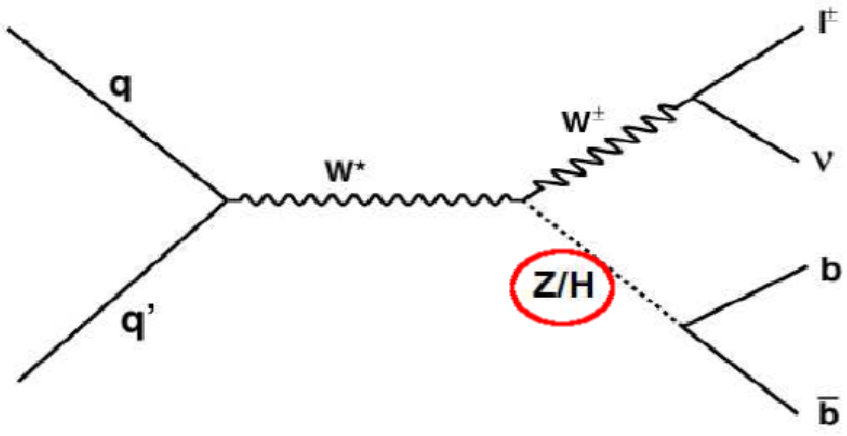
Evidence for $WW/WZ \to l\nu + Heavy\ Flavor$ Jets at CDF

Federico Sforza
on behalf of the CDF Collaboration

INFN & Università di Pisa

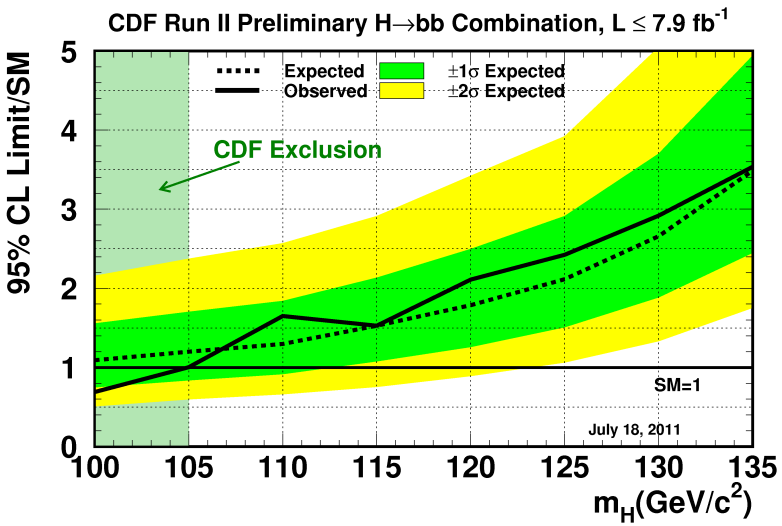
DPF 2011 – 12 August 2011

Evidence for $WW/WZ \rightarrow l\nu + HF$: Outline



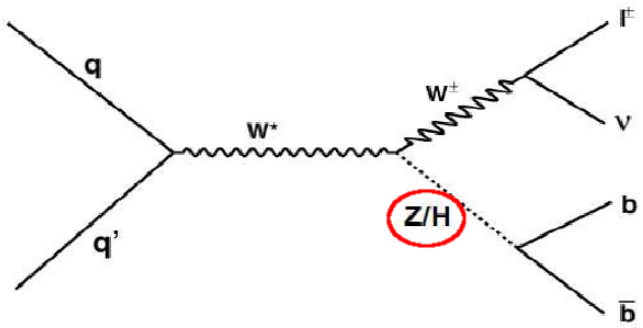
Diboson/Higgs production in the $l\nu+$ Heavy Flavor (HF) Jets Channel.

Dibosons and Higgs$\rightarrow b\bar{b}$



$H \rightarrow b\bar{b}$:

- dominates low mass Higgs exclusion limits;
- decay mode still difficult at LHC;
- test of the SM couplings;
- most important production mode: *WH*



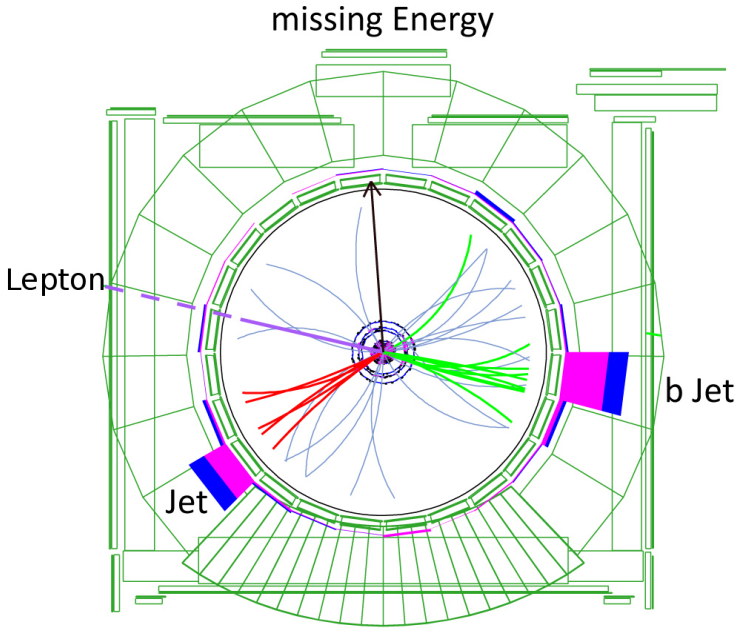
- $WZ \rightarrow l\nu + b\bar{b}$ has same final state as *WH*:
  $\Rightarrow$ larger yield (but still rare, $\sigma_{WZ} \approx 3$ pb)
  $\Rightarrow$ different resonance mass.

- SM processes with Heavy Flavor *to be observed* before a *low mass Higgs*.

Analysis Strategy: $WW/WZ \rightarrow l\nu HF$

Goal: Observe a rare SM resonance in the $l\nu + HF$ Channel:

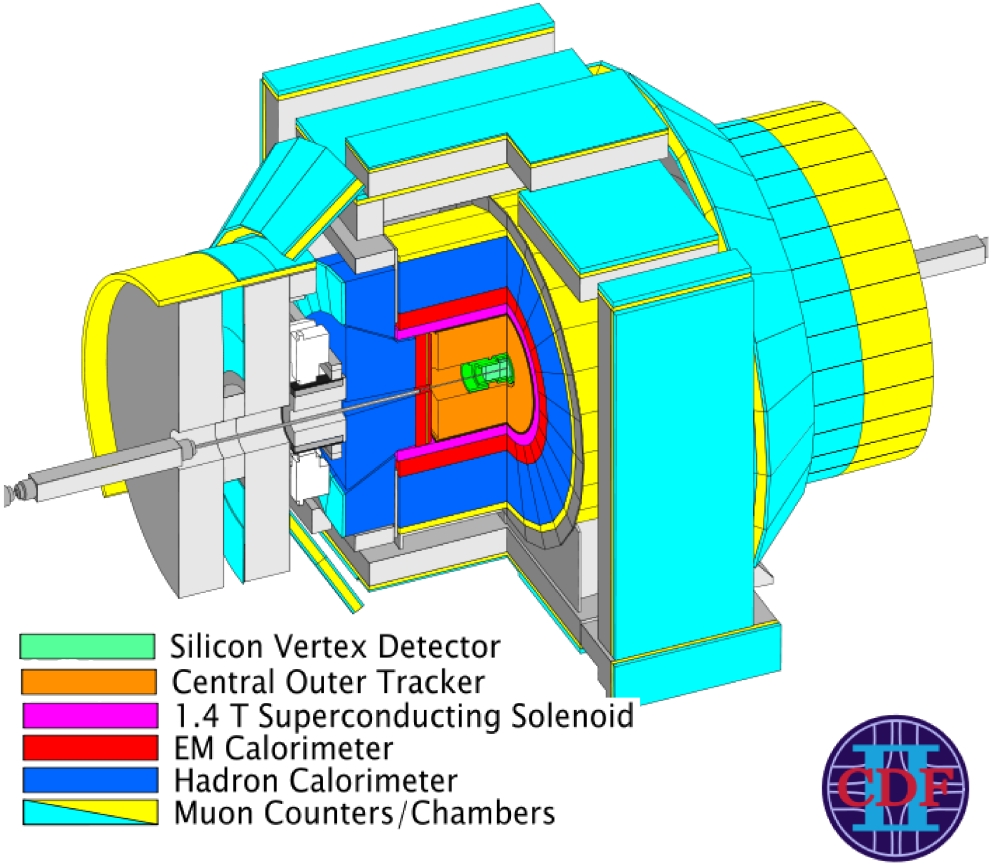
- collect $W \rightarrow l\nu$ candidate with max efficiency:
  $\Rightarrow$ several trigger strategies;
  $\Rightarrow$ release lepton identification requirements.
- Di-jet Heavy Flavor enriched sample selection:
  $\Rightarrow$ require two high energy jets;
  $\Rightarrow$ $b-$hadron identification: secondary decay vertices.
- Signal discrimination based on *di-jet mass*.



The CDF II Experiment

Multipurpose detector:

- **Integrated tracking:**
  ⇒ silicon detector;
  ⇒ wire chamber (COT).
- **Calorimeter system:**
  ⇒ ElectroMagnetic; (EM)
  ⇒ HADronic (HAD)
- **Muon detection:**
  ⇒ outer muon chambers.



- First data was in 2002.
- This analysis uses: $\int \mathscr{L} = 7.5\ \text{fb}^{-1}$ (March 2011).

$W \to l\nu$: Multiple Trigger Strategies

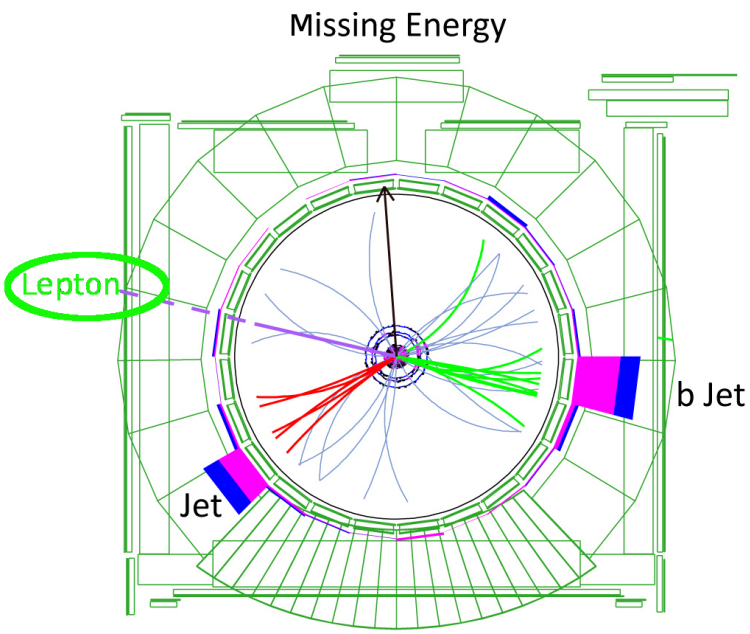
- Need to collect *all* the interesting events:
  $\Rightarrow$ more trigger strategies to increase the acceptance.

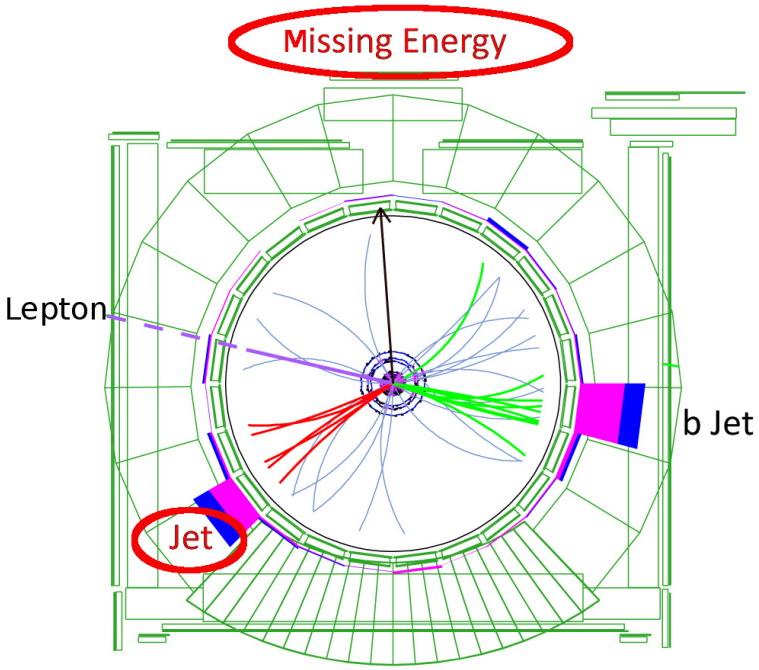
Strategy I:

- single object triggers
- high energy lepton: electron or muon.

Strategy II:

- multiple object triggers
- Missing Energy + Jet.





$W \to l\nu$: Offline Selection

Neutrino: imbalance in the total transverse energy: $\not{E}_T > 20$ GeV.

Tight Leptons: standard set of CDF leptons (e, μ) cuts:

Electrons:

- Energy cluster in EM calorimeter ($E_T > 20$ GeV)
- Energy deposit isolated in the calorimeter.
- Good quality track pointing to the cluster.

Muons:

- Hits in the muon chambers.
- Track pointing to the hits ($P_T > 20$ GeV/c).
- MIP energy deposit isolated in the calorimeter.

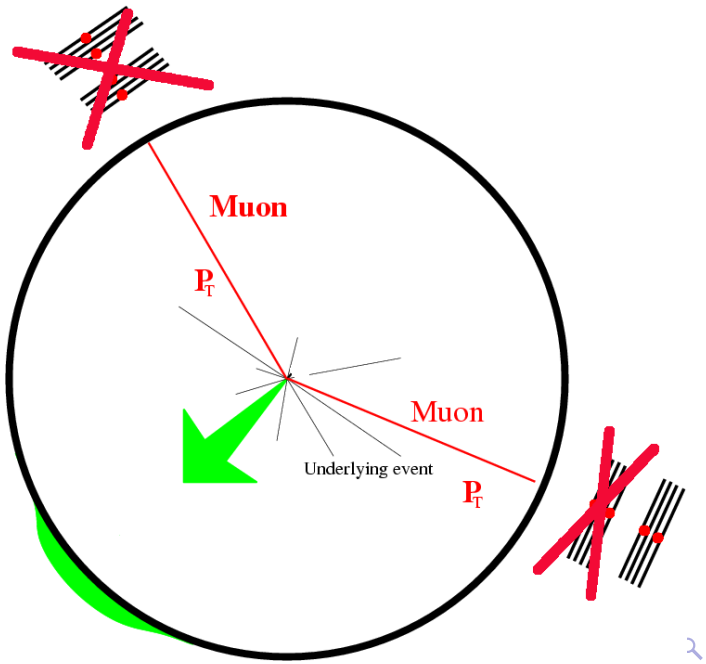
Extended μ Category (EMC): 30% acceptance gain:

Loose Muons:

- few or missing hits in the muon chambers.
- MIP energy deposit isolated in the calorimeter.
- Track pointing to the MIP ($P_T > 20$ GeV/c).

Isolated Tracks:

- Good quality track ($P_T > 20$ GeV/c).
- No other track within $\Delta R = 0.4$.

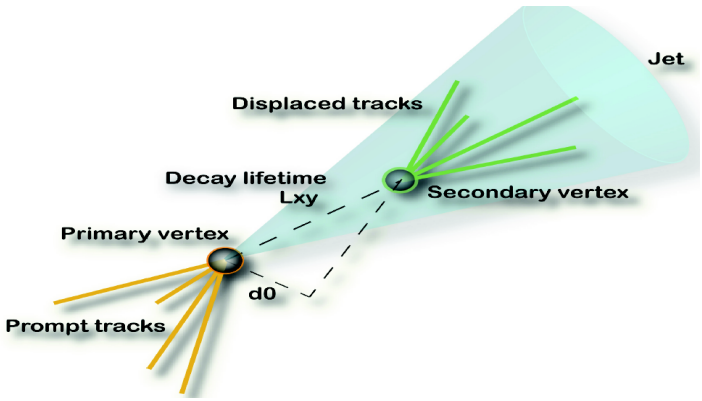


Heavy Flavor Jets Identification

High P_T di-jet sample selection:

- CDF *standard* jet identification: cone algorithm (`JETCLU04`, $\Delta R < 0.4$);
- two central energy clusters: $|\eta| < 2.0$, $E_T > 20$ GeV (corrected for detector effects).

`SecVtx` algorithm



Heavy Flavor hadrons identification:

- "b-tag": secondary decay vertex.
- *b*-tag efficiency: $\sim 40\%$;
- *c*-tag: efficiency $\sim 6\%$;
- fake tags $\sim 1\%$ *(background)*.

3 selection regions

1. **Pretag:** no `SecVtx` request, control region.
2. **Single Tag:** exactly 1 `SecVtx` tagged jet, signal region.
3. **Double Tag:** exactly 2 `SecVtx` tagged jet, signal region.

WW vs WZ Selection

Interested in $WZ \to l\nu b\bar{b}$ but not sensitive yet...

- Must use more advanced techniques (some already in use in CDF WH analysis):
  $\Rightarrow$ improve $b-$tag efficiency
  $\Rightarrow$ use multivariate discriminants;
  $\Rightarrow$ ....
- Ongoing effort to obtain 3σ significance together with D0 for WZ alone.

Di-jet invariant mass resolution does not allow to distinguish WW from WZ.

- $WW \to l\nu + cs$ contributes to Heavy Flavor signal:
  - $c-$quark produces long living hadrons (6% tagging efficiency);
  - higher cross section for WW associate production than WZ;
  - $W \to cs$ branching ratio $\approx$ 30%;
  - Large acceptance w.r.t WZ: both W can decaying to $c-$quarks or to leptons.

We can probe high$-P_T$ behavior of $c-$quarks too!

Background Composition

Signal topology: *exclusive 1 lepton* $+ \not{E}_T + 2$*jets(1 or 2 tags)*

background estimate: Monte Carlo simulation & data control samples

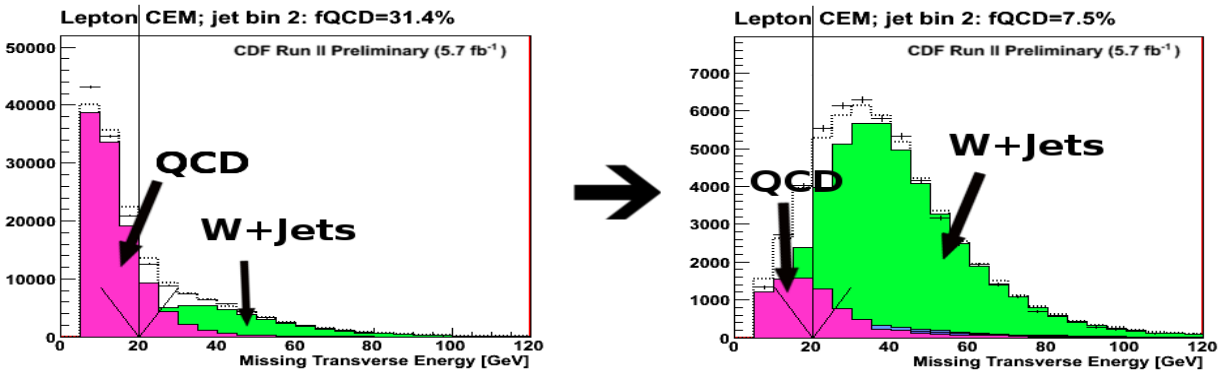
$\Rightarrow$ **4 background components:**

1. **MC based**: $t\bar{t}$, single top, Z+jets $\Rightarrow$ estimated from MC .
2. **QCD**: lepton and $\not{E}_T$ are faked by mis-reconstructed jets.
   $\Rightarrow$ measured in data.
3. **W+HF**: Heavy Flavors $\Rightarrow$ major background, large uncertainty.
   - Total W+jets Normalization obtained from data (pretag control region);
   - Heavy Flavor Fractions $= \frac{W+HF}{W+jets}$ estimated from MC.
4. **W+LF**: Light Flavors, W+fake tags $\Rightarrow$ parametrized on jet data.

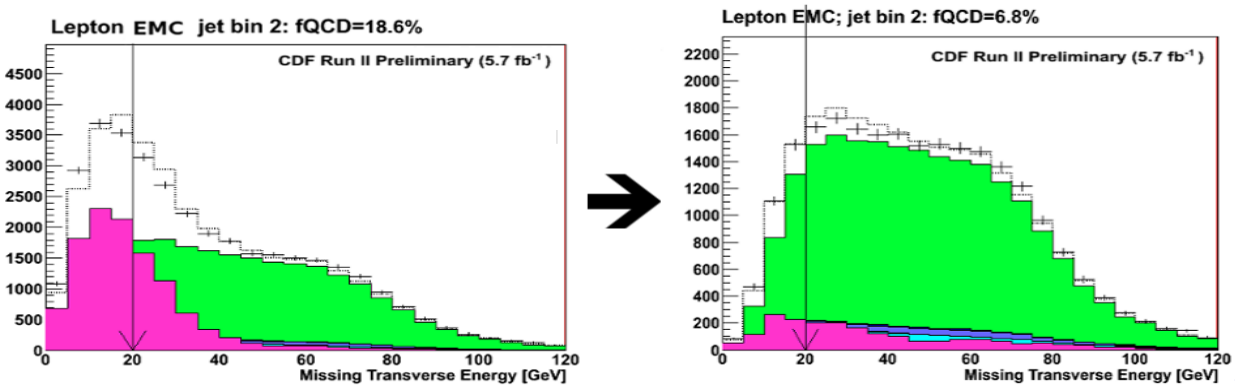
QCD Modeling and Rejection

- Fake W models obtained reversing some of the lepton identification cuts.
- Multivariate tool (based Support Vector Machine) drastically removes QCD.
- Normalization extracted from data: QCD, $W + jets$ two component fit on $\not{E}_T$.

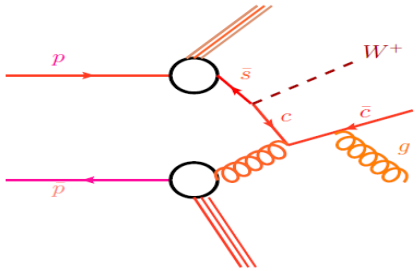
Tight electron QCD estimate before and after QCD rejection:

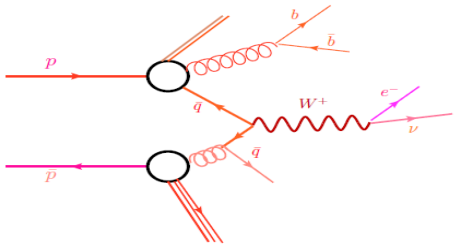


EMC leptons QCD estimate before and after QCD rejection:



W+Heavy Flavor Background





$W + b\bar{b}$, $W + c\bar{c}$, $W + c$ estimate

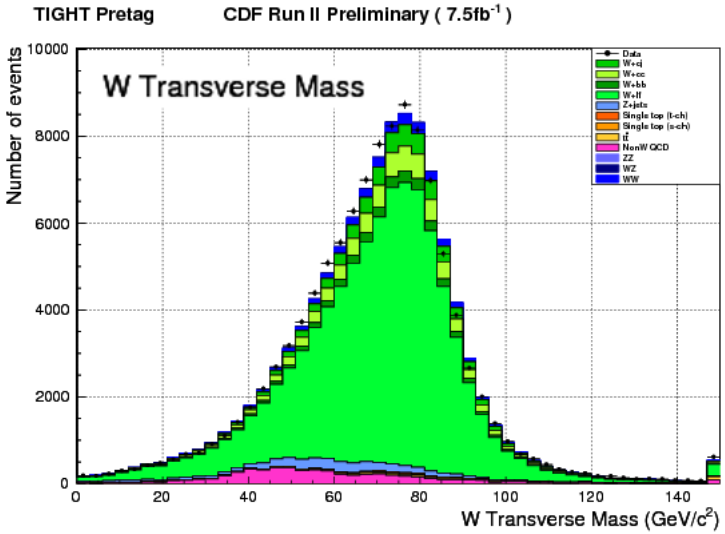
- Large theoretical uncertainty on σ_{W+jets}.
- Normalization (N^{W+jets}) from the *pretag* data sample:

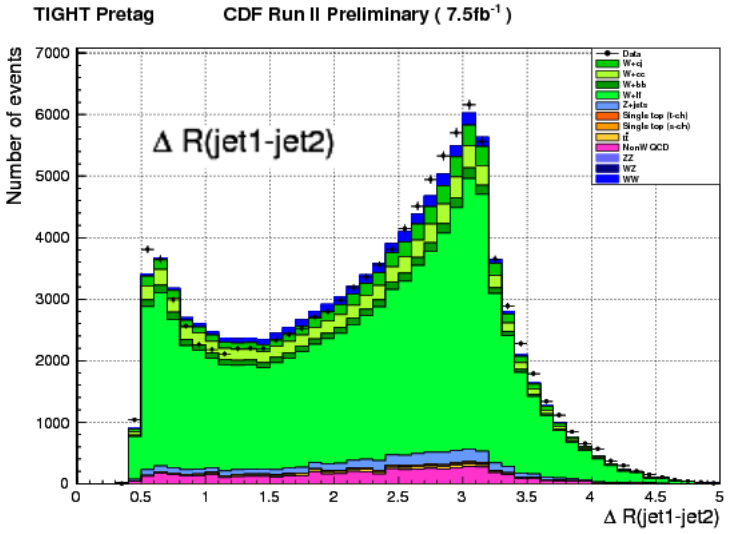
$$N^{W+jets} = N^{Data}_{Pretag}(1 - F^{\mathrm{QCD}}) - N^{MC}$$

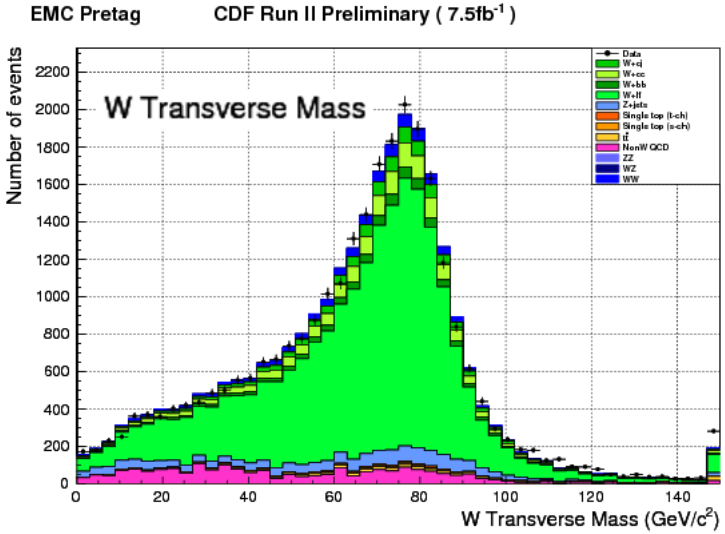
- Heavy Flavor Fractions: $f^{HF} = \frac{W+HF}{W+jets}$ derived from MC (Alpgen, LO)
  $\Rightarrow$ Technically difficult $\approx$ 90 exclusive MCs used.
- Final normalization: $N^{HF} = N^{W+jets} \times f^{HF} \times K$
- $K = 1.4 \pm 0.4$, correction factor to HF production rate in MC:
  $\Rightarrow$ derived from $W + 1$ jet control sample.

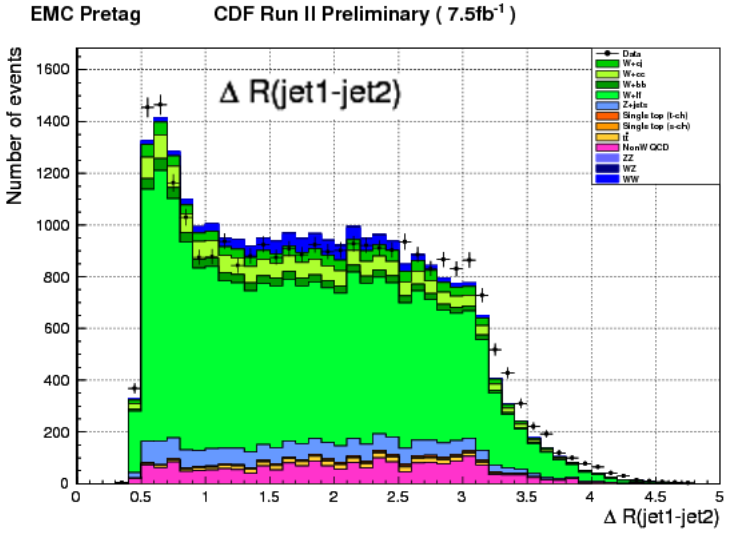
Understanding the Sample: Pretag

- Control region;
- Very large statistic;
- Dominated by $W + LF$



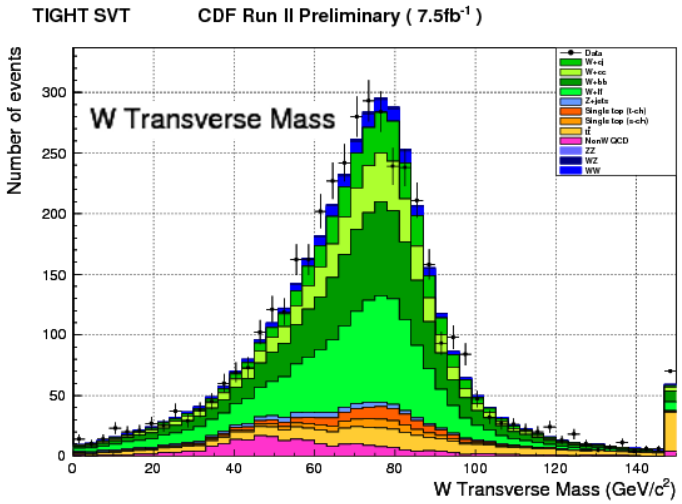


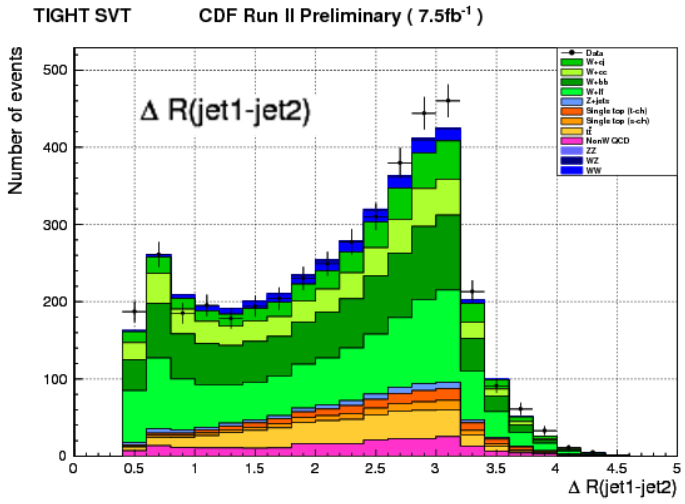


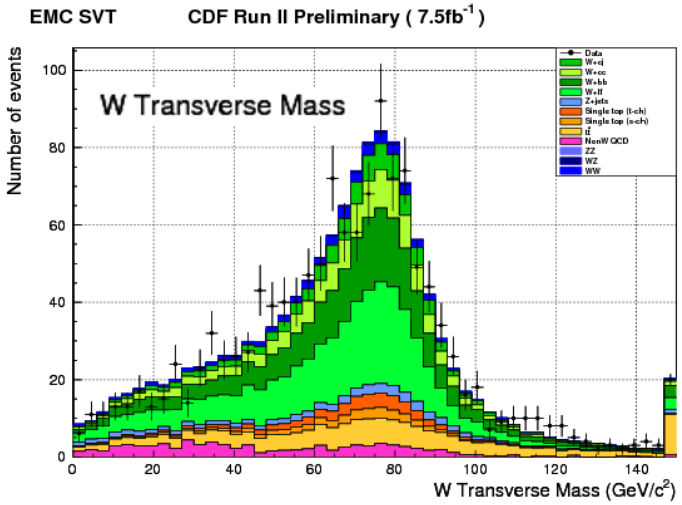


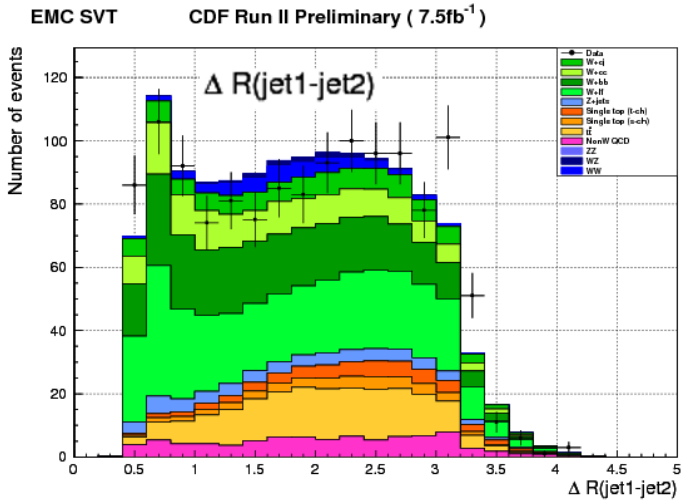
Understanding the Sample: Single Tag

- Signal region;
- Large statistic;
- $W + c$, $W + b\bar{b}$, $W + LF$, top



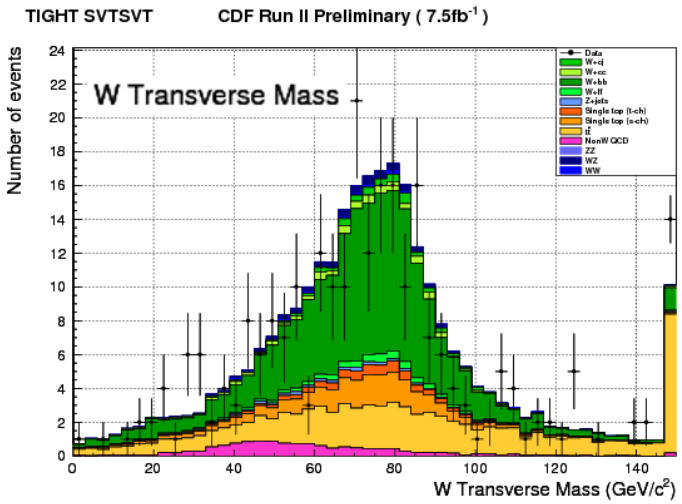


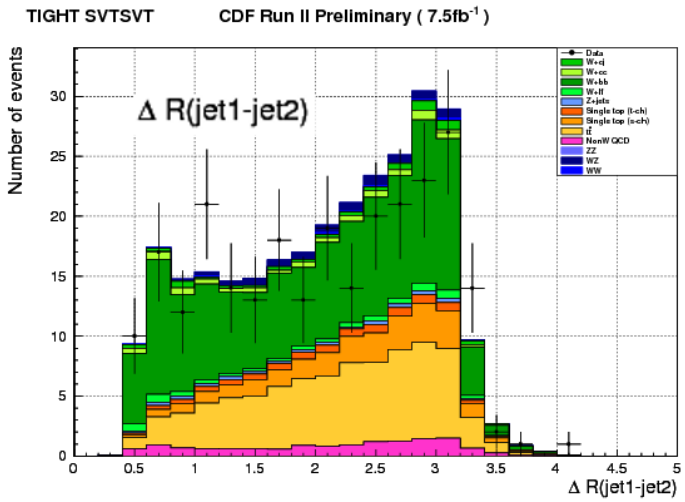


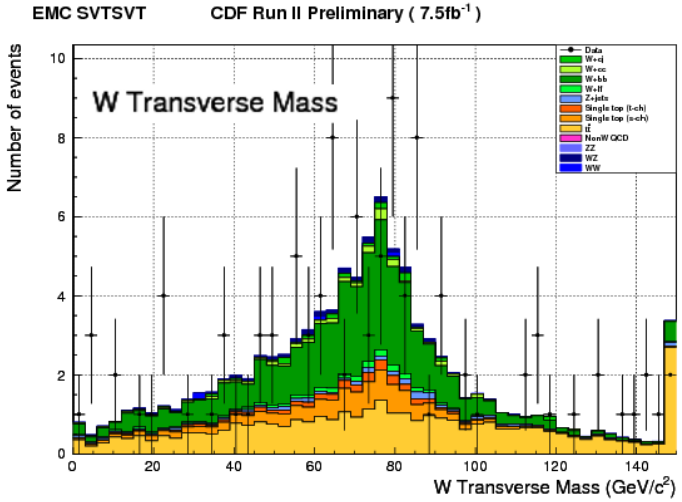


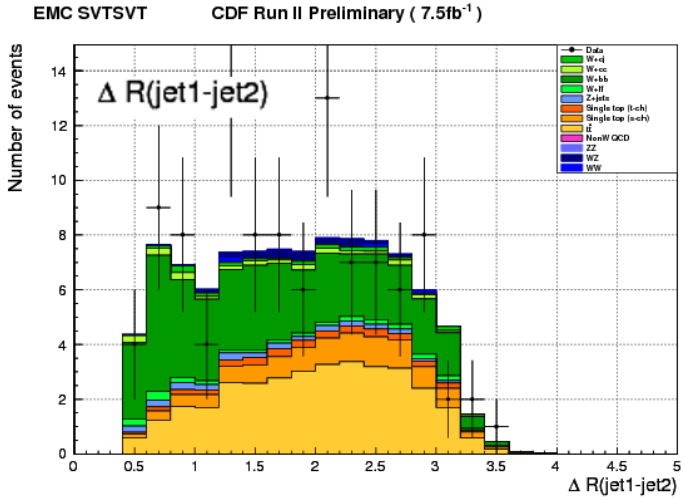
Understanding the Sample: Double Tag

- Signal region;
- Low statistic;
- Dominated by $W + b\bar{b}$ and top





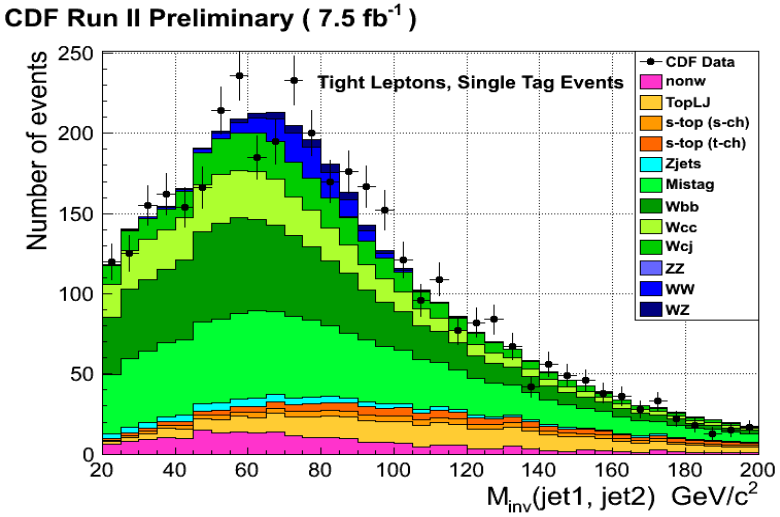


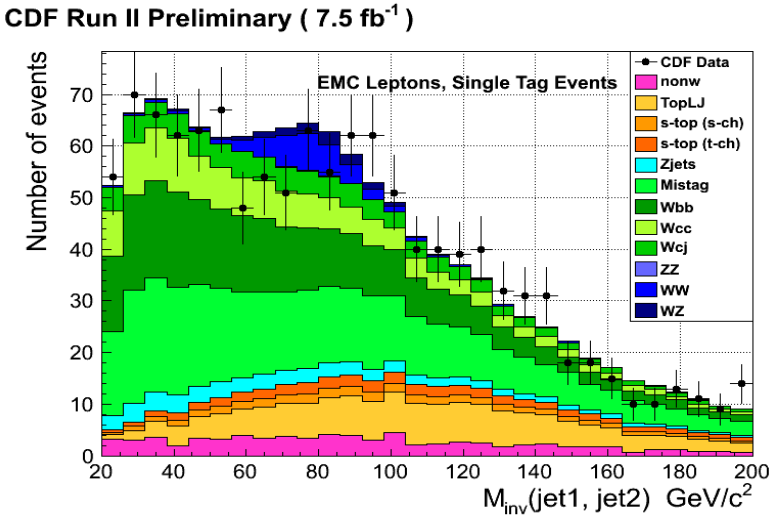


RESULTS

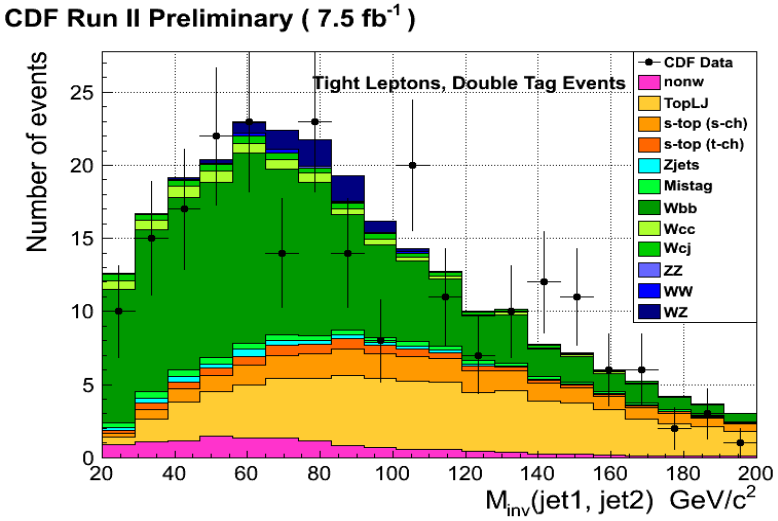
Di-jet Mass Spectrum: 4 Channels

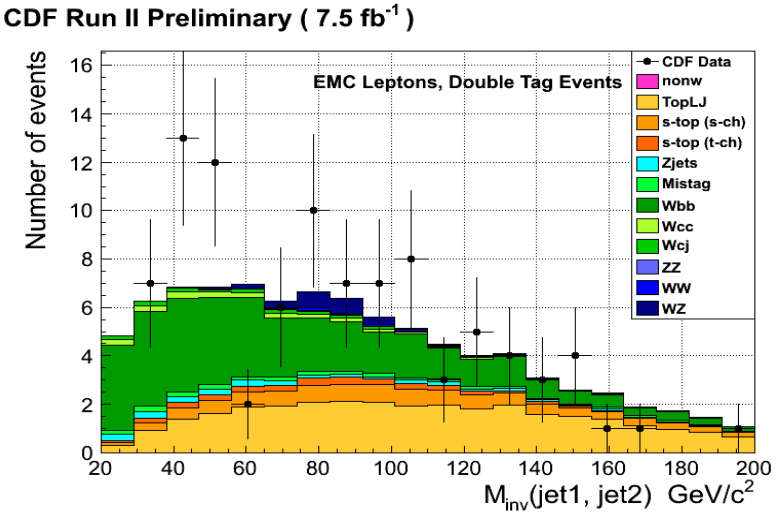
Single Tag (**Tight + EMC**): sizable signal excess over W+jets mass spectrum.





Double Tag (**Tight + EMC**): *WZ* main signal contribution: still statistically limited.





Systematics Inclusion and Fitting Procedure

- Shape and rate systematics are considered in the measurement:

(some of the) systematics	rate (priors)	shape
Trigger	1-3%	no
Lepton ID	1-4%	no
Luminosity	6%	no
$b-$tag SF	5-10%	no
QCD Normalization	40%	no
JES	5-10%	yes
Alpgen Q^2	3-6%	yes
HF K factor	30%	no

- Some systematics are large due to extrapolation from control regions:
  - K factor: 30%, QCD: 40%;
- Shape systematics can shift background and signal peak regions.
- A maximum binned likelihood fit is performed to enhance the sensitivity.
  - Prior uncertainties constrained by data.

Significance of the Observation

Significance:

1. Define a test statistics (Likelihood ratio):

$$-2lnQ = -2ln\frac{p(data|H_{Test},\hat{\theta})}{p(data|H_{Null},\hat{\hat{\theta}})} \qquad \theta\text{: best fit nuisance parameters}$$

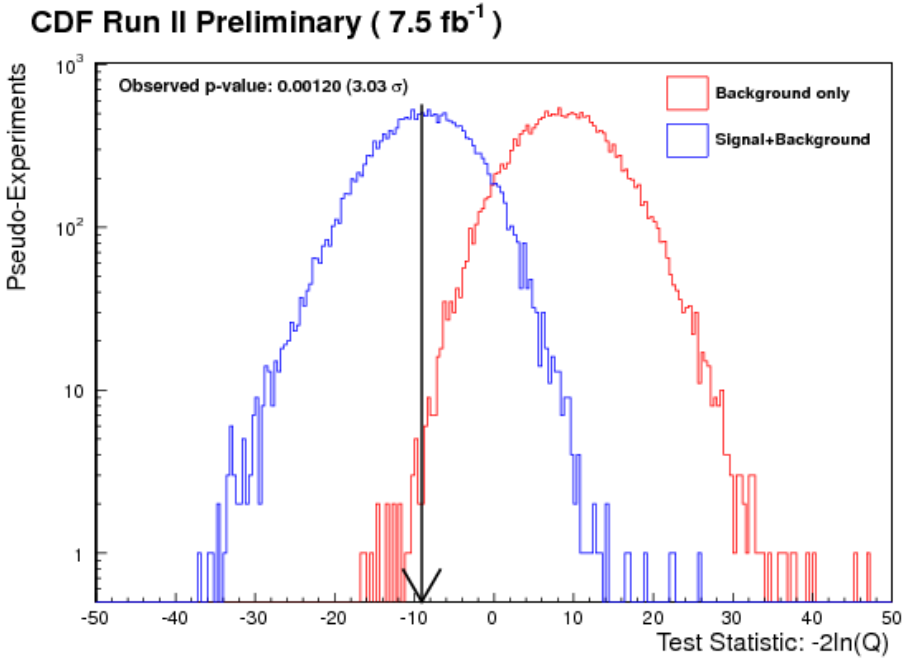
2. Throw background only (H_{Null} distribution) ***pseudo-experiments***.
3. compare **data** $-2lnQ$ against *pseudo-experiments* $-2lnQ$.

- Expected $-2lnQ$: $-8.85^{+6.2}_{-6.4}$ (Test hypothesis)

$$p - value = 0.00125 \Rightarrow 3.02\sigma$$

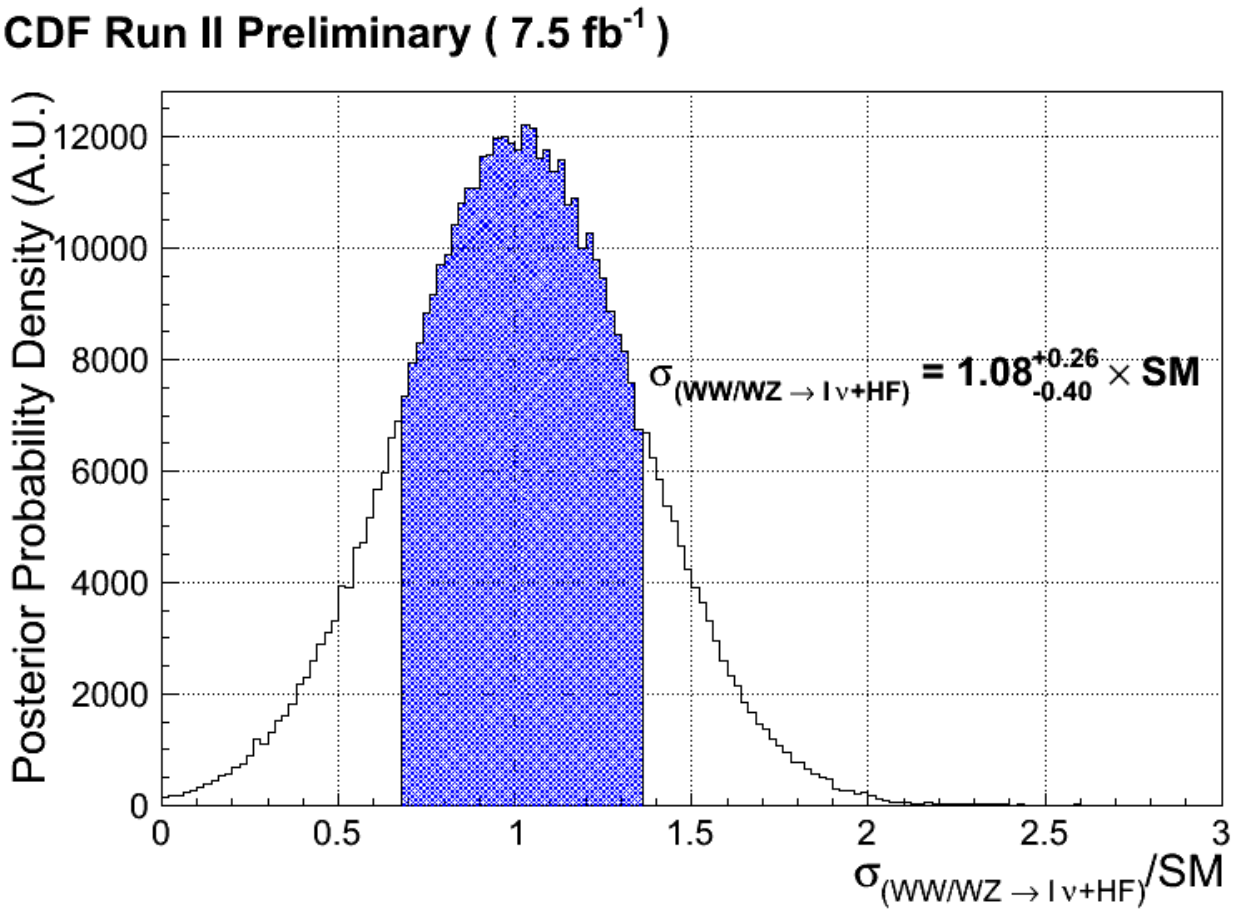
- Observed $-2lnQ$: -9.0446.

$$p - value = 0.00120 \Rightarrow 3.03\sigma$$



Cross section measurement:

- Central value ⇒ maximum of the Bayesian posterior.
- Error ⇒ minimum interval covering 68% of the Bayesian posterior.



Conclusions

Analysis consclusions:

- obtaied 3σ evidence of WW/WZ production in the tagged, *heavy flavor* enriched channel using di-jet mass as signal discriminant;
- measured cross section in good agreement with SM and other diboson channels measurements:

$$\sigma_{(WW/WZ \to l\nu bb/l\nu cs)} = 1.085^{+0.26}_{-0.40} \times SM;$$

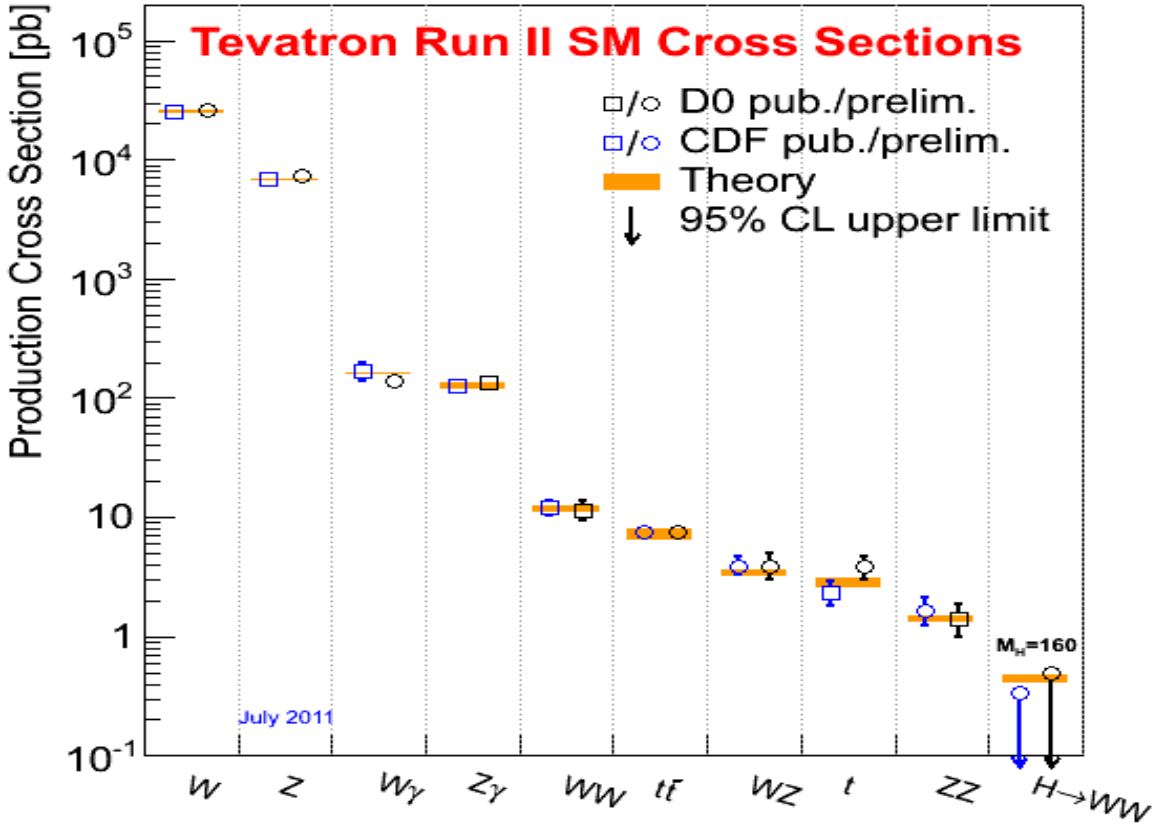
- strong support to the CDF *WH* analysis strategy and results;
- public material available at:

  `http://www-cdf.fnal.gov/physics/new/hdg/Results_files/results/wzlnubb_071911/`

Back Up Slides

Looking for the Rarest Processes

SM diboson production cross sections at the TeVatron:



Extremely rare processes:

$\sigma_{p\bar{p}\to W^+W^-}$
11.34 ± 0.7 pb

$\sigma_{p\bar{p}\to W^\pm Z^0}$
3.22 ± 0.2 pb

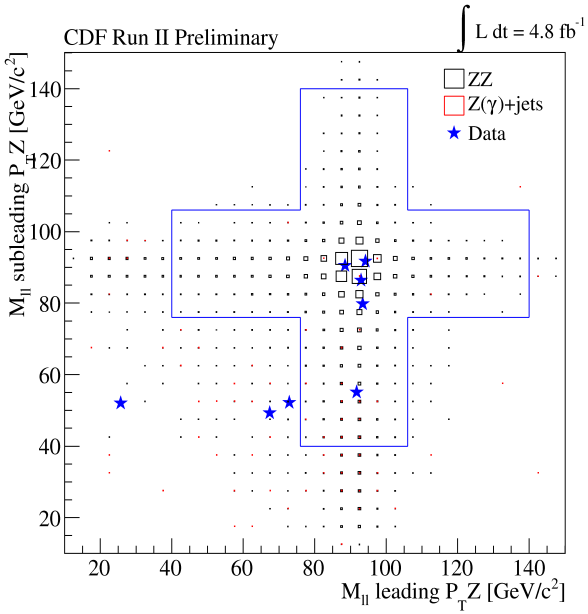
$\sigma_{p\bar{p}\to Z^0Z^0}$
1.2 ± 0.07 pb

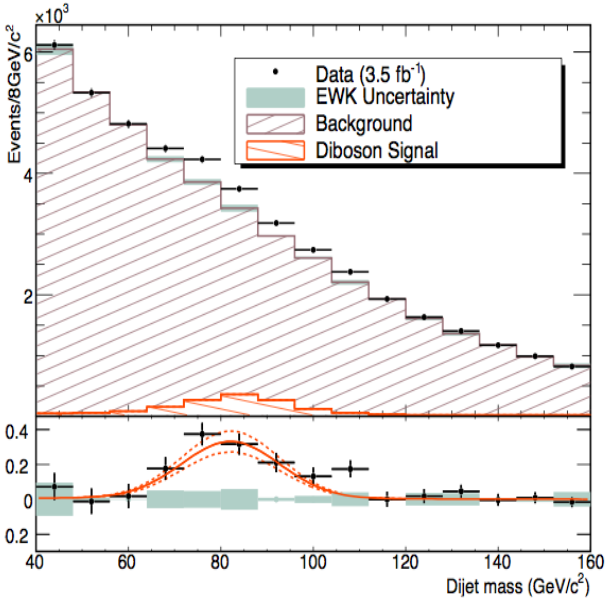
(arXiv:1107.5518)

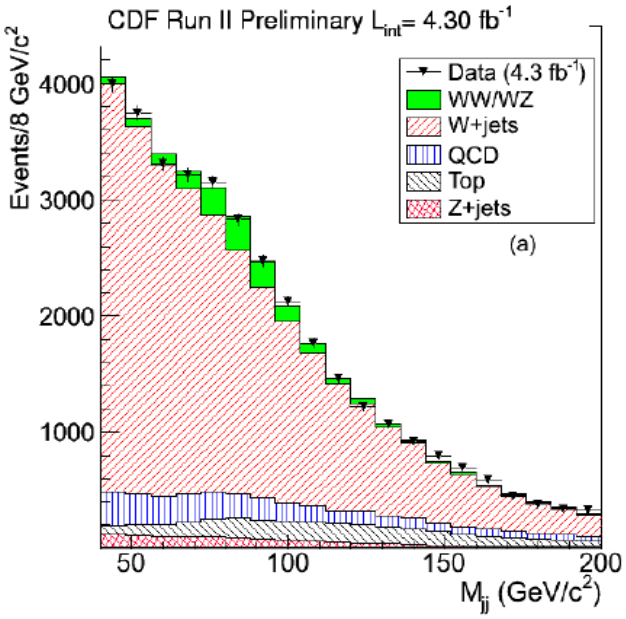
TeVatron experiments already observed all the diboson production modes!

Diboson Observations at CDF

Observation in both leptonic and hadronic mode:







- $ZZ \to llll$;
- four leptons final state;
- clean signature;
- CDF Public Note 9910 (2009).

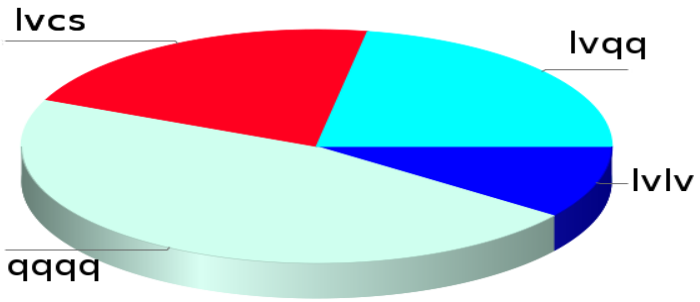
- $WW/WZ/ZZ \to \nu\nu + 2\text{Jets}$
- invisible energy + hadronic final state
- PRL 103, 091803 (2009)

- $WW/WZ \to l\nu + 2\text{Jets}$
- semileptonic decay + hadronic final state
- PRL 104, 101801 (2010)

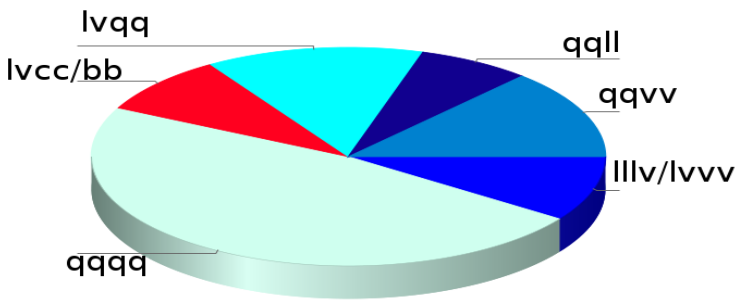
Next Challenge: Lepton + Neutrino + Heavy Flavor

Dibosons have a variety of final states:

WW



WZ

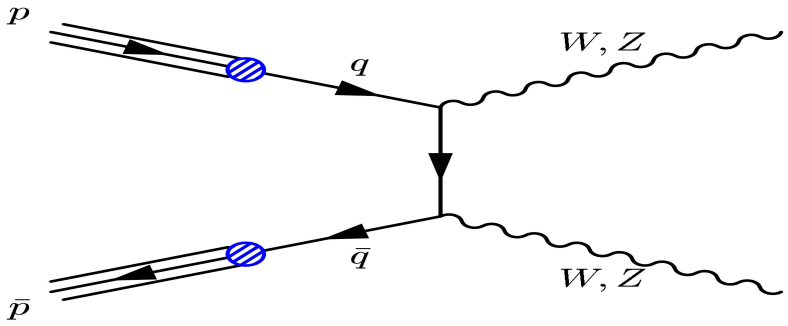


Experimental challenges:

- Selection of long living heavy hadrons in jets:
  - identification of *secondary* decay vertex $\Rightarrow$ low efficiency.
- Reconstruct mass resonances in Heavy Flavor enriched final states.
  - hadronic environment $\Rightarrow$ low energy resolution;
  - background dominated by large irreducible components.

Triple Gauge Couplings

t-channel:

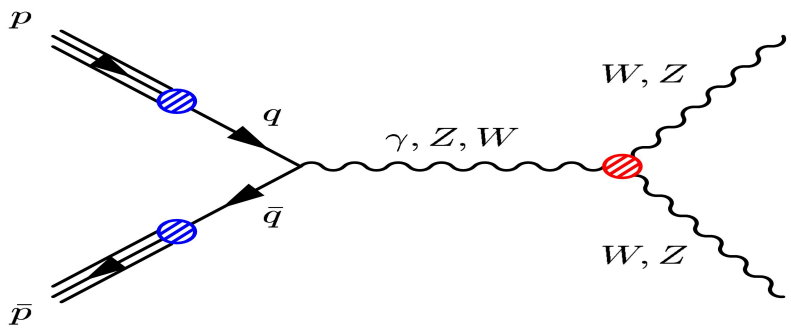


Absent at LEP

$q\bar{q}' \to W^{(*)}$	$\to W\gamma$:	$WW\gamma$ only
$q\bar{q}' \to W^{(*)}$	$\to WZ$:	WWZ only
$q\bar{q} \to Z/\gamma^{(*)}$	$\to WW$:	$WW\gamma, WWZ$
$q\bar{q} \to Z/\gamma^{(*)}$	$\to Z\gamma$:	$ZZ\gamma, Z\gamma\gamma$
$q\bar{q} \to Z/\gamma^{(*)}$	$\to ZZ$:	$ZZ\gamma, ZZZ$

Absent in SM

s-channel:



- s-channel sensible to Triple Gauge Couplings (TGC);
- Cross section can be enhanced by New Physics.

Selection Overview

- High energy lepton ($E_T(P_T) > 20$ GeV(Gev$/c$)). *Maximize acceptance:*
  - TIGHT: CEM, CMUP, CMX.
  - LOOSE: BMU, CMU, CMP, CMXNT, CMIO, SCMIO.
  - ISOTRK: $\Delta R > 0.1$ w.r.t other categories.
  - LOOSE+ISOTRK classified in Extended Muon Category (**EMC**)
- LOOSE+TIGHT Dilepton Veto (TopTools definition);
- LOOSE+TIGHT+TRACK Z Veto (TopTools definition);
- $\not\!\!E_T$ corrected for μ, track MIP, $L5$ jet corrections, Primary Vtx:
  - $\not\!\!E_T > 20$ GeV for CEM and EMC (standard cut).
  - $\not\!\!E_T > 10$ GeV for CMUP and CMX (**relaxed cut**).
- QCD rejection based on Support Vector Machine Algorithm.
- 2 TIGHT Jets ($E_T^{L5Cor} > 20$ GeV, $|\eta| < 2.0$).
- 1 or 2 SecVtx TIGHT Heavy Flavor Tags.

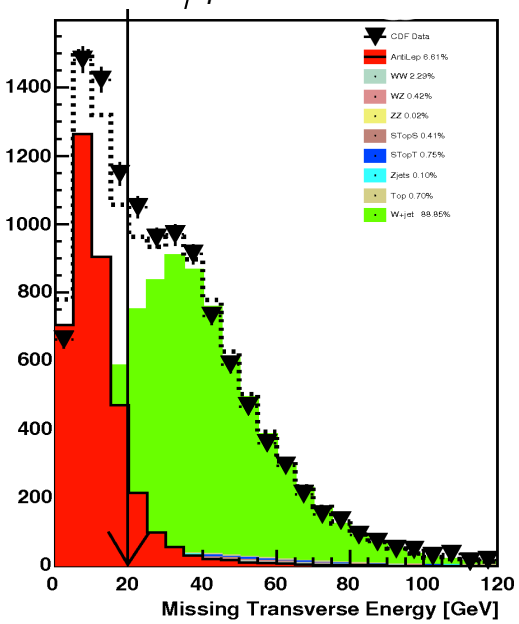
QCD Background (Multi-jet Events)

Lepton and $\not\!\!E_T$ faked by mis-reconstructed jets: hard to model instrumental effects.

- Fake W models obtained reversing some of the lepton identification cuts.
- Normalization measured from data using a two component fit on $\not\!\!E_T$ (QCD, W+Jet).
- Key element for total W+Jet estimate:

$$N^{\mathrm{W+jets}} = N^{Data}_{Pretag}(1 - F^{\mathrm{QCD}}) - N^{MC}$$

Example of a $\not\!\!E_T$ fit on a muon sample:



QCD normalization and modelling have large uncertainties:

- important to reduce the QCD contribution.
- New multivariate tool (based Support Vector Machine) drasticaly removes QCD.

QCD and Multivariate Techniques

- **Electrons sample**: large multi-jet contamination.
- **Modeling fake** W: "anti-electron" sample $\Rightarrow$ reverse $\geq$ 2 out of 5 cuts for the shower-id;

Is it possible to use multivariate techniques to solve this problem?

- Support Vector Machines algorithm supposed to be optimal in this case.
  - sample statistically limited ($\approx 12k$ events);
  - partial bias due to inverted selection.
- **Results**: optimal sample of *6 kinematic* variables:
  - W related: P_T^{ele}, $\not{E}_T^{cor}$, $\Delta\phi(\mathrm{ele}, \not{E}_T^{raw})$
  - Jets related: Jet2 E_T^{raw}, Jet2 E_T^{cor}, $\not{E}_T$Significance.
- $F^{\mathrm{QCD}} \lesssim 10\%$;
- signal efficiency: $\varepsilon_{W(e,\nu)+2\mathrm{jets}} \approx 95\%$, $\varepsilon_{WZ} \approx 97.5\%$.

More details in CDF NOTE 10197

- Software and results presented at CHEP2010 (22 October, Taipei)

SVM Discriminant

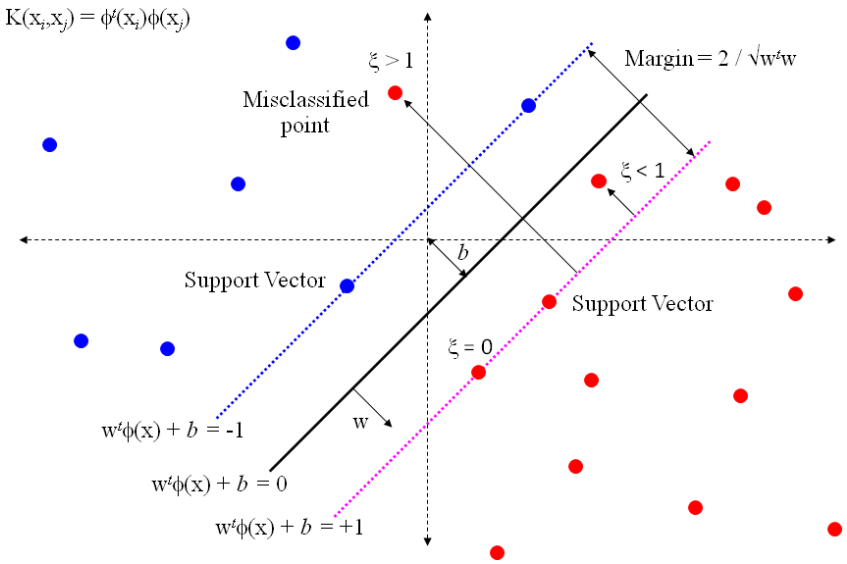
Concept: best hyper-plane dividing two classes of vectors.

The Support Vector Machines (or SVM) are optimal in low statistics separable samples.

- Minimization of $|\vec{w}|^2$ ($\vec{w} \equiv$ normal to the plane) with constrain:

$$y_i(\vec{x_i} \cdot \vec{w} + b) - 1 \geq 0 \quad \begin{cases} y_i = +1; & i \in \text{signal} \\ y_i = -1; & i \in \text{bkg} \end{cases}$$

- Equivalent to maximize:
  $L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \vec{x_i} \cdot \vec{x_j}$



- Non-linear separation obtained with a transformation on the scalar products:

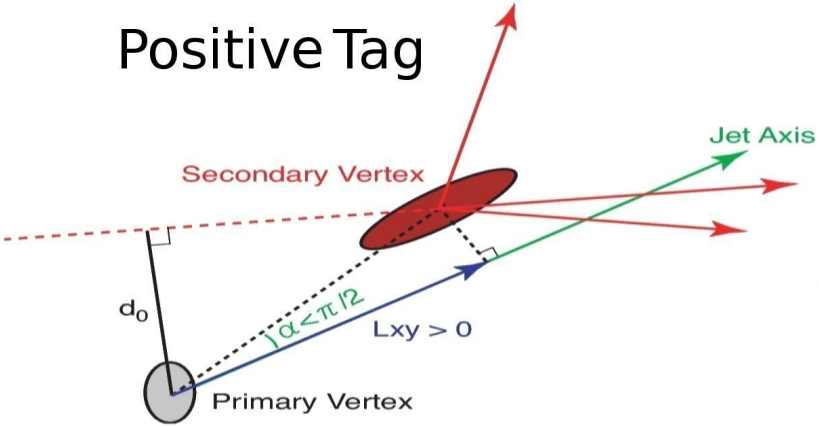
$$\mathbf{K}(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad \text{with } \phi : \Re^n \mapsto \mathcal{H} \;\; \text{K = Kernel function}$$

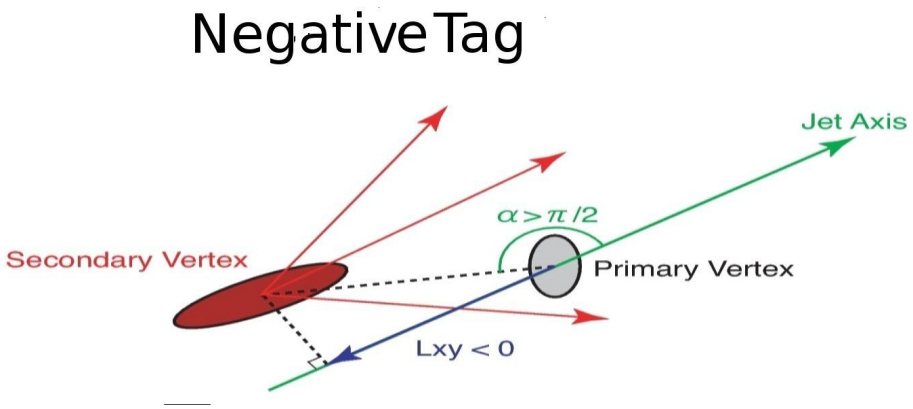
Work performed in collaboration with V. Lippi, PhD student of the Scuola Superiore S. Anna.

Software used: LibSVM C++ (by C.-C. Chang, C.-J. Lin).

(4) W+LF (Mistags) Background

Wrongly reconstructed tracks and long living LF hadrons can produce fake $b-$tags:





- symmetric distribution around primary vertices:
  $\Rightarrow$ negative vertices: opposite to the jet direction.
- per-jet fake probability parametrizated on a jet control sample.
- correction factor due to long living Light Flavor hadrons.

Full Background Estimate

Four independent channels:

- 2 Lepton categories, **Tight** and **EMC**:
  $\Rightarrow$ homogeneous kinematics.
- 2 Tag categories, **Single** and **Double** tagged jets:
  $\Rightarrow$ homogeneous background composition.

Channel	Single Tag		Double Tags	
	Tight	EMC	Tight	EMC
$t\bar{t}$	366 ± 37	172 ± 17	75 ± 11	34 ± 5
Single Top $s+t$	224 ± 25	87 ± 9	33 ± 5	13 ± 2
Z + jets	98 ± 13	65 ± 8	4.1 ± 0.6	3.1 ± 0.4
QCD	223 ± 89	101 ± 40	12.8 ± 5.6	0.0 ± 0.5
W + $b\bar{b}$	1135 ± 455	309 ± 124	148 ± 60	45 ± 18
W + $c\bar{c}$	582 ± 235	164 ± 66	8 ± 3	3 ± 1
W + cj	463 ± 187	106 ± 43	6.6 ± 2.7	1.8 ± 0.7
Mistag	1070 ± 124	346 ± 39	5.7 ± 1.2	2.2 ± 0.4
ZZ	1.51 ± 0.15	0.86 ± 0.08	0.27 ± 0.04	0.16 ± 0.02
WW	123 ± 17	38 ± 5	1.0 ± 0.2	0.33 ± 0.08
WZ	41 ± 4.5	14 ± 2	7.3 ± 1.1	2.4 ± 0.4
Prediction	4325 ± 895	1405 ± 243	301.6 ± 68.5	105.3 ± 21.5
Observed	4168	1318	260	106

Likelihood Definition

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

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