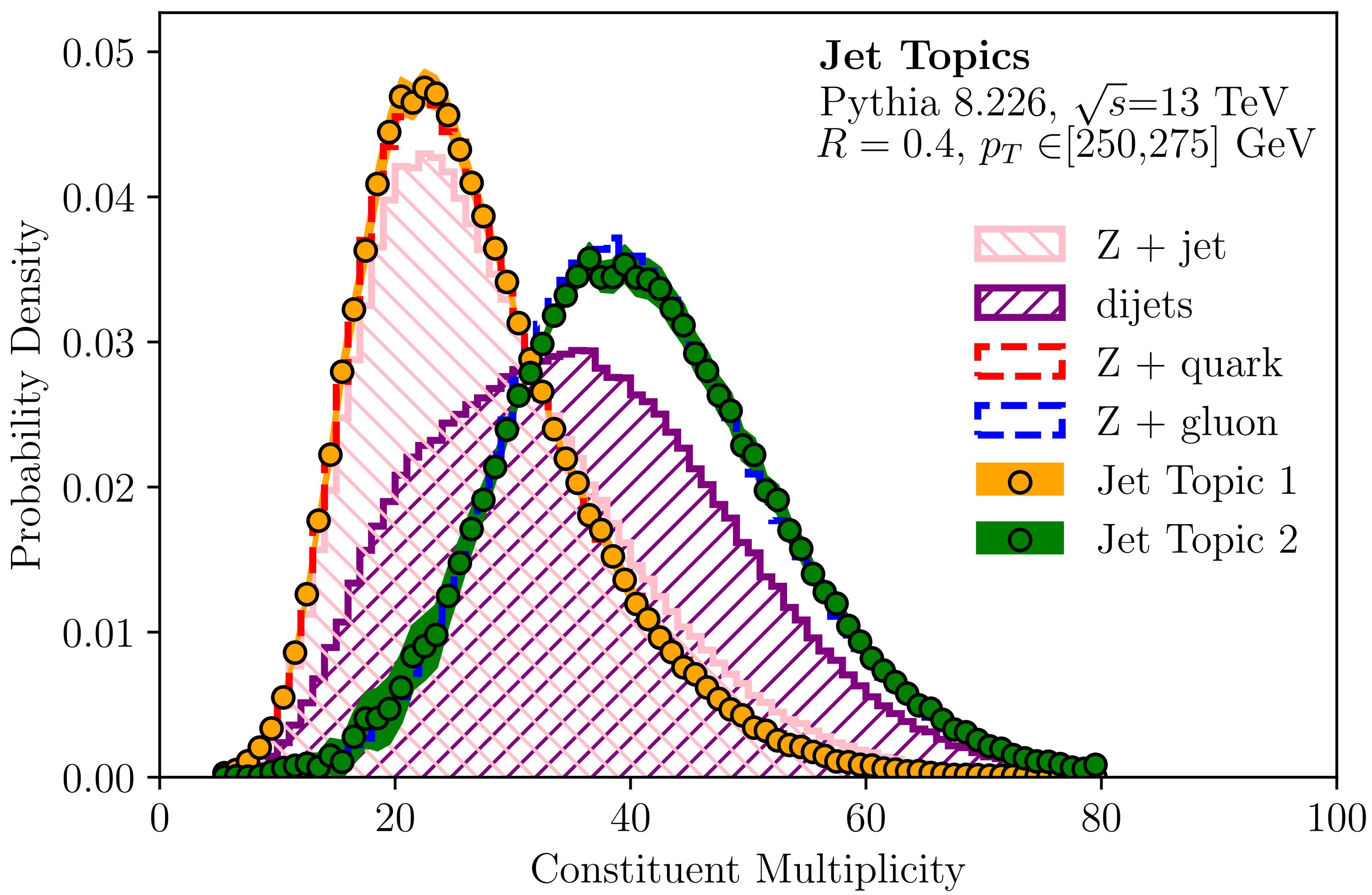
g

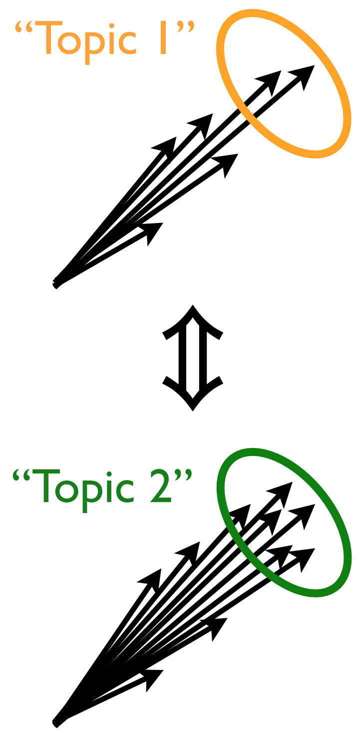


0++

Jet Topics

Extract jet categories from data…

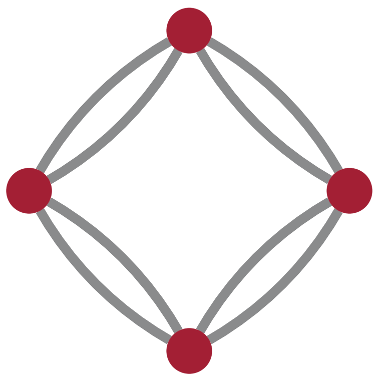




…solely* from the assumption they exist

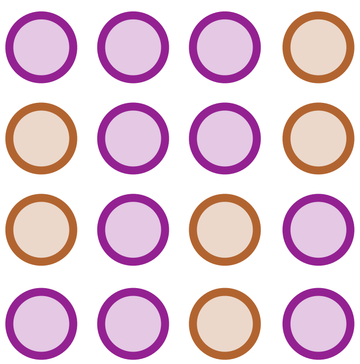
[Metodiev, JDT, 1802.00008]

Outline



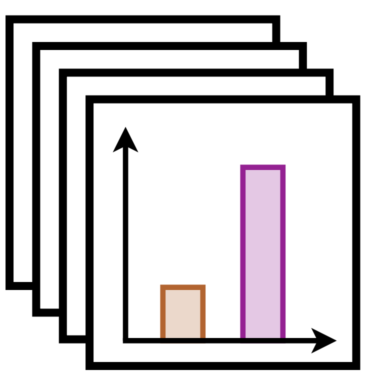
A Basis for Jet Substructure

"Solving" the problem of jet classification



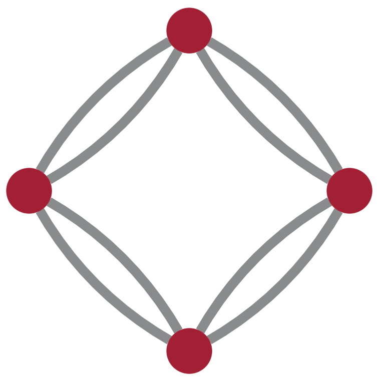
Learning Without Labels

Trustable training samples from data



Introducing Jet Topics

Well-defined categories by construction



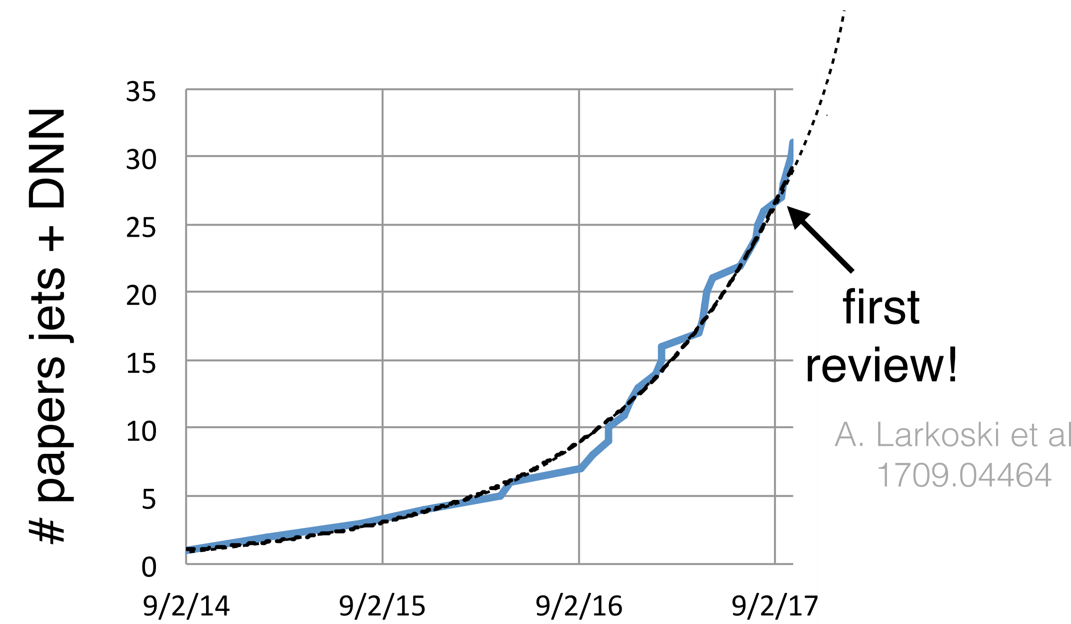
A Basis for Jet Substructure

Learning Without Labels

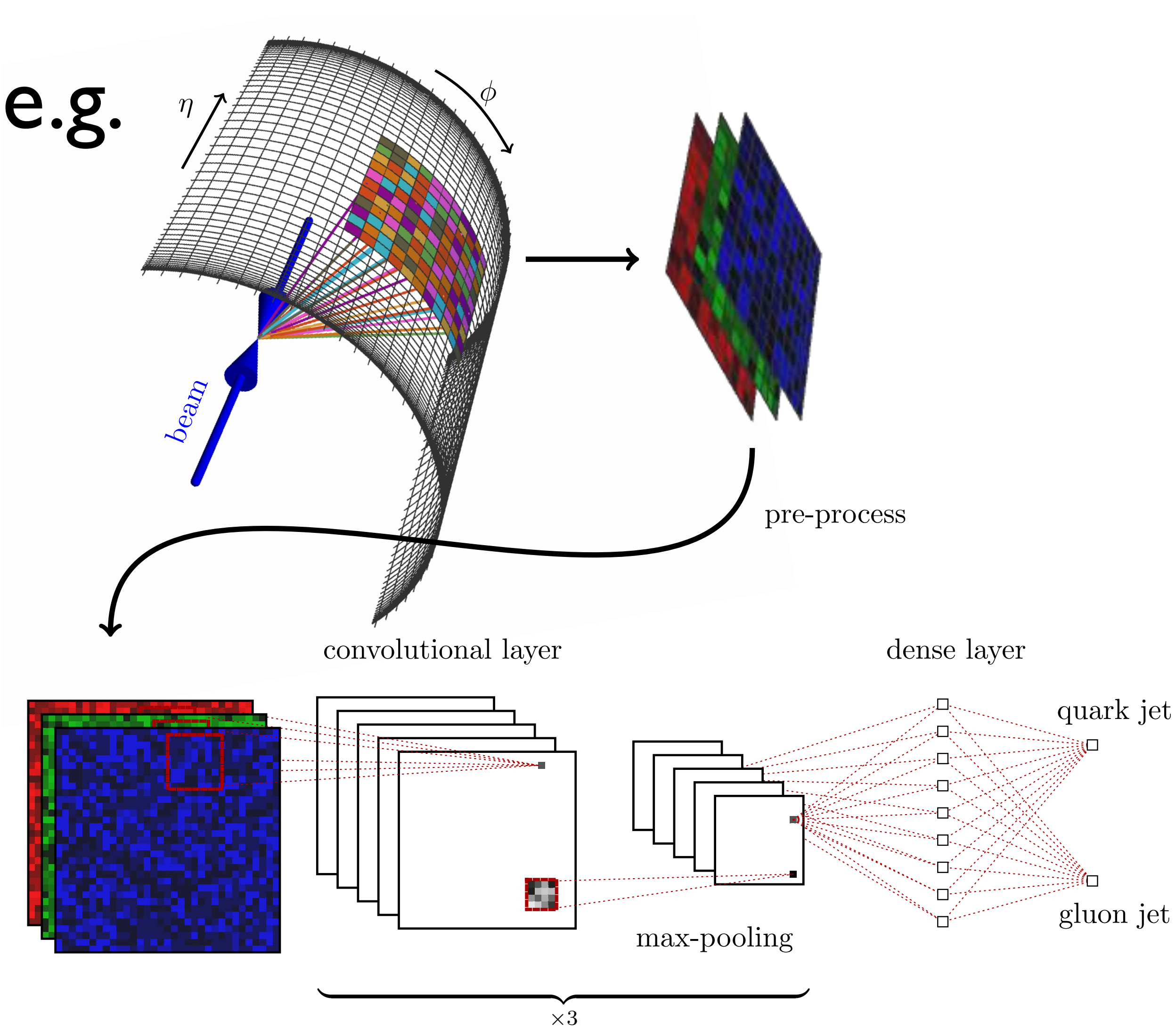
Introducing Jet Topics

10 Years of Jet Substructure!

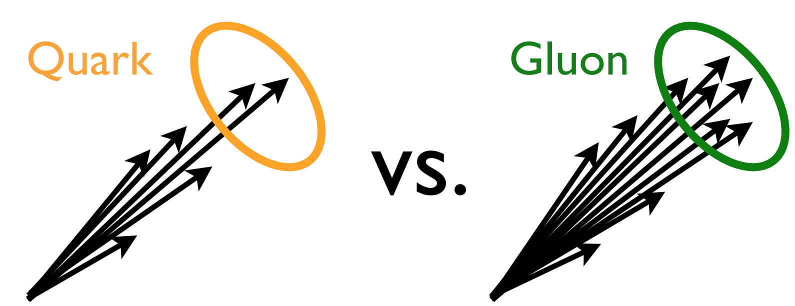
The Rise of Machine Learning for Jets



e.g.



$\Rightarrow$



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

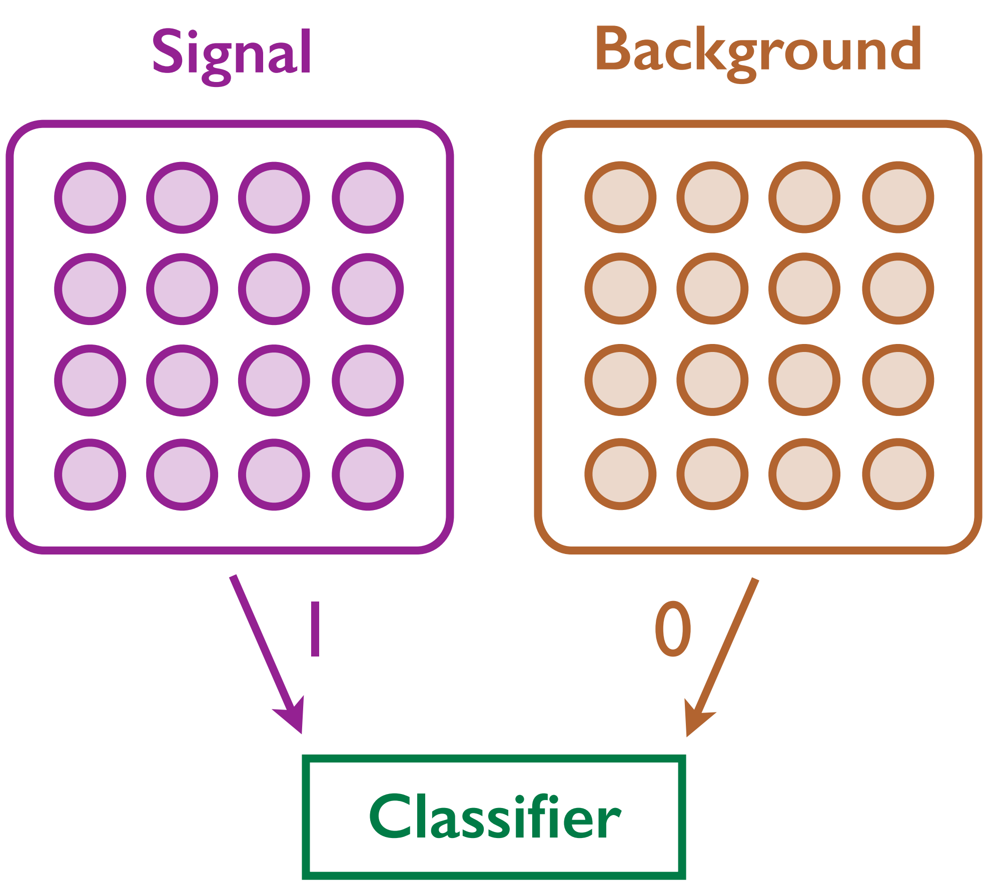
&

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

A Cartoon of Machine Learning

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

Set of observables



Minimize Loss Function

(assuming infinite training sets)

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

Optimal Classifier (Neyman–Pearson)

A Cartoon of Machine Learning

$\ell_{\text{MSE}} =$

background

S

Signal

Function

ning sets)

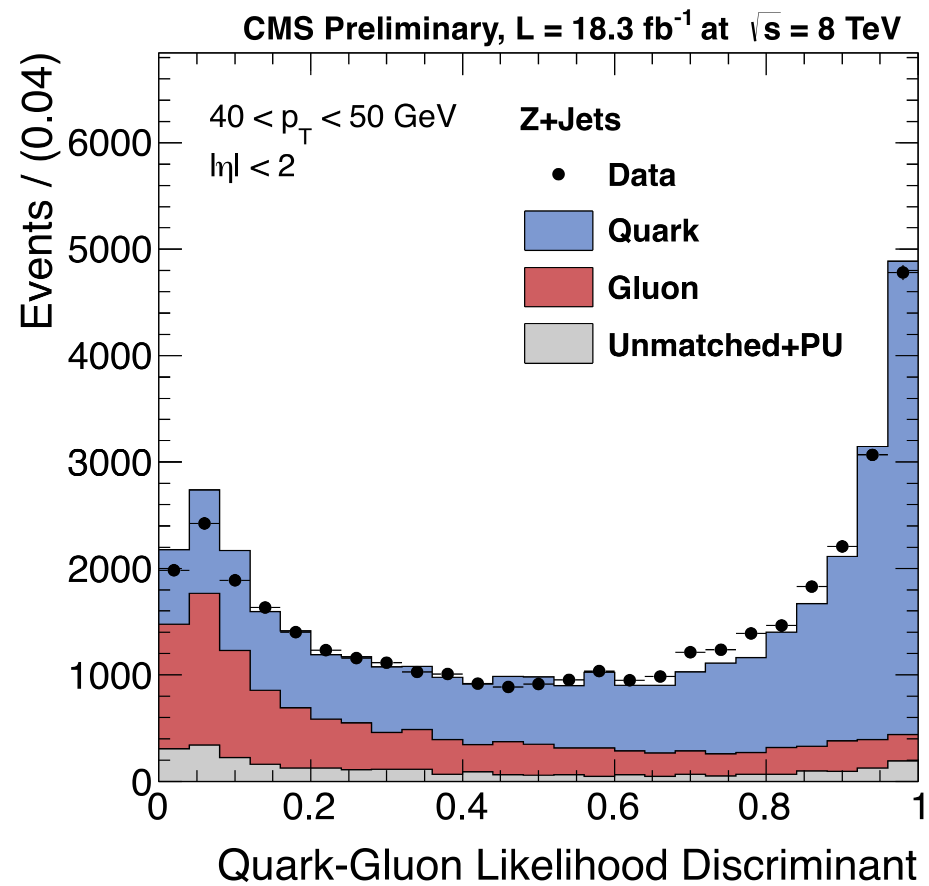
$(\vec{x})$

$p_{\text{bkgd}}(\vec{x})$

an–Pearson)

CMS Quark/Gluon Classifier

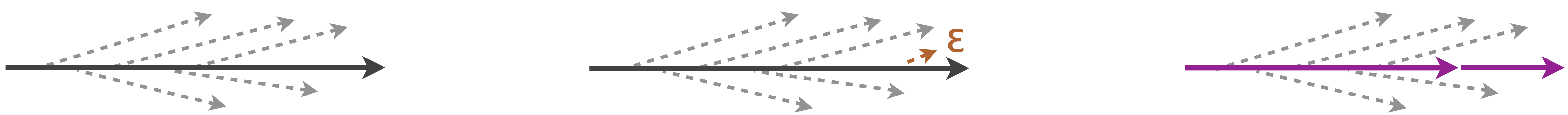
Three Observables (n_{had}, $p_T{}^D$, σ_2)



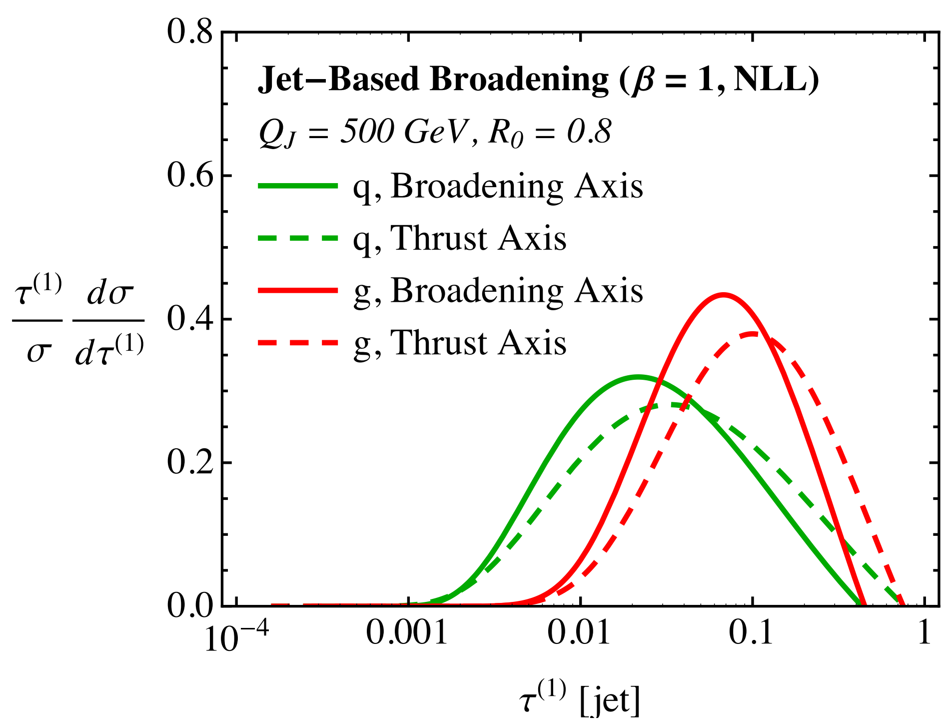
[CMS, JME-13-002]

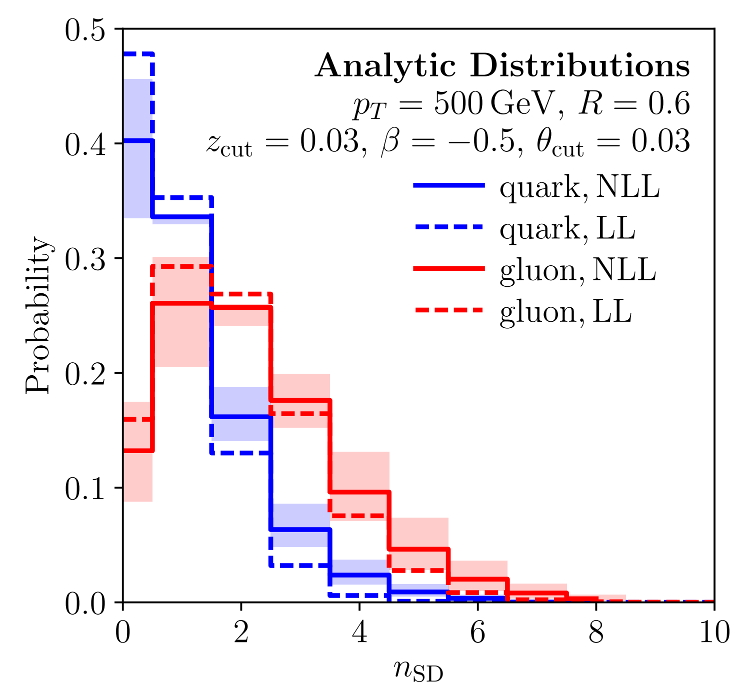
A Cartoon of Infrared/Collinear Safety

Original Jet | Infrared | Collinear



IRC Safe Observable: Insensitive to IR or C emissions





[e.g. Larkoski, Neill, JDT, 1401.2158; Frye, Larkoski, JDT, Zhou, 1704.06266]

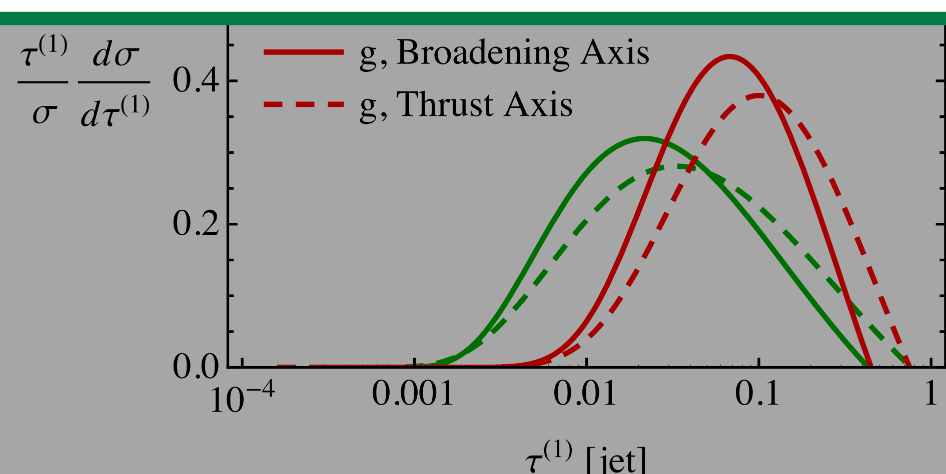
A Cartoon of Infrared/Collinear Safety

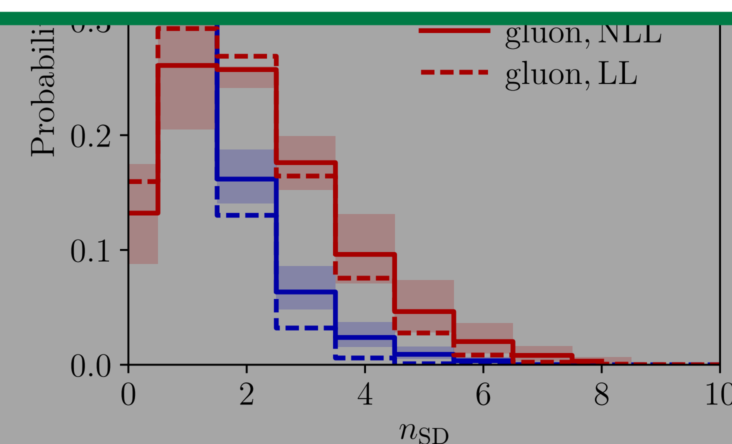
Original Jet

Infrared

Collinear

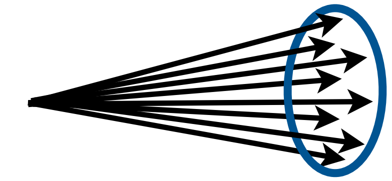
What if *optimal* jet classifier is IRC safe observable?





[e.g. Larkoski, Neill, JDT, 1401.2158; Frye, Larkoski, JDT, Zhou, 1704.06266]

A Systematic Expansion



Expand* any IRC safe observable in small energy limit

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

$$+ \sum_{ijk} E_i E_j E_k \, f_3^{\mathcal{S}}(\hat{n}_i, \hat{n}_j, \hat{n}_k) + \ldots$$

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]

Introducing the Energy Flow Polynomials

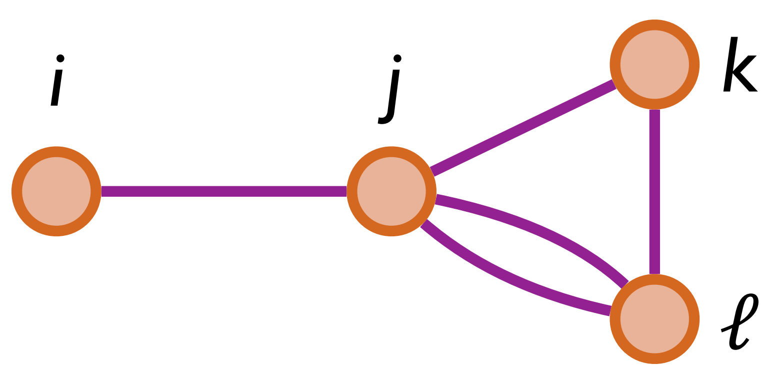
Multigraph *Angular Scaling*

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

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]

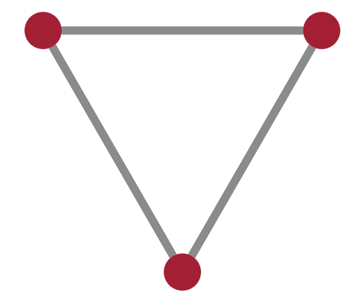
Down the Rabbit Hole

Known Structures:

${}_1e_2$

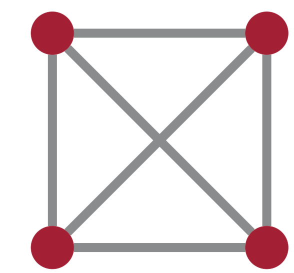
width (β=1), thrust (β=2)

${}_3e_3$



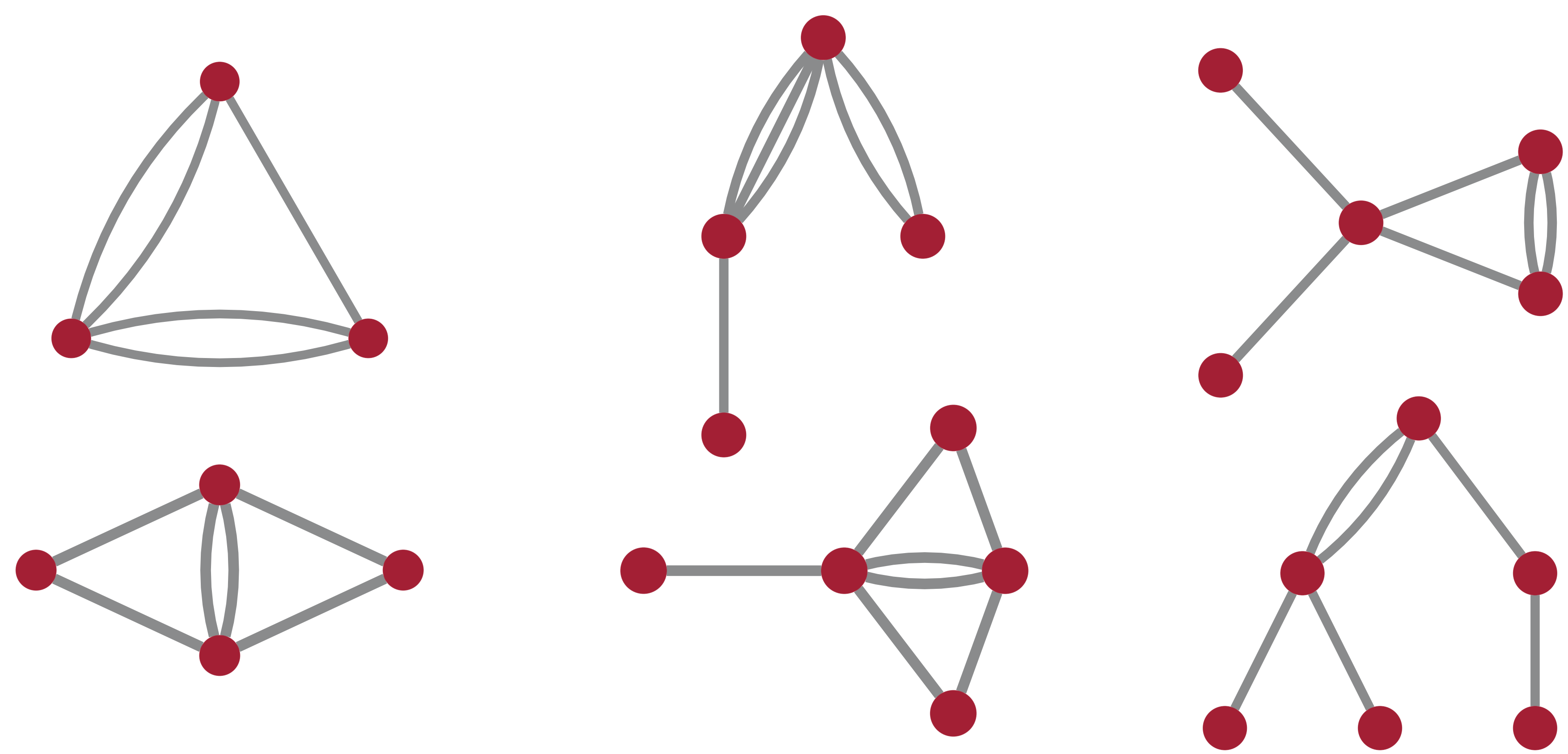
used in D_2

${}_6e_4$



used in C_3

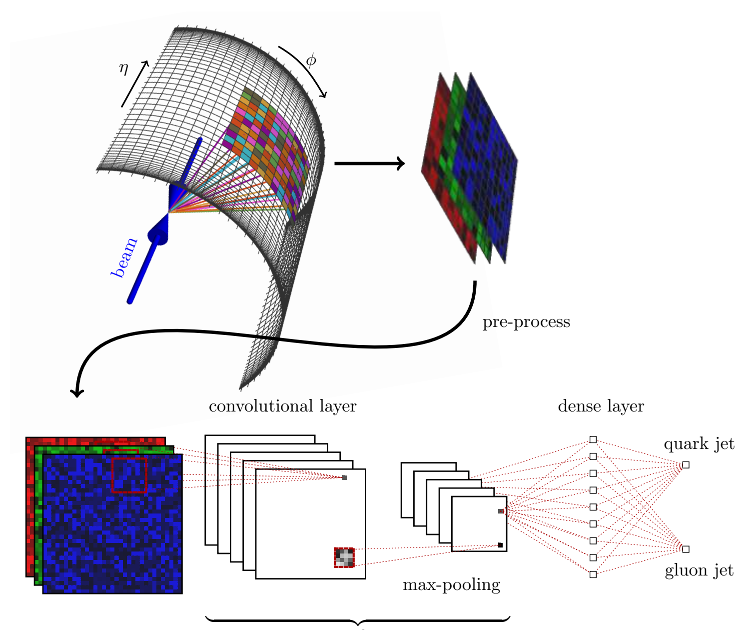
No Idea:

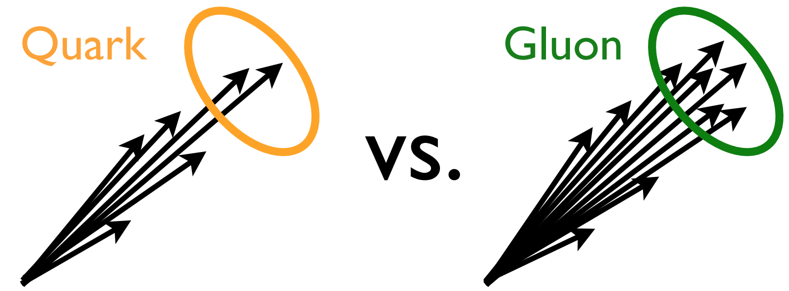


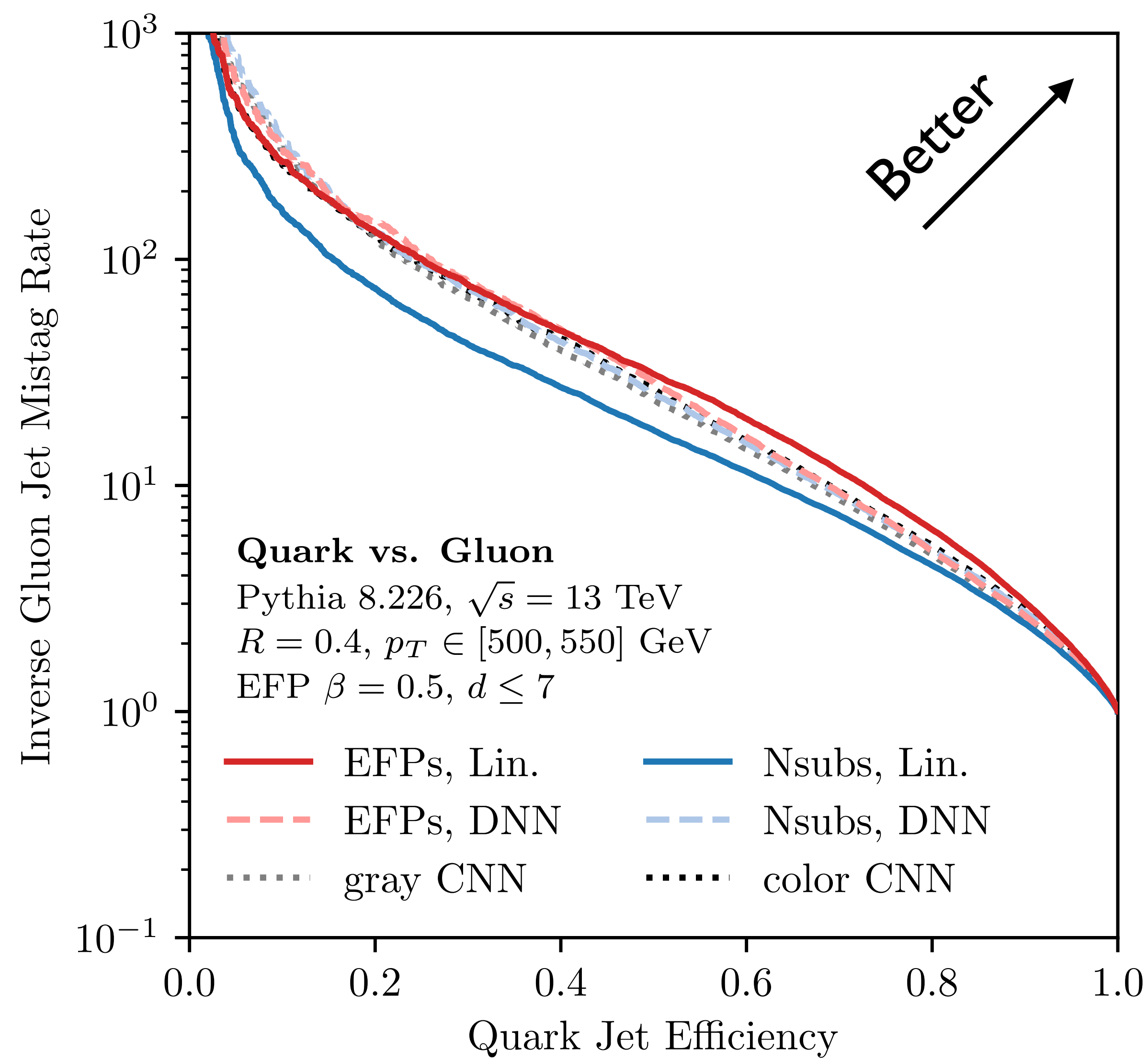
Linear Regression or Neural Network?

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

$$\approx$$

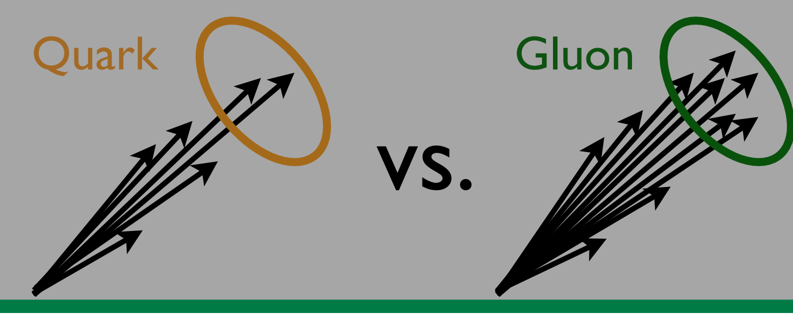






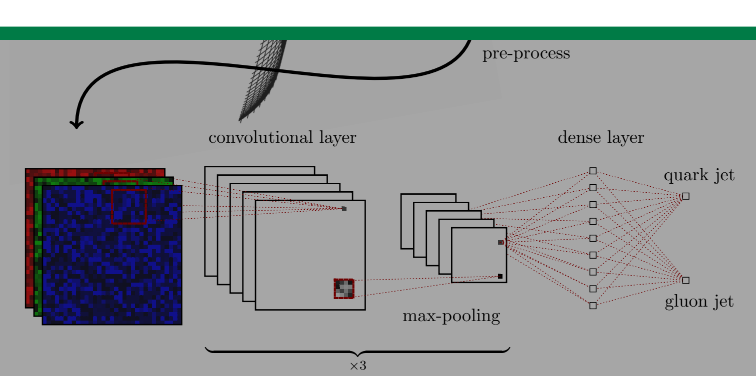
[Komiske, Metodiev, JDT, 1712.07124; Komiske, Metodiev, Schwartz, 1612.01551]

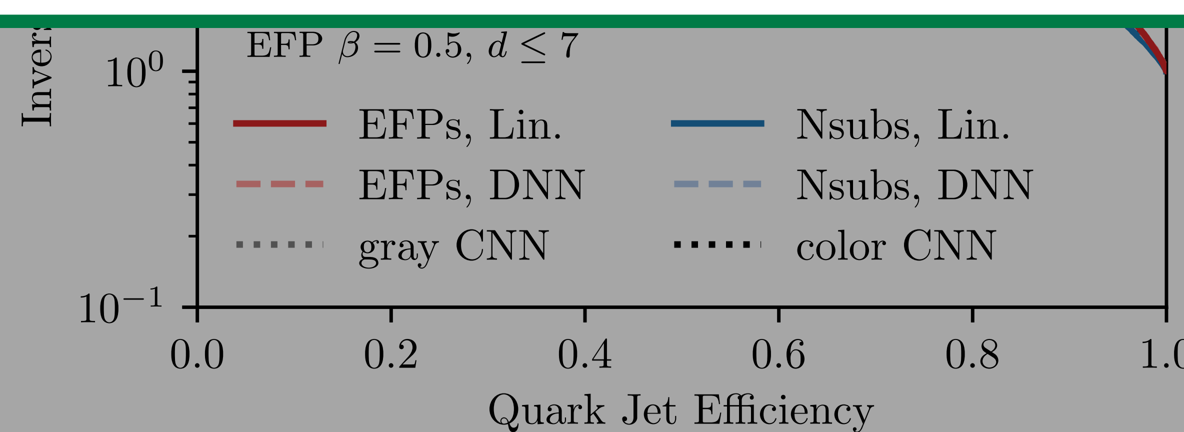
Linear Regression or Neural Network?



Bottom Line:
Jet Classification is "Solved"

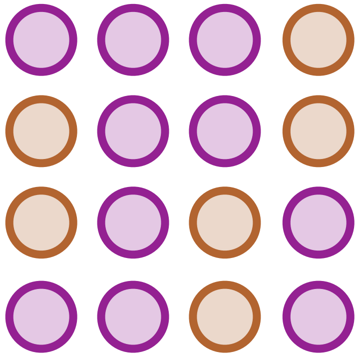
Assuming trustable training samples, well-defined categories, etc.





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

A Basis for Jet Substructure

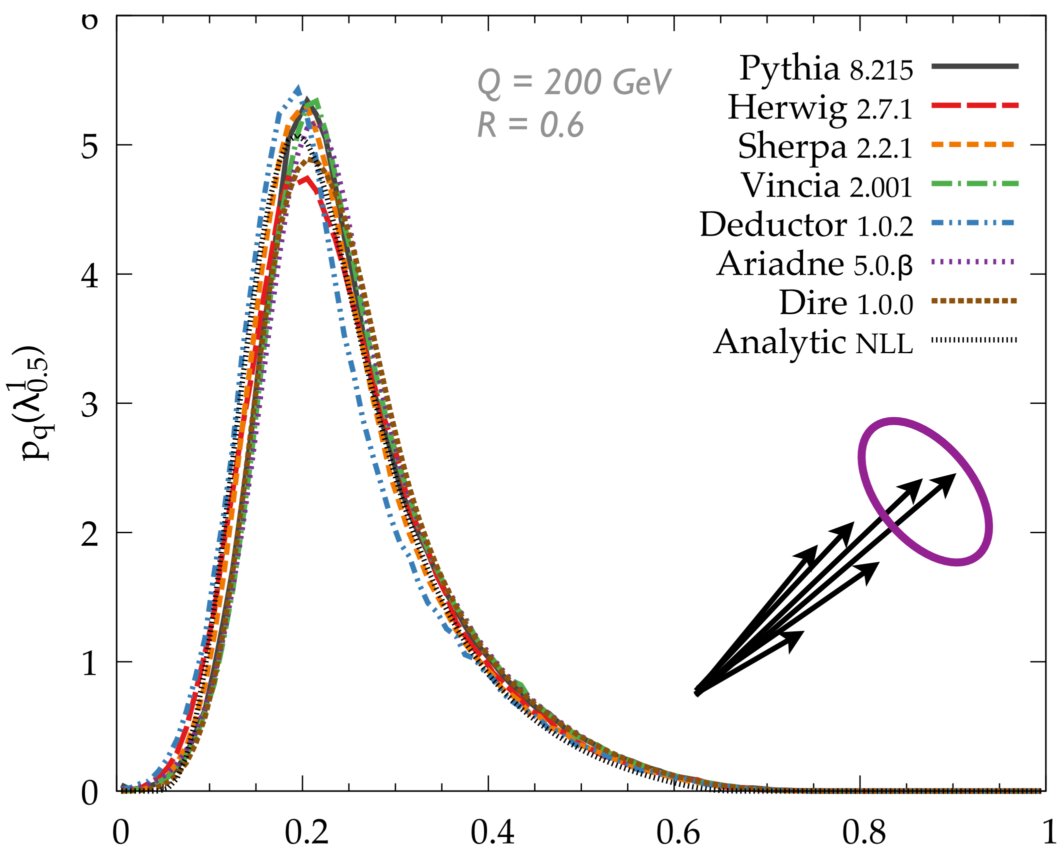


Learning Without Labels

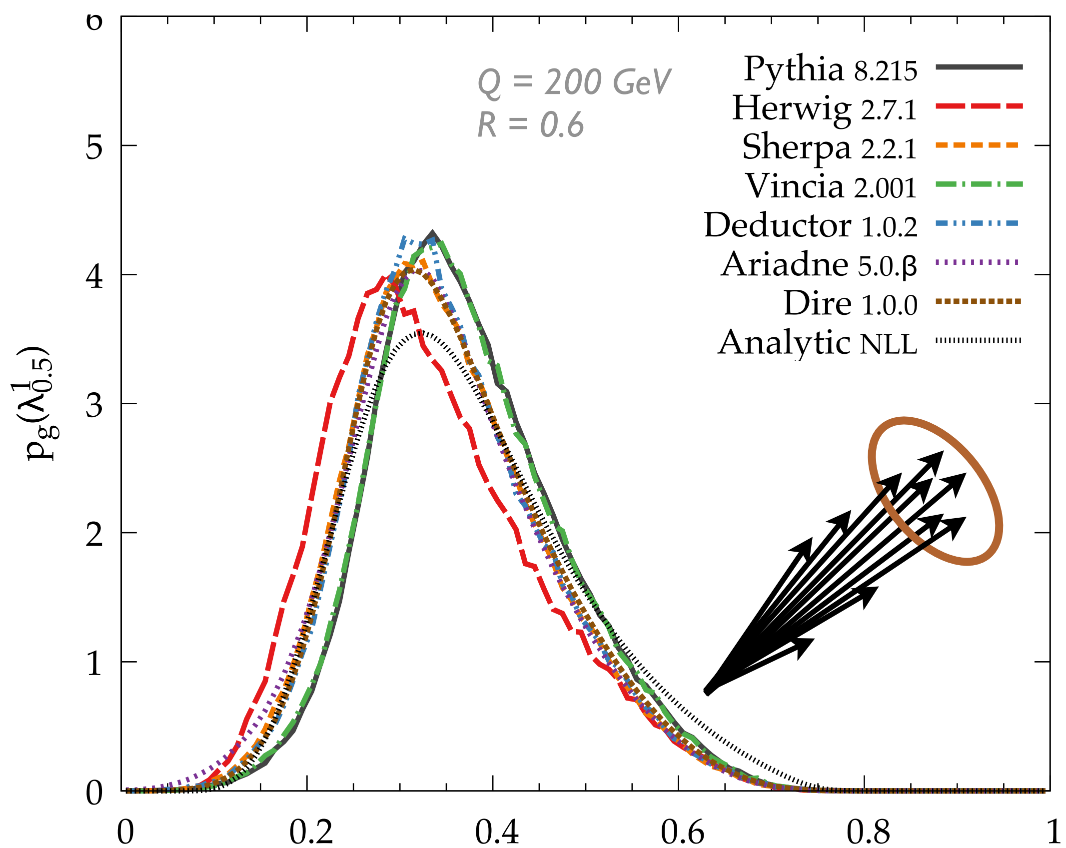
Introducing Jet Topics

Trustable Training Samples?

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



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

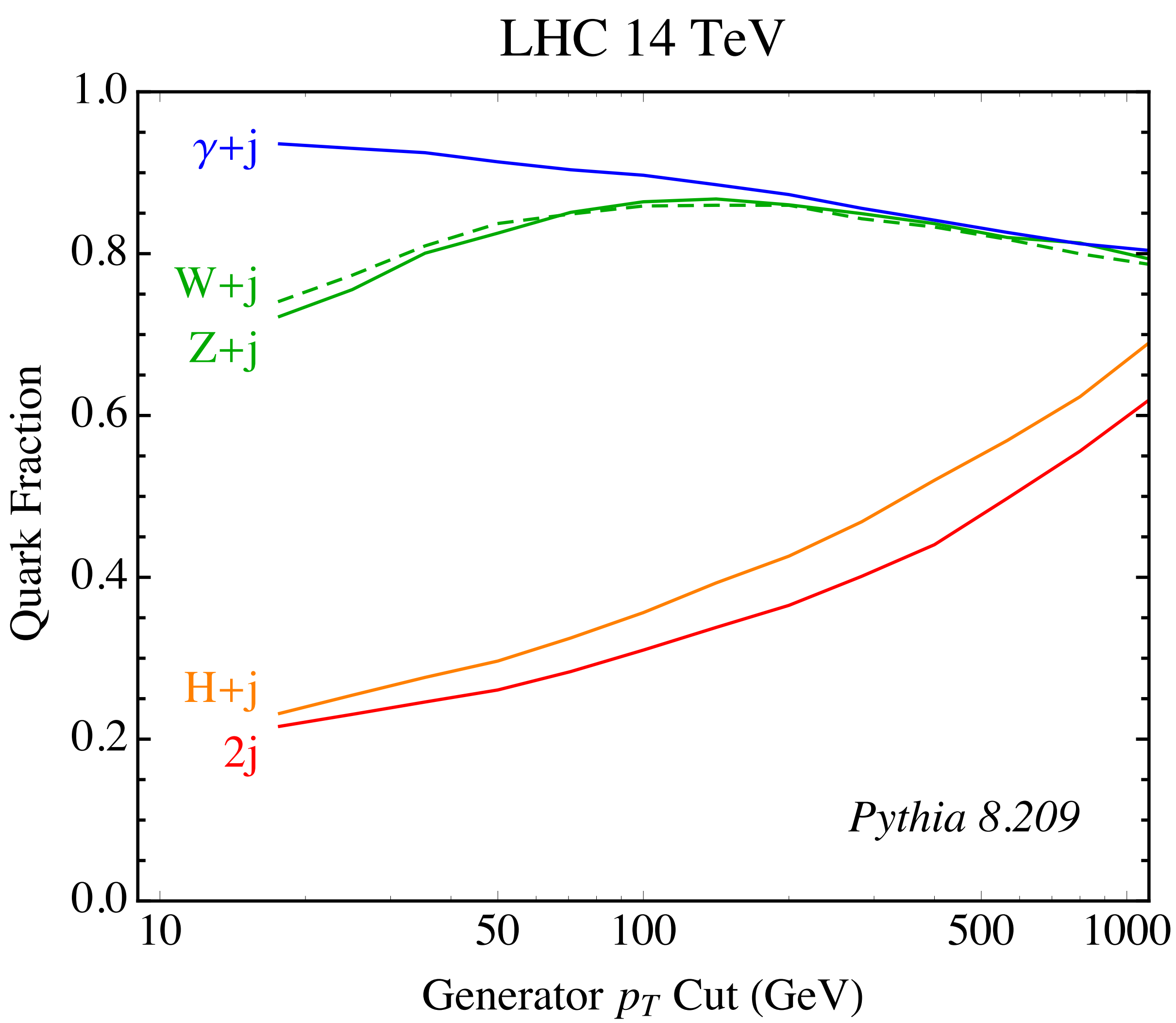


$$\mathrm{LHA} = \sum_i z_i \sqrt{\theta_i}$$

Large 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]

Quark vs. Gluon from Data?



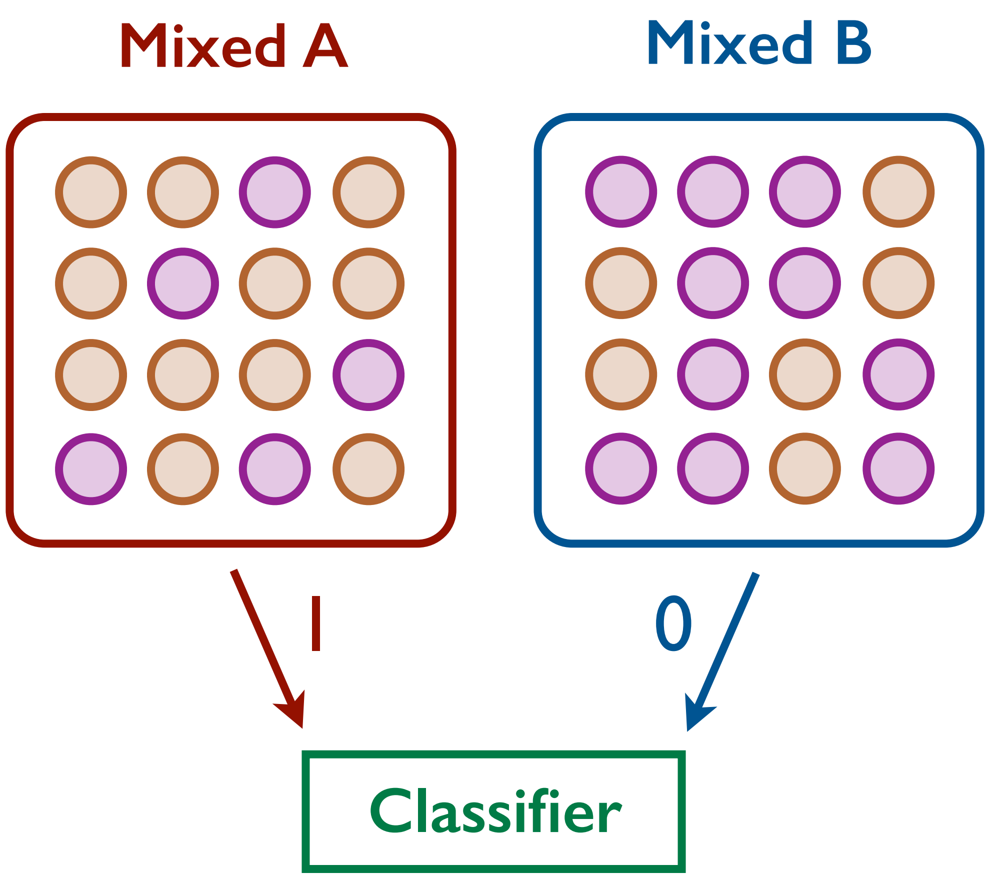
Plenty of (mixed) jets to study!

(Though plenty of uncertainties on quark/gluon fractions…)

[Gras, Hoeche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, 1704.03878; see also Gallicchio, Schwartz, 1104.1175]

Key Challenge: Mixed Samples are Mixtures

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



Mixed Classifier?

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

$$\neq$$

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

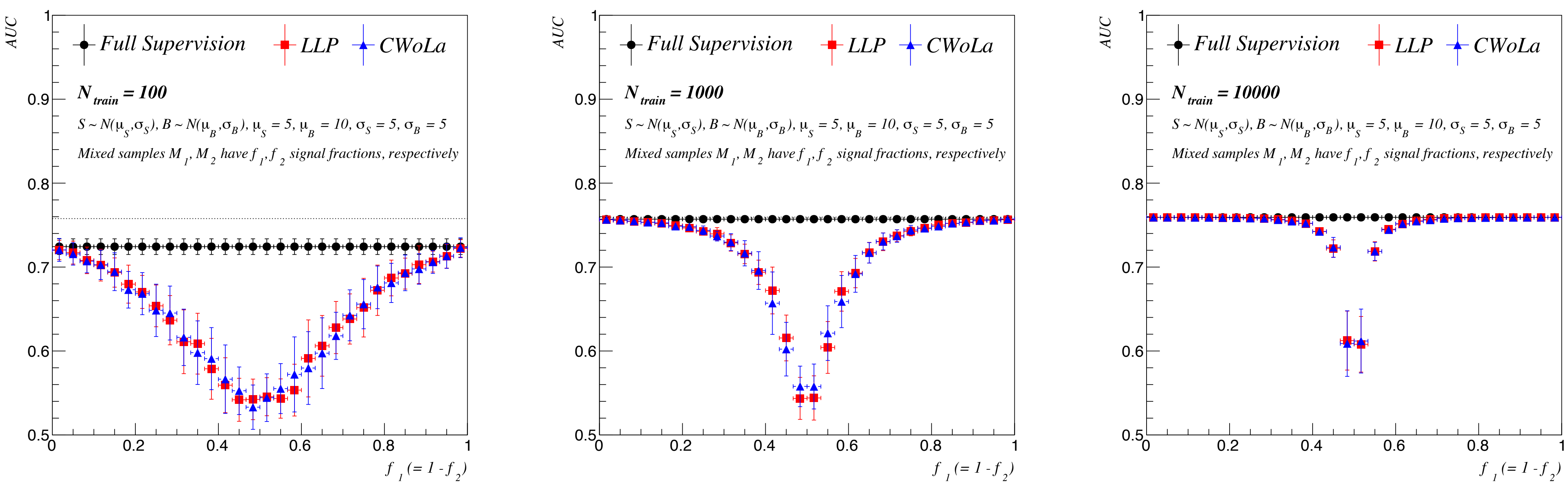
but... $$\frac{\partial h_{\text{mixed}}(\vec{x})}{\partial h_{\text{pure}}(\vec{x})} > 0$$

[Metodiev, Nachman, JDT, 1708.02949; see also Cranmer, Pavez, Louppe, 1506.02169; Blanchard, Flaska, Handy, Pozzi, Scott, 2016; Dery, Nachman, Rubbo, Schwartzman, 1702.00414; Cohen, Freytsis, Ostdiek, 1706.09451]

Key Challenge: Mixed Samples are Mixtures

Classification Without Labels

Slower training, but same ultimate performance



(Subtlety: Some fraction information needed to calibrate classifier)

[Metodiev, Nachman, JDT, 1708.02949; see also Cranmer, Pavez, Louppe, 1506.02169; Blanchard, Flaska, Handy, Pozzi, Scott, 2016; Dery, Nachman, Rubbo, Schwartzman, 1702.00414; Cohen, Freytsis, Ostdiek, 1706.09451]

Key Assumption: Mixed Samples are Mixtures

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

Sensible? No! Well, ok…

Sample Dependence

"Quark jet" in dijets vs. Z+jets are different
because of color correlations with rest of event

Approximate Sample Independence

Differences are power suppressed with small radius jets
Differences can be mitigated using jet grooming

[see Banfi, Dasgupta, Khelifa-Kerfa, Marzani, 1004.3483; Frye, Larkoski, Schwartz, Yan, 1603.06375, 1603.09338]

Key Assumption: Mixed Samples are Mixtures

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

Bottom Line:

Jet Classification is "Solved"

with trustable mixed training samples from data

Assuming sample independence, well-defined categories, etc.

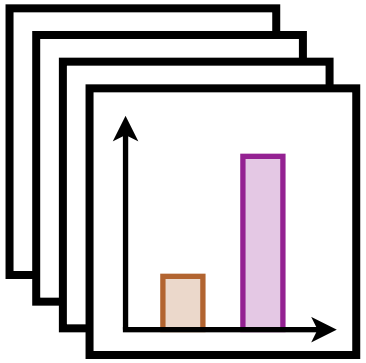
Approximate Sample Independence

Differences are power suppressed with small radius jets
Differences can be mitigated using jet grooming

[see Banfi, Dasgupta, Khelifa-Kerfa, Marzani, 1004.3483; Frye, Larkoski, Schwartz, Yan, 1603.06375, 1603.09338]

A Basis for Jet Substructure

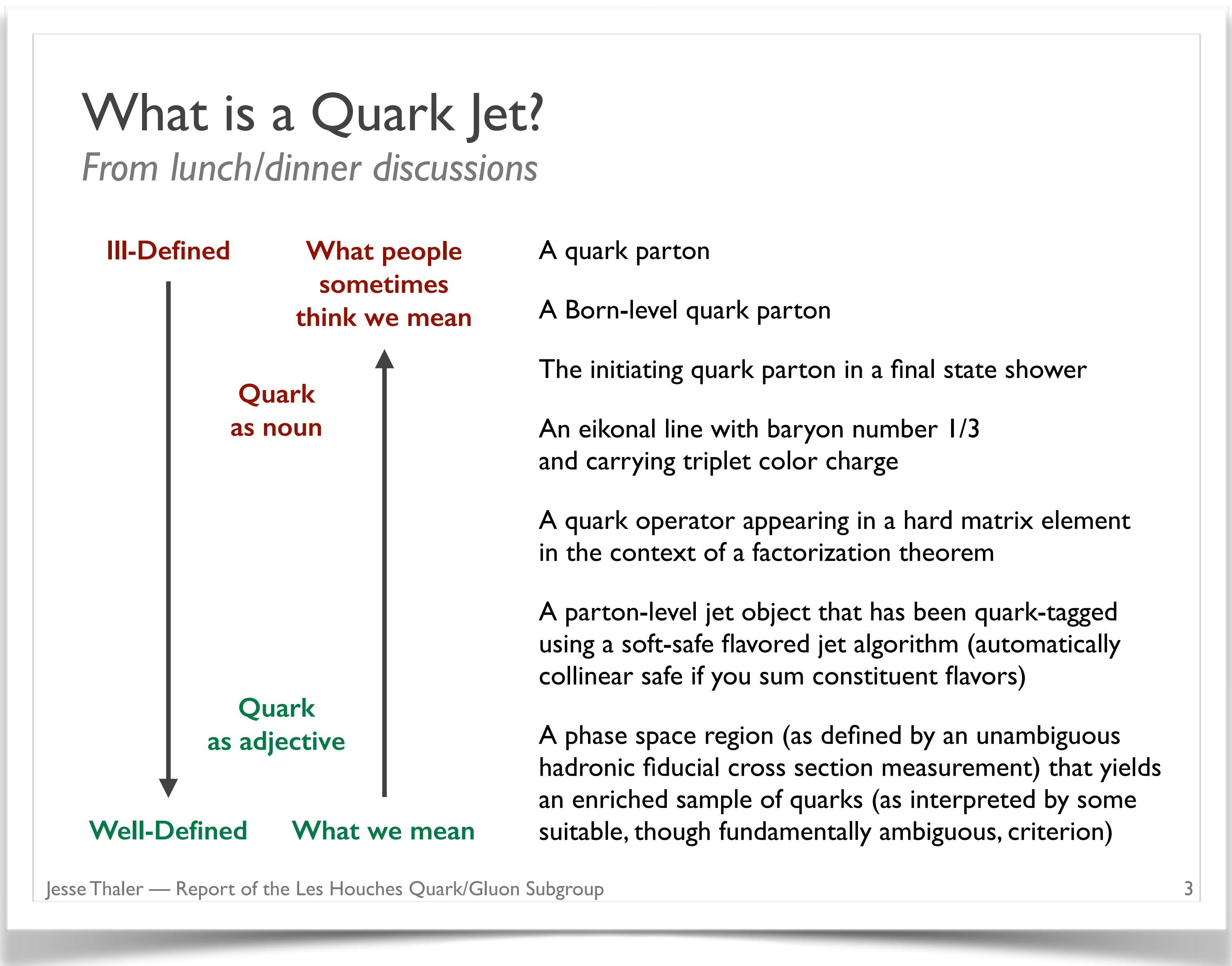
Learning Without Labels



Introducing Jet Topics

Well-Defined Categories?

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



[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]

Assume "Quark" and "Gluon" Exist

i.e. Sample Independence

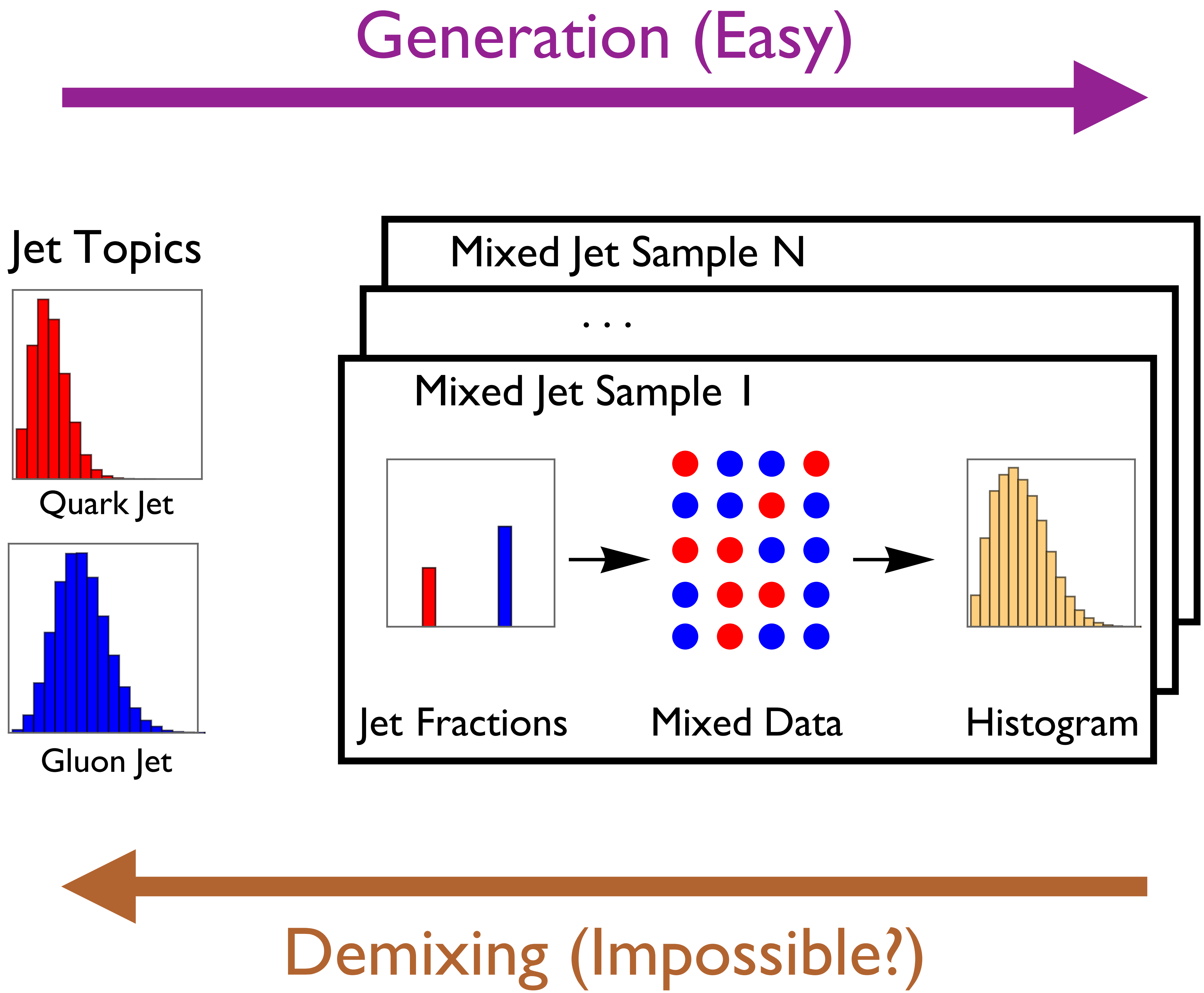
$$p_{\text{mixed A}}(\vec{x}) = f_q^A \, p_{\text{quark}}(\vec{x}) + (1 - f_q^A) \, p_{\text{gluon}}(\vec{x})$$
$$p_{\text{mixed B}}(\vec{x}) = f_q^B \, p_{\text{quark}}(\vec{x}) + (1 - f_q^B) \, p_{\text{gluon}}(\vec{x})$$

If you can extract these…

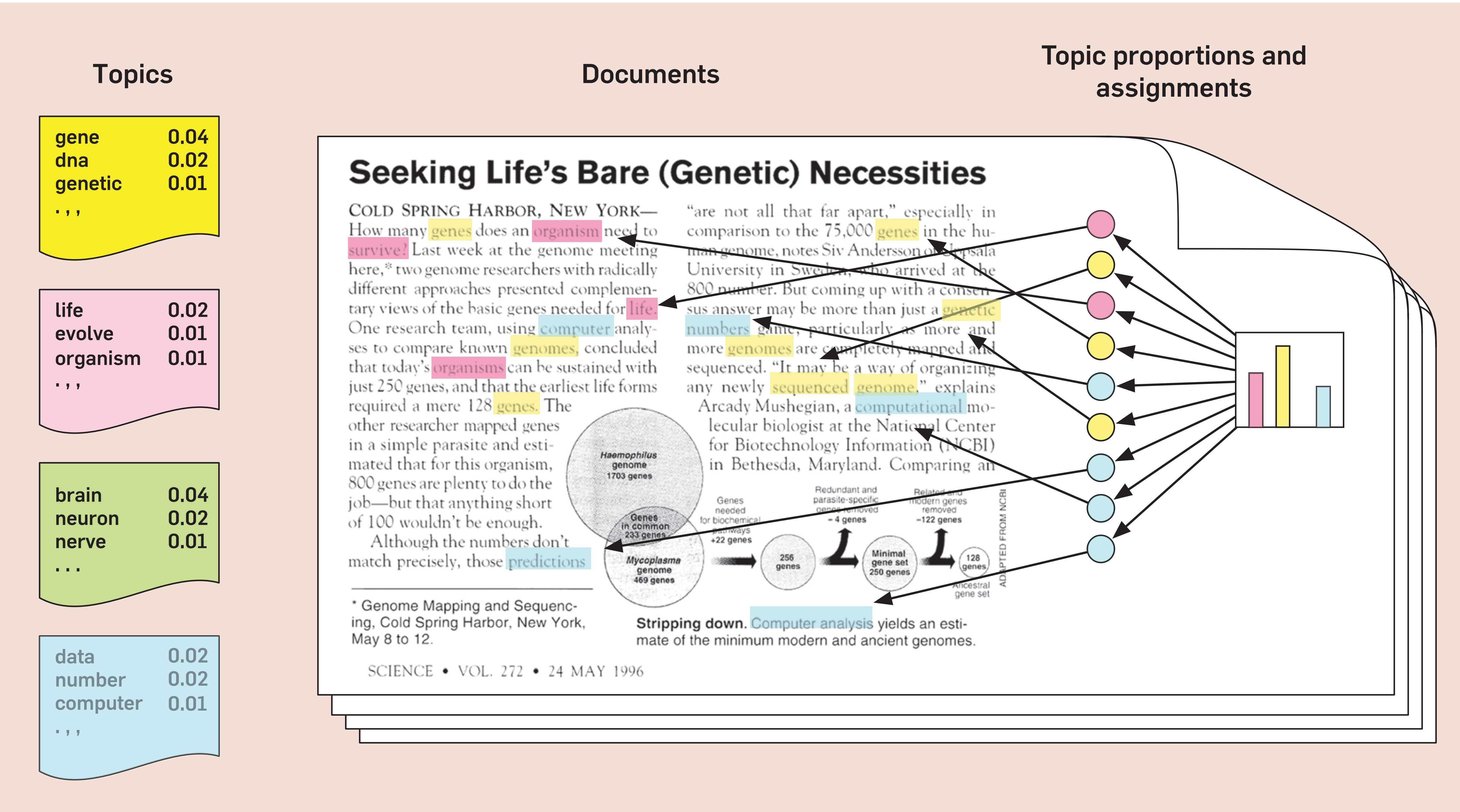
$$f_q^A \quad f_q^B \quad p_{\text{quark}}(\vec{x}) \quad p_{\text{gluon}}(\vec{x})$$

…then you have effectively defined "quark/gluon"

Too good to be true? Or already solved?

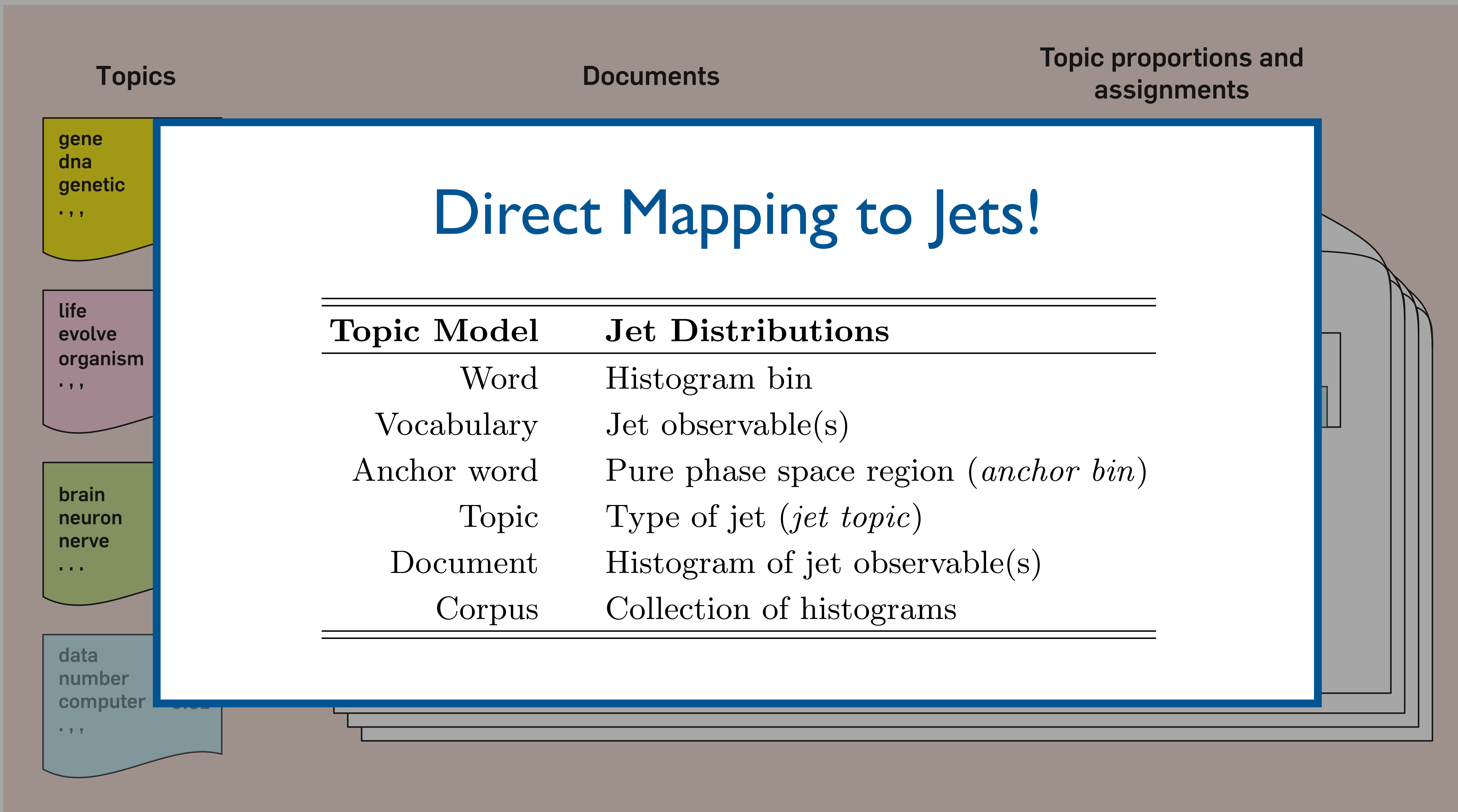

Generation (Easy)
Jet Topics
Quark Jet
Gluon Jet
Mixed Jet Sample N
. . .
Mixed Jet Sample 1
Jet Fractions
Mixed Data
Histogram
Demixing (Impossible?)

Topic Modeling



[Blei, 2012]

Topic Modeling

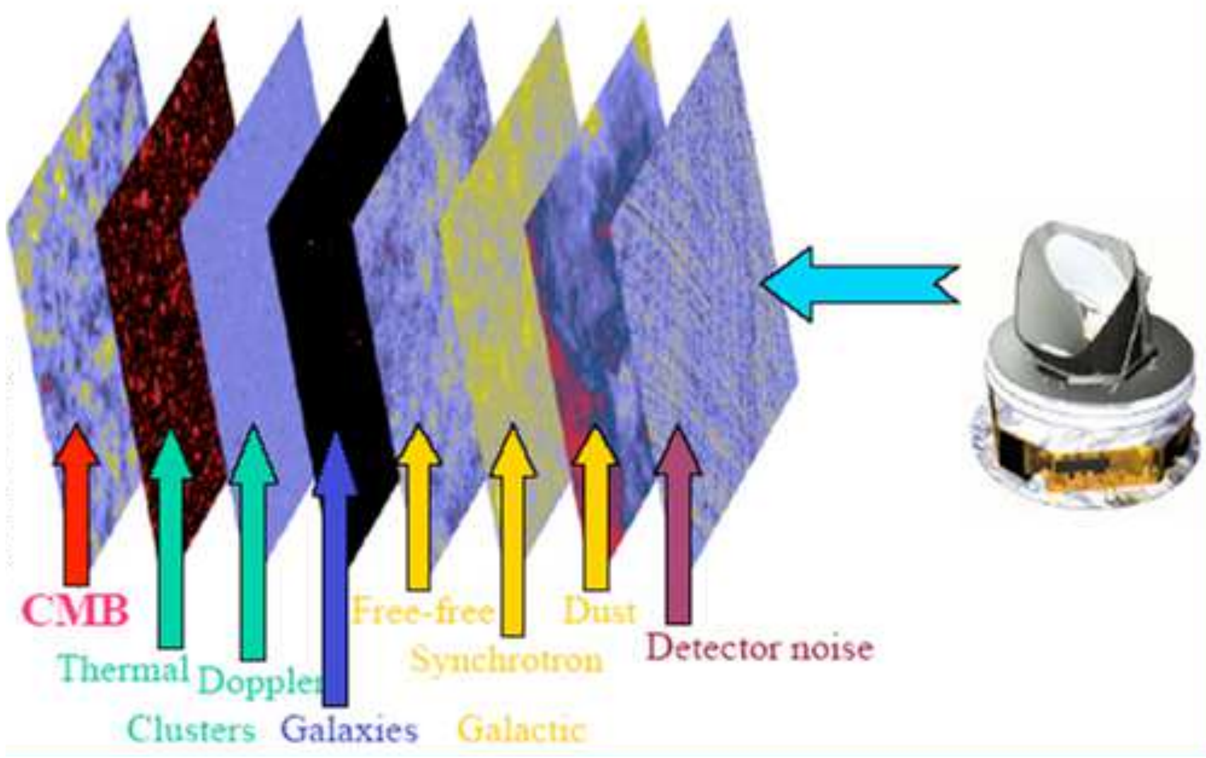


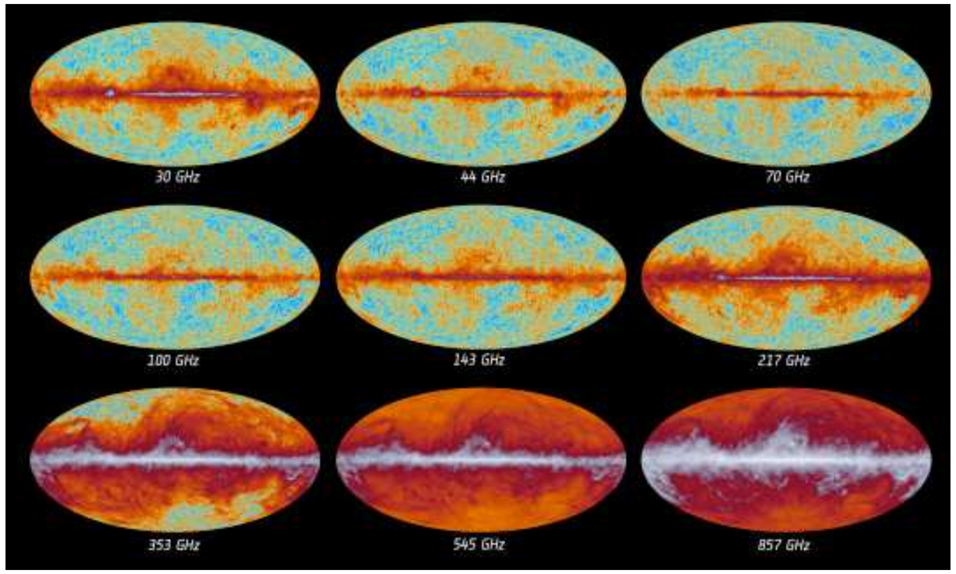
Direct Mapping to Jets!

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

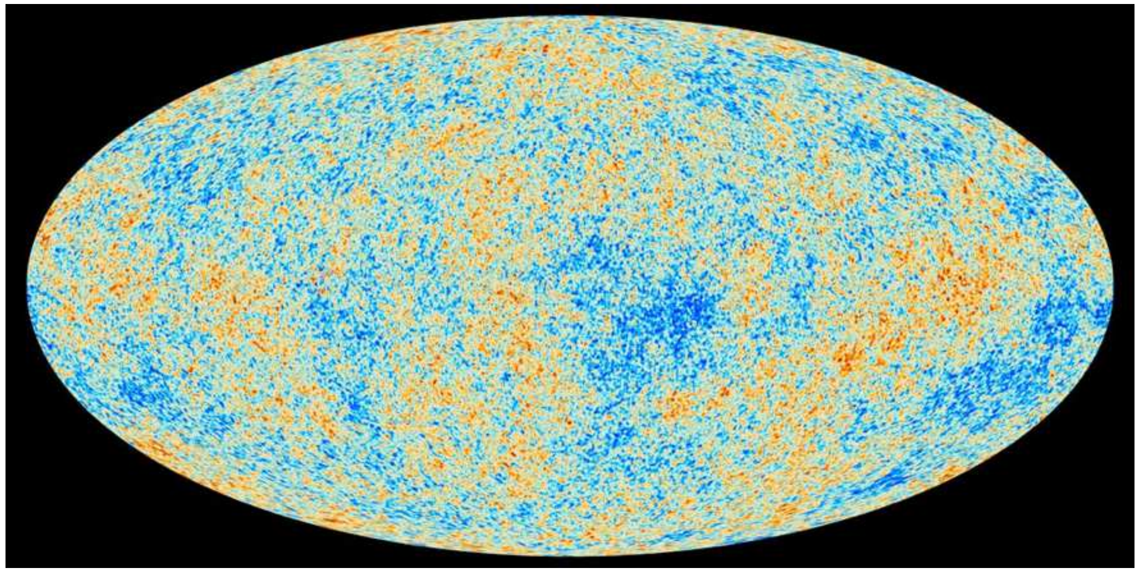
[Blei, 2012]

Related to CMB Foreground Separation





$\Rightarrow$



[Planck Outreach]

The Demix Algorithm

Simplifying to two mixtures of two topics

Just subtract the mixed distributions!

$$p_{\mathrm{T1}}(\vec{x}) = \frac{p_A(\vec{x}) - p_B(\vec{x})\, \kappa_{A|B}}{1 - \kappa_{A|B}}$$

$$p_{\mathrm{T2}}(\vec{x}) = \frac{p_B(\vec{x}) - p_A(\vec{x})\, \kappa_{B|A}}{1 - \kappa_{B|A}}$$

Reducibility Factors

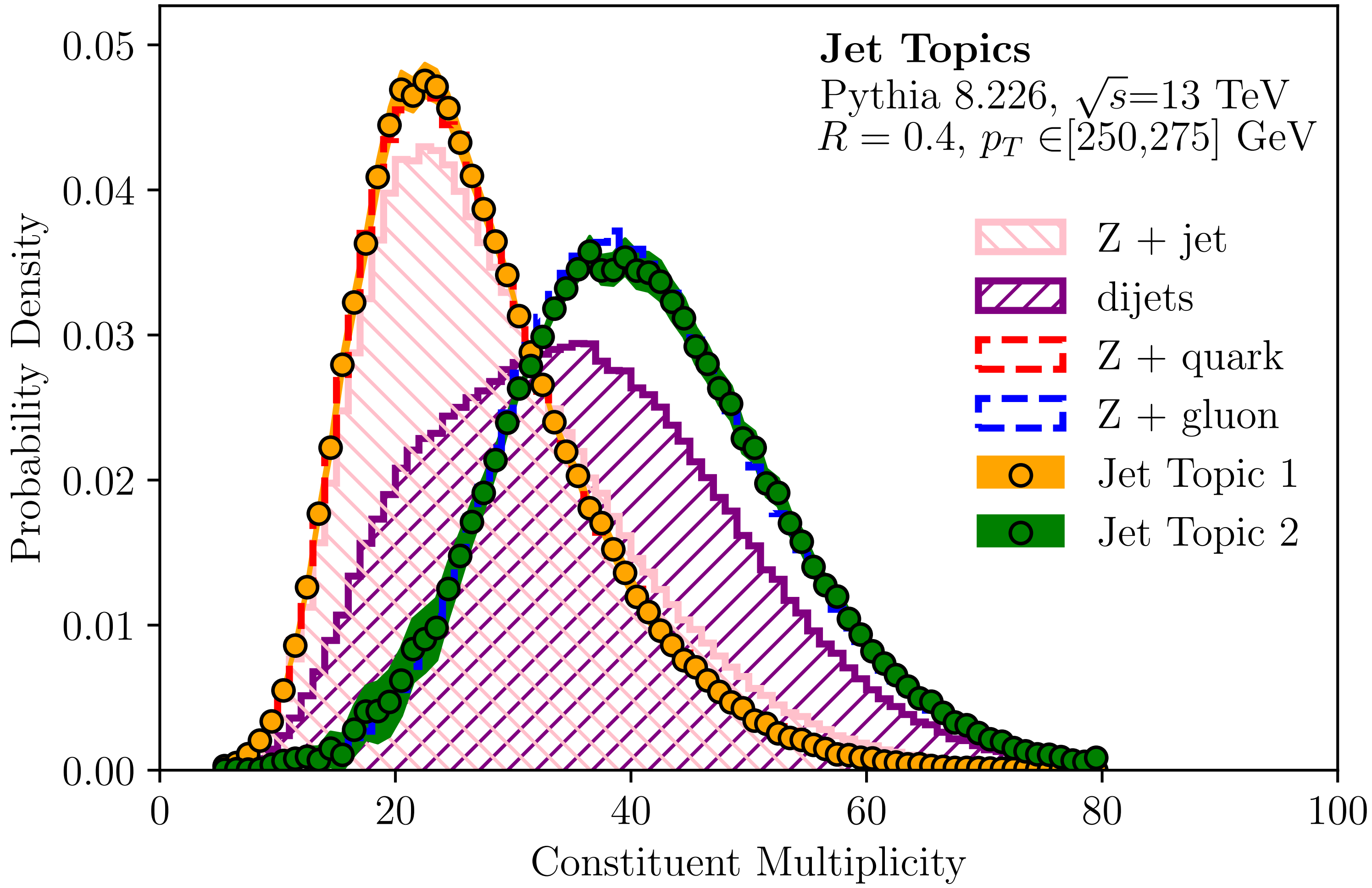
Requires Anchor Bins / "Mutual Irreducibility"

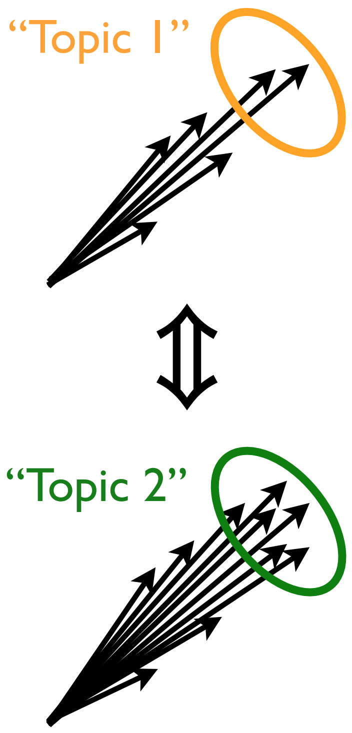
Region of 100% purity for each topic (even if tiny efficiency)
Probabilities are positive, so make κ as large as possible

[Katz-Samuels, Blanchard, Scott, 1710.01167]

Jet Topics

Extract jet categories from data…



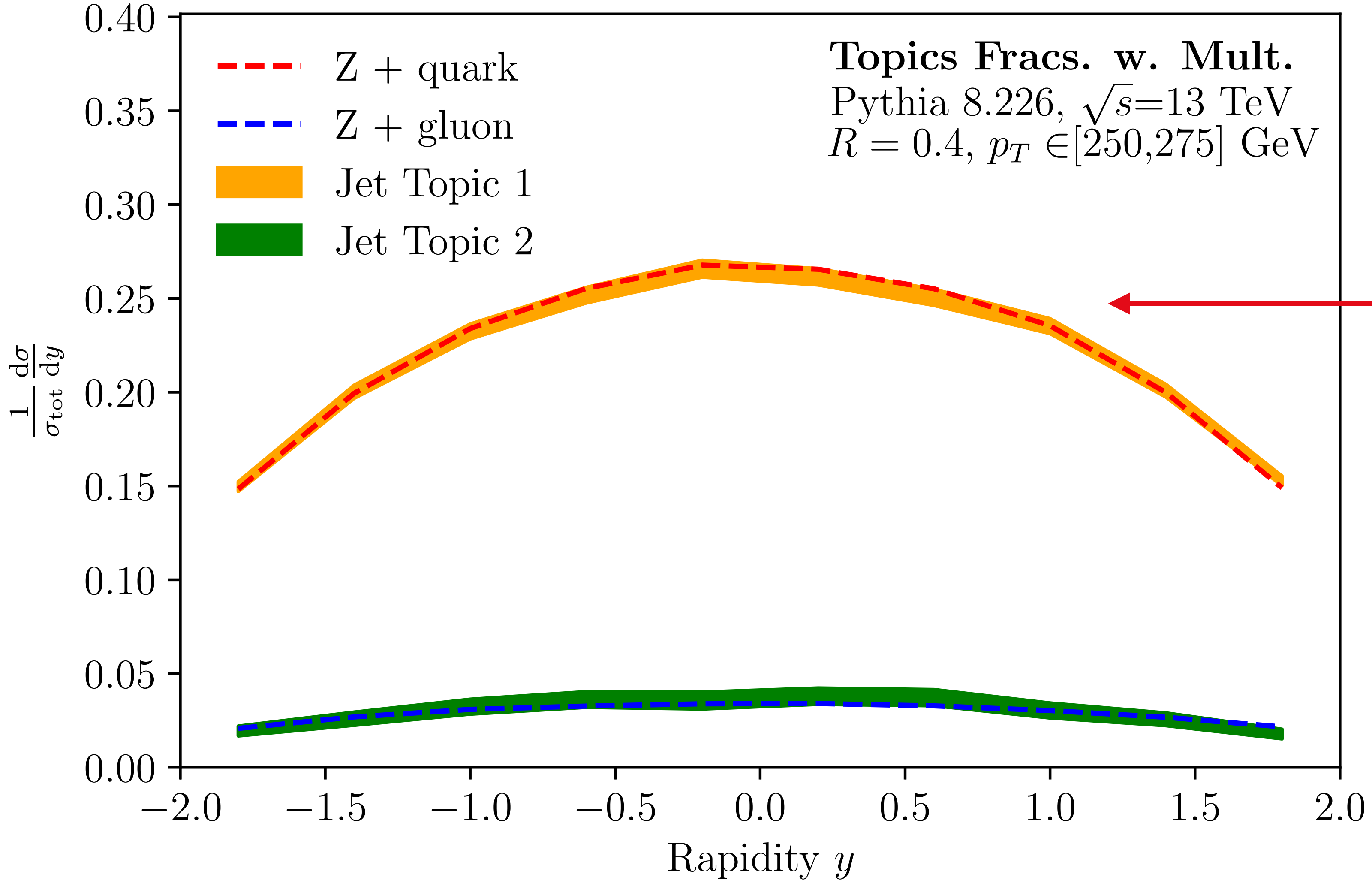


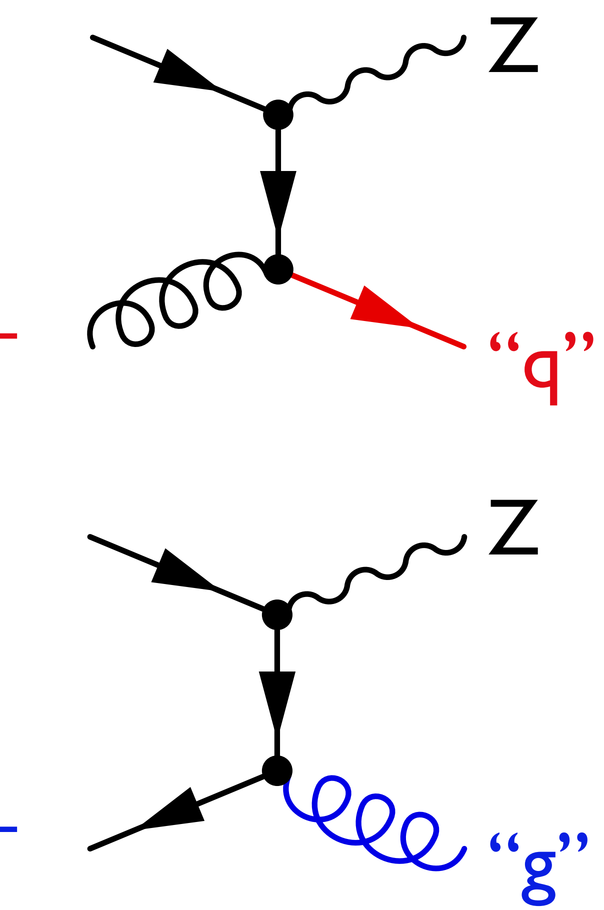
…solely* from the assumption they exist

Sample Independence, Different Fractions, Mutual Irreducibility

[Metodiev, JDT, 1802.00008]

“Parton”-Labeled Cross Sections?



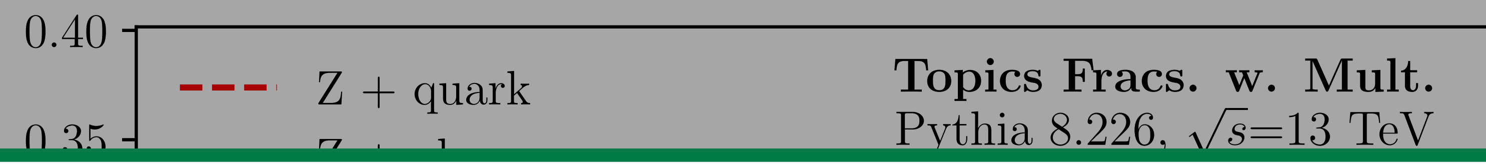


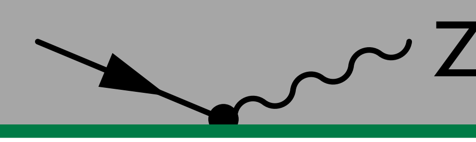
Implications for PDF extraction?

Key challenge: Defining jet topics at fixed order

[Metodiev, JDT, 1802.00008]

“Parton”-Labeled Cross Sections?





Bottom Line:
Jet Classification is “Solved”

with trustable mixed training samples from data
and well-defined categories from high-purity regions

Assuming sample independence, mutual irreducibility, etc.

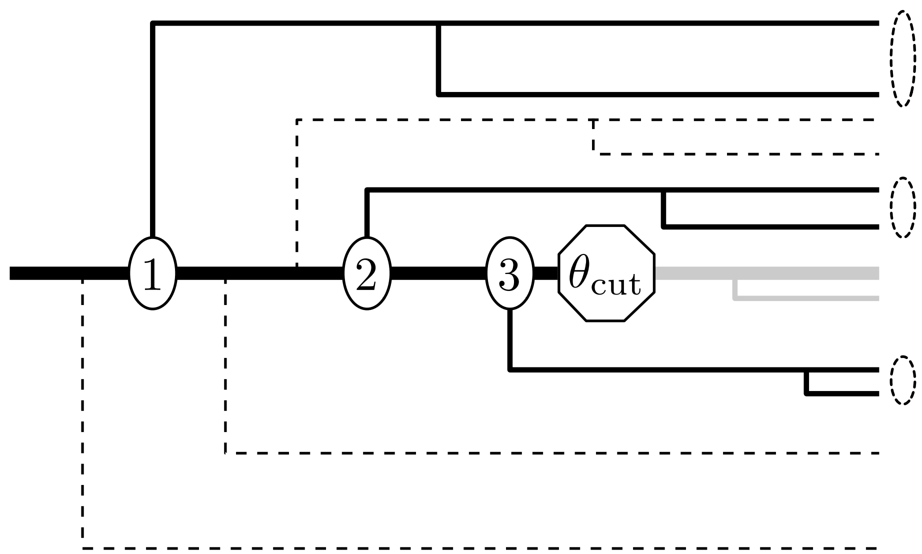
Implications for PDF extraction?

Key challenge: Defining jet topics at fixed order

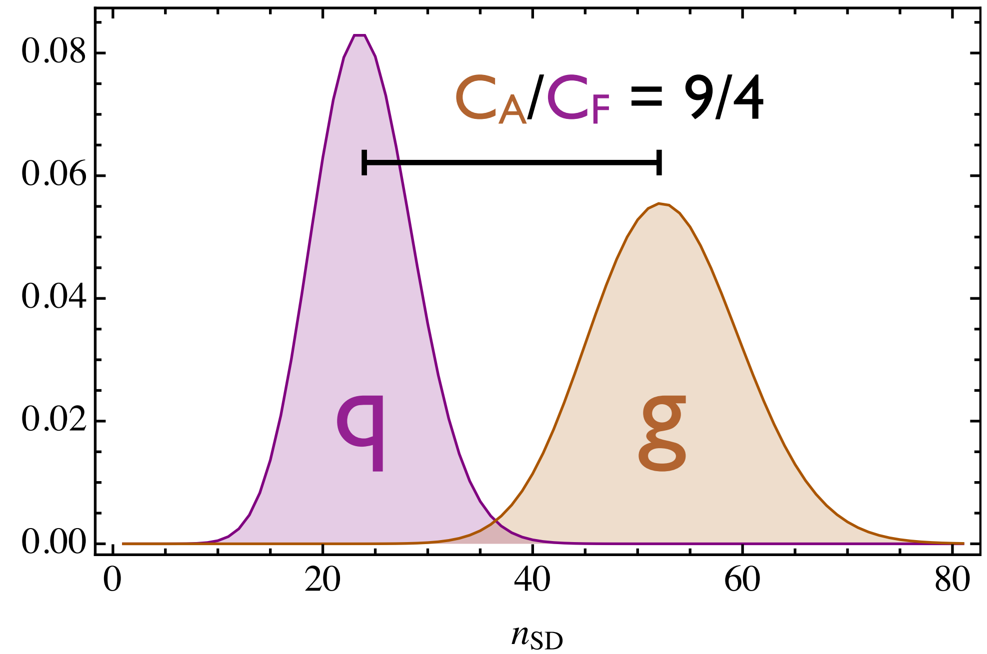
[Metodiev, JDT, 1802.00008]

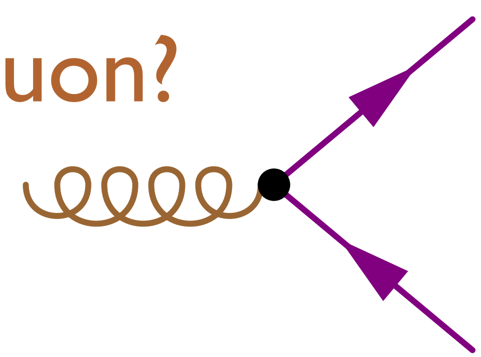
Mutual Irreducibility from QCD?

Count emissions using
"soft drop multiplicity" (IRC safe)



Asympotes to Poissonians
in high energy limit

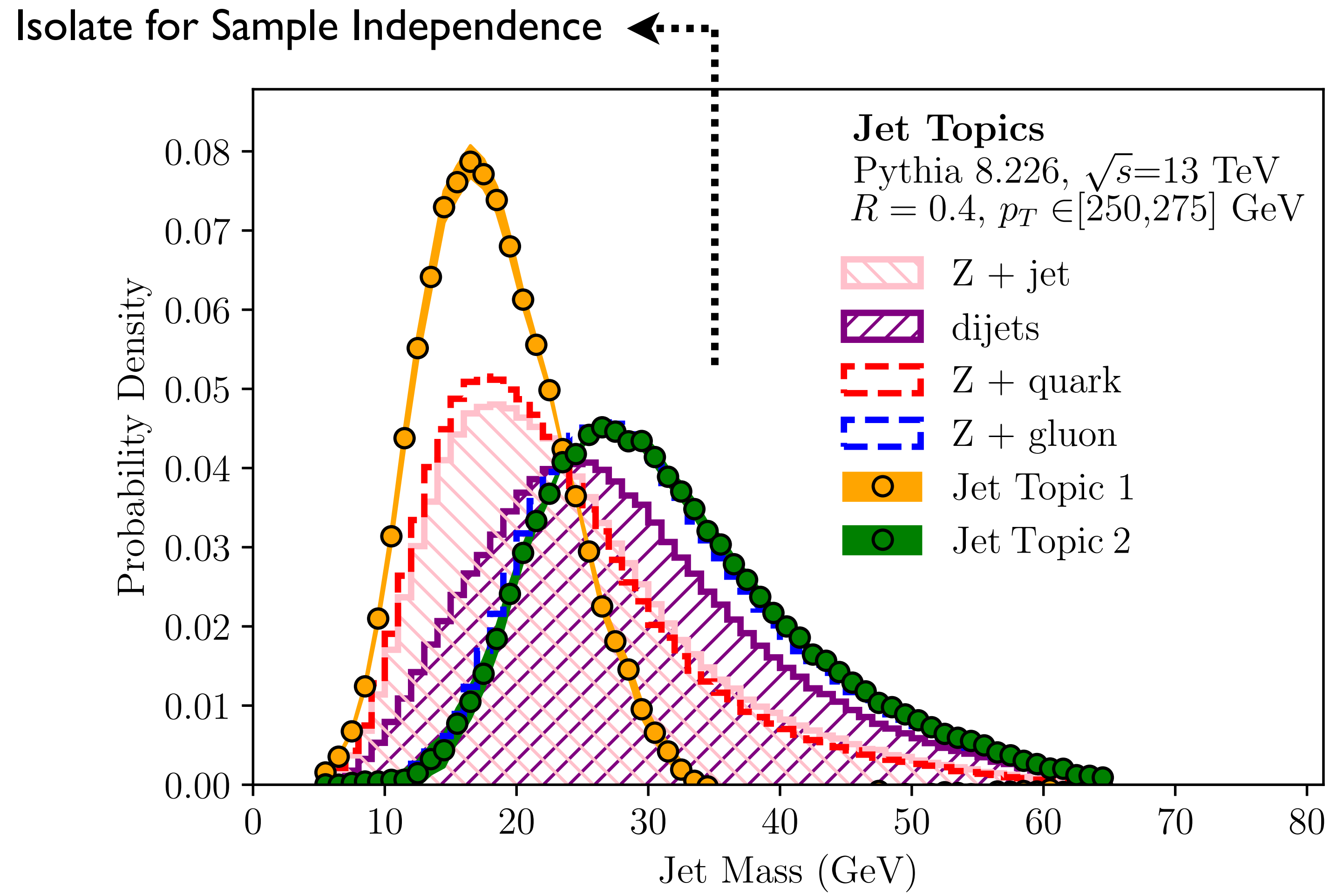




One solution:
Define "quark"/"gluon"
by mutual irreducibility

[Frye, Larkoski, Thaler, Zhou, 1704.06266]

Jet Mass is not Mutually Irreducible



Casimir Scaling at LL

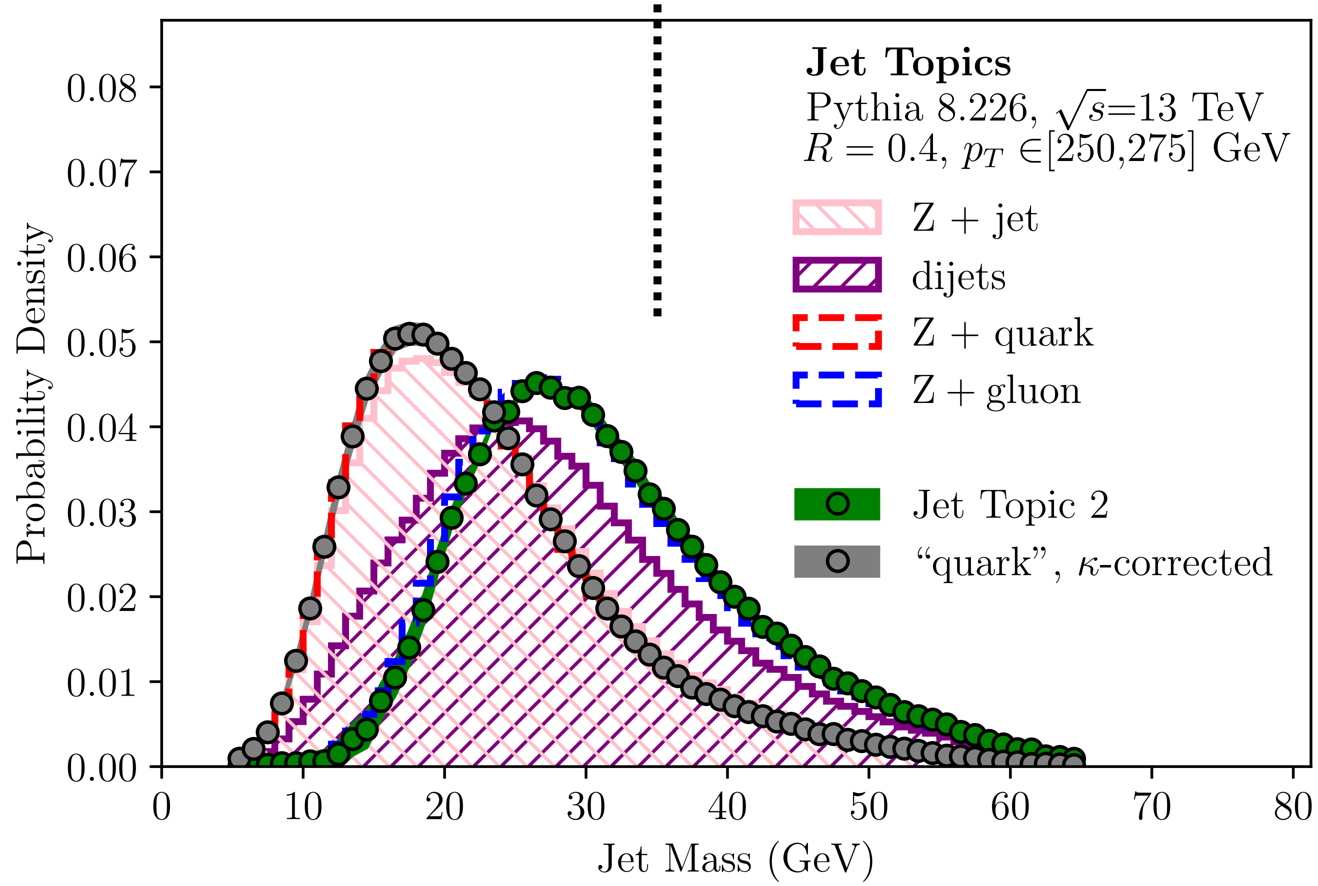
$$\kappa(g|q) = \frac{C_A}{C_F}\,\min \Sigma_q^{\frac{C_A}{C_F}-1} = 0$$ "Gluon" Topic is Pure

$$\kappa(q|g) = \frac{C_F}{C_A}\,\min \Sigma_q^{1-\frac{C_A}{C_F}} = \frac{C_F}{C_A}$$ "Quark" Topic is Distorted

[Metodiev, JDT, 1802.00008]

Jet Mass is not Mutually Irreducible

Isolate for Sample Independence



Casimir Scaling at LL

$$\kappa(g|q) = \frac{C_A}{C_F} \min \Sigma_q^{\frac{C_A}{C_F}-1} = 0$$ “Gluon” Topic is Pure

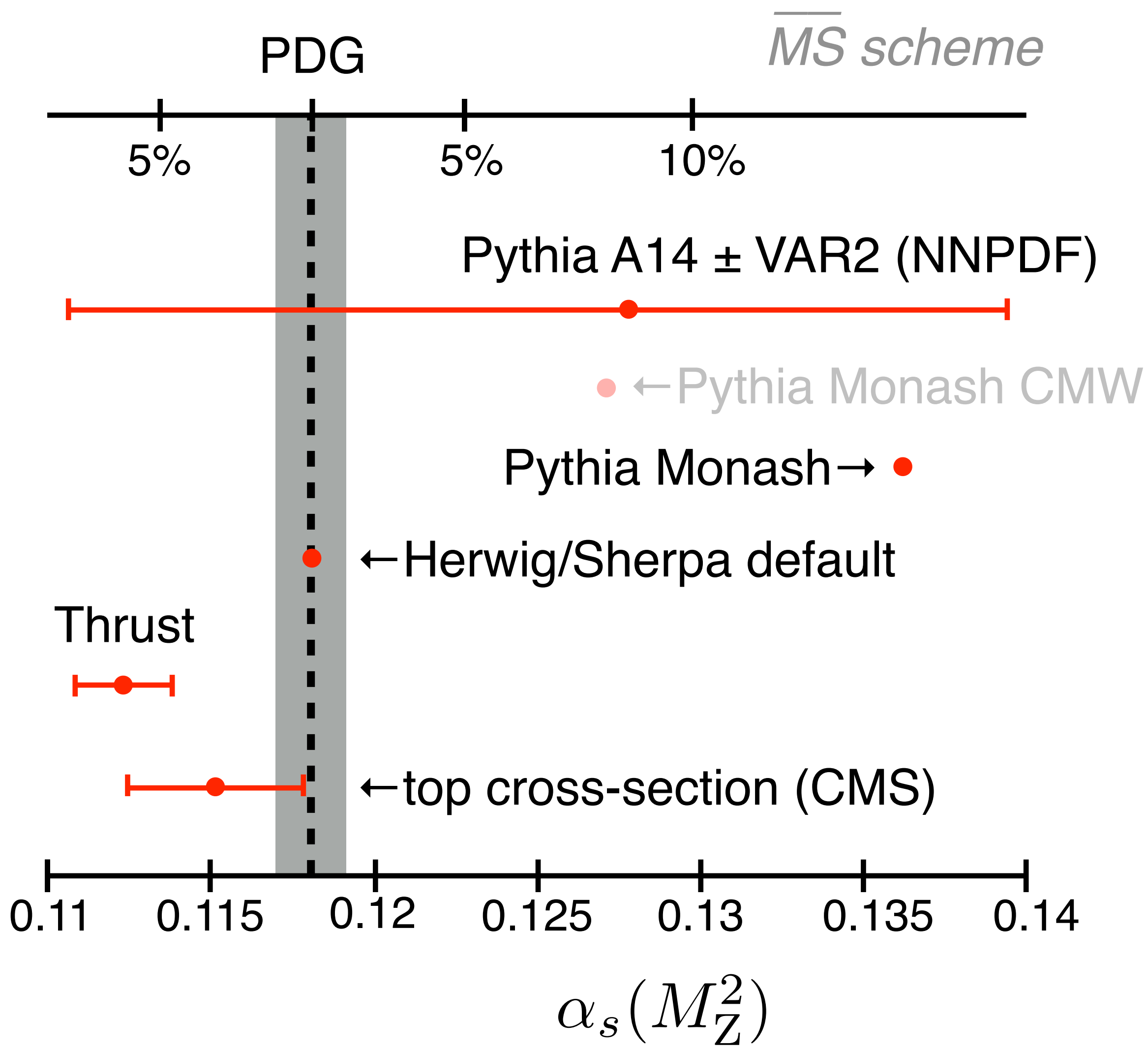
$$\kappa(q|g) = \frac{C_F}{C_A} \min \Sigma_q^{1-\frac{C_A}{C_F}} = \frac{C_F}{C_A}$$ “Quark” Topic is Distorted

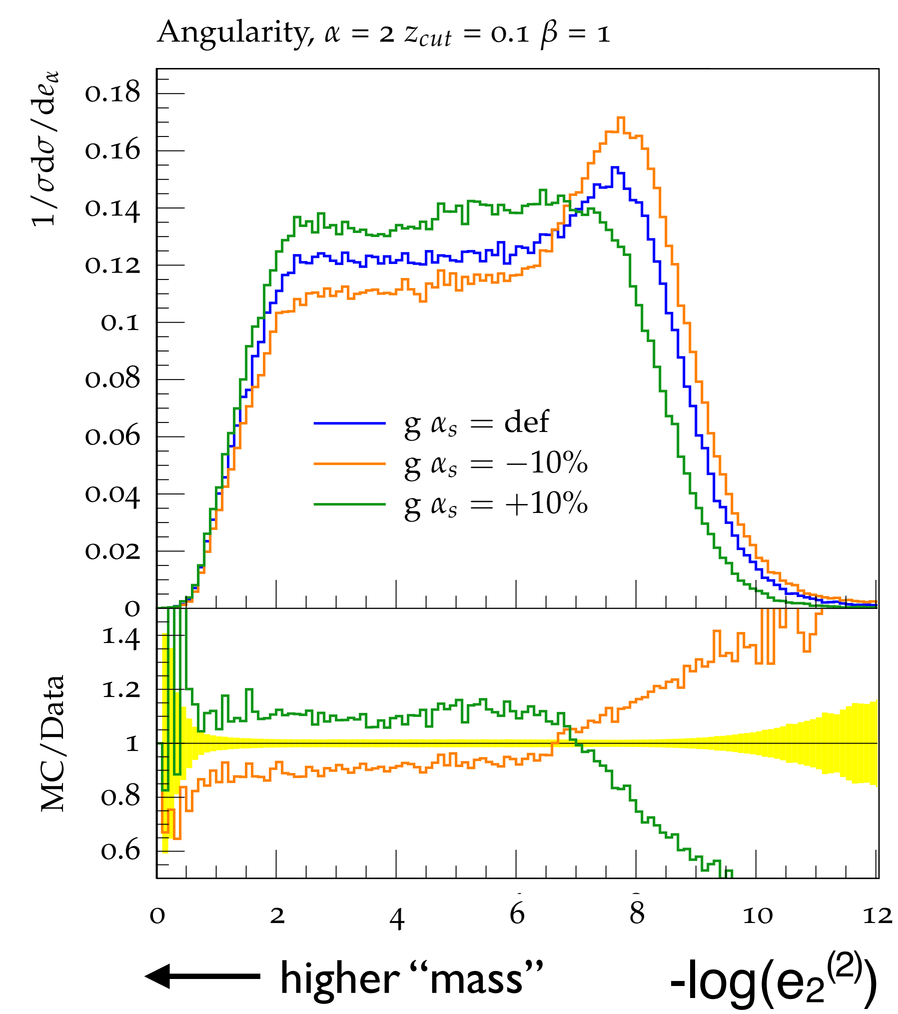
If you know κ… “Quark” Topic can be Corrected

[Metodiev, JDT, 1802.00008]

The Next Precision Frontier

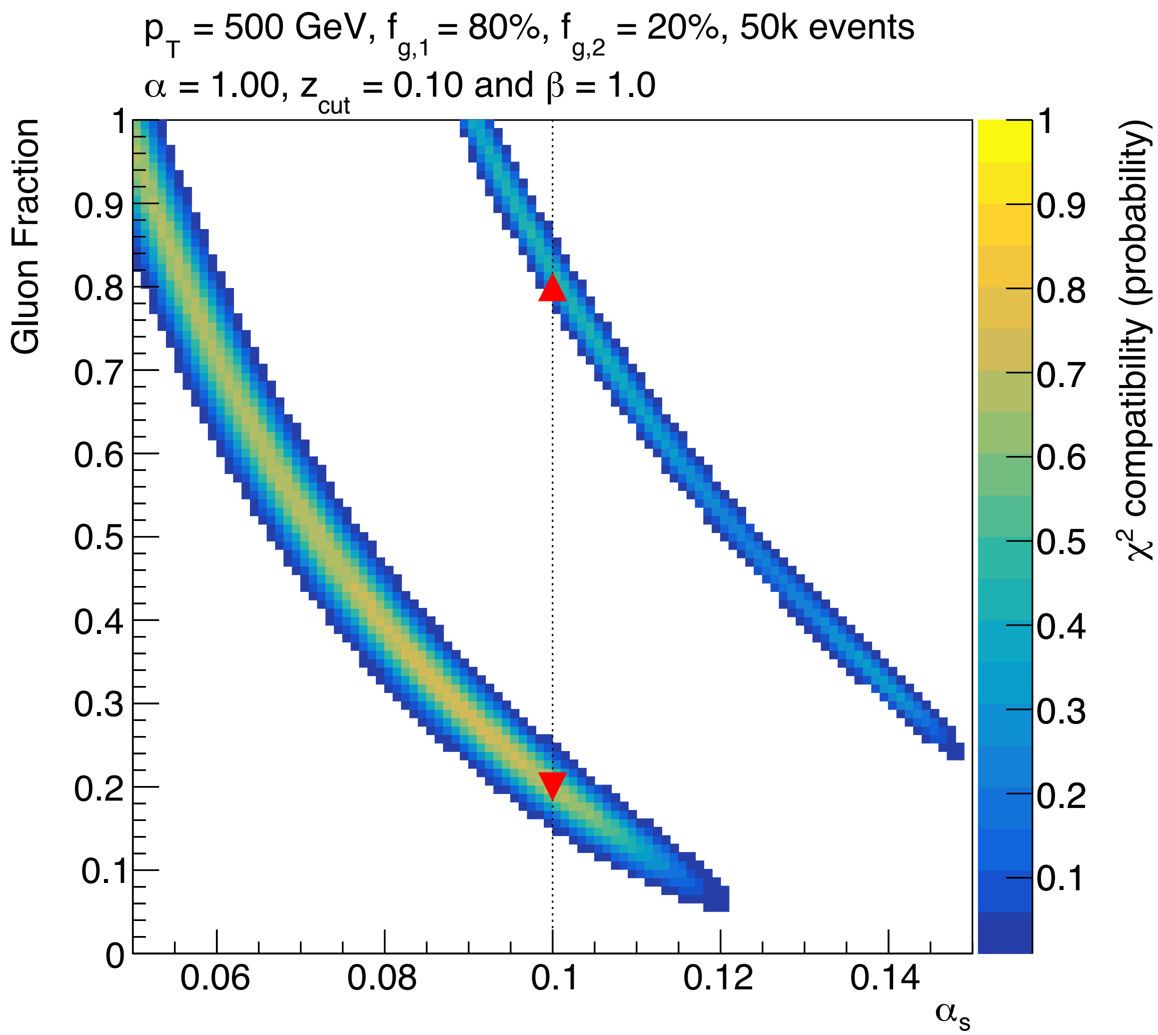
Extract Strong Coupling Constant from Jet Substructure





[see Moult, Nachman, Soyez, JDT, Chatterjee, Dreyer, Vittoria Garzelli, Gras, Larkoski, Marzani, Siódmok, Papaefstathiou, Richardson, Samui, in 1803.07977]

Key Issue for Precision Extraction



Correlation between quark/gluon fraction and α_s

$$\Sigma(\lambda) \simeq \exp\left[-\frac{\boxed{\alpha_s C_i}}{\pi}\log^2(\lambda)\right]$$

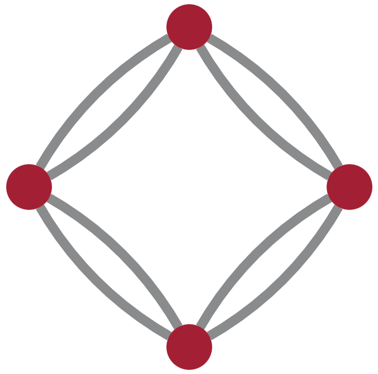
Introduces residual dependence on PDFs

By construction, jet topics are fraction independent

With or without mutual irreducibility

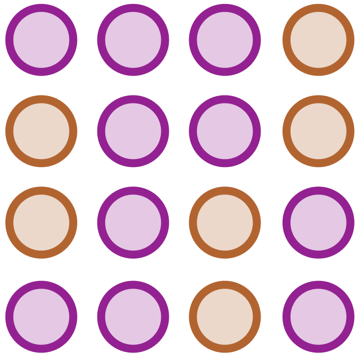
[see Moult, Nachman, Soyez, JDT, Chatterjee, Dreyer, Vittoria Garzelli, Gras, Larkoski, Marzani, Siódmok, Papaefstathiou, Richardson, Samui, in 1803.07977]

Summary



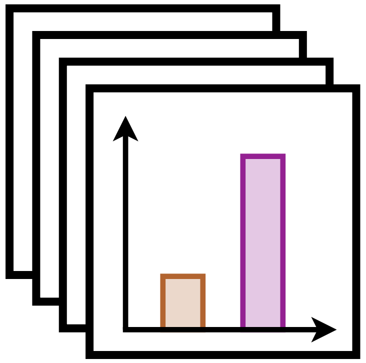
A Basis for Jet Substructure

Energy flow polynomials for linear classification



Learning Without Labels

Data-driven classifiers from mixed samples



Introducing Jet Topics

Defining jet categories by mutual irreducibility

&

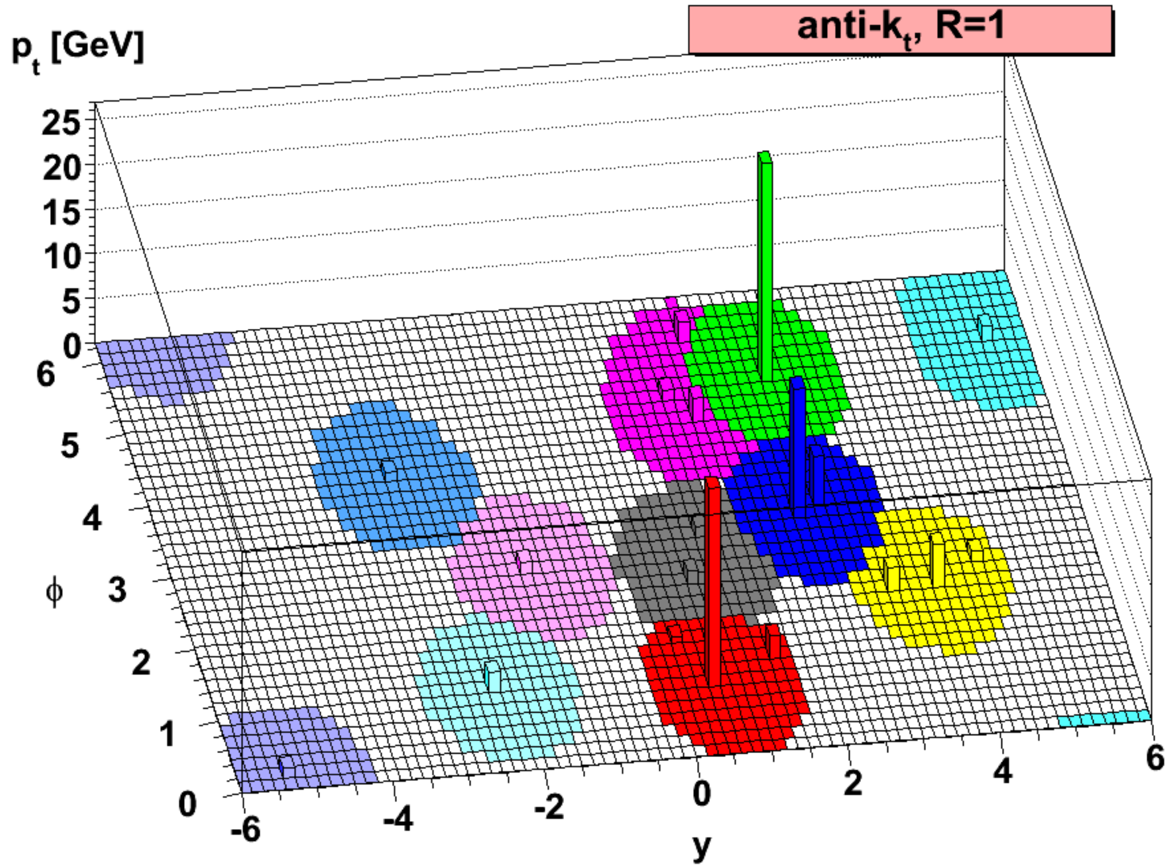
"Deep Learning" ~~vs.~~ "Deep Thinking"

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

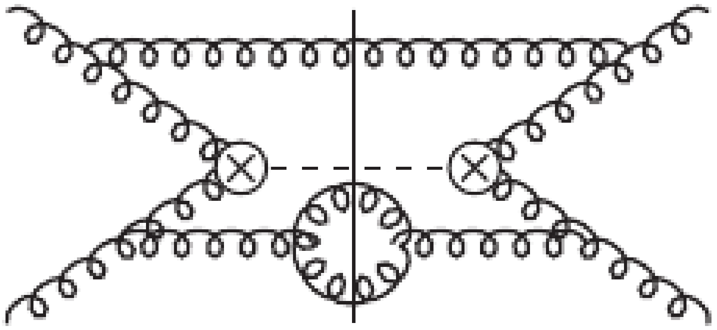
Backup Slides

A QCD Renaissance

Theory c. 2008–present



New Jet Algorithms



Loop/Leg/Log Explosion

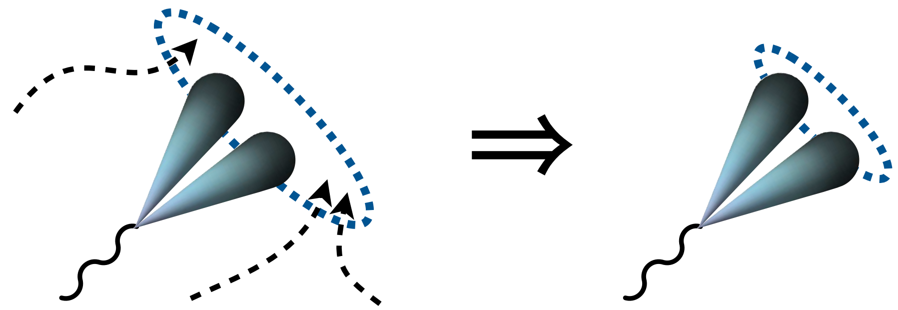
Jet Substructure

[Anti-k$_T$: Cacciari, Salam, Soyez, 2008; see also Delsart, 2006] [N^{3}LO: Anastasiou, Duhr, Dulat, Herzog, Mistlberger, 2015]
[BDRS: Butterworth, Davison, Rubin, Salam, 2008; see also Seymour, 1991, 1994]

The Substructure Toolbox

Grooming:

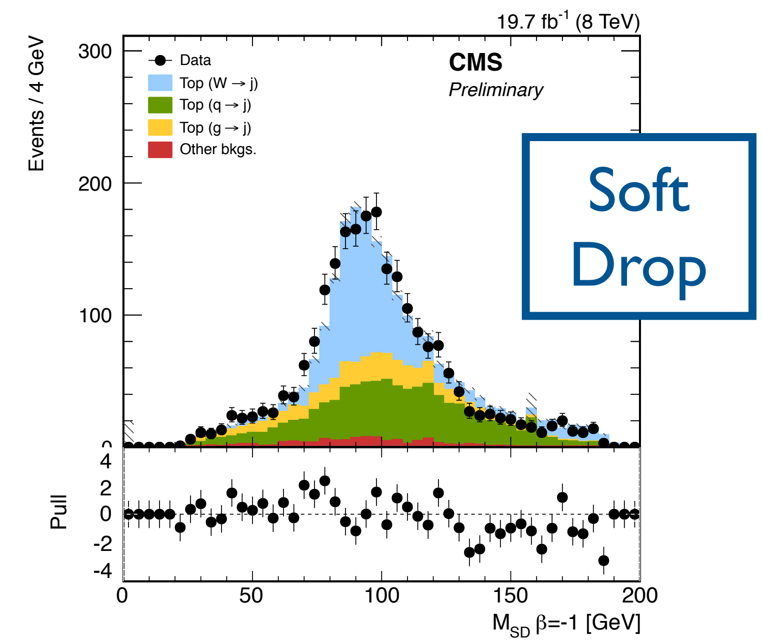
e.g. ISR/UE/pileup



[Mass Drop/Filtering, Trimming, Pruning, Soft Drop, Jet Reclustering…; for pileup: Area Subtraction, Jet Cleansing, SoftKiller, PUPPI, Constituent Subtraction, PUMML…]

W/Z-Tagging @ CMS

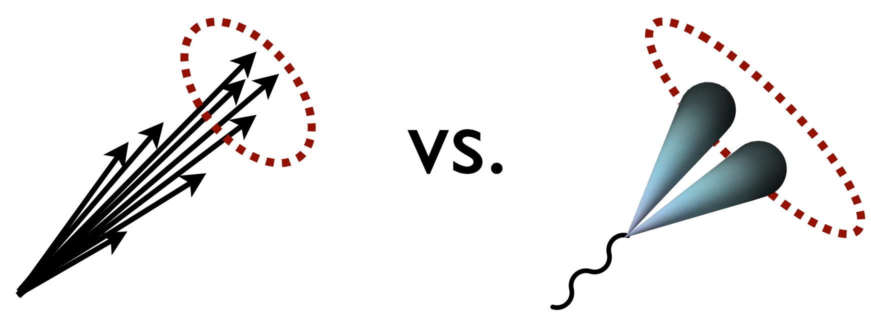
[JME-14-002, CMS-PAS-EXO-15-002]



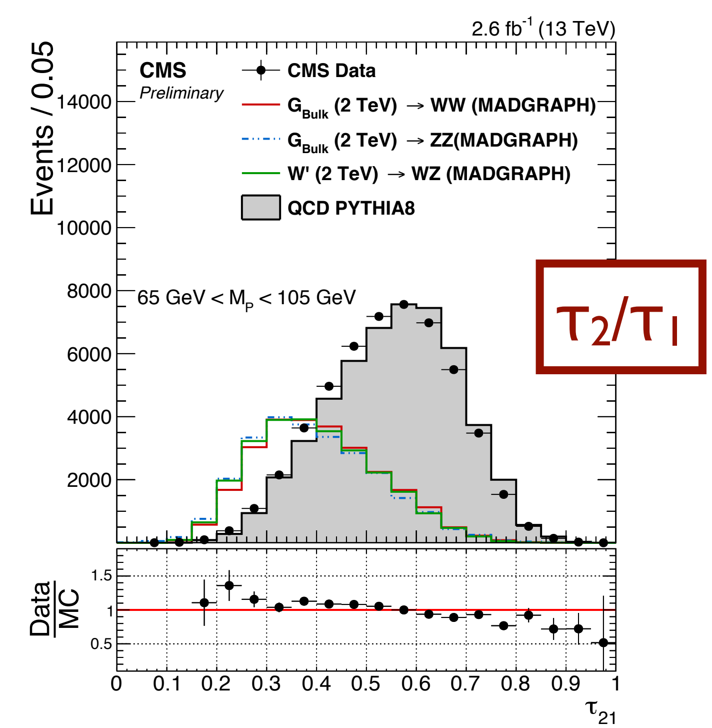
[using Larkoski, Marzani, Soyez, JDT, 1402.2657]

Discrimination:

e.g. 1-prong vs. N-prong



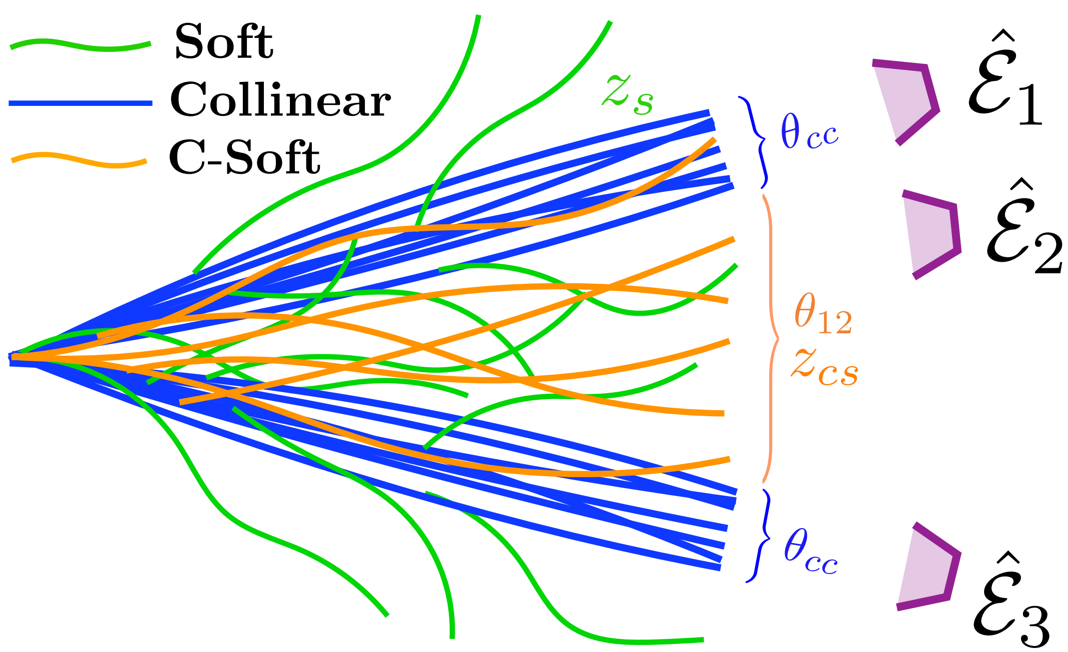
[p_T Balance, Y-splitter, Angularities, Planar Flow, N-subjettiness, Angular Structure Functions, Jet Charge, Jet Pull, Energy Correlation Functions, Dipolarity, p_T^D, Zernike Coefficients, LHA, Fox-Wolfram Moments, JHU/CMSTopTagger, HEPTopTagger, Template Method, Shower Deconstruction, Subjet Counting, Wavelets, Q-Jets, Telescoping Jets, Deep Learning…]



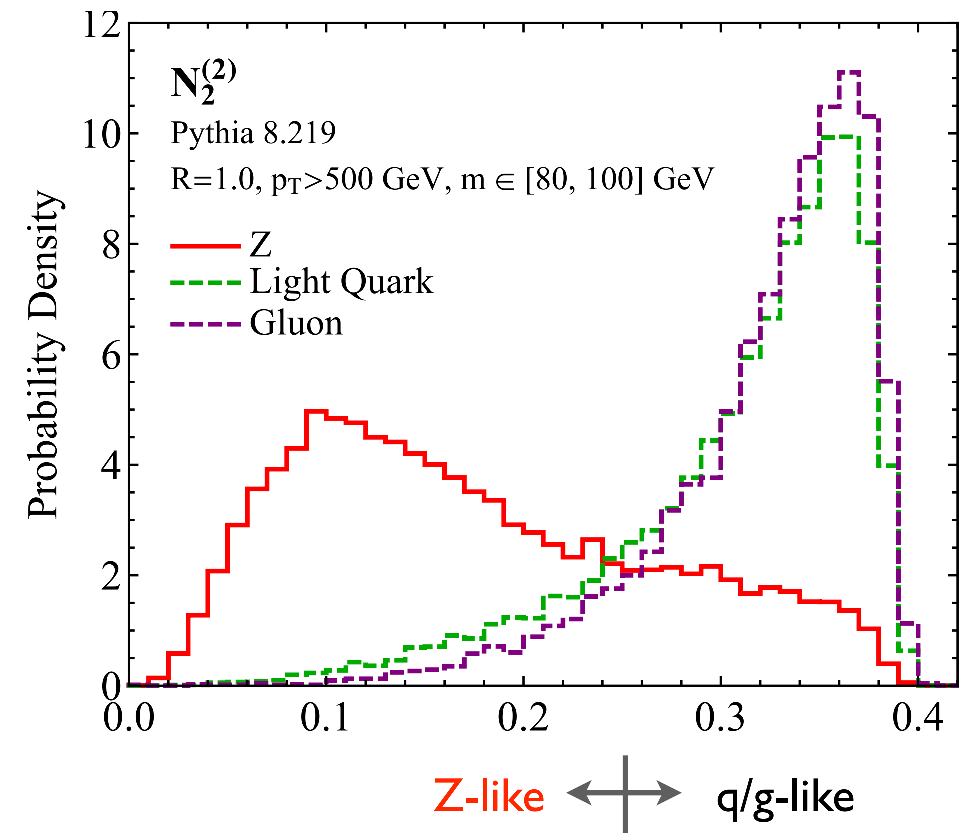
[using JDT, Van Tilburg, 1011.2268, 1108.2701]

2-prong Discrimination with Energy Correlators

$$\boxed{N_2} = \frac{\sum_{i<j<k} p_{Ti}\, p_{Tj}\, p_{Tk} \min\left\{(R_{ij}R_{jk})^2, (R_{jk}R_{ki})^2, (R_{ki}R_{ij})^2\right\}}{\left(\sum_{i<j} p_{Ti}\, p_{Tj}\, R_{ij}^2\right)^2 \Big/ \sum_i p_{Ti}}$$

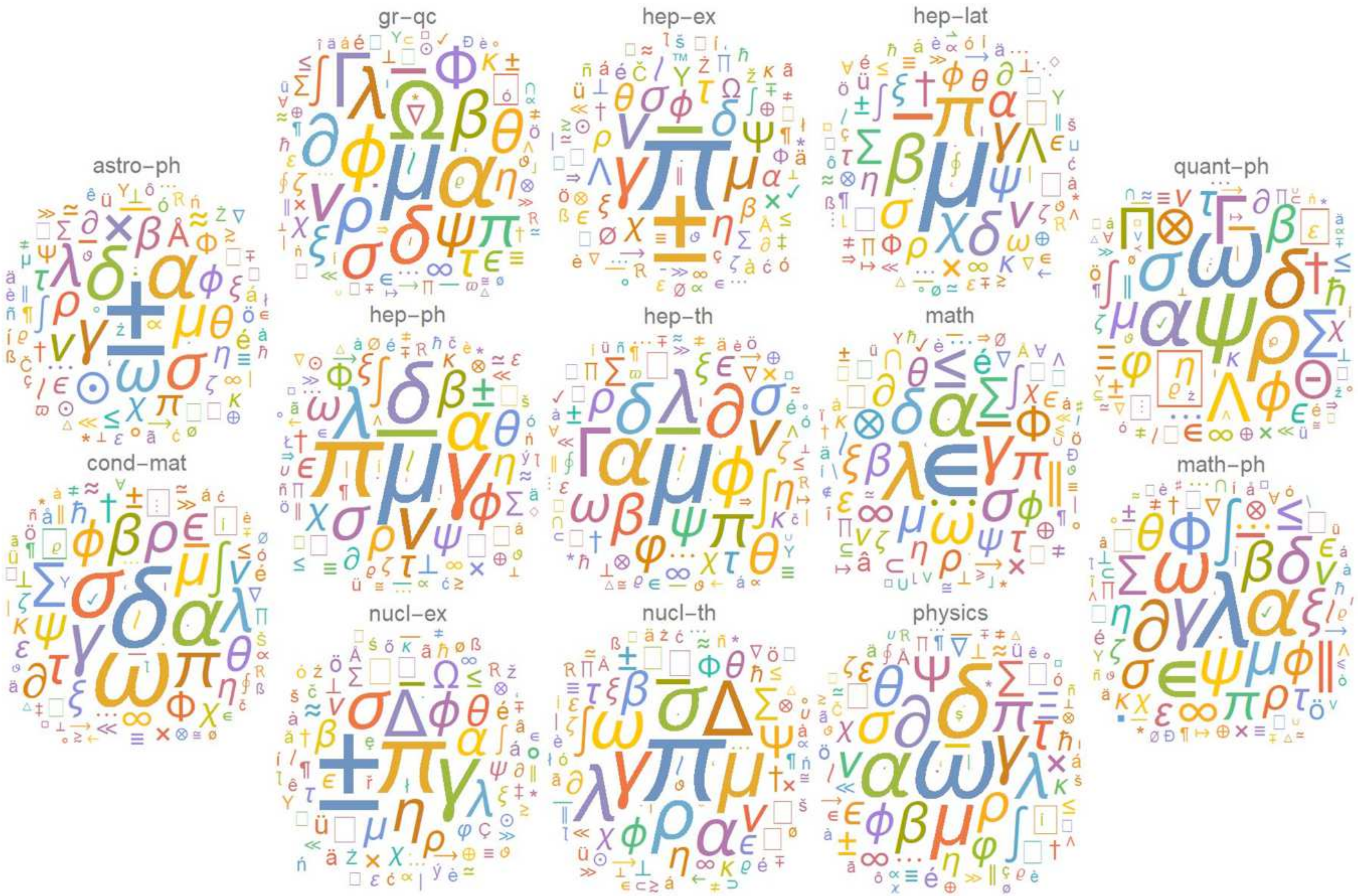


$\Rightarrow$



[Moult, Necib, JDT, 1609.07483; based on Larkoski, Salam, JDT, 1305.0007]

Frequency of Symbols on the arXiv



arXiv 2.0: Determine categories just from documents?
(Without training from hep-ph, hep-ex, etc.)

[Wolfram Summer School, 2017]