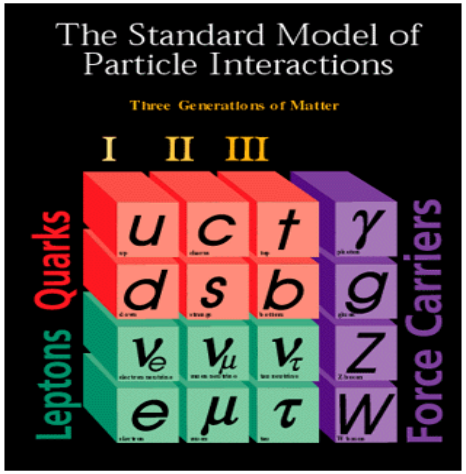
Model Independent Search for New Physics in Leptonic Final States at DØ

Joel Piper
Michigan State University

Fermilab
November 17, 2009

The Standard Model is incomplete. At the energy frontier at the Fermilab Tevatron, we must address a central question:

Do we see what we expect?



Outline

I. Introduction

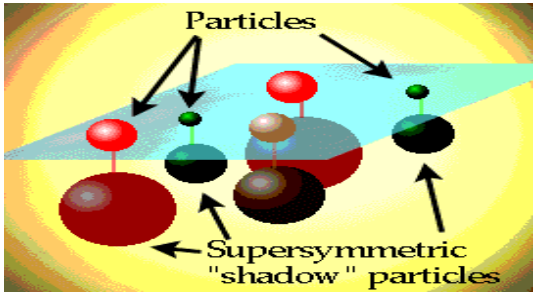
II. Strategy

III. Results

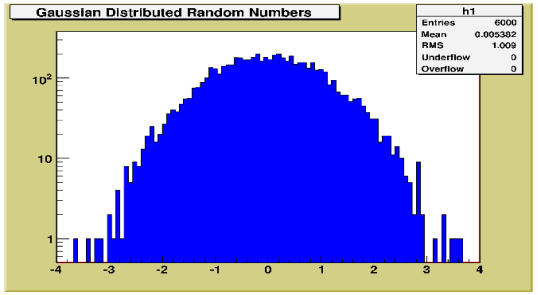
?

II. Introduction

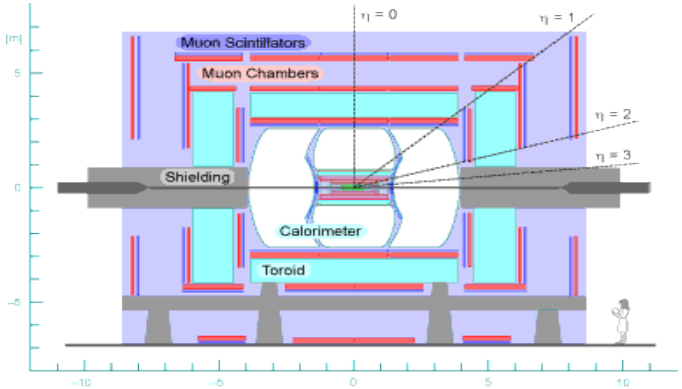
A. Motivation



B. What is a Model-Independent Search?



C. The D0 Detector



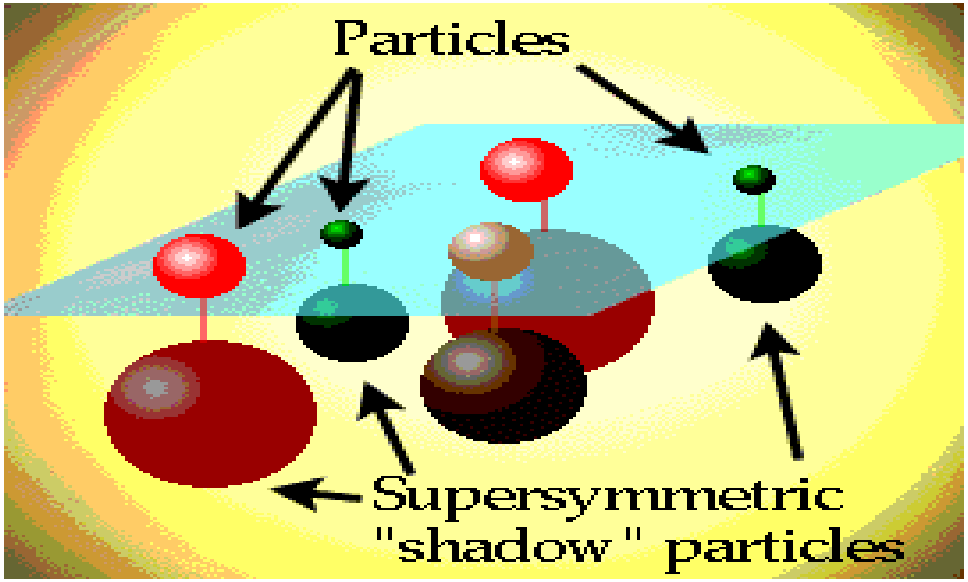
Searching for physics beyond the Standard Model in 1 fb^{-1}

Standard Model is successful yet incomplete

- Gravity not incorporated
- Higgs mass fine-tuned
- No unification of forces at high energy

We are at the edge of an energy scale where we expect to find something new... but what?

Many theories devised to explain these problems with unique signals and characteristics

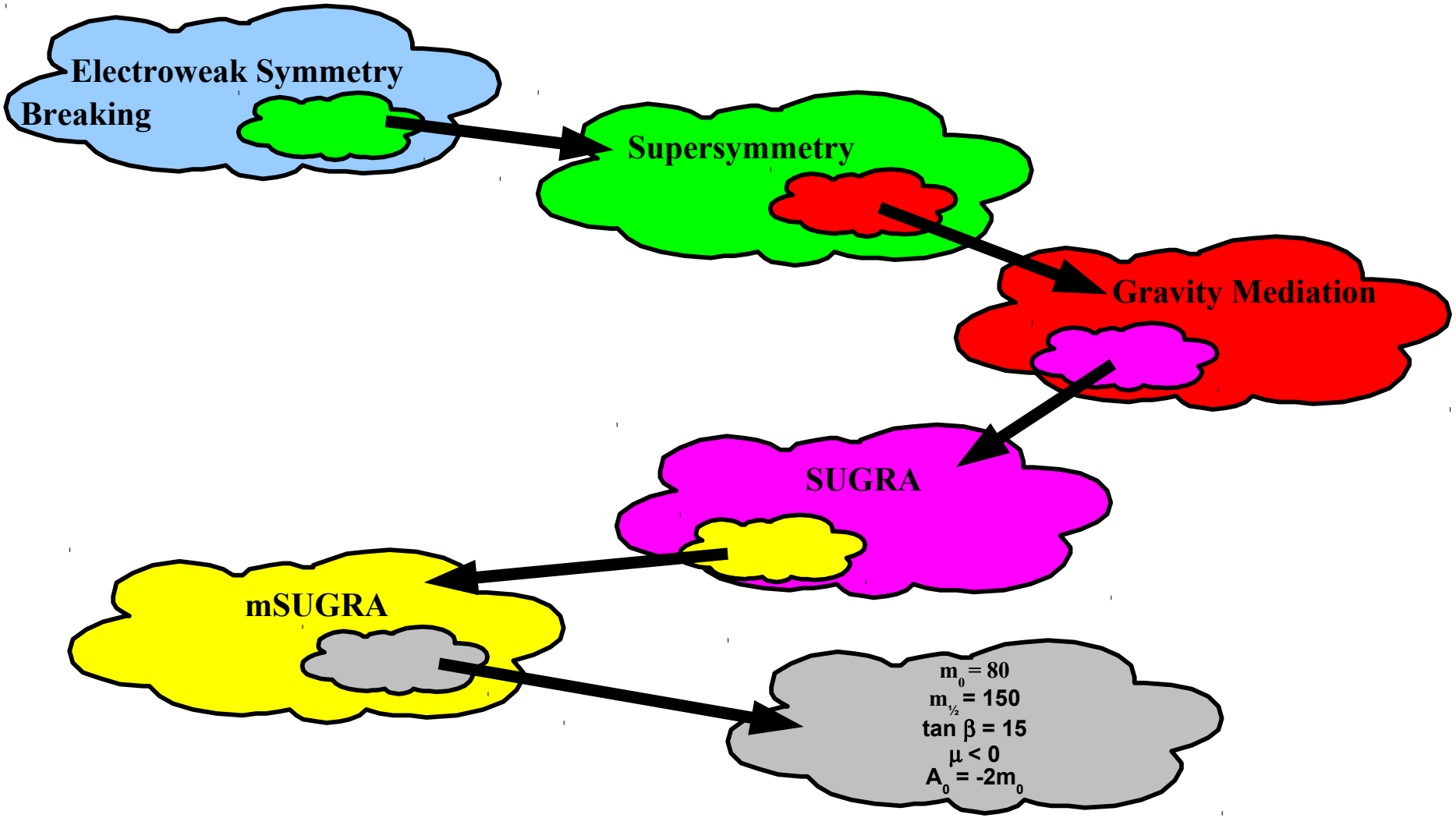


We must choose where to invest our resources

Most interesting?

Reasonable acceptance?

A broad search complements many targeted ones



What is a global search?

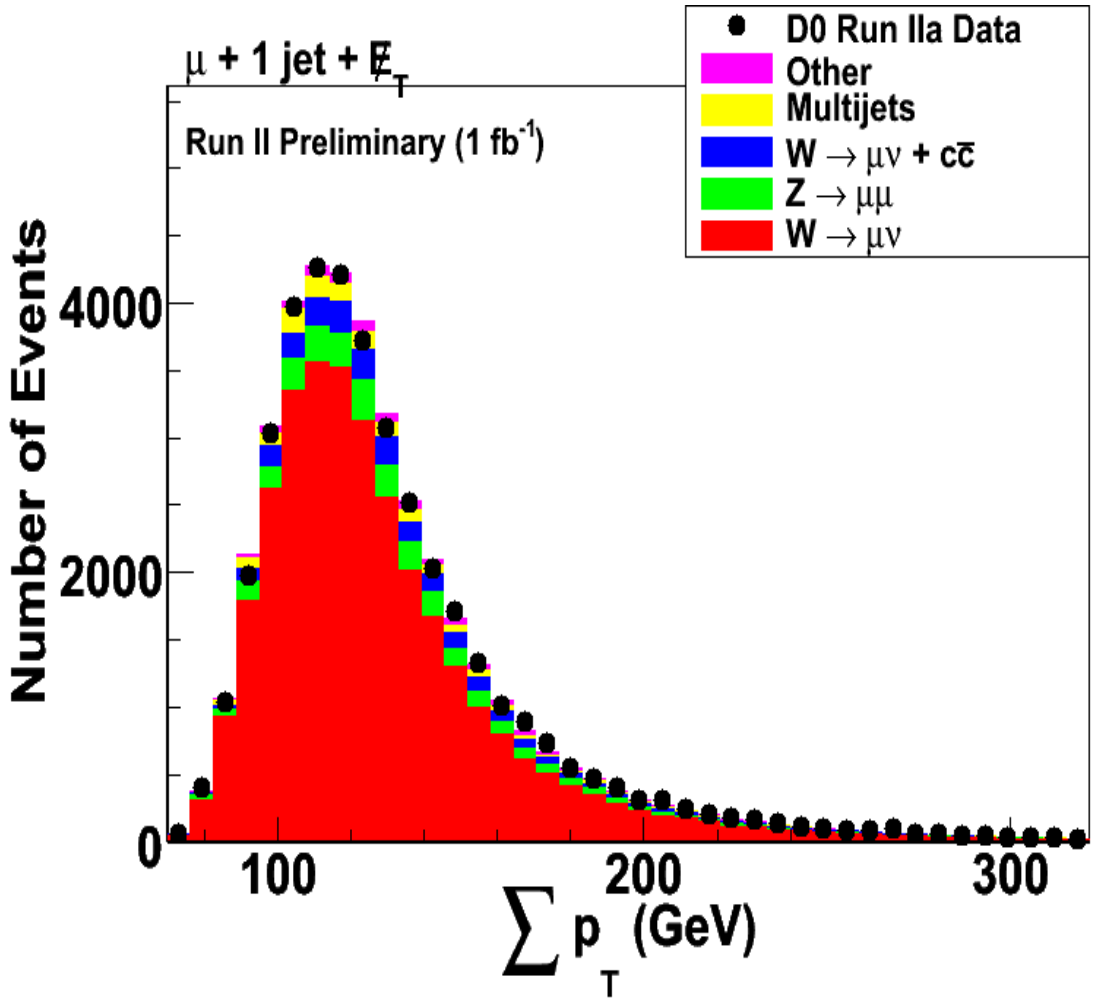
Need to a specific strategy....

What are the goals?

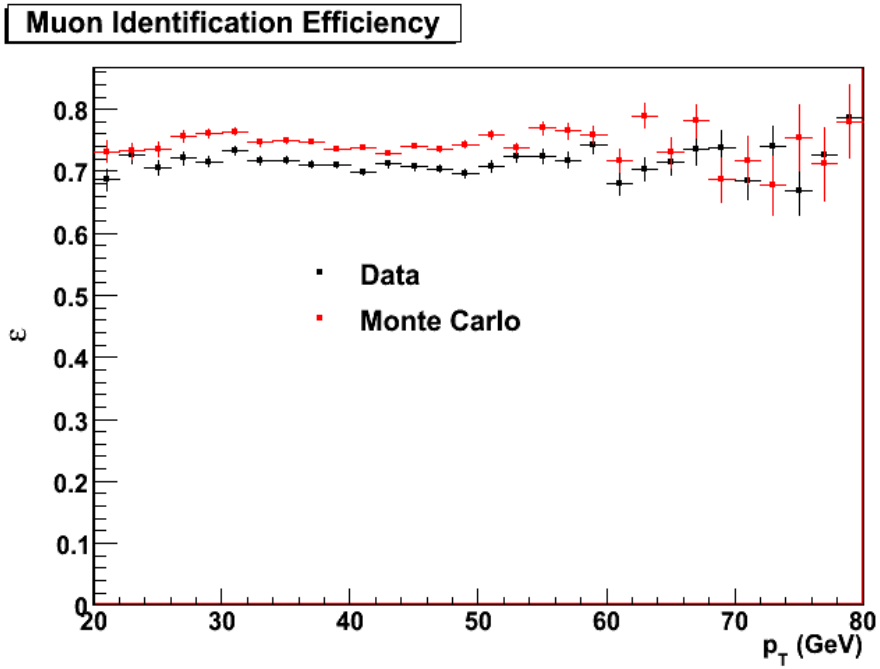
- Large breadth
- Check for EW symmetry breaking in high-p_T final states

So, where do we look?

- Event counts in final states
- Shape agreement over many distributions
- Event $\Sigma\, p_T$ distributions



A Model Independent Search is a search for a deviation from SM predictions



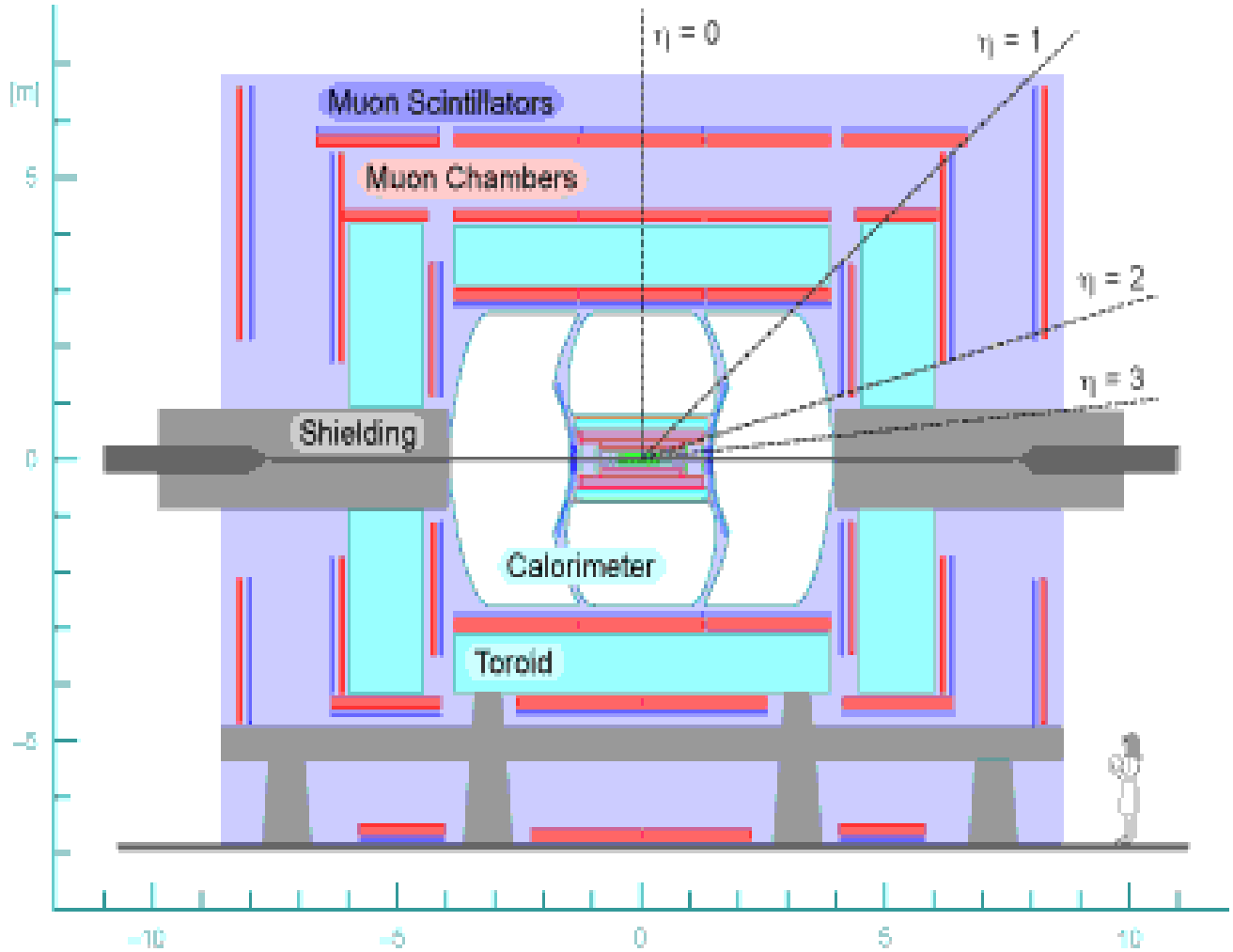
What can cause a deviation?

- Statistical fluctuation
  - Keep careful count of how many places checked
  - Trials Factor
- MC simulation accurate here?
- Detector modeling or event reconstruction?
- Could this be new physics?

To minimize MC and detector modeling issues:

- Use common MC with standard object definitions
- Apply common collaboration-wide scale factors
- Fit for normalization factors

The D0 Detector is used in study of high energy $p\bar{p}$ collisions



Three level trigger system used to choose interesting events

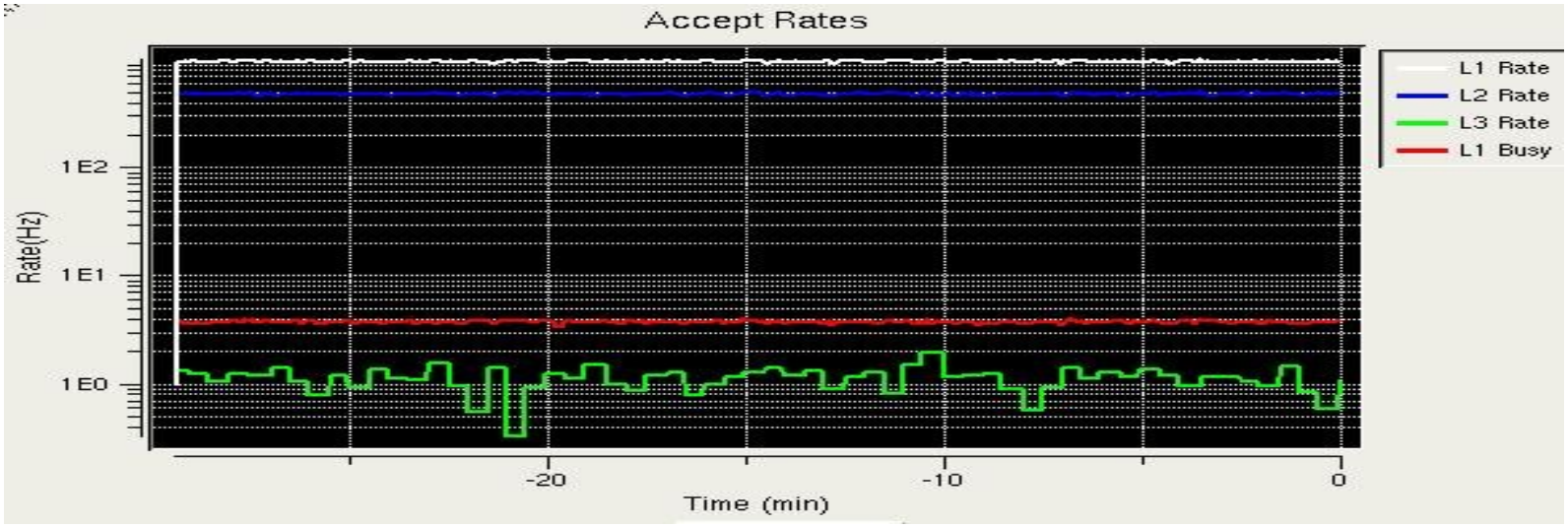
Collisions occur at max rate of 2.5 MHz

Level 1 reduces this to ~1.5 kHz

Level 2 cuts this rate in half for input to Level 3

- Uses five processors for physics object refinement
- One processor to combine event-wide information (L2 Global)

Level 2 Global makes Level 2 trigger decision



Level 2 Global first part of triggering that combines objects from all parts of detector

My work was as resident Level 2 Global expert

Check algorithm results from analysis group against implementation in online code

- Take algorithms from analysis group code used to test trigger rates
- Implement algorithm in Level 2 environment
- Ensure that properties assumed in initial testing are equivalent to online code results

Implementation projects included:

- b-quark jet events
- τ objects
- sphericity/acoplanarity

Key member of team in system critical for daily operations

Implemented new code and fixed problems with matching information from leptons and tracks at Level 2 Global

- Insidious τ track-matching issue- found cause and fixed
- Added clarity in μ track matching
- Removed ambiguities in electron tracking

Studied use of multiple processors for Level 2 objects

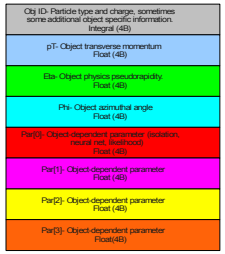
- Found to be unnecessary
- Single processing board sufficient for even very high luminosities

Crucial member of Level 2 team

- Week-long, 24-hour/day pager shifts
- On-site co-leader during several month interim between L2 coordinators
- Data acquisition expert

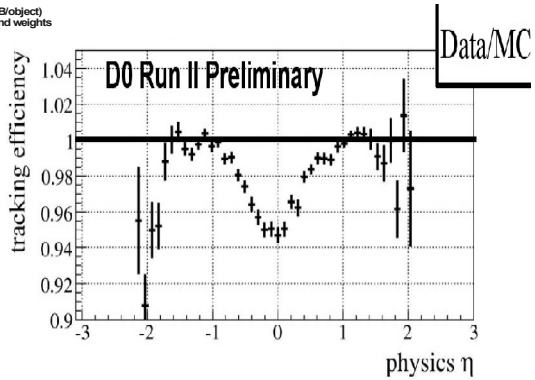
II. D0 Model Independent Search Strategy

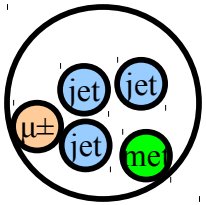
A. D0 High-p_T Dataset



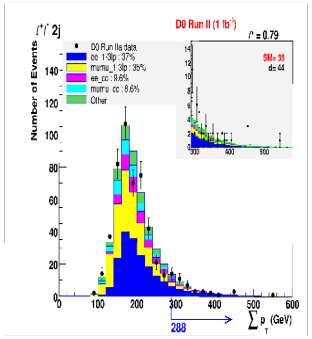
Total Object Size: 32 B
Also save:
Weight information (~12B/object)
Global event information and weights

B. D0 Model Independent Search Analysis Packages & Fits



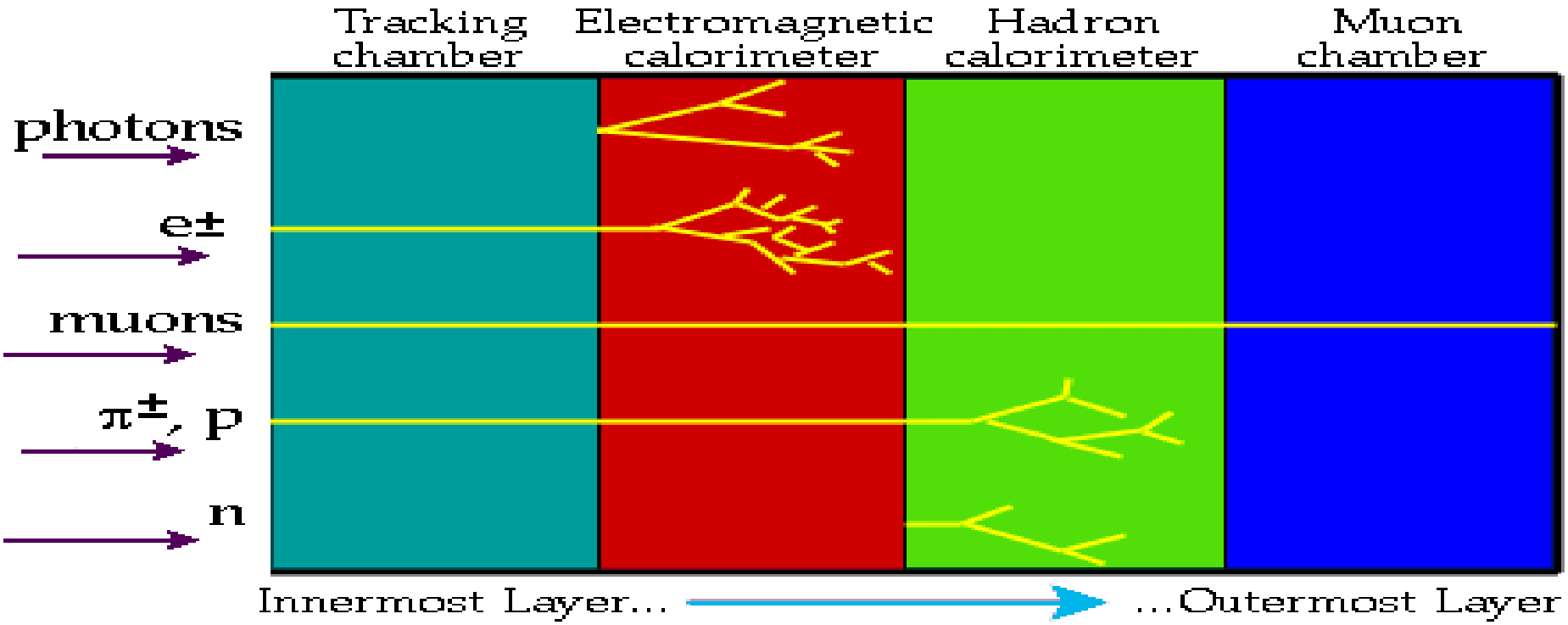


C. Vista Exclusive Model Independent Comparison



D. Sleuth Comparison at High-p_T Tails

Stored detector tracks and energies must be translated into physics objects



Object ID through detector signature

A model-independent search must ensure no double counting

Also single-track taus and b-quark jets

D0 High-P_T Dataset for Model Independent Search

- **After reconstruction, events saved in data skims**
  - MU inclusive skim- (eg. one 8 GeV μ w/loose cuts)
  - EM inclusive skim- (eg. one 20 GeV EM obj w/loose cuts)
- **Monte Carlo**
  - Using Pythia matched Alpgen or just Pythia generators
  - Use common GEANT-based detector simulation code
- **Model Independent Search considers many final states, so many processes contribute**
  - OR of many possible object contributions
  - Running over huge dataset and many inclusive Monte Carlo processes →
    - Small changes in analysis can take weeks to implement
- **Solution: Create a reduced dataset that captures all events w/high-p_T objects and save only basic object & event info**

Only most basic quantities stored for high-p_T events

- Full rerun from 1 week → < 2 hours
- Full dataset storage from 19.3 TB → 7.1 GB
- I can implement a change in analysis strategy in ~1 day
- Also, I implemented reduced trigger info for high-p_T
- I only perform high-p_T rerun roughly every 3-4 months
- Last 18 months, charged with maintaining full high-p_T code

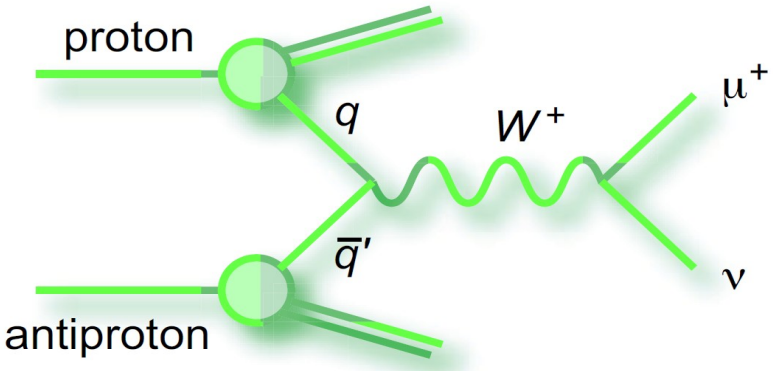
Obj ID- Particle type and charge, sometimes some additional object specific information. Integral (4B)
pT- Object transverse momentum Float (4B)
Eta- Object physics pseudorapidity. Float (4B)
Phi- Object azimuthal angle Float (4B)
Par[0]- Object-dependent parameter (isolation, neural net, likelihood) Float (4B)
Par[1]- Object-dependent parameter Float (4B)
Par[2]- Object-dependent parameter Float (4B)
Par[3]- Object-dependent parameter Float(4B)

Total Object Size: 32 B
Also save:
Weight information (~12B/object)
Global event information and weights

Standard Model Background Processes

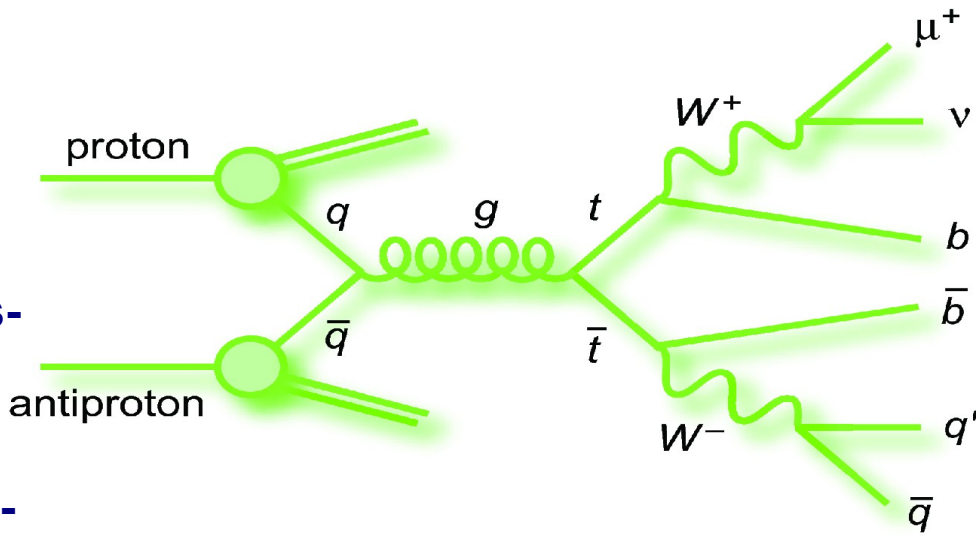
Monte Carlo with real zero-bias overlay from many processes

- Alpgen/Pythia
  - W boson decay
  - Drell-Yan processes
  - ttbar
- Pythia
  - diboson



Tevatron dominated by all-jet processes (multijets)

- Modeling relatively rare jet mis-ID processes is difficult
- Use data with loosened cuts
- For each type of jet-lepton mis-ID needs different sample

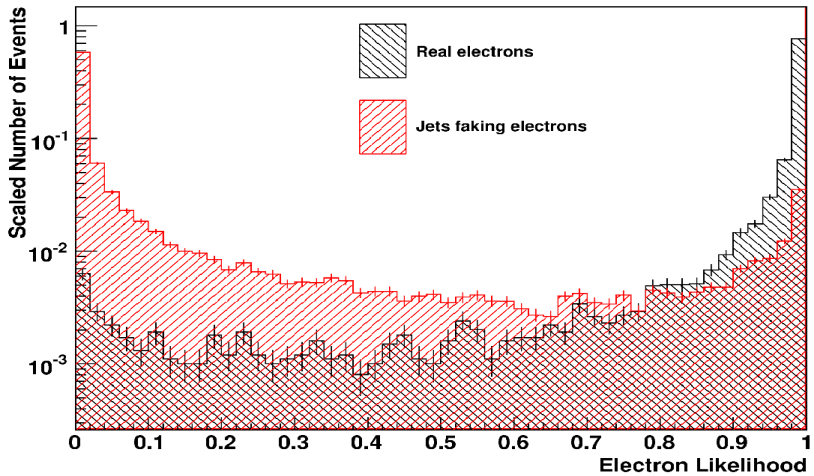


Electron multijet background uses loosened likelihood cuts

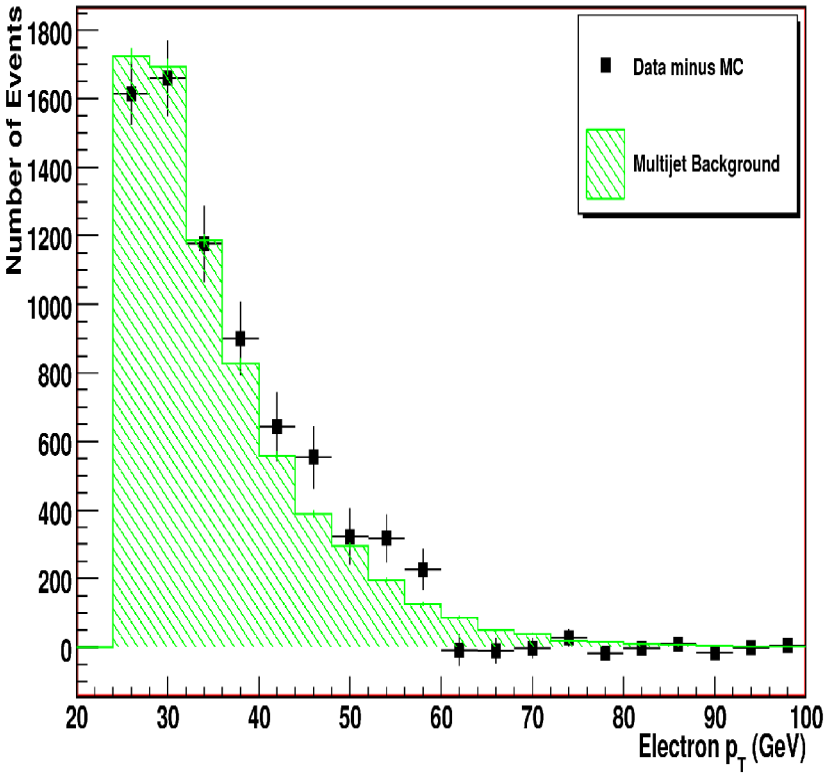
Jet faking mechanism

- Large energy fraction in EM calorimeter
- Single dominant track

Electrons use likelihood variable to separate from jets



Work in progress



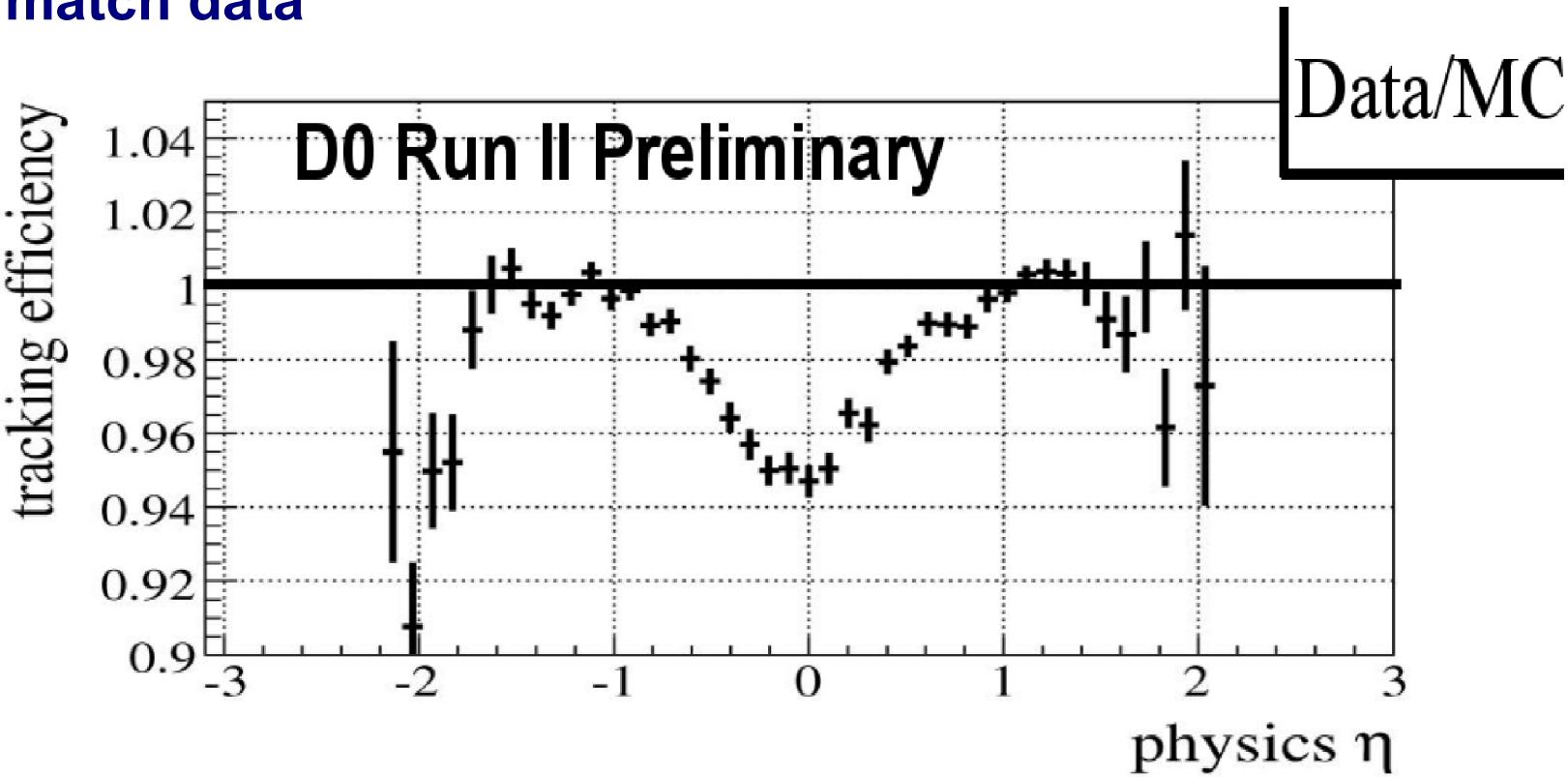
Multijets use

0.25 < likelihood < 0.8

With complete background, need to account for differences with expectation

Incorporate known D0 data/MC correction factors (weights)

Make sure these are used to correct a known issue and not just to match data



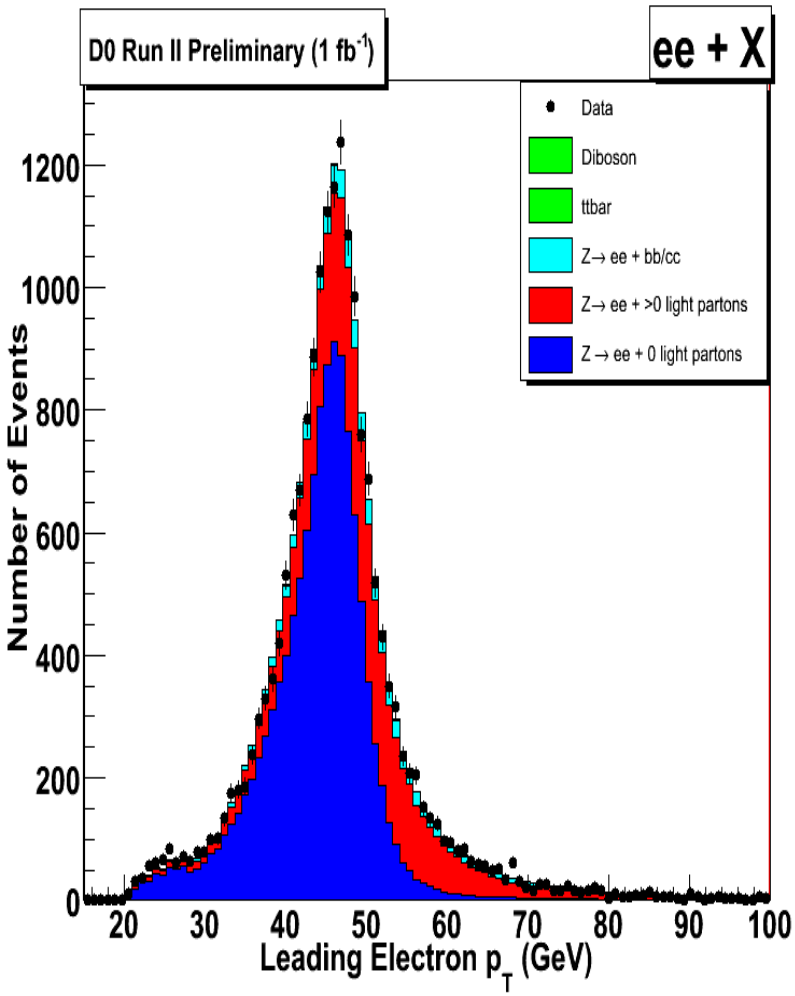
Use high-p_T dataset to feed into 7 MIS analysis final state fits

- MIS analysis final states inclusive in jets (unlike Vista final states)
- After incorporation of correction factors, one more normalization
  - Incorporates K-factors, multijet pollution, etc.
  - One factor for each process
- Rare final states excluded to test strategy

$$(WW, WZ, ZZ, t\bar{t})$$

- Subset of histograms (excluding p_T tails) picked for overall fit

$$(p_T, \eta, cos(\phi_{obj}, \phi_{MET}))$$



Total of seven inclusive final states used in normalization fits

MIS Final State	Object	Min p_T (GeV)	Max η
e + jets + X	e	35	1.1
	jet	20	2.5
	MET	20	NA
μ + jets + X	μ	25	1.5
	jet	20	2.5
	MET	20	NA
ee + X	e	15	1.1
	jet	20	2.5
$\mu\mu$ + X	μ	15	1.5
	jet	20	2.5
$\mu\tau$ + X	μ	15	1.5
	τ	15	2.5
	jet	20	2.5
eτ + X	e	15	2.5
	τ	15	2.5
	jet	20	2.5
μe + X	μ	15	1.5
	e	15	2.5
	jet	20	2.5

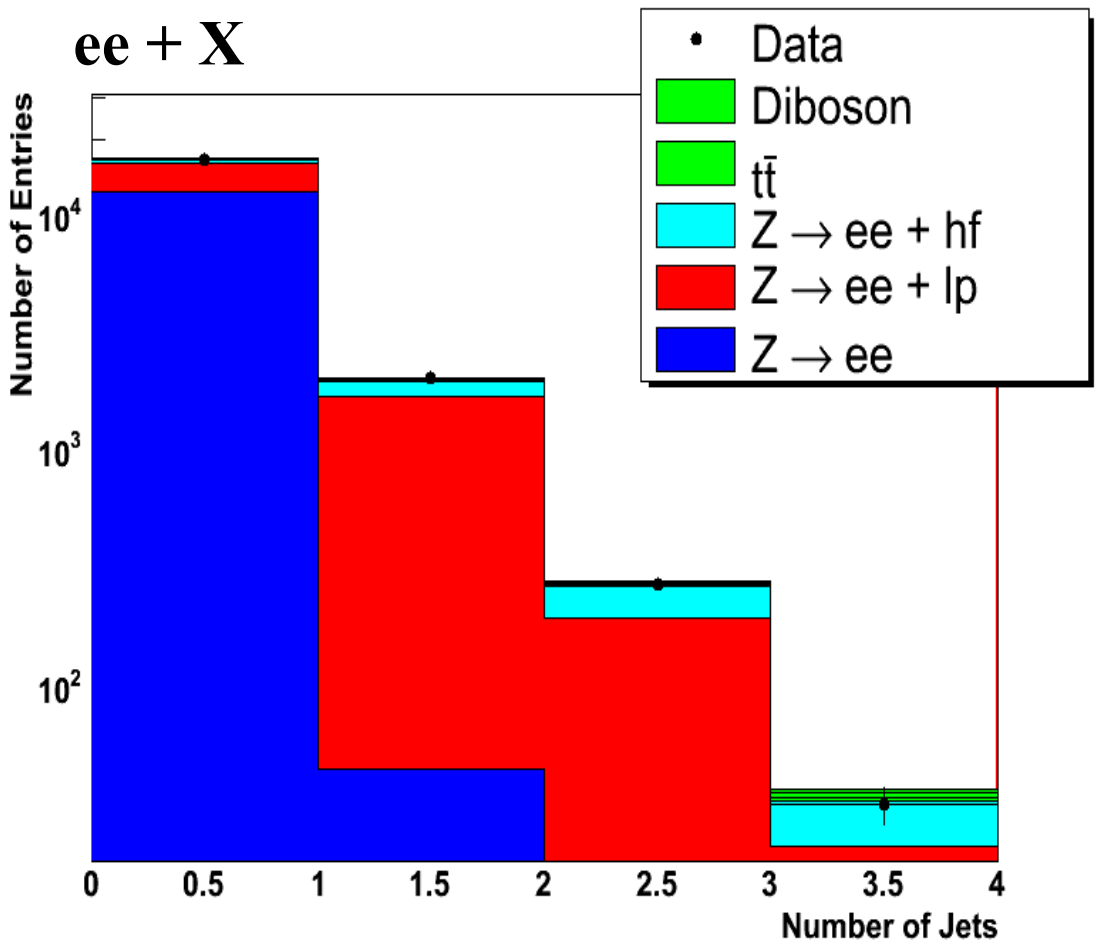
D0 Run II Preliminary

Additional objects X

Object	Min p_T (GeV)	Max η
e	15	2.5
μ	15	1.5
τ	15	2.5
γ	15	1.1

D0 Run II Preliminary

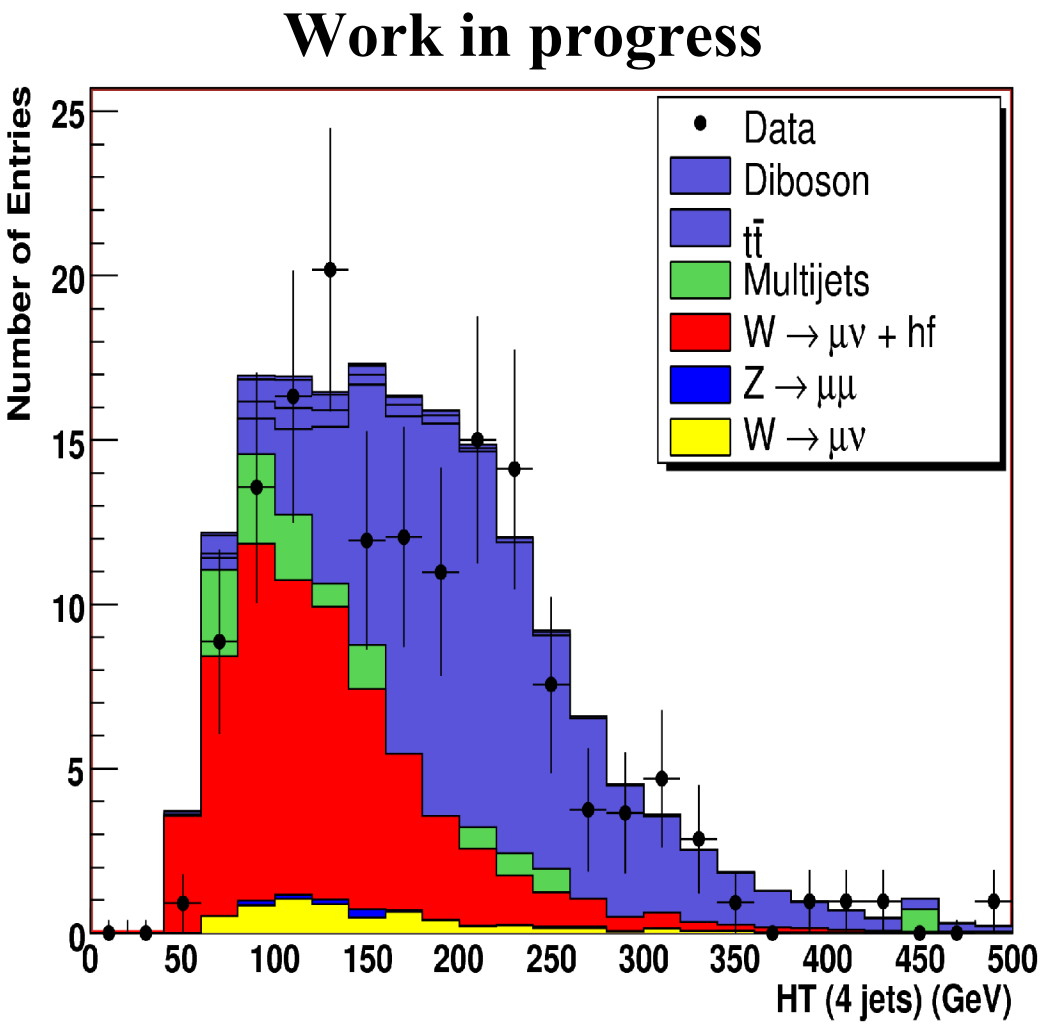
Perform checks and pass on to experiment-independent programs



- Pick another set of histograms to check
- Looking for qualitative agreement
- KS test for each plot for quantitative check
- Search for significant discrepancies
  - Use Vista for more subtle issues
  - Are basic processes well-modeled?

Qualitatively check incorporation of rare states

- μ + jets with at least four jets
- Σ of all jet p_T's
- The top quark Monte Carlo is shown in the dark blue
- Necessary for reasonable agreement



Model independent searches at CDF use a different approach

They cover as many final states as possible

Live with discrepancies if poor modeling

Use Multijets (QCD) Monte Carlo

All data manipulation is done at Vista level

CDF Run II Preliminary (2.0 fb^{-1})

Code	Category	Explanation	Value	Error	Error(%)
0001	luminosity	CDF integrated luminosity	1990	50	2.6
0002	k-factor	cosmic_ph	0.83	0.05	6.0
0003	k-factor	cosmic_j	0.192	0.006	3.1
0004	k-factor	$1\gamma 1j$ photon+jet(s)	0.92	0.04	4.4
0005	k-factor	$1\gamma 2j$	1.26	0.05	4.0
0006	k-factor	$1\gamma 3j$	1.61	0.08	5.0
0007	k-factor	$1\gamma 4j+$	1.94	0.16	8.3
0008	k-factor	$2\gamma 0j$ diphoton(+jets)	1.6	0.08	5.0
0009	k-factor	$2\gamma 1j$	2.99	0.17	5.7
0010	k-factor	$2\gamma 2j+$	1.2	0.09	7.5

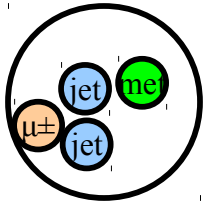
Vista checks number of events and shapes in exclusive channels

Divide data/MC for each *exclusive* final state

- Based on high- p_T objects: μ, e, τ, γ, jets, b-jets, MET

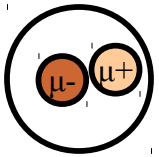
Check numbers of events

- Significance diluted by number of final states



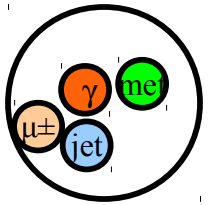
Compare shapes

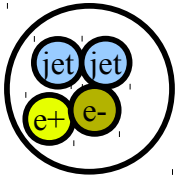
- Using KS probabilities
- Significance diluted by number of histograms
- Sample variables: (p_T,M(all combinations),M_T(w/MET),ΔR,$\Delta\phi$,η)

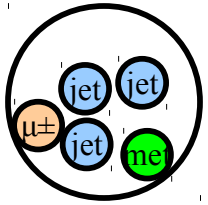


Report significant discrepancies

- > 3σ after trials factor

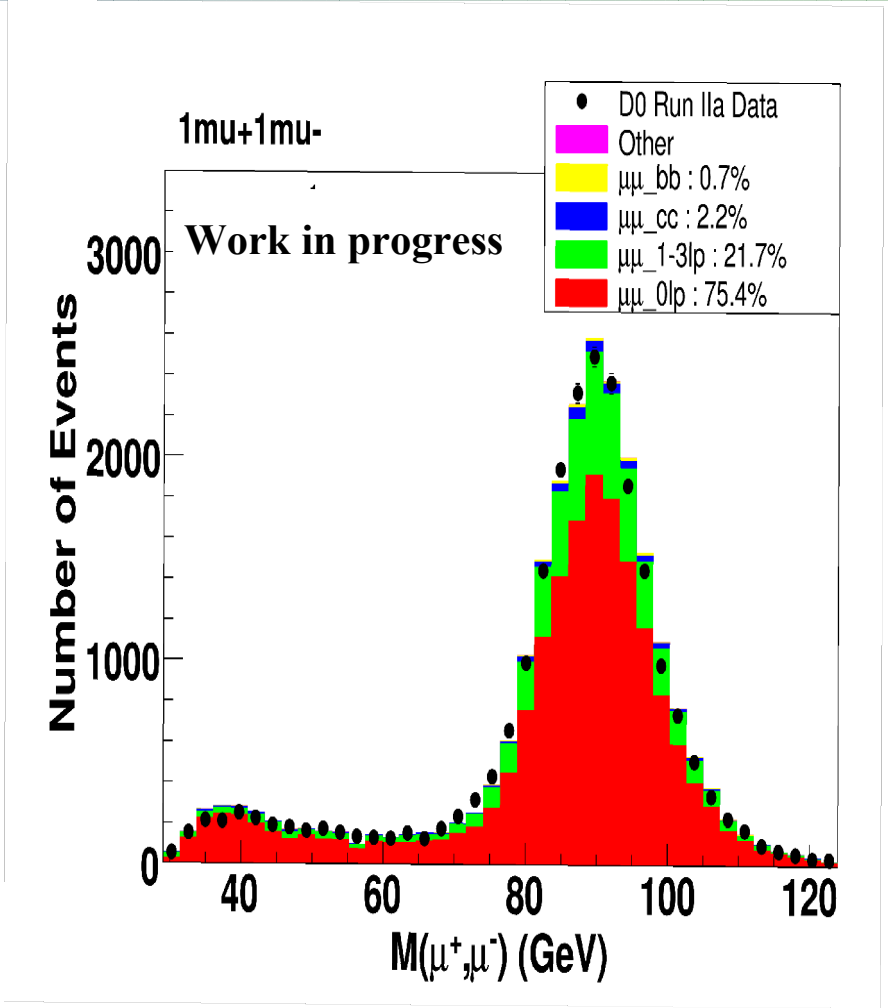




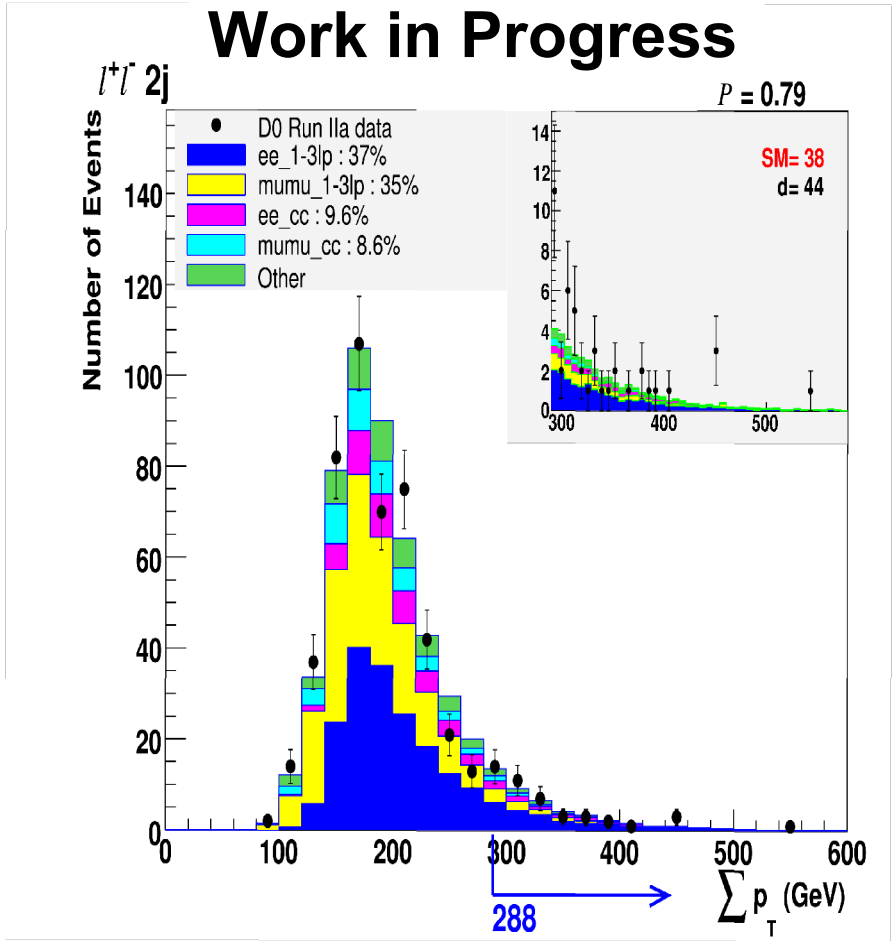


Vista checks > 9000 distributions from 180 exclusive final states

- **When a discrepancy is found, it is investigated**
  - Monte Carlo issue?
  - Detector simulation?
  - Generally within expected systematic errors
- **Some final states combined before passing to Sleuth**
  - Limits trials factor
  - Global charge conjugation
  - 1st /2nd generation equivalence



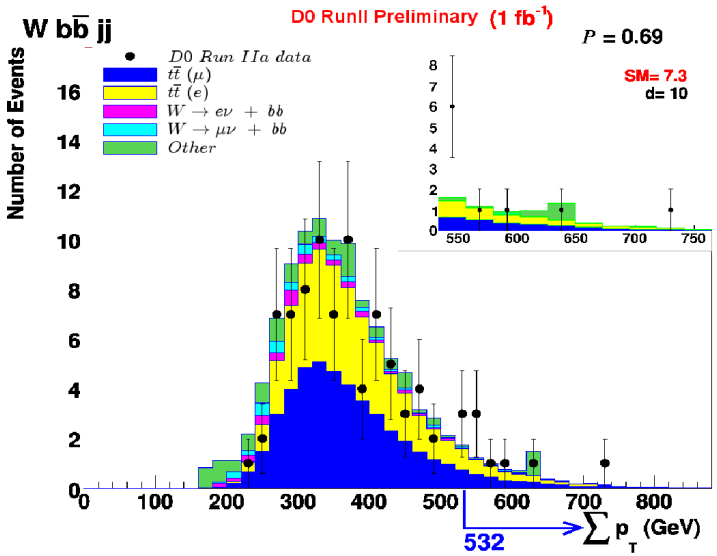
Sleuth performs a quasi-model independent search

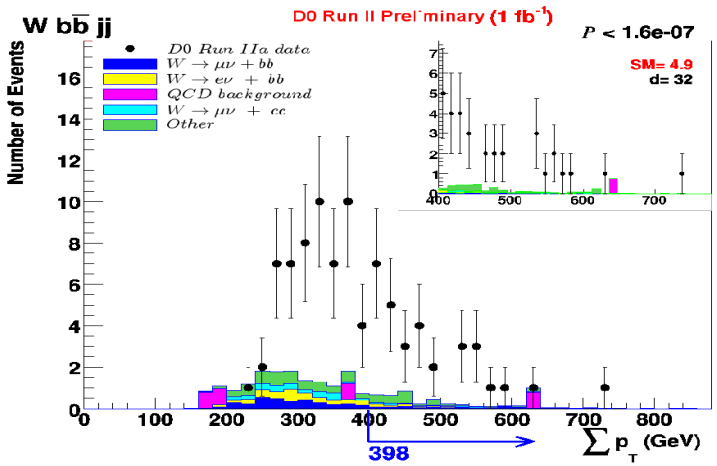


- Focuses on tails of Σp_T for each final state
- Searches for excesses
- Finds cut that maximizes discrepancy
- Simulates experiments to quantify significance
  - Use SM Monte Carlo to perform pseudo-experiments
  - How many to see something as interesting as data?

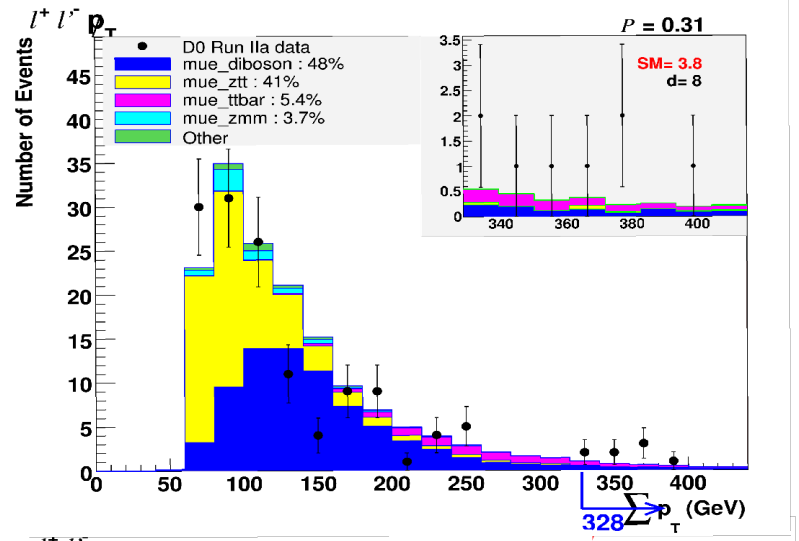
Additional checks for rare states to ensure Sleuth would find them

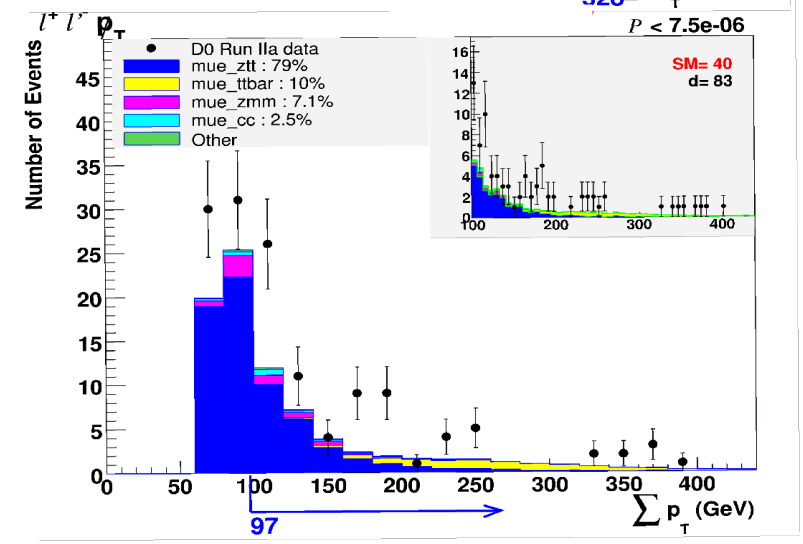
$t\bar{t}$ cross-check





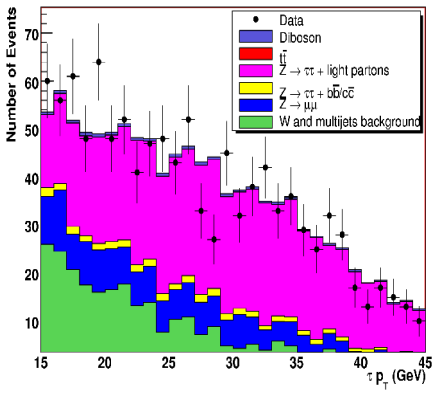
Diboson cross-check



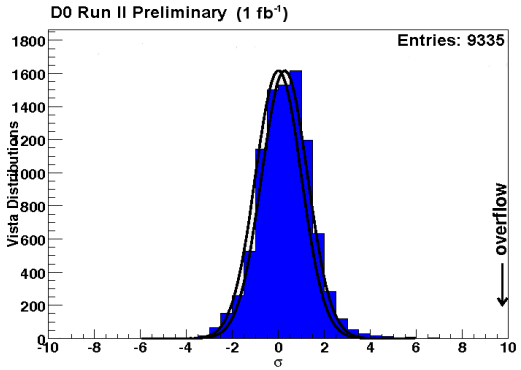


III. Results

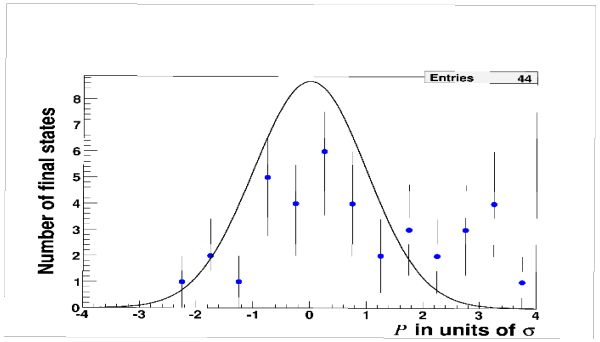
A. Qualitative D0 MIS Analysis Package Fits



B. Vista Model Independent Comparison Results



C. Sleuth "Quasi" Model Independent Comparison Results



Inclusive fitting and checking is done separately for each final state

Final State	Parameters	Fitting Histograms	Checking Histograms
$e + X$	2.5	19	22
$\mu + X$	2.5	19	21
$ee + X$	2	22	18
$\mu\mu + X$	2	22	22
$e + \tau + X$	3	13	8
$\mu + \tau + X$	2	13	8
$e + \mu + X$	2	13	6

Single isolated μ + jets dominated by W boson production

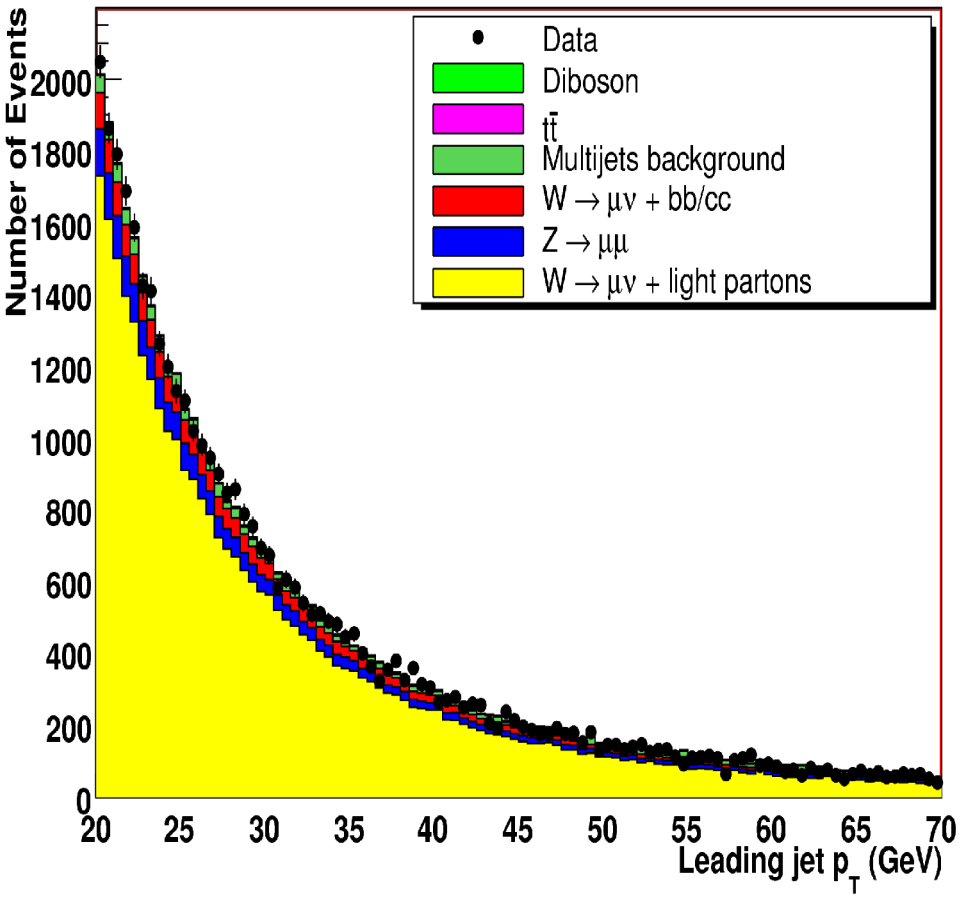
Fit parameters

- Multijets Background
- W w/light partons
- W w/heavy flavor

Event Cuts

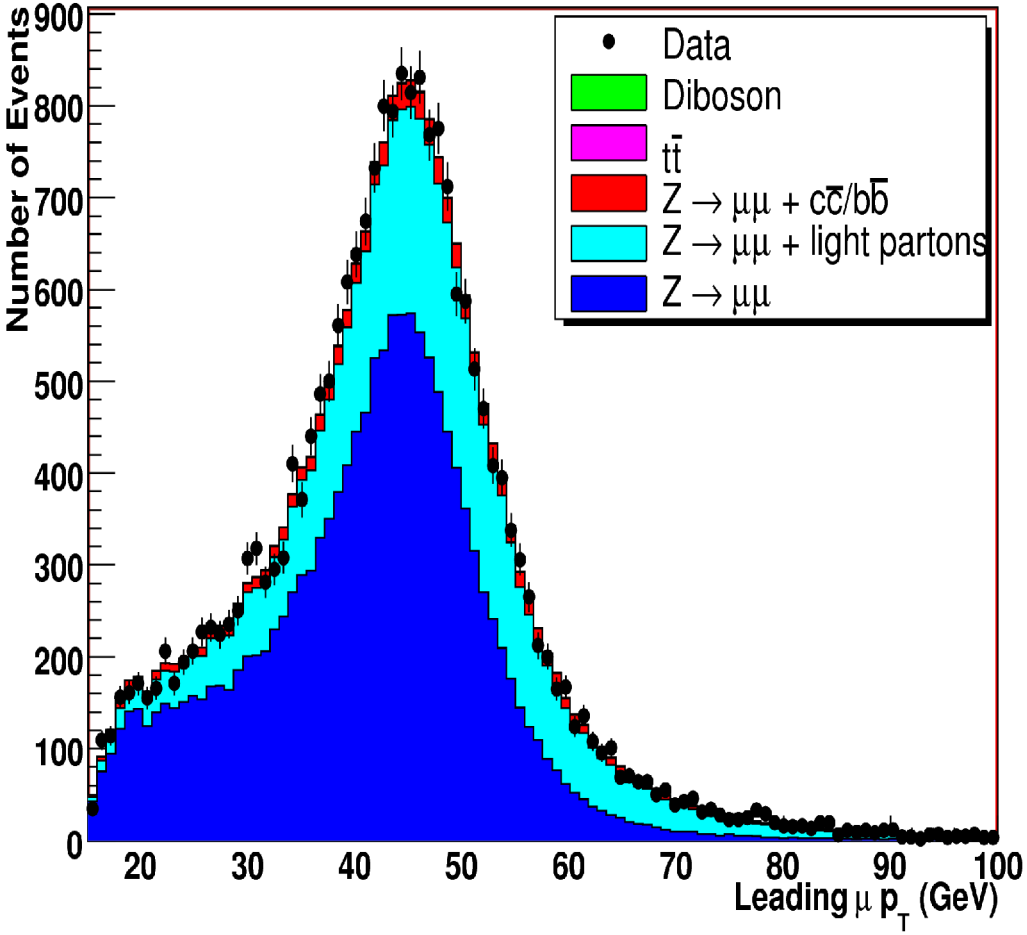
- Single μ or e w/ at least one jet
  - μ p_T > 25 GeV
  - e p_T > 35 GeV
  - jet p_T > 20 GeV
  - MET > 20 GeV

Work in progress



Lower p_T cut for dileptons, otherwise similar to lepton + jets states

Work in progress



Fit Parameters

Drell-Yan

- w/no jets
- w/at least one light parton
- w/at least one heavy parton

Event Cuts

- Same iso cuts as single μ data
- μ p_T > 15 GeV
- e likelihood > 0.85
- e p_T > 20 GeV

$\mu + \tau$ and $e + \tau$ use single-prong, hadronic τs

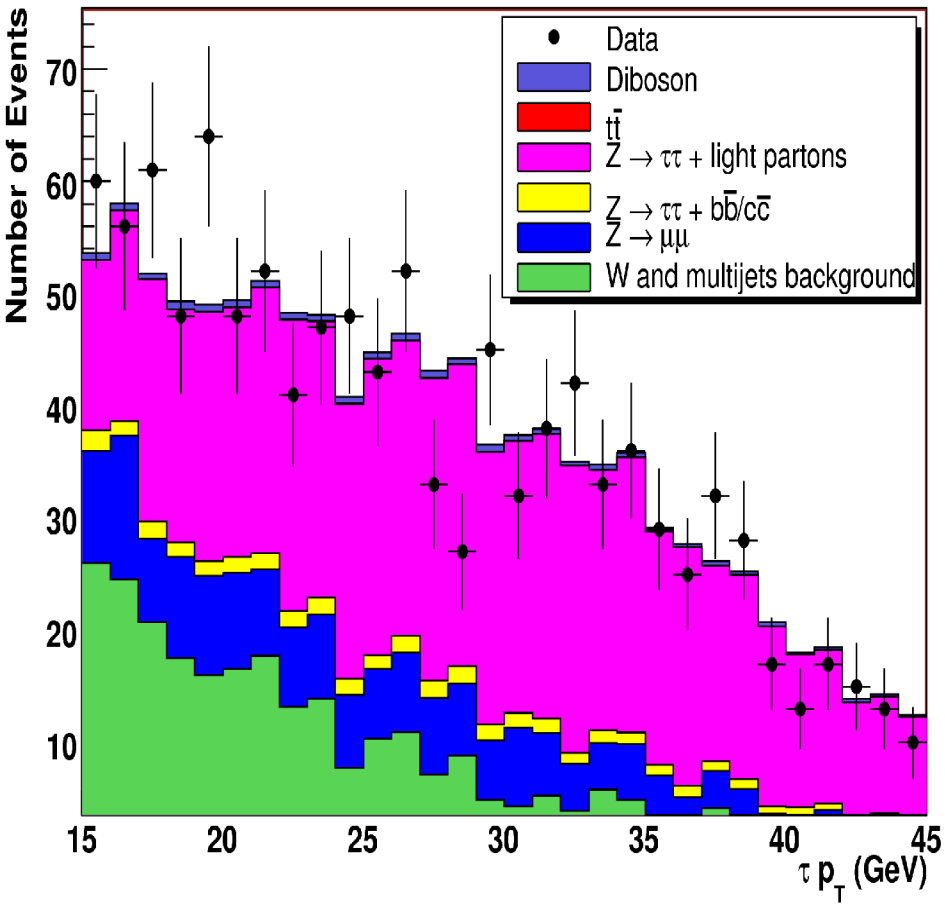
Fit Parameters

- Combine multijet and W backgrounds from data
  - Cannot differentiate from pure multijet & $W \rightarrow \mu/e\ \nu$ + jet mis-ID
- Drell-Yan $\mu\mu$, ee: fixed
- Drell-Yan $\tau\tau$

Event cuts

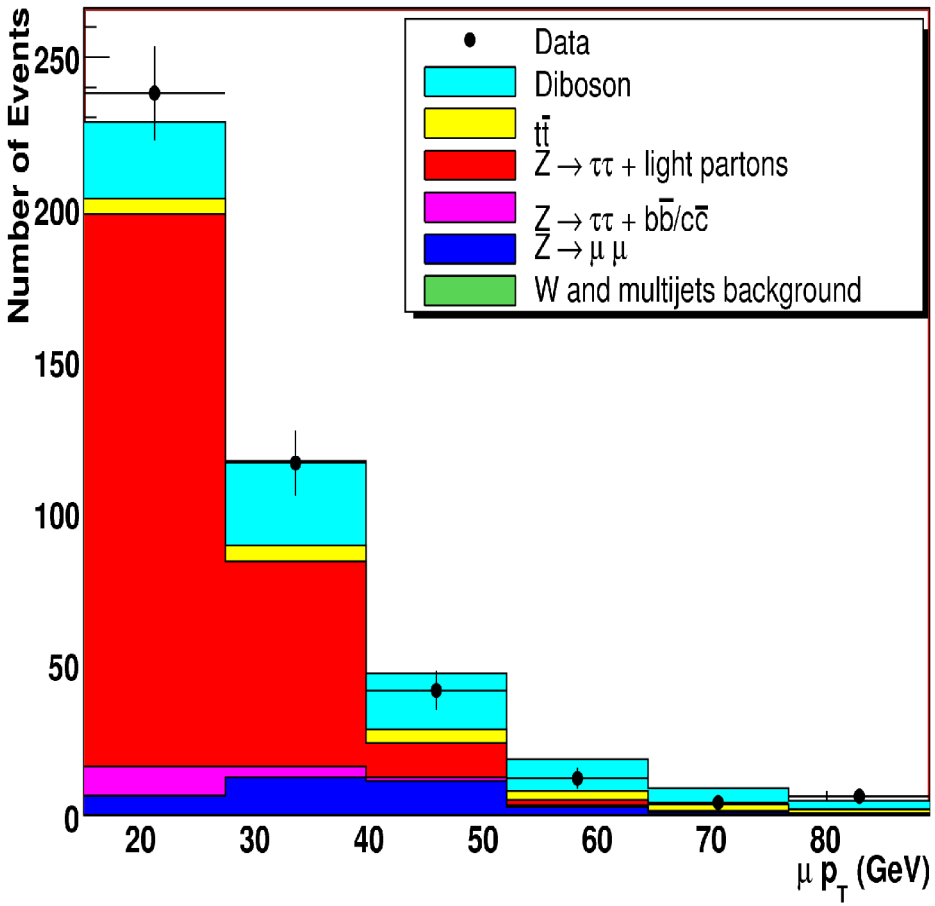
- μ,e $p_T > 15$ GeV
- τ $p_T > 15$ GeV
- Separate NNs for single π and dipion decays (e sep)

Work in progress



Final state of μ + e also dominated by Drell-Yan to ditaus

Work in progress



Fit Parameters

- Drell-Yan μμ: fixed
- Drell-Yan ττ

Event cuts

- μ p_T > 15 GeV
- μ η < 1.5
- e p_T > 15 GeV
- e η < 1.1
- Use background of real μs and fake e's
  - Cuts eliminate most of the fake background

Vista Results

Large majority (>97%) of event counts and histograms agree, but we focus on disagreements

DØ sees 4/180 event count and 24/9335 shape discrepancies

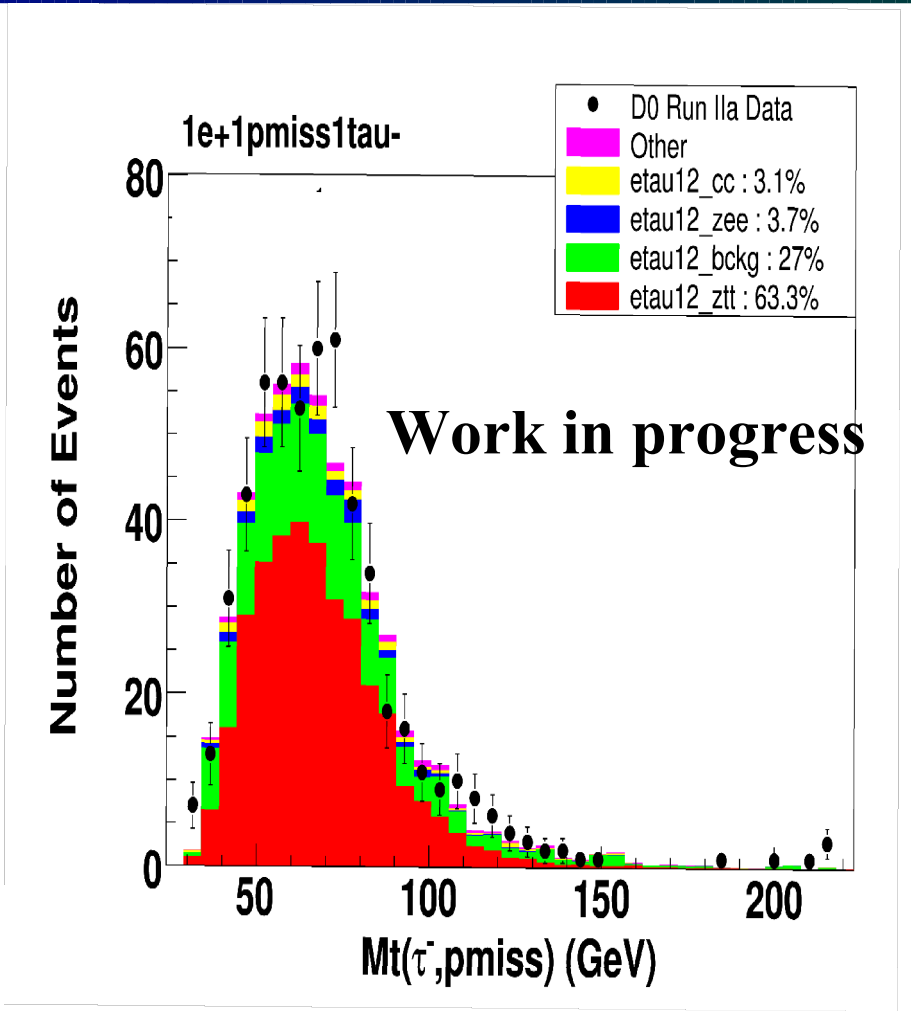
1 fb^{-1}

3 known modeling issues

- **η-dependent trigger efficiency in μ + jets + MET**
- **Muon resolution in $\mu\mu$ + MET**
- **Jets misidentified as photons in γ states**

DØ Event Count Discrepancies	σ After Trials Factor
$\mu + 2$ jets $+ \not{E}_T$	9.3
$\mu + \gamma + 1$ jet $+ \not{E}_T$	6.6
$\mu^{\pm}\mu^{\mp} + \not{E}_T$	4.4
$\mu^{\pm}\mu^{\mp} + \gamma$	4.1

Most histograms are within expected statistical fluctuations

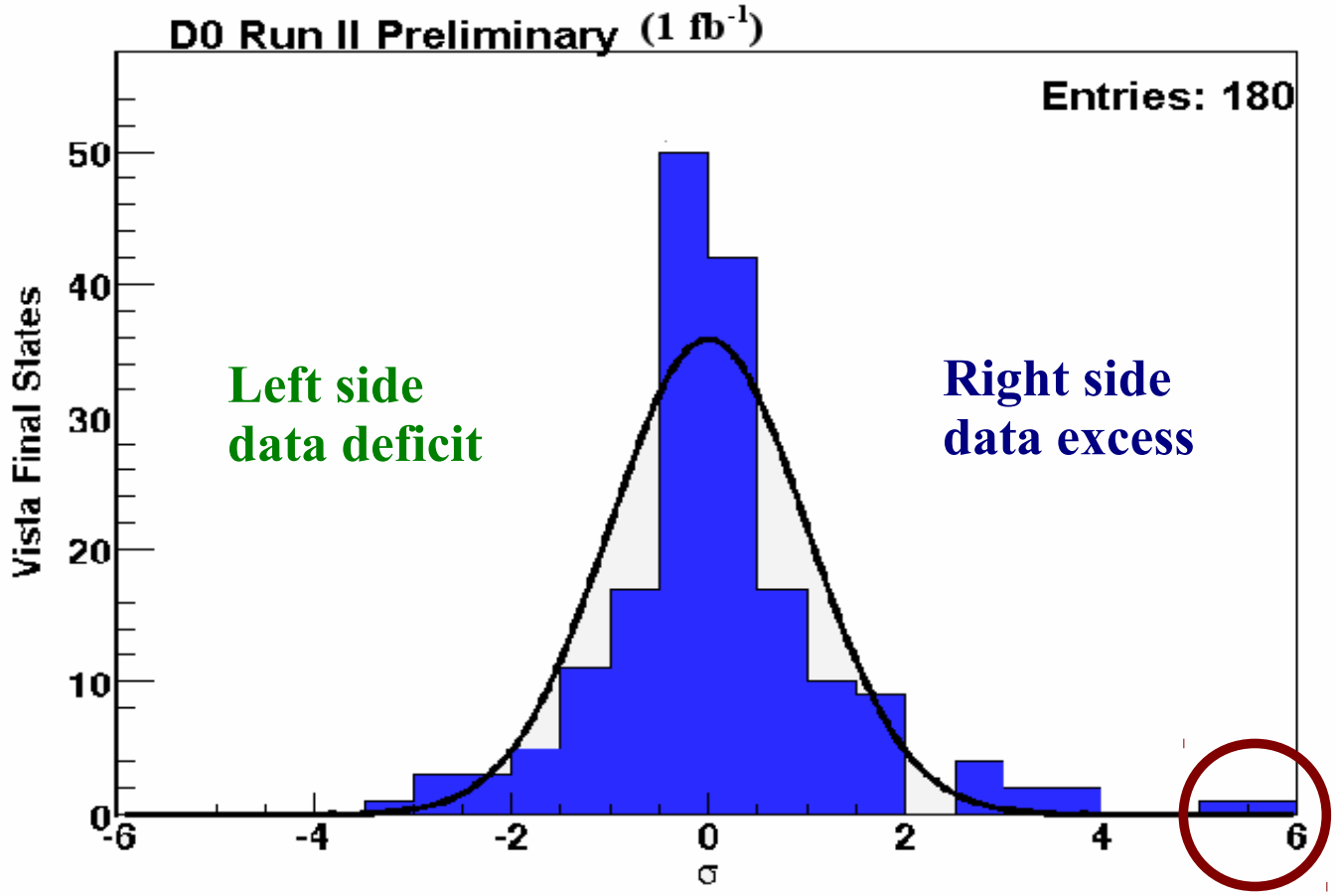


Note we are using a coarse correction model

Many plots show minor differences

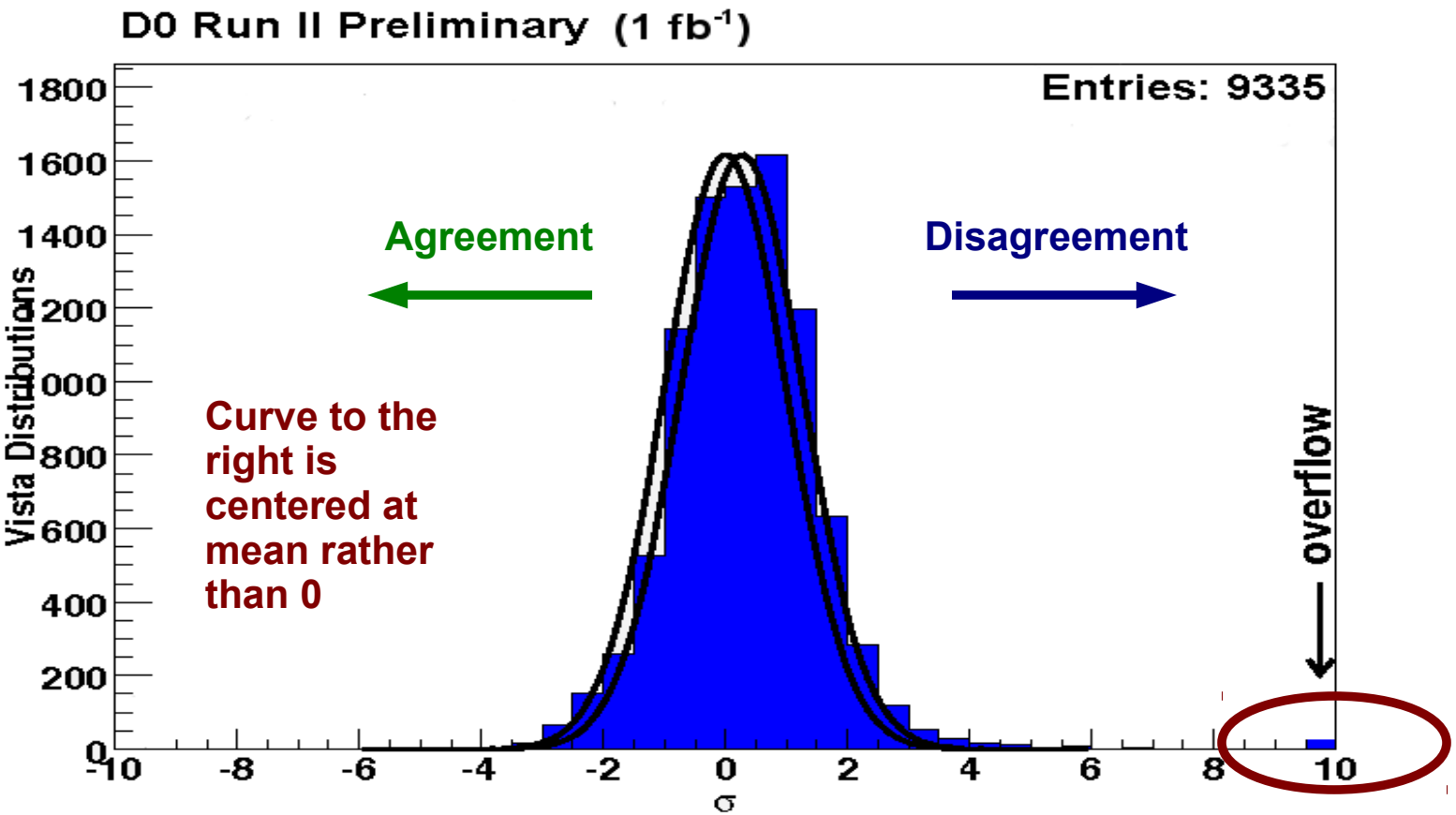
Use KS to determine signficance

Distributions of Final States' Data/MC Count Agreement



DØ Curve: Gaussian

Distributions of KS Probabilities in Histogram Shapes

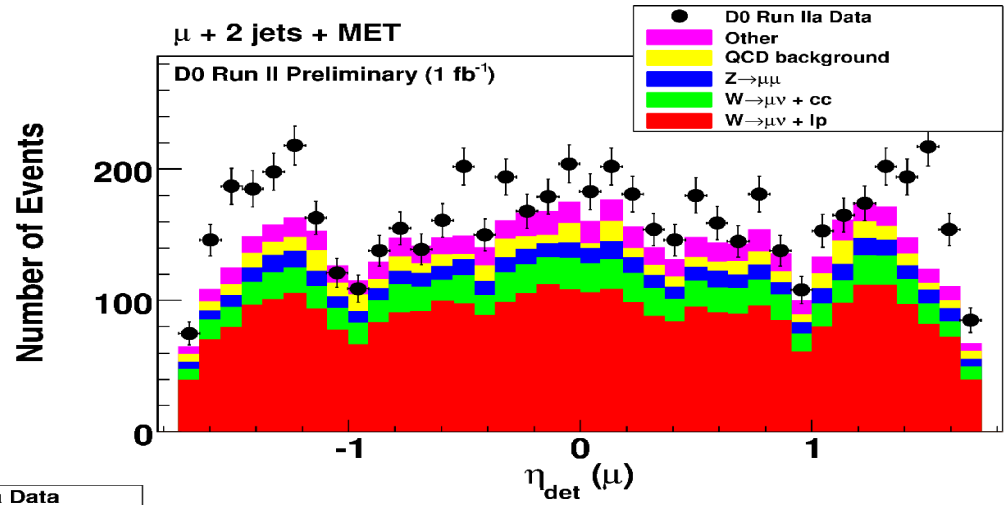


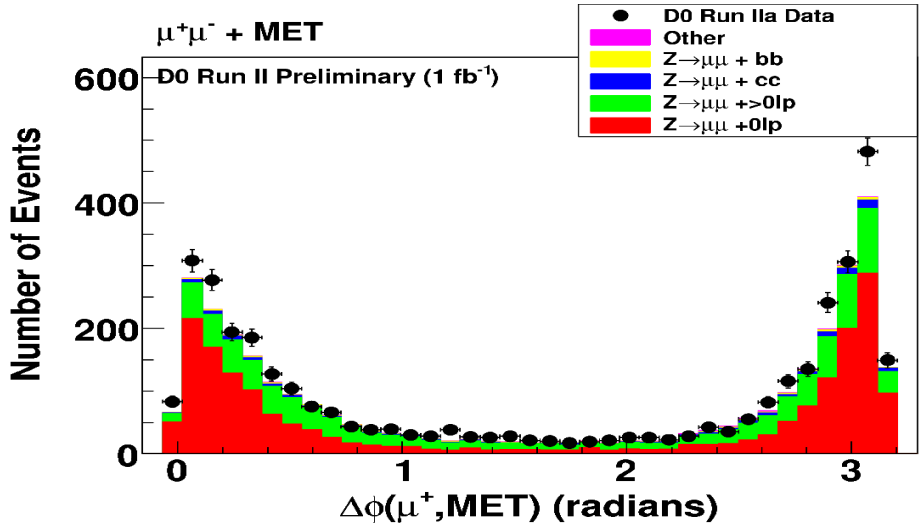
24 histograms show discrepancies after trials

Two number discrepancies point to subtle modeling issues

η-dependent trigger efficiency

- single muon
- muon + jets





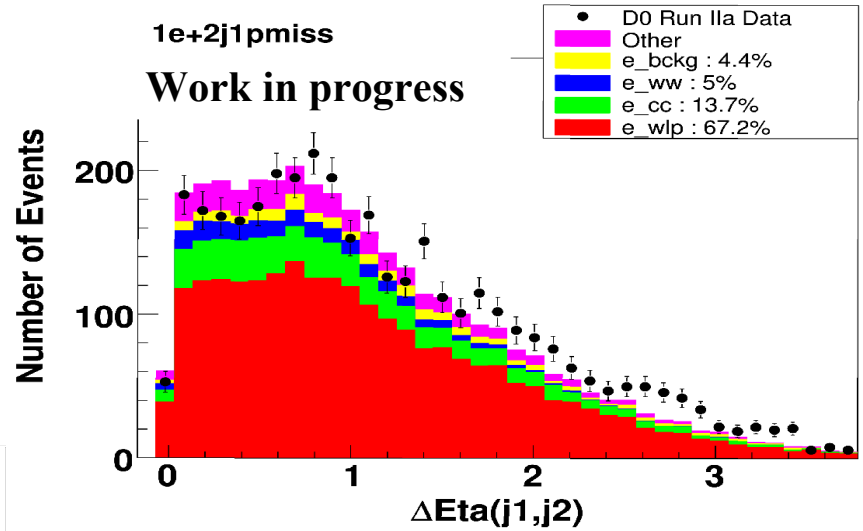
Track resolution modeling

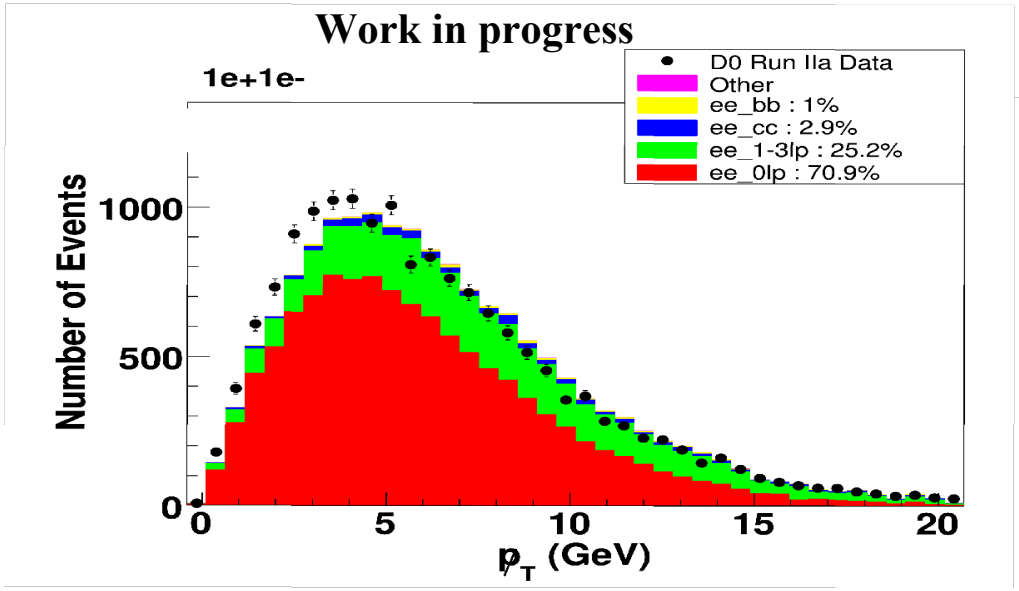
- high-p_T μ
- muon and missing transverse energy are back-to-back
- appears again in Sleuth

Two of the most discrepant histograms based on shape

Alpgen-Pythia jet Δη disagreement

- Related distributions also affected
- Also seen in μ + jets



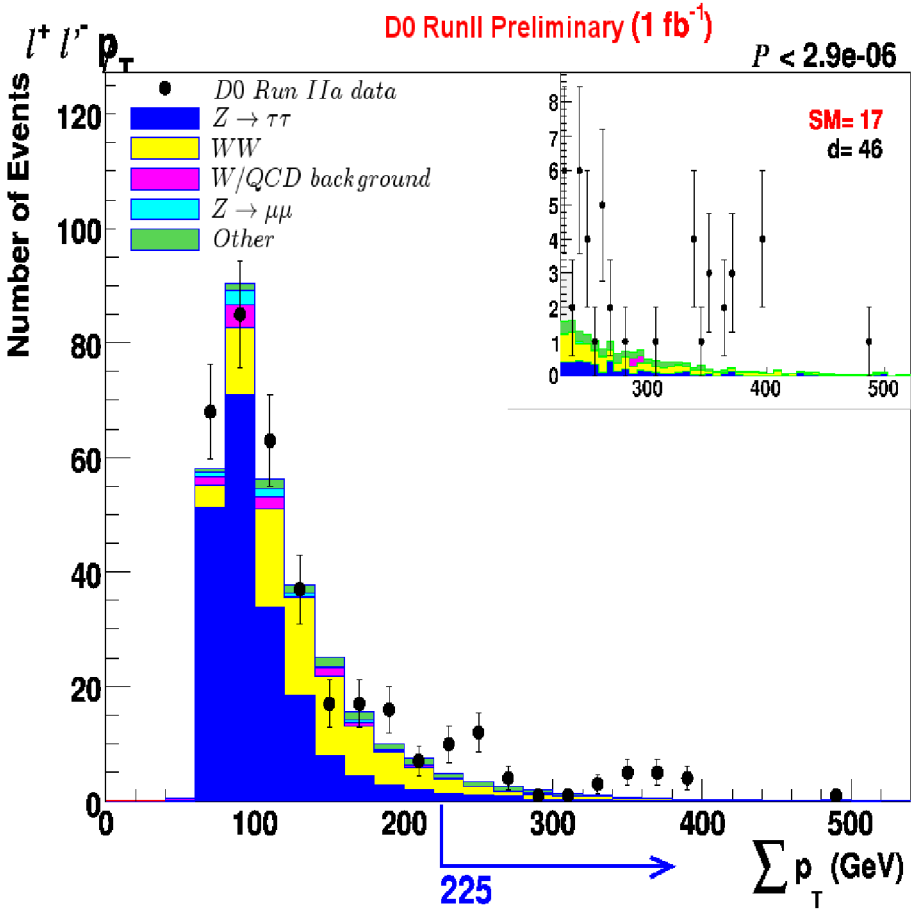


Some disagreement at very low MET

- Little effect on high-p_T distributions
- Not particularly relevant in Sleuth
- Possible imperfection in detector modeling

DØ Sleuth Results

Sleuth looks for data excess in high-p_T tails



Opposite sign μ e MET

Events in tails

- large μ p_T
- back-to-back with MET
- same problem as μμ + MET

Other highly populated final states w/ MET

- similar excess w/high-p_T μ

Vista number excesses also trigger Sleuth "discoveries"

- μ + 2 jet + MET
- bulk of distribution

Distribution of Sleuth results show common theme

D0 Run II Preliminary

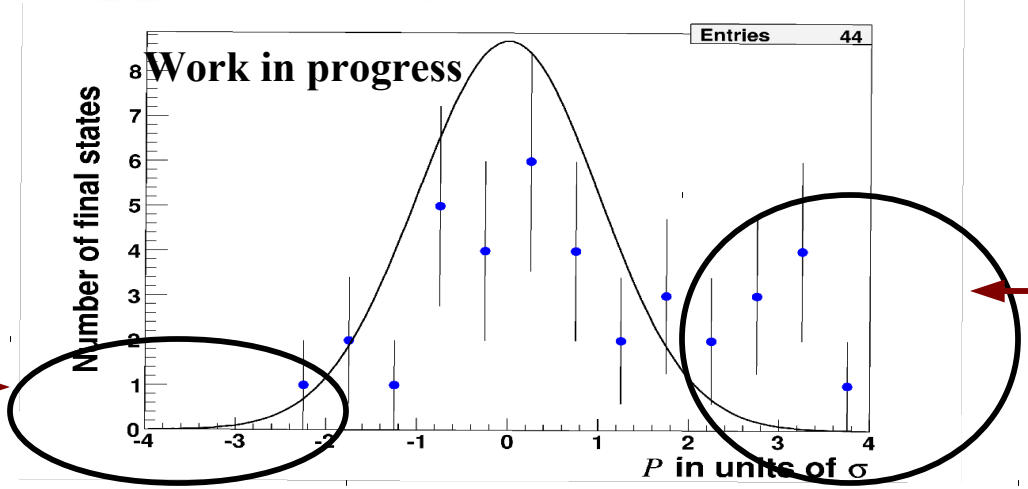
Final State	$\mathcal{P}$
$\ell^+\ell'^-$ + MET	2.9 E-6
ℓ + MET	.00082
$\ell^+\ell'^-$	.0031
$\ell^+\tau^-$ + MET	0.006
$\ell^+\tau^+$	0.0066

Most significant states all show issues with μ resolution

Search concentrates on a single most significant state

Just leading state crosses threshold after trials... but systematic problem

Few states on far left as would be expected for a background overestimate



Cluster of states with high σ; many related to μ resolution issue

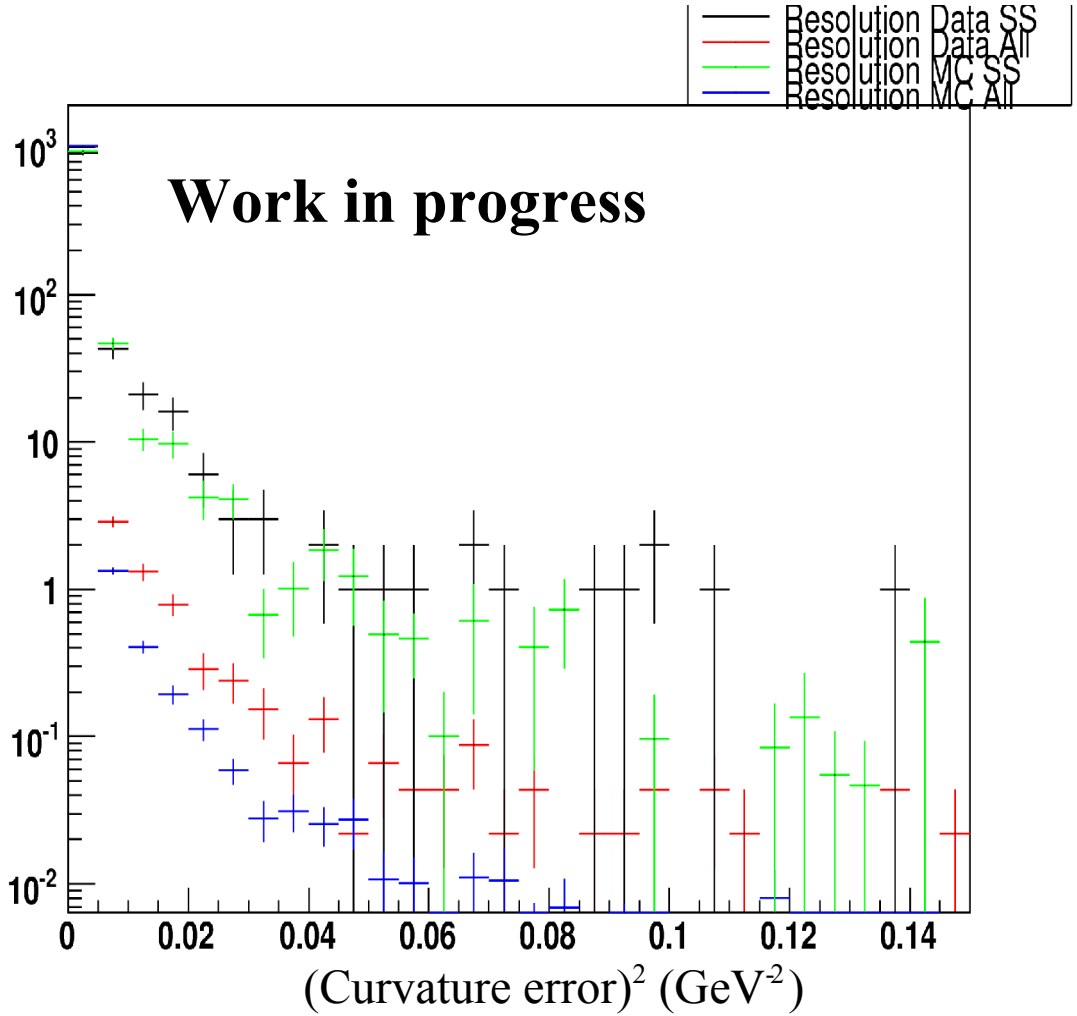
Conclusions

- Vast majority of event counts and distributions are in agreement at DØ
- Vista considers event counts and the shapes of 1-D histograms
  - Searched over 1 fb^{-1}
  - Four event count discrepancies
  - 24 shape discrepant distributions
  - All of the given discrepancies at this time point to modeling issues
- Sleuth examines specific distributions for high-p_T excesses
  - One Sleuth discrepancy
  - Probably related to muon momentum resolution
- While no compelling case for new physics found in the search..
  - Background implementation found to describe the detector well
  - Pointed to interesting modeling features that should be explored

Backups

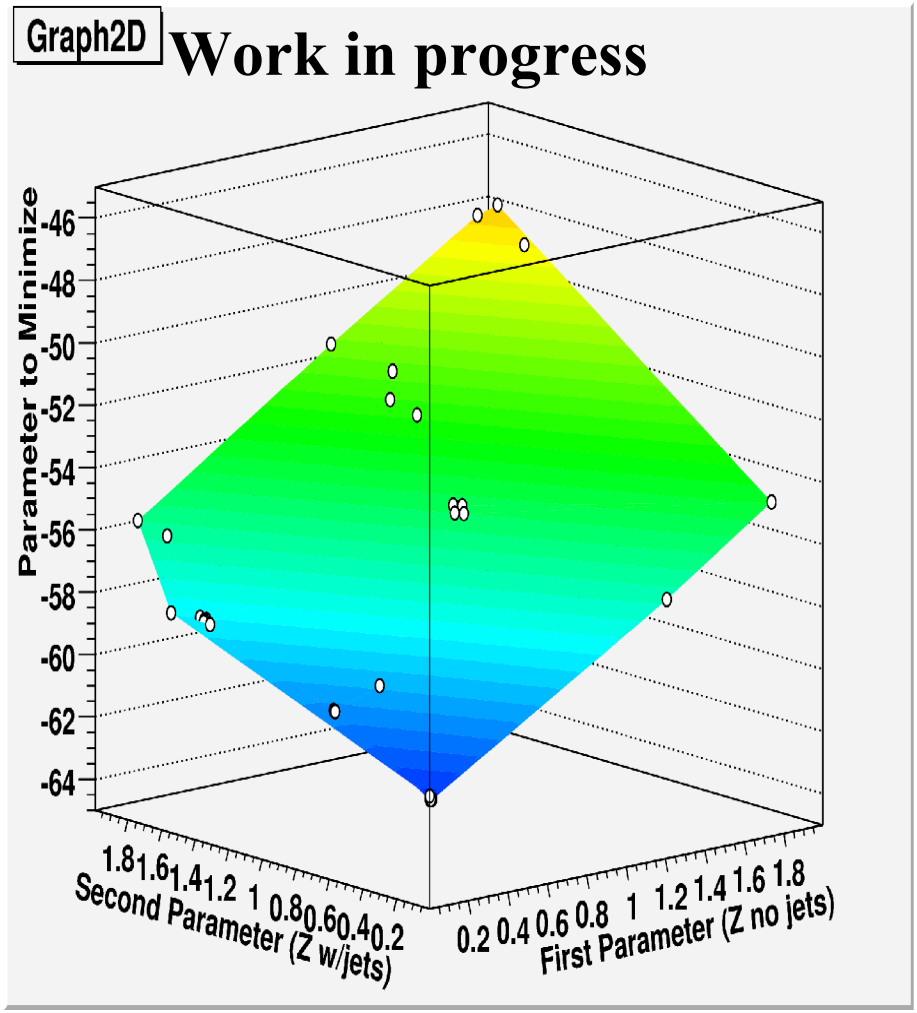
Same sign final states are crucial in final state count-based search

- Error in curvature shown at right
- Found to be best predictor of sign error
- Tight cuts used to ensure a Z boson dominated sample
  - Should be opposite sign
- All inputs normalized to same sign data
- Simulation underestimates sign error rate w/standard scale factor

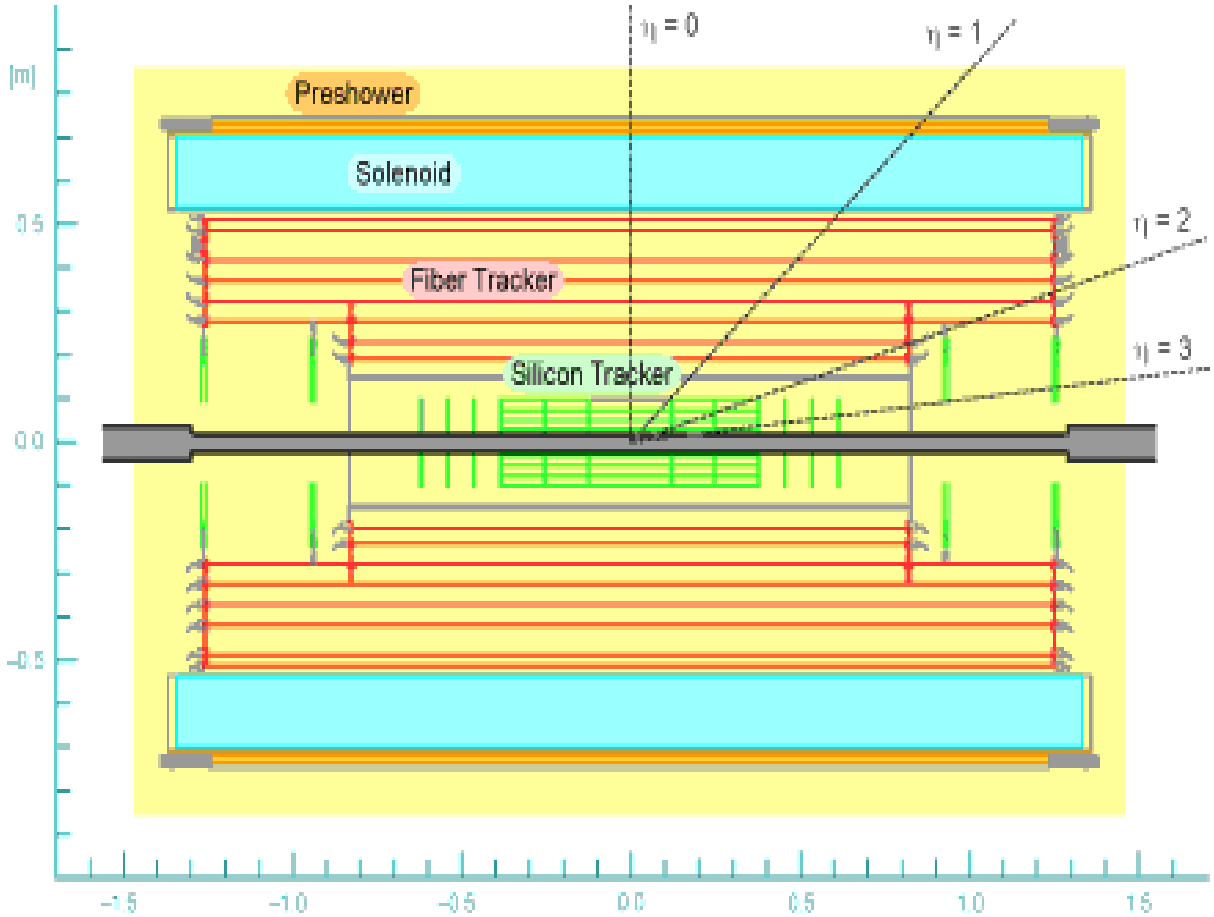


Use log-likelihood fit method

- Each final state has two or three floating parameters
- Some fixed parameters: tt, diboson, output of other fits
- Loop over all fitting histograms
- Minimize
  - Minimum of negative log-likelihood
  - Minuit to minimize input log-likelihood function

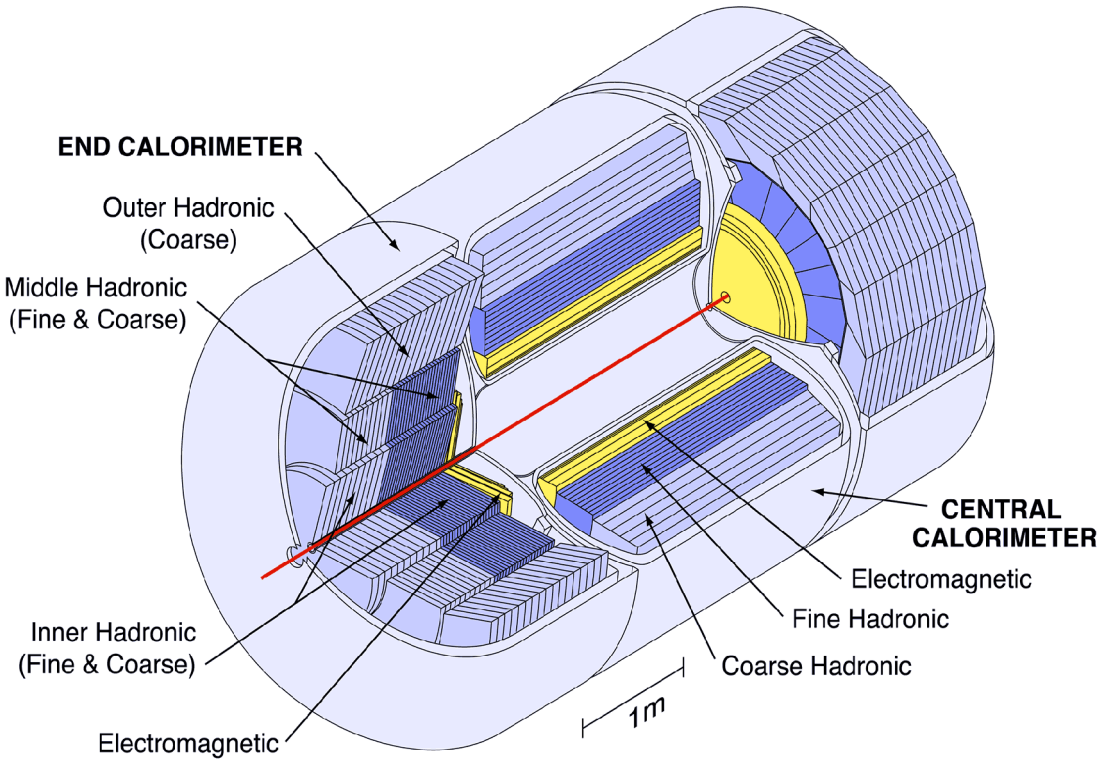


Tracking



- **Silicon microstrip tracker**
  - Tracking and vertexing
- **Central fiber tracker**
  - Scintillating fibers
  - Fast and continuous readout
- **Central and forward muon systems**
  - Drift tubes and scintillators
  - $\eta \approx 2$

Calorimeters



Central and end calorimeters

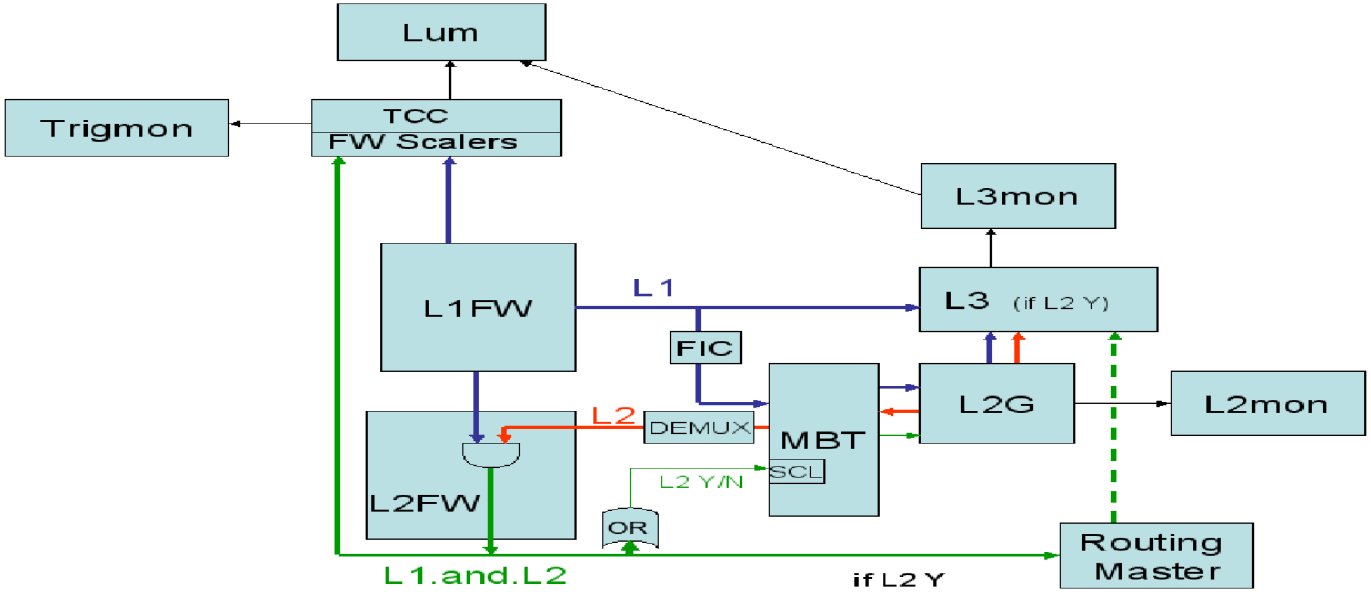
- **Uranium/LA**
- **$\eta \approx 4$**
- **Identifies e, γ, τ, jets and missing transverse energy**

The path of Level 2 Decisions

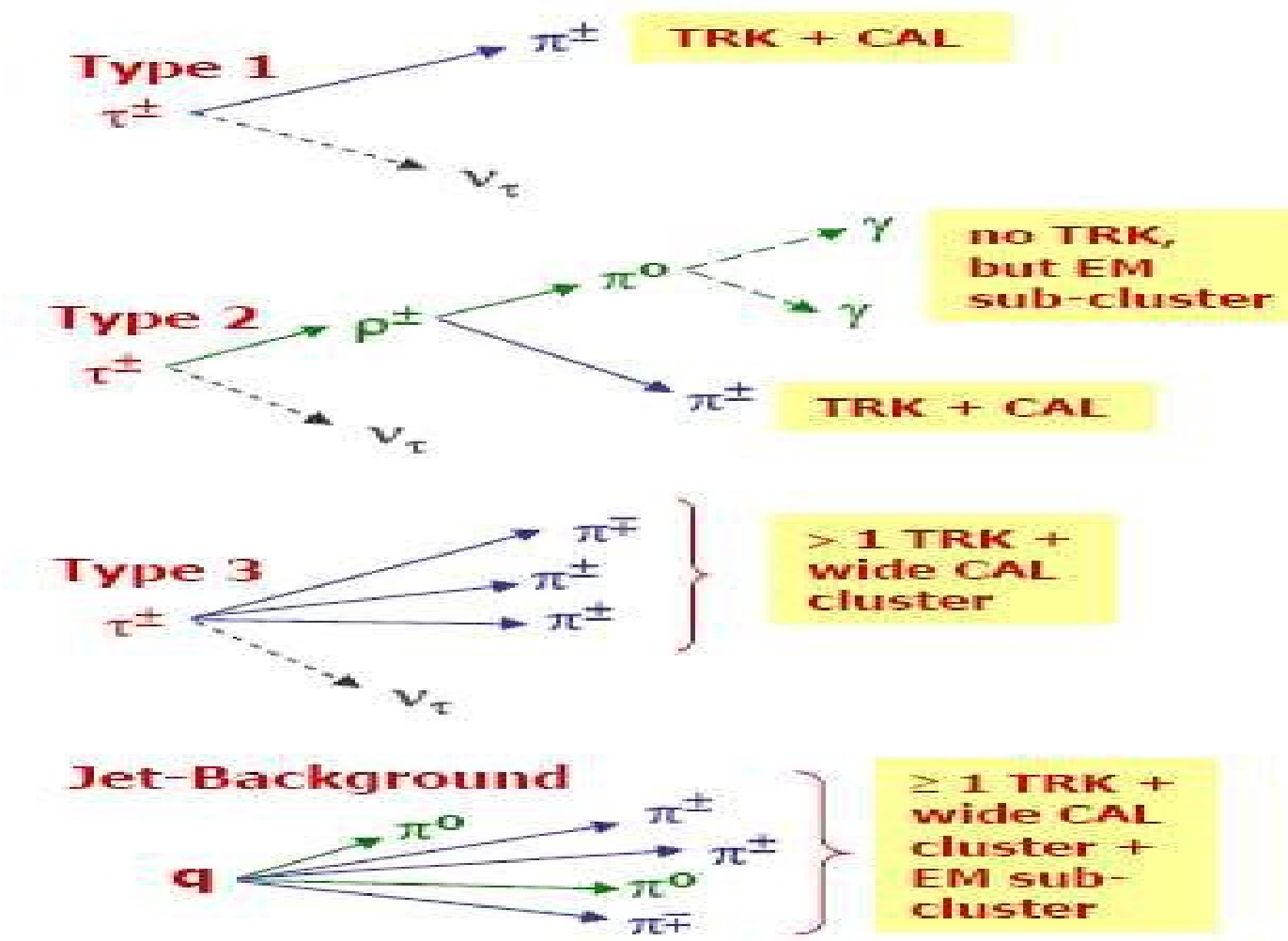
Level 2 Global makes the L2 trigger decision

This flows out to trigger framework

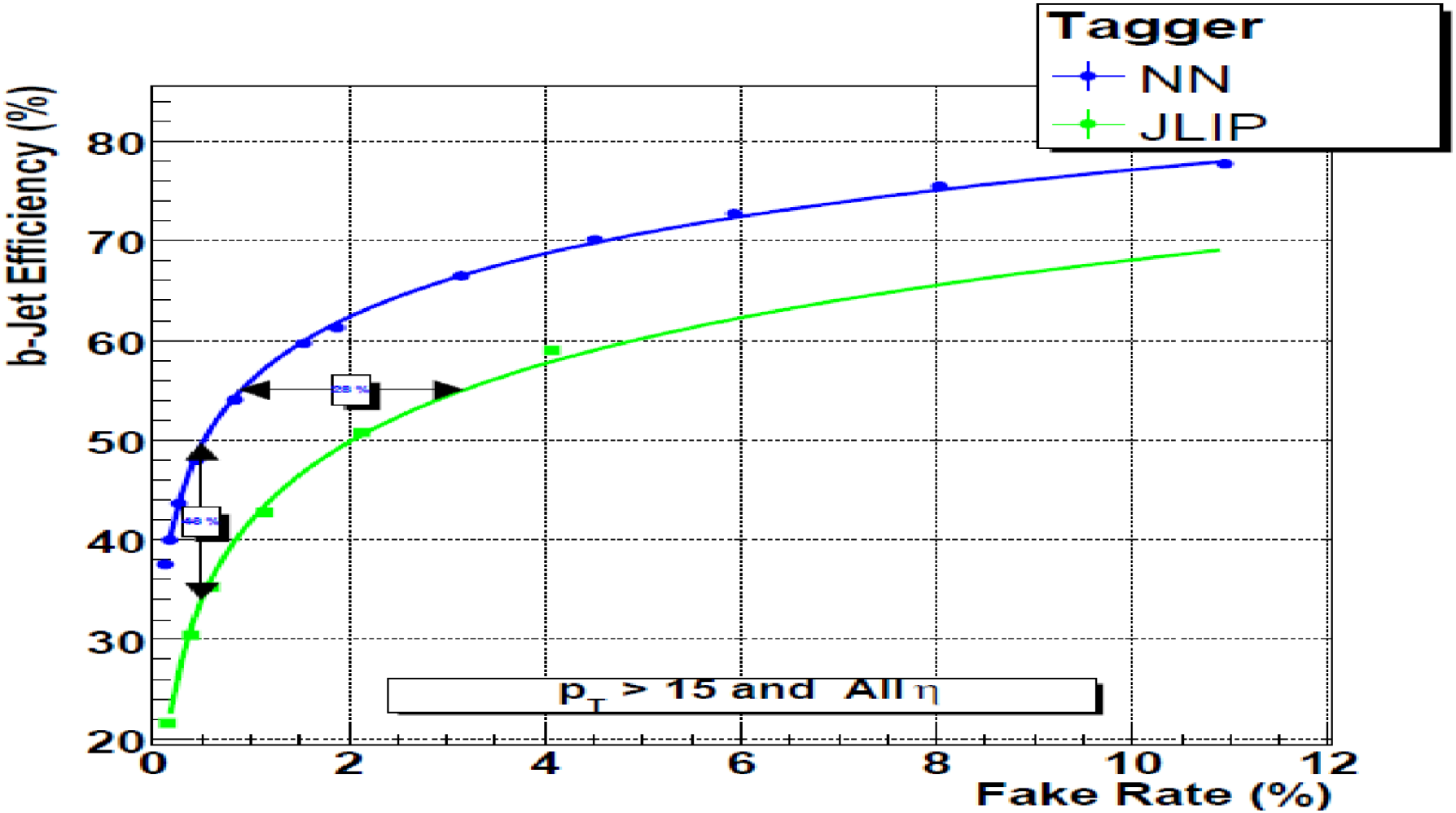
From trigger framework, decision sent to all other crates



Tau Decays


Type 1
$\tau^{\pm}$
$\pi^{\pm}$
ν_{τ}
TRK + CAL
Type 2
$\tau^{\pm}$
$\rho^{\pm}$
π^{0}
γ
γ
$\pi^{\pm}$
ν_{τ}
no TRK, but EM sub-cluster
TRK + CAL
Type 3
$\tau^{\pm}$
$\pi^{\mp}$
$\pi^{\pm}$
$\pi^{\pm}$
ν_{τ}
> 1 TRK + wide CAL cluster
Jet-Background
q
π^{0}
$\pi^{\pm}$
$\pi^{\pm}$
π^{0}
$\pi^{\mp}$
≥ 1 TRK + wide CAL cluster + EM sub-cluster

b NN vs. JLIP

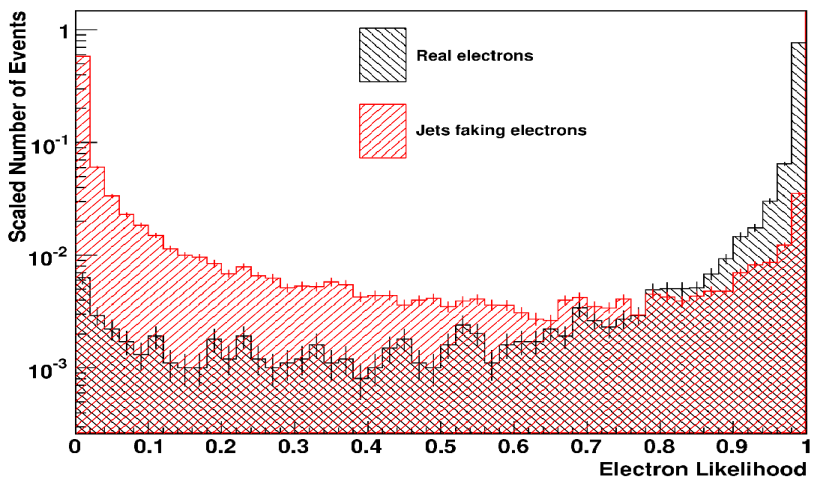

Tagger
NN
JLIP
b-Jet Efficiency (%)
80
70
60
50
40
30
20
0
2
4
6
8
10
12
Fake Rate (%)
p_T > 15 and All η

Electron multijet background uses loosened likelihood cuts

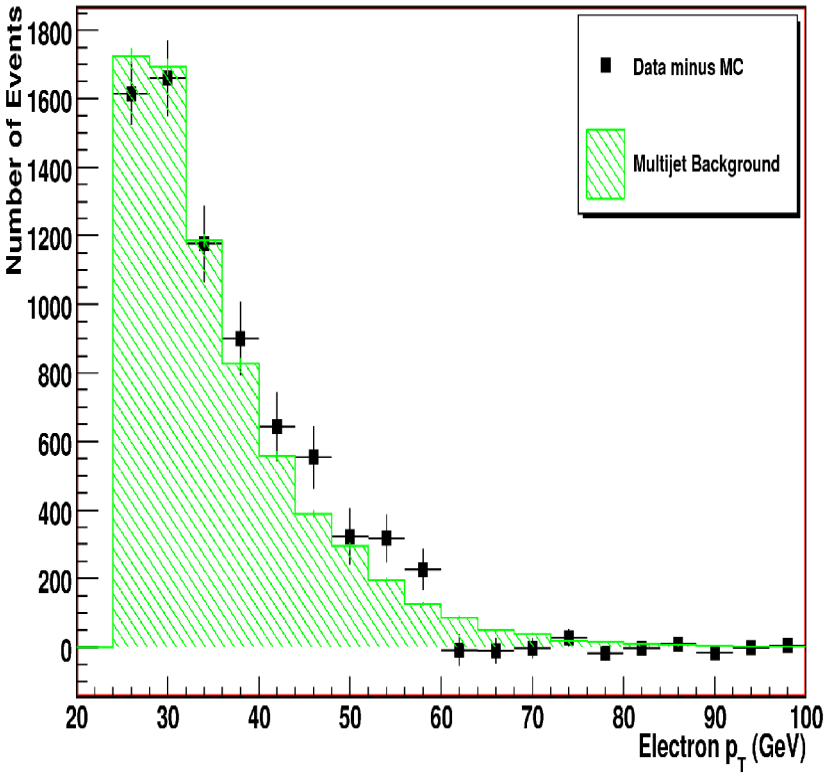
Jet faking mechanism

- Large energy fraction in EM calorimeter
- Single dominant track

Electrons use likelihood variable to separate from jets



Work in progress



Multijets use

0.25 < likelihood < 0.8

Muon multijet background uses isolation variables

Jet faking mechanism

- Heavy jets decay to real muons
- Very little punch-through at D0

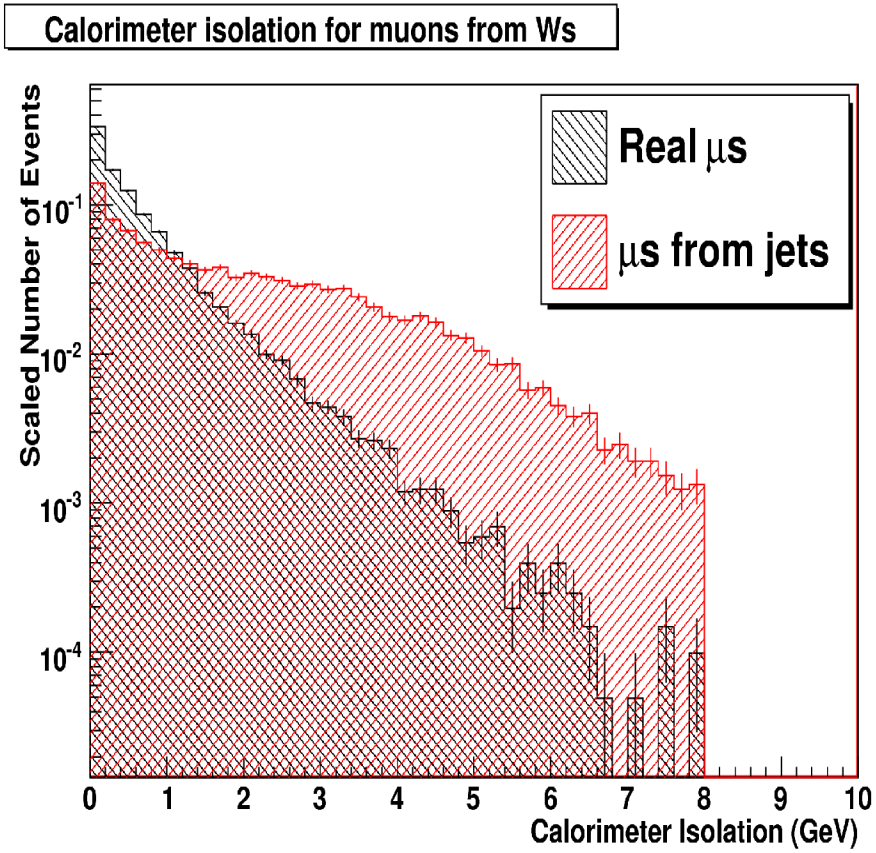
Is muon associated with a jet?

- Hollow cone in calorimeter
- Cone around matched track

Multijets use

- 4.0 < Cal Iso < 8.0 GeV
- 4.0 < Trk Iso < 8.0 GeV

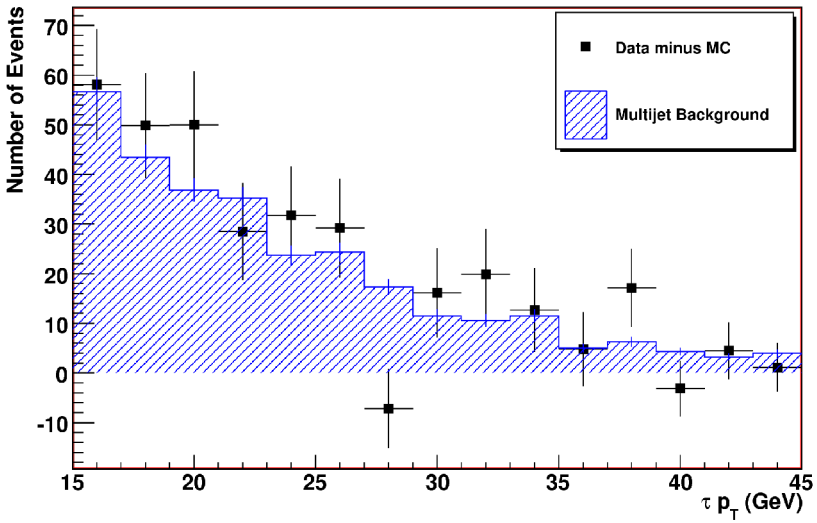
Work in progress

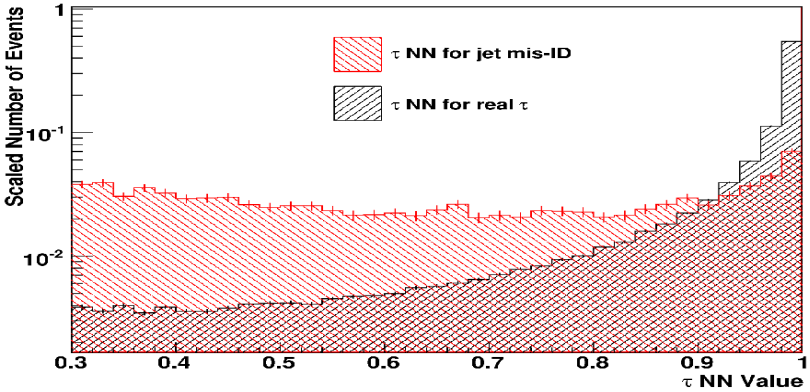


Hadronic tau multijet background determined from neural network values

- **Jets fake hadronic taus**
  - … fairly easily
  - Narrow energy deposit in calorimeter
  - Single track w/ or w/o EM cal deposit
- **Each type of hadronic tau decay has NN**
  - Includes (cal, preshower, and tracking vars)
- **$\tau \rightarrow \pi^{\pm} + \pi_0$ include electron separation NN**
- **Multijets background**
  - 0.3 < NN < 0.8

Work in progress





Final fitting must incorporate constraints

Ratio of heavy flavor/light flavor scale factors for single process should be flavor independent

Process jet multiplicity scale factors also flavor independent

Other analyses should see similar results

Jet multiplicity flavor independence:

$$SF = SF_{ee}\frac{\#events(ee)}{\#events(ee+\mu\mu)} + SF_{\mu\mu}\frac{\#events(\mu\mu)}{\#events(ee+\mu\mu)}$$

Triggers passed by event stored as list: 1 bit / trigger

List of trigger names stored internally

If event passes particular trigger, bit is set

All physics triggers included on list

Use standard efficiency for trigger suites

Inclusive Final State	Trigger Suite
$e + X$	Single e
$\mu + X$	Single μ
$ee + X$	No trigger selection
$\mu\mu + X$	Single μ
$e + \tau + X$	Single e
$\mu + \tau + X$	Single μ
$e + \mu + X$	Single μ

After final weighting and checks, create text files for Vista

Vista uses text file input for standardization across experiments

Simple conversion from high-p_T format to text file

Info saved:

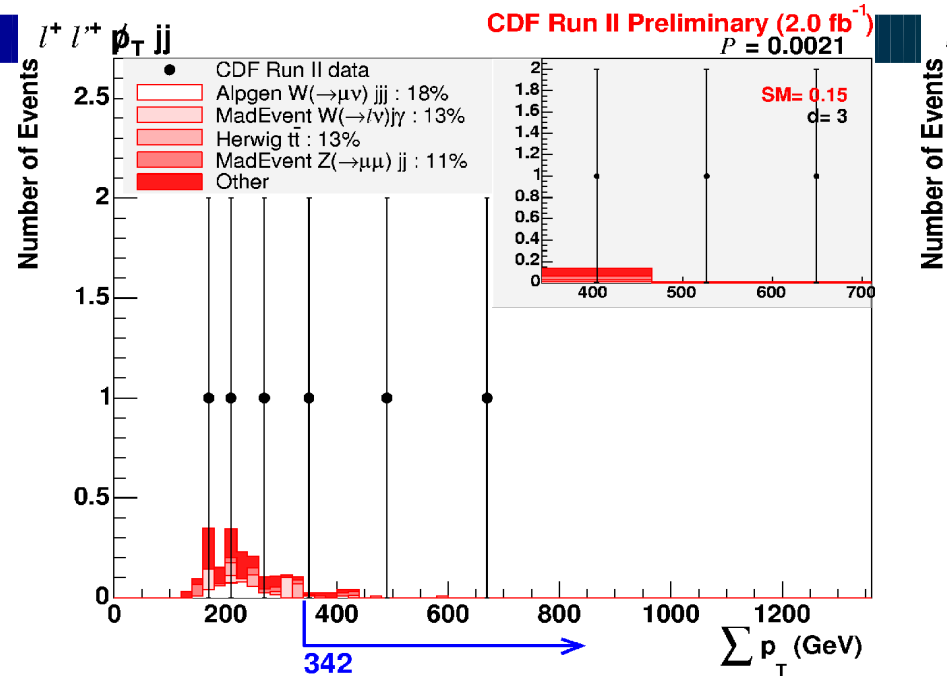
- run #, event #, vertex, other global info
- object ID, charge, 4-vector

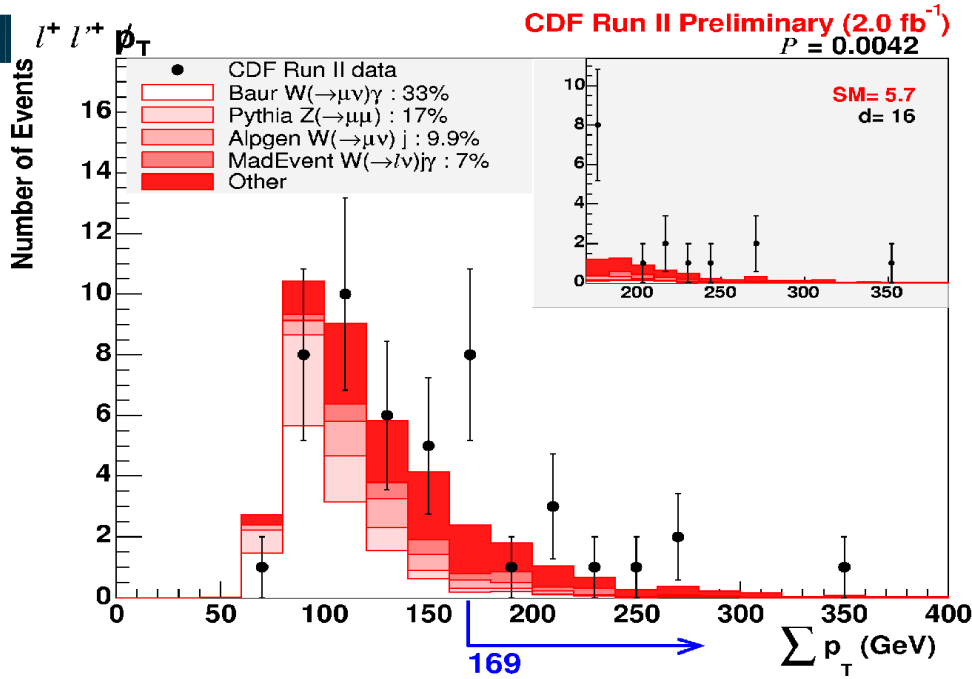
Text file creation

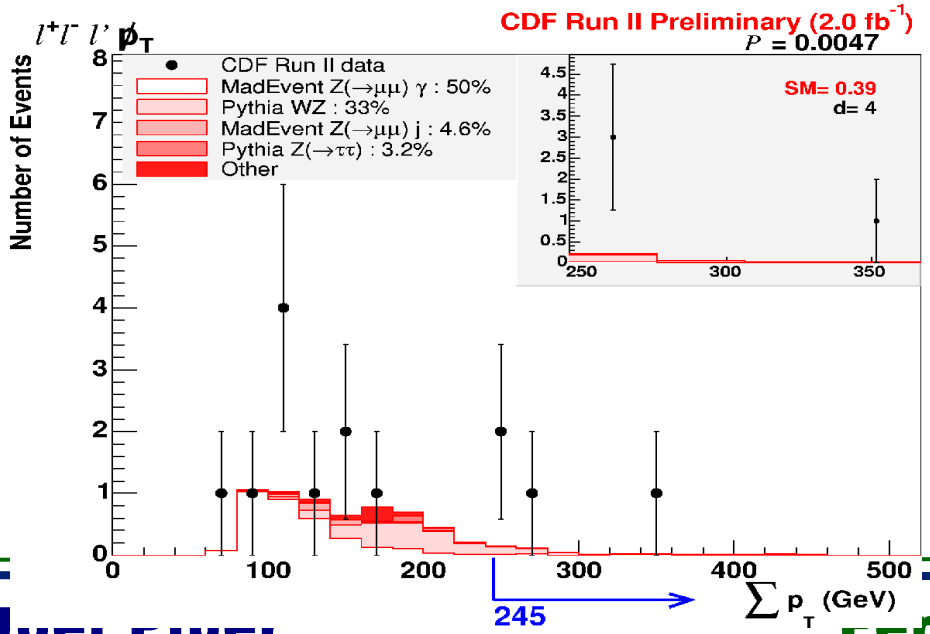
- Same process as histogram creation
- Now one additional weight from fit

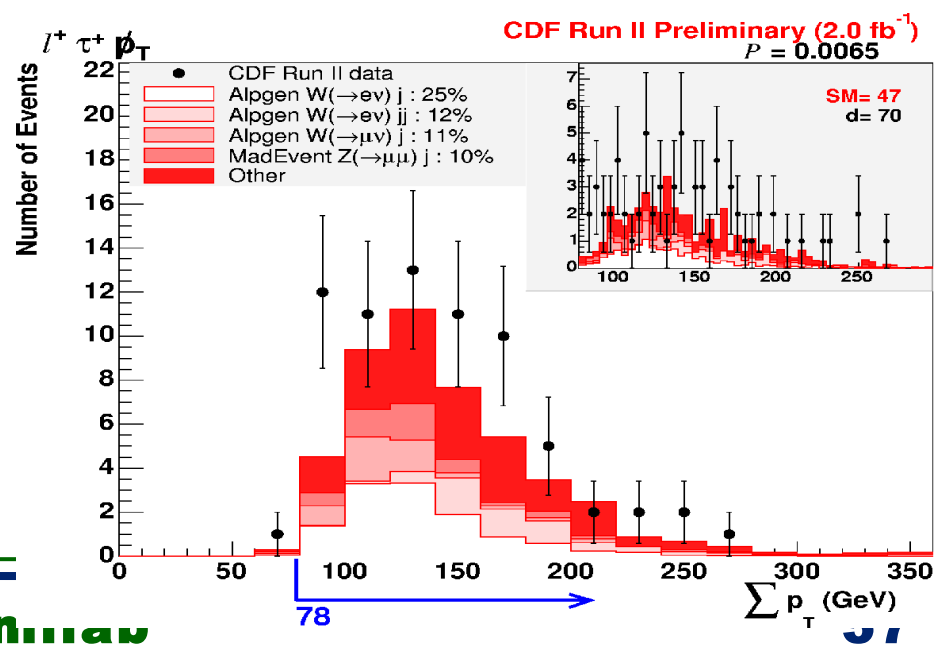
One Event:

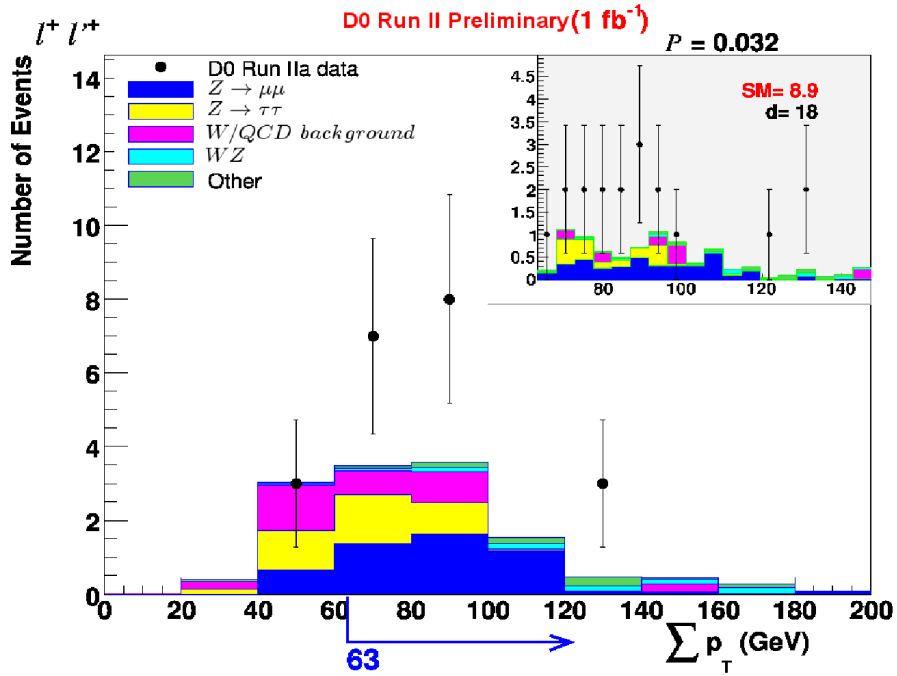
```
mutau12_ztt 1.113 0.000239137 ppbar 1960 409.057 13.9419
mu- 61.563 0.135986 139.865 tau+ 162.489 -1.20636 317.824 j 0
21.319 -1.64867 131.604 uncl 0 41.058 0 -37.7612 ;
```

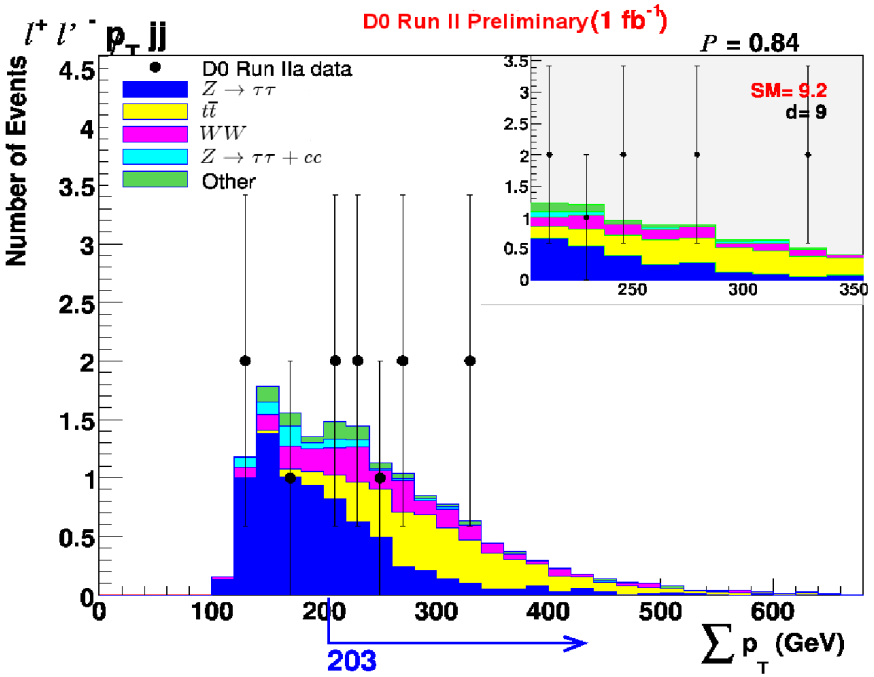

CDF Run II Preliminary (2.0 fb^{-1})
$l^+ l'^+ \not{p}_T$ jj
P = 0.0021
CDF Run II data
Alpgen W(→μν) jjj : 18%
MadEvent W(→lν)jγ : 13%
Herwig tt̄ : 13%
MadEvent Z(→μμ) jj : 11%
Other
SM= 0.15
d= 3
Number of Events
342
$\sum p_T$ (GeV)

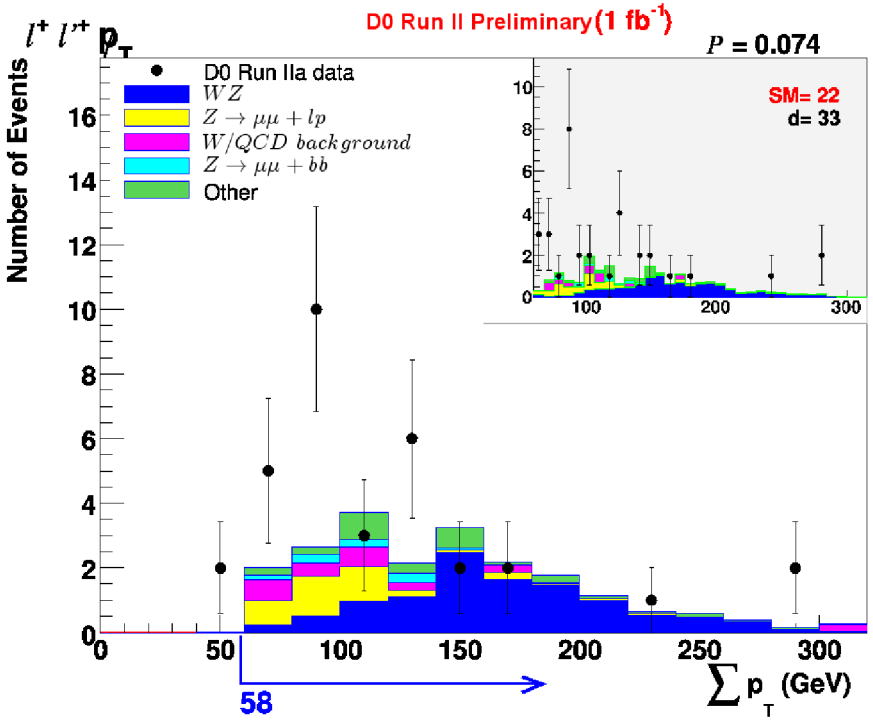

CDF Run II Preliminary (2.0 fb^{-1})
$l^+ l'^+ \not{p}_T$
P = 0.0042
CDF Run II data
Baur W(→μν)γ : 33%
Pythia Z(→μμ) : 17%
Alpgen W(→μν) j : 9.9%
MadEvent W(→lν)jγ : 7%
Other
SM= 5.7
d= 16
Number of Events
169
$\sum p_T$ (GeV)

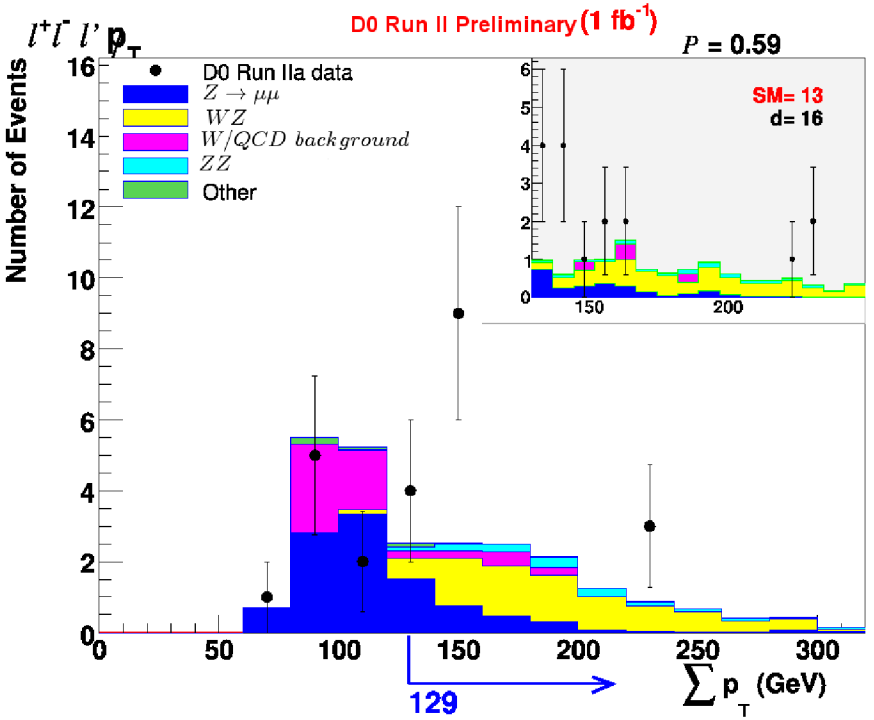

CDF Run II Preliminary (2.0 fb^{-1})
$l^+ l^- l' \not{p}_T$
P = 0.0047
CDF Run II data
MadEvent Z(→μμ) γ : 50%
Pythia WZ : 33%
MadEvent Z(→μμ) j : 4.6%
Pythia Z(→ττ) : 3.2%
Other
SM= 0.39
d= 4
Number of Events
245
$\sum p_T$ (GeV)


CDF Run II Preliminary (2.0 fb^{-1})
$l^+ \tau^+ \not{p}_T$
P = 0.0065
CDF Run II data
Alpgen W(→eν) j : 25%
Alpgen W(→eν) jj : 12%
Alpgen W(→μν) j : 11%
MadEvent Z(→μμ) j : 10%
Other
SM= 47
d= 70
Number of Events
78
$\sum p_T$ (GeV)

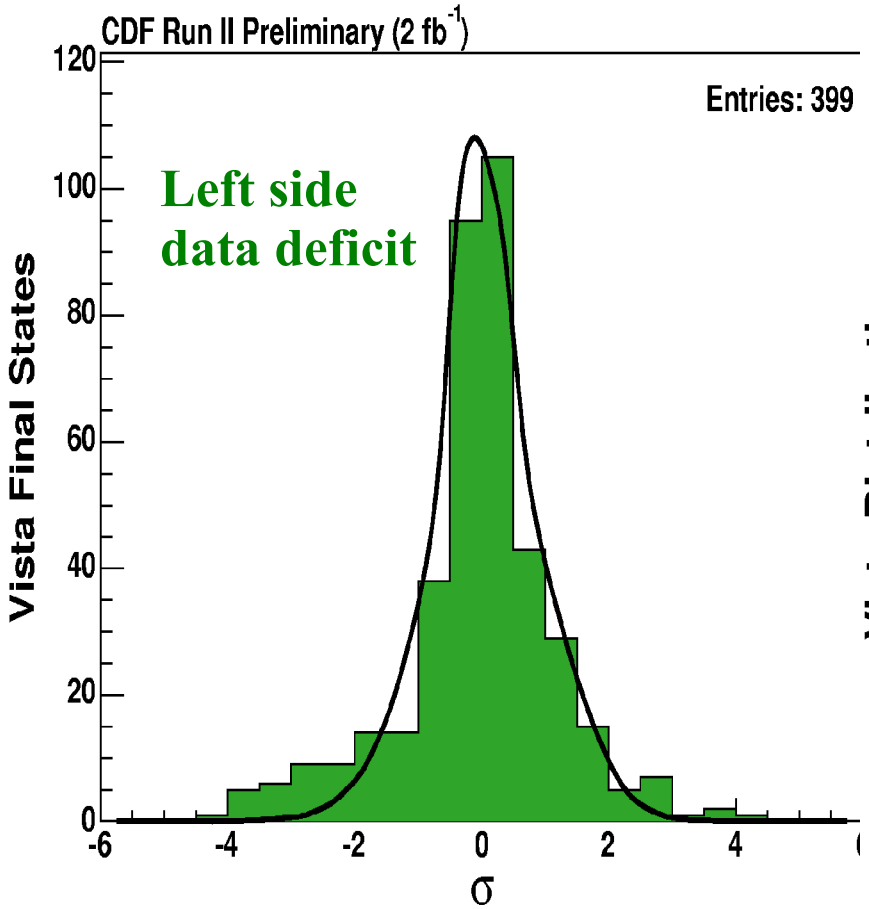

$l^+ l'^+$
D0 Run II Preliminary(1 fb^{-1})
P = 0.032
D0 Run IIa data
$Z \to \mu\mu$
$Z \to \tau\tau$
W/QCD background
WZ
Other
SM= 8.9
d= 18
Number of Events
63
$\sum p_T$ (GeV)


$l^+ l'^- p_T jj$
D0 Run II Preliminary(1 fb^{-1})
P = 0.84
D0 Run IIa data
$Z \to \tau\tau$
$t\bar{t}$
WW
$Z \to \tau\tau + cc$
Other
SM= 9.2
d= 9
Number of Events
203
$\sum p_T$ (GeV)

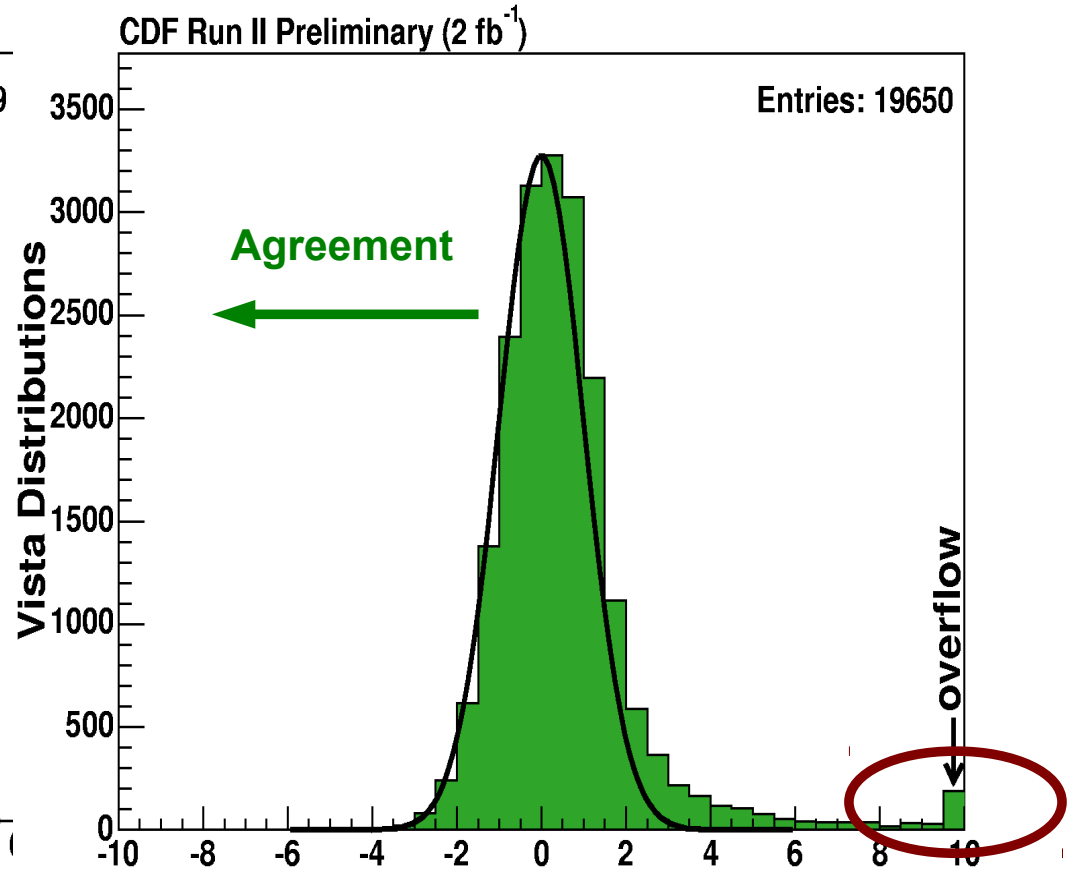

$l^+ l'^+ p_T$
D0 Run II Preliminary(1 fb^{-1})
P = 0.074
D0 Run IIa data
WZ
$Z \to \mu\mu + lp$
W/QCD background
$Z \to \mu\mu + bb$
Other
SM= 22
d= 33
Number of Events
58
$\sum p_T$ (GeV)


$l^+ l^- l' p_T$
D0 Run II Preliminary(1 fb^{-1})
P = 0.59
D0 Run IIa data
$Z \to \mu\mu$
WZ
W/QCD background
ZZ
Other
SM= 13
d= 16
Number of Events
129
$\sum p_T$ (GeV)

This compares to CDF plots



Curve: pseudoexperiments



555 discrepancies mostly in multijets dominated states

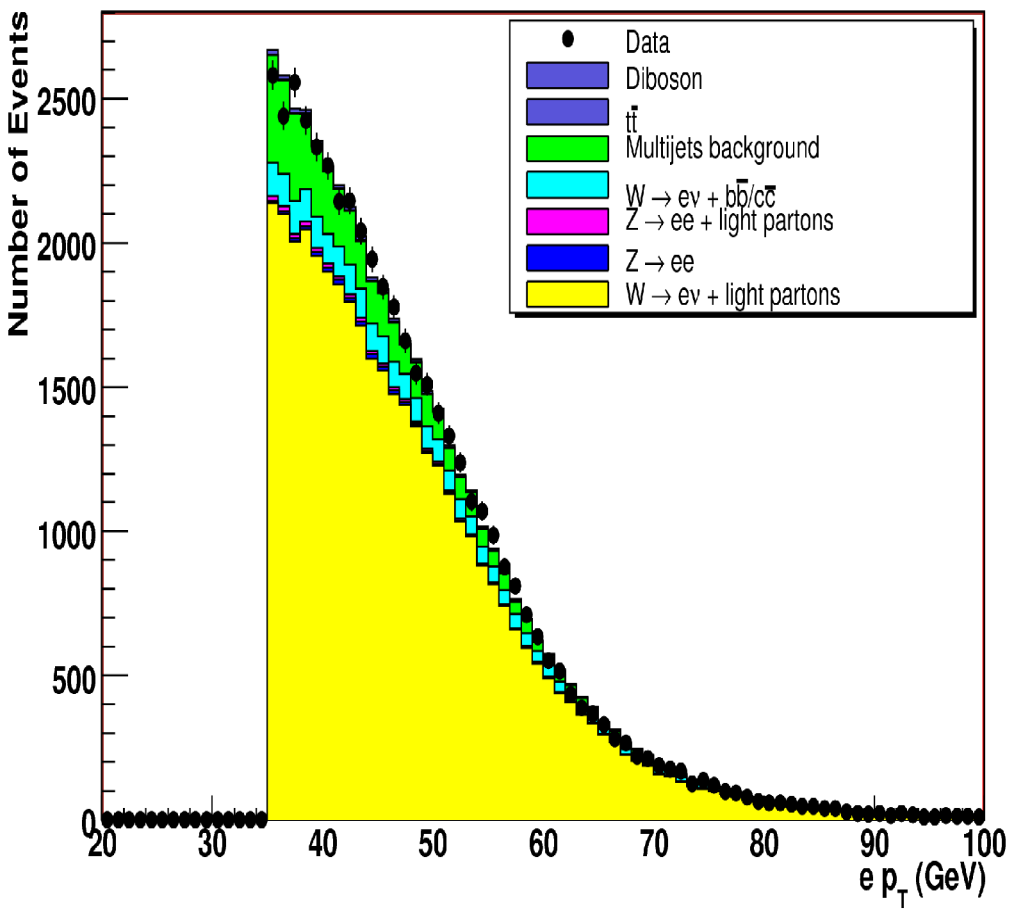
Single e + jets requires tighter cuts to avoid being dominated by multijet background

Same fitting parameters as μ + jets

Event Cuts

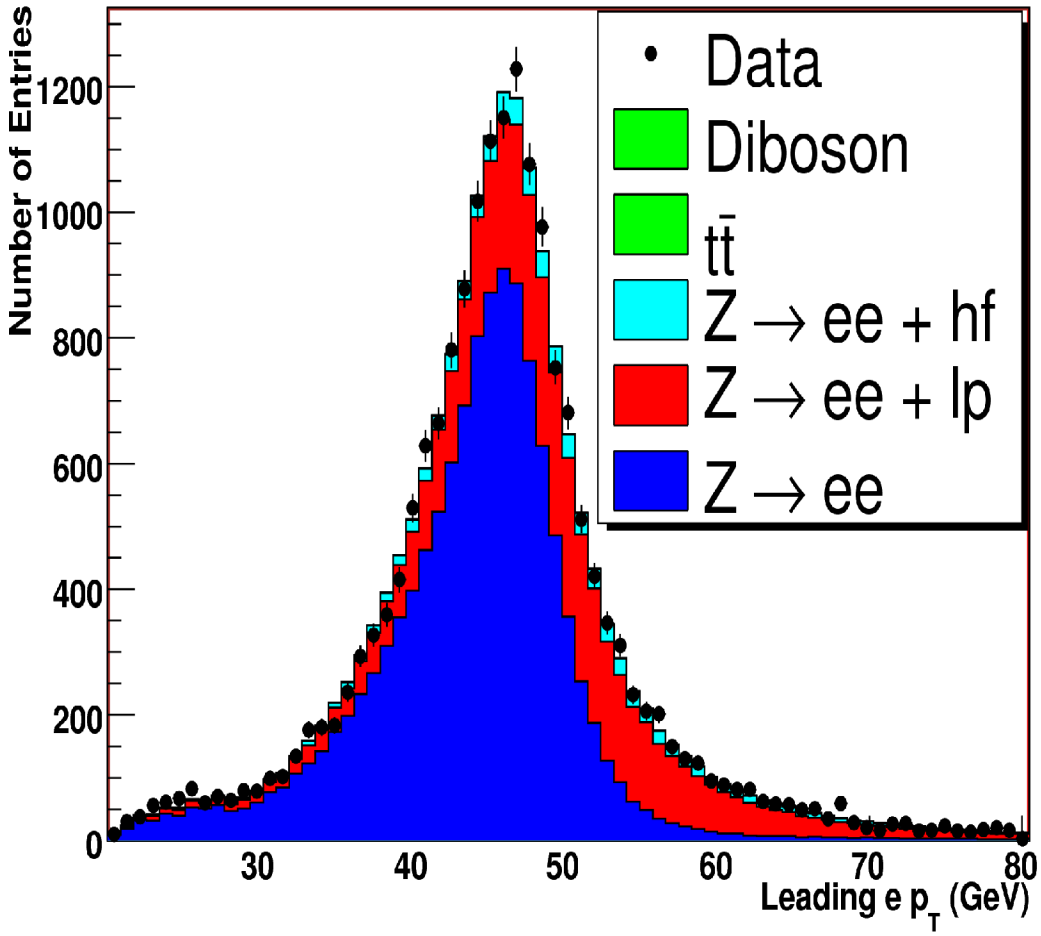
- e with at least one jet
  - e p_T > 35 GeV
  - e η < 1.1
  - jet p_T > 20 GeV
- Likelihood cut of 0.95
  - Multijets between 0.3 and 0.85

Work in progress



Dielectrons loosen p_T and likelihood wrt e + jets in relative absence of multijets

Work in progress



Fit Parameters

- Drell-Yan
  - w/no jets
  - w/at least one light parton
  - w/at least one heavy parton

Event Cuts

- e likelihood > 0.85
- e p_T > 20 GeV
- e η > 1.1
- Other cuts same as e + jets

e + τ state uses same τ parameters with loose electron

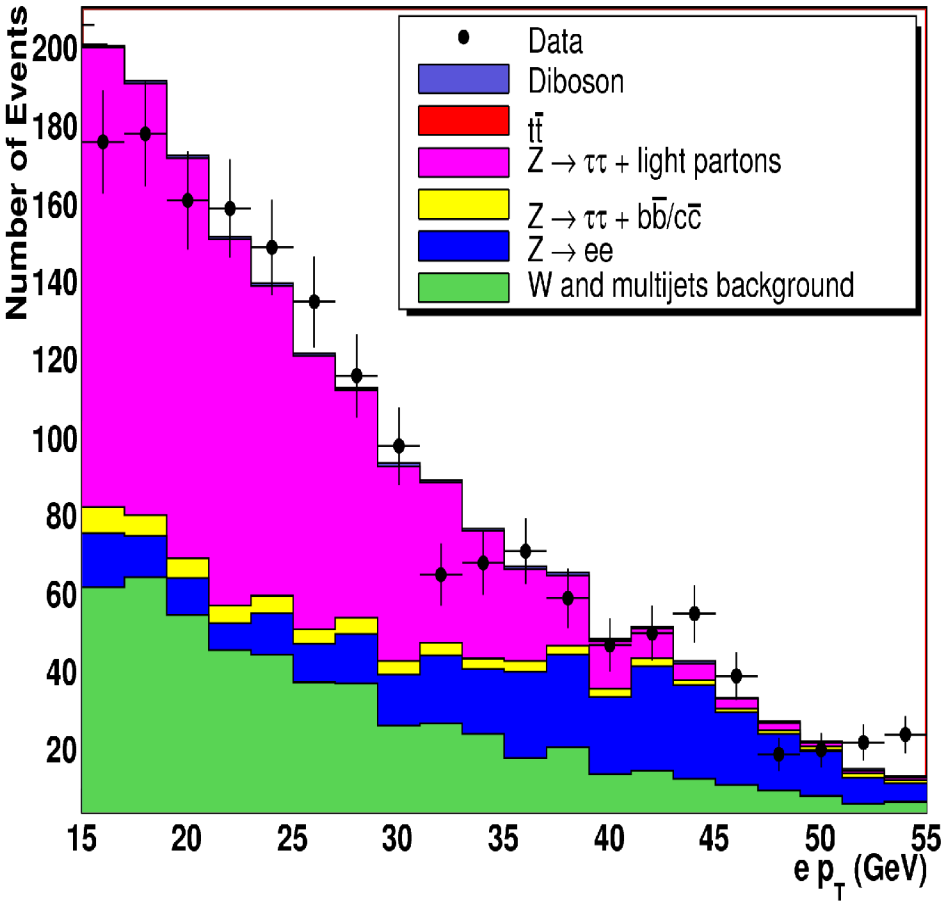
Parameters

- Drell-Yan ee
  - Allowed to float to incorporate mis-ID sf
- Drell-Yan ττ
- Low statistics final states heavy-flavor factor fixed

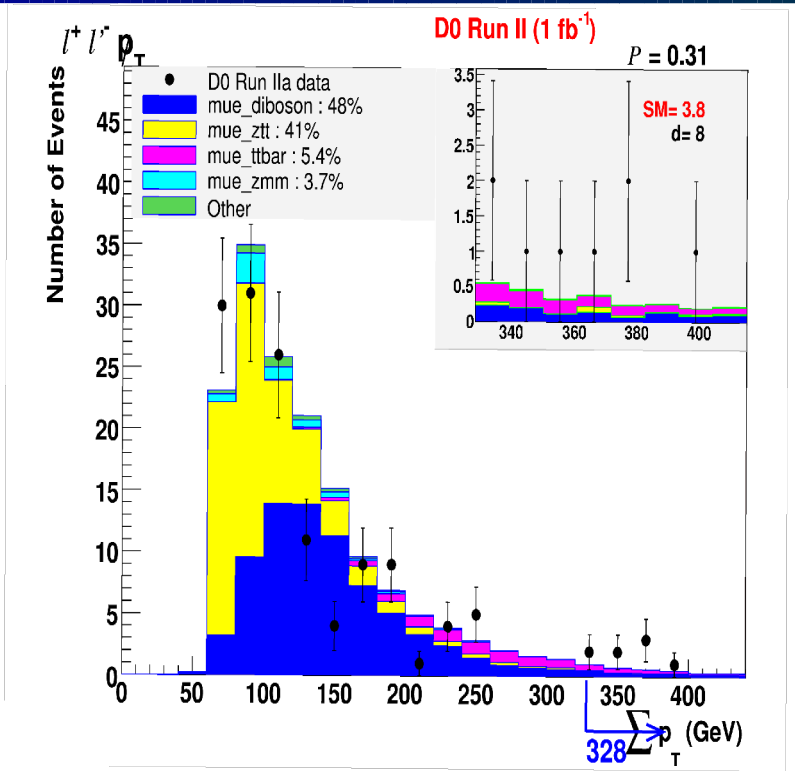
Event cuts

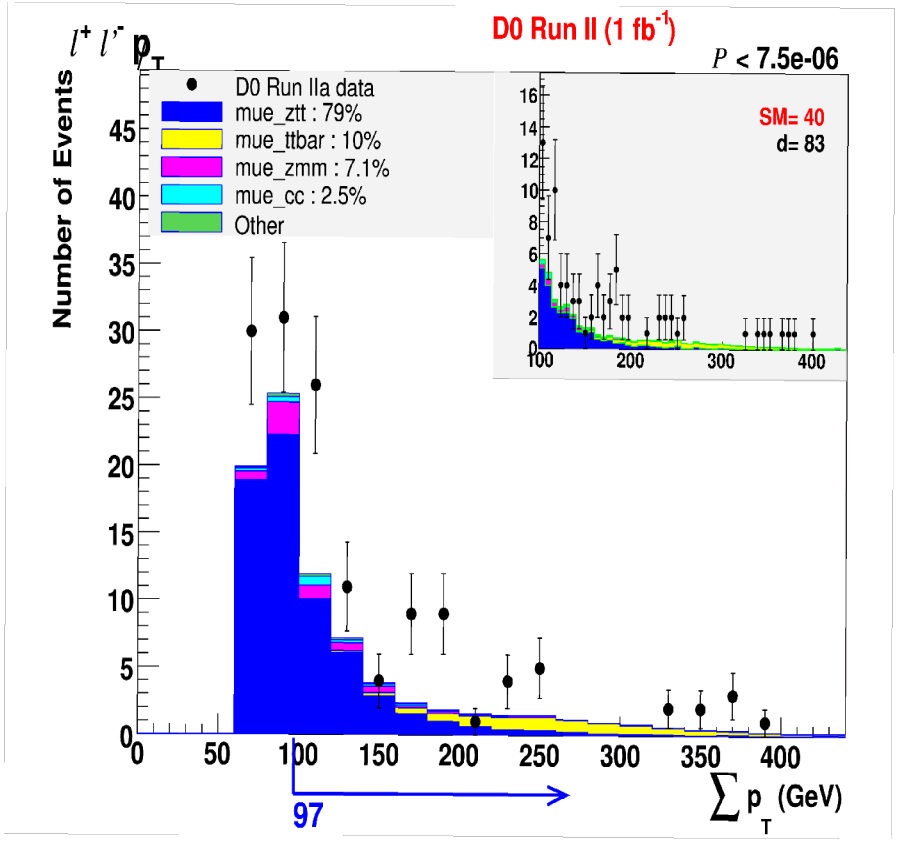
- e p_T > 15 GeV
- e η < 1.1
- τ p_T > 15 GeV
- Again, only single-prong τs

Work in progress



Diboson


D0 Run II (1 fb^{-1})
$l^+ l'^- p_T$
P = 0.31
D0 Run IIa data
mue_diboson : 48%
mue_ztt : 41%
mue_ttbar : 5.4%
mue_zmm : 3.7%
Other
SM= 3.8
d= 8
Number of Events
$\sum p_T$ (GeV)
328


D0 Run II (1 fb^{-1})
$l^+ l'^- p_T$
P < 7.5e-06
D0 Run IIa data
mue_ztt : 79%
mue_ttbar : 10%
mue_zmm : 7.1%
mue_cc : 2.5%
Other
SM= 40
d= 83
Number of Events
$\sum p_T$ (GeV)
97