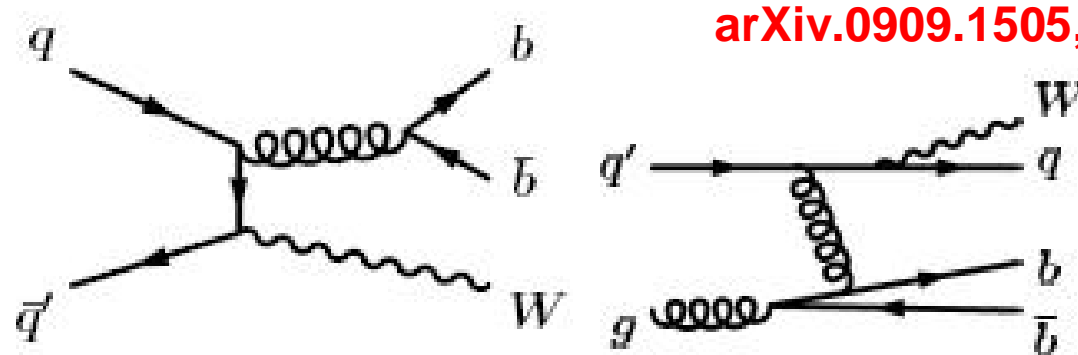


Measurement of the b -jet Cross Section in Events with a W Boson in $ppbar$ Collisions at $\sqrt{s} = 1.96$ TeV



Chris Neu (U. Virginia)

Evelyn Thomson, Joel Heinrich (U. Pennsylvania)

Northwest Terascale Research Projects
W/Z+ b-quark physics at the LHC
September 14-18 2009
University of Oregon

Motivation

Accurate model of $W+b$ -jet production is essential for searches for rare processes that have the same signature

Calculation	Prediction (pb)
MCFM NLO	1.22±0.14
PYTHIA	1.10
ALPGEN	0.78 $Q^2 = M_W^2 + p_{T,W}^2$

Progress in NLO calculations

Wbb 1999, 2006 Ellis, Veseli PRD 60, 011501
Cordero, Reina, Wackerroth PRD 74, 034007

Wbj 2007 Campbell, Ellis, Maltoni,
Willenbrock PRD 75, 054015

Wb 2009 Campbell, Ellis, Cordero, Maltoni,
Reina, Wackerroth, Willenbrock PRD 79,
034023

Note well: rare $WH \rightarrow \ell v b b$ searches (0.042 pb at $m_H=115 \text{ GeV}/c^2$):

- $W+b$ -jet uncertainty same size as 10x expected WH events
- Multivariate discrimination depends on model of $W+b$ -jet kinematics from ALPGEN (leading-order)

We present first measurement of the $W+b$ -jet production rate with better than 20% accuracy (arXiv.0909.1505, submitted to PRL)

How we quote the measurement

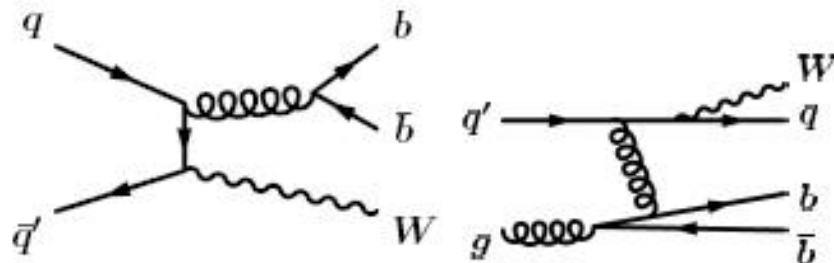
NB: We quote our measurement in a restricted region of phase space

§ Avoids large extrapolation of our measurement to experimentally inaccessible regions (*i.e.* low jet E_T , large jet η)

§ Any theorist can compare their new result with our measurement based on a clear definition of the phase space

We also quote our measurement as a jet-level cross section

§ Avoids model-dependent correction for number of b jets per event



Phase space requirements are made on stable final state particles from $W+b$ -jet production and are coincident with our reconstructed event selection:

✓ Lepton $p_T > 20$ GeV/c with $|\eta| < 1.1$, and neutrino $p_T > 25$ GeV/c

✓ One or two hadron-level jets with $E_T > 20$ GeV and $|\eta| < 2.0$

✓ Hadron-level jets (SPARTYJET) use final state particles and same cone algorithm as our reconstructed jets

Theoretical predictions are jet-level & have same phase requirements imposed

✓ *i.e.* $1.22 \text{ pb} \times 1900 \text{ pb}^{-1} = 2320$ b jets in events passing phase space requirements

Event Selection

Focus on leptonic W decays, $W \rightarrow \ell \nu$ where $\ell = e, \mu$

Online event trigger:

§ 18 GeV $|\eta| < 1.1$ electron OR

§ 18 GeV $|\eta| < 0.6$ muon OR

§ 18 GeV $0.6 < |\eta| < 1.0$ muon

W selection:

§ $p_T > 20$ GeV/ c isolated central lepton

§ Large missing transverse energy: MET > 25 GeV

Jet selection:

§ Exactly 1 or 2 $E_T > 20$ GeV, $|\eta| < 2.0$ jets

§ JetClu clustering with $R=0.4$

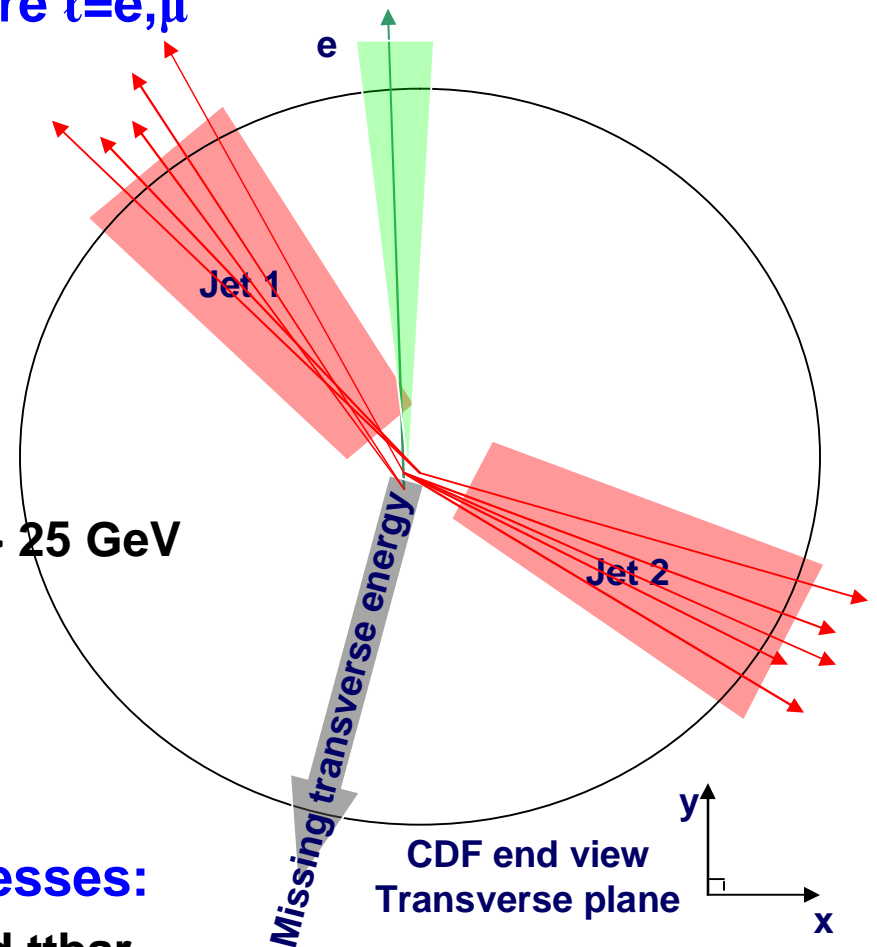
Do not consider events from other processes:

§ Veto events w/ 2 high p_T leptons to avoid $t\bar{t}$

§ Reject $Z \rightarrow \ell\ell$ production with one lepton not fully reconstructed

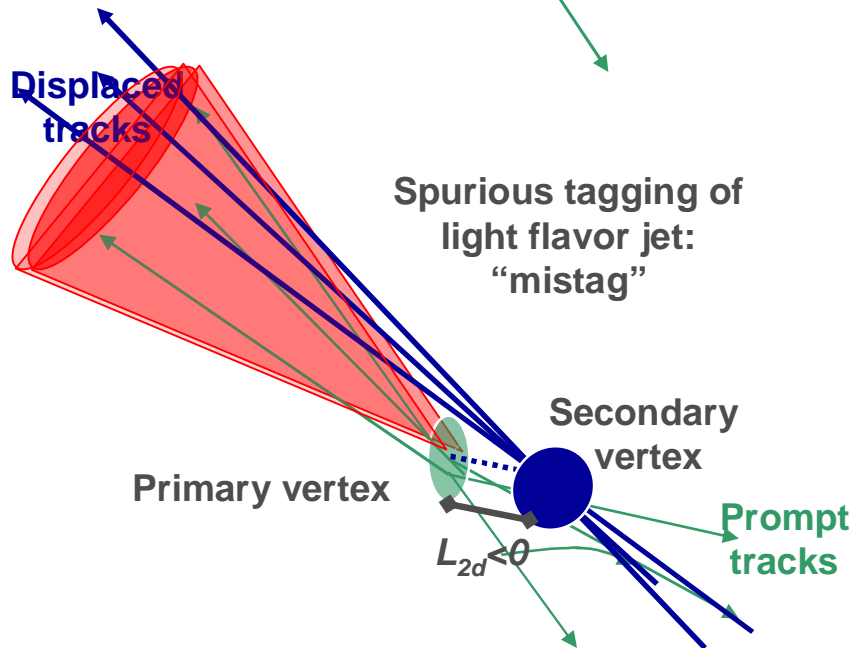
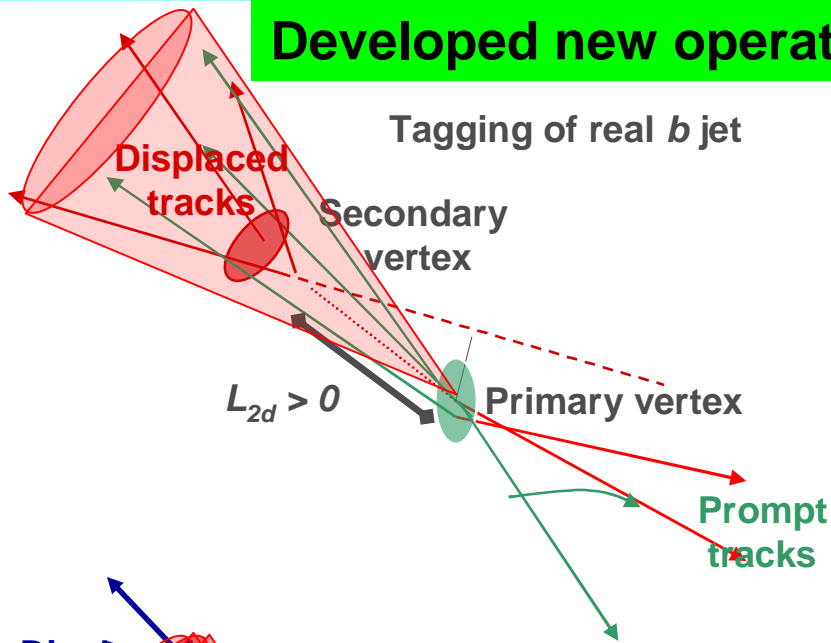
§ Remove cosmic rays, events with objects from different interactions

§ Veto fake W events

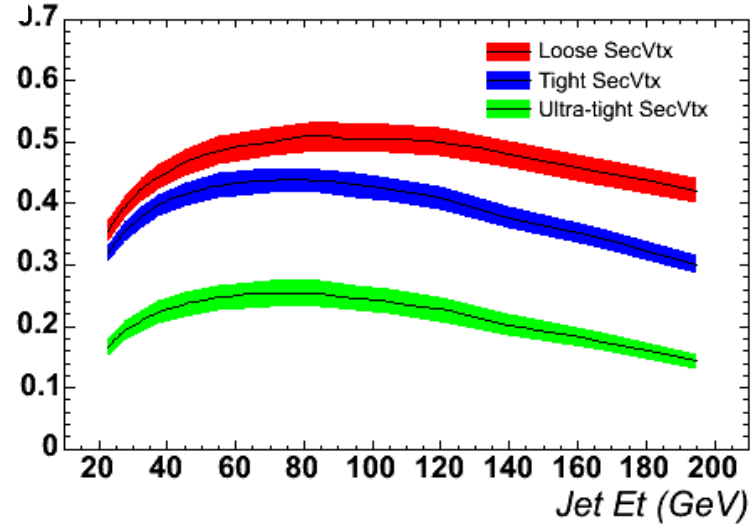


b-jet identification: Ultra-tight

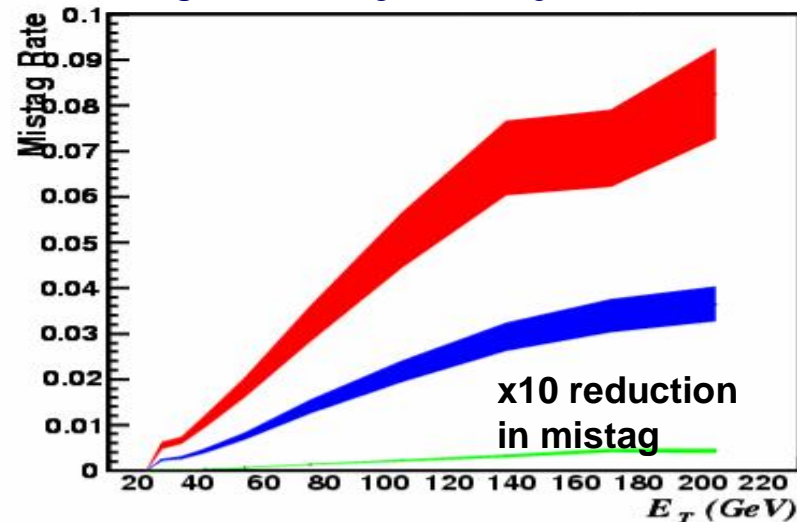
Developed new operating point of SecVtx to reduce mistags



Tag efficiency for *b* jets

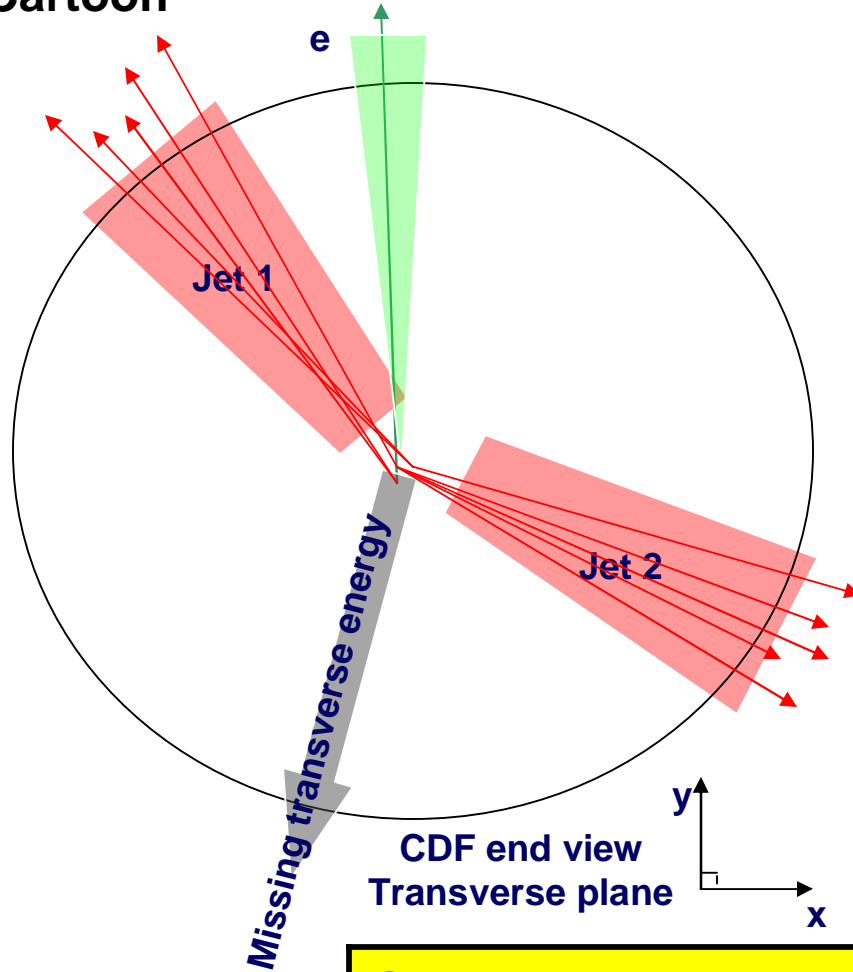


Tag efficiency for LF jets

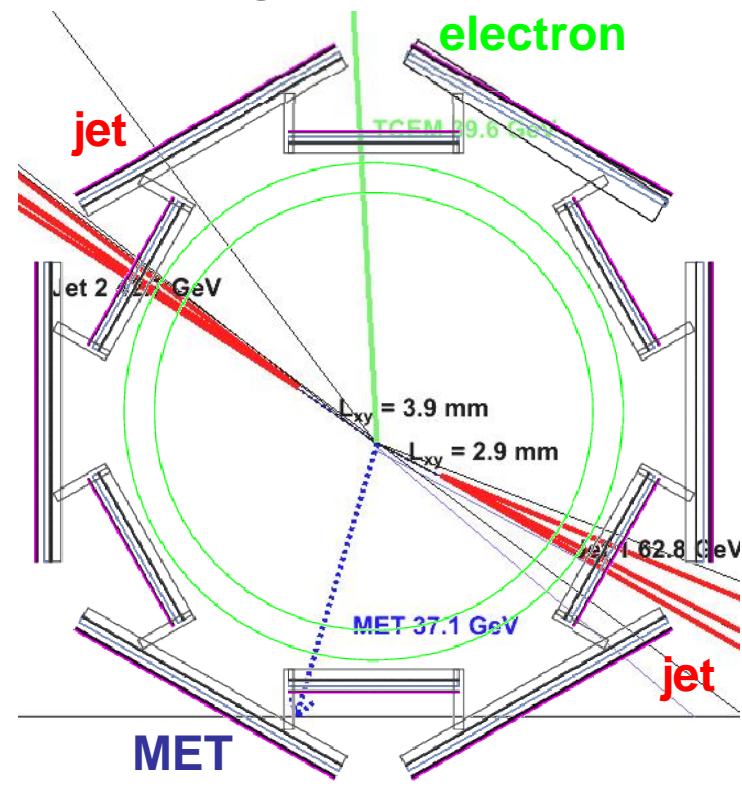


Events & jets in 1.9 fb⁻¹

Cartoon



The real thing: event recorded 10/2005



Event has 2 tagged jets!

Selected events	175,712
Total jets	199,670
Tagged jets	943

Ultratight SECVTX (UT)
has lower yield but
increased purity.

W+b jets: Measurement Strategy

$$\sigma_{b \text{ jets}} \times B(W \rightarrow \nu) = \frac{n_{\text{tag}} \cdot f^b - n_{\text{bkg}}^{b \text{ jets}}}{\sum_{e, \mu} L \cdot A_{W+b}^{b \text{ jets}} \cdot \epsilon_{\text{tag}}^b \cdot \epsilon}$$

Where do various pieces come from?

Discriminate b jets from c /LF jets in tagged sample using **vertex mass**

$$n_{\text{tag}} \cdot f^b$$

Determine contribution from **background** with tagged b jets and subtract from overall yield

$$n_{\text{bkg}}^{b \text{ jets}}$$

Calculate **acceptance** for $W+b$ -jet events

$$A_{W+b}^{b \text{ jets}}$$

Measure **tag efficiency** for b jets in $W+b$ -jet production in MC and correct to match data

$$\epsilon_{\text{tag}}^b$$

Luminous region, trigger efficiency, lepton id SF

$$\epsilon$$

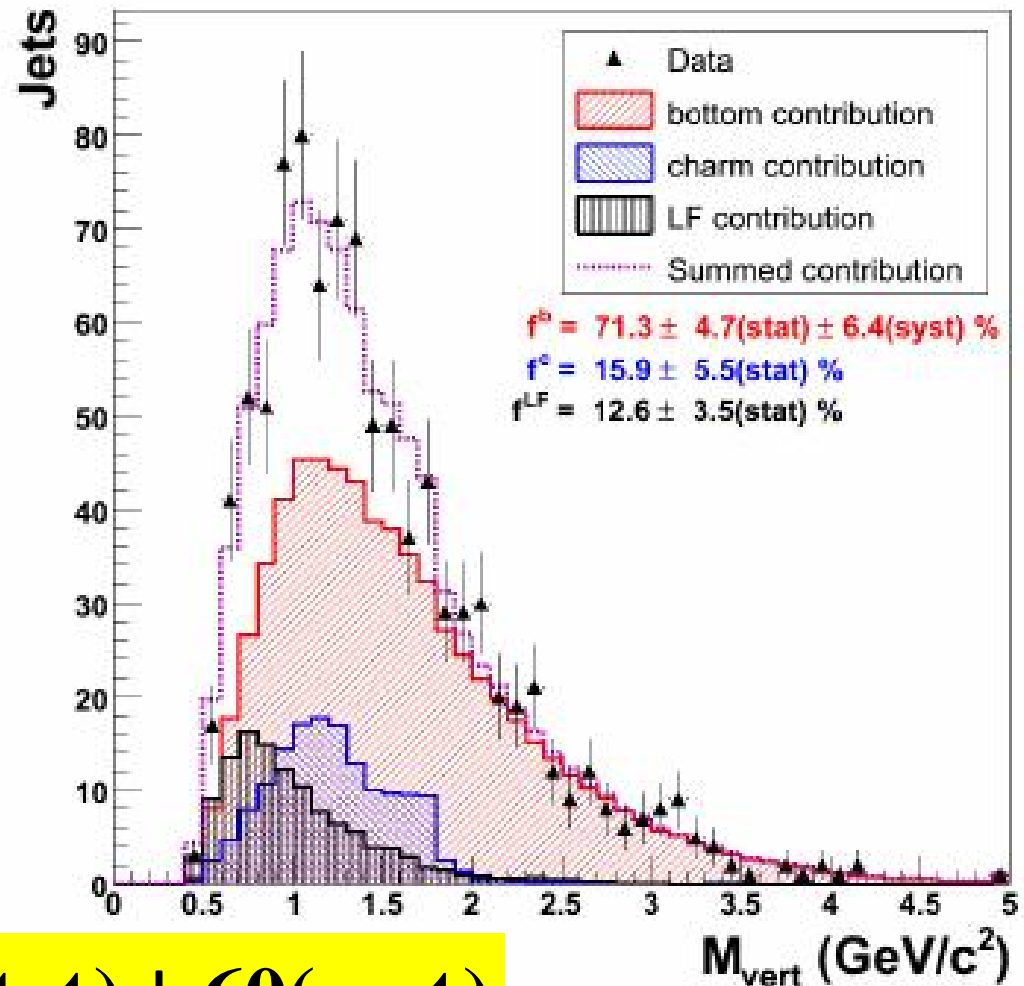
Discrimination with vertex mass

Invariant mass of charged particle tracks in secondary vertex correlated to mass of parent hadron

b-jet distribution: b-matched tagged jets in simulation for W+b jets, ttbar, single top with weighted contributions, all have very similar shapes

c-jet distribution: c-matched tagged jets in simulation for W+c-jets (W+c, W+c cbar)

LF-distribution: non-HF matched tagged jets in simulation for dijets (Jet50) for higher statistics



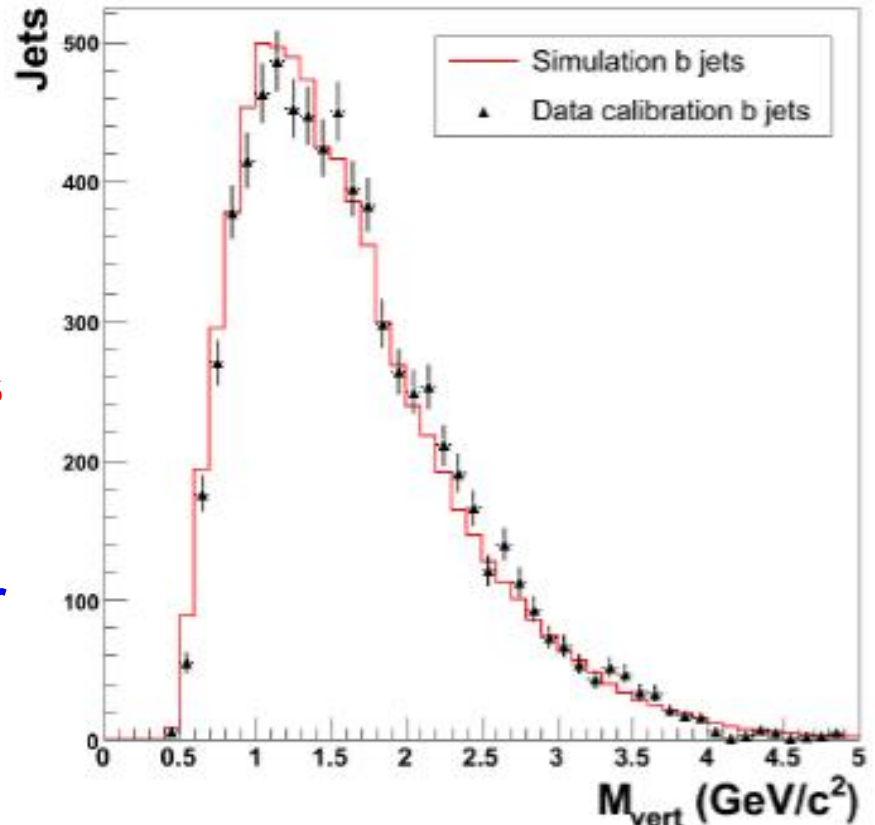
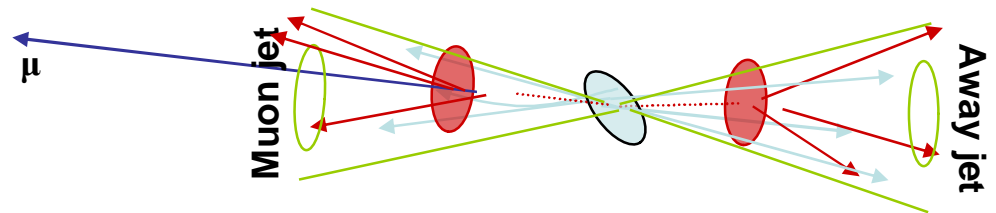
$$n_{\text{tag}} \cdot f^b = 670 \pm 44(\text{stat}) \pm 60(\text{syst})$$

Calibration & Fit systematics

Data-based calibration of vertex mass distribution for b jets

Can construct a pure sample of Ultratight tagged b jets in data:

- Trigger: $p_T > 8$ GeV/c muon
- Back-to-back dijet system:
 - Muon jet: UT-tag, $M_{vert} > 1.7$ GeV/c²
 - Away jet: UT-tag
- Away-jet b purity > 99% in Pythia
- Data-simulation b -jet difference gives a relative $\delta f_b/f_b$ systematic of 8%
- c, LF shape systematics have smaller effects of 1% and 3%



Backgrounds with b jets

Of the 670 tagged b jets, some are from other processes with b jets

$$n_{\text{bkg}}^{b \text{ jets}} = 177 \pm 22$$

MC-based estimate for top, diboson, Z+jets normalized to theoretical predictions

§ Small contribution from $W \rightarrow \tau\nu$ treated as background





Data-based estimate for non-W

§ Eliminated as much as possible with non-W veto

§ Use anti-electron sample as MET model

§ Fit MET distribution to find non-W

§ Fit vertex mass distribution in anti-electron sample to find fraction of tagged jets that are really b jets

Process		n_{W+12j}^b
$t\bar{t}$		73.1 ± 10.1
s-channel		22.2 ± 9.6
t-channel		33.4 ± 15.0
WZ		9.1 ± 0.9
ZZ		0.28 ± 0.03
WW		0.83 ± 0.12
$W + bb+Np, W \rightarrow \tau\nu$		7.3 ± 0.8
$Z + bb+Np, Z \rightarrow e^+e^-$		0.67 ± 0.08
$Z + bb+Np, Z \rightarrow \mu^+\mu^-$		4.1 ± 0.4
$Z + bb+ \geq Np, Z \rightarrow \tau^+\tau^-$		1.48 ± 0.20
Non-W		24.5 ± 8.4
Total		176.8 ± 22.3

Acceptance and Phase Space

$$\mathcal{A}_{W+b}^{b \text{ jets}} = \frac{\# \text{reconstructed } b \text{ - matched jets in events passing selection} \quad \& \text{ phase space}}{\# \text{hadron } b \text{ - matched jets in events passing phase space}} = 0.68 \pm 0.03$$

- Acceptance factor from ALPGEN $W+b$ jets MC corrects for two effects:
- ✓ Smearing from fragmentation effects, CDF detector resolution, lower energy for b -jets with semileptonic decays
 - ✓ Reduction in sample through selection cuts that isolate signal
 - ✓ MC events only migrate out due to event vetoes

- Recall, phase space requirements are made on stable final state particles and are coincident with our reconstructed event selection:
- ✓ Lepton $p_T > 20$ GeV/c with $|\eta| < 1.1$, and neutrino $p_T > 25$ GeV/c
 - ✓ One or two hadron-level jets with $E_T > 20$ GeV and $|\eta| < 2.0$
 - ✓ Hadron-level jets (SPARTYJET) use final state particles and same cone algorithm; b jet is a jet matched within $\Delta R < 0.4$ to a b quark (after showering)

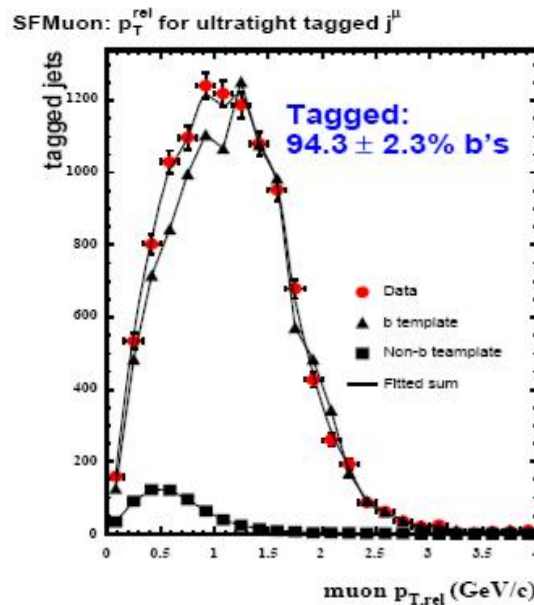
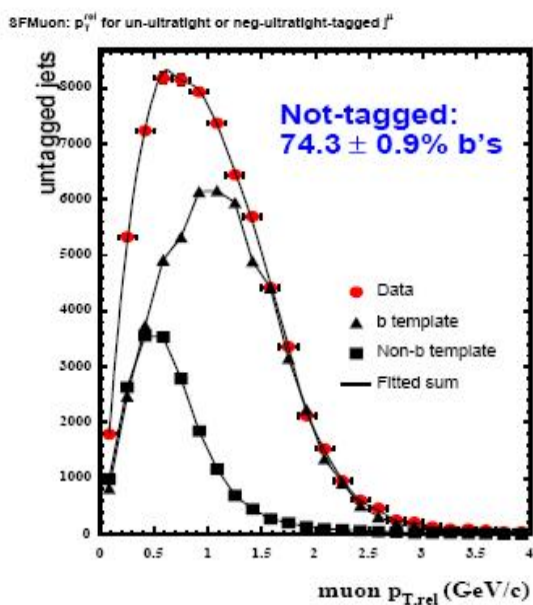
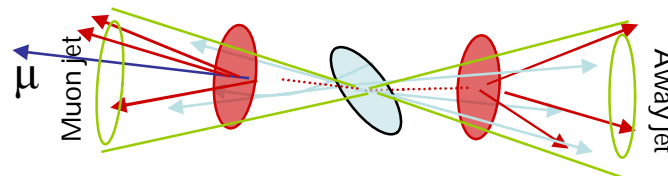
b-jet Identification Efficiency

Measure tag efficiency for reconstructed *b* jets in ALPGEN *W*+*b* jet simulated samples passing event selection and correct with MC-to-data scale factor

$$\epsilon_{\text{tag}}^b = 0.16 \pm 0.01 \text{ (syst)}$$

Measuring the tag efficiency in the data

- Dijet events
- Enhance HF content with UT away-jet
- Require muon in probe jet
- Muon's momentum relative to jet discriminates *b* from non-*b*
- Fits for *b* fraction in tagged, untagged probe jets to extract efficiency



CDF Note 8640

W+b jets: Result

$$\sigma_{b \text{ jets}} \times B(W \rightarrow \nu) = \frac{n_{\text{tag}} \cdot f^b - n_{\text{bkg}}^{b \text{ jets}}}{L \cdot A_{W+b}^{b \text{ jets}} \cdot \epsilon_{\text{tag}}^b \cdot \epsilon} = 2.74 \pm 0.27_{\text{stat}} \pm 0.42_{\text{syst}} \text{ pb}$$

Where do various pieces come from?

Discriminate b jets from c /LF jets in tagged sample using **vertex mass**

$$n_{\text{tag}} \cdot f^b = 670 \pm 44$$

Determine contribution from **background** with tagged b jets and subtract from overall yield

$$n_{\text{bkg}}^{b \text{ jets}} = 177 \pm 22$$

Calculate **acceptance** for $W+b$ -jet events

$$A_{W+b}^{b \text{ jets}} = 0.68 \pm 0.03$$

Measure **tag efficiency** for b jets in $W+b$ -jet production in MC and correct to match data

$$\epsilon_{\text{tag}}^b = 0.16 \pm 0.01$$

Luminous region, trigger efficiency, lepton id

$$\epsilon = 0.88 \pm 0.006$$

Summary of systematics

Source	$\frac{\delta\sigma_{b\text{-jets}} \times BR}{\sigma_{b\text{-jets}} \times BR}$ (%)
<i>b</i> shape modeling	8
<i>c</i> shape modeling	1
LF shape modeling	3
UT tag efficiency	6
Luminosity	6
Top Cross Sections	2
Fake $W^\pm \cancel{E}_T$ fits	1
Tagged Fake W^\pm <i>b</i> fraction	1
Jet Energy Scale	3
Q^2	3
PDF	2
$ z_0 $ efficiency	<1
Trigger efficiency	<1
Lepton ID efficiency	<1

Reduced modeling systematics thanks to ultra-tight b-tag that yields purer sample 6% on 71% b fraction instead of 33% b fraction

Very small uncertainty from non-W thanks to veto, anti-electron method, fit for b-fraction

Conclusions & Future Work

First publication on $W+b$ -jets production from CDF, better than 20% precision with 1.9 fb^{-1}

$$\sigma_{b \text{ jets}} \times B(W \rightarrow \nu) = 2.74 \pm 0.27_{\text{stat}} \pm 0.42_{\text{syst}} \text{ pb}$$

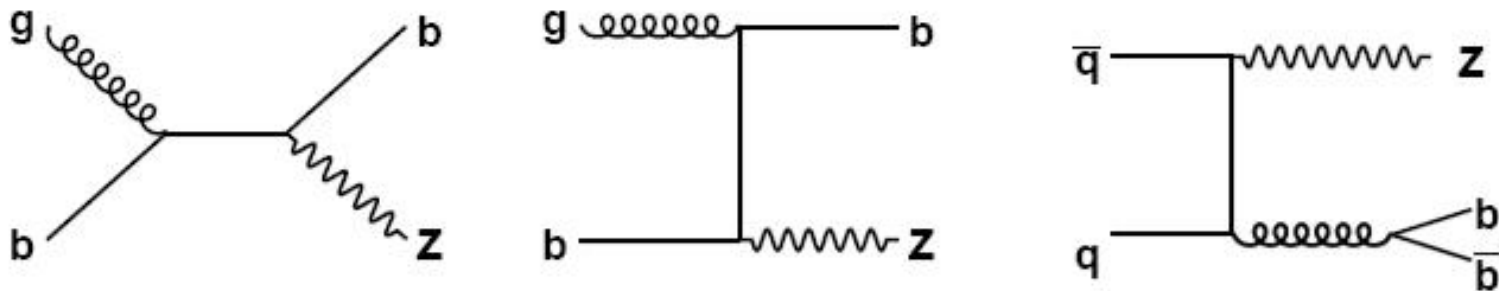
Find 2-3 times larger than predictions

- § MCFM NLO: $1.22 \pm 0.14 \text{ pb}$ (private communication with authors)
- § PYTHIA: 1.10 pb
- § ALPGEN: 0.78 pb (CDF default Q^2 scale)

Next plan to measure differential cross section with 5 fb^{-1}

- § Important test of model of $W+b$ -jets kinematics: an accurate model is critical for robust searches for WH using multivariate techniques that exploit kinematic differences
- § CDF $Z+b$ -jets measurement found indications that default CDF scale set for ALPGEN is not a good model: is the same true for $W+b$ jets?

Measurement of the b -jet Cross Section in Events with a Z Boson in ppbar Collisions at $\sqrt{s} = 1.96$ TeV



Dedicated measurement of production rate and ratio to inclusive Z production, with differential rates, published in **PRD 79, 052008 (2009)**

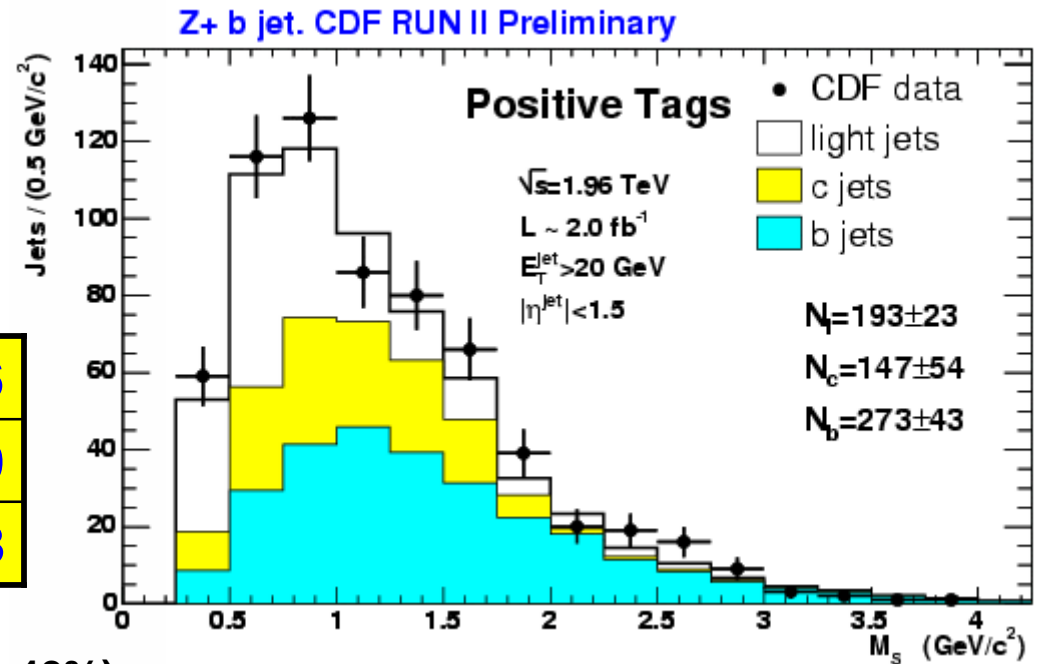
$$(Z+b \text{ jet})/Z = 3.32 \pm 0.53 \text{ (stat)} \pm 0.42 \text{ (syst)} \times 10^{-3}$$

Z+b-jets

§ Dilepton ($ee, \mu\mu$) with invariant mass [76,106] GeV/c^2

§ Jet selection: $E_T > 20 \text{ GeV}$ and $|\eta| < 1.5$ with cone size of 0.7

Selected Z events	295,716
Selected Z+j events	47,450
Tagged jets	648



Using default **SECVTX**, x2 efficiency (30-40%) but x4 contamination from charm and x10 from light flavor relative to ultratight (note “46%” b -jet fraction here)

But can exploit by also fitting negative tags to constrain light flavor

Smaller backgrounds from other processes with b-jets (35.2 ± 8.6 jets) so constrain in fit rather than subtracting after fit as for W+jets

b-jet backgrounds	e^+e^-	$\mu^+\mu^-$
ZZ	6.5 ± 1.3	4.3 ± 0.9
$t\bar{t}$	1.3 ± 0.3	1.4 ± 0.3
$Z \rightarrow \tau^+\tau^- / WW$	0.2 ± 0.1	0.1 ± 0.1
fake lepton	16.4 ± 8.2	5.0 ± 2.2

Z+b-jets: Systematics

Source	Uncertainty (%)
E_T, η model	8.0, 2.8
Track efficiency	5.7
b fragmentation	0.8
b to bb jets, c to cc	3.8
Light flavor	1.7
b-tag efficiency	5.3
Jet energy scale	2.4
Fake lepton bkg	1.9
Other bkg	0.8
Total	12.7

Vertex mass model: 11%
 No independent data
 check for b model. Instead
 re-weight simulation to
 match data jet E_T, η ; also
 vary reconstruction and
 simulation parameters

Results

Ratio of Z+b-jet to Z:

$$\frac{\sigma^{\text{jet}}(Z + b\text{jet})}{\sigma(Z)} = (3.32 \pm 0.53(\text{stat.}) \pm 0.42(\text{syst.})) \times 10^{-3}.$$

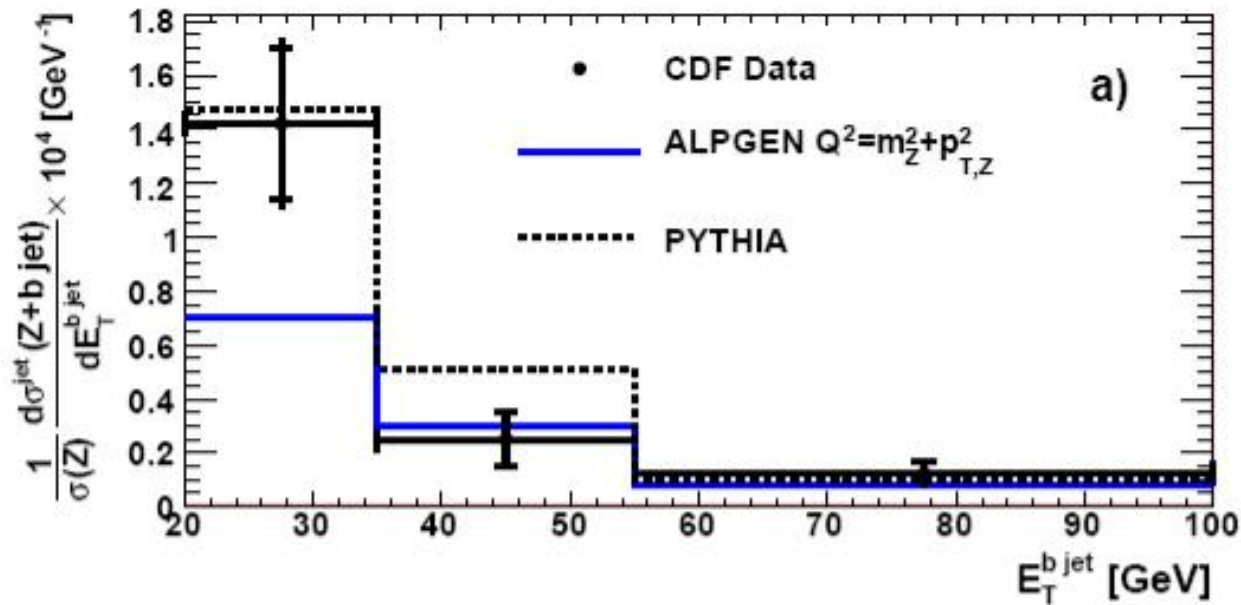
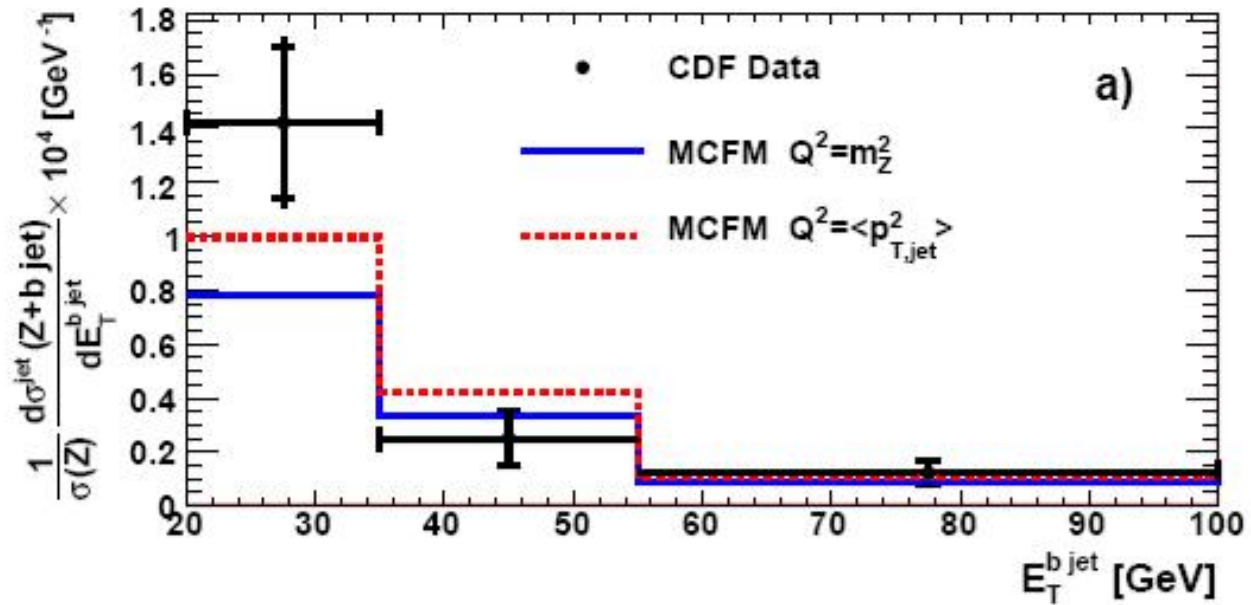
In reasonable agreement with theory, preference for a lower scale

§ MCFM:	2.3×10^{-3}	$Q^2 = M_Z^2 + p_{T,Z}^2$
§ MCFM:	2.8×10^{-3}	$Q^2 = \langle p_{T,\text{jet}}^2 \rangle$
§ ALPGEN:	2.1×10^{-3}	$Q^2 = M_Z^2 + p_{T,Z}^2$
§ PYTHIA:	3.5×10^{-3}	

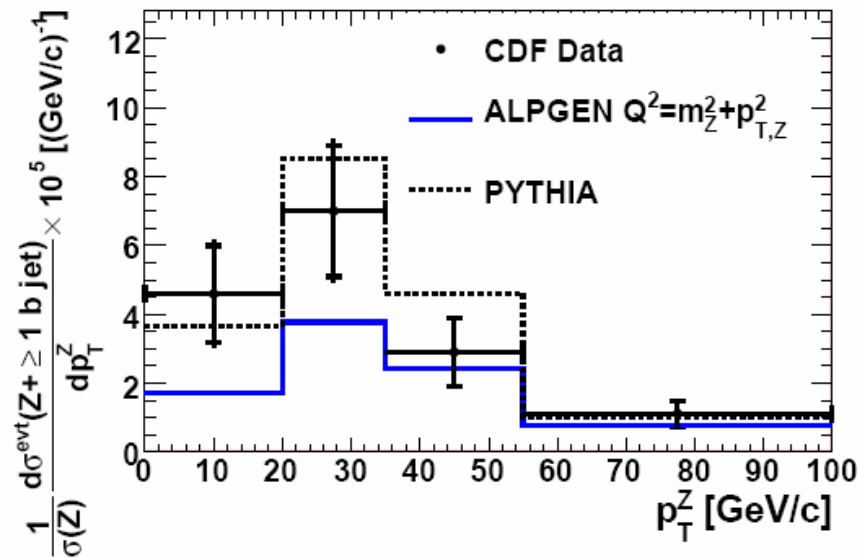
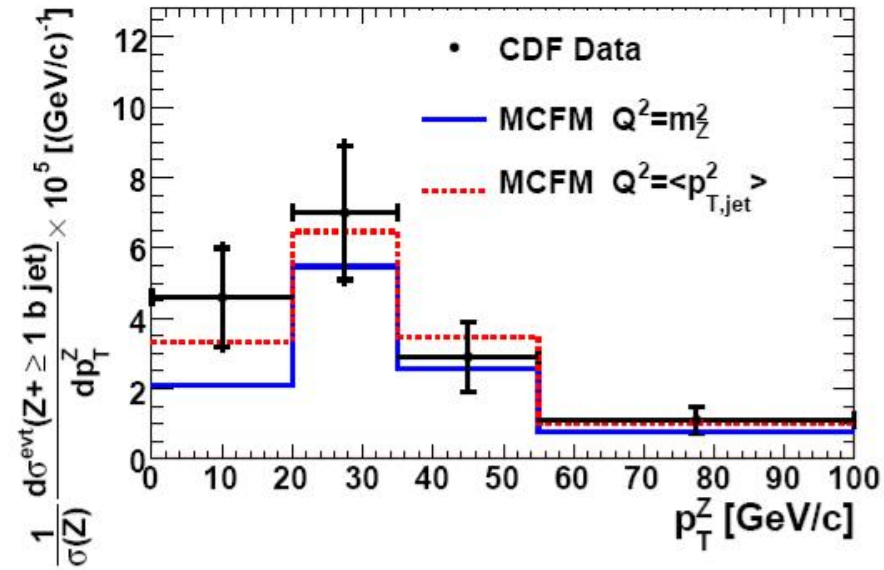
$$\sigma_{b\text{ jets}} \times B(Z \rightarrow \gamma\gamma) = \frac{\sigma^{\text{jet}}(Z + b\text{-jet})}{\sigma(Z)} \times 254.9 = 0.85 \pm 0.14_{\text{stat}} \pm 0.12_{\text{syst}} \text{ pb}$$

For jets with $E_T > 20$ GeV and $|\eta| < 1.5$ and cone of 0.7

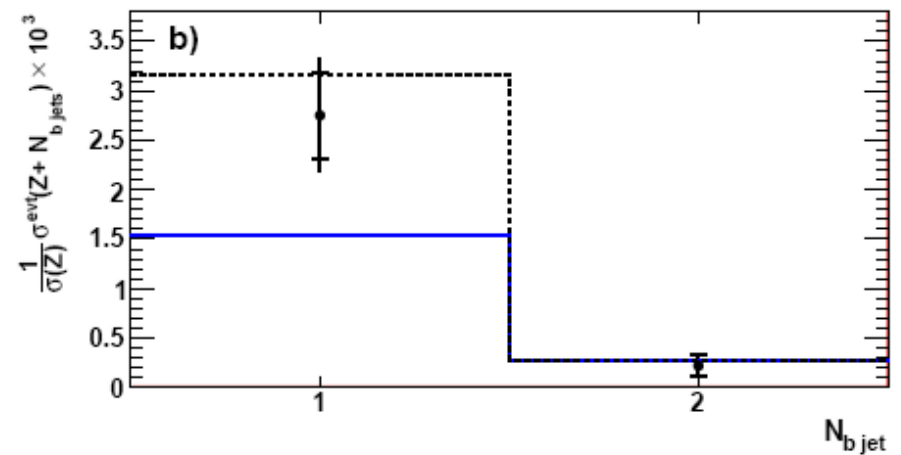
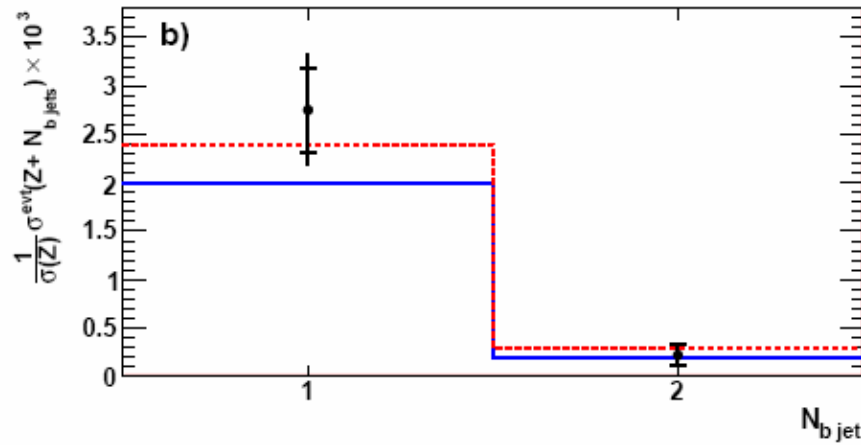
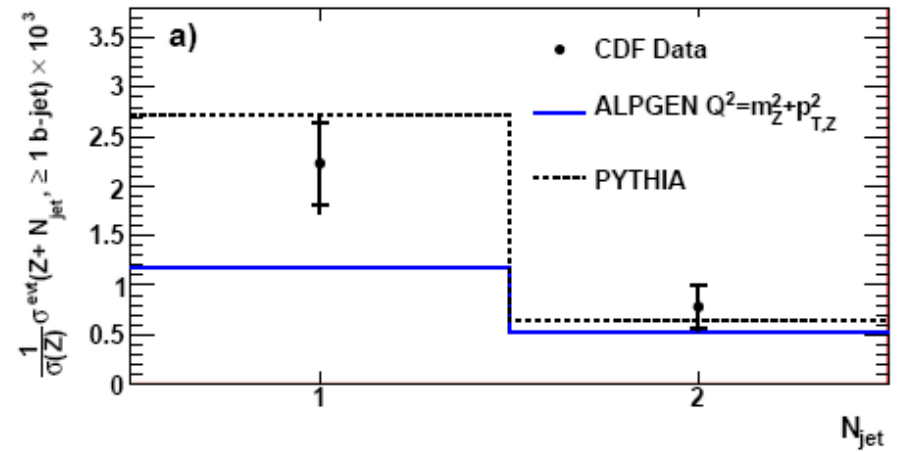
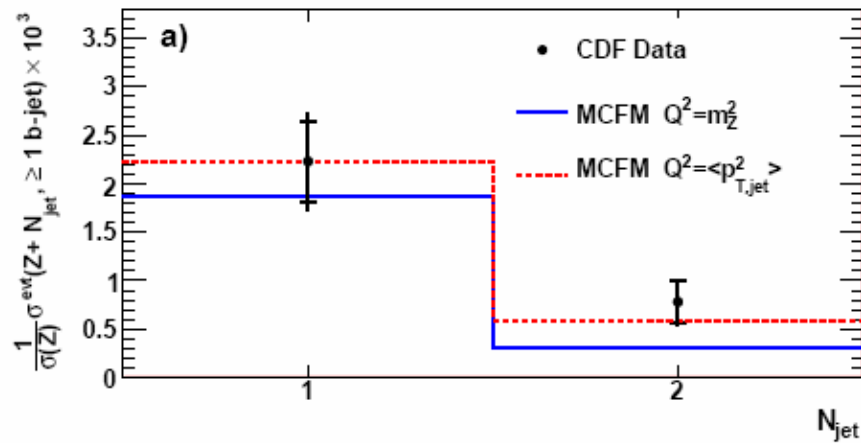
Differential rate vs jet E_T



Differential rate vs Z boson p_T



Differential rate vs Number of jets



Summary

§ **W+b-jets rate 2-3 times higher than predictions**

§ **Differential rates will be measured soon**

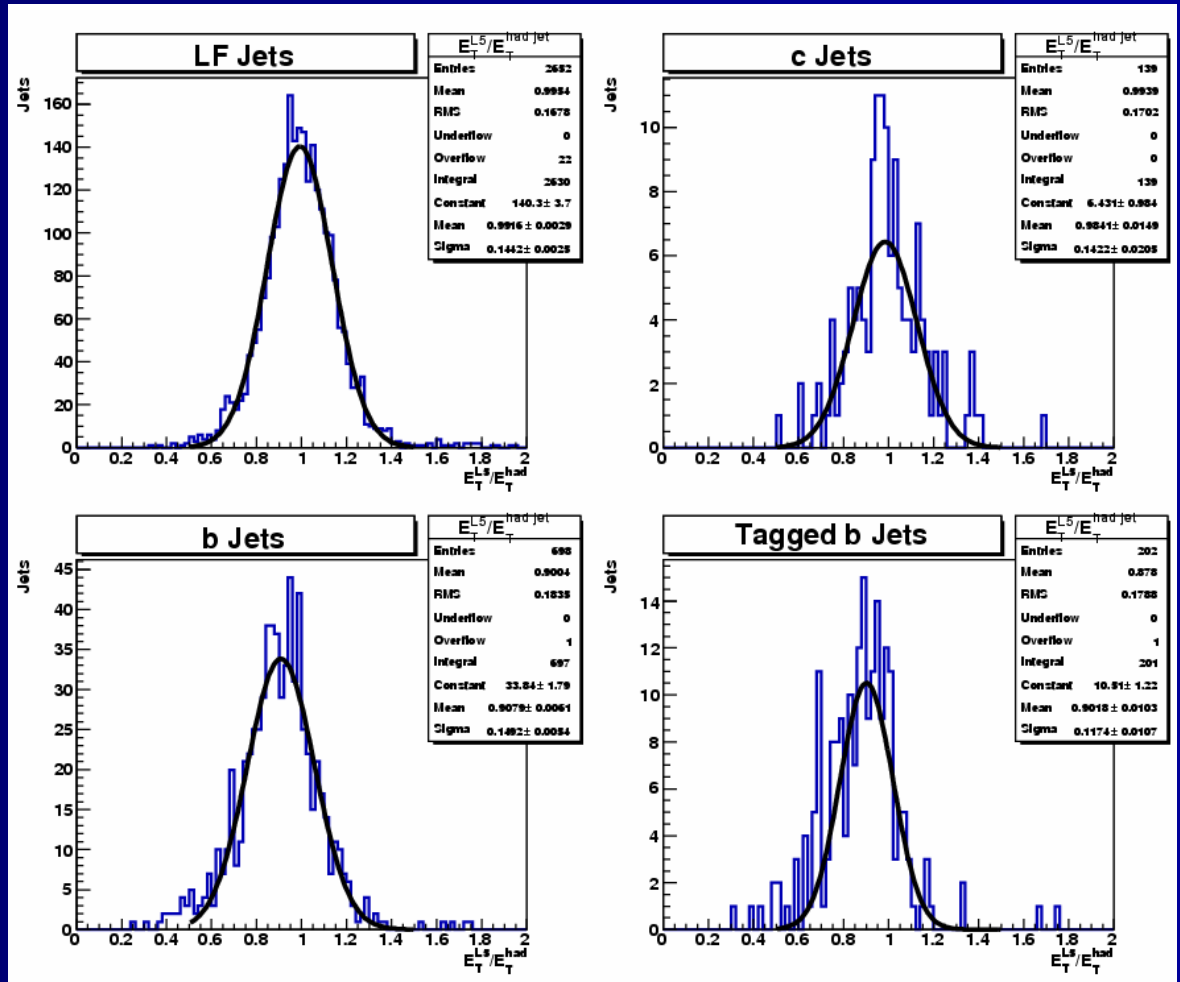
§ **Z+b-jets rate in reasonable agreement with predictions**

§ **Preference for lower scale than $Q^2 = M_Z^2 + p_{T,Z}^2$**

§ **Jet E_T quite different than default hadron collider ALPGEN prediction, which is used to construct multivariate discriminant for ZH search**

Hadron-level Jets

- Jets without the calorimeter!
 - SpartyJet: Software provides jet clustering on raw particles
 - Some knowledge of CDF geometry
 - Glimpse of “truth” jets
- Convention: exclude W daughters but make jets out of everything else
- Natural mismatch wrt measured jet E_T s
 - effect largest for b jets



Measured b-jet energies are ~10% low on average wrt “truth”. Agreement better for LF, c jets.

Veto of Events with Fake W (aka NonW, aka QCD)

Keep event if...

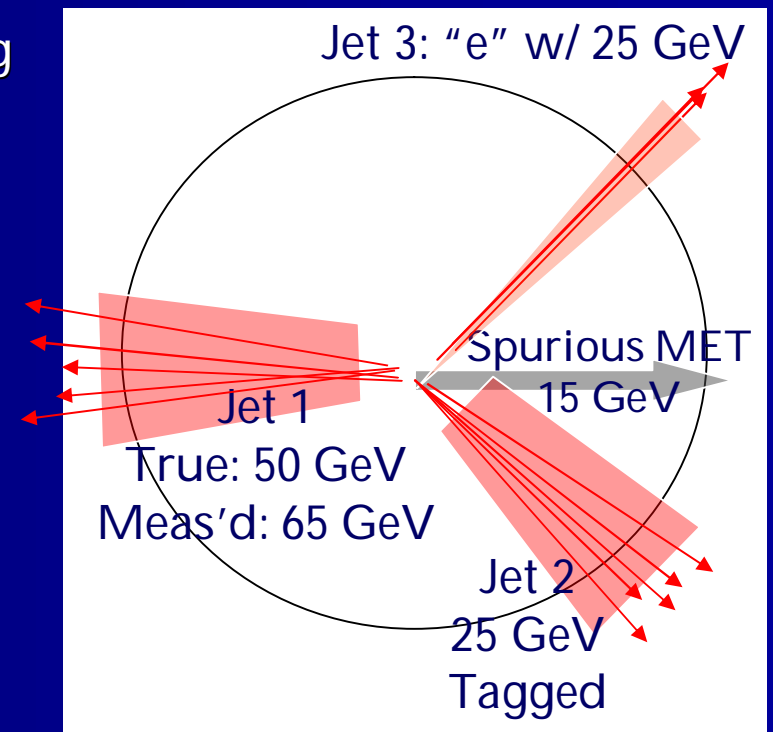
	1 Jet	2 Jet
CEM	<ul style="list-style-type: none"> $M_T(W) > 20 \text{ GeV}$ $S_{\text{MET}} \geq -0.05 * M_T(W) + 3.5$ $S_{\text{MET}} \geq -7.6 + 3.2 * \Delta\phi(\ell, j1)$ 	<ul style="list-style-type: none"> $M_T(W) > 20 \text{ GeV}$ $S_{\text{MET}} \geq -0.05 * M_T(W) + 3.5$ $S_{\text{MET}} \geq 2.5 - 3.125 * \Delta\phi(\text{MET}, j2)$
CMUP	<ul style="list-style-type: none"> $M_T(W) > 10 \text{ GeV}$ $\text{MET} \geq -145 + 60 * \Delta\phi(\ell, j1)$ 	<ul style="list-style-type: none"> $M_T(W) > 10 \text{ GeV}$
CMX	<ul style="list-style-type: none"> $M_T(W) > 10 \text{ GeV}$ 	<ul style="list-style-type: none"> $M_T(W) > 10 \text{ GeV}$

- Seek to eliminate fake W events – mostly QCD multijets – hard to model
- Effective non- W removal developed by Karlsruhe group for 1.5+/fb single top analyses
- Exploits features of fake W events:
 - Low transverse mass of spurious W
 - MET from spurious W is less significant
 - Correlations between jets and leptons and MET

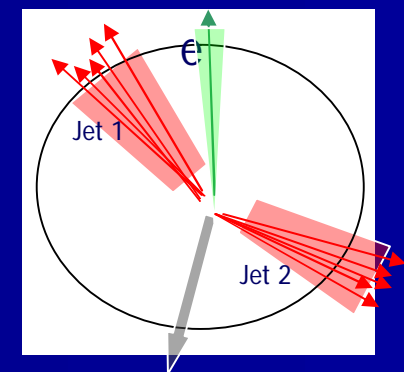
Background Sources of b Jets: b Jets in Fake W Events

- What are fake W events?
 - Mostly QCD multijet production mimicking isolated lepton w/ spurious missing energy from mismeasured jets
 - Tagged jets found elsewhere in the event
 - Characterized by:
 - n small MET
 - n large MET error
 - n small W transverse mass
- Strategy here:
 - Remove as much as possible from the start
 - Model what remains using data
- Model for fake W : “antielectrons”
 - Most fake electrons *just barely* satisfy electron identification
 - Construct a sample of objects that nearly satisfy electron ID - marginal failures

Fake W Event:



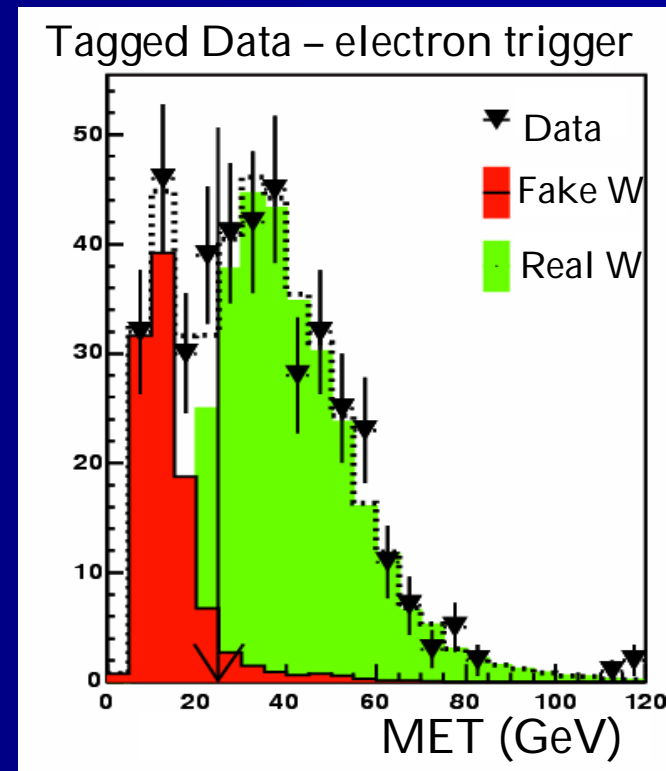
Recall, picture of real W event:



Background Sources of b Jets: b Jets in Fake W Events

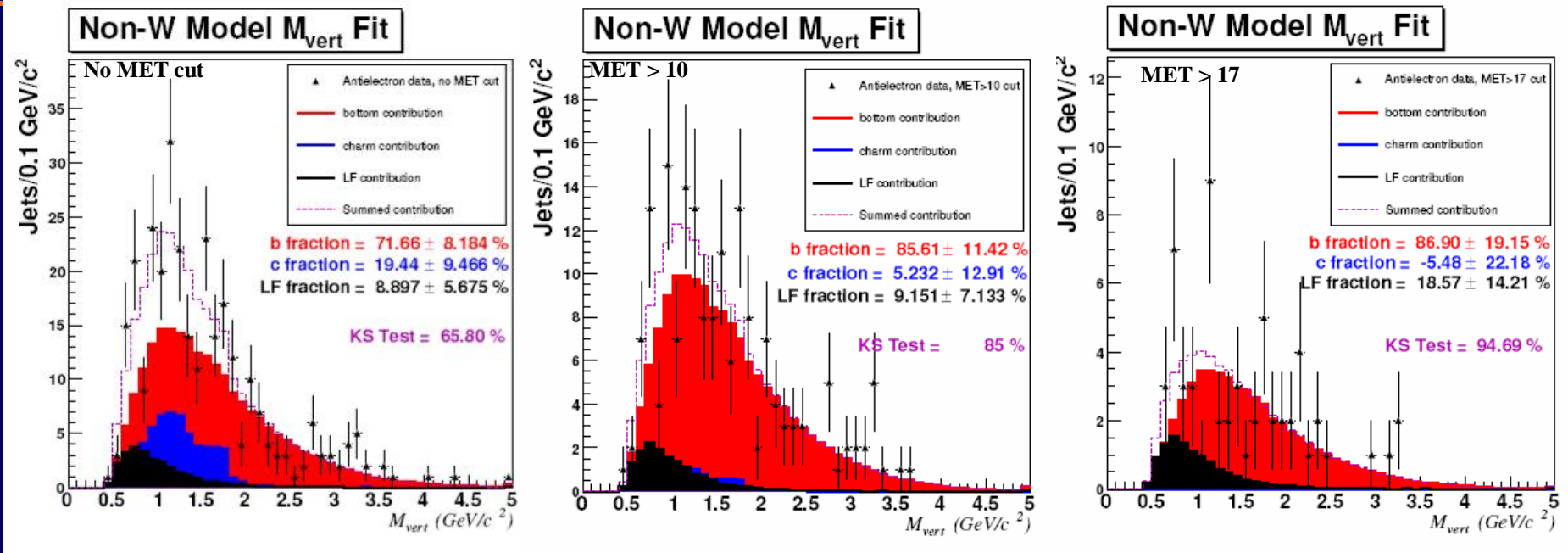
- With model in place can now determine how many tagged jets come from fake W
- Procedure:
 1. Use MET – discriminates between real and fake W events
 2. Relax MET cut (for lever arm)
 3. Fit entire data MET dist to shapes from top, single top, W +jets, Fake- W
 4. Return to MET>25 cut after fit and obtain Fake- W fraction
 5. Fit vertex mass of tagged jets to get b fraction

NB: Here antielectron shape used to model fake W 's in the muon trigger sample as well.



From this fit, fake W is responsible for 2.9% of tagged jets in electron trigger data.

Background Sources of b Jets: b Jets in Fake W Events



- Step 5 fails – insufficient stats in antielectron sample with MET>25
- Step through different MET cuts, examine behavior
- As one tightens the MET cut f_b^{QCD} increases
- Reasonable choice: $f_b^{FakeW} = 0.8 \pm 0.2$

	W + 1 jet	W + 2 jet	W + 1,2 jet
Fake W tags	11.8 +- 3.6	18.8 +- 6.3	30.6 +- 7.4
Fake W tagged b	9.4 +- 3.7	15.1 +- 6.3	24.5 +- 8.4