Search for a low mass SM Higgs boson in the di-tau decay channel at CDF

PHENOMENOLOGY 2010 Symposium

Madison, Wisconsin

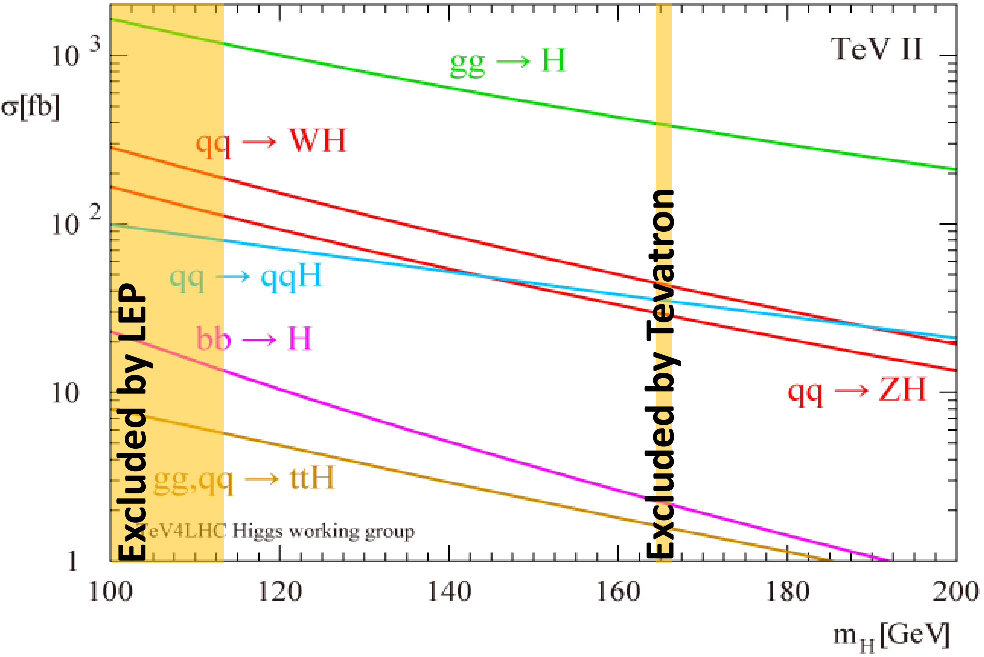
Pierluigi Totaro, INFN and University of Trieste

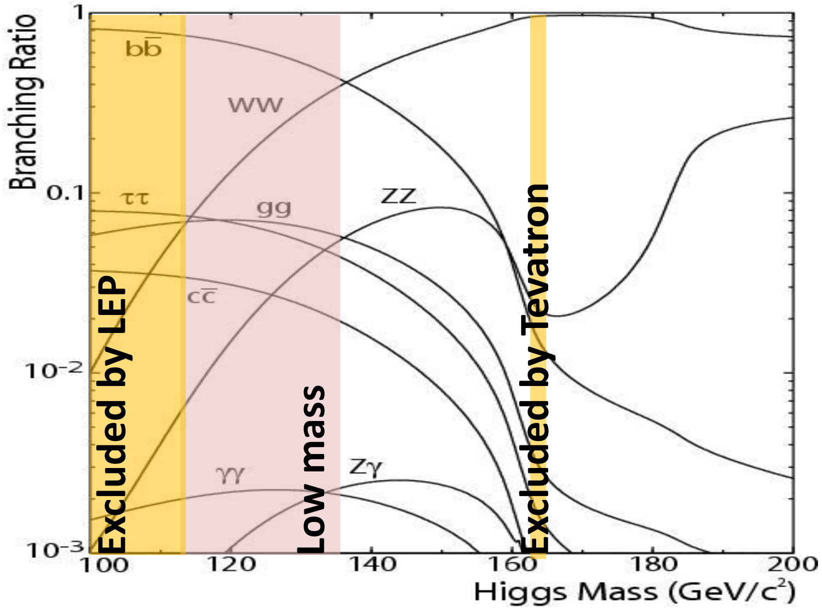
On behalf of the CDF collaboration

Outline

- Low mass Standard Model Higgs at Tevatron
- Motivation of the $H \rightarrow \tau\tau$ search
- Analysis strategy
  - Event selection
  - Background estimation
  - BDT multivariate technique
  - Results
- Prospects and summary

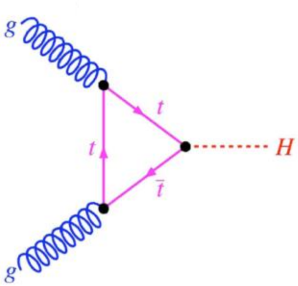
Higgs production and decay at Tevatron



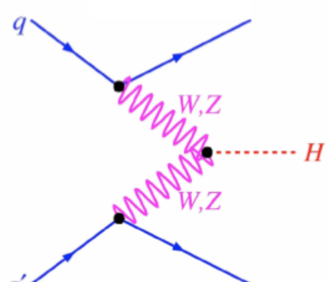


Primary production modes are:

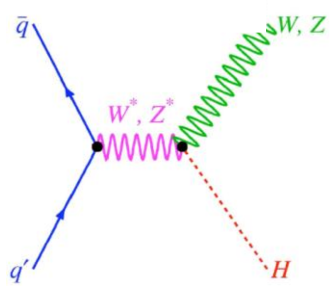
Low mass Higgs (M_H < 135 GeV/c^2) :
H→bb is the dominant decay channel



0.2 ~ 1.0 pb



0.02~ 0.1 pb

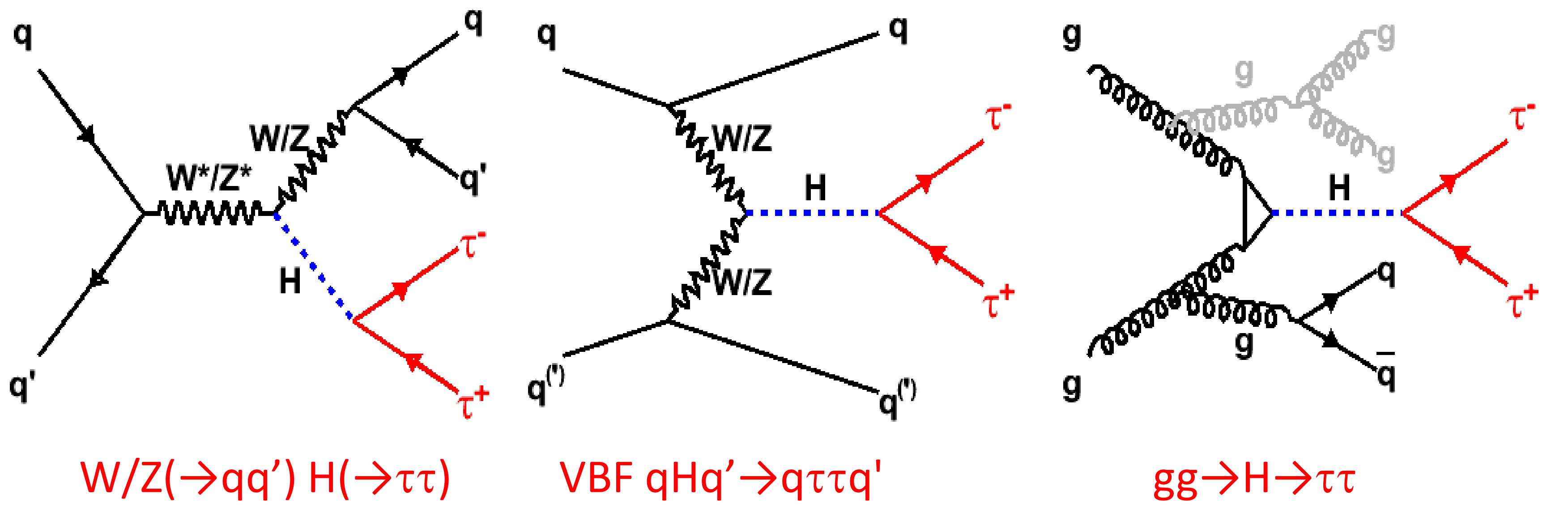


0.01~ 0.3 pb

- gg→H→bb is overwhelmed by QCD multijet background, thus search of associated production through a virtual W or Z boson is preferred
- H → ττ complementary channel

H→ττ search: motivation

- H→ττ complementary to H→bb signature
- small H→ττ B.R.(<10%) but 4 signal processes considered:

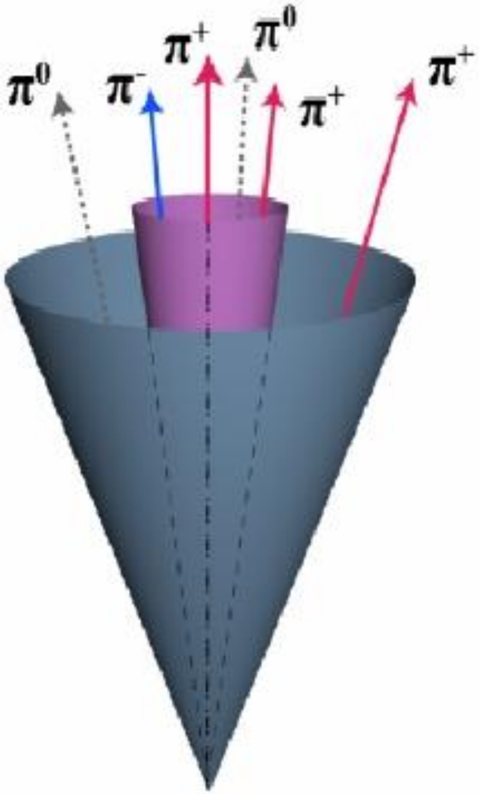


- acceptance increase by including the W/Z→jj final state in the ass. prod.
- direct production and VBF become accessible
- Total σ x B.R. comparable to other Higgs analyses

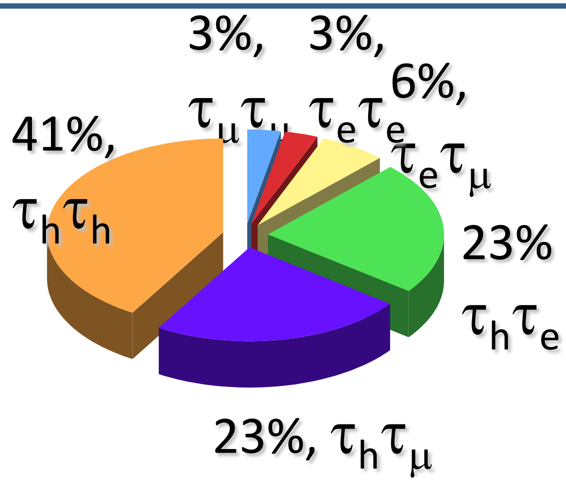
About tau leptons

- Heavy particles: 1.78 GeV/c^2
- Short lived: mean lifetime 291 ps ($c\tau$=87 μm)
- Decay modes:
  - $\tau \rightarrow \nu_\tau \nu_e e$ (B.R.~17%)
  - $\tau \rightarrow \nu_\tau \nu_\mu \mu$ (B.R.~17%)
  - $\tau \rightarrow \nu_\tau X_h$ (B.R.~65%) (X_h mainly $\pi^{\pm 0}$,small frac. of K)

- **Hadronic tau decays** appear in the detector as narrow jets with low tracks and neutral multiplicity
- **Hadronic tau ID at CDF relies on a two-cone algorithm**:
  - Signal cone around "seed" track, reconstruct P_{had}(p,E)
  - Isolation annulus for g/q→jet veto
  - In this analysis: standard **cut-based ID** is replaced by a **multivariate selection** based on a set of BOOSTED DECISION TREES trained to separate hadronic taus (MC) from QCD jets. An additional 20% of jet→τ fakes is rejected with respect to CDF standard ID.



Event selection



τ decay modes included in the analysis:

$\tau_h\tau_e+\tau_h\tau_\mu$

(46% B.R.)

- One central isolated lepton (e/μ) with $p_T > 10$ GeV/c
- One central hadronic tau with visible $p_T > 15$ GeV/c
- Opposite charged leptons
- At least one energetic calorimeter jet:
  - transverse energy: $E_T > 20$ GeV
  - EM fraction < 0.9
  - pseudorapidity: $|\eta| < 2.5$

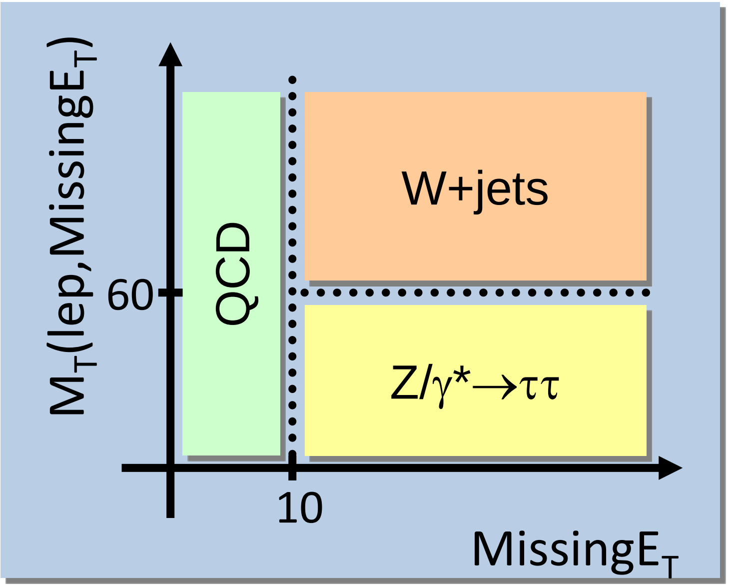
Background estimation

IRREDUCIBLE PHYSICS CONTRIBUTIONS

$Z \to \tau\tau$, top-antitop, diboson: from Monte Carlo

BACKGROUND FROM MISIDENTIFIED LEPTONS:

γ + jet, QCD multijet, W+jets: data driven technique



Events with N_{jet} = 0 subdivided in 3 orthogonal control regions for background modeling test

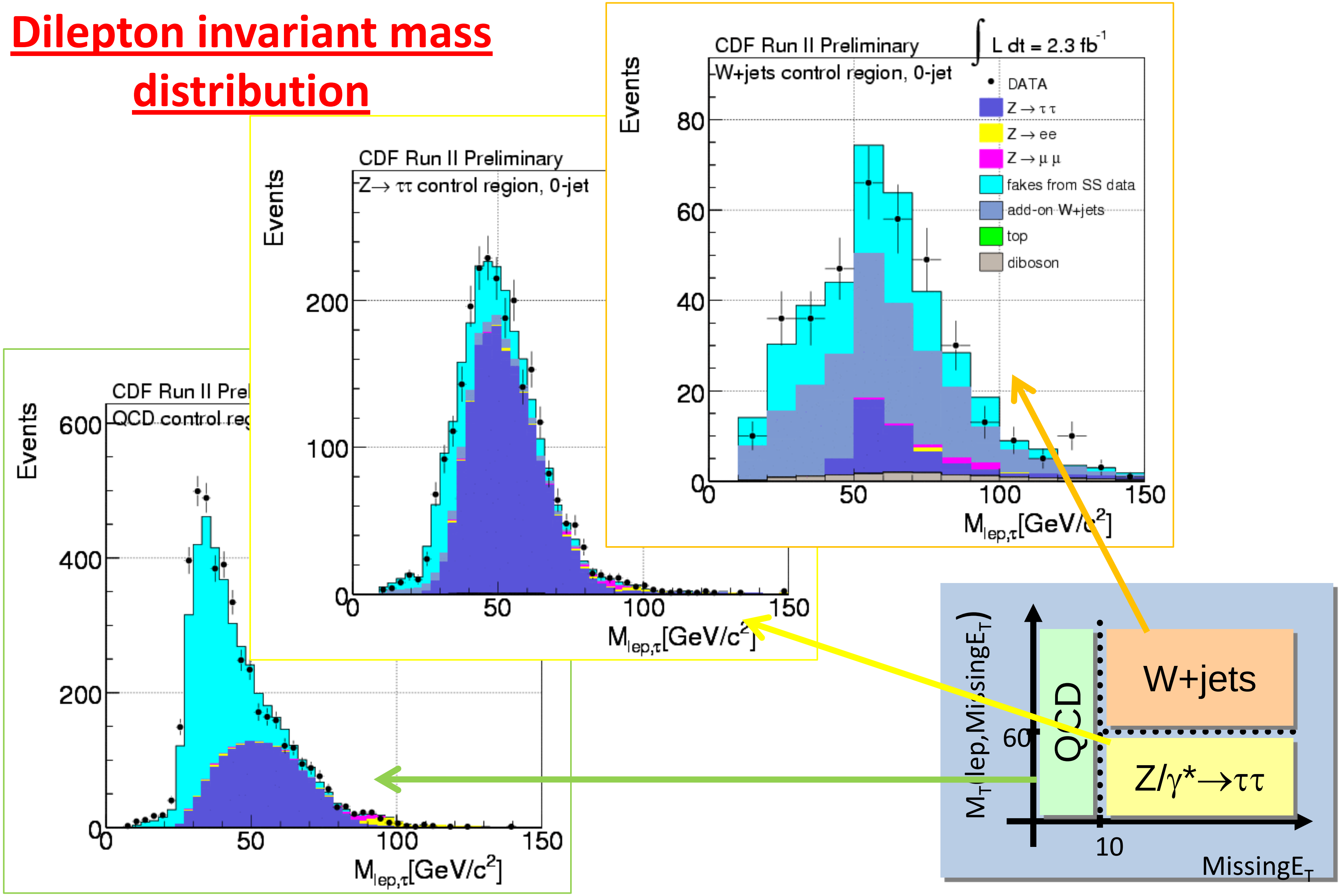
$Z \to \tau\tau$ region: - MET > 10 GeV
- M_T(lep,MET) < 60 GeV

W+jets region: - MET > 10 GeV
- M_T(lep,MET) > 60 GeV

QCD region: - MET < 10 GeV

Background estimation

Dilepton invariant mass distribution



Systematic uncertainties

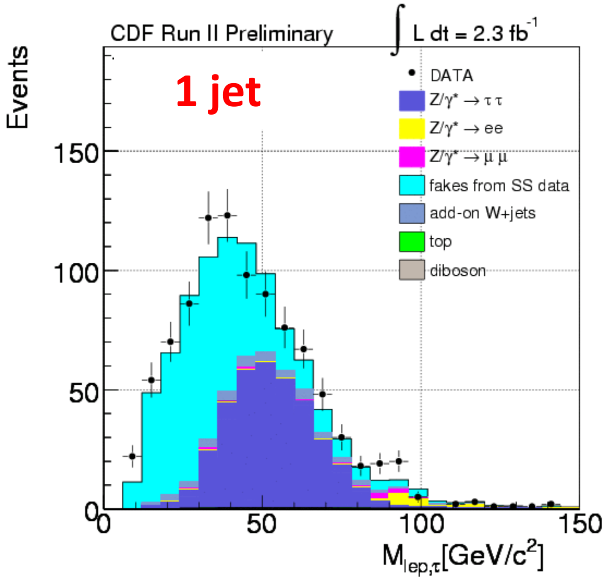
This search relies on a good jet multiplicity modeling.

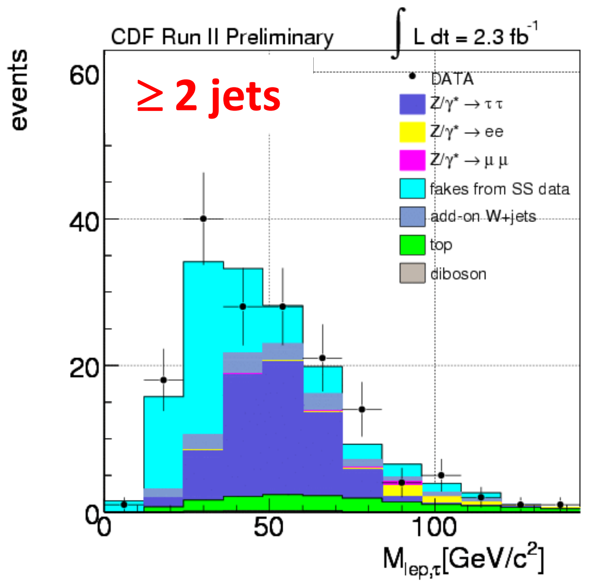
Thus, the main source of systematics for MC-derived processes is the uncertainty on the the **Jet Energy Scale (JES)**

Other sources which have been taken into account are:

- Cross section and MC acceptance
- Parton Distribution Function (PDF) modeling
- W+JETS and QCD modeling
- Initial State Radiation (ISR)
- Final State Radiation (FSR)
- Tau ID scale factor

Signal channels with 2.3 fb^{-1} of CDF data





Signal channels $M_H = 120\ GeV/c^2$

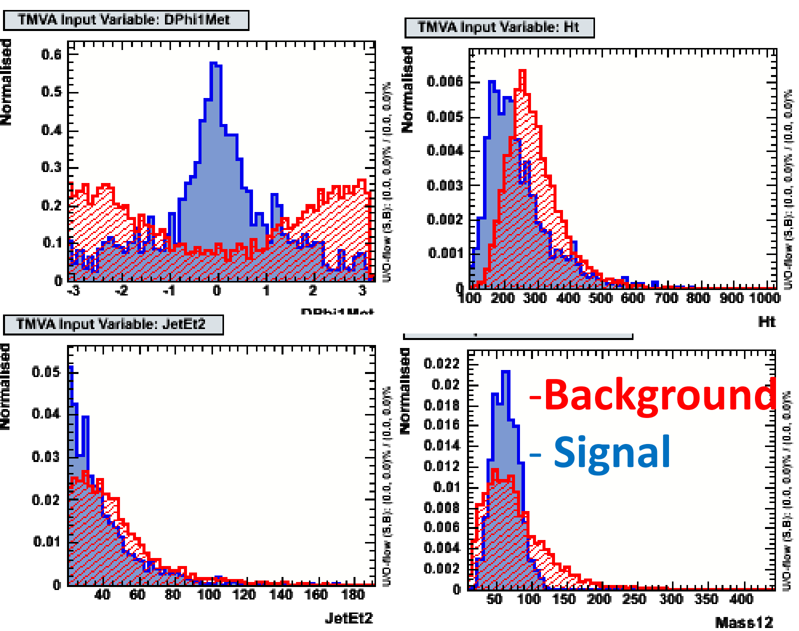
Background source	1 JET			≥ 2 JETS		
$Z/\gamma^* \to \tau\tau$	357.9	±	32.9	59.2	±	8.8
$Z/\gamma^* \to ee/\mu\mu$	26.3	±	2.0	4.8	±	0.7
WW/WZ/ZZ	4.3	±	0.6	1.0	±	0.2
$t\bar{t}$	4.4	±	0.7	15.5	±	2.3
fakes from SS data	483.0	±	48.3	64.0	±	8.0
add-on W+jets	45.8	±	8.2	14.1	±	4.2
Total Background	921.8	±	60.2	158.7	±	13.9
Data		965			166	
ggH	0.572	±	0.164	0.153	±	0.109
WH	0.069	±	0.009	0.150	±	0.015
ZH	0.039	±	0.004	0.100	±	0.009
VBF	0.075	±	0.009	0.104	±	0.014
Total Signal	0.755	±	0.165	0.507	±	0.110

- S/B is small
- Expected signal is much smaller than background uncertainties.

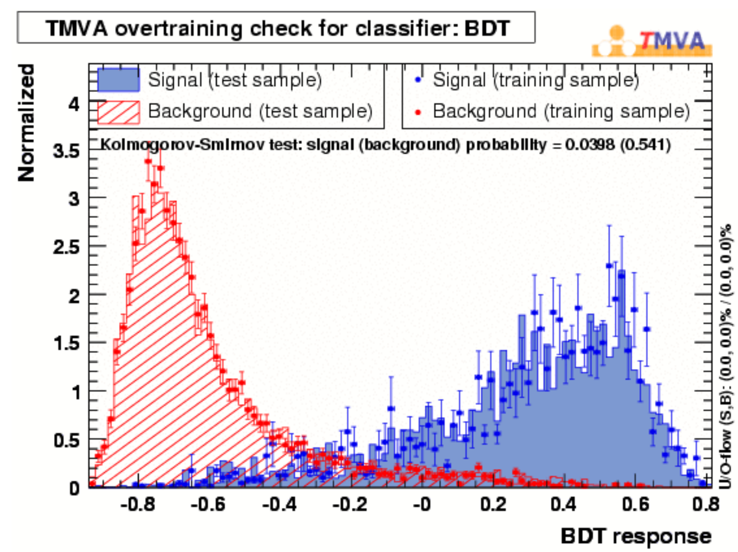
Multivariate techniques

- S/B is small → counting experiment is not possible.
- Need to exploit all the event information to extract the small signal from data

A multivariate technique allows us to combine the **discriminating power** of different kinematical and topological distribution into **one single variable**



MULTIVARIATE ALGORITHM



Building the final discriminant

We build a MULTIVARIATE DISCRIMINANT by combining a set of **Boosted Decision Trees** trained with a choice of 23 kinematical and topological variables

SIGNAL CHANNEL A (1 JET)

BDT1 $H\to\tau\tau$ vs $Z\to\tau\tau$

BDT2 $H\to\tau\tau$ vs QCD

SIGNAL CHANNEL B (≥ 2 JETS)

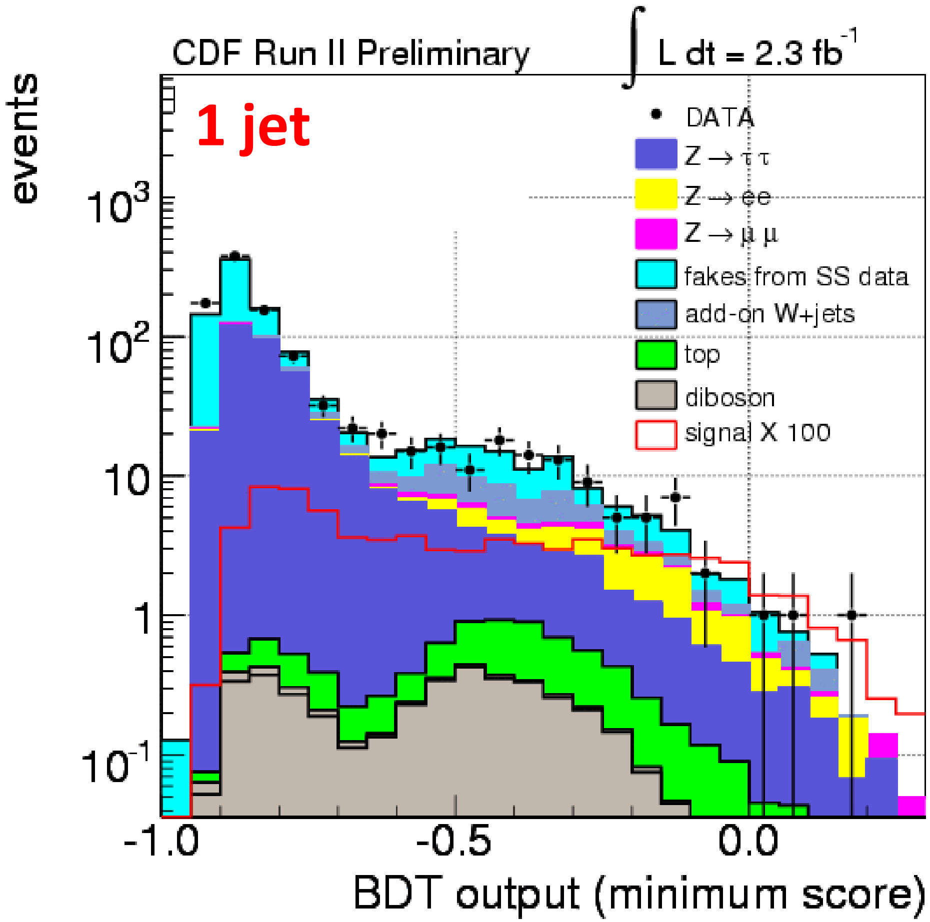
BDT3 $H\to\tau\tau$ vs $Z\to\tau\tau$

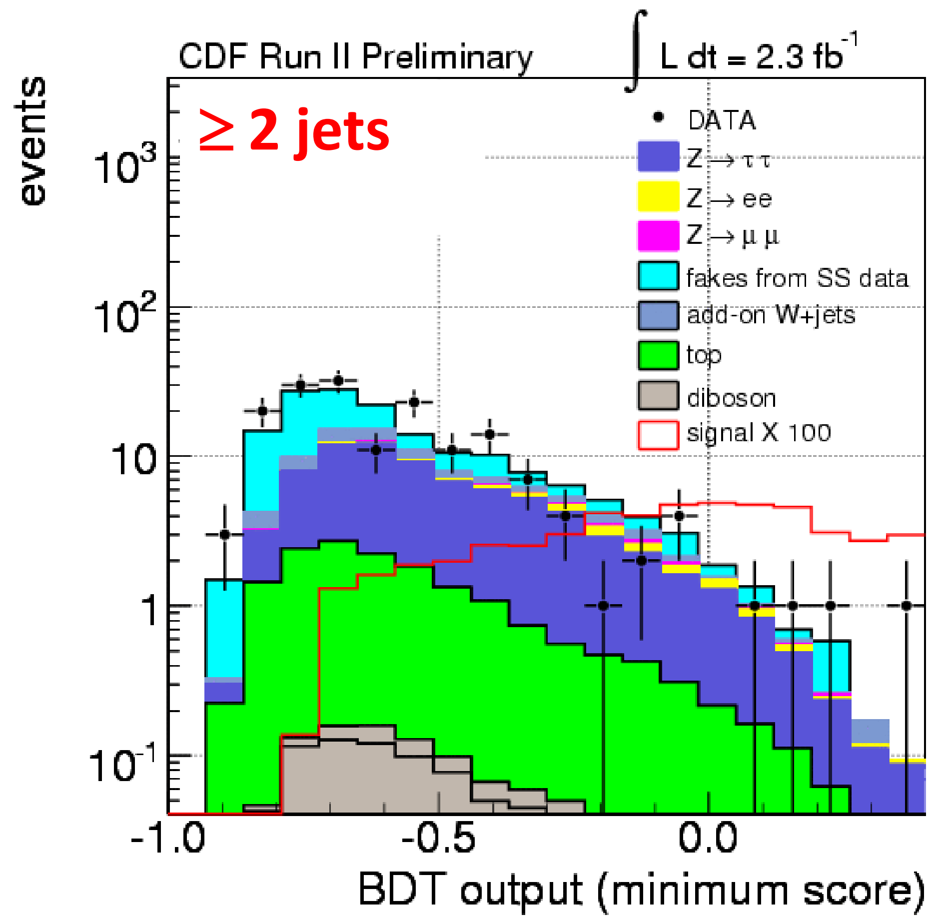
BDT4 $H\to\tau\tau$ vs QCD

BDT5 $H\to\tau\tau$ vs top-antitop

Results: final discriminant

Higgs mass hypothesis: 120 GeV/c^2

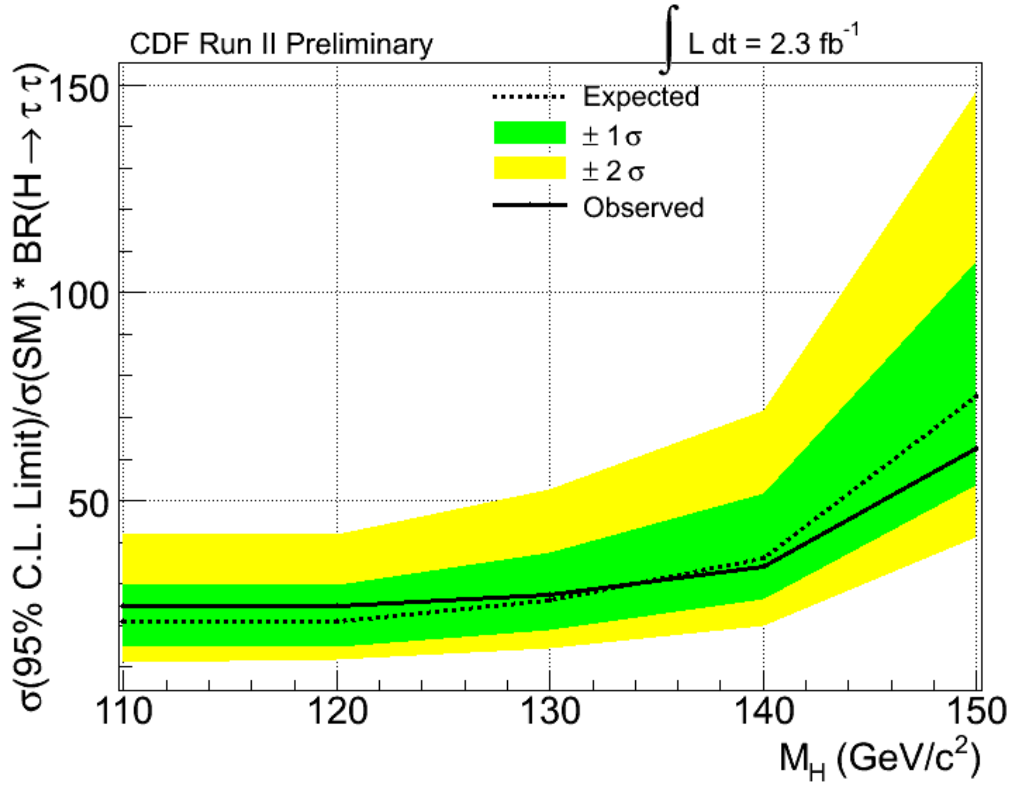




No significant excess observed

Results: 95% C.L. upper limit

Higgs mass	Exp.Limit/SM	Obs.Limit/SM
110	21.0	24.7
120	20.8	24.8
130	26.2	27.4
140	36.3	34.1
150	75.2	62.5



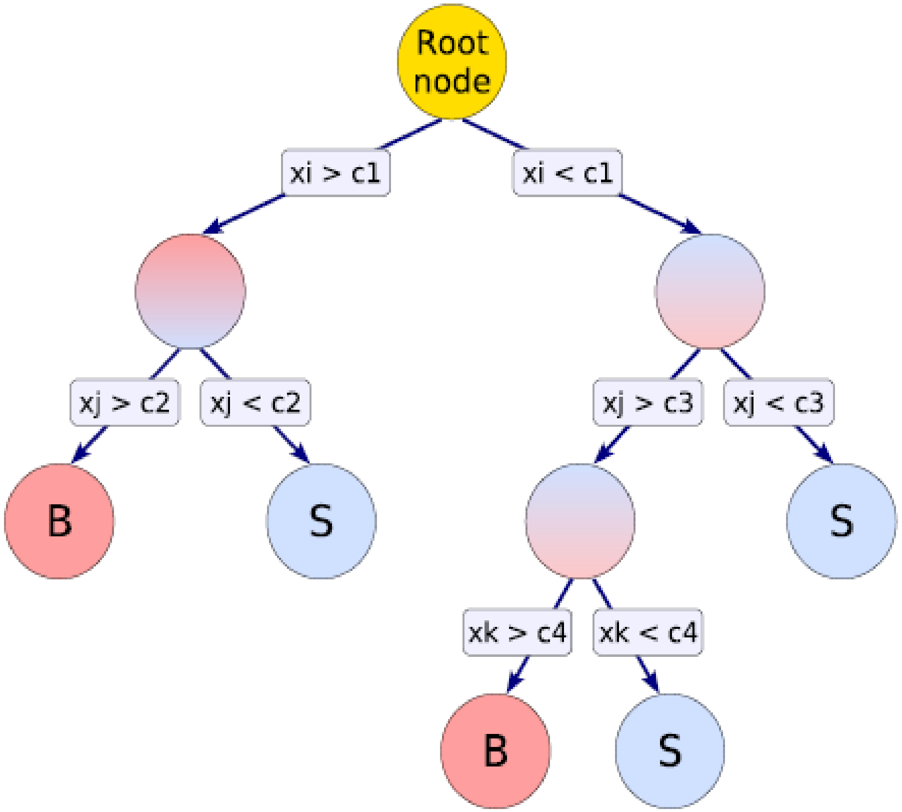
The **net sensitivity improvement** with respect to the previous CDF analysis ranges **from 10% to 40%**

Summary

- A SM Higgs search with improved analysis techniques performed with 2.3 fb^{-1} of CDF data in the di-tau decay mode with:
  - an increased acceptance on signal events → "1 jet channel" included
  - a more performing hadronic tau ID algorithm based on the BDT method
- The sensitivity improvement with respect to the previous CDF analysis ranges from 10% to 40%
- The results will be included in the CDF limit combination for the summer 2010 conferences
- Expect soon the update with 5.0 fb^{-1}!

BACK-UP SLIDES

The Boosted Decision Tree method



A DECISION TREE: a sequence of rooted binary splits

Ingredients : 1) a **training sample** for signal and background
2) a set of **discriminating variables**

At the end of a splitting, leaves are classified as signal-like (event score +1) or background-like (event score -1), accordingly to the purity.

BOOSTING: N trees are created. Events misclassified in the N-th tree, are given an increased weight in the (N+1)th tree.

An event final score is given by the weighted average of different tree outputs