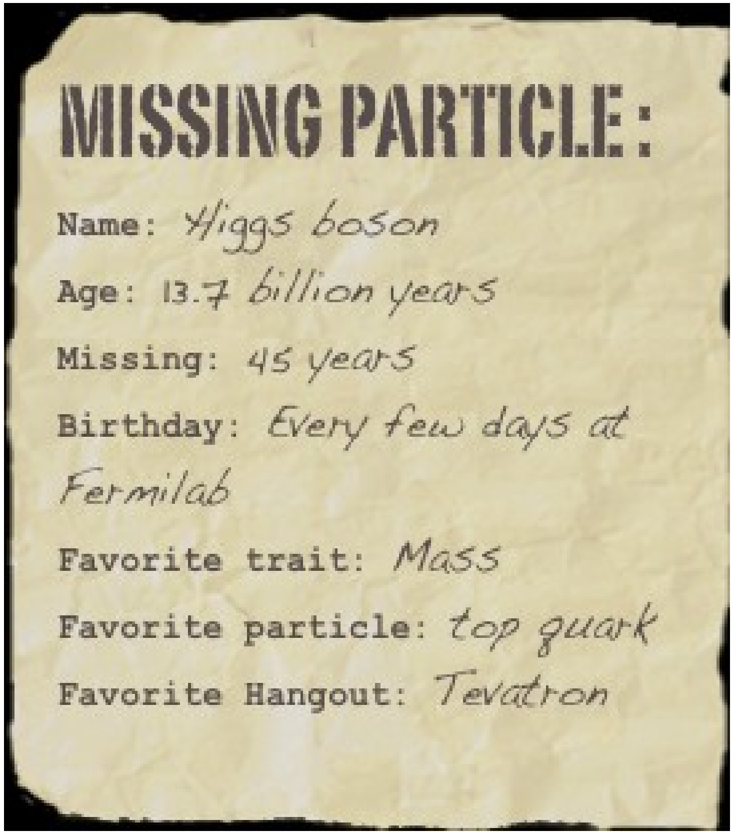
Updated Search for H WW at CDF with 6 fb^{-1}

Jennifer Pursley, University of Wisconsin-Madison

Fermilab Wine & Cheese Seminar, June 18, 2010

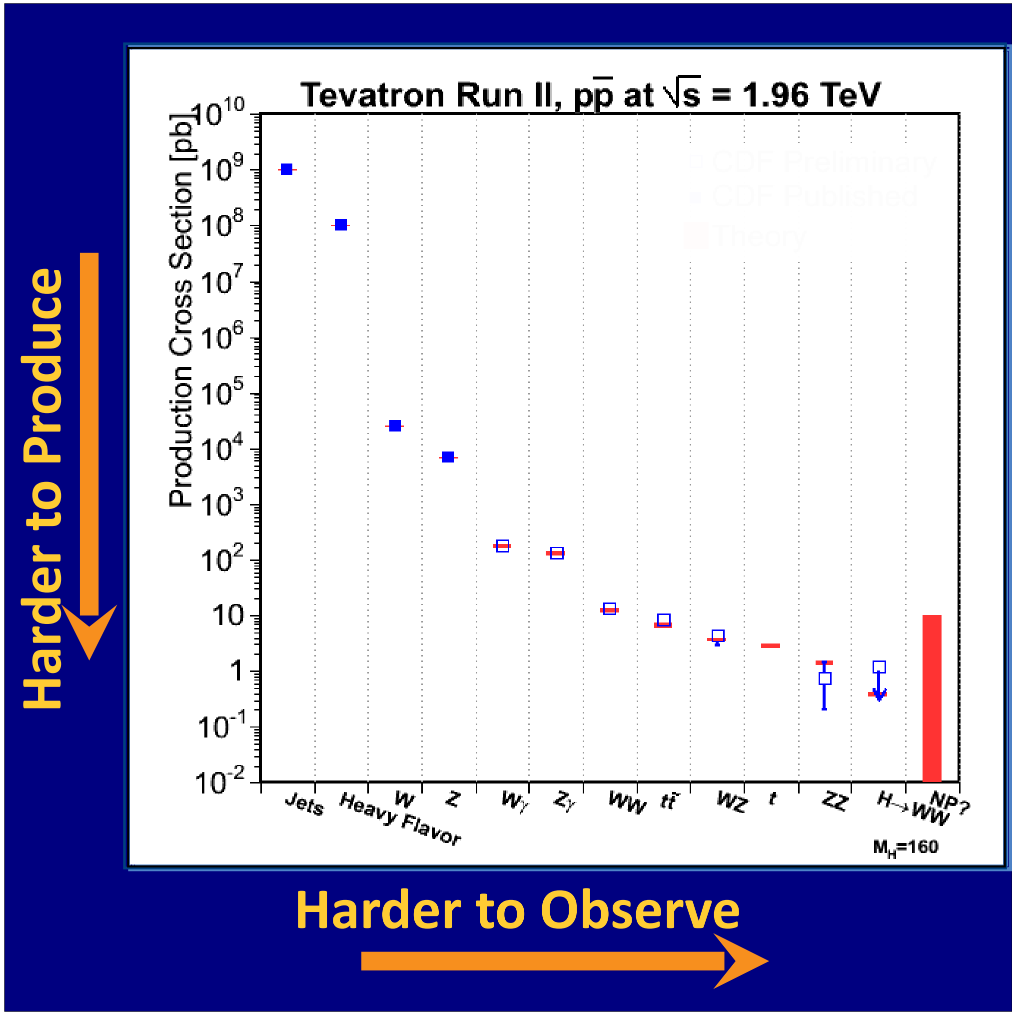
Outline

- Overview
- Analysis Strategy
  - Cross-checks
- Theoretical and systematic uncertainties
- Results and conclusions
  - Future research
- Many recent summaries of Tevatron Higgs programs:
  - Jay Dittmann, Users Meeting, June 2, 2010
  - Matthew Herndon, Wine & Cheese on March 12, 2010
  - Sergo Jindariani, Wine & Cheese on March 13, 2009



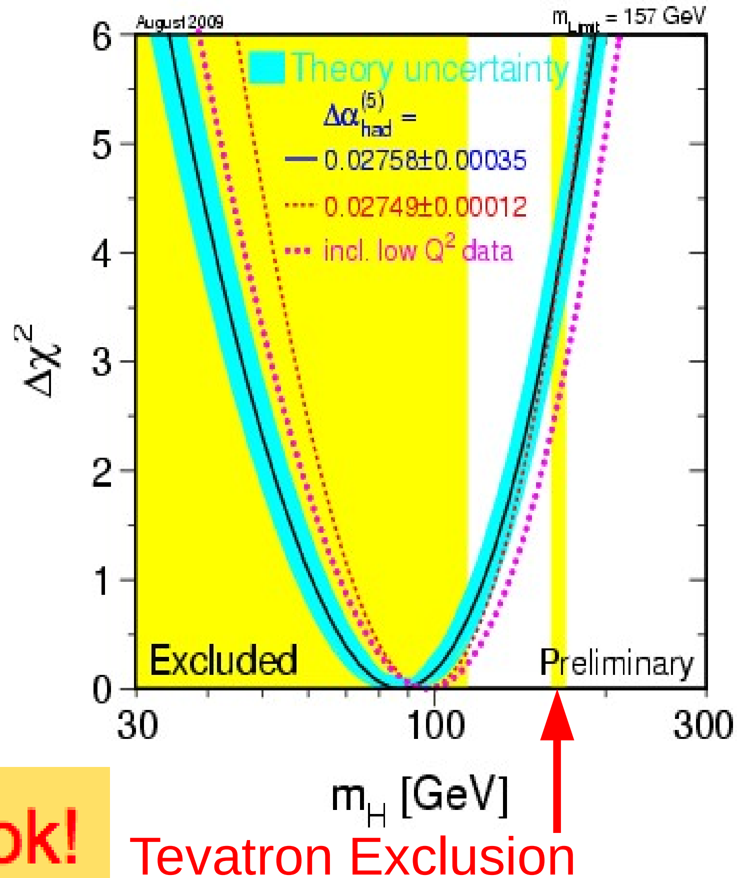
Collider Physics at the Tevatron

- Tevatron experiments probe processes many orders of magnitude apart in cross section
- Precision measurements and new discoveries
  - WZ, ZZ, single top
- Now reaching sub-picobarn cross section sensitivity
  - Standard Model Higgs boson is within reach!



The Higgs Boson

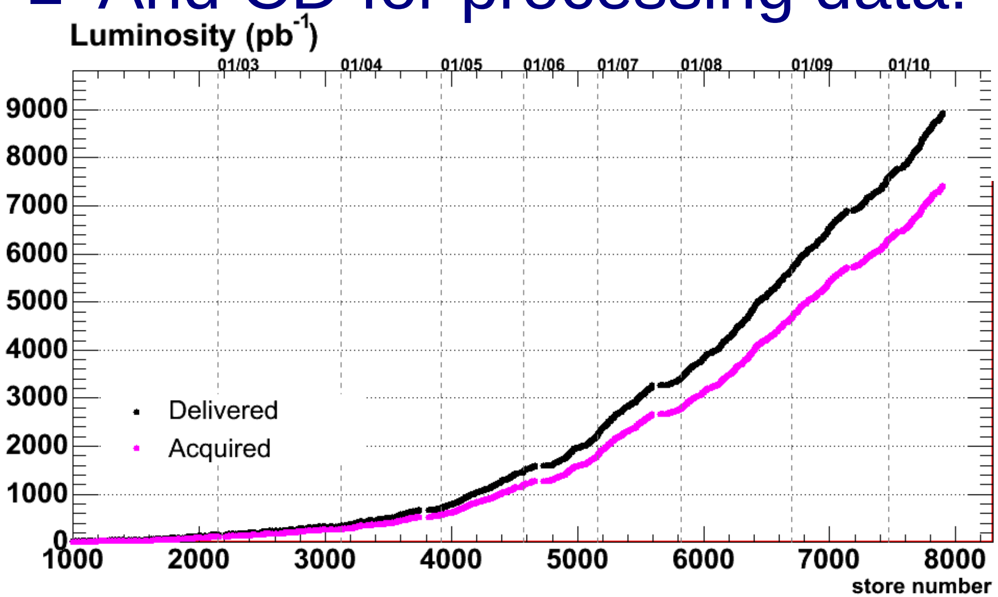
- Higgs Mechanism generates the mass of particles
  - … yet, gives no hint of what the Higgs boson mass is
- If Higgs boson exists, its mass must be determined experimentally
- LEP direct searches excluded m_H < 114.5 GeV at 95% C.L.
  - Tevatron excludes 162-166 GeV from CDF+D0 H   WW searches
- Indirect electroweak constraints prefer light Higgs (< 154)
  - Combined with LEP results upper limit of m_H < 185

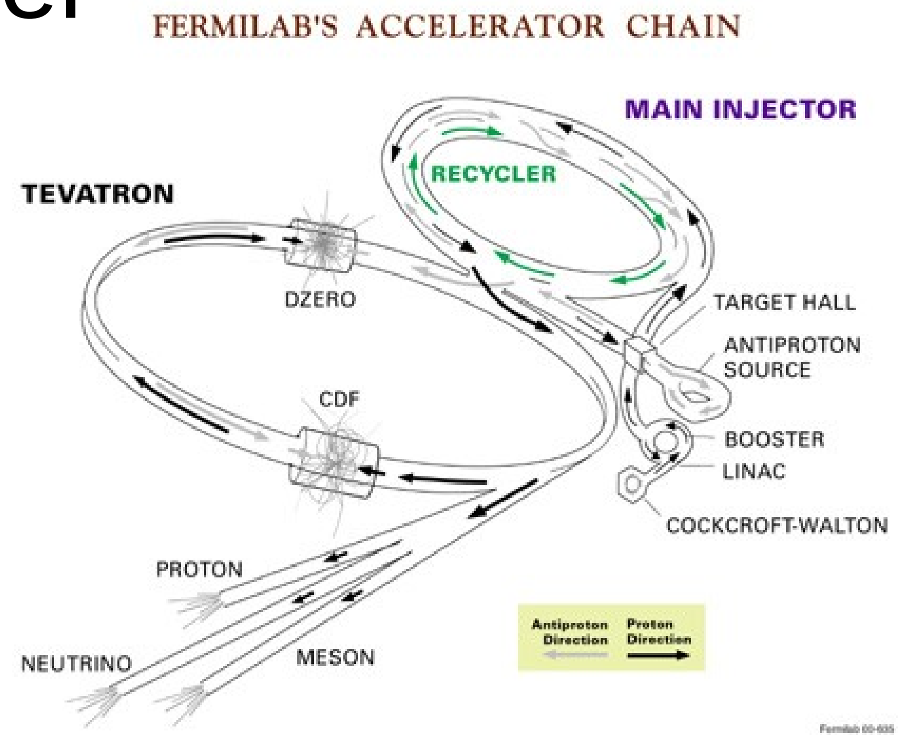


We know where to look!

The Tevatron Collider

- Collides $p\bar{p}$ at $\sqrt{s}$ = 1.96 TeV
- Thanks to AD for delivering luminosity...
- And CDF for keeping the detector running well...
- And CD for processing data!

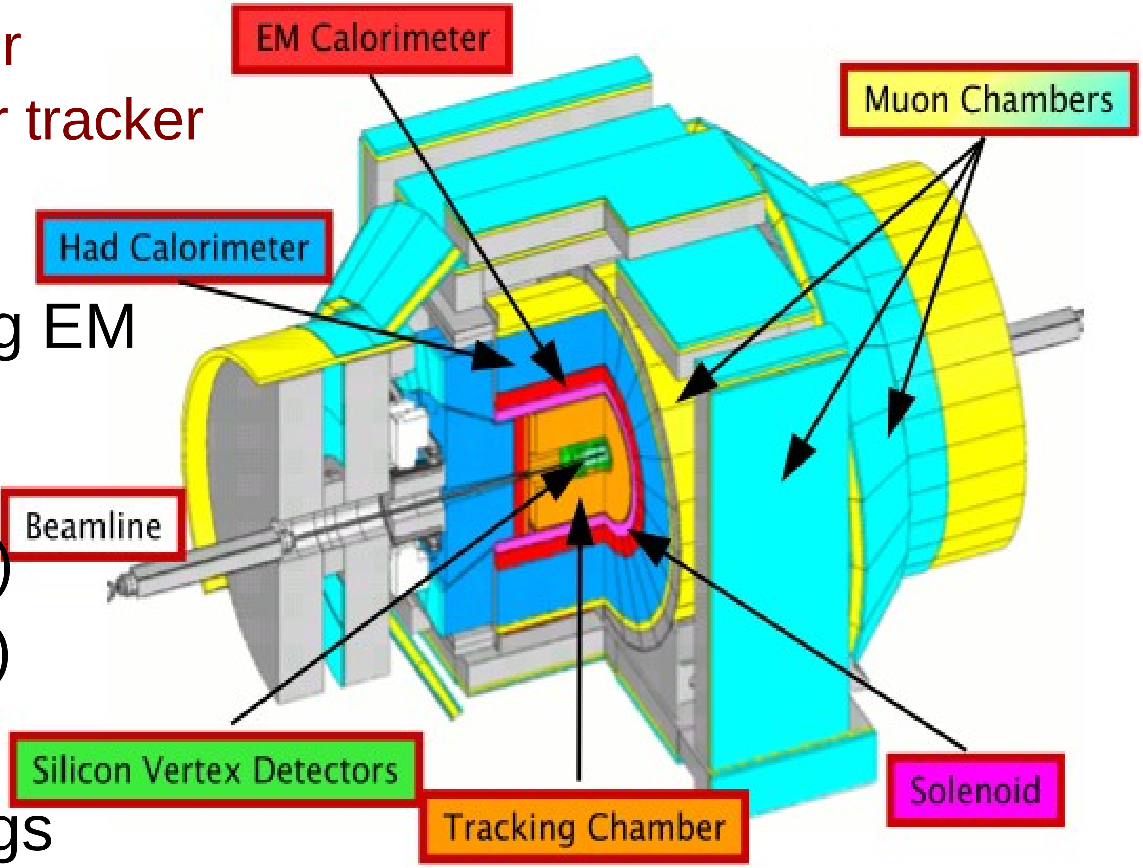




- CDF acquired luminosity ~7.4 fb^{-1}
  - Using 5.9 fb^{-1} for today's result

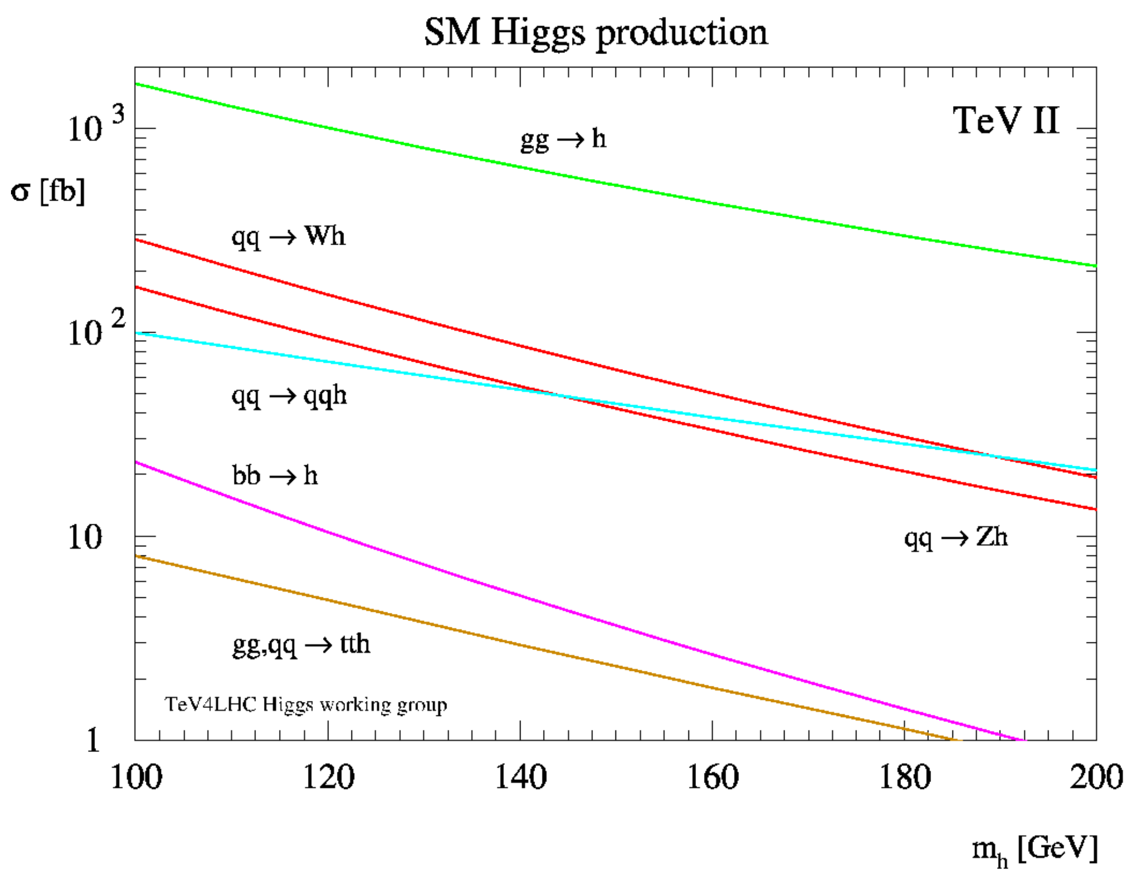
The CDF II Detector

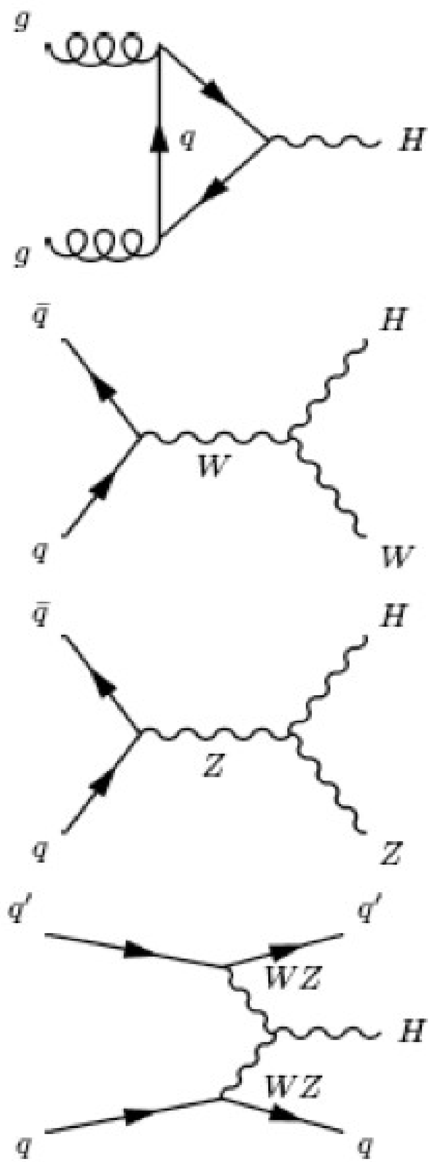
- General multipurpose detector
  - Excellent tracking and mass resolution:
    - Silicon inner tracker
    - Drift chamber outer tracker
- Calorimeters
  - Segmented sampling EM and Hadronic
- Muon chambers
  - CMU/CMP (| | < 0.6)
  - CMX (0.6 < | | < 1.0)
- Complex geometry
  - Try to maximize Higgs acceptance



SM Higgs Boson at the Tevatron

- Four main production mechanisms
  - Gluon fusion is the dominant process



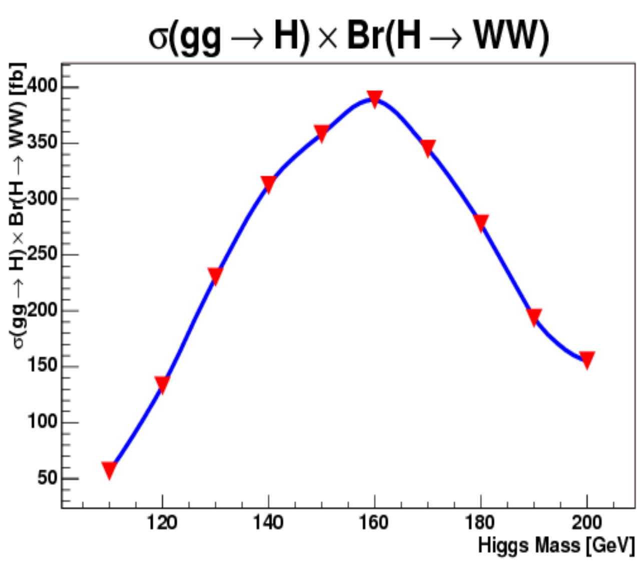


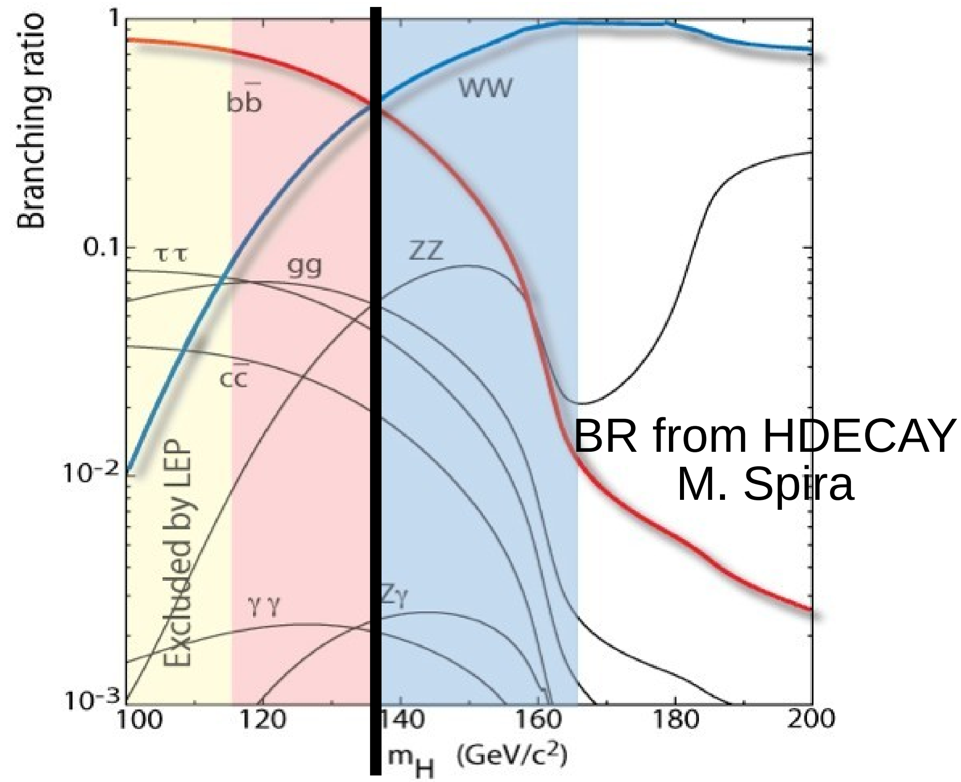
SM Higgs Boson Decay

- Higgs decay mode depends on Higgs mass m_H
  - Low Mass: H bb
  - High Mass: H WW
- For gg H WW,
  - Peak sensitivity m_H~165

Low Mass

High Mass





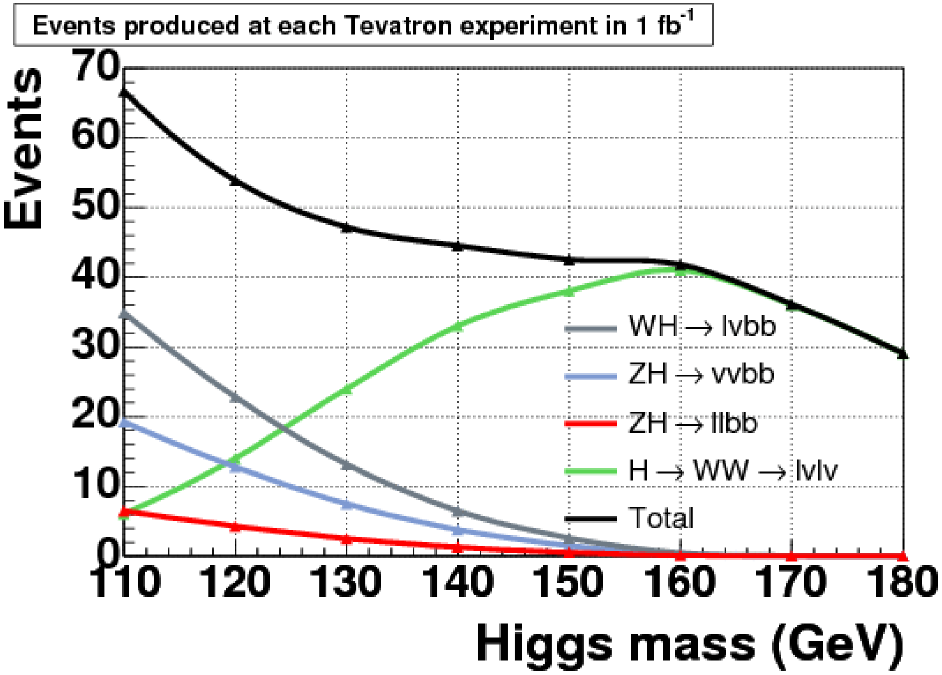
135 GeV

Analysis Strategy

- Final State Signature
- Cross Checks
- Signal Regions

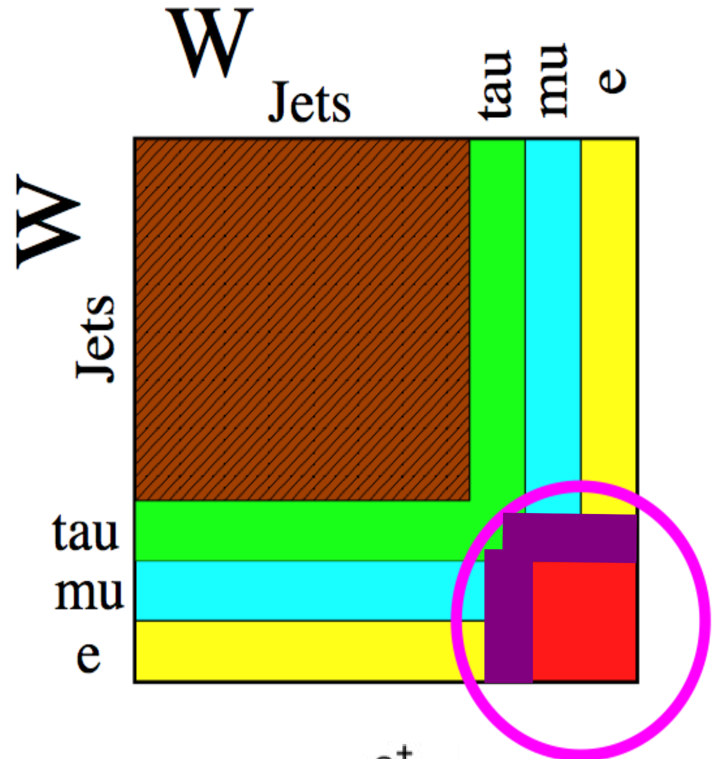
General Analysis Approach

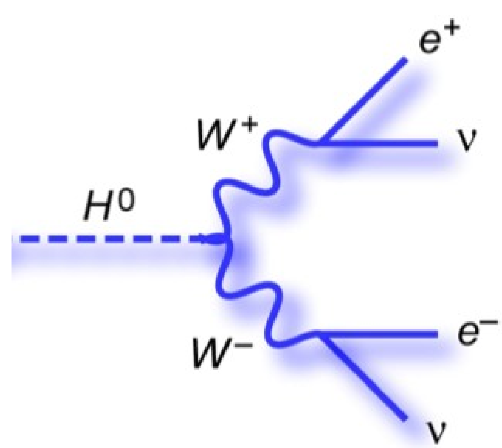
1. Select inclusive event sample that maximizes acceptance for Higgs signal
   - For m_H = 165 GeV, CDF reconstructs ~7 events per inverse fb
2. Model all backgrounds and cross check with data using control regions
3. Use advanced analysis tools to separate signal from background based on event kinematics



H WW Final States

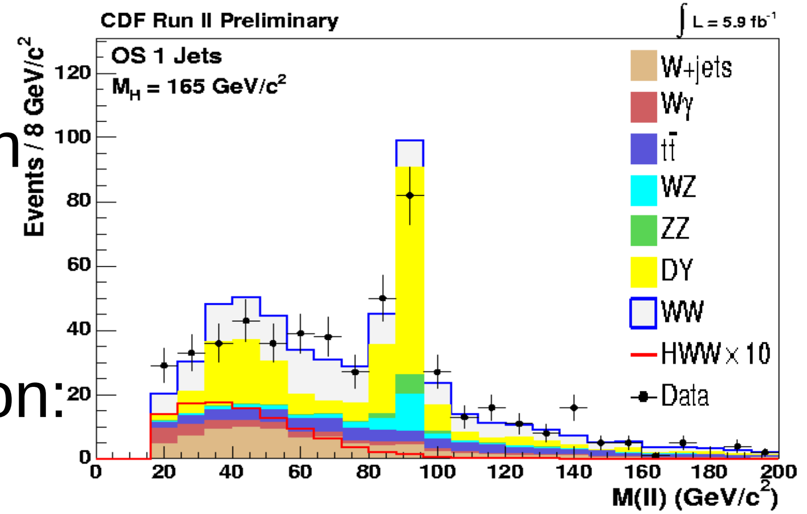
- BR(W hadrons) ~ 68%
  - Large QCD backgrounds
  - Investigate adding channels with one leptonic W and one hadronic W
- Dilepton (e,): BR ~ 6%
  - Sensitive to (e,)
  - Small BR, but...
    clean, easy to trigger
- Lepton + $_{had}$: BR ~ 4%
  - Recently added at CDF

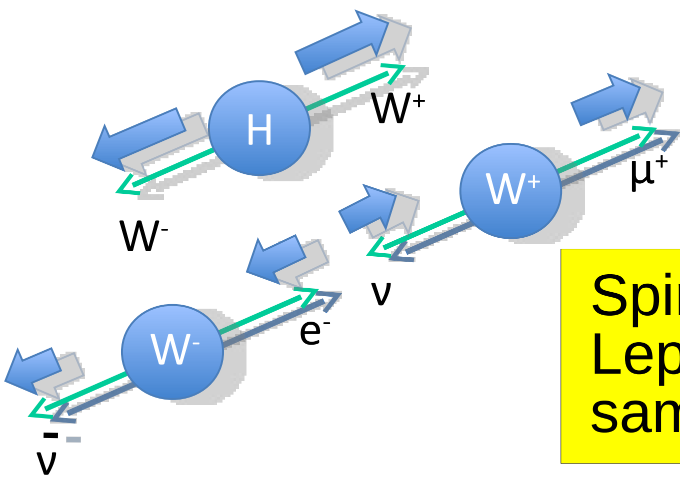




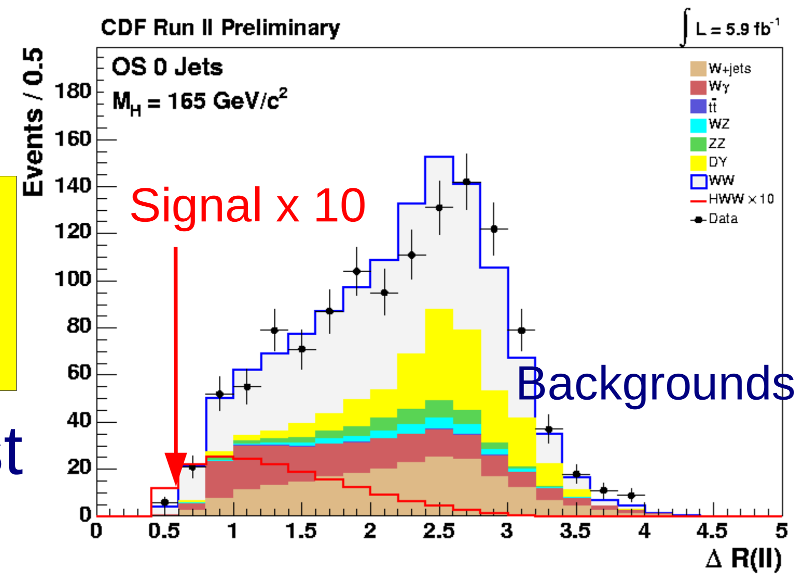
H WW ℓℓ Signature

- Decay kinematics
  - 2 opposite-sign leptons with high transverse momentum (p_T)
  - Missing energy from neutrinos
  - WW pair from spin-0 Higgs boson:





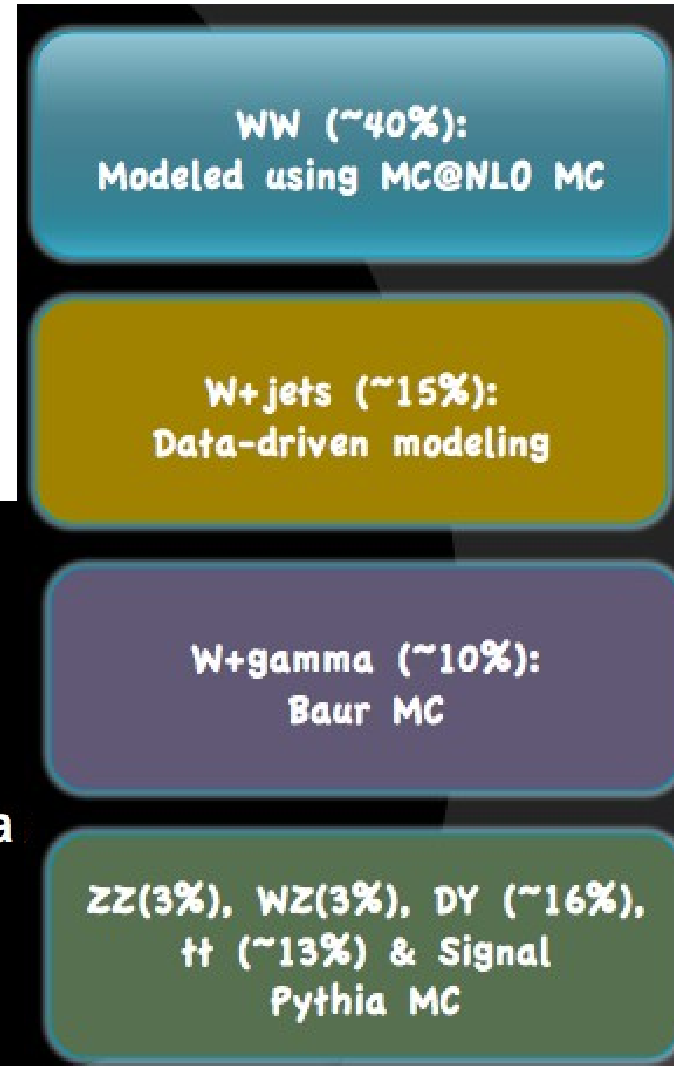
Spin correlation: Leptons go in the same direction

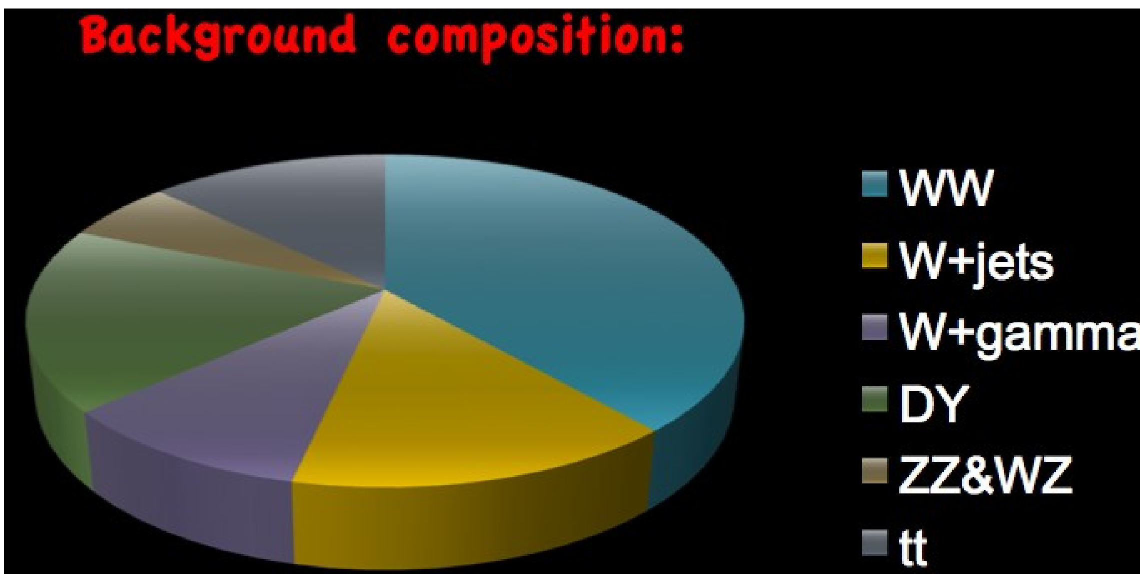


- Dilepton opening angle strongest background discriminant

H WW $\ell\ell$ Backgrounds

- SM processes with similar final states considered backgrounds
- All cross sections measured by Tevatron experiments
  - Many discovery analyses:
    - WW, WZ, ZZ





Analysis Strategy

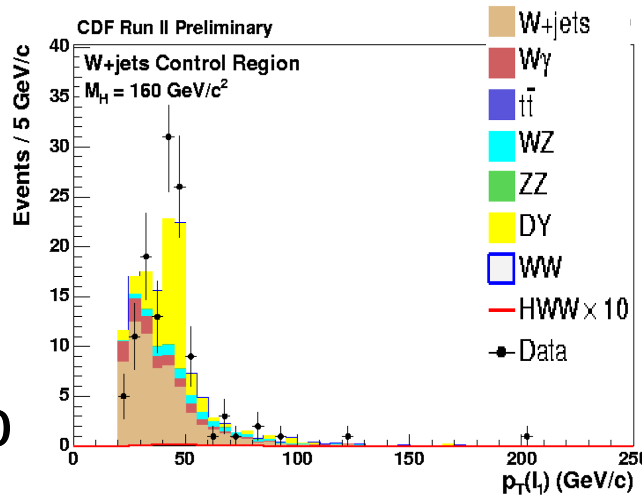
- Final State Signature
- Cross Checks
- Signal Regions

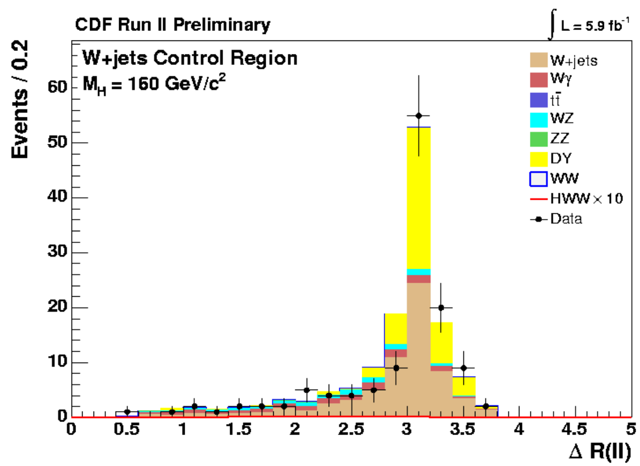
Strategy

- Simple event counting won't work
  - S/B = 0.015 in most sensitive search channel
- Use multivariate analysis (MVA) techniques to discriminate between signal and background
  - Matrix Elements (ME), Artificial Neural Networks (NN), Boosted Decision Trees (BDT)
  - Typically add 10-20% in sensitivity beyond that achieved using the best 1-2 variables
- Since we rely on kinematic shapes to separate potential signal from backgrounds, important aspect of these searches is how well we model these shapes
  - Specific control regions designed to test modeling of individual backgrounds (whenever possible)

W+jets Background

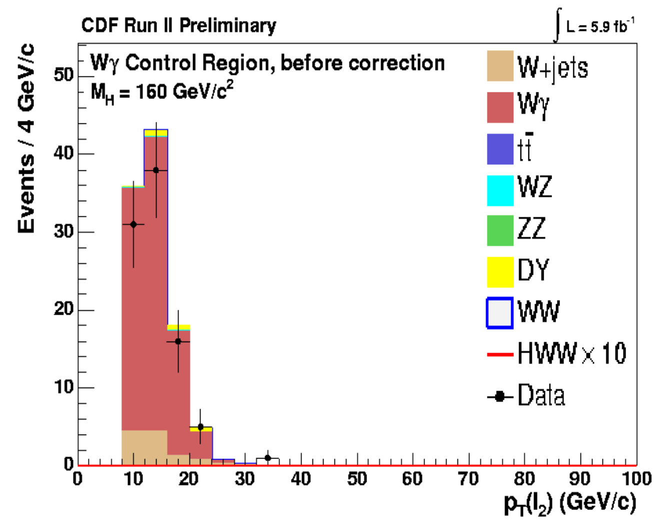
- W+jets events enter dilepton sample when the W decays leptonically, and a jet is misidentified as a lepton
- Model with data, not MC
  - Use jet-triggered data samples to measure rate at which jets are misidentified
- Check modeling in same-sign dilepton events with zero jets
  - Excellent modeling of kinematic variable shapes

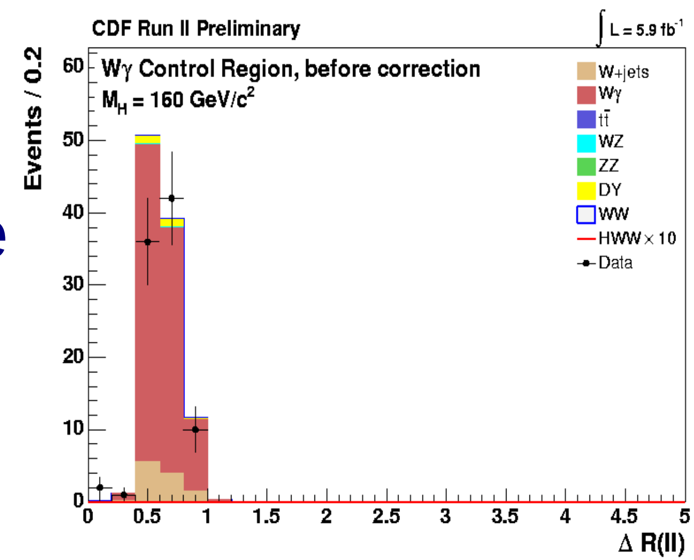




W Background

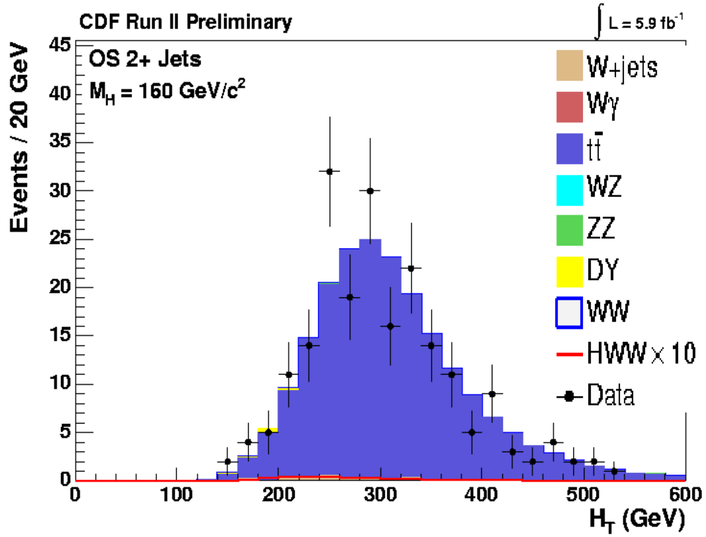
- W enters dilepton sample when W decays leptonically and photon converts in detector material
  - Modeled by Baur MC
- Powerful control region: same-sign leptons with dilepton invariant mass < 16 GeV
  - 90% composed of W
  - Above 16 GeV, W+jets dominate
- Control region is used to determine a scale factor for W normalization
  - Scale W by 0.87
  - Excellent modeling of kinematic variable shapes

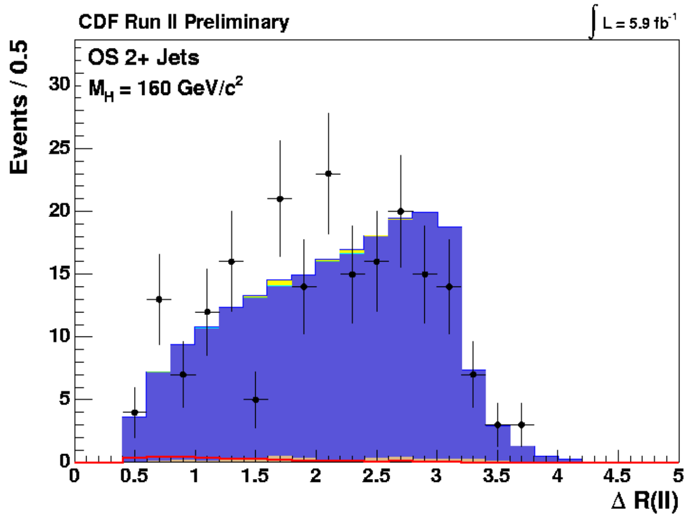




Top-quark Background

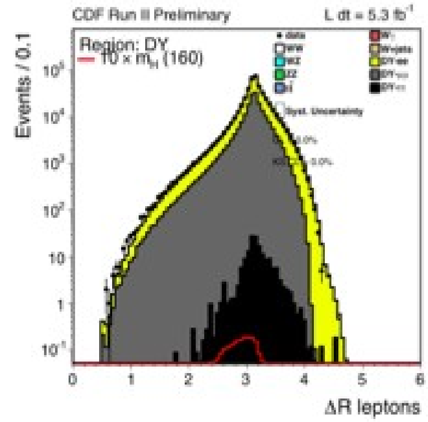
- Dominant background for opposite-sign dilepton events with two or more jets
  - Modeled by PYTHIA Tune A
- Remove events with b-tagged jets as a control region
  - Tight secondary vertex tagger
  - Almost entirely top-quark pairs
- Measure $t\bar{t}$ cross section consistent with theory and CDF top-quark measurements
  - Excellent modeling of kinematic variable shapes





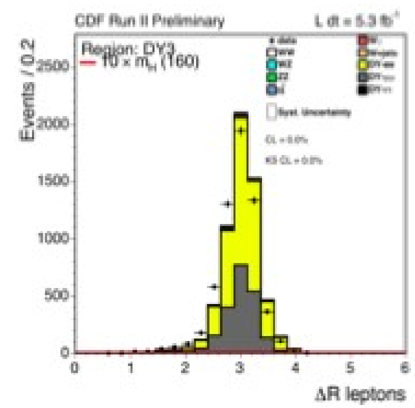
Drell-Yan Background

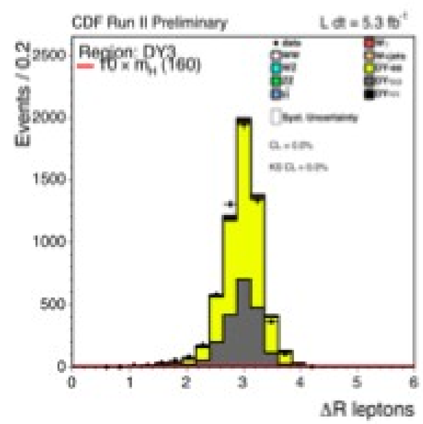
- Model Z $\ell\ell$ (ℓ = e, ,) with PYTHIA
  - Good match with inclusive Z p_T (boost) observed in data
- However, requiring missing E_T leads to disagreement between data and MC
  - Due to mismodeling of the underlying event and jets
- Correct MC in intermediate missing E_T range
  - Obtain good modeling of kinematic variable shapes



Inclusive Sample

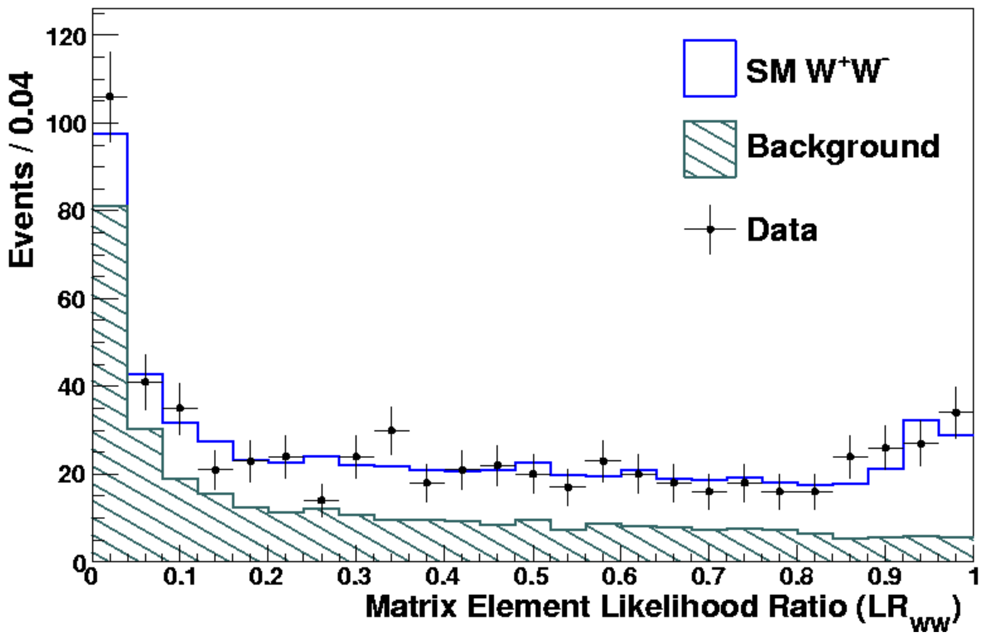
Intermediate missing E_T region





After tuning

WW Cross Section



PRL 104, 201801 (2010)

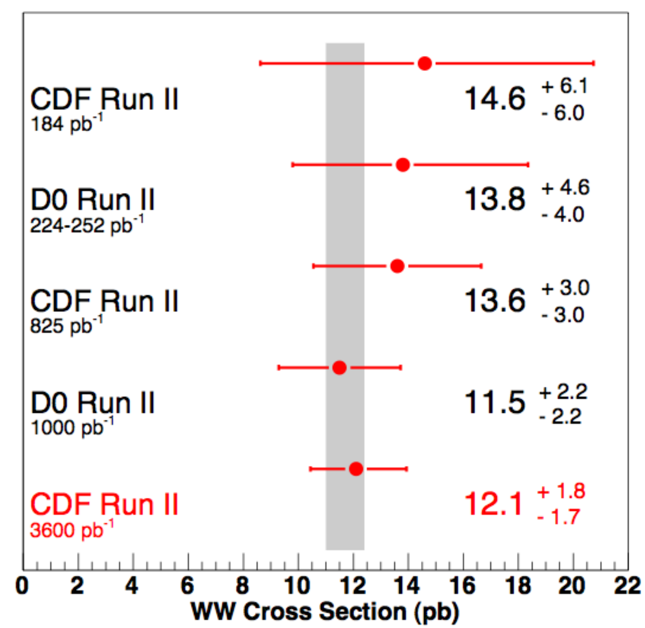
World's best measurement!

- Good agreement with theory

$$\sigma(p\bar{p} \rightarrow WW) = 12.1 \pm 0.9\ (\text{stat.})^{+1.6}_{-1.4}\ (\text{syst.})\ [\text{pb}]$$

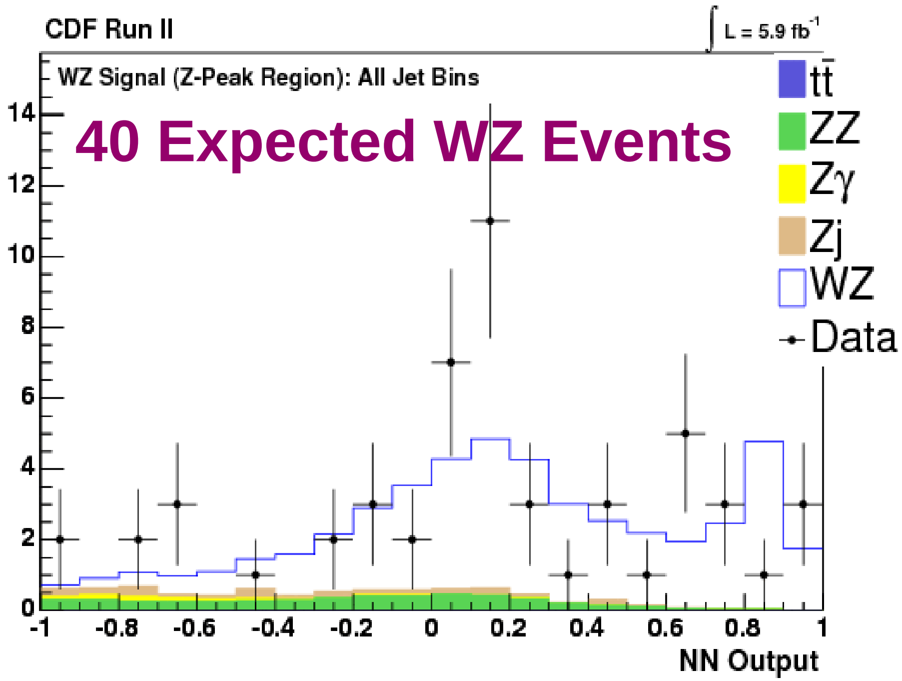
Syst. **includes** 5.9% luminosity uncertainty

- Measure WW cross section in 0 jet signal region
  - Two opp-sign leptons, high missing energy
- Binned maximum likelihood fit to ME likelihood ratio distribution

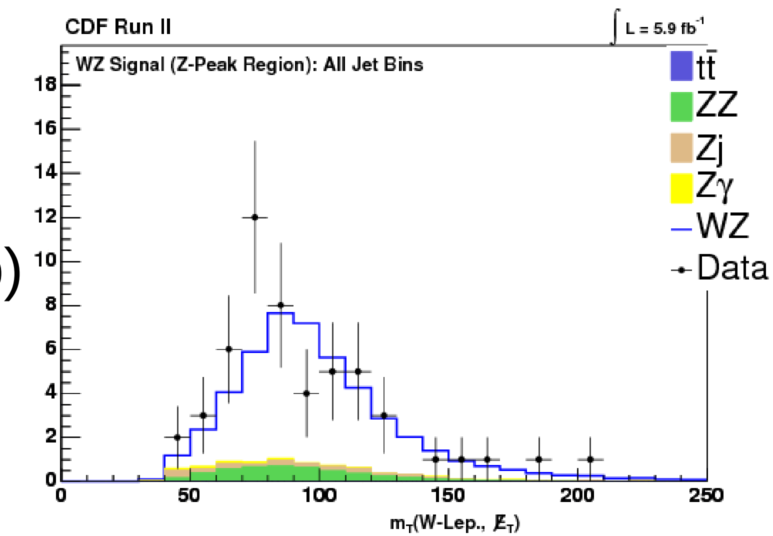


WZ Cross Section

New Result!



- Measure WZ cross section in trilepton signal region
  - WZ ℓ ℓℓ
- Use NN to separate WZ
- Binned max. likelihood f t to NN template



World's best measurement!

- Good agreement with theory (3.46 pb)

$$\sigma(p\bar{p} \to WZ) = 3.7 \pm 0.6(\mathrm{stat})^{+0.6}_{-0.4}(\mathrm{syst})(\mathrm{pb})$$

Analysis Strategy

- Final State Signature
- Cross Checks
- Signal Regions

Summary of Signal Regions

Channel	Main Signal	Main Background	Most important kinematic variables
OS dileptons 0-jets	gg H	WW	LR_{HWW}, R_{ll}, H_T
OS dileptons 1-jet	gg H	WW, DY	R_{ll}, $M_T(ll,\not{E}_T)$, $\not{E}_T$
OS dileptons 2+ jets	Mixture	$t\bar{t}$	H_T, P_{ll}, M_{ll}
OS dileptons low M_{ll}, 0+1 jets	gg H	W	$p_T(l_2)$, $p_T(l_1)$, $E(l_1)$
SS dileptons 1+ jets	WH	W+jets	N_{jets}, $\not{E}_T$ signif, H_T
Trileptons, no Z-cand, all jets	WH	WZ	$M_T(l,\not{E}_T)$, R_{ll}^{close}, Flavor
Trileptons, Z-cand and 1-jet	ZH	WZ	$\not{E}_T$, R_{ll}(W-lep,j), $E_T(j)$
Trileptons, Z-cand and 2+ jets	ZH	WZ, Z+jets	R_{ll}(W-lep,j), M_{jj}, M_W
OS dilepton, e + hadronic	gg H	W+jets	R_l , id variables
OS dilepton, + hadronic	gg H	W+jets	R_l , id variables

No Channel Left Behind

Summary of Signal Regions

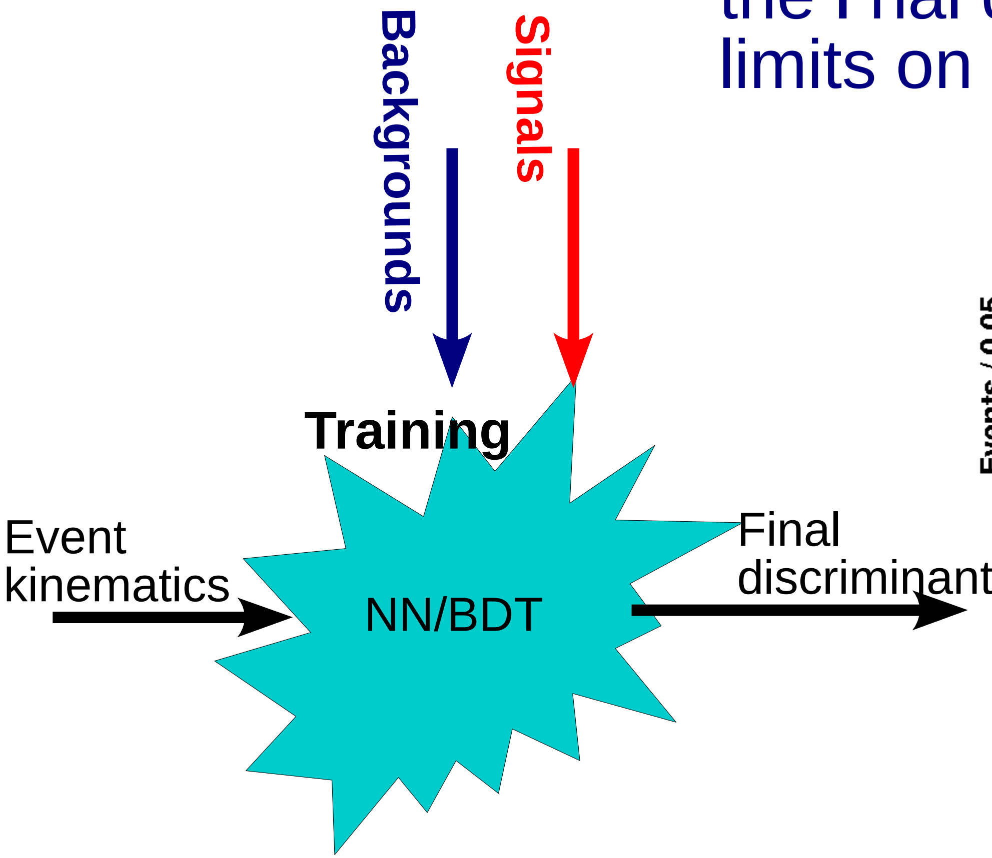
Channel	Main Signal	Main Background	Most important kinematic variables
OS dileptons 0-jets	gg H	WW	LR_{HWW}, R_{ll}, H_T
OS dileptons 1-jet	gg H	WW, DY	R_{ll}, $M_T(ll,\not{E}_T)$, $\not{E}_T$
OS dileptons 2+ jets	Mixture	$t\bar{t}$	H_T, P_{ll}, M_{ll}
OS dileptons low M_{ll}, 0+1 jets	gg H	W	$p_T(l_2)$, $p_T(l_1)$, $E(l_1)$
SS dileptons 1+ jets	WH	W+jets	N_{jets}, $\not{E}_T$ signif, H_T
Trileptons, no Z-cand, all jets	WH	WZ	$M_T(l,\not{E}_T)$, R_{ll}^{close}, Flavor
Trileptons, Z-cand and 1-jet	ZH	WZ	$\not{E}_T$, R_{ll}(W-lep,j), E_T(j)
Trileptons, Z-cand and 2+ jets	ZH	WZ, Z+jets	R_{ll}(W-lep,j), M_{jj}, M_W
OS dilepton, e + hadronic	gg H	W+jets	R_l , id variables
OS dilepton, + hadronic	gg H	W+jets	R_l , id variables

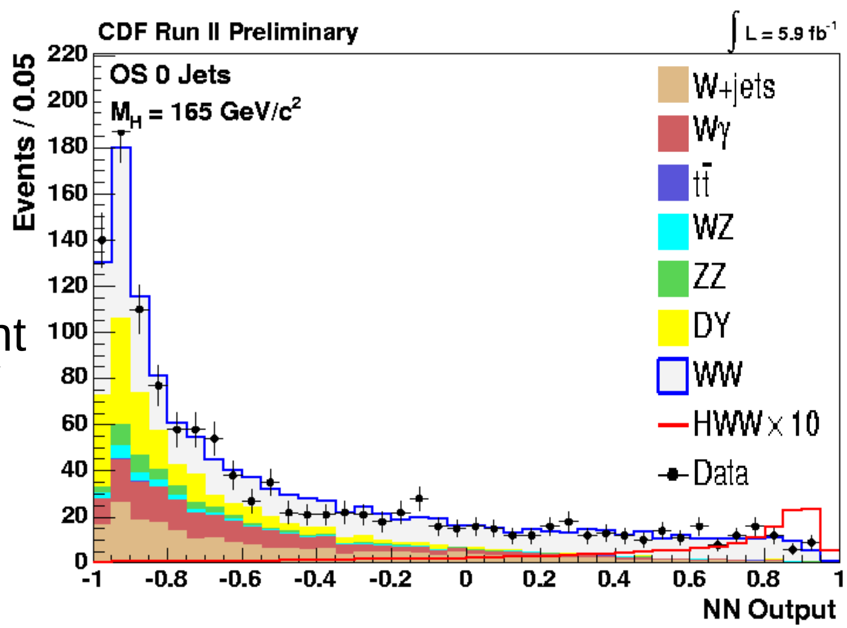
No Channel Left Behind

See S. Jindariani's W&C for details

Final Discriminants

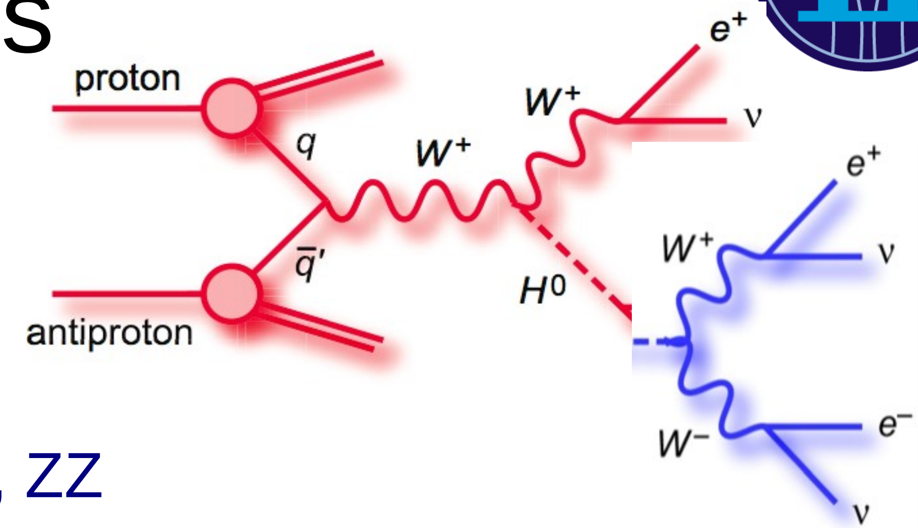
- NN or BDT outputs (templates) are the final discriminant used to set limits on Higgs production





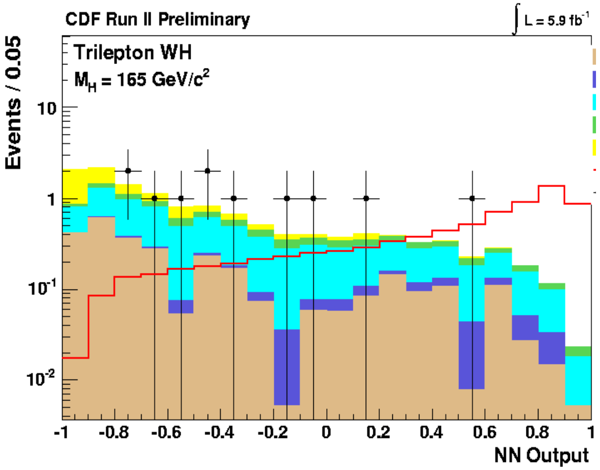
Trilepton Searches

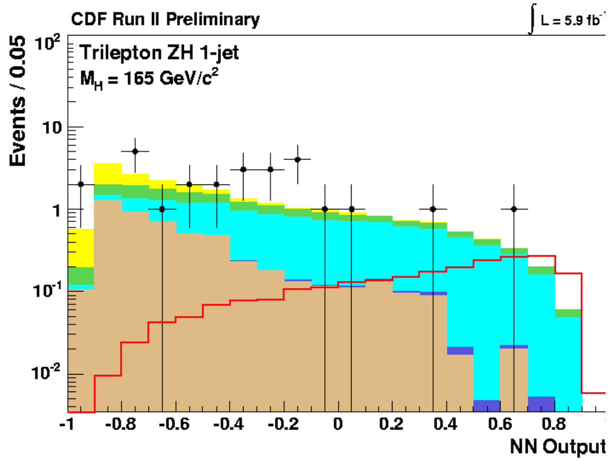
- Trilepton signature occurs in associated production
  - WH   WWW   ℓ ℓ ℓ
  - ZH   ZWW   ℓℓℓ + X

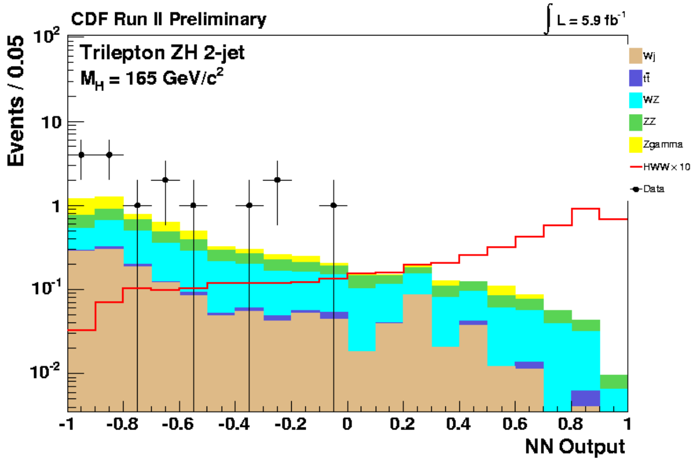


- Dominant background is WZ, ZZ
- Divide events by whether two opposite-sign, same-flavor leptons form a Z-candidate
  - Isolate WH and ZH signal regions

Expected sensitivity at 165 GeV ~4.6xSM

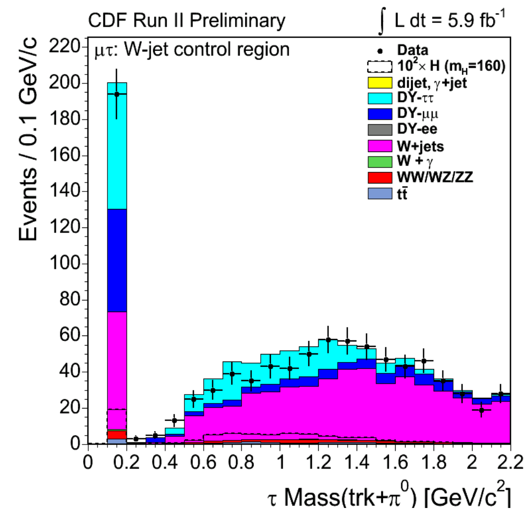


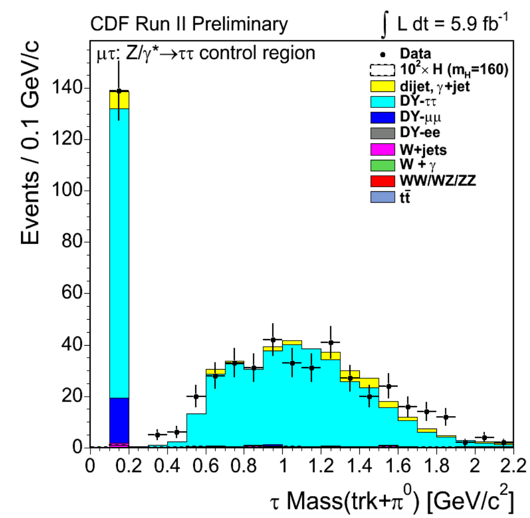


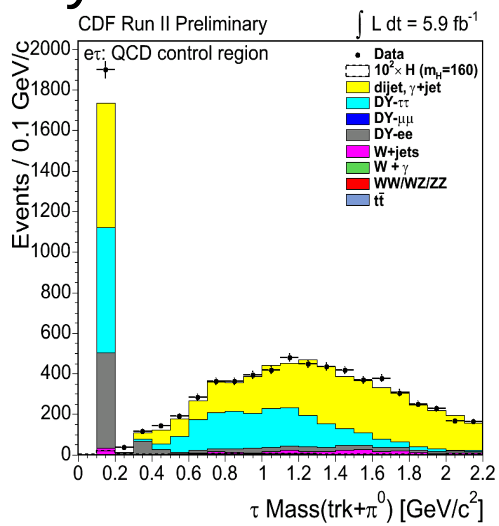


Hadronic Tau Backgrounds

- Very different mix of backgrounds for events with one hadronically decaying tau lepton
  - QCD and Z backgrounds
- Unique: rely on ID variables as well as kinematics to discriminate between signal and background
  - Need cross checks to verify both
- Form orthogonal control regions to study:
  - W+jets (both e and)
  - QCD (for e)
  - Z (for)

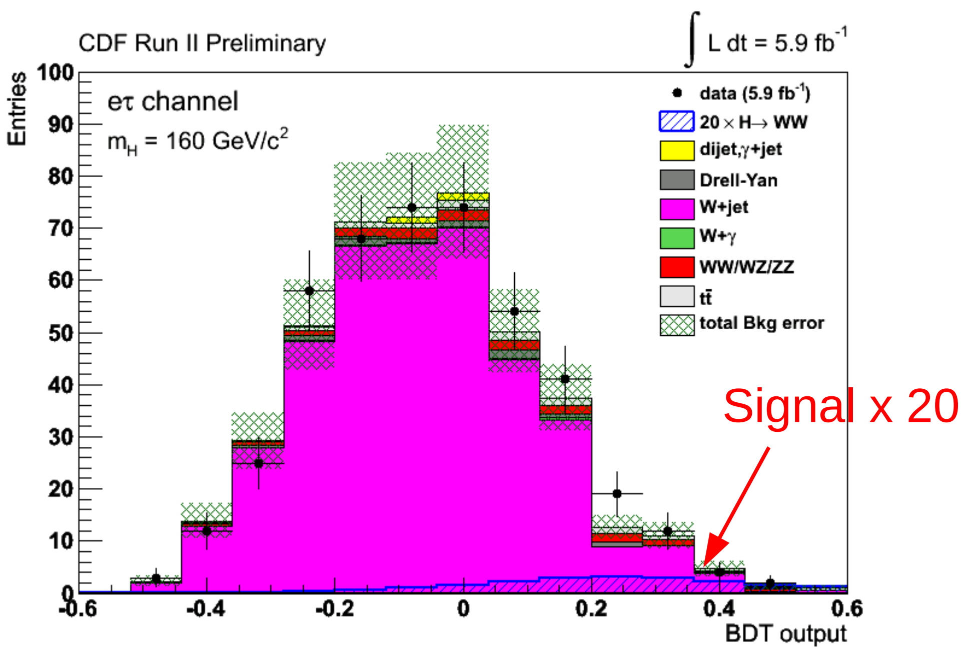


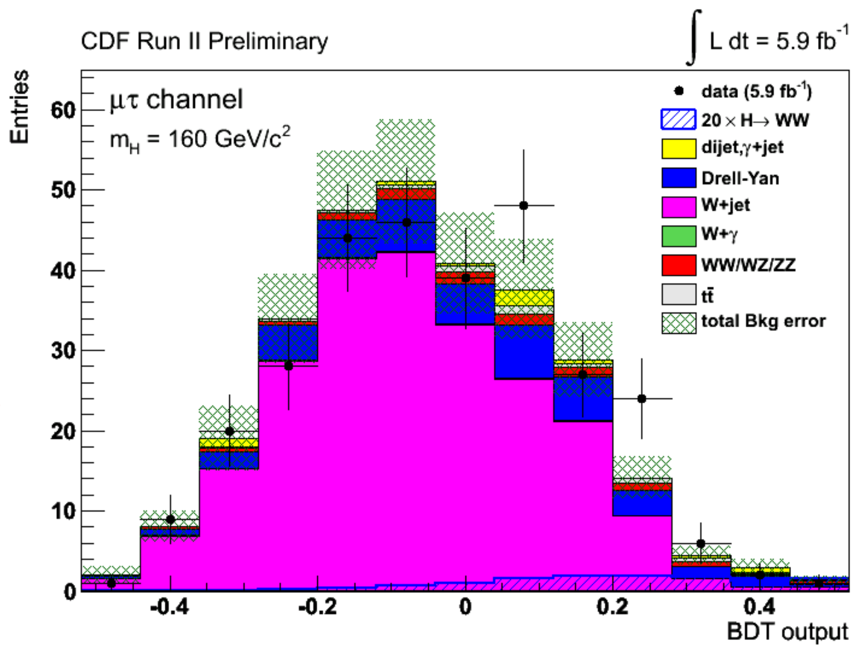




Hadronic Tau Searches

- Dominant background W+jets
  - Modeled by ALPGEN MC instead of with data
- Use different MVA technique
  - Boosted Decision Trees instead of NN
- Overall good modeling in both e and
- Expect ~1.5 signal events, ~730 backgroud
  - Expected sensitivity at 165 GeV is ~15 x SM





Improvements since March 2010

- Updated all search channels to 5.9 fb^{-1}
- Drell-Yan missing E_T correction
- Tightened electron selection for the same-sign dilepton search
  - Reduces W+jets events with minimal impact on signal
- New WW MC@NLO sample
  - Old sample had limited statistics
  - Updated to latest version of MC@NLO and generated 10x more statistics
- Improved treatment of systematic uncertainties
  - More sophisticated determination of both rate and shape uncertainties

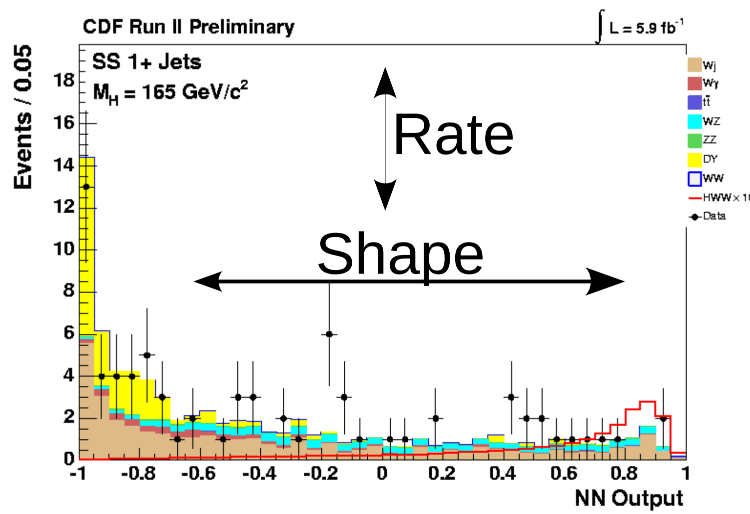
Theoretical and Systematic Uncertainties

- Overview
- Signal Uncertainties
- Background Uncertainties

H WW $\ell\ell$ Systematics

Systematic (%)	Sig	Bkgs
Cross section	5-12	5-10
Conversions	0	10-20
NLO diagrams	3-10	10
PDF model	3-12	1-5
Jet energy scale	5-10	1-30
Lepton ID	2	2
Trigger eff ciency	2-3	2-3
Luminosity	6	6

- Two classes of systematics:
  - Rate
    - Affect only normalization
  - Shape
    - Modify output of discriminant



Determination of Uncertainties

- Two main categories of systematics
  - Cross Section: theoretical uncertainty on the production cross section for a process
    - Rate systematic only
  - Acceptance: uncertainty on our modeling of the acceptance or kinematic variables for a process
    - Rate and shape systematics
- For today, touch on the main signal and background uncertainties
  - Gluon fusion signal: theoretical uncertainties affect both cross section and acceptance
  - Also look at theory uncertainties for the WW background
  - Example of shape effects for the jet energy scale

Theoretical and Systematic Uncertainties

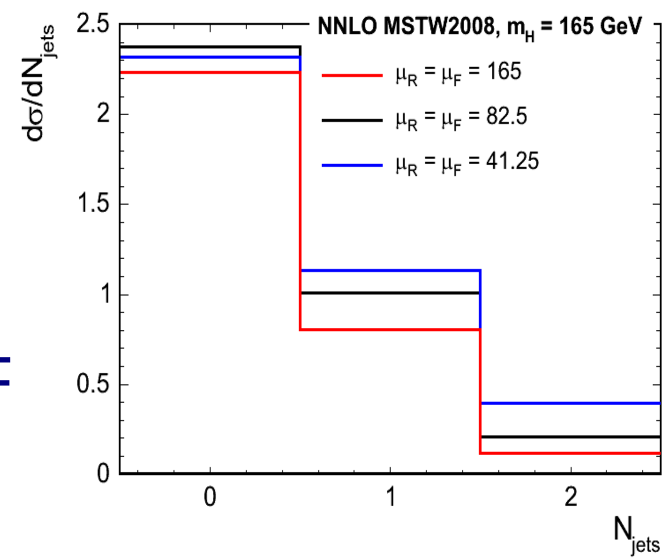
- Overview
- Signal Uncertainties
- Background Uncertainties

Gluon Fusion Cross Section

- Limits depend significantly on theoretical Higgs production cross sections
  - Gluon fusion, the dominant production process, has the largest uncertainties!
- Currently use inclusive cross section calculations of de Florian and Grazzini (arXiv:0901.2427v2)
  - Soft-gluon resummation to NNLL
  - Proper treatment of b-quarks to NLO
  - Inclusion of two-loop electroweak effects
  - MSTW2008 Parton Density Functions
- In good agreement with calculations of Anastasiou, Boughezal, and Petriello (arXiv:0811.3458v2)
  - Fixed-order calculation up to NNLO

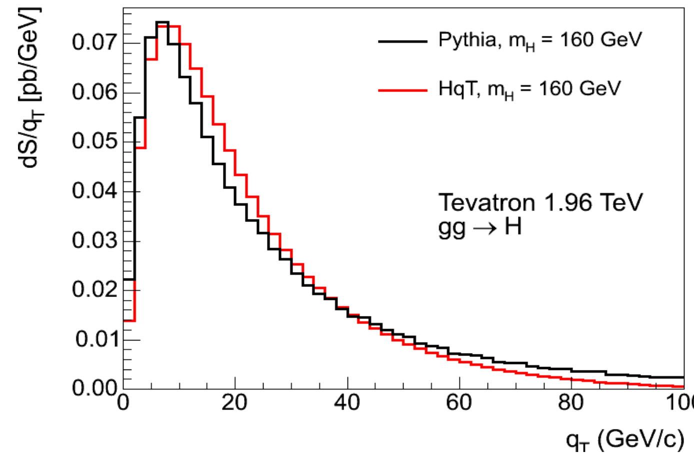
Cross Section Uncertainties

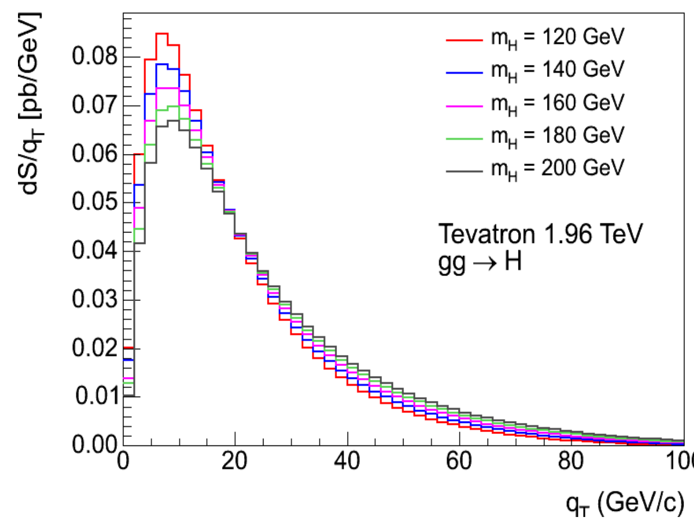
- Dominant sources of theoretical uncertainty:
  - Higher-order QCD radiative corrections (Scale)
  - Parton density functions (PDF)
- Because we separate on number of reconstructed jets, must determine topology-dependent scale factors
- Estimate scale uncertainties by varying renormalization and factorization scales between $m_H/4$ and m_H
  - $m_H/2$ is central value for fixed-order
- Use MSTW2008 alternative error sets which vary both $_s$ and 20 PDF fit parameters



Modeling Gluon Fusion Signal

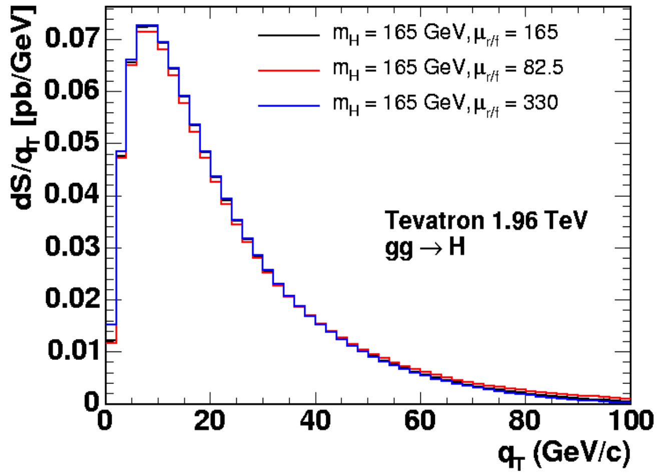
- Use PYTHIA, which is LO but has its own mechanisms for including effects of soft gluon radiation
  - Generate samples in 5 GeV steps from 110 up to 200
- Because kinematics are important, re-weight PYTHIA events at generator-level to match Higgs p_T spectrum obtained from full NNLL calculation
  - Self-consistent with normalizing to NNLL inclusive cross section
- Signal acceptance is determined from re-weighted sample

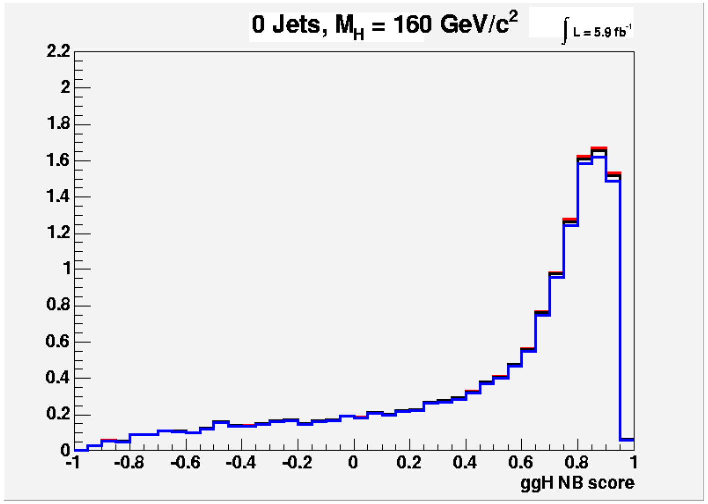




Acceptance Uncertainties

- We assign scale and PDF uncertainties on the acceptance, in addition to the cross section
  - Quantify variations in Higgs p_T and rapidity spectra as a function of scale and PDF choices
  - Apply additional reweightings until PYTHIA samples match variations
  - Assign uncertainties based on observed changes in signal acceptance by channel
- Also allows us to assign shape uncertainties to signal templates



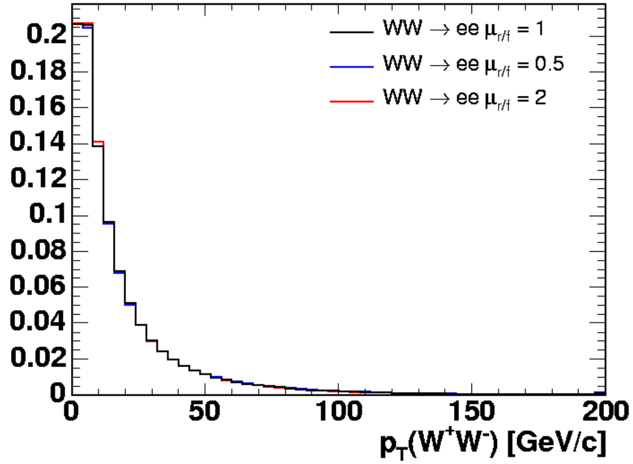


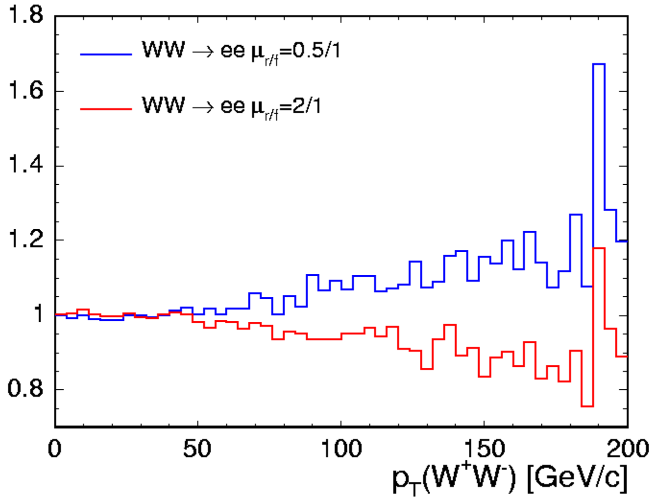
Theoretical and Systematic Uncertainties

- Overview
- Signal Uncertainties
- Background Uncertainties

WW Uncertainties

- Dominant background in most regions where gluon fusion is dominant signal
  - Want to model kinematics as well possible
  - Use NLO Monte Carlo: MC@NLO
- Treat uncertainties in same manner as for gluon fusion
  - Use the WW p_T and rapidity spectra to re-weight and assign uncertainties based on changes in acceptance

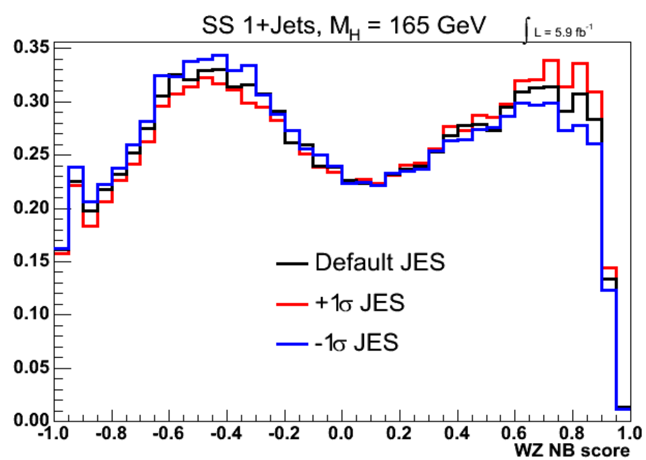


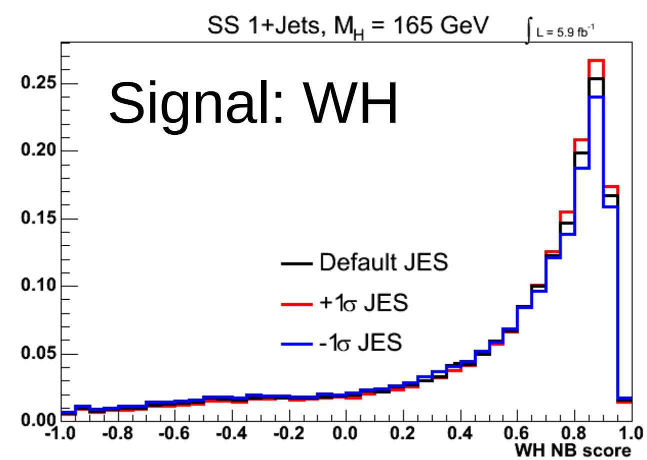


Other Shape Systematics: JES

- Negligible shape effects in regions where we separate by jet multiplicity or use all jets
  - Rate uncertainty: moves events between jet bins
  - Affects both backgrounds and signals
- Shape effects come in regions which reject events based on jet multiplicity
  - Same-sign dileptons: remove 0 jet events

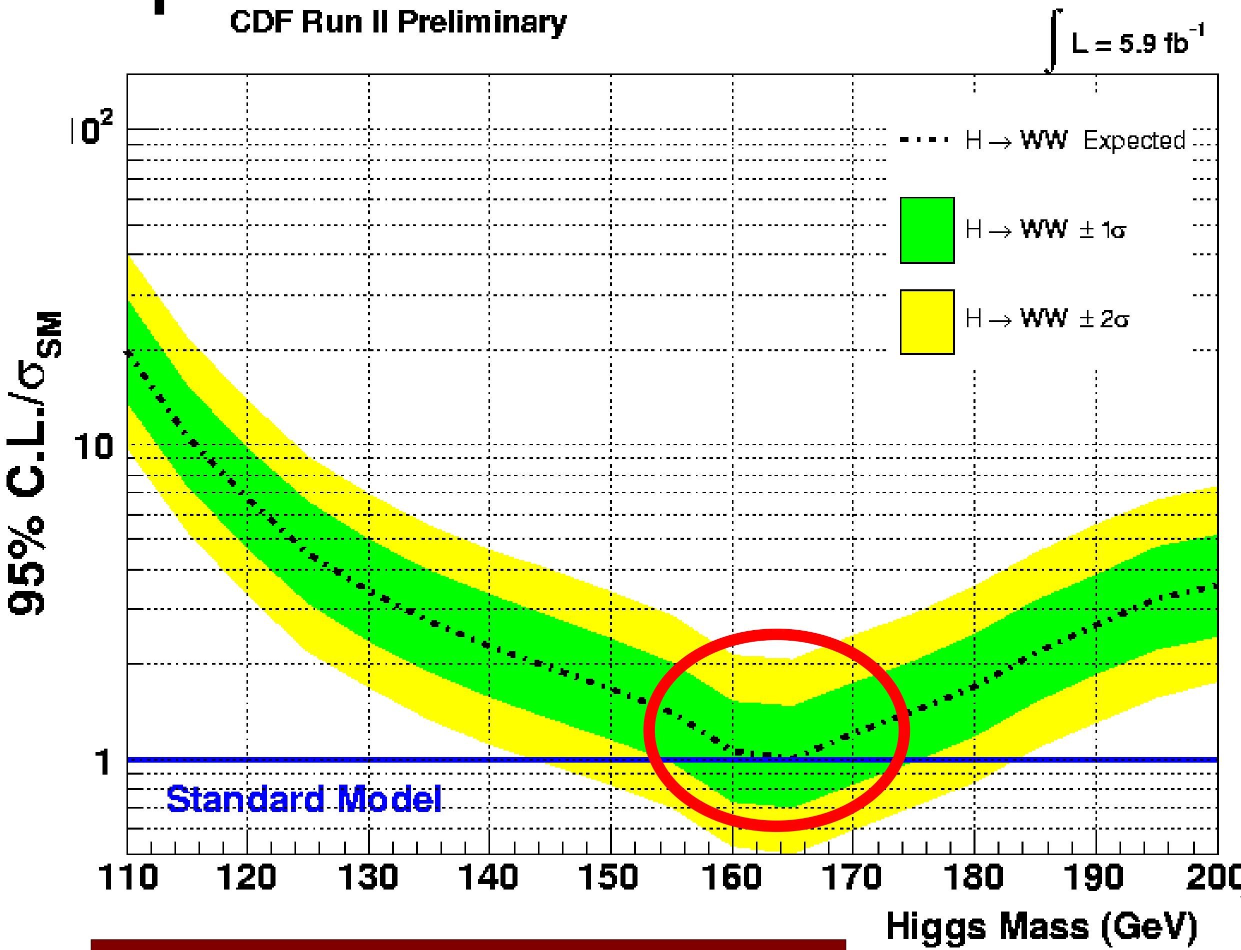
Background: WZ





Results

Updated CDF H WW Result

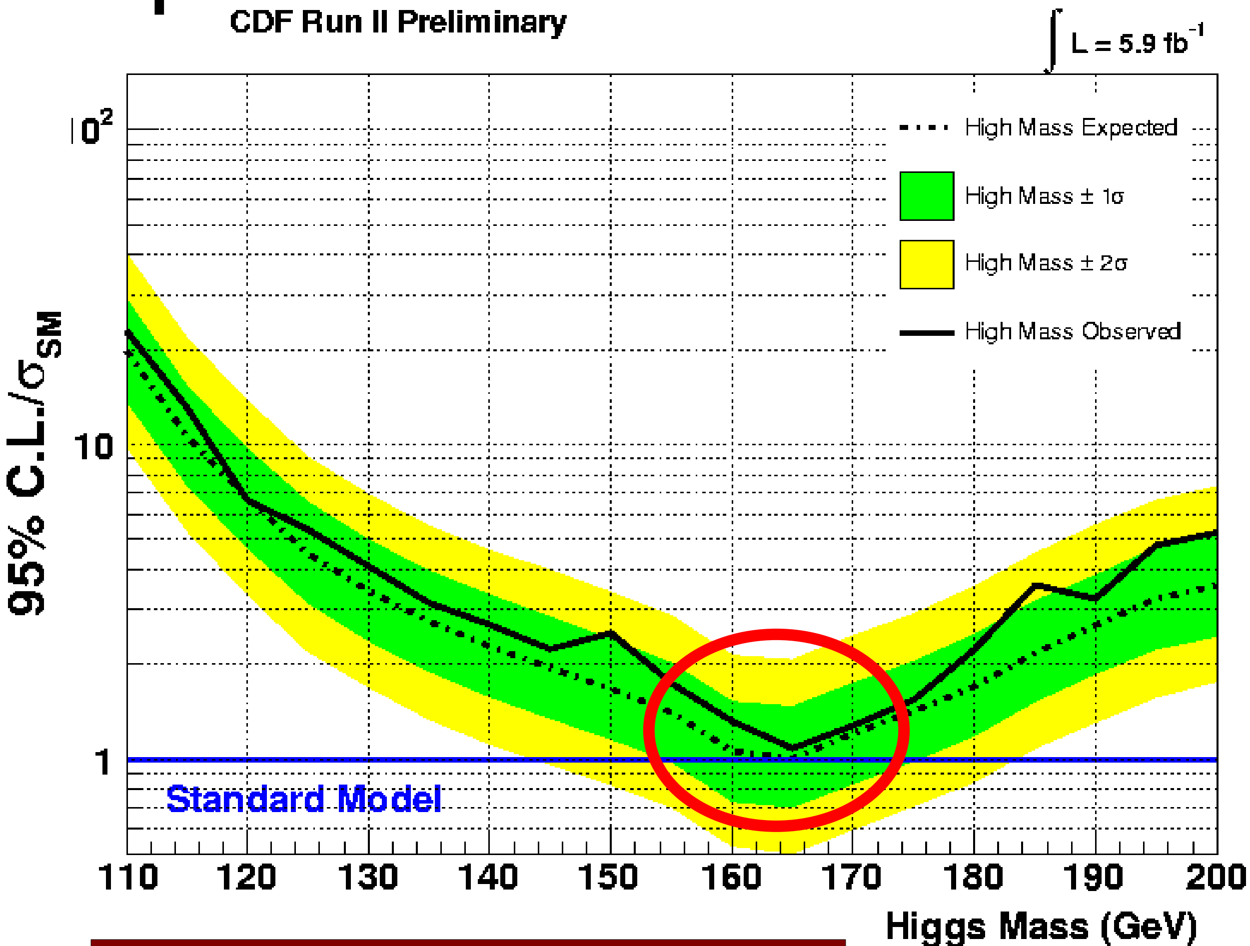


- Expected ~5% in sensitivity from adding luminosity
- Additional systematics reduced this to ~2-3%

Reaching SM sensitivity with a single experiment!

Expected Limits at
160: 1.05 x SM
165: 1.00
170: 1.20

Updated CDF H WW Result

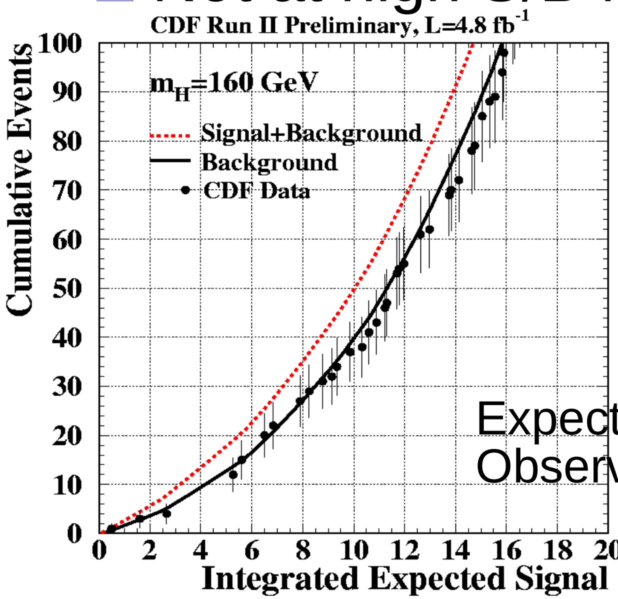


- Observed limit is slightly higher than expected over the mass range
- Previously, expected and observed followed each other closely

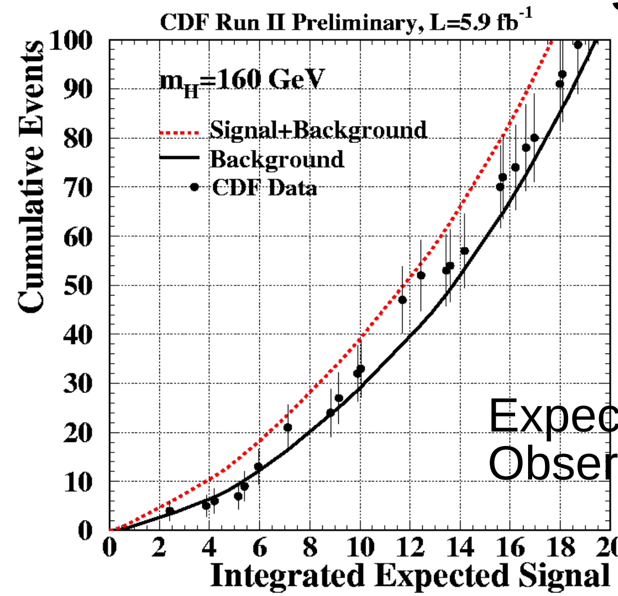
Observed (Expected) Limits at
160: 1.32 (1.05)
165: 1.08 (1.00)
170: 1.28 (1.20)

Observed limits at 160 GeV

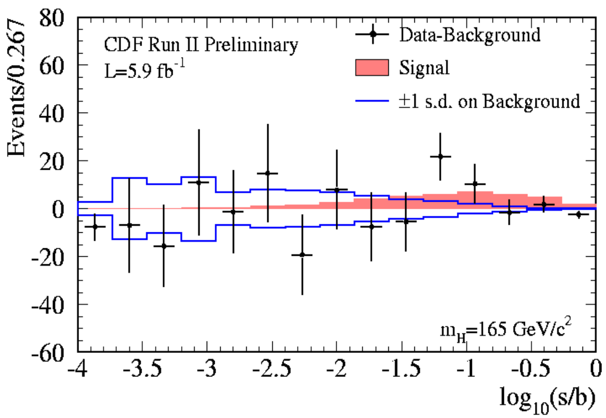
- Number of events in data increased more than expected
  - Not at high S/B NN output, but over the entire range



Expected 1.26
Observed 1.27

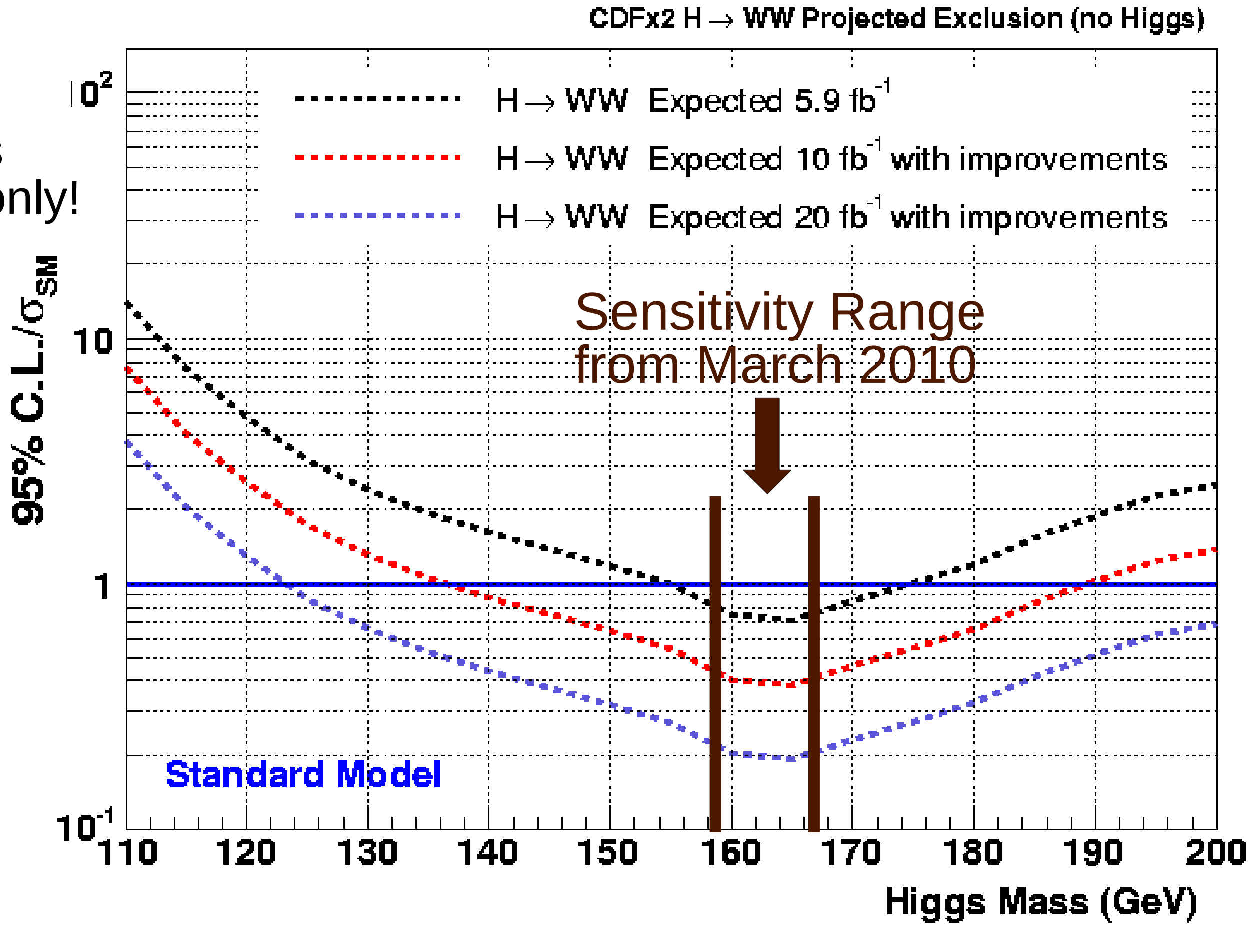


Expected 1.05
Observed 1.32



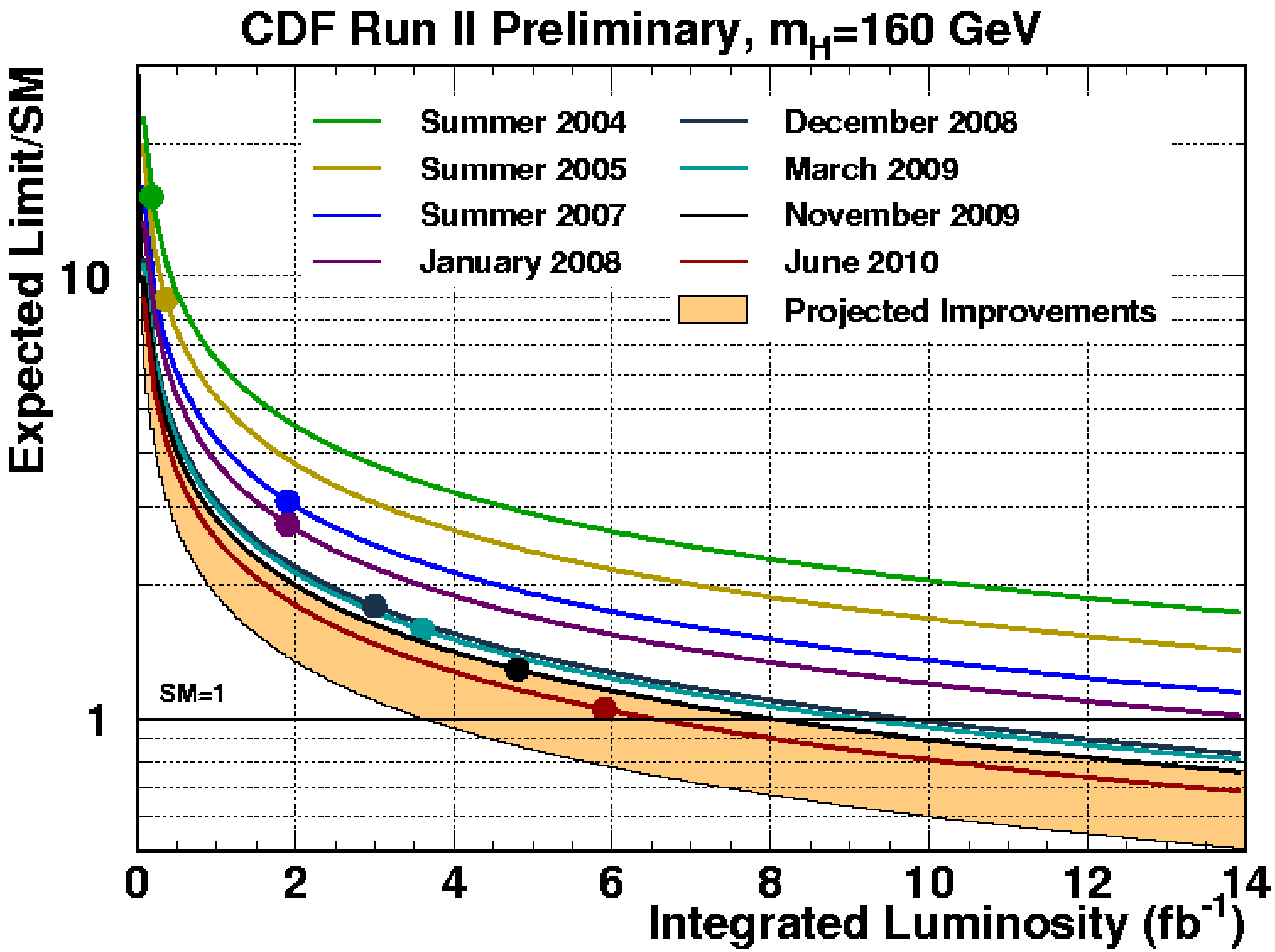
Sensitivity Projections

High Mass
analyses only!



- Improvements will push Tevatron sensitivity!

Projected Improvements

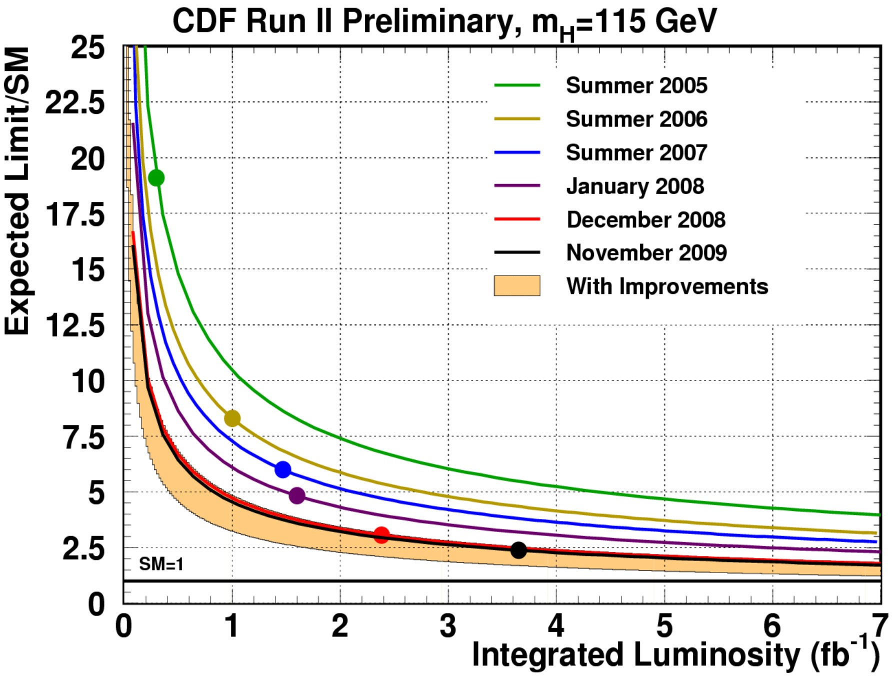


How to Improve?

- Add more acceptance
  - New search channels:
    - H   WW ℓ jj (in progress)
  - Addition of lower p_T leptons and triggers
  - Investigate loosening isolation cut on leptons
    - Higgs leptons very close together, could lie in each other's isolation cones (especially for low M_{ll} events)
    - Need to understand rate of lepton fakes with isolation
- Improve analysis techniques
  - Still many ideas!
    - Optimize neural networks for low/high mass separately, improve missing E_T description, study lepton isolation...

Summary

- Exciting times for Higgs boson searches!
- Tevatron making great strides in high mass searches
  - Sensitivity continues to improve faster than luminosity scaling
  - Low mass searches also approaching SM sensitivity
    - At m_H = 115, 2.4 x $_{SM}$
  - Soon “high mass” will become important to probe intermediate mass range
- Current Tevatron exclusion in the Higgs mass range 162-166 GeV
  - More to come!

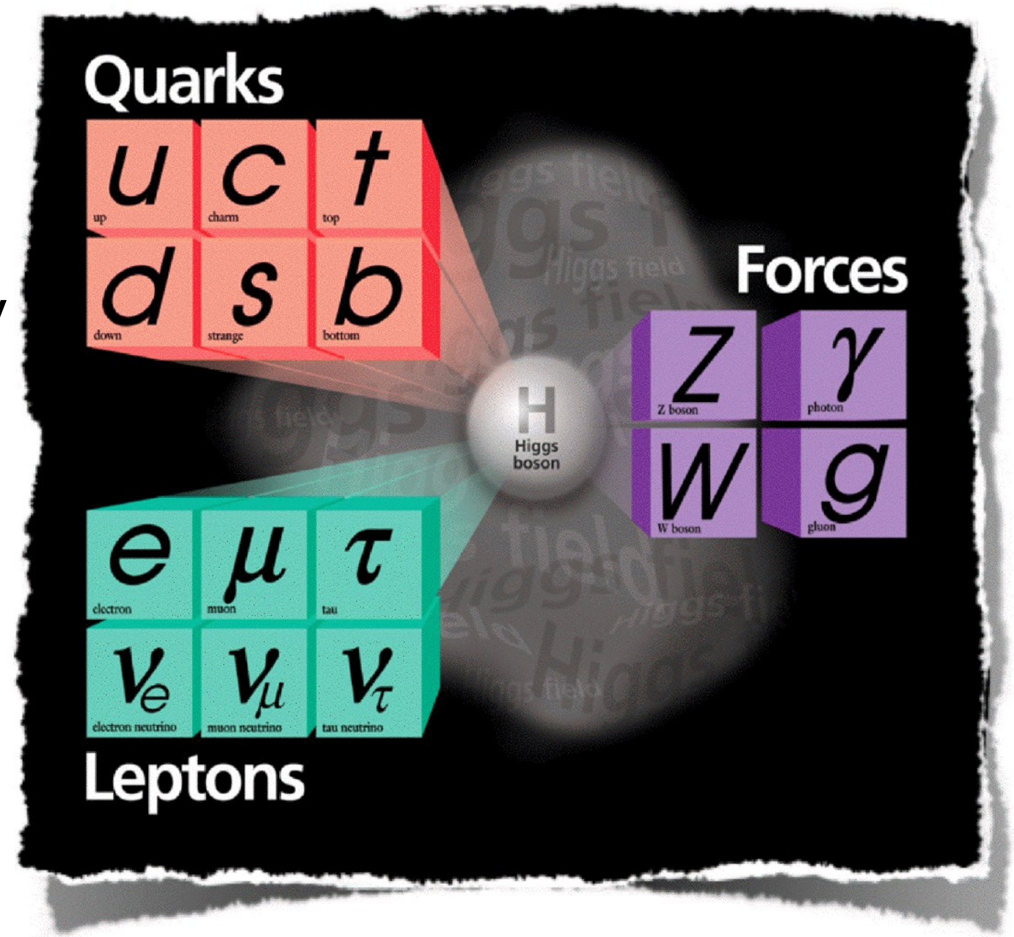


- To Fermilab and the Accelerator Division for providing the data
- To the CDF Collaboration for collecting the data with high efficiency
- And especially, the dedicated members of the H WW analysis group at CDF for analyzing the data!
  - Particularly Massimo Casarsa, Eric James, Sergo Jindariani, Thomas Junk, Jason Nett, Rick St. Denis, Geumbong Yu

Extra Slides

Standard Model of Particle Physics

- At high energies, weak and electromagnetic forces can be unif ed into one force – electroweak
  - But at low energies, they behave very differently
  - Photon is massless while W and Z bosons are heavy
- How does electroweak symmetry breaking occur?
  - In the SM, via the Higgs Mechanism

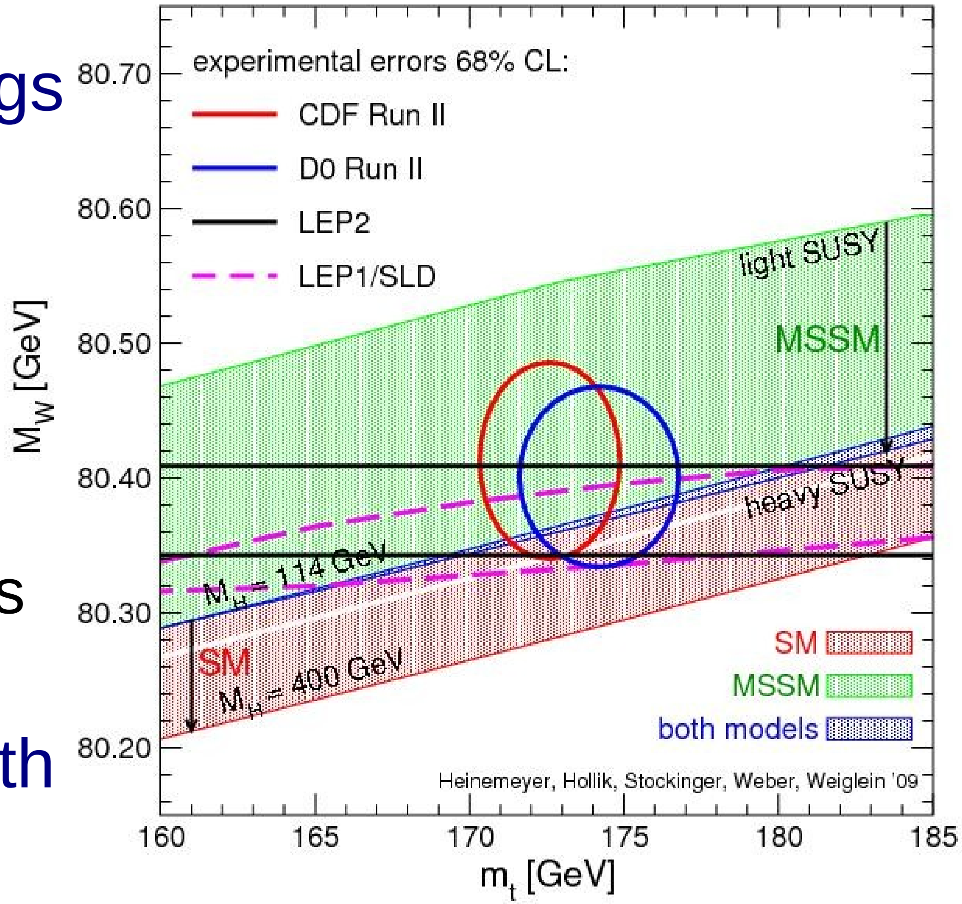


SM Higgs Mechanism

- To break the symmetry of the electroweak force:
  - Electroweak force is a gauge theory – SU(2)   U(1)
    - Interactions follow from symmetries   4 massless gauge bosons
  - Introduce nonzero scalar field permeating all space
    - To preserve gauge invariance, 3 of the 4 gauge bosons gain mass (W^+, W^-, Z^0)
- One remaining degree of freedom:
  - Manifests as a massive, spin-0 particle associated with the scalar field
    - The Higgs boson! – but no prediction for its mass
  - Finding the Higgs boson would directly test the theory

Searches for SM Higgs Boson

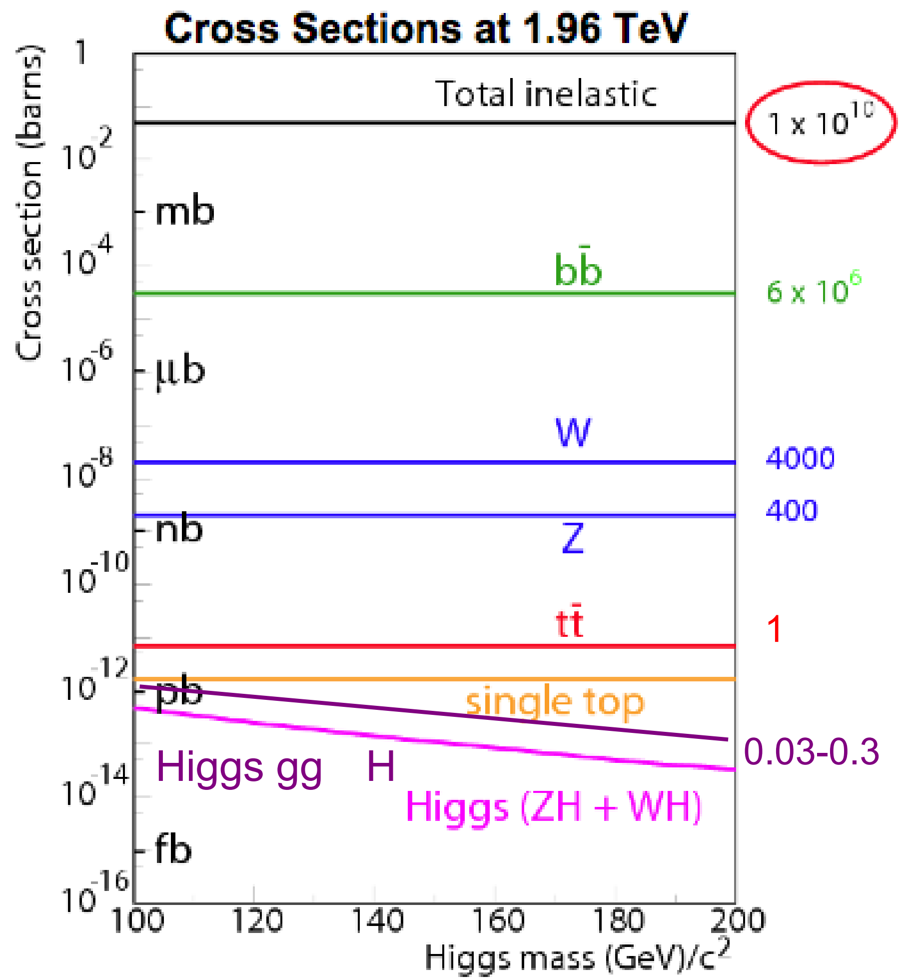
- In late 1990s, CERN made direct searches for SM Higgs
  - Excluded m_H < 114.5 GeV at 95% C.L.
- Indirect constraints from electroweak data prefer lighter Higgs (m_H < 154)
  - Combined with LEP results upper limit of m_H < 185
- Now Tevatron continues with direct searches
- We know where to look!



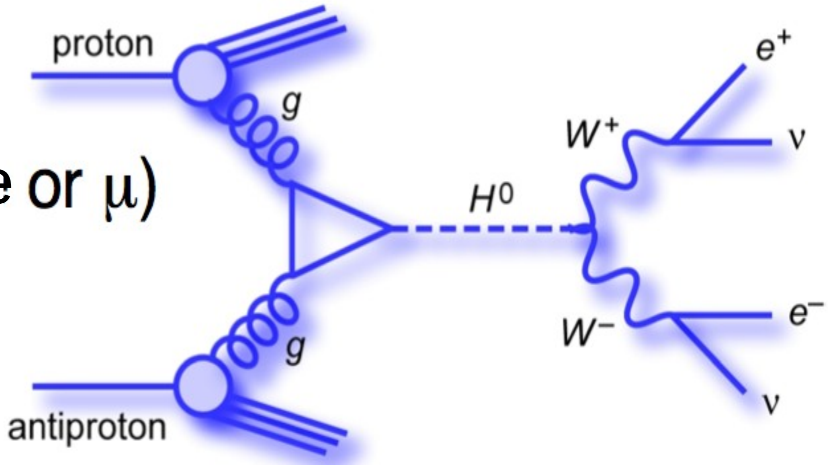
Plot from Tommaso Dorigo's blog

H WW ℓℓ Triggers

- Extract handful of Higgs events from background 11 orders of magnitude larger!
- High p_T lepton triggers
  - Central electrons
  - Muons (CMUP, CMX)
  - Forward electrons + Met
  - One lepton must satisfy trigger requirements
- Use luminosity ~4.8 fb^{-1}
  - Require good detector performance



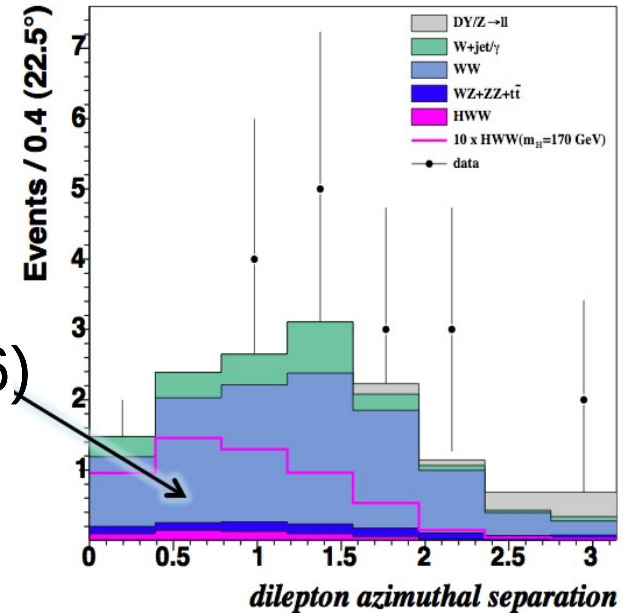
H WW $\ell\ell$ Selection

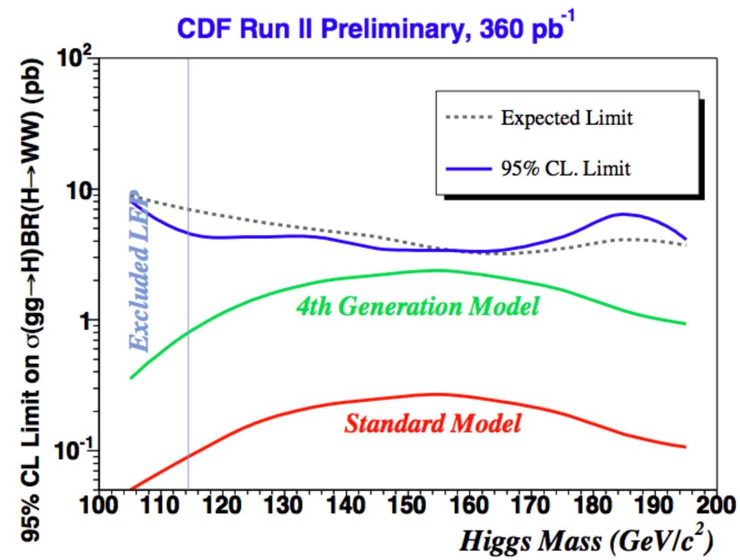


- Trigger on high p_T lepton
  - Two opposite-charge leptons (e or μ)
    - $p_T(l_1) > 20$, $p_T(l_2) > 10$ GeV
  - Dilepton mass $M_{ll} > 16$ GeV
    - Suppress low mass backgrounds
  - Require large missing transverse energy (Met)
    - Backgrounds can mimic Met if the energy of a jet or lepton is mismeasured in the detector
- Classify events by the number of reconstructed jets
  - Three categories: 0 jet, 1 jet, and 2 or more jets
  - Each has a different background composition
    - Better to optimize for signal in each separately

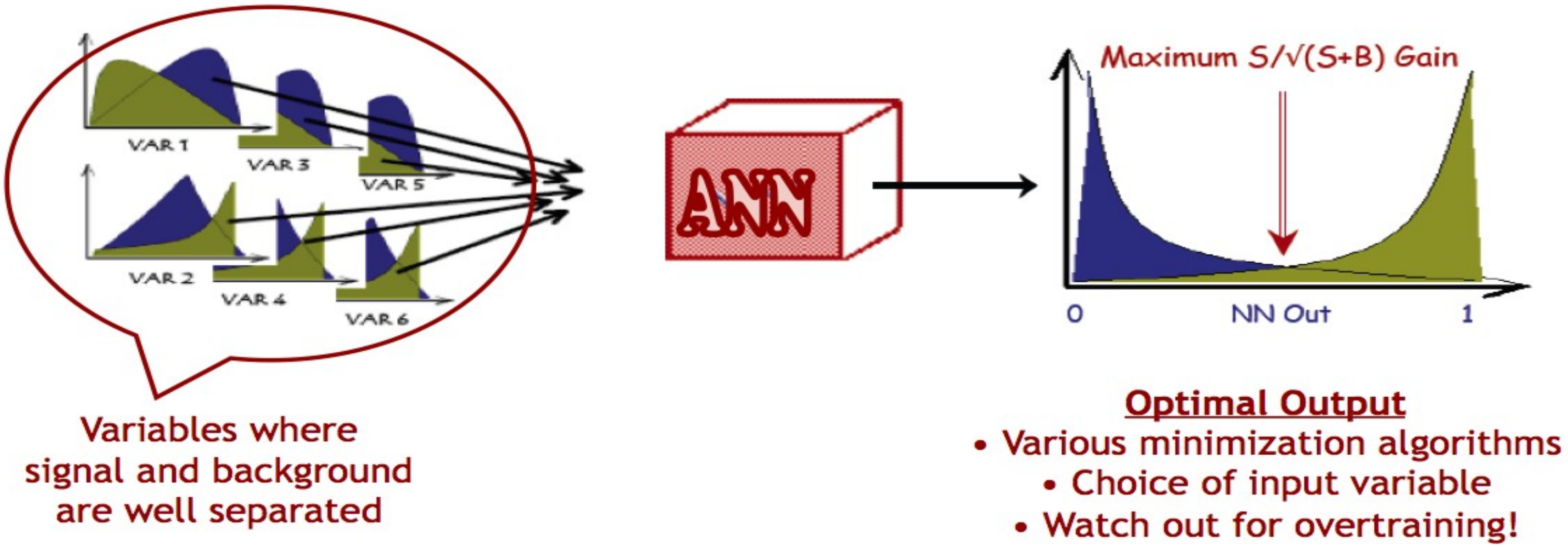
H WW Analysis, 4 years ago

- Based on 360 pb^{-1} of data
  - Considered only gluon fusion Higgs production
  - Used dilepton as discriminant
  - Published: PRL 97, 081802 (2006)
- With 5 fb^{-1} using this method,
  - Expected limit for m_H = 160 GeV: ~3 x $_{SM}$
- To reach SM sensitivity, need to improve the method!
  - Increase lepton acceptance
  - Optimize signal separation
  - Multivariate analysis techniques





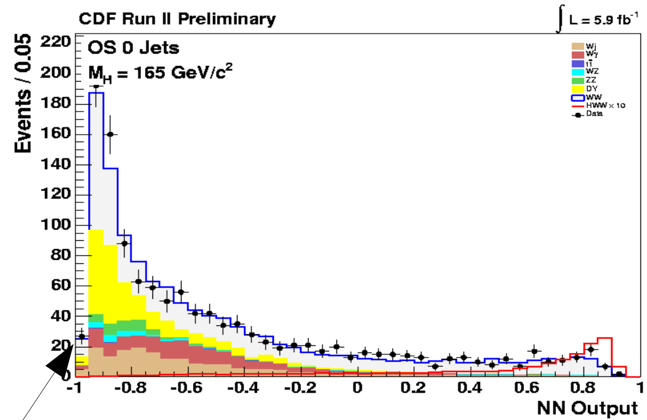
Neural Network

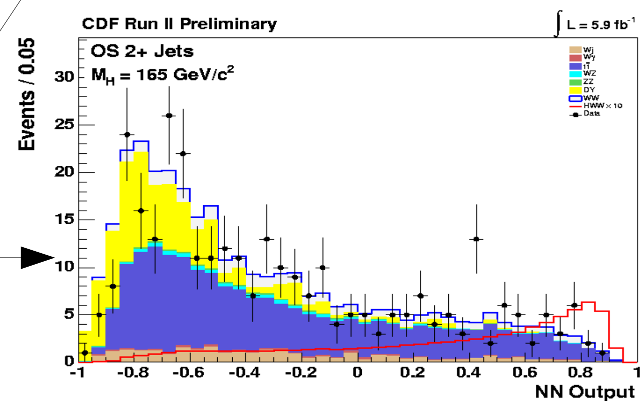


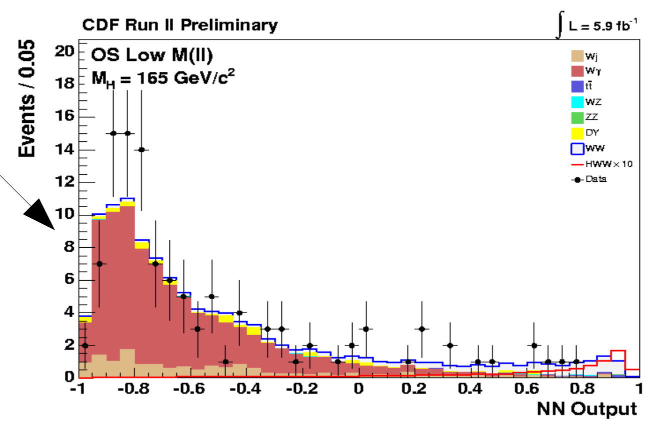
- Use NeuroBayes neural networks
  - Commercial NN package with fast, robust training methods
  - Each network has 3 layers:
    - Input layer (*n* nodes), hidden layer (*n+1*), output layer (*1*)
  - Trained on a weighted combination of signal + background
  - Excess of data at high NN score would indicate signal!

Signal Regions

- See S. Jindariani's wine & chees from March 2009 for more details on our primary search regions
- Opposite-sign dileptons divided by number of reconstructed jets
  - 0-jet: WW and gluon fusion dominate
  - 1-jet: WW and DY backgrounds
  - 2+ jets: t-tbar dominates
- Also consider separately a low-dilepton mass region (M_{ll} < 16 GeV)
  - W background, gluon fusion signal
- Same-sign dileptons
  - W+jets background, VH signal







Matrix Elements

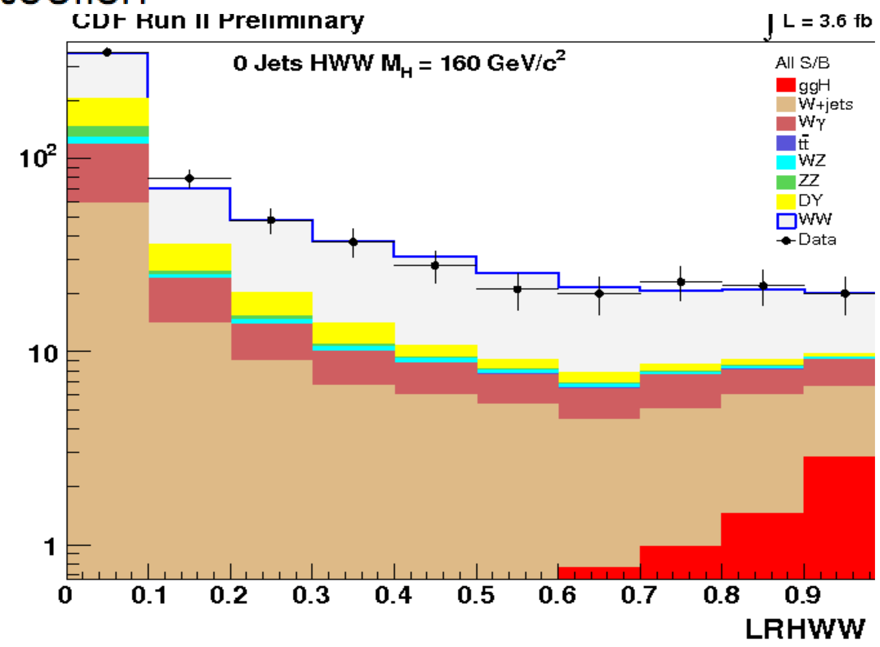
$$P(\vec{x}_{obs}) = \frac{1}{\langle\sigma\rangle} \int \frac{d\sigma_{th}(\vec{y})}{d\vec{y}} \, \varepsilon(\vec{y}) \, G(\vec{x}_{obs}, \vec{y}) \, d\vec{y}$$

- Event-by-event probability density

$\vec{x}_{obs}$	Observed leptons and $\not\!\!E_T$
$\vec{y}$	True lepton 4-vectors (l, ν)
σ_{th}	Leading order theoretical cross-section
$\varepsilon(\vec{y})$	Efficiency & acceptance
$G(\vec{x}_{obs}, \vec{y})$	Resolution effects
$1/\langle\sigma\rangle$	Normalization

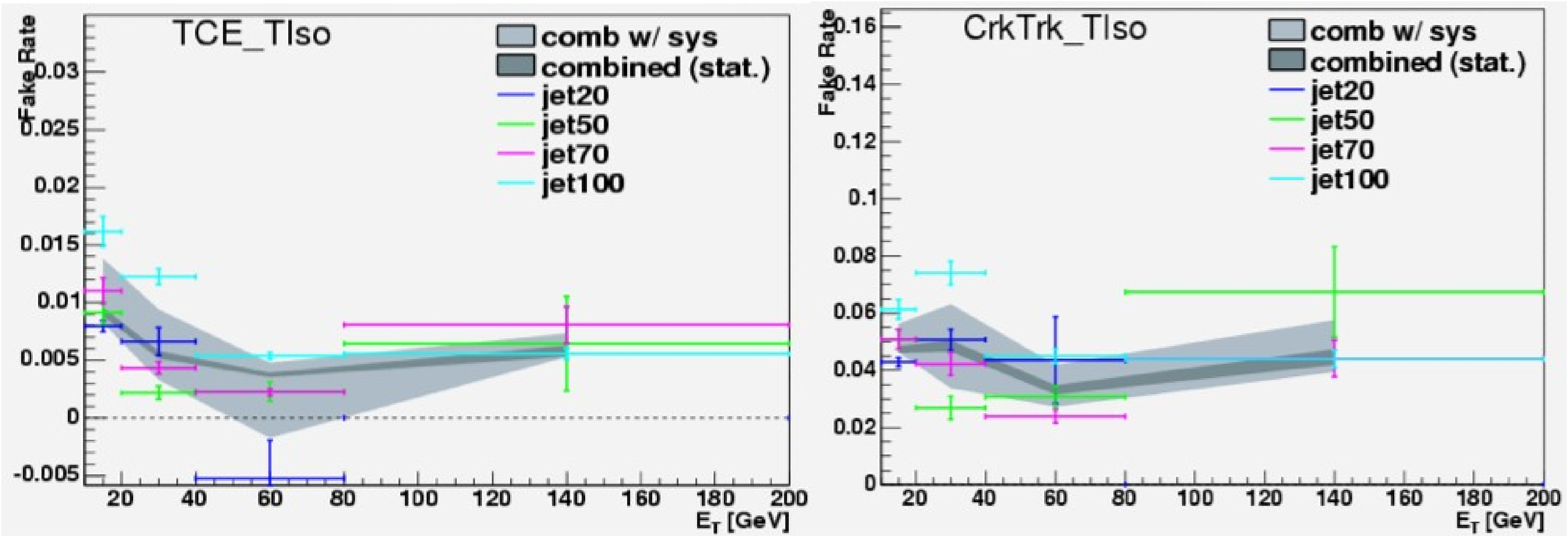
$$LR_m = \frac{P_m(\vec{x}_{obs})}{P_m(\vec{x}_{obs}) + \sum_i k_i P_i(\vec{x}_{obs})}$$

- Model 5 modes:
  - HWW, WW, ZZ, W , W+jet
- Construct Likelihood Ratio

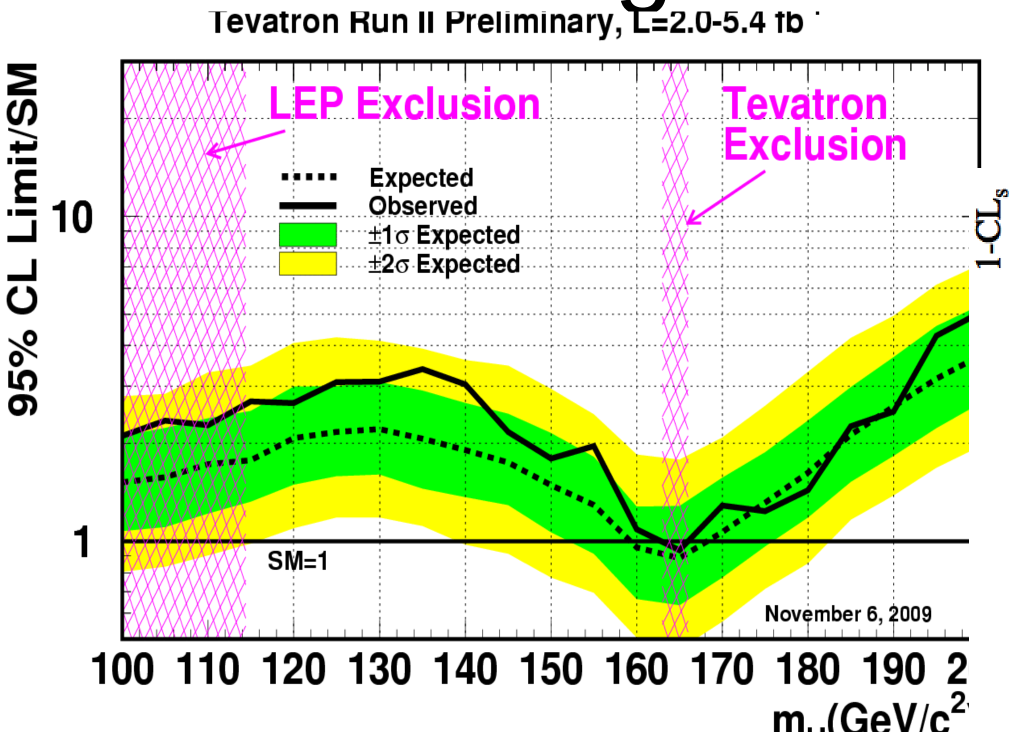


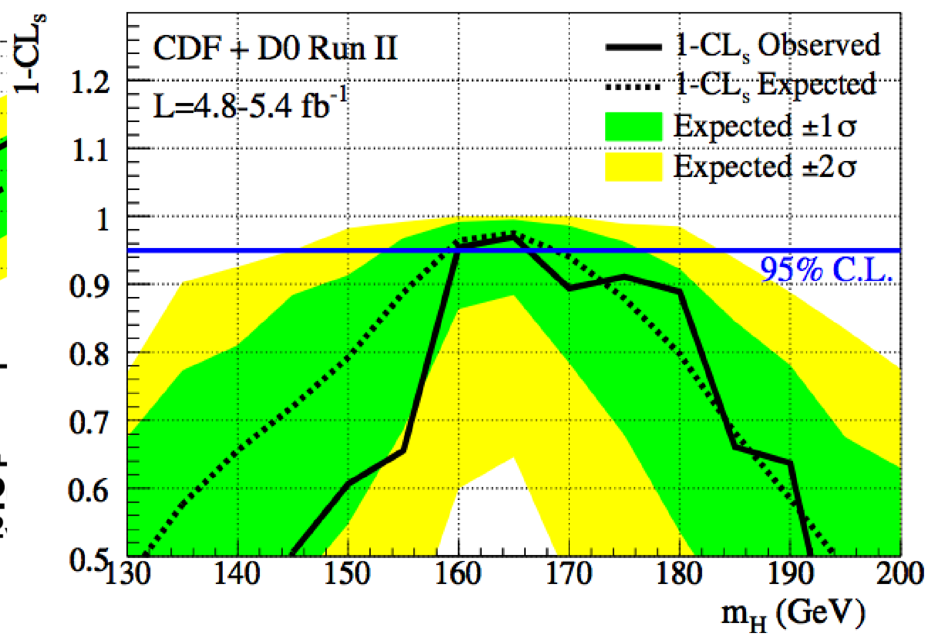
Background Modeling

- Most backgrounds modeled by Monte Carlo
  - WW by MC@NLO, others by Pythia or Baur (W), except...
- W+jets uses data-driven estimate of fake leptons:
  - Select identifi ed leptons (numerator) and "fakeable objects" (denominator) in jet data samples
  - Subtract ewk contributions from Z ee/ and W e/ MC
  - Calculate ratio – for each lepton category



Tevatron High Mass Combination





- Combine results into an overall Tevatron Higgs limit
  - Calculate both Bayesian and CL_s limits (similar results)
  - Exclude SM Higgs with mass 162-166 GeV at 95% CL

Setting a limit

- Use Bayesian limits calculator
  - Tom Junk's MCLimit program
  - Prior is f at in the number of Higgs boson events
  - Return the 95% credibility upper limit (C.L.)
- Input distributions for each channel:
  - 1 NN output template for each event hypothesis:
    - gg H, ZH, WH, VBF, WW, WZ, ZZ, W , W+jets, DY, tt
    - Total of 8 (11) histograms at each Higgs mass
  - For a combined limit, use templates for all channels being combined
- Include all systematic uncertainties as nuisance parameters using pseudo-experiments