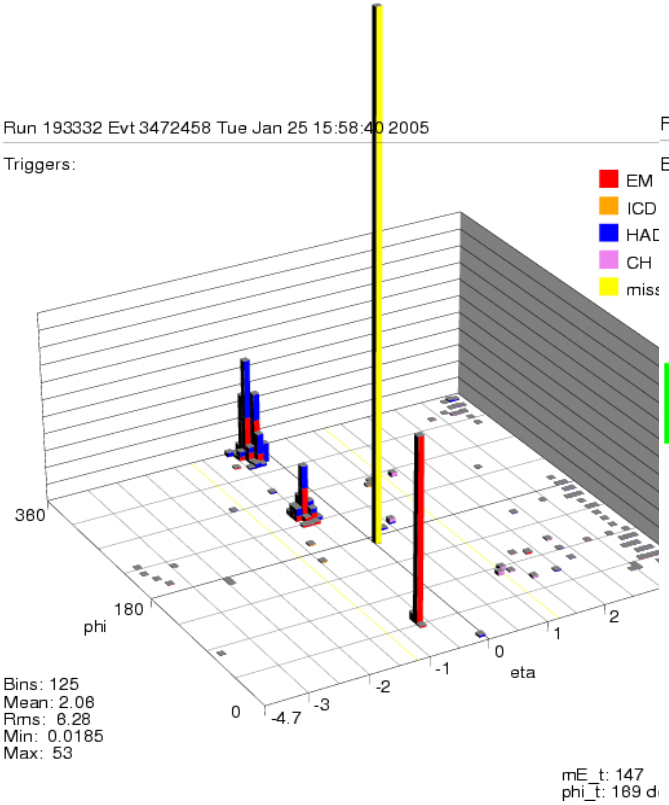
Top Workshop IN2P3 : from Dzero to ATLAS

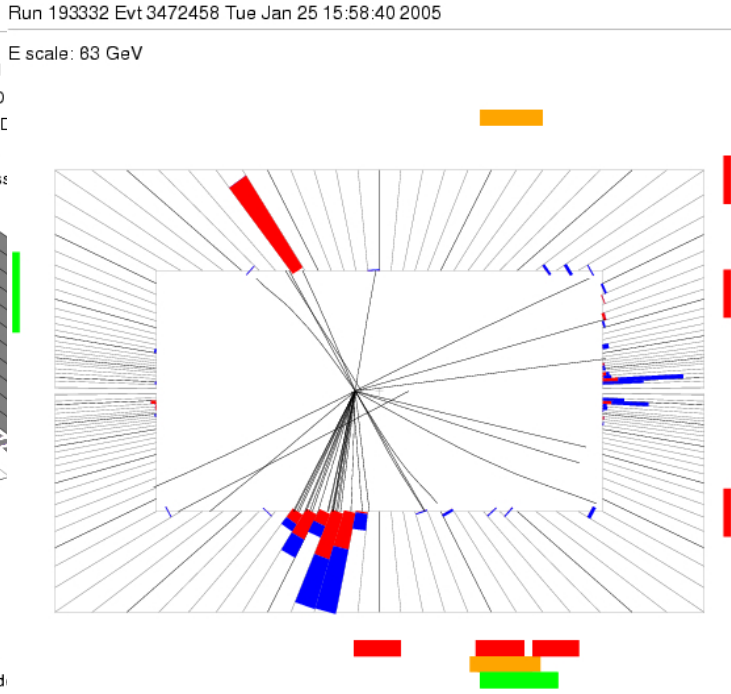
Grenoble, October 9-10, 2006

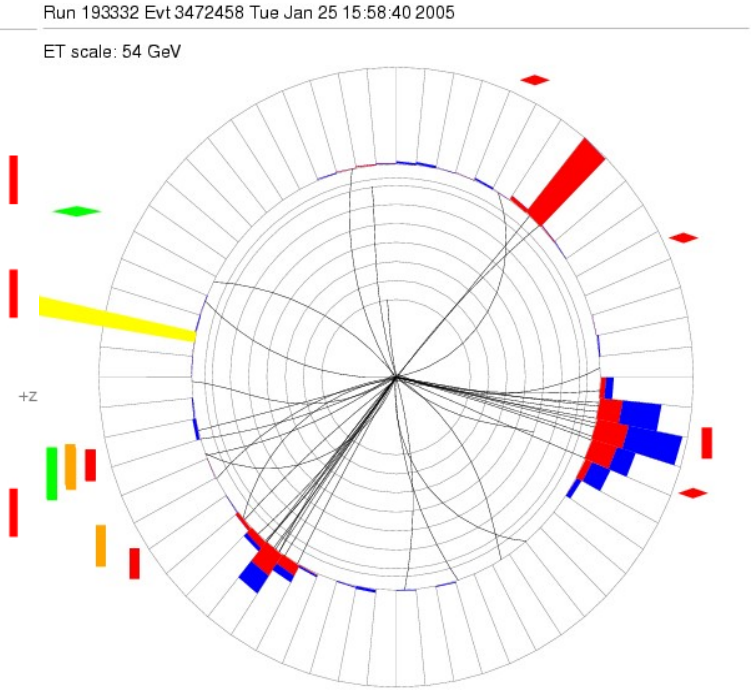
Measurements of the Top Quark Pair Production Cross Section in the Dilepton and All-Jets Final States

Viatcheslav Sharyy

D0 collaboration

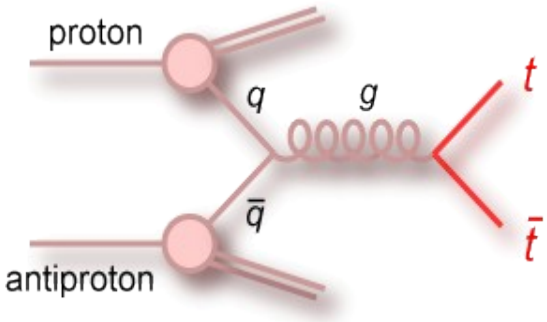






Top Pair Production at Tevatron

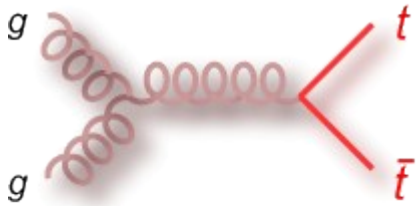
- Two production modes
  - qq annihilation 85%, gluon fusion 15%
  - LHC: 10% / 90%

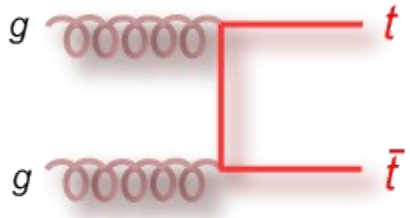


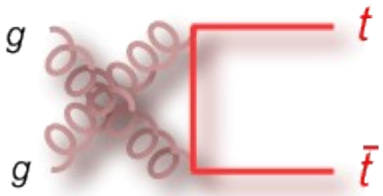
- Cross-section

$$\sigma(s, m_t^2) = \sum_{i,j} \int dx_i dx_j \; f_i(x_i, \mu^2) f_j(x_j, \mu^2) \; \hat{\sigma}_{ij}[\hat{s}, m_t, \mu^2]$$

  - i,j are the input partons index
  - $f(x, \mu^2)$ is a parton density function (PDF)
  - Factorization and renormalization scale are usually chosen to be the same $\mu \sim m_t$. To estimate systematics uncertainties the $0.5m_t$ and $2m_t$ scales usually are used







Cross-section

- The NLO+NLL calculation yield ~ 7 pb, e.g. *BCMN hep-ph/9801375, updated in Cacciari et al. Hep-ph/0303085 (CTEQ6, MRST):*

$6.8^{+0.6}_{-0.7}$ @ 1.98 TeV, m_t = 175 GeV

(~830 @14 TeV)

- *Calculated* cross-section has a top mass dependence and the *measured* cross-section depends on top mass due to the acceptance. The SM prediction on the cross-section - mass dependence could be tested (but no mass measurements via the cross-section).

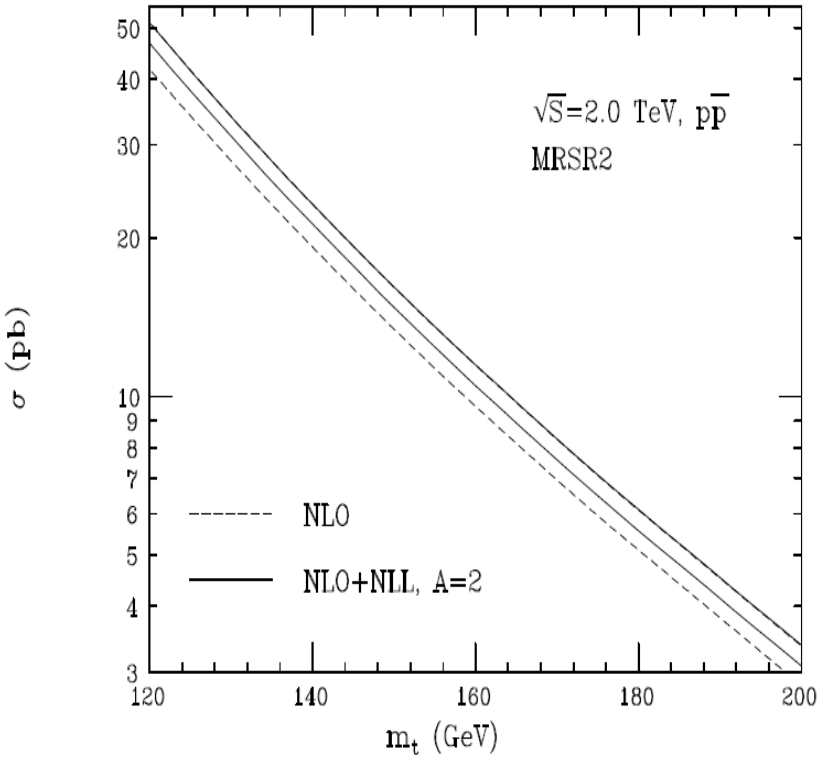
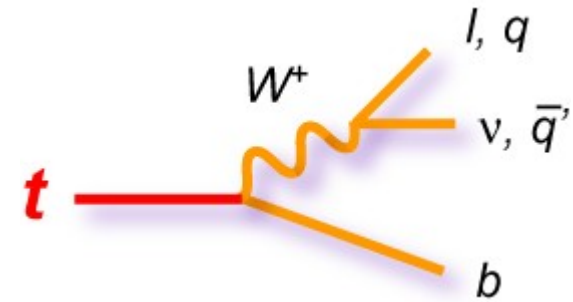


Figure 12: Total $t\bar{t}$ production cross-section in $p\bar{p}$ collisions at $\sqrt{S}$ = 2.0 TeV, as a function of the top-quark mass. Dashed lines: NLO result; solid lines: NLO+NLL result. Upper lines: $\mu = m_t/2$; lower lines: $\mu = 2m_t$.

Decay modes

- In SM dominant 2-body decay t→Wb. Assuming unitarity of 3-generation CKM matrix: $|V_{tb}|$ = 0.9990-0.9992 @ 90% CL, B(t→Wb) ~ 100%



- Γ_t ~ 1.4 GeV @m_t = 175 GeV: top decays before top-flavored hadrons or tt-quarkonium bound states can form. Top quark spin efficiently transferred to the final state.

- Top final state signature:
  - All jets: 46%, huge background
  - Lepton +jets: 30% moderate background
  - Dileptons: ~5%, small background
  - Tau + X: ~14%, significant background

Top Pair Decay Channels

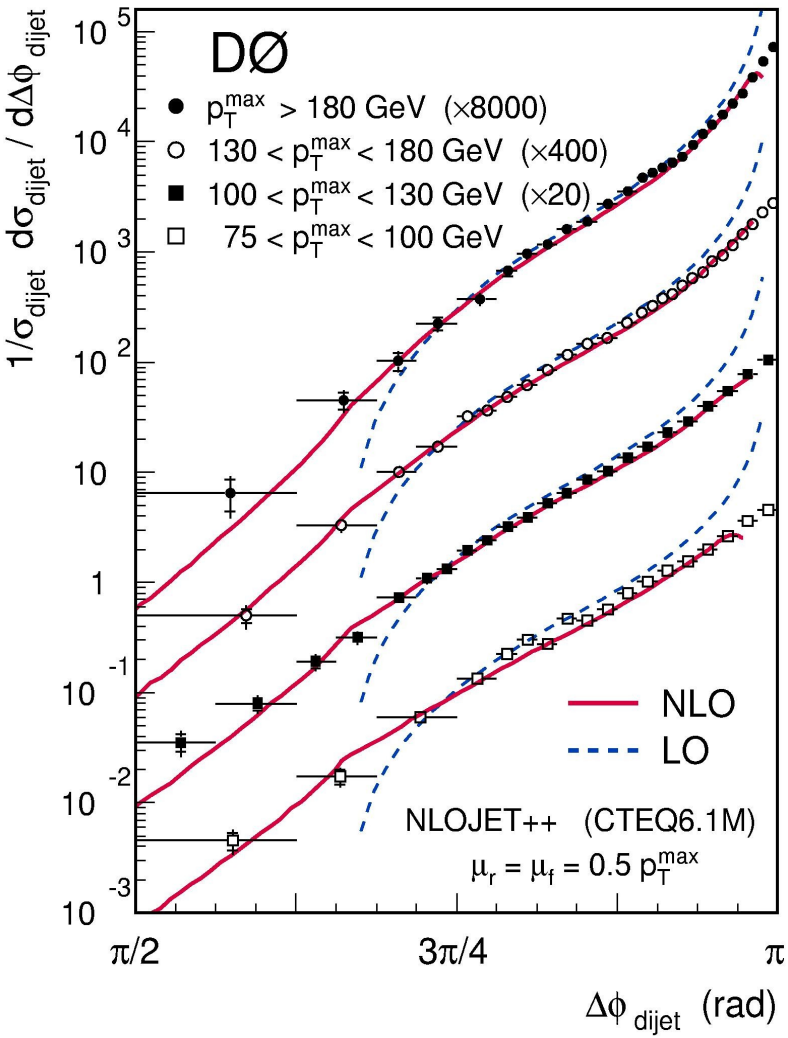
W decay	e^+	μ^+	τ^+	$u\bar{d}$	$c\bar{s}$
$\bar{c}s$	electron+jets	muon+jets	tau+jets	all-hadronic	
$\bar{u}d$					
τ^-	$e\tau$	$\mu\tau$	$\tau\tau$	tau+jets	
μ^-	$e\mu$	$\mu\mu$	$\mu\tau$	muon+jets	
e^-	ee	$e\mu$	$e\tau$	electron+jets	

dileptons

Monte Carlo

- Aplgen generator is used usually in the cross-section measurement. Fixed order results are not reliable in the divergent region (QCD radiative effects).
- Solutions:
  - Analytic resummation (collect divergent terms and sum to all orders) are not available for many processes.
  - Monte Carlo "resummation" : use a model to represent activity below a cut off (turned to data), e.g. Pythia together with fixed order generator, e.g. Alpgen

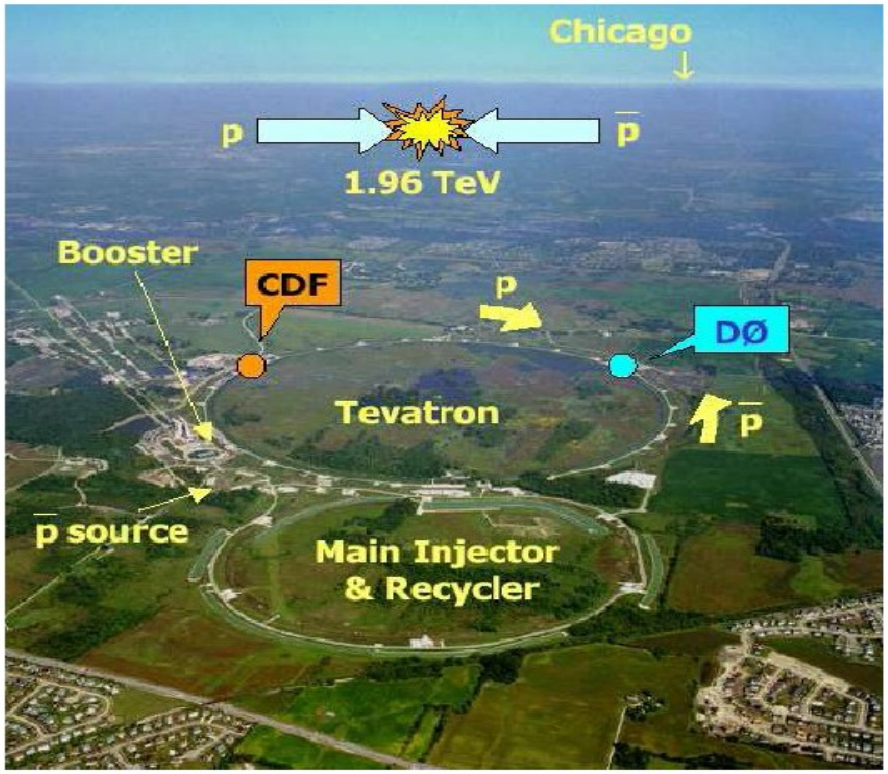
PRL 94, 221801 (2005)

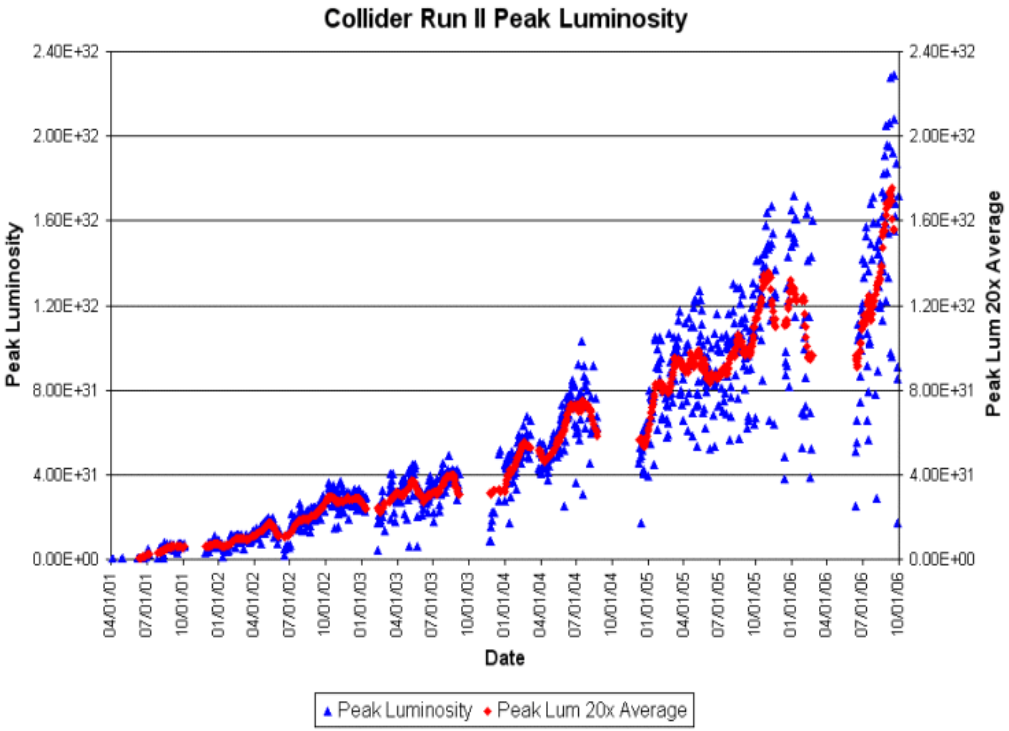


Monte Carlo (2)

- Double counting of contributions is an issue. Use existing scheme to solve this problem *(e.g. S. Hoeche et al., hep-ph/0602031).*
- MC generation.
  - Matrix element generation: generate samples with different parton multiplicities which fill the phase space except for the parton levels cuts used to avoid divergences.
  - Parton shower application: divergent phase space filled by the parton shower, however, multiplicity bins are not overlap
  - Matching scheme (MLM in Alpgen): parton shower constrained to prevent overlap between multiplicity bins
  - Using: merge different final state weighted with a corresponding cross-section

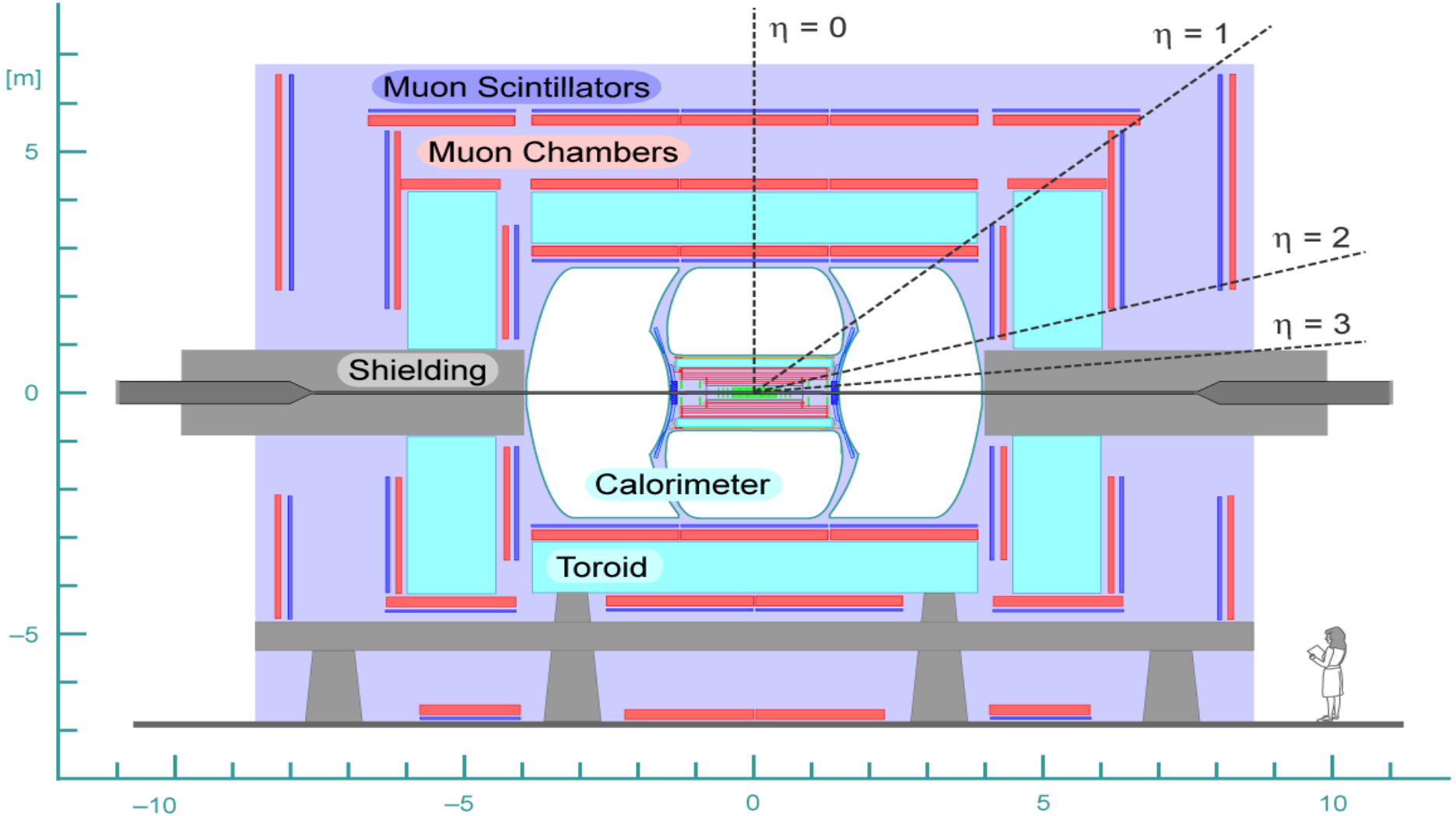
TEVATRON





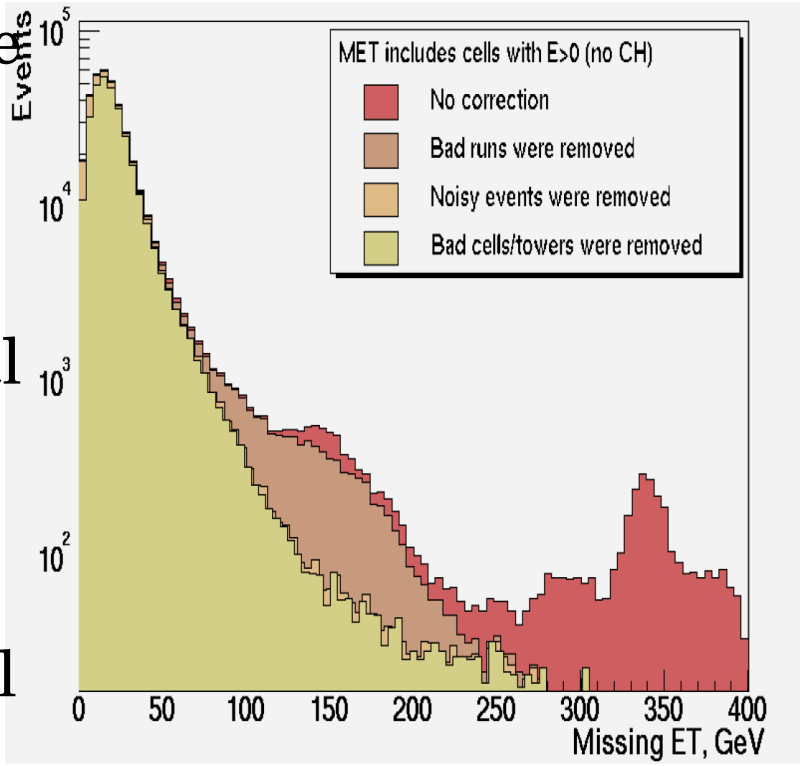
- Current performance: Instantaneous luminosity **1.5×10^{32} cm^{-2}s^{-1}**; delivered integrated luminosity: ~**1.3 fb^{-1}**. Meeting design performance in 2004 and 2005.
- Long term luminosity plan :
  - Instantaneous luminosity: 3×10^{32} cm^{-2}s^{-1} by 2007
  - Delivered integrated luminosity: 4.1 fb^{-1} (Base) - 8.2 fb^{-1} (Design)

D0 detector


η = 0
η = 1
η = 2
η = 3
[m]
5
0
–5
Muon Scintillators
Muon Chambers
Shielding
Calorimeter
Toroid
–10
–5
0
5
10

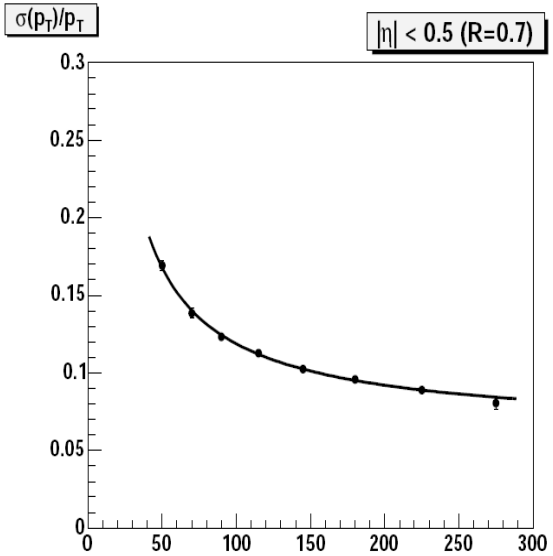
Data Quality: calorimeter example

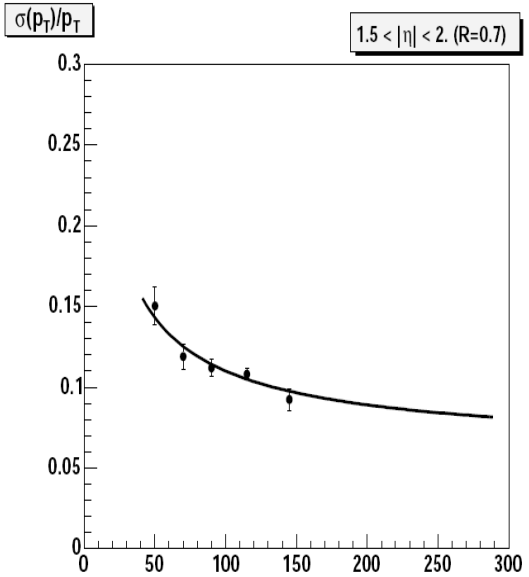
- The very first step in analysis is a data quality selection
  - Removing "bad" runs (hardware failure)
  - Remove noisy cells or tower before reconstruction
  - Remove noisy events
- Additional algorithms applied to improve quality of reconstructed calorimeters objects
  - Remove isolated cells with high signal
  - Remove "noisy" region: cells with small signal, when no neighbors present with significantly large signal



Jet identifications

- Iterative cone algorithm is used for the jets reconstructions (split and merge proto-jets).
- Apply jet identifications criteria: EM fraction not too high; jet is not purely hadronic; no jet formed by single cluster or one tower; jet should be confirmed by L1 trigger readout.
- Correct jet energy with JES is the crucial part of many analysis, e.g. top mass measurement (*see C. Royon JES presentation later today*)

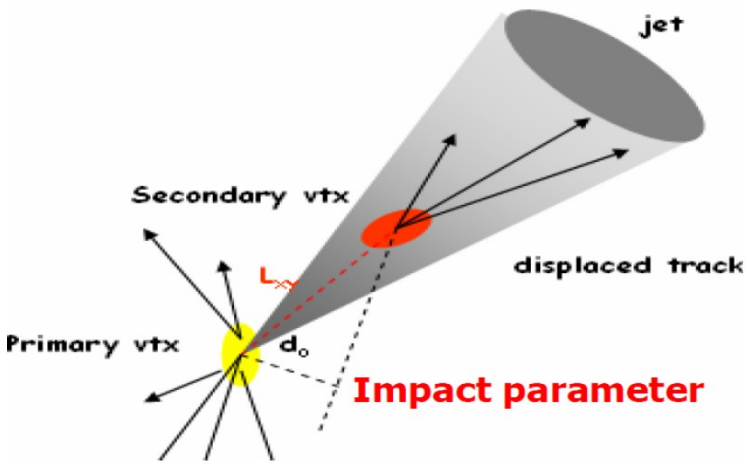


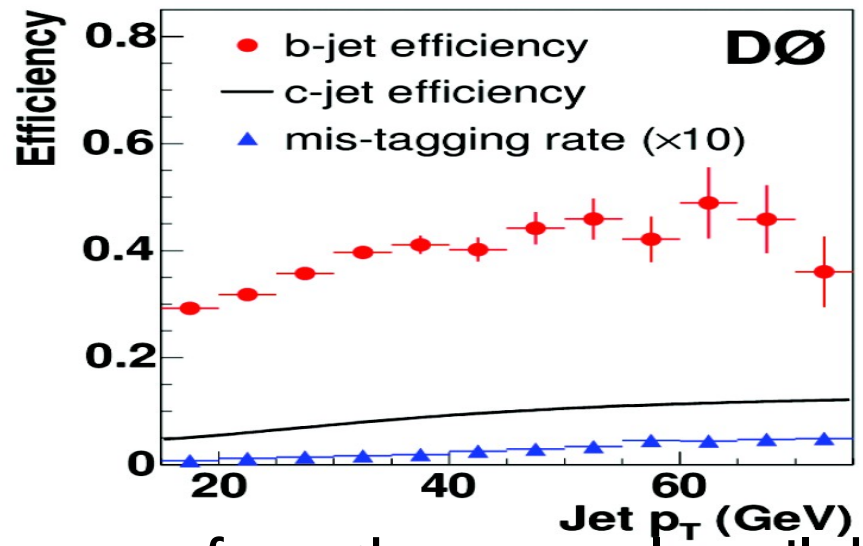


Object ID and trigger efficiencies

- The object ID and trigger efficiencies is measured in data: $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, γ+jet and compared with MC
- For a perfect agreement need to introduced additional scale factors and object over-smearing. Scale factors produced for the object with "standard" selection and have dependence from the chosen parameters (p_T, η, φ, Z_{PV}, ...). The typical average corrections are 1 – 10 %.
- Another product of this comparison are the efficiencies systematics. Usually less than several percent.

b-tagging

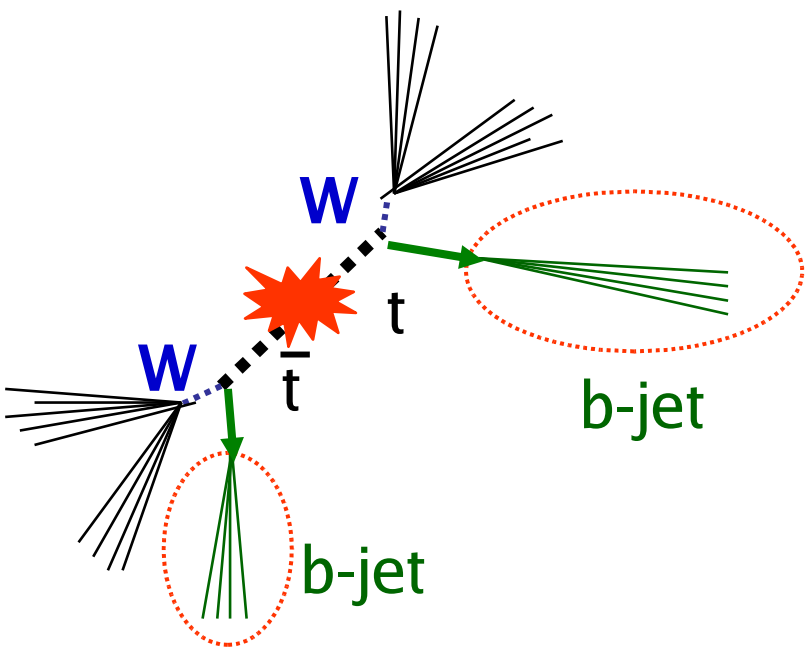




- Before: secondary vertex tagging was one from the several available at D0
- Now: Neural Network tagger is under certifications. Use more inputs: *decay length significance of the Secondary Vertex (SV), weighted combination of the N_{Tracks} with high IP significance; probability that the jet originates from the PV; χ^2 of the SV, number of tracks used to reconstruct the SV; mass of the SV; number of SV found in the jet;*
- Efficiency rise to 60% and more per jet.

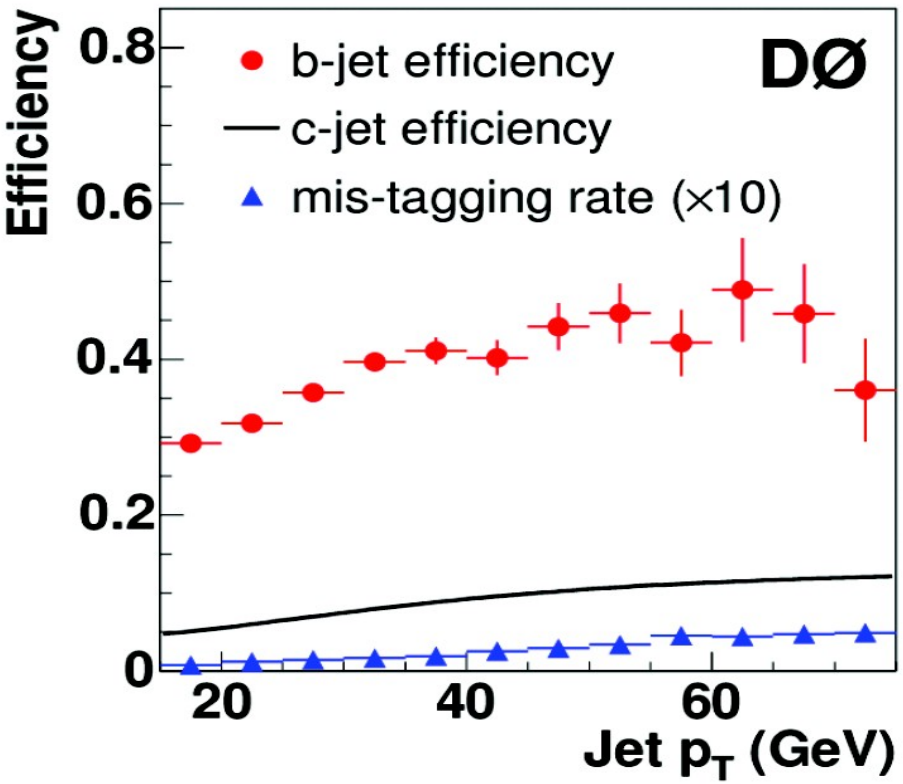
ALL-JETS FINAL STATE

- Final state: at least 6 jets (with 2 b-jets). Large branching ratio (46%), but huge QDC background $\sigma_{QCD} > 5000\ \sigma_{tt}$



- Preselection: at least 6 jets (p_T>15 – 45 GeV), isolated lepton veto, jets should come from the same primary vertex
- Use b-tagging to suppress the QCD background
- Final selection, 2 approaches:
  - Neural network discriminant based on the kinematic variables
  - Mass peak discriminant

SECONDARY VERTEX TAGGING



- Typical efficiency to have at least one b-tag in a $t\bar{t}$ event is 61%. Use muon-tagged jets to measure data / MC efficiency correction factor and then apply this factor to MC.
- 1 or 2 b-tagged jets required for NN analysis
- 2 b-tag + not-tag for all other jets required in mass peak analysis

NN: background estimation

- Tag rate in background data events (TRF) is 3%. This includes: b-jets, c-jets, mis-tag rate for light jets.
- TRF was measured using the pre-selected 6 jet sample (contains mainly background) and parametrized in η, φ, p_T of the jets, primary vertex Z, sum of jet p_T (H_T).
- Corrected background prediction for the double tag events, because bb+jets contributions increased.
- Use NN output outside the tt peak for the normalization.

NEURAL NETWORK DISCRIMINANT

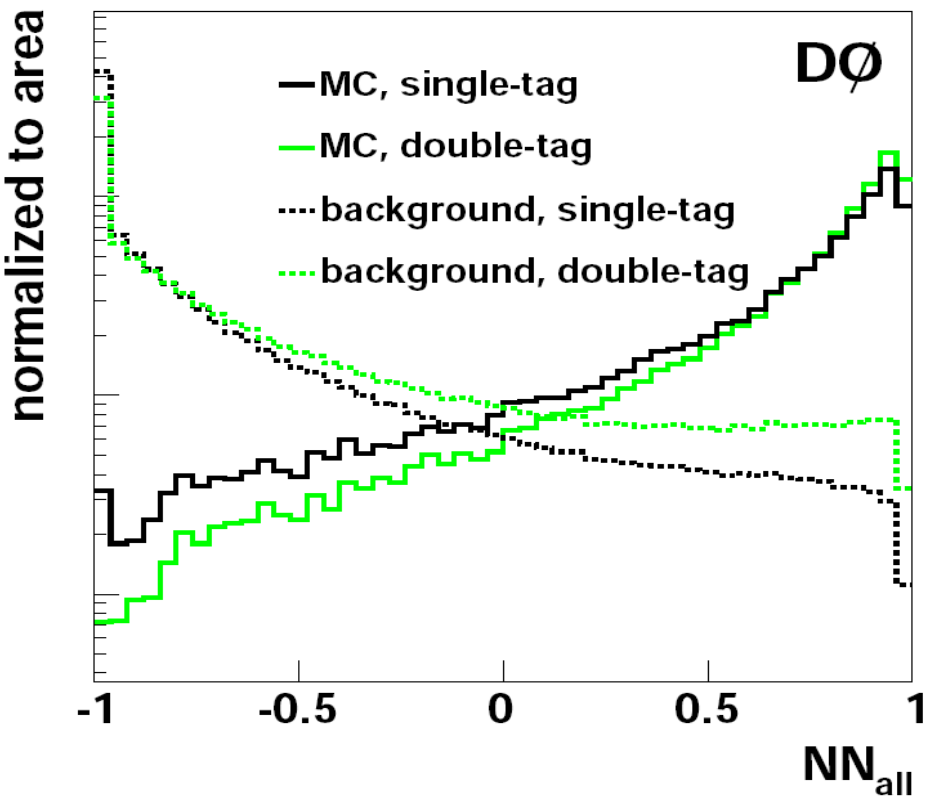


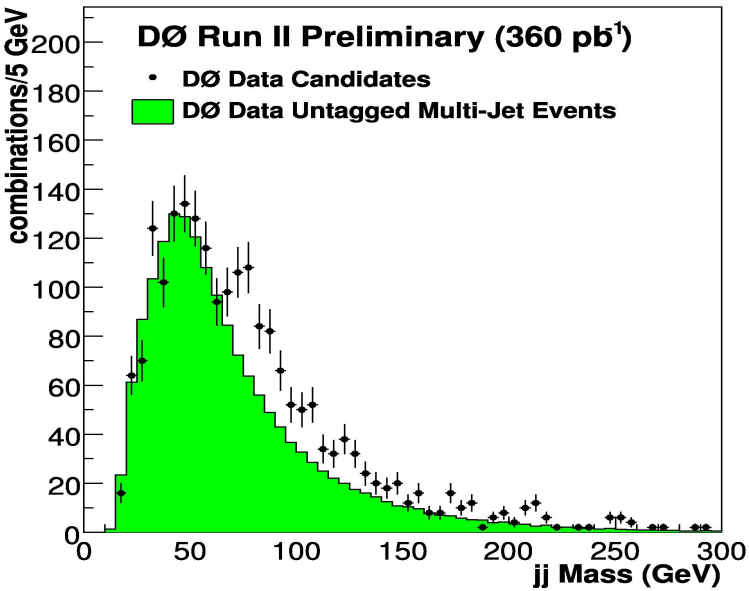
FIG. 7: The output discriminant of NN_{all}, an artificial neural network with six input nodes. NN_{all} is optimized to distinguish between fully hadronic $t\bar{t}$ Monte Carlo events (signal) and multijet production (background).

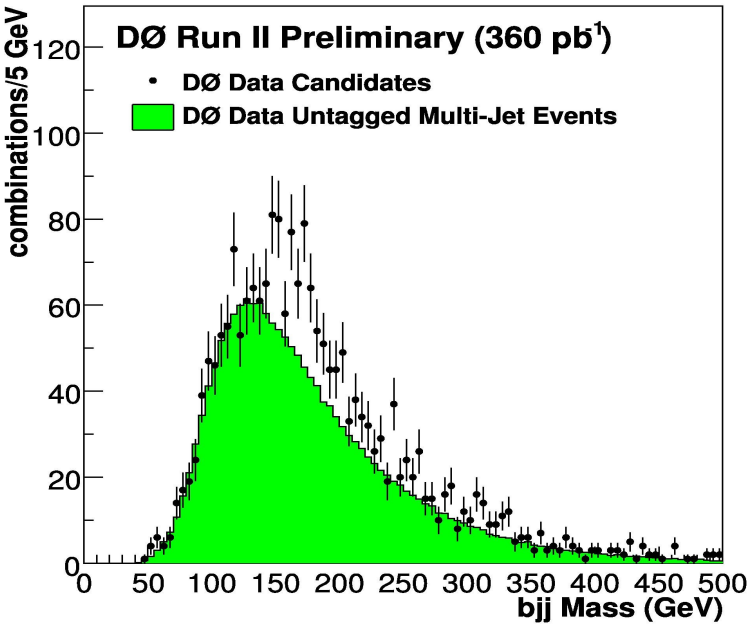
$$M = \frac{(M_{W_1} - M_W)^2}{\sigma_W^2} + \frac{(M_{W_2} - M_W)^2}{\sigma_W^2} + \frac{(m_{t_1} - m_{t_2})^2}{\sigma_t^2}$$

1) Scalar sum of the p_T of four leading jet (H_T)
2) Aplanarity: a linear combination of the eigenvalues of a normalized momentum tensor
3) The geometric mean of the transverse energies of the fifth and six leading jets (E_{T56})
4) The weighted RMS of η of the six leading jets ($<\eta^2>$)
5) The second smallest di-jet mass in the event (M^{34}_{min})
6) The minimal mass likelihood value over all combination (M)

Mass peak discriminant

- Cuts on aplanarity, sphericity, centrality and distance between 2 b-taged jets are used for selection.
- Background estimation:
  - Multijets QCD events + W+jets background distributions derived from data (random “b-tag”)
  - Additional correction applied to take into account correlation between b-jets
  - Normalized background sample using jj mass spectrum (m < 65 GeV)



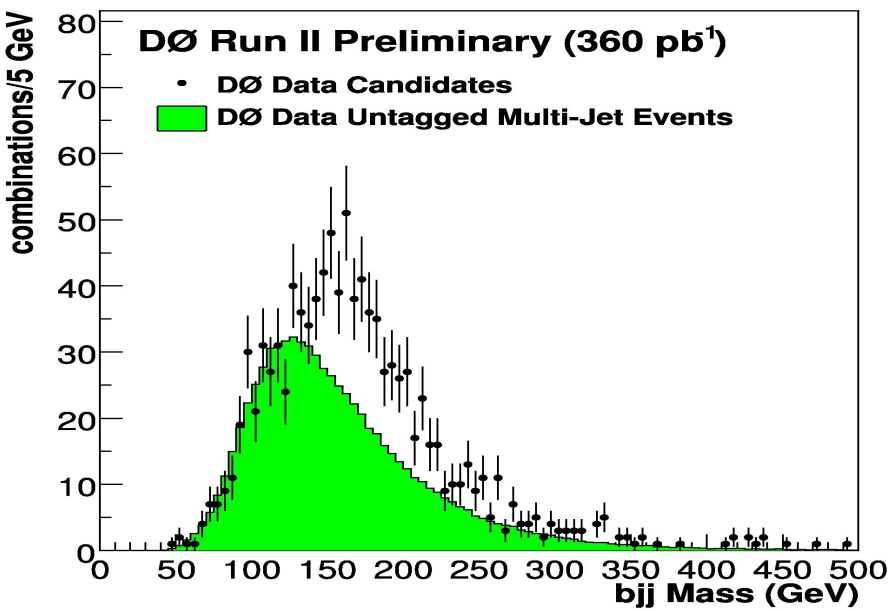


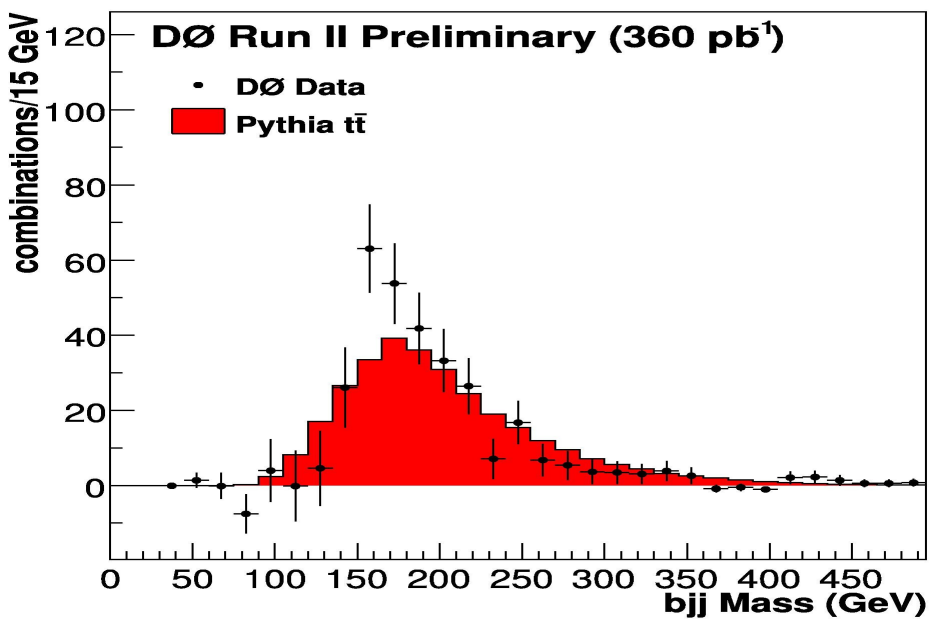
Results

- NN: use cut on NN, subtract the background contribution, extract cross-section separately for the one tag and double tag events and combine them.
- Mass peak discriminant (main syst. is a backg. prediction):

$$\sigma = 12.1 \pm 4.1(stat) \pm 4.6\ (syst)\ pb,\ m_t = 175\ GeV,\ L \sim 360\ pb^{-1}$$

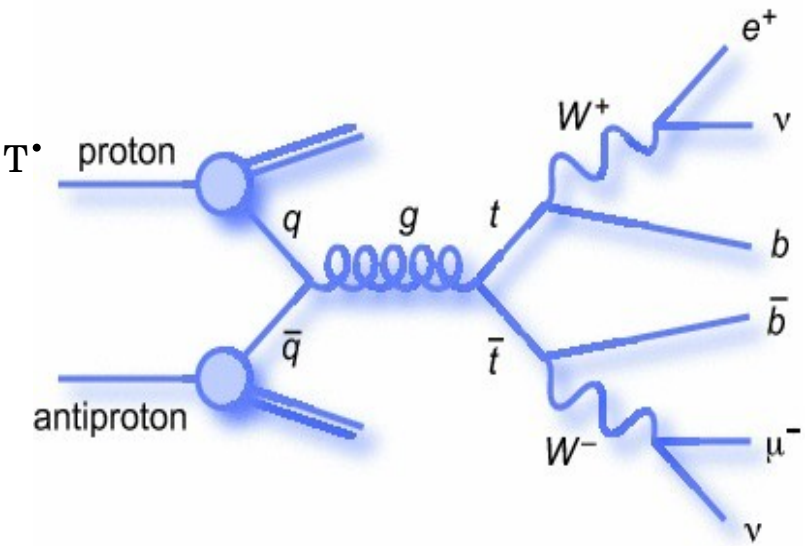
- CDF: $8.3 \pm 1.0\ (stat)\ ^{+2.0}_{-1.5}(syst)$, $L \sim 1000\ pb^{-1}$





DILEPTON FINAL STATE

- Signature: 2 b jets; 2 high p_T leptons (electron(s), muon(s)); large missing E_T.

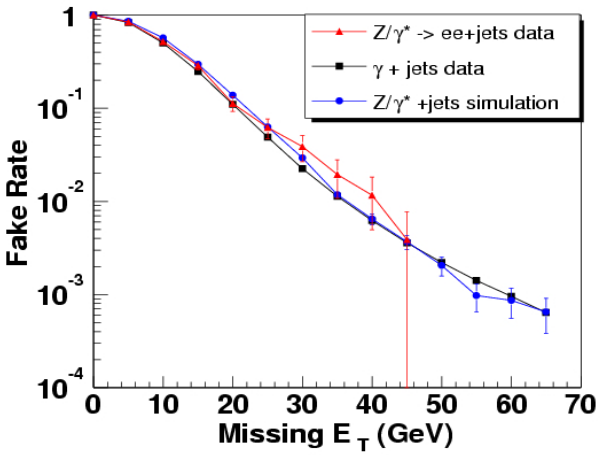


- The "cleanest" channel for the top production, but small branching ratio (~6%)
- Main backgrounds:
  - "physics" backgrounds WW, WZ and Z production (ee, μμ, eμ);
  - "fake" electron, muon and missing E_T backgrounds;
- 4 separate analyses: ee (1.6%), μμ (1.6%), eμ (3.2%), lepton+track. *Branching ratios include τ decays to electron and muon.*

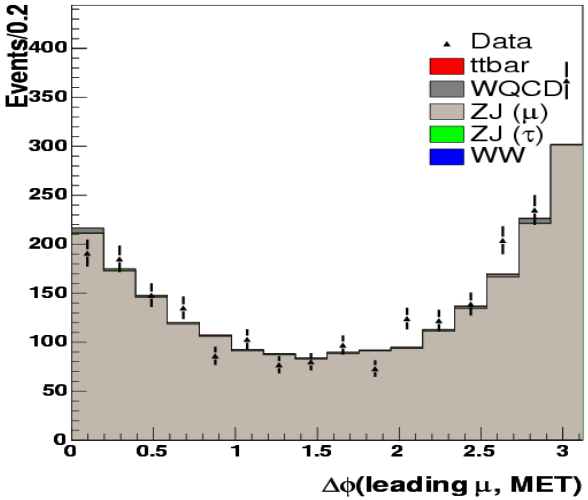
EVENTS SELECTION

- **Electron:** isolated electron-like shower in calorimeter, matched to the track in the tracking system, $p_T > 15$ GeV, $|\eta| < 1.1$ or $1.5 < |\eta| < 2.5$
- **Muon:** track in the muon system matched to the track in the tracking system, isolated both in the calorimeter and in the tracking system, $p_T > 15$ GeV, $|\eta| < 2.0$
- **Jets:** jets-like shower in the calorimeter $p_T > 20$ GeV, $|\eta| < 2.5$
- **Missing E_T:** sum of the calorimeter cell transverse energies, corrected by jet, electron, muon energy depositions. Typical cut value 35 – 40 GeV

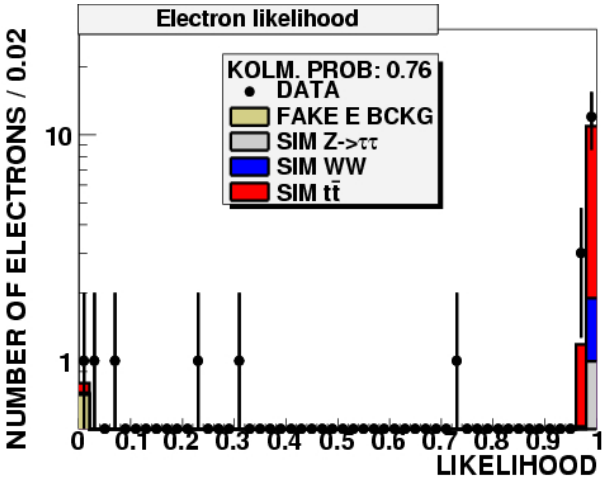
INSTRUMENTAL BACKGROUND



- **Di-electron** channel has a multijet and Z production backgrounds due-to mismeasured missing E_T and misidentified electron. These background estimated from data.



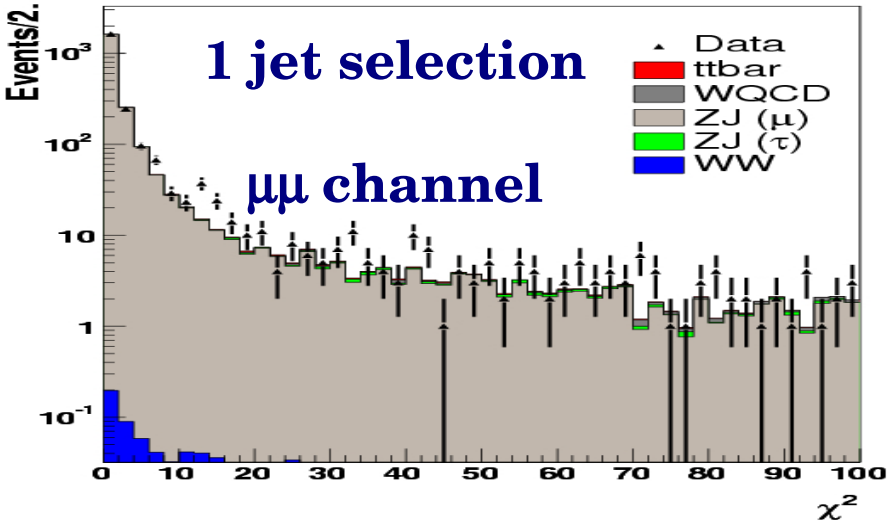
- **Di-muon** channel has a significant contribution from Z→μμ. It is estimated with simulation corrected by experimental resolutions.

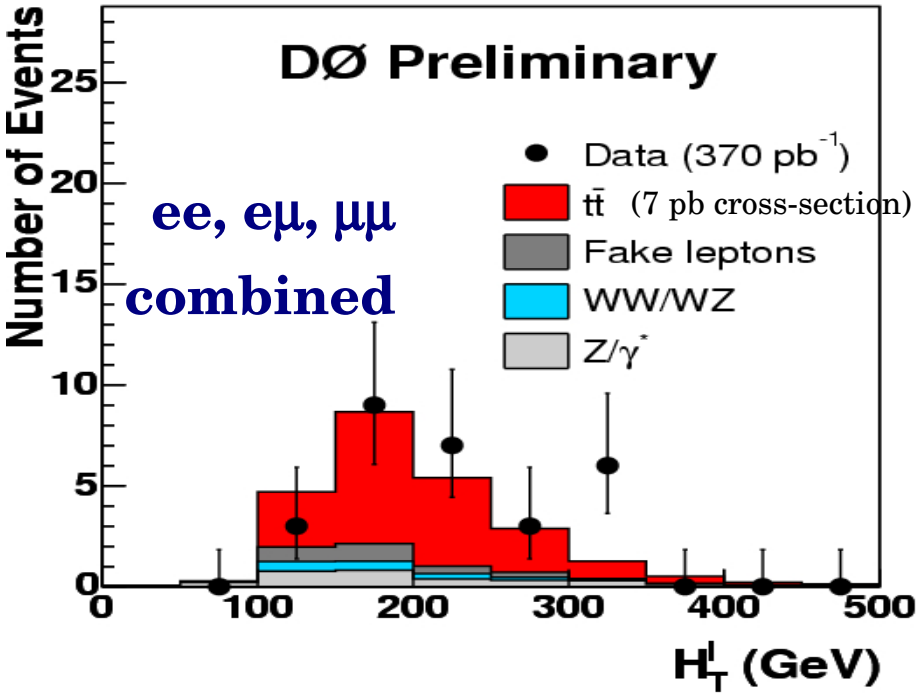


- **Electron-muon** channel has a contribution from multijet production due to the misidentified electron. This contribution is subtracted using the electron likelihood distribution with signal and background shapes measured from data

FINAL SELECTION

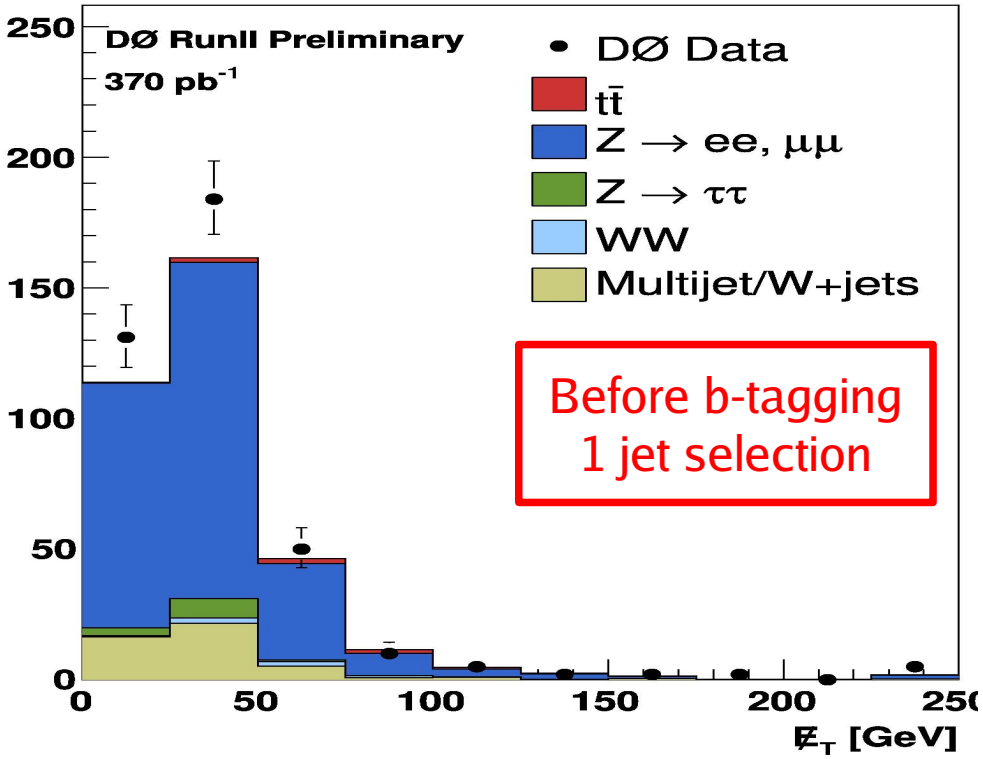
- **Di-electron** channel uses cut on sphericity (sum of 2 leading eigenvalues of a normalized momentum tensor)
- **Di-muon** channel uses cut on the χ^2 value calculated in $Z \rightarrow \mu\mu$ hypothesis ($\chi^2 > 2$)
- **Electron-muon** channel uses cut on the sum of transverse momenta of 2 leading jet and leading lepton ($H^l_T > 122$ GeV)

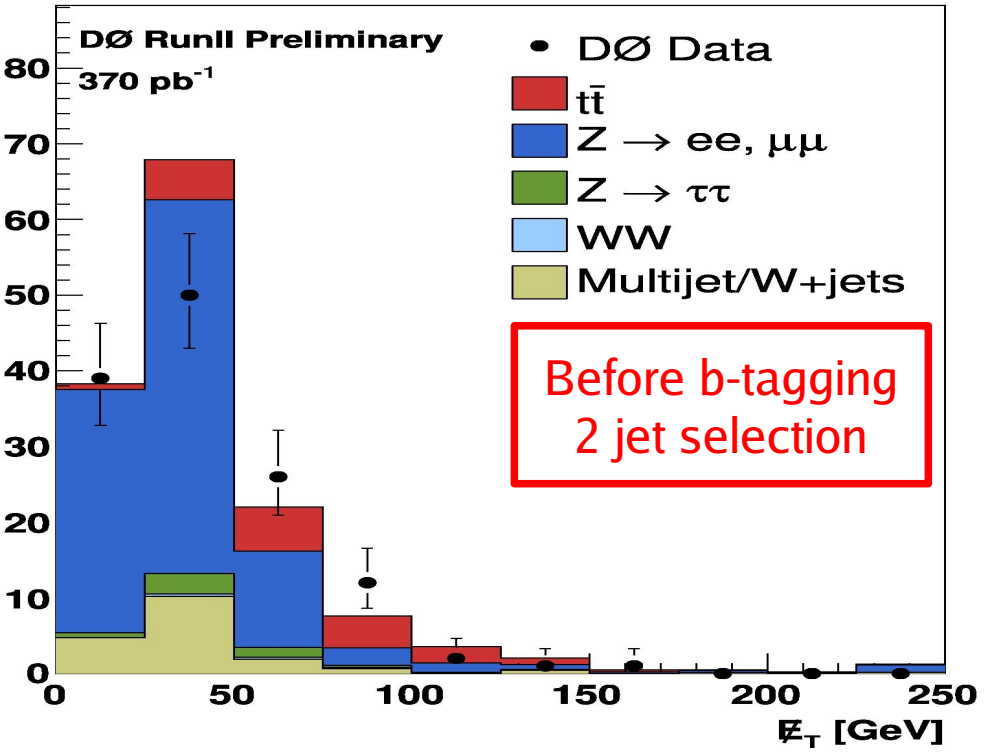




Lepton + track: event selection

- One identified lepton + track (p_T > 15 GeV) + 2 jets as before
- MET (cut differs in the e and muon channel, inside or outside Z-mass peak), 15 – 35 GeV
- At least one b-tagged jet. Veto on eμ final state.





Lepton + track background

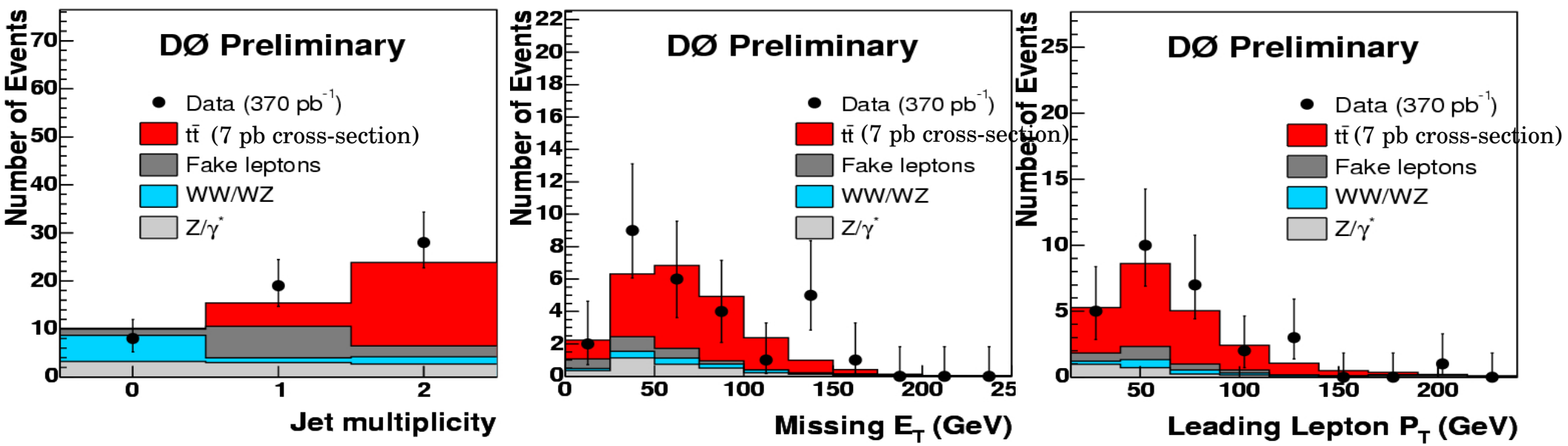
- Physics background. $Z \rightarrow \tau\tau$, WW estimated from MC. b-tagging efficiency for $Z \rightarrow \tau\tau$ the same as for the $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, for WW the same as for W+jets
- Missing E_T instrumental background. Mainly $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ events. Normalized using the ratio of the observed to predicted number of events in the low MET region.
- Lepton, track instrumental background arise from QCD and W+jets events. Estimated using samples with different data / background fractions.

More about physics background.

- Z and WW background is simulated using Pythia hadronization, but results are not perfect. Jet selection require additional correction
- Z background: Z p_T reweighting, cross-section available NNLO could be corrected with data.
- WW cross-section available for NLO, but generator is only LO. Due to the small statistics it is difficult to normalized WW+jets to data. Lead to the big systematics error on this background contribution.

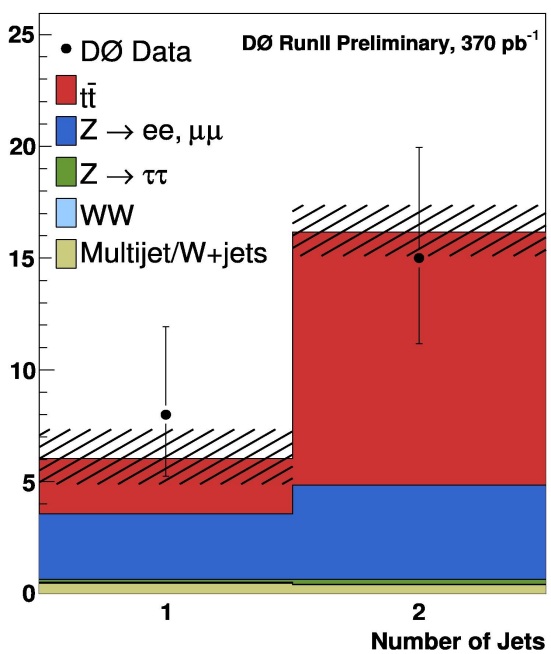
Selected events

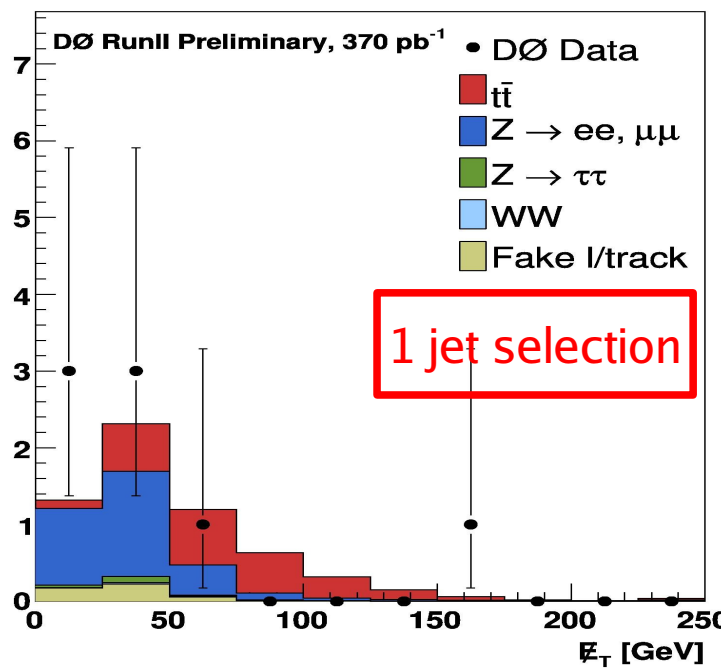
Category	ee	$\mu\mu$	$e\mu$	$\ell\ell$
Integrated luminosity (pb^{-1})	384	363	368	–
Z/γ^*	$0.75^{+0.18}_{-0.21}$	$1.01^{+0.22}_{-0.34}$	$1.22^{+0.33}_{-0.39}$	$2.98^{+0.43}_{-0.55}$
WW/WZ	$0.20^{+0.10}_{-0.14}$	$0.20^{+0.08}_{-0.07}$	$1.13^{+0.45}_{-0.48}$	$1.53^{+0.47}_{-0.50}$
Instrumental leptons	0.09 ± 0.03	0.13 ± 0.04	$2.13^{+2.50}_{-1.66}$	$2.35^{+2.50}_{-1.67}$
Total background	$1.0^{+0.2}_{-0.3}$	$1.3^{+0.3}_{-0.4}$	$4.5^{+2.6}_{-1.8}$	$6.8^{+2.6}_{-1.8}$
Signal efficiency	0.082	0.064	0.139	–
Expected signal	3.5 ± 0.4	2.5 ± 0.3	$11.3^{+1.2}_{-1.4}$	$17.3^{+1.3}_{-1.5}$
Signal + background expectation	$4.5^{+0.4}_{-0.5}$	$3.8^{+0.4}_{-0.5}$	$15.8^{+2.8}_{-2.3}$	$24.1^{+2.9}_{-2.4}$
Selected events	5	2	21	28

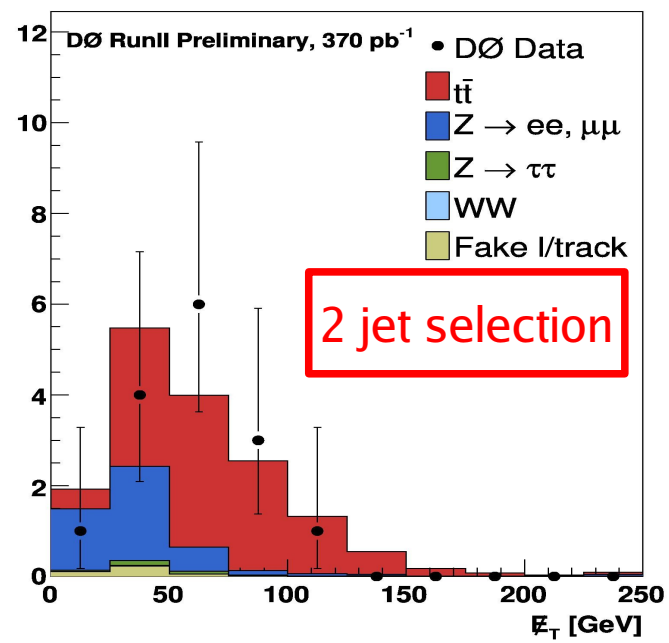


Lepton + track events

	e + track		μ + track	
	Njets = 1	Njets $\geq$ 2	Njets = 1	Njets $\geq$ 2
	Expected number of tagged events			
WW	0.037 ± 0.002	0.010 ± 0.002	0.016 ± 0.001	0.009 ± 0.002
$Z/\gamma^* \to \tau\tau$	0.09 ± 0.02	0.13 ± 0.02	0.03 ± 0.01	0.09 ± 0.02
$Z/\gamma^* \to ee, \mu\mu$	1.49 ± 0.04	2.35 ± 0.06	1.44 ± 0.04	1.86 ± 0.06
Multijet/W+jets	0.36 ± 0.06	0.35 ± 0.07	0.08 ± 0.02	0.05 ± 0.03
Total background	1.97 ± 0.08	2.83 ± 0.09	1.57 ± 0.05	2.00 ± 0.07
Tot. uncertainty (stat+syst) on bkg	+0.91 −0.85	+0.87 −0.64	+0.77 −0.77	+0.51 −0.49
$t\bar{t}$	1.55 ± 0.03	6.59 ± 0.07	0.92 ± 0.02	4.74 ± 0.06
Total	3.53 ± 0.08	9.4 ± 0.1	2.49 ± 0.05	6.74 ± 0.09
Tot. uncertainty (stat+syst) on $t\bar{t}$ + bkg	+0.99 −0.86	+0.99 −0.85	+0.83 −0.77	+0.67 −0.64
	Observed number of tagged events			
Data	7	9	1	6







RESULTS

Lepton + track and eμ combined (L ~ 360 pb):

$\sigma = 8.6\ ^{+1.9}_{-1.7}$ *(stat) ± 1.1 (syst) ± 0.6 (lum) pb*

CDF (2 leptons, L ~ 750 pb) : 8.3± 1.5 (stat) ± 1.1 (syst)

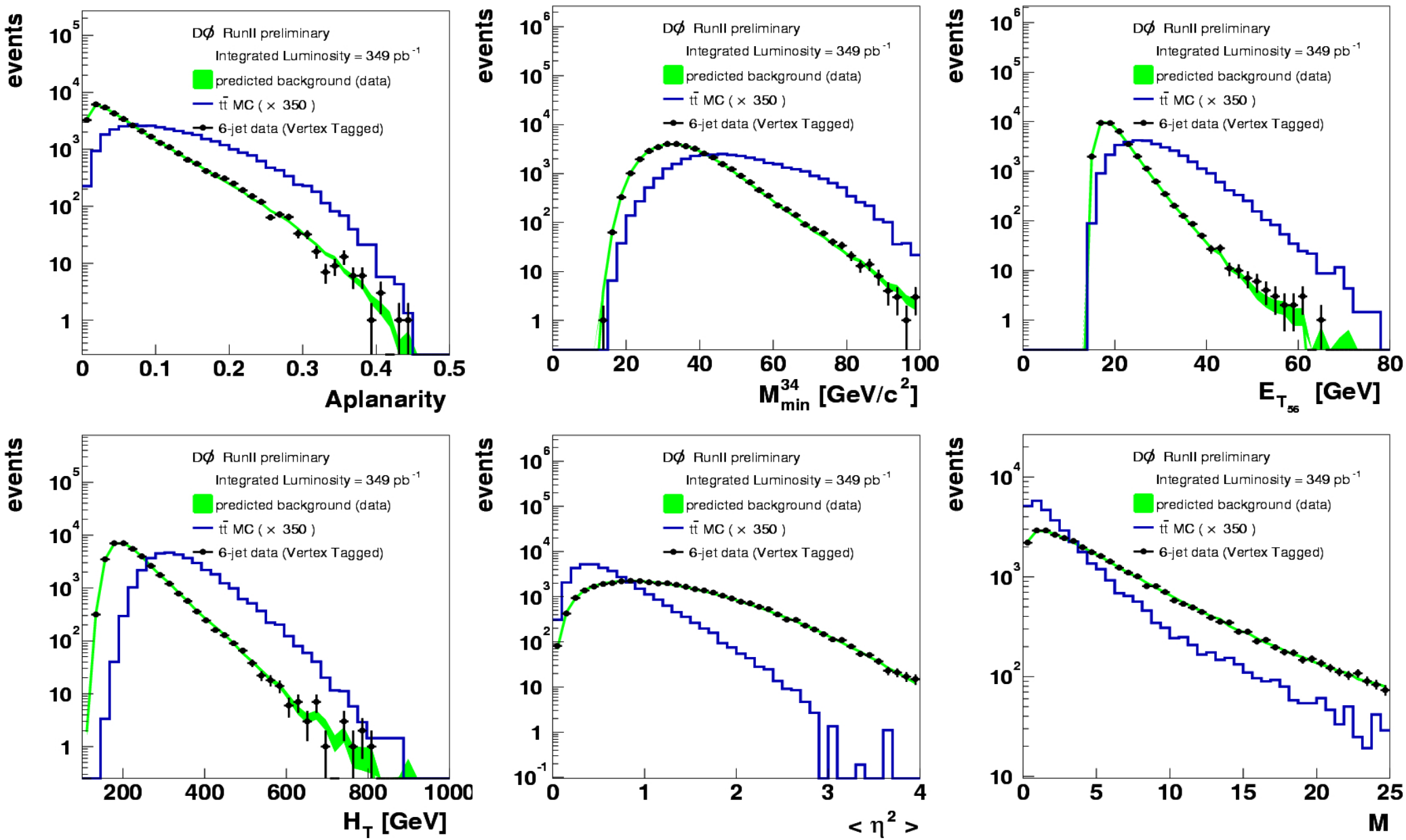
SYSTEMATICS UNCERTAINTY	VALUE, %	
Jet energy calibration	+6.6	-8.5
Jet identification	+2.2	-3.2
Muon identification	+4.5	-4.4
Electron identification	+3.7	-2.9
Trigger	+3.7	-5.8
Other	+9.0	-4.5

CONCLUSION

- The most crucial part for the analysis is the data quality, objects identifications, b-tagging, jet energy scale, in some cases MC simulation (hadronization).
- Despite many efforts the "pure" MC is not sufficient to determine ID and trigger efficiencies, background contribution. Need to use appropriate correction or data based methods
- More than 1 fb^{-1} available for analysis. In all final states the cross-section measurement will be limited by systematics. With current analysis ~10% error is reachable

BACKUP

MLM Matching Scheme

- generate parton-level configuration for a given multiplicity bin withcuts $p_T > p_{T\ min}$ and $R > R_{min}$
- perform parton showering using HERWIG or PYTHIA
- process showered event before hadronization with jet algorithm (UA1, though kT and Run 2 cone yield similar results)
- match partons and parton-shower jets in $\eta - \varphi$:
  - a jet can only be matched to a single parton
  - exclusive: every parton matched to a jet with $N_{jet} = N_{parton}$
  - inclusive: all partons matched to jets
  - good to LL precision

NN KINEMATIC DISTRIBUTIONS