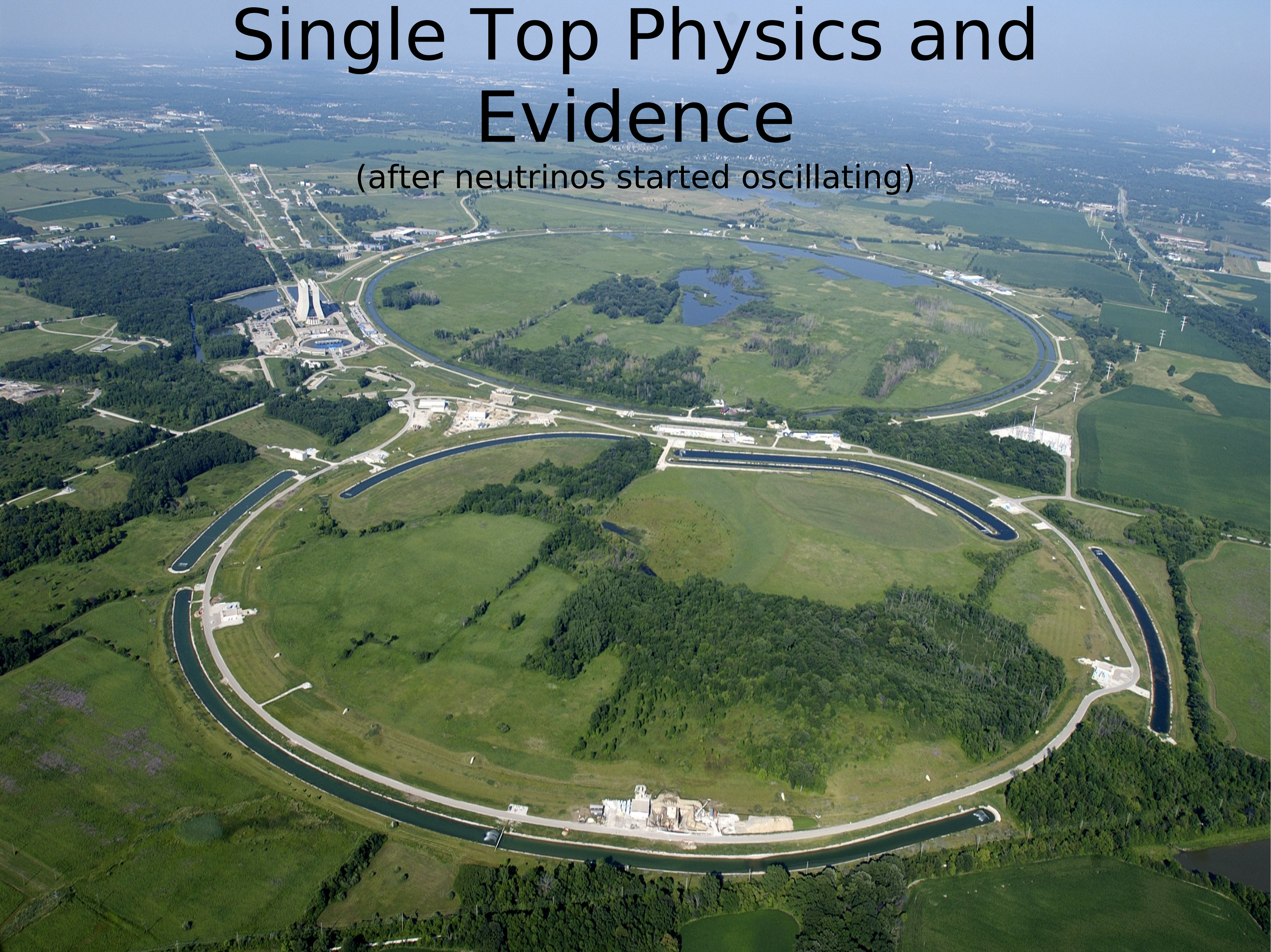
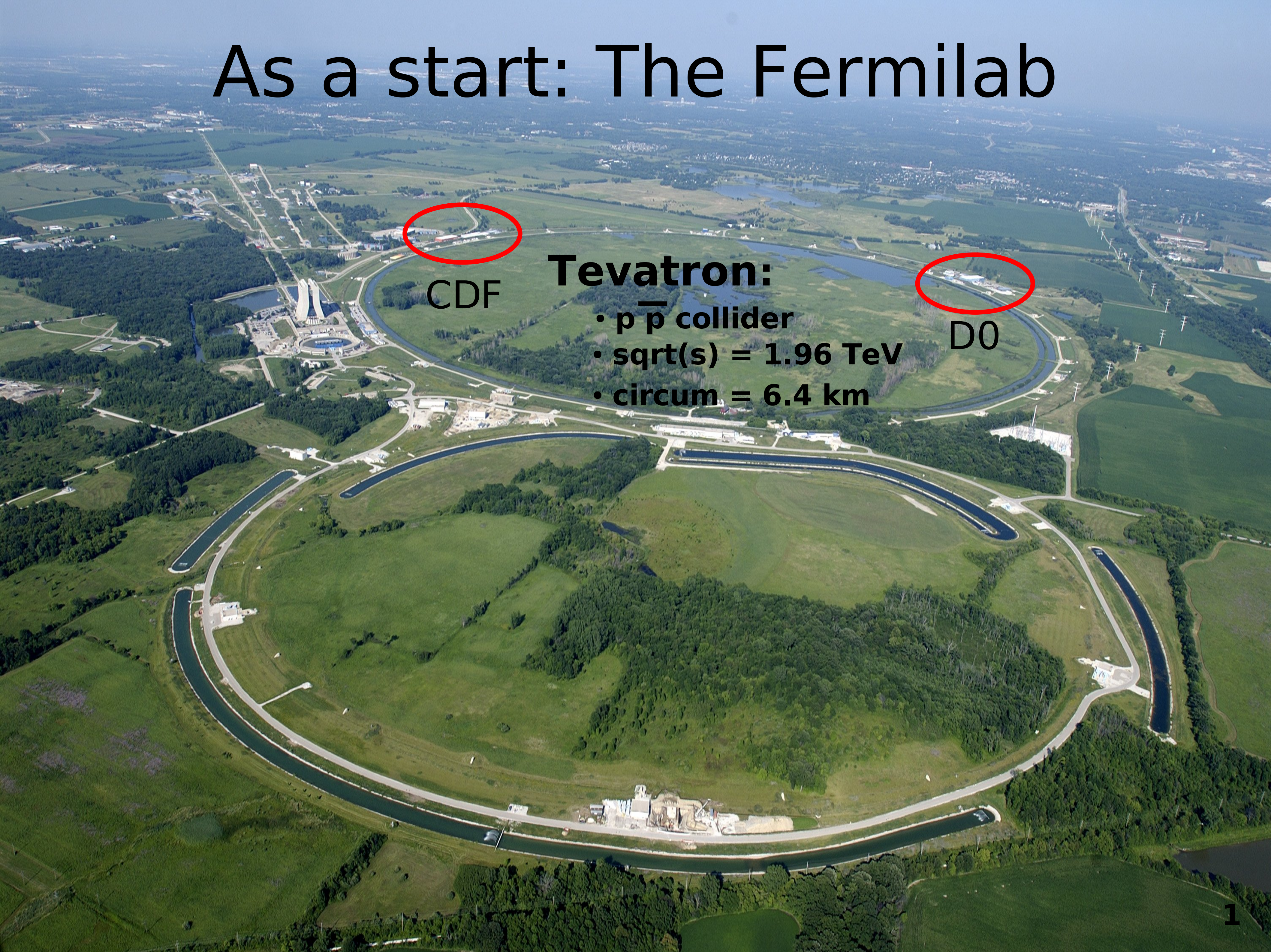
Single Top Physics and Evidence

(after neutrinos started oscillating)



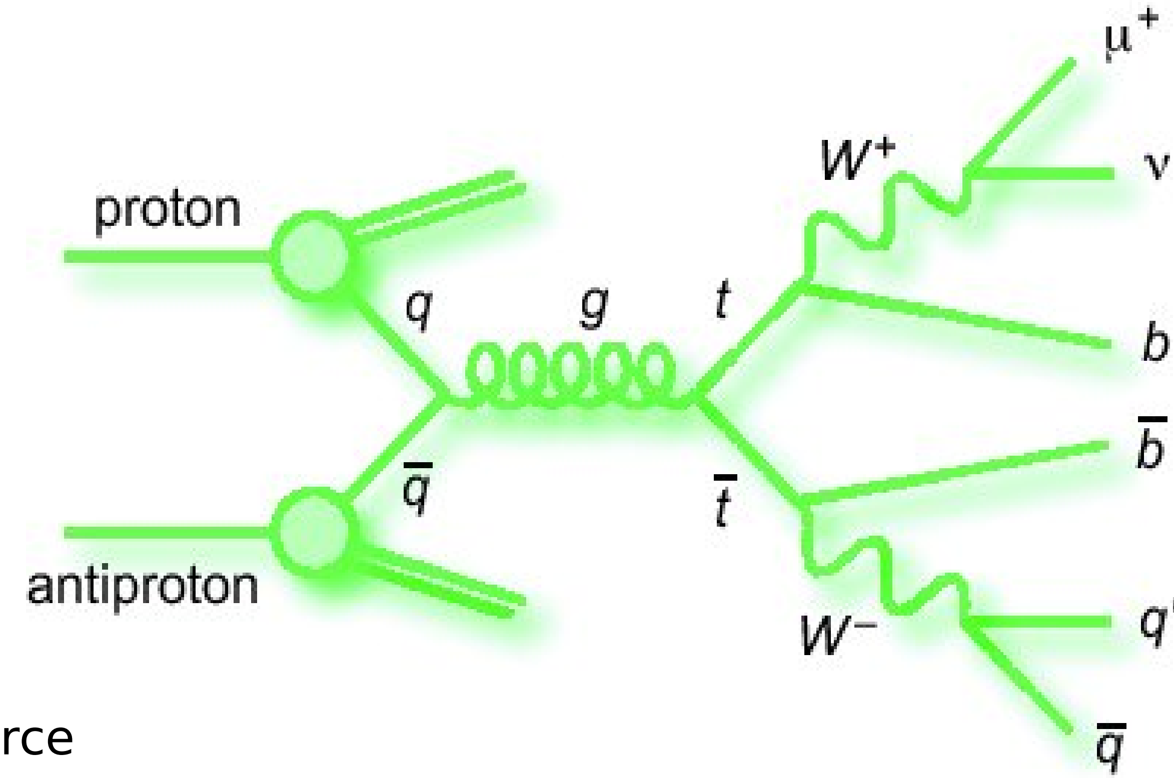
As a start: The Fermilab



Tevatron:

- $p\bar{p}$ collider
- sqrt(s) = 1.96 TeV
- circum = 6.4 km

Short Reminder: Top pairs

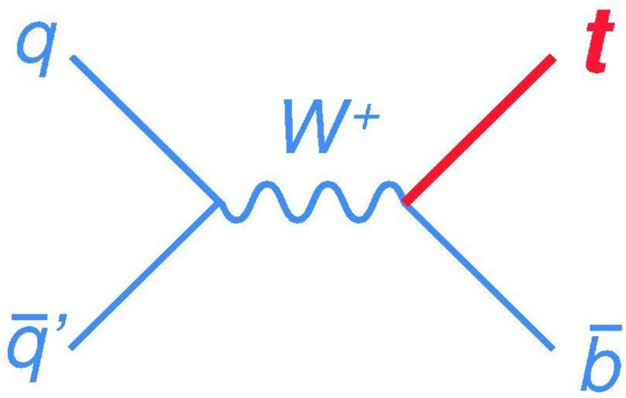


- strong force
- "typically" 4 jets, with 2 b jets
- cross section: ~6.5 pb
- first observation with dataset of ~**60 pb^{-1}** in 1995 by CDF and D0

Now, what is this "Single-Top" then?

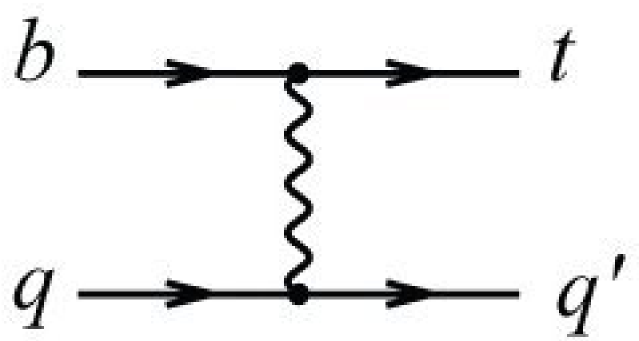
- **Electroweak** top quark production via **Wtb**-vertex

- SM process BUT not yet observed



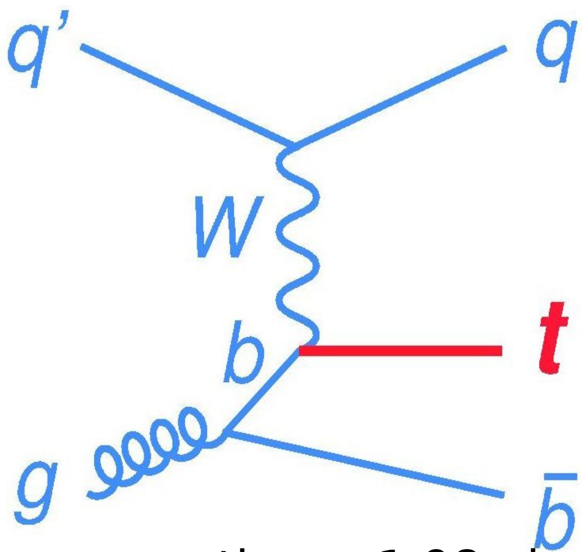
cross section = 0.88 pb

s-channel aka. tb



cross section = 0.08 pb

associated production



cross section = 1.98 pb

t-channel aka. tqb

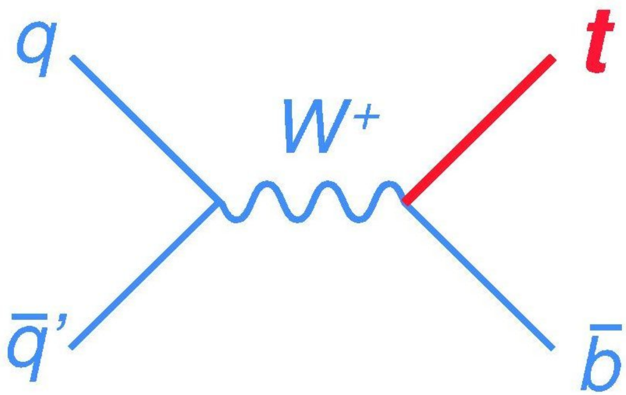
- Top is „standard candle" for LHC

- Signal of today is background of tomorrow

Now, what is this “Single-Top“ then?

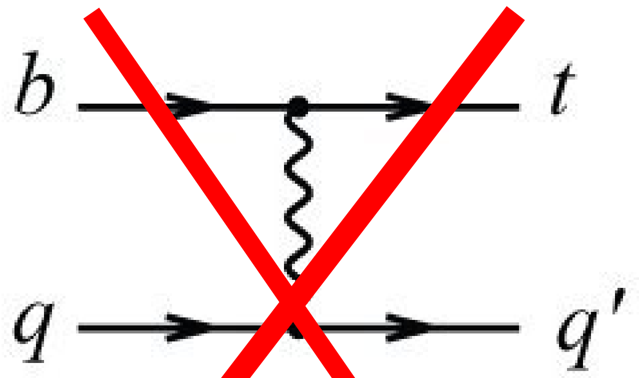
- **Electroweak** top quark production via **Wtb**-vertex

- SM process BUT not yet observed



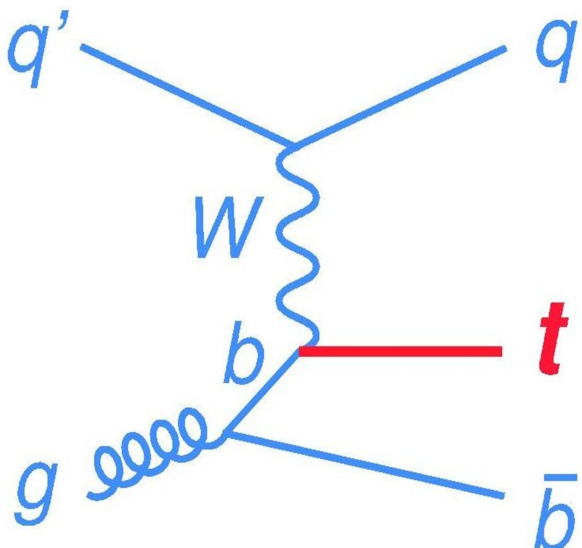
cross section = 0.88 pb

s-channel aka. tb



cross section = 0.08 pb

associated production



cross section = 1.98 pb

t-channel aka. tqb

- Top is „standard candle“ for LHC

- Signal of today is background of tomorrow

Why is it interesting?

- Lifetime is of order 10^{-24} s, decays before it hadronizes
    - Possible to really measure its (kinematic) mass!
    - Measure Spin of top directly

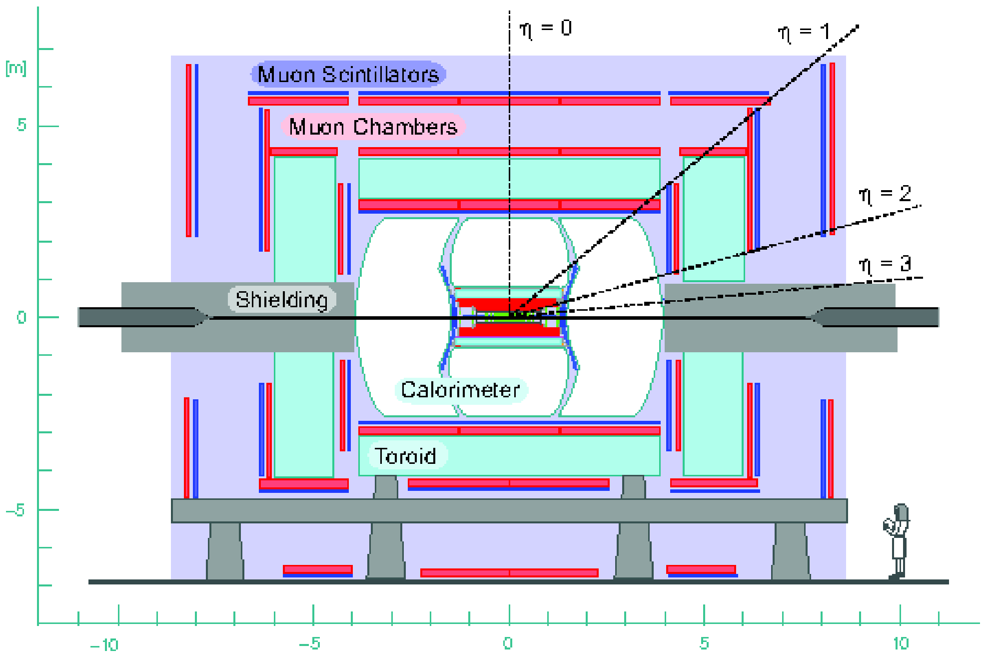
- Not yet observed!

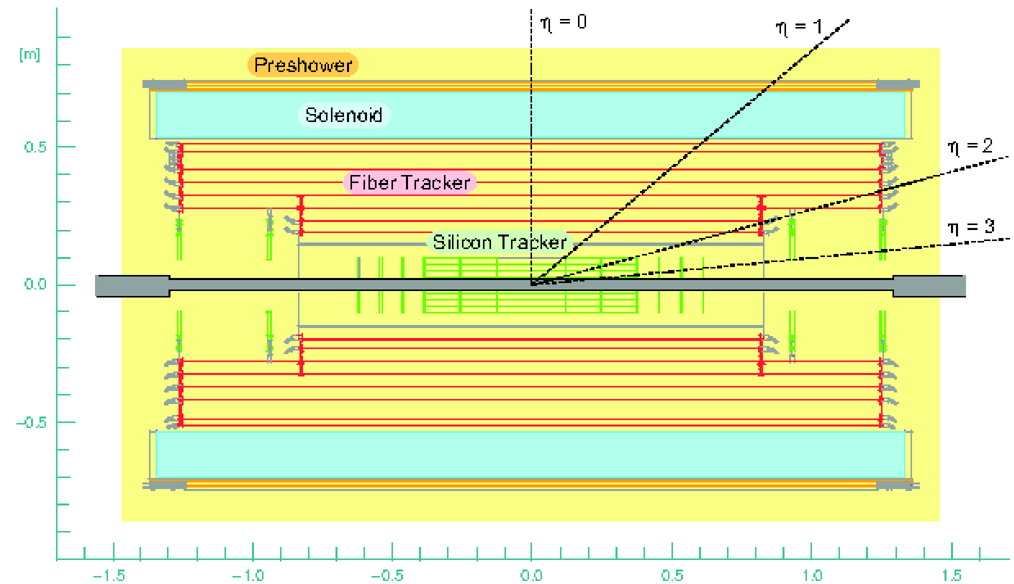
- Only way to measure $|V_{tb}|$ directly
  and with little assumptions

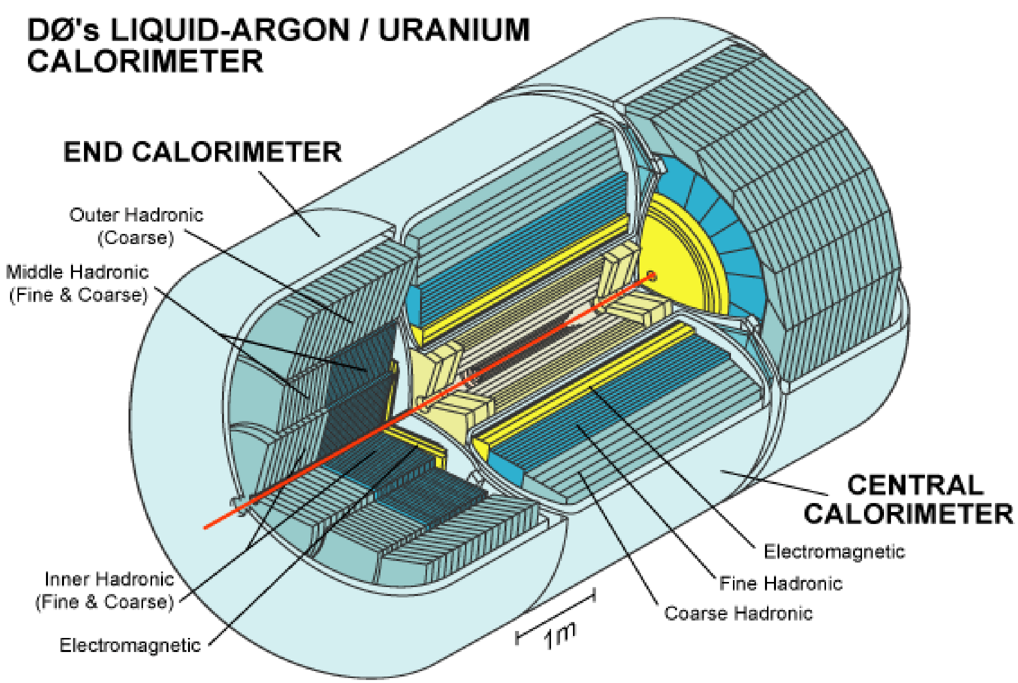
When Trish discovers Ned works exclusively with top quarks, she will be putty in his hands.

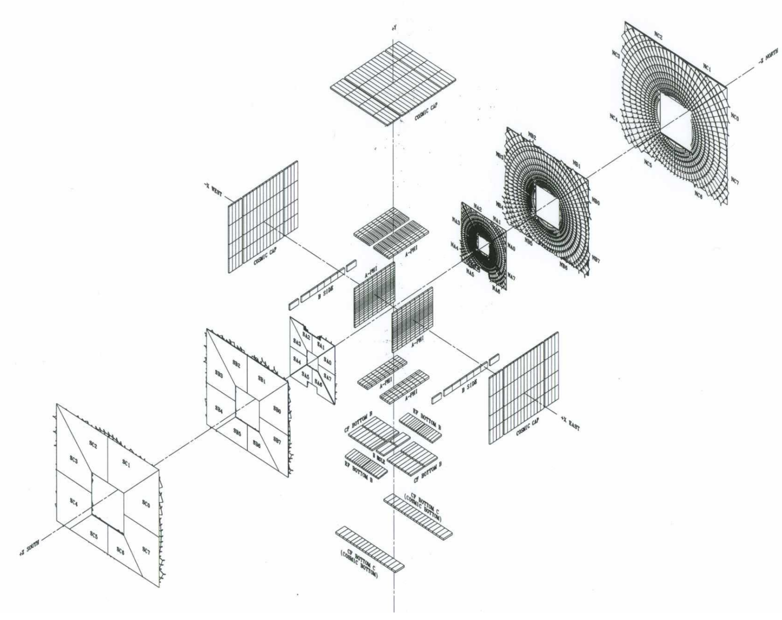
- Because it is sexy, obviously! :)

Just quickly: The D0 Detector



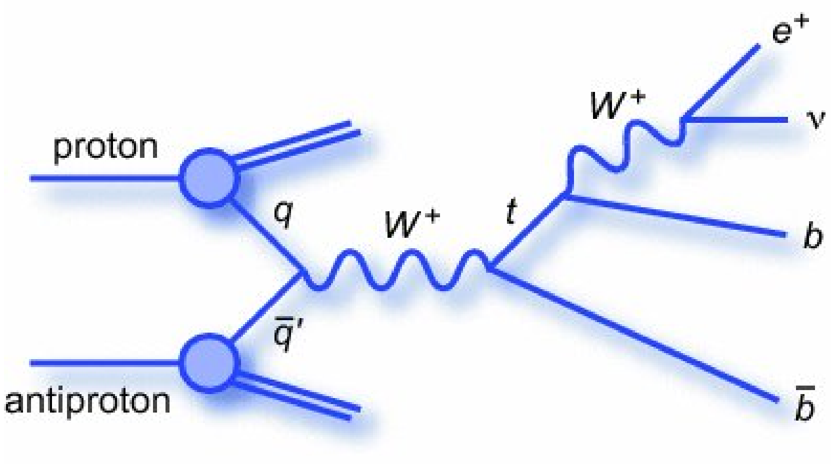






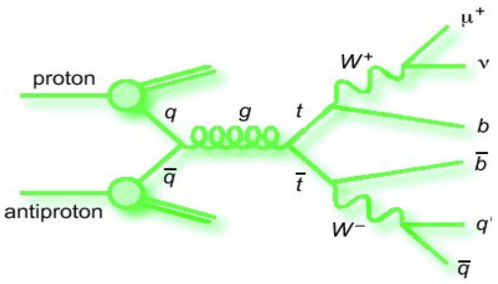
Why is it so difficult?

- **First**, what is signature in detector => what is background?



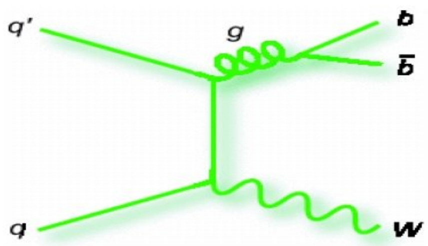
- one isolated lepton
- $\text{miss}E_T$
- 2-3 jets
- at least one b tags

Typical background:



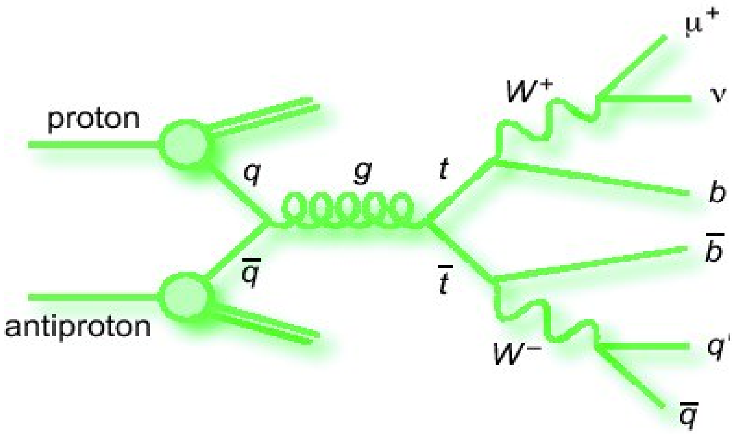
ttbar: x-section: ~ 6.5 pb

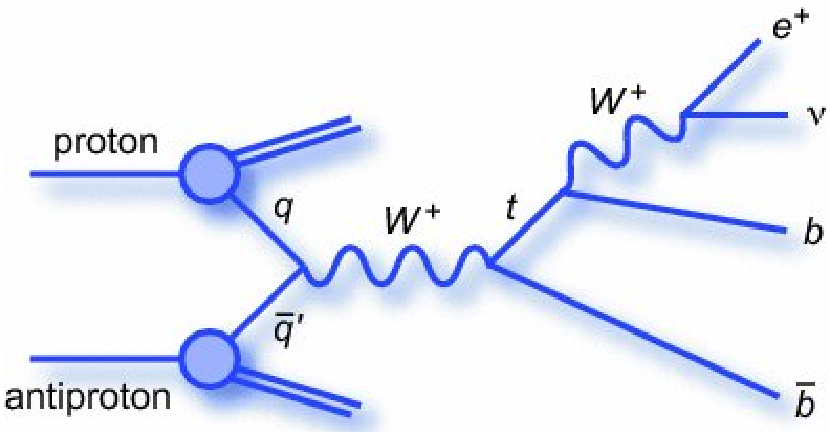
- **QCD bkg**



W+jets: x-section: ~ 80 - 200 pb

Why is it so difficult?





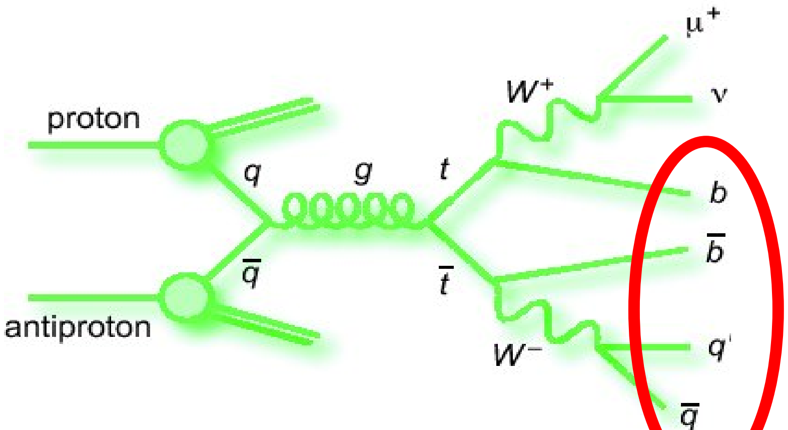
- cross section: ~6.5 pb

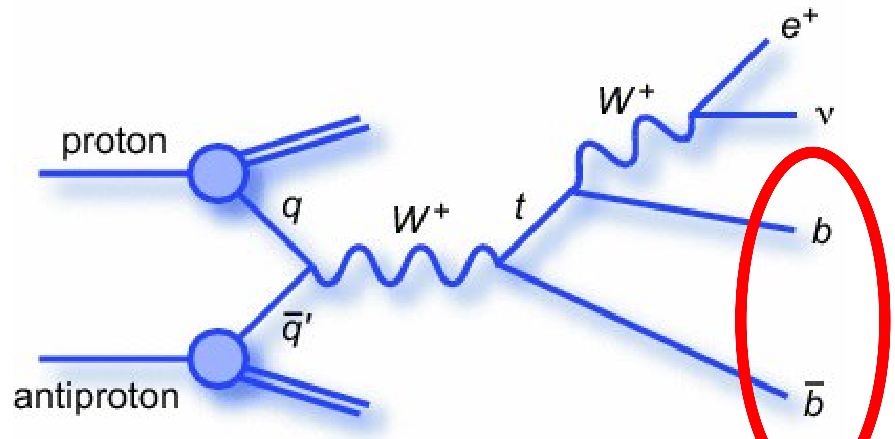
first observation, ~**60 pb^{-1}** data in 1995

- cross section: ~3 pb

first evidence with dataset of ~**900 pb^{-1}** in 2006

Why is it so difficult?



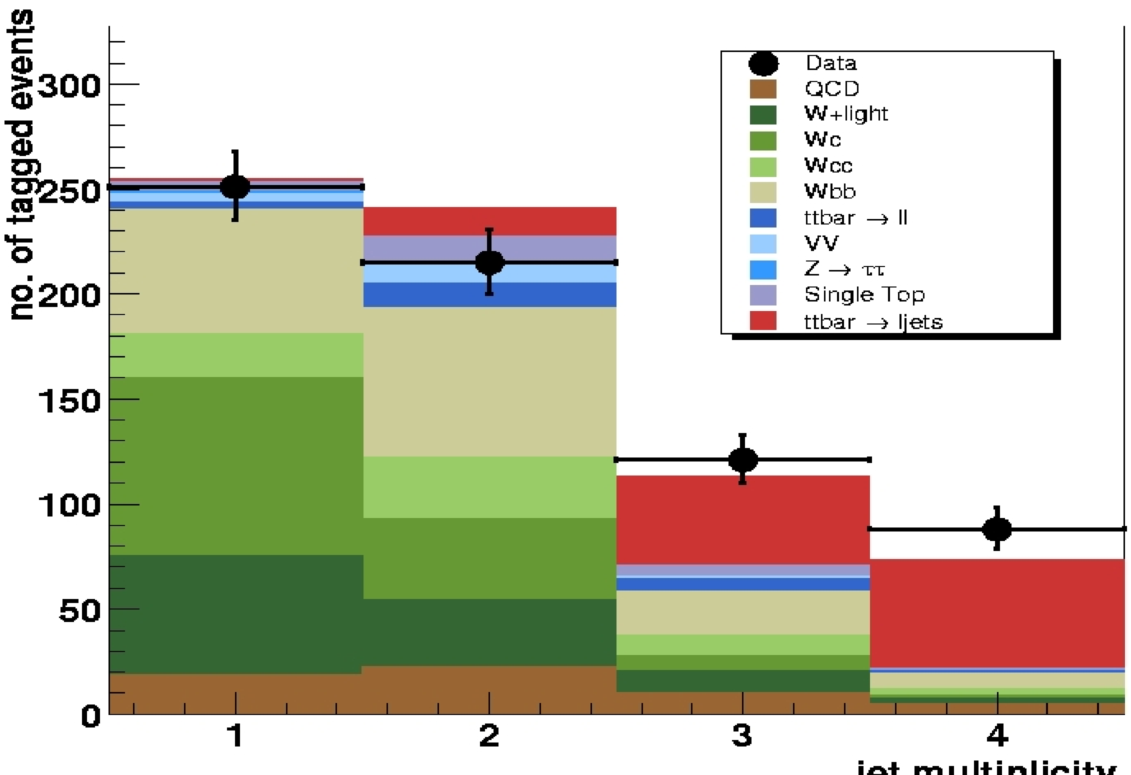


- cross section: ~6.5 pb

first observation, ~**60 pb^{-1}** data in 1995

- cross section: ~3 pb

first evidence with dataset of ~**900 pb^{-1}** in 2006



Ok, how did they do it?

3 different multivariate analyses:

- Boosted Decision Trees (DT)
- Bayesian Neural Network (BNN)
- Matrix Element Analysis (ME)

- all yield a discriminant: **$D(x) = p(x|S) / (p(x|S) + p(x|B))$**
- Each analysis based on different numerical method to approximate the D(x)

Ok, how did they do it?

3 different multivariate analyses:

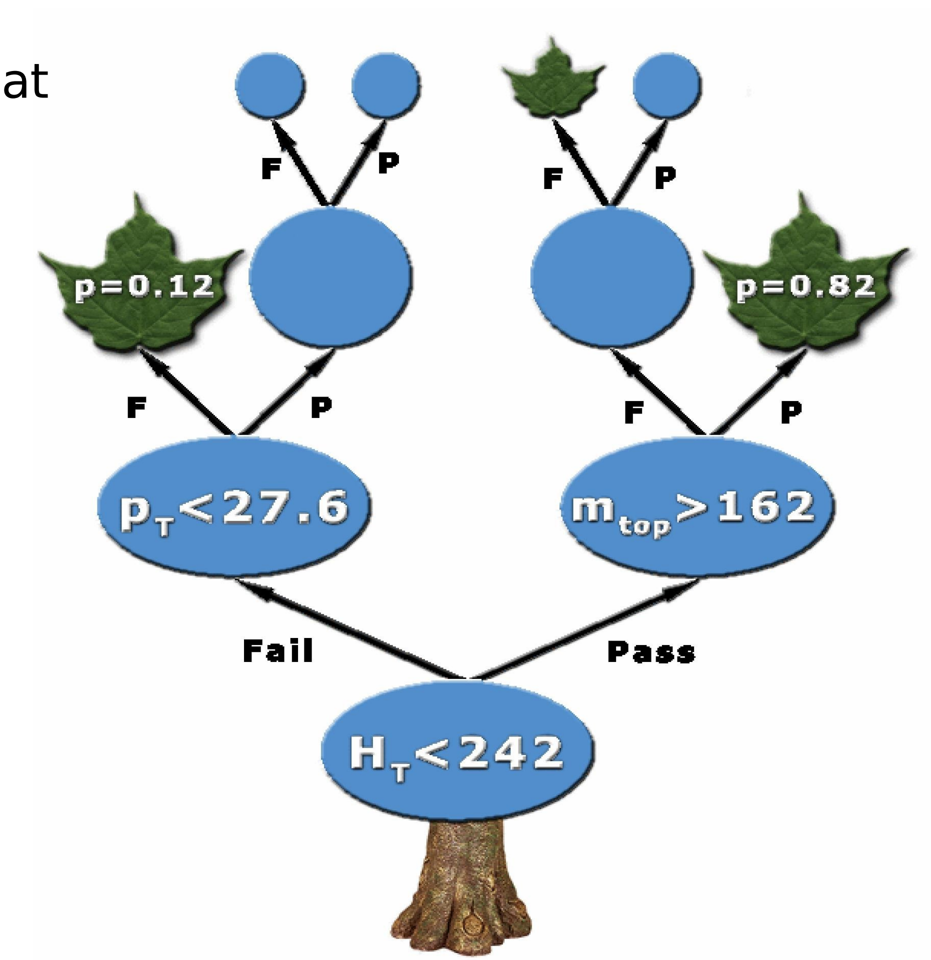
- Boosted Decision Trees (DT)
- Bayesian Neural Network (BNN)
- Matrix Element Analysis (ME)

- all yield a discriminant: $\mathbf{D(x) = p(x|S) \,/\, (\, p(x|S) + p(x|B)\,)}$
- Each analysis based on different numerical method to approximate the D(x)

Ok Boosted DTs, wth?

- A decision trees employs a machine-learning technique that extends a cut based analysis into a multivariate algorithm

- Boosting is a process that can be used on any weak classifier (defined as any classifier that does little better than random guessing)



- DT is created by creating 2 branches at each non terminal node

- Terminal nodes called leaves

- Each leaf has a purity value p (= D)

$$\mathbf{p = s / (s + b)}$$

s = sum w_i(signal)

b = sum w_i(background)

(Dis)Advantages of DTs

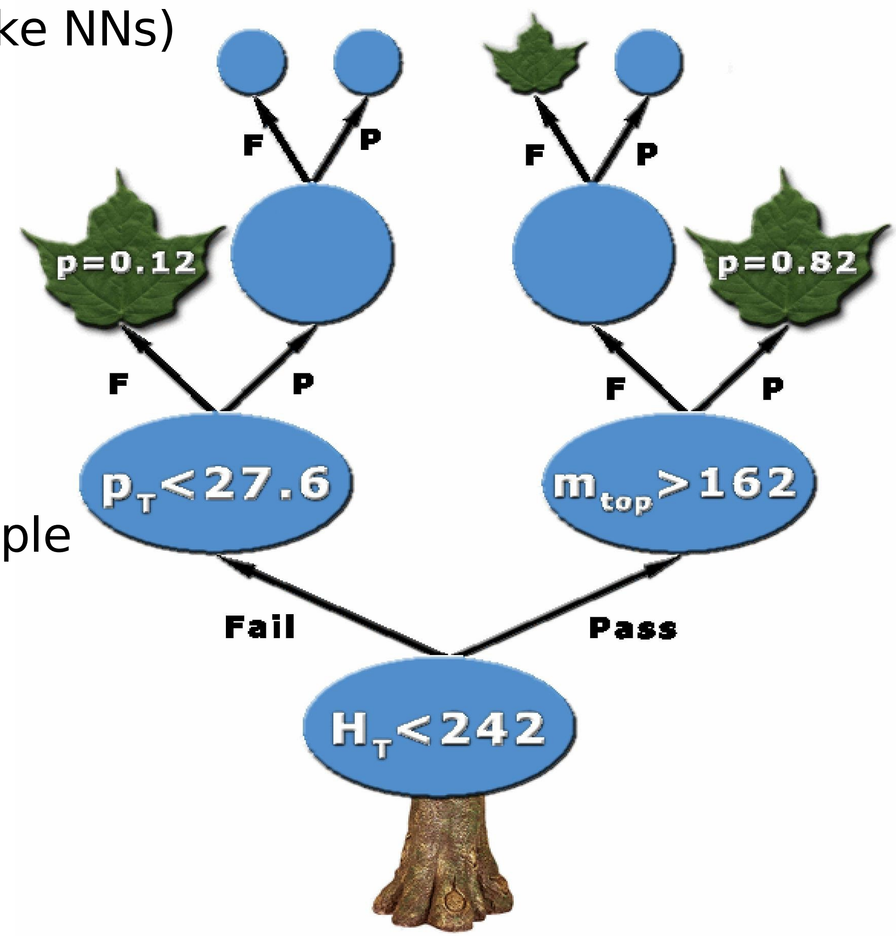
Pros:

- Events which fail a cut are still considered in Analysis!

- Tree is human readable! => know why event is called Bkg or Signal

- DT is insensitive to extra varibales! (unlike NNs)

- Training is fast!



Cons:

- Tree depends extremely on training sample
  take care what you do!

- Discrete discriminant,
  since number of leaves is finite

Training of DTs

1. Normalize signal trainings sample to background trainings sample
   i.e. $\mathrm{Sum}(w_S) = \mathrm{Sum}(W_B)$

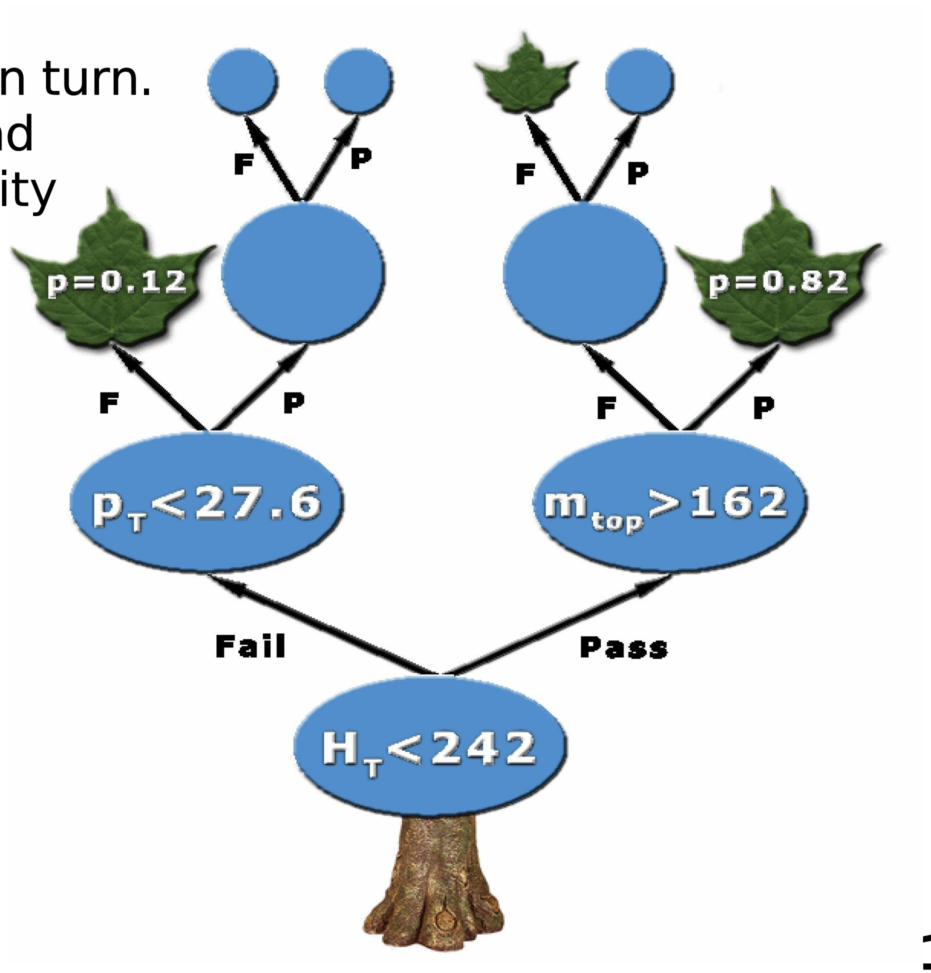
2. Create first node containg full sample

3. Sort events according to each variable in turn.
   best splitting value is found
   i.e. get highest/lowest purity

4. Sample is devided into 2 sub-samples

5. If statistics are too low,
   node becomes a leaf

6. Repeat 3-5



How to split...

Goal: find split S that:
maximizes purity
or maximizes decrease of impurity

Splitting: some measure i of impurity at node t

- should be max for equal mix of s and b
- should be min for either only s or b
- symmetric in s and b
- concave to always reward purer nodes

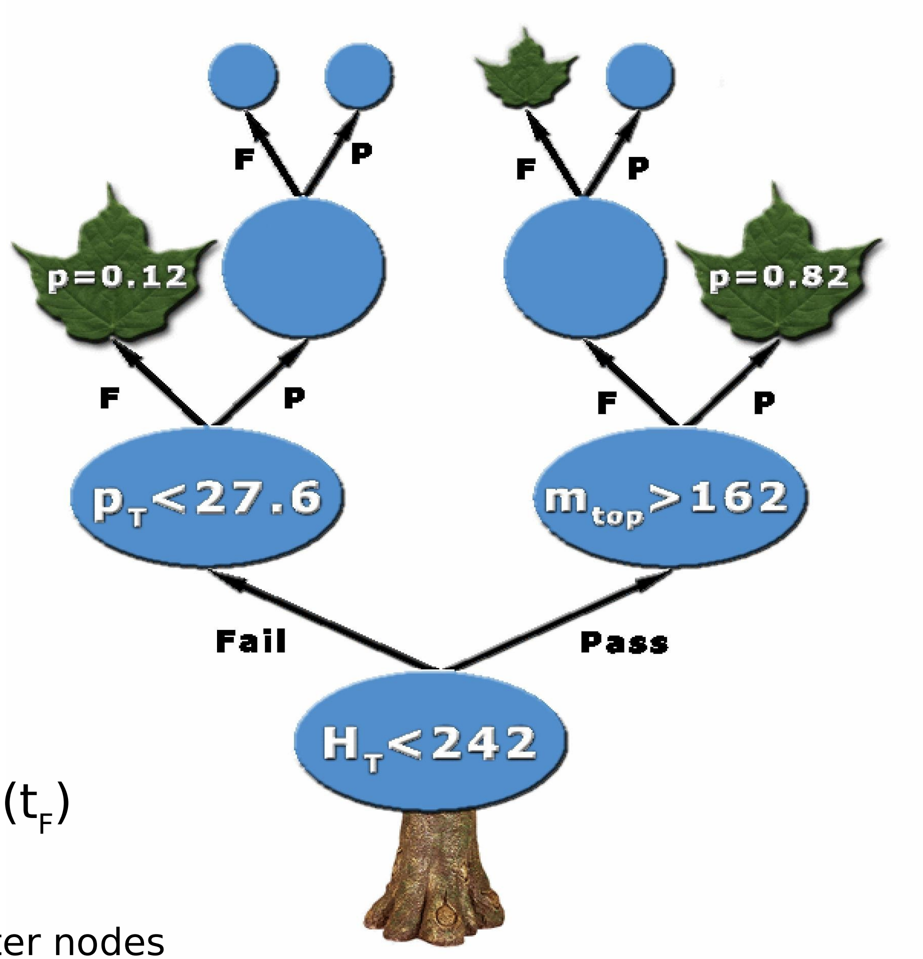
=> Gini index (measure of impurity):

$$i_{Gini} = 2*sb/(s+b)^2$$

decrease of **impurity**:

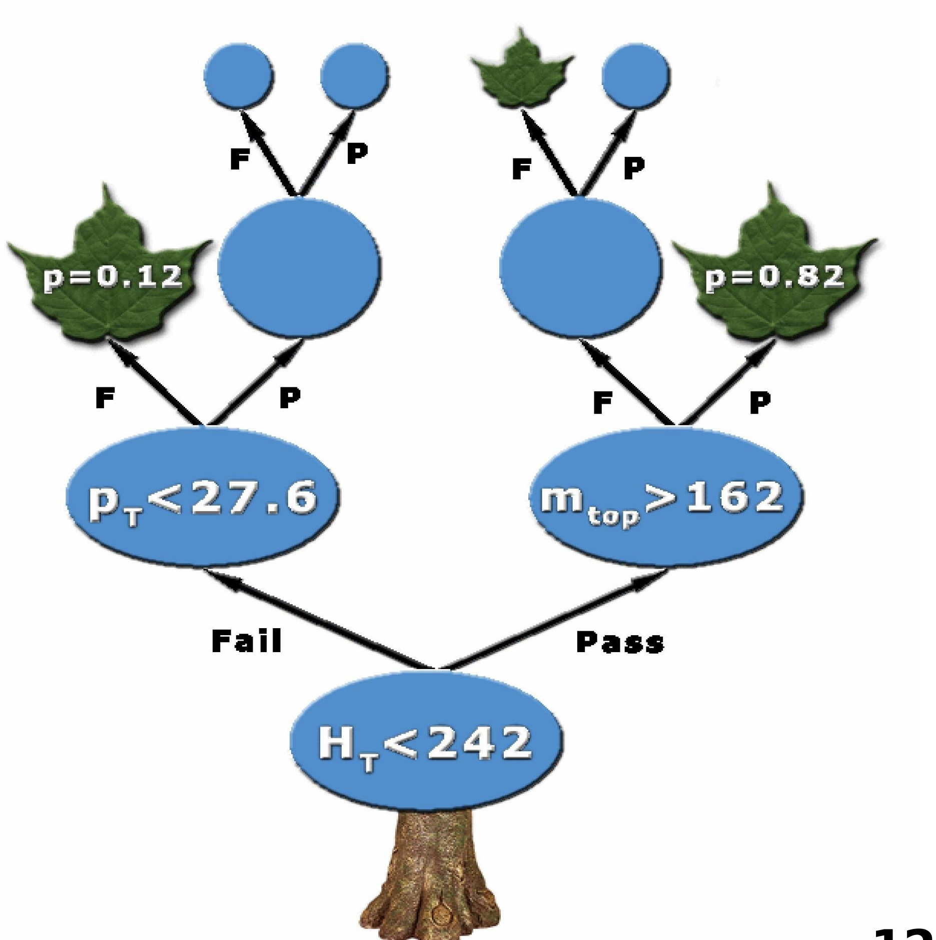
$$\text{Delta } i_{Gini}(S,t) = i_{Gini}(t) - p_P\, i_{Gini}(t_P) - p_F\, i_{Gini}(t_F)$$

$\mathbf{t}_{P/F}$ daughter node, $\mathbf{p}_{P/F}$ fractions of events in daughter nodes



...and boost

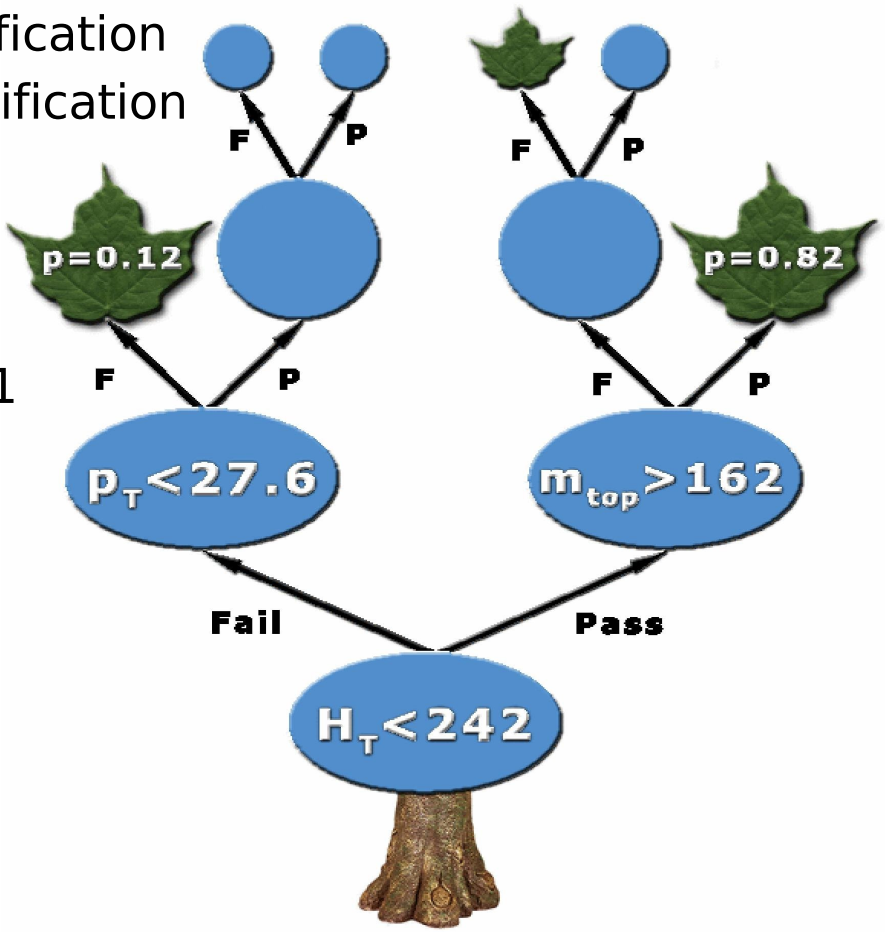
- let one tree run
- reweight events
- make new tree with new weights
- new tree will work “harder“ on missclassified events
- calculate discrimnant



...and boost (a bit complicated...)

- x_i set of PID variables
- y_i = 1 if signal event
  = 0 if bkg event
- initial weight of each event is 1/N
- $T_m(x_i)$ = -1 if event on bkg leaf
  = 1 if event on signal leaf

$I(y_i, T_m(x_i))$ = 1 if $y_i \neq T_m(x_i)$ wrong classification
= 0 if $y_i = T_m(x_i)$ correct classification



- define error function

$$err_m = \frac{\sum_i w_i * I(y_i, T_m(X_i))}{\sum_i w_i}$$

$0 \le err_m \le 1$

- define “boost“ function

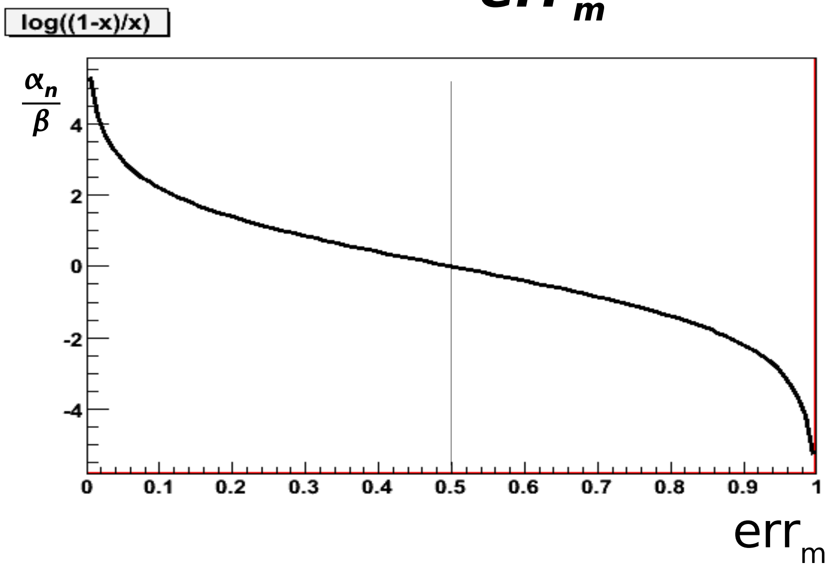
$$\alpha_n = \beta * \ln\left(\frac{1 - err_m}{err_m}\right)$$

...and boost (a bit complicated...)

$$err_m = \frac{\sum_i w_i * I(y_i, T_m(X_i))}{\sum_i w_i}$$

if better than random: $err_m < 0.5$

$$\alpha_n = \beta * \ln\left(\frac{1 - err_m}{err_m}\right)$$



- each missclassified event gets new weight according to $alpha_m$

$$w_i \rightarrow w_i * e^{\alpha_n * I(y_i, T_m(X_i))}$$

- create new tree with new weights of events
- each tree tries to optimize purity in leaves through differences in Gini index

$$i_{Gini} = 2*sb/(s+b)^2$$

s = sum w_i(signal)

b = sum w_i(background)

- since each misclassified event gets bigger weight, next tree tries harder to classify this event correctly

Redefine Discriminant: $$D(x_i) = \frac{1}{\sum_n \alpha_n} \sum_n \alpha_n D_n(x_i)$$

What is the input?

3 classes of input variables: Overall 49 (!) variables

Object kinematics:

p_T(jet1), p_T(jet2), p_T(tag1), p_T(l), ...

Angluar variables:

cos(jet1,l)$_{lab}$, delR(jet1,jet2), cos(jet2,alljets)$_{alljets}$,...

Event kinematics:

missE$_T$,M$_T$(W),M(alljets),H(alljets), Centrality,...

What is the input?

3 classes of input variables: Overall 49 (!) variables

Object kinematics:

p_T(jet1), p_T(jet2), p_T(tag1), p_T(l), ...

Angluar variables:

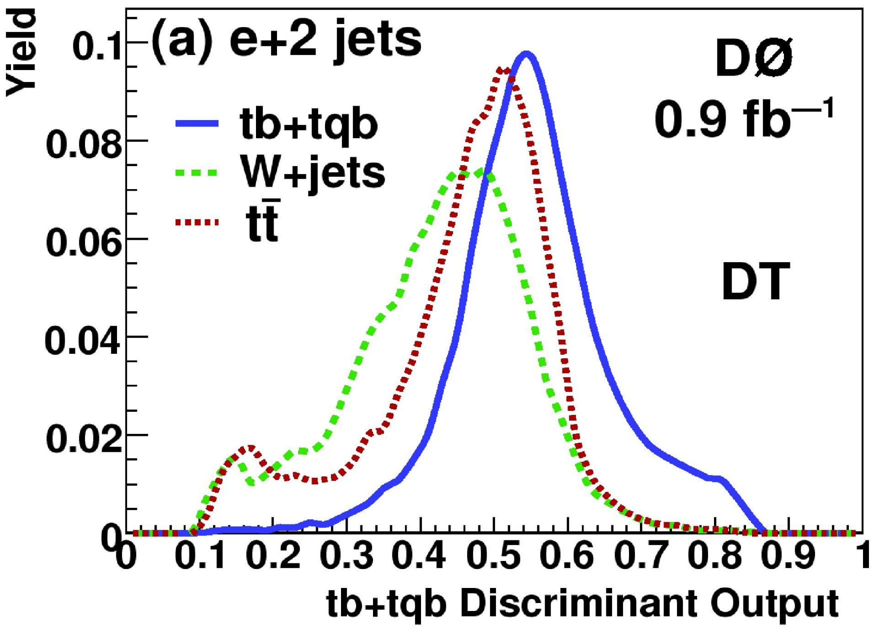
$\cos(\text{jet1,l})_{\text{lab}}$, delR(jet1,jet2), $\cos(\text{jet2,alljets})_{\text{alljets}}$,...

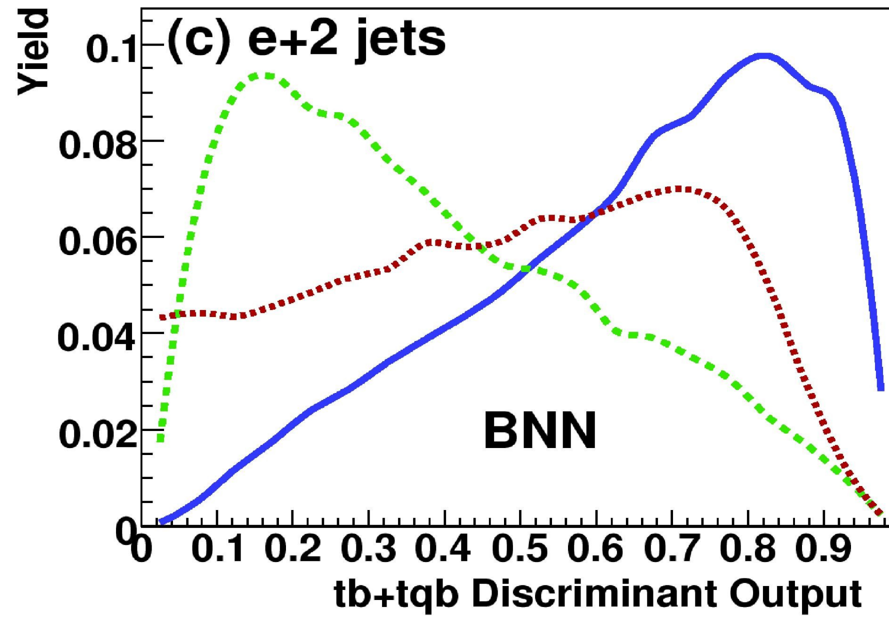
Event kinematics:

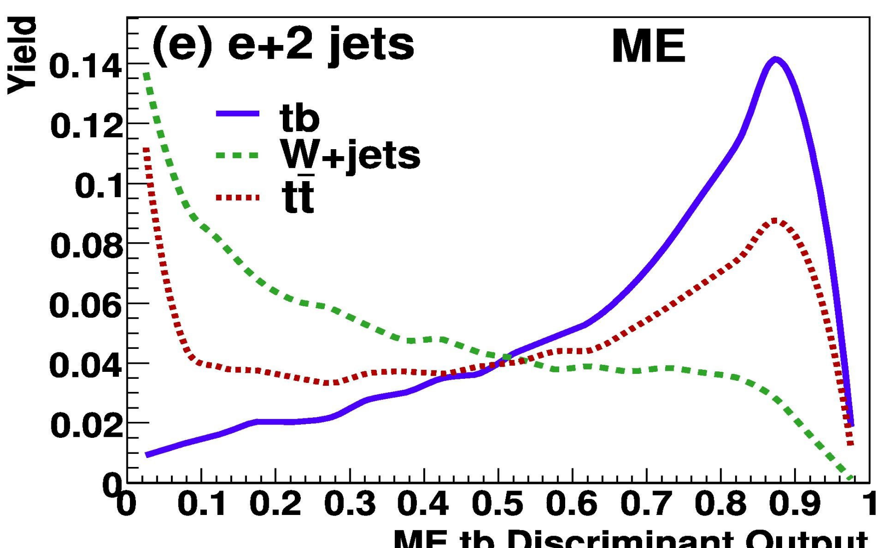
missE$_T$,M$_T$(W),M(alljets),H(alljets), Centrality,...

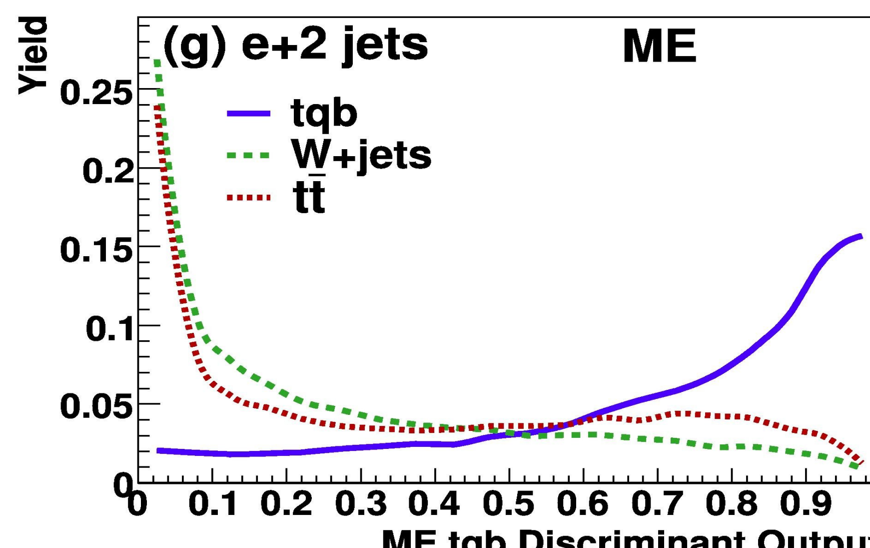
- Why use so many variables if ~ handfull of right variables give you **ALL** information?
  - **True** there is no more information in derived var. then in fundamental ones
  - **BUT** for some **numerical approx**. methods it is easier to provide an accurate D(x) if use constructed variables

And what is the output?

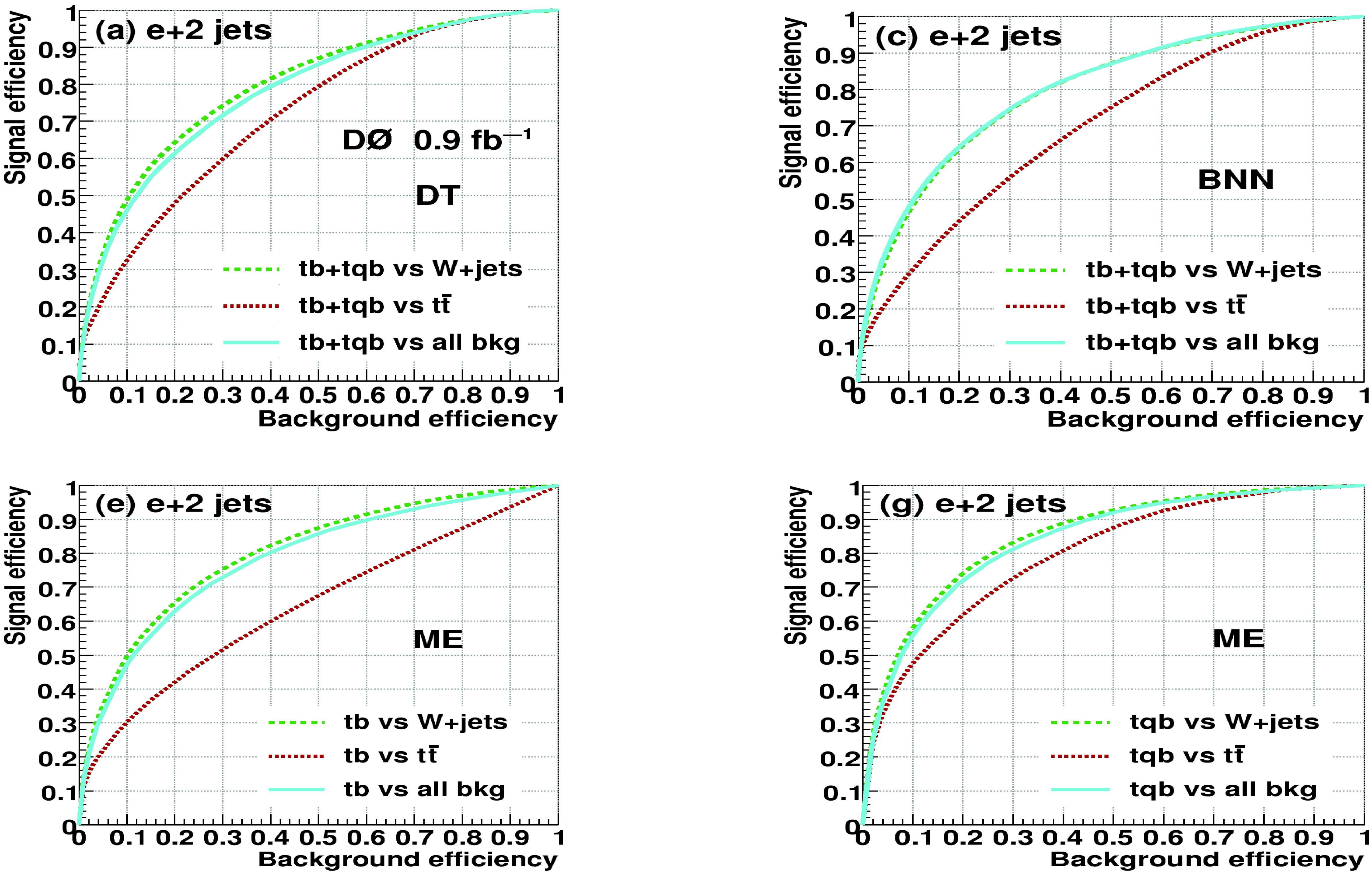




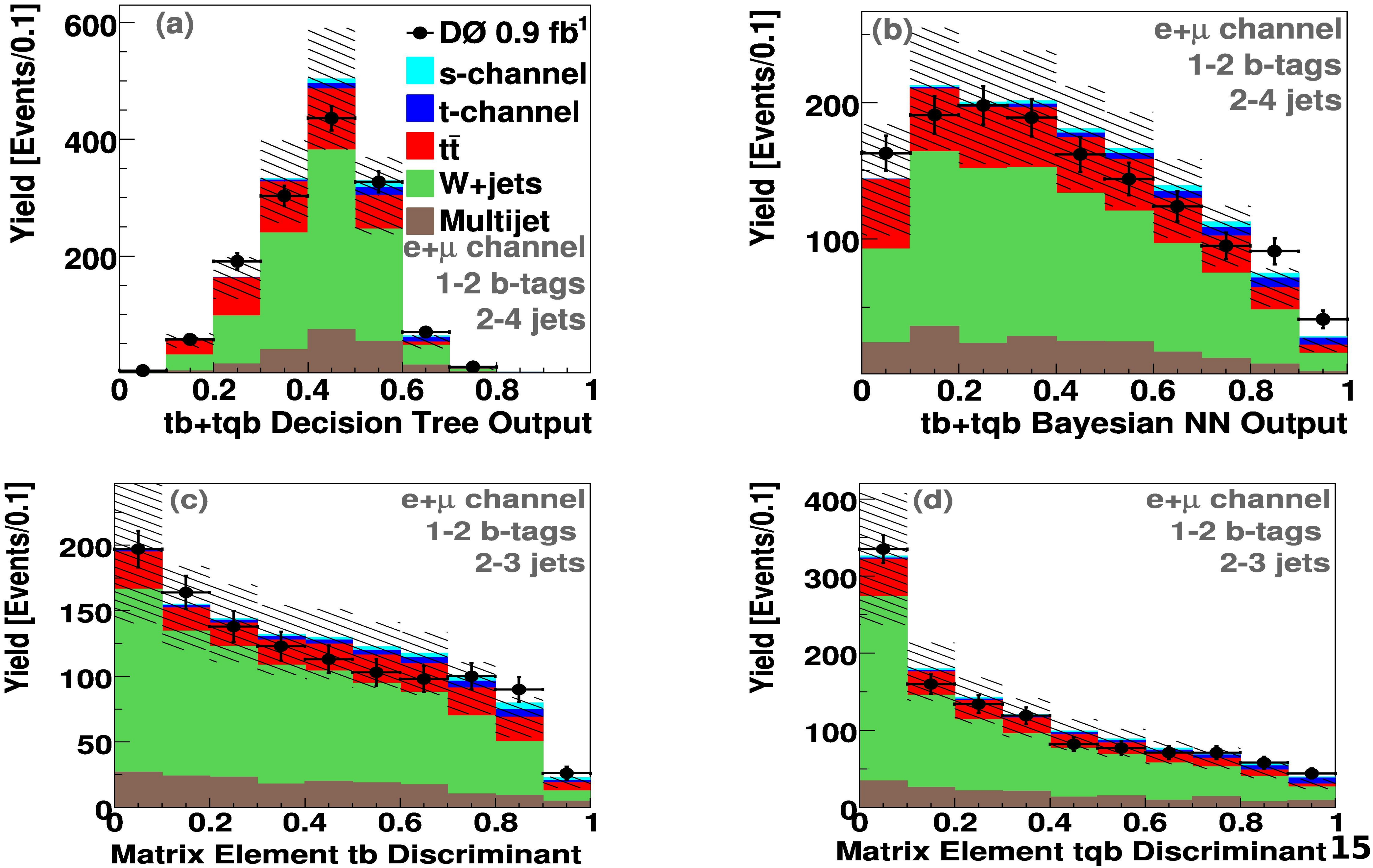




And what is the output?



How does it compare to data?



Finally, Results!

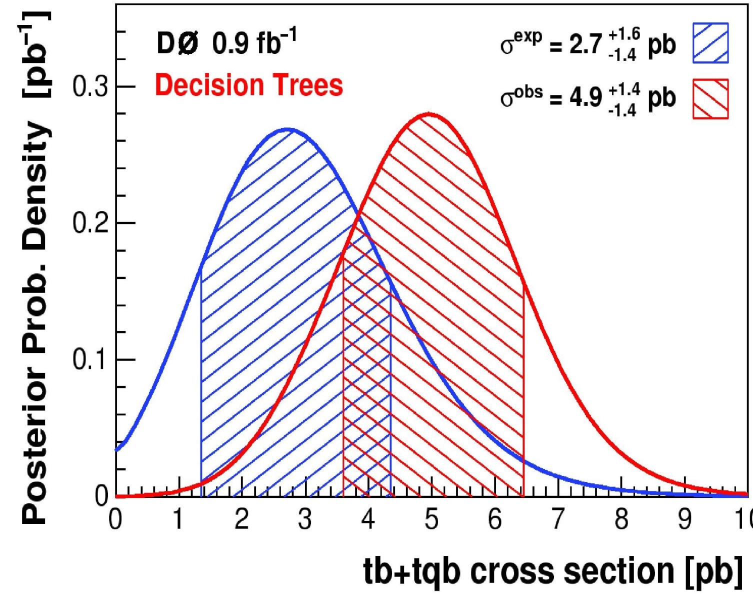
Applying a Bayesian Analysis:

- find *expected cross section* by setting N_{data} = (theoretical) $N_{Sig} + N_{Bkg}$

  - sigma(exp) = $2.7^{+1.6}_{-1.4}$ pb (DT)

    $2.7^{+1.5}_{-1.5}$ pb (BNN)

    $2.8^{+1.6}_{-1.4}$ pb (ME)

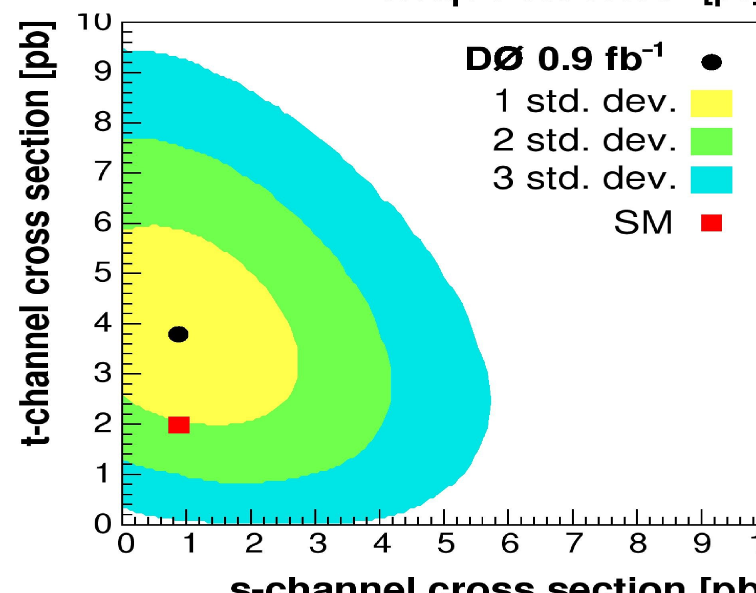


- *measured (observed) cross section* :

  - sigma(obs) = $4.9^{+1.4}_{-1.4}$ pb (DT)

    $4.4^{+1.6}_{-1.4}$ pb (BNN)

    $4.8^{+1.6}_{-1.4}$ pb (ME)



Maybe its all only background?

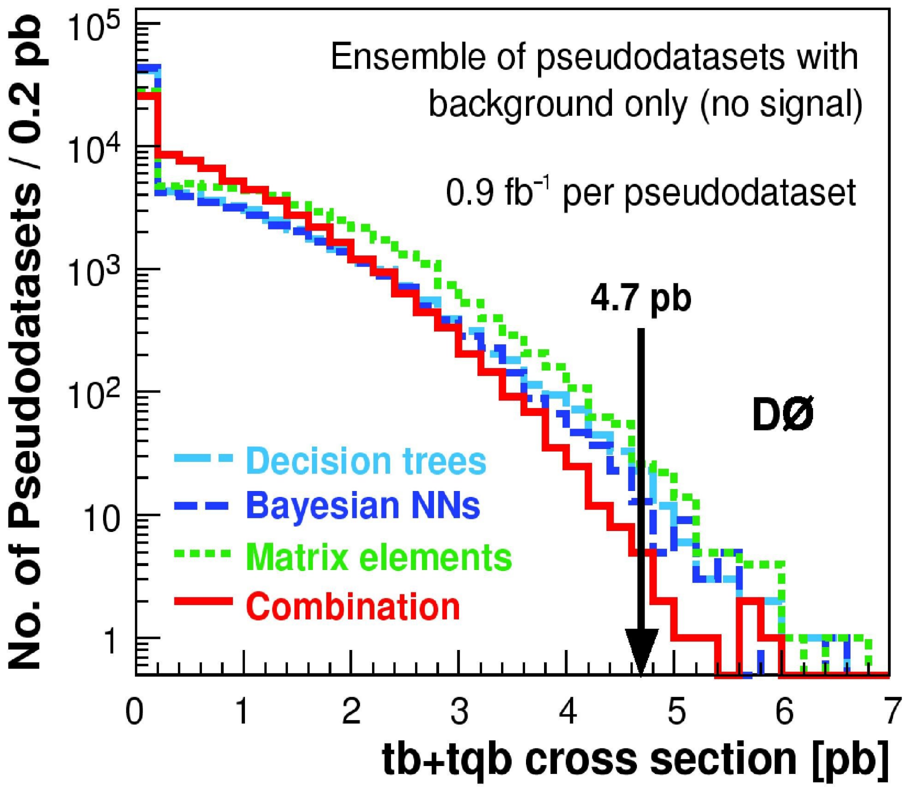
- find out the **significance** of the expected / observed x-section with **Ensemble-Tests**
  - use background only MC ensemble with all uncertainties
  - set background to estimated yield value
  - run analysis and count how often background only yields exp/obs x-section
    - calculate p-value => significance

Expected Results

Analysis	Expected cross section [pb]	Expected p-value	Expected significance (std. dev.)
DT	2.7	0.018	2.1
BNN	2.7	0.016	2.2
ME	2.8	0.031	1.9
Combined	2.8	0.011	2.3

Observed Results

Analysis	Measured cross section [pb]	Measured p-value	Measured significance (std. dev.)
DT	4.9	0.00037	3.4
BNN	4.4	0.00083	3.1
ME	4.8	0.00082	3.2
Combined	4.7	0.00014	3.6



Defines EVIDENCE, more than 3 sigma!

Is it compatible with signal?

- find out with **Ensemble-Tests**
    - use background and signal MC ensemble with all uncertainties
    - set background and signal to estimated yield value
    - run analysis and count how often ensemble yields observed x-section
    - In ~ 10% of all cases the observed x-section is seen
      => It is **compatible** with background + signal hypothesis
    - => D0 was lucky

I remember something about $|V_{\mathbf{tb}}|$

Reminder:

- **with** 3 generations and unitarity: $|V_{\mathbf{tb}}| = 0.999100\ ^{+0.000034}_{-0.000004}$
- **without** 3 generations and **without** unitarity: $0.07 <= |V_{\mathbf{tb}}| <= 0.993$

- to extract $|V_{\mathbf{tb}}|$ is straight forward, since sigma ~ $|V_{\mathbf{tb}}|^2$

  assumptions:

  - $|V_{td}|^2 + |V_{ts}|^2 << |V_{tb}|^2$ ok since measured before
  - Vertex Wtb is V-A type, but allowed for anomalous strength $f_1^L > 1$!

    $$|V_{tb} * f_1^L| = 1.31^{+0.25}_{-0.21}$$

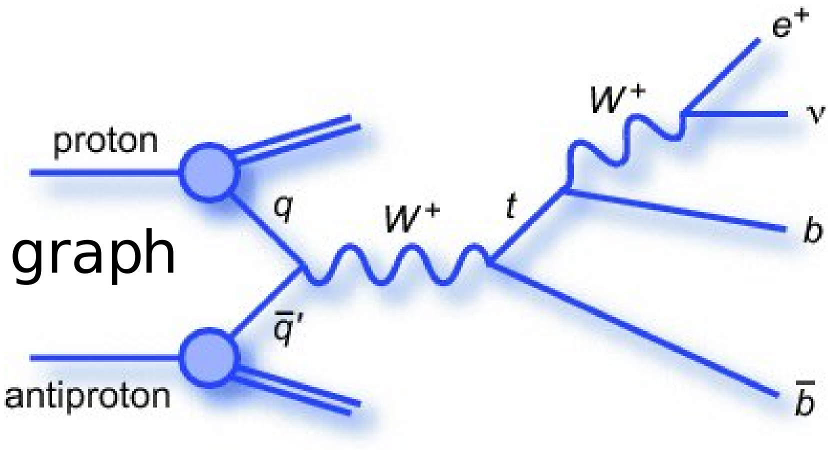
  - Vertex Wtb is V-A type, but $f_1^L = 1$!

    $$|V_{tb}| = 1.00^{+0.00}_{-0.12}$$

    95% C.L. lower limit: $|V_{tb}| > 0.68$

Summary

- This is a “Single Top“ feynman graph



- There is **evidence** for ST with 3.6 sigma significance
- The **observed** x-section is 4.7+- 1.3 pb (combined)
- First direct measurements of $|V_{tb}|$
  - $|V_{tb} f_1^L| = 1.31^{+0.25}_{-0.21}$
  - $0.68 < |V_{tb}| <= 1$ at 95% C.L.
- AFTER analysis was done, new theoretical x-section published: x-sect = 3.21 +- 0.21

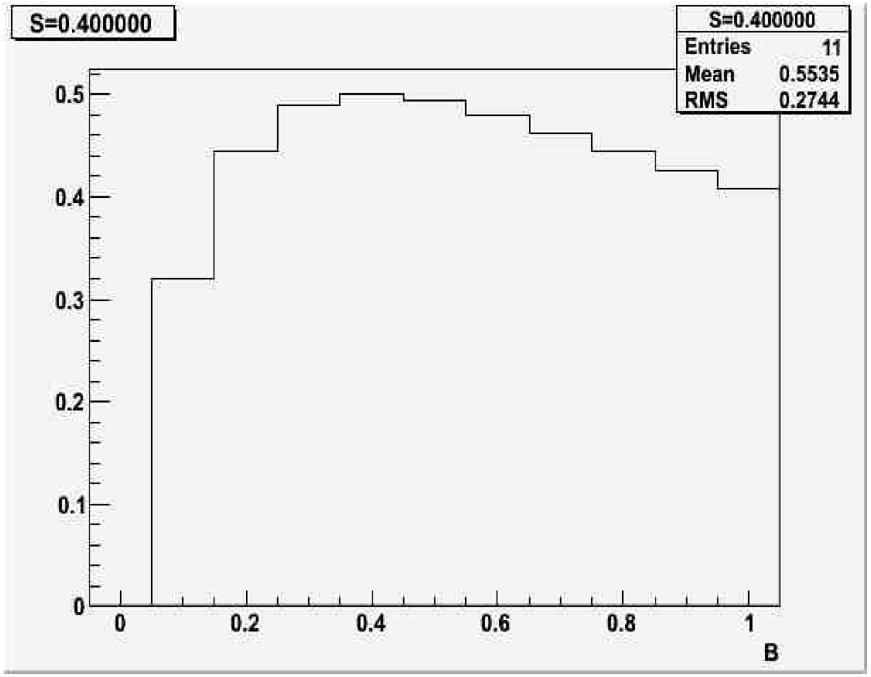
Lunch!

Enjoy your meal !

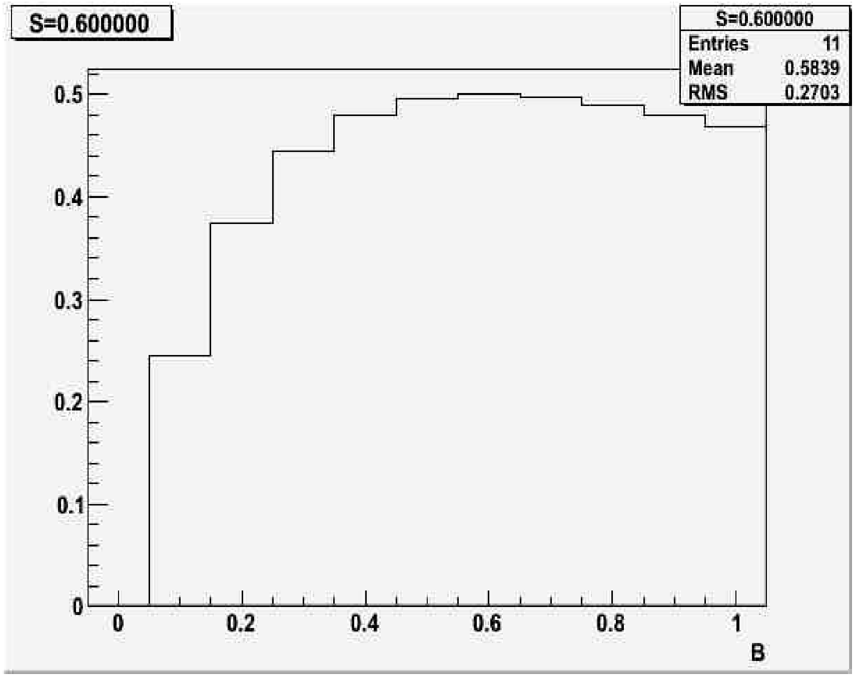
Backup

Gini index: $i_{Gini} = 2*sb/\ (s+b)^2$

s fixed and b varied [0,1]



s = 0.4



s = 0.6

Gini index max. for same amount signal and bkg