Computer-Vision Jet Identification

Josh Cogan, Michael Kagan, **Emanuel Strauss**, Ariel Schwartzman

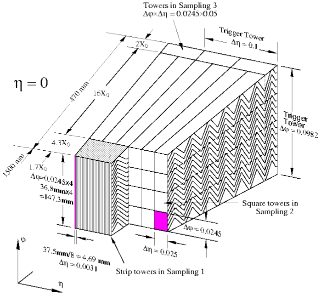
SLAC NATIONAL ACCELERATOR LABORATORY

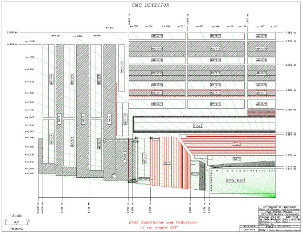
September 17, 2013

Overview

Motivation of Boosted Searches

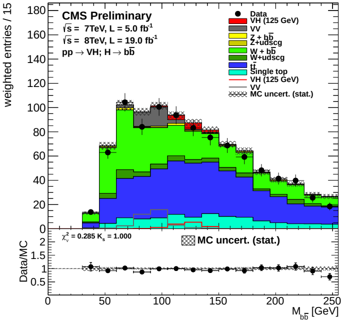
- Idea gained traction in 2008 by Butterworth, Davison, Rubin, and Salam as a way to recover $V(H \to b\bar{b})$ channels.
    - Causes system to be more central (better tagging performance)
    - Additional background handles (additional jet activity)
- By consequence, overlapping signal jets: $\Delta R_{b\bar{b}} \simeq \frac{1}{\sqrt{z(1-z)}} \frac{m_H}{p_T}$





- Fine segmentation of ATLAS and CMS calorimeters allows exploration of jet substructure.

Boosted Objects at the LHC



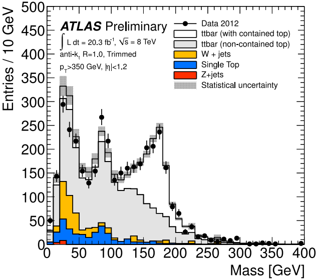
The future played out differently, but boost is a real part of LHC analyses:

- CMS (ATLAS) $V(H \to b\bar{b})$, three (five) Vp_T bins

http://arXiv.org/abs/0802.2470v2

Boosted Objects at the LHC

- ATLAS triggers on "Fat" jets with $\Delta R = 0.1$
- Boosted top analyses have their best mass resolution in the high p_T bin.
  - All three decay products fully contained in a single jet
  - QCD p_T spectrum falls off rapidly



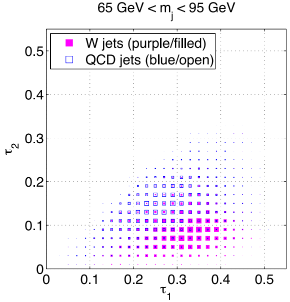
(a) $p_T > 350$ GeV (b) $p_T > 500$ GeV

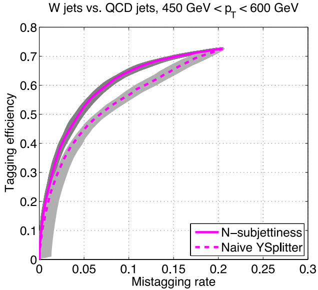
- MC Simulation does an admirable job of describing the data (due to immense amount of validation and performance work).

http://cds.cern.ch/record/1571040

Approaches to Substructure: N-Subjetiness

- Designed to identify boosted hadronically-decaying objects like EW bosons and top quarks.
  - Boosted bosons: 2 energetic, narrow, sub-jets.
  - QCD jets: several broader sub-jets.
- $\tau_N = \frac{1}{d_0} \sum_{k} p_{T,k} \min_{A} \Delta R_{A,k}$
  - k – sum over jet constituents
  - A – Minimize distance to candidate subjet axes
- W/Z/H tagging with τ_2/τ_1, top tagging with τ_3/τ_2
- "Intuitively" measures likelihood system is compatible with n-prong structure.



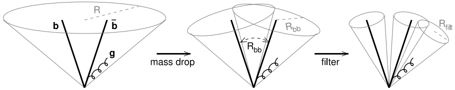


http://arxiv.org/abs/1011.2268, http://arxiv.org/abs/0806.0023v2

Other Approaches

Mass Drop

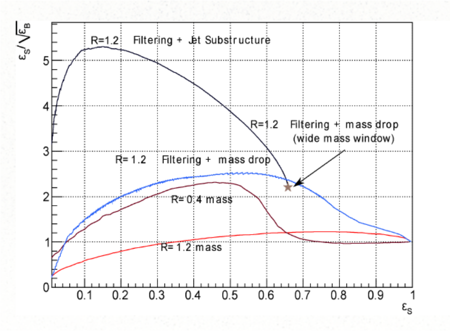
- Decompose the jet evolution.
- Given a hard jet, perform iterative decomposition.
  - Using Cambridge-Achen jet clusering
- Find a subjets with $m_{j1} < \mu m_{j2}$
- And $y = \frac{\min(p_{T1}, p_{T2})(\Delta R_{12})^2}{m_{\mathrm{jet}}^2}$



- Originally proposed to improve $H \to b\bar{b}$ selection.

MVA Tagger

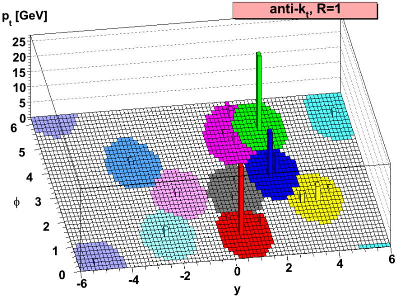
- Attack the dimensionality problem.
- Many (well motivated) features describe a reconstructed jet.
- Use planar flow, jet shapes, etc. . .



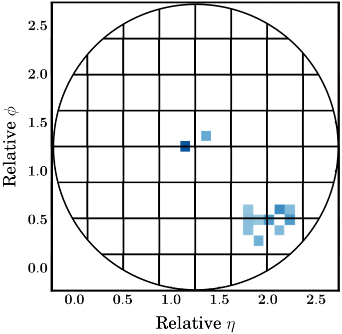
http://arxiv.org/abs/0802.2470, http://arxiv.org/abs/1012.2077

Jets as Images

Most focus on a top-down approach (variables motivated by an analytical understanding of the problem), we tackle the problem from the bottom up.



Jet identification and sub-structure is generally treated as a form of cluster analysis.



Towers can just as easily be treated as the pixels of an image.

Uniform basis, with much lower dimensionality than machine vision (625 cells for $\Delta\eta \times \phi = 0.1 \times 0.1$, instead of $\sim$1M for a picture)

From Computer Vision To Jet Classification

Techniques for producing numerical values or decisions from high-dimensional data.

Mapping gender recognition onto jet identification:

- Face detection
  - Face region of interest finding
  - Eye detection
- Face preprocessing
  - Geometric transformation
  - Equalization
  - Noise reduction / smoothing
  - Masking
- Facial recognition / discrimination
  - Transformations
  - Training and using a discriminator

- Jet finding
  - Jet clustering
  - Grooming
  - Principal axis finding
- Jet preprocessing
  - Geometric transformation
  - Normalization
  - Masking
- Jet tagging / discrimination
  - Transformations
  - Training and using a discriminator

Demonstrate proof of principle using Monte Carlo with an idealized detector

Practically Speaking

- Picking a representation
  - Considering only calorimeter towers (no track information)
  - Enforce a uniform representation by projecting the calorimeter energies onto a $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ grid
  - Perform clustering (Anti-k_T with $\Delta R = 1.2$) to find jets, apply jet trimming with radius 0.3 and $f = 0.05$.
  - Keep 25×25 cells, centered around the jet.
  - Each jet is described by an uncurled row vector of cell energies (length = 625).

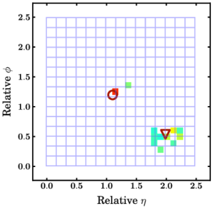
- Each jet is now uniformly defined by a single feature vector.
- Same number of variables per jet.
- Complete representation of the physics (within tower granularity).
- Set of feature vectors defines a feature space.
- Easily lends itself to feature extraction and dimensionality reduction.

Case Study: W vs Light Jets

- Great system for algorithm development and validation
- Well known system with several attractive features:
  - W decays to two quarks should be composed of two distinct hard sub-jets.
  - Decays have a fundamentally different energy pattern than QCD jets (for comparable jet masses).
  - Thoroughly explored in the litterature (N-Subjetiness, MVA Tagger, etc...).

If the method works here, it gives us good reason to try other systems as well.

Image Processing



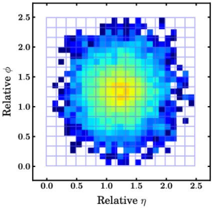
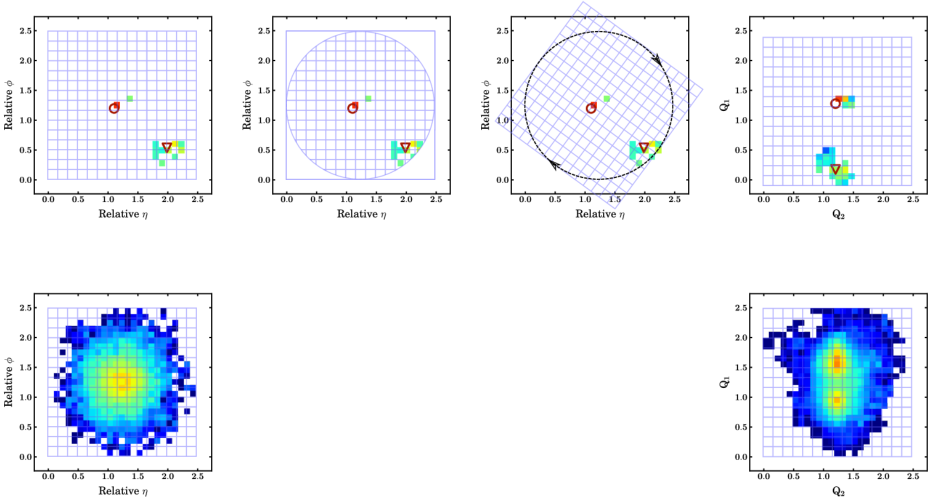
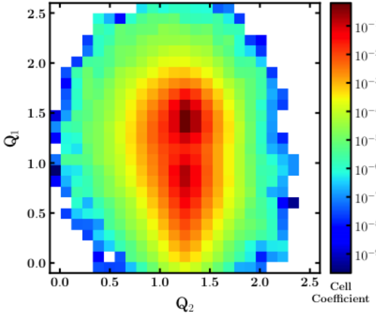


Image Processing

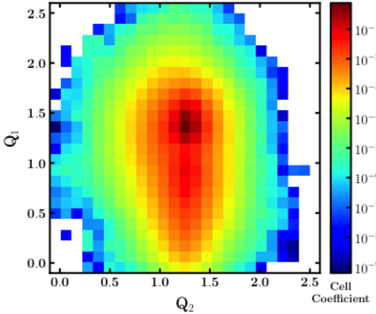


Building the Discriminant

Consistent “image” representation for calorimeter objects.



Average W Jet



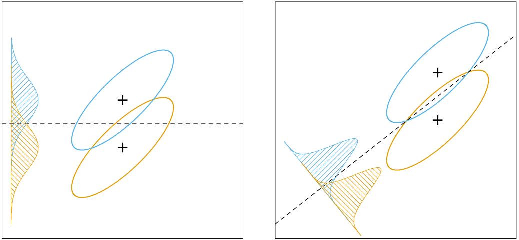
Average Light Jet

Now, turn this information into a number we can cut on.

Fisher's Linear Discriminant

Finding a subspace which represents most of the data variance:

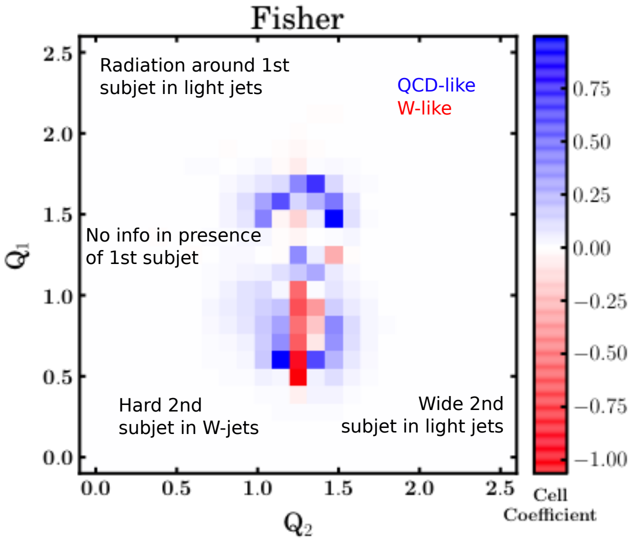
- Map same class sample vectors in a single spot of the feature representation.
  - Minimize within class differences: $S_w = \sum_{i=1}^{C} \sum_{j=1}^{N} (x_j - \mu_i)(x_j - \mu_j)^T$
- Map those of different classes as far apart from each other as possible.
  - Maximize between class differences: $S_b = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T$



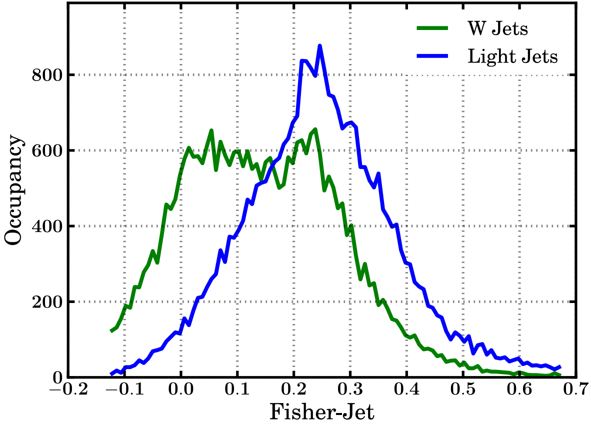
Find the basis vector V for: $S_b V = (S_w + \sigma^2 I) V \lambda$

(Actually, even easier in the two class case.)

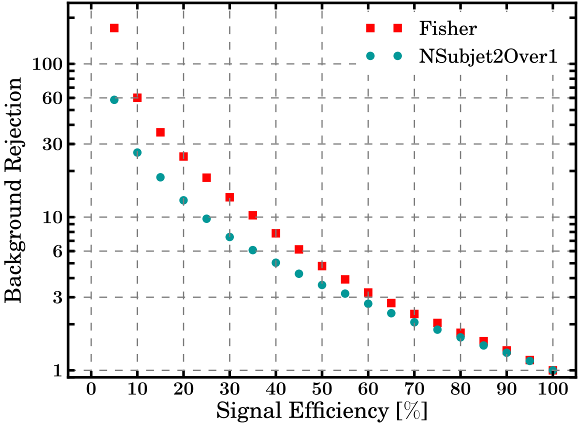
Visualizing the Classifier



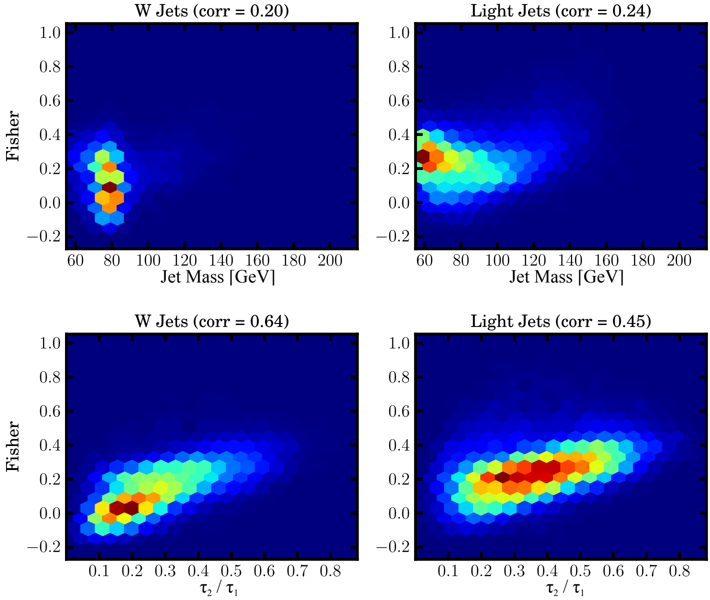
Classifier Performance



Classifier Performance



Unique Information

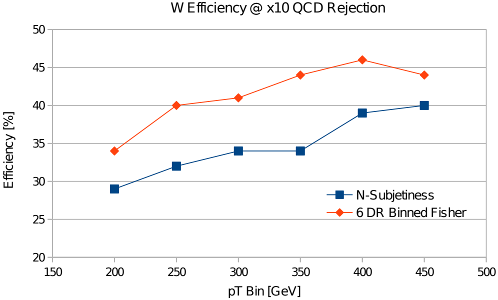


Robustness

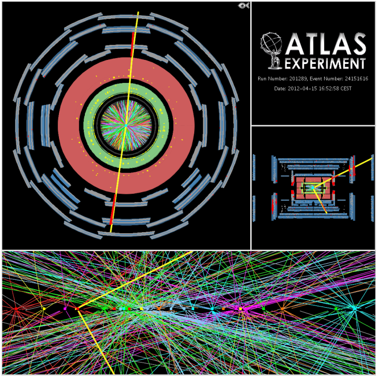
- Facial identification techniques are designed to perform well in the presence of deteriorated face images.
  - Gaussian noise
  - Bluring
  - Salt and Pepper
  - Partial Input
- A priori expect the method to work well based on the success in facial identification litterature.

p_T Dependence

- Fisher discriminant is trained in 50 GeV bins
  - Improves the image "resolution"
  - Easier to generate MC to probe the full p_T spectrum
- Binning could be optimized, or extended into other variables.
- Performance is robust against jet p_T



Effects of Pileup

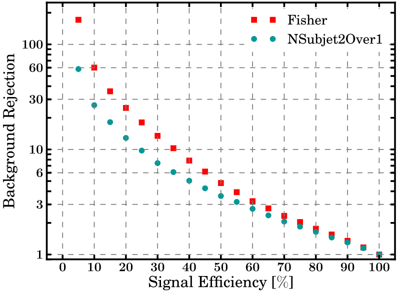


- One of the main challenges at the LHC
  - Additional interactions per bunch crossing (in-time)
  - Calorimeter integration time over multiple crossings (out-of-time)
- Additional event energy
- Increased fluctuations in the jet resolution noise term.
- Increased number of "fake" jets

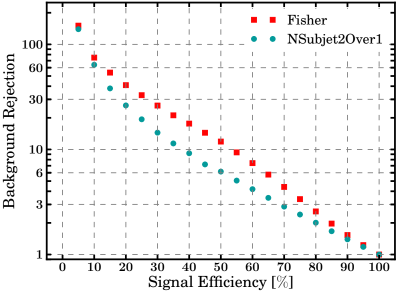
Pileup

Investigate effect of pileup by overlaying additional Pythia Minbias events.

- Train on samples without pileup
- Test on W and L samples with $<\mu>= 0$ and 30.



$<\mu>= 0$

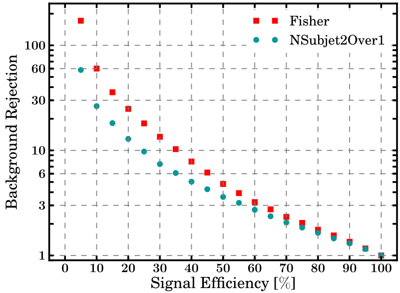


$<\mu>= 30$

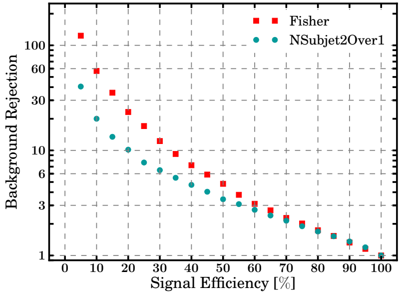
- Note: statistically independent and differently distributed samples
  - Pre-selection for sig (bkg) increases by a factor of 1.5 (7).

Radial Energy Smearing

- Simulate the energy spread introduced by electromagnetic and hadronic showers in calorimeters.
- Use an ATLAS-like setup with EM and HAD showers.



No Smearing



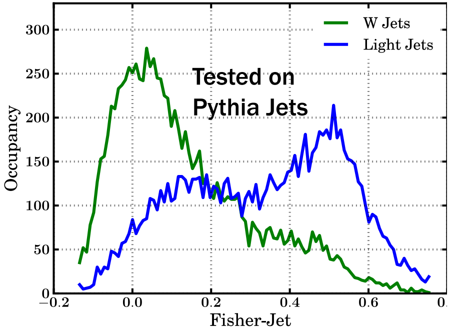
Radial Energy Smearing

http://atlas.physics.arizona.edu/~loch/

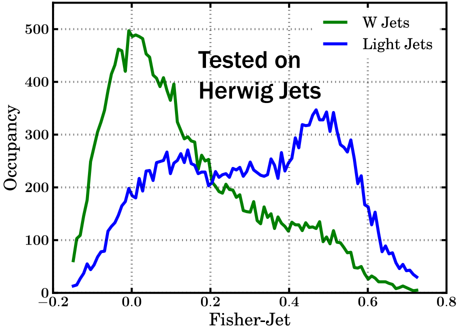
Generator Model

- Check Pythia vs Herwig: want to see no sensitivity to differences
- Use the same discriminant (trained on Pythia jets)
- Similar performance when tested on Pythia or Herwig jets

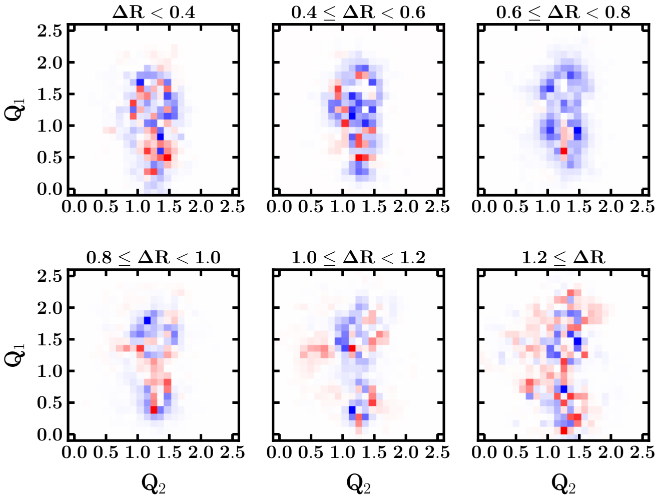
W eff @ QCD Rejection
71% @ x2
18% @ x10
10% @ x20



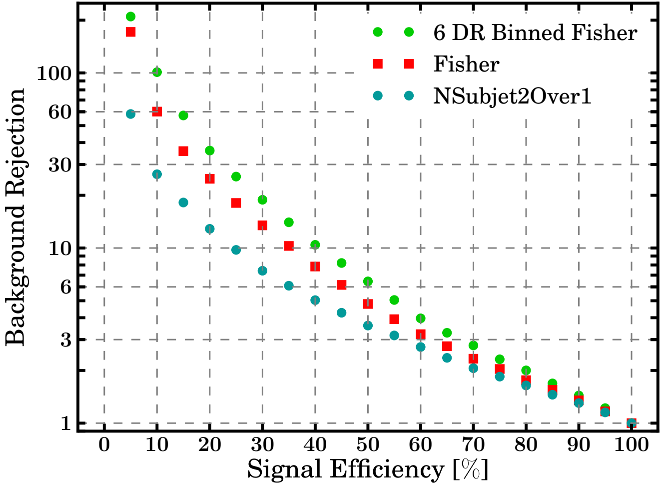
W eff @ QCD Rejection
70% @ x2
20% @ x10
11% @ x20



Binning The Classifier

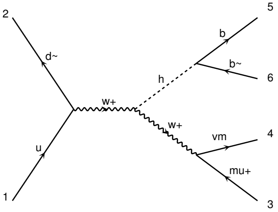


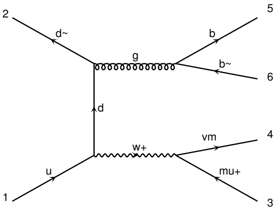
Binning The Classifier



Case Study: $H \to b\bar{b}$ vs $g \to b\bar{b}$

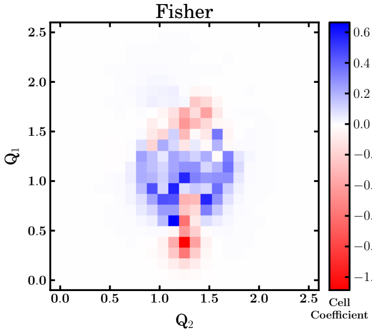
- The canonical example par excellence
  - Jet substructure looks very different, even if mass, p_T, etc. . . are the same.
- Potential for large boost in 14 TeV dataset.
  - Natural extension for moderately boosted analysis at 8 TeV



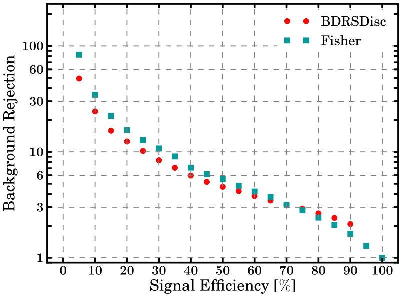


- Most powerful handle to date: *BDRS* corrected invariant jet mass
  - Potential to exploit additional b-tagging.

Discriminant Performance



$300 < H_{p_T} < 350$ GeV, $110 < m_{jj} < 125$ GeV,
inclusive ΔR classifier



- Performance is about the same before and after cuts on BDRS
- Can get x10 rejection for ∼70% acceptance using both.

Case Study: q vs g

Most new physics is expected to produce quark jets, rather than gluon-initiated jets.

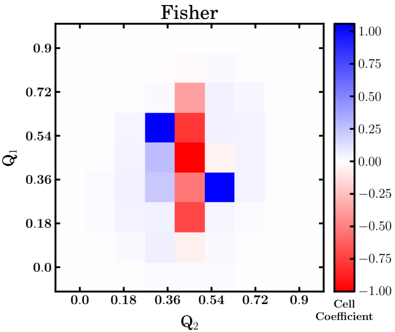
- SUSY cascade decays: $\tilde{g} \rightarrow q\bar{q}\chi_1^0$
- Heavy $Z' \rightarrow q\bar{q}$ or $W' \rightarrow qq'$

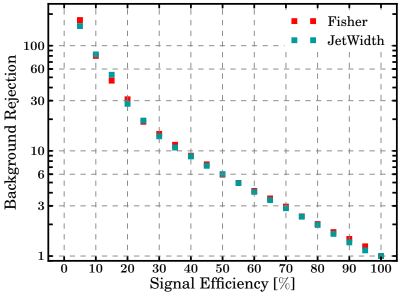
And interesting SM physics tends also to produce quark-jets.

- $t\bar{t}$ to 4 or 6 quarks
- Vector Boson Fusion
- Boosted top, W, and Higgs!

- Switch from “fat” jets $\Delta R = 1.2 \rightarrow 0.6$

q vs g Performance





- Bottom-up approach performs identically to jet width.

Wrap Up

- Generic method for calorimeter tagging using computer vision techniques
  - Many other systems to study: top vs light, particle ID (e.g. photon vs ϕ^0), b-tagging
  - Pileup detection and suppresion
  - Most important conceptual step is the image paradigm, the rest is "simple"
- Provides a powerful tool for gaining physics insight
- In cases studied, performance is as good as, or better, than analytically motivated variables.
- Many parameters to varry in pre-processing
  - Normalization, center fixing, scaling, boosting, whitening
  - Alternative basis (e.g. cell p_T pairs, instead of cell p_T)
- Lots of room for extensions to the method
  - The contents of this talk covers up to the year ∼2000 in the litterature
    - Not limited to using linear classifiers in a two-class problem
  - Baseline method mature enough for application to LHC analyses
  - Documentation is being written up (expect it on the arxiv soon)

Backup Slides

Processing Time

- Method is extremely fast – $\mathcal{O}(1ms)$ per jet