Quark and Gluon Jet Discrimination

Matthew Schwartz
Harvard University

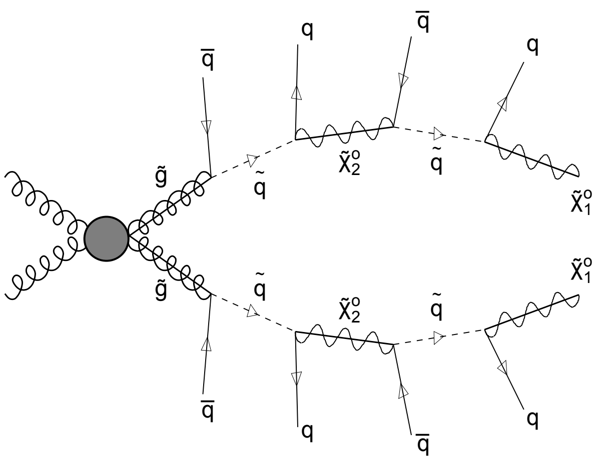
Based on work with J. Gallicchо arXiv:1211.7038. 1106.3076, 1104.1175
and P. Komiske and E. Metodiev arXiv:1612.01551

See also ATLAS arXiv:1405.6583, ATLAS-CONF-2016-034
Larkoski et al. arXiv:1405.6583
Schwartzman et al. arXiv:1407.5675, arXiv:1511.05190

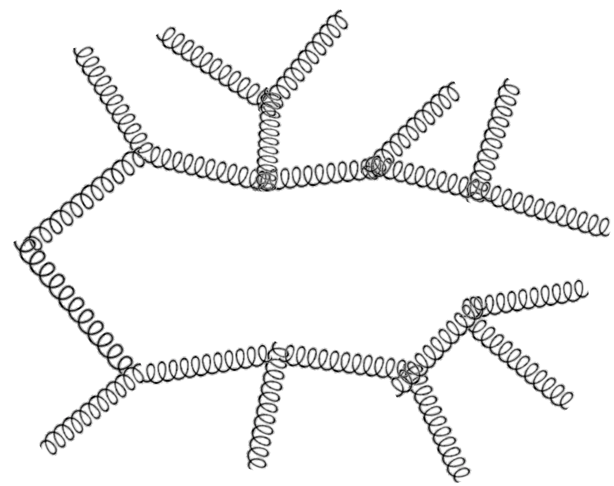
Why do we care?

1. BSM searches:

New physics mostly **quark jets**



Backgrounds mostly **gluon jets**



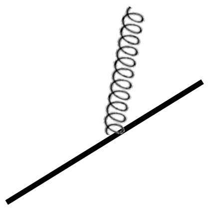
2. SM searches
   - Gluonic backgrounds to e.g. hadronic top decays

3. Improve Monte Carlos
   - Gluon jet modeling limits accuracy of current simulations

4. Test precision QCD
5. For the challenge: can we do it?

Quark/Glue basics

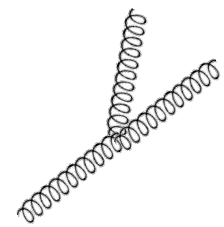
Probability of quark radiating:



$$P(q \to qg) = \frac{\alpha_s}{2\pi} C_F(\cdots)$$

$$C_F = \frac{4}{3} = 1.3$$

Probability of quark radiating:



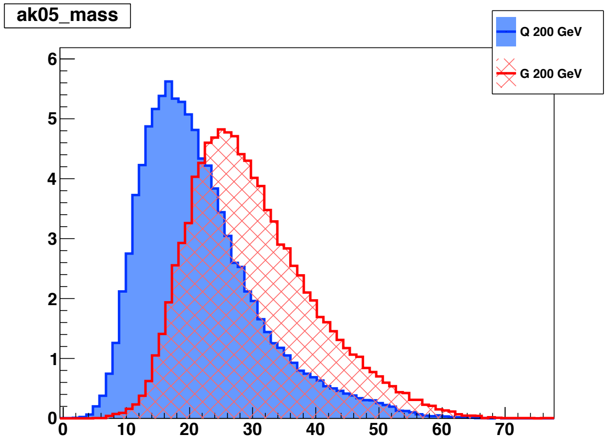
$$P(g \to gg) = \frac{\alpha_s}{2\pi} C_A(\cdots)$$

$$C_A = 3$$

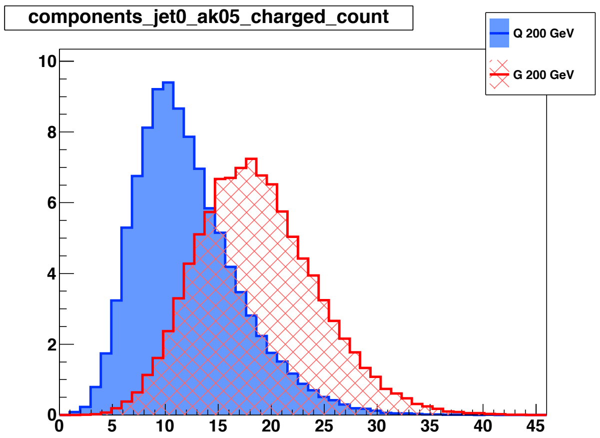
- Gluons around twice as likely to radiate than quarks
  - Gluon jets are fatter
  - Gluon jets are more massive
  - Gluon jets have more particles
  - …

Example distributions

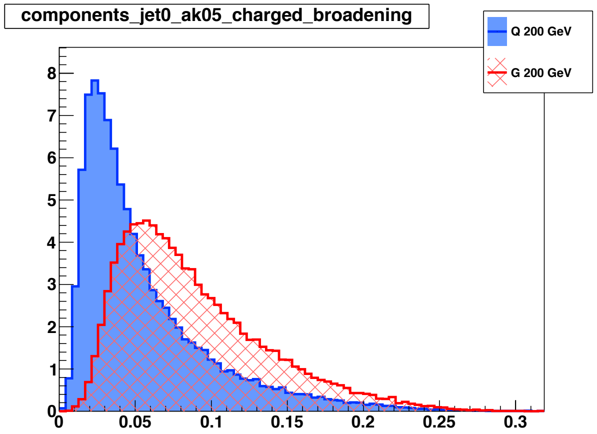
Jet mass



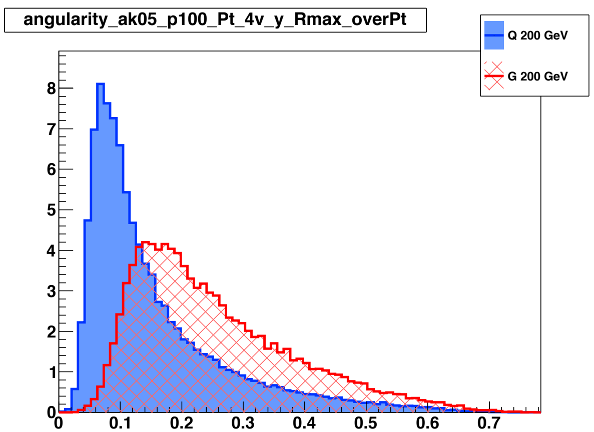
Charged particle count



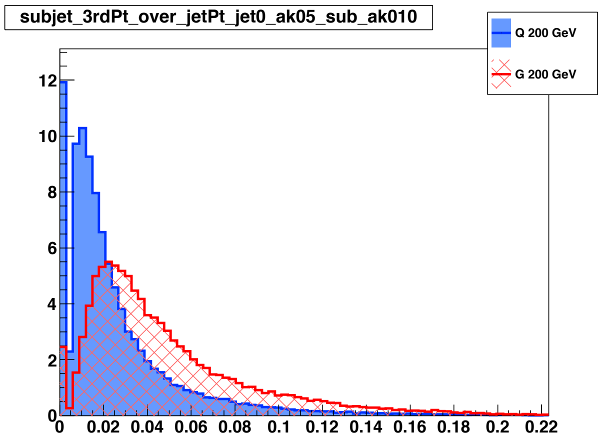
Jet broadening



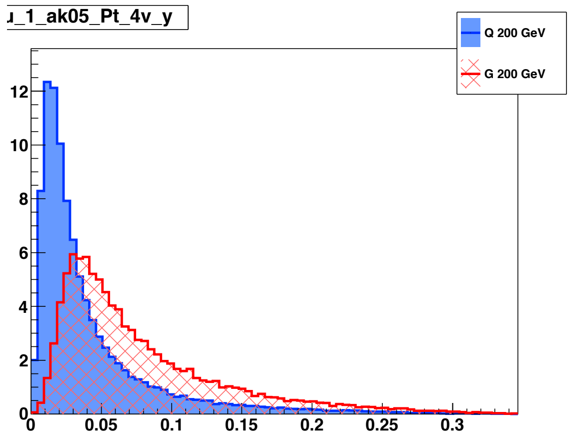
Jet angularity



P_T fraction of 3rd hardest subjet

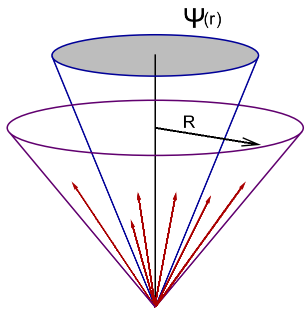


Moment of Hu

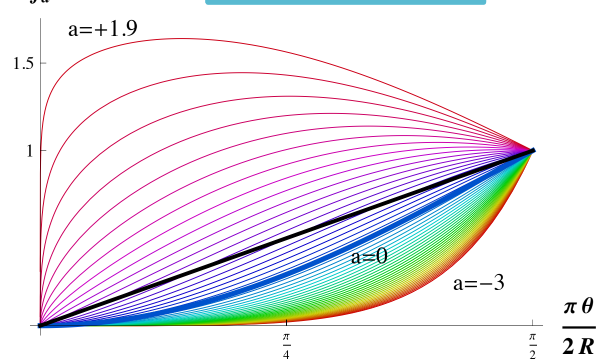


We looked at 10,000 variables

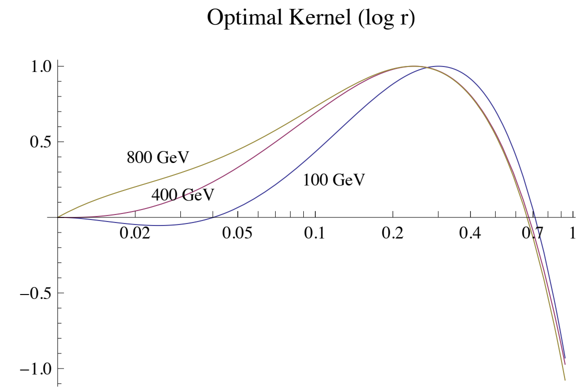
Integrated/differential "Jet Shape"



Jet angularities



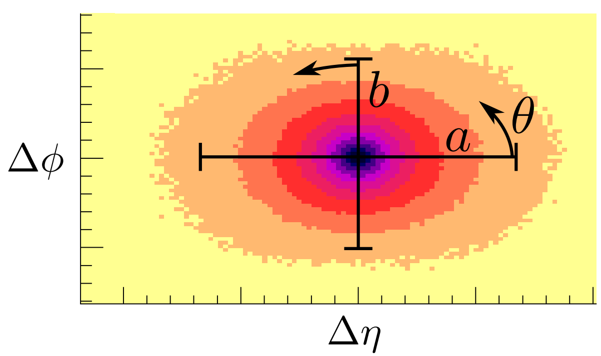
Iteratively optimized radial profile



Subjet counts and properties

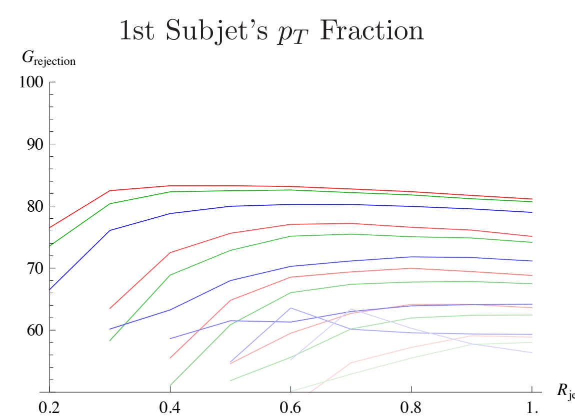
Properties of Covariance tensor

$$\mathbf{C} = \sum_{i\in \text{jet}} \frac{p_T^i}{p_T^{jet}} \begin{pmatrix} \Delta\eta_i\Delta\eta_i & \Delta\eta_i\Delta\phi_i \\ \Delta\phi_i\Delta\eta_i & \Delta\phi_i\Delta\phi_i \end{pmatrix}$$



Combination of Eigenvalues

Eigenvalues:	$a > b$
Quadratic Moment:	$g = \sqrt{a^2+b^2}$
Determinant:	$det = a \cdot b$
Ratio:	$\rho = b/a$
Eccentricity:	$\epsilon = \sqrt{a^2-b^2}$
Planar Flow:	$pf = \frac{4ab}{(a+b)^2}$
Orientation:	θ



http://jets.physics.harvard.edu/qvg

We looked at 10,000 variables

The best two variables in Pythia are:

1 Charged particle count

- Better spatial and energy resolution works better
  - e.g. particles > calorimeter clusters > subjets

and

2 Linear radial moment (girth)

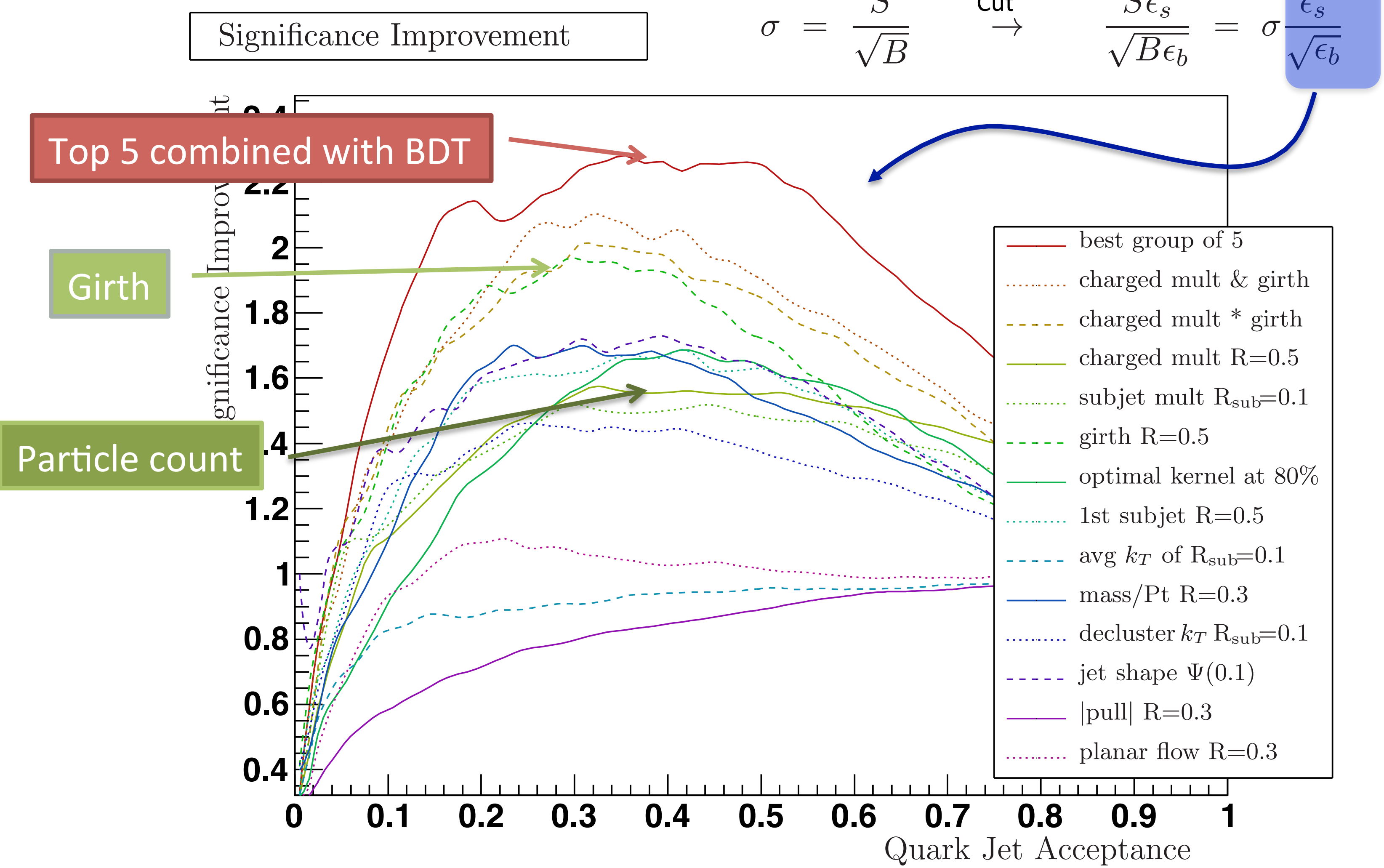
- Similar to jet broadening

$$g = \frac{1}{p_T^{\text{jet}}} \sum_{i \in jet} p_T^i |r_i|$$

- Many variables have similar performance

Quark and gluon jet substructure

$$\sigma = \frac{S}{\sqrt{B}} \xrightarrow{\text{Cut}} \frac{S\epsilon_s}{\sqrt{B\epsilon_b}} = \sigma\frac{\epsilon_s}{\sqrt{\epsilon_b}}$$

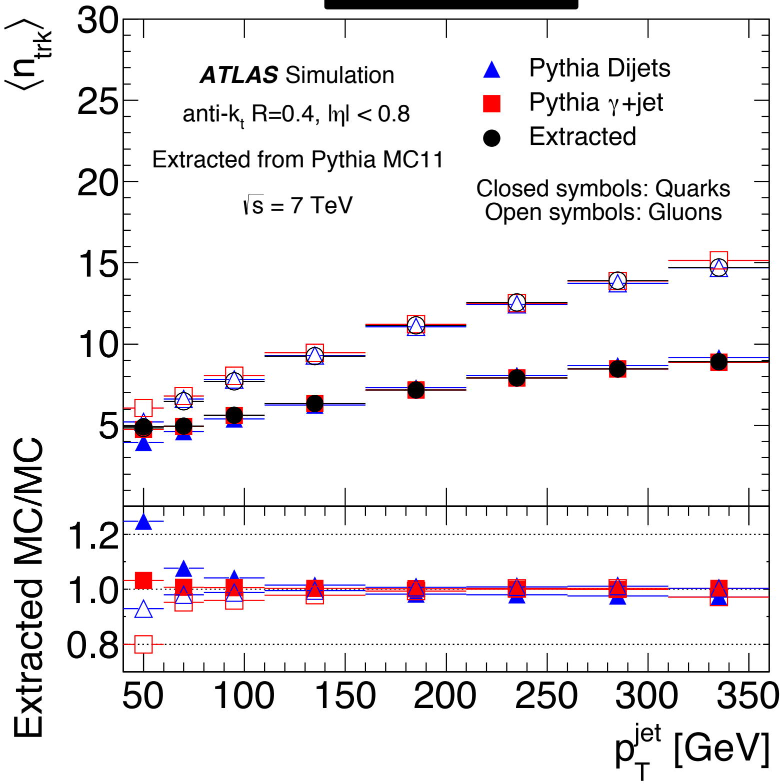


100 GeV jets, Pythia

ATLAS 7 TeV

- ATLAS developed procedure to disentangle quark and gluon jets
- Used relatively pure samples (dijets for gluon, γ + jet for quark)

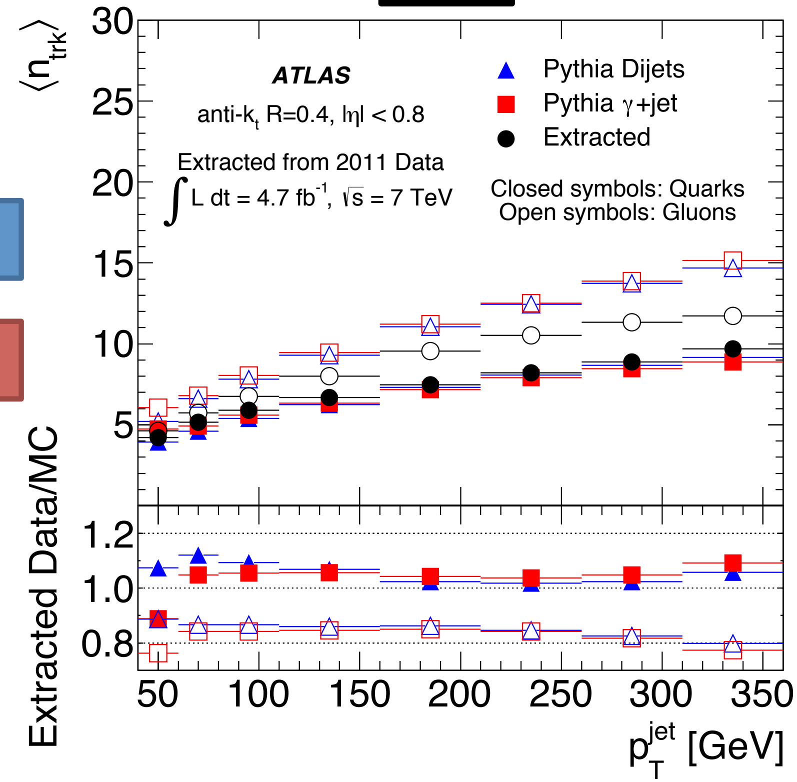
Simulation



gluon

quark

Data

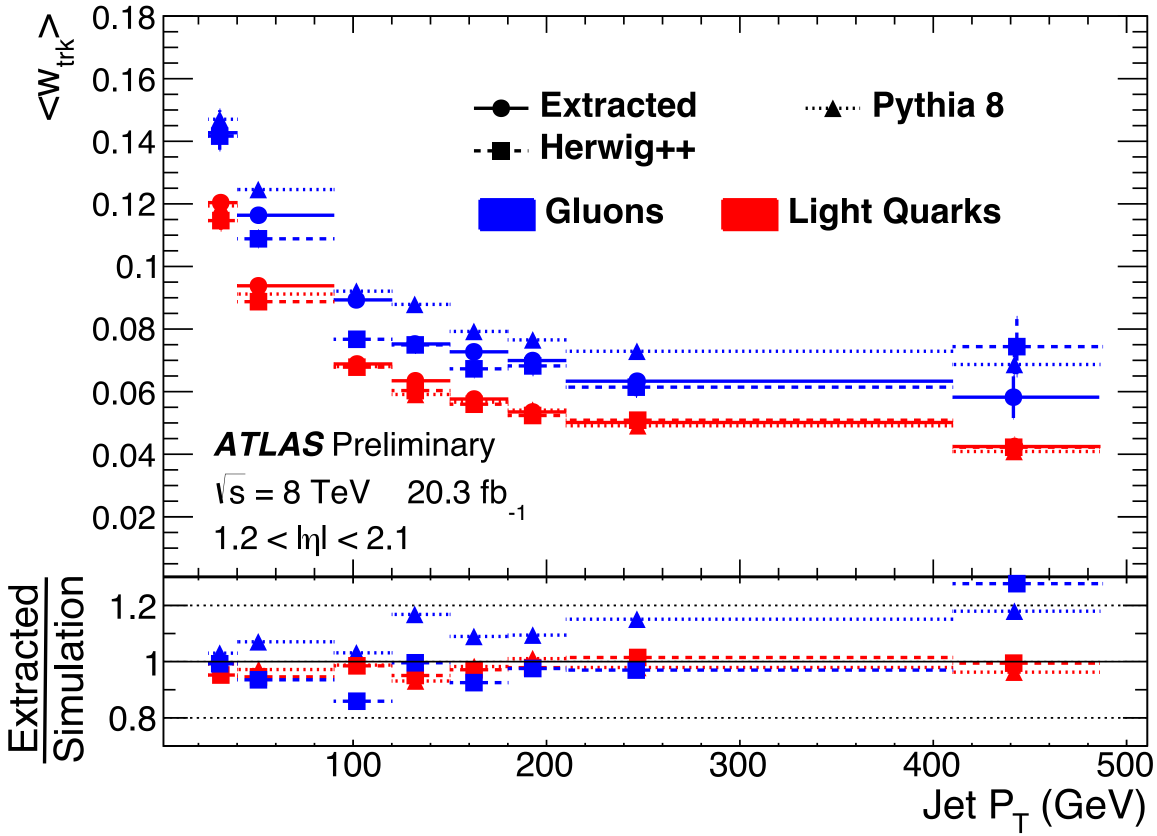


Extracted gluon jet properties closer to quark than in pythia

ATLAS 8 TeV

ATLAS-CONF-2016-034

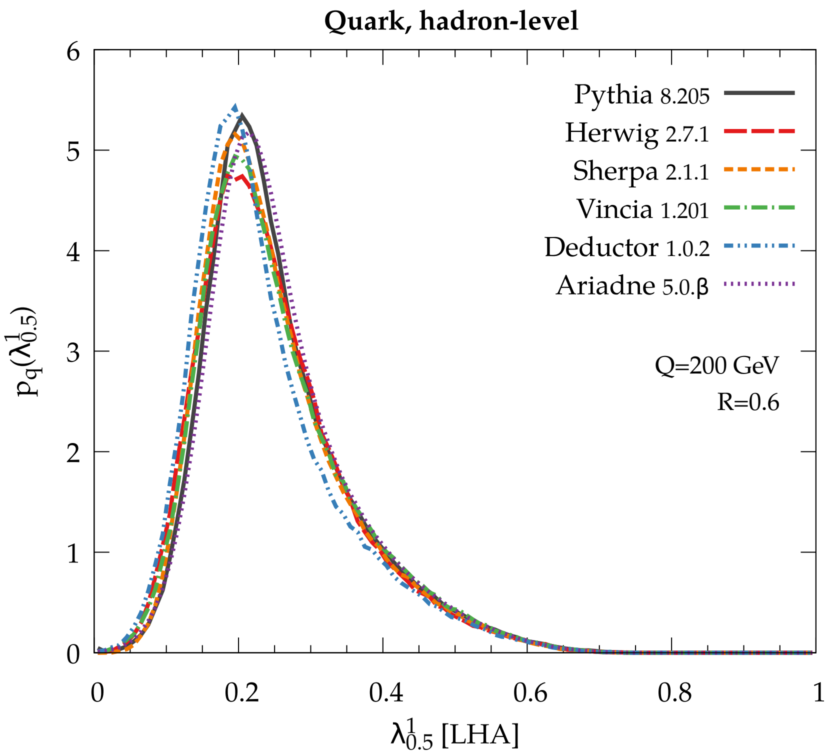
Track width

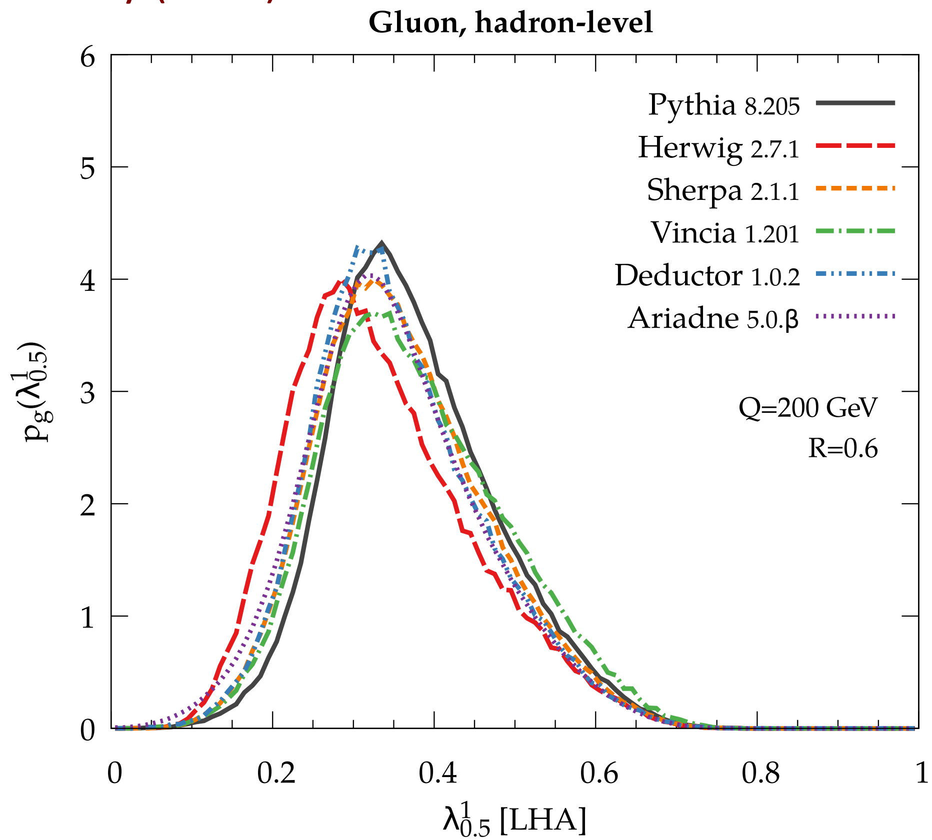


Data appears to be between Pythia and Herwig

Monte Carlos can be improved…

Les Houches study (2016)



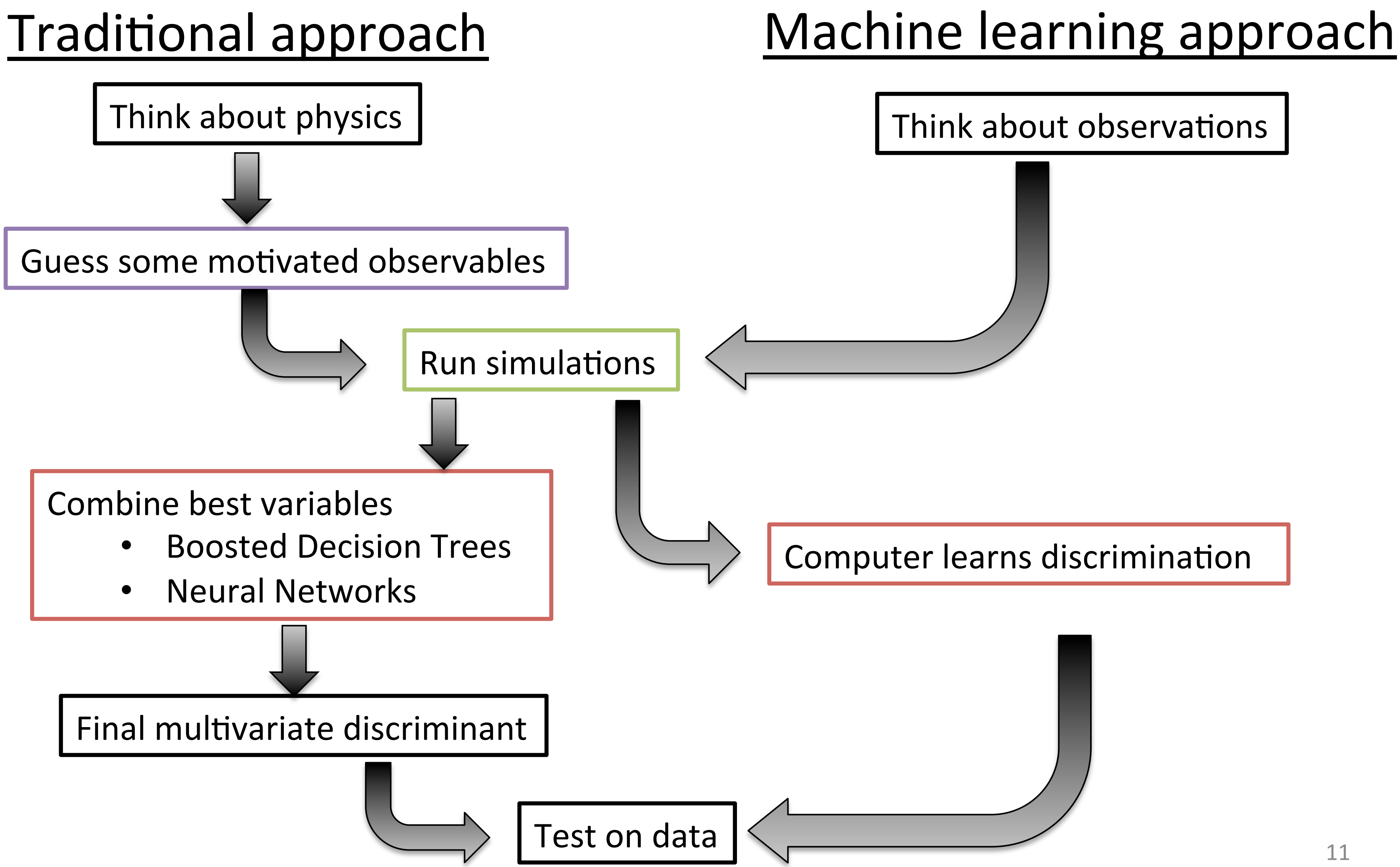


Monte Carlo's all seem to agree on quarks

Improved shower MCs in between Herwig and Pythia

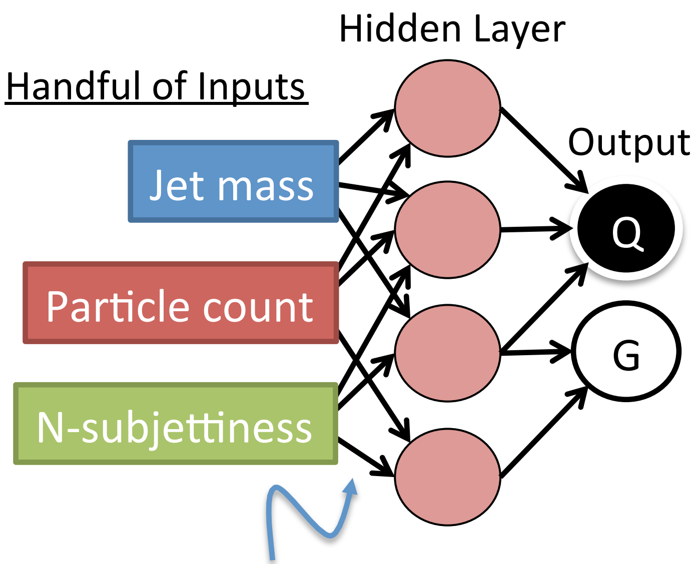
- Promising sign for future progress

Deep learning



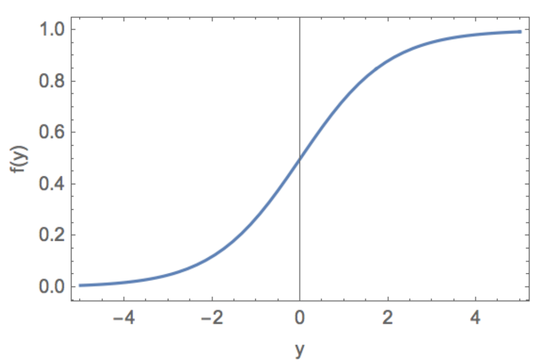
Neural networks

Traditional (shallow) neural networks
Useful for multivariate analysis

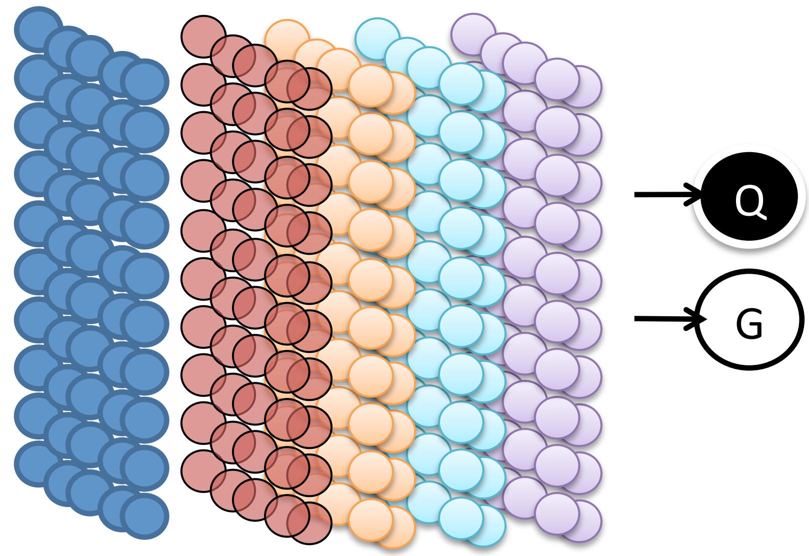


$$f(y) = \frac{1}{1+\exp(-y)}$$ (sigmoid)

Activation function inspired by biology



Deep networks



- Many inputs
- Many hidden layers

Recent advances allowing Deep Learning

- In algorithms:
  - New activation functions to avoid issues such as saturation
  - New model regularizations
    - **Dropout**: Randomly selected fraction *p* of units are ignored during each weight-update.
  - Network architecture adapted to application
    - Drastically fewer elements to optimize

- In computing:
  - Faster computing capabilities
    - Graphics Processing Units (GPUs)
  - Easier Usability
    - Keras Deep Learning Python Library

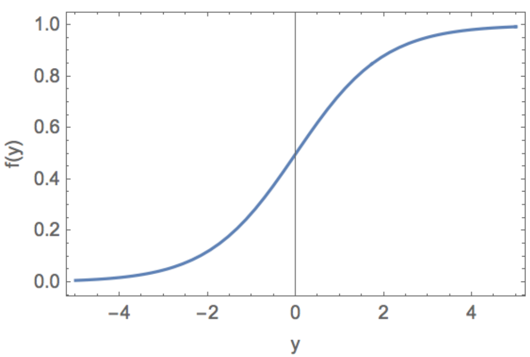
Activation Functions

Traditional

New!

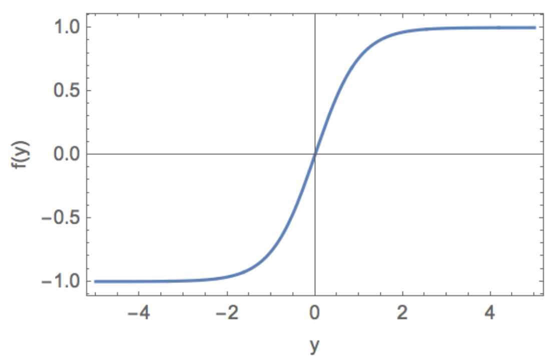
- Sigmoid

$$f(y) = \frac{1}{1+\exp(-y)}$$



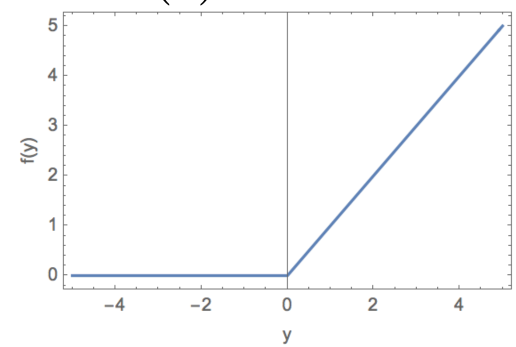
- Tanh

$$f(y) = \tanh(y)$$



- Rectified Linear Unit (ReLU)

$$f(y) = \max(0, y)$$

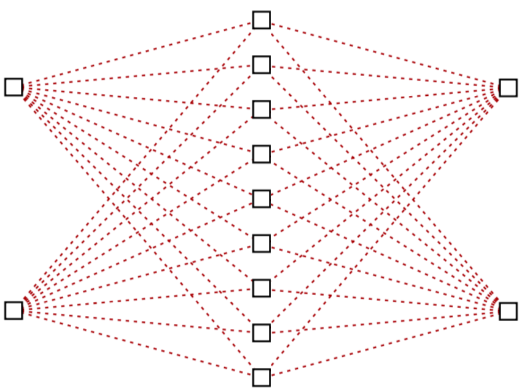


- Sigmoid and Tanh can **saturate**, whereby the gradient becomes vanishingly small for inputs far from zero, making training difficult.

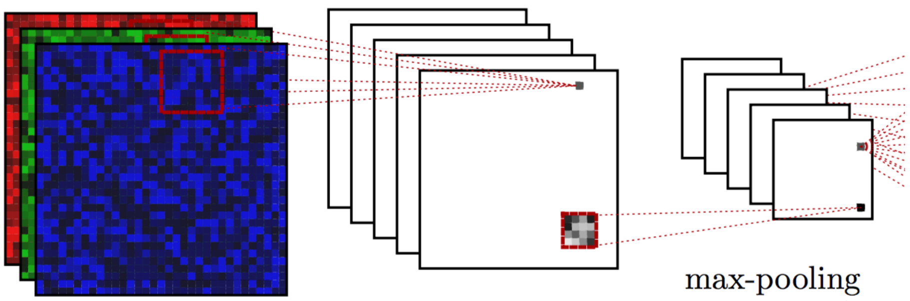
- ReLUs avoid this saturation problem and have an easily computable gradient.

How to link up units?

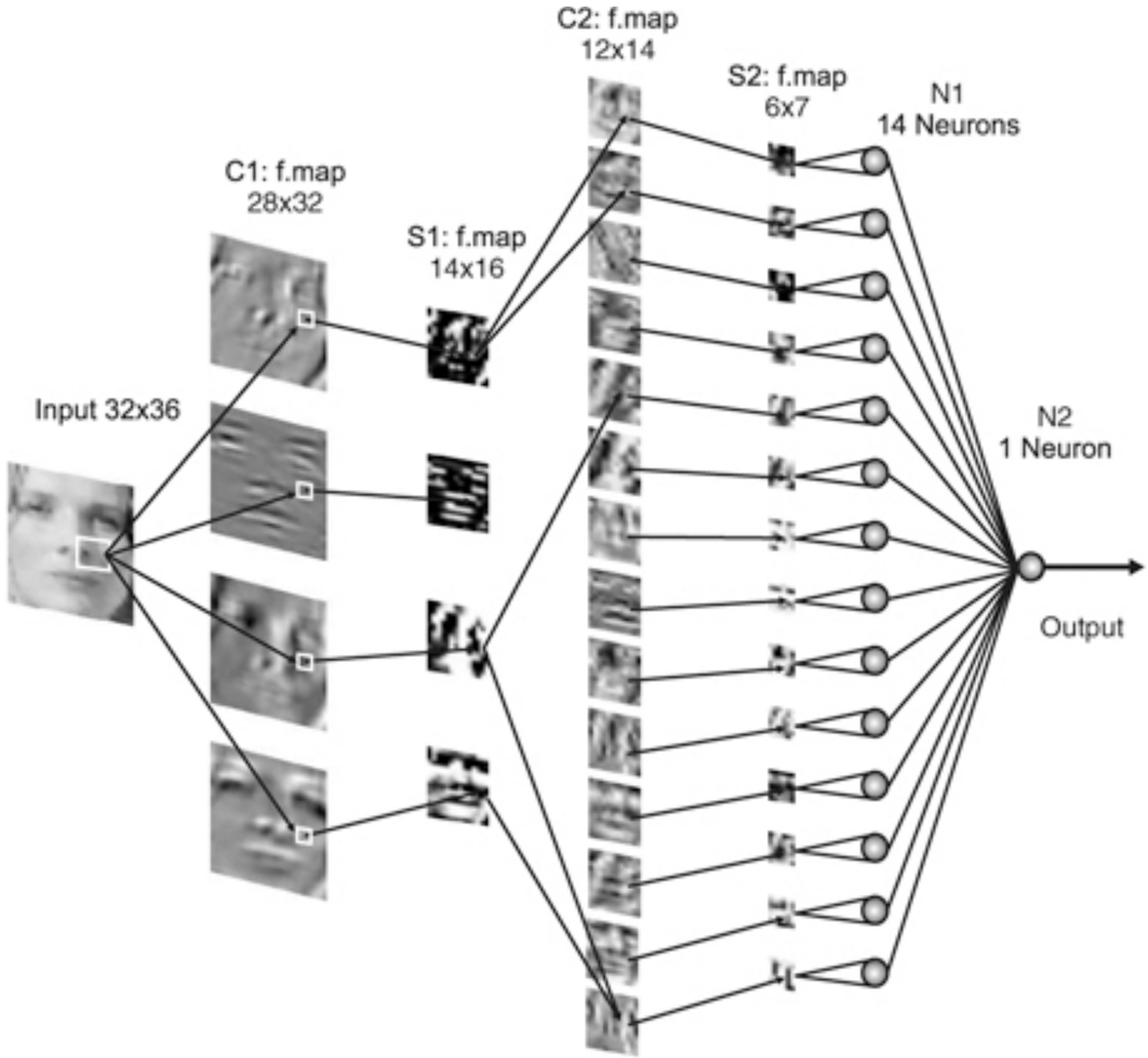
- Dense (Fully Connected) Layer
  - Each unit is linked to every unit in the previous layer.



- Convolutional Layer
  - Each unit is linked to an $n \times n$ patch of the previous layer.
  - Units are downsampled to $n/2 \times n/2$ patches with a **max-pooling** layer.
  - Can handle multiple **channels**.
    E.g. RGB images.



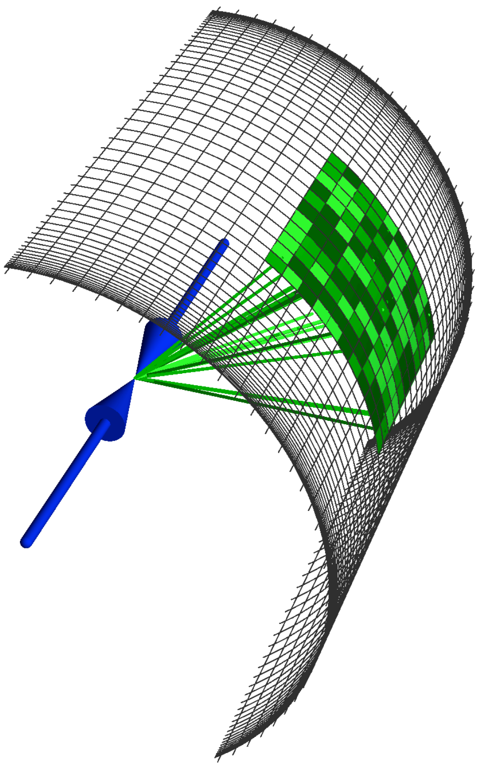
Applications to Image Recognition



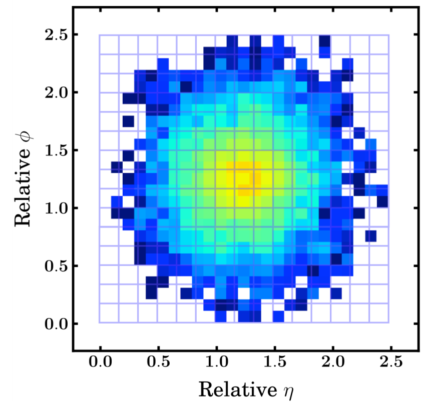
Jet Images

Cogan et al. arXiv:1407.5675

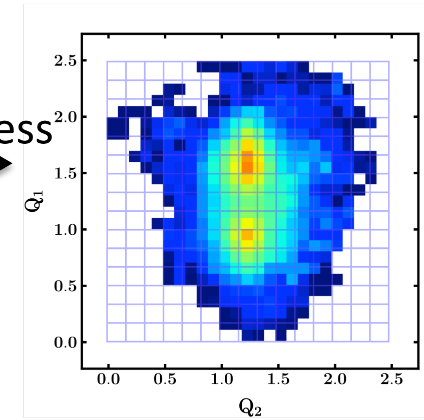
- Treat energy deposits as image

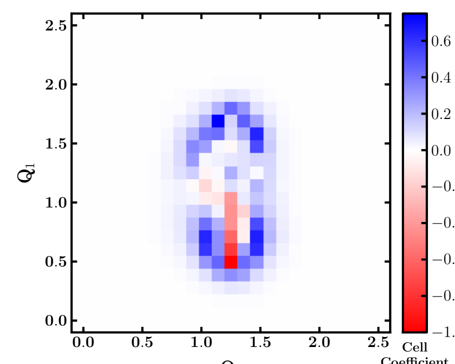


Application to boosted W tagging



preprocess

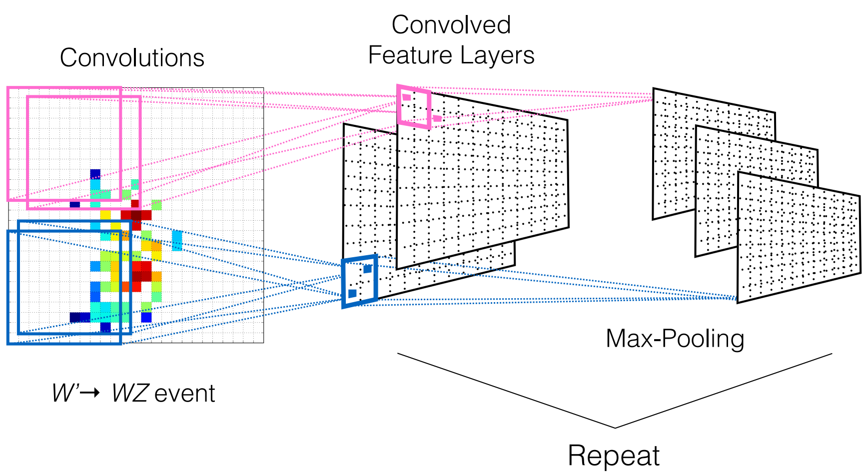


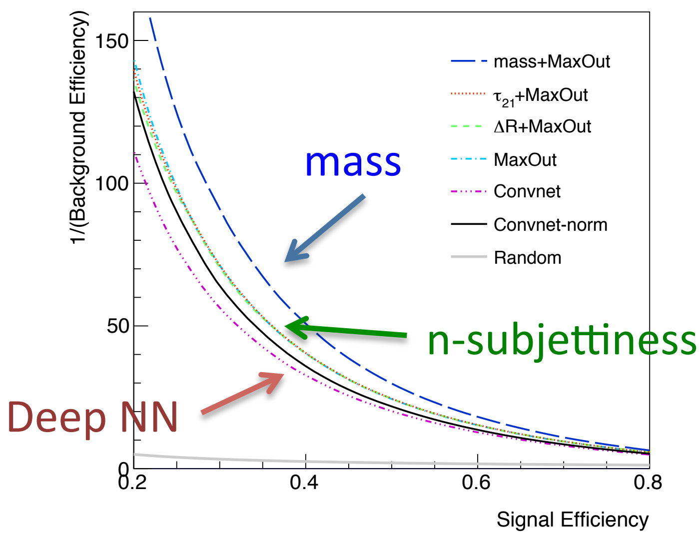


Fisher discriminant

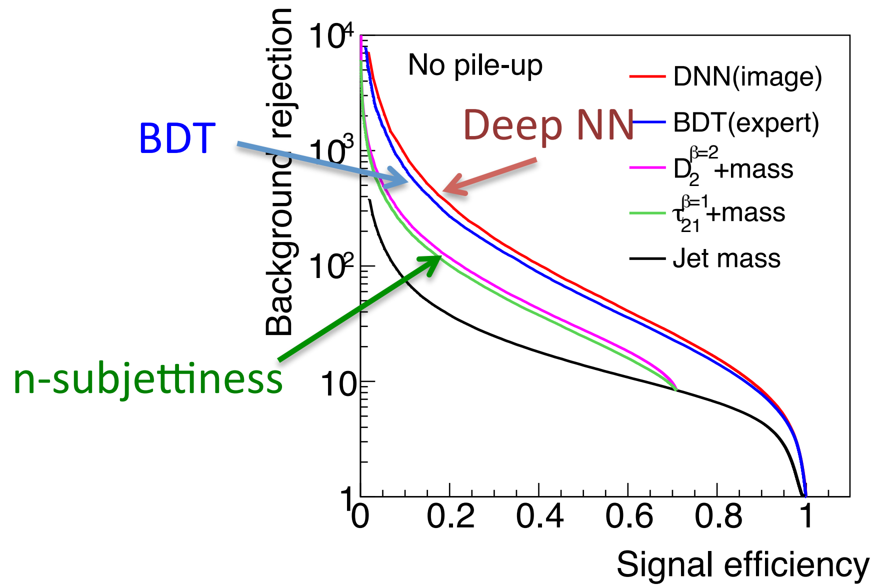
Boosted W's and jet images

de Olivera et al. arXiv:1511.05190



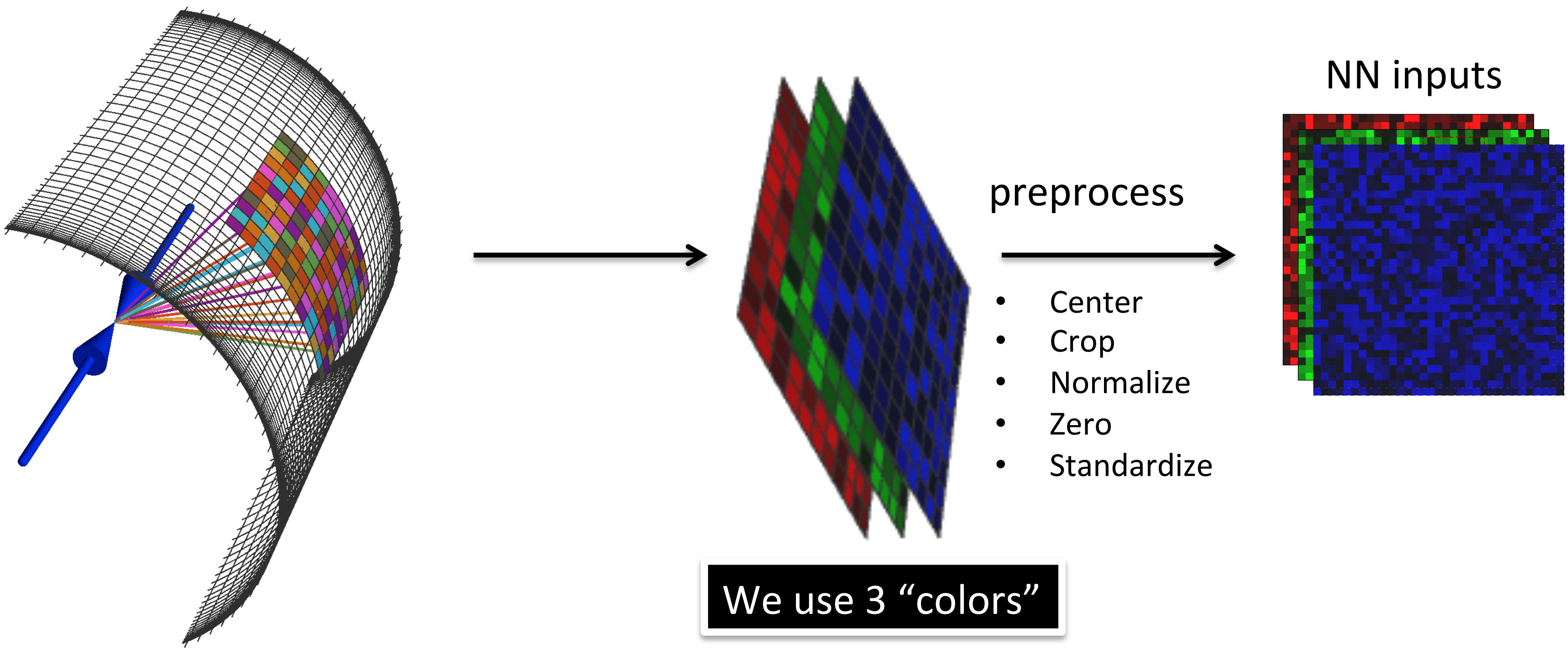


Baldi et al. arXiv:1603.09349



- Deep NN does better than single variables
- Deep NN does about as well as BDT of 6 good discriminants

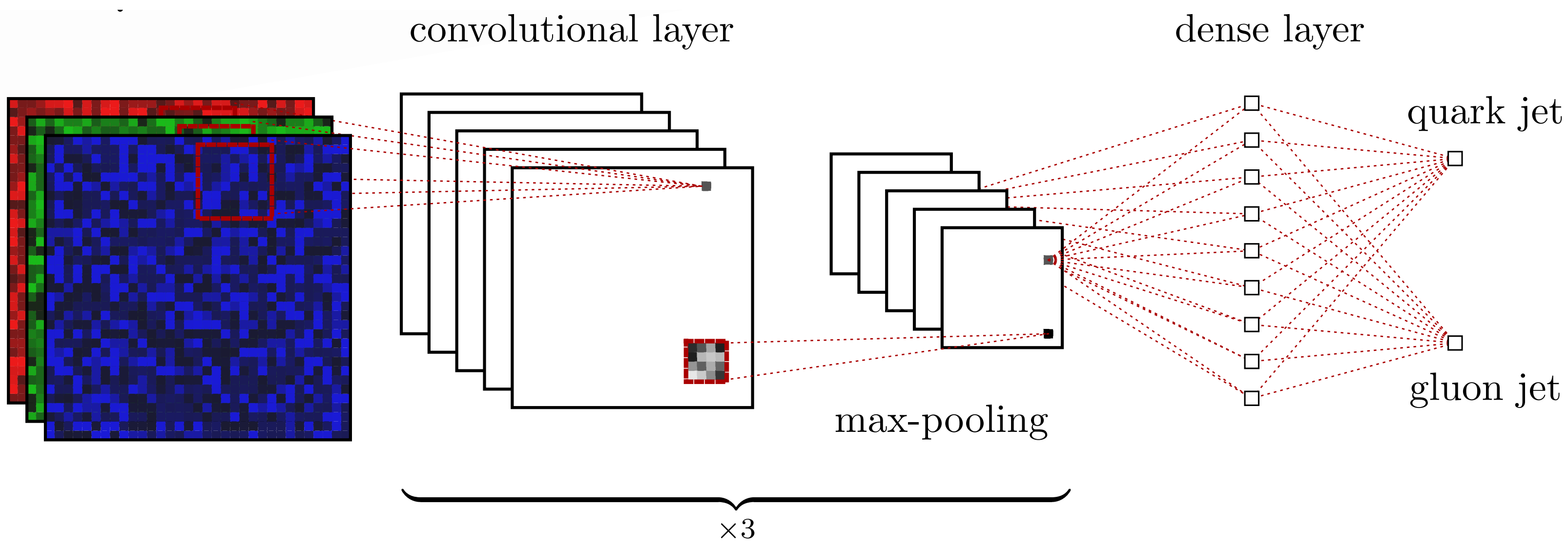
Deep learning for Q vs G



- Find anti-k_T R=0.4 jets
- Extract square grid around jet center
- Pixelate into $\Delta\eta \times \Delta\phi = 0.024 \times 0.024$ cells
  - Produces 33x33 image

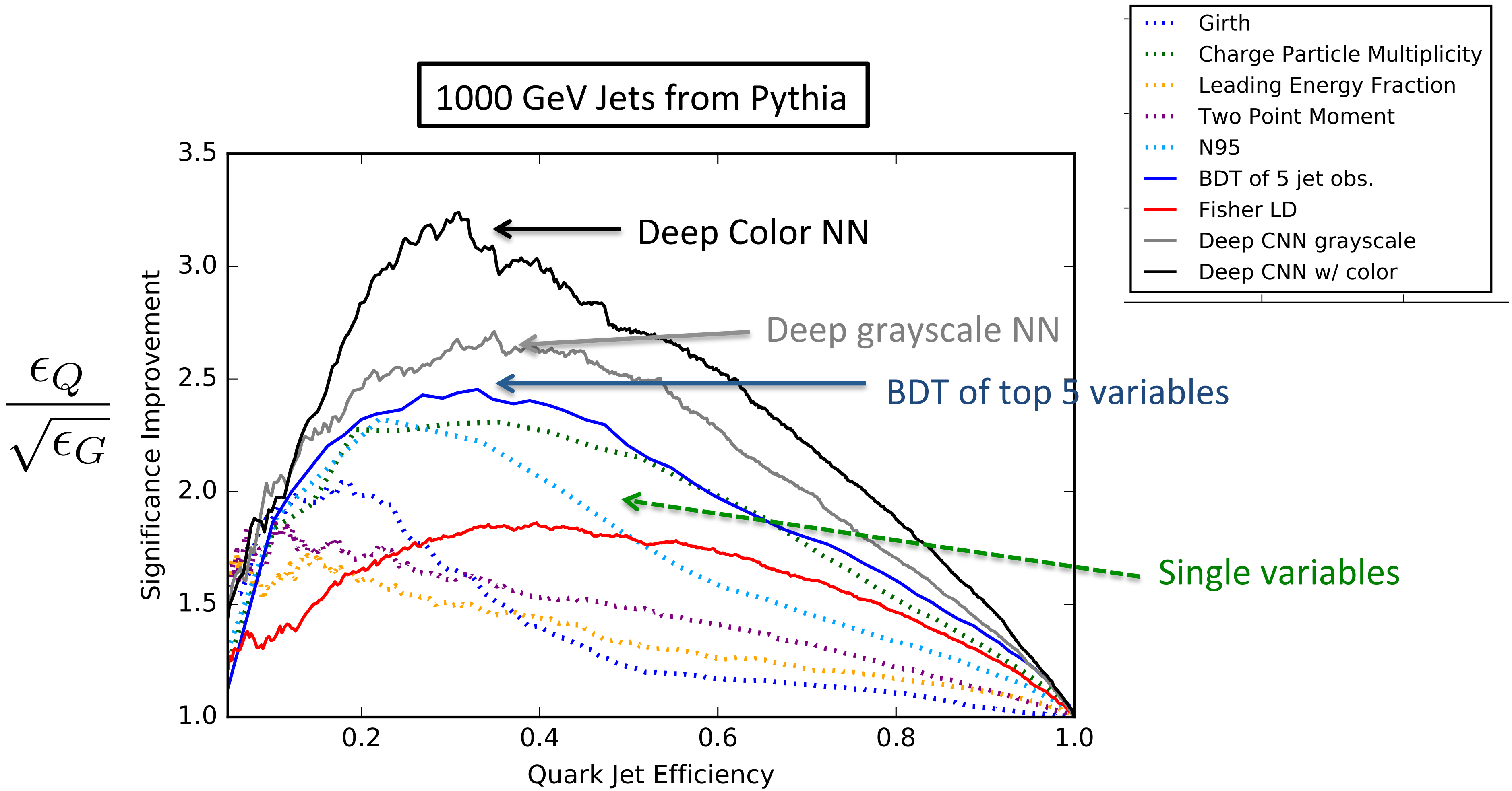
- Red = p_T of charged particles
- Green = p_T of neutral particles
- Blue = charged particle multiplicity

Deep NN architecture



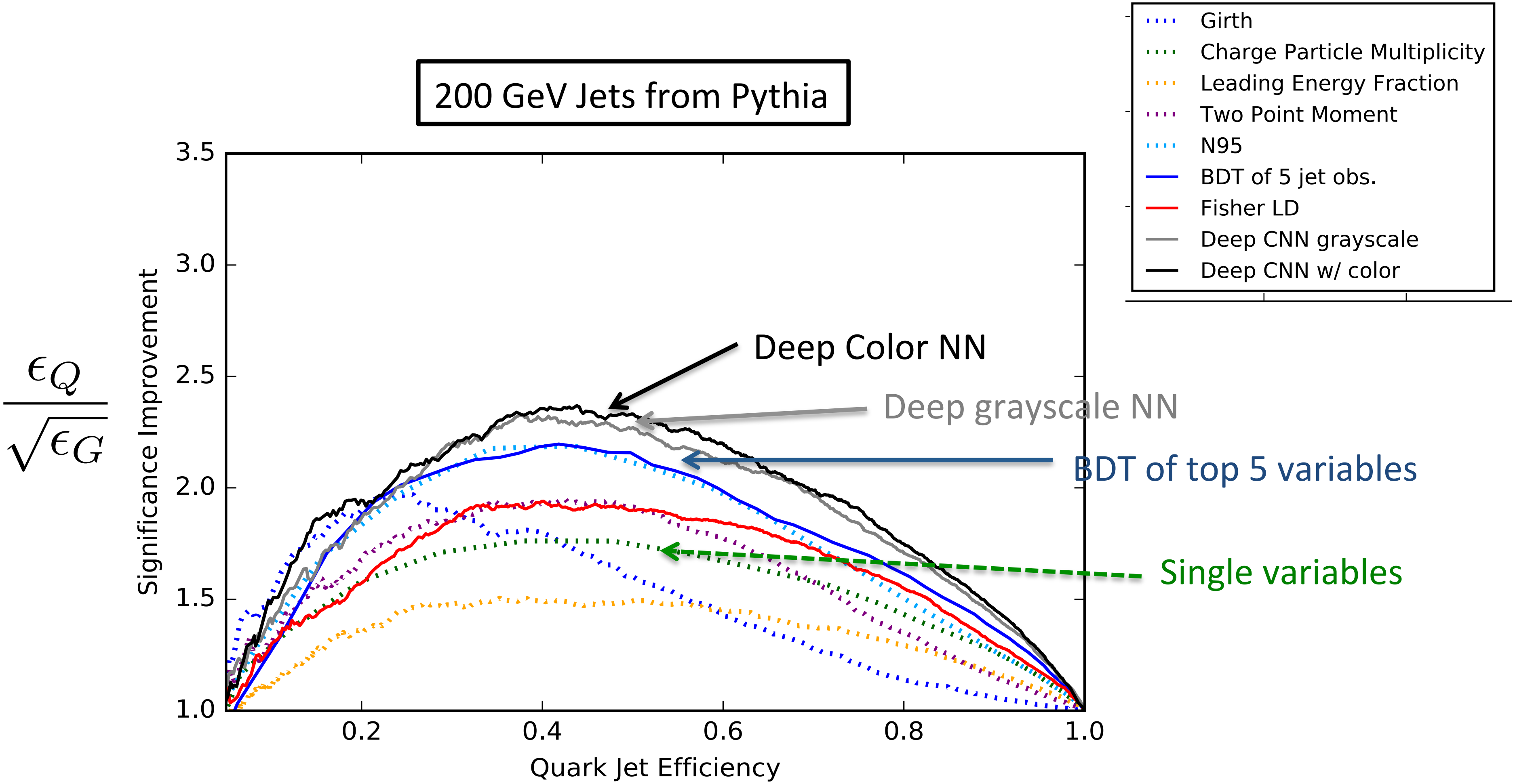
- Convolution layers apply 8x8 pixel filters to images
    - 4x4 filters used for for 2nd and 3rd conv. layers
    - We use 64 independent filters in each layer
- Max-pooling reduces layer size by 4
- Final layer is densely connected to all final filters

Results



Works really well – especially considering we don't put in any physics!

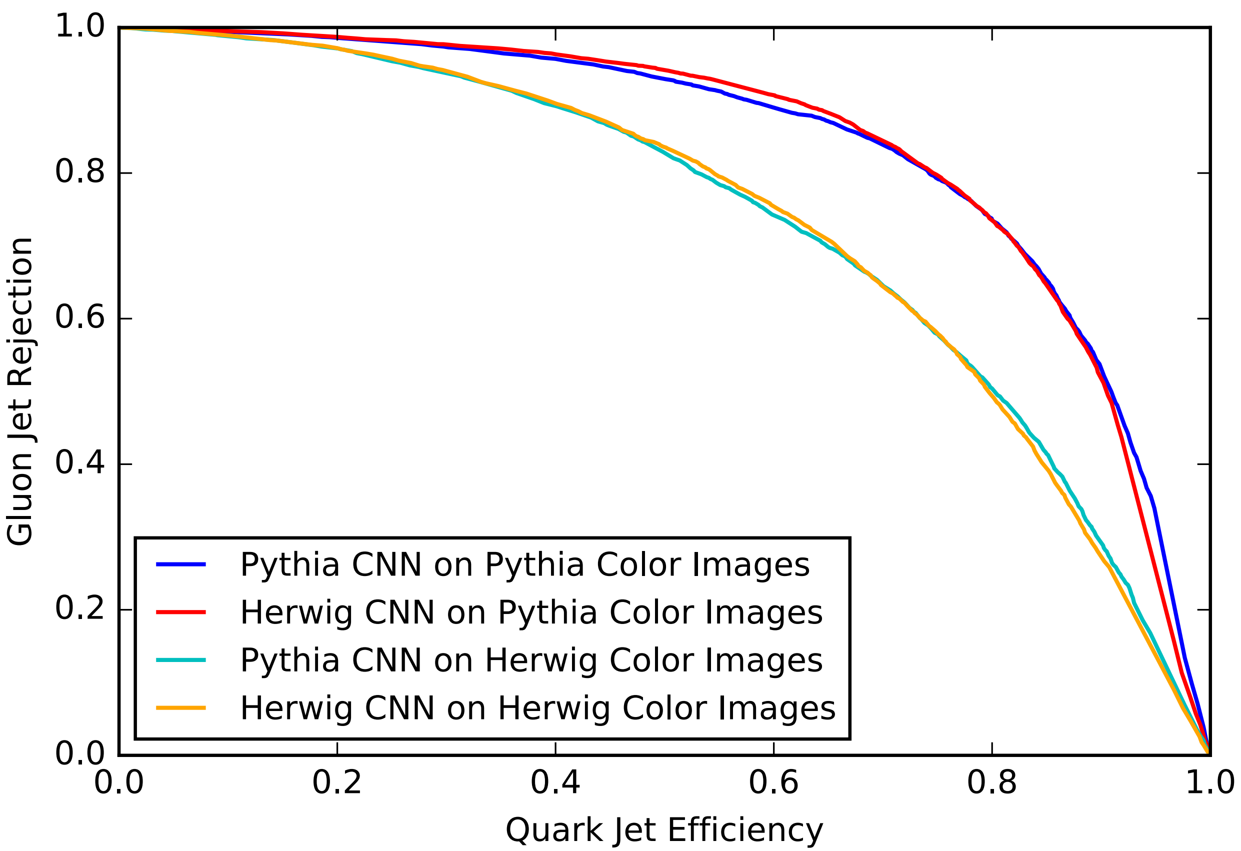
Results



Works really well – especially considering we don't put in any physics!

Comparing Pythia and Herwig

- Discrimination worse in Herwig
  - Gluon and quark jets are more similar
  - Consistent with previous studies



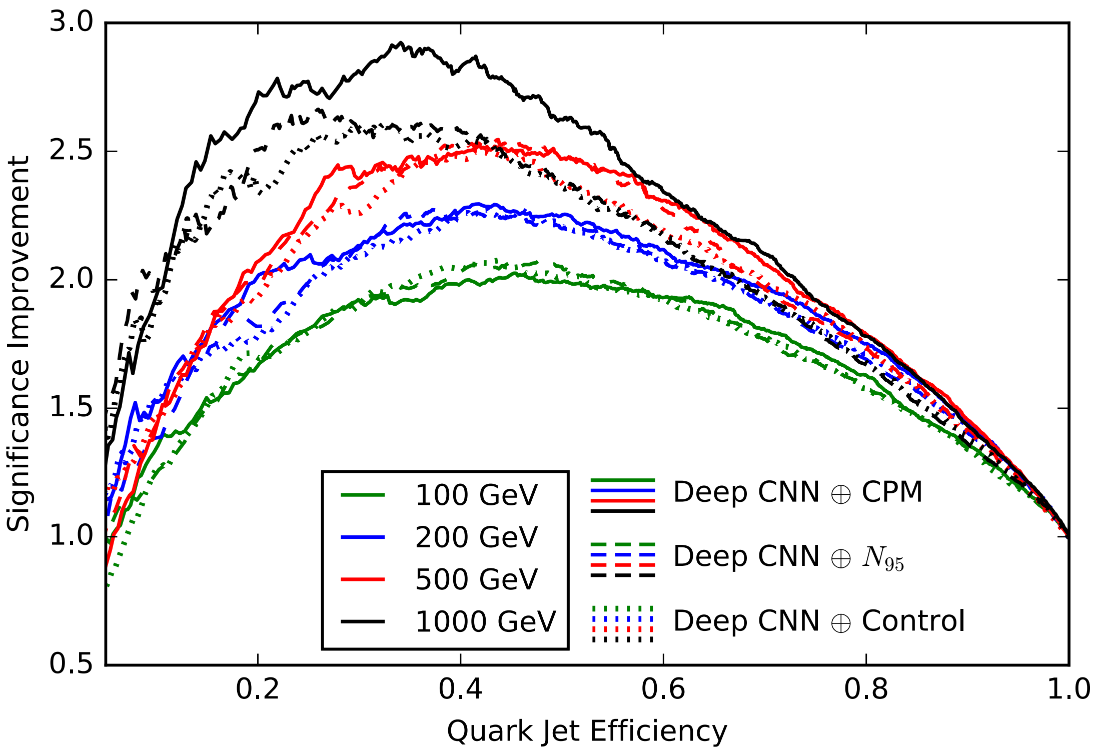
Network performance independent of MC used to train

- Indicates robustness
- May work on data

Is it learning physics?

Add in observables

- CPM = charged particle multiplicity
- N_{95}= a useful discriminant (minimum number of pixels with 95% of jet p_T) [Pumplin 1991]



Except at very high p_T,
no benefit from adding observables

May indicate that NN has "learned" physics

Conclusions

- **Quark and gluon jets can be distinguished** by radiation patterns
  - Pythia and Herwig have significant differences, particularly for gluons
  - Improved parton showers (e.g. vincia) look promising

- **Traditional variables**
  - Two types: **shape** (mass, girth, n-jettiness) and **count** (# particles, # subjets)
  - Marginal gains from exploiting correlations of >2 variables using BDTs

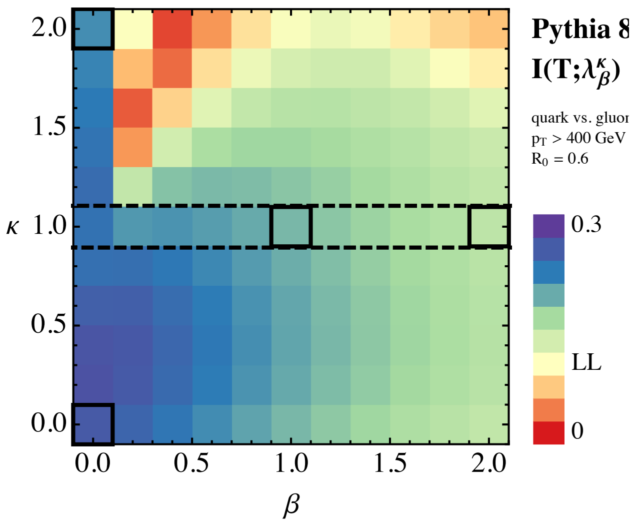
- **Deep learning** approach
  - Use image-recognition technology to avoid thinking
  - Does better than traditional approach!
  - Relies heavily on simulations, *but*
    - Performance independent of Pythia or Herwig training

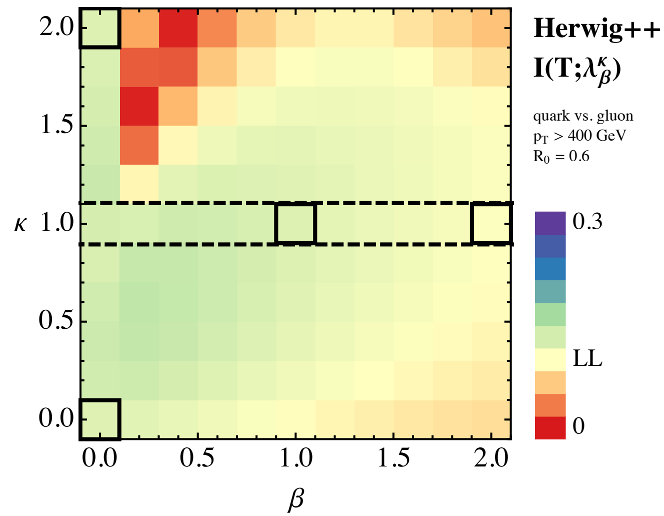
Domo Arigato!

BACKUP

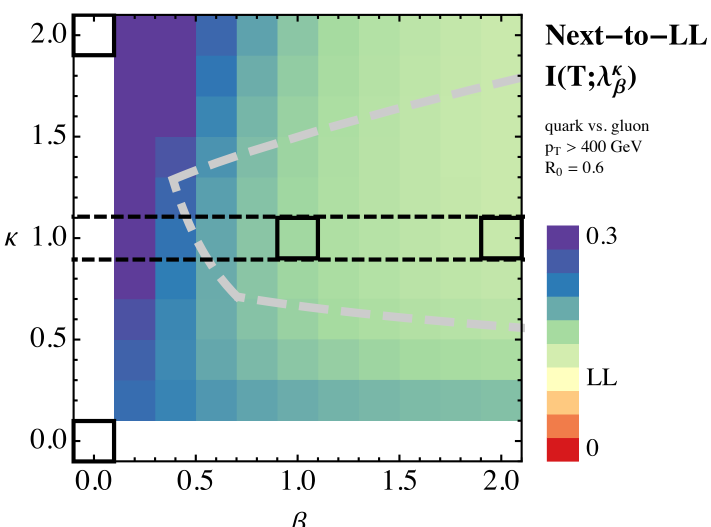
Analytic approach to correlations

Larkoski et al. arXiv:1408.3122





Monte-Carlo simulations

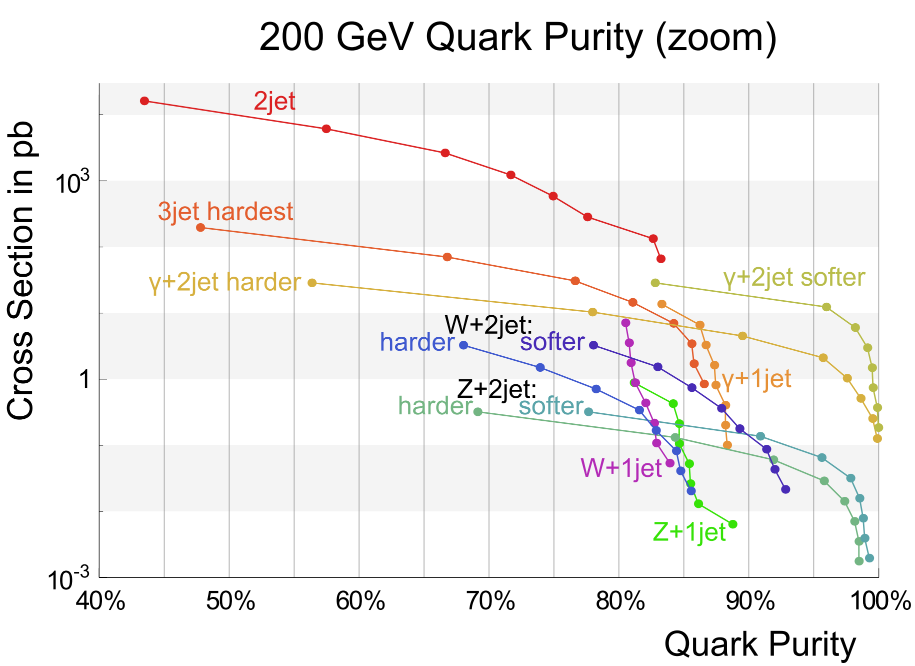


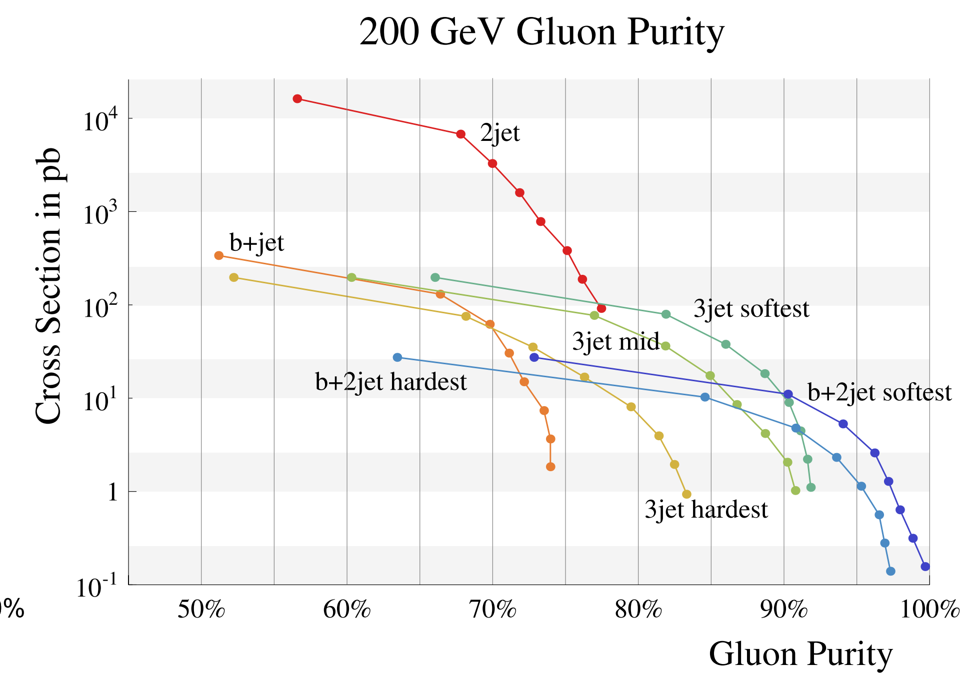
Analytic approach to generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta \, .$$

- Challenging
- Not impossible
- Complementary to MCs

Data: where are the quark and gluon jets?



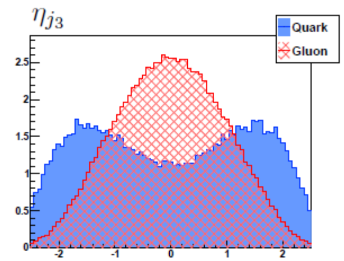


Photon + jet samples

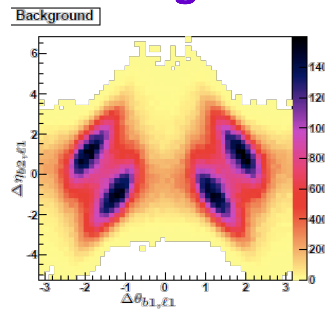
- Jet closer to photon likely quark

- b + 2 jet: high purity low s
  - One jet is b other is gluon
- Dijet: high cross section, low purity

Multivariate approach



- We can think about and visualize **single variables**



- Two variables are harder

- Multidimensional distributions are not well-suited for visualization.

- There are things that **computers are just better** at.

- Multivariate approaches let you figure out how well you could **possibly do**

EFFICIENCY

Save you the trouble or looking for good variables (project killer)

FRAMING

See if simple variables can do as well (establishes the goal)

POWER

Sometimes they are really necessary (e.g. b tagging)

Multivariate methods

Lots of methods

- Boosted Decision Trees
- Artificial Neural Networks
- Fischer Discriminants
- Rectangular cut optimization
- Projective Likelihood Estimator
- H-matrix discriminant
- Predictive learning/Rule ensemble
- Support Vector Machines
- K-nearest neighbor
- …

Useful in many areas of science

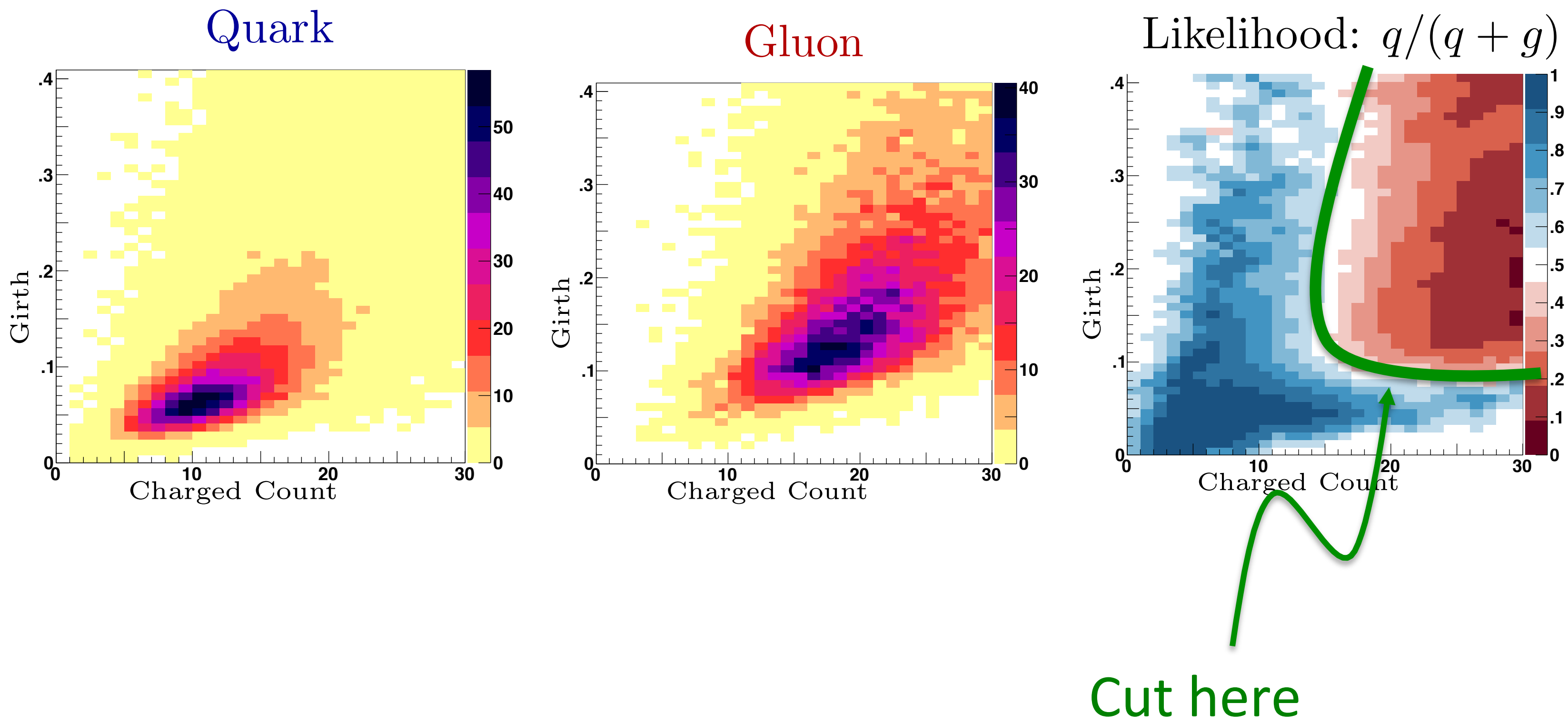
For particle physics, **Boosted Decision Trees** are best suited for combining variables

Easy to understand

Train fast

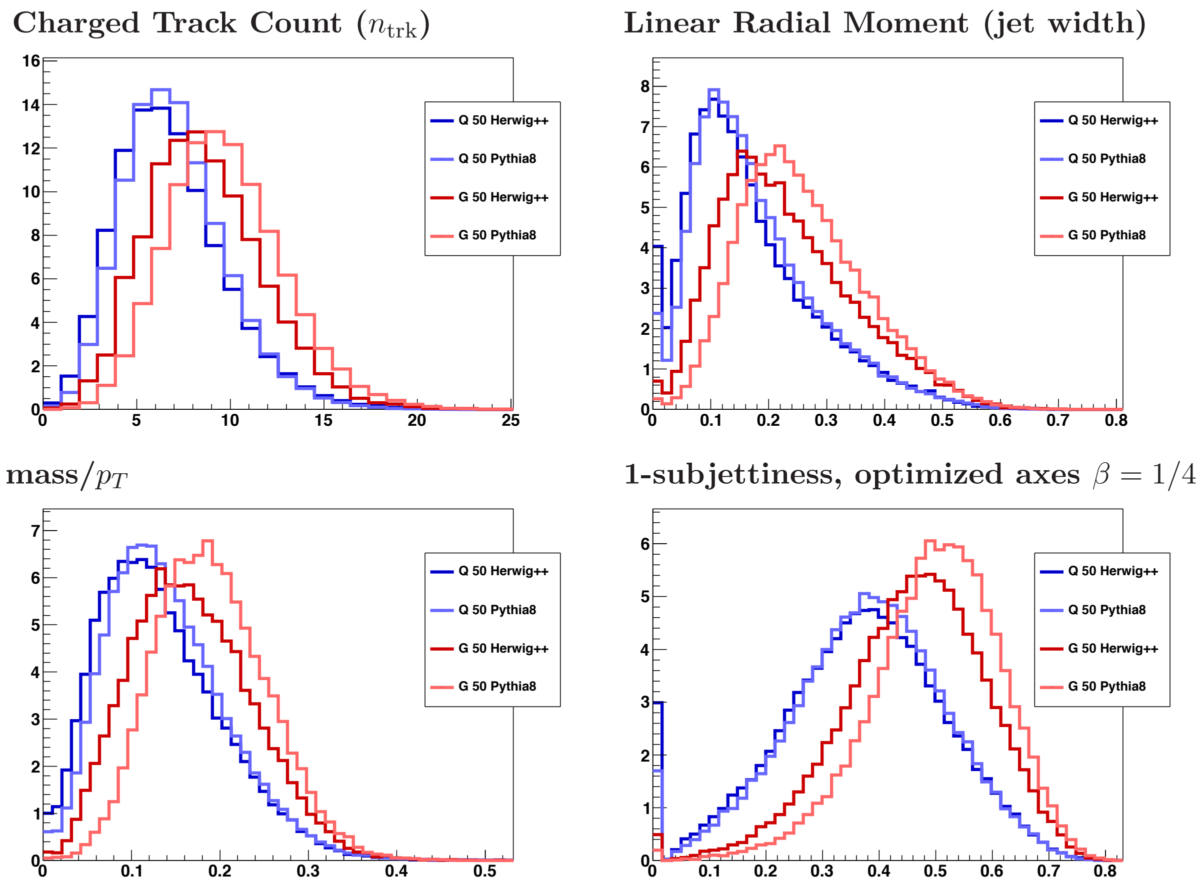
Nearly optimal efficiencies

Are they correlated?



- Not completely
- Can get more discrimination from 2D cuts

Pythia vs Herwig



- Pythia and Herwig qualitatively similar
- Discrimination power with Herwig ++ universally worse

Quark and gluon tagging: results

	Gluon Efficiency % at 50% Quark Acceptance	**50 GeV**				**200 GeV**			
		Particles		Tracks		Particles		Tracks	
		P8	H++	P8	H++	P8	H++	P8	H++
Single variables	2-Point Moment β=1/5	8.7*	17.8*	13.7*	22.8*	8.3	15.9	13.2	19.6
	1-Subjettiness β=1/2	9.3	18.5	14.2	22.9	7.6	16.2	12.3	19.4*
	2-Subjettiness β=1/2	9.2	18.6	13.9	23.6	6.8	15.7*	9.8	18.7
	3-Subjettiness β=1	9.1	19.3	14.6	24.4	5.9*	16.7	8.6*	19.5
	Radial Moment β=1 (Girth)	10.3	20.5	16.1	24.9	11.2	18.9	15.3	21.9
	Angularity $a = +1$	10.3	20.0	15.8	24.5	12.0	19.3	14.0	21.6
	Det of Covariance Matrix	11.2	21.2	18.1	27.0	9.4	20.9	13.5	24.6
	Track Spread: $\sqrt{< p_T^2 >}/p_T^{\rm jet}$	16.5	25.3	16.5	25.3	9.3	20.1	9.3	20.1
	Track Count	17.7	26.4	17.7	26.4	8.9	21.0	8.9	21.0
	Decluster with k_T, ΔR	15.8	24.5	20.1	28.4	13.9	20.1	16.9	23.4
	Jet m/p_T for R=0.3 subjet	13.1	25.9	16.3	27.7	11.9	24.2	14.8	26.2
	Planar Flow	28.7	34.4	28.7	34.4	39.6	42.9	39.6	42.9
	Pull Magnitude	37.0	39.0	32.9	35.6	30.6	30.2	29.6	30.6
Pairs of variables	Track Count & Girth	9.9	20.1	13.4	23.2	7.1	17.3	7.7*	18.7
	R=0.3 m/p_T & R=0.7 2-Point β=1/5	7.9*	17.7	12.2*	22.1	5.7	14.4*	8.5	17.9
	1-Subj β=1/2 & R=0.7 2-Point β=1/5	8.5	17.3*	12.9	22.1	6.0	14.6	8.6	17.7*
	Girth & R=0.7 2-Point β=1/10	12.6	21.9	12.6	21.9*	9.2	18.0	9.2	18.0
	1-Subj β=1/2 & 3-Subj β=1	8.9	18.0	14.0	23.2	5.6*	15.0	8.4	18.4
3,4,5 variables	Best Group of 3	7.5	17.0	11.0	20.9	4.7	14.0	6.9	16.6
	Best Group of 4	7.1	16.7	10.6	20.5	4.5	13.7	6.2	16.3
	Best Group of 5	6.9	16.4	10.4	20.0	4.3	13.3	6.1	15.9