$Ht\bar{t}$ in Run 2

Joe Incandela
University of California, Santa Barbara

With all of the work done by
A.Colijn, J.Goldstein, P.Merkel, T.Nelson, S.Parke, D.Stuart, D.Rainwater
Fermilab
And
C. Hill
University of California, Davis

Outline

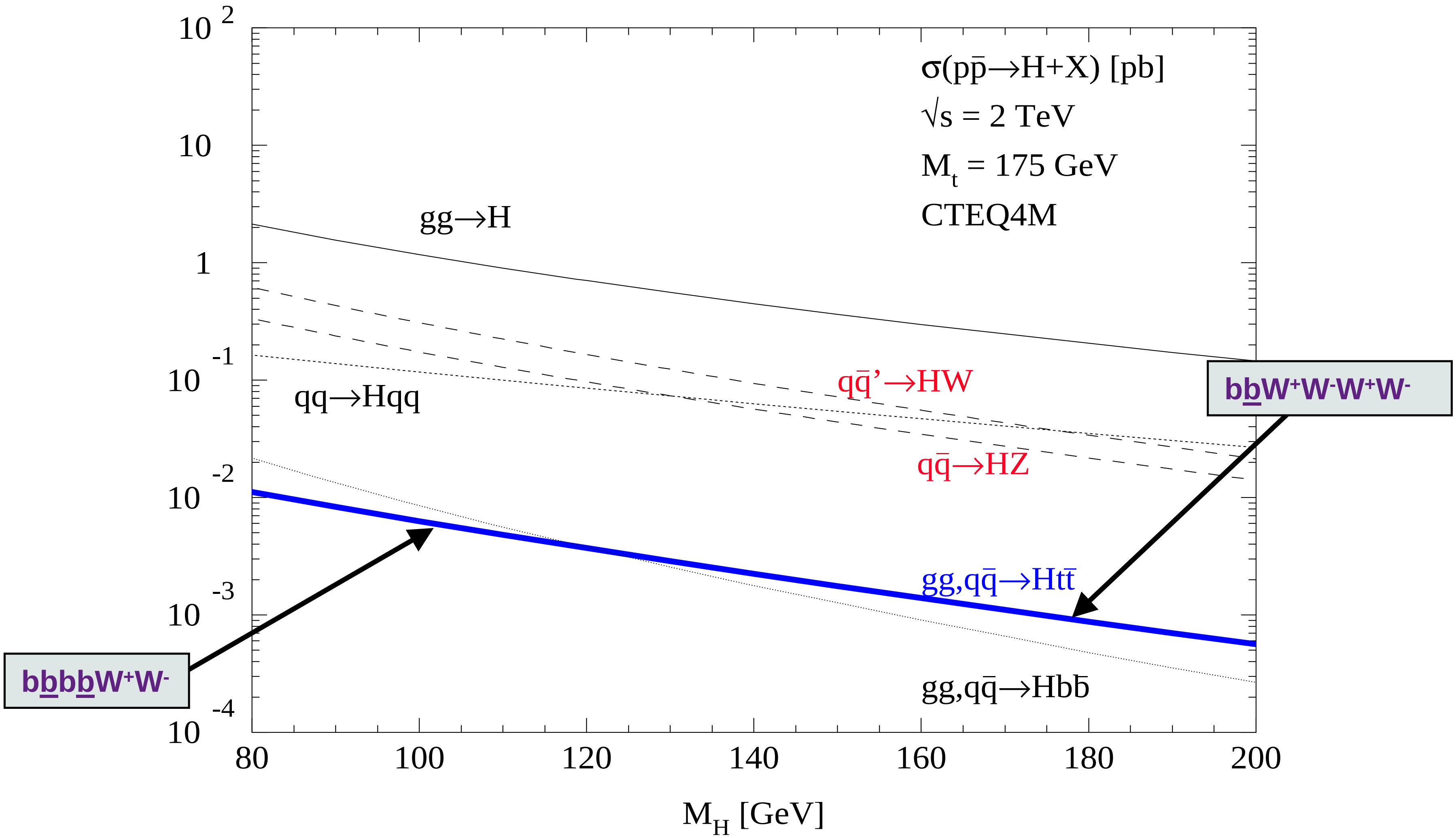
- Introduction & Overview
- Comparison of $Ht\bar{t}$ in Run 2 to $t\bar{t}$ in Run 1a
- Event topologies and backgrounds
- The obvious approach: Reconstruct $t\bar{t}$
- A less obvious approach (a work in progress)
- Impact on Higgs sensitivity (preliminary)
- What's next…

Introduction

- Tev2000 : Low mass higgs ?
  - Need to pull out all the stops: Higgs was a major part of the motivation for:
    - **CDF ISL, proposed 1996: Increase range for b tagging from $|\eta| < 1$ to $|\eta| < 2$**
    - **CDF L00, proposed 1998: Improve in Impact Parameter (IP) resolution**

- SUSY/HIGGS workshop: Low mass higgs within reach
  - Need to combine experiments and channels
  - Need as much data as possible
  - Need the detectors to work optimally.
- i.e.… it looks kind of marginal
  - Is there anything else we can do ?

Htt̲

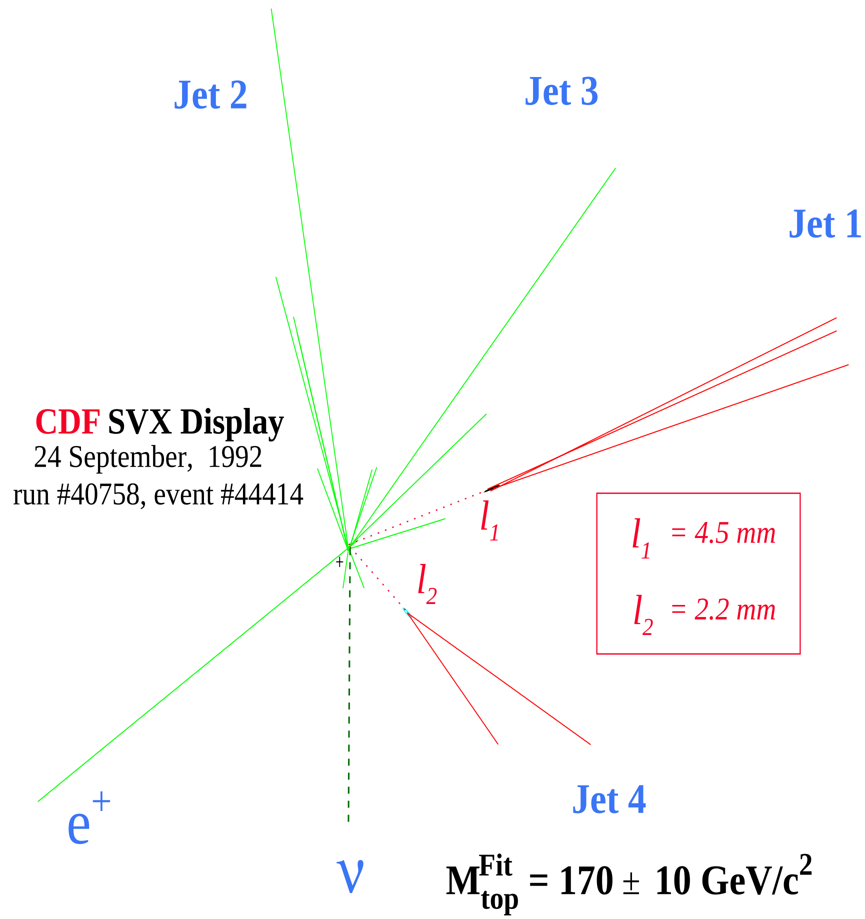


First considered by J.Gunion et. al. in the context of SSC, LHC

Phys Rev Lett 71 2699 (1993)

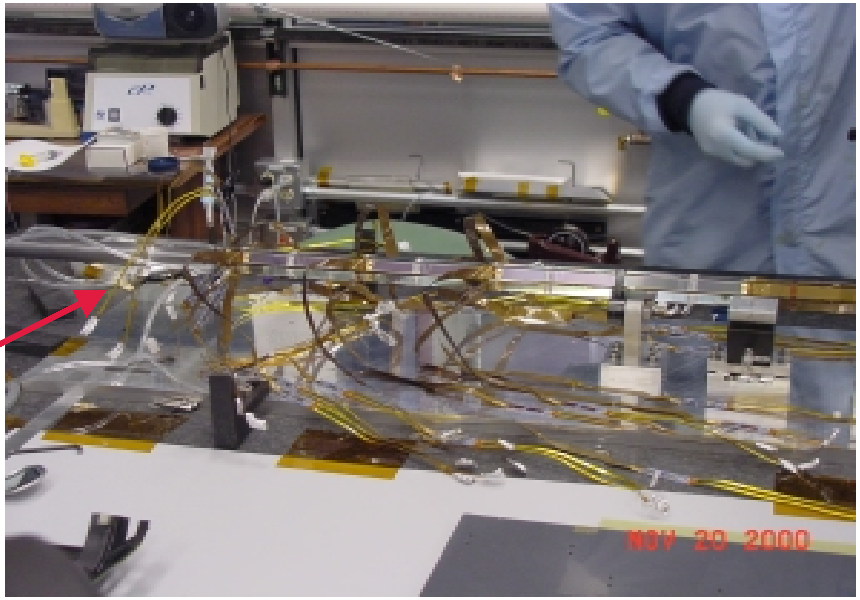
Comparison with top in Run 1A

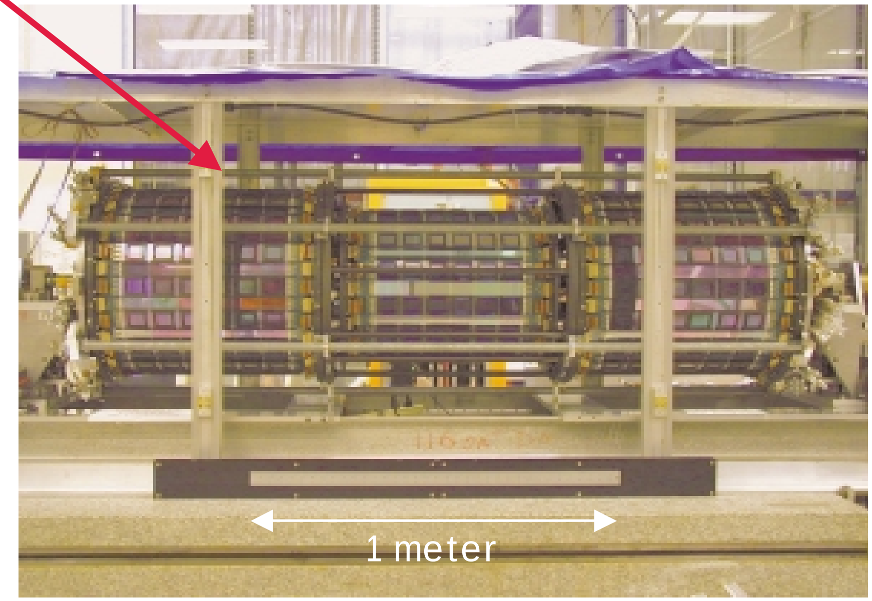
- Events are rare but distinctive
  - $b\underline{b}b\underline{b}W^+W^-$ or $b\underline{b}W^+W^-W^+W^-$
- Event Count
  - Run 1a ~ 100 $t\underline{t}$ in 20 pb^{-1}
  - Run 2 ~ 100+ $t\underline{t}H$ in 15 fb^{-1}
- Background
  - Run 1a $\rightarrow$ W+jets
  - Run 2 $\rightarrow$ $t\underline{t}$+jets
- Comparisons
  - Con: Events are more crowded
  - Pro: Detector is better
- Conclude: this channel could add something:
  - If nothing else, at least a few great events like the first CDF $t\underline{t}$ event …
    - Example: Large missing E_T, 4 high E_T b tagged jets and 2 opposite sign, isolated, high E_T Leptons most likely of different flavor
    - Example: Large missing E_T, 2 high E_T b tagged jets, 3 isolated leptons or 2 like-sign leptons…



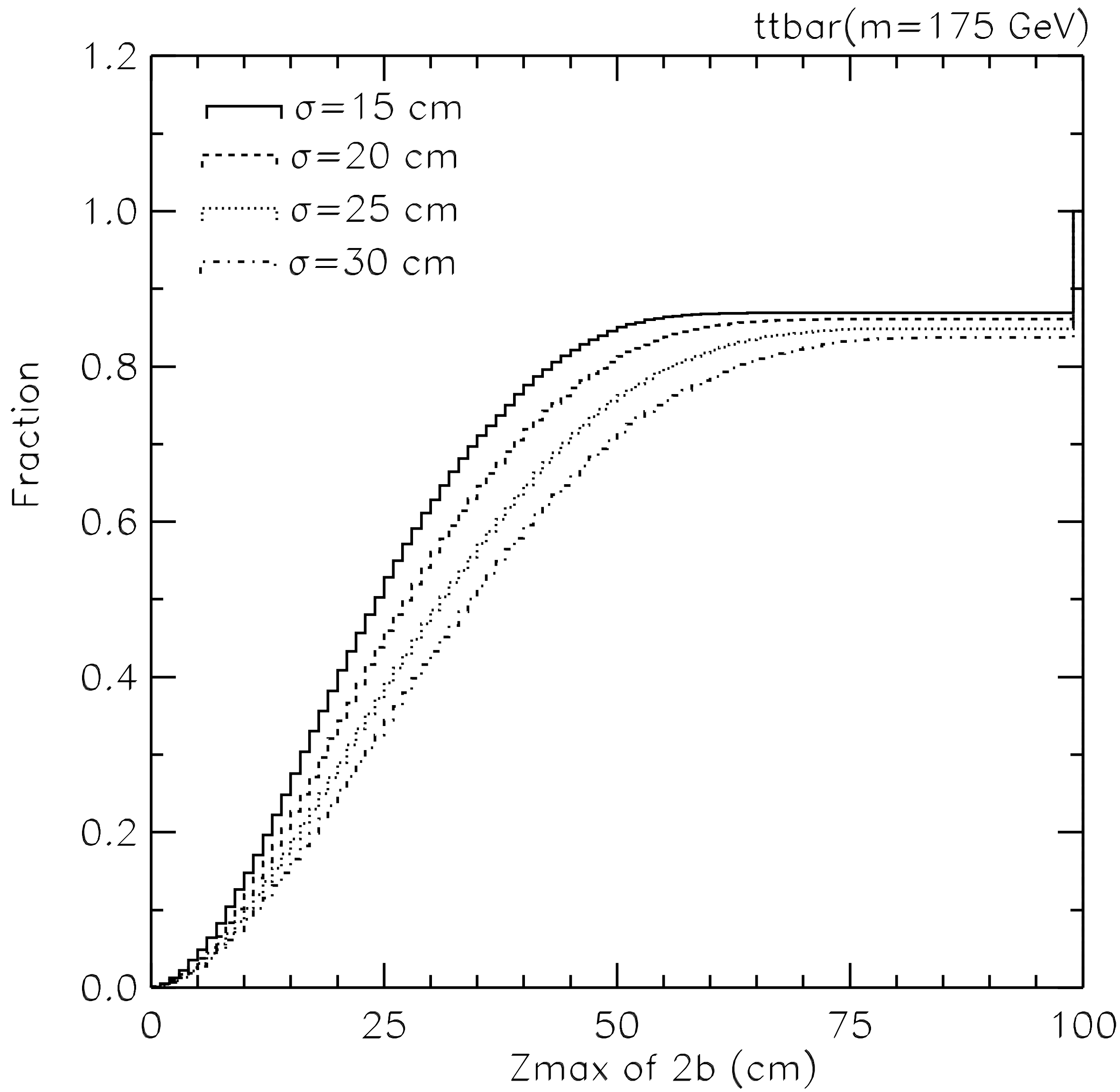
Leveraging what we have…

- $t\bar{t}$ the primary benchmark for optimization of CDF II.
- We will learn a lot from large $t\bar{t}$ data samples in Run 2.
- B-tagging Technology continues to improve. We expect higher efficiency and purity
  - Run 2A Silicon
    - **Better IP resolution (L00)**
    - **Increased coverage (ISL)**
    - **Better z-matching (SVXII)**
    - **2 track IP trigger (SVT)**
  - Run 2B Silicon
    - **Will likely have**
      - **Less mass**
      - **Better resolution**
      - **More robust pattern recognition**
    - **Smaller luminous region will also produce much higher acceptance for the same detector scale**





Acceptance in Run 2b



Plot courtesy of Weiming Yao/LBNL

Signal Topologies and Backgrounds

$Ht\bar{t} \rightarrow W^+W^-b\bar{b}b\bar{b}$ events		
$jjjjb\bar{b}b\bar{b}\ \not{E}_T$	~55 %	Backgrounds ?
$l\ jjb\bar{b}b\bar{b}\ \not{E}_T$	~38 %	Focus of our study
$l^+l^-\ b\bar{b}b\bar{b}$	~7 %	Few events ?

- Mainly investigated lepton+jets
- $t\bar{t}$ + jets by far the largest background
  - Reduce first with B tagging & cuts
    - At least 5 high E_T jets and 3 b tags
  - Further separation via kinematics
- $t\bar{t}b\bar{b}$ (subset of $t\bar{t}$ + jets)
  - Same components as $Ht\bar{t}$ but the 2 additional b jets are softer.

∃ scale uncertainties in matrix element calculations

$t\bar{t}$+jets sample in Run II will be large enough to calibrate the MC

THE BACKGROUNDS

K-factor of 1.33 included for all backgrounds.

backgrounds to $H \rightarrow b\bar{b}$	σ (fb)
$t\bar{t} + jj \quad (\Delta R(jj) > 0.4)$	1030
$t\bar{t} + b\bar{b}$	27
$t\bar{t} + Z, Z \rightarrow b\bar{b}$	1.5
$WZ + jj, Z \rightarrow b\bar{b}, W \rightarrow e\nu, \mu\nu$	10.4

backgrounds to $H \rightarrow W^+W^-$	σ (fb)
$t\bar{t} + jj \quad (\Delta R(jj) > 0.4)$	1030
$t\bar{t} + W$	17
$t\bar{t} + Z, Z \rightarrow \ell^+\ell^-$	0.9

Sequence of studies: from idealized to realistic

Low mass case ($H \rightarrow b\bar{b}$)

1. Quasi generator level study with $t\bar{t}$ reconstruction
   - Selection of at least 6 jets, lepton, $\not{E}_T$, 3 or more b tags
     - **CDF run 1 calorimeter used for jets.**
     - **Parametric simulation of run 2 tracker**
     - **Run 1 b tagger**
   - Assume you can reconstruct the hadronic top decay or both tops
     - **Unrealistic but leads to some new thoughts**

2. As close to full simulation as we could muster for now
   - MC Events: 140k $t\bar{t}H$ with M_H =120 and 3.4M $t\bar{t}X$
   - Jets: Run 1 full simulation, no assumed improvements in jet energy resolution
   - Leptons: Run 2 acceptance but Run 1 efficiencies
   - Tracking and B tagging:
     - **Parametric model of tracker with resolutions taking into account material**
     - **We studied optimization of b tagging for additional tracker capabilities but we partially discount capabilities to compensate for pattern recognition effects.**

Event Generation

- Signal and background: Pythia 6.129
- Signal normalization given by exact tree-level matrix elements
  - Generated by MADGRAPH
  - COMPHEP
  - NLO corrected decay rates of Higgs via HDECAY
  - K-factor taken to be ~1.33
- $t\bar{t}$+j, $t\bar{t}$+jj parton level cross sections calculated using exact tree-level matrix elements
  - $\sigma_{ttjj}/\ \sigma_{tt} \approx 1/7$
  - Agrees with Pythia within matrix element uncertainty

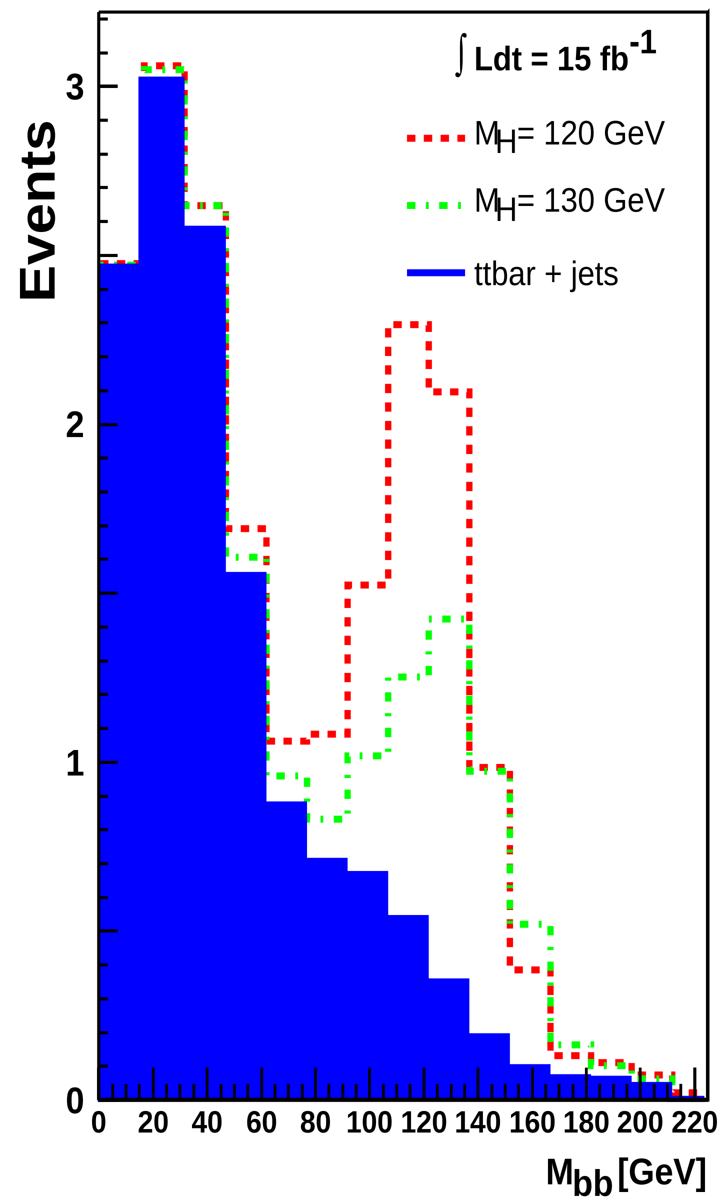
Run 2 Detector Simulation

- Calorimeter:
  - GFLASH and CDF Calorimetry code for simulation and reconstruction.
  - Jet clustering with default CDF clustering algorithm (jetclu)
    - cone size of 0.4 and a minimum jet Et of 7 GeV.
- Parametric model of CDFII tracking:
  - Gaussian smearing of generator-level particle information.
  - Estimates of the full tracking covariance matrices for smearing were obtained from studies of the expected resolution of the Run 2a silicon system after including all material effects.
  - Pattern recognition and other tracking inefficiencies were not included
- B tagging
  - For the 1st study we used the Run 1 SECVTX algorithm.
  - For the 2nd study we defined new tight and loose tagging levels but did not use the highest efficiencies we saw in order to take into account real pattern recognition and tracking losses.

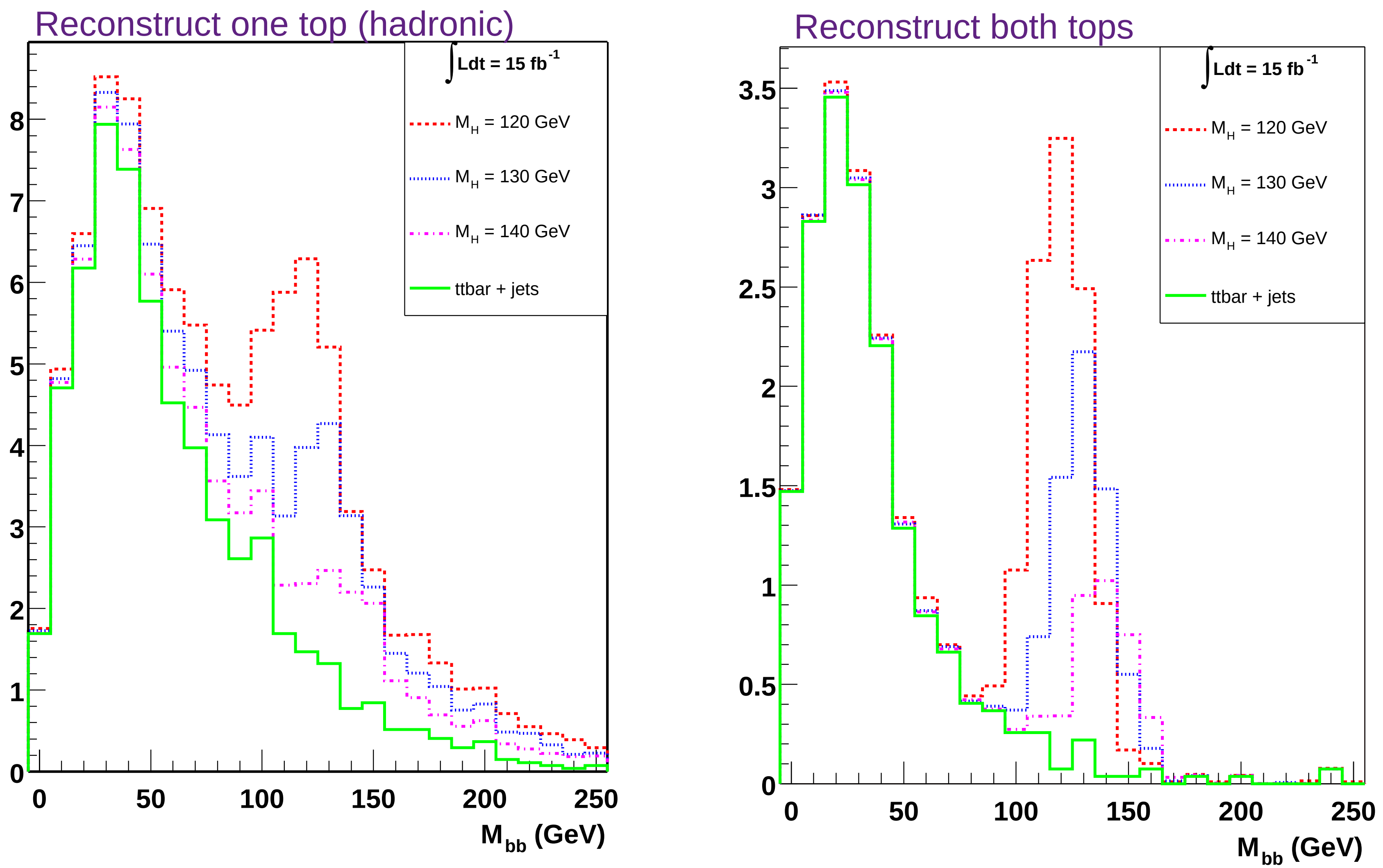
IF you could group tt̲ products with high efficiency

- Selection Cuts
  - 1 isolated lepton > 15 GeV
  - Missing ET > 15 GeV
  - 4 jets greater than 15 GeV
  - 2 additional jets > 10 GeV
    - Required to reconstruct tt̲
  - B Tag w/Run 1 algorithm
    - $\varepsilon_b \sim 60\%$, $\varepsilon_c \sim 25\%$, $\varepsilon_J \sim 0.5\%$
  - 3 b-tags to Reject tt̲jj background
- M_H = 120 Gev, 15fb^{-1}, CDF only

  ~7 signal events and ~14 tt̲bb̲/cc̲ background
- Grouping top products
  - tt̲H has many more combinations but also more b tags and *you don't care if you reconstruct top correctly, you just want to group all the tt̲ products*
  - Efficiency depends on No. and purity of b-tags (CDF Run 1 for 2 b tags ~ 60%)
  - Assuming very high grouping efficiency, you get a clear excess in the dijet mass plot



Improved Mass Resolution*



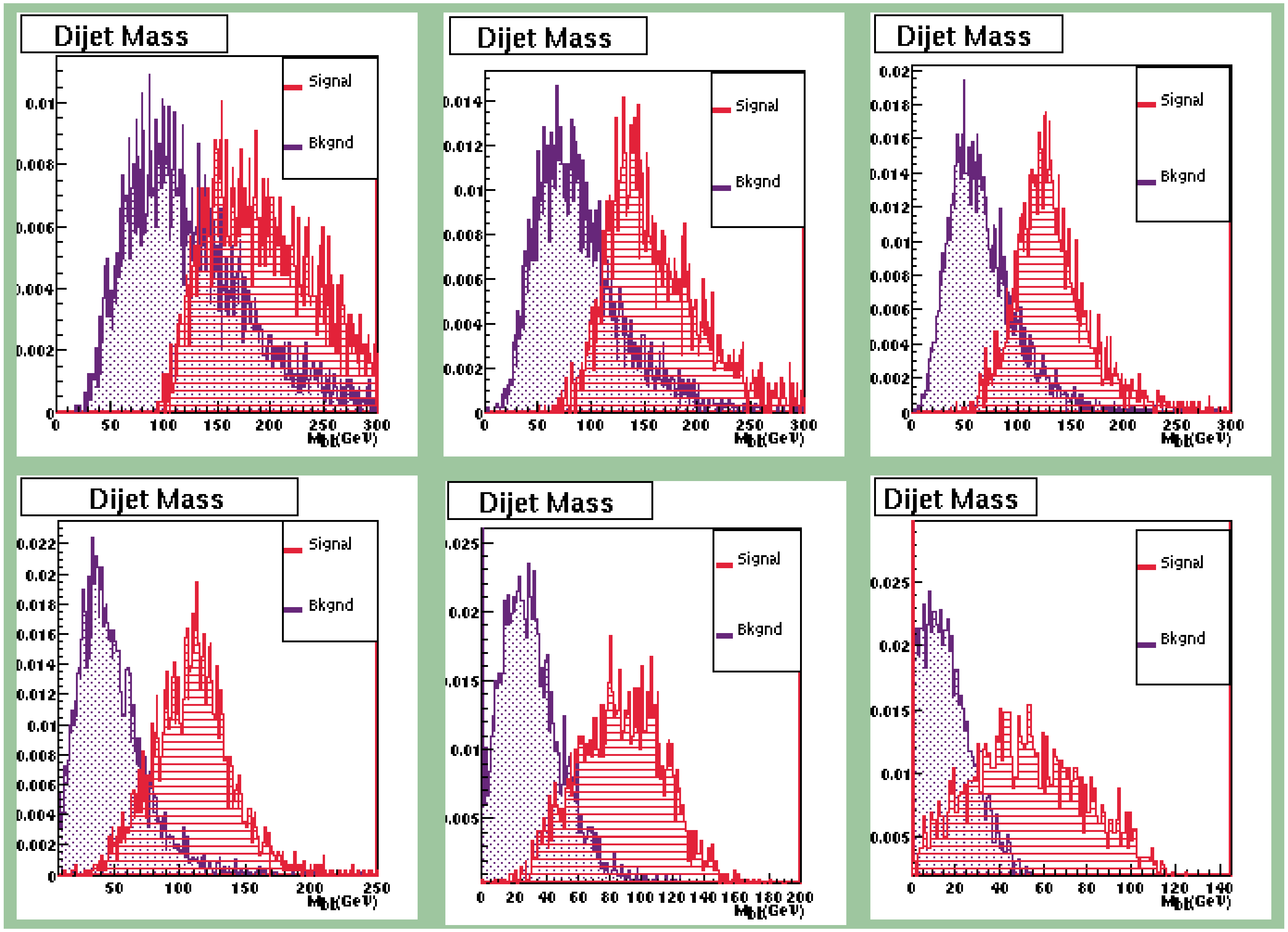
* 10% m_{bb} resolution: SUSY/HIGGS Report section III

Next Considerations

- High reconstruction efficiency is difficult to achieve
  - In any case we didn't want to spend our time working on this before we got real $t\bar{t}$+jets data …
  - M_{bb} plot does however tell us something interesting:
    - the non-$t\bar{t}$ part of the event which is the only difference between signal and background is in fact *quite* different

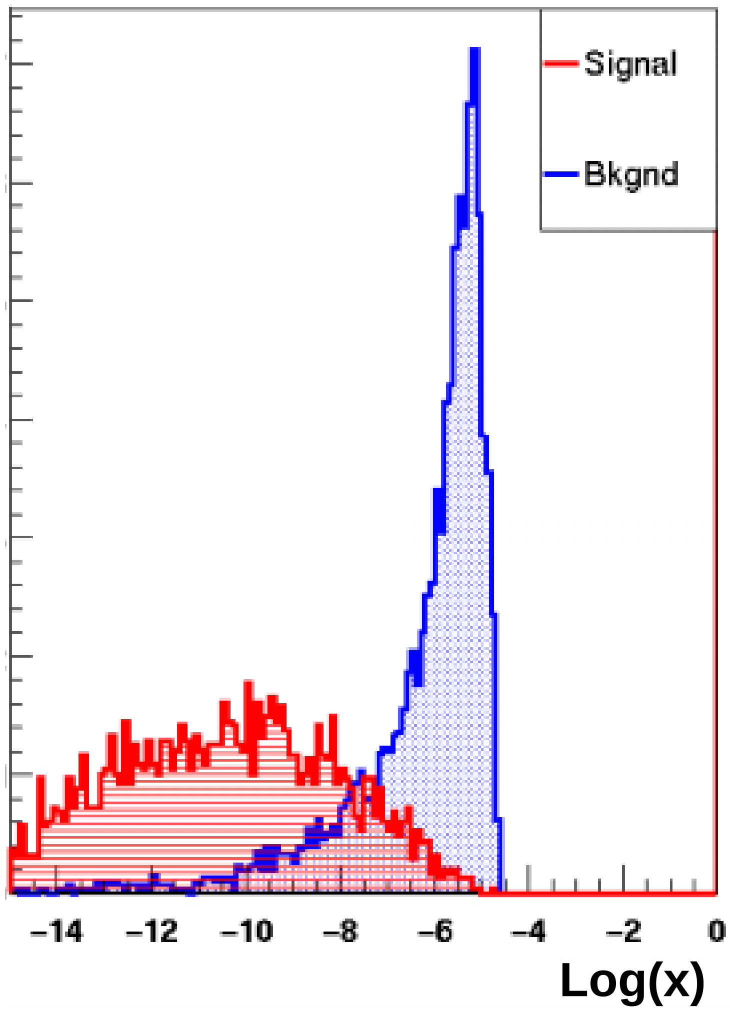
- Subsequent Approach: Try to exploit this difference.
  - We want to use no *a priori* knowledge of jet origins and minimal fitting.
  - Consider the following:
    - Require at least 3 b tagged jets
    - If only 3, choose a 4^{th} jet at random and assume it is a b jet
    - Form the 6 invariant mass pairs of these 4 jets and order them
    - Plot the 6 ordered mass distributions for signal and background
  - Based on the previous plot of M_{bb} for signal and background, we expected these distributions to show some observable differences between signal and background.

First look at the ordered dijet mass pairs



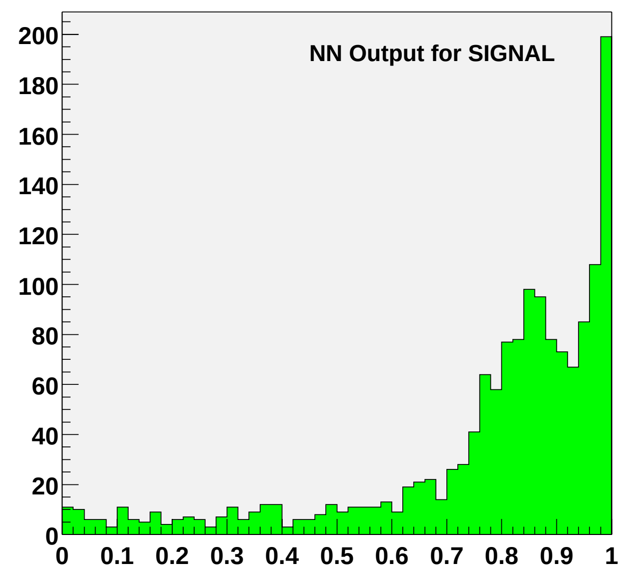
- $Ht\bar{t}$ compared to $t\bar{t}b\bar{b}$ (no special normalization)
  - Smeared partons (use the standard CDF Run 1 resolution without improvements)
  - Run 1 b tagger and parametric Run 2 tracking

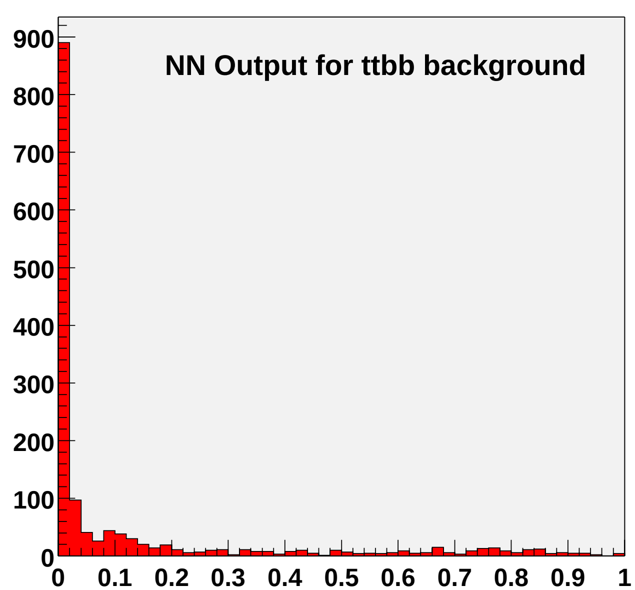
First try at a discriminant



- Form Templates
  - Unit normalize the plots from the $t\bar{t}b\bar{b}$ background (arbitrary binning). The resultant 6 curves are then treated as functions $f_j(m)$
- Calculate an event discriminant
  - For any event, (signal or background) calculate the product x of the 6 amplitudes you get from the functions $f_j\,(m_j)$ using the 6 ordered masses m_j
- Compare Log(x) distributions

Neural Net



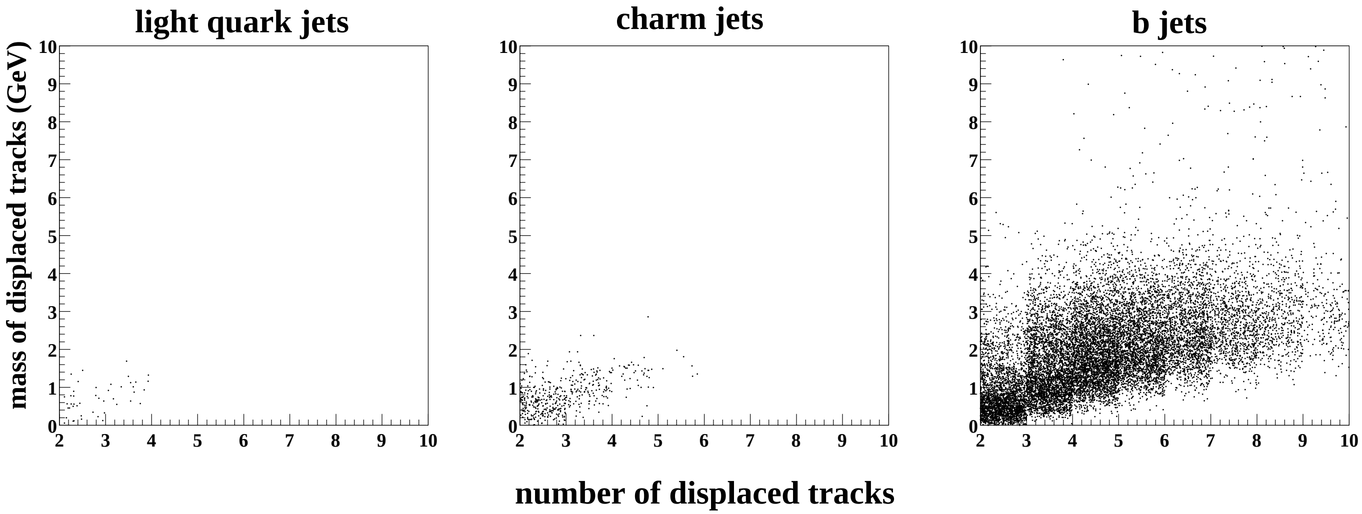


- How to define an optimal discriminant?
  - No time to think about it so we let a Neural Net do the thinking for us…
- Neural Net
  - $t\bar{t}H$ and $t\bar{t}b\bar{b}$ as described above to train NN
    - Used Root_Jetnet [CDF note 5434]. to interface to NNs based on the FORTRAN program JETNET [L. Lonnblad, C. Peterson, H. Pi, T. Rognvaldsson, Comput. Phys. Commun. 81, 185 (1994)] can be trained within ROOT.
    - Use back propagation algorithm with one input layer, one hidden layer and one output node.

Retain ~50% signal for > 98% background rejection

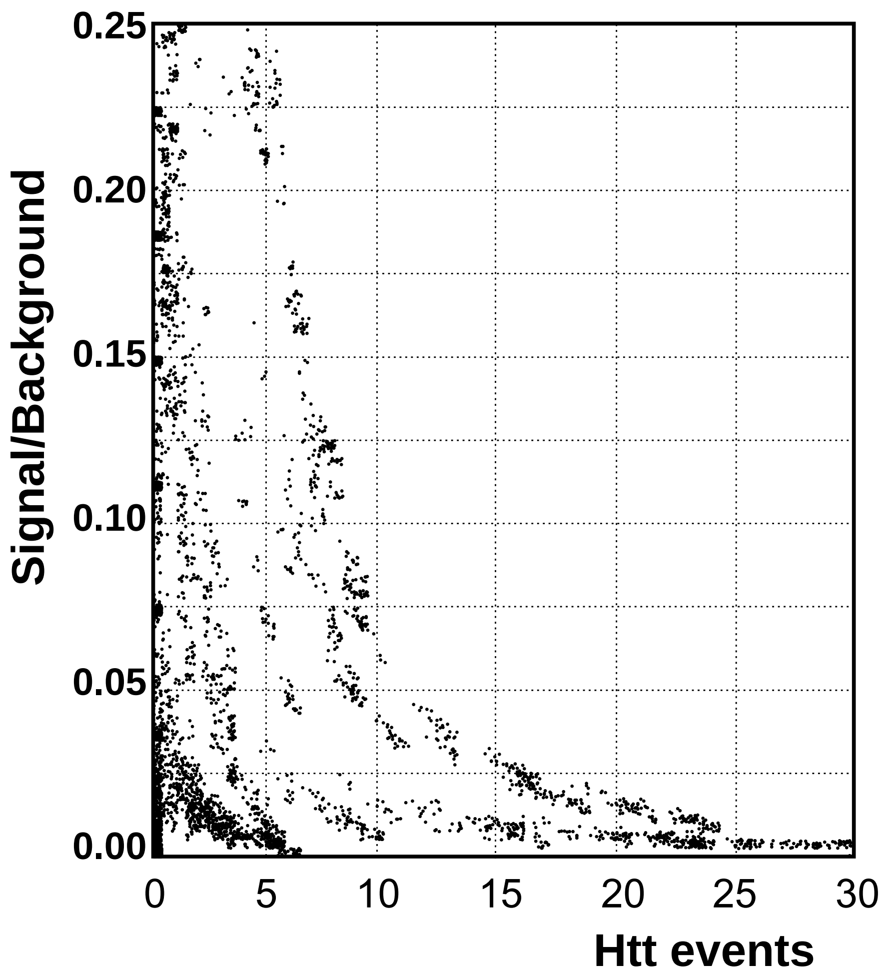
- Time to try a realistic analysis
  - Generate millions of $t\bar{t}X$
  - Use crappy energy resolution.
  - Assume a bit better b tagging for Run 2.

Study B tagging in Run 2



- Light and c quark tagging has to be controlled
  - ttjj problematic due to potential tag of charm from W decay.
  - Studied how to discriminate against charm.
- Displaced track multiplicity and vertex mass
  - N ≥ 2 displaced tracks eliminates most u,d,s quark jets
  - To cut charm: require M_{VTX} > 2 GeV for N = 2
- Efficiency and Discrimination
    - Loose: ε_b ~ 75% ε_c~30%
    - Tight: ε_b ~ 50% ε_c~4%
  - Significantly better results were seen but we do not use them.
  - Other methods (multivariate, NN) could pay off …
    - ALEPH achieved 87%/85% efficiency/purity

Event Counts and S/B After Selection

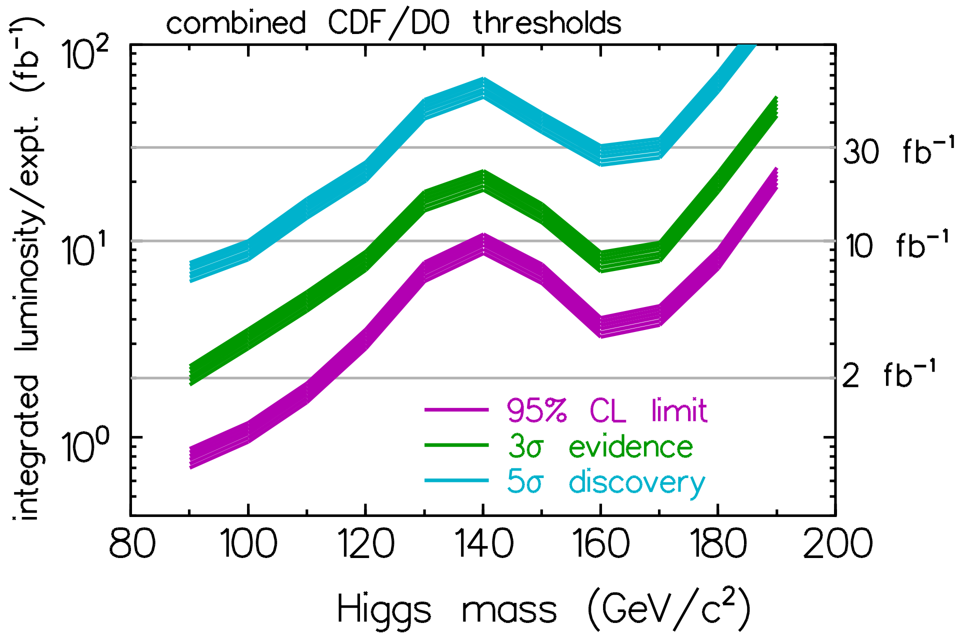


- Initial Selection (15 fb^{-1} M_H=120)
  - Use PYTHIA to generate tt and do standard calorimeter simulation.
  - Vary selection cuts and look at S/B versus Htt event count.
- Some examples (approx.)
  - 20 Signal on 1000 background
  - 10 signal on 120 background
  - 8 signal on 65 background
- For 50% (98%) S (B) these would become:
  - 10 signal on 20 background
  - 5 signal on 2.4 background
  - 4 signal on 1.3 background

Preliminary Results

- t$\underline{t}$jj as opposed to t$\underline{t}$b$\underline{b}$
  - Often the third tag is a charm jet.
  - Alters the m_{bb} distributions by bringing a harder W jet into the mix
    - the distinction between signal and background is then diminished.
  - NN: 90% background rejection for 50% signal acceptance
- Improvements should be possible
  - Use tagging information to select the tags that go into the templates
  - Use t$\underline{t}$ fitter to help find the b jets (see next slide)
- We haven't yet attempted ANY optimization of 4 b jets selection
  - *Guess* 95% background rejection should be achievable
- For now 90% case in 15 fb^{-1} and for 120 GeV Higgs yields
    - 10 signal on 100 background, or
    - 5 signal on 12 background, or
    - 4 signal on 6.5 background

Impact on Higgs Search

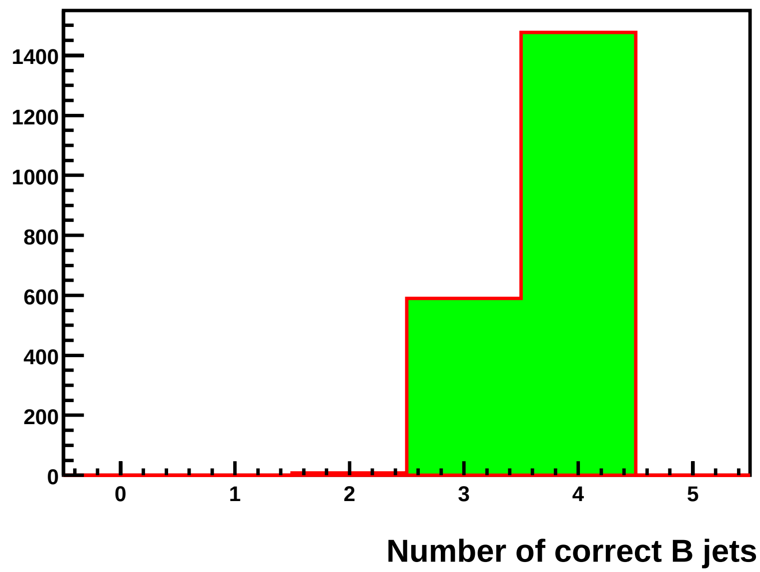


What's the impact ?

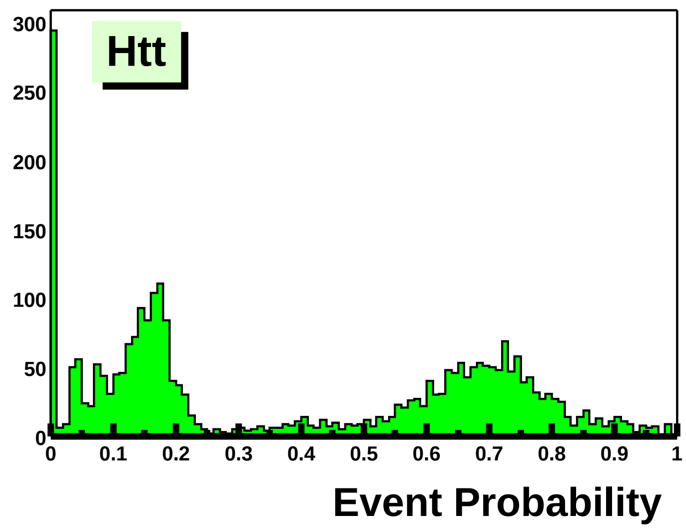
- If each experiment gets 4 ttH on 6.5 background in 15 fb^{-1}, then the amount of W/Z+H data needed for 95% CL exclusion or 3 and 5 σ observations of 120 GeV Higgs is scaled by ~0.85
- With optimization, one may do better
- Contingency for the W/Z+H searches

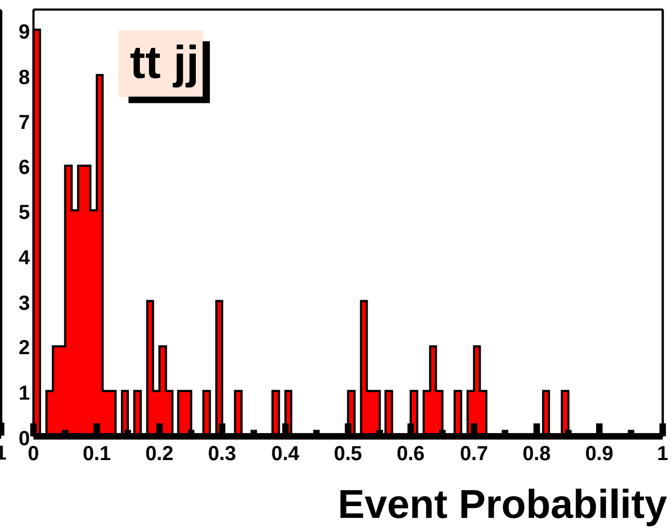
Signal	Background	95% CL	3 Sigma	5 Sigma
4	4	0.81	0.8	0.79
8	8	0.68	0.66	0.67
4	8	0.88	0.87	0.87
8	16	0.77	0.78	0.81

Using tt fitter to help select 4 b jets



- Studied 6 jet/3 tag bin only so far
  - Lots of combinations
  - 2 jets left out of best tt fit are from higgs only ~25% of the time
- Use tt fitter for b jet finding instead of tt reconstruction!
  - Starting with 3 tight tags, you find all 4 b jets correctly in 1476 of 2074 events
- We are also studying the use of the tt fitter to define an event probability for Htt and ttjj





High Mass Higgs

THE $H \to W^+W^-$ CHANNEL

This is clearly more difficult.

1 $\geq 3\ell$ $(2j)$ $b\bar{b}$ $\not{p}_T$ – ~ 2 events before tagging

~ 1 event per experiment w/ minimal cuts $+ 1$ b tag

event(s) would be spectacular, but need several to convince!

2 2ℓ $4j$ $b\bar{b}$ $\not{p}_T$ – ~ 7 events before tagging

$\sim 3 - 4$ events per experiment w/ minimal cuts $+ 1$ b tag

better: same sign leptons $\to \sim 1 - 2$ events per experiment

Conclusions

"realistic analysis" indicates $t\bar{t}H$ channel may contribute

- **Preliminary result**: Reasonable assumptions indicate ~15% reduction in luminosity threshold for discovery at 120 GeV
- Can probably be further optimized

- The limiting factors for this channel are
  - machine luminosity
  - b jet acceptance and b tagging efficiency/purity
- Our Plans
  - Optimize/refine study. Firm up results and document it by June.
  - Try this out on LHC Monte Carlo samples.
- There are many uncertainties
  - Delivered luminosity, detector performance, b acceptance, signal and background cross-sections
- Really need Run 2a Data
  - Not much point in working a lot more on the methodology until we have real data to look at….
  - We'll know much more about $t\bar{t}jj$ rates, kinematics, b tagging
    - Measure tt+jets cross-sections, scale etc.