Improvements in the Jet Energy Resolution for H$\rightarrow$bb at D0

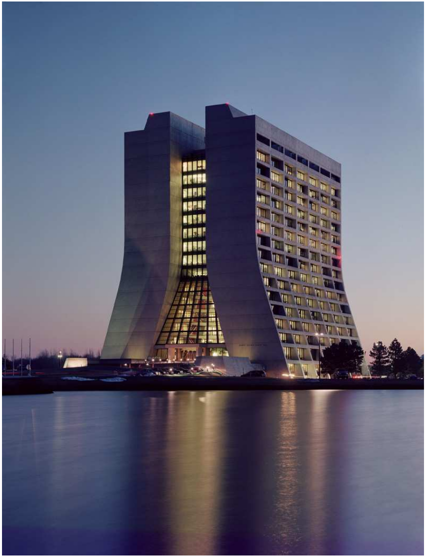
Savanna Shaw

Michigan State University

February 25, 2014

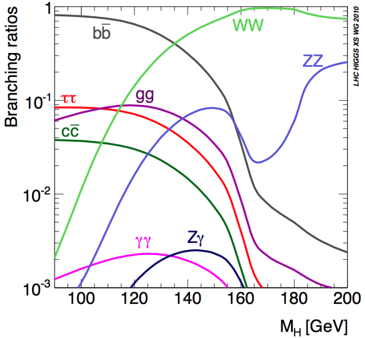
Overview

- Improvements to the Jet Energy Resolution
- Propagation to Higgs analysis: WH$\rightarrow \ell\nu b\bar{b}$
- Summary

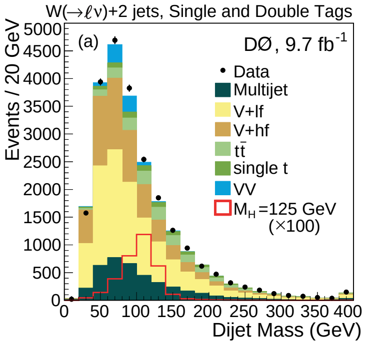


Motivation

- Dominant decay mode for Higgs boson with $m_H = 125$ GeV is $H \to b\bar{b}$.

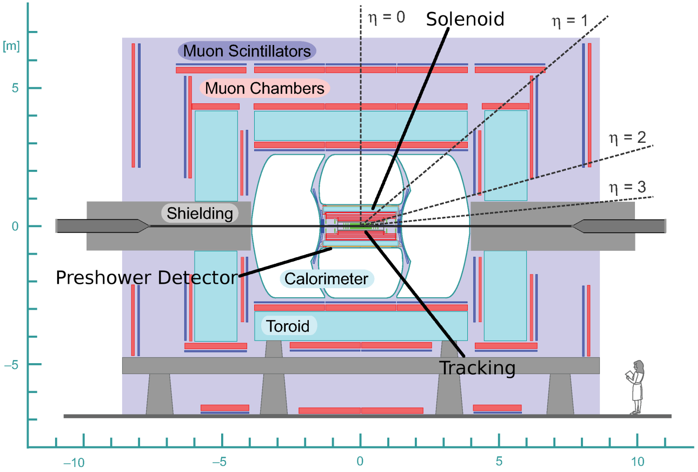


- Most of the sensitivity in $H \to b\bar{b}$ analyses comes from the dijet mass.

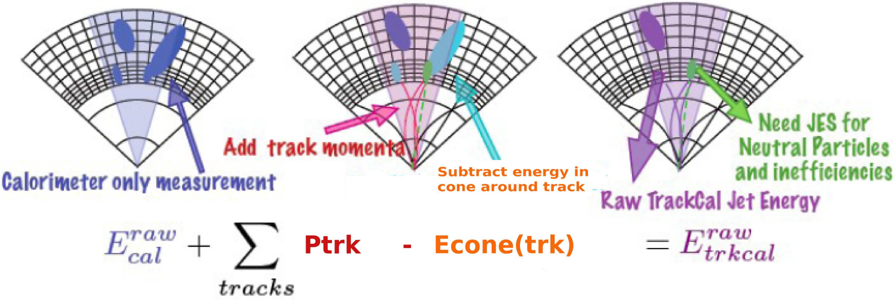


- Phys. Rev. D 88, 052008 (2013)

The D0 Detector

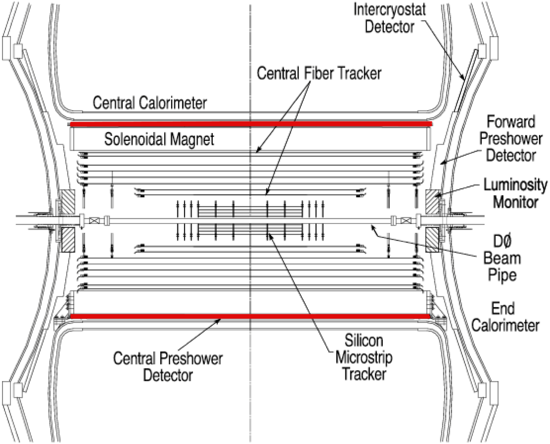


- At low energies, tracker has better resolution than calorimeter, so we can try to use the track momentum measurement instead of the calorimeter energy measurement.
- $E_{trkcal\ jet} = \text{raw jet E} + \Sigma_{trks}(P_{trk} - E_{cone}(trk))$

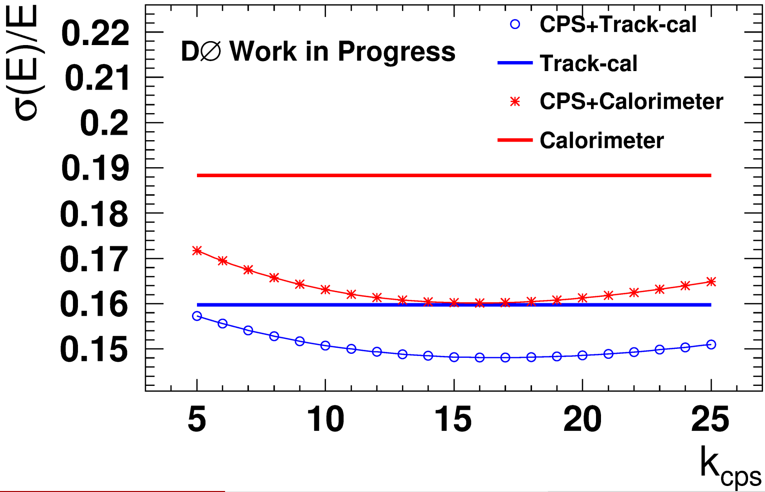


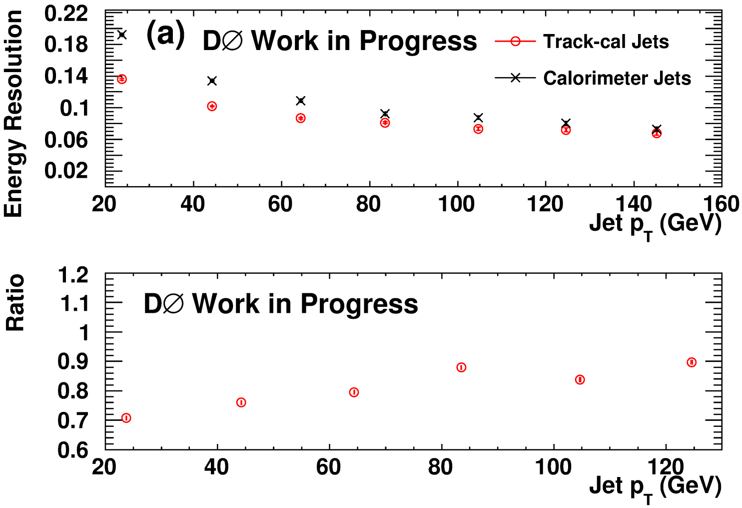
Central Preshower Detector Energy

- Jets are reconstructed using only calorimeter information
- Can add in energy from central preshower (CPS) detector
    - This corrects for energy lost in solenoid and upstream material.

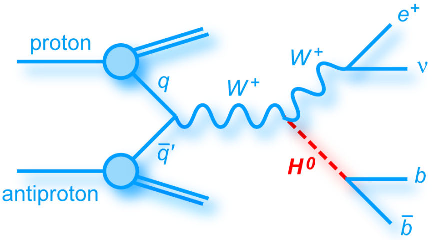


- Define our corrected energy as $E = E_{trkCal} + K_{CPS} * E_{CPS}$.
- K_{CPS} is chosen to minimize resolution $(K_{CPS} = 18)$.



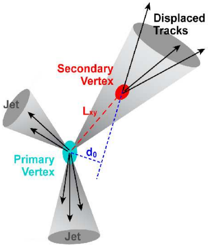


(a) $\sigma(E)/E$, $|\eta_{jet}| < 0.4$

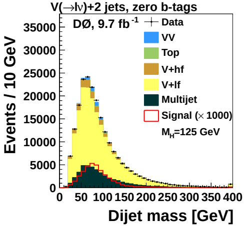


- One charged lepton (e or μ) with $p_T > 15$ GeV
- $\not{E}_T > 15$ (e), 20 (μ) GeV
- Jet $p_T > 20$ GeV, $|\eta| < 2.5$.

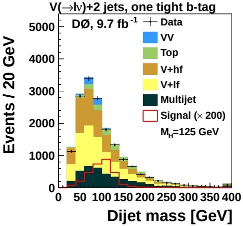
WH Analysis Overview - b-tagging



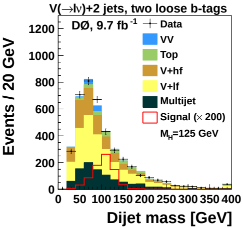
(a) B Hadron Decay



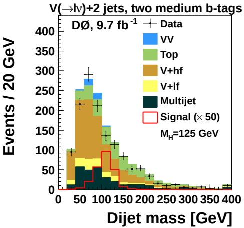
(b) 0 b-tags



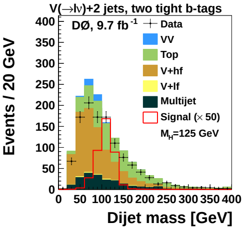
(c) 1 tight b-tag



(d) 2 loose b-tags



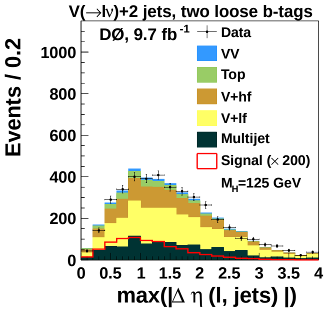
(e) 2 medium b-tags



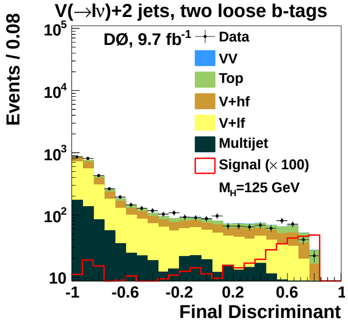
(f) 2 tight b-tags

WH Analysis Overview - Multivariate Analysis

- Use a boosted decision tree to distinguish between signal and background
- Separate boosted decision tree trained for each jet, b-tag and lepton multiplicity to distinguish signal from total background

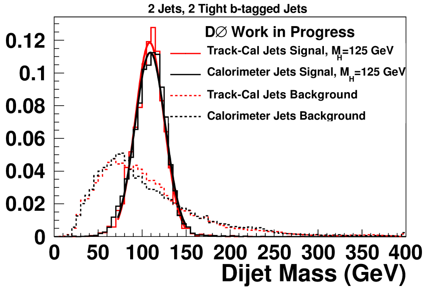


(a) Max $|\Delta\eta(\ell, jets)|$

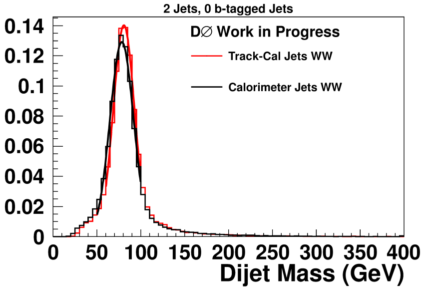


(b) BDT Output

Dijet Mass



(a) 2 jets, 2 tight b-tags



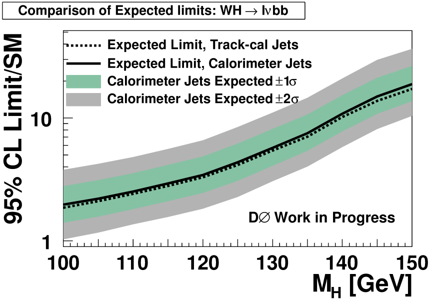
(b) 2 jets, 0 b-tags

- Track-cal jets width: 16.4
- Calorimeter jets width: 17.6

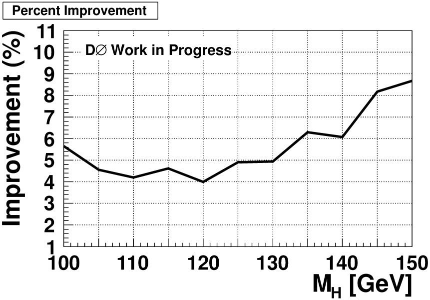
- Track-cal jets width: 12.0
- Calorimeter jets width: 12.9

7% improvement in dijet mass resolution for $H \to b\bar{b}$ and W$\to jj$ events!

Limit Improvement



(a) Expected 95% C.L. limits



(b) Percent improvement

~5% improvement in expected limits!

Summary

- Implemented a correction to the jet energy for charged particles within a jet
- Propagated this correction to D0 WH analysis
  - Improved jet energy resolution by $\sim 20\%$,
  - dijet mass resolution by $\sim 7\%$,
  - and expected upper limit on WH$\rightarrow \ell\nu b\bar{b}$ cross section by $\sim 5\%$.

Backup

TrkCal Jets Backup

$$E_{track-cal} = E_{raw}^{calo} + \sum_{tracks} (P_{track} - E_{track}^{\Delta R=0.15}) + 18E_{CPS},\ if\ |\eta_{jet}| < 1.2,$$
$$E_{track-cal} = E_{raw}^{calo} + 13E_{CPS},\ if\ 1.2 < |\eta_{jet}| < 1.6,$$
$$E_{track-cal} = E_{raw}^{calo},\ if\ |\eta_{jet}| > 1.6.$$

- Need to choose a cone size so that we subtract the energy coming from the track.

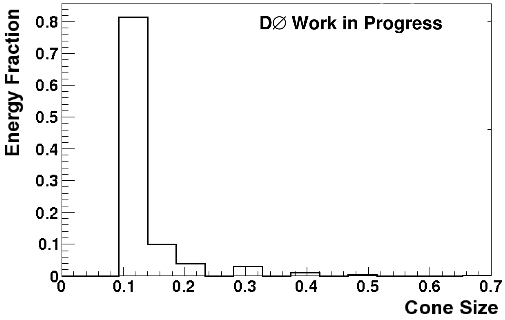
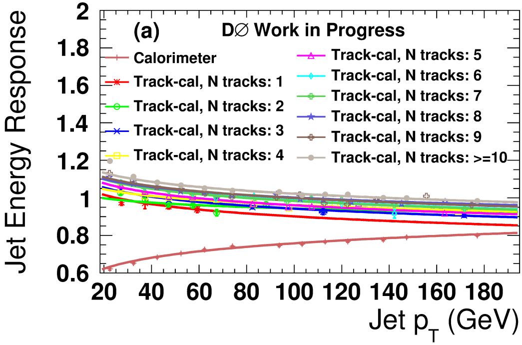


Figure: Fraction of energy deposited vs cone size

Track Selection

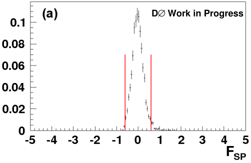
- Track selection was optimized to select high resolution tracks, and cut down on fake tracks
  - —track Z - PV Z—< 0.5 cm
  - Radial DCA (PV vs. Track)< 0.05 cm
  - Fake Track Killer quality> 0
  - Nsmt hits> 0
  - significance of curvature> 5
    - Sig curv $= \frac{P_{T trk}}{\sqrt{\sigma_{trk P_T}}}$

- Calculate response as $p_T^j / p_T^{j\ truth}$ in jet p_T bins using γ+jet events.
- Plot mean as a function of true jet p_T to obtain a response function.

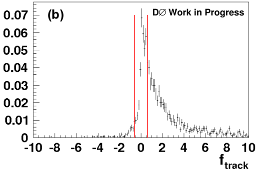


(a) $R = p_T^j / p_T^{j\ truth}$, $|\eta_{jet}| < 0.4$

- Need to be careful not to subtract out too much energy (for example if a neutral particle deposits energy near the track)
  - Did a study in single pion events to have some sort of metric of when too much energy is too much:
  - Considered σ=(E_{cone} - $P_{trk} * R(P_{trk}))/P_{trk}$,
    - $R(P_{trk})$=single pion response.
  - σ parameterized as a function of $P_{trk} * R(P_{trk})$, and jet η.
- Only apply algorithm if $|E_{cone}$ - $P_{trk} * R(P_{trk})|/P_{trk} < 2\sigma$



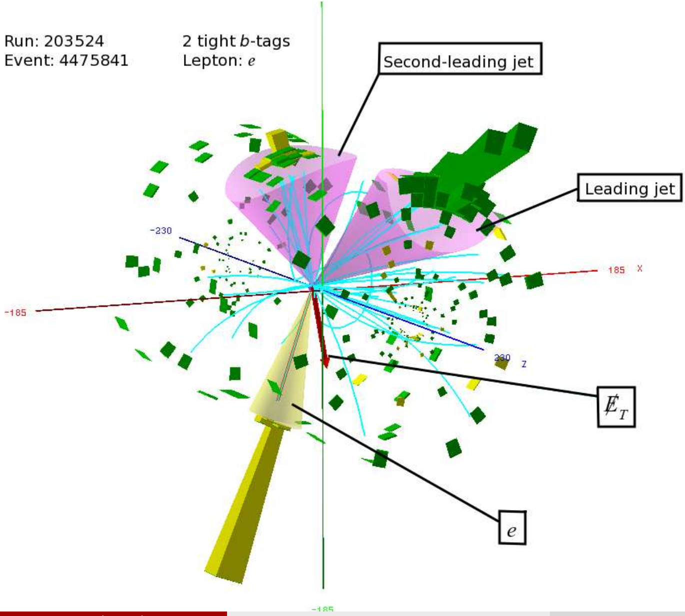
(a) Single Pions, 5 GeV



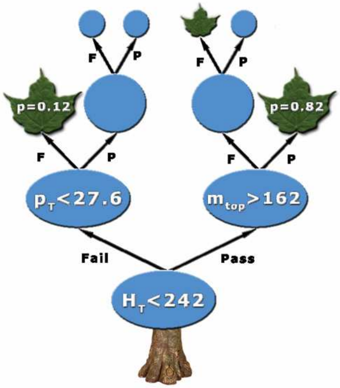
(b) photon+jet, $4.5 < P_{trk} < 5$ GeV

WH Backup

Event Selection



- Start with a list of well modelled variables that have some separation between signal and at least one background.
- Select best variable, and best cut on that variable, to maximally separate signal and background.
- Repeat for signal and background separated samples
- Keep repeating until a state of pure background or signal is reached, or until a minimum number of events is reached.



- Boosting retrains tree to correct for events that were misclassified in the previous tree.
- This builds up a forest of trees which and returns a weighted response:
- $w_i = \frac{1}{\sum_{j=1}^{Ntrees} \alpha_j} \log(\alpha_i) p_i$,
    - α is the tree weight, p is the node purity.

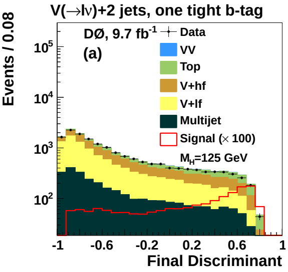
Multijet MVA Input Variables

MVA_{MJ} Input Variables	Description
η^{ν}	Pseudorapidity of the missing E_T vector
$\not{E}_T^{Sig}$	$\not{E}_T$ significance, a measure of the consistency of the observed $\not{E}_T$ with respect to zero $\not{E}_T$, accounting for the uncertainty in the calorimeter objects that contribute to $\not{E}_T$
$\Delta\eta(\ell,\nu)$	Separation in η between the lepton and the reconstructed neutrino, $\lvert\eta^{\ell}-\eta^{\nu}\rvert$
$\mathcal{T}_{W\rightarrow\ell\nu}$	Twist of the $\ell\nu$ system $= \arctan(\Delta\phi(\ell,\nu)/\Delta\eta(\ell,\nu))$
$\cos\theta(\ell)_{\ell\nu CM}$	Cosine of the angle between the charged lepton and the proton beam axis in the CM of $\ell\nu$ system
$\mathcal{V}(j_{12})$	Velocity of the dijet system
m^{Asym}	Mass asymmetry between $\ell\nu$ system and the dijet system: $(M_{\ell\nu}-m_{j_{12}})/(M_{\ell\nu}+m_{j_{12}})$
$\mathcal{C}$	Centrality is $(\sum_i p_T^i)/(\sum_i \lvert\vec{p_i}\rvert)$, where i runs over ℓ and all jets
$\not{E}_T$	Missing transverse energy
p_T^{VIS}	Magnitude of the vector sum of the $\vec{p}_T$ of the visible particles
$\max\lvert\Delta\eta(\ell,\{j_1\ or\ j_2\})\rvert$	Maximum $\Delta\eta$ between the charged lepton and the leading or second leading jet

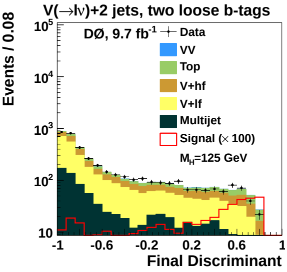
Final MVA Input Variables

Variable	Description	2T	2M	2L	1T
$\mathrm{MVA_{MJ}}(VH)$	Output of the multivariate discriminant trained to distinguish $WH \rightarrow \ell\nu b\bar{b}$ from the MJ background	1	1		
m_{b1b2}	The dijet invariant mass	2	4	3	1
$p_T^W/(p_T^\ell + \not{E}_T)$	The transverse momentum of the reconstructed W divided by the sum of the transverse momentum of the lepton and the missing transverse energy	3	6	4	2
b_{ID}^{j12}	Averaged b-jet identification output for the highest energy b-tagged jets	4	13	1	4
$\cos(\chi^*)$	$\chi^* = \angle(\ell,$ the direction of the W boson spin) in $\ell\nu$ system CM frame [?]	5	3		
$\max\|\Delta\eta(\ell, \{j_1\ or\ j_2\})\|$	Maximum $\Delta\eta$ between the charged lepton and the leading or second leading jet	6	11	2	3
$q^\ell \times \eta^\ell$	Product of the lepton charge and its pseudorapidity	7	2	6	6
$\Delta\mathcal{R}(\ell, j_1)$	$\Delta\mathcal{R}$ between the charged lepton and the leading jet	8	5		
$\min[\mathcal{SIG}(j_{12}, \{j_1\ or\ j_2\})]$	Minimum $\mathcal{SIG}$ of the leading or second leading jet defined as $p_T^{\min}(j_1, j_2)\Delta\mathcal{R}(j_1, j_2)/\sum_{i=1}^{2} p_T^{j_i}$ with respect to the dijet system. Based on the pull variables described in Ref. [?]	9	15	9	5
$q^\ell \times \eta^{j1}$	Product of the the lepton charge and pseudorapidity of the leading jet	10	7	11	9
$\mathcal{V}(j_{12})$	Velocity of the dijet system	11	12	7	11
$\cos(\theta^*)$	$\theta^* = \angle(W, incoming\ \ u - type\ quark)$ in the Higgs CM frame [?]	12	10		
$m_{\ell\nu j_2}$	Invariant mass of the system consisting of the charged lepton, reconstructed neutrino, and second leading jet	13	16	12	13
m_T^{j12}	Transverse mass of the leading and second leading jets	14	14		
$\mathcal{C}$	Centrality is $(\sum_i p_T^i)/(\sum_i \|\vec{p_i}\|)$, where i runs over ℓ and all jets	15	8	8	10
$\sum(p_T)^{VIS}$	Scalar sum of the p_T of the visible particle	16	9		
m^{Asym}	Mass asymmetry between $\ell\nu$ system and the dijet system: $(M_{\ell\nu} - m_{j_{12}})/(M_{\ell\nu} + m_{j_{12}})$			5	8
$\mathcal{A}$	Aplanarity is $3\lambda_3/2$ where λ_3 is the smallest eigenvalue of the normalized momentum tensor $S^{\alpha\beta} = (\sum_i p_i^\alpha p_i^\beta)/(\sum_i \|\vec{p_i}\|^2)$, where $\alpha, \beta = 1, 2, 3$ correspond to the x, y, z momentum components, and i runs over selected objects. Without arguments, it is calculated for all visible objects			10	12
p_T^{j2}	p_T of the second leading jet			13	7

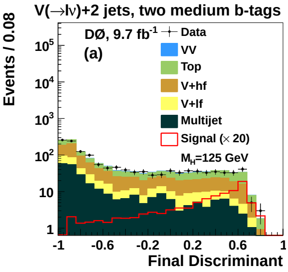
MVA Output Distributions



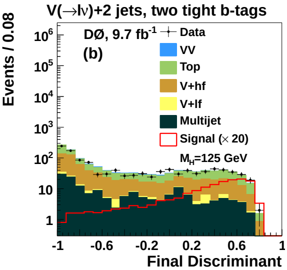
(a) 1 tight b-tag



(b) 2 loose b-tags



(c) 2 medium b-tags



- Consider two hypotheses: Background only (H_0), and Signal + Background (H_1))
- Assume data comes randomly from a Poisson distribution, and generate pseudo data for two hypotheses
- Use Log-Likelihood Ratio (LLR) test statistic to evaluate statistical significance.
  - LLR is the -2 times the log of the ratio of likelihoods for the test and null hypotheses:
  - LLR $= -2\log\left(\frac{LH(H_1)}{LH(H_0)}\right)$.
  - Most powerful test statistic for distinguishing between two hypotheses.
  - Calculated for each set of pseudo-data for background only and signal + background hypotheses.

$$LLR = \sum_{i=0}^{Nc} \sum_{j=0}^{Nbins} s_{ij} - d_{ij} \ln\left(1 + \frac{s_{ij}}{b_{ij}}\right) \quad (1)$$