Model Independent Search in Leptonic Final States at DØ

Joel Piper
Michigan State University

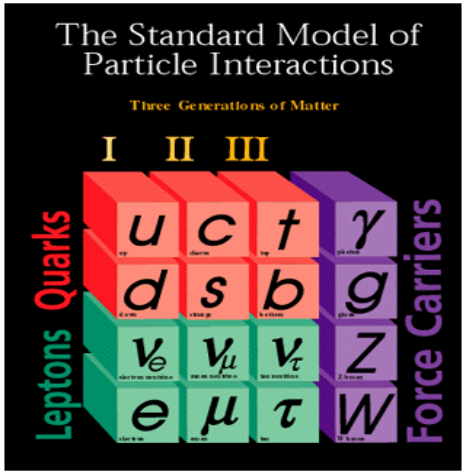


Brookhaven Particle Physics Seminar
October 29, 2009



The Standard Model is incomplete. With shrinking resources at Fermilab, we must address a central question:

Do we see what we expect?



Outline

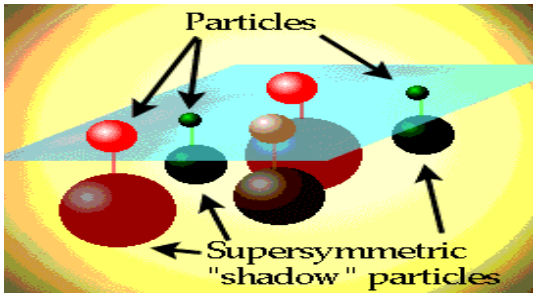
I. Introduction

II. Strategy

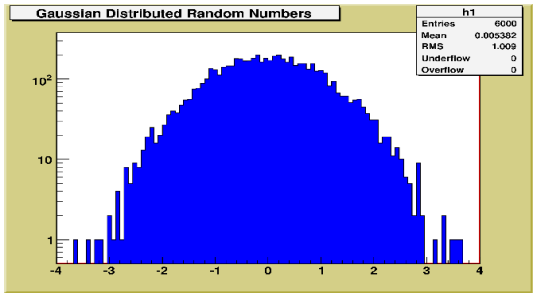
III. Results

II. Introduction

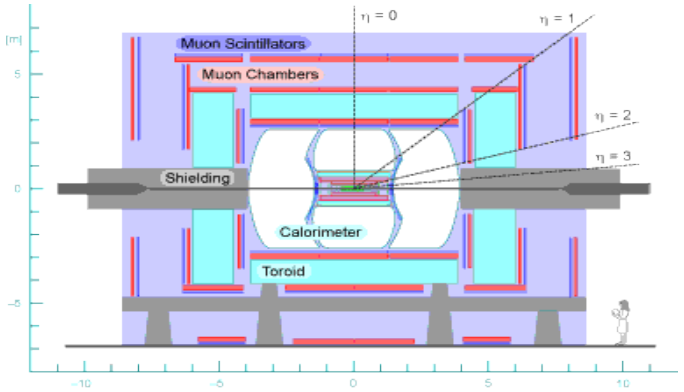
A. Motivation



B. What is a Model-Independent Search?



C. The D0 Detector



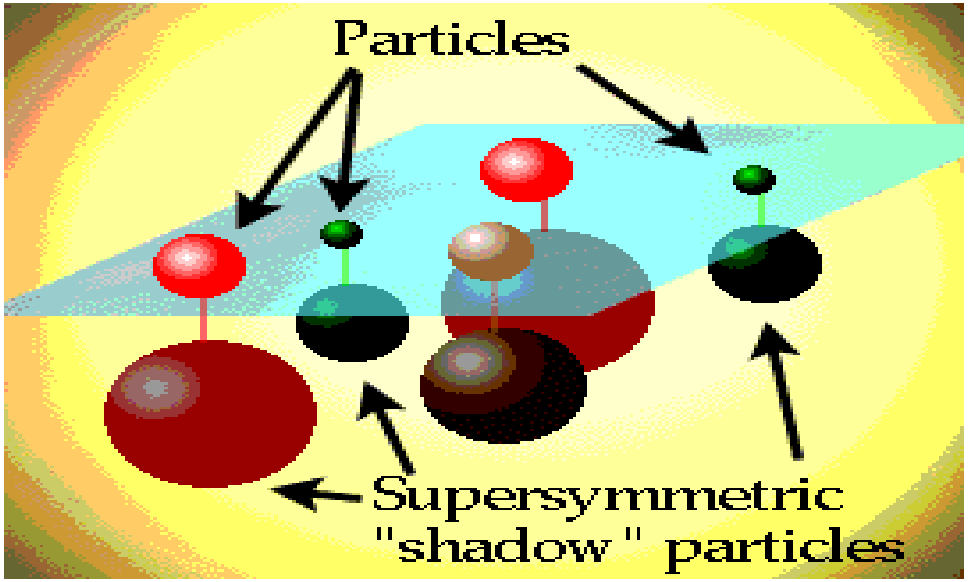
Particle physics is moving into a new era

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... but what?

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

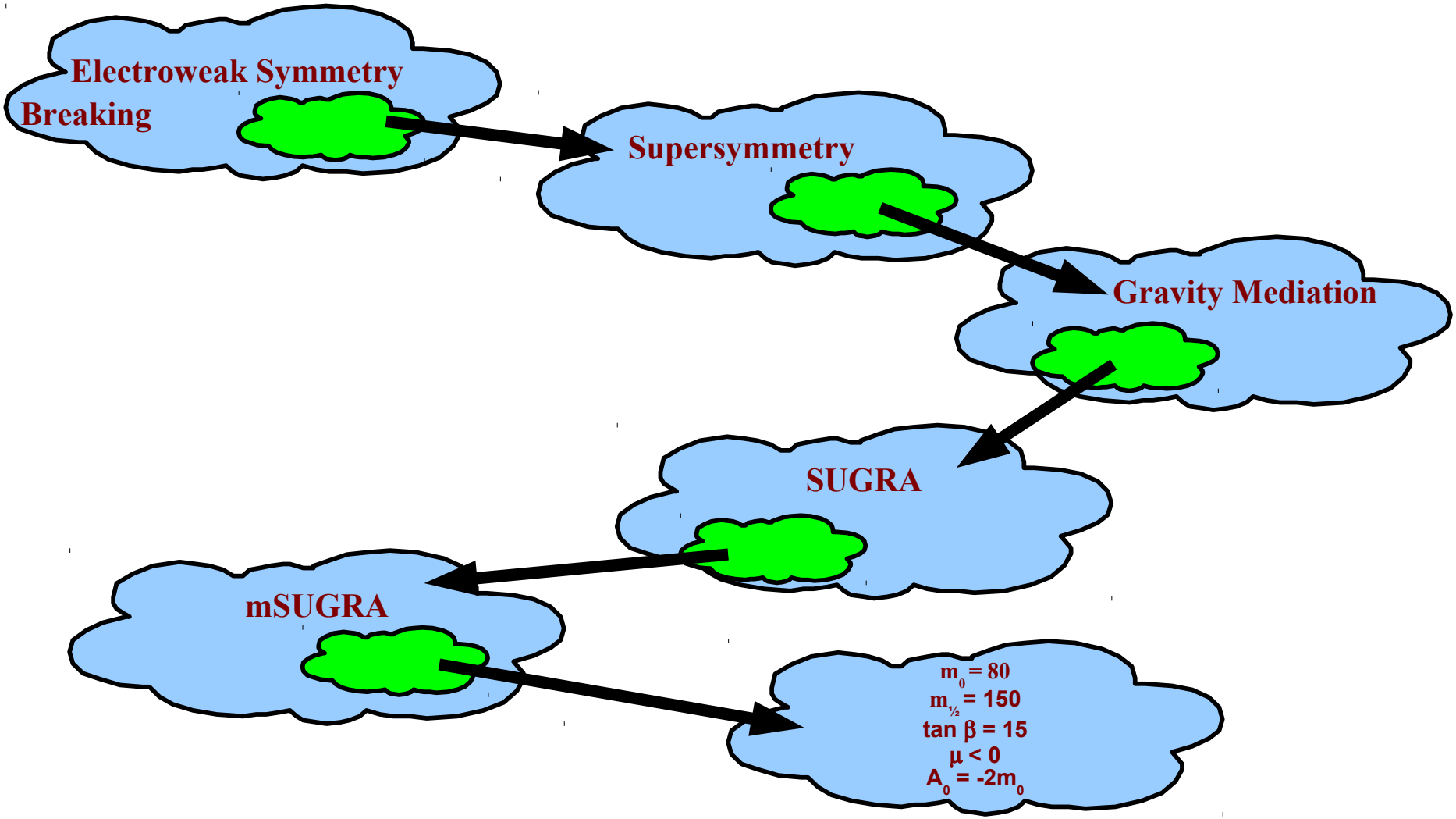


We must choose where to invest our resources

Most interesting?

Reasonable acceptance?

A broad search complements many narrower ones



A global search is an involved process with complicated results

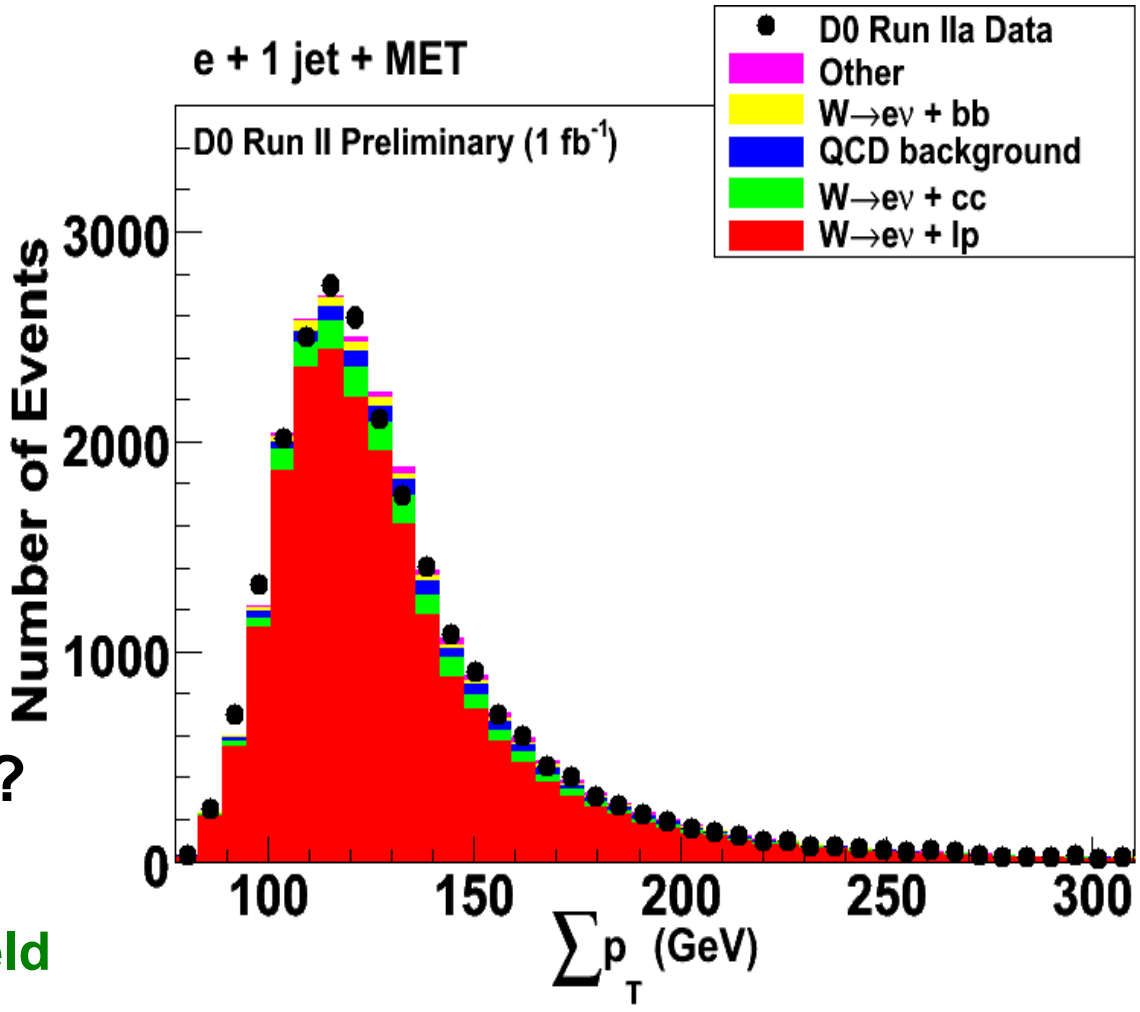
Impossible to check all data for any discrepancy, so...

Where do we look?

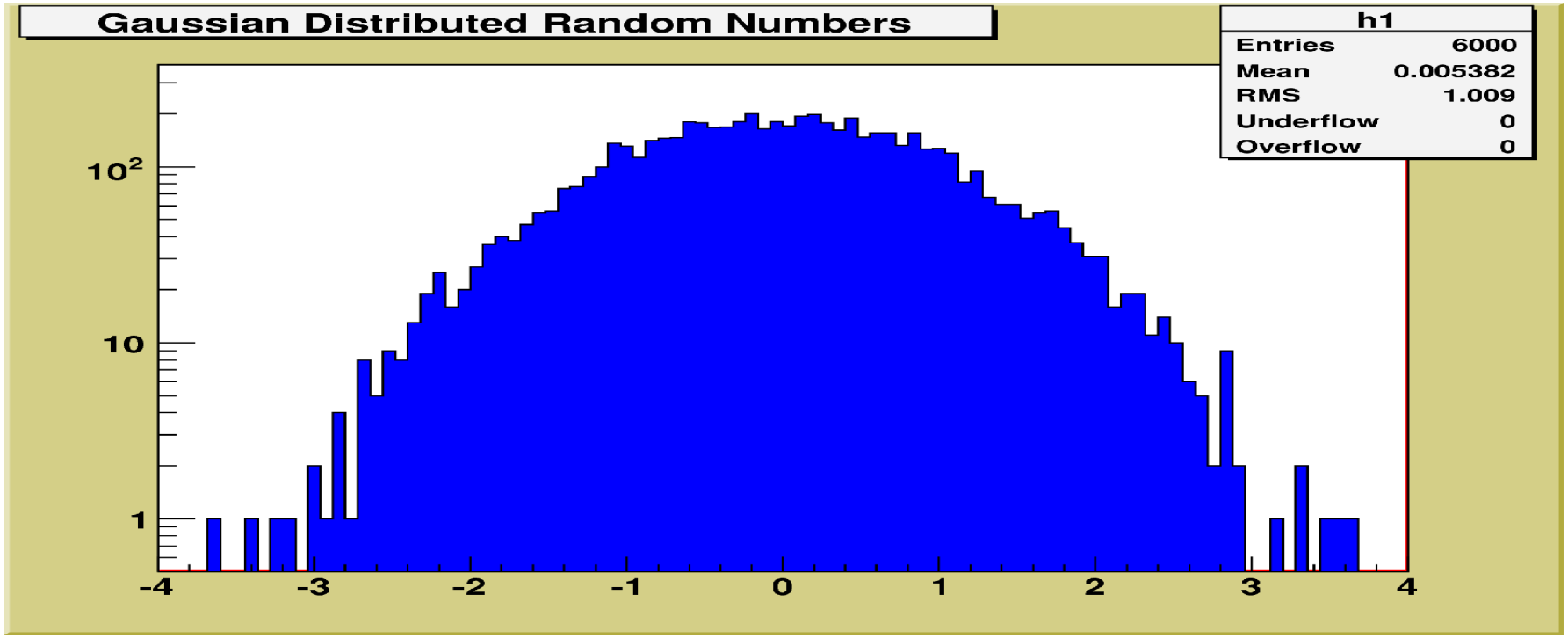
- Event counts in final states
- Shape agreement over many distributions
- Event Σp_T distributions

Why these variables?

- Large breadth
- Many EW symmetry breaking theories yield high-p_T final states



A Model Independent Search is a search for a deviation



What can cause a deviation?

- Statistical fluctuation
- Keep careful count of how many places checked
  - Trials Factor

After applying trials factor, still need to isolate sources of remaining disagreement

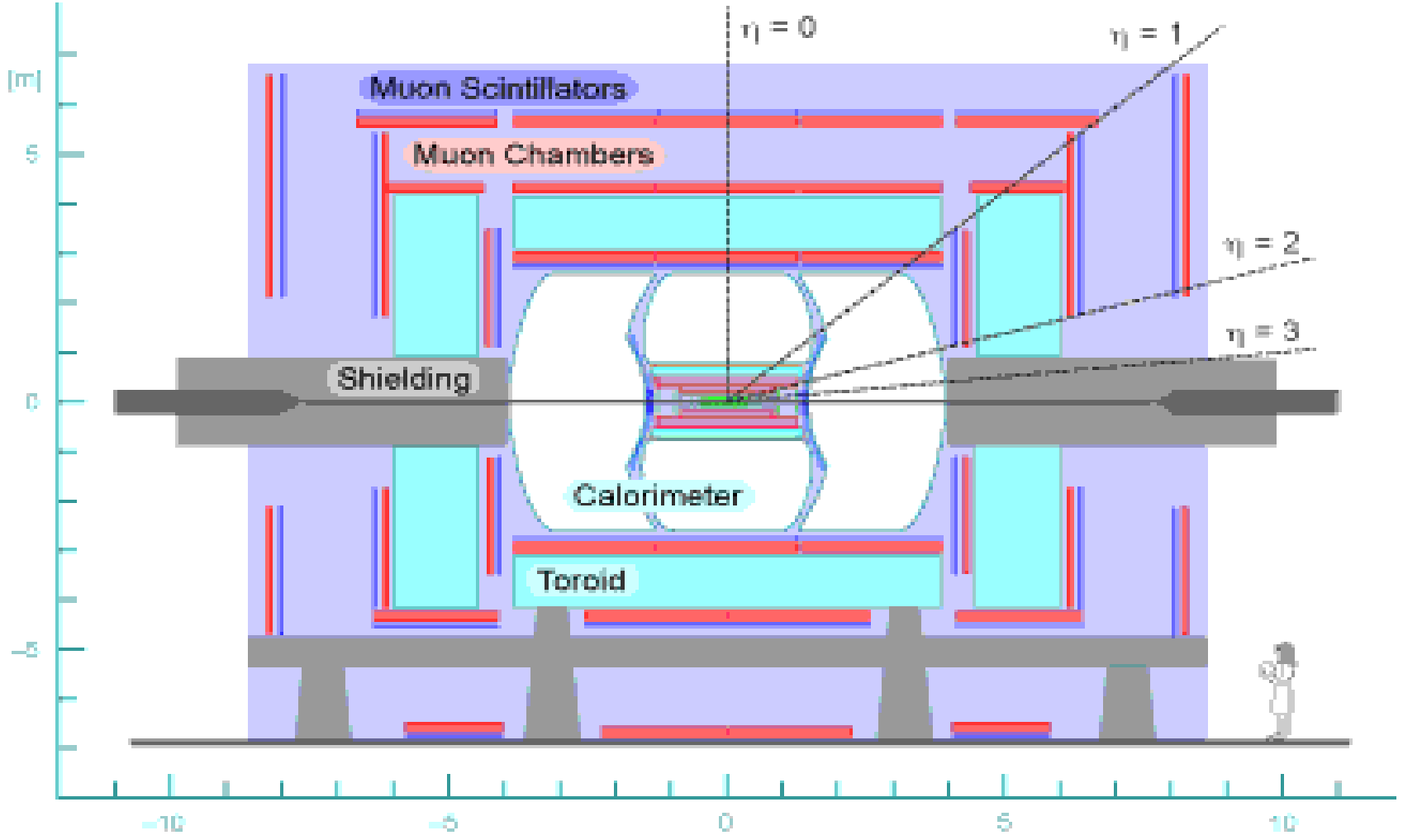
What if we see something we do not expect?

- Is MC simulation accurate here?
- Could detector modeling or event reconstruction cause this?
- Could this be new physics?

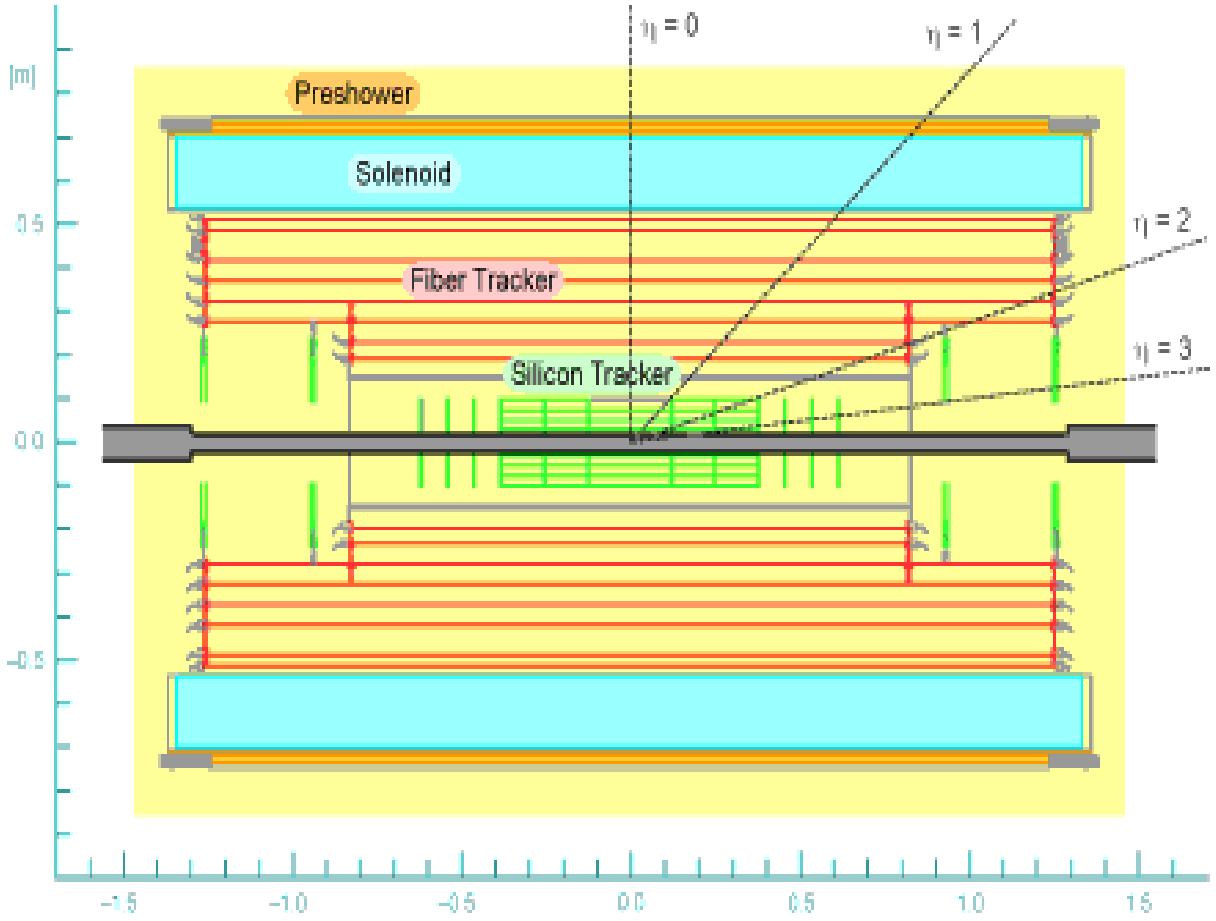
To minimize MC and detector modeling issues:

- Use common MC
  - PYTHIA and ALPGEN
  - Multijets background from data
- Apply common collaboration-wide scale factors
  - Can be bin-by-bin or several parameter functions
  - Use phase space dominated by SM processes
- Fit for normalization factors

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

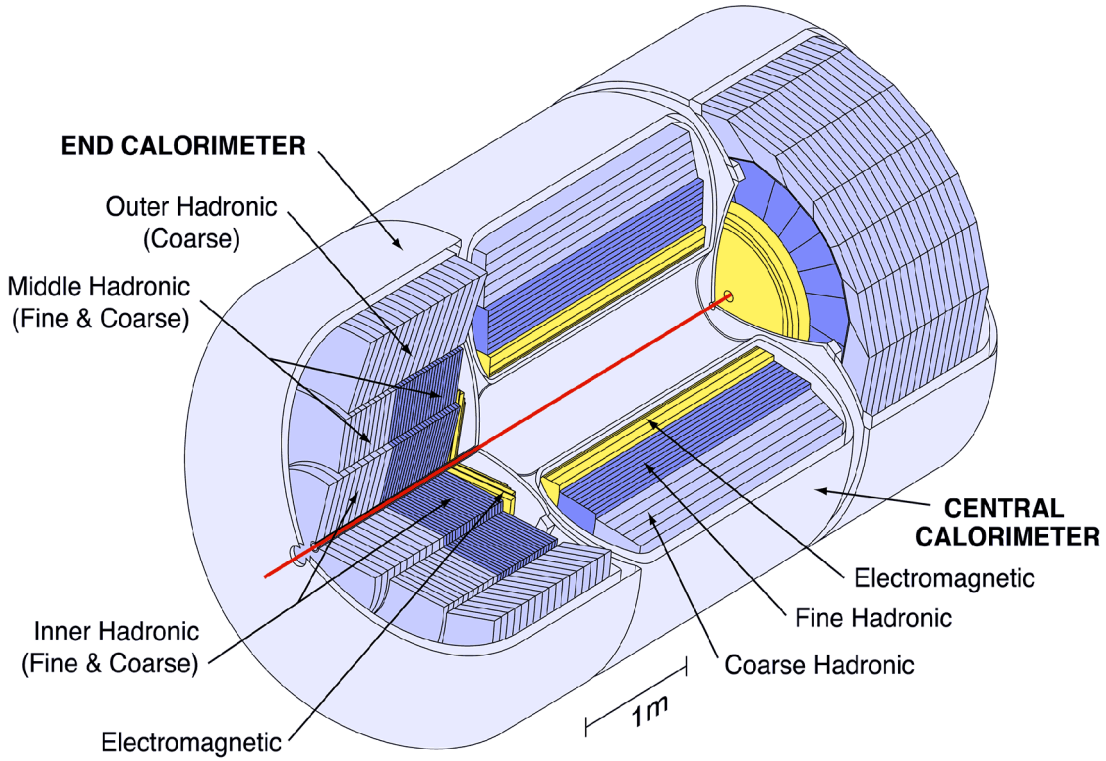


Tracking



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

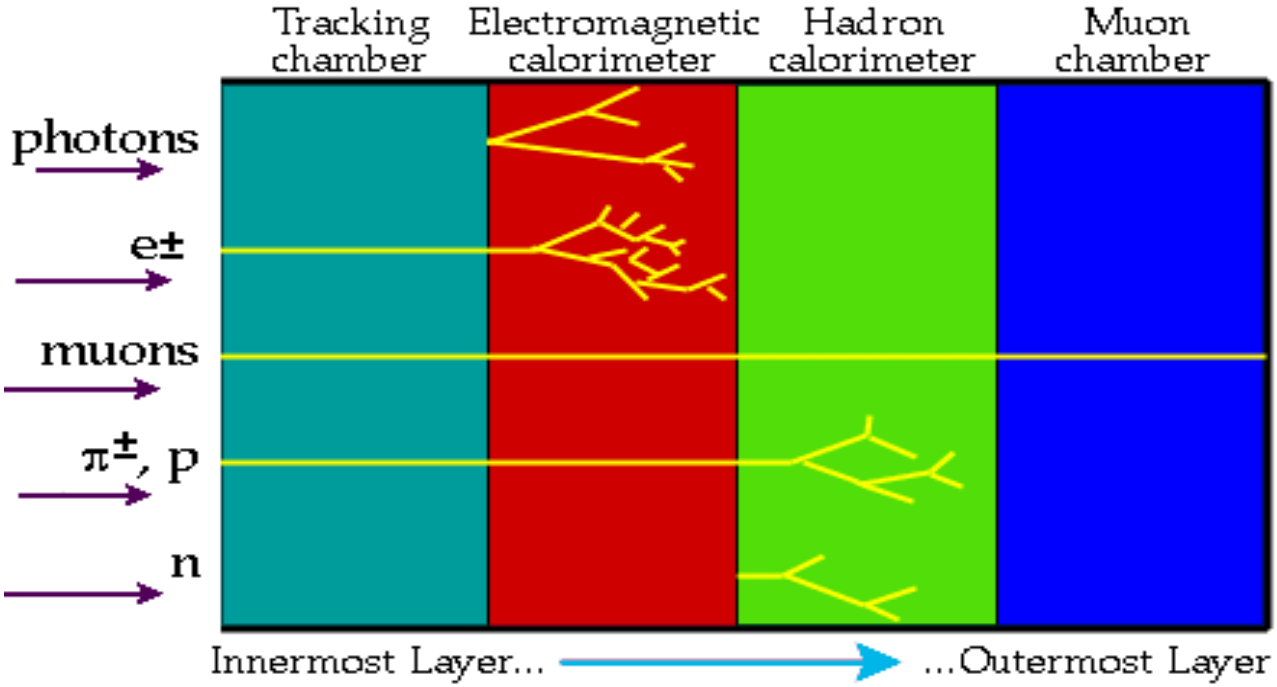
Calorimeters

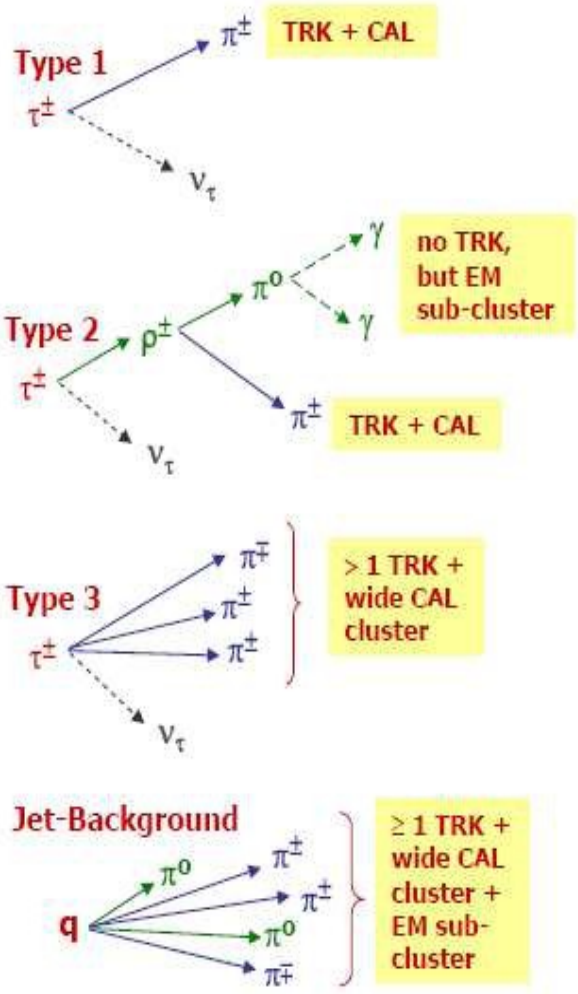


Central and end calorimeters

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

Object Identification





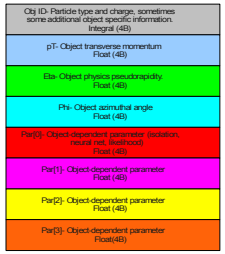
Object ID through detector signature

A model-independent search must ensure no double counting

Also b-jets from displaced vertex

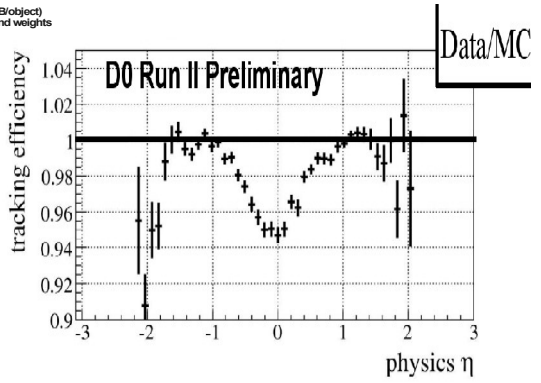
II. D0 Model Independent Search Strategy

A. D0 High-p_T Dataset

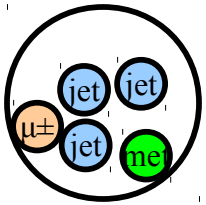


Total Object Size: 32 B
Also save:
Weight information (~12B/object)
Global event information and weights

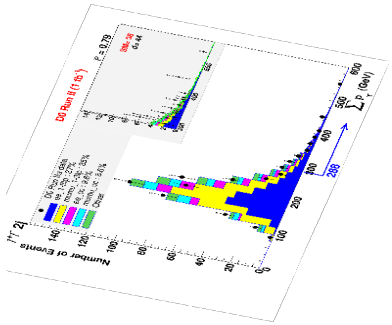
B. D0 MIS Analysis Packages & Fits



C. Vista Exclusive Model Independent Comparison



D. Sleuth "Quasi" Model Independent Comparison at High-p_T Tails



New reduced D0 High-P_T Dataset for MIS

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

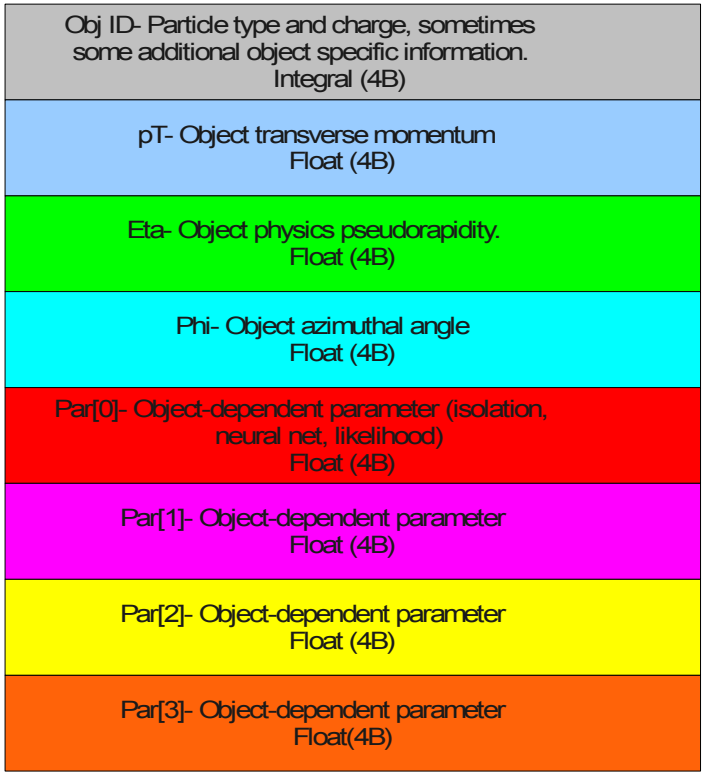
MIS search considers many final states, so many processes contribute

- 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 quantites stored in high-p_T events

- Full rerun of using normal data format → 1 week
- Full Run IIa dataset is 19.3 TB
- Using high-p_T format, rerun takes < 2 hours
- Storage done locally using only 7.1 GB
- Testing a change in analysis strategy takes ~1 day to implement



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

Monte Carlo and Data Used

Monte Carlo with real zero-bias overlay from many processes

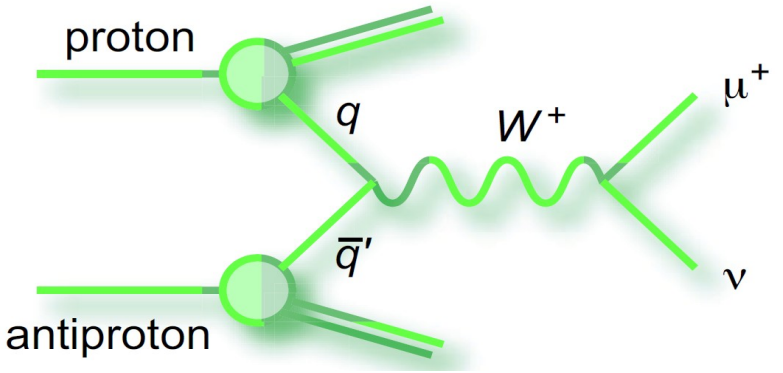
Alpgen/Pythia

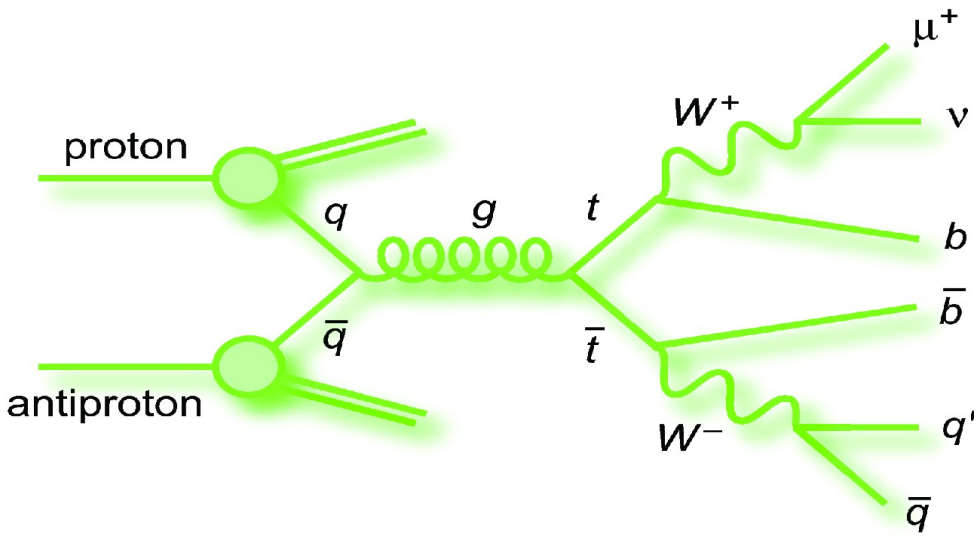
- **W boson decay**
- **Drell-Yan processes**
- **ttbar**

Pythia

- **diboson**

Processes resulting in multijet states extracted from data



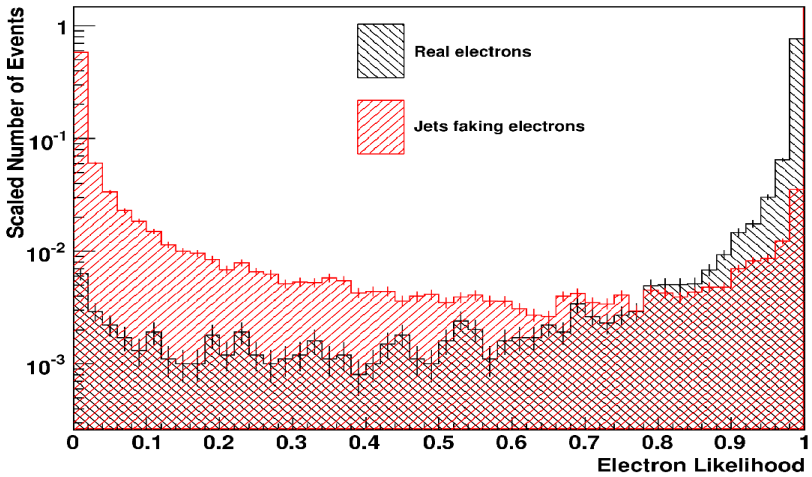


Electron multijet background uses loosened likelihood cuts

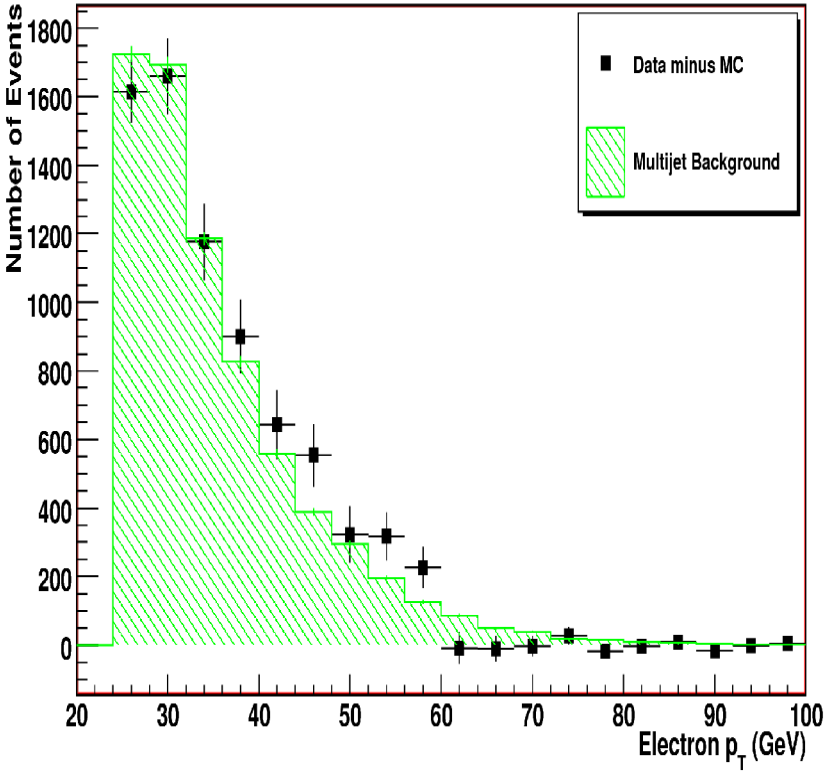
Jet faking mechanism

- 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

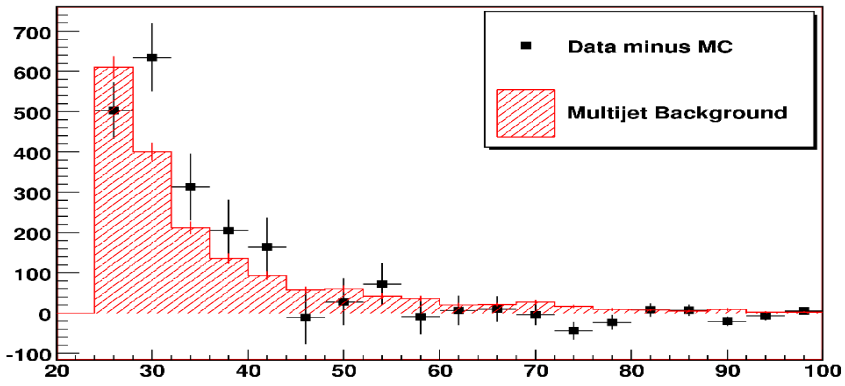
Muon multijet background uses isolation variables

Jet faking mechanism

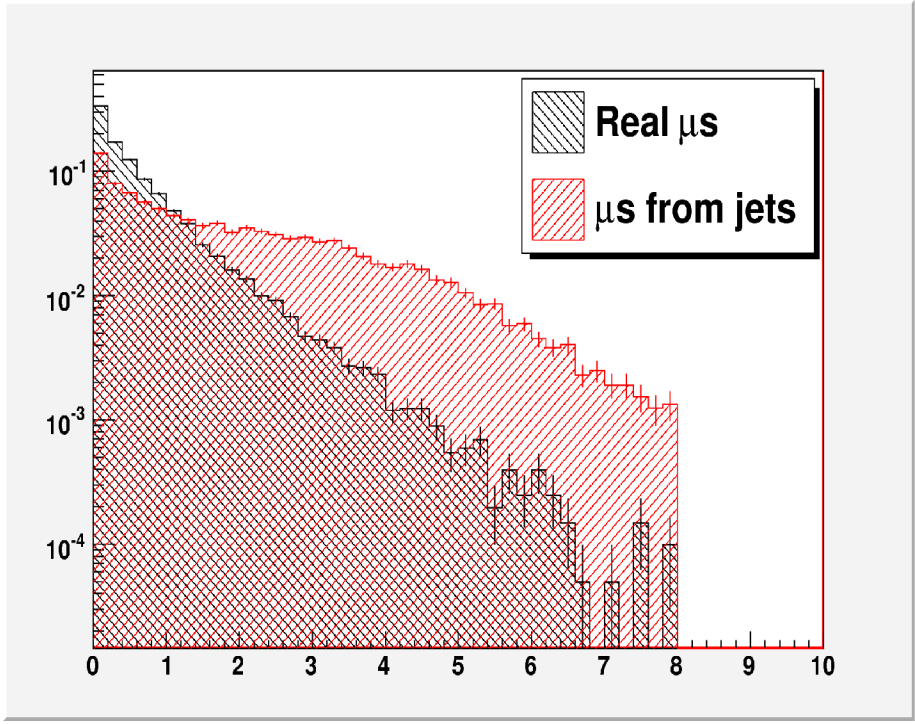
- Heavy jets decay to real muons
- Very little punch-through at D0

Is muon associated with a jet?

- Hollow cones in calorimeter, tracker



Work in progress



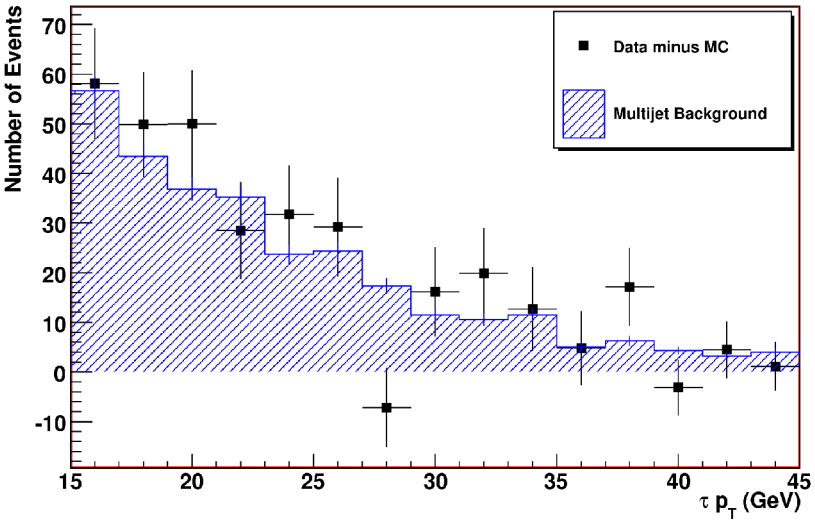
Multijets use

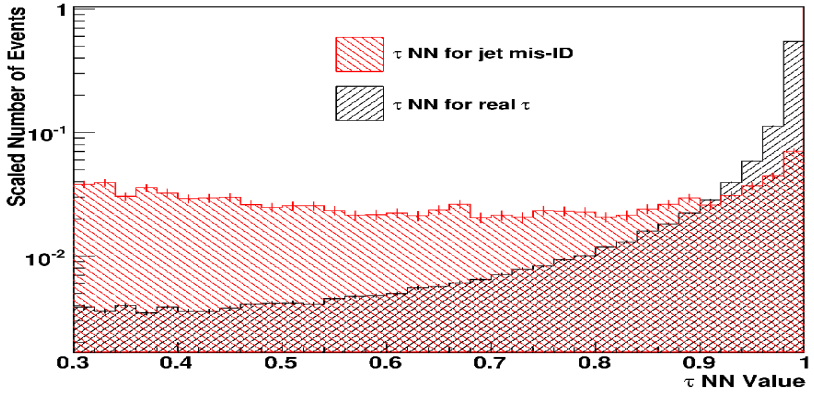
- 4.0 < Cal Iso < 8.0 GeV
- 4.0 < Trk Iso < 8.0 GeV

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

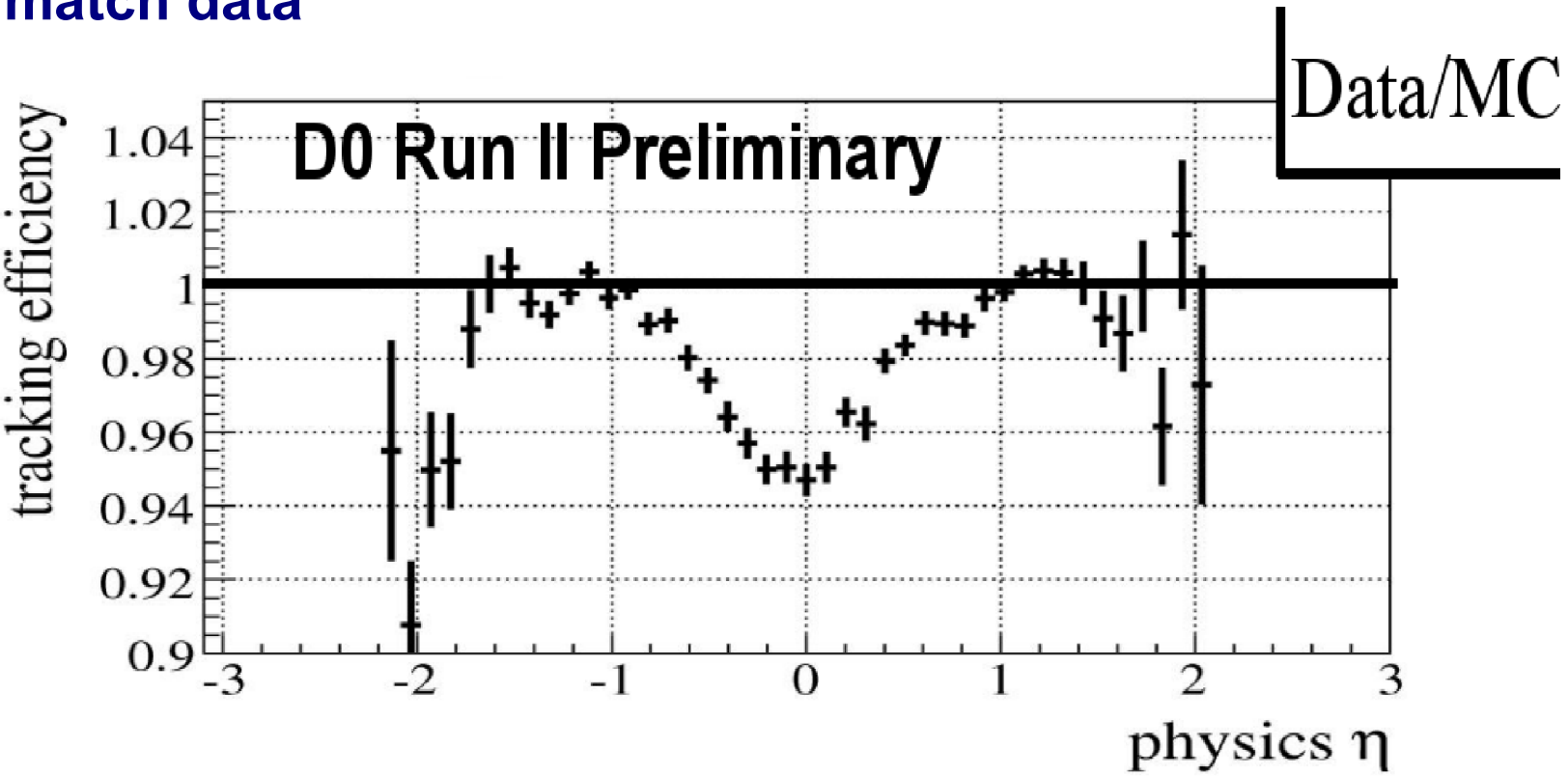




Need to fix problems with expectation without biasing results

Incorporate known D0 data/MC correction factors (weights)

Make sure these are used to solve known problem and not just to match data



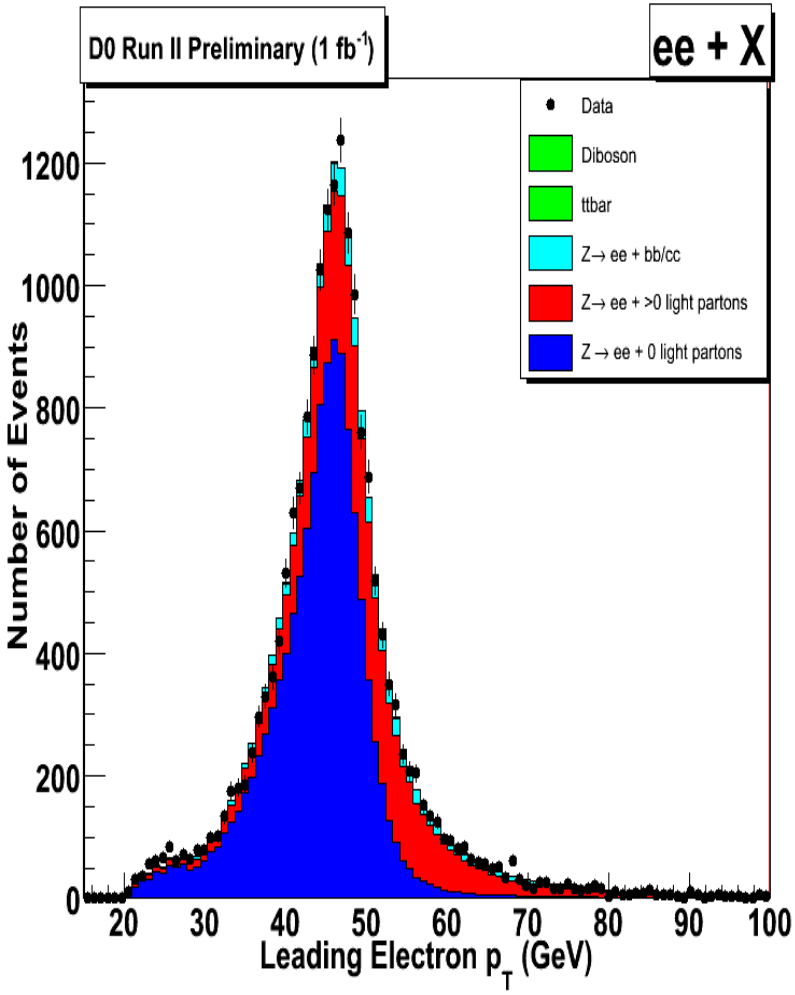
Use high-p_T files to feed into MIS analysis inclusive state fits

- States dominated by single process used to constrain normalization
- Rare final states excluded to test strategy

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

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

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



D0: Inclusive final states and object selection

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

(GeV)	Ma
5	2
5	1
5	2
5	1

D0 Run II Preliminary

μ's have additional 300 GeV maximum p_T cut

Final fitting must incorporate constraints

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

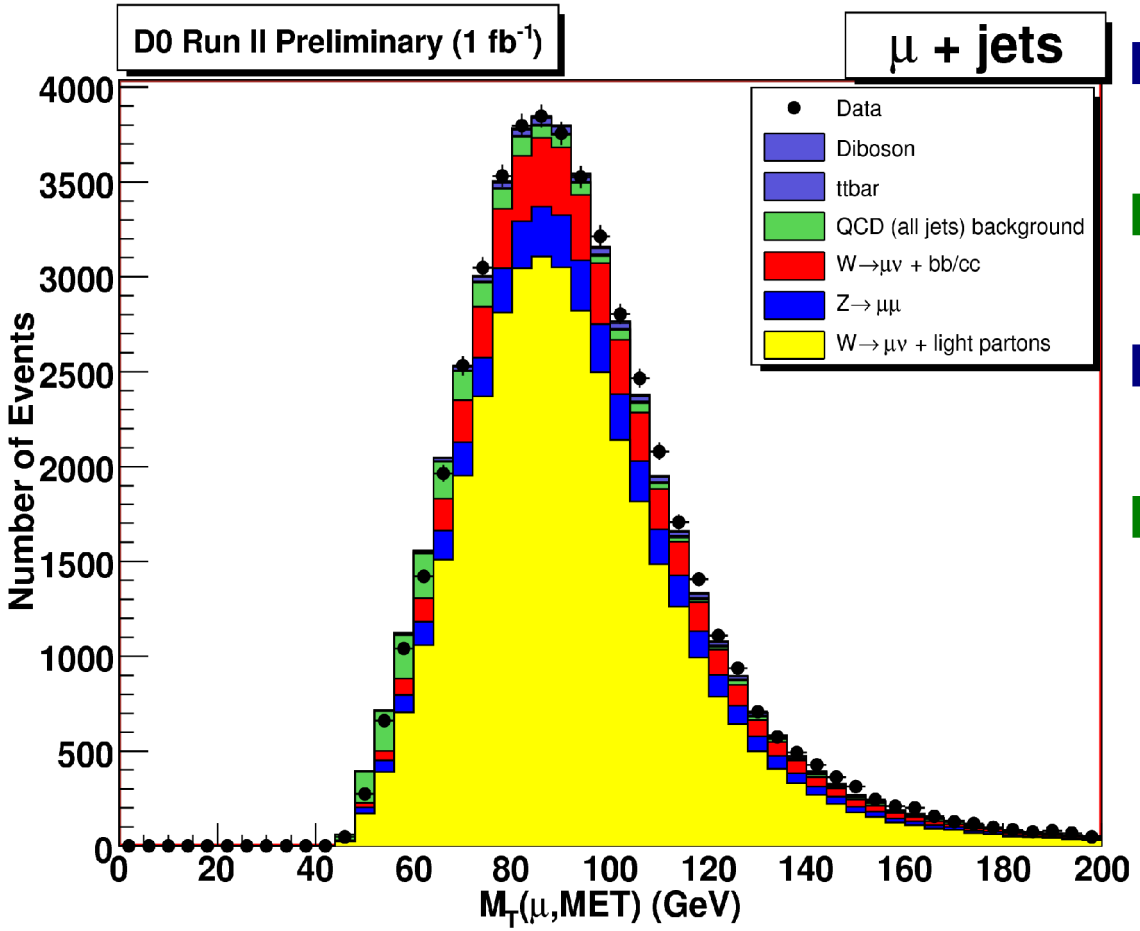
Process jet multiplicity scale factors also flavor independent

Other analyses should see similar results

Jet multiplicity flavor independence:

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

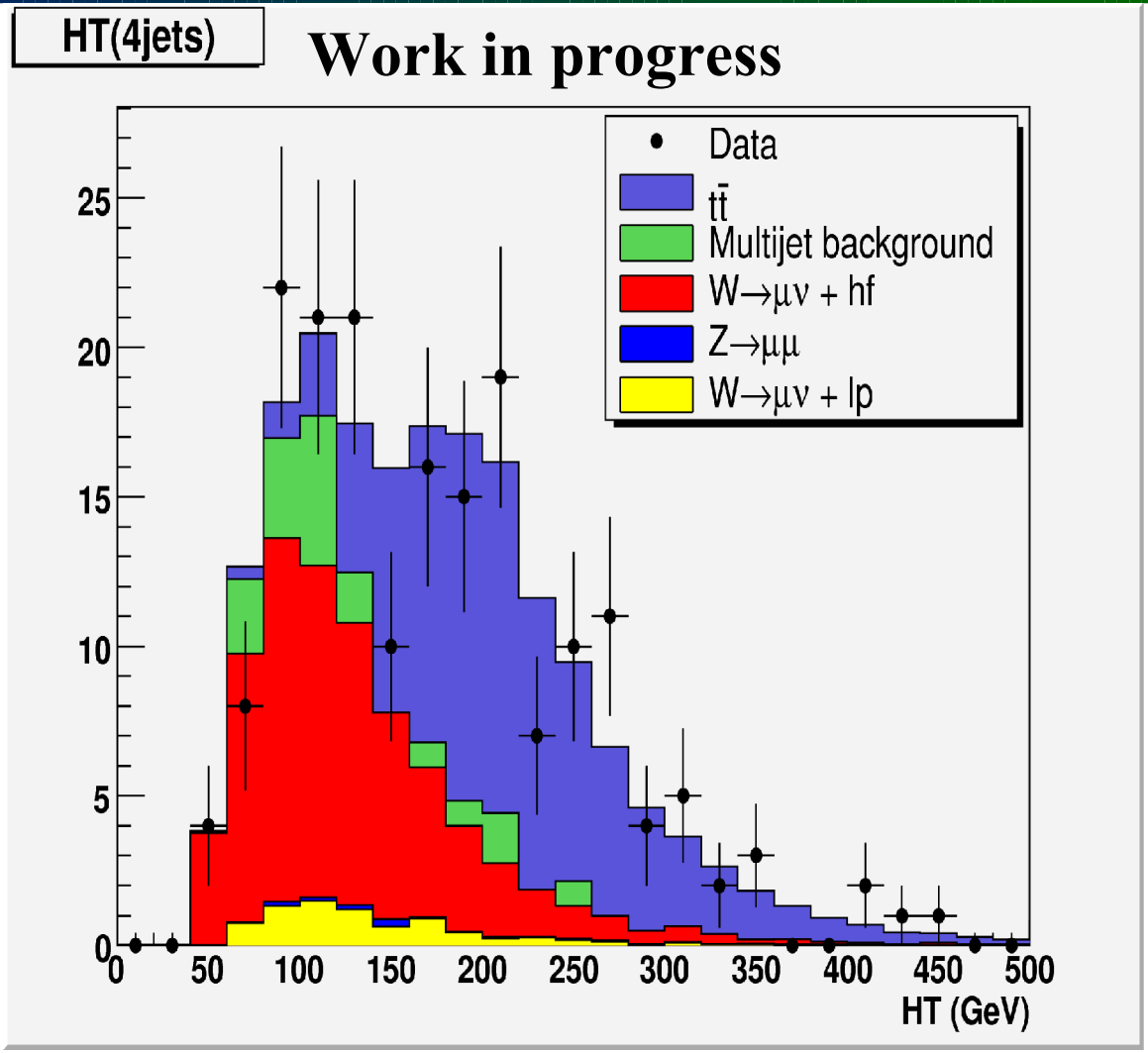
Perform checks and pass on to experiment-independent programs



- Pick another set of histograms to check
- Looking for qualitative agreement
- KS test for each plot for quantitative check
- Look for large problems at this stage
  - 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 Monte Carlo is shown in the dark blue
- Necessary for reasonable agreement



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 ;
```

Model independent searches at CDF use a different approach

They cover as many final states as possible

Live with discrepancies if poor modeling

Use QCD Monte Carlo

they can use exact cuts as data, use single numbers for mis-ID

All data manipulation is done at Vista level

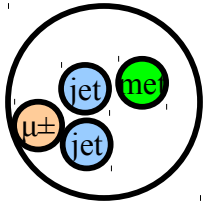
CDF Run II Preliminary ($2.0\ fb^{-1}$)

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

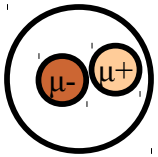


Check numbers of events

- Significance diluted by number of final states

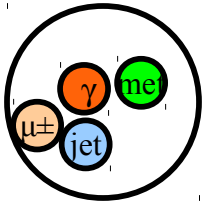
Compare shapes

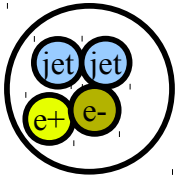
- Using Kolmogorov-Smirnov probabilities
- Significance diluted by number of histograms
- Sample variables: (p_T,M(all combinations),M_T(w/MET),ΔR,$\Delta\phi$,η)

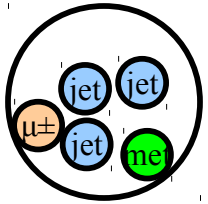


Report significant discrepancies

- > 3σ after trials factor

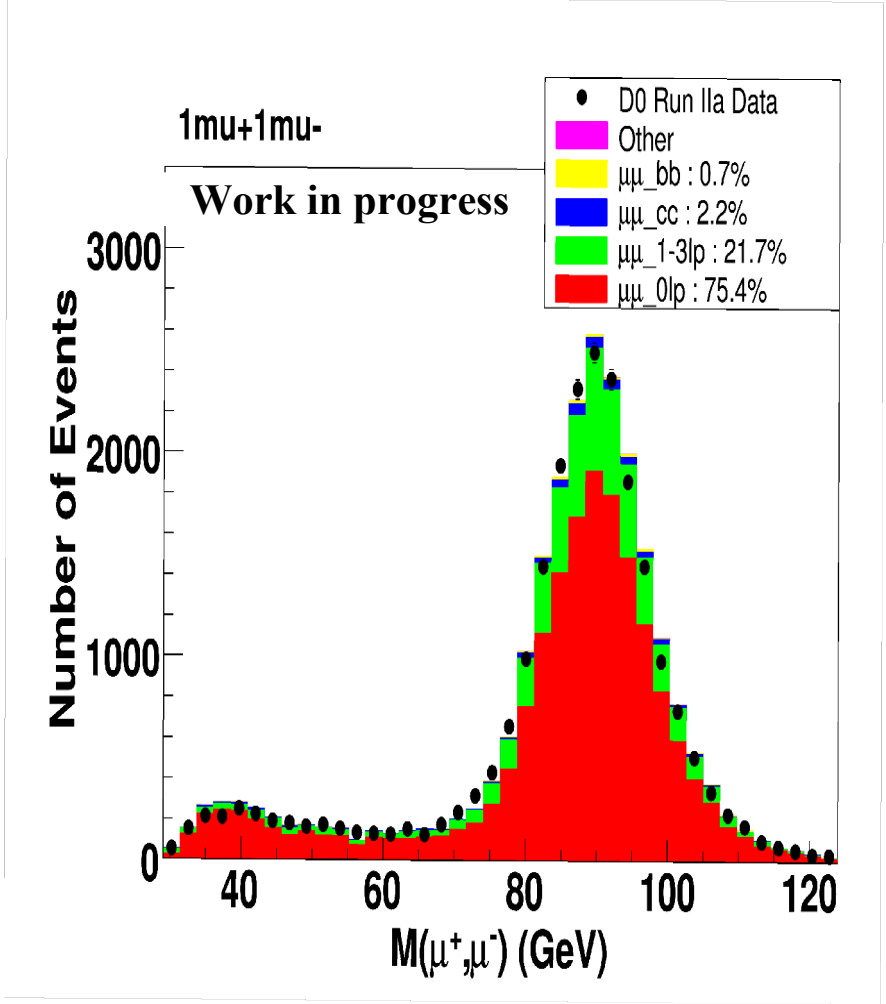




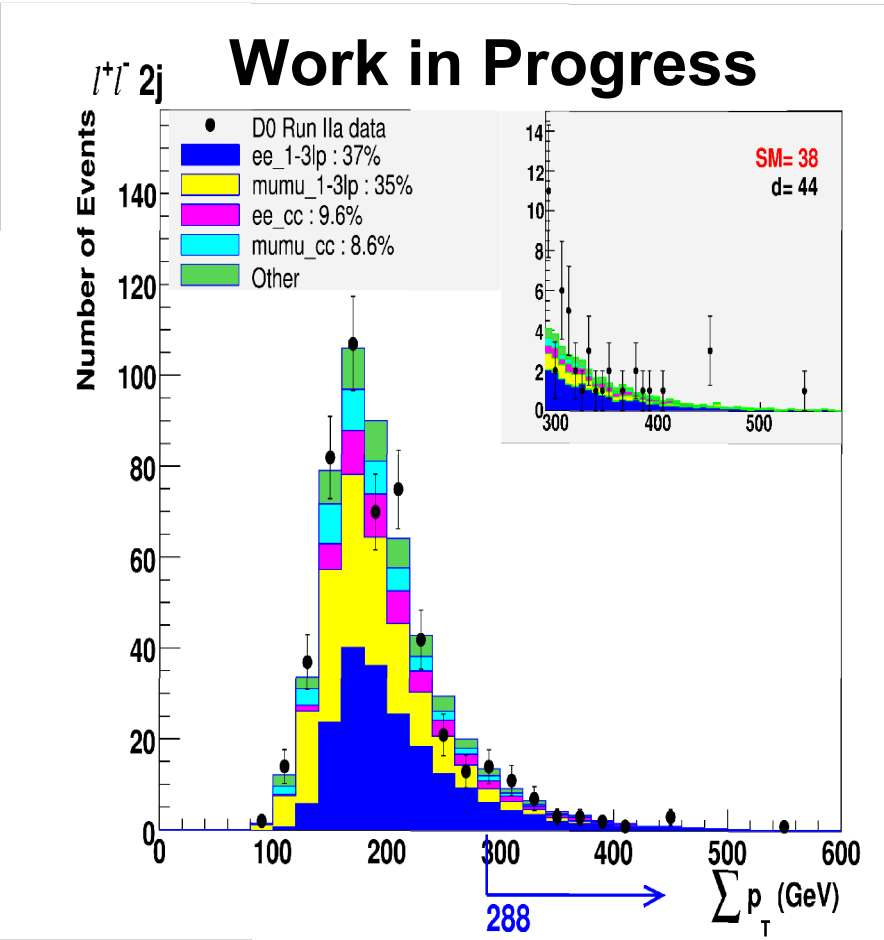


Vista checks > 9000 distributions from 180 exclusive final states

- 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



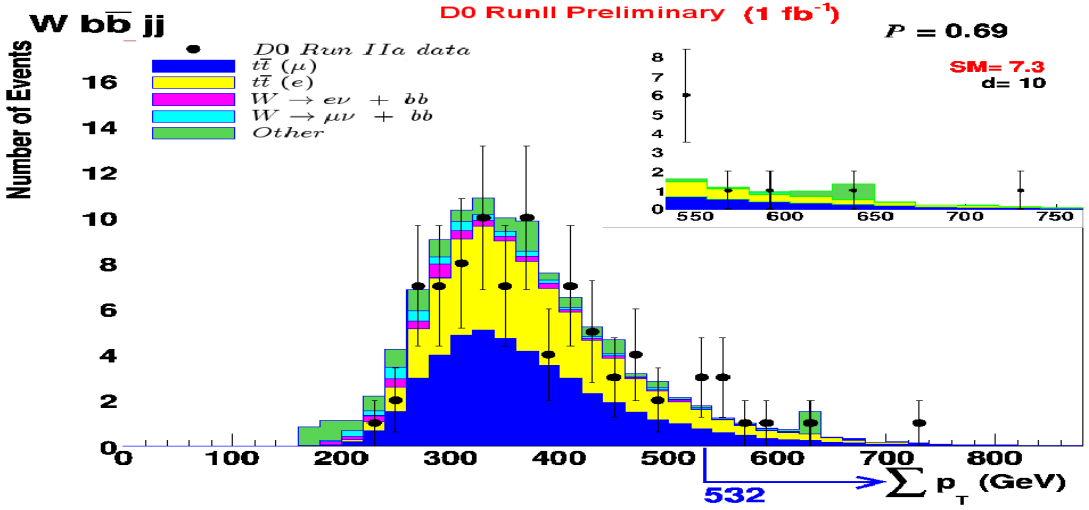
Sleuth performs a quasi-model independent search

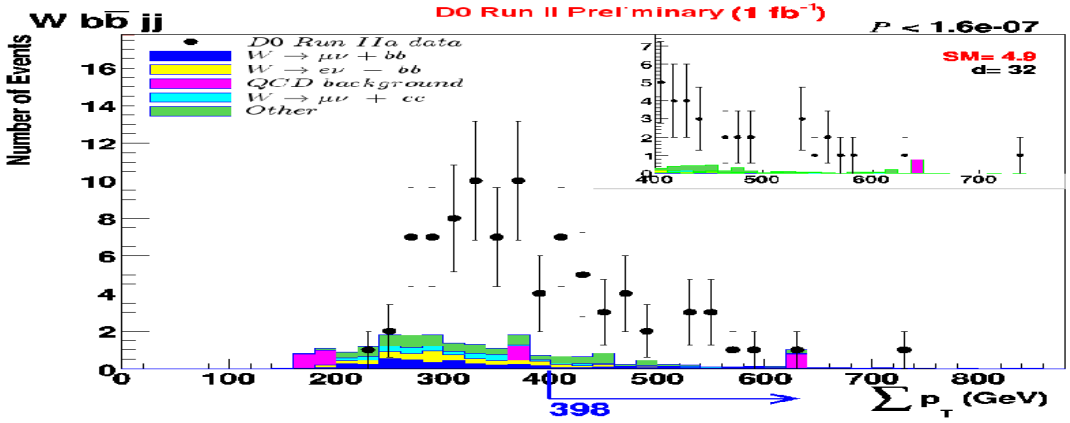


- Focuses on tails of Σp_T for each final state
- Searches for excesses
- Finds cut that maximizes discrepancy
- Simulates experiments to quantify significance
  - Use SM Monte Carlo to perform pseudo-experiments
  - How many to see something as interesting as data?

Additional checks for rare states to ensure Sleuth would find them

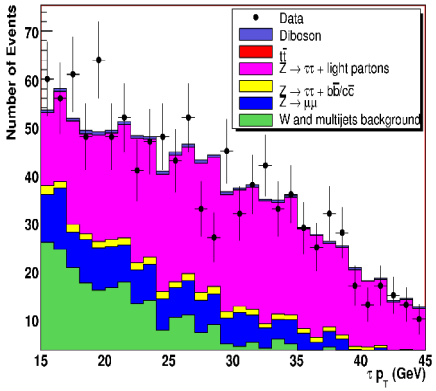
$t\bar{t}$ cross-check



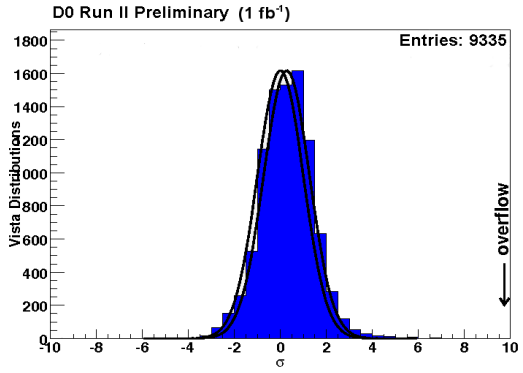


III. Results

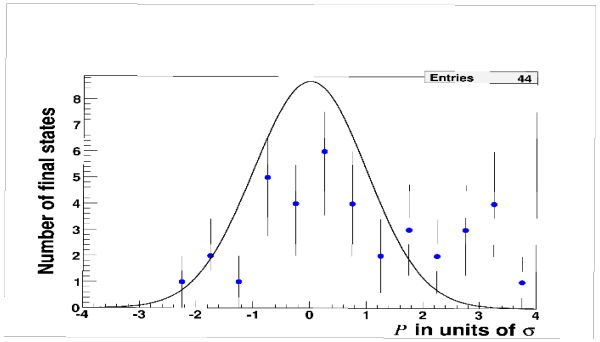
A. Qualitative D0 MIS Analysis Package Fits



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 process

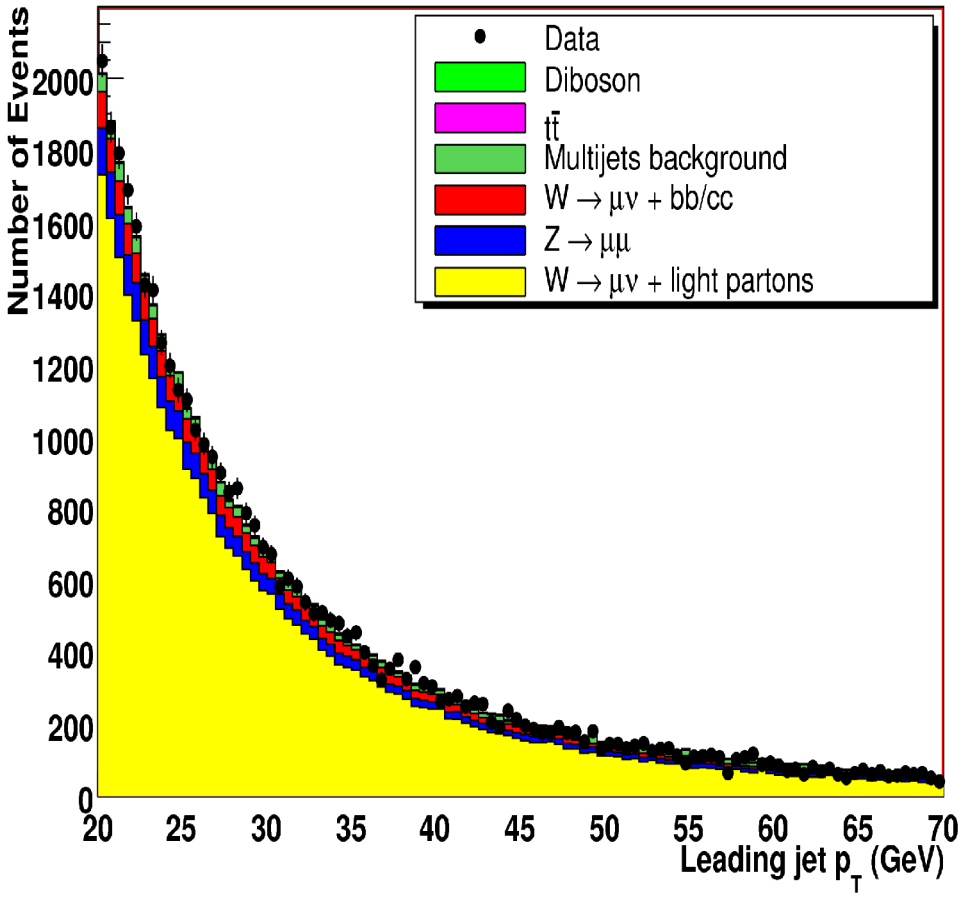
Fit parameters

- Multijet Bkg
- W w/light partons
- W w/heavy flavor

Event Cuts

- Single muon w/ at least one jet
  - μ p_T > 25 GeV
  - jet p_T > 20 GeV
  - MET > 20 GeV
- Tight cuts in muon system and tracker
- Isolated in cal and tracker

Work in progress



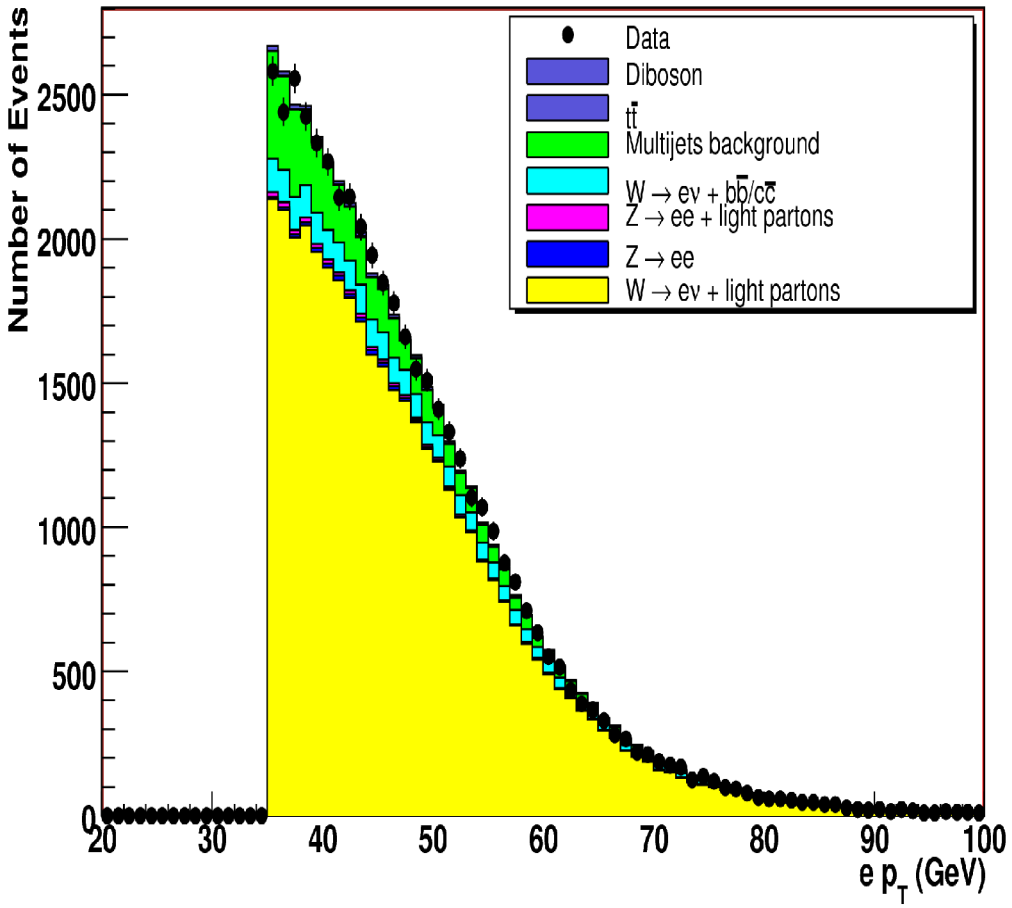
Single e + jets requires tighter cuts to avoid being dominated by multijet background

Same fitting parameters as μ + jets

Event Cuts

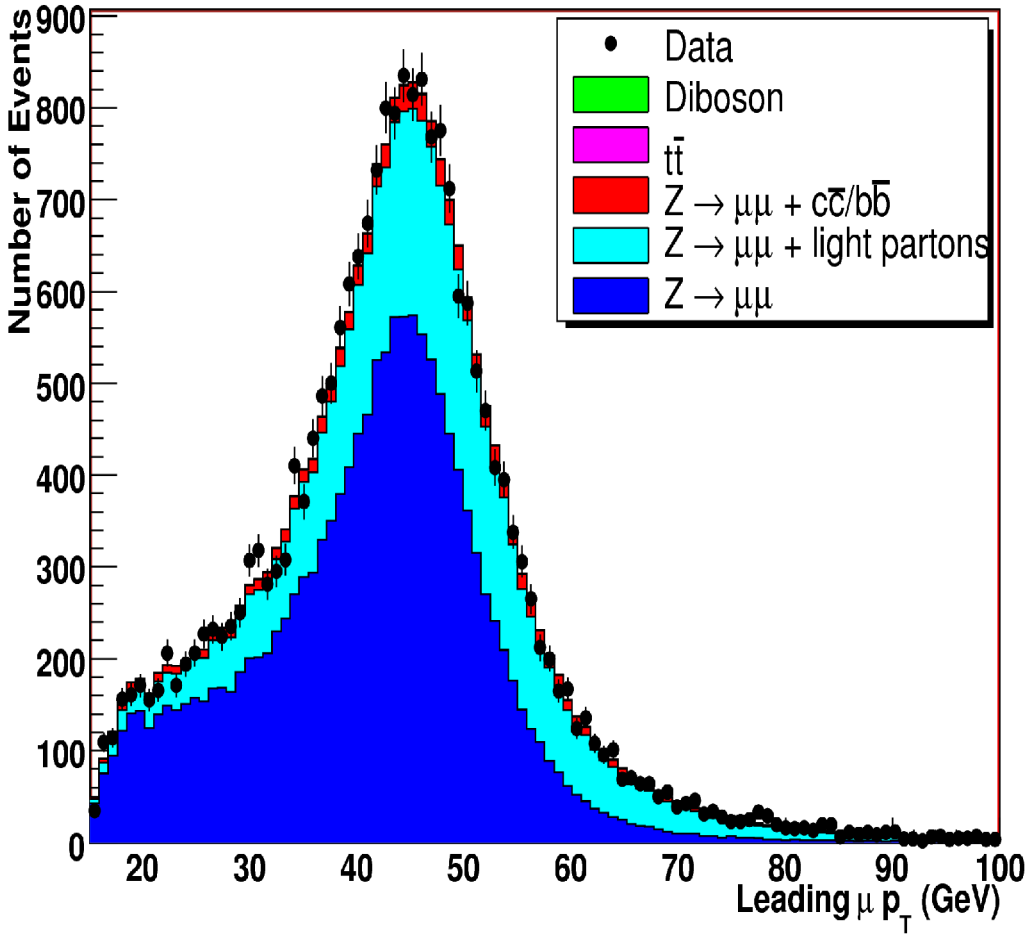
- **e with at least one jet**
  - e p_T > 35 GeV
  - e η < 1.1
  - jet p_T > 20 GeV
- **Strong likelihood cut of 0.95**
  - Multijets between 0.3 and 0.85
- **Basic EM calorimeter fraction cut (0.9)**

Work in progress



Lower p_T cut for dimuons, otherwise similar to μ + jets

Work in progress



Parameters

Drell-Yan

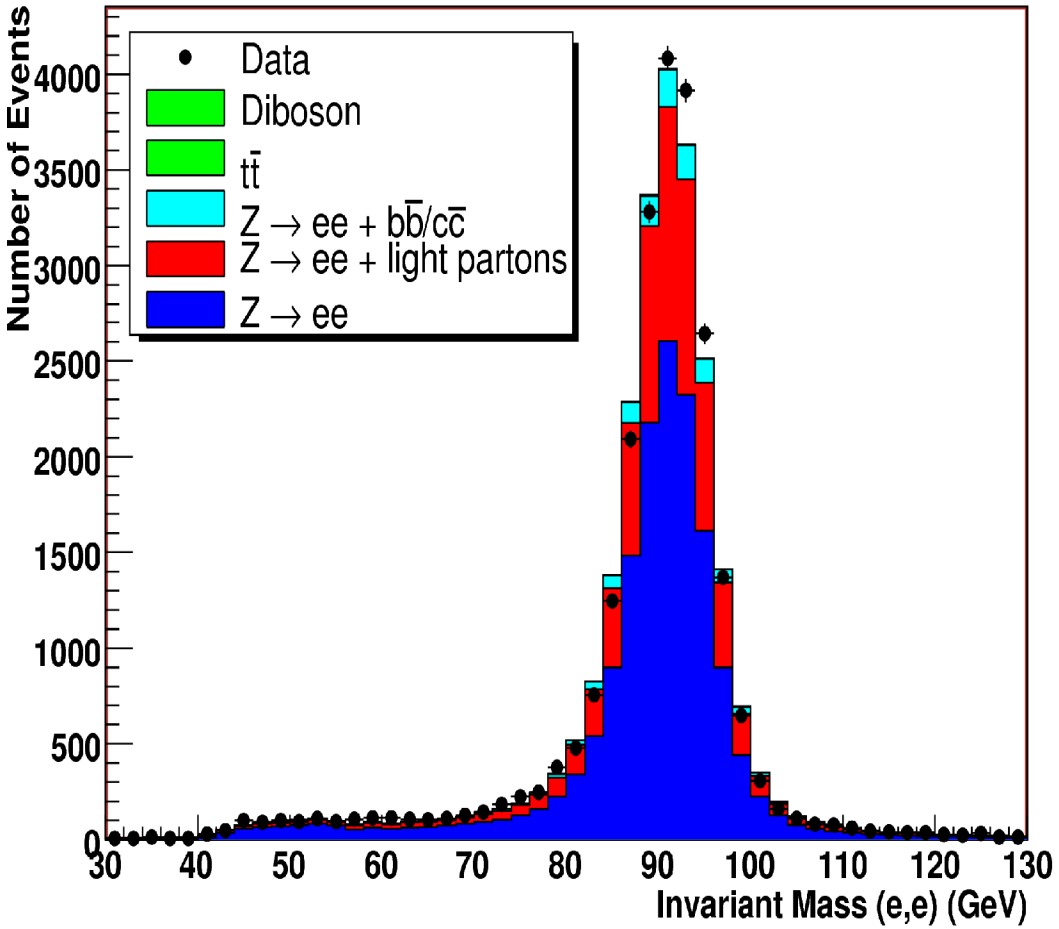
- w/no jets
- w/at least one light parton
- w/at least one heavy parton

Event Cuts

- Same iso cuts as single μ data
- μ p_T > 15 GeV
- μ η < 1.5

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

Work in progress



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

μ + τ uses single-prong, hadronic τs

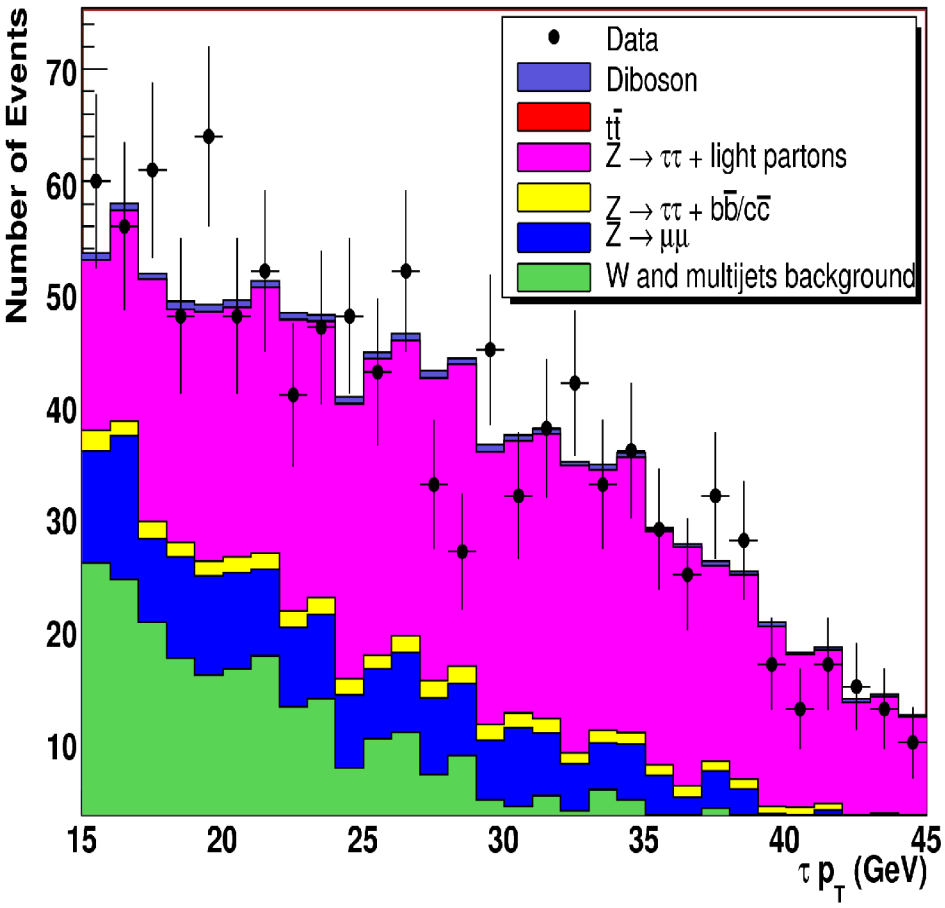
Parameters

- Combine multijet and W backgrounds from data
  - Cannot differentiate from pure multijet & $W \rightarrow \mu\nu$ + jet mis-ID
- Drell-Yan μμ: fixed
- Drell-Yan ττ

Event cuts

- μ p_T > 15 GeV
- μ η < 1.5
- τ p_T > 15 GeV
- Separate NNs for single π and dipion decays (e sep)

Work in progress



e + τ state uses same τ parameters with loose electron

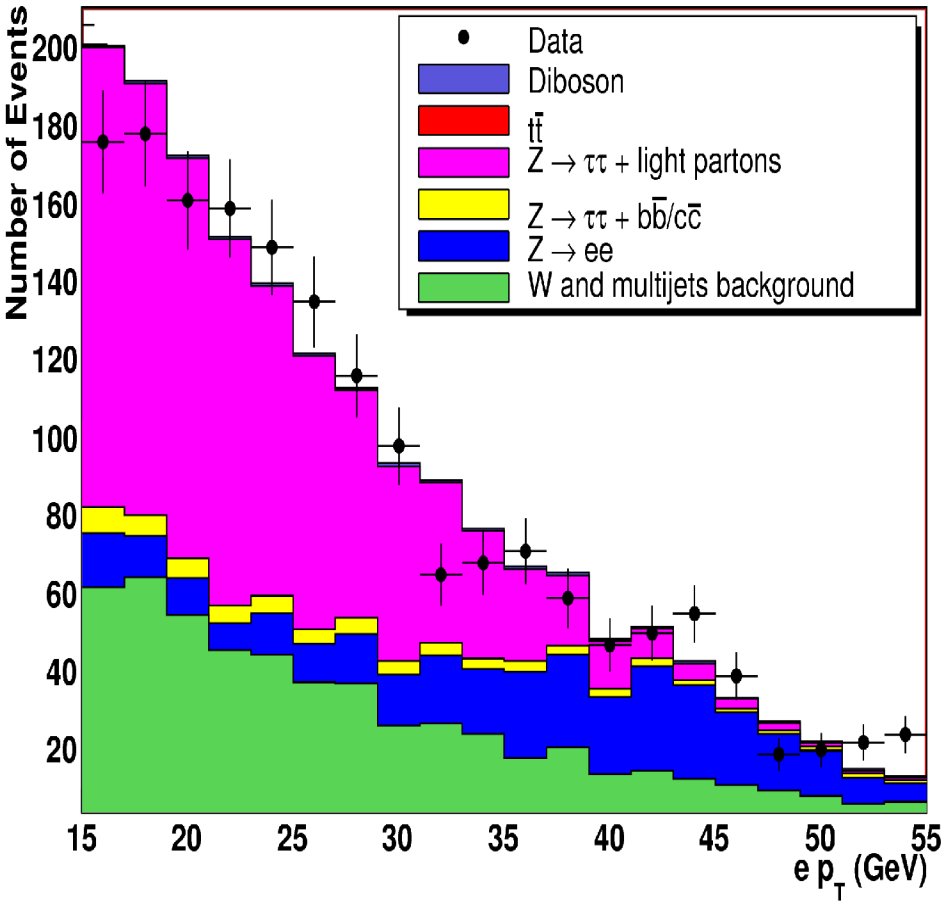
Parameters

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

Event cuts

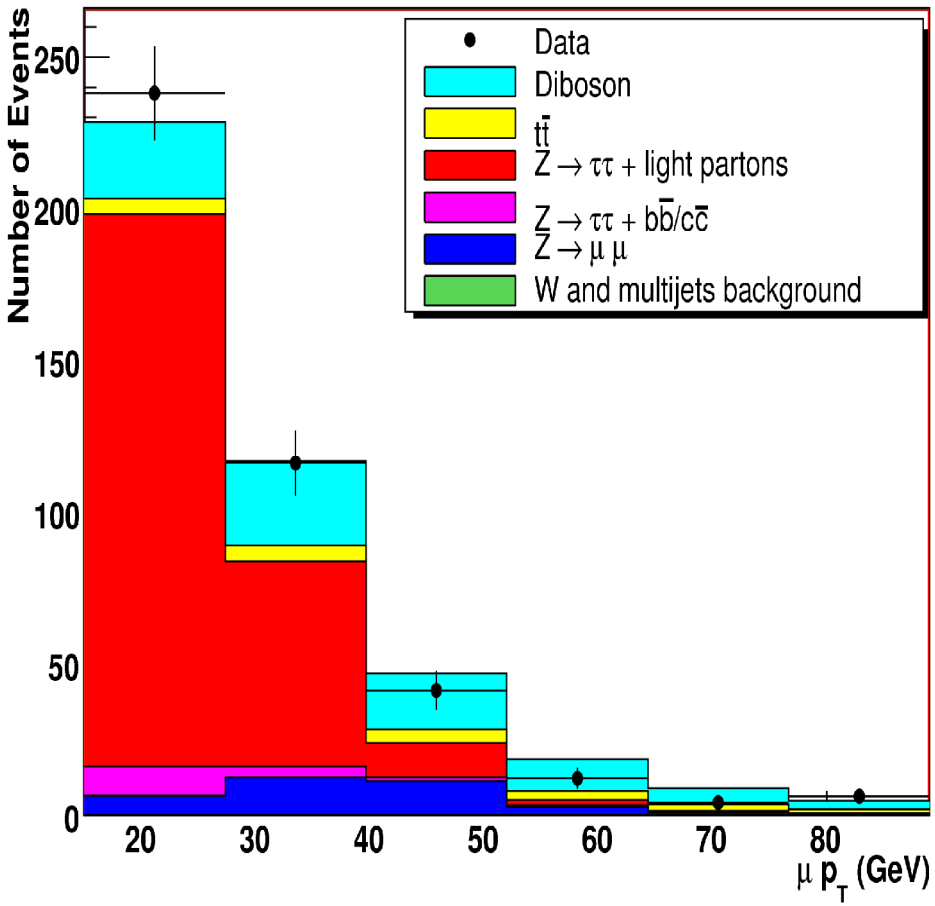
- e p_T > 15 GeV
- e η < 1.1
- τ p_T > 15 GeV
- Again, only single-prong τs

Work in progress



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

Work in progress



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 states and histograms agree, but focus is on disagreement

DØ sees 4/180 number and 24/9335 shape discrepancies

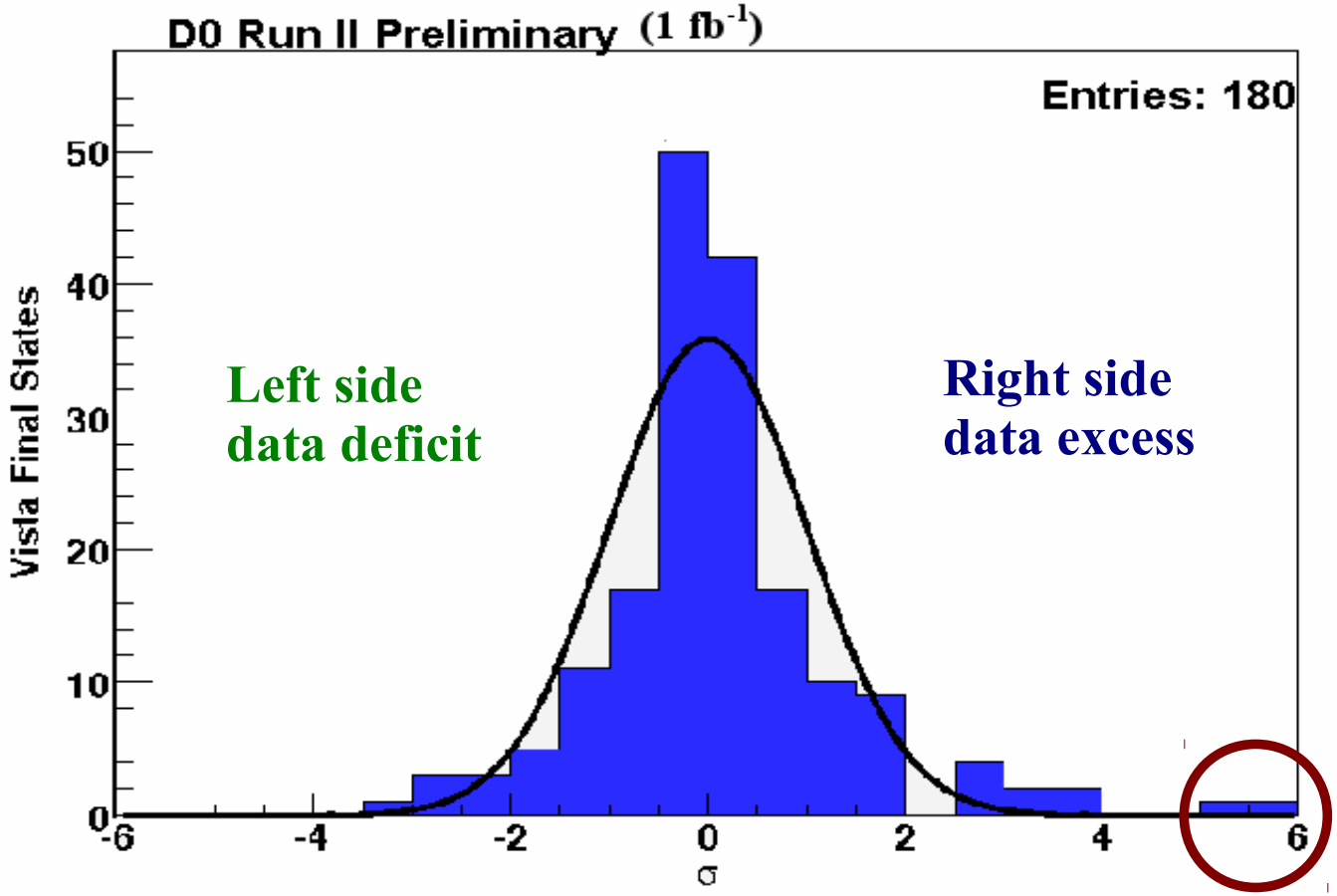
3 basic modeling issues

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

1 fb^{-1}

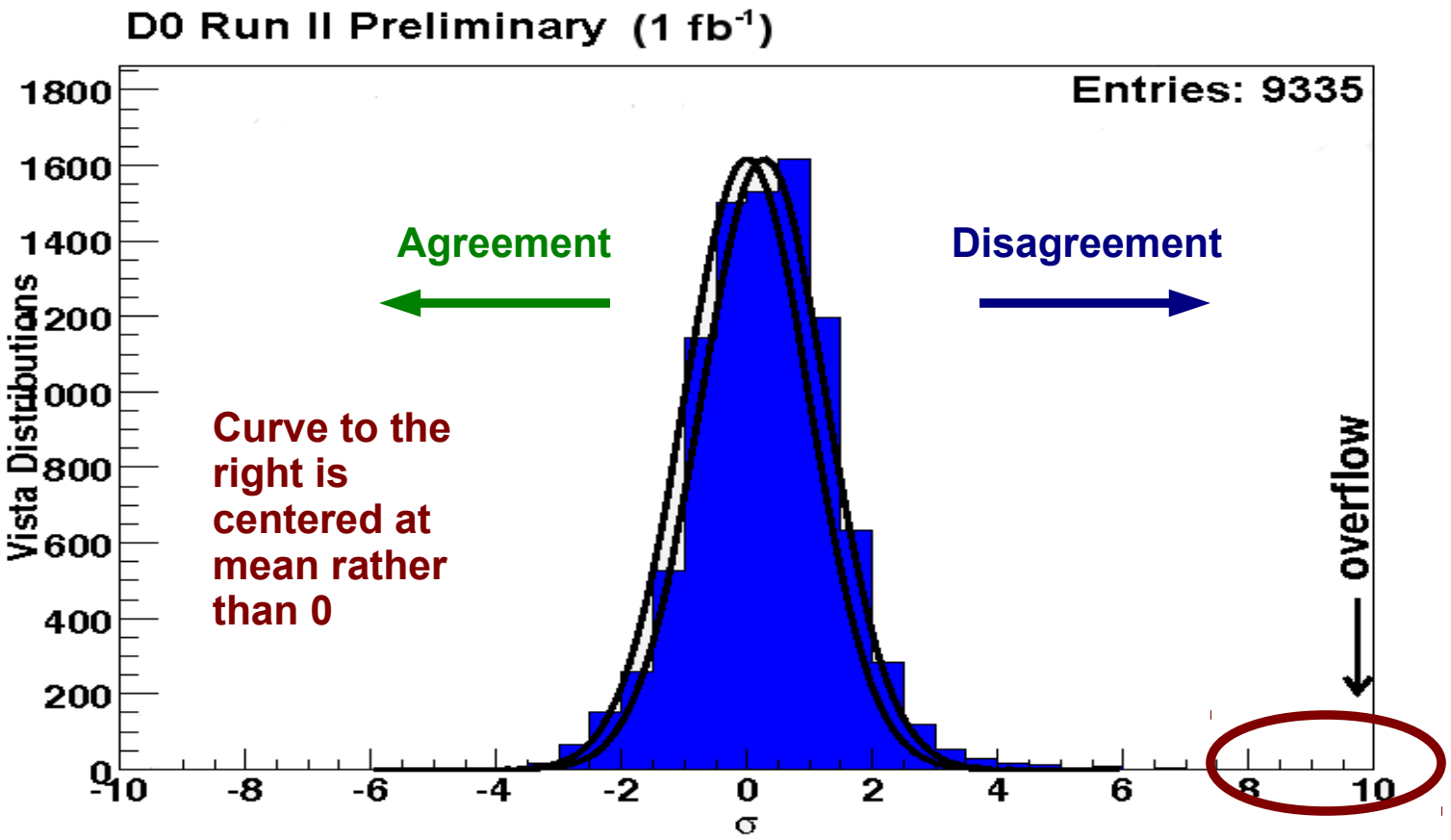
DØ Final State Population Discrepancies	σ After Trials Factor
$\mu + 2\,\text{jets} + \not\!\!E_T$	9.3
$\mu + \gamma + 1\,\text{jet} + \not\!\!E_T$	6.6
$\mu^{\pm}\mu^{\mp} + \not\!\!E_T$	4.4
$\mu^{\pm}\mu^{\mp} + \gamma$	4.1

Distributions of Final States' Data/MC Count Agreement



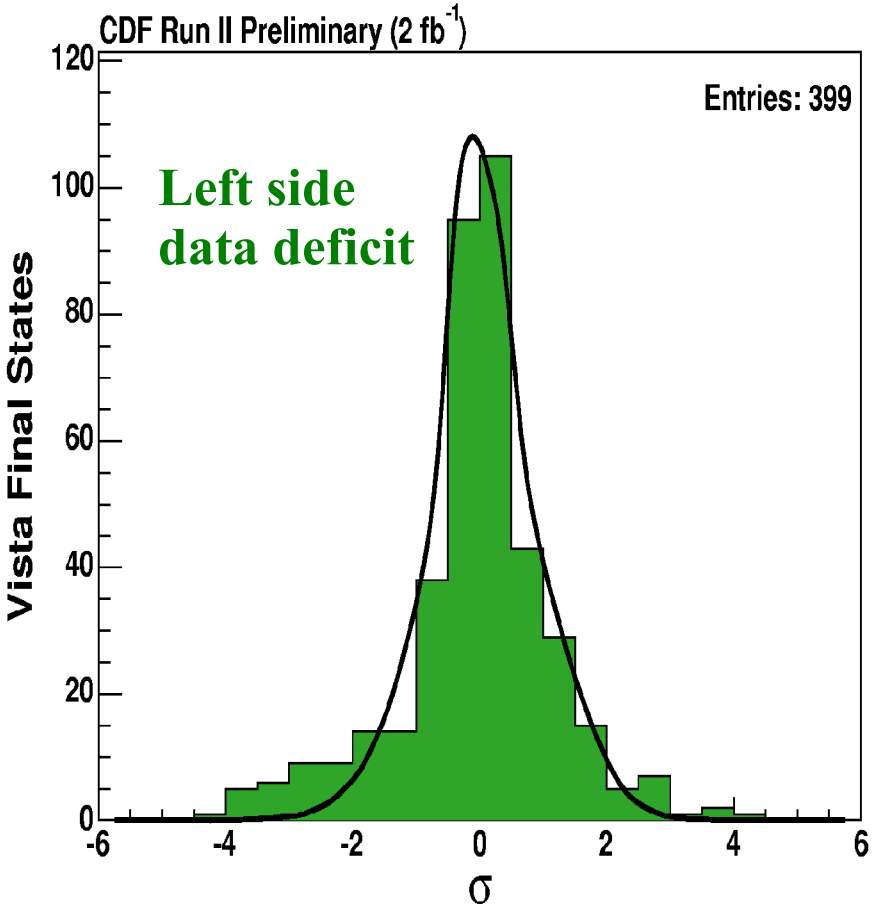
DØ Curve: Gaussian

Distributions of KS Probabilities in Histogram Shapes

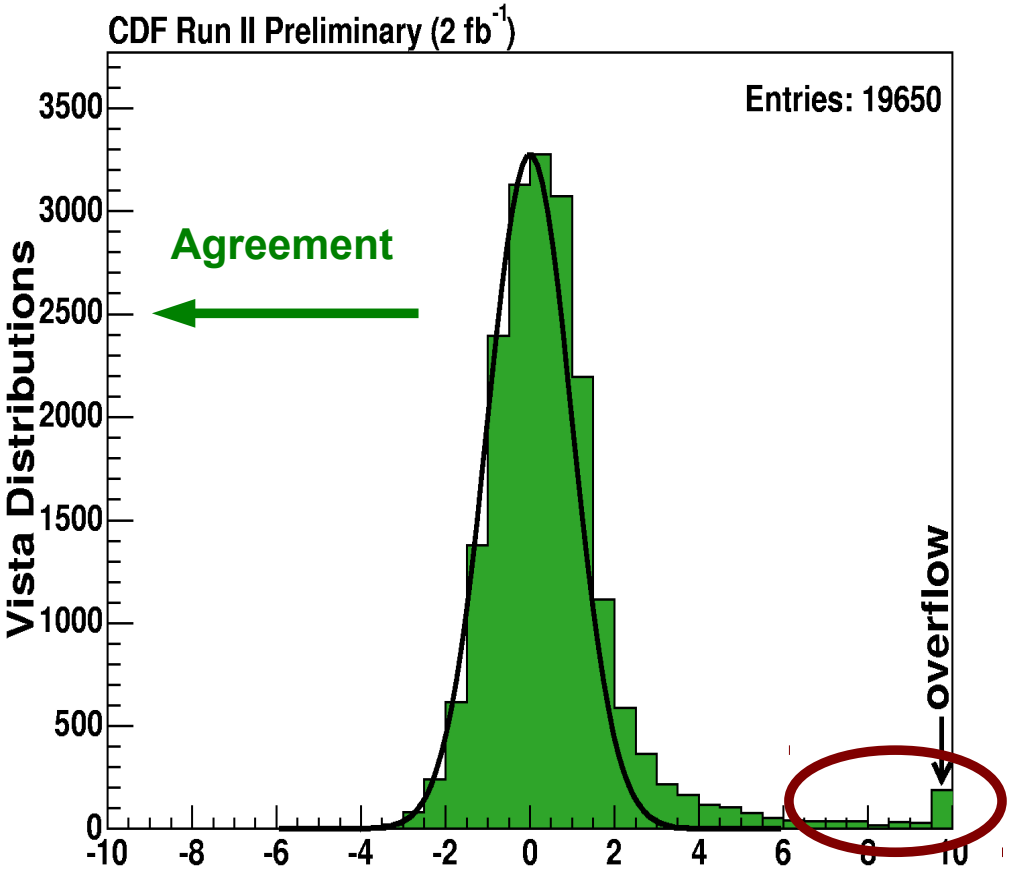


24 histograms show discrepancies after trials

This compares to CDF plots



Curve: pseudoexperiments

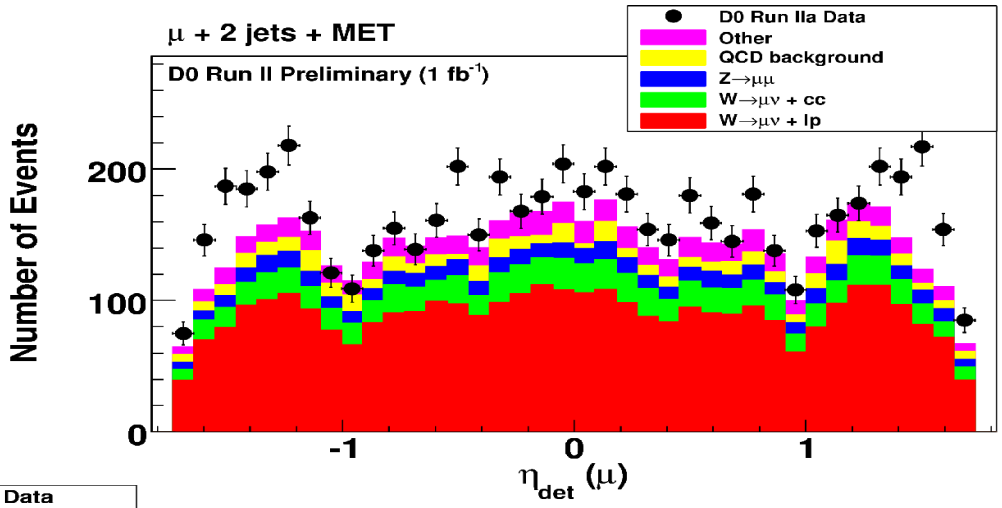


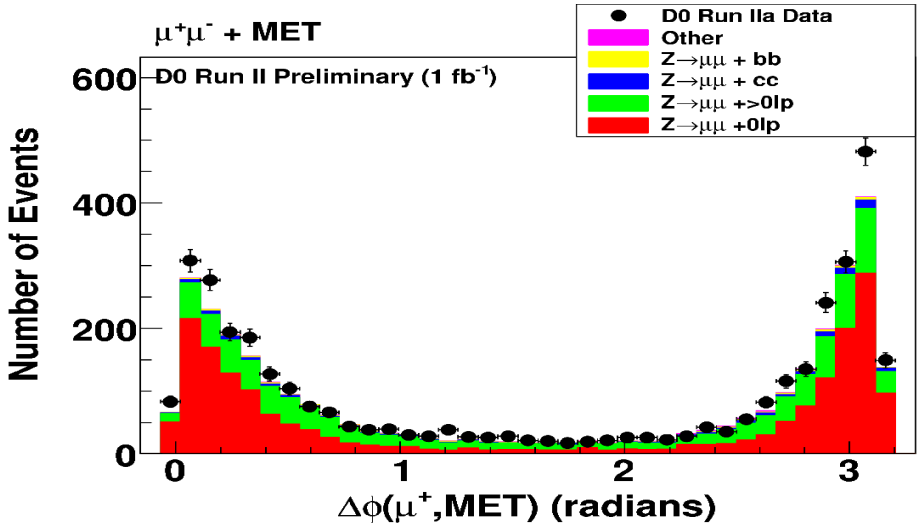
555 discrepancies σ mostly in multijets dominated states

Two number discrepancies point to subtle modeling issues

η-dependent trigger efficiency

- single muon
- muon + jets





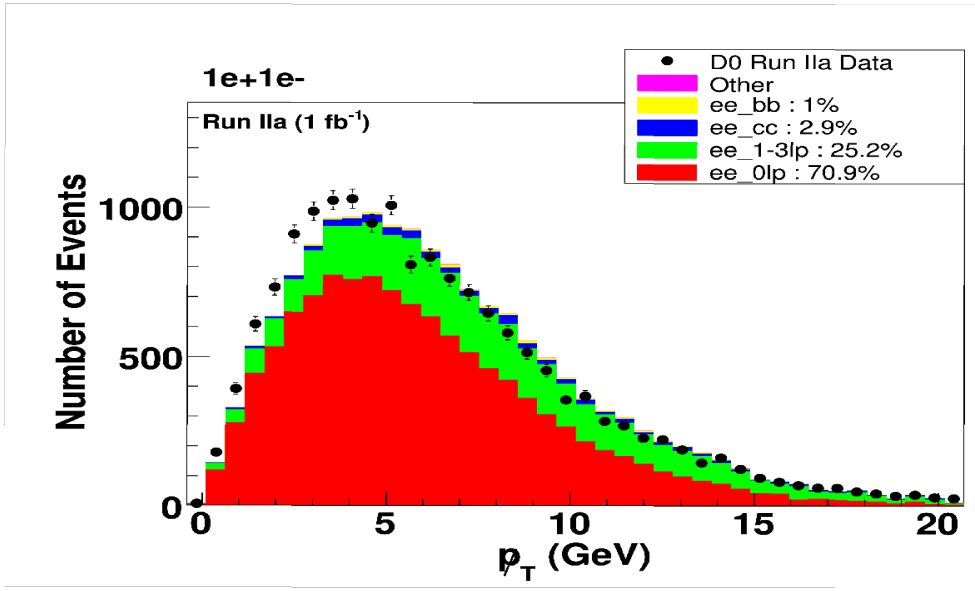
Track resolution modeling

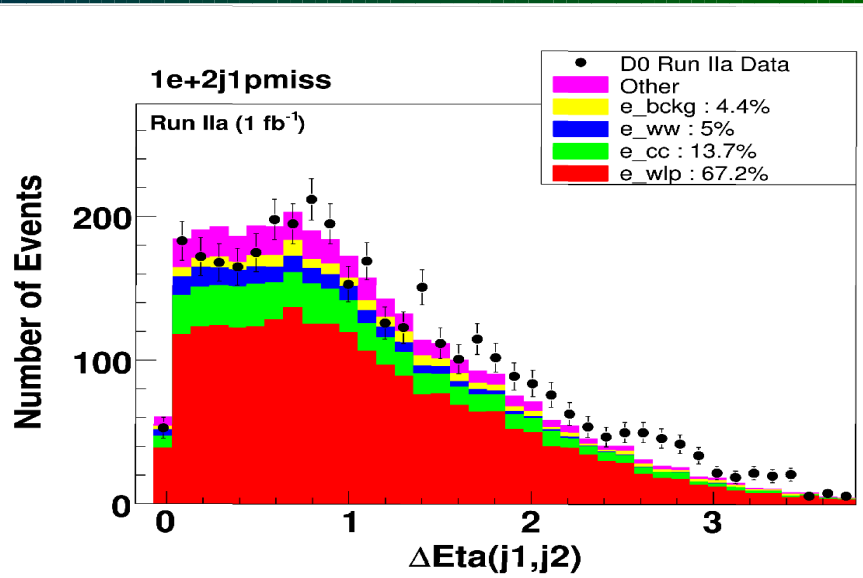
- high p_T μ
- muon and missing transverse energy are back-to-back
- appears again in Sleuth

Two of the most discrepant histograms in non-discrepant states

Alpgen-Pythia jet Δη disagreement

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

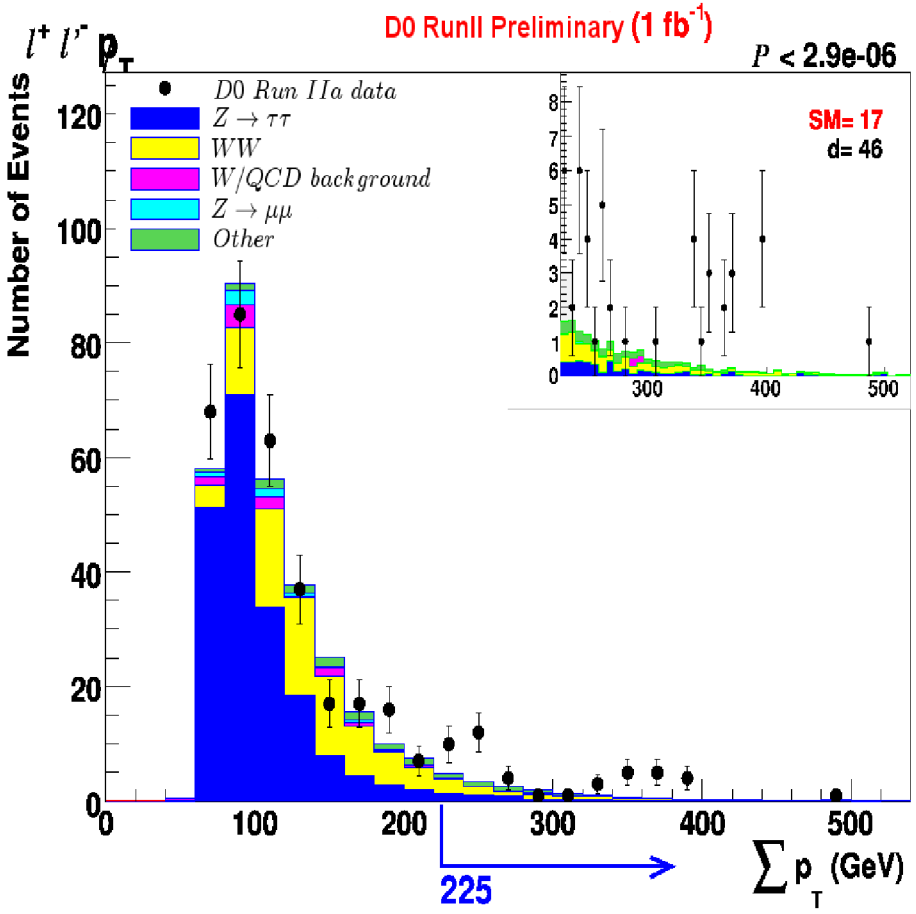




Some disagreement at very low MET

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

DØ Sleuth Results



OS 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

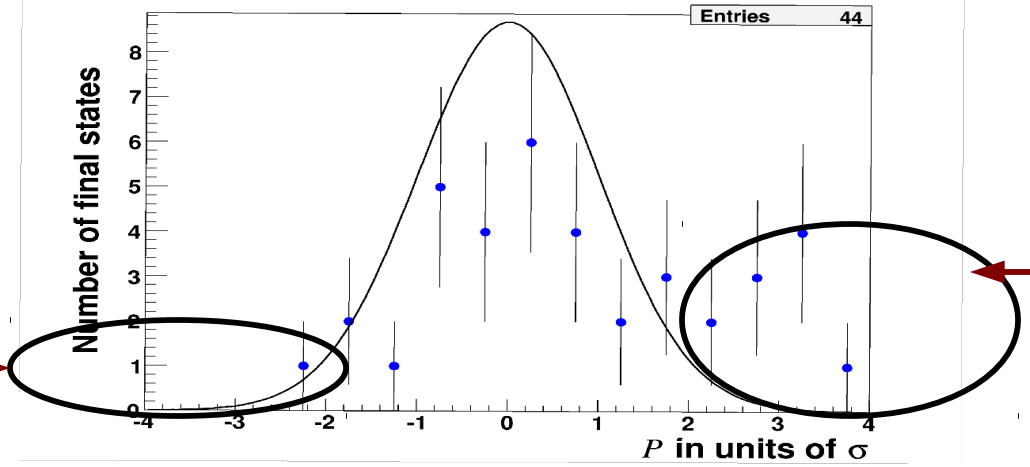
Distribution of Sleuth results show common theme

D0 Run II Preliminary

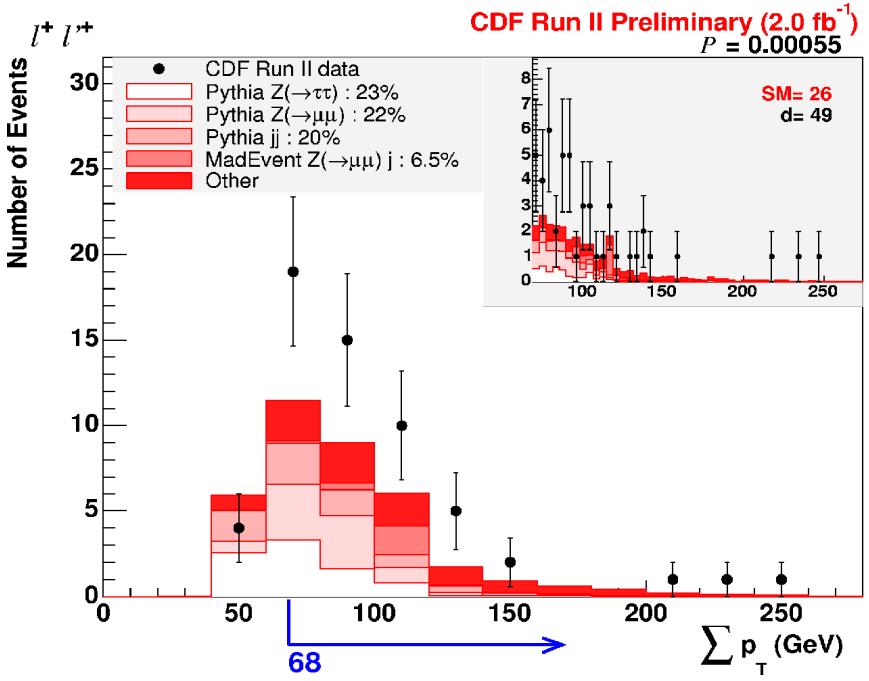
Final State	$\mathcal{P}$
$\ell^+\ell'^-$ + MET	2.9 E-6
ℓ + MET	.00082
$\ell^+\ell'^-$	.0031
$\ell^+\tau^-$ + MET	0.006
$\ell^+\tau^+$	0.0066

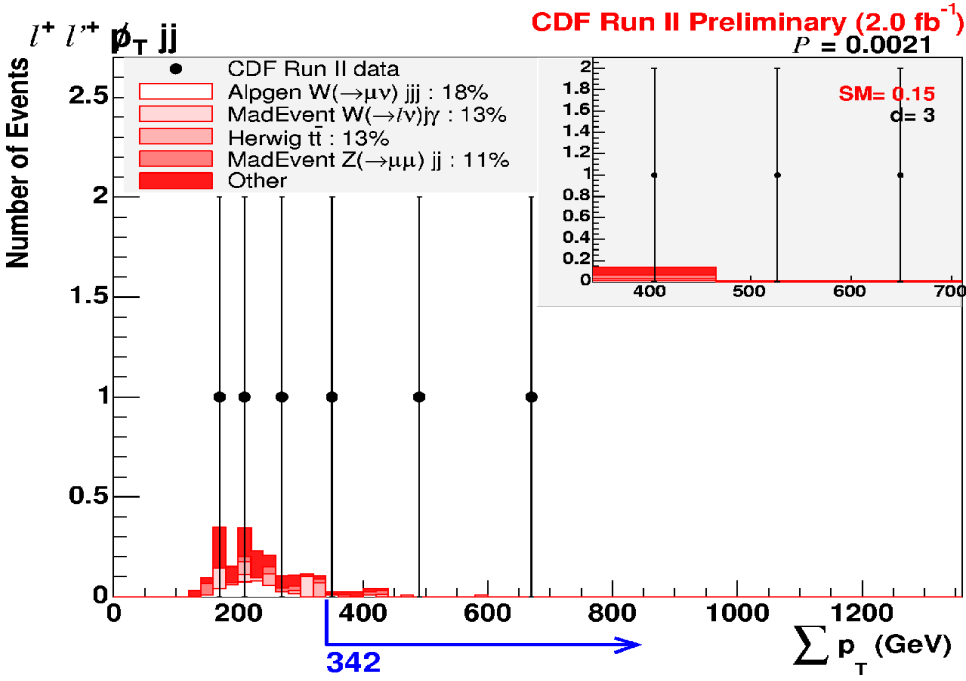
- Most significant states all show issues with μ resolution
- Search concentrates on a single most significant state
- Just leading state crosses threshold after trials... but systematic problem

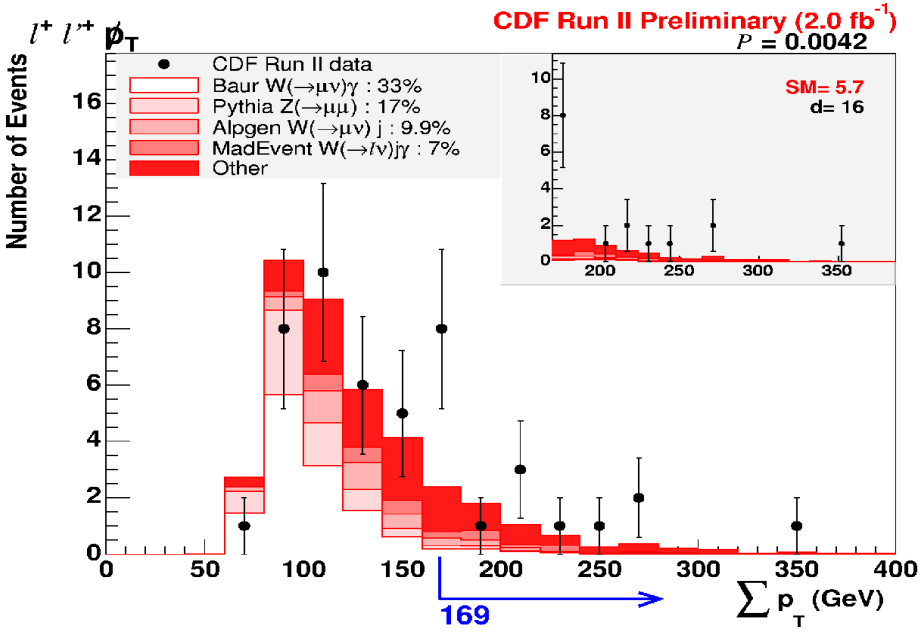
Few states on far left as would be expected for a background overestimate

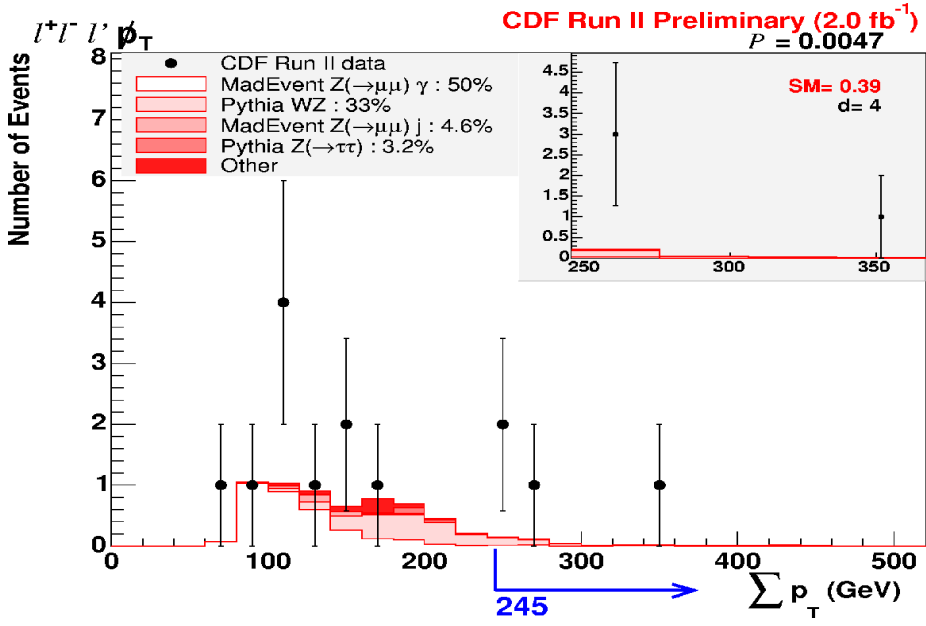


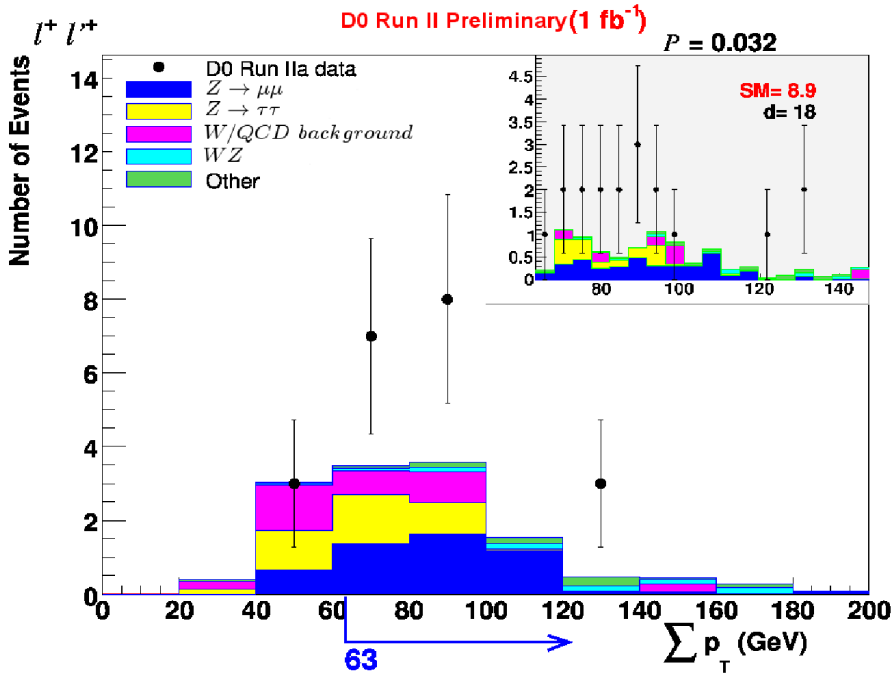
Cluster of states with high σ; many related to μ resolution issue

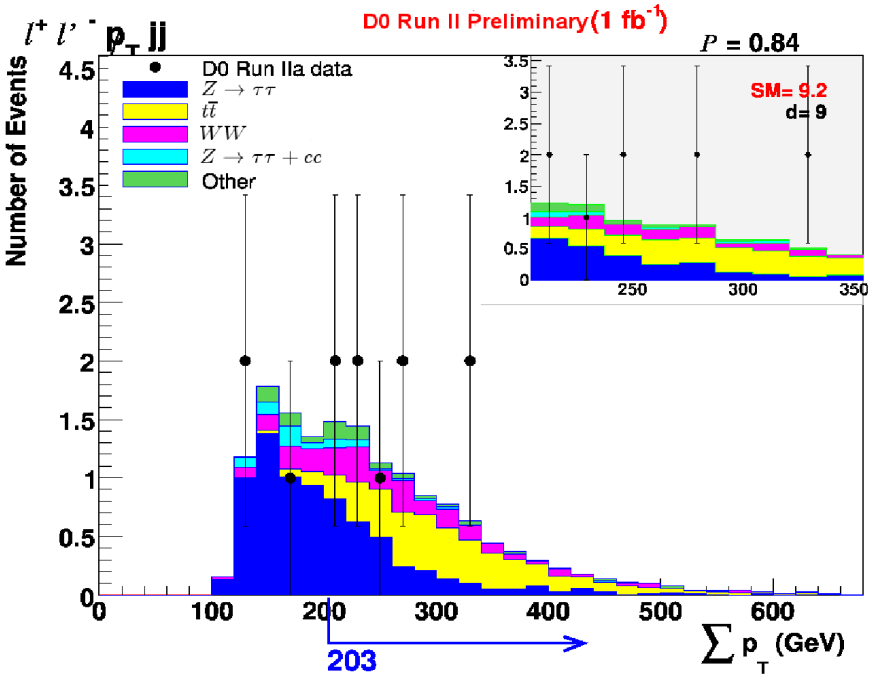

$l^+ l'^+$
CDF Run II Preliminary (2.0 fb^{-1})
P = 0.00055
CDF Run II data
Pythia Z(→ττ) : 23%
Pythia Z(→μμ) : 22%
Pythia jj : 20%
MadEvent Z(→μμ) j : 6.5%
Other
SM= 26
d= 49
Number of Events
$\sum p_T$ (GeV)
68

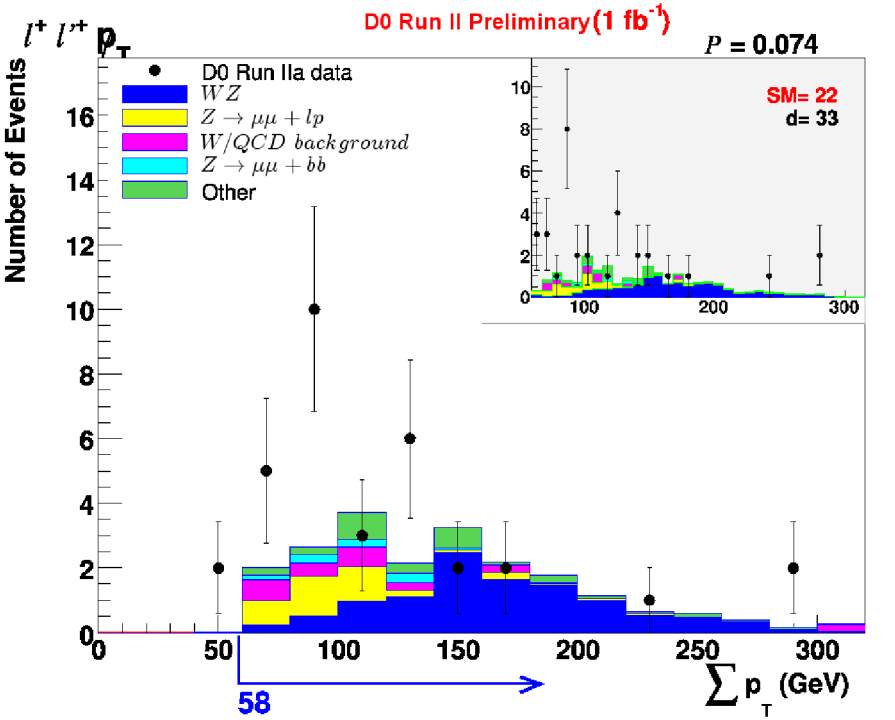

$l^+ l'^+ \not{p}_T$ jj
CDF Run II Preliminary (2.0 fb^{-1})
P = 0.0021
CDF Run II data
Alpgen W(→μν) jjj : 18%
MadEvent W(→lν)jγ : 13%
Herwig tt̄ : 13%
MadEvent Z(→μμ) jj : 11%
Other
SM= 0.15
d= 3
Number of Events
$\sum p_T$ (GeV)
342

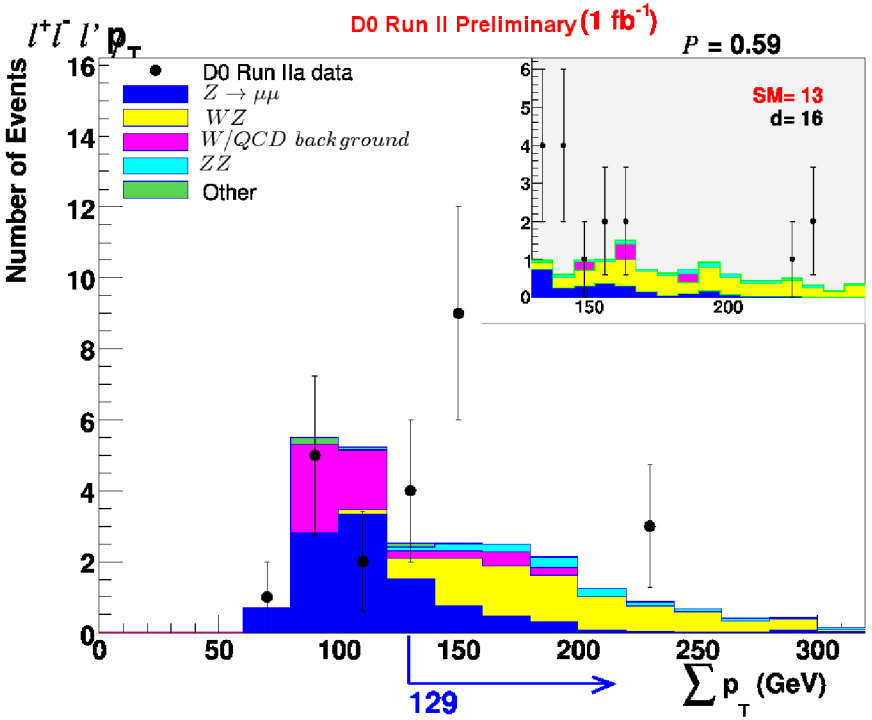

$l^+ l'^+ \not{p}_T$
CDF Run II Preliminary (2.0 fb^{-1})
P = 0.0042
CDF Run II data
Baur W(→μν)γ : 33%
Pythia Z(→μμ) : 17%
Alpgen W(→μν) j : 9.9%
MadEvent W(→lν)jγ : 7%
Other
SM= 5.7
d= 16
Number of Events
$\sum p_T$ (GeV)
169


$l^+ l^- l' \not{p}_T$
CDF Run II Preliminary (2.0 fb^{-1})
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
$\sum p_T$ (GeV)
245


$l^+ l'^+$
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
$\sum p_T$ (GeV)
63


$l^+ l'^- \not{p}_T$ 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
$\sum p_T$ (GeV)
203


$l^+ l'^+ \not{p}_T$
D0 Run II Preliminary(1 fb^{-1})
P = 0.074
D0 Run IIa data
WZ
$Z \to \mu\mu + lp$
W/QCD background
$Z \to \mu\mu + b\bar{b}$
Other
SM= 22
d= 33
Number of Events
$\sum p_T$ (GeV)
58


$l^+ l'^- \not{p}_T$
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
$\sum p_T$ (GeV)
129

Conclusions

- Vast majority of states and distributions are in agreement at DØ
- Vista considers final state populations and the shapes of 1-D histograms
  - Searched over 1 fb^{-1}
  - Four population discrepancies
  - 24 shape discrepant distributions
  - All of the given discrepancies point to modeling difficulties
- Sleuth examines specific distributions for high-p_T excesses
  - One Sleuth discrepancy
  - Probably related to μ resolution
- No hints yet... but with 6.0 fb^{-1} of data already collected at each experiment, much more data are available for searches

Backups

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

List of trigger names stored internally

If event passes particular trigger, bit is set

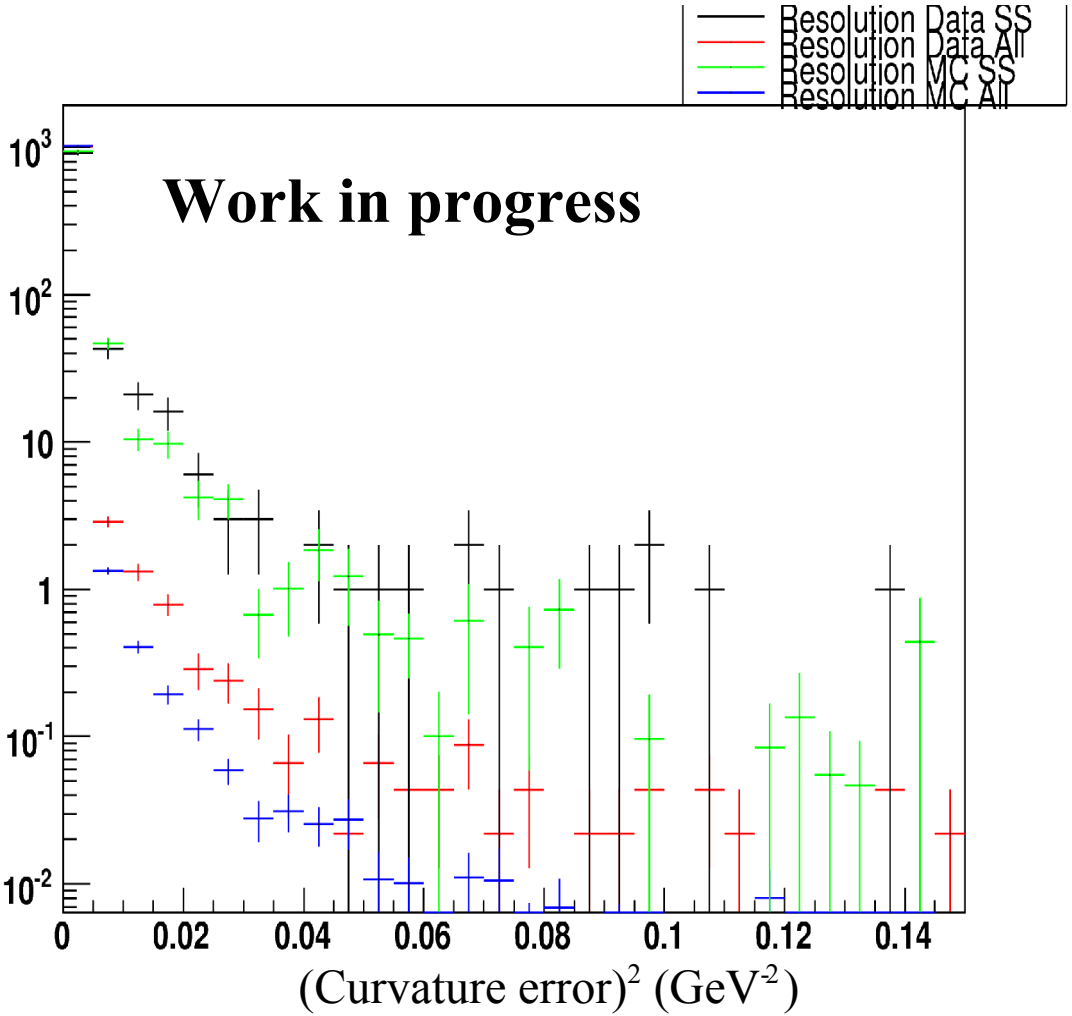
All physics triggers included on list

Use standard efficiency for trigger suites

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

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

