

## A global analysis of $1 \text{ fb}^{-1}$ of CDF Run II data

**Vista**

(Spanish, Italian)

panoramic view  
model-independent

**Sleuth**

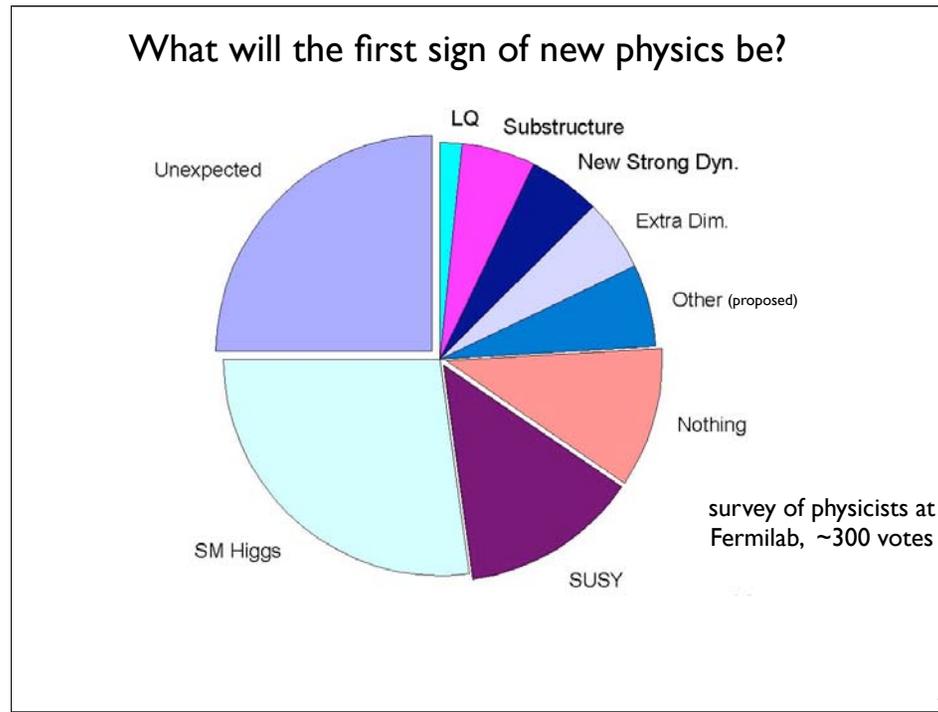
(British)

detective  
quasi-model-independent

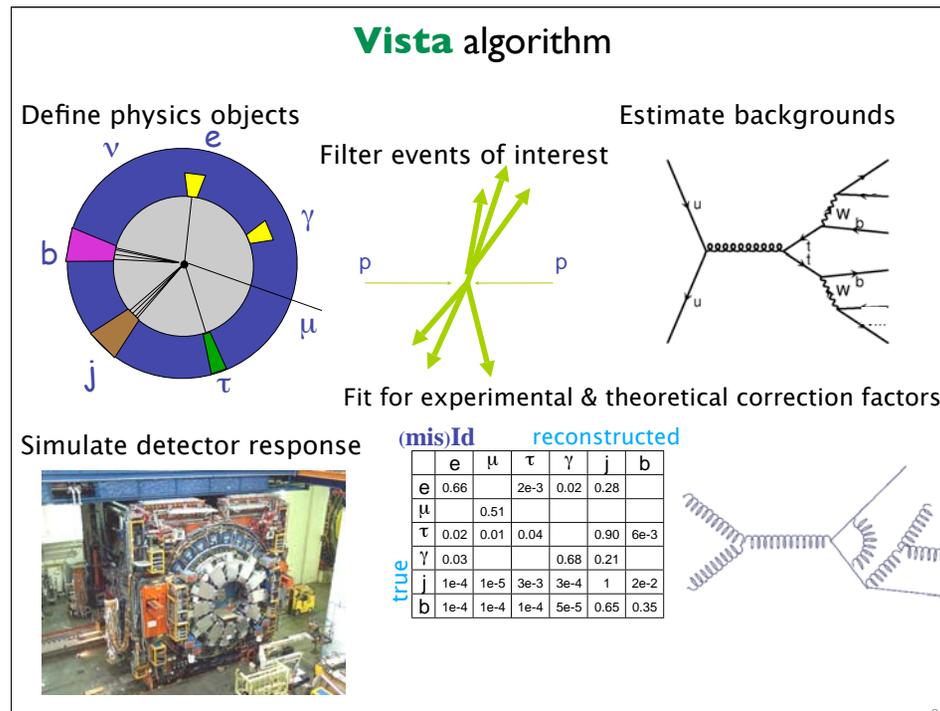
CDF Collaboration      April 2007

1

This talk presents a global analysis of  $\sim 1 \text{ fb}^{-1}$  of CDF Run II high- $p_T$  data. An algorithm named “Vista” (Spanish or Italian for “view” or “sight”) focuses on obtaining a panoramic view of the entire high- $p_T$  data landscape, testing whether there are discrepancies in the bulk of the data. An algorithm named “Sleuth” (British for “detective”) focuses on searching for an excess of data over Standard Model prediction at high  $\Sigma p_T$ .



Although we have good reason to believe that the Standard Model is extended at or just above the electroweak scale, we do not know what form this new physics will take. This pie chart shows (from a survey of over 300 students, postdocs, and faculty) what professional and aspiring physicists think may lie beyond the Standard Model. Note that 25% of those surveyed think the first sign of new physics will be something unexpected (light purple wedge, not yet appearing in the literature), while another 8% think the first sign of new physics will be other (medium blue wedge, something appearing in the literature, but not one of the usual suspects). Given this uncertainty, it is clearly important to cast as wide a net as possible.



The Vista algorithm begins by defining basic physics objects, using standard CDF particle identification criteria to identify energetic ( $p_T > 17$  GeV) and isolated electrons, muons, taus, photons, jets, b-jets, and missing transverse energy. Vista then imposes an offline trigger to filter the roughly two million most interesting high- $p_T$  events. Standard Model processes are estimated using a number of event generators, and the response of the CDF detector to these events is simulated.

Recognizing that the calculation of the Standard Model prediction and the simulation of the detector response are both imperfect, a correction model is constructed to account for possible deficiencies. This correction model, including object (mis)identification efficiencies and the effect of quantum corrections at higher order, is obtained from a global fit of Standard Model prediction to data. The correction model includes only 44 correction factors, all with intuitive physics interpretation. The correction model is intended to be as crude (unsophisticated) as possible, in the interest of simplicity and transparency in the event of a possible discovery.

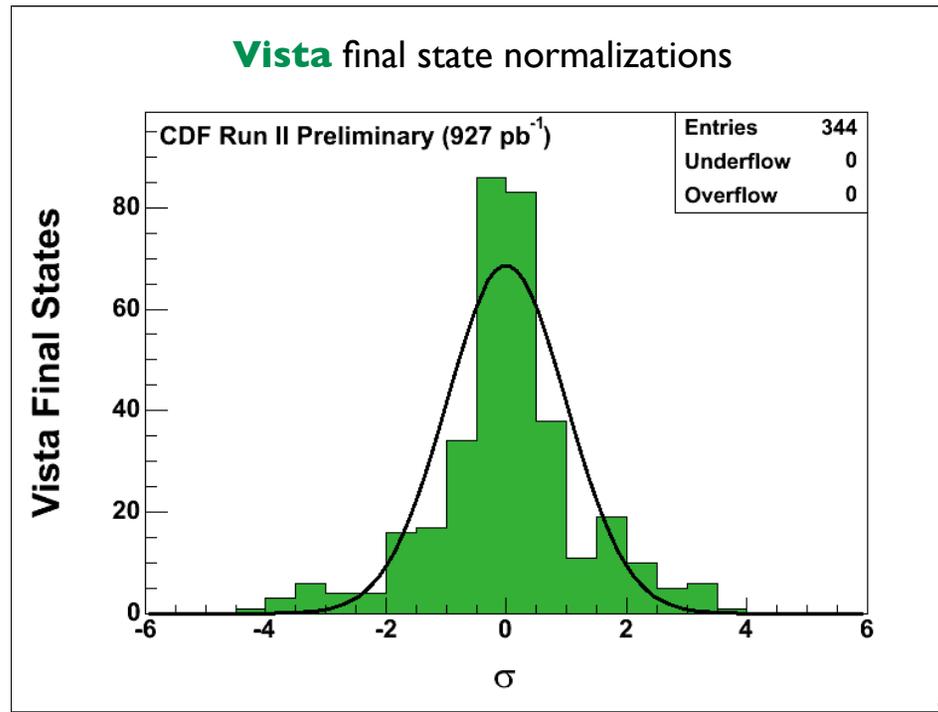


# Vista output

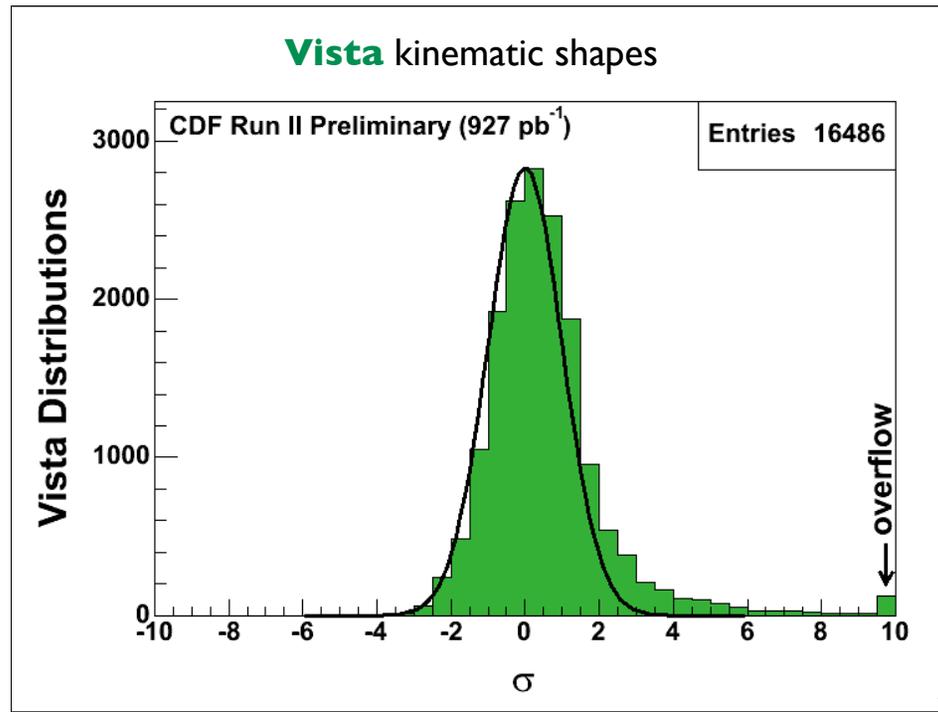
CDF Run II preliminary (927 pb<sup>-1</sup>)

Final State	Data	Background	Final State	Data	Background	Final State	Data	Background
3p++	71	113.7 ± 3.6	2e+1	13	18.8 ± 2.2	++p#	141	144.2 ± 6.0
3j	1461	1002.0 ± 55.8	2e+2	12	4.8 ± 1.2	++p#	34	42.6 ± 2.7
2p++	233	206.5 ± 5.6	2e+	23	36.1 ± 3.8	++p+e	13	15.0 ± 1.3
ee+j	2207	2015.4 ± 28.7	3e E <sub>miss</sub> > 400 GeV	227	235.8 ± 7	++e-	153	137.0 ± 4.2
3j E <sub>miss</sub> < 400 GeV	25430	27204.0 ± 524.3	3e E <sub>miss</sub> < 400 GeV	187	173.1 ± 7.1	e+e-	266880	39214 ± 5031.8
++3p#	1954	1751.0 ± 42	2e3j E <sub>miss</sub> < 400 GeV	28	23.5 ± 3.5	++2p-	14	15.9 ± 2.0
ee+2j	799	695.3 ± 19.3	2e2j E <sub>miss</sub> < 400 GeV	355	326.3 ± 8.4	++j--	79	79.3 ± 2.9
3j# E <sub>miss</sub> > 400 GeV	811	907.5 ± 38.4	2e1j E <sub>miss</sub> < 400 GeV	56	80.2 ± 5	e+j-	102	148.8 ± 7.6
++p-	28	11.0 ± 2.1	2e2j2p-	15	15.5 ± 3.5	++p#	24636	37531.7 ± 6613.6
++e-	636	551.2 ± 11.2	2e1e-	27	31.7 ± 4.8	++2p#	52	75.2 ± 9
++3j	28056	27261.5 ± 405.2	2e1 E <sub>miss</sub> > 400 GeV	415	403.8 ± 9.1	++2e#	22	13.1 ± 1.7
3e1	131	95 ± 4.7	2e1 E <sub>miss</sub> < 400 GeV	101	105.5 ± 8.3	++2e-	28	26.8 ± 2.3
3p+-	60	85.8 ± 8.2	3e1# E <sub>miss</sub> > 400 GeV	28	43 ± 2.8	++e1	102	119.5 ± 6.8
j+e-	74	125 ± 18.6	2e1e-	25	24.7 ± 4.3	++e3j	456	473 ± 14.6
3p E <sub>miss</sub> > 400 GeV	10	26.3 ± 4.6	2e++2j#	15	12.3 ± 1.6	++2j#	30	39 ± 4.6
e1j+	286	269.4 ± 21.1	2e++2j	39	39.5 ± 2.5	++e2j	2149	2152 ± 101.1
++p#e-	29	14.2 ± 1.8	2e++1j	28	29.1 ± 2.8	++e++	14	11.1 ± 2
2j E <sub>miss</sub> < 400 GeV	94502	92437.3 ± 1354.5	2e++	48	45.2 ± 3.7	++e#	491	487.0 ± 12
3e+3j	355	208.0 ± 7.7	e+e-	498	438.5 ± 22.7	++e-#	127	122.3 ± 4.2
8j	11	6.1 ± 2.5	+++	177	204.4 ± 5.4	++e1	10720	10669.3 ± 123.5
7j	57	35.0 ± 4.9	-p#	1052	1045.8 ± 77.1	++e1#	157	144 ± 11.2
6j	255	208 ± 14.7	++e+	18	19.8 ± 2.3	++e1e-	20	45.6 ± 6.7
4j E <sub>miss</sub> < 400 GeV	20665	20008.8 ± 640.2	++e+	121	170.1 ± 4.7	++e+	26344	26373.6 ± 608.9
4j E <sub>miss</sub> > 400 GeV	4241	4409.7 ± 144.7	++e#	221351	200500 ± 3475.5	l6j	24	15.5 ± 2.3
4j2e-	38	57.5 ± 11	++e#e-	22	26.8 ± 2.7	4e1 E <sub>miss</sub> > 400 GeV	13	9.2 ± 1.8
4j2e+	20	36.0 ± 2.4	++e#e+	269	285.5 ± 5.3	4e1 E <sub>miss</sub> < 400 GeV	464	409.2 ± 12.4
4j# E <sub>miss</sub> > 400 GeV	516	525.2 ± 34.5	++e#p#	209	292.2 ± 6.6	4e1 E <sub>miss</sub> < 400 GeV	5354	5285 ± 72.4
4j#e-	28	53.8 ± 11	++e#e-	49	61.4 ± 3.5	4e1 E <sub>miss</sub> < 400 GeV	1639	1558.0 ± 24.1
4j#e+	3653	3627.2 ± 112.1	++e#e+	22	29.9 ± 2.6	4e1# E <sub>miss</sub> > 400 GeV	111	116.8 ± 11.2
4j#e-	576	568.2 ± 26.1	++e#e-	10648	10845.8 ± 96	4e1#	182	184.1 ± 8.8
4j#e+	292	294.7 ± 8.5	2p-	3196	2909.3 ± 85.9	4e1#e-	37	34.1 ± 2
4j#e+	17	20.1 ± 2.5	2p#	38	37.3 ± 3.2	4e1#e+	47	52.2 ± 3
3e-	13	24.3 ± 3	2e-	543	545.7 ± 16.2	4e-	15	14.6 ± 3.1
3j E <sub>miss</sub> > 400 GeV	75894	75939.2 ± 1943.9	2p E <sub>miss</sub> > 400 GeV	4183	4209.1 ± 56.1	4e2 E <sub>miss</sub> > 400 GeV	8812	8076.2 ± 97.9
3j2e-	145	178.1 ± 7.4	2e-	49052	48742 ± 540.3	4e2 E <sub>miss</sub> < 400 GeV	4491	4048.2 ± 57.7
3j2e+	20	30.3 ± 14.4	2e+	108	104 ± 4.1	4e2# E <sub>miss</sub> > 400 GeV	158	205.2 ± 8.3
3j2e-	12	11 ± 2.2	2p#	913	965.2 ± 21.5	4e2#	429	425.1 ± 12.1
3j#e-	83	102.0 ± 11.1	2e++	33462	34020.7 ± 610.1	4e2#e-	46	49.1 ± 2.7
3j#e+	11424	11506.4 ± 190.6	2e++e-	29	37.5 ± 4.5	4e2#e+	86	80.6 ± 3.4
3j#e+	1114	1118.7 ± 27.1	2e++e+	19	16.6 ± 2.1	4e2#e+	19	19.8 ± 2.2
3j#e+	61	84.5 ± 9.2	2e++e+	45728	46516.4 ± 568.2	4e2#	976	1034.8 ± 16.0
3j#e+	212	2168.7 ± 64.2	2e++e#	78	69.8 ± 6.9	4e2#	18	16.7 ± 3.1
3e1 E <sub>miss</sub> > 400 GeV	14	9.3 ± 1.9	2e++e#	79	88.5 ± 12.1	4e2#	302	261.5 ± 7.8
2p+	316	290.8 ± 24.2	2e++e#	1977	2063.3 ± 74.7	4e2#	204	218.1 ± 6.4
2p#	161	176 ± 9.1	++4j	7144	6661.0 ± 147.2	4e2 E <sub>miss</sub> < 400 GeV	9600	9273.7 ± 87.8
2p-	8492	8349.1 ± 84.1	++4j#	403	369 ± 39.9	4e2 E <sub>miss</sub> < 400 GeV	7236	7020.8 ± 74
2j E <sub>miss</sub> > 400 GeV	93408	92799.5 ± 1138.2	++4j#e-	11	7.6 ± 1.6	4j2p-	13	17.6 ± 3.9
2j2e-	643	612.6 ± 18.8	++4j#e+	27	21.7 ± 3.4	4j2p+	13	12.9 ± 1.8
2j2e+	15	25 ± 3.5	++2p-	87	74.5 ± 5	4j# E <sub>miss</sub> > 400 GeV	53	60.4 ± 10.9
2j# E <sub>miss</sub> > 400 GeV	74	100 ± 7.8	++2j	126605	122437 ± 1672.6	4j#	937	989.4 ± 201.6
2j# E <sub>miss</sub> < 400 GeV	43	37.7 ± 100.2	++2j#	53	37.3 ± 5.5	4j#e-	34	30.5 ± 4
2j#	23084	23209.9 ± 397.6	++2j#e+	29	24.7 ± 2.3	4j#e+	104	112.0 ± 4.4
2j#e-	48	11.4 ± 3.4	++2j#e-	10451	10190.1 ± 339.4	4j#e+	179	141.4 ± 3.8
2j#e+	403	425.2 ± 29.7	++2j#e+	101	84.8 ± 6.1	4e-3j#	68	82.2 ± 3.2
2j#e+	7287	7320.5 ± 118.9	++e-	609	555.9 ± 10.2	4e-2j#	87	85 ± 3.3
2j#e#	13	12.6 ± 2.7	++e+	225	212.2 ± 4.7	4e-e#	203	247.2 ± 6.9
2j#e+	41	35.7 ± 6.1	++e#	476424	476372 ± 5863.2	4e-e#	211	176.6 ± 3
2j#e+	374	304.2 ± 24.8	++e#e-	48	35 ± 2.7	4e-e#e1	22	34.6 ± 2.6
2j#e+	3013	3002.3 ± 166.8	++e#e+	29	18.7 ± 1.9	4e-e#e+	62	55 ± 3.1

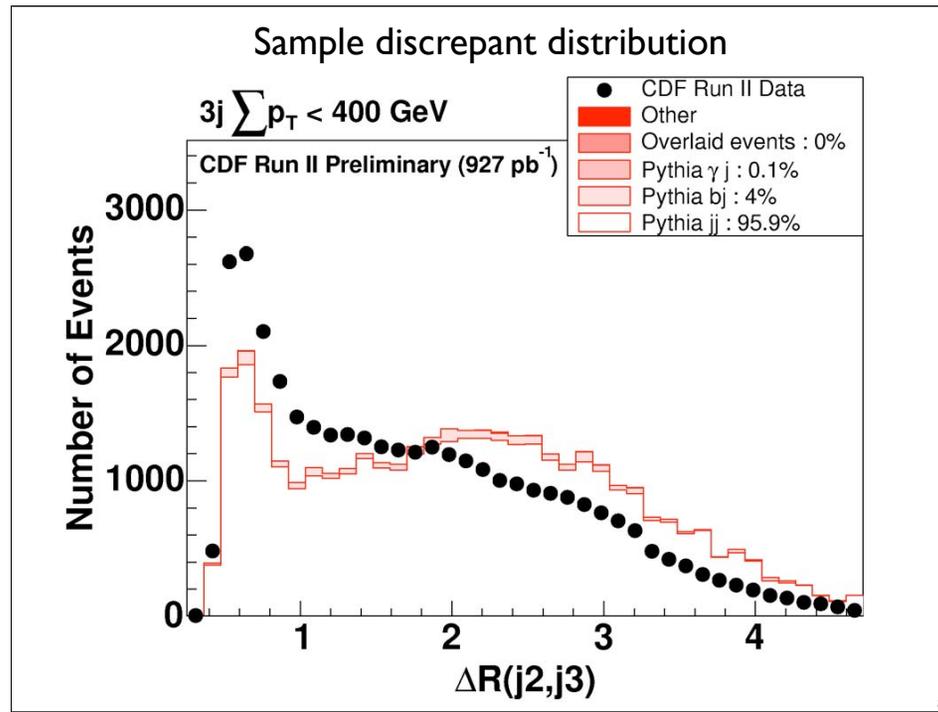
This table shows a subset of the Vista@CDF comparison between Tevatron Run II data and Standard Model prediction. Events are partitioned into exclusive final states based on standard CDF particle identification criteria. Final states are labeled in this table according to the number and types of objects present, and are ordered according to decreasing discrepancy between the total number of events expected and the total number observed in the data. (Final states that do not exhibit notable discrepancies have been listed in inverted alphabetical order, to allow the reader to quickly find a particular final state of interest.) The error on the Standard Model prediction reflects statistical (Monte Carlo) uncertainty only.



The amount (in units of standard deviation  $\sigma$ ) by which the number of events observed in each final state differs from the number of events predicted is shown here as a histogram for Vista's 344 populated exclusive final states. The distribution is centered at zero with unit width. One or more outliers was hoped for, but no outliers are seen.



The amount (in units of standard deviation  $\sigma$ ) by which the shape of data differs from Standard Model prediction is shown here as a histogram for each of 16,486 kinematic distributions considered by Vista. The distribution is centered at zero with unit width, as expected. Roughly 400 distributions show significant disagreement ( $>5\sigma$ ).



Of the Vista@CDF distributions that show significant disagreement between data and Standard Model prediction, roughly 90% derive from the discrepancy shown in its purest form here, reflecting an apparent difference between the field's state-of-the-art showering algorithms and CDF data. Significant effort has gone into trying to understand this discrepancy, with no conclusive result yet obtained.

- A significant Vista discrepancy (such as this one) can motivate a new physics claim if an argument can be made that the effect:
1. is not a statistical fluctuation (corresponding to a probability  $< 0.001$  after the appropriate trials factor is accounted for);
  2. is not due to a problem with our modeling of the CDF detector response;
  3. is not due to a problem with our implementation of the Standard Model prediction; and
  4. has a plausible new physics interpretation.

This Vista shape discrepancy certainly satisfies #1, and with high confidence also satisfies #2, but the case for new physics falters on #3 and #4. None of the remaining Vista shape discrepancies is believed to have a plausible new physics interpretation.

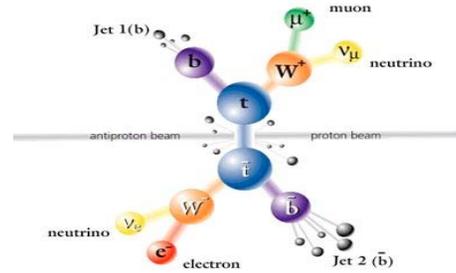
# Sleuth



a quasi-model-independent search strategy for new physics

Assumptions:

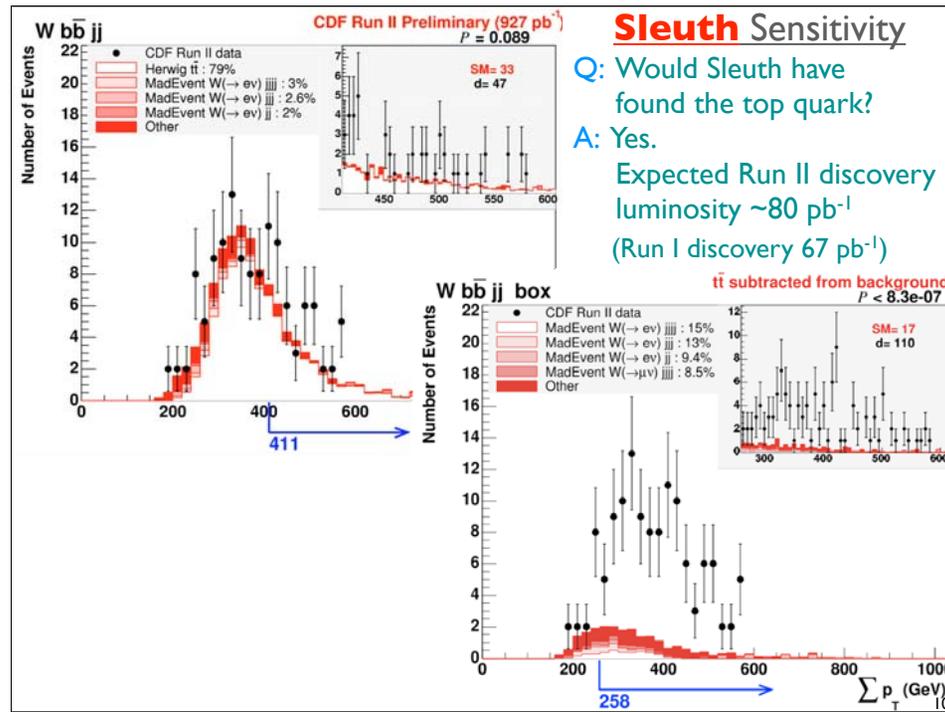
1. Exclusive final state
2. Large  $\sum p_T$
3. An excess



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

Rigorously compute  
the trials factor  
associated with  
looking everywhere

Vista having found no discrepancy on which to base a new physics claim, attention is turned to the high- $\Sigma p_T$  tails. If Vista is described as being model-independent, Sleuth is perhaps best described as being quasi-model-independent, where “quasi” entails three fundamental assumptions: that new physics will appear (1) predominantly in one exclusive final state, (2) at large  $\Sigma p_T$ , and (3) as an excess of data over Standard Model prediction. These three assumptions follow by considering recent discoveries (including the top quark, upper right) and by performing an integration over new physics predictions from recent hep-ph postings (lower left). Crucial to the algorithm is rigorously computing the trials factor associated with the many regions Sleuth considers (lower right). A discrepancy corresponding to a probability  $< 0.001$  after this trials factor is accounted for is considered to be statistically significant.



## Sleuth Sensitivity

Q: Would Sleuth have found the top quark?

A: Yes.

Expected Run II discovery luminosity ~80 pb<sup>-1</sup>  
 (Run I discovery 67 pb<sup>-1</sup>)

To test Sleuth's sensitivity, Sleuth searches the CDF Run II data with the full Standard Model prediction (upper left) and with the Standard Model prediction feigning ignorance of top quark pair production (lower right). Sleuth easily identifies an excess in 1 fb<sup>-1</sup> of CDF Run II data.

Scaling back the integrated luminosity of the sample (continuing to assume 2007 knowledge of backgrounds), Sleuth finds top quark pair production in an integrated luminosity comparable to that of Tevatron Run I. Even if the top quark had not been known to exist, Sleuth likely would have found it in Tevatron Run I.

For a known Standard Model signal like top quark pair production (or Standard Model Higgs boson production), a targeted search is typically the best strategy. Sleuth's strength lies in the possibility of finding something genuinely surprising.

## Sleuth@CDFIIa result

(top 5)

SLEUTH Final State	$\mathcal{P}$	
$b\bar{b}$	0.0055	$\tilde{\mathcal{P}} = 0.46$ <ul style="list-style-type: none"> <li>• Sleuth finds no significant excess in 1fb<sup>-1</sup> of CDF Run II high-pr data</li> <li>• 46% of pseudo experiments are expected to be as interesting</li> <li>• This does <b>not</b> prove there is no new physics present</li> </ul>
$j\cancel{p}$	0.0092	
$\ell^+\ell'^+\cancel{p}jj$	0.011	
$\ell^+\ell'^+\cancel{p}$	0.016	
$\tau\cancel{p}$	0.016	

11

The result of applying Sleuth to CDF Run II data is shown here. The table shows Sleuth's top 5 final states, together with the fraction of pseudo experiments that would produce a region as interesting as the most interesting region observed in that final state. The Sleuth  $b\bar{b}$  final state leads this list, followed by the monojet final state. The third and fourth final states in this list contain events with an electron and muon of the same sign, missing transverse energy, and two or zero jets, respectively. The fifth final state contains one identified tau, significant missing transverse energy, and zero or one reconstructed jets. In the Sleuth  $b\bar{b}$  final state, a fraction 0.0055 of pseudo experiments in this final state alone would produce something as interesting as what is observed. After taking into account the roughly 100 Sleuth final states, the fraction of pseudo experiments that would produce a region in any final state as interesting as the most interesting region in this final state is  $\tilde{\mathcal{P}}=0.46$ , shown at upper right. All pseudo experiments involve pseudo data drawn from the Standard Model prediction assuming an accurate detector simulation and Vista correction model.

Sleuth's result of course does not prove that there is no new physics present, merely that this particular search strategy does not find a discrepancy motivating a new physics claim.

## Summary

A global analysis of  $1 \text{ fb}^{-1}$  of CDF Run II data has been performed

### Vista

model-independent  
searches the bulk of distributions

### Sleuth

quasi-model-independent  
searches the high- $\Sigma p_T$  tails

This global analysis has revealed no new physics in  $1 \text{ fb}^{-1}$   
Tevatron Run II will provide 10x more data yet to be searched

12

A global analysis of  $\sim 1 \text{ fb}^{-1}$  of CDF Run II data has been performed. Vista provides a model-independent search of the bulks of distributions. Sleuth provides a quasi-model-independent search for excesses at large scalar summed transverse momentum. Sleuth is broadly comparable to targeted searches for models satisfying the assumptions on which Sleuth is based. Sleuth may miss evidence for models that do not result in an excess of data above Standard Model prediction in some final state at large summed scalar transverse momentum.

This global search on  $\sim 1 \text{ fb}^{-1}$  has revealed no indications of new physics. (This result of course does not prove that no hint of new physics lies in these data, merely that no such hint has been revealed by this particular analysis.) This global analysis will continue throughout the remainder of Tevatron Run II, aiming for discovery.