Search for the Standard Model Higgs Boson in Missing Energy and b-jet Final States at the Tevatron

Tim Scanlon

On behalf of the CDF and DØ Collaborations

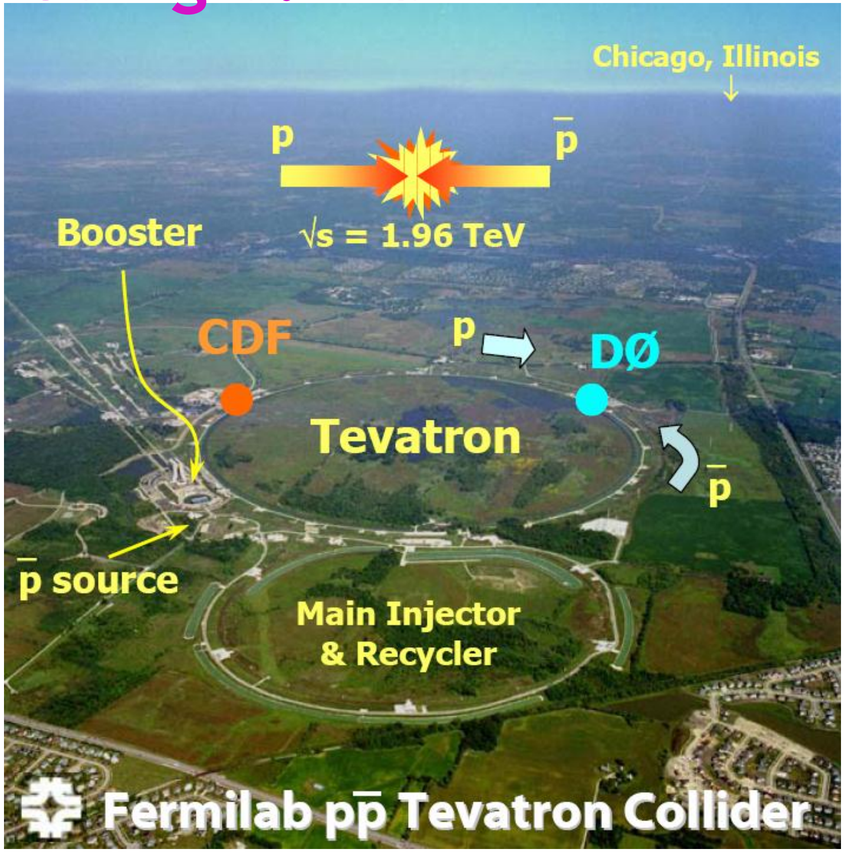
Outline

- Search overview
  - ➢ Signal signature
  - ➢ Background
- Analysis Technique
  - ➢ Event selection
  - ➢ b-jet identification
  - ➢ Multivariate selection
- Limits
- Conclusions

Delivered > 11.5 fb^{-1}

Recorded > 10 fb^{-1}

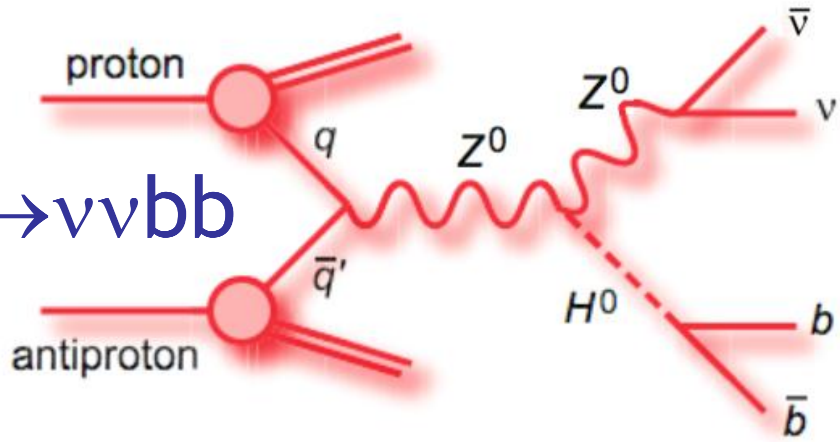
Using 8.4 fb^{-1}



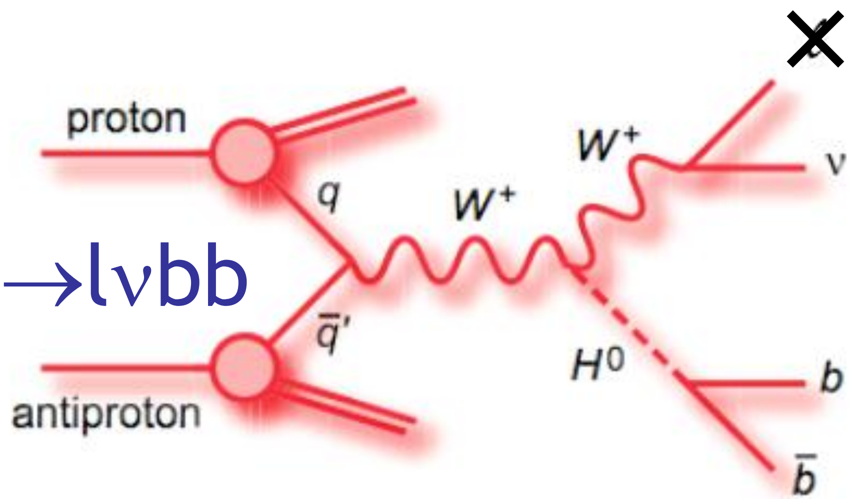
[Thanks to all my Tevatron colleagues]

Signal Signature

$ZH \to \nu\nu bb$

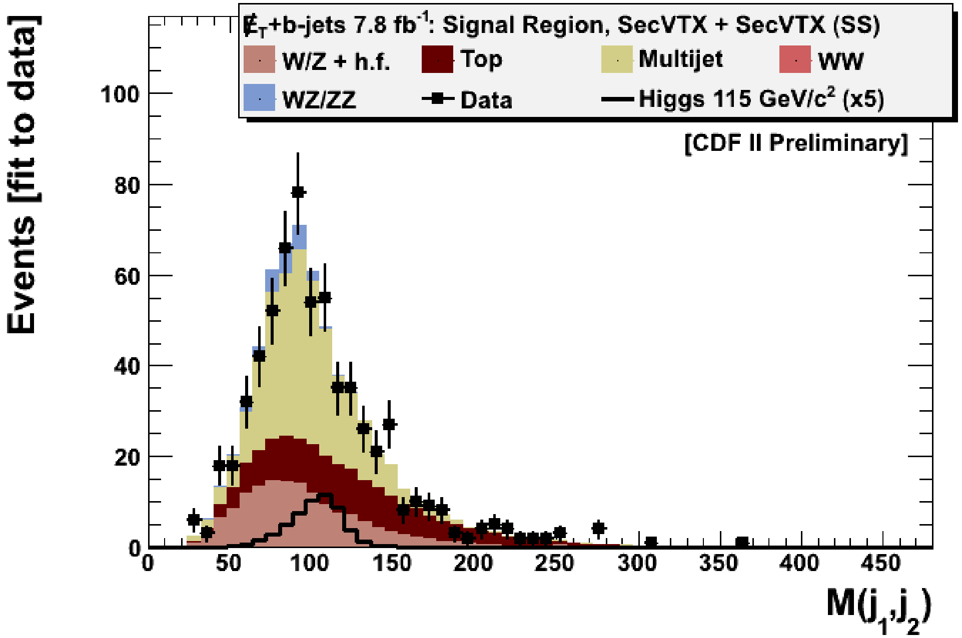


$WH \to l\nu bb$

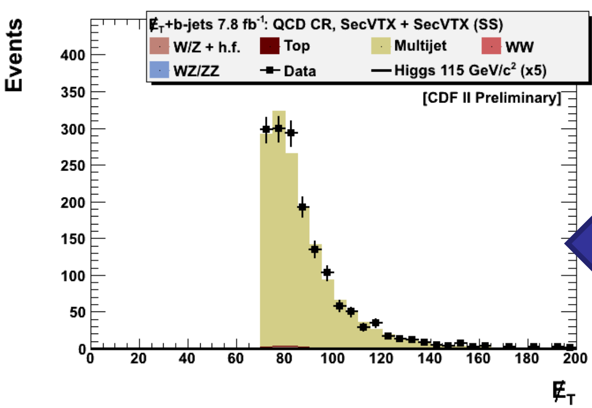


- Searching for:
  - Large missing energy
  - Two b-jets
  - No reconstructed isolated leptons
- Two dominant channels
  - $Z \to \nu\nu$
  - $W \to l\nu$ with missing lepton
- Two searches
  - D0 8.4 fb^{-1}
  - CDF 7.8 fb^{-1}

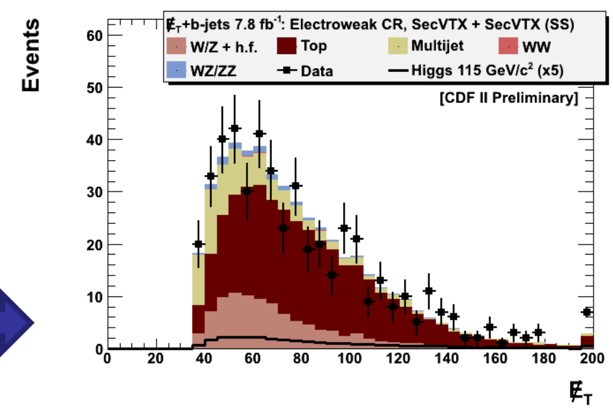
Backgrounds



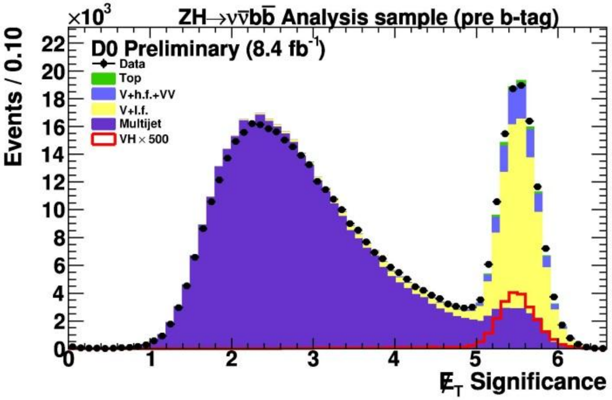
- Instrumental
  - Multijet (MJ)
  - Large background
  - Difficult to model
    - Estimated from data
- Physics
  - V+jets (V=W/Z)
  - Diboson
  - Top



Define four samples:
Signal sample
Multijet model
Multijet Control
Physics Control

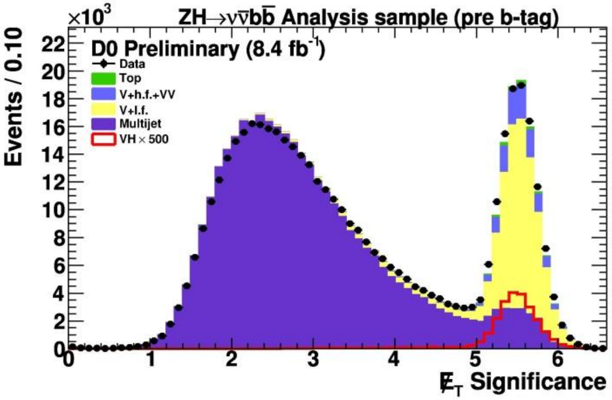


Analysis Strategy



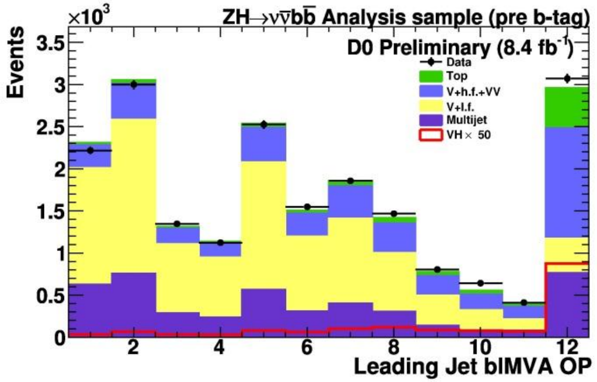
Kinematic event Selection

Analysis Strategy

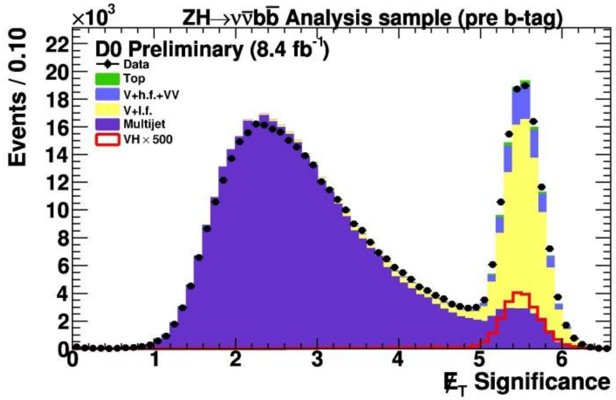


Kinematic event Selection

b-jet Identification



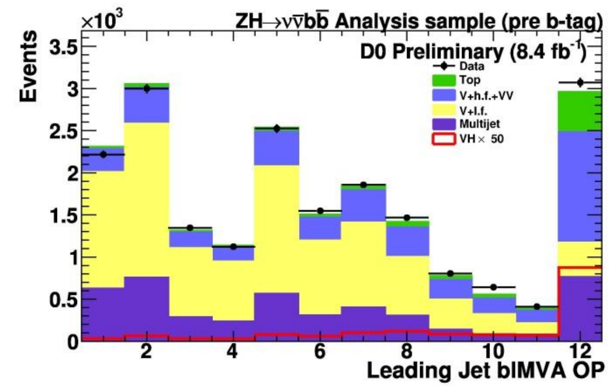
Analysis Strategy

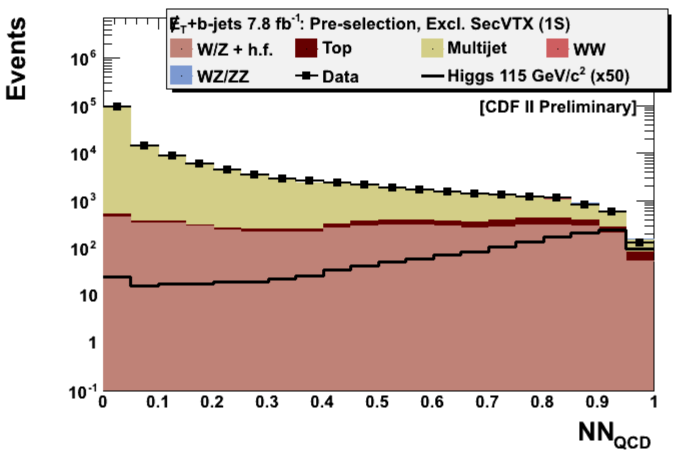


Kinematic event Selection

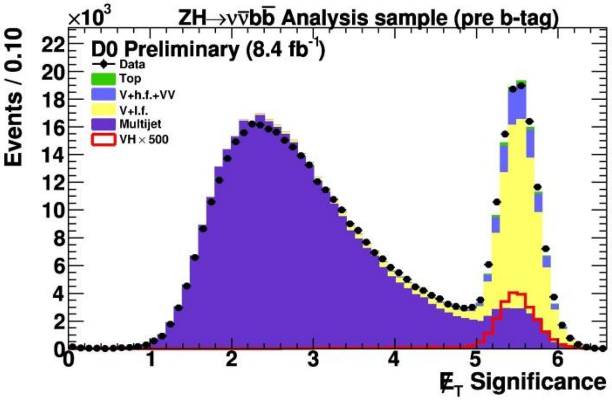
b-jet Identification

Multijet Removal





Analysis Strategy

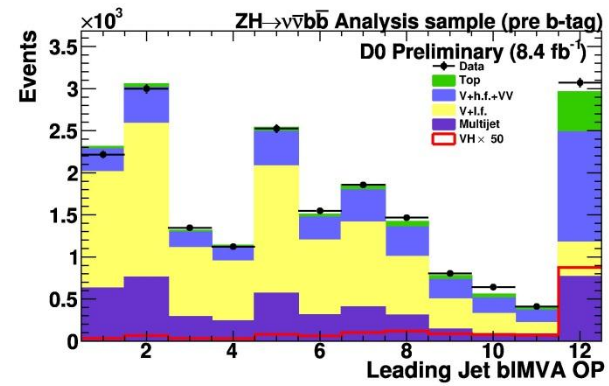


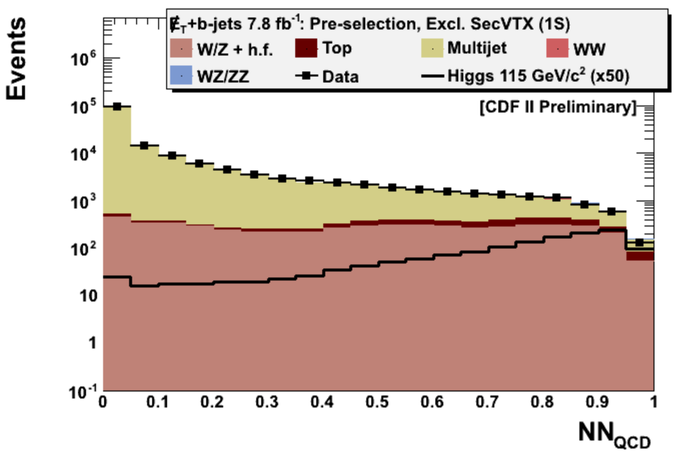
Kinematic event Selection

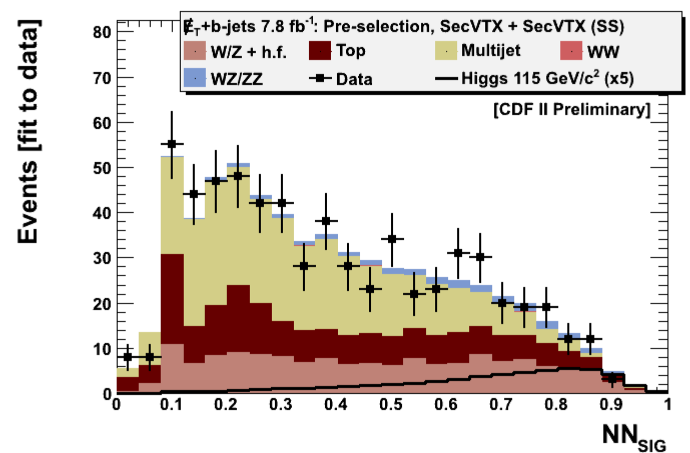
b-jet Identification

Multijet Removal

Final Discriminant





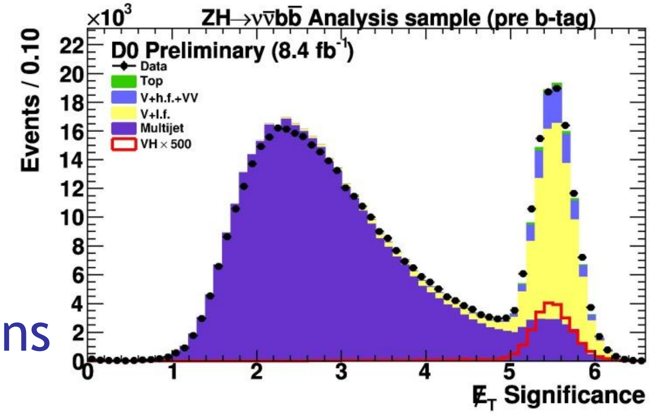


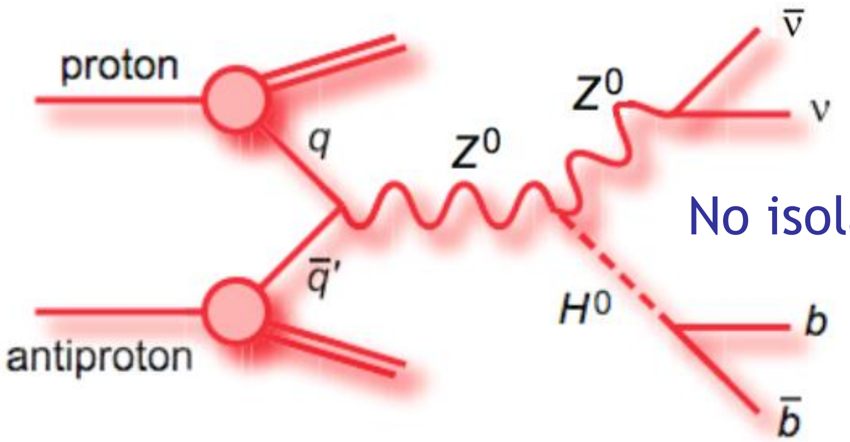
Imperial College London

Signal Sample

Large missing energy (MET)

- CDF: MET > 35 GeV
- D0: MET > 40 GeV and MET Significance > 5





No isolated leptons

Two or more jets

- D0: 2 jets not back to back
- CDF: 2 or 3 jets

CDF relaxed requirements

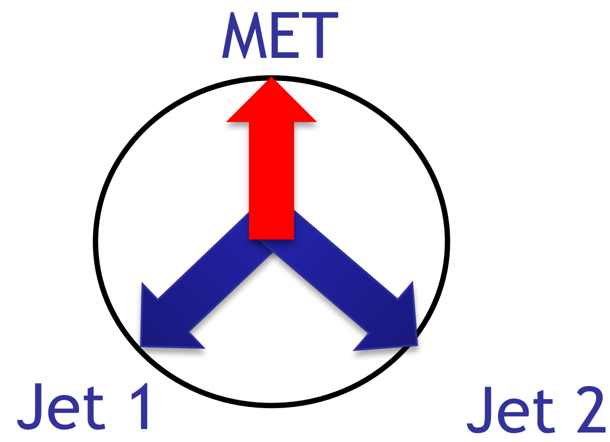
- MET 50→35 GeV
- Lead jet pT 35→25 GeV
- Increase acceptance 30-40%

Imperial College London

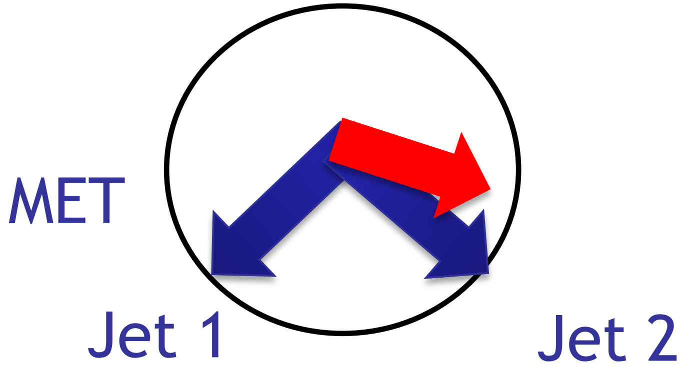
Multijet Model Sample

- Multijet events
  - ➢ MET tends to align with jets
  - ➢ Use as handle on MJ events
- Define two regions
  - ➢ Signal
    - MET and jets well separated in respect of azimuthal angle
  - ➢ MJ model
    - MET and jets closely aligned
    - Negligible signal
- Events in MJ region used to model MJ in signal region
  - ➢ Validated in control region

Signal

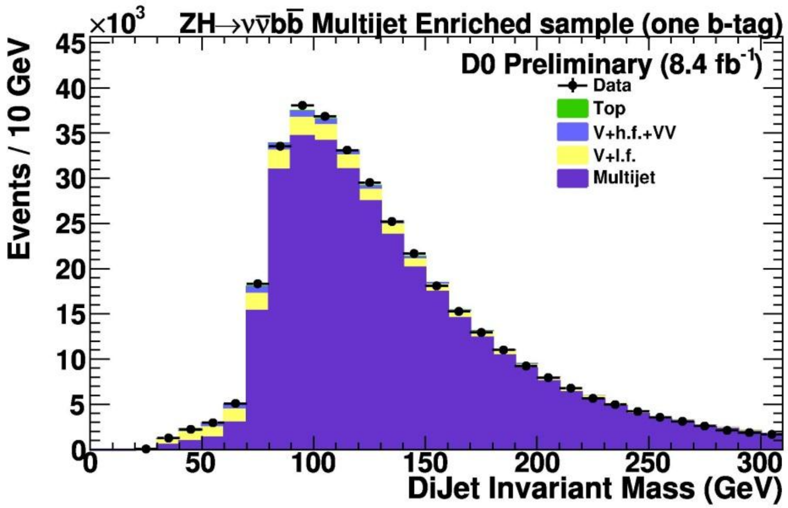


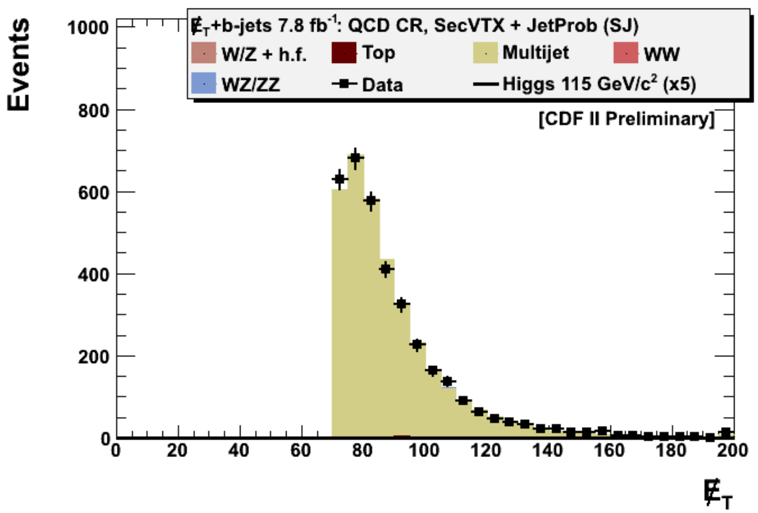
Multijet



Multijet Control Sample

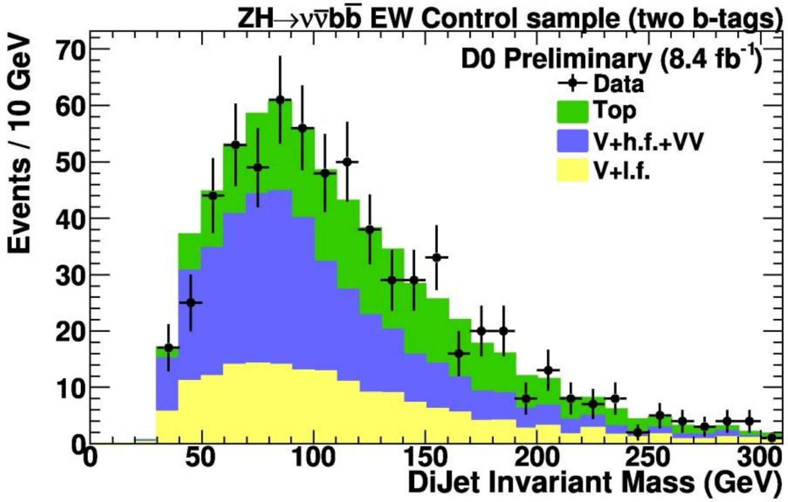
- Sample dominated by MJ
  - Kinematically similar to signal sample
  - D0
    - Relax MET cut
    - No MET significance cut
  - CDF
    - High MET
    - MET aligned with jet
- Validates modelling of MJ background
  - All variables well modelled

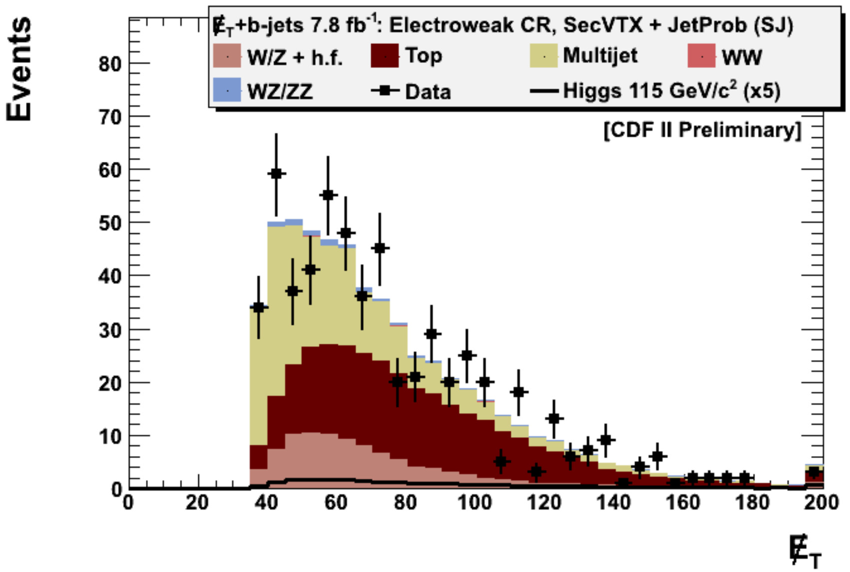




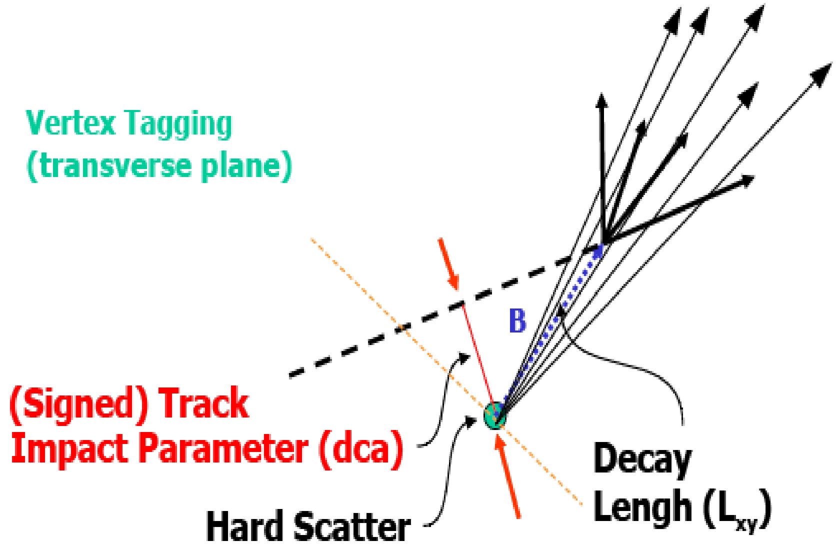
Physics Control Sample

- Form sample enriched in W+jets/top events
- Require:
  - D0: Isolated muon
  - CDF: Isolated electron/muon
  - Keep other cuts the same as signal region
- Validates modelling of EW/top backgrounds
  - All variables well modelled

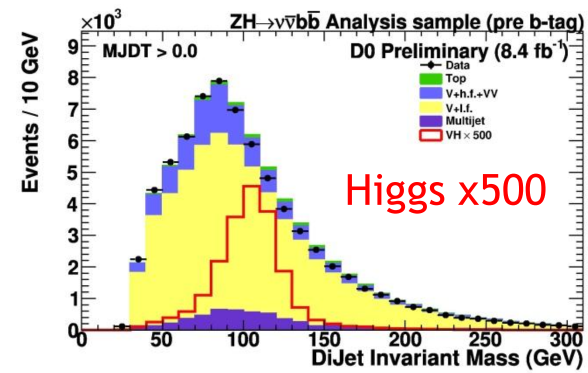




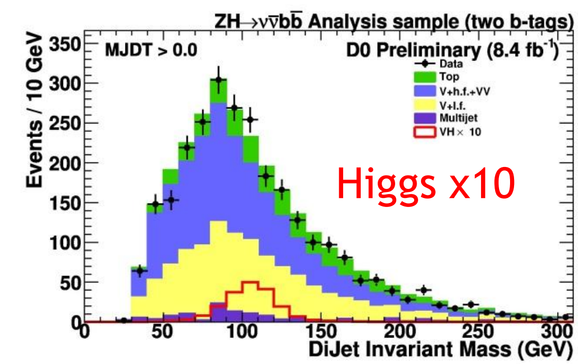
b-Jet Identification



- Tools
  - ➢ Secondary vertex
  - ➢ Jet lifetime probability
  - ➢ Counting impact parameters
  - ➢ Combine all information using multivariate technique

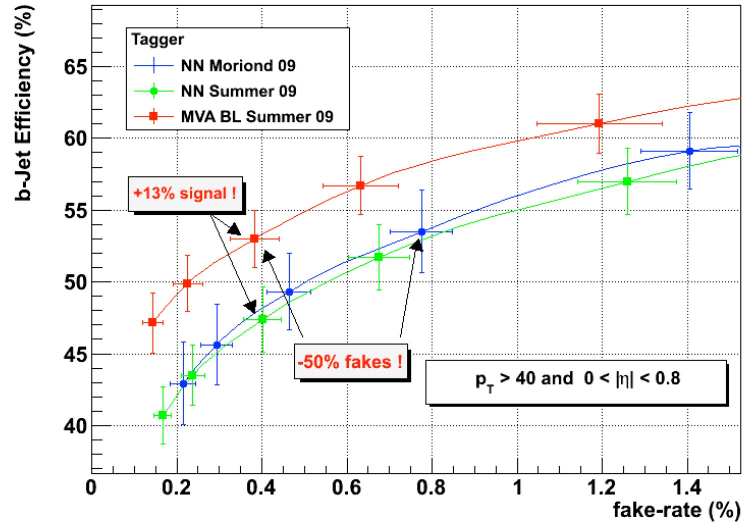


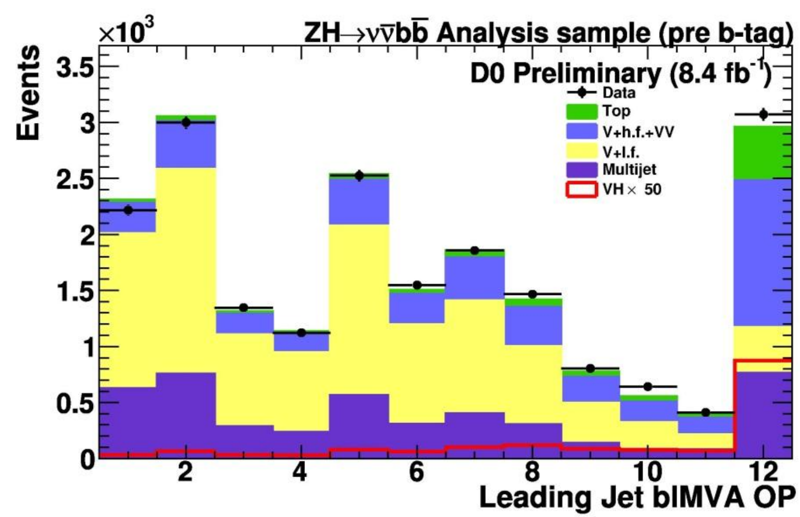
Increase in
S/√B > x10



b-Jet Identification

- CDF – Secondary vertex and lifetime probability
  - ➢ Three exclusive tagging channels
    - One single tag/two double tag
- D0 – Multivariate (MVA) tagging
  - ➢ Very loose single and double tag channels
    - ~80% b-jets, ~10% light-jets
  - ➢ Use MVA output as input to final discriminant
    - ~10% gain in sensitivity





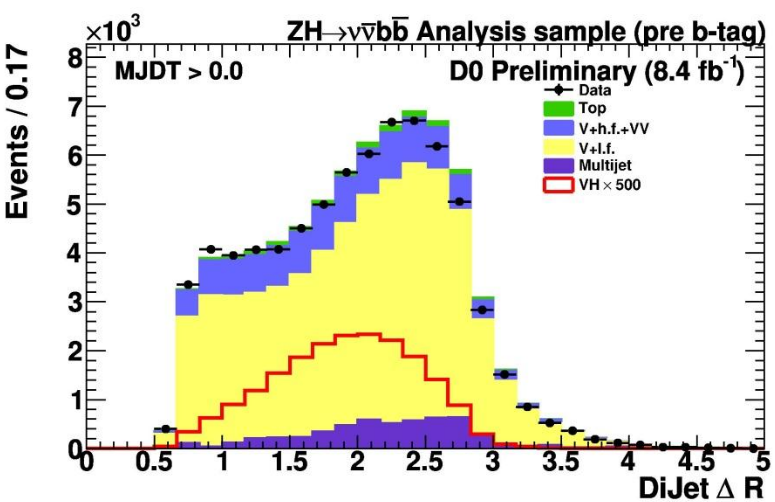
Multijet Removal

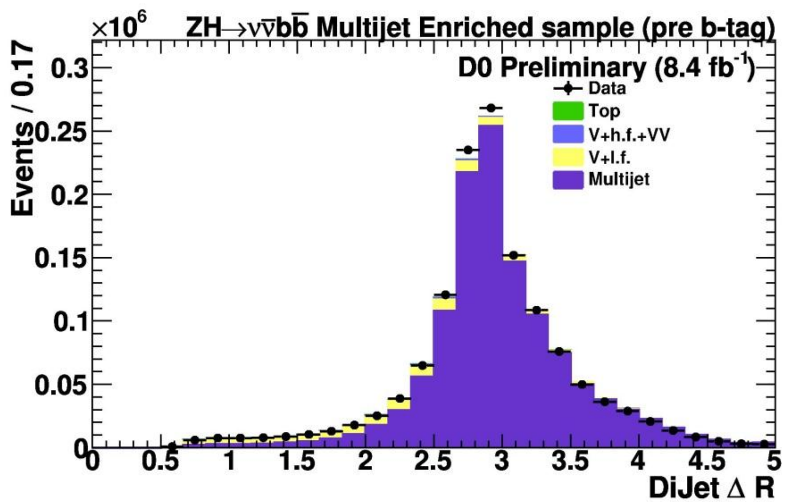
- Train multivariate technique to remove large MJ background
  - D0 Decision Tree (DT)
  - CDF Neural Network (NN)
  - Kinematic variables used as input
    - D0 (30), CDF(14)

Example variables:

- MET from jets, tracks and calorimeter
- Relative size of MET
- Relative position of jets/MET
- Event shape

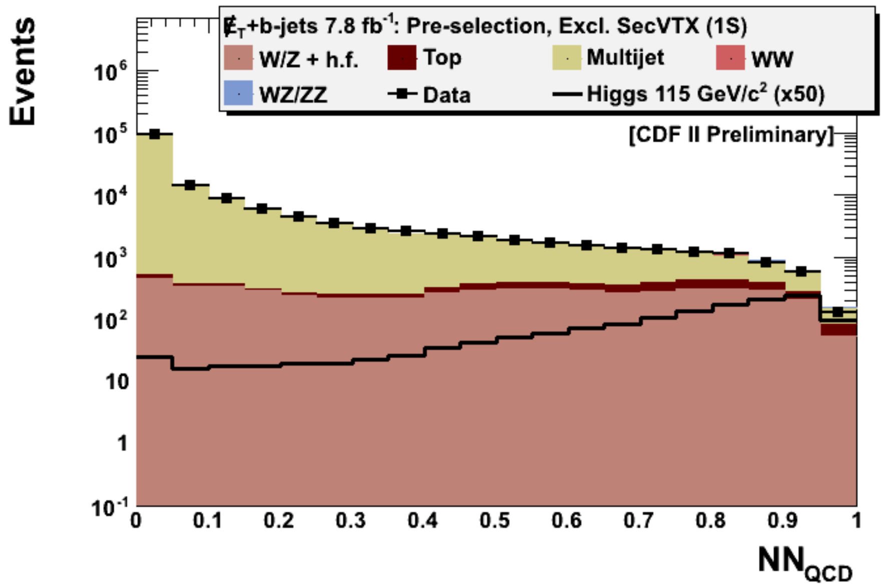
- Trained in pretag samples





Multijet Removal

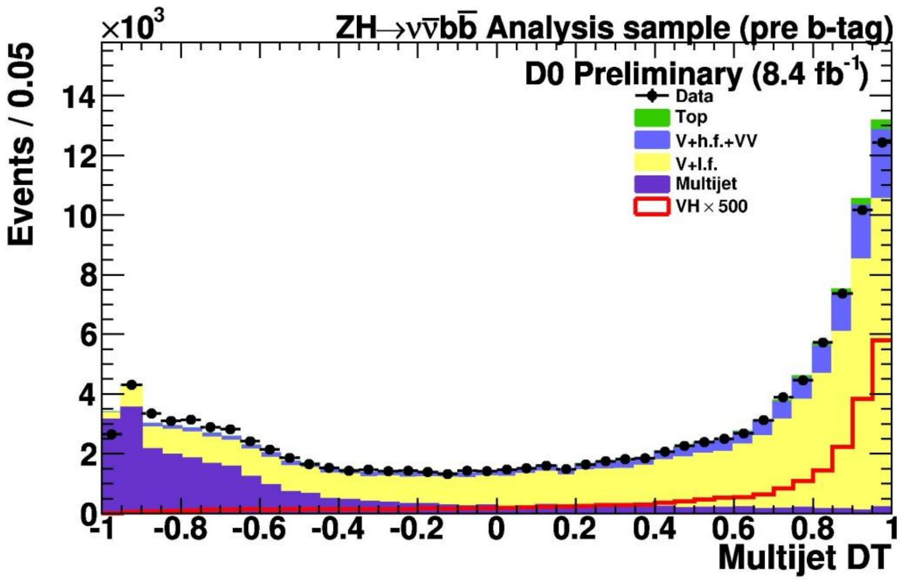
CDF



Signal efficiency ~90%
MJ Rejection ~87%

→ MJ 44% of double tag signal sample

D0

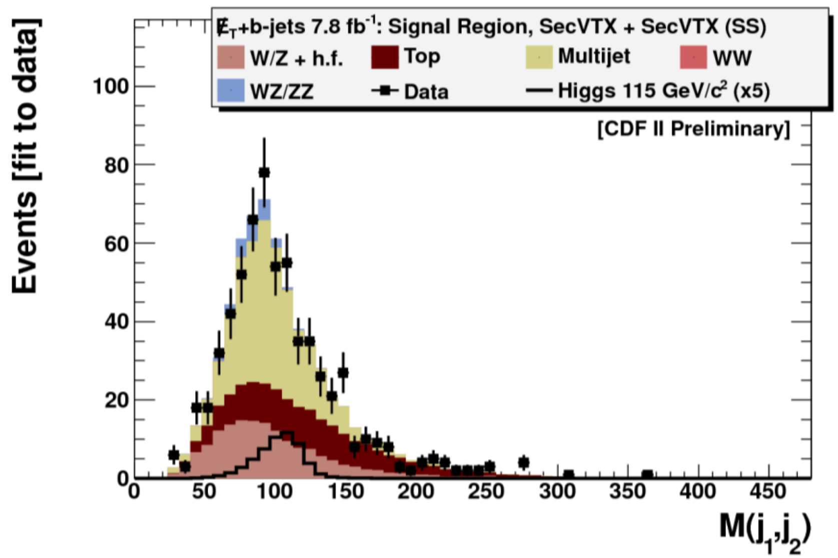


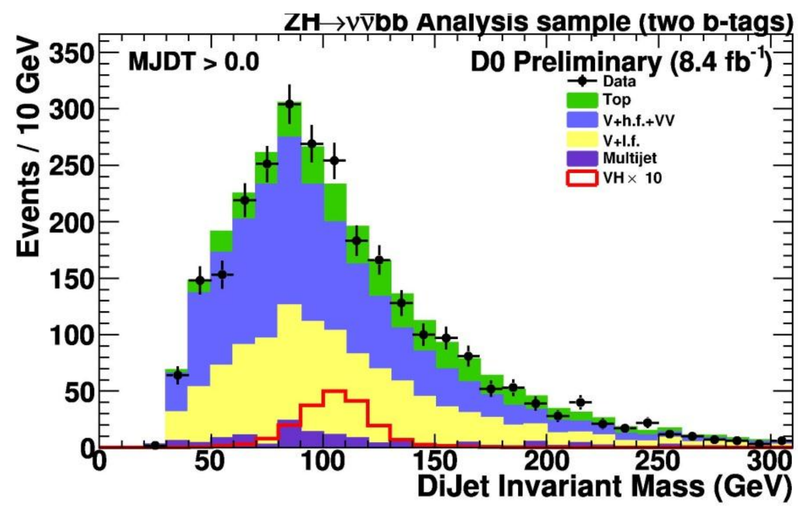
Signal efficiency ~88%
MJ Rejection ~80%

→ MJ 4.5% of double tag signal sample

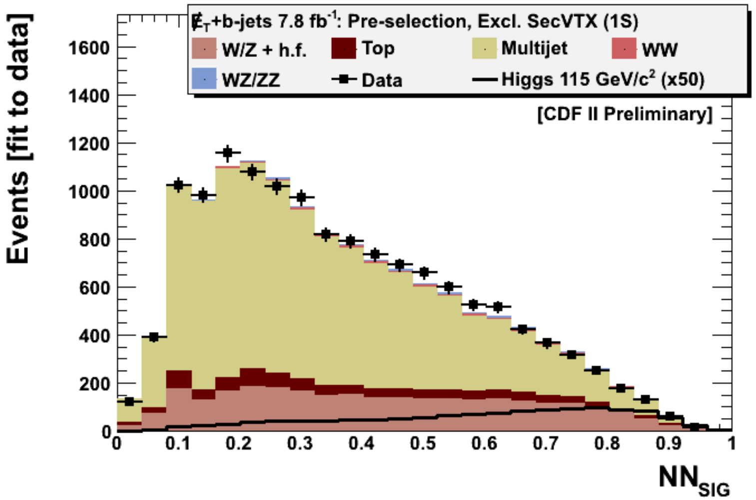
Final Discriminant

- Train DT (D0) or NN (CDF) to separate signal from signal-like backgrounds
- Trained separately in
  - Each b-tagging/jet channel
- Variables
  - D0
    - Inputs to MJ DT plus b-tagging and mass related variables
  - CDF
    - Output of MJ NN, mass, direction of jets and MET related variables

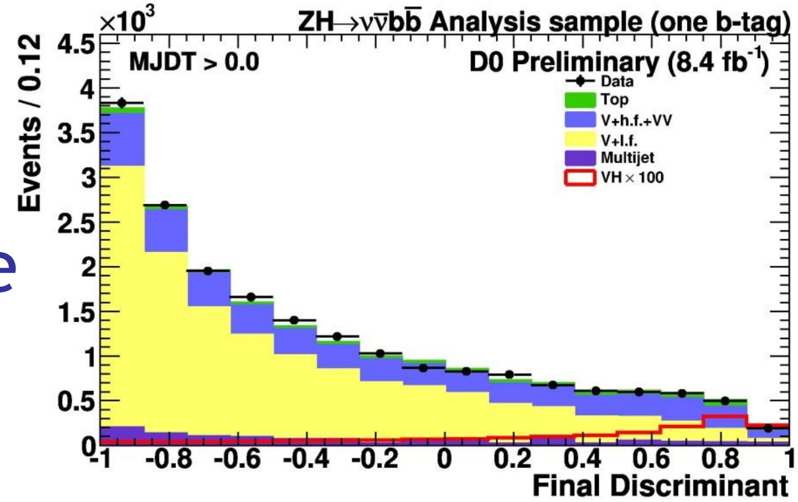


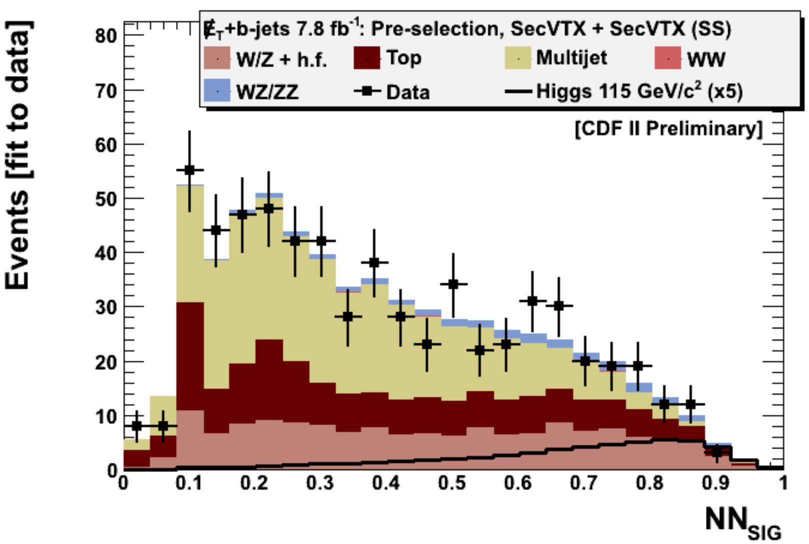


Final Discriminant

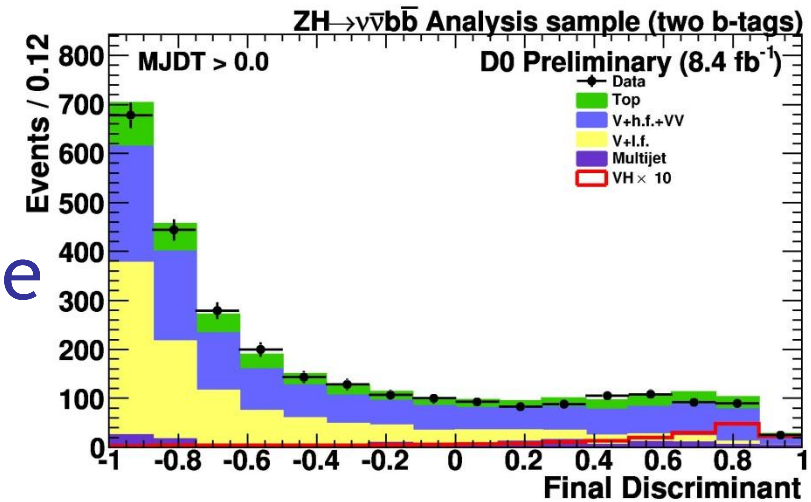


Single Tag



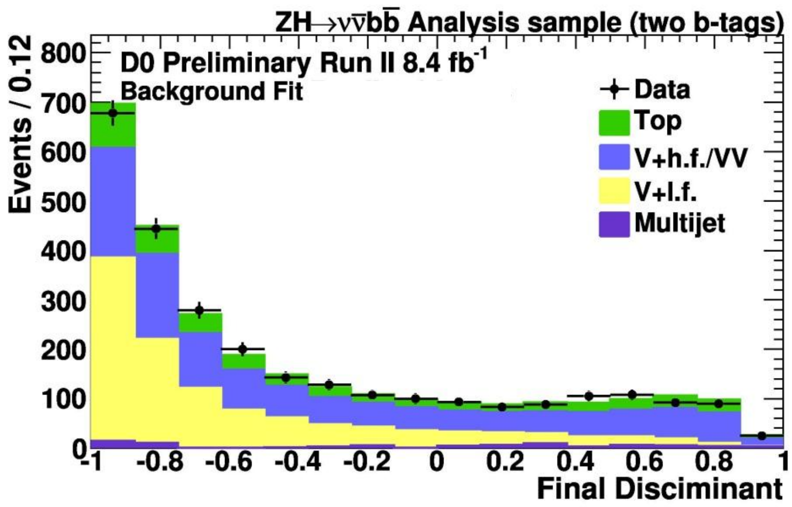


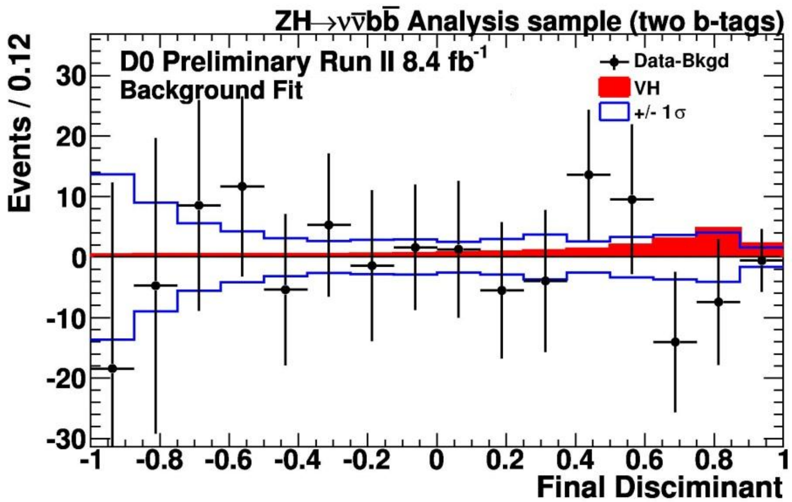
Double Tag



Systematic Uncertainties

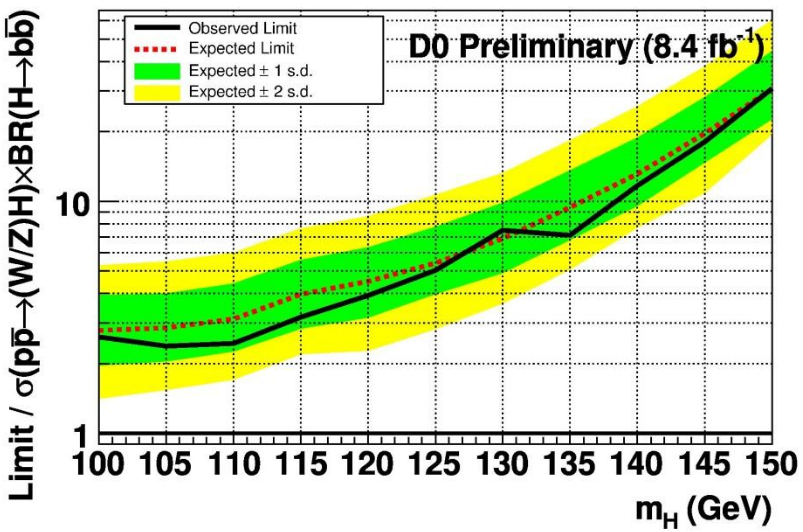
- Uncertainties included on
  - Shape
  - Normalisation
- Dominant uncertainties
  - Jet energy
  - b-tagging
  - Background cross section
  - Luminosity
- Impact reduced by constraining uncertainties during limit setting

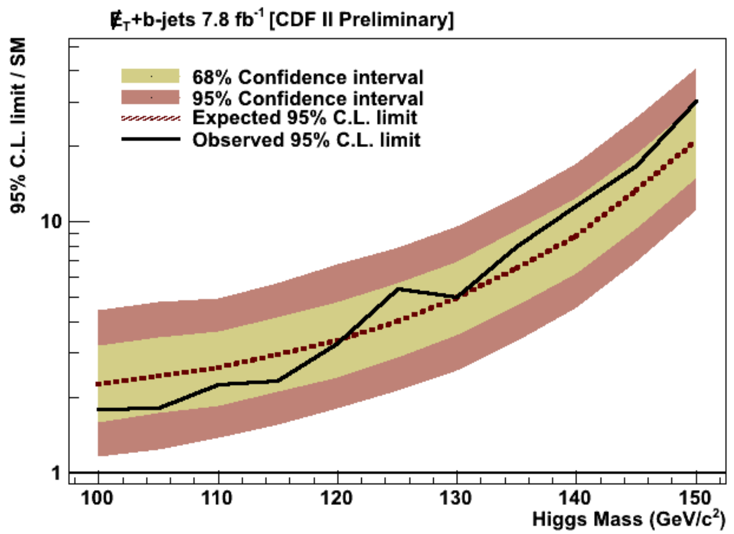




Limits

- Limits set at 95% CL relative to SM cross section





Expected: 4.0 Observed: 3.2	Limits @115 GeV	Expected: 3.0 Observed: 2.3

Conclusions

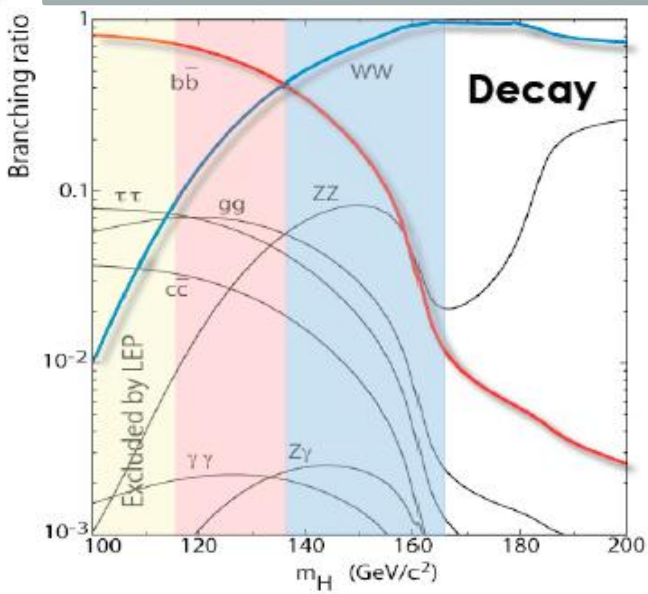
- SM Higgs search for missing energy and b-jets is one of the most sensitive at the Tevatron
  - Vital channel in low mass Higgs search
- Large increase in sensitivity of searches
  - Using up to 8.4 fb^{-1} of data
  - More intelligent use of b-tagging
  - Loosened kinematic cuts
- Many further improvements in pipeline
  - More data
  - Reduced systematic errors
  - Increased acceptance

The SM Higgs is running out of places to hide!

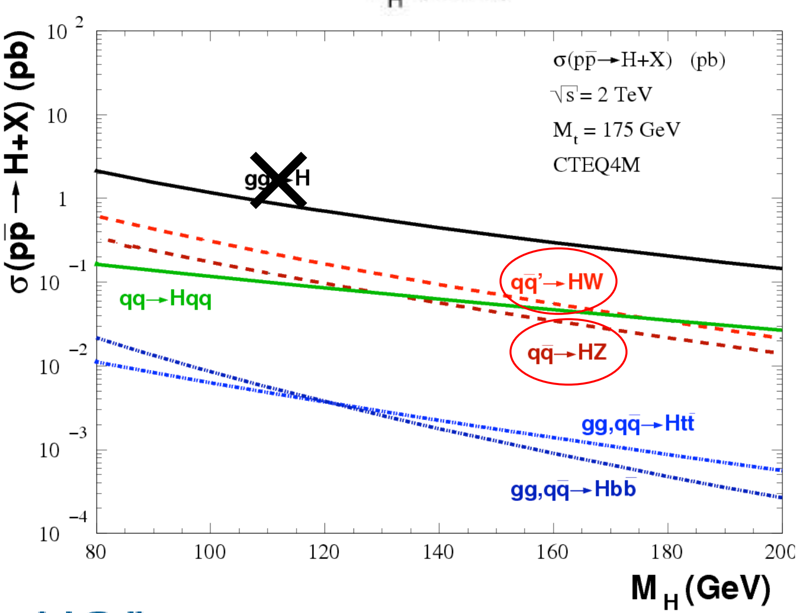
Watch this space

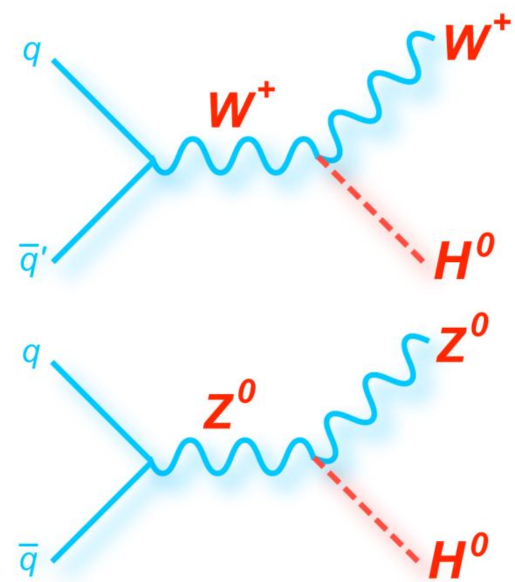
Backup slides

Low Mass Higgs Search



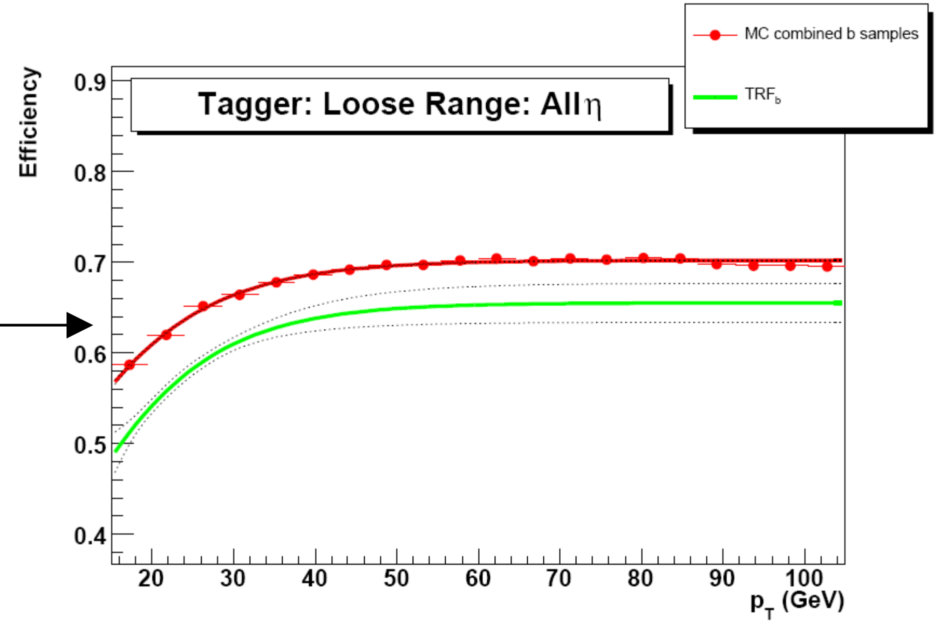
- Low mass Higgs (m_H < 135 GeV)
  - Dominant decay to two b-quarks
  - Large heavy flavour jet background
- Associated production provides more distinctive decays
  - Search for various decay channels of W/Z





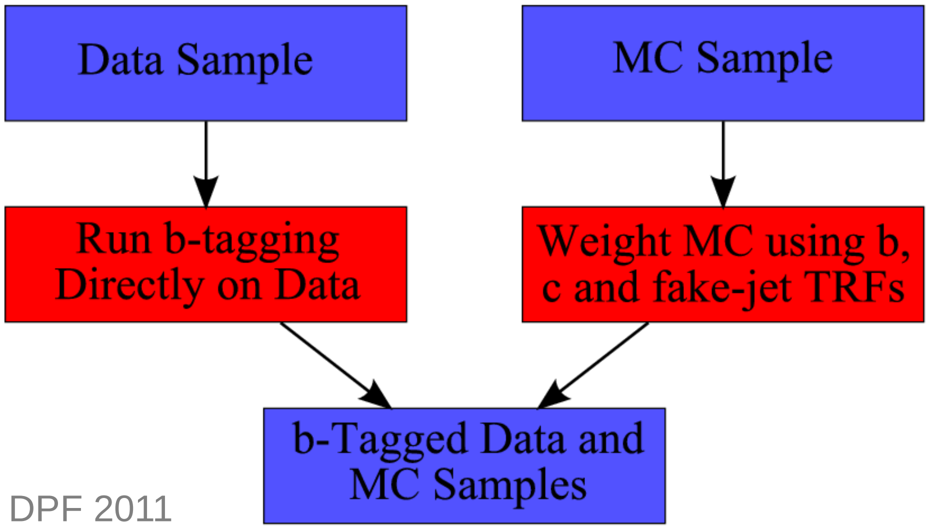
B-tagging – (DØ) Certification

- Have MC / data differences – particularly at a hadron machine
  - ➢ Measure performance on data
    - ▪ Tag Rate Function (TRF) Parameterized efficiency & fake-rate as function of p_T and η
  - ➢ Use to correct MC b-tagging rate



- b and c-efficiencies
  - ➢ Measured using a b-enriched data sample

- Fake-rate
  - ➢ Measured using multi-jet data



Region Definitions

CDF

QCD CR1	EWK CR	QCD CR2	Signal region
No lepton	At least one lepton	No lepton	No lepton
$\not{E}_T > 70$ GeV$/c^2$	$\not{E}_T > 35$ GeV$/c^2$	$\not{E}_T > 35$ GeV$/c^2$	$\not{E}_T > 35$ GeV$/c^2$
$\varphi(j_2, \not{E}_T) \leq 0.4$	$\varphi(j_2, \not{E}_T) > 0.4$	$\varphi(j_2, \not{E}_T) > 0.4$	$\varphi(j_2, \not{E}_T) > 0.4$
	$\varphi(j_3, \not{E}_T) > 0.4$	$\varphi(j_3, \not{E}_T) > 0.4$	$\varphi(j_3, \not{E}_T) > 0.4$
	$\varphi(j_1, \not{E}_T) > 1.5$	$\varphi(j_1, \not{E}_T) > 1.5$	$\varphi(j_1, \not{E}_T) > 1.5$
		$NN_{QCD} < 0.45$	$NN_{QCD} > 0.45$

TABLE IV: Main kinematic selection requirements for each of the control regions and the signal region.

D0

$$\mathcal{D} = (\Delta\phi(\not{p}_T, \mathrm{Jet_L}) + \Delta\phi(\not{p}_T, \mathrm{Jet_NL}))/2$$

Signal region D > $\pi/2$

MJ model region D < $\pi/2$

MJ Rejection Variables

CDF

Variable
Magnitude of $\vec{\not{E}}_T$
Magnitude of $\vec{\not{p}}_T$
$\not{E}_T/\sqrt{\sum E_T}$
$\not{E}_T/H_T$
$\not{H}_T$ / $\not{E}_T$
$M(\vec{\not{E}}_T, \vec{j}_1, \vec{j}_2)$
$\Delta\varphi$ between $\vec{\not{E}}_T$ and $\vec{\not{p}}_T$
Maximum of $\Delta\varphi$ between any two jets
Maximum of ΔR between any two jets
Minimum of $\Delta\varphi$ between the $\vec{\not{E}}_T$ and $\vec{j}_i$
Minimum of $\Delta\varphi$ between the $\vec{\not{p}}_T$ and $\vec{j}_i$
$\Delta\varphi(\vec{j}_1, \vec{j}_2)$ in the 2-jet rest frame
Sphericity
Centrality

D0

Variables used in the MJ DT and in the SM DT
$\Delta\eta(j_1, j_2)$
$\Delta\phi(j_1, j_2)$
$\Delta R((j_1, j_2))$
η of j_1
η of j_2
p_T weighted $\Delta R(j_1, j_{\rm all})$
p_T weighted $\Delta R(j_2, j_{\rm all})$
$\not{E}_T$
$\not{E}_T$ significance
$\Delta\phi(\not{E}_T, j_1)$
$\Delta\phi(\not{E}_T, j_2)$
$\Delta\phi(\not{E}_T, \text{dijet})$
$\min\Delta\phi(\not{E}_T, j_{\rm all})$
$\max\Delta\phi(\not{E}_T, j_{\rm all}) + \min\Delta\phi(\not{E}_T, j_{\rm all})$
$\max\Delta\phi(\not{E}_T, j_{\rm all}) - \min\Delta\phi(\not{E}_T, j_{\rm all})$
$\not{H}_T$(vectorial sum of $j_{\rm all}$ p_T)
$\not{H}_T/$ H_T (with H_T the scalar sum of $j_{\rm all}$ p_T)
Asymmetry between $\not{E}_T$ and $\not{H}_T$
$\not{E}_T$ component along the thrust axis
$\not{E}_T$ component perpendicular to the thrust axis
Sum of the signed components of the dijet and recoil momenta along the thrust axis
Sum of the signed components of the dijet and recoil momenta perpendicular to the thrust axis
Dijet p_T
Scalar sum of j_1 and j_2 p_T
Centrality (ratio of the scalar sum of j_1 and j_2 p_T to the sum of j_1 and j_2 energy)
Effective mass (sum of $\not{E}_T$ and of the scalar sum of j_1 and j_2 p_T)
θ angle of j_1 boosted to the dijet rest frame
θ angle of the dijet system
Polar angle of j_1 boosted to the dijet rest frame with respect to the dijet direction in the laboratory
Azimuthal angle of j_1 boosted to the dijet rest frame with respect to the dijet direction in the laboratory
Color flow j_1
Color flow j_2

Final MVA Variables

D0

MJ DT Variables plus

Dijet mass
Dijet transverse mass
j_1 p_T
j_2 p_T
H_T
j_1 b-tagging output
j_2 b-tagging output

CDF

Variable
Invariant mass of the two leading jets in the event (M_{jj})
Invariant mass of $\vec{\not{E}}_T$, $\vec{j_1}$ and $\vec{j_2}$
Difference between the scalar sum of transverse energy of the jets (H_T) and $\not{E}_T$
Difference between the vector sum of transverse energy of the jets ($\not{H}_T$) and $\not{E}_T$
The output of the TRACKMET neural network
Maximum of the difference in the $\eta - \phi$ space between the directions of two jets, taking two jets at the time
The output of NN_{QCD}

TABLE VI: Input variables to the final discriminant neural network.