University of Rochester — Physics Colloquium

Rochester, December 13, 2007

Top quarks can do it alone

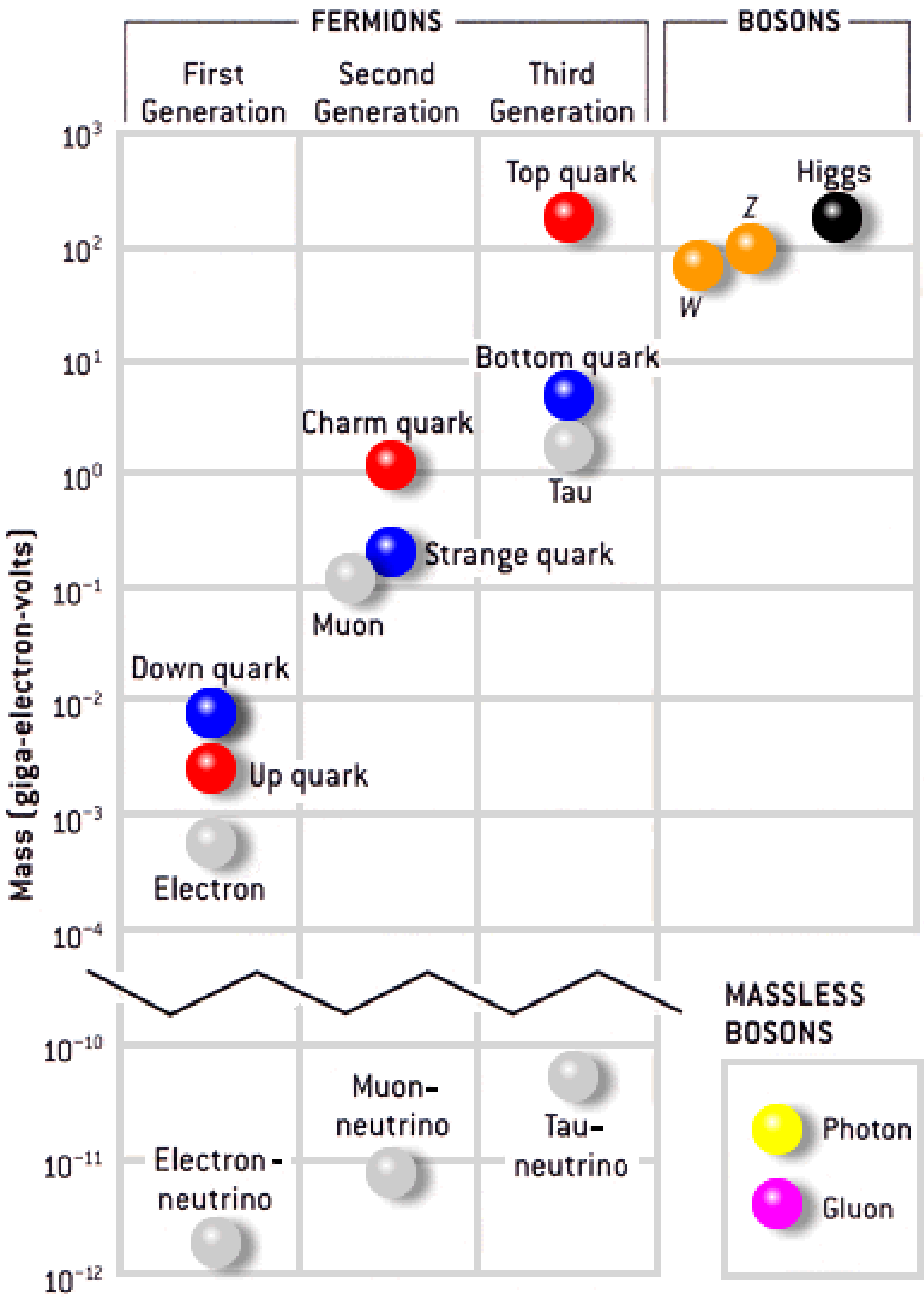
- The Standard Model and top quarks
- The Tevatron and DØ
- Searching for single top quarks
- Multivariate methods
  - Decision Trees
- Evidence
- Prospects at the LHC
- Summary

UNIVERSITY OF WASHINGTON

Arán García-Bellido

DØ

Top quark: not just the sixth quark



- Discovered in 1995 at CDF and DØ
- Heaviest known particle

  40 times heavier than b (~Au atom)
- Only quark that decays before hadronization

  t→Wb in ~10^{-25}s
- Couples strongly to Higgs boson

  Related to the origin of mass?
- Unique laboratory to study the SM and beyond

The SM under attack

- The Standard Model is a fantastic success:

  Predictions confirmed by discoveries (c, b, t, W, Z) and precise measurements

- But recently: Neutrino masses, dark matter
- So we know it is not a complete description of Nature
- Many unanswered questions:
  - Why three generations?
  - What is the mechanism responsible for particles' masses?
  - Why that hierarchy of masses?
  - What’s with so many free parameters?
  - Gravity is not in the picture
  - Unification of three couplings is not possible

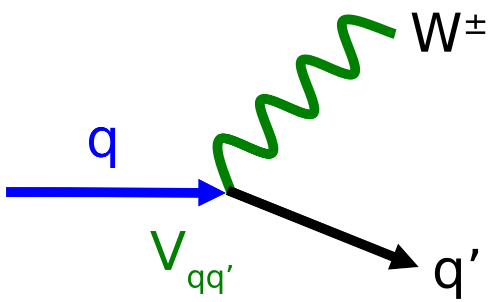
It's all dubbya's fault

- Studying the electroweak sector is crucial to test the SM... and understand the asymmetry of matter and antimatter in the Universe

- Weak interactions treat matter and antimatter differently ...only possible because there are three families!

- Weak interaction and mass eigenstates aren't the same ➔ **Mixing** (Cabibbo-Kowayashi-Maskawa matrix)

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



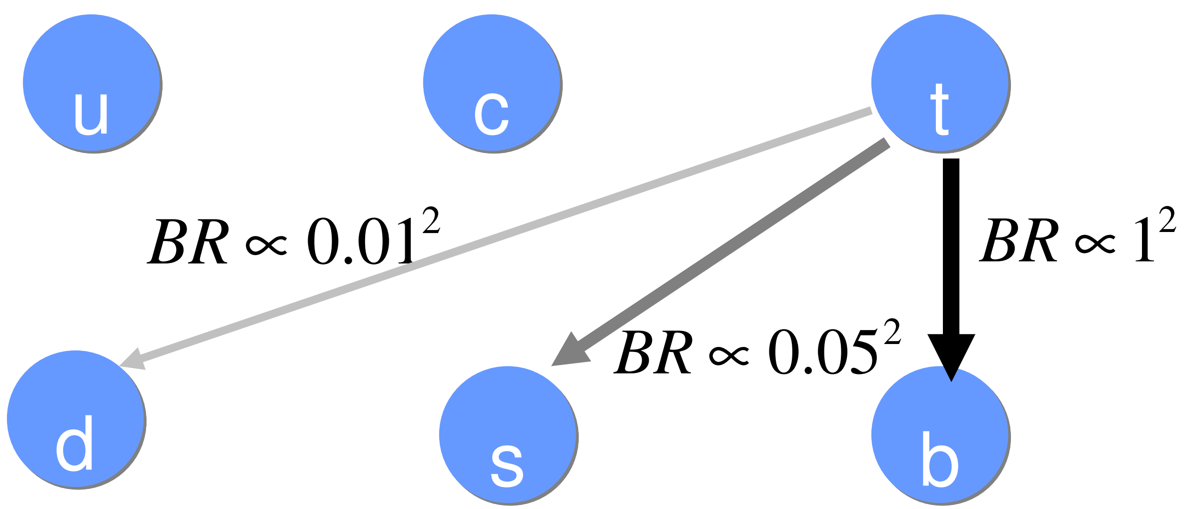
Only element not measured directly yet

- The CKM matrix is being scrutinized from many different angles: B-factories, Tevatron, nuclear experiments...

Flavor changing interactions

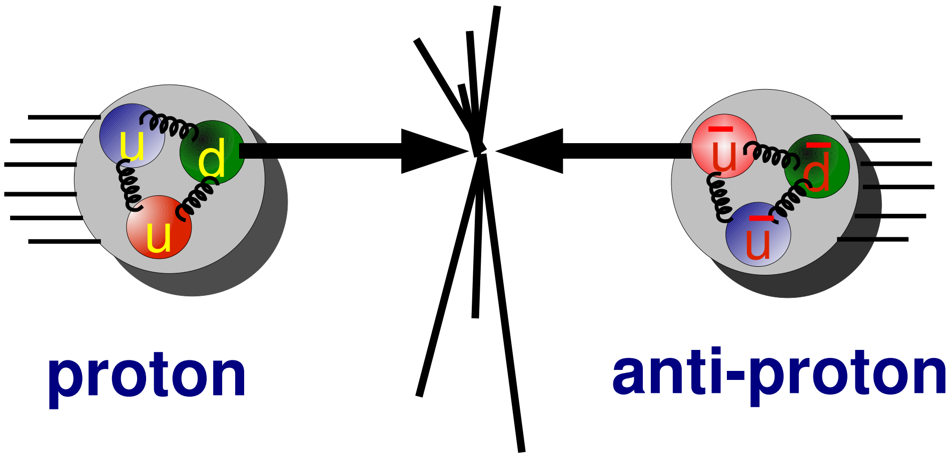
$$\begin{pmatrix} 0.97383 & 0.2272 & 0.00396 \\ 0.2271 & 0.97296 & 0.04221 \\ 0.00814 & 0.04161 & 0.999100 \end{pmatrix}$$

- Observe hierarchy in flavor-changing transitions
- Probability of transition (branching ratio) within one family is the largest
- Transitions between families are suppressed:



Tools of the trade

- Particle physicists use high energy colliders to probe physics at small distances



Note on units: N[collisions]=σ[pb]L[pb^{-1}]

- **Picobarns** (pb) are a measure of "cross section" (σ=interaction probability). 1 barn = 10^{-24} cm^2.

- **Inverse picobarns** (pb^{-1}) are a measure of the "integrated luminosity" (L=collected data)

Example: 100 pb^{-1} = sufficient data to observe 100 events of a process having 1 pb cross section

- **GeV** are used interchangeably for mass, energy and momentum

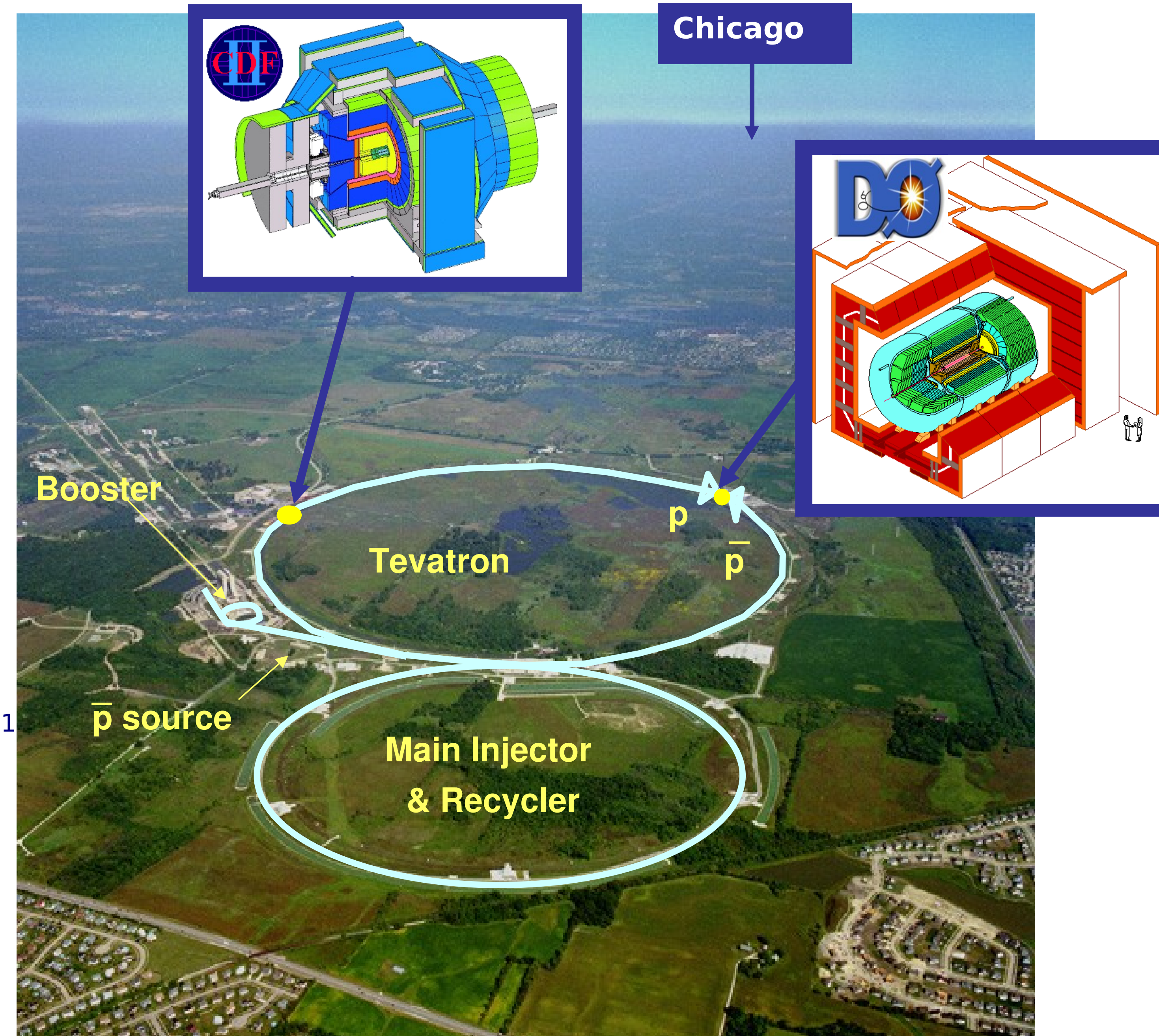
The Tevatron

The highest energy particle accelerator in the world!

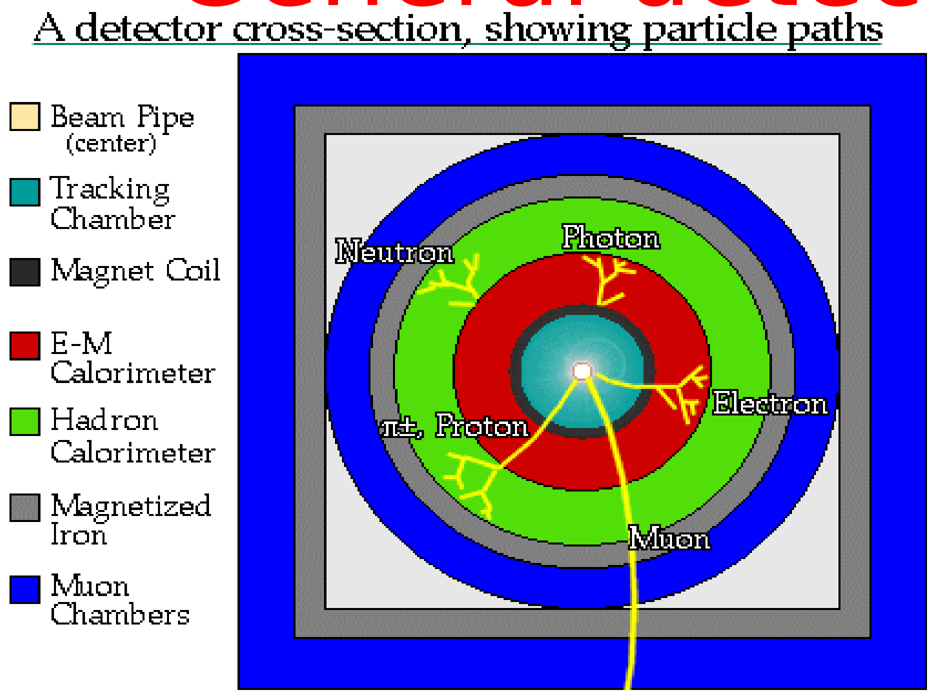
Proton-antiproton collider 1km radius

Run I 1992-1995
Top quark discovered!

Run II 2001-09(?)
$\sqrt{s}$ = 1.96 TeV
Δt = 396ns
>3fb^{-1} delivered
Peak Lum: $3 \cdot 10^{32} cm^{-2}s^{-1}$

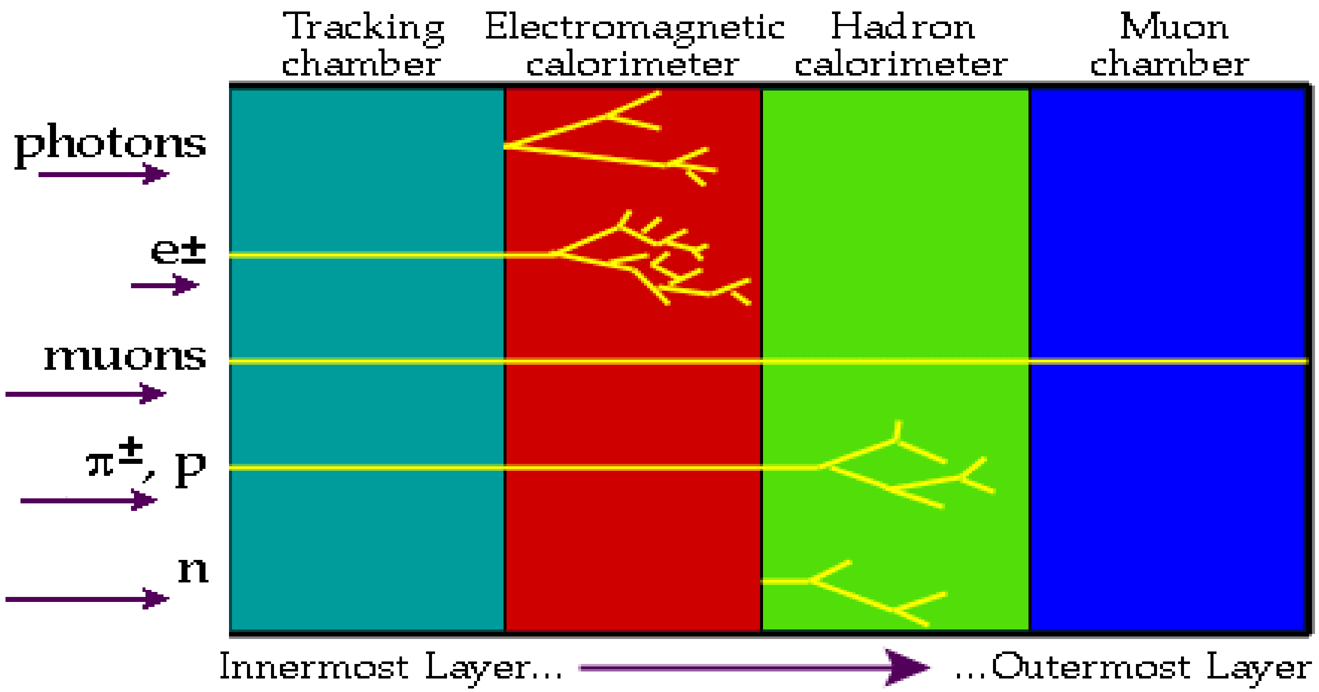


General detector and particle ID

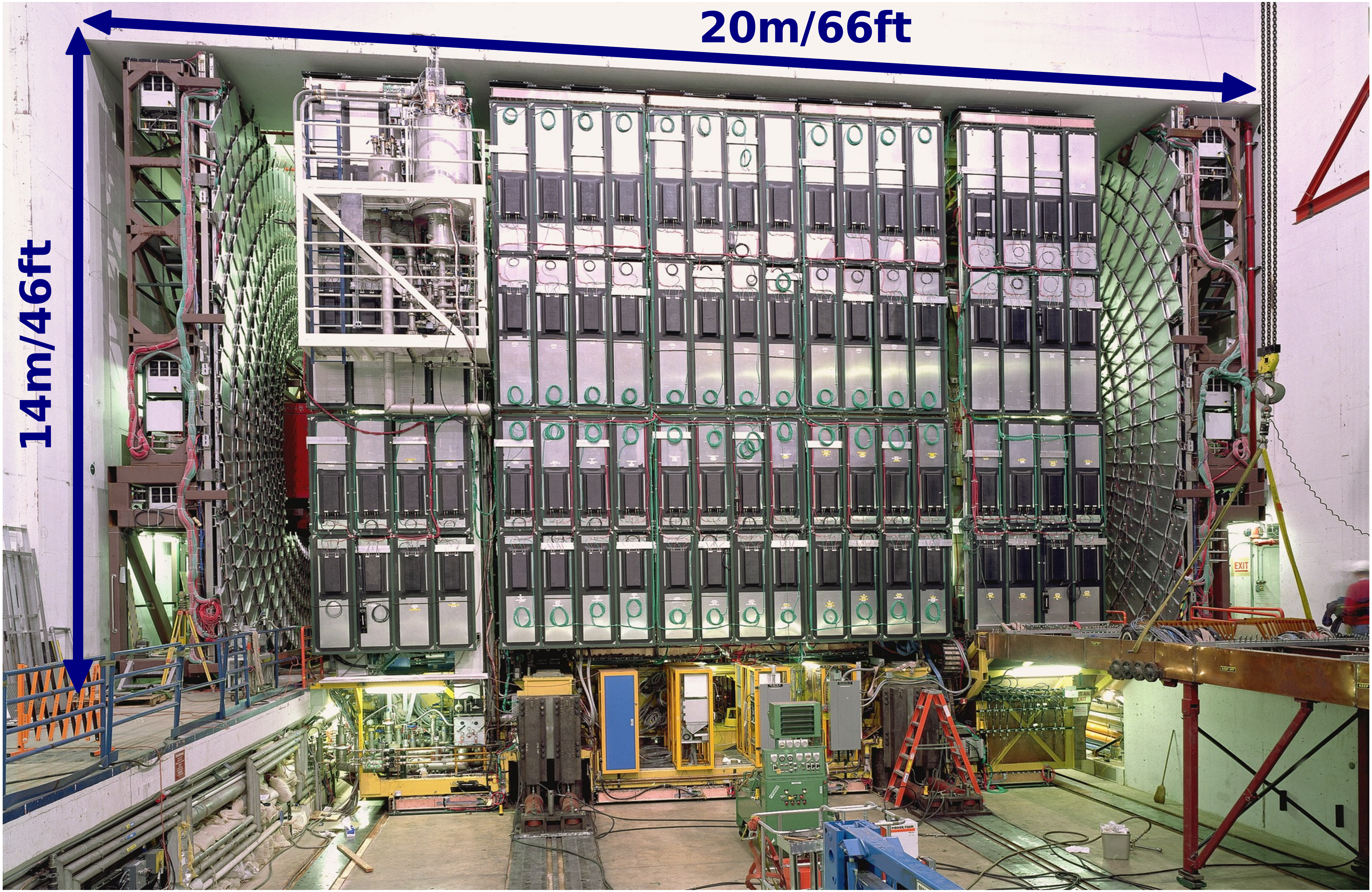


We detect particles by the EM and strong interaction fingerprints they leave behind

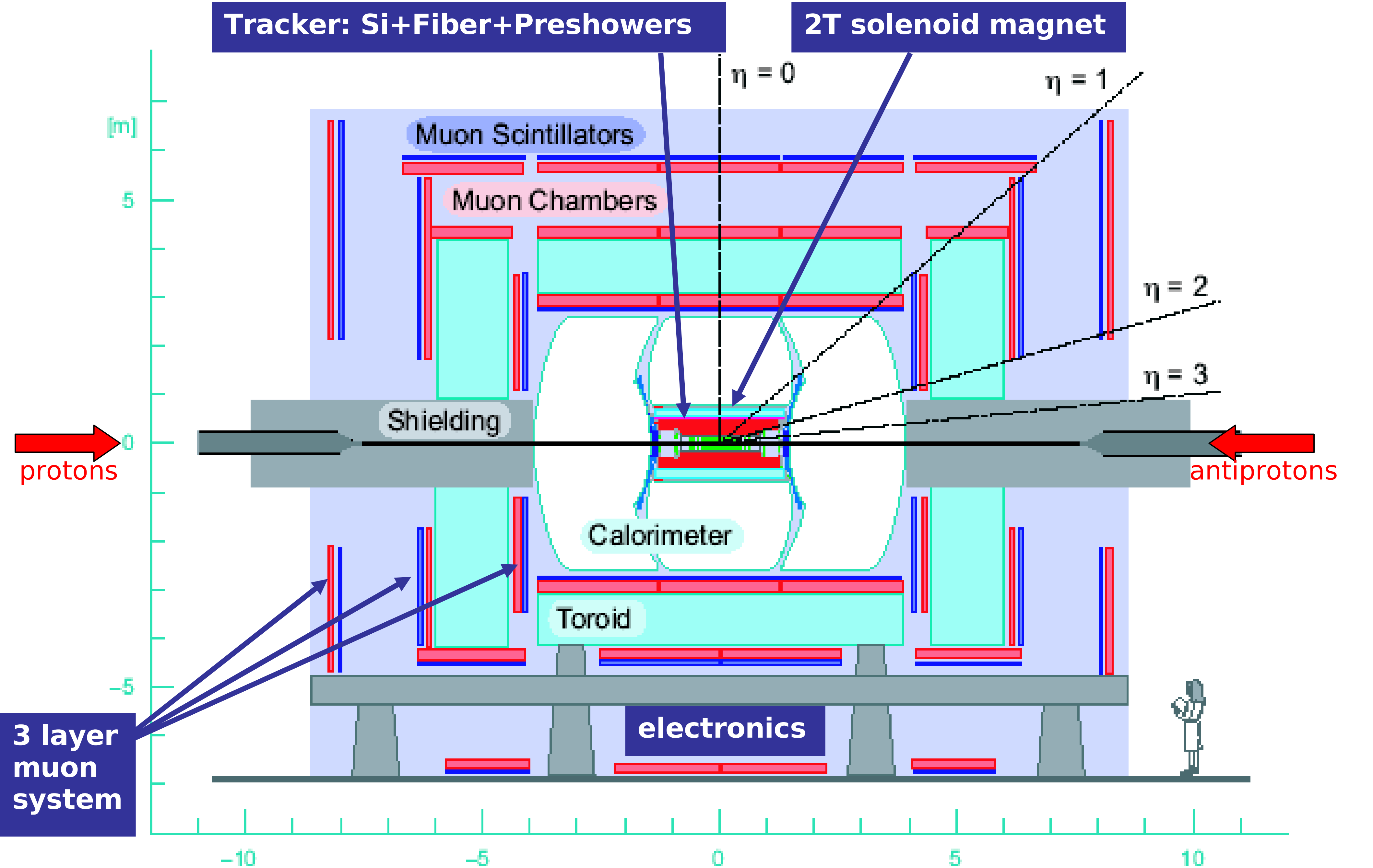
- Tracking is first (measure p_T)
- Calorimetry (EM and hadronic)
- Muons
- All the rest is neutrinos



The real thing: the DØ detector

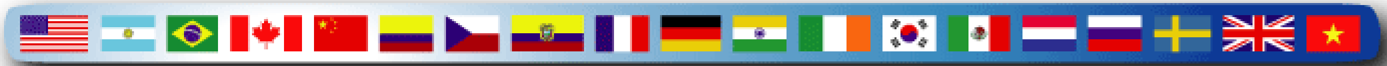


DØ for Run II


Tracker: Si+Fiber+Preshowers
2T solenoid magnet
η = 0
η = 1
η = 2
η = 3
[m]
Muon Scintillators
Muon Chambers
Shielding
Calorimeter
Toroid
electronics
protons
antiprotons
3 layer muon system
5
0
–5
–10
–5
0
5
10

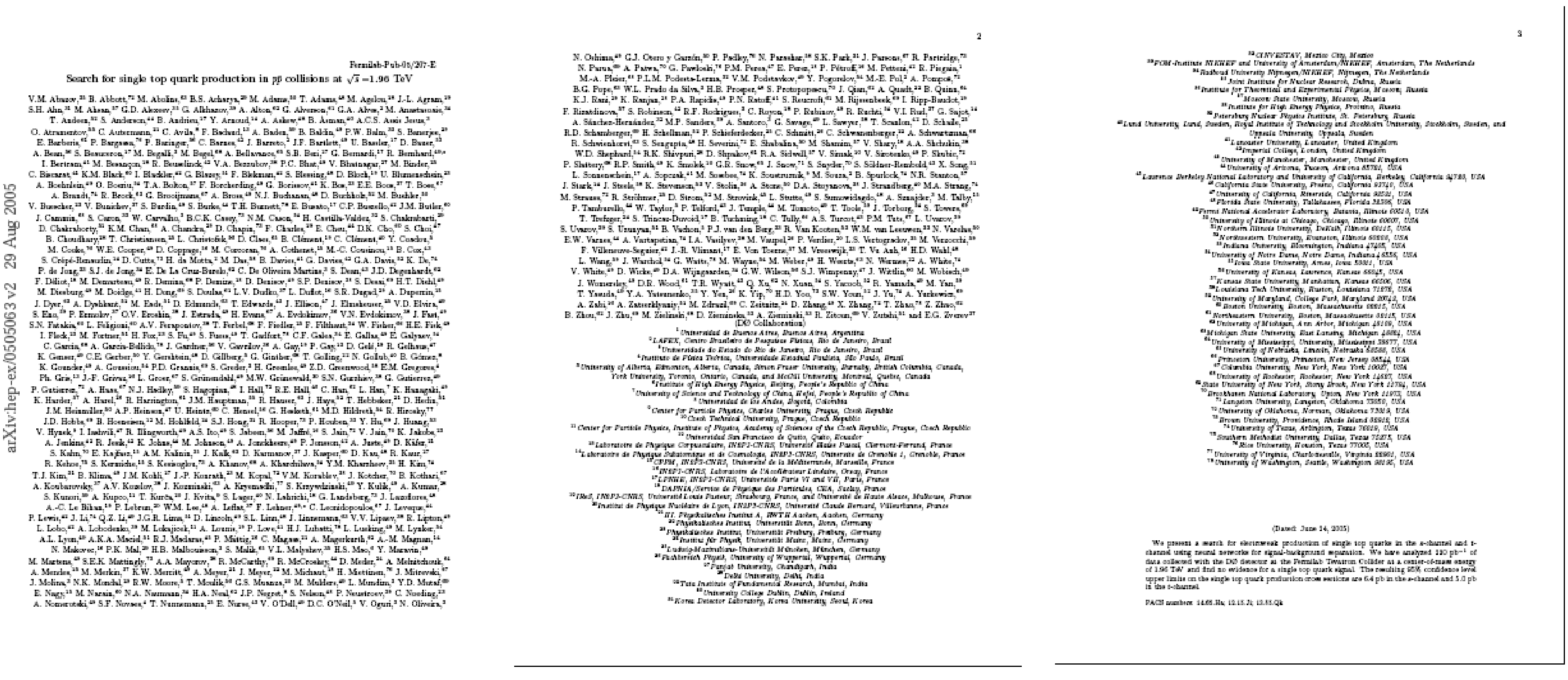
Many, many people running it

19 countries, 80 institutions, 670 physicists



A lot of convincing to do…

Since we are all signing the papers together you have to convince them all that what you are doing is sensible and deserves to be published!


arXiv:hep-ex/0505063 v2 29 Aug 2005

Fermilab-Pub-05/207-E

Search for single top quark production in $p\bar{p}$ collisions at $\sqrt{s}$ =1.96 TeV

(DØ Collaboration)

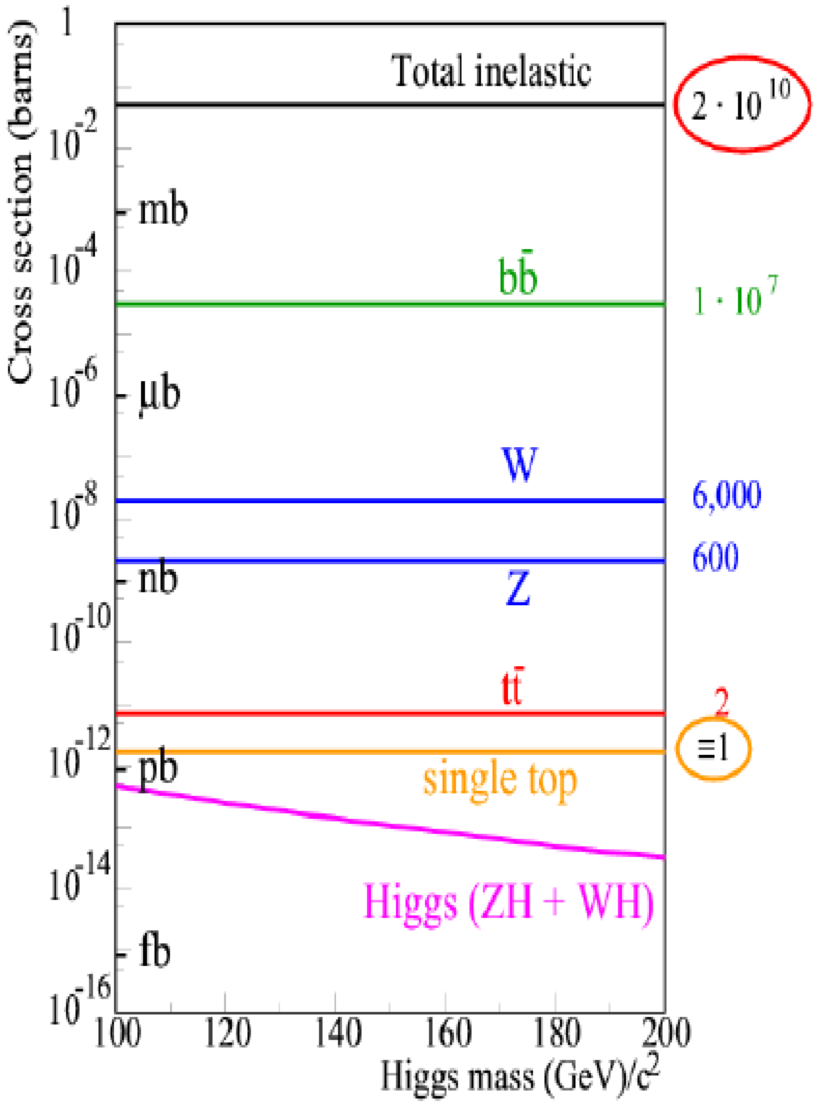
(Dated: June 14, 2005)

We present a search for electroweak production of single top quarks in the s-channel and t-channel using neural networks for signal-background separation. We have analyzed 230 pb^{-1} of data collected with the DØ detector at the Fermilab Tevatron Collider at a center-of-mass energy of 1.96 TeV and find no evidence for a single top quark signal. The resulting 95% confidence level upper limits on the single top quark production cross sections are 6.4 pb in the s-channel and 5.0 pb in the t-channel.

PACS numbers: 14.65.Ha; 12.15.Ji; 13.85.Qk

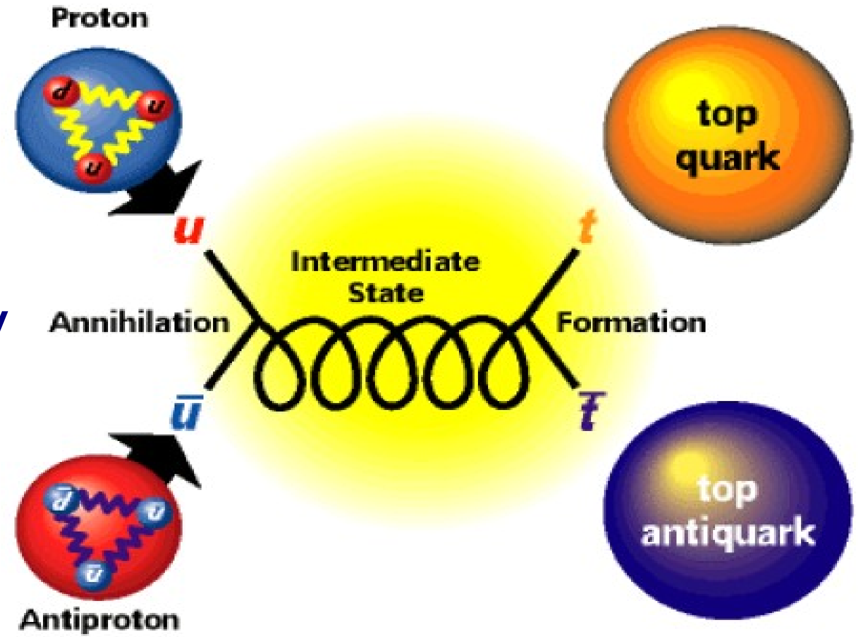
Physics at a hadron collider is like… drinking from a fire hose

- Collision rate is huge
  Every 396 ns → ~1.7 MHz (live crossings)
- Total cross section ~0.1b
  2-3 interactions per collision at L=10^{32}
- But W, Z, t, H are rare!
  Around 20 single top events per day
- Need trigger system to select interesting events
  Only store manageable size ~25MB/s

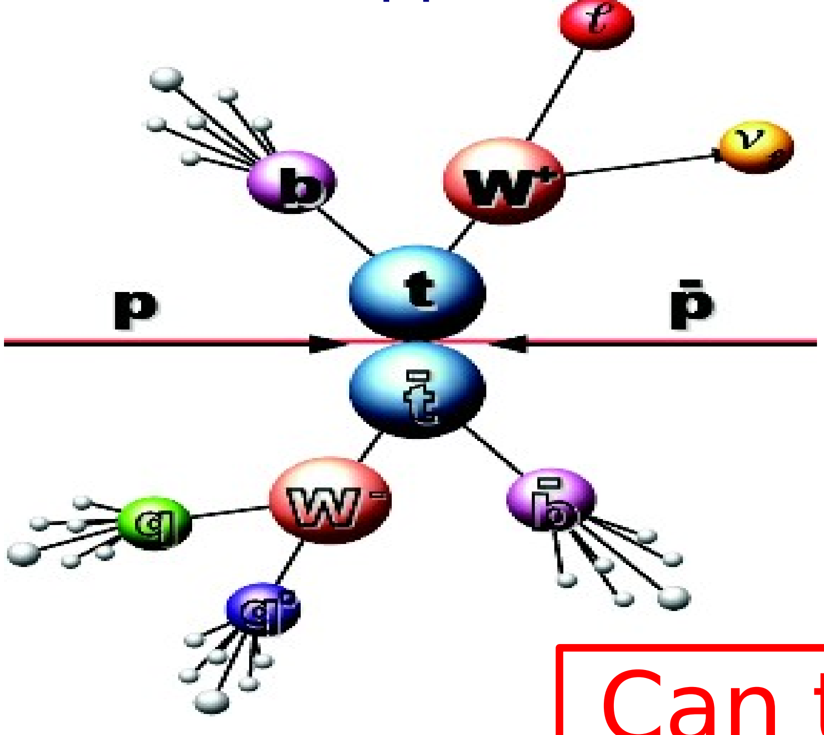


Close encounters of the 3rd generation

- Top quarks have only been seen so far produced in pairs of top and anti-top



- Then each top quark decays quickly into a W boson and a b-quark
- The W can then decay into $\ell\nu$ (30%) or qq' (70%)



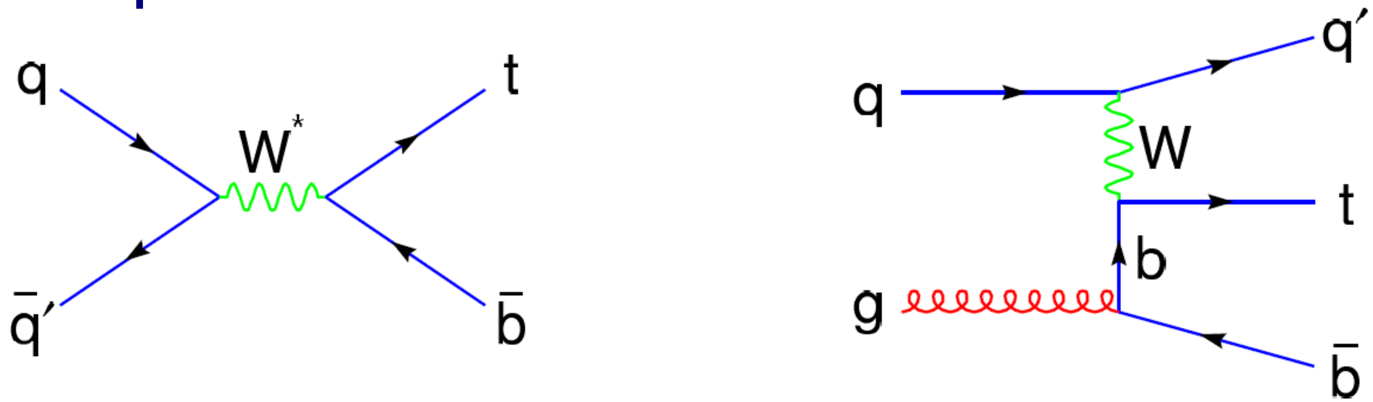
- We have measured the pair production cross section: $\sigma = 8.2^{+0.9}_{-0.8}\,pb \quad (D0 \;\; L = 0.9\text{fb}^{-1})$
- And its mass: $m_t = 170.9 \pm 1.8\text{GeV} \quad (CDF + D0)$
- And some of its properties...

Can they be produced alone?

Yes! Top quarks can be lonely!

Electroweak production of single top quarks

Two main production modes at the Tevatron:



s-channel σ_s~1pb

t-channel σ_t~2pb

Why do we care?

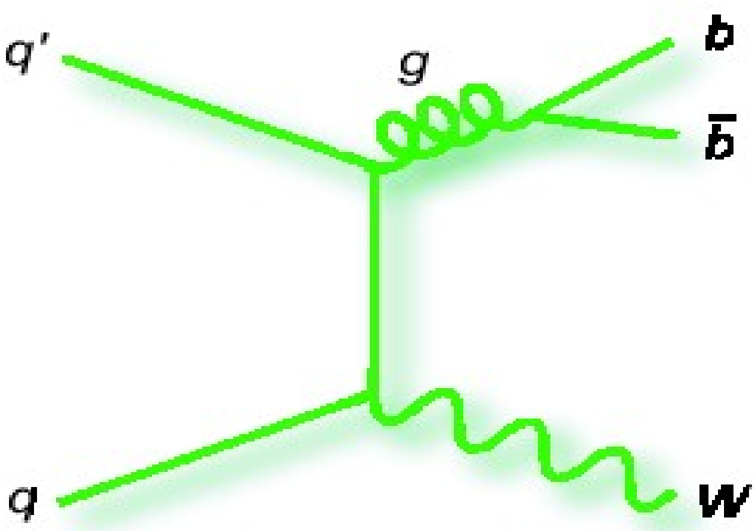
- Had not been seen before!
- Challenging signature!
- Probe V_{tb} at production
- Sensitive to new physics
- Necessary step towards Higgs discovery

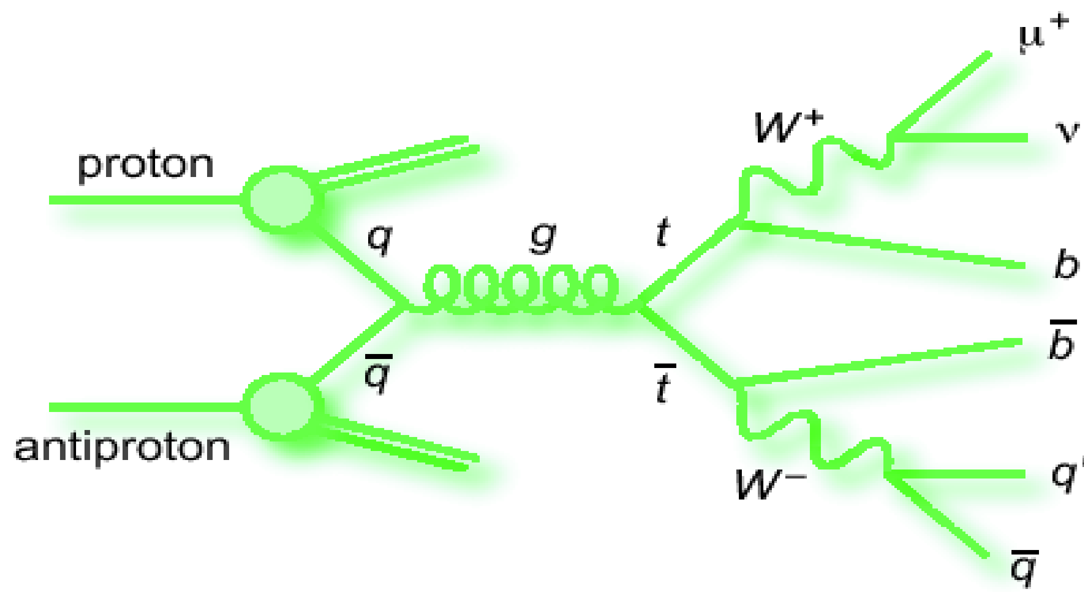
Single top vs top pairs events

- Have less total energy
- Are less spherical
- Are produced less often
- Only have two jets, live in a higher noise environment

How do we find single tops?

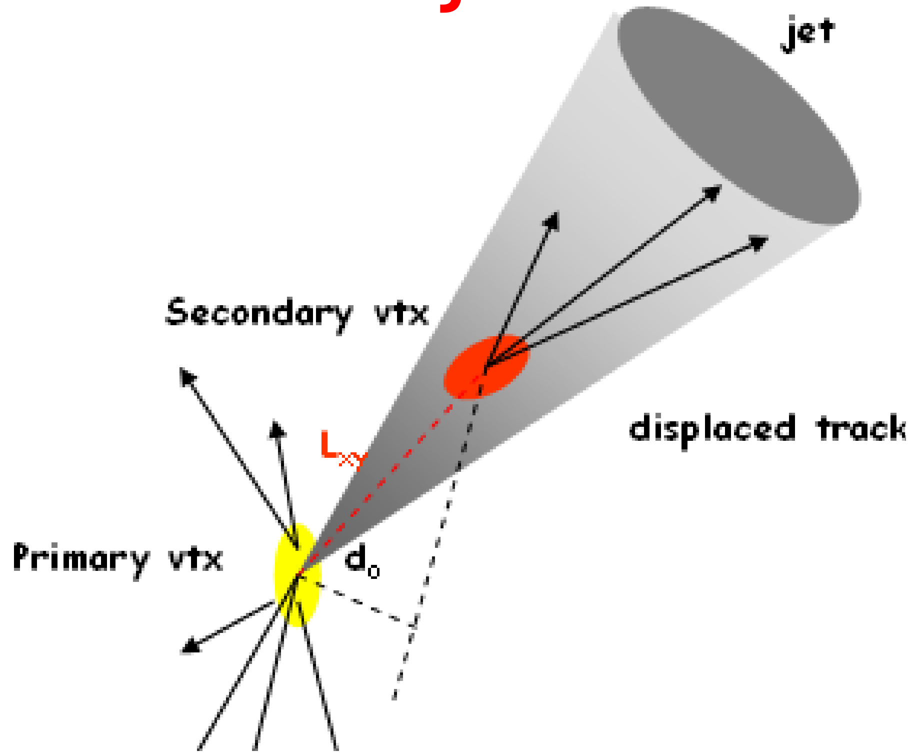
- It's not easy!
- Out of ~1 billion recorded events
  we are looking for ~100 signal events
- And there are many other processes that mimic single top events: W+jets, tt, multijets
- Our final state consists of 2, 3, or 4 jets (with at least one of them b) + lepton + neutrino (missing E_T)

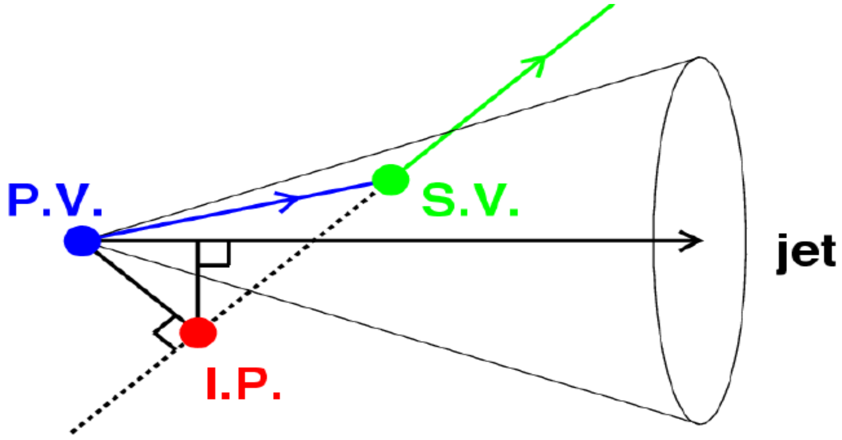


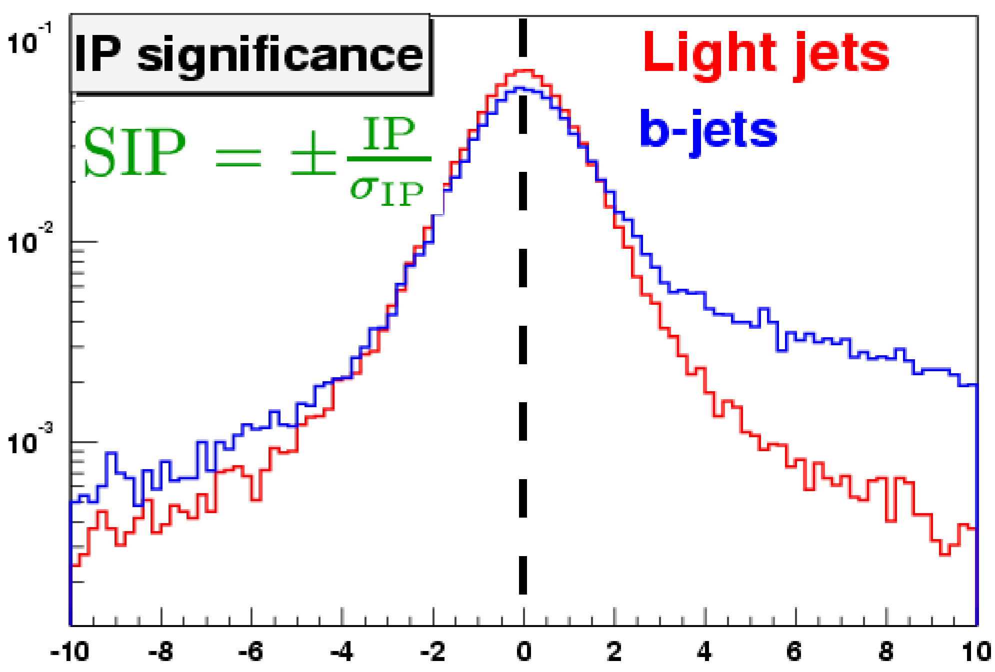


Did you see that bottom jet?

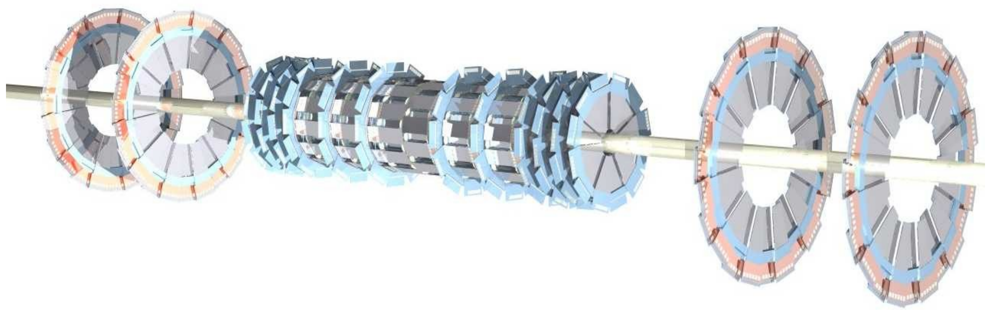
- Top quarks decay into b quarks
  → can we tell the difference between a b jet and any other jet originated from u, d, s or a gluon?
- b-quarks have a lifetime ~10^{-12}s
  → they travel ~500μm before decaying
- Look for tracks coming from a common vertex displaced from the original pp collision
- These tracks have a positive signed impact parameter with respect to the collision point



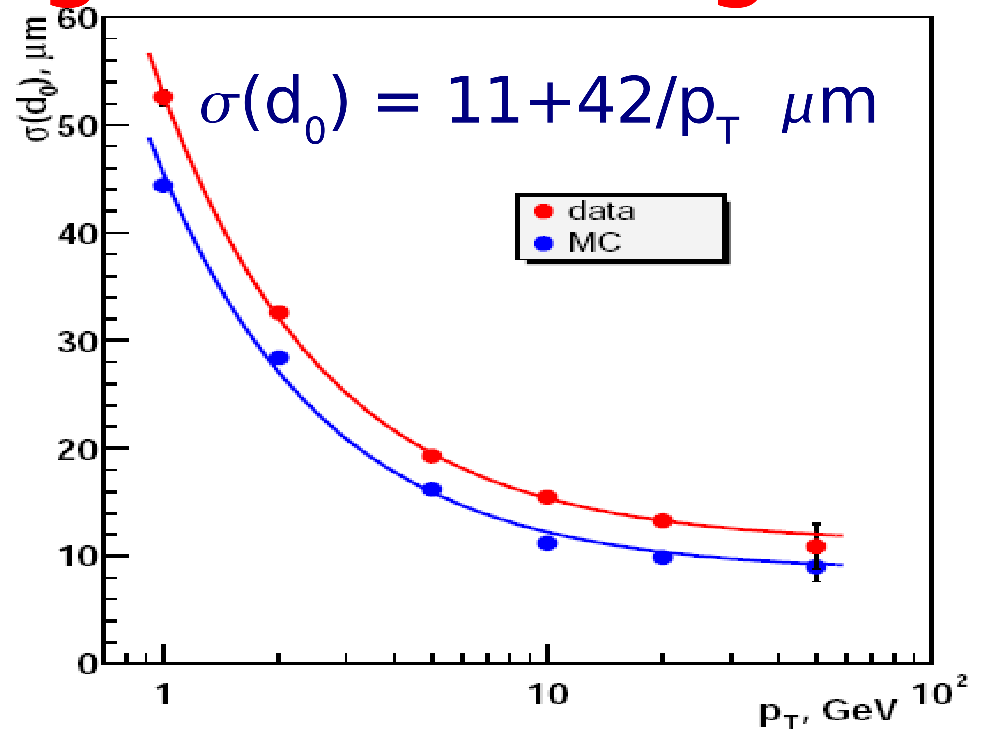


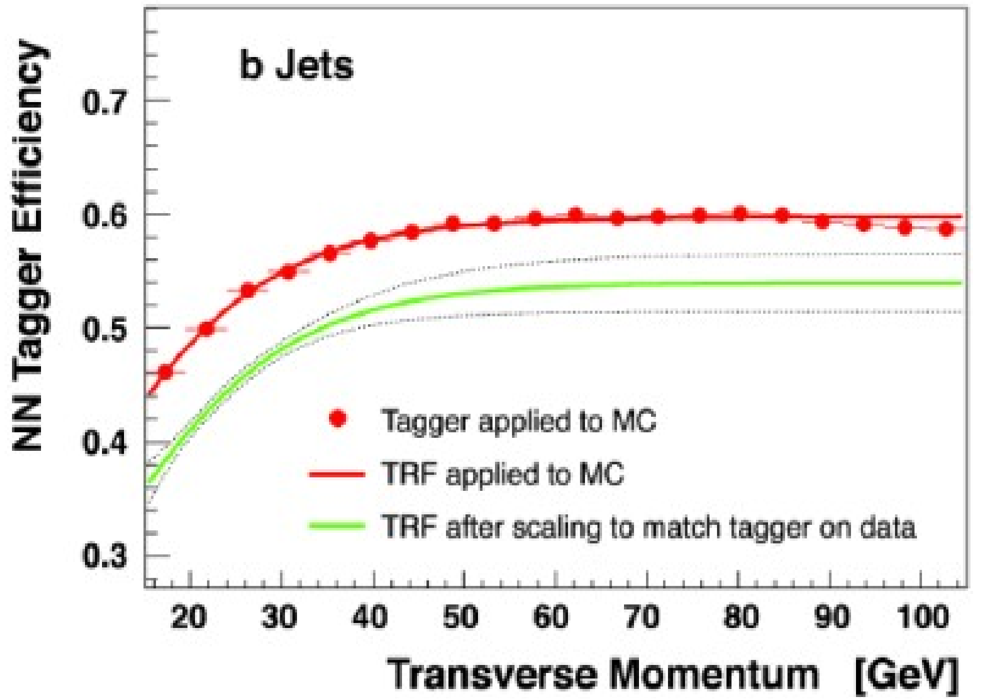


You better have good tracking



- The Silicon Microstrip Tracker allows resolutions of ~10 μm
- Inner radius: 1.7cm away from the interaction point
- Combine displaced tracks properties into a Neural Net
- Efficiency to identify a b-quark jet ~50%
- Mistag-rate ~0.5%





Analysis strategy

1) Event selection

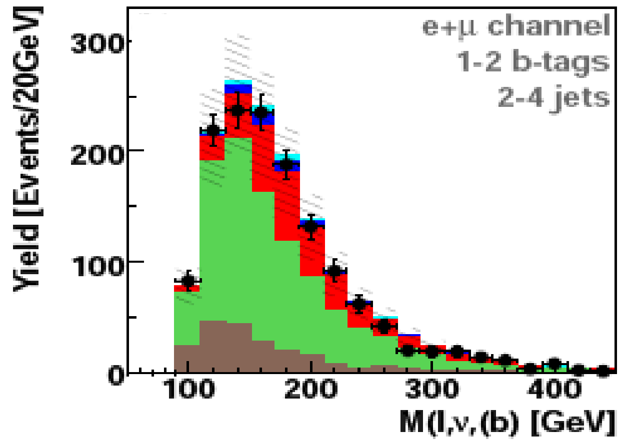
- Select W+jets like events
- Maximize acceptance
- Model backgrounds well

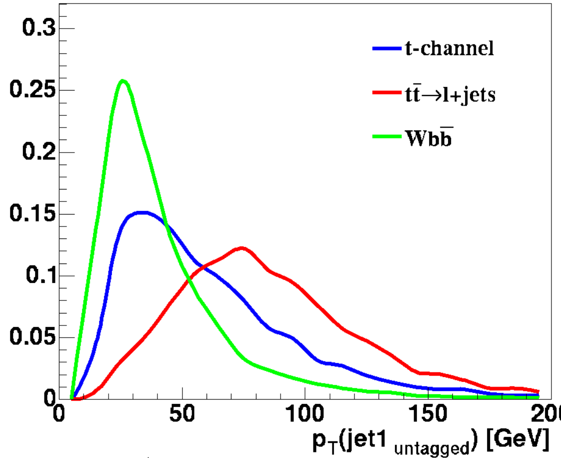
2) Separate signals from backgrounds

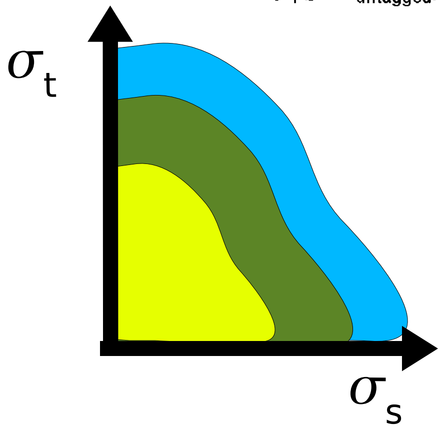
- Find discriminating variables
- Multivariate analysis

3) Measure the cross section

- Use shape information
- Bayesian statistical analysis
- Make sure this is not a fluctuation!



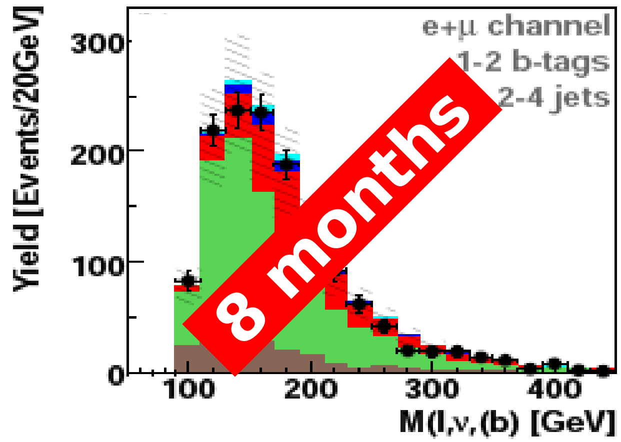




Analysis strategy

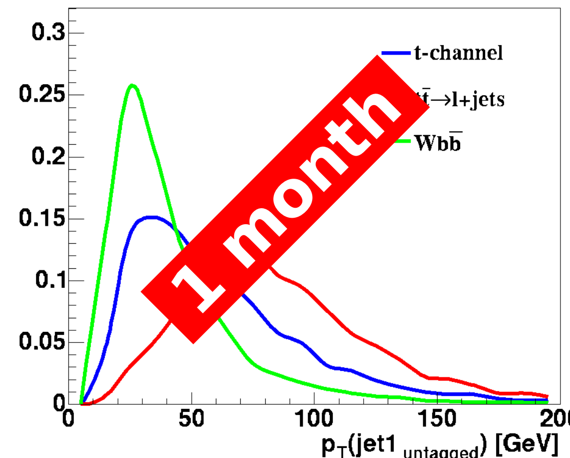
1) Event selection

- Select W+jets like events
- Maximize acceptance
- Model backgrounds well



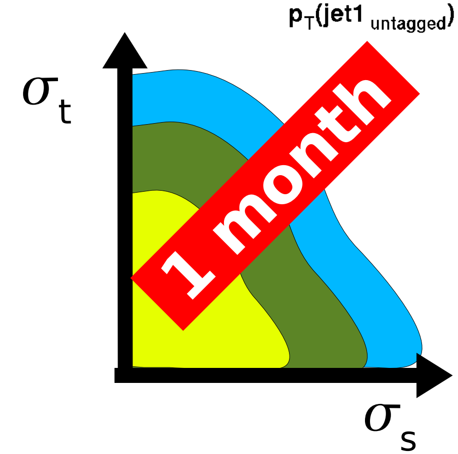
2) Separate signals from backgrounds

- Find discriminating variables
- Multivariate analysis

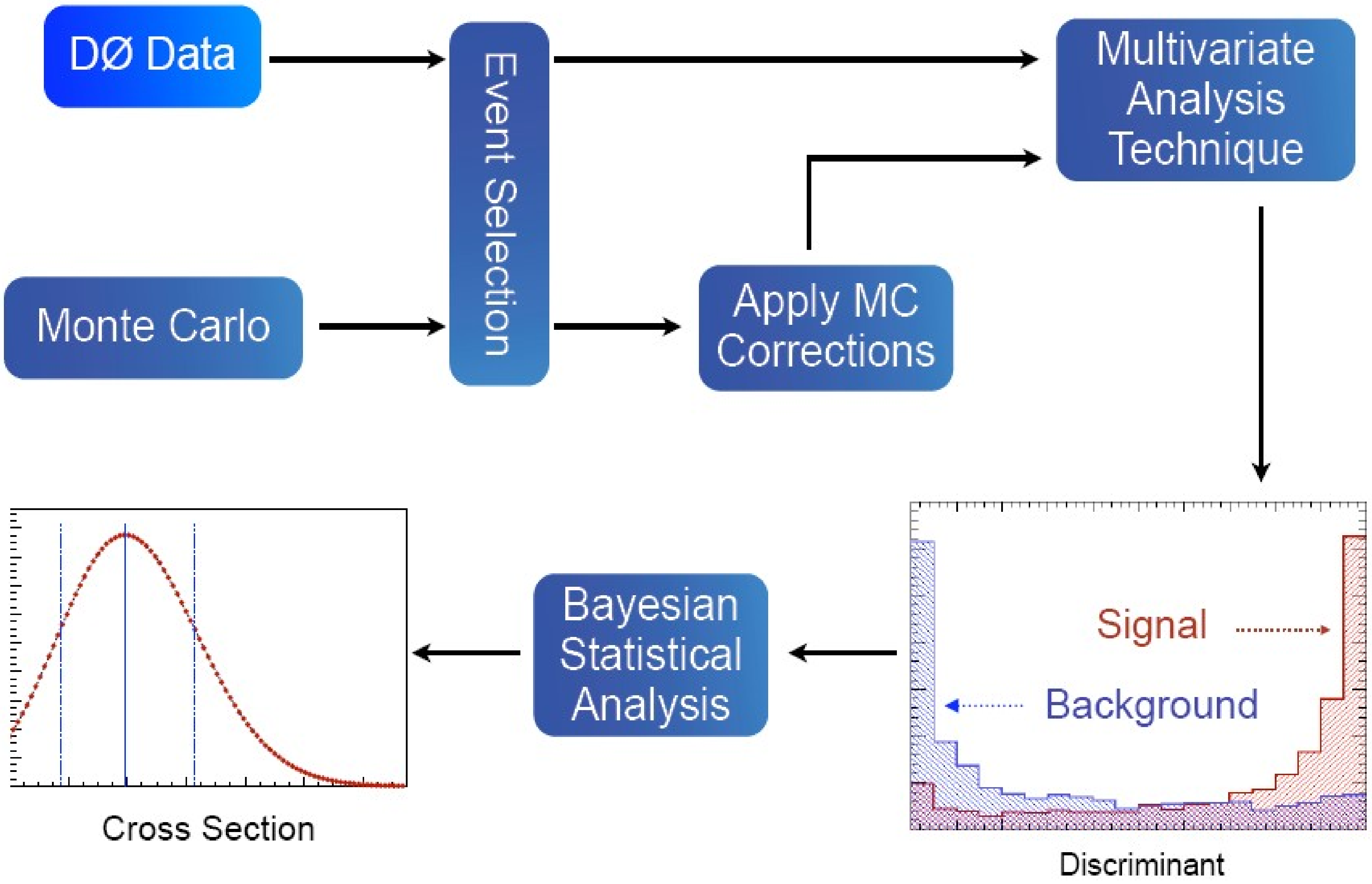


3) Measure the cross section

- Use shape information
- Bayesian statistical analysis
- Make sure this is not a fluctuation!



Analysis flow



1) Event Selection

- 2 ≤ Njets ≤ 4, p_T>25,20,15 GeV
- 1 lepton p_T>15 GeV
- MET>15 GeV

Source	Event Yields in 0.9 fb^{-1} Data Electron+muon, 1tag+2tags combined 2 jets	3 jets	4 jets
tb	16 ± 3	8 ± 2	2 ± 1
tqb	20 ± 4	12 ± 3	4 ± 1
$t\bar{t} \rightarrow ll$	39 ± 9	32 ± 7	11 ± 3
$t\bar{t} \rightarrow l$+jets	20 ± 5	103 ± 25	143 ± 33
$W+b\bar{b}$	261 ± 55	120 ± 24	35 ± 7
$W+c\bar{c}$	151 ± 31	85 ± 17	23 ± 5
$W+jj$	119 ± 25	43 ± 9	12 ± 2
Multijets	95 ± 19	77 ± 15	29 ± 6
Total background	686 ± 41	460 ± 39	253 ± 38
Data	697	455	246

Before b-tagging

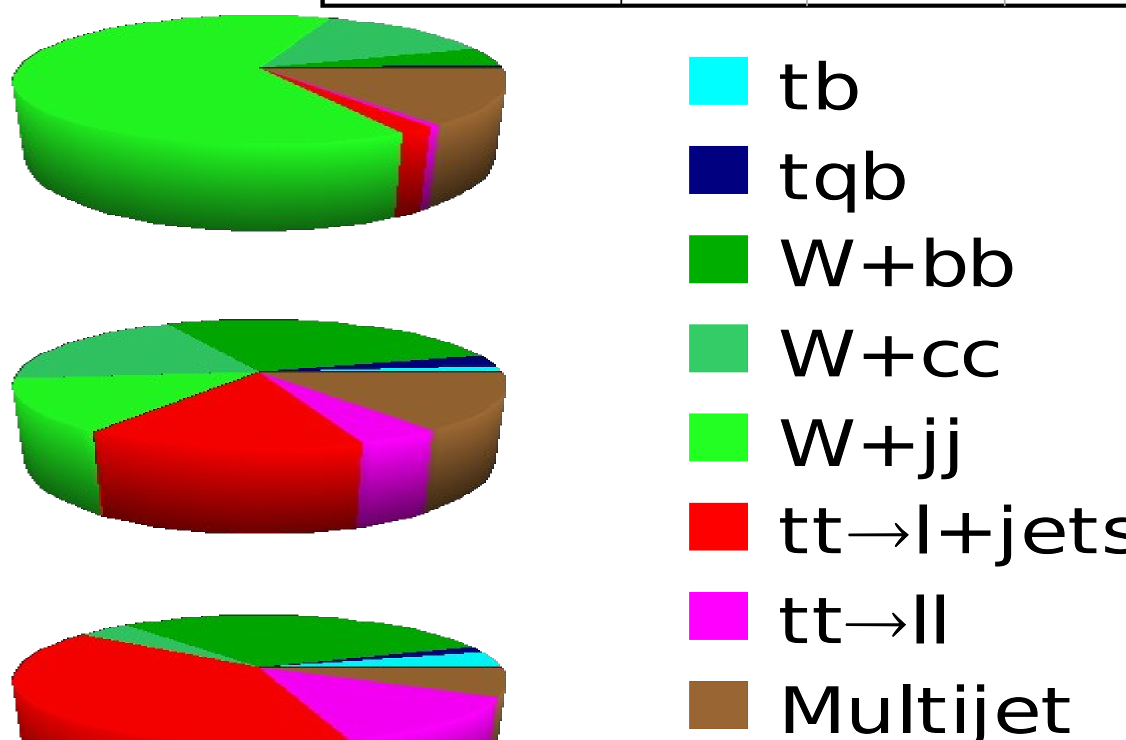
21,918 events (121 signal)

=1 b-tagged jet

1,227 events (53 signal)

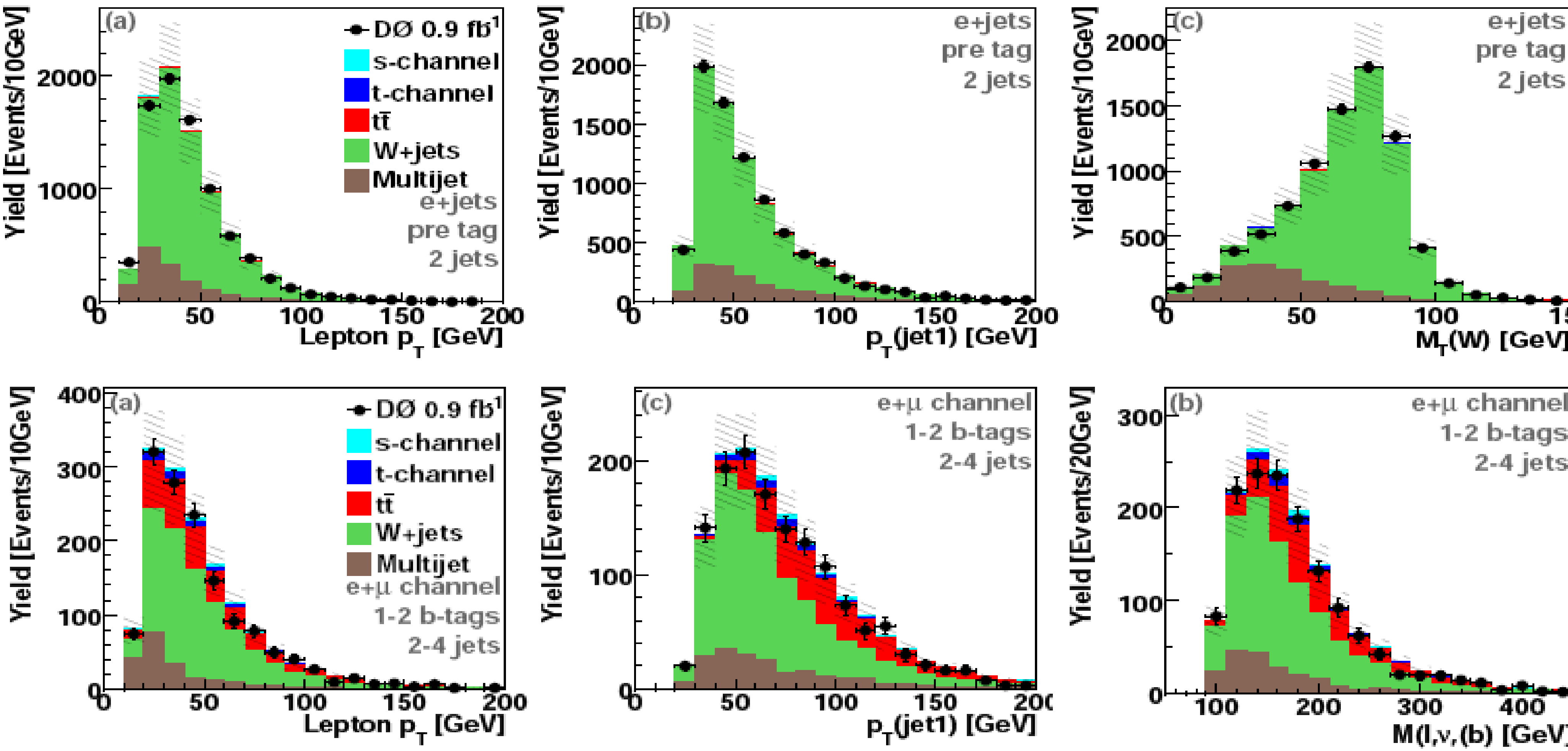
≥2 b-tagged jets

171 events (9 signal)



Data-Background agreement

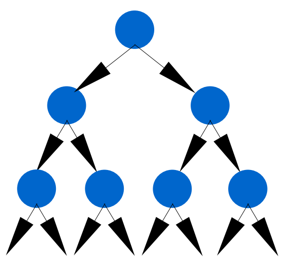
- The most challenging part of this analysis is to get an appropriate model for the backgrounds
- Kinematics are obtained from simulation
- We have used the data to normalize the main backgrounds
  - Before b-tagging: get multijet & W+jets composition
  - After b-tagging: estimate how much Wbb+Wcc



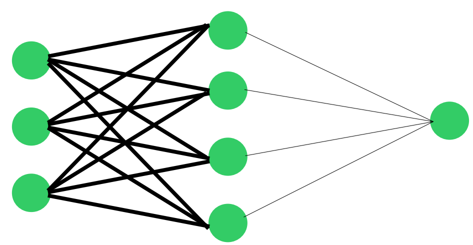
2) Separate signals from backgrounds

- Once we understand our data, need to measure the signal
- We cannot use simple cuts to extract the signal: use **multivariate techniques**
- DØ has implemented three analysis methods to extract the signal from the **same dataset**:

Decision Trees



Bayesian NNs



Matrix Elements

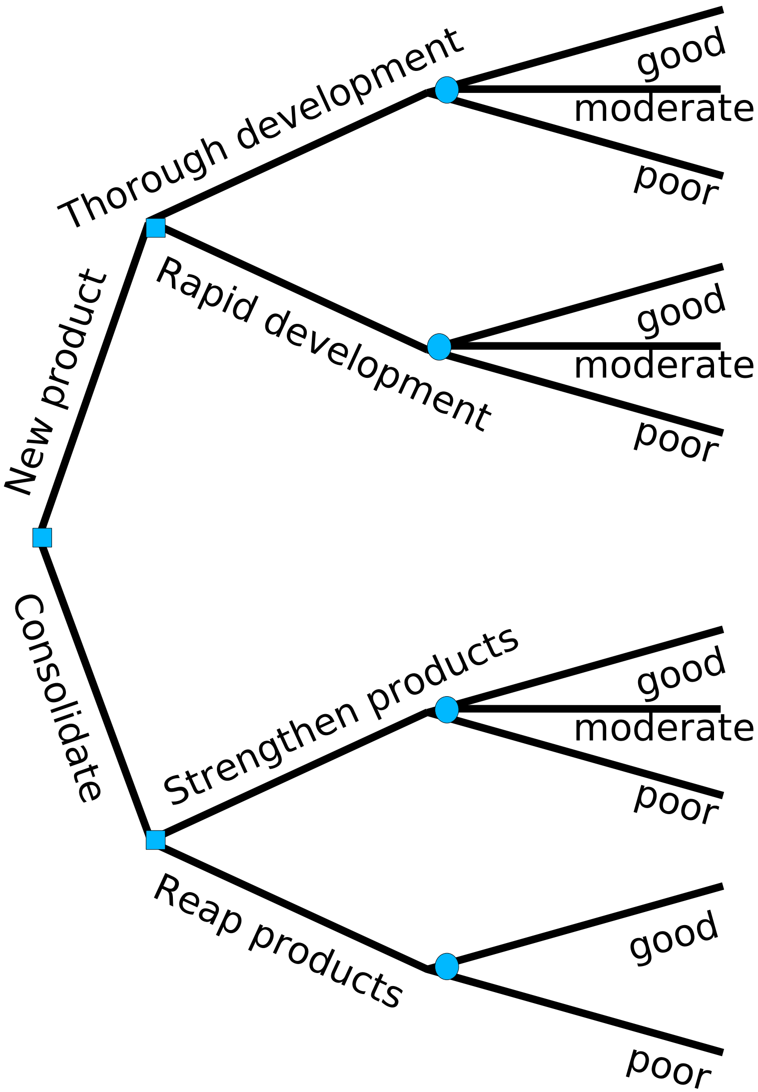
$$\int M$$

- **DT**: Simple cuts to obtain continuous distribution based on purity
- **BNN**: Average many different Neural Networks to be more efficient
- **ME**: Uses 4-vectors of reconstructed objects and full kinematic info
- Use same pool of discriminating variables for DT and BNN
- Optimized separately for s-channel, t-channel and s+t

Decision Trees Introduction

Machine learning technique widely used in social sciences

In finance:
Should we consolidate or not?



Decision Trees Introduction

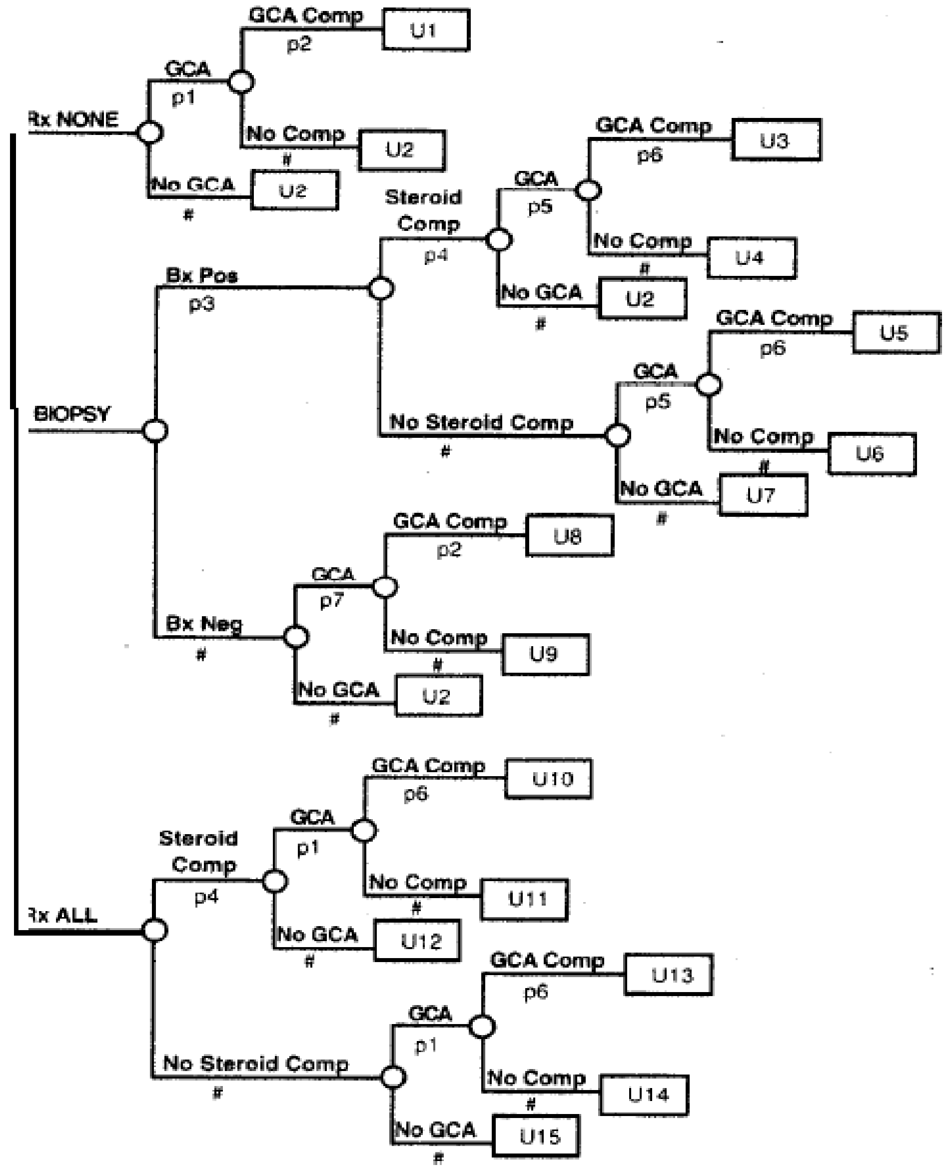
Machine learning technique widely used in social sciences

GCA = giant cell arteritis
GCA comp = giant cell arteritis complication (i.e., blindness)
Steroid comp = steroid complication

p1 = probability patient has GCA
p2 = probability of GCA complication without steroids (pCGA comp) given that the patient has giant cell arteritls
P3 = probability of positive temporal artery biopsy = sens X p1 + (1 − spec) X (1 − p1)
sens = sensitivity of biopsy
spec = specificity of biopsy
P4 = probability of steroid comp
P5 = probability of GCA in patients wlth positive temporal artery biopsy = (sens X p1) ÷ p3
p6 = probability of GCA comp given that the patient has GCA and is treated with steroids = p2 X (1 − e)
P7 = probability of GCA for patients with negative biopsy = [(1 − sens) X pll ÷ (1 − p3)
U1 to u15 = utilities for the individual states

Bx Pos = biopsy positve for GCA
Bx Neg = biopsy negative for GCA

In medicine: strategies for patients with suspected GCA

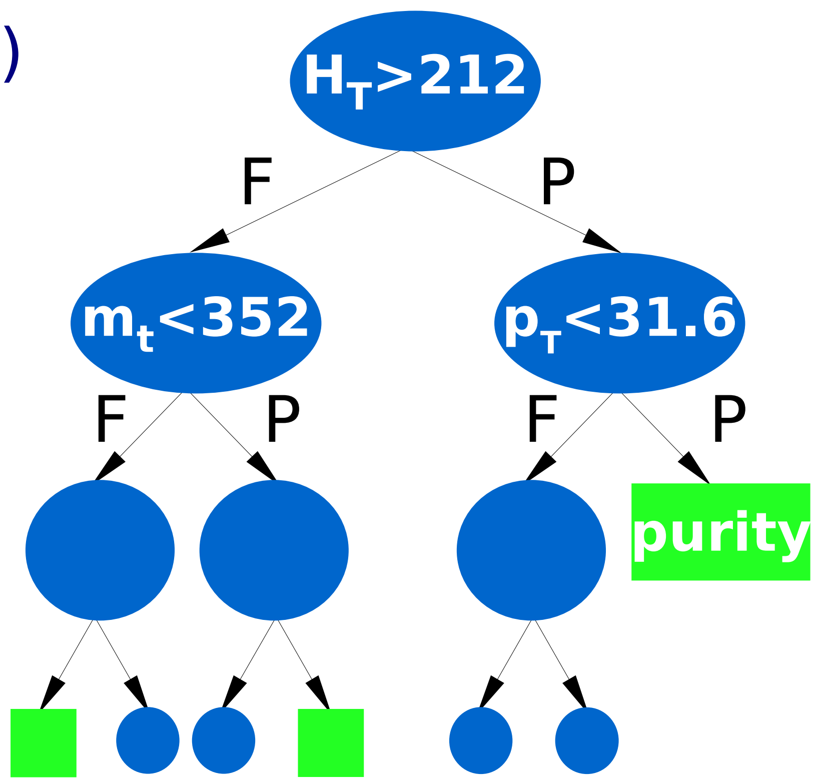


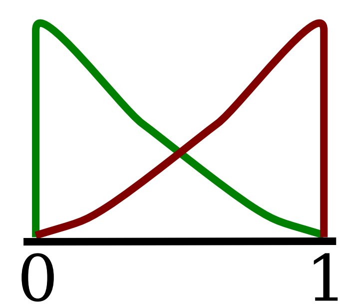
Decision Trees

DØ has been the first to apply DTs to a search in HEP

Idea: recover events that fail criteria in cut-based analysis

- Start with all events (first node)
- For each variable, find the splitting value with best separation between children
- Select best variable and cut: produce **P**ass and **F**ailed branches
- Repeat recursively on each node
- Stop when improvement stops or when too few events left
- Terminal node: leaf with purity = $N_S/(N_S+N_B)$
- Output: purity for each event



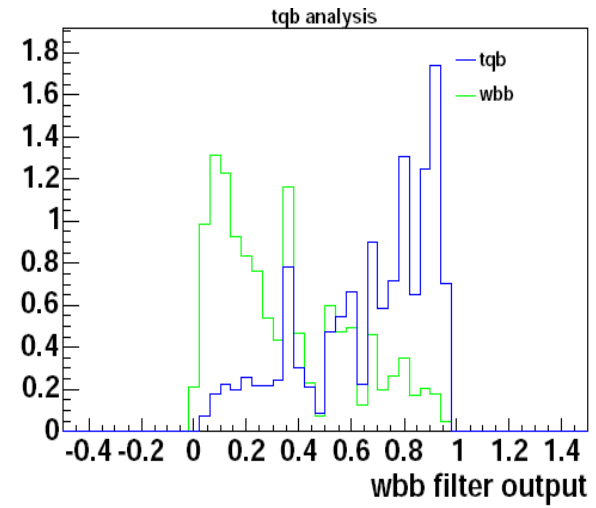


Decision Trees + Boosting

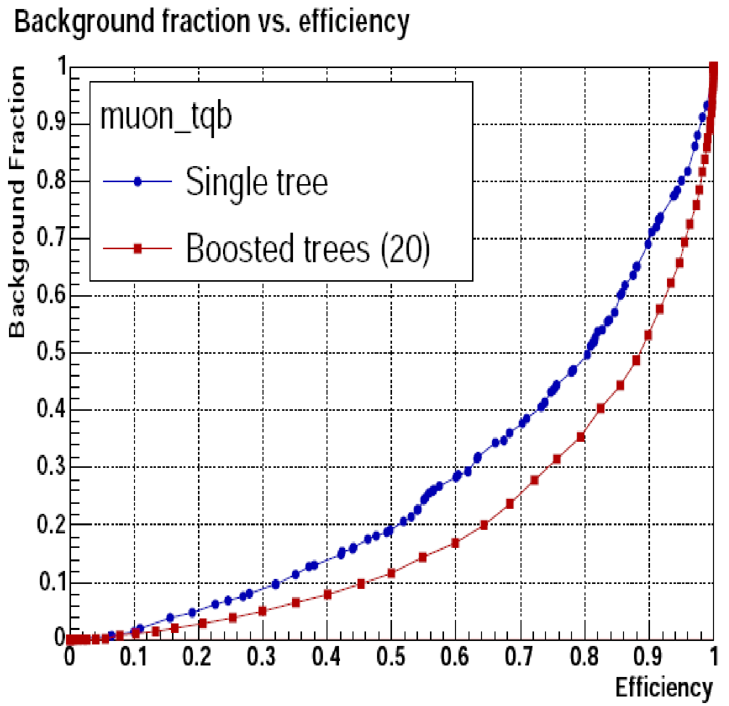
Boosting is a recent technique to improve the performance of any weak classifier: recently used in DTs by GLAST and MiniBooNE

AdaBoost algorithm: adaptive boosting

1) Train a tree T_k
2) Check which events are **misclassified** by T_k
3) Derive tree weight α_k
4) Increase weight of misclassified events
5) Train again to build T_{k+1}



- Single trees can have spikes, even with enough statistics of training events
- We use the weighted sum of 20 trees
  - Smoother distributions
  - Better separation
  - More stability



Decision Trees: 49 variables

Object Kinematics

p_T(jet1)
p_T(jet2)
p_T(jet3)
p_T(jet4)
p_T(best1)
p_T(notbest1)
p_T(notbest2)
p_T(tag1)
p_T(untag1)
p_T(untag2)

Angular Correlations

ΔR(jet1,jet2)
cos(best1,lepton)$_{\mathrm{besttop}}$
cos(best1,notbest1)$_{\mathrm{besttop}}$
cos(tag1,alljets)$_{\mathrm{alljets}}$
cos(tag1,lepton)$_{\mathrm{btaggedtop}}$
cos(jet1,alljets)$_{\mathrm{alljets}}$
cos(jet1,lepton)$_{\mathrm{btaggedtop}}$
cos(jet2,alljets)$_{\mathrm{alljets}}$
cos(jet2,lepton)$_{\mathrm{btaggedtop}}$
cos(lepton,Q(lepton)$\times z$)$_{\mathrm{besttop}}$
cos(lepton$_{\mathrm{besttop}}$,besttop$_{\mathrm{CMframe}}$)
cos(lepton$_{\mathrm{btaggedtop}}$,btaggedtop$_{\mathrm{CMframe}}$)
cos(notbest,alljets)$_{\mathrm{alljets}}$
cos(notbest,lepton)$_{\mathrm{besttop}}$
cos(untag1,alljets)$_{\mathrm{alljets}}$
cos(untag1,lepton)$_{\mathrm{btaggedtop}}$

Event Kinematics

Aplanarity(alljets,W)
M(W,best1) ("best" top mass)
M(W,tag1) ("b-tagged" top mass)
H_T(alljets)
H_T(alljets−best1)
H_T(alljets−tag1)
H_T(alljets,W)
H_T(jet1,jet2)
H_T(jet1,jet2,W)
M(alljets)
M(alljets−best1)
M(alljets−tag1)
M(jet1,jet2)
M(jet1,jet2,W)
M_T(jet1,jet2)
M_T(W)
Missing E_T
p_T(alljets−best1)
p_T(alljets−tag1)
p_T(jet1,jet2)
Q(lepton)$\times\eta$(untag1)
$\sqrt{\hat{s}}$
Sphericity(alljets,W)

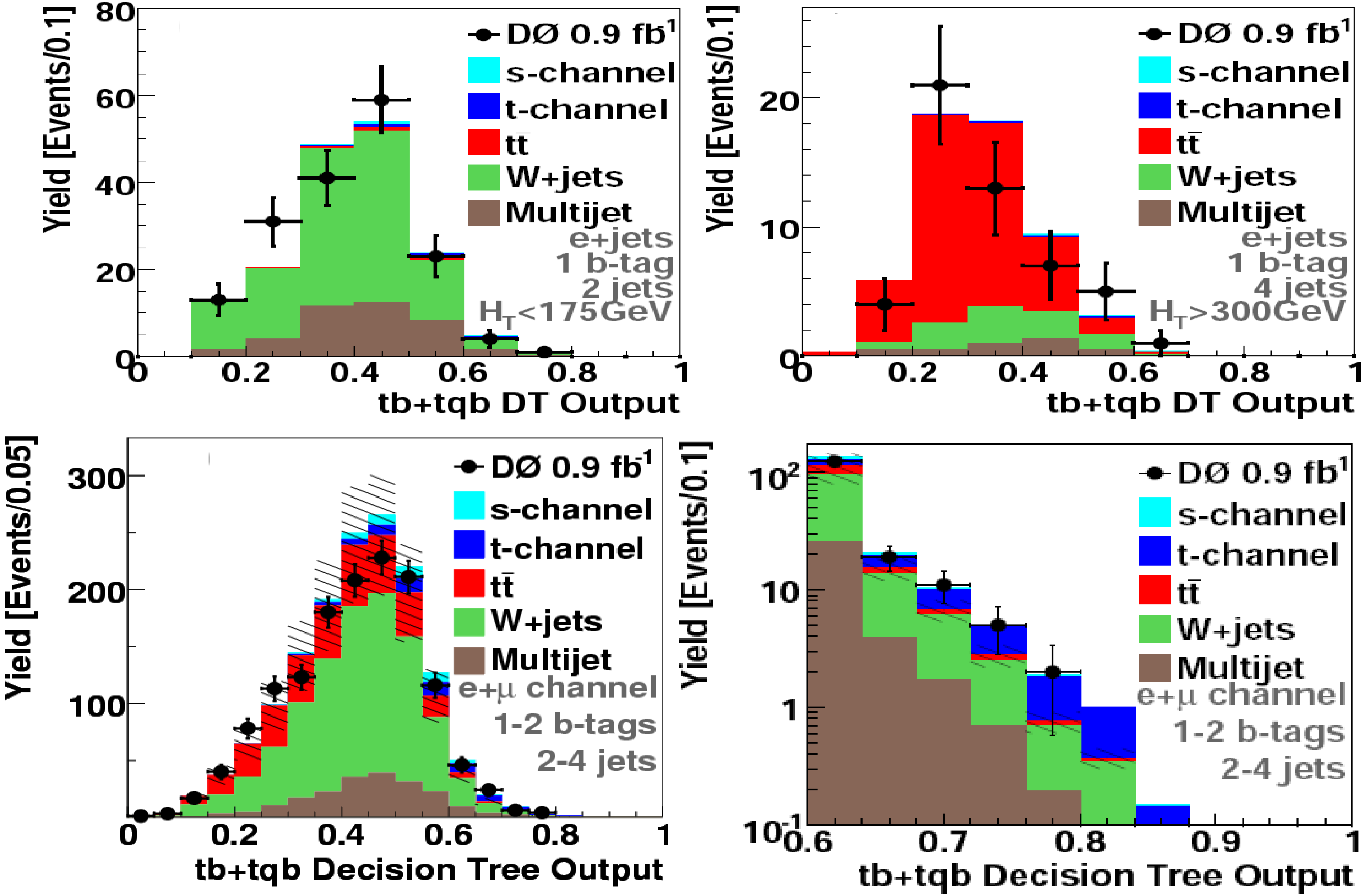
<u>Most discrimination:</u>

M(alljets)
M(W,tag1)
cos(tag1,lepton)$_{\mathrm{btaggedtop}}$
Q(lepton) x η(untag1)

- Adding variables does not degrade performance
- Tested shorter lists, lose some sensitivity
- Same list used for all channels

DT cross checks

Check the description of the data in the DT output

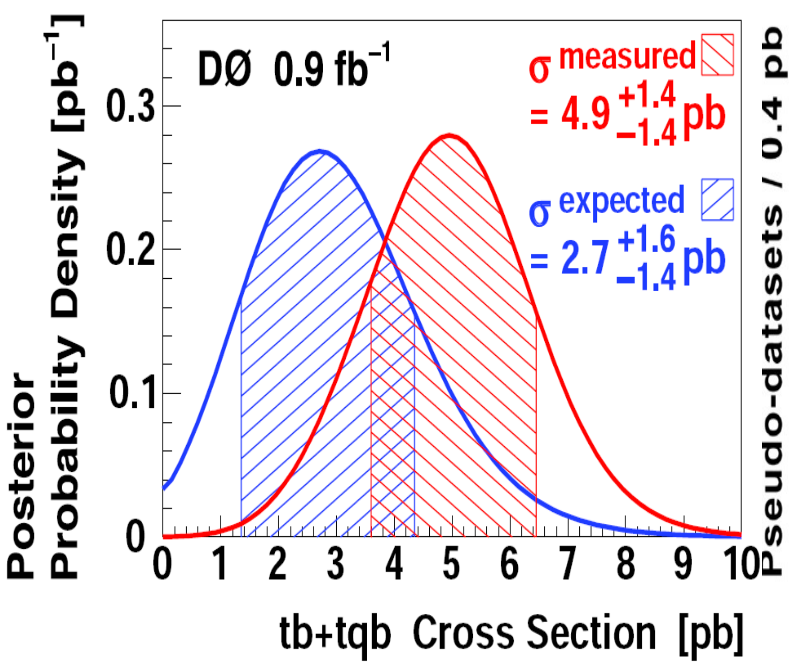


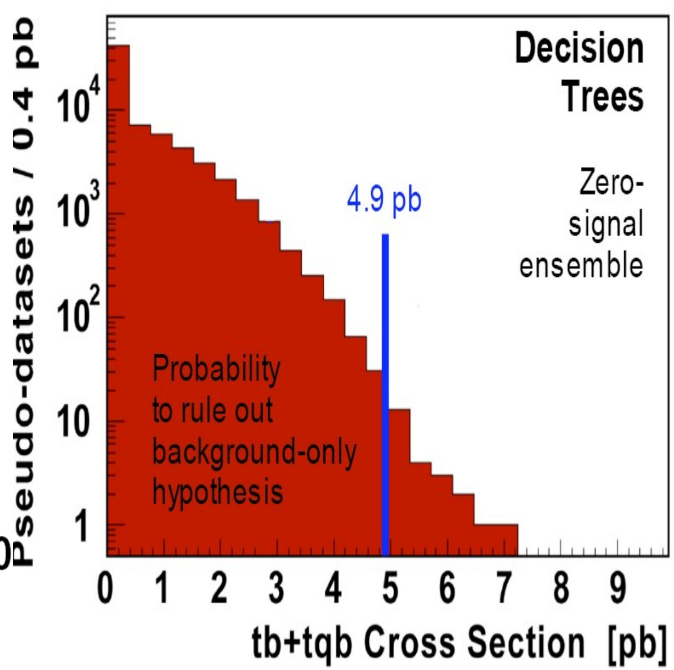
3) Measure cross section

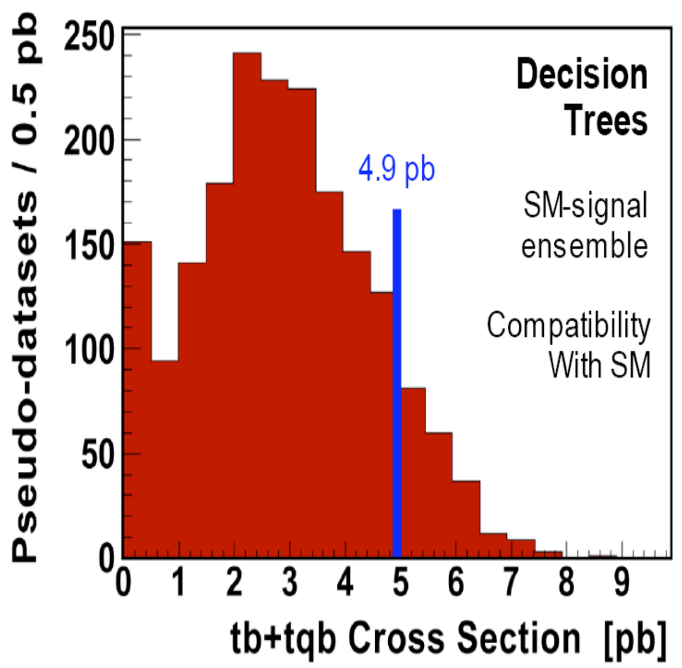
	Decision Trees		Matrix Elements		Bayesian NN	
	Expected	Observed	Expected	Observed	Expected	Observed
σ(tb+qb) [pb]	$2.7^{+1.6}_{-1.4}$	**4.9±1.4**	$2.8^{+1.6}_{-1.4}$	$\mathbf{4.8^{+1.6}_{-1.4}}$	$2.7^{+1.5}_{-1.5}$	$\mathbf{4.4^{+1.6}_{-1.4}}$
p-value×1000	17.7	**0.37**	30.7	**0.81**	10.5	**0.14**
significance	2.1σ	**3.4σ**	1.9σ	**3.2σ**	2.2σ	**3.2σ**

All three analyses measure >3σ! Evidence for single top production!

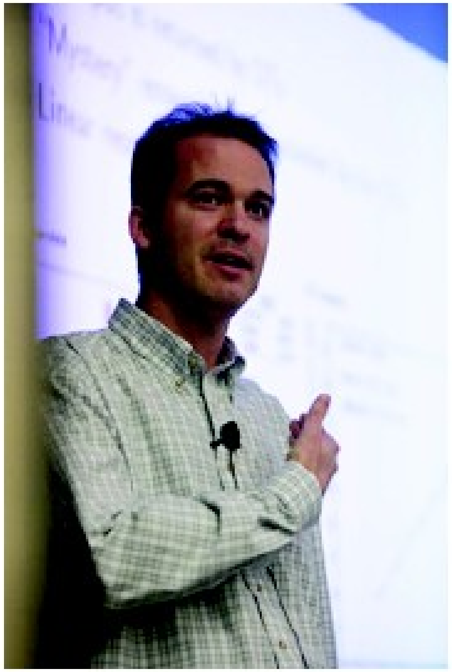
- Results are compatible with the SM at ~1 std. dev.







Announcement



SCIENTIFIC AMERICAN

PHYSICS

Alone at the Top

CLOSER TO GOD: FERMILAB MAKES SOLO TOP QUARKS BY ALEXANDER HELLEMANS

CERN COURIER

Scientists at the D0 experiment discover new path to the top

DZero finds evidence of rare single top quark; Observation marks a step closer to finding Higgs boson

Batavia, Ill.--Scientists o
Fermi National Accelera
December 8, 2006 the fi
subatomic process invol
predictions made by par
In the longer term, the te
search for an even more

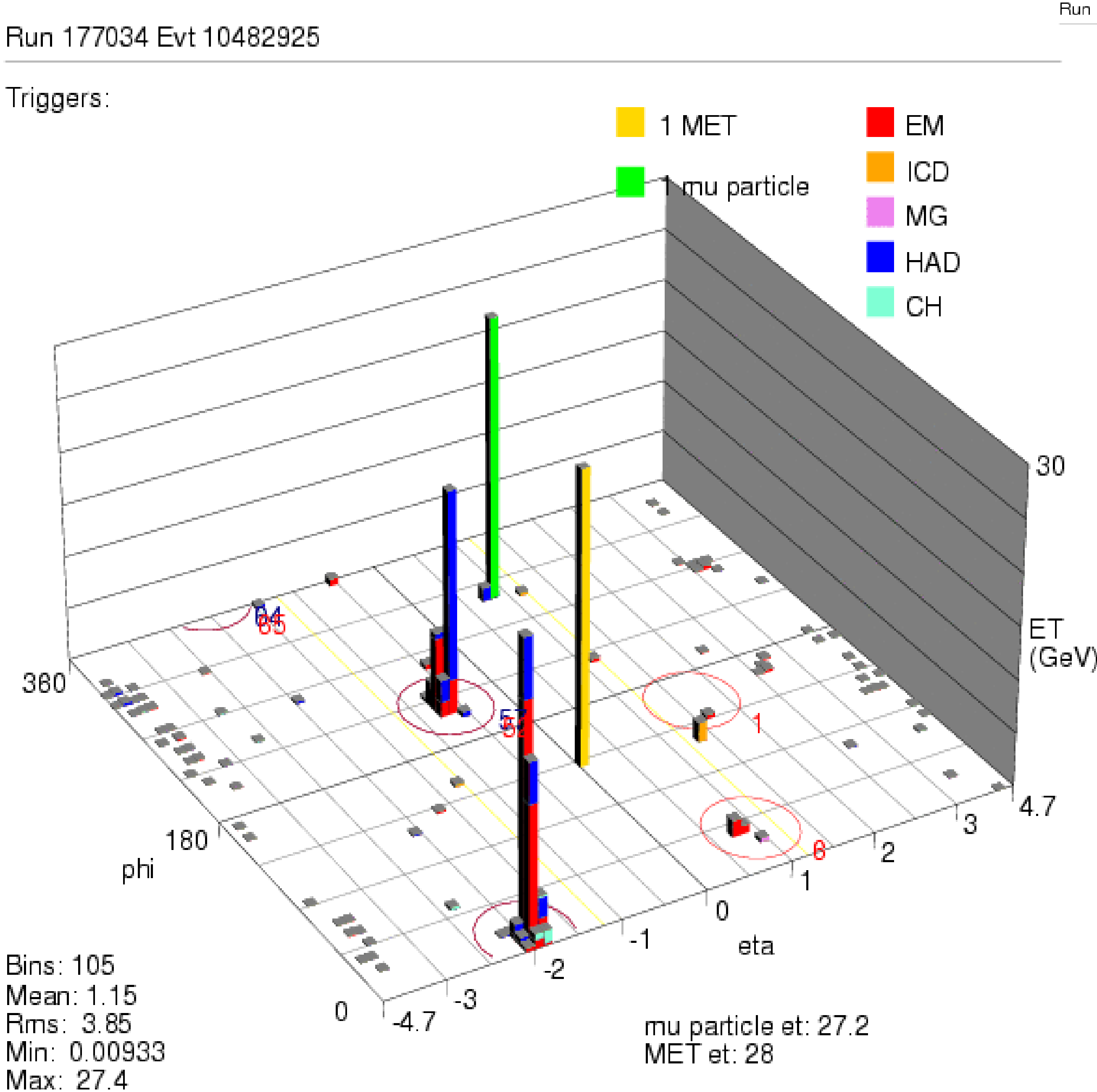
HIGH-ENERGY PHYSICS

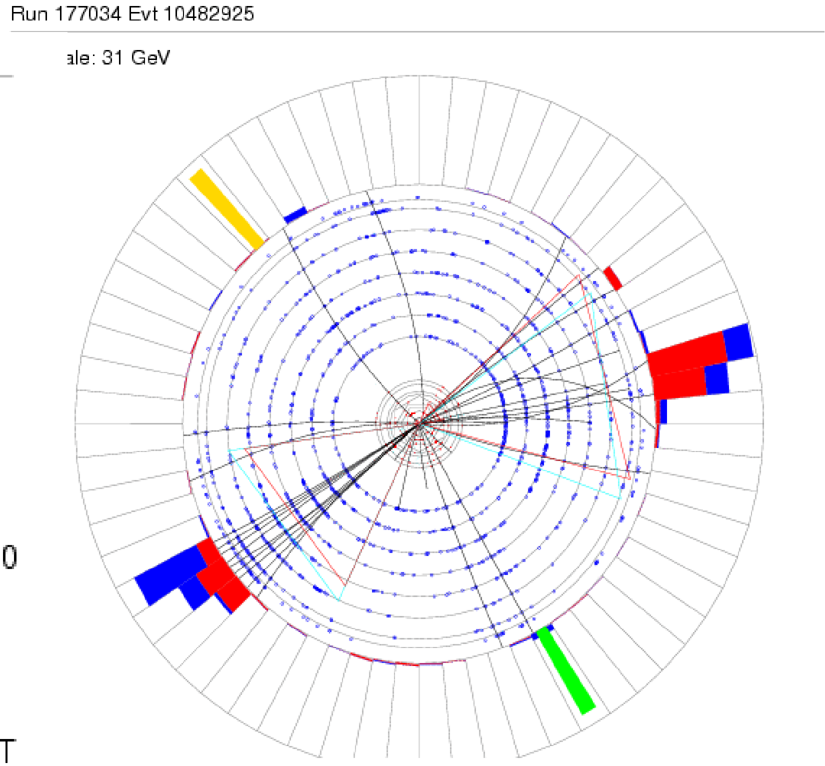
Top quarks go it alone

nature physics

The first, long-sought evidence for the production of single top quarks, by the weak interaction, has been reported from a sophisticated analysis of a large number of proton–antiproton collisions at the Tevatron.

A candidate event



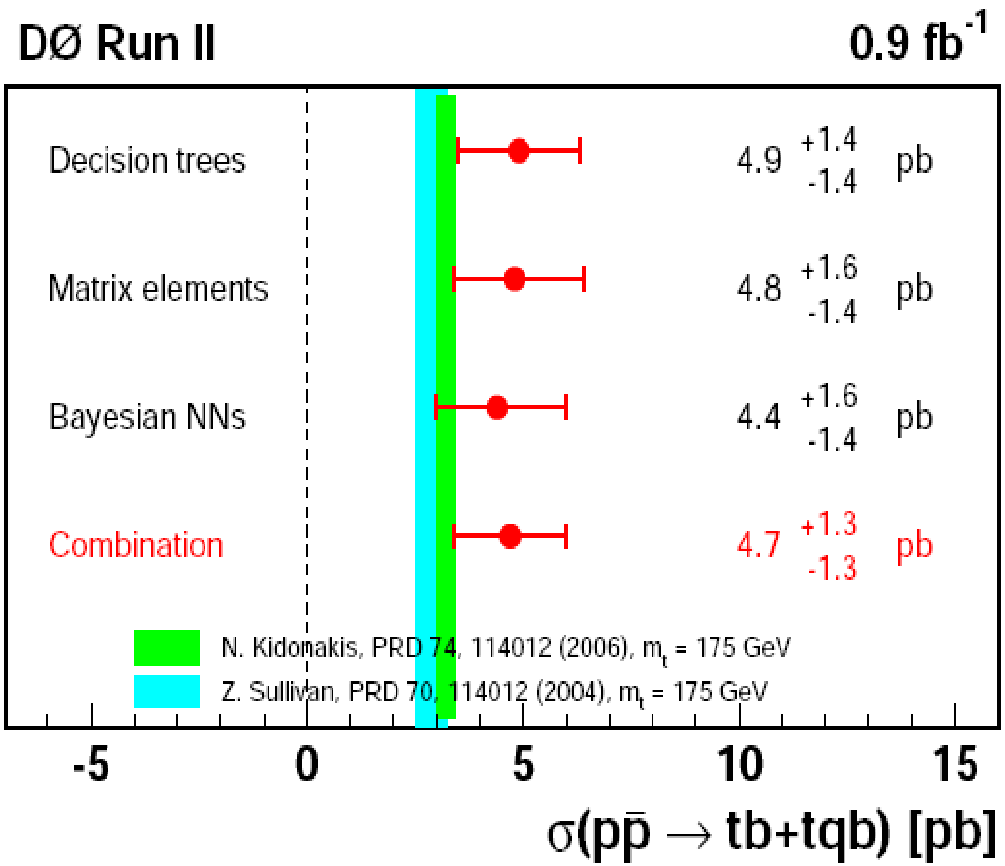


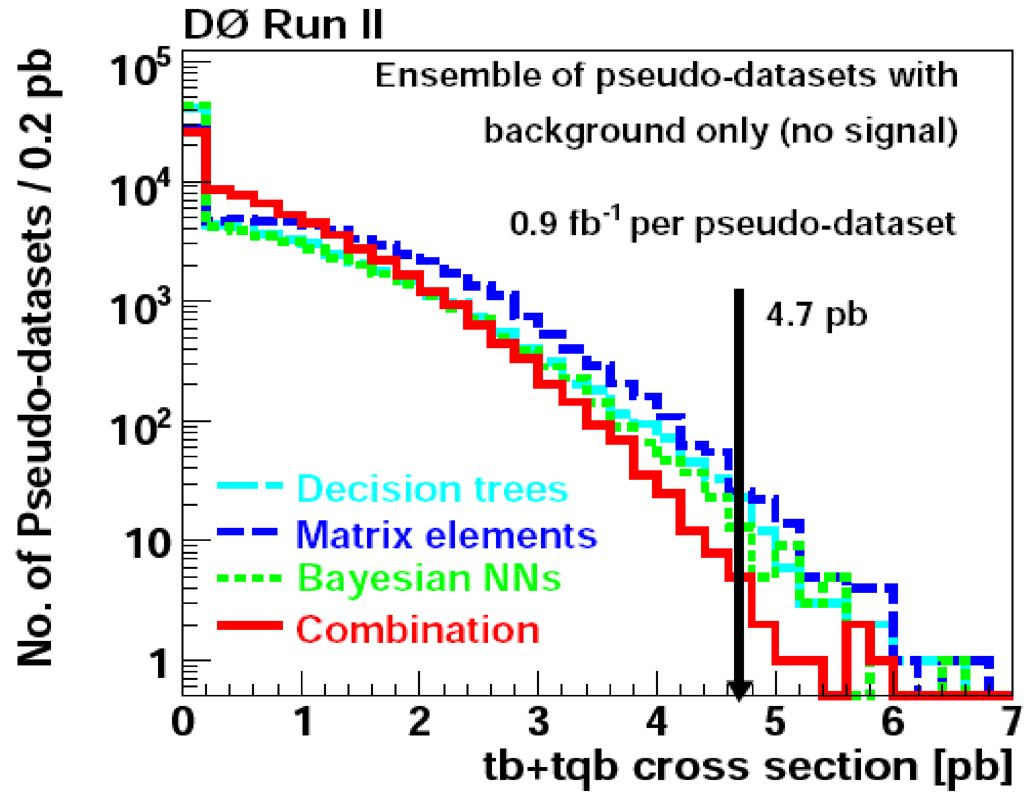
Ranked 3rd in ME, 4th in DT

- $M_T(\ell,\nu)$ = 82 GeV
- $M(\ell,\nu,b)$ = 177 GeV
- $Q\text{x}\eta$ = 1.88

Combination of analyses

Combined result: 4.7 ± 1.3 pb ➔ Significance of 3.6 std. dev.

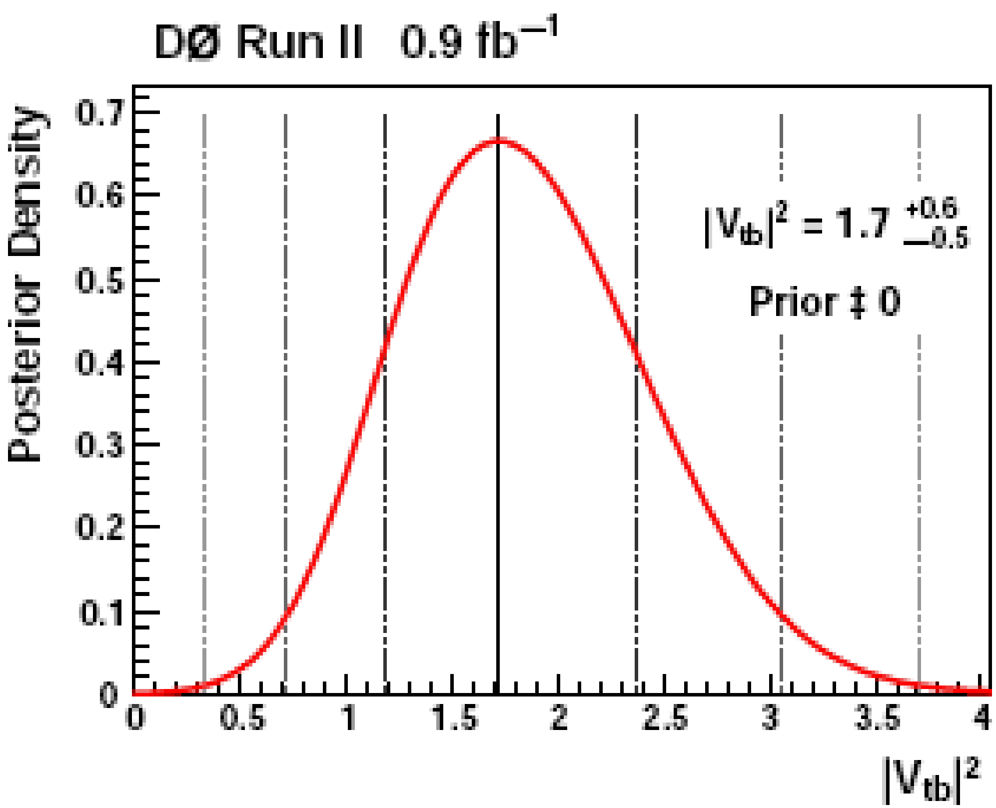


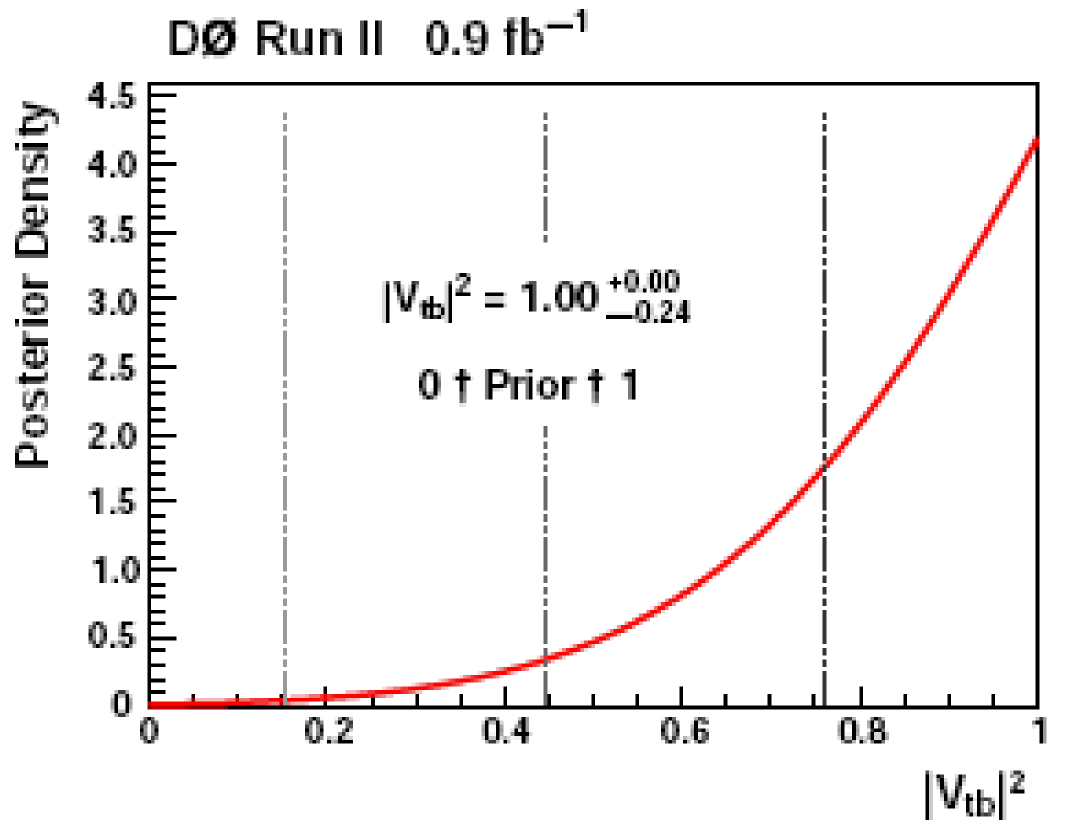


- The three multivariate methods are highly correlated
- But ME and DT look at different kinds of events
  - 50% overlap in highest ranked data events

First direct measurement of $|V_{tb}|$

- Directly translate the σ into a $|V_{tb}|$ measurement: $\sigma \propto |V_{tb}|^2$





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

f_1^L free parameter (=1 in SM)

$|V_{tb}| > 0.68$ @ 95 C.L.

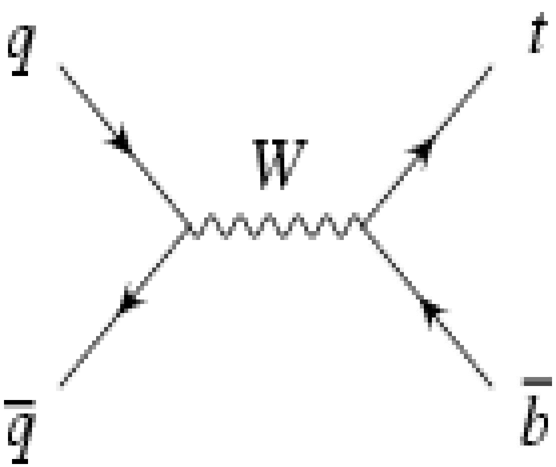
(assuming $f_1^L=1$)

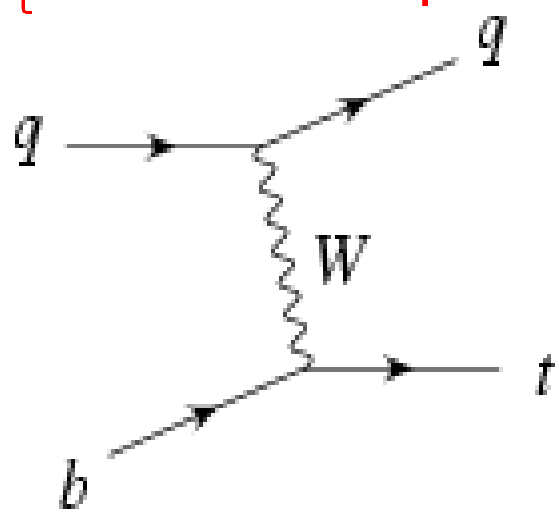
This measurement does not assume 3 generations or unitarity

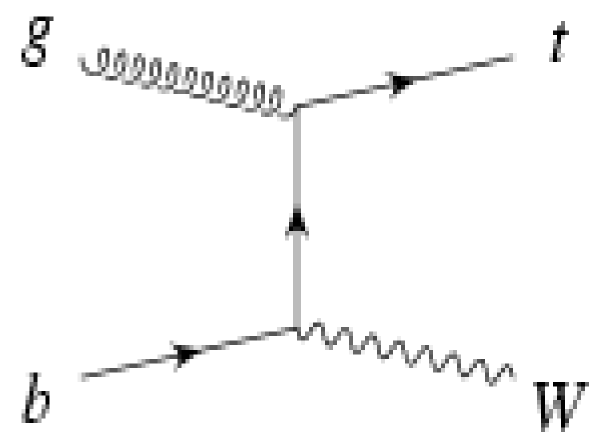
Single top prospects

- In 2008 work on the discovery, possible observation of t-channel alone
- Then the LHC will start with huge production rates:

σ_s=10.6±1.1 pb σ_t=246.6±17 pb σ_{tW}=62.0$^{+16.6}_{-3.6}$ pb

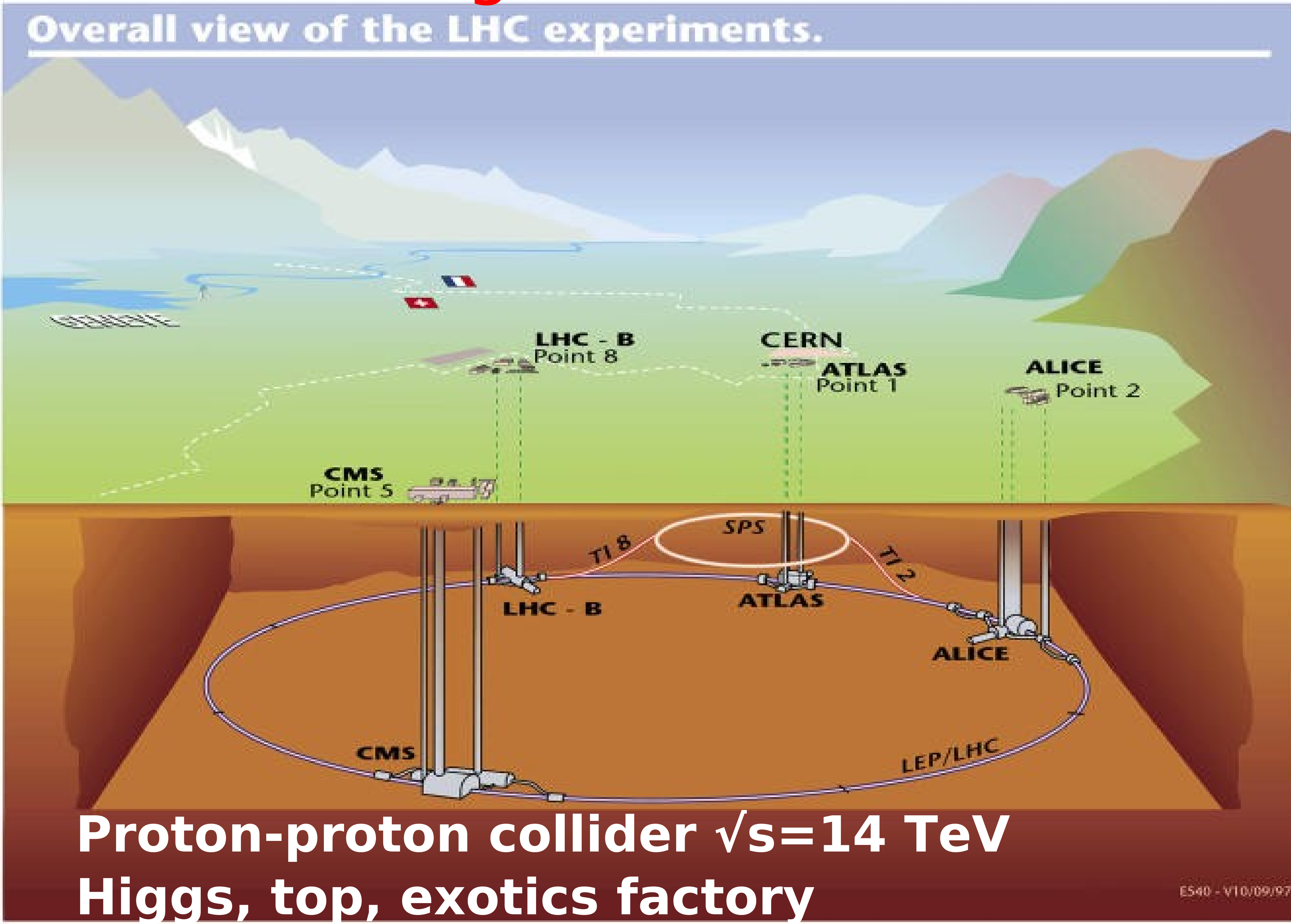






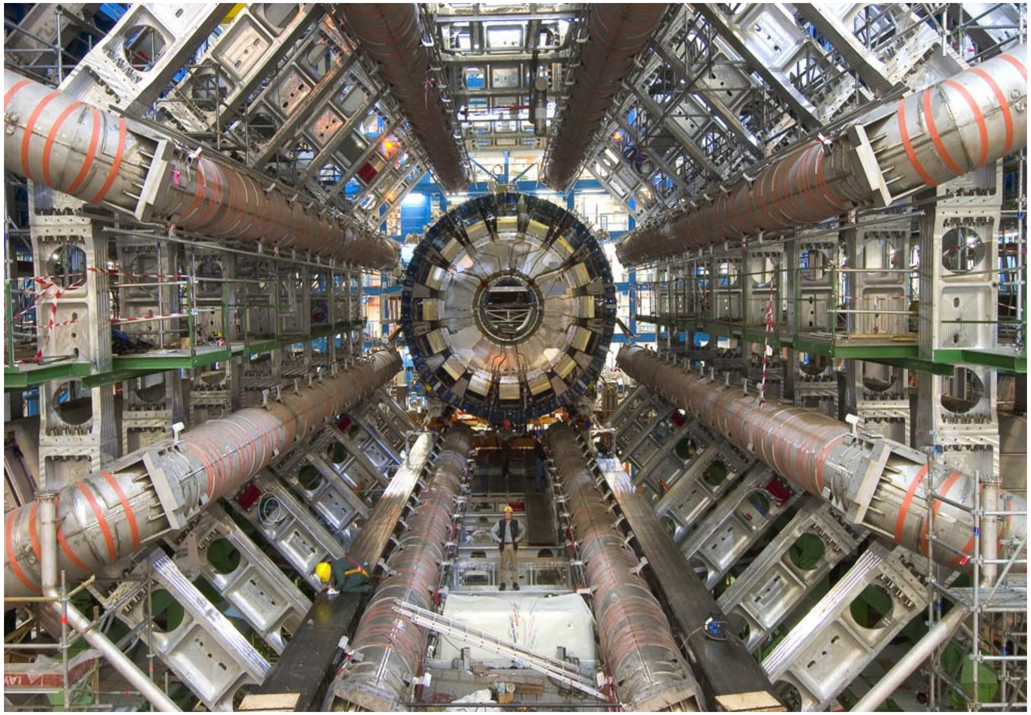
- Observe all three channels (s-channel will be tough)
- tW mode offers new window into top physics
- Measure V_{tb} to a few %
- Large samples: study properties

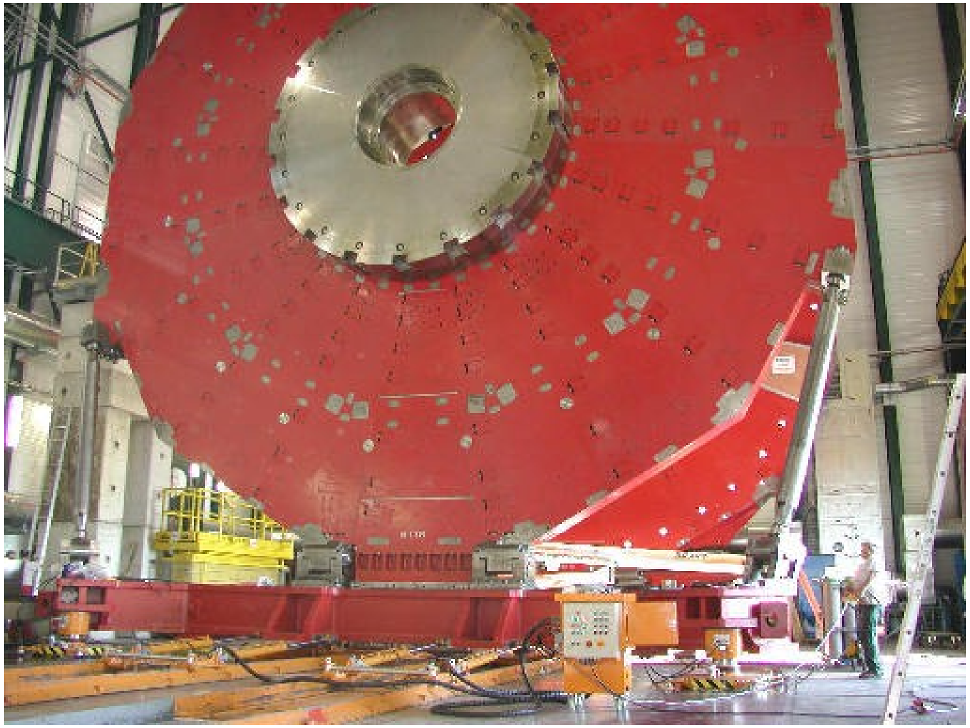
The Large Hadron Collider

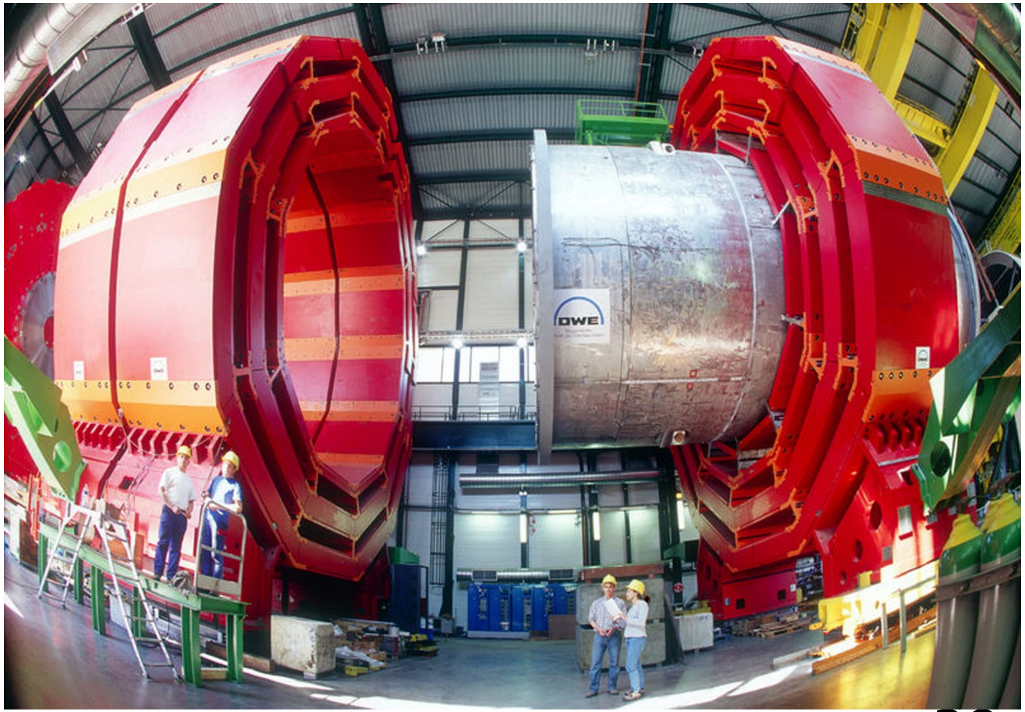


Proton-proton collider √s=14 TeV
Higgs, top, exotics factory

Starts next year!







Conclusions

- First evidence for electroweak production of top quarks and direct measurement of $|V_{tb}|$
- It is a challenging measurement, where the modeling of the large backgrounds is key
- Innovative multivariate techniques have been used to separate the small signal from the backgrounds
- Opened the way to many other analyses, like searches for the Higgs boson

$$\sigma(s+t) = 4.7 \pm 1.3 \text{ pb}$$

$$3.6\sigma \text{ significance!}$$

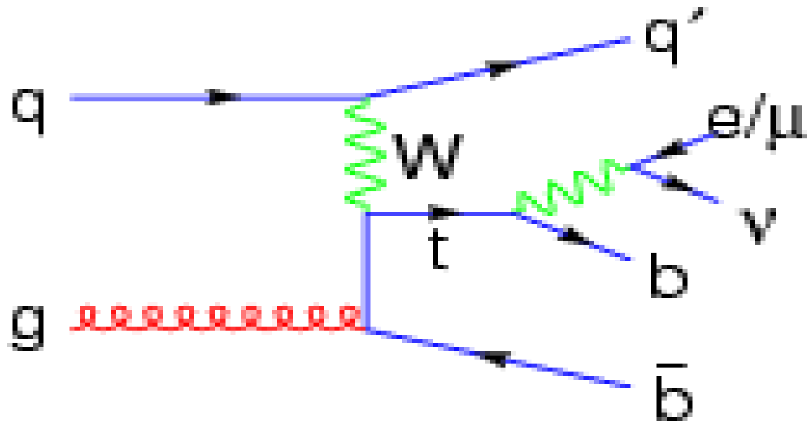
$$|V_{tb}| > 0.68 \text{ @ } 95\% \text{C.L.}$$

Published in PRL 98, 181802 (2007)

Extra slides

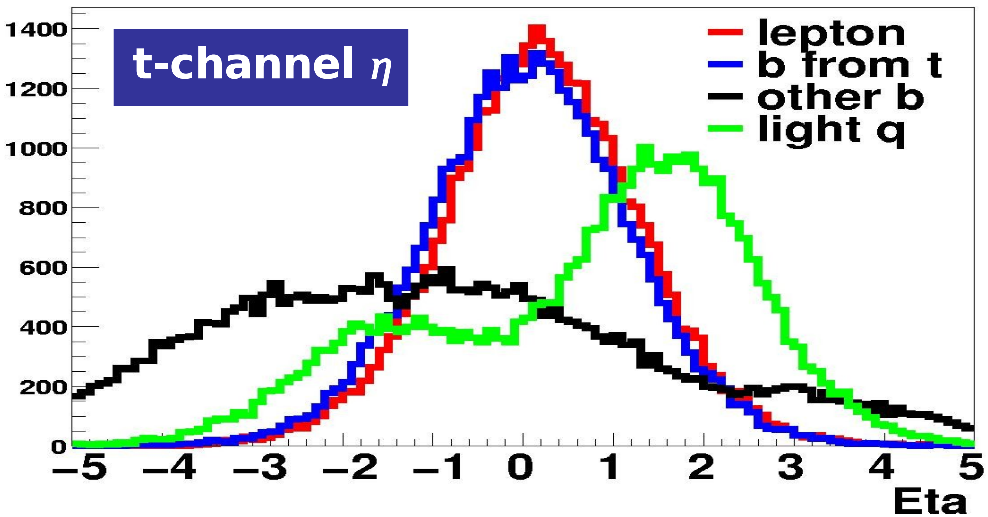
For more information:
http://www-d0.fnal.gov/Run2Physics/top/public/fall06/singletop/

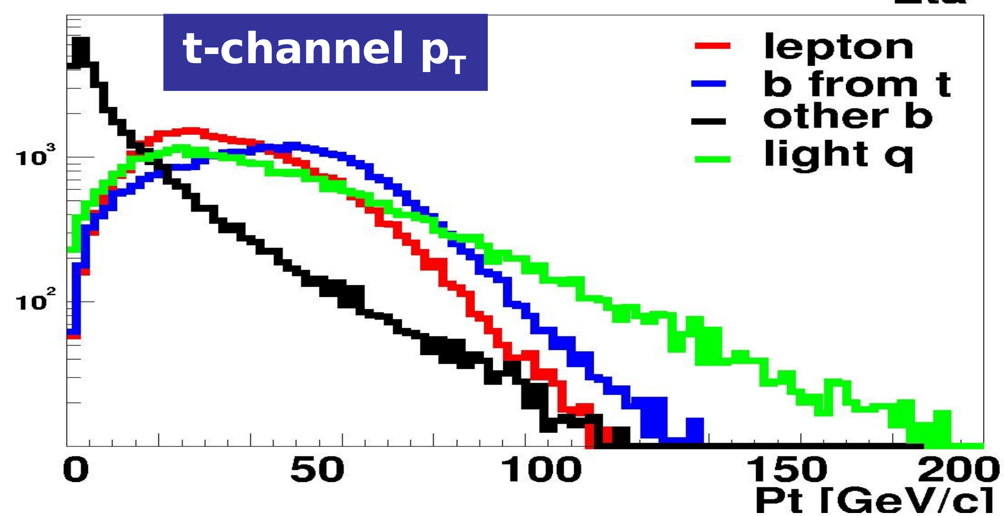
Signal selection



Signature:

- One high p_T isolated lepton (from W)
- MET (ν from W)
- One b-quark jet (from top)
- A light flavor jet and/or another b-jet





Event selection:

- Only one tight (no loose) lepton:
  - e: p_T >15 GeV and $|\eta^{det}|$<1.1
  - μ: p_T >18 GeV and $|\eta^{det}|$<2.0
- MET > 15 GeV
- 2-4 jets: p_T >15 GeV and $|\eta^{det}|$<3.4
  - Leading jet: p_T>25GeV ; $|\eta^{det}|$<2.5
  - Second leading jet: p_T >20 GeV
- One or two b-tagged jets

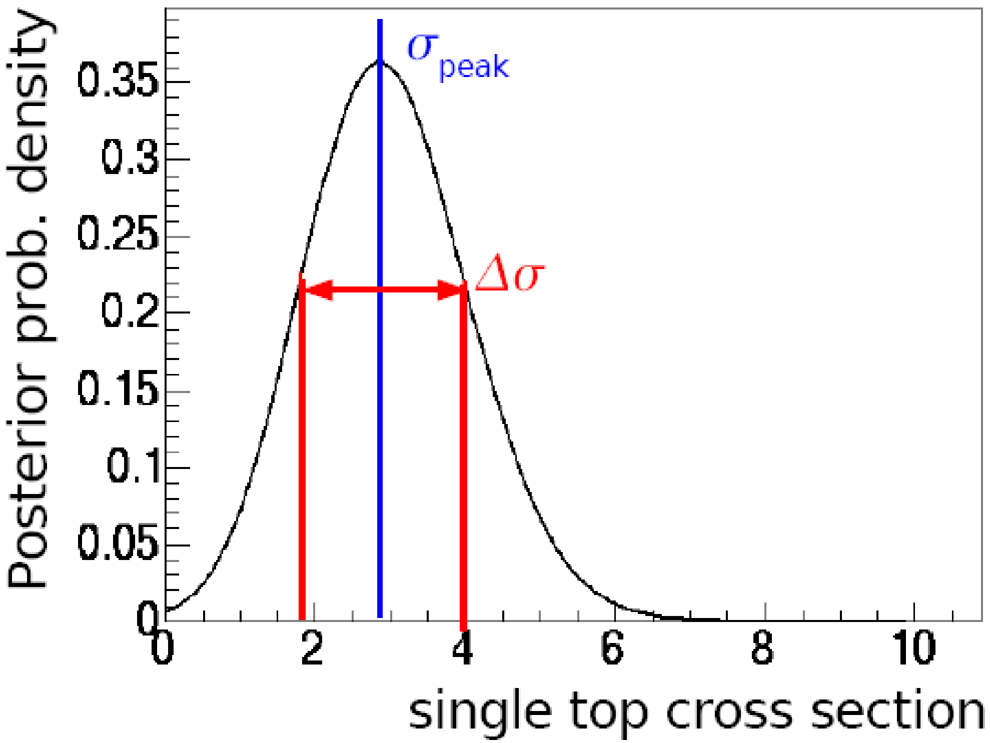
3) Measuring the cross section

- We form a binned likelihood from the discriminant outputs
- Probability to observe data distribution D, expecting y:

$$y = \alpha \mathscr{L} \sigma + \sum_{s=1}^{N} b_s = a\sigma + \sum_{s=1}^{N} b_s$$

signal — bkgd.

$$P(D|y) \equiv P(D|\sigma, a, b) = \prod_{i=1}^{nbins} P(D_i|y_i)$$



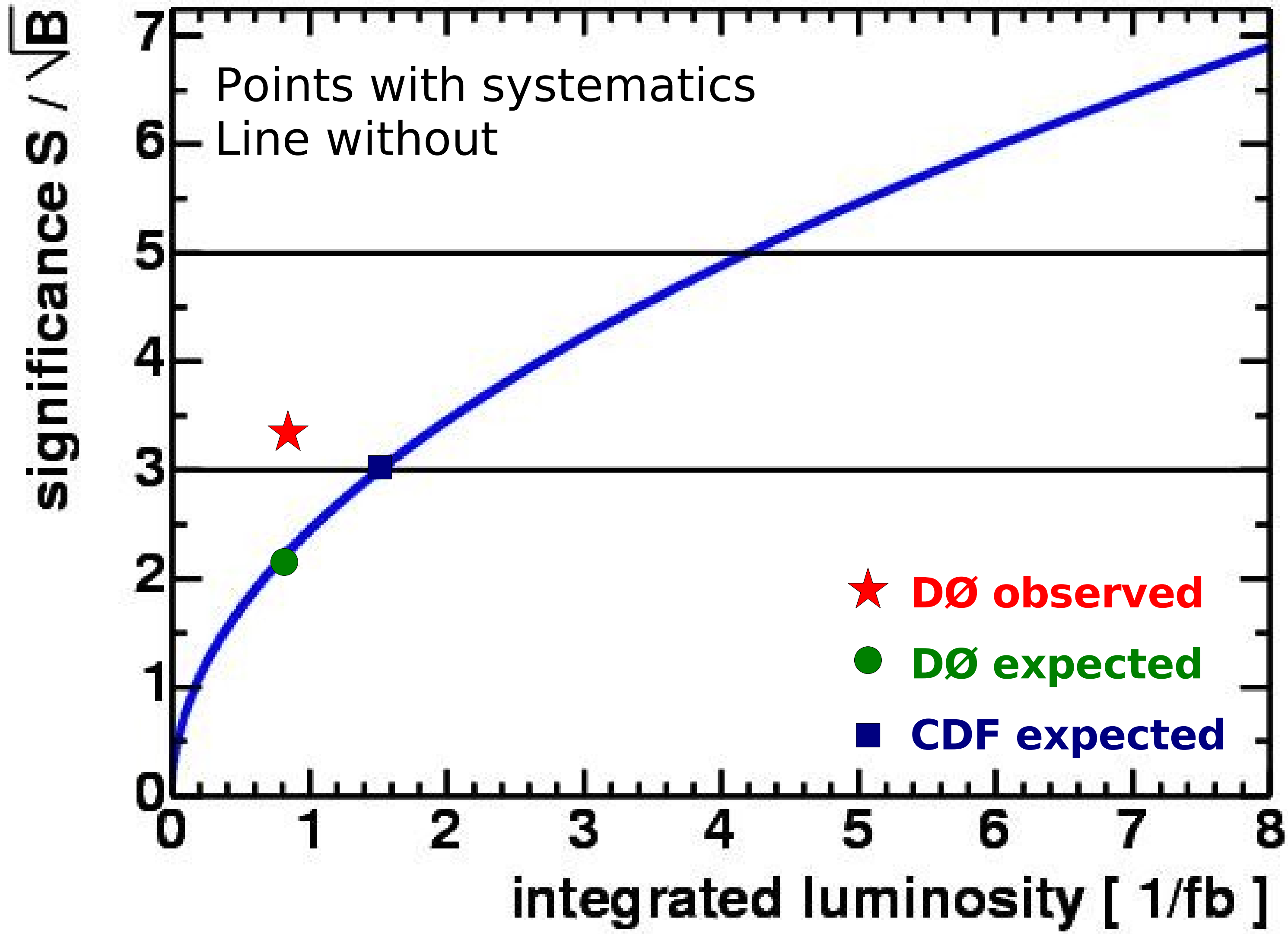
- And obtain a Bayesian posterior probability density as a function of the cross section:

$$Post(\sigma|D) \equiv P(\sigma|D) \propto \int_a \int_b P(D|\sigma, a, b) Prior(\sigma) Prior(a, b)$$

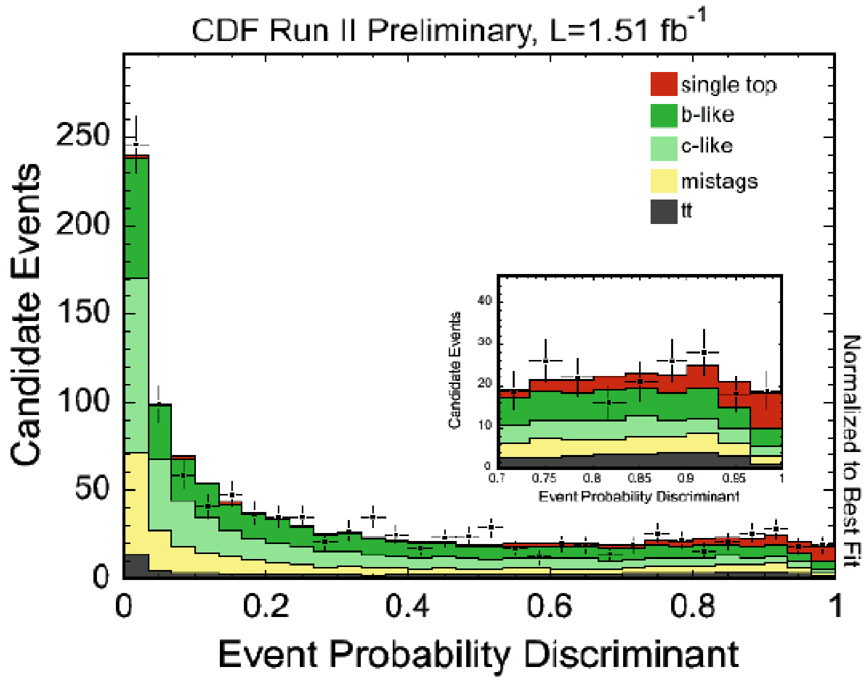
- Shape and normalization systematics treated as nuisance parameters
- Correlations between uncertainties properly accounted for
- Flat prior in signal cross section

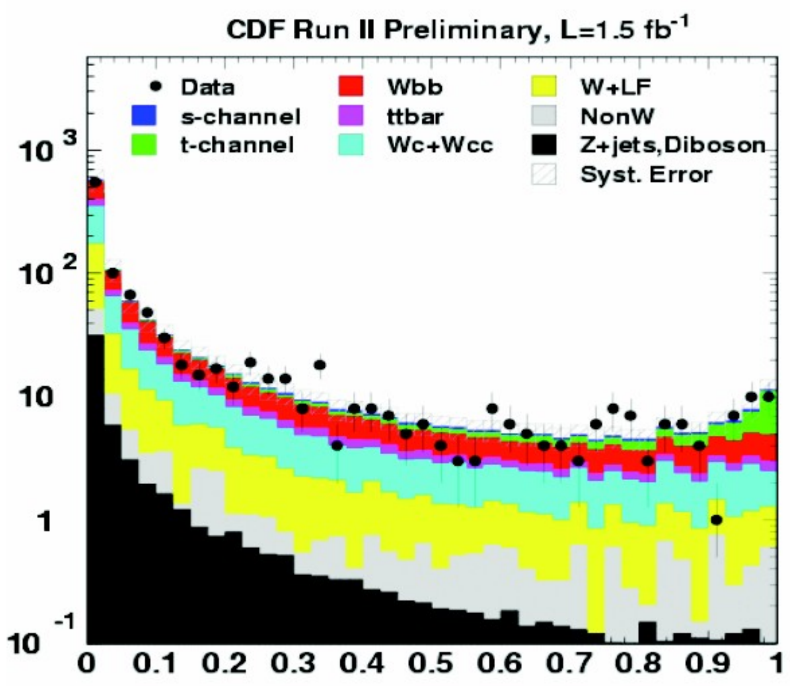
Projections for s+t

Projection by CDF for P5 in 2005



CDF's latest results





$\sigma_{s+t} = 3.0^{+1.2}_{-1.1}$pb

3.0σ expected

3.1σ observed

$\sigma_{s+t} = 2.7^{+1.3}_{-1.1}$pb

2.9σ expected

2.7σ observed

Preparing the way for the LHC

Studies at the Tevatron will help the LHC:

- Wbb measurement (will also help WH search) (DØ: hep-ex/0410062)
  - Current limit at 4.6 pb for p_T(b)>20GeV
- In general, W+jets background determination techniques
  - tt will be main background, but large uncertainties come from W+jets
  - Effect of jet vetoes (N_{jet}=2), check other methods planned in LHC analyses
- Study charge asymmetries (Bowen, Ellis, Strassler: hep-ph/0412223)
  - Signal shows asymmetry in $(Q_\ell \times \eta_j \, , \, Q_\ell \times \eta_\ell)$ plane at TeV
- Study kinematics of forward jets in t-channel (WW→H at LHC)
- Even measure asymmetry in production rate (Yuan: hep-ph/9412214) (probe CP-violation in the top sector):

$$A_t = \frac{\sigma(p\bar{p} \to tX) - \sigma(p\bar{p} \to \bar{t}\,X)}{\sigma(p\bar{p} \to tX) + \sigma(p\bar{p} \to \bar{t}\,X)}$$

TeV4LHC workshop report: 0705.3251 [hep-ph]

Crash course in Bayesian probability

Bayes' theorem expresses the degree of belief in a hypothesis A, given another B. "Conditional" probability P(A|B):

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

In HEP: B→$N_{observed}$, A→$n_{predicted}$=n_{signal}+n_{bkgd} , n_s=Acc*L*σ

P(B|A): "model" density, or likelihood: $L(N_{observed}|n_{predicted}) = n^N e^{-n}/N!$

P(A): "prior" probability density $\Pi(n_{pred}) = \Pi(Acc*L, n_b)\,\Pi(\sigma)$

$\Pi(n_s, n_b)$ multivariate gaussian ; $\Pi(\sigma)$ assumed flat

P(B): normalization constant Z: $P(N_{observed})$

P(A|B): "posterior" probability density $P(n_{predicted}|N_{observed})$

$$P(n_{predicted}|N_{observed}) = 1/Z\; L(N_{observed}|n_{predicted})\,\Pi(n_{pred})$$

W+jets normalization

- Find fractions of real and fake isolated ℓ in the data before b-tagging. Split samples in loose and tight isolation:

$$N^{loose} = N^{loose}_{fake} + N^{loose}_{real}$$

$$N^{tight} = \varepsilon_{fake} N^{loose}_{fake} + \varepsilon_{real} N^{loose}_{real}$$

Obtain: N^{loose}_{real} and N^{loose}_{fake}

- Normalize the MC Wjj and Wbb samples to the real ℓ yield found in data, after correcting for the presence of tt events:

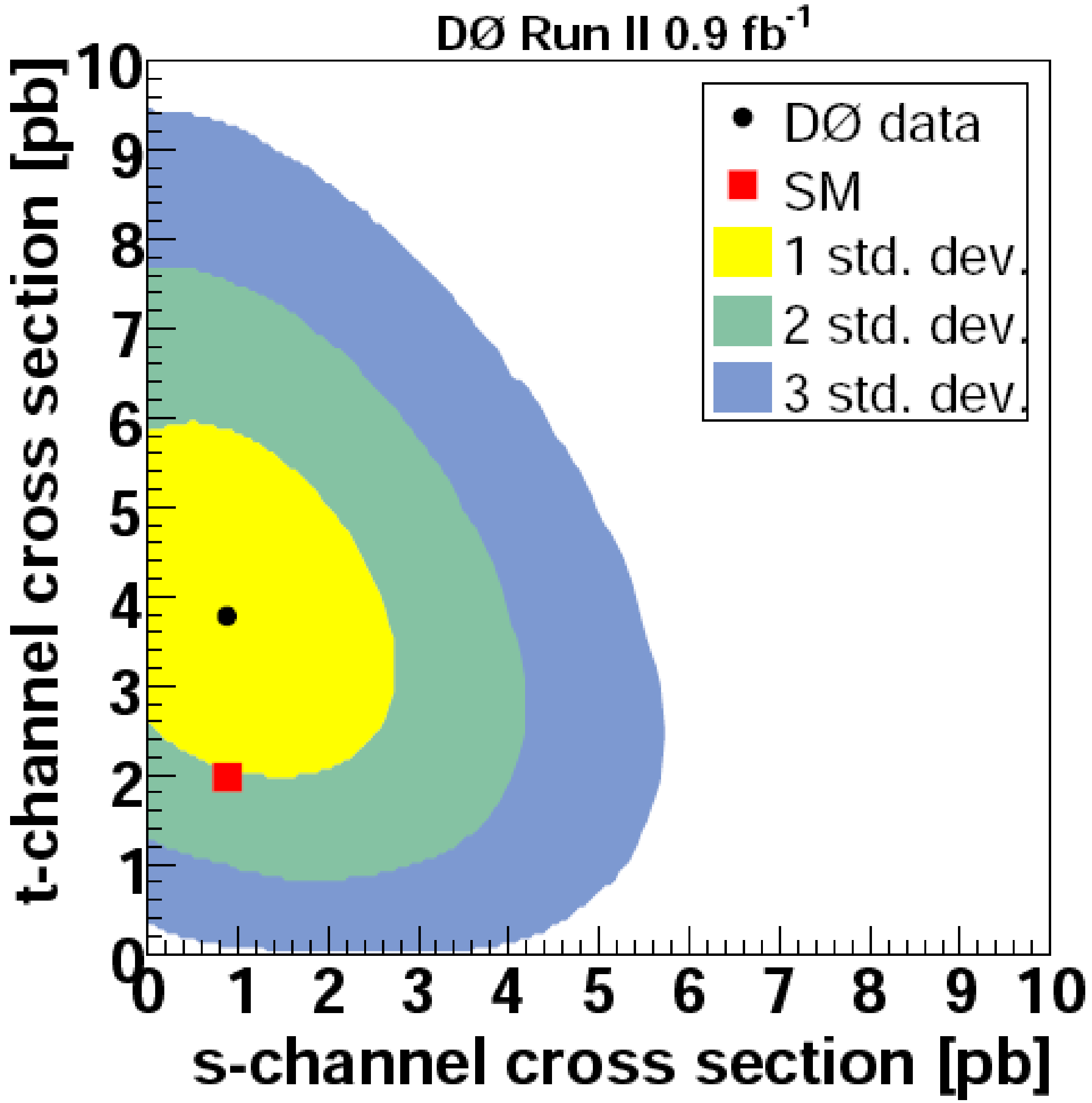
$$\varepsilon_{real} N^{loose}_{real} = SF\left[Y(Wjj) + Y(Wb\bar{b}) + Y(Wc\bar{c})\right] + Y(t\bar{t})$$

SF=1.4

- The sum Y(Wjj)+Y(Wbb)+Y(Wcc) is done according to the ratio of (Wbb+Wcc)/Wjj found in 0-tag data → 1.5±0.5
- Then apply b-tagging
  - Greatly reduce W+jets background (Wbb ~1% of Wjj)
  - Shift distributions, changes flavor composition

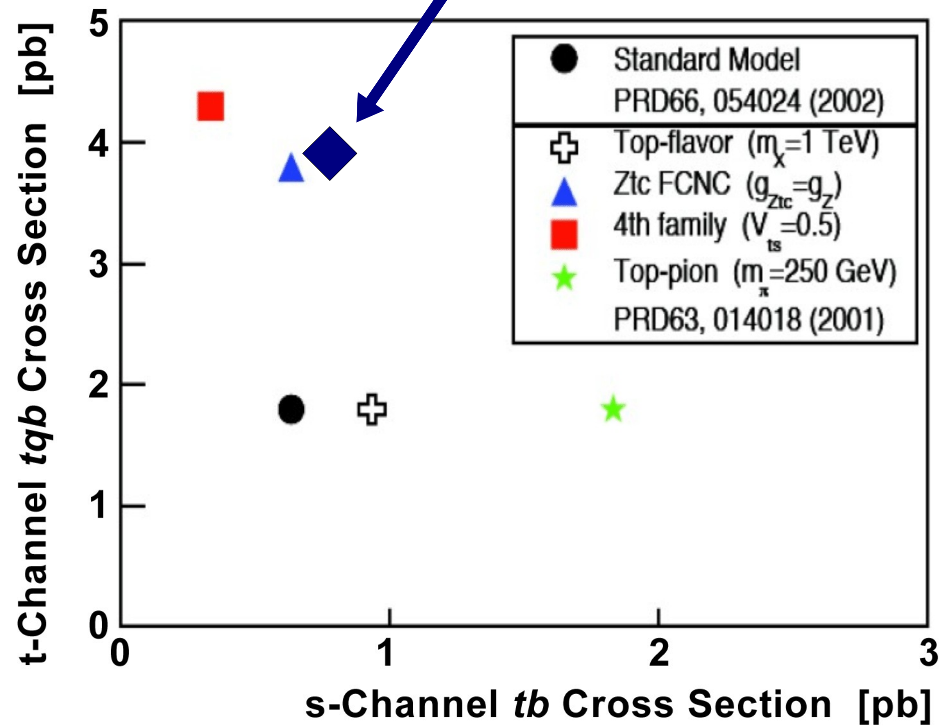
tb and tqb separately

- Remove the constraint of SM s:t ratio
- Measure model independent s- and t-channel cross sections



$$\sigma(tqb) = 3.8\ ^{+1.8}_{-1.4}\ \text{pb}$$

$$\sigma(tb) = 0.9 \pm 0.9\ \text{pb}$$



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

Systematic uncertainties

- Uncertainties are assigned per background, jet multiplicity, lepton channel, and number of tags
- Uncertainties that affect both the **normalization** and the **shapes**: JES and tag rate functions
- Correlations between channels and sources are taken into account

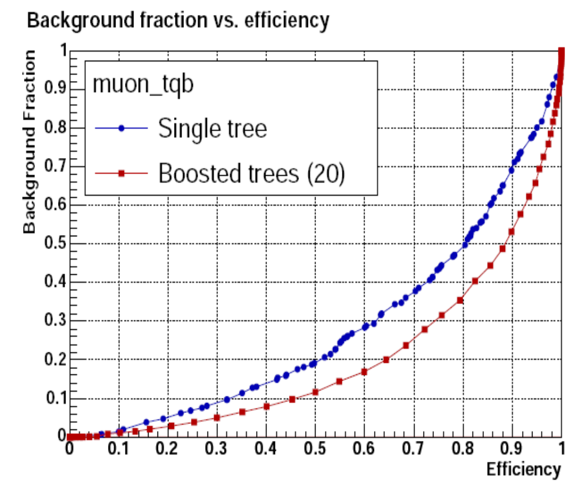
Relative systematic uncertainties

Component	Size
W+jets & QCD normalization	18 – 28%
top pair normalization	18%
Tag rate functions (+shape)	2 – 16%
Jet energy scale (+shape)	1 – 20%
Luminosity	6%
Trigger modeling	3 – 6%
Lepton ID	2 – 7%
Jet modeling	2 – 7%
Other small components	few%

Decision Trees + Boosting

Boosting is a recent technique to improve the performance of any weak classifier: recently used in DTs by GLAST and MiniBooNE

AdaBoost algorithm: adaptive boosting



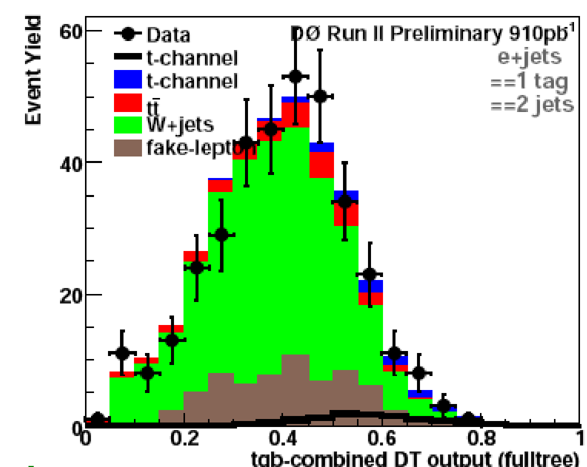
1) Train a tree T_k

2) Check which events are **misclassified** by T_k

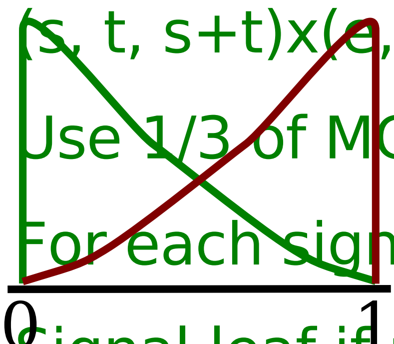
3) Derive tree weight α_k

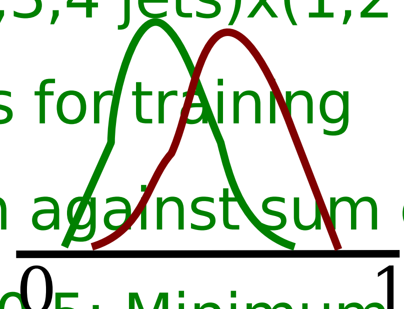
4) Increase weight of misclassified events

5) Train again to build T_{k+1}



- We have trained 36 separate trees: (s, t, s+t)x(e,mu)x(2,3,4 jets)x(1,2 tags)
- Use 1/3 of MC events for training
- For each signal, train against sum of backgrounds
- Signal leaf if purity>0.5; Minimum leaf size=100 events; Goodness of split: Gini factor; Adaboost β=0.2; boosting cycles=20





Bayesian Neural Networks

A different sort of NN (http://www.cs.toronto.edu/radford/fbm.software.html):

- Instead of choosing one set of weights, find posterior probability density over all possible weights
- Averages over many networks weighted by the probability of each network given the training data
- Use 24 variables (subset of the DT variables) and train against sum of backgrounds

Advantages:

- Less prone to overfitting, because of Bayesian averaging
- Network structure less important: can use large networks!
- Optimized performance

Disadvantages:

- Computationally demanding!