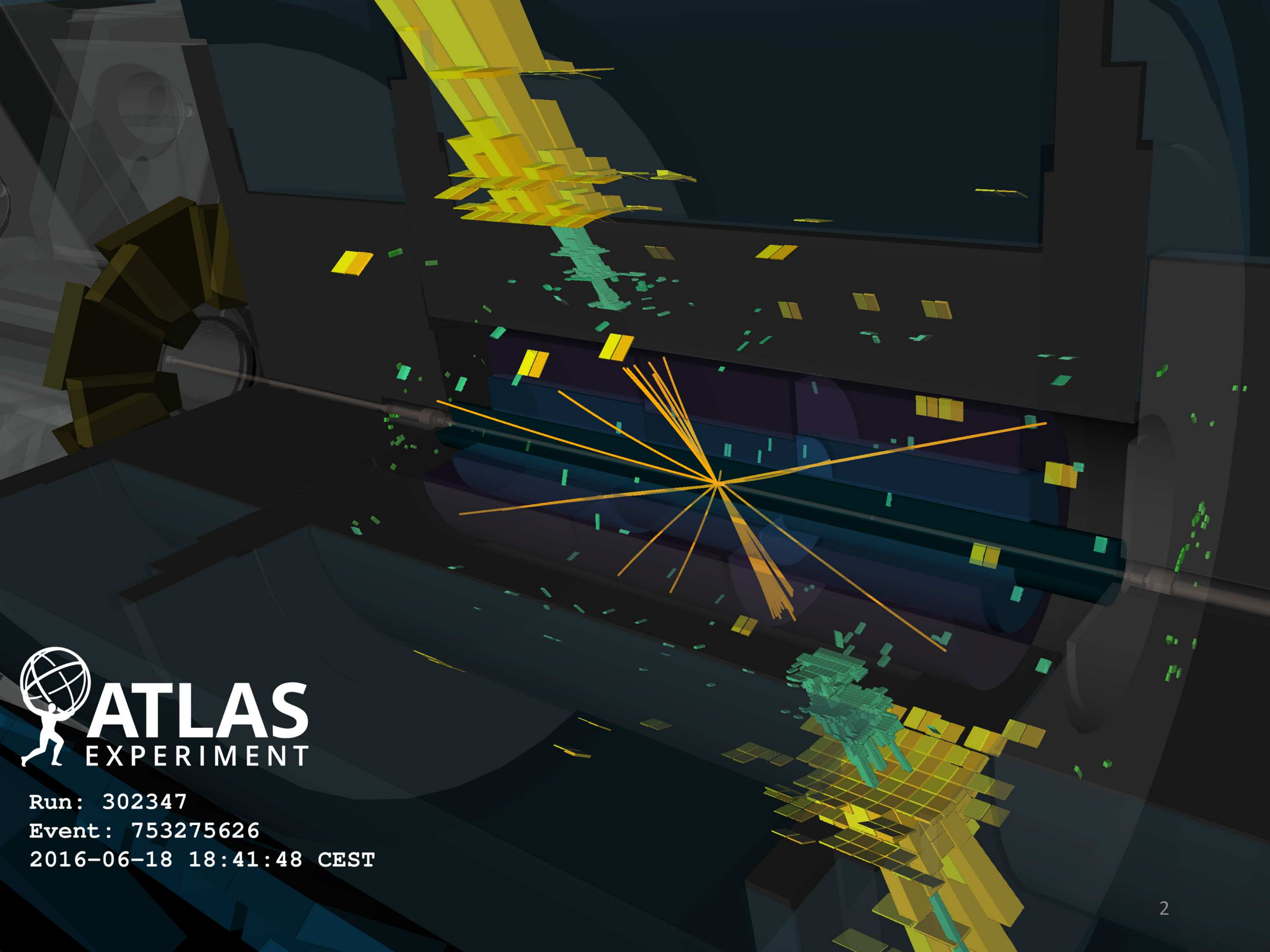
Jet Physics & Modern Machine Learning

Harvard Physics Lunch Talk

Patrick T. Komiske and Eric M. Metodiev
Center for Theoretical Physics
Massachusetts Institute of Technology

Collaborators: Benjamin Nachman, Matthew D. Schwartz, and Jesse Thaler

February 7, 2018


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

Jets in Theory

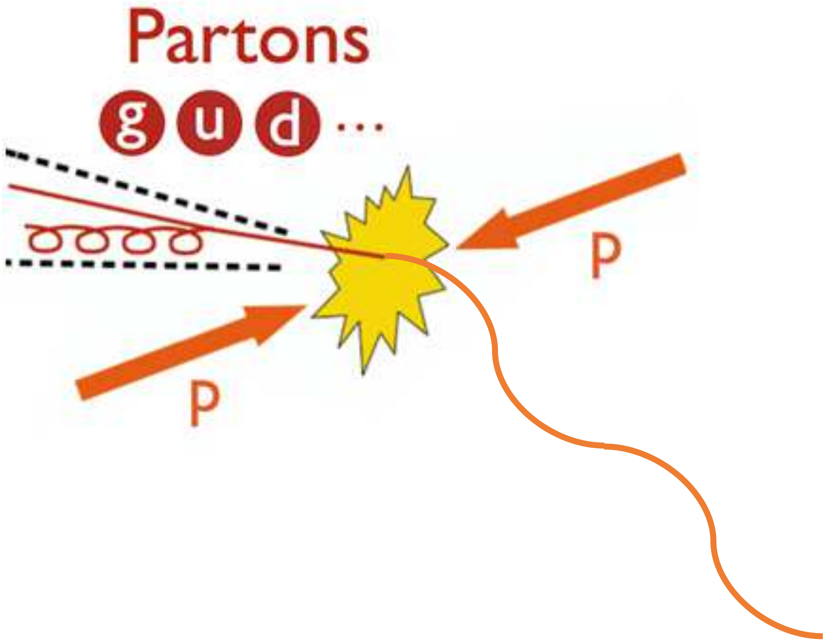


Diagram inspired by J. Thaler and M.D. Schwartz

Jets in Theory

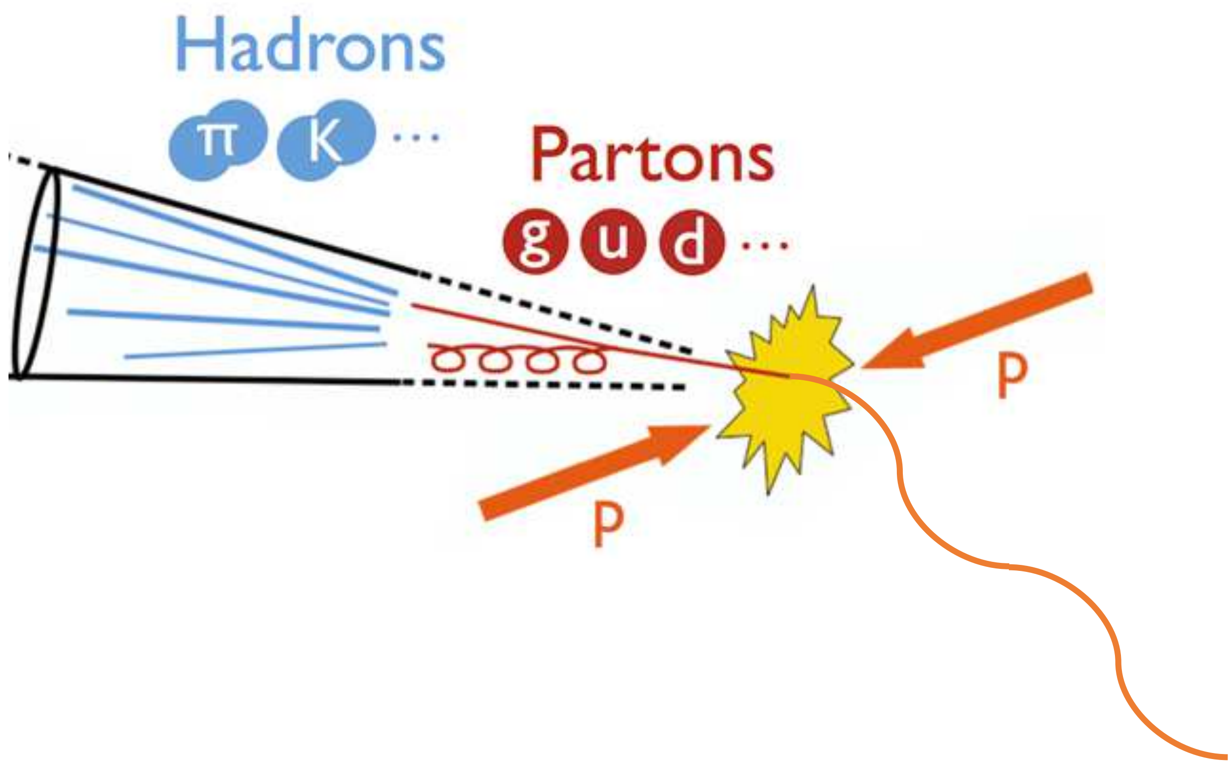


Diagram inspired by J. Thaler and M.D. Schwartz

Jets in Theory in Practice

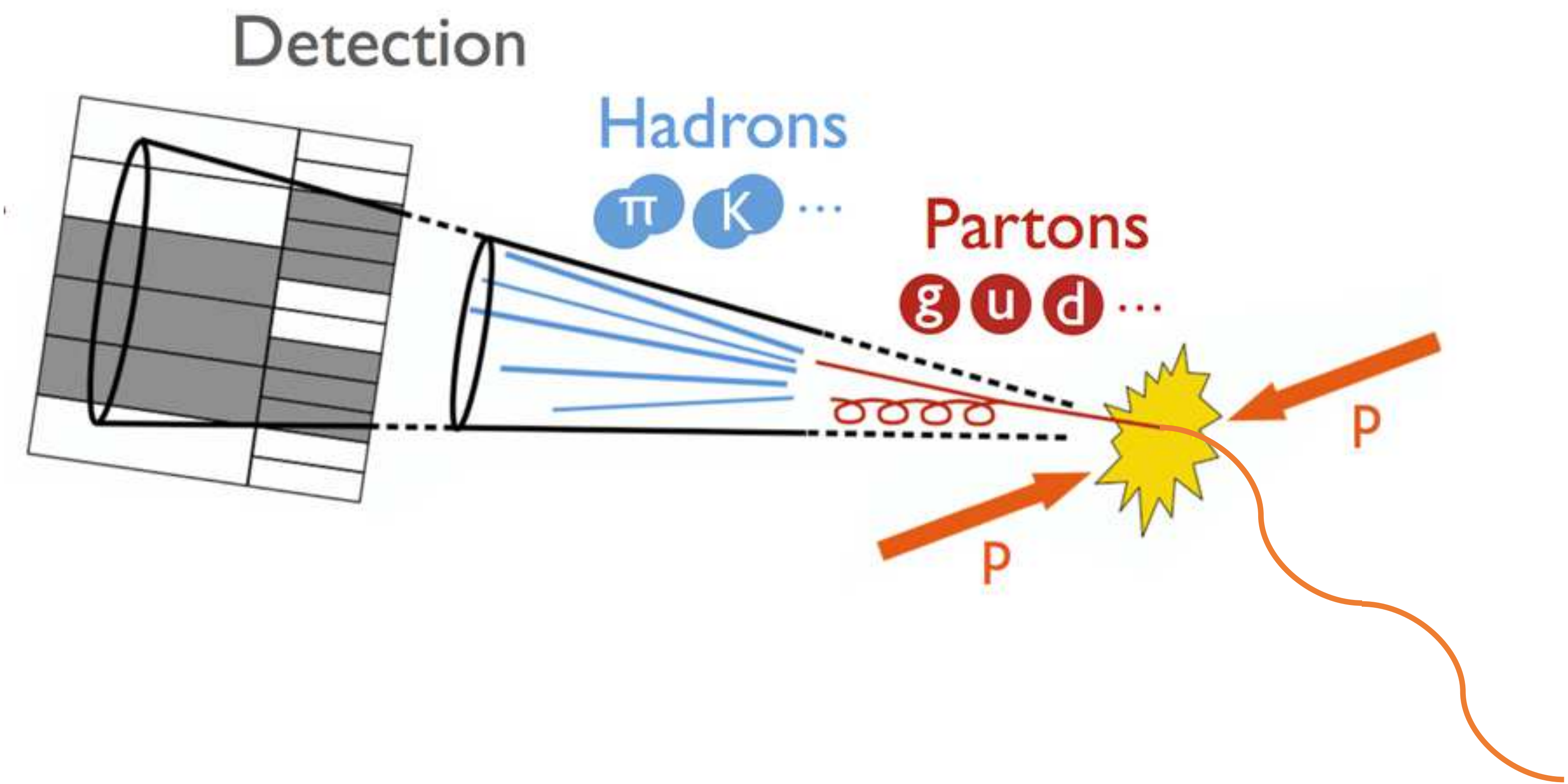


Diagram inspired by J. Thaler and M.D. Schwartz

Jets in Theory in Practice in Theory

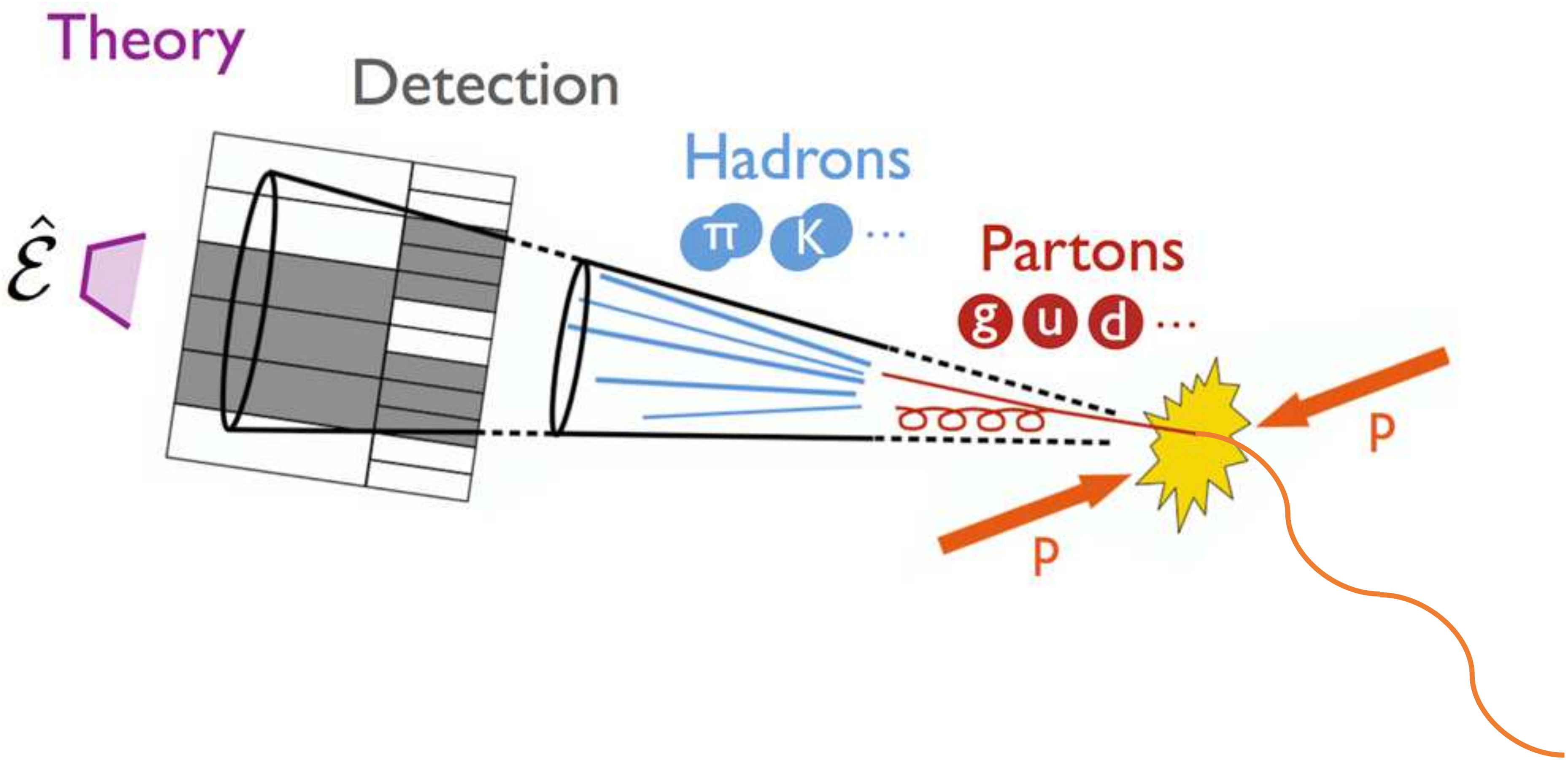


Diagram inspired by J. Thaler and M.D. Schwartz

Jets in Theory in Practice in Theory in Practice

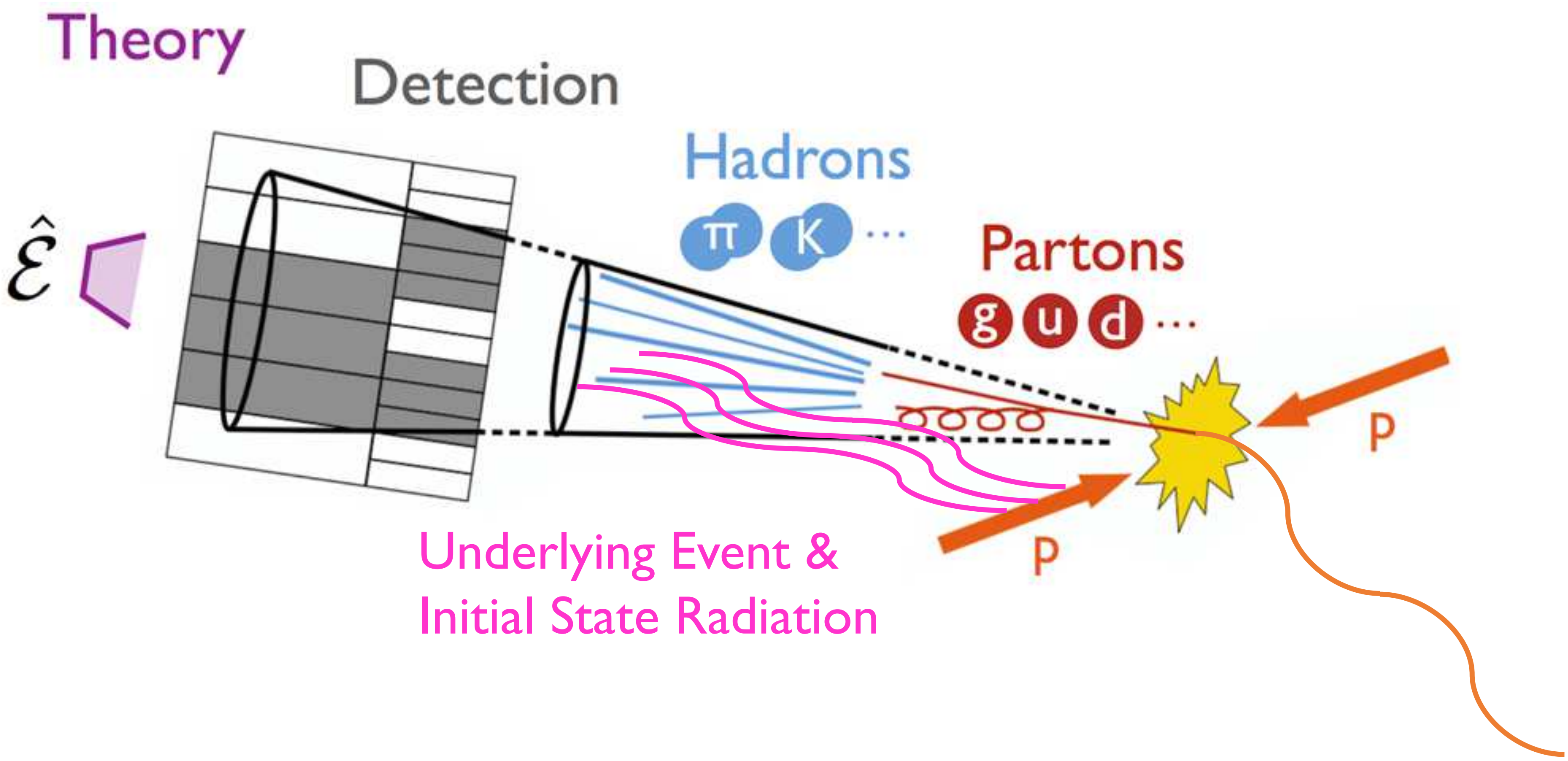


Diagram inspired by J. Thaler and M.D. Schwartz

Jets in Theory in Practice in Theory in Practice… ☹

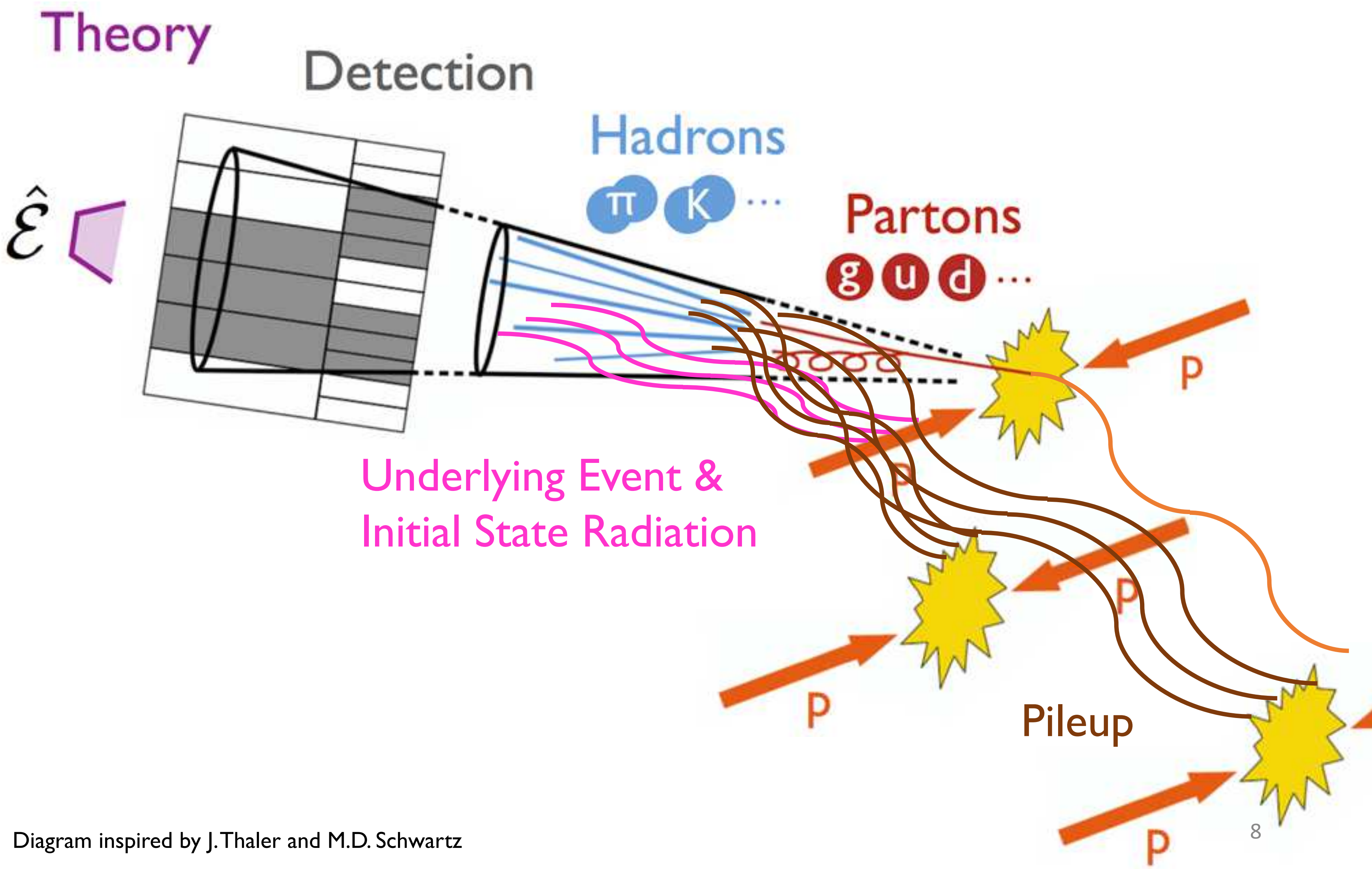


Diagram inspired by J. Thaler and M.D. Schwartz

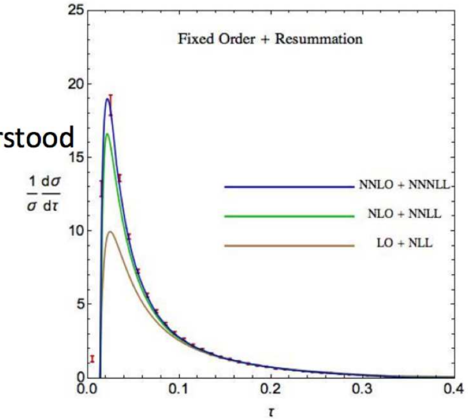
Jets in Theory in Practice in Theory in Practice…

Can ML help?

We need to master:

- Final state radiation
- Soft radiation from other jets
- Hadronization
  - Single scale Λ_{QCD}
  - Universal power corrections
  - Shape-function models

Present, studied, and fairly well understood in e^+e^-



- Initial state radiation
  - Soft radiation into jets understood
  - Collinear radiation understood with beam functions
- Underlying event
  - Modeled, but no systematic theory
- Pileup
  - Stochastic
  - Uncorrelated with jet shapes

Many powerful methods developed

- Area subtraction, cleansing, PUPPI, …

- Non-global logarithms
  - Extra scales ruin factorization
  - Some progress on resummation
  - Active area of research

NLL resummation of NGLs

- Coherent branching approach [Dasgupta&Salam, Banfi, Marchesini,Smye]
- EFT approach [Becher&Neubert, Larkoski, Neill, Moult]

- Factorization-violating effects
  - When is factorization violated?
  - How do we separate perturbative from non-pertubative effects?
  - Are there super-leading logarithms?

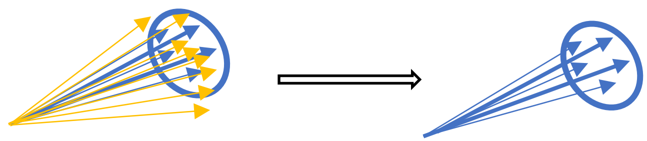
Can ML help?

Slide from overview talk by Matthew Schwartz at 2017 ML for Jets workshop at LBNL

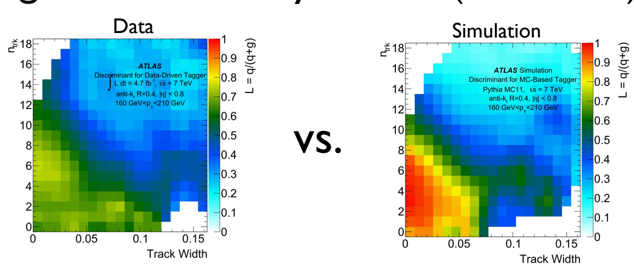
Jet Tasks We'll Talk About

Jet Tagging: How can we distinguish a quark jet vs. a gluon jet? A *W* jet vs. a QCD jet?

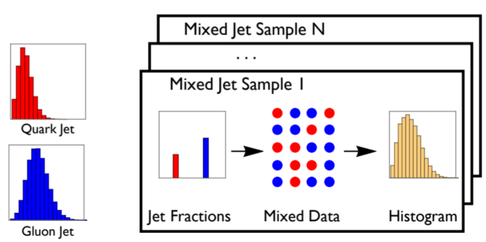
Pileup Mitigation: Can we decontaminate the jet radiation from soft, diffuse pileup?



Data vs. Simulation: Do we really need simulations to provide labeled training data? Or are there ways to train algorithms directly on the (unlabeled) data?

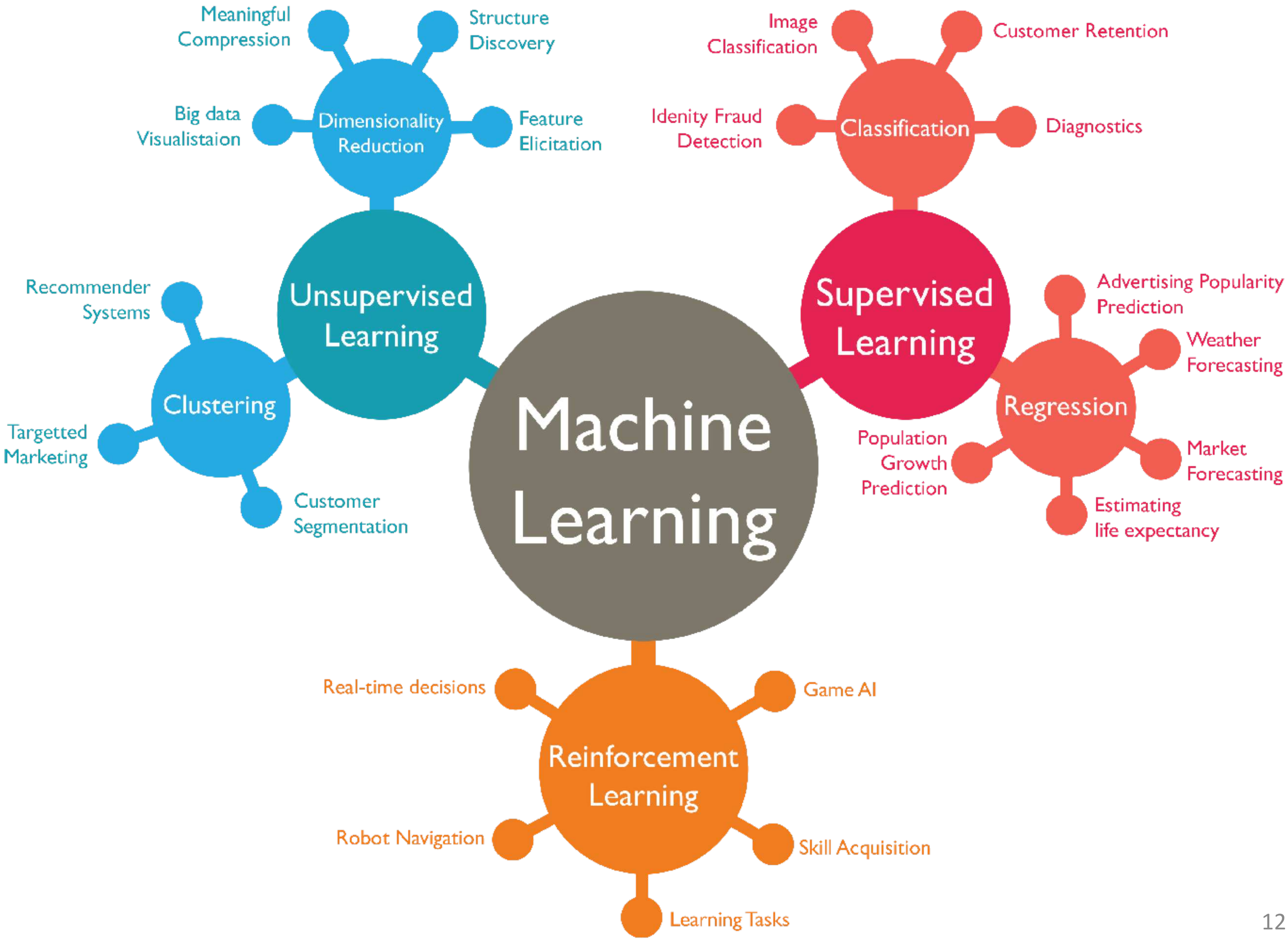


Measuring Jet Observables: Do we need to perfectly classify quark and gluon jets to separately measure quark and gluon jet observable distributions?

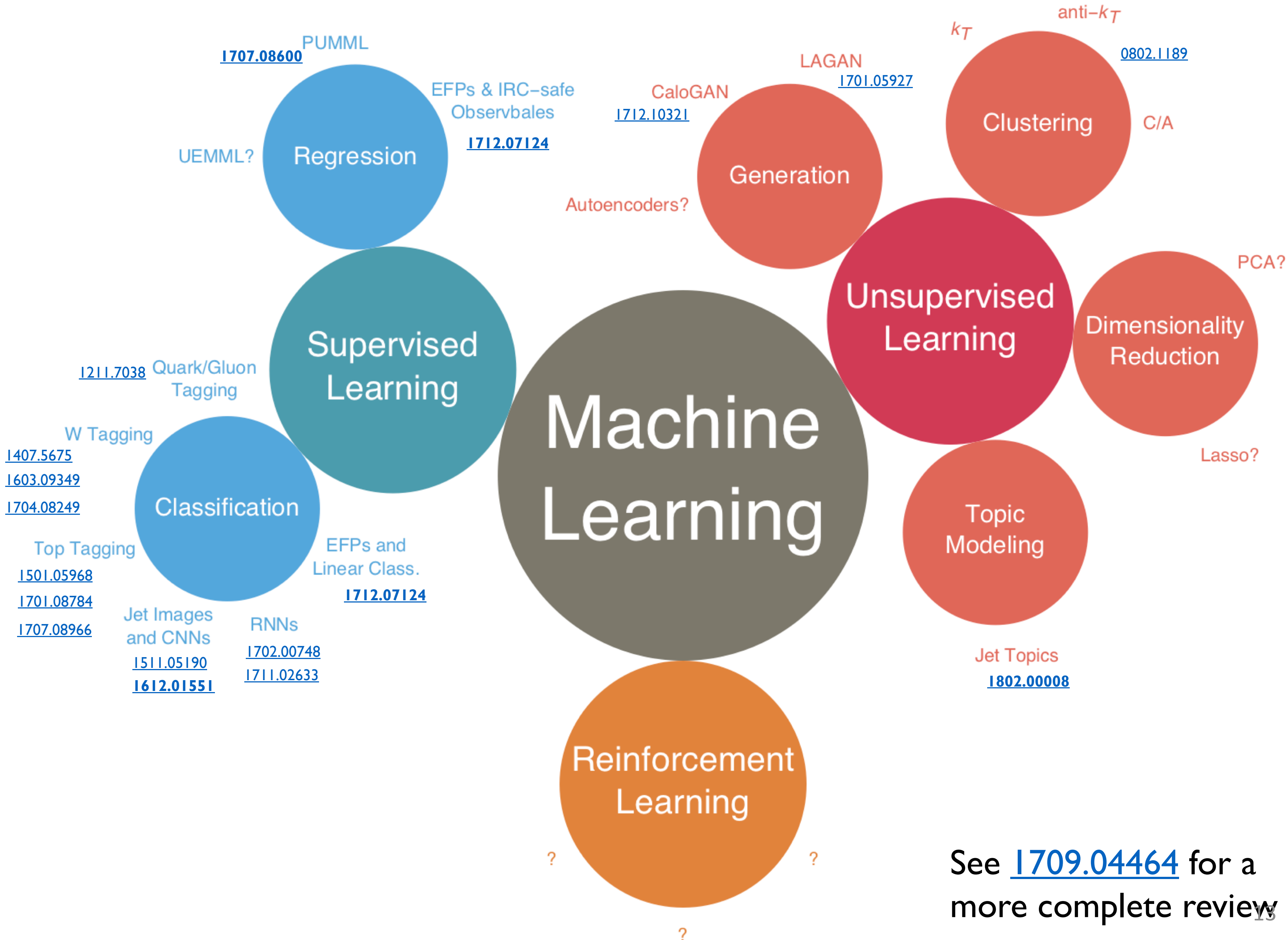


Machine Learning

Machine Learning



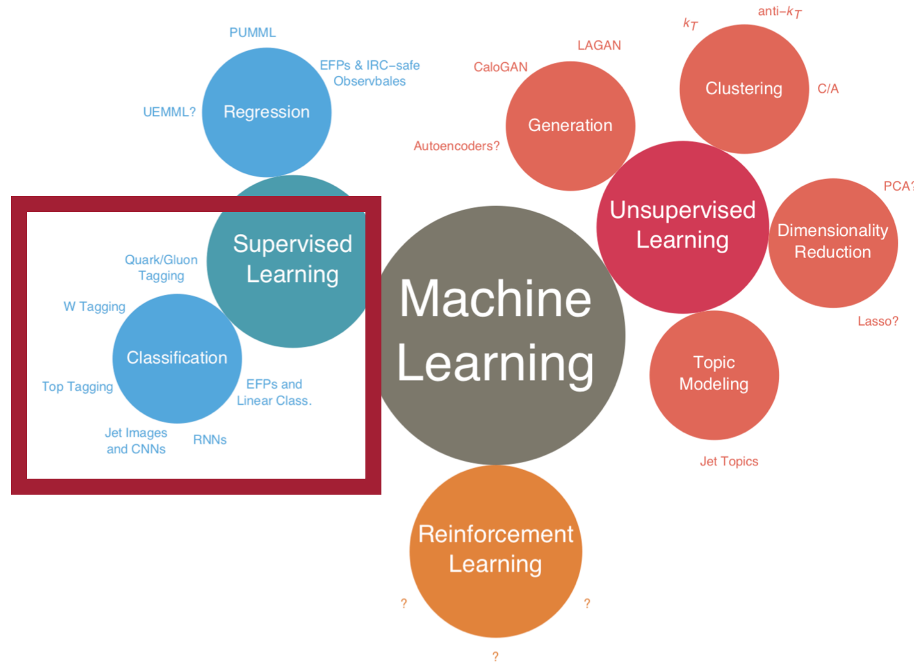
Machine Learning in High Energy Physics



See 1709.04464 for a more complete review

Quark vs. Gluon Jet Tagging

[**PTK, EMM**, M.D. Schwartz, 1612.01551]



For many BSM processes:

Quark = Signal

Gluon = Background

Quark charge: C_F = 4/3
Gluon charge: C_A = 3 $\Longrightarrow$ Gluons radiate more than quarks and are "wider"

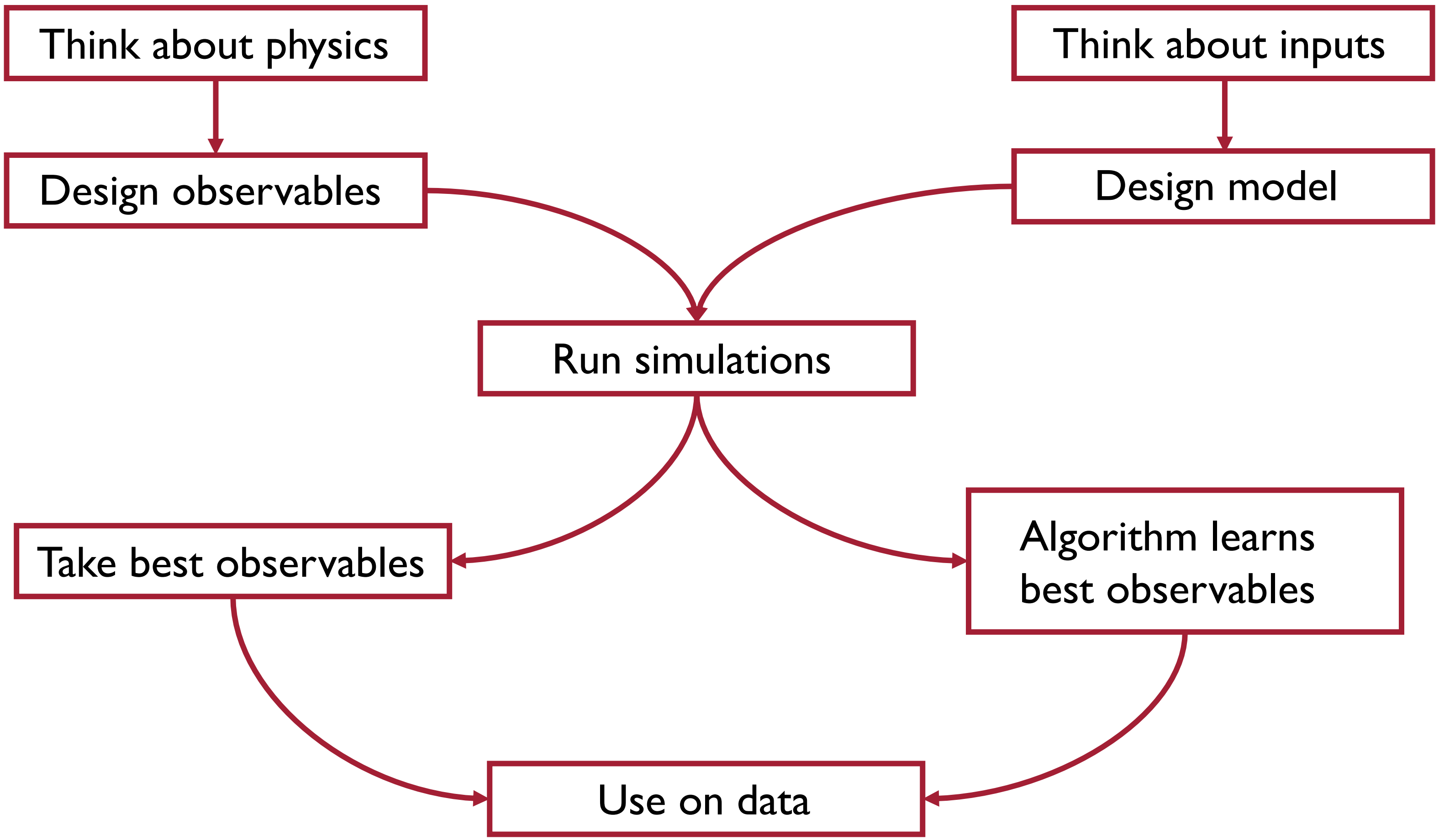
Inherently difficult problem for conventional taggers (both are one-pronged jets)

$\Downarrow$

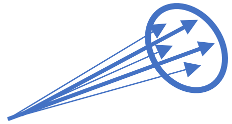
Machine learning to the rescue!

Traditional Approach

Machine Learning Approach



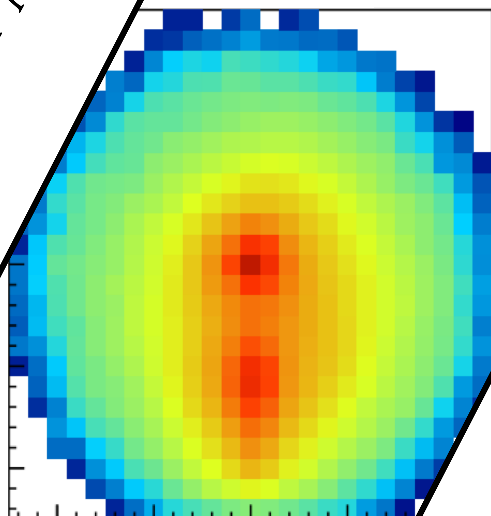
Representing a Jet



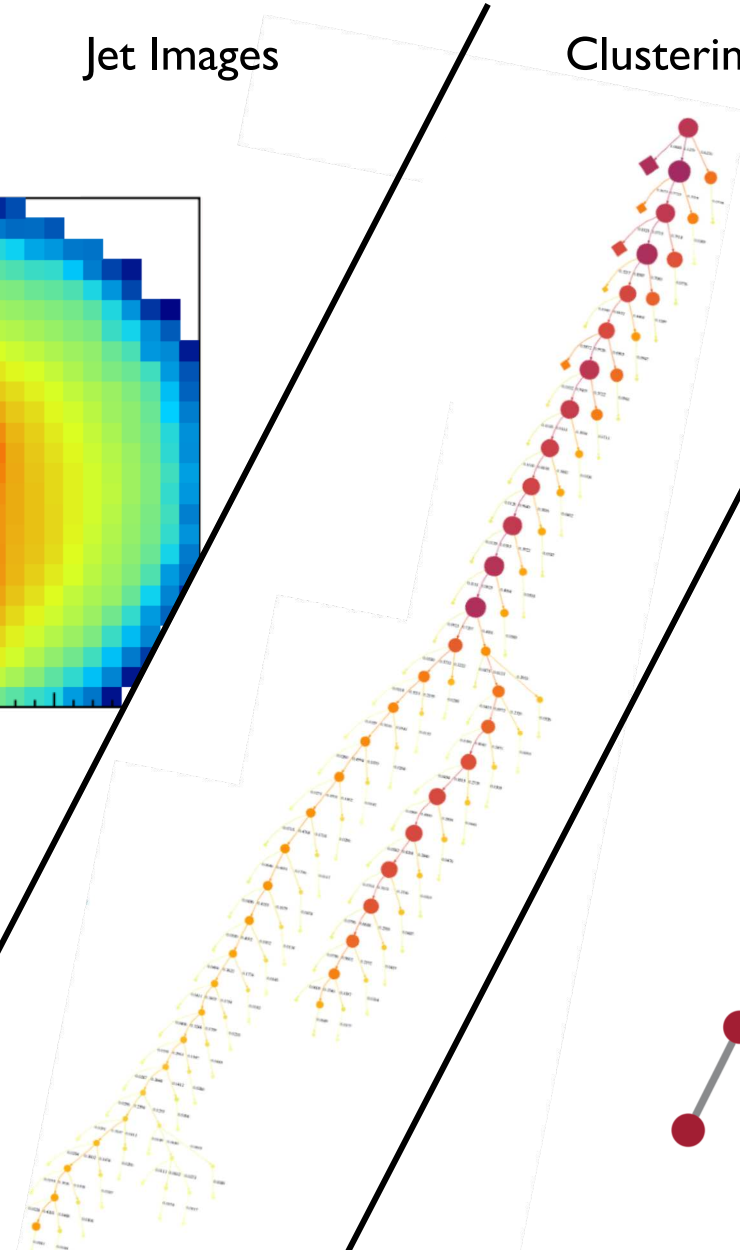
List of Particles

$$\text{Jet} = \{p_1^\mu, p_2^\mu, \ldots, p_M^\mu\}$$

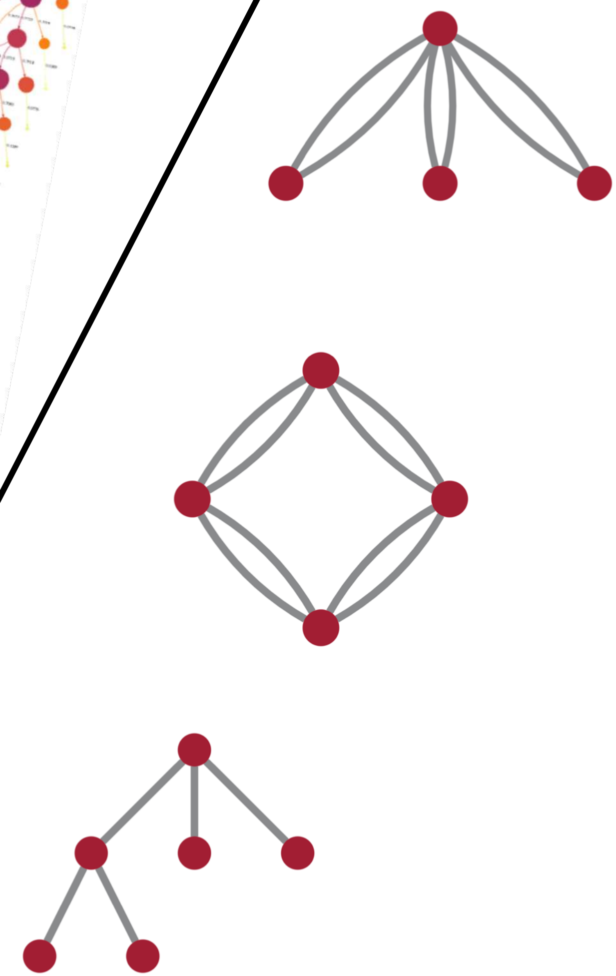
Jet Images



Clustering Trees



Energy Flow



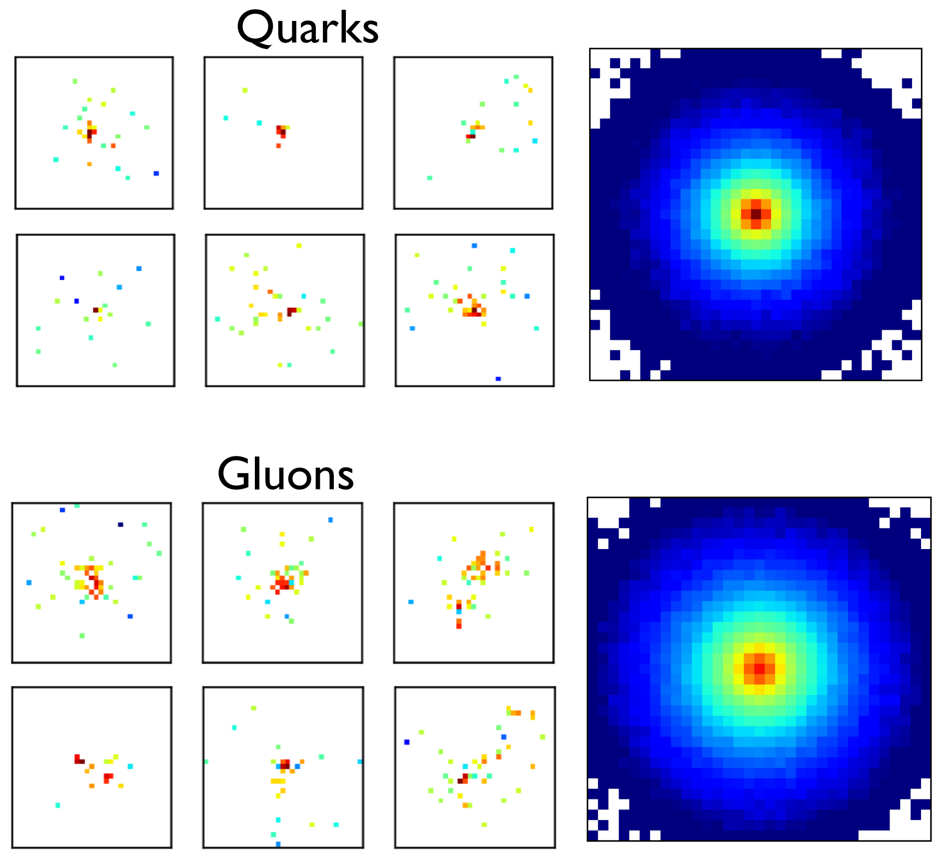
Jet Images

Center on patch of the pseudorapidity-azimuth plane containing a jet

Treat energy/transverse momentum deposits in calorimeter as pixel intensities

Additional input channels possible:

- **Red**: p_T of charged particles
- **Green**: p_T of neutral particles
- **Blue**: charged particle multiplicity

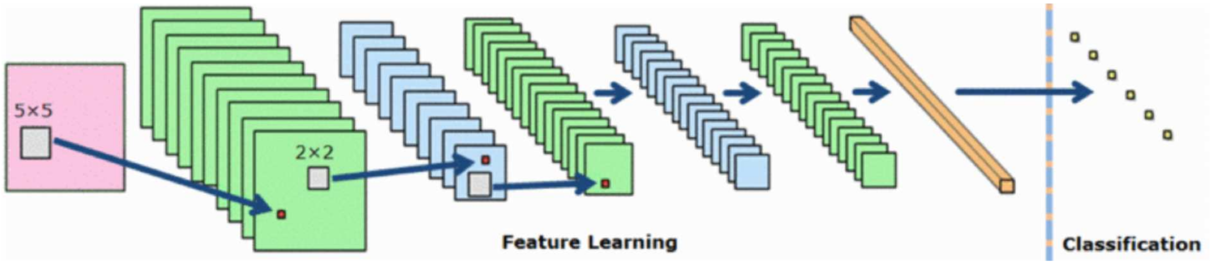


Jet images are sparse

Gluons wider than quarks

Convolutional Neural Networks

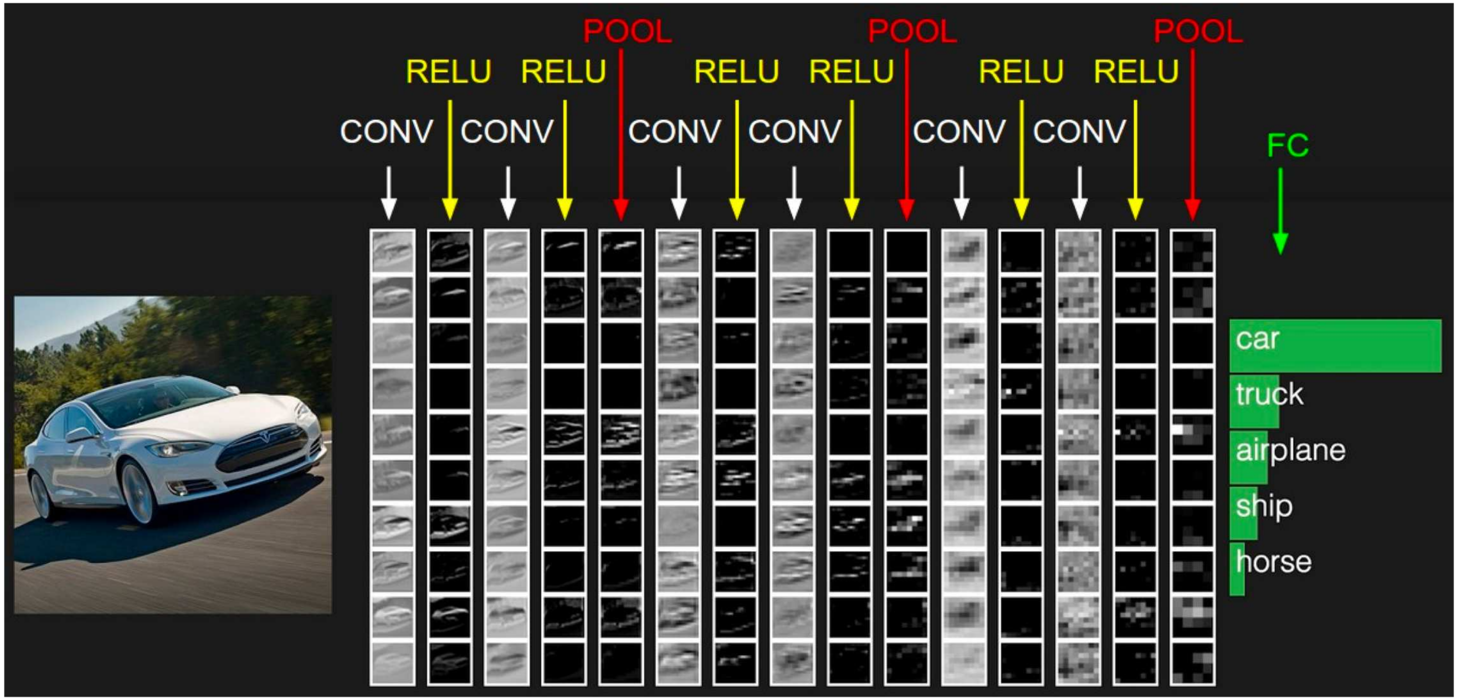
Standard ML method for image classification



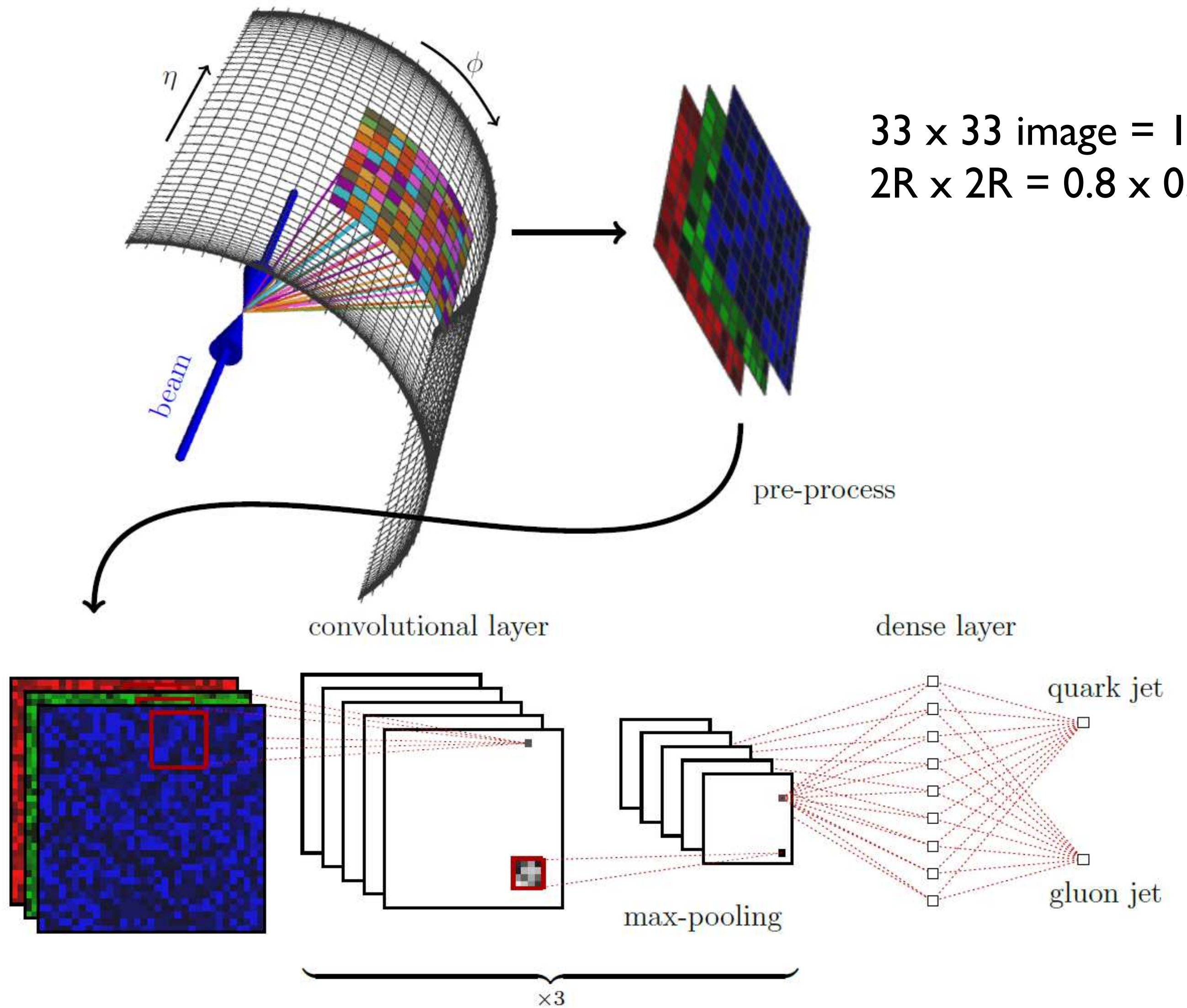
Learns *filters* which extract features

Encodes translation invariance

Natural to use with jet images



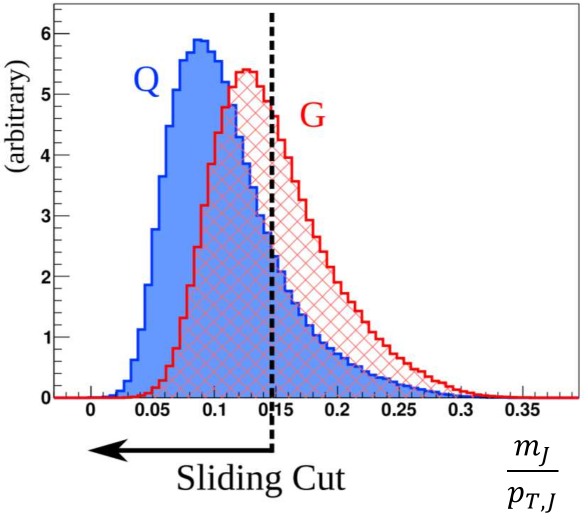
Convolutional Net for QG



33 x 33 image = 1089 inputs
2R x 2R = 0.8 x 0.8 in (y, ϕ)

Quantifying a Classifier

Receiver Operating Characteristic (**ROC**) curve:
True negative rate of the classifier at different true positive rates



ROC curve for Jet Mass

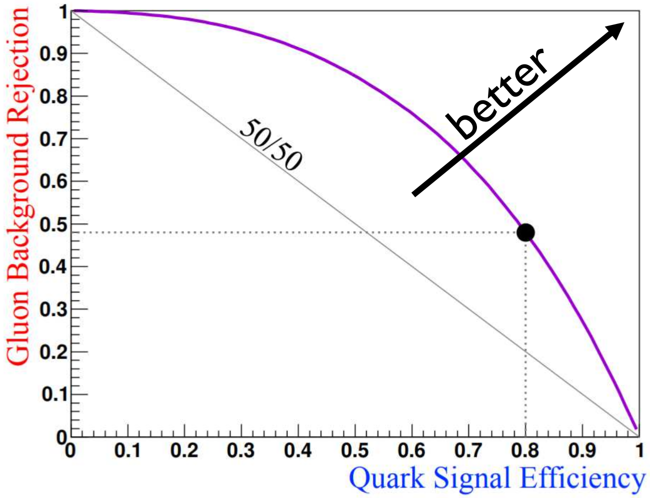
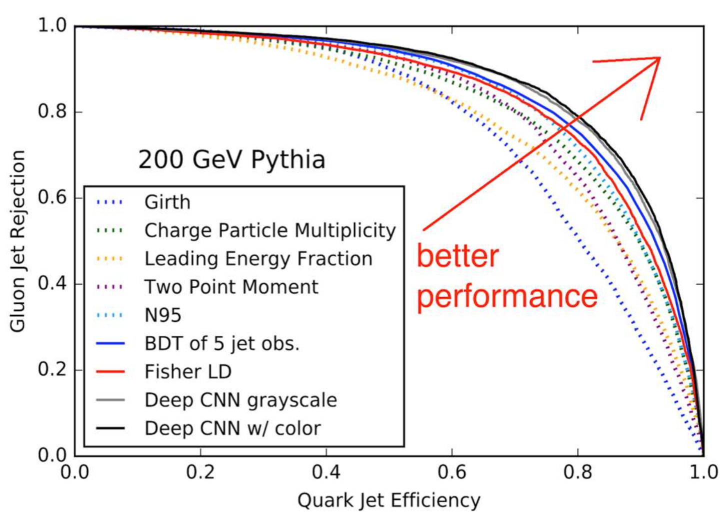
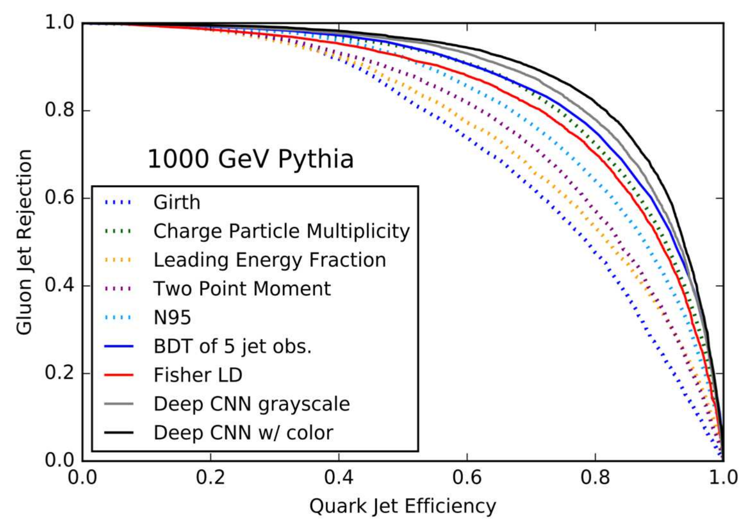


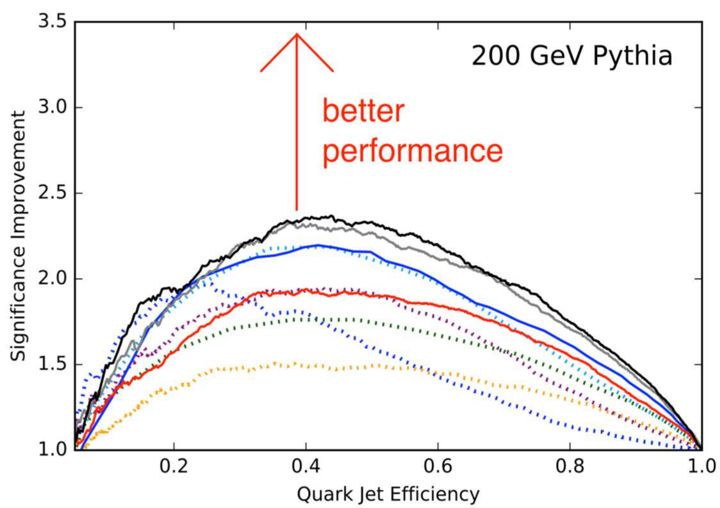
Figure from [1211.7038]

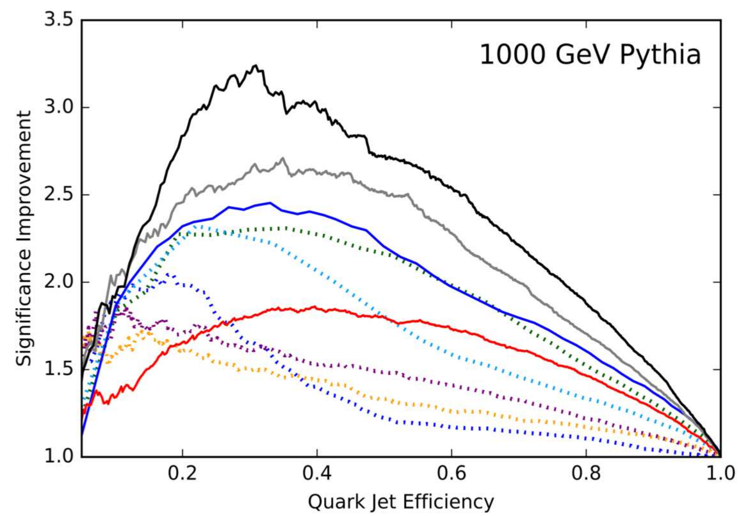
Area Under the ROC Curve (**AUC**) captures the classifier performance in a number.

Classification Performance





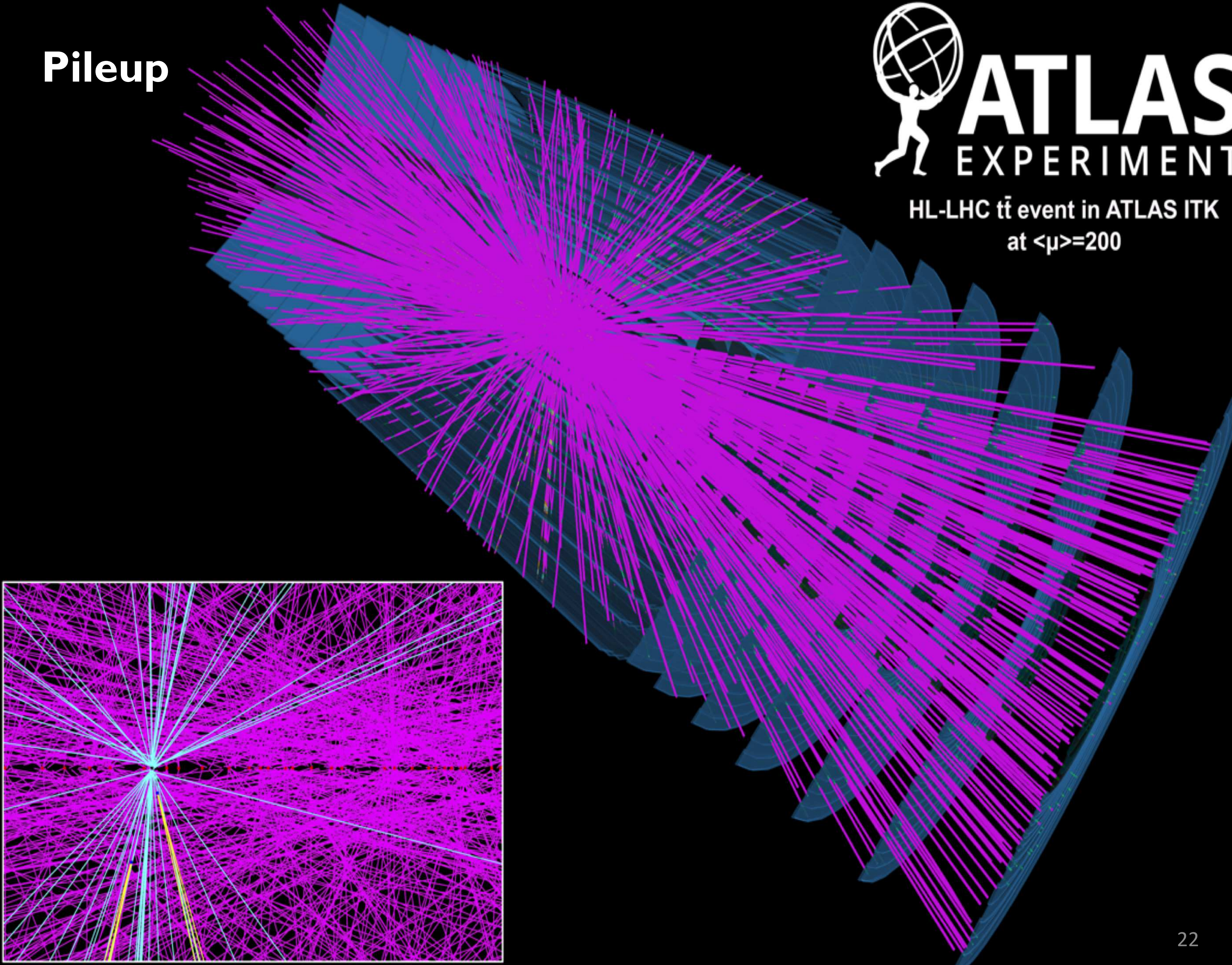




CNN outperforms expert observables!

Multi-channel images help at high p_T

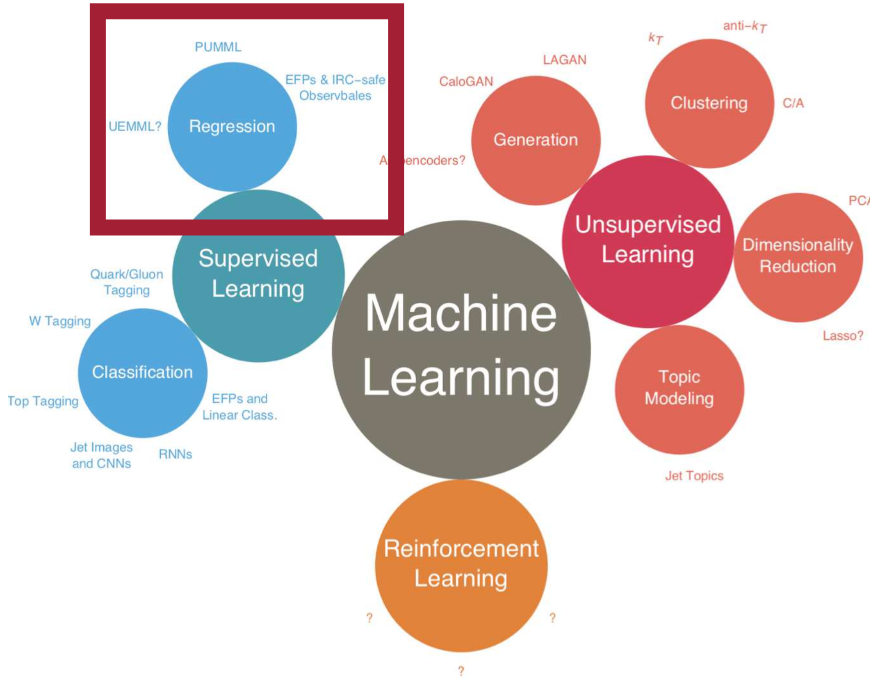
Pileup



HL-LHC $t\bar{t}$ event in ATLAS ITK
at <μ>=200

Pileup Mitigation with Machine Learning (PUMML)

[**PTK**, **EMM**, B. Nachman, M.D. Schwartz, 1707.08600]

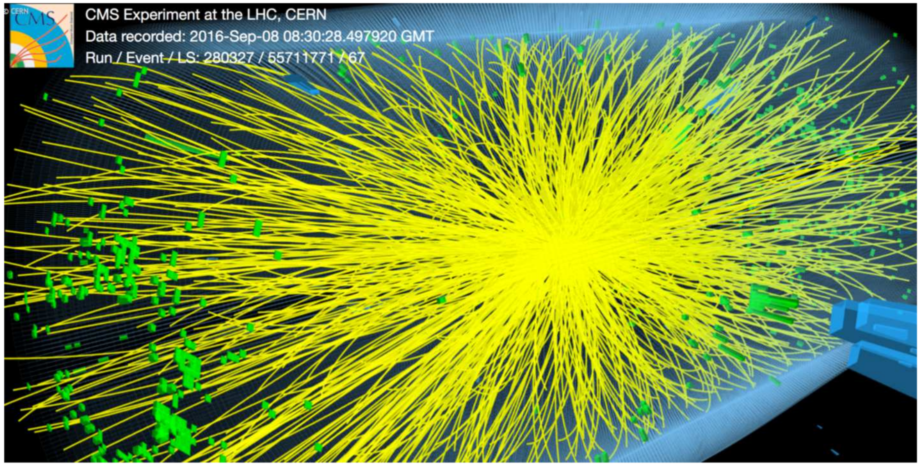


Pileup comes from additional interaction vertices

Soft and uniform (on average) noise

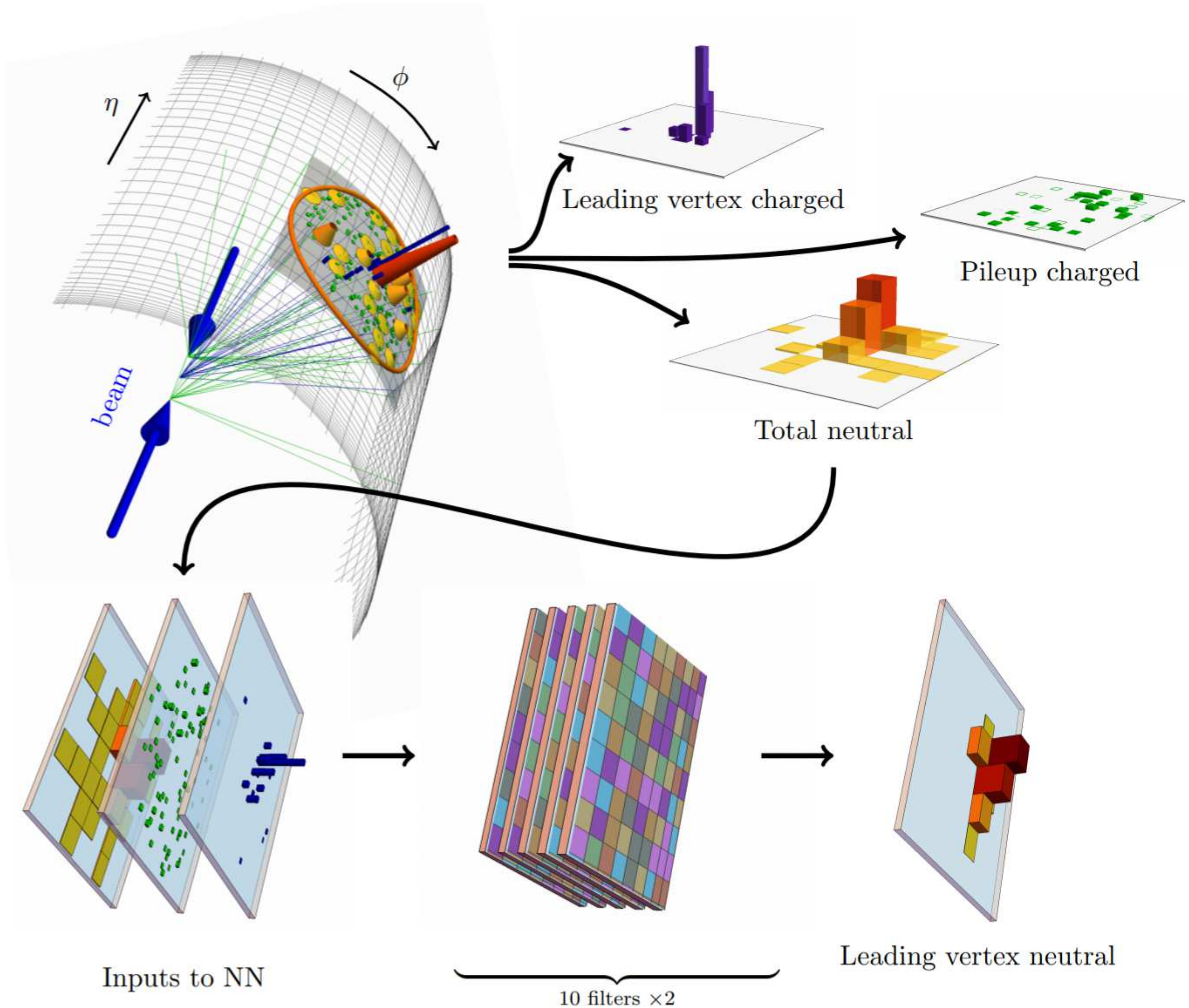
Want to remove pileup to be sensitive to high energy effects

PUMML is first application of regression in particle physics



CMS event with 86 pileup vertices

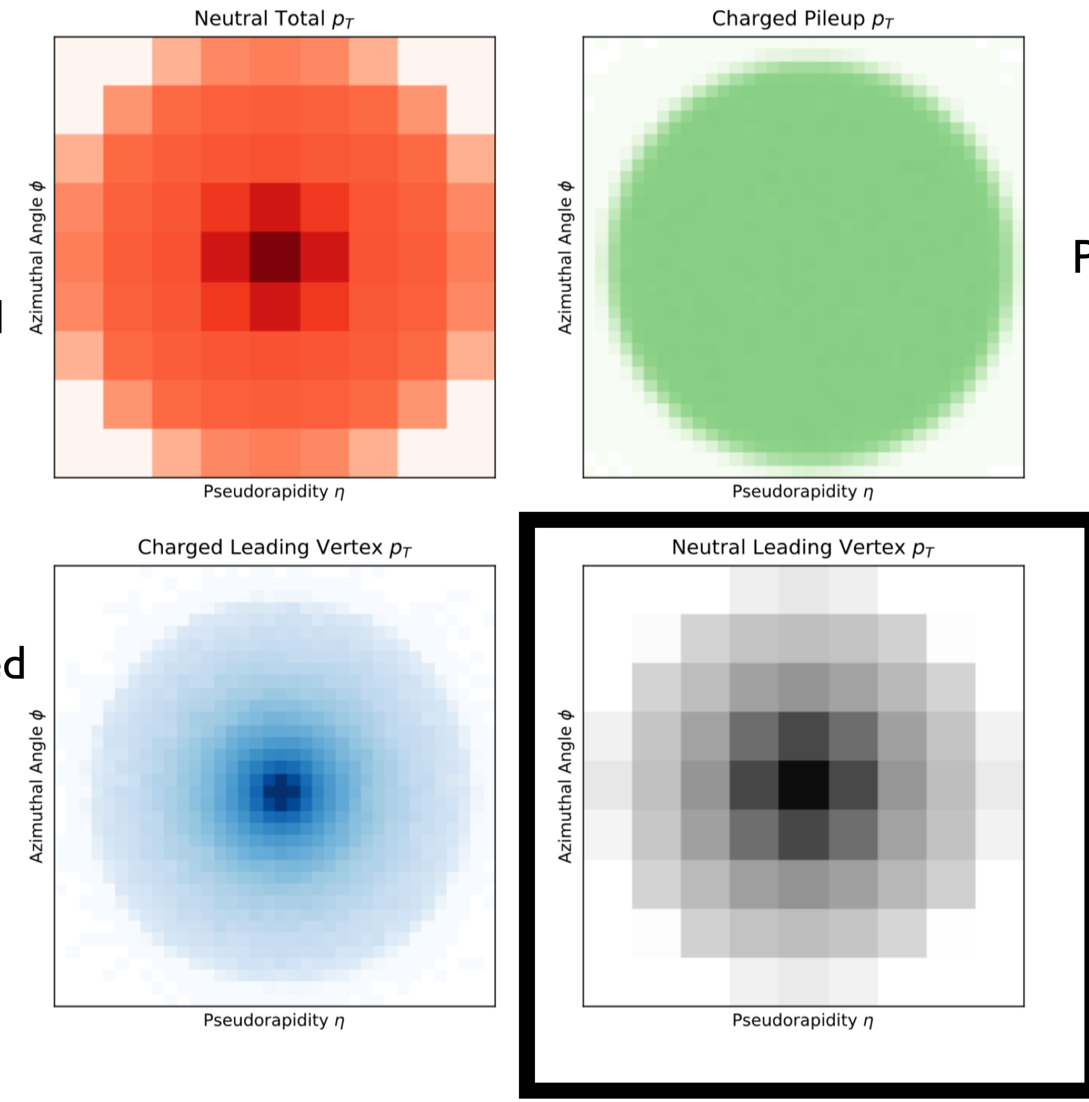
Pileup Mitigation with Machine Learning (PUMML)



Average PUMML Jet Image Inputs

Lower neutral resolution

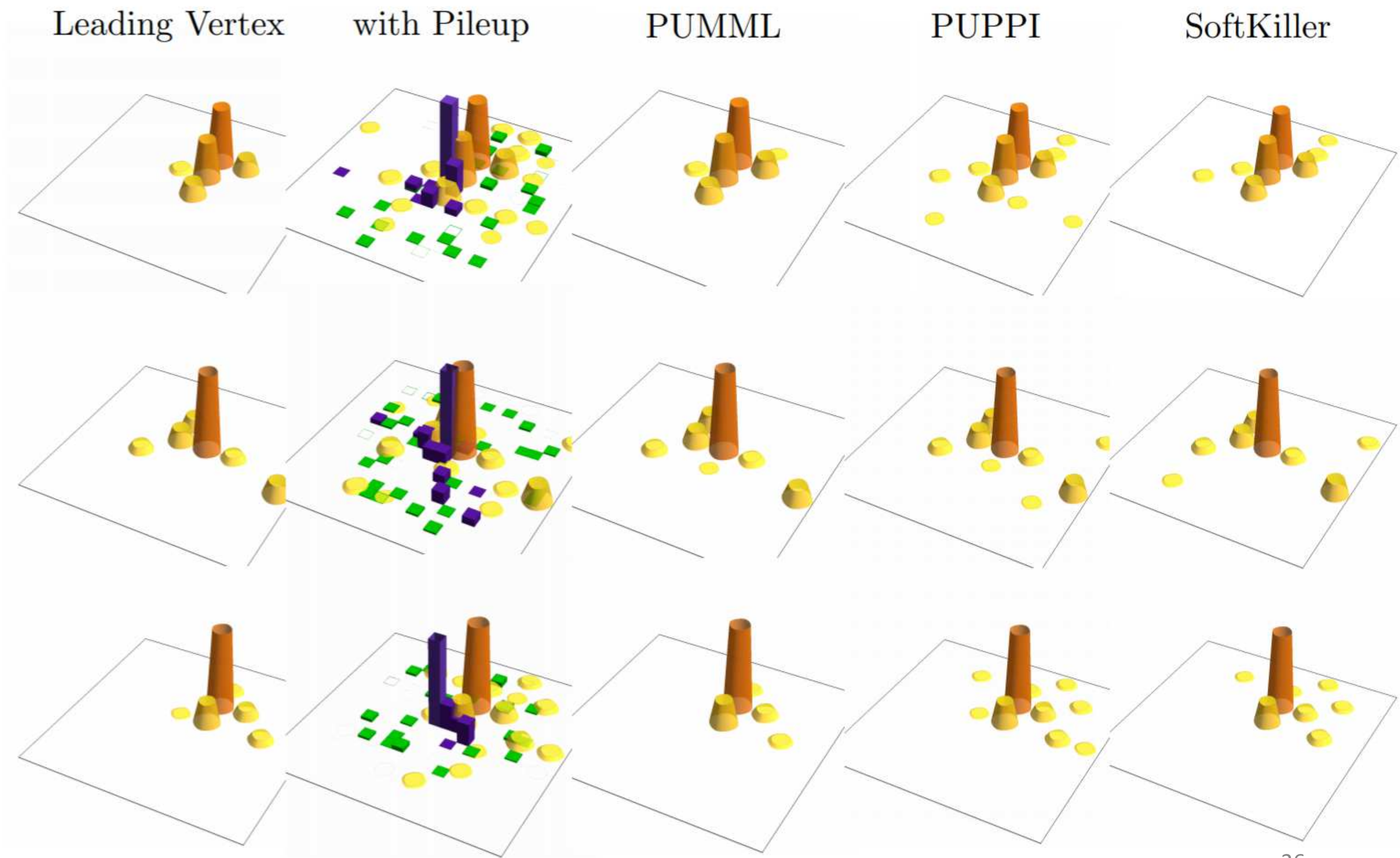
Pileup is uniform



Higher charged resolution

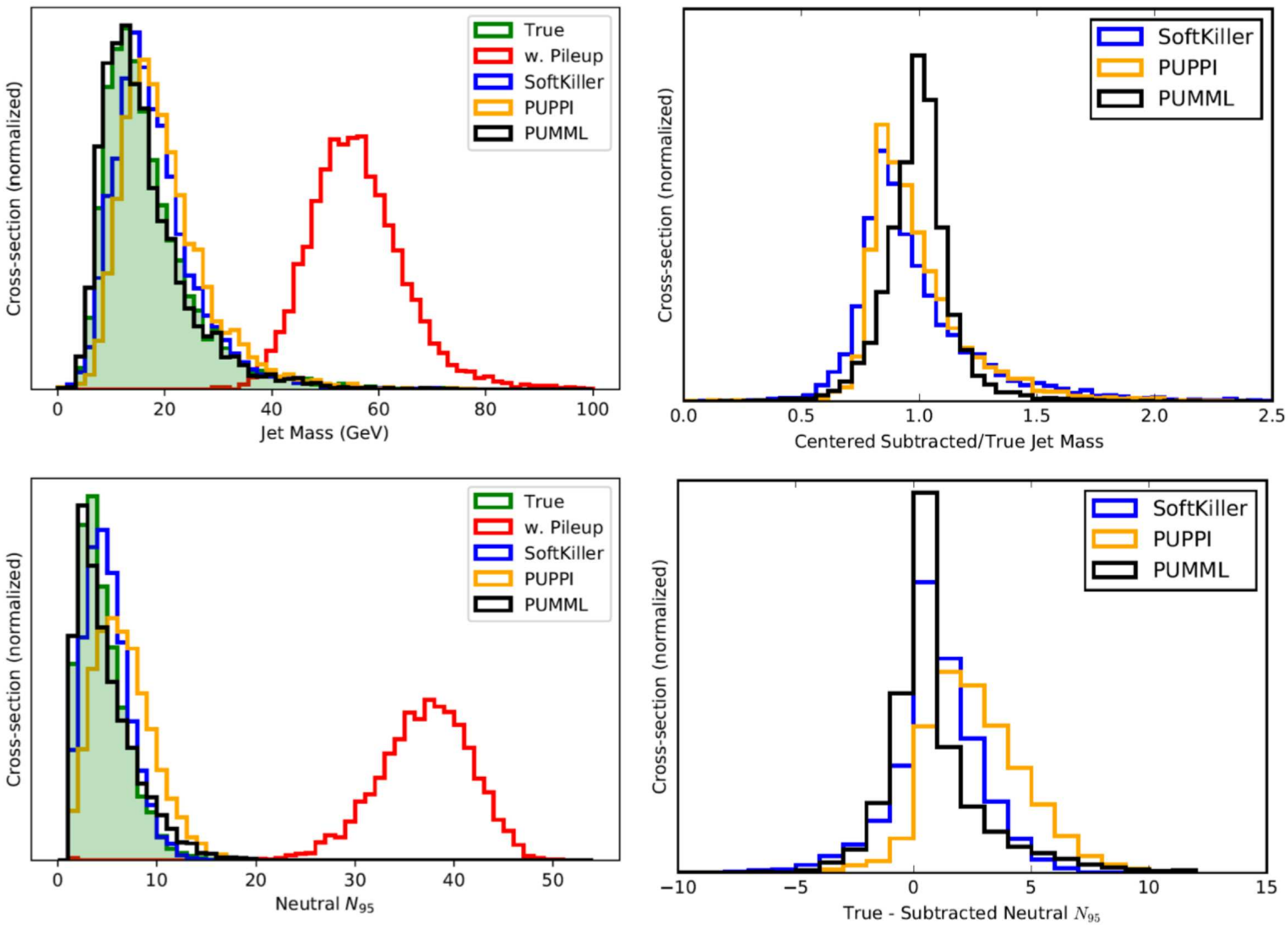
PUMML tries to predict this

Example Pileup Removal Comparisons



Comparison of Pileup Removal Methods

PUMML compares favorably to other existing pileup mitigation methods!



Back to Observables

Jet mass

Angularities

Subjet Count

Multiplicity

N-subjettiness

Geometric Moments

Energy Correlation Functions

What is IRC Safety?

Infrared (IR) safety – observable is unchanged under addition of a soft particle:

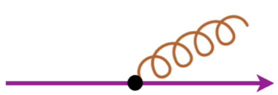
$$S(\{p_1^\mu, \dots, p_M^\mu\}) = \lim_{\epsilon\to 0} S(\{p_1^\mu, \dots, p_M^\mu, \epsilon p_{M+1}^\mu\}), \qquad \forall p_{M+1}^\mu$$

Collinear (C) safety – observable is unchanged under collinear splitting of a particle:

$$S(\{p_1^\mu, \dots, p_M^\mu\}) = \lim_{\epsilon\to 0} S(\{p_1^\mu, \dots, (1-\lambda)p_M^\mu, \lambda p_M^\mu\}), \qquad \forall \lambda \in [0,1]$$

A necessary and sufficient condition for soft/collinear divergences of a QFT to cancel at each order in perturbation theory (KLN theorem)

Divergences can be seen in QCD splitting function:

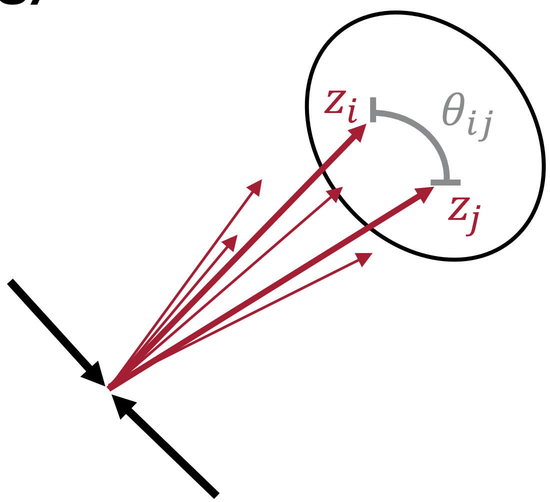


$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{d\theta}{\theta}\frac{dz}{z}$$

$$C_q = C_F = 4/3$$
$$C_g = C_A = 3$$

IRC-safe observables probe high energy structure while being insensitive to low energy modifications

Energy Flow



At the heart is the Energy Flow Operator:

Energy Flow to infinity

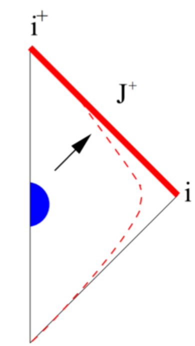
$$\hat{\mathcal{E}}(\hat{n}, v) = \lim_{t\to\infty} \hat{n}_i T^{0i}(t, vt\hat{n})$$

in the $\hat{n}$ direction
at velocity v

[N. Sveshnikov and F. Tkachov, hep-ph/9512370]
[V. Mateu, I.W. Stewart, and J. Thaler, arXiv:1209.3781]

Progress has been made in computing correlations of $\hat{\mathcal{E}}(\hat{n}, v)$ in conformal field theory

[D. Hofman and J. Maldecena, 0803.1467]



IRC-safe observables are built out of energy correlators:

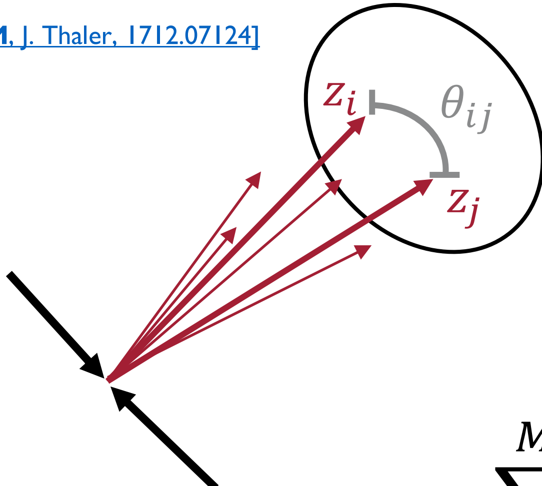
[F. Tkachov, hep-ph/9601308]

Rigid energy structure   Arbitrary angular function f

$$C_f = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} E_{i_1} E_{i_2} \cdots E_{i_N} f(\hat{p}_{i_1}, \cdots, \hat{p}_{i_N})$$

Energy Flow Polynomials (EFPs)

[PTK, EMM, J. Thaler, 1712.07124]



	Energy Fraction	Pairwise Angular Distance
e^+e^-:	$z_i = \frac{E_j}{\sum_k E_k}$,	$\theta_{ij} = \left(\frac{2p_i^\mu p_{j\mu}}{E_i E_j}\right)^{\frac{\beta}{2}}$
Hadronic:	$z_i = \frac{p_{Tj}}{\sum_k p_{Tk}}$,	$\theta_{ij} = \left(\Delta y_{ij}^2 + \Delta\phi_{ij}^2\right)^{\frac{\beta}{2}}$

In equations:

$$\underset{\text{multigraph}}{\mathrm{EFP}_{\mathrm{G}}} = \underbrace{\sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M}}_{\text{Correlator}} \underbrace{z_{i_1} z_{i_2} \cdots z_{i_N}}_{\text{Energies}} \underbrace{\prod_{(k,l)\in G} \theta_{i_k i_l}}_{\text{Angles}}$$

In words:

Correlator of **Energies** and **Angles**

Correlator: Sum over all N-tuples of particle in the event

Energies: Product of the N energy fractions

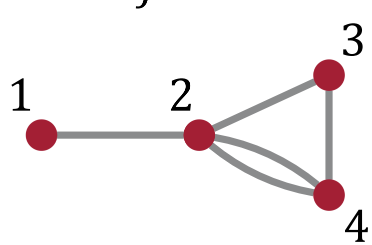
Angles: One $\theta_{i_k i_l}$ for each edge in $(k,l) \in G$

In pictures:

$j \longleftrightarrow z_{i_j}$

k — $l \longleftrightarrow \theta_{i_k i_l}$

(e.g.)



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4}\, \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_3 i_4} \theta_{i_2 i_4}^2$$

(any index labelling works)

Organization of the basis

EFPs *linearly* span all IRC-safe observables!

EFPs are truncated by angular degree *d*, the order of the angular expansion.

Online Encyclopedia of Integer Sequences (OEIS)

A050535 # of multigraphs with *d* edges
# of EFPs of degree *d*

A076864 # of connected multigraphs with *d* edges
# of prime EFPs of degree *d*

Exactly 1000 EFPs up to degree *d*=7!

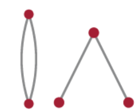
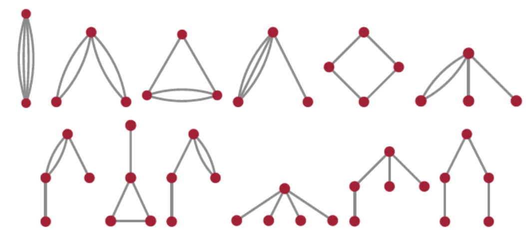
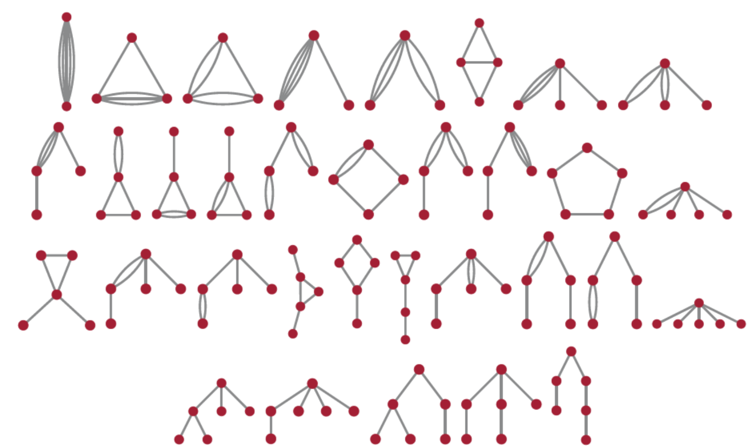
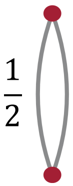
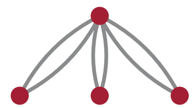
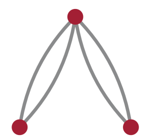
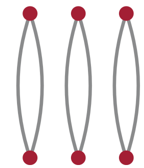
Degree	Connected Multigraphs
$d = 1$	
$d = 2$	
$d = 3$	
$d = 4$	
$d = 5$	

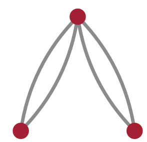
Image files for all of the prime EFP multigraphs up to d = 7 are available here.

Jet Substructure Observables as EFPs

Scaled Jet Mass: $\frac{m_J^2}{p_{TJ}^2} = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} (\cosh \Delta y_{i_1 i_2} - \cos \Delta\phi_{i_1 i_2}) = \frac{1}{2}$  $+ \cdots$

Jet Angularities: $\lambda^{(\alpha)} = \sum_{i}^{M} z_i \theta_i^{\alpha}$

$\lambda^{(6)} =$  $-\frac{3}{2}$   $+\frac{5}{8}$ 

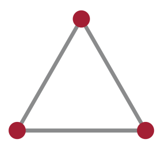
$\lambda^{(4)} =$  $-\frac{3}{4}$ 

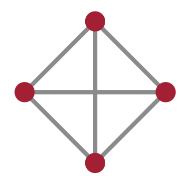
[C. Berger, T. Kucs, and G. Sterman, hep-ph/0303051]
[S. Ellis, *et al.*, arXiv:10010014]
[A. Larkoski, J. Thaler, and W. Waalewijn, arXiv:1408.3122]

Energy Correlation Functions(ECFs): $e_N^{(\beta)} = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} z_{i_2} \cdots z_{i_N} \prod_{k<l \in \{1,\cdots,N\}} \theta_{i_k i_l}^{\beta}$

[A. Larkoski, G. Salam, and J. Thaler, arXiv:1305.0007]

$e_2^{(\beta)} =$ 

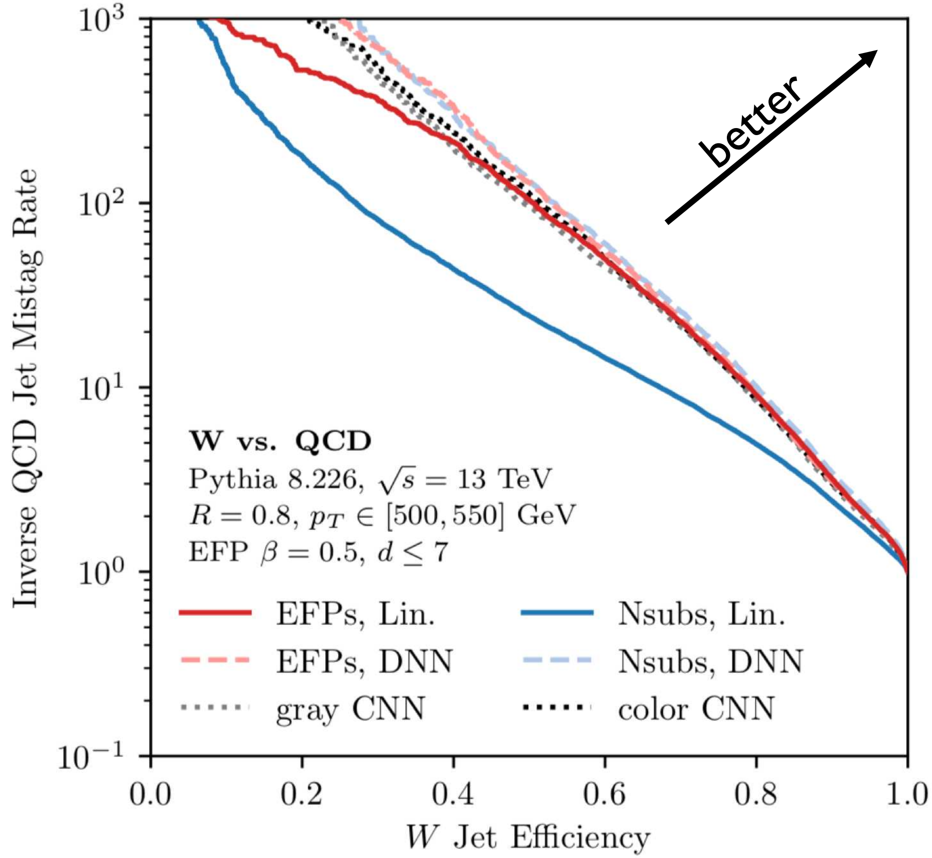
$e_3^{(\beta)} =$ 

$e_4^{(\beta)} =$ 

and many more…

Jet Tagging Comparison

ROC curves for W jet vs. QCD jet tagging

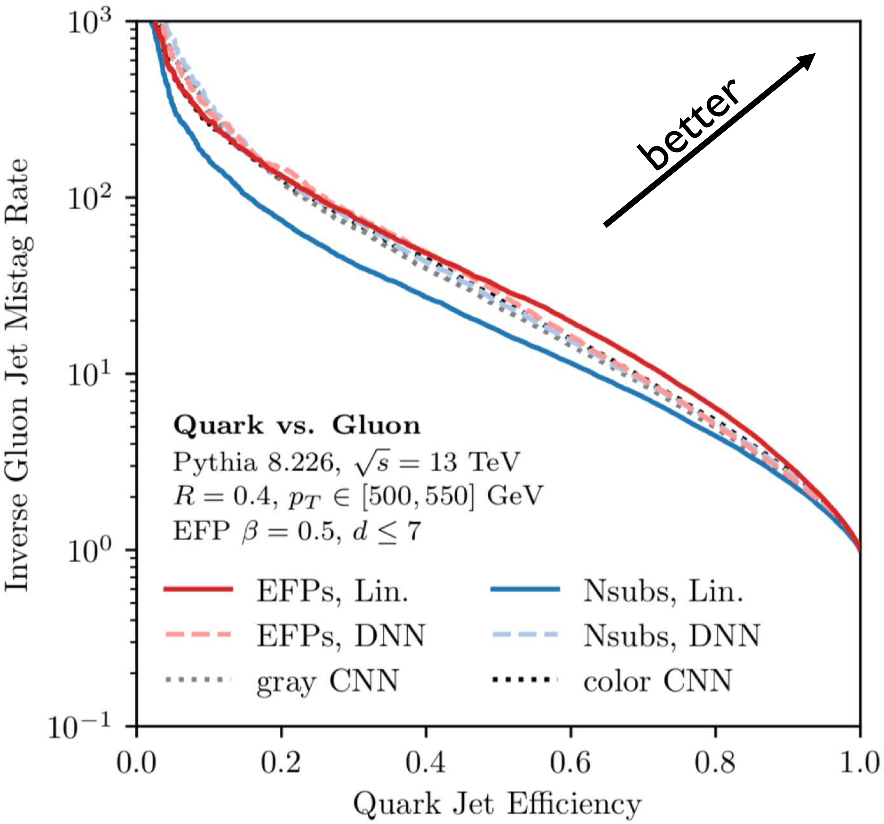


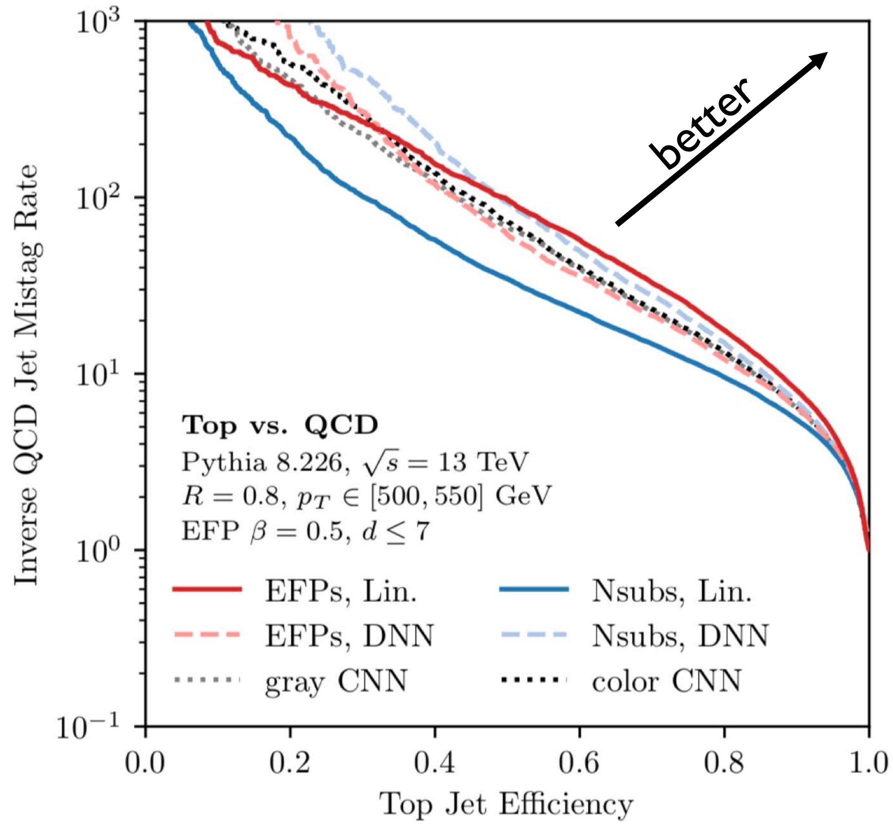
(Linear classification with EFPs) ~ (MML) for efficiency > 0.5!

N-subjettiness: [1011.2268](), *N*-subjettiness basis: [1704.08249](), NN Review: [1709.04464]()

Jet Tagging Comparison

ROC curves for quark vs. gluon tagging and top tagging





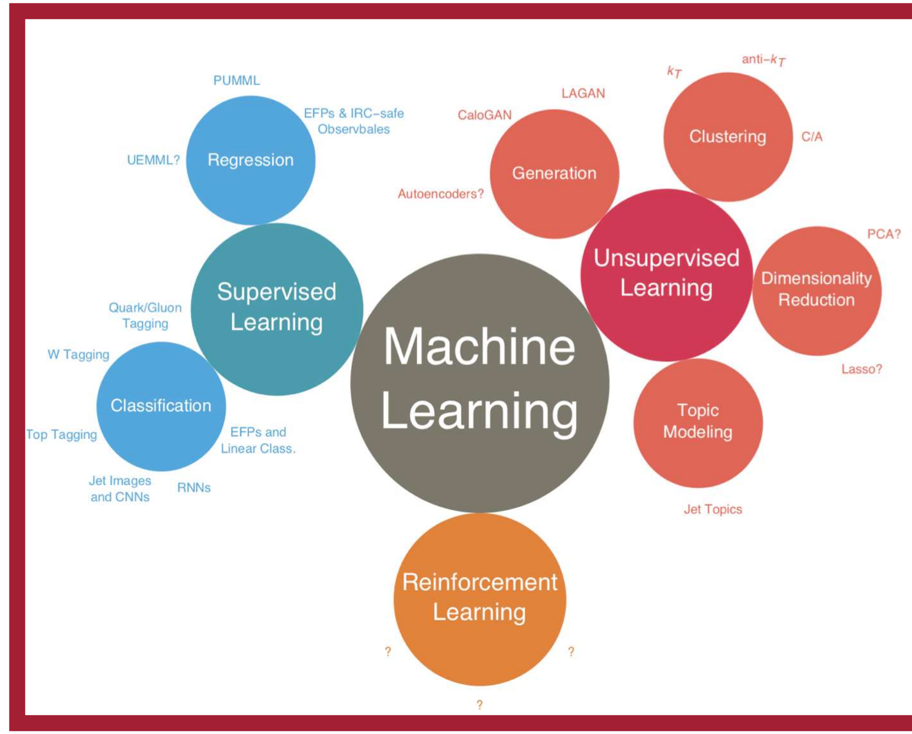
(Linear classification with EFPs) ~ (MML) for efficiency > 0.5!

Escaping the Simulation

Simulation vs. Data

In physics, we usually don't have access to labelled training data.

If we knew which jets were quark and gluon jets… we wouldn't need a tagger!



In collider physics, we usually rely on (imperfect) simulations to provide labelled examples.

PYTHIA

DELPHES fast simulation

Modern machine learning exploits subtle correlations. The simulations do not fully capture all of the complex correlations. Is this a fundamental obstacle to all ML in Physics?

Simulation vs. Data

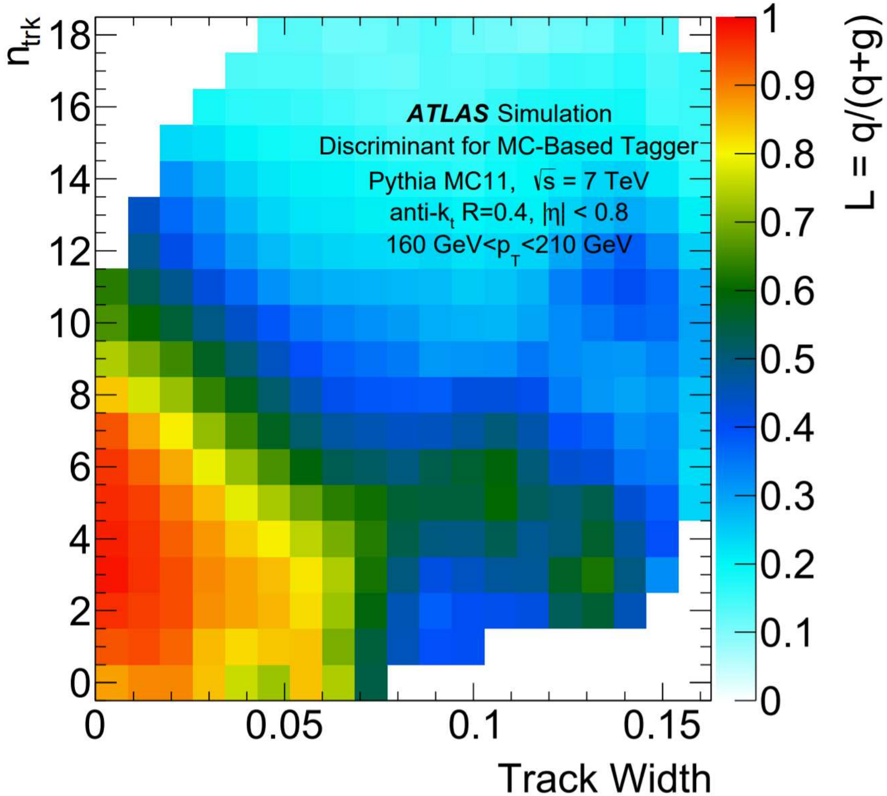
Quark/Gluon Discrimination

Using two features: Width and Number of tracks.

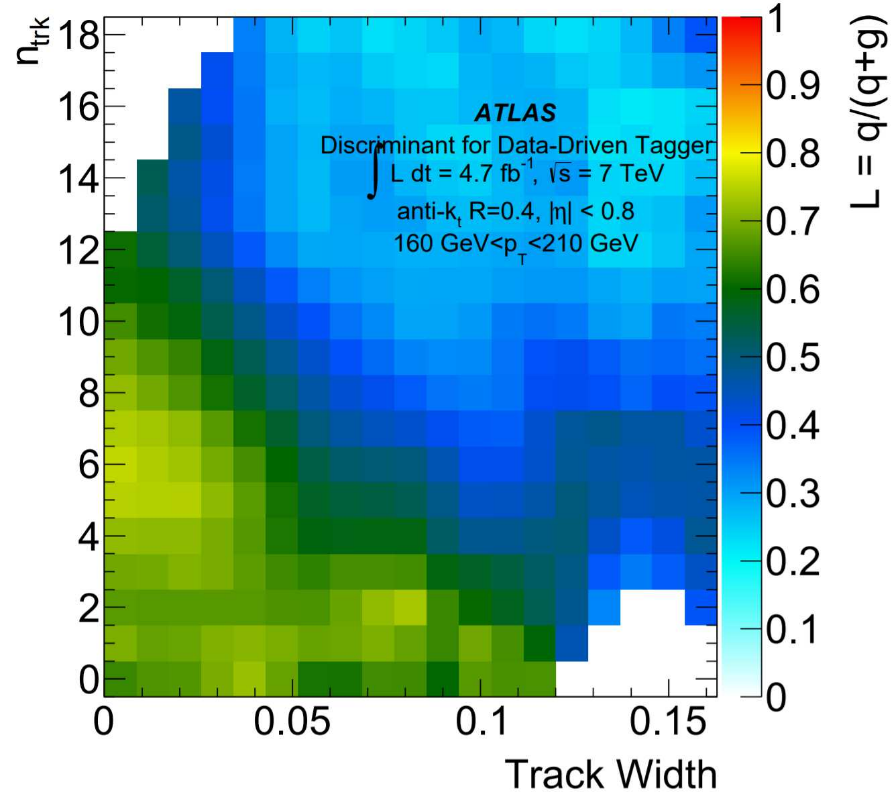
Signal (Q) vs. Background (G) likelihood ratio

[ATLAS Collaboration, arXiv: 1405.6583]

Simulation



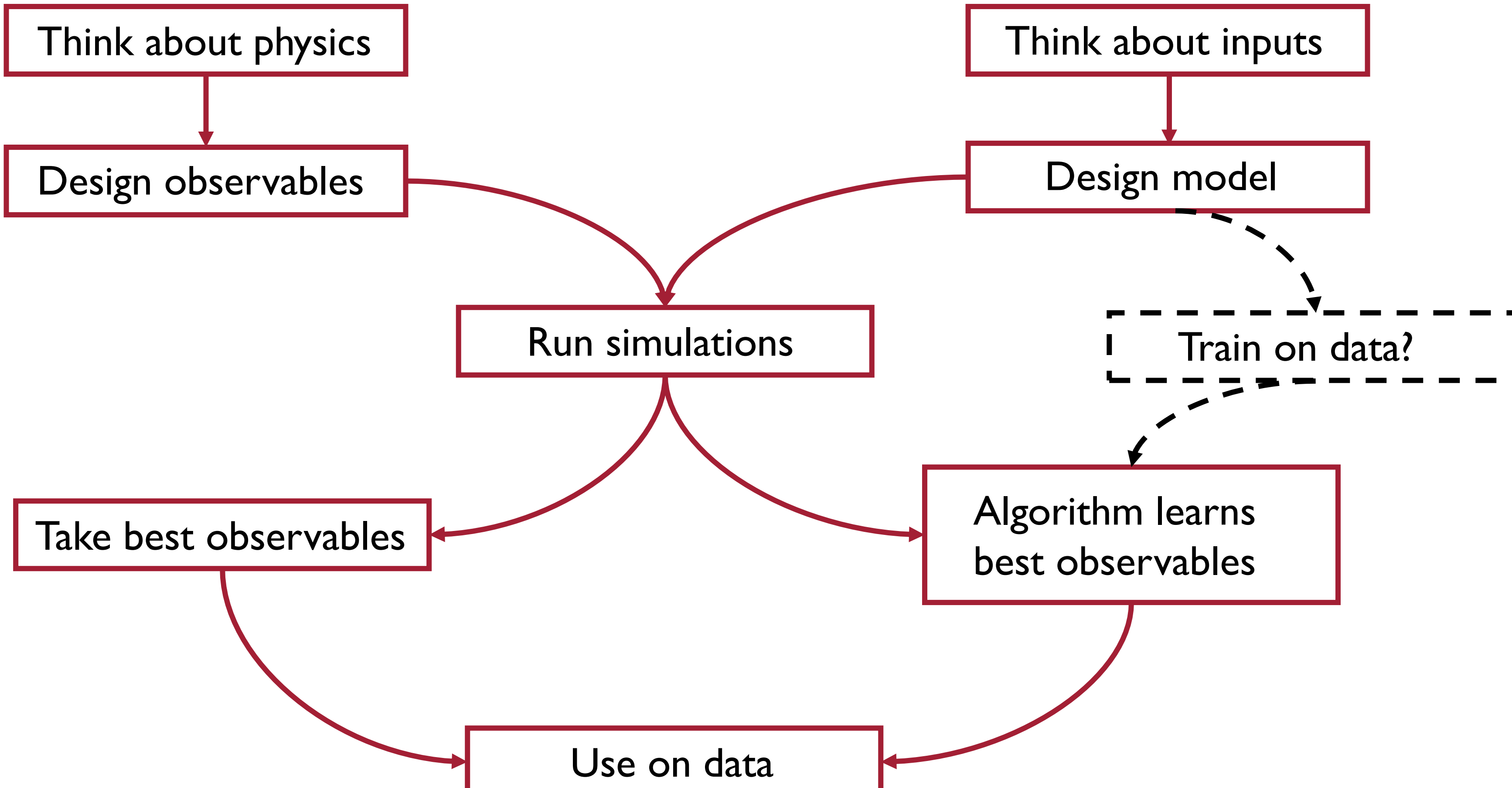
Data



Important differences between simulation and data even for simple observables!

Traditional Approach

Machine Learning Approach



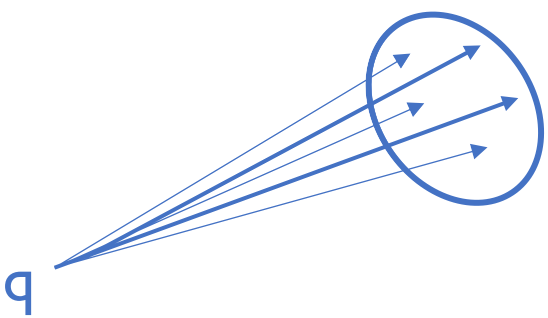
“Physics ML”

This is relatively new territory for Machine Learning.

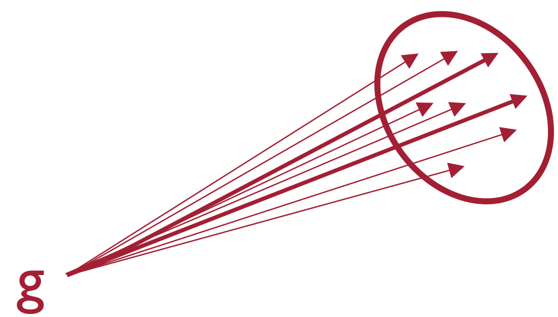
In “Usual ML”: Automate a task that is possible but time consuming for humans (e.g. cat jet vs dog jet).

vs.

In “Physics ML”: Automate a task that is impossible for humans (e.g. quark jet vs gluon jet)



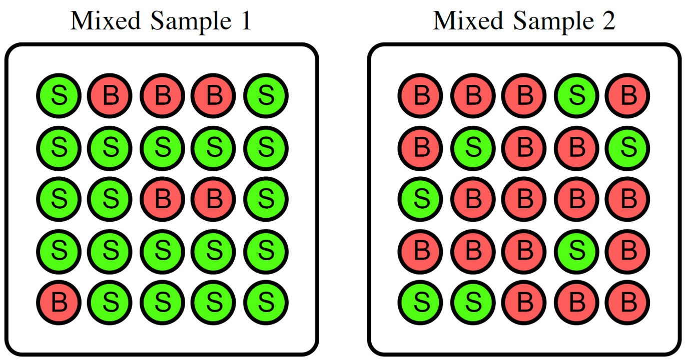
vs.



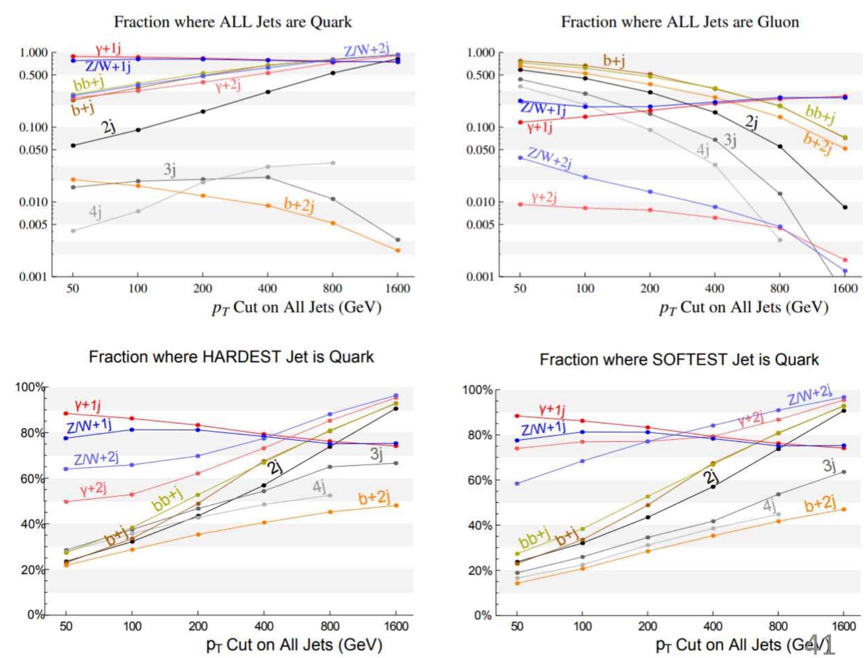
Mixed Samples

Key: Data does not have pure labels, but does have mixed samples!

Some caveats apply. See e.g. [P. Gras, *et al.*, arXiv: 1704.03878]



$$p_{M_a}(x) = f_a\, p_S(x) + (1 - f_a)\, p_B(x)$$

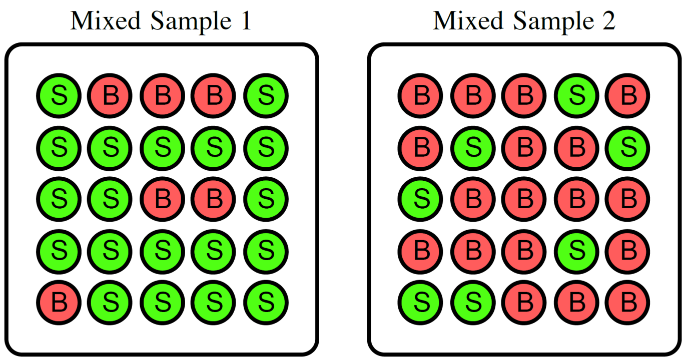


Fractions of quark and gluon jets studied in detail in:
[J. Gallicchio and M.D. Schwartz, arXiv: 1104.1175]

Mixed Samples

Data does not have pure labels, but does have mixed samples!

Some caveats apply. See e.g. [P. Gras, *et al.*, arXiv: 1704.03878]



$$p_{M_a}(x) = f_a\, p_S(x) + (1 - f_a)\, p_B(x)$$

Sample Independence: The *same* signal and background in all the mixtures.

Different Purities: $f_a \neq f_b$ for some a and b.

(Known Fractions): The fractions f_a are known.

Weak Supervision

ML Umbrella term for any classification framework using partial label information.

Collection of supervision models.

Model	References	Description
Full-supervision	[9,24,34,43]	For each example, complete class information is provided.
Unsupervision	[24]	No class information is provided with the examples.
Semi-supervision	[5]	Part of the examples are provided fully supervised. The rest are unsupervised.
Positive-unlabeled	[4,10,21,32]	Part of the examples are provided fully supervised, all of them with the same categorization. The rest are unsupervised.
Candidate labels	[7,13,16]	For each example, a set of class labels is provided. In this set, the class label(s) that compose the real categorization of the example are included.
Probabilistic labels	[18]	For each example, the probability of belonging to each class label is provided. This probability distribution is expected to assign high probability to the real label(s).
Incomplete	[3,33,42]	For each example, a subset of the labels that compose its real categorization is provided (SIML or MIML, Table 1).
Noisy labels	[2,44]	For each example, complete class information is provided, although its correctness is not guaranteed.
Crowd	[30,40]	For each example, many different non-expert annotators provide their (noisy) categorization.
Mutual label constraints	[19,20,31]	For each **group** of examples, an explicit relationship between their class labels is provided (e.g., all the examples have the same categorization).
Candidate labeling vectors	[22]	For each **group** of examples, a set of labeling vectors (including the real one) is provided. A labeling vector provides a class label for each examples of a group.
Label proportions	[15,25,28]	For each **group** of examples, the proportion of examples belonging to each class label is provided.

J. Hernández-González et al. / Pattern Recognition Letters 69 (2016) 49–55

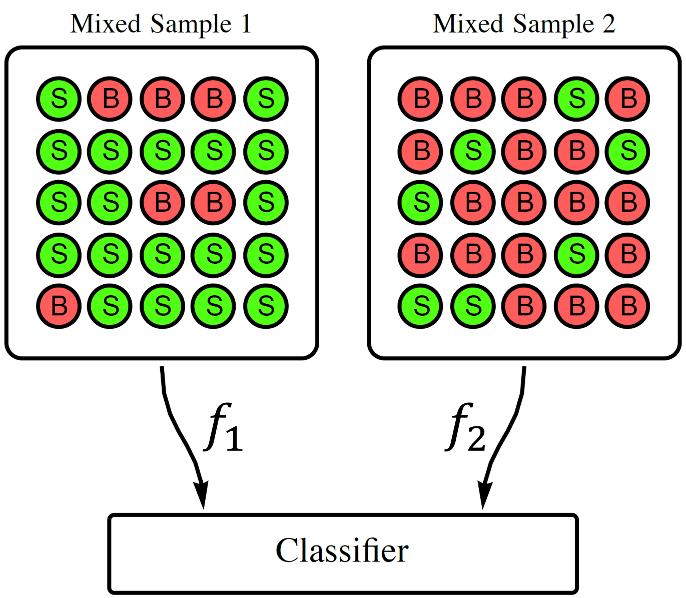
No exact weak supervision framework for the physics (mixture) use-case.

An opportunity to develop new ML tools for the job!

Learning from Label Proportions (LLP) (LoLiProp)

[L. Dery, *et al.*, arXiv: 1702.00414]

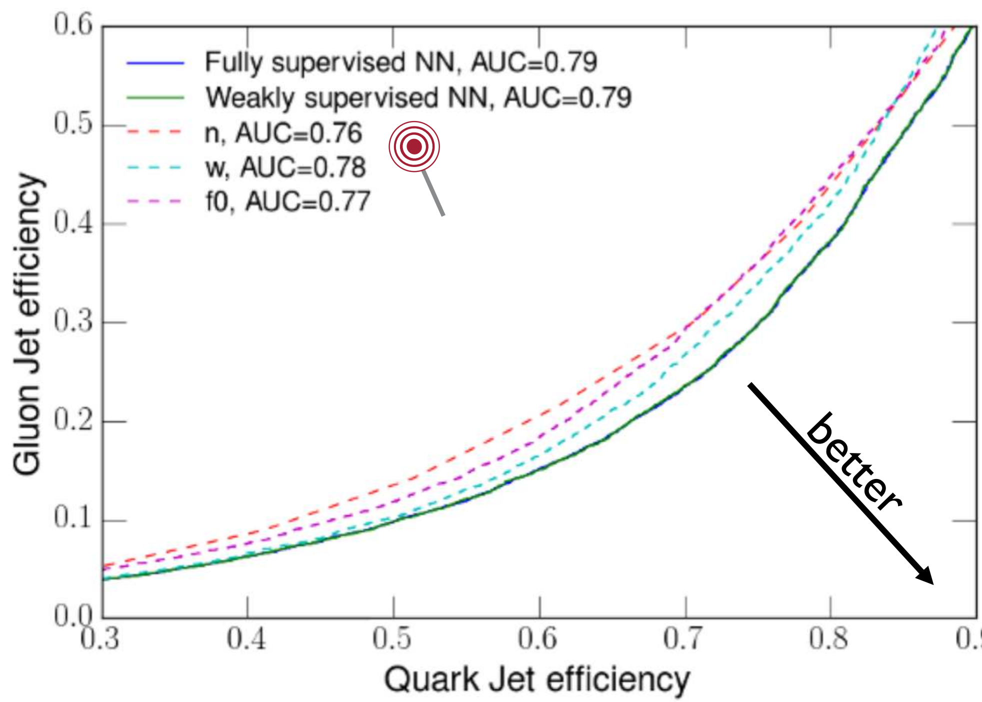
Try to match the signal fractions in aggregate



$$\ell_{\text{LLP}} = \sum_a \ell\left(f_a, \frac{1}{N_a}\sum_{i=1}^{N_a} h(x_i)\right)$$

$\ell_{MSW}, \ell_{CE}, \ldots$

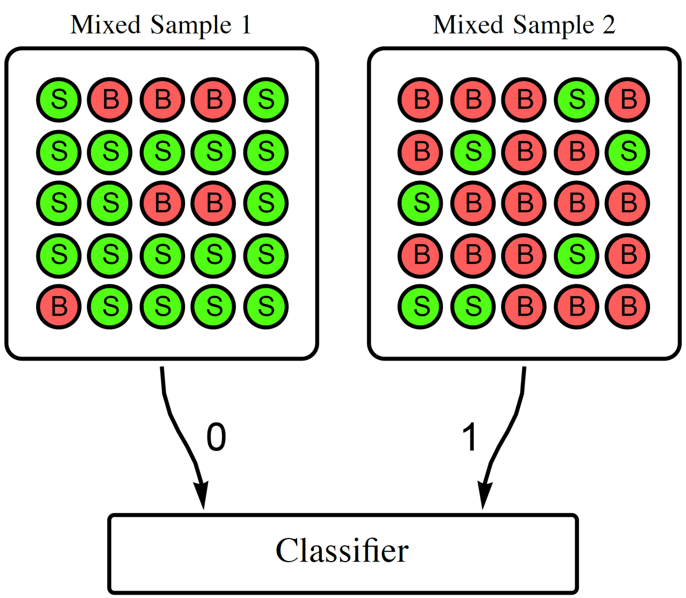
Q/G LLP with **3** inputs works



Classification Without Labels (CWoLa, “koala”)

[**EMM**, B. Nachman, and J. Thaler, arXiv: 1708.02949]
[T. Cohen, M. Freytsis, and B. Ostdiek, arXiv: 1706.09451]
[**PTK**, **EMM**, B. Nachman, and M.D. Schwartz, arXiv: 1801.10158]
See also: [G. Blanchard, M. Flaska, G. Handy, S. Pozzi, and C. Scott, arXiv:1303.1208]

Classify mixed samples from each other

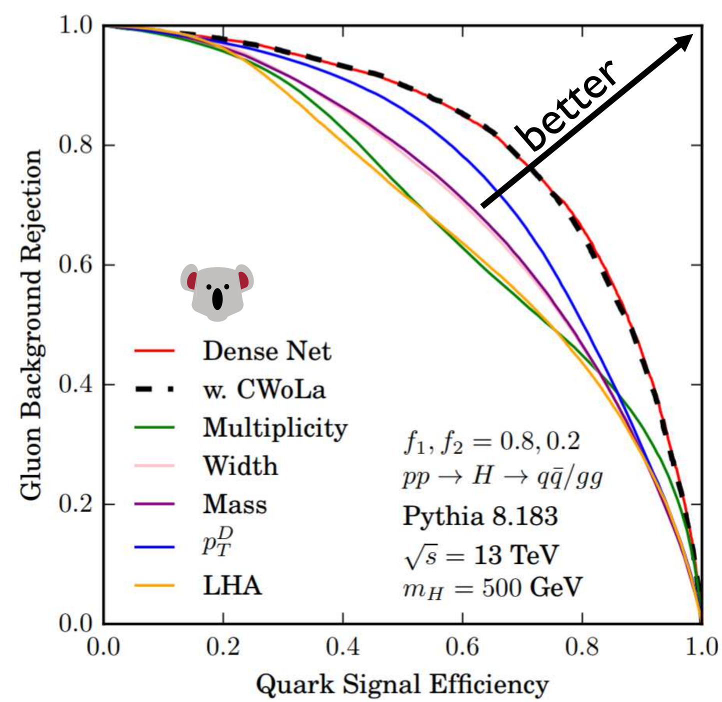


No label proportions needed during training!

Smoothly connected to the fully supervised case as $f_1, f_2 \to 0,1$

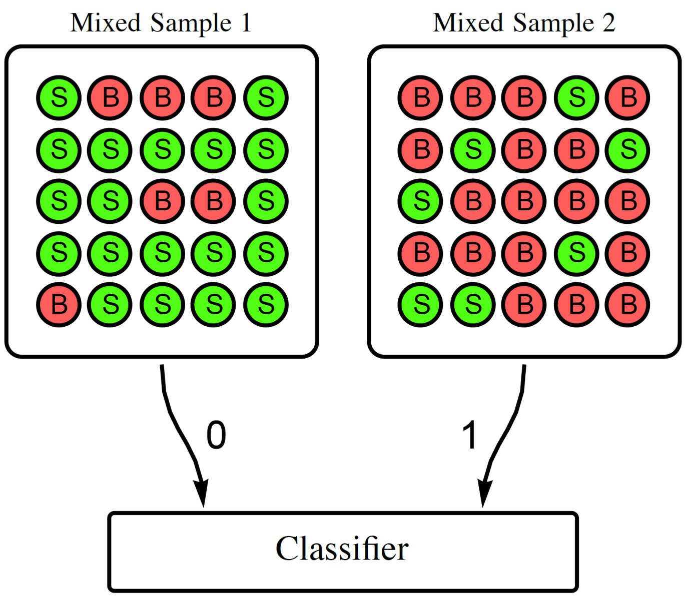
Note: Need small test sets with known signal fractions to determine the ROC.

Q/G WS with **5** inputs works



Classification Without Labels (CWoLa, "koala")

Why does CWoLa work?



Neyman-Pearson Lemma: There is an optimal binary classifier: the likelihood ratio.

$$L_{S/B}(\boldsymbol{x}) = \frac{p_S(\boldsymbol{x})}{p_B(\boldsymbol{x})}.$$

The mixed-sample likelihood ratio is related to the signal/background likelihood ratio by:

$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_2}} = \frac{f_1 p_S + (1-f_1)p_B}{f_2 p_S + (1-f_2)p_B} = \frac{f_1 L_{S/B} + (1-f_1)}{f_2 L_{S/B} + (1-f_2)}.$$

This is a monotonic rescaling of the signal/background likelihood ratio!

Therefore Mixture 1 vs. Mixture 2 and Signal vs. Background define the same classifier. They have the same ROC curves.

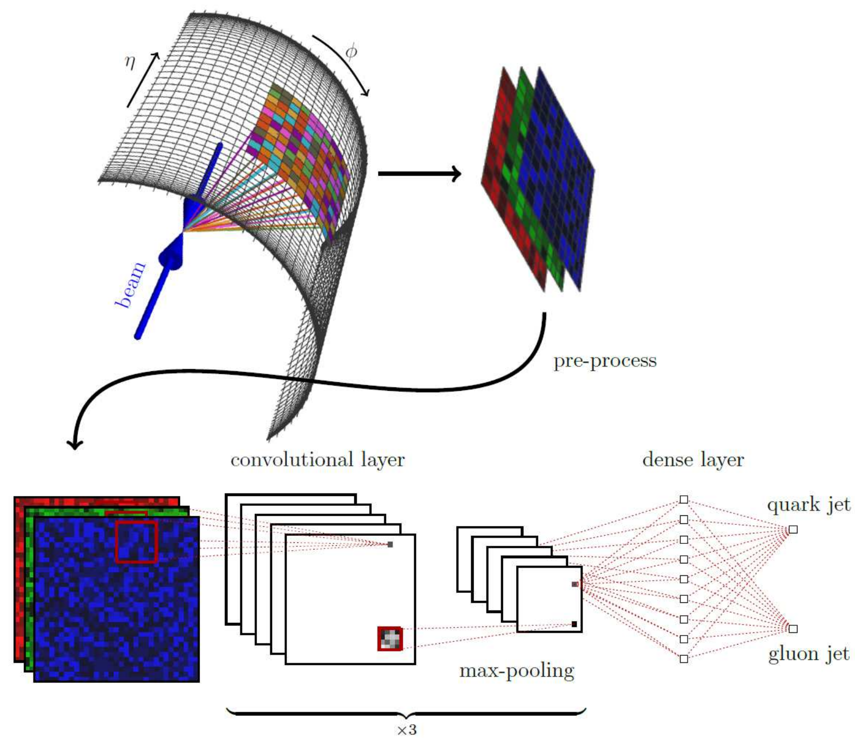
Learning to Classify from Impure Samples

[**PTK**, **EMM**, B. Nachman, and M.D. Schwartz, arXiv: 1801.10158]

CWoLa and LLP have been shown to work for simple architectures and small inputs.

Can these weak supervision methods be used for real deep learning applications in collider physics? Are they ready for the big leagues?

To answer this question, we did our quark/gluon tagging with jet images using only *mixtures* of quarks and gluons – *no labels*.

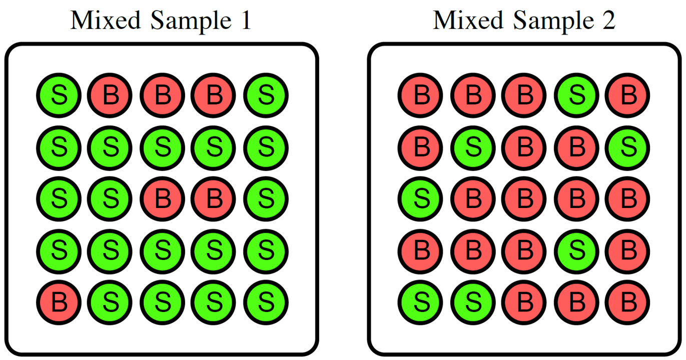


Short answer: CWoLa generalizes very well
LLP needs tuning, but it works

Potential to train on data!

Purity and Number of Data

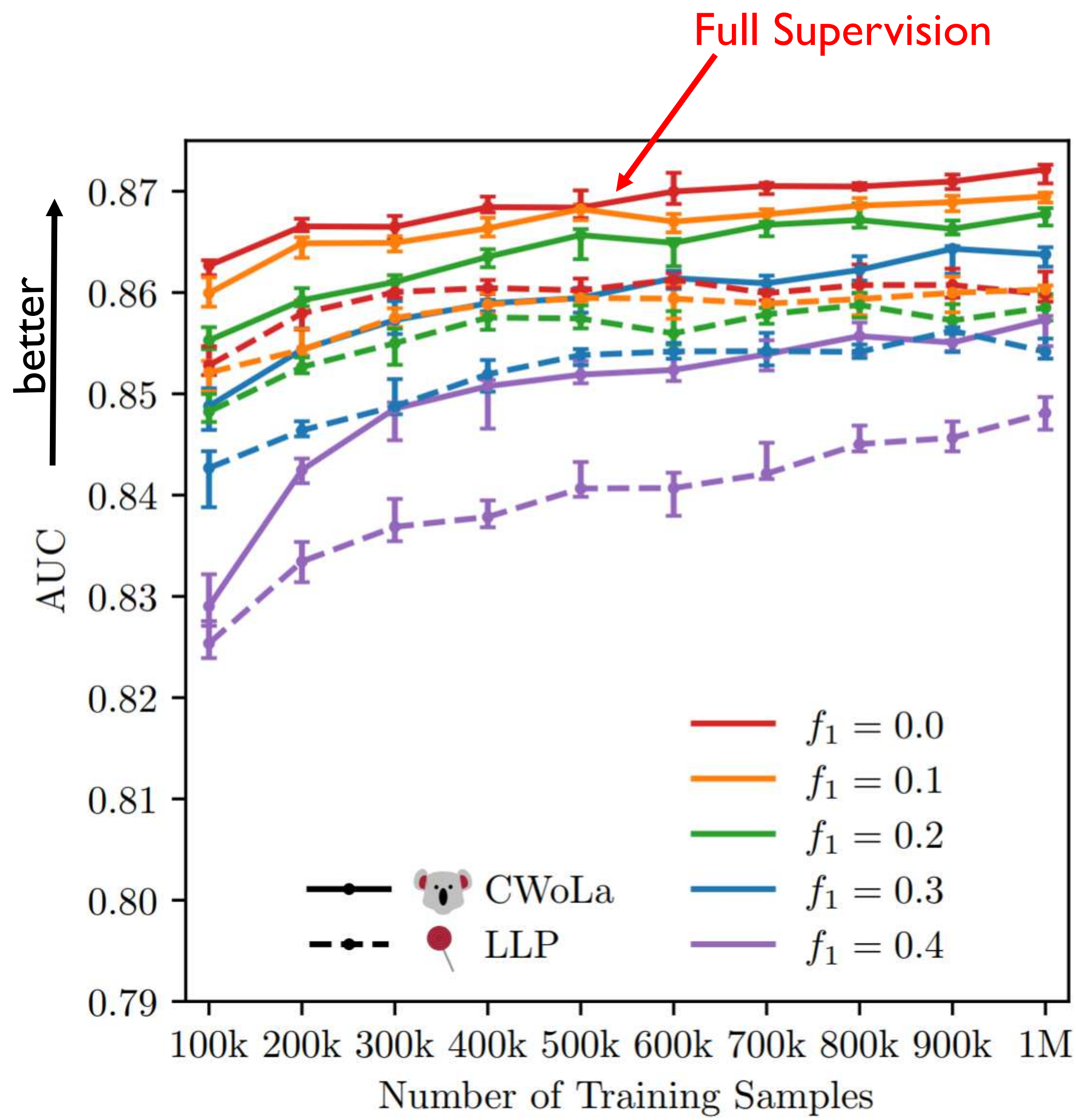
Two mixed samples: $f_1, 1 - f_1$



Purity/Data plot can characterize tradeoffs in a weak learning method

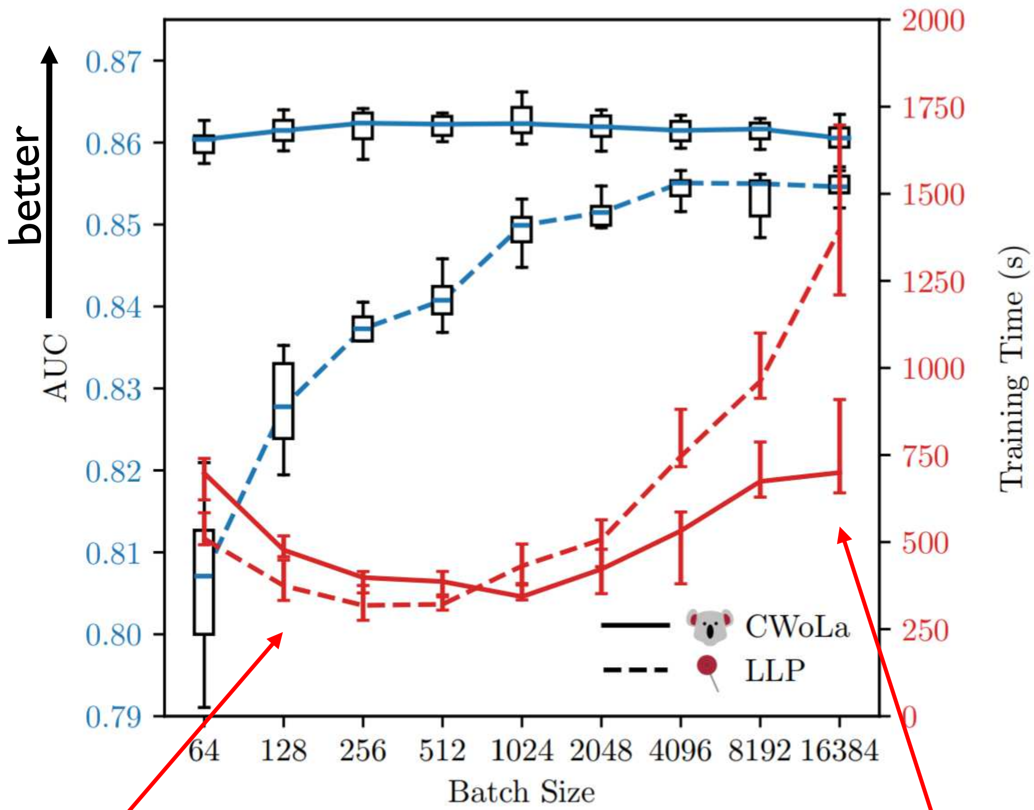
CWoLa performs near full supervision if the samples are relatively pure.

LLP lags behind but still achieves good classification performance.



Batch Size and Training Time

We explored hyperparameters, training times, and other lessons from using the methods in practice.



time/epoch increases

# of epochs increas

Batch size

As usual for CWoLa

Need large batch size for LLP

Batch Size > 1000

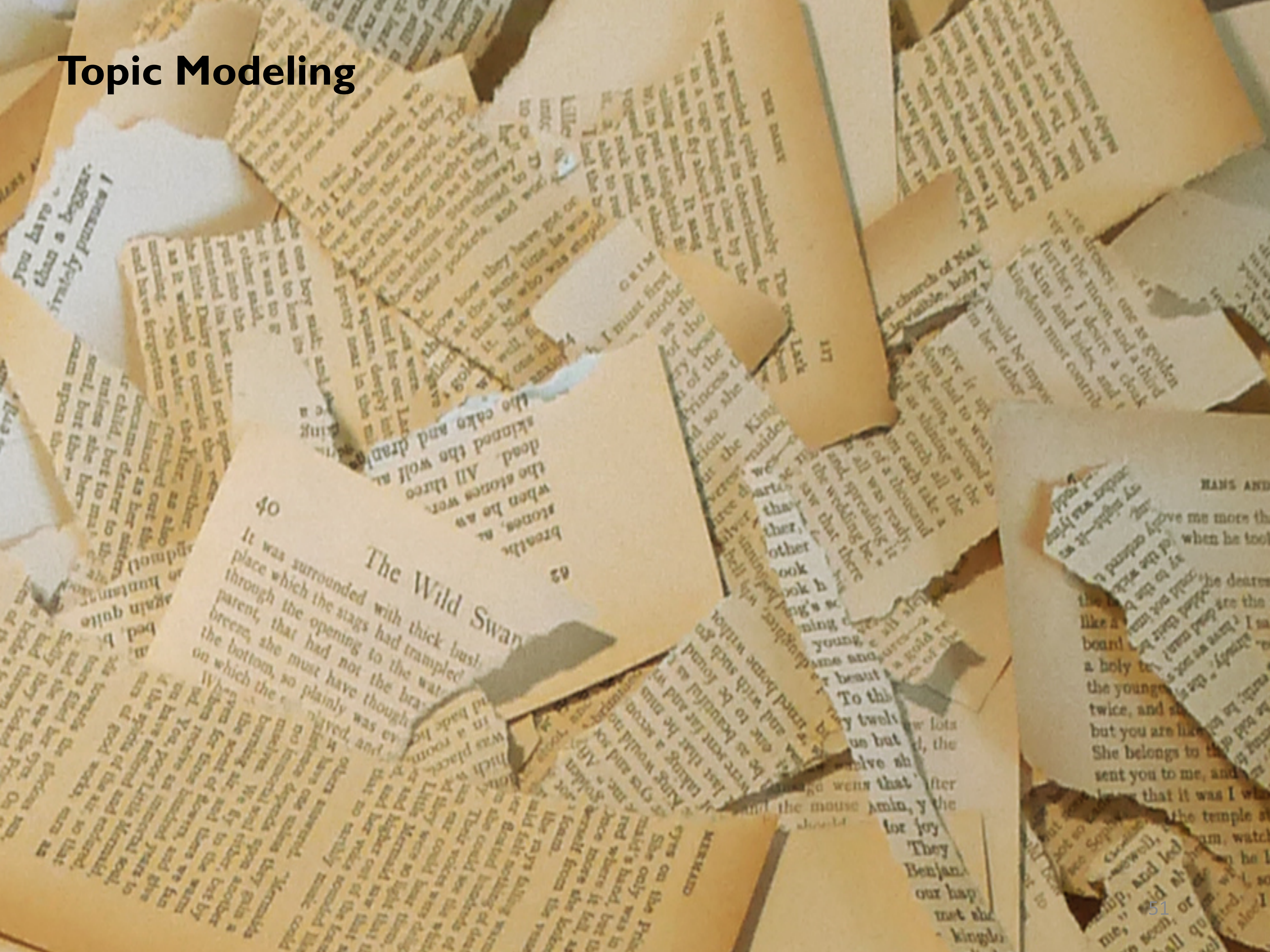
$$\ell_{\text{LLP}} = \sum_{a} \ell\left(f_a, \frac{1}{N_a}\sum_{i=1}^{N_a} h(x)\right)$$

Weak Supervision in Summary

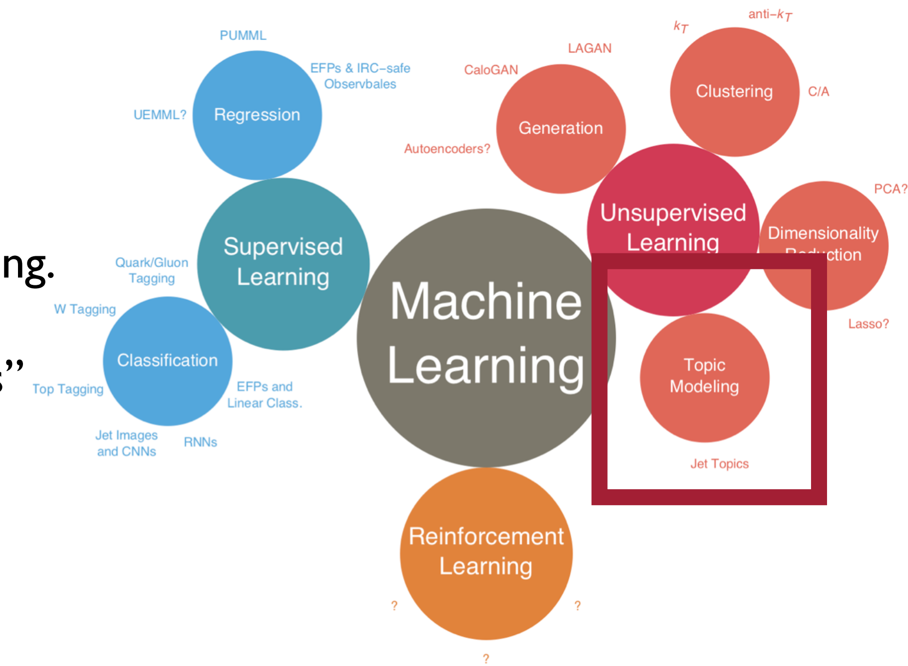
We now have two candidate methods to train ML algorithms directly on jet data!

Property	 **LLP**	 **CWoLa**
No need for fully-labeled samples	✓	✓
Compatible with any trainable model	✓	✓
No training modifications needed	✗	✓
Training does not need fractions	✗	✓
Smooth limit to full supervision	✗	✓
Works for > 2 mixed samples	✓	?

Topic Modeling



Topic Modeling



A statistical model from natural language processing.

Used to discover the emergent themes or "topics" in a collection of documents or "corpus".

A Topic Model View of the World:

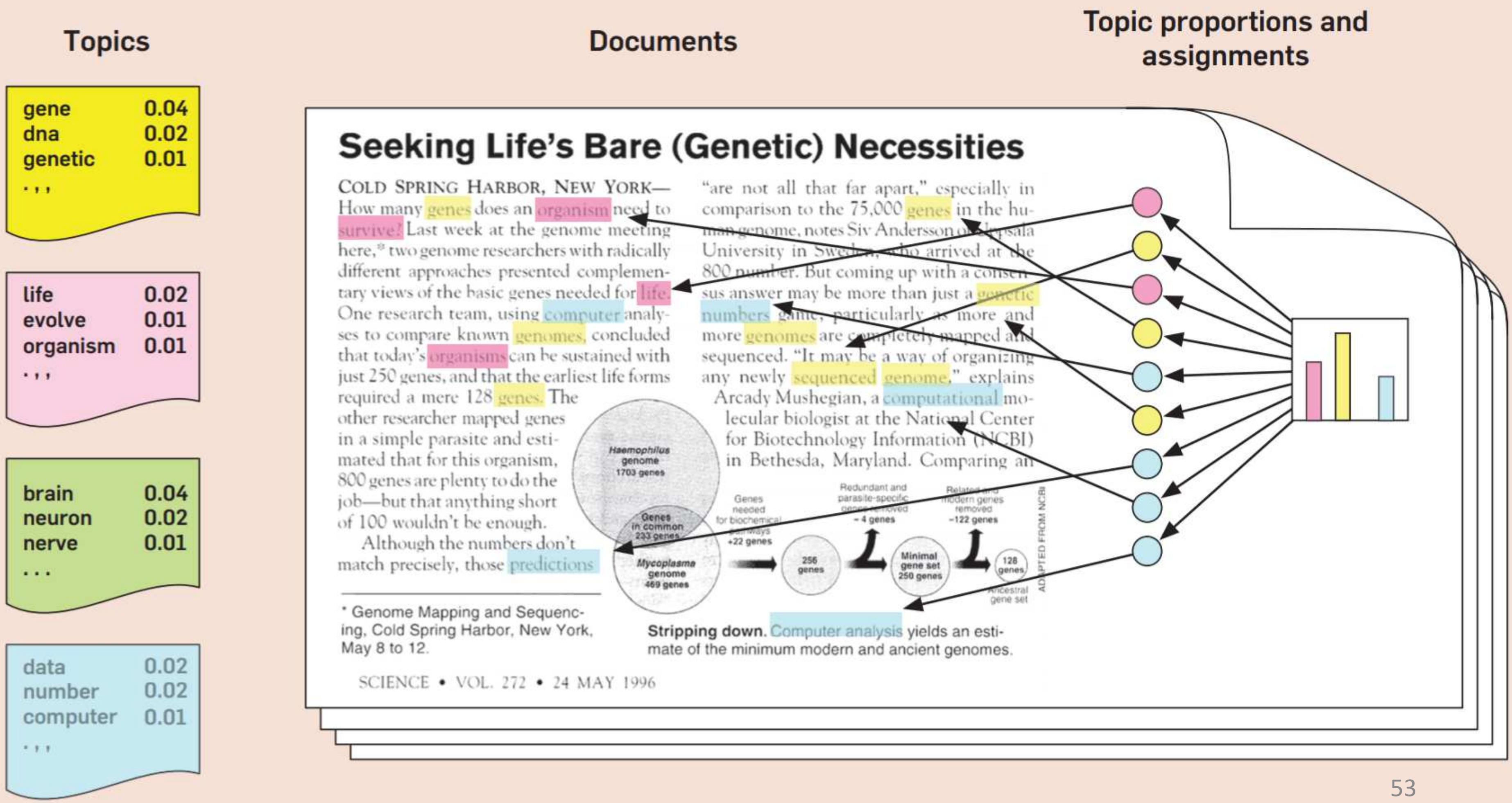
Document (e.g. newspaper article) = Bag of words.
Corpus (e.g. collection of articles) = Bag of documents.
Topic (e.g. "Health") = Distribution over words.
Each document is comprised of mixtures of topics.
The goal of topic modeling is to find the *topics* and the *mixture proportions*.

For example:
"Sports" topic: {Score, game, football, baseball, soccer, tie, win, lose, …}
"Finance" topic: {Interest, dividends, crash, buy, sell, price, …}
"Politics" topic: {Law, Congress, President, election, campaign, …}

A newspaper article might be 80% politics, 20% finance, and 0% sports.

Topic Modeling

The machine learning community has a zoo of methods for topic modeling. Some even with theoretical guarantees!



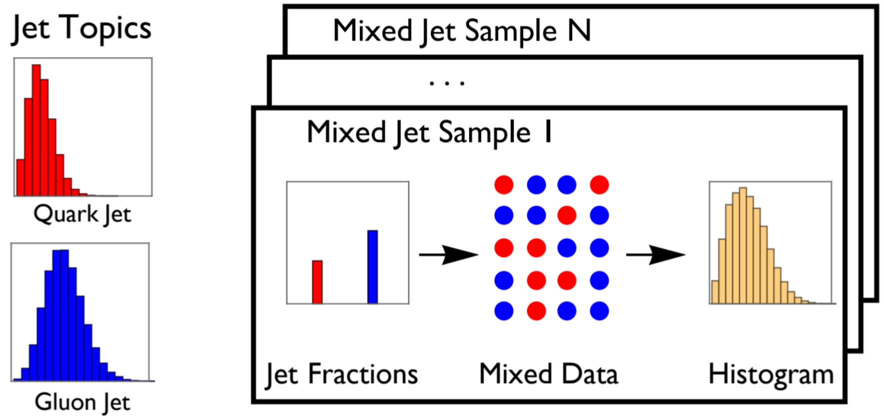
Jet Topics

[EMM and J. Thaler, arXiv: 1802.00008]

How do jets come in?

Jet observable distributions are *mixtures* of the quark and gluon distributions.

$$p_{M_a} = \sum_{k=1}^{K} f_k^{(a)} p_k(\boldsymbol{x})$$



Jet observables have the same generative model as *documents*!

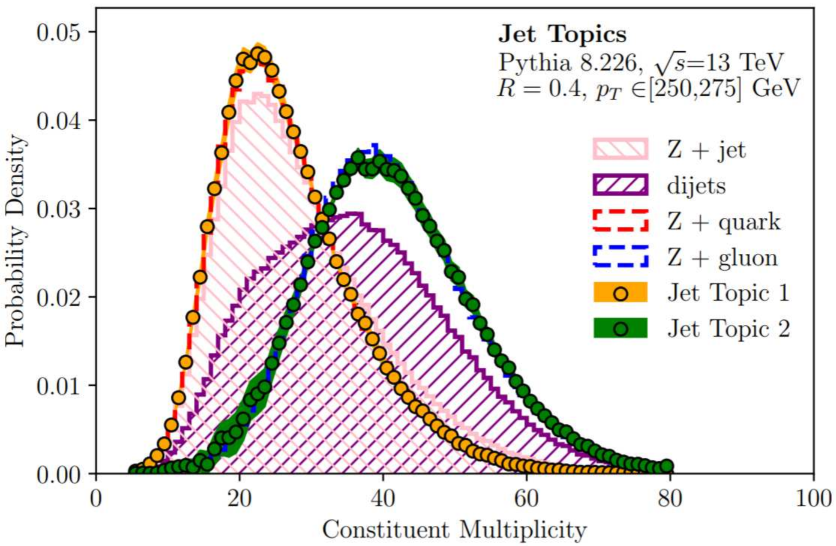
Document-Jet Correspondence

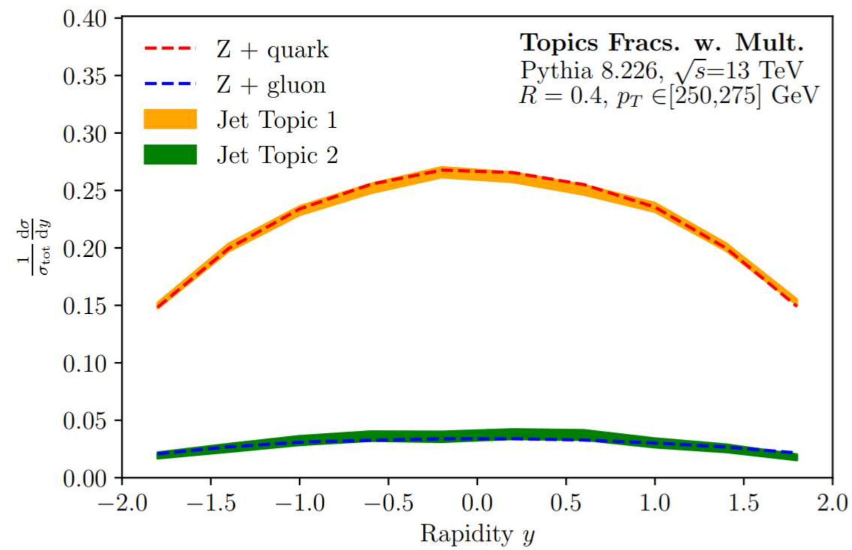
Topic Model	Jet Distributions
Word	Histogram bin
Vocabulary	Jet observable
Topic	Type of jet (i.e. *jet topic*)
Document	Histogram of jet observable(s)
Corpus	Collection of histograms

Jet Topics

What is topic modeling with jets good for?

We can use topic modeling methods to extract the topics (quark and gluon distributions) and the mixture proportions (quark and gluon fractions).





Jet topics sheds light on defining "quark" and "gluon" in theory & in experiment.
Extract the notion of "quark" and "gluon" from the data itself.
The jet topics method can be used directly on data!

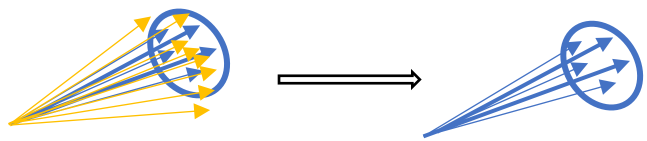
Jet Tasks We'll Talk About

Jet Tagging: How can we distinguish a quark jet vs. a gluon jet? A *W* jet vs. a QCD jet?

Classification

[**[PTK, EMM**, M.D. Schwartz, 1612.01551]]()

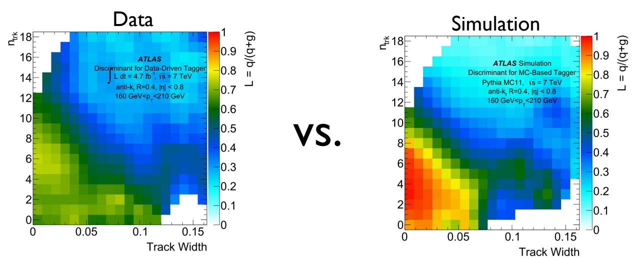
Pileup Mitigation: Can we decontaminate the jet radiation from soft, diffuse pileup?



Denoising

[**[PTK, EMM,** B. Nachman, and M.D. Schwartz, 1707.08600]]()

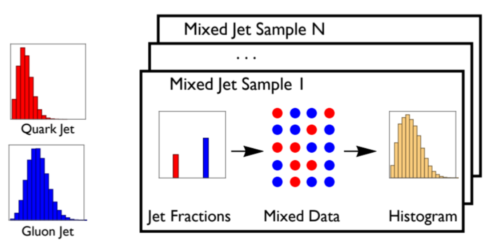
Data vs. Simulation: Do we really need simulations to provide labeled training data? Or are there ways to train algorithms directly on the (unlabeled) data?



Weak Supervision

[**[PTK, EMM**, B. Nachman, and M.D. Schwartz, 1801.10158]]()

Measuring Jet Observables: Do we need to perfectly classify quark and gluon jets to separately measure quark and gluon jet observable distributions?



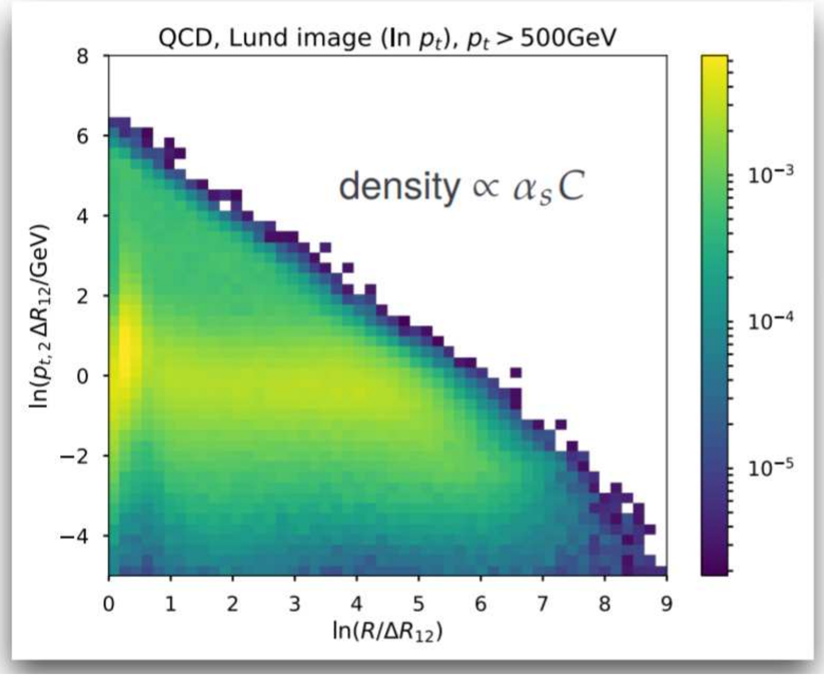
Topic Modeling

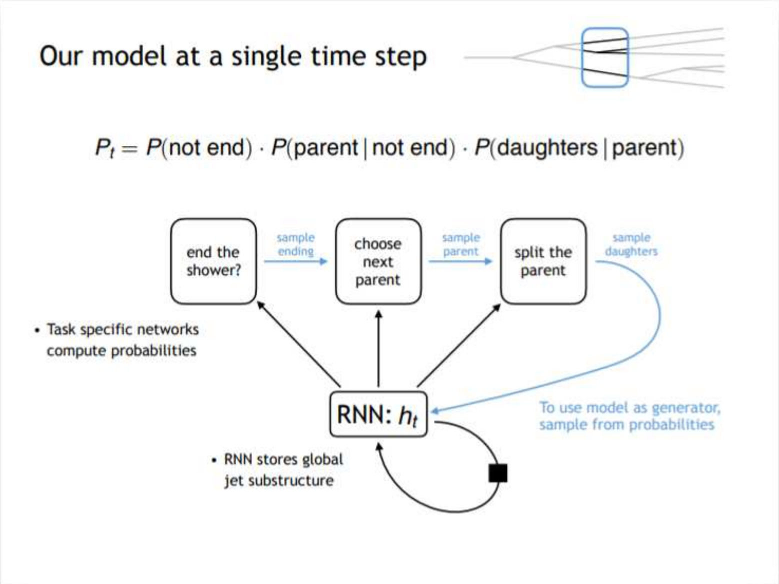
[**[EMM** and J. Thaler, 1802.00008]]()

Many Interesting Ideas Out There!

A wealth of new ways to directly access physics with machine learning methods!

F. Dreyer, G. Salam, G. Soyez





A. Andreassen, C. Frye, I. Feige, M. Schwartz

Slide from B. Nachman.

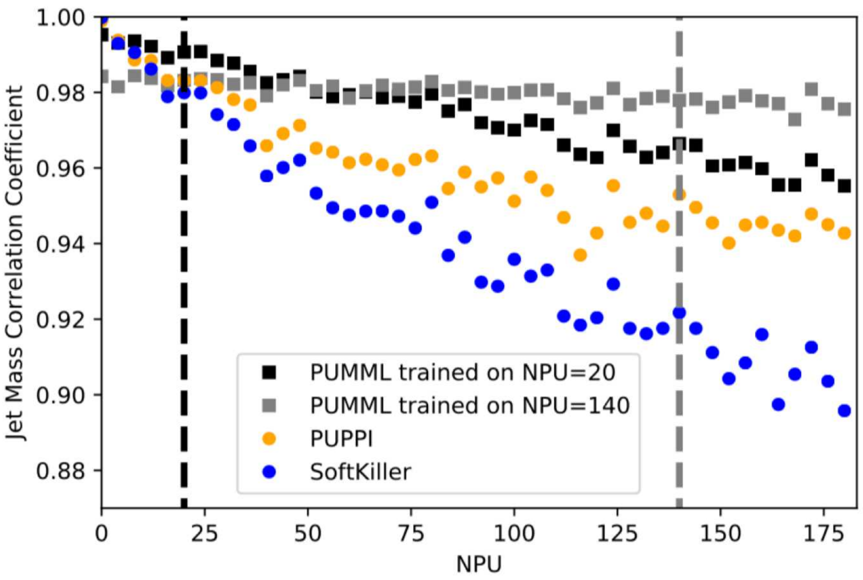
Even more waiting to be developed!

Thank you!

Backup Slides

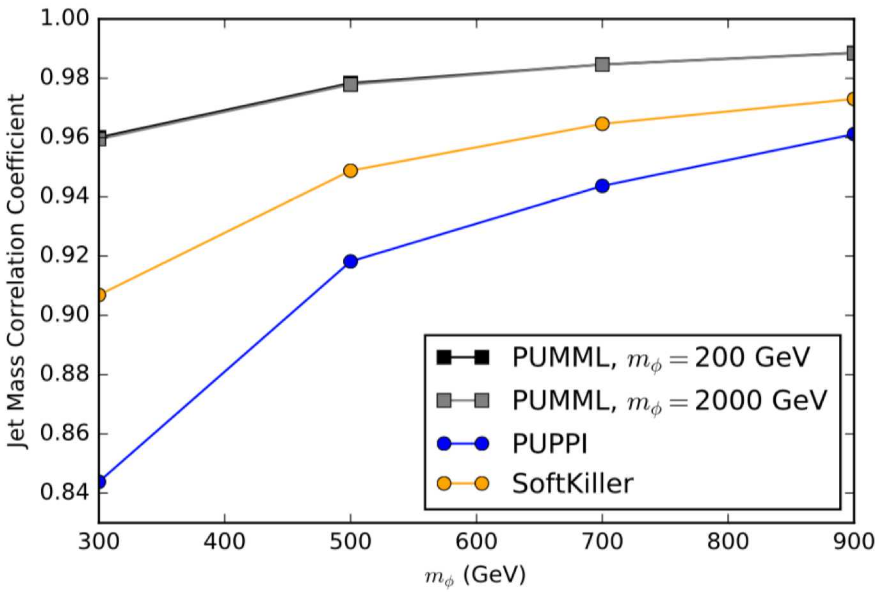
Robustness of PUMML

Train and test on different amounts of pileup



PUMML more robust than PUPPI and SK across a wide amount of pileup!

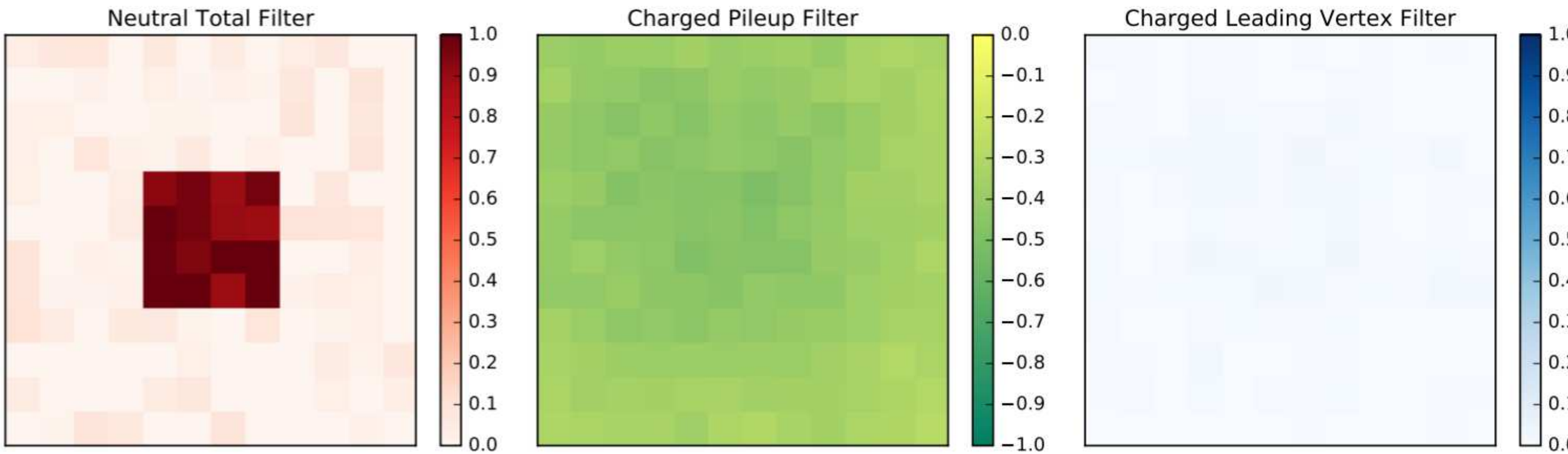
Train and test on different processes



PUMML demonstrates process independence!

What is PUMML Learning?

Train PUMML on a simplified architecture

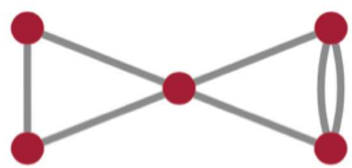


Approximately learns linear cleansing!

$$p_T^{N,LV} = p_T^{N,tot} - \left(\frac{1}{\overline{\gamma_0}} - 1\right) p_T^{C,PU}$$

Multigraph/EFP Correspondence

Multigraph $\longleftrightarrow$ EFP



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} z_{i_5} \theta_{i_1 i_2}\theta_{i_2 i_3}\theta_{i_1 i_3}\theta_{i_1 i_4}\theta_{i_1 i_5}\theta_{i_4 i_5}^2$$

$\longleftrightarrow$ z_{i_j}

$\longleftrightarrow$ $\theta_{i_k i_l}$

N	Number of vertices	$\longleftrightarrow$	N-particle correlator
d	Number of edges	$\longleftrightarrow$	Degree of angular monomial
χ	Treewidth + 1	$\longleftrightarrow$	Optimal VE Complexity

Connected $\longleftrightarrow$ Prime

Disconnected $\longleftrightarrow$ Composite

$\vdots$

EFPs linearly span IRC-safe observables

IRC-safe Observable

Energy Expansion: Expand/approximate the observable in polynomials of the particle energies

IR safety: Observable unchanged by addition of infinitesimally soft particle

C safety: Observable unchanged by the collinear splitting of a particle

Relabeling Symmetry: All ways of indexing particles are equivalent

New, direct argument from IRC safety
See also: F. Tkachov, hep-ph/9601308
N. Sveshnikov and F. Tkachov, hep-ph/9512370

Energy correlators linearly span IRC-safe observables

Angular Expansion: Expansion/approximation of angular part of correlators in pairwise angular distances
Analyze: Identify the unique analytic structures that emerge as non-isomorphic multigraphs/EFPs

Similar expansions & emergent multigraphs in:
M. Hogervorst *et al.* arXiv:1409.1581
B. Henning *et al.* arXiv:1706.08520

EFPs linearly span/approximate IRC-safe observables!

Linear Regression and IRC-safety

$\frac{m_J}{p_{TJ}}$: IRC safe. No Taylor expansion due to square root.

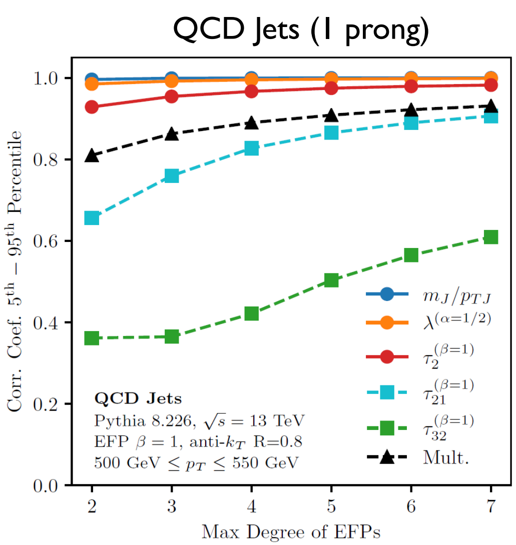
$\lambda^{(\alpha=1/2)}$: IRC safe. No simple analytic relationship.

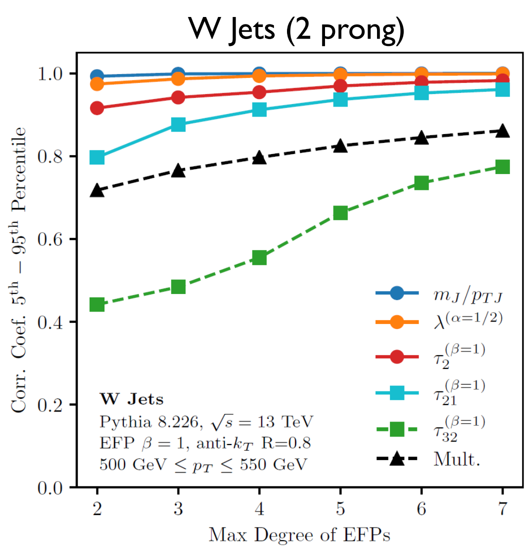
τ_2: IRC safe. Algorithmically defined.

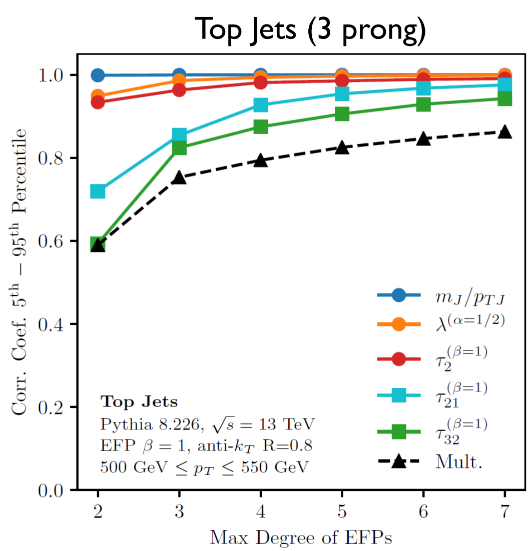
τ_{21}: Sudakov safe. Safe for 2-prong jets and higher. [A. Larkoski, S. Marzani, and J. Thaler, 1502.01719]

τ_{32}: Sudakov safe. Safe for 3-prong jets and higher.

Multiplicity: IRC unsafe.





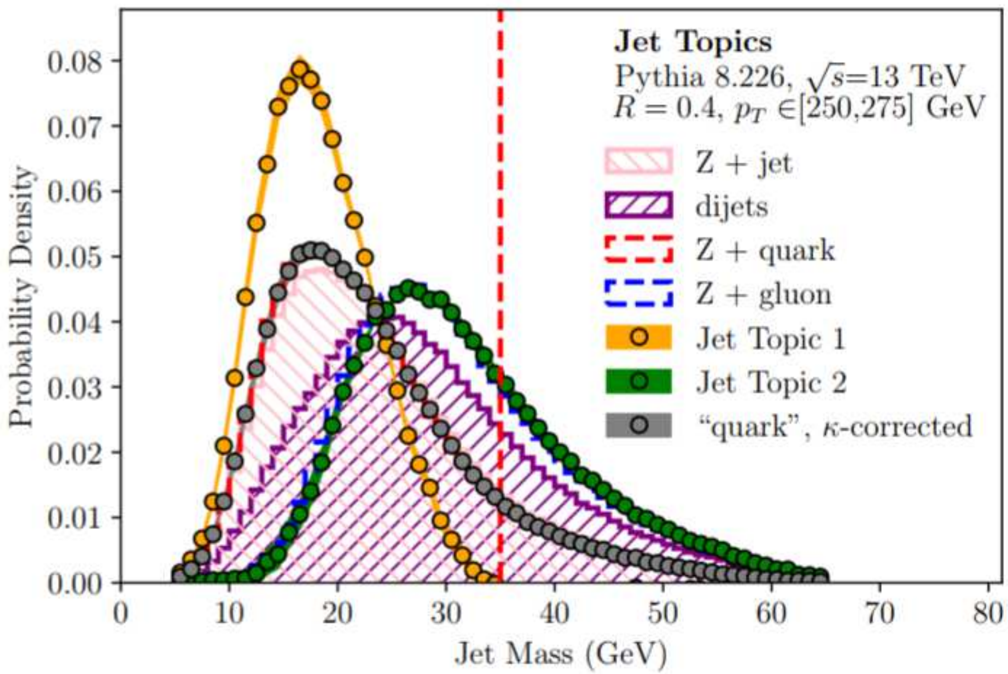


Expected to be IRC safe = Solid.
Expected to be IRC unsafe = Dashed.

Jet Topics

[**EMM** and J. Thaler, arXiv: 1802.00008]

Caveats apply: Only works "out of the box" for certain observables with "mutual irreducibility".
Need some additional theory input for other observables.



Can understand the behavior with a leading logarithmic calculation of the jet mass topics:

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

$$\kappa(q|g) = \frac{C_F}{C_A} \min \Sigma_q^{1-\frac{C_A}{C_F}} = \frac{C_F}{C_A}$$