Low Mass Higgs Search with $H \rightarrow b\bar{b}$ decay at the Tevatron

Adrian Buzatu, University of Glasgow

On behalf of the **CDF** and **DZero** Collaborations

Higgs Hunting 2012, 19 July 2012

RA (now)

MSc, PhD at CDF

Outline

- **Standard Model Higgs boson searches**
  - Higgs boson searches
  - Higgs decays to bottom quark ($b\bar{b}$) pairs

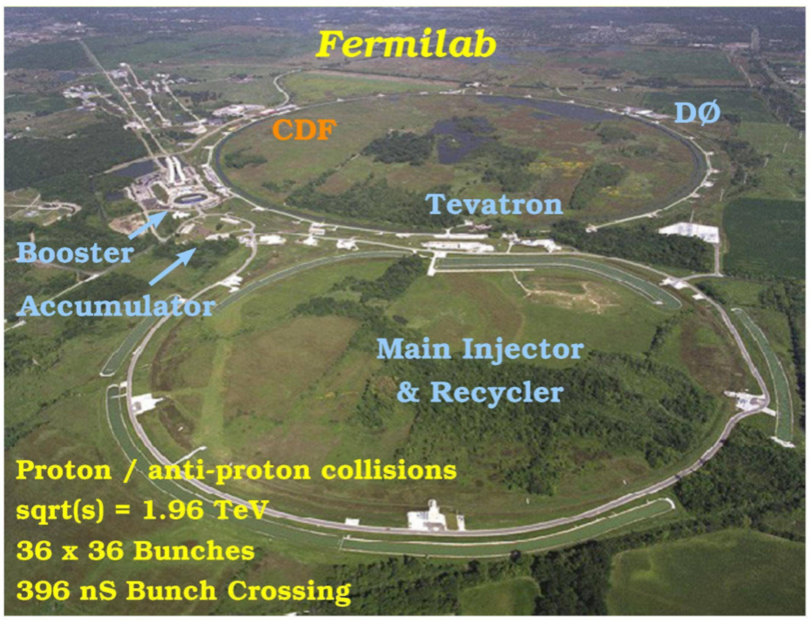
- **Tevatron Combination for H to bb decay**
  - Statistical Inference & Systematic Uncertainties
  - Log Likelihood Ratio and p-values
  - Expected and Observed Upper Limits
  - Interpretation

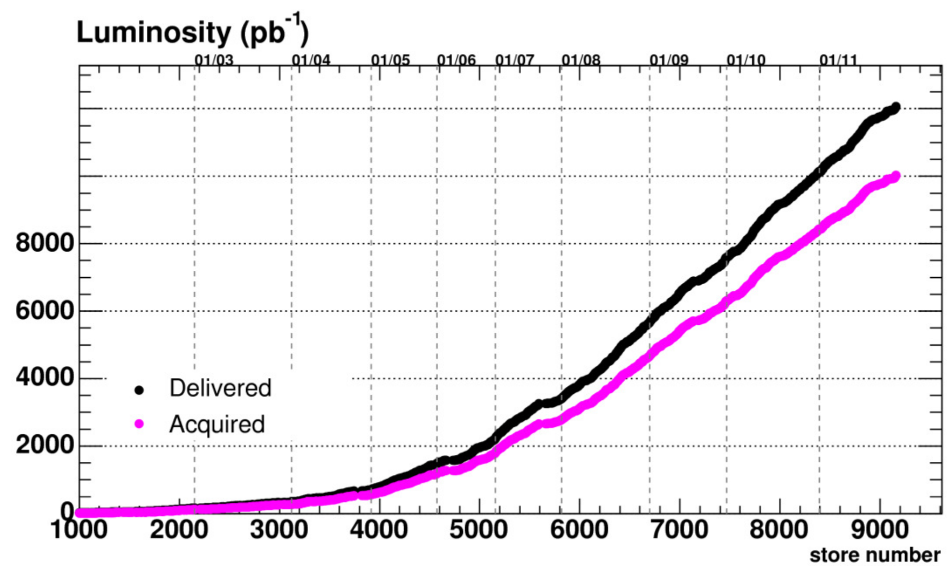
Standard Model Higgs Boson

- **Motivation**
  - Only elementary particle predicted by the Standard Model (SM) not yet observed. Predicted in 1964, it explains
    - spontaneous symmetry breaking
    - the masses of the electroweak bosons, the masses of fermions
- **Mass – only free parameter for the SM Higgs boson**
  - Before ICHEP 2012, direct searches at ATLAS and CMS at the LHC superseded previous limits at the Tevatron and LEP experiments
    - **masses < 115.5 GeV/c^2 and > 127 GeV/c^2 excluded at 95% CL**
  - Indirect electroweak fits
    - masses > 152 GeV/c^2 excluded at 95% CL
  - ATLAS & CMS announced the observation **a new elementary particle with mass around 125 GeV/c^2**, in decays to di-photon and di-Z-boson, **consistent with the SM Higgs boson; but is it?**
- **Higgs production is a very rare process**

The Tevatron Accelerator

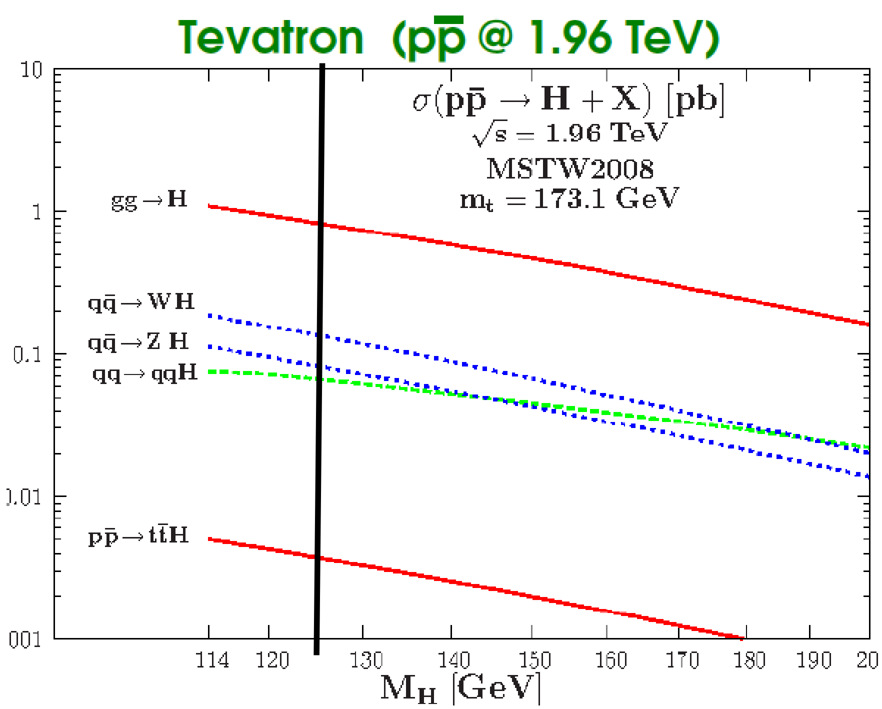
- p-pbar collisions at 1.96 TeV, $L_{peak} = 4.3 \times 10^{32}\ cm^{-1}s^{-1}$
- Delivered ~12 fb^{-1} full data set before shutdown on 30 Sept 2011
- Results presented today **use the full data after data quality requirements ~ 10 fb^{-1}**

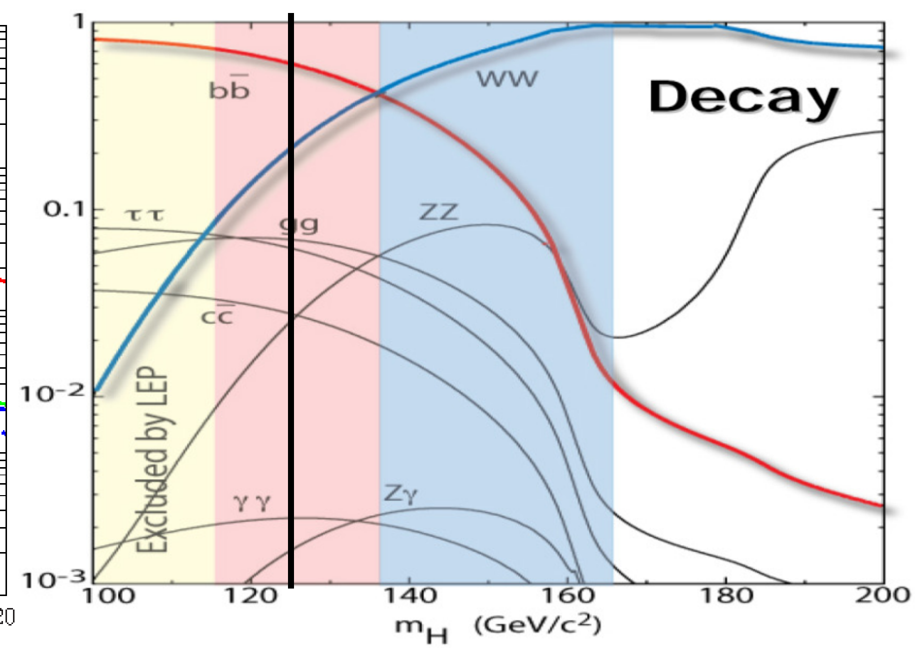




Higgs Production and Decay

- For masses smaller than 135 GeV/c^2 the Higgs boson decays mostly to bottom quark (bb) pairs
- **At 125 GeV/c^2, BR: 57.8% ($b\bar{b}$),** 21.6% (WW), 2.67%(ZZ), 0.23%(γ γ)
- **Need to observe the bb decay of the newly observed ~ 125 GeV/c^2 boson** by ATLAS and CMS, to test if it is indeed the SM Higgs boson





Divide, Conquer, Combine

- Divide data in different analysis channels based on topology
- **Optimize each channel individually**

- **Primary s/b discriminant is the dijet invariant mass**
- **Heavy use of multivariate techniques**, adding other kinematic information in addition to dijet invariant mass
  - Artificial Neural Networks
  - Boosted Decision Trees, etc
- **Combine all channels**, both from CDF and DZero

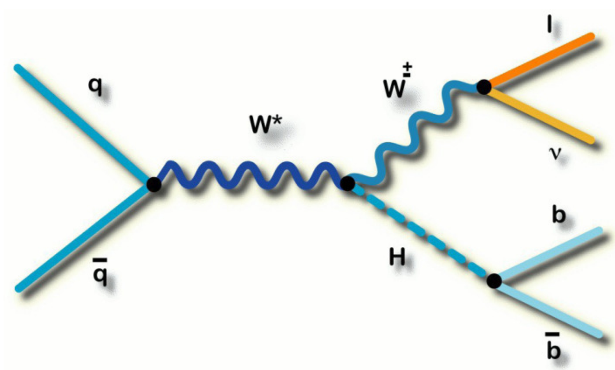
- **See Friday 10 am talk by J. Haley on Tevatron SM Higgs combination**

H -> b$\bar{b}$ Searches

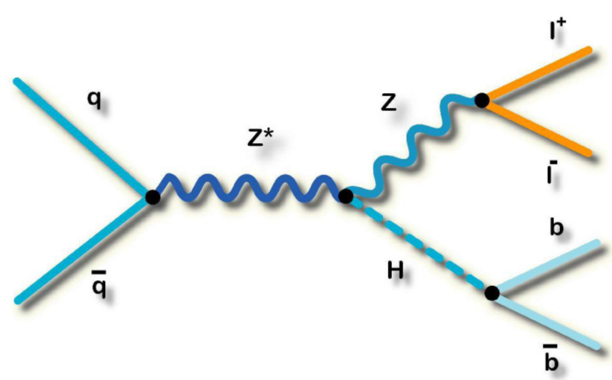
- CDF/DZero search for SM H -> b$\bar{b}$ using Tevatrons's full data set

- **Production mechanisms**
  - Gluon fusion not feasible for bb decays: 10^9 more multijet bkg
  - **Associated production VH (in this talk),** also part of the **full Tevatron dataset combination: hep-ex:1207.0449**

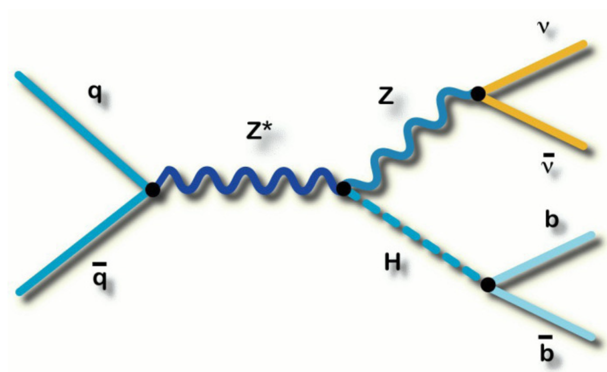
WH to lvbb
CDF (9.45 fb^{-1})
DZero (9.7 fb^{-1})



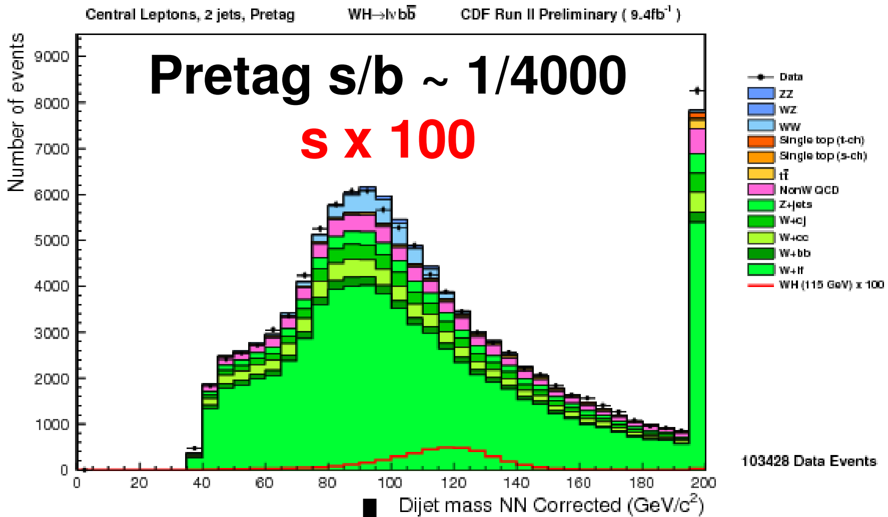
ZH to llbb
CDF (9.45 fb^{-1})
DZero (9.7 fb^{-1})



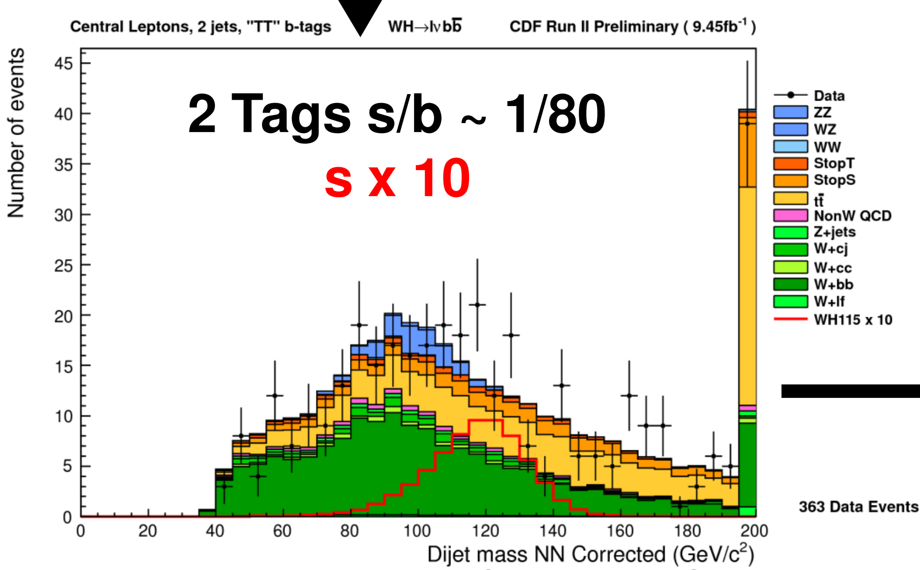
ZH to vvbb
CDF (9.45 fb^{-1})
DZero (9.5 fb^{-1})

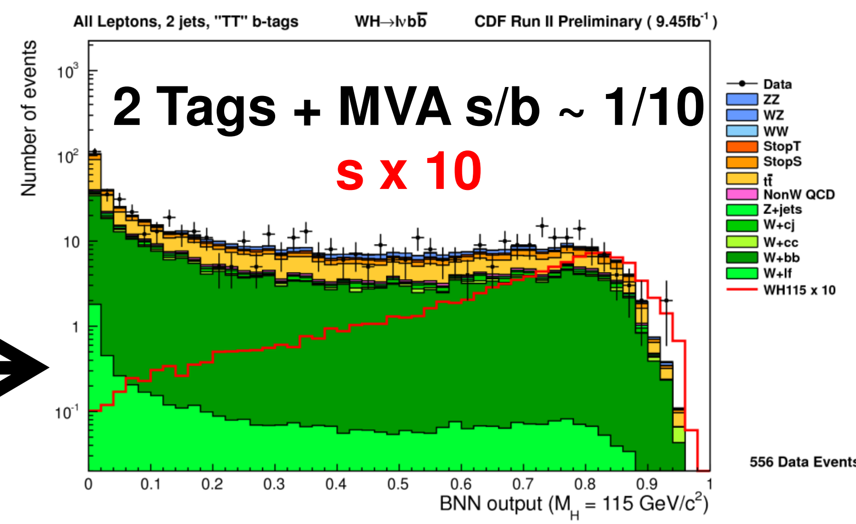


Analysis Strategy - 1



- WH as an example
- Two b-tagged jets
- Corrected dijet invariant mass (mjj)
- Multivariate final discriminant (MVA) using mjj and other 5-6 other variables: MET, HT, SumtET, etc





Analysis Strategy - 2

- **Several b-tagging categories grouping similar s/b bins** increases sensitivity over all categories summed up

- **Increased charged lepton efficiency** by introducing new categories with less stringent reconstruction criteria

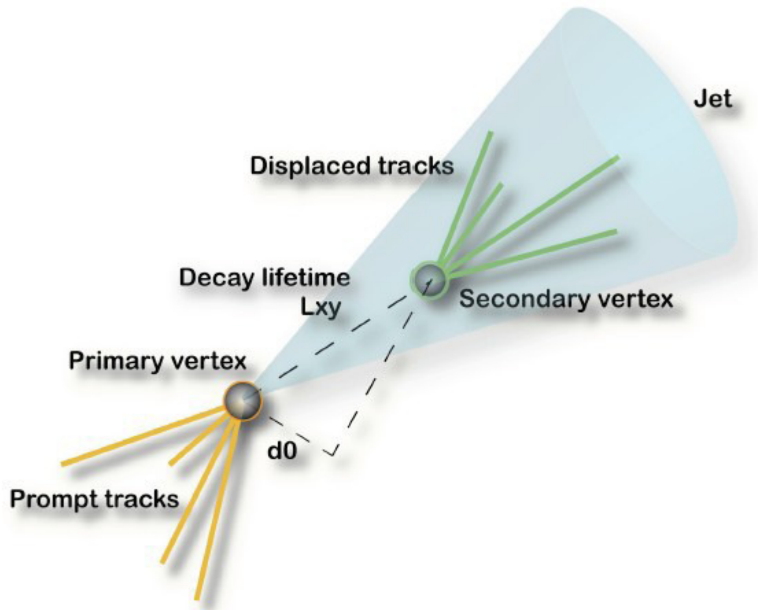
- **Increased trigger efficiency** by including new triggers (e.g. novel trigger combination method **arXiv:1206.4813 - CDF**)

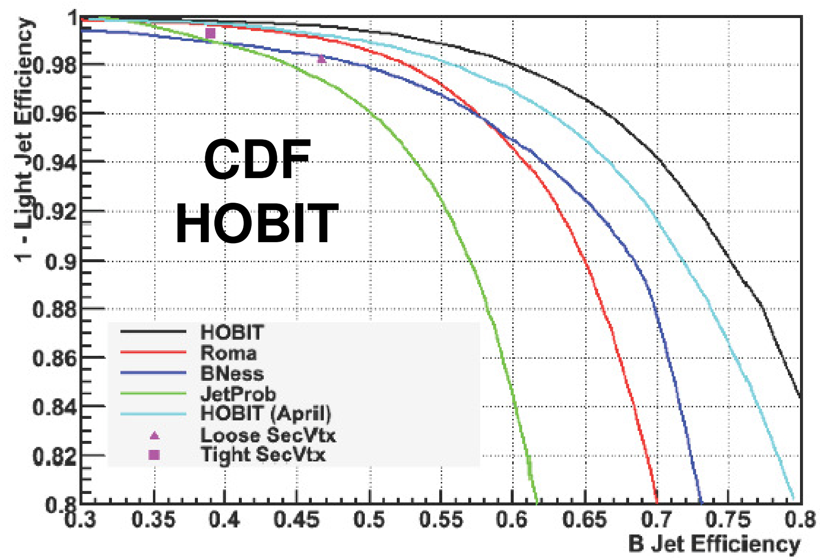
- **Search for a broad excess in the mjj distribution**
- **Improving the mjj resolution is key** (e.g. neural-network-based correction to the b-jet energy **arXiv:1107.3026 - CDF**)

- **Use shape of multivariate discriminants** obtained by taking as input mjj (best input), and also other kinematic information in the event

b Tagging

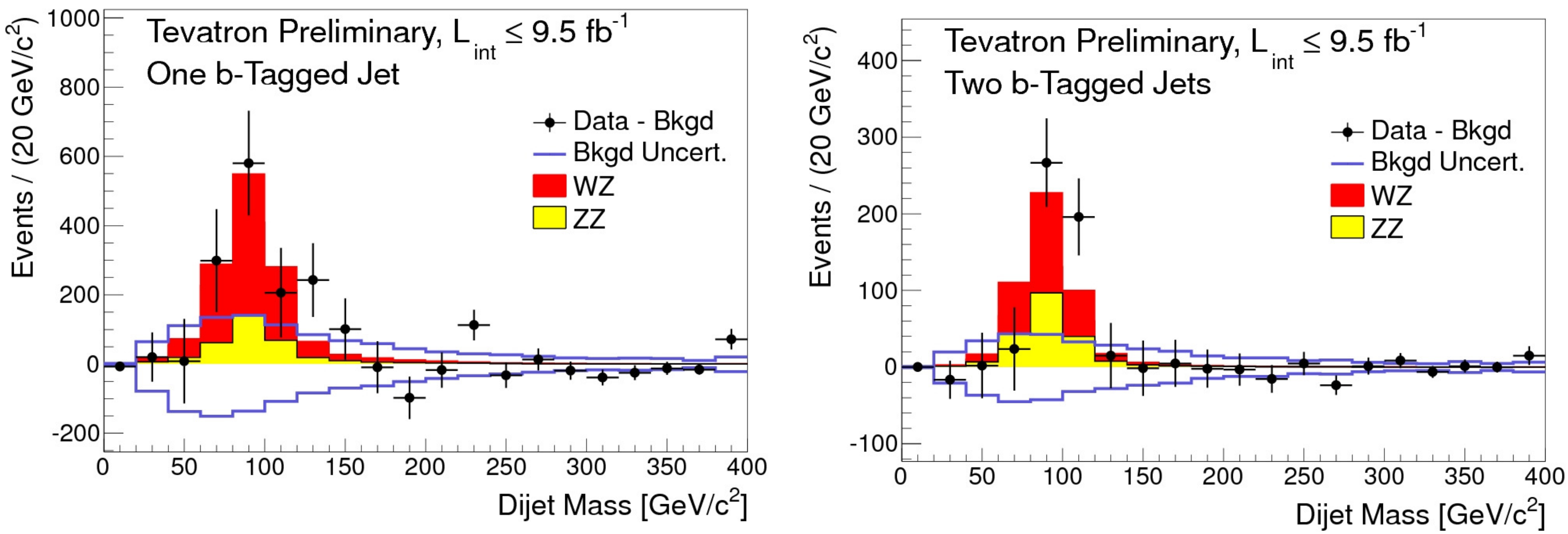
- **Both CDF (new) & DZero use multivariate techniques (MVA) to improve the b-tagging algorithms**
- Displaced vertices; large track impact parameters, muon in jet
- Typical efficiencies: 40-70%; Typical mistag rates: 1-5%
- Since last iteration, CDF combined existing b tagging algorithms into a MVA tagger, called **HOBIT** (**arXiv:1205.1812**), which improves efficiency by 20% while keeping the same mistag rate.
  - **Since double-tagged category is already very pure, the efficiency is crucial**



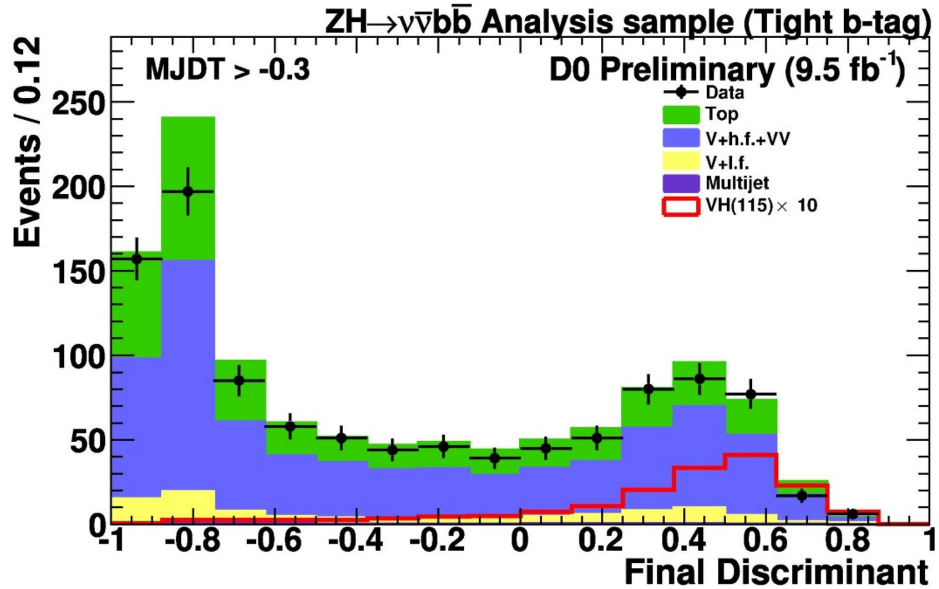


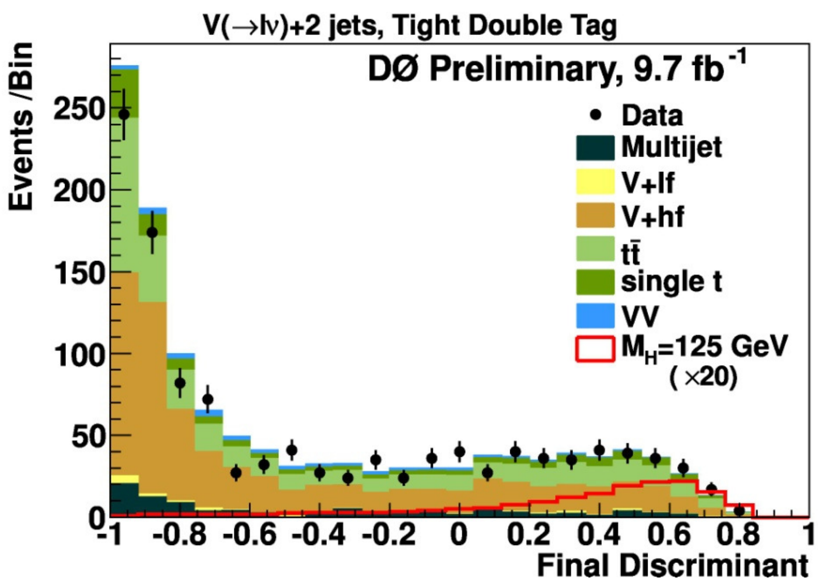
Validation in Z -> $b\bar{b}$ Searches - 1

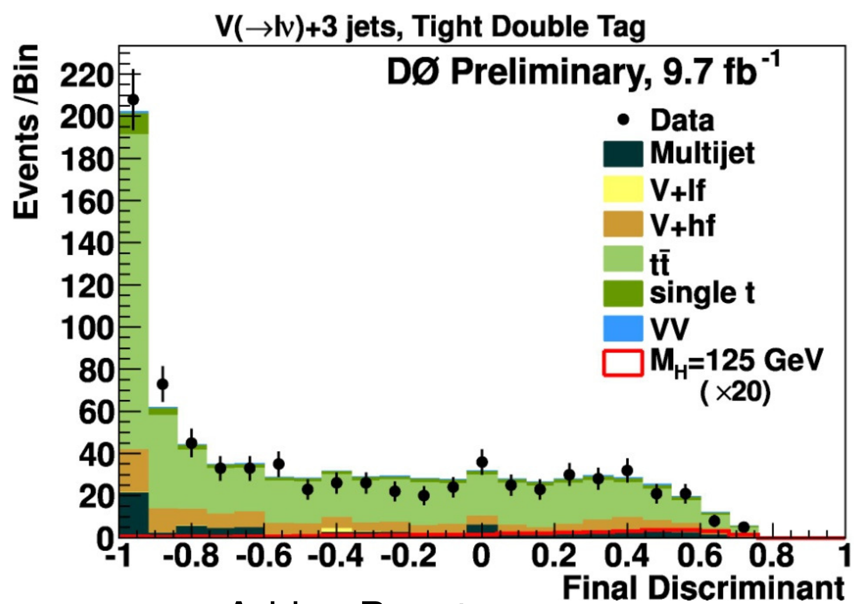
- Search for a bump in the dijet invariant mass
  - Sum all lvbb, llbb, vvbb channel
- Then subtract backgrounds from data
  - **Except WZ and ZZ**, bkg fit to data assuming no Higgs signal



Other MVA Examples

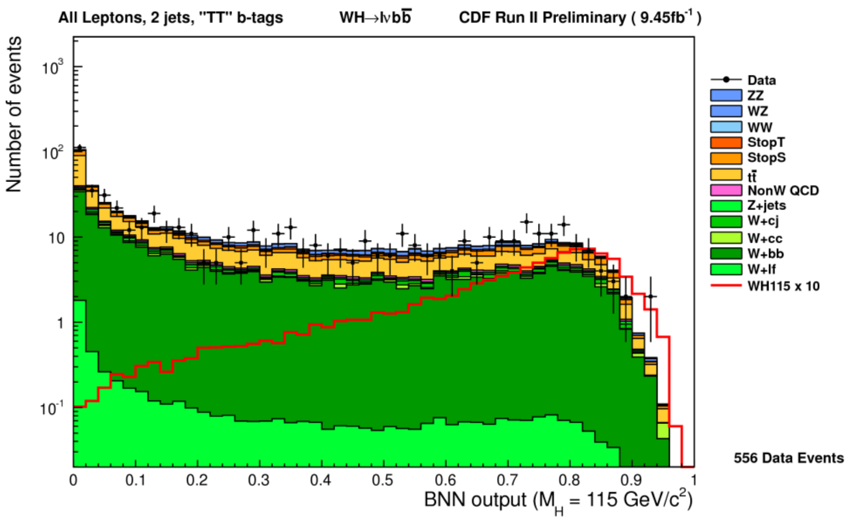




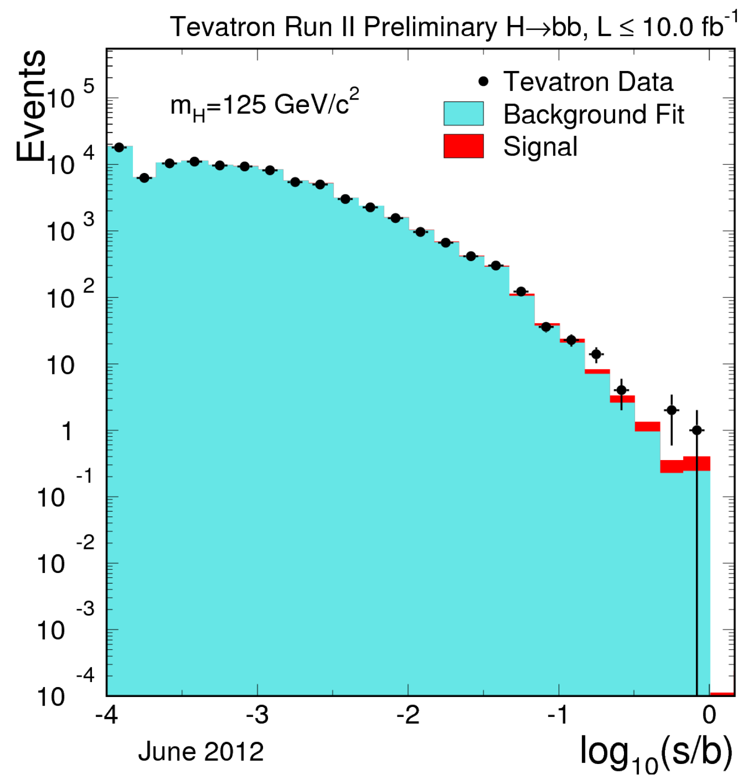


Signal (s) to background (b) Ratio

- Sum final discriminants after rebinning by log(s/b)
  - Fill a histogram with the **log(s/b) value weighted by the number of data** points for each bin of the MVA discriminants of each channel
  - Fit bkg to data within bkg rate and shape systematics for each channel, then **sum lvbb, llbb, vvbb channels**
  - **data agrees with background**
  - **best Higgs data candidates at high s/b**
- Shown for m_H=125 GeV/c^2

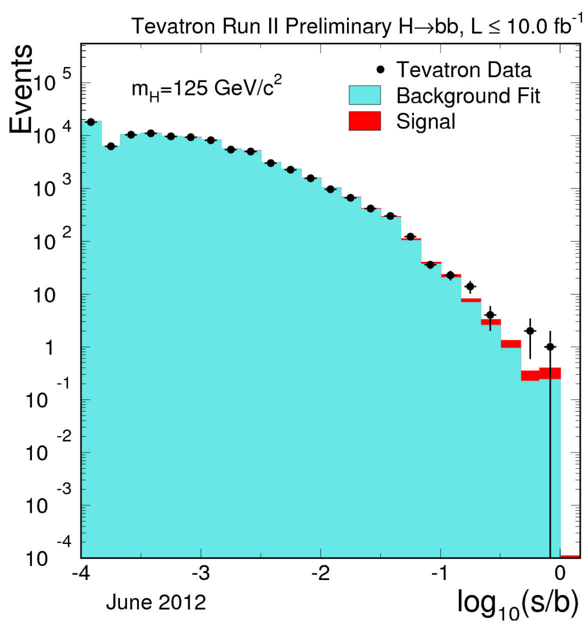


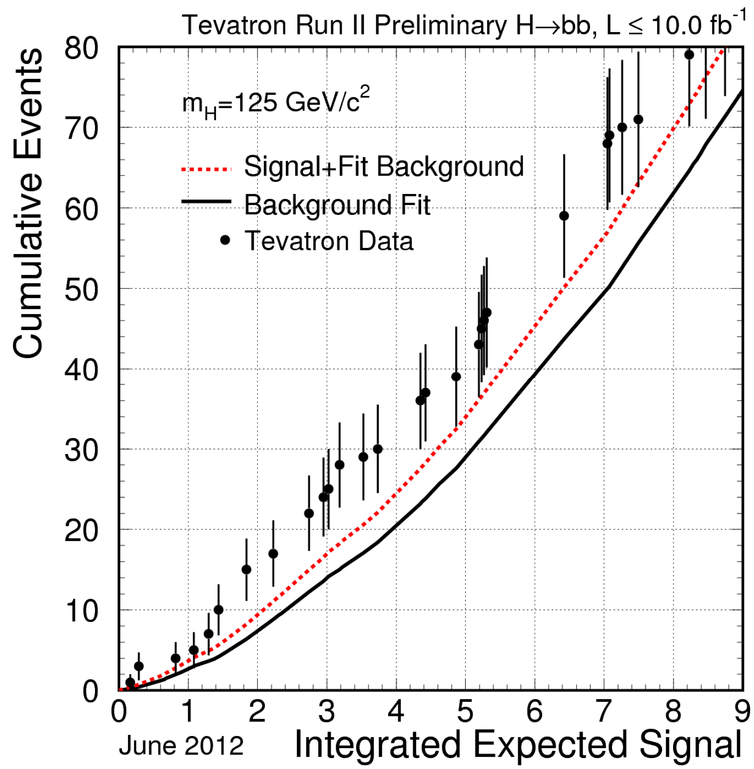
Sum all MVAs



Integrated Expected Signal

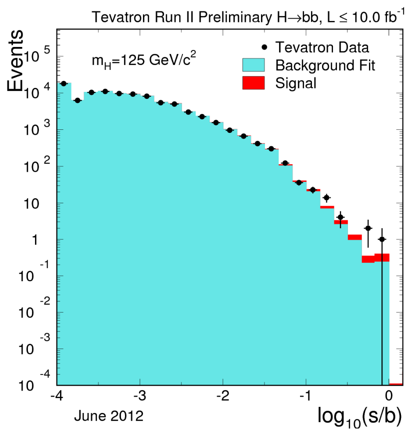
- Integrated distributions of s/b, starting from the high-s/b-value side (**most pure data in the left bins below**)
  - Statistical-only errors on data, correlated point to point
- For a signal given by m_H=125 GeV/c^2, **the data agree more with the s+b than the b-only hypothesis**

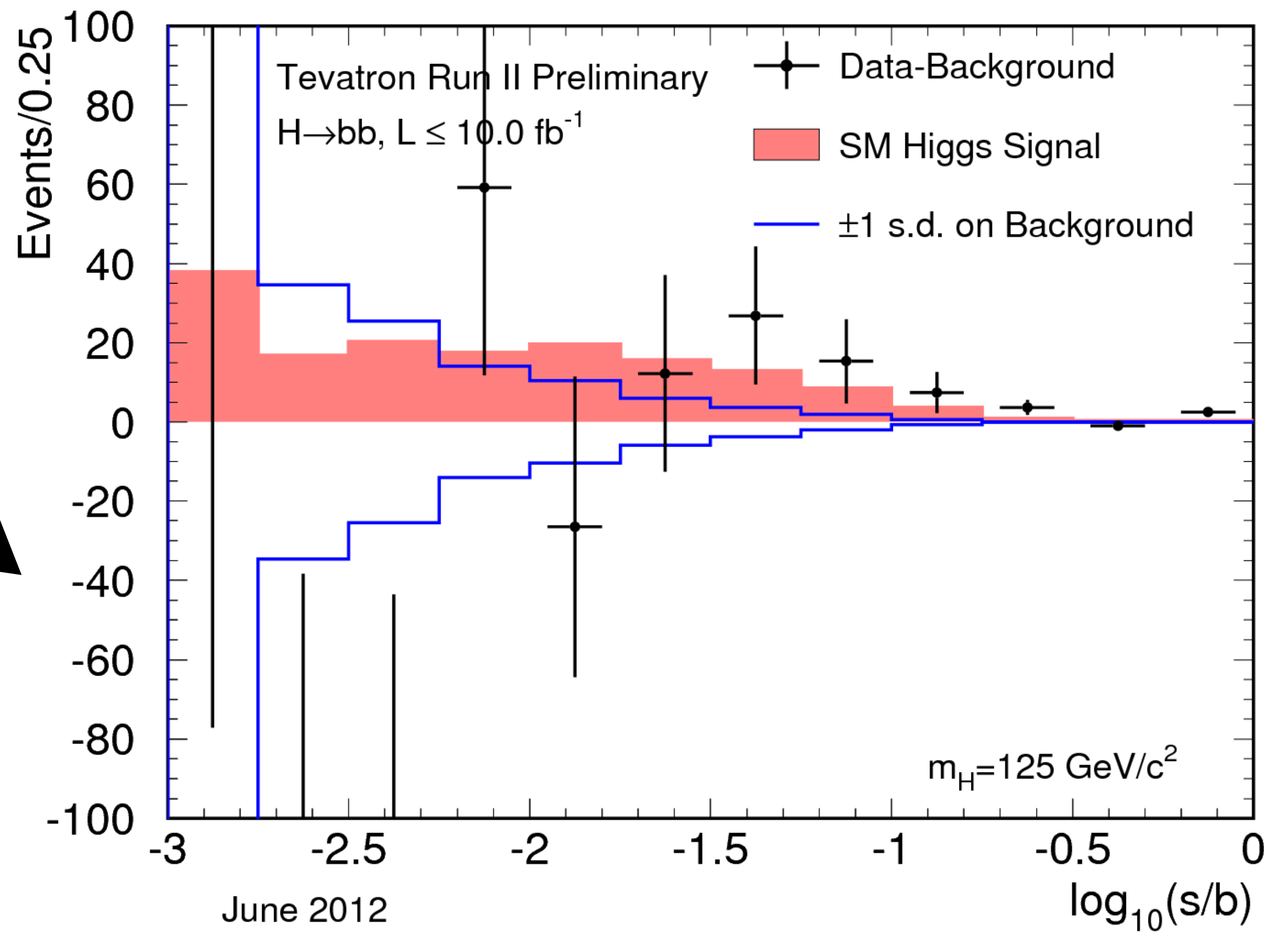




Signal (s) to background (b) Ratio

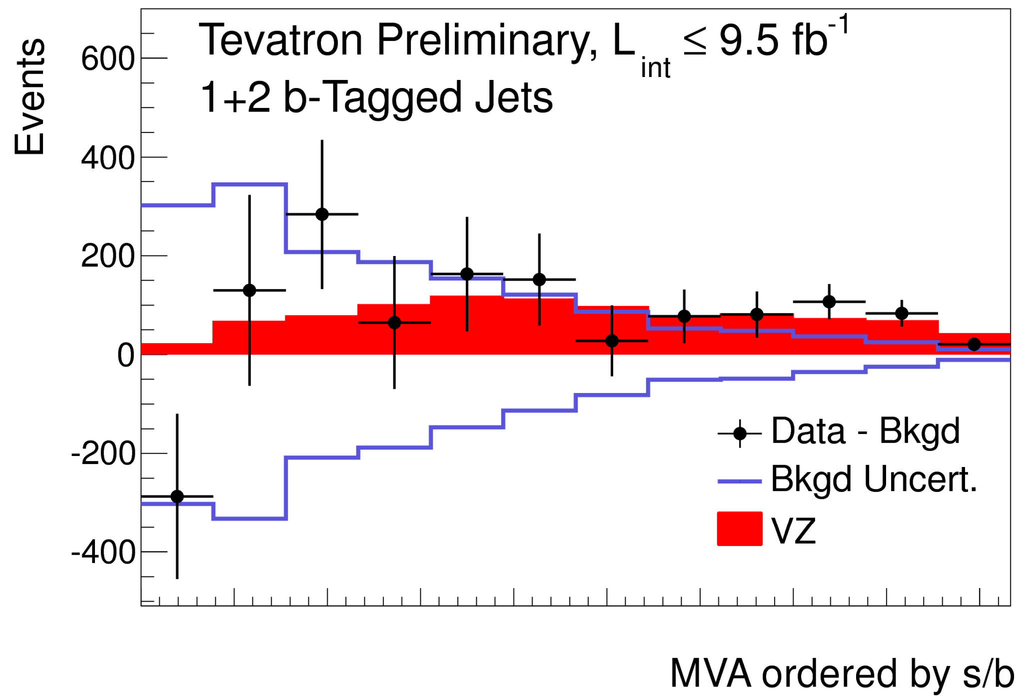
- **Subtract backgrounds from data** in the $\log_{10}(s/b)$ plot
- Shown for m_H=125 GeV/c^2





Validation in Z -> $b\bar{b}$ Searches - 2

- Analysis (hep-ex:1203.3782)
  - **VZ (V=W,Z) with Z->bb, using the same lvbb, llbb, vvbb channels and their respective analysis techniques**
  - **Treat WW as background. Assume SM WZ and ZZ σ ratio**
- Measure σ(WZ)+σ(ZZ)=4.47+/−0.64(stat)+/−0.73(syst)
  - **Consistent with the SM**
  - **4.6 standard deviations** above the background-only hypothesis

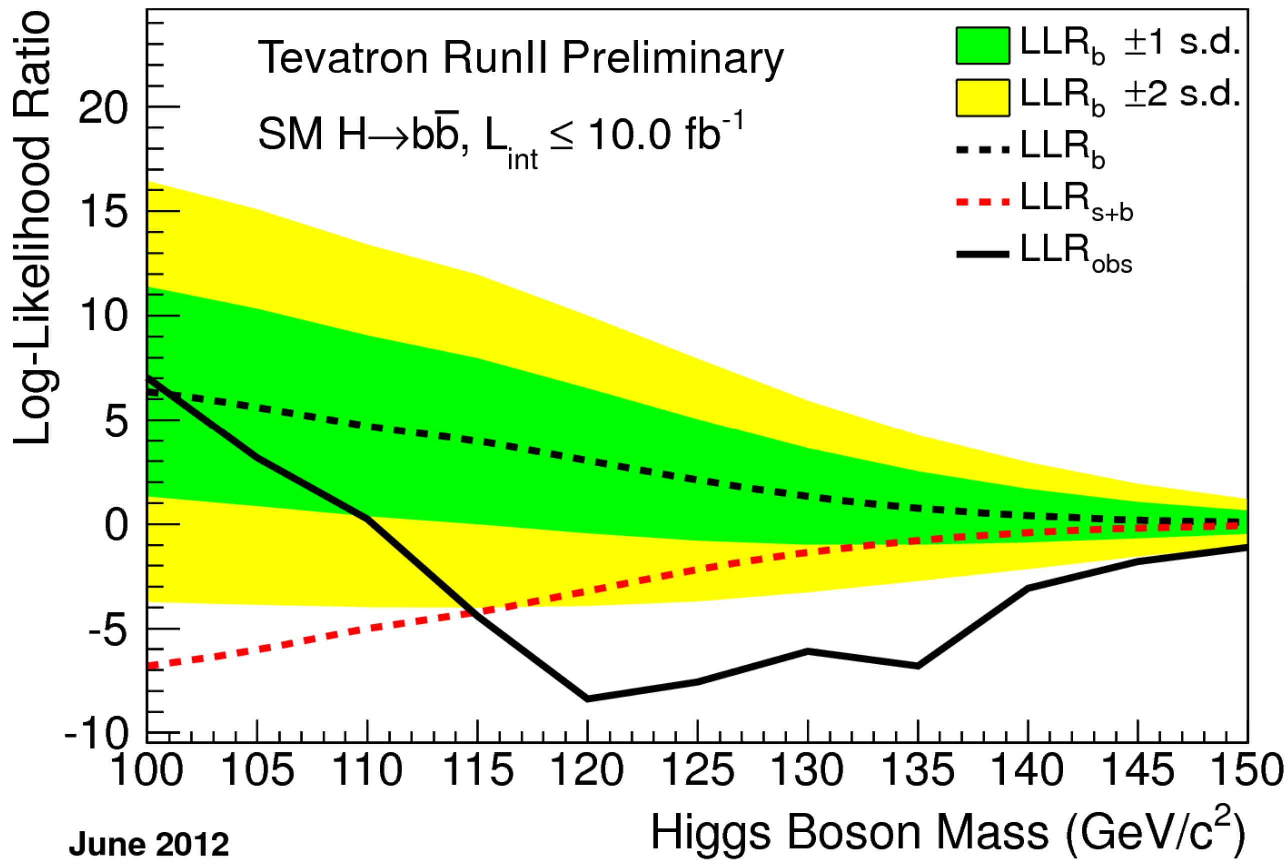


Statistical Approaches

- **Bayesian and Modified Frequentist methods cross-check each other**
  - At each Higgs boson mass, they agree within 10%
  - On average, they agree within 1%
  - Both rely on distributions of the final discriminants
    - not counting experiments
- **Rate and shape systematics**
  - **Some correlated across experiments and channels**
    - Theoretical uncertainties on predictions for signals and backgrounds
    - Luminosity
  - **Rest either correlated among sub-channels**
    - Charged lepton, trigger, b-tagging efficiencies
  - **Or not**
    - Fake object identification, data-driven background modelling

Log Likelihood Ratio (LLR) from CL_s

- Separation between LLR_b and LLR_{s+b} medians illustrates the search sensitivity. Closeness of LLR_{obs} to LLR_b (LLR_{s+b}) illustrates a preference for the b-only (s+b) hypothesis for that mass point
- **Broad (115-145 GeV/c²) excess larger than 2 standard deviations**



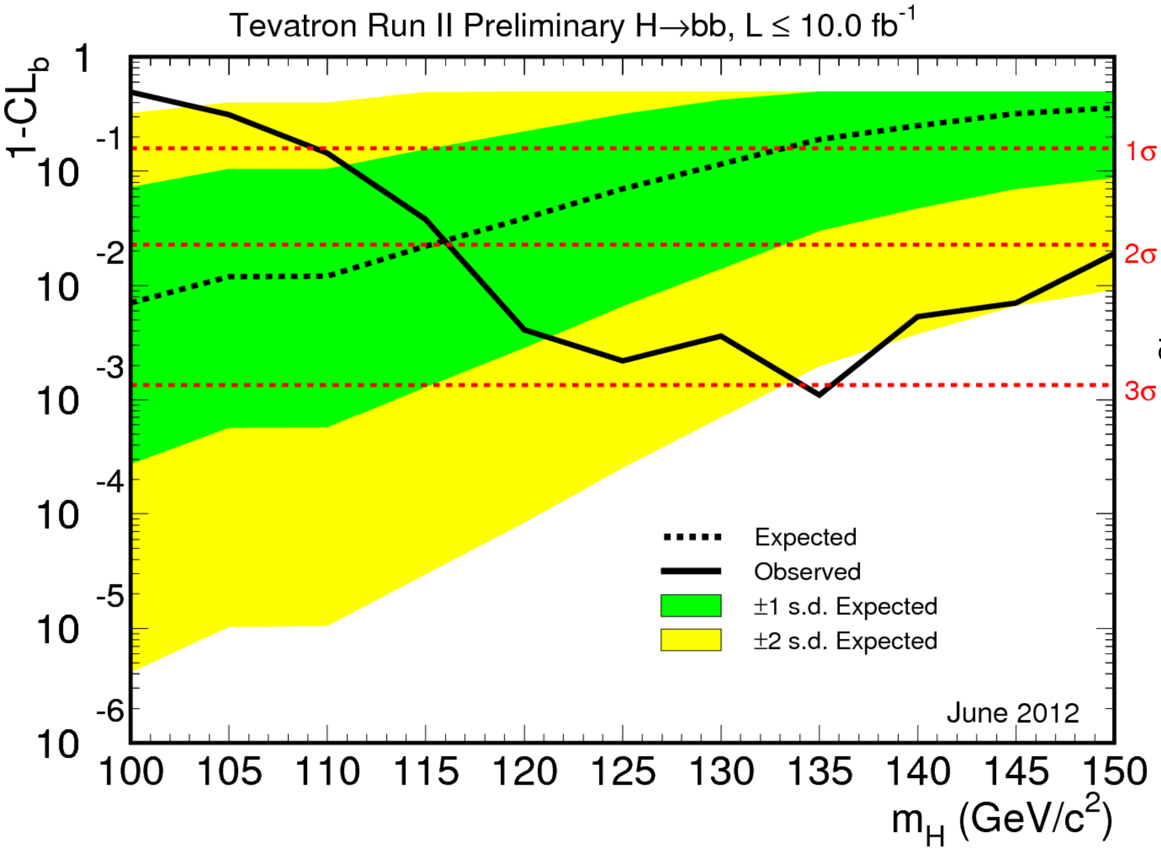
b-only expected

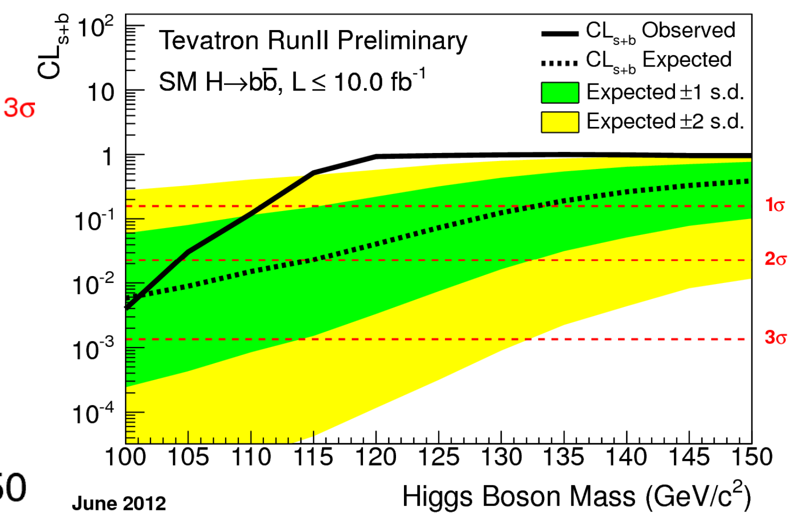
s+b expected

observed

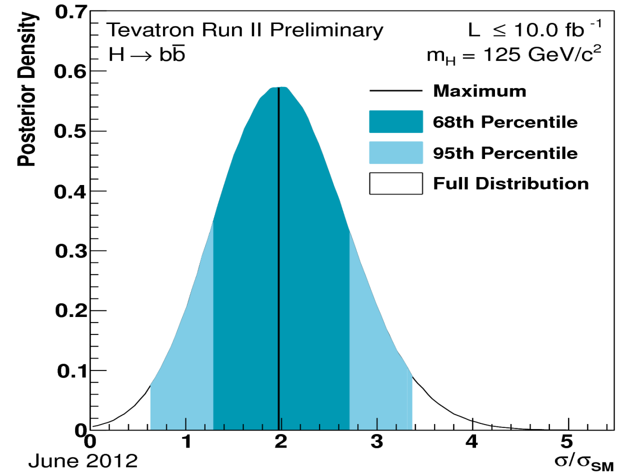
Local p-values for 1-CL_b and CL_{s+b}

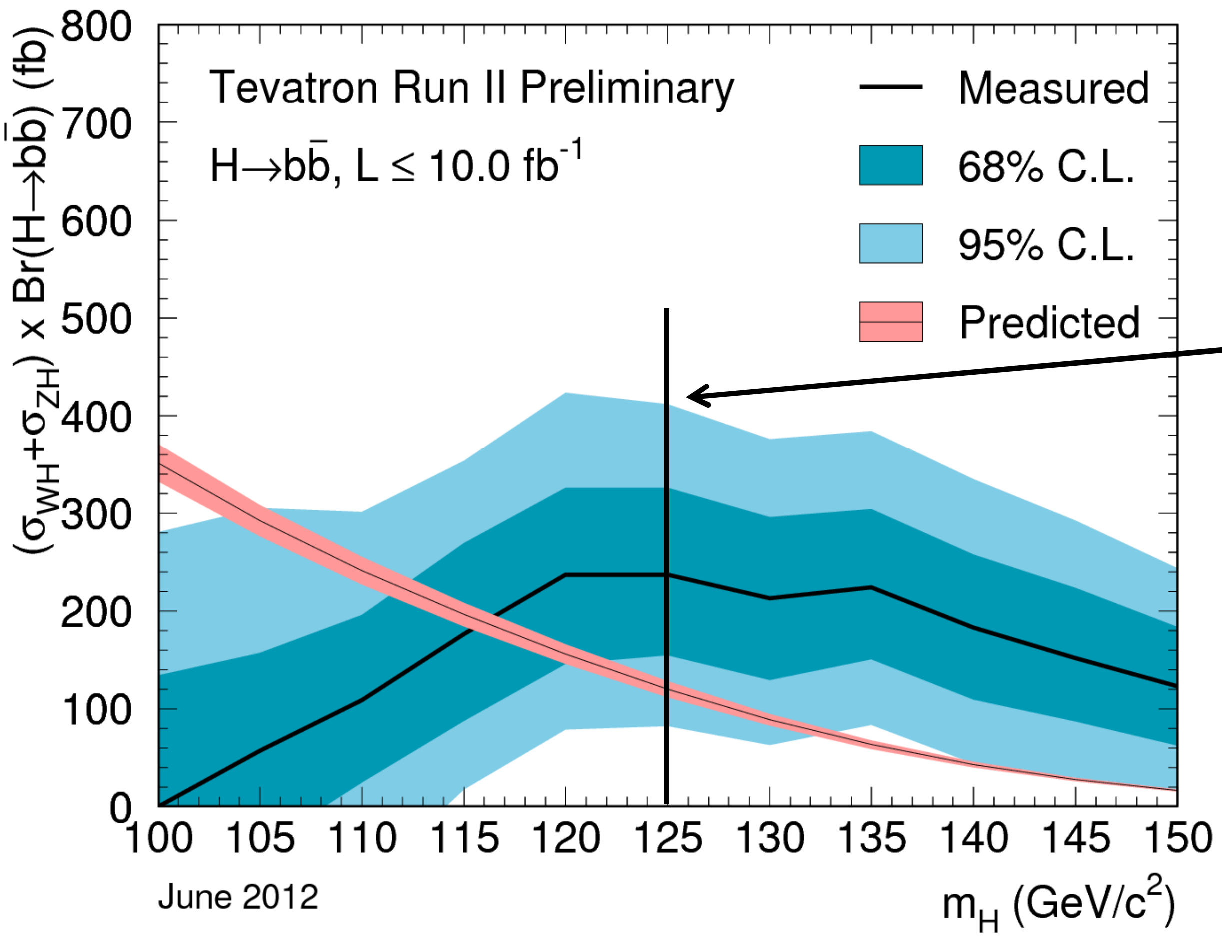
- Minimum local p value for 1-CL_b ~ 8.10^{-4} (3.2 local significance, 2.9 global significance) at m_H=135 GeV/c^2
- CL_{s+b} shows the same broad excess at 115-145 GeV/c^2





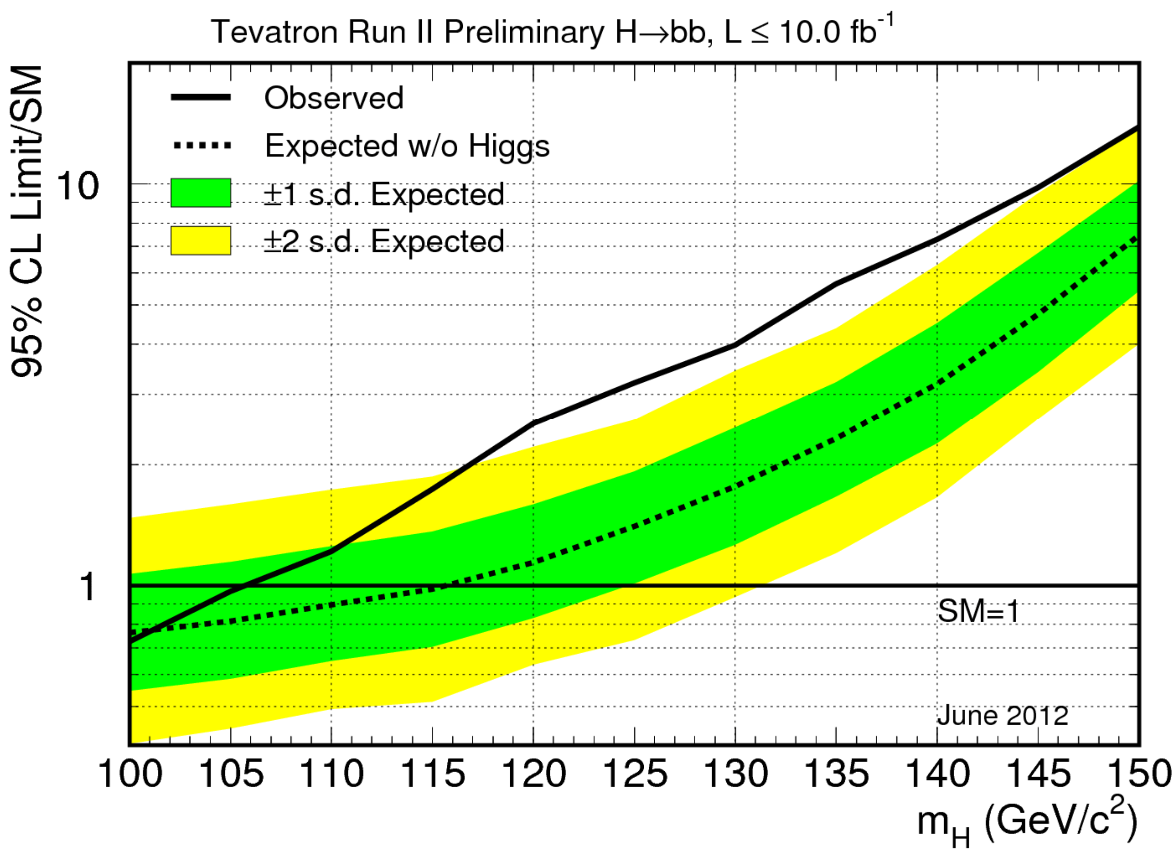
Fit H to bb signal σ·BR to data





95% CL Upper Limits

- ❑ Expect to exclude at 95% CL: [100 - 116] GeV/c^2
- ❑ Exclude at 95% CL: [100 - 106] GeV/c^2
- ❑ Tevatron combination still most sensitive in $H \to b\bar{b}$



Conclusions

- CDF and DZero combined Higgs to $b\bar{b}$ searches using **the full Tevatron dataset**, ~ 10 fb^{-1} / experiment
- **Validated the techniques in a VZ search with Z->bb**
  - 4.6 standard deviations excess above bkg, measured cross section consistent with SM
- **Combined three VH channels (lvbb, llbb, vvbb)**
  - **Most sensitive search in this channels**
  - **Expected** 95% CL exclusion: [**100 - 116**] GeV/c^2
  - **Observed** 95% CL exclusion: **[100 - 106]** GeV/c^2
  - **A broad excess larger than 2 s.d. is seen in 115-145 GeV/c^2**
- **Result consistent with a 125 GeV/c^2 SM Higgs boson**
- **The bb decay is essential** to establish if the SM Higgs-like boson observed by both ATLAS and CMS is indeed the SM Higgs boson

Backup Slides

Bayesian Method

Bayesian Posterior Probability

$$p(R|\vec{n}) = \frac{\int\int d\vec{s}d\vec{b}L(R,\vec{s},\vec{b}|\vec{n})\pi(R,\vec{s},\vec{b})}{\int\int\int dRd\vec{s}d\vec{b}L(R,\vec{s},\vec{b}|\vec{n})\pi(R,\vec{s},\vec{b})} \Rightarrow \int_0^{R_{0.95}} p(R|\vec{n})dR = 0.95$$

$R = (\sigma \times BR)/(\sigma_{SM} \times BR_{SM}),\ R_{0.95}$: 95% Credible Level Upper Limit

$\vec{s},\vec{b},\vec{n} = s_{ij}, b_{ij}, n_{ij}$ (# of signal, background and observed events in j-th bin for i-th channel)

π : Bayes' prior density

Combined Binned Poisson Likelihood

$$L(R,\vec{s},\vec{b}|\vec{n}) = \prod_{i=1}^{N_{\text{channel}}} \prod_{j=1}^{N_{\text{bin}}} \frac{\mu_{ij}^{n_{ij}} e^{-\mu_{ij}}}{n_{ij}!}$$

Principle of ignorance

- for the number of higgs events (instead of higgs Xsec)

$$\pi(R,\vec{s},\vec{b}) = \pi(R)\pi(\vec{s})\pi(\vec{b}) = s_{tot}\theta(Rs_{tot})\pi(\vec{s})\pi(\vec{b})$$

$s_{tot} = \Sigma_{i,j} s_{ij}$: Total number of signal prediction

$\pi(x) = G(x|\hat{x},\sigma_x) \quad (x = s, b)$ $\quad$ $\hat{x}$: expected mean, σ_x: total uncertainty

Modified Frequentist Method

- H_1 = test hypothesis (signal + background)
- H_0 = null hypothesis (background only)
- LLR = -2 ln $p(data|H_1)/p(data|H_0)$
- 2 p-values: 1-CL_b and CL_{s+b}
- 1-CL_b = probability for an upward background fluctuation to create the background-plus-signal-like data excess
  - 1-CL_b = $p(LLR<=LLR_{obs}|H_0)$
- CL_{s+b} = probability for a downward signal-plus-bkg fluctuation to create a background-only-like data
  - CL_{s+b} = $p(LLR>=LLR_{obs}|H_1)$
- $CL_s = CL_{s+b}/CL_b$. Exlusion of H_1 if $CL_s< 0.05$

Sensitivity at 125 GeV/c^2

- Sensitivity (95% CL expected upper limits) in H->bb in VH, where V decays to two charged leptons and H->bb
- **Tevatron (CDF + Dzero) 1.40 x SM**
- **CMS 1.64 x SM**
- ATLAS no VH-only combo yet

- For CMS 37% more luminosity (at the same analysis techniques) will be enough to catch up withTevatron
  - This will be achieved in 2012 data taking