Measurement of Single Top Quark Production at D0 Using Matrix Elements

---

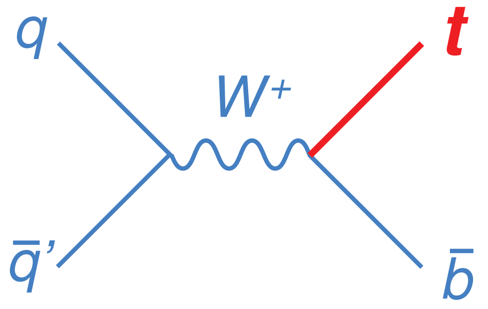
Jovan Mitrevski
Columbia University
(representing the D0 collaboration)

May 7, 2007

Electroweak Top Quark Production

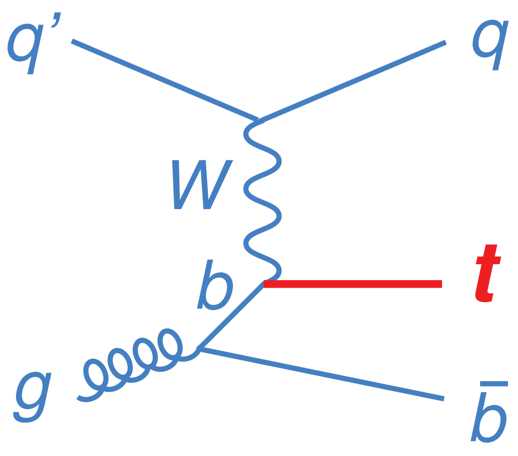
V. M. Abazov *et al.*, Phys. Rev. Lett. **98**, 181802 (2007).

s-channel



$$\sigma_{tb} = 0.88 \pm 0.07\,\mathrm{pb}$$

t-channel



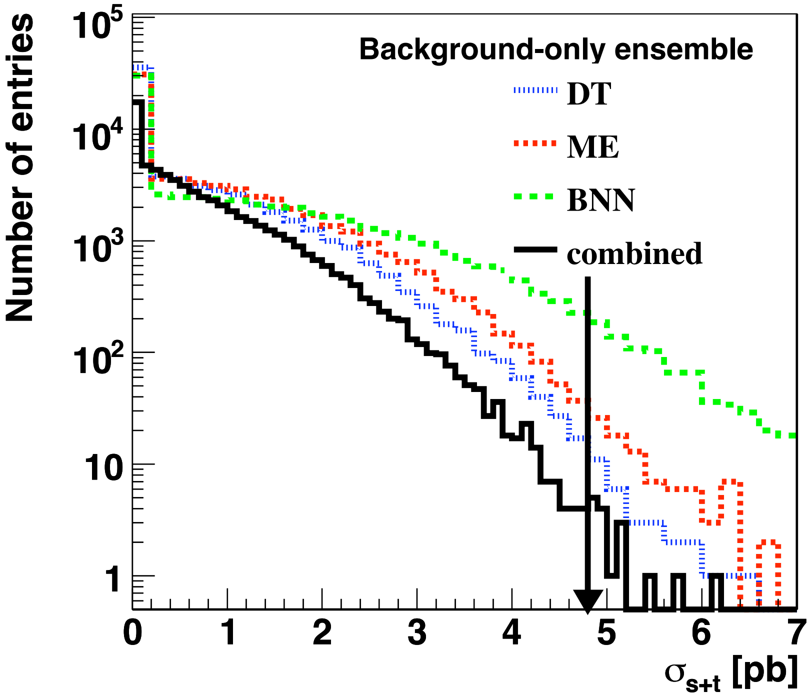
$$\sigma_{tqb} = 1.98 \pm 0.23\,\mathrm{pb}$$

NLO cross sections at m_t = 175 GeV, Phys. Rev. D **70** 114012 (2004)

DØ Results with 0.9 fb^{-1}

Methodology	s+t-channel	observed p-value
BNN (orig.)	$\sigma = 5.0 \pm 1.9$ pb	0.89% (2.4 σ)
ME (orig.)	$\sigma = 4.6^{+1.8}_{-1.5}$ pb	0.21% (2.9 σ)
DT	$\sigma = 4.9 \pm 1.4$ pb	0.04% (3.4 σ)

Combination using BLUE Method:



$$\sigma\left(p\bar{p} \to tb + tqb + X\right) = 4.8 \pm 1.3\,\mathrm{pb}$$

3.5σ significance

The Main Idea of the Matrix Element Method

- Assume a particular process (e.g. t-channel single top, *W*+jets).
  - ➡ The probability density to observe a particular configuration of jets and leptons (*x*) given that process:

$$P(x|\text{process}_i) = \frac{1}{\sigma_i}\frac{d\sigma_i}{dx}$$

- Can use Bayes's Theorem to invert the relation:

$$P(\text{signal}|x) = \frac{P(x|\text{signal})P(\text{signal})}{P(x|\text{signal})P(\text{signal}) + P(x|\text{background})P(\text{background})}$$

- We use the related discriminant:

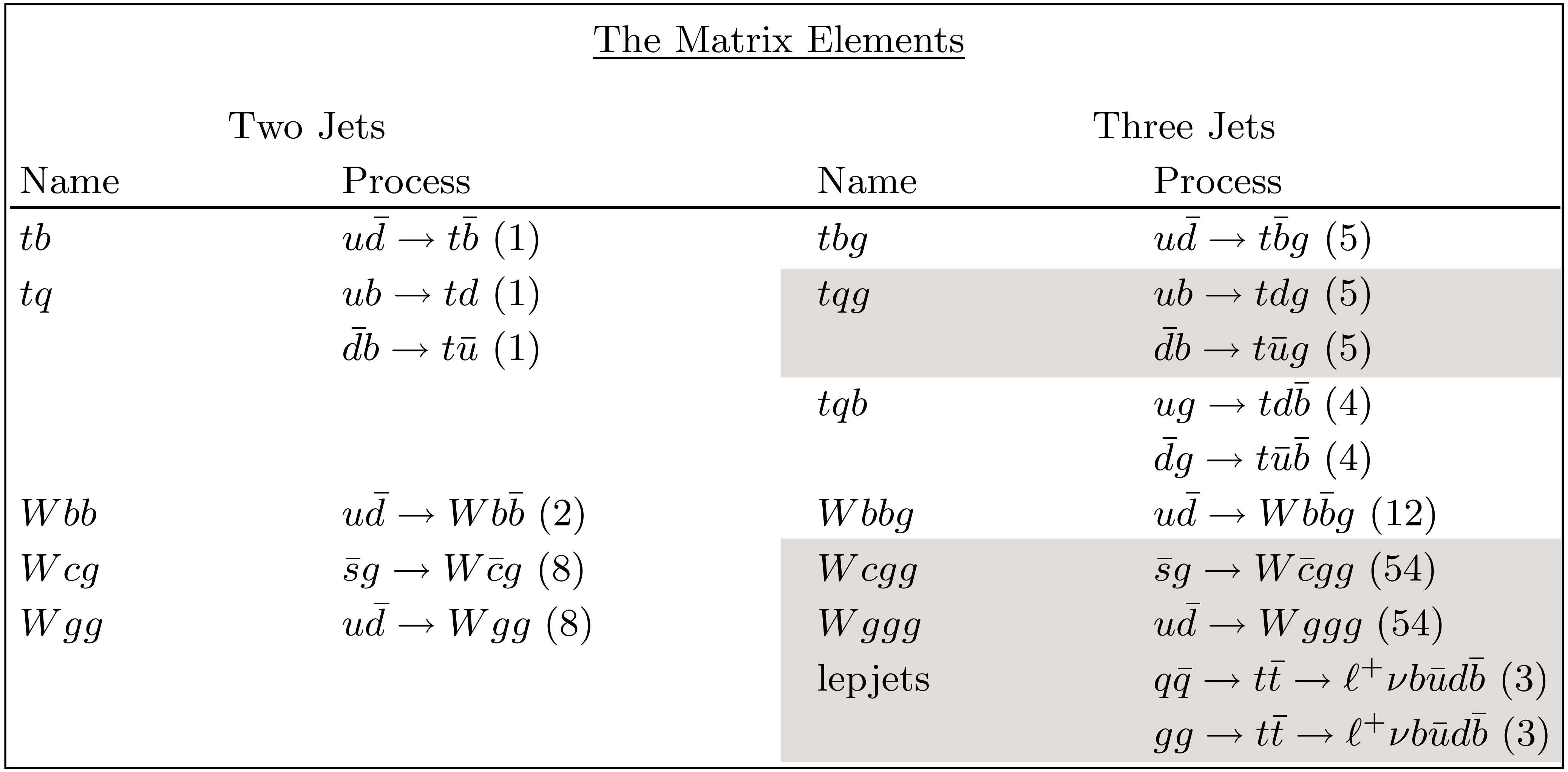
$$D(x) = \frac{P(x|\text{signal})}{P(x|\text{signal}) + P(x|\text{background})}$$

The Differential Cross Section

$$\frac{d\sigma}{dx} = \sum_j \int dy \left[f_{1,j}(q_1, Q^2)\, f_{2,j}(q_2, Q^2)\, \frac{d\sigma_{hs,j}}{dy}\, W_j(x,y)\, \Theta_{\mathrm{parton}}(y) \right]$$

- The event configuration ***x*** is the reconstruction-level event configuration, but the MEs are defined at the parton-level.
- Need to integrate / sum over the parton-level values (***y, j***) to relate them to the reconstruction-level values (***x***). The parts are:
  - The parton-level cross section, containing the MadGraph ME: ***dσ/dy.***
  - The transfer function to relate the parton-level information of the final state particles to the reconstructed objects: ***W***
  - The PDFs to relate the incoming protons to the initial state partons: ***$f_{1,j}$, $f_{2,j}$***
  - Parton-level cuts, if necessary: **Θ**

The Matrix Element Discriminants



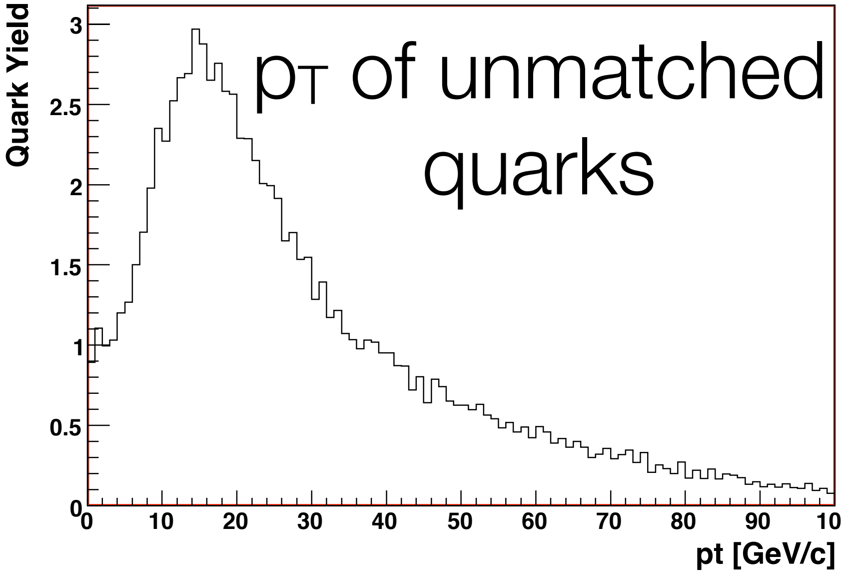
The Matrix Elements

Two Jets		Three Jets	
Name	Process	Name	Process
tb	$u\bar{d} \to t\bar{b}$ (1)	tbg	$u\bar{d} \to t\bar{b}g$ (5)
tq	$ub \to td$ (1)	tqg	$ub \to tdg$ (5)
	$\bar{d}b \to t\bar{u}$ (1)		$\bar{d}b \to t\bar{u}g$ (5)
		tqb	$ug \to td\bar{b}$ (4)
			$\bar{d}g \to t\bar{u}\bar{b}$ (4)
Wbb	$u\bar{d} \to Wb\bar{b}$ (2)	$Wbbg$	$u\bar{d} \to Wb\bar{b}g$ (12)
Wcg	$\bar{s}g \to W\bar{c}g$ (8)	$Wcgg$	$\bar{s}g \to W\bar{c}gg$ (54)
Wgg	$u\bar{d} \to Wgg$ (8)	$Wggg$	$u\bar{d} \to Wggg$ (54)
		lepjets	$q\bar{q} \to t\bar{t} \to \ell^+\nu b\bar{u}d\bar{b}$ (3)
			$gg \to t\bar{t} \to \ell^+\nu b\bar{u}d\bar{b}$ (3)

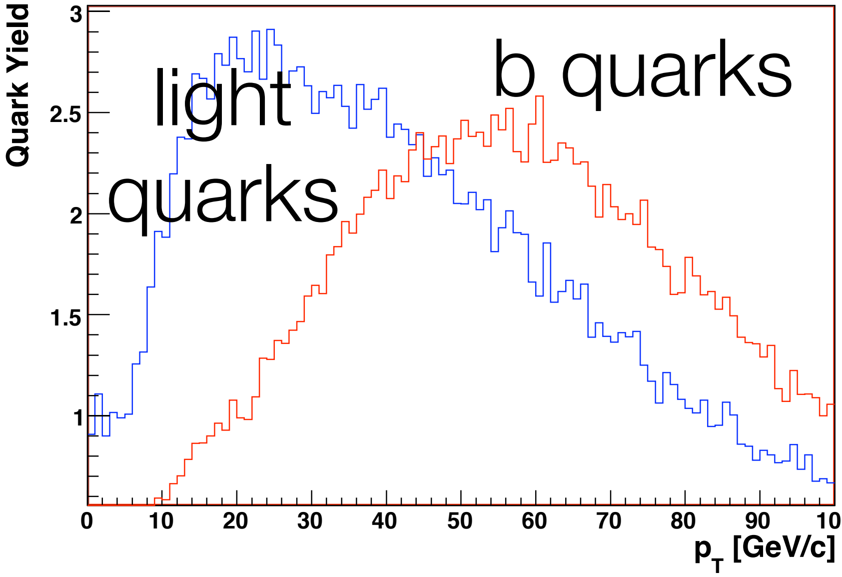
- Also use charge conjugate processes
- Use the same MEs for muon channel, and for different input pairs ($u\bar{d}$, $c\bar{s}$, etc.)
- The main change from the PRL version: extra MEs for 3-jet events (shaded).

A Closer Look at the Lepjets Matrix Element

- In the 3-jet bin, $t\bar{t} \rightarrow \ell$+jets is 22% of the background for single-tag e+jets, and 17% for single-tag μ+jets.
- $t\bar{t} \rightarrow \ell$+jets decays into $\ell\nu b$ quark from one top quark, *qq'b* from the other
  - ➡ 1:1 quark-jet matching: 4-jet bin. For the 3-jet bin, we need to lose a jet.



- looking at our $t\bar{t} \rightarrow$ *e*+jets MC sample, jets are lost without merging 80% of the time, and light quark jets are lost without merging at 1.7 × the rate of the *b*-jets.



- As a simplification:
  - ➡ assume light quark is lost.
  - ➡ In usual case, use transfer function to predict probability to have jet energy below 15 GeV.

Permutation Weights: *B*-Tagging and Muon Charge

- We use *b*-tagging to weigh the different jet-parton assignments differently:

$$W_{btag}(\text{perm}) = \prod_{\text{jets } i} w_{btag}(\text{tag}_i \,|\, \text{flavor}_i, p_{\text{T}i}, \eta_i)$$

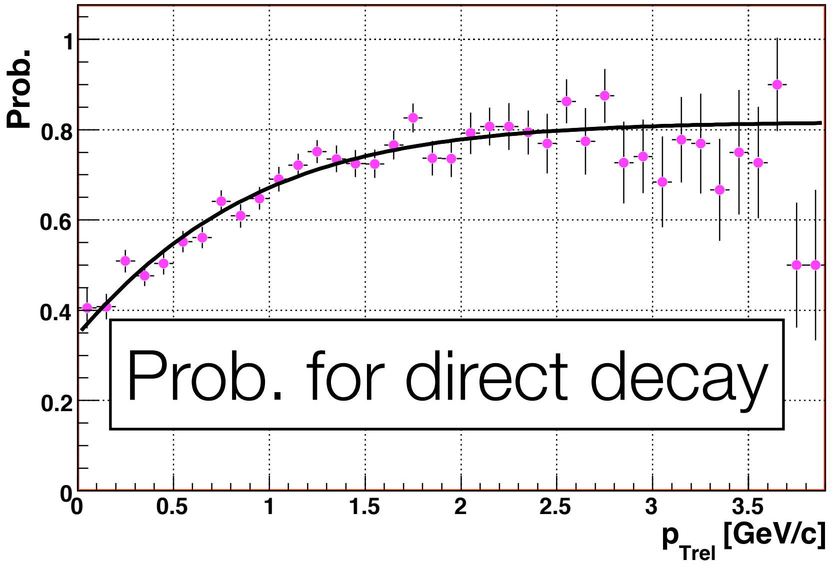
- For example, for the t-channel process, *bu* → *e*$^+$ν*bd*, in the single-tag two-jet bin:

$$W_{btag}(a) = w_{btag}(\text{tagged}|b, p_{\text{T}b}, \eta_b)\, w_{btag}(\text{untagged}|d, p_{\text{T}d}, \eta_d)$$

$$W_{btag}(b) = w_{btag}(\text{tagged}|d, p_{\text{T}d}, \eta_d)\, w_{btag}(\text{untagged}|b, p_{\text{T}b}, \eta_b)$$

- If a *b*-quark decays muonically we can use the muon charge:
  - direct: $\boldsymbol{b} \rightarrow \mu^- \bar{\nu} c$ $\quad \bar{\boldsymbol{b}} \rightarrow \mu^+ \nu \bar{c}$
  - but also: $\boldsymbol{b} \rightarrow \bar{x} c \rightarrow x \bar{x} \mu^+ \bar{\nu} s$ $\quad \bar{\boldsymbol{b}} \rightarrow x \bar{x} \bar{c} \rightarrow x \bar{x} \mu^- \nu \bar{s}$

- Use p_{Trel}, or the p_{T} of the muon relative to the jet. Muons from *c*-quarks tend to have a lower p_{Trel}.

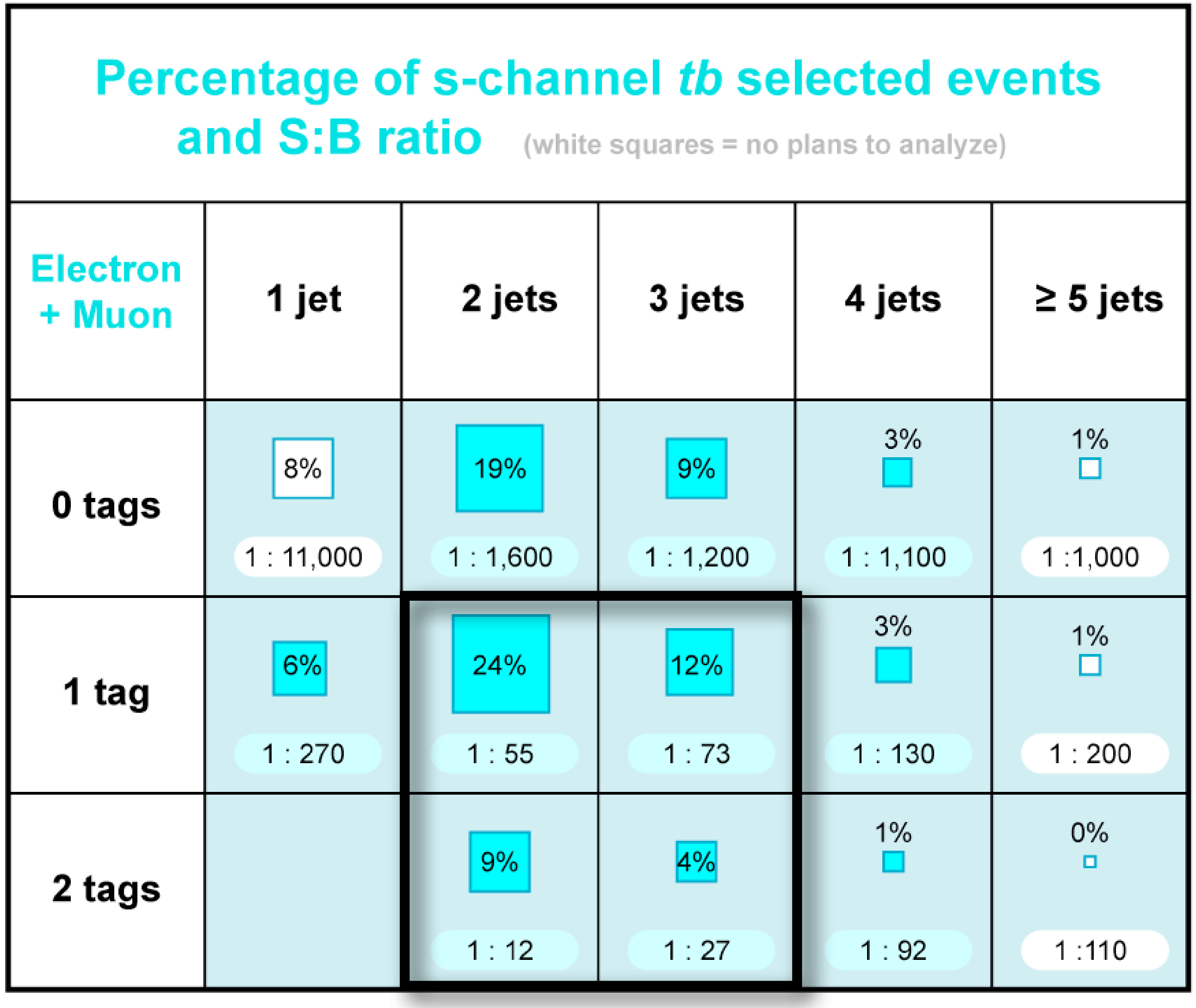


The Selection: Unchanged from PRL

- The same 0.9 fb^{-1} data set as for the PRL.
  - Good data quality
  - Good primary vertex
  - lepton+jets triggered data
  - Leptons: “tight” electron with $p_T > 15$ GeV, $|\eta| < 1.1$,
    or “tight” muon with $p_T > 18$ GeV, $|\eta| < 2.0$.
  - Veto on second charged lepton
  - Jets: leading $p_T > 25$ GeV, second jet $p_T > 20$ GeV, others $p_T > 15$ GeV.
    leading $|\eta| < 2.5$, $|\eta| < 3.4$ for subsequent jets.
  - 15 GeV $< E_T <$ 200 GeV
  - “Triangle” cuts: don’t take events that have the missing E_T aligned or anti-aligned with the lepton or the leading jet

The Analysis Channels

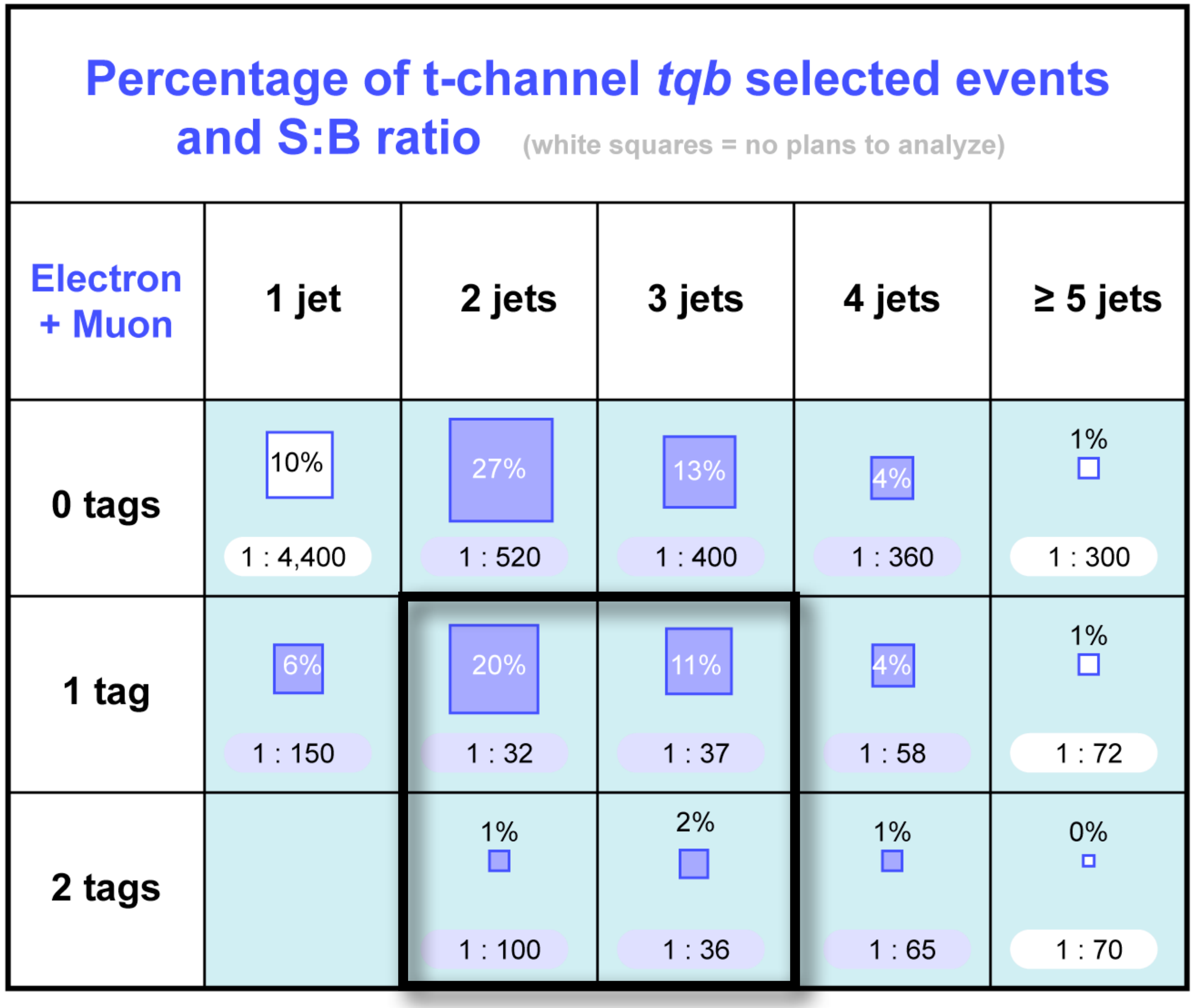
s-channel



Percentage of s-channel *tb* selected events and S:B ratio (white squares = no plans to analyze)

Electron + Muon	**1 jet**	**2 jets**	**3 jets**	**4 jets**	**≥ 5 jets**
0 tags	8% 1 : 11,000	19% 1 : 1,600	9% 1 : 1,200	3% 1 : 1,100	1% 1 :1,000
1 tag	6% 1 : 270	24% 1 : 55	12% 1 : 73	3% 1 : 130	1% 1 : 200
2 tags		9% 1 : 12	4% 1 : 27	1% 1 : 92	0% 1 :110

t-channel

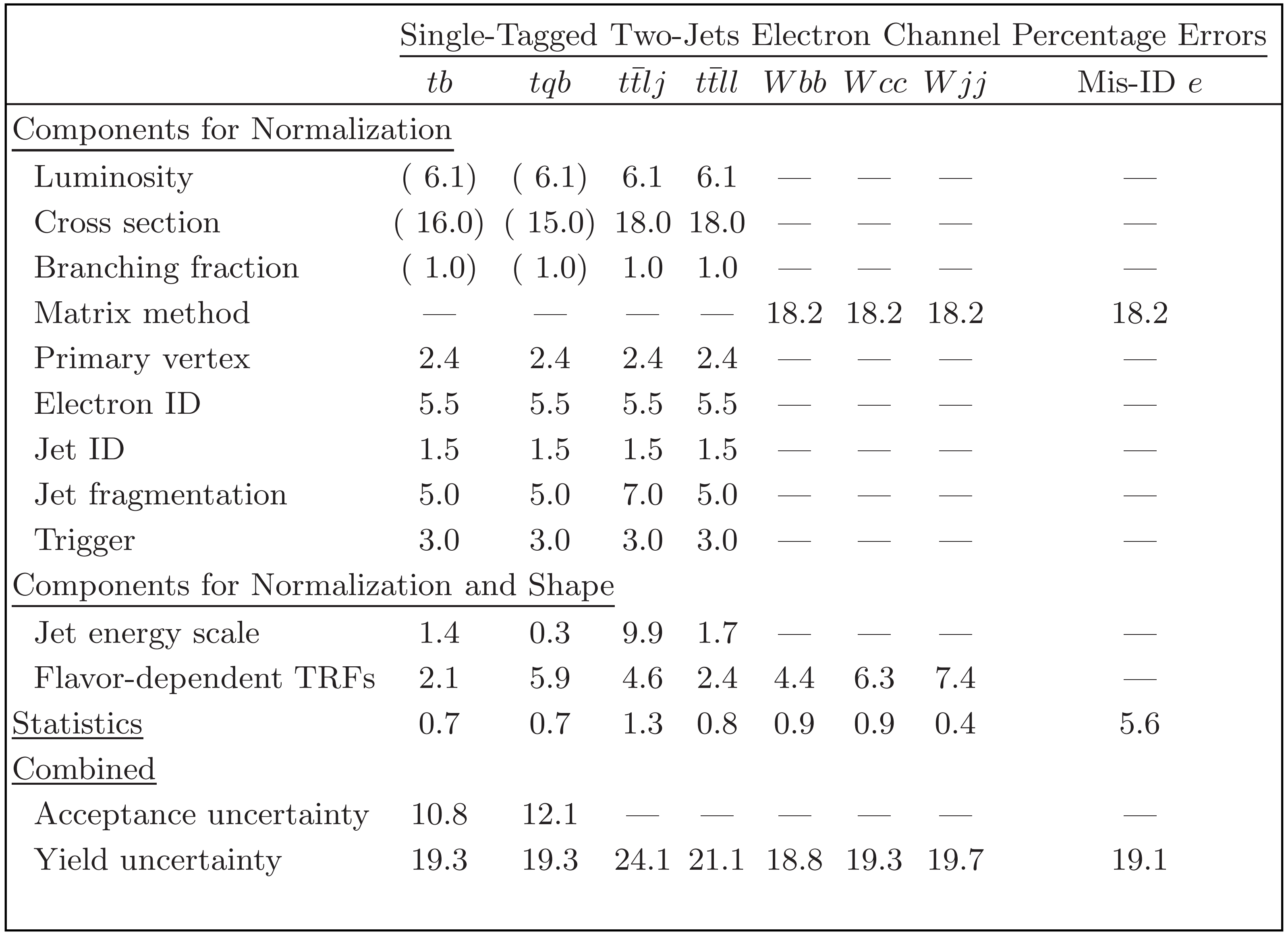


Percentage of t-channel *tqb* selected events and S:B ratio (white squares = no plans to analyze)

Electron + Muon	**1 jet**	**2 jets**	**3 jets**	**4 jets**	**≥ 5 jets**
0 tags	10% 1 : 4,400	27% 1 : 520	13% 1 : 400	4% 1 : 360	1% 1 : 300
1 tag	6% 1 : 150	20% 1 : 32	11% 1 : 37	4% 1 : 58	1% 1 : 72
2 tags		1% 1 : 100	2% 1 : 36	1% 1 : 65	0% 1 : 70

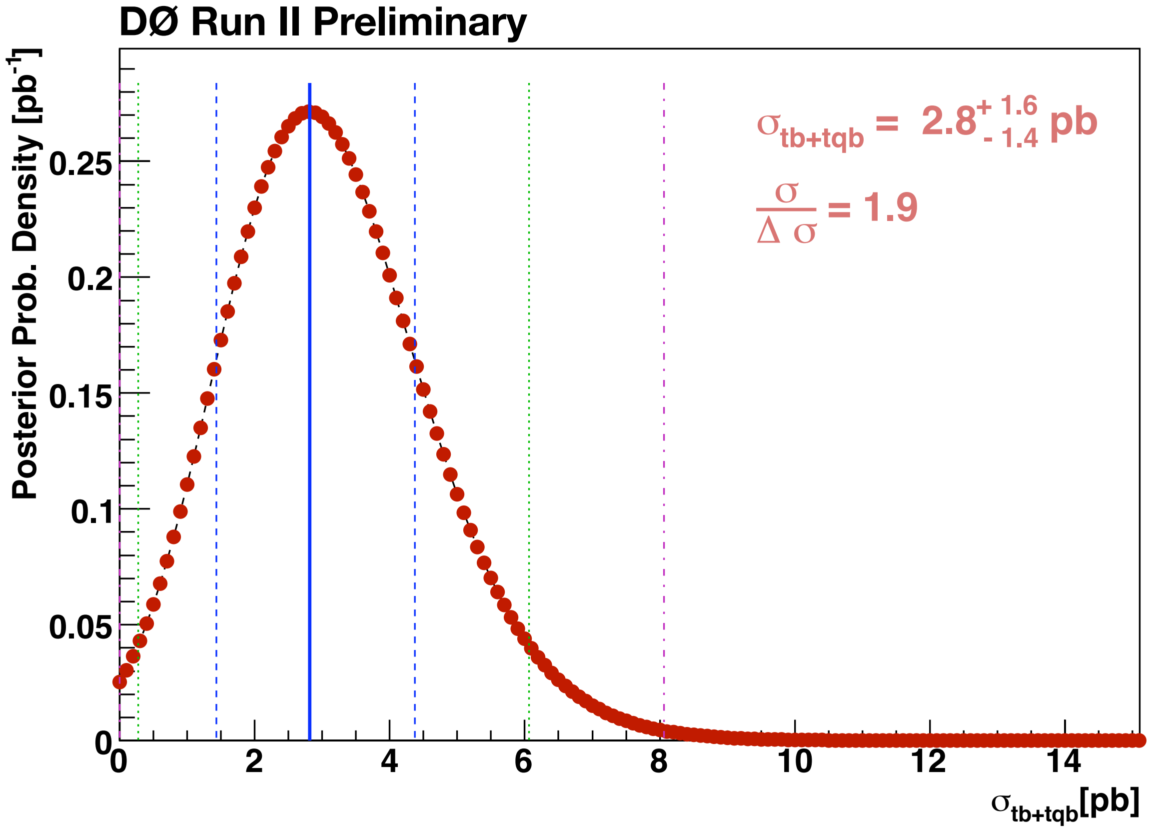
Systematics and Extracting a Result

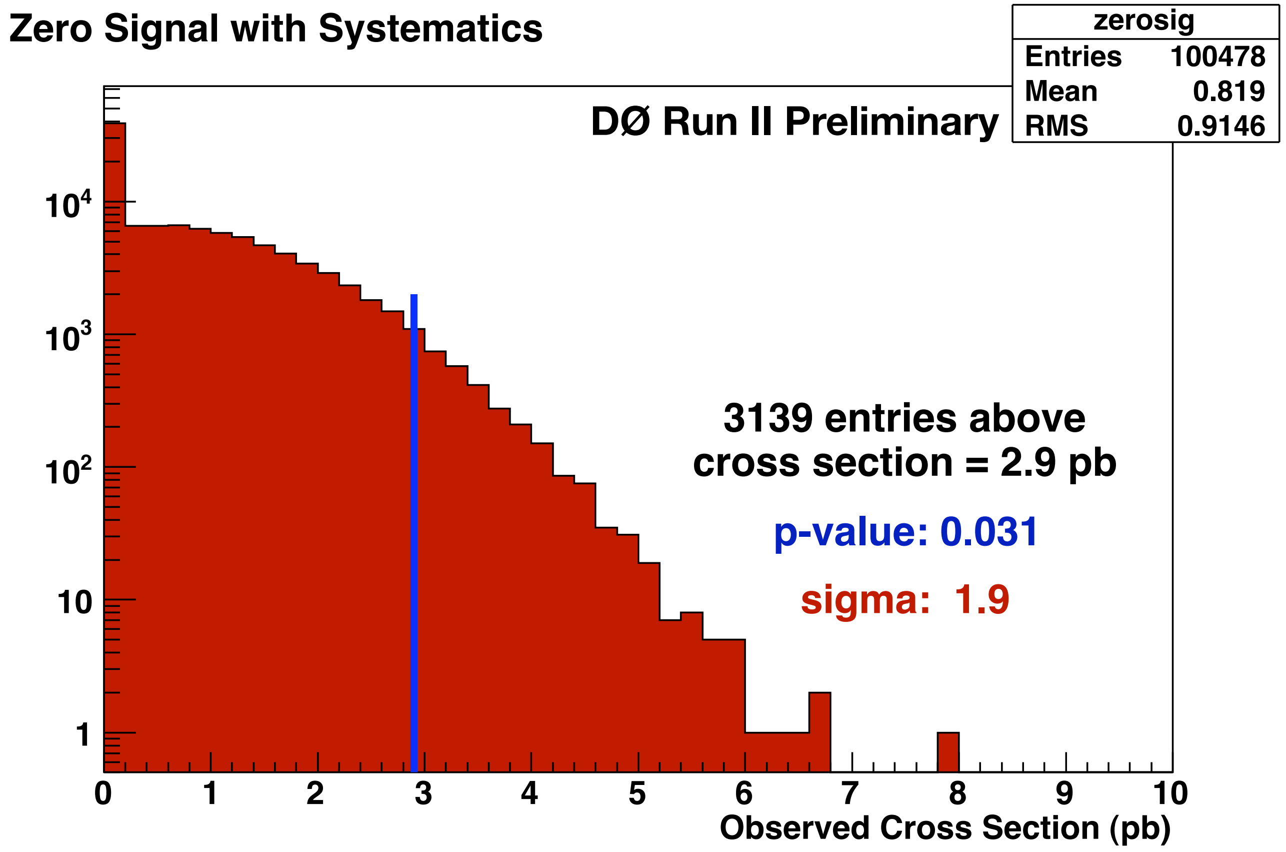
- Build a 2-dimensional histogram: s-disc × t-disc
- Integrate over shifts to yields, acceptances, and luminosity (Gaussian priors) to simulate systematics
  - ➡ Table on the right shows example uncertainties. (We are still statistics dominated.)
- Extract a measurement using a Bayesian approach.



	Single-Tagged Two-Jets Electron Channel Percentage Errors							
	tb	tqb	$t\bar{t}lj$	$t\bar{t}ll$	Wbb	Wcc	Wjj	Mis-ID e
Components for Normalization								
Luminosity	(6.1)	(6.1)	6.1	6.1	—	—	—	—
Cross section	(16.0)	(15.0)	18.0	18.0	—	—	—	—
Branching fraction	(1.0)	(1.0)	1.0	1.0	—	—	—	—
Matrix method	—	—	—	—	18.2	18.2	18.2	18.2
Primary vertex	2.4	2.4	2.4	2.4	—	—	—	—
Electron ID	5.5	5.5	5.5	5.5	—	—	—	—
Jet ID	1.5	1.5	1.5	1.5	—	—	—	—
Jet fragmentation	5.0	5.0	7.0	5.0	—	—	—	—
Trigger	3.0	3.0	3.0	3.0	—	—	—	—
Components for Normalization and Shape								
Jet energy scale	1.4	0.3	9.9	1.7	—	—	—	—
Flavor-dependent TRFs	2.1	5.9	4.6	2.4	4.4	6.3	7.4	—
Statistics	0.7	0.7	1.3	0.8	0.9	0.9	0.4	5.6
Combined								
Acceptance uncertainty	10.8	12.1	—	—	—	—	—	—
Yield uncertainty	19.3	19.3	24.1	21.1	18.8	19.3	19.7	19.1

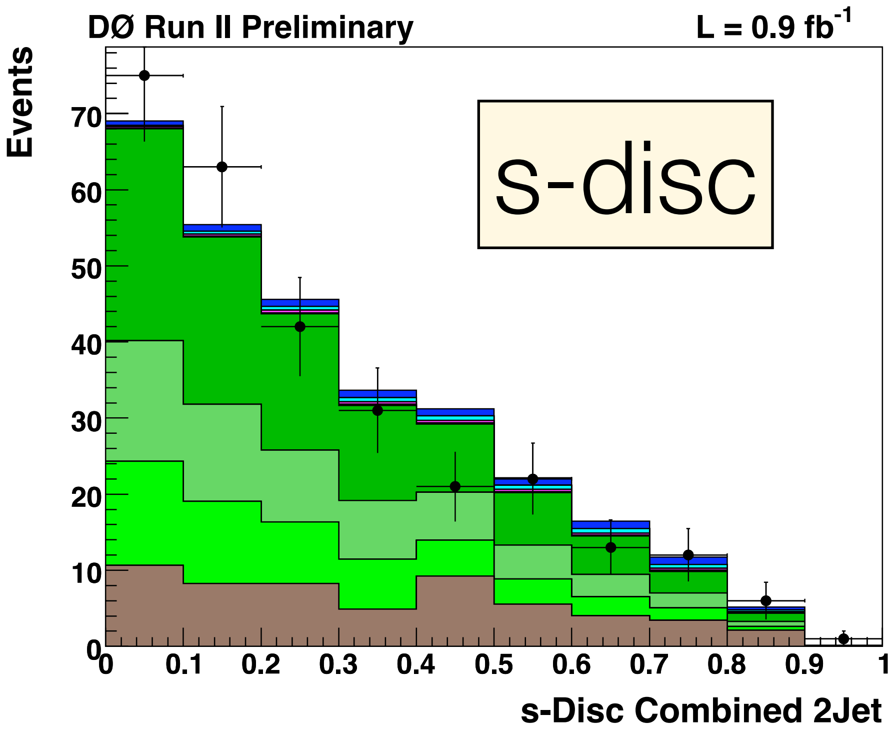
Expected Results

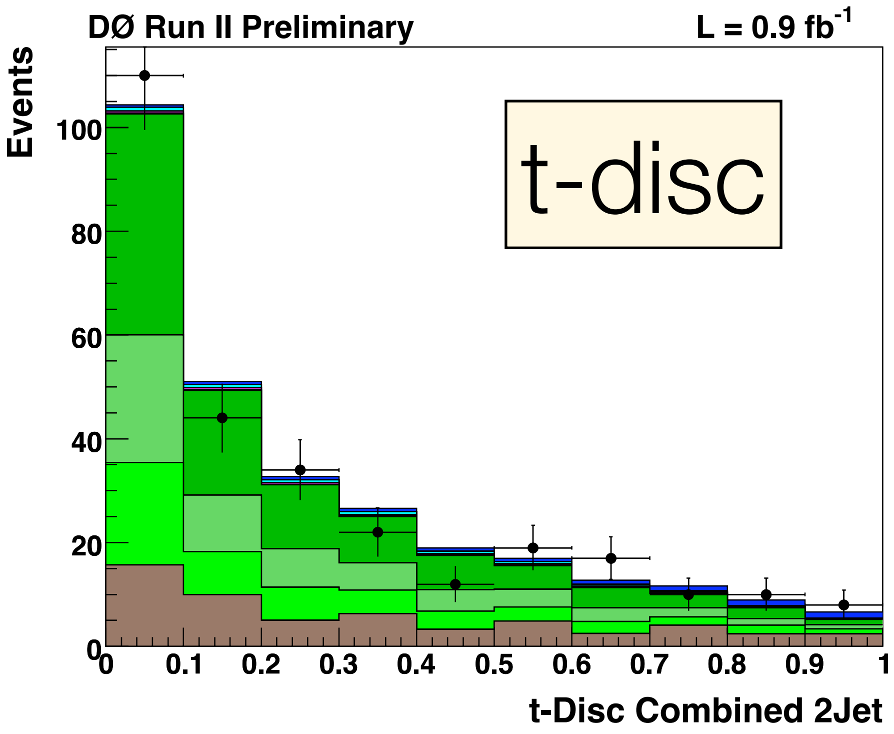




- We get back the Standard Model value of the cross section when we set the “data” to the background + SM signal yield.
- Expected significance: 1.9σ. There is a 3.1% chance for background only to result in a measurement of 2.9 pb or higher.

Cross Check Plots



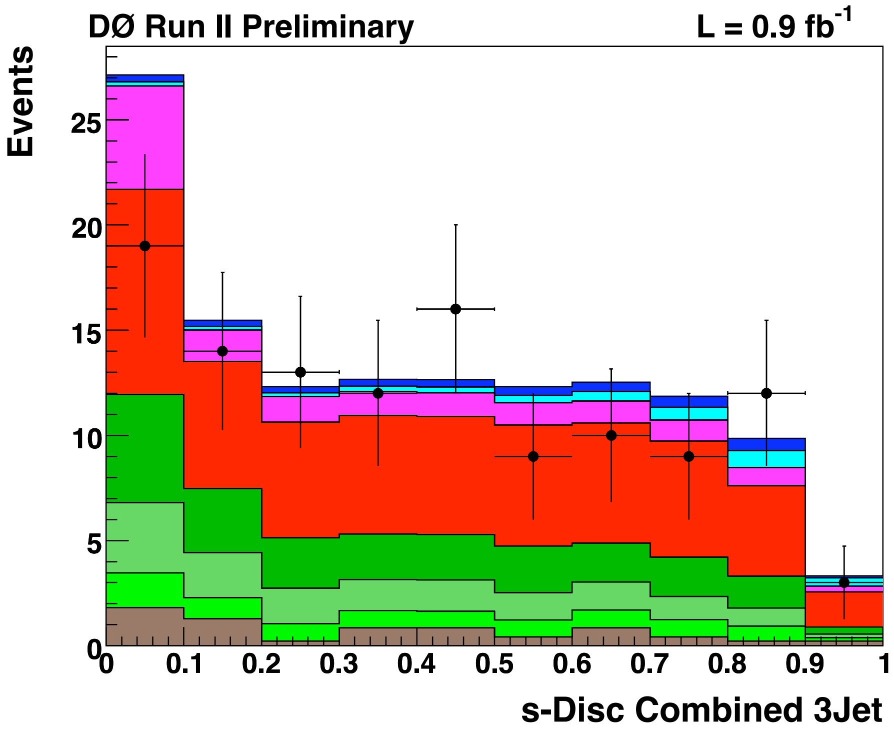


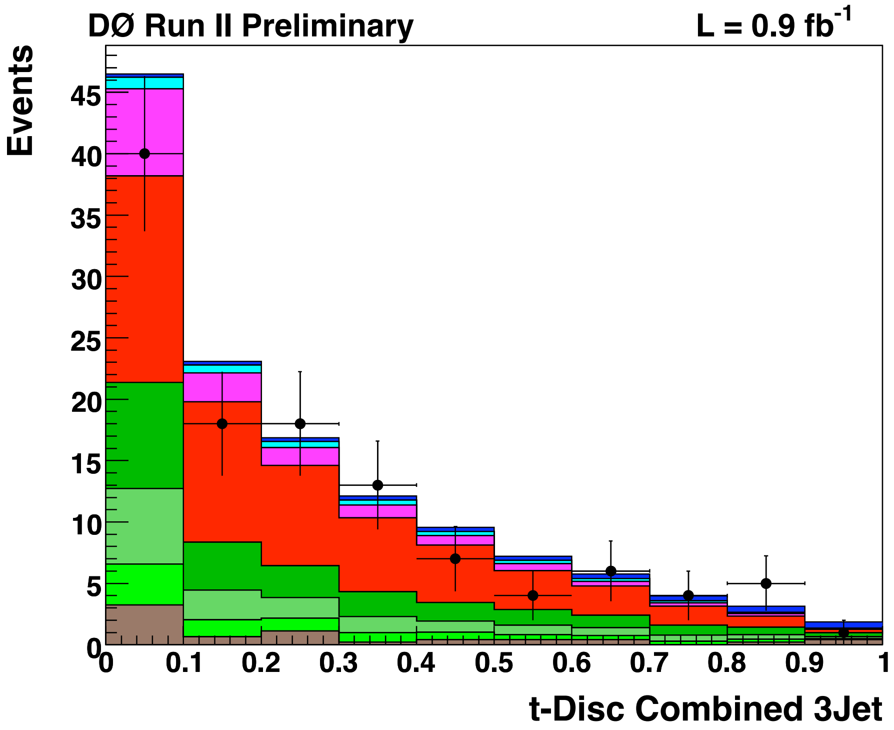
H_T < 175 GeV, 2 jets

Key

- 🟫 QCD
- 🟩 W + light jets
- 🟩 Wcc + jets
- 🟩 Wbb + jets
- 🟥 $t\bar{t} \rightarrow$ lep + jets
- 🟪 $t\bar{t} \rightarrow$ dilepton
- 🟦 signal: tb
- 🟦 signal: tqb
- • DATA

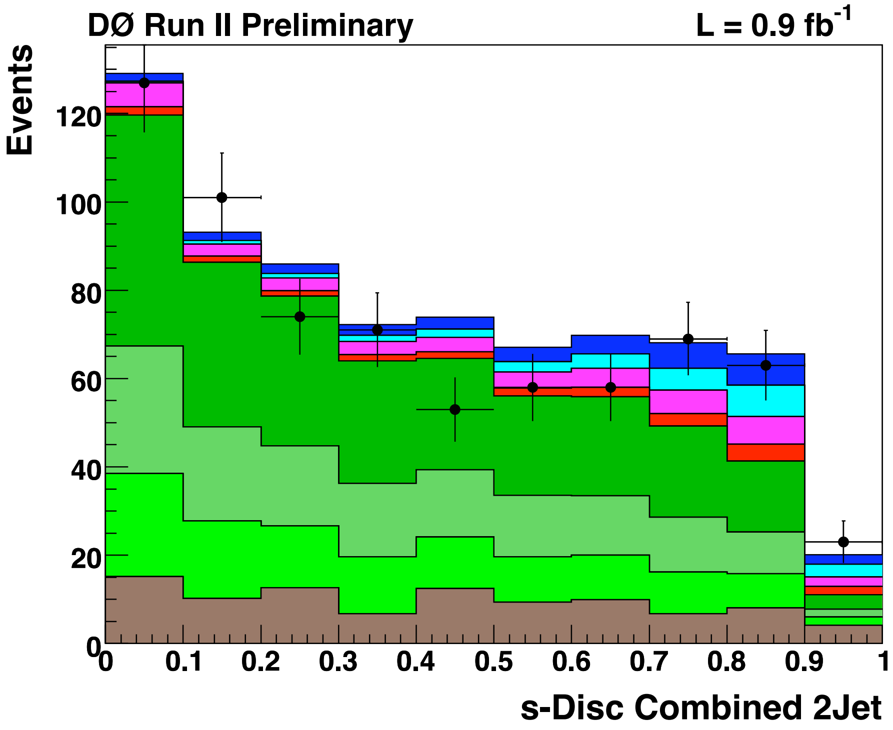
H_T > 300 GeV, 3 jets



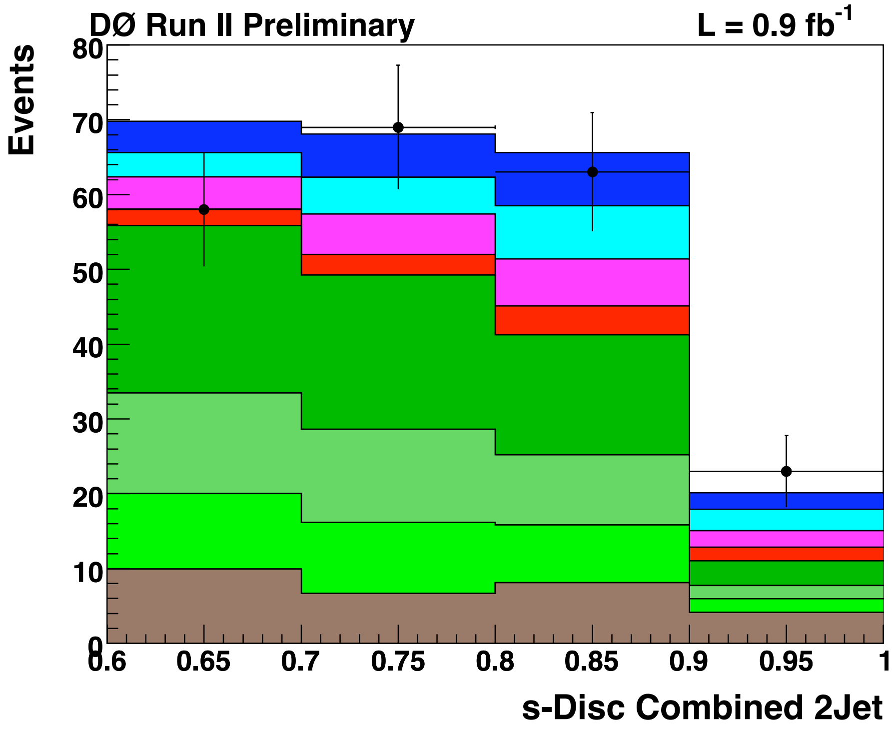


Discriminant Results (2 Jets)

Single top scaled to measured cross section.

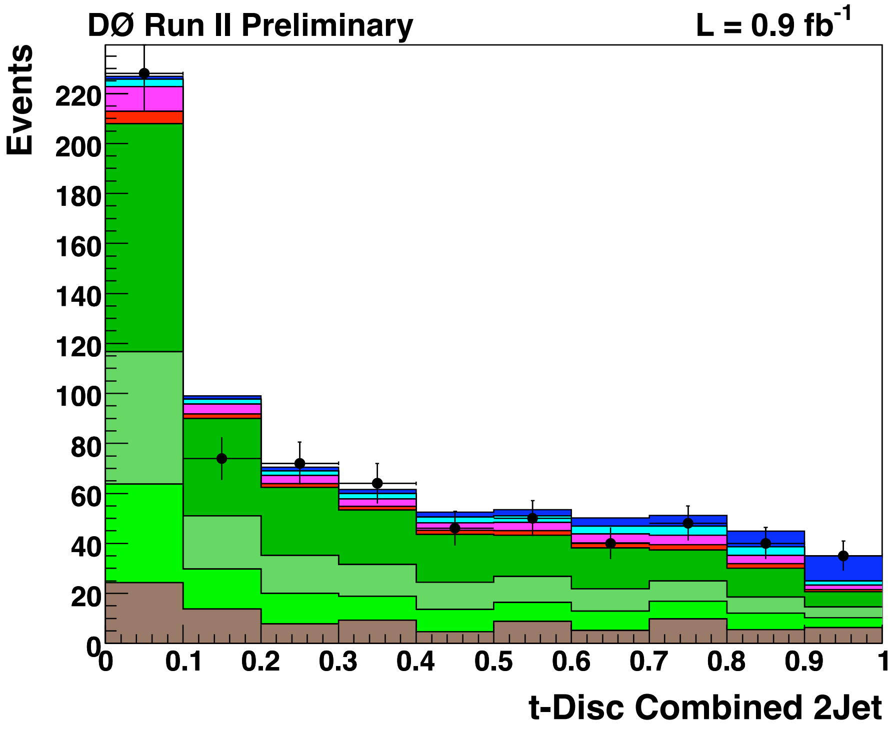


s disc

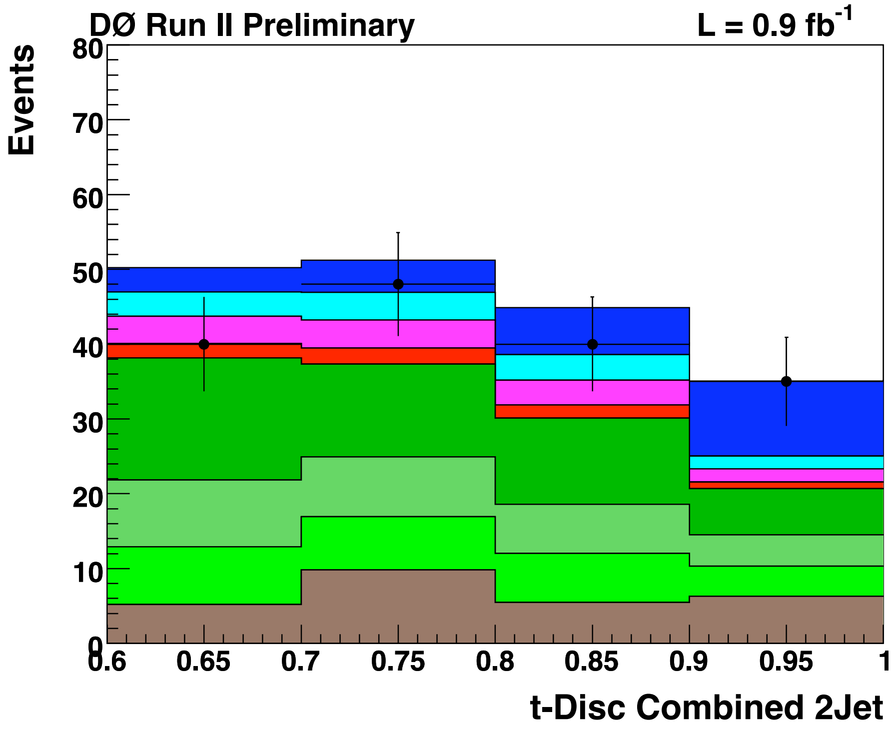


Key

🟫 QCD
🟩 W + light jets
🟩 Wcc + jets
🟩 Wbb + jets
🟥 $t\bar{t} \to$ lep + jets
🟪 $t\bar{t} \to$ dilepton
🟦 signal: tb
🟦 signal: tqb
● DATA



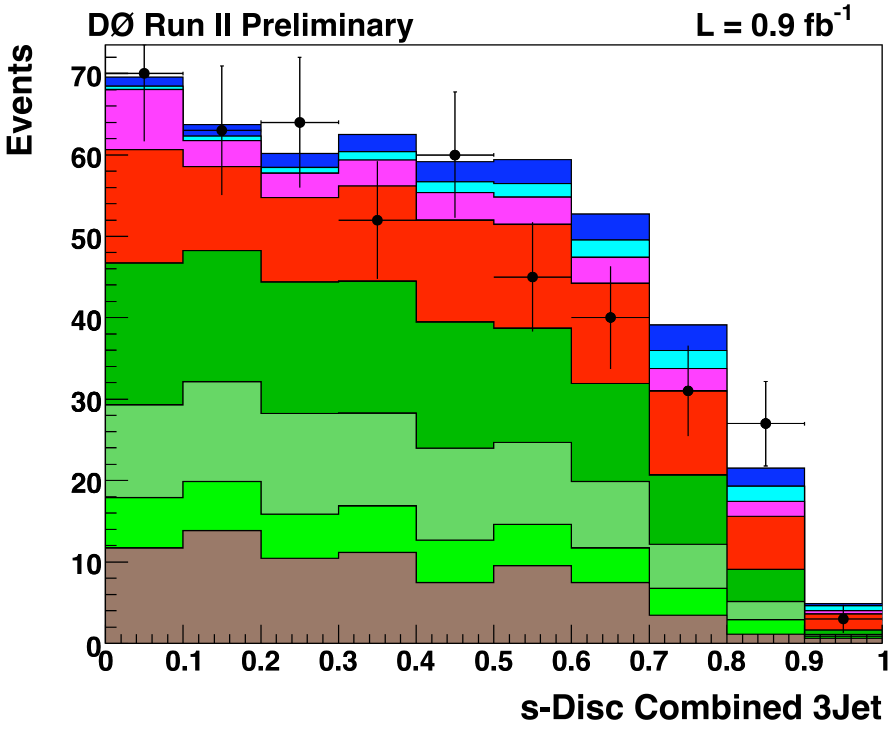
t disc

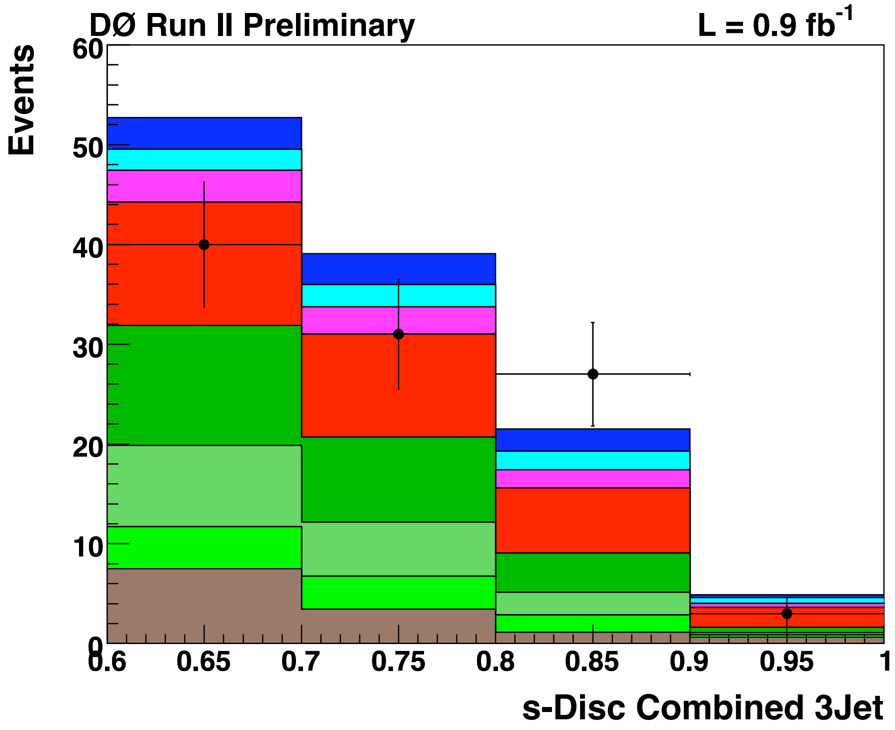


Discriminant Results (3 Jets)

Single top scaled to measured cross section.

s disc

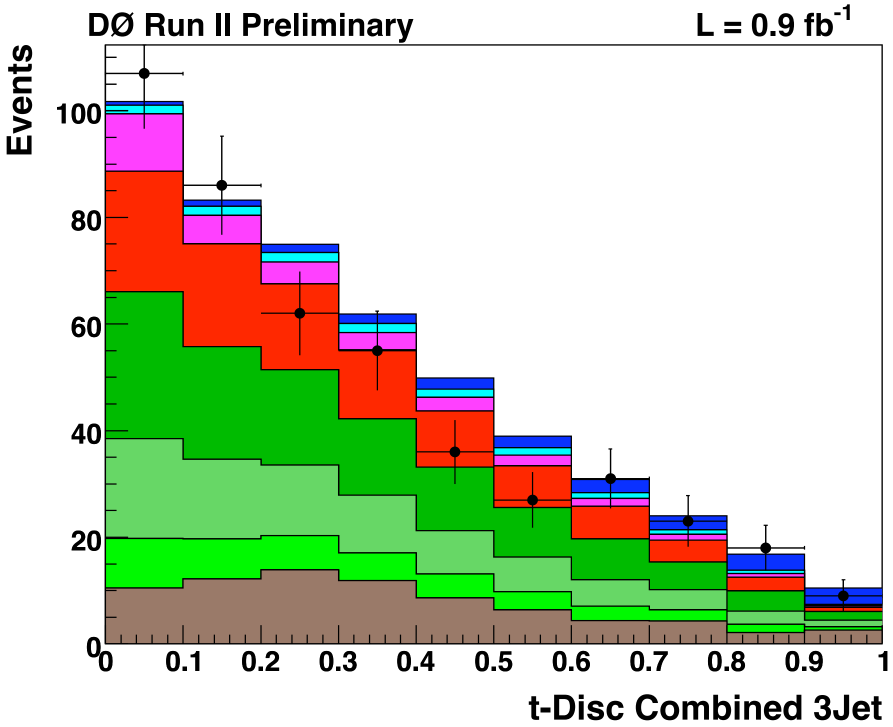


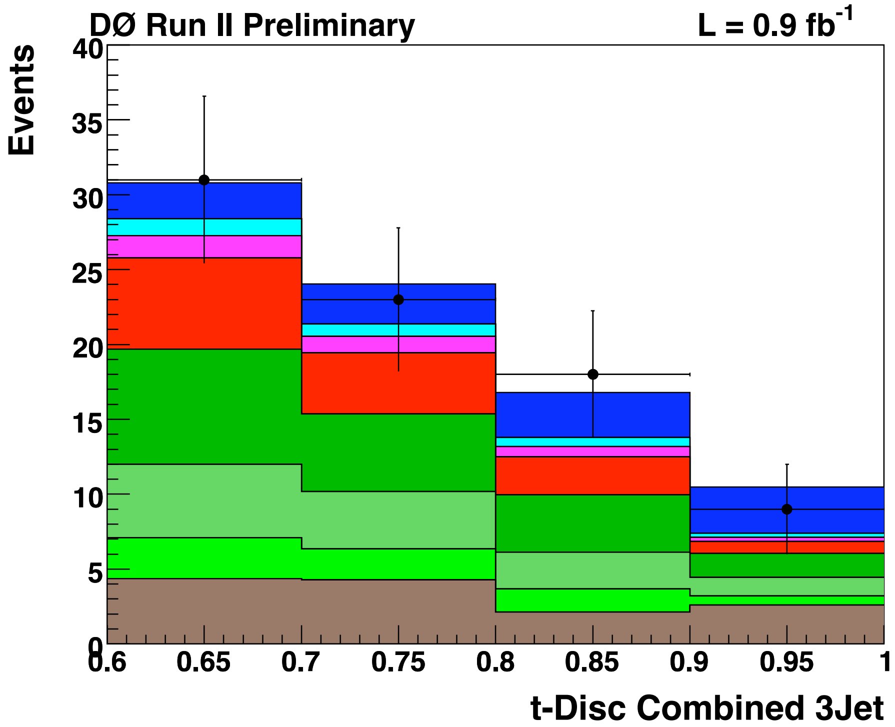


Key

- 🟫 QCD
- 🟩 W + light jets
- 🟩 Wcc + jets
- 🟩 Wbb + jets
- 🟥 $t\bar{t} \rightarrow$ lep + jets
- 🟪 $t\bar{t} \rightarrow$ dilepton
- 🟦 signal: tb
- 🟦 signal: tqb
- ● DATA

t disc

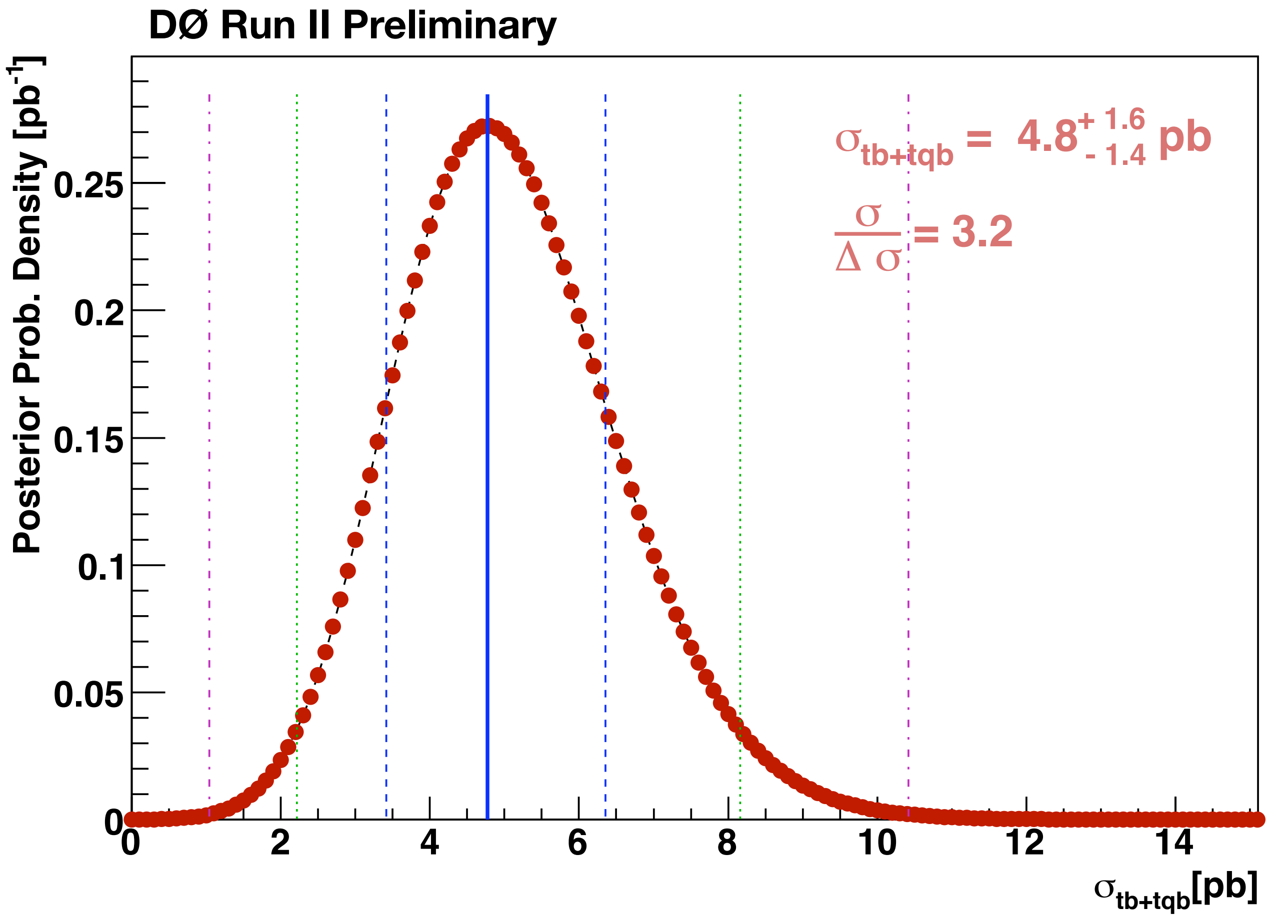




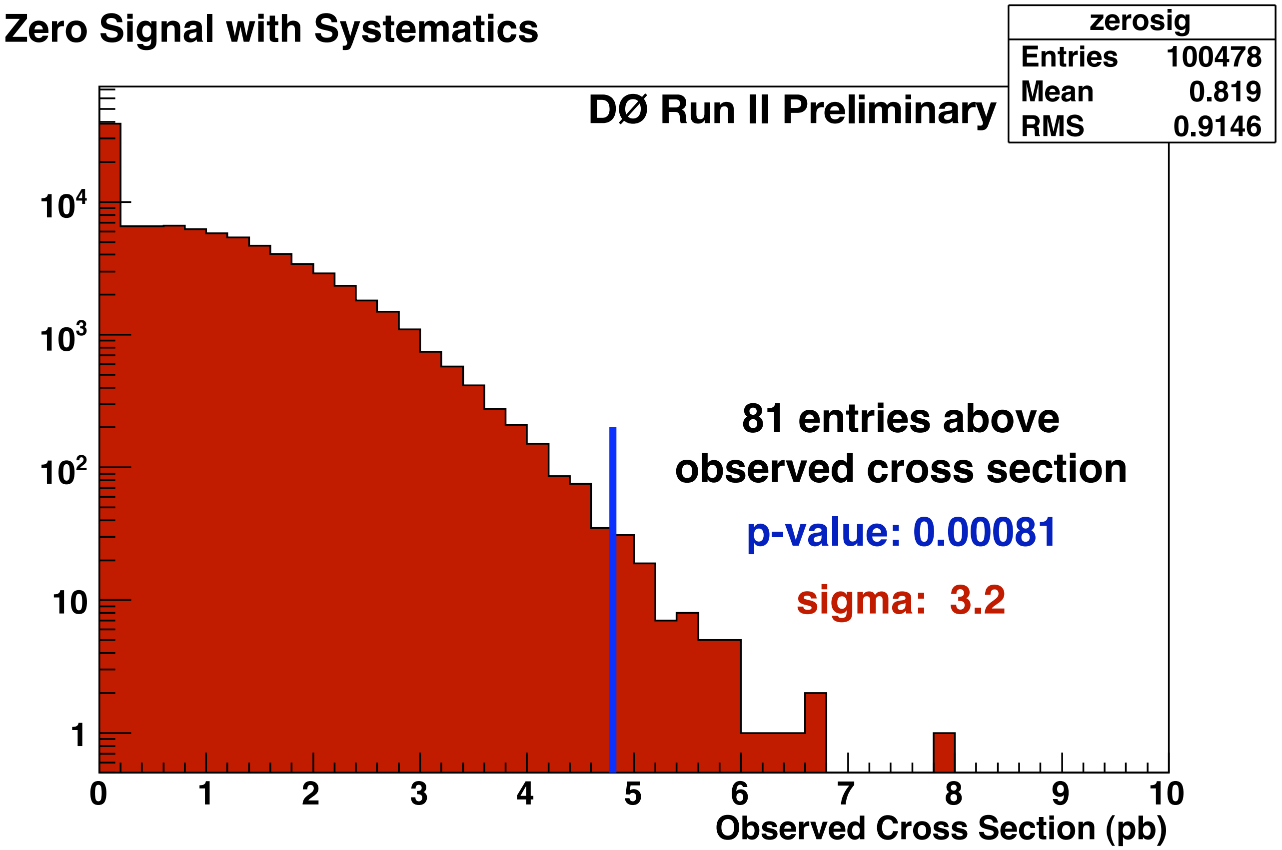
Result

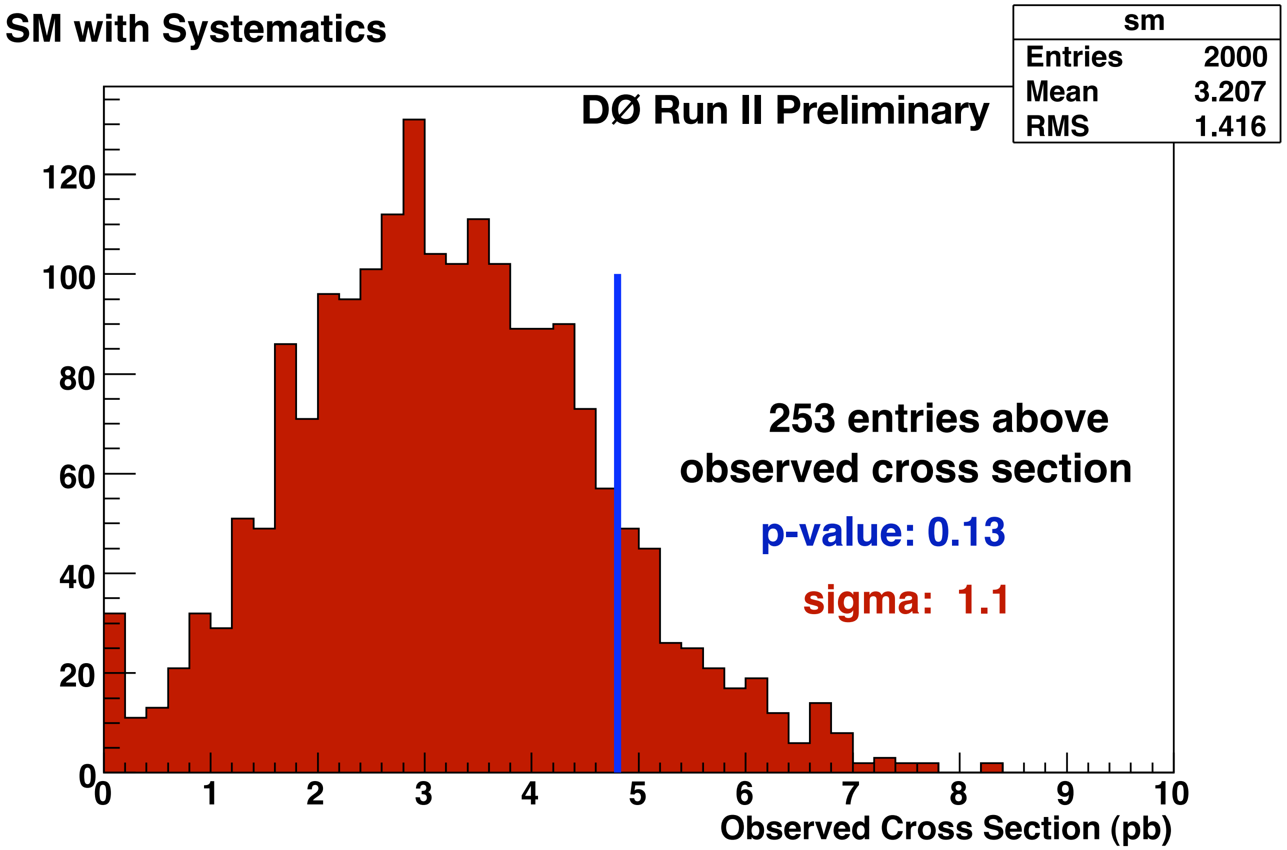
Assuming a SM cross section ratio of $\sigma_s/\sigma_t = 0.44$

$$\boxed{\sigma\left(p\bar{p} \rightarrow tb + tqb + X\right) = 4.8^{+1.6}_{-1.4}\ \mathrm{pb}}$$



Significance





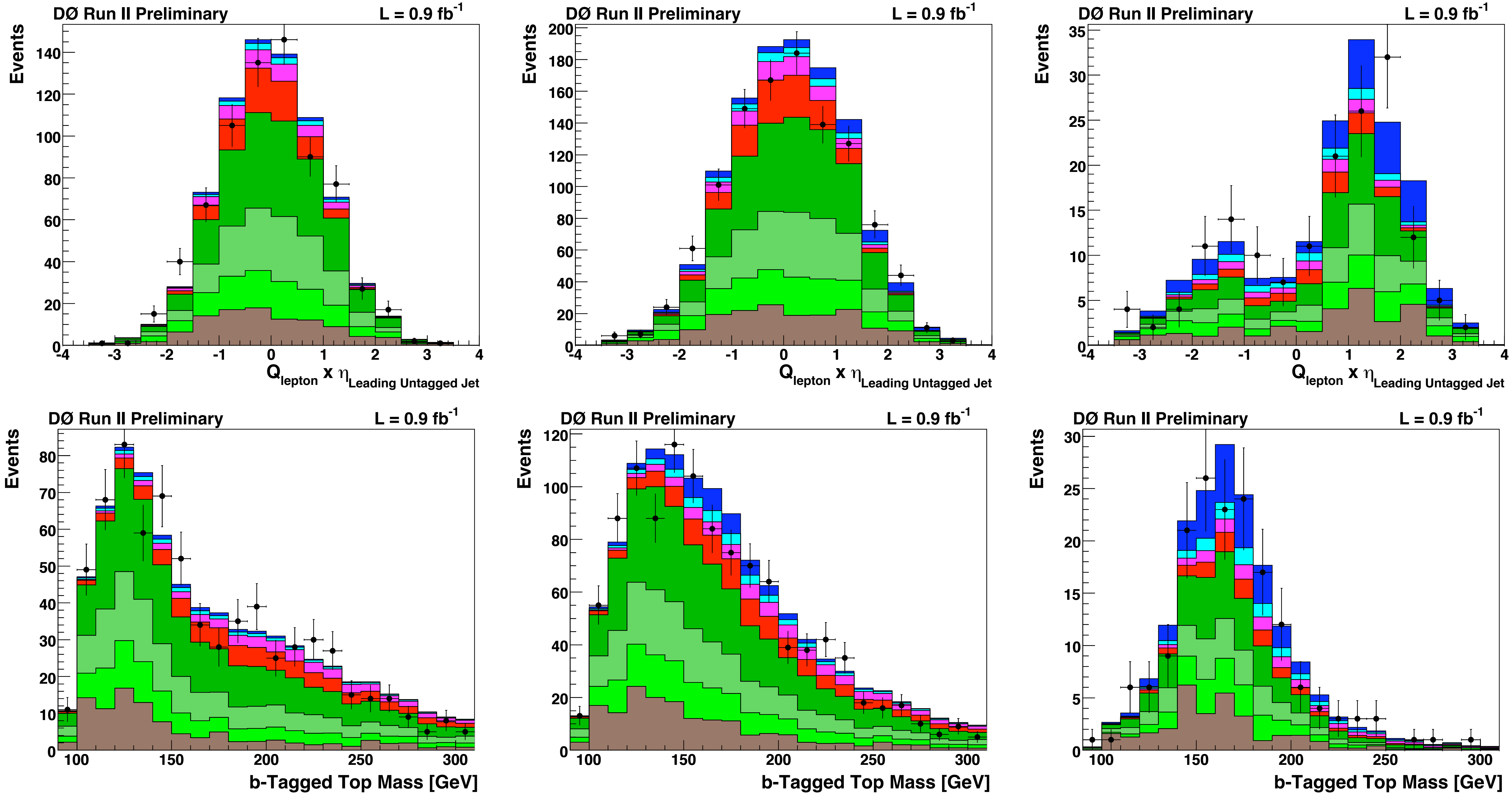
- Significance: 3.2σ. There is only a 0.08% chance for zero signal to fluctuate up to what we measure or higher.
- There is a 13% chance for a 2.9 pb signal to result in our measurement or higher.

Distributions (t-channel discriminant cut)

$D_t < 0.4$

all events

$D_t > 0.7$



Conclusion

- Made a post-PRL iteration of the ME analysis, with a number of improvements, the main one being the addition of a $t\bar{t} \rightarrow$ lepjets matrix element for the 3-jet bin. The measured cross section is:

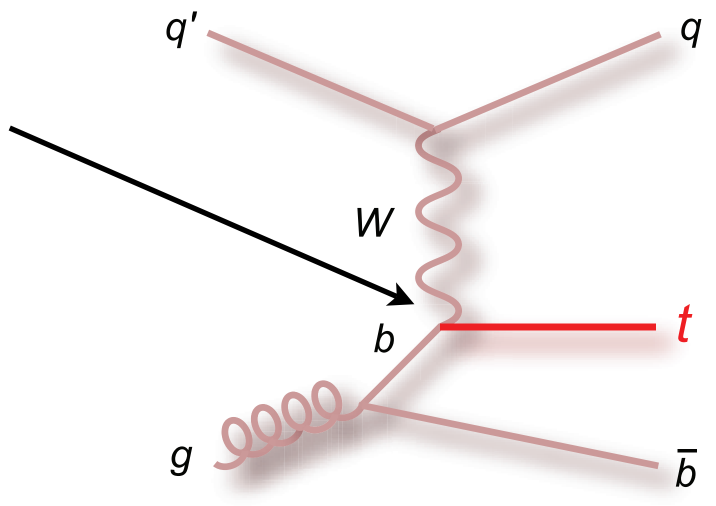
$$\sigma\left(p\bar{p} \longrightarrow tb + tqb + X\right) = 4.8^{+1.6}_{-1.4}\ \mathrm{pb}$$

  - p-value: 0.08%: 3.2σ Gaussian equivalent significance.

- An updated combination including the DT, new BNN, and new ME, using the BLUE method, is coming.

Backups

Why is Electroweak Production Interesting?



- Electroweak production is directly proportional to $|V_{tb}|^2$

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

  - ➡ Assuming unitarity:

$$|V_{tb}| = 0.999100^{+0.000034}_{-0.000004}.$$

W.-M. Yao *et al,* J. Phys. G **33**, 1 (2006)

  - ➡ Without that assumption, it can be significantly smaller:

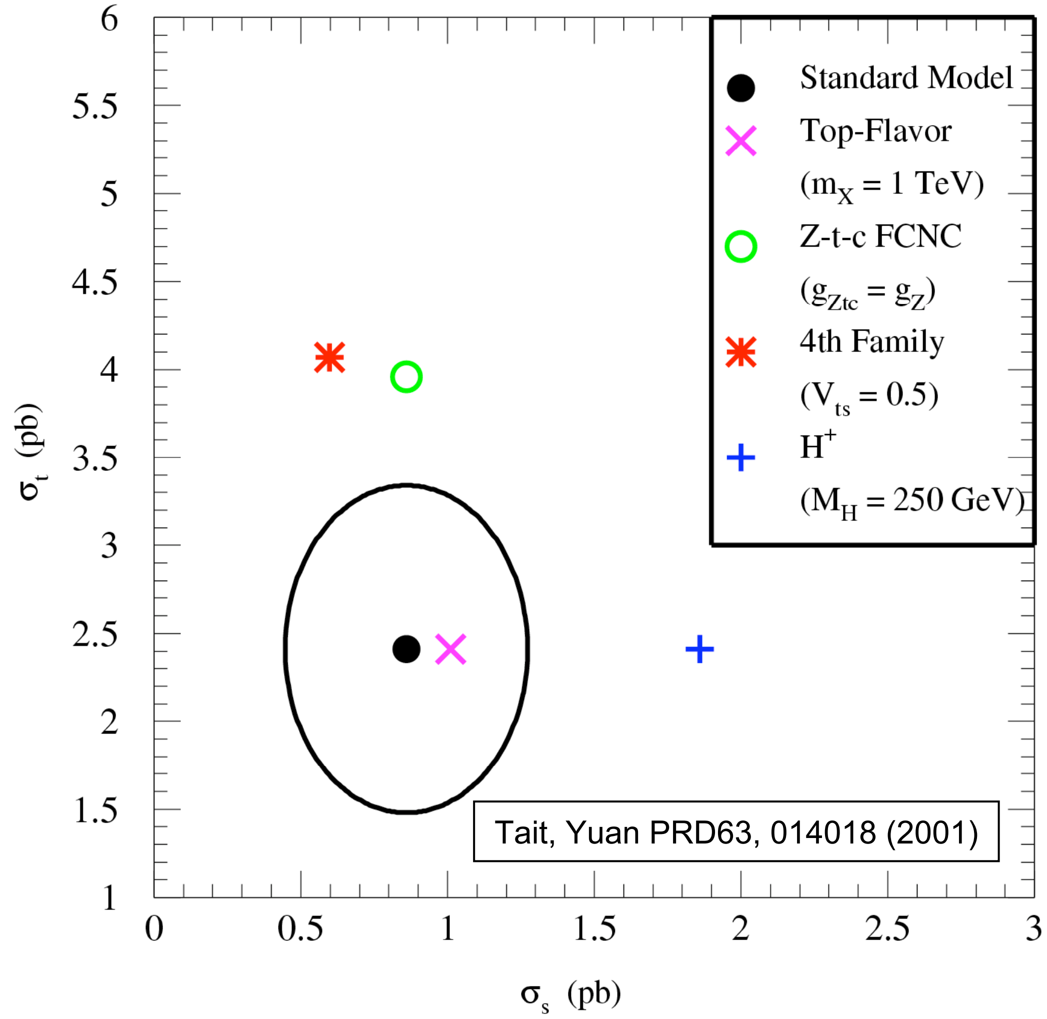
J. Alwall *et al,* arXiv:hep-ph/0607115

  - ➡ Single top production tests that assumption

- Good place to study the V–A charged current interaction
  - ➡ Because the top quark decays before it has time to hadronize, it preserves its polarization

Why is Electroweak Production Interesting?

- Sensitive to new physics.
- s-channel and t-channel have different sensitivities.
    - The s-channel is more sensitive to charged resonances, like top pions or charged Higgs particles.
    - The t-channel is more sensitive to FCNC and other new interactions.

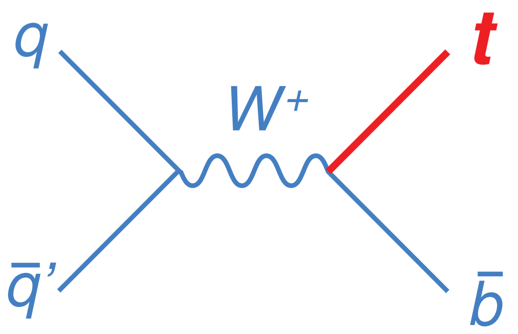


Tait, Yuan PRD63, 014018 (2001)

NLO cross sections (+ higher order soft gluon correction) at m_t = 170 GeV, Phys. Rev. D **74** 114014 (2006)

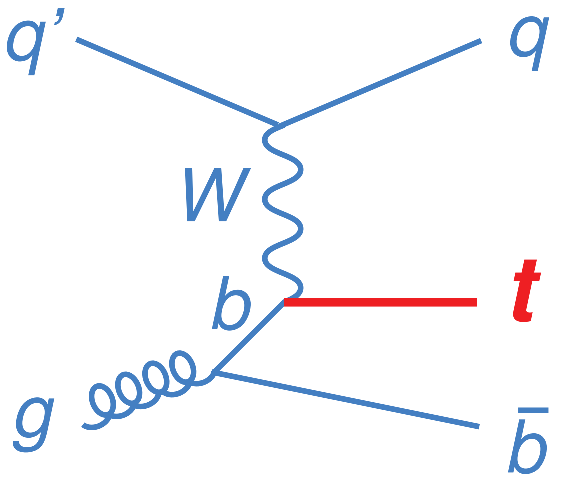
Electroweak Top Quark Production

s-channel



$$\sigma_{tb} = 1.12 \pm 0.07\,\mathrm{pb}$$

t-channel



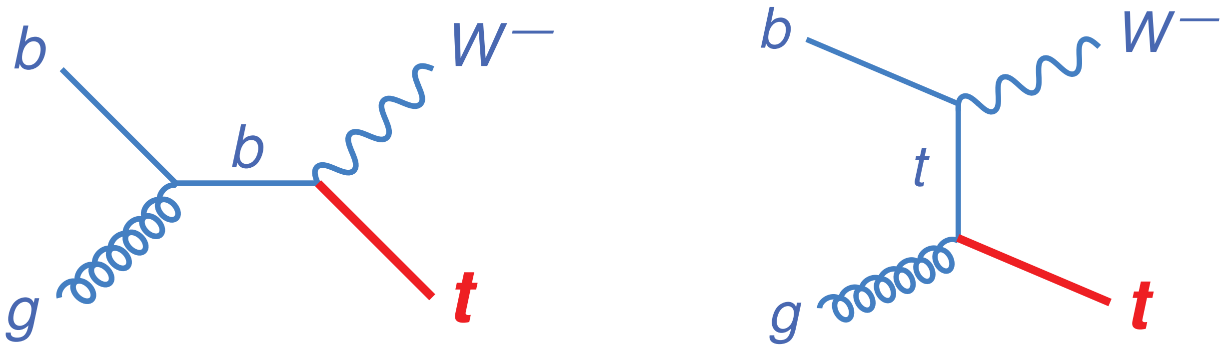
$$\sigma_{tqb} = 2.34 \pm 0.12\,\mathrm{pb}$$

DØ Results with 0.9 fb^{-1}

Methodology	s+t-channel	observed p-value
BNN (orig)	σ = 5.0 ± 1.9 pb	0.89% (2.4 σ)
ME (orig)	$\sigma = 4.6^{+1.8}_{-1.5}$ pb	0.21% (2.9 σ)
DT	σ = 4.9 ± 1.4 pb	0.04% (3.4 σ)

V. M. Abazov *et al.*, Phys. Rev. Lett. **98**, 181802 (2007).

tW associated production



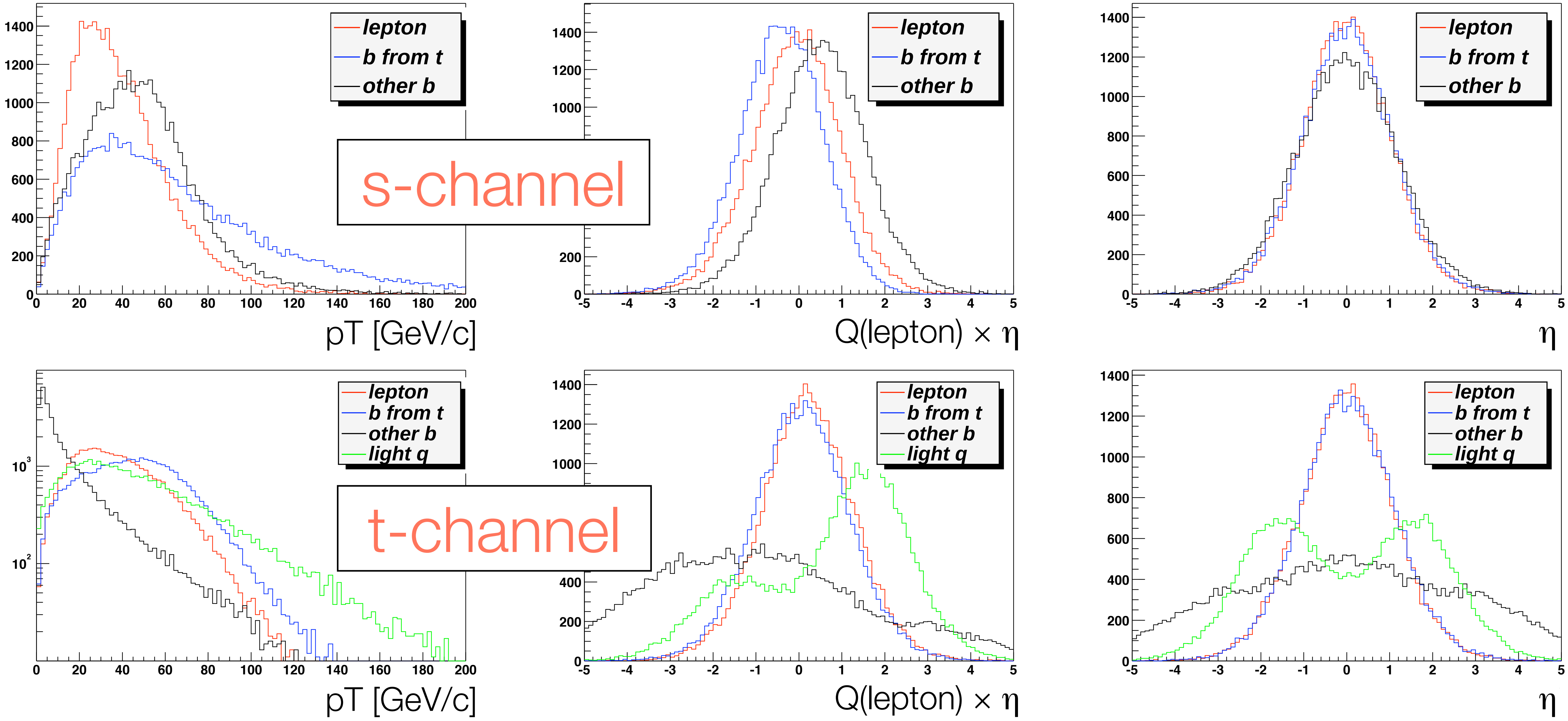
$$\sigma_{tW} = 0.30 \pm 0.06\,\mathrm{pb}$$

CDF Results with 955 pb^{-1}

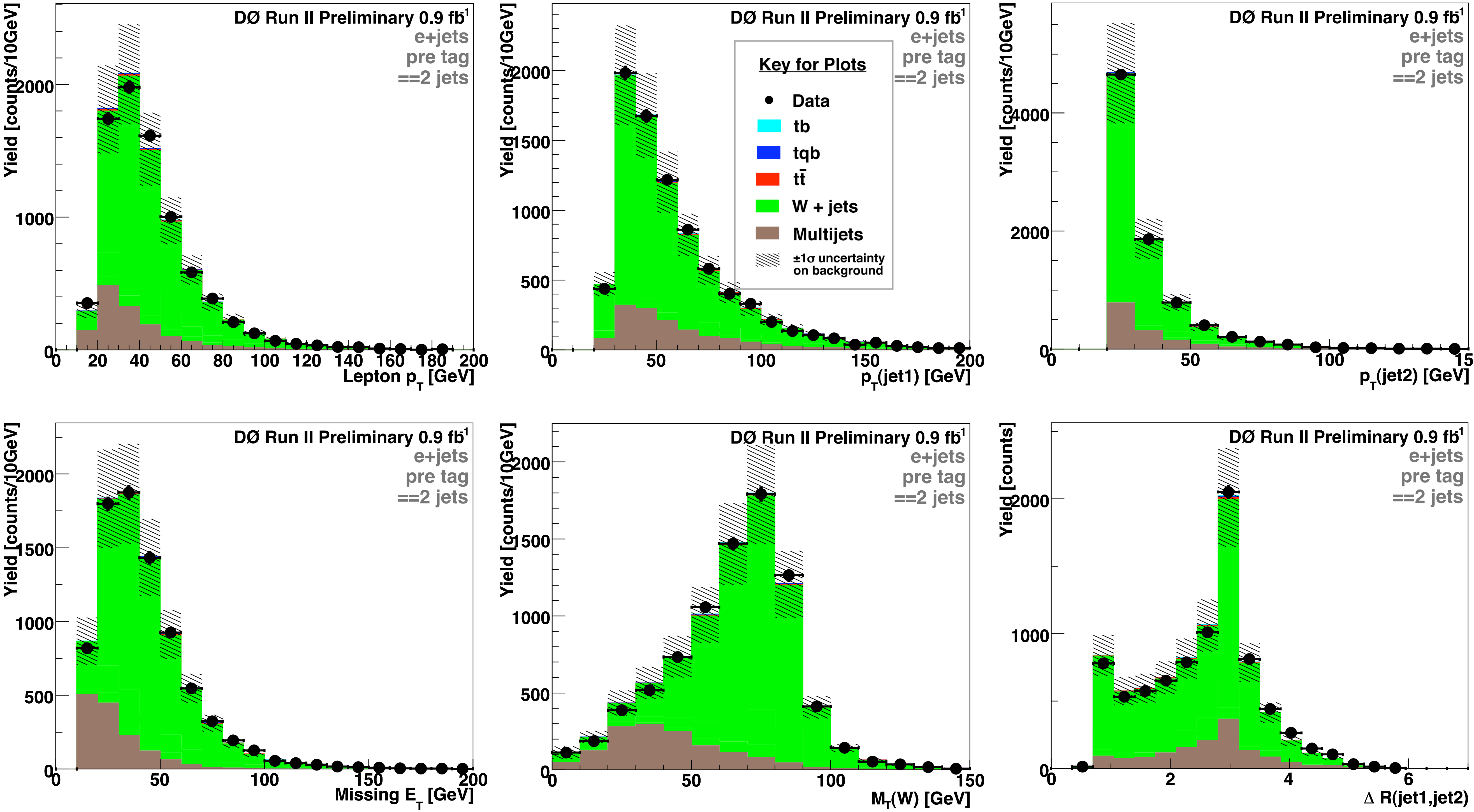
Methodology	s+t-channel	extra info
Neural Network	$\sigma < 2.6$ pb @ 95% CL	$\sigma_t = 0.2^{+1.1}_{-0.2}$ pb $\sigma_s = 0.7^{+1.5}_{-0.7}$ pb
Likelihood	$\sigma < 2.7$ pb @ 95% CL	best fit t-channel = 0.2 pb best fit s-channel = 0.1 pb
Matrix Element	$\sigma = 2.7^{+1.5}_{-1.3}$ pb	p-value: 1.0% (2.3 σ)

Compatibility of NN (both 1D and 2D), LF and ME data results is 0.65%

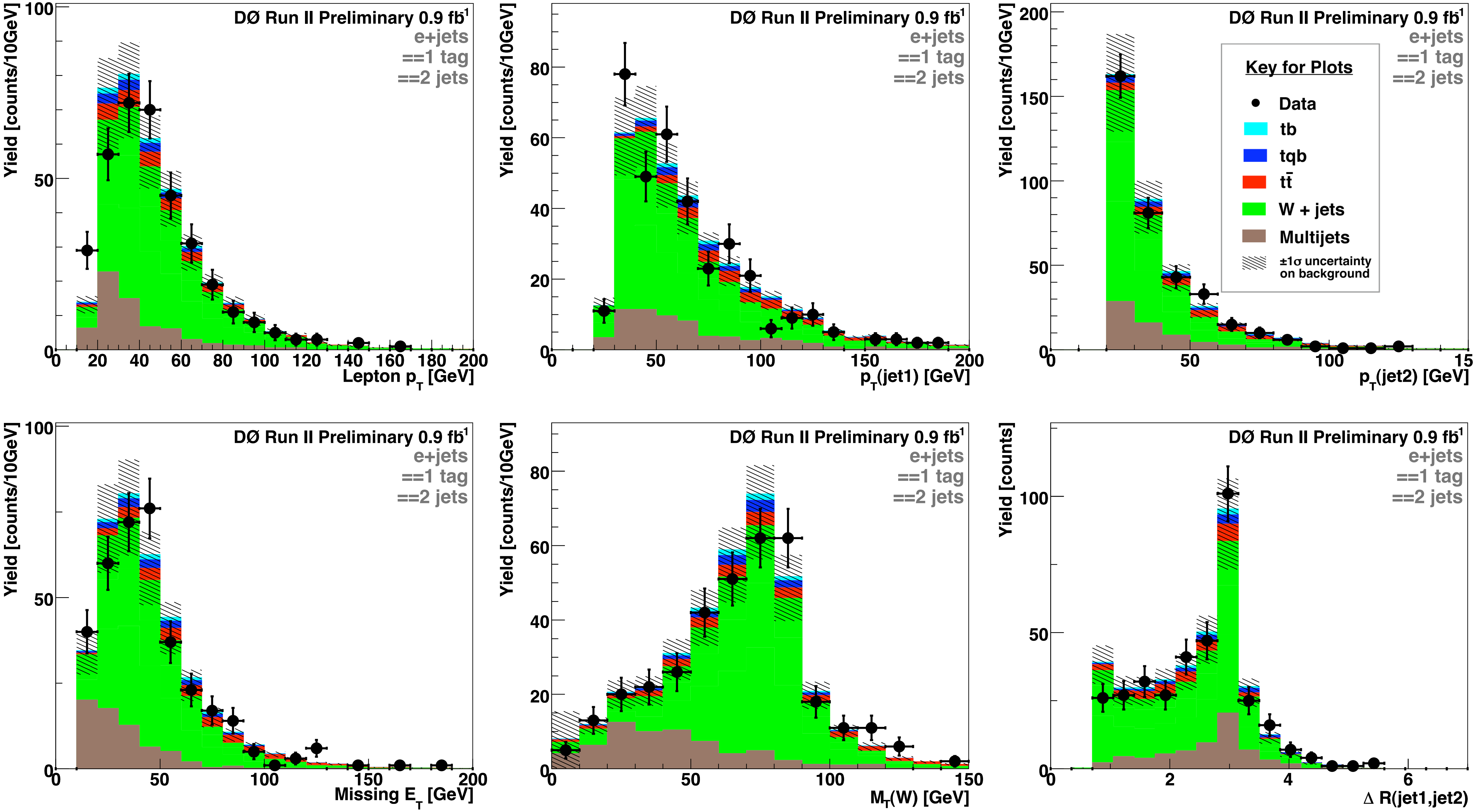
Single Top Parton Distributions



Data/MC Comparisons Before *b*-Tagging (2 jet bin, electron channel)



Data/MC Comparisons After *b*-Tagging (2 jet bin, electron channel, one tag)



Event Yields

	Yields with One b-Tagged Jet									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	5+ jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signals										
tb	2	7	3	1	0	1	5	2	1	0
tqb	3	11	6	2	1	2	9	5	2	0
$tb+tqb$	5	18	9	3	1	3	14	7	2	1
Backgrounds										
$t\bar{t}\rightarrow ll$	4	16	13	5	2	2	13	10	4	1
$t\bar{t}\rightarrow l$+jets	1	11	47	58	30	0	6	32	45	20
$Wb\bar{b}$	188	120	50	14	2	131	110	56	16	4
$Wc\bar{c}$	81	74	36	9	1	64	74	46	13	2
Wjj	175	61	20	5	1	125	58	23	6	2
Multijets	36	66	48	18	7	17	26	24	8	2
Background Sum	484	348	213	110	43	340	286	191	93	30
Data	445	357	207	97	35	289	287	179	100	38

- Try to discriminate against $t\bar{t}$ → ℓ+jets in the three-jet bin.

ME Weights

$$D(x) = \frac{P(x|\text{signal})}{P(x|\text{signal}) + P(x|\text{background})}$$

- One issue has always been how do we combine the various MEs to determine P(x | background) and P(x | signal).

$$P(x|B) = \sum_i w_i P(x|B_i)$$

- In the old analysis, the weights, w_i, are optimized by grid search.

- To be more physics-motivated, we decided to choose weights based on the relative yields. Not so easy in practice because we don't have all the matrix elements.

- For P(x | t-channel) in the 3-jet bin:
  - $w_{tqb} = 0.6$, $w_{tqg} = 0.4$ in 1-tag
  - $w_{tqb} = 1.0$, $w_{tqg} = 0.0$ in 2-tag

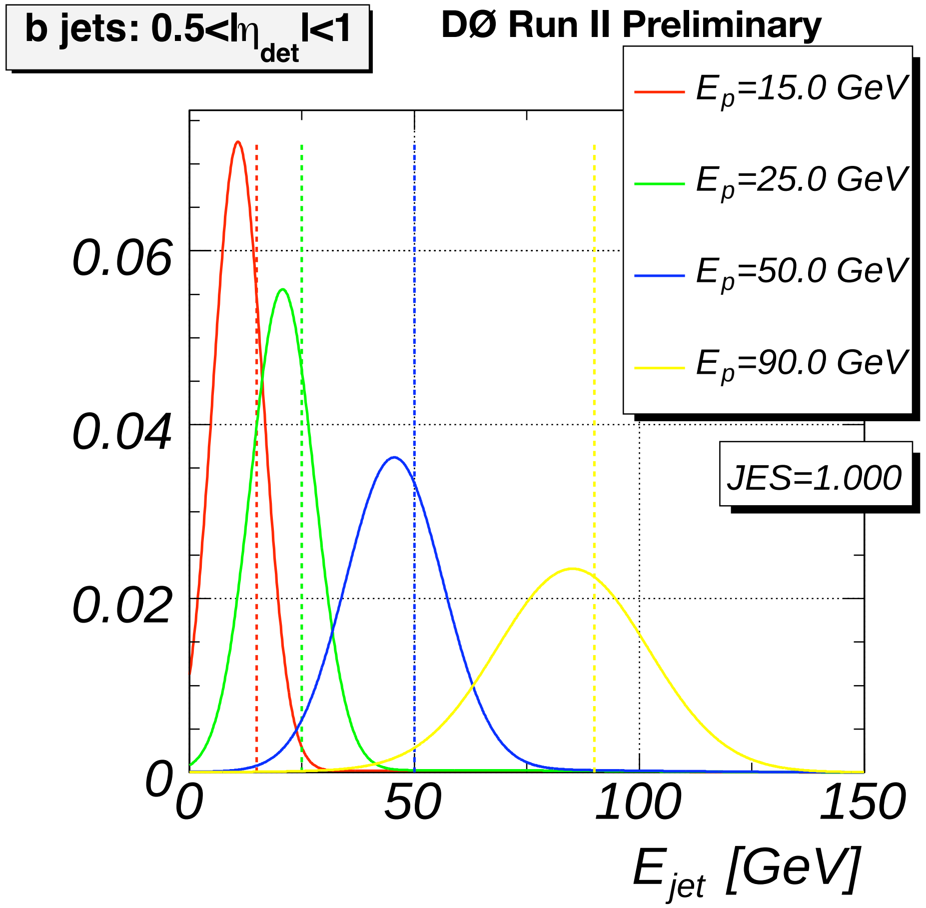
Background Fractions

	1 tag		2 tags	
	Electron	Muon	Electron	Muon
w_{wbb}	0.55	0.60	0.83	0.87
w_{wcg}	0.15	0.15	0.04	0.04
w_{wgg}	0.35	0.30	0.13	0.09
w_{wbbg}	0.35	0.45	0.30	0.40
w_{wcgg}	0.10	0.10	0.02	0.03
w_{wggg}	0.30	0.25	0.13	0.10
w_{lepjets}	0.25	0.20	0.55	0.47

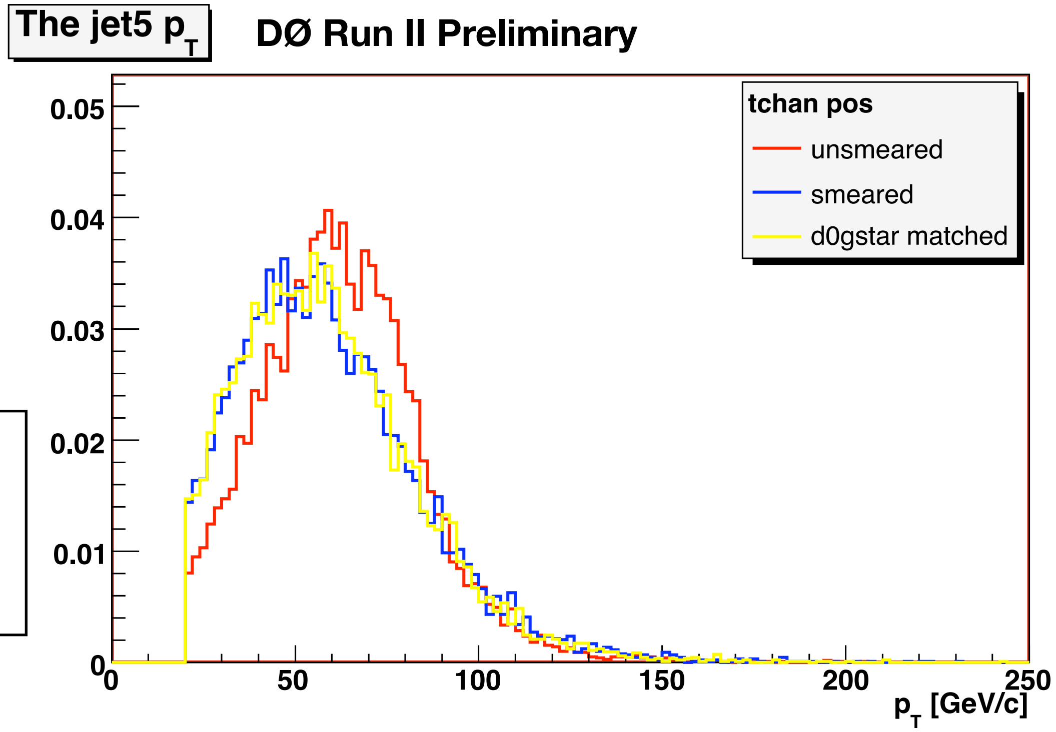
TABLE 3: Background fractions chosen for each analysis channel in two-jet and three-jet events.

Transfer Functions

- We measure reconstructed values, but the Matrix Element uses parton values.
  - ➡ Transfer Functions
- We assume:
  - can use per-object transfer functions
  - the angles are perfectly measured



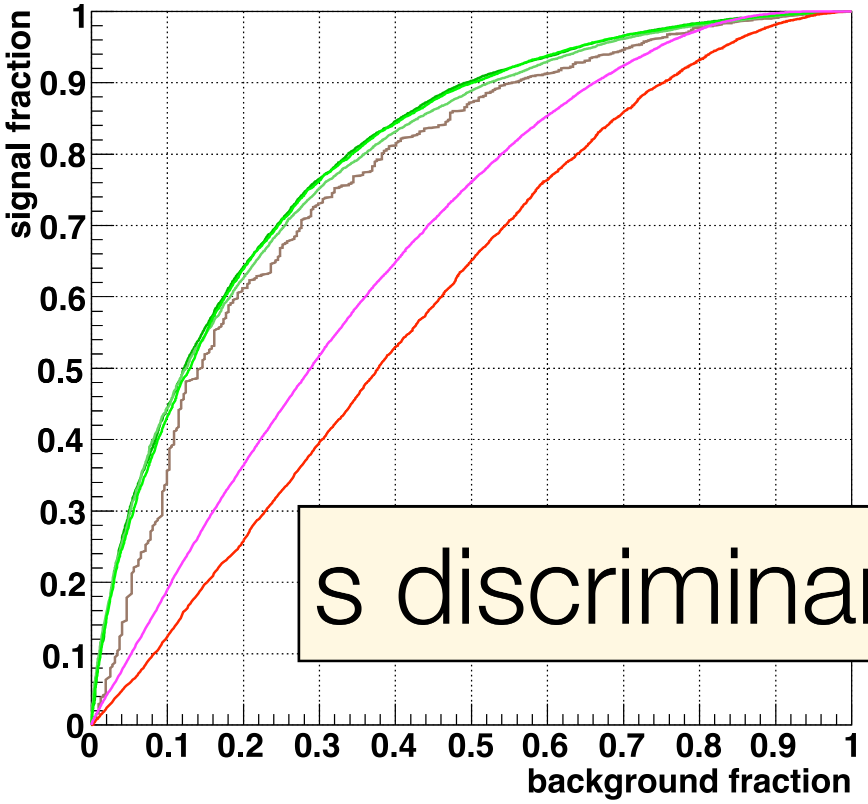
The jet transfer function for b-jets



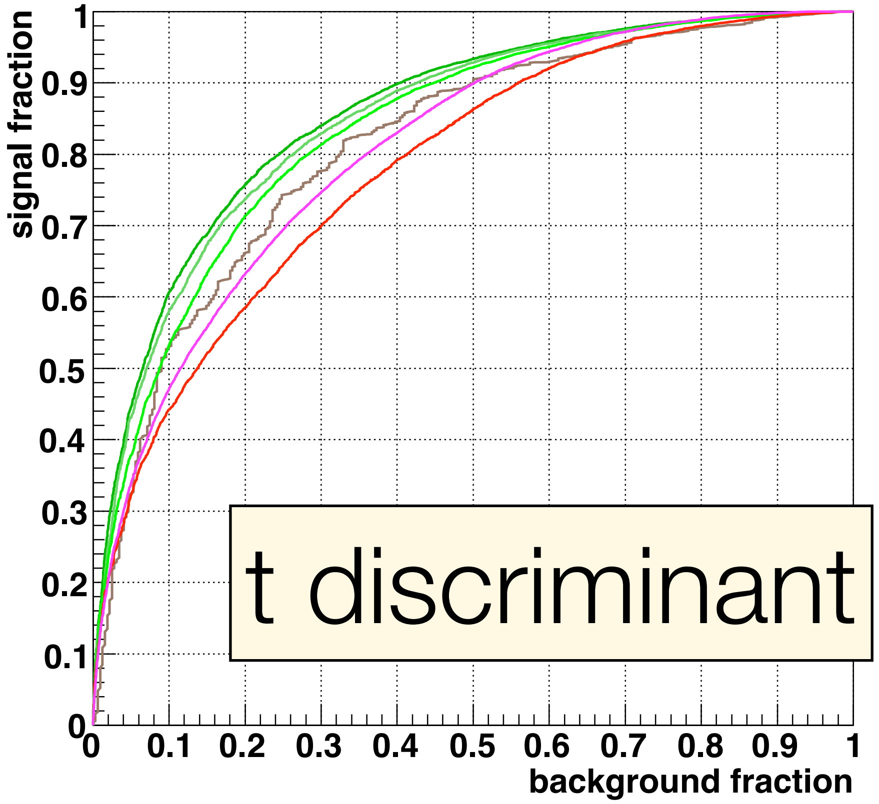
Applying the jet transfer function on the bottom quark from the top decay (blue) vs. full GEANT simulation (yellow)

Discriminant Performance (Electron, One Tag)

2 Jet



s discriminant

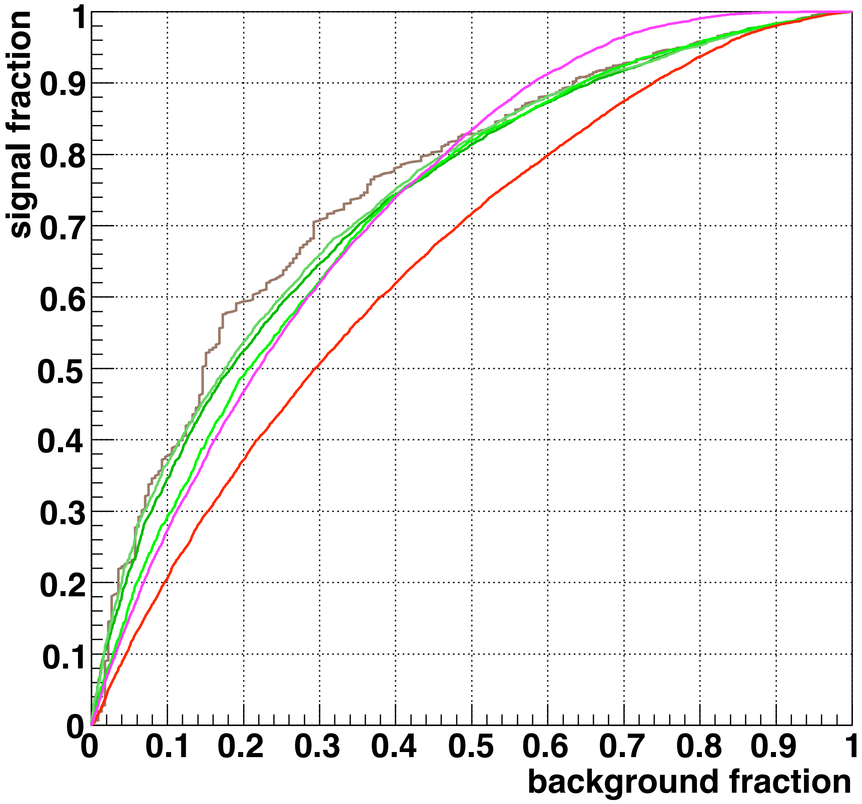


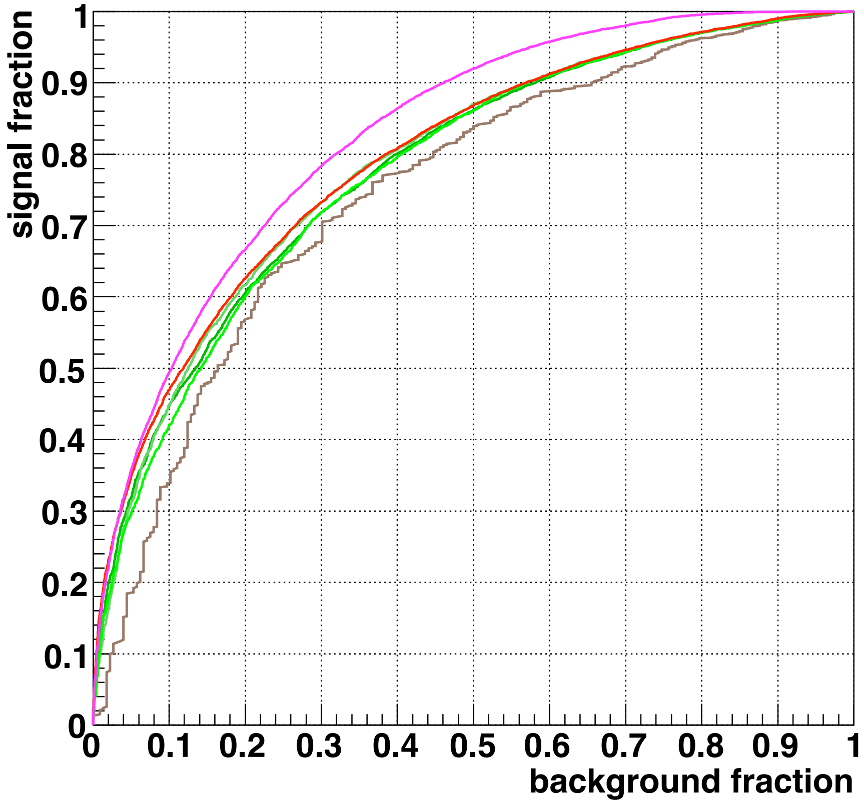
t discriminant

Key

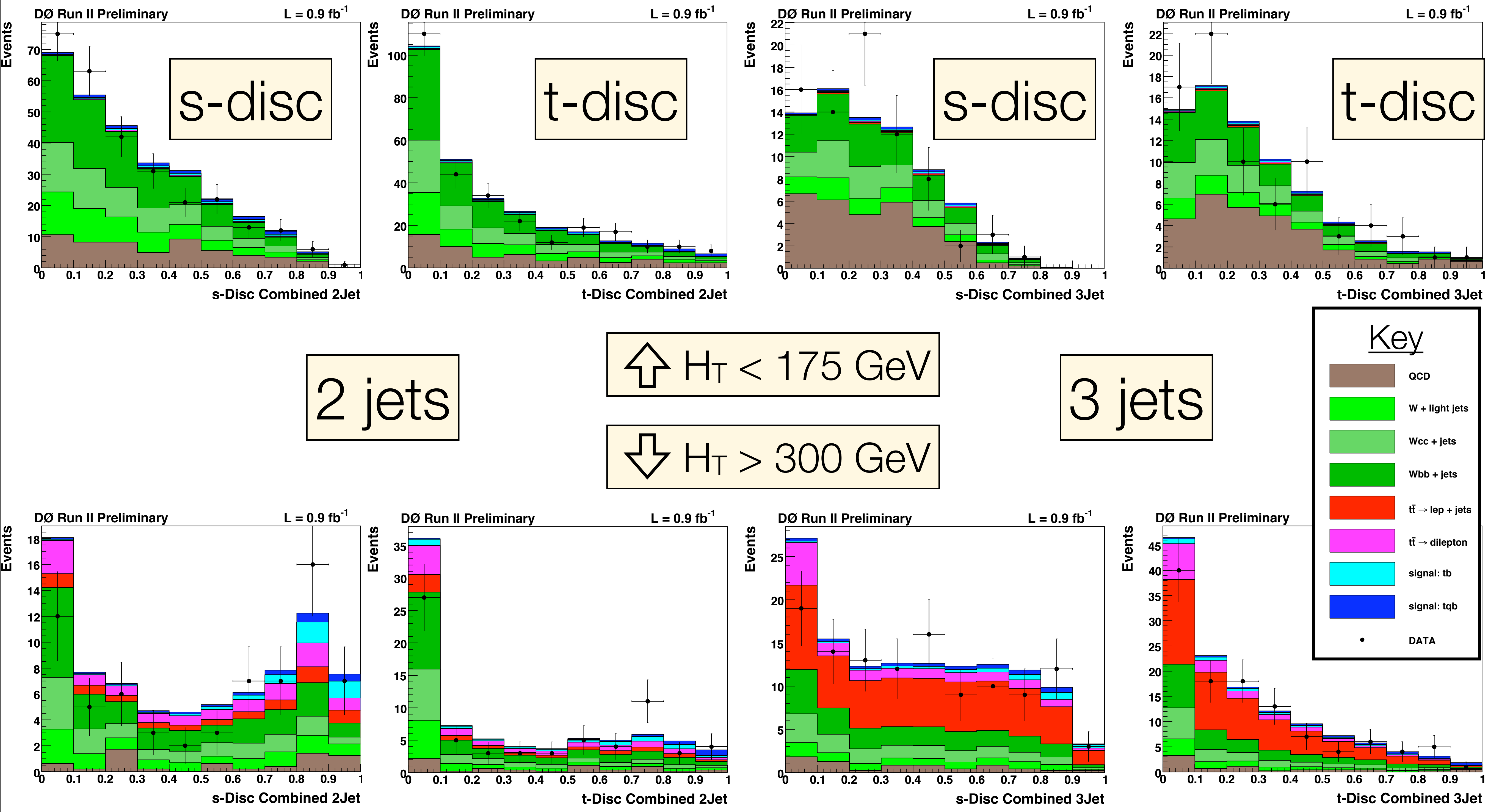
- 🟫 **QCD**
- 🟩 **W + light jets**
- 🟩 **Wcc + jets**
- 🟩 **Wbb + jets**
- 🟥 **$t\bar{t} \rightarrow$ lep + jets**
- 🟪 **$t\bar{t} \rightarrow$ dilepton**

3 Jet





Cross Check Plots



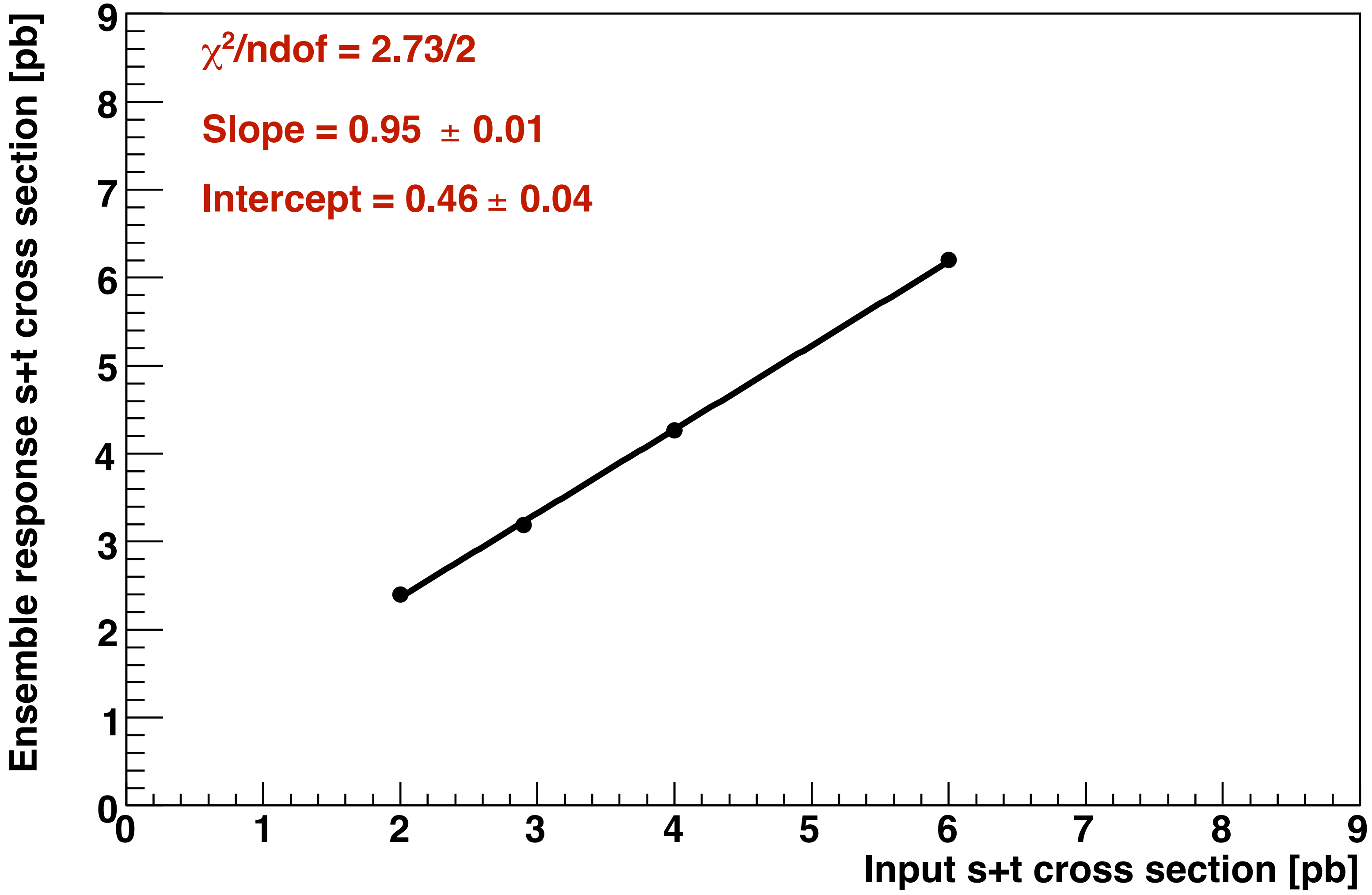
2 jets

⇧ $H_T < 175$ GeV

⇩ $H_T > 300$ GeV

3 jets

Calibration



input 2.9 pb → measure 3.2 pb

input 4.5 pb → measure 4.8 pb

The Algorithm to Lose a Jet

- Assume, for simplicity, that we lose only light quark jets.
- The algorithm requires figuring out which quark to lose and assigning a weight reflecting the probability to lose that jet. It proceeds as follows:
  - If the two light quarks are within ΔR < 0.6, it is assumed that they merge. No merging with *b*-jets is supported. The weight returned is 1.
  - Randomly choose which light parton to lose.
  - If the lost parton has |η| > 3.4, it is assumed that the associated jet is not found with probability 1.
  - Otherwise, (and this should be the main method) the returned weight is:

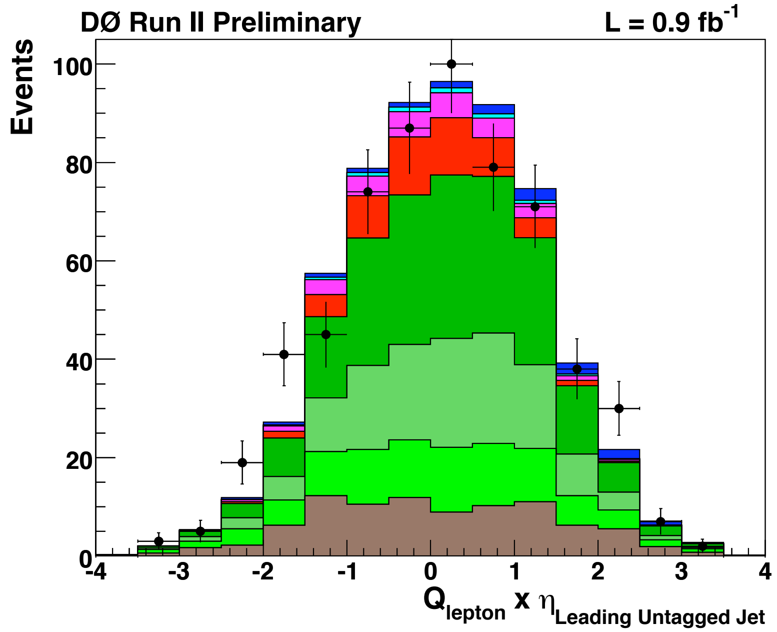
$$w(E_{\mathrm{T,parton}}) = \max\left\{\int_0^{15} dE_{\mathrm{T,reco}} W_{jet}(E_{\mathrm{T,reco}}|E_{\mathrm{T,parton}}), 0.05\right\}$$

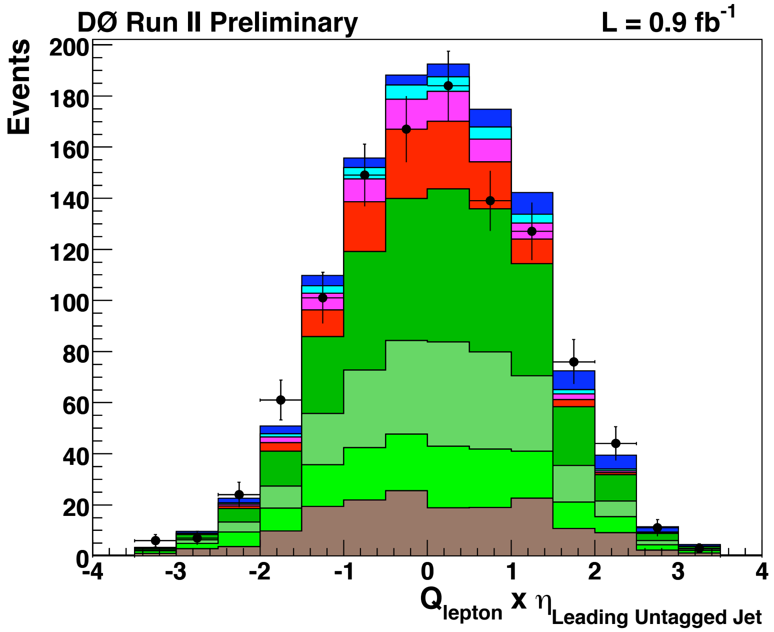
Distributions (s-channel discriminant cut)

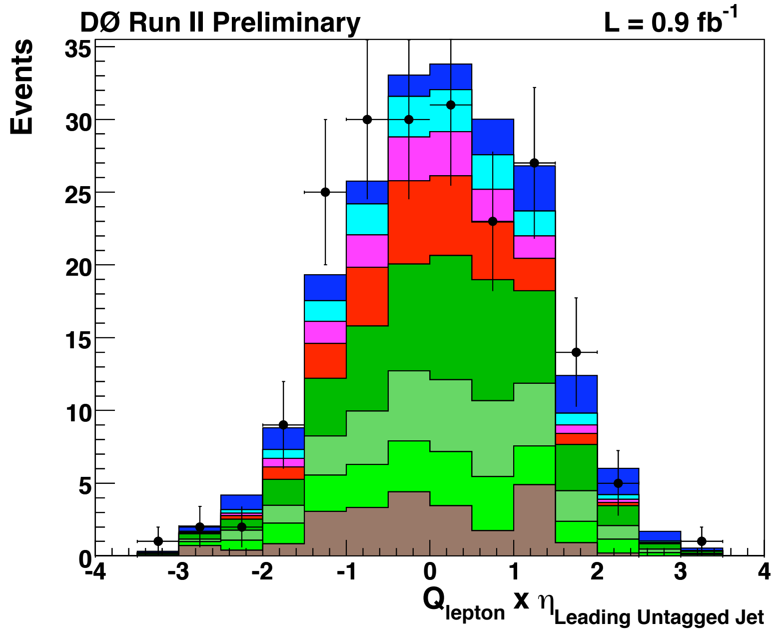
$D_S < 0.4$

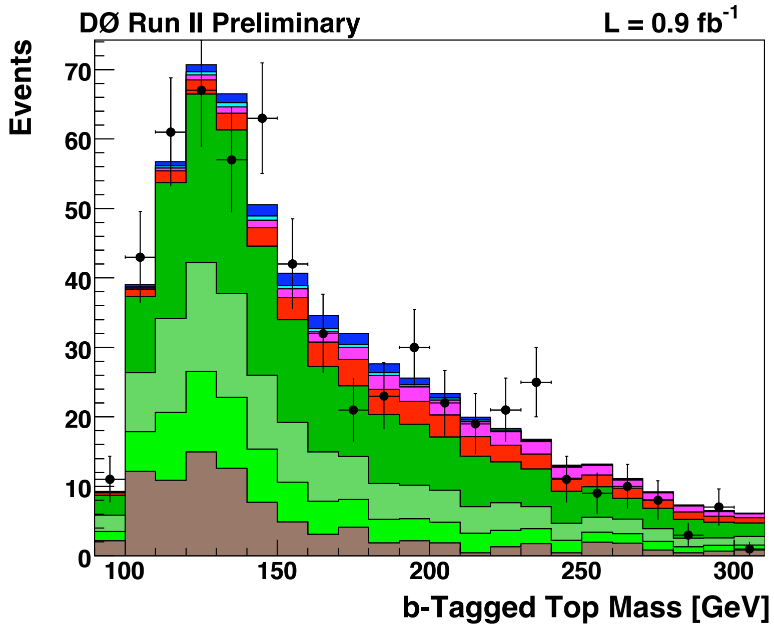
all events

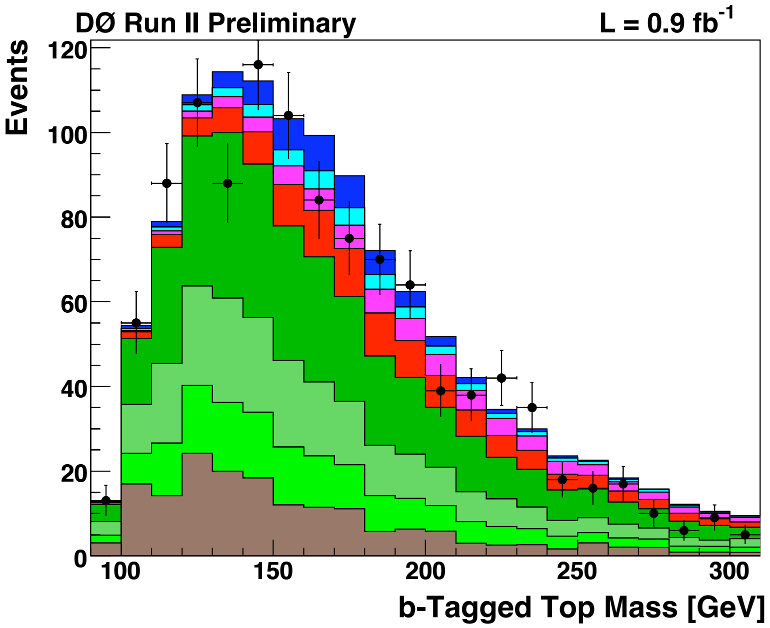
$D_S > 0.7$

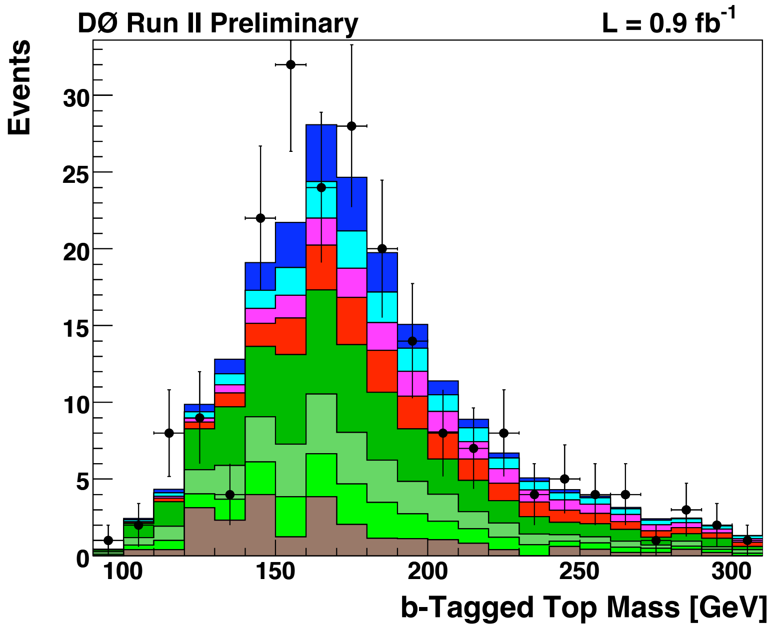












Combining using the BLUE method

- BLUE method:

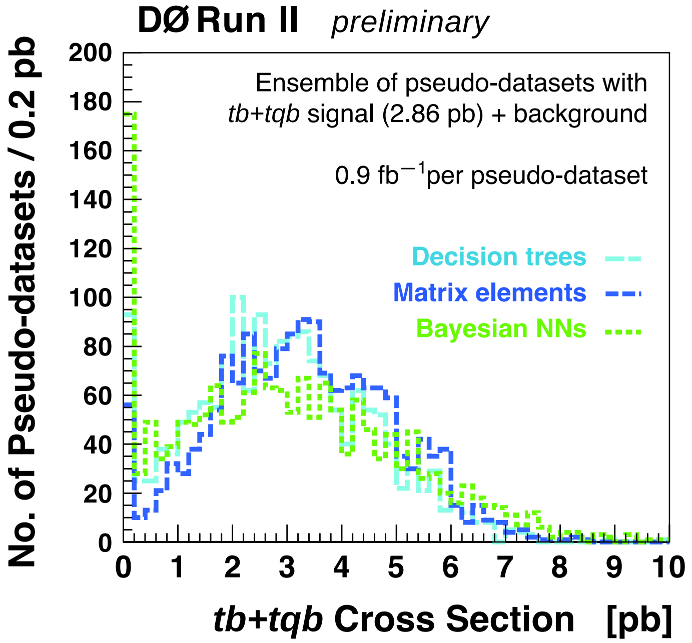
$$\sigma_{\text{comb}} = \sum_j w_j \sigma_j \qquad \Delta\sigma_{\text{comb}} = \sqrt{\sum_i \sum_j w_i w_j \rho_{ij} \Delta\sigma_i \Delta\sigma_j}$$

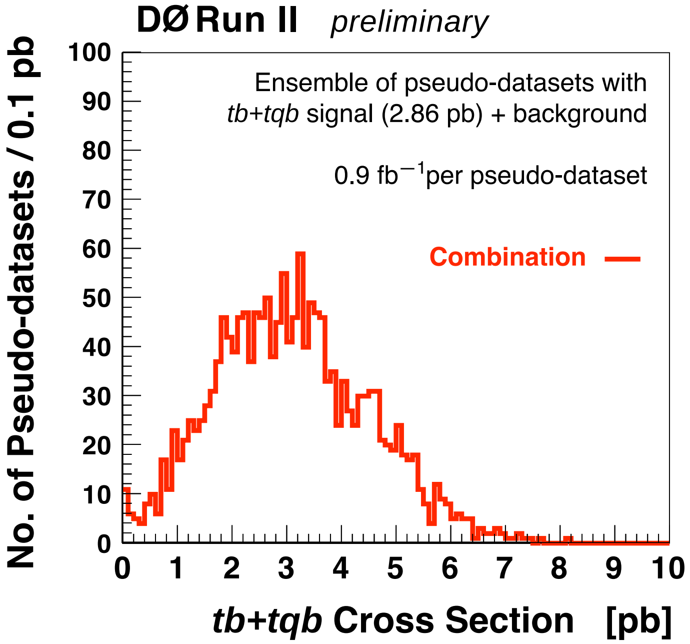
- Minimize variance by choosing:

$$w_i = \frac{\sum_j \text{Cov}^{-1}(\sigma_i, \sigma_j)}{\sum_k \sum_l \text{Cov}^{-1}(\sigma_k, \sigma_l)}$$

- Correlation matrix:

$$\rho_{ij} \equiv \frac{\text{Cov}(i,j)}{\sqrt{\text{Var}(i)\text{Var}(j)}}$$





$$\rho = \left(\begin{array}{cccc} DT & ME & BNN & \\ 1 & 0.57 & 0.51 & DT \\ 0.57 & 1 & 0.45 & ME \\ 0.51 & 0.45 & 1 & BNN \end{array}\right)$$

Combining using the BLUE method (cont.)

- From SM Ensembles:

Analysis	Mean σ [pb]	RMS $\Delta\sigma$ [pb]	$\sigma/\Delta\sigma$
Decision trees (DT)	2.9	1.6	1.8
Matrix elements (ME)	3.3	1.6	2.1
Bayesian neural networks (BNN)	3.0	2.1	1.4
Combined	3.1	1.4	2.2

- The following weights are chosen:

  $w_{DT} = 0.401$, $w_{ME} = 0.452$, $w_{BNN} = 0.146$

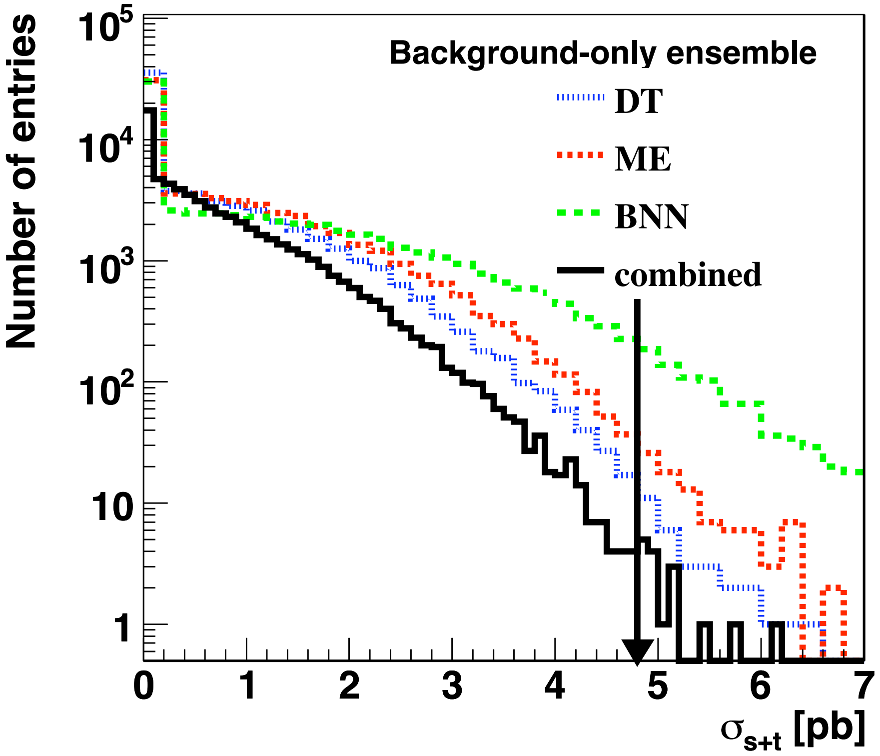
- Expected Significance:

Analysis	Expected p-value	Expected significance [std. dev.]
Decision trees (DT)	0.0177	2.1
Matrix elements (ME)	0.0358	1.8
Bayesian neural networks (BNN)	0.0992	1.3
Combined	0.0137	2.2

Combination Results

$$\sigma\left(p\bar{p} \rightarrow tb + tqb + X\right) = 4.8 \pm 1.3\,\mathrm{pb}$$

Analysis	Measured cross section [pb]	p-value	Significance [std. dev.]
Decision trees (DT)	4.9	0.00040	3.4
Matrix elements (ME)	4.6	0.00201	2.9
Bayesian neural networks (BNN)	5.0	0.01157	2.3
Combined	4.8	0.00027	3.5



3.5σ significance