Classifaction of $gg \to gh$ against $qg \to qh$

Jet Variables, Machine Learning, Box vs Triangle Diagrams

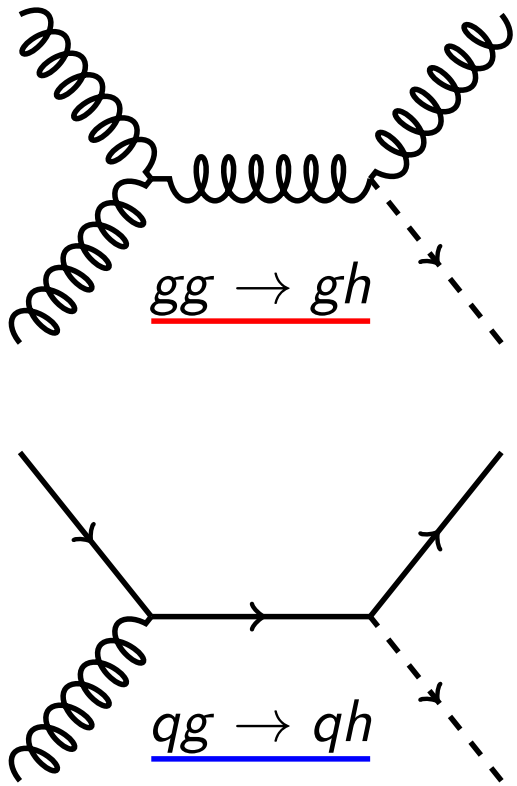
Joshua Lin + Ben Nachman
joshua.z.lin@gmail.com

LBNL, Berkeley

2018 January 8-12 Conference, LBNL

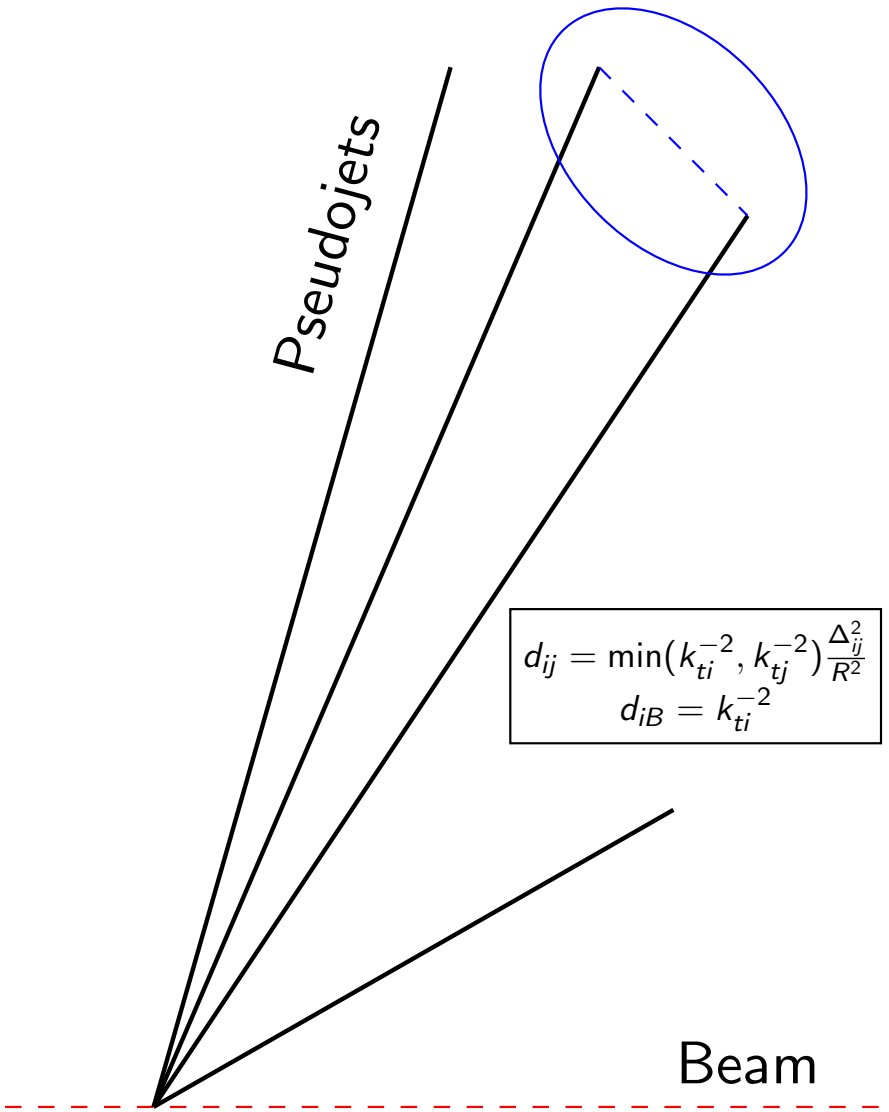
- These days, we understand very well the Higgs decay channels $H \to \gamma\gamma$, $H \to 4l$.
- The high pT regime holds great promise to resolve loops in production (boxes vs triangles). We may find New Physics here!
- Question: Can we already do this at moderate pT regime; by first differentiating different Higgs production mechanisms?

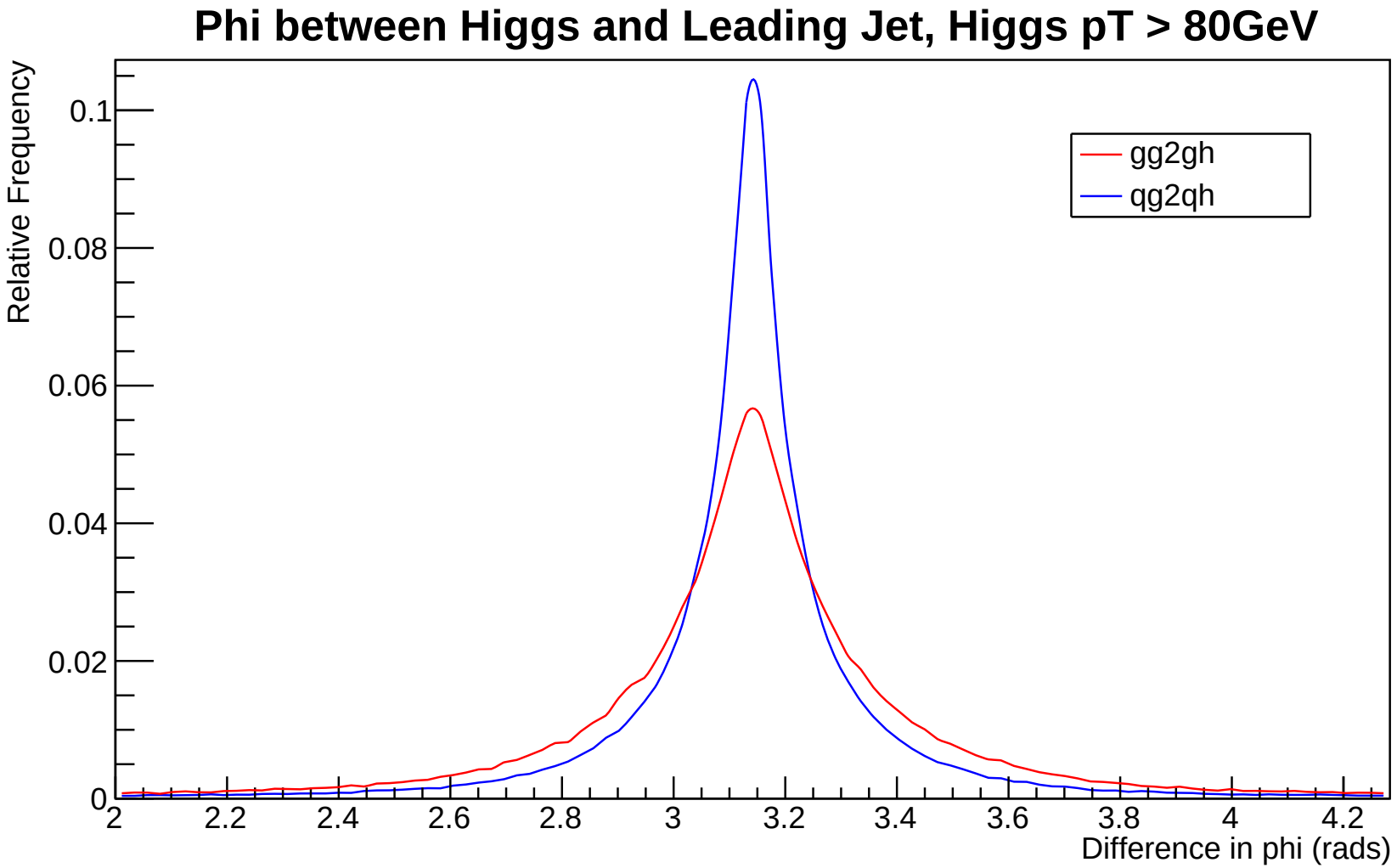
Two Higgs production mechanisms
(Loops not shown)



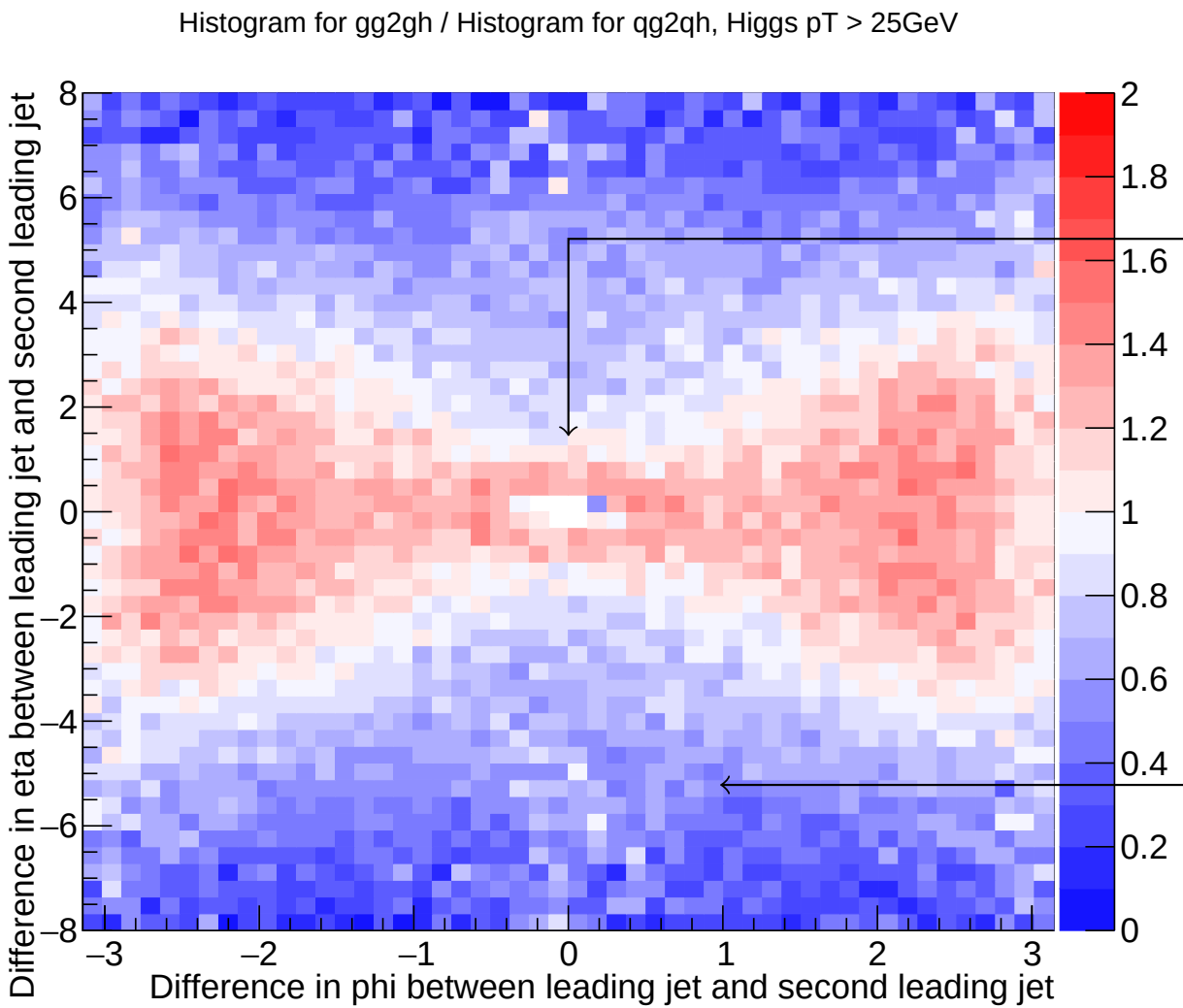
GOAL: Use ISR/FSR quark/gluon tagging to differentiate higgs production processes.

- *The idea of this project is to do a simulation-based study as a proof-of-principle*
- We used Madgraph amC@NLO to generate events to NLO, and Pythia 8.226 to shower the events. We simulated the collisions as *pp* collisions at 13TeV.
- We turned off neutrinos, and also the FSR from the Higgs particle ejected for simplicity.
- The resulting events were then clustered with FastJet with anti-kT, with R=0.4.




Phi between Higgs and Leading Jet, Higgs pT > 80GeV
Relative Frequency
0.1
0.08
0.06
0.04
0.02
0
2
2.2
2.4
2.6
2.8
3
3.2
3.4
3.6
3.8
4
4.2
Difference in phi (rads)
gg2gh
qg2qh

Angle between Leading and Subleading Jets

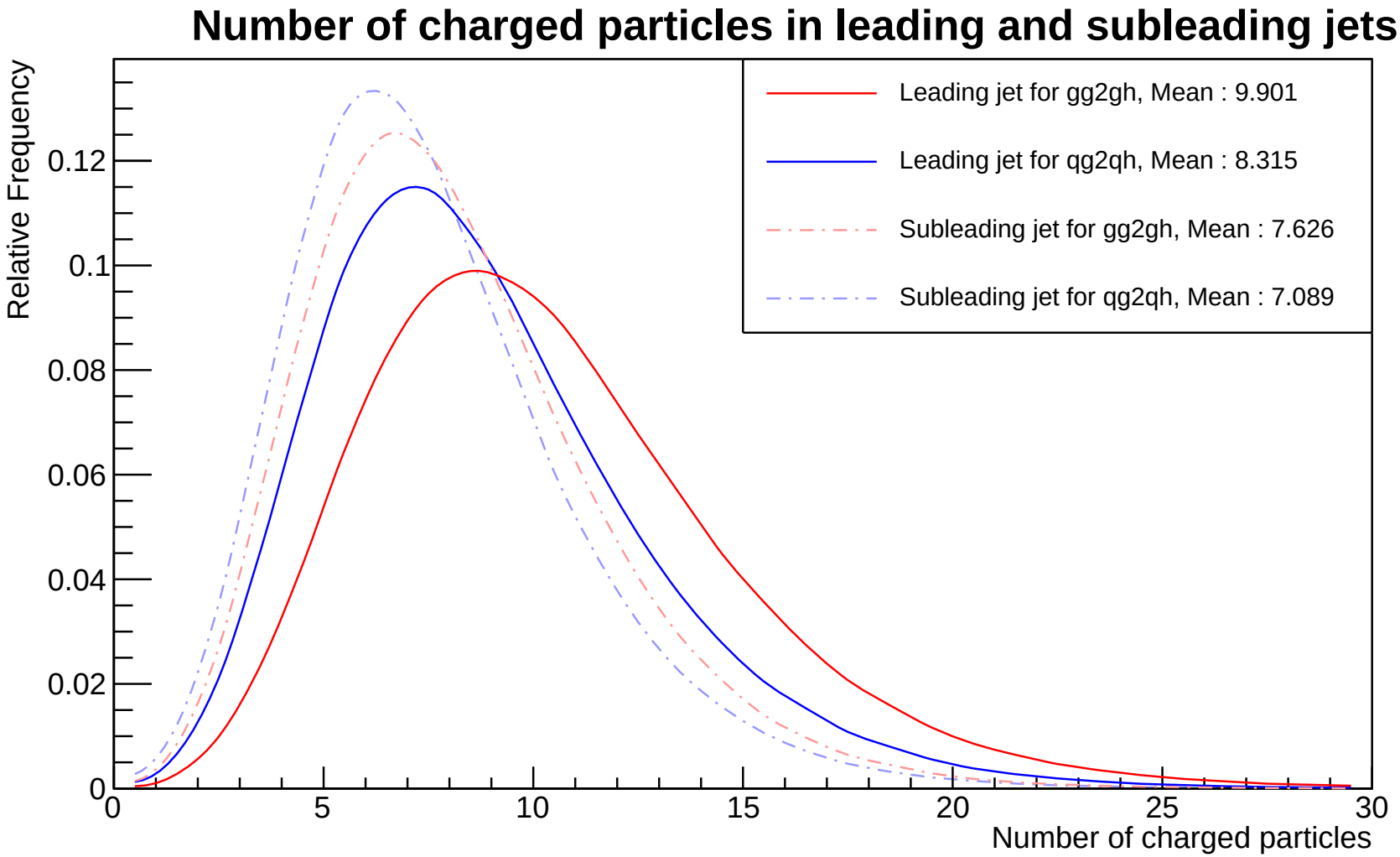


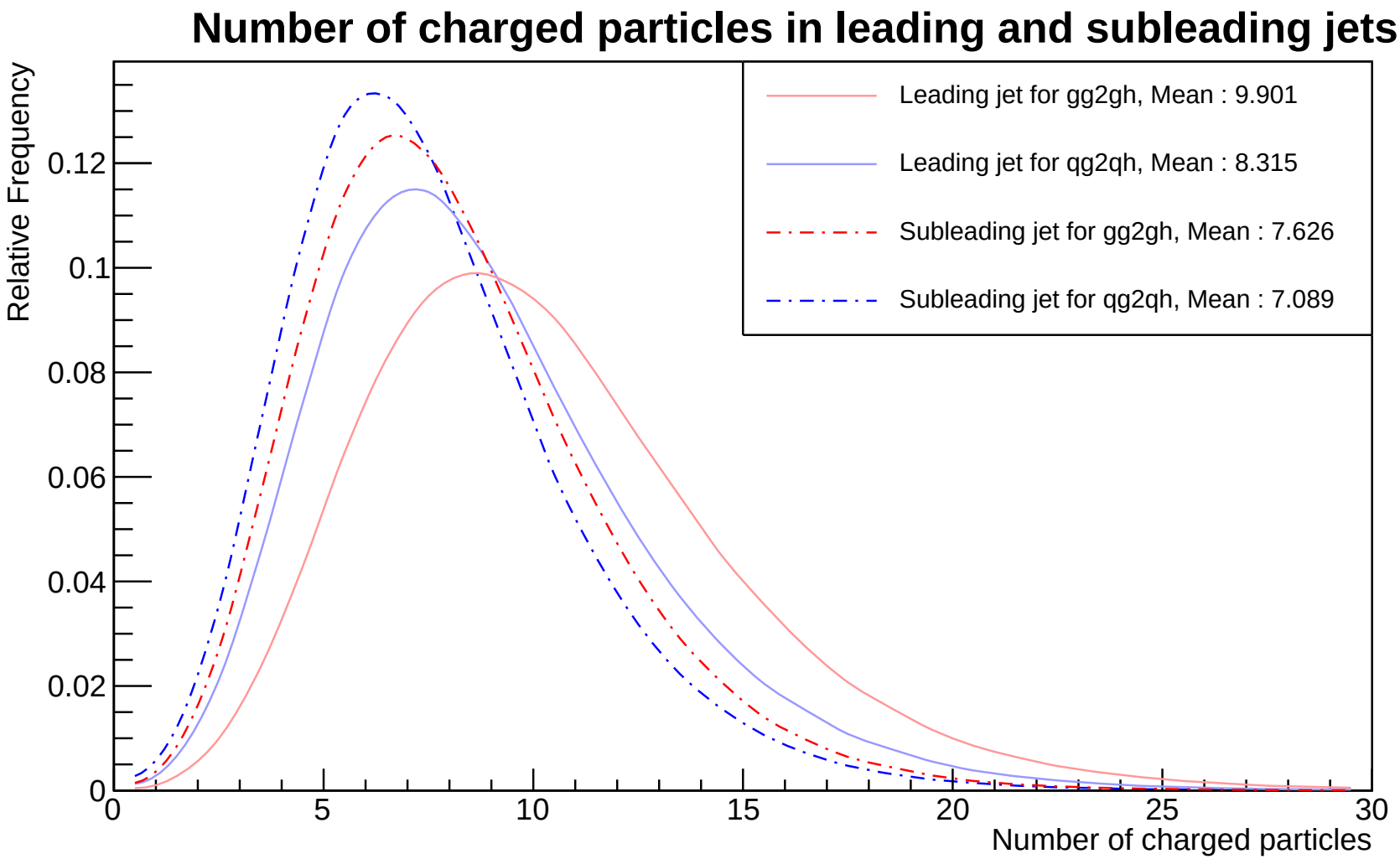
Red means more $gg2gh$

Jets generated by $gg2gh$ are clustered around the Leading Jet

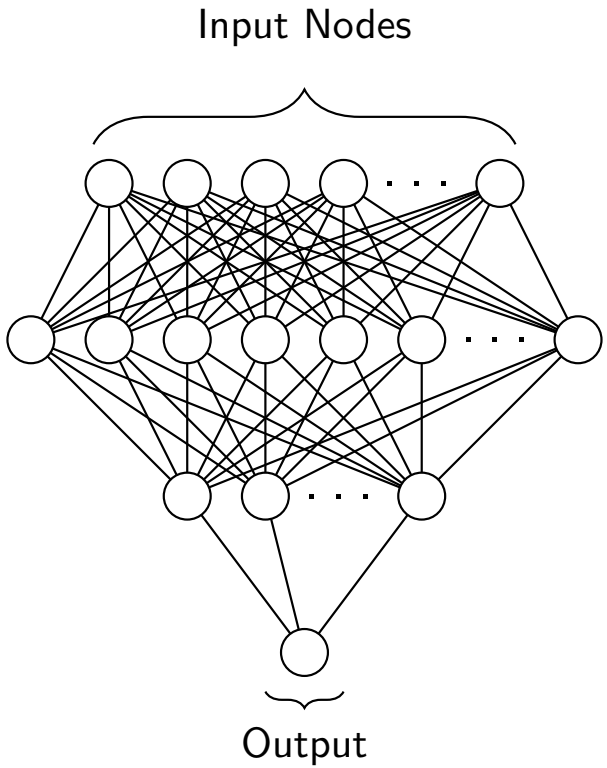
Jets generated by $qg2qh$ have a lot more variation in η

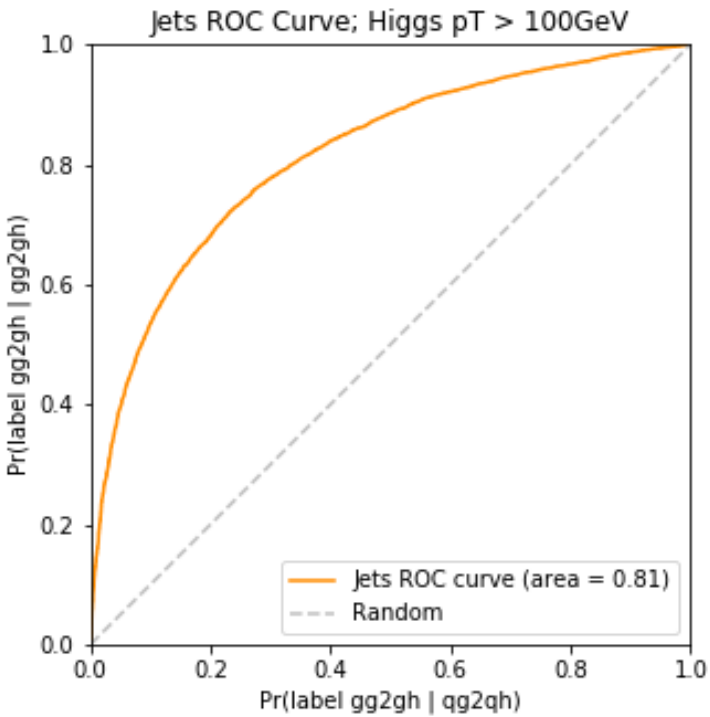
Blue means more $qg2qh$


Number of charged particles in leading and subleading jets
Relative Frequency
0
0.02
0.04
0.06
0.08
0.1
0.12
0
5
10
15
20
25
30
Number of charged particles
Leading jet for gg2gh, Mean : 9.901
Leading jet for qg2qh, Mean : 8.315
Subleading jet for gg2gh, Mean : 7.626
Subleading jet for qg2qh, Mean : 7.089


Number of charged particles in leading and subleading jets
Relative Frequency
0
0.02
0.04
0.06
0.08
0.1
0.12
0
5
10
15
20
25
30
Number of charged particles
Leading jet for gg2gh, Mean : 9.901
Leading jet for qg2qh, Mean : 8.315
Subleading jet for gg2gh, Mean : 7.626
Subleading jet for qg2qh, Mean : 7.089

- After clustering into jets using FastJet (anti-kT, R=0.4), the variables we keep are the (pT, η, ϕ) of the Higgs, and the (pT, η, ϕ, Charged Multiplicity) of the ten hardest jets
- Feed this into a dense feed forward network, with two hidden layers (the first with 30 neurons, the second with 15 neurons)





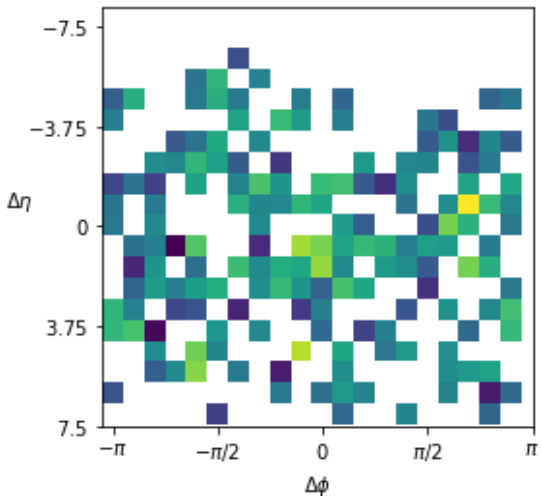
Possible Further Investigation:

- The effect of different jet clustering algorithms on classification performance (e.g. different Jet Radius, different choice of metric)

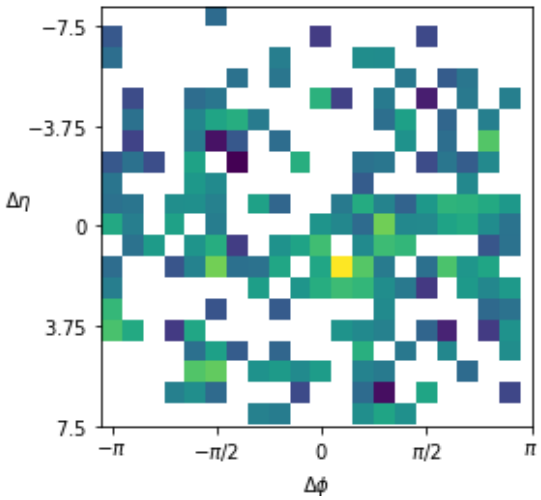
- Apart from using Jets, there have been two other ways of viewing the problem:
- Viewing the event as an image (Convolutional Neural Nets)- https://arxiv.org/pdf/1612.01551.pdf - Deep Learning in Color - towards automated quark/gluon jet discrimination (P.T. Komiske, E.M. Metodiev, M.D. Schwartz)
- Viewing the event as a structured sequence of particles (Recursive Neural Nets)- https://arxiv.org/pdf/1702.00748.pdf - QCD-Aware Recursive Neural Networks for Jet Physics (G Louppe, K Cho, C Becot, K Cranmer)

- We 'centered' all the particles around the ejected Higgs, in the sense that the particles pT were normalised by the Higgs pT, and the ϕ, η measured relative to the Higgs ϕ, η.
- Three colour channels:

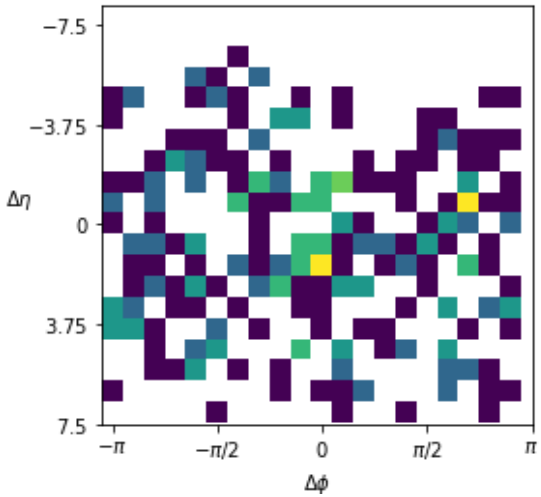
pT of charged particles



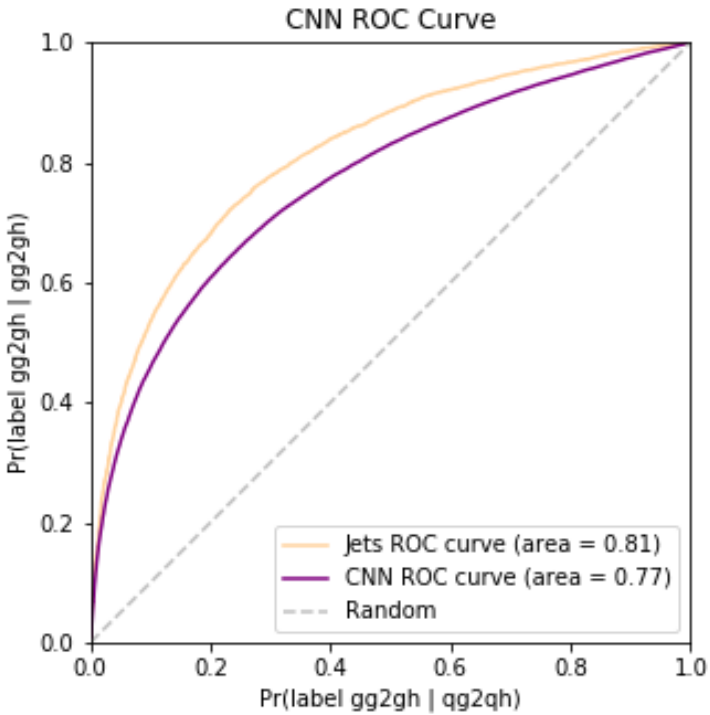
pT of neutral particles



charged particle multiplicity.



(a $gg \to gH$ event)



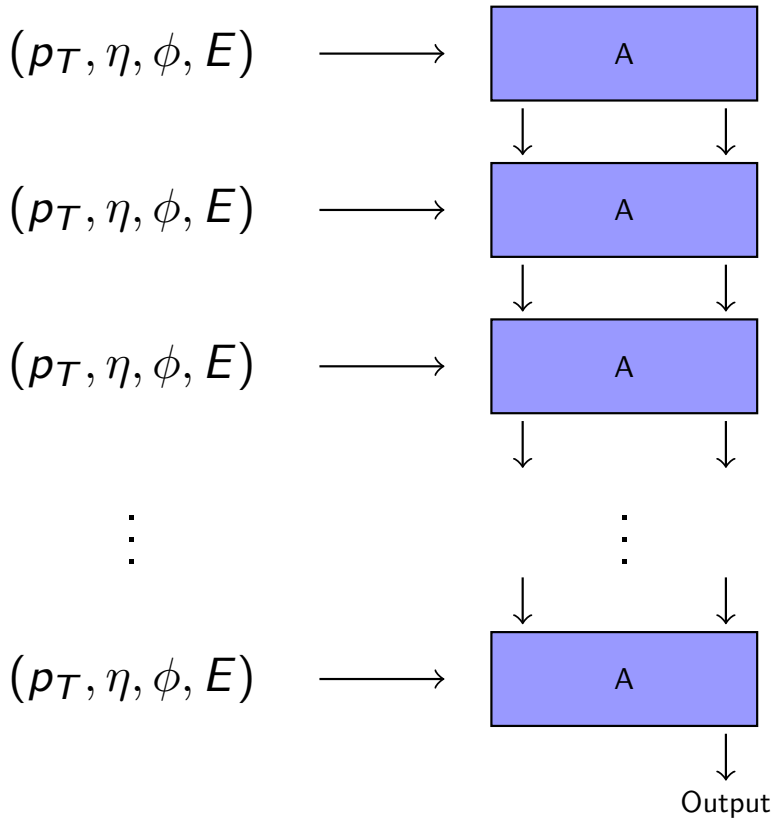
Possible Improvements

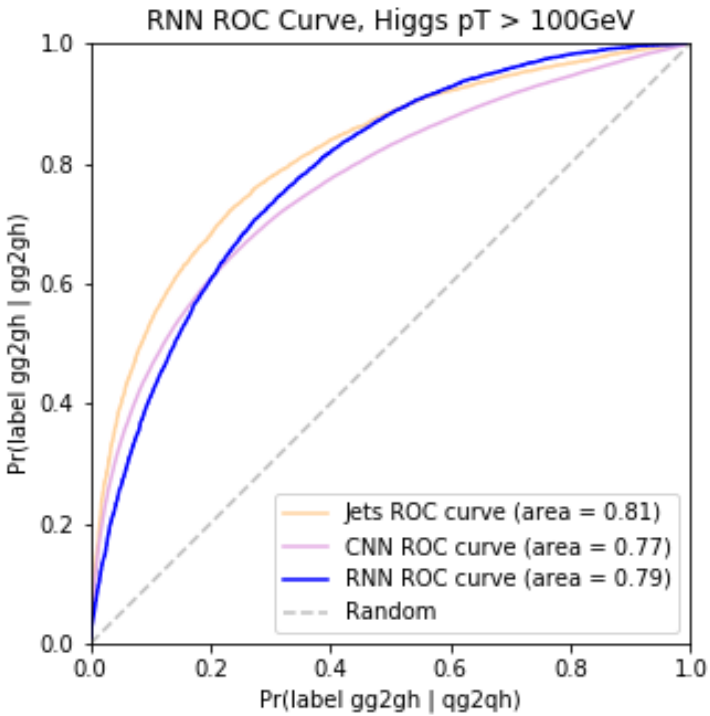
- More fine resolution for the jet images.
- Additional convolutional/pooling layers.

Possible Areas of Further Investigation

- Which variables can be used as additional colour channels?

- We tried a 'naive' implementation of an RNN to classifying our events; by feeding particle by particle (sorted by pT) into an RNN network. The particles 4-momenta were used as the information for the event
- We used LSTM neurons.



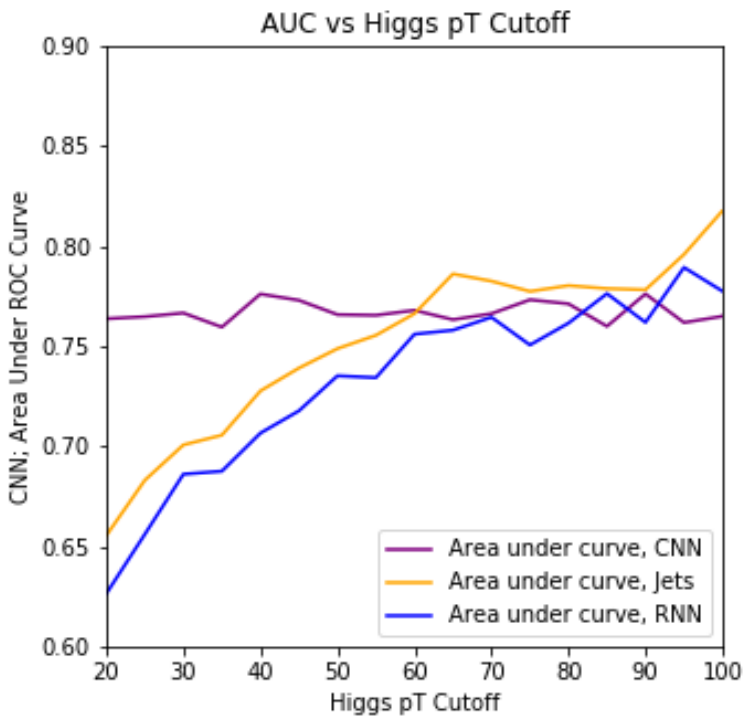


Possible Improvements

- Feeding the entire list of final particles; instead of the first few hardest particles. (We had a pT cutoff of 5GeV)

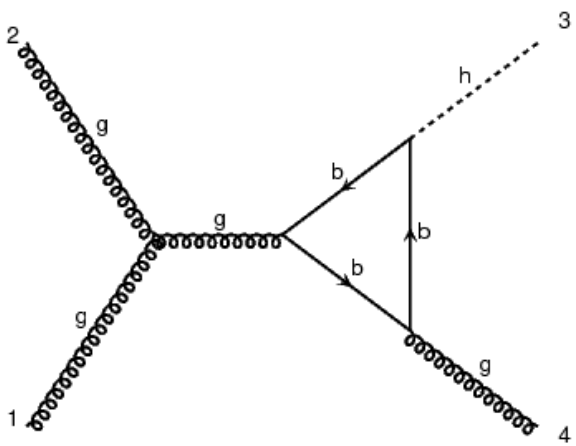
Possible Further Investigation

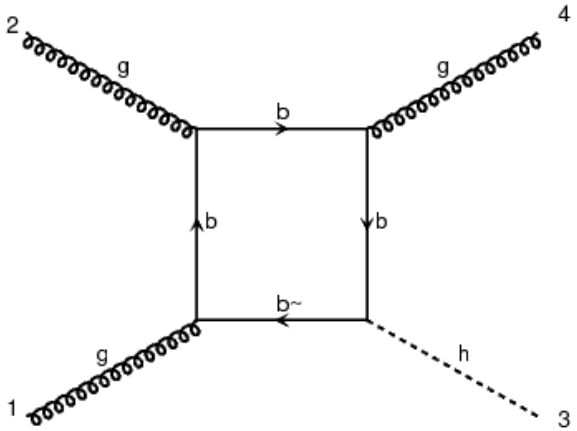
- Investigating the effects of different orderings for the particles - e.g. ordering by jet clustering algorithms

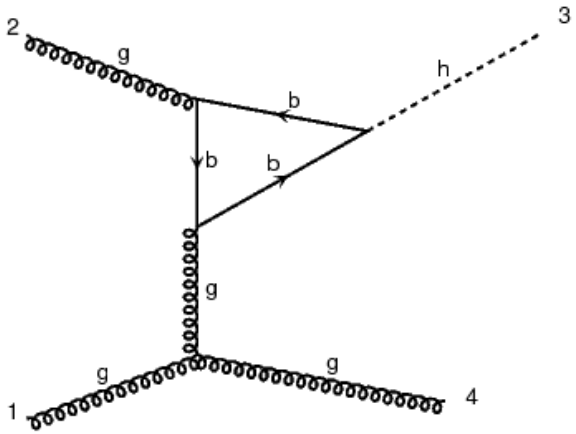


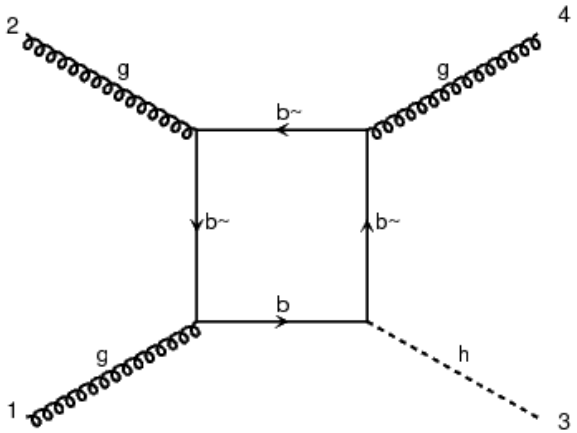
- Note that the performance of the CNN is roughly constant as a function of Higgs pT; whereas both the Jets method and the RNN increase performance as the Higgs pT increases. Perhaps this is because we normalised all our particles with respect to the Higgs; or perhaps just due to our low resolution of images.

Box vs Triangle diagrams in gg2gh

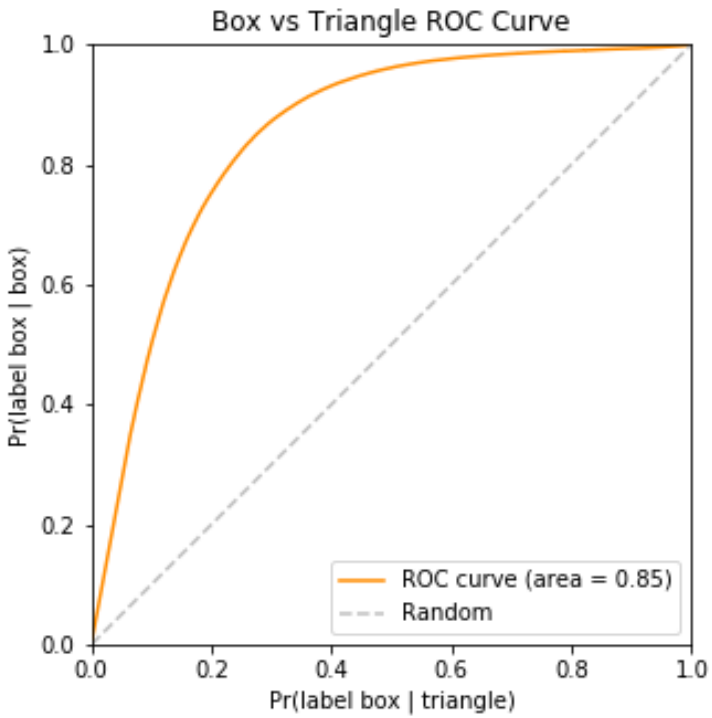








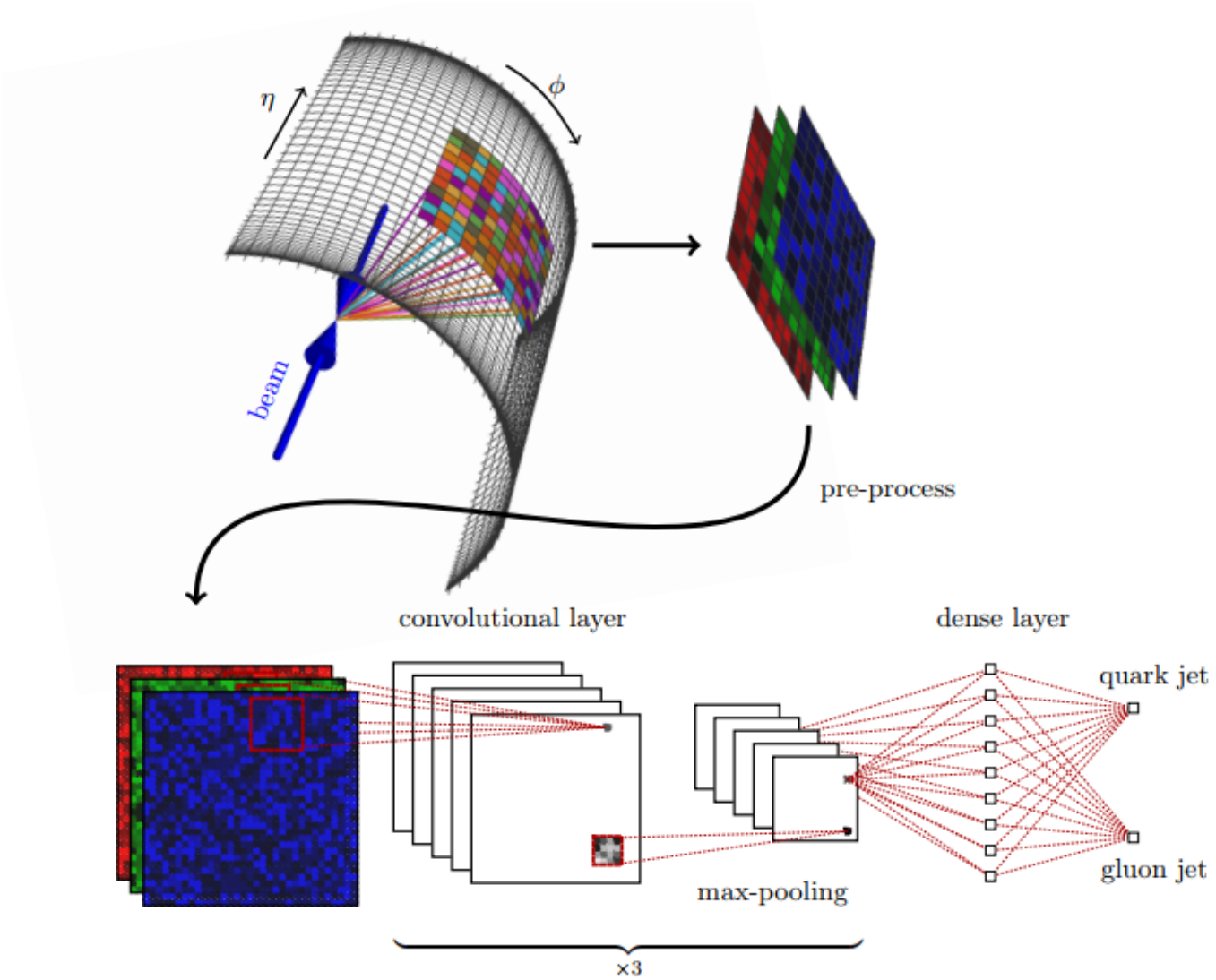
Box vs Triangle ROC curve



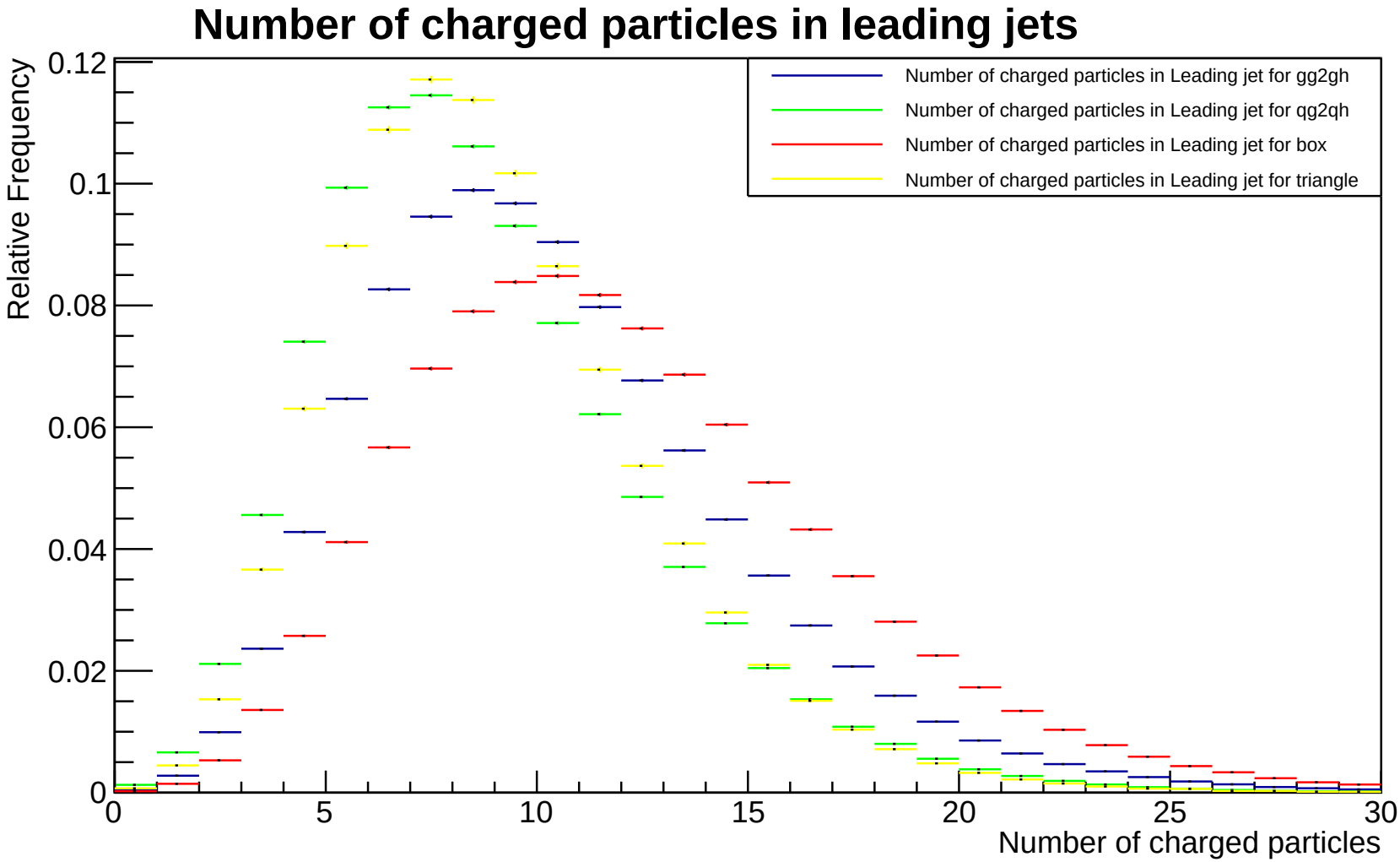
- Surprising results! We can distinguish between triangle diagram processes and box diagram processes fairly well!
- (Work in progress)

- We can look at ISR vs FSR radiation; and whether using one or the other is more effective in classification. ISR is usually considered a nuisance (background radiation) so it would be interesting to see it's capabilities.
- Investigation of Box vs Triangle diagrams; can we resolve these diagrams at moderate pT?
- Using the Machine Learning models developed to discover the important features used in classification.
- Estimate the actual sensitivity of our models to discovering New Physics; e.g. as a contribution to Higgs production that couples differently to boxes vs triangles.
- (Collaborating with Ian and Frank on further progress in classification)

- Backup Slides



https://arxiv.org/pdf/1612.01551.pdf


Number of charged particles in leading jets
Relative Frequency
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
Number of charged particles
Number of charged particles in Leading jet for gg2gh
Number of charged particles in Leading jet for qg2qh
Number of charged particles in Leading jet for box
Number of charged particles in Leading jet for triangle