Expecting the Unexpected

A global analysis of Run2 data

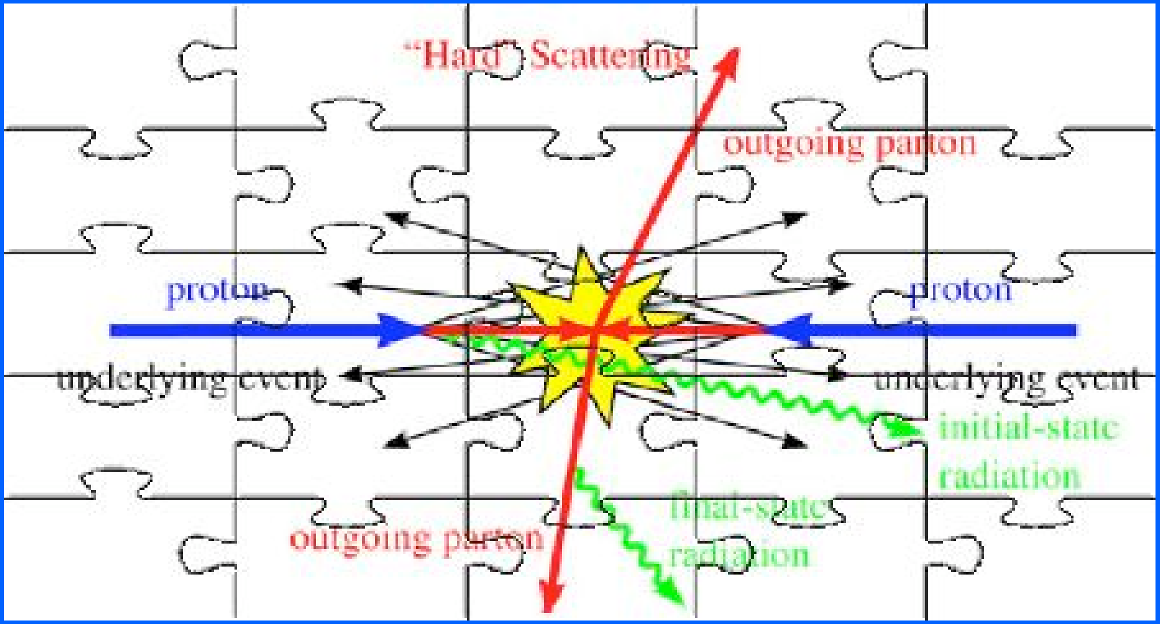
Stephen Mrenna
Scientist I
Computing Division
Fermilab

Understanding High-Pt Physics: many pieces to the puzzle

LO, NLO and NNLLO, NLO and NNLO calculations
K-factors

Benchmark cross sections and pdf correlations

PDFs with uncertainties



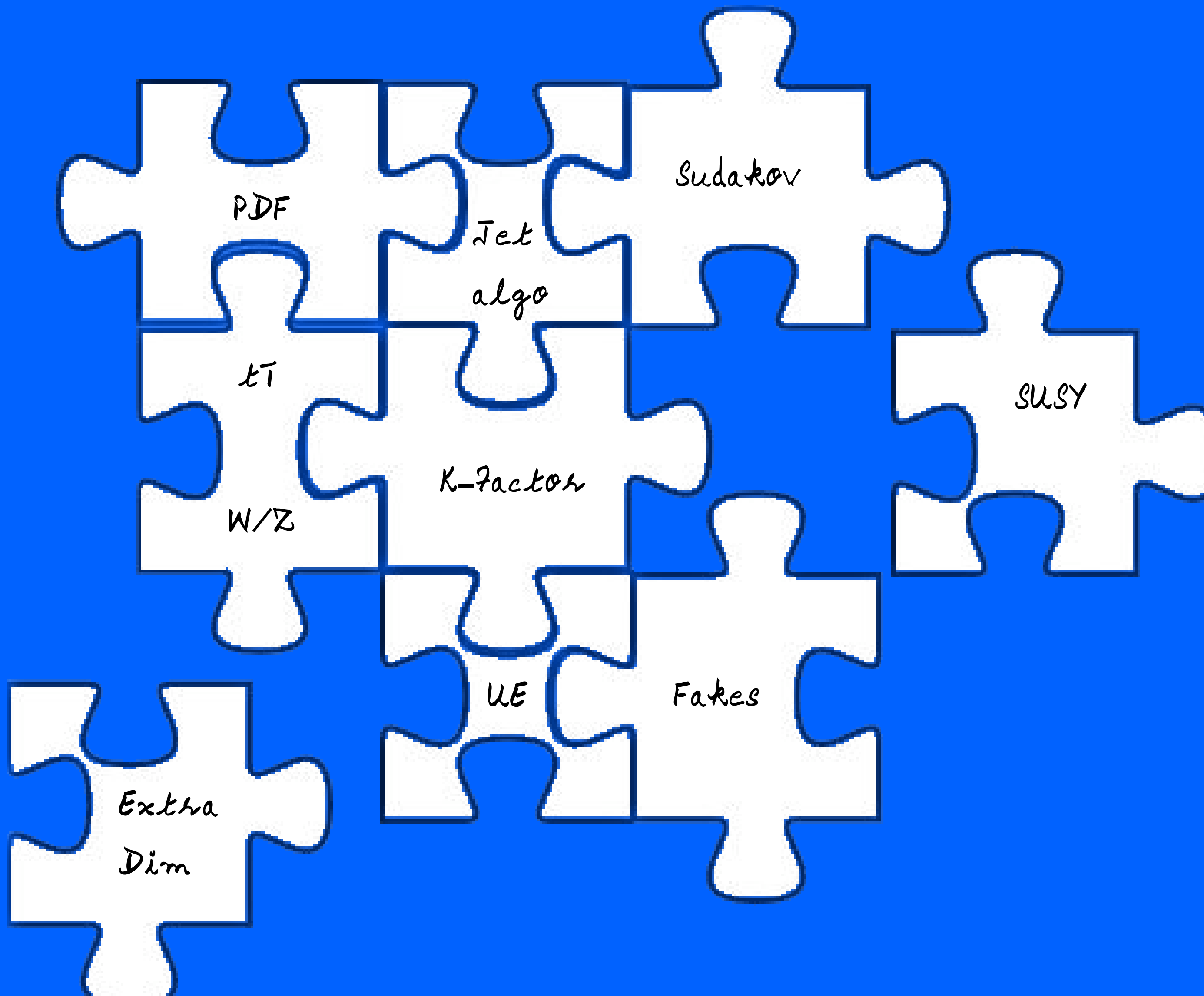
Underlying event and minimum bias

Fragmentation/Hadronization

Sudakov form factors

Jet algorithms and jet reconstruction

(How) will the puzzle pieces fit together?



Missing Energy Events are Confirmed

1) Events are real

2) Not conventional physics
$W \to \tau$, $Q\bar{Q}$, $Z^0 g$
$\hookrightarrow \nu\nu$

3) Not $X \to Z^0 +$ jet(s)
$\hookrightarrow \nu\nu$

4) Not $Z^0 \to X_1 X_2$
$\hookrightarrow \nu$'s or stable
$\hookrightarrow$ jet(s)

(ORIGINAL TRANSPARENCY FROM 1986 UA1 'DISCOVERY' OF SUSY
ASPEN CONF, 1986

Table 1. Predicted rates for processes giving large missing transverse energy events passing all event selection cuts.

Data

Process	Events (total)	Events with $L_\tau < 0$	Events with $L_\tau < 0$ and $E_T^{jet} < 40$ GeV
$W \to e\nu$ $W \to \mu\nu$ $W \to \tau\nu \to$ leptons	3.6	2.0	1.4
$W \to \tau\nu$ $\to \nu\bar{\nu}$ + hadrons	34.7	8.0	7.1
$W \to c\bar{s}$	<0.1	<0.1	<0.1
$Z^0 \to \tau^+\tau^-$	0.5	0.1	0.1
$Z^0 \to \nu\bar{\nu}$ (3 neutrino species)	7.4	7.1	5.6
$Z^0 \to c\bar{c}$ and $b\bar{b}$	<0.1	<0.1	<0.1
$c\bar{c}$ and $b\bar{b}$ (direct production)	0.2	0.2	0.2
Jet fluctuations (fake missing energy)	3.8	3.4	3.4
TOTAL	52.2	20.8 ± 5.1 ± 1.0	17.8 ± 3.7 ± 1.0

56

24

How well do we understand the Standard Model (@ high pT)?

It seems to work very well ... how well?

- Global Analysis
- Null search for New Physics

What does that mean for the LHC?

#1: Global Analysis @ High Pt

Define high-p_T objects reconstructed in experiment (CDF in this case)

Generate-Simulate Monte Carlo events and reconstruct same objects

Introduce a correction model (fakes, K-factors, uncertainties) and refine

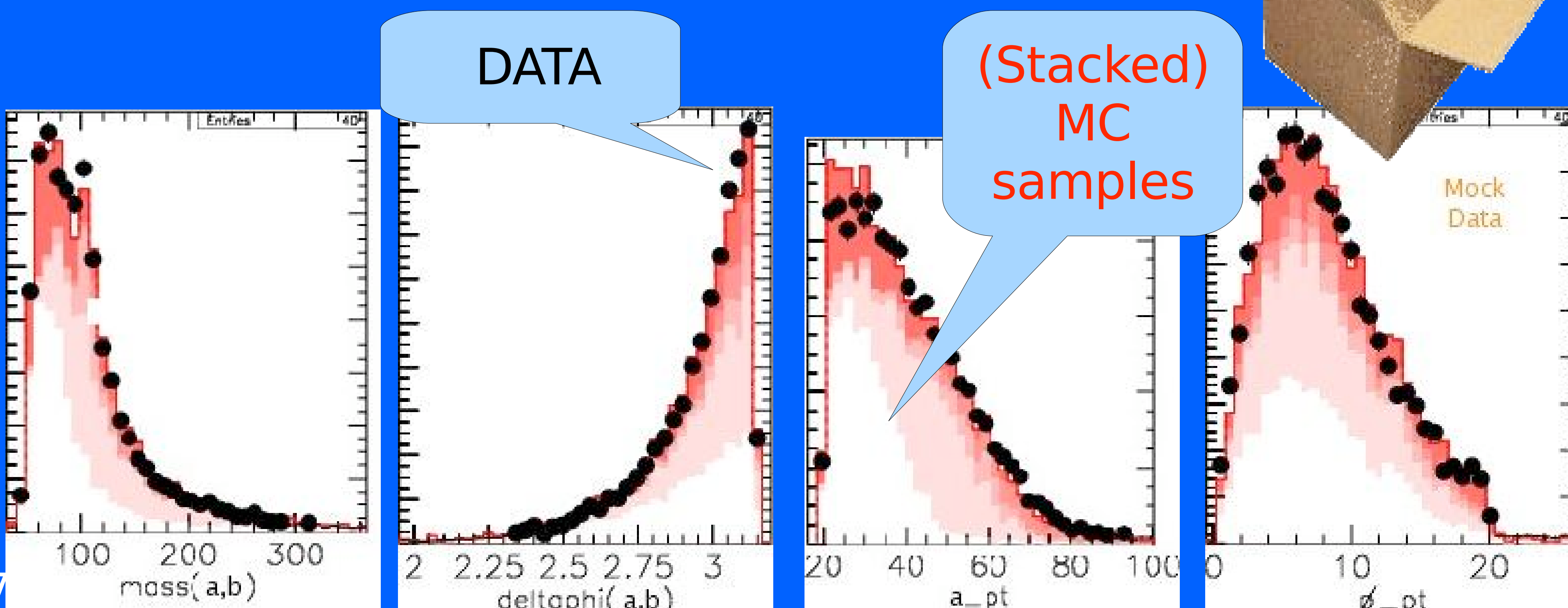
Compare counts and shapes in different final states

Final State: 1a 1b 1pmiss

1 high-P_T object "a"+ any number low-P_T

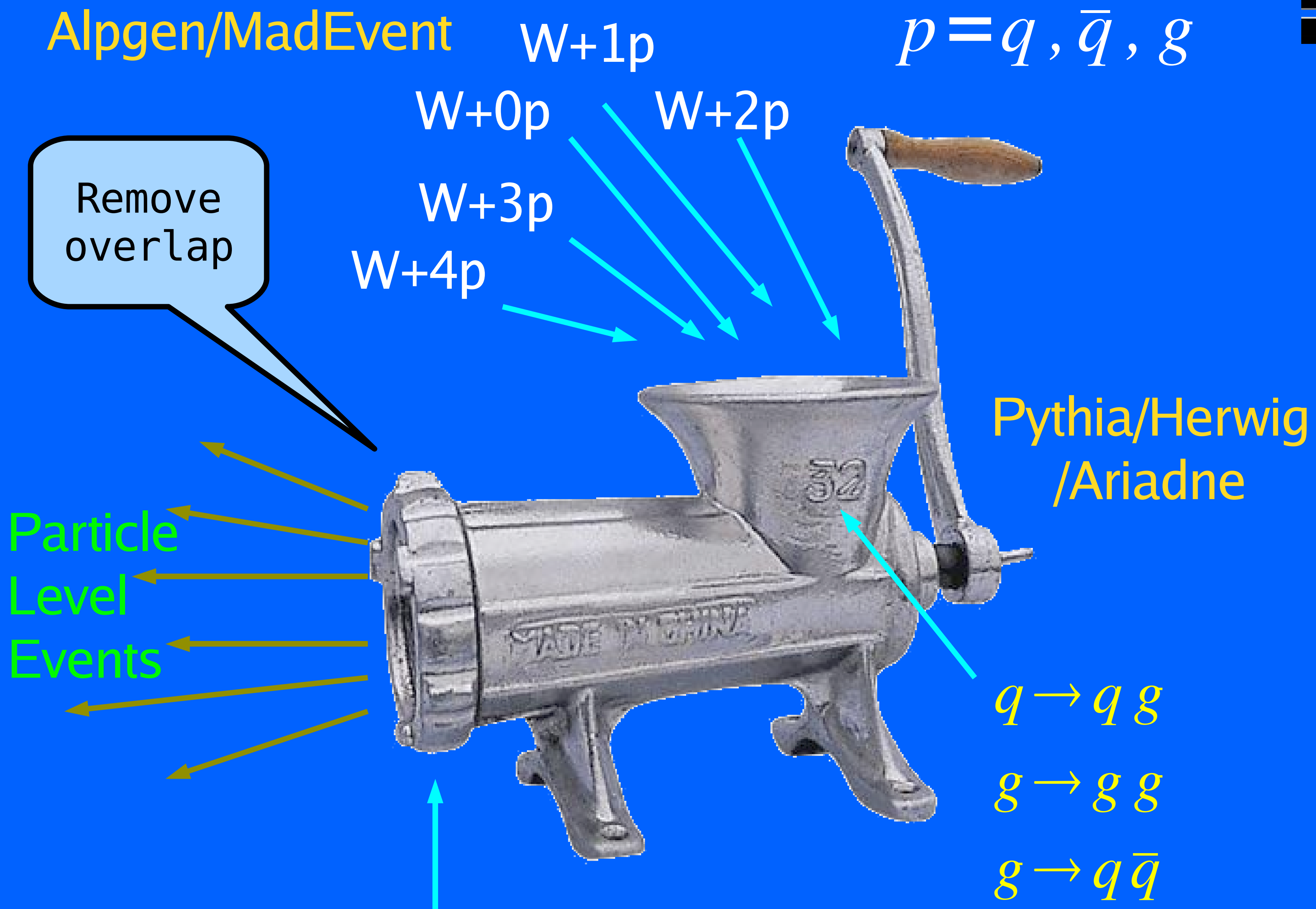
1 high-P_T object "b"+ any number low-P_T

Significant missing -P_T

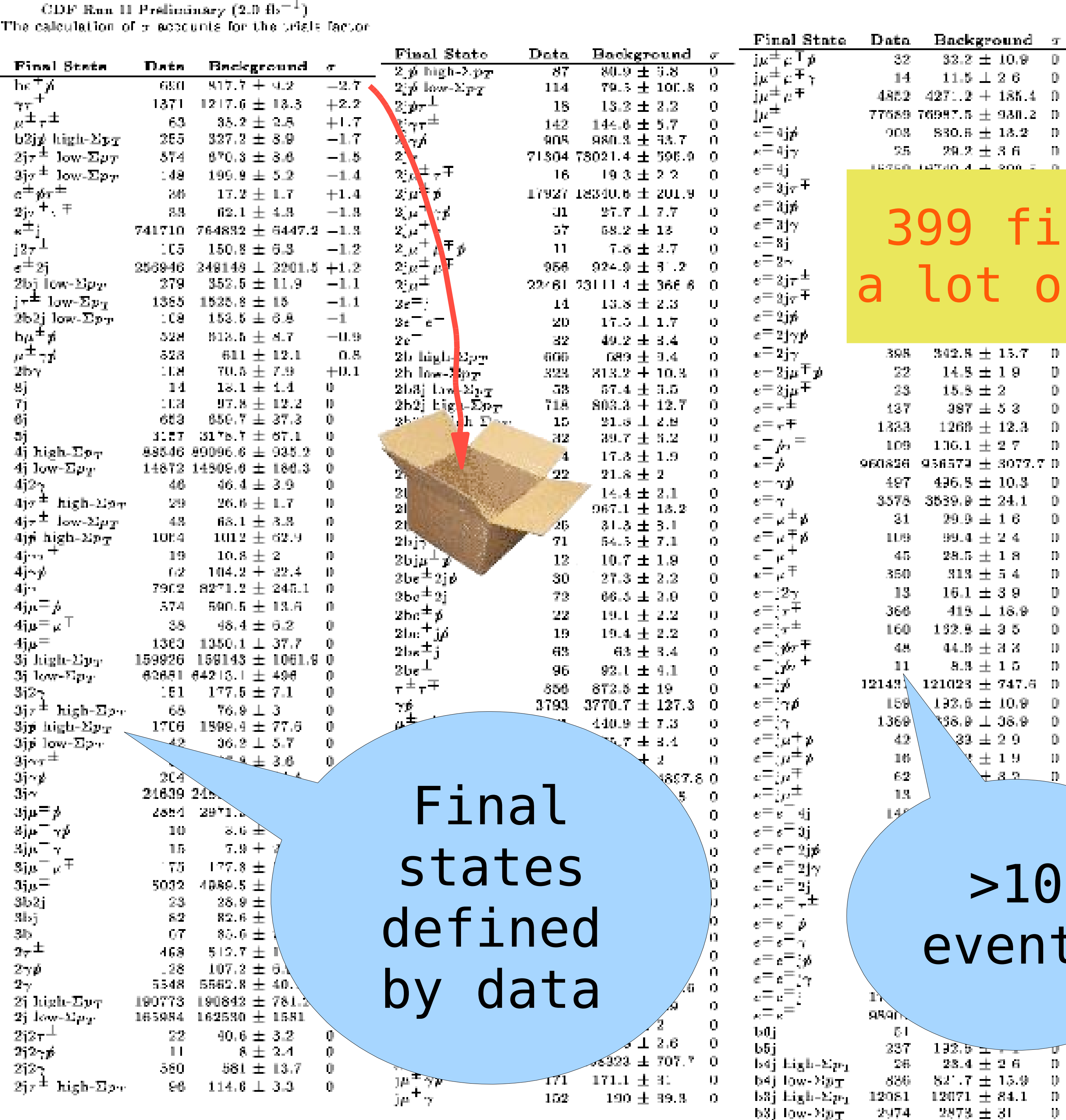


Modeling the SM in practice

- We know the importance of PDFs, NLO …
- In practice, we try to use the data to calculate all orders, pert and non
- Data(Y) = MC(Y)/MC(X) * Data(X)
    - Other theoretical developments are used mainly for cross checks or to model signals
- Like mixing cocktails or making sausage



$$W + 4\mathrm{p} \rightarrow W + 4\mathrm{j} + \text{softer stuff}$$



399 final states:
a lot of information

$2\,fb^{-1}$

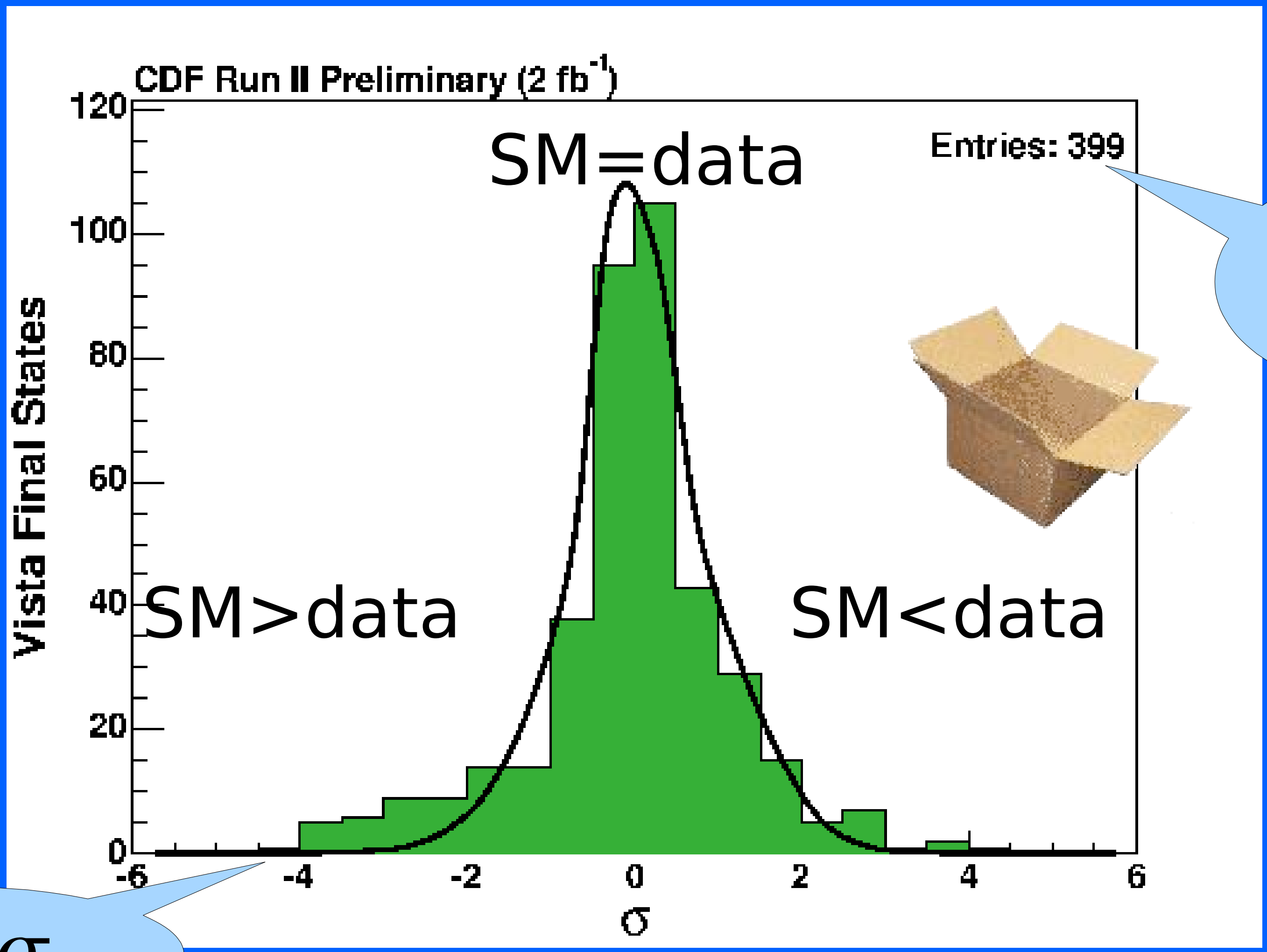
Final states defined by data

>10 events

ArXiv:0712.1311 CDF+...SM, GC, RC, CH, BK, SX

Vista final state normalizations

CDF RunII 2 fb^{-1}



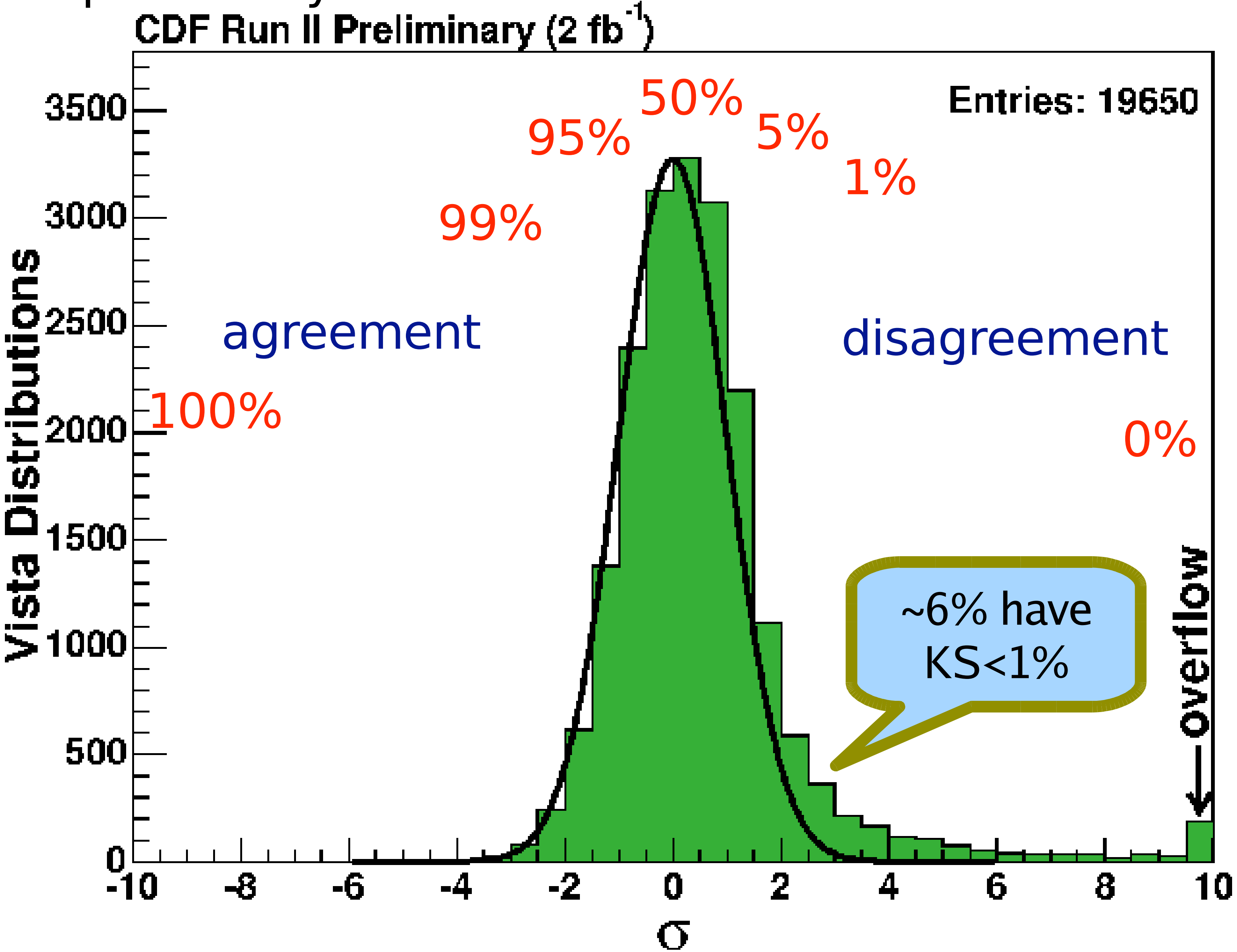
Trials factor

2.7σ

G. Choudalakis, R. Culbertson, C. Henderson, B, Knuteson, Si Xie

Vista kinematic shapes

KS probability



Quantitative Results

Event counts are distributed as you expect when you look at 399 final states

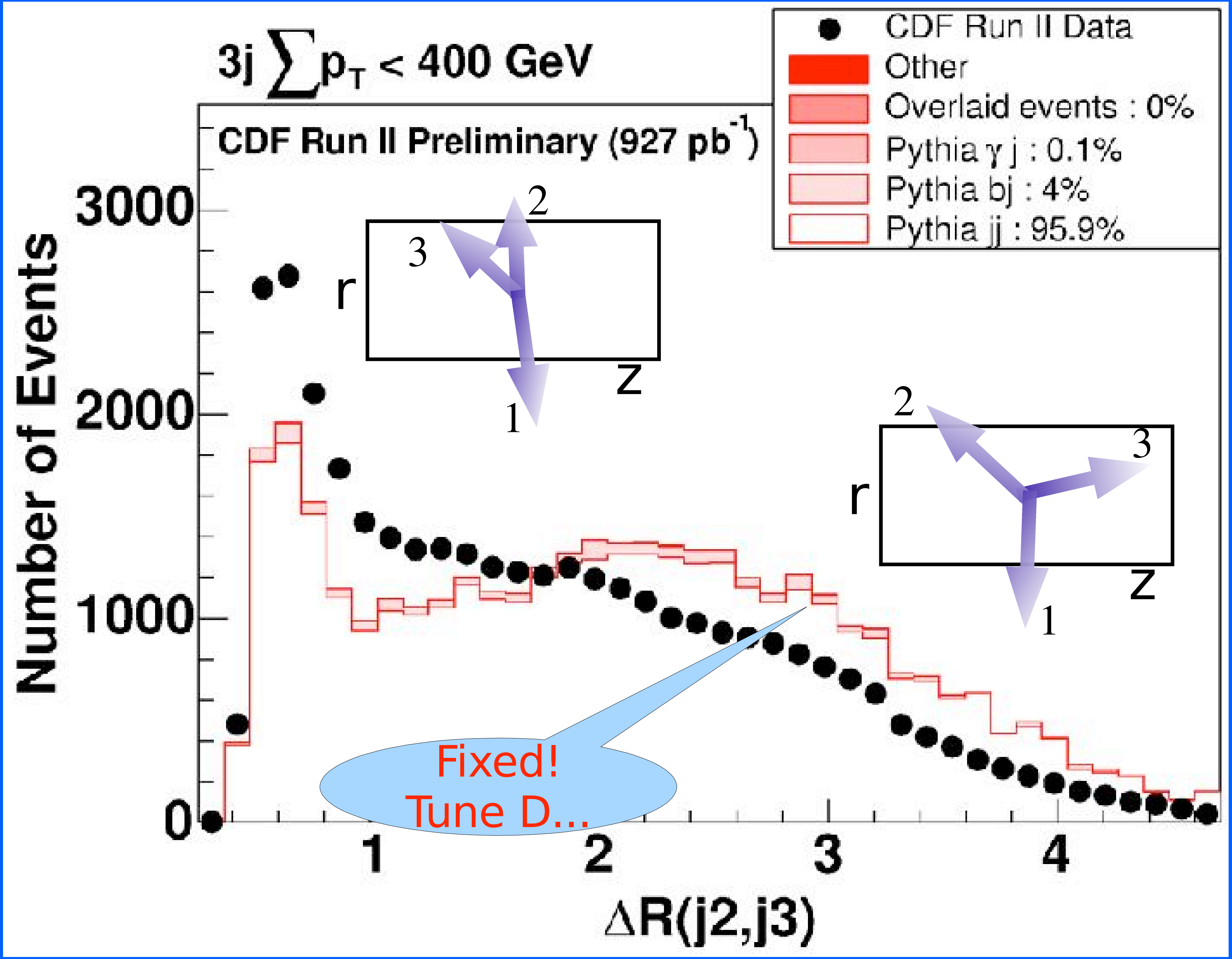
Largest discrepancy is a 2.7sigma deficit

Several % of all distributions disagree:

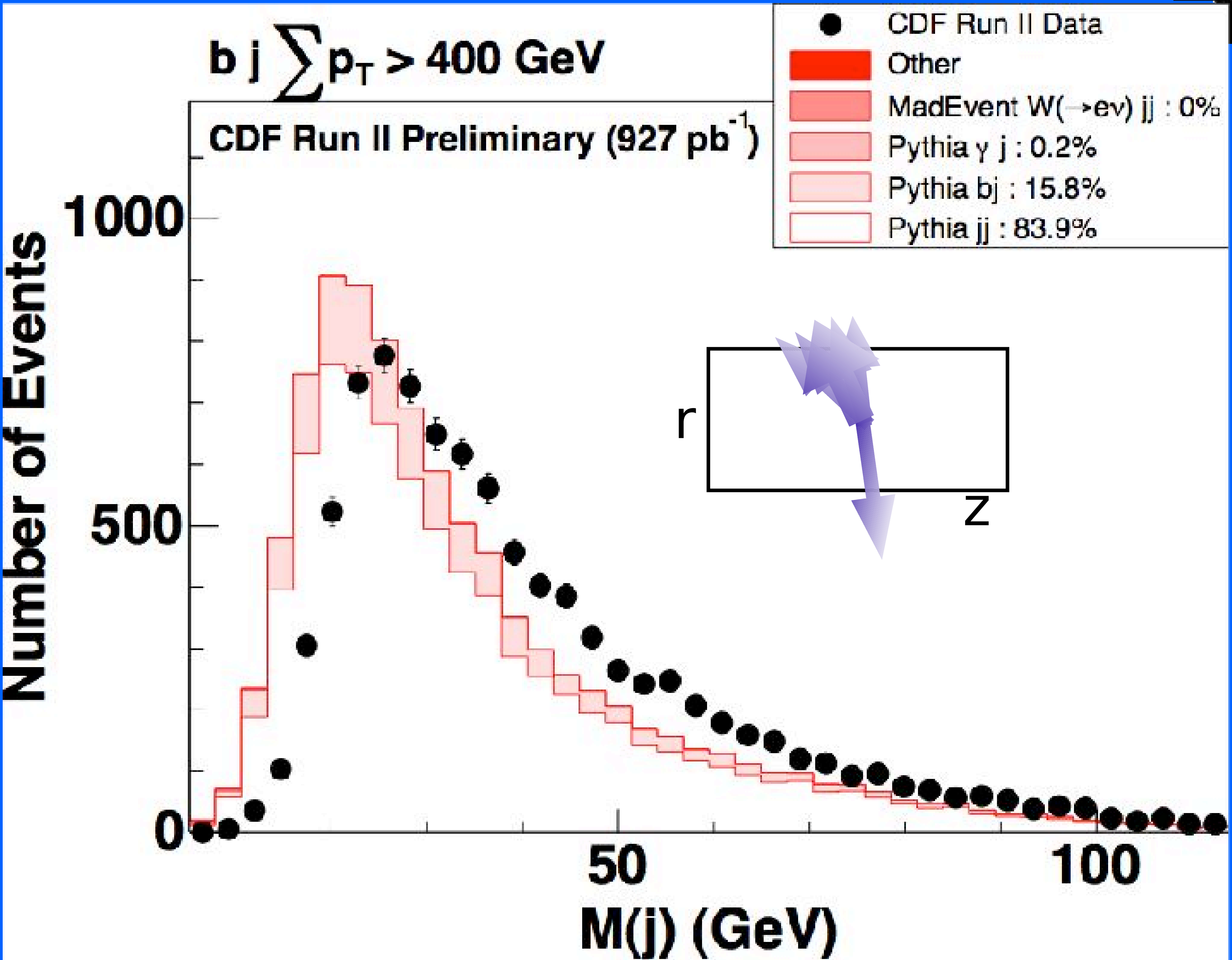
1% is typical of the systematic expected in event generators

about 6% of distributions have KS<1%, but there are many commonalities

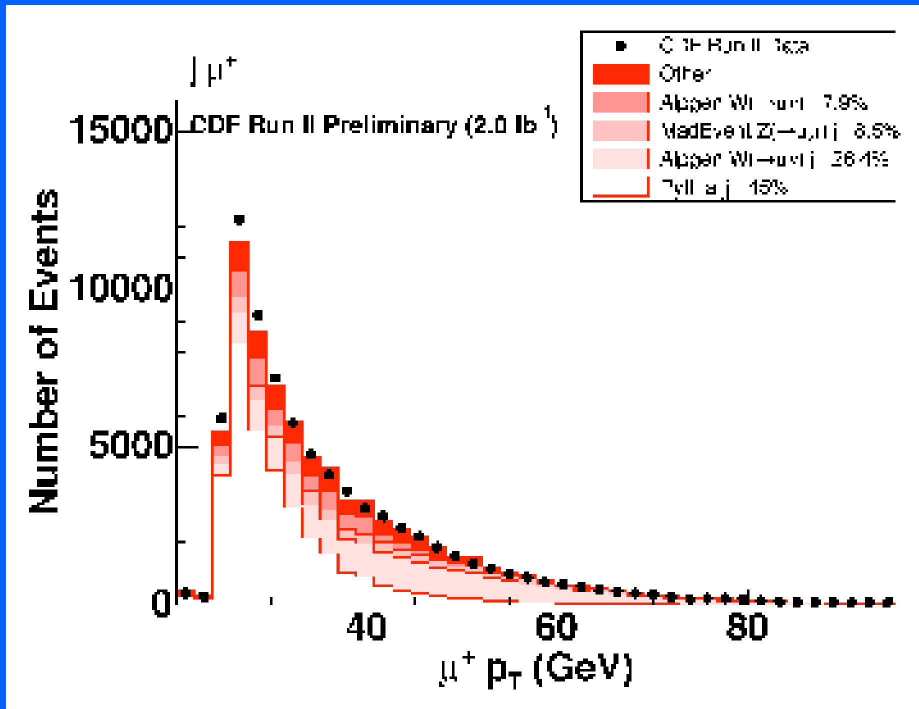
Sample discrepant distribution

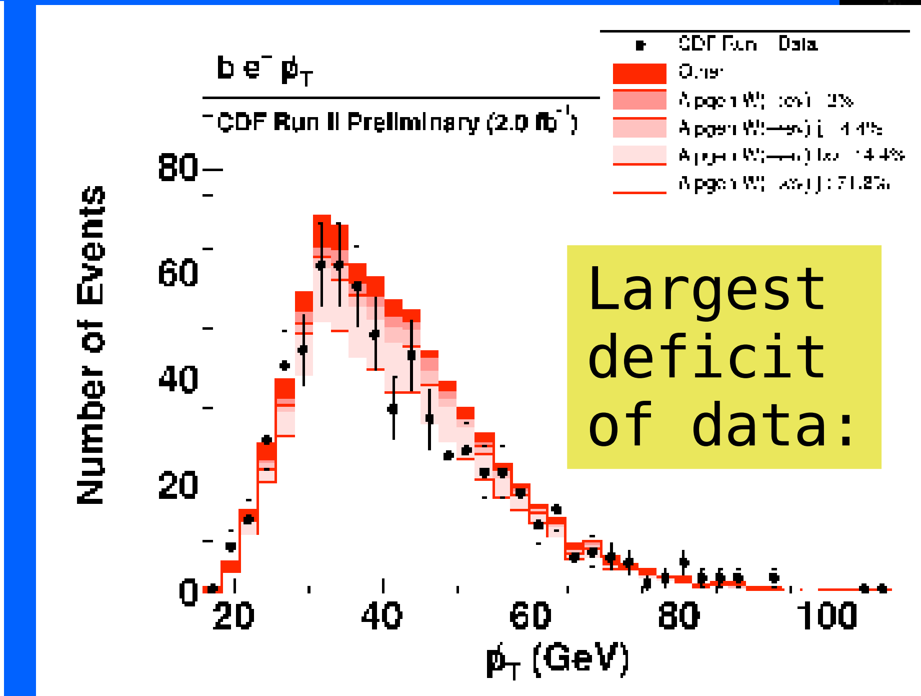


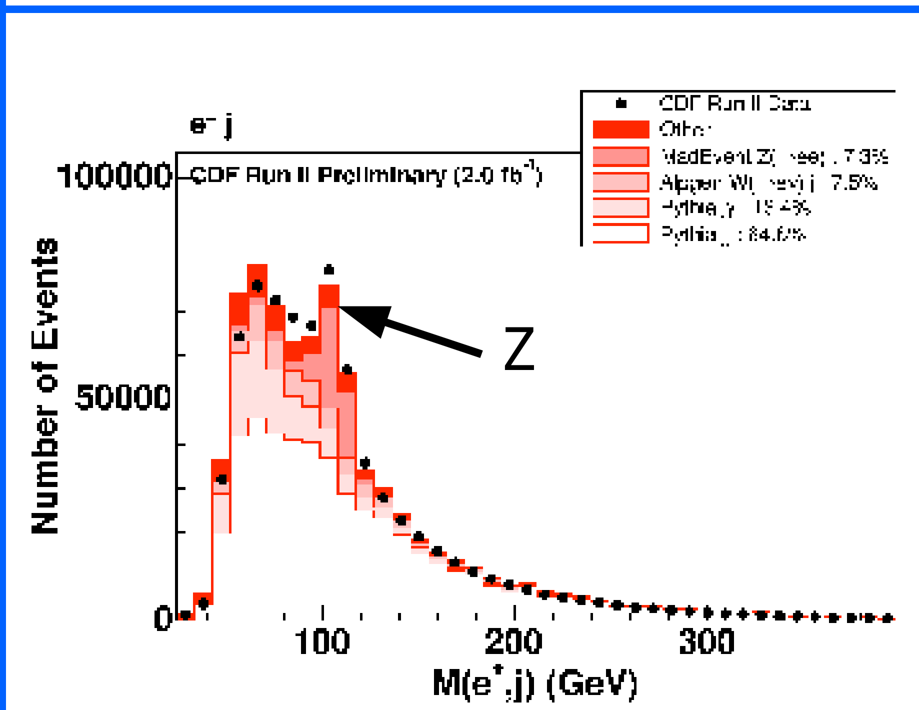
Related discrepant distribution

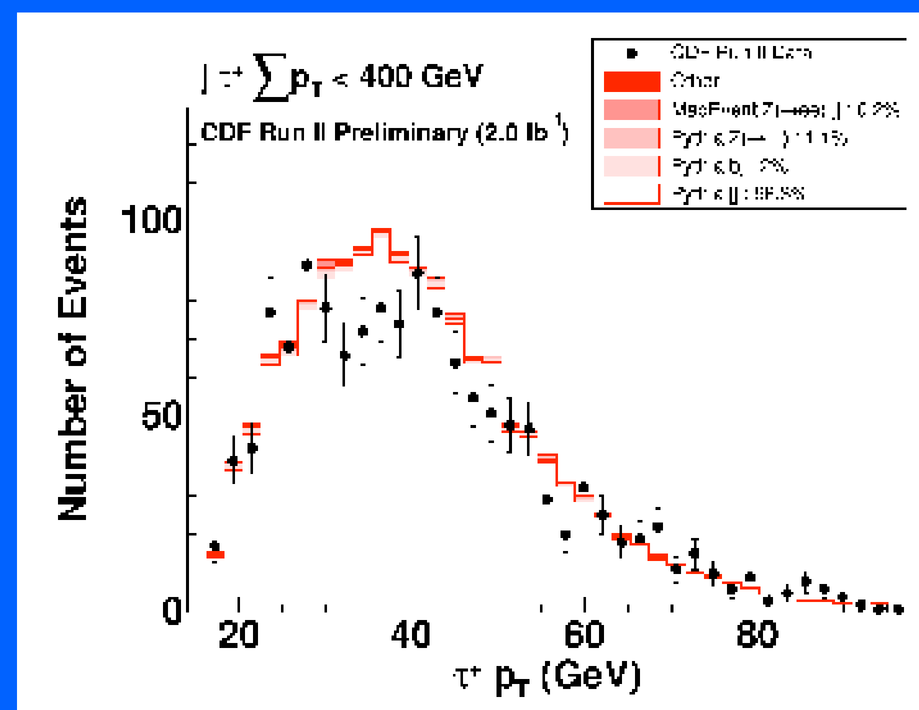


Many things described well!









Dissecting the SM cocktail

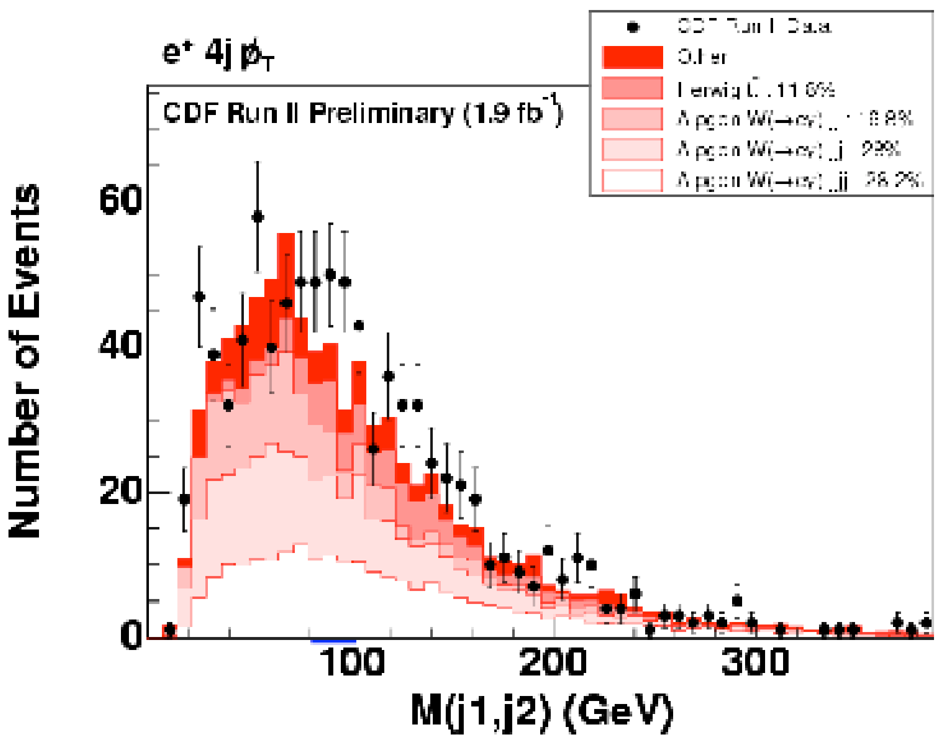
Much of the Monte Carlo is default Pythia/Herwig (simple processes + parton showers)

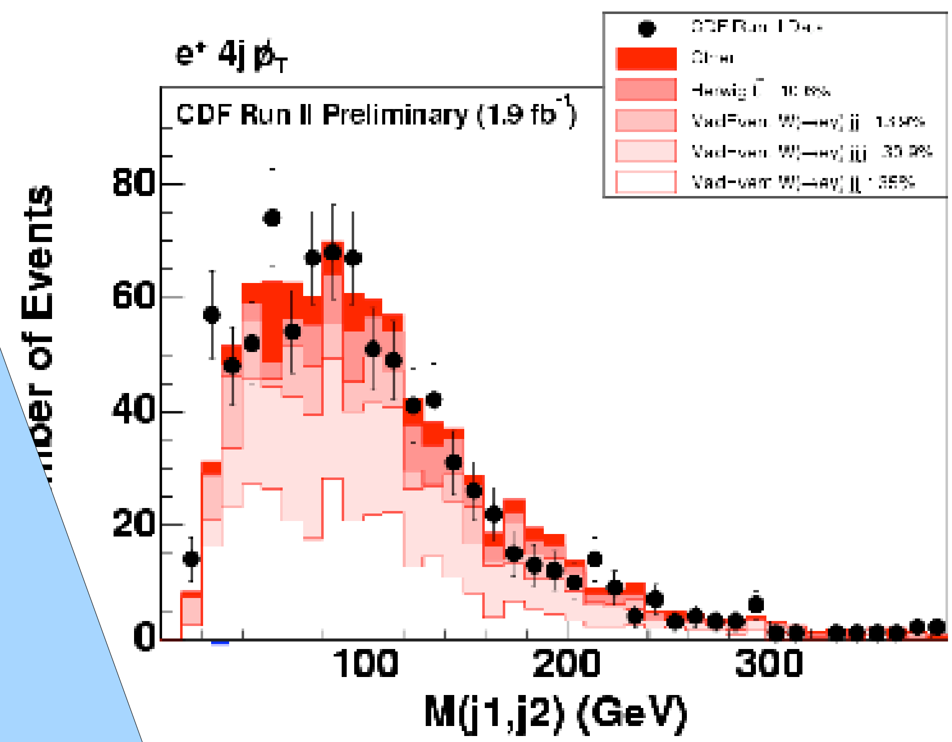
Some processes like W/Z/γ+jets combine Matrix Elements with parton showers

Such calculations are necessary for the LHC

We can remix our cocktail with different implementations of the Standard Model theory

Change W+4j model:Goodness of fit unchanged





MLM matching

SM matching

		Alpgen	SM
k-factor	W0j	1.379	1.452
k-factor	W1j	1.329	1.20
k-factor	W2j	2.007	1.23
k-factor	W3j	2.109	1.18

Relevant K-Factors

K=NLO/LO

6M/6L1

6M/6M

Process	Typical scales		Tevatron K-factor			LHC K-factor		
	μ_0	μ_1	$\mathcal{K}(\mu_0)$	$\mathcal{K}(\mu_1)$	$\mathcal{K}'(\mu_0)$	$\mathcal{K}(\mu_0)$	$\mathcal{K}(\mu_1)$	$\mathcal{K}'(\mu_0)$
W	m_W	$2m_W$	1.33	1.31	1.21	1.15	1.05	1.15
W+1jet	m_W	p_T^{jet}	1.42	1.20	1.43	1.21	1.32	1.42
W+2jets	m_W	p_T^{jet}	1.16	0.91	1.29	0.89	0.88	1.10
WW+jet	m_W	$2m_W$	1.19	1.37	1.26	1.33	1.40	1.42
$t\bar{t}$	m_t	$2m_t$	1.08	1.31	1.24	1.40	1.59	1.48
$t\bar{t}$+1jet	m_t	$2m_t$	1.13	1.43	1.37	0.97	1.29	1.10
$b\bar{b}$	m_b	$2m_b$	1.20	1.21	2.10	0.98	0.84	2.51
Higgs	m_H	p_T^{jet}	2.33	–	2.33	1.72	–	2.32
Higgs via VBF	m_H	p_T^{jet}	1.07	0.97	1.07	1.23	1.34	1.09
Higgs+1jet	m_H	p_T^{jet}	2.02	–	2.13	1.47	–	1.90
Higgs+2jets	m_H	p_T^{jet}	–	–	–	1.15	–	–

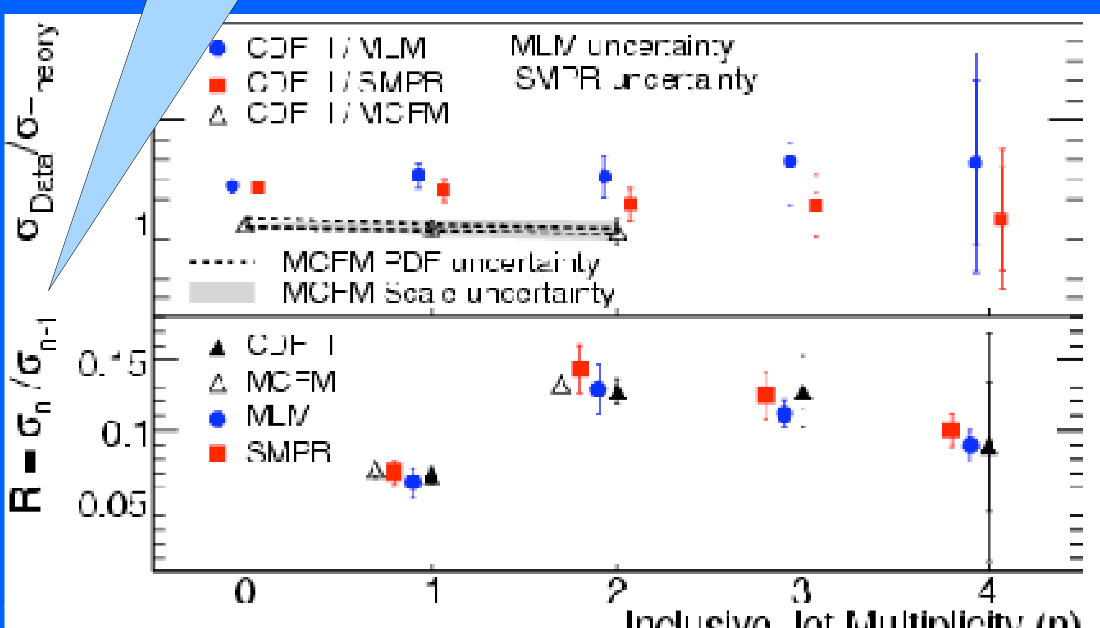
ArXiv:0711.4044 CDF...(SM)

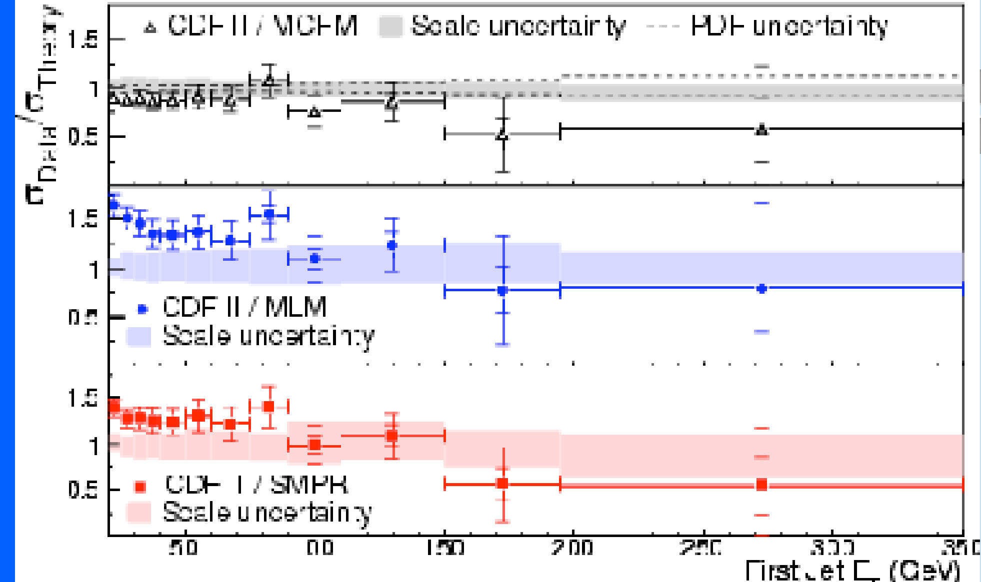
Traditional Analysis

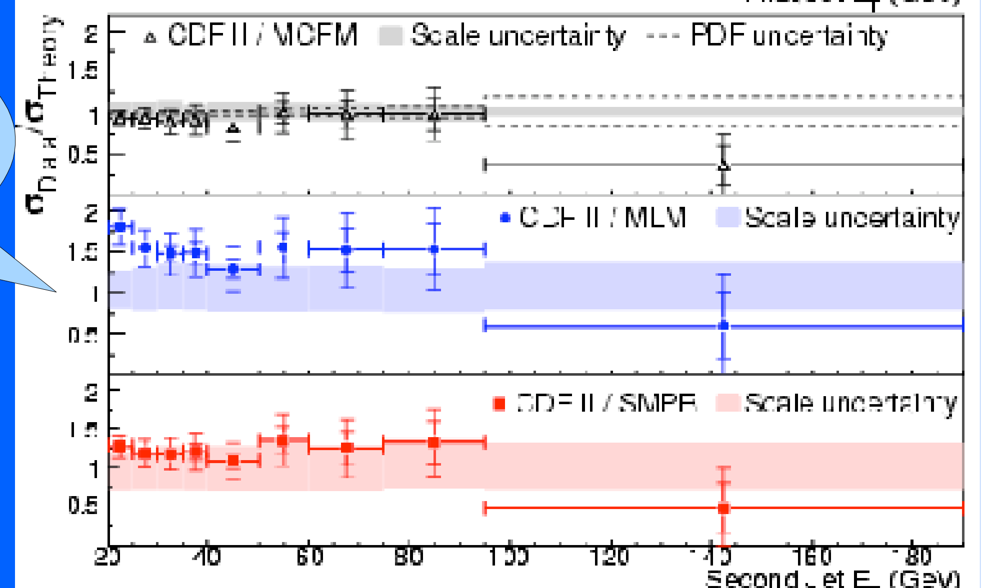
Data corrected (unfolded)
back to the particles
(this it the output of Pythia)

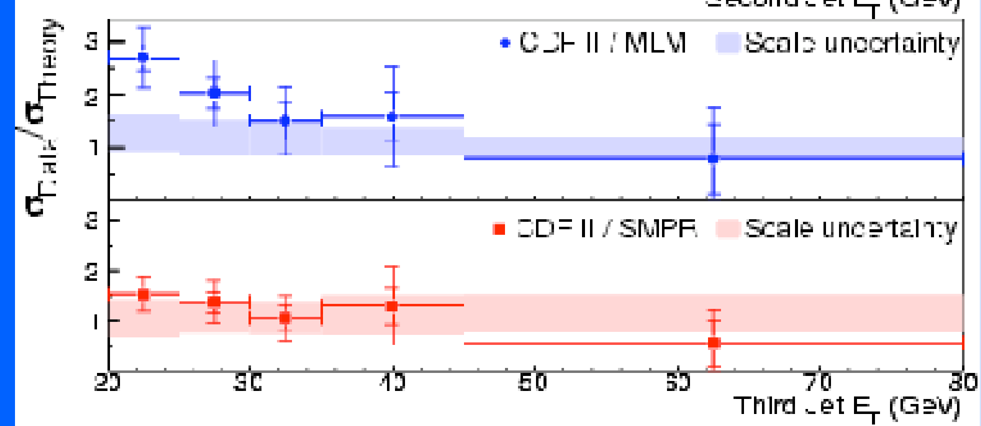
Comparison of relative event counts

Comparison of relative shapes









“ … All distributions show good agreement with the data … “

- Uncovered several other modeling issues, including:
  - Back-to-back jets from unshowered multi-parton interactions
  - Need for consistent tune of W/Z pT with underlying event

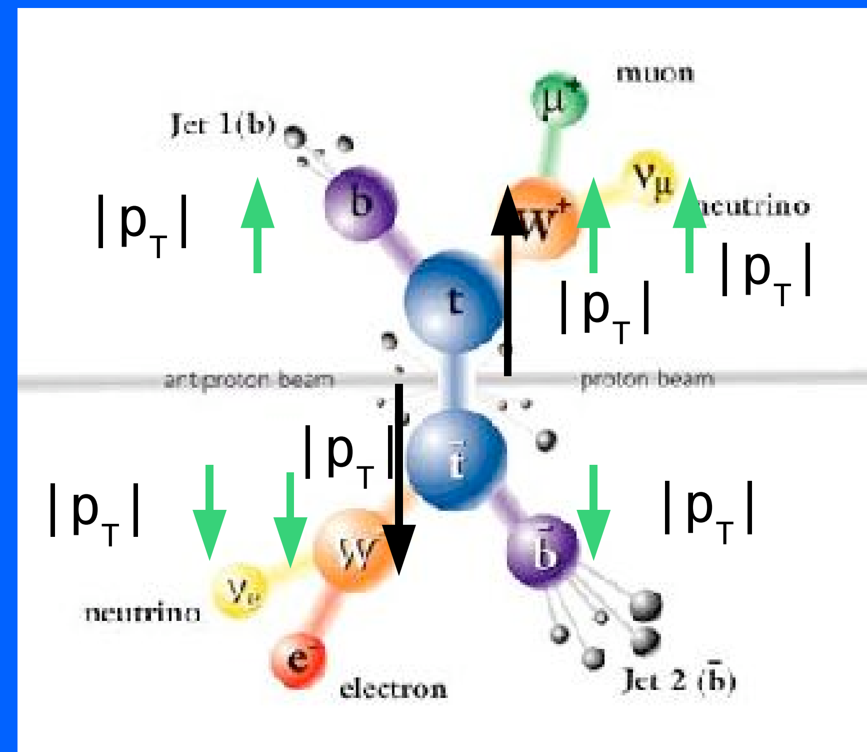
#2: Sleuth, a model independent search strategy for new physics

Exclusive final states

Large $\sum|p_T|$

An excess

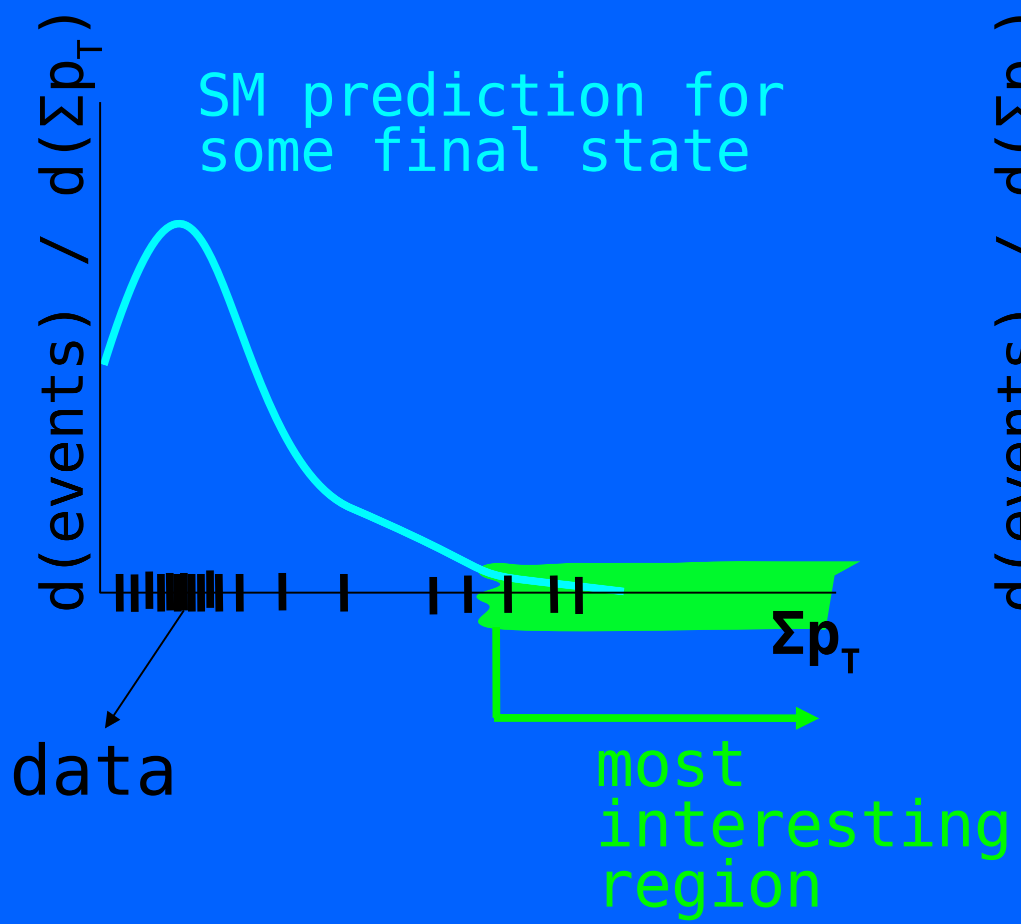
Rigorously compute the trials factor associated with looking everywhere

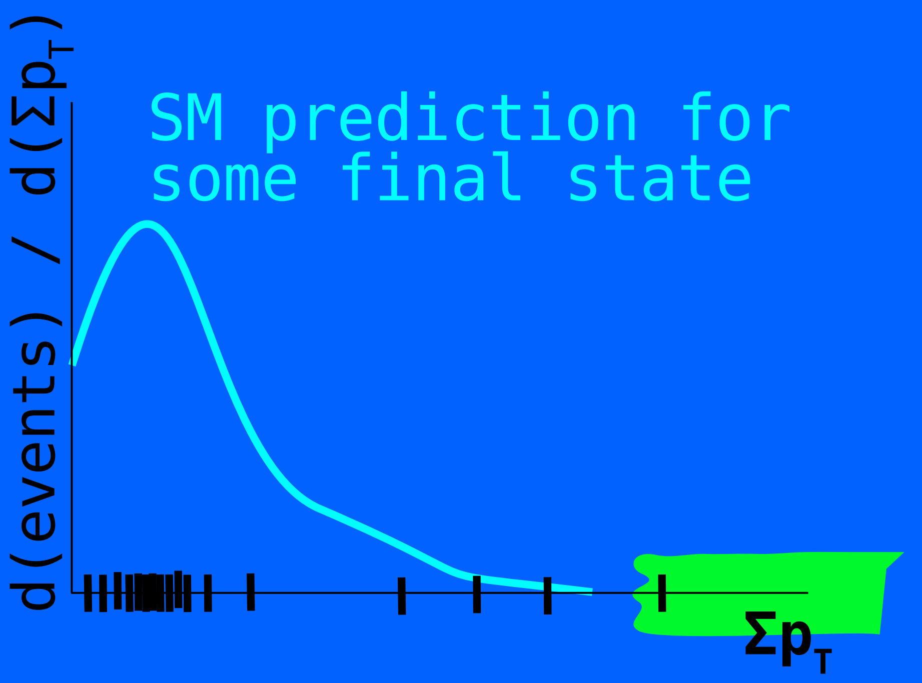


$$\int_{0001001}^{today} d(hep-ph)(prediction)$$

Goal of Sleuth

Identify statistically significant excess of data in the high-Σp_T tails.





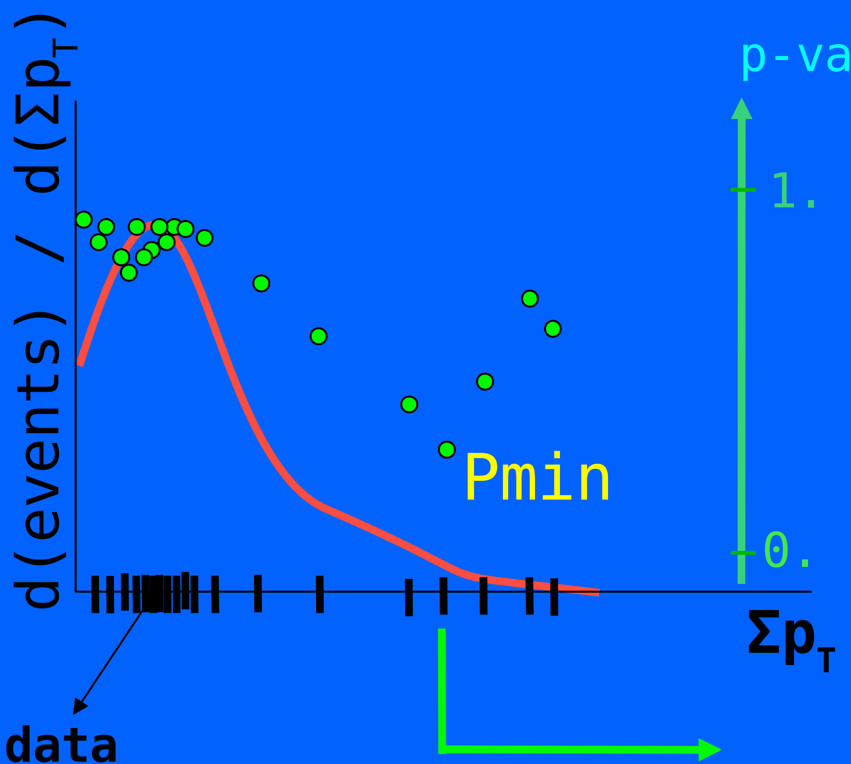
Pmin calculation

● : p-value

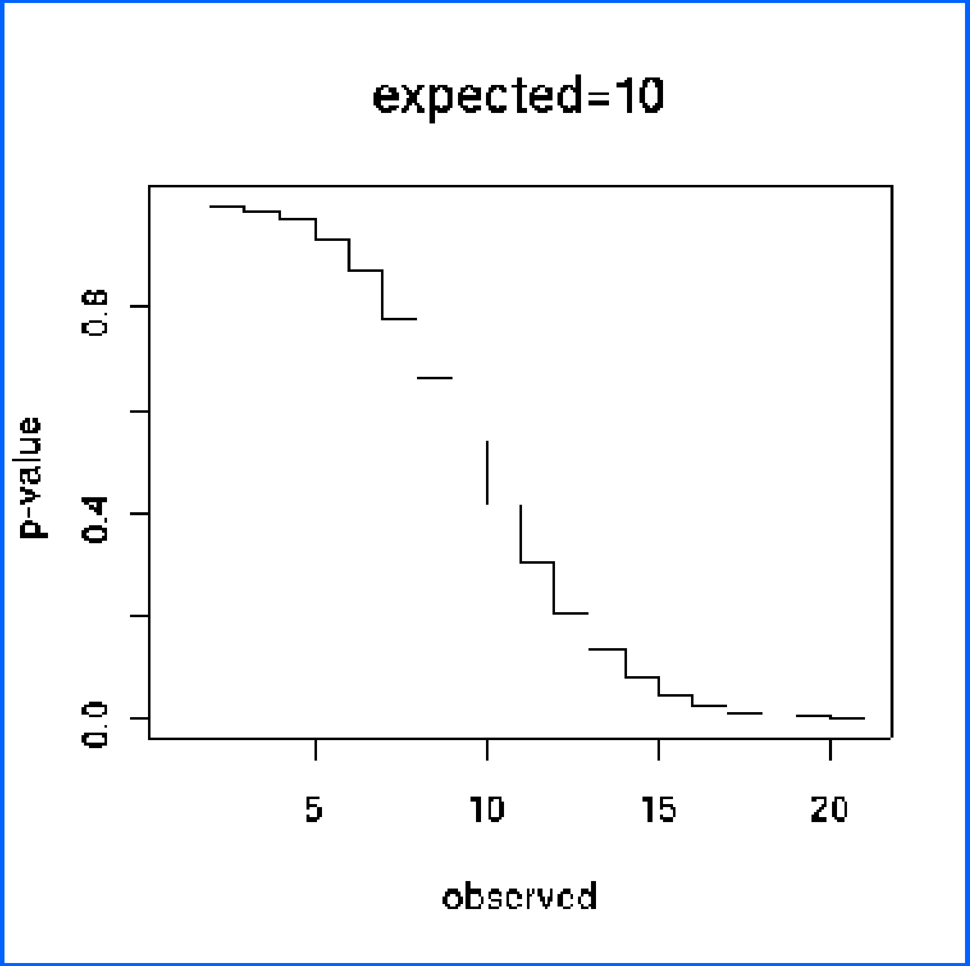
p-value ==

Prob(>=o|b)

$$\sum_{\nu=d}^{\infty} \frac{b^{\nu}}{\nu!} \exp(-b)$$



The most interesting region, characterized by Pmin.



Pmin → scriptP

- How unusual is this Pmin?
- Generate pseudo-data to see how often this (or something more interesting) would happen.
  - Fraction == scriptP
- Smaller scriptP → more interesting

Trials Factor

Each final state → scriptP

N final states:

tildeScriptP = 1 – (1-min{scriptP})^{N}

Prob. fluctuation in *any region in any final state* is as or more interesting

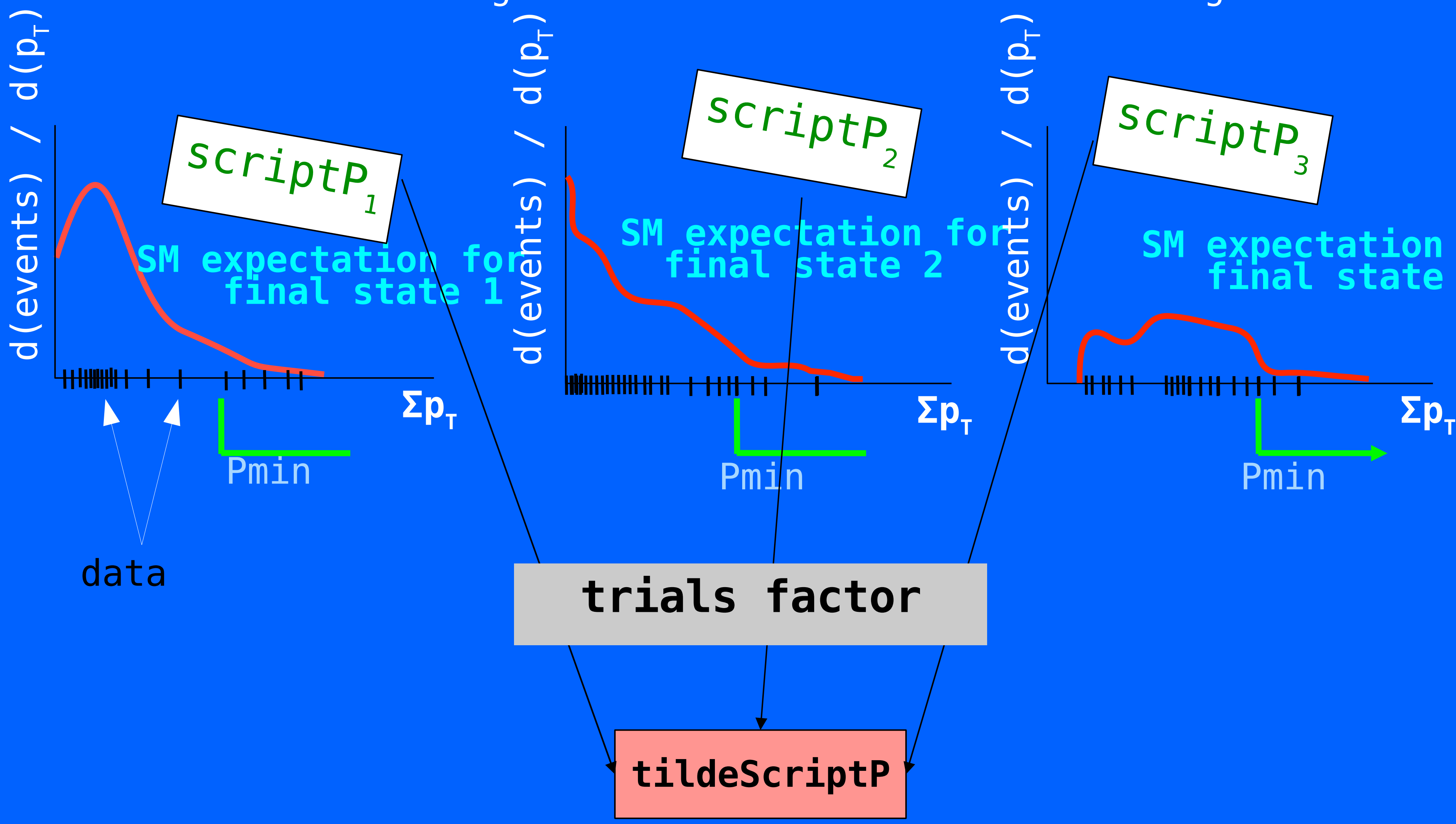
scriptP → all regions

tildeScriptP → all final states

tildeScriptP < 0.001 → >3 sigma effect

Recap: Sleuth Algorithm

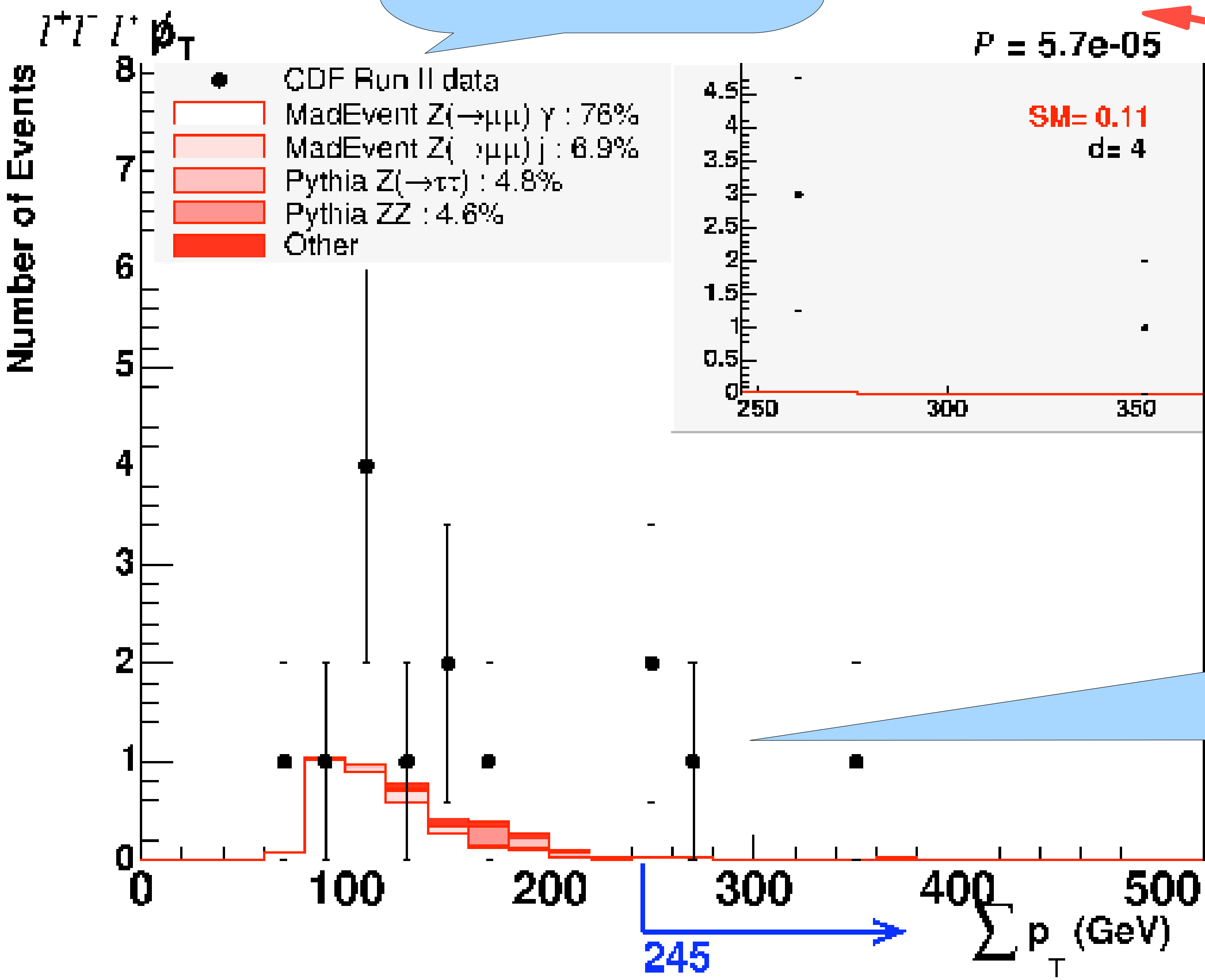
ScriptP = % of pseudo-experiments where this final state has any tail more interesting than the actual most interesting one.



TildeScriptP = % of pseudo-experiments that would produce any tail in any final state, that would be more interesting than *the* most interesting tail actually observed.

W+Z removed

ScriptTildeP =.01 =2.6sigma



Vista “discovery” of W+Z

Sleuth @CDFII result

$$\tilde{\mathcal{P}} = 0.08$$

8% of pseudo-experiments should be as interesting

(top 5)

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

SLEUTH Final State	$\mathcal{P}$
$\ell^+\ell'^+$	0.00055
$\ell^+\ell'^+ \not{p} jj$	0.0021
$\ell^+\ell'^+ \not{p}$	0.0042
$\ell^+\ell^-\ell' \not{p}$	0.0047
$\ell^+\tau^+ \not{p}$	0.0065

No significant excess

This does not prove no new physics!

Sleuth @CDFIIa result

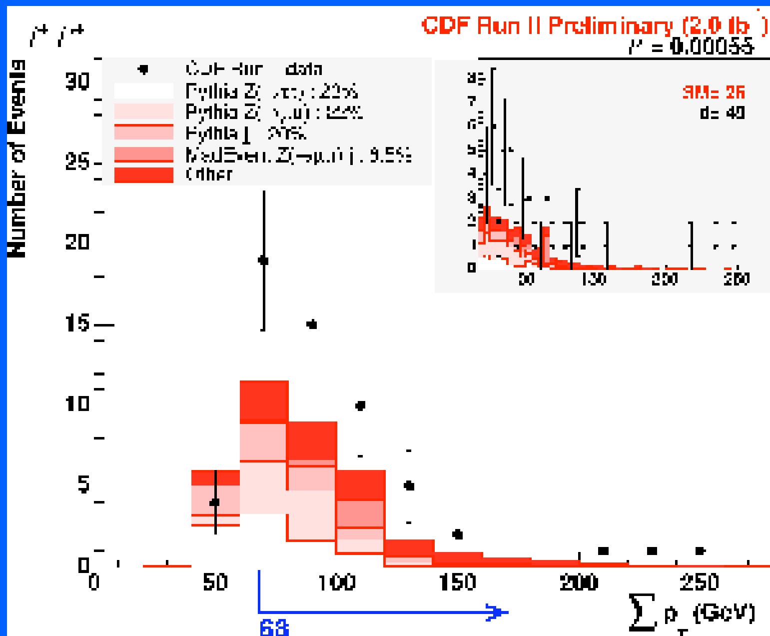
$$\tilde{\mathcal{P}} = 0.46$$

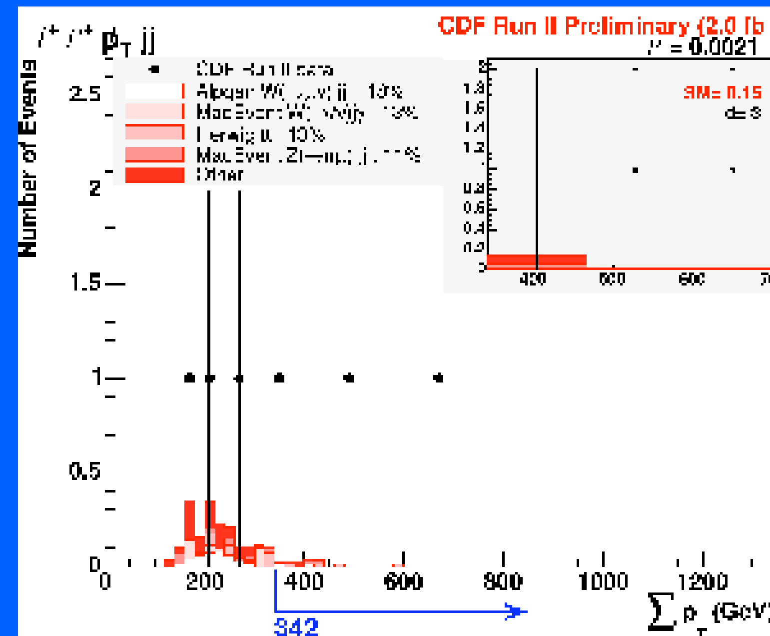
Sleuth Final State	$\mathcal{P}$
$b\bar{b}$	0.0055
$j \not{p}$	0.0092
$\ell^+ \ell'^+ \not{p} j j$	0.011
$\ell^+ \ell'^+ \not{p}$	0.016
$\tau \not{p}$	0.016

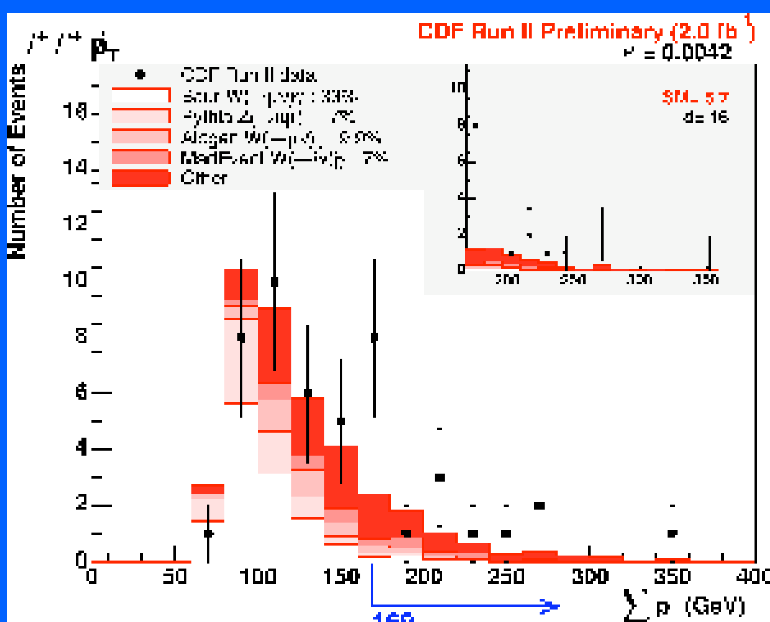
CDF Run II Preliminary (2.0 fb^{-1})

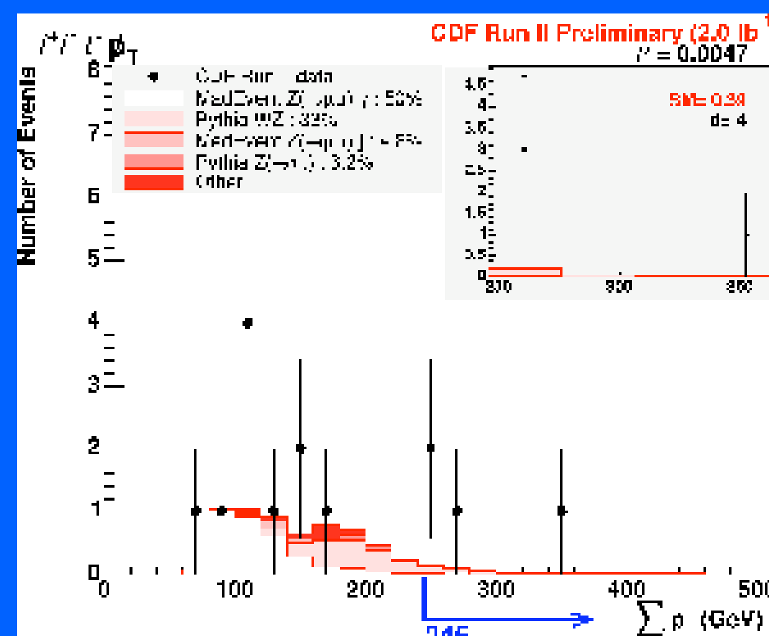
Sleuth Final State	$\mathcal{P}$
$\ell^+ \ell'^+$	0.00055
$\ell^+ \ell'^+ \not{p} j j$	0.0021
$\ell^+ \ell'^+ \not{p}$	0.0042
$\ell^+ \ell^- \ell' \not{p}$	0.0047
$\ell^+ \tau^+ \not{p}$	0.0065

http://www-cdf.fnal.gov/physics/exotic/r2a/20080228.vista_sleuth/publicPage.html









The greatest limitation to this blind new physics search is mis-modeling of backgrounds

Note: this analysis does NOT incorporate PDF, showering uncertainties:

these are “fit” using correlations between different final states (e.g. K-factors from data)

What can we expect at the LHC?

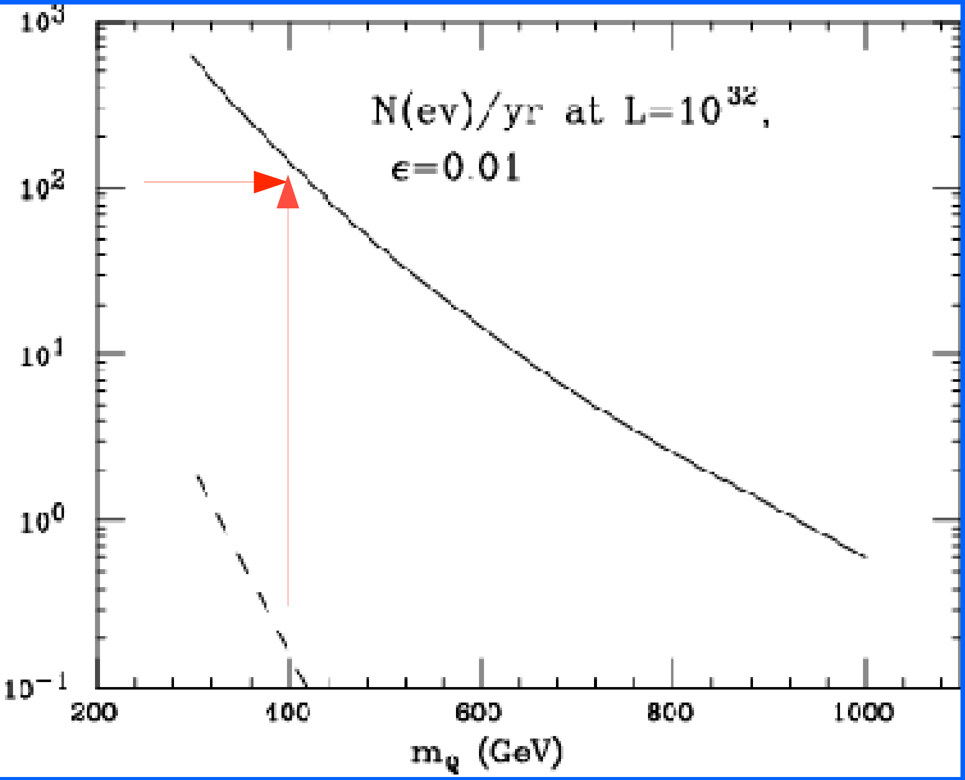
Can we understand it?

Greatest Concerns in TeV → LHC Extrapolation

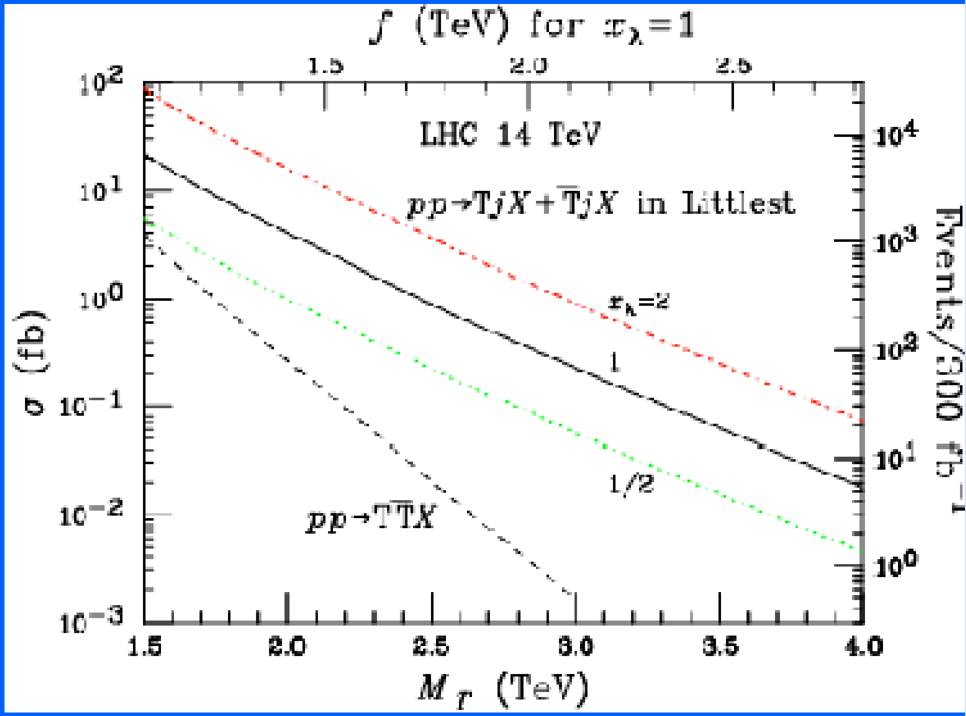
- Exploring new kinematic regimes
  - Not so much an issue, except UE, small x
- Complicated topologies
- Studying gluons instead of quarks

Heavy Quark Production @ LHC

Huge phase space in an interesting kinematic region

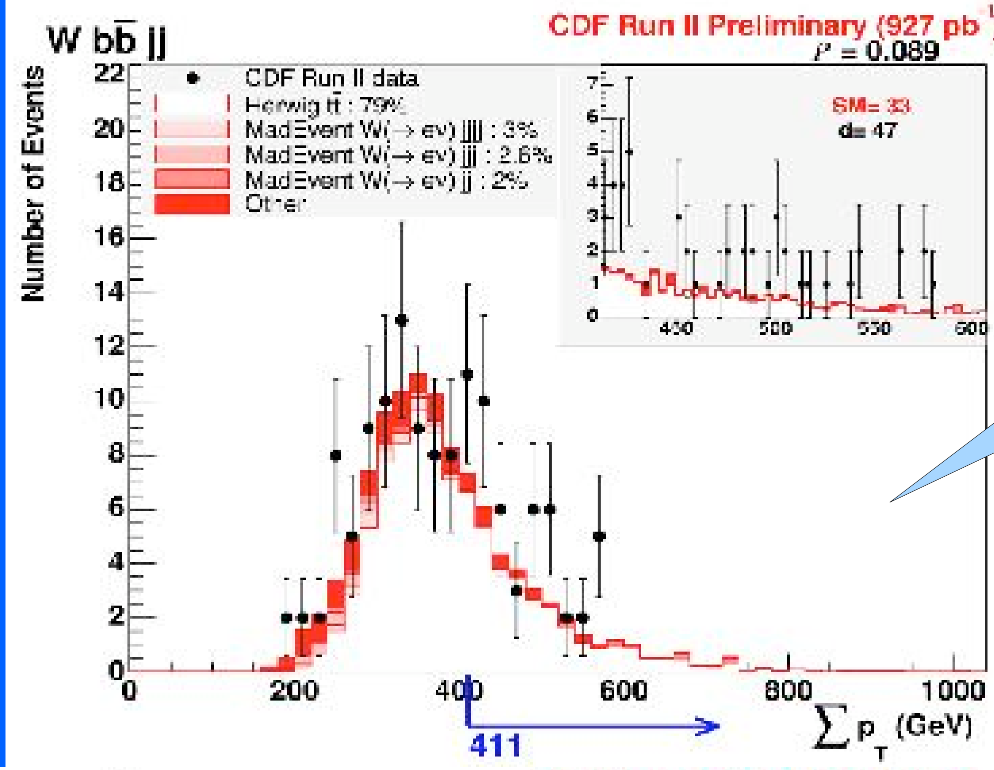


MLM

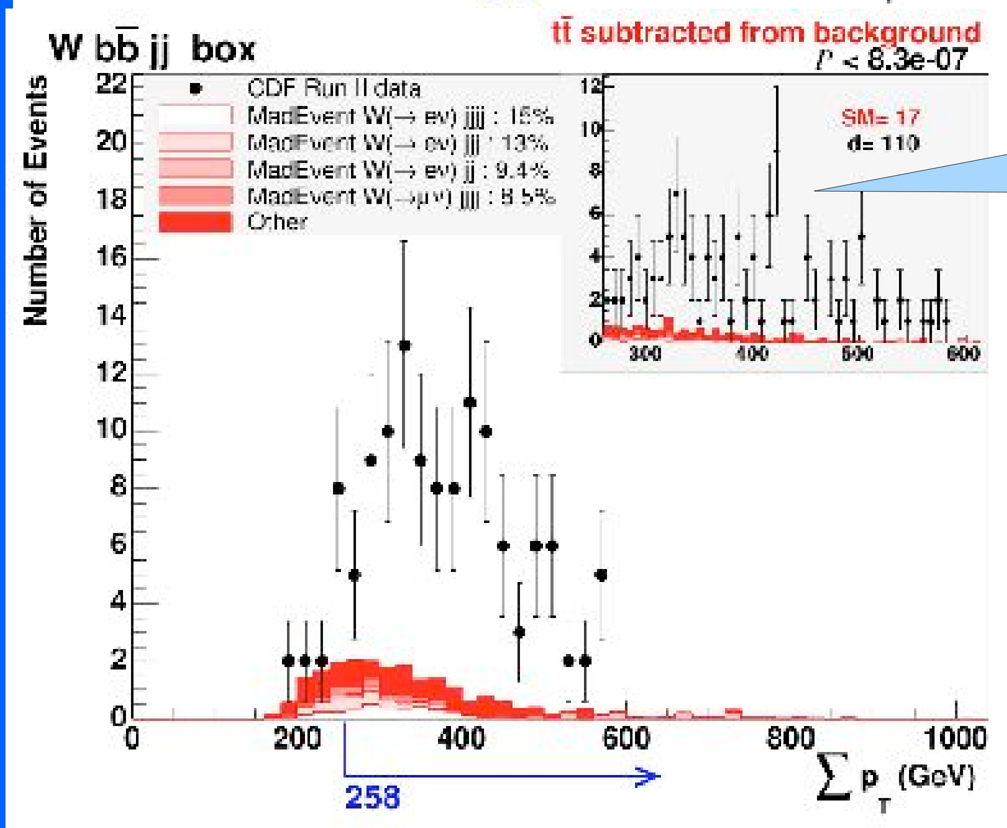


Logan, Han, Wang

Something new can appear very quickly



Would Sleuth find the top quark?



Yes in 80 pb-1
vs Run1: 67 pb-1

Possible LHC Outcomes

Something so striking you can't miss it $Z^{'} \rightarrow \mu^{+}\mu^{-}$ $BH \rightarrow 100\ \mathrm{Z/W/t/h}$	~100 GeV particles with cascade decays
New exotica (quirks, hidden valley,...)	Nothing (except marginal WW scattering)

Consequences

Easy Use sideband data as your "Monte Carlo" (probably something else to complete the picture)	**Challenging** (Control regions are all mixed up)
More Challenging Requires detailed understanding of SM (and detector) tails	**Most Challenging** When do you give up?

Conclusions

At the Tevatron, we have qualitative AND quantitative measures of how well we understand the SM.

Raises confidence for the challenging case @LHC.
We have a methodology for understanding the data.
We can improve our current tools with manpower and mindpower.

For early tests, data-driven Monte Carlo tools should be sufficient. But now is a good time to start worrying about higher orders and more logs

At the Tevatron, the story continues ...

D0 Vista/Sleuth with 1.07 fb^-1

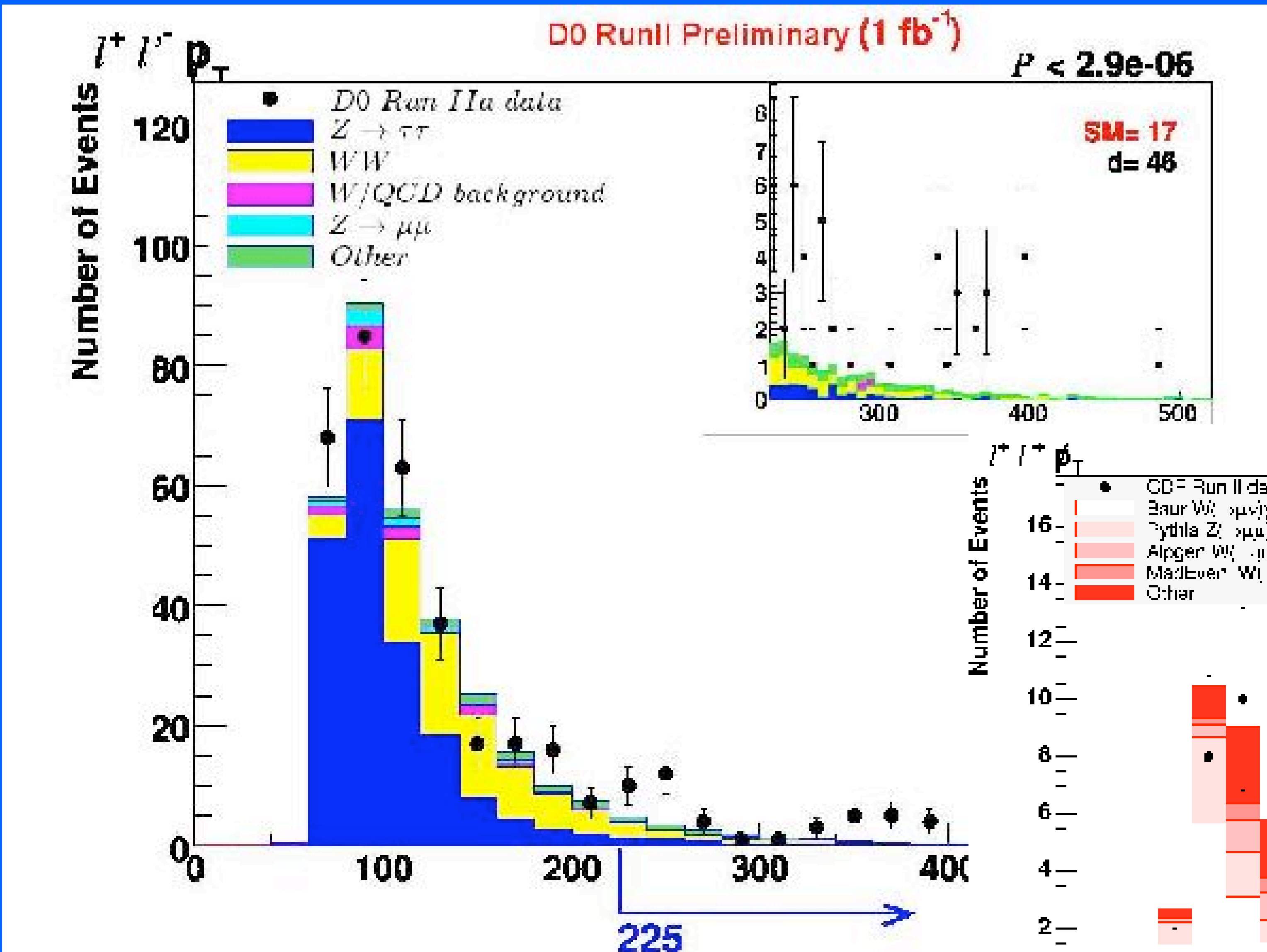
http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/NP/N65/N65.pdf

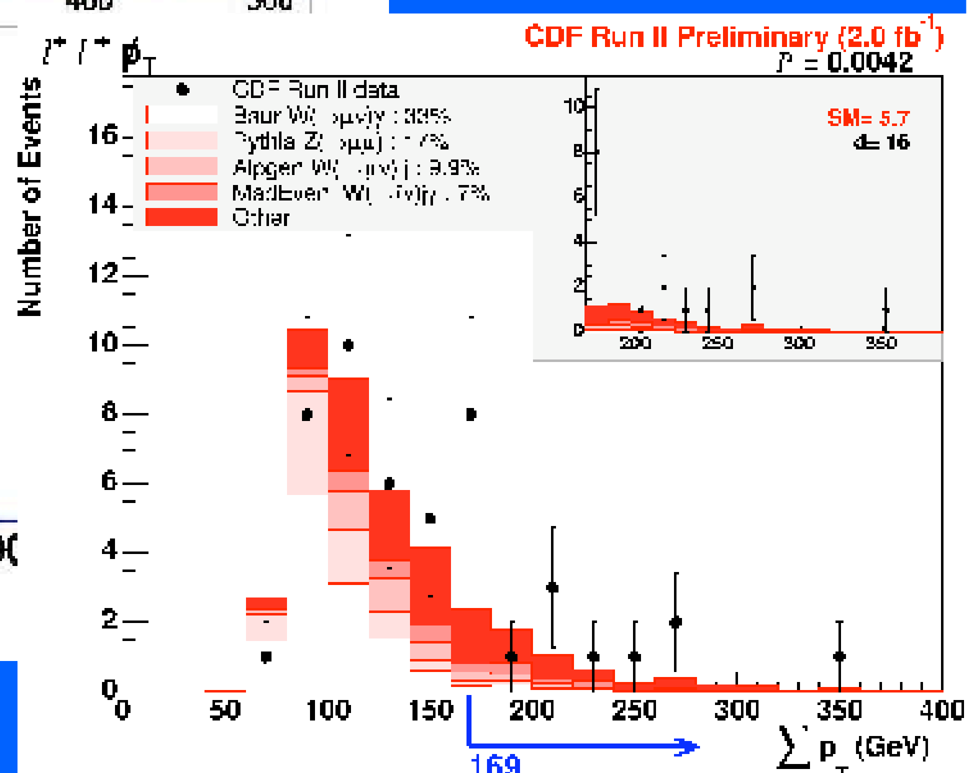
- Focus on final states with leptons
- Use 7 non-overlapping final states to set cocktail normalization and apply standard corrections
- Merge these “normalized” samples as input to Vista
  - Discrepancies in 1mu2jpmiss(9.3σ), 1mu1ph1jpmiss(6.6σ),1mu+1mu-pmiss (4.4σ), 1mu+1mu-1ph(4.1σ)

D0 Vista/Sleuth (cont)

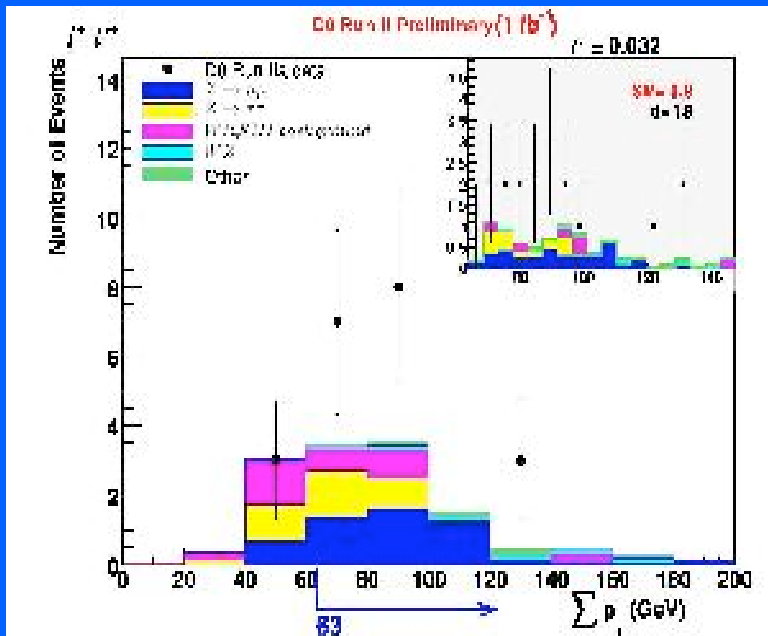
- 2 related to oversimplification of photon fake modeling
- Others related to muon triggering for |eta|>1

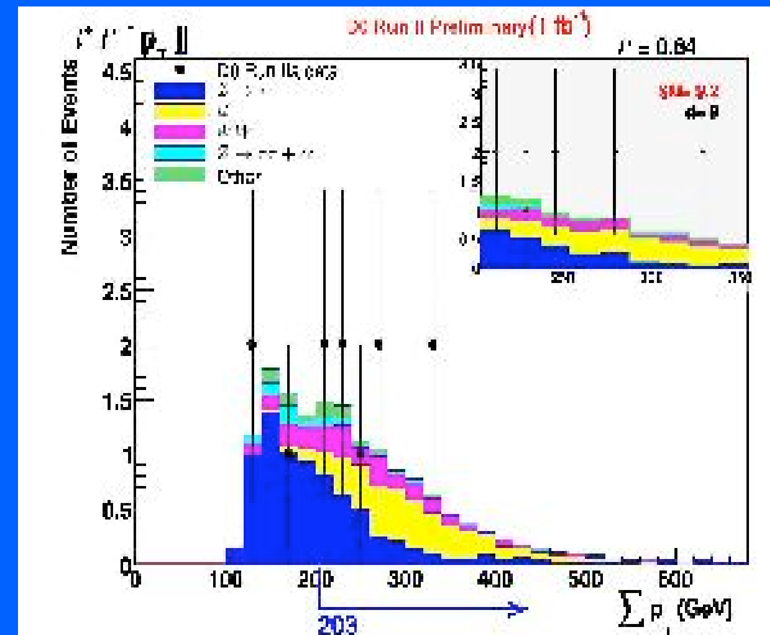
Most significant(*) discrepancy

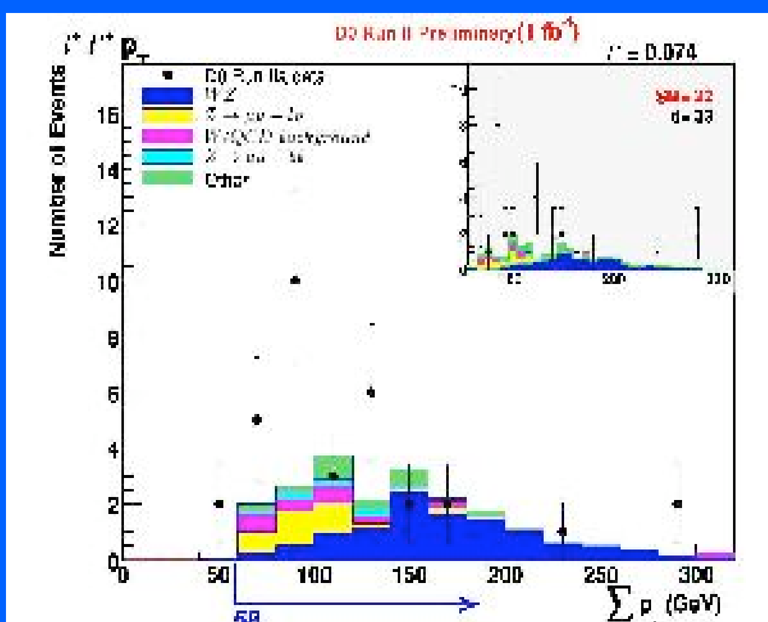


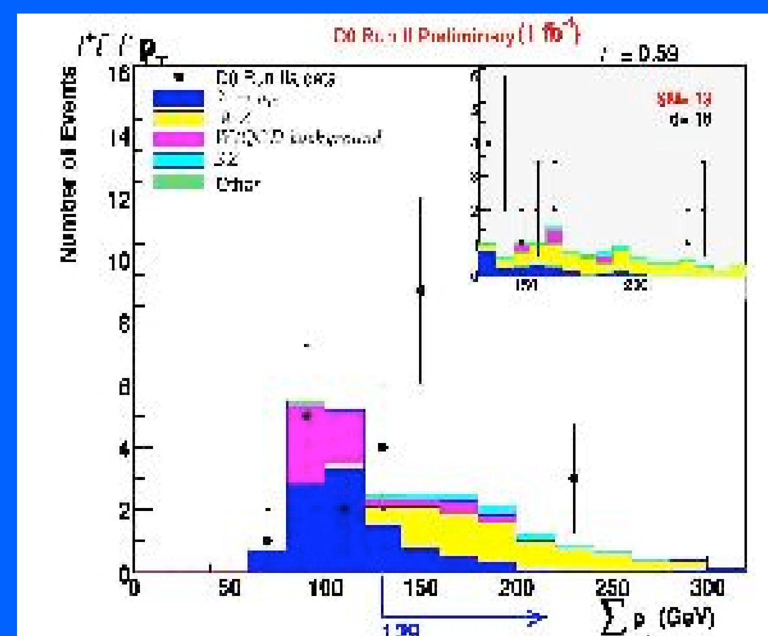


Comparison with CDF Sleuth

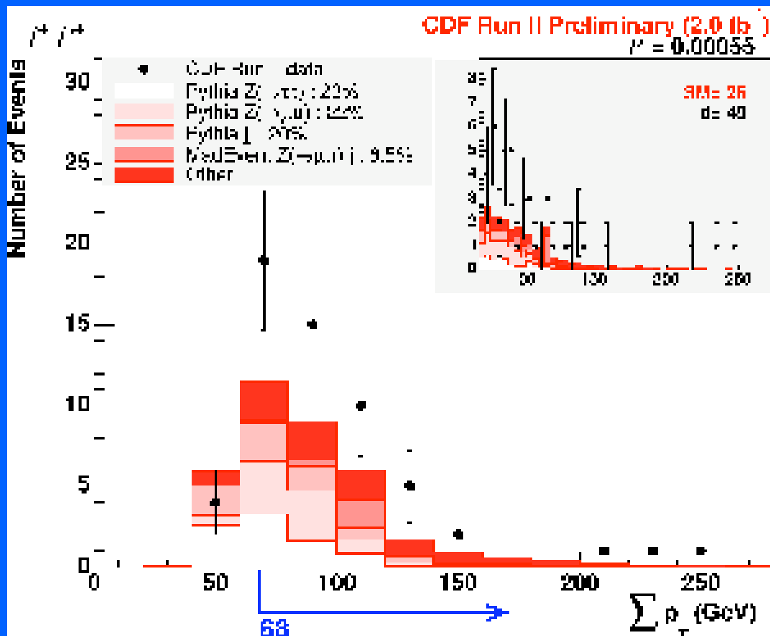


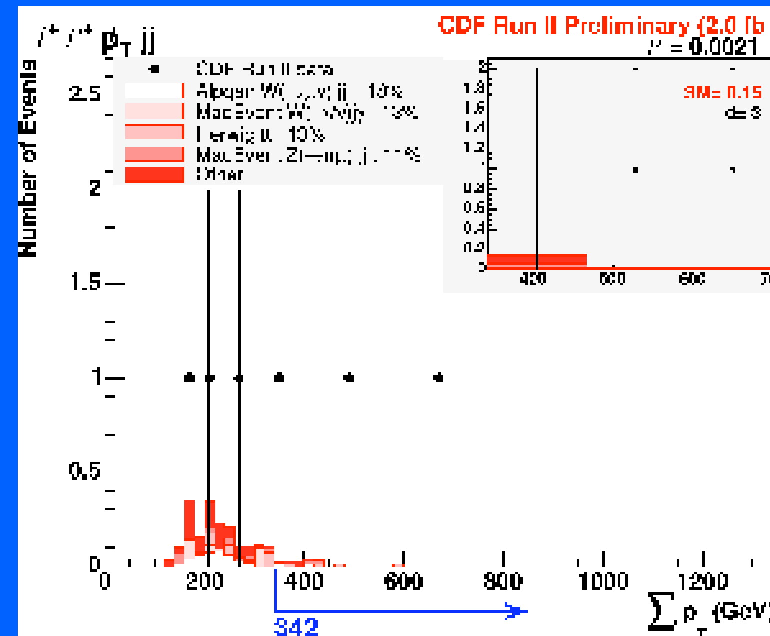


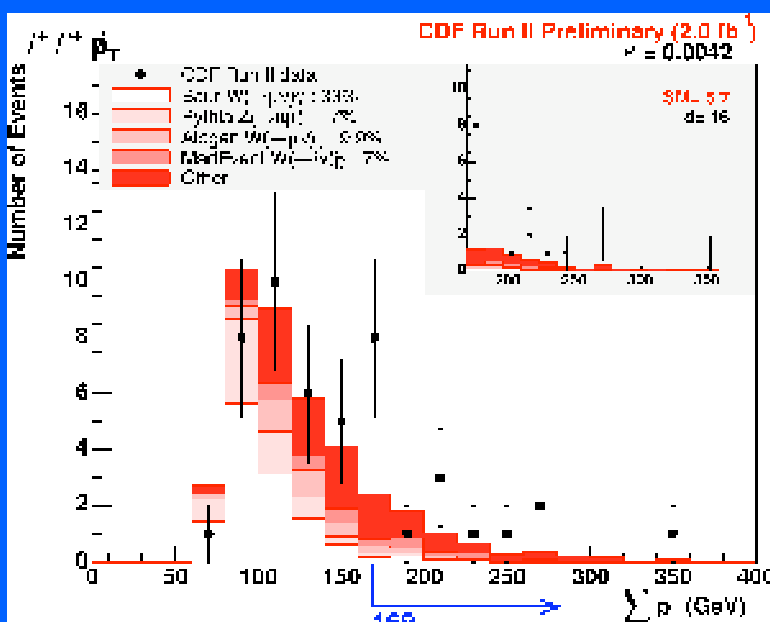


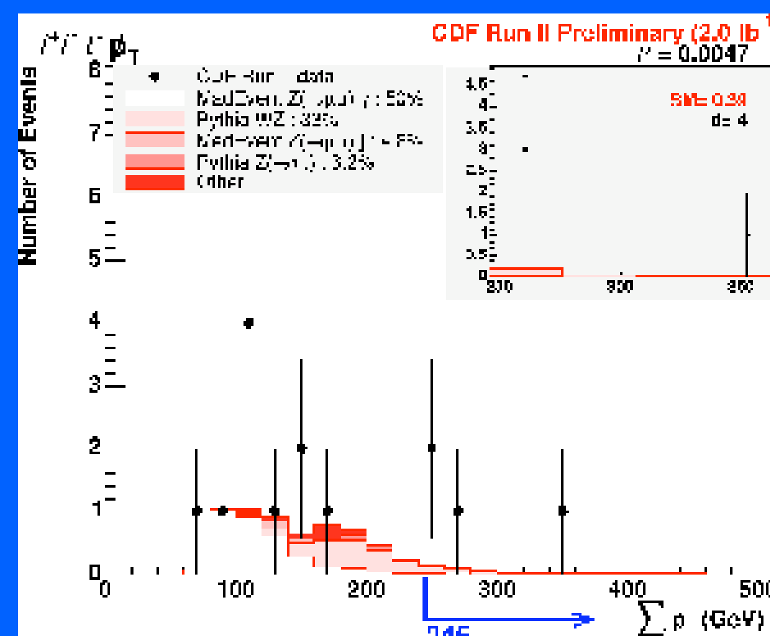


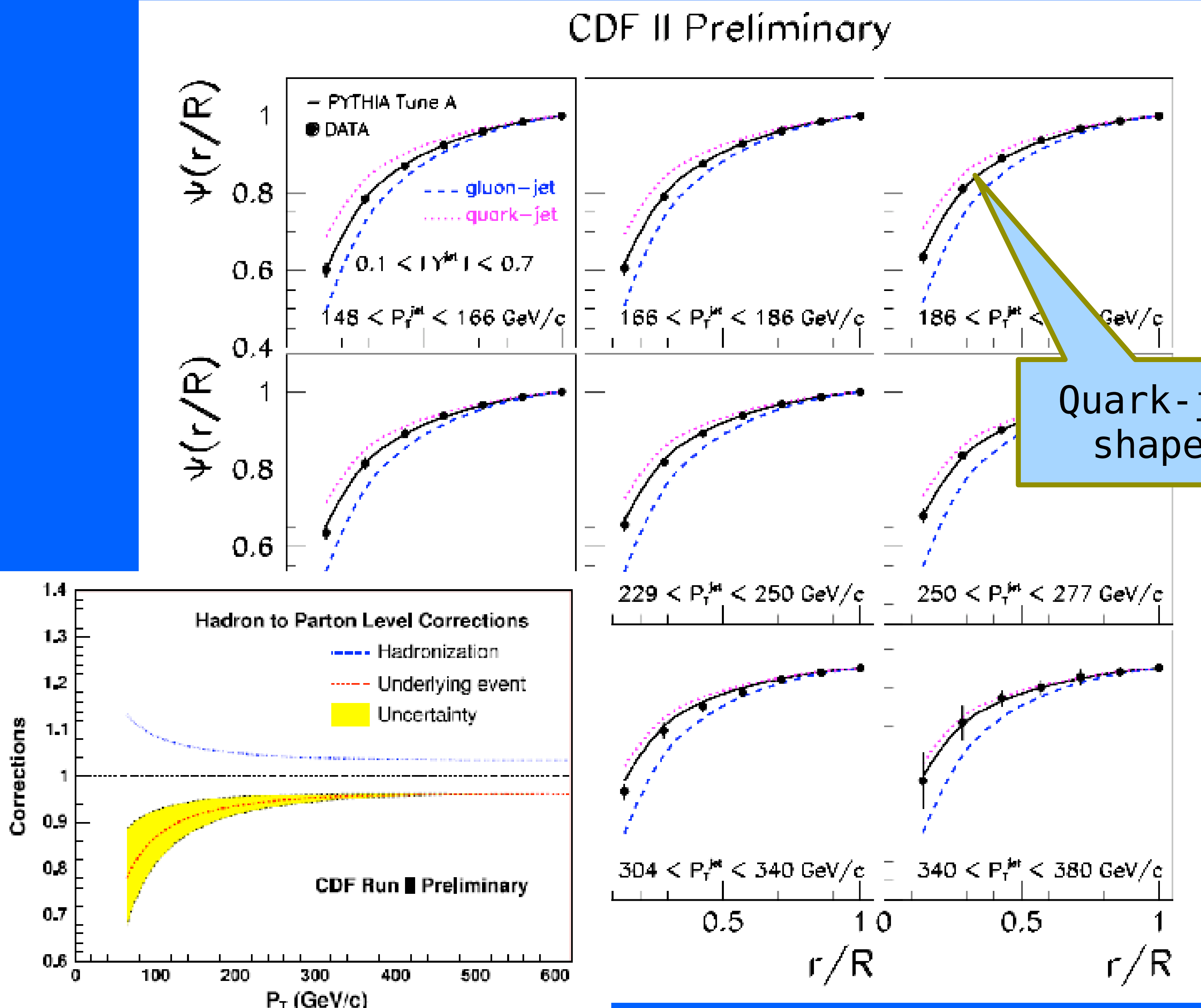
http://www-cdf.fnal.gov/physics/exotic/r2a/20080228.vista_sleuth/publicPage.html





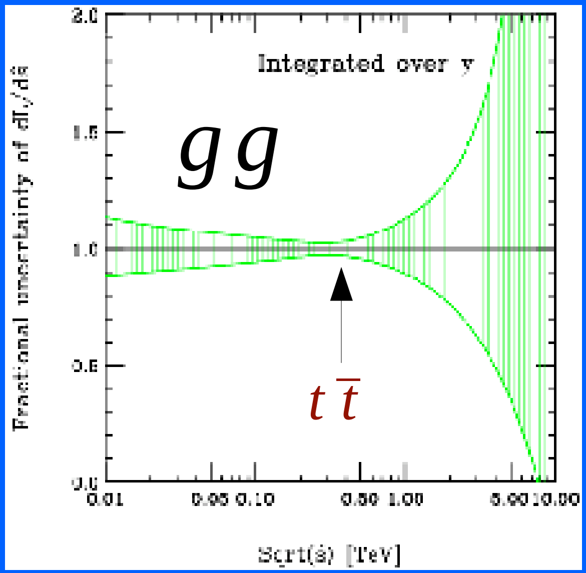






Quark-jet shapes

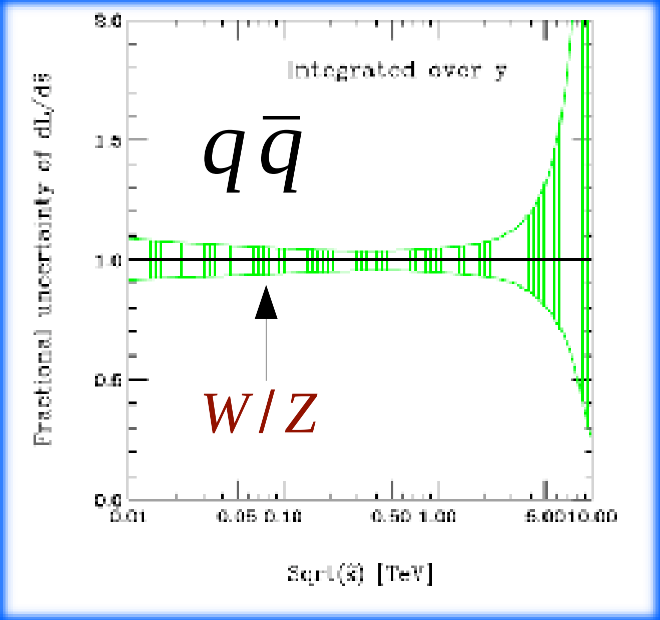
PDF uncertainties at the LHC



Under 1 TeV, PDF lumi known to 10%

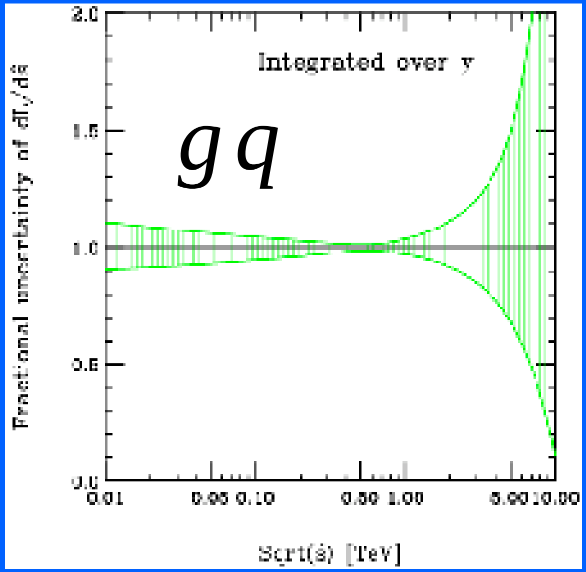
Need similar precision in theory calculations

Limits when LHC data will impact PDF fits



Pdf uncertainties for W/Z cross sections are not the smallest

Top uncertainty is of the same order as W/Z production



Large PDF uncertainty:
- final state likely not well-studied

Small PDF uncertainty:
- initial state known, but not necessarily final state

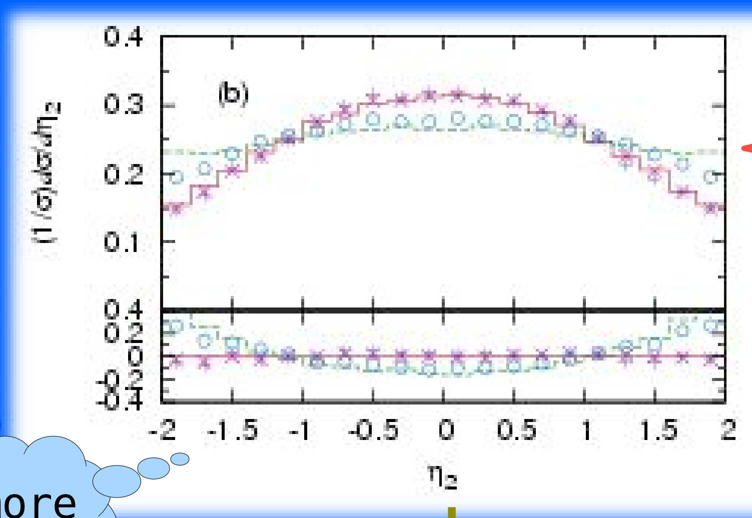
Topological overlap

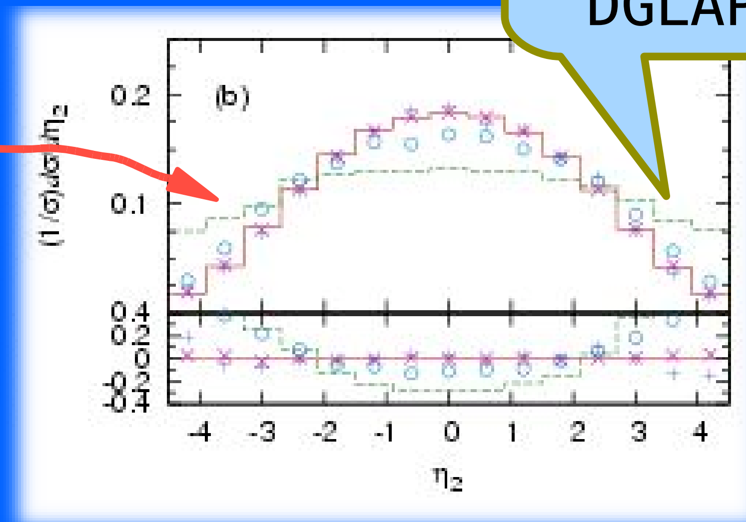
W+4 partons			
TEVATRON		LHC	
Graph	Cross Sect (fb)	Graph	Cross Sect (pb)
Sum	1035.004	Sum	577.948
ug_e+vedggg	112.250	gu_e+vedggg	89.815
gux_e-vexdxggg	112.040	ug_e+vedggg	89.603
uux_e-vexudxgg	112.010	gd_e-vexuggg	45.522
uux_e+veuxdgg	111.900	dg_e-vexuggg	45.342
dux_e-vexddxgg	46.423	uu_e+veudgg	34.174
udx_e+veuuxgg	46.388	dxg_e+veuxggg	15.346
dux_e-vexuuxgg	46.349	gdx_e+veuxggg	15.341
udx_e+veddxgg	46.330	uxg_e-vexdxggg	10.868
gdx_e+veuxggg	40.234	gux_e-vexdxggg	10.866
dg_e-vexuggg	40.122	gg_e+veuxdgg	9.920
udx_e+vegggg	30.906	gg_e+vescxgg	9.907
dux_e-vexgggg	30.867	gg_e-vexsxcgg	9.907
ddx_e-vexudxgg	15.189	gg_e-vexudxgg	9.842
ddx_e+veuxdgg	15.171	du_e+veddgg	8.903
...	...	...	...

W+jets

TeV → LHC

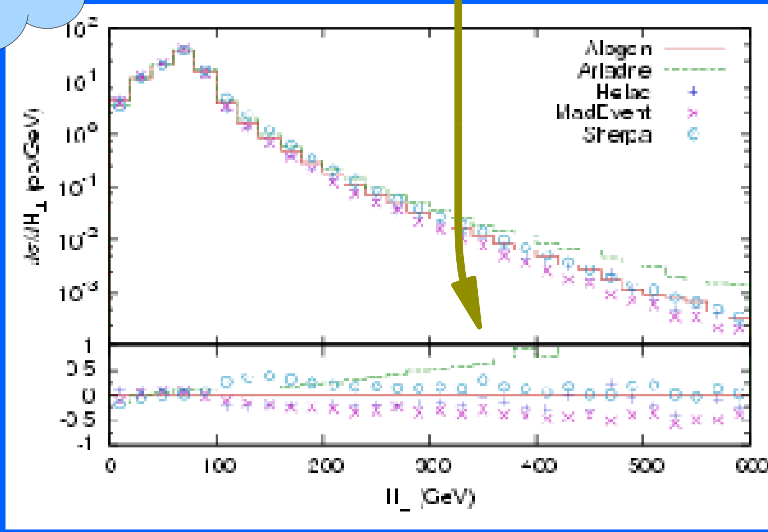
Eta (2nd Jet)

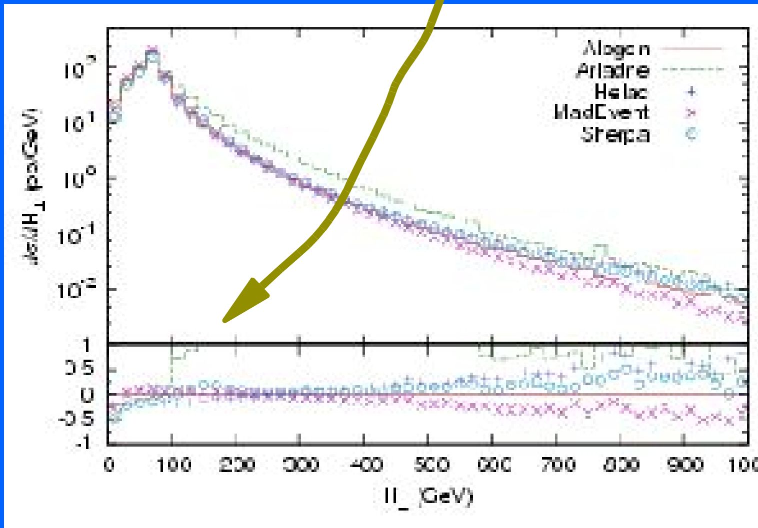




Ariadne more than DGLAP

Data more Like MLM





Sum H_T

W+charm

Direct charm + gluon splitting

- Different game once heavy quarks are included: isolating clean event samples is not so easy.
- Serious studies only recently undertaken.

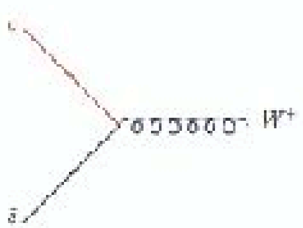
Abazov et al., PLB666, 23 (2008)

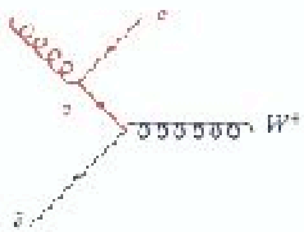
$$\frac{\sigma\,[W + c\text{-jet}]}{\sigma\,[W + \text{jets}]} = 0.074 \pm 0.019(\text{stat.})^{+0.012}_{-0.014}(\text{syst.})$$

$$\left.\frac{\sigma[W + c\text{-jet}]}{\sigma[W + \text{jets}]}\right|_{ALPGEN} = 0.044$$

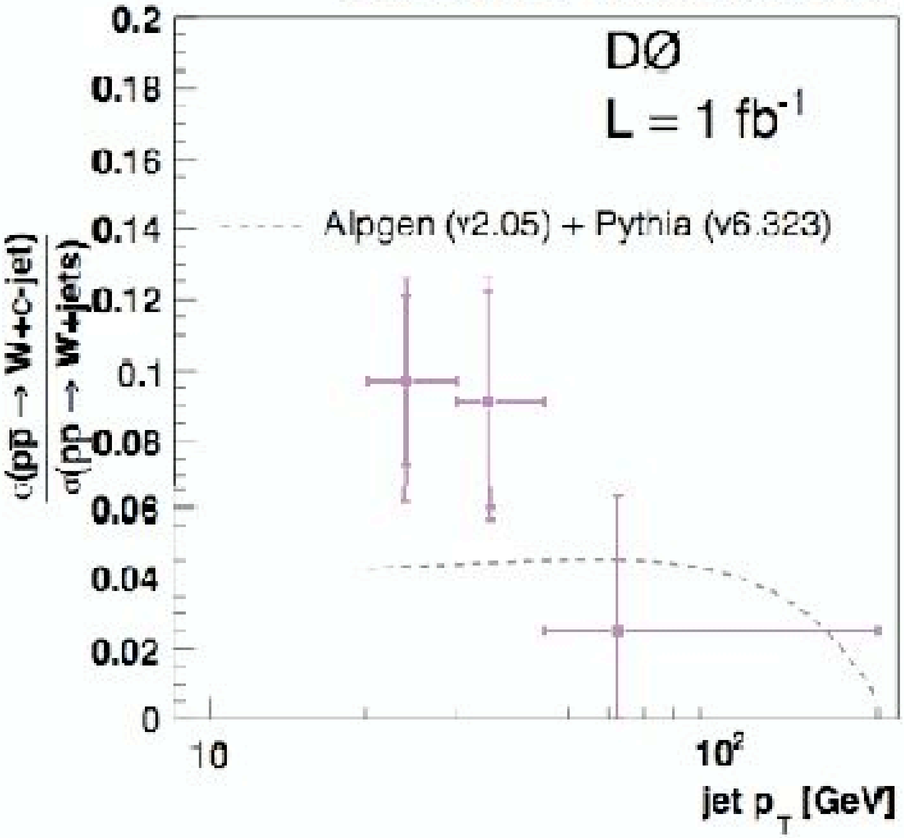
$$\left.\frac{\sigma[W + c\text{-jet}]}{\sigma[W + \text{jets}]}\right|_{MCFM} = 0.045$$

theory error ~ 10%





NB: large logs and charm PDF



- Jury still out, kinematic study essential.

W+bottom

Gluon splitting

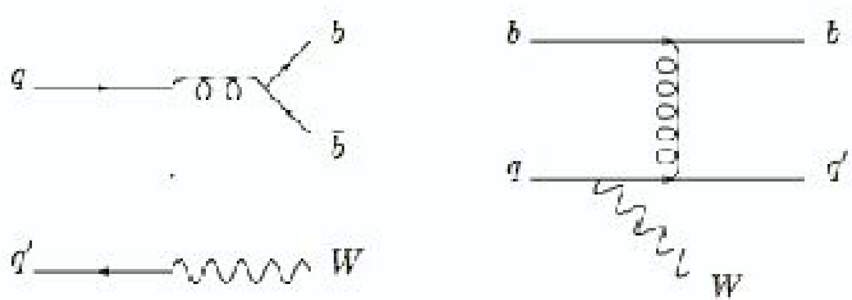
- W+1 or 2 jets, either or both of which may be b-tagged.
- Most important for single top study.
- CDF measurement:

$$\sigma_{b-\text{jets}}(W+b-\text{jets}) \times BR(W \to \ell\nu) = 2.74 \pm 0.27(stat) \pm 0.42(syst)\text{pb}$$

$$\sigma_{b\text{-jets}}(W+b\text{-jets}) \times BR(W \to \ell\nu)_{ALPGEN} = 0.78 \text{ pb}$$

CDF Note 9321

- Ongoing work to compare with ACOT formalism combining (at NLO) two sources of W+b events.



JC et al., arXiv:0809.3003 [hep-ph]

... but still hard to explain factor of 3-4

(NB: role of bottom PDF again)