Localized 4σ and 5σ Dijet Mass Excesses in ALEPH LEP2 Four-Jet Events

Jen Kile

July 15, 2019

QCD@LHC

Buffalo, NY

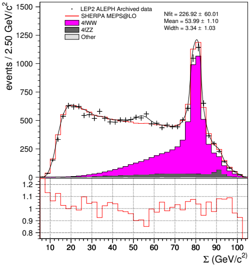
Based on JHEP 1810 (2018) 116, 1706.02242, 1706.02269,
JK, J. von Wimmersperg

How this situation arose

- LEP was an e^+e^- collider that ran at $\sqrt{s} = M_Z$ (LEP1) and then at higher energies (LEP2, $\sqrt{s} = 130 - 209$ GeV). 4 experiments: ADLO.
- Aleph left behind data policy allowing former members to do analyses, so we got ahold of the data.
- Initially had no concrete plan on what we were going to do.
- Did data-MC comparisons with very loose cuts (hadronic preselection) to make sure code working properly, MC normalized correctly, etc.
- Clustered events into 4 jets, paired jets to minimize dijet mass difference. Plotted $\Sigma = (M_1 + M_2)/2$. (M_1, M_2: dijet masses. 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 $e^+e^- \to W^+W^- \to qq\bar{q}\bar{q}$.
- Added all LEP2 data (~ 735 pb^{-1}) together on one plot.

How this situation arose

- And we saw a $\sim (3-4)\sigma$ bump around $\Sigma \sim 55$ GeV.

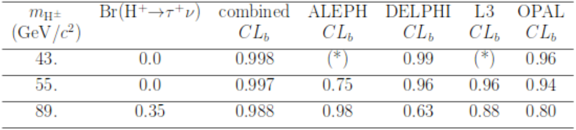


(This plot after refinements in analysis.)

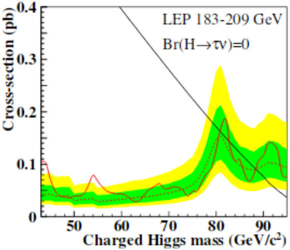
- Not the first time a bump had been seen in this location: Aleph had seen a 55-GeV bump in Σ in 4-jet search for *hA* at $\sqrt{s} = 130 - 136$ GeV (5.7 pb^{-1}) (Buskulic, Z. Phys. C 71, 179). Not confirmed by other experiments, not seen in higher energy data.
- We decided this merited further study, started developing analysis tools.

How this situation arose

- Afterward, saw the LEP2 ADLO charged Higgs combination (2013!)



$m_{H^\pm}$ (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).
- Strategy: catalog features of excess, give nominal results for the analysis choices we made (QCD simulation, jet clustering algorithms, etc), and see how results would change if we made other choices.
- Hope DLO can perform blind analyses.

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, e.g. $e^+e^- \to W^+W^-, ZZ \to q\bar{q}q\bar{q}$.
- The electroweak processes are pretty easy to simulate well; tend to have well-separated jets.
- QCD not so easy: $q\bar{q}$ radiates gluons at low energies, angles.
- Excess occurs in region dominated by $e^+e^- \to Z/\gamma \to$ hadrons.
- 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. Some MEs can be generated @NLO.
- We simulated QCD w/Sherpa using MEs for final states of up to 6 partons.

MC tuning

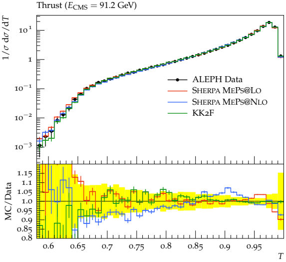
- QCD generators have parameters that can be dialed which control e.g. showering and hadronization.
- To optimally choose parameter values, need input from experiment: MC tuning!
  - Generate many (10^2-10^3) different MC samples, with different values of generator parameters.
  - Compare these to many distributions of observables in data.
  - Fit to find parameter values which produce MC with best agreement to data.
  - After you have your optimal parameter values, ready to generate MC.
- Used publicly-available unfolded LEP1 data in Rivet (Buckley et al, 2013; Cacciari et al, 2012). Fit done with Professor (Buckley et al, 2010).
- Unfolded data: experimental effects like selection cuts, detector effects, ISR removed to produced idealized data sample.

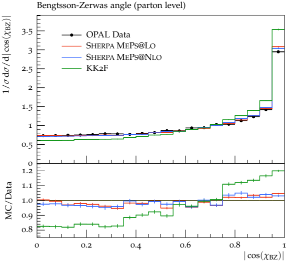
MC tuning

- 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 $q\bar{q}$ ME generator, interfaced to PYTHIA for parton shower.
- Compare at both Rivet level and at full simulation, at both LEP1 and LEP2.

MC tuning

Unfolded,
Rivet level:





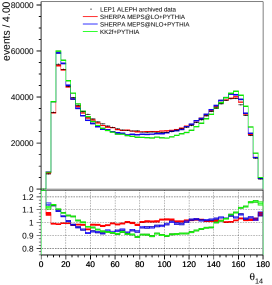
Sherpa generations better for distributions related to clustering into jets.

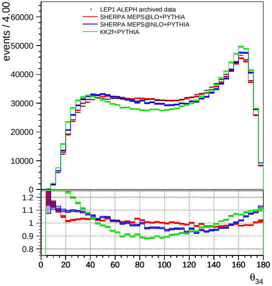
On other event-shape variables, older KK2f generation similar to new MC.

MC tuning

Inter-jet angles particularly important. *Closely* related to reconstructed dijet masses.

Full simulation:





Jets energy-ordered, Jet 1 most energetic.

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

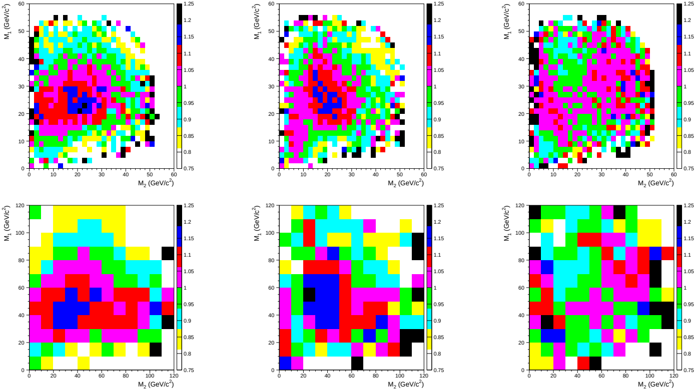
MC reweighting

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

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

LEP1

LEP2

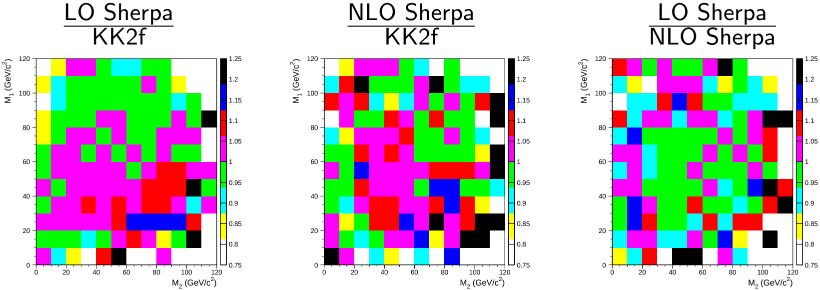


Similar behavior at LEP1, LEP2, up to overall energy scale.
Wanted to see if could use this to get even better MC.

MC reweighting

- Idea: Reweight LEP2 QCD MC samples with multiplicative factors which would bring MC, data into agreement at LEP1.
- Need to map M_1, M_2 at LEP1 to M_1, M_2 at LEP2. Except at low masses, map is pretty close to just multiplying by $\sqrt{s}/M_Z$.

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



Disagreement between KK2f and other samples much reduced, take as indication reweighting reduces systematics.

Take reweighted LO Sherpa as most trusted SM QCD estimation, retain others for systematic studies.

Preselection, jet clustering, and jet rescaling

Preselection:

- Want to retain hadronic events while removing two-photon events and events with hard ISR.
- Most cuts identical/similar to standard cuts in ALEPH 4-jet analyses.

Jet clustering:

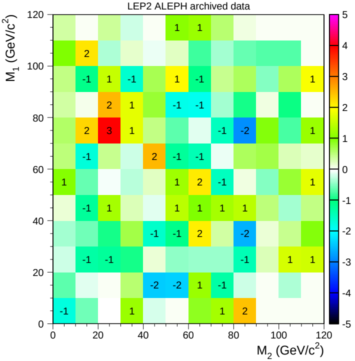
- We use the jet-clustering algorithm Luclus, as it has been noted to have better resolution for jet angles and energies.
- Most common algorithm at LEP was Durham; will compare several.

Jet rescaling:

- 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 (typical resolution $\sim 1^\circ$).
- For this reason, dijet masses determined mostly from inter-jet angles.
- Rescaled momenta such that jet masses were kept constant.

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").

No similar bump seen in LEP1 data.

Our nominal result

- We decided to fit the excess to two 2D gaussians.
- Gaussians in $\Sigma = (M_1 + M_2)/2$ and $\Delta = (M_1 - M_2)$, not M_1, M_2.
- Fitting to gaussians in Σ, Δ easier. Resolution in Σ better than that in Δ.

Our nominal result, without systematic uncertainties:

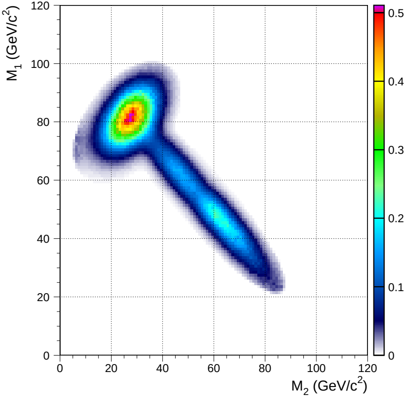
Parameter	Value ALEPH Archived Data
N_A	121 ± 33
μ_A	53.1 ± 1.7
δ_A	53.2 ± 2.3
$\sigma_{\Sigma A}$	5.80 ± 1.28
$\sigma_{\Delta A}$	7.04 ± 2.71
p-value(Region A)	$1.8^{+0.6}_{-0.5} \times 10^{-8}$
Significance(Region A)	$5.51^{+0.06}_{-0.05}\sigma$
N_B	138 ± 43
μ_B	54.6 ± 0.9
$\sigma_{\Sigma B}$	2.38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

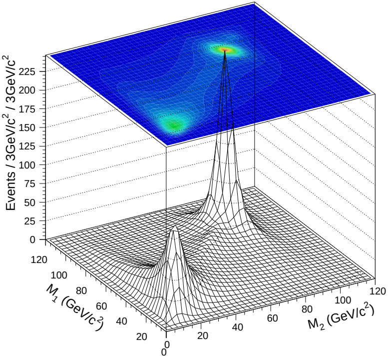
Excess roughly split between two Regions, but Region B very wide in Δ.

Our nominal result

The fitted S/B:

SM:





Changing the QCD Simulation and Jet Clustering Algorithm

We have 6 different estimations of the SM QCD. What if we switched to another one?

MC sample	LO_{unrew}	LO_{rew} (Nominal)	NLO_{unrew}	NLO_{rew}	KK2f_{unrew}	KK2f_{rew}
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
Significance (Region B)	3.5	4.2	3.5	4.1	5.25 ± 0.06	4.6

All show highly significant excesses.

We also tried 6 different jet-clustering algorithms.

Algorithm	LUCLUS (Nominal)	DURHAM	JADE	DICLUS	DMLR	LMNR
Significance (Region A)	$5.51^{+0.06}_{-0.05}$	4.5	$5.02^{+0.04}_{-0.03}$	> 5.59(stat) 5.73(ex)	5.15 ± 0.05	5.3 ± 0.1
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 $3 \to 2$.

Whatever is happening in the data is not just exploiting a feature of a single algorithm.

More jet-clustering & jet-rescaling features

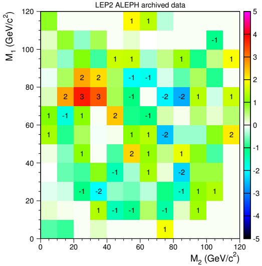
Can also try fixed-velocity jet rescaling instead of fixed-mass jet rescaling:

Parameter	Fixed-Mass Rescaling (Nominal)	Fixed-Velocity Rescaling
Significance (Region A)	$5.51^{+0.06}_{-0.05}$	4.8
Significance (Region B)	4.2	3.3

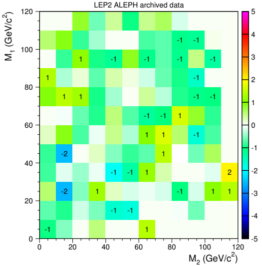
Seems to reduce significance in both regions.
Perhaps gives some clue about jet structure of these events.

What happens if you ask that Luclus and Durham agree on Σ?

$|\Sigma_{LUC} - \Sigma_{DUR}| < 5$ GeV



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



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

Systematics

We break sources of systematic uncertainty into 2 groups:

- Those which can be estimated by comparing reweighted LO Sherpa, NLO Sherpa, and KK2f QCD MC samples (omitted higher-order terms, merging scale, value of α_s, showering, QCD MC statistics).

  Take $1/2$ of the difference between LO and other reweighted samples as uncertainty from these sources.
- Those which cannot (Luminosity, cross-sections, beam backgrounds, modelling of photons, hadronization, all 4-fermion MC uncertainties).

  This set requires dedicated systematic studies.

Must consider corrections and uncertainties at both LEP1 and LEP2.

Systematics

Hadronization likely dominant systematic uncertainty:

- Compare LO Sherpa sample to another LO Sherpa w/AHADIC++.
- Difference before reweighting: $2-3\%$ (A), $\sim 1\%$ (B). After: $\sim 1\%$.
- Also some studies with LEP-era generators.
- Take 3% for Region A and 2% for Region B as upper bounds, applied to reweighted QCD MC samples.
- Quote significance as function of hadronization error.

Systematics

Corrections and uncertainties to MC samples:

Source	SHERPA		KK2f		Four-fermion
	LEP1	LEP2	LEP1	LEP2	LEP2 Only
Luminosity	±0.12%	±0.5%	±0.12%	±0.5%	±0.5%
Cross-section	±0.1%	±0.2%	±0.1%	±0.2%	±0.4%
MC Statistics	Included in MC comparison				±1.3% (A) ±0.38% (B)
Reweighting	±0.3% (A) ±0.25% (B)	N/A	±0.3% (A) ±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)	±0.035%	±0.35%	±0.035%	±0.35%	N/A
$\lvert p_{zmis} \rvert$ cut variation	N/A	±0.07%	N/A	±0.07%	N/A
Hadronization	±0%,±1%, ±2%, ±3% on LEP2 after reweighting				±0.2%

Plus systematics from comparing QCD samples: 1.2% (Region A), 0.3% (Region B).

Include correlations.

Total systematics: 1.4% – 3.1% (Region A), 0.7% – 1.5% (Region B).

Results

Final results:

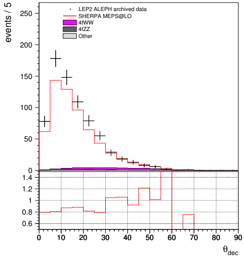
Hadronization uncertainty	0%	±1%	±2%	±3%
Toy MC generated	8.0×10^8	4.0×10^8	1.2×10^8	4×10^7
p-value/(10^{-8}) (Region A)	1.6 ± 0.5	$2.5^{+1.0}_{-0.8}$	16^{+4}_{-3}	113 ± 17
Significance (Region A)	$5.53^{+0.06}_{-0.05}$	$5.45^{+0.07}_{-0.06}$	$5.11^{+0.05}_{-0.04}$	4.73 ± 0.03
p-value/(10^{-6}) (Region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

Basic features of the excess

What do these events look like?

- 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:



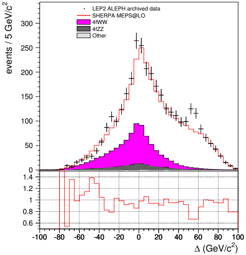
- Both features consistent w/less-energetic jet in the 80-GeV system being softer than would be expected from a genuine 80-GeV particle.
- Does not look very resonance-like. Looks a lot like QCD background.

Sanity Checks

Some basic checks:

- Excess events showing activity in any particular part of detector?
  → Polar, azimuthal angles of thrust axis in accord with background.
- Missing energy in events?
  → 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 $\Delta = M_1 - M_2$:
  →Hard to say anything conclusive.



- Excess dependence on $\sqrt{s}$?
  → Compatible w/being proportional to QCD bkg, but error bars large.

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.
- Even if it is QCD mismodelling, we have reason to think this excess has affected LEP analyses. Could recur at future lepton colliders.

Conclusions

- Likely need QCD experts to see if QCD is (could be?) the culprit.
- 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. Any effects that would be missing in all 3 SM QCD generations? INPUT WELCOME!

Backup

Preselection

LEP1 cuts:

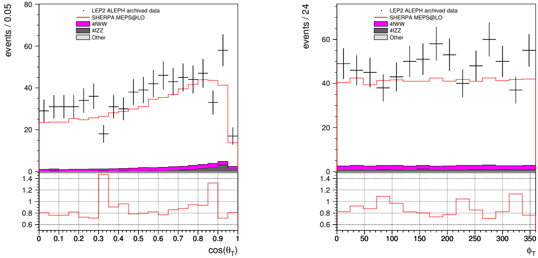
- Require 7 good charged tracks in event.
- Force events into 4 jets; require each jet have at least one good charged track.
- Sum of jet transverse momenta $p_{tsum} > 25\%\sqrt{s}$.
- Rescale energy, momenta of four jets, keeping directions and masses fixed, so that sum of four-momenta are $(\sqrt{s},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

LEP2 cuts:

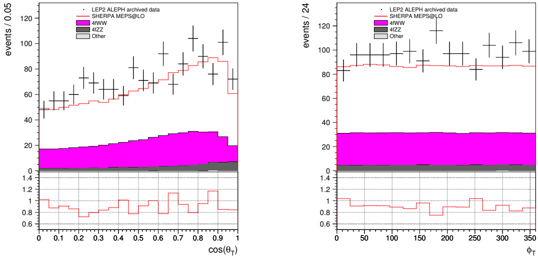
- above cuts
- No jet with more than 80% of its EM energy in a 1° cone around an energy flow object.
- $|p_{zmis}| < 1.5(m_{vis} - 90)$.

Excess Thrust Polar & Azimuthal Angles

Region A:

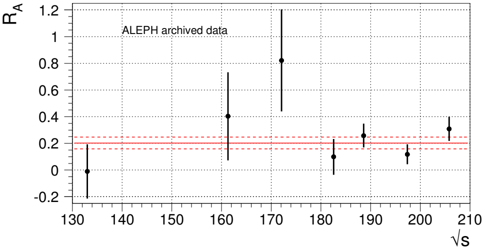


Region B:

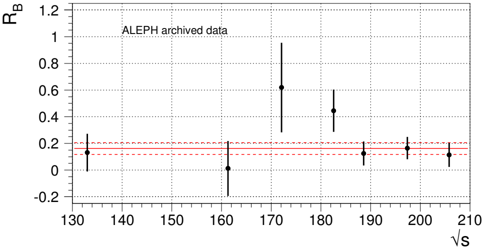


Ratio of fitted number of events to QCD Bkg

Region A:



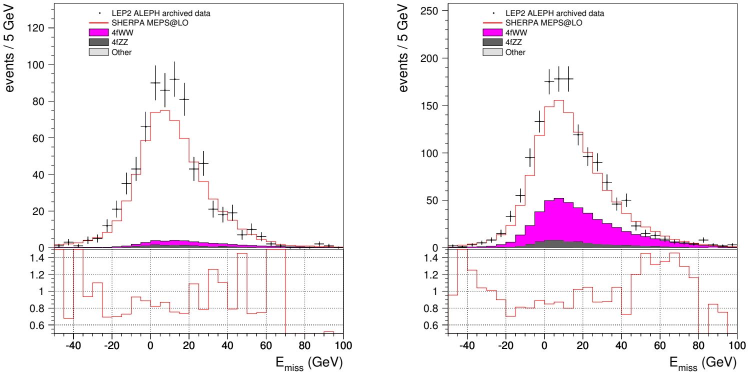
Region B:



Amount of QCD used is that contained in ellipse containing 90% of fitted excess.

Missing Energy

Region A: Region B:



Systematics

- Hadronization likely dominates systematic uncertainty.
  QCD samples all used Pythia hadronization (w/different parameters).
  Want hadronization uncertainty for LO Sherpa. Two strategies:
  - Compare KK2f (Pythia string hadronization) to Herwig (cluster hadronization). Take difference as uncertainty.
    Issue: KK2f, Herwig also have different parton showers, which control the structure of the event. May greatly overestimate hadronization uncertainty for Sherpa samples.
    Diff after reweighting: $\sim 3\%$ (Region A), $\sim 2\%$ (Region B).
  - Compare LO Sherpa sample to another LO Sherpa w/AHADIC++.
    Not clear if should take difference before ($2-3\%$ (A), $\sim 1\%$ (B)) or after ($\sim 1\%$) reweighting.

  Take 3% for Region A and 2% for Region B as upper bounds, applied to reweighted QCD MC samples.
  Quote significance as function of hadronization error, in case better estimates become available.

Systematics

Details of these sources of corrections/uncertainties:

- Beam backgrounds:
  Responsible for extra energy deposits in detector not modelled in MC.
  Can cause events to fail rescaling requirement or 80% EM energy cut.
  Will cause efficiencies in MC to be too large.
  Estimated by adding random-trigger energy flow objects to MC.
- Modelling of photons:
  ISR, FSR photons not well-modelled in Sherpa. Much better in KK2f.
  Basic idea: Correct efficiencies of anti-photon and rescaling cuts to agree w/KK2f, other LEP-era generators with good photon modelling.

Assumptions and Interpretation

Excesses are robust, but similarity to SM QCD (especially θ_{dec}) is confusing. Assumptions which could skew interpretation:

- Regions A and B separate excesses, or one continuous excess?
- Other events hiding in data?
- Should we be considering other jet pairings?

The last two points imply that we could be missing relevant events. Any of the above issues could cause skew distributions, make interpretation difficult.

- Is there something invisible in the event? (Rescaling procedure assumes any invisible particles collinear with jets.)
- Should events be forced into 4 jets?

So, it is possible that we have not found the optimal way to look at the excess.

Shouldn't this have been seen?

Lots of LEP 4-jet analyses. How could this have stayed hidden until now?

- Here, we looked at all LEP2 data at preselection at once. Not typical during LEP era. If combination of different years of running made, usually at end stage of analysis, after tight cuts.
- Excess looks much like QCD. Most analyses placed hard cuts (thrust, sphericity, y_{34}) to eliminate almost all (hard to model) QCD bkg. Typically kills Region A very effectively, does damage to Region B.
- Searches for equal-mass dijets would select small fraction of excess.
- SUSY searches usually involved $\not{E}$; Higgs searches required b jets.
- Possibly relevant: flavor-independent Higgs searches for hZ, hA.
  - No b-tagging, low masses considered, analyses started near end of LEP running, so data all analysed roughly simultaneously.
  - Limits on cross-sections of order that which we see for our excesses.
  - Several intriguing small excesses in regions where we would expect, but difficult to translate their results to our situation.