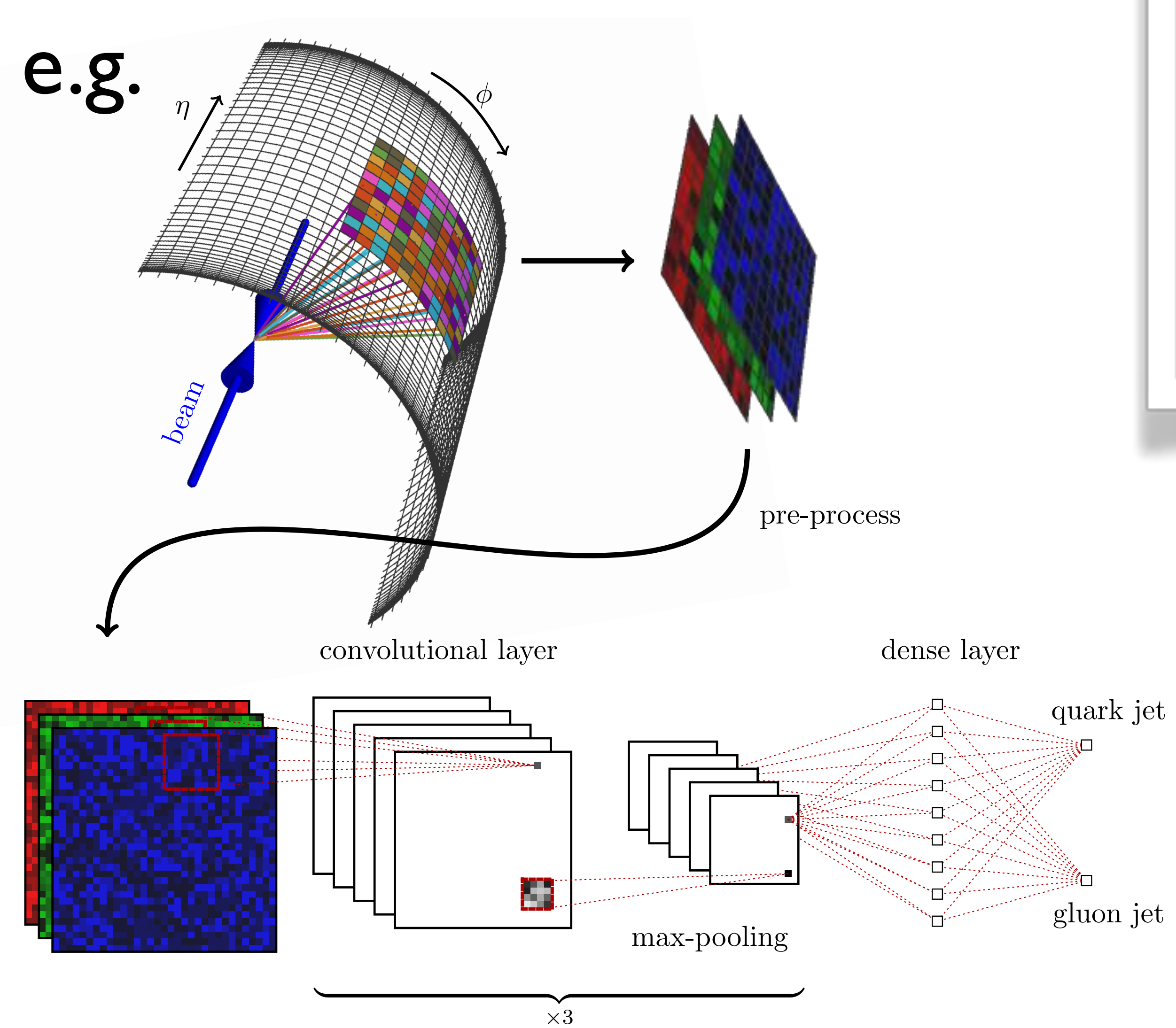
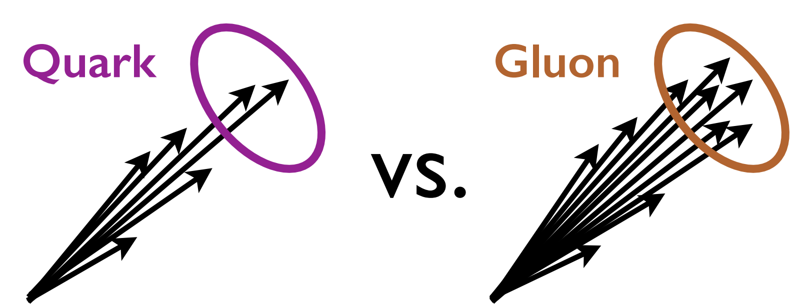
A Theorist's Perspective on Machine Learning for Jets

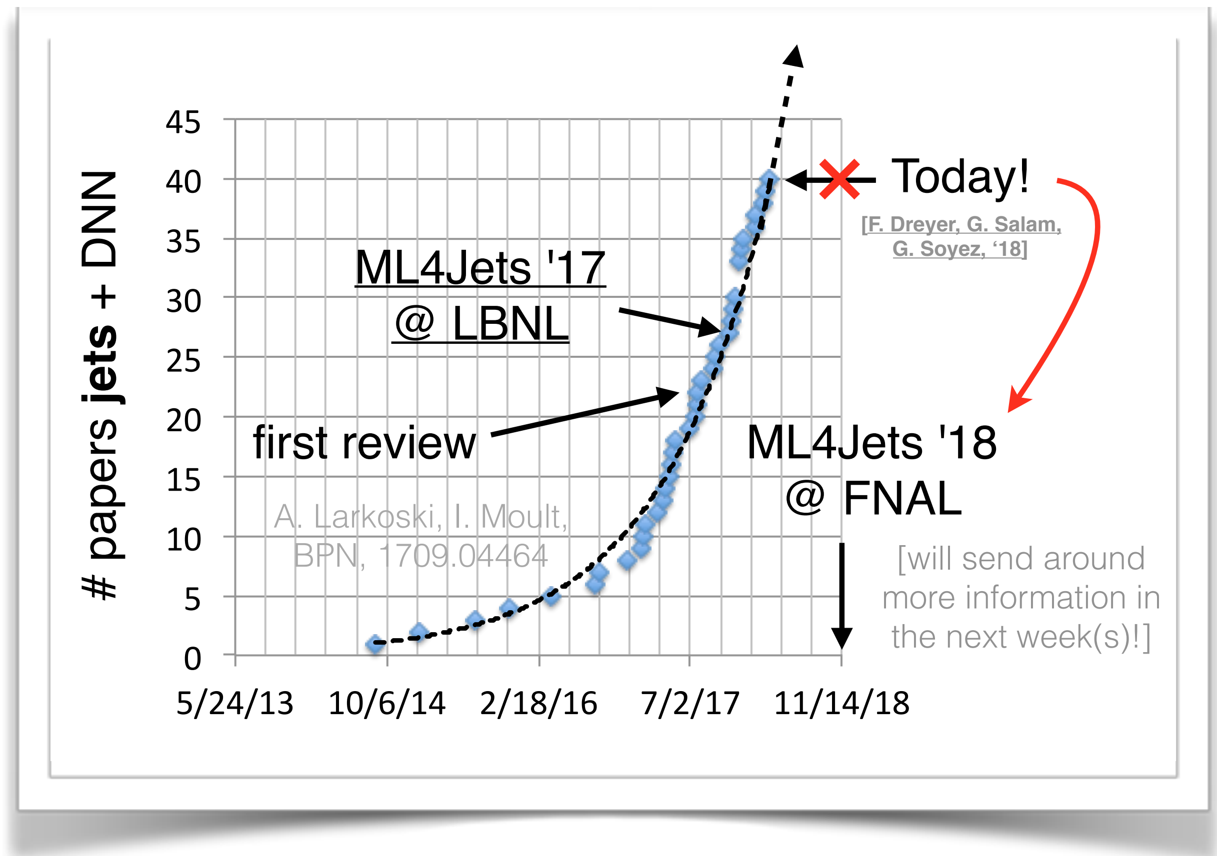
Jesse Thaler

Machine Learning for Jet Physics, Fermilab — November 14, 2018

The Rise of Machine Learning for Jets







[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]

My Perspective c. 2016

"Deep Learning"

My Perspective c. 2016

“Deep Learning” vs. “Deep Thinking”

My Perspective c. 2018

BOOST 2018

10th International Workshop on Boosted Objects
Phenomenology, Reconstruction and Searches

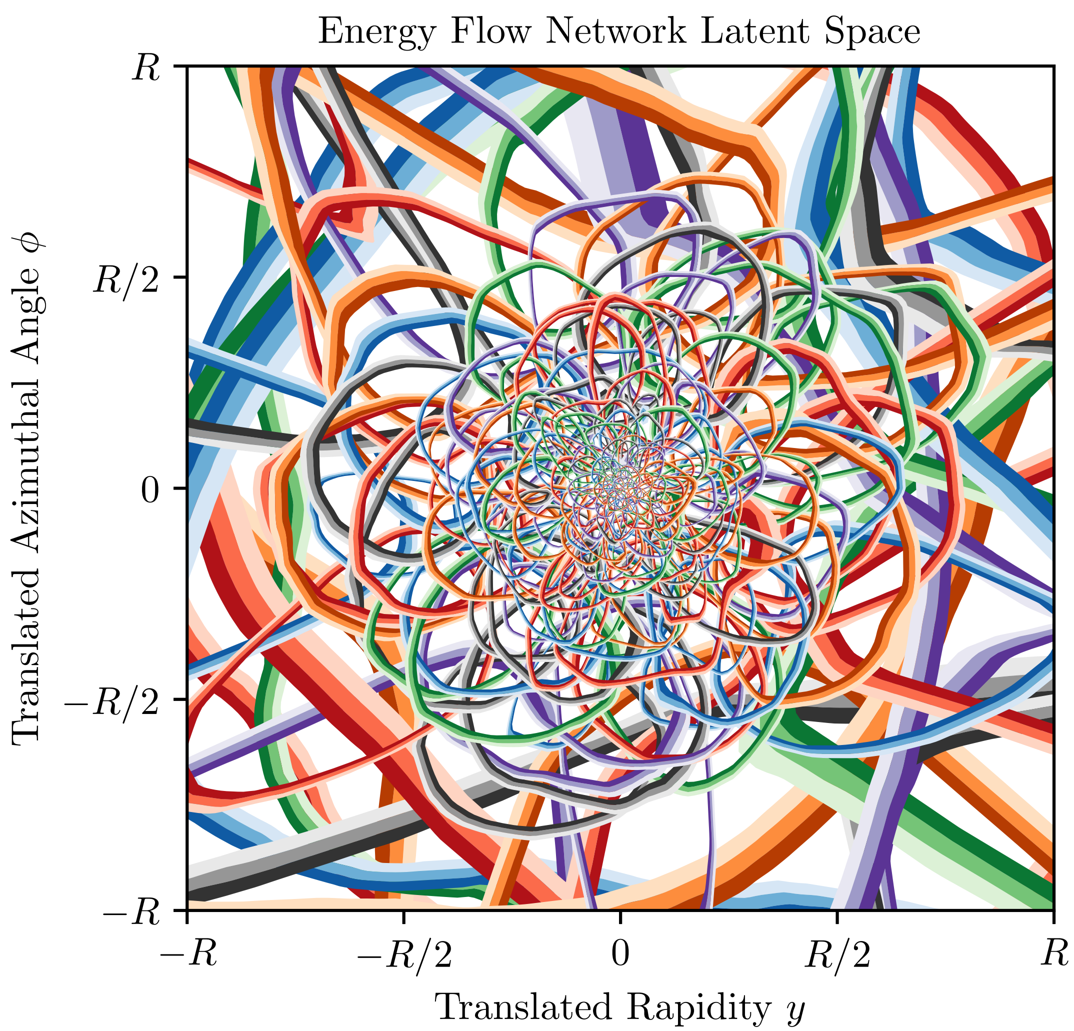
"Deep Learning" ~~vs.~~ & "Deep Thinking"

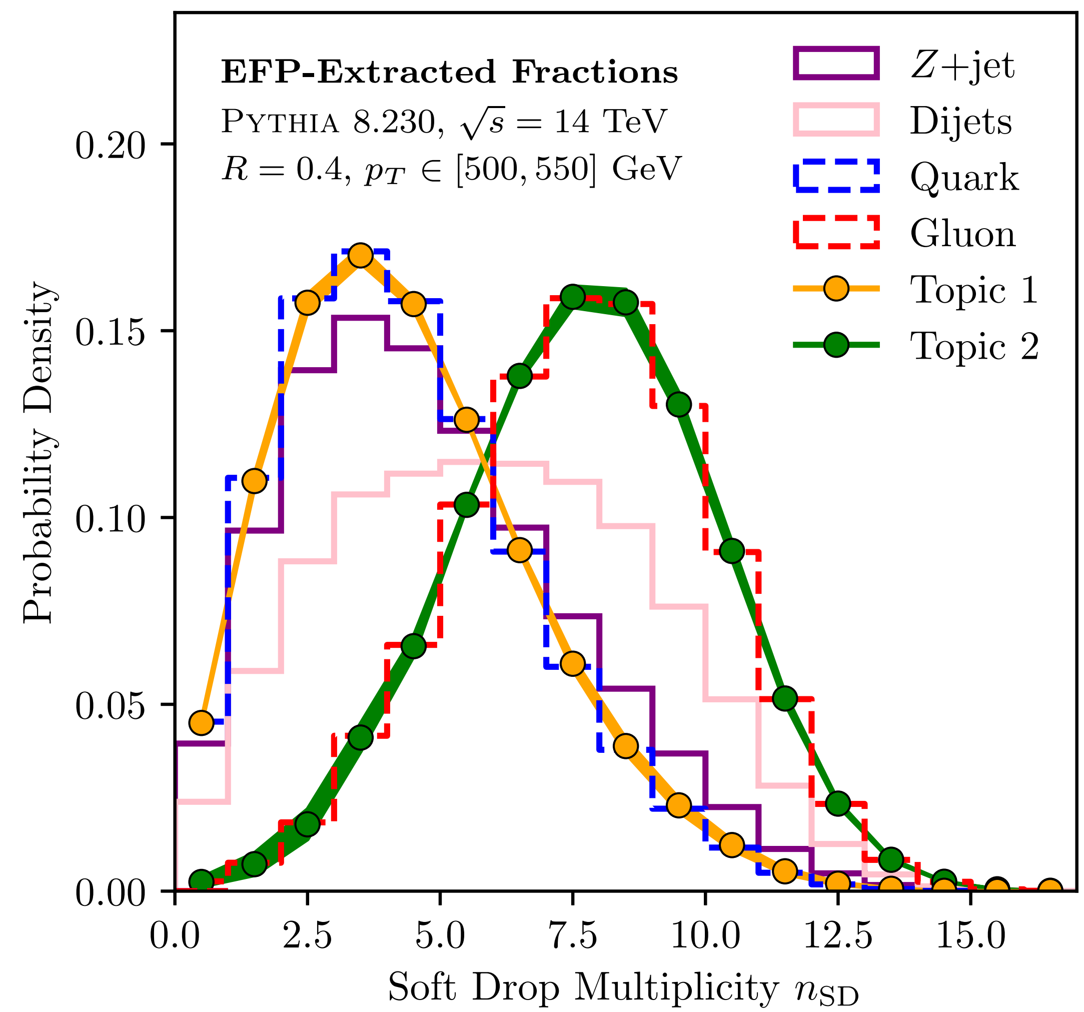
New first-principles studies in collider physics
facilitated by advances in
mathematics, statistics, and computer science

Proximate Reasons for My Conversion

Patrick Komiske
Thursday, 9:30am

Eric Metodiev
Friday, 10:20am



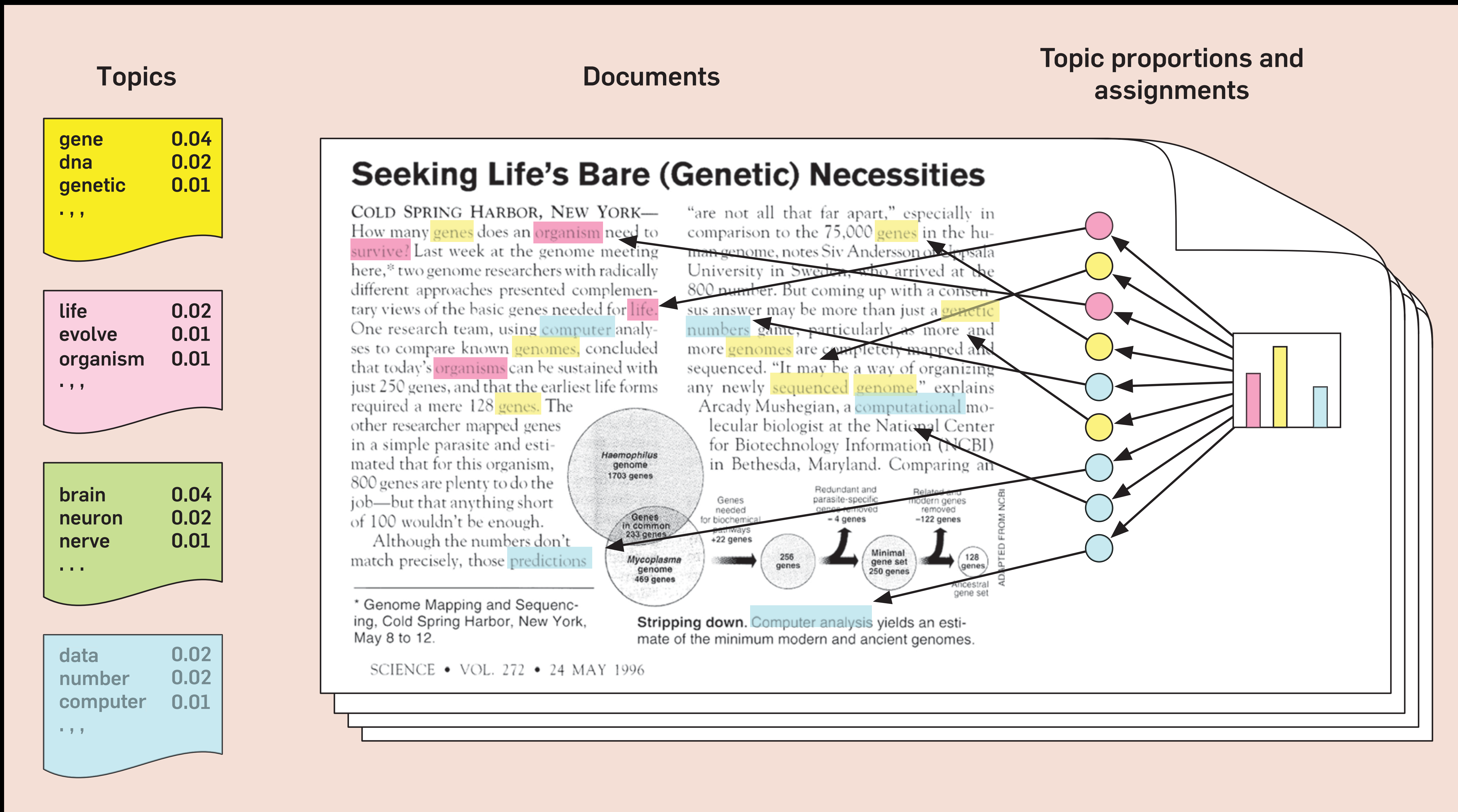


plus Ben Nachman, Kyle Cranmer, Daniel Whiteson, Mike Williams, Matt Schwartz, Dan Roberts, Phiala Shanahan, …

Point Clouds

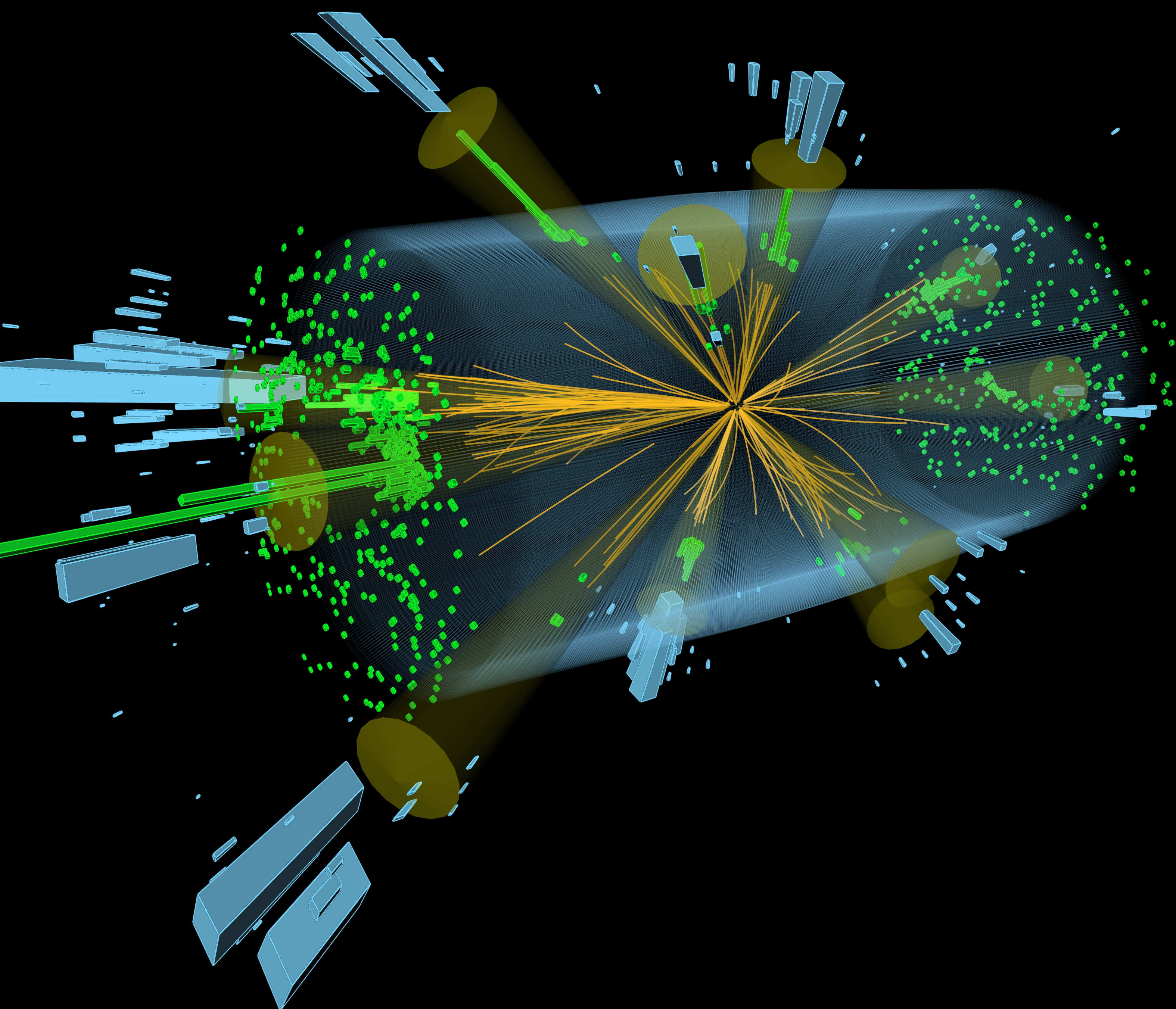
[Popular Science, 2013]

Topic Modeling



[Blei, 2012]

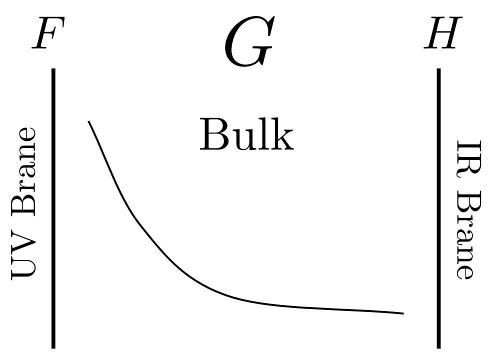
Debris Taxonomy



T	E	H	M			
	●			γ	photon	elementary
●	●			$e^{\pm}$	electron	elementary
●	•	•	●	$\mu^{\pm}$	muon	elementary
●	•	●		$\pi^{\pm}$	pion	composite
●	•	●		$K^{\pm}$	kaon	composite
	•	●		K_L^0	K-long	composite
●	•	●		$p/\overline{p}$	proton	composite
	•	●		$n/\overline{n}$	neutron	composite

Evolution of a "Model Builder"

$\boxed{\mathcal{L}}$

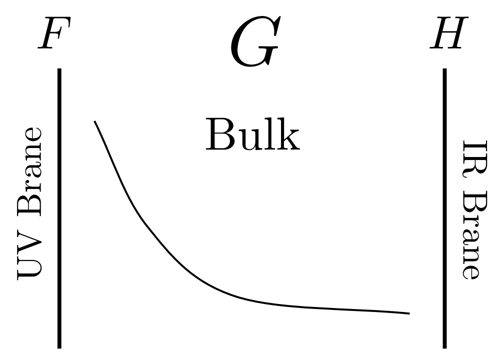


Build Lagrangians & Paradigms

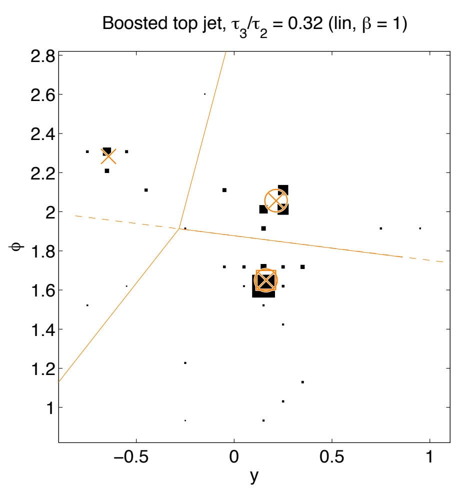
[images from JDT, hep-ph/0502175; JDT, Van Tilburg, 1108.2701; Komiske, Metodiev, JDT, 1810.05165]

Evolution of a "Model Builder"

$\boxed{\mathcal{L}}$ $\boxed{\mathcal{O}}$



Build Lagrangians & Paradigms

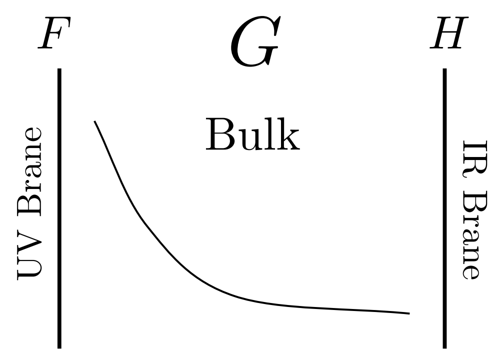


Build Observables & Algorithms

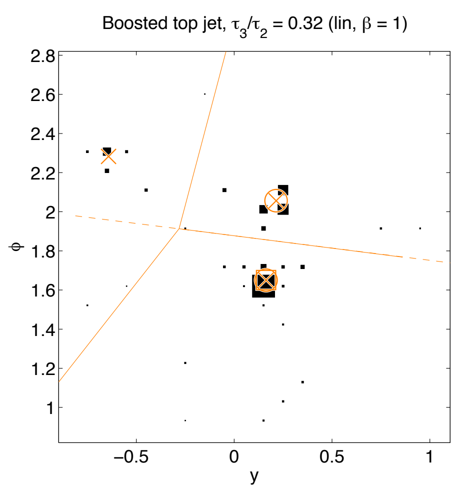
[images from JDT, hep-ph/0502175; JDT, Van Tilburg, 1108.2701; Komiske, Metodiev, JDT, 1810.05165]

Evolution of a "Model Builder"

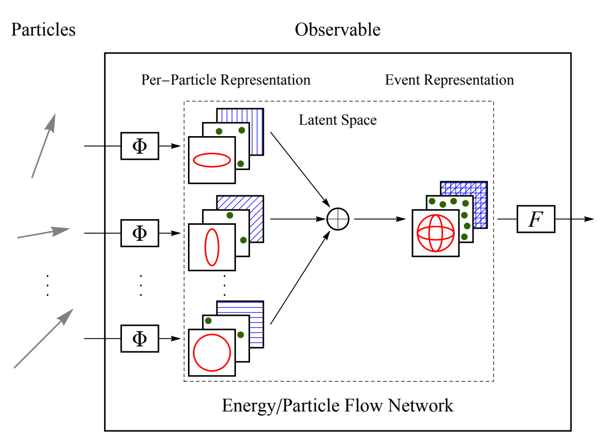
$\mathcal{L}$ $\mathcal{O}$ $\mathcal{M}$



Build Lagrangians & Paradigms



Build Observables & Algorithms



Build Models & Loss Functions

[images from JDT, hep-ph/0502175; JDT, Van Tilburg, 1108.2701; Komiske, Metodiev, JDT, 1810.05165]

Scrutinizing the Theory Toolbox

$\boxed{\mathcal{L}}$ $\boxed{\mathcal{O}}$ $\boxed{\mathcal{M}}$

Given current status of the LHC, which strategy:

Makes maximal (verifiable) use of collider data?

Can scale up to the challenges of HL-LHC?

Offers new insights into fundamental physics?

My view: Machine learning has enormous potential,
but onus is on ML4Jets community to answer these questions

My Goals for this Workshop

$\mathcal{L}$ $\mathcal{O}$ $\mathcal{M}$

Learn about state-of-the-art ML and novel HEP applications (esp. beyond fully-supervised classification)

Press issue of verifiability (i.e. robustness, transparency, calibration, unfolding, visualization)

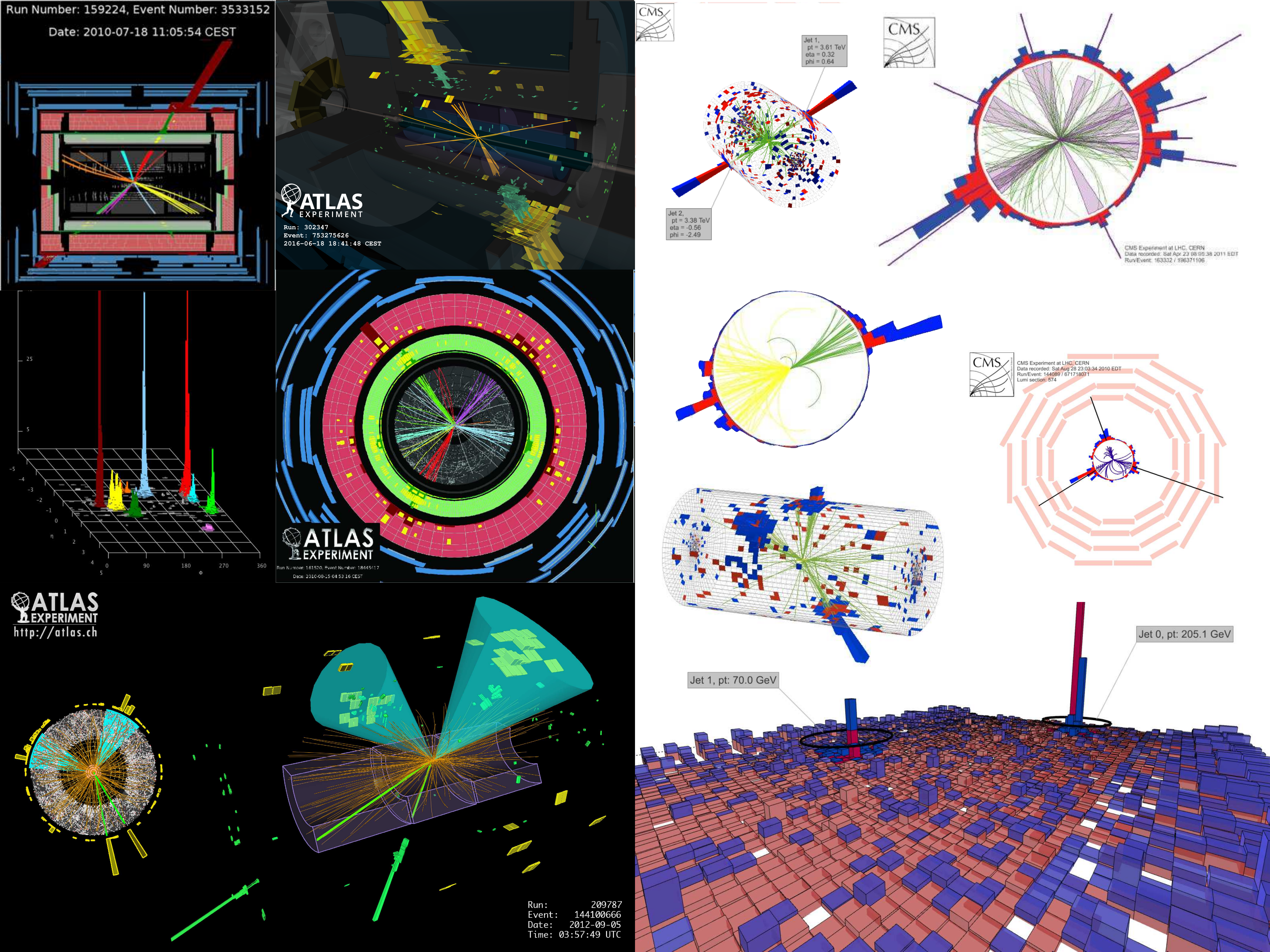
Explore possibilities for factorization (i.e. versatility, component reuse, symmetries, transfer learning)

Seek opportunities to push boundaries of fundamental physics

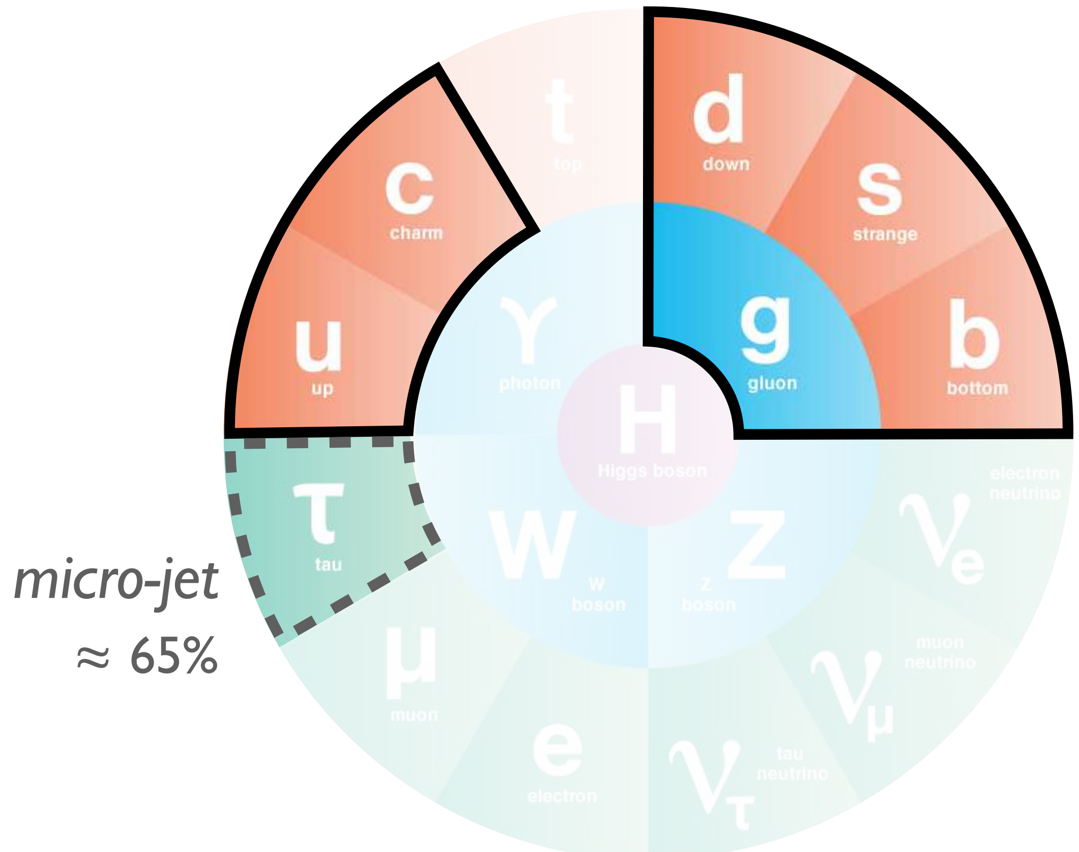
And have fun! ML4Jets is a young, vibrant, and creative subfield

A Brief Introduction to Machine Learning and Jet Classification

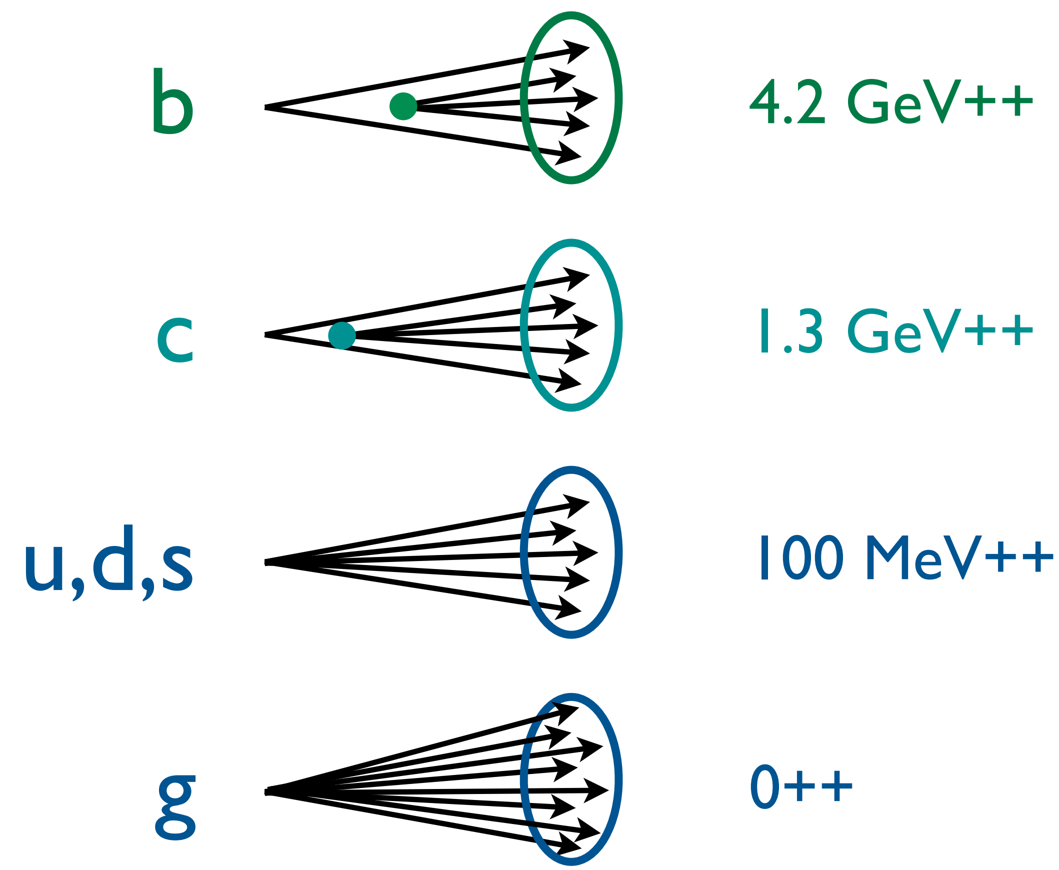
(For the benefit of any newcomers)



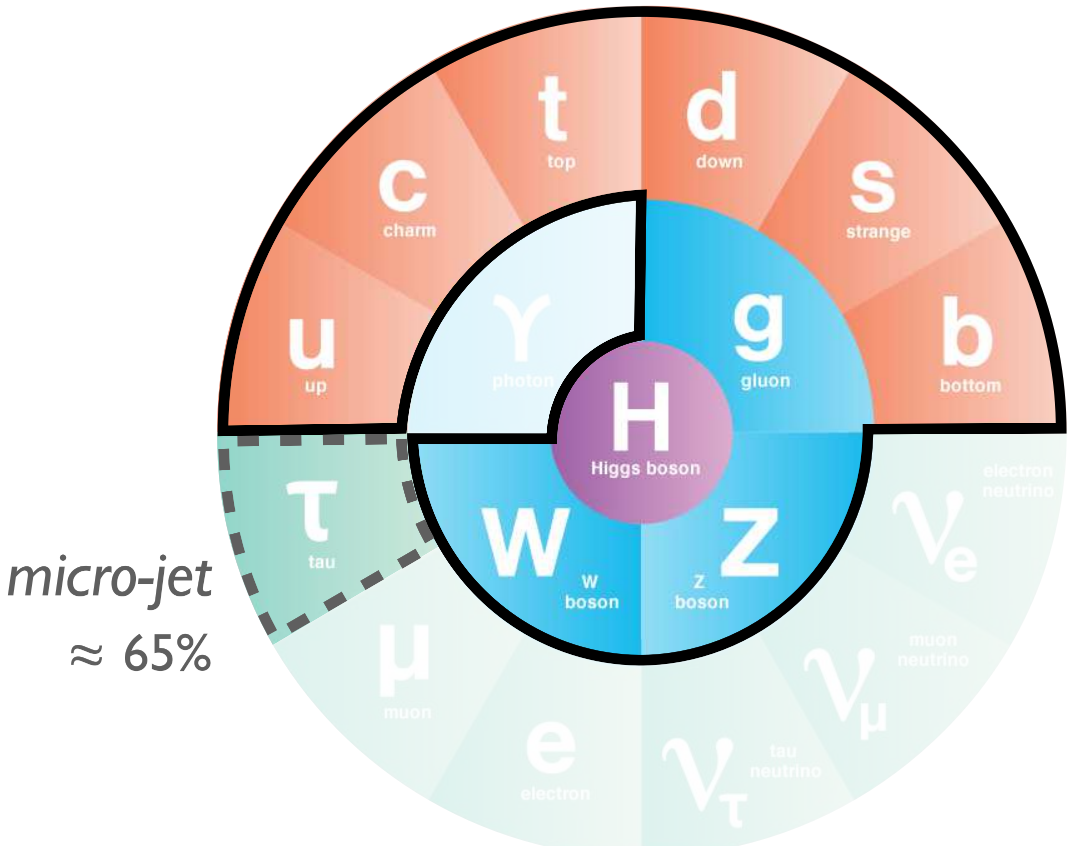
Run Number: 159224, Event Number: 3533152
Date: 2010-07-18 11:05:54 CEST
ATLAS EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST
CMS
Jet 1, pt = 3.61 TeV eta = 0.32 phi = 0.64
Jet 2, pt = 3.38 TeV eta = -0.56 phi = -2.49
CMS Experiment at LHC, CERN
Data recorded: Sat Apr 23 08:05:38 2011 EDT
Run/Event: 163332 / 196371106
ATLAS EXPERIMENT
Run Number: 161520, Event Number: 18445417
Date: 2010-08-15 04:53:16 CEST
CMS Experiment at LHC, CERN
Data recorded: Sat Aug 28 23:03:34 2010 EDT
Run/Event: 144089 / 671718071
Lumi section: 574
ATLAS EXPERIMENT
http://atlas.ch
Run: 209787
Event: 144100666
Date: 2012-09-05
Time: 03:57:49 UTC
Jet 0, pt: 205.1 GeV
Jet 1, pt: 70.0 GeV



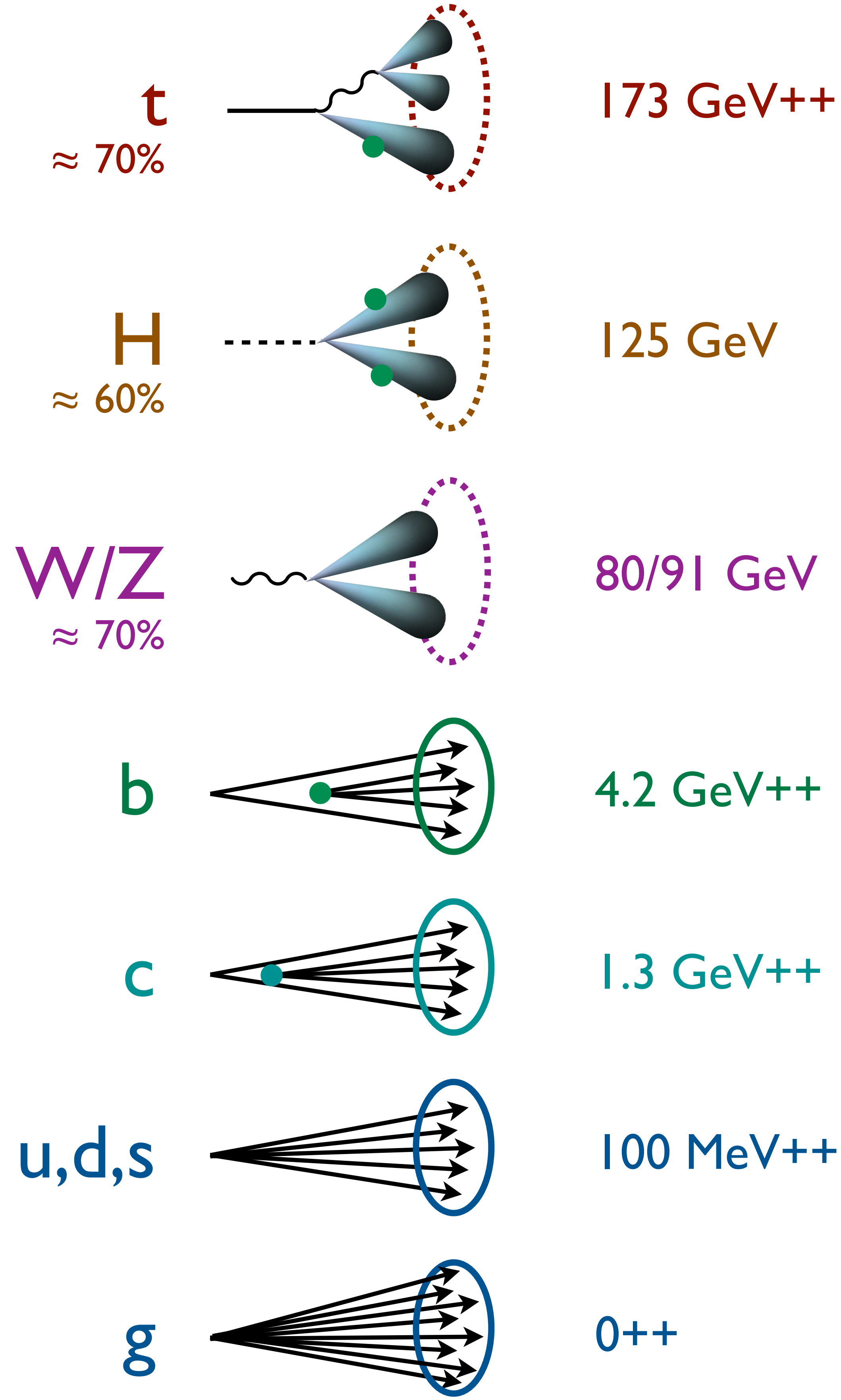
Jets from the Standard Model



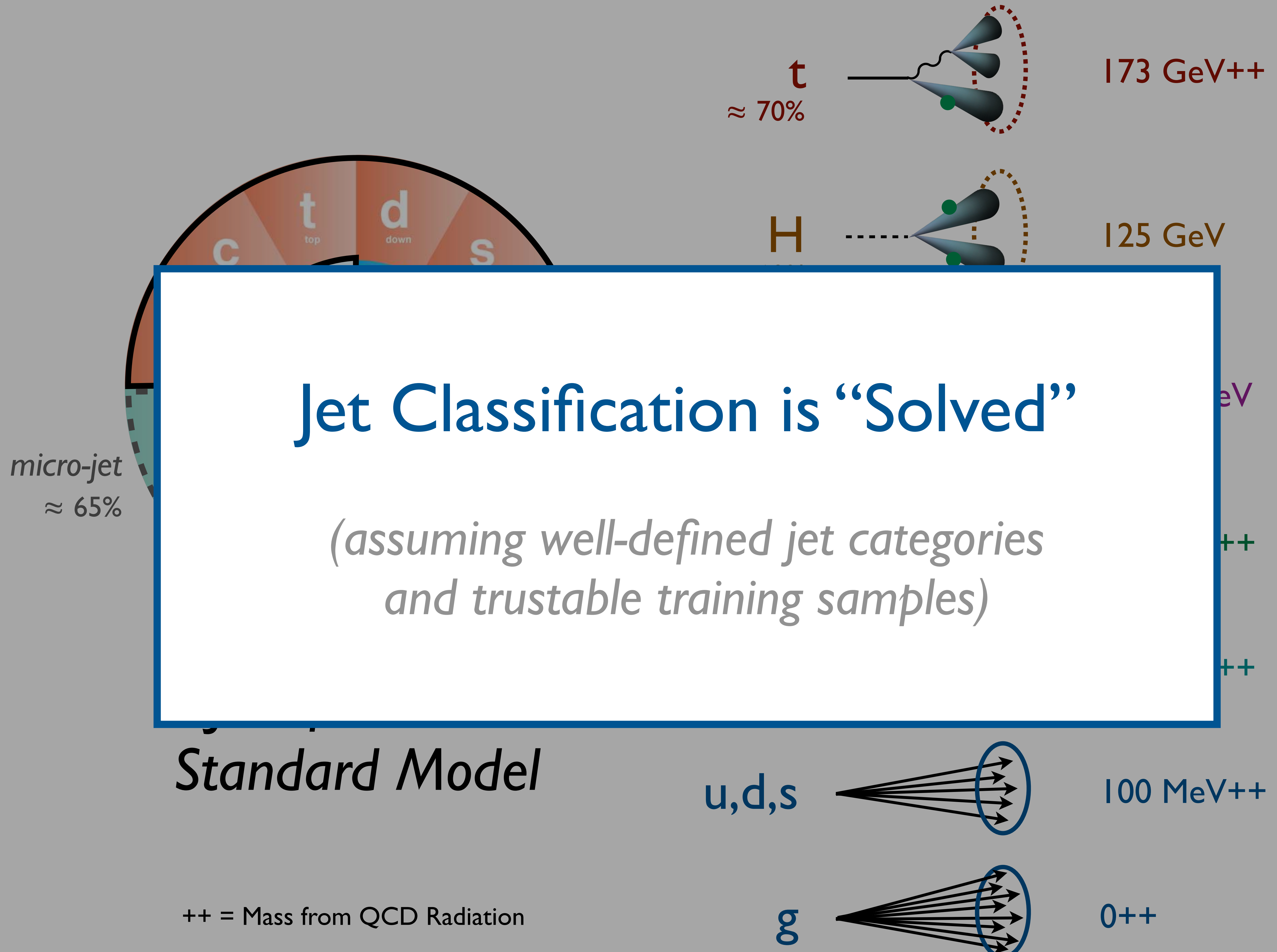
++ = Mass from QCD Radiation



Jets from the Standard Model



++ = Mass from QCD Radiation



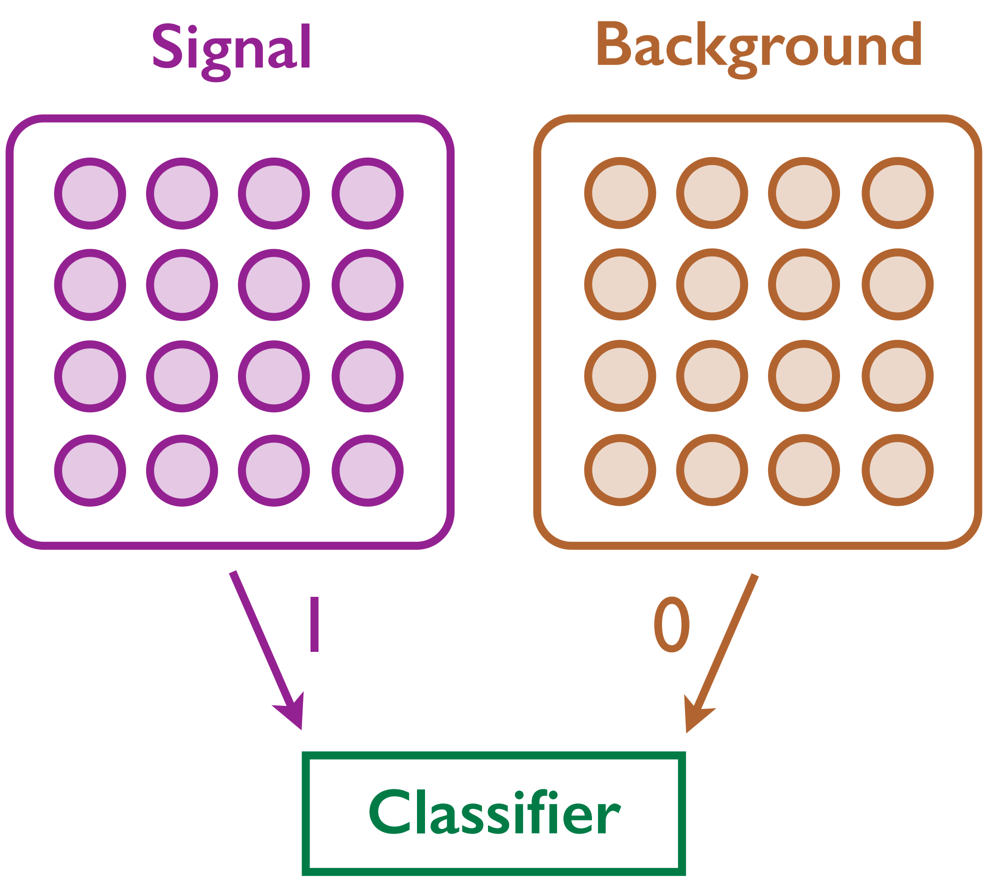
Jet Classification is "Solved"

(assuming well-defined jet categories and trustable training samples)

A Cartoon of Machine Learning

For fully-supervised jet classification

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



Minimize Loss Function

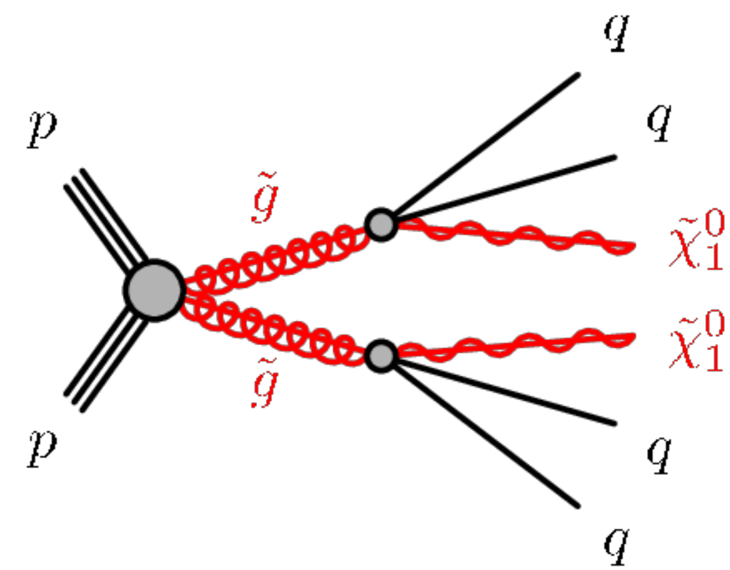
(assuming infinite training sets, and flexible enough functional form)

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

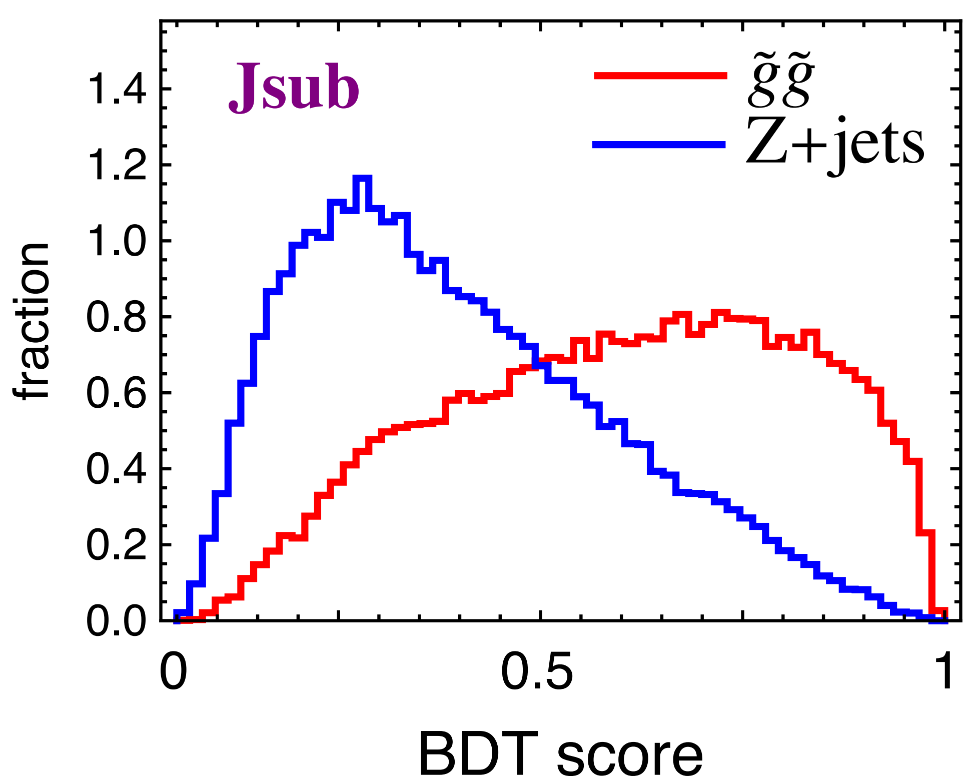
Optimal Classifier (Neyman–Pearson)

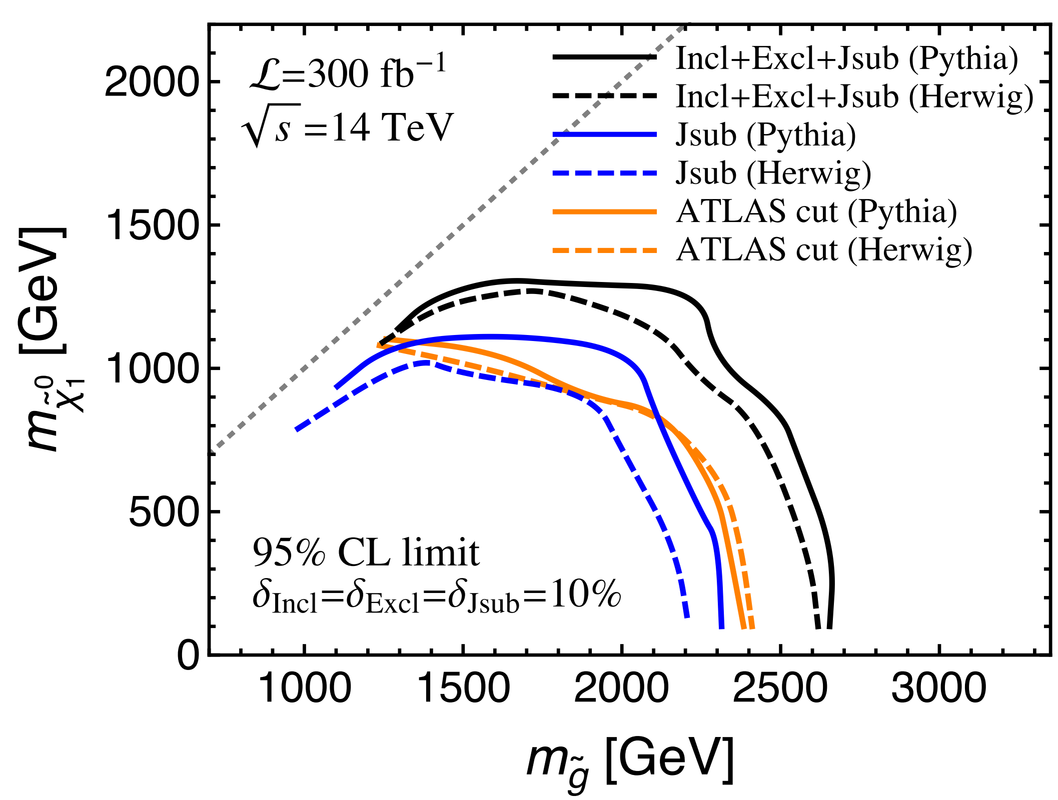
E.g. SUSY Search for Gluino Pairs

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

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

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.	…

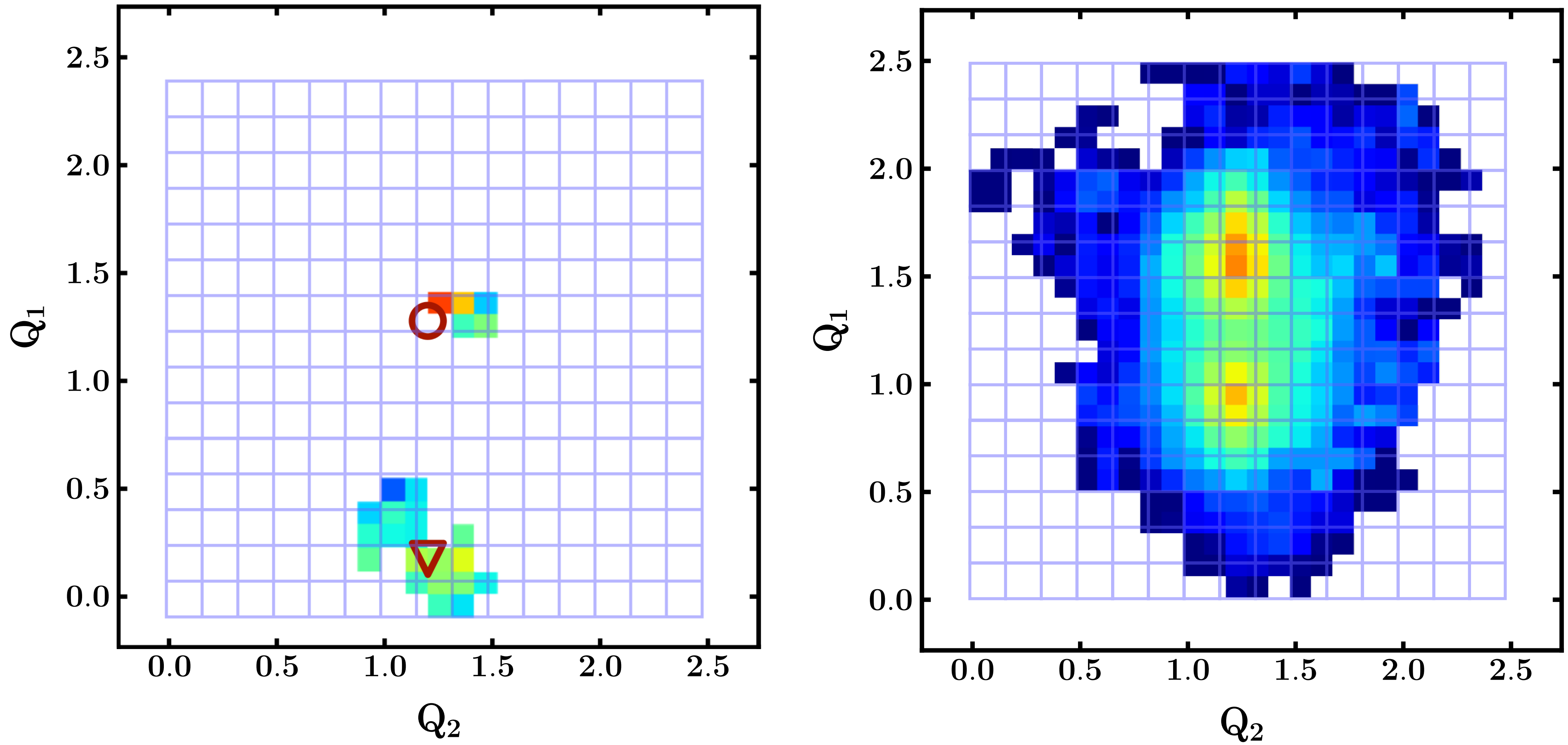
[Lönnblad, Peterson, Rögnvaldsson, 1990, …, Cogan, Kagan, Strauss, Schwartzman, 1407.5675; Almeida, Backović, Cliche, Lee, Perelstein, 1501.05968; de Oliveira, Kagan, Mackey, Nachman, Schwartzman, 1511.05190; Baldi, Bauer, Eng, Sadowski, Whiteson, 1603.09349; Conway, Bhaskar, Erbacher, Pilot, 1606.06859;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

Jet Classification Studies

Mix and match

Standard CNN input: Jet images

Individual W jet | Ensemble average

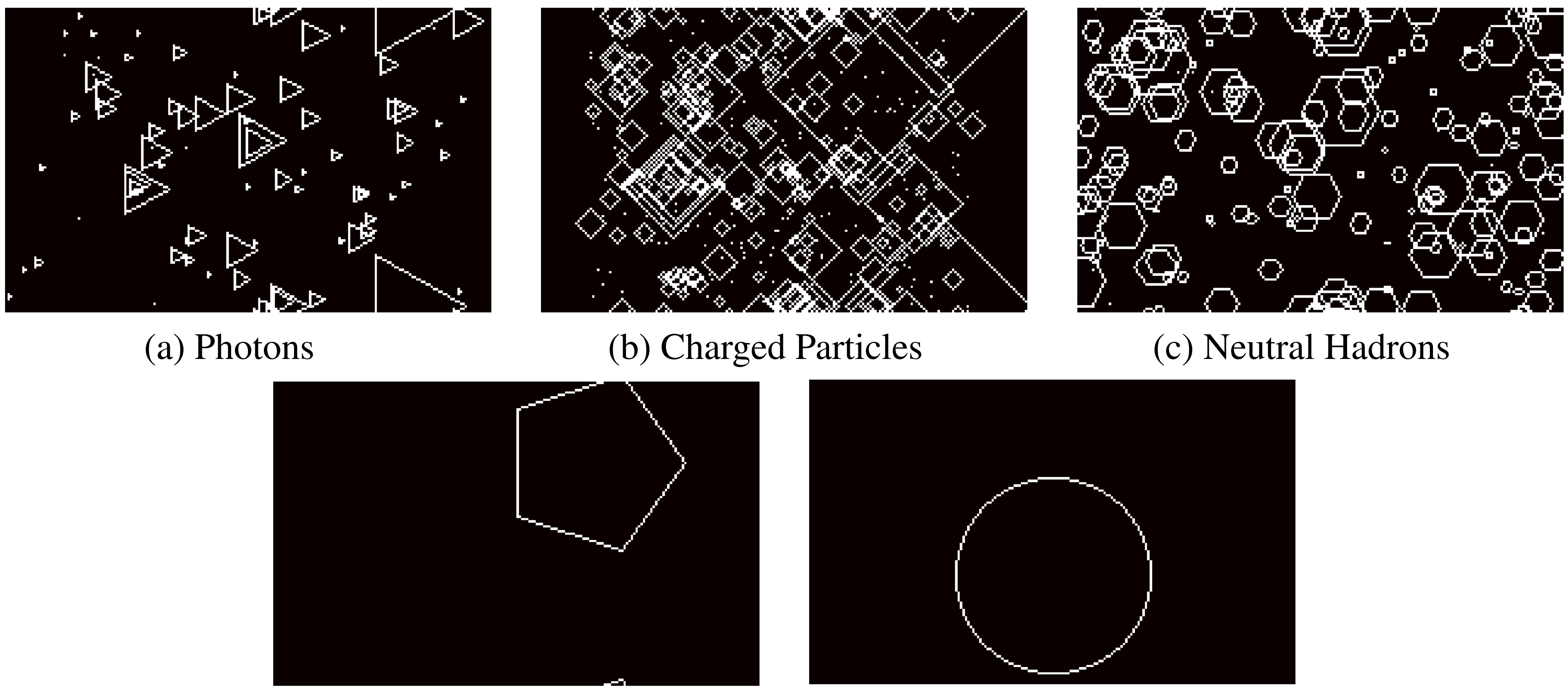


[Cogan, Kagan, Strauss, Schwartzman, 1407.5675]

Jet Classification Studies

Mix and match

Novel CNN input: Abstract event images



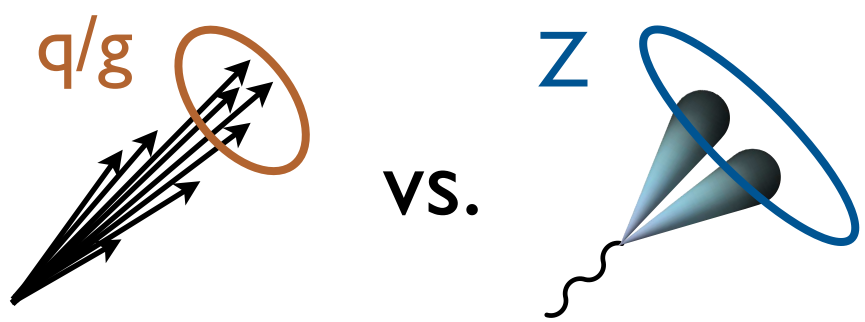
(a) Photons (b) Charged Particles (c) Neutral Hadrons

(d) Lepton (e) E_T^{miss}

Addresses sparsity problem of standard energy-to-intensity mapping

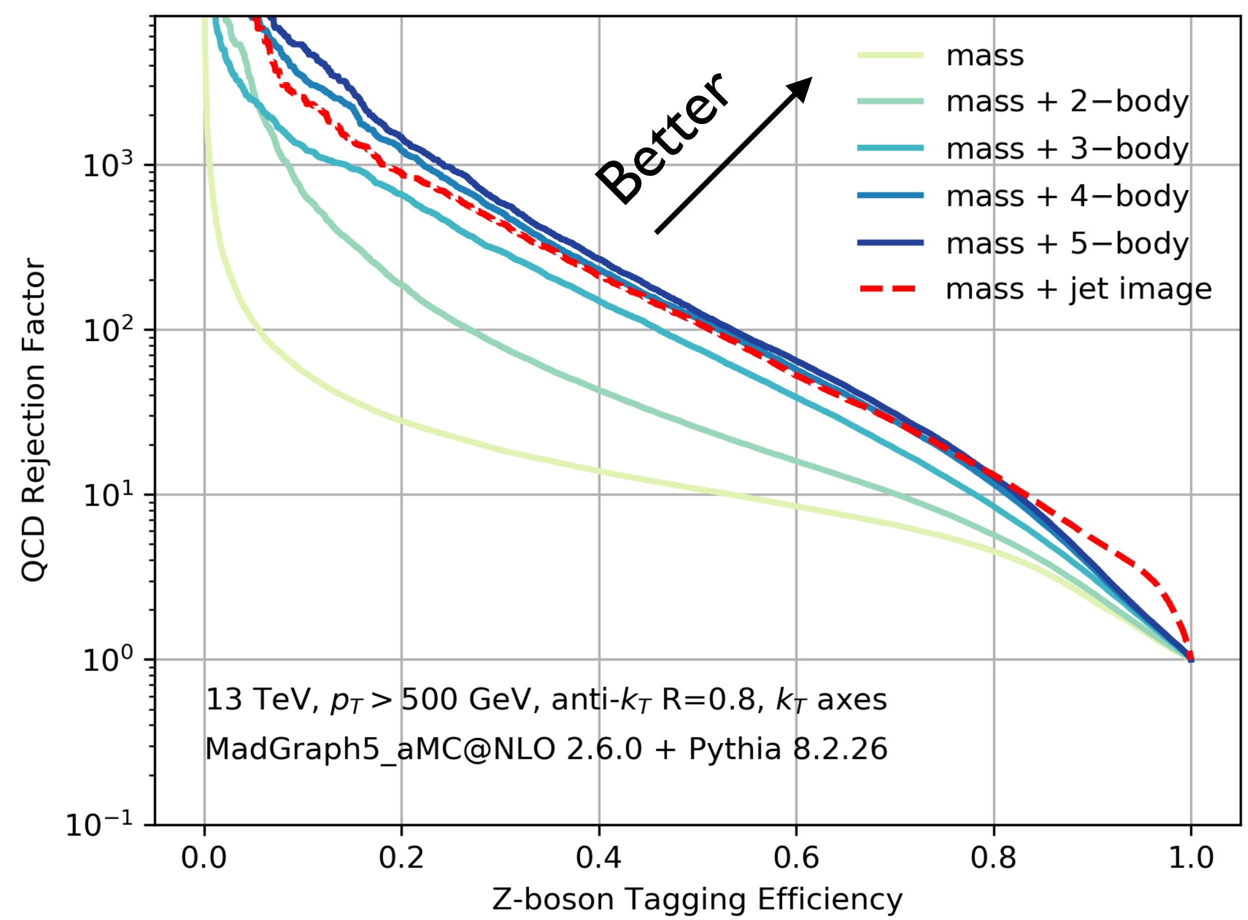
[Nguyen, Weitekamp, Anderson, Castello, Cerri, Pierini, Spiropulu, Vlimant, 1807.00083; using Fernández Madrazo, Heredia Cacha, Lloret Iglesias, Marco de Lucas, 1708.07034]

Evidence for Performance Saturation



"Any sufficiently advanced technology is indistinguishable from magic"

Jet Images CNN ≈ "Expert" BDT



Next frontier is robustness, versatility & transparency

[plot from Moore, Nordström, Varma, Fairbairn, 1807.04769; see also Datta, Larkoski, 1704.08249]

Flashback to BOOST 2016

Open Questions for Deep Learning

Hyper-variate vs. multi-variate?

Raw image processing or preprocessed "basis" inputs?

Ultimate performance boundary?

Saturated by physics in parton shower? (or go data driven?)
Approximately equivalent to BDT of existing discriminants?

Multi-category classification?

Natural in deep learning to go beyond S vs. B
e.g. for diboson excess: q, g, W^+, Z^0, W^- (and $_L$ vs. $_T$?)

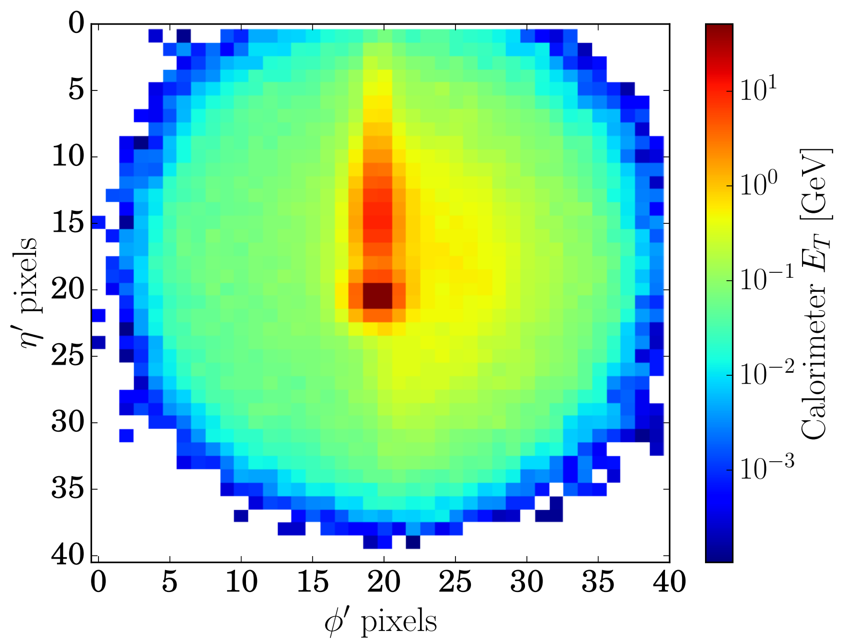
Deep thinking via deep learning?

How to understand/visualize what has been learned?
Could next uni-variate technique come from neural network study?

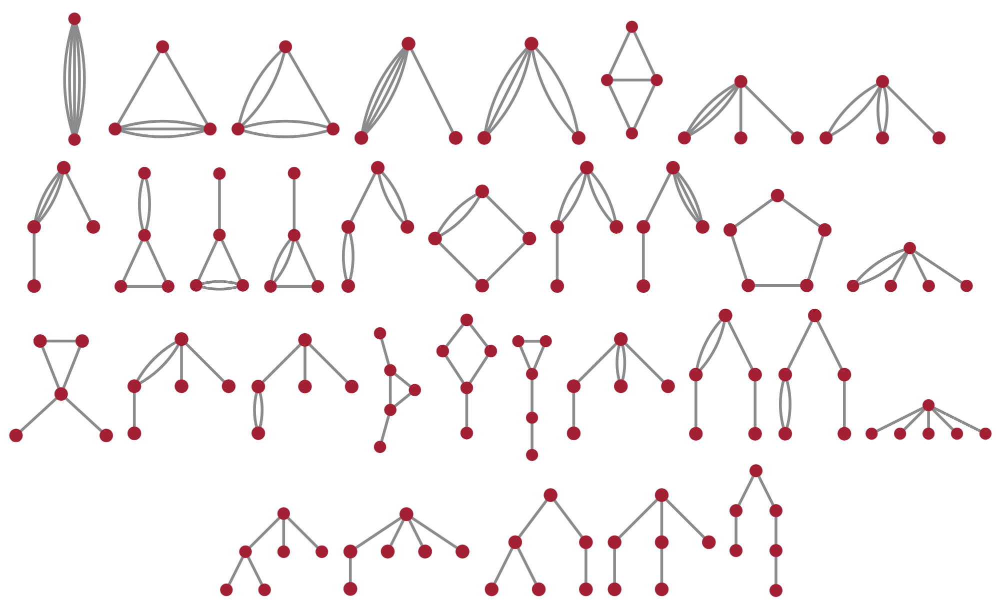
Hyper-variate vs. multi-variate?

Raw image processing or preprocessed "basis" inputs?

O(1000) pixel image



O(1000) d ≤ 7 EFPs



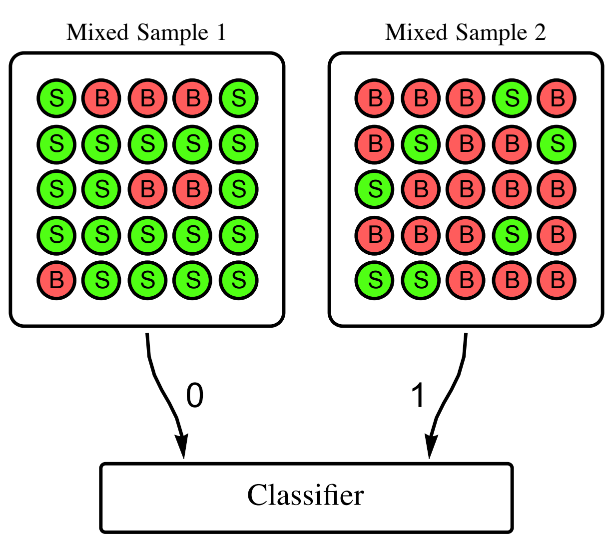
False dichotomy: Many jet encodings yield comparable performance
Which ones are easiest to validate, calibrate, unfold, reuse?

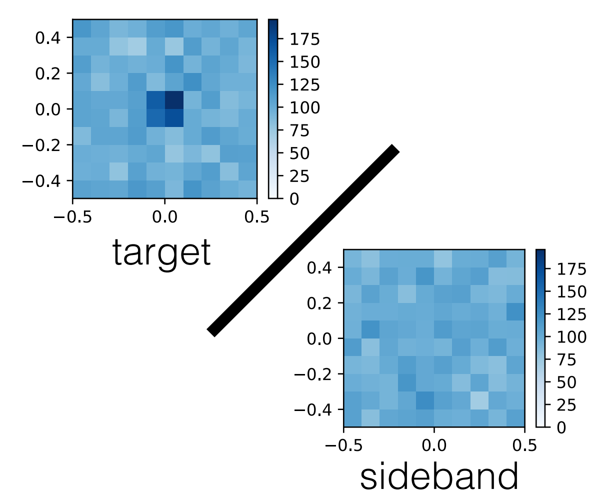
[Kasieczka, Plehn, Russell, Schell, 1701.08784; Komiske, Metodiev, JDT, 1712.07124]

Ultimate performance boundary?

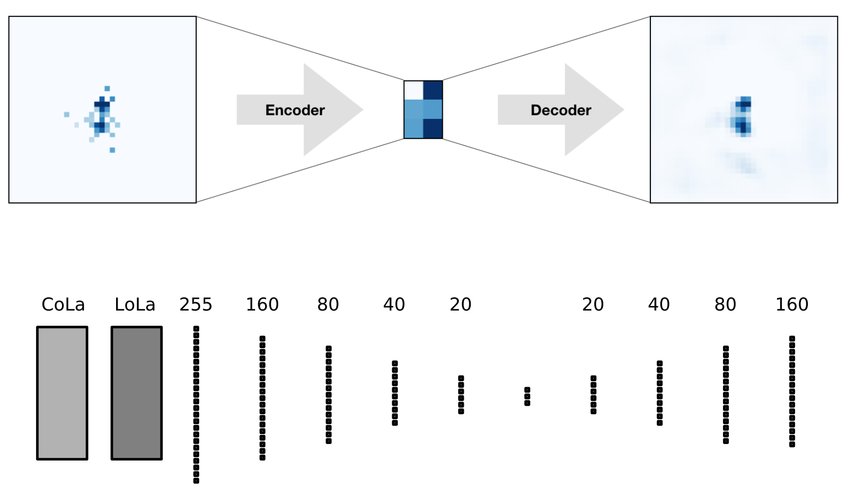
Saturated by physics in parton shower? (or go data driven?)
Approximately equivalent to BDT of existing discriminants? *(yes)*

Weak Supervision





Anomaly Detection

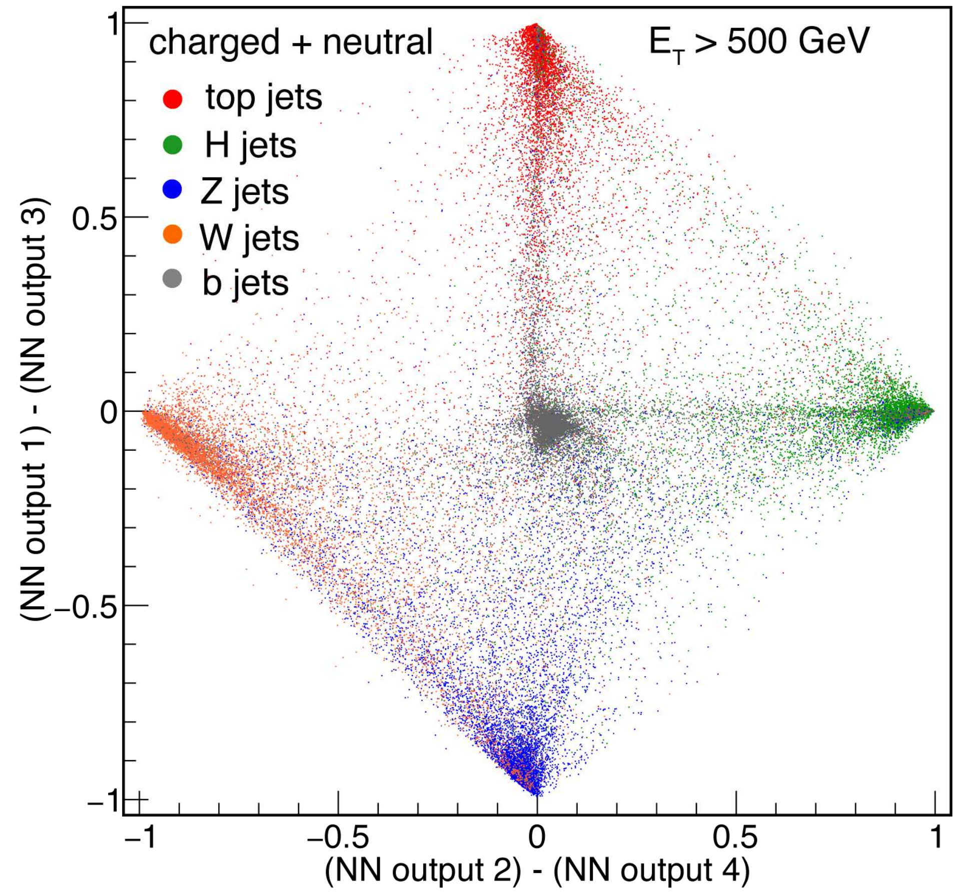


Shifting the burden from simulation to assumptions (e.g. sample independence including detector effects)

[Metodiev, Nachman, JDT, 1708.02949; Collins, Howe, Nachman, 1805.02664; Heimel, Kasieczka, Plehn, Thompson, 1808.08979; Farina, Nakai, Shih, 1808.08992]

Multi-category classification?

Natural in deep learning to go beyond S vs. B
e.g. for diboson excess: q, g, W^+, Z^0, W^- (and $_L$ vs. $_T$?)



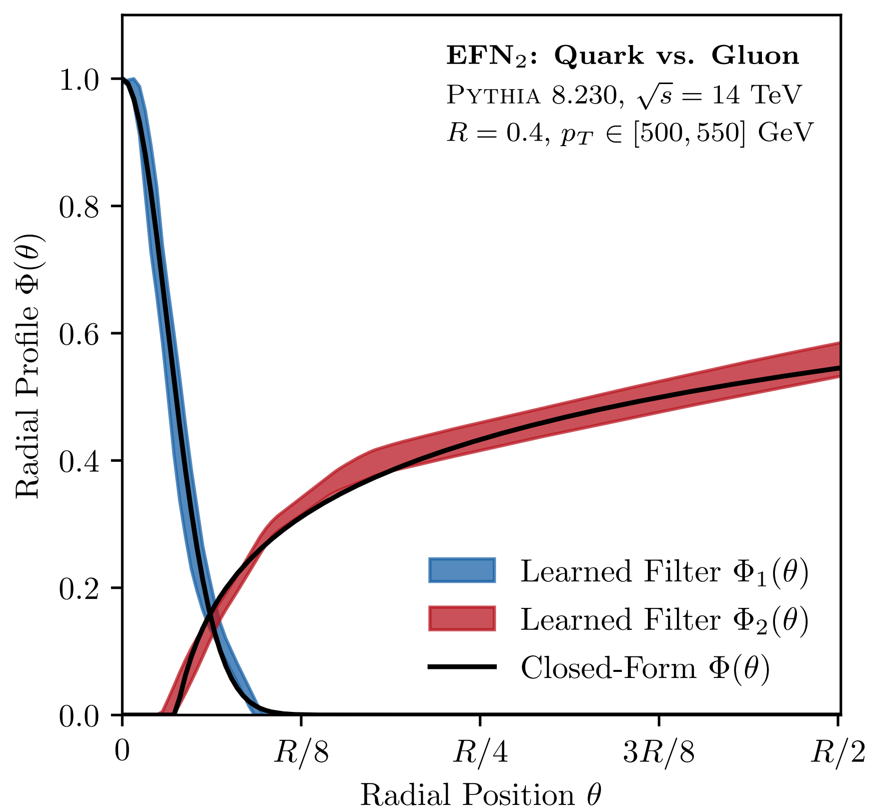
Where is my 11-category quark/gluon tagger?
Polarization-dependent W/Z/top tagger? Multi-category CWoLa?

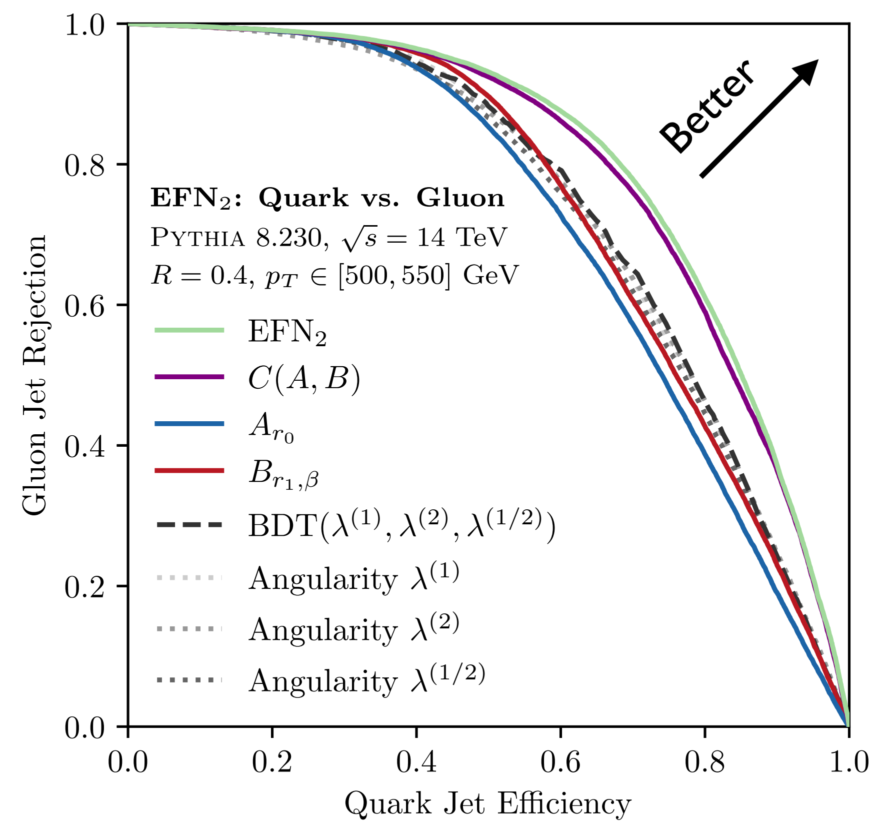
[Conway, Bhaskar, Erbacher, Pilot, 1606.06859; see also Stoye, Kieseler, Qu, Gouskos, Verzetti, DLPS 2017]

Deep thinking via deep learning?

How to understand/visualize what has been learned?

Could next *bi*-variate technique come from neural network study?





A neural network outsmarted a domain expert (i.e. me)

But we reverse engineered the machine

[Komiske, Metodiev, JDT, 1810.05165; see talk on Thursday]

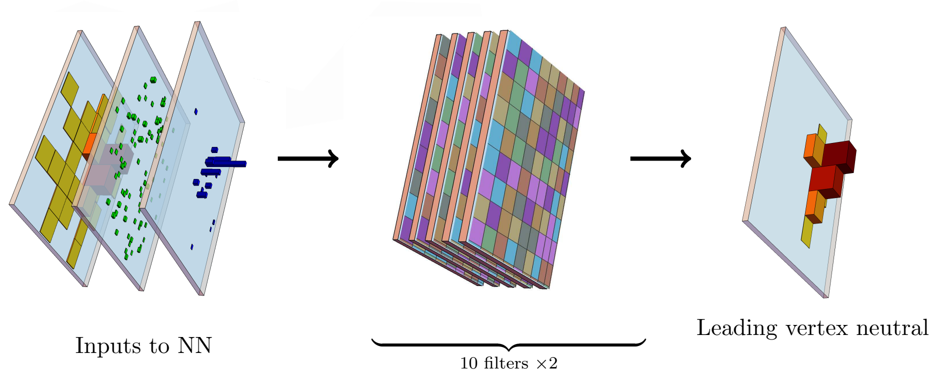
Predictions for ML4Jets 2019

More ML Applications Beyond Classification

Regression

e.g. PUMML for pileup mitigation

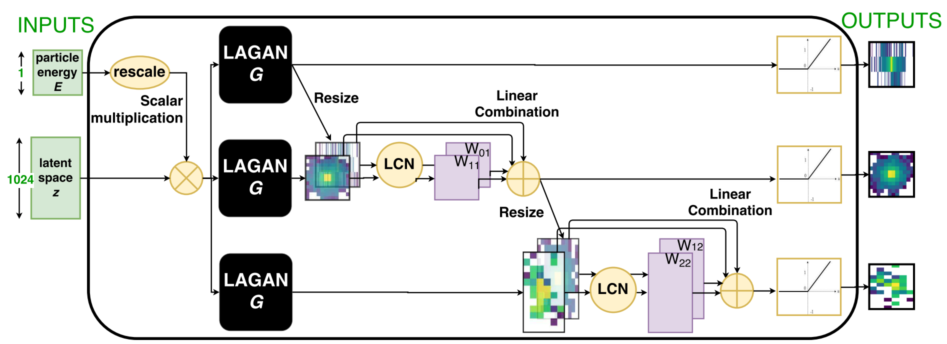
[Komiske, Metodiev, Nachman, Schwartz, 1707.08600;
see also Arjona Martínez, Cerri, Pierini, Spiropulu, Vlimant, 1810.07988]



Generation

e.g. CaloGAN for fast detector simulation

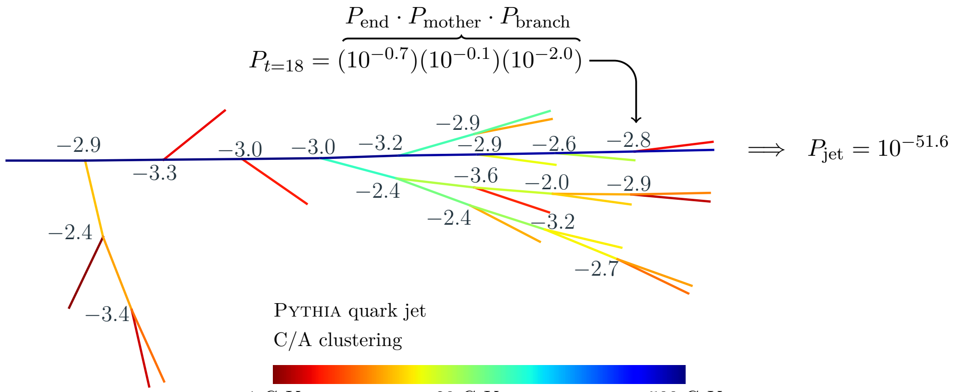
Paganini, de Oliveira, Nachman, 1705.02355, 1712.10321;
see also de Oliveira, Michela Paganini, Nachman, 1701.05927]



Probability Modeling

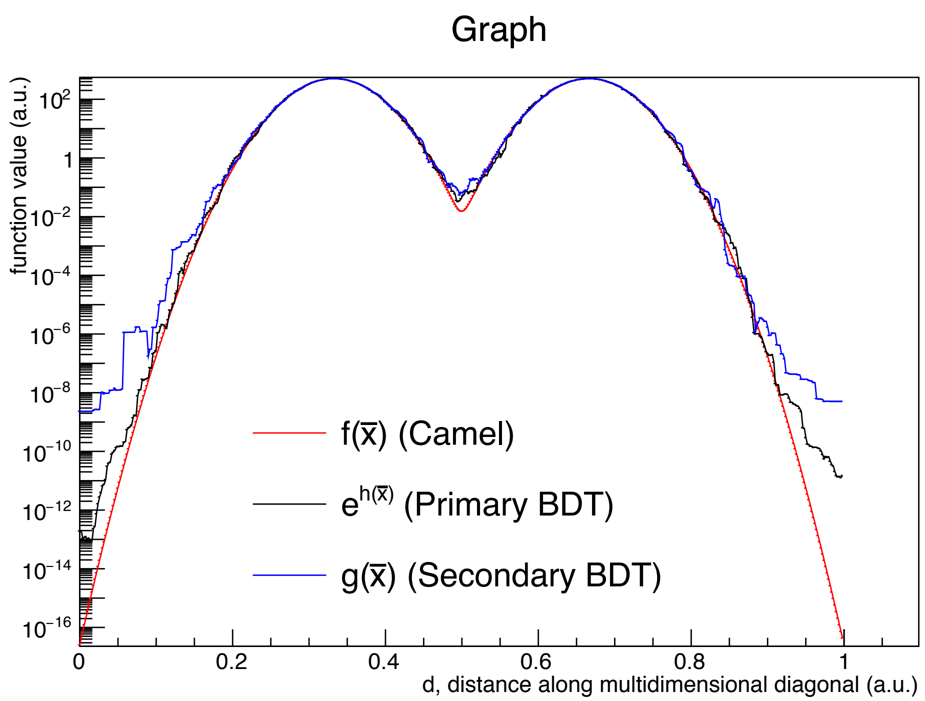
e.g. JUNIPR to characterize parton shower

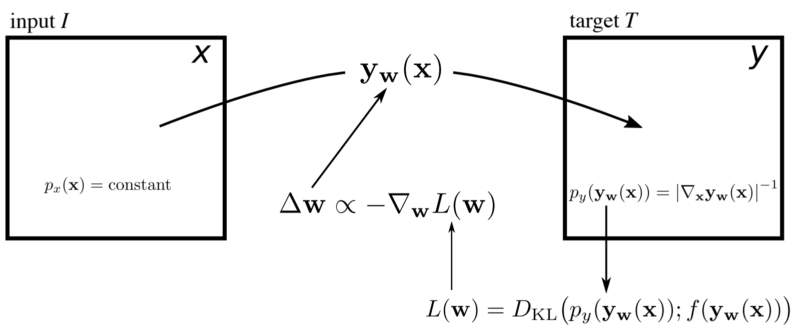
[Andreassen, Feige, Frye, Schwartz, 1804.09720;
see also Monk, 1807.03685]

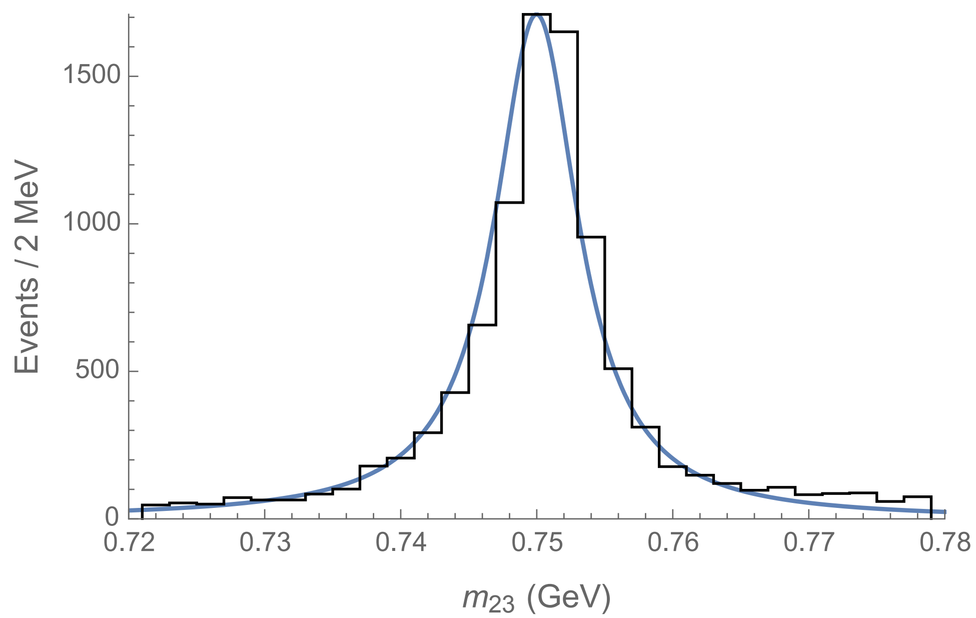


More "Uncontroversial" Uses of ML

Phase Space Integration





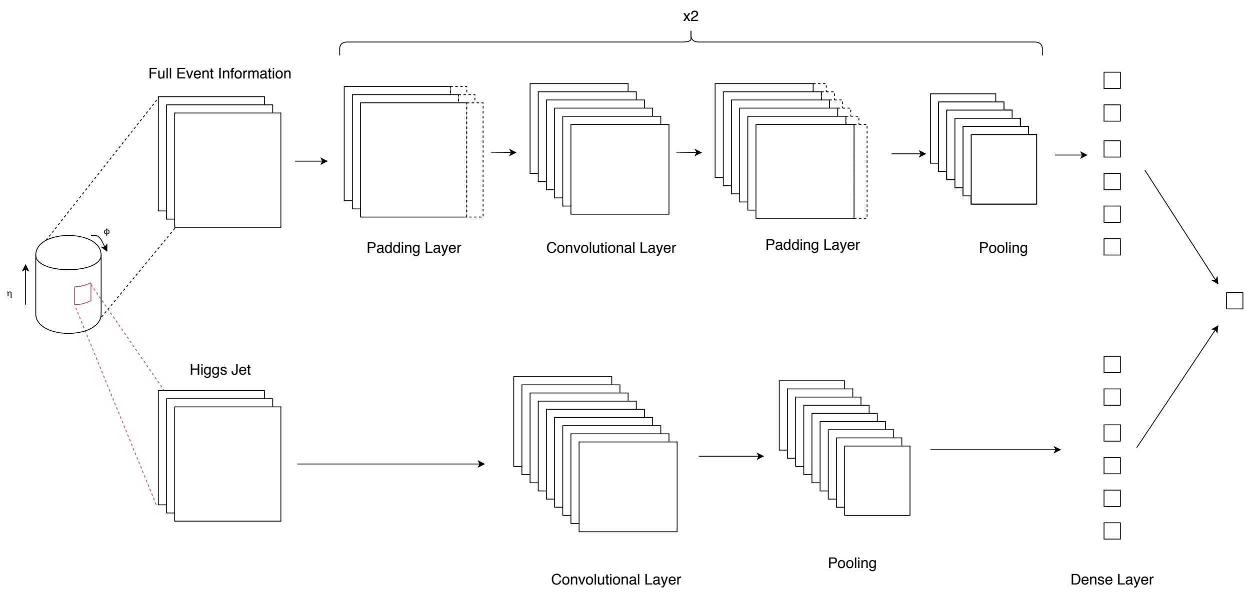


Applications like where ML improves efficiency with no change in formal asymptotic behavior

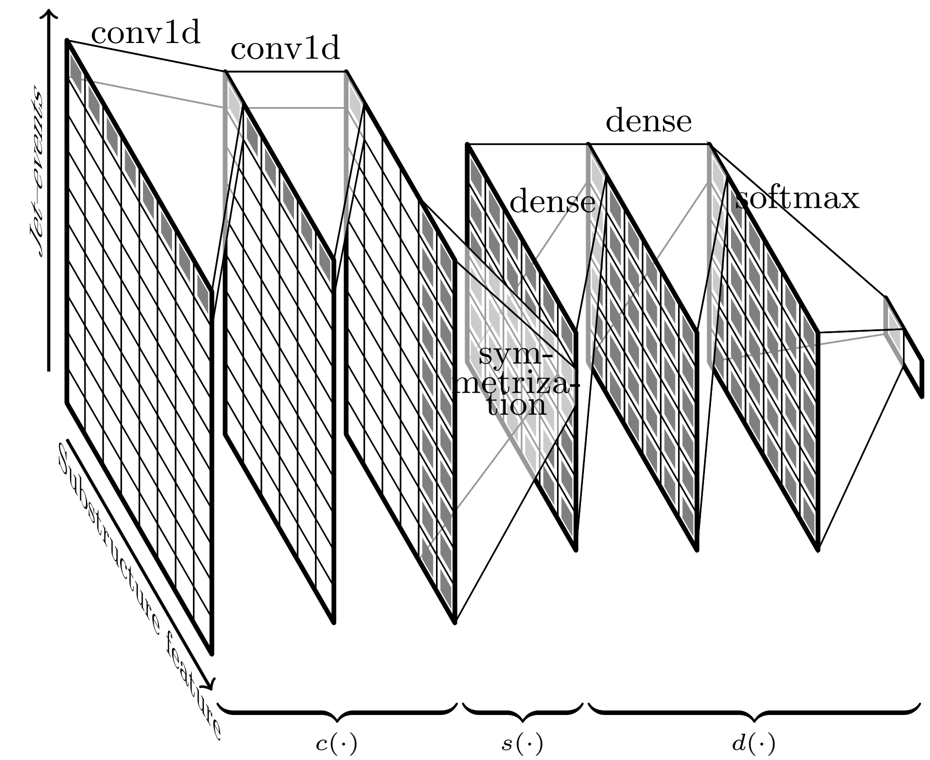
[Bendavid, 1707.00028; Klimek, Perelstein, 1810.11509]

More Event/Ensemble Learning

Boosting H → bb



QCD Plasma Temperature

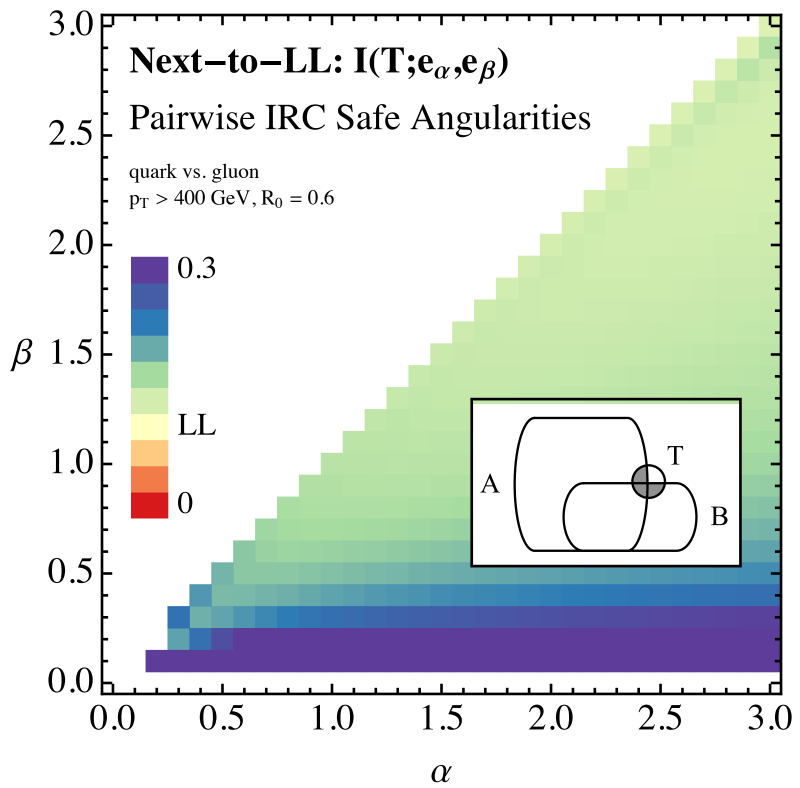


Can we learn about jets more effectively/efficiently from training on entire events? On sets of events?

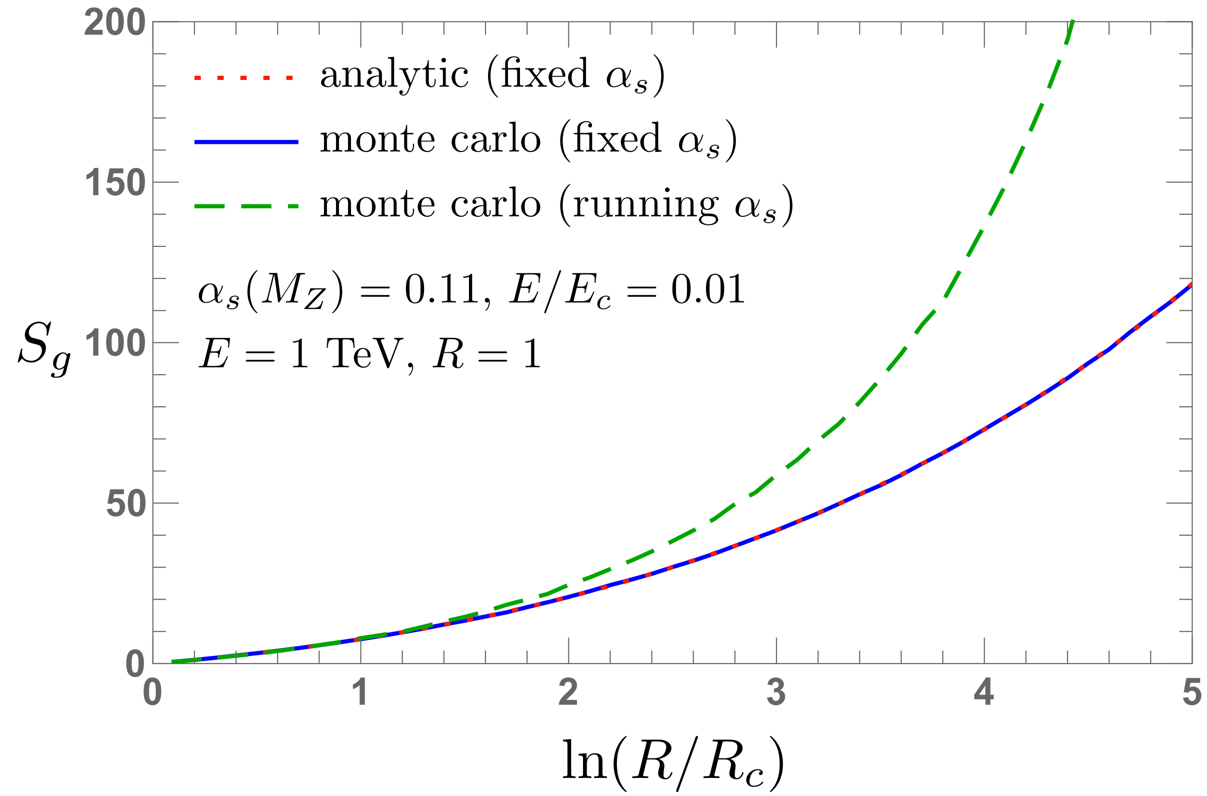
[Lin, Freytsis, Moult, Nachman, 1807.10768; Lai, 1810.00835]

More First-Principles Calculations

Mutual Information



Von Neumann Entropy

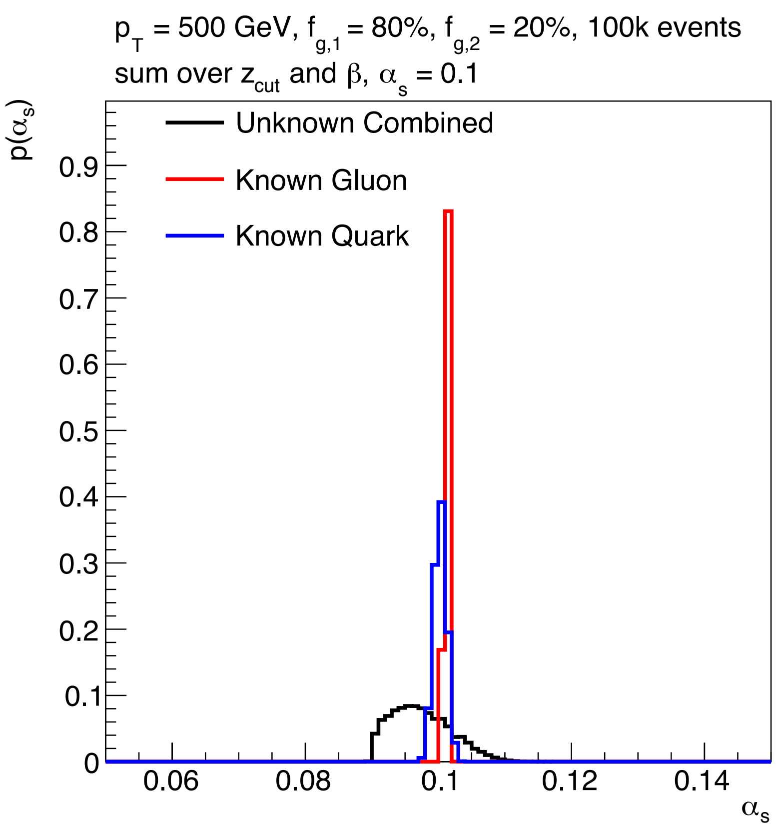


Can we improve ML with a better understanding of jet information theory?

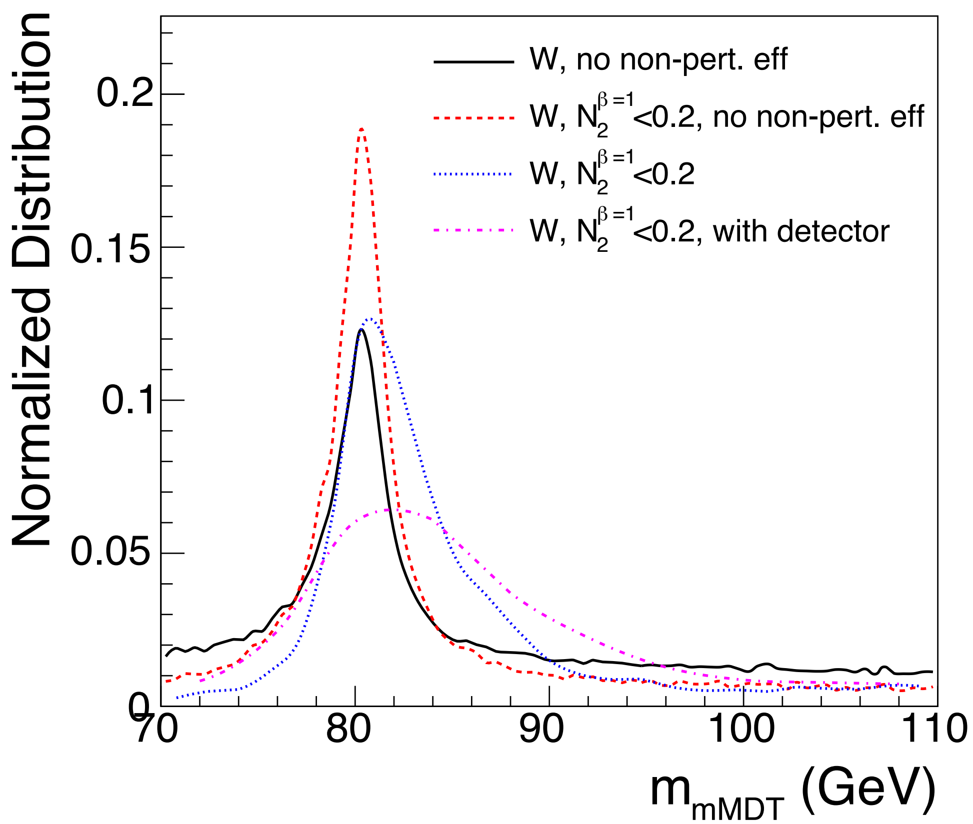
[Larkoski, JDT, Waalewijn, 1408.3122; Neill, Waalewijn, 1811.01021]

More Standard Model Measurements

Strong Coupling Constant



Hadronic W Mass

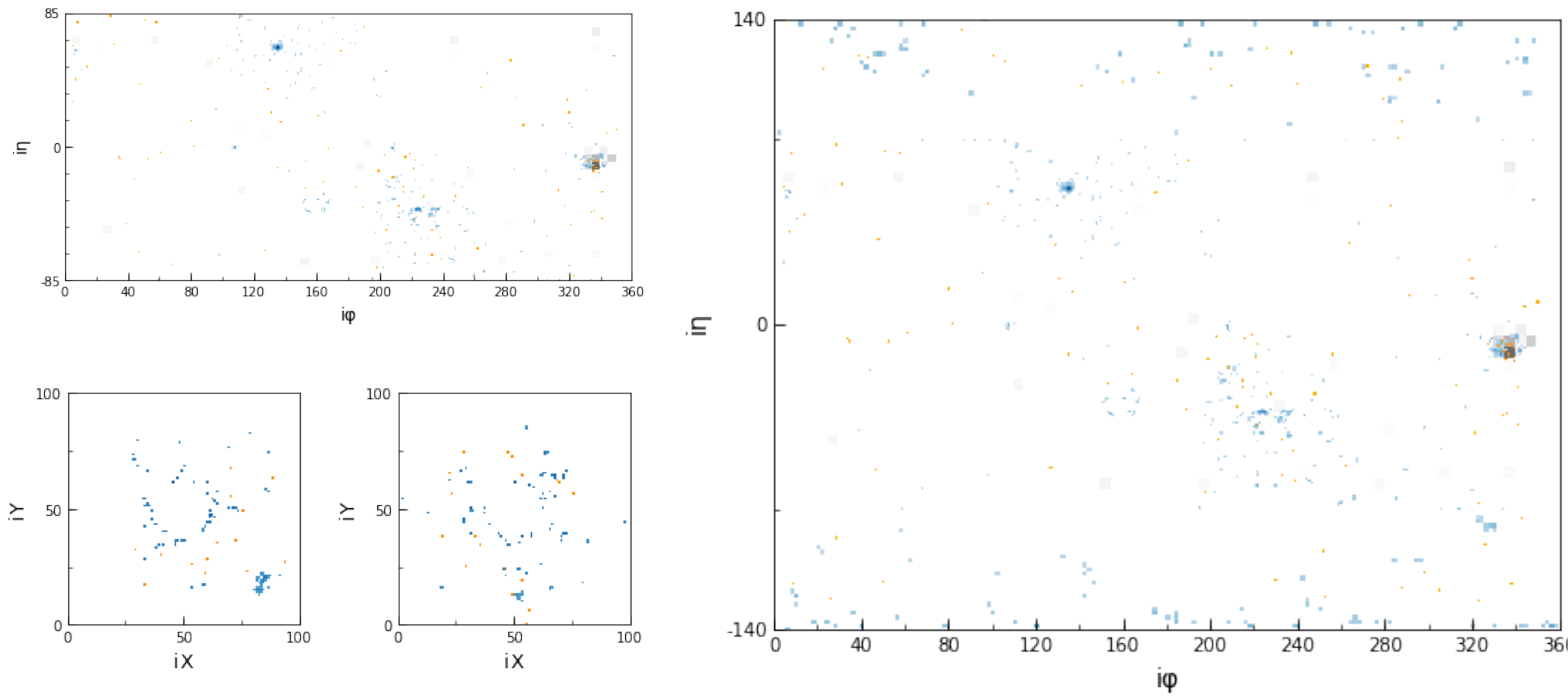


If we set (overly) ambitious goals, then we might reach them
Can ML help for precision measurements?

[Moult, Nachman, Soyez, JDT, et al., 1803.07977; Freytsis, Harris, Hinzmann, Moult, Tran, Vernieri, 1807.07454]

More Crossing Boundaries

Shower Shapes for $H \to \gamma\gamma$



CMS Open Simulation: Opportunity to pursue end-to-end deep learning at experiment/theory & HEP/CS interfaces

[Andrews, Paulini, Gleyzer, Poczos, 1807.11916]

Using All *Five* LHC Interaction Points

P1 P2 P5 P8 R1

Reiterating My Goals

$\mathcal{L}$ $\mathcal{O}$ $\mathcal{M}$

Learn about state-of-the-art ML and novel HEP applications (esp. beyond fully-supervised classification)

Press issue of verifiability (i.e. robustness, transparency, calibration, unfolding, visualization)

Explore possibilities for factorization (i.e. versatility, component reuse, symmetries, transfer learning)

Seek opportunities to push boundaries of fundamental physics

And have fun! ML4Jets is a young, vibrant, and creative subfield