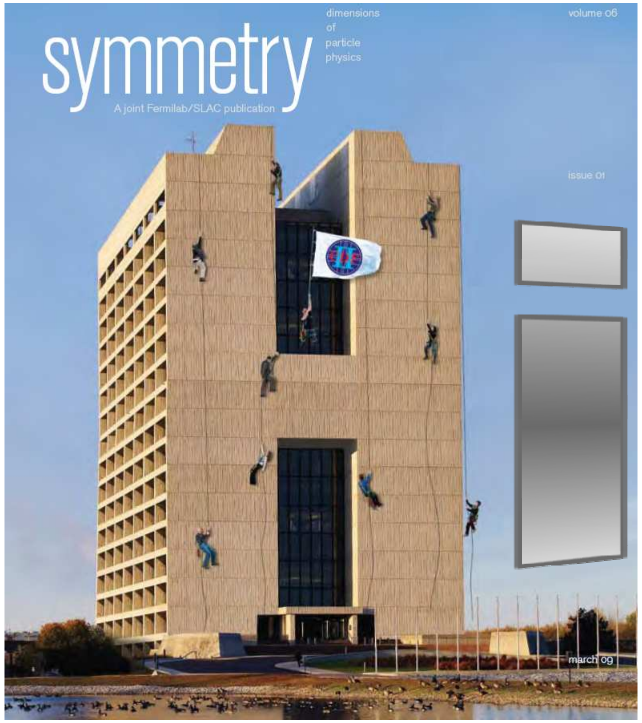
Search for WZ→lνbb at CDF: Proving ground of the Hunt for the



iggs Boson

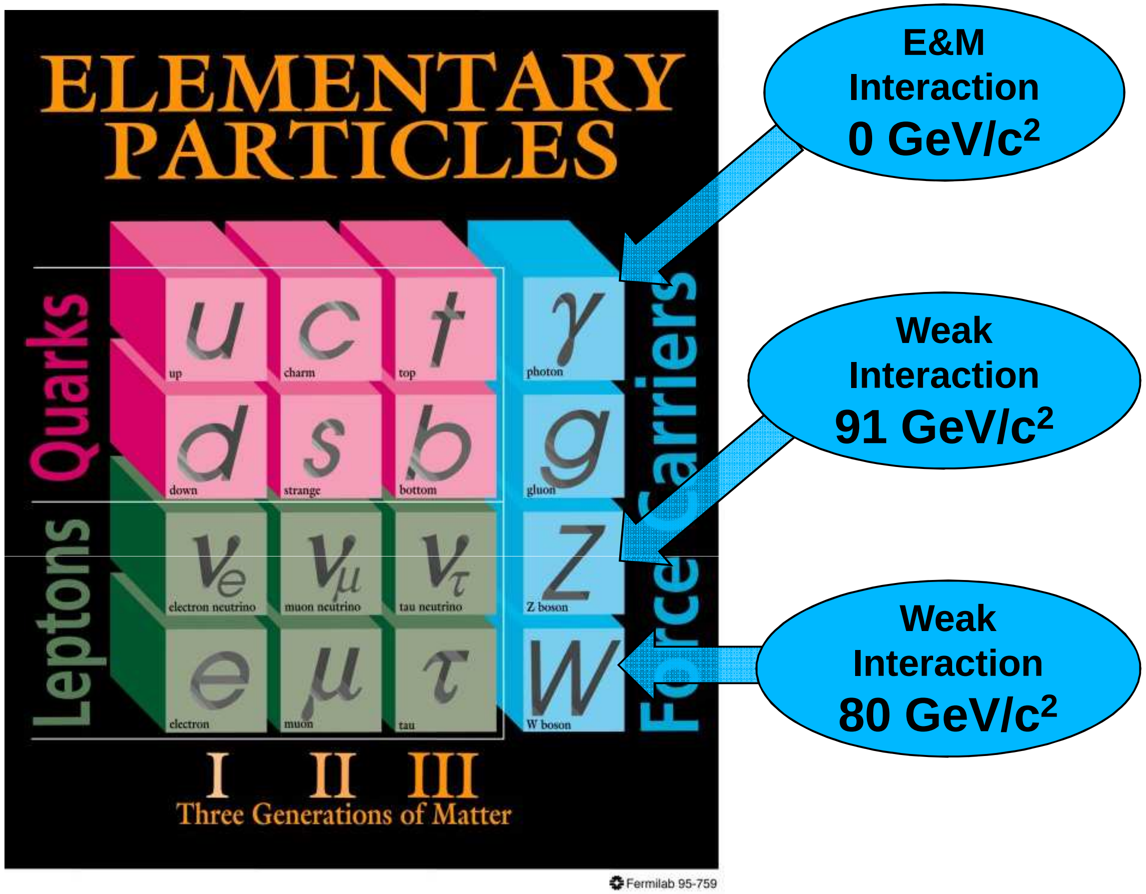
Justin Keung
University of Pennsylvania

March 24th, 2010
HEP Seminar @ University of Virginia

Outline

- Motivation: Higgs Boson Search
  - Importance of b-tagging
  - Comparison of signatures from WZ and WH
- b-tagging improvements
  - Calibration of efficiency
  - Calibration of mis-tag rate
  - Comparison
- Search for WZ→lνbb
  - Selection
  - Discrimination
- Conclusion & Future Plans

Electroweak Symmetry Breaking



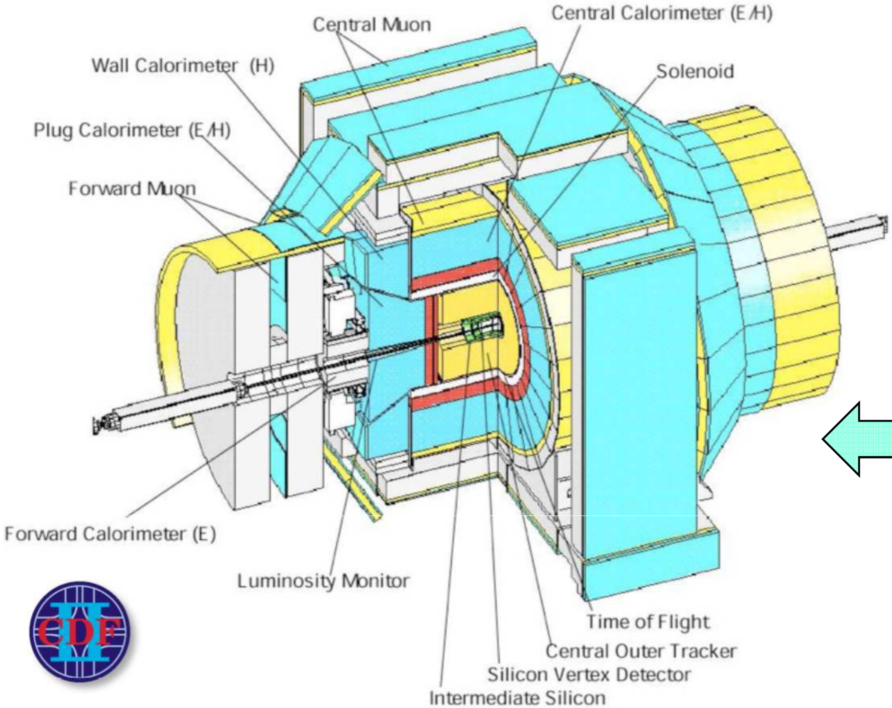
- One of the main goals of particle physics is to understand electroweak symmetry breaking
- The Higgs mechanism breaks the electroweak symmetry, giving mass to the W and Z bosons but not the photon

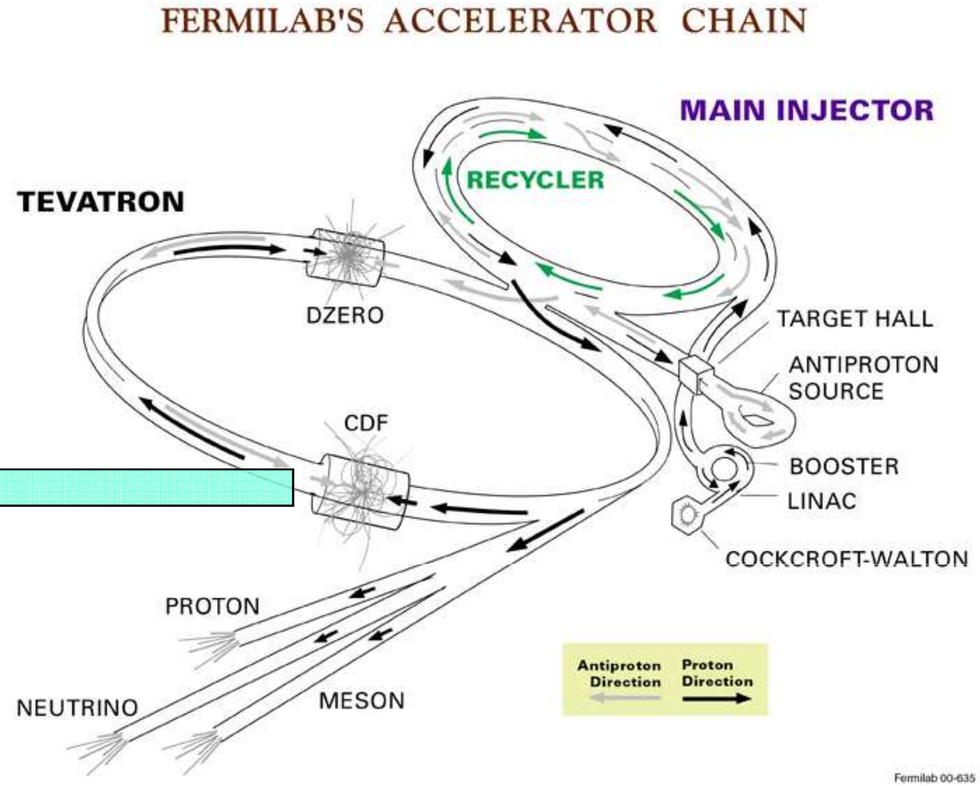
- Higgs boson mass isn't predicted, and it hasn't been observed yet, despite 40+ year hunt since 1968
- SM can predict very well possible production and decay mechanisms since we have measured masses of other particles in SM

Standard Model and Higgs

- Peter Higgs proposed that particles acquire mass by interacting with the Higgs field, which has non-zero vacuum expectation value
  - Analogy: vacuum is like a party, celebrities (particles) passing through interact with groupies (Higgs field) and is slowed down (acquires mass)

Fermilab Tevatron and CDF

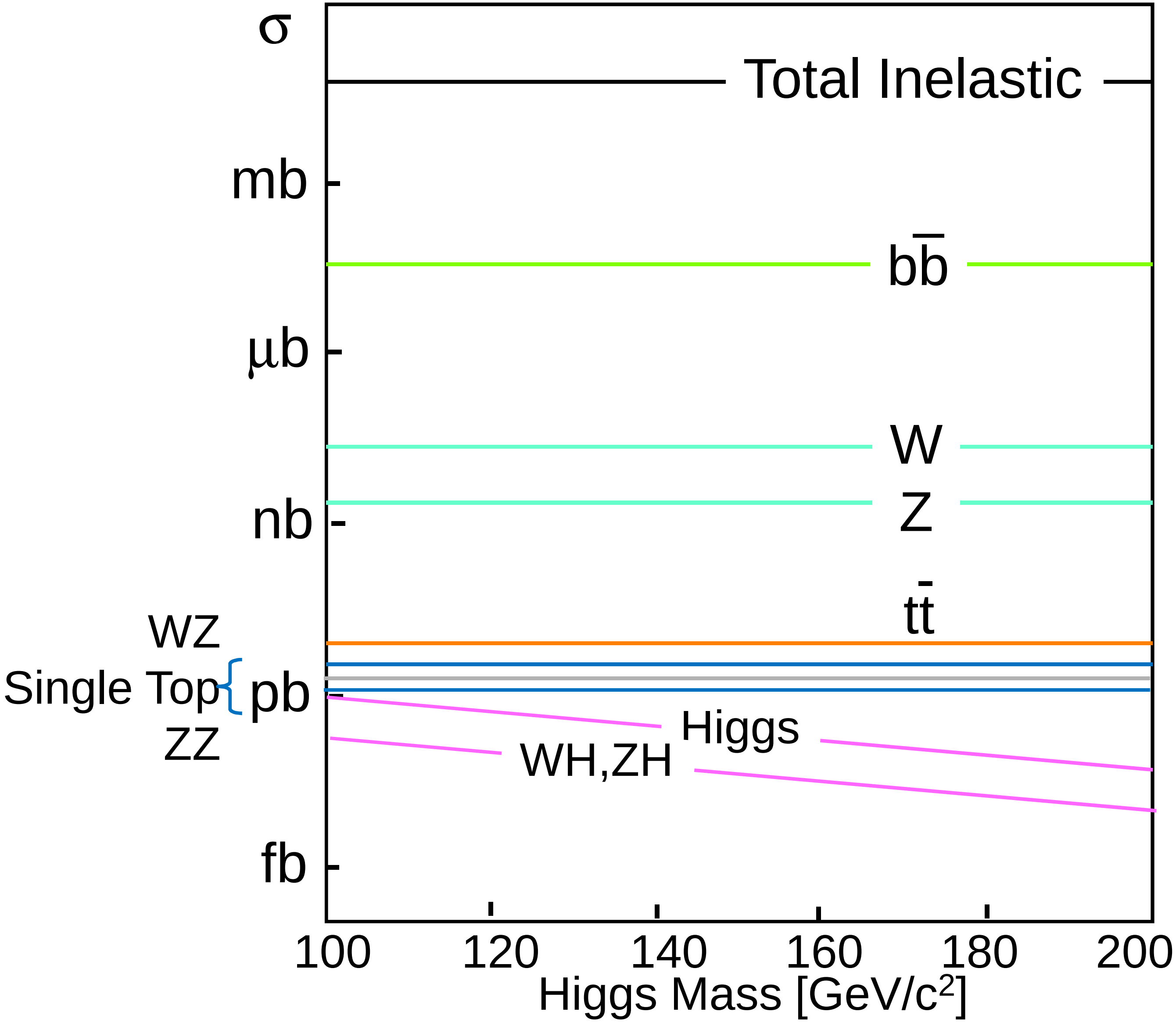




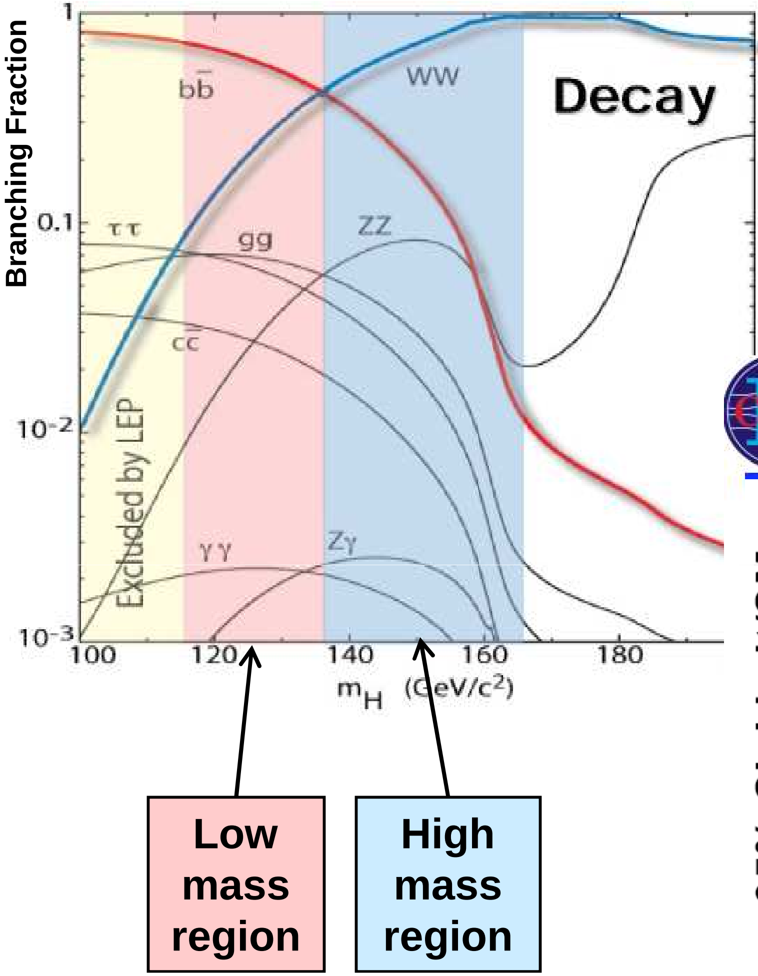
- We collide protons(p) and anti-protons(pbar) accelerated to high energies with the Tevatron, collisions detected at the 3 stories tall CDF
- Tevatron ring has radius of 1km
- Produces p-pbar collisions 1.7 million times per second, in 36 bunches with 250 billion protons and 50 billion antiprotons each
- Running since 2001 at center-of-mass of 1.96 TeV
- Projected to deliver $12fb^{-1}$ by end of 2011

Road to the Higgs

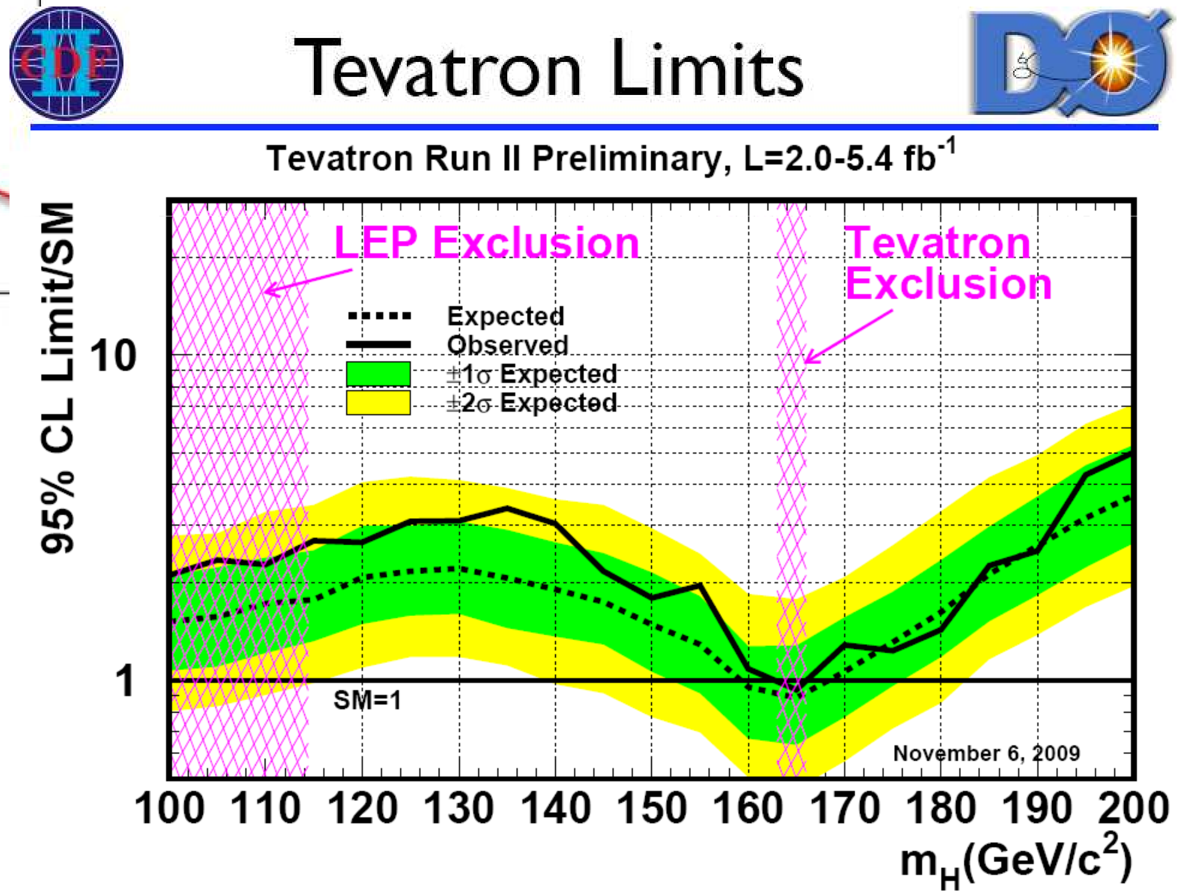
- Unfortunately, Higgs production is rare while backgrounds are huge



Searching for Higgs @Tevatron

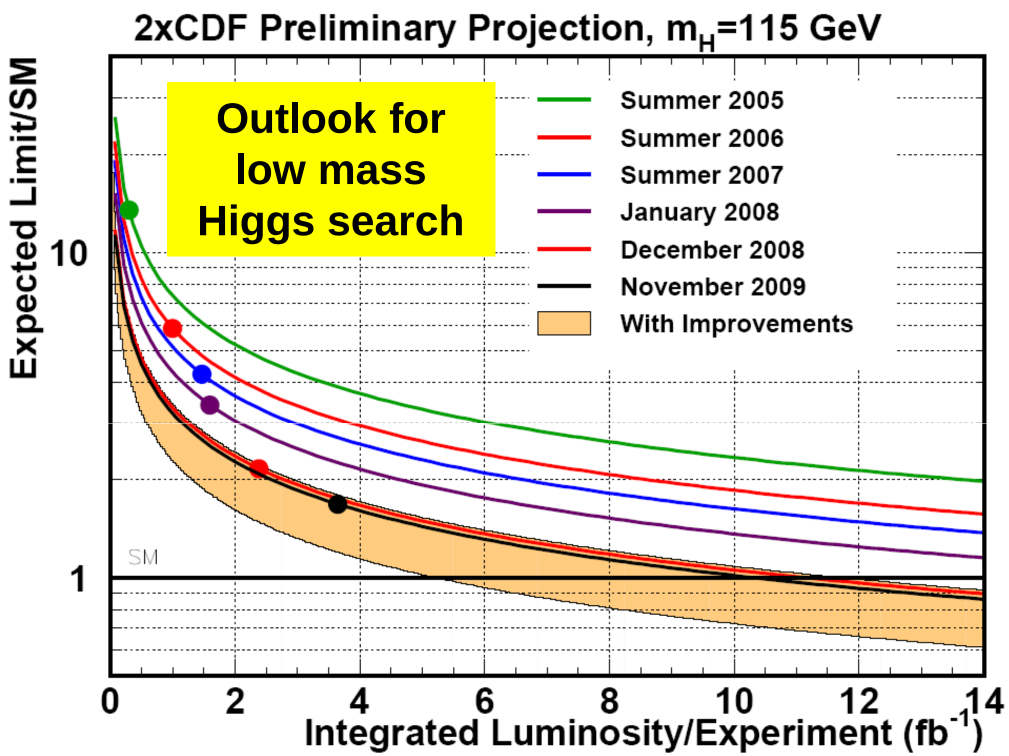


- Our sensitivity to the Higgs boson depends on how it decays in the detector
- Tevatron needs more sensitivity at low mass region (H→bb): important to increase b identification power



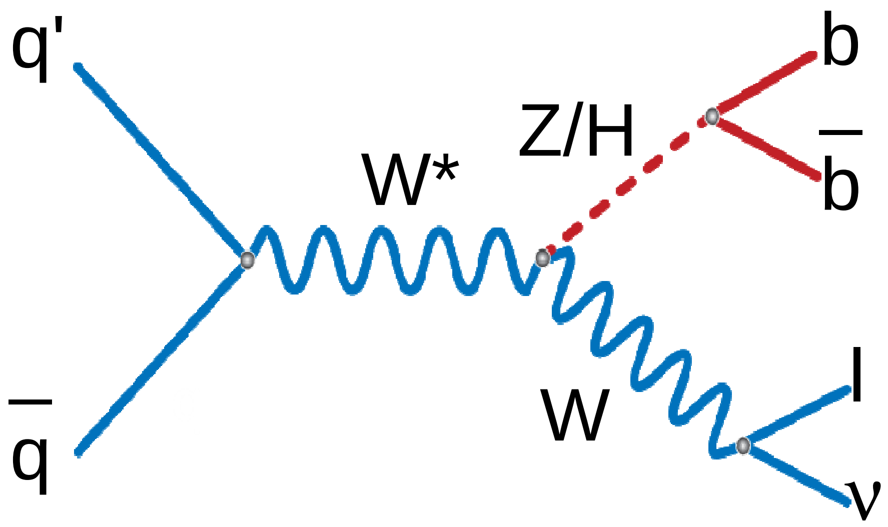
Higgs Status and Prospects

- Projection with all CDF channels, assuming that D0 has same sensitivity



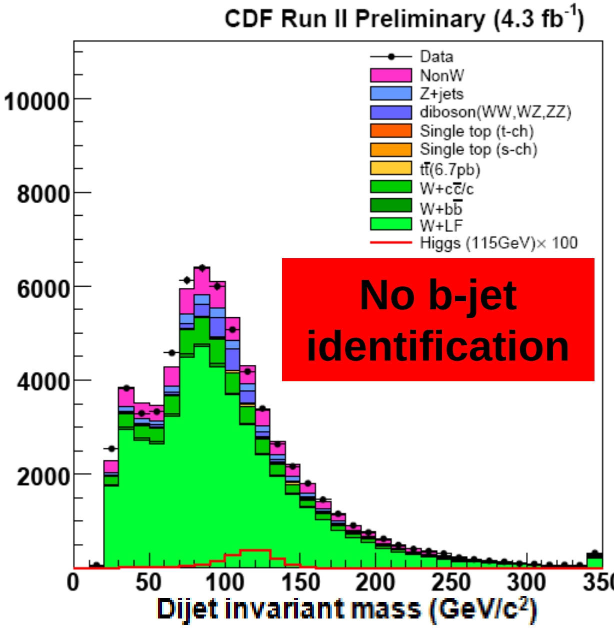
- My contributions are on the November 2009 line, only in one channel
  - ➢ Will be usable for all low mass Higgs channels

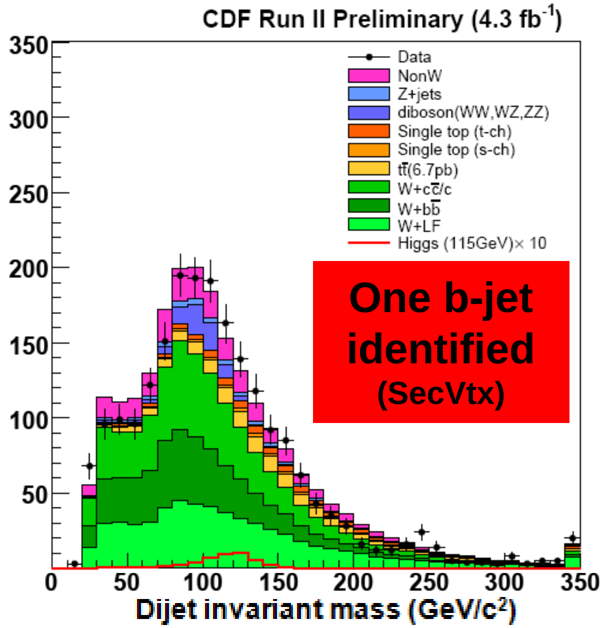
My Channel

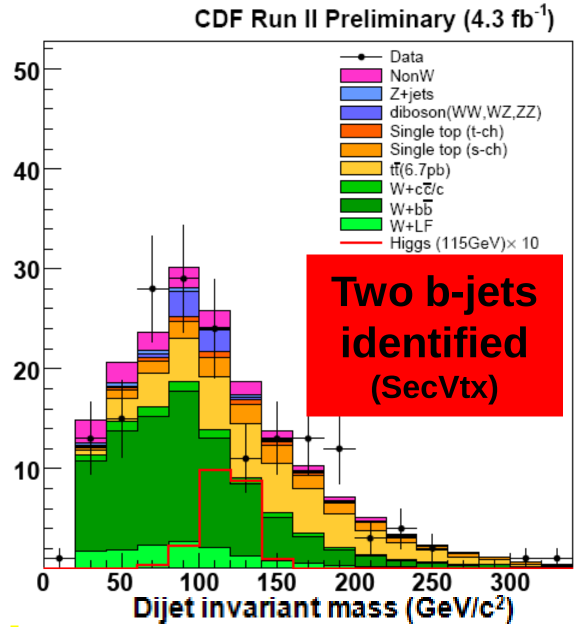


- Standard Model Z boson produced in association with a W boson
  - Identification of b-jets is important in this channel

Why Identify the b-jets?

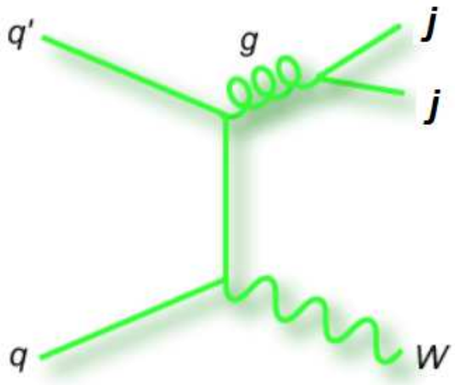






WH Limit = 15x SM
WW: 76%, WZ: 23%, ZZ 1%

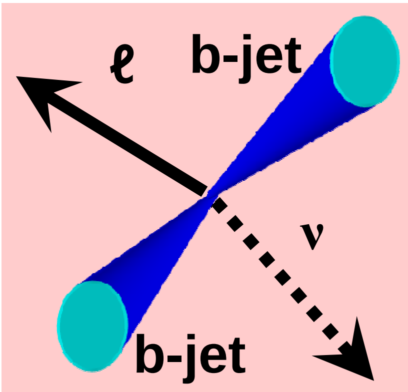
WH Limit = 4x SM
WW: 4%, WZ: 64%, ZZ 32%



- Identification of b-jets improves the signal to background ratio
- Important to improve b-jet identification efficiency to gain more signal in the best signal to background channel
  - If per-jet efficiency increases 10%, then the number of events with 2 identified b-jets increases 21%

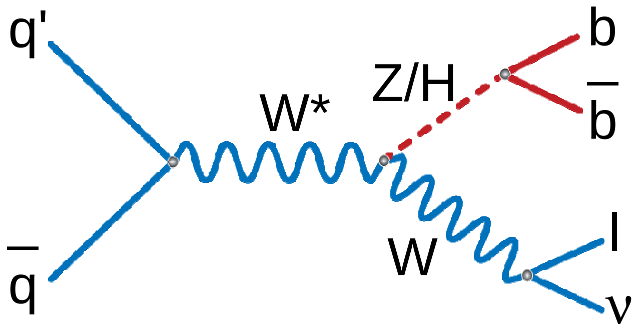
WZ vs WH Search

- WZ search is an excellent test of WH search tools
  - Same final state and similar topology
  - WZ→lvbb has effective cross section 5x higher than that predicted for WH→lvbb (H @ 120 GeV/c^2)



	WZ	WH
Production cross section	3.96pb	0.16pb
W->lν (e or μ) branching fraction	0.21	0.21
Z/H->bb branching fraction	0.15	0.7
XSec x BR(W→lν) x BR(Z/H→bb)	0.125pb	0.024pb

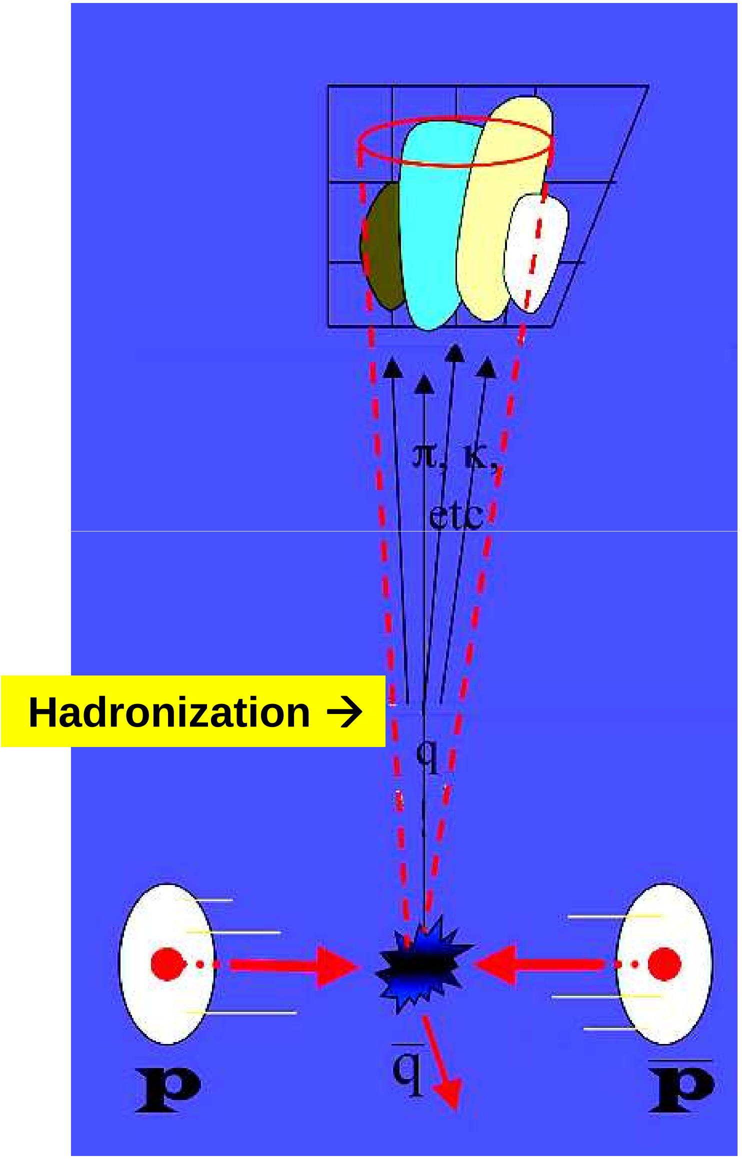
- Set a limit WZ production in order to test the b-jet identification tools and sophisticated search techniques used for WH



Outline

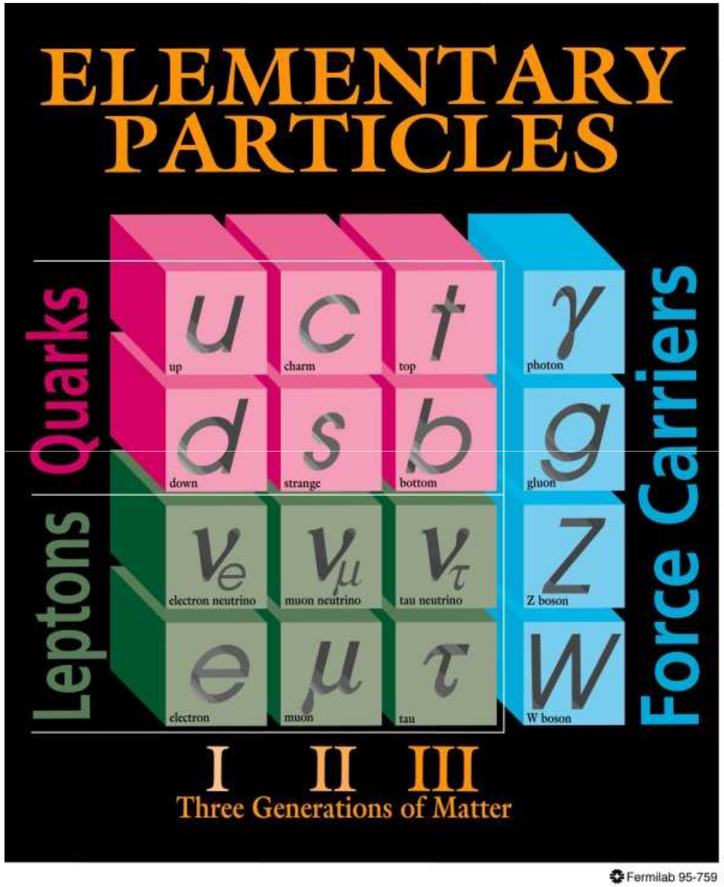
- Motivation: Higgs Boson Search
  - Importance of b-tagging
  - Comparison of signatures from WZ and WH
- b-tagging improvements
  - Calibration of efficiency
  - Calibration of mis-tag rate
  - Comparison
- Search for WZ→lvbb
  - Selection
  - Discrimination
- Conclusion & Future Plans

Jets



- In proton-antiproton collisions, quarks and gluons are produced
- After quarks are produced, they hadronize into mesons and baryons
- If the quark was produced with significant energy, the hadrons produced will be in a narrow cone
- This spray of particles is called a jet

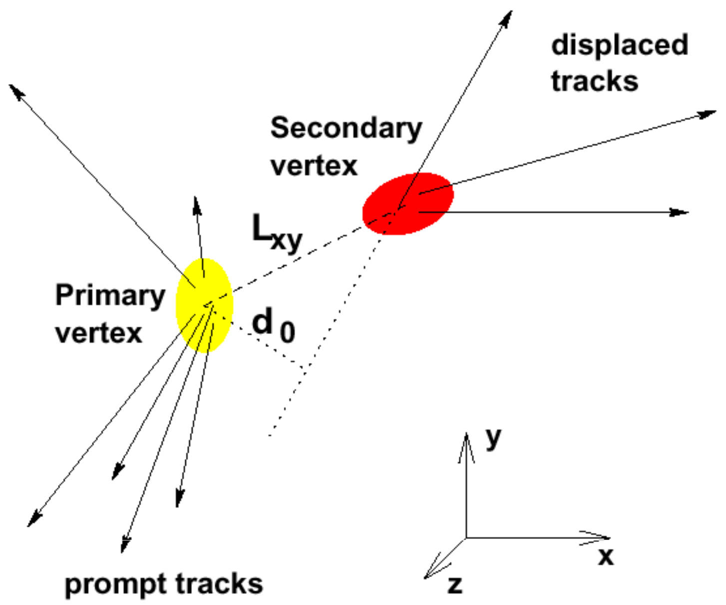
Bottom Quarks



- The bottom quark is the second heaviest quark
  - B-mesons have masses ~5.3GeV/c^2
    B-baryons have masses ~5.6GeV/c^2
- B-hadrons have mean lifetime of ~1.5ps
  - Example: a 53 GeV b-jet has γ=10, and travels almost at c, so on average it travels 4.5mm before decaying
- We want to identify those jets that have a B hadron within them
  - This identification is called "b-tagging", or to "tag" the b-jet

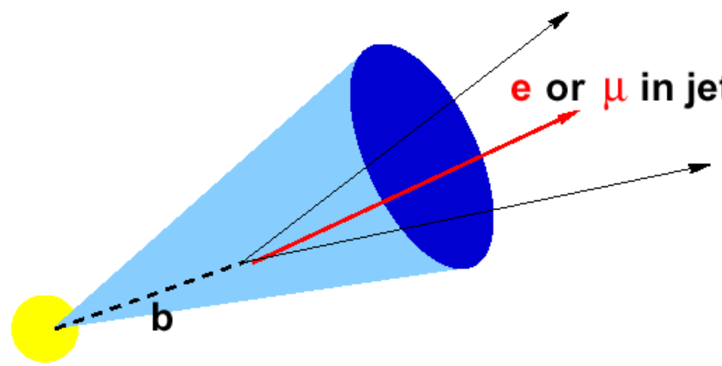
Bottom Jets Identification

- "Long"-lived and massive
  - Secondary vertex significantly displaced from primary vertex
  - More tracks with large impact parameters
    - CDF Silicon detector has track hit resolution of 10um, impact parameter resolution is 30um



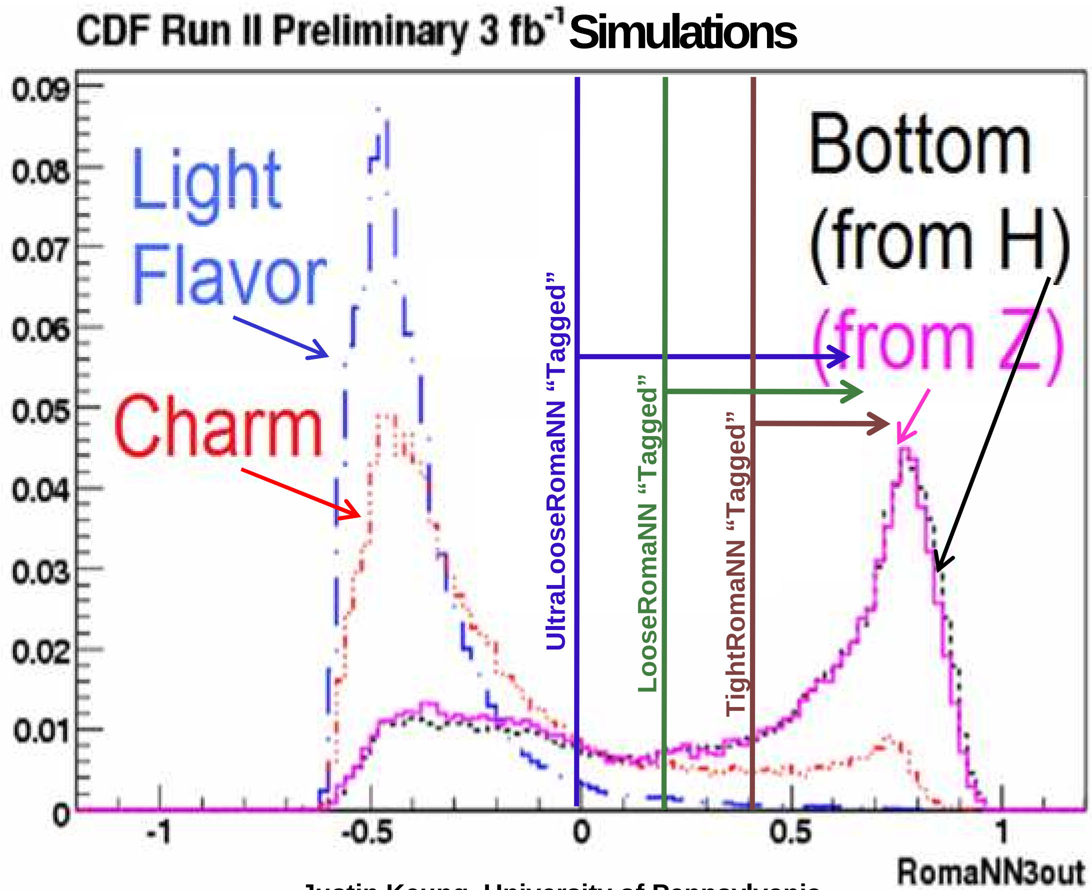
- Decays semileptonically
  - 40% of b-jets has a μ/e within the jet (uses μ only)

- **New algorithm RomaNN** uses both of these characteristics



RomaNN

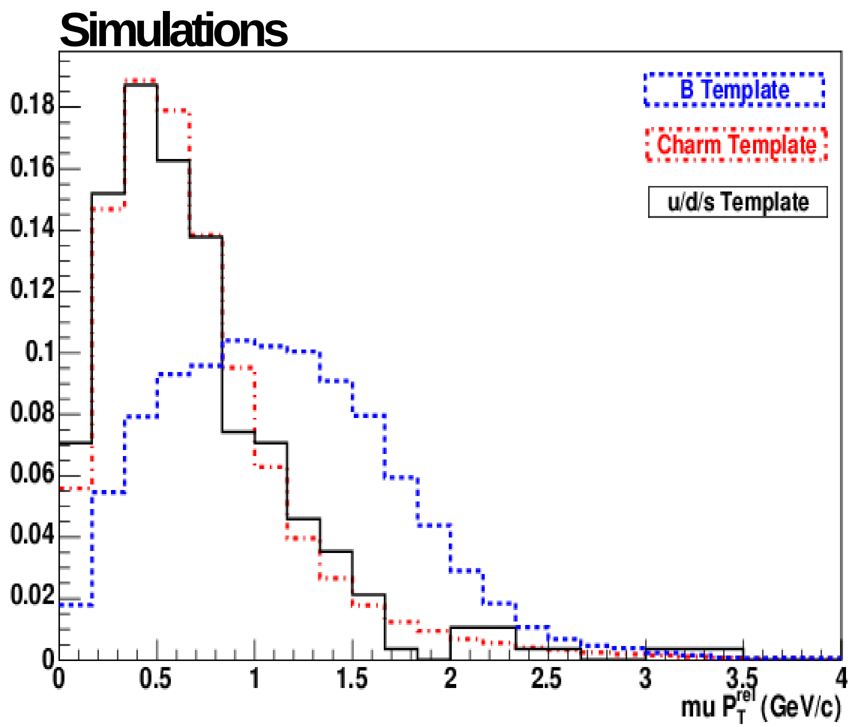
- New algorithm, provides an indicator of how consistent the jet is with coming from a b-quark → performed well during initial testing
- I adapted it to perform b-tagging (binary mode), and commissioned it in order for it to become another standard CDF b-tagging algorithm
- Three cuts: UltraLoose/Loose/Tight, different b-purity and efficiency

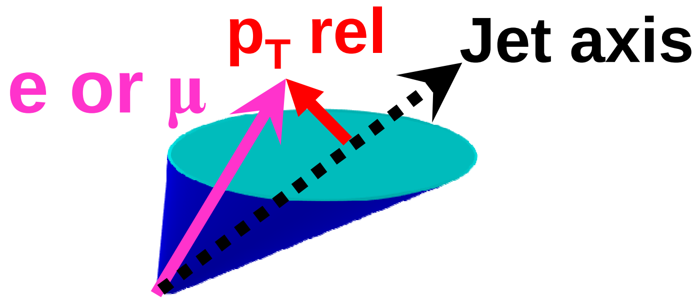


b-tagging Calibration: Efficiency

- Measure b-tag efficiency in data, use lepton P_T^{rel} to discriminate b from charm and light flavor jets

- To calibrate efficiency, we need a sample of b jets in the data: choose di-jet data sample
  1. Away-jet: tagged to improve b purity of the sample
  2. Probe-jet: jet containing a lepton





P_T^{rel}: lepton transverse momentum relative to jet axis

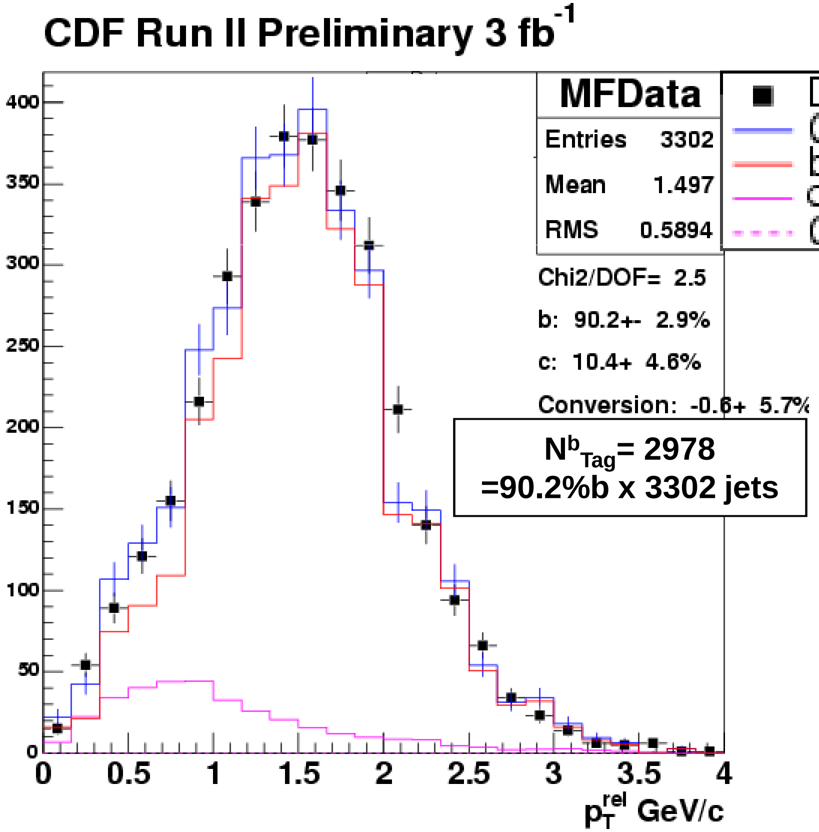
$$lepton\ p_T^{rel} = \left|\vec{P}_{lepton}\right| \sqrt{1 - \left(\frac{\vec{P}_{lepton} \cdot \vec{P}_{jet}}{\left|\vec{P}_{lepton}\right| \left|\vec{P}_{jet}\right|}\right)^2}$$

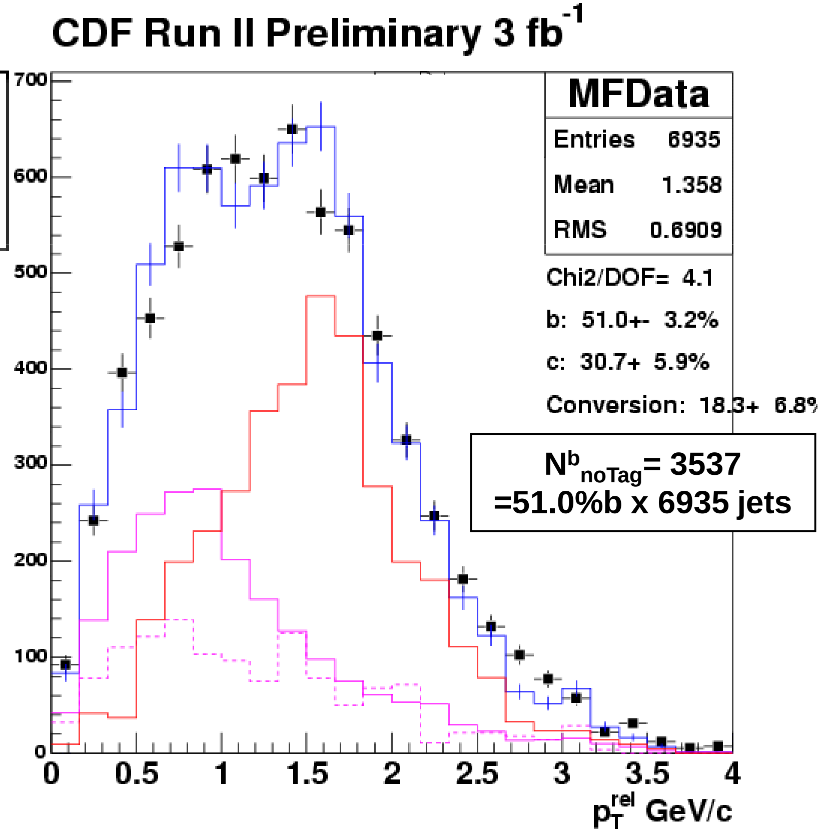
- Due to the large b mass, the P_T^{rel} is larger for bottom jets than for charm and light jets

Data Efficiency Measurement: Example

$$\epsilon_b = \frac{N^b_{Tag}}{N^b_{total}} = \frac{N^b_{Tag}}{N^b_{Tag} + N^b_{NoTag}}$$

- Split a sample of jets (with b/c/light fractions unknown) into 2 subsets
  jets tagged (left), ---and--- jets not-tagged (right)

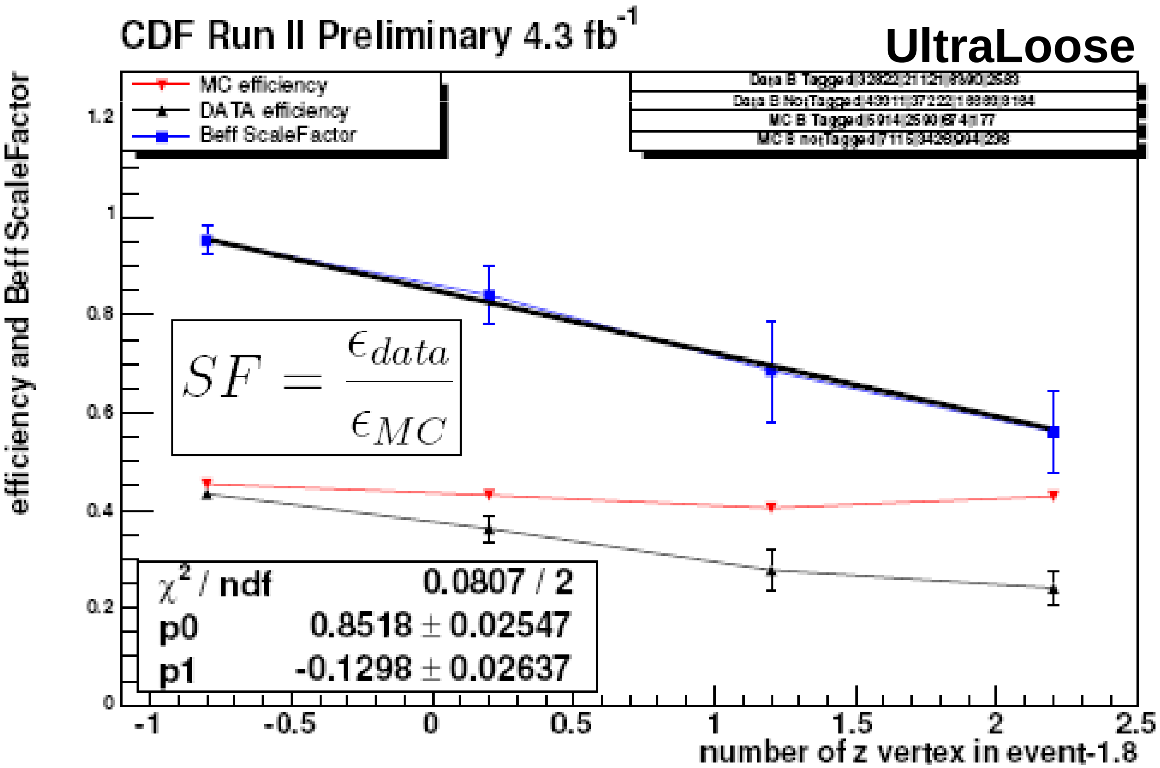


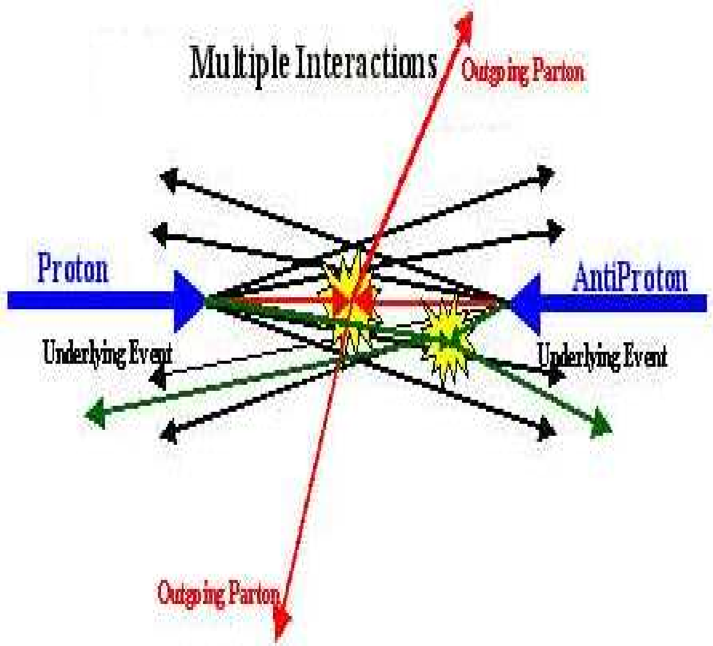


in this example, ϵ_b=2978/(2978+3537)=46%

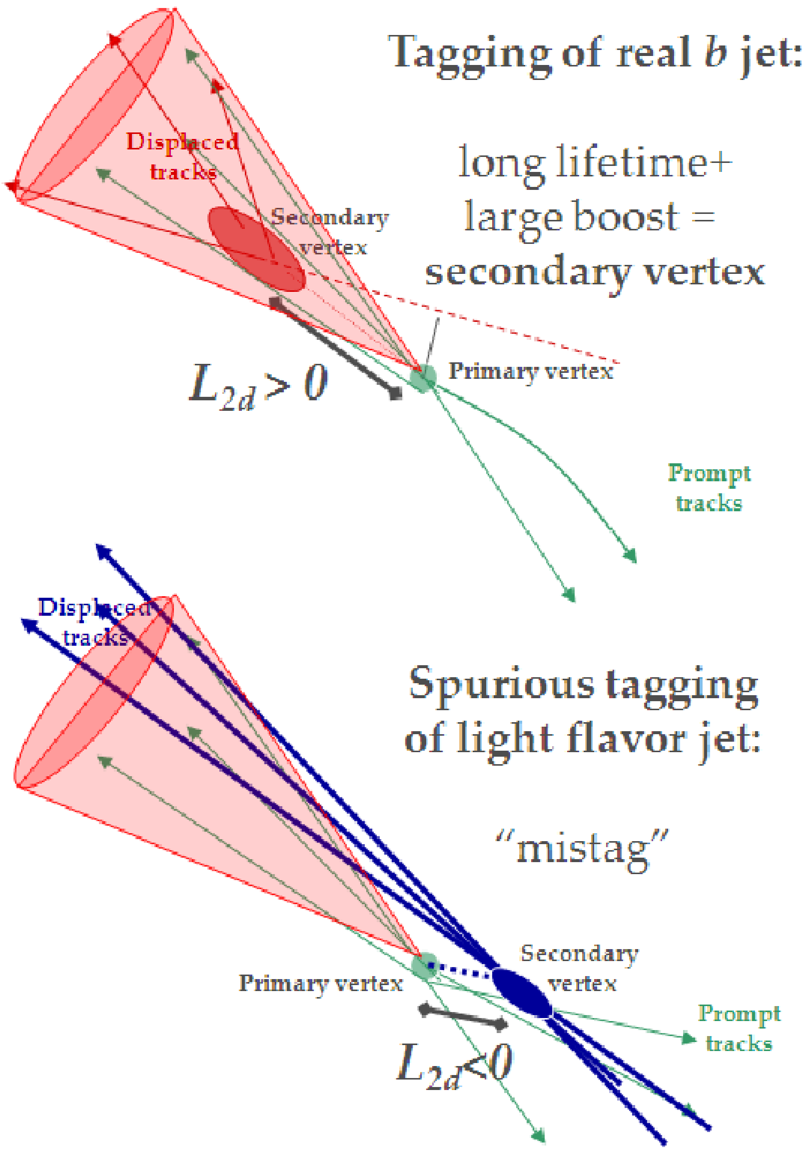
Data Efficiency Result

- After measuring efficiency in data, it is time to compare it to simulation
- Scale factor (SF) corrects the simulation efficiency to Data efficiency
  - Needed to predict yields
- SF clearly decreases as the number of z vertices increases
  - # z vertices is a measure of multiple collisions per bunch crossing
    - ➢ Produce extra tracks in the detector, important because our simulation does not model well its dependence





b-tagging Calibration: Misidentification Rate



- Misidentifications are due to spurious large impact parameter tracks
  - From limited detector resolution, long-lived light particle decays, and material interactions
- For SecVtx (an algorithm that searches only for a secondary decay vertex), misidentification due to the limited detector resolution is expected to be symmetric in their L_{2d}
  - Signed 2D displacement of the vector separating the primary and secondary vertices

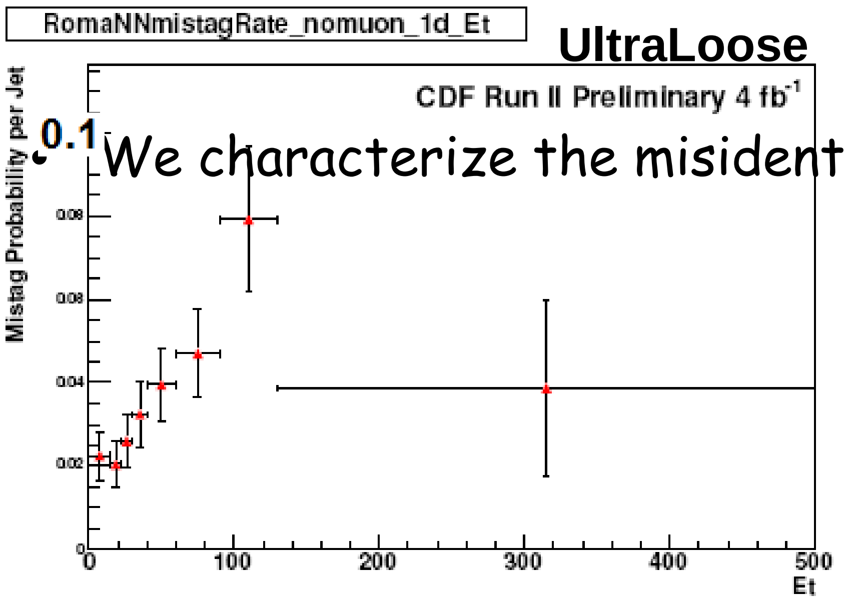
Misidentification Rate Measurement

- For RomaNN, the misidentifications due to the limited detector resolutions cannot be expected to be symmetric in any single variable

- The strategy used is to measure the overall tag rate, then subtracting from it the tag rate due to real heavy jets

$$rate^{mistag}_{RomaNN} = rate^{TotalTag}_{RomaNN} - rate^{b} \times (\epsilon^{b}_{RomaNN;MC} \times ScaleFactor_{RomaNN})$$

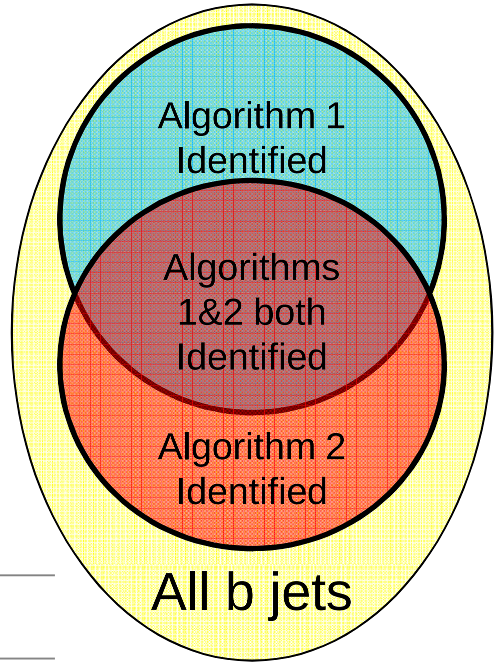
$$rate^{b} = \frac{(rate^{+}_{SecVtx}) - \alpha\beta(rate^{-}_{SecVtx})}{\epsilon^{b}_{SecVtx;MC} \times ScaleFactor_{SecVtx}}$$

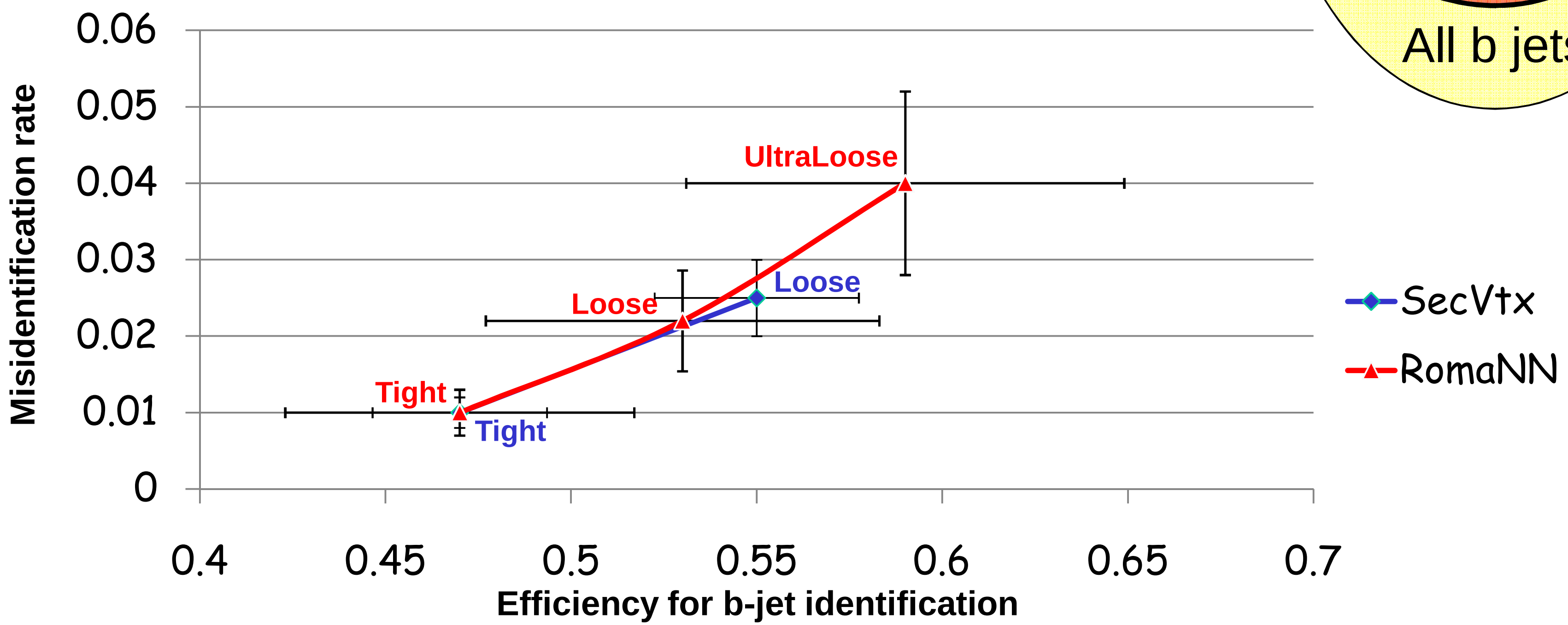


- We characterize the misidentification rate using a matrix
  - based on several of its parameters, such as its energy and location within the detector

b-tagging Performance Comparison

- SecVtx algorithm performs similar to new algorithm
  - Improvements from initial studies by developers in simulation, not seen after calibration
- Use new b-tag in conjunction with default increases signal efficiency
  - CDF WH search summer 2009: addition of UltraLooseRomaNN increased signal acceptance by 20% in the double-tag category

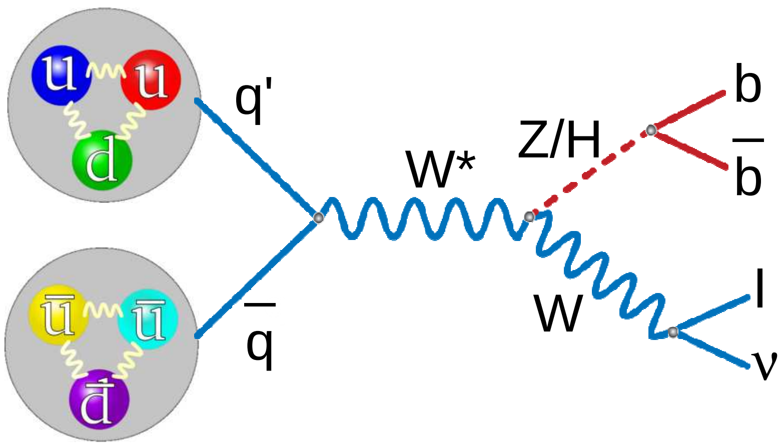




Outline

- Motivation: Higgs Boson Search
  - Importance of b-tagging
  - Comparison of signatures from WZ and WH
- b-tagging improvements
  - Calibration of efficiency
  - Calibration of mis-tag rate
  - Comparison
- Search for WZ→lνbb
  - Selection
  - Discrimination
- Conclusion & Future Plans

WZ/WH Event Selection



WZ→lνbb

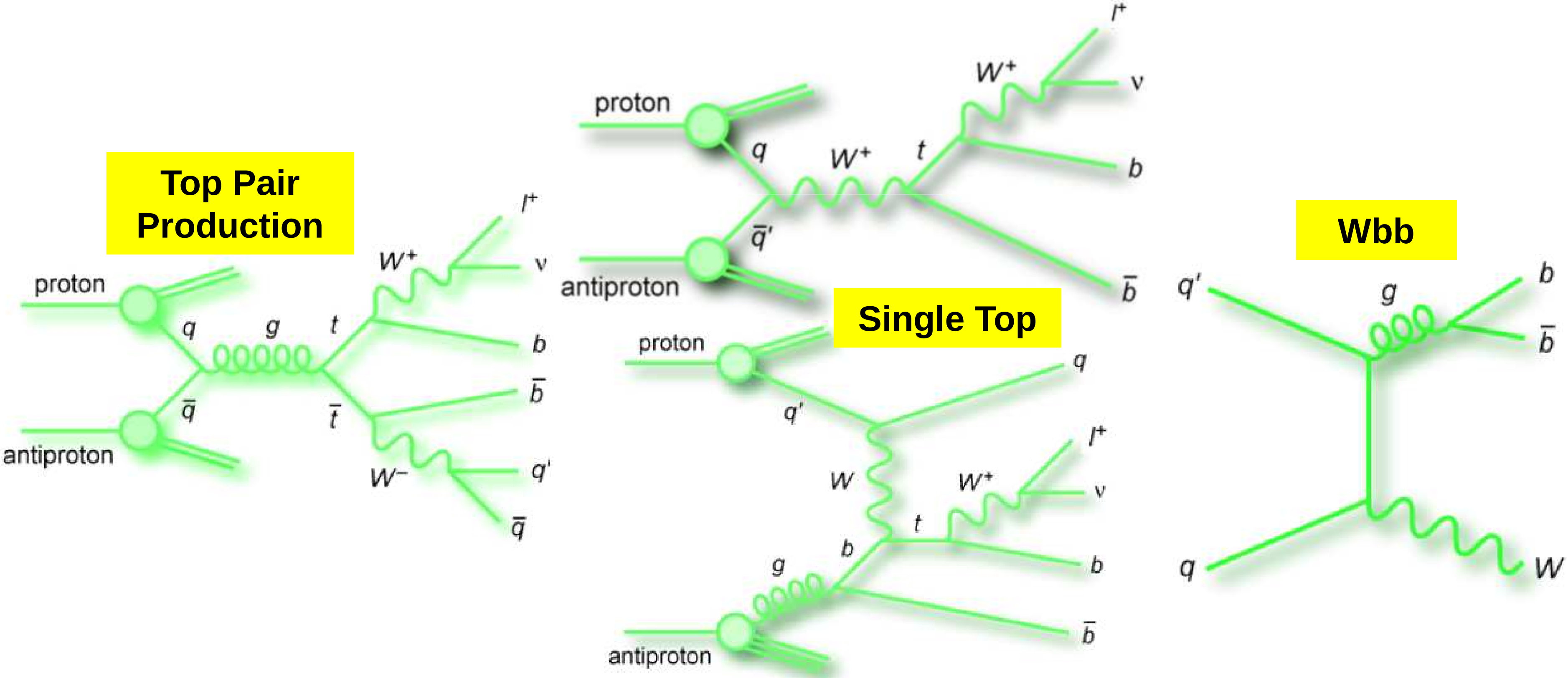
- **High p_T lepton**
  - p_T > 20 GeV
  - |η| < 1.1
- **Missing Transverse Energy**
  - MET > 20 GeV
- **Two *b*-jets**
  - E_T > 20 GeV
  - |η| < 2.0
  - Jet cone size 0.4
  - identification of b-jet

WZ	% efficiency	percent of initial
Fiducial Lepton	60.3	60.3
Lepton E_T>20GeV	84.2	50.8
Reconstructed & Identified	58.7	29.8
MET > 20GeV	88.3	26.3
2 jets, both E_T>20GeV and Fiducial	37.4	9.9
Both identified as b-jet (SecVtx)	11.6	1.1
Both identified as b-jet (RomaNN)	**18.0**	**1.7**

- Identify both jets using UltraLoose tag, increases acceptance 55% over SecVtx selection

lvbb Backgrounds

- b-tagging applied to remove large W+light flavor background
- Unfortunately, many physics processes can lead us to the identification of a lepton, a neutrino, and two b-quark jets
  - 70% is irreducible background
  - top quark pair production, single top quark production, Wbb

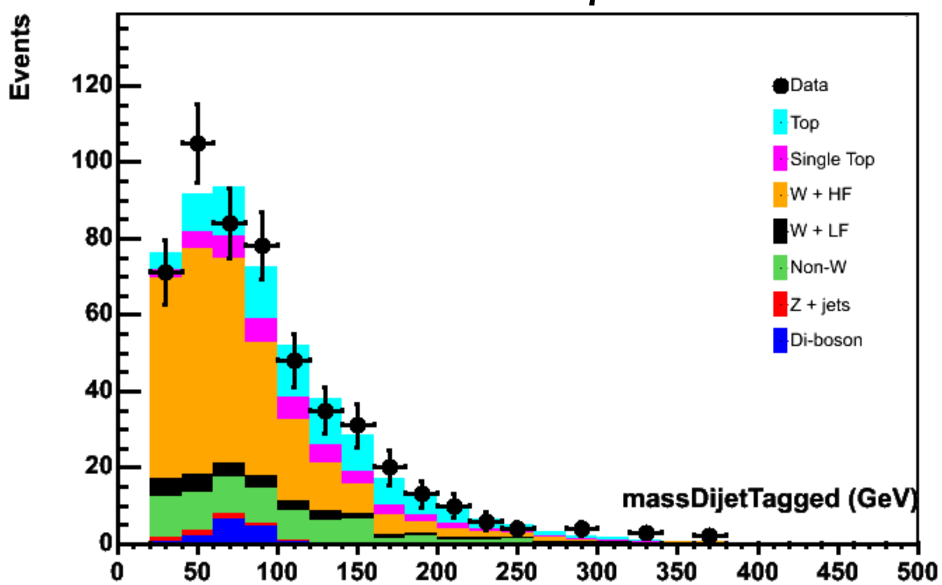


- The other 30% of background: Wcc, Wc, Wlf, Zlf, QCD

Signal and Background Expectation

- Expected WZ yield is 2% of all lvbb events
  - Need to use additional information in the events to distinguish between signal and background
  - ➢ We expect the invariant mass of the two b-jets (M_{bb}) from WZ to resemble the Z mass peak (91 GeV), whereas the background is more diffuse
    - Performing a fit using the entire M_{bb} histogram, the expected 95% Confidence Level limit is 4.0x SM cross section with $4.3fb^{-1}$

2 UltraLoose tags data/MC normalized to equal area

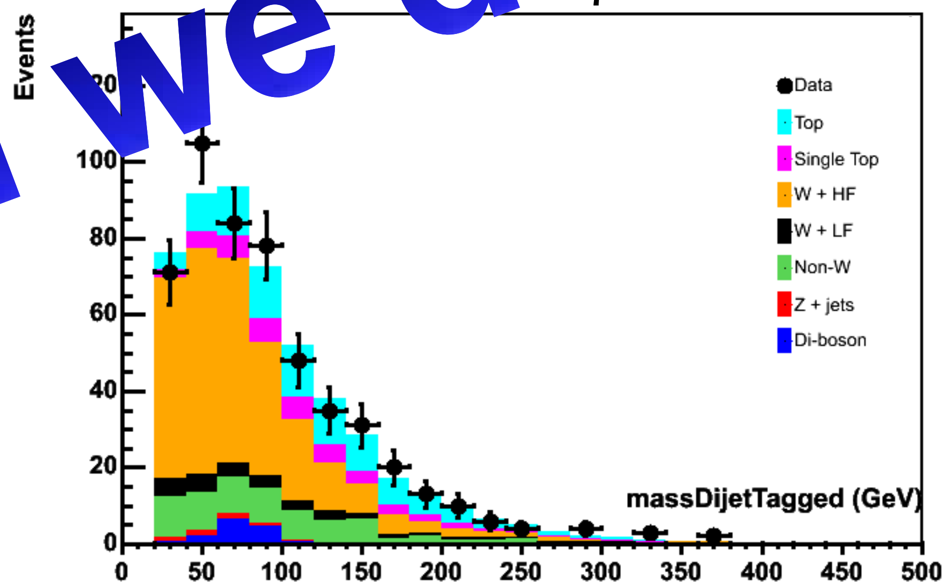


Signal and Background Expectation

- Expected WZ yield is 2% of all lvbb events
  - Need to use additional information in the events to distinguish between signal and background
  - ➢ We expect the invariant mass of the two b-jets (M_{bb}) from WZ to resemble the Z mass peak (91 GeV), whereas the background is more diffuse
    - Performing a fit using the entire M_{bb} histogram, the expected 95% Confidence Level limit is 4.0x SM cross section with 4.3 fb^{-1}

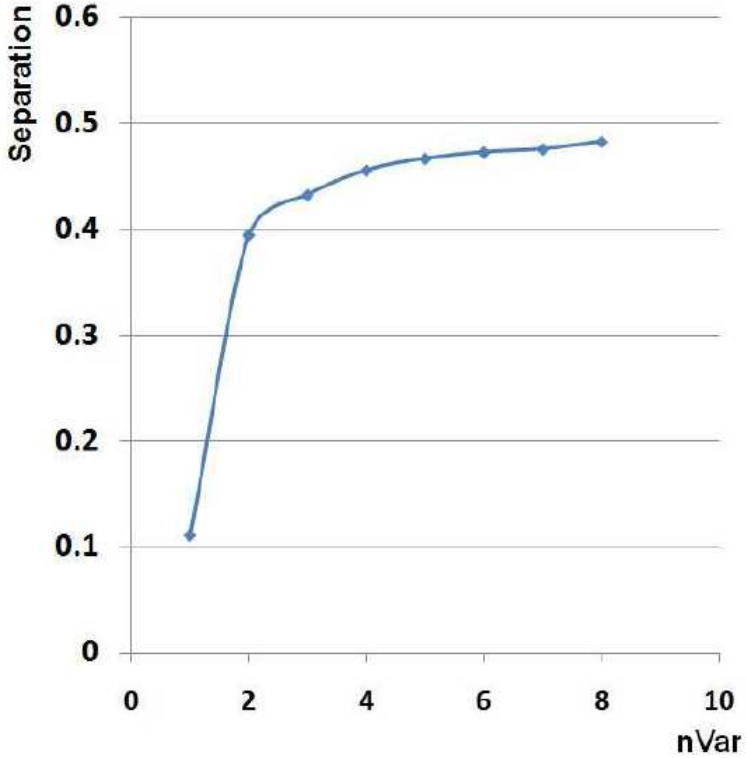
Can we do better?

2 UltraLoose tags data/MC normalized to equal area



Neural Network

- Try to improve sensitivity by using a stronger discriminating variable
  - Construct this using a TMVA Neural Network
    - ➢ Combines information from several variables
- Train and test to search for the best neural network with the fewest input variables



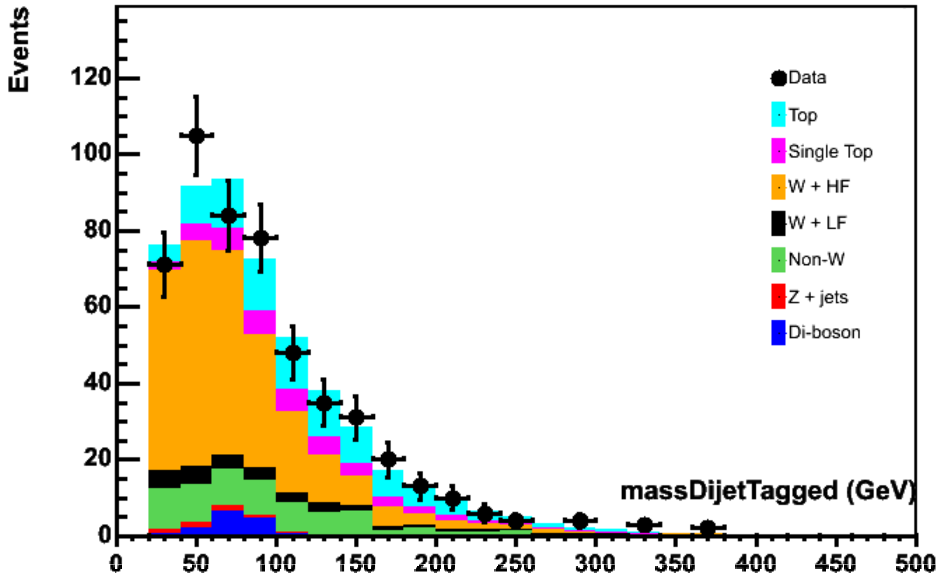
Separation:

$$< S^2 > = \frac{1}{2}\int \frac{(P_S(y) - P_B(y))^2}{P_S(y) + P_B(y)} dy$$

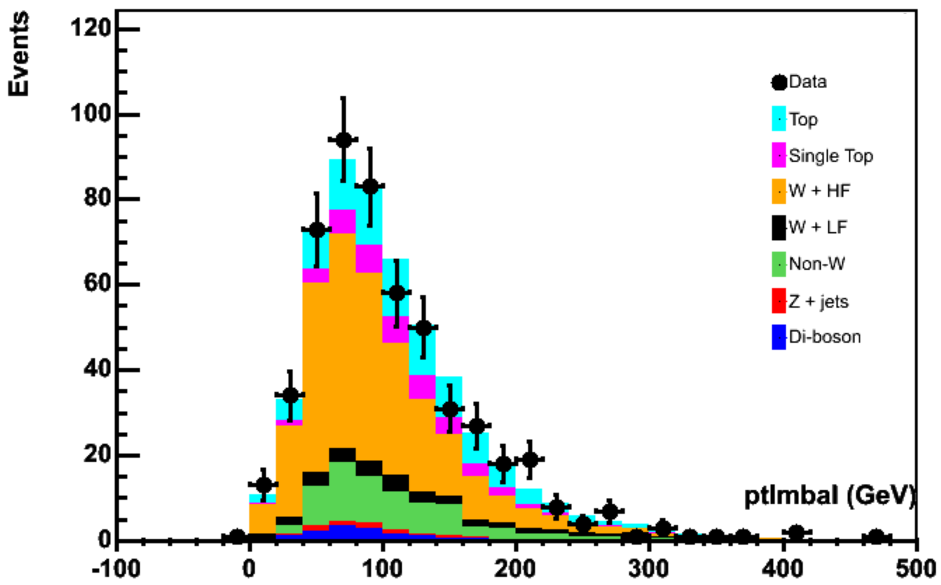
Where P_S, P_B are signal and background probability distributions

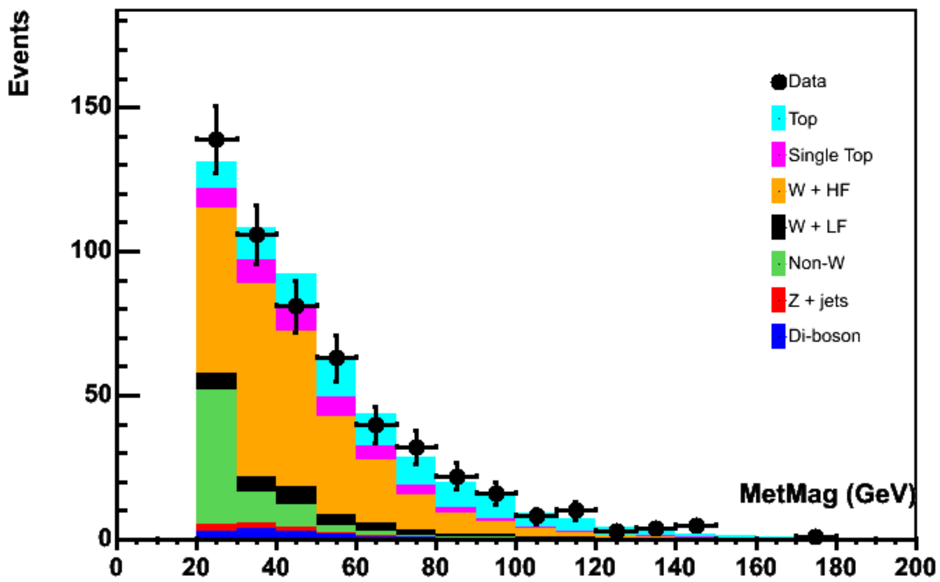
- For the additional complexity involved in using more inputs, not much separation gain after three variables

Input to TMVA Neural Net



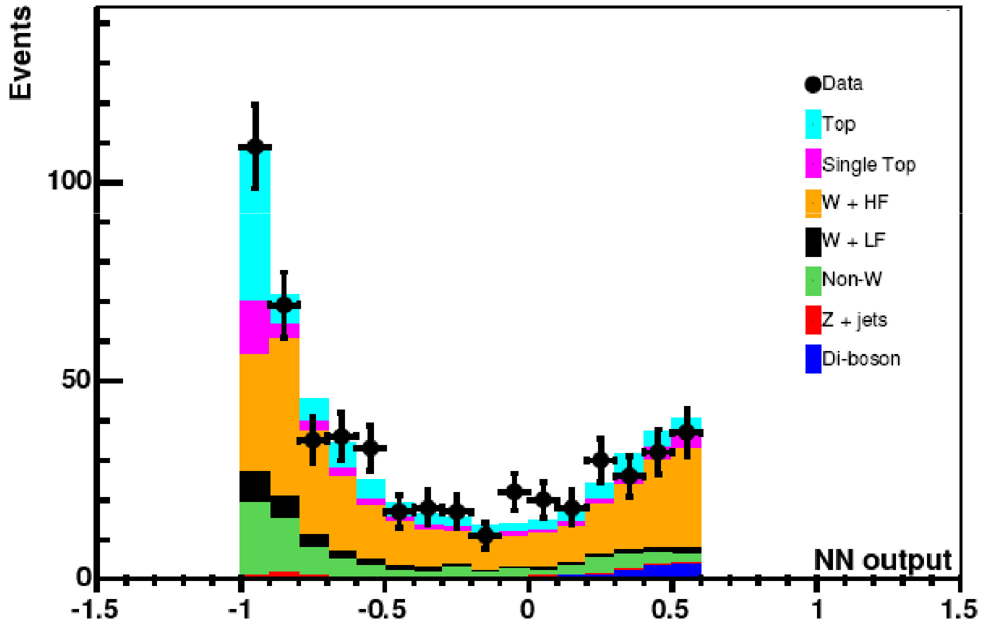
- M_{bb} : the invariant mass calculated from the two jets
- ptImbal : the difference between the scalar sum of the P_T of all measured objects and the MET
    - Specifically, it is calculated as $P_T(jet1)+P_T(jet2)+P_T(lep)-MET$
- MetMag : |MET|





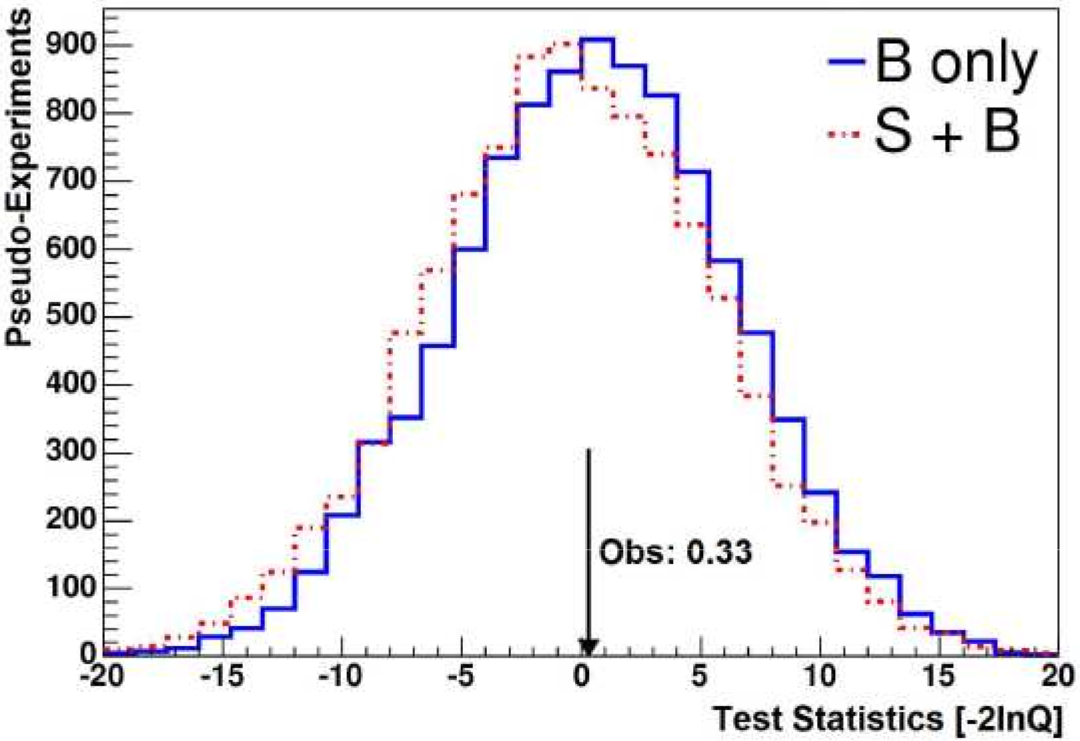
Signal Limit Estimate With Neural Network

- Performing a fit using the Neural Network discriminant, expected 95% Confidence Level limit is now 3.8x standard model cross section
  - Improvement from 4.0x with M_{bb} alone
- Apart from M_{bb}, other kinematical quantities contribute to a 5% improvement in limit



Limit					
-2σ	-1σ	Median	+1σ	+2σ	**Observed**
2.0	2.7	3.8	5.3	7.1	**3.6**

Test Statistics



$$Q = \frac{P(\mathrm{data}|H_1)}{P(\mathrm{data}|H_0)}$$

- H_1 is the model with signal
- H_0 is the null hypothesis

- large overlap between the results for experiments with signal and without signal present
  - This analysis has little sensitivity to the presence of a signal
- P-value of 0.48
- There is 48% probability of the background fluctuating to give a value of -2lnQ lower than 0.33

Summary

- Searched for WZ→lvbb, part of the Higgs search effort
  - Improved b-jet identification
  - Utilized Neural Network to improve signal sensitivity

  - Using 4.3fb^{-1}, expected a 95% Confidence Level limit of 3.8x, and measured 3.6x standard model WZ cross section

  - Will publish in PRD