Boosting $H \to b\bar{b}$ measurement with ML

Joshua Lin[1,2], Marat Freytsis[3], Ian Moult[4,5], and Benjamin Nachman[1]

[1]Physics Division, Lawrence Berkeley National Laboratory, Berkeley
[2]Department of Physics, University of California, Berkeley, Berkeley
[3]Institute of Theoretical Science, University of Oregon, Eugene
[4]Berkeley Center for Theoretical Physics, University of California, Berkeley
[5]Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley

BERKELEY LAB

NERSC

Figure 1

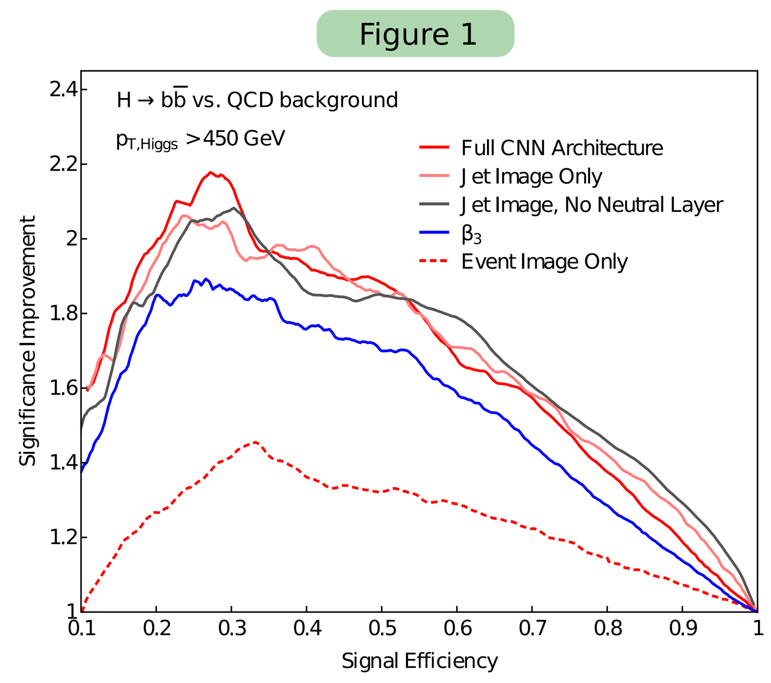


Figure 2

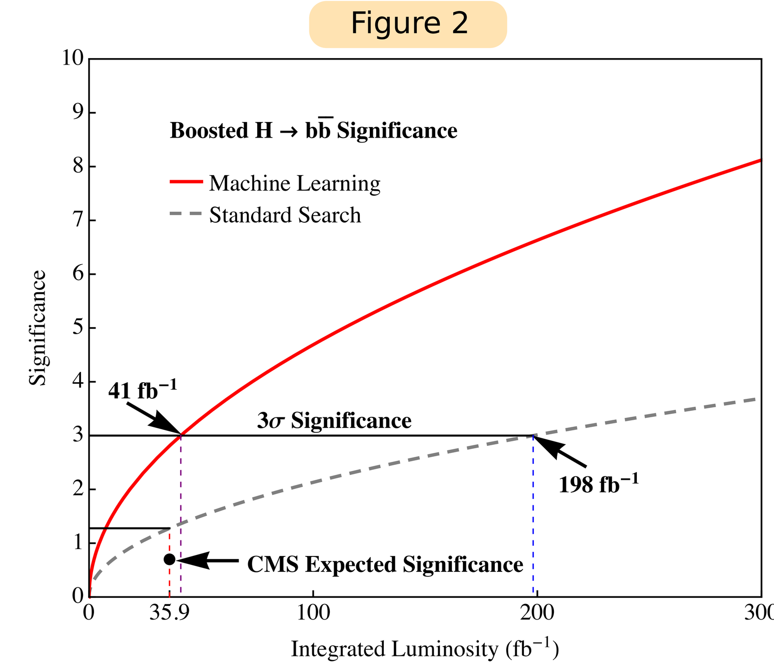


Figure 1 : Signal Improvement Characteristic (SIC)

- With p_T > 450 GeV, calculated with a binned likelihood fit, with bins of 7 GeV in mass spectrum
- Note that CNN's perform better than classical variables such as β_3
- The CNN without neutral layer also performs well, suggesting CNN is resistant to pileup

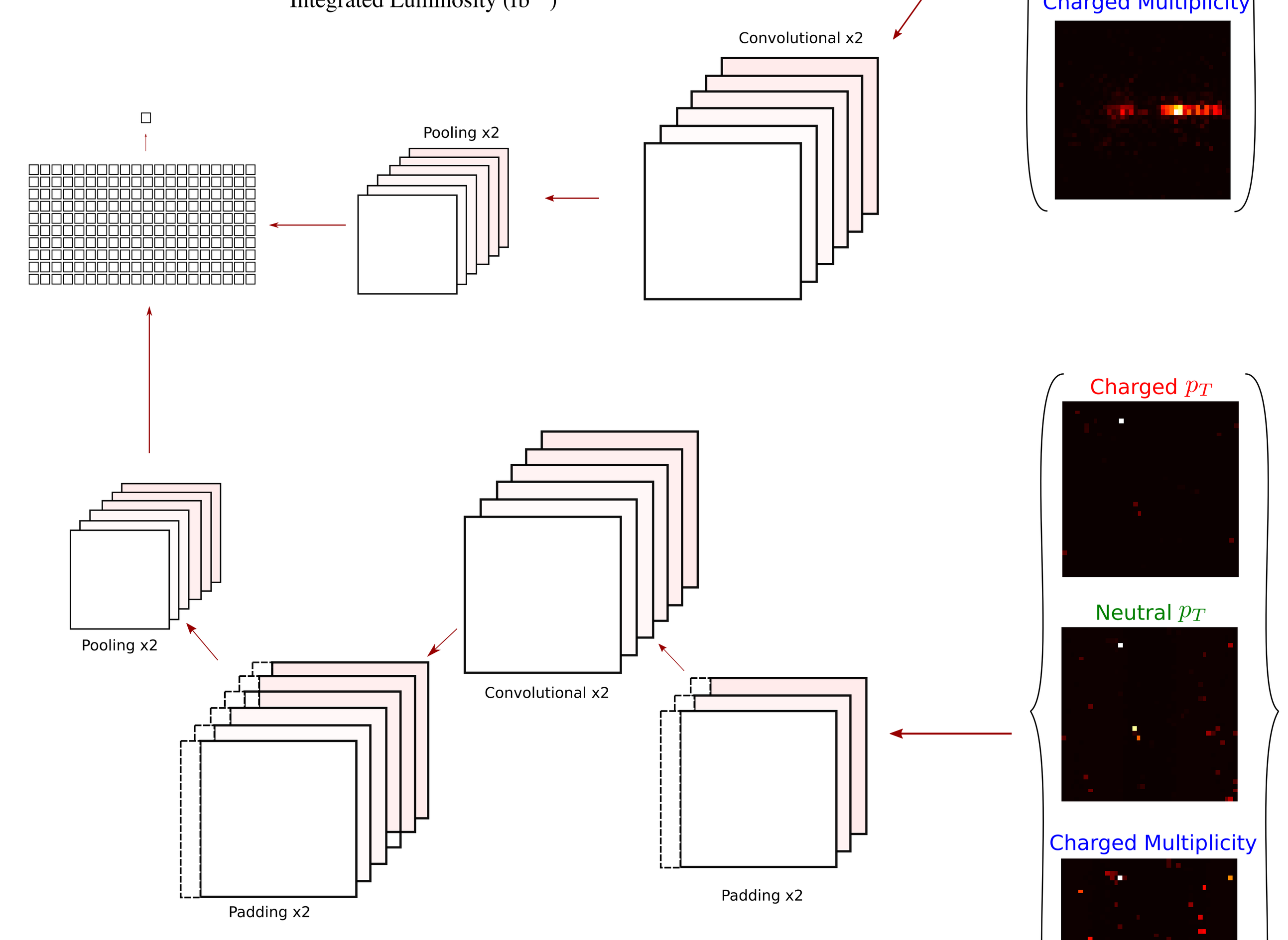
Figure 2 : Significance of H -> bb decay as a function of Integrated Luminosity (amount of data)

- With ML (red curve), we reach a 3 sigma (observation) much faster than without ML. To give scale, Run 2 has given us ~ 100 fb^{-1} so far, and we expect 300 fb^{-1} by the end of run 3

To the Right : Schematic of Neural Network

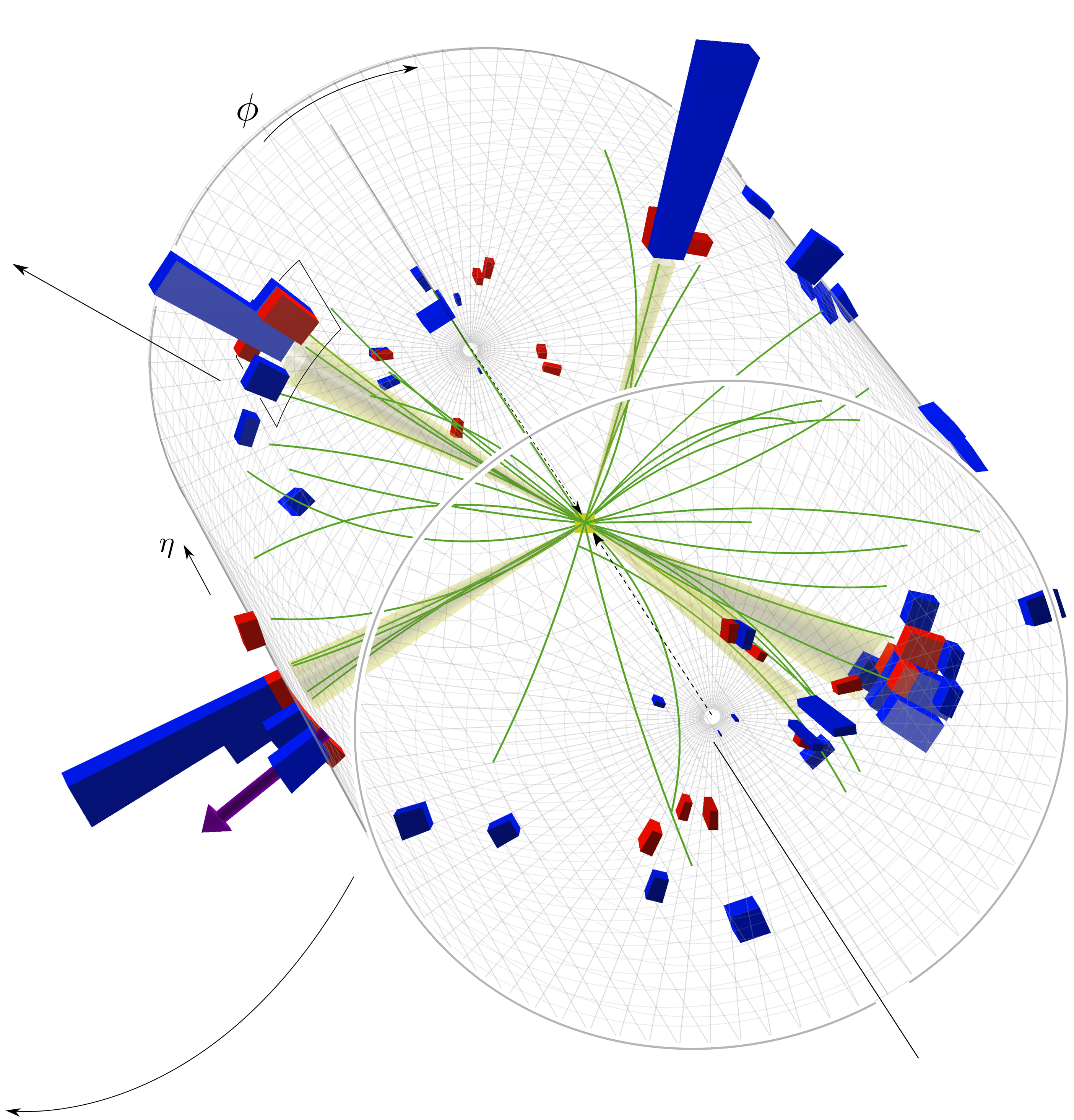
- We use a two-stream neural network to attempt to gain information about both the full event and the jet image.
- We introduce 'padding layers' to account for phi invariance, and we discuss their usefulness below.

Three Color Channel Jet Images
(arxiv 1612.01551)



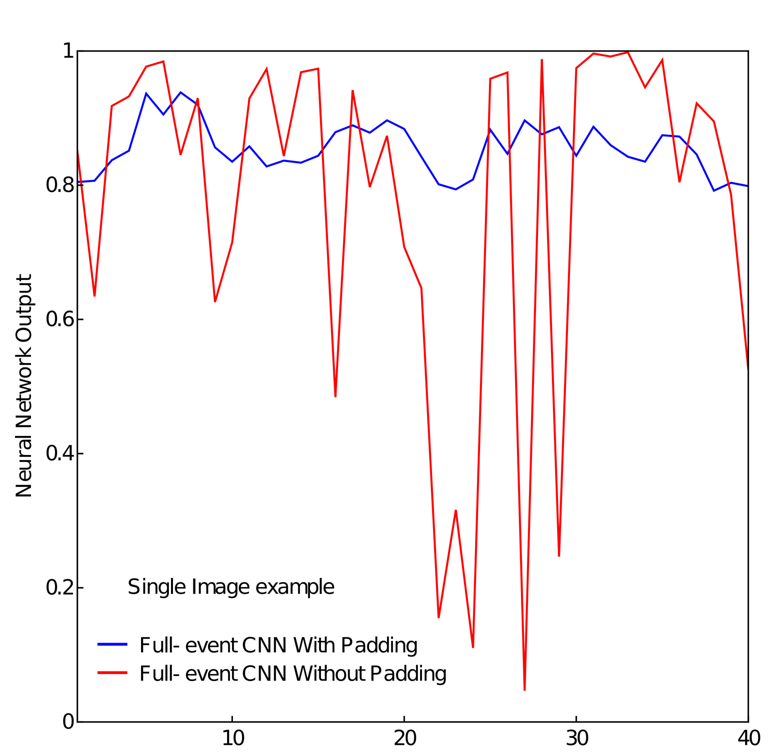
Keras (backend Tensorflow) on Nvidia 1080 Ti GPUs
AdaDelta, Early Stopping regularization, 2.6 million parameters

MADGRAPH5_aMC@NLO 2.6.2 -> PYTHIA 8.226
Signal : pp > Hj [QCD], Hjj[QCD]   Background : pp > jj,jjj,jjjj



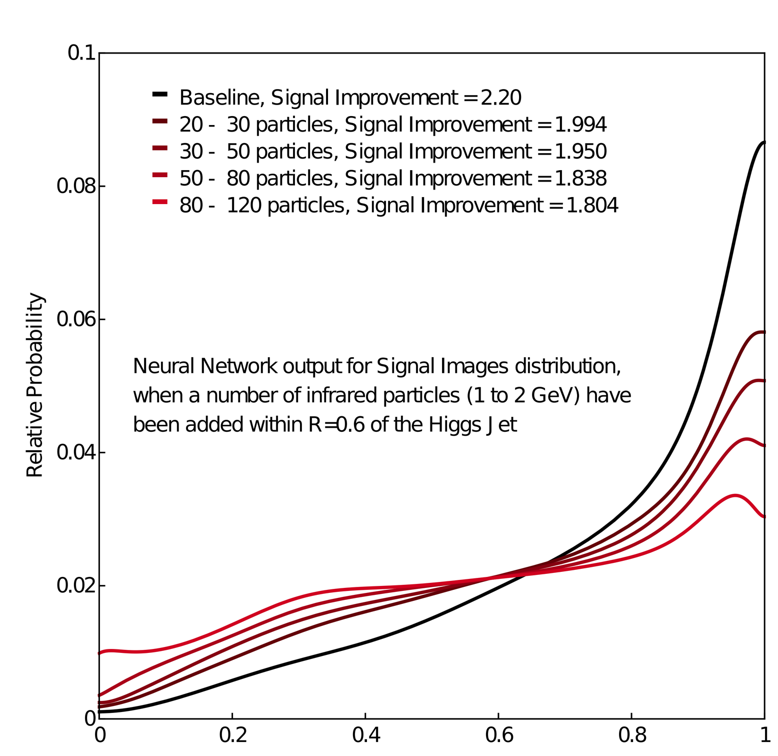
Event Display taken from CMS Fireworks/cmsShow. This particular event shown is for demonstration purposes only.

Investigating Our Neural Network



In this plot, we demonstrate that the padding layers on our Convolutional Neural Network have the desired effect of reducing the variance in the neural network output as the images are rotated in the phi direction.

The reason that the padded CNN is not completely rotationally invariant is that after all the convolutional/pooling layers there is a dense layer that breaks the symmetry.



In this study, we tried adding different amounts of infrared noise to each of the pp collisions, by generating some amount of 1 - 2 GeV pT particles and place them uniformly randomly in an 0.6 x 0.6 square around the Higgs Jet. We then feed these noisy images into the Neural Network that we had previously trained, and see how it affects the neural network performance. This plot shows that even at relatively high amounts of noise, we retain good signal improvement (> 1.8 factor)

Classical Jet Substructure Variables

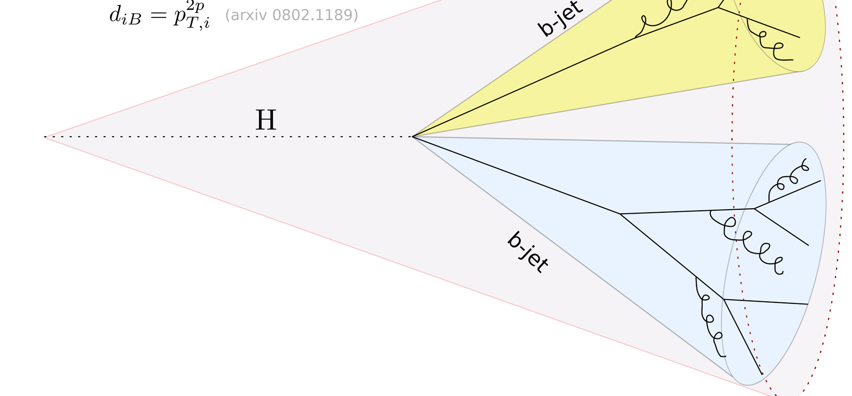
First, particles are usually clustered into jets, based on a specific choice of metric between particles. The algorithm proceeds by clustering the pair of particles closest to each other, until you remain with your jets. The most popular metric is shown right, with p = -1 (anti-kT).

After clustering into jets, various physically motivated variables can be defined with input as a jet's constituent particles 4-vectors. These are mainly built with the intention of separating 'interesting jets' (Usually Higgs, vector bosons,..) from 'uninteresting jets' (QCD background)

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{R_{ij}^2}{R_0^2}$$

$$d_{iB} = p_{T,i}^{2p}$$

(arxiv 0802.1189)



$$\vec{r}_i = (\Delta y_i, \Delta\phi_i) \qquad \vec{t} = \sum_{i \in \text{ Jet}} \frac{p_{T,i}|r_i|}{p_{T,\text{Jet}}} \vec{r}_i$$

(arxiv 1001.5027)

The Jet Pull vector is a variable that attempts to capture the color flow information available in a jet. Specifically, it shows how much a jet (or subjet) is 'pulled off-center' by radiation; which is different for Higgs compared to QCD.

$$\tau_N^{(\beta)} = \frac{1}{p_{T,\text{Jet}}} \sum_{i \in \text{Jet}} p_{T,i} \min\left\{ R_{1i}^\beta, R_{2i}^\beta, \dots, R_{Ni}^\beta \right\}$$

(arxiv 1011.2268)

The N-subjettiness variables are designed to measure how much the jet looks like it is made out of N distinct components. They are useful in a wide variety of applications, specifically in this study as our signal involves a splitting of H to two b-quarks; we expect τ_2 to be high.

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

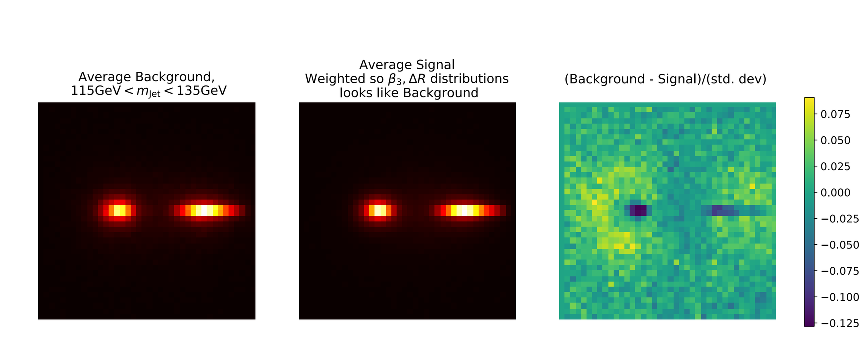
(arxiv 1710.01305)

Built off the N-Subjettiness variables, the β_3 (modern) variable was built specifically to classify Higgs to bb jets apart from gluon to bb jets (which are the main source of QCD background in our study). The parameters are all optimized by means of a neural network to give optimal improvement in significance.

$$N_2^{(\beta)} = \frac{\sum_{1 \le a_1 < a_2 < a_3 \le n_J} \frac{p_{T,a_1}}{p_{T,\text{Jet}}} \frac{p_{T,a_2}}{p_{T,\text{Jet}}} \frac{p_{T,a_3}}{p_{T,\text{Jet}}} \min\left\{ R_{a_1 a_2}, R_{a_1 a_3}, R_{a_2, a_3} \right\}}{\sum_{1 \le a_1 < a_2 \le n_J} \frac{p_{T,a_1}}{p_{T,\text{Jet}}} \frac{p_{T,a_2}}{p_{T,\text{Jet}}} R_{a_1, a_2}}$$

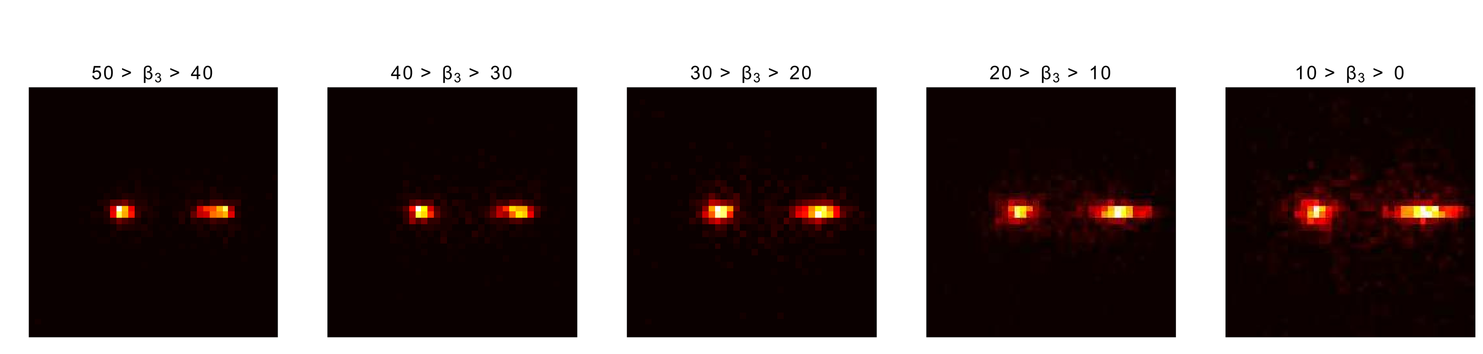
(arxiv 1609.07483)

The N_2 variable is a (modern) discriminant built off of the energy correlation functions, which are designed to behave similarly to the N-subjettiness variables but with the benefit that they are defined without respect to a choice of subjet axes.



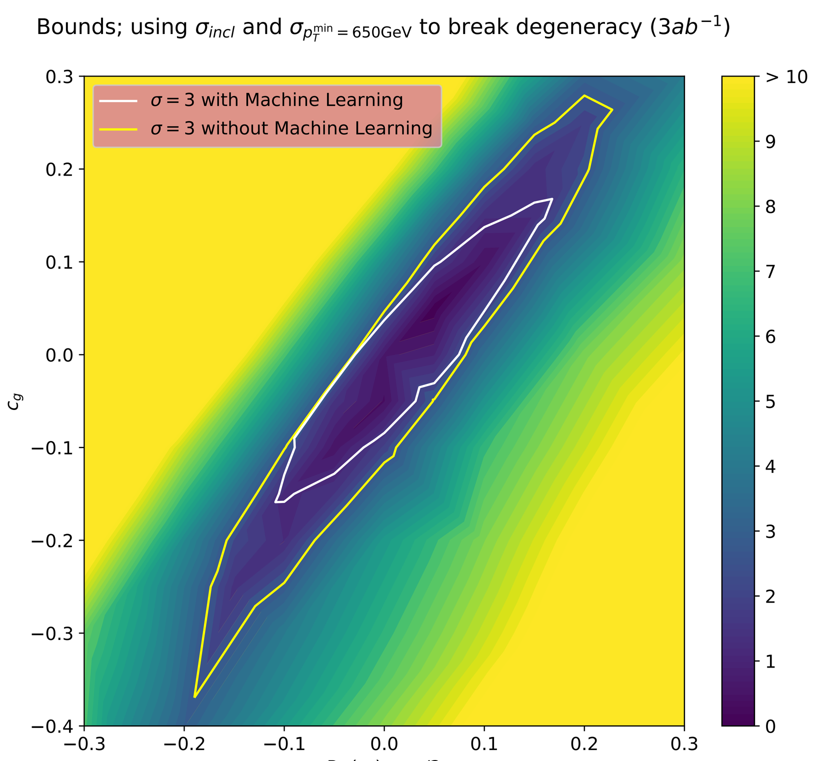
Left : To investigate what other information is available apart from β_3, we can weight the signal images to have the same β_3 distribution as the background. After taking the difference, we find that there is still colour pull information to be learnt, shown by the radiation patterns remaining.

Bottom: Images to show what different values of β_3 correspond to.



Probing physics Beyond the Standard Model

Bounds; using σ_{incl} and $\sigma_{p_T^{cut}=650\text{GeV}}$ to break degeneracy ($3ab^{-1}$)



$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} - \kappa_t \frac{m_t}{v} t\bar{t}h + i\tilde{\kappa}_t \frac{m_t}{v} \bar{t}\gamma_5 th + \kappa_g \frac{\alpha_s}{12\pi} \frac{h}{v} G_{\mu\nu}^a G^{\mu\nu a} + \tilde{\kappa}_g \frac{\alpha_s}{8\pi} \frac{h}{v} G_{\mu\nu}^a \tilde{G}^{a\mu\nu}$$

$$\kappa_t = 1 - \text{Re}(c_y) - \frac{c_H}{2}, \quad \kappa_g = c_g, \quad \tilde{\kappa}_t = \text{Im}(c_y), \quad \tilde{\kappa}_g = \tilde{c}_g$$

Shown left are our bounds on the Wilson Coefficients that control the coupling of the Higgs Boson to tops and gluons. Changing these coefficients introduces two effects, one of which is an overall difference in the inclusive cross-section (which gives us the degeneracy band going from bottom-left to top right), and the other is an overall hardening of the p_T spectrum at high p_T, which allows us to 'break the degeneracy' and constrain the coefficients to a small region. We see that with ML, we achieve a much better constraint than without ML.