Search for rare SM processes in the $\not{E}_T+b$-jets signature at CDF

Karolos Potamianos

PURDUE UNIVERSITY

April, 8 2011

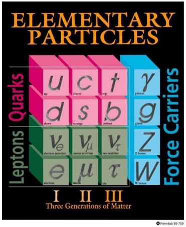
IIHE (ULB-VUB) Seminar

Outline

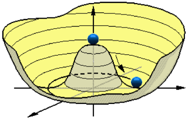
Particle physics at the end of the 20th century.

The Standard Model of particle physics

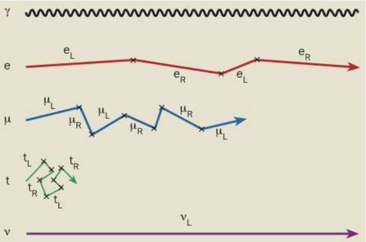
- QED and QCD are very successful effective theories predicting the world around us;
- Yet, these theories cannot accommodate massive elementary particles without loosing gauge invariance and becoming un-renormalizable (and loosing their predictive power);
- In 1964, several physicists proposed a solution to give a mass to the elementary particles: the BEHHGK mechanism;
- This mechanism introduces a field with which particles interact and acquire mass;
- However, a predicted particle remains yet (as of 2011) unobserved: **the Higgs boson**;



+

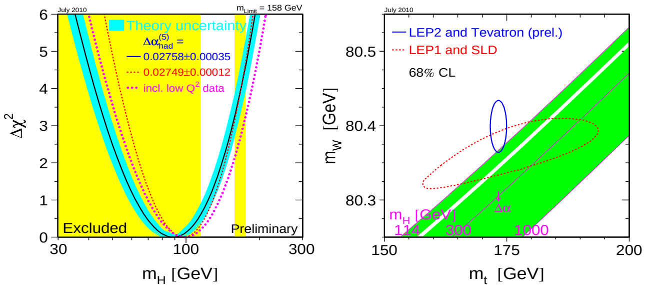


=



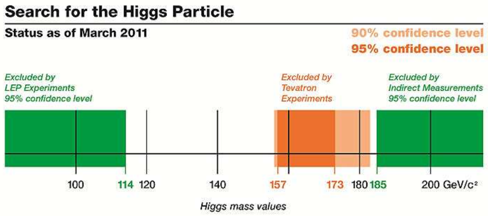
The search for the SM Higgs boson

- The Higgs boson is a key piece of the SM;
- The LEP experiments and indirect searches constrain the Higgs mass;
- EW fits and direct searches: $114 < m_H < 185$ GeV$/c^2$ at 95% C.L.



The search for the SM Higgs boson

- The search for a SM Higgs boson is a central part of the Tevatron Program (and, of course, at the LHC);
- Recently, the observation of single top and diboson production have brought us closer to elusive particles such as the Higgs;
- Sine two years know, the Tevatron experiments have been putting new constraints on the Higgs: it cannot hide much longer;



The SM Higgs is (if it exists) (possibly) within Tevatron's (and LHC's) reach !!!

Yet unsolved puzzles

- In the 1980's, it was thought that the SM picture was complete;
- Later on, experiments discovered dark matter and dark energy, telling us that we actually don't know what 95% of the Universe is made of;
- Also, we know since 1998 that neutrinos oscillate and are massive, something which cannot be accounted for by the SM. Moreover, they have a tiny mass;

There are several unsolved puzzles in nature, fortunately for us!

A possible path to a solution to this puzzle

Good experiments and powerful analysis techniques

- With the **Tevatron** not having said its last word, and the **LHC** just starting to speak, we have many reasons for expectations;
- The involved collaborations are actively developing **new tools and techniques to discover new particles**;

Many theories are available

- Today, an experimental physicist faces **many theories** that can provide a (partial) solution to the puzzle:
  - SUSY, extra dimensions, technicolor, extra generations, heavy vector bosons...

A powerful signature: missing transverse energy ($\not{E}_T$)

- These unobserved particles **either decay into neutrinos** or **do not decay into known matter**, leaving an **energy imbalance in the detector**;

The $\not\!\!E_T$+b-jets signature

Interesting signature in searching for **both SM and BSM physics**

- $ZH \to \nu\nu b\bar{b}$ is one of the most sensitive decay modes for a low mass Higgs;
- $ZZ \to \nu\nu b\bar{b}$ is on the road to the Higgs;
- Sensitive to single top, mainly through hadronic τ decays;
- SUSY: $\tilde{b}\bar{\tilde{b}} \to b\tilde{\chi}^0\bar{b}\tilde{\chi}^0$;
- Technicolor: $\rho_T^\pm \to Z\pi_T^\pm \to \nu\nu b\bar{q}$;

In this talk, I'll present the analysis technique developed to search for particles in this signature and the analyses it made possible.

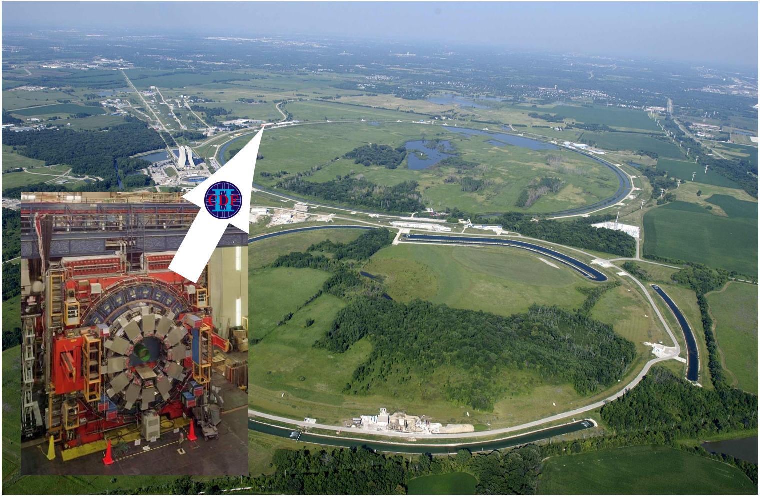
Though the technique itself does not rely on a specific detector, it was developed while analyzing CDF II data.

The Tevatron

A proton-antiproton collider with a center of mass energy of 1.96 TeV.

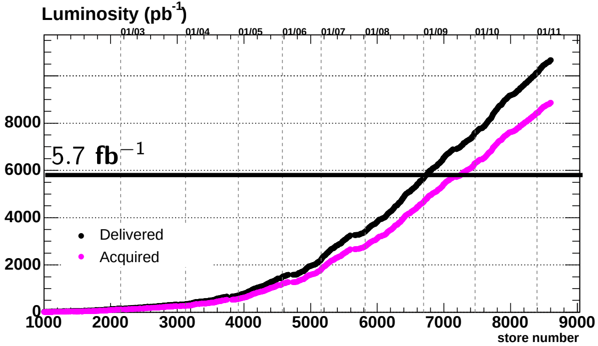
The Tevatron

The CDF experiment.



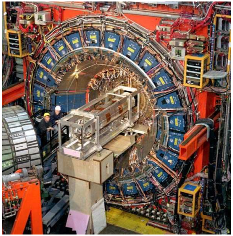
The Tevatron

The Tevatron and CDF are outstanding performers in delivering and acquiring quality data.



I will discuss and present results analysing up to 5.7 fb^{-1}.

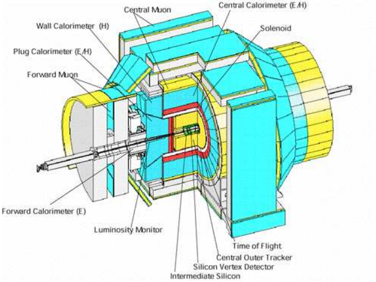
The CDF II detector



- A well-understood multi-purpose detector;
- Relatively compact: 10m×10m×15m;
- Basic detection sub-systems: two tackers, a calorimeter, and muon chambers;

The CDF sub-detectors

Silicon Detector:	Precision vertex detection and heavy flavor tagging up to $\|\eta\| \sim 2$;
Drift chambers:	Charged particle p_T measurement up to $\|\eta\| \sim 1.5$;
Calorimeters:	Electromagnetic and hadronic, with 4π coverage up to $\|\eta\| \sim 3.5$;
Muon chambers:	Muon identification up to $\|\eta\| \sim 1.5$;



Coverage is not very good compared to LHC experiments.

Charged lepton ID

Electrons/Positrons

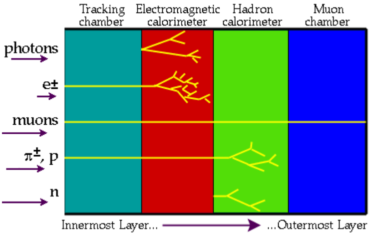
- A match between a track in the central tracker and EM calorimeter deposit;
- A shower maximum (to reject π^0); An isolation (to reject showers from quarks);

Muons

- A match between a central track and the muon chambers or an isolated track;

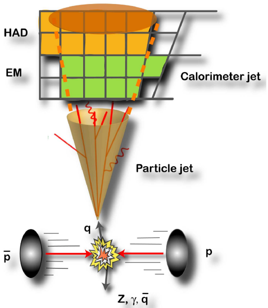
Taus

- Until recently, no explicit ID. We do not consider any ID in the analyses presented here;



Jet reconstruction at CDF

- Bare quarks hadronize and produce particle jets;
- Jets are reconstructed with a seeded cone algorithm that loops over the calorimetric towers;
- The reconstruction efficiency is nearly 100%, with an **almost complete angular coverage**;
- The jet energy scale is known to $\sim 3\%$ and is the cause to systematically limited analyses;
- The jet energy **resolution is driven by** that of **the hadronic calorimeter**: $80\%/\sqrt{E_T}$;
- Un-instrumented regions of the calorimeter lead to underestimation of a jet's energy, yielding an **energy imbalance** in the detector;

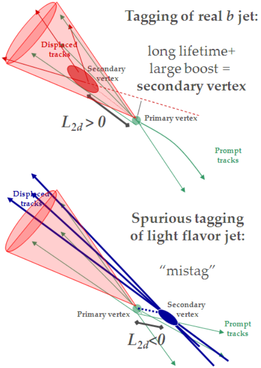


The analyses presented here use additional information from the tracker, using the **H1 algorithm**.

After identifying jets in the calorimeter, we further use detector information to isolate jets originating from b-quarks.

b-quark identification (a.k.a. b-tagging)

The main goal of b-tagging is to enhance the presence of the $ZH \to \nu\nu b\bar{b}$ signal by identifying jets originating from b-quarks in the final state.



How to identify b-jets?

- Bare b-quarks hadronize into B hadrons (long-lived);
- These hadrons travel before decaying ($L_{xy} \simeq 500\mu$m), yielding a **secondary vertex**;
- The decay products yields wider jets with high track multiplicities;
- We need high precision tracking in order to be able to identify the tracks and reconstruct the secondary vertex;

How to estimate the rate of false positives (a.k.a. mis-tags)?

- Limited detector resolution and poor track reconstruction can lead on an incorrect b-tagging of a jet;
- We use secondary vertices with a negative impact parameter to estimate the rate of false positives;

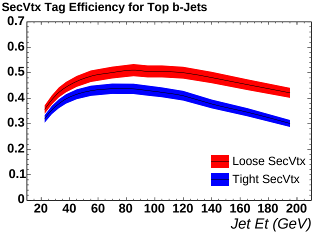
b-quark identification (a.k.a. *b*-tagging)

Two complementary algorithms to find *b*-jets

SECVTX This algorithm uses good quality tracks to **identify the secondary vertex**;

SECVTX

- The main *b*-tagging algorithm at CDF. We use it at the tight operating point;
- Very low fake rate ($< 0.5\%$) but a limited efficiency (40%);



b-quark identification (a.k.a. *b*-tagging)

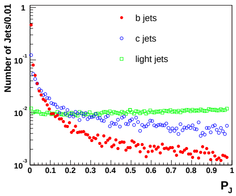
Two complementary algorithms to find *b*-jets

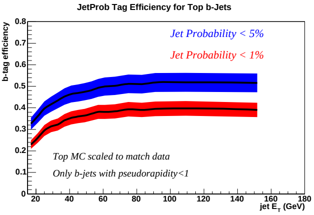
SecVtx — This algorithm uses good quality tracks to **identify the secondary vertex**;

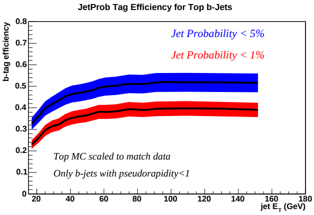
JetProb — This algorithms uses the impact parameters of the tracks to compute the **probability** that tracks within a jet are consistent with **coming from the primary vertex**;

JetProb

- Complementary algorithm: increasing acceptance when used in combination with SecVtx;
- This algorithms yields a higher fake rate (5%) for a larger efficiency (50%);

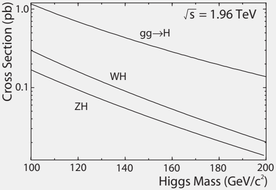


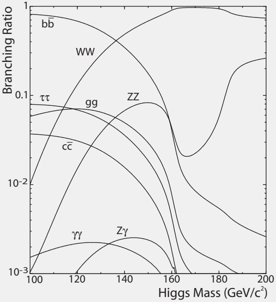


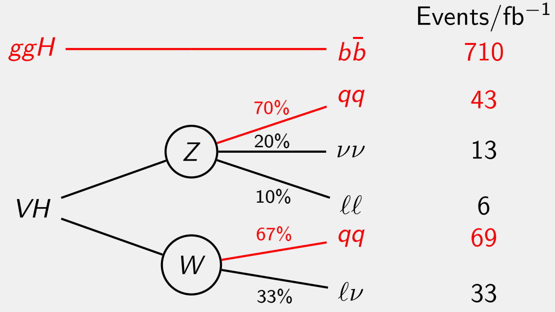


Now that we are familiar with the physics potential and our detector, we are ready to search for a Higgs signal!

Higgs production at the Tevatron



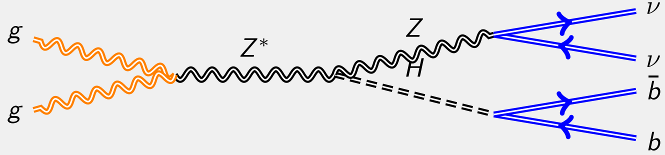




For $m_H < 135$ GeV/c^2 (low mass), $H \to b\bar{b}$:

- Hadronic final states are almost impossible to distinguish from the QCD background;
- Associated production provides a leverage through leptonic decays, but only for ~ 50 events per fb^{-1};

$ZH \to \nu\nu b\bar{b}$

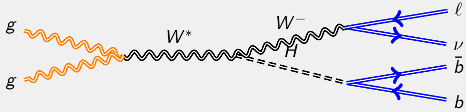


- Main motivation to study the $\not{E}_T$+b-jets signature in the SM sector;
- At low mass, about 13 events per fb^{-1} expected (before any selection);
- Striking signature: $\not{E}_T$ and to b-jets;

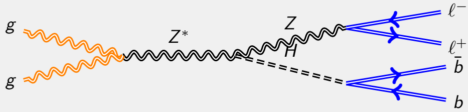
Event selection

- For the neutrinos: $\not{E}_T > 50$ GeV/c^2;
- For the Higgs: two high-p_T b-jets ($E_T^{j_{1(2)}} > 35(25)$ GeV/c^2) with $|\eta| < 2$;
- At least one central jet ($|\eta| < 0.9$), i.e. a jet with good tracker coverage;

Acceptance to leptonic decays



- CDF's lepton coverage reliably goes up to $|\eta| \sim 1$ (1.5) for electrons (muons);
- Tau ID is not very efficient;
- We can thus collect many leptonic W decays, about 50% of them in fact;
- About half of our events come from $Z \to \nu\nu$ and half from $W \to \ell\nu$;
- We also consider the much smaller contribution from $ZH \to \ell\ell b\bar{b}$;



More on the leptonic decays we accept

We collect e/μ events from:

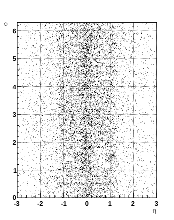
- Un-instrumented regions of the calorimeter;
- Gaps between the muon chambers;
- Outside the tracking coverage;

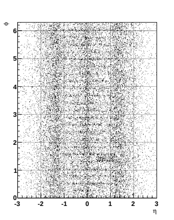
The τ we collect come from hadronic decays;

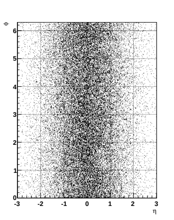
- Their distribution is unbiased;

Distribution of the W decays we collect: the detector structure is clearly visible

	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$
All events	19%	30%	51%
Two jet events	20%	33%	47%
Three jet events	17%	23%	60%

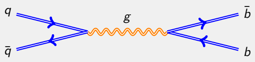






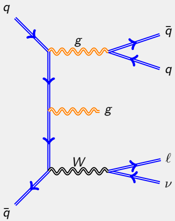
Other SM processes in the $\not{E}_T$+b-jets signature

QCD production



- By far largest background at a hadron collider;
- Only SM background **not involving neutrinos** ;
- It takes a lot of CPU to model QCD with Monte Carlo, and high order effects are difficult to include;

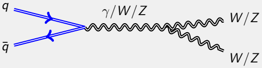
The production of a W/Z boson in association with jets

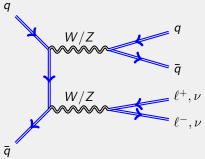


- The $\not{E}_T$ comes from the W/Z decays;
- Thus, it is very much **signal-like**, and hard to separate;
- These processes are **complex to model**, due to the variety of final states and non-negligible NLO effects;
- They enter our selection through two contributions:
  - W+$b\bar{b}$+jets events, which naturally belong to our final state;
  - W+$c\bar{c}$+jets events, which are **wrongly** ***b*-tagged**;
- Overall, the production rate of W/Z+jets is second to QCD;
- However, it is the **largest in our final analysis region**;
- In fact, the sensitivity of our analyses in this channel highly depends on our ability to properly model this background;

Other SM processes in the $\not{E}_T$+b-jets signature

Production of a pair of vector bosons (diboson): $WW/WZ/ZZ$





- Final state similar to W+jets, i.e. with a neutrino;
- Observation in leptonic signature in 2007 [1.1fb^{-1}];
- Observation in semileptonic signature in 2009 [2.7fb^{-1}];
- Observation in $\not{E}_T$+jets (without b-ID) in 2009 [3.5fb^{-1}];
- Evidence in $\not{E}_T$+b-jets within reach at the Tevatron;
- These processes are a challenging **benchmark on the road to the SM Higgs boson**;
- They allow to **test the sensitivity of our tools**, both in the hadronic and leptonic decay modes;

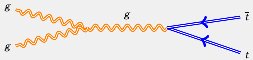
This background is 40 to 140 times larger than our signal.

Other SM processes in the $\not{E}_T+b$-jets signature

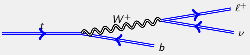
Top Pair production

Observed by CDF & DZero in 1995.

The top quark production rate is about 75 times larger than that of the SM Higgs.



Top quark decay

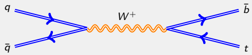


The top quark decays to *Wb*, just like our signal, which then produces a neutrino.

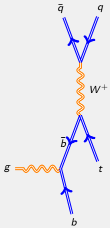
Single top production

Observed by CDF & DZero in 2009.

The production rate of single top is about 20 times that of the SM Higgs.



[s-channel]



[t-channel]

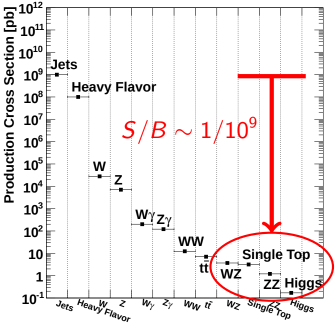
At $m_t = 175\ \ \text{GeV}/c^2$:

- $\sigma^t_{NLO} = 1.98 \pm 0.25$ pb;
- $\sigma^s_{NLO} = 0.88 \pm 0.11$ pb;

An quick review of the $\not{E}_T+b$-jets signature

Many SM processes populate the $\not{E}_T+b$-jets final state

- We require appropriate models for the many backgrounds;
- Many processes share the same decay paths, **limiting our options to discriminate them**;
- Moreover, there are **several orders of magnitude between signal and background**;
- These considerations make this signature **very challenging**;



Very different sources of background

- **Involving neutrinos or missed leptons ($\not{E}_T$):**
  - All the electroweak backgrounds;
- **Involving "detector" effects (e.g. mis-measurement):**
  - Multi-jet QCD production;
- **Involving our identification algorithms:**
  - W/Z+jets with wrong b-jet ID;

Not just the SM Higgs

- There are other interesting SM processes besides the Higgs;
- Single top and diboson production are two obvious examples;

Prerequisites for searching for small signals

Today's signal is tomorrow's background.

- More than 16 years ago, top pair production was the signal sought after;
- Today, we are searching for signals a few times smaller, up to 100 times;
- At this scale, the **uncertainty** on the background could **hide the signal**;
- This has two important consequences:
  - Our **background modeling must be accurate**, and we need to accurately measure our uncertainties, and avoid over-estimating them;
  - A cut-based analysis using a counting experiment requires **large statistics**. Until then, **multivariate analysis** is our only hope for sensitivity;

Of course, we can only probe a given theory with an appropriate description of its predictions, i.e. a proper model for the signal.

Overview of the modeling

We use several techniques to model our signal and our backgrounds, assessing the needs of the analysis and what simulations are available.

Process	Simulation Method	Details
QCD Multi-jet (MJ)	Data-driven	
W/Z+jets	Monte Carlo	ALPGEN (GEN) +PYTHIA (PS)
W/Z mis-tags	Combined	Mis-tag rate applied to Monte Carlo
Top Pair	Monte Carlo	PYTHIA (GEN+PS)
Diboson	Monte Carlo	PYTHIA (GEN+PS)
Single Top	Monte Carlo	MADEVENT (GEN) + PYTHIA (PS)
SM Higgs	Monte Carlo	PYTHIA (GEN+PS)

- GEN: Event Generation ; PS: Parton Shower;
- W/Z mis-tags are light flavor jets (i.e. non-b) falsely tagged as b-jets;

Data-driven models

How we model the multi-jet background

- A pre-tag multi-jet sample is composed of four components:
  - (A) QCD heavy flavor (HF);
  - (B) QCD mis-tags, from light flavor (LF);
  - (C) EWK (non-QCD) HF;
  - and (D) EWK mis-tags from LF;
- We use a **Tag-Rate-Matrix** to describe the QCD contribution (A+B) and a **Mis-Tag** matrix to describe the LF mis-tags (D);

	QCD	EWK
HF	(A)	(C)
LF	(B)	(D)

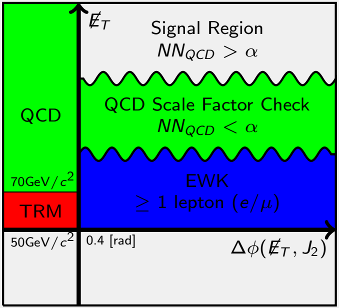
Data-driven model for multi-jet (MJ) production

- **Why?** Efficiency is so low that we would need a very large QCD Monte Carlo sample;
- **Data-driven method**: deriving a (4D) **Tag-Rate-Matrix** from QCD MJ sample ($> 99.9\%$);
  - Using the transverse momentum imbalance ($\not{p}_T$), the scalar sum of the jets in the event (H_T) and the ratio between the $(\sum p_T)/p_T^{j_i}$. where p_T is the sum of the good tracks of jet i;
- Applying the matrix to the (pre-tag) data to get b-tagging probability for each event;
  - We apply the matrix to the Monte Carlo and subtract to avoid double counting;
- This technique provides **excellent shape agreement**;
- The appropriate **normalization** is obtained using a **scale factor** derived in a control region;

Making sure we do it right

We carefully check our background modeling in several **control regions**. Only when we are certain that everything is under control and shows good agreement with data do we proceed to the next step and **unblind the analysis**.

Control regions



QCD-rich regions for QCD MJ model & cross-check:

- **TRM**: sample used to derive the Tag-Rate-Matrix;
- **QCD**: cross-check for the data-driven model;

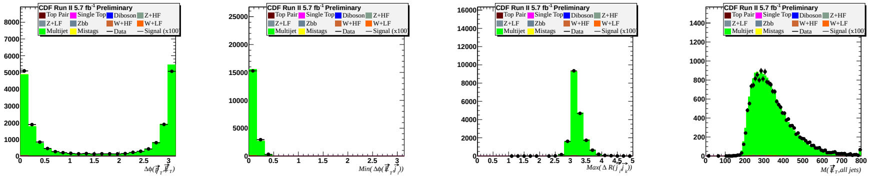
Cross-checks for other backgrounds:

- **EWK**: cross-check for the EWK backgrounds (MC);
- **QCD Scale Factor Check**: derivation of the QCD MJ scale factor (~ 1);
- New: extra regions with high *NN* output in pre-tag;

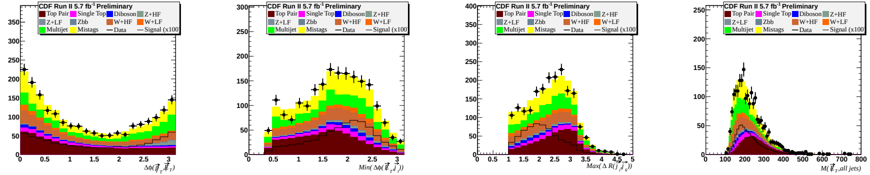
Note: The value of α depends on the NN_{QCD} and is chosen to retain about 90% of the signal.

A few plots from our control regions

From a QCD-rich region (no leptons) where we check our data-driven model



From a region with an **identified lepton** where we check the EWK modeling

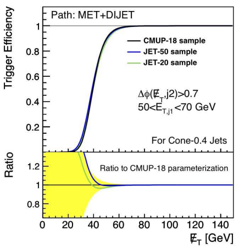


Excellent agreement throughout our control regions.

Searching for rare SM processes in $\not\!\!E_T$+b-jets

	Data	[%]	Higgs	[%]	Single Top	[%]	Diboson	[%]
Good Runs	1885312	100.0	94.3	100.0	1126.1	100.0	1958.9	100.0
Trigger driven selections	**1529702**	**81.1**	**76.2**	**80.8**	**1055.6**	**93.7**	**1414.4**	**72.2**

The $\not\!\!E_T$ triggers



- CDF uses a three level trigger:
  - L1: $\not\!\!E_T > 28$ GeV/c^2 and one 10 GeV/c^2 calorimeter tower;
  - L2: $\not\!\!E_T > 28$ GeV/c^2 and two jets ($E_T > 3$ GeV/c^2);
  - L3: $\not\!\!E_T > 30$ GeV/c^2;
- We use a similar (prescaled) trigger path for early run II data;
- We parametrize the trigger turn-on considering:
  - The $\not\!\!E_T$ of the event, and the E_T and η of the two leading jets;
  - The separation between the two leading jets: $\Delta R(j_1, j_2)$;
  - Three different samples: to extract the turn-on and an uncertainty;
- Our **trigger-driven cuts** select events from **fully efficient** regions;
- Future plans include **relaxing** these cuts to cover less efficient regions;

Searching for rare SM processes in $\not\!\!E_T+b$-jets

	Data	[%]	Higgs	[%]	Single Top	[%]	Diboson	[%]
Good Runs	1885312	100.0	94.3	100.0	1126.1	100.0	1958.9	100.0
Trigger driven selections	1529702	81.1	76.2	80.8	1055.6	93.7	1414.4	72.2
Lepton veto	**1486862**	**78.9**	**53.4**	**56.6**	**689.8**	**61.3**	**1117.2**	**57.0**

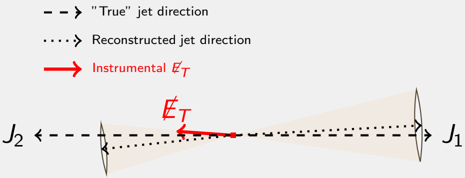
Lepton veto

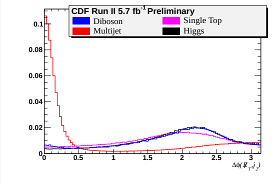
- We explicitly reject events with a lepton (e or μ);
- We also reject events with electrons mis-identified as jets;
- The main motivation is to remain orthogonal to other analyses at CDF;
- However, we use these events to **validate our Monte Carlo background model**;
- After this cut, we remain with 55-60% of our signal, keeping 79% of the main backgrounds;

Searching for rare SM processes in $\not\!\!E_T$+b-jets

	Data	[%]	Higgs	[%]	Single Top	[%]	Diboson	[%]
Good Runs	1885312	100.0	94.3	100.0	1126.1	100.0	1958.9	100.0
Trigger driven selections	1529702	81.1	76.2	80.8	1055.6	93.7	1414.4	72.2
Lepton veto	1486862	78.9	53.4	56.6	689.8	61.3	1117.2	57.0
$\not\!\!E_T$ not collinear to any jet	**191311**	**10.1**	**48.6**	**51.6**	**623.3**	**55.4**	**1026.5**	**52.4**

Removing obviously mis-measured QCD multi-jet events





- QCD dijet events that are **mis-measured** yield **instrumental $\not\!\!E_T$**;
- We reject most of them requiring $\Delta\phi(\not\!\!E_T, j_2) > 0.4$, with a small loss of signal acceptance;
  - We have similar requirements for the other jets;
- This cut alone **reduces the background by about one order of magnitude**;

Searching for rare SM processes in $\not\!\!E_T$+b-jets

	Data	[%]	Higgs	[%]	Single Top	[%]	Diboson	[%]
Good Runs	1885312	100.0	94.3	100.0	1126.1	100.0	1958.9	100.0
Trigger driven selections	1529702	81.1	76.2	80.8	1055.6	93.7	1414.4	72.2
Lepton veto	1486862	78.9	53.4	56.6	689.8	61.3	1117.2	57.0
$\not\!\!E_T$not collinear to any jet	191311	10.1	48.6	51.6	623.3	55.4	1026.5	52.4
SS	**493**	**0.02**	**6.3**	**6.7**	**40.4**	**3.6**	**15.3**	**0.8**
SJ	**898**	**0.04**	**4.9**	**5.2**	**33.1**	**2.9**	**13.2**	**0.7**
1S	**11056**	**0.58**	**14.4**	**15.2**	**213.3**	**18.9**	**64.9**	**3.3**

b-jet identification (a.k.a. *b*-tagging)

- Requiring jets originating form a *b*-quark **drastically reduces the backgrounds**;
- We combine two different tagging algorithms and form three **orthogonal** categories;
  - SecVtx+SecVtx (SS), SecVtx+JetProb (SJ), and exclusive SecVtx (1S);

Status after the kinematic cuts

- We have removed >2 orders of magnitude of background (data);
- The S/B has improved significantly;
- **But still low S/B:** $\sim 1/500$ **(Higgs),** $\sim 1/50$ **(Single Top),** $\sim 1/150$ **(Diboson)**;
- Need to further reject the (instrumental) background: using a dedicated **Neural Network**;

Further rejecting the QCD background

- Now we have pre-selected our events, it is time to remove the remaining backgrounds and increase our sensitivity to the signal;
- We have extensively searched for and studied variables that can **separate** the QCD multi-jet background from our signal;
- We've come up with several variables which provide a good separation power;
- However, we **cannot afford** to "cut" on these: we will **loose too much signal**;
- The only viable solution is to **combine the variables** using **multivariate analysis**. exploiting the different correlations among variables in the signal and the background;
- Because of the complexity of this scheme, we have to be careful and provide only variables that are **properly reproduced by our modeling**;

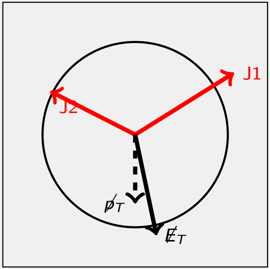
Multivariate analysis is a great tool that must be used with care!

Intrinsic $\not{E}_T$ vs. instrumental $\not{E}_T$

How we measure $\not{E}_T$

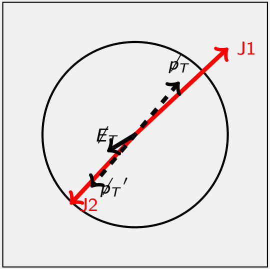
- Typically provided by the transverse energy imbalance ($\not{E}_T$) in the **calorimeter**;
- We also use the **transverse momentum flow imbalance** ($\not{p}_T$) from the **spectrometer**;
  - $\not{p}_T$ largely correlated with $\not{E}_T$ in presence of neutrinos (or $\tilde{\chi}^0$, etc.);
  - Very different for instrumental $\not{E}_T$: $\not{p}_T$ and $\not{E}_T$ <u>either correlated or anti-correlated</u>;

Exemple: $ZZ \to \nu\nu b\bar{b}$



[$\not{E}_T$ aligned to $\not{p}_T$]

Example: QCD $b\bar{b}$



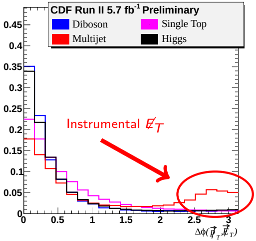
[$\not{p}_T$ is not aligned to $\not{E}_T$]

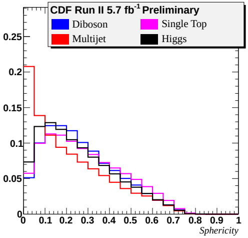
More inputs considered

Variable	
Magnitude of $\vec{\not{E}}_T$	Magnitude of $\vec{\not{p}}_T$
$\not{E}_T/\sqrt{\sum E_T}$	$\not{E}_T/H_T$ (scalar sum of the jet energies)
$\not{H}_T$ / $\not{E}_T$	$M(\vec{\not{E}}_T, \vec{j}_1, \vec{j}_2)$
$\Delta\phi$ between $\vec{\not{E}}_T$ and $\vec{\not{p}}_T$	Maximum of $\Delta\phi$ between any two jets
Maximum of ΔR between any two jets	Minimum of $\Delta\phi$ between the $\vec{\not{E}}_T$ and $\vec{j}_i$
Minimum of $\Delta\phi$ between the $\vec{\not{p}}_T$ and $\vec{j}_i$	$\Delta\phi(\vec{j}_1, \vec{j}_2)$ in the 2-jet rest frame
Sphericity	

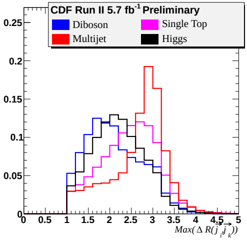
Many variables were carefully reviewed and checked. We only keep the most sensitive in separating multi-jet from the signal.

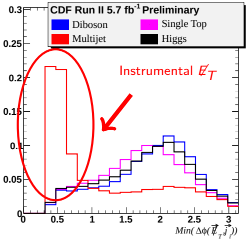
A Neural Network to reject QCD



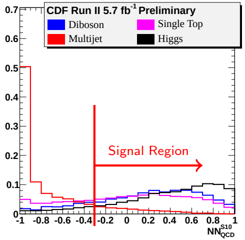


Adding more variables





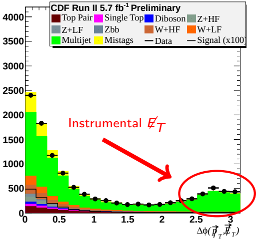
We combine novel variables to identify instrumental $\not{E}_T$ and distinguish it from "real" $\not{E}_T$.

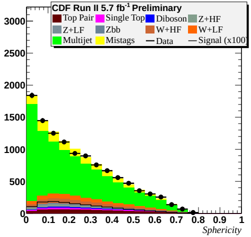


Performance

Signal acceptance	90-95%
Multi-jet rejection	$\sim$ 90%

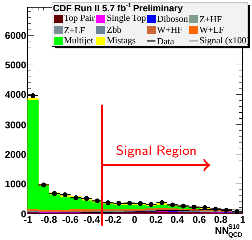
A Neural Network to reject QCD



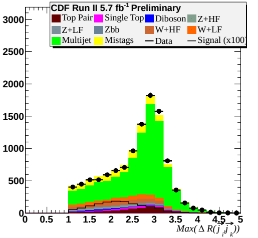


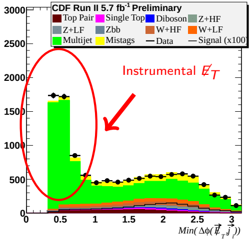
S/B

Single Top	1/20	(×2.5)
WZ/ZZ	1/50	(×3.0)
SM Higgs	1/200	(×2.5)



Every input variable is validated in several control regions.
The neural network output is also checked for mis-modeling.





Performance

Signal acceptance	90-95%
Multi-jet rejection	∼ 90%

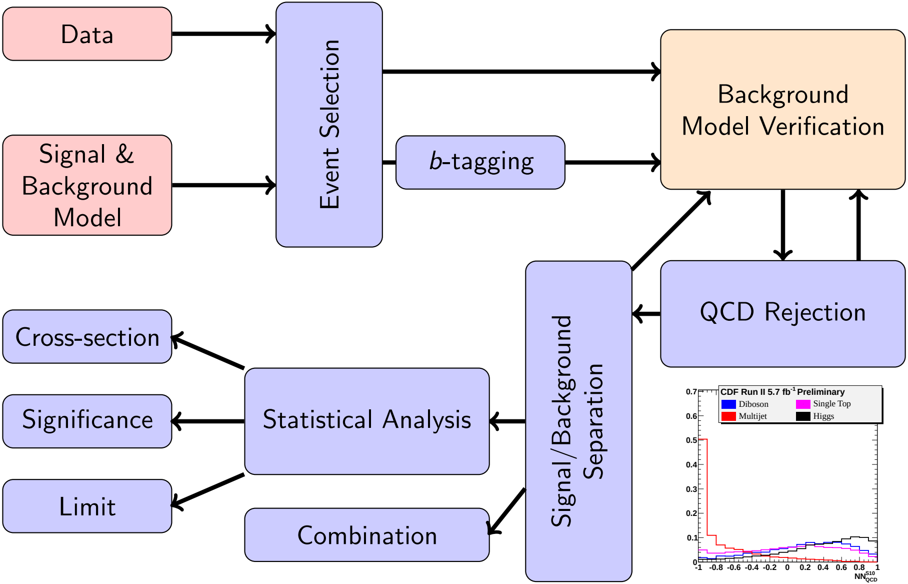
CDF analyses in the $\not\!\!E_T$+b-jets signature

- The QCD-rejecting NN is very powerful in **increasing the significance** of an analysis;
- It has the **same effect as a lepton ID**, but with much **larger acceptance**;

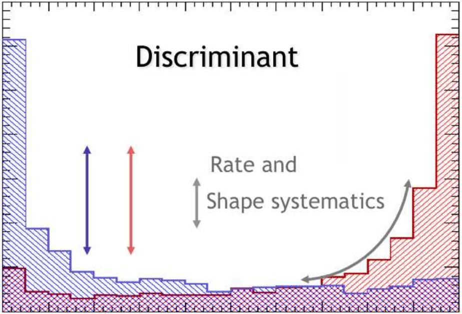
Analyses using this technique

- Three CDF analyses in this signature have benefited from this technique and would have otherwise be much less sensitive:
  - A single top cross-section measurement, which was part of the 2009 observation;
  - A top pair cross-section measurement, validating our tools with a well-known process;
  - A search for the SM Higgs boson, among the most sensitive;

The analysis scheme



Systematic Uncertainties



Systematic uncertainties can affect rate and template shape

- Rate systematics only affect normalization
- Shape systematics allow template to slide bin by bin

Sources of systematic uncertainties: Jet Energy, PDF, luminosity, cross-section, b-jet identification efficiency, Initial/Final State Radiation, background modeling

Bayesian method: integrating over likelihoods

- Systematics as nuisance parameters (truncated Gaussian)

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)

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

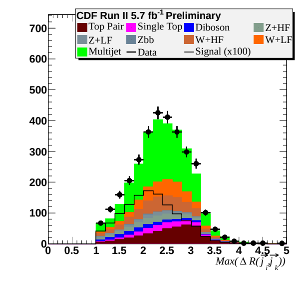
$s_{tot} = \Sigma_{i,j} s_{ij}$: Total number of signal prediction

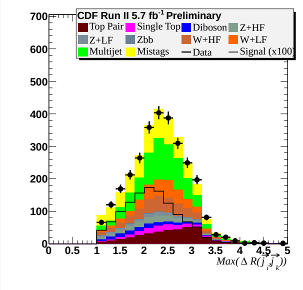
$\pi(x) = G(x|\hat{x},\sigma_x)$ $(x = s, b)$ $\hat{x}$: expected mean, σ_x: total uncertainty

A preliminary remark before we proceed with results

The analyses presented here contain an old version of our data-driven background model.

- The multi-jet model in the past analyses jointly modeled QCD + EWK mis-tags;
- We are now using a new modeling in which the EWK mis-tags are modeled separately;





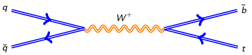
We are currently in the process of updating these analyses with the latest improvements (including relaxing the kinematic cuts to increase acceptance).

Single top cross-section measurement

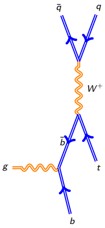
Why measure the single top production ?

What can we learn?

- Single top production gives us access to the W-t-b vertex, allowing to probe the V-A structure and the top quark spin;
- We also get to measure $|V_{tb}|^2$, allowing to
  - test the unitarity of the CKM matrix;
  - determine whether the CKM is 3×3, or if a 4th generation is possible;



[s-channel: $\sigma_{NLO} = 0.88 \pm 0.11$]

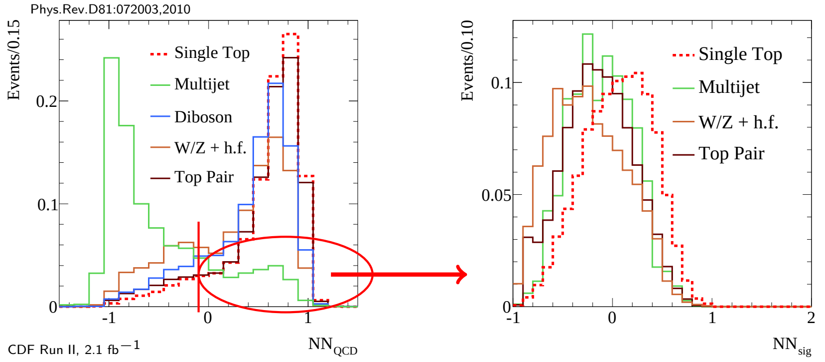


[t-channel: $\sigma_{NLO} = 1.98 \pm 0.25$]

$$\sigma_{ST} \propto |V_{tb}|^2$$

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} & ? \\ V_{cd} & V_{cs} & V_{cb} & ? \\ V_{td} & V_{ts} & V_{tb} & ? \\ ? & ? & ? & ? \end{pmatrix}$$

$\not\!\!E_T+b$-jets: A challenging final state



- The final state does not allow to reconstruct the top quark;
- This makes it hard to discriminate the single top signal from the background;

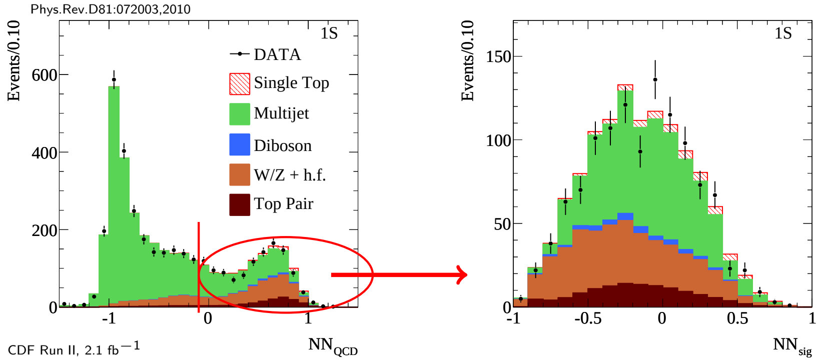
This channel is totally independent to the others and increases the signal acceptance by 33%.

Event yields after QCD rejection

Process	1b-tag (1S)	2 b-tags(2S)	2 b-tags(SJ)
s-channel	15.7±2.0	7.6±0.9	6.3±0.8
t-channel	31.2±4.9	1.7±0.2	1.6±0.2
$t\bar{t}$	125±23	30.3±5.8	29.2±5.7
$WW/WZ/ZZ$	33.0±6.5	4.9±0.6	4.2±0.6
W + h.f.	269±113	12.7±7.5	22.7±13.7
Z + h.f	105±53	11.8±5.8	11.8±6.0
Multijet	592±27	28.9±3.8	58.5±5.8
Total	1172±169	98±15	134±21
Observed	1167	113	131

- The background is mainly composed of multi-jet (mostly mis-tags);

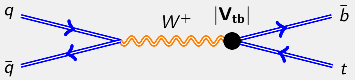
Cross-section measurement in the $\not{E}_T$+b-jets signature



$$\sigma_{s+t} = 4.9^{+2.5}_{-2.2} \text{ pb}$$

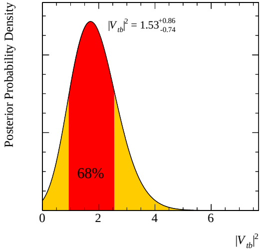
The result is compatible with the SM within the (large) uncertainties.

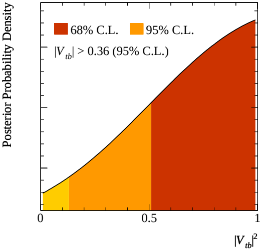
Measurement of V_{tb}



$$\left|V_{tb}^{\text{meas.}}\right|^2 = \frac{\sigma_{meas.}}{\sigma_{SM}} \left|V_{tb}^{SM}\right|^2$$

- Assuming $|V_{td}|^2 + |V_{ts}|^2 << |V_{tb}|^2$;
- No assumption of the number of quark families;
- No assumption of unitarity for the CKM matrix;

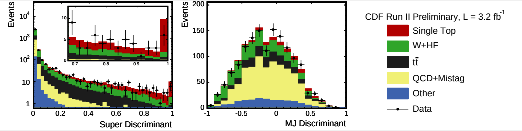




Not a big constraint but consistent with the Standard Model.

A necessary combination with other analyses

CDF Combination (5σ observed p-value, $> 5.9\sigma$ expected)



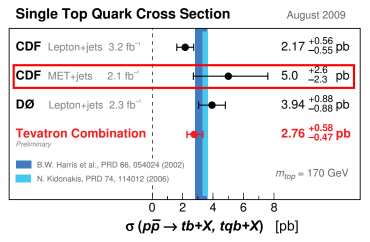
Phys.Rev.Lett.103:092002,2009

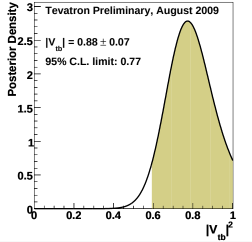
$$\sigma_{s+t} = 2.3^{+0.6}_{-0.5}\ \text{pb}$$

$$|V_{tb}| = 0.91 \pm 0.11(\text{exp.}) \pm 0.07(\text{th.})$$

$$|V_{tb}| > 0.71 \text{ at } 95\% \text{ C.L.}$$

Tevatron Combination: observation 14 years after the top quark!

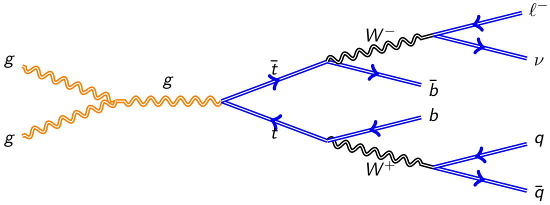




Top pair cross-section measurement

Top pair production as a background to new physics

- Top pair production is one important background in this signature;
- We want to cross-check our tools using a well-known process;
- This first measurement is totally independent from others;
  - Top pair measurements have ≥ 4 jets to reduce the backgrounds;
  - Can be combined to improve the world average;



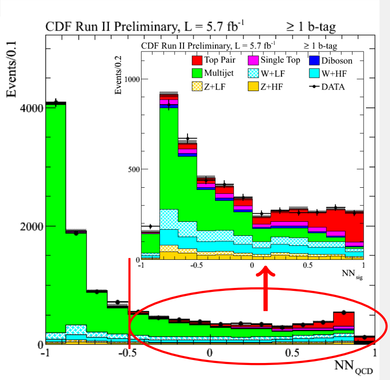
Event yields in the signal region

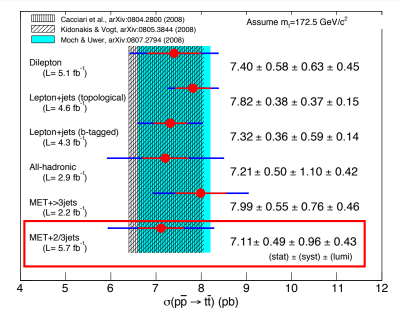
CDF Run II Preliminary $L = 5.7fb^{-1}$									
	1S			SS			SJ		
Single Top	195.7	±	33.7	39.0	±	7.5	30.4	±	6.0
Diboson	95.7	±	12.7	13.8	±	2.4	11.5	±	2.0
W+LF	494.7	±	152.2	5.1	±	1.9	17.9	±	6.1
W+HF	405.0	±	124.5	41.7	±	14.1	44.6	±	14.6
Z+LF	155.9	±	47.9	2.6	±	0.9	6.1	±	2.0
Z+HF	184.2	±	56.6	27.6	±	9.1	25.1	±	8.2
Multijet	1732.0	±	103.9	93.6	±	15.3	145.1	±	18.7
Top Pair	503.4	±	33.4	117.5	±	11.8	102.9	±	11.0
Tot. Exp.	3766.6	±	239.5	340.9	±	26.8	383.6	±	28.9
Data	3814			290			401		

Table: Acceptance table in Signal Region.

Top pair cross-section measurement

A cross-check using well understood signal.





More information

http://www-cdf.fnal.gov/physics/new/top/2010/xsection/METjets/

Search for the SM Higgs boson

Search for the SM Higgs boson in the $\not{E}_T$+b-jets signature

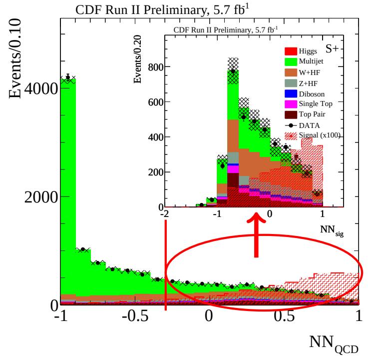
- Most of the tools presented today were initially designed for the SM Higgs analysis, which is the "main" analysis in this signature;
- After rejecting 90% of the QCD (and $\sim$ 60% of the overall background) while keeping 90-95% of our signal, we have the following yields:

CDF Run II Preliminary, 5.7 fb^{-1}

Process		Excl. ST	ST+ST	ST+JP
Top Pair		381.1 $\pm$ 46.4	89.7 $\pm$ 13.0	76.7 $\pm$ 13.0
Single Top		136.4 $\pm$ 24.5	32.3 $\pm$ 6.3	24.6 $\pm$ 5.3
Diboson		106.5 $\pm$ 17.3	14.2 $\pm$ 2.6	12.4 $\pm$ 2.5
Z+HF		399.7 $\pm$ 129.6	32.7 $\pm$ 11.1	35.0 $\pm$ 12.3
W+HF		1065.2 $\pm$ 356.0	49.7 $\pm$ 17.4	68.5 $\pm$ 24.9
Multi-jet		1108.0 $\pm$ 113.8	43.5 $\pm$ 9.7	120.3 $\pm$ 13.3
Exp. Background		3196.9 $\pm$ 501.8	262.2 $\pm$ 33.5	337.7 $\pm$ 42.0
Observed		3220	237	301
$ZH \to llb\bar{b}$	$(m_H = 115$ GeV$)$	0.3 $\pm$ 0.02	0.2 $\pm$ 0.02	0.1 $\pm$ 0.02
$WH \to l\nu b\bar{b}$	$(m_H = 115$ GeV$)$	5.9 $\pm$ 0.3	2.8 $\pm$ 0.2	2.1 $\pm$ 0.3
$ZH \to \nu\nu bb$	$(m_H = 115$ GeV$)$	5.6 $\pm$ 0.2	2.8 $\pm$ 0.2	2.0 $\pm$ 0.2
Exp. Signal		11.8 $\pm$ 0.5	5.7 $\pm$ 0.5	4.3 $\pm$ 0.5

Search for the SM Higgs boson in the $\not\!\!E_T$+b-jets signature

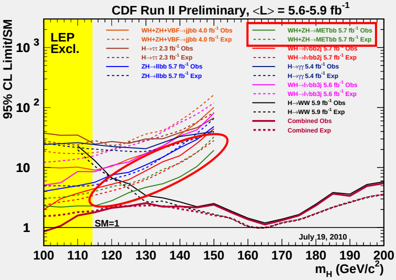
Reaching 3× the SM prediction at low mass

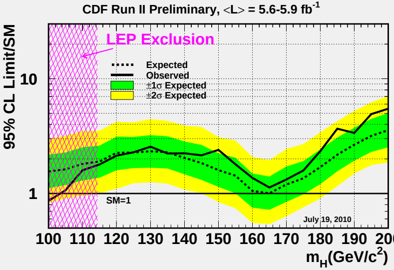


We re-train our final discriminant NN for each mass point.

Combine & Conquer (CDF Combination)

One of the most sensitive low mass ($< 135\ \ \text{GeV}/c^2$) channels

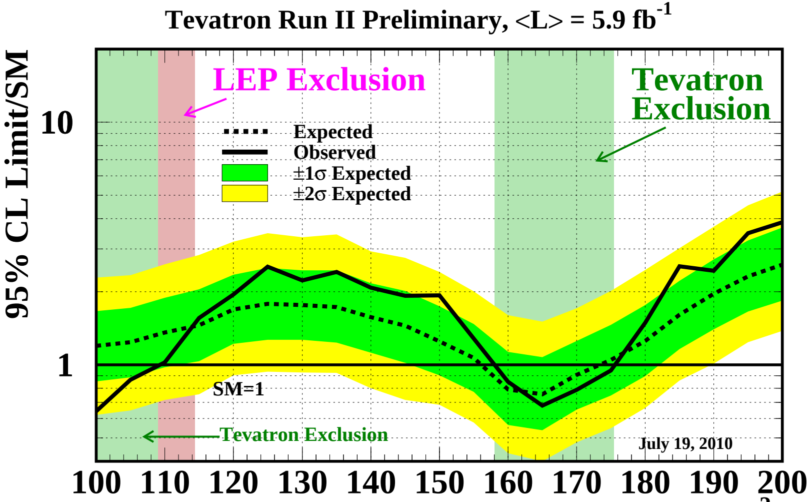




ICHEP2010

The **combined** observed (expected) limit is **1.9 (1.8) times the SM prediction** at $m_H = 115\ \ \text{GeV}/c^2$ and 1.0 (1.1) times the SM prediction at $m_H = 165\ \ \text{GeV}/c^2$.

Combine & Conquer (Tevatron Combination)



Future prospects

Currently going on in this signature (in the SM sector)

- Some new features (since last summer) were already presented today;
- In order to increase acceptance, we will have to relax our kinematic selections;
- The expected gain is about 30% for Higgs and single top, and up to 60% for diboson;
- We are carefully cross-checking our background modeling in this new territory (for us);
- Of course, we'll keep adding new data as it comes;

New challenge besides the SM Higgs

- We would like to measure the diboson production with b-jets;

Summary

$\not\!\!E_T$+b-jets: an interesting channel for SM and BSM physics

- Many SM and BSM yield $\not\!\!E_T$+b-jets;
- This channel has very large acceptance;
- Very sensitive **provided we get rid of the large QCD multi-jet background**;

A powerful tool to reject QCD multi-jet

- Novel combination of kinematic variables (exploits correlations);
- The technique is very **generic**: works with many different signals;
- It is **as powerful as a lepton ID**, in a channel with much larger acceptance (at CDF);
- Made three SM analyses possible at CDF (and similarly at DZero);
- We plan to use this technique to measure $\sigma(WZ/ZZ \to \not\!\!E_T + b\bar{b})$;

Thank You

Backup Slides

Technical aspects of the multi-jet (MJ) model

Combination procedure

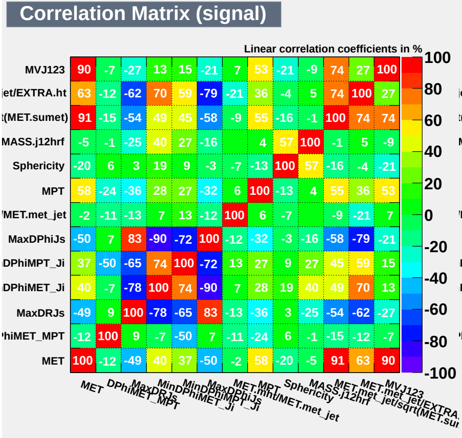
- The TRM predicts the probability for an event to be in either tagging category;
- We apply the TRM to HF EWK MC and subtract to avoid double counting (C);
- We apply the TRM to LF EWK MC and subtract (D);
- EWK mis-tags (D) are modeled using a mis-tag matrix;
- We apply the official mis-tag matrices to LF EWK MC;

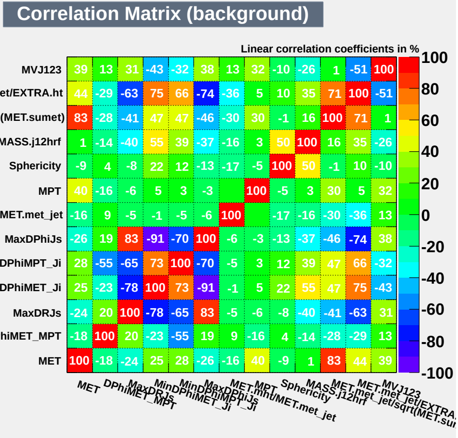
	QCD	EWK
HF	(A)	(C)
LF	(B)	(D)

Why a mis-tag matrix ?

- The *b*-tagging algorithms use information on the position of the secondary vertex w.r.t. the primary vertex;
- The limited resolution of the tracks cause light flavor jets to be wrongly tagged;
- We can estimate the mis-tag probability from the fact that the resolution effects yield equal probability for a light flavor jet to get a positive or negative tag;
- We apply the mis-tag matrices for the SecVtx and JetProb algorithms, and further include a factor to account for the asymmetry between negative and positive tags (effects from Λ's and K's);

Correlation between variables: NN_{QCD}





Effect of QCD rejection in numbers

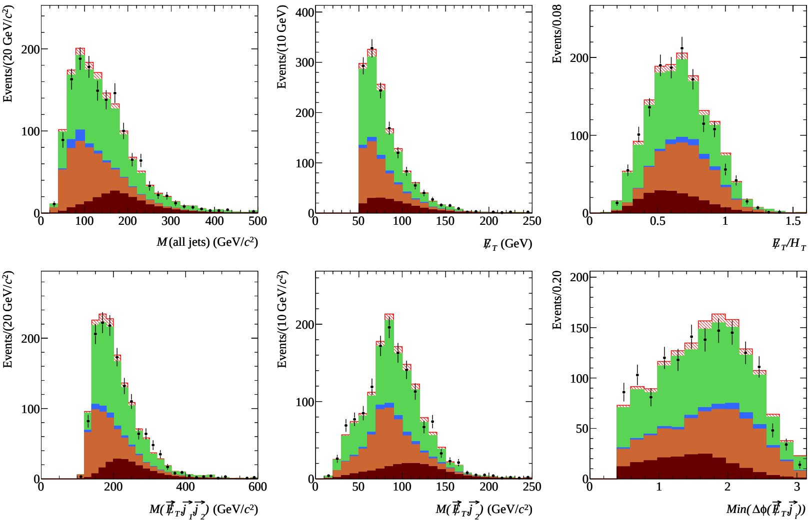
CDF Run II Preliminary $L = 5.7fb^{-1}$			
	1S	SS	SJ
Single Top	236.5 ± 40.7	41.8 ± 8.1	34.1 ± 6.7
Diboson	128.8 ± 17.1	15.3 ± 2.7	13.8 ± 2.4
W+LF	869.3 ± 267.3	6.5 ± 2.4	26.4 ± 8.8
W+HF	610.1 ± 187.5	50.1 ± 16.9	59.4 ± 19.5
Z+LF	241.6 ± 74.3	3.1 ± 1.1	8.2 ± 2.7
Z+HF	278.9 ± 85.7	33.2 ± 10.9	32.9 ± 10.7
Multijet	7819.6 ± 469.2	215.3 ± 35.1	586.8 ± 75.7
Top Pair	536.2 ± 35.6	120.7 ± 12.1	107.4 ± 11.5
Tot. Exp.	10721.0 ± 585.5	486.0 ± 43.2	869.0 ± 80.6
Data	10721	486	869

Table: Acceptance table in PreSelection.

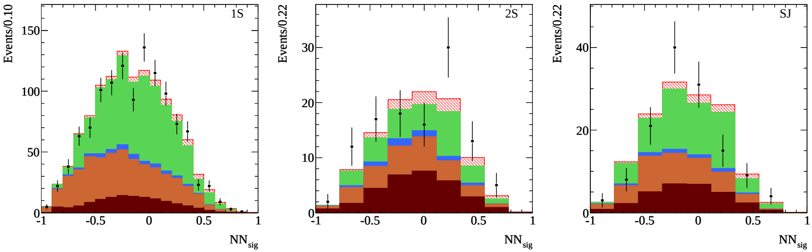
CDF Run II Preliminary $L = 5.7fb^{-1}$			
	1S	SS	SJ
Single Top	195.7 ± 33.7	39.0 ± 7.5	30.4 ± 6.0
Diboson	95.7 ± 12.7	13.8 ± 2.4	11.5 ± 2.0
W+LF	494.7 ± 152.2	5.1 ± 1.9	17.9 ± 6.1
W+HF	405.0 ± 124.5	41.7 ± 14.1	44.6 ± 14.6
Z+LF	155.9 ± 47.9	2.6 ± 0.9	6.1 ± 2.0
Z+HF	184.2 ± 56.6	27.6 ± 9.1	25.1 ± 8.2
Multijet	1732.0 ± 103.9	93.6 ± 15.3	145.1 ± 18.7
Top Pair	503.4 ± 33.4	117.5 ± 11.8	102.9 ± 11.0
Tot. Exp.	3766.6 ± 239.5	340.9 ± 26.8	383.6 ± 28.9
Data	3814	290	401

Table: Acceptance table in Signal Region.

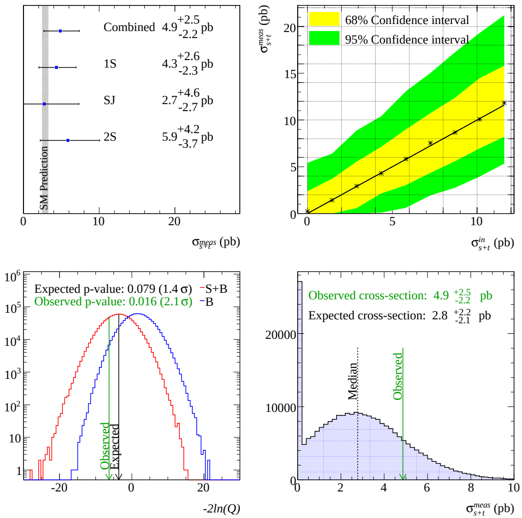
Single Top


Events/(20 GeV/c^2)
M(all jets) (GeV/c^2)
Events/(10 GeV)
$\not{E}_T$ (GeV)
Events/0.08
$\not{E}_T/H_T$
Events/(20 GeV/c^2)
$M(\vec{\not{E}}_T \vec{j}_1 \vec{j}_2)$ (GeV/c^2)
Events/(10 GeV/c^2)
$M(\vec{\not{E}}_T \vec{j}_2)$ (GeV/c^2)
Events/0.20
$Min(\Delta\phi(\vec{\not{E}}_T \vec{j}_i))$

Single Top


Events/0.10
1S
150
100
50
0
-1
-0.5
0
0.5
1
NN sig
Events/0.22
2S
30
20
10
0
-1
-0.5
0
0.5
1
NN sig
Events/0.22
SJ
40
20
0
-1
-0.5
0
0.5
1
NN sig

Single Top

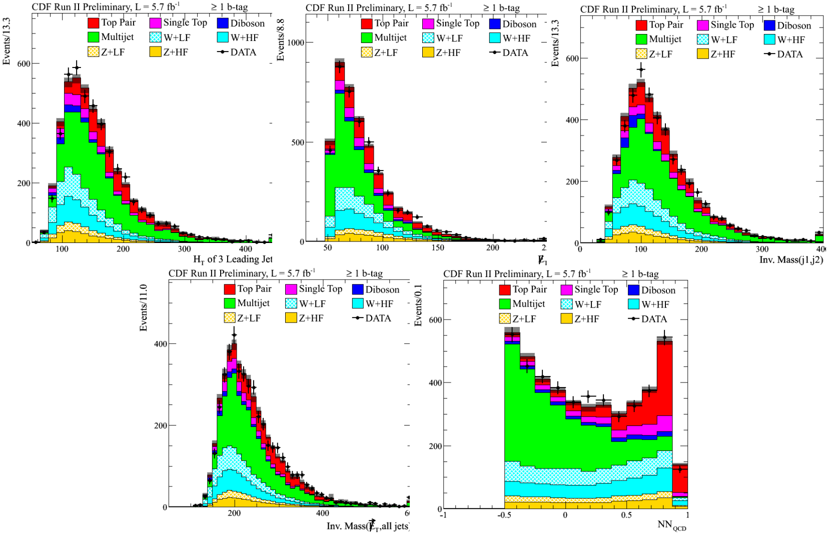


Combined 4.9 +2.5 -2.2 pb
1S 4.3 +2.6 -2.3 pb
SJ 2.7 +4.6 -2.7 pb
2S 5.9 +4.2 -3.7 pb
SM Prediction
68% Confidence interval
95% Confidence interval
Expected p-value: 0.079 (1.4 σ)
Observed p-value: 0.016 (2.1 σ)
S+B
B
Observed
Expected
-2ln(Q)
Observed cross-section: 4.9 +2.5 -2.2 pb
Expected cross-section: 2.8 +2.2 -2.1 pb
Median
Observed

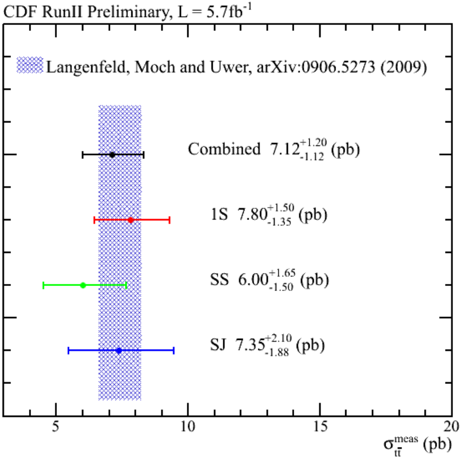
Systematic uncertainties

Systematic	Rate	Shape	Comment
Luminosity	6%	-	Not for QCD multijet
PDF	+0.7% / -0.4% (s-chan)	X	
	+2.3% / -2.1% (t-chan)	X	
	+0.8% / -1.4% (top pair)	-	
	2% (MC backgrounds)	-	
JES	-13.9% . . . 23%	X	
QCD multijet normalization	4.5% (ST) 13.1% (ST + ST) 10.0% (ST + JP)	- - -	Only QCD multijet
B tagging	4.3% (ST) 8.6% (ST + ST) 12.3% (ST + JP)	- - -	
W & Z + h.f. cross-section	40%	-	
Di-bonson cross-section	11.5%	-	
Top pair cross-section	±12.4%	-	
Trigger Efficiency	±0% . . . ± 2.6%	X	Not for QCD multijet
Lepton Veto	2%	-	
ISR/FSR	-4.5% . . . 16.5%	X	Only top-processes
Top mass dependence	-16.4% . . . 7.5%	X	Top mass only for V_{tb} and p-value
TRF	-	X	Only QCD multijet
Signal contamination	-	X	Only QCD multijet
Background scaling	2%	-	
Signal cross-section	± 12.4% (s-chan) ± 12.6% (t-chan)	- -	Only for p-value and V_{tb}

Top Pair: NN_{SIG} and its inputs

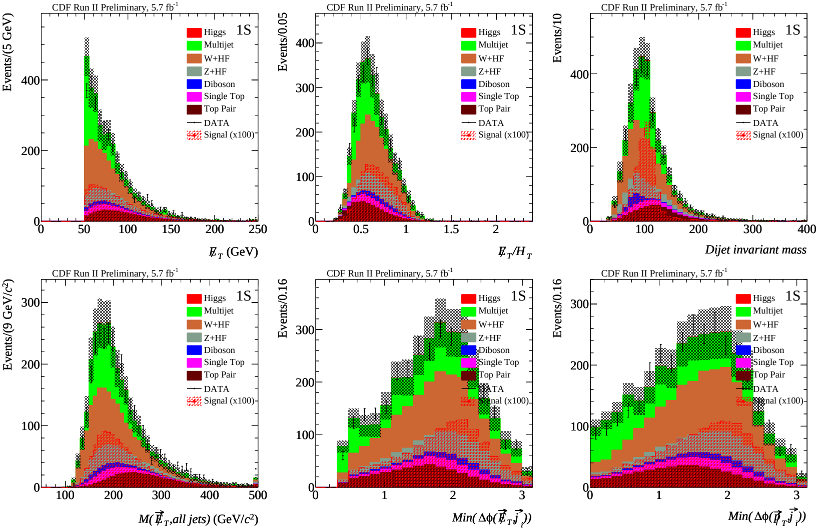


Top Pair



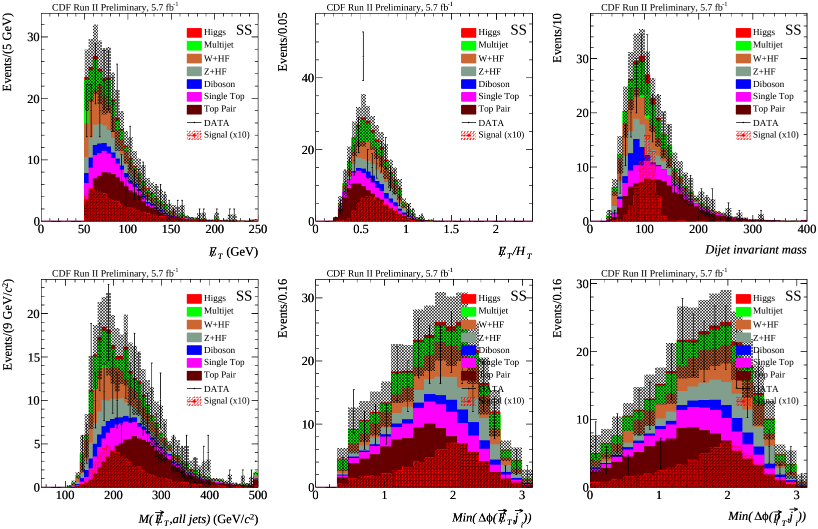
CDF RunII Preliminary, L = 5.7fb-1
Langenfeld, Moch and Uwer, arXiv:0906.5273 (2009)
Combined 7.12 +1.20 -1.12 (pb)
1S 7.80 +1.50 -1.35 (pb)
SS 6.00 +1.65 -1.50 (pb)
SJ 7.35 +2.10 -1.88 (pb)
5
10
15
20
σ meas tt̄ (pb)

SM Higgs

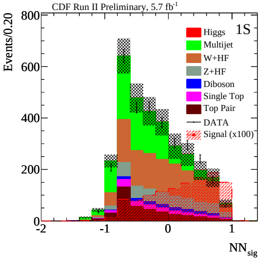


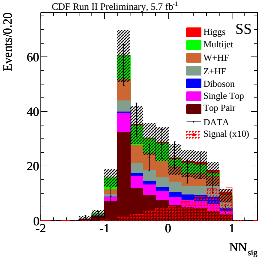
CDF Run II Preliminary, 5.7 fb^{-1}
1S
Higgs
Multijet
W+HF
Z+HF
Diboson
Single Top
Top Pair
DATA
Signal (x100)
Events/(5 GeV)
$\not\!\!E_T$ (GeV)
Events/0.05
$\not\!\!E_T/H_T$
Events/10
Dijet invariant mass
Events/(9 GeV/c^2)
$M(\vec{\not\!\!E}_T$,all jets) (GeV/c^2)
Events/0.16
Min($\Delta\phi(\vec{\not\!\!E}_T,\vec{j}_i)$)
Min($\Delta\phi(\vec{\not\!p}_T,\vec{j}_i)$)

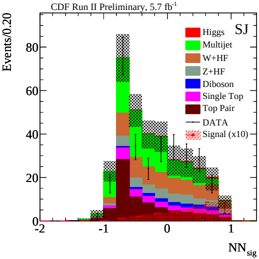
SM Higgs


CDF Run II Preliminary, 5.7 fb^{-1}
SS
Higgs
Multijet
W+HF
Z+HF
Diboson
Single Top
Top Pair
DATA
Signal (x10)
Events/(5 GeV)
$\not{E}_T$ (GeV)
Events/0.05
$\not{E}_T/H_T$
Events/10
Dijet invariant mass
Events/(9 GeV/c^2)
$M(\vec{\not{E}}_T$,all jets) (GeV/c^2)
Events/0.16
$Min(\Delta\phi(\vec{\not{E}}_T,\vec{j}_i))$
Events/0.16
$Min(\Delta\phi(\vec{\not{p}}_T,\vec{j}_i))$

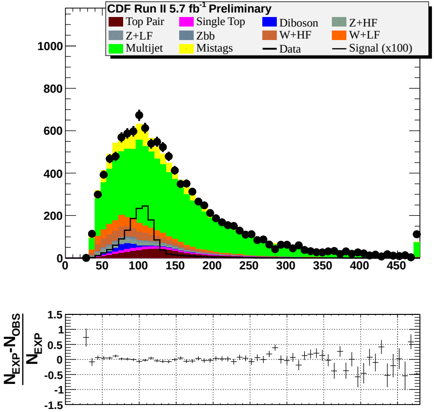
SM Higgs

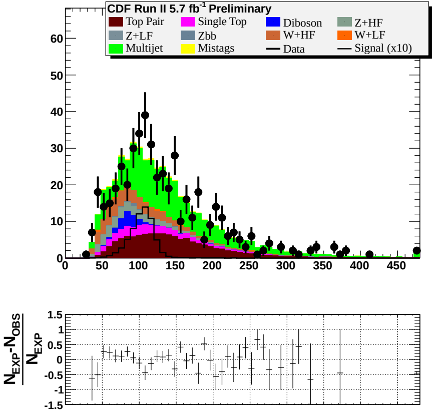

CDF Run II Preliminary, 5.7 fb-1
1S
Events/0.20
NNsig
Higgs
Multijet
W+HF
Z+HF
Diboson
Single Top
Top Pair
DATA
Signal (x100)
800
600
400
200
0
-2
-1
0
1


CDF Run II Preliminary, 5.7 fb-1
SS
Events/0.20
NNsig
Higgs
Multijet
W+HF
Z+HF
Diboson
Single Top
Top Pair
DATA
Signal (x10)
60
40
20
0
-2
-1
0
1


CDF Run II Preliminary, 5.7 fb-1
SJ
Events/0.20
NNsig
Higgs
Multijet
W+HF
Z+HF
Diboson
Single Top
Top Pair
DATA
Signal (x10)
80
60
40
20
0
-2
-1
0
1

Dijet mass distribution





Cross-sections and Branching Fractions

m_H (GeV/c^2)	$\sigma_{gg\to H}$ (fb)	σ_{WH} (fb)	σ_{ZH} (fb)	σ_{VBF} (fb)	$B(H \to b\bar{b})$ (%)	$B(H \to \tau^+\tau^-)$ (%)	$B(H \to W^+W^-)$ (%)
100	1861	286.1	166.7	99.5	81.21	7.924	1.009
105	1618	244.6	144.0	93.3	79.57	7.838	2.216
110	1413	209.2	124.3	87.1	77.02	7.656	4.411
115	1240	178.8	107.4	79.07	73.22	7.340	7.974
120	1093	152.9	92.7	71.65	67.89	6.861	13.20
125	967	132.4	81.1	67.37	60.97	6.210	20.18
130	858	114.7	70.9	62.5	52.71	5.408	28.69
135	764	99.3	62.0	57.65	43.62	4.507	38.28
140	682	86.0	54.2	52.59	34.36	3.574	48.33
145	611	75.3	48.0	49.15	25.56	2.676	58.33
150	548	66.0	42.5	45.67	17.57	1.851	68.17
155	492	57.8	37.6	42.19	10.49	1.112	78.23
160	439	50.7	33.3	38.59	4.00	0.426	90.11
165	389	44.4	29.5	36.09	1.265	0.136	96.10
170	349	38.9	26.1	33.58	0.846	0.091	96.53
175	314	34.6	23.3	31.11	0.663	0.072	95.94
180	283	30.7	20.8	28.57	0.541	0.059	93.45
185	255	27.3	18.6	26.81	0.420	0.046	83.79
190	231	24.3	16.6	24.88	0.342	0.038	77.61
195	210	21.7	15.0	23	0.295	0.033	74.95
200	192	19.3	13.5	21.19	0.260	0.029	73.47

References for the analyses

Websites

- Higgs: http://www-cdf.fnal.gov/physics/new/hdg/Results_files/results/vhmetbb_100705/;
- Single top: http://www-cdf.fnal.gov/physics/new/top/2008/singletop/METbb/;
- Top pair: http://www-cdf.fnal.gov/physics/new/top/2010/xsection/METjets/;

Publications

- T. Aaltonen *et al.* [The CDF Collaboration], "Search for single top quark production in pbar p collisions at $\sqrt{s} = 1.96$ TeV in the missing transverse energy plus jets topology", Phys. Rev. D **81**, 072003 (2010), arXiv:1001.4577 [hep-ex].
- T. Aaltonen *et al.* [The CDF Collaboration], "A Search for the Higgs Boson Using Neural Networks in Events with Missing Energy and b-quark Jets in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV", Phys. Rev. Lett. **104**, 141801 (2010), arXiv:0911.3935 [hep-ex].
- T. Aaltonen *et al.* [The CDF Collaboration], "First Observation of Electroweak Single Top Quark Production", Phys. Rev. Lett. **103**, 092002 (2009), arXiv:0903.0885 [hep-ex]. [100+ citations]