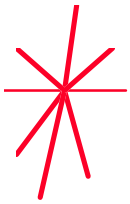


Top to All-Jets Production at DØ

Norman Amos

University of Michigan

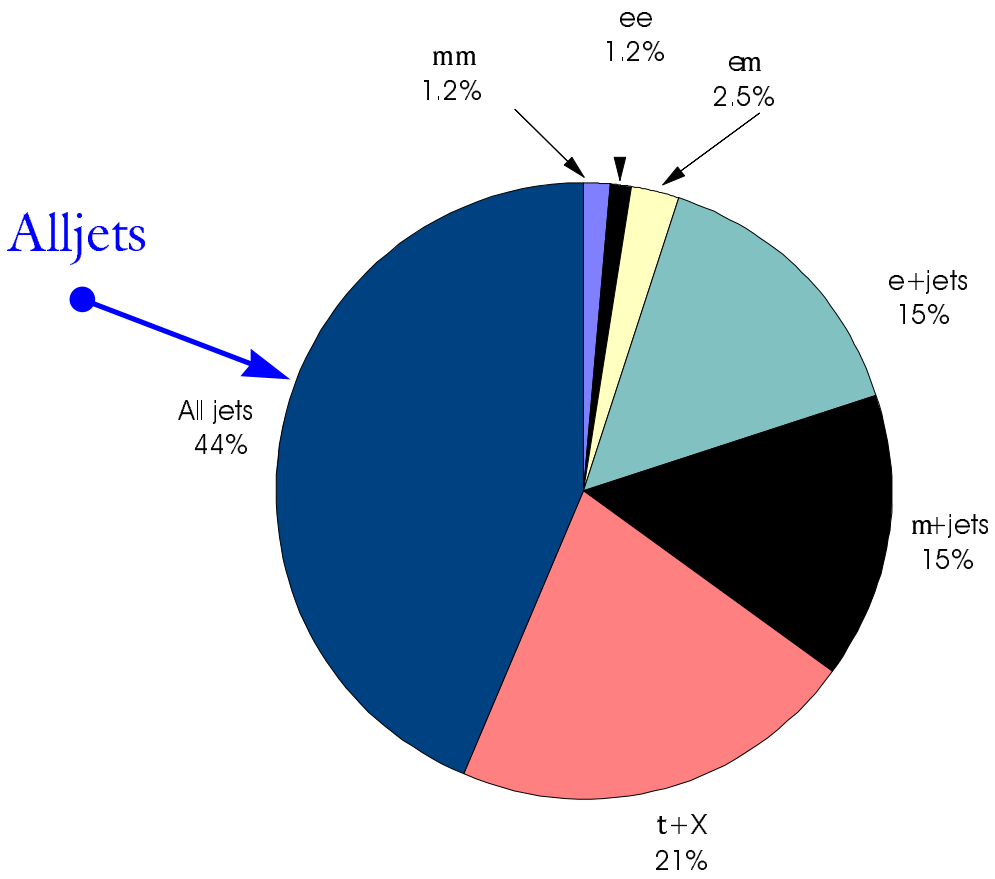
for the DØ collaboration

Fermilab Wine & Cheese

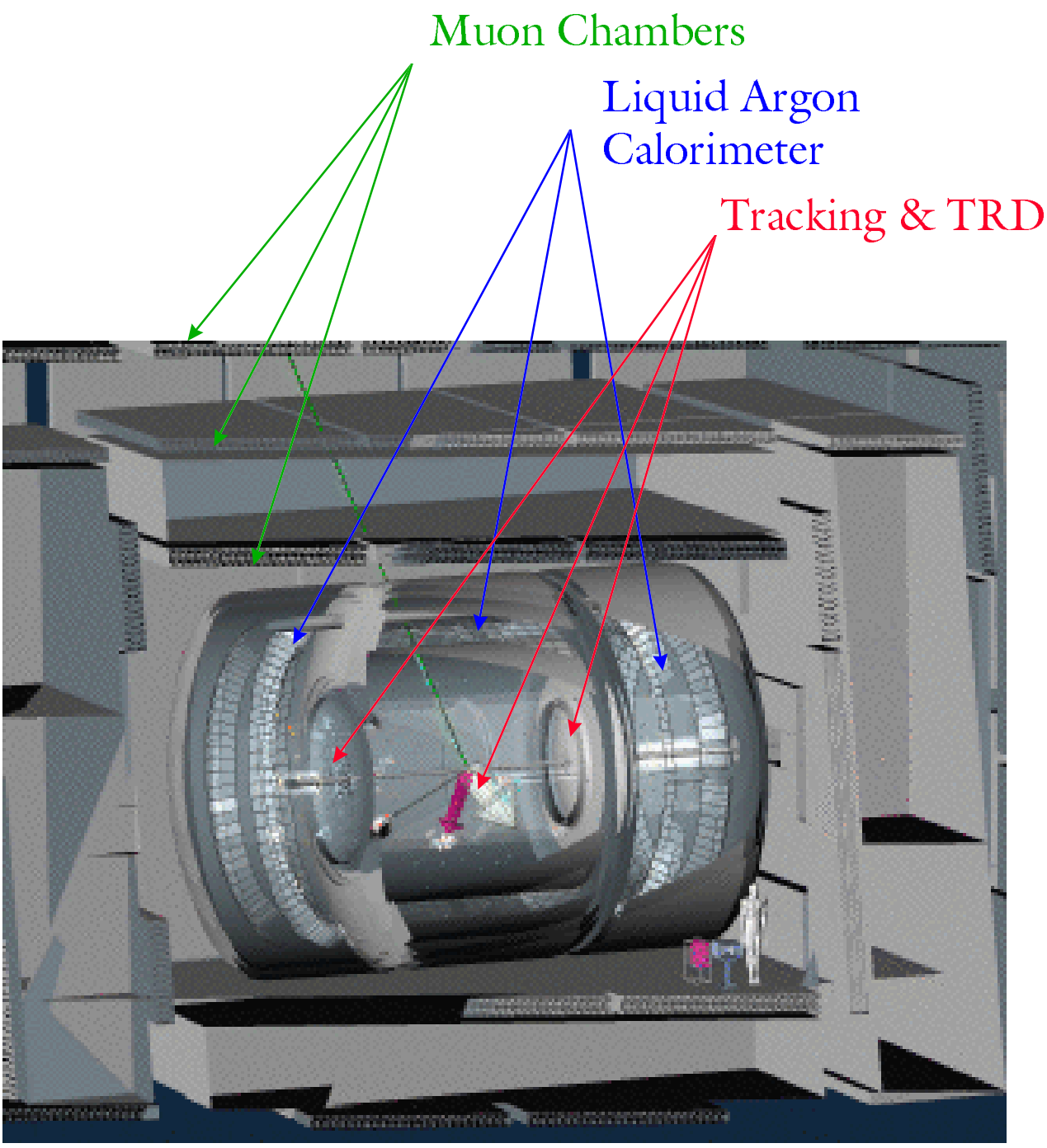
June 20, 1997

Introduction

- Observation of the top quark (CDF & DØ) on Feb. '95
- At Tevatron, top quarks are largely pair produced.
- Each top decays hadronically:
  - $t \rightarrow W + b$, $W \rightarrow$ jets
  - Branching fraction for Alljets channel ~44%.

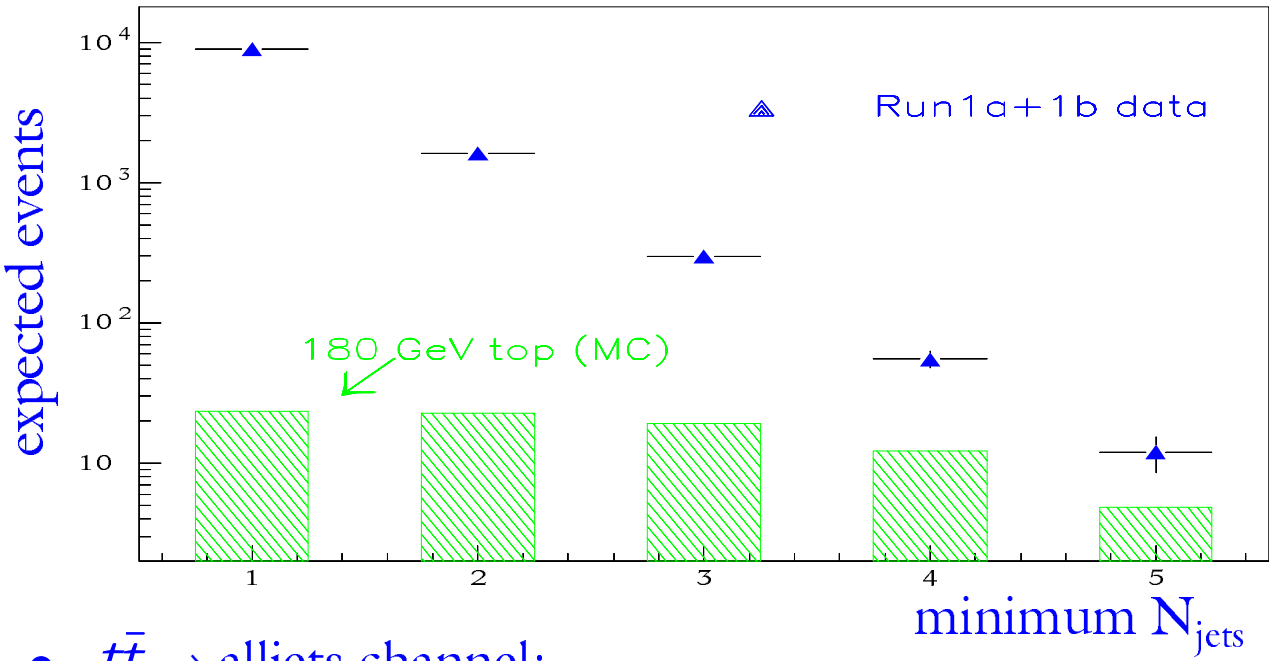


the DØ Detector

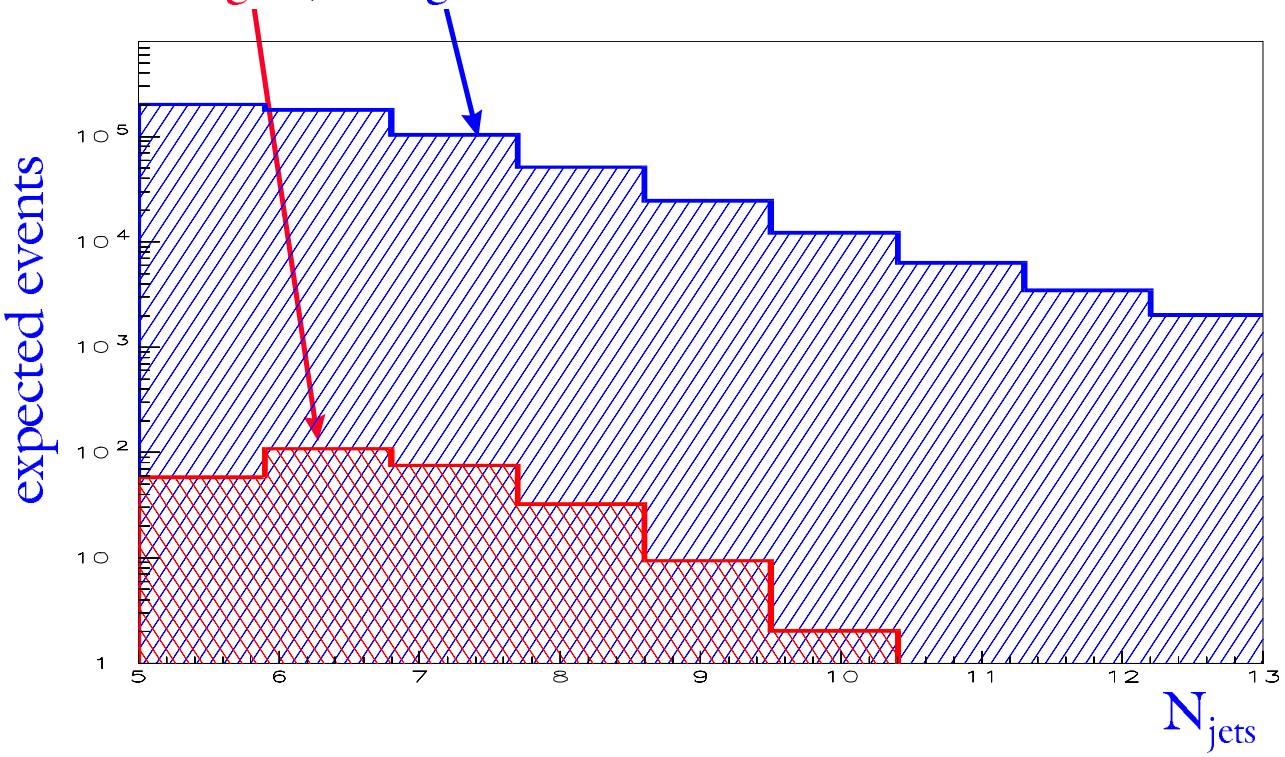


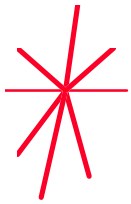
Signal / Background

- $t\bar{t} \rightarrow e$ + jets channel (*mainly QCD & W+jets*):



- $t\bar{t} \rightarrow$ alljets channel:
  - Primary background is from QCD multijets
  - Signal/Background ~ **4 x 10^{-4}**





Needle in a Haystack

- Very challenging channel!
- Enormous background.
- Need all of the standard techniques.
- Plus new techniques.

- Strategy:
  - kinematic and topological variables
  - quark/gluon separation (*new*)
  - *b*-tagging
  - almost undiluted background (use data to model the background)
  - Neural Networks (*already used in Top mass & LQ search*)

Trigger & Offline

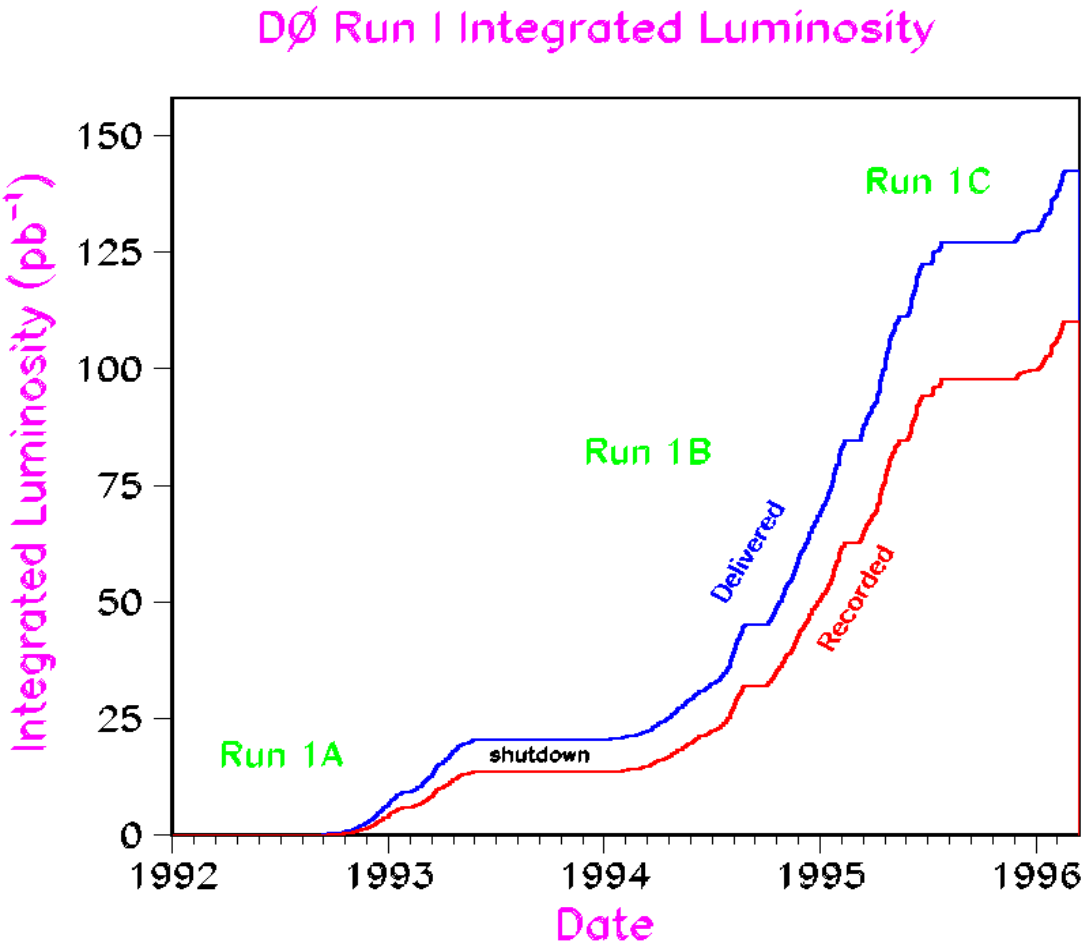
- Level 1:
  - Three large calorimeter trigger tiles ($\Delta\eta \times \Delta\phi = 0.8 \times 1.6$) with $E_T > 15$ *GeV*
- Level 2:
  - Five jets ($\Delta R = 0.3$ cone) with $|\eta| \leq 2.5$ and $E_T > 10$ *GeV*
  - Sum of transverse jet energy > 115 *GeV*
- Reconstruct jets using cone algorithm:
  - $\Delta R = 0.3$ cone, better for resolving jets (less merging of adjacent jets)
  - $\Delta R = 0.5$ cone, better energy calibration (less out-of-cone energy)

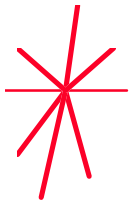
General Condition	Requirement	Effective Cross Section	Efficiency for Top
Level 1	3 large tiles	0.4 ± 0.1 μb	0.98
Level 2	5 jets with $E_T > 10$ *GeV*	20 ± 5 nb	0.93
Level 2	$H_T > 115$ *GeV*	5.3 ± 1.3 nb	0.87

Data Sample

- Data from 1992 - 1996 runs.
- Total Integrated Luminosity ~ 110 pb^{-1}.
- Reconstructed events ($N_{jets} \geq 5$):
  - total number of events ~ 600 k
  - expected top yield ~ 240 events

} 0.04% !





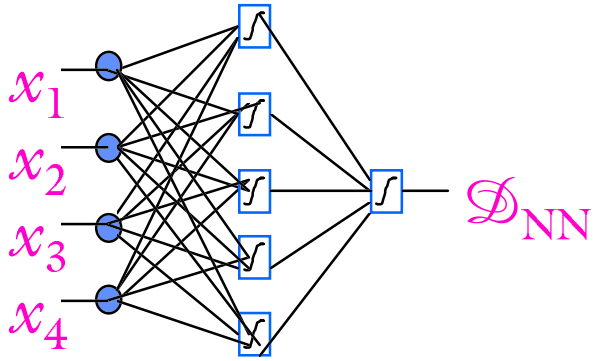
Offline Analysis

- Basic Selection Criteria:
  - No isolated leptons (orthogonal to other channels)
  - Six or more 0.3-cone jets (removes ~14% of top)
    - Improves S/B (removes ~36% of bkg.)
    - E_T of 6th jet used in analysis
  - Less than nine 0.5-cone jets (removes ~5% of top)
    - Improves S/B (removes ~13% of bkg.)
    - Improves the calculation of some variables by reducing combinatorics

- Neural Network Analysis:
  - Only modest discrimination from any single variable
  - Improve signal/background by using many variables
  - Traditional cuts are not effective with large number of variables which are highly correlated

Neural Networks (NN)

- Neural Network is a Multivariate Discriminant which:
  - Sum over input sets × weights
  - Sigmoid function @ some threshold
  - Training ⇒ setting weights & thresholds



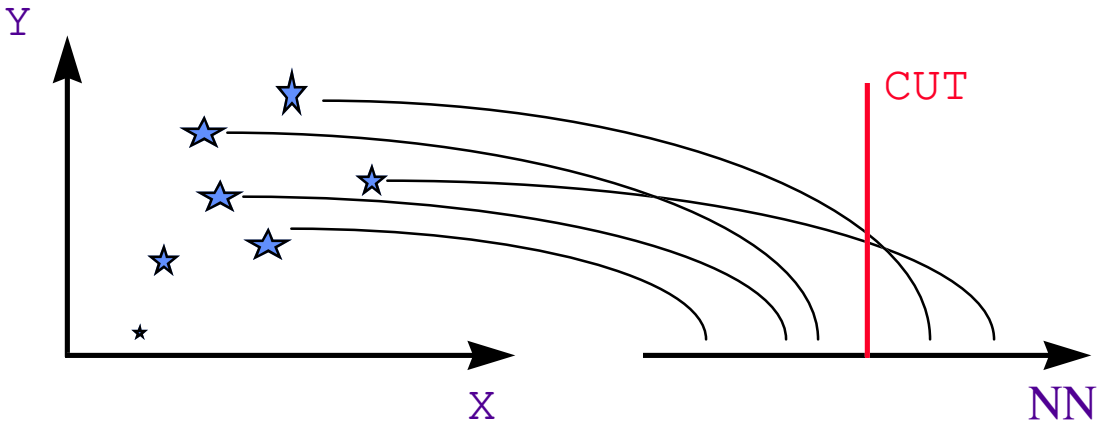
- NN is trained on top Monte Carlo and on background.
- $\mathcal{D}_{NN}$ directly approximates signal probability:

$$D_{NN} = \frac{s(\vec{x})}{s(\vec{x}) + b(\vec{x})}$$

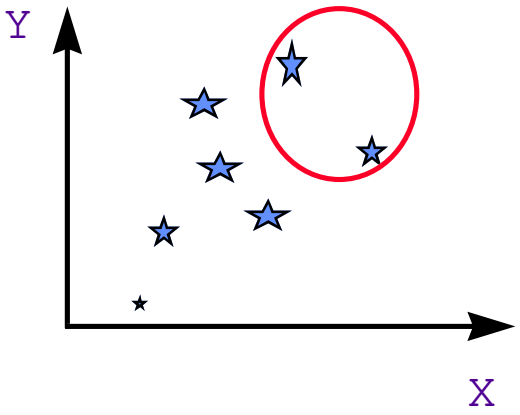
- Signal region: $\mathcal{D}_{NN} \Rightarrow 1$.
- Background region: $\mathcal{D}_{NN} \Rightarrow 0$.

Why Neural Networks?

- Tool that properly handles nonlinear correlations between input variables.
- Improves signal efficiency by handling correlations.
- Describes likelihood of being top *vs.* background.
- Simplifies analysis by mapping multi-dimensional space onto 1 output dimension.

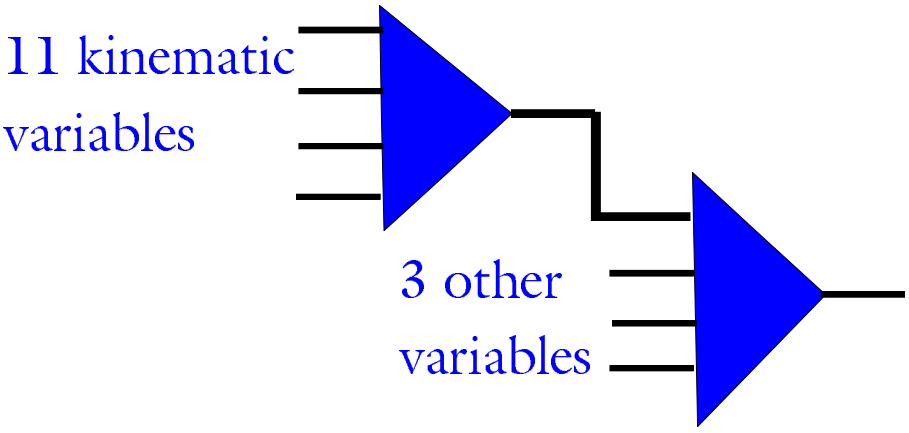


- Simply a cut in multi-dimensional space.
- May take the NN output at face value as long as:
  - NN training is **independent** of the data
  - One has a **reliable** way to calculate background



Neural Network Analysis

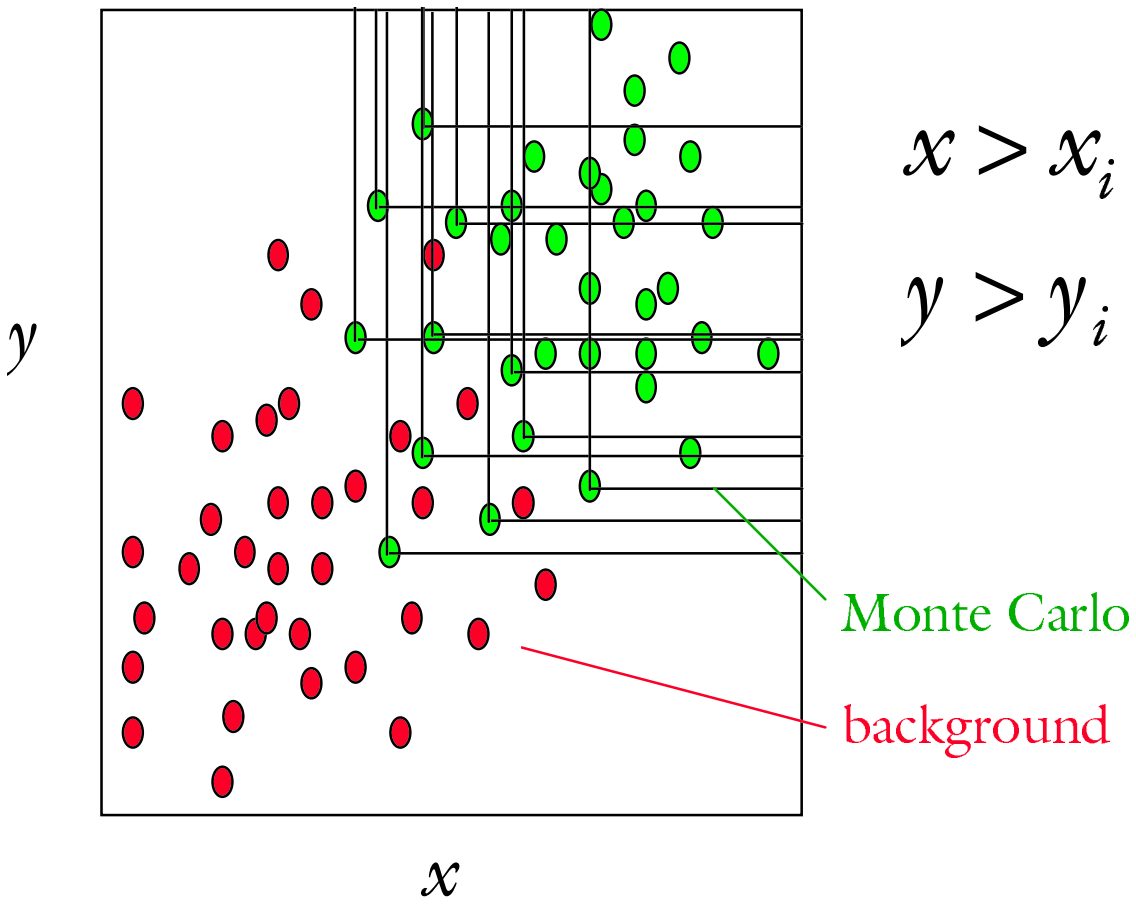
- Independence maintained:
  - Train only on **untagged** (tag=μ-tag) events.
  - Apply to **tagged** events to extract signal.
- Train with Herwig top (signal) + untagged data (background).
- Inputs to first Neural Network:
  - 11 kinematic variables
- Inputs to final Neural Network:
  - Output of first Neural Network
  - P_T of muon
  - Jet width discriminant
  - Likelihood discriminant



- 14 variables in total.
- One final Neural Network output.
- Extract $\sigma_{t\bar{t}}$ from NN output distributions.

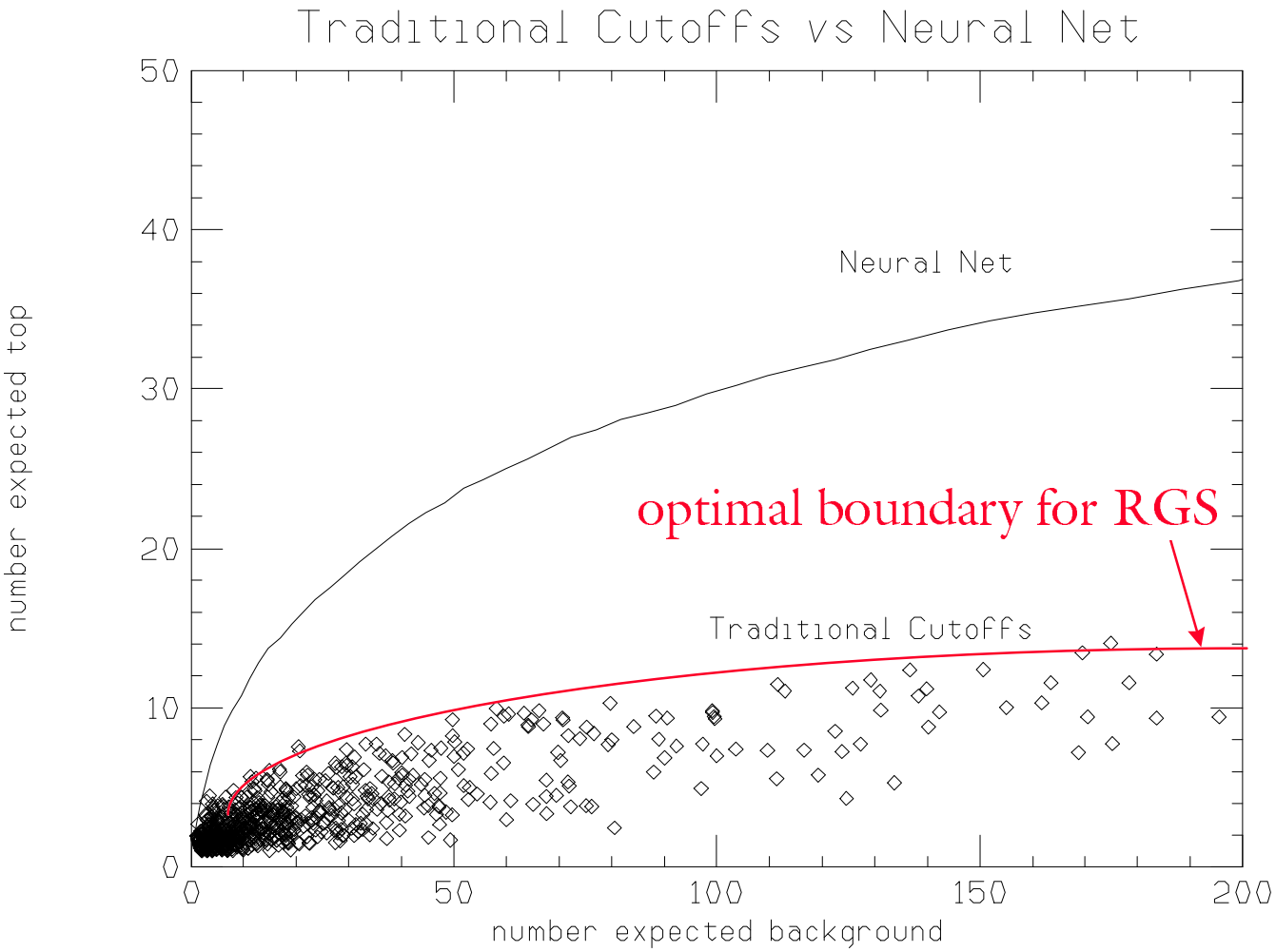
Random Grid Search (RGS)

- Technique to search for optimal cuts:
  - Take each point (x_i,y_i) of the signal Monte Carlo as a cutoff in variables (x,y)
  - Apply this cut and determine the expected signal and expected background
  - Observe that we have a set of cuts that comprise an "optimal boundary"
  - Choose cuts which fall on the "optimal boundary"

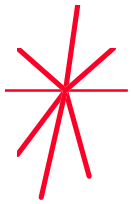


Comparison of NN to Random Grid Search Technique

- The Neural Network can be compared to the results achievable using traditional cuts (*RGS*):



- The Neural Network gains ~ factor of 3 in signal at an equivalent background.



Summary of Neural Networks

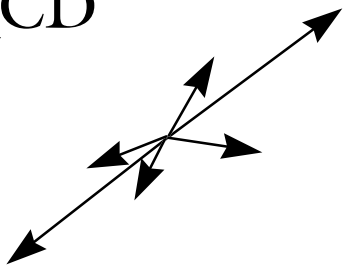
- Neural Networks have been successfully employed in the DØ top mass analysis and in the DØ Lepto-Quark search.

- Neural Networks give superior discrimination when handling a large number of correlated input variables.

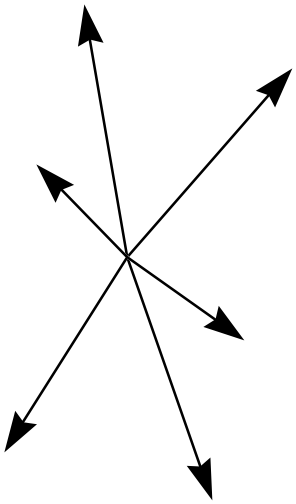
Event Characteristics

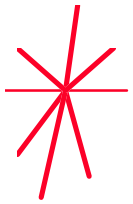
- QCD Multijets:
  - largely $2 \rightarrow 2$ process + gluon brems
  - jets are softer
  - jets are more planar
  - jets are more forward-backward in detector
- TOP:
  - harder jets
  - event topology is more spherical
  - jets are more central

QCD



$t\bar{t}$





11 Kinematic Variables

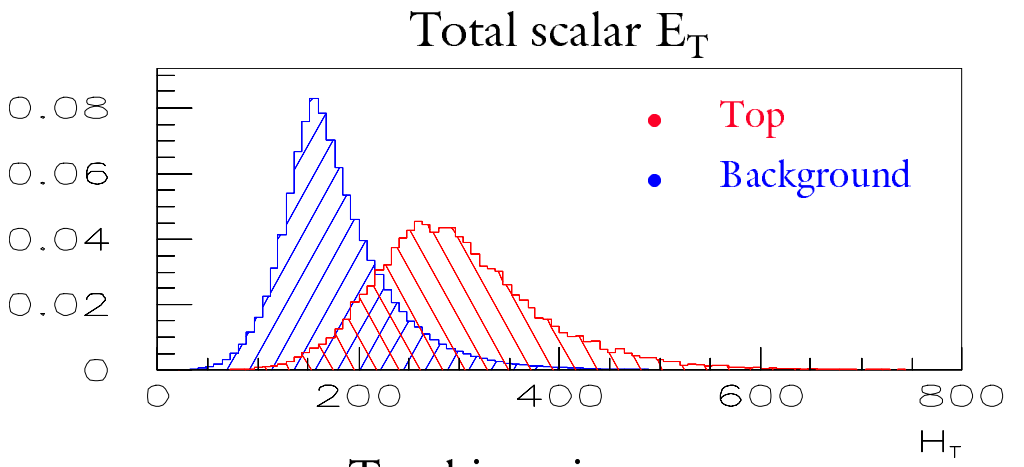
- Top events are:
  - harder
  - more spherical (less planar)
  - more central
  - with stiffer non-leading jets {*i.e.*, jets 3,4,5…}

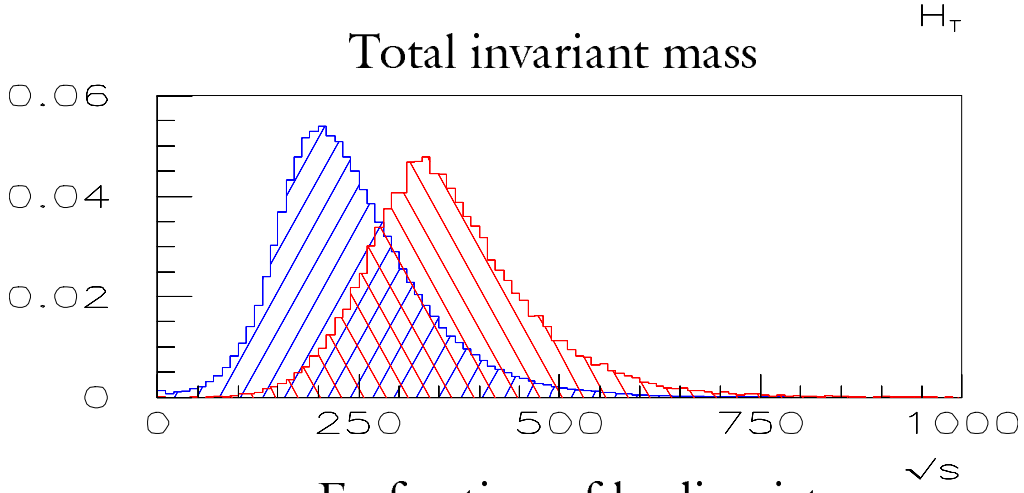
  than QCD multijets.

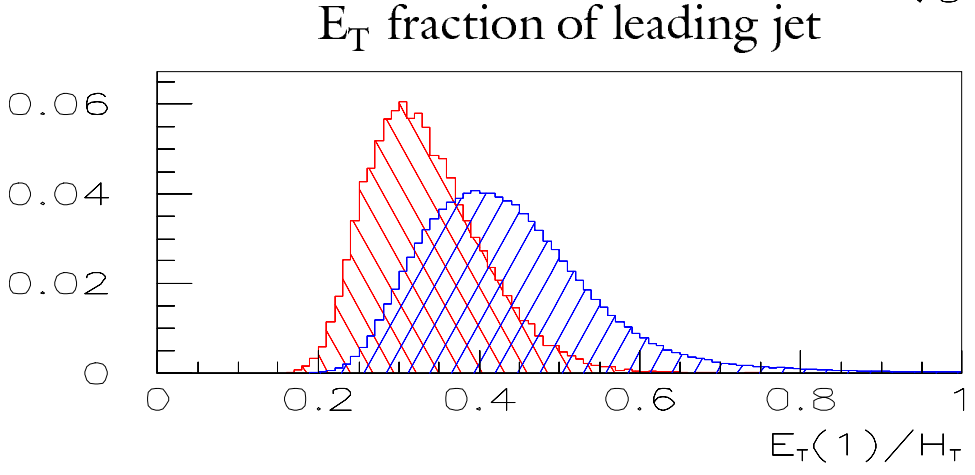
- Choose ~two kinematic vars for each property:
  - H_T Total scalar E_T
  - $\sqrt{s}$ Total invariant mass
  - $E_{T(1)}/H_T$ E_T fraction carried by leading jet
  - $\mathcal{A}$ Aplanarity
  - $\mathcal{S}$ Sphericity
  - $\mathcal{C}$ Centrality = H_T / H_E
  - η_{RMS} weighted RMS in η
  - $|\eta_5\ \eta_6|$ ηs of 5th & 6th jets
  - H_T^{3j} Total scalar E_T of non-leading jets
  - N_j^{ave} Threshold weighted number of jets
  - $\sqrt{E_{T(5)}\ E_{T(6)}}$ Average E_T of 5th & 6th jets

Variables Sensitive to Hardness of the Event

- $H_T = \Sigma E_{T(i)} \Rightarrow$ total scalar E_T
- $\sqrt{s} \Rightarrow$ total invariant mass
- $E_{T(1)}/H_T \Rightarrow E_T$ fraction carried by leading jet

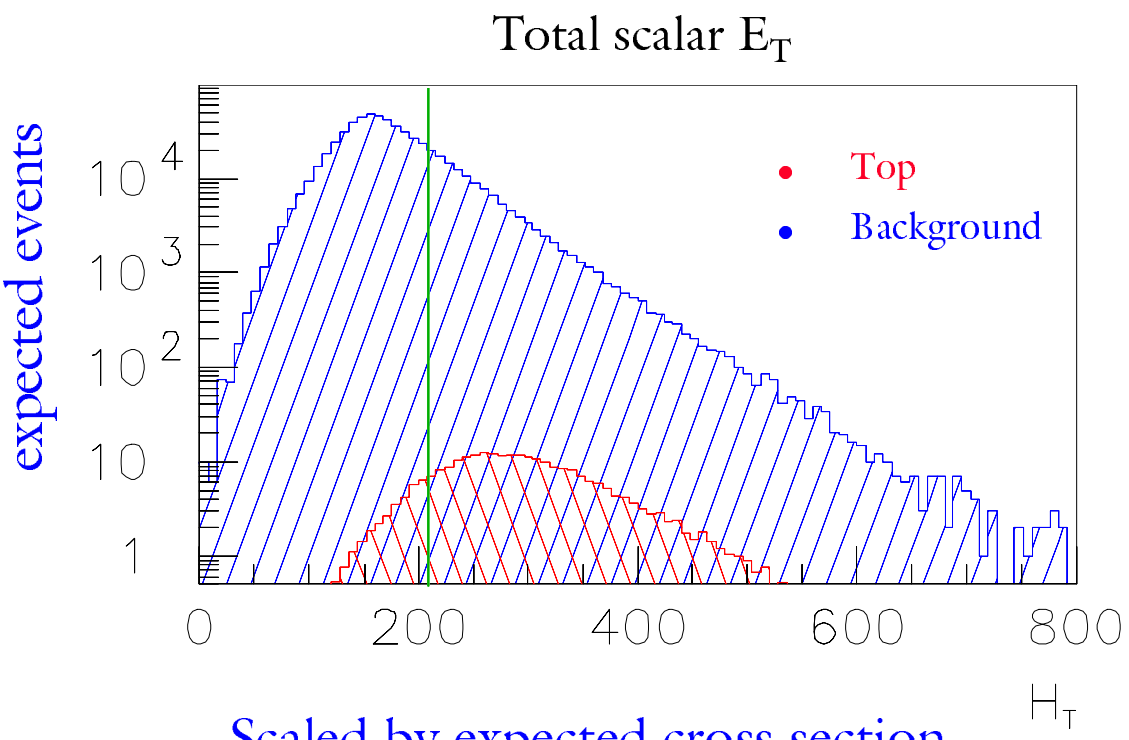






Total Scalar E_T Variable

- $H_T = \Sigma E_{T(i)} \Rightarrow$ total scalar E_T



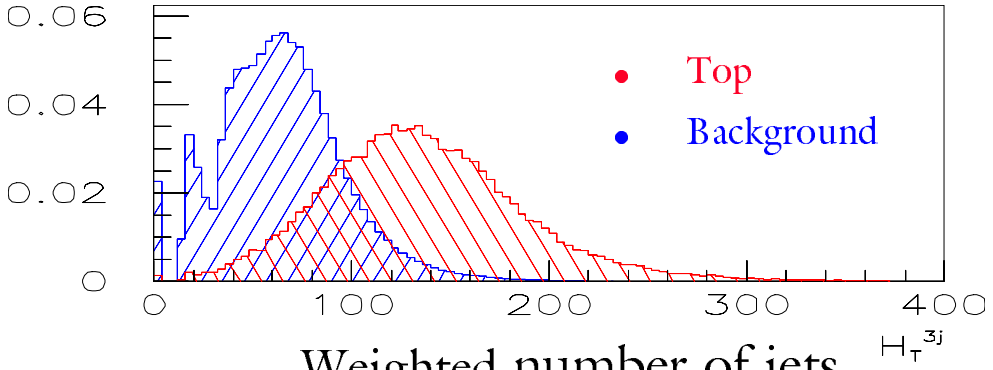
Scaled by expected cross section.

- Only modest discrimination from any single variable!

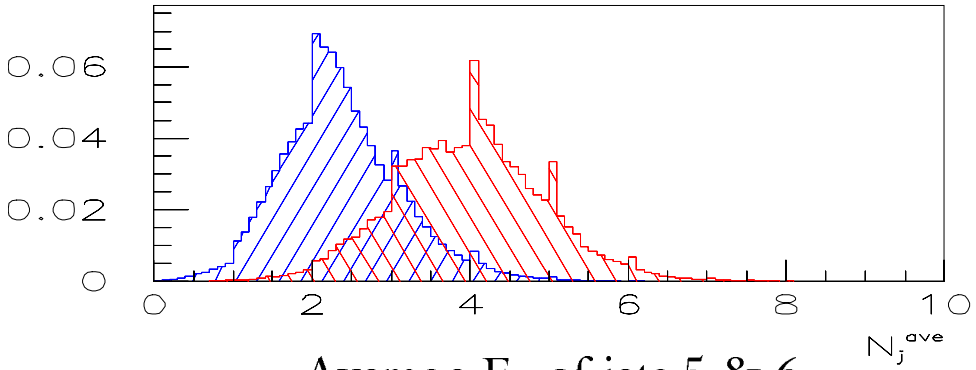
Variables Sensitive to Additional Radiation

- $H_T^{3j} = \Sigma E_{T(i)} - E_{T(1)} - E_{T(2)} \Rightarrow$ Scalar E_T of non-leading jets
- $N_j^{ave} = \frac{\int E_T \cdot N_j(E_T)\, dE_T}{\int E_T\, dE_T} \Rightarrow$ Weighted number of jets
- $E_{T(5,6)} = \sqrt{E_{T(5)}\, E_{T(6)}} \Rightarrow$ Average E_T of 5th & 6th jets

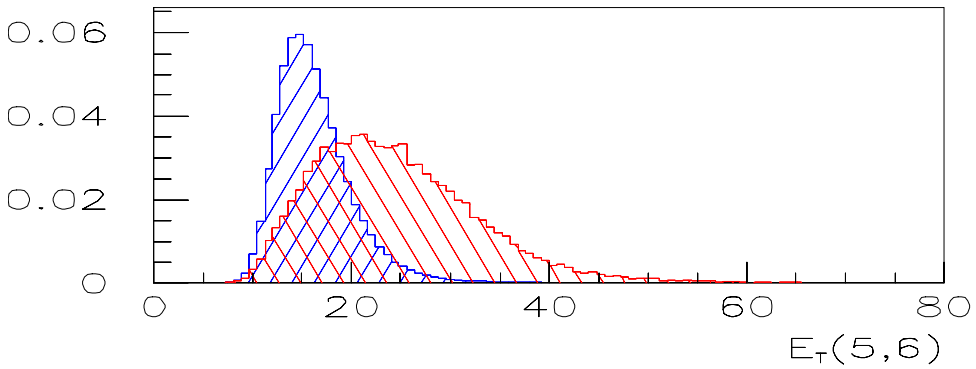
Scalar E_T of non-leading jets



Weighted number of jets



Average E_T of jets 5 & 6

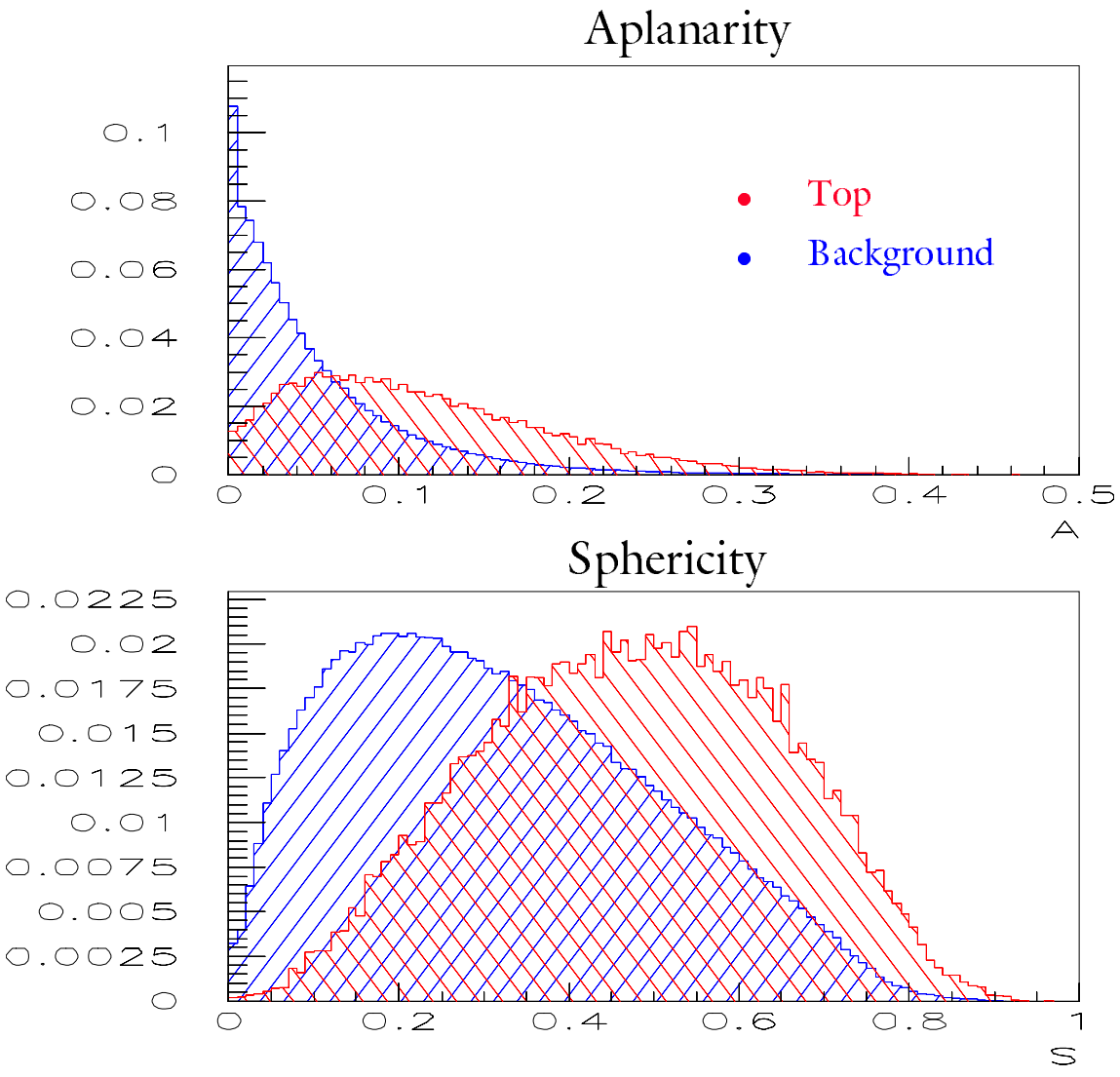


Aplanarity and Sphericity

$M^{ab} = \Sigma p^a p^b / \Sigma p^2$, where (Q_1, Q_2, Q_3) are the M^{ab} eigenvalues with $0 \le Q_1 \le Q_2 \le Q_3$

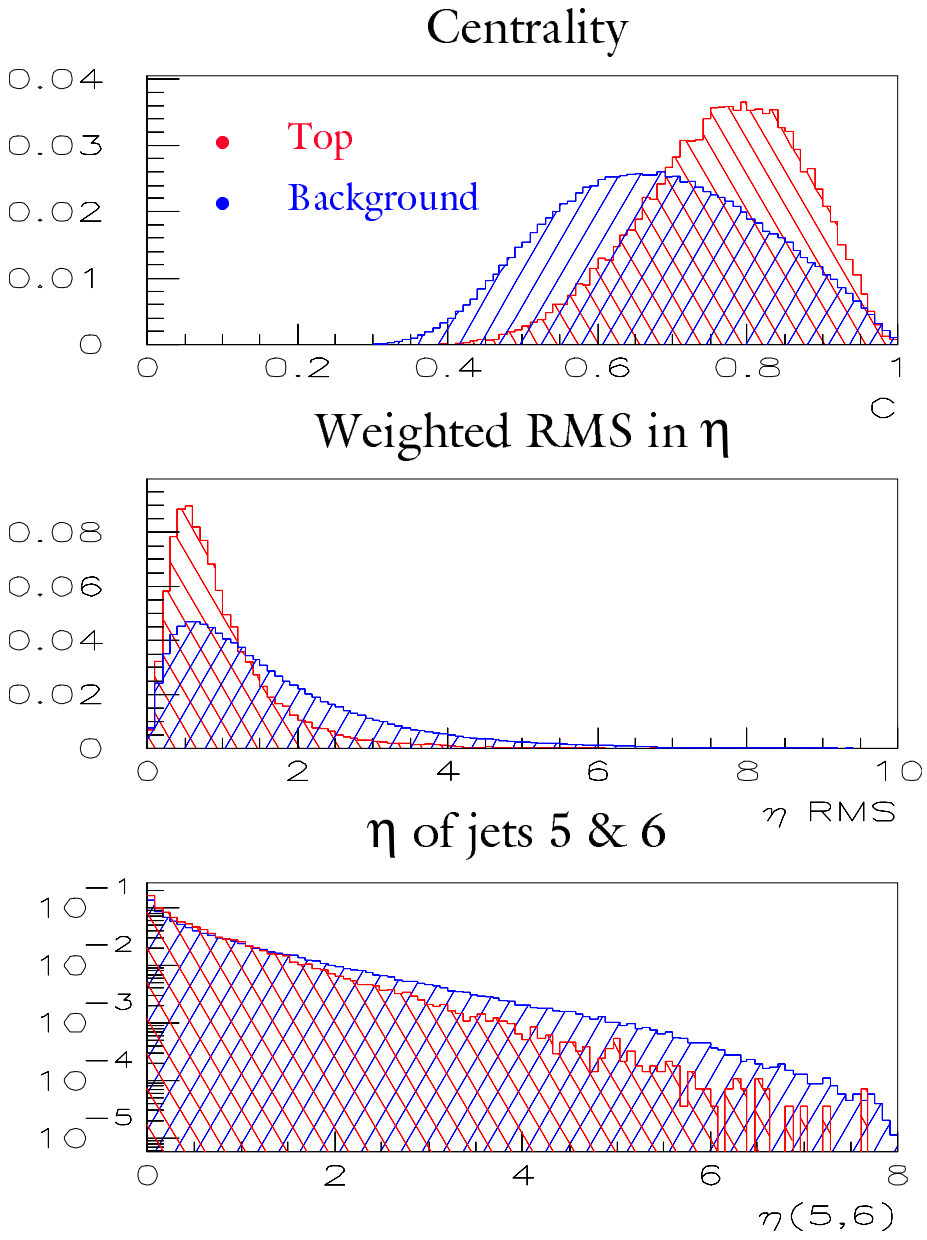
$\mathcal{A} = 3/2 \; x \; (Q_1)$

$\mathcal{S} = 3/2 \; x \; (Q_1 + Q_2)$



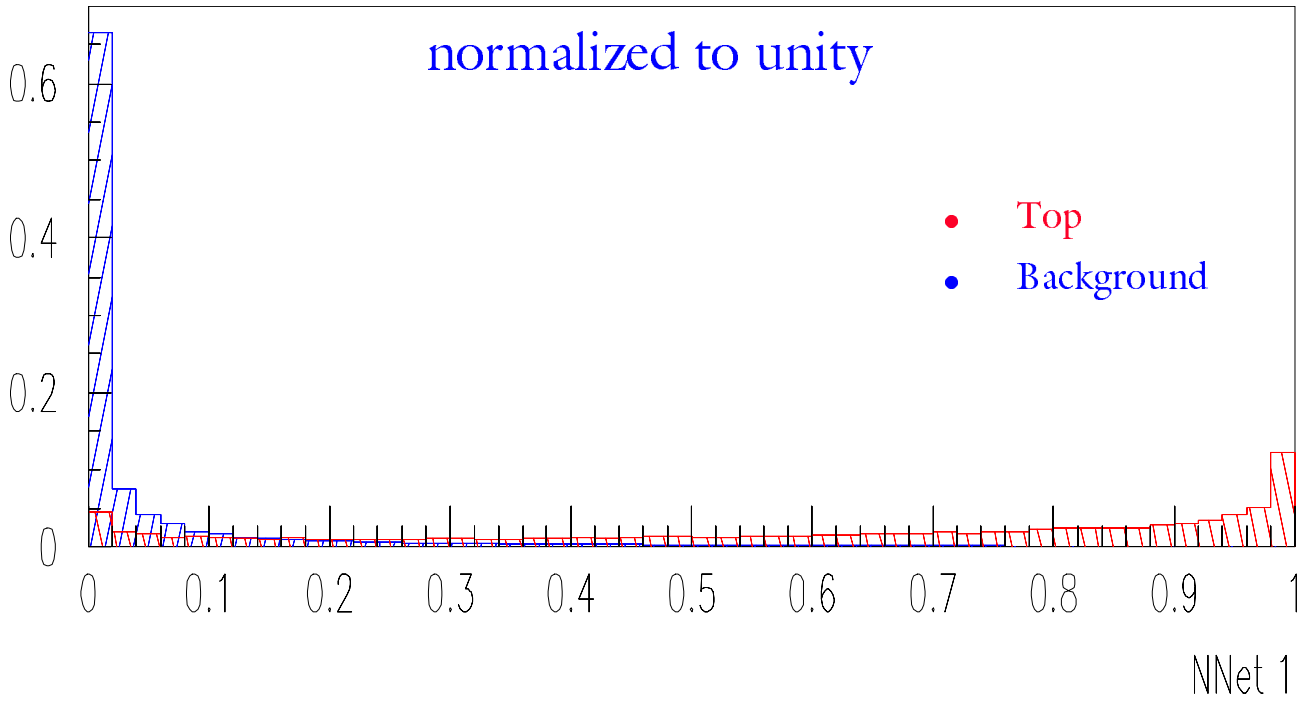
Variables Sensitive to Rapidity Distributions

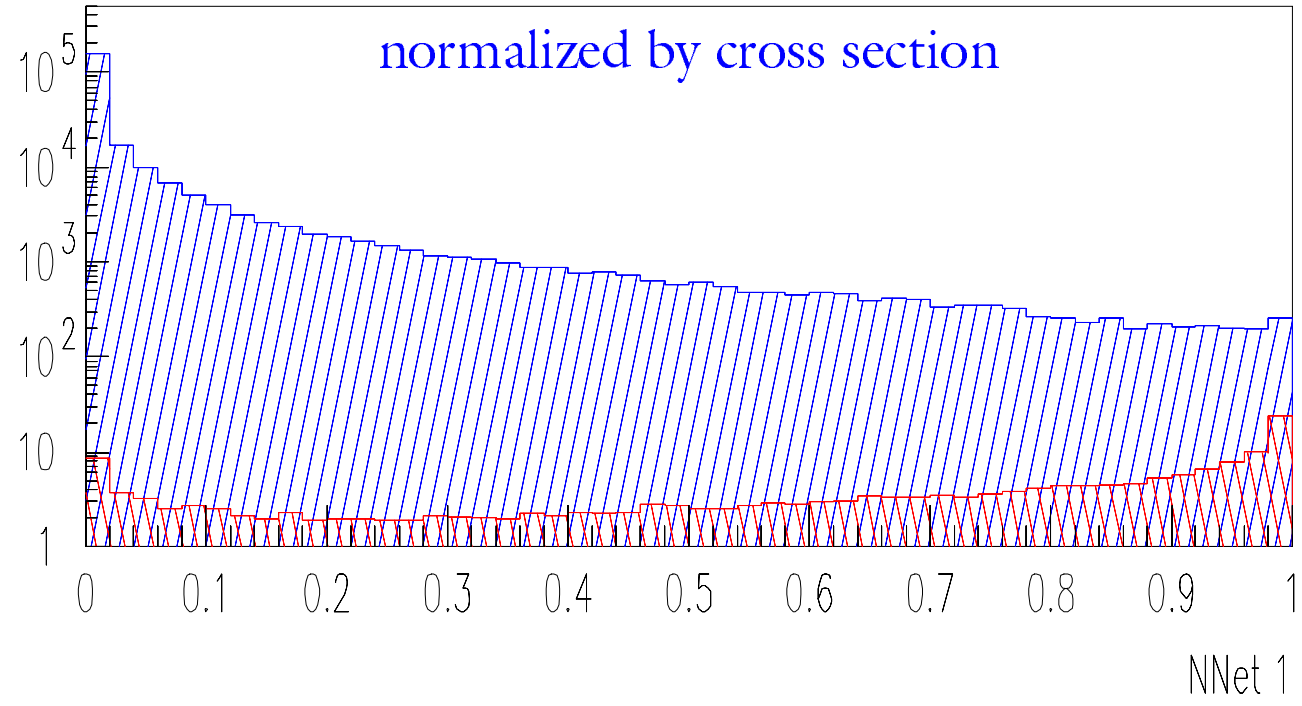
- $\mathscr{C} = \Sigma E_{T(i)}/\Sigma E_{(i)} \Rightarrow$ Centrality
- $\eta_{RMS} \Rightarrow$ Weighted RMS in η
- $\eta_{(5,6)} = |\eta_5\ \eta_6| \Rightarrow \eta$ of jets 5 & 6

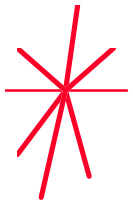


First Neural Network

- Neural Network based on 11 kinematic variables.

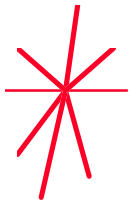






Summary of Kinematic Variables

- Eleven kinematic variables, based on the expected characteristics of top and background, are selected.
- A Neural Network, based on these variables, gives good discrimination between top and background.
- Discrimination based on this Neural Network, alone, is insufficient to extract a significant top signal.
- Add three new and orthogonal parameters.



Likelihood Parameter

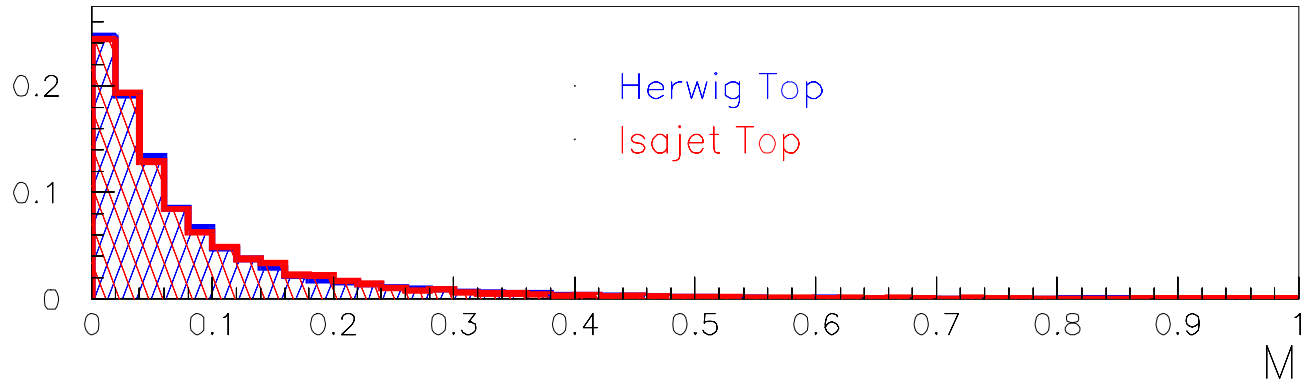
- Parameter based on top decay model.
- Use M as an input variable to the **Neural Network.**
- Loop to find best jet combination with smallest M.
- χ^2 -like object :

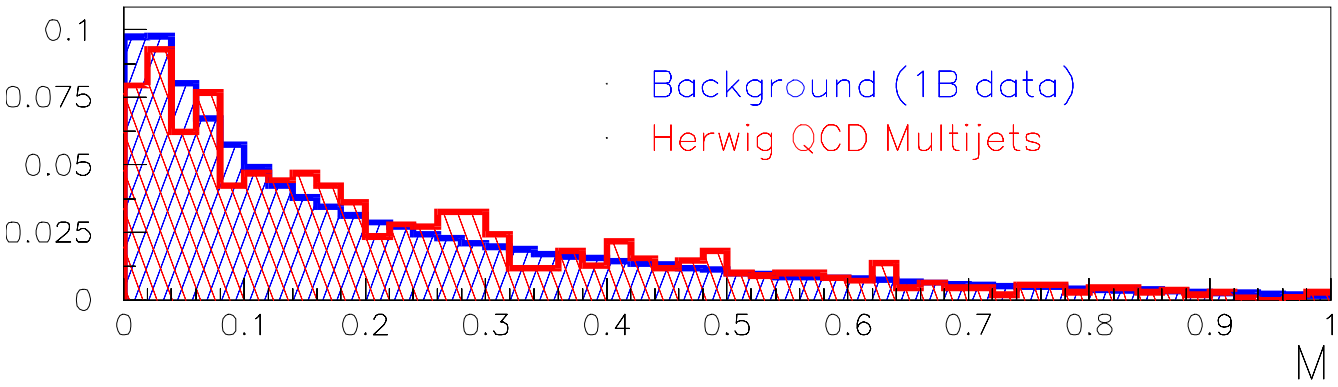
$$M = \frac{(M_{W1} - 80)^2}{16^2} + \frac{(M_{W2} - 80)^2}{16^2} + \frac{(M_{T1} - M_{T2})^2}{62^2}$$

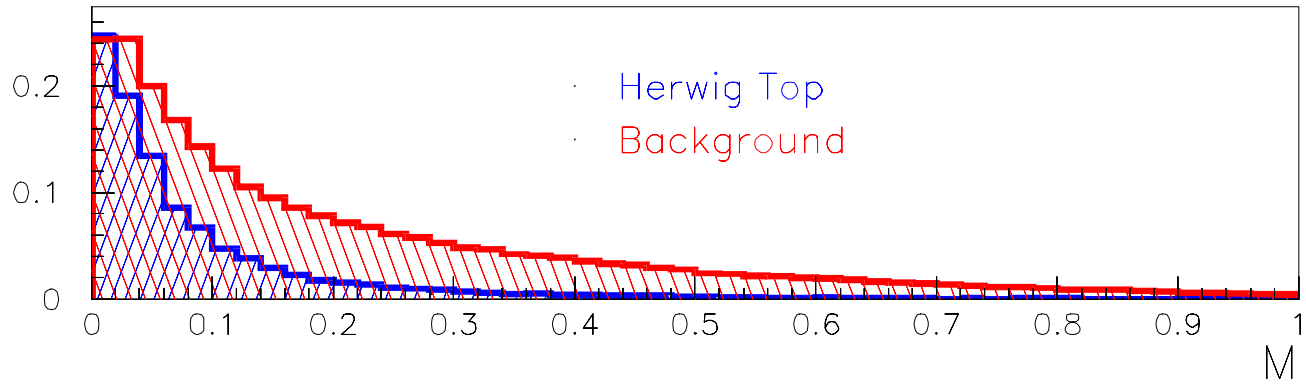
- Handle tagged & untagged events in the same way; μ-tagged events are **not** required to be *b* jets.

Likelihood Variable

- Good agreement between Herwig top and Isajet top.
- Good agreement between Herwig QCD multijets and data.
- Discriminates between top and background.

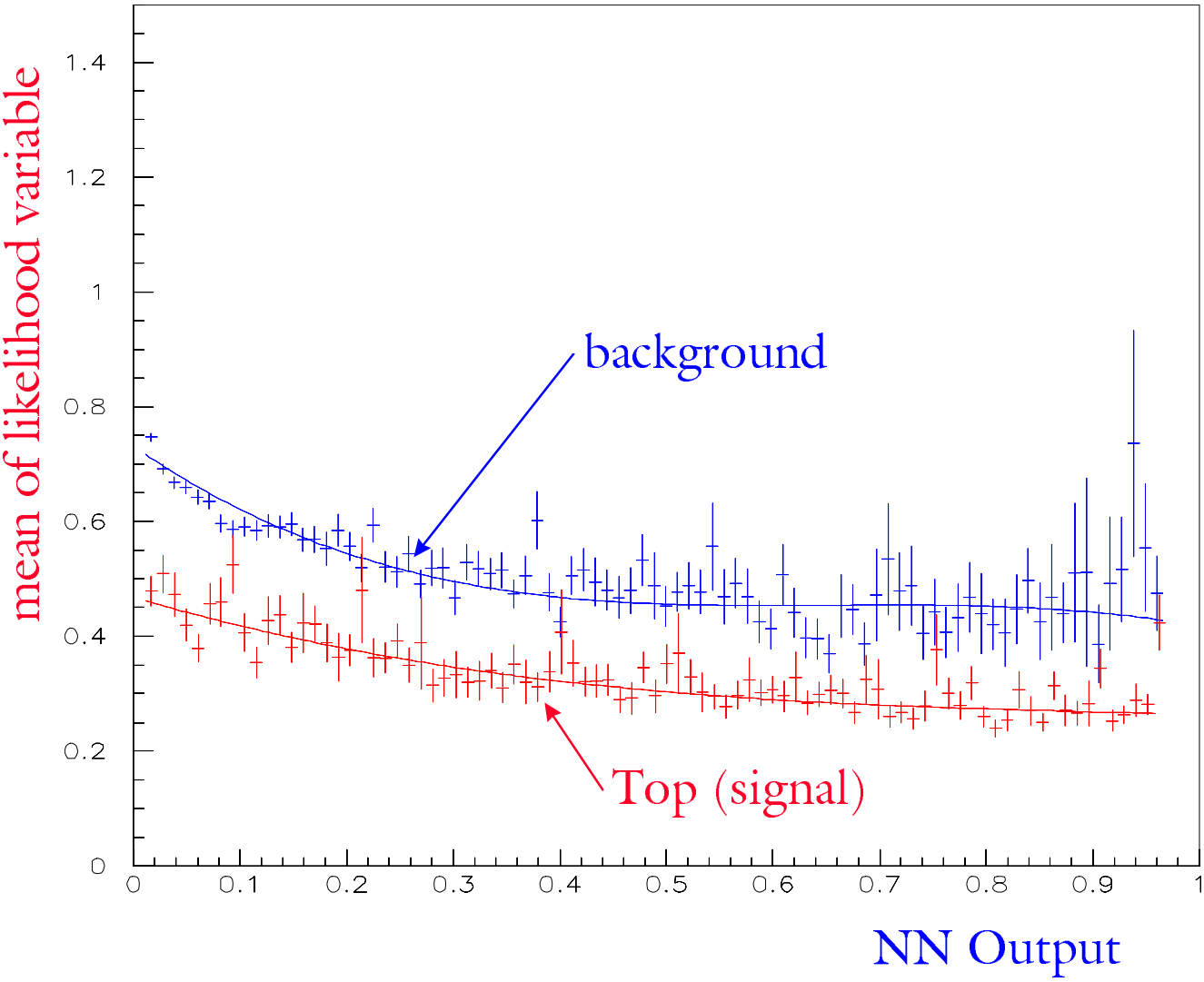


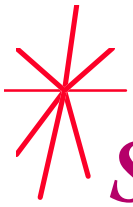




Likelihood Variable

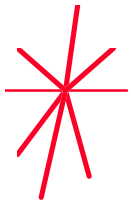
- Likelihood variable is reasonably independent of the kinematic variables (NN output).
- Essentially no loss of discrimination between background region ($\mathcal{D}_{NN} \Rightarrow 0$) and the signal region ($\mathcal{D}_{NN} \Rightarrow 1$).





Summary of Likelihood Parameter

- The shape of the likelihood parameter does not depend on the choice of signal generator.
- Herwig QCD multijets Monte Carlo and background (data) agree well.
- The likelihood parameter, based on the top decay model, is useful in differentiating top from background.

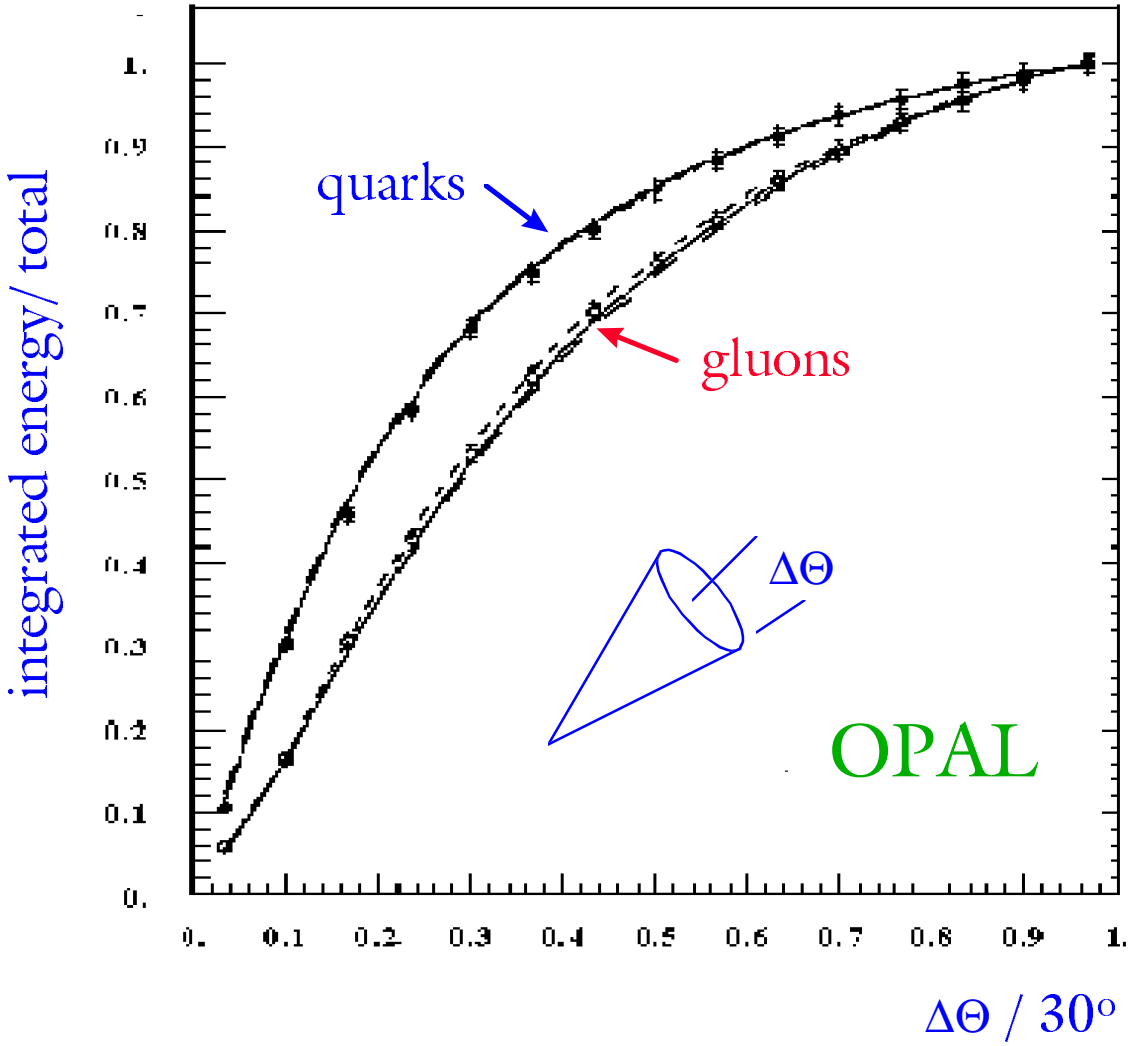


Differences in Widths of Jets

- Background to Top is QCD Multijets.
- Simplest Case:
  - Top ≡ 6 quark-jets
  - QCD Multijets ≡ 6 gluon-jets
- Will:
  - show the properties of jet widths in $p\bar{p}$ collisions
  - examine quark/gluon differences in jet widths ⇒ Herwig Monte Carlo
  - find a method for discrimination per event
- Define jet width: E_T weighted RMS in η & ϕ, in quadrature, for an 0.5-cone jet.

Quark/Gluon Differences:

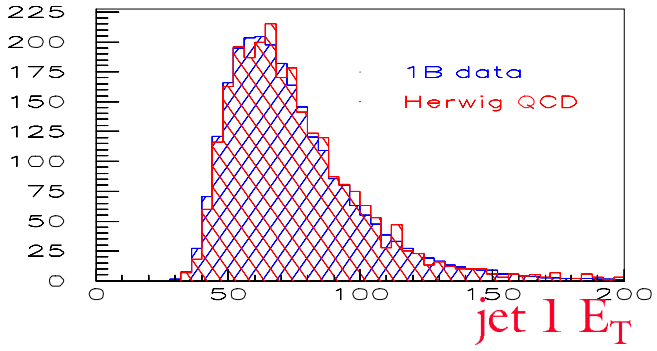
- Direct observation at LEP & AMY.
- LEP in good agreement with HERWIG.



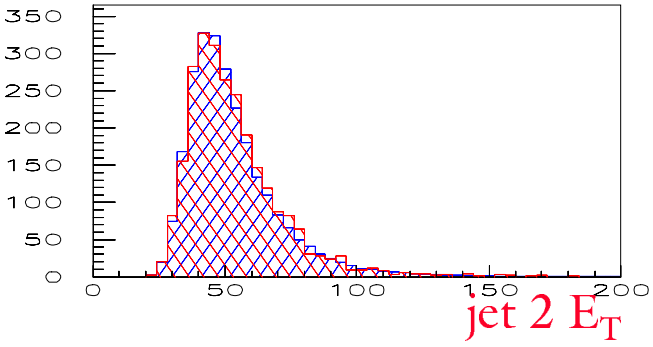
- Use HERWIG to establish expected widths for quark and gluon jets.

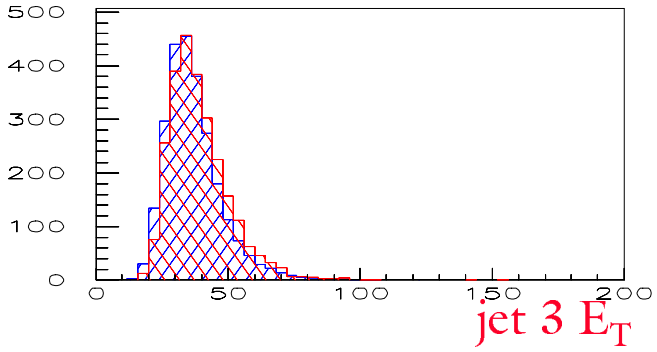
Herwig QCD/Data Comparison

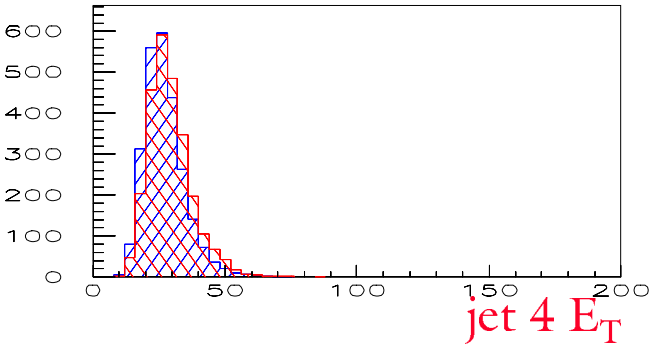
- jet E_T distributions:



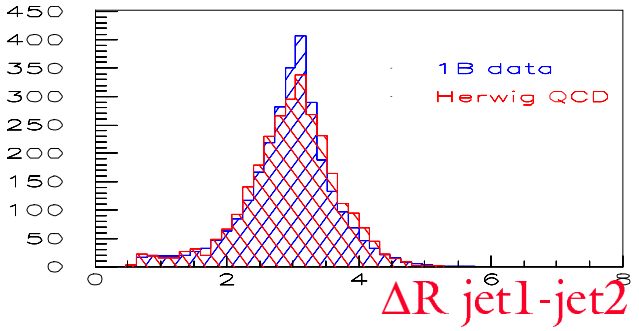
jet 1 E_T


jet 2 E_T

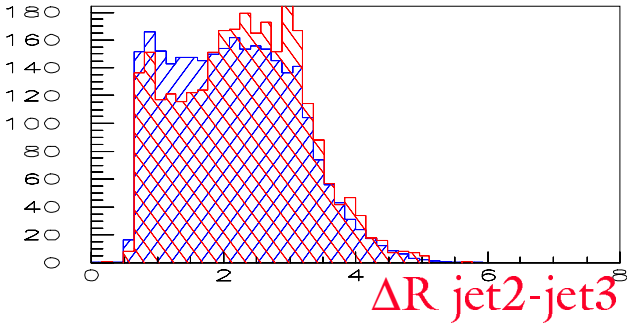

jet 3 E_T

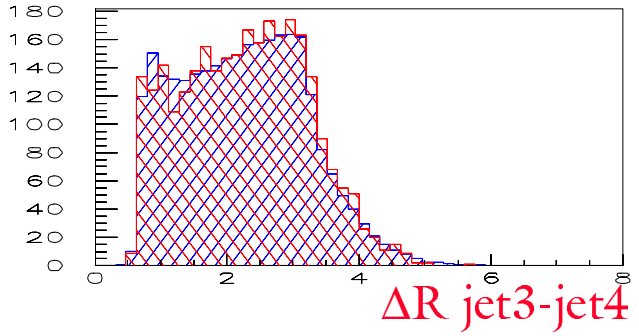

jet 4 E_T

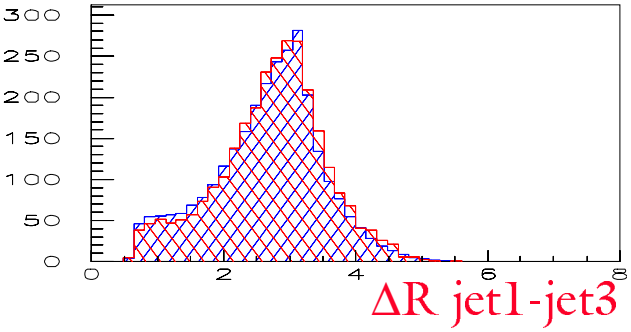
- the ΔR between jets:



ΔR jet1-jet2


ΔR jet2-jet3

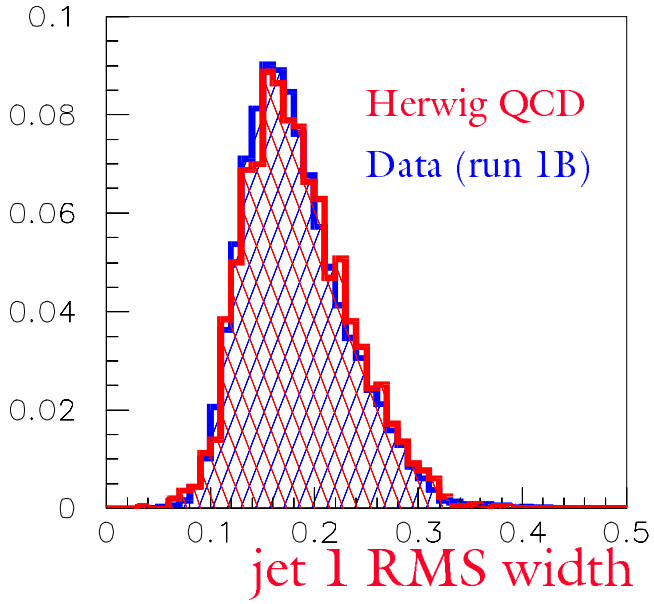

ΔR jet3-jet4

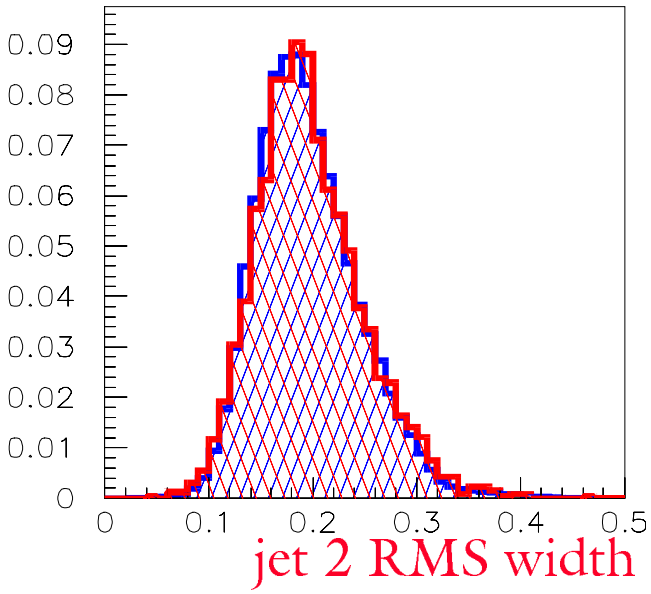

ΔR jet1-jet3

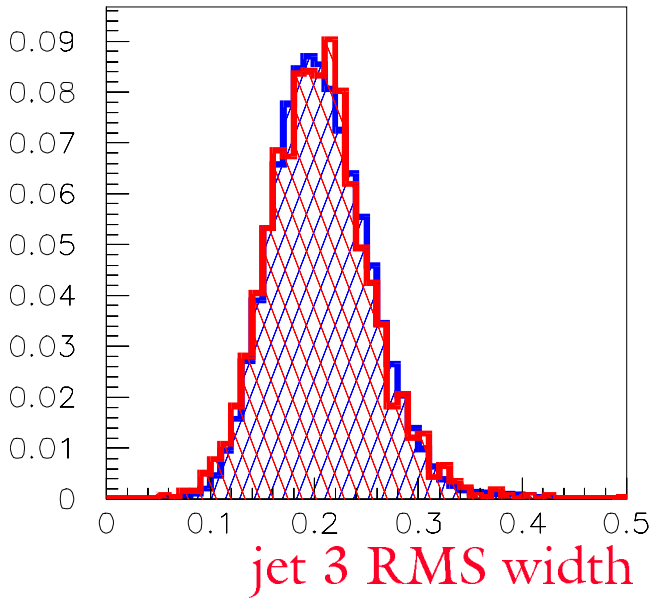
- Herwig QCD & Data: reasonable agreement

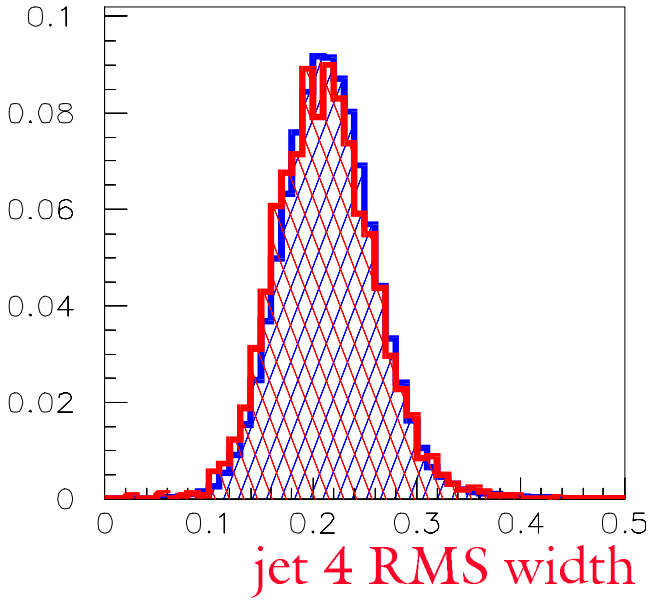
Herwig QCD/Data Comparison

- jet RMS width distributions:
  (Herwig was broadened by ~3% to match data)





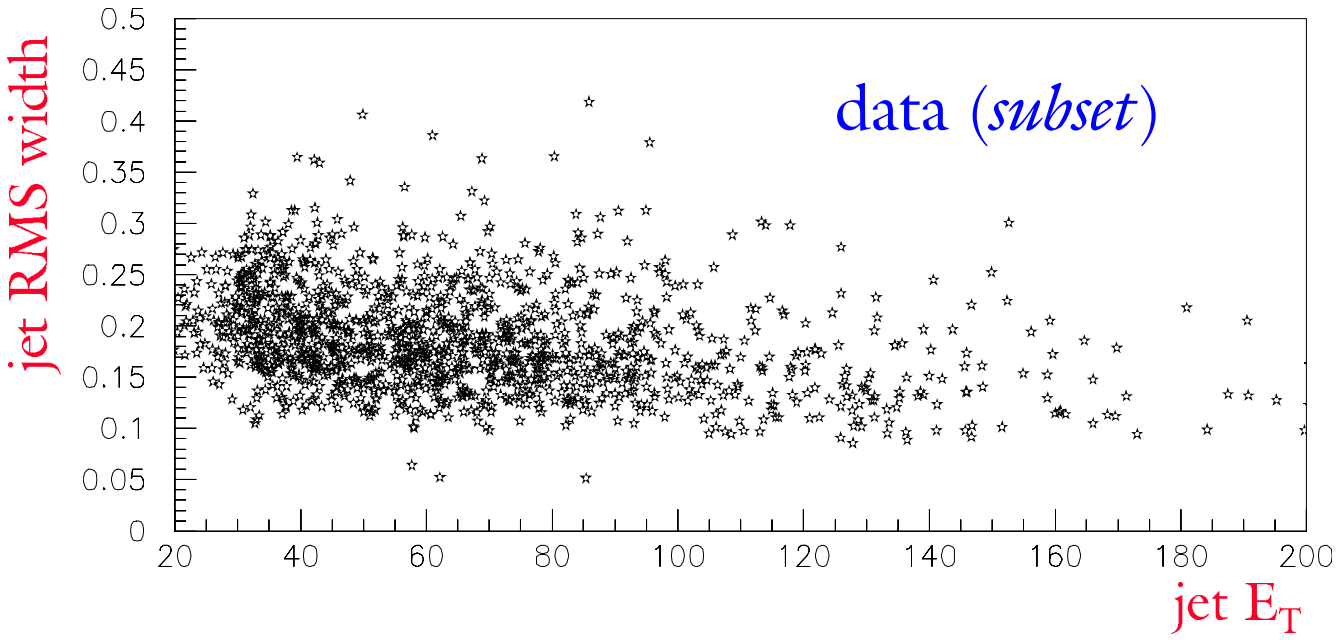


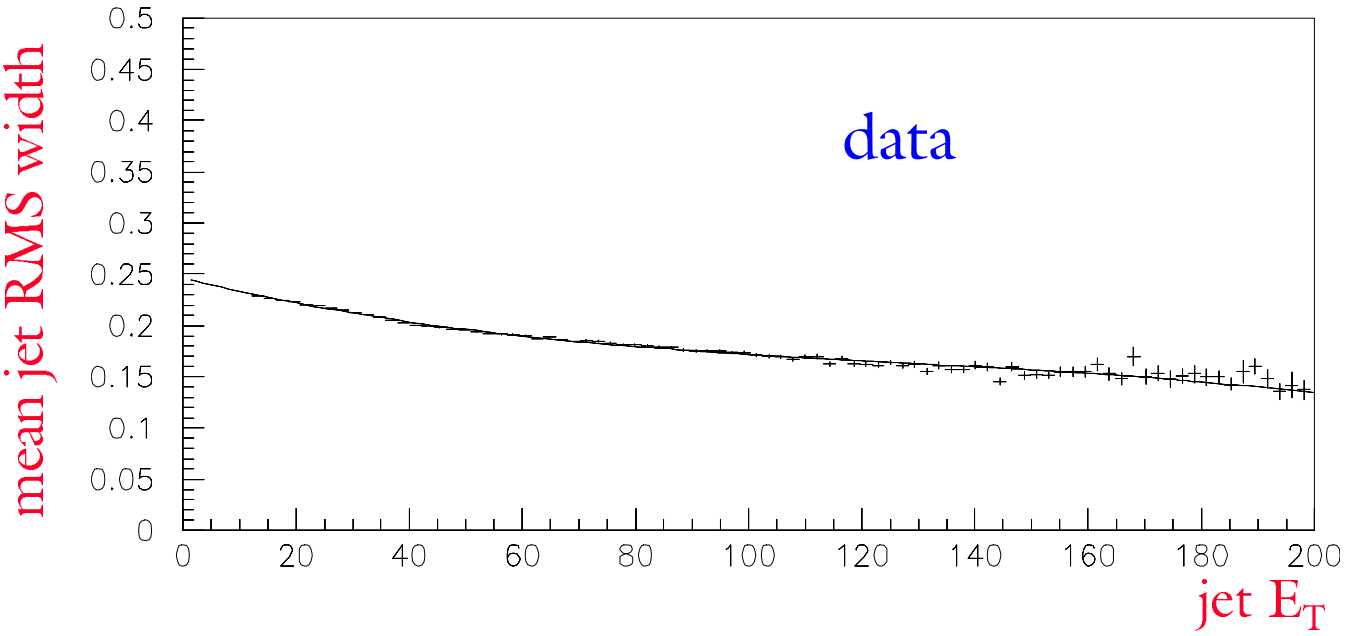


- Good agreement between Herwig QCD and data.

Width vs Jet E_T

- jet RMS width *vs* jet E_T:

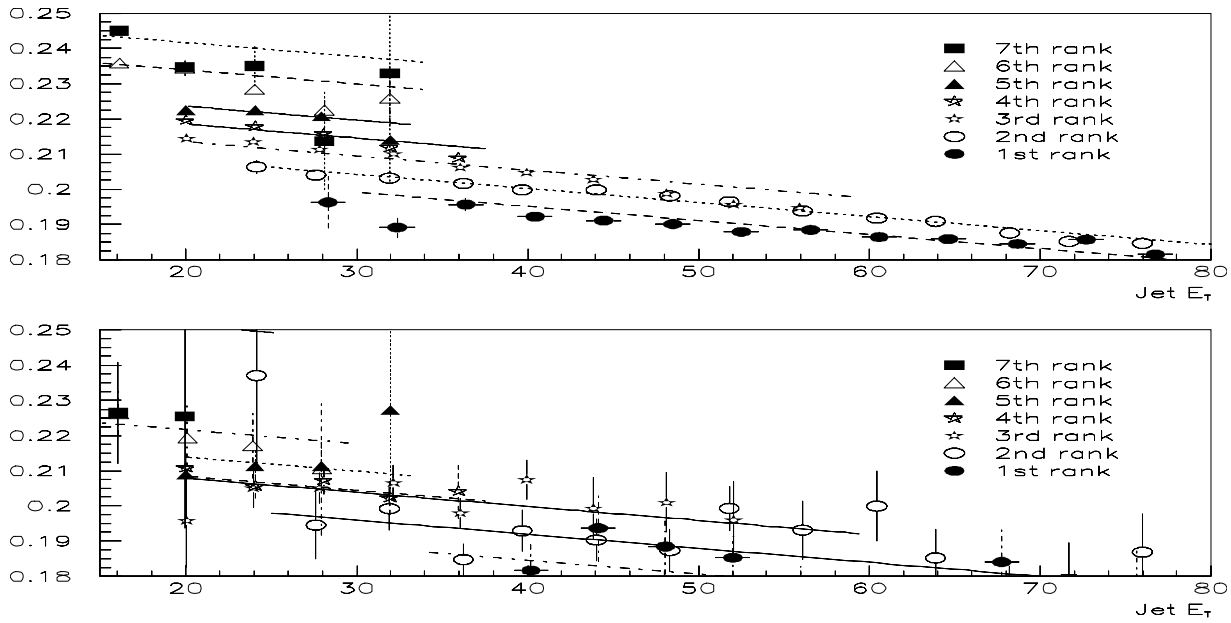




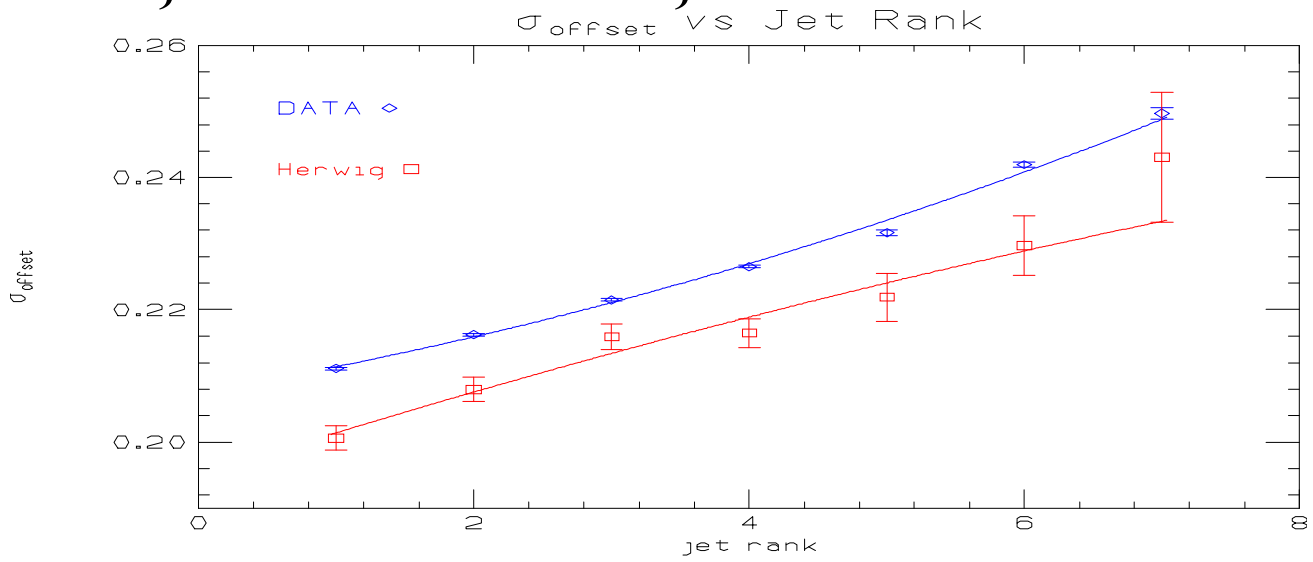
- Boost along jet axis narrows jets.
- Choose: E_T, $\Delta\phi$, $\Delta\eta$ (*Lorentz invariant along beam*).

Width vs Jet Rank

- average jet RMS width *vs.* jet E_T, by rank:



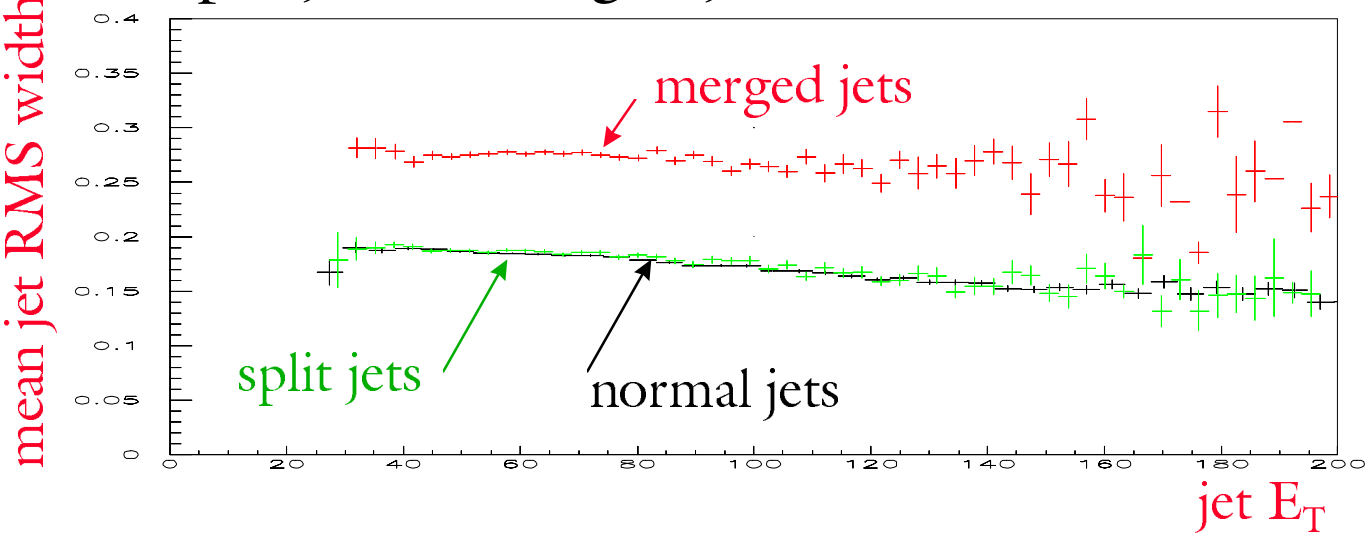
- jet RMS offsets *vs.* jet rank:



- Herwig QCD & Data show similar rank dependence.
- Herwig about ~3% too narrow.

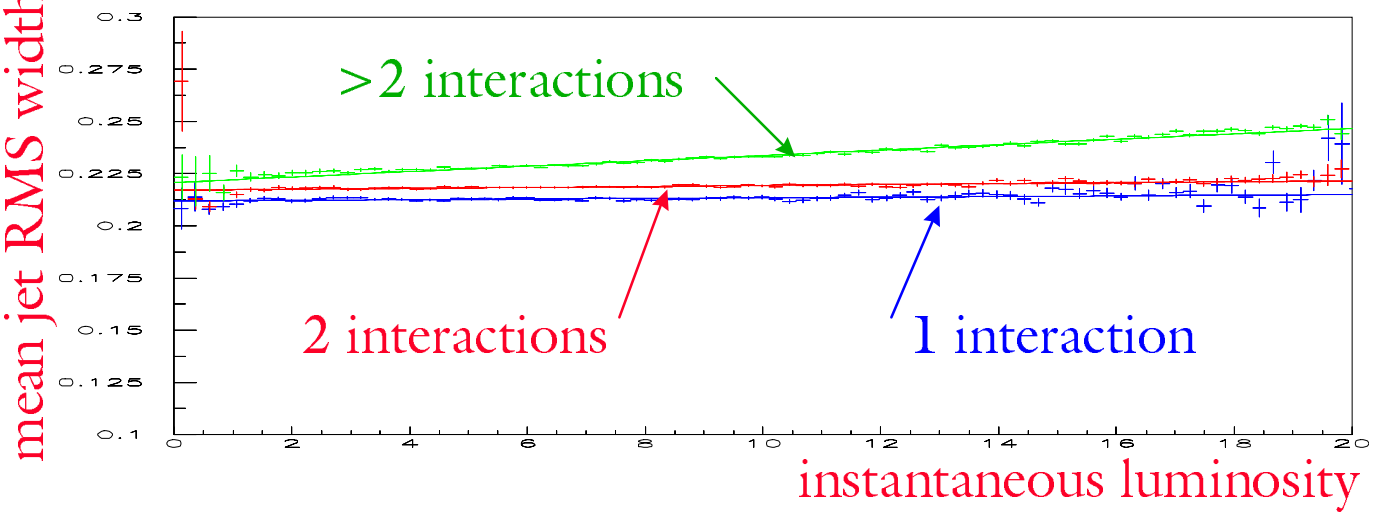
Jet Quality Cuts

- Split jets vs merged jets:



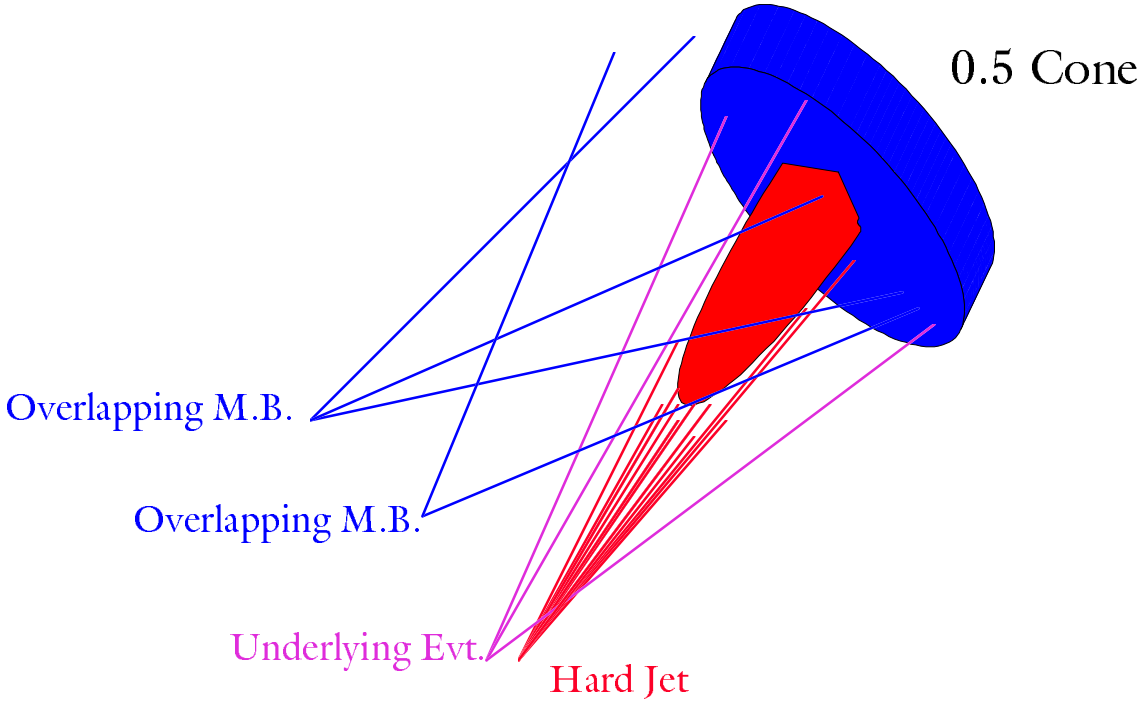
- split jets OK, but **not** merged jets
- merged jets (~8%) are rejected in this analysis.

- Multiple interactions:



- overlapping minbias events → correction needed.
- underlying event must be important, and is included in Herwig Monte Carlo.

Underlying Event & Overlapping Minbias Events

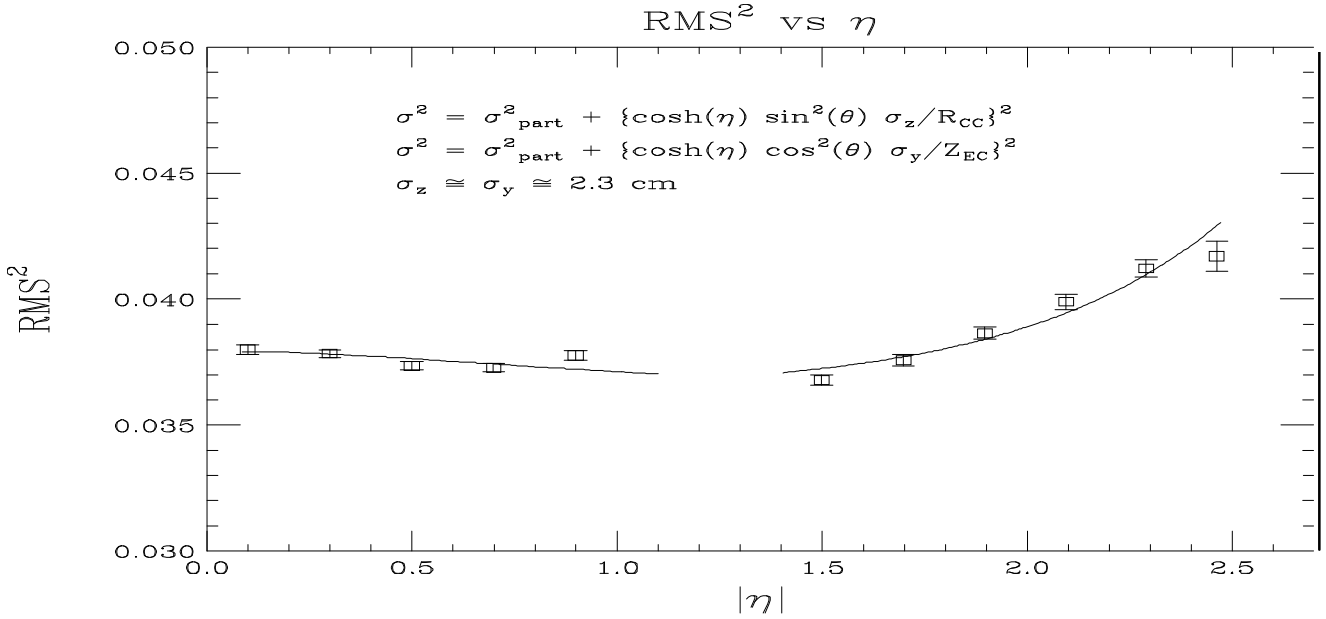


$$\text{Measured RMS}^2 = \frac{E_{T_{jet}} \bullet \sigma^2_{jet} + E_{T_{MB}} \bullet \sigma^2_{MB}}{E_{T_{jet}} + E_{T_{MB}}}$$

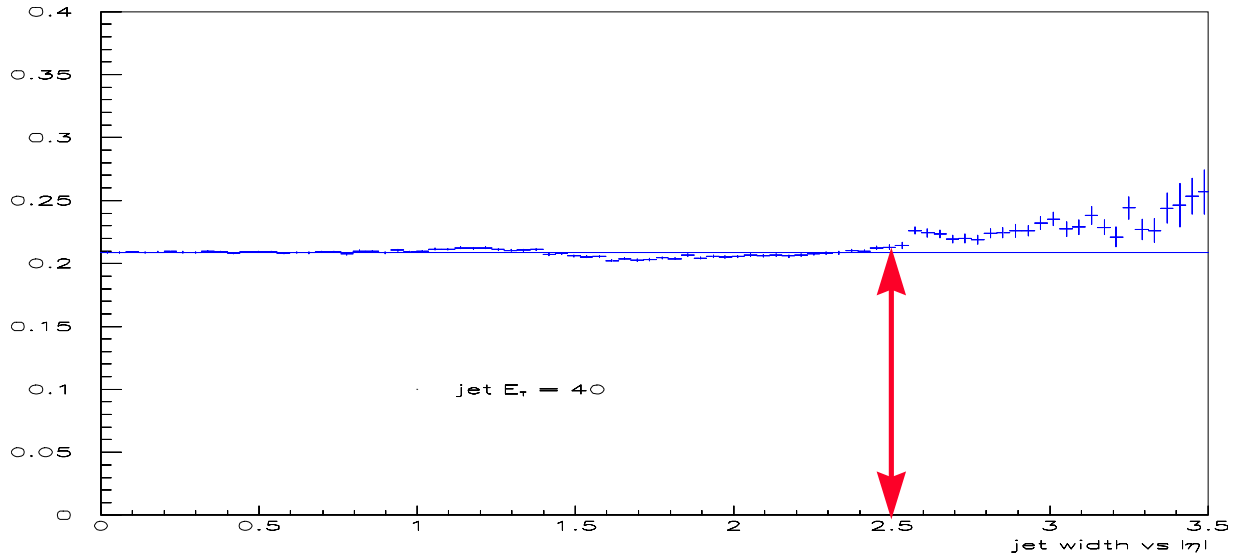
- overlapping minbias event can be subtracted.

Jet Quality Cuts (cont.)

- particle showering in detector:
  - constant spreading in real space {*i.e.* cm.}
  - but η bins contract in forward direction



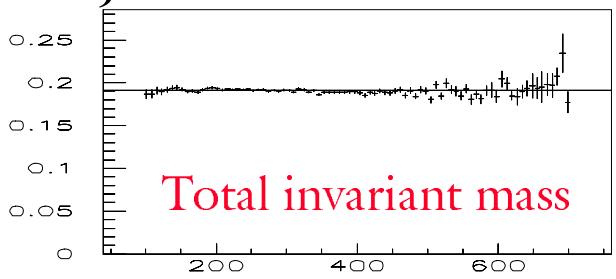
- jet RMS width *vs* |η|:

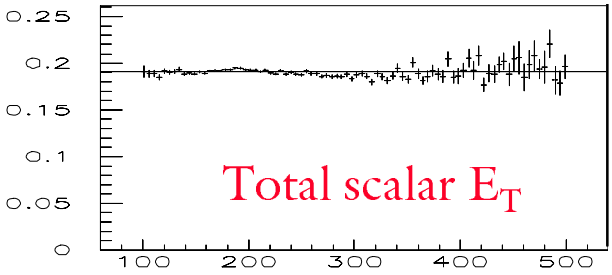


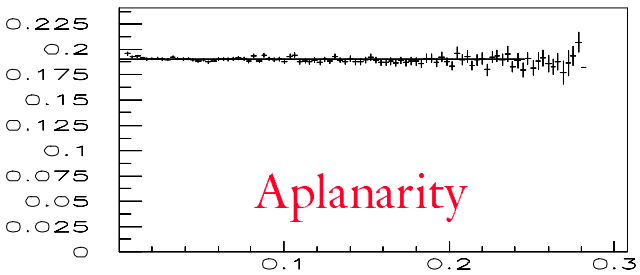
- Choose: $|\eta| \leq 2.5$

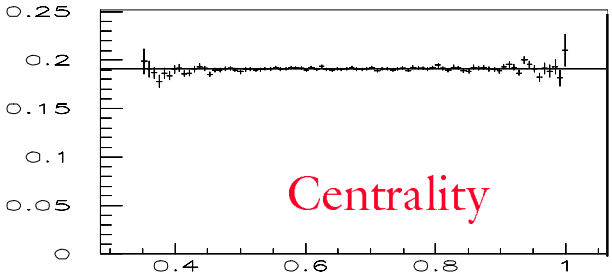
Jet Width Studies:

- jet width *vs.* kinematic variables:



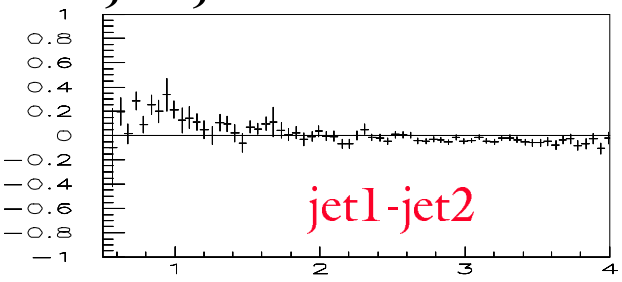


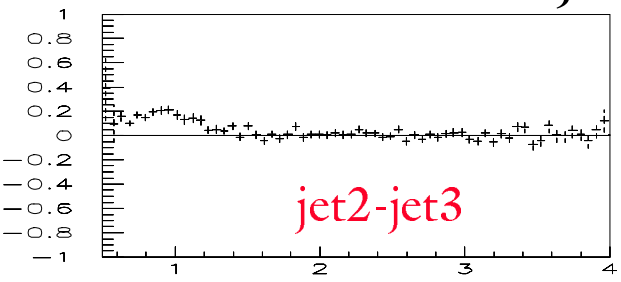


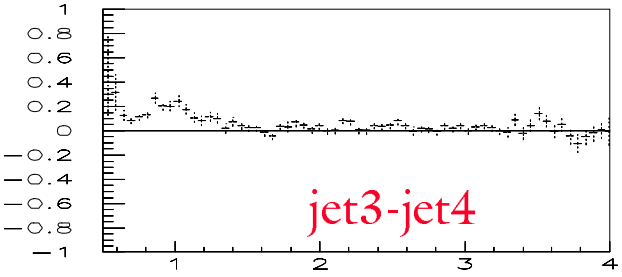


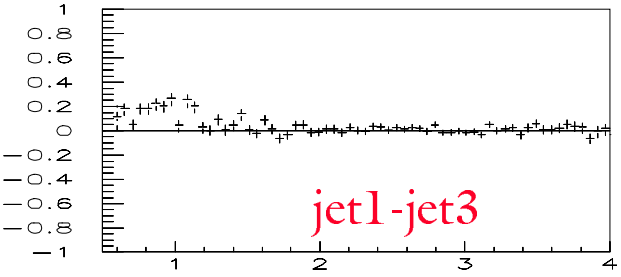
- Jet hadronization ~ independent of kinematic properties of the event {*i.e.* of hard scatter}

- jet-jet width correlations *vs.* ΔR between jets:





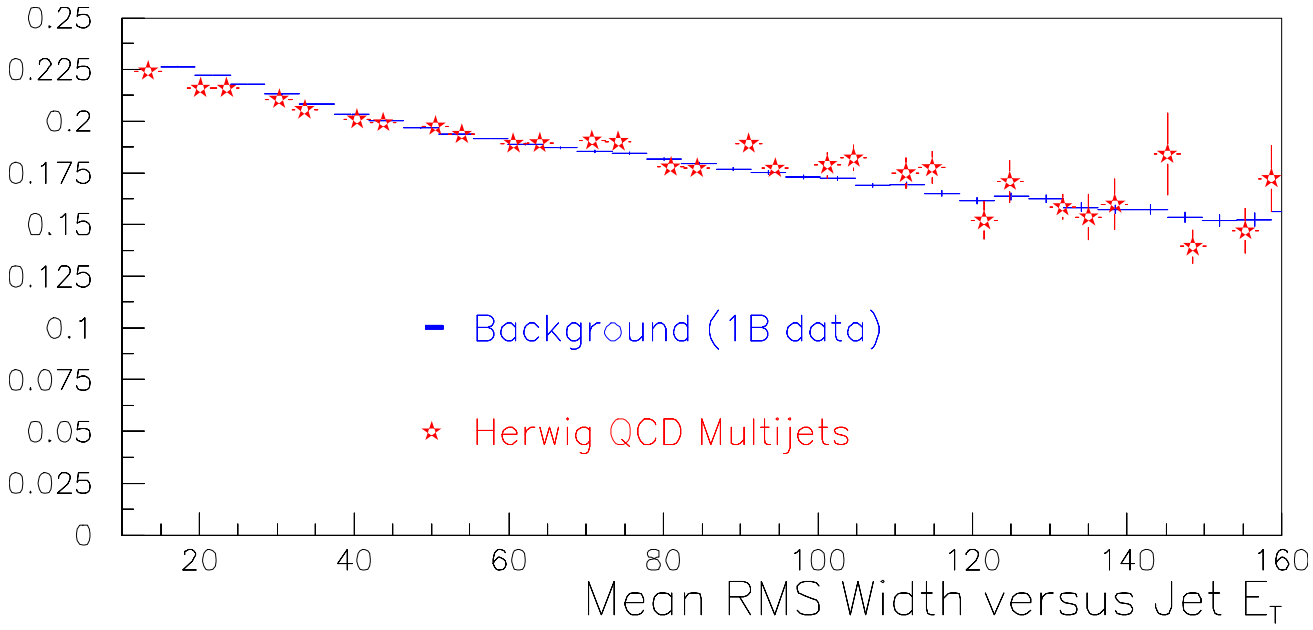


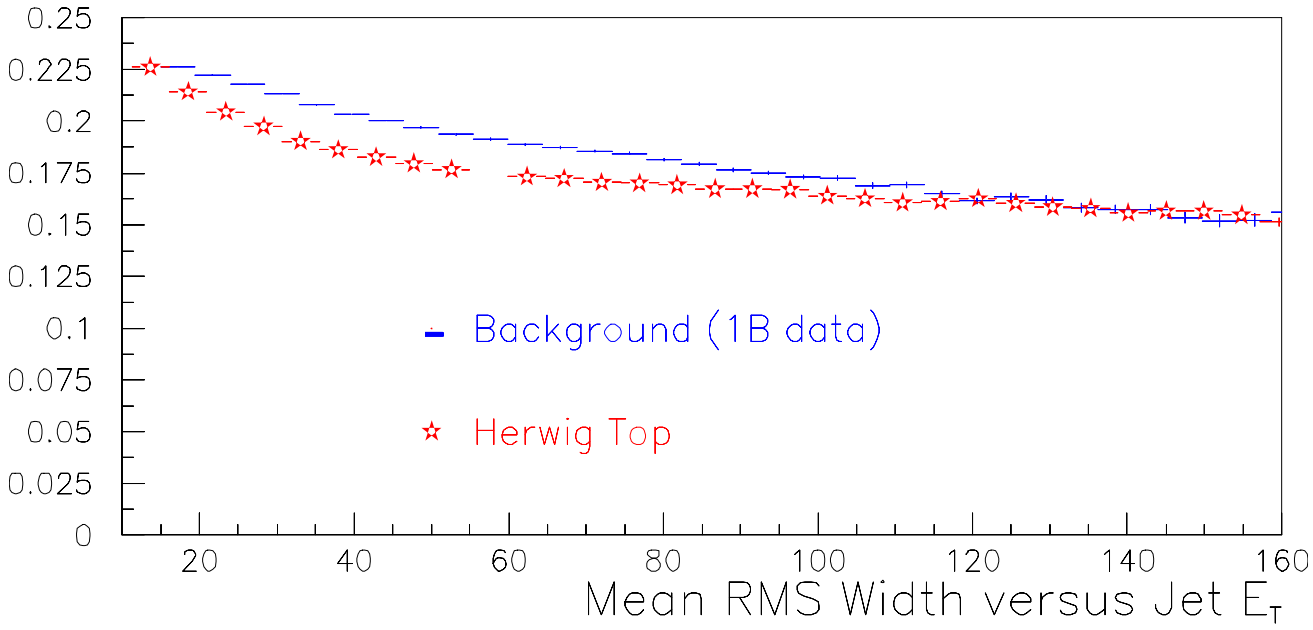


- correlations between widths of two jets is present when the jets are adjacent, otherwise small.

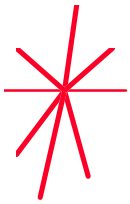
Herwig/Data Comparison

- mean jet RMS width *vs* jet E_T:





- Herwig top narrower than background.



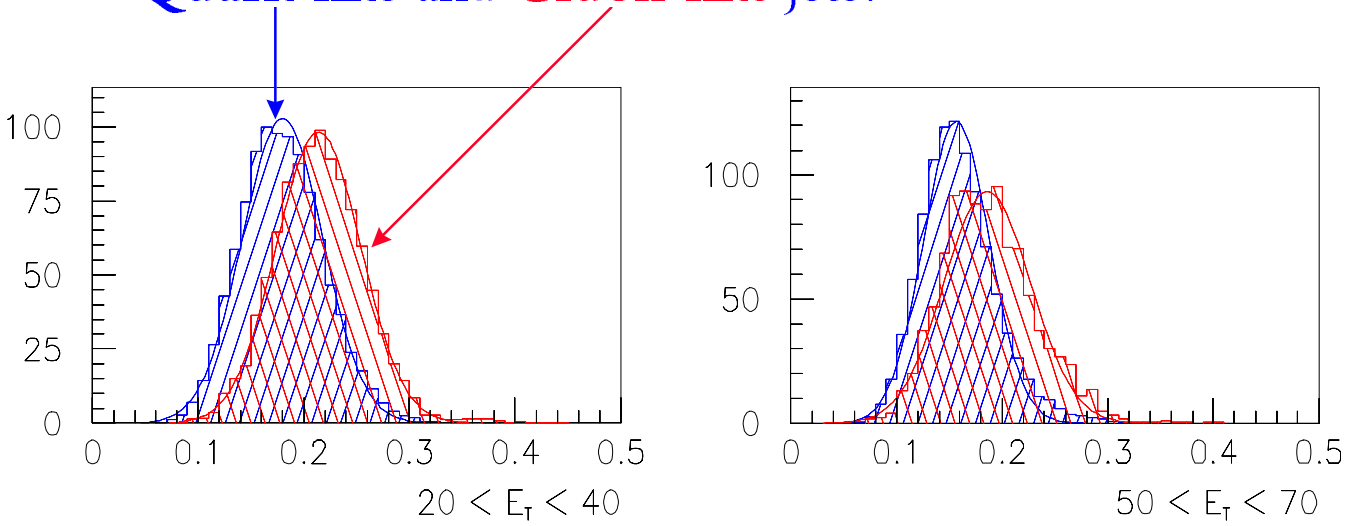
Herwig Quark/Gluon Selection

- Study Quark/Gluon differences *via* Herwig.
- **Difficult** to trace quarks → particles → jets in Herwig.
- Use Herwig Top: with 6 quarks (and extra gluons) in final state.
- Match quarks to jets:

- **define Quark-like jets:**
  - jet quality cuts: non-merged jets and $|\eta| < 2.5$
  - the ΔR between quark & jet < 0.05
  - the ΔE between quark & jet $< \sqrt{E_{quark}}$
- **define Gluon-like jets:**
  - jet quality cuts: non-merged jets and $|\eta| < 2.5$
  - **not** a quark-like jet
  - the ΔR between quarks & jet > 1

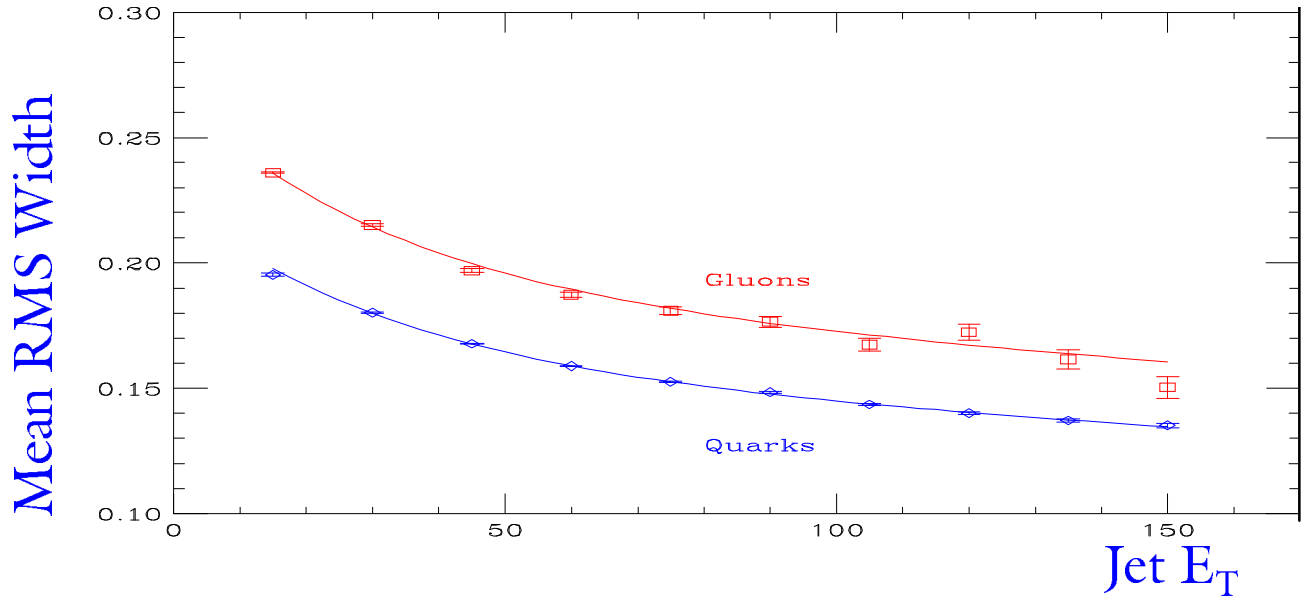
Herwig Quark/Gluon Selection

- For these definitions of **Quark-like** and **Gluon-like** jets:



- distributions are nearly Gaussian, which suggests using a Fisher Discriminant:

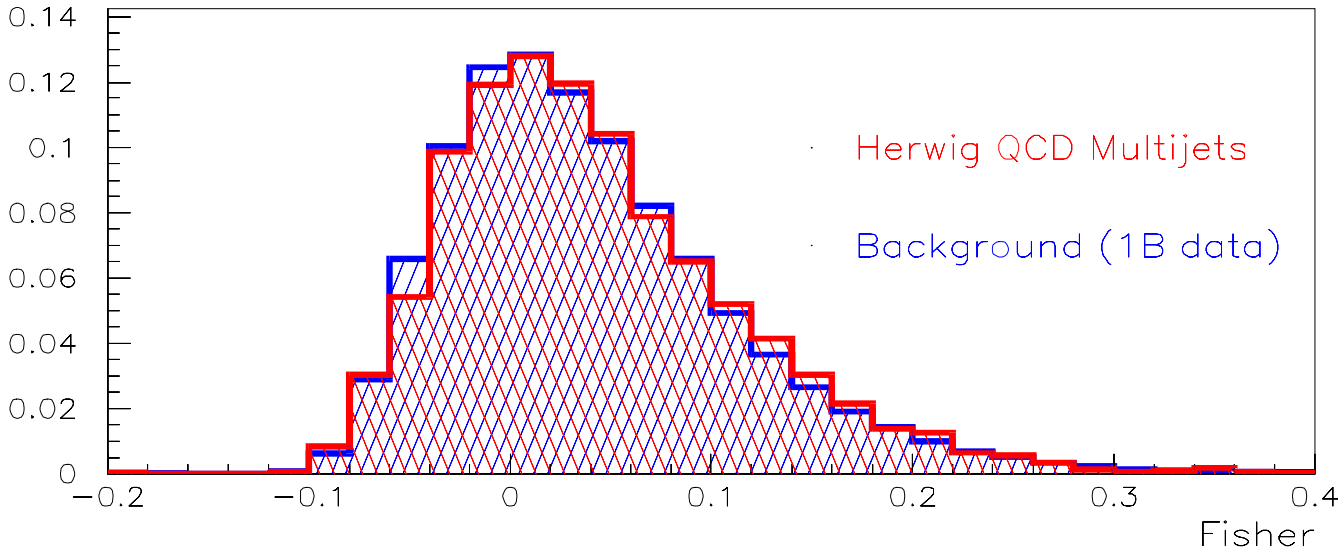
$$F \propto \frac{(\sigma_{jet} - \sigma_{quark})^2}{\sigma^2_{quark}} - \frac{(\sigma_{jet} - \sigma_{glue})^2}{\sigma^2_{glue}}$$

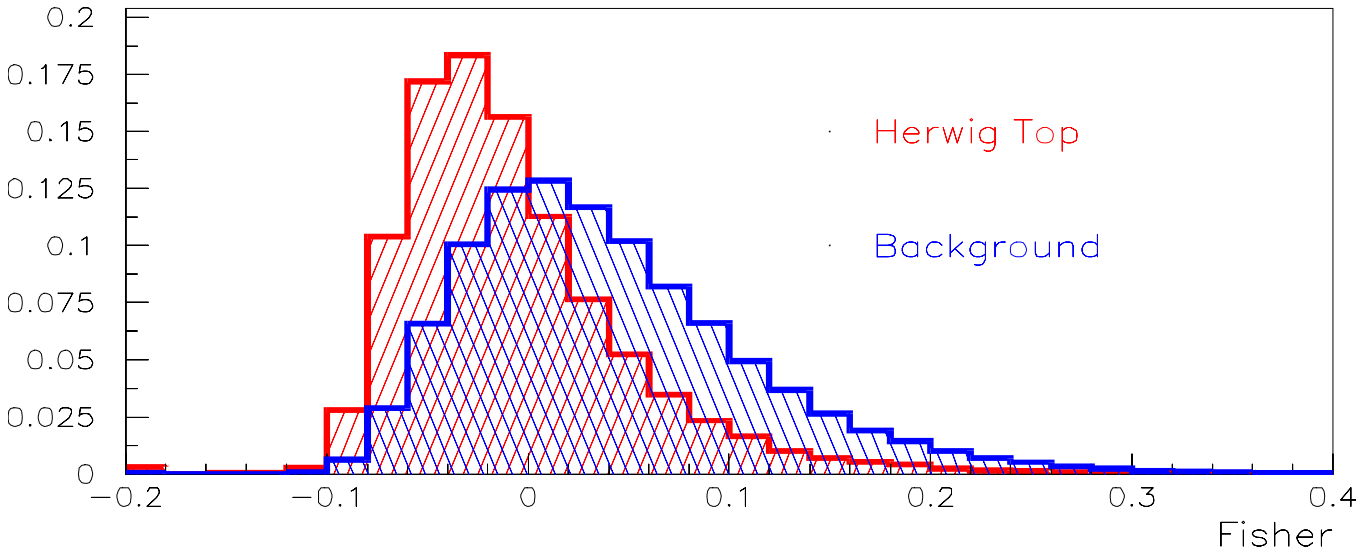


- "Quarks" ≈ 20% narrower than "gluons."

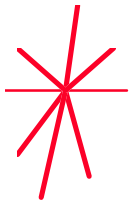
Fisher Discriminant

- A measure of how quark-rich (*signal-like*) or gluon-rich (*background-like*) the event is.
- Sum over all jets, keep 4 smallest (*most quark-like*) values.





- Herwig QCD multijets & Data in good agreement.
- Top considerably narrower.

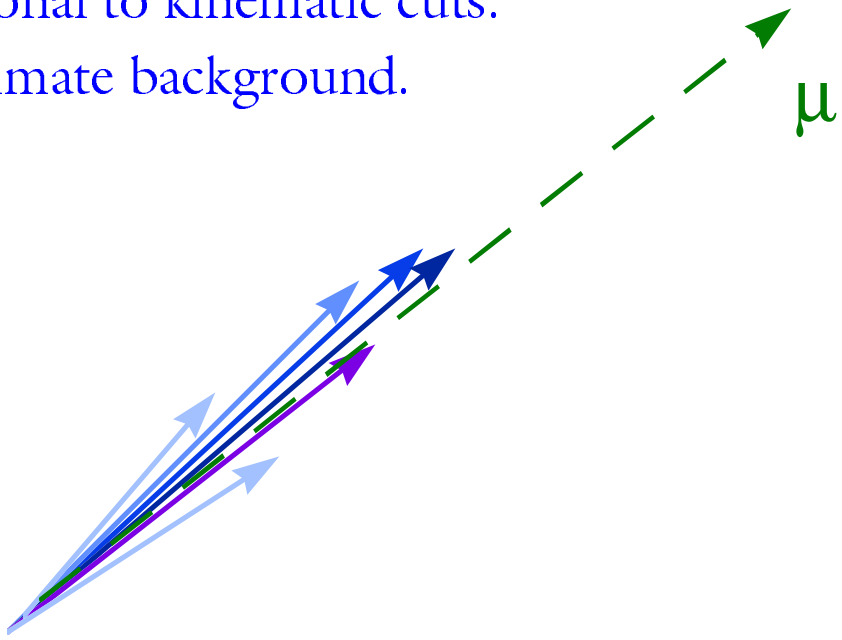


Summary of Jet Widths

- Herwig QCD multijets Monte Carlo describes the widths of jets observed in the DØ data.
- Jets in top Monte Carlo are, on average, narrower than the background.
- A Fisher discriminant, based on the differences in the widths of jets, is useful in differentiating top from background.

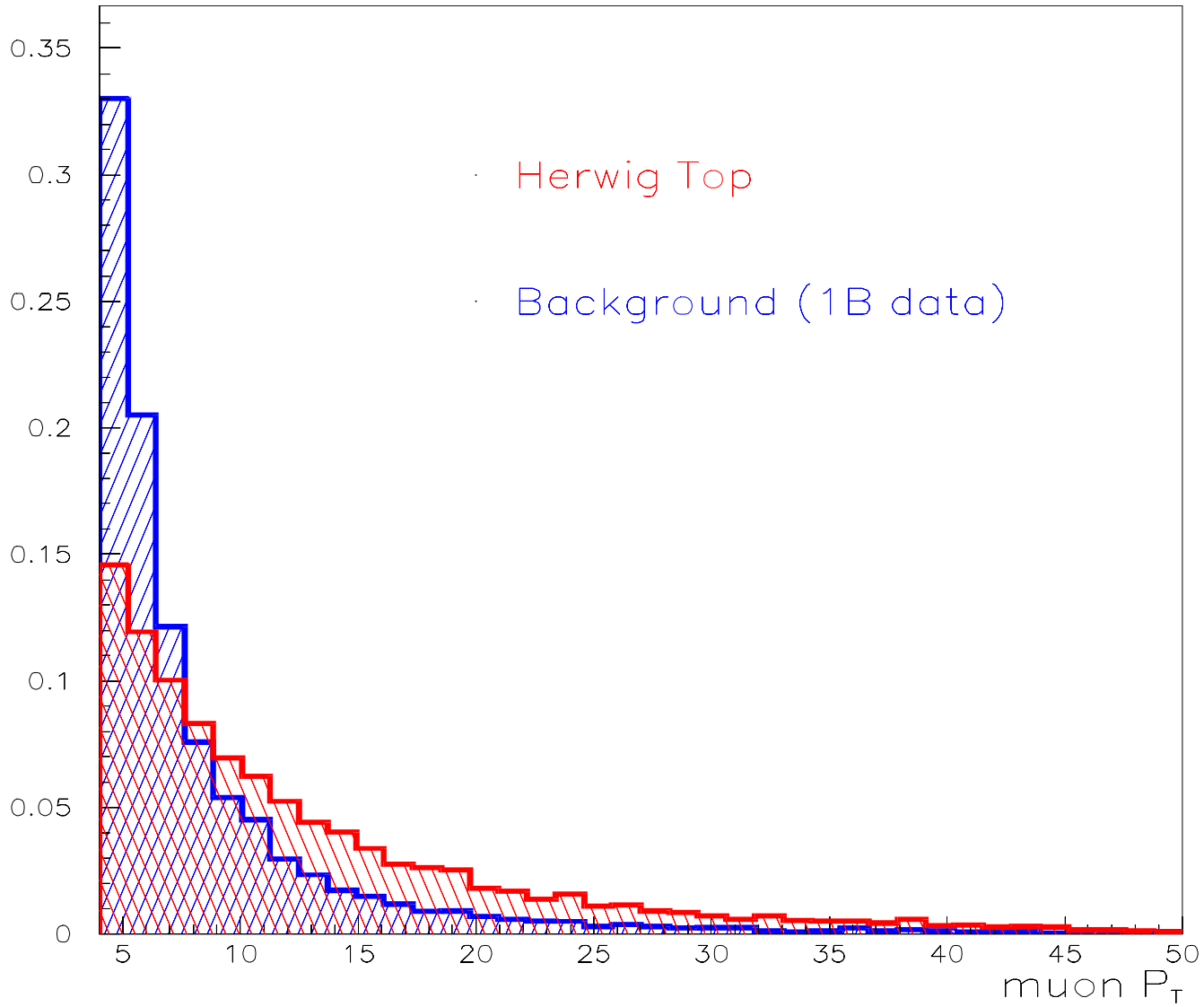
b-quark tagging with "soft" muons in jets

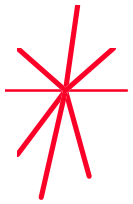
- Soft μ tags indicate possible presence of *b*-quark.
- Used in the DØ $t\bar{t} \rightarrow$ lepton + jets analyses:
  - same μ-tag selection criteria
- Increase background rejection :
  - top tagged ~ 20 %
  - background tagged ~ 3 %
- Orthogonal to kinematic cuts.
- Can estimate background.



Muon P_T

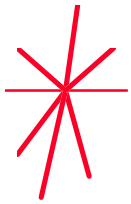
- P_T of tagged μ is stiffer in top than background.
- Use μ P_T as input variable to **Neural Network.**
- Must model μ P_T for the background calculation.





Philosophy for Background

- Untagged: QCD multijets.
- Tagged: QCD multijets with one jet μ-tagged.
- Basic process is the same, therefore can use as background as long as:
  - know probability for μ-tagging ⇒ "tag rate function"
- Tag Rate Function properties:
  - scales with # jets
  - depends on jet E_T
  - detector efficiency:
    - detector η
    - run number (radiation damage)
- Background sample is very clean:
  - Top fraction in untagged events < 4 %
  - Corrections for the 4 % are applied



Background in Alljets

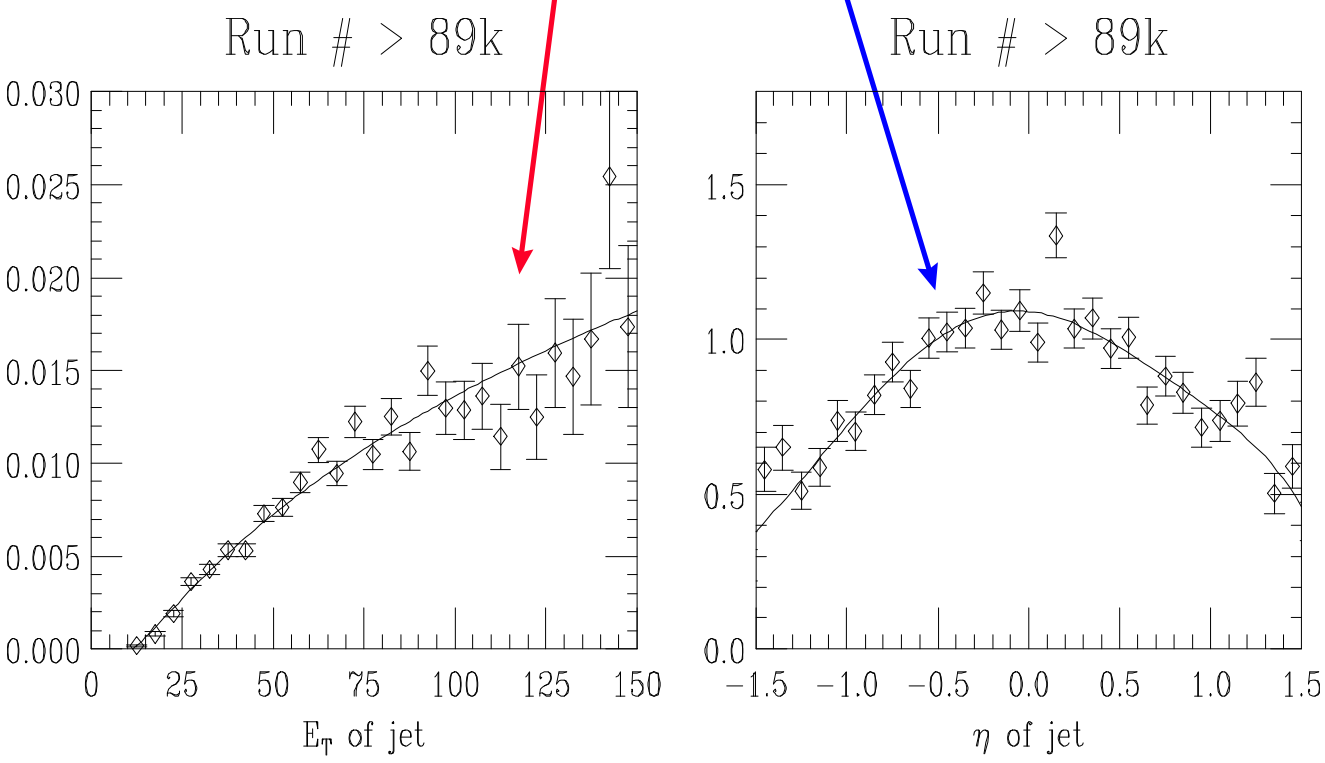
- Model Signal with **tagged** Herwig Top.
- Model Background using the **untagged** data × tag rate function.
- Because background is modeled with data, don't have to worry about:
  - The Physics (QCD `LO/NLO/NNLO`)
  - The Model (Herwig / Isajet / Pythia)
  - Detector Simulation (Geant / Noise / Resolution)
  - Energy scale

    …

μ Tag Rate

- Probability that a μ is observed in a jet of given E_T and given η (tag rate function):

$$P_\mu(E_T, \eta) = g_{jet}(E_T) \bullet h_{det}(\eta, \text{run\#})$$



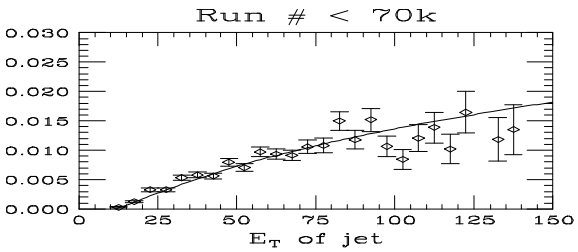
μ Tag Rate

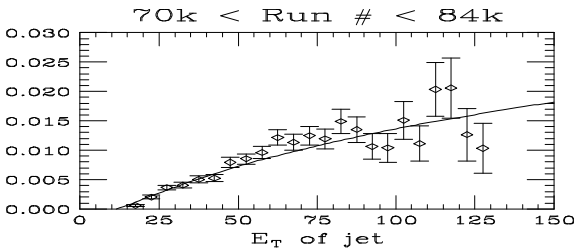
- Tag rate function:

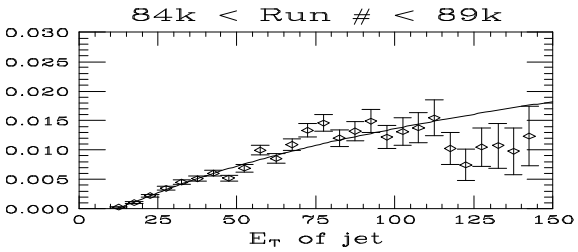
$$P_\mu(E_T, \eta) = g_{jet}(E_T) \bullet h_{det}(\eta, \text{run\#})$$

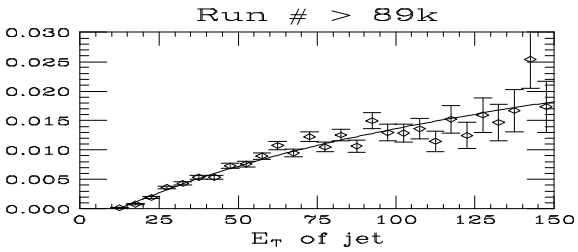
- Two possible assumptions on shape of $g_{jet}(E_T)$:

Without Saturation (used in this analysis):

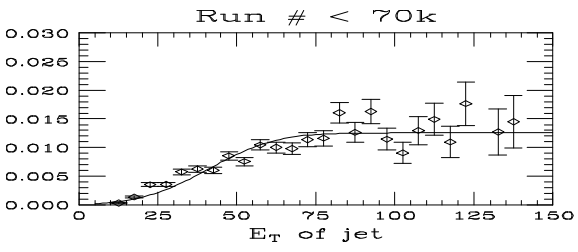


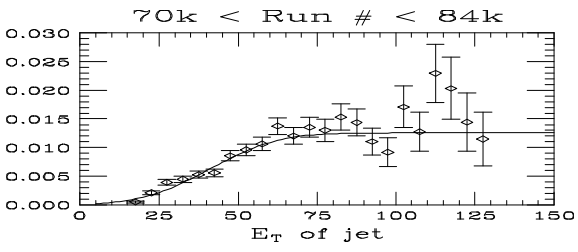


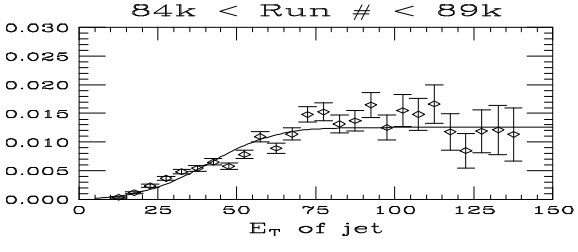


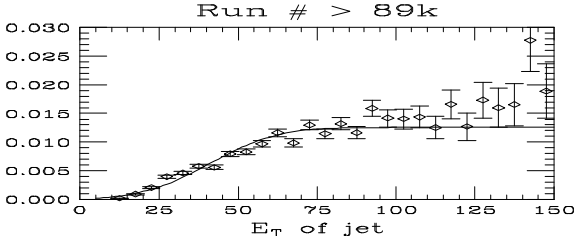


With Saturation (used for systematics study):





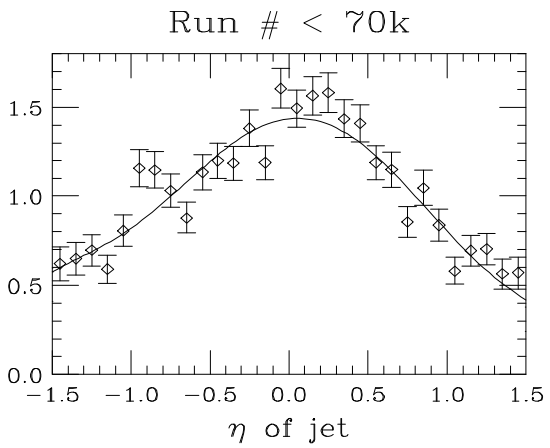


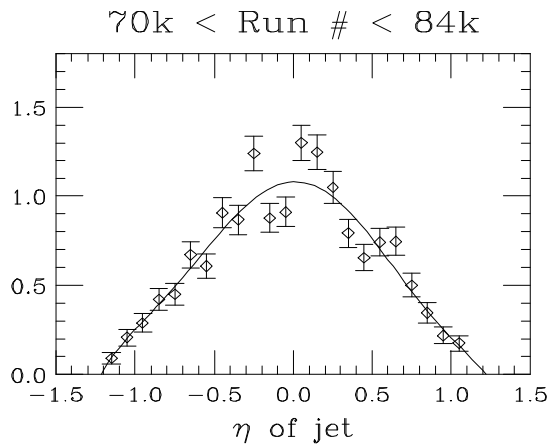


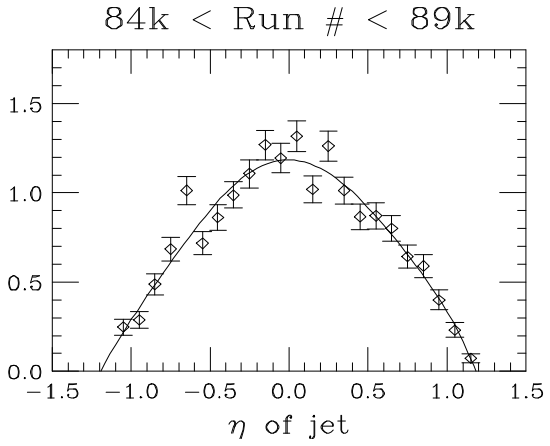
μ Tag Rate (cont.)

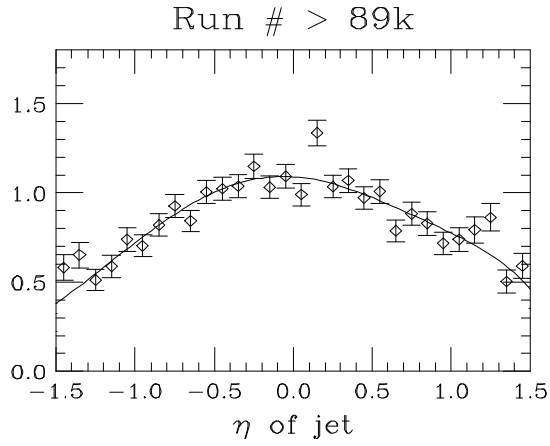
- Probability that μ is observed in a jet of E_T and η:

  $P_\mu(E_T, \eta) = g_{jet}(E_T) \bullet h_{det}(\eta, \text{run\#})$
- Using $g_{jet}(E_T)$, measure $h_{det}(\eta)$.
- Using $h_{det}(\eta)$, measure $g_{jet}(E_T)$ at η=0.
- Iterate until stable.
- $h_{det}(\eta)$ is run dependent:



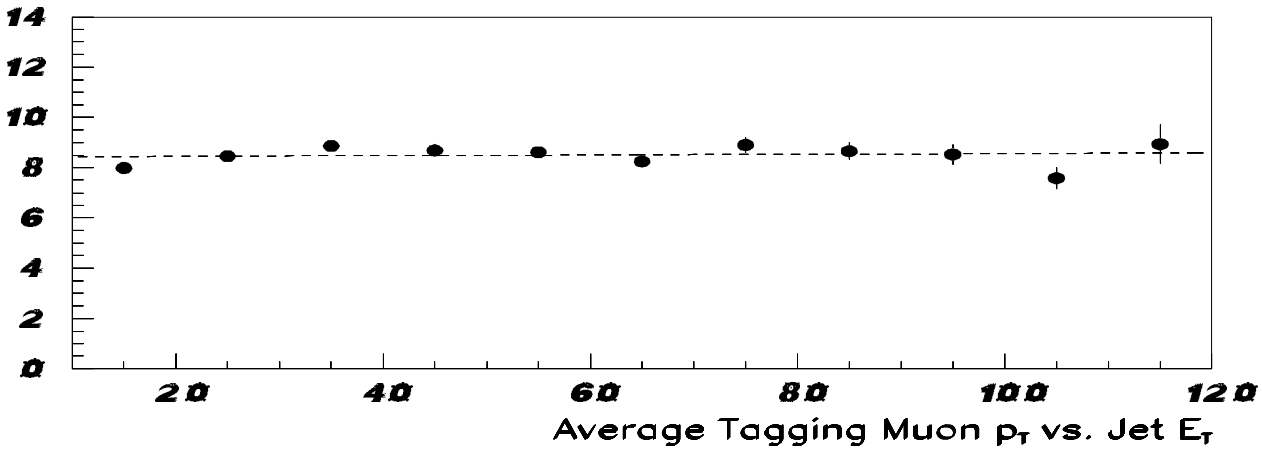


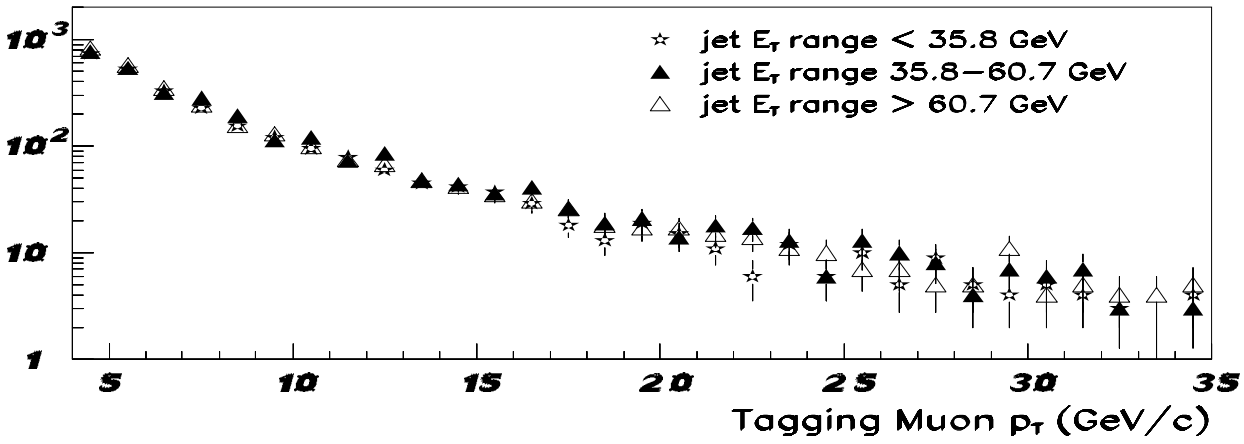


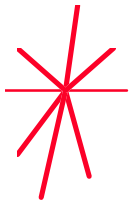


Dependence of μ Tag Rate on P_T of muon

- We observe that:
  - P_T of μ is **independent** of the observed E_T of the tagged jet.
  - Hence, P_T of μ is independent of kinematics of event and factorizes.
  - Probability to find μ in jet **does** depend on the observed E_T of the jet {*i.e.*, tag rate function}.
  - $P_\mu(P_T, E_T, \eta) = f_\mu(P_T) \bullet g_{jet}(E_T) \bullet h_{det}(\eta)$





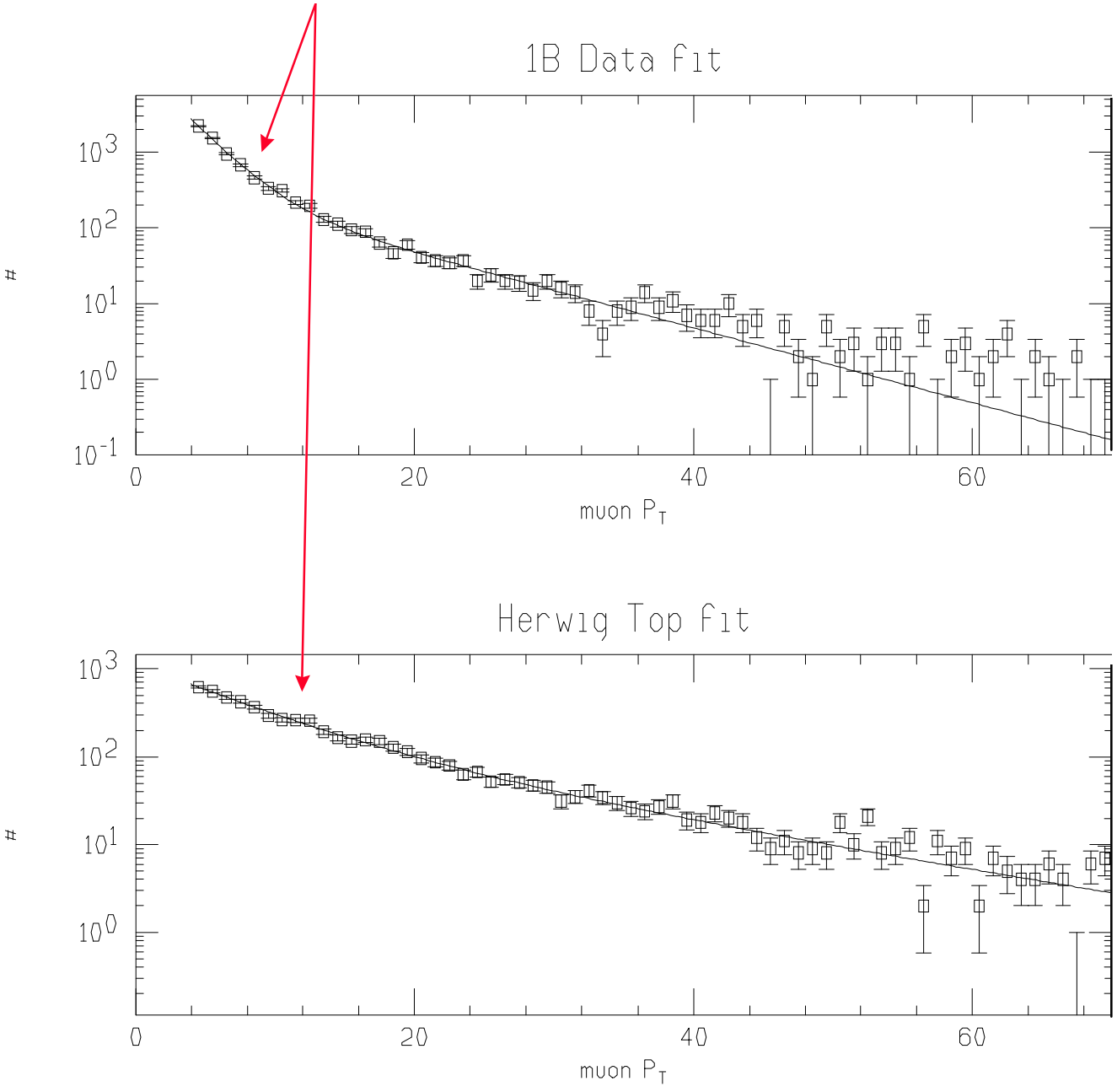


P_T of μ is <u>independent</u> of the observed E_T of the jet

- P_T of the μ depends on the fragmentation of the meson. For *b*-quarks, this is much harder.
- The E_T of the jet is <u>uncorrected</u> for the momentum of the μ or the ν.
- Gradual decorrelation in momentum between: measured jet $\Rightarrow$ *b* $\Rightarrow$ meson $\Rightarrow$ μ $\Rightarrow$ measured μ

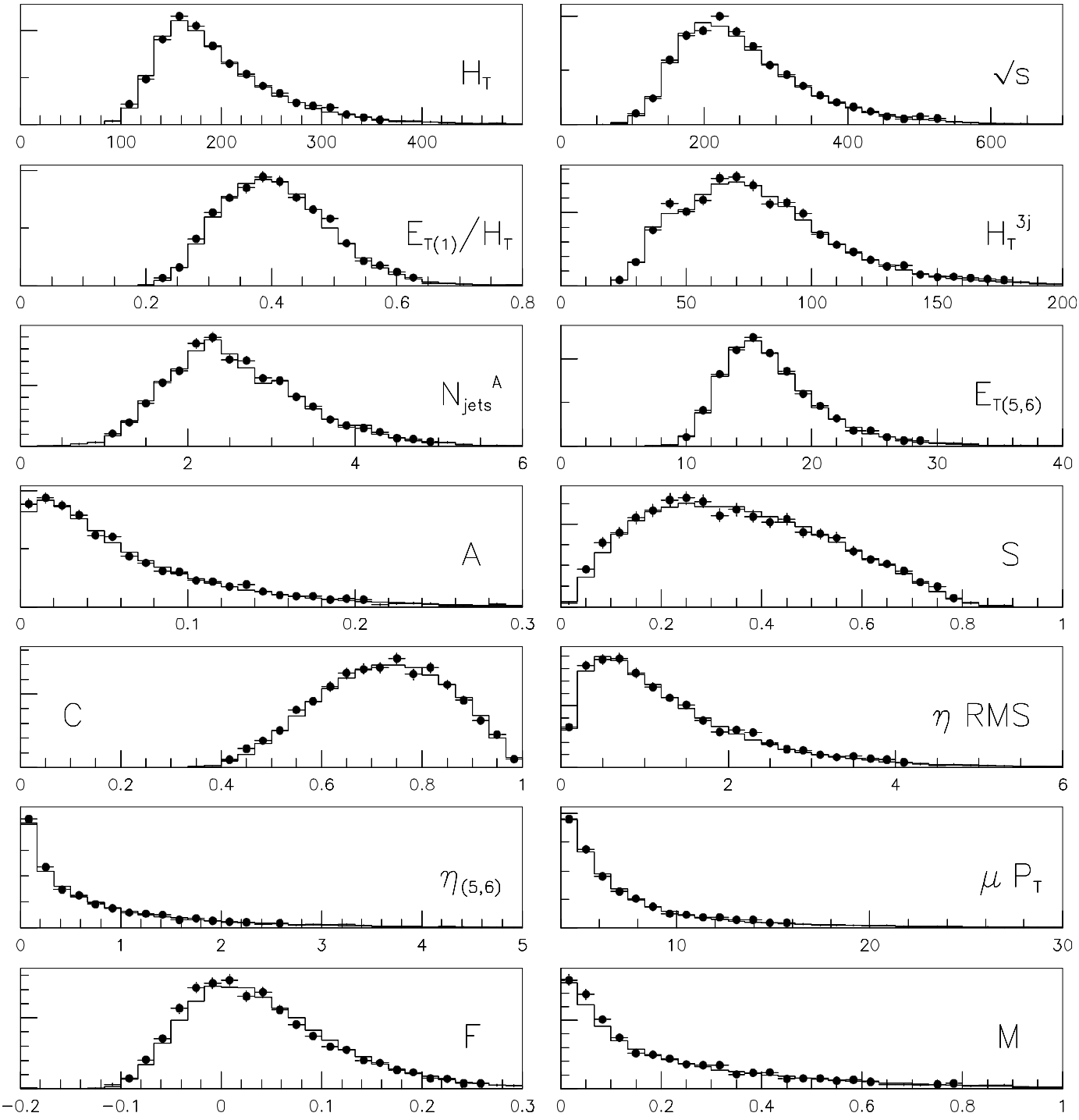
Background Method

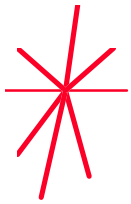
- Background: untagged data events + "fake" μ.
- "fake" μ's generated according to:
  - $P_\mu(P_T, E_T, \eta) = f_\mu(P_T) \bullet g_{jet}(E_T) \bullet h_{det}(\eta)$
  - $f_\mu(P_T)$ is the fitted P_T spectrum of the μ



Background Method (continued)

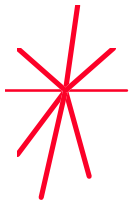
- Observed (•) and modeled (*hist.*) distributions for tagged events agree well in all 14 parameters:





Background Summary

- Tagged background is well modeled using the untagged data and the tag rate function.
- Have a method to include the P_T of the μ in the background model.



Estimate of Excess

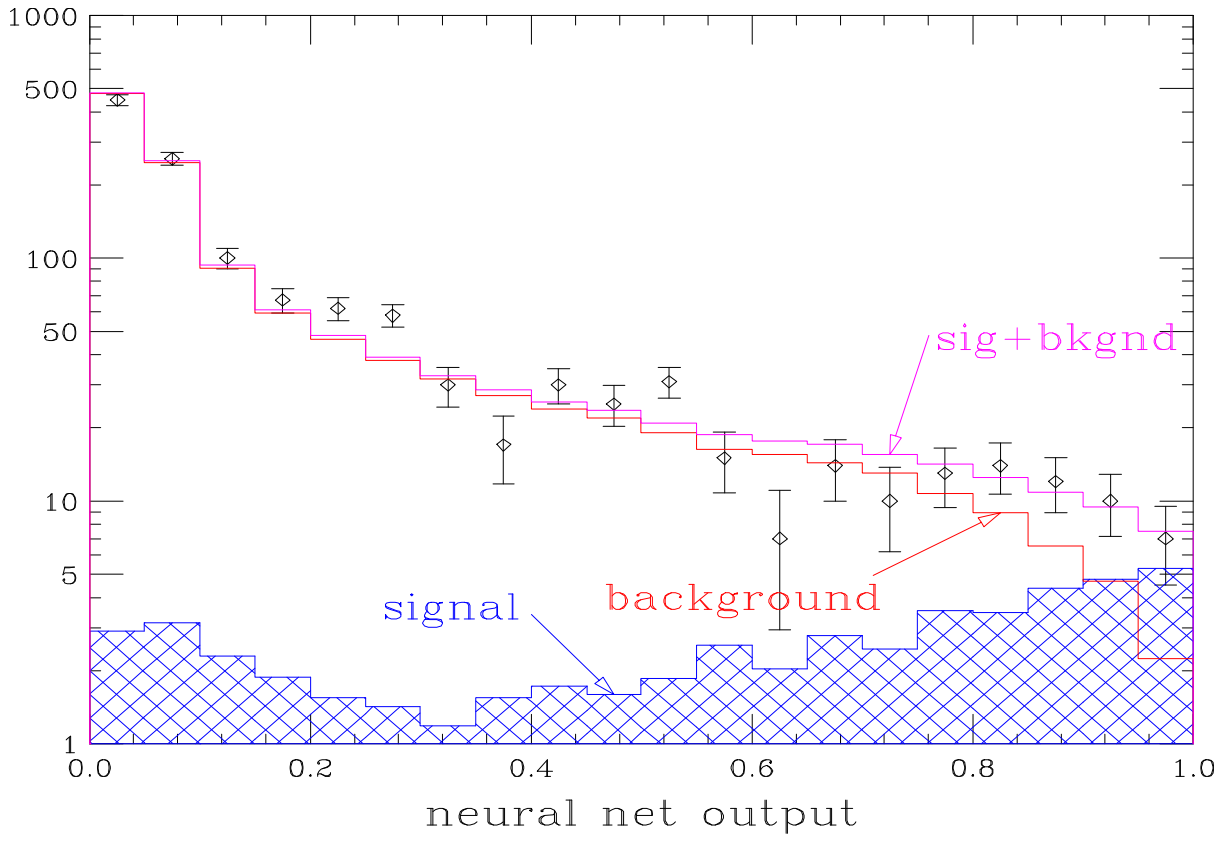
- Event excess is the difference between the observed tagged events and the background model.
- Event excess **depends** on the background estimate being correct.
- Background estimate is **independent** of:
  - Neural Network
  - Energy scale
  - Top Model

    …
- **Do** worry that:
  - tag rate function has the correct form and fits well

Combined Fit to Alljets

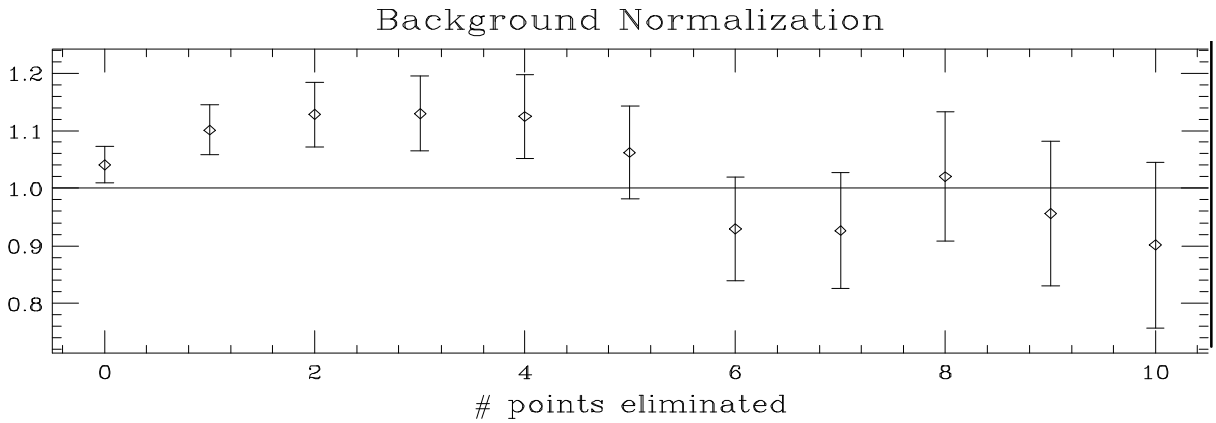
- Fit the normalized tagged data (*points*) to the expected shapes for signal & background:
  - Signal: tagged Herwig Top
  - Background: untagged data+"fake" μ
- Simultaneously fit: background normalization & top cross section.
- Preliminary cross section:

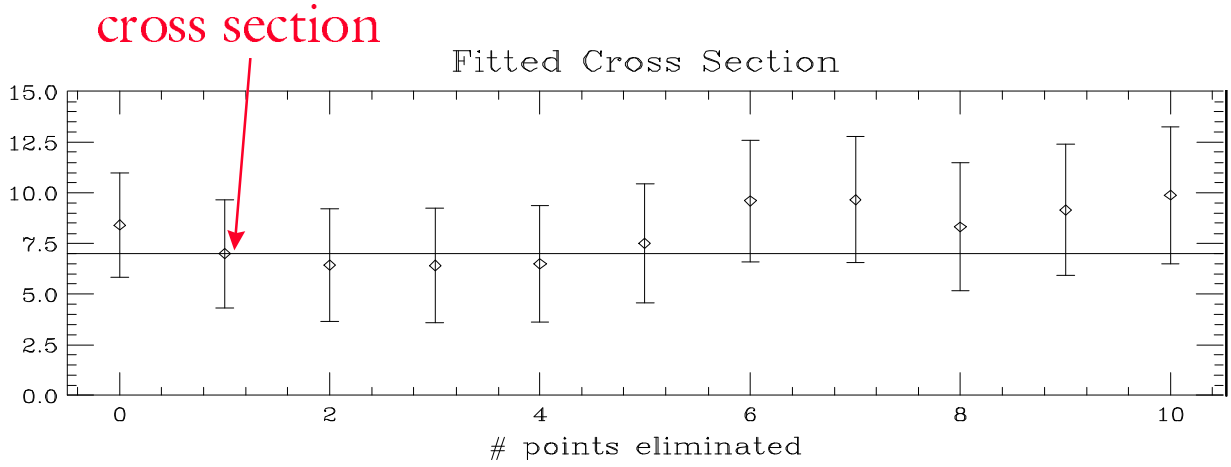
$\sigma_{t\bar{t}} = 7.9 \pm 3.1$ (stat.) ± 1.7 (syst.) pb
@ $m_t = 172$ GeV/c^2

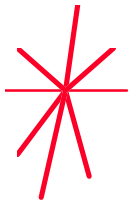


Combined Fit to Alljets (cont.)

- Check fit by sequentially eliminating bins at lowest neural network output.
- Refit for Signal & Background:
  - Statistical error on cross section includes normalization error on the background
  - No variation of cross section outside errors
  - Quote cross section eliminating lowest bin







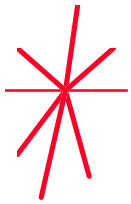
Uncertainties for Cross Section

- For background:

4.0 %	statistical
5.4 %	correction for top
6.8 %	tag parameterization
11.0 %	background model

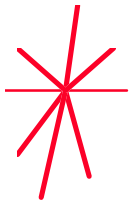
- For acceptance:

2.0 %	Fisher discriminant
3.4 %	Monte Carlo statistics
5.0 %	trigger acceptance
5.4 %	luminosity
5.7 %	(4%+1GeV) MC energy-scale
6.0 %	$b \rightarrow \mu$ branching fractions
6.2 %	Model: Herwig-Isajet
7.0 %	muon chamber efficiency
7.0 %	P_T spectrum of μ



Significant Excess?

- Preliminary Results:
  - Cut at Neural Network > 0.78
  - Events = 44
  - Background = 25.3 ± 3.1
  - Excess = 18.7 ± 7.3
  - Expected = 11.6 ± 4.5
    (Using Herwig Top, m_t=*172* & σ_{tt}=*5.6 pb*)
- Significance of excess ~ 3 sigma for Gaussian equivalent probability.



Conclusion

- Preliminary $t\bar{t}$ cross section from $\bar{t}t \rightarrow$ alljets @ $m_t = 172$ GeV/c^2

  $\sigma_{t\bar{t}} = 7.9 \pm 3.1$ (stat.) ± 1.7 (syst.) pb.
- In good agreement with DØ semi-leptonic result: $\sigma_{t\bar{t}} = 5.6 \pm 1.8$ pb
- Significance ~ 3 sigma for Gaussian equivalent probability.