APPLYING COMPUTER VISION TO JET FLAVOR IDENTIFICATION

JOSH COGAN
MICHAEL KAGAN
EMANUEL STRAUSS
ARIEL SCHWARTZMAN

BOOST 2013
AUGUST 15, 2013

SLAC NATIONAL ACCELERATOR LABORATORY

MOTIVATION

- Computer vision: high-dimensional data → usable information
  - Gender recognition, optical character recognition
  - Similar to a myriad of HEP problems—object identification

- Here, focus on calorimeter jet flavor classification

- Use linear methods on *all available* information to
  - Perform feature extraction/dimensional reduction
  - Use transparency of these methods to inspect them physically

SAMPLES STUDIED

- Event Generation: pp @ $\sqrt{s}$ = 8 TeV
  - W vs Light: Pythia W(μν)W(qq) vs W(μν)j
  - Cross Checks Performed:
    - Pile up
    - Fast detector simulation (Loch)
    - Compare generators
- Calorimeter: 0.1x0.1 ΔηxΔϕ, |η| < 2.5
- C/A R=1.2 Jet Finding:
  - Save highest pT jet / event, $|\eta_j| < 2.5$
  - Trimmed (5% kt0.3)
  - Only keep towers that survived trimming
  - Study jets in 6 bins of subjet-ΔR
  - Jet Mass > 55 GeV

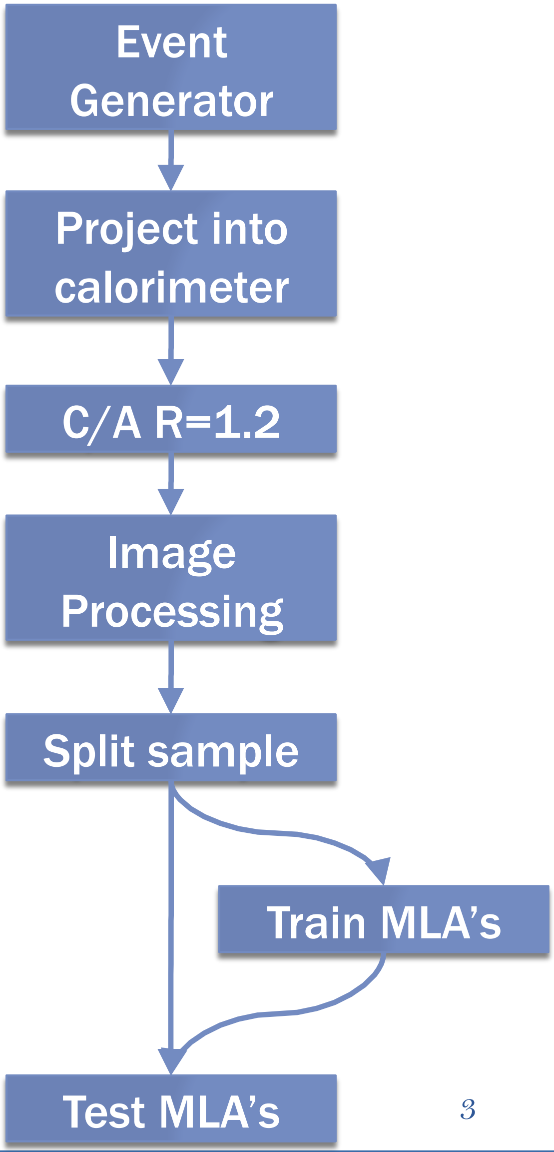
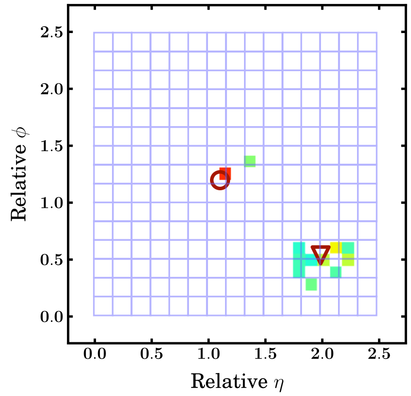
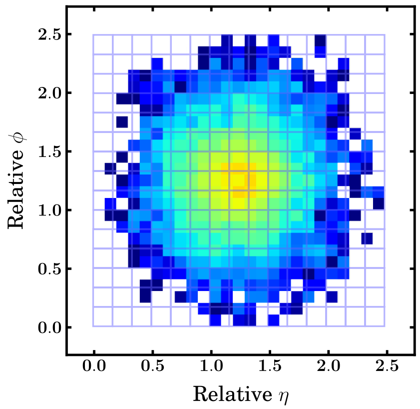


IMAGE PROCESSING



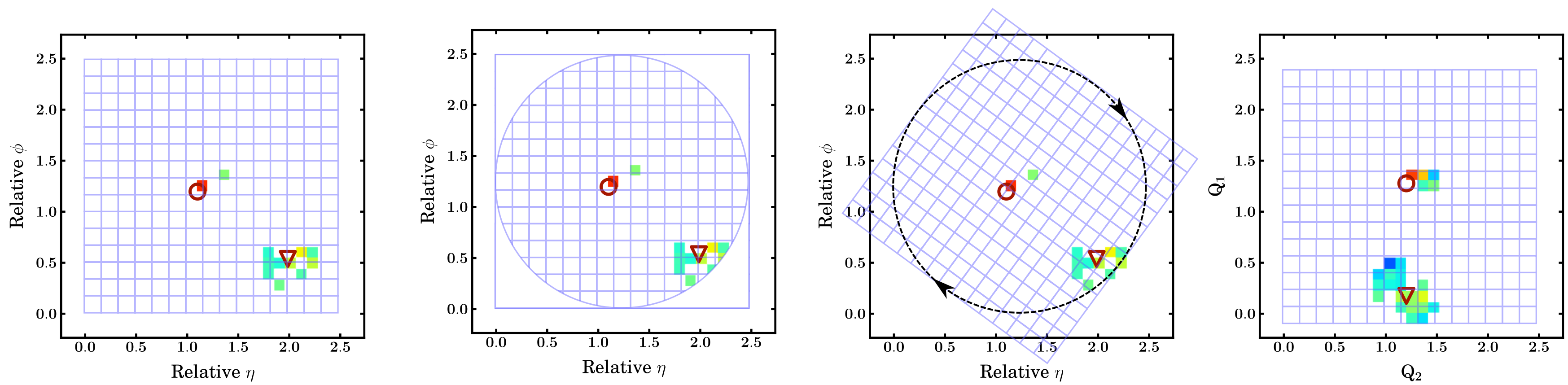
Example W jet

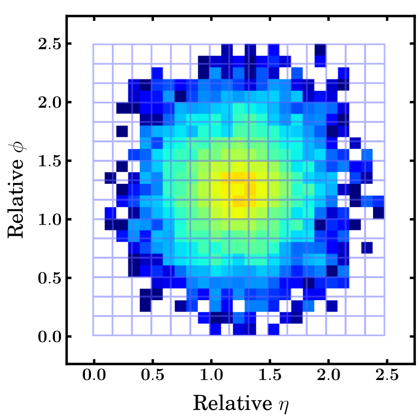


Average of W jets
Not much info!

IMAGE PROCESSING

- Center and rotate jet-images before training MVA
- Introduces small smearing, but huge gain in discrimination!



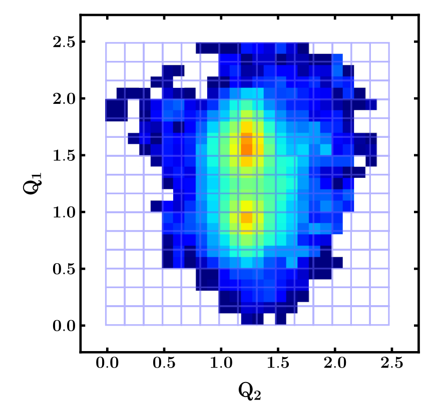


Average of *unrotated* W jet

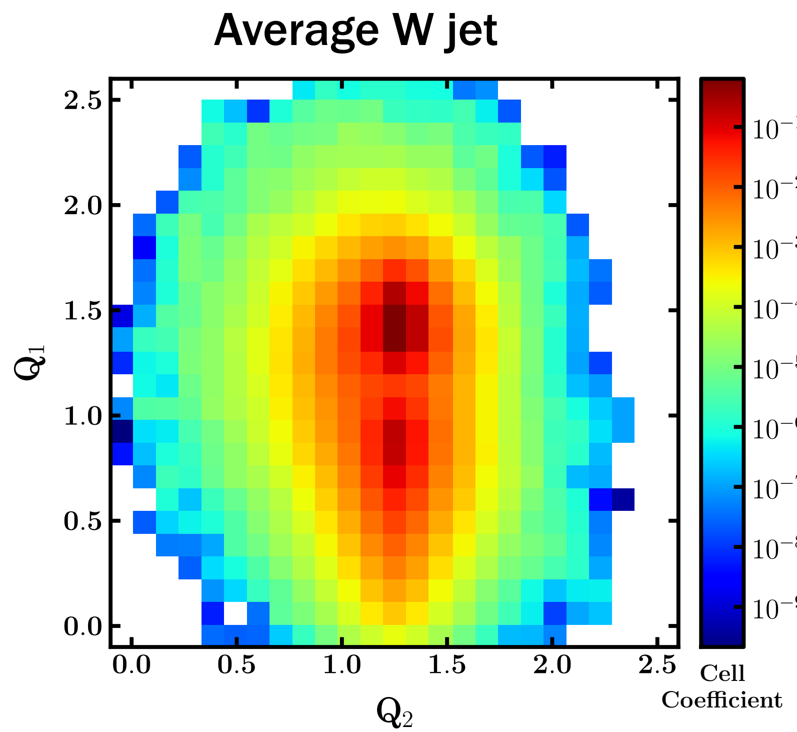
Not much info!

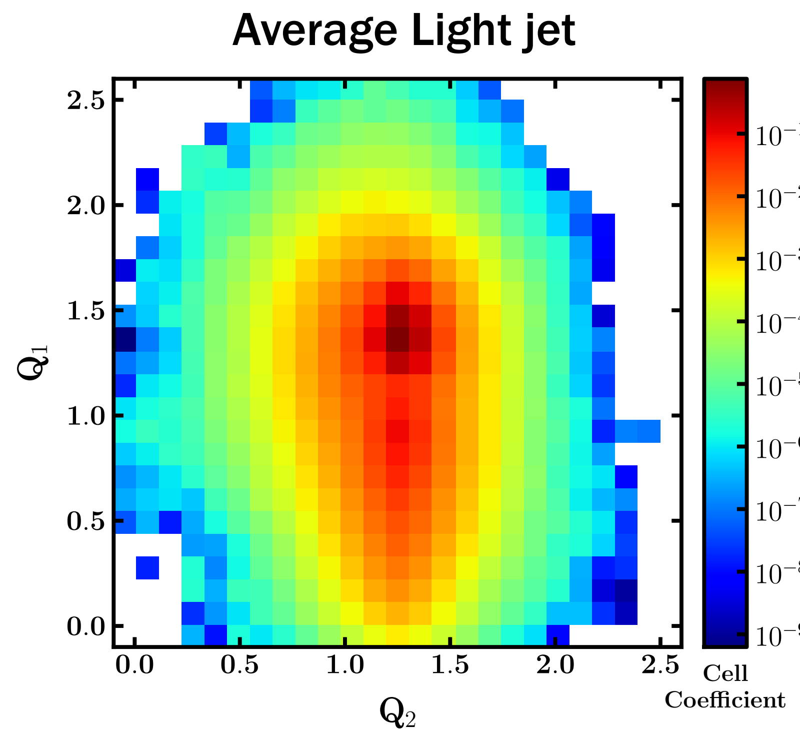
Average of *rotated* W jet

Much better!



CLASS AVERAGES





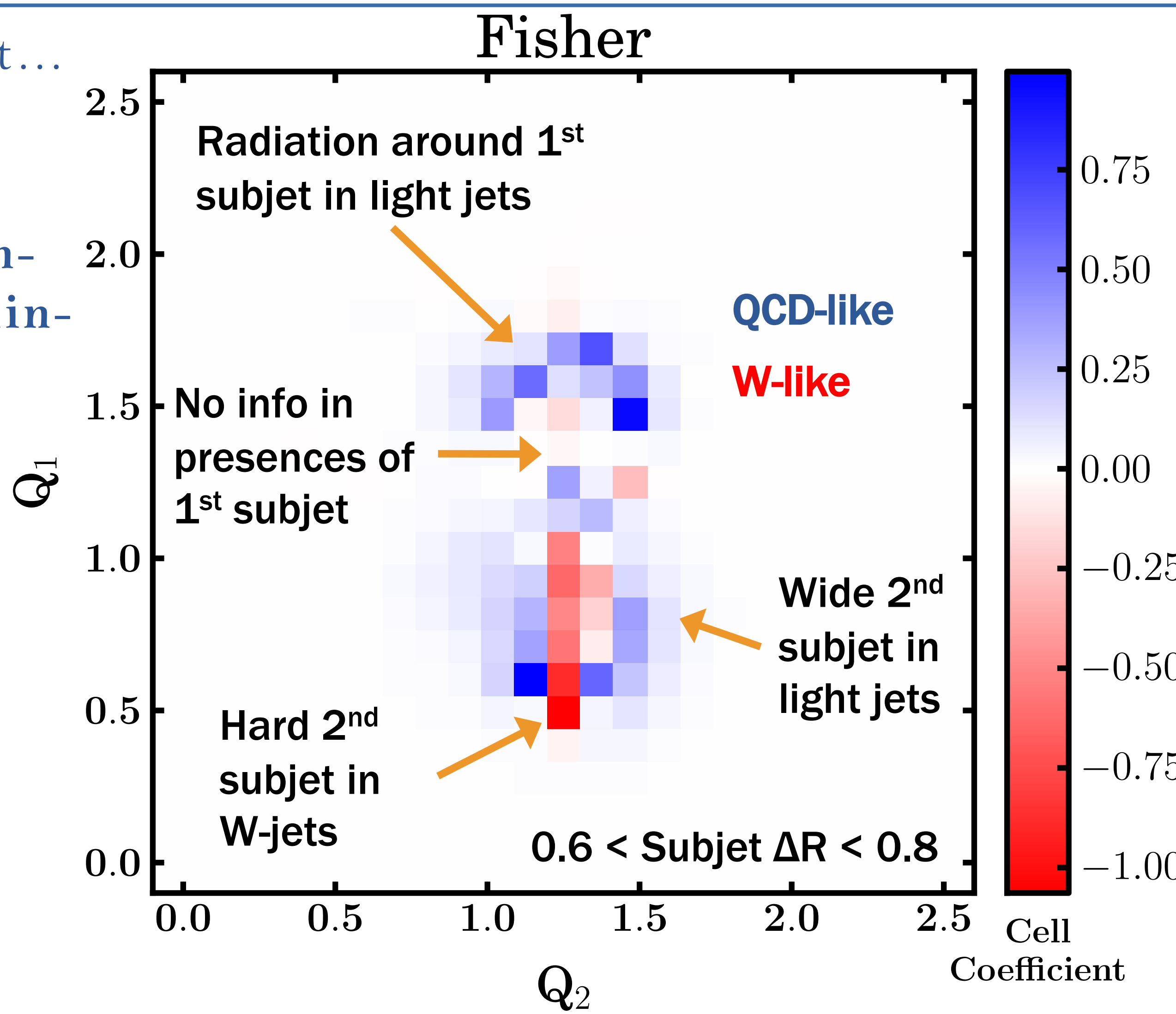
How can we extract the important features?
How can we convert this into discrimination power?

FISHER'S LINEAR DISCRIMINANT

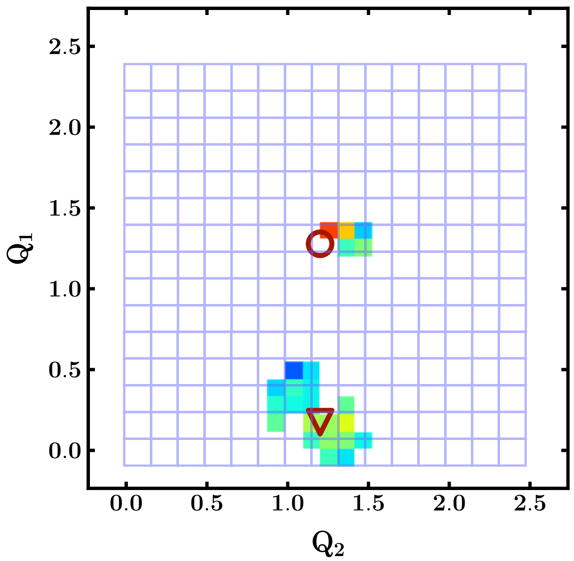
- Finds direction that…
- Maximizes **between-class scatter** / **within-class scatter**
- Extracts the single "most important" feature

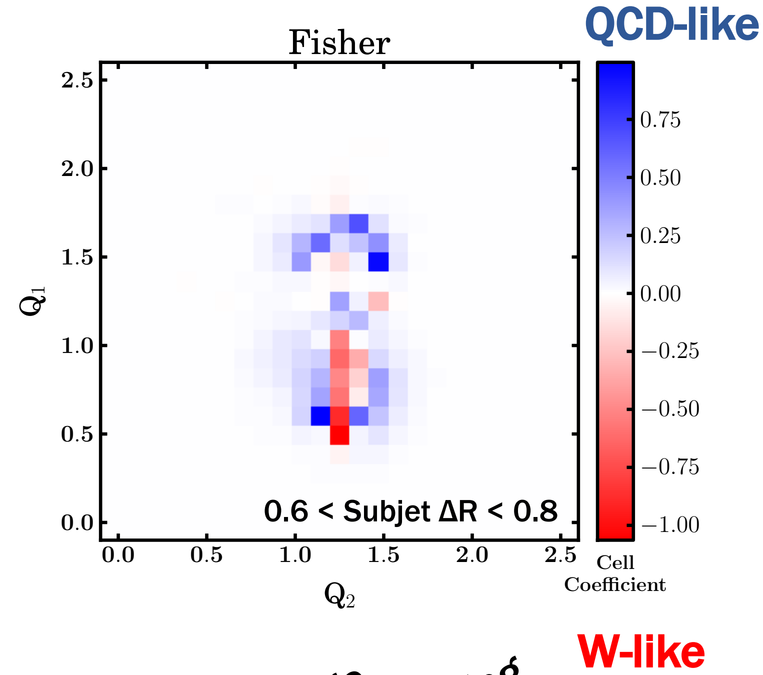
FISHER'S LINEAR DISCRIMINANT

- Finds direction that…
- Maximizes **between-class scatter** / **within-class scatter**
- Extracts the single "most important" feature



EXTRACTING A VALUE

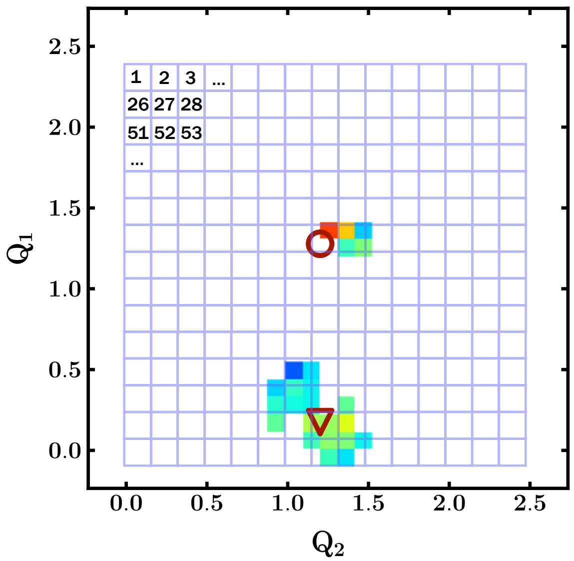




Jet-image being tested

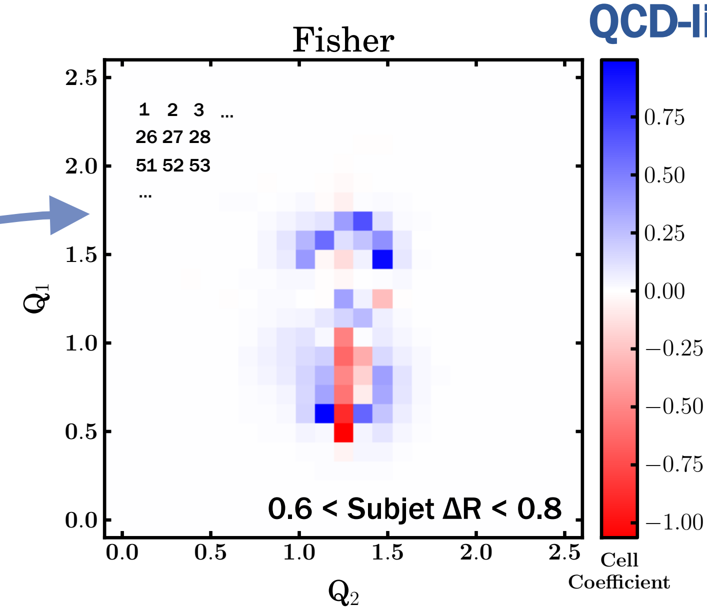
Jet-image doing the testing

EXTRACTING A VALUE



$$D = \sum p^T_i \lambda_i$$

QCD-like

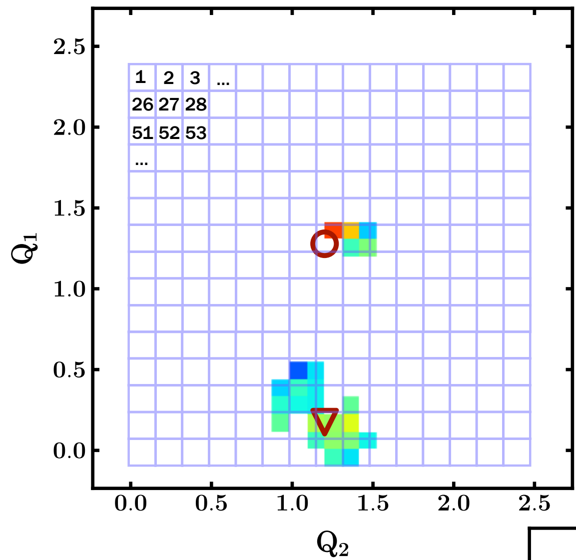


W-like

Jet-image being tested

Jet-image doing the testing

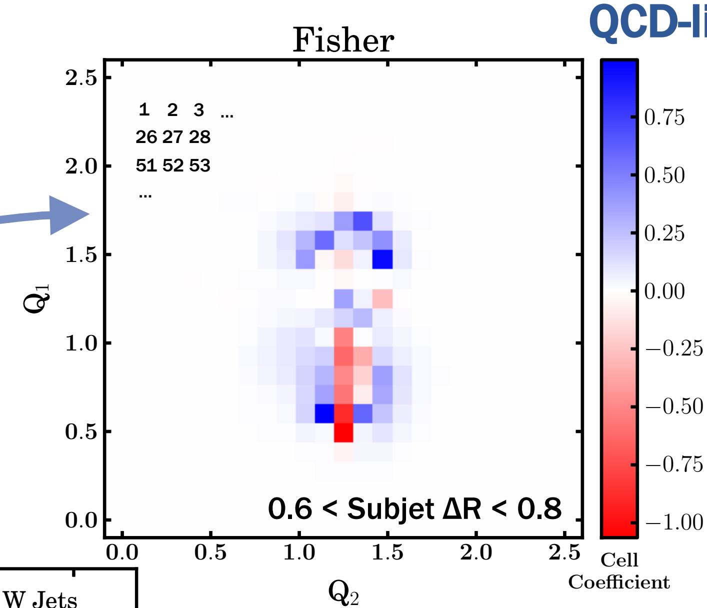
EXTRACTING A VALUE



Dot product of two jets-images!

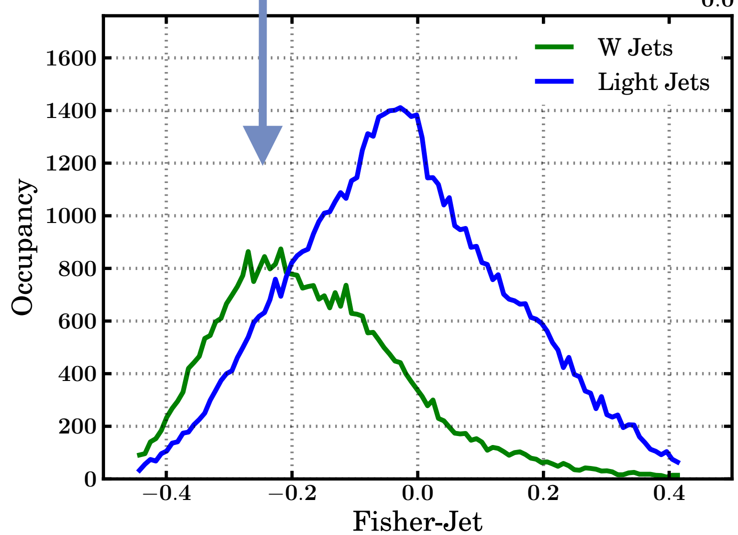
$$D = \Sigma p^T_i \lambda_i$$

QCD-like



W-like

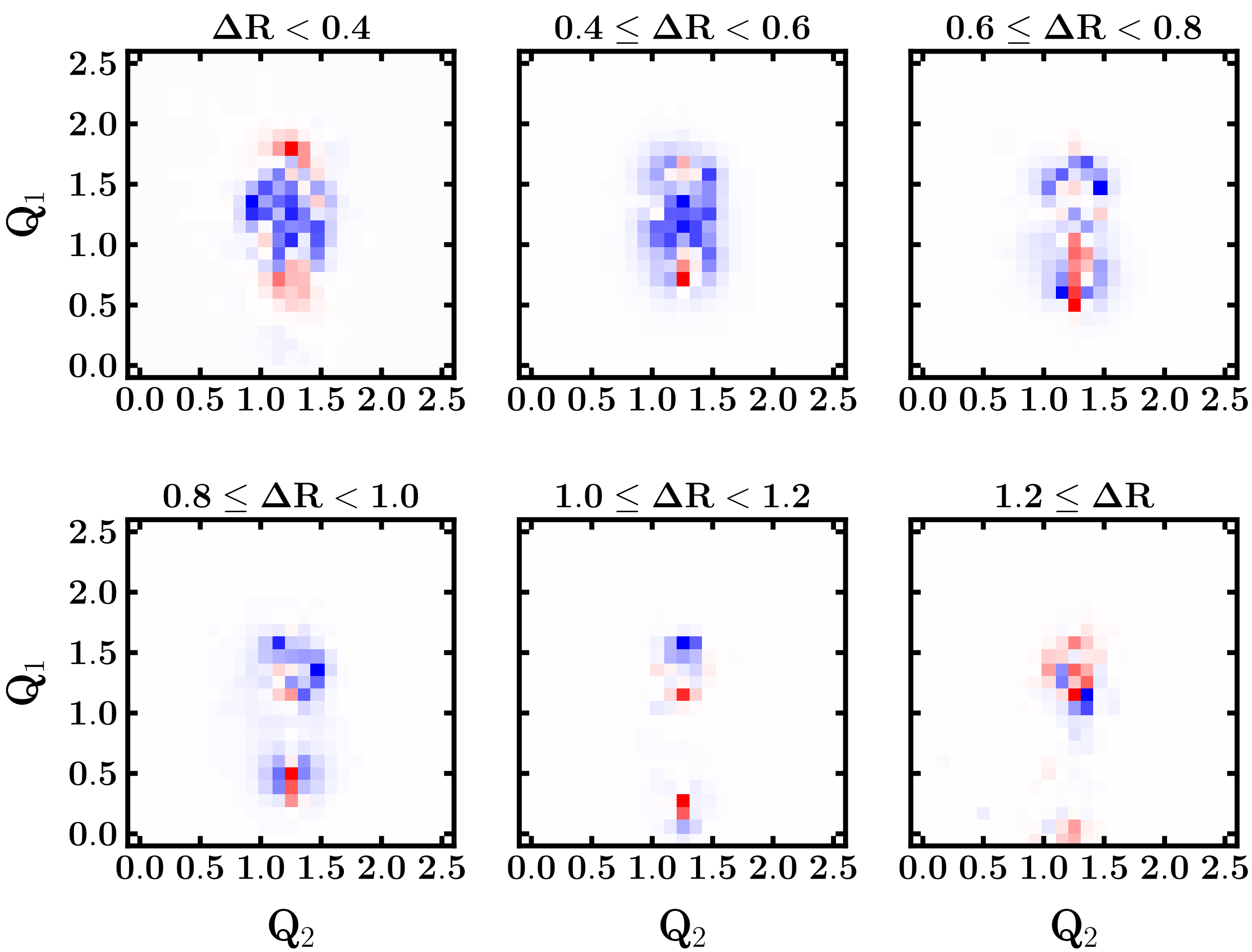
Jet-image being tested



Jet-image doing the testing

BINNED FISHER

Subjet ΔR-Binned Fisher-Jets QCD-like W-like

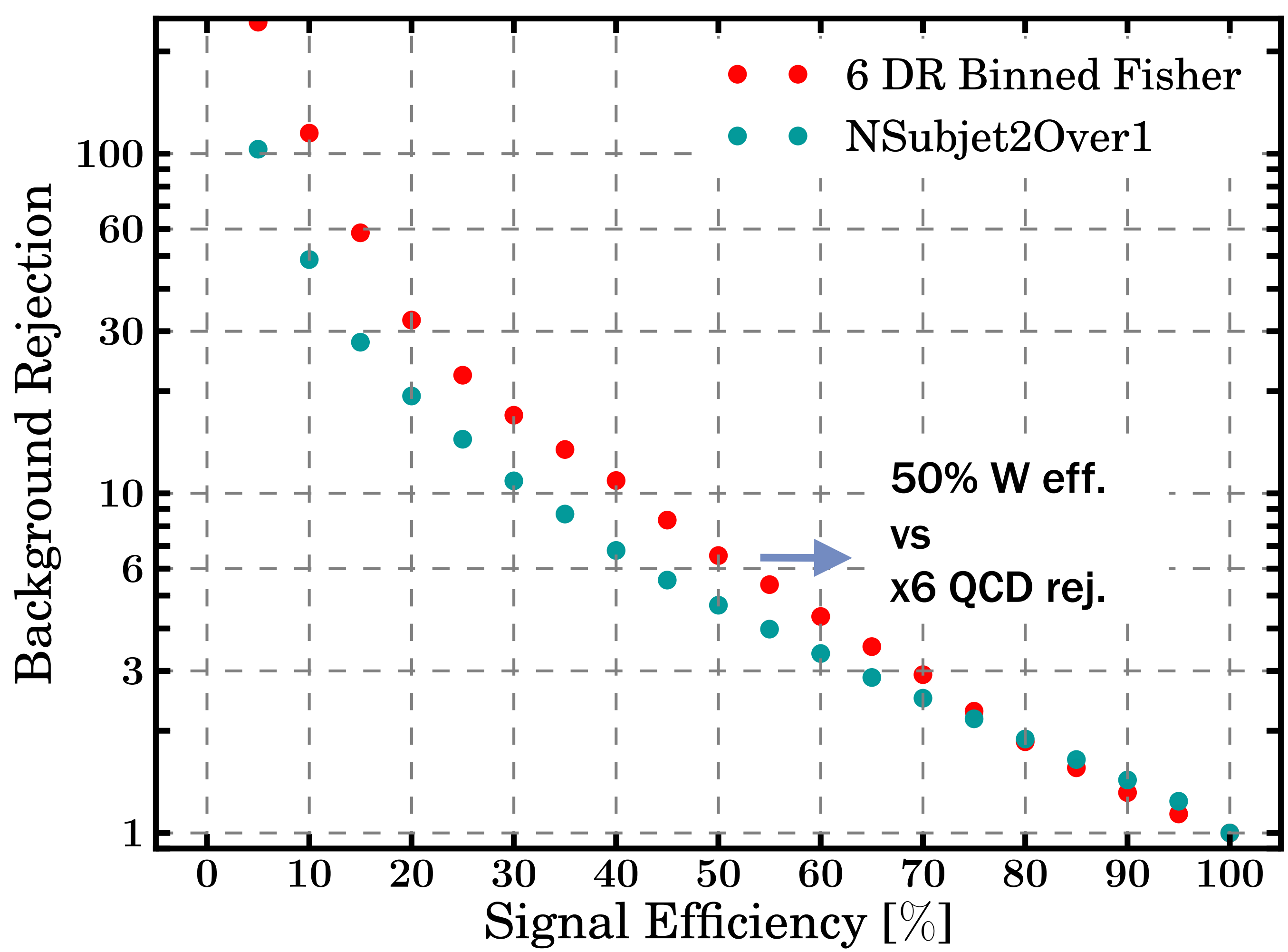


- Hadronic W vs light
- 250<pT/GeV<300
- 55<Mass/GeV<105

Subjet ΔR causes much image variation without too much discrimination

PERFORMANCE

- Use the Fisher' single feature a a discriminant?
- Comparable results to $\beta=1\ \tau_2/\tau_1$



DIFFERENT PT BINS

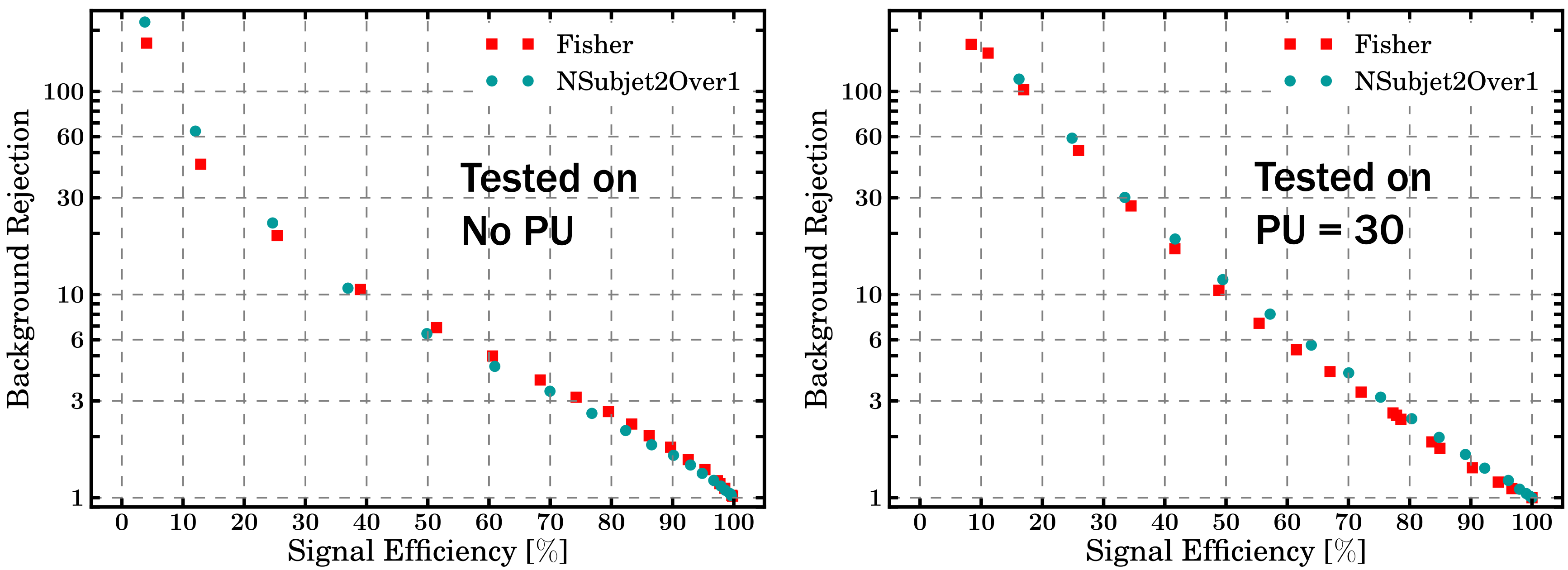
- Train new discriminant for each 50 GeV jet pT window
- More complex binning/image processing could be useful, but performance is robust against jet pT

W Efficiency @ x10 QCD Rejection

Minimum Jet pT	200	250	300	350	400	450
NSubjettiness	29%	32%	34%	34%	39%	40%
6 DR Bin Fish	34%	40%	41%	44%	46%	44%

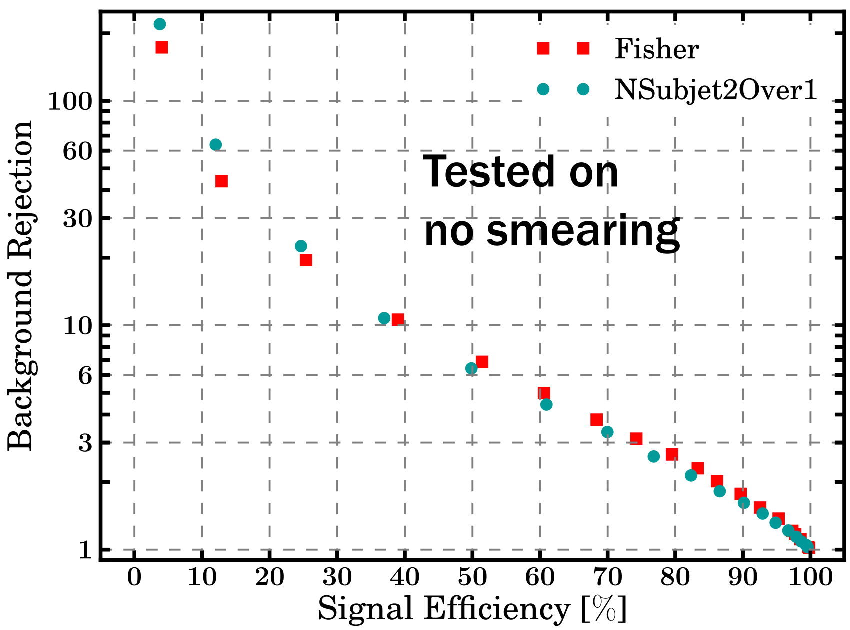
PRELIMINARY CROSS CHECKS

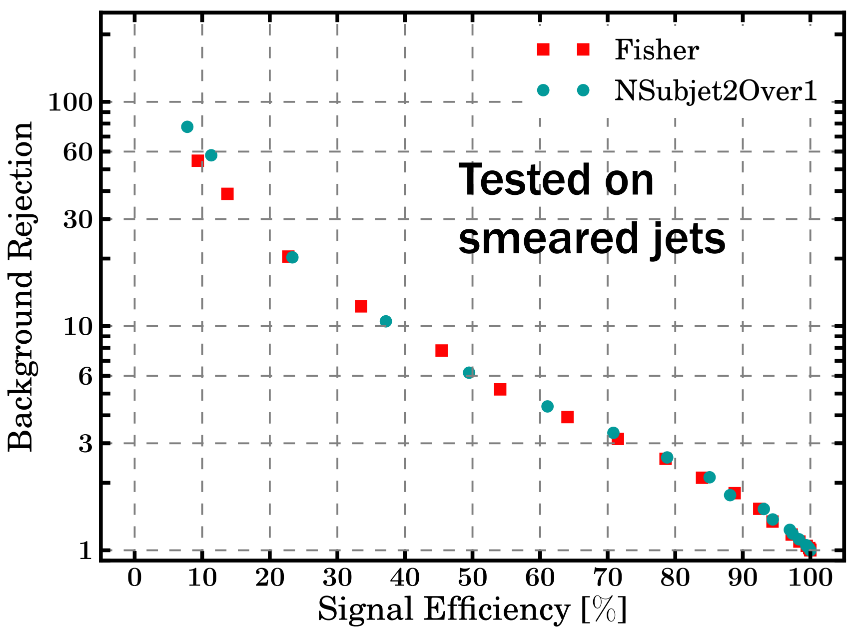
- Train on no PU samples
- Test on samples of W and light jets with μ=30
- Our performance gets "better"
  - Only because we add jets to the bkg sample that are very easy to remove
  - Happens for N-subjettiness too



CROSS CHECK: SMEARING

- As a first pass, lets try applying Peter Loch's "smearing"
  - Produces a cell-to-cell smearing by simulating particle in material
  - No significant effect on performance

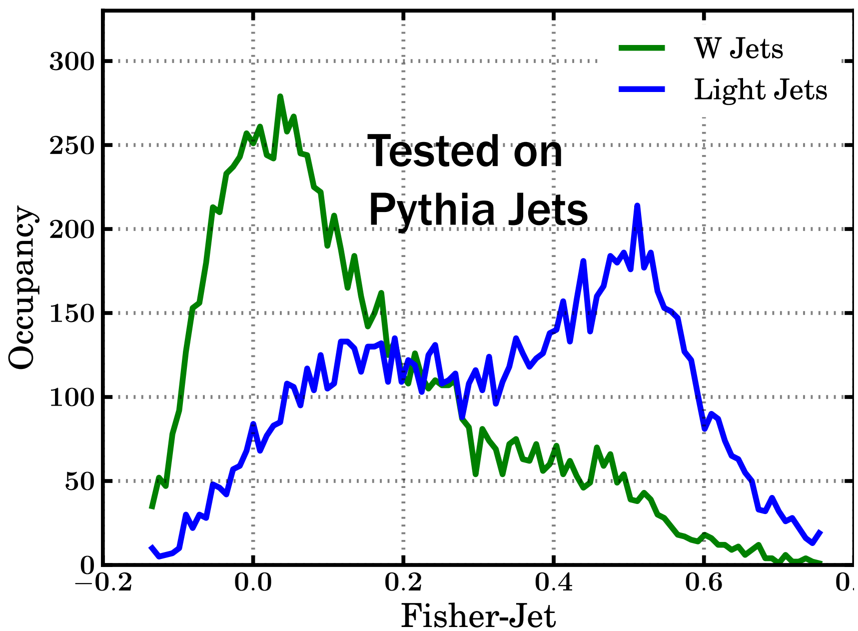




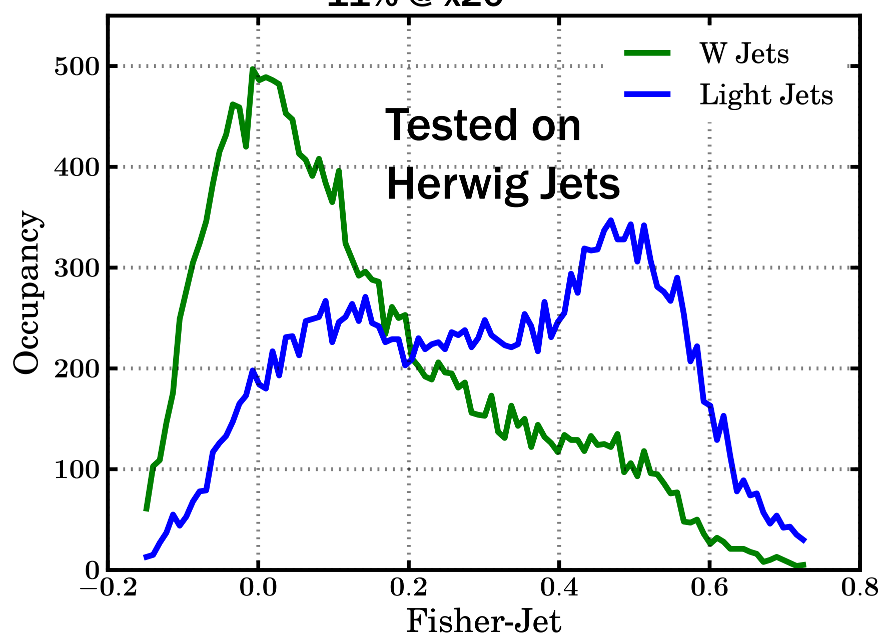
CROSS CHECKS: GENERATORS

- 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



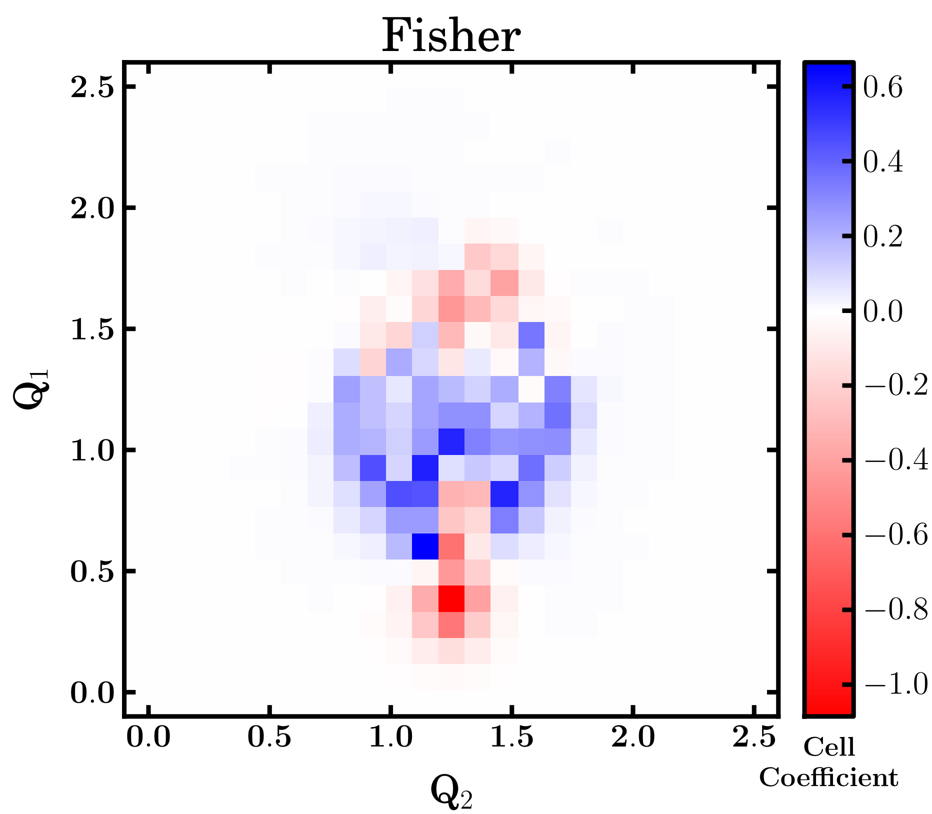
GENERALIZE

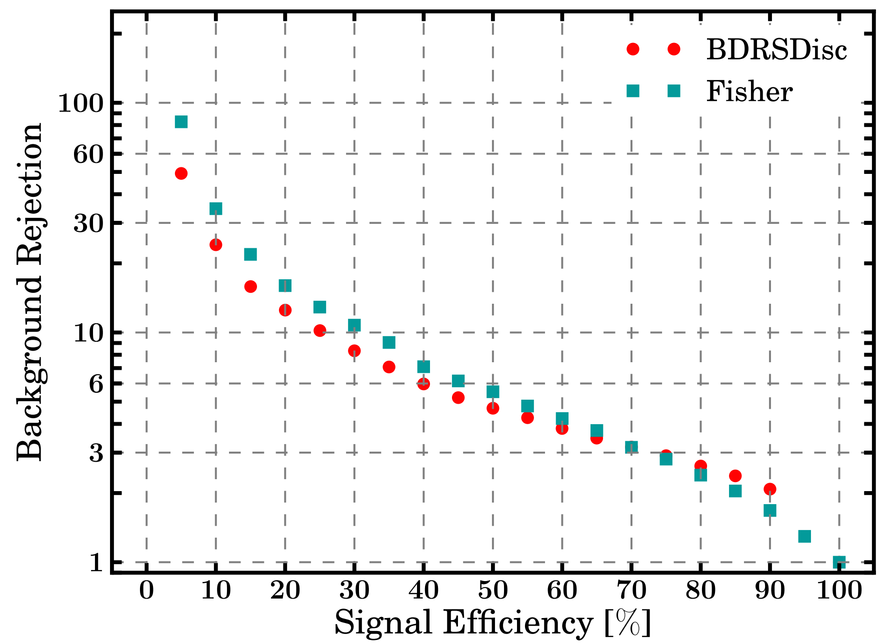
Use exact same technology

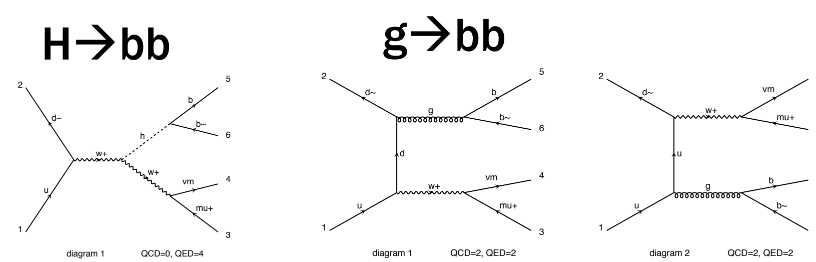
for many different systems!

Lets take a *peek*…

HIGGS TO B-QUARKS

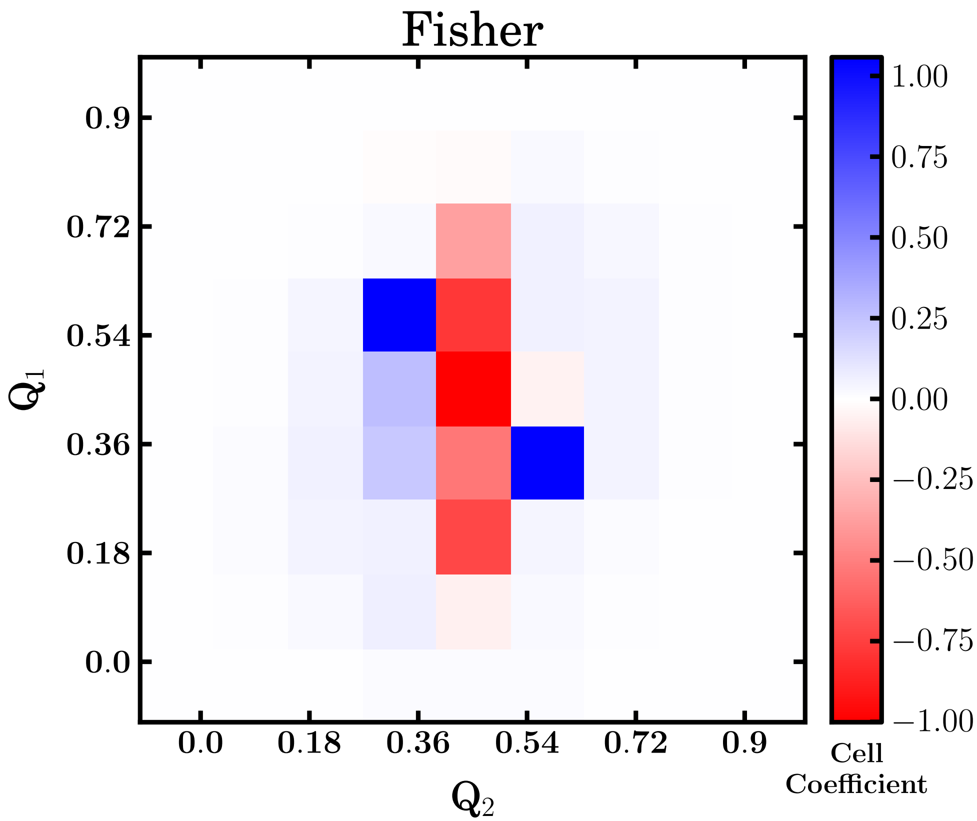


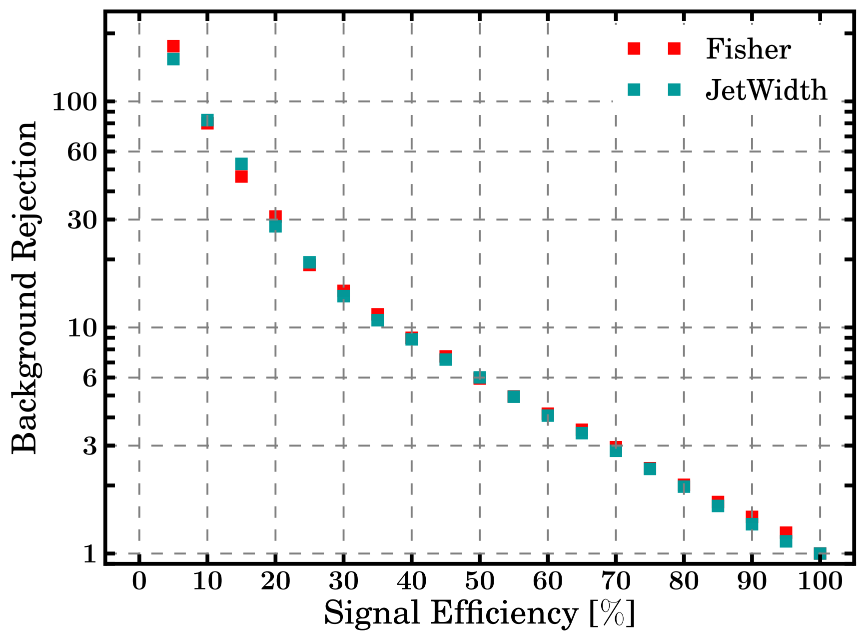




Generate with MG5
Showered with Pythia8
Two G.A. btagged subjets
125 > Jet Mass > 110

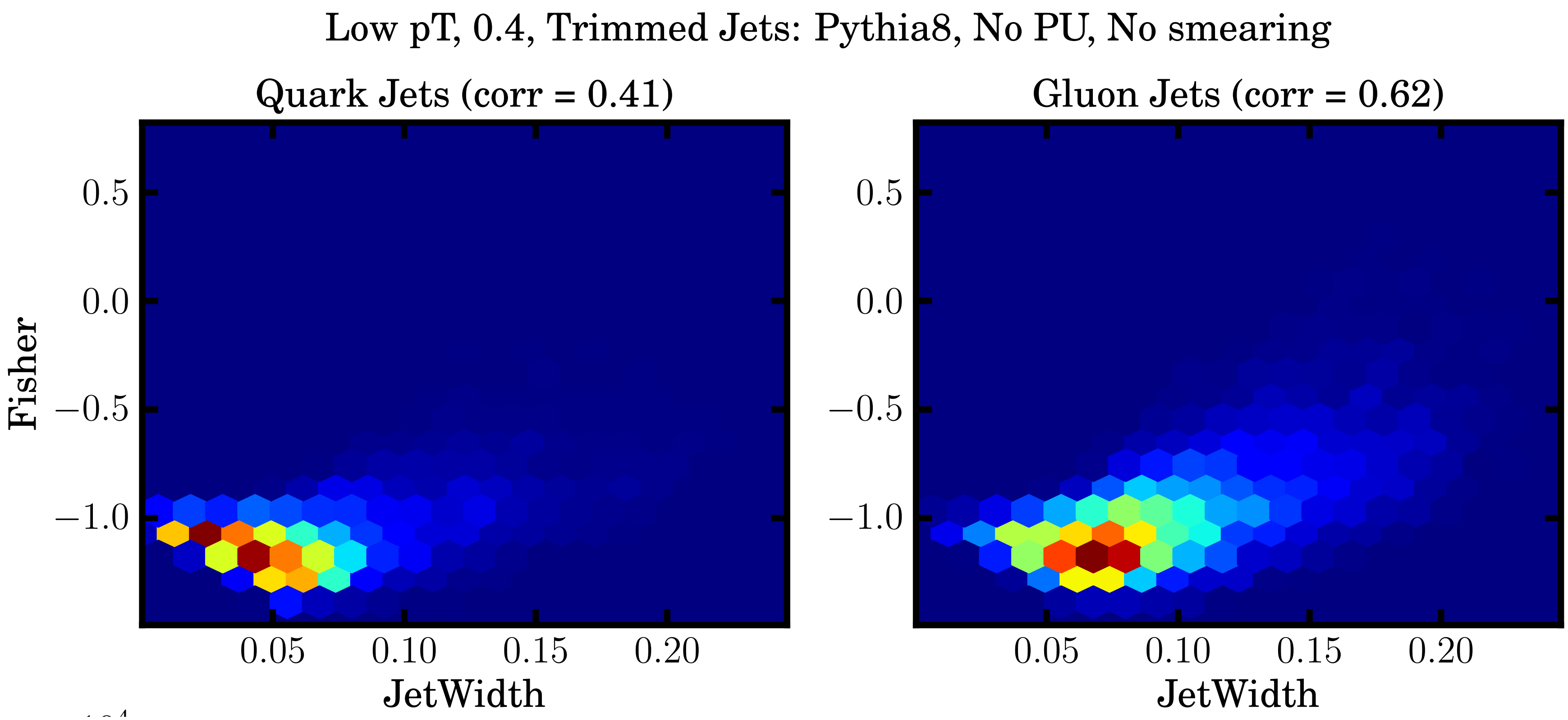
QUARK VS GLUON





Identical performance to width

QUARK VS GLUON



Width and Fisher have the same performance, but aren't completely correlated

FUTURE OPPORTUNITIES

- Many other systems to study
  - Top vs light, hadronic higgs, q/g
  - Hadronic taus, electrons in jets, photon vs π^0
  - Whole calorimeter event-images for high multiplicity signatures
  - Pile up quantification is the perfect linear problem

- Not sure preprocessing is optimized
  - Normalization, center fixing, scaling, boosting, whitening…
  - Starting from other physical bases (eg. cell pT pairs vs cell pTs)

- Today's methods only took us the year 2000 in the CS literature
  - Facial recognition is very active and new methods abound pouring out!
  - More powerful linear methods, non-linear dimensional reduction, etc…

CONCLUSION

- Conceptual mapping between **jets and images** allowed us to exploit well-studied **CS techniques** to jet flavor tagging
- Using easy-to-examine **Fisher**'s linear discriminant, produced W vs light **separation at or above N-subjettiness**
  - No need to define 'good' variables *a priori* → **flexible method!**
- Discriminant shows little sensitivity to changing: pile up, detector smearing, event generator

BACK UP

MATH OF FISHER

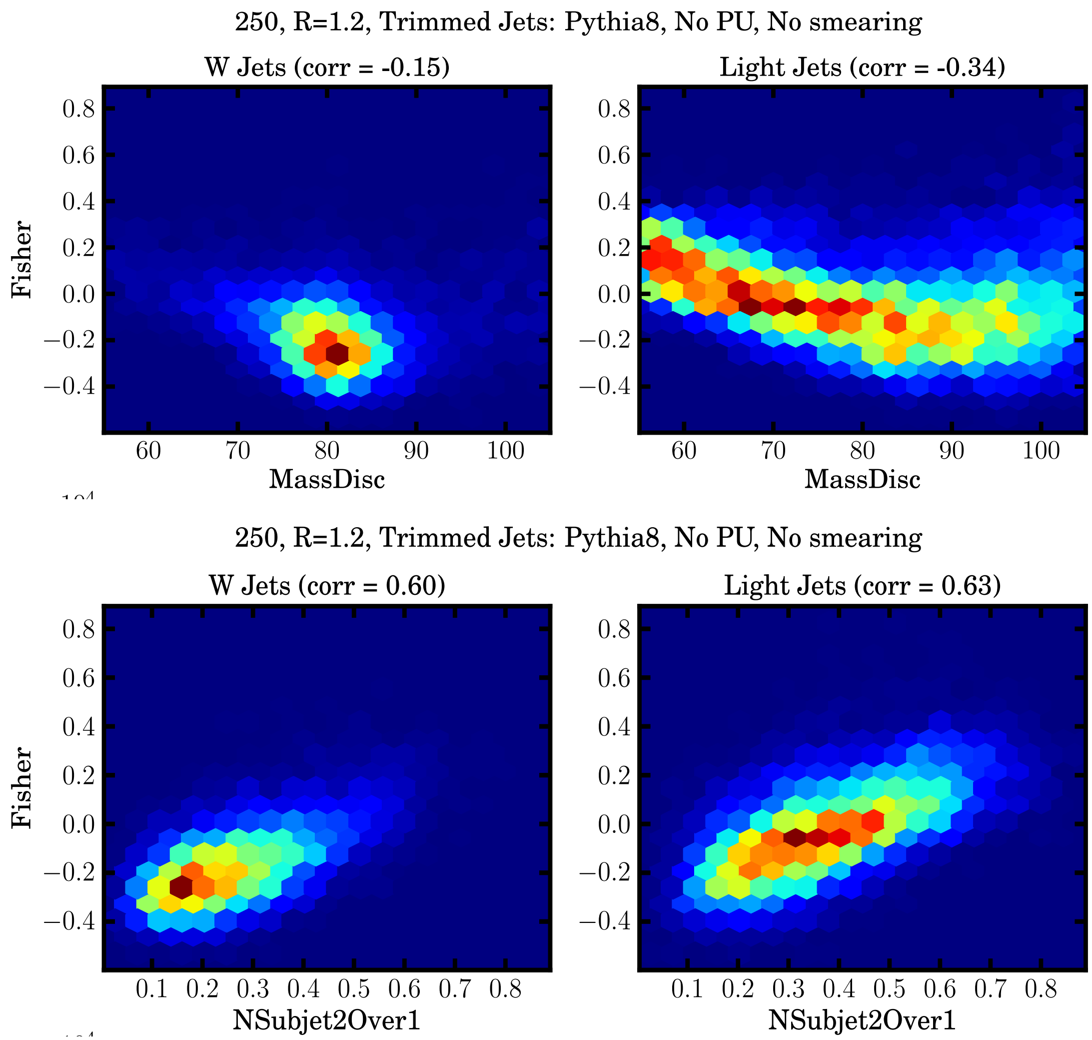
- Simplification in the case of only 2 classes…

$$[\vec{\mu}_W]_k = \frac{1}{N_{jets}} \sum_{W\ jets} p_k^T \qquad [C_W]_{i,k} = \frac{1}{N_{jets}} \sum_{W\ jets} (p_k^T - \mu_k)(p_i^T - \mu_i)$$

Average W jet **Covariance of cell p^Ts in W jets**

$$\vec{F} = \left(C_W + C_L\right)^{-1}\left(\vec{\mu}_W - \vec{\mu}_L\right)$$

FISHER 2D



0.6 < Subjet ΔR < 0.8

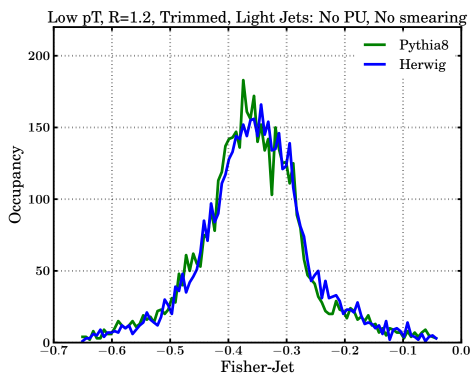
55 < Mass< 105

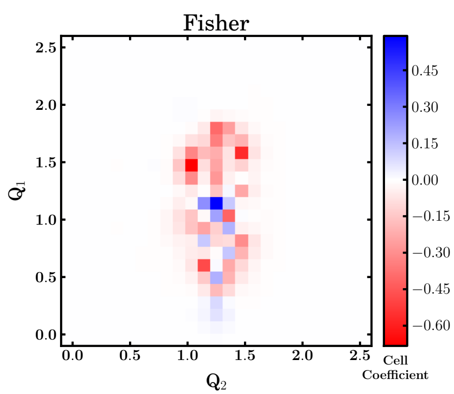
250 < jet pt < 300

COMPARING GENERATORS

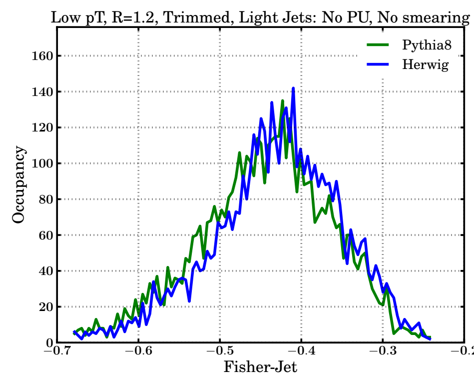
- Is there linear discrimination between Pythia / Herwig?
- Nope! But this is how they differ linearly

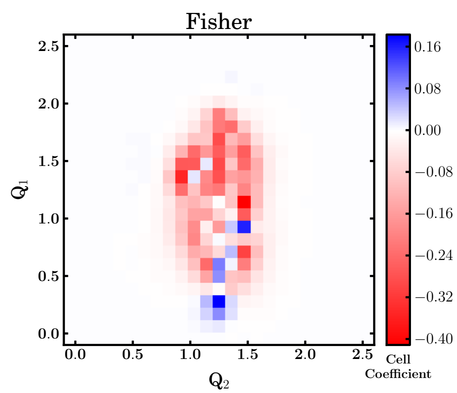
W Jets





Light Jets





Train on Pythia VS Herwig

- There are some small but measurable differences
- Further study needed but I'd hypothesize its about the pT balance

Pythia-like
Herwig-like