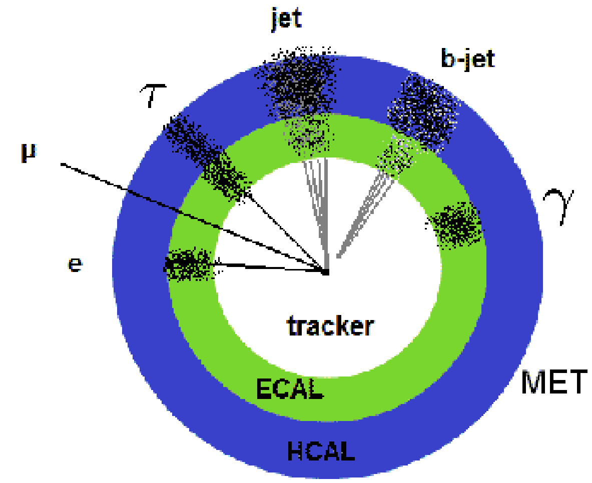
Bump Hunter
in 2 fb^{-1} of CDF data

Georgios Choudalakis

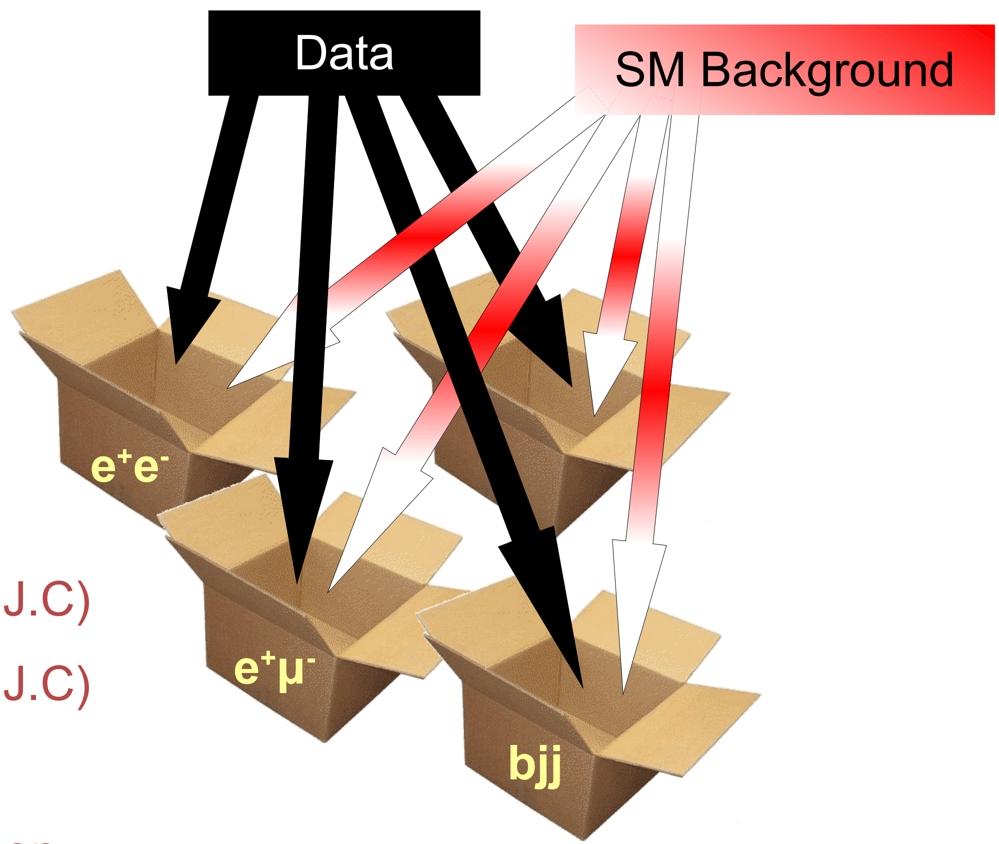
CDF Collaboration

PHENO-08
U. of Wisconsin-Madison
2008/04/29

Context



- CDF Global Search for New Physics
- Si Xie's talk 
- Implement globally the SM background
- Partition in exclusive final states
- Search for discrepancies in populations, shapes, Σp_T tails, and now also in mass bumps.



Bibliography

http://arxiv.org/abs/0712.2534 (submitted to PRL)

http://arxiv.org/abs/0712.1311 (PRD)

http://arxiv.org/abs/0710.2372 (SUSY07,Eur.Phys.J.C)

http://arxiv.org/abs/0710.2378 (SUSY07,Eur.Phys.J.C)

My PhD Thesis (MIT,2008)

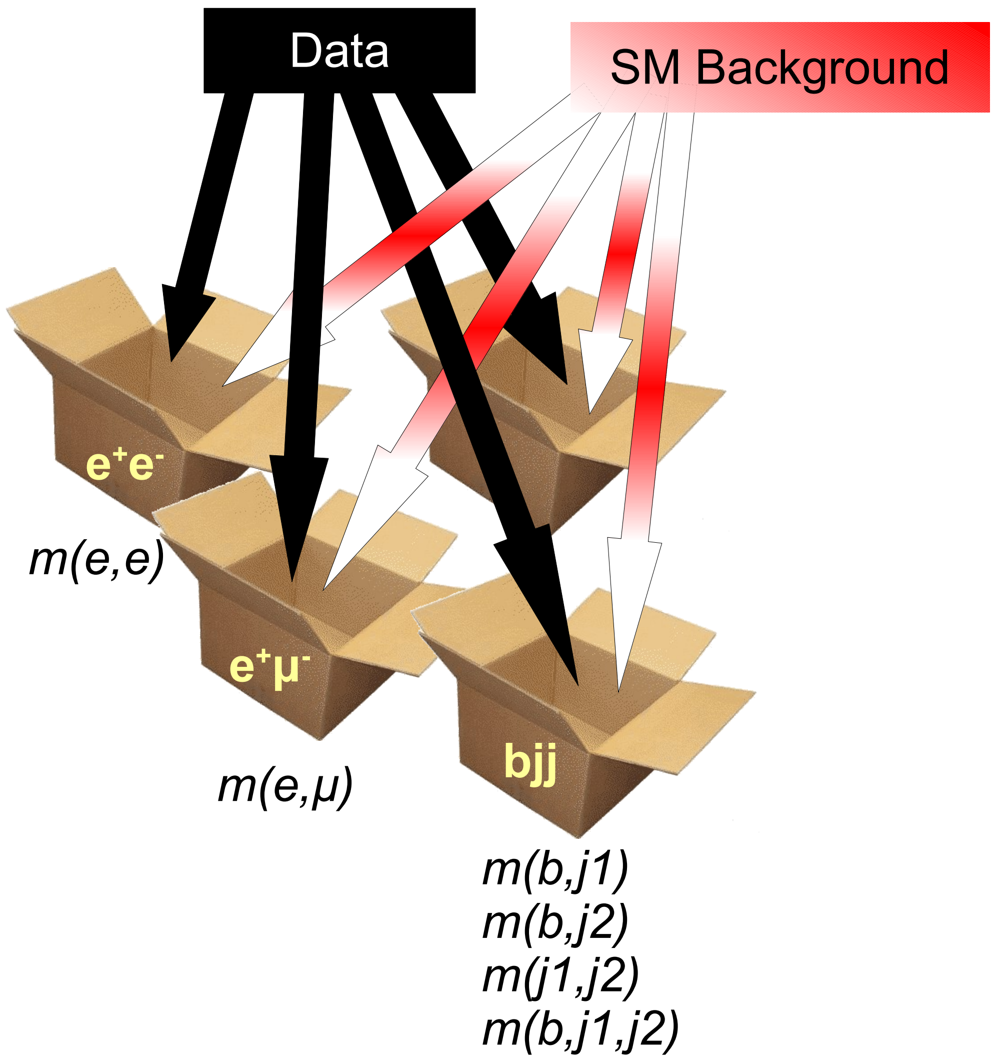
Upcoming Moriond Proceedings by Conor Henderson

Goal of the Bump Hunter

- Find narrow mass resonances
- Evaluate their statistical significance

Scope

- All mass variables in all exclusive final states.



399 final states; a lot of information

- Table including all Vista final states with at least 10 data events
- Background uncertainties are statistical.

CDF Run II Preliminary (2.0 fb^{-1})
The calculation of σ accounts for the trials factor

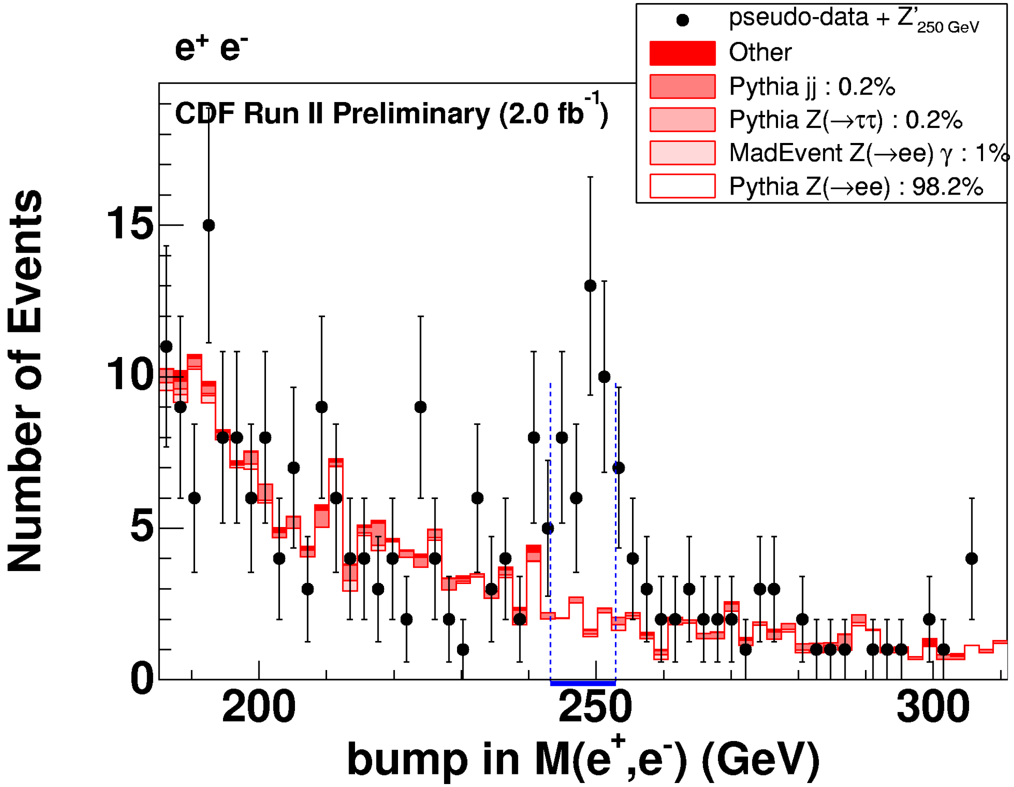
Final State	Data	Background	σ
$b e^{\pm} \not p$	690	817.7 ± 9.2	−2.7
$\gamma \tau^{\pm}$	1371	1217.6 ± 13.3	+2.2
$\mu^{\pm} \tau^{\pm}$	63	35.2 ± 2.8	+1.7
$b2j\not p$ high-Σp_T	255	327.2 ± 8.9	−1.7
$2j\tau^{\pm}$ low-Σp_T	574	670.3 ± 8.6	−1.5
$3j\tau^{\pm}$ low-Σp_T	148	199.8 ± 5.2	−1.4
$e^{\pm}\not p\tau^{\pm}$	36	17.2 ± 1.7	+1.4
$2j\tau^{\pm}\tau^{\mp}$	33	62.1 ± 4.3	−1.3
$e^{\pm}j$	741710	764832 ± 6447.2	−1.3
$j2\tau^{\pm}$	105	150.8 ± 6.3	−1.2
$e^{\pm}2j$	256946	249148 ± 2201.5	+1.2
2bj low-Σp_T	279	352.5 ± 11.9	−1.1
$j\tau^{\pm}$ low-Σp_T	1385	1525.8 ± 15	−1.1
2b2j low-Σp_T	108	153.5 ± 6.8	−1
$b\mu^{\pm}\not p$	528	613.5 ± 8.7	−0.9
$\mu^{\pm}\gamma\not p$	523	611 ± 12.1	−0.8
$2b\gamma$	108	70.5 ± 7.9	+0.1
8j	14	13.1 ± 4.4	0
7j	103	97.8 ± 12.2	0
6j	653	659.7 ± 37.3	0
5j	3157	3178.7 ± 67.1	0
4j high-Σp_T	88546	89096.6 ± 935.2	0
4j low-Σp_T	14872	14809.6 ± 186.3	0
$4j2\gamma$	46	46.4 ± 3.9	0
$4j\tau^{\pm}$ high-Σp_T	29	26.6 ± 1.7	0
$4j\tau^{\pm}$ low-Σp_T	43	63.1 ± 3.3	0
$4j\not p$ high-Σp_T	1064	1012 ± 62.9	0
$4j\gamma\tau^{\pm}$	19	10.8 ± 2	0
$4j\gamma\not p$	62	104.2 ± 22.4	0
$4j\gamma$	7962	8271.2 ± 245.1	0
$4j\mu^{\pm}\not p$	574	590.5 ± 13.6	0
$4j\mu^{\pm}\mu^{\mp}$	38	48.4 ± 6.2	0
$4j\mu^{\pm}$	1363	1350.1 ± 37.7	0
3j high-Σp_T	159926	159143 ± 1061.9	0
3j low-Σp_T	62681	64213.1 ± 496	0
$3j2\gamma$	151	177.5 ± 7.1	0
$3j\tau^{\pm}$ high-Σp_T	68	76.9 ± 3	0
$3j\not p$ high-Σp_T	1706	1899.4 ± 77.6	0
$3j\not p$ low-Σp_T	42	36.2 ± 5.7	0
$3j\gamma\tau^{\pm}$	39	37.8 ± 3.6	0
$3j\gamma\not p$	204	249.8 ± 24.4	0
$3j\gamma$	24639	24899.4 ± 372.4	0
$3j\mu^{\pm}\not p$	2884	2971.5 ± 52.1	0
$3j\mu^{\pm}\gamma\not p$	10	3.6 ± 1.9	0
$3j\mu^{\pm}\gamma$	15	7.9 ± 2.9	0
$3j\mu^{\pm}\mu^{\mp}$	175	177.8 ± 16.2	0
$3j\mu^{\pm}$	5032	4989.5 ± 108.9	0
3b2j	23	28.9 ± 4.7	0
3bj	82	82.6 ± 5.7	0
3b	67	85.6 ± 7.7	0
[illegible]$\tau^{\pm}$	498	512.7 ± 14.2	0
[illegible]$\gamma\not p$	128	107.2 ± 6.9	0
[illegible]γ	5548	5562.8 ± 40.5	0
[illegible] high-Σp_T	190773	190842 ± 781.2	0
[illegible] low-Σp_T	165984	162530 ± 1581	0
[illegible]	22	40.6 ± 3.2	0
[illegible]$\gamma\not p$	11	8 ± 2.4	0
$2j2\gamma$	580	581 ± 13.7	0
$2j\tau^{\pm}$ high-Σp_T	96	114.6 ± 3.3	0

Final State	Data	Background	σ
$2j\not p$ high-Σp_T	87	80.9 ± 6.8	0
$2j\not p$ low-Σp_T	114	79.5 ± 100.8	0
$2j\not p\tau^{\pm}$	18	13.2 ± 2.2	0
$2j\gamma\tau^{\pm}$	142	144.6 ± 5.7	0
$2j\gamma\not p$	908	980.3 ± 63.7	0
$2j\gamma$	71364	73021.4 ± 595.9	0
$2j\mu^{\pm}\tau^{\mp}$	16	19.3 ± 2.2	0
$2j\mu^{\pm}\not p$	17927	18340.6 ± 201.9	0
$2j\mu^{\pm}\gamma\not p$	31	27.7 ± 7.7	0
$2j\mu^{\pm}\gamma$	57	58.2 ± 13	0
$2j\mu^{\pm}\mu^{\mp}\not p$	11	7.8 ± 2.7	0
$2j\mu^{\pm}\mu^{\mp}$	956	924.9 ± 61.2	0
$2j\mu^{\pm}$	22461	23111.4 ± 366.6	0
$2e^{\pm}j$	14	13.8 ± 2.3	0
$2e^{\pm}e^{\mp}$	20	17.5 ± 1.7	0
$2e^{\pm}$	32	49.2 ± 3.4	0
2b high-Σp_T	666	689 ± 9.4	0
2b low-Σp_T	323	313.2 ± 10.3	0
2b3j low-Σp_T	53	57.4 ± 6.5	0
2b2j high-Σp_T	718	803.3 ± 12.7	0
$2b2j\not p$ high-Σp_T	15	21.8 ± 2.8	0
$2b2j\gamma$	32	39.7 ± 6.2	0
$2b2j\mu^{\pm}\not p$	14	17.3 ± 1.9	0
$2b2j\mu^{\pm}$	22	21.8 ± 2	0
$2b\mu^{\pm}\not p$	11	14.4 ± 2.1	0
2bj high-Σp_T	891	967.1 ± 13.2	0
$2bj\not p$ high-Σp_T	25	31.3 ± 3.1	0
$2bj\gamma$	71	54.5 ± 7.1	0
$2bj\mu^{\pm}\not p$	12	10.7 ± 1.9	0
$2be^{\pm}2j\not p$	30	27.3 ± 2.2	0
$2be^{\pm}2j$	72	66.5 ± 2.9	0
$2be^{\pm}\not p$	22	19.1 ± 2.2	0
$2be^{\pm}j\not p$	19	19.4 ± 2.2	0
$2be^{\pm}j$	63	63 ± 3.4	0
$2be^{\pm}$	96	92.1 ± 4.1	0
$\tau^{\pm}\tau^{\mp}$	856	872.5 ± 19	0
$\gamma\not p$	3793	3770.7 ± 127.3	0
$\mu^{\pm}\tau^{\mp}$	381	440.9 ± 7.3	0
$\mu^{\pm}\not p\tau^{\mp}$	60	75.7 ± 3.4	0
$\mu^{\pm}\not p\tau^{\pm}$	15	12 ± 2	0
$\mu^{\pm}\not p$	734290	734296 ± 4897.8	0
$\mu^{\pm}\gamma$	475	469.8 ± 12.5	0
$\mu^{\pm}\mu^{\mp}\not p$	169	198.5 ± 8.2	0
$\mu^{\pm}\mu^{\mp}\gamma$	83	60 ± 3.1	0
$\mu^{\pm}\mu^{\mp}$	25283	25178.5 ± 86.5	0
$j2\gamma\not p$	36	30.4 ± 4.2	0
$j2\gamma$	1822	1813.2 ± 27.4	0
$j\tau^{\pm}$ high-Σp_T	52	56.2 ± 2.5	0
$j\tau^{\pm}\tau^{\mp}$	203	252.2 ± 8.7	0
$j\not p$ high-Σp_T	4432	4431.7 ± 45.2	0
$j\gamma\tau^{\pm}$	526	476 ± 9.3	0
$j\gamma\not p$	1882	1791.9 ± 72.3	0
$j\gamma$	103319	102124 ± 570.6	0
$j\mu^{\pm}\tau^{\mp}$	71	98 ± 3.9	0
$j\mu^{\pm}\tau^{\pm}$	15	12 ± 2	0
$j\mu^{\pm}\not p\tau^{\mp}$	26	30.8 ± 2.6	0
$j\mu^{\pm}\not p$	109081	108323 ± 707.7	0
$j\mu^{\pm}\gamma\not p$	171	171.1 ± 31	0
$j\mu^{\pm}\gamma$	152	190 ± 39.3	0

Final State	Data	Background	σ
$j\mu^{\pm}\mu^{\mp}\not p$	32	32.2 ± 10.9	0
$j\mu^{\pm}\mu^{\mp}\gamma$	14	11.5 ± 2.6	0
$j\mu^{\pm}\mu^{\mp}$	4852	4271.2 ± 185.4	0
$j\mu^{\pm}$	77689	76987.5 ± 930.2	0
$e^{\pm}4j\not p$	903	830.6 ± 13.2	0
$e^{\pm}4j\gamma$	25	29.2 ± 3.6	0
$e^{\pm}4j$	15750	16740.4 ± 390.5	0
$e^{\pm}3j\tau^{\mp}$	15	21.1 ± 2.2	0
$e^{\pm}3j\not p$	4054	4077.2 ± 63.6	0
$e^{\pm}3j\gamma$	108	79.3 ± 5	0
$e^{\pm}3j$	60725	60409.3 ± 723.3	0
$e^{\pm}2\gamma$	41	34.2 ± 2.6	0
$e^{\pm}2j\tau^{\pm}$	37	47.2 ± 2.2	0
$e^{\pm}2j\tau^{\mp}$	109	95.9 ± 6.8	0
$e^{\pm}2j\not p$	25725	25403.1 ± 209.4	0
$e^{\pm}2j\gamma\not p$	30	31.8 ± 4.8	0
$e^{\pm}2j\gamma$	398	342.8 ± 15.7	0
$e^{\pm}2j\mu^{\mp}\not p$	22	14.8 ± 1.9	0
$e^{\pm}2j\mu^{\mp}$	23	15.8 ± 2	0
$e^{\pm}\tau^{\pm}$	437	387 ± 5.3	0
$e^{\pm}\tau^{\mp}$	1333	1266 ± 12.3	0
$e^{\pm}\not p\tau^{\mp}$	109	106.1 ± 2.7	0
$e^{\pm}\not p$	960826	956579 ± 3077.7	0
$e^{\pm}\gamma\not p$	497	496.8 ± 10.3	0
$e^{\pm}\gamma$	3578	3589.9 ± 24.1	0
$e^{\pm}\mu^{\pm}\not p$	31	29.9 ± 1.6	0
$e^{\pm}\mu^{\mp}\not p$	109	99.4 ± 2.4	0
$e^{\pm}\mu^{\pm}$	45	28.5 ± 1.8	0
$e^{\pm}\mu^{\mp}$	350	313 ± 5.4	0
$e^{\pm}j2\gamma$	13	16.1 ± 3.9	0
$e^{\pm}j\tau^{\mp}$	386	418 ± 18.9	0
$e^{\pm}j\tau^{\pm}$	160	162.8 ± 3.5	0
$e^{\pm}j\not p\tau^{\mp}$	48	44.6 ± 3.3	0
$e^{\pm}j\not p\tau^{\pm}$	11	8.3 ± 1.5	0
$e^{\pm}j\not p$	121431	121023 ± 747.6	0
$e^{\pm}j\gamma\not p$	159	192.6 ± 10.9	0
$e^{\pm}j\gamma$	1389	1368.9 ± 38.9	0
$e^{\pm}j\mu^{\mp}\not p$	42	33 ± 2.9	0
$e^{\pm}j\mu^{\pm}\not p$	16	9.2 ± 1.9	0
$e^{\pm}j\mu^{\mp}$	62	63.8 ± 3.2	0
$e^{\pm}j\mu^{\pm}$	13	8.2 ± 2	0
$e^{\pm}e^{\mp}4j$	148	159.1 ± 7	0
$e^{\pm}e^{\mp}3j$	717	743.6 ± 24.4	0
$e^{\pm}e^{\mp}2j\not p$	32	41.4 ± 5.6	0
$e^{\pm}e^{\mp}2j\gamma$	10	11.4 ± 2.9	0
$e^{\pm}e^{\mp}2j$	3638	3566.8 ± 72	0
$e^{\pm}e^{\mp}\tau^{\pm}$	18	16.1 ± 1.7	0
$e^{\pm}e^{\mp}\not p$	822	831.8 ± 13.6	0
$e^{\pm}e^{\mp}\gamma$	191	221.9 ± 5.1	0
$e^{\pm}e^{\mp}j\not p$	155	170.8 ± 12.4	0
$e^{\pm}e^{\mp}j\gamma$	48	45 ± 3.9	0
$e^{\pm}e^{\mp}j$	17903	18258.2 ± 204.4	0
$e^{\pm}e^{\mp}$	98901	99086.9 ± 147.8	0
b6j	51	42.3 ± 3.8	0
b5j	237	192.5 ± 7.1	0
b4j high-Σp_T	26	23.4 ± 2.6	0
b4j low-Σp_T	836	821.7 ± 15.9	0
b3j high-Σp_T	12081	12071 ± 84.1	0
b3j low-Σp_T	2974	2873 ± 31	0

Method

Caution: This is NOT real data!



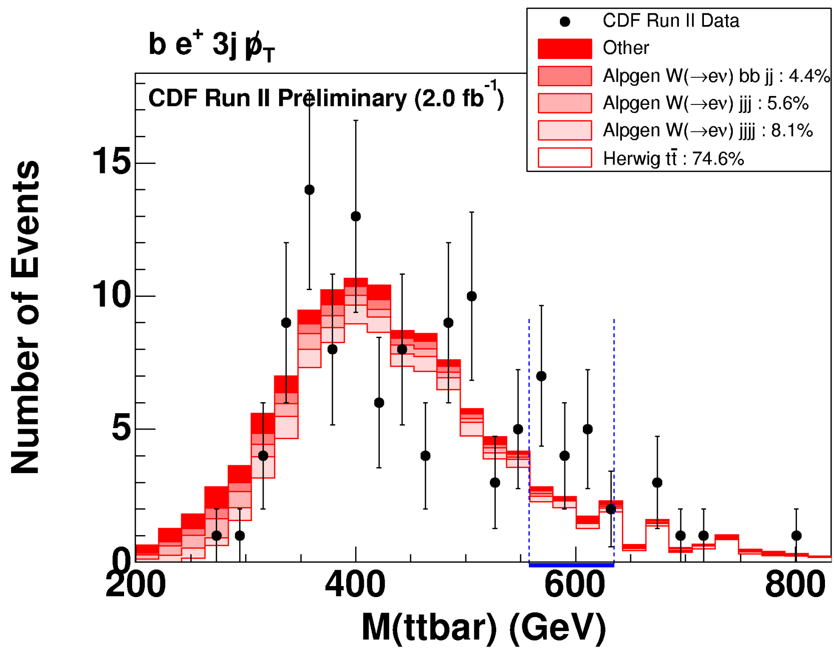
Scan all mass spectra with a window.

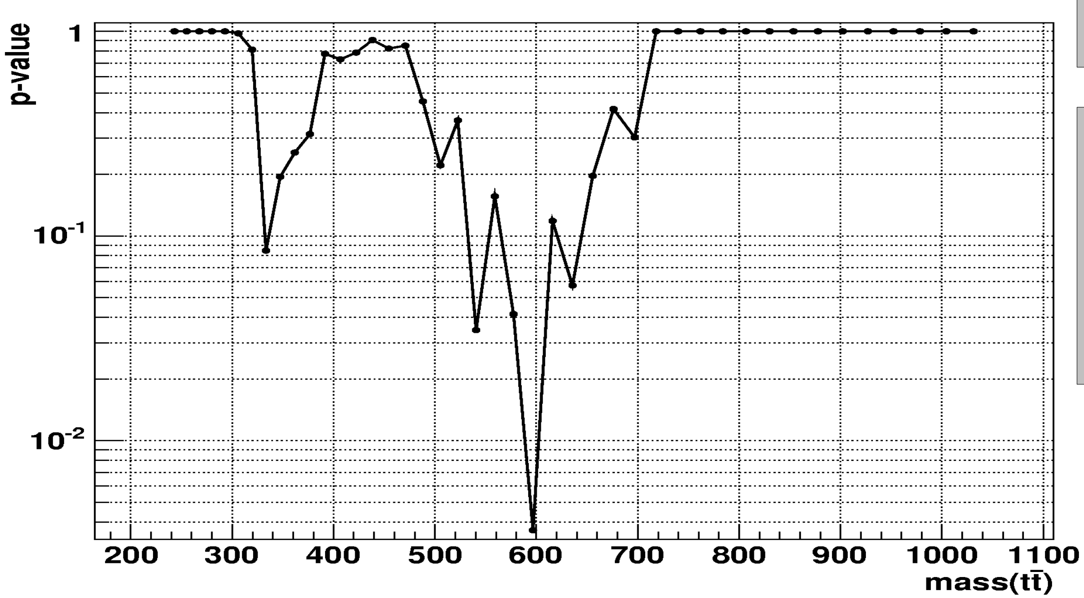
Window size follows mass resolution.

$$m = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2} \Rightarrow \Delta m$$

- Consider bumps with ≥ 5 data events.
- Sidebands have to agree more than center, and not be too discrepant (5σ).

Statistical Significance





Each bump has a p-value

$$\sum_{n=d}^{\infty} \frac{b^n}{n!} e^{-b}$$

Most interesting bump: p-val_{min}

Use pseudo-data to find

P_a = The probability that a $\text{p-val} \le \text{p-val}_{min}$ would appear by coincidence.

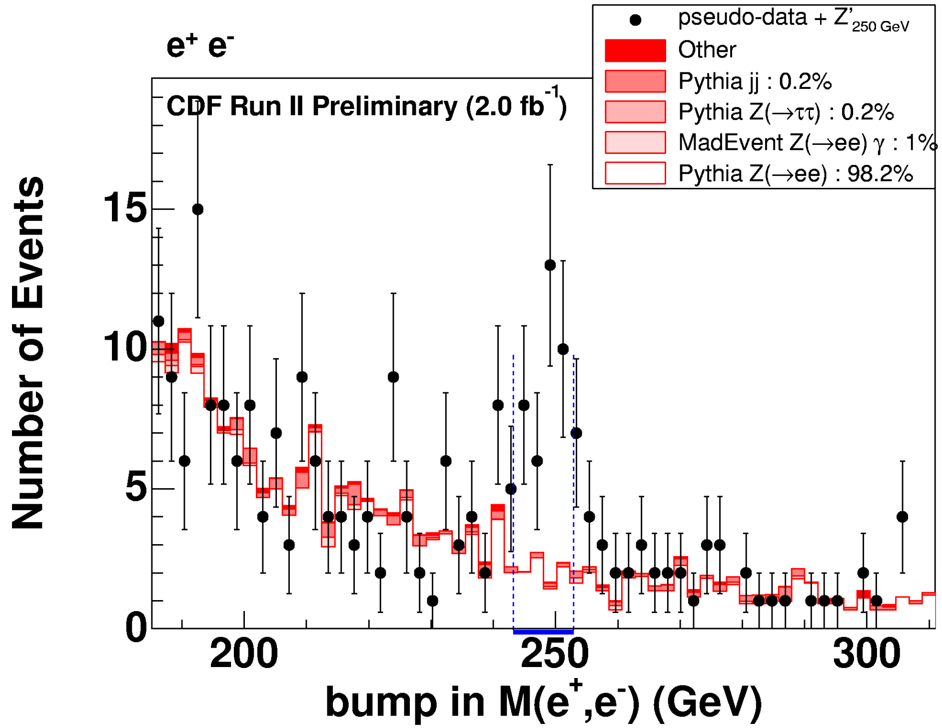
$P_b = 1 - (1-P_a)^{\text{5000 distributions}}$

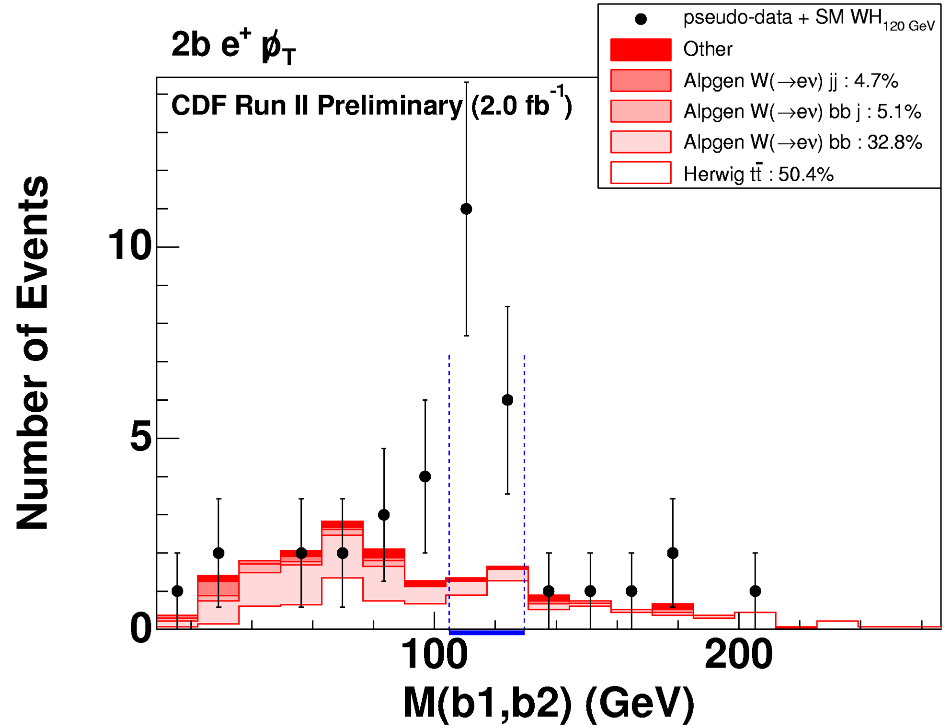
= probability at least one mass would have such a small P_a by coincidence.

Discovery threshold:

$P_b = 3\sigma \leftrightarrow P_a = 5\sigma$

Sensitivity





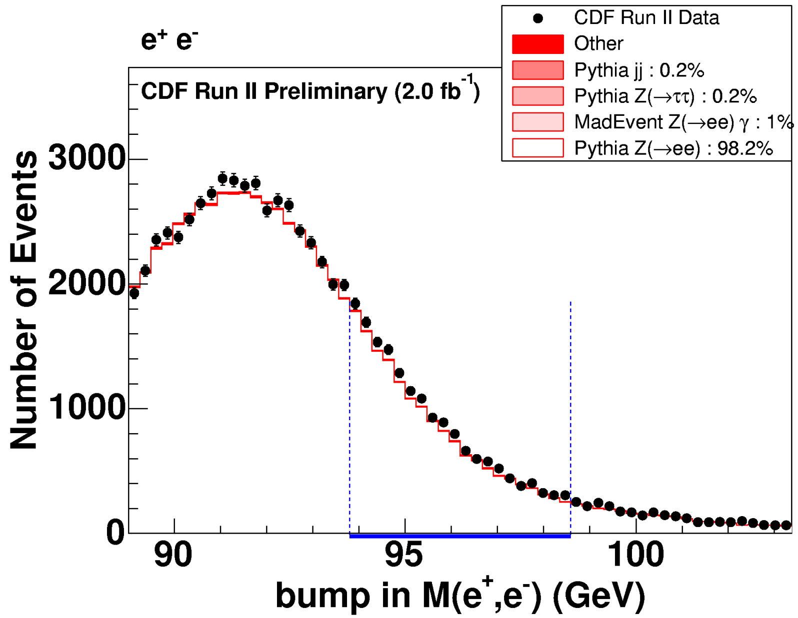
$Z'_{250} \to$ charged leptons
5σ discovery if $\sigma \times BR \approx 0.325$ pb.

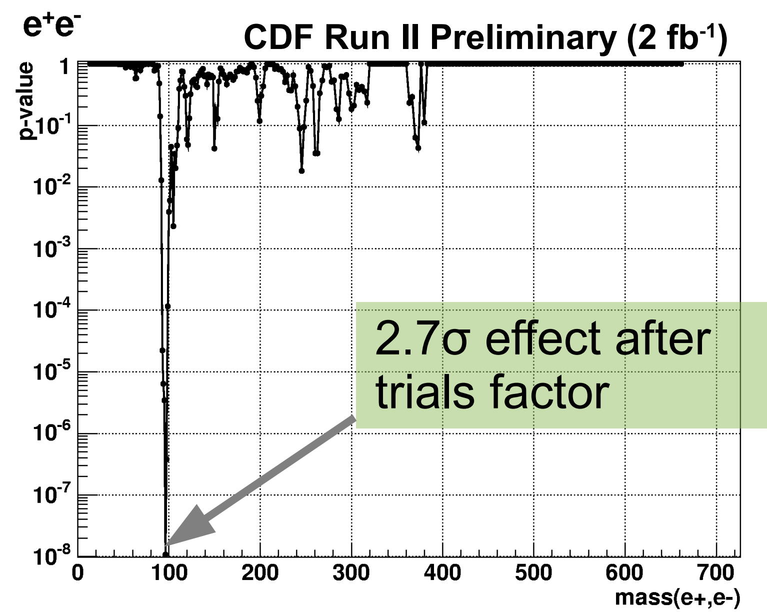
$WH_{120} \to \ell\nu b\bar{b}$
5σ discovery if $\sigma \approx 14$ pb.

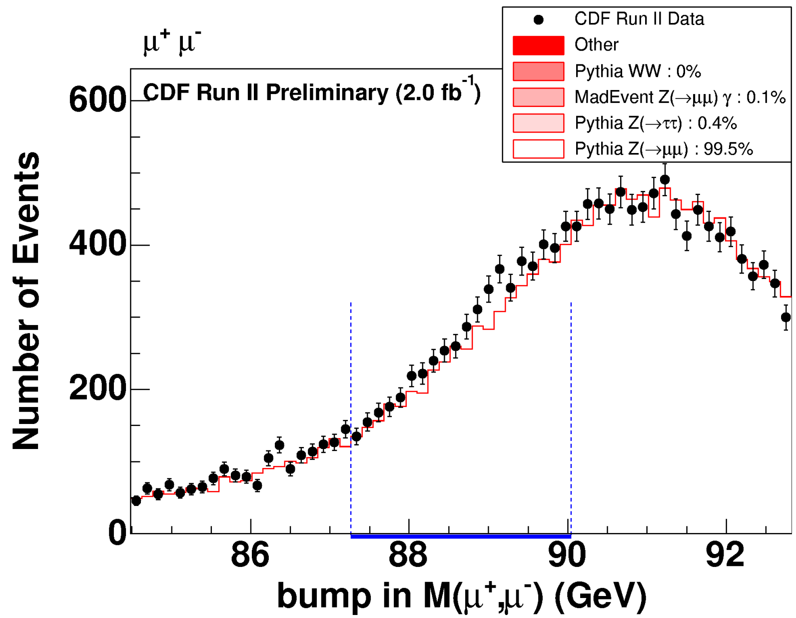
(Standard Model predicts 0.159 pb.)

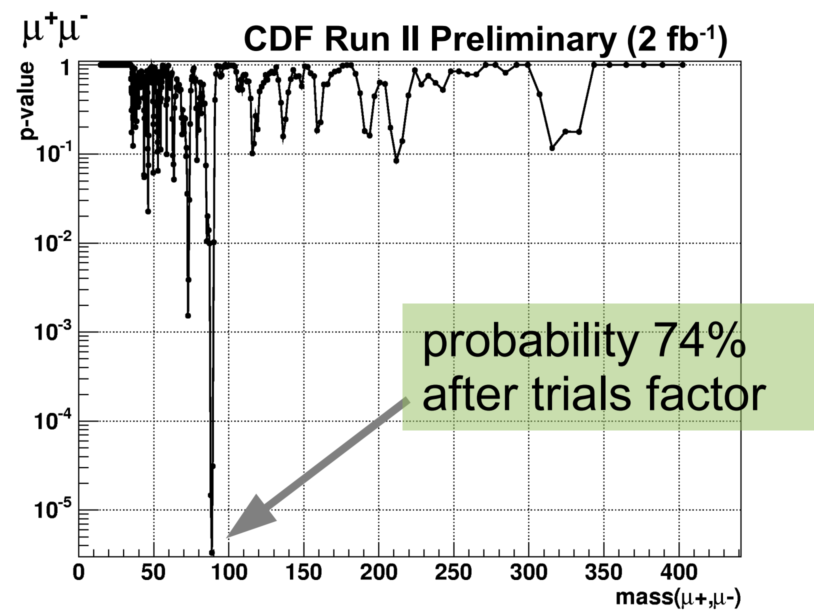
Results

Anything like Z' → e^+e^- or $\mu^+\mu^-$?

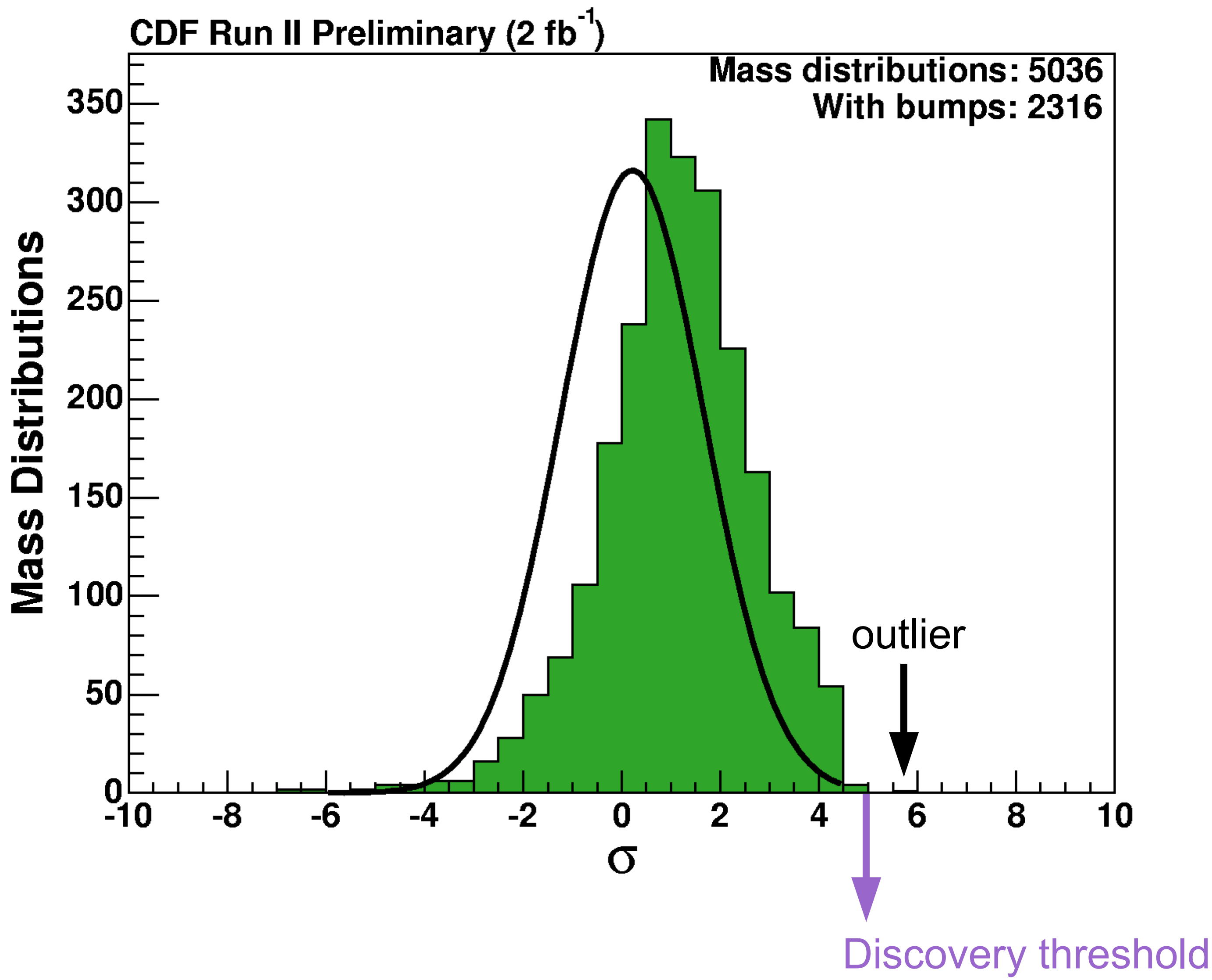




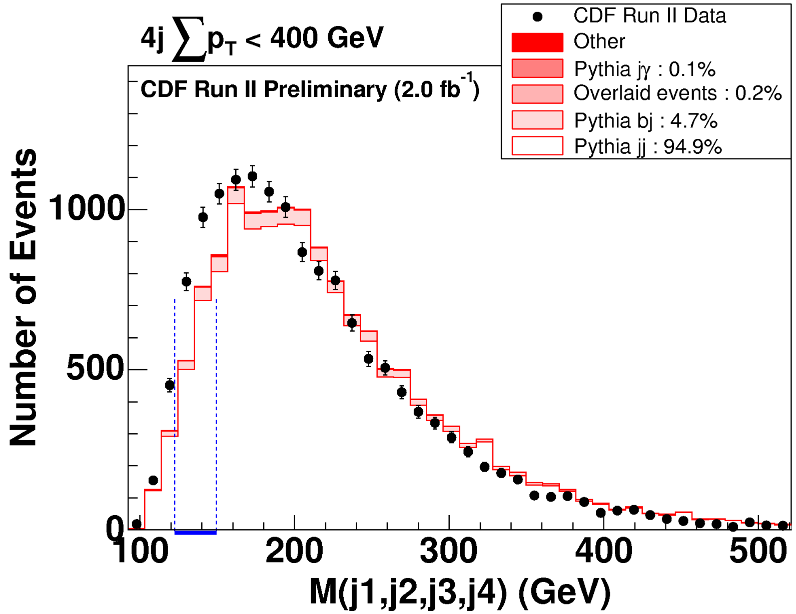




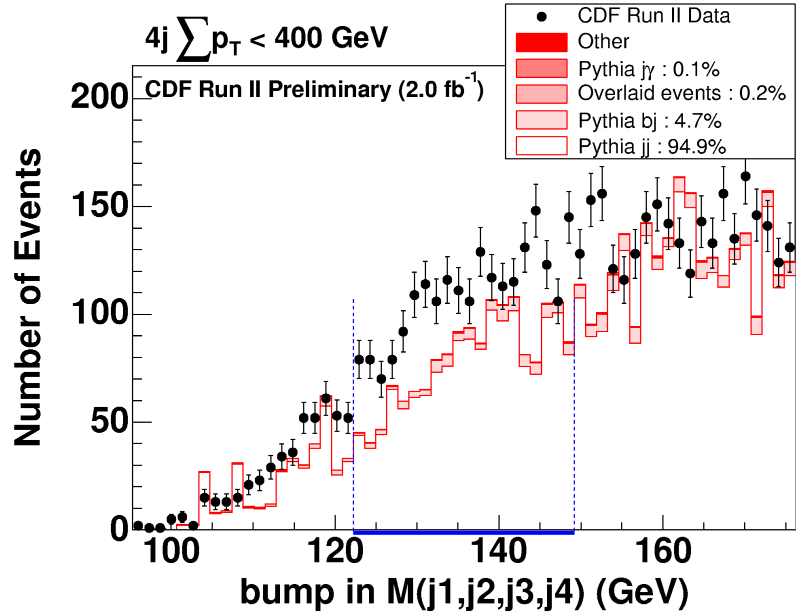
Summary of Mass Bumps



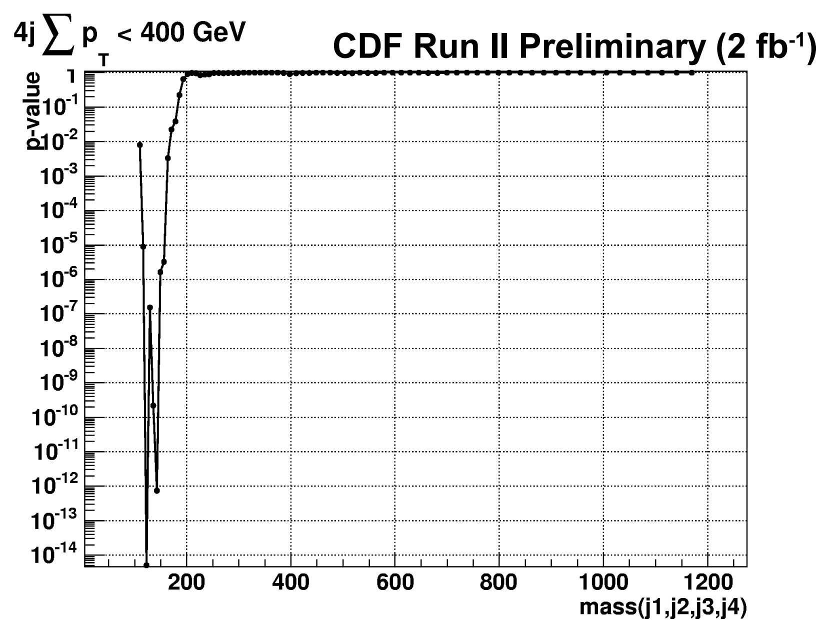
The outlier



$4j \sum p_T < 400$ GeV
CDF Run II Preliminary (2.0 fb^{-1})
CDF Run II Data
Other
Pythia jγ : 0.1%
Overlaid events : 0.2%
Pythia bj : 4.7%
Pythia jj : 94.9%
Number of Events
M(j1,j2,j3,j4) (GeV)

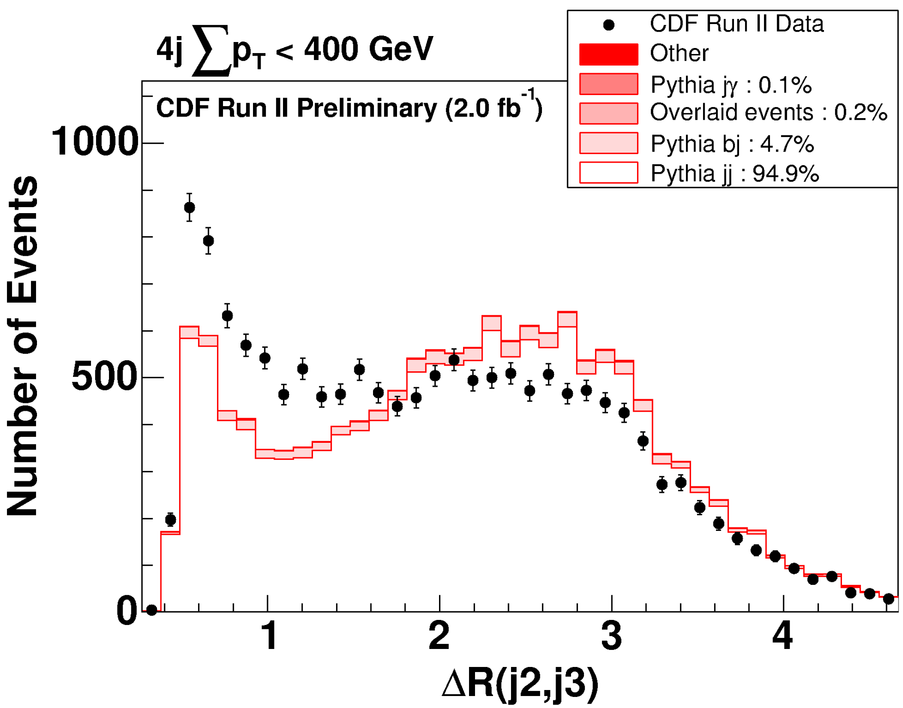


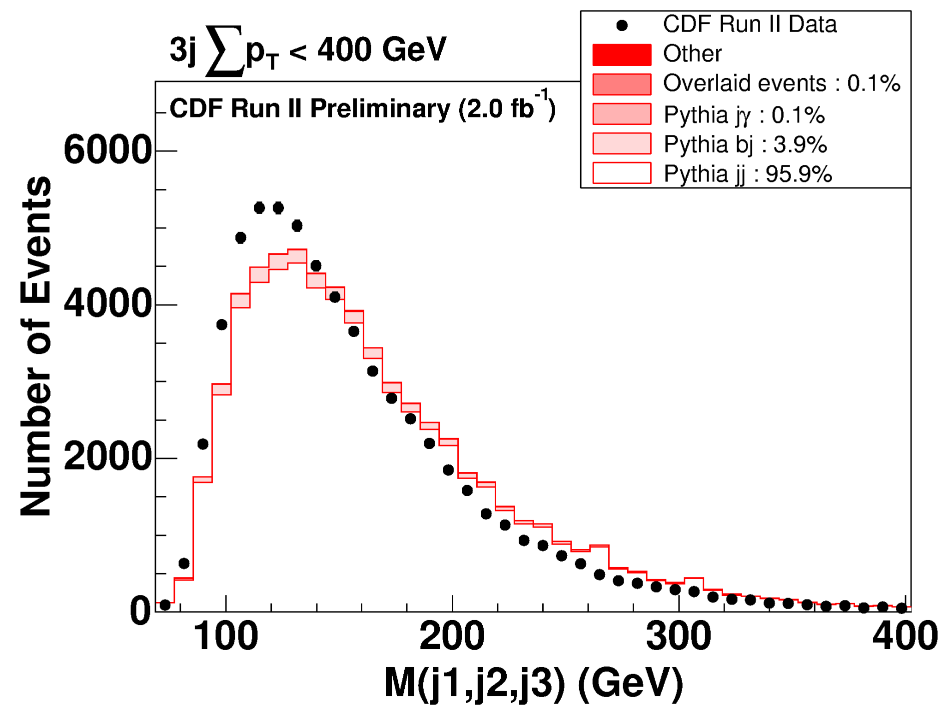
$4j \sum p_T < 400$ GeV
CDF Run II Preliminary (2.0 fb^{-1})
CDF Run II Data
Other
Pythia jγ : 0.1%
Overlaid events : 0.2%
Pythia bj : 4.7%
Pythia jj : 94.9%
Number of Events
bump in M(j1,j2,j3,j4) (GeV)



$4j \sum p_T < 400$ GeV
CDF Run II Preliminary (2 fb^{-1})
p-value
mass(j1,j2,j3,j4)

Attributed to PYTHIA's parton showering





The infamous "3-jet" effect, seen first in the 3-jet final state.
The data have more "close-together" jets than PYTHIA predicts.

More details in Si Xie's talk.

Summary

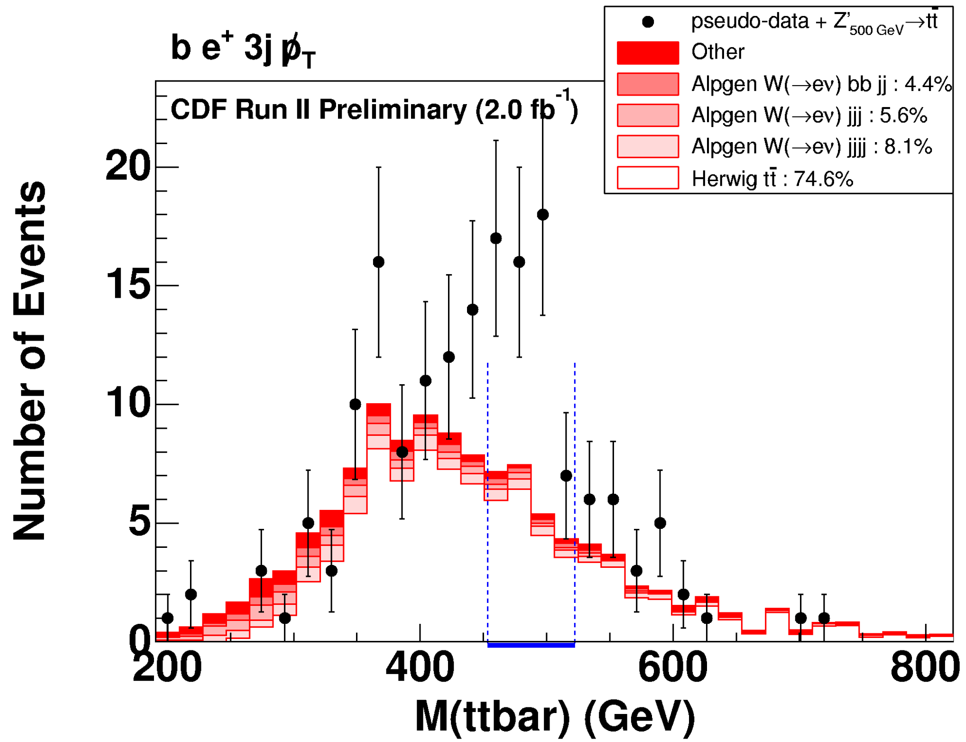
- Bump Hunter : A new algorithm, suited for global searches
- Used to expand the CDF Global Search (a.k.a. Vista/Sleuth)
- Scanned about 5,000 mass distributions found in 399 final states

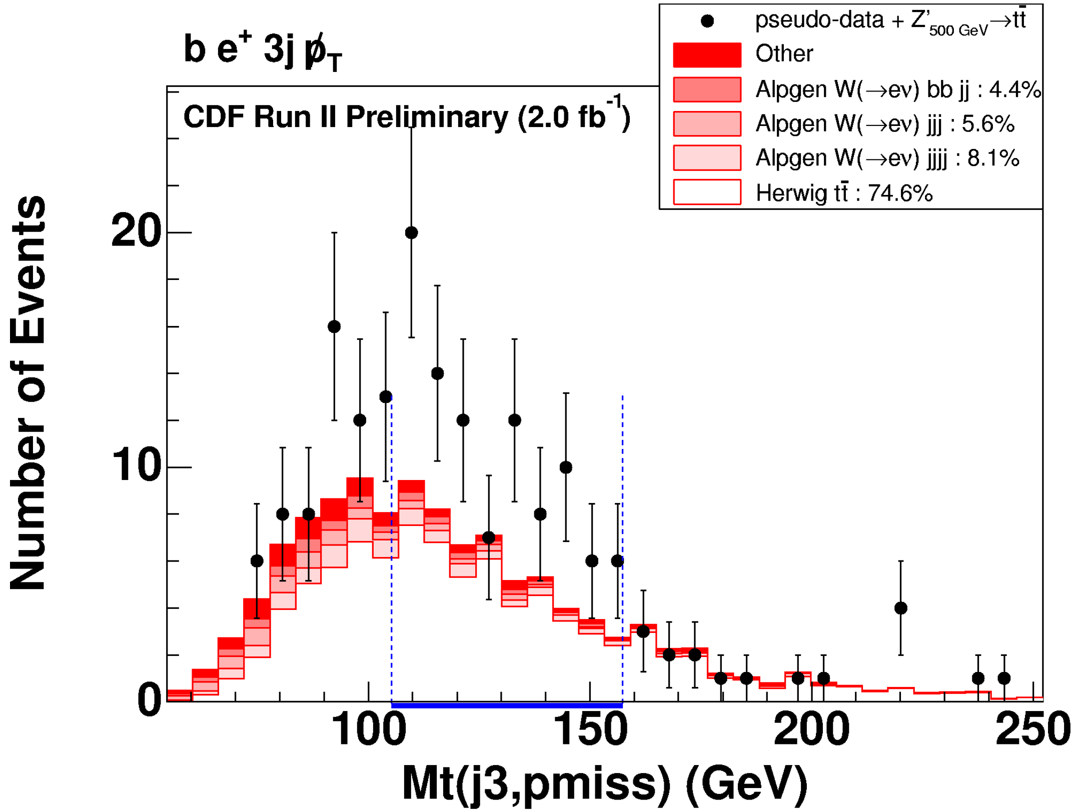
- Remarkable agreement with the SM in “smoking gun” final states (such as e^+e^- and $\mu^+\mu^-$)
- One outlier found, but attributed to the “3-jet” effect

- Unfortunately, this tool found no new physics in 2 fb^{-1} of CDF II data.

Appendix

Z' (500GeV) → $t\bar{t}$

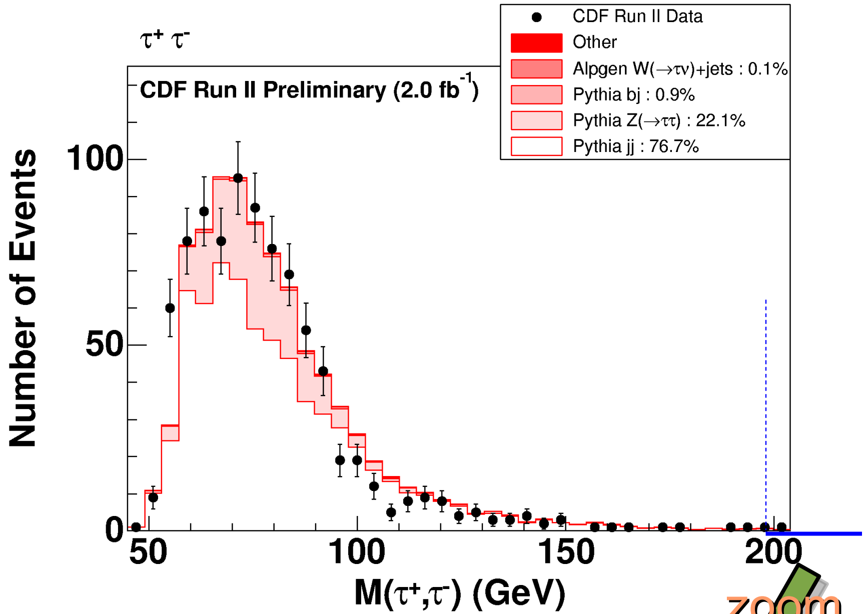


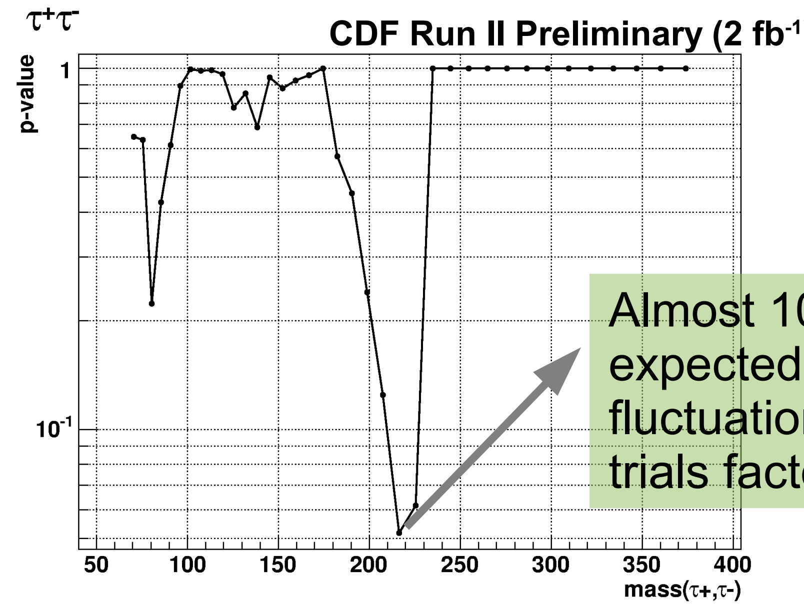


5σ discovery if σ×BR=2.2pb

The same signal can appear in different variables.

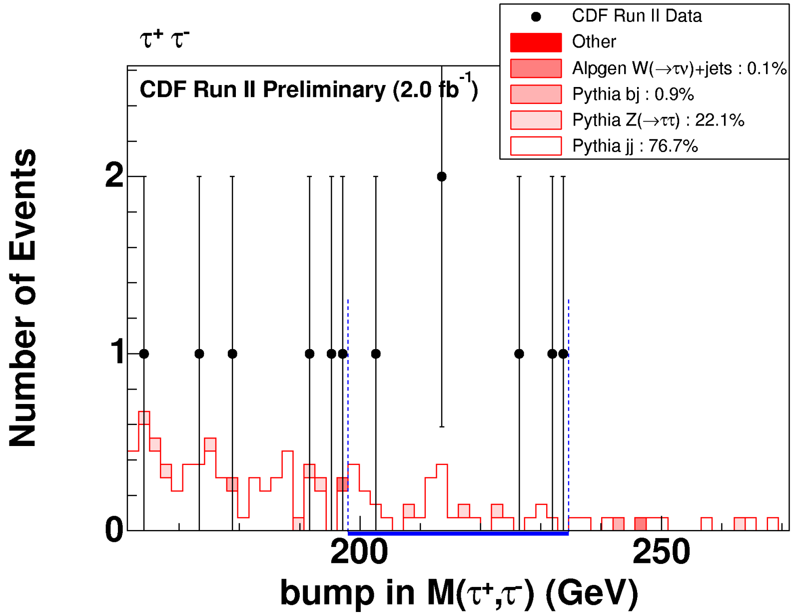
same for $\tau^+\tau^-$



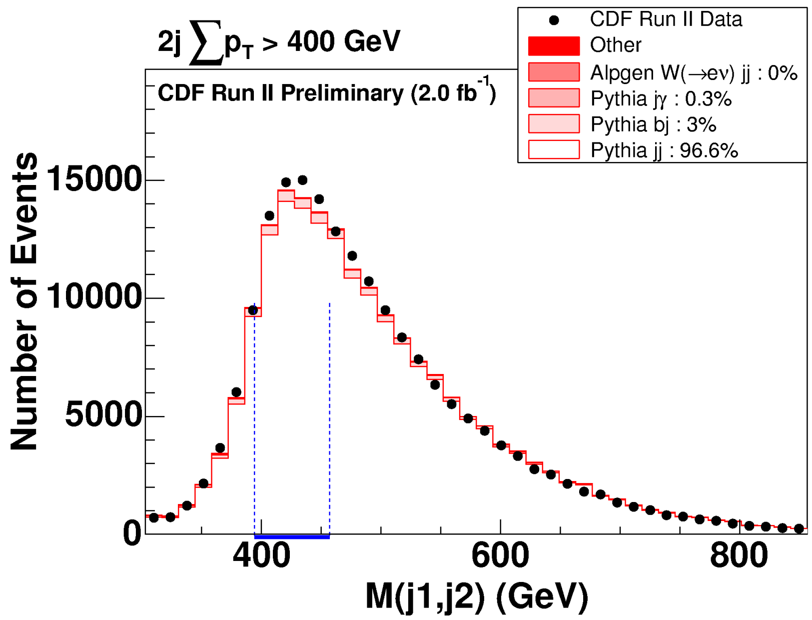


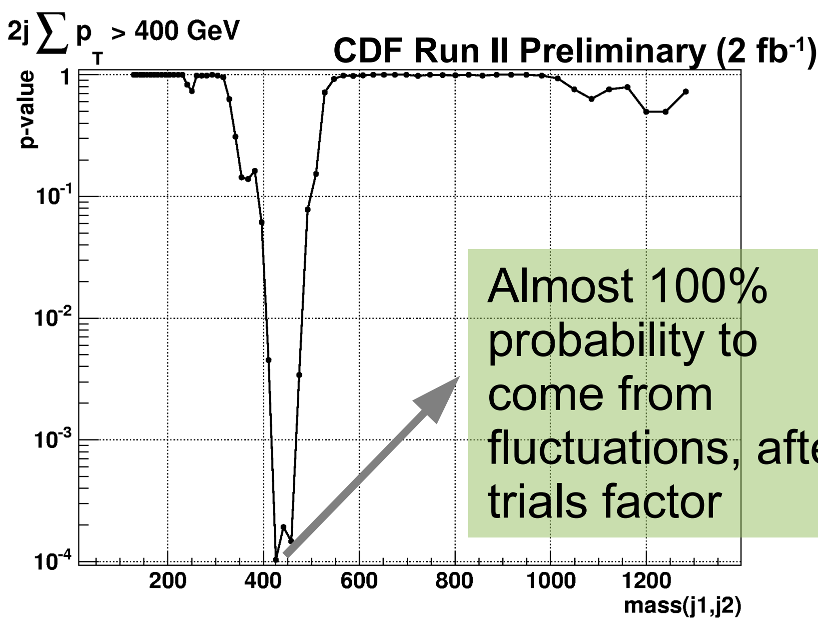
Almost 100% expected from fluctuations, after trials factor

zoom



high-Σp_T dijets?

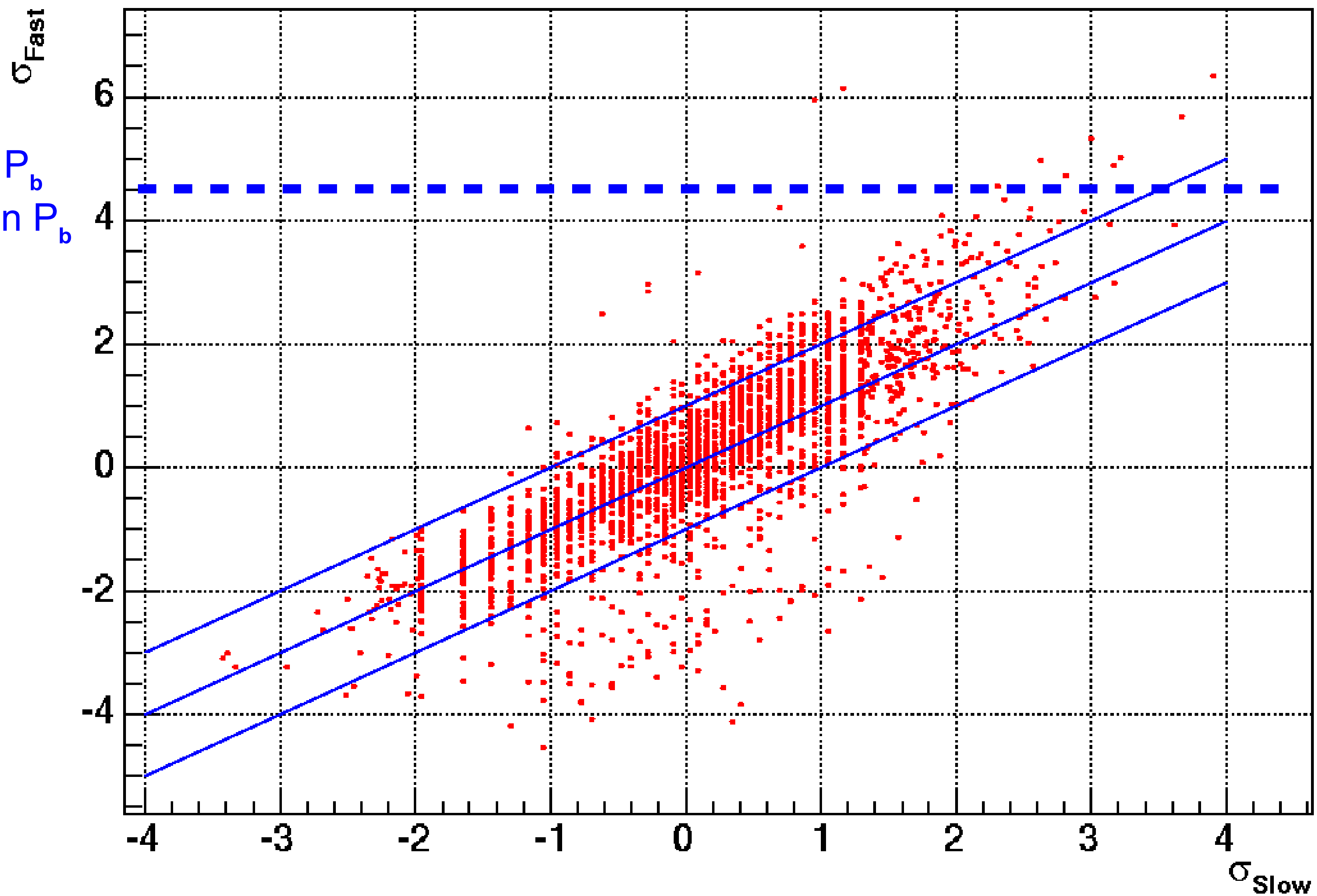




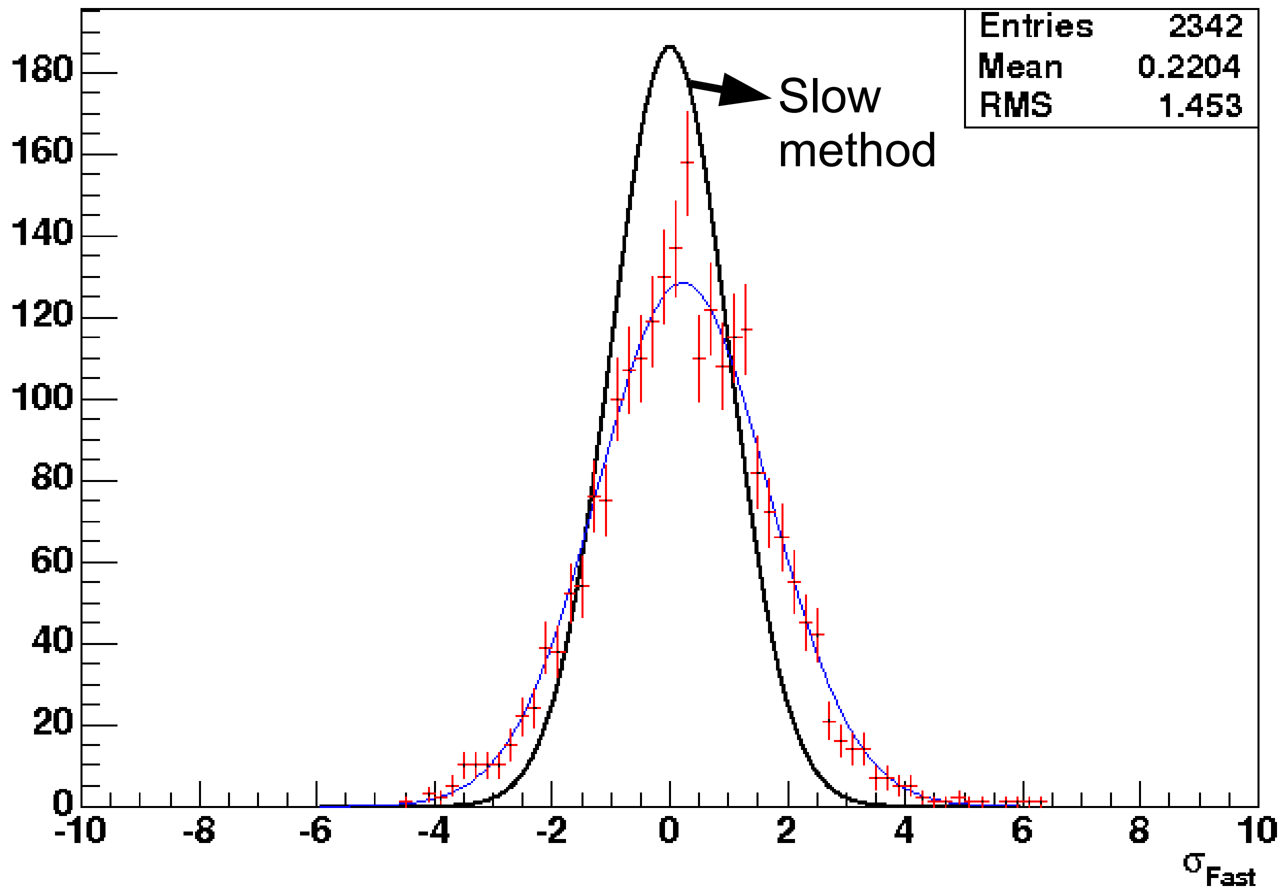
Almost 100% probability to come from fluctuations, after trials factor

Fast vs Slow method to estimate P_a

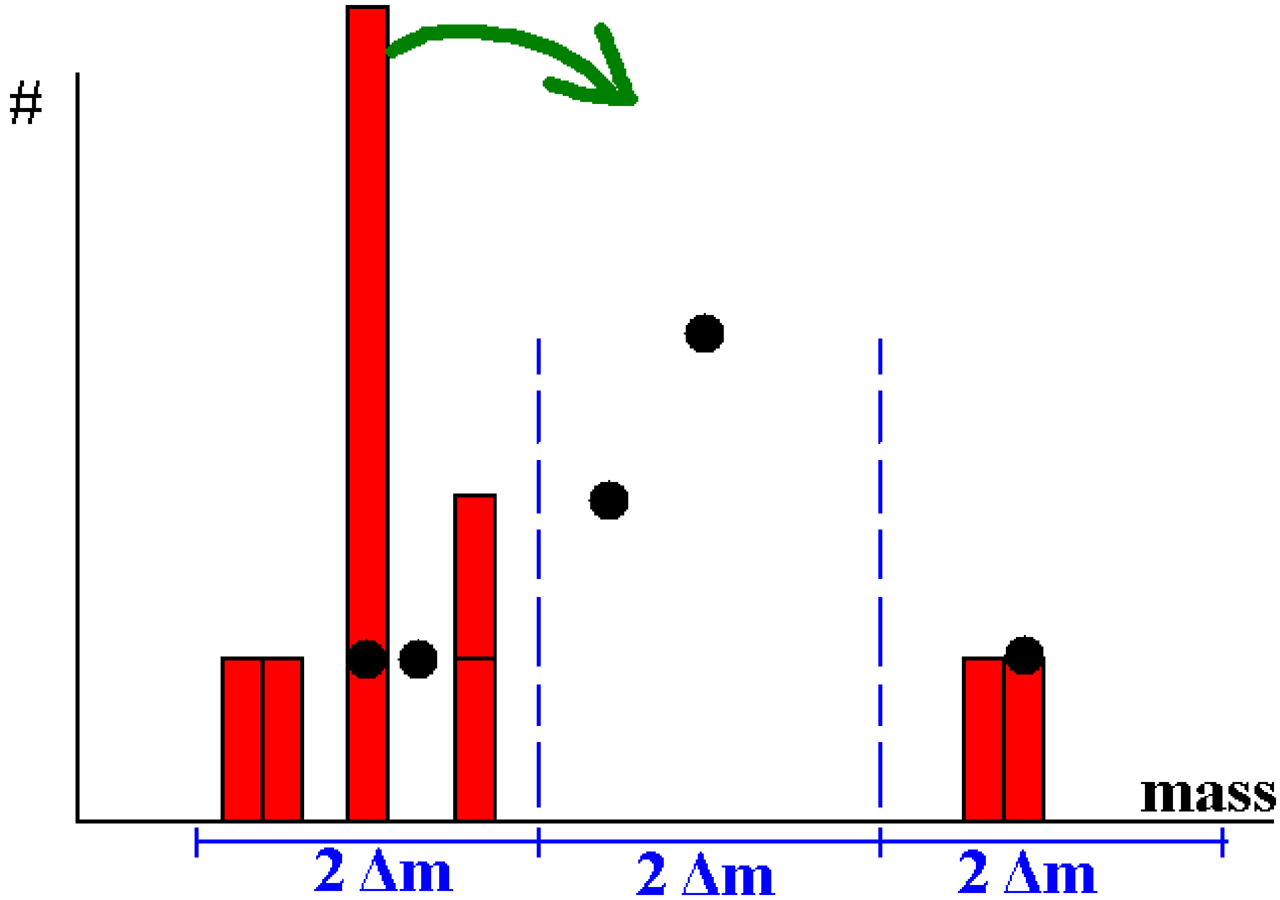
5σ in $P_a \rightarrow 3\sigma$ in P_b
4.5σ in $P_a \rightarrow 2.1\sigma$ in P_b



Expected P_a


Entries 2342
Mean 0.2204
RMS 1.453
Slow method
180
160
140
120
100
80
60
40
20
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
σ_{Fast}

The need for spike treatment



Spikes

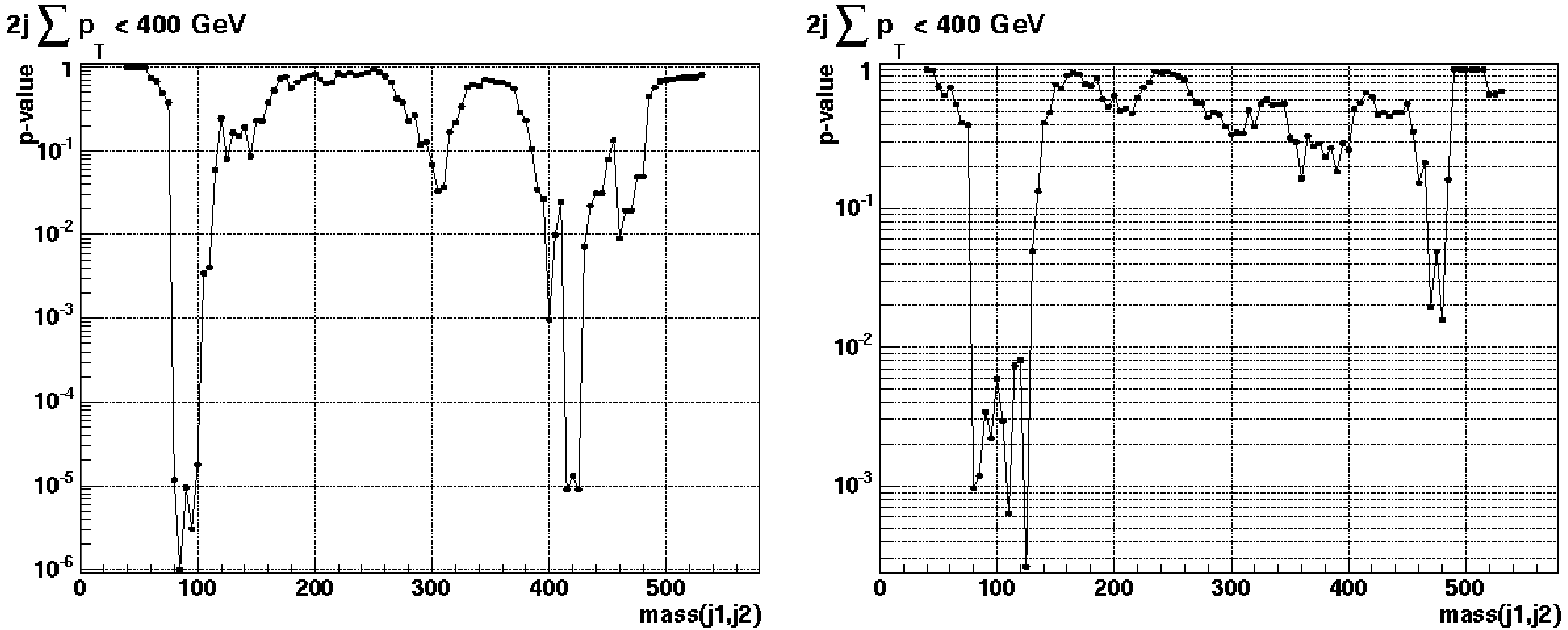


Figure 4-10: *(Left)* The p-val of each bump candidate, as a function of the location of each window's center, along mass$(j1, j2)$ in final state $2j\ \sum p_T < 400$ GeV. Bump candidates failing quality criteria have p-val=1. The most significant bump has p-val $\sim 10^{-6}$, which translates to $P_a \sim 3 \times 10^{-5}$ and $P_b \sim 0.15$, therefore all local excesses are insignificant. (Right) For demonstration, we apply the conservative anti-spike treatment to all bump candidates. The result of anti-spike treatment is to have larger p-values and the reduction of significance is greater in regions like around 400 GeV, where Monte Carlo statistics are poorer, therefore spikes contribute more.

Potential for Improvement

- Search for wider resonances
- Combine leptons & jet multiplicities
- Dynamic optimization of window width
- Use of only data