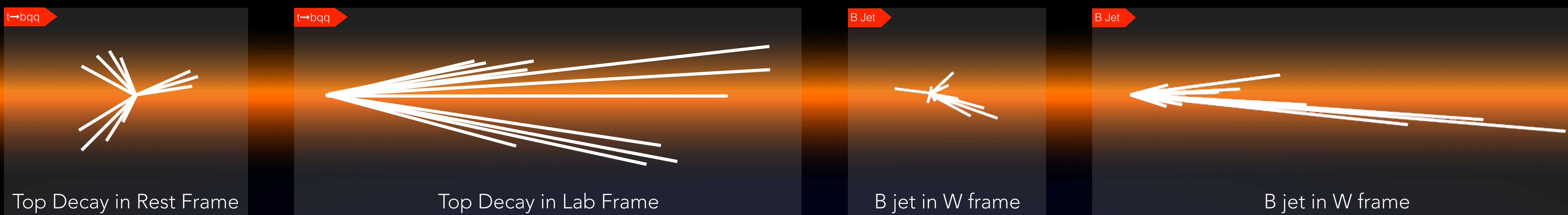
Identification of High-Momentum Top Quarks, Higgs Bosons, and W & Z Bosons using Boosted Event Shapes

Ramya Bhaskar, J. S. Conway, R. D. Erbacher, J. Pilot. University of California at Davis



Motivation

LHC collaborations ATLAS and CMS have developed powerful new analysis techniques to identify high-momentum ("highly boosted") hadronically decaying top quarks, Higgs bosons, and W & Z bosons, while rejecting background from multi-jet final states produced in QCD processes.

We present a new approach in identifying boosted massive particles. Our goal is a generic algorithm which can be used to determine the origin of individual "fat" jets with large transverse energy (E_T) produced in p-p collisions at the LHC.

Ideally this will optimize tagging t quarks, H bosons, and W & Z bosons, while discriminating against jets from light quarks and gluons.

Boosted Event Shapes

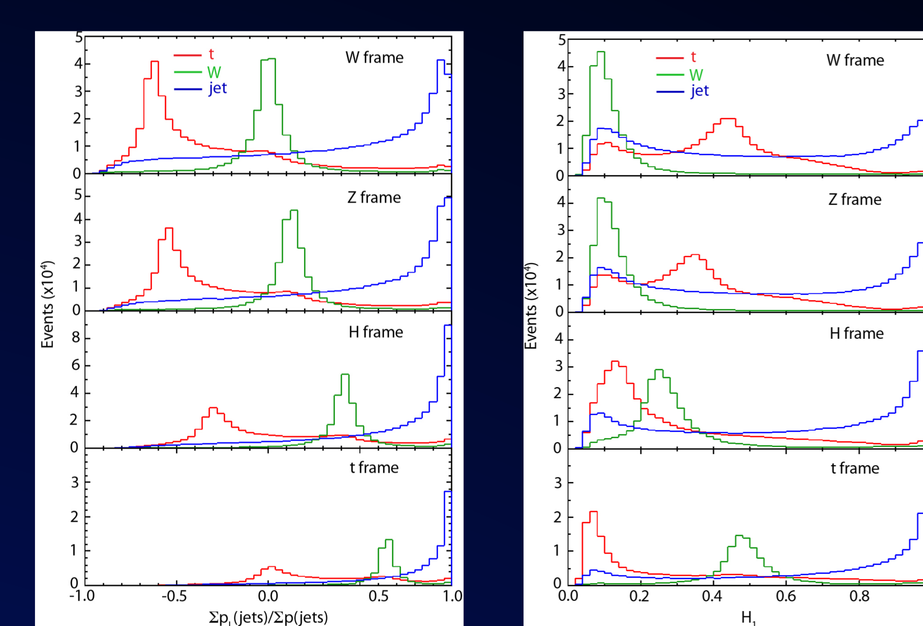
To generate Boosted Event Shapes (BES):

- Start with fat (R=0.8) jets
- Lorentz-boost jet constituents into t/H/Z/W frames

  Calculate:
  - Fox-Wolfram moments
  - sphericity tensor eigenvalues
  - thrust, aplanarity
  - k_T jets in boosted frame
- In correct frame expect two or three jets, balanced momentum
- For each jet use 50 variables:
  - E_T, eta
  - H_1 - H_4, I_1 - I_3, T, S_{pL}, S_p for each frame
  - E_{T1} - E_{T4}, m_{12}, m_{13}, m_{23}, $m_{1\text{-}4}$ for W frame

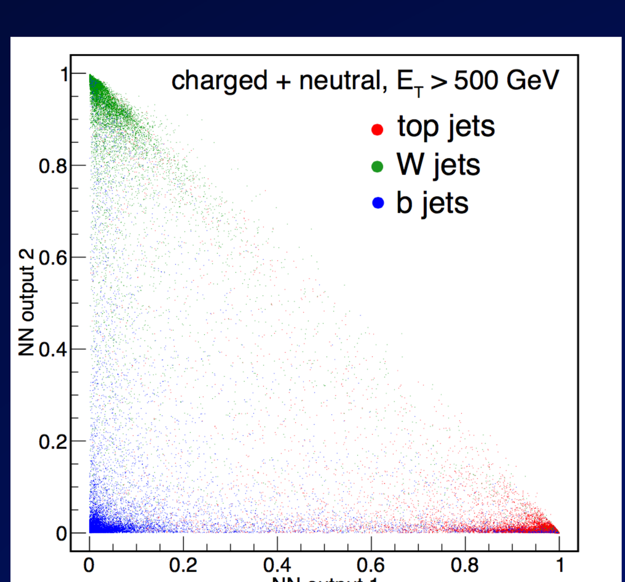
Simulations

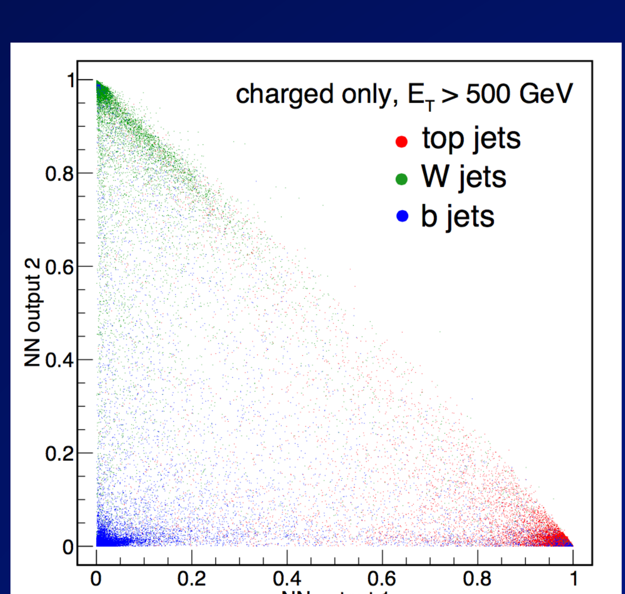
- We use PYTHIA 3 TeV Z′ → $t\bar{t}$, HZ, W^+W^-, $b\bar{b}$
- Just used generator-level final state particles
- Impose 90% efficiency
- 2 GeV p_T threshold



Distributions of Fox-Wolfram moment H_1 (left) and ratio of longitudinal to total momentum of anti-k_T jets (right)

Neural Network BES Discriminator





ANN outputs of jets originating from all hadronic top quark decay, W decay, and b quarks for nets trained with charged & neutral (top) and only charged constituents (bottom)

Due to the large number of variables, we deploy artificial neural networks (ANNs) for the analysis. For tagging 3 jet types t, W, and b:

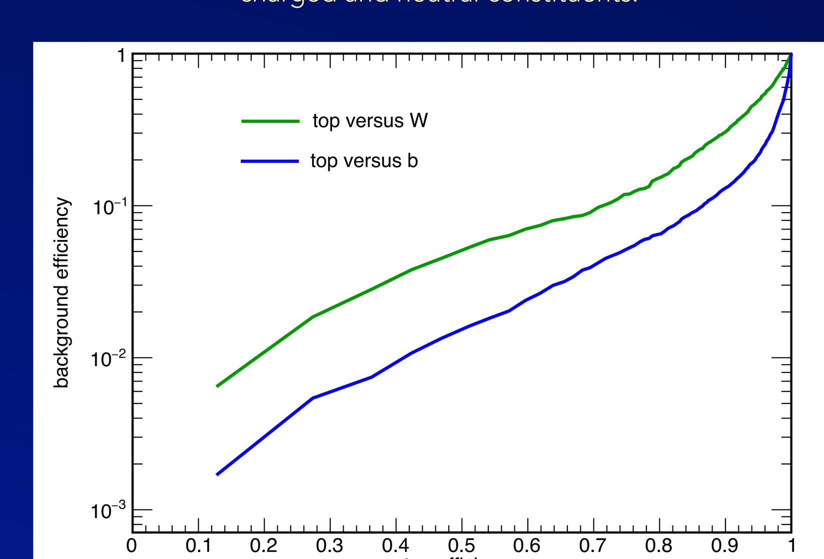
- Create training and test samples with 3.0 TeV Z′ events
  - 50 variables per event
- Feed-forward net trained with vanilla back-propagation
  - learning strength 0.001; momentum 0.5
  - 3 outputs: b, W, and t jets
  - Stability achieved well before 1000 epoch training cycle
  - Varying layers & nodes minimally affected performance
- Tested BES algorithm with charged tracks only
  - more confusion between neutral particles at high momenta

Performance of top quarks separating from W bosons remains nearly the same, but is degraded for top from b jets.

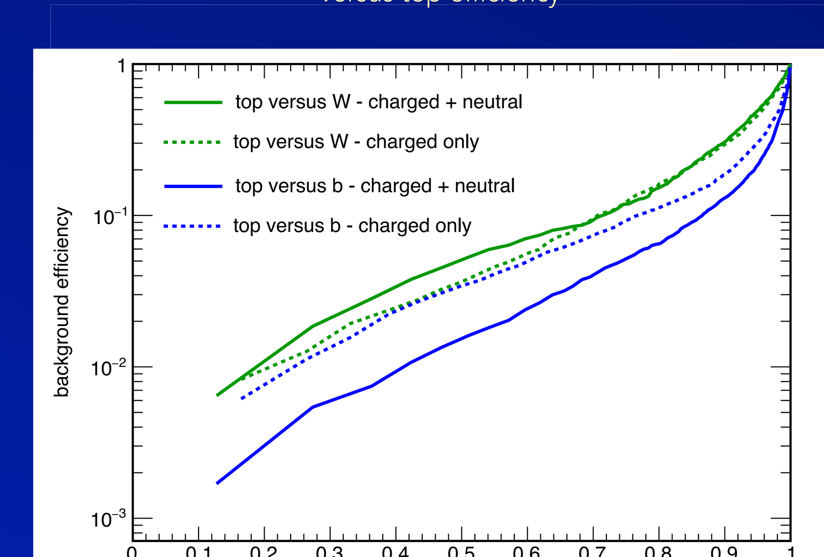
Though results are encouraging, an actual LHC analysis would not achieve such high performance levels—largely because jet constituents are not as accurate as in this prototype study.

Generated:	t	W	b
Classified as t:	84.2%	3.9%	10.2%
Classified as W:	5.6%	86.8%	5.9%
Classified as b:	10.4%	9.3%	83.9%

Probabilities for classification of t, W, and b jets, while training on charged and neutral constituents.



Initial result ROC curves show background rejection versus top efficiency



Efficiency of tagging jets from top decay (horizontal axis) versus efficiency for W jets and b jets

Also investigated was whether the new algorithm could tag all-hadronic top quark decays, H Boson → $b\bar{b}$, Z → $q\bar{q}$, and W → $q\bar{q}$ when decaying particles are highly boosted.

For tagging 5 jet types:

- Use PYTHIA again
- Pass through PGS
- 4 outputs: top, H, Z, W

We represent the results of this 5 jet problem as a 4 dimensional plot projected onto 2 dimensions.

〚 Event Shape Variables 〛

$$H_l = \sum_{i,j} \frac{|\mathbf{p}_i||\mathbf{p}_j|}{E_{vis}^2} P_l(cos\theta_{ij})$$

〔Fox Wolfram Moments〕

$$T = \max_{|\mathbf{n}|=1} \frac{\sum_i |\mathbf{n}\cdot\mathbf{p}_i|}{\sum_i |\mathbf{p}_i|}$$

〔Thrust〕

$$S^{\alpha\beta} = \frac{\sum_i p_i^{\alpha} p_j^{\beta}}{|\mathbf{p}_i|^2}$$

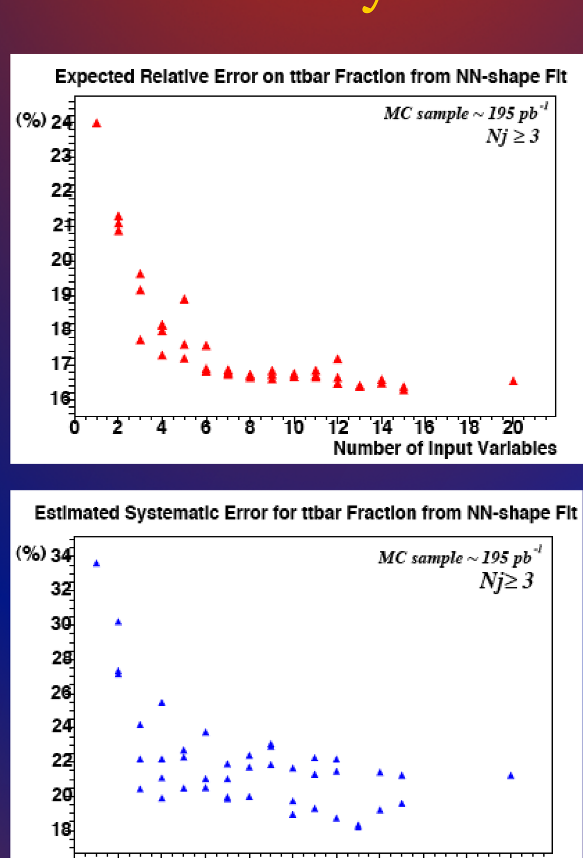
〔Sphericity〕

$$S = \frac{3}{2}(\lambda_2 + \lambda_3)$$

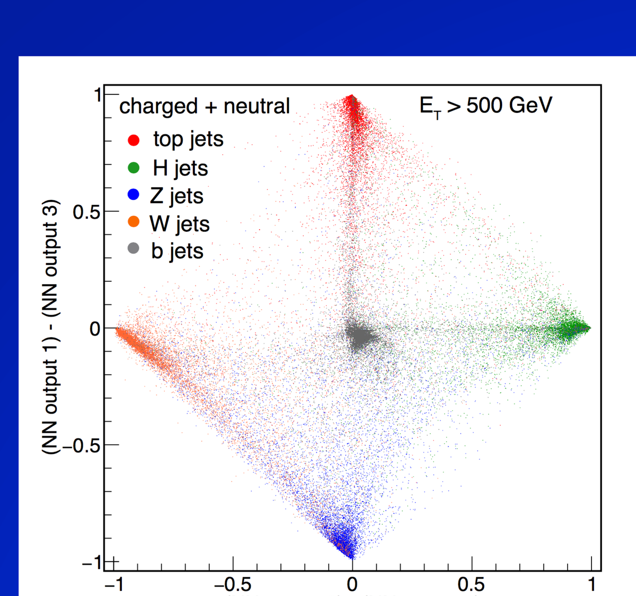
$$A = \frac{3}{2}\lambda_3$$

〔Sphericity〕

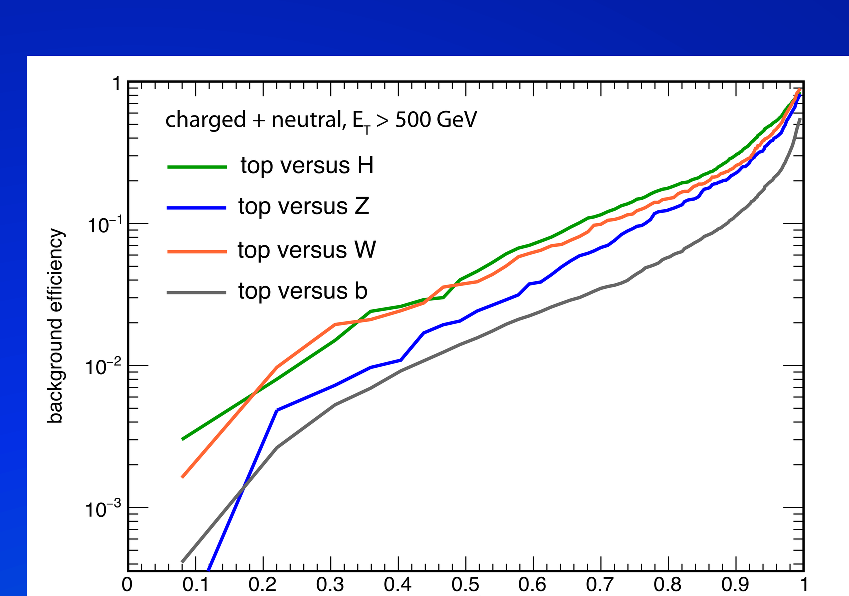
Worried about Systematics?



R. Marginean, Ph.D. thesis, Ohio State, CDF 2006



Distributions of ANN outputs for a 4-output net, tagging jets from all-hadronic top quark decay, H → $b\bar{b}$ decay, Z → $q\bar{q}$ decay, W → $q\bar{q}$ decay, and b quarks



Efficiency of tagging jets from top decay (horizontal axis) versus efficiency for H, Z, W, and b jets

Summary

We have a new algorithm to identify very high momentum hadronically decaying top quarks, Higgs bosons, and vector bosons, by boosting into rest frames of hypothetical particle masses and then calculating a set of event shape quantities.

The algorithm is capable of separating jets which physically overlap in laboratory momentum space. In the idealized scenario, we see excellent separation of hadronic jets, largely maintained by using only charged jet constituents. Variables in this algorithm can be readily combined with other boosted particle tagging techniques.