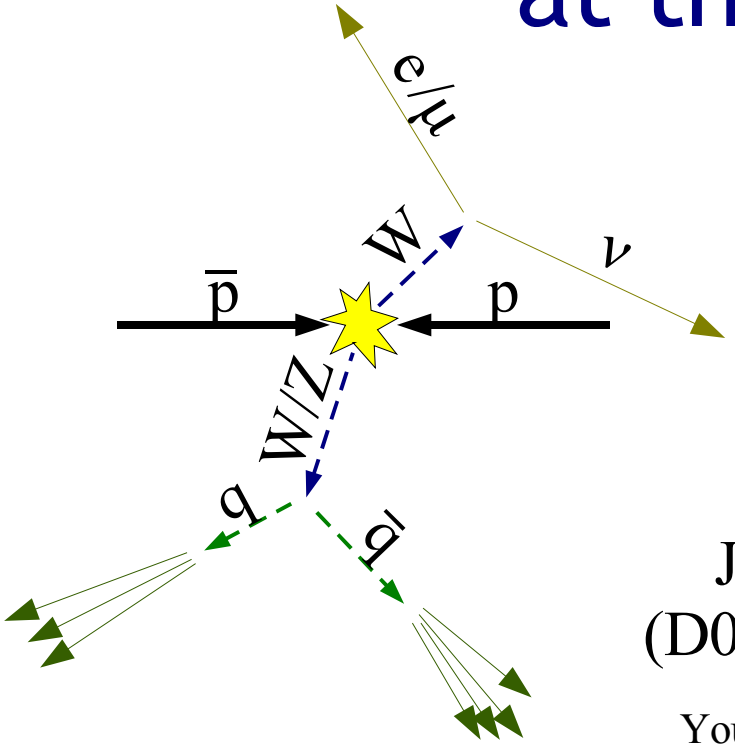
First Evidence of $WW/WZ \to l\nu q\bar{q}$ at the Tevatron



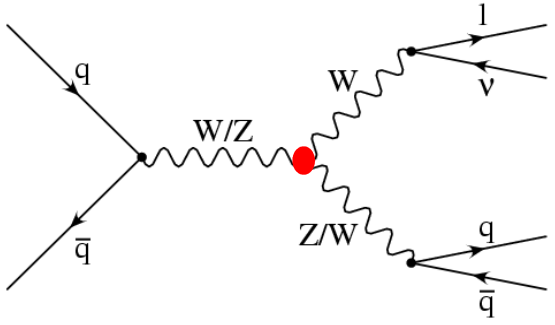
- Motivation
- Selection
- Multivariate Classification
- Cross Section Measurement
- Significance

Joseph Haley
(D0 Collaboration)

Young Scientists Forum
Rencontres de Moriond – EW
March 2009

Motivation

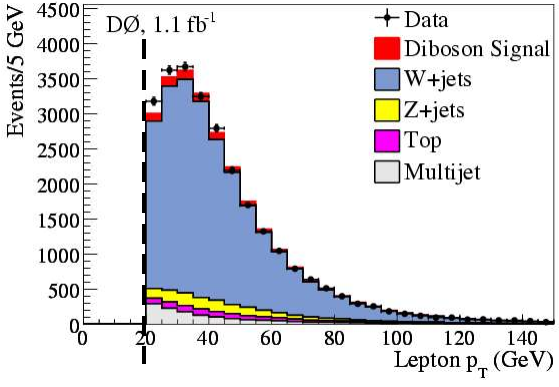
- Probe of the Electroweak sector of the SM at the Tevatron
  - Sensitive to anomalous trilinear gauge-boson couplings
  - Complimentary to measurements in the fully leptonic channels

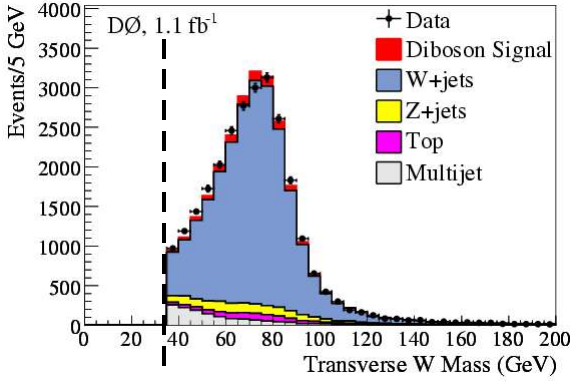


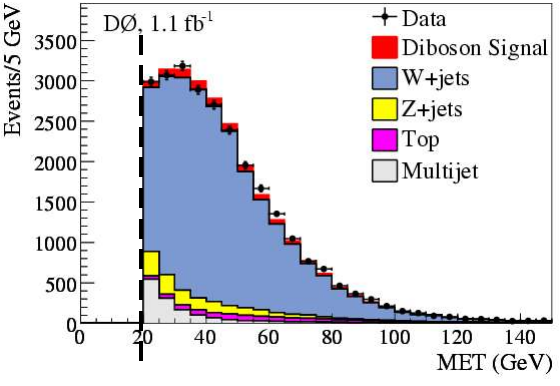
- Background to Higgs searches
- Benchmark for low mass Higgs (m_H < 135 GeV/c^2)
  - Very similar to *WH→lνbb*
    - Same challenge of large *W*+jets background
    - Similar Multivariate classifiers, statistical treatment

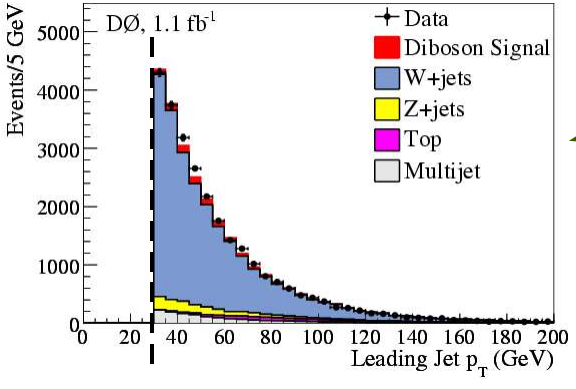
⇒ Proving ground for these analysis techniques

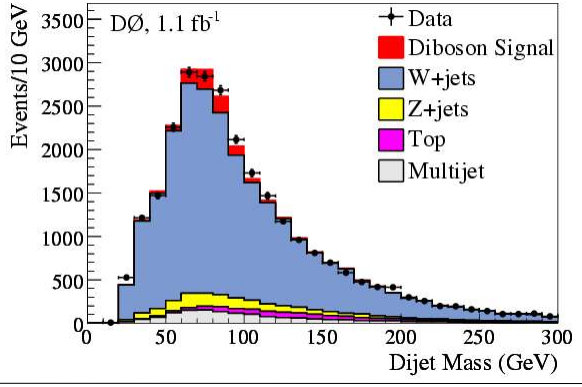
Selection

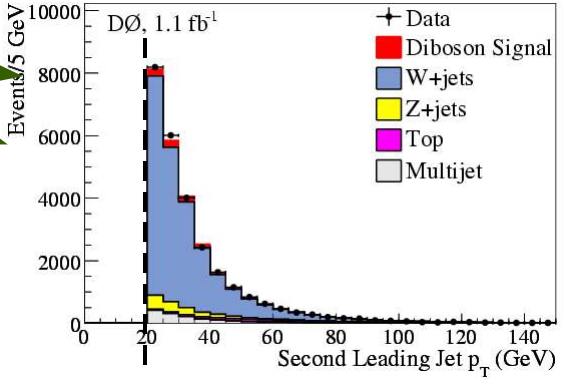

DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/5 GeV
Lepton pT (GeV)

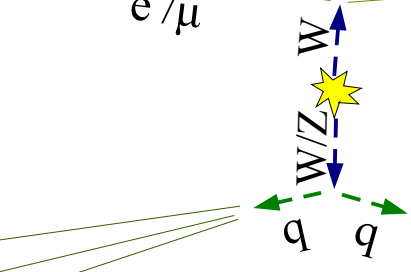

DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/5 GeV
Transverse W Mass (GeV)


DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/5 GeV
MET (GeV)


DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/5 GeV
Leading Jet pT (GeV)

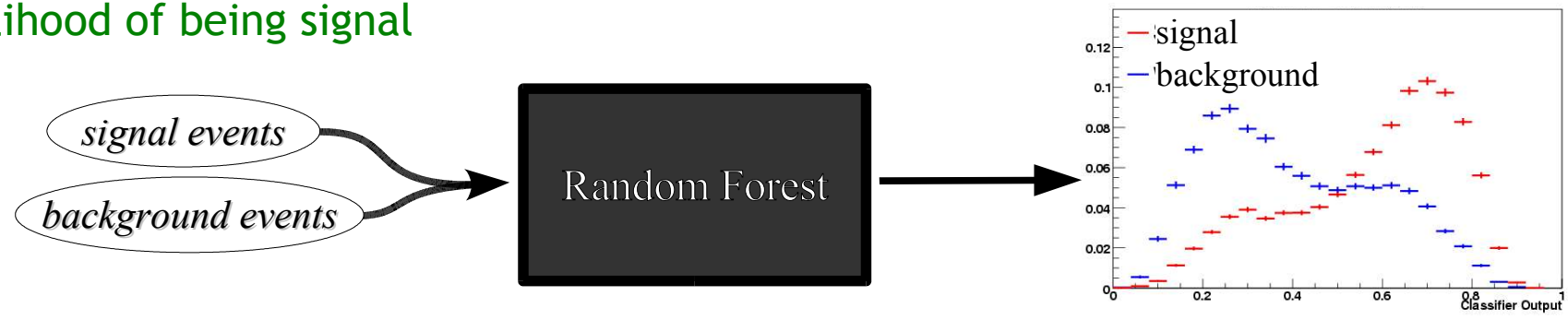

DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/10 GeV
Dijet Mass (GeV)


DØ, 1.1 fb-1
Data
Diboson Signal
W+jets
Z+jets
Top
Multijet
Events/5 GeV
Second Leading Jet pT (GeV)


e /μ
ν
W
W/Z
q
q

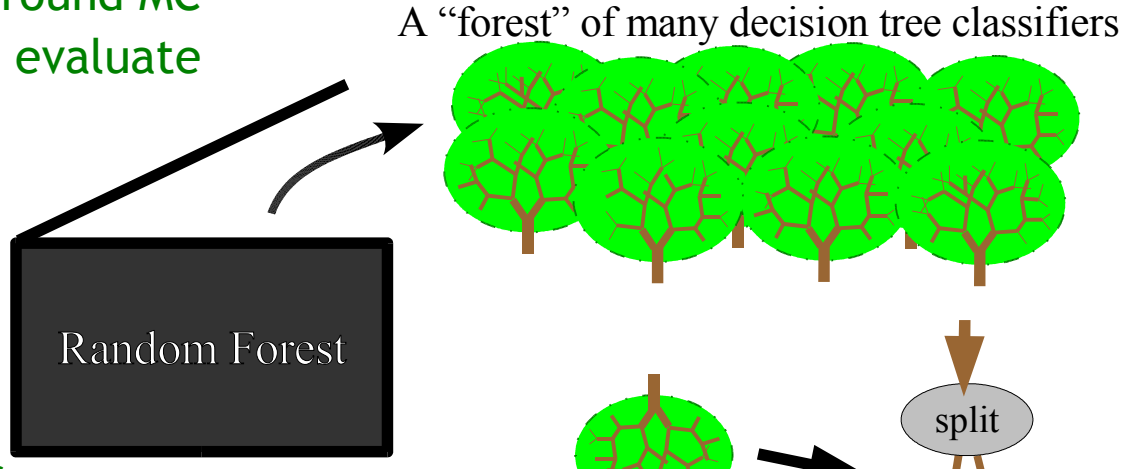
Multivariate Classification

- Improve signal and background separation w/ a multivariate classifier
  - Found Random Forest (RF) classifier to be the most powerful and robust
- From outside (black box), RF is like other classifiers (e.g. NN)
  - Trained with signal and background MC
  - Use trained Random Forest to evaluate new events → determine the likelihood of being signal

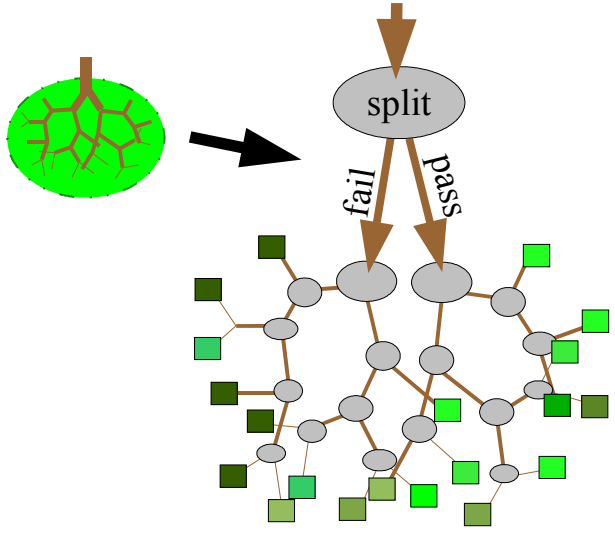


Multivariate Classification

- Improve signal and background separation w/ a multivariate classifier
  - Found Random Forest (RF) classifier to be the most powerful and robust
- From outside (black box), RF is like other classifiers (e.g. NN)
  - Trained with signal and background MC
  - Use trained Random Forest to evaluate new events → determine the likelihood of being signal



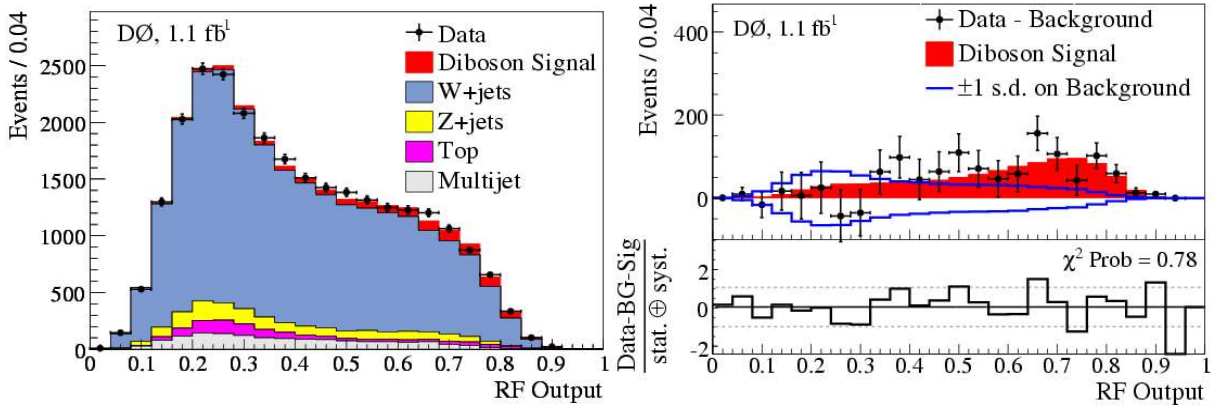
- Inside the RF
  - Many different tree classifiers
    - Each tree classifier performs a series of optimized cuts to separate signal from background
  - The RF output averages the output from all the trees
    - Fluctuations and over-training are reduced because each tree will fluctuate differently



Cross Section Measurement

- Measuring WW+WZ cross section
  - Perform "best-fit" of the signal and backgrounds to data using the RF distribution
    - Minimize Poisson χ^2 between data and MC
    - Allow signal and background distributions to fluctuate within uncertainties
    - Signal cross section is a free parameter determined by fit
    - Also allow fit to determine W+jets normalization

RF distributions after best-fit



$\Rightarrow$

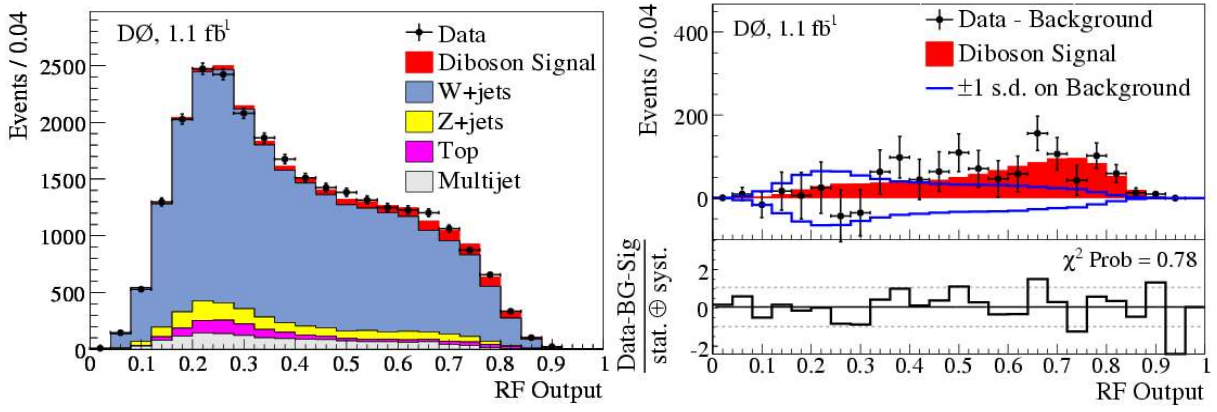
Channel	Measured Signal Cross Section [pb]
$e\nu q\bar{q}$	18.0 $\pm$3.7(stat) $\pm$5.2(sys) $\pm$1.1(lum)
$\mu\nu q\bar{q}$	22.8 $\pm$3.3(stat) $\pm$4.9(sys) $\pm$1.4(lum)
Combined	20.2 $\pm$2.5(stat) $\pm$3.6(sys) $\pm$1.2(lum)

- SM prediction: $\sigma(WW+WZ)$ = 16.1 ± 0.9 pb

Cross Section Measurement

- Measuring WW+WZ cross section
  - Perform "best-fit" of the signal and backgrounds to data using the RF distribution
    - Minimize Poisson χ^2 between data and MC
    - Allow signal and background distributions to fluctuate within uncertainties
    - Signal cross section is a free parameter determined by fit
    - Also allow fit to determine W+jets normalization

RF distributions after best-fit



$\Rightarrow$

Channel	Measured Signal Cross Section [pb]
$e\nu q\bar{q}$	18.0 ±3.7(stat) ±5.2(sys) ±1.1(lum)
$\mu\nu q\bar{q}$	22.8 ±3.3(stat) ±4.9(sys) ±1.4(lum)
Combined	20.2 ±2.5(stat) ±3.6(sys) ±1.2(lum)

- SM prediction: $\sigma(WW+WZ) = 16.1 \pm 0.9$ pb

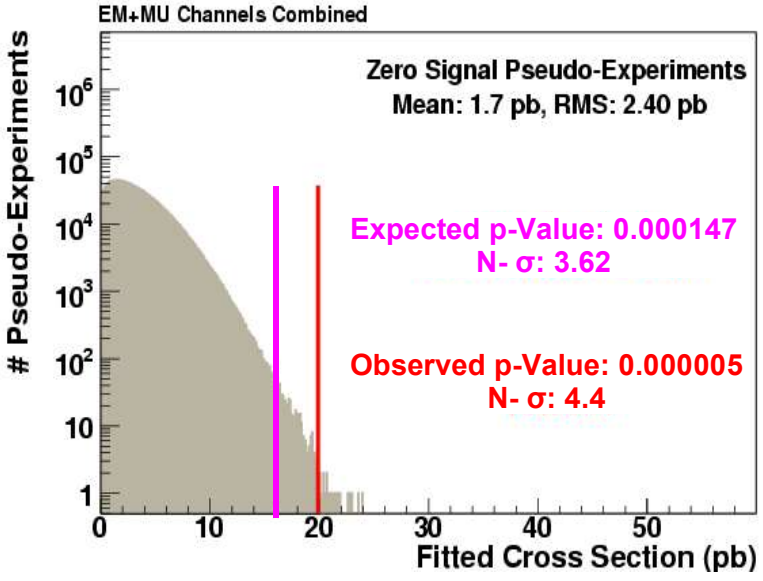
- Measured using the Dijet Mass: $\sigma(WW+WZ) = 18.5 \pm 2.8(\text{stat}) \pm 4.9(\text{sys}) \pm 1.1(\text{lum})$ pb

Significance

- Estimation of significance
  - Frequentist approach: count fraction of times the background-only scenario would mimic a signal cross section as large as we measured (or expect)
    - Generate pseudo-data for background-only scenario
      - randomly sample the systematics from their priors, then drawing Poisson trials for each bin
    - Fit the background-only pseudo-data outcomes in the same was as real data

⇒ Expected significance: 3.6σ (p-value = $1.5 \cdot 10^{-4}$)

⇒ Observed significance: 4.4σ (p-value = $5.4 \cdot 10^{-6}$)



Significance

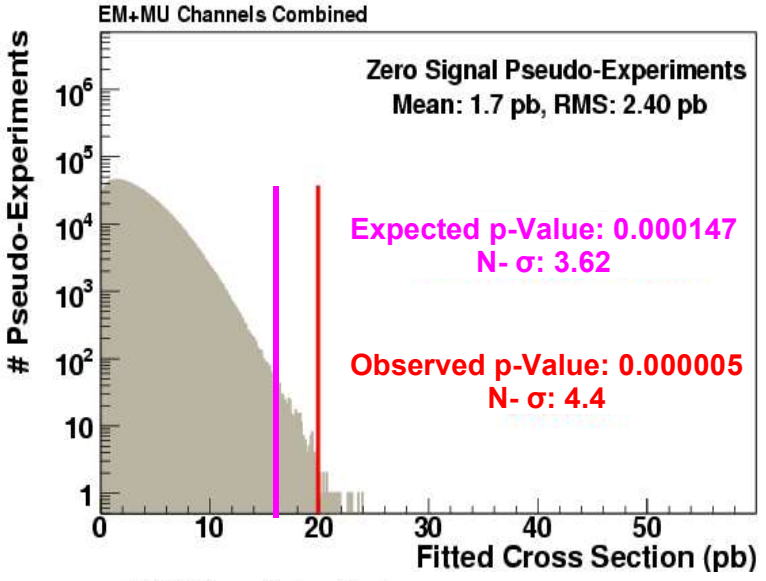
- Estimation of significance
  - Frequentist approach: count fraction of times the background-only scenario would mimic a signal cross section as large as we measured (or expect)
    - Generate pseudo-data for background-only scenario
      - randomly sample the systematics from their priors, then drawing Poisson trials for each bin
    - Fit the background-only pseudo-data outcomes in the same was as real data

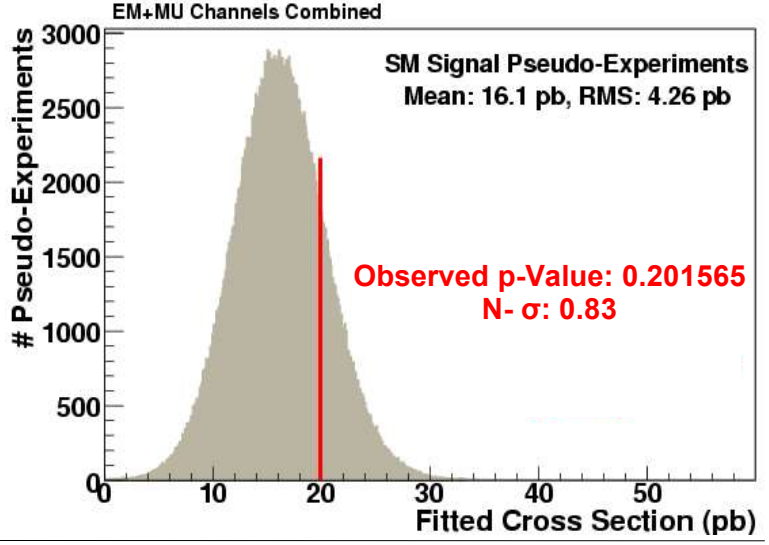
⇒ Expected significance: 3.6σ (p-value = $1.5 \cdot 10^{-4}$)

⇒ Observed significance: 4.4σ (p-value = $5.4 \cdot 10^{-6}$)

- Compatibility with the SM
  - Same approach, but with pseudo-data for signal+background scenario

⇒ 0.83σ above SM prediction



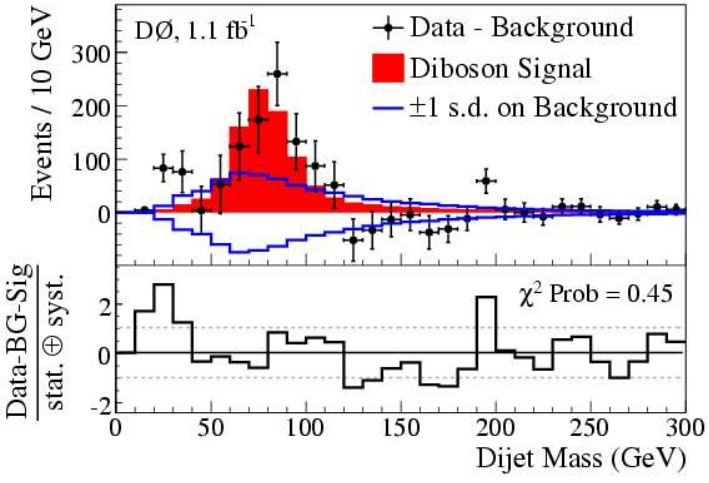


Summary

- First Evidence of $WW/WZ \to l\nu q\bar{q}$ at a hadron collider
- Benchmark for similar Higgs analyses
- Results:
  - $\sigma(WW+WZ) = 20.2 \pm 4.9$ pb
  - Observed significance of 4.4σ
  - Consistent with the SM prediction at 0.83σ

thank you

Background subtracted dijet mass distributions after best-fit of RF

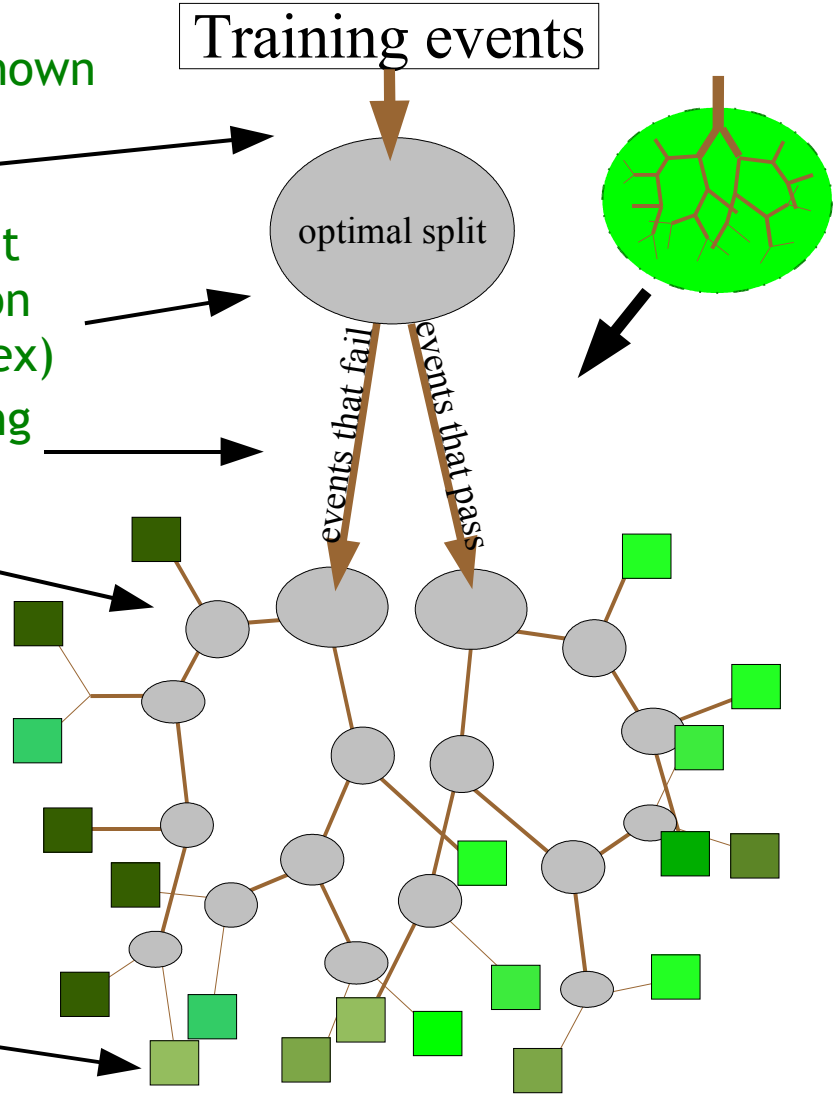


arXiv:0810.3873 [hep-ex], FERMILAB-PUB-08-457-E, Submitted to PRL 2008/10/21

backup slides

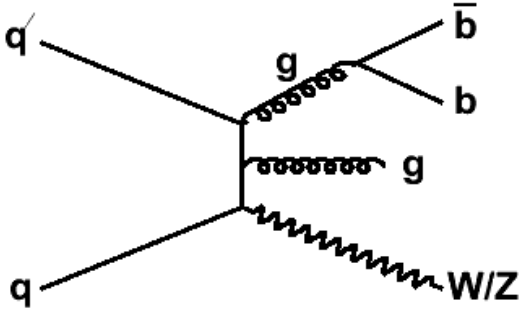
Decision Tree Classifier

- Decision tree is trained/grown using a set of known signal and background training events
  ⇒ These events go into the root node
- Algorithm looks at all possible splits on all input variables and applies split giving best separation between signal and background (using Gini index)
- Events pass to one of two child nodes depending on whether they pass of fail split
- This process is repeated for each child node until
  - a node contains all or no signal events
  - the number of events per node is less than a some specified amount (must be optimized for each analysis)
- Output for an unknown event is determined by the signal purity of the terminal node that the event ends up in



MC Corrections

- Selected sample is dominated by W+jets background
  - *W*+jets is ~30 times larger than the signal!
  - Need to accurately model *W*+jets at the percent level



- Potential problems with modeling of *W*+jets
  - Angles
    - Jet η and ΔR between jets is not modeled well by Alpgen
    - To correct these differences we re-weight the MC to agree with data
  - Energies
    - Alpgen does a fairly good job of modeling energies, however, energy distributions can be affected by Alpgen generator and matching parameters
    - Studied the effects of these parameters to
      - Determine the uncertainty from these parameters
      - Determine best values for agreement with data
  - Cross section
    - The *W*+jets cross section is floated when measuring diboson cross section