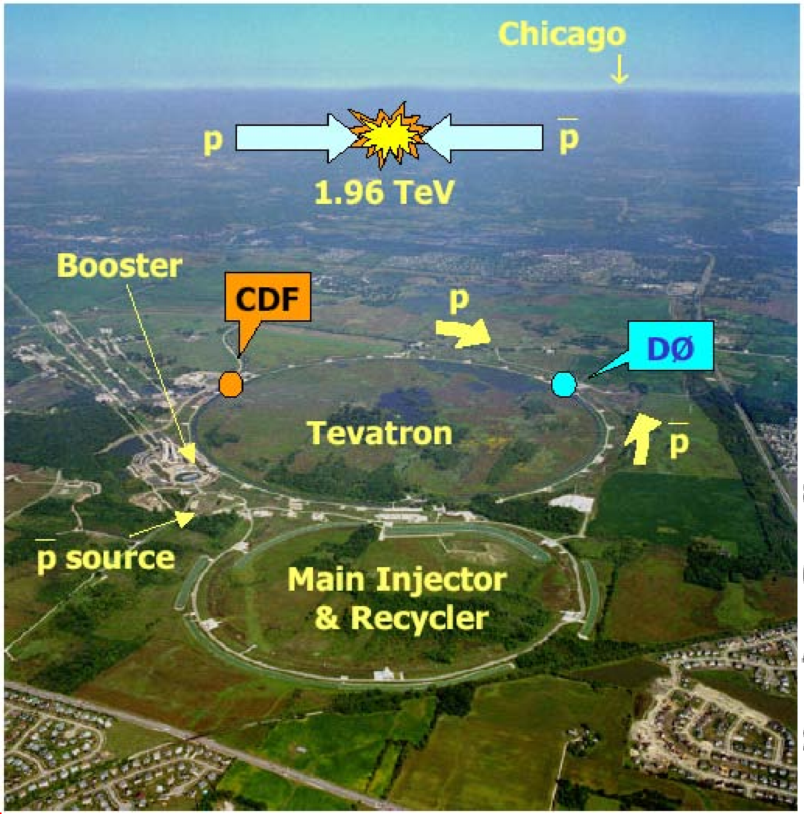
Measurement of WW and WZ production in the lepton plus heavy flavor jets final state at CDF

Sandra Leone
(INFN Pisa)
on behalf of the CDF Collaboration

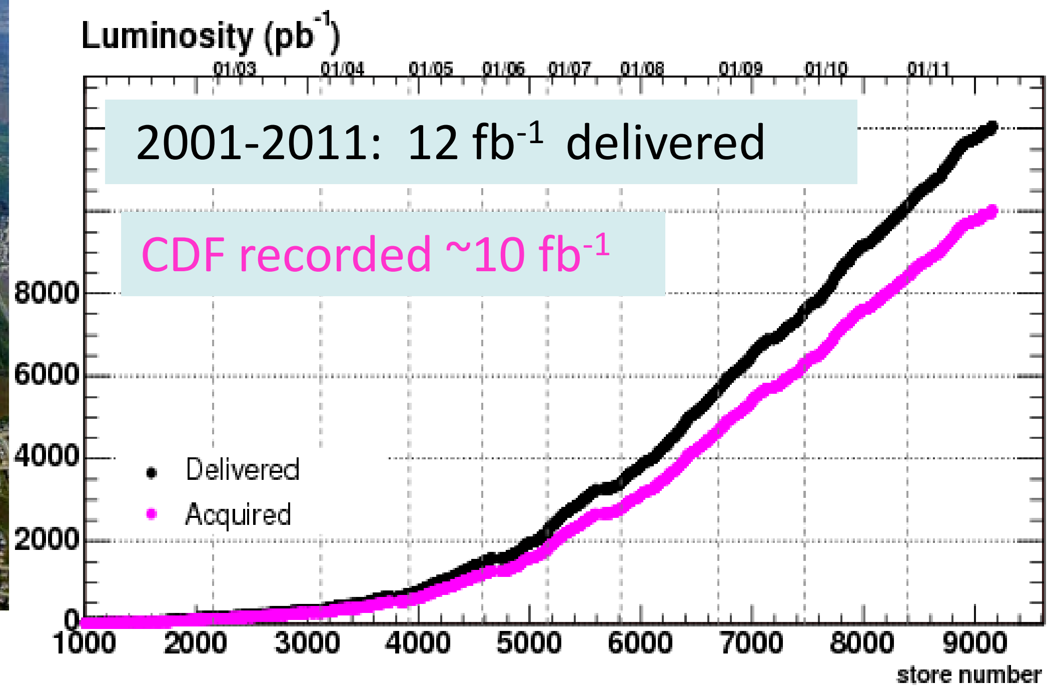
ICHEP2016CHICAGO

ICHEP 2016 Chicago, August 3-10

The Tevatron and CDF



Run II: √s = 1.96 TeV
Tevatron stopped providing p-pbar collisions on september 30, 2011



Among other interesting Standard Model results, Tevatron experiments took actively part to the hunt for the Higgs boson

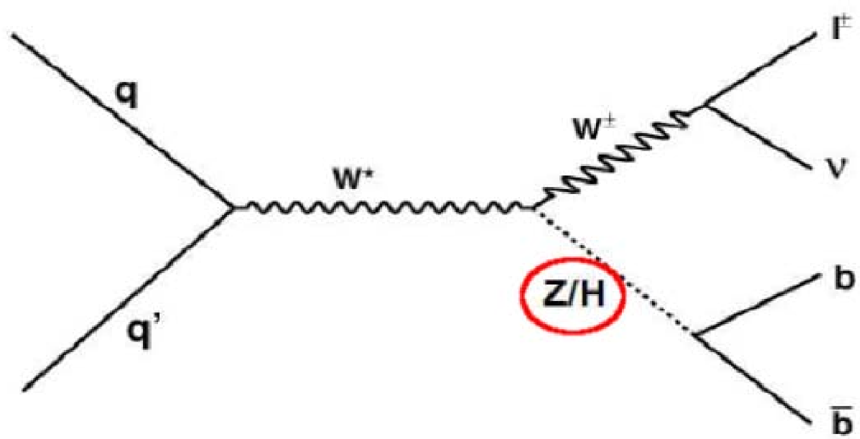
Motivation for Diboson Search in $l\nu$ + heavy-flavor jets

- Diboson production is a theoretically well known process
- Probe of SM couplings $\Rightarrow$ a significant excess would open a window on new physics
- Often used as "benchmark" of experimental sensitivity to rare processes, in a variety of final states

$\Rightarrow$ one of the ways in which the Higgs was hunted is through its associated production with W bosons:

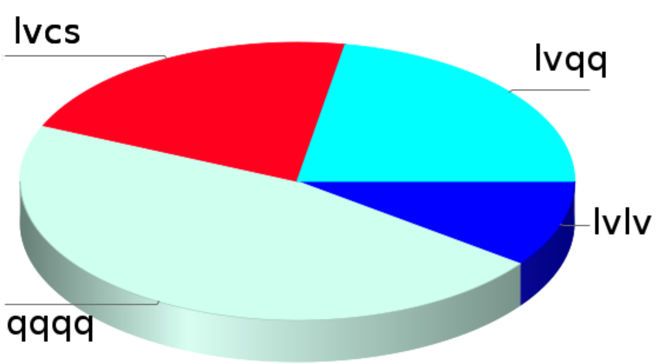
➢ WH $\rightarrow$ l ν + bb and WZ $\rightarrow$ l ν + bb share same final state

PRL 109 (2012) 111802

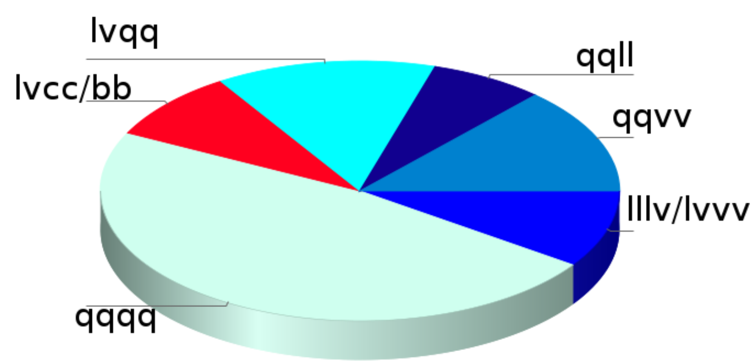


Motivation for Diboson Search in lν + heavy-flavor jets

- Leptonic final states:
  - ⇒ Clean signature, low background, small BR
  - ⇒ Measured with good precision at LHC and Tevatron
- Semi-leptonic final states:
  - ⇒ Experimentally challenging both at Tevatron and LHC
  - ⇒ Large non-resonant background: V+jets QCD production
  - ⇒ Poor di-jet mass resolution: no W-to-Z separation
- No precise measurement of WZ in semi-leptonic final state:
  - ⇒ Feasible using WW and WZ heavy-flavor (HF) decays



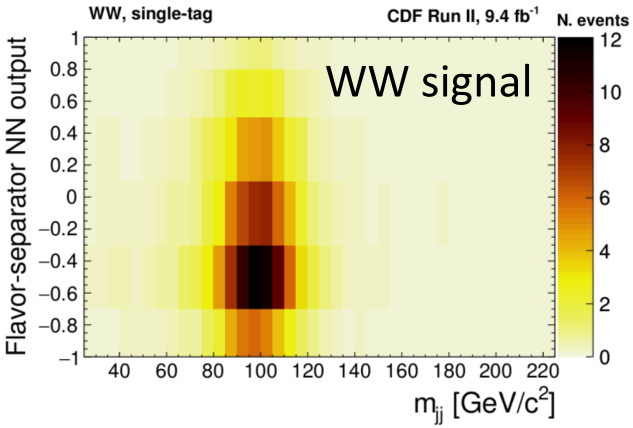
WW final state BRs

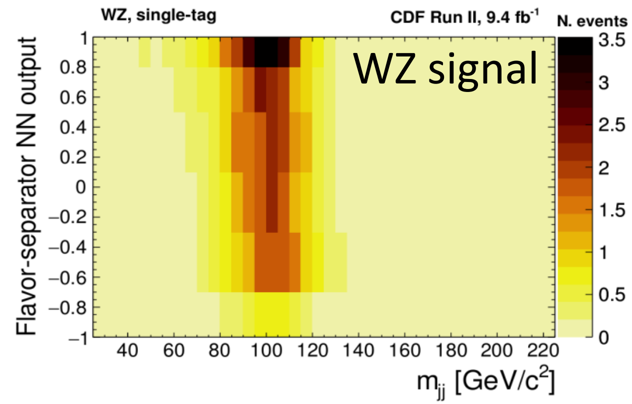


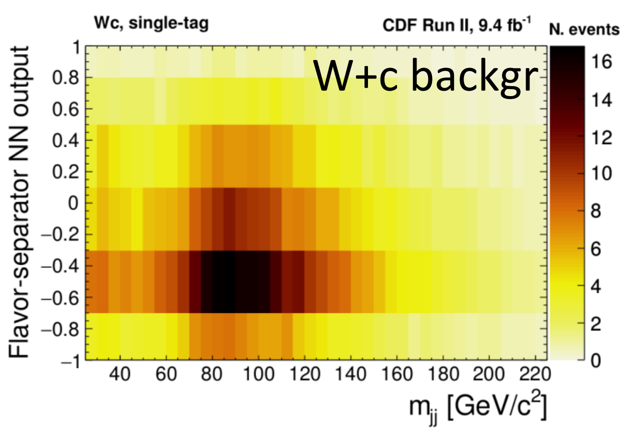
WZ final state BRs

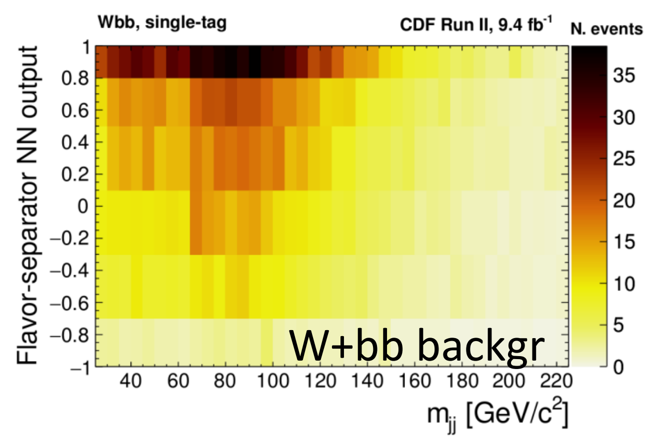
Analysis Strategy

- High-acceptance lepton-plus-two-jets selection (similar to single-top and WH):
  - $\Rightarrow$ Support Vector Machine discriminant used to suppress multi-jet (MJ) background
- Secondary-vertex jet tagging to enrich sample in HF and reduce W+jets background
- Search of a peak over large non-resonant background $\Rightarrow$ use m_{jj} as discriminant
- WW $\rightarrow l\nu$+cs versus WZ $\rightarrow l\nu$ +cc/bb $\Rightarrow$ 1-tag vs 2-tag and Flavor-separator NN









Event Selection

Full Run II data set: L =9.4 fb^{-1} with 6% uncertainty

- **Trigger: 4 event categories:**
  - ⇒ Central electrons
  - ⇒ Forward electrons
  - ⇒ Central muons
  - ⇒ Extended muons (using MET+jets)

- **Offline lepton selection:**
  - ⇒ Exactly 1 e/μ candidate: E_T (P_T) > 20 GeV
  - ⇒ Use of 10 lepton-ID classes:
    - ✓ central and forward electrons
    - ✓ 2 tight central muons
    - ✓ 5 loose muons
    - ✓ isolated tracks

- **Selection of pretag control sample:**
  - ⇒ Exactly two jets: $E_T > 20$ GeV, $|\eta| < 2{:}0.$
  - ⇒ $\not{E}_T > 15$ GeV,

- **HF-tagged signal samples:**
- One-tag:
  - ⇒ 1 jet tagged by SecVtx-tight
- Two-tags:
  - ⇒ Both jets tagged by SecVtx-tight or SecVtx-loose working points

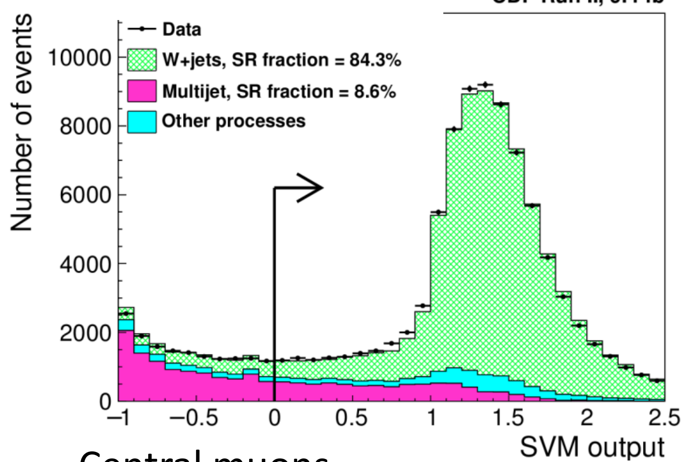
Background Estimate

- Main contributions from:
  - ⇒ W + heavy flavor, main source of irreducible background
  - ⇒ W + light flavor, mistakenly identified as a HF
  - ⇒ EWK: contributions from processes with a real lepton and HF jets
  - ⇒ Multi Jet (MJ): giving a boson-like signature and false missing ET
- Templates for EWK and W+jets backgrounds from simulation
- Normalization of the W+jets simulation determined in each lepton category using data before requiring b-tagging
- Data-driven models for MJ background:
  - ⇒ Muons: reverse Isolation cut (Iso > 0.1)
  - ⇒ Electrons: reverse at least 2 (out of 5) shower-id cuts (anti-e sample)
- MJ and W+jets template normalizations are left free in the fit

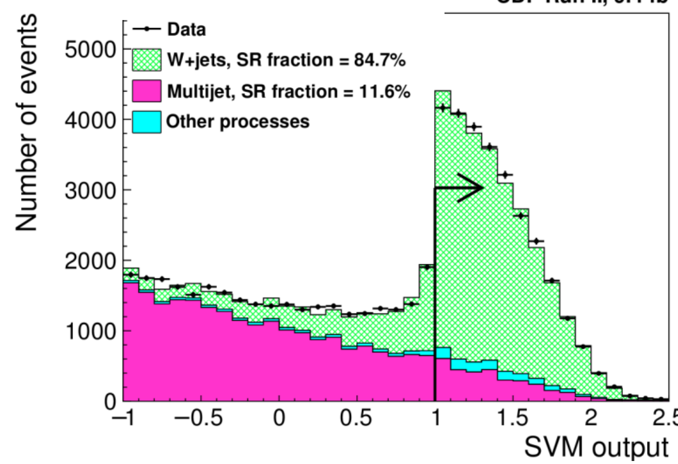
Multi-Jet Rejection and Normalization Extraction

- Data is superposition of multi-jet and W+jets contribution
- SVM multi-variate discriminant used for MJ rejection, and for templates of normalization fit

Central electrons

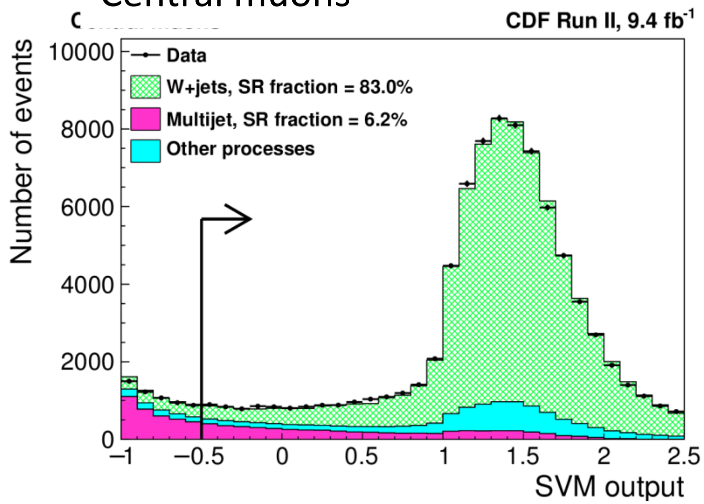


Forward electrons

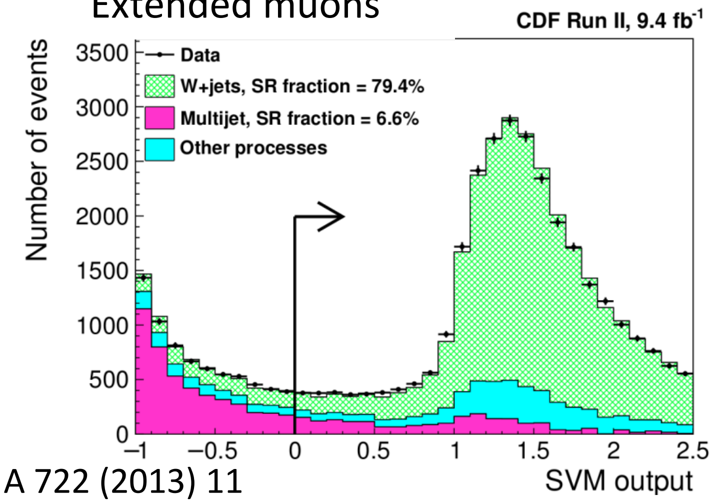


Arrows define the signal region

Central muons



Extended muons



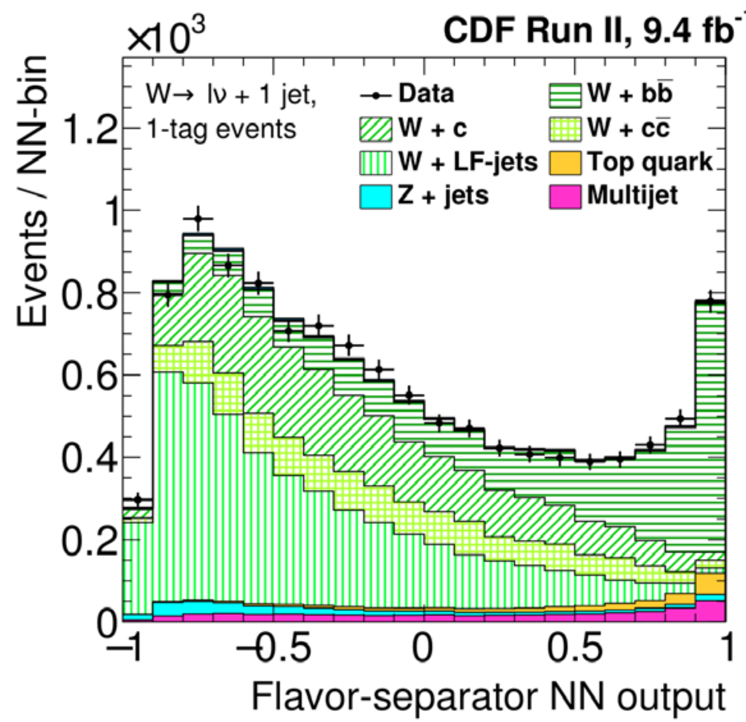
NIM A 722 (2013) 11

W + Heavy Flavor Estimate

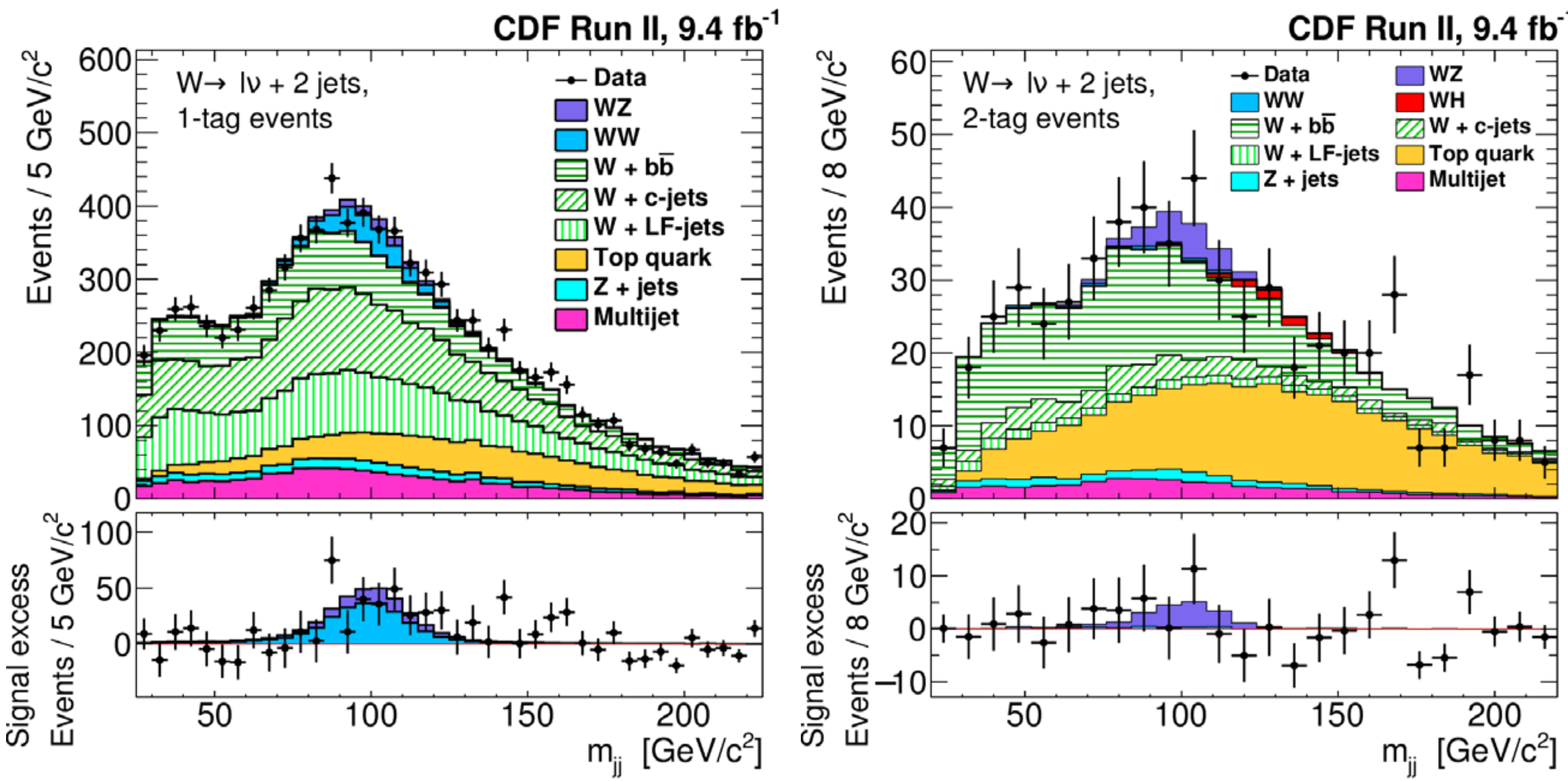
- Normalize W+jets yield to data
- Fraction of W+jets events with heavy flavor estimated from Alpgen MC
- Calibrate HF fractions using W+1 jet sample

$$N_{Wbb}^{data} = \left(\frac{N_{Wbb}}{N_{W+jets}}\right)^{MC} \cdot K_{HF} \, N_{W+jets}^{data}$$

$K_{cc} = K_{bb} = 1.24 \pm 0.25, \; K_c = 1.0 \pm 0.3$



Di-jet Invariant Mass Distributions

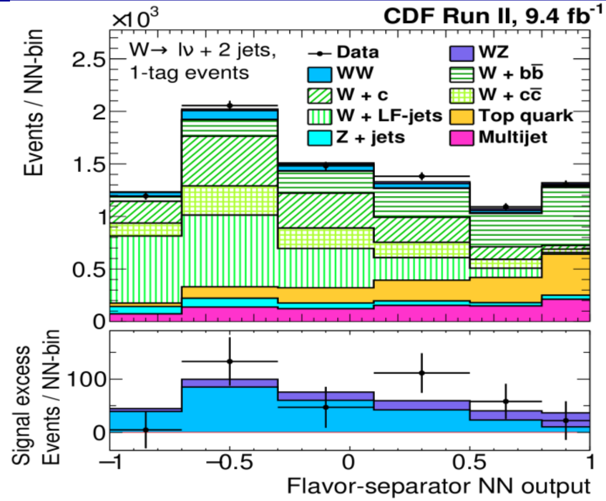


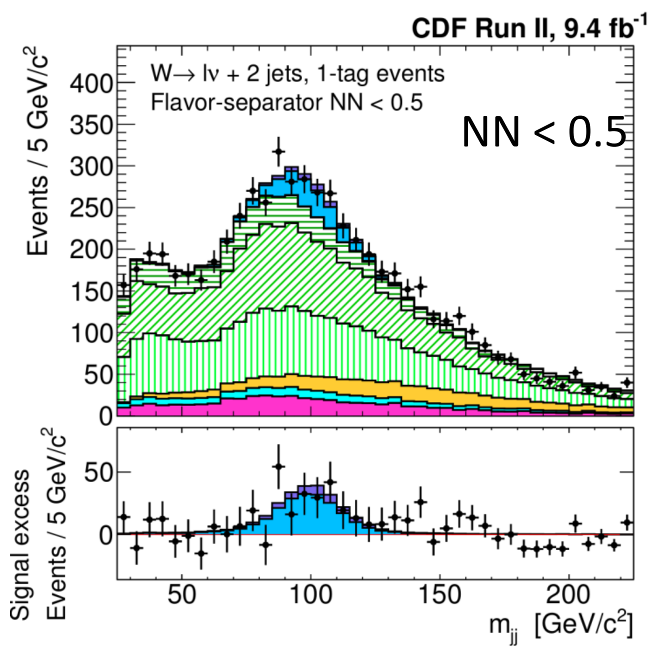
Single-tag events

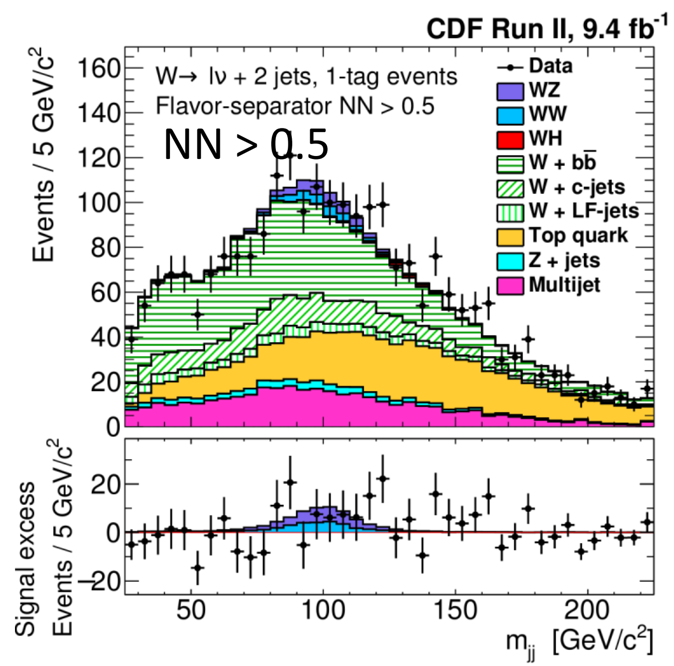
Double-tag events

Signal extraction using 2-dim variable

- Using flavour-separator NN to obtain b-quark versus c-quark separation
- 2-dimensional m_{jj} vs flavour-separator NN for single-tag events
- Different signal and background composition across NN values





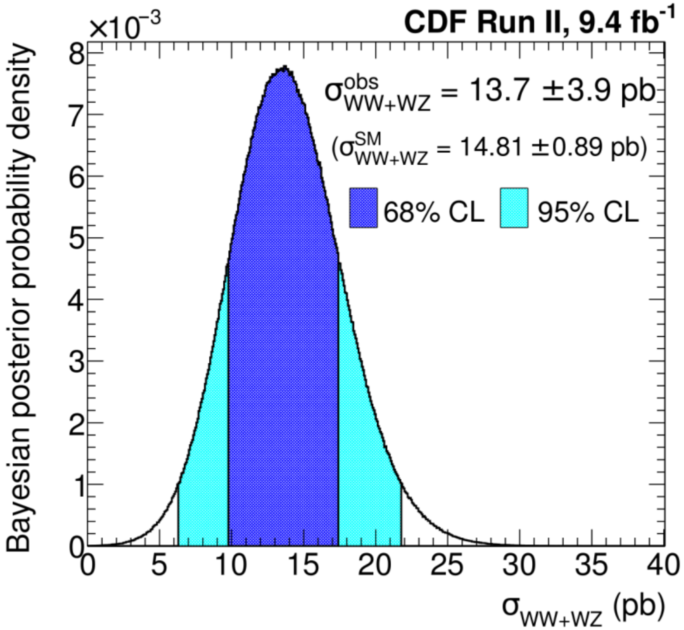


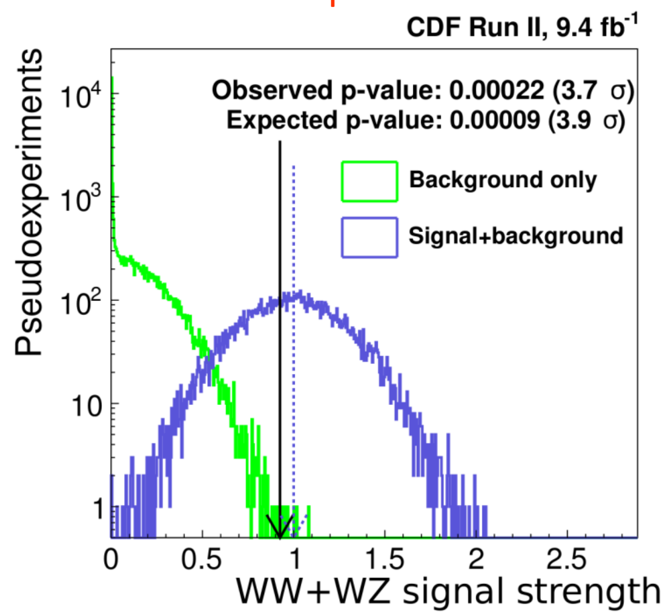
Measured WW+WZ Cross Section

- Likelihood function built with signal and background yield and shape predictions combining:
  - ⇒ Four lepton-analysis categories
  - ⇒ m_{jj} vs flavour-separator NN distributions for 1-tag
  - ⇒ m_{jj} distribution for 2-tag events

arXiv:1606.06823
Accepted by PRD

- Cross section extraction using Bayesian analysis:
  - ⇒ marginalize the posterior probability distribution over nuisance parameters





Posterior WW + WZ cross section:

$$\sigma^{obs}_{WW+WZ} = 13.7 \pm 2.4(\mathrm{stat}) \pm 2.9(\mathrm{syst}) = 13.7 \pm 3.9\ \mathrm{pb}$$

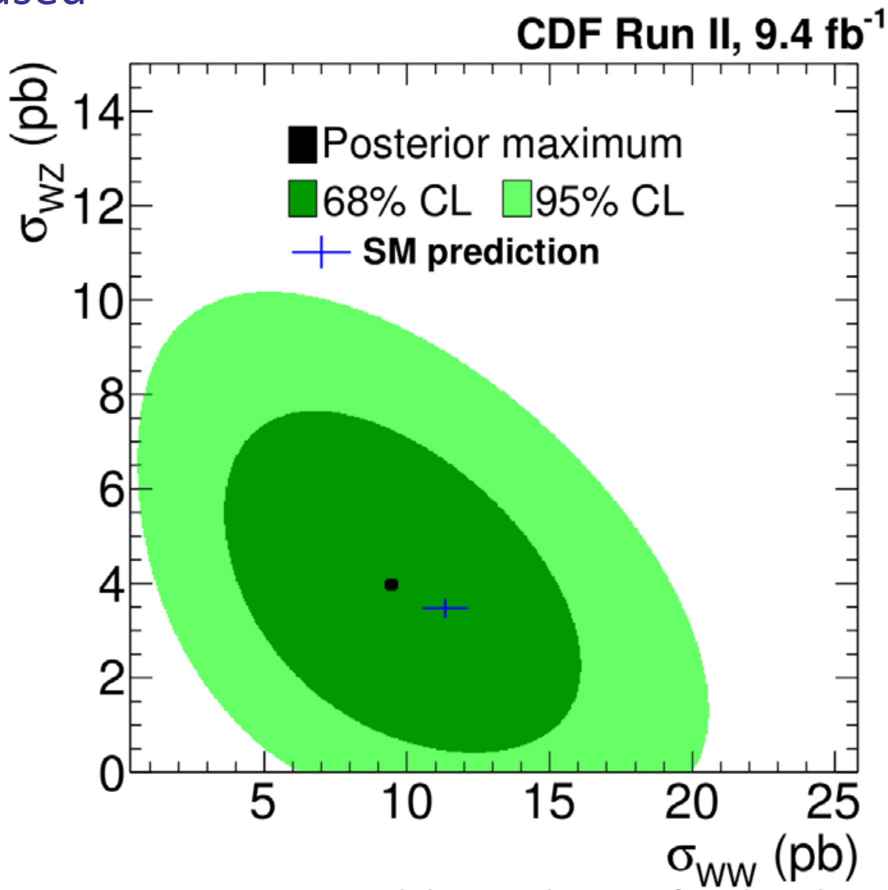
σ_{SM} = 14.8 ± 0.9 pb

Pseudoexperiments of background-only hypothesis compared to measurement:

WW + WZ signal significance: 3.7 σ
(3.9 σ expected)

Separate WW and WZ Cross Section Measurements

- WW vs WZ simultaneous signal extraction ⇒ 2-dimensional posterior distribution is used



- Measured cross sections compatible with SM for both WW and WZ signals detected in HF-enriched final state

WW and WZ Signal Results

- Measured integrating one or the other cross-section variable of 2-dim. posterior:

arXiv:1606.06823
Accepted by PRD

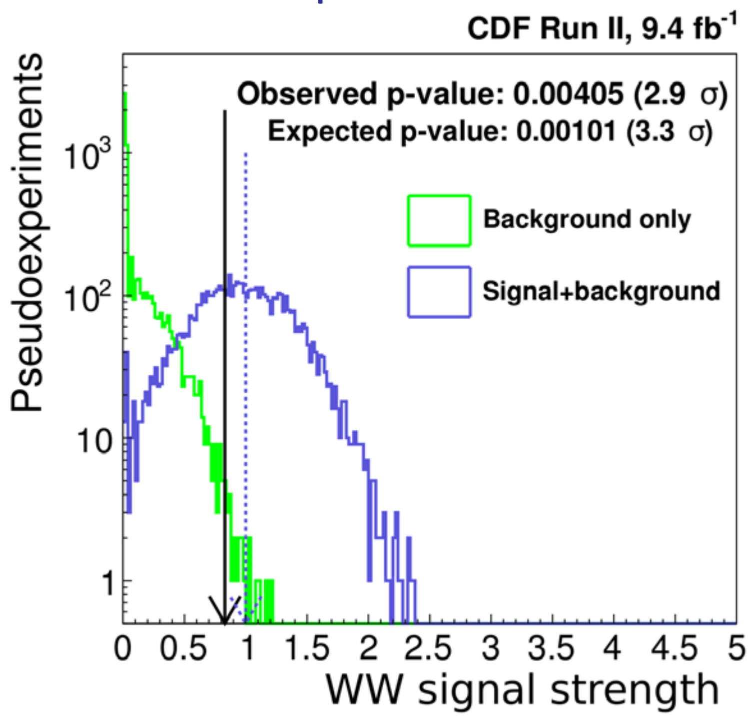
$$\sigma^{obs}_{WW} = 9.4^{+3.0}_{-3.0}(\text{stat})^{+2.9}_{-2.9}(\text{syst}) = 9.4 \pm 4.2 \text{ pb}$$

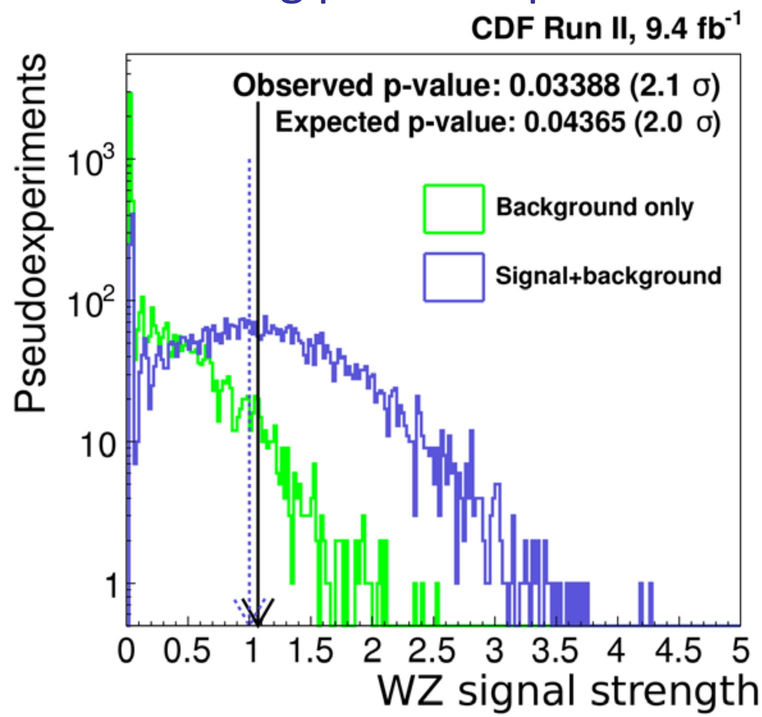
$$\sigma^{obs}_{WZ} = 3.7^{+2.0}_{-1.8}(\text{stat})^{+1.4}_{-1.2}(\text{syst}) = 3.7^{+2.5}_{-2.2} \text{ pb}$$

σ_{WW}^{SM}=11.7±0.9pb

σ_{WZ}^{SM}=3.5±0.2pb

- Evaluation of separate WW and WZ significance using pseudoexperiments:





- Significance: 2.9 σ for WW (3.3 σ expected), 2.1 σ for WZ (2.0 σ expected)

Conclusions

- WW and WZ diboson production have been measured in a semi-leptonic final state enriched in HF jets on the full Run II CDF data set
- Di-jet invariant mass and a flavour-separator NN have been used to extract the total and separate WW and WZ signal cross sections
- Total diboson cross section measured with a precision of about 30%, comparable with other experiment measurements in semi-leptonic final-states
- Separate WW and WZ cross sections measured with a precision of 45% and 60% respectively, with WZ measurement being the most precise in this final state
- Almost 5 years after the shutdown of the Tevatron, still digging out interesting results from the Tevatron data!!!

Conclusions

- WW and WZ diboson production have been searched in a semi-leptonic final state enriched in HF jets on the full Run II CDF data set
- Di-jet invariant … have been used to extract the tota… …oss sections
- Total diboson cross sectio… …h a precision of about 30%, comparable with other experiment measurements in semi-leptonic final-states
- Separate WW and WZ cross sections measured with a precision of 45% and 60% respectively, with WZ measurement being the most precise in this final state
- Almost 5 years af… still digging out interesting resu…

Thank you!

Thanks to all the CDF collaborators and the Fermilab/Tevatron staff for many years of real fun!!

Backup

Event Selection

Full Run II data set: L =9.4 fb^{-1} with 6% uncertainty

- **Trigger strategy and 4 event categories:**
  - ⇒ Central electrons
  - ⇒ Forward electrons
  - ⇒ Central muons
  - ⇒ Extended muons (using MET+jets)
- Events selected by common trigger have homogeneous kinematic and background composition

- **Offline lepton selection:**
  - ⇒ Exactly 1 e/μ candidate: E_T (P_T) > 20 GeV
  - ⇒ Use of 10 lepton-ID classes:
    - ✓ central and forward electrons
    - ✓ 2 tight central muons
    - ✓ 5 loose muons
    - ✓ isolated tracks
- All leptons isolated in calorimeter, except iso-tracks isolated in tracking

- **Selection of pretag control sample:**
  - ⇒ Exactly two jets: E_T > 20 GeV, $|\eta|$ < 2:0.
- Jet energy corrections: JES and Quark-Gluon response in MC
- $\not{E}_T$ > 15 GeV, corrected for muon track, JES, primary-vertex

- **Definition of HF-tagged signal samples:**
- One-tag:
  - ⇒ 1 jet tagged by SecVtx-tight
- Two-tags:
  - ⇒ Both jets tagged by SecVtx-tight or SecVtx-loose working points

Background and Signal Yield Estimates

Process	Pretag	one-tag	two-tag
MJ	$18\,100 \pm 2700$	800 ± 330	30 ± 14
W+LF	$161\,700 \pm 3700$	2440 ± 350	29.5 ± 6.8
$W + c\bar{c}$	$13\,400 \pm 1700$	1190 ± 290	33 ± 16
$W + c$	$11\,600 \pm 2200$	930 ± 310	12.5 ± 5.5
$W + b\bar{b}$	6370 ± 930	2190 ± 520	313 ± 125
Z + jets	9400 ± 1900	281 ± 42	13.5 ± 2.1
$t\bar{t}$	1600 ± 230	663 ± 94	137 ± 22
Single-top ($s + t$ channels)	1109 ± 42	441 ± 23	70.8 ± 8.4
ZZ	93.4 ± 4.4	10.1 ± 0.7	2.0 ± 0.3
WH+ZH	40.0 ± 1.4	17.6 ± 0.8	5.4 ± 0.6
WW	5530 ± 400	240 ± 30	3.0 ± 0.7
WZ	904 ± 53	91.4 ± 7.6	17.2 ± 2.1
Total prediction	$229\,900 \pm 5800$	9300 ± 1200	670 ± 140
Observed data	$232\,145$	9074	604

- Rate systematics uncertainties included in the table:

  $\Rightarrow$ theory, luminosity, trigger-efficiency, lepton-ID SF, HF K-factor, SecVtx SF, MJ rate
- Additional shape and rate systematics considered:

$\Rightarrow$ JES, ALPGEN Q^2, flavour-separator response to c-jets, Light Flavor jets and Multi-Jet events

Background Estimate

- Method developed for l+jets, HF-tagged analyses. The key points are:
- W+jet pretag normalization extracted from template-fit of multi-jet (MJ) vs W+jets:

$$N^{W+Jets} = N^{Data}_{Pretag}(1 - F^{MJ}) - N^{MC}$$

- W+jets line-shape and HF component from ALPGEN LO, multi-leg simulation
- Estimate of W + HF (W + bb, W + cc, W + c) normalization:

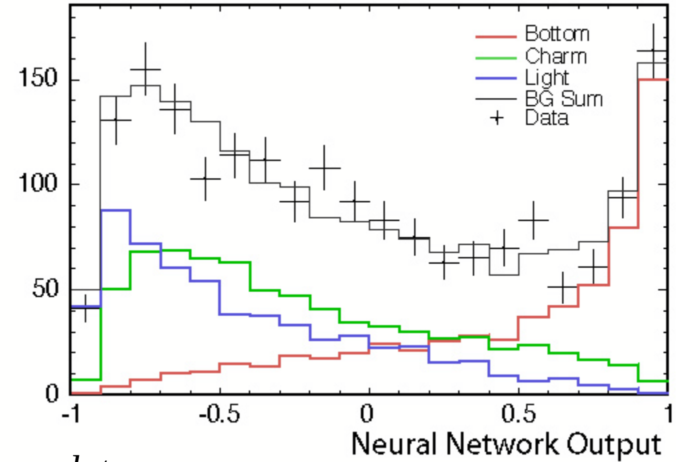
$$N^{HF} = N^{W+jets} \times f^{HF} \times \epsilon_{tag} \times K^{HF}$$

- $f^{HF} = \frac{W+HF}{W+jets}$: HF fractions derived from MC
- ϵ_{tag} : tagging efficiency derived from MC and corrected by per-jet SF_{tag}
- $K^{HF} = \frac{f^{HF}_{data}}{f^{HF}_{MC}}$: correction to HF production rate in MC:

$\Longrightarrow$ Extracted from W + 1 jet control sample

W + Heavy Flavor Estimate

- Normalize W+jets yield to data
- Fraction of W+jets events with heavy flavor estimated from Alpgen MC
- Calibrate HF fractions using W+1 jet sample



$$N_{Wbb}^{data} = \left(\frac{N_{Wbb}}{N_{W+jets}}\right)^{MC} \cdot K_{HF} \cdot N_{W+jets}^{data}$$

Correct data for non W+jets events

$$N_{W+jets}^{data} = N_{Candidates}^{data} - N_{non-W} - N_{EWK}$$

$$N_{EWK} = \sigma_{EWK} \cdot A \cdot L$$

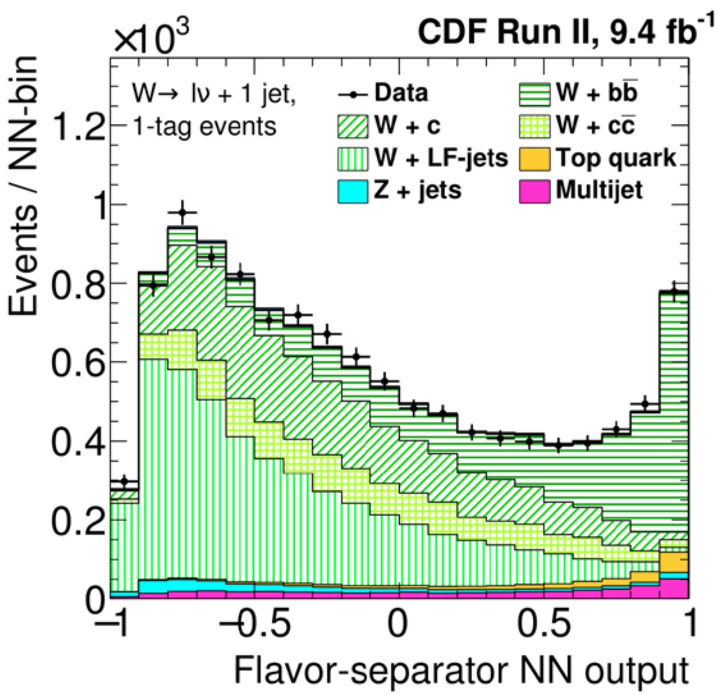
Heavy flavor fractions and b-tagging efficiencies from LO ALPGEN Monte Carlo

Calibrate ALPGEN heavy flavor Fractions by comparing W + 1jet data with ALPGEN Monte Carlo

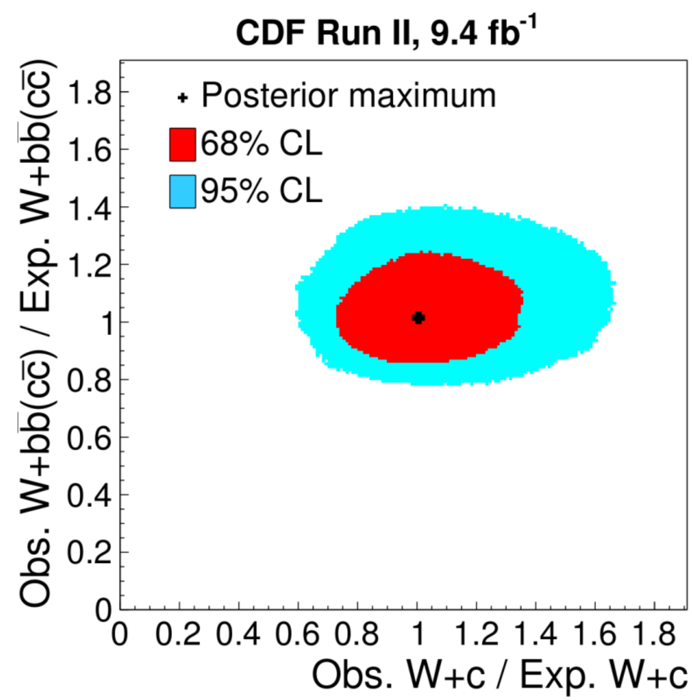
Evaluation of HF Correction Factors

- W + bb/cc and W + c K-factors extraction from W + 1 jet control sample:
- Analysis repeated for central tight leptons, 1-jet selection, pretag and 1-tag cat.
- Simultaneous extraction of Kcc = Kbb and Kc using flavour-separator NN
- Iterative measurement ⇒ Kcc = Kbb and Kc re-included in successive iterations

Flavor-separator NN for 1-jet events



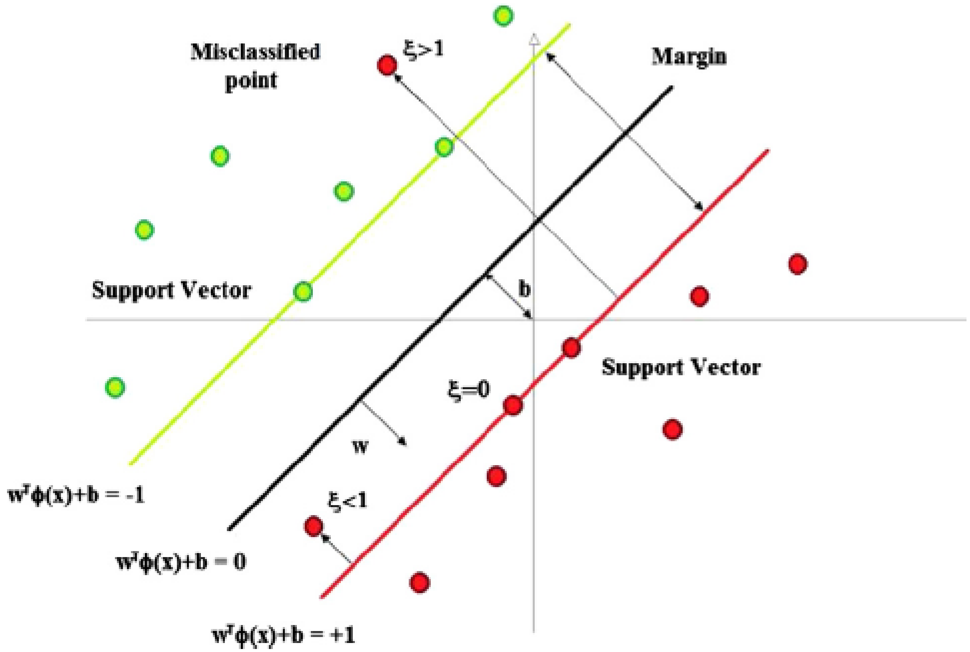
Last iteration of correction factors extraction



- **Result: Kcc = Kbb = 1.24, Kc = 1.0, 20% and 30% uncertainties respectively**

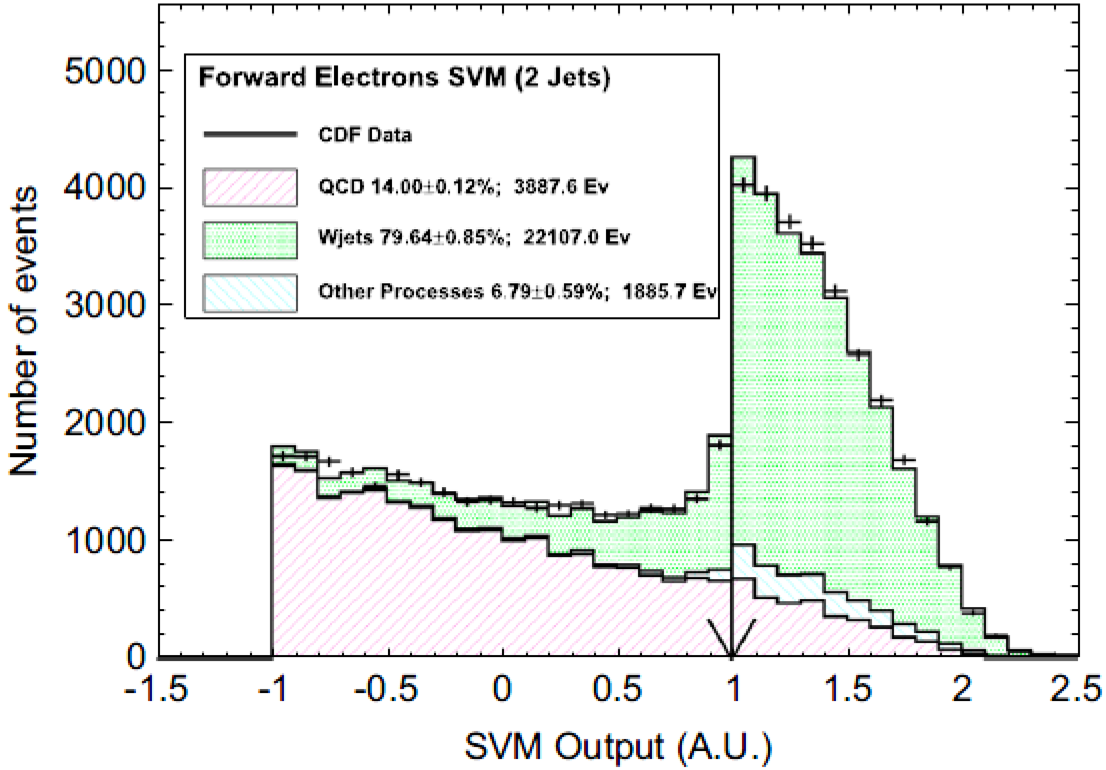
Support vector machine

How the linear classification problem can be formalized:



Two linearly separable classes of vectors are represented with red and blue dots.The plane leading to a maximum separation is defined by the weight vector w and the constant term b. NIM A 722 (2013) 11-19.

Multi jet background rejection in W+jets data sample



Contribution of the different physics processes to the shape of the SVM output distribution D used during the forward electron sample selection.The multi-jet background fraction (in magenta) is extracted from the fit together with the total W+ jets component (in green). The remaining physics processes are normalized to the expected production cross-sections. The SVM selection threshold for the final signal region identification is D=1

NIM A 722 (2013) 11-19

SVM input variables

All the possible input variables used for the SVM training and optimization.

Possible input variables							
1	p_T^{lep}	7	$E_T^{raw,jet1}$	13	$\Delta\phi(\not{p}_T, lep)$	19	$\Delta R(lep, jet2)$
2	$\not{E}_T$	8	$E_T^{raw,jet2}$	14	$\Delta\phi(\not{p}_T, \not{E}_T)$	20	$\Delta R(\nu^{min}, jet1)$
3	$\not{E}_T^{raw}$	9	$E_T^{cor,jet1}$	15	$\Delta\phi(\not{p}_T, \not{E}_T^{raw})$	21	$\Delta R(\nu^{min}, jet2)$
4	$\not{p}_T$	10	$E_T^{cor,jet2}$	16	$\Delta\phi(lep, \not{E}_T)$	22	$\Delta R(\nu^{min}, lep)$
5	M_T^W	11	$\Delta\phi(jet1, \not{E}_T)$	17	$\Delta\phi(lep, \not{E}_T^{raw})$	23	$\Delta R(\nu^{max}, jet1)$
6	$MetSig$	12	$\Delta\phi(jet2, \not{E}_T)$	18	$\Delta R(lep, jet1)$	24	$\Delta R(\nu^{max}, jet1)$

Final input variables used for the central and forward SVM multi-jet discriminant:

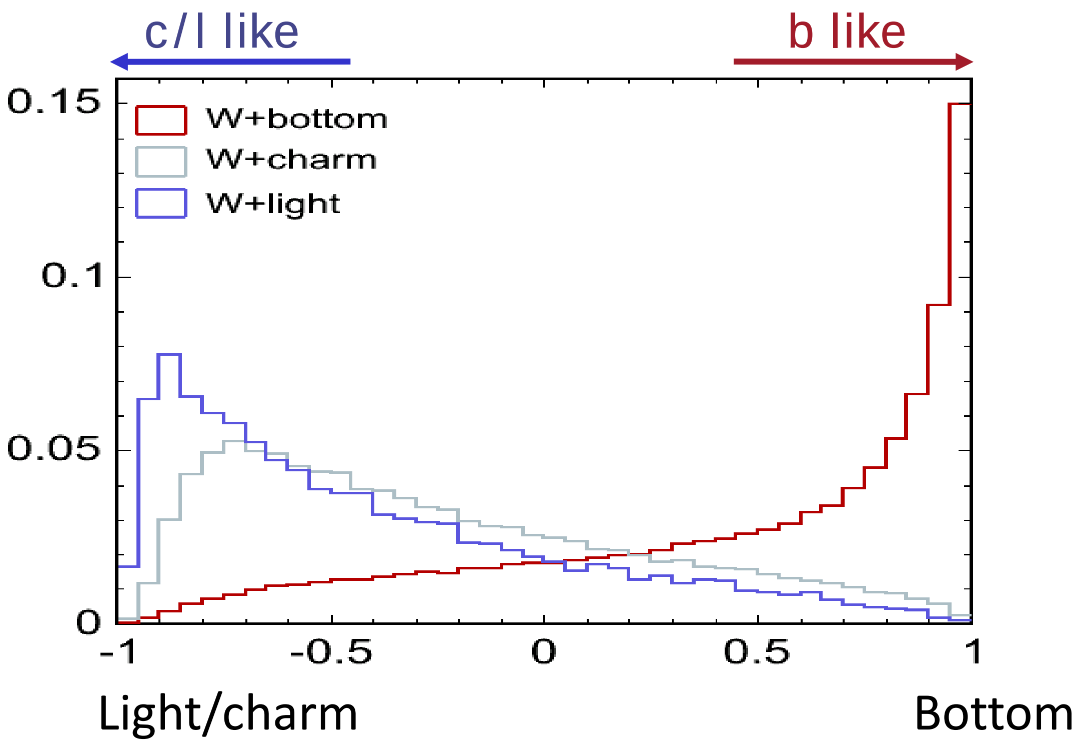
Final SVM input variables			
Central SVM:	M_T^W	$\not{E}_T^{raw}$	$\not{p}_T$
	$MetSig$	$\Delta\phi(\not{p}_T, \not{E}_T)$	$\Delta\phi(lep, \not{E}_T)$
	$\Delta R(\nu^{Min}, lep)$	$\Delta\phi(jet1, \not{E}_T)$	
Forward SVM:	M_T^W	$\not{E}_T^{raw}$	$\not{p}_T$
	$MetSig$	$\Delta\phi(\not{p}_T, \not{E}_T)$	$\Delta\phi(\not{p}_T, \not{E}_T^{raw})$

NIM A 722 (2013) 11-19.

Flavor Separator NN

- Developed for single top search
- Train Neural Network with jet and secondary vertex tracking information (25 input variables) for bottom/charm/light flavor separation

  $\Rightarrow$ Lxy, vertex mass, track multiplicity, impact parameter, semi-leptonic decay information, etc...
- Replaces Yes-No tag decision by a continuous variable *(0<b<1)*



Systematic uncertainties

- Affecting the rate:
  - ⇒ luminosity (6%)
  - ⇒ lepton acceptance (including trigger efficiencies, lepton reconstruction scale factors, from 2% to 5%)
  - ⇒ b and c tagging efficiency (from 3% to 10%)
  - ⇒ PDFs and radiative corrections (approx. 4%)
  - ⇒ Theory uncertainties on EWK backgrounds (from 5% to 40% for Z+jets)
  - ⇒ Mistag estimate (15% and 23% for single and double tag)
  - ⇒ W + HF fractions corrections (from 20% to 40%)
- Affecting both rate+shape
  - ⇒ Flavor separator NN
  - ⇒ Multi-Jet model
  - ⇒ Jet energy corrections
  - ⇒ W+jets Q^2