Search for the Higgs Boson in the $ZH \rightarrow \mu\mu bb$ Channel at CDF Using Novel Multivariate Techniques

Justin R. Pilot
Department of Physics, The Ohio State University

Thesis Dissertation Defense
November 16, 2011

Personal Background

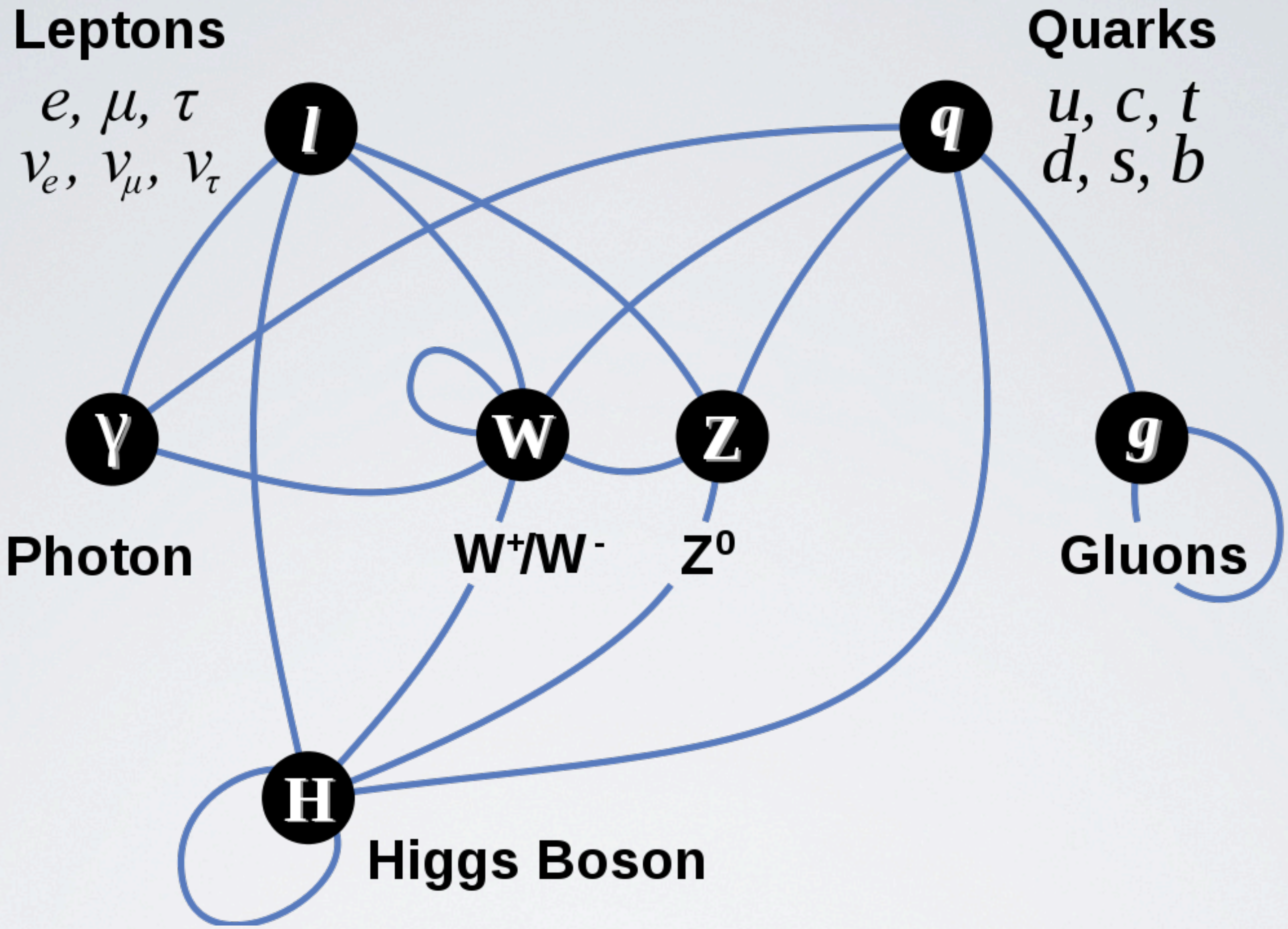
- B.S., Physics -- 2007
  - Michigan State University
  - Started graduate school at OSU that fall
- Passed candidacy exam Summer 2009 (M.S.)
- Involved in CDF activities
  - Detector operations
  - Analysis groups

Outline

- The Standard Model of Particle Physics
  - Open issues
  - The Higgs Boson and Higgs Mechanism
- Higgs boson search strategies
- Experimental Facilities
  - Fermilab
  - The Tevatron
  - The CDF Detector
- Analysis Techniques
  - Object identification
- Event Selection
- Background and Signal properties
- Signal Discrimination Techniques
- Systematic Uncertainties
- Results, Conclusions, and Outlook

Summary

- Theory of the Standard Model does not include particle masses
  - There is a missing ingredient
- The Higgs boson is a particle postulated to explain the origin of particle masses
  - My research focuses on searching for this particle
- The search is performed using the CDF detector at Fermilab
- My research includes several new techniques that greatly improve the search for the Higgs boson
- No discovery was made, so we set limits on the possible number of Higgs bosons seen with the CDF detector



The Standard Model

and Higgs Mechanism

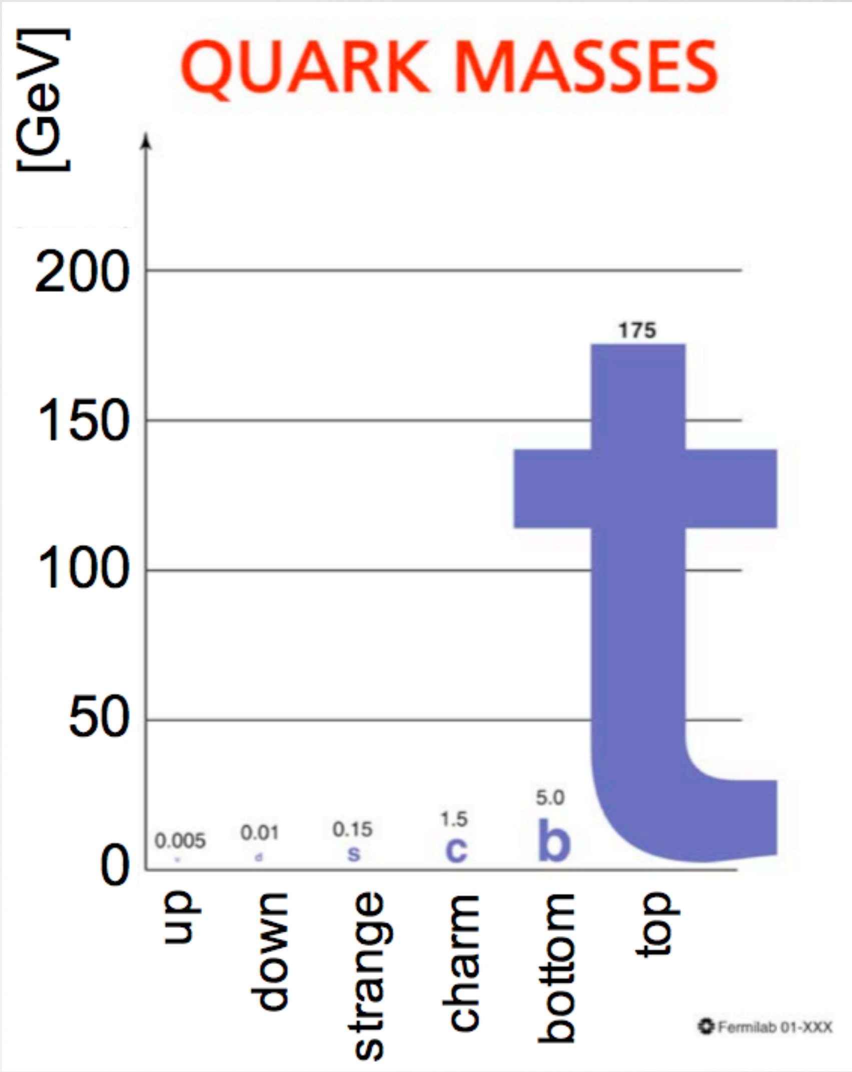
The Standard Model

- Describes motion and interaction of particles
  - Weak, strong, and electromagnetic forces

- Two classes of particles
  - **Fermions**
    - Make up everyday matter
    - Quarks, leptons
  - **Bosons**
    - ‘Force-carrying’ particles responsible for interactions

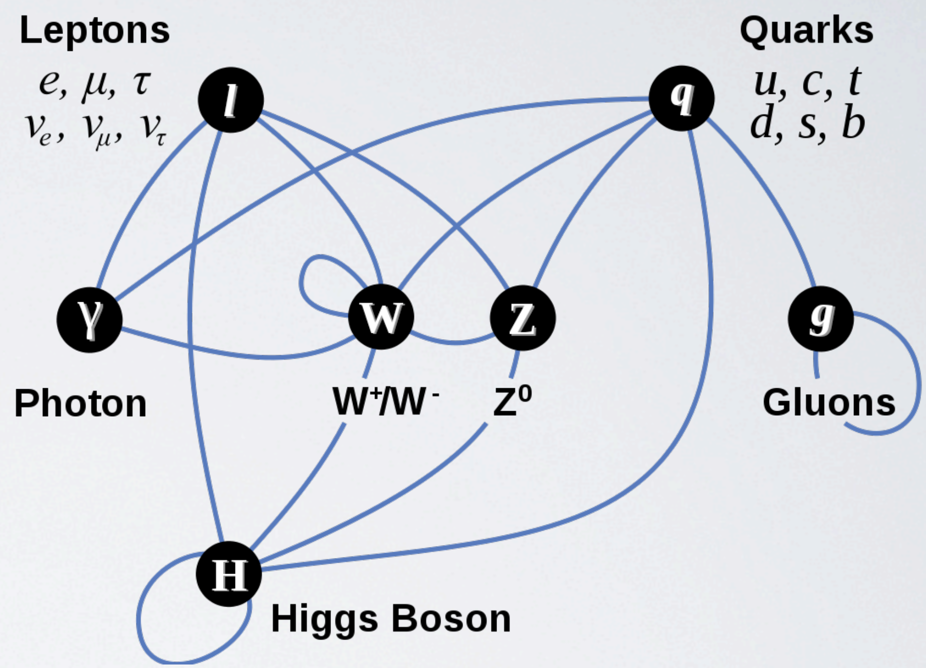
The Problem

- The Standard Model is formulated on symmetries of the universe
  - This gives massless particles
- In nature, we have observed massive particles
  - Why are they massive?
  - How do they ‘acquire’ their mass?
- Higgs boson is postulated to answer these questions



The Higgs Boson

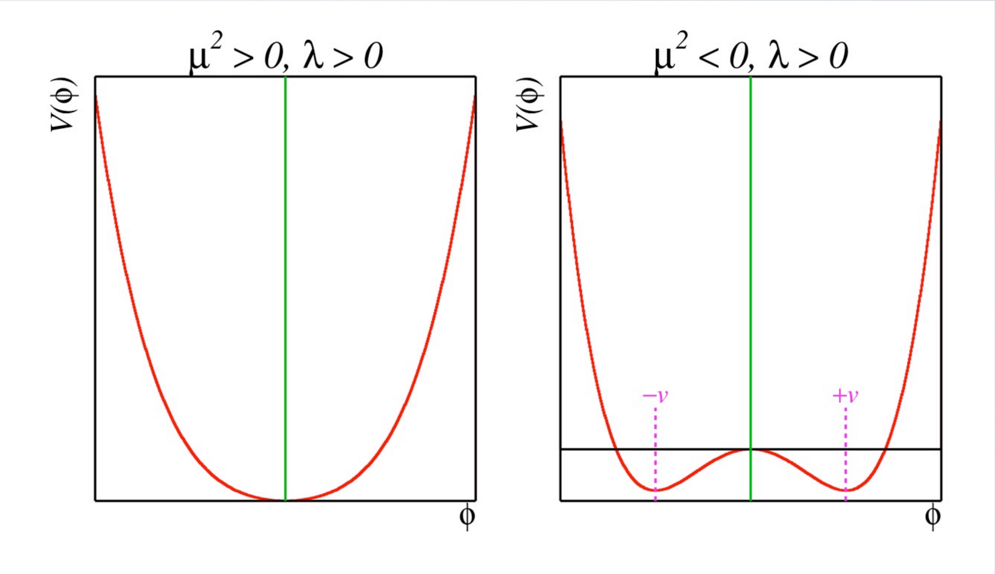
- An additional particle postulated by Peter Higgs and others in 1964
  - Explains origin of particle masses in the SM
- Introduces a new field permeating the universe
  - Higgs field
- Interactions with this field give particle masses



The Higgs Mechanism

- Higgs field contains a symmetric potential function
  - Ground state spontaneously breaks symmetry
  - Non-zero field value permeating the universe
    - Experimental evidence confirms this
- Mass of Higgs boson not predicted by the SM

$$V(\phi) = \mu^2\phi^2 + \lambda\phi^4$$



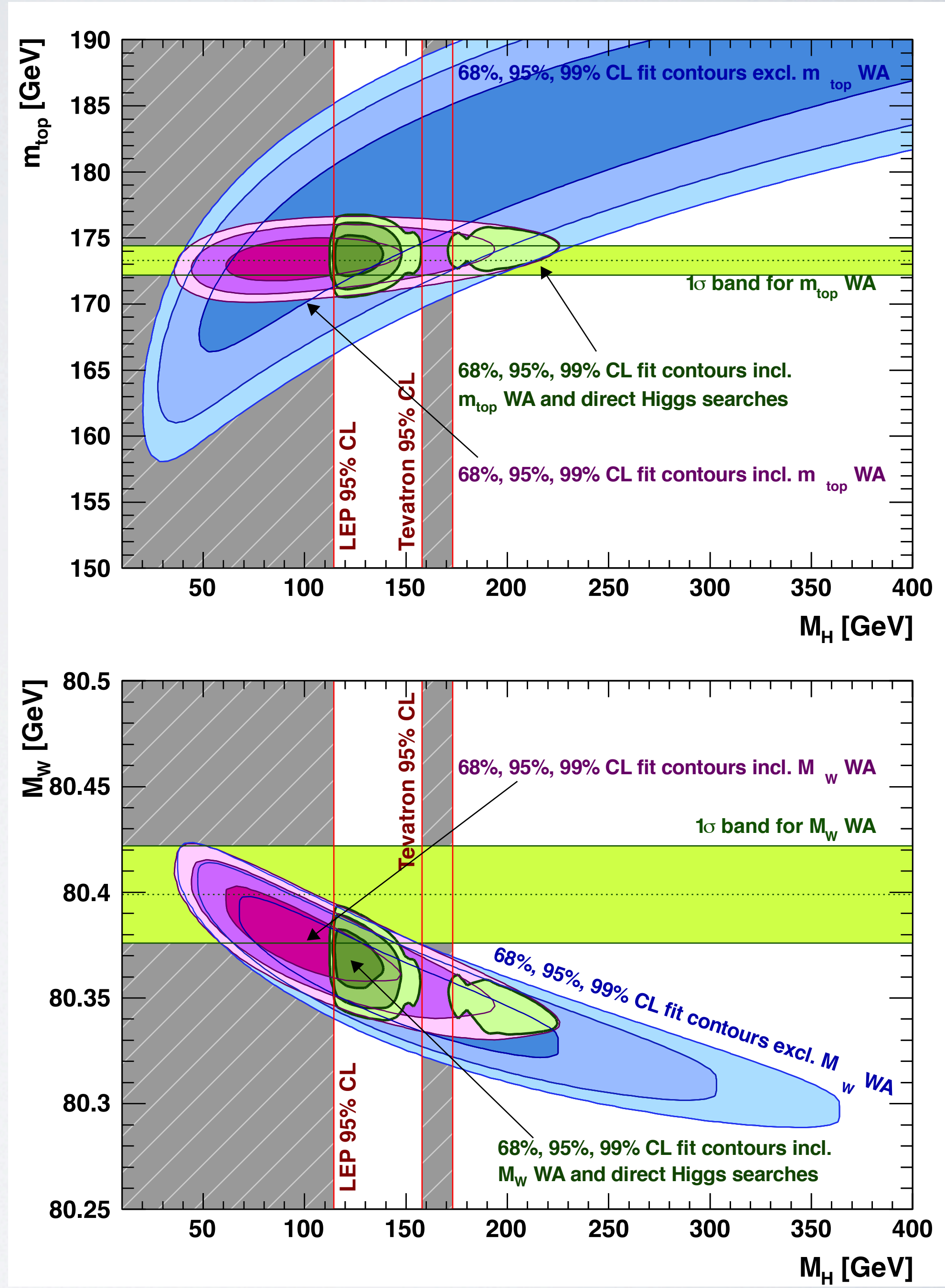
$$v = \sqrt{\frac{-\mu^2}{\lambda}} = 246 \text{ GeV}$$

$$m_H = \mu\sqrt{2} = ???$$

Higgs Boson Search Strategy

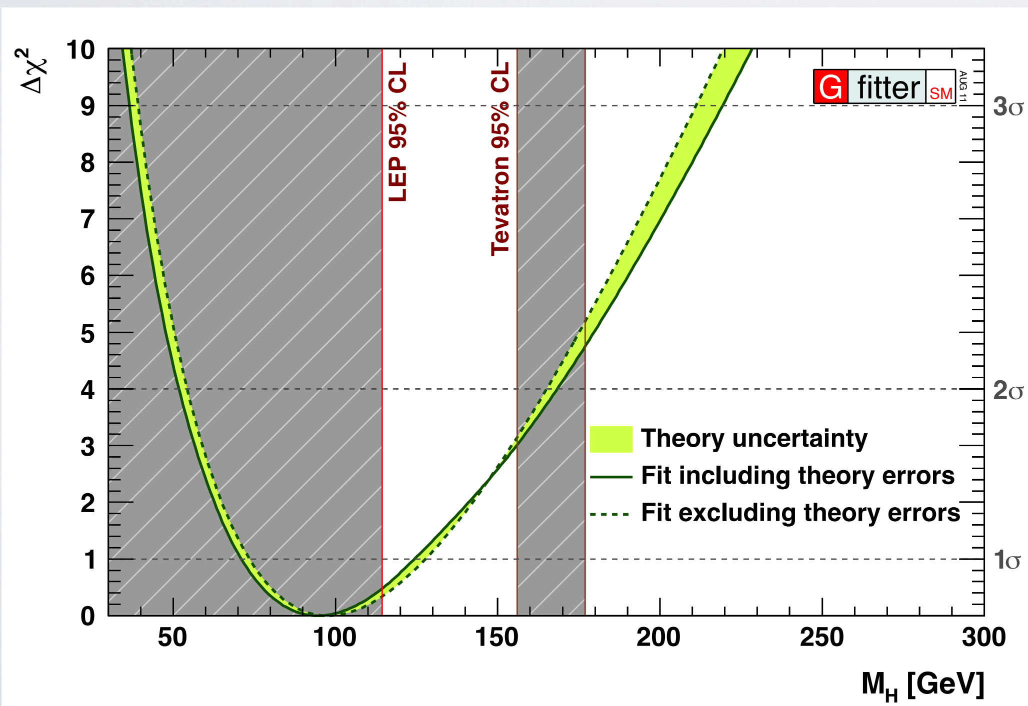
What We Know

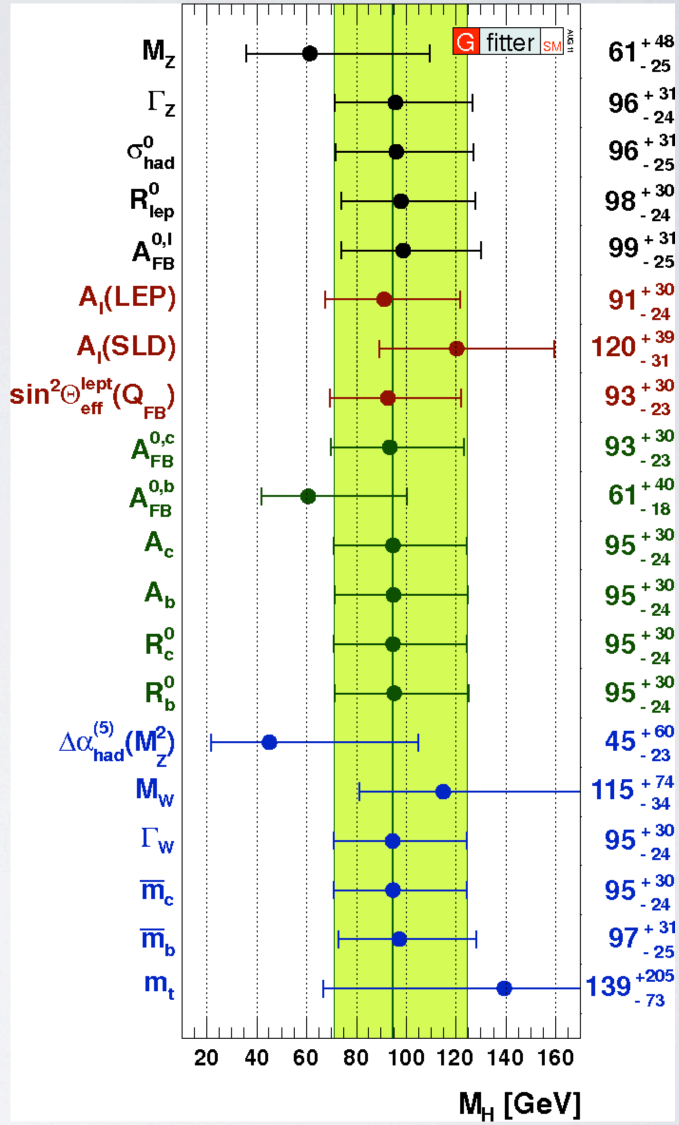
- Mass of Higgs boson not predicted by SM
  - There still are many constraints that can be determined
- Measurements of other SM parameters can determine most likely range of Higgs masses
  - Two important measurements
    - Top quark mass
    - W boson mass



Furthermore

- Can use all measured parameters of the SM to fit for most likely Higgs mass
- From this fit, it appears the Higgs is likely to have a mass ~100 GeV





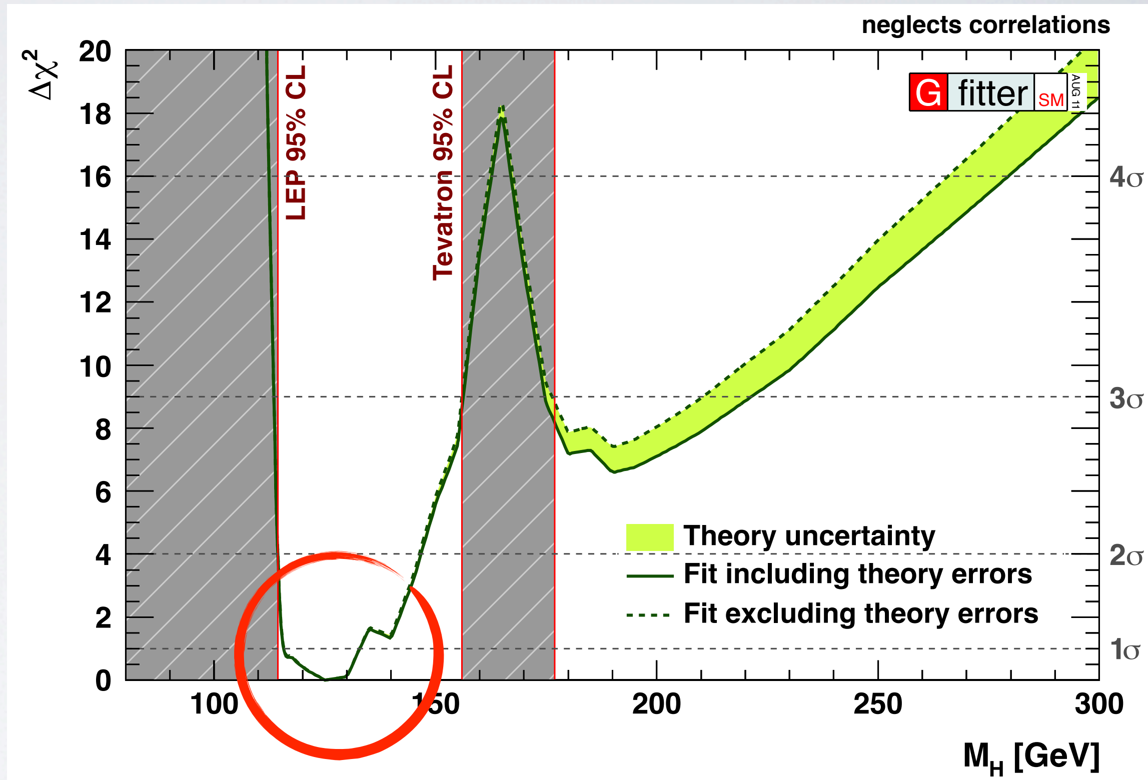
Direct Experimental Searches

- There have been previous direct searches for the Higgs
  - Some masses already excluded

- LEP experiment
  - $m_H > 114.4$ GeV

- Tevatron experiments (CDF+D0)
  - m_H excluded in [156,177] GeV

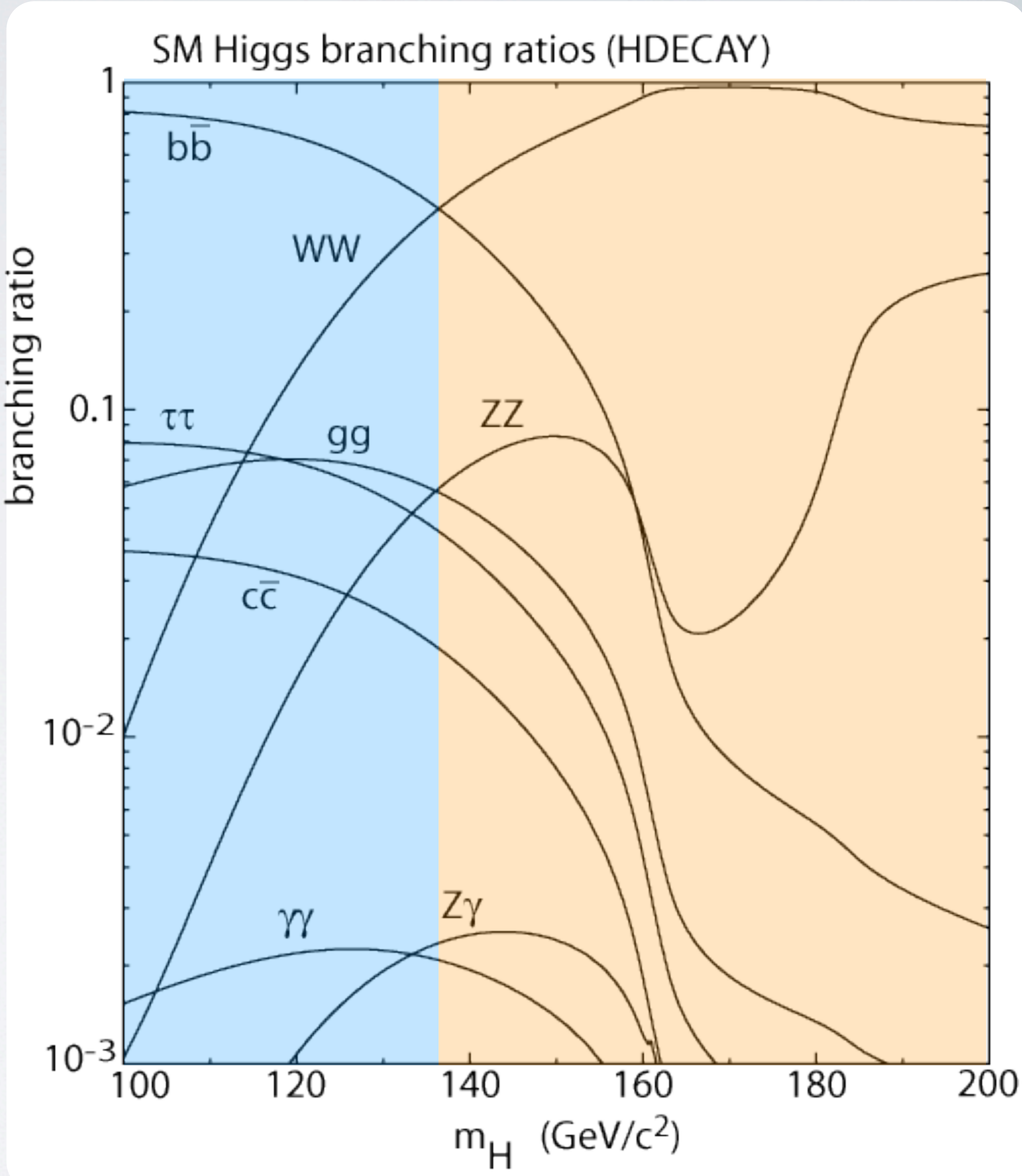
- LHC experiments exclude $m_H >$ 145 GeV



Higgs should be ~115-130 GeV !

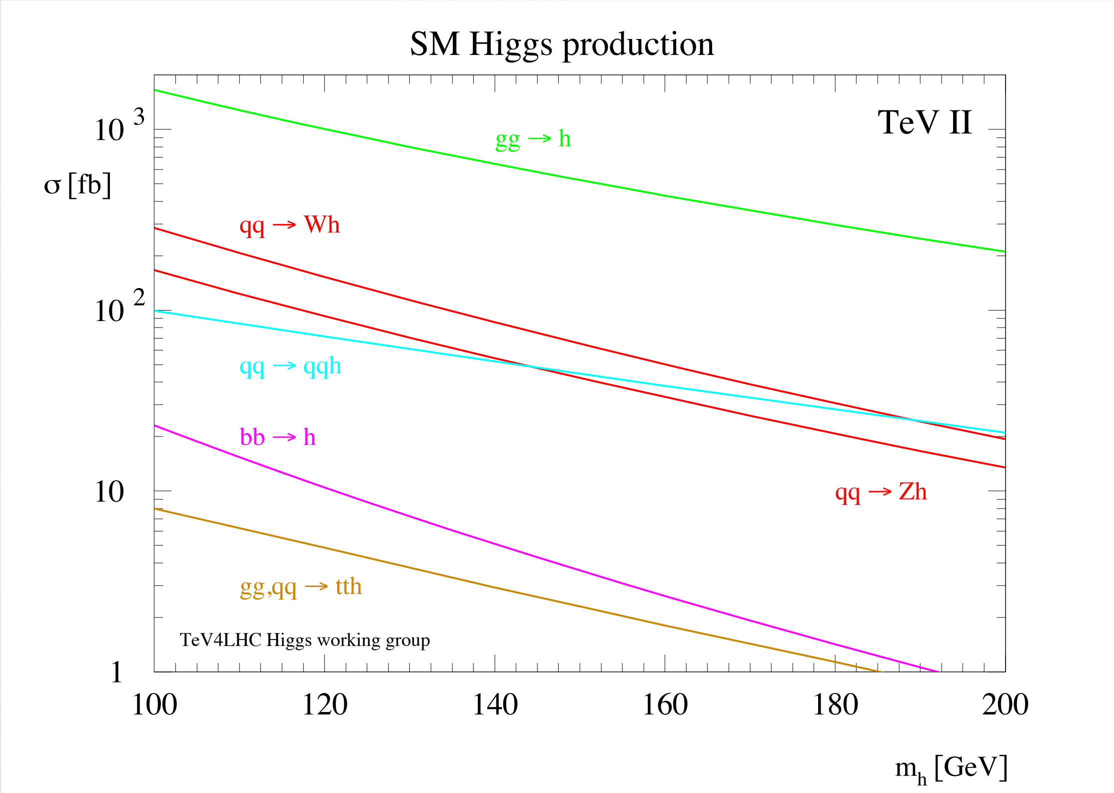
Higgs Searches at the Tevatron

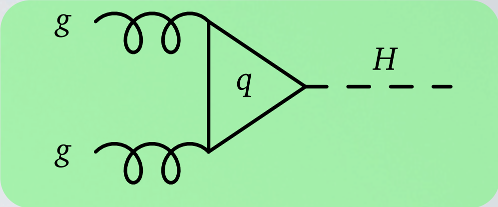
- Divided into two regimes
  - Low-mass
    - < 135 GeV/c^2
    - H→bb is dominant decay mode
    - H→γγ, H→ττ contribute
  - High-mass
    - > 135 GeV/c^2
    - H→WW is dominant, H→ZZ also contributes

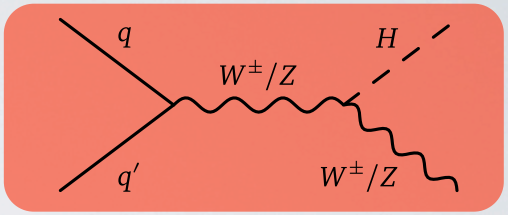


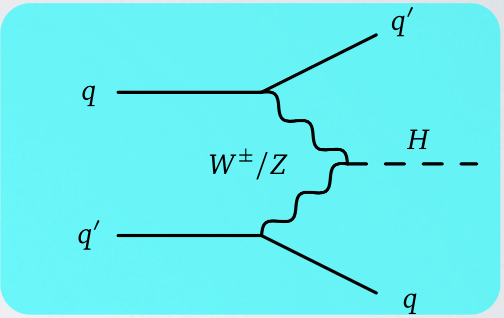
Higgs Production Mechanisms

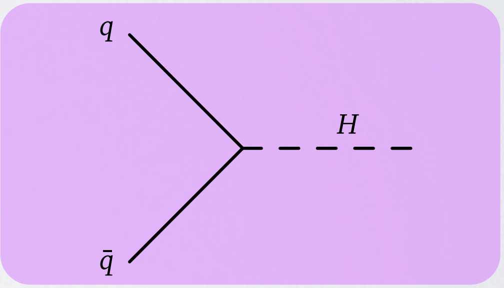
- Higgs boson can be produced by a variety of mechanisms
  - Each has advantages and disadvantages when combined with decay modes
    - Optimize for high/low mass search

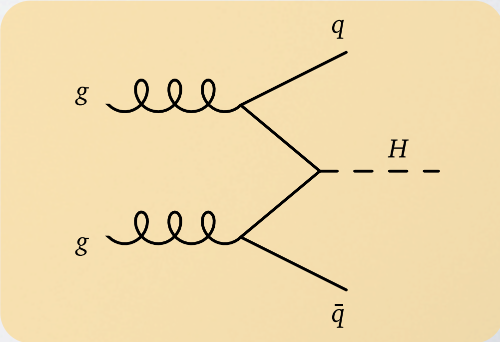






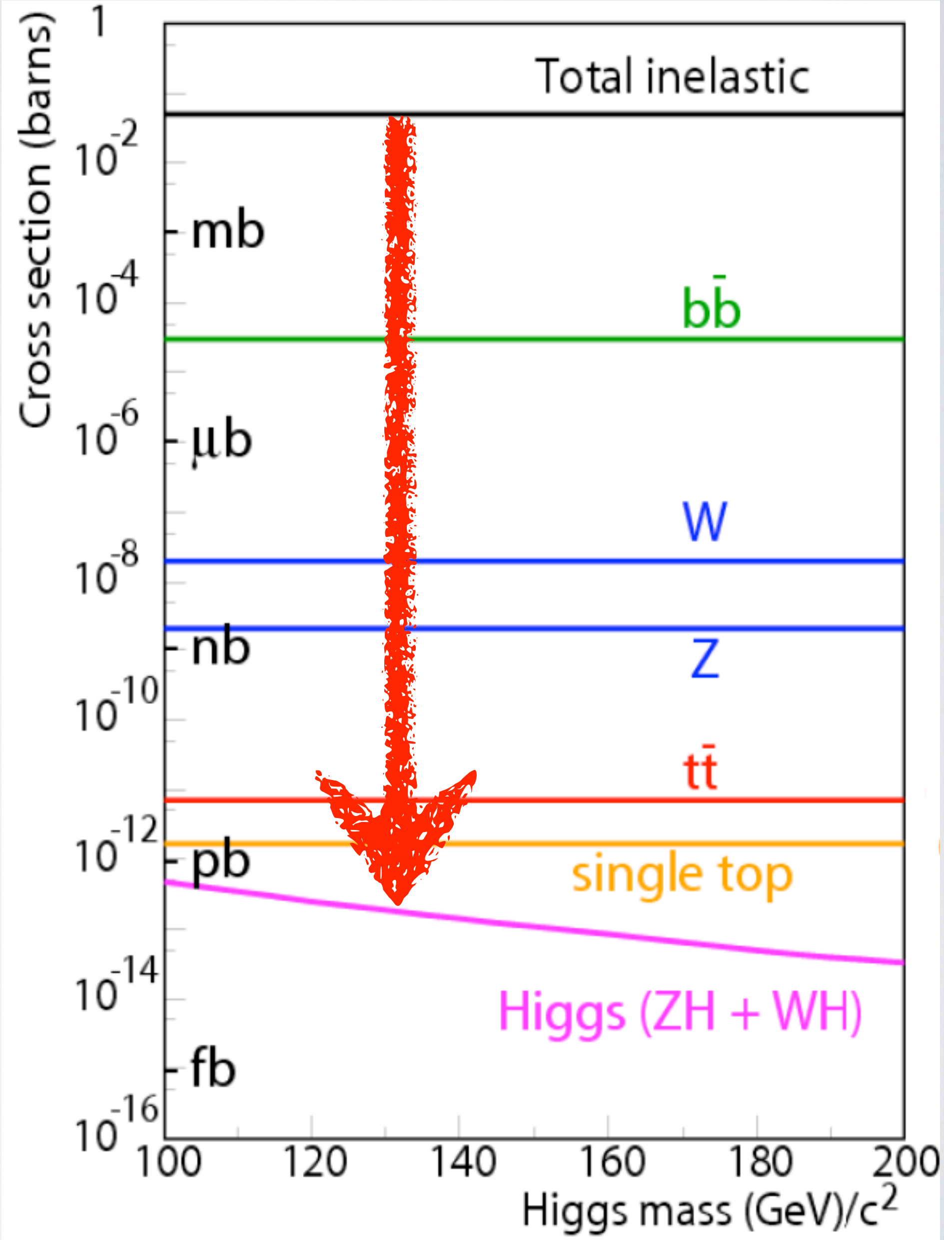




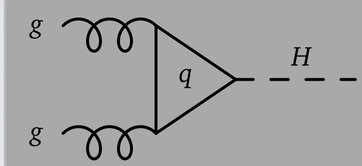
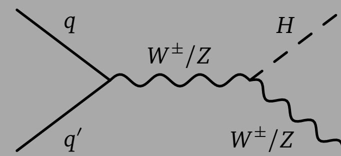
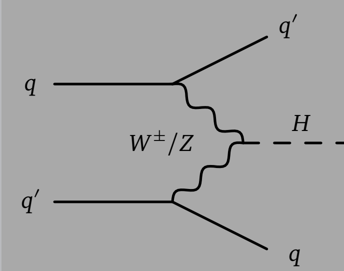
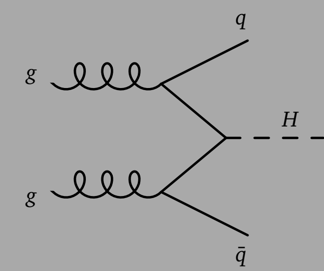


Process Comparison

- Higgs signatures have similar characteristics to other SM processes
  - W, Z production
  - *b*-quark production
- It is a challenge to distinguish the Higgs signal from the overwhelming SM background
- Must choose appropriate combination of production and decay



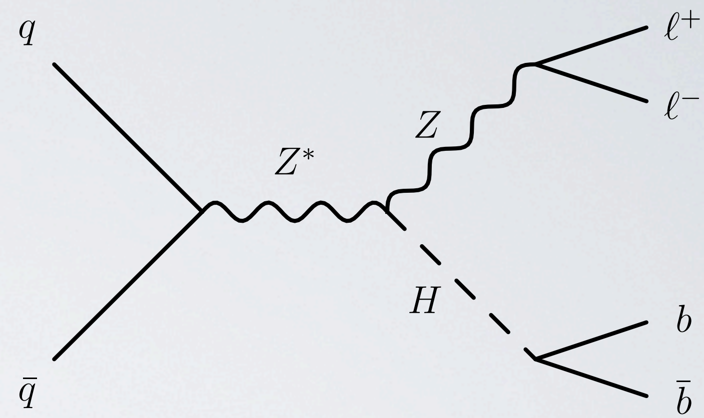
Production and Decay Modes

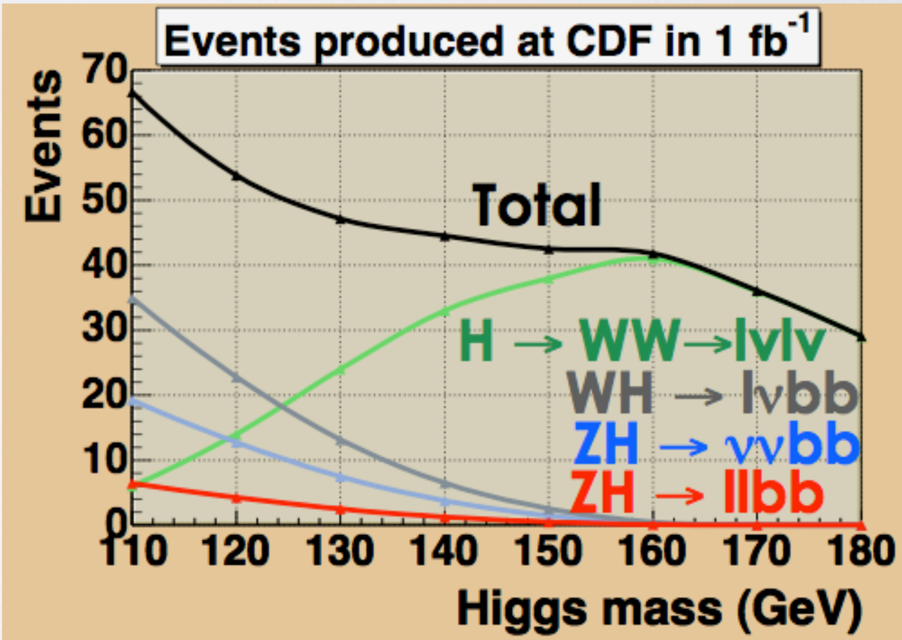
	Gluon Fusion	W/Z Assoc.	Vector Fusion	Di-top Assoc.
				
H→bb LOW MASS	QCD background	Charged lepton(s)	QCD background	Large jet multiplicity; more b-jets
H→WW HIGH MASS	Charged lepton(s)	Low cross section	Charged lepton(s)	
H→γγ LOW MASS	Isolated photons	Low cross section and branching ratio		
H→ττ LOW MASS	Charged lepton(s) Low branching ratio			

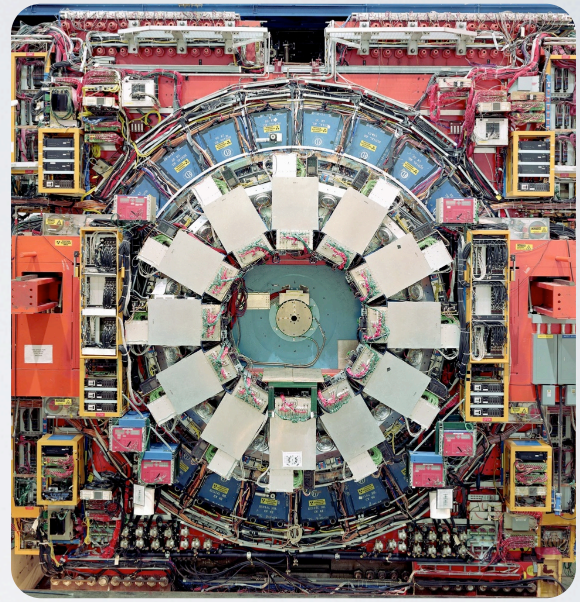
My Analysis Channel

- My research focuses on searching for the Higgs in the associated production channel
  - Z boson decaying to two muons
  - Higgs boson decaying to a pair of b-quarks

- Low signal yield but also low background contribution







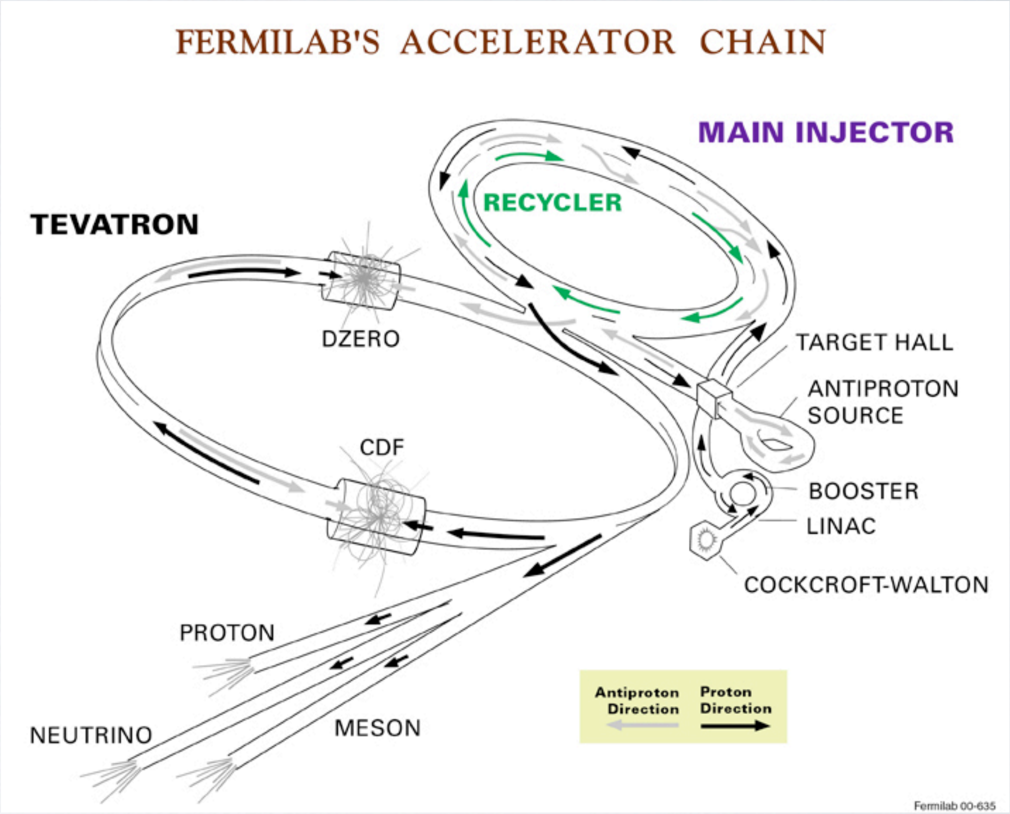
Experimental Facilities

Fermilab, the Tevatron, and CDF

Fermilab

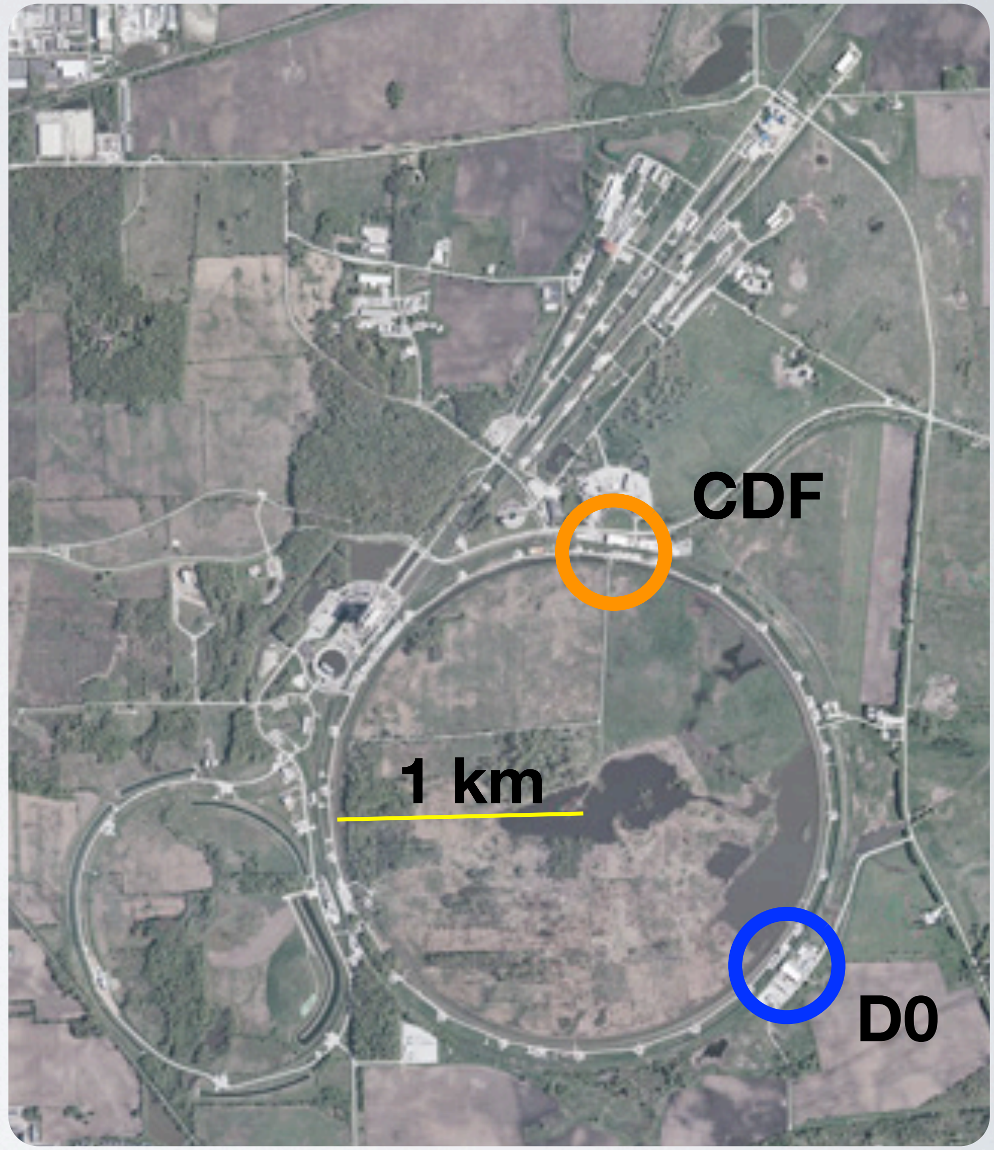
- Nation's premier particle physics laboratory
  - Located ~40 miles West of Chicago, IL
  - Many discoveries through the years: top, bottom quarks, tau neutrino

- Complex of many accelerators used to produce and collide proton and antiproton beams
  - Largest is the Tevatron



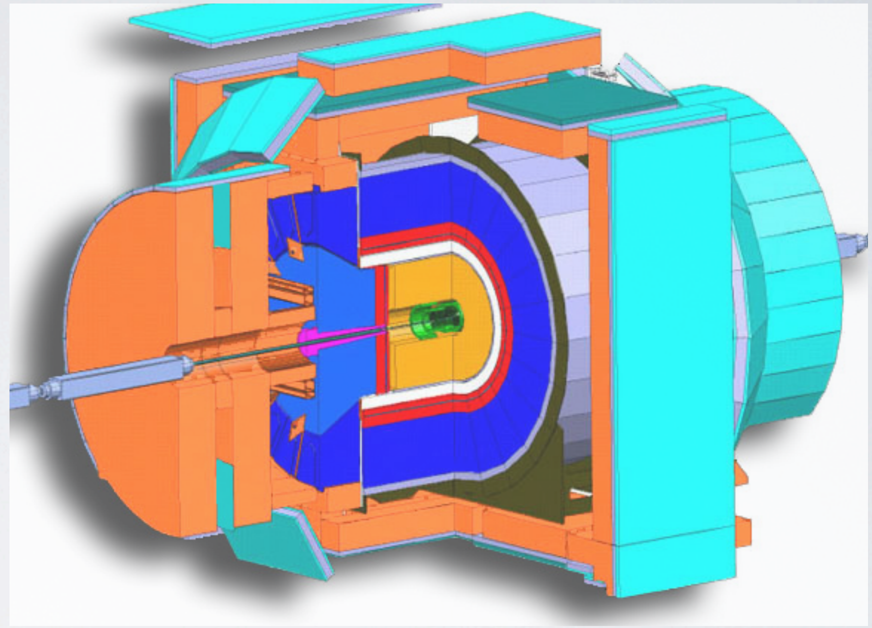
The Tevatron

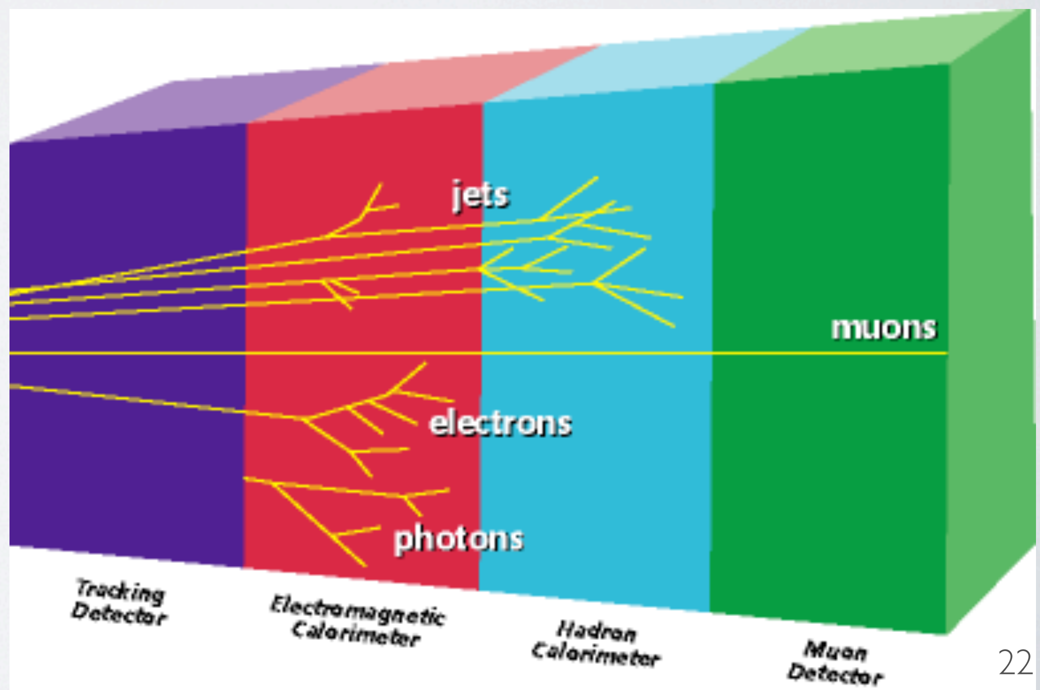
- Proton-antiproton collider
  - ~4 miles in circumference
  - Center-of-mass energy 1.96 TeV
  - Crossing rate 396 ns
- Two collision points, each with general-purpose detectors
  - CDF, D0
- Ended operations September 30, 2011
  - Still much to do!



CDF Detector

- A general purpose detector built to study many types of physics processes
- High-precision silicon tracking system
  - Full coverage for $|\eta| < 2.0$
- Gas-filled wire drift chamber for charged particle tracking
  - Full coverage for $|\eta| < 1.0$
  - Partial coverage for $|\eta| < 2.0$
- Calorimeters
  - EM -- Lead absorber + scintillator
  - HAD -- Steel absorber + scintillator
- Muon detectors
  - Scintillators and wire drift chambers



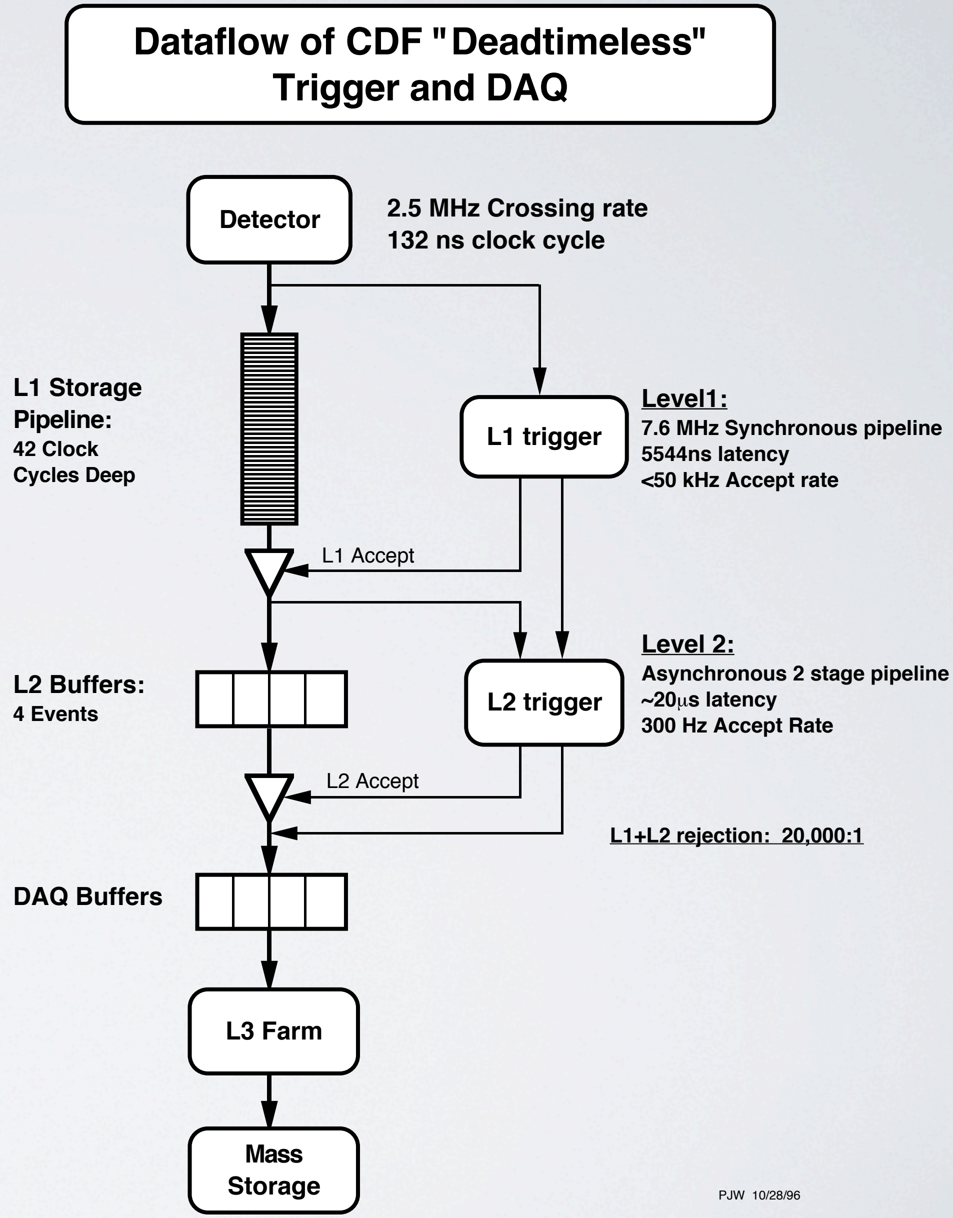


Data Acquisition

- Cannot take every event
  - Must utilize real-time selection
  - "Trigger" system

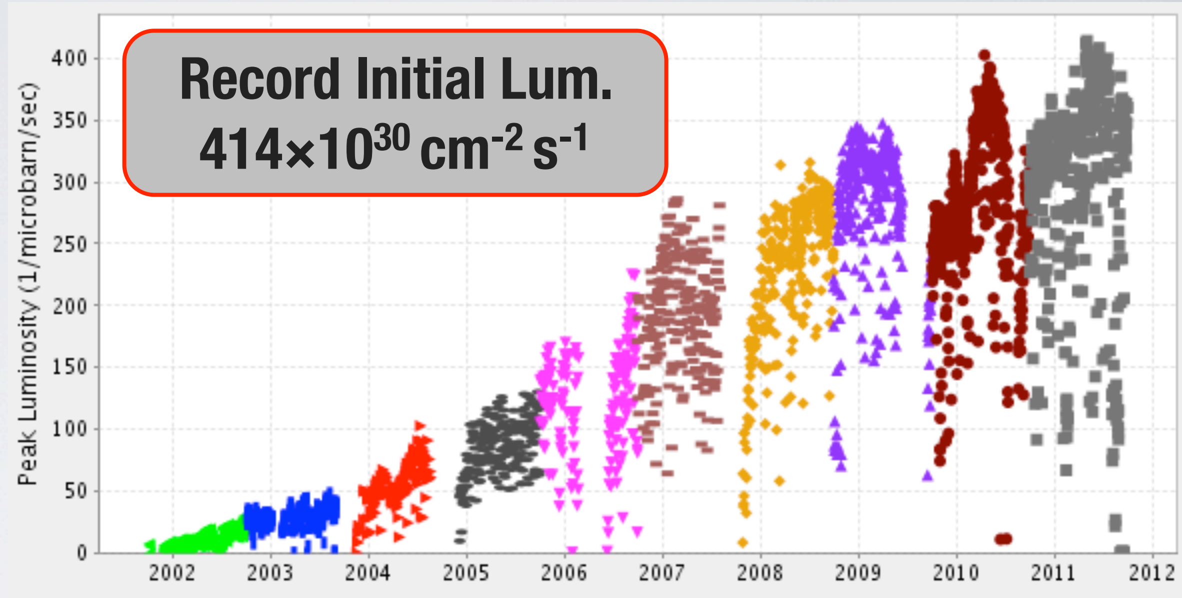
- CDF has a 3-level trigger system
  - 2.5 MHz crossing rate
  - ~100 Hz output to tape for final storage

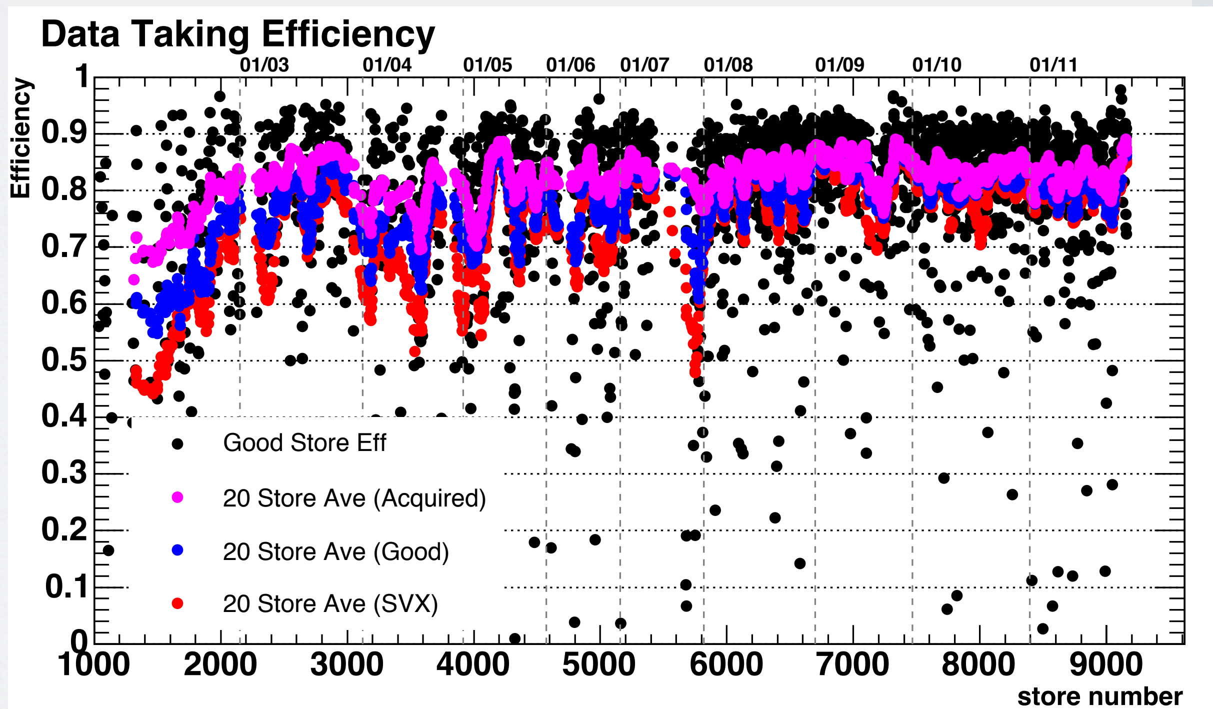
- Choose interesting events to save
  - Leptons, MET, jets



Experimental Performance

- Tevatron delivered ~12 fb^{-1}
- CDF has maintained high data-taking efficiency
  - ~10 fb^{-1}
- Large dataset useful for searches with small expected signal
  - Higgs !!

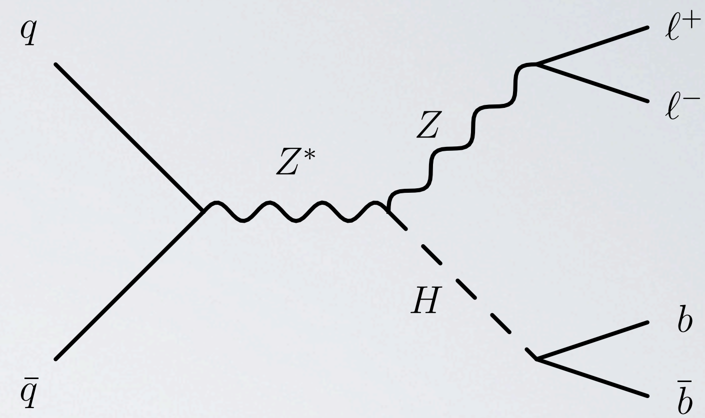


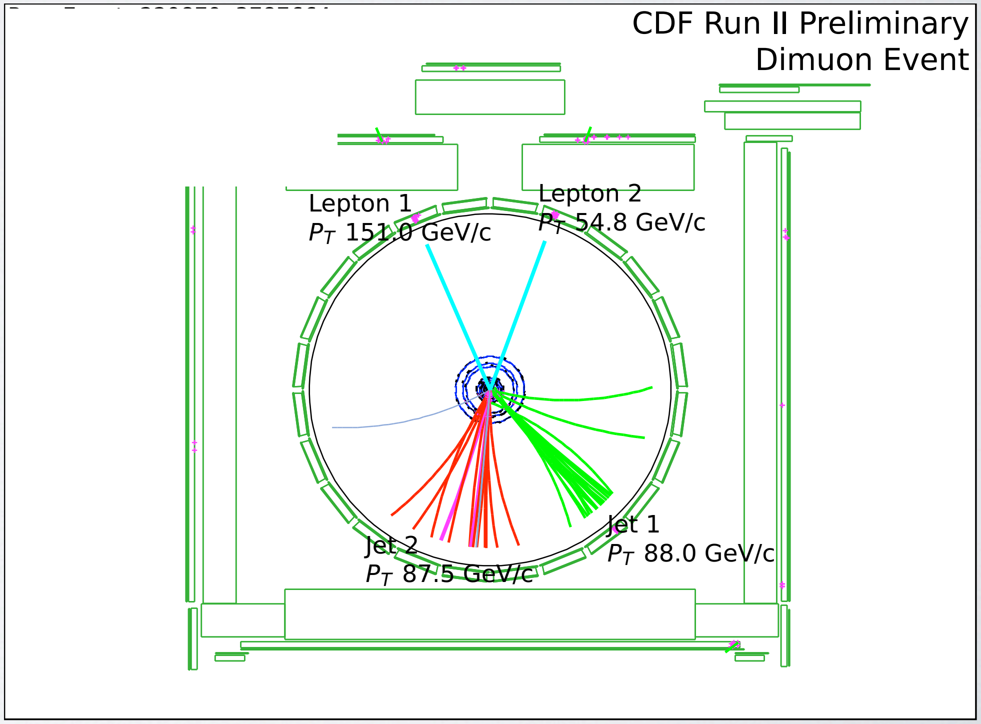


Analysis Methods

Event Selection

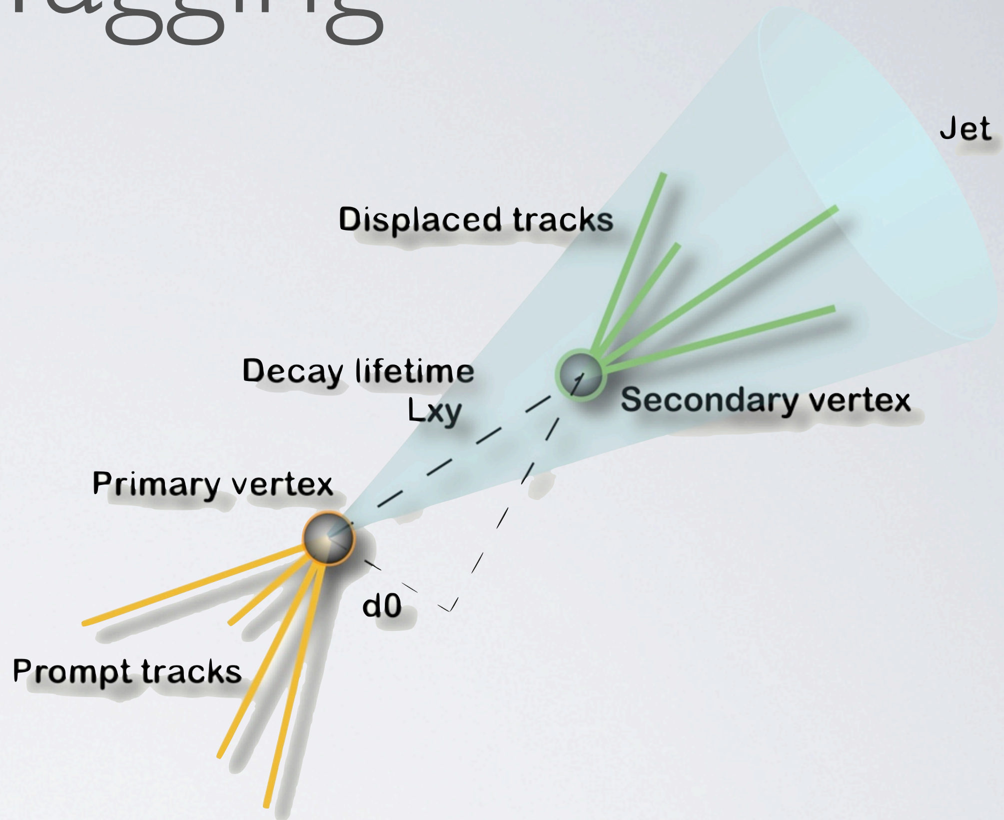
- Select events consistent with the expected final state
  - 2 oppositely charged muons
    - Invariant mass in the 'Z window' -- [76,106] GeV
  - 2 or more jets
    - Leading jet $E_T > 25$ GeV
    - Other jets $E_T > 15$ GeV
    - Also look for '*b*-tags'



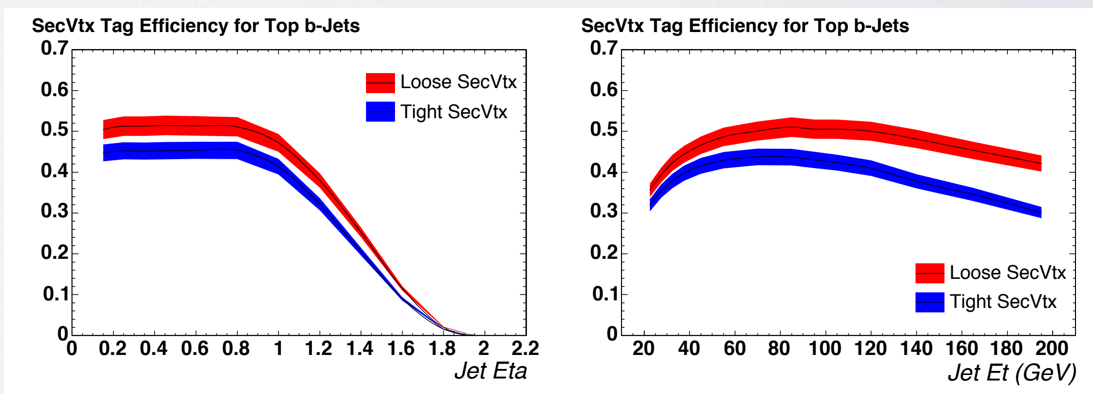


b-jet Tagging

- *b*-quarks hadronize and travel some distance before decaying
  - Gives a displaced vertex
  - Reconstructing a displaced vertex can signify a *b*-quark

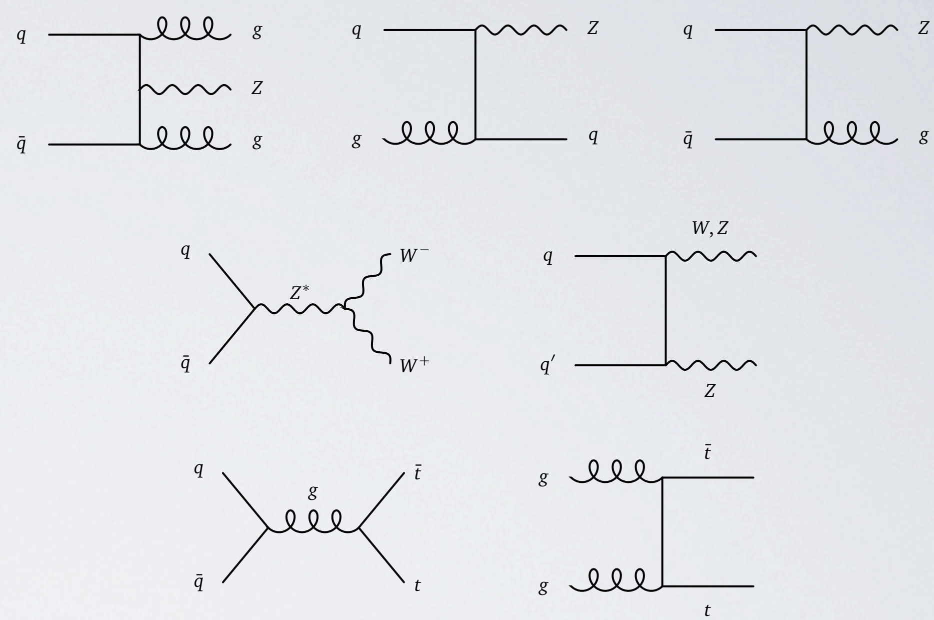


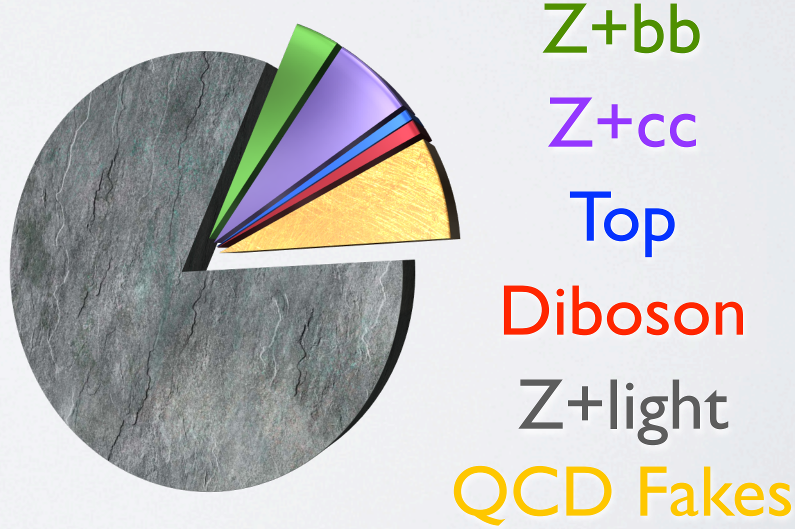
- This analysis uses two algorithms
  - Secondary vertex algorithm
    - Searches for a displaced vertex
  - Jet probability algorithm
    - Uses tracks in jets to form likelihood of originating from primary vertex



Background Composition

- Several processes have same final state
- **Monte Carlo simulations used for:**
  - Z+light flavor jets
  - Z+heavy flavor jets (*b*, c)
  - Top-pair production
  - Diboson (*ZZ*, *WZ*, *WW*)
- **Data-derived method used for:**
  - QCD fake muon production
    - Select two same-sign muons in the Z window



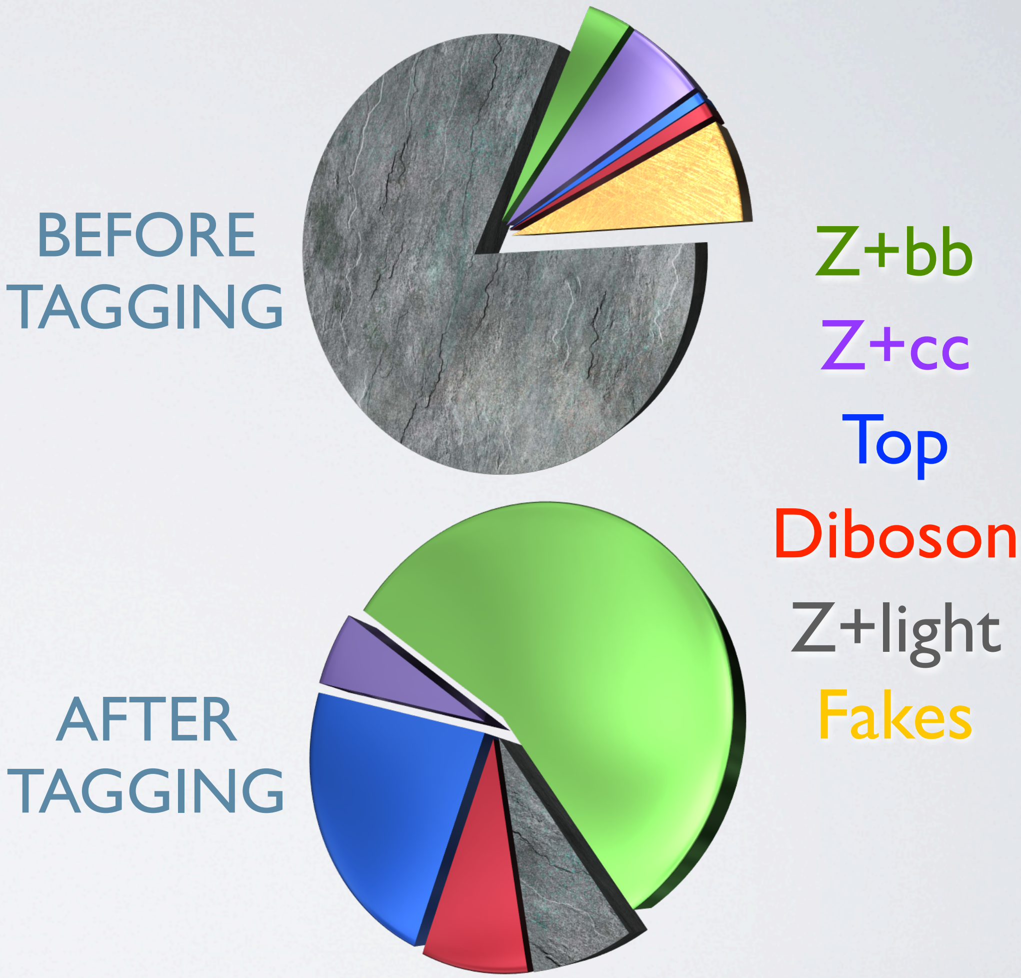


BEFORE TAGGING

Background Composition

- Requiring *b*-tags can remove majority of Z+light flavor background
- We use categories with both 1 and 2 *b*-tags
- As the Higgs signal has 2 *b*-quark jets, the S:B ratio dramatically improves!

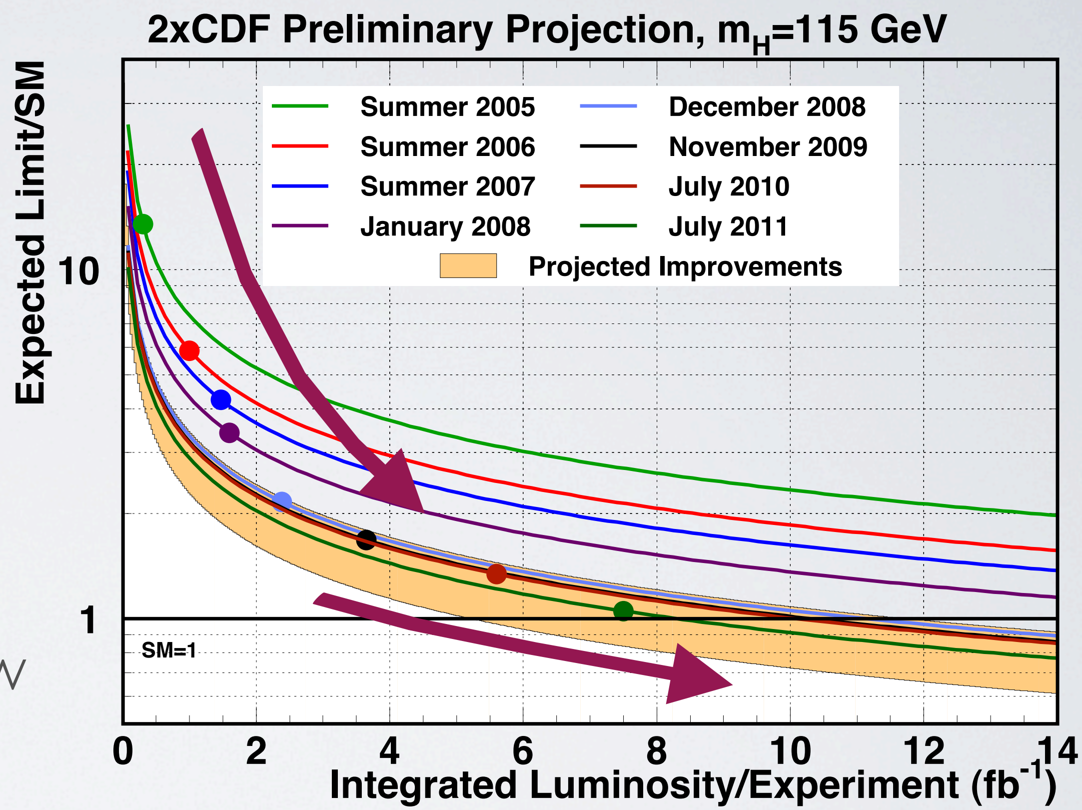
Selection	S : B
None	1 : 10^{12}
Z + ≥ 2 jets	1 : 10,000
+ 1 b-tag	1 : 500
+ 2 b-tags	1 : 50



Focus on getting even better!

How can S:B be improved ?

- Many techniques have been incorporated into the analysis over time
  - Harder to gain sensitivity
- Several strategies
  - Increase acceptance
  - Enhance signal discrimination
- I have incorporated several new techniques to this channel
  - Many multivariate methods
  - Large gains in sensitivity

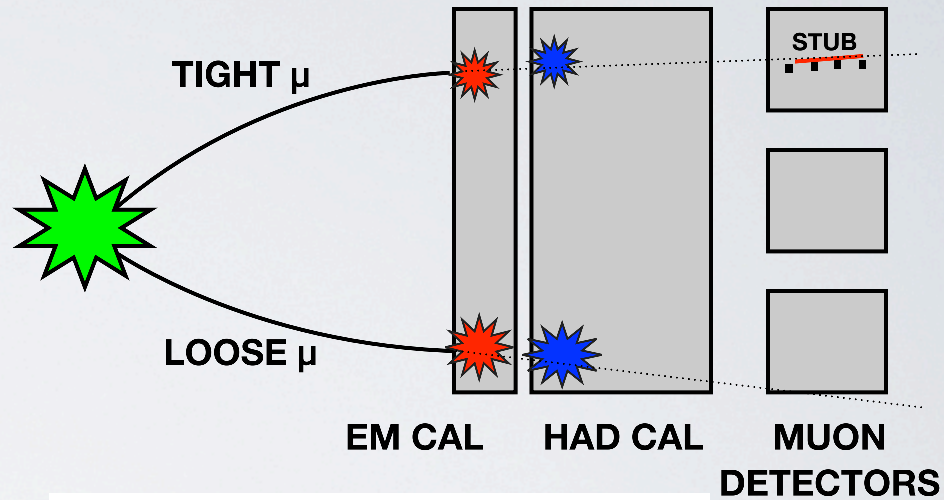


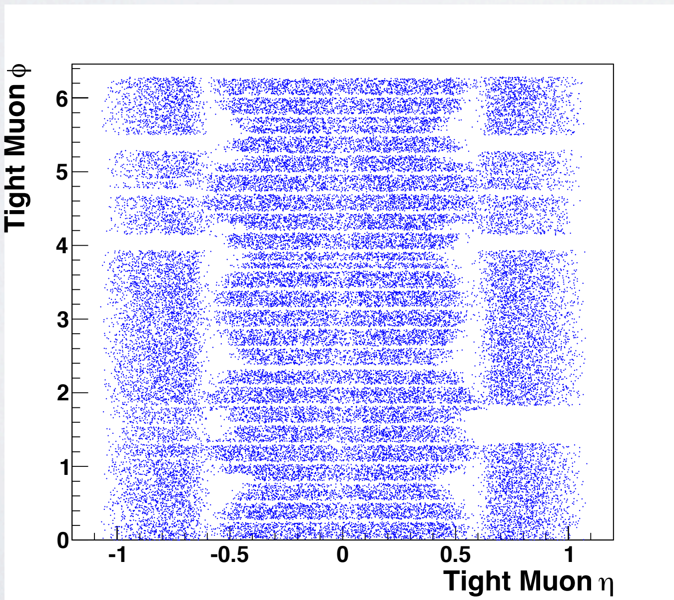
Muon Acceptance

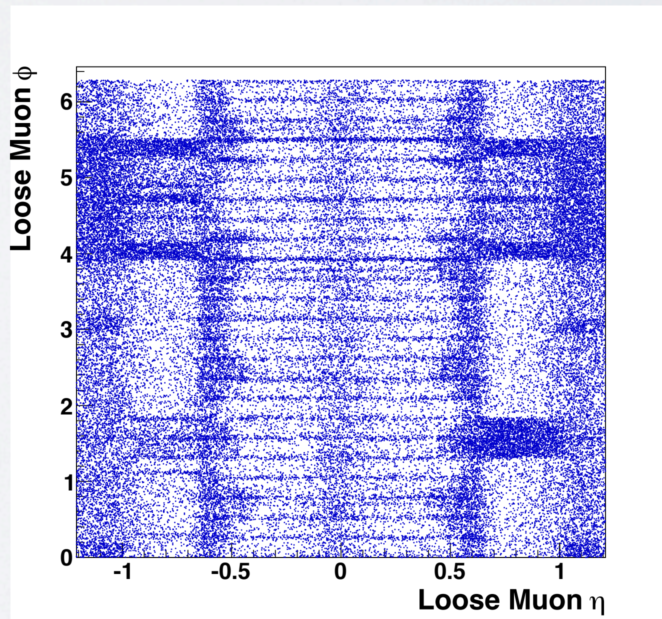
- One of my first contributions was increasing the muon acceptance

- I introduced a new category consisting of 2 'loose' muons
  - Missed muon detectors
  - Failed some reconstruction cuts

- This resulted in > 30% acceptance gain!





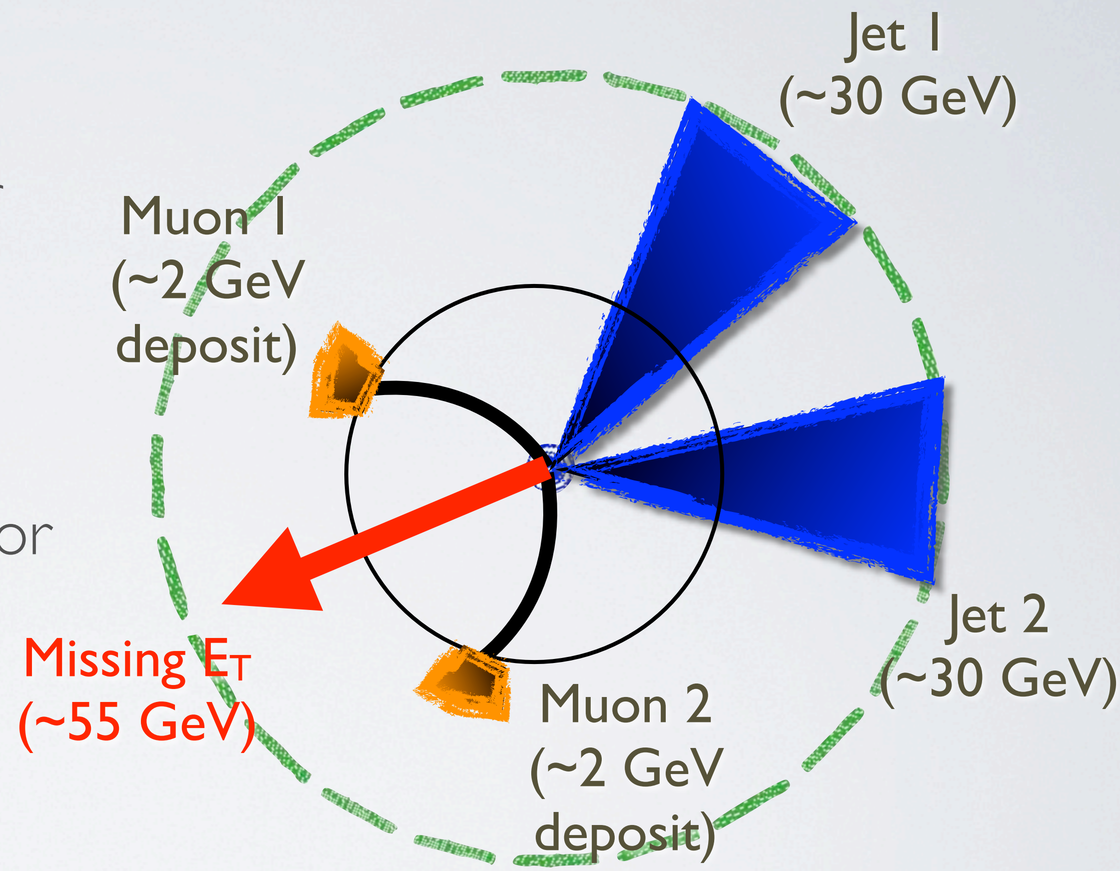


New Trigger Selection

- Previous searches in this channel used only a high-p_T muon trigger

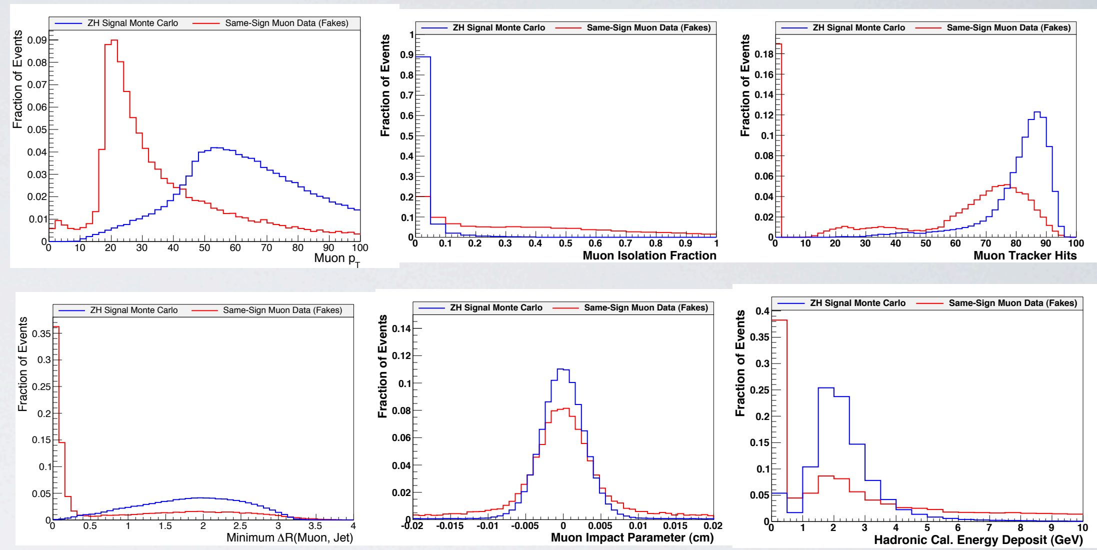
- As muons are minimal-ionizing particles, there is a signature of missing energy (MET) at the calorimeter level
  - Can trigger on this!
  - The MET is later corrected for the presence of the reconstructed muons

- My analysis was the first to use MET triggers in this channel!

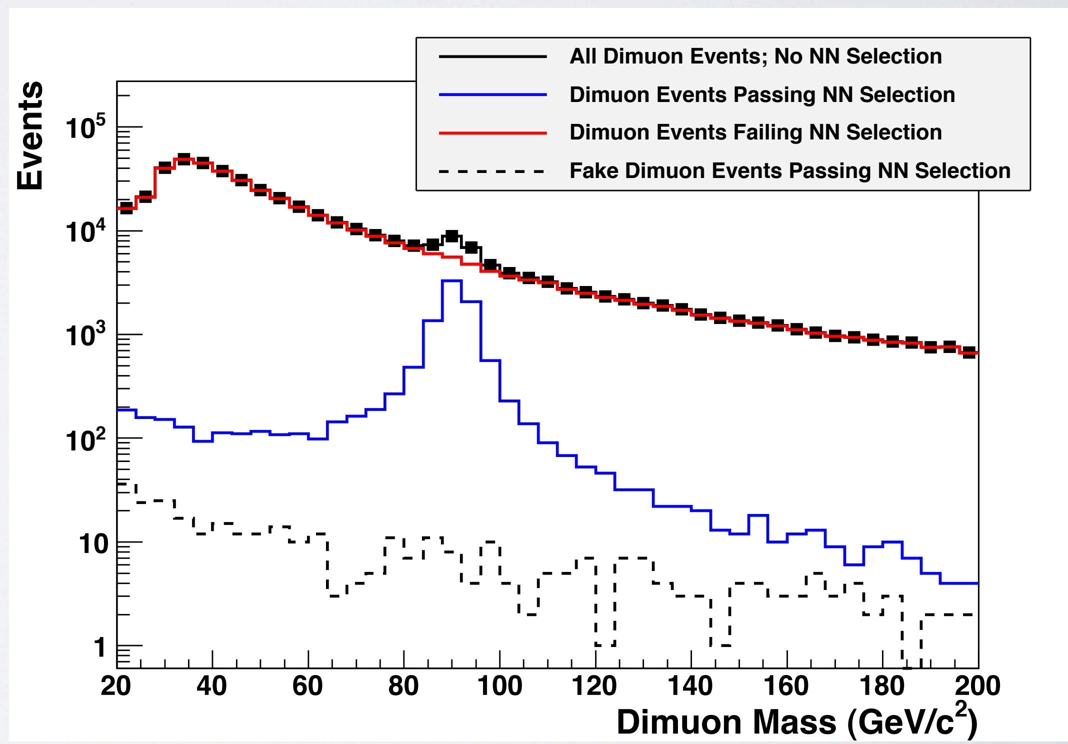


Multivariate Muon Selection

- To go one step further, I developed a multivariate muon selection function
  - Removes all reconstruction cuts
  - A single function to select high-quality muon candidates



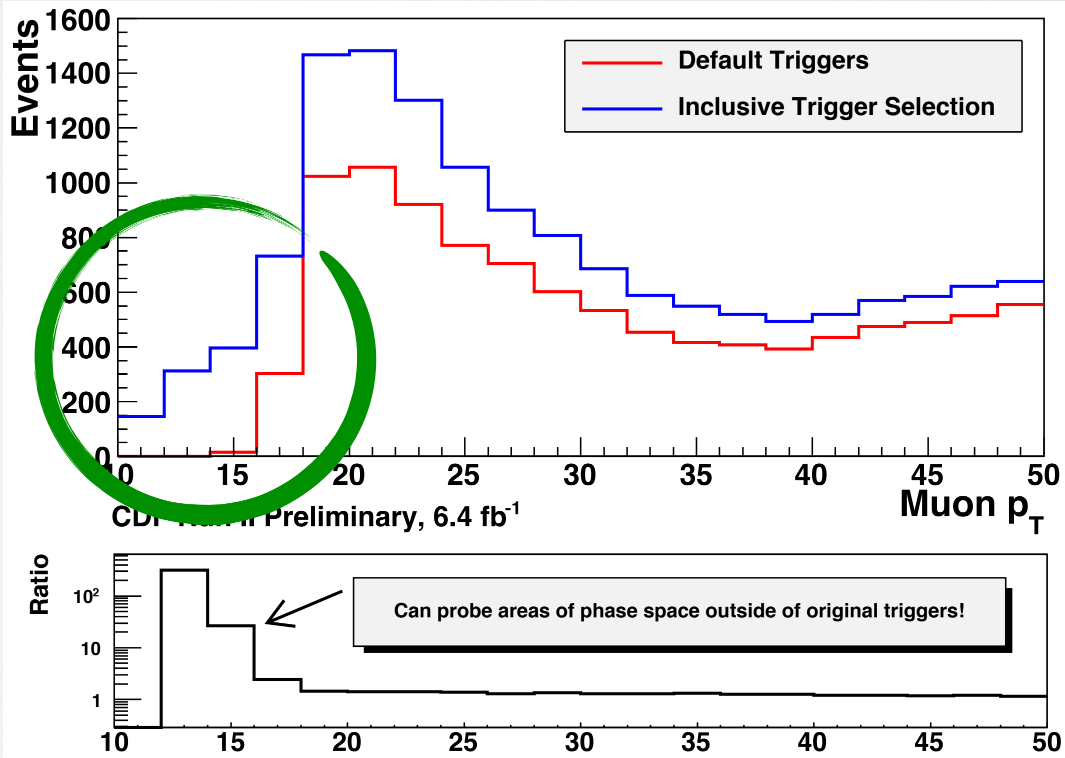
- Artificial neural network
  - Trained using several kinematic properties and detector responses of the muons



- 96% Z selection efficiency
- 94% fake rejection

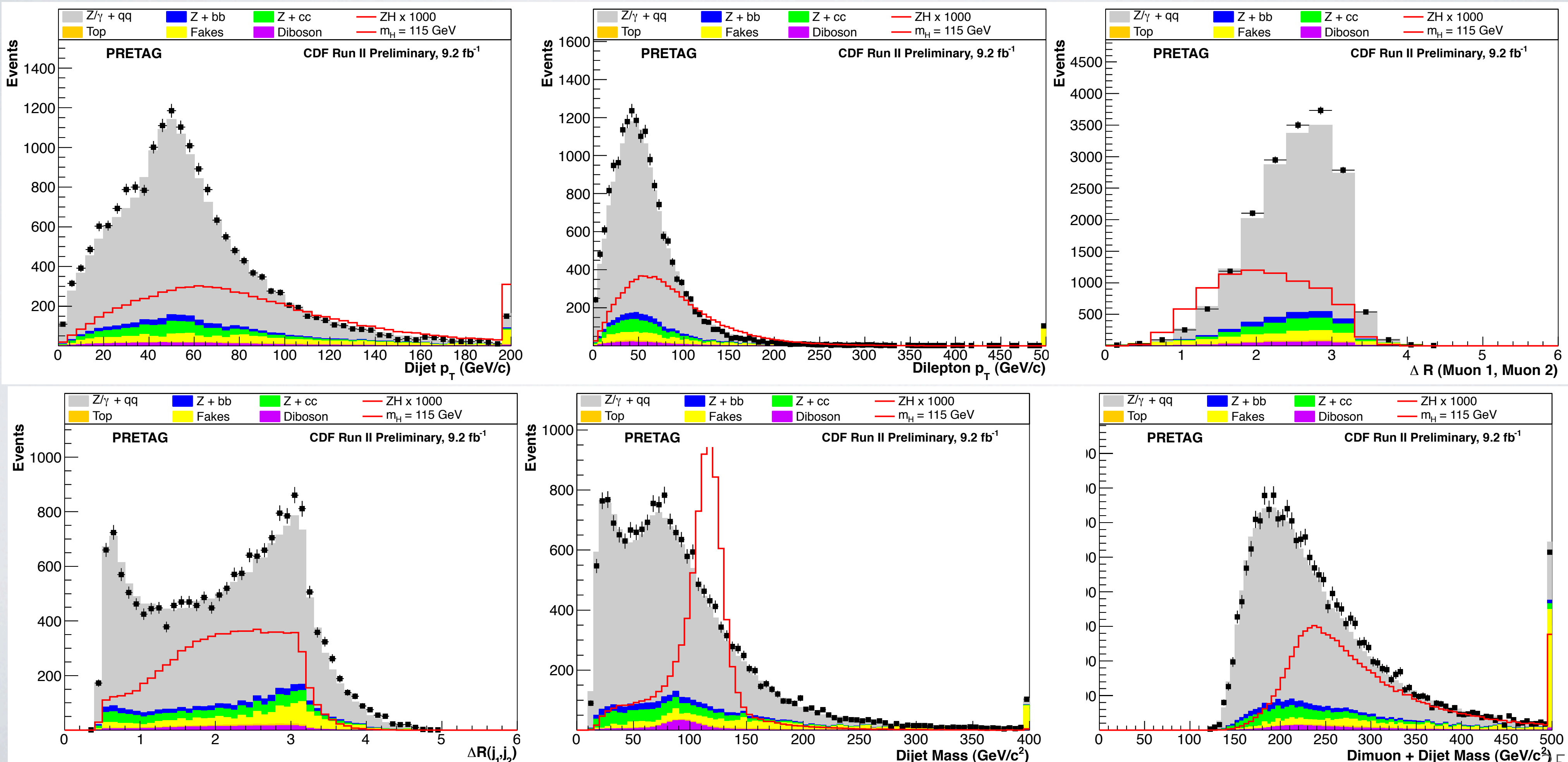
Inclusive Trigger Selection

- Usually, analyses use a single trigger to select data events
  - Lose events due to trigger selection efficiency
- This analysis was first to use an inclusive trigger selection
- Uses all available triggers!
- Increases acceptance by ~10%
- How to model trigger efficiencies?
  - Use a regression-based multivariate function
- Measured using independent datasets
- This technique has been extended to other CDF Higgs searches



Kinematic Distributions

- Check modeling of kinematic distributions before using them to discriminate signal from background
- Model includes all background processes, both MC and data derived



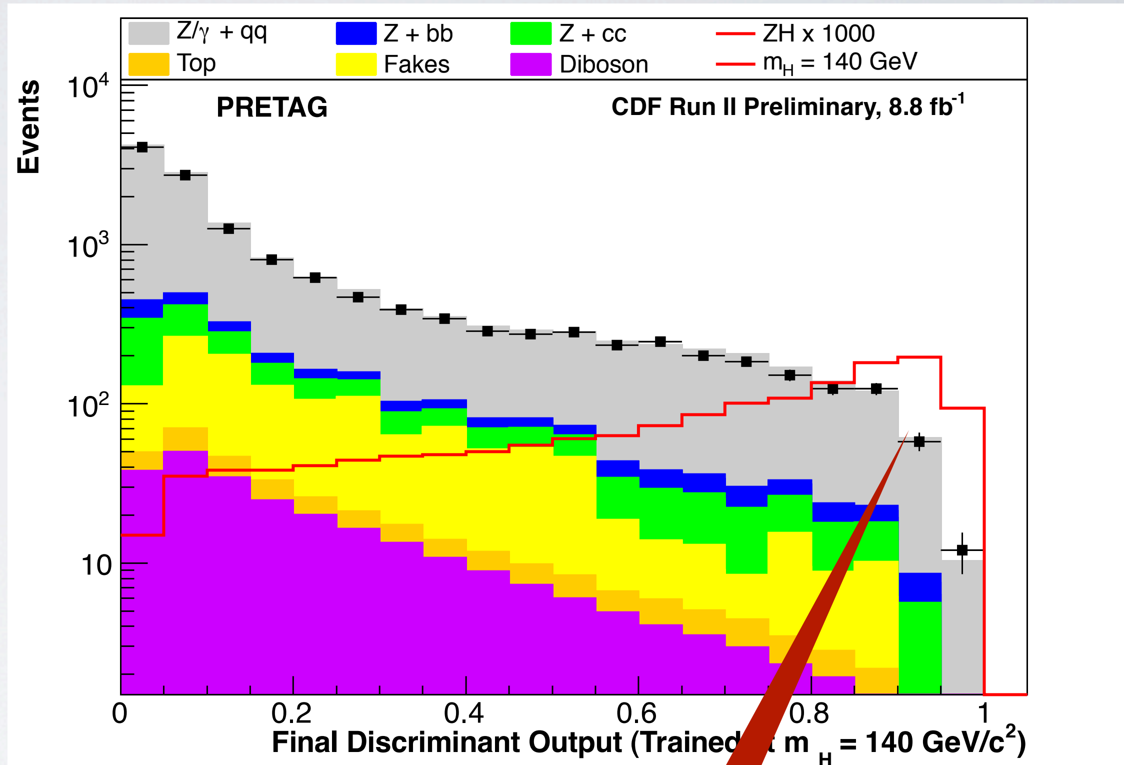
Forming a Discriminant

- Use kinematic variables with best signal separation as inputs to a neural network discriminant
- NN is trained using Monte Carlo
  - Mixture of processes for background
  - Specific ZH signal sample for each mass point
    - 11 independent networks
    - m_H = [100,105, ... , 150] GeV/c^2
    - Same set of input variables for each

DISCRIMINANT INPUTS
Number of Jets
MET
MET projection on Dijet System
Z P_T
Sum Jet E_T
Dijet Mass
Delta R(j_1 , j_2)
Delta R(Z , dijet system)
Sphericity
Total Mass
Dijet P_T
Muon 1 P_T
Muon 2 P_T
Delta R(μ_1 , μ_2)

Usual Technique

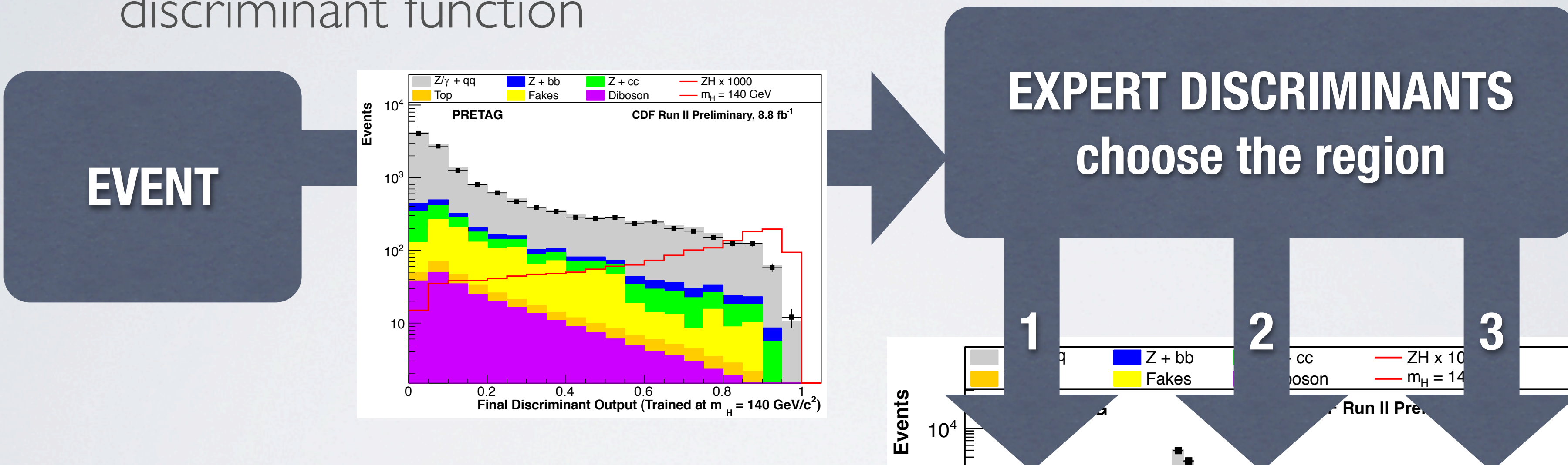
- Usually there is a single function to separate signal from background
- More information is available
- Kinematically distinct process can be separated
  - Top / Z+jets
  - *b* / light flavor jets
- I have implemented new tools to take advantage of these differences and enhance discrimination

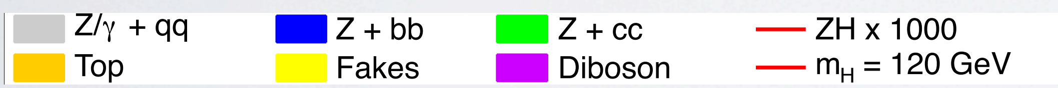


S:B greatly enhanced in these bins

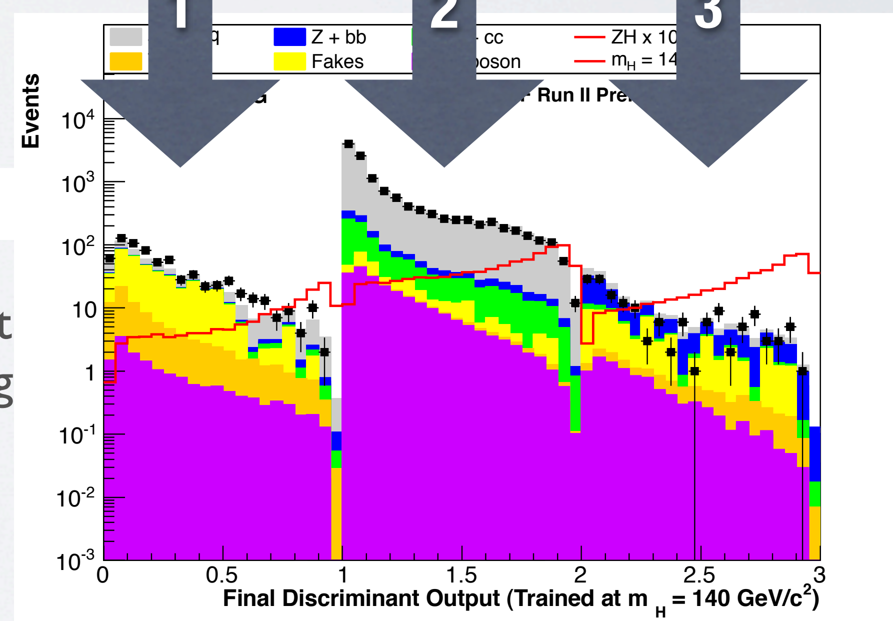
Multi-Layer Discriminant

- Events are divided into three regions based on these differences
  - Still use the same final signal discriminant function
- Expert Discriminants determine in which region an event falls



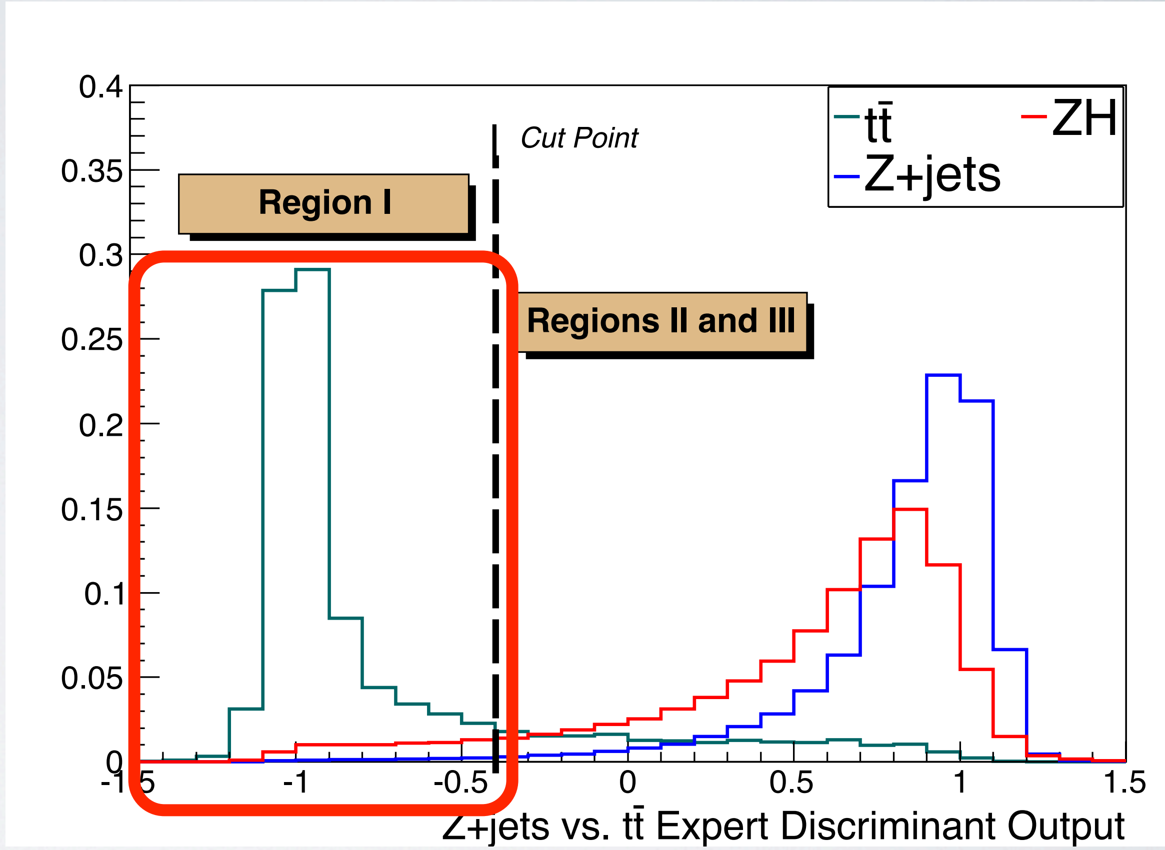


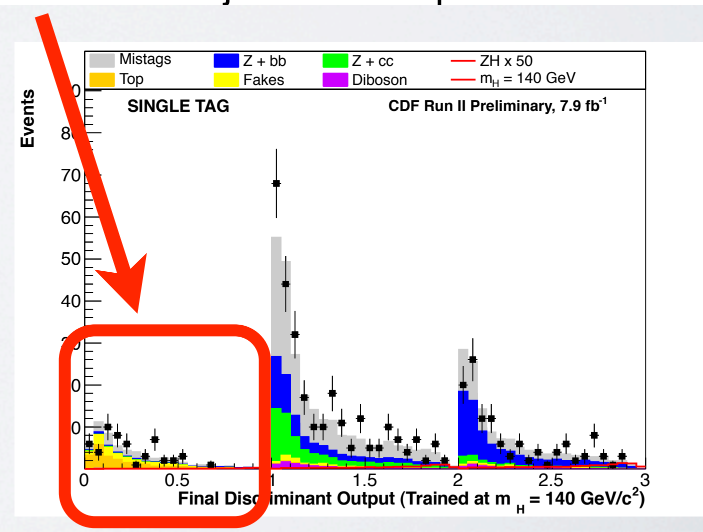
3 versions of the event discriminant containing different background compositions



Expert Discriminant

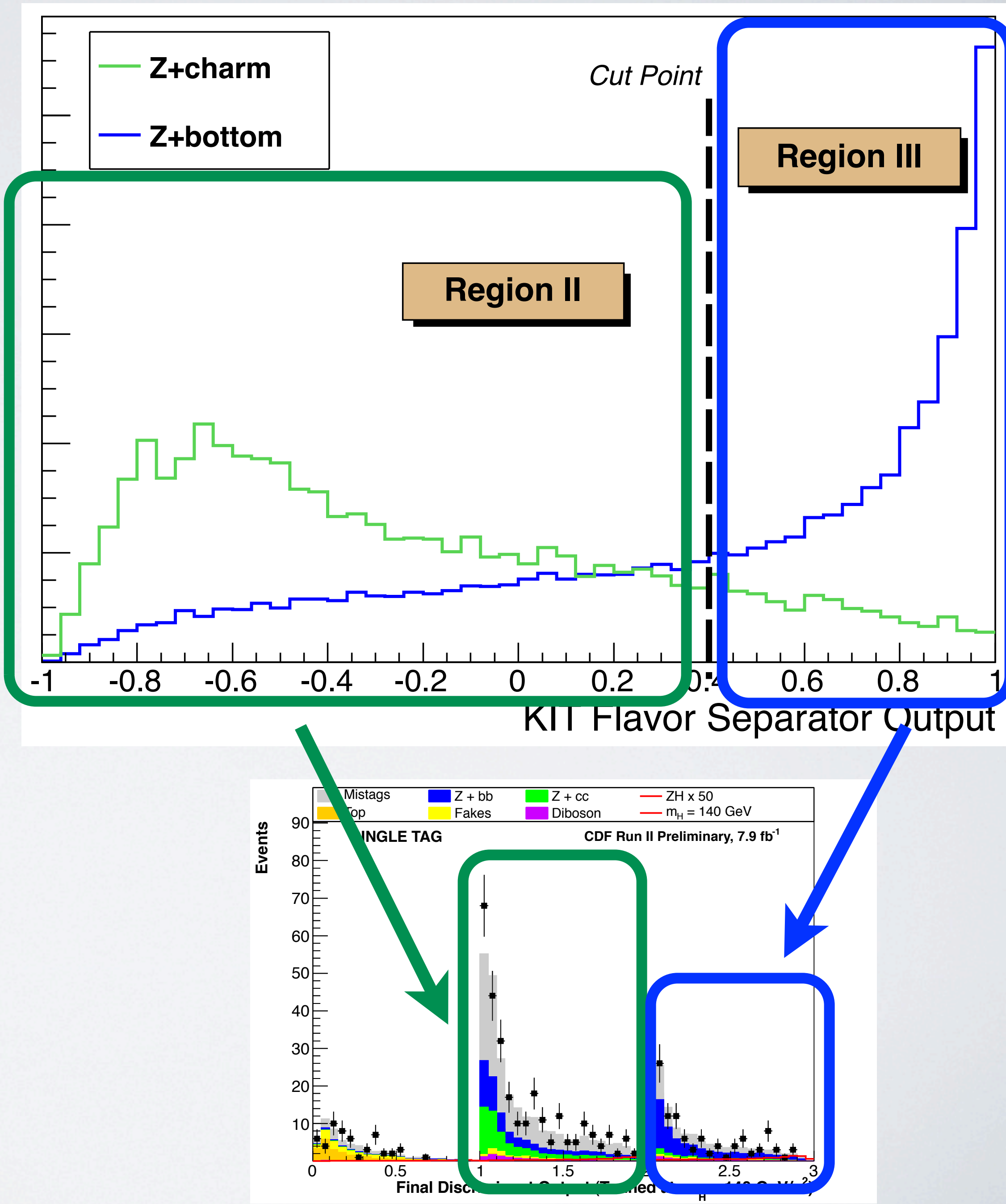
- First step involves distinguishing top pair events from Z+jets events
- An 'expert' discriminant is trained to separate these two processes
- A cut on the output of this function defines a top-pair enriched region
- Events below the cut are placed in the top-pair region (I)
- Events above the cut are evaluated further





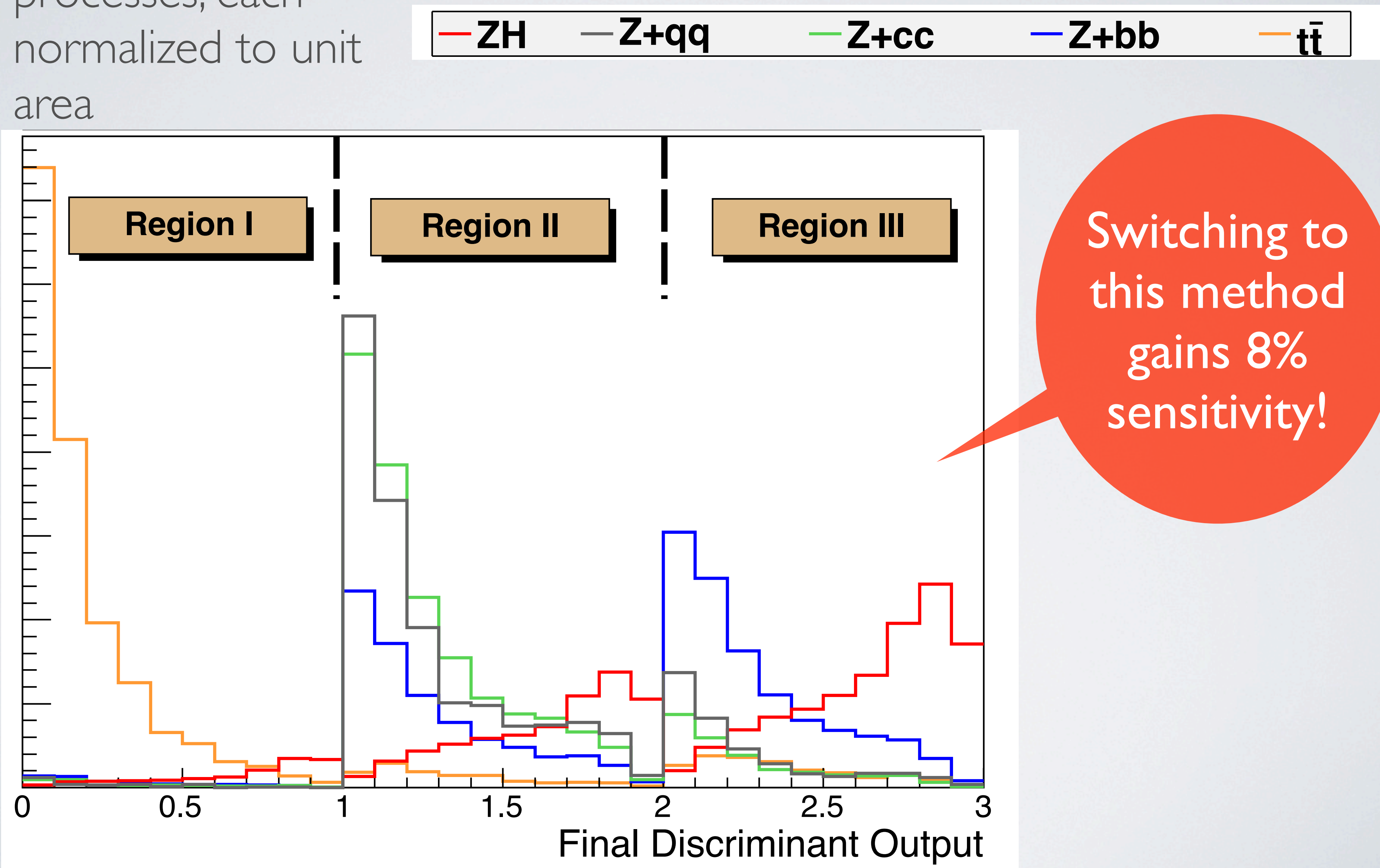
Flavor Separator

- Next step separates Z+charm and Z+light flavor events from Z+bottom and signal events
- A flavor separator is used here
  - Developed by Karlsruhe Institute of Technology
- A cut on this output defines regions 2 and 3
  - 2 -- Light flavor + charm jets
  - 3 -- Bottom jets
- Highest signal purity in region 3
  - Dominant background is Z+bb, almost no contribution from any other process



Summary of Method

- Output shapes of processes, each normalized to unit area



Switching to this method gains 8% sensitivity!

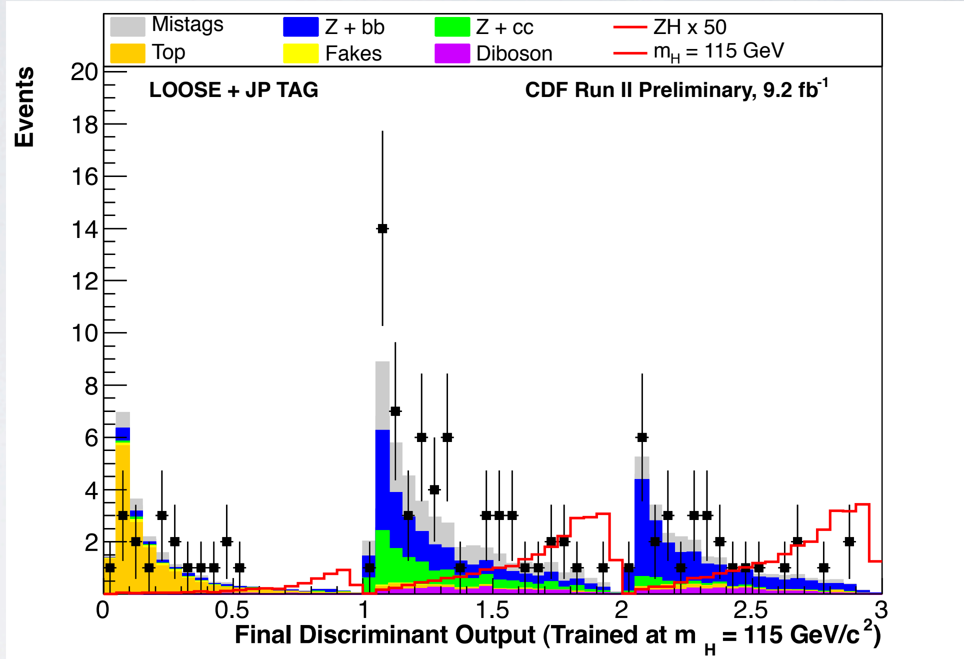
Signal Region

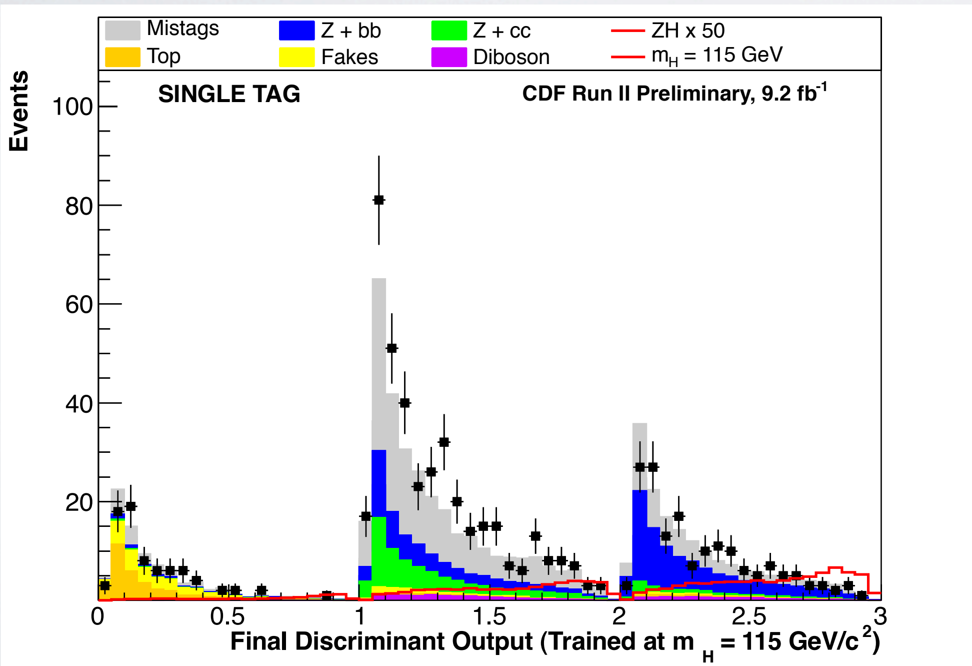
- We define three independent signal regions to set limits on Higgs production
  - Using the two *b*-tagging algorithms
    - Secondary Vertex Tight and Loose (SVT, SVL)
    - Jet Probability (JP)
- The three *b*-tag categories are:
  - 2 x SVT
  - 1 x JP + 1 x SVL
  - 1 x SVT
- Expected events shown in table
  - Data in agreement with expectation

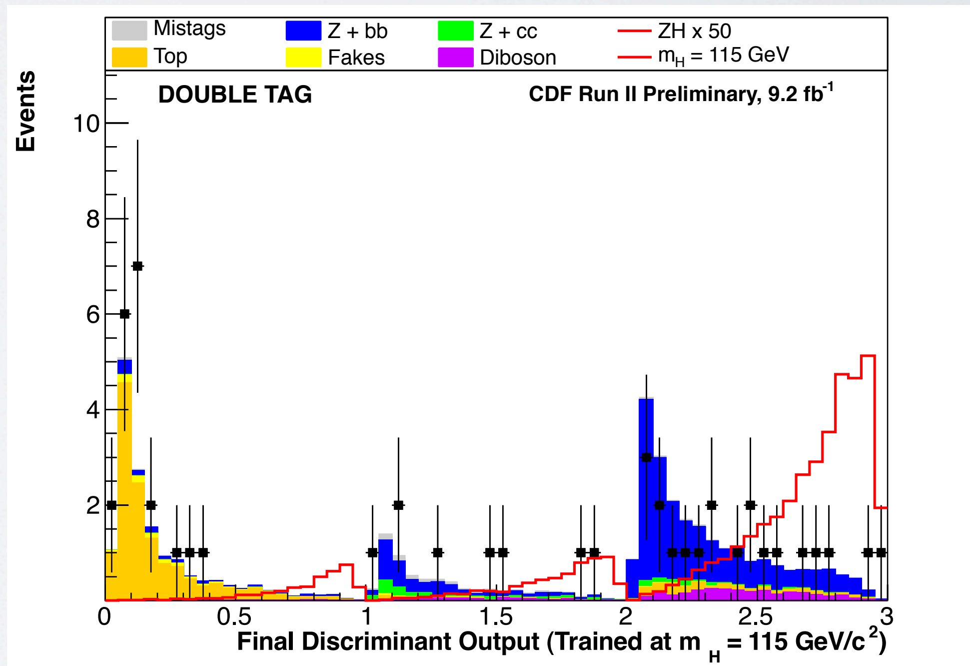
Process **Expected Events**	SVT	SVL+JP	SVT+SVT
$Z \to \mu\mu +$ Mistags	260.0 ± 35.1	22.1 ± 6.2	1.3 ± 0.4
$Z \to \mu\mu + c\bar{c}$	71.7 ± 28.7	11.8 ± 4.7	2.0 ± 0.8
$Z \to \mu\mu + b\bar{b}$	148.3 ± 59.3	33.4 ± 13.4	22.0 ± 8.8
$t\bar{t}$	40.2 ± 4.0	18.6 ± 1.9	15.3 ± 1.5
ZZ	15.1 ± 0.9	4.9 ± 0.3	3.4 ± 0.2
WZ	5.4 ± 0.3	0.5 ± 0.01	–
WW	0.3 ± 0.01	–	–
Misidentified $Z \to \mu\mu$	33 ± 1.7	1 ± 0.1	1 ± 0.1
Total Background	574.0 ± 74.8	92.2 ± 15.6	45.1 ± 9.0
Observed Data	631	107	48

Discriminant Distributions

- These distributions used to set limits on Higgs production
  - Different for each Higgs mass hypothesis

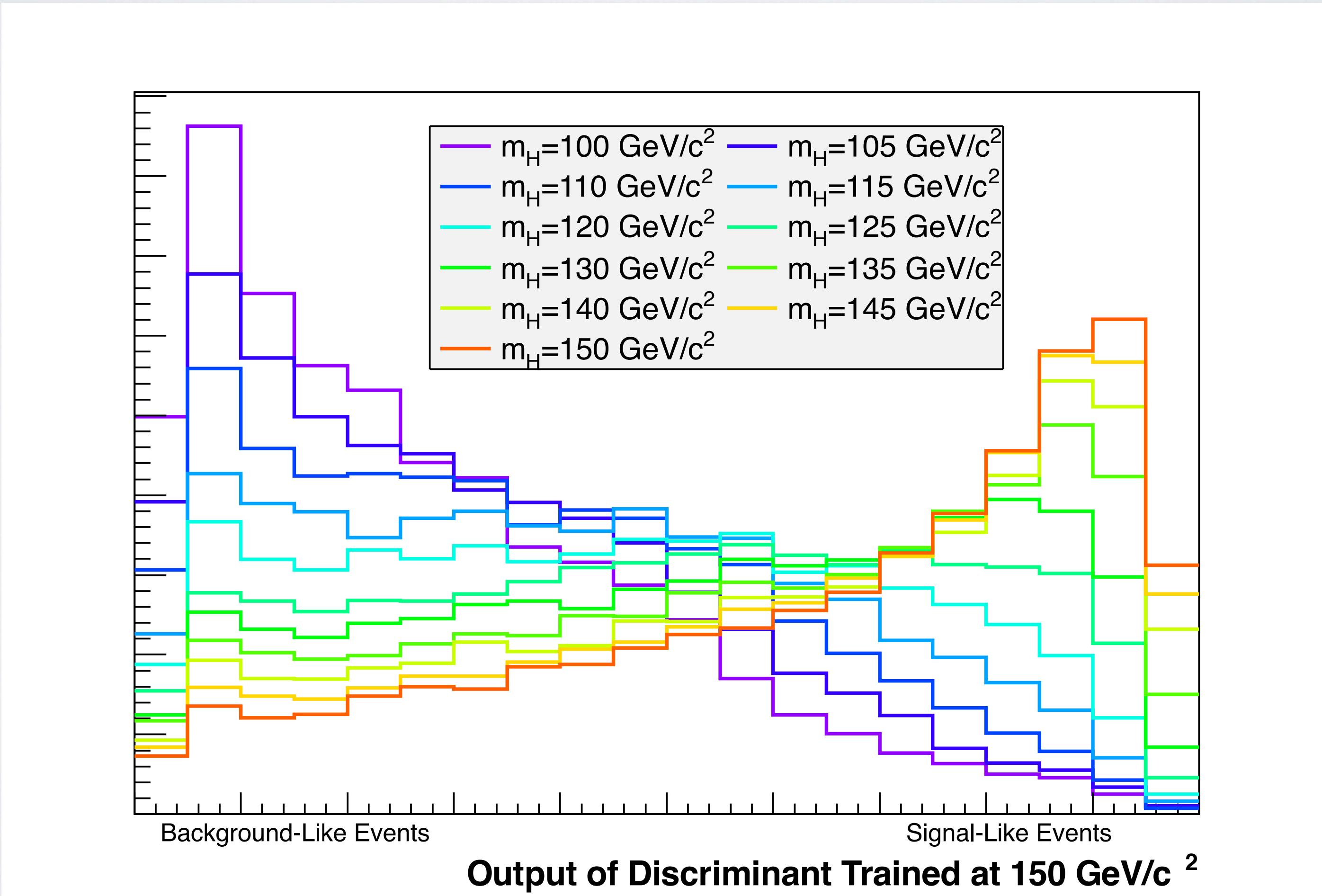






Higgs Mass Variation

- Output scores vary depending on the Higgs mass hypothesis
  - Optimization for each Higgs mass hypothesis



Systematic Uncertainties

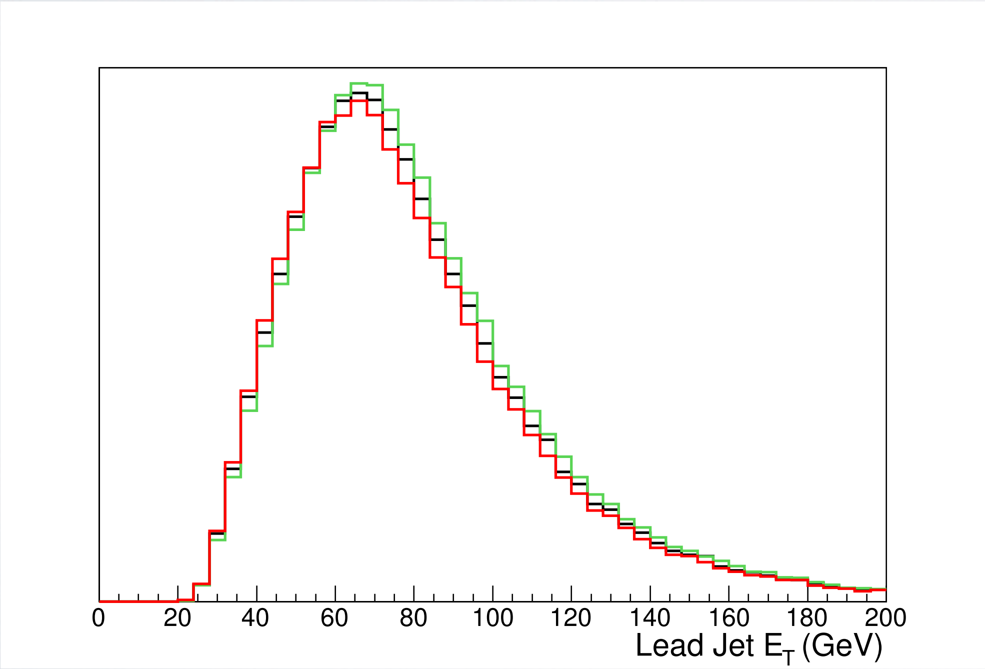
- No significant excess is observed
  - Proceed to set upper limit on Higgs production cross section
- Must account for several systematic uncertainties
- Rate uncertainties
  - Cross sections
  - Luminosity
  - Reconstruction factors
- Shape uncertainties
  - Jet Energy Scale
  - Mistag Estimate

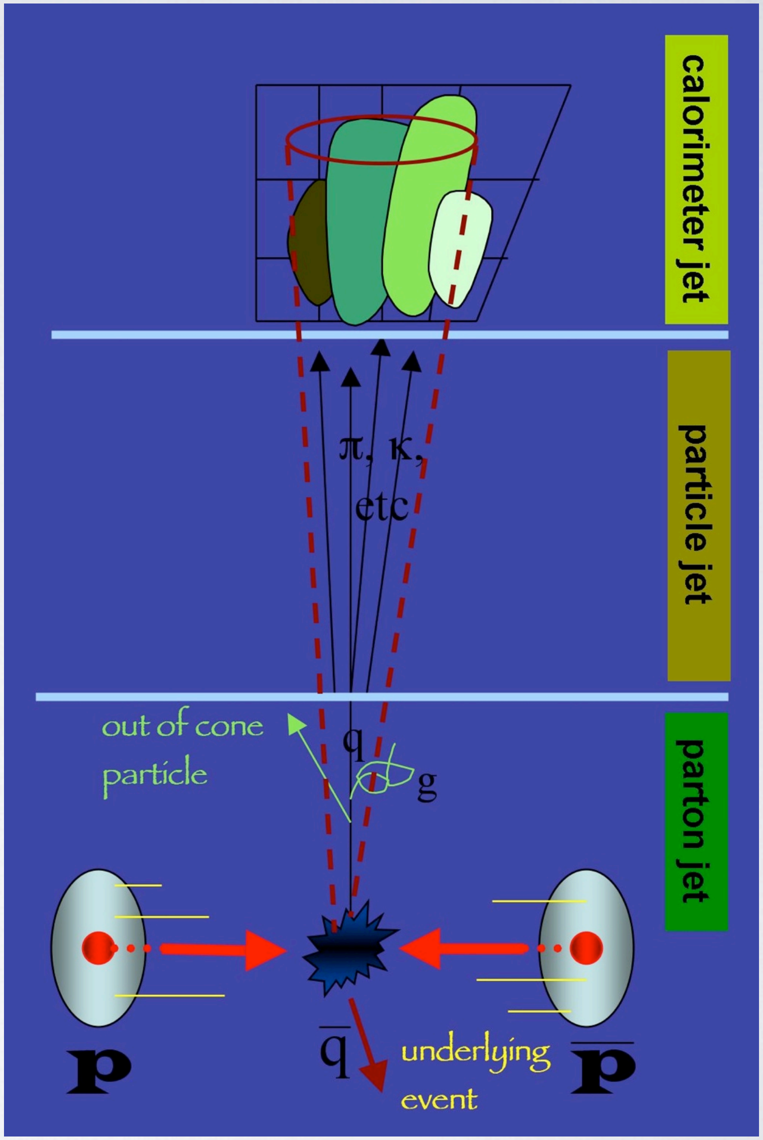
Systematic	VALUE (%)	Applied To
Luminosity (CDF)	4.4	All MC
Luminosity (TeV)	3.8	All MC
Trigger Model	5.0	All MC
Lepton ID	1.0	All MC
HF Cross Section	40.0	Z+bb, Z+cc MC
VV Cross Section	6.0	WW, WZ, ZZ MC
tt Cross Section	10.0	tt MC
Fakes	5.0	Fake Template
b-Tag Scale Factor	4.0	All MC
Jet Energy Scale	Shape	All MC
ISR / FSR	5.0	ZH MC
Mistag Matrix	Shape	Mistag Template

Jet Energy Scale

- Largest Shape Systematic
  - Rate component ~5%

- Shifts jet energies up and down
  - Changes 4-vectors
  - Changes kinematic variables
  - Changes output discriminant

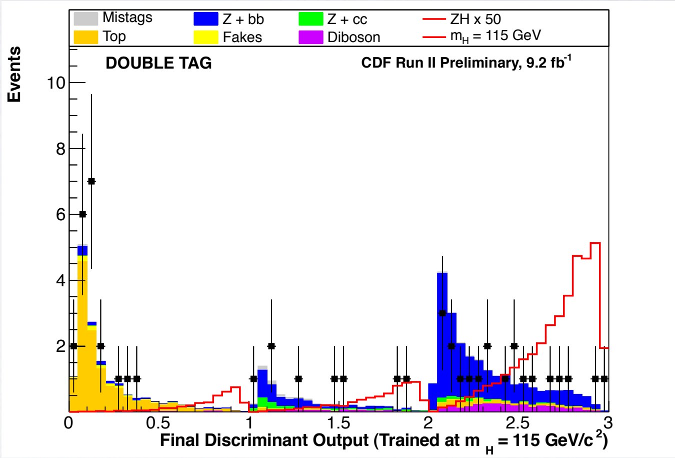




Setting Limits

- Use a counting experiment in each bin of the final output distributions

$$P(\text{data} \mid \text{signal, background}) = \prod_{i=1}^{N} \frac{e^{-(s\epsilon_i + b_i)}(s\epsilon_i + b_i)^{n_i}}{n_i!}$$



- To account for systematic and statistical errors, form posterior distribution for signal
  - Sample many different configurations of statistical and systematic choices
    - $M \sim 1000$
    - Gaussian priors for rate systematics
    - Shape templates for JES

$$p(s) = \frac{1}{M \cdot \mathcal{N}} \sum_{j=1}^{M} \left[\prod_{i=1}^{N} \frac{e^{-(s\epsilon_{j,i} + b_{j,i})}(s\epsilon_{j,i} + b_{j,i})^{n_i}}{n_i!} \right]$$

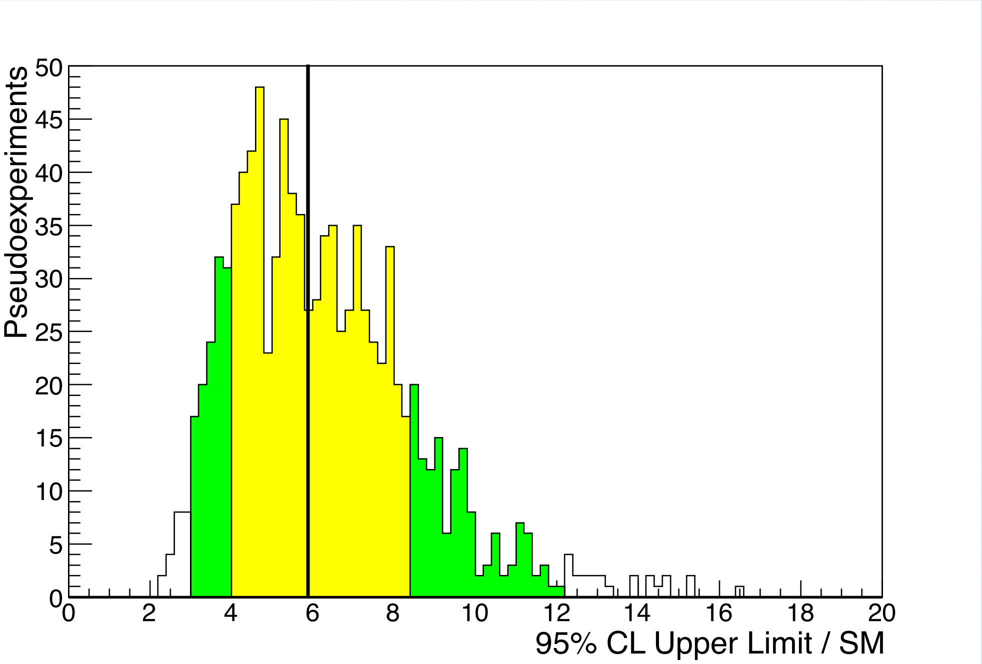
Setting Limits

- With the signal posterior formed, we integrate to find the upper limit at the 95% CL

$$I(s') = \int_{s'}^{\infty} p(s)\,\mathrm{d}s$$

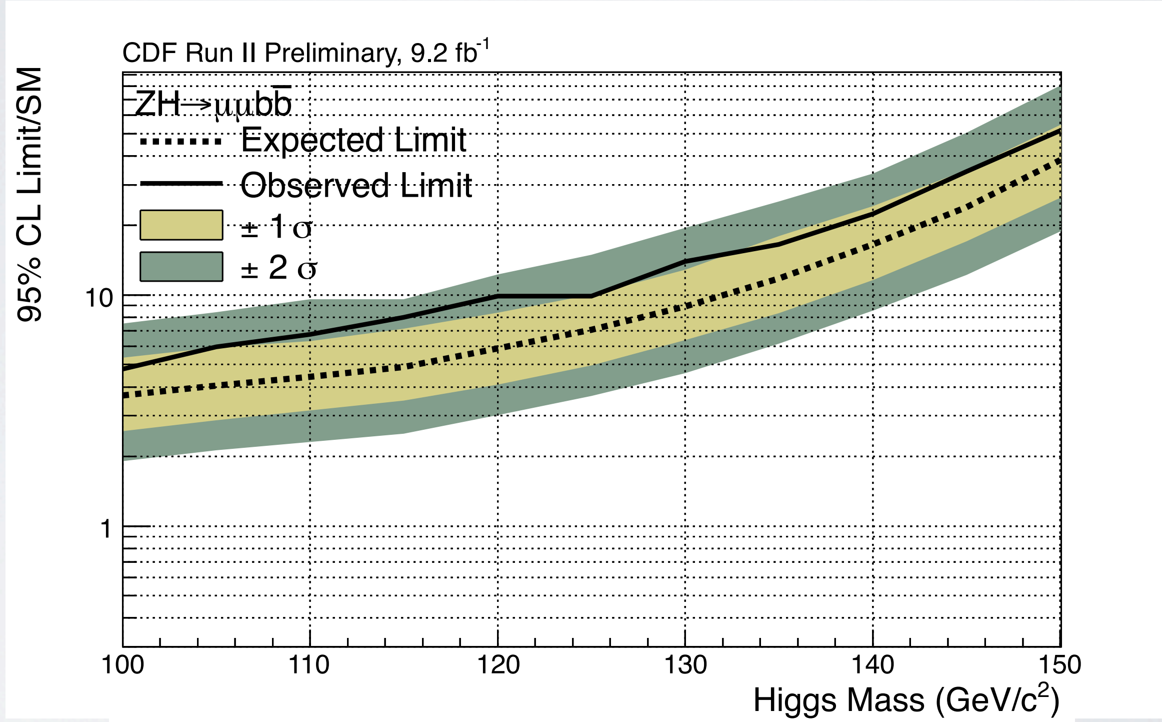
- To estimate the sensitivity of the analysis, we perform 1000 pseudoexperiments
  - Pseudodata formed from sampling the background-only hypothesis
  - Procedure from previous slide repeated for each pseudoexperiment
  - Gives a distribution of upper limits
    - Median, 1, 2, sigma bands

$$I(s_{\mathrm{upper}}) = 0.05 \cdot I(0)$$



Setting Limits

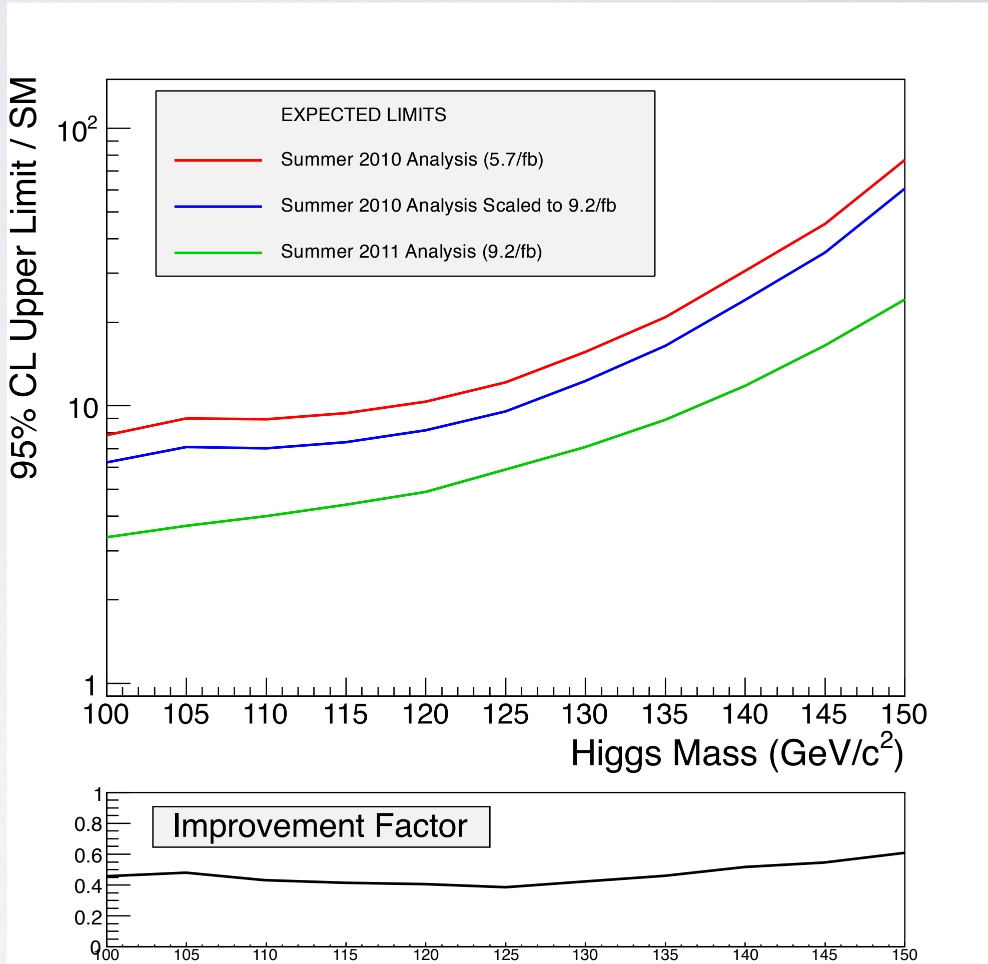
- We plot the expected limits along with the 1, 2 sigma uncertainty bands
- Observed limit comes from actual data distribution
  - No pseudodata
- Upper limits shown as a ratio to the Standard Model expected value
  - 1xSM = Higgs sensitivity



m_H (GeV/c^2)	Observed Limit	Expected Limits				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
100	**4.8**	1.9	2.6	**3.7**	5.4	7.5
105	**6.0**	2.1	2.9	**4.0**	5.9	8.4
110	**6.8**	2.3	3.2	**4.4**	6.3	9.6
115	**8.0**	2.5	3.5	**4.9**	7.1	9.6
120	**9.9**	3.0	4.1	**5.9**	8.4	12.3
125	**9.9**	3.7	4.9	**7.1**	10.0	14.9
130	**14.0**	4.6	6.4	**8.9**	12.8	19.5
135	**16.5**	6.2	8.4	**11.8**	17.9	25.4
140	**22.5**	8.5	11.6	**16.5**	24.2	33.6
145	**34.2**	12.2	17.1	**24.1**	34.7	50.5
150	**51.6**	18.8	26.4	**38.4**	54.6	80.7

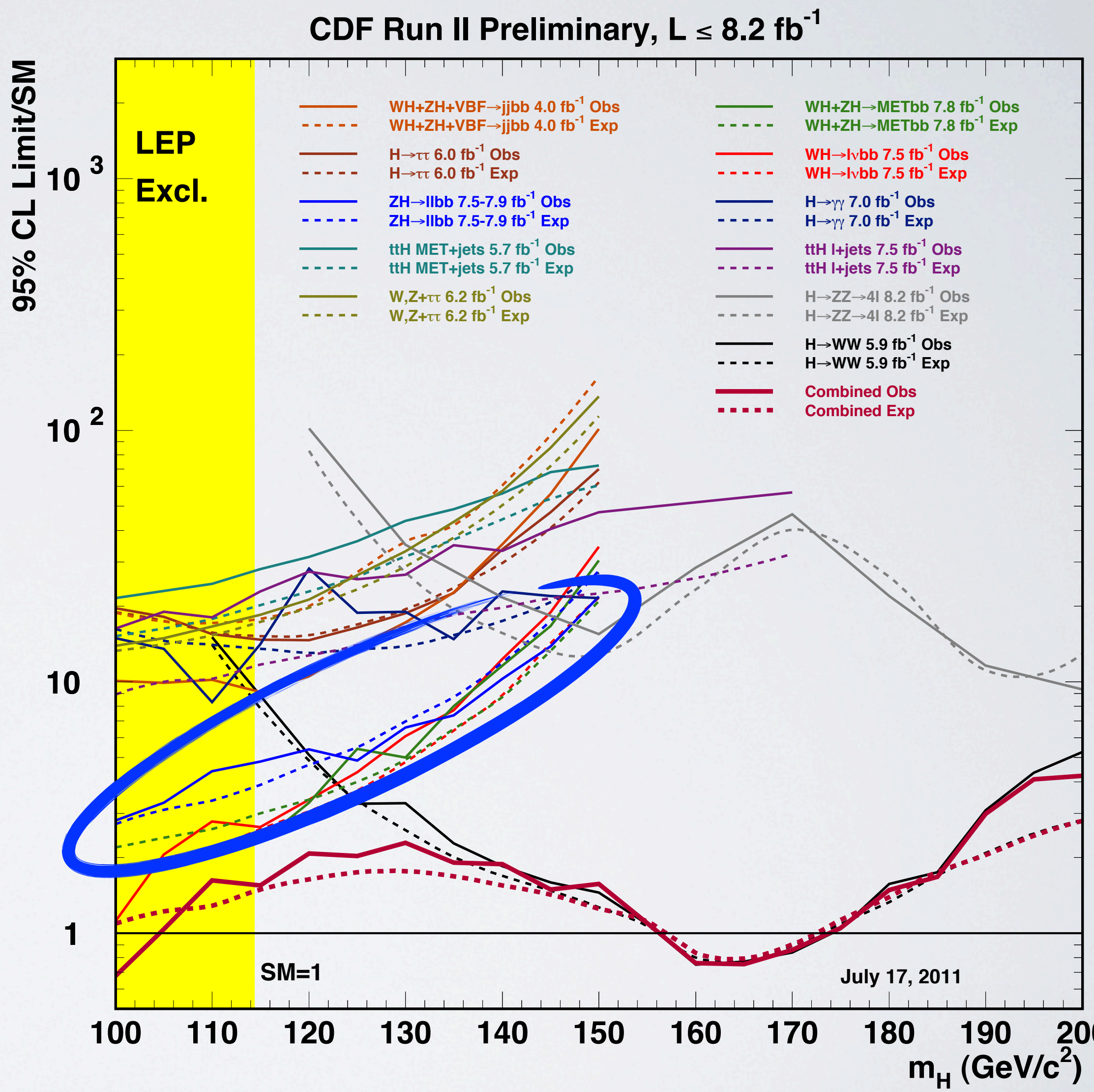
Total Improvement

- The techniques I have implemented have greatly improved the analysis sensitivity
- Over 50% increase in sensitivity from the new techniques alone!
  - Does not include the larger dataset
- My contributions have made a direct impact on Higgs searches at CDF
  - Many of the techniques being propagated to other search channels for similar gains



Context - CDF

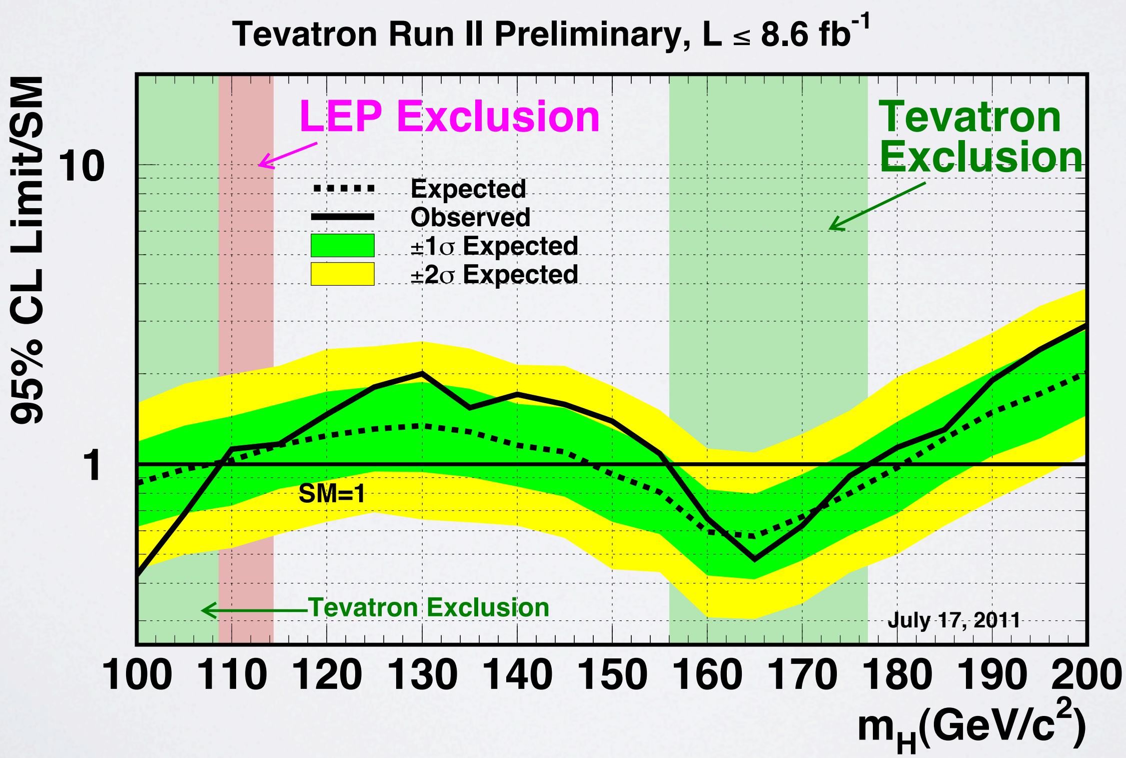
- This analysis can be combined with other Higgs search channels at CDF
    - *ZH* dilepton channel is third most sensitive channel at CDF
- Combination sensitivity is < 2.0 x SM in the low mass range
- Two excluded mass ranges
    - 100-104.5 GeV/c^2
    - 156.5 - 173.7 GeV/c^2



Context - Tevatron

- Similar combination between all of CDF and D0 Higgs searches
- Immense effort of many people to produce this result!

- Approaching Standard Model sensitivity
  - < 1.3 x SM in low mass range
- Two exclusion regions
  - 100-108 GeV/c^2
  - 156-177 GeV/c^2

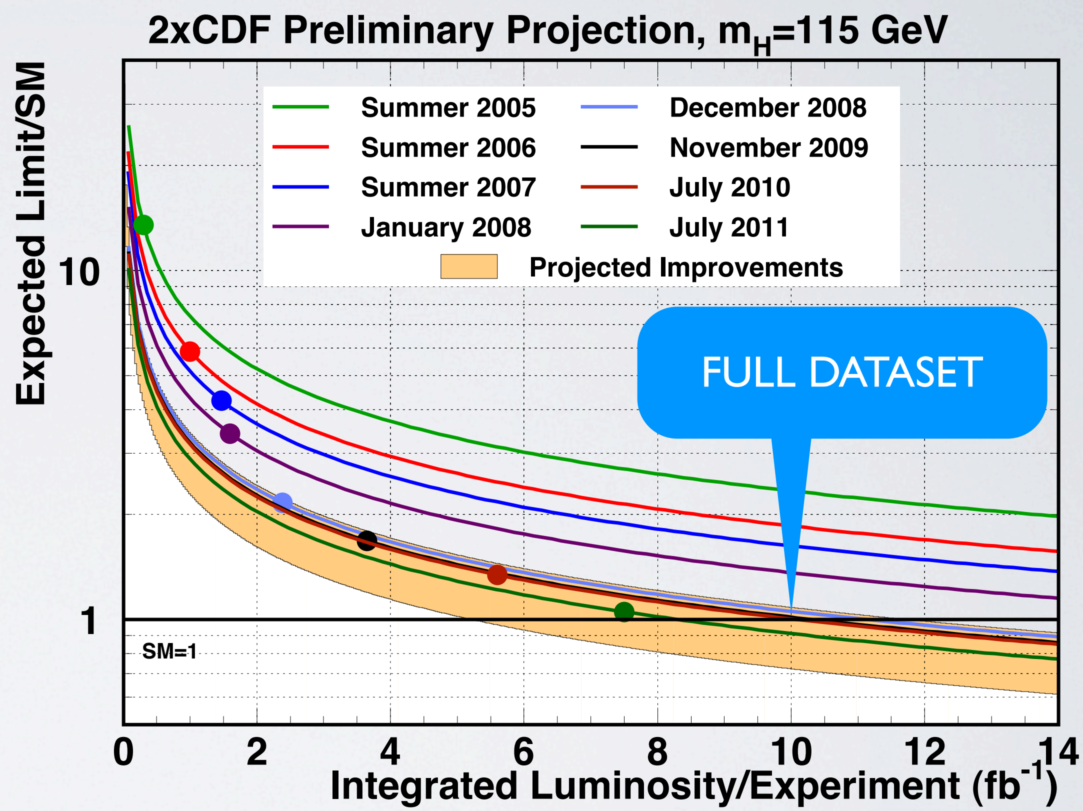


Still Work To Do!

- Analyses are working on improvements for the final analyses with the full Tevatron dataset

- Hope to reach Standard Model sensitivity across the entire search range
  - Start to see a Higgs signal?
  - Rule out the Higgs?

- Stay tuned for future results!
  - Planning on having final Tevatron result ready by 2012 Winter conferences



Conclusions

- Higgs searches at the Tevatron involve small signals in the presence of large backgrounds
  - This provides the opportunity to develop many new techniques to increase the sensitivity of analyses
  - Similar techniques will be useful for LHC searches!
- My thesis result implements several multivariate methods in the ZH→llbb analysis
  - Muon selection
  - Trigger efficiency modeling
  - Expert discriminants
  - Flavor separators
  - NN ensembles
- These techniques alone have improved the sensitivity of the analysis by over 50% relative to previous versions
  - Still some improvements in the pipeline for later updates
  - Always room for improvement!