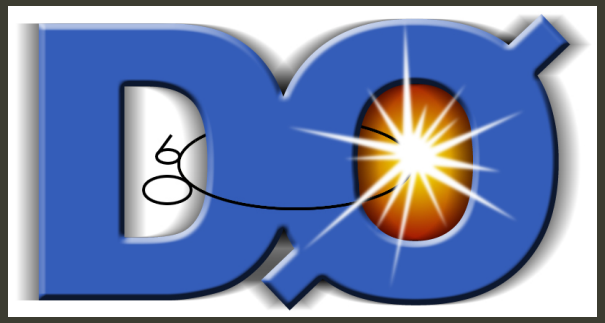


OPTIMIZATION OF MULTIVARIATE DISCRIMINATORS IN THE WH→LVBB CHANNEL AT DØ

5 August 2013

Stephanie Hamilton
Michigan State University

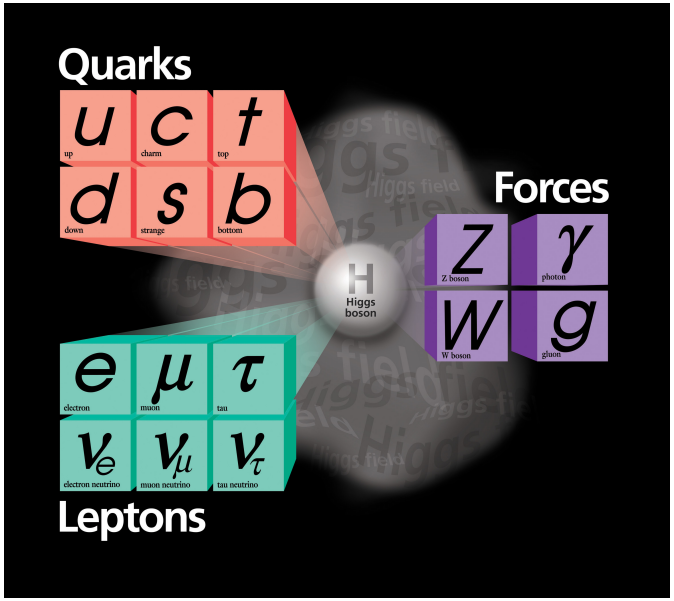
Introduction

The Standard Model (SM)

The SM Higgs Boson

The Standard Model (SM)

- Current theory of known fundamental particles and their interactions via the exchange of gauge bosons
- Extremely successful!
  - Predicted the existence of the top quark, W and Z bosons

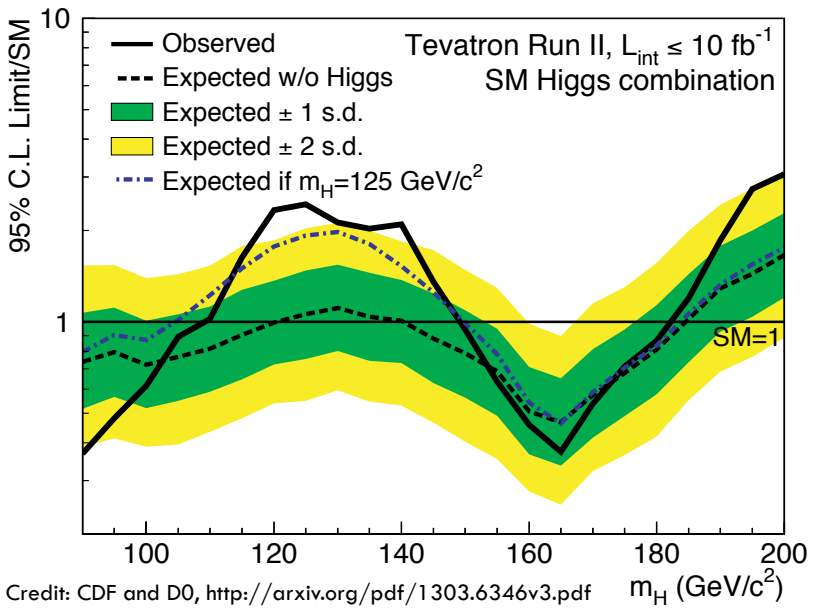


Why do we need a Higgs boson?

- A Higgs mechanism is an essential part of the SM
  - Gives mass to most particles – without it, the SM would not describe life as we know it
  - Provides explanation for electroweak symmetry breaking in the early universe
- A victory for the Standard Model!
  - A Higgs boson was discovered by ATLAS and CMS at CERN in July 2012
  - Simultaneously saw evidence for a new particle in the WH→lν *bb* channel at the Tevatron

The 95% Confidence Level Limit

- WH→l ν $b\bar{b}$ is one of six analyses combined for this plot
  - Want to improve sensitivity because the Higgs boson has not been established in this channel yet
- Expected production cross-section over predicted SM cross-section => a measure of how many more events we need to exclude or confirm the particle
  - A measure of our sensitivity
    - Greater than 1 => cannot give a definite answer
    - Less than 1 => can definitively say whether or not the particle is there



Credit: CDF and D0, http://arxiv.org/pdf/1303.6346v3.pdf

How do we search for a Higgs?

The SM Higgs Boson at the Tevatron

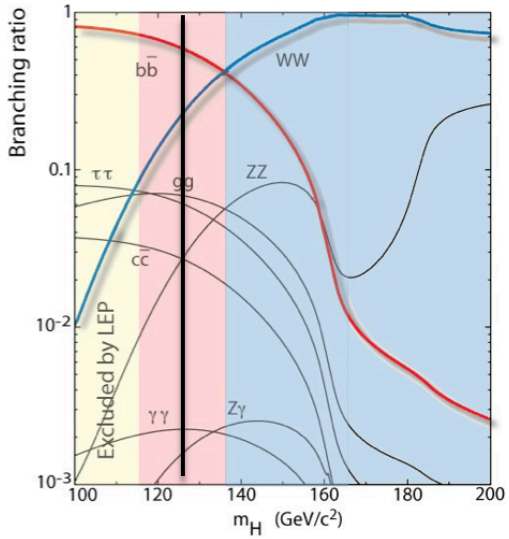
The DØ Detector

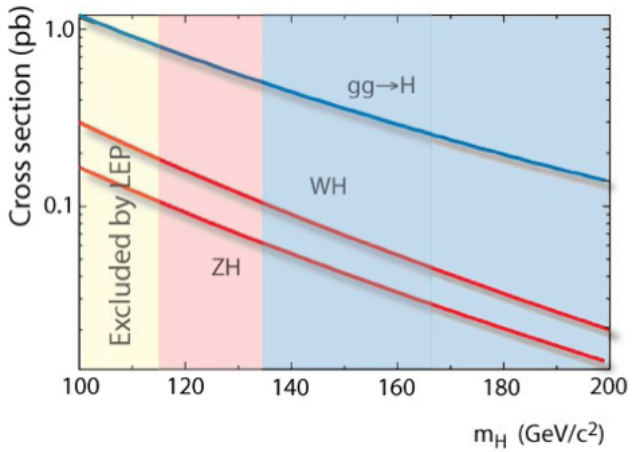
The WH→l ν $b\bar{b}$ Channel

TMVA and Multivariate Analysis

The SM Higgs Boson at the Tevatron

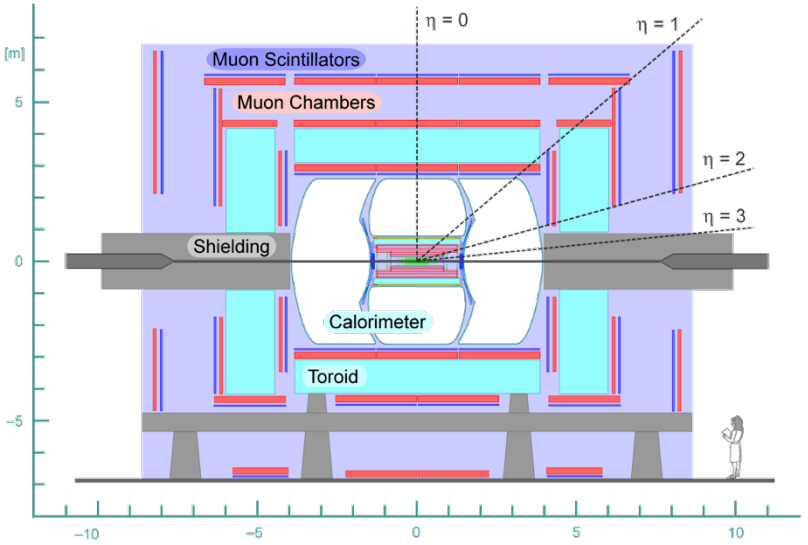
- Direct search at $\sqrt{s} = 1.96$ TeV
- Two primary means of production
  - Gluon fusion
  - Associated production
- Decay branching ratios depend on the mass

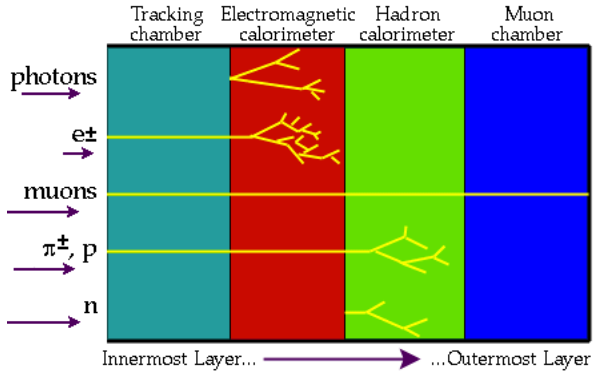




The DØ Detector

- Multiple subdetectors
  - Tracking system
    - Silicon Microstrip Tracker
    - Central Fiber Tracker
  - Calorimeter
  - Muon system
- Neutrinos identified as missing transverse energy

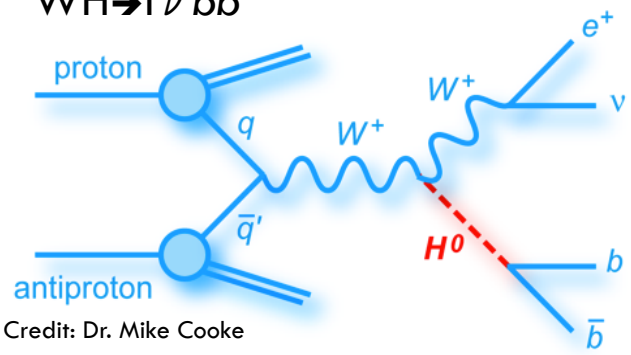




The WH→l ν $b\bar{b}$ Channel

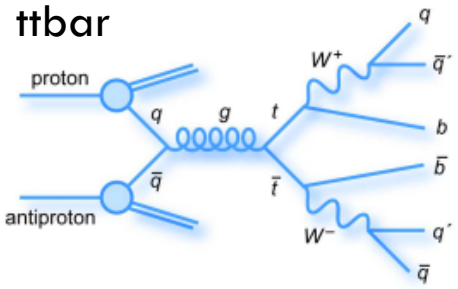
- Tiny Higgs signal against huge backgrounds
- Reducing the huge background
  - b-tagging, Multivariate techniques

WH→l ν $b\bar{b}$

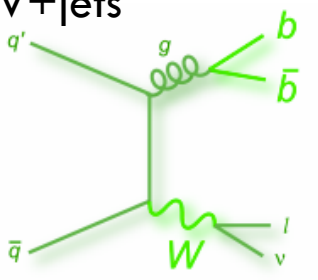


Credit: Dr. Mike Cooke

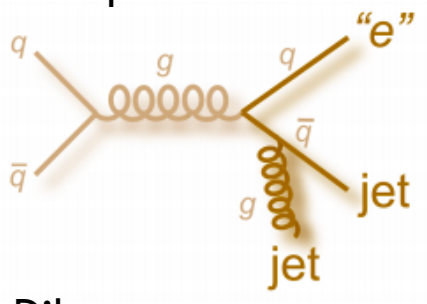
ttbar



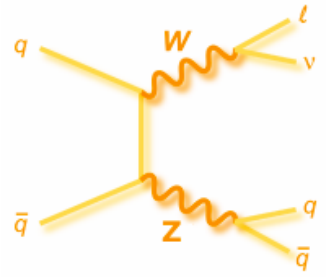
V+jets



Multijet

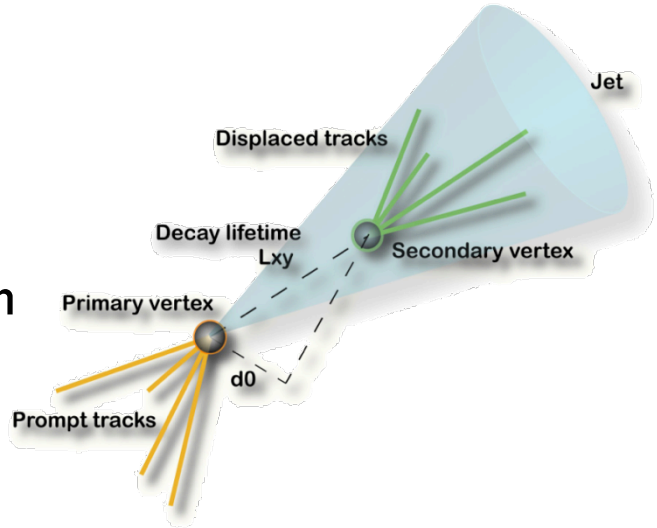


Diboson



What is b-tagging?

- First, what is a jet?
  - Attempting to separate a pair of quarks - takes less energy to create a spray of new particles
  - Charged particles leave tracks in the tracker and the spray leaves a wide deposit of energy in the calorimeter
- Identifying bottom quark jets
  - Look for:
    - A secondary vertex displaced from the primary vertex
    - Displaced impact parameter



Multivariate Techniques

TMVA and Multivariate Analysis

TMVA Method Options

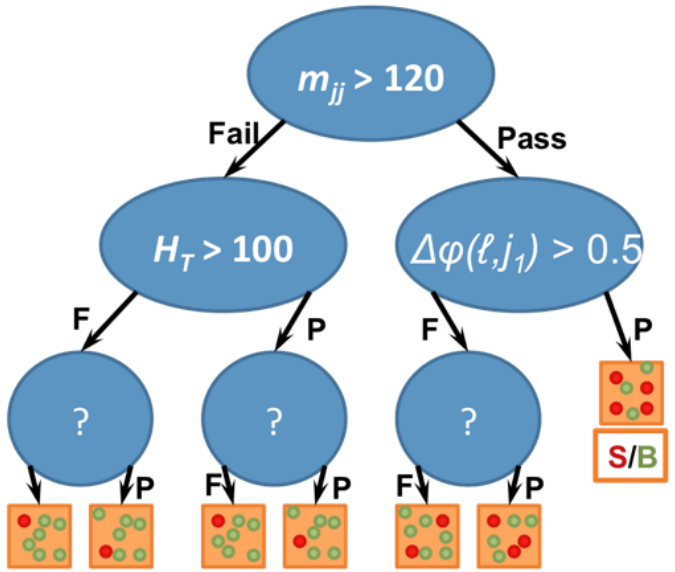
TMVA Output

TMVA and Multivariate Analysis

- Toolkit for Multivariate Analysis (TMVA)
  - A library of ROOT, the statistical analysis framework used by most of the high energy physics community to analyze data
- Multivariate Analysis (MVA)
  - Combining several moderately discriminating variables into one strongly discriminating variable
    - Discriminating => background distribution of the variable tends toward left of histogram, while signal tends toward right
  - Secondary MVAs
    - Higgs vs. specific background (ttbar, V+jets, diboson, multijet)
  - Final MVA
    - Higgs vs. all background

Multivariate Techniques

- Decision Trees (DT)
  - Subsequent cuts are made on different input variables until a stop criterion is reached
  - Each leaf has a specific signal-to-background ratio
- Boosted Decision Trees (BDT)
  - A "forest" of many DTs
  - The signal-to-background ratios are used as weights for misclassified events to train the next trees



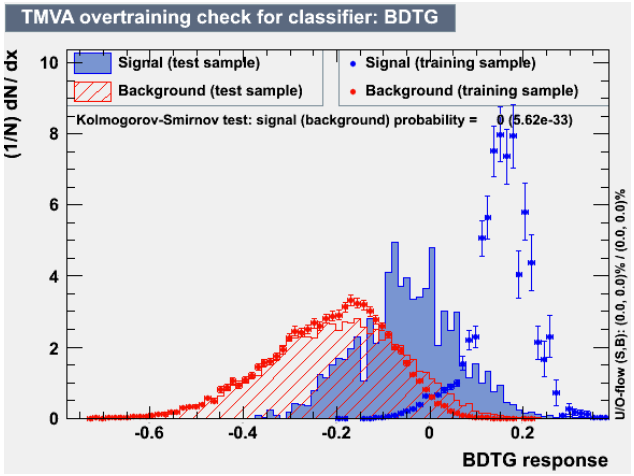
Credit: Dr. Mike Cooke

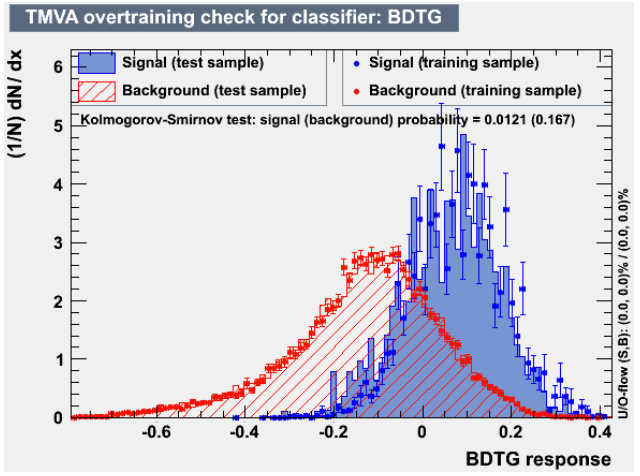
TMVA Method Options

- Possible to vary
  - `BoostType` – defines how TMVA uses the signal-to-background ratios as weights for the next trees
  - `NTrees` – number of trees in the random forest
  - `Shrinkage` – defines the learning rate of the boosting algorithm
  - `NNodesMax` – maximum number of nodes any tree is allowed to have
  - `MaxDepth` – how many "levels" a tree is allowed to have
  - `GradBaggingFraction` – defines the fraction of events that will be used in each iteration of growing a tree, when one is using random subsamples of all events.
  - And many more…

TMVA Output

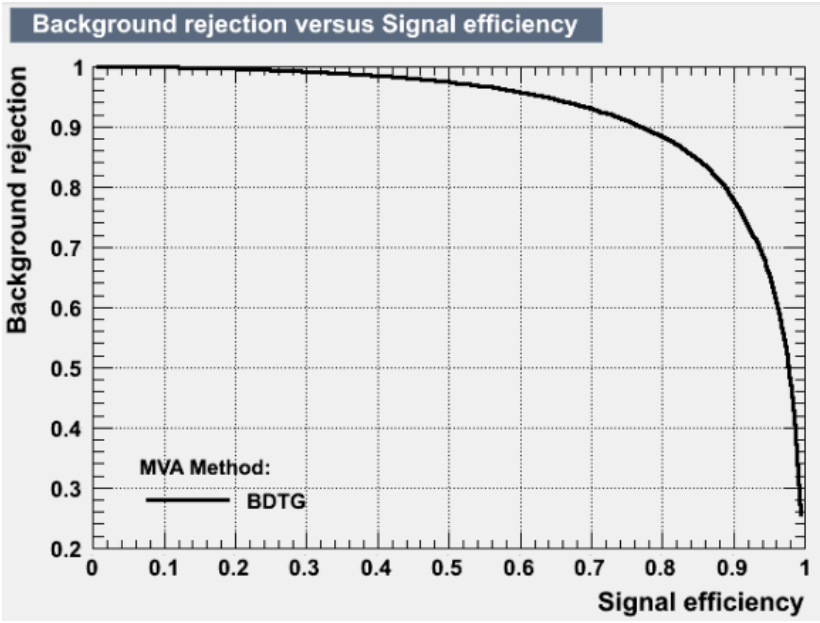
- Overtraining
  - TMVA begins to cut on statistical fluctuations rather than on the physics properties of the data
  - Compare "train" and "test" subsamples to determine the probability that they originated from same sample
    - KS test – considered passed if both background and signal results were above 1%





TMVA Output (cont'd)

- Background Rejection vs. Signal Acceptance Curve
  - How much signal is being kept after a certain amount of background is rejected?



Summer Work

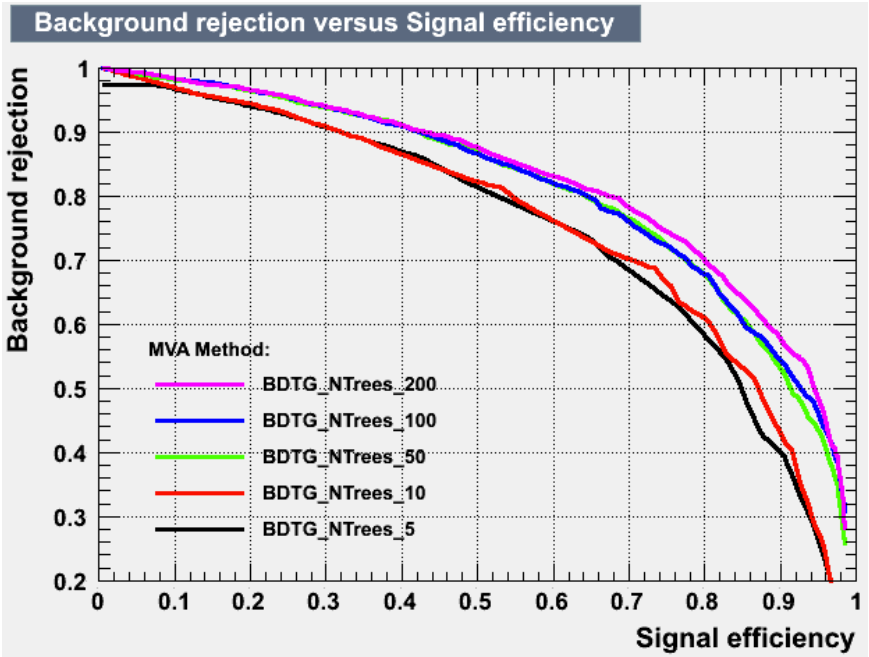
Optimization of Multivariate Discriminators

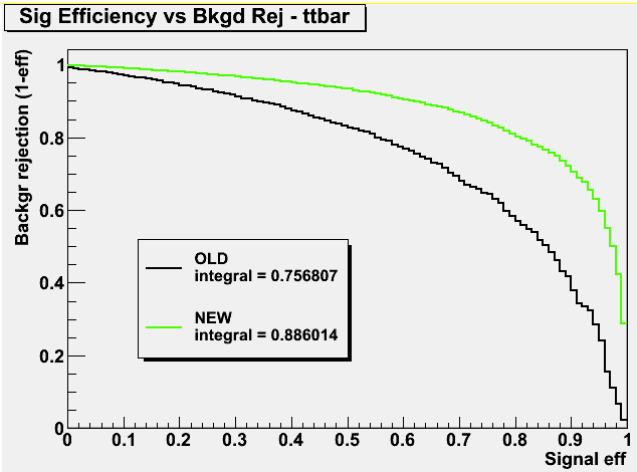
Results

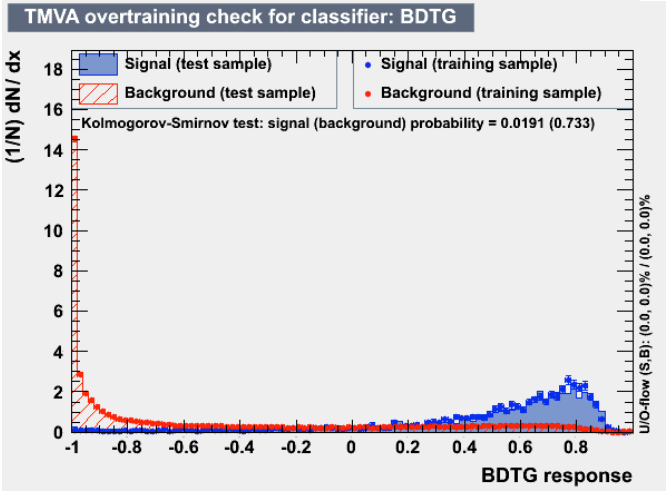
Optimization of Multivariate Discriminators

- When run, the optimization process would vary
  - `NTrees`
  - `Shrinkage`
  - `NNodesMax`
  - `GradBaggingFraction`
- Signal Acceptance vs. Background Rejection curve integral and overtraining plots used to determine which combination was the best

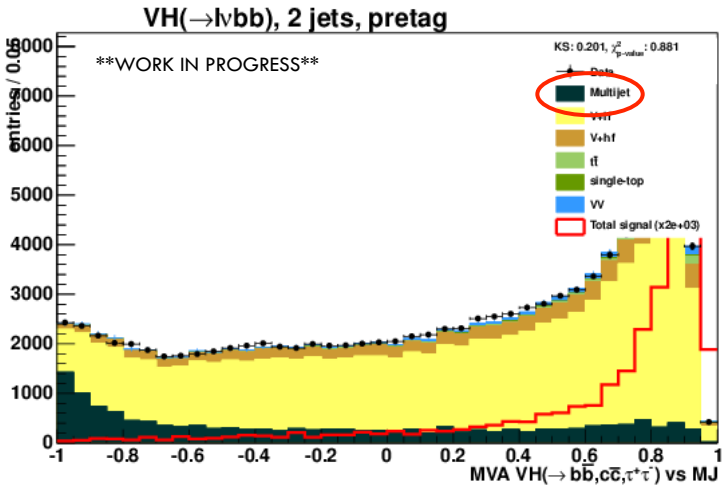
Improvements in MVAs

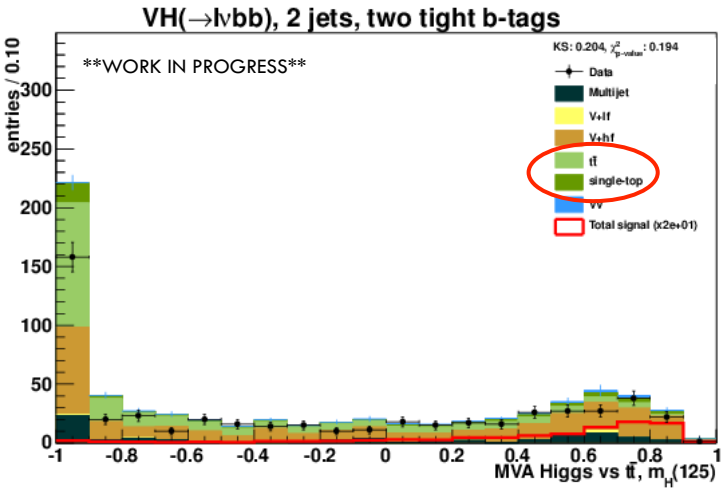


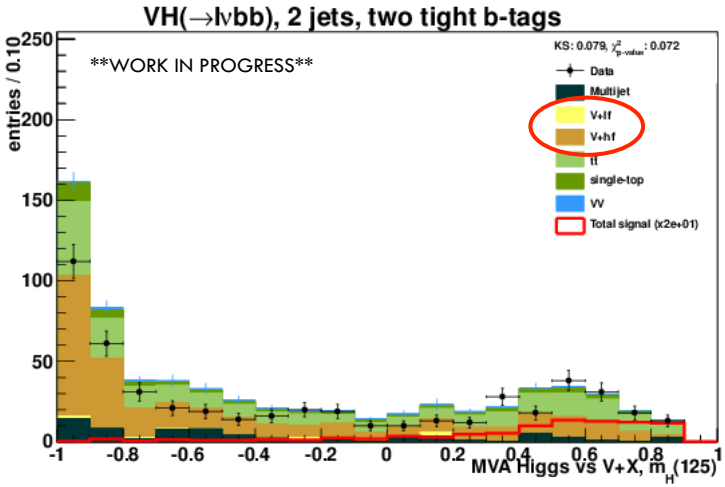


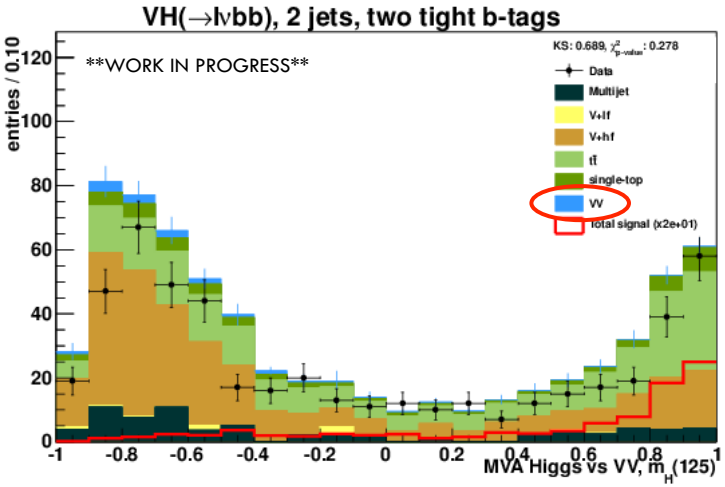


Results

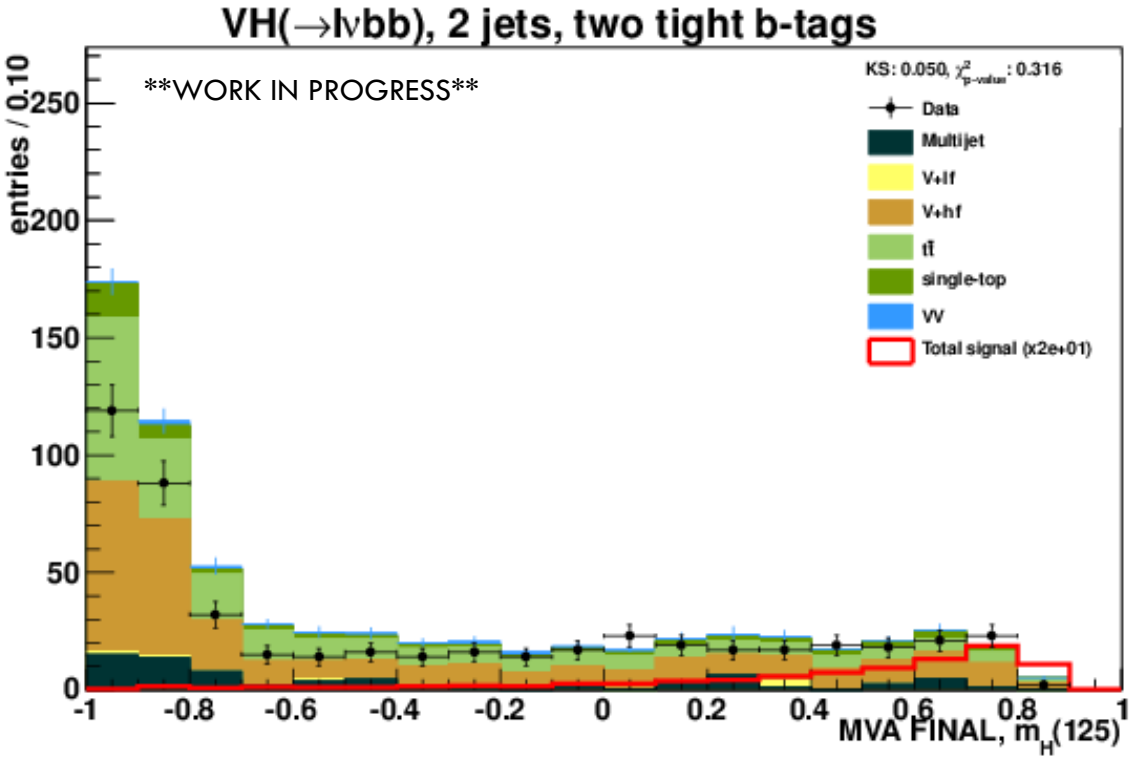

VH(→lνbb), 2 jets, pretag
WORK IN PROGRESS
KS: 0.201, $\chi^2_{p\text{-value}}$: 0.881
Data
Multijet
V+lf
V+hf
tt̄
single-top
VV
Total signal (x2e+03)
entries / 0.05
MVA VH(→ bb̄,cc̄,τ⁺τ⁻) vs MJ


VH(→lνbb), 2 jets, two tight b-tags
WORK IN PROGRESS
KS: 0.204, $\chi^2_{p\text{-value}}$: 0.194
Data
Multijet
V+lf
V+hf
tt̄
single-top
VV
Total signal (x2e+01)
entries / 0.10
MVA Higgs vs tt̄, m_H(125)


VH(→lνbb), 2 jets, two tight b-tags
WORK IN PROGRESS
KS: 0.079, $\chi^2_{p\text{-value}}$: 0.072
Data
Multijet
V+lf
V+hf
tt̄
single-top
VV
Total signal (x2e+01)
entries / 0.10
MVA Higgs vs V+X, m_H(125)


VH(→lνbb), 2 jets, two tight b-tags
WORK IN PROGRESS
KS: 0.689, $\chi^2_{p\text{-value}}$: 0.278
Data
Multijet
V+lf
V+hf
tt̄
single-top
VV
Total signal (x2e+01)
entries / 0.10
MVA Higgs vs VV, m_H(125)

Results (cont'd)



Results (cont'd)

- Significant improvements in our expected sensitivity to the SM Higgs boson cross-section

95% C.L. Limits on the Higgs Boson Production Cross-Section

	Before Summer 2013	After Summer 2013	Percent Increase
MVA el	6.28	5.70	9.24%
MVA mu	6.52	5.88	9.51%
MVA el+mu	4.42	4.02	9.05%

Summary

- New optimization tools for Multivariate Analysis were developed
  - Varies the values of different options used for training BDTs
- These tools played an important part in the over-9% increases from the pre-Summer 2013 starting point

Thanks

- Dr. Michael Cooke
- Dr. Ryuji Yamada
- My fellow summer students and the rest of the *WH* group
- The SIST Committee
  - Linda Diepholz
  - Dianne Engram
  - Dr. Davenport
- The DØ Collaboration
- Fermi National Accelerator Laboratory