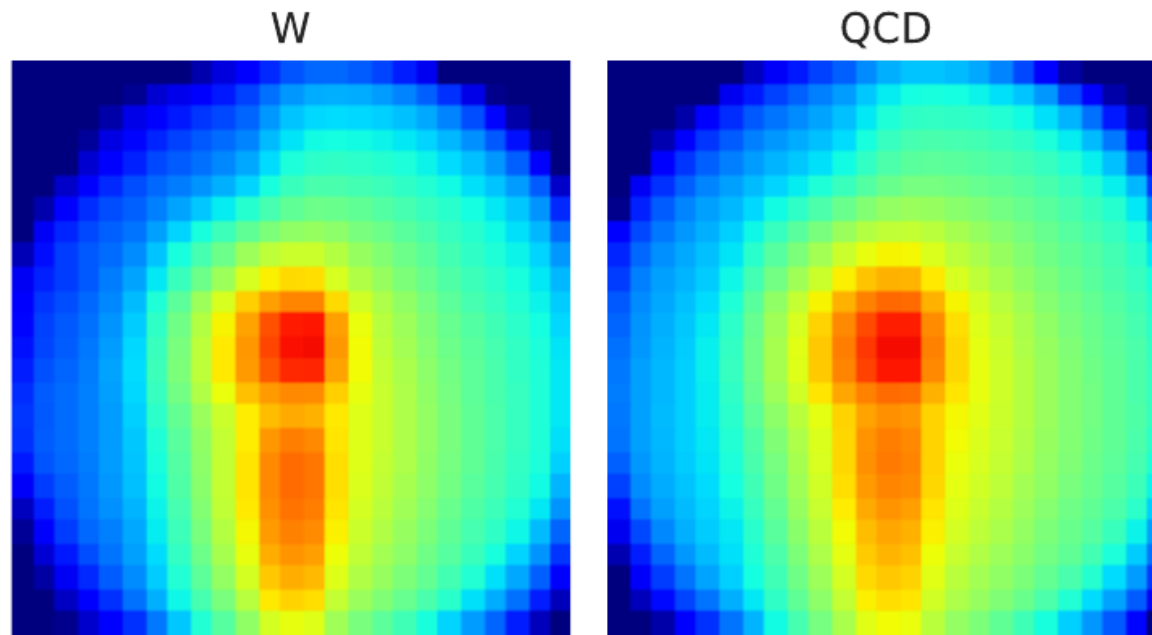
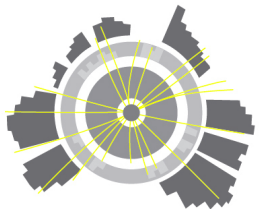


# Jet-Tagging: Overview



Matthew Dolan  
University of Melbourne  
Centre of Excellence for Particle Physics at the Terascale



**CoEPP**  
ARC Centre of Excellence for  
Particle Physics at the Terascale



THE UNIVERSITY OF  
**MELBOURNE**

# This talk

- Meant to be an ‘inspiration and overview’
- Stands between you and other, more interesting talks
  - Does not involve new ATLAS/CMS results
    - Does not involve back-flipping robots

# Recent themes

- Taggers for everything
- Supervised vs unsupervised learning
- Jet representation: Images, sequences
- Networks: CNNs, RNNs, LSTM, GANs,
  - What is the machine learning
- How can we exploit that analytically

# Jet-Tagging

In many situations we want to specify the origin of a jet, from either a single state

- Quark vs gluon
- Heavy flavour vs light (b,c-tagging)

Or the decay products of a heavier one:

- SM resonances: Higgs, W/Z, top
  - Some BSM resonance

# Machine Learning and Jet-Tagging

- An old idea: quark vs gluon, 1990
- JETNET, 1991
- Further quark vs gluon, flavour tagging
- Event generator dependence: JETSET, ARIADNE, HERWIG
- Hadronisation studies
- Utility in track-finding and triggering

[Lonnblad, 1990](#)

[Lonnblad, 1991](#)

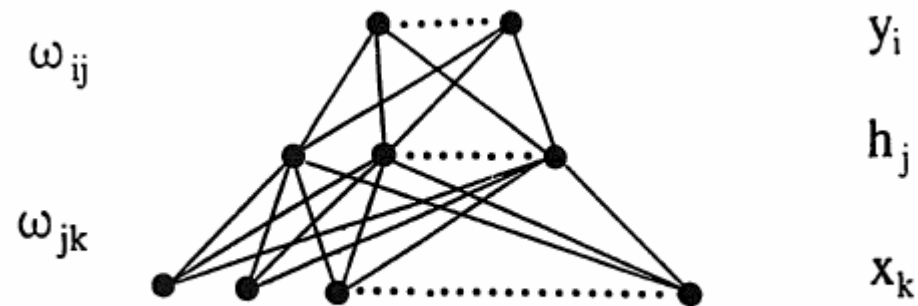
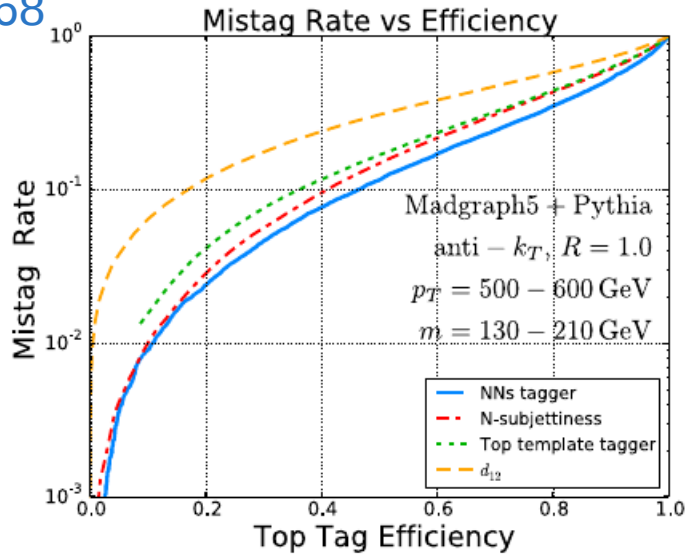


Fig. 1. A feed-forward neural network with one layer of hidden units.

# Taggers for Everything: MC

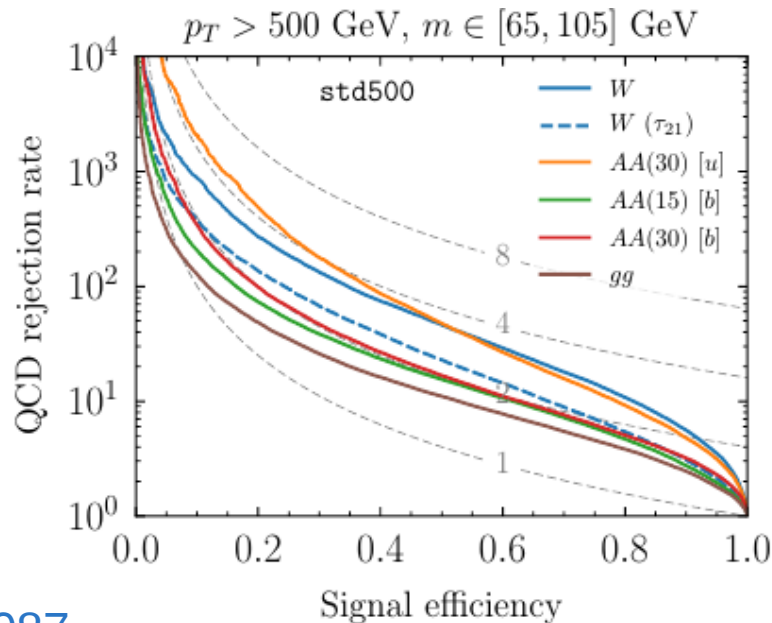
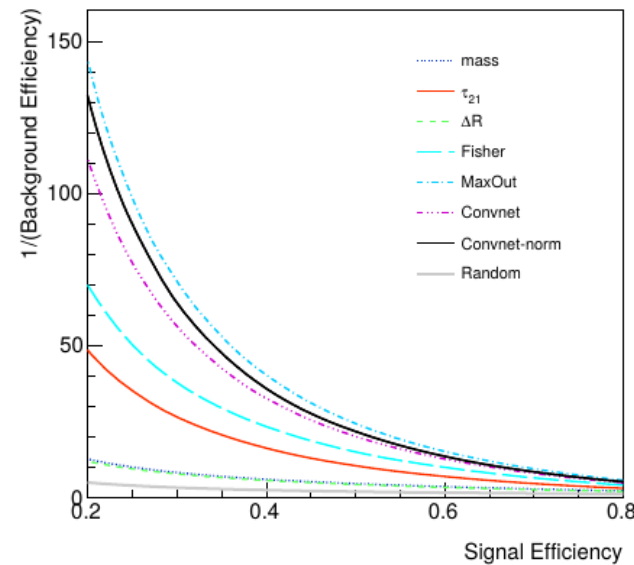
- Quark vs gluon, W/Z, Higgs, top, anti-QCD and BSM...

1501.05968

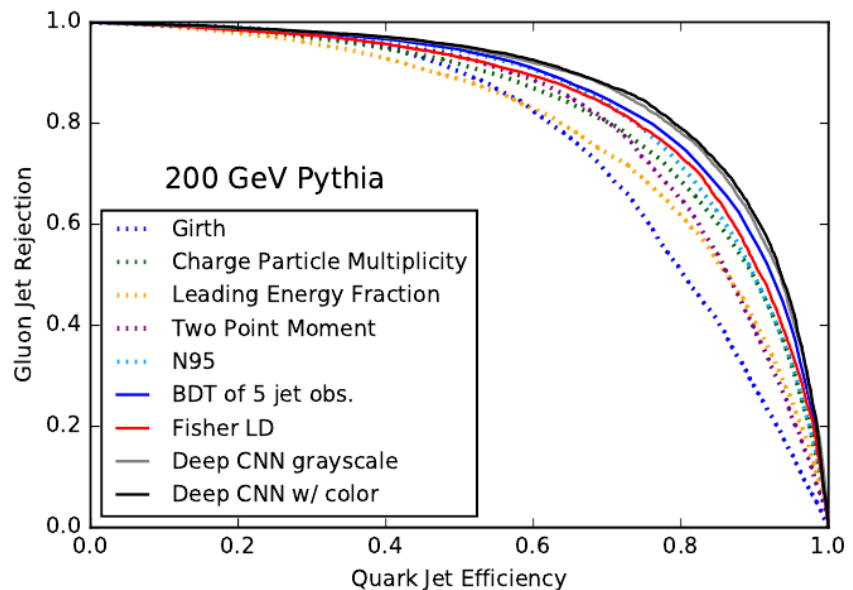


250 <  $p_T$ /GeV < 300 GeV, 65 < mass/GeV < 95  
 $\sqrt{s} = 13$  TeV, Pythia 8

1511.05190



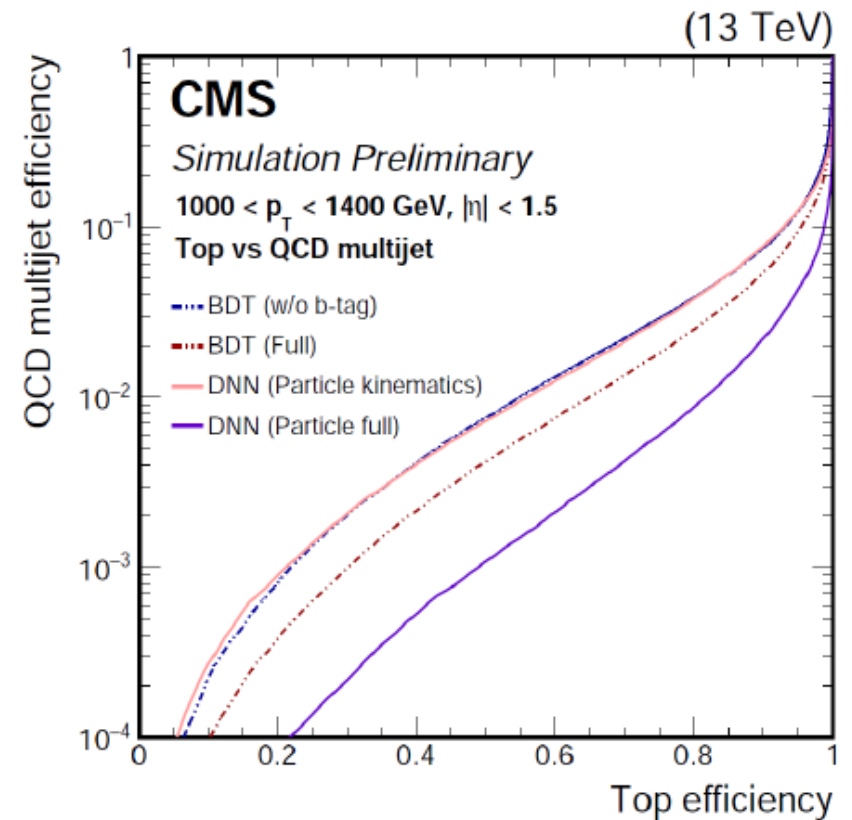
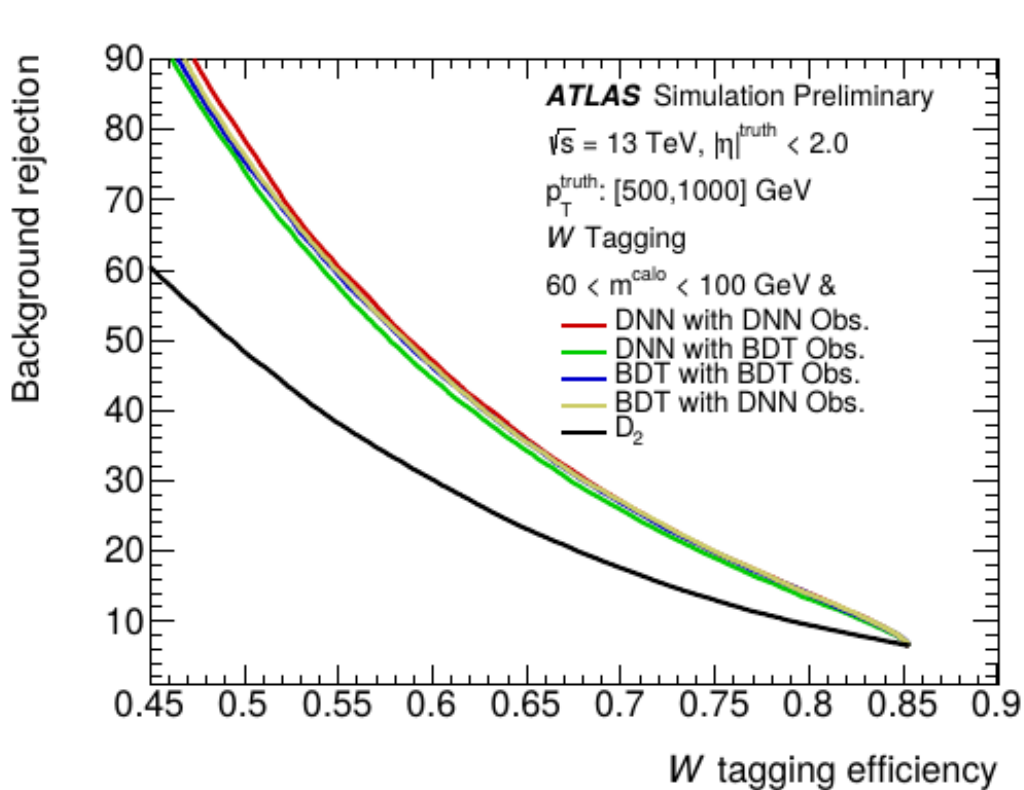
1612.01551



1709.01087

# I like your manifesto, put it to the test-o

- ATLAS and CMS studies + more to come
- CMS OpenData and MC: theorists can play too!

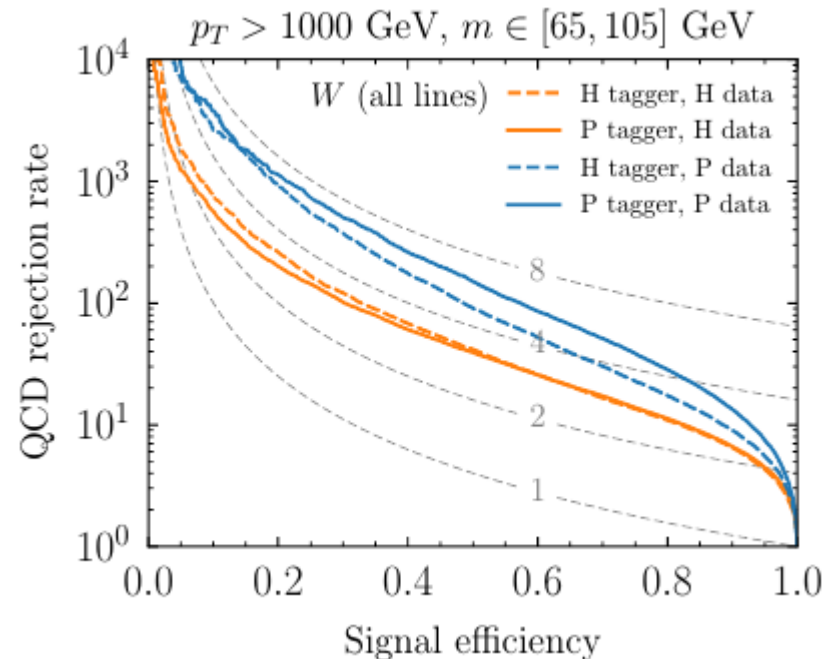
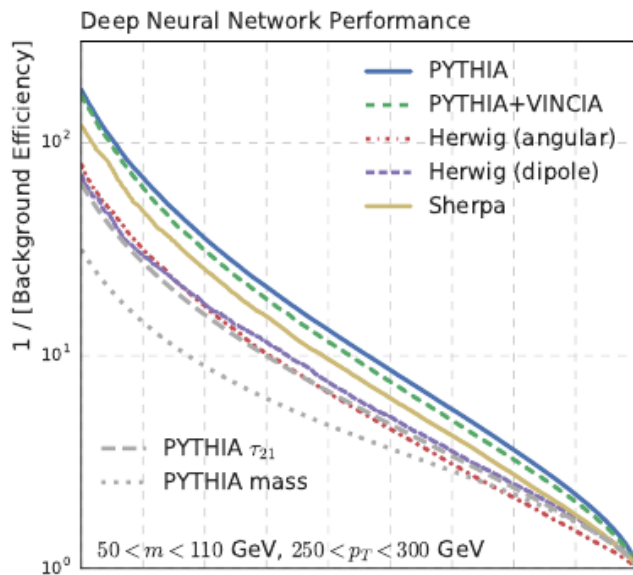


ATLAS-PHYS-PUB-2017-004

CMS-DP-2017-049

# Supervised vs Unsupervised

- Uncertainties from training on MC.
- Can we train taggers directly on the data?



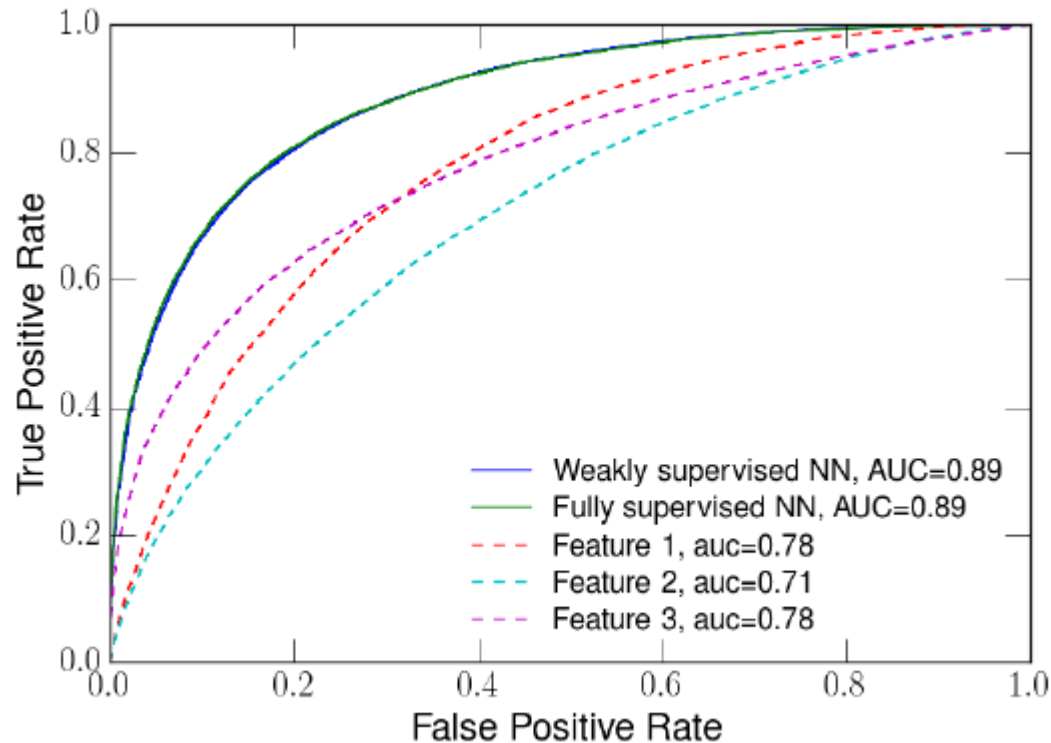
1609.00607

1709.01087



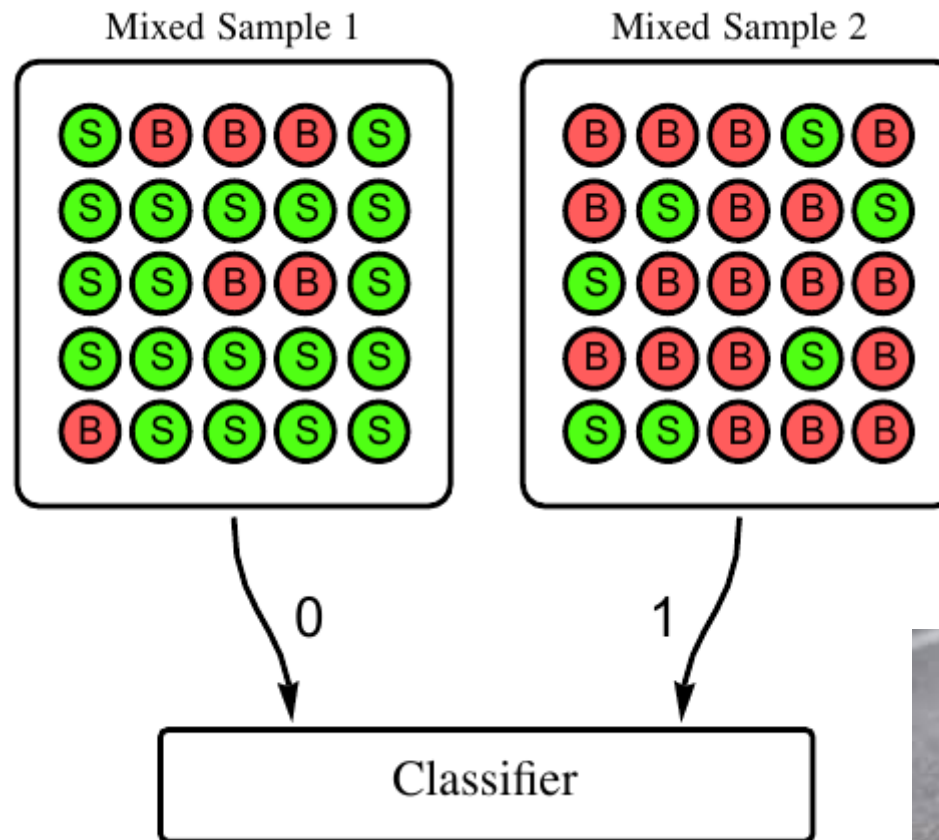
# Supervised vs Unsupervised

- Weak supervision: training with labels
- Know the fractions of S and B in training samples
- Looking at W-tagging



# Supervised vs Unsupervised

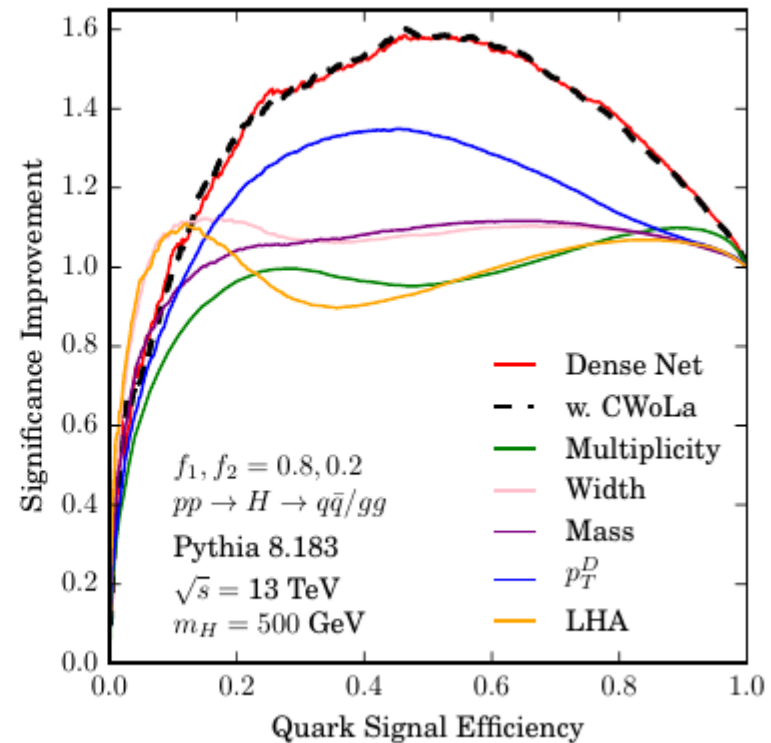
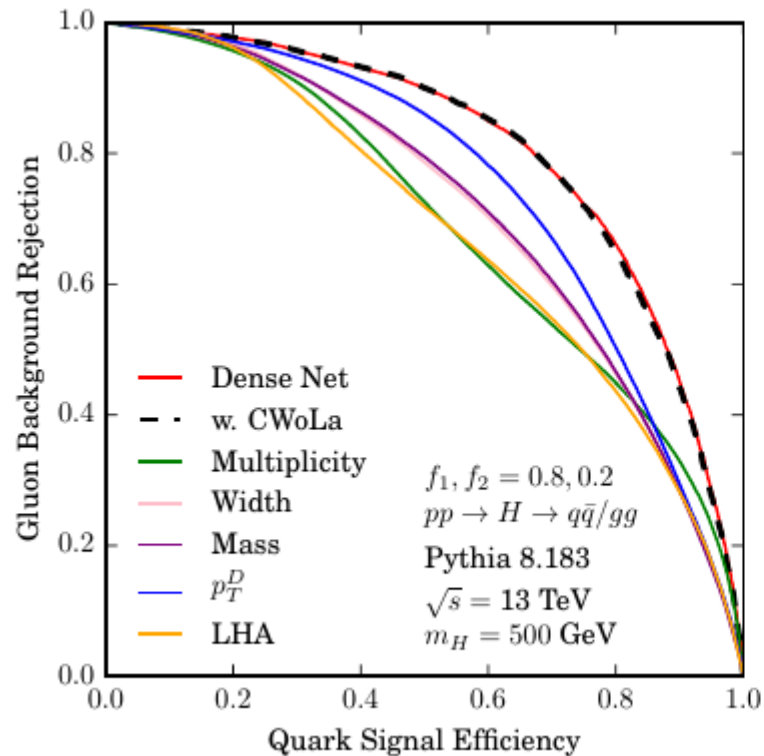
- Classification without labels (CWoLa) [1708.02949](#)



# Supervised vs Unsupervised

- Classification without labels

1708.02949

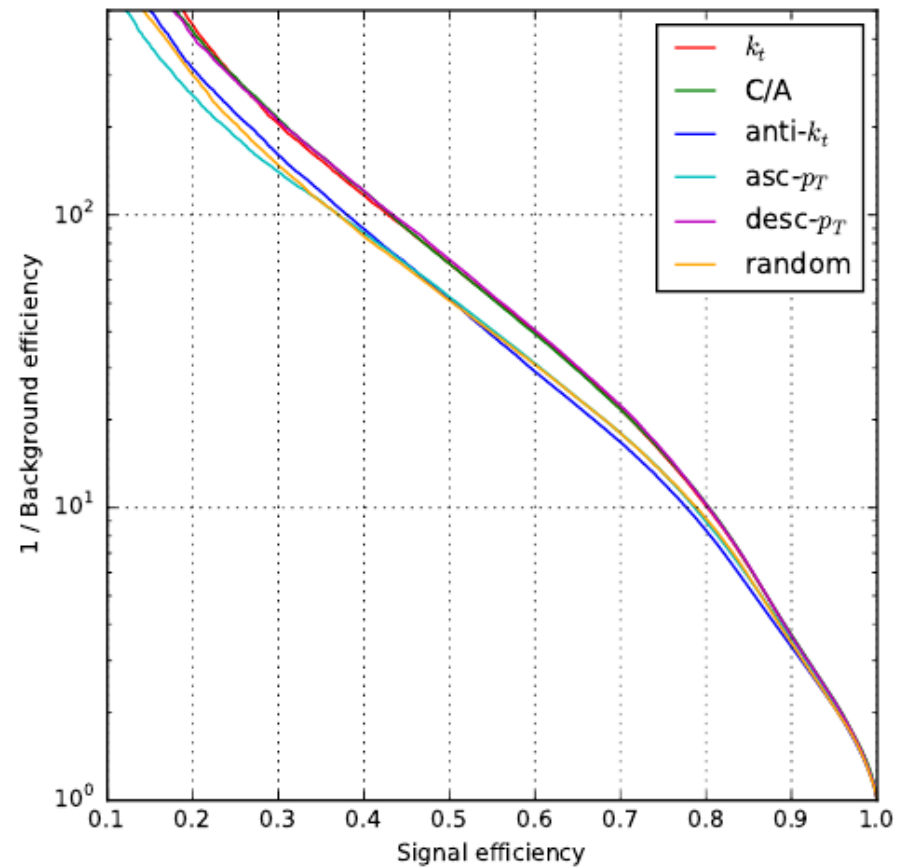
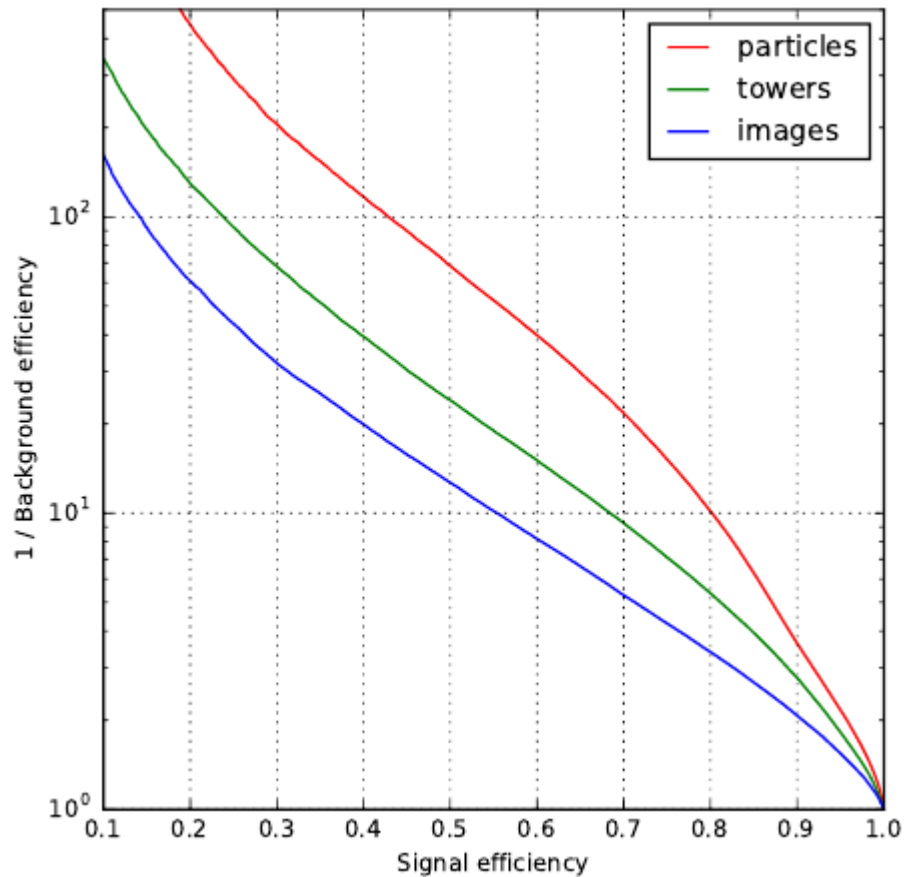


# Jet Representations

- What are we tagging?
- Is an image a good representation of a jet?
- Use 4-vectors and relational information?
- How much information is in a jet?
- High-level variables?

# Jet Representations

- Jets as vectors + clustering history



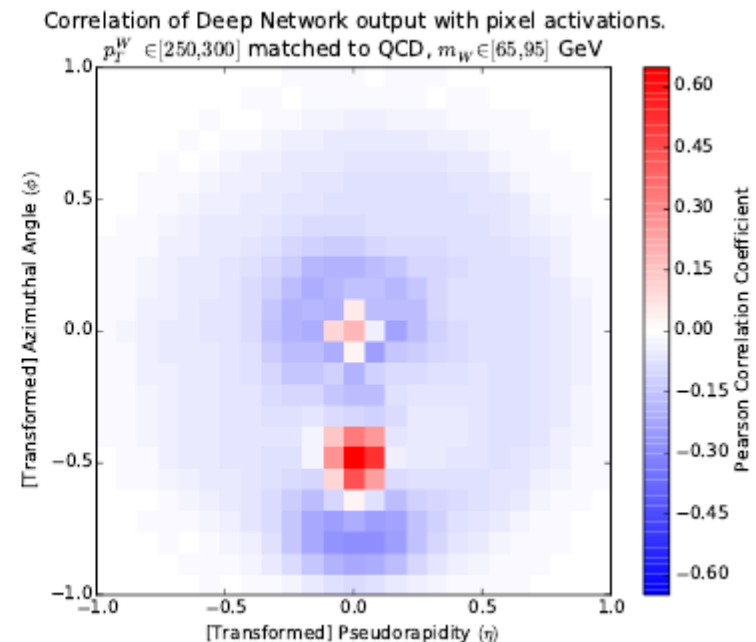
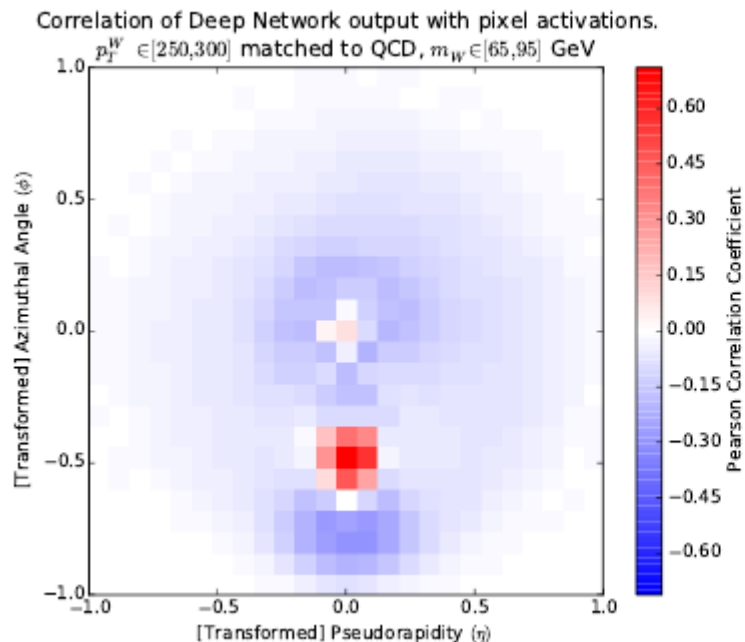
1702.00748

# What is the machine learning?

- What information are these algorithms exploiting?
- Improvements over tagging variables from exploiting colour-flow and radiation patterns
- “Plane’ away information contained in standard variables

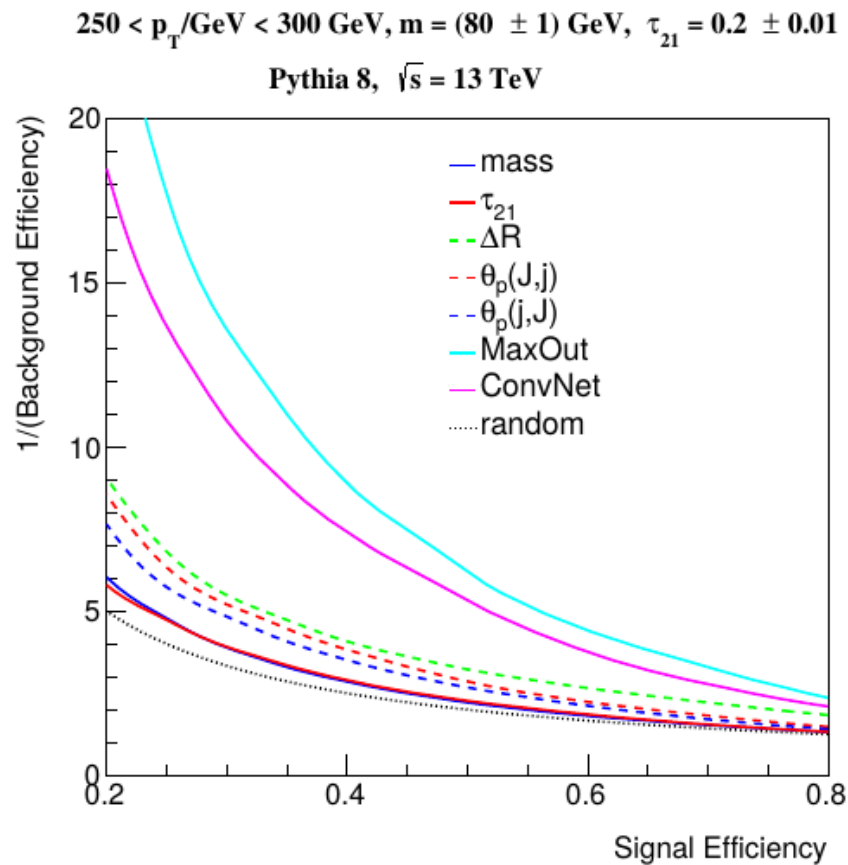
1709.10106

Oregon tag team



# What is the machine learning?

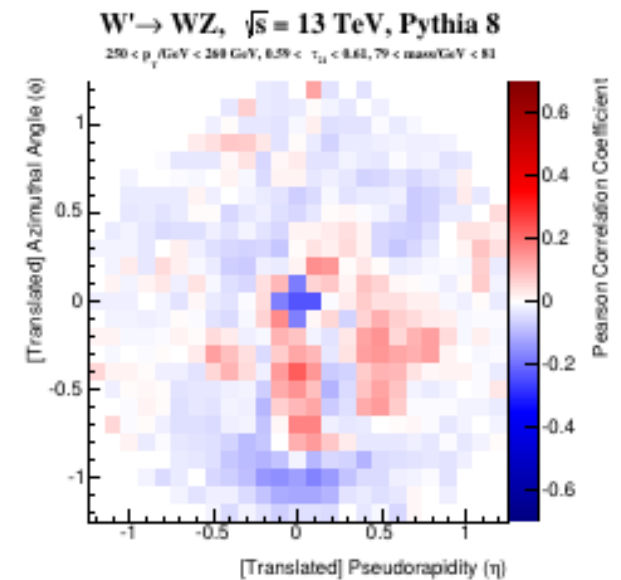
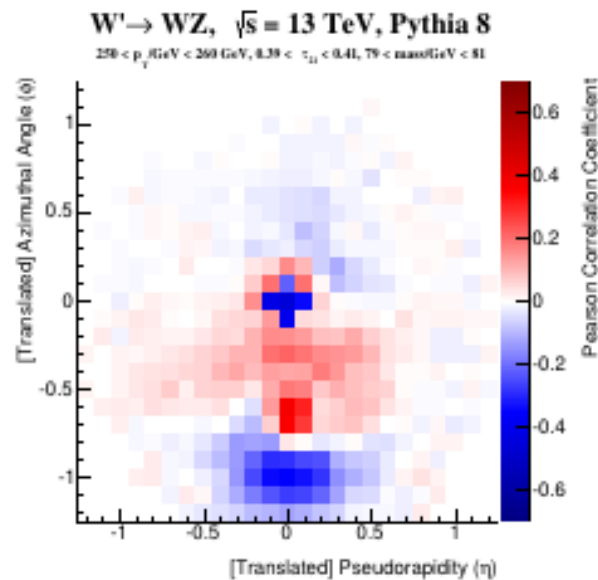
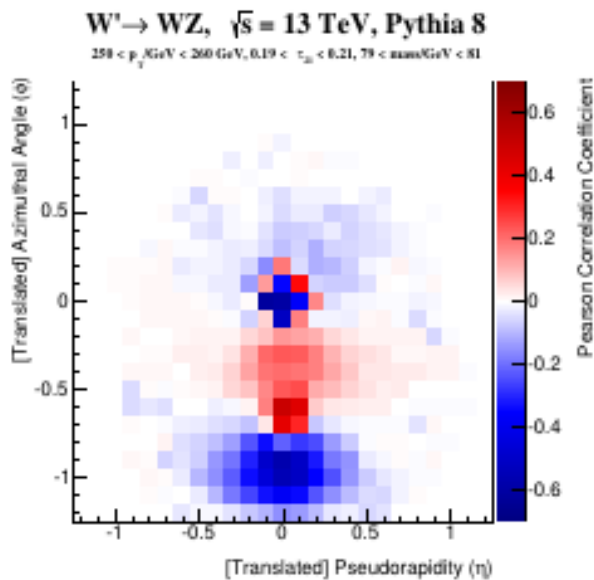
- Look in restricted phase space: W vs QCD.





# What is the machine learning?

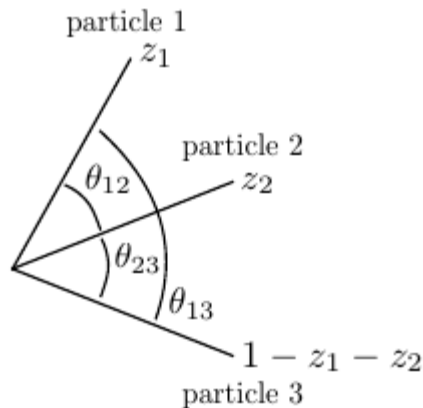
- Look in restricted phase space: W vs QCD.
- Pearson Correlation Coefficients





# Exploiting what the machine learns

- Design and study new tagging variables?
- Probe higher-order emissions? [1609.07483](#)
- Form a basis of variables to probe available phase space of jet?
- N-subjettiness
- Energy flow polynomials MIT tag team

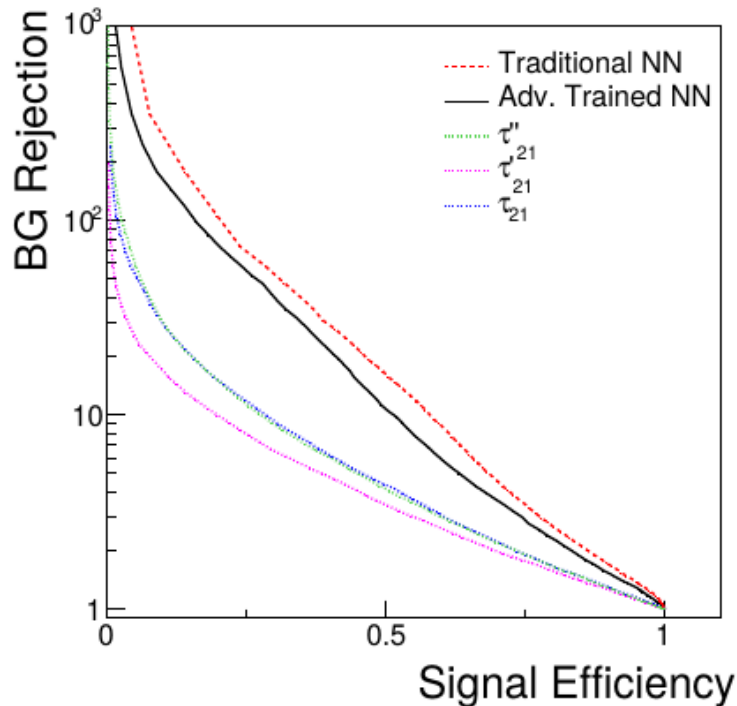


$$\beta_3 \equiv \left(\tau_1^{(0.5)}\right)^a \left(\tau_1^{(1)}\right)^b \left(\tau_1^{(2)}\right)^c \left(\tau_2^{(1)}\right)^d \left(\tau_2^{(2)}\right)^e.$$

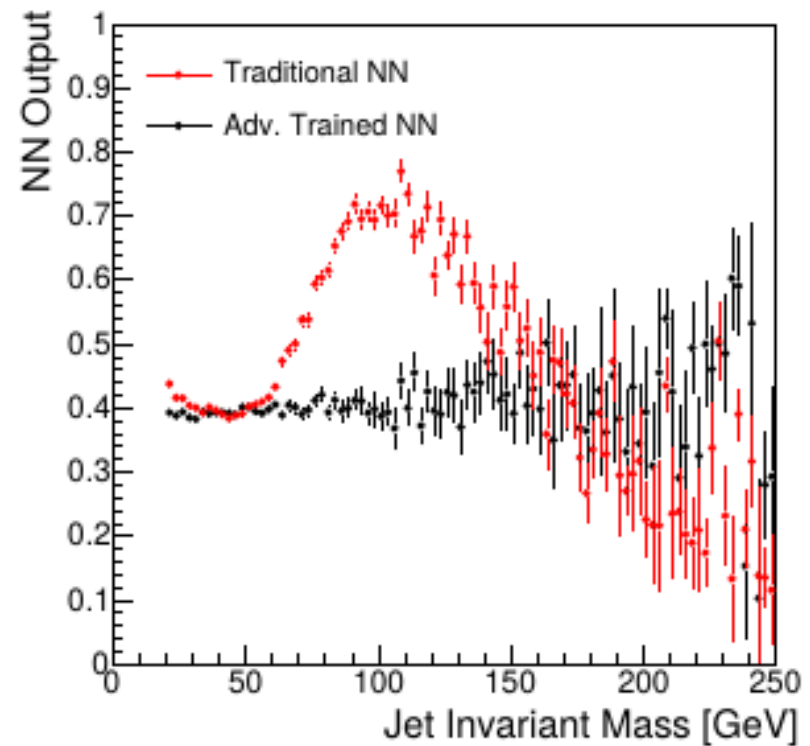
[1710.01305](#)

# Yes we GAN! [1611.01046](#)

- Use adversarial networks to decorrelate network performance from jet mass.



[1703.03507](#)



Analytic methods: [1603.00027](#) [1710.06859](#)

Also in anti-QCD tagger: [1709.01087](#)

# Conclusions?

- Taggers for everything?
- Supervised vs unsupervised learning?
- Jet representation: Images, sequences?
- Networks: CNNs, RNNs, LSTM, GANs?
  - What is the machine learning?
- How can we exploit that analytically?
  - Your favourite topic?

# Inspirational Back-Up Slide

## real life

### Sausage-addicted kookaburra too fat to fly

Story summary:

- Bird became fat on BBQ handouts
- Too fat to escape mauling by dogs
- She weighed a whopping 540g

**TOO** many tasty sausages almost killed this kookaburra.

But health worries weren't the threat. She became so obese from barbecue handouts she could not fly when attacked by dogs in a Mosman park.

After weighing in at 540g, 40 per cent heavier than a typical adult bird, the kookaburra has been sent to bird boot camp to shed weight, prompting Taronga Zoo to warn we are killing our wildlife with kindness.

