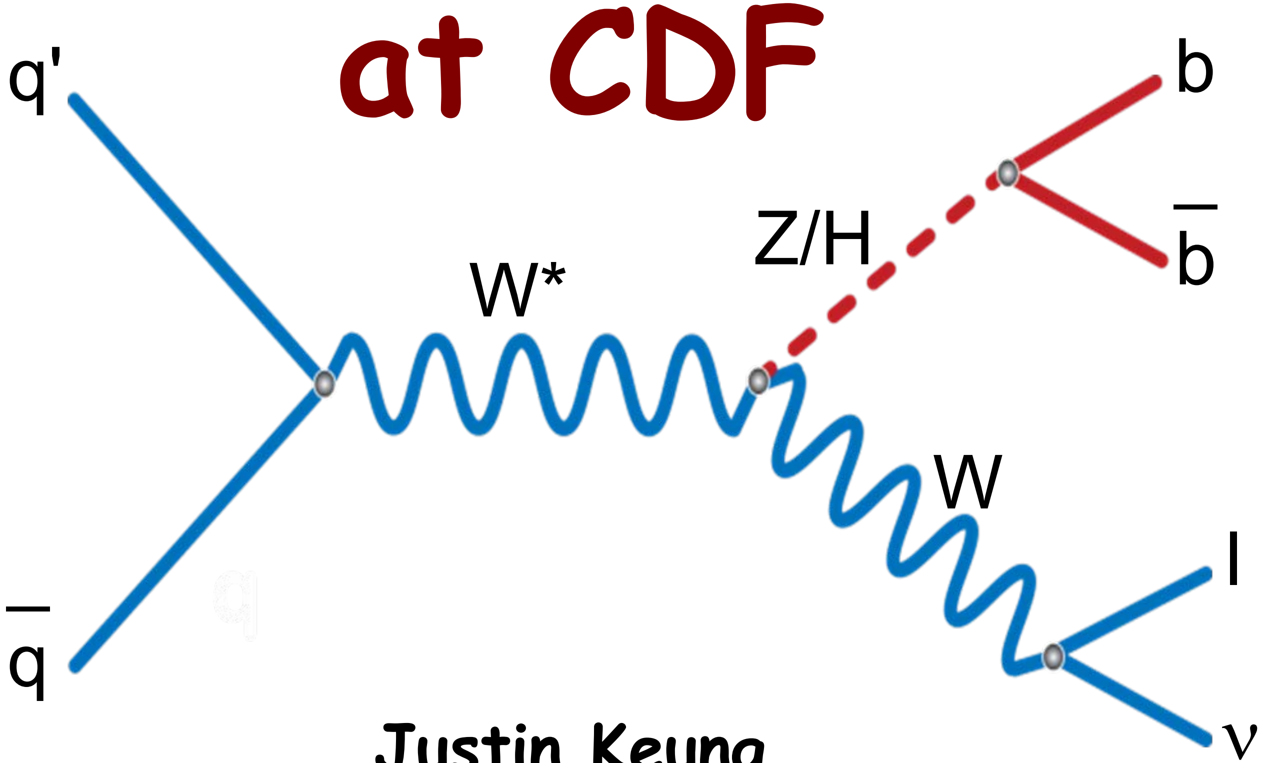
Search for WZ→lvbb at CDF



Justin Keung

University of Pennsylvania

November 24th, 2009

Experimental HEP Seminar @ University of Pennsylvania

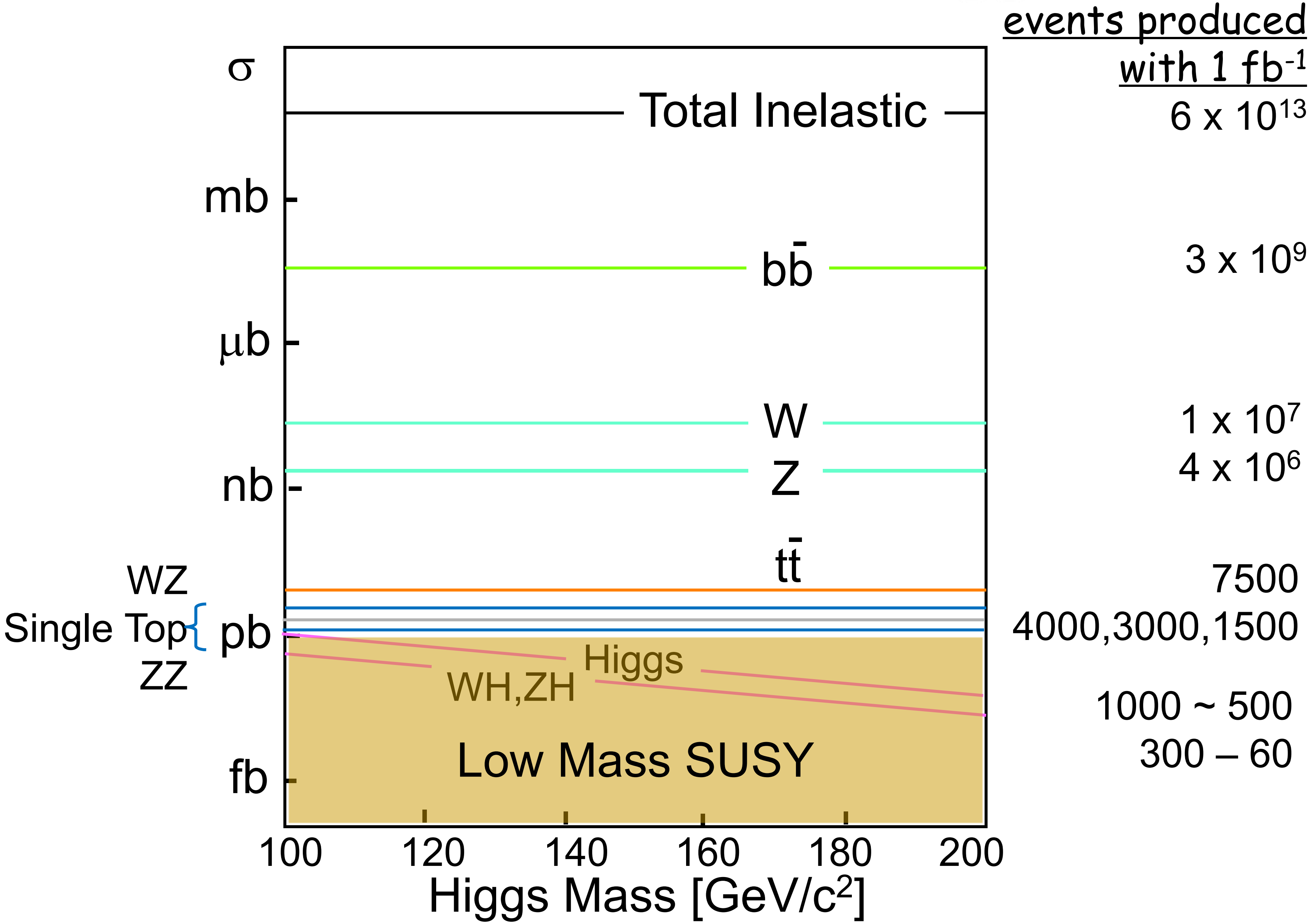
Outline

- Why WZ→lvbb ?
  - Higgs search effort
  - Fermilab, Tevatron and CDF

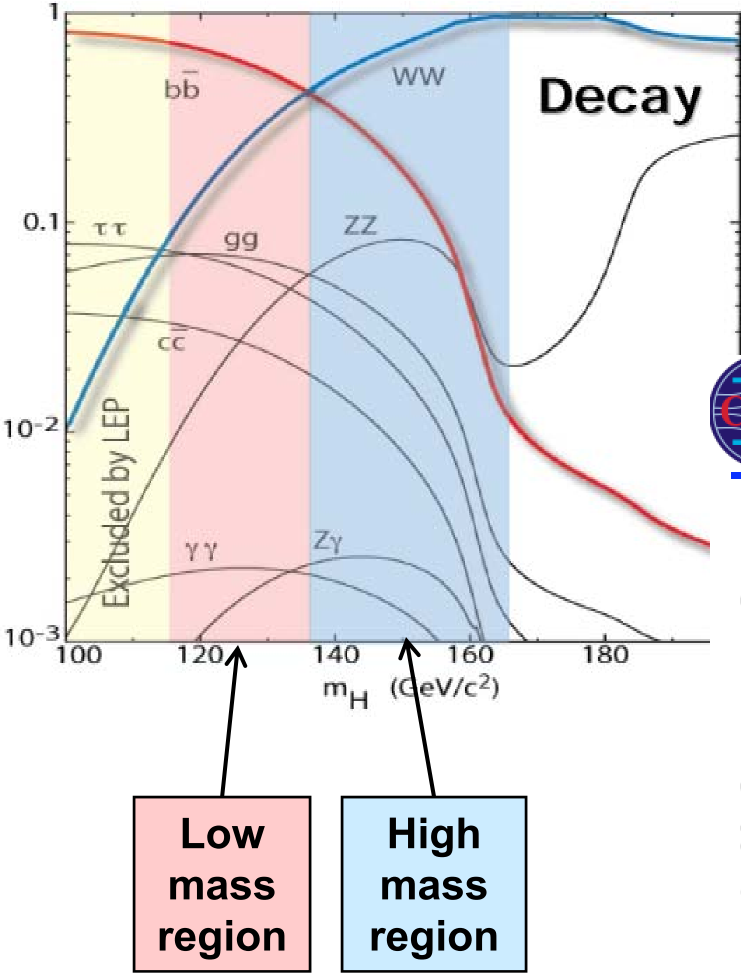
- **Improvements in b-jet identification**
  - Importance in the WZ/WH→lvbb search
  - Measurement of per-jet efficiency and background rate in data

- **Search for the WZ**
  - Background composition
  - Sensitivity estimates
  - Neural Network

Road to the Higgs



Searching for Higgs @Tevatron

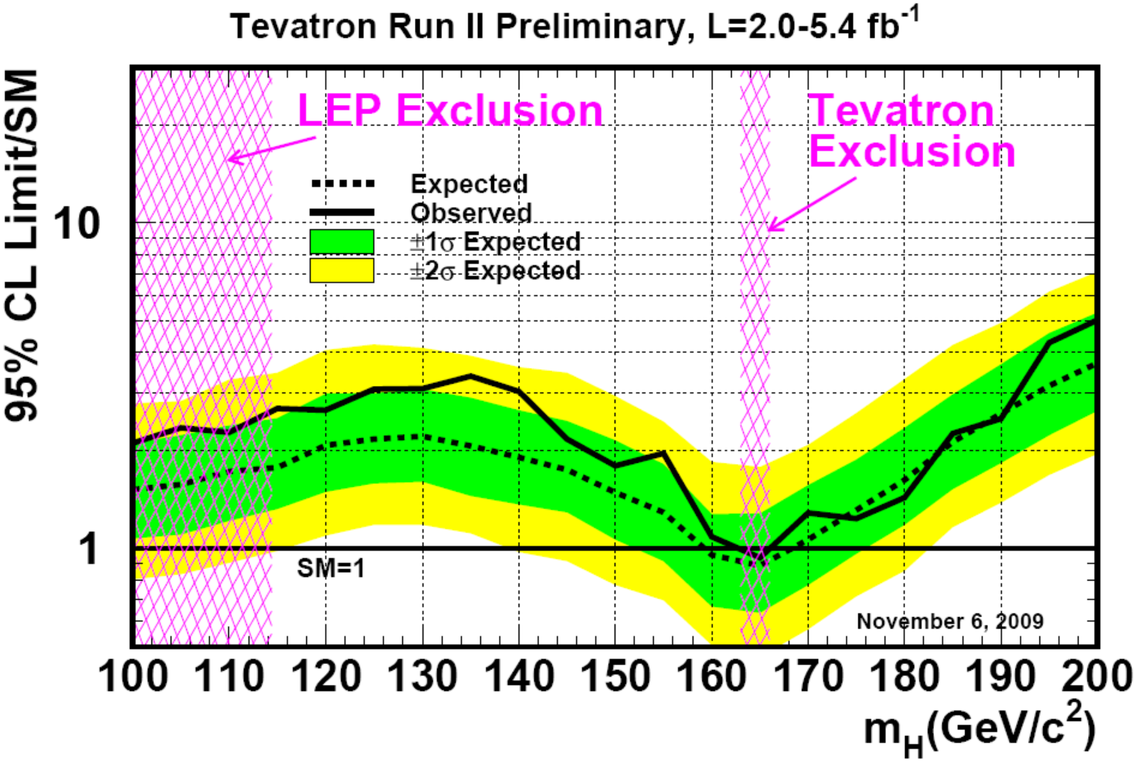


Low mass region

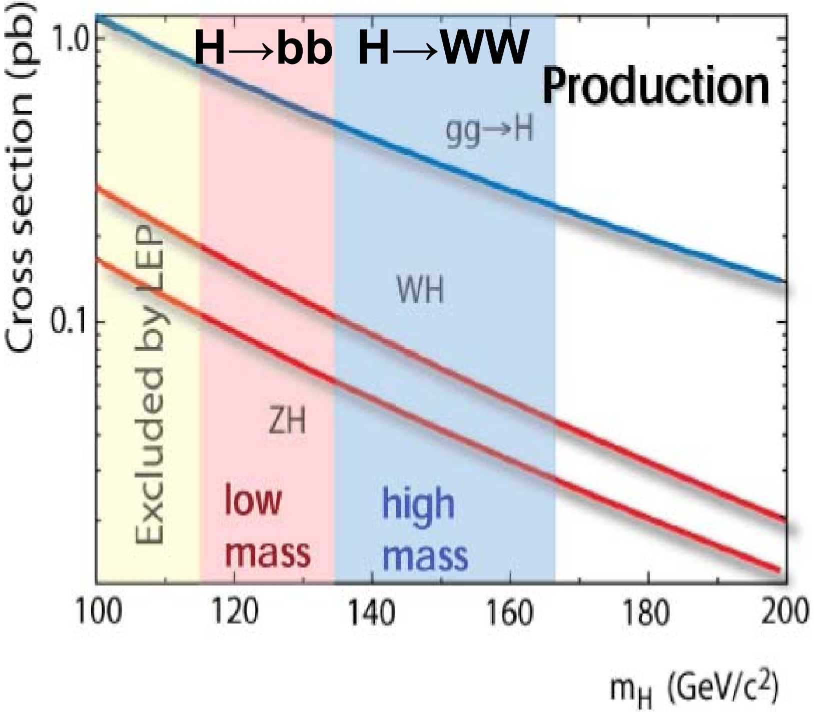
High mass region

- Tevatron needs more sensitivity at low mass region (H→bb)
- One important improvement: increase b identification power

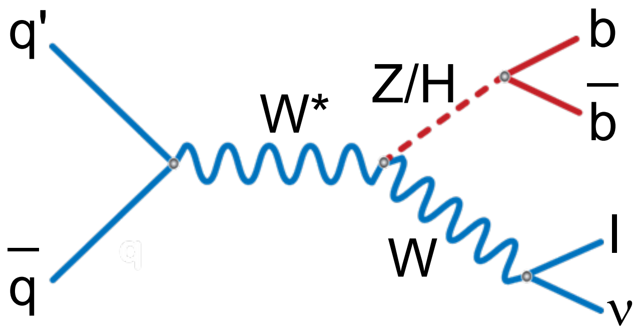
Tevatron Limits

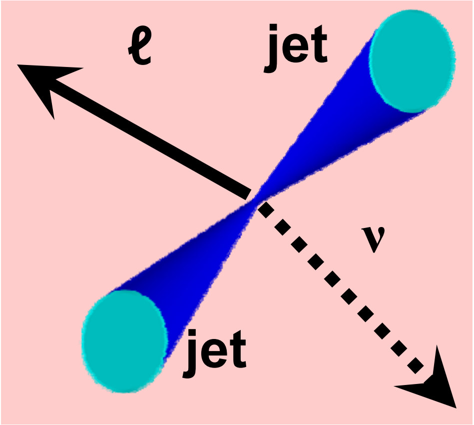


Higgs Production @Tevatron



- Direct production gg→H→bb swamped by huge (~10^7x) background QCD bb
- Associated production WH→lνbb reduces QCD background rate
  - Leptonic decay (into either electron or muon) gives a distinct signature for efficient triggering
  - Most sensitive process for low mass Higgs searches
- Z resonant production is similar to H resonant production





WZ→lvbb Search

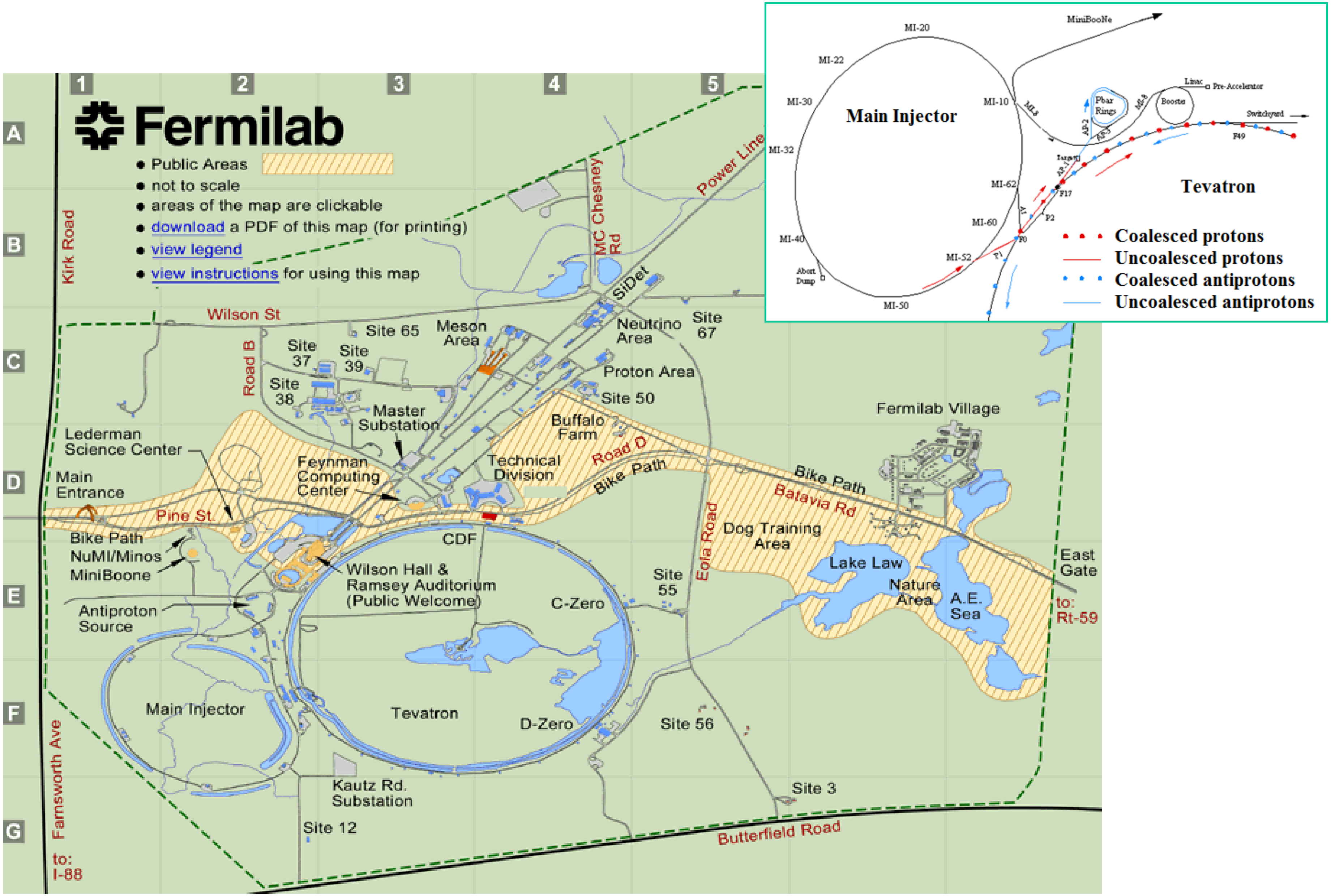
- WZ→lvbb is an excellent test of WH search tools
  - Same final state and similar topology
  - WZ→lvbb has effective cross section 4x higher than that predicted for WH→lvbb (H @ 120 GeV/c^2)

	WZ	WH (120 GeV/c^2)
Production cross section	3.96pb	0.16pb
W->lv (e or μ) branching fraction	0.21	0.21
Z/H->bb branching fraction	0.15	0.8
XSec x BR(W→lv) x BR(Z/H→bb)	0.12pb	0.03pb

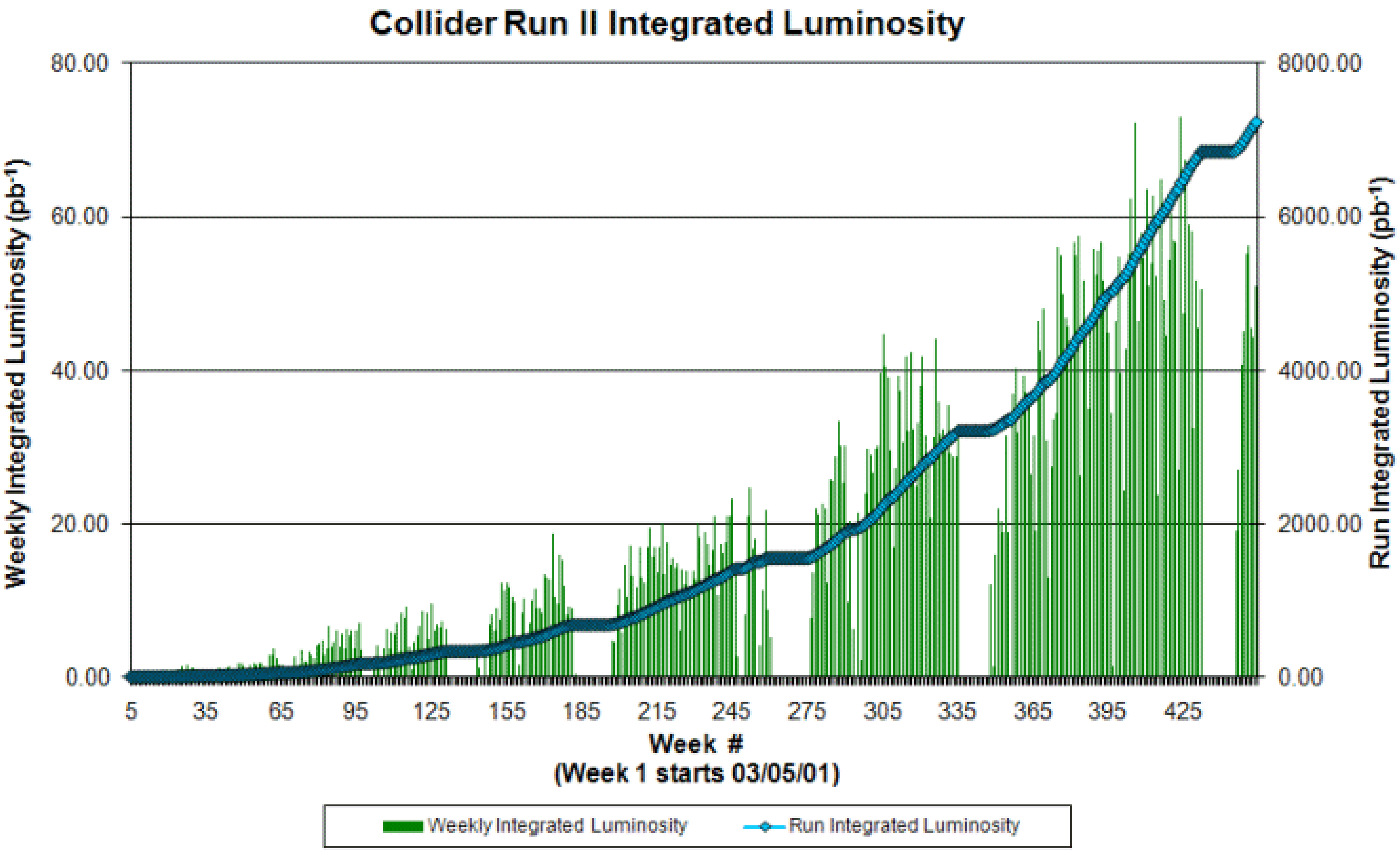
- Plan to set a limit/observe WZ production in order to test the b-jet identification tools and sophisticated search techniques used for WH

- CDF recently observed WW/WZ→lvjj without identifying the b-jet, with a signal significance of 5.4σ (16±3.3pb)

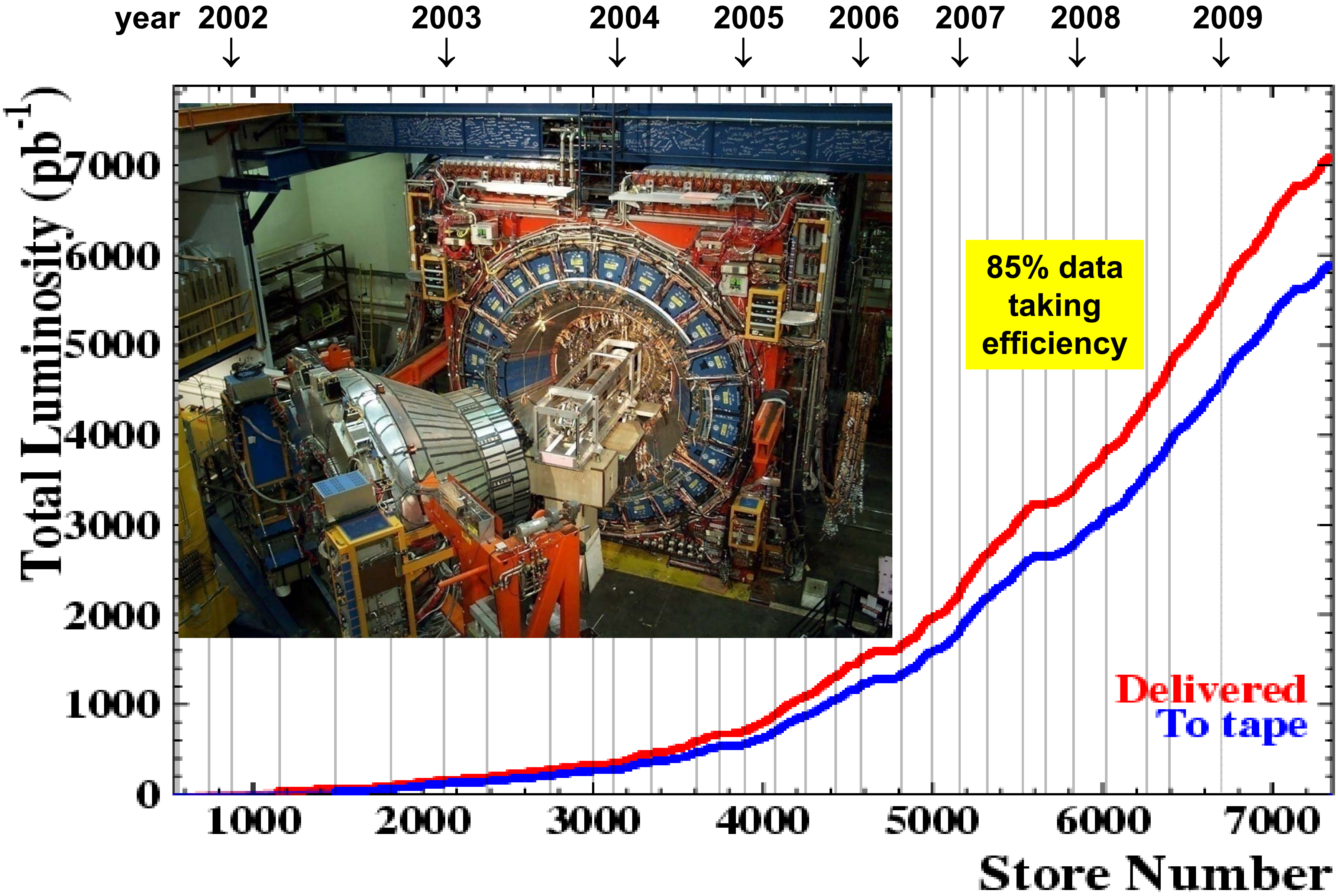
Fermilab at a Glance



Tevatron Performance



CDF Performance



85% data taking efficiency

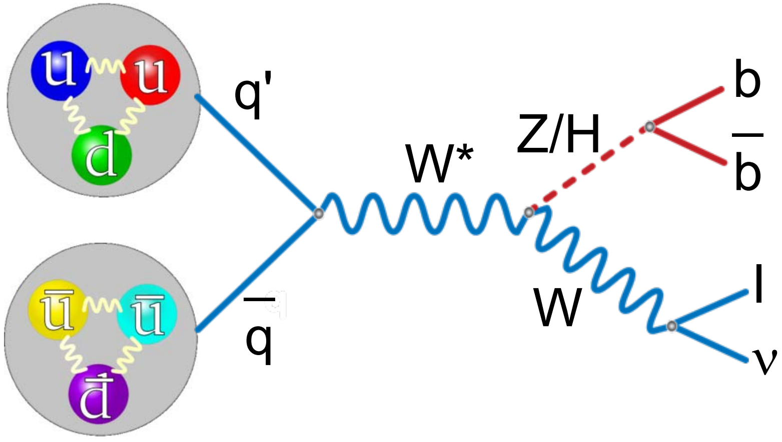
Outline

- Why WZ→lvbb ?
  - Higgs search effort
  - Fermilab, Tevatron and CDF

- **Improvements in b-jet identification**
  - Importance in the WZ/WH→lvbb search
  - Measurement of per-jet efficiency and background rate in data

- Search for the WZ
  - Background composition
  - Sensitivity estimates
  - Neural Network

WZ/WH Event Selection

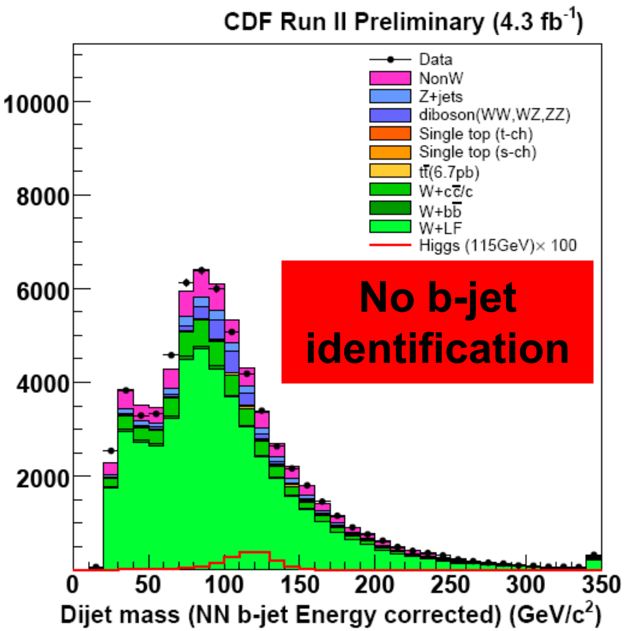


- **High p_T lepton**
  - $p_T > 20$ GeV
  - $|\eta| < 1.1$
- **Missing Transverse Energy**
  - MET > 20 GeV
- **Two *b*-jets**
  - $E_T > 20$ GeV
  - $|\eta| < 2.0$
  - Jet cone size 0.4
  - identification of b-jet

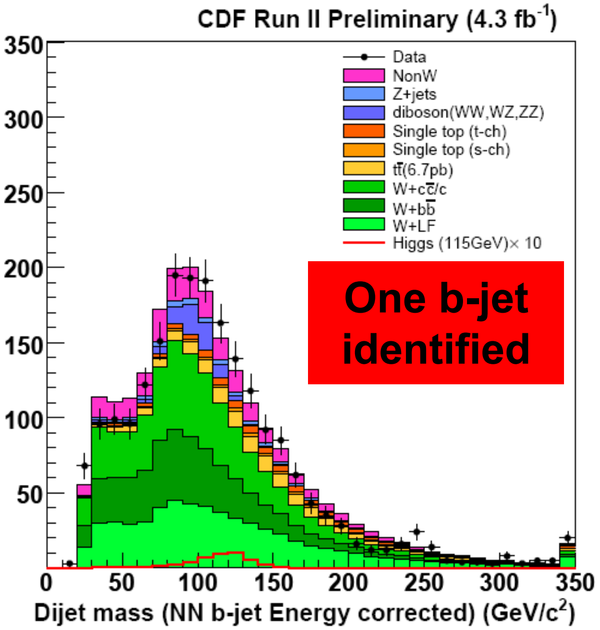
WZ	% efficiency	percent of initial
Fiducial Lepton	60.3	60.3
Lepton E_T>20GeV	84.2	50.8
Reconstructed & Identified	58.7	29.8
MET > 20GeV	88.3	26.3
2 jets, both E_T>20GeV and Fiducial	37.4	9.9
Both identified as b-jet	11.6	1.1

- My contribution: improving b-jet identification

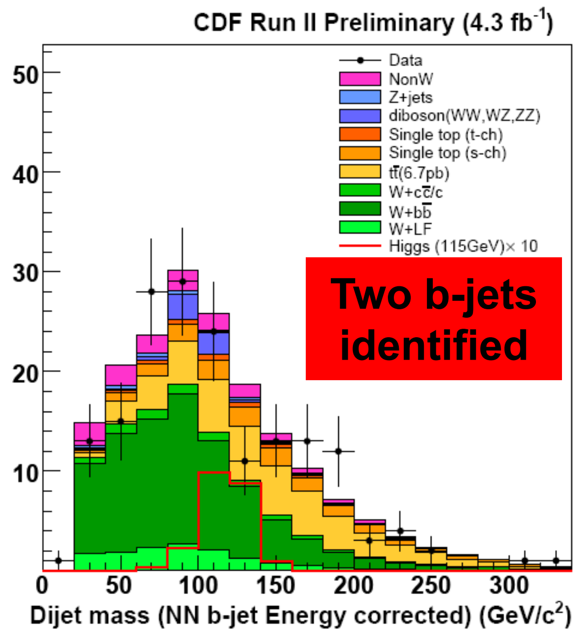
Why Identify the b-jets


CDF Run II Preliminary (4.3 fb^{-1})

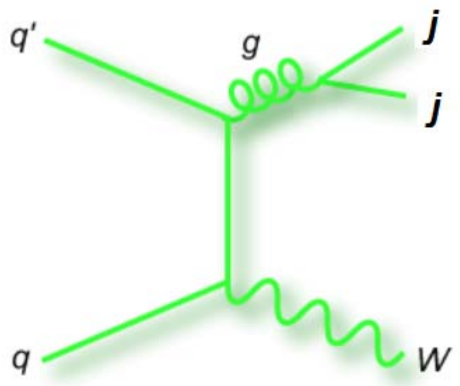
No b-jet identification


CDF Run II Preliminary (4.3 fb^{-1})

One b-jet identified


CDF Run II Preliminary (4.3 fb^{-1})

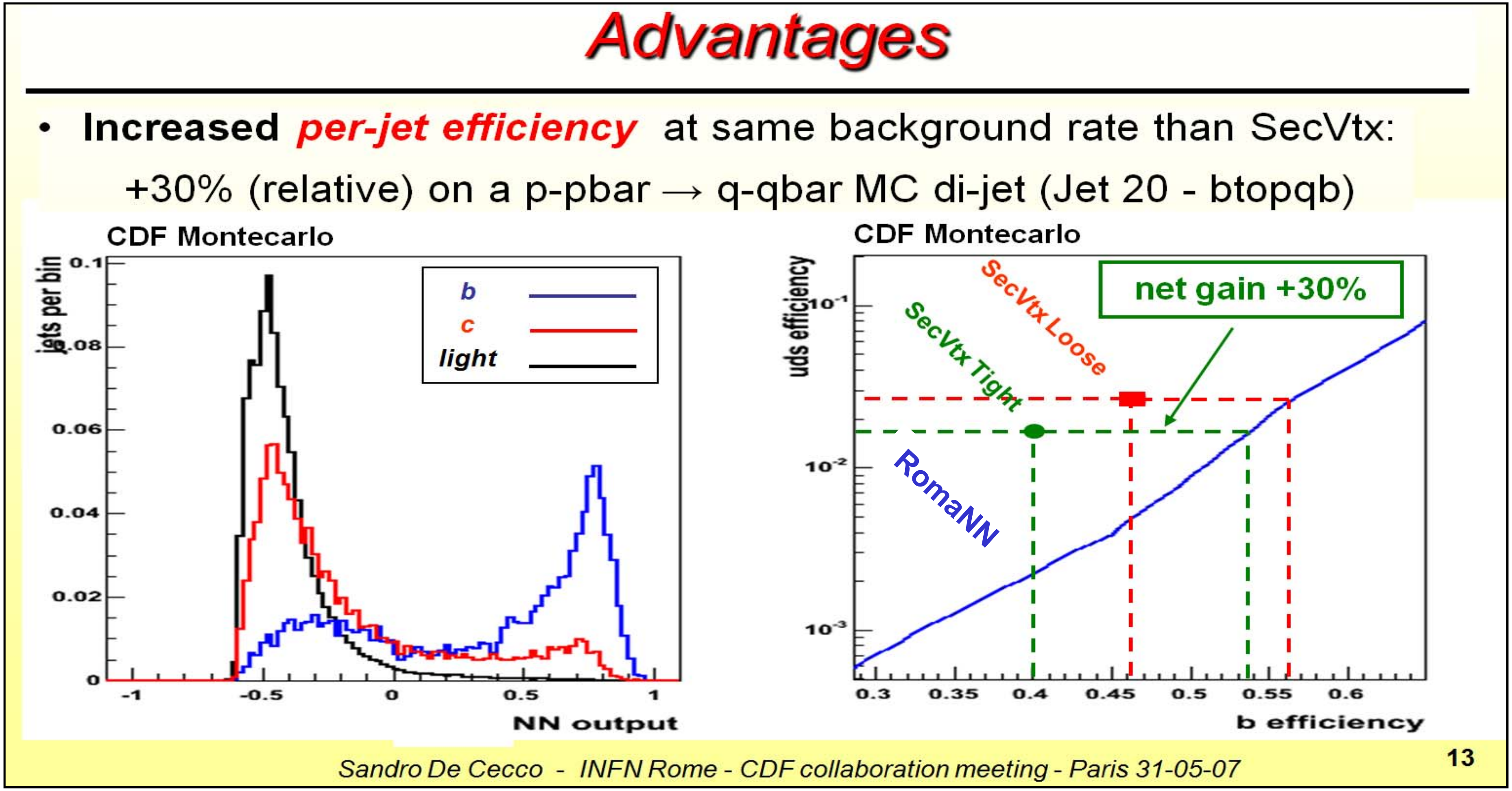
Two b-jets identified



- Identification of b-jets improves the signal to background ratio
- Important to improve b-jet identification efficiency to gain more signal in the best signal to background channel
  - If per-jet efficiency increases 10%, then the number of events with 2 identified b-jets increases 21%

New b-jet Identification Algorithm

- Roma Neural Network
- **Simulations** claims a 30% increase in per-jet efficiency for same background rate as default algorithm
- I measured this per-jet efficiency and background rate **in data**



Advantages

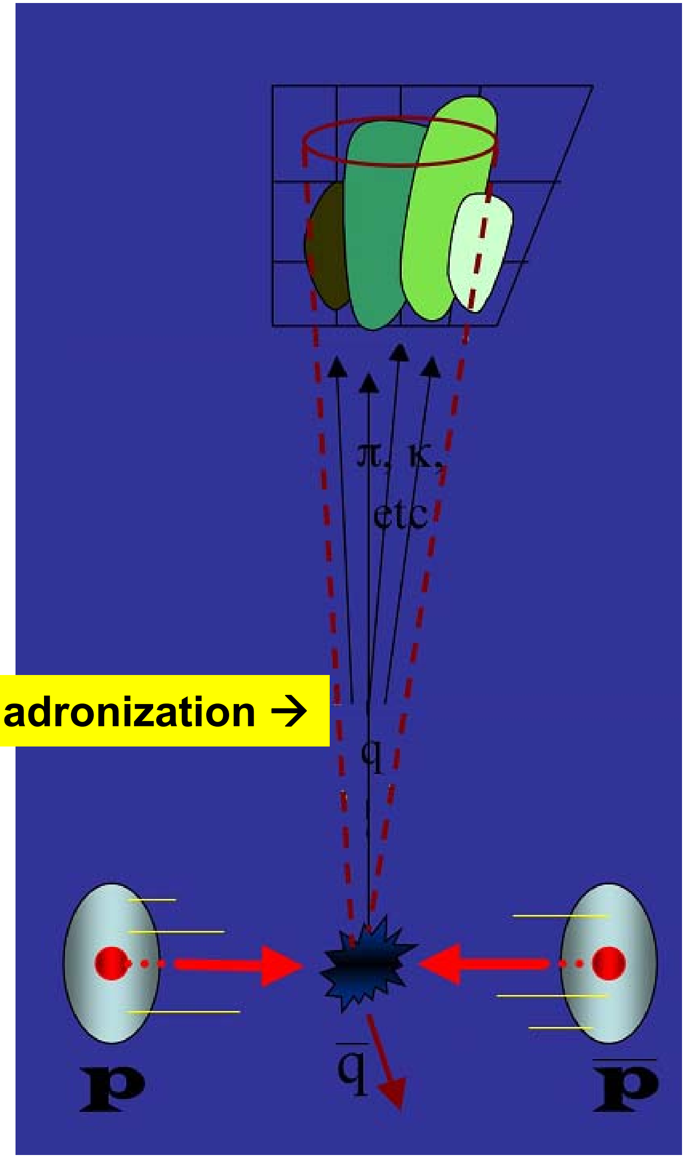
- **Increased** ***per-jet efficiency*** at same background rate than SecVtx:

+30% (relative) on a p-pbar → q-qbar MC di-jet (Jet 20 - btopqb)

Sandro De Cecco - INFN Rome - CDF collaboration meeting - Paris 31-05-07

13

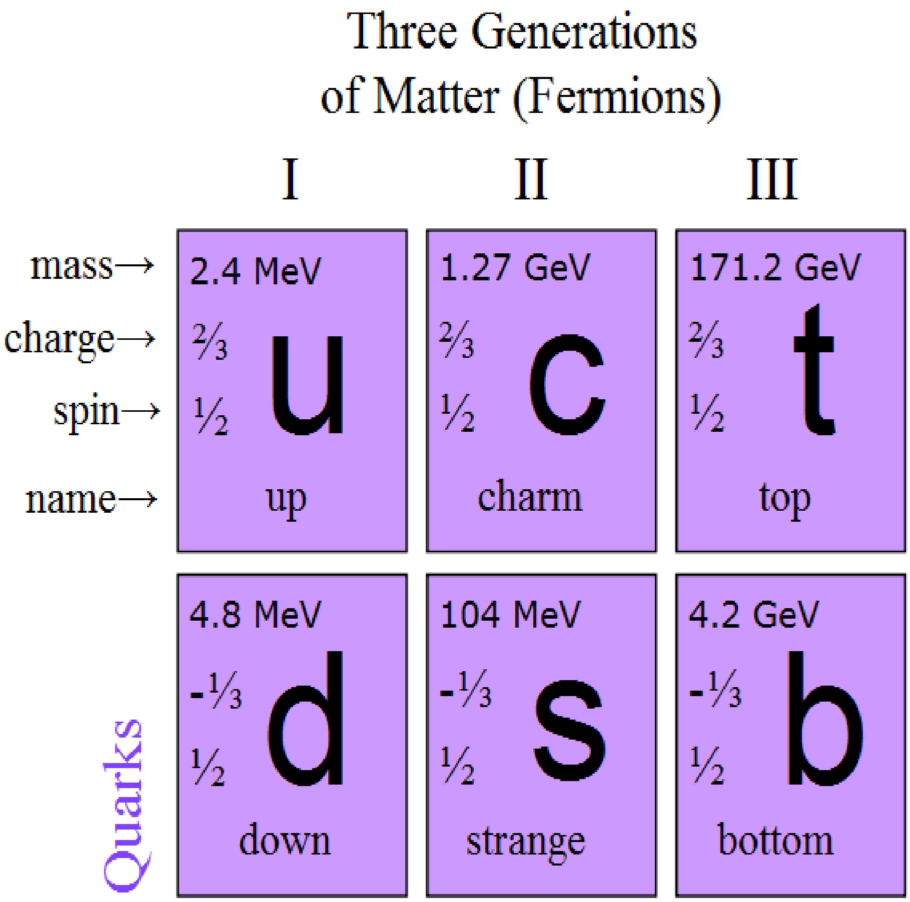
Jets



Hadronization →

- In proton-antiproton collisions, quarks and gluons are produced
- After quarks are produced, they hadronize (also called fragment) into mesons and baryons
- If the quark was produced with significant energy, the hadrons produced will be in a narrow cone
- This spray of particles is called a jet

Bottom Quarks



Three Generations of Matter (Fermions)

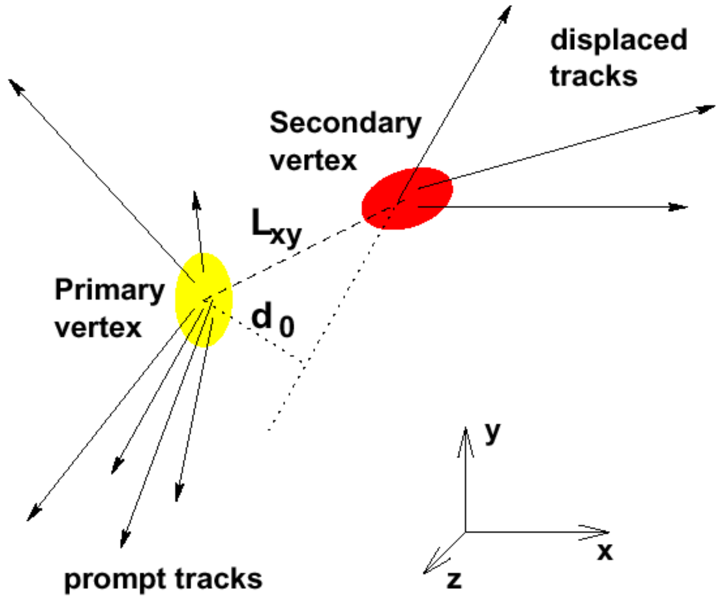
	I	II	III
mass→	2.4 MeV	1.27 GeV	171.2 GeV
charge→	⅔	⅔	⅔
spin→	½	½	½
	u	c	t
name→	up	charm	top
Quarks	4.8 MeV	104 MeV	4.2 GeV
	-⅓	-⅓	-⅓
	½	½	½
	d	s	b
	down	strange	bottom

- The bottom quark is the second heaviest quark
  - B-mesons have masses ~5.3GeV/c^2 and up, B-baryons have masses ~5.6GeV/c^2 and up
- B-hadrons have mean lifetime of ~1.5ps
  - Example: a 53 GeV b-jet has γ=10, and travels almost at c, so on average it travels 4.5mm before decaying
- We want to identify with some certainty that the jet has within it a B hadron
  - This identification is called "b-tagging", or to "tag" the b-jet

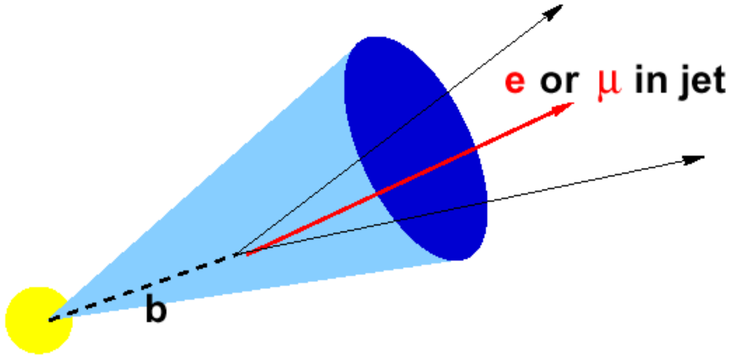
Bottom Jets Identification

Identify b-jets from the fact that B hadrons are

“long”-lived and massive
(τ=1.5ps) (>5.3 GeV)



decays semileptonically 40%



- $b \rightarrow e\nu_e X$ (BR ~ 10%)
- $b \rightarrow \mu\nu_\mu X$ (BR ~ 10%)
- $b \rightarrow cX \rightarrow e\nu_e X$ (BR ~ 10%)
- $b \rightarrow cX \rightarrow \mu\nu_\mu X$ (BR ~ 10%)

Charm hadrons are
“long”-lived and massive
(τ=1.0ps $D^\pm$) (>1.9 GeV)

Strange hadrons are
“long”-lived and massive
(τ>90ps) (~0.5 GeV)

Charm hadrons decay semileptonically 20%

- $c \rightarrow e\nu_e X$ (BR ~ 10%)
- $c \rightarrow \mu\nu_\mu X$ (BR ~ 10%)

Bottom Jets Identification

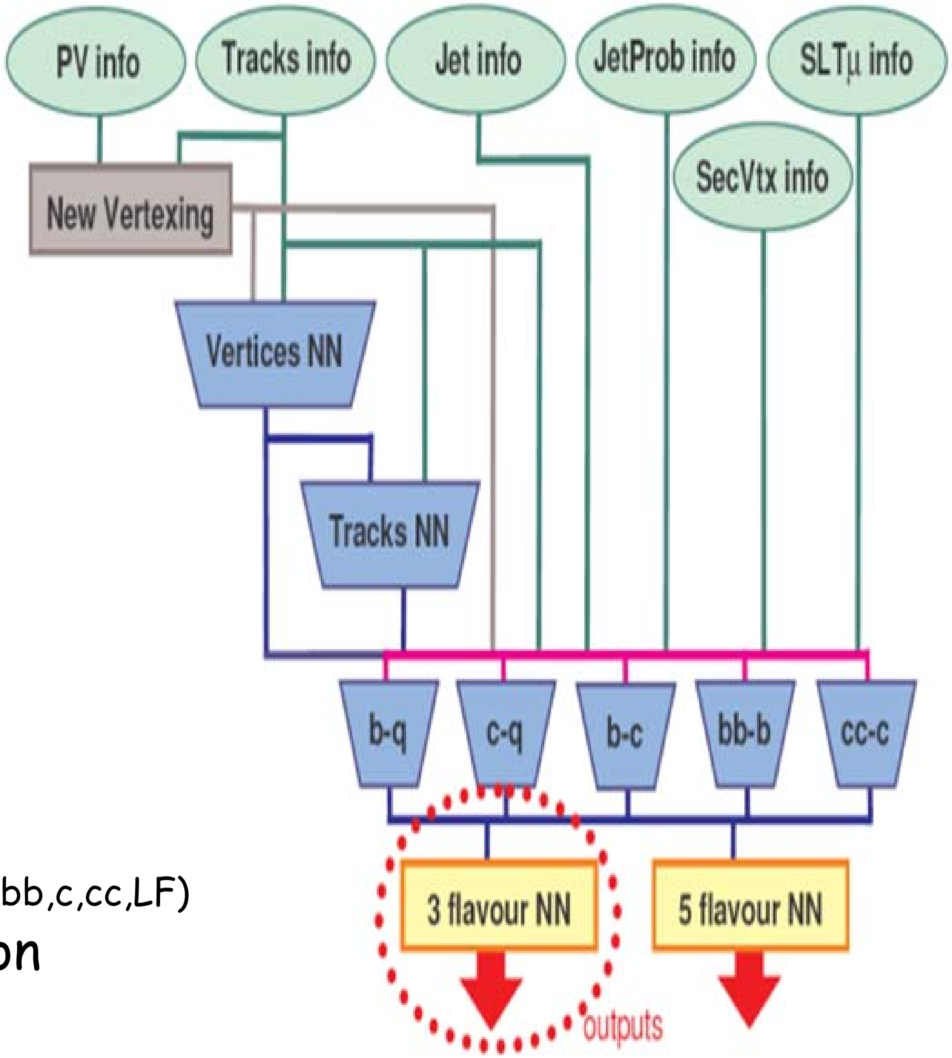
- "Long"-lived and massive
  - On average ~5 charged tracks per b-jet
  - Secondary vertex significantly displaced from primary vertex → **SecVtx**
  - More tracks with large impact parameters → **JetProb**
  - CDF Silicon detector has track hit resolution of 10um, impact parameter resolution is 30um

- Decays semileptonically → **SLT**
  - 40% of b-jets has a muon/electron within the jet

- **RomaNN uses all of this information to enhance identification efficiency while keeping misidentification rate manageable**

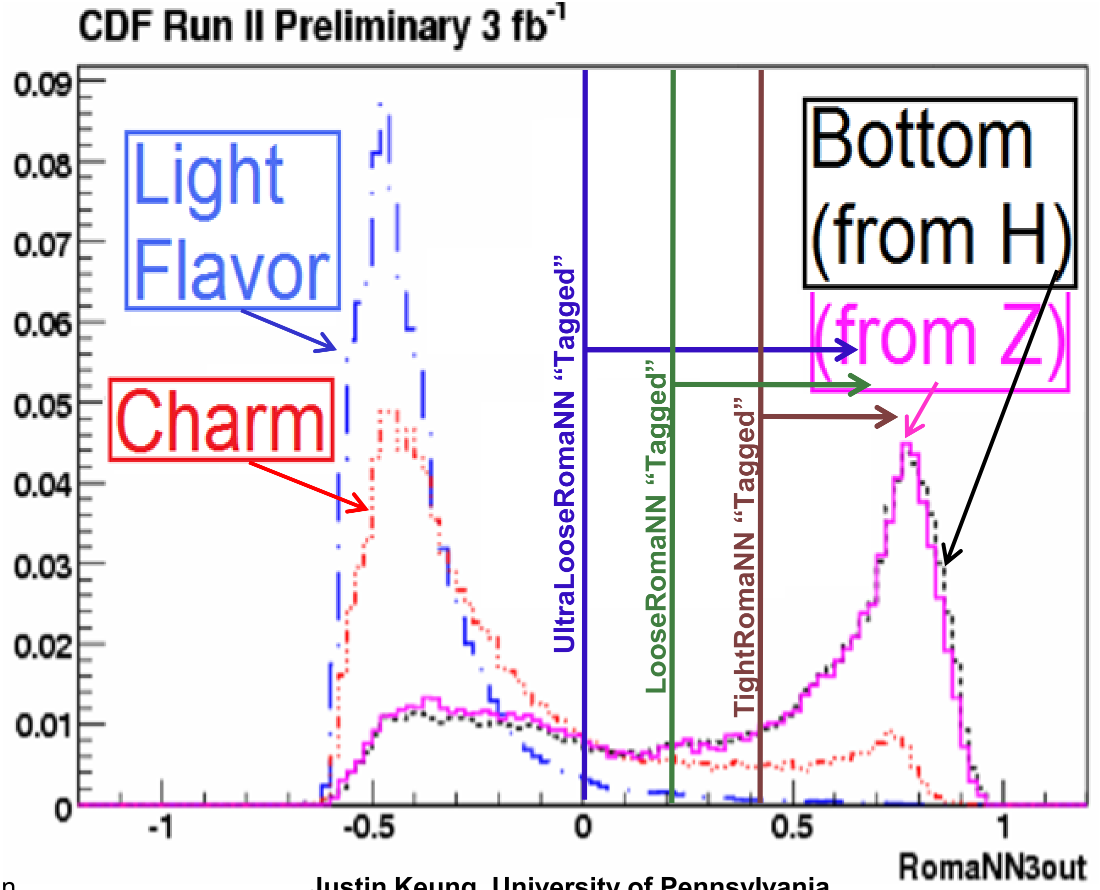
Neural Network Based b-jet Identification Algorithm

- 4 cascaded levels of Neural Networks trained by INFN-Roma
  - Vertex identification
    - Look for all vertices
    - SecVtx looks for only one
  - Track identification
    - Tell unvertexed tracks from prompt tracks
  - One by one "expert" flavor separation
    - Combines SecVtx, JetProb, SLT with Vertex/Track NNs
  - Final separation:
    - 3 flavor (b,c,LF) or 5 flavor (b,bb,c,cc,LF)
    - Using 3 flavor NN separation
      - "NN3out": outputs [-1,1]



RomaNN Usage

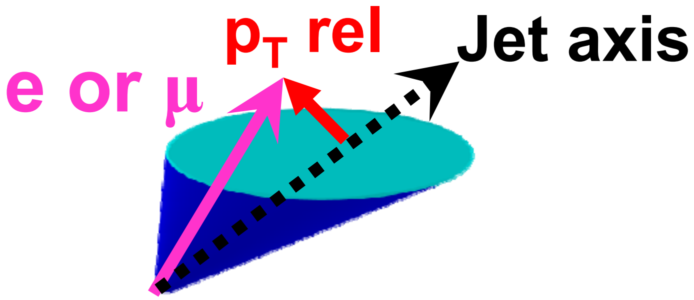
- Require NN output greater than a cut
- Three calibrated cuts: UltraLoose/Loose/Tight
  - They have different b-purity and efficiency
- Two important quantities need to be calibrated
  - ➢ Efficiency for b-jet identification
  - ➢ Misidentification rate (rate of false positives)

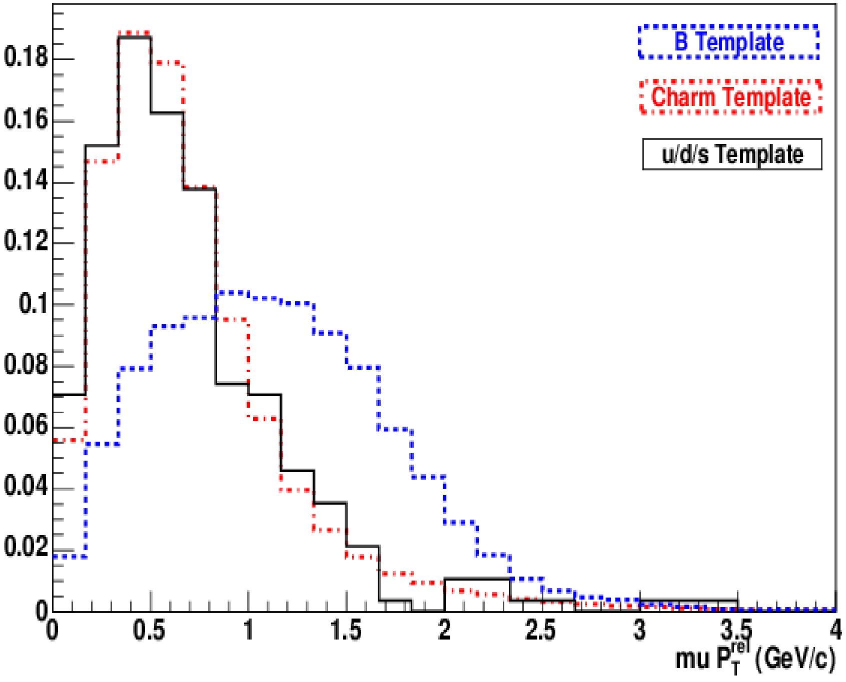


b-tagging Calibration: Efficiency

- Select dijet events (bb) with
  1. A jet tagged by loose SecVtx to improve b purity of the sample
  2. A jet containing a lepton

$$\text{lepton } p_T^{rel} = \left|\vec{P}_{lepton}\right| \sqrt{1 - \left(\frac{\vec{P}_{lepton} \cdot \vec{P}_{jet}}{\left|\vec{P}_{lepton}\right| \left|\vec{P}_{jet}\right|}\right)^2}$$

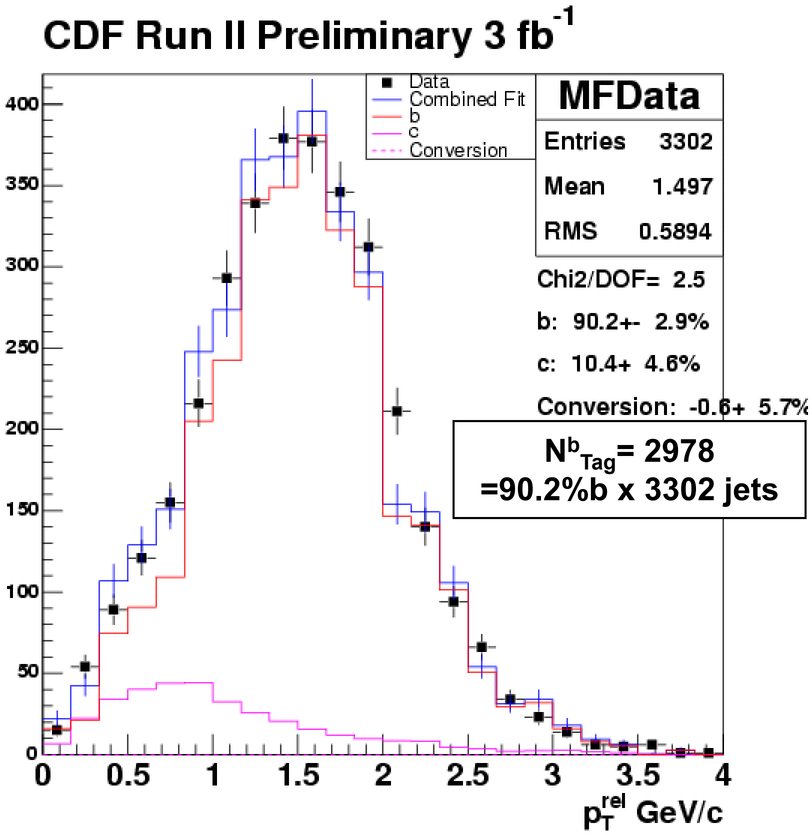




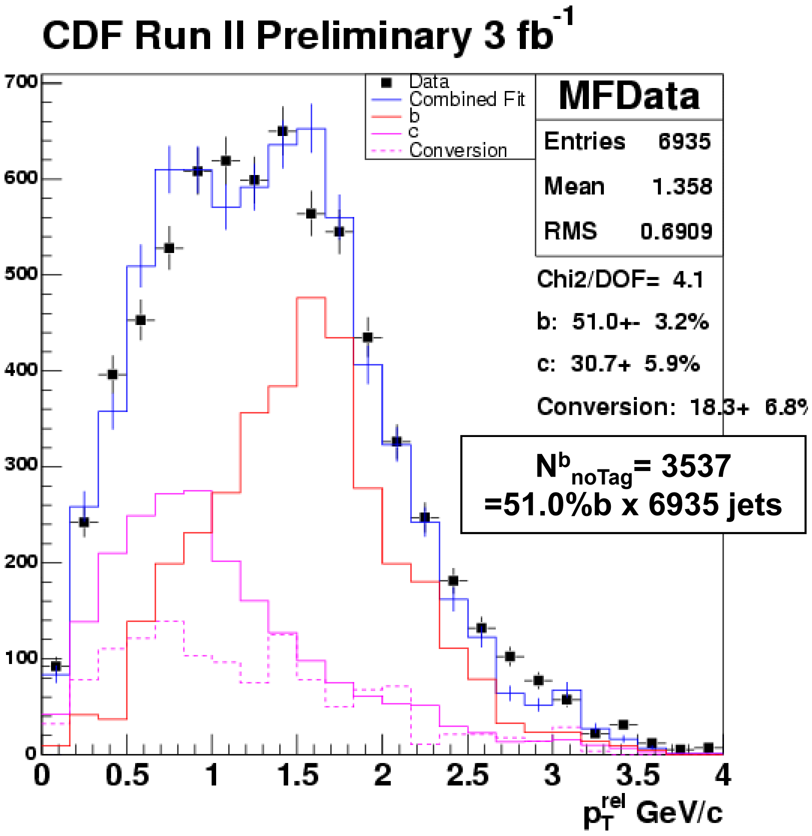
- Measure b efficiency in Data using Lepton P_T^{rel}
  - Due to the large b mass, the lepton transverse momentum relative to jet axis (P_T^{rel}) is larger than for charm and light jets

Data Efficiency Measurement

- Split sample into two subsets
  - *electron jets tagged (left), electron jets not-tagged (right)*



N^b_{Tag}= 2978
=90.2%b x 3302 jets

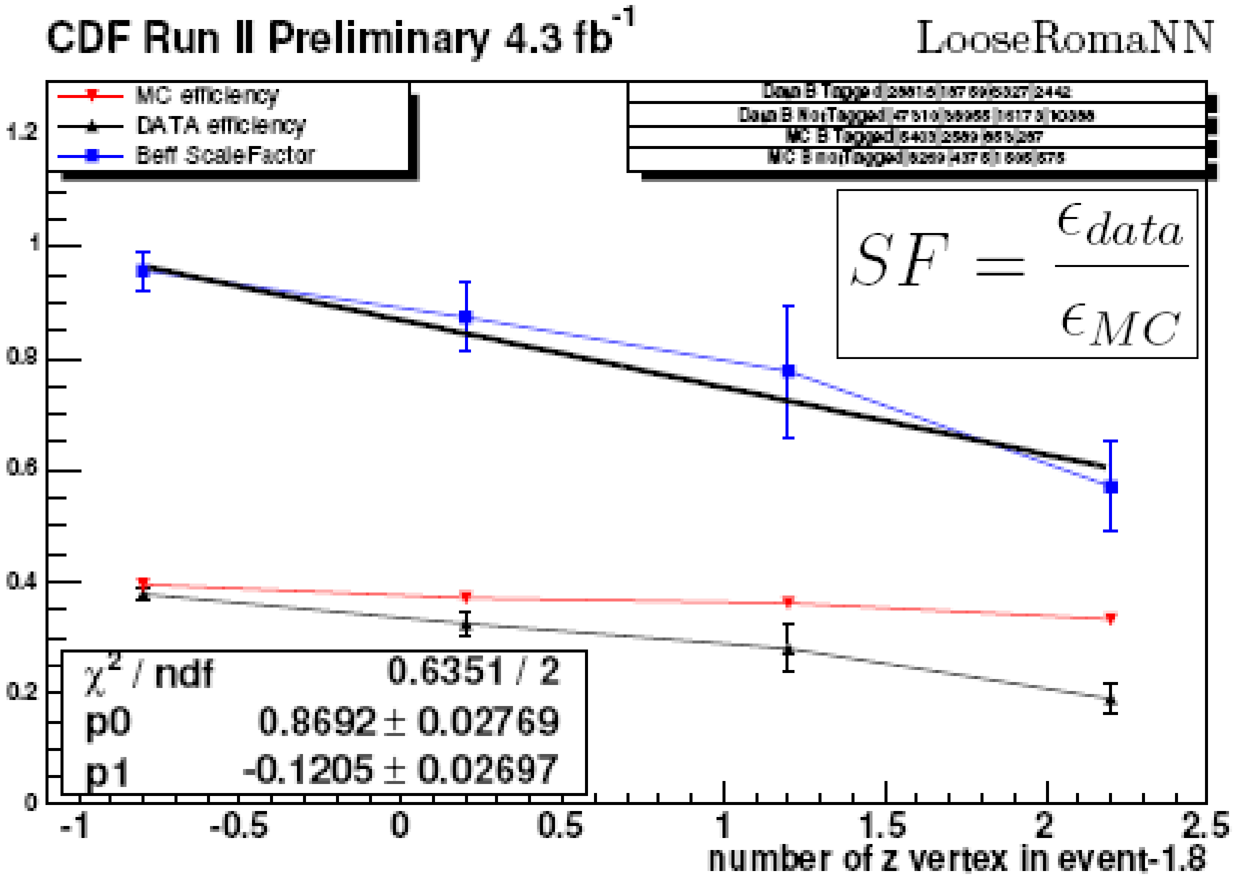


N^b_{noTag}= 3537
=51.0%b x 6935 jets

$$\epsilon_b = \frac{N^b_{Tag}}{N^b_{total}} = \frac{N^b_{Tag}}{N^b_{Tag} + N^b_{NoTag}}$$

Data Efficiency Result

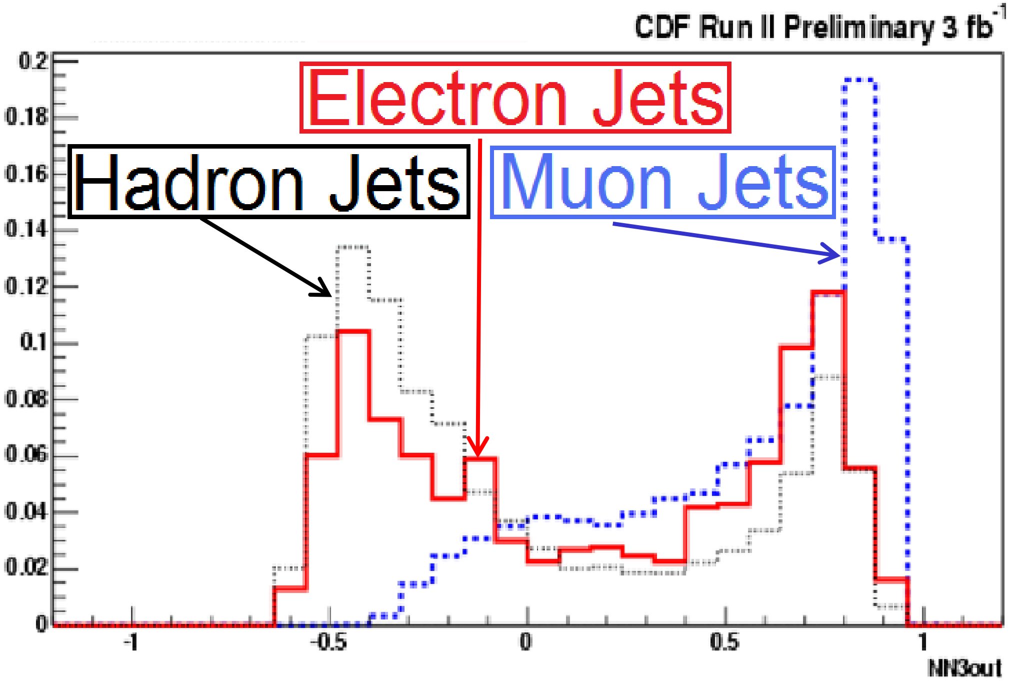
- Scale factor (SF) corrects the MC efficiency to Data efficiency
  - needed to find acceptances and yields in b-based analyses
- SF clearly decreases as the number of z vertices increases
- SF does not vary significantly with Jet E_T, Jet Eta etc



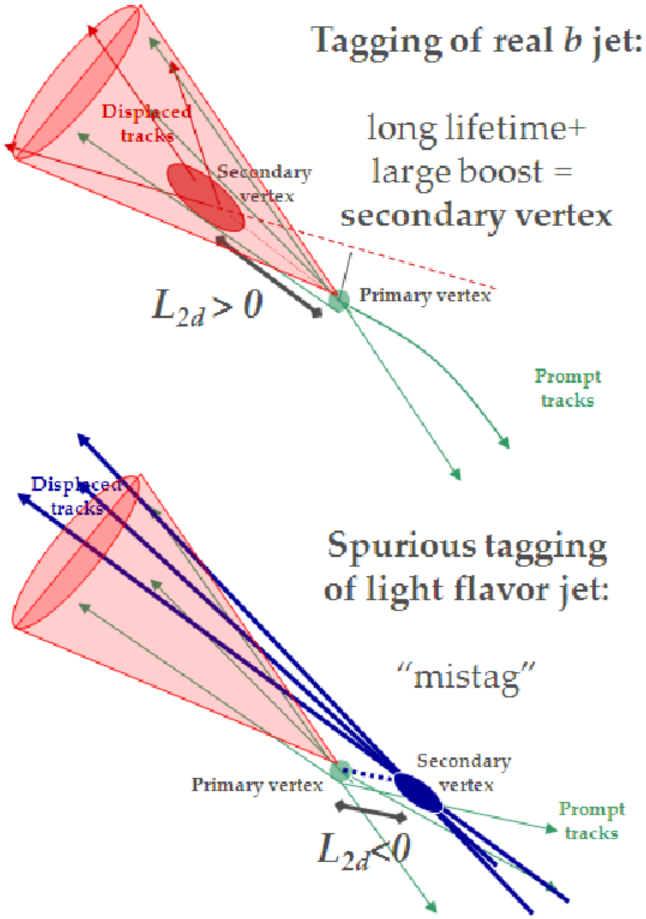
Note on Efficiency Calibration

- Electron and hadron-jets have similar NN output
  - Can calibrate efficiency for generic jets with electron jets
  - First time electrons P_T^{rel} used in CDF

- Muon and hadron-jets have different NN outputs
  - Cannot calibrate efficiency for generic jets with muon jets
  - Must calibrate efficiency for muon jets separately



b-tagging Calibration: Misidentification Rate


Tagging of real *b* jet:

long lifetime+
large boost =
secondary vertex

Spurious tagging
of light flavor jet:

“mistag”

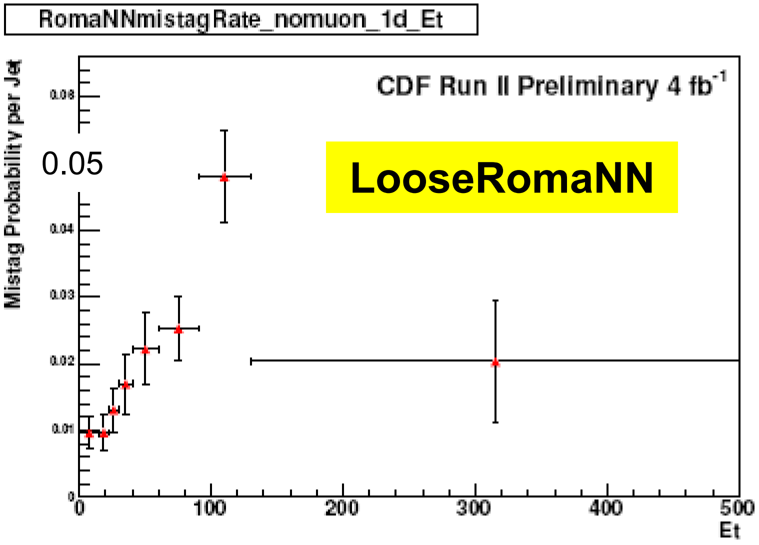
- Misidentifications are due to spurious large impact parameter tracks
  - From limited detector resolution, long-lived light particle decays, and material interactions
- For simple b-tag algorithms, misidentification due to the limited detector resolution is expected to be symmetric in their signed 2D displacement of the vector separating the primary and secondary vertices

b-tagging Calibration: Misidentification Rate

- For a more complex b-tag algorithm, the misidentifications due to the limited detector resolutions cannot now be expected to be symmetric in any single variable
- The strategy used is to measure the overall tag rate, then subtracting from it the tag rate due to real b-jets

$$\text{rate}^{\text{mistag}}_{\text{RomaNN}} = \text{rate}^{+}_{\text{RomaNN}} - \text{rate}^{\text{heavy}} \times (\varepsilon^{b}_{\text{RomaNN}} \times \text{ScaleFactor}_{\text{RomaNN}})$$

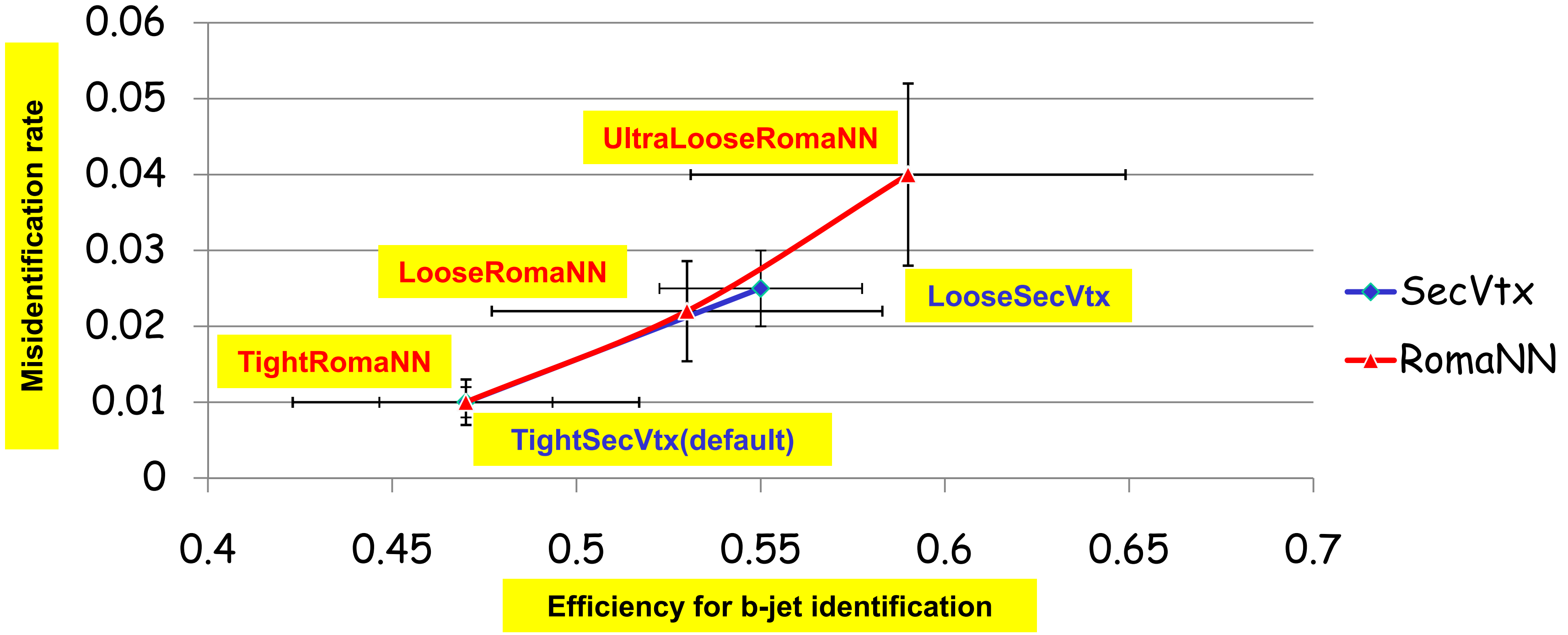
$$\text{rate}^{\text{heavy}} = \frac{(\text{rate}^{+}_{\text{SecVtx}}) - \alpha\beta(\text{rate}^{-}_{\text{SecVtx}})}{\varepsilon^{b}_{\text{SecVtx}} \times \text{ScaleFactor}_{\text{SecVtx}}}$$



- We characterize the misidentification rate using a matrix, predicting the misidentification rate for each jet based on several of its parameters, such as its energy and location within the detector

b-tagging Performance Comparison

- Compare the two important calibrated quantities
  - Efficiency: use b-jets from WH simulation
  - Misidentification rate: use jets from WH data sample
- Default b-tag tool performs similar to new b-tag tool
  - 30% increase in per-jet efficiency for same background rate seen in simulation was not seen in data
- UltraLooseRomaNN useful to increase signal efficiency
  - CDF WH search summer 2009: increased signal acceptance by 20% in the double-tag category



Outline

- Why WZ→lvbb ?
  - Higgs search effort
  - Fermilab, Tevatron and CDF

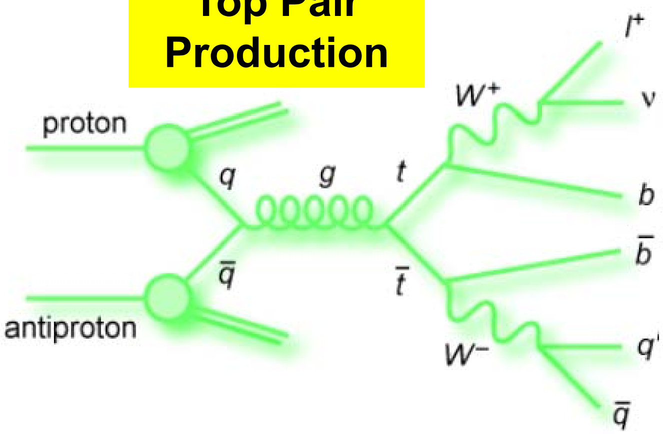
- Improvements in b-jet identification
  - Importance in the WZ/WH→lvbb search
  - Measurement of per-jet efficiency and background rate in data

- **Search for the WZ**
  - Background composition
  - Sensitivity estimates
  - Neural Network

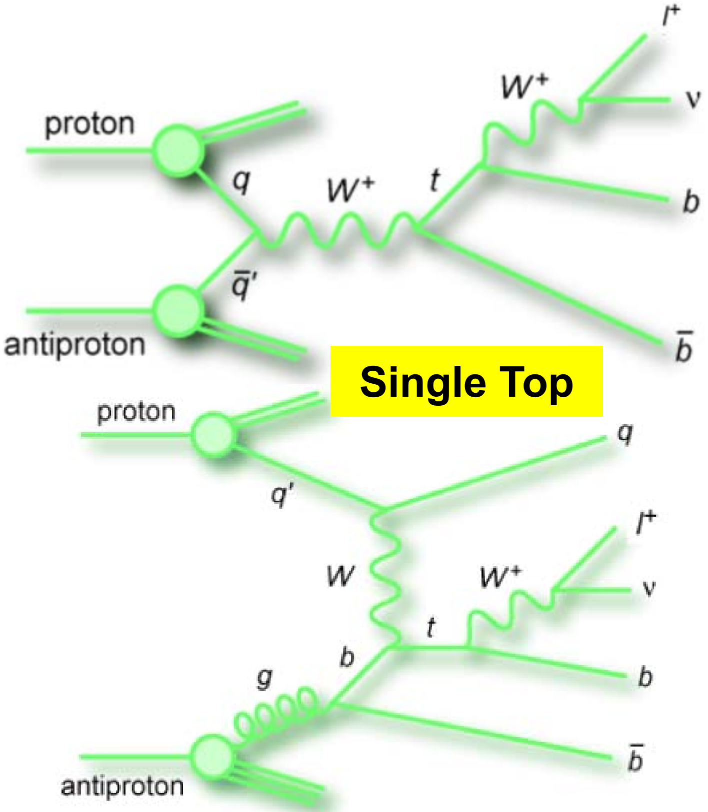
lvbb Backgrounds

- Unfortunately, many physics processes can lead us to the identification of a lepton, a neutrino, and two b-quark jets
  - 90% is irreducible background
  - top quark pair production, single top quark production, Wbb

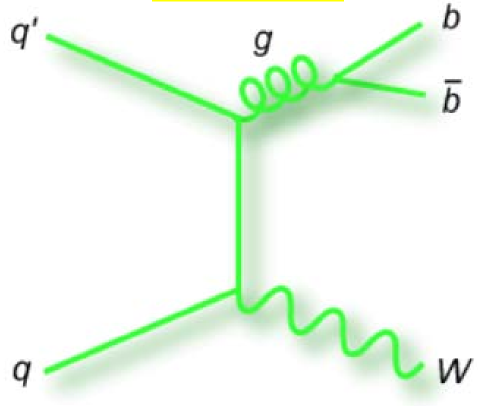
Top Pair Production



Single Top



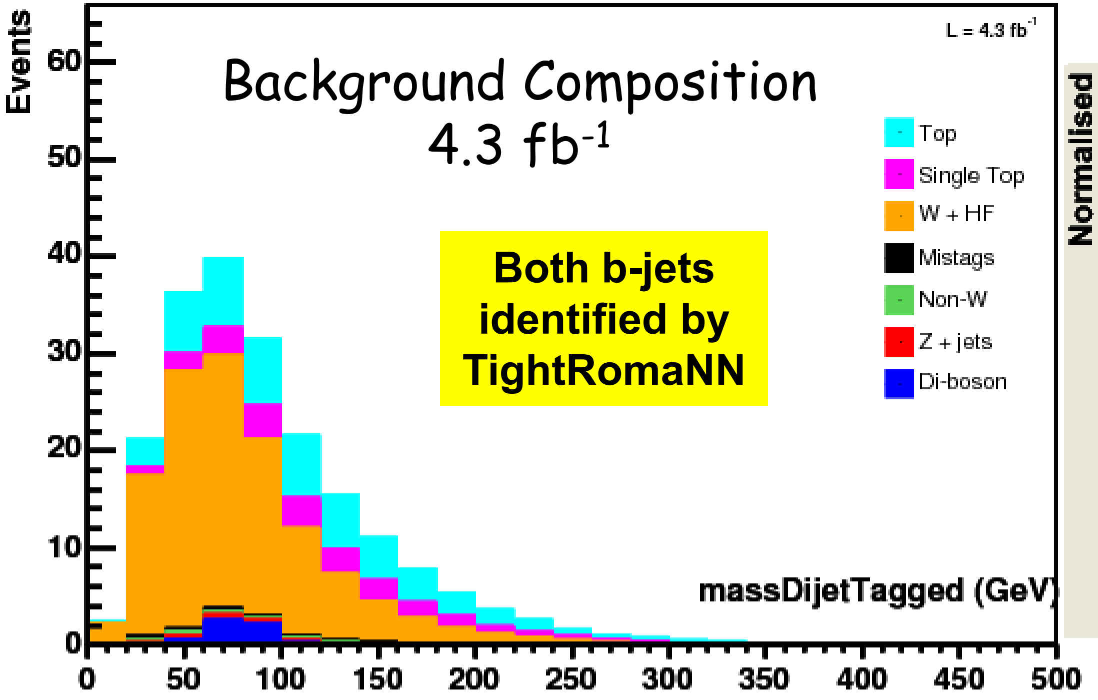
Wbb

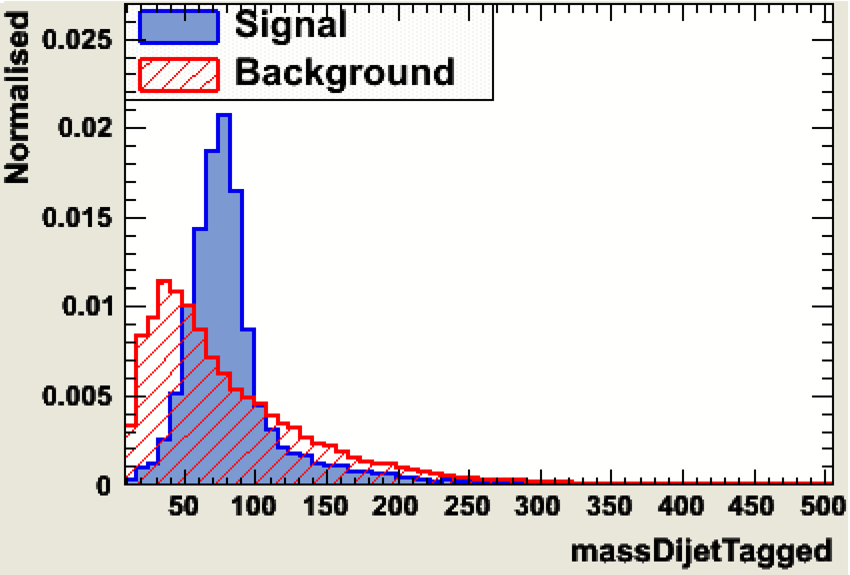


- The other 10% of background: Wcc, Wc, Wlf, Zlf, QCD

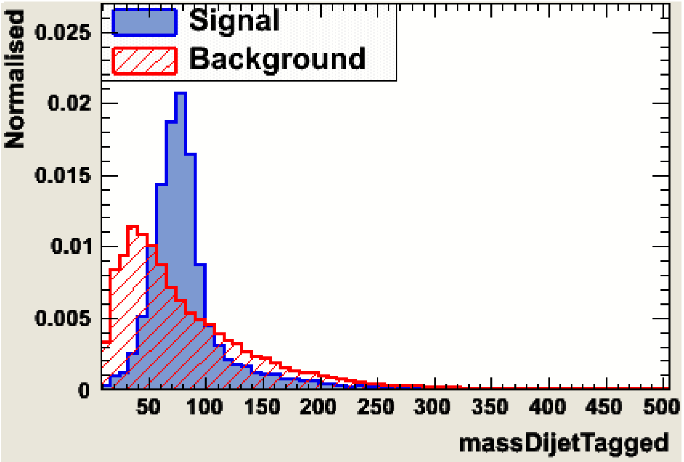
Signal and Background Expectation

- Expected WZ yield is 3% of all lvbb events, counting experiment will not give enough sensitivity to observe WZ
  - Need to use additional information in the events to distinguish between signal and background
  - For example, we expect the invariant mass of the two b-jets from WZ to resemble the z mass peak, whereas the background is more diffuse
    - Using M_{bb}, expected 95% Confidence limit is 3.6x standard model cross section with $4.3 fb^{-1}$ for WZ

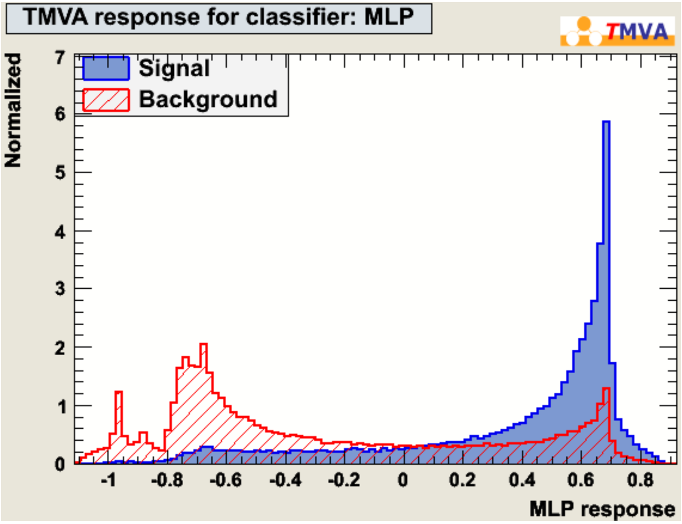


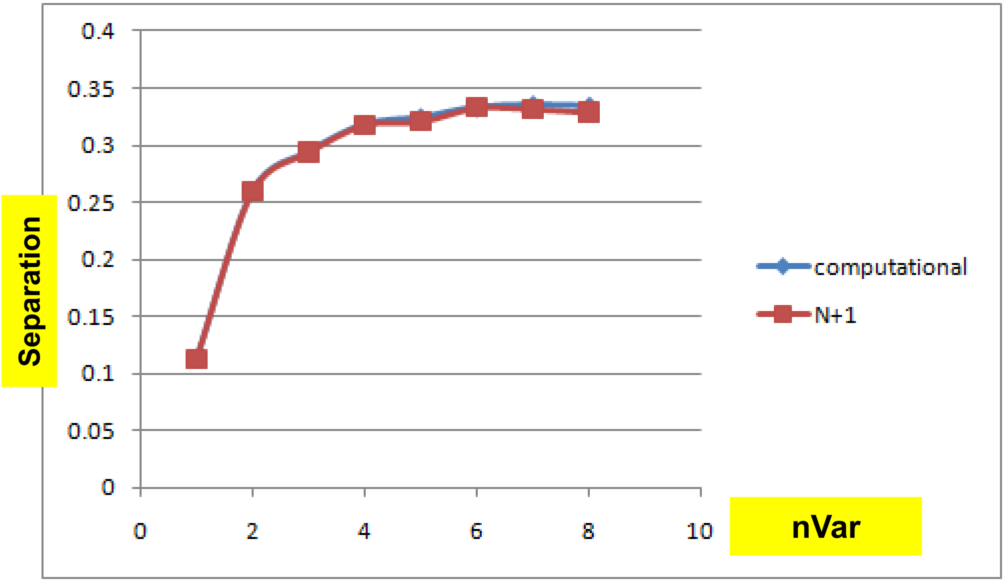


Neural Network



- Try to improve sensitivity by using a stronger discriminating variable: Neural Network
- Train and test to search for the best neural network with the fewest input variables
- For the additional complexity involved in using more inputs, not much separation gain after four variables

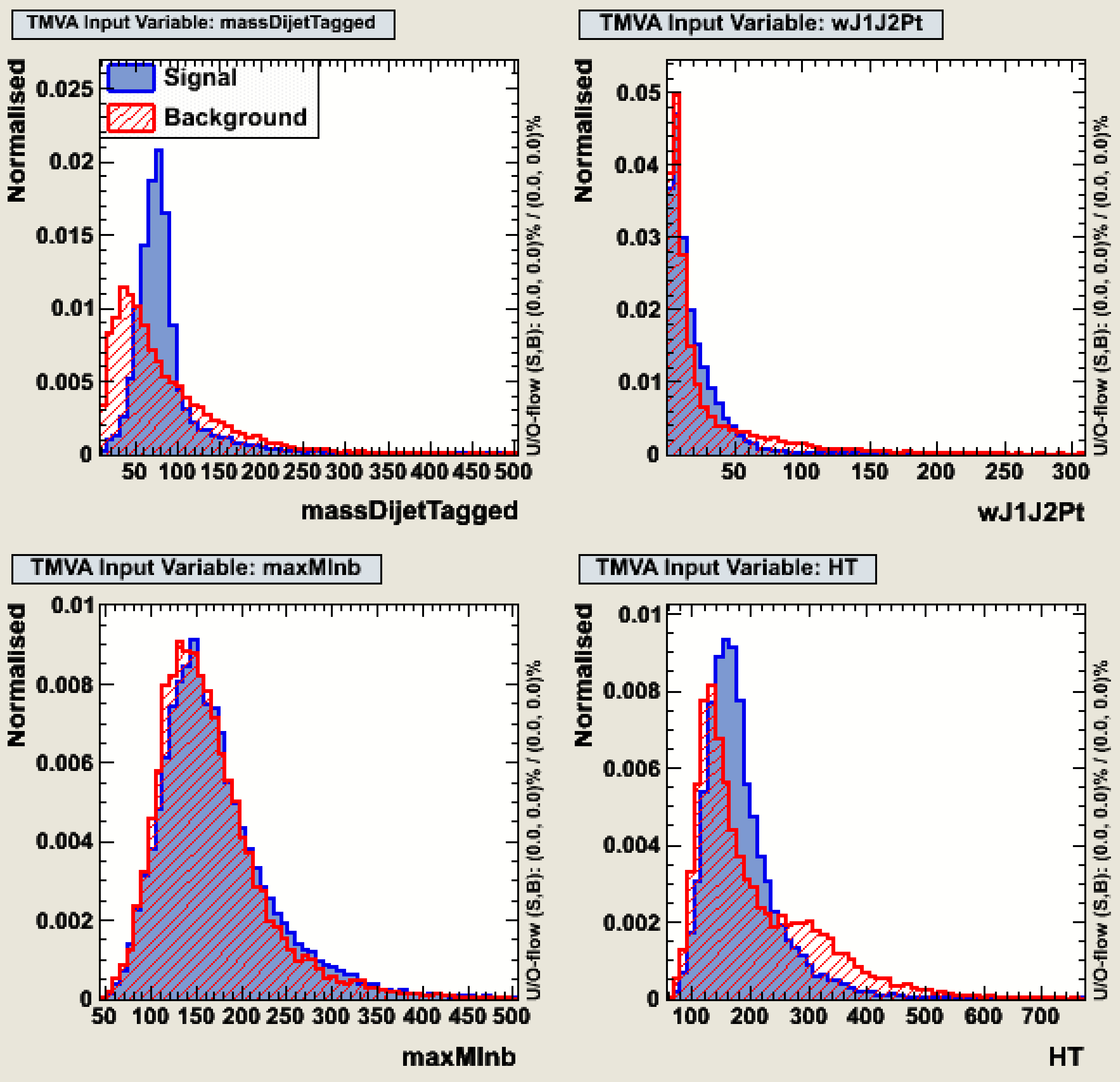




Separation: $< S^2 >= \frac{1}{2}\int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$

where $\hat{y}_S$ $\hat{y}_B$ are signal and background PDFs of y.

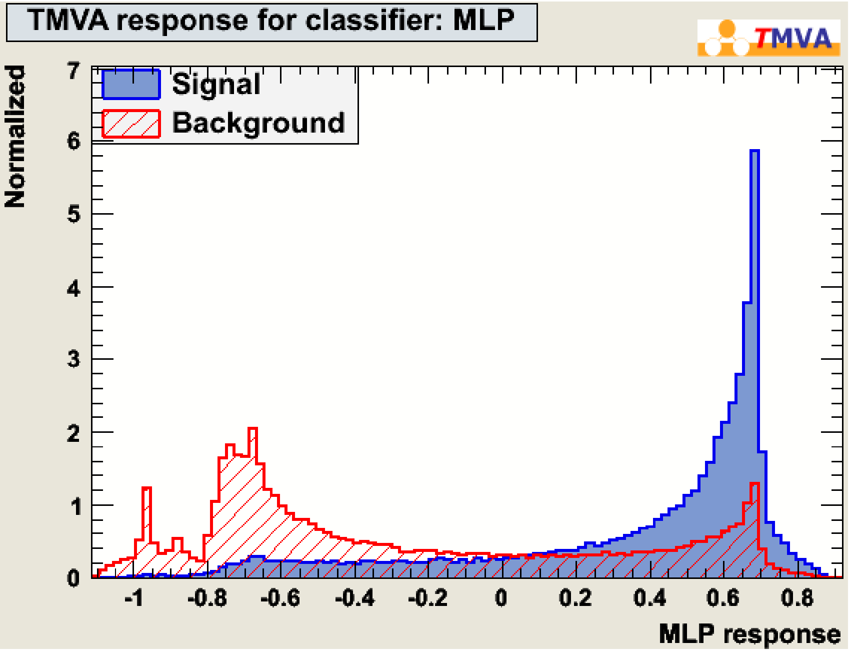
Input to TMVA Neural Net, Normalized

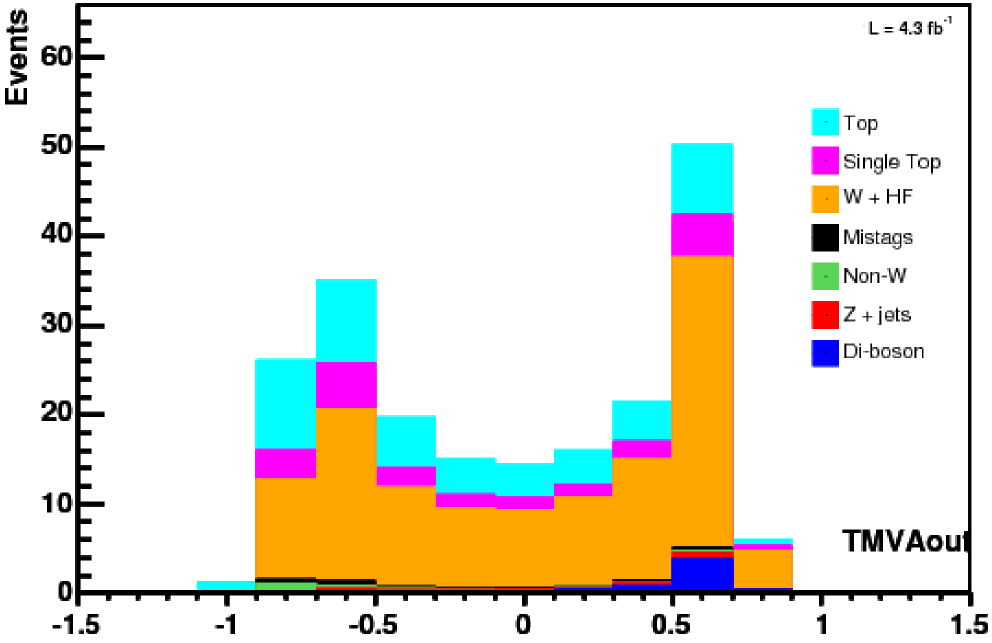


Signal Limit Estimate With Neural Network

- After training Neural Network, the most sensitive bins are transferred to the right
- Expected 95% Confidence limit is now 3.4x standard model cross section (improvement from 3.6x with M_{bb} alone)

- Apart from M_{bb}, other kinematical quantities and angular distributions contribute to a 6% improvement in limit





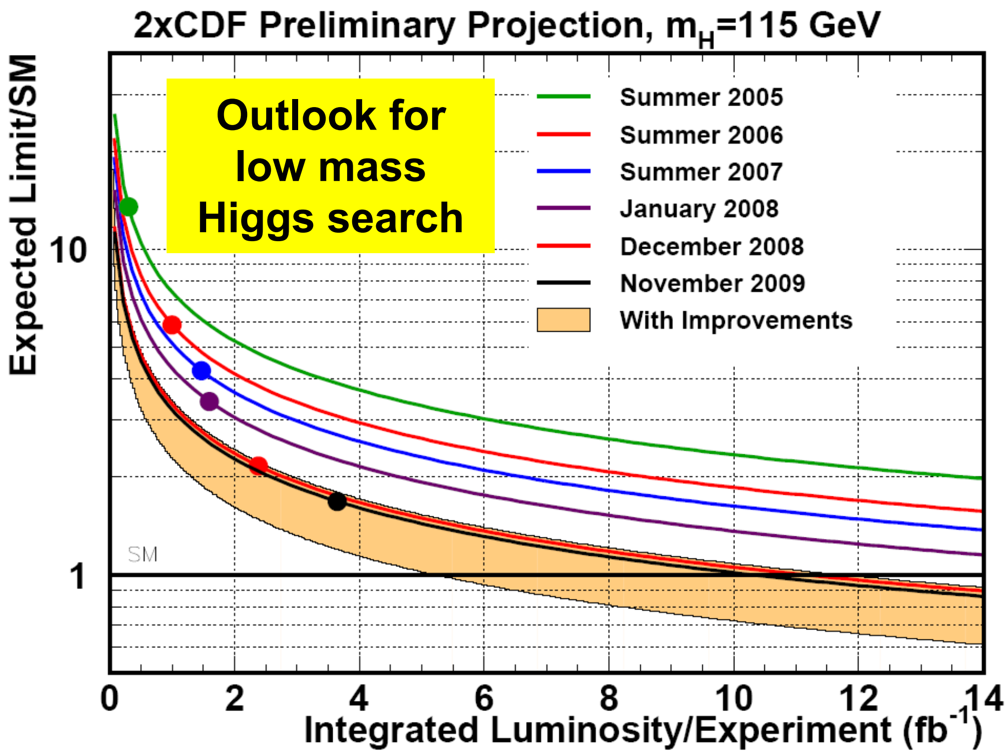
Improvements

- My work in progress
  - UltraLooseRomaNN in use already to obtain 20% acceptance gain
  - Measurement of top quark pair production to validate this new b-jet identification algorithm
  - Optimizing for the b-jet identification purity operating point, by comparison of expected WZ limit

- Other improvements from the WH working group
  - Use additional "loose lepton" categories (>10% acceptance gain)
  - Improve b-jet energy resolution (>5% improvement in limit)
    - Goal is to improve dijet invariant mass resolution
  - Add τ leptons (>5% acceptance gain) → Elisabetta Pianori @UPenn

Summary

- Searching for WZ→lvbb, part of the Higgs search effort
  - Improved b-jet identification
  - Utilized Neural Network to improve signal sensitivity
- Expect a limit of 3.4x standard model WZ cross section using 4.3fb^{-1}
- Expect total of ~10fb^{-1} after 2 more years



Outlook for low mass Higgs search

Backup

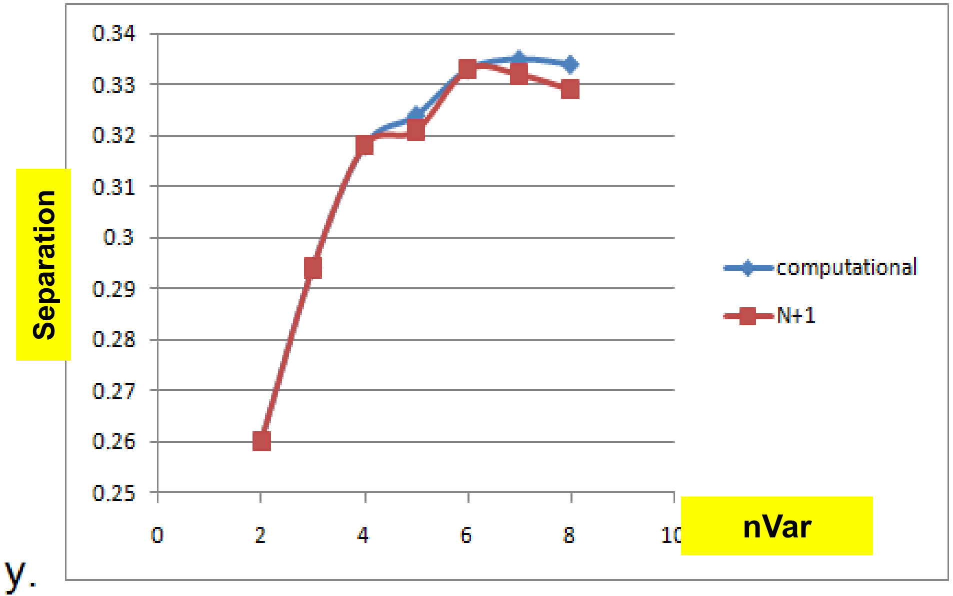
Training a Neural Network

- Train and test to search for the best neural network with the fewest input variables
- For the additional complexity involved in using more inputs, not much separation gain after four variables

nVar	vars	Separ		Legend		
9	abcdefghi	0.334		a	:	massDijetTagged
8	abcdefgi	0.334		b	:	wJ1J2Pt
7	abcdfgh	0.335		c	:	dRMetMaxLep
6	abefgh	0.333		d	:	sumLooseJetEt
5	abcgh	0.324		e	:	minMlnb
4	abfg	0.318		f	:	maxMlnb
3	abg	0.294		g	:	HT
2	ab	0.26		h	:	ptW
1	a	0.113		i	:	MetMag

Separation: $\langle S^2 \rangle = \frac{1}{2}\int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$

where $\hat{y}_S$ $\hat{y}_B$ are signal and background PDFs of y.

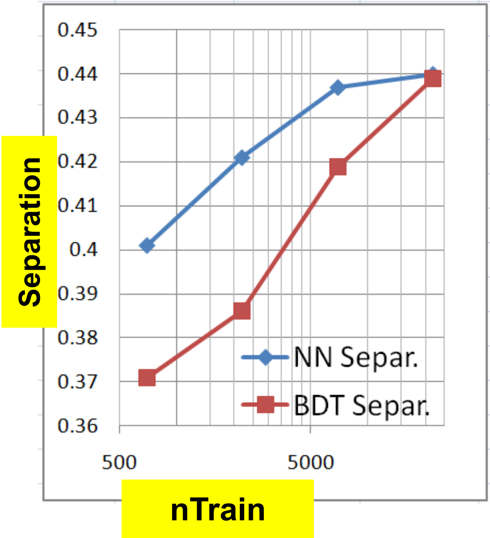


- Use 9 variables known to have the most separation power
  - Using TMVA Neural Network (MLP) with WZ as signal, and as background the cocktail from the background estimation table
- For all 511 NNs, same training set and testing set of events were used
  - (18k training events, 54k testing events)

Better Sensitivity with Neural Network

- Neural Network uses more information → better sensitivity
    - 2 studies: training sample size, and number of variables used

nTrain	NN Separ.	BDT Separ.
700	0.401	0.371
2200	0.421	0.386
7000	0.437	0.419
22000	0.44	0.439



- Insufficient training sample reduces separation power
    - Using off the shelf TMVA Neural Network and Boosted Decision Tree with different training sample size, fixed 89000 testing size
    - Want (nTest>>nTrain) to measure separation power without statistical jitter from small testing sample size
        - E.g. for NN, with nTrain=nTest=700, Separation = 0.474
    - This study used events from WZ vs Wbb (50% of bkgnd)
    - Included events with >=2 jets to increase statistics
    - Separation: $< S^2 > = \frac{1}{2}\int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$

      where $\hat{y}_S$ $\hat{y}_B$ are signal and background PDFs of y.

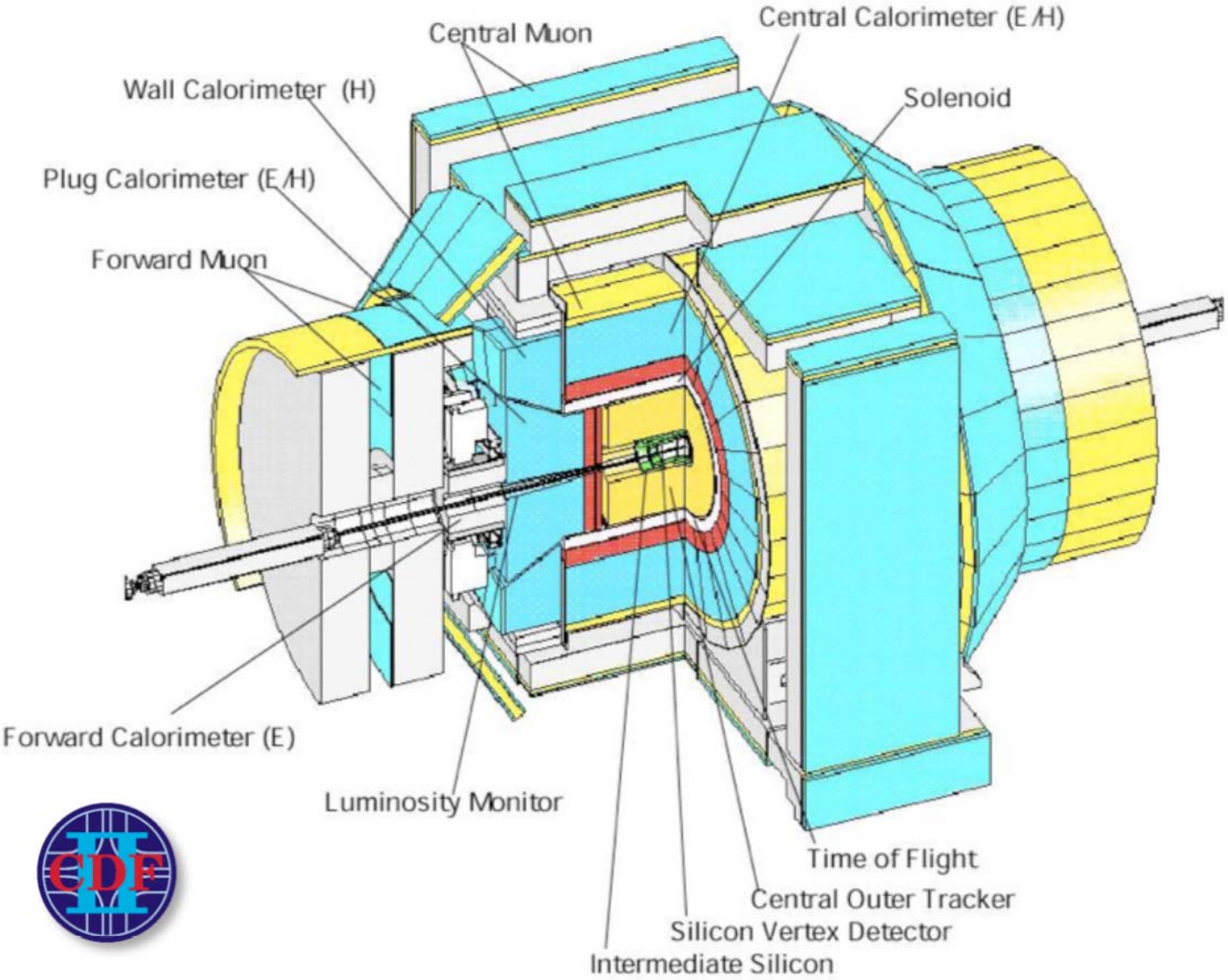
- Too much information provide leads to statistical noise, weakens the resulting Neural Network discriminant
    - Seek to provide information to the neural network in an optimal manner
- But how to choose input variables for the Neural Network?

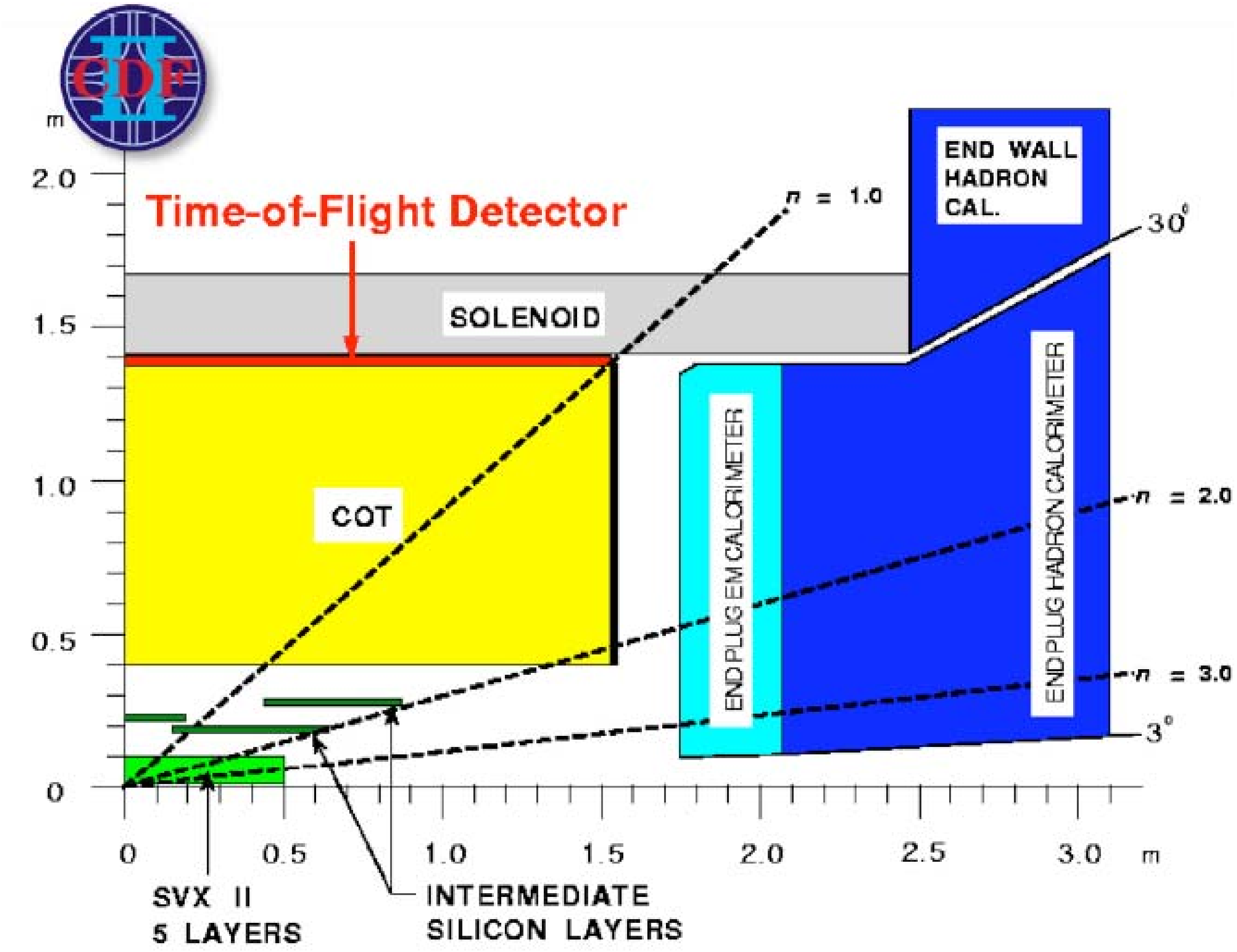
CEM Acceptance

- WZ vs WH CEM acceptance (Wbb is bkgd)
  - W→e+nu and b pair required in HEPG

	#events				this stage only, %				percent of initial		
all jet bins	WZ	WH	Wbb		WZ	WH	Wbb		WZ	WH	Wbb
Stage 0 Initial	2100	3407	10000		100	100	100		100	100	100
Stage 1 HEPG e |Eta|<1.1	1266	2428	5899		60.29	71.27	58.99		60.29	71.27	58.99
Stage 2 HEPG e ET>20GeV	1066	2161	5102		84.2	89	86.49		50.76	63.43	51.02
Stage 3 RECO+ID e in CEM	626	1267	2974		58.72	58.63	58.29		29.81	37.19	29.74
Stage 4 MET > 20GeV	553	1114	2713		88.34	87.92	91.22		26.33	32.7	27.13
Stage 5 2 jets ET>20GeV |Eta|<2.0	207	592	210		37.43	53.14	7.741		9.857	17.38	2.1
Stage 6 both SecVtx Tight	24	76	18		11.59	12.84	8.571		1.143	2.231	0.18

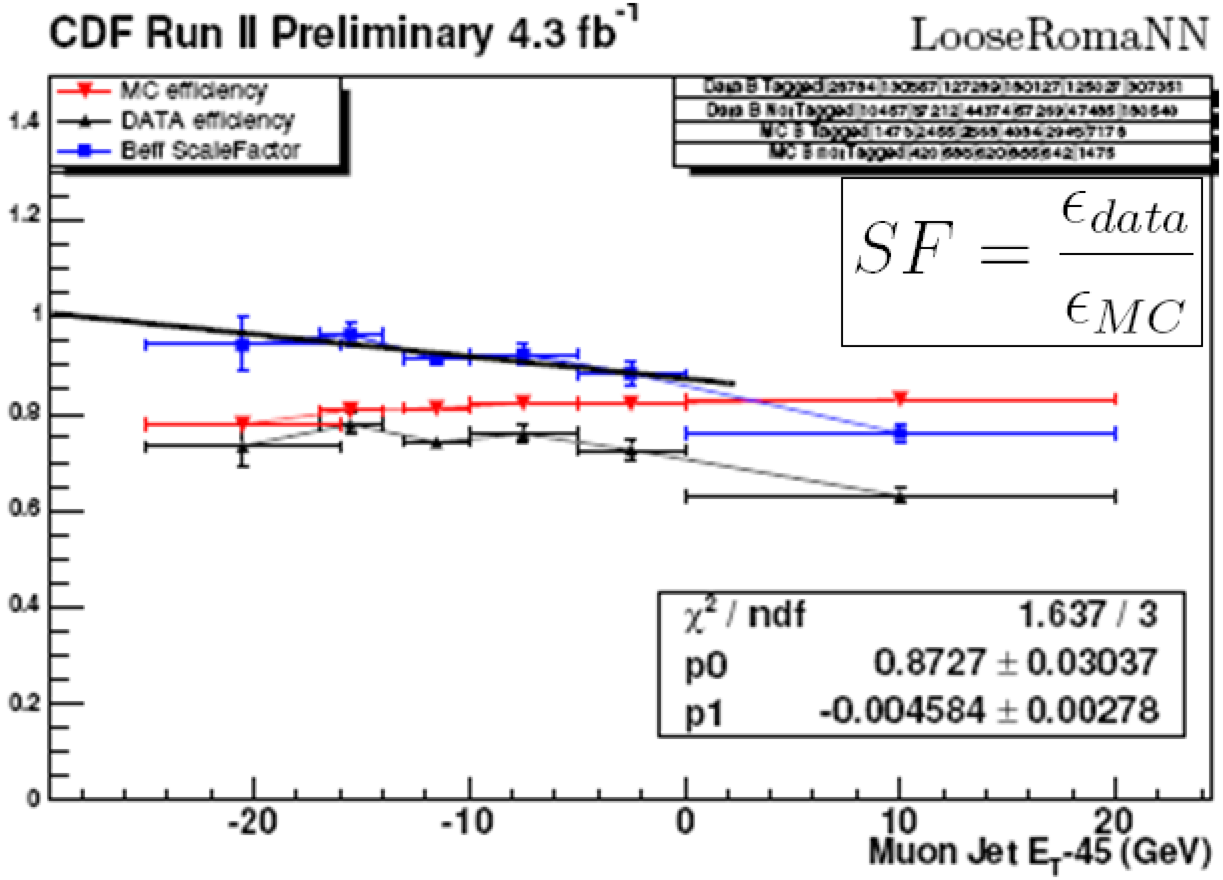
  - Stage 1+2 : e from WZ has lower average ET than WH
  - Stage 5 : jets from Z have lower average ET than from H, since Z mass is lower than H mass (120 GeV/c^2 here)
  - Stage 6 : jets from Z have lower average tag rate than from H due to lower ET


Central Muon
Central Calorimeter (E/H)
Wall Calorimeter (H)
Solenoid
Plug Calorimeter (E/H)
Forward Muon
Forward Calorimeter (E)
Luminosity Monitor
Time of Flight
Central Outer Tracker
Silicon Vertex Detector
Intermediate Silicon


Time-of-Flight Detector
m
2.0
1.5
1.0
0.5
0
SOLENOID
COT
END WALL HADRON CAL.
η = 1.0
30°
END PLUG EM CALORIMETER
END PLUG HADRON CALORIMETER
η = 2.0
η = 3.0
3°
0
0.5
1.0
1.5
2.0
2.5
3.0
m
SVX II
5 LAYERS
INTERMEDIATE
SILICON LAYERS

Data Efficiency Result: Muon Jets

- Scale factor (SF) corrects the MC efficiency to Data efficiency
  - needed to find acceptances and yields in b-based analyses
- SF clearly decreases as the Jet E_T increases
- SF does not vary significantly with # z vertex, Jet Eta etc



Backup: Legend in more detail

- tight-jet: E_T > 20GeV and |eta| < 2.0
- loose-jet: 12GeV < E_T < 20GeV and |eta| < 2.4

	Legend	
a	:	massDijetTagged
b	:	wJ1J2Pt
c	:	dRMetMaxLep
d	:	sumLooseJetEt
e	:	minMlnb
f	:	maxMlnb
g	:	HT
h	:	ptW
i	:	MetMag

- DijetMass: Largest vector sum mass of all pairs of tagged-jet
- WJ1J2P_T: Vector sum P_T of (lepton + Met + dijet)
- dRMetMaxLep: deltaR between lepton and Met
  - |Met_z| is chosen such that the Mass(lepton+Met) = W mass
  - sign(Met_z) chosen such that it maximizes dRMetMaxLep
- sumLooseJetE_T: Scalar sum of all loose-jets
- minMlnb: minimum invariant mass of the lepton, Met, and 1 of the b-jet
- maxMlnb: maximum invariant mass of the lepton, Met, and 1 of the b-jet
- HT: Scalar sum E_T of all tight-jets and loose-jets and lepton and Met
- P_TW: P_T of the vector sum lepton and Met
- MetMag: |Met|