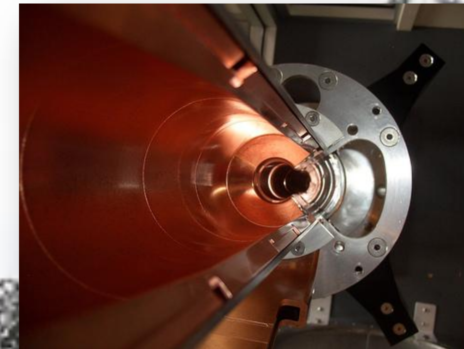
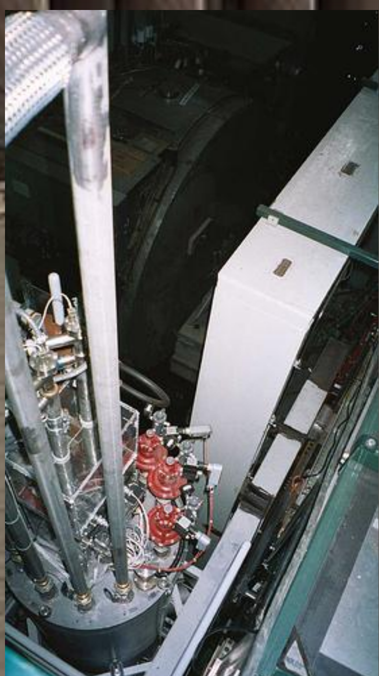


EVIDENCE FOR SINGLE TOP QUARK PRODUCTION AT DØ



G. Watts
University of Washington

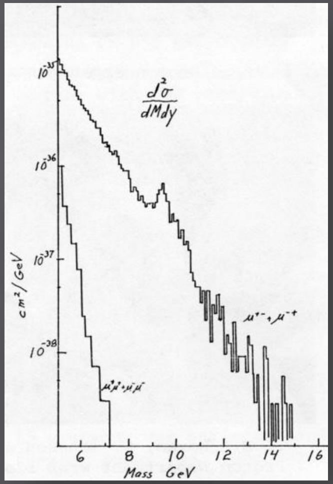
History Of The Top Quark

- 1961: Gell-Mann proposes 3 quarks
  (classification of the hadron zoo)

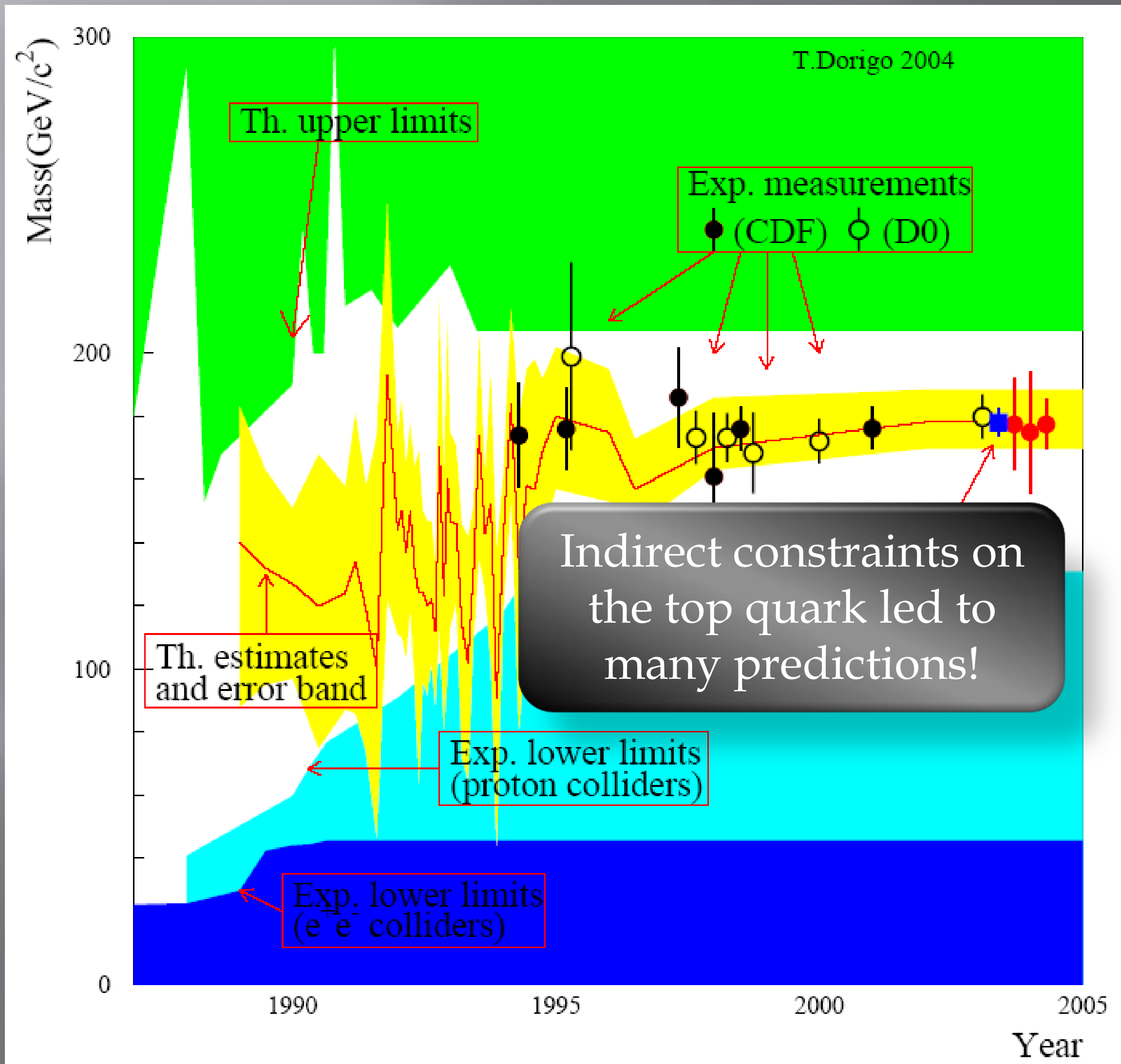
- 1973: Glashow, Iliopoulos, Mani propose charm
  (address FCNC non-existence)

- 1973: Kobayashi and Maskawa propose 3rd generation
  (CP violation in the Kaon system)

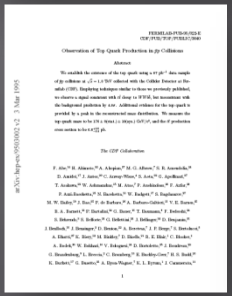
- 1977: Lederman & co. discover the b



1995: Top Quark Discovery



Indirect constraints on the top quark led to many predictions!



The start of a long program of top physics

Top Quark

It is 10 years on… We still have a lot to learn

$m_t = 171.4 \pm 2.1$ GeV/c^2!

Close to a gold atom

Heaviest fundamental particle

Coincidence or EWSB?

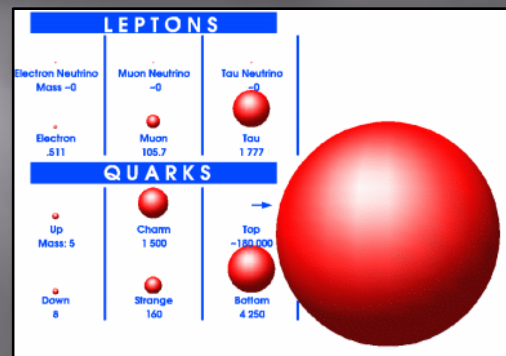
$m_t \sim vev/\sqrt{2}$, $\lambda_t \sim 1$

A Unique Laboratory

Lifetime $\tau = 5 \times 10^{-25}$ s Λ_{QCD}

➡ Decays as a free quark!

Passes spin information directly off to its decay products.



Fermilab is the only place to produce them in any quantity!

LHC is up next…

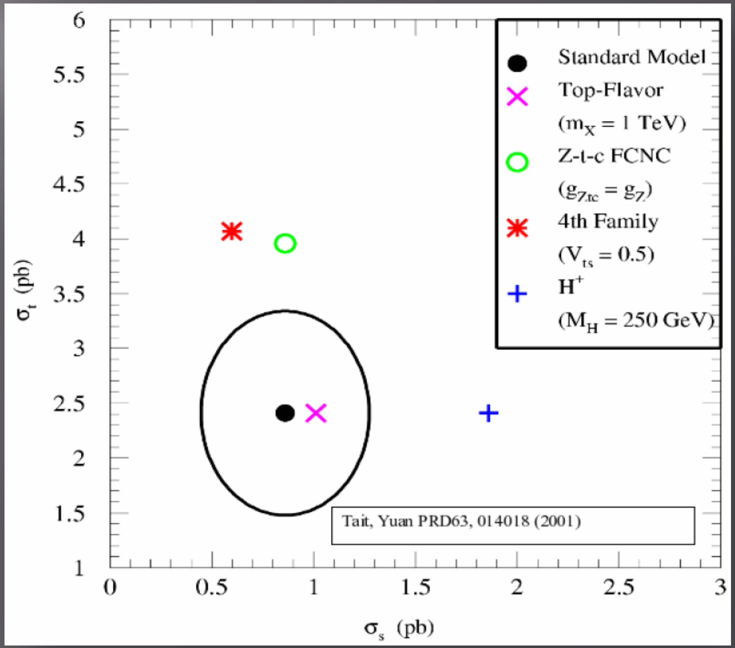
Single Top Physics

What Do We Know About the Top Quark?

- Cross Section for Pair Production
- Mass
- BR(t→Wb) ~ 1 assuming the SM
- Charge

Plenty is Unknown

- Decay Width
- Lifetime
- Spin
- BR not assuming the SM
- Direct measurement of V_{tb}

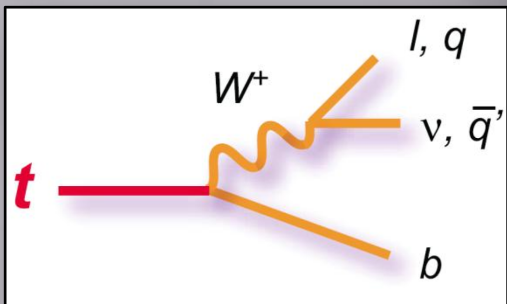


Measuring σ_s and σ_t

- Cross Sections for s and t are sensitive to different types of new physics
  - t-channel is sensitive to FCNC
  - s-channel is sensitive to new resonances

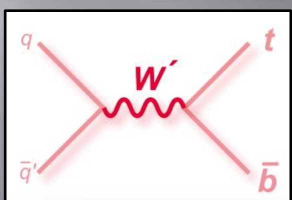
It is important to measure the rates independently

The Single Top Lab

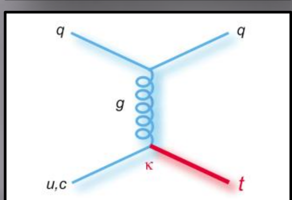


Direct Access to the W-t-b coupling (σ_{st})

Measure V_{tb} of the CKM directly
CKM Unitarity



s-channel sensitive to new resonances:
W', top pions, SUSY, etc.



t-channel sensitive to FCNC, anomalous couplings

Also

- Polarized top quarks
- Backgrounds to Higgs!

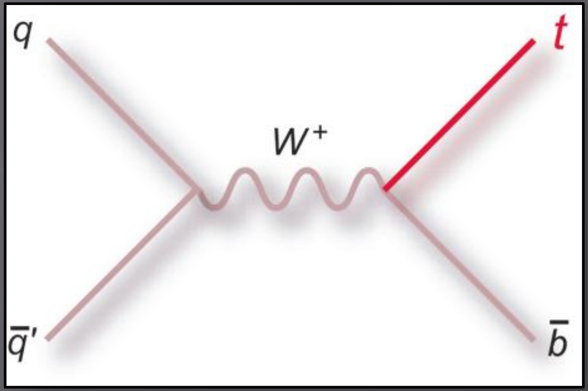
This is for after discovery!

Single Top Production

Top Quark Also Produced by the EW Process

Smaller Cross Section

0.88 ± 0.11 pb
"s-channel"



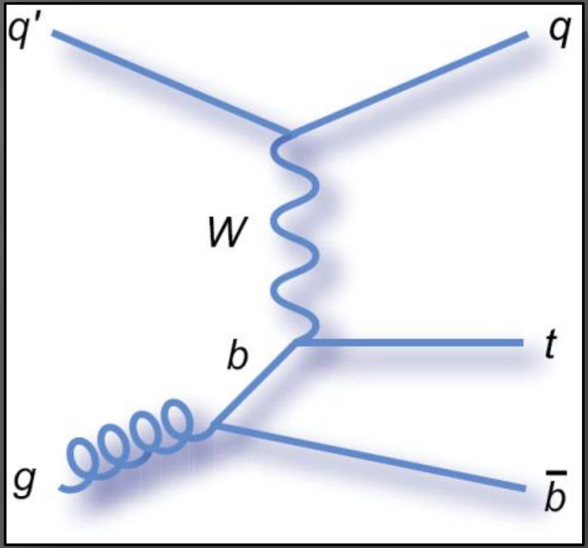
Top Decays to Wb ~ 100%

Require Isolated High p_T e, μ

- W→jj - Dijet decay backgrounds too large
- W→τν included only when it decays to a isolated lepton

Signature: Lepton, Missing E_T, jets

1.98 ± 0.25 pb
"t-channel"



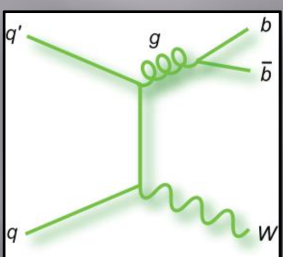
Where Is It?

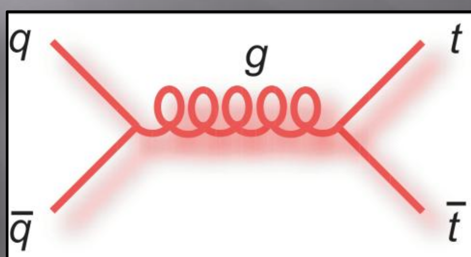
Single Top Final State

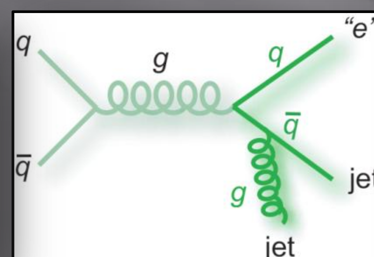
Lepton, missing E_T, and jets

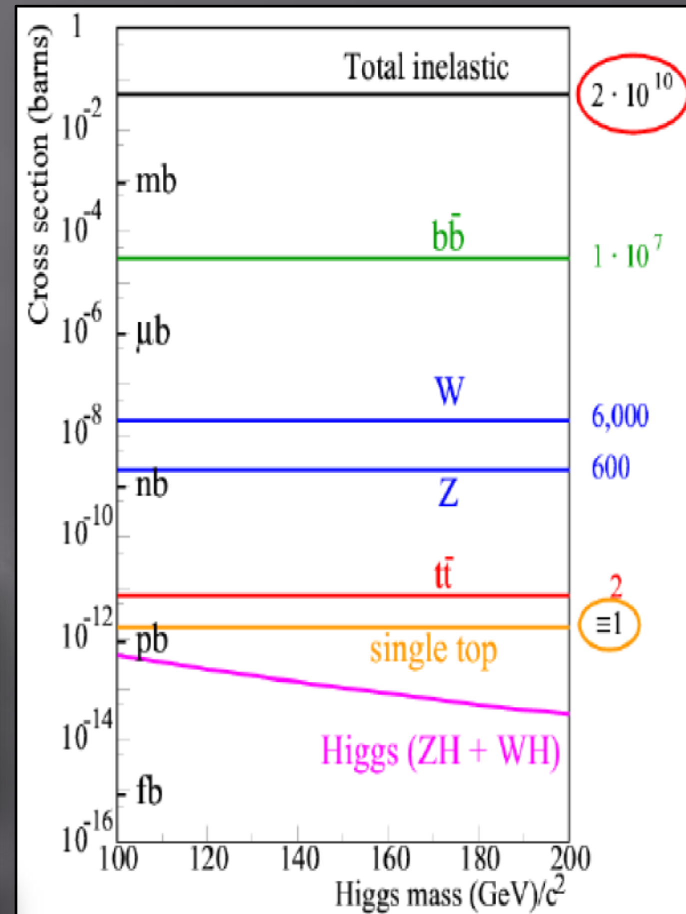
Backgrounds

W+Jets – σ = 1000 pb
tt – σ = 7 pb
QCD multi-jet background/jet mistaken ID









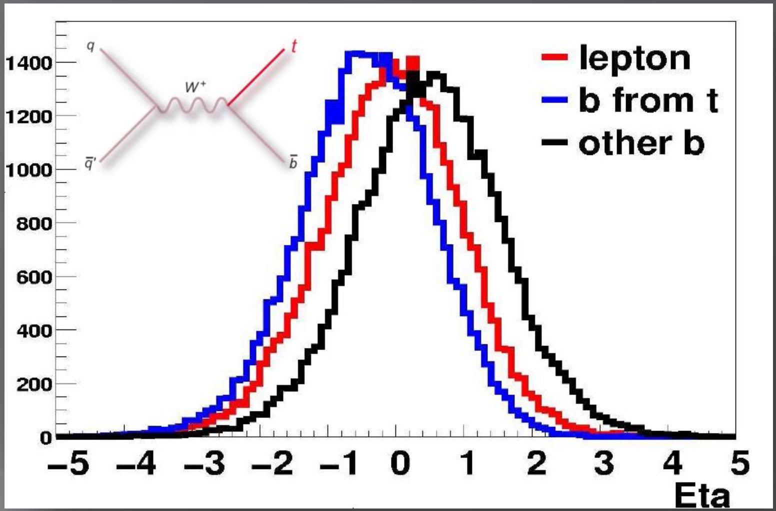
Most recent D0 result: 370 pb^{-1} σ_s<5.0 pb, σ_t<4.4 pb

Improvements: Better MC modeling (PS/ME Matching), new calibrations, jet energy scale, etc., new b-tagger, split analysis by S:B, combined s+t channel search

Single Top Final State

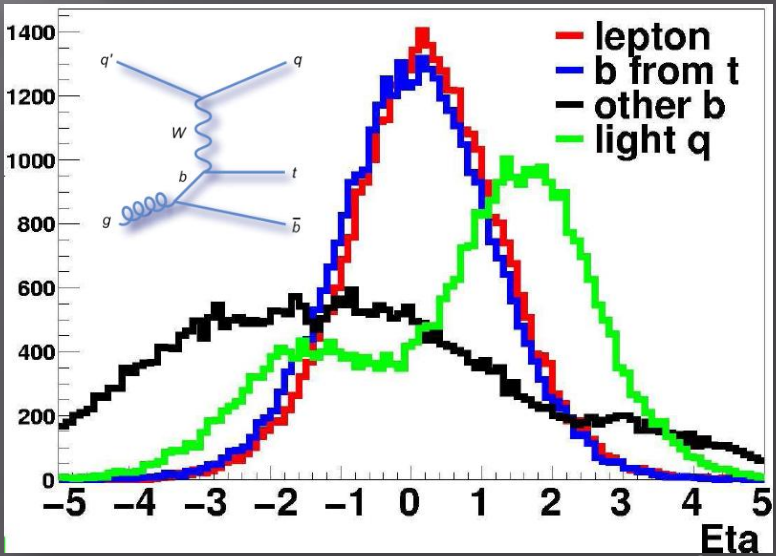
s-channel

- The top decay products and the b tend to all be central
  - Lepton, neutrino, and two b-quark jets

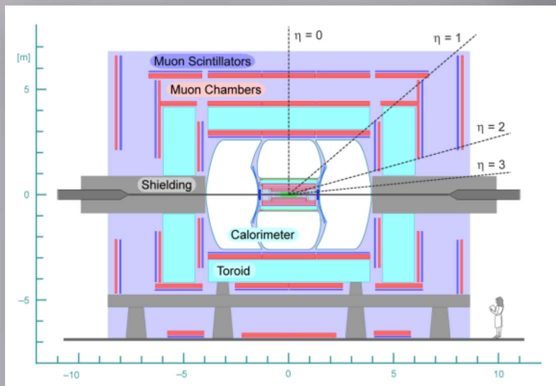


t-channel

- The b-bar tends to be very close to the beam pipe
  - Lepton, neutrino, and one b-quark jets (second only if you are lucky!)



The Search



Selection Cuts to remove background not well modeled

~1 year

Background Model – MC and Data

Scale Factors, etc.

~9 months

~1.5 years

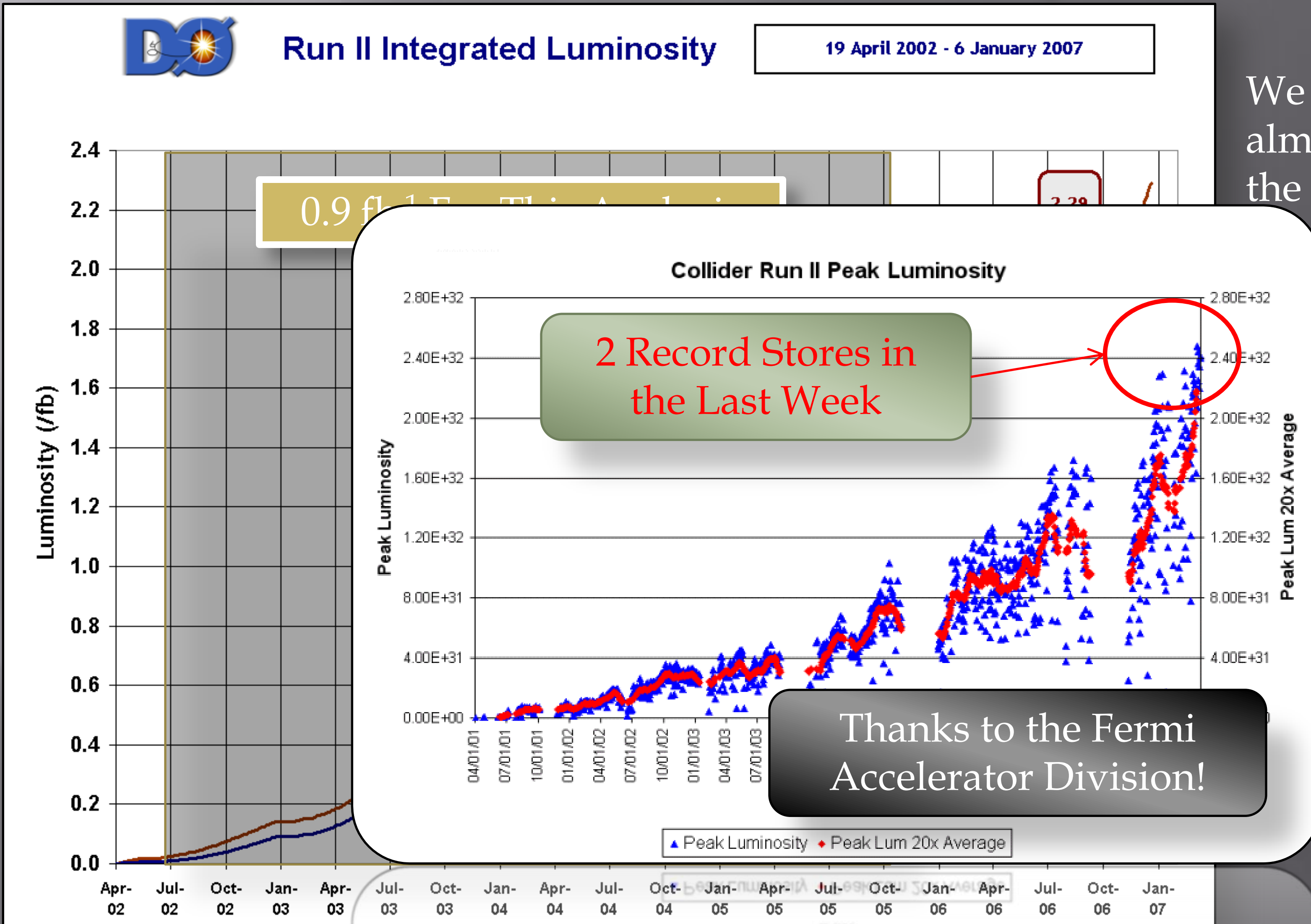
b-tagging

~4 months

Signal and Background Separation

Cross Section Estimation

Data – Fermilab/DØ

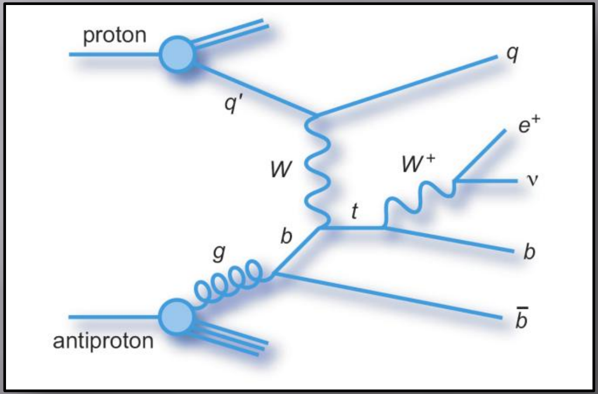


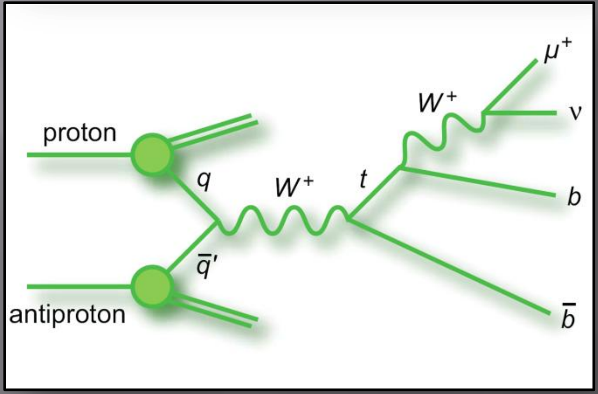
We have another almost 1 fb^{-1} in the can.


DEW

Selecting the Data Sample

We are not trying to select the signal as much as get a data sample that is well understood and modeled *and* includes as much signal as possible.





- One tight isolated lepton (from the W):
  - Muon pT>18 GeV and $|\eta_{det}|<2.0$
  - Electron pT>15 GeV and $|\eta_{det}|<1.1$
  - No other *loose* leptons allowed
- $ME_T > 15$ GeV (from the W)
- 2-4 jets
  - $p_T > 15$ GeV and $|\eta_{det}|<3.4$
  - Leading jet $p_T>25$ GeV, $|\eta_{det}|<2.5$
  - Second leading jet $p_T>20$ GeV

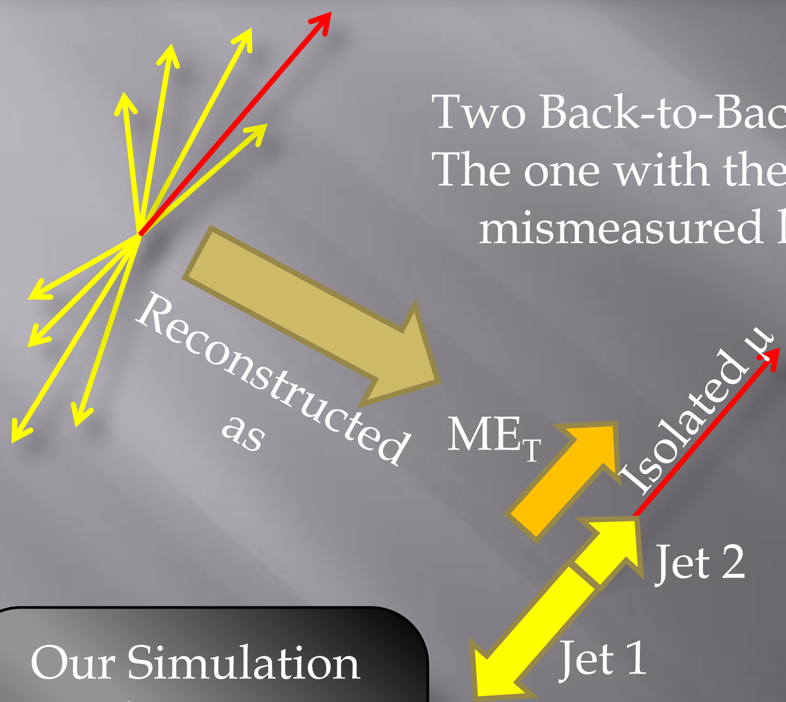
There are regions of our detector and event topologies that we do not model well… So we remove them…

Cleaning Up The Data an example…

Triangle Cuts

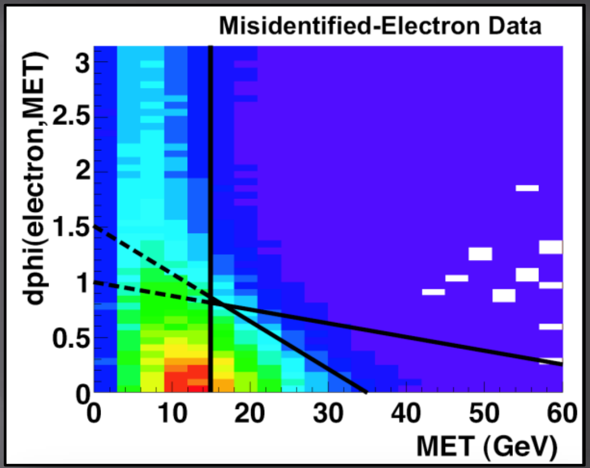
We have cuts to clean up particularly pathological backgrounds like badly mismeasured muons or noise in the calorimeter.

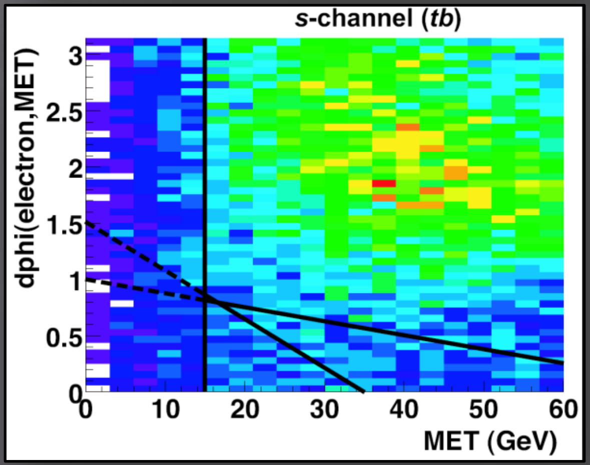
Two Back-to-Back Jets
The one with the muon is mismeasured low



All objects (jets, e, μ) can be at the source of this effect

Our Simulation does not reproduce this effect so we remove it





Boos, et al. Nucl.Instrum.Meth. A534 (2004) 250-259
Boos, Dudko, et al., CMSNote 2000/065.

Monte Carlo Samples

Signal

CompHEP-SingleTop + Pythia

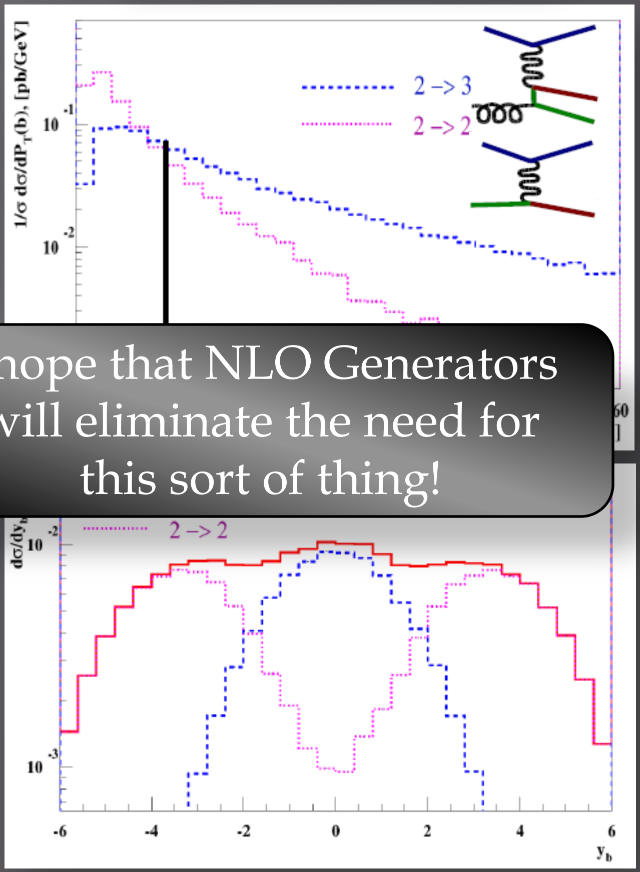
Backgrounds

- Wjj, Wbb – ALPGEN 2.0 + Pythia
  - *Parton Shower↔Jet Matching* to avoid double counting
  - Heavy Flavor fractions from data
  - Normalization from data
- tt – ALPGEN 2.0 + Pythia
  - Matching done
  - Normalize to NNLO σ
- Multijet Events (mis-id of lepton)
  - From Data

MC/Data Differences

Event weights applied to account for differences in vertex finding, jet reconstruction eff, etc.

Getting the NLO t-channel shape



I hope that NLO Generators will eliminate the need for this sort of thing!

Background Normalization

$$\text{Data} = N_{QCD} + N_{W+Jets}$$

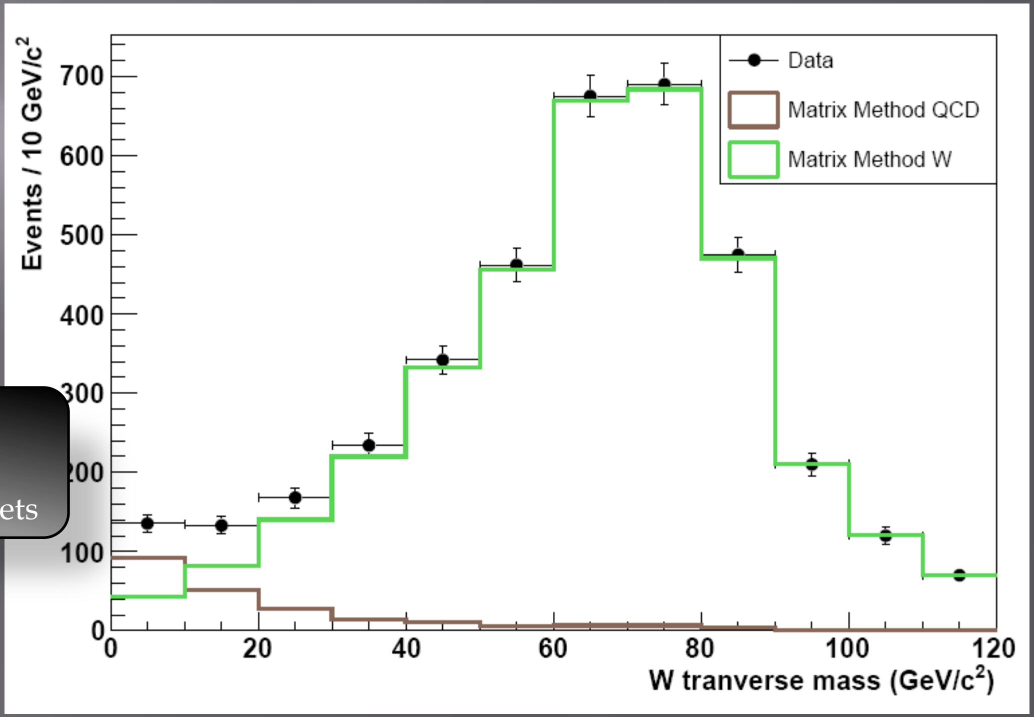
Need to Know Fractions because b-tagging rates are different, affects kinematic distributions, etc.

1. Define a loose and tight isolated lepton sample
2. Determine the Probability of seeing an isolated lepton in each sample (a fake in QCD, and a real one in W+Jets)

$$\text{Data} = N_{QCD} + N_{WJets}$$
$$\text{Isolated Data} = \varepsilon_{QCD} \cdot N_{QCD} + \varepsilon_{WJets} \cdot N_{WJets}$$

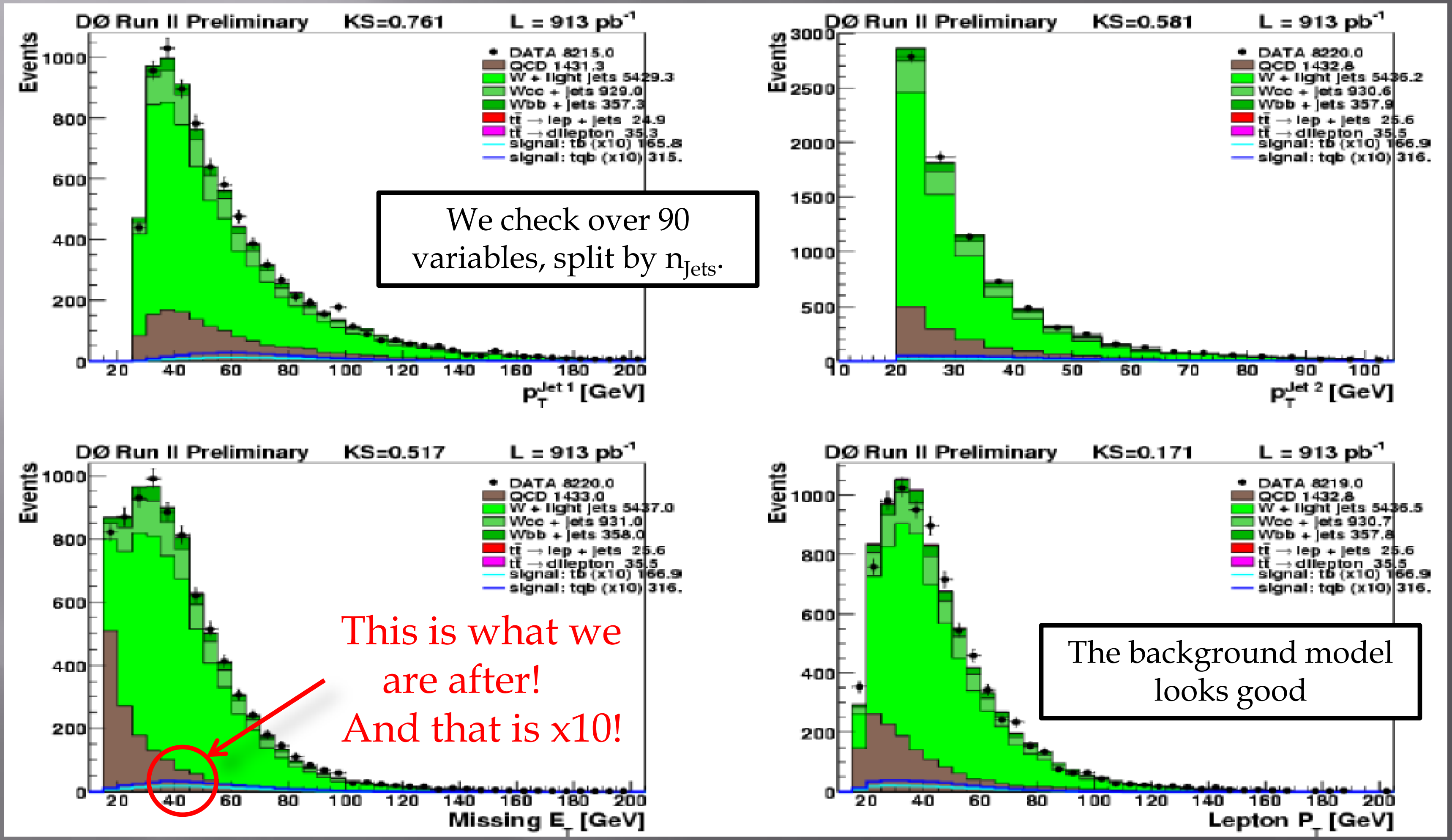
Known, Unknown

ε_{QCD} and ε_{WJets} are determined on sample with relaxed isolation criteria.



Fake rate dependence as a function of η is taken into account

Agreement Before b-Tagging

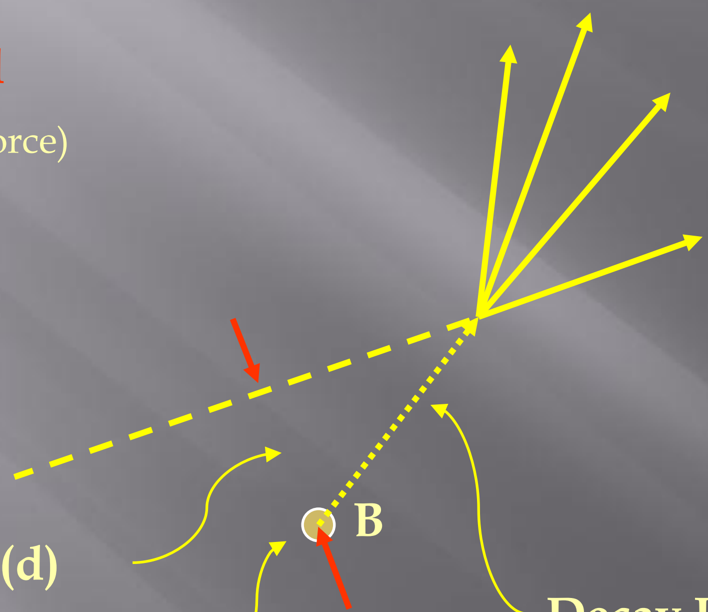


B Tagging

A B is Long Lived

(decays via weak force)

- Top, Higgs contain b-quark jets
  - Most backgrounds do not
- Jets look like any light quark jet
- Other than contain a B meson
  - Has finite life time
  - Travels some distance from the vertex before decaying
    - ~ 1mm
    - With charm cascade decay, about 4.2 charged tracks



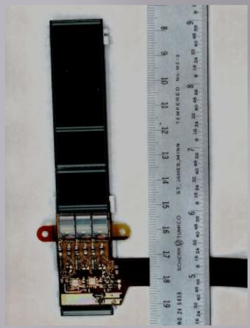
Impact Parameter (d)

Hard Scatter

Decay Lengh (L_{xy})

➡ Impact Parameter Resolution	$d/\sigma(d)$
➡ Decay Length Resolution	$L_{xy}/\sigma(L_{xy})$

All algorithms take advantage of these basic features



An Event

"Proton View"

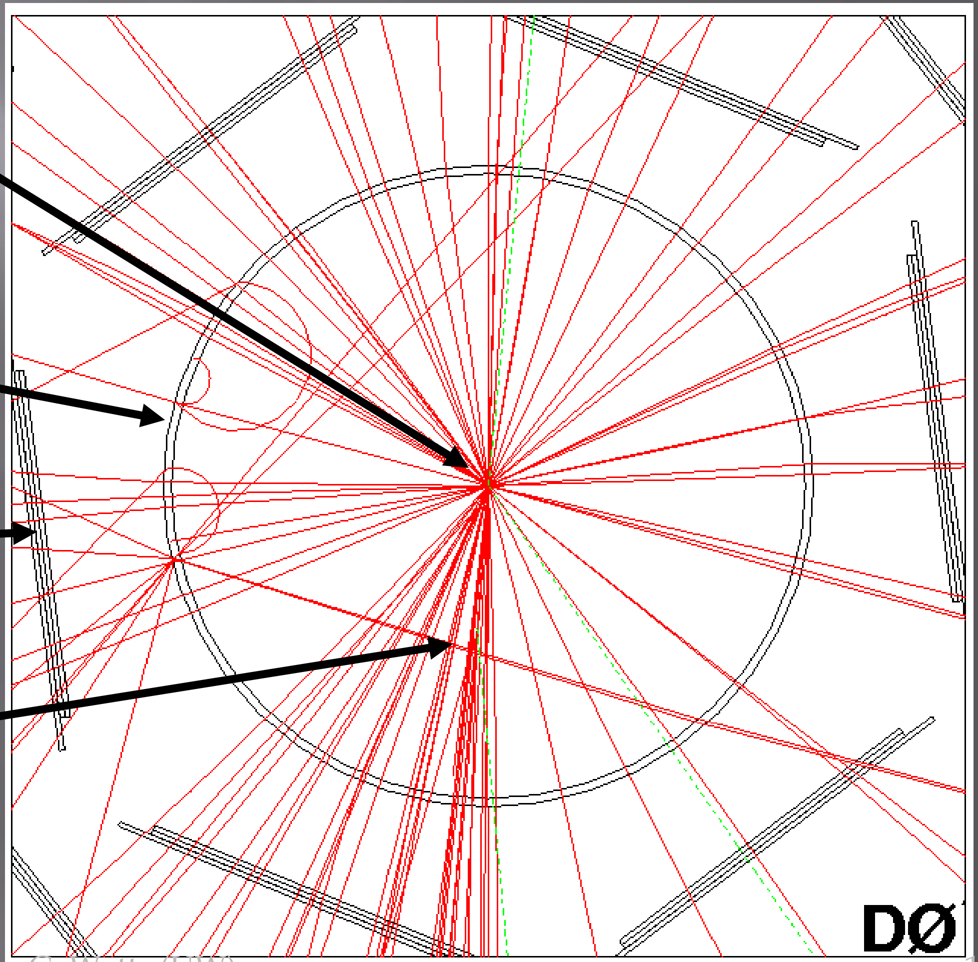
Hard Scatter
(Primary Vertex)

Beampipe
(2.3 in diameter)

Layer Of Silicon

Reconstructed
Secondary Vertex

Green Track Is Displaced

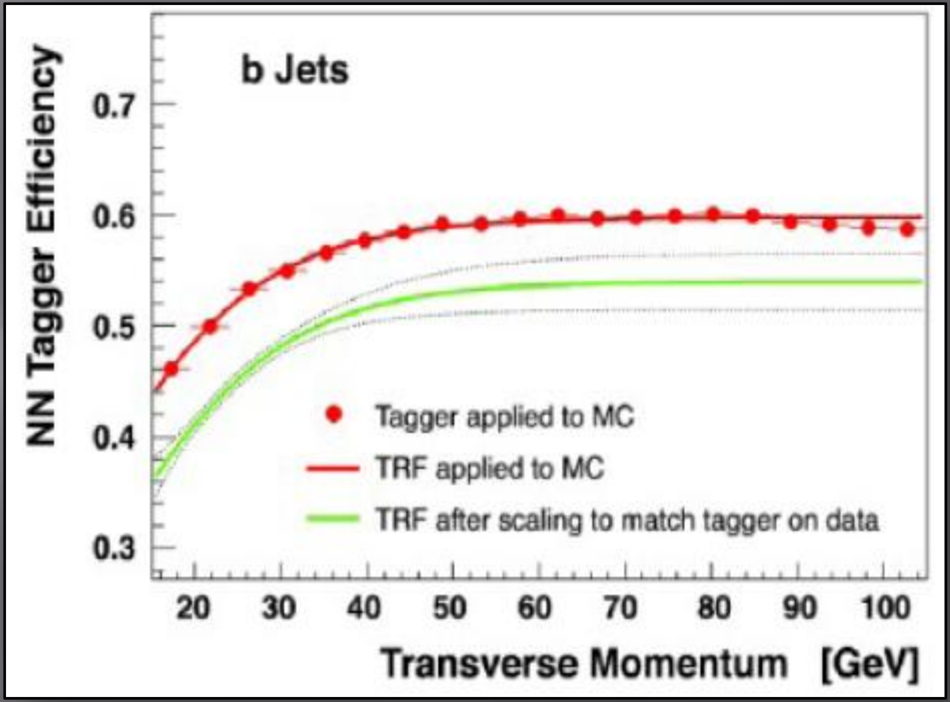


NN Algorithm & Performance

NN Algorithm

- Use 3 older tagging algorithms as input
  - Vertex reconstruction based
  - Probability Based
- Mass, decay length, etc.
- Trained on Monte Carlo
- Performance measured on data

30% performance improvement over individual taggers



Systematics

Function of the jet p_T and η!

- V0's (K_s, Λ, etc.)
- Tracking Resolution/MC matching
- Charm content
- Gluon Splitting to bb

Tagging in Data is easy… Monte Carlo is a bit trickier

Apply Tagging To MC

Tag Rate in MC and Data

b-tagging in MC is 15-20% more efficient

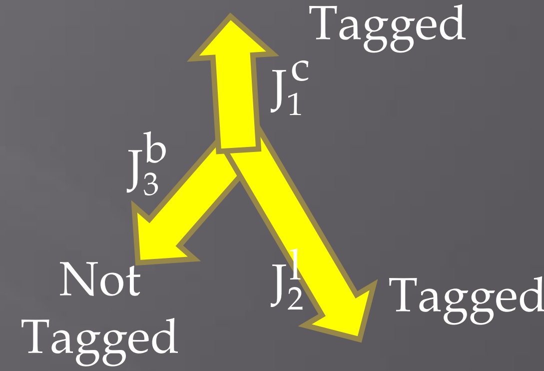
Final Variables Require Tagged Jets

Can't just weight the event. Either:

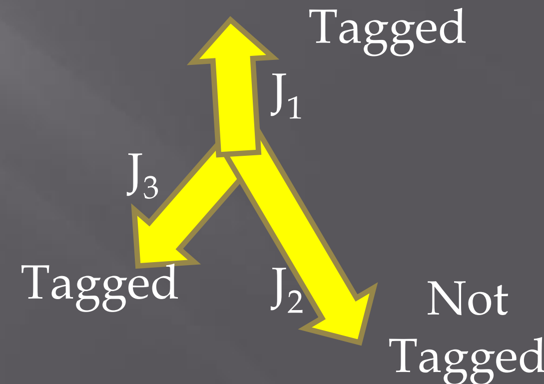
A. Run tagger on MC and apply Data/MC Scale factor on a jet-by-jet basis.
   - Requires large statistics to model light quark tags

★ B. Permute the event through every possible tag configuration
   - Assign weight based on probability of that configuration.

Same Event appears multiple times in sample with different tagging configuration and event weight.



$W=P_T(J_1)P_T(J_2)P_{NT}(J_3)$



$W=P_T(J_1)P_{NT}(J_2)P_T(J_3)$

Splitting Data by S:B

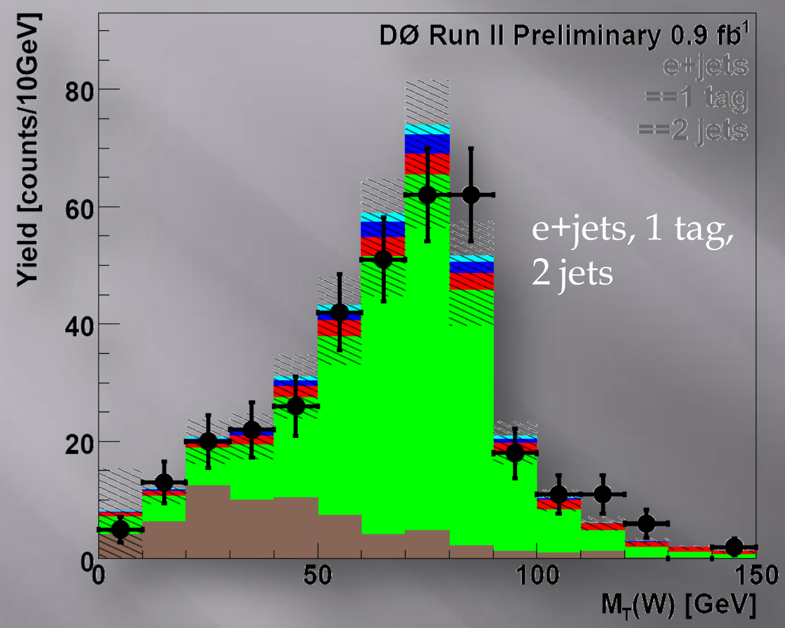
Partitioning our dataset by S:B will prevent backgrounds from contaminating especially sensitive regions of parameter space

Event selection and S:B

Percentage of single top *tb+tqb* selected events and S:B ratio (white squares = no plans to analyze)

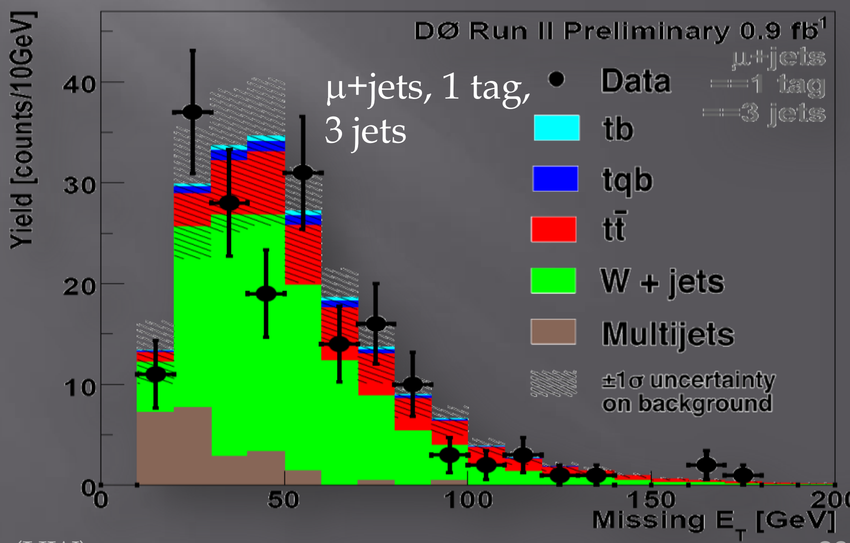
Electron + Muon	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
0 tags	10% 1 : 3,200	25% 1 : 390	12% 1 : 300	3% 1 : 270	1% 1 : 230
1 tag	6% 1 : 100	21% 1 : 20	11% 1 : 25	3% 1 : 40	1% 1 : 53
2 tags		3% 1 : 11	2% 1 : 15	1% 1 : 38	0% 1 : 43

Post-Tagging Agreement



Sample	# of Events
s&t-channel Signal	62
Wjj	174
tt→l+jets	266
Wbb & Wcc	675
Mis-ID's leptons	201
Diboson,tt→ dileptons	82

Totals	2 Jets	3 Jets	4 Jets
Data	697	455	246
Total Background	685	460	253
Signal	36	20	6



Systematic Errors

Many Sources of Error

- Theoretical cross sections
- Heavy flavor fraction
- Luminosity
- Jet energy scale
- b-tag rate

Assigned Per Sample/Channel

- By channel (each lepton, jet, tag bin)
- By sample/source
- Correlations between samples are accounted for.

Relative Systematic Uncertainties

$t\bar{t}$ cross section	18%	Primary vertex	3%
Luminosity	[illegible]	[illegible]	[illegible]
Electron trigger	[illegible]	[illegible]	[illegible]
Muon trigger	[illegible]	[illegible]	[illegible]
Jet energy scale	[illegible]	[illegible]	[illegible]
Jet efficiency	[illegible]	[illegible]	[illegible]
Jet fragmentation	5–7%	$\varepsilon_{\mathrm{real}-\mu}$	2%
Heavy flavor ratio	30%	$\varepsilon_{\mathrm{fake}-e}$	3–40%
Tag-rate functions	2–16%	$\varepsilon_{\mathrm{fake}-\mu}$	2–15%

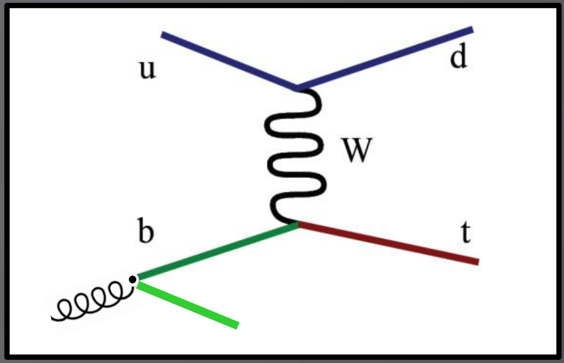
But it is hard to judge which is most important from this table – its impact on the final result depends on how large the same is that it applies to, or how much an effect it has on the sample we are looking at.

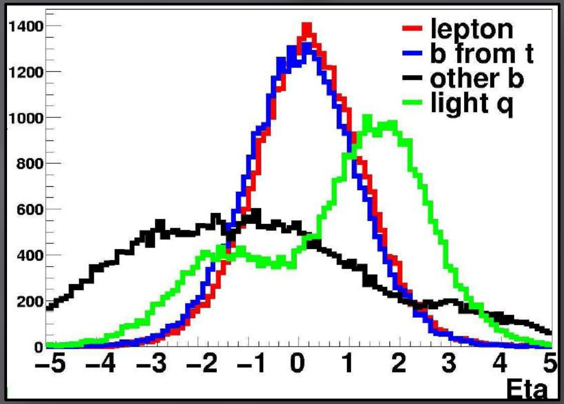
Separating Signal and Background

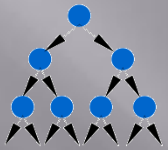
We have a well understood sample with large signal acceptance

- S:B is 1:20 and in some channels 1:40
- Large irreducible physics backgrounds (e.g. Wbb)
- Signal and background have different shapes due to production kinematics

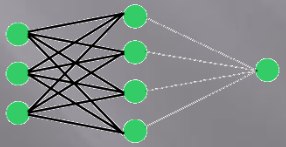
⇨ Take advantage of shape and extract the signal using multivariate techniques







Boosted Decision Trees
Trained, discriminating variables

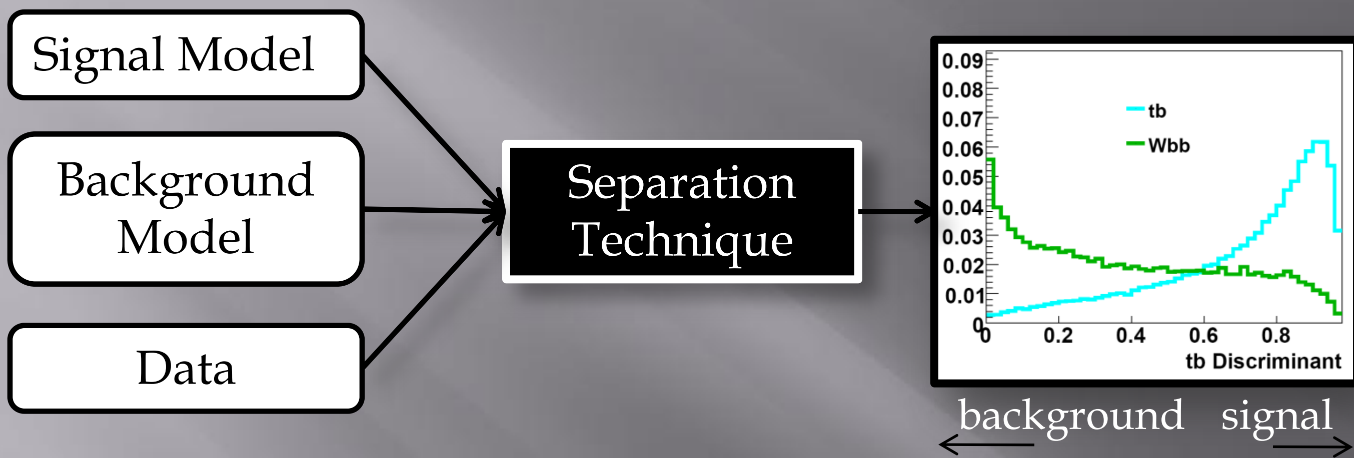


Bayesian Neural Networks
Trained, discriminating variables

$\int M$

Matrix Element
4 vectors and MC LO matrix elements

Black Box



If we get our background model right the separation technique doesn't matter

We care about the separation technique only in as much as any correlations it counts on correctly modeled background model.

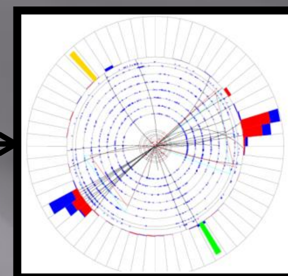
- Check background model on ~50 variables
- Cross check against orthogonal data samples
- Does our data behave as expected vs. separation parameter?

Matrix Element

Monte Carlo Generator

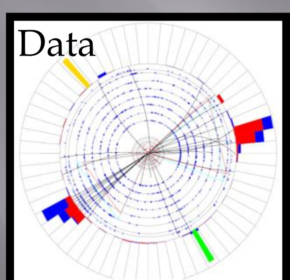
e.g. MadGraph

Random Number Generator → Matrix Element →



Produces an event topology according to ME probability

Reverse Monte Carlo Generator



→ Matrix Element → Probability of event topology

One for each background and signal type

Problem:

- ME deals in final state partons and PDF's.
- Data has detector and reconstruction effects!

Matrix Element Introduction

The probability a measured detector topology (x) is a particular process (M):

$$P(\vec{x}) = \frac{1}{\sigma} \int \underbrace{f(q_1;Q)dq_1\, f(q_2;Q)dq_2}_{\text{CTEQ6 Parton Distribution Functions}} \times \underbrace{|M(\vec{y})|^2 \phi(\vec{y}) dy}_{\text{Leading Order ME from MadGraph and phase space \& parton level cuts}} \times \underbrace{W(\vec{x},\vec{y})}_{\text{Transfer Function}}$$

Every possible final state parton configuration

$$\int f(q_1;Q)dq_1\, f(q_2;Q)dq_2 \times |M(\vec{y})|^2 \phi(\vec{y}) dy$$

Transfer function is probability a particular object 4-vector could have come from a final state parton

$$W(\vec{x},\vec{y})$$

Matrix Element Introduction

Transfer function is probability a particular object 4-vector could have come from a final state parton $W(\vec{x}, \vec{y})$

Matrix Element Introduction

Transfer function is probability a particular object 4-vector could have come from a final state parton $W(\vec{x}, \vec{y})$

Object Matching

Number of jets must match number of partons!
→ Simulating missing jets is very difficult.

Integration is Expensive

Using 4-vectors of all reconstructed leptons and jets
Using b-tag information to help decide which quark is a b-quark
Assume masses and momentum and energy conservation
End up with 4 independent variables
→ It still takes >60 seconds per event!
Don't do ttbar in 3-jet bin
Don't look at 4-jet bin

Have to run on every MC event!!

Detector Response

Transfer Functions

Assume detector response is separable

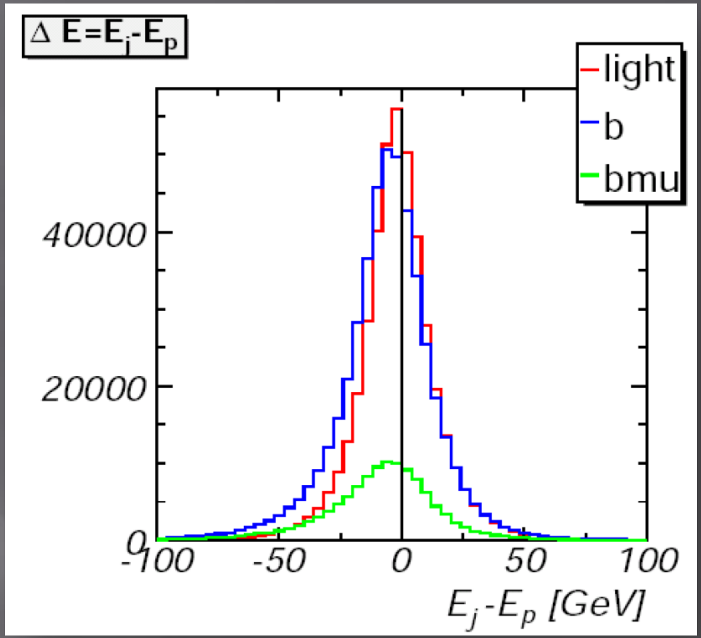
$$W(x,y) = W_{jet}(x,y)W_{electron}(x,y)$$

Determined From Monte Carlo

Jets
- By flavor, E, and η.

Electrons
- By E and η.

Muons
- By $1/p_T$, Silicon Hit (or not).

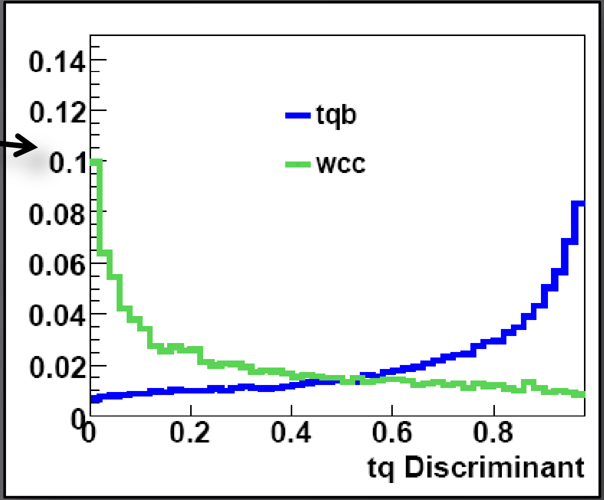


Shared!

Expensive to calculate: same ones as used by the top mass analysis

ME Discrimination

$$P(S|\vec{x}) = \frac{P_S(\vec{x})}{P_S(\vec{x}) + P_B(\vec{x})}$$

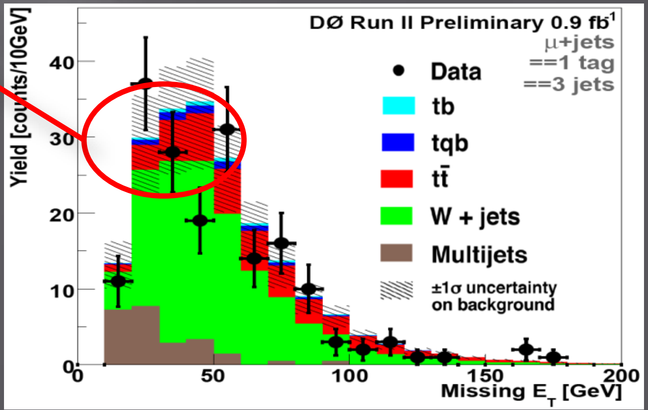


W+ 2 Jet Events: ME for Wbg, Wcg, and Wgg
W+ 3 Jet Events: ME for Wbbg

ttbar is a major background in 3 jet events

- Currently work in progress to add it
- Severely limits the separation in the 3 jet bin as ttbar and s-channel look very similar!

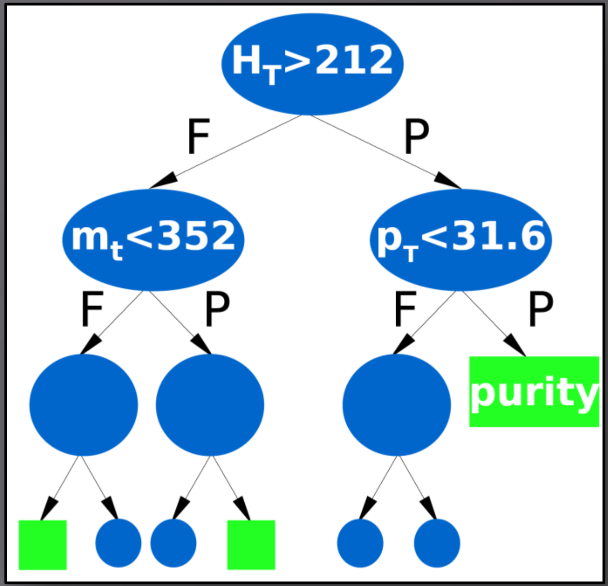
μ+jets, 1 tag, 3 jets



Decision Tree

Machine Learning Technique

- Train on 1/3 of our background and signal model
- Creates a tree, with a simple straight cut at every branch point
- Each leaf classifies an event with a purity
- Performance measured on the other 2/3's of our signal and background model



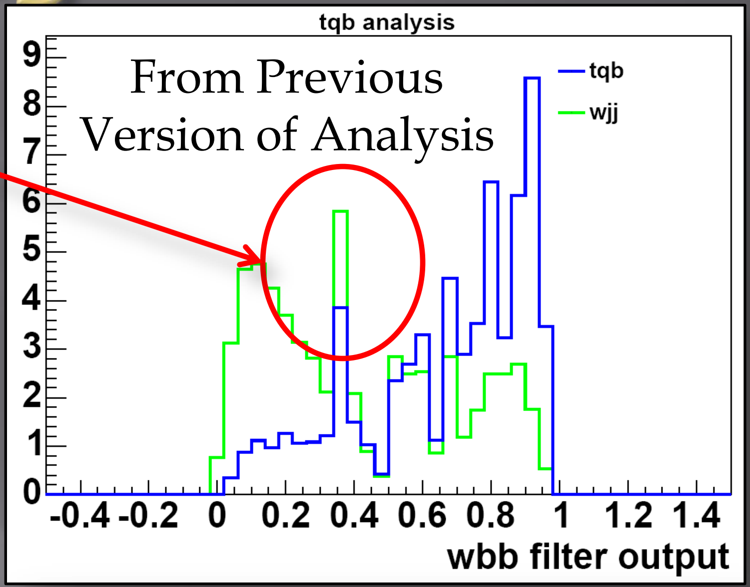
Input Variables

- A DT is not good at finding complex correlations because of its straight cut methodology
- We used the standard input variables (H_T, ME_T, M_{Top}, M_{WT} etc.)
- Use more complex angular variables motivated by leading order matrix elements
- 45 in total – most important are tagging related, H_T, etc.
- DT automatically sorts out which ones are interesting

Boosting

Single decision tree has some problems

- Leaves are discrete – can lead to funny spikes
- That plot contains more than ample statistics!
- Misclassifies more events than it needs to

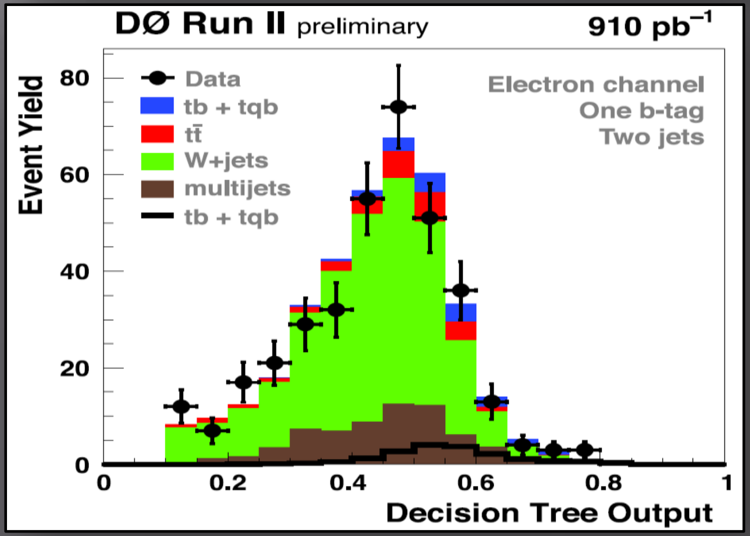


Boosted Decision Trees

Boost the weight of misclassified events and train to derive a new tree.

The result is the weighted sum of 20 trees

- Smoother distributions
- Better separation
- More stability

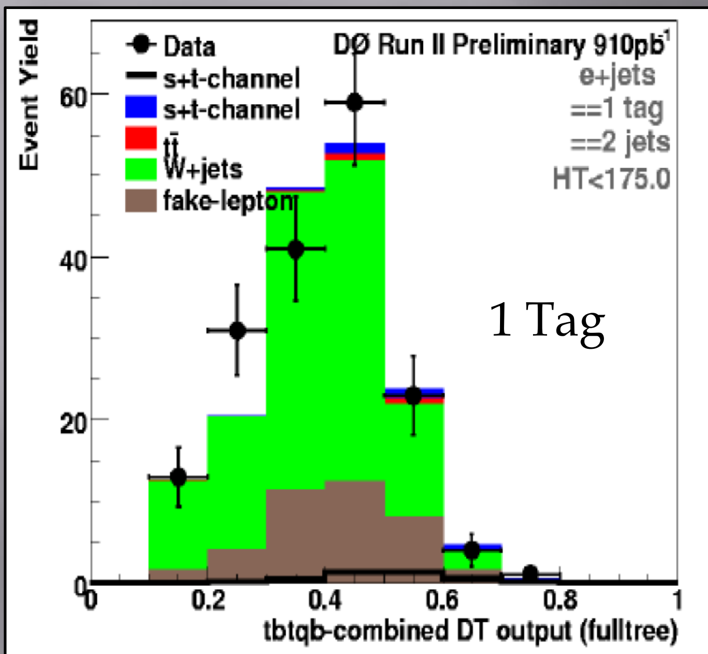


Matrix Element vs. Decision Trees

Decision Tree	Matrix Element
You must come up with the important variables and correlations to separate	All separation power is encoded in the matrix element
Very fast: retraining the entire analysis is less than an hour. Ideal for rapid turn around	Really slow. Adding a new matrix element can be weeks of processing time. Don't make a mistake!!
Trivially extendable to NLO generators	Will take some work to extend to NLO guys
As good as your input variables	All things equivalent will probably be able to squeeze more out of your data
Fairly easy to understand the mechanics; training parameters are well studied by the statistics community	Complex to explain, details (transfer function, parton level cuts, etc.) can be arcane.
Train against all background samples at once.	Requires separate ME for each process to discriminate against

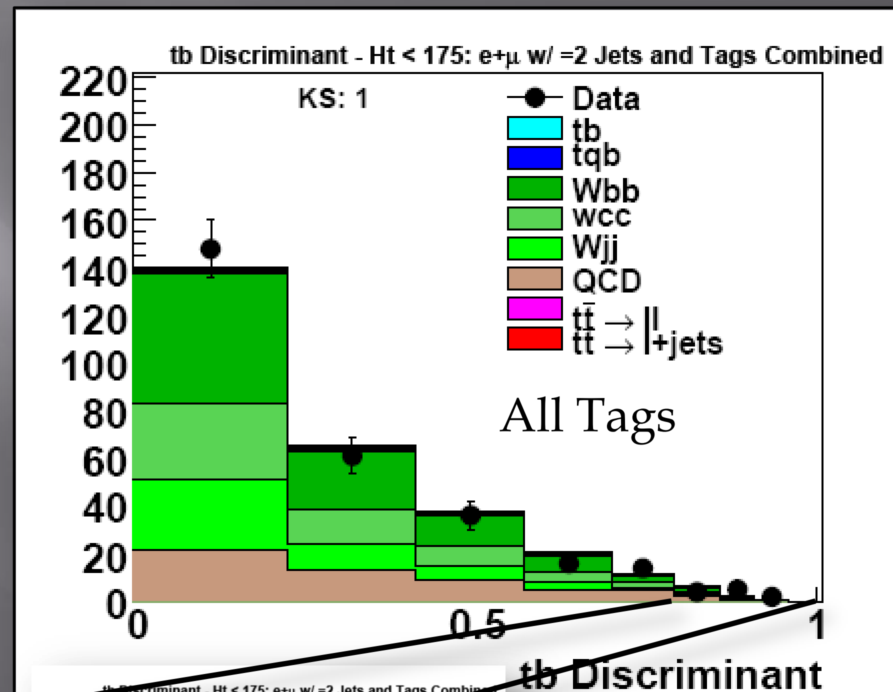
Cross Check: W+Jets Like

2-jets, H_T<175 GeV

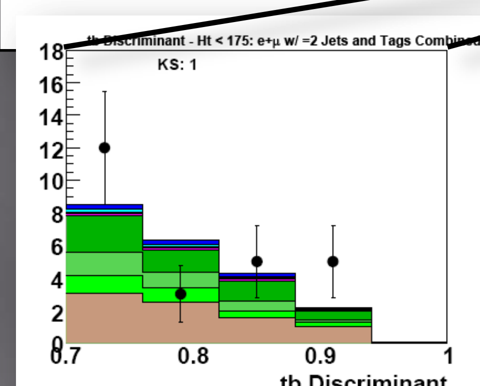


1 Tag

Decision Tree Output



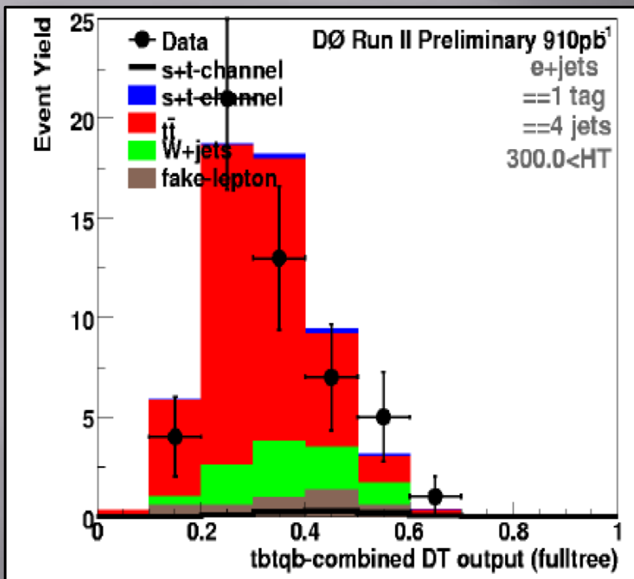
All Tags



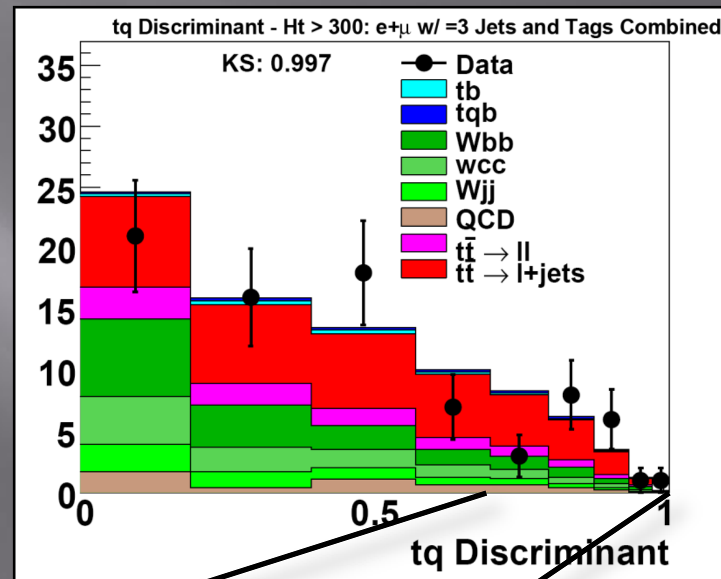
Matrix Element Output

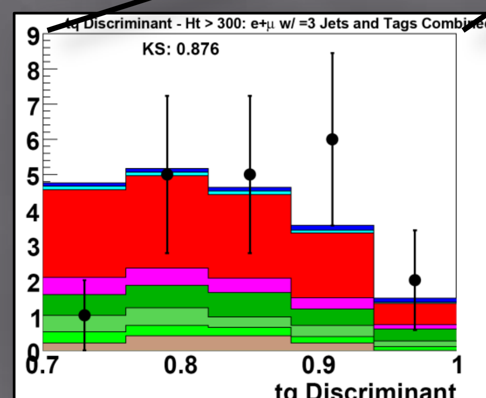
Cross Check: ttbar Like

H_T>300 GeV



Decision Tree Output

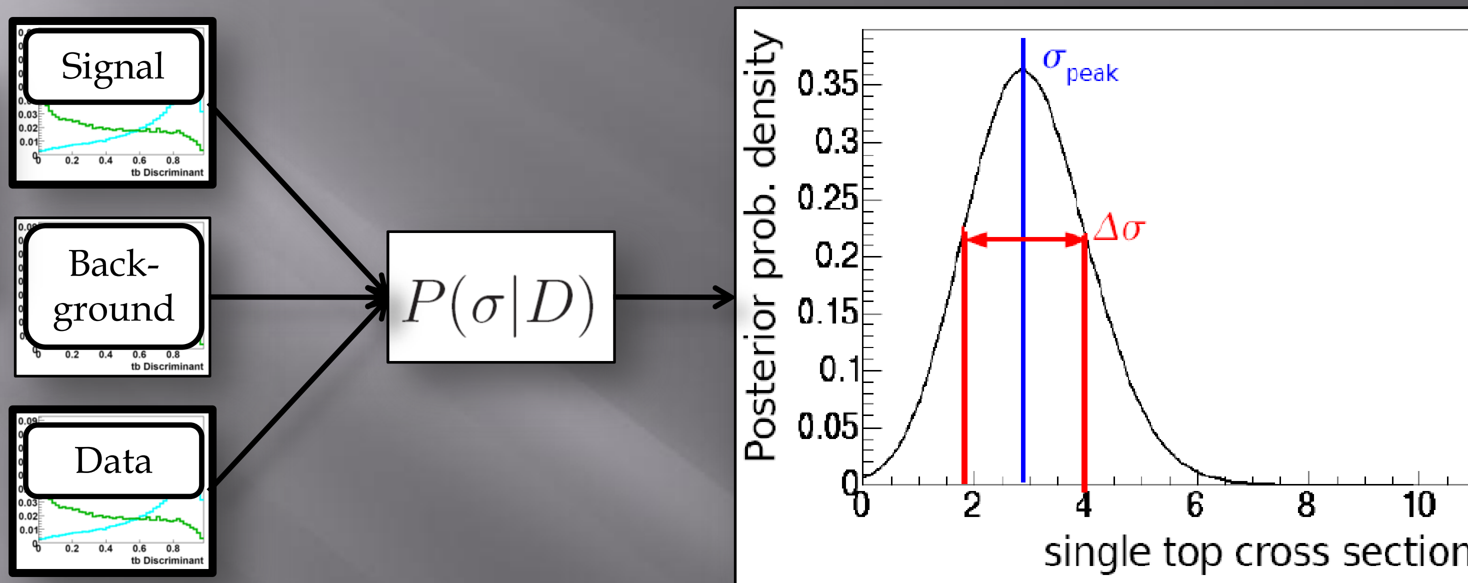




Matrix Element Output

Cross Section Determination

Bayesian calculation of the cross section



$$P(\sigma|D) \propto \int\int \ldots \int L(D|\sigma, a, \mathbf{b})\pi(a, \mathbf{b})\, da\, d\mathbf{b}.$$

Cross Section

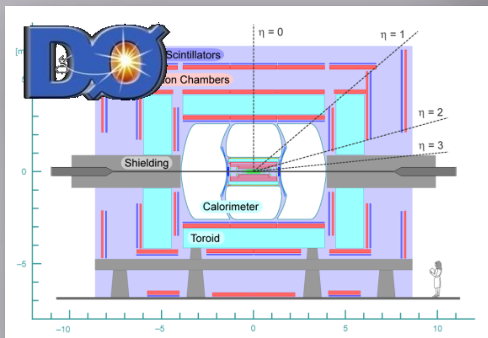
Observed Data

Signal Acceptance

Probability of this signal and background

Background

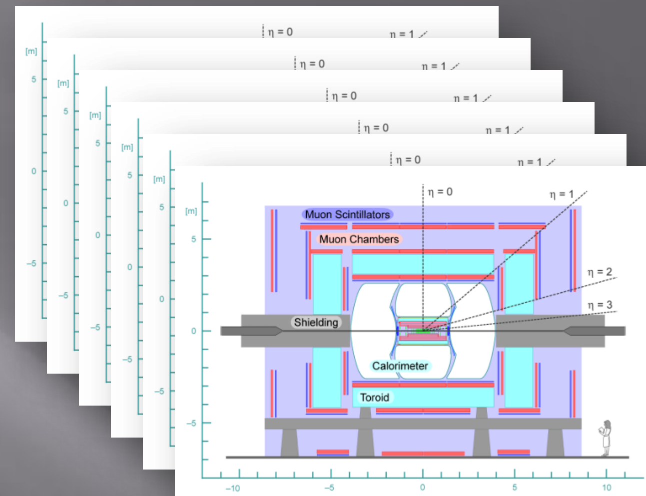
Ensemble Tests



THE Result

For a given expected signal cross section

1. Poisson sample from signal and background sample # of events seen in real experiment
   - Take into account systematics
   - Take into account correlations
2. Run the full analysis

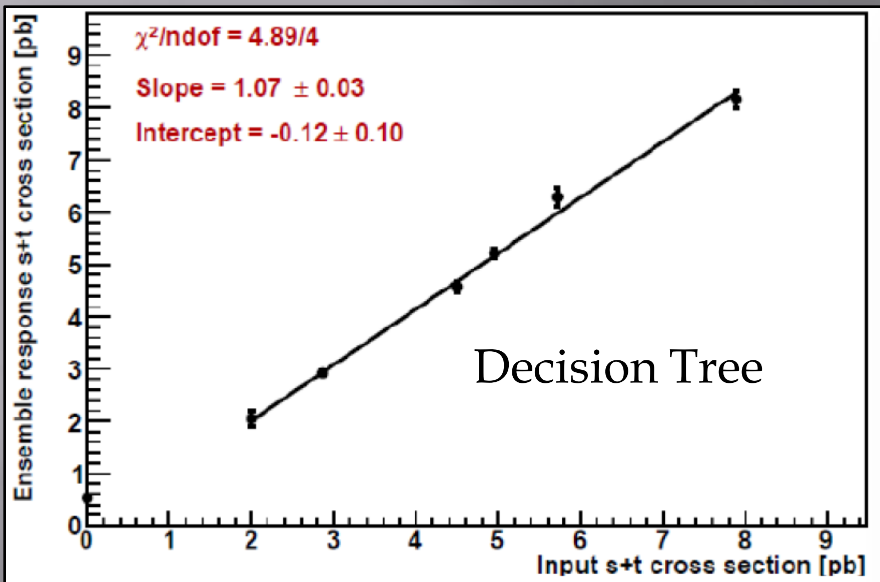


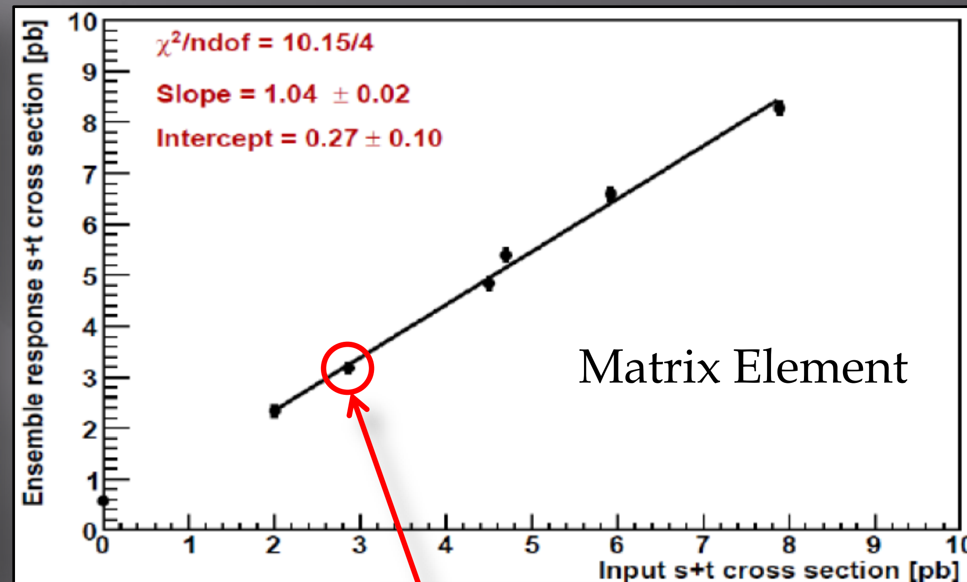
Distribution of Results

Good at analysis method and statistics test.

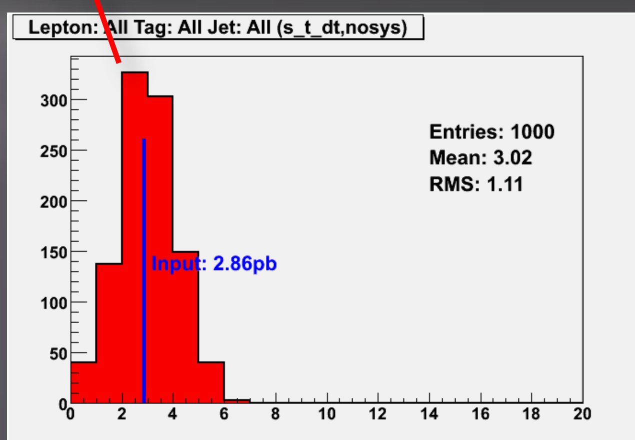
Won't detect a missing error or fatal flaw in background model

Linear Response in Cross Section





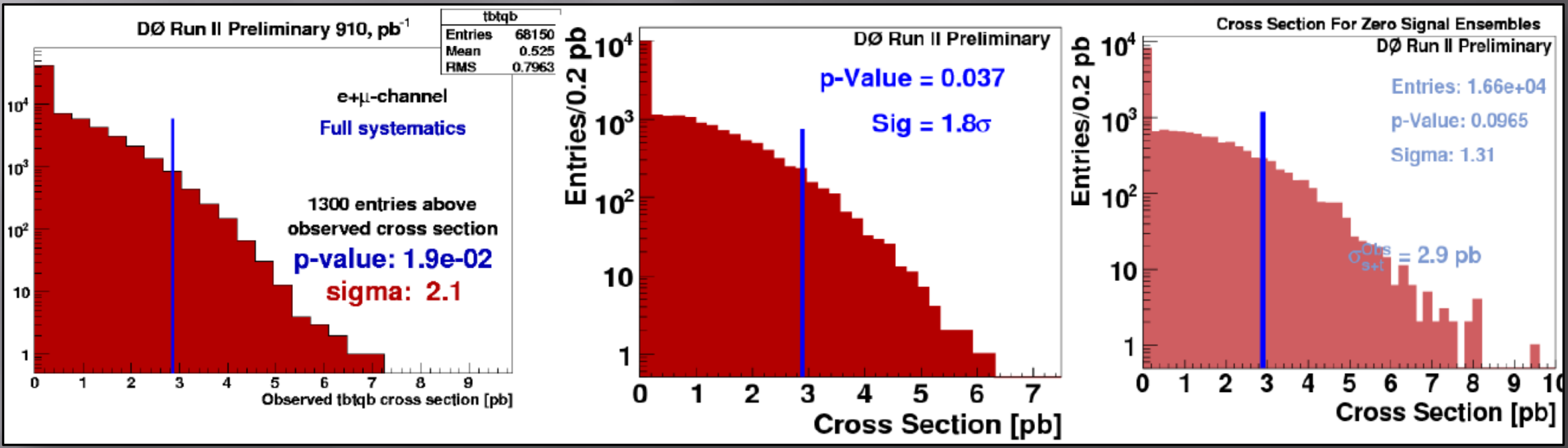
- Some of the input samples were blind (had unknown cross sections)
- All three analysis methods (DT, ME, and BNN) are close to linear.



Expected Sensitivity

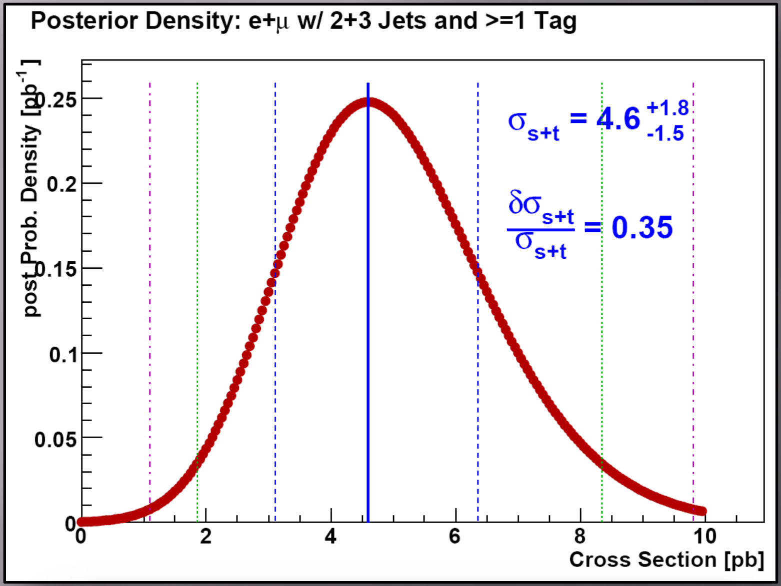
➡ A large zero signal ensemble can answer a number of crucial questions

Q: What fraction of the zero signal datasets have a measured cross section of a least 2.9 pb?



Decision Tree	Matrix Element	Bayesian NN
1.9%	3.7%	6.5%

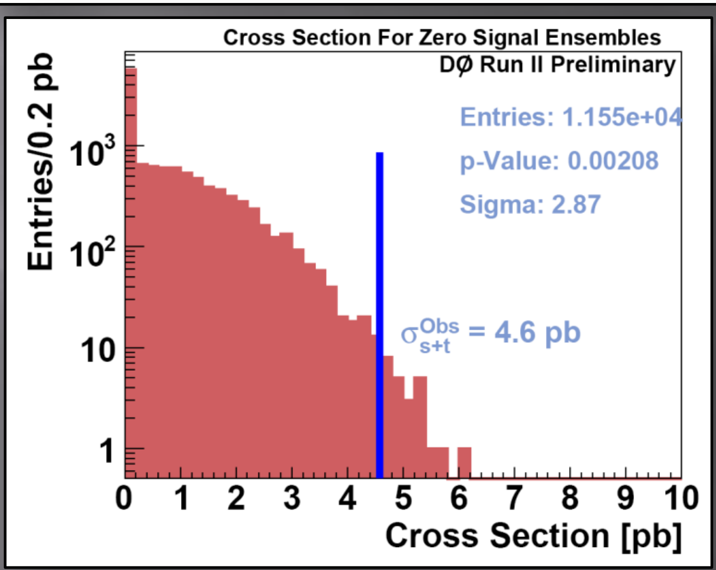
Matrix Element Results

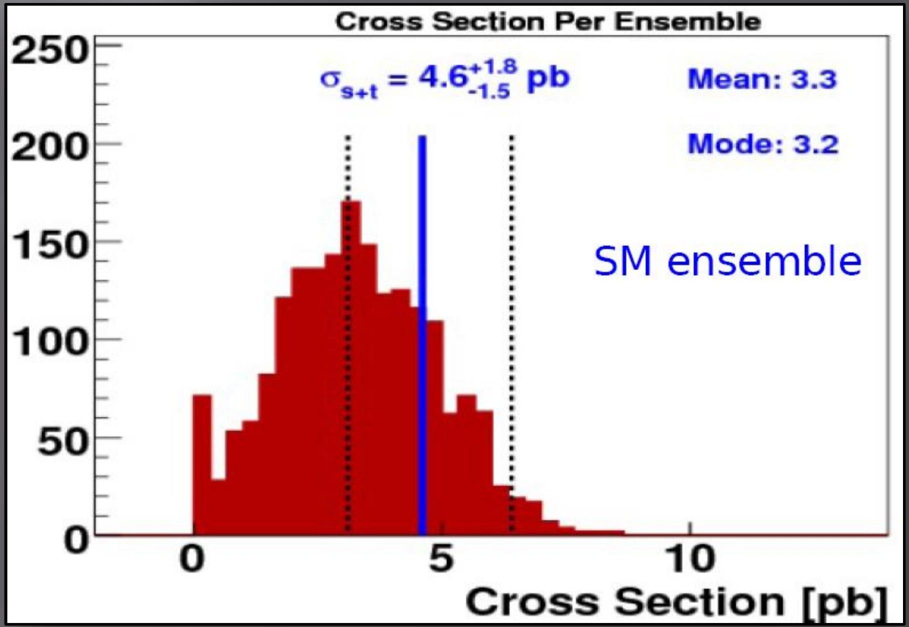


σ_{s+t} = 4.6 +1.8 -1.5 pb

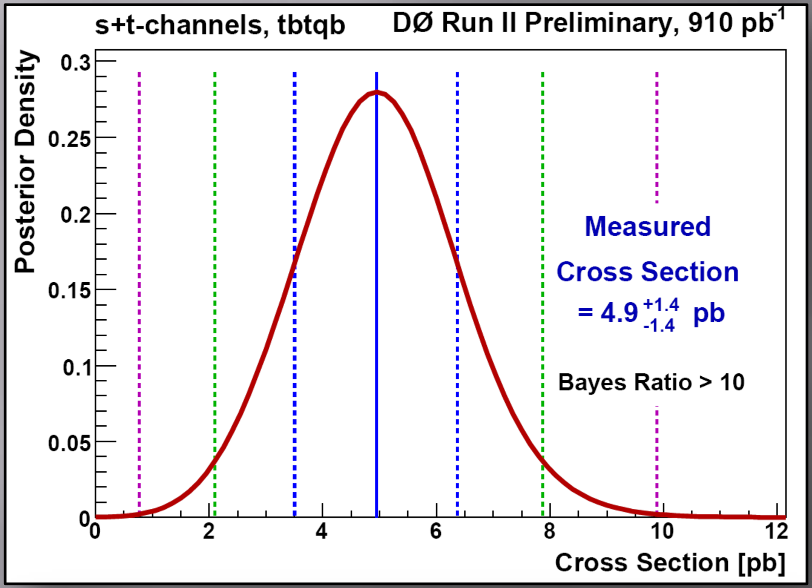
Significance: 2.9σ!

21 % of the SM ensemble is above 4.6 pb.



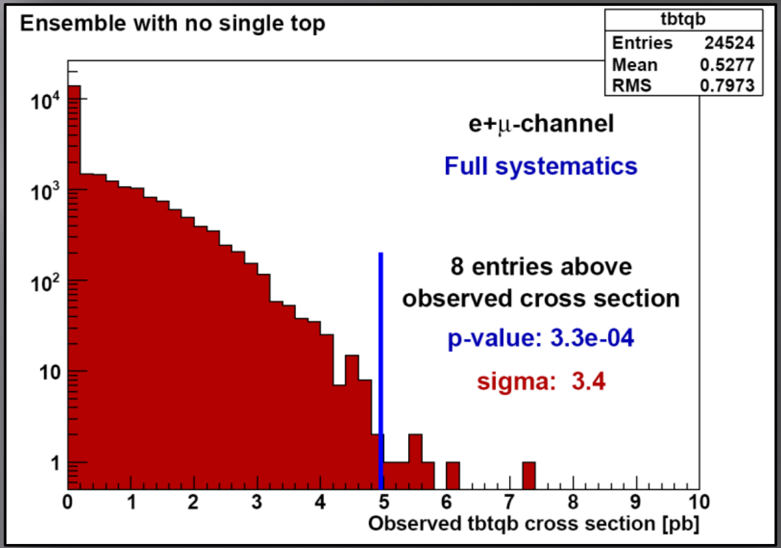


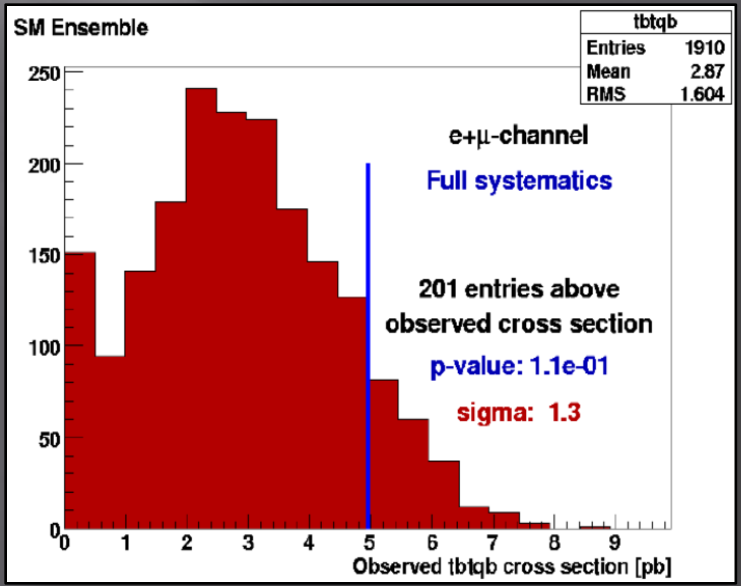
Decision Tree Results



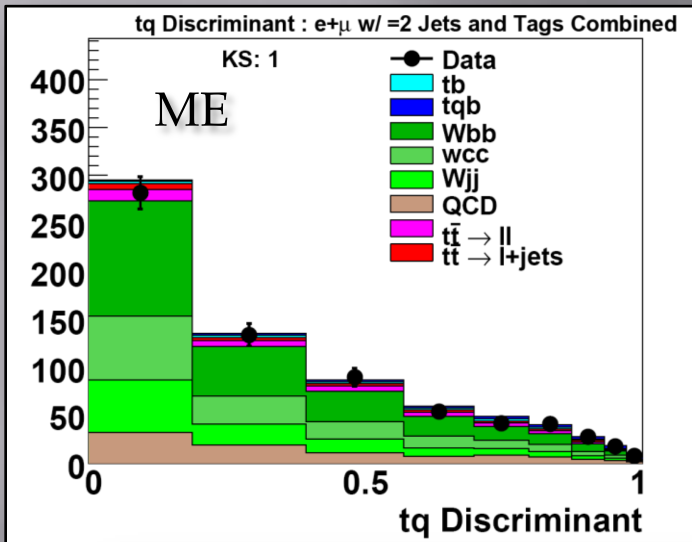
σ_{s+t} = 4.9 +1.4 -1.4 pb
Significance: 3.4σ!!

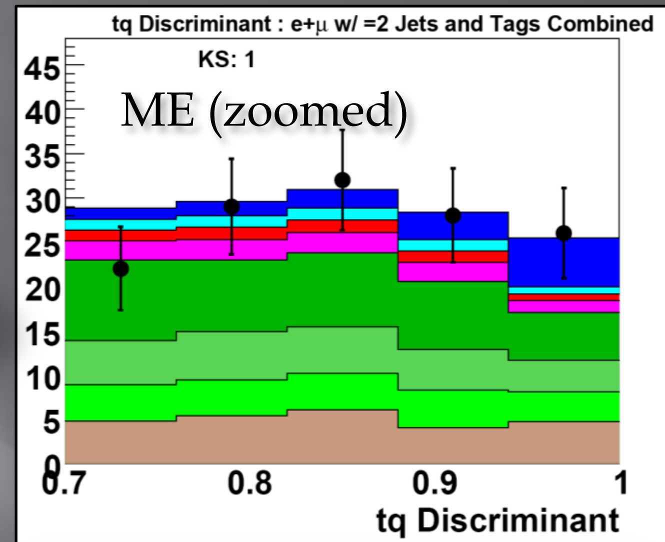
11 % of the SM ensemble is above 4.9 pb.

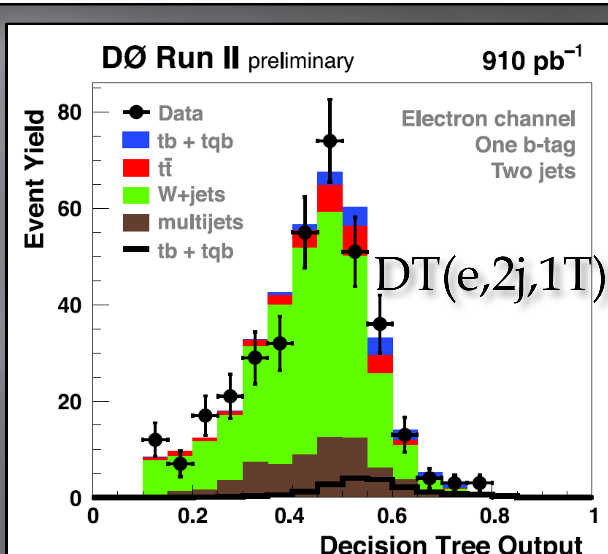


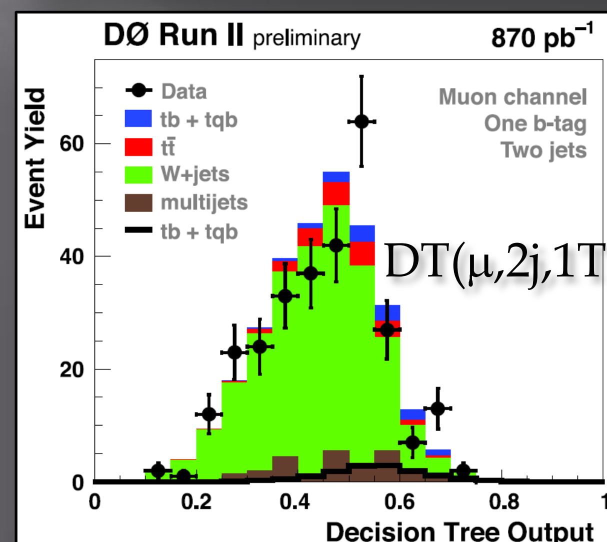


Sample Discriminate Outputs

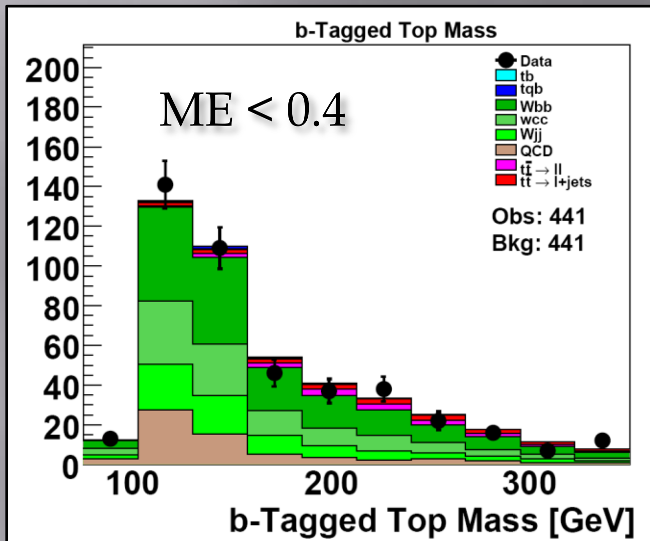


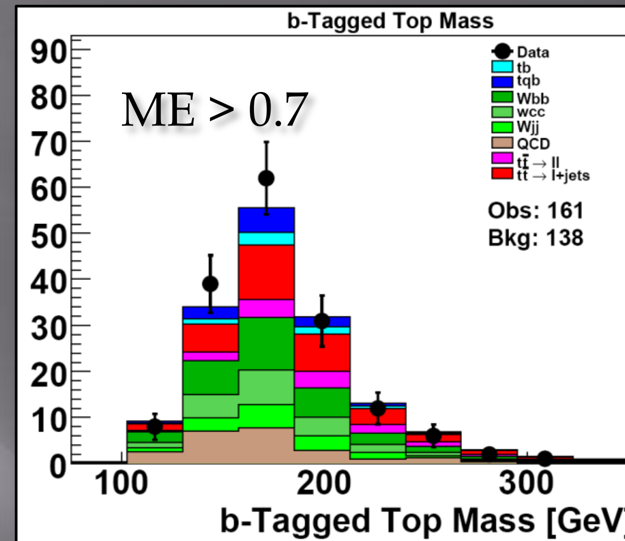


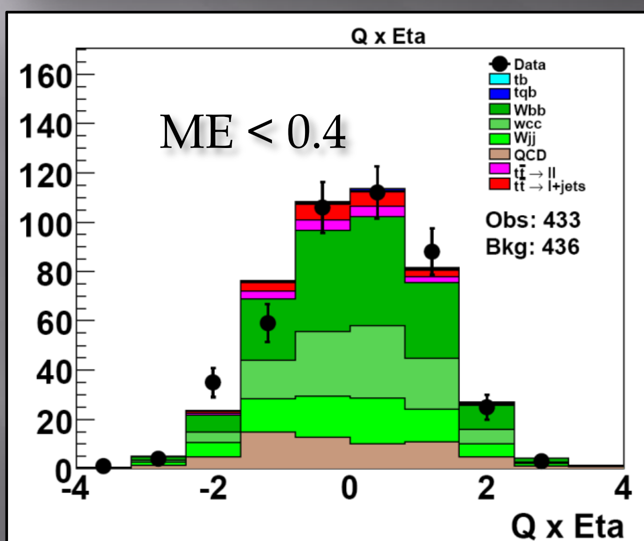


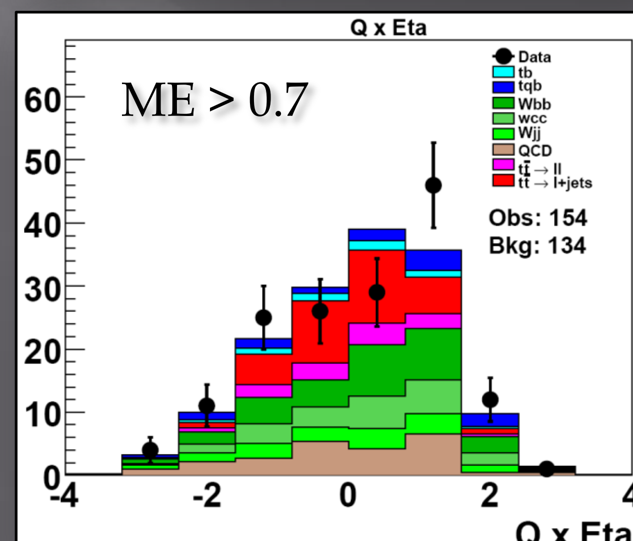


Look At The Data: ME

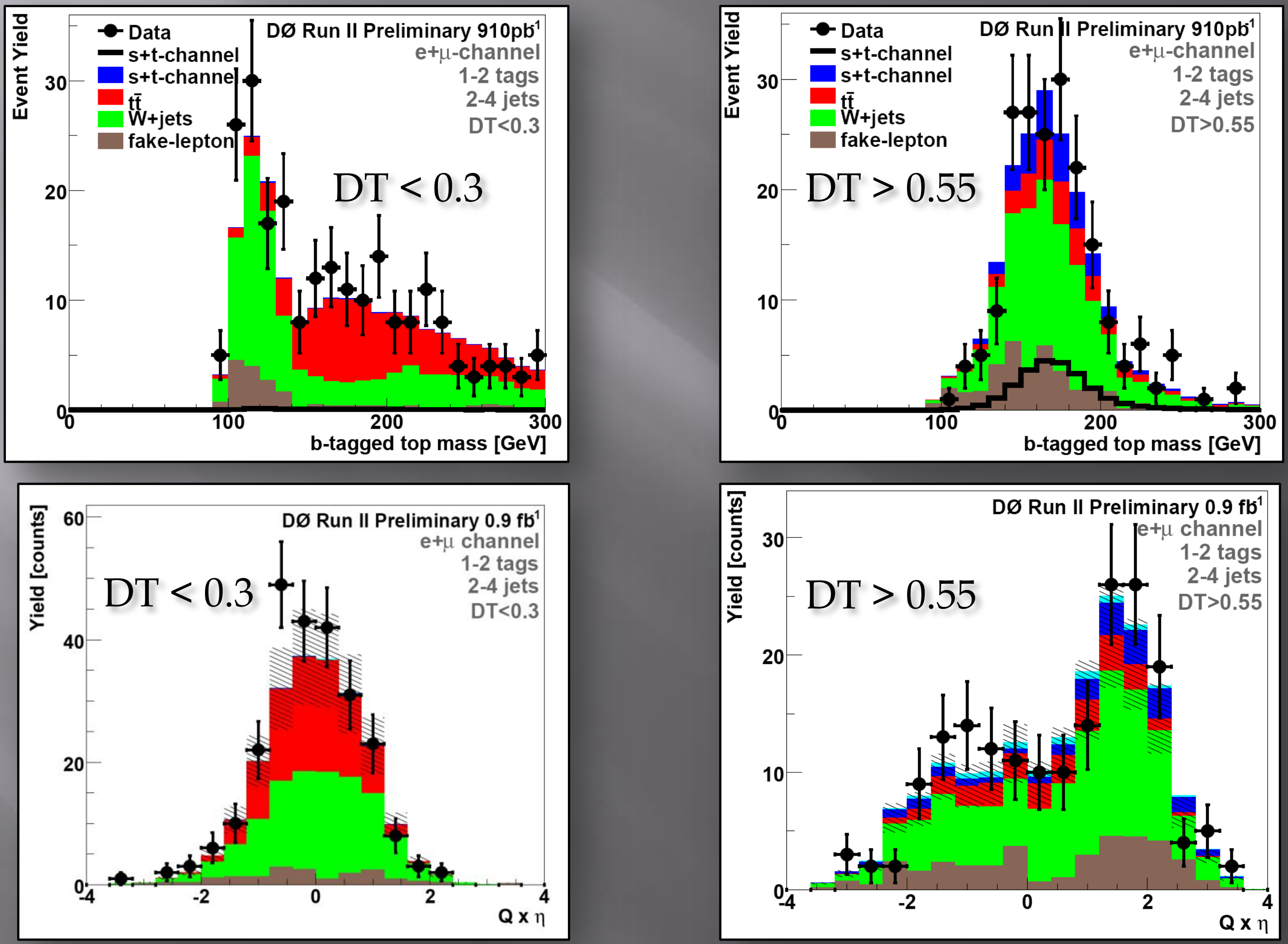




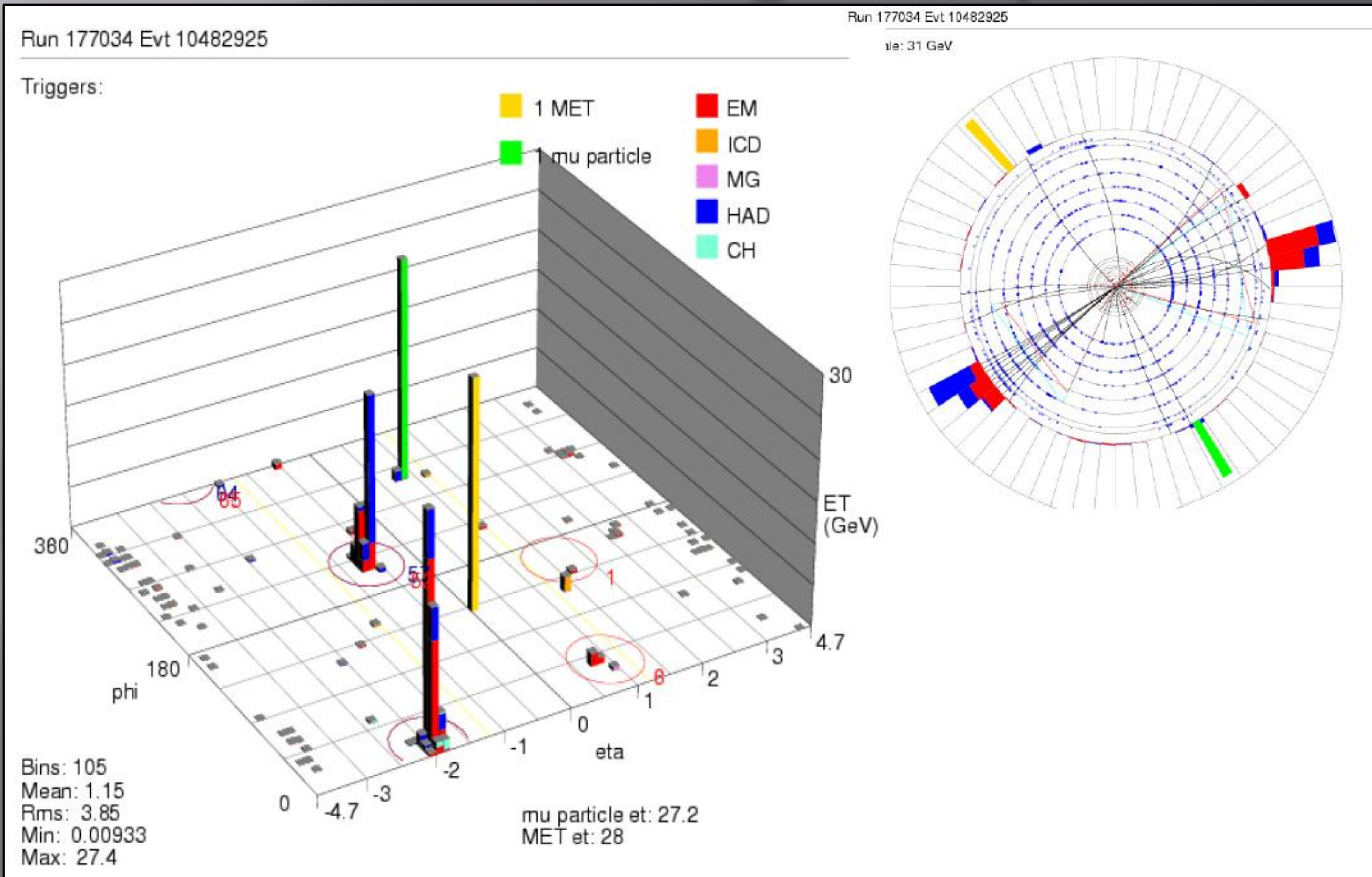




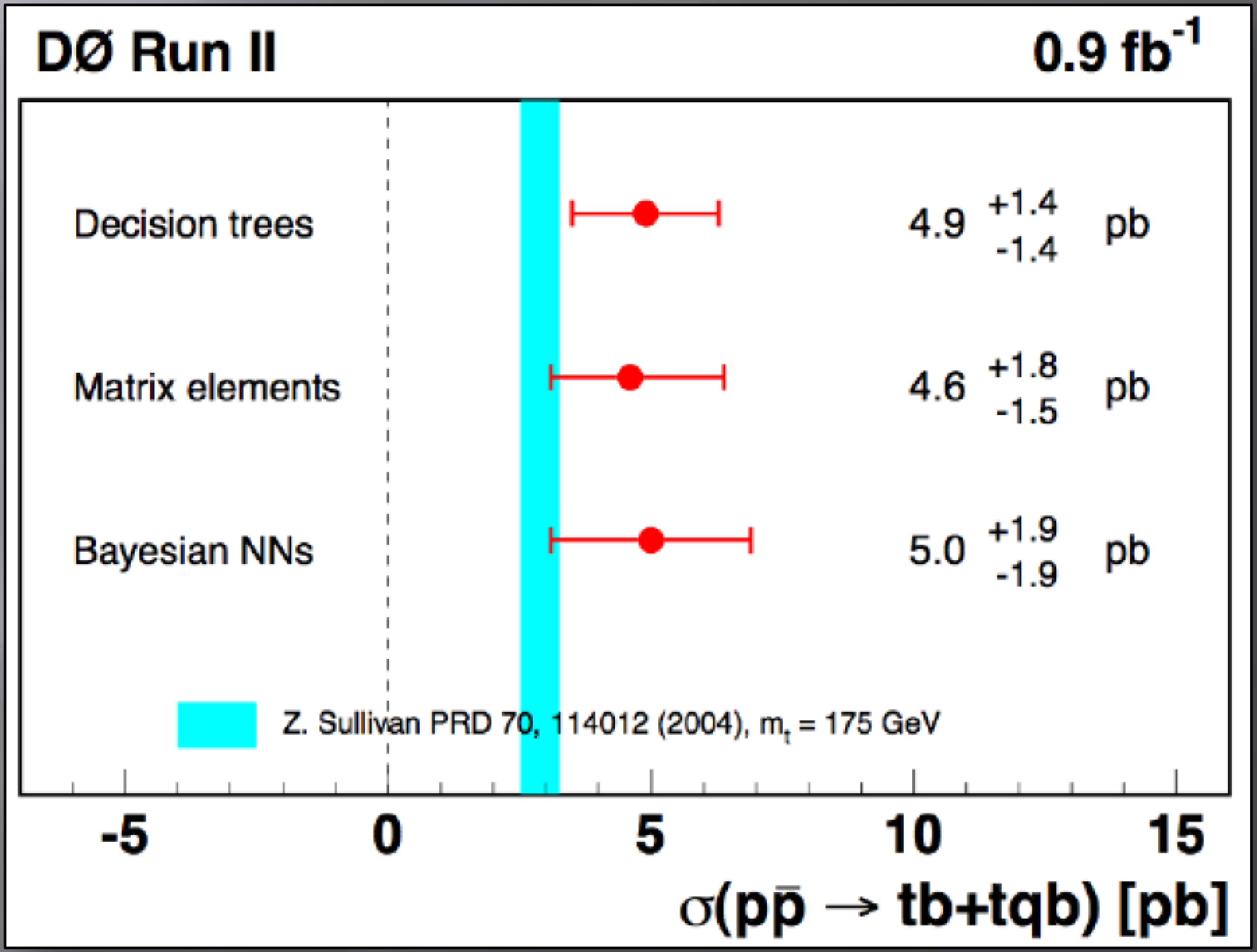
Look at the Data: DT



A Candidate Single Top Event

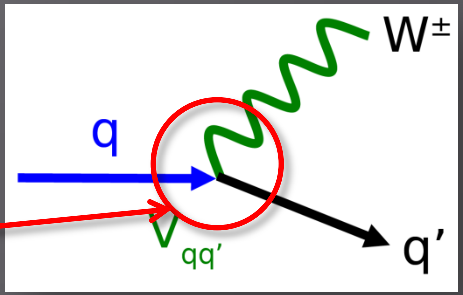


Search for Single Top Summary



The CKM and V_{tb}

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



Weak interaction eigenstates and mass eigenstates are not the same: mixing occurs between the quarks

The CKM matrix

Standard Model Top Decays

$V_{td}^2 + V_{ts}^2 + V_{tb}^2 = 1$

V_{td} and V_{ts} well constrained: $V_{tb} > 0.998$

Unitarity and 3 generations:

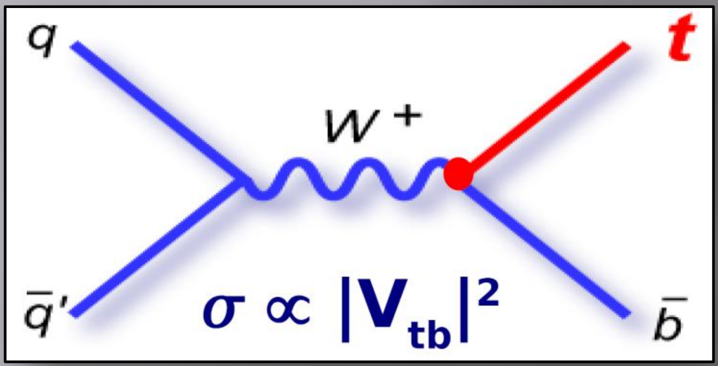
$Br(t \rightarrow Wb) \sim 100\%$

New Physics

$V_{td}^2 + V_{ts}^2 + V_{tb}^2 < 1$

V_{tb} must be measured!

Measuring $|V_{tb}|$



Use the same procedure to determine $|V_{tb}|$ as we did the cross section

We have to assume SM decays of the top quark

General Form of the Vertex:

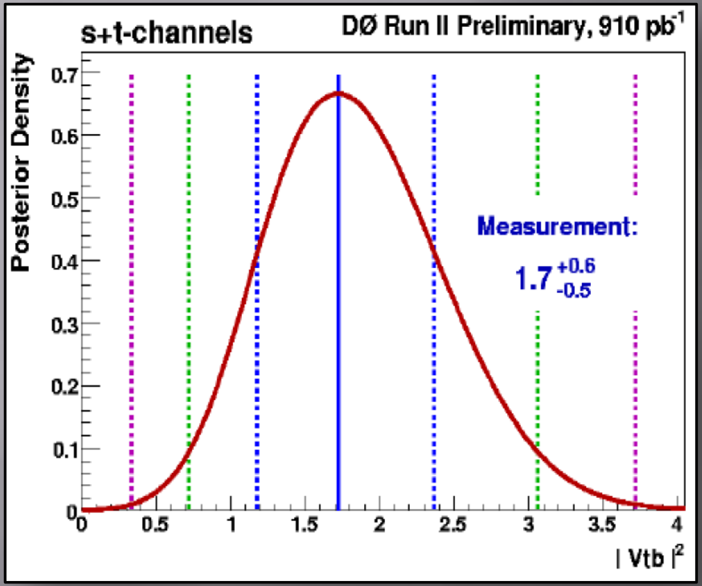
$$\Gamma^{\mu}_{tbW} = -\frac{g}{\sqrt{2}} V_{tb} \left\{ \gamma^{\mu} \left[f_1^L P_L + f_1^R P_R \right] - \frac{i\,\sigma^{\mu\nu}}{M_W} (p_t - p_b)_\nu \left[f_2^L P_L + f_2^R P_R \right] \right\}$$

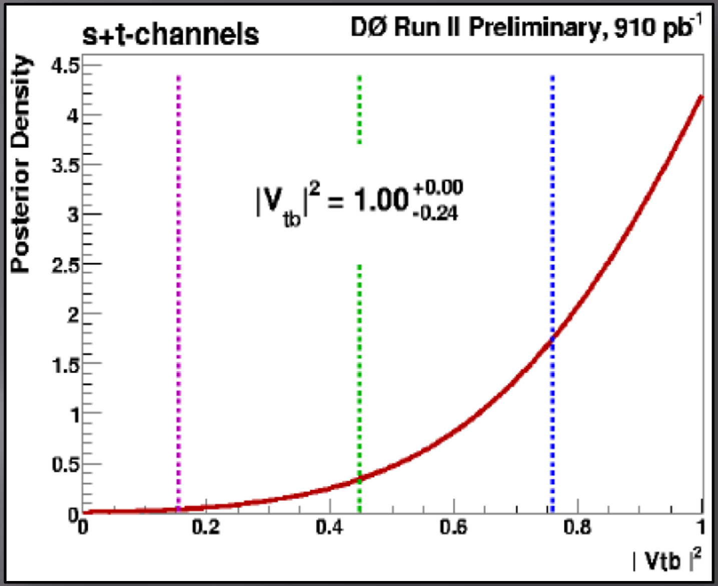
In the SM: 1   0   0   0

CP Conserved

Measuring $|V_{tb} f_1^L|$: strength of the V-A coupling

$|V_{tb}|$ Results





$|V_{tb}f_1^L| = 1.3 \pm 0.2$

$|V_{tb}| > 0.68$ at 95% CL
Assuming $f_1^L = 1$

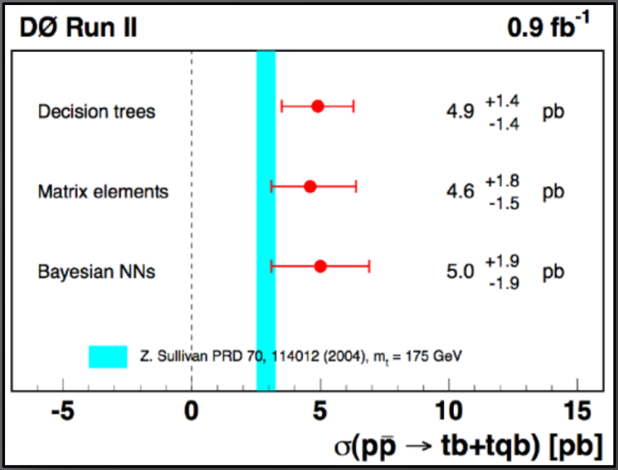
Conclusion

We see 3.4σ evidence for single top production!

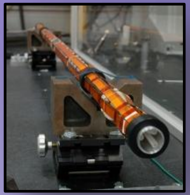
$|V_{tb} f_1^L| = 1.3 \pm 0.2$

$|V_{tb}| > 0.68$ at 95% CL
Assuming $f_1^L = 1$

We were very lucky!!



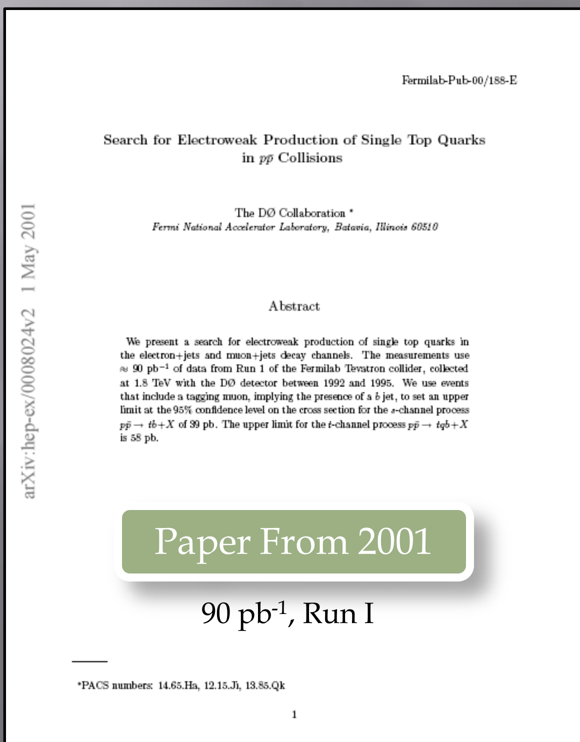
This is just a start!
- Correlation between analyses is not 100%: we are hard at work on the combination
- We have 1 fb^{-1} on "tape"
- New trigger installed
- New Layer 0 of silicon (20-30% improvement in b-tagging hopefully)
- Further analysis improvements



LHC Physics 2008
- Huge production rate
- W+Jets backgrounds are more manageable!
- V_{tb} to a few percent…
- tW production mode to explore

The Start of Single Top

A subgroup to search for single top quark production was formed almost the day of the 1995 top quark discovery announcement.



arXiv:hep-ex/0008024v2 1 May 2001

Fermilab-Pub-00/188-E

Search for Electroweak Production of Single Top Quarks in $p\bar{p}$ Collisions

The DØ Collaboration *
Fermi National Accelerator Laboratory, Batavia, Illinois 60510

Abstract

We present a search for electroweak production of single top quarks in the electron+jets and muon+jets decay channels. The measurements use $\approx$ 90 pb^{-1} of data from Run 1 of the Fermilab Tevatron collider, collected at 1.8 TeV with the DØ detector between 1992 and 1995. We use events that include a tagging muon, implying the presence of a b jet, to set an upper limit at the 95% confidence level on the cross section for the s-channel process $p\bar{p} \rightarrow tb+X$ of 39 pb. The upper limit for the t-channel process $p\bar{p} \rightarrow tqb+X$ is 58 pb.

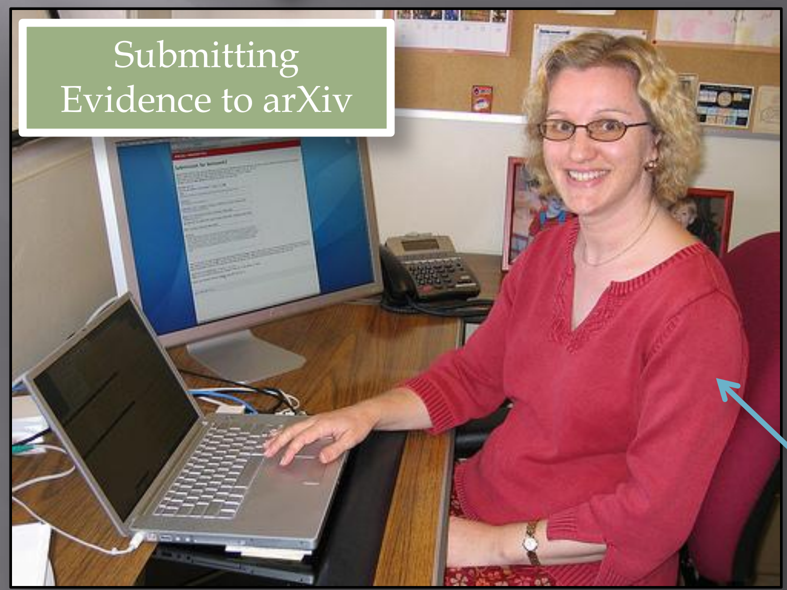
*PACS numbers: 14.65.Ha, 12.15.Ji, 13.85.Qk

1

Paper From 2001

90 pb^{-1}, Run I

Submitting Evidence to arXiv



"It was after I read a paper by CP Yuan…"
"Thought it would be easy…"

Aran Garcia-Bellido (UW)

Ann Heinson (UCR)

(conveners)