New Physics in Old Data? Localized 4σ and 5σ Dijet Mass Excesses in ALEPH LEP2 Four-Jet Events

Jen Kile

Pheno, May 7, 2018

Based on 1706.02242, 1706.02269, 1706.02255, JK, J. von Wimmersperg

How this situation arose

- LEP was an e^+e^- collider that ran at the $\sqrt{s} = M_Z$ (LEP1) and then at higher energies (LEP2, $\sqrt{s} = 130 - 209$ GeV). Many BSM searches, precision measurements at Z pole. Stopped running in 2000 to make way for LHC.
- Aleph left behind data policy allowing former members to do analyses, so we got ahold of the data.
- We had no concrete plan on what we were going to do. (Ideas: leptonically-interacting DM, excited leptons.)
- Did data-MC comparisons with very loose cuts (hadronic preselection) to make sure code working properly, MC normalized correctly, etc.
- Added all LEP2 data (~ 735 pb^{-1}) together on one plot.
- Clustered events into 4 jets, paired jets to minimize dijet mass difference. Plotted $\Sigma = (M_1 + M_2)/2$. (Dijet 1 defined to contain most energetic jet.)
- Σ has better resolution than M_1, M_2. Gives good differentiation between $e^+e^- \to Z/\gamma \to$ hadrons and $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$.

How this situation arose

- And we saw a ($\sim 3\sigma$) bump around $\Sigma \sim 55$ GeV.
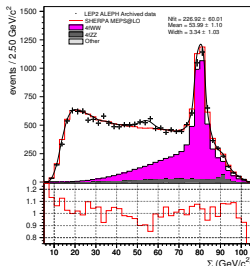
(This plot after
refinements in analysis.)

- Not the first time a bump had been seen in this location:
  Aleph had seen a 55-GeV bump in a search for hA at $\sqrt{s} = 130 - 136$
  GeV (5.7 pb^{-1}) (Buskulic, Z. Phys. C 71, 179).
  Not confirmed by other experiments, not seen in higher energy data.

- We decided this merited further study, started developing analysis
  tools.

How this situation arose

- 6 months later, we saw the LEP2 ADLO charged Higgs combination

$m_{H^\pm}$ (GeV/c^2)	Br(H$^+\to\tau^+\nu$)	combined CL_b	ALEPH CL_b	DELPHI CL_b	L3 CL_b	OPAL CL_b
43.	0.0	0.998	(*)	0.99	(*)	0.96
55.	0.0	0.997	0.75	0.96	0.96	0.94
89.	0.35	0.988	0.98	0.63	0.88	0.80
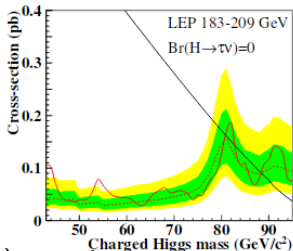
- All three other (DLO) experiments see an excess at 55 GeV. (Amusingly, Aleph least significant!)
  Only seen in purely hadronic $e^+e^- \to H^+H^- \to c\bar{s}\bar{c}s$ channel.
- Could be small remnant of our excess; further motivation to get to the bottom of this.

Challenges of this project

- We had no model with which to design analysis cuts (no signal MC).
- Had to develop analysis machinery (choose QCD simulation, preselection cuts, jet clustering & jet rescaling algorithms) after knowing of excess.
- If experiment were still running, solution would be to design cuts to select excess, take more data, see if excess persists in new data.
- Not possible here. But, there are 3 other LEP2 datasets (DLO).
- Our strategy: catalog features of excess, hope DLO can perform blind analyses.
- To prevent bias:
  We do not design cuts to select excess.
  Give nominal results for the analysis choices we made, and see how results would change if we made other choices (QCD simulation, jet clustering algorithms, etc).
  Try to present as many results as possible.

MC Simulation of SM QCD

- Almost all of the SM production at preselection level is
  $e^+e^- \to Z/\gamma \to q\bar{q} \to$ hadrons (QCD),
  & electroweak processes, such as
  $e^+e^- \to W^+W^-, ZZ \to q\bar{q}q\bar{q}$.
- The electroweak processes are pretty easy to simulate well; tend to have well-separated jets.
- The QCD is much messier: $q\bar{q}$ radiates gluons at low energies, angles.
- In LEP era, QCD simulated by interfacing $e^+e^- \to Z/\gamma \to q\bar{q}$ ME to parton shower. (Also some matching to $q\bar{q}g$ ME). No 4-parton ME.
- MC generators have advanced a lot since.
- Forcing events into 4 jets; very desirable to have 4-parton ME.
- Sherpa lets us do this; can generate MEs for final states with different numbers of partons, interface them to parton shower, correctly merge them into one inclusive sample.
- We simulated QCD w/Sherpa using MEs for final states of up to 6 partons.

MC tuning

- But, shouldn't just use MC generator straight out of the box.
- They have parameters that can be dialed which control e.g. showering and hadronization. (*Tuning.*)
- We did 2 tunes of Sherpa using MEs for up to 6 final-state partons.
- For first (LO) tune, all MEs were at LO.
- For other (NLO), MEs for final states of up to 4 partons were at NLO using BlackHat.
- PYTHIA used for hadronization.
- We placed an emphasis on being able to reproduce event-shape variables and multi-jet distributions.
- Compare MC generated with our 2 tunes to data, to each other, and to QCD generated with KK2f, a LEP-era parton shower.

MC tuning

Inter-jet angles particularly important. *Closely* related to reconstructed dijet masses. At preselection level:
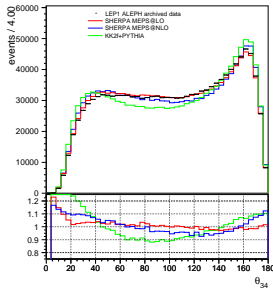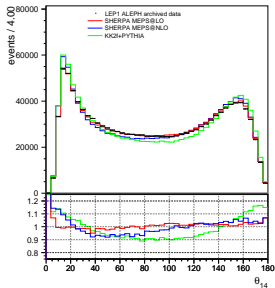
Sherpa performs much better on distributions directly related to clustering events into four jets.

Event-shape variables similar for Sherpa, KK2f.

Overall, LO tune seemed to have best agreement with data.

Used LO tune, reweighted using LEP 1 data & MC, as our QCD estimation.

Analysis Choices

- Preselection:
  - Want to retain hadronic events while removing two-photon events and events with hard ISR.
  - Most cuts identical to or very similar to standard cuts in ALEPH 4-jet analyses.
- Jet clustering:
  - Most common jet-clustering algorithm at LEP was Durham.
  - We use the jet-clustering algorithm Luclus, as it has been noted to have better resolution for jet angles and energies.
- Jet rescaling:
  - At a lepton collider, after cuts on initial-state radiation, you have a pretty good handle on the center-of-mass energy of an interaction.
  - Directions of jets well-measured (typical resolution $\sim 1°$).
  - Resolution on masses greatly improved if jet four-momenta are rescaled so that they add up to $(\sqrt{s}, 0, 0, 0)$. Directions kept fixed.
  - Rescaled momenta such that jet masses were kept constant. Constant-velocity rescaling more common.

Our nominal result

Using above choices for QCD MC sample, preselection cuts, and jet clustering and rescaling, we see this in the M_1-M_2 plane at LEP2:
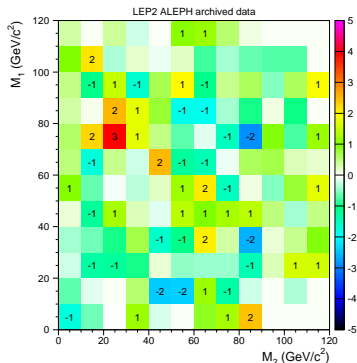
Significance of data-MC.

M_1 = mass of dijet containing most energetic jet.

Systematics not included.

Excess at $M_1 + M_2 \sim 110$ GeV, with concentration around $M_1 \sim 80$ GeV, $M_2 \sim 25$ GeV ("Region A").
$M_1 \sim M_2 \sim 55$ GeV ("Region B").
We fit the excess in each Region to a 2D gaussian.

Our nominal result

Our nominal result, without systematic uncertainties:

Parameter	Value ALEPH Archived Data
N_A	121 ± 33
μ_A	53.1 ± 1.7
δ_A	53.2 ± 2.3
$\sigma_{\Sigma A}$	5.80 ± 1.28
$\sigma_{\Delta A}$	7.04 ± 2.71
p-value(Region A)	$1.8^{+0.6}_{-0.5} \times 10^{-8}$
Significance(Region A)	$5.51^{+0.06}_{-0.05}\sigma$
N_B	138 ± 43
μ_B	54.6 ± 0.9
$\sigma_{\Sigma B}$	2.38 ± 0.75
$\sigma_{\Delta B}$	21.1 ± 3.7
p-value(Region B)	$1.62^{+0.02}_{-0.01} \times 10^{-5}$
Significance(Region B)	4.2σ

Excess roughly split between two Regions, but Region B very wide in $\Delta \equiv M_1 - M_2$.

Other possible analysis choices

- We have 6 different estimations of the SM QCD (LO Sherpa, NLO Sherpa, KK2f, all either reweighted with LEP1 data/MC or not). Significant excesses remain for all choices of QCD simulation.

- We tried 6 different jet-clustering algorithms.
  Significant excess in both regions for all six algorithms!
  Whatever is happening in the data is not just exploiting a feature of a single algorithm.
  We also checked what happens to the excess if Luclus and Durham agreed/disagreed on $(M_1 + M_2)/2$: agreement within 5 GeV seems to select excess–may hint that excess is more four-jetty than expected from QCD.

- Also looked at fixed-velocity jet rescaling. Excesses still significant, but reduced (4.8σ and 3.3σ)–may give some clue about jet structure of these events.

Systematics

We break sources of systematic uncertainty into 2 groups:

- Those which can be estimated by comparing QCD MC samples (omitted higher-order terms, merging scale, value of α_s, showering, QCD MC statistics).
- Those which cannot (Luminosity, cross-sections, beam backgrounds, modelling of photons, hadronization, all 4-fermion MC uncertainties).

Latter set requires dedicated studies of the uncertainties.

When corrections can be reliably estimated, they are applied.

Systematics from QCD hadronization difficult to estimate reliably, and likely dominant. Quote final results as function of hadronization uncertainty.

Results

Final results:

Hadronization uncertainty	0%	$\pm 1\%$	$\pm 2\%$	$\pm 3\%$
Toy MC generated	8.0×10^8	4.0×10^8	1.2×10^8	4×10^7
p-value$/(10^{-8})$ (Region A)	1.6 ± 0.5	$2.5^{+1.0}_{-0.8}$	16^{+4}_{-3}	113 ± 17
Significance (Region A)	$5.53^{+0.06}_{-0.05}$	$5.45^{+0.07}_{-0.06}$	$5.11^{+0.05}_{-0.04}$	4.73 ± 0.03
p-value$/(10^{-6})$ (Region B)	4.2 ± 0.1	6.7 ± 0.1	21.9 ± 0.4	N/A
Significance (Region B)	4.5	4.4	4.1	N/A

Significance ranges from 4.7σ to 5.5σ for Region A, 4.1σ to 4.5σ for Region B.

Sanity Checks

Some basic checks:

- Excess events showing activity in any particular part of detector?
  $\rightarrow$ Polar, azimuthal angles of thrust axis in accord with background.
- Missing energy in events?
  $\rightarrow$ No evidence of anomalous $\not{E}$, jet rescaling factors also looked fine. But, can't rule out invisible object w/$E \lesssim$ few GeV.
- Regions A and B separate excesses, or 1 continuous excess?

Take 45 GeV $< \Sigma < 61$ GeV,
look at $\Delta = M_1 - M_2$:
$\rightarrow$Hard to say anything conclusive.
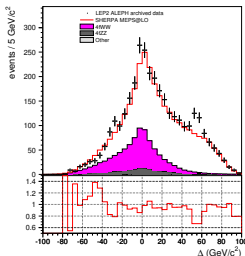
- Excess dependence on $\sqrt{s}$?
  $\rightarrow$ Compatible w/being proportional to QCD bkg, but error bars large.

Basic features of the excess

What do these events look like? No detailed study yet, but:

- Topology of Region A events is very $1 - 3$, with one jet of $E \sim \sqrt{s}/2$ in one hemisphere, and three jets in the other.
- Jets paired to minimize dijet mass difference. Sensible for 55-GeV pair-production; less so for production of 80-GeV, 25-GeV resonances.
- Also, decay angle θ_{dec} of 80-GeV dijet is strongly peaked near 0:
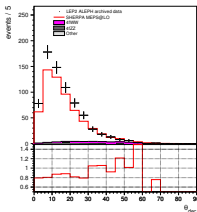
- Both features consistent w/less-energetic jet in the 80-GeV system being softer than would be expected from a genuine 80-GeV particle.
- We caution against jumping to the conclusion that Region A is indicative of 80-GeV, 25-GeV resonances.

Conclusions

- Hadronic events in Aleph archived LEP2 data show excesses relative to MC simulation: Region A: 4.7σ-5.5σ, Region B: 4.1σ-4.5σ (local)
- No analogous feature at LEP1.
- Excesses extremely robust against changes in SM QCD MC simulation, jet-clustering algorithm.
- Excess events (especially Region A) look much like SM QCD. Most reasonable conventional hypothesis is residual QCD mismodelling.
- Important to understand this, whether it is BSM physics or QCD mismodelling. *Will not be easy.*
- We hope we can convince the other LEP experiments to attempt to confirm/refute our results. INPUT WELCOME!
- Would like QCD experts to weigh in on whether or not further MC tuning (or other simulation modifications) could reproduce LEP2 data without ruining agreement at LEP1. INPUT WELCOME!
- If answer to latter question is negative, need to consider new physics explanations.

MC tuning

First, we compared to unfolded LEP1 and LEP2 data.

Which sample was best varied depending upon the variable being examined (or even what part of the range one is looking at).
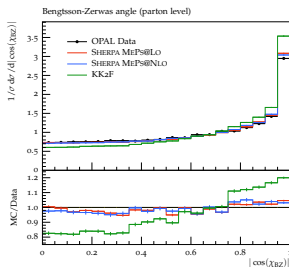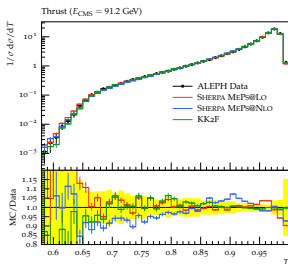
Sherpa generations better for distributions related to clustering into jets.

On other event-shape variables, older KK2f generation similar to new MC.

Overall, LO tune seemed to give the best agreement with data.

Our nominal result

The fitted S/B:
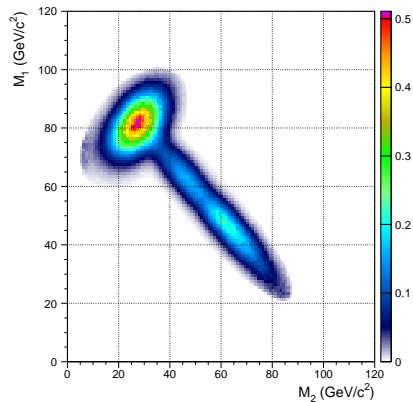
SM:
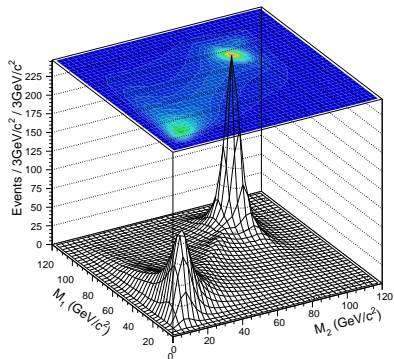
MC reweighting

What we really care about are dijet masses M_1, M_2.
Look at ratios of MC expectation in M_1-M_2 plane:

$$\frac{\text{LO Sherpa}}{\text{KK2f}} \qquad \frac{\text{NLO Sherpa}}{\text{KK2f}} \qquad \frac{\text{LO Sherpa}}{\text{NLO Sherpa}}$$

LEP1

LEP2

Similar behavior at LEP1, LEP2, up to overall energy scale.
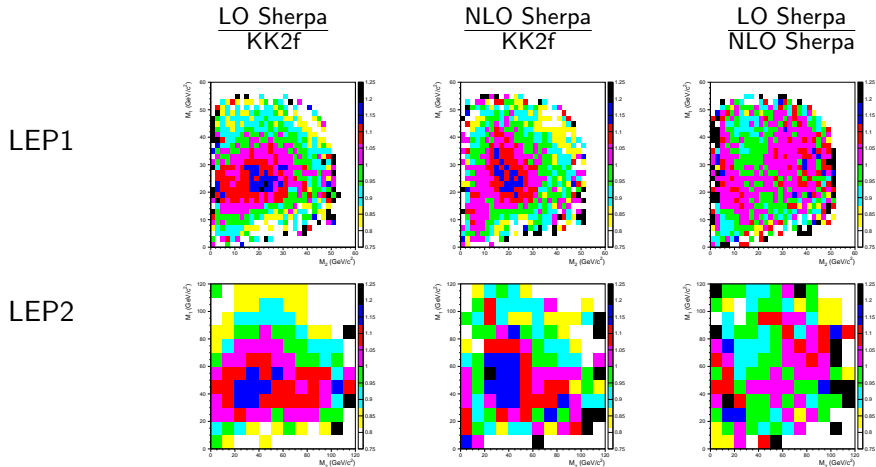
MC reweighting

Comparison of LEP2 MC samples in M_1-M_2 plane after reweighting procedure:

$$\frac{\text{LO Sherpa}}{\text{KK2f}} \qquad \frac{\text{NLO Sherpa}}{\text{KK2f}} \qquad \frac{\text{LO Sherpa}}{\text{NLO Sherpa}}$$
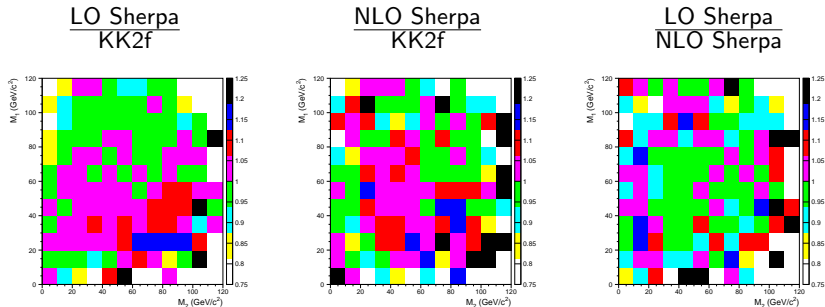
Disagreement between KK2f and other samples much reduced, take as indication reweighting reduces systematics.

Will take reweighted LO Sherpa as our most trusted SM QCD estimation, retain other samples for systematic studies.

Preselection

Purpose of preselection is to retain hadronic events while removing two-photon events and events with hard ISR. LEP1 cuts:

- Require 7 good charged tracks in event.
- Force events into 4 jets; require each jet have at least one good charged track.
- Sum of jet transverse momenta $p_{tsum} > 25\%\sqrt{s}$.
- Rescale energy, momenta of four jets, keeping directions and masses fixed, so that sum of four-momenta are $(\sqrt{s}, 0, 0, 0)$. Require all rescaling factors to be positive.

LEP2 cuts:

- above cuts
- No jet with more than 80% of its EM energy in a $1°$ cone around an energy flow object.
- $|p_{zmis}| < 1.5(m_{vis} - 90)$.

Most of our preselection cuts are identical to or very similar to preselection cuts used by other Aleph analyses.

Changes in QCD simulation

We have 6 different estimations of the SM QCD (LO Sherpa, NLO Sherpa, KK2f, all either reweighted with LEP1 data/MC or not).

Parameter	LO_{unrew}	LO_{rew} (Nominal)	NLO_{unrew}	NLO_{rew}	$KK2f_{unrew}$	$KK2f_{rew}$
N_A	131 ± 34	121 ± 33	92 ± 25	111 ± 26	110 ± 32	118 ± 31
μ_A	53.6 ± 1.7	53.1 ± 1.7	53.2 ± 2.6	53.5 ± 2.0	53.3 ± 1.7	52.8 ± 1.7
δ_A	53.3 ± 2.1	53.2 ± 2.3	54.2 ± 1.8	53.8 ± 1.6	52.2 ± 2.2	53.3 ± 1.9
$\sigma_{\Sigma A}$	5.69 ± 0.98	5.80 ± 1.28	6.30 ± 1.47	6.45 ± 1.20	5.40 ± 1.01	5.67 ± 1.01
$\sigma_{\Delta A}$	7.15 ± 2.10	7.04 ± 2.71	4.93 ± 1.05	5.59 ± 0.94	6.48 ± 1.91	6.30 ± 1.78
p-value/(10^{-8}) (Region A)	< 0.8(stat) 0.53(ex)	$1.8^{+0.6}_{-0.5}$	360 ± 28	$7.7^{+2.9}_{-2.5}$	14 ± 3	$2.4^{+1.6}_{-1.0}$
Significance (Region A)	> 5.66(stat) 5.72 (ex)	$5.51^{+0.06}_{-0.05}$	4.49	$5.25^{+0.07}_{-0.06}$	$5.13^{+0.05}_{-0.03}$	5.5 ± 0.1
N_B	109 ± 70	138 ± 43	100 ± 33	125 ± 37	203 ± 56	154 ± 56
μ_B	54.9 ± 3.0	54.6 ± 0.9	54.9 ± 0.8	54.7 ± 0.7	54.1 ± 1.7	54.7 ± 2.0
$\sigma_{\Sigma B}$	2.07 ± 0.85	2.38 ± 0.75	1.70 ± 0.67	1.80 ± 0.63	3.04 ± 0.64	2.38 ± 0.76
$\sigma_{\Delta B}$	21.1 ± 4.2	21.1 ± 3.7	20.6 ± 4.0	20.4 ± 3.5	20.3 ± 3.0	21.7 ± 3.4
p-value/(10^{-6}) (Region B)	236 ± 1	$16.2^{+0.2}_{-0.1}$	266 ± 2	$24.2^{+0.5}_{-0.4}$	$7.7^{+2.5}_{-2.2} \times 10^{-2}$	1.8 ± 0.1
Significance (Region B)	3.5	4.2	3.5	4.1	5.25 ± 0.06	4.6

All show highly significant excesses.
Results which differ most from Nominal are from unreweighted samples.

Changing the jet-clustering algorithm

We tried 6 different jet-clustering algorithms.

| Parameter	LUCLUS (Nominal)	DURHAM	JADE	DICLUS	DMLR	LMNR
N_A	121 ± 33	134 ± 44	81 ± 24	128 ± 31	134 ± 37	119 ± 32
μ_A	53.1 ± 1.7	57.4 ± 2.2	56.4 ± 1.5	56.8 ± 1.5	54.2 ± 2.4	54.5 ± 1.8
δ_A	53.2 ± 2.3	55.7 ± 4.9	56.1 ± 2.5	52.1 ± 2.0	51.8 ± 2.3	54.3 ± 2.3
$\sigma_{\Sigma A}$	5.80 ± 1.28	5.97 ± 1.38	4.84 ± 1.51	5.04 ± 1.17	7.04 ± 1.85	5.99 ± 1.19
$\sigma_{\Delta A}$	7.04 ± 2.71	13.54 ± 4.52	7.18 ± 1.74	6.69 ± 1.59	7.20 ± 2.12	7.28 ± 1.78
p-value/(10^{-8}) (Region A)	$1.8^{+0.6}_{-0.5}$	293 ± 16	25.2 ± 4.7	$< 1.1\text{(stat)}$ 0.50(ex)	$13.0^{+3.8}_{-3.2}$	$6.1^{+2.9}_{-2.4}$
Significance (Region A)	$5.51^{+0.06}_{-0.05}$	4.5	$5.02^{+0.04}_{-0.03}$	$> 5.59\text{(stat)}$ 5.73(ex)	5.15 ± 0.05	5.3 ± 0.1
N_B	138 ± 43	139 ± 68	102 ± 43	113 ± 47	118 ± 46	143 ± 43
μ_B	54.6 ± 0.9	53.1 ± 2.2	55.3 ± 1.3	58.5 ± 1.6	52.7 ± 1.5	54.2 ± 1.2
$\sigma_{\Sigma B}$	2.38 ± 0.75	2.51 ± 0.62	2.31 ± 1.00	1.48 ± 0.22	3.02 ± 0.95	2.96 ± 0.72
$\sigma_{\Delta B}$	21.1 ± 3.7	22.5 ± 5.5	18.7 ± 5.4	16.8 ± 3.8	20.4 ± 4.7	16.7 ± 4.7
p-value/(10^{-6}) (Region B)	$16.2^{+0.2}_{-0.1}$	15.9 ± 0.4	877 ± 3	$25.5^{+0.4}_{-0.5}$	531 ± 2	50 ± 1
Significance (Region B)	4.2	4.2	3.1	4.05 ± 0.01	3.3	3.9

Significant excess in both regions for all six algorithms!

Note: 5 of these are binary joining algorithms. Diclus clusters $3 \rightarrow 2$.
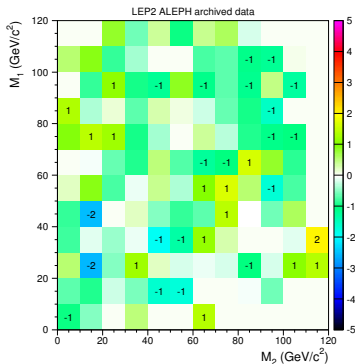
Whatever is happening in the data is not just exploiting a feature of a single algorithm.

Changing the jet-clustering algorithm

What happens if you ask that Luclus and Durham agree on Σ?
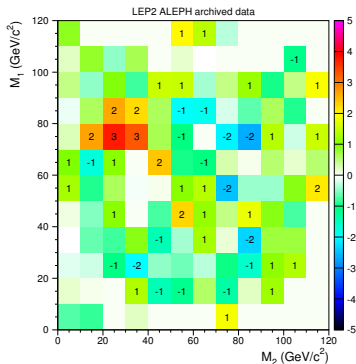
(c) $|\Sigma_{LUC} - \Sigma_{DUR}| < 5$ GeV

(d) $|\Sigma_{LUC} - \Sigma_{DUR}| > 5$ GeV

4-fermion MC tends to have better agreement between Luclus & Durham than QCD events do.

May indicate excess more "four-jetty" than expected from background.

Changing the jet-rescaling algorithm

We changed from fixed-mass jet rescaling to fixed-velocity jet rescaling:

Parameter	Fixed-Mass Rescaling (Nominal)	Fixed-Velocity Rescaling
N_A	121 ± 33	120 ± 42
μ_A	53.1 ± 1.7	54.4 ± 2.3
δ_A	53.2 ± 2.3	53.0 ± 2.0
$\sigma_{\Sigma A}$	5.80 ± 1.28	7.27 ± 2.11
$\sigma_{\Delta A}$	7.04 ± 2.71	6.84 ± 2.29
p-value$/(10^{-8})$ (Region A)	$1.8^{+0.6}_{-0.5}$	83.5 ± 7.0
Significance (Region A)	$5.51^{+0.06}_{-0.05}$	4.8
N_B	138 ± 43	126 ± 49
μ_B	54.6 ± 0.9	55.0 ± 1.6
$\sigma_{\Sigma B}$	2.38 ± 0.75	3.22 ± 0.95
$\sigma_{\Delta B}$	21.1 ± 3.7	21.2 ± 6.0
p-value$/(10^{-6})$ (Region B)	$16.2^{+0.2}_{-0.1}$	496.4 ± 1.7
Significance (Region B)	4.2	3.3

This does seem to reduce the significance in both regions.
Understanding this left for future work.
Perhaps gives some clue about jet structure of these events.

Corrections and uncertainties to MC samples:

Source	SHERPA		KK2f		Four-fermion
	LEP1	LEP2	LEP1	LEP2	LEP2 Only
Luminosity	$\pm 0.12\%$	$\pm 0.5\%$	$\pm 0.12\%$	$\pm 0.5\%$	$\pm 0.5\%$
Cross-section	$\pm 0.1\%$	$\pm 0.2\%$	$\pm 0.1\%$	$\pm 0.2\%$	$\pm 0.4\%$
MC Statistics	Included in MC comparison				$\pm 1.3\%$ (A) $\pm 0.38\%$ (B)
Reweighting	$\pm 0.3\%$ (A) $\pm 0.25\%$ (B)	N/A	$\pm 0.3\%$ (A) $\pm 0.25\%$ (B)	N/A	N/A
Beam Background	$-0.35 \pm 0.05\%$	$-0.50 \pm 0.10\%$	$-0.35 \pm 0.05\%$	$-0.50 \pm 0.10\%$	$-0.15 \pm 0.05\%$
LEP1 ISR	$-0.52 \pm 0.02\%$	N/A			
FSR (uncorrelated)	$-0.459 \pm 0.034\%$	$-1.77 \pm 0.16\%$	$+0.085 \pm 0.034\%$	$+0.59 \pm 0.13\%$	N/A
FSR (correlated)	$\pm 0.035\%$	$\pm 0.35\%$	$\pm 0.035\%$	$\pm 0.35\%$	N/A
$|p_{zmis}|$ cut variation	N/A	$\pm 0.07\%$	N/A	$\pm 0.07\%$	N/A
Hadronization	$\pm 0\%, \pm 1\%, \pm 2\%, \pm 3\%$ on LEP2 after reweighting				$\pm 0.2\%$

We vary the SM background expectation, taking above numbers as 1σ error bars.
Include correlations.