Daily Research Logs

Yi Chen
California institute of technology
chen.yi.first@gmail.com

November 9, 2010

Contents

Chapter 1

September 2010

1.1 6263 Log (Septemper 29, 2010)

1.1.1 Goals

1. Catch up with Maria
2. Get some rest

1.1.2 Summary List

1. Back from vacation. Trying to catch up.
2. Copied exotica hotline code to CMSDetNoiseLine package. Nothing modified.
3. Setup new logbook in latex.
4. Start reading exotica hotline code.

1.1.3 Latex logbook

The main goal is to have a logbook that is easily searchable and scalable. Original handwritten logbooks have the advantage of sketching ideas, but is not suitable for searching or write texts with a lot of revisions. In the near future I might start using scanners to scan sketches as pictures and include them in the latex logbook. To make it scalable, each day is to have its own tex segment which can be included in a tex file that does the structuring. The title and (sub-) sections are newly defined commands that can be reassigned in the structuring tex file.

The structure of each day is as follows:

1. Daily goals.
2. Summary of things done.
3. For each non-trivial item, write something about it.
4. Meeting notes.
5. Anything else worth noting.
6. Reflection. What was done and what could be done better.
7. Progress on studying, summary on paper reading.
8. Minimum goal for the next workday.

Not all of them need to be filled in.

1.1.4 Reading exotica hotline code

The code is in package `UserCode/ExoticaHotLine/src/HotlineSkimCode/RecoSkim`. In the final configuration file, each filter is a module, and there are various paths assembling them together. In the end the events are kept using the `SelectEvents` field in `PoolOutputModule`.

Even though it need not be the case, it appears that all the filters are implemented together as a `EDFilter` named `RecoSkim`. Different filters are the same module with different parameters. For Hcal noise we definitely can implement multiple filter modules.

There are two modules in the hotline code directory. One is the aforementioned `RecoSkim` filter, which looks like basic cut-based selections with cut values specified in the configuration file. The other one is an analyzer `HotlineSummary`, and it appears to be printing various summary values from edm collections. The printout is long....this module is probably only for debugging purposes.

1.1.5 Reflection

Need to think through the purpose of hcal noise hotline. I want to be able to estimate noise rate (of various type) for any given run from the hotline. Also it will be good to include some kind of correlation with beam luminosity and/or triggers.

On latex logbook, need to think about possible types of extensions and how to implement them. In principle the current framework should be enough.

1.1.6 Goals for next work day

1. Skim through Hcal noise meetings
2. Skim through vecbos meetings
3. Catch up with progress on the candle note and make a list of items to do
4. Move the daily latex logbook to svn
5. Catch Maria

1.2 6264 Log (Septemper 30, 2010)

1.2.1 Goals

1. Go through vecbos meetings in september
2. Go through Hcal meetings in september
3. Move the logbook to subversion
4. Make a list of things to do for the candle note

1.2.2 Summary List

1. Skimmed through vecbos meetings in espace and V+Jet meetings
2. Skimmed through hcal WG meetings and DPG meetings
3. Moved logbook to subversion
4. Update VecbosApp to newest version, test run on the current muon list (up to run 144114). No obvious problem spotted.

1.2.3 Go through vecbos meetings in September

1.2.3.1 Espace meetings

1. September 8, "Thresholds" by Maria Spiropulu. Default value: CaloJet 30, UncorrectedCalo 20, Track 15, PF 30
2. September 8, "Lucas Fit" by Lukas Vanelderen.
   (a) Fit MT for W, t+X, other
   (b) Fix shape to MC for W, t+X, and float the other
   (c) "W and top+X separated well and unbiassed from other"
   (d) Fit W+LF vs. W+HF with t+X, and use the HF fraction from MC to recover W yield
3. September 15, "btag" by Lukas Vanelderen. Control sample for HF from data. Need to read about b-tagging algorithms.
4. September 15, "Vecbos Meeting" by Matthias Ulrich Mozer. Revisit uncertainties on AlphaL and AlphaR.
   (a) Traditional fit: fix alpha to best known value, and redo fit with different alpha to get uncertainty
   (b) Nuisance parameter: constrain alpha by a gaussian centered at the best known value.
   (c) 7-fit plot.
5. September 22, "Vecbos Meeting". "W and Z + jets" by E. di Marco in General EWK meeting.
6. September 22, "Vecbos Meeting". Lukas updated results on WJet fit.

7. September 22, "Vecbos Meeting". Will Reece updated on trigger efficiencies.
8. September 29, "W fit strategy, flavor part" by Lukas Vanelderen. Estimate PDF from b-tag variables from control samples for t+X and W+LF. Seems to have problem in the 2Jet bin.

1.2.3.2 V+Jet meetings

1. September 7. Lukas on fit strategy in W+Jets (sane as the one in espace). Z candle analysis status report (with toys).
2. September 21, "Introduction on Zbb issues and current plans" by Alexandre Nikitenko. Z+b is similar to H+b
3. September 21, "Task list overview" by Vitaliano Ciulli and Ilaria Segoni.
4. September 28, "Report on Zbb analysis" by Anne-Marie Magnan.
5. September 28, "Report on Zb(b) analysis" by Natalie Heracleous.
6. September 28, "Update on Z(ee)+jets and W(enu) +jet studies" by Sarah Malik. (....)
7. September 28, "Status on PFlow Z+Jets Analysis" by Anil Pratap Singh.

1.2.4 Go through Hcal meetings on noise

1.2.4.1 Hcal Noise WG

1. September 9, "HF Flags in 3_8 (slides for Maria)". Some notes on HF reconstruction and flagging.
2. September 9, "Isolated Noise Filtering" by John Paul Chou. Summary of the isolation-based noise filter. Performance on ttbar and Ztautau. Suggests going on to JetID. Reviewed reconstruction chain.
3. September 9, "HPD Pulseshape Discriminators" by Jason St. John. Included HE. MC shape needs work.
4. September 9, "Hits in a Jet" by Hongxuan Liu. Good hits and PMT window hit could overlap.
5. September 9, "HBHE Timing and Noise Studies" by Phil Dudero. Derive time envelope from collisions. Plots for time envelope with/without low energy hits as well as square filter (energy independent).
6. September 9, "Impact on MET due to ECAL masked/dead cells" by Hongxuan Liu. Jet response 2% quantile map. Holes correspond to dead cells. Jet energy recovery algorithm.
7. September 23, "Isolated Noise Filter: Performance" by John Paul Chou. Update his filter to be used as a hit cleaner and not a event filter.
8. September 23, "HBHE Timing and Noise Studies" by Phil Dudero. Some error/problem two weeks ago. Updated square filter results.

1.2.4.2 Hcal DPG

Note: Talks that have nothing to do with noise are omitted here.

1. September 13, "HCAL QIE Offsets" by a list of people. The new setting is consistent with old setting (with a overall constant shift) for HB and HE.

2. September 13, "HCAL Noise" by Maria. A summary to be used in Bodrum.

3. September 27, "TP Energy Scale" by Patrick Tseng. He recalibrated and checked TP energy.

4. September 27, "QIE hardware offset and time reco" by Pawel de Barbaro. Validated new QIE settings. Overall good. Time spread is smaller. Some channels (not many) are off.

5. September 27, "Precise time correction" by Jeremiah Mans. An independent analyses to derive time corrections. Compared with those from Pawel et al. and looked at channels that disagree.

6. September 27, "An Isolated HB/HE Noise Filter" by John Paul Chou. Same talk as in Hcal noise WG.

7. September 30, "Phi calibration of HB, HE - initial results" by Igor Vodopiyanov. Intercalibration using non-ZS data. Not clear from the presentation what "E1" is.

The QIE hardware timing offset is adjusted since runs 146XXX!

1.2.5 Updating VecBosApp to newest version and test run on data (2.66/pb)

1. Everything went fine on cvs update and merging versions.

2. Test run on ZJetsMADGRAPH sample, all jobs finished successfully, though castor was busy for one job. Rerun does the job.

3. Copying from castor back to local disk gives "`. : Invalid argument`". Maybe castor was extremely busy.

4. ps. the error means that disk quota was exceeded.

5. No problem spotted in ZJetsMADGRAPH sample from the QM plots.

6. Test run on current dataset (up to run 144114, 2.66/pb reported). While submitting jobs, encountered one instance of "LSF js on lxbsp0901.cern.ch: LFS js: no AFS token" error. It doesn't seem to be related to the updating of VecBosApp. It doesn't seem to be affecting anything either. Jobs are successful.

7. The castor-friendly safety sleep time (10s) is getting annoying now that there is more statistics. Let's try to reduce it to 3 seconds.

8. Data looked OK at first glance.

9. The mass of any two global muons looks nice, see figure 1.1.

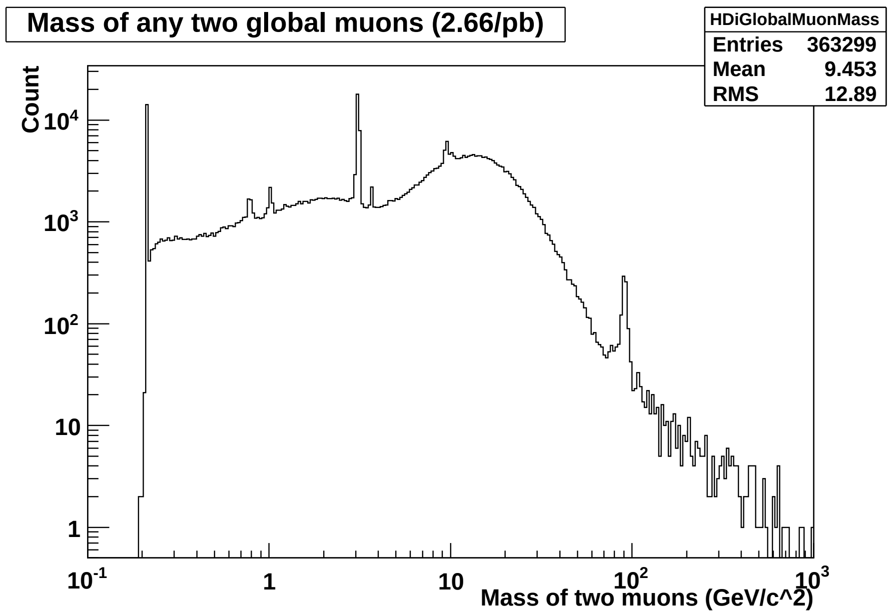


Figure 1.1: Mass of any two global muons from all processed data so far (up to run 144114). Peaks from right to left are speculated to be Z (~90), Upsilon family (~10), J/Psi(1s, 2s) (~3), phi (~1), rho/omega (0.7~0.8), and muonium (~0.2). (ps. The last one was just kidding. It's probably from doubly reconstructed ghost muons. Though further investigation is needed.)

1.2.6 Meeting notes

1.2.6.1 Caltech group meeting

1. There is some narrow peak discovered (!?)
2. Maria: the comment system needs to be rethought. Actual commitment is needed. Comments on physics, not styles.
3. Artur gave a presentation on the recent drama on Hcal. Accidental unmasking of hcal bad channels, severity level in HLT
4. Piotr reports on the peak of opposite-sign dimuons around 244 GeV.
5. Update from Jan. Z$\rightarrow \mu\mu$ vs. $\mu\mu + \gamma$, Energy scale of photon.
6. Action items for next Tuesday to be emailed out by Dorian

1.2.7 Reflection

To fully understand hcal noise, we need to have real categories (instead of the simplistic ion/hpd/rbx picture), and monitor the change over time to obtain a control sample estimate of the amount of noise of each type for all RBXs.

1.2.8 Goals for next work day

1. Sort out goals for Hcal noise line
2. Make sure how prescale works with multiple triggers
3. Make a list of to-do items for candle analysis
4. Review/summarize progress so far on pulse shape variables
5. Check strategy on Z shape fit, find out ways to contrain RooFormulaVar

Chapter 2

October 2010

2.1 6265 Log (October 1, 2010)

2.1.1 Goals

1. Check the opposite-sign dimuon spectrum and note anything interesting, especially the "muonium" peak.
2. How does the trigger prescale work?
3. Sort out the purpose of Hcal noiseline and how/what to implement
4. Summarize work on noise characterization so far
5. Make a list of to-do items on candle note
6. How to do the fit on Z shape?

2.1.2 Summary List

1. Noiseline meeting. Not too productive without a clear-cut goal.
2. Checked quickly opposite-sign di-muon mass spectrum.

2.1.3 Opposite-sign dimuon (global) spectrum

The result of almost all data is shown in figure 2.1. I still need to think about the overall underlying shape of the curve. For example, what is the wide bump around 2 GeV and 14 GeV. The resonances are nice however. The tail also bends after the Z peak.

2.1.4 Work so far on noise characterization

2.1.4.1 Things tried

1. First three TS should be compatible to zero.
2. The maximum N continuous time slices. Generalization of E2.
3. Number of time slices required to achieve P%. Useful to pick out sharp noise and broad noise.
4. RMS vs. mean of the 10 TS.
5. Linear fit of the pulse shape. Potentially useful to pick out flat pulses.
6. Two-step fit.

2.1.4.2 Late pulses - produced at interaction

Skimmed through PDG tables and estimate what might be late, and how much energy they will deposit. Assume distance to go is 1.4m. See table 2.1 for numbers.

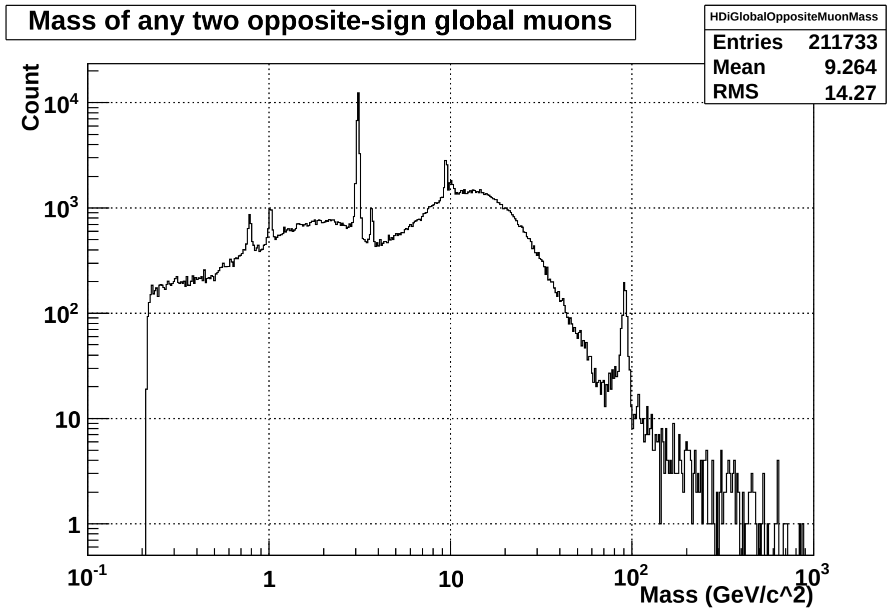


Figure 2.1: Opposite-sign dimuon (global) mass spectrum for almost full statistics (full = 2.66/pb). The last job is still running, and I don't think it matters for this. The muonium peak goes away, which is nice.

Particle	Mass (GeV)	τ (ns)	min(β)	max delay (ns)	Energy (GeV)
$\pi+$	0.13957	26.033	0.176	26.45	0.142
$K+$	0.493677	12.38	0.353	13.23	0.528
K_{0L}	0.497614	51.16	0.0908	51.37	0.500
p	0.938	∞	0.000	∞	0.938
n	0.940	∞	0.000	∞	0.940
$\Omega-$	1.67245	0.0821	0.9998	4.667	95.08
Ξ_0	1.31486	0.29	0.998	4.676	21.20
Ξ_-	1.32171	0.1639	0.9994	4.670	37.656

Table 2.1: Table for maximum delay (constrained by particle lifetime) of particles produced at interaction point and travelled 1.4 m. The "maximum delay" is calculated using the mean lifetime. Other particles decay too quickly, the delay would be tiny.

Out of these, only $\pi+$, $K+$, K_{0L}, proton and neutron have the chance to reach 1.4m at 12.5 ns or more. The equivalent total energy would be 0.15, 0.53, 0.54, 1.01, 1.01 GeV, respectively. Longer time delay means smaller energy - these are the ballpark numbers we expect. If the allowed length is 3m, then the total energy will be 0.23, 0.82, 0.83, 1.56, 1.57 GeV. So from the slow (relatively) stable particles produced at interaction we expect O(1 GeV) energy deposit at most, if they were half TS or more late.

2.1.4.3 Late pulses - decay-in-flight

The idea here is that there might be some heavy stuff produced at small β, and decay after a while to light, energetic particles. Lab frame lifetime is $\gamma\tau$, and let's take an assumption that the decay product is extremely energetic, ie, $\beta \sim 1$. Let's do the calculation when the parent particle stays on average 12.5 ns in lab frame, and the minimum required energy of it.

$$\begin{aligned}
T &= 12.5ns \\
\gamma &= \frac{T}{\tau} \\
\beta &= \sqrt{1-\gamma^{-2}} = \sqrt{1-\frac{\tau^2}{T^2}} \\
\text{Distance travelled} &= \beta\gamma c\tau = \beta c T = \beta \times 3.74m
\end{aligned}$$

Unless τ is close enough to T, the particle won't stay in the detector for 12.5 ns. Which means that under the assumption that the decay product is relativistic, decay-in-flight particles will at most deposit same order of energy as the ones produced at production. If the decay product has classical velocity, it won't deposit much energy anyways.

2.1.4.4 Other items - moved to next work day to think about

After the afternoon noiseline meeting, it seems that I have a lot on my plate already. Let me make a definite plan on how to tackle each of them first....

1. Late pulses - late hadronic/EM shower-developement....what's the typical shower develop time? What's the chance that a shower fragment makes an ion-feedback noise?
2. Out-of-time pulses - radiation from other sub-detectors...how is the strength of radiation related to dosage history (for different material)? What's the expected dosage for certain instantaneous luminosity? Can we estimate the radiation from the environment? What's the radiation content? Mostly photons? How about hadrons?
3. Out-of-time pulses - beam background
4. Out-of-time pulses - cosmic muons
5. Things that might worth trying

2.1.5 Meeting notes

2.1.5.1 6265 Morning Noiseline Meeting

1. Maria had a car accident. I hope it's not too serious.
2. The main thing to clarify is the purpose of the noiseline. Everything goes from there.
3. Adi mentioned a few possible use cases:
   (a) Find noise that won't be caught otherwise
   (b) Correlation between detectors (more like DQM plots)
   (c) Radiation damage?
4. What I want is some kind of noise trend monitoring, finer than the current ion/HPD/RBX categories
5. To begin with, Artur will send me examples of DQM codes so that I can play with it.

2.1.5.2 6265 Afternoon Noiseline Meeting

1. Until we have the first result, the noiseline should be the same as exotica hotline - keep events so that we can look at it. We want to integrate into DQM and P5 event display.
2. Have one firework display that constantly show noises + physics. (And one for normal events, one for exotica hotline.)
3. Eventually it will become a skim.
4. To start with, we should see what is meaningful. Find noise overlapping with physics signature (muons, etc.).
5. Artur: integrate into Hcal DQM?
6. JetMET? Homework for Artur?
7. What is Muon DPG doing? Homework for Piotr.
8. Now we should put whatever we have in (HCAL, ECAL).
9. Maria: We need to think about the workflow to rereco without noise cleaning.
10. How to catch new forms of noise?
11. Piotr will show dimuon results! He will check noise in muon system and report.
12. Exotica hotline spots possible types of new noise, and we follow up on them.
13. Integrate Shuichi's monitoring to DQM?
14. Run first on exotica hotline files and see how many we keep.

15. As a first step, try to run the exotica hotline workflow.
16. Aim to have a prototype in the next two weeks.
17. Artur wants to have Hcal noise DQM done in the next two weeks. (Attack for bonus point!)

2.1.6 To-do's for next week

2.1.6.1 Z+Jet Candle

1. Find Matthias and update on the status of fit...strategy, etc.
2. Find out how to constrain `RooFormulaVar` to be greater than zero.
3. Check out a copy of the Z candle note and make a list of items to produce (and automize).
4. Check with Maurizio and see if I miss anything.

2.1.6.2 W+Jet Fit Without b-tagging

1. Check with Chris to discuss on the strategy. Maybe there is a working fit from electrons.

2.1.6.3 Hcal Noise Characterization

1. Continue doing subtraction from noise sample.
2. Make a signal root file and see where the signal lands.
3. Get the most recent timing correction from Jeremy et. al.
4. Condense into a few categories of noise shape and make `EDFilter` of them.
5. Then it's ripe to integrate into DQM

2.1.6.4 Hcal Noiseline Project

1. Check out the exotica hotline twiki (`https://twiki.cern.ch/twiki/bin/view/CMS/ExoticaHotline`) and follow the steps to get a working version.
2. Read and make a map of various paths in the exotica hotline to see what physics signatures are included.
3. Check what kinds of noise cleaning are done in the hotline.
4. If noise filter is not there, include a simple one (ICHEP JP filter) to start.

2.1.6.5 Hcal DQM Integration

1. Get code structure from Artur, make a working private copy.
2. Learn how the structure is in the DQM.
3. Add a simple practice plot to the structure.
4. Somehow find out where Shuichi's code is, and extract the requirements on different categories.
5. Put Shuichi's monitoring tool into DQM plots.
6. Integrate ICHEP JP filter variables into this private DQM.
7. Integrate JP isolation filter-related variables into this private DQM.

2.1.7 Reflection

.....

2.1.8 Goals for next work day

See previous section "To-do's for next week"

2.2 6267 Log (October 3, 2010)

2.2.1 Goals

1. Understand the filtering logic behind exotica hotline
2. Make the hotline code work
3. Make a list of people to bug

2.2.2 Summary List

1. Read exotica hotline code and list the physics signatures it is catching.
2. Checked the event cleaning in exotica hotline.

2.2.3 Reading exotica hotline code

2.2.3.1 "Physics signature" in exotica hotline

Paths available in the `hotlineSkim_AllFilters_cfg.py`. All paths are `basicFilters(NoPV) + ak5CaloJetsL2L3 + HBHENoiseFilter` and then the target filter.

1. High Ht (`NoPV`). `HtJetThreshold` = 30, `HtMin` = 700
2. High Met. `PFMetMin` = 300
3. High Pt Electron. `gsfElectron`, PT Min = 100
4. High Pt Photon. 150 GeV, `photons` collection
5. High Pt Jet. 350 GeV. `ak5CaloJet`.
6. High Pt Muon. 80 GeV. Standard muon collection.
7. Multi Track (`NoPV`). 600 tracks. `generalTracks` collection
8. Multi Muon. Two muons above 45 GeV.
9. Multi Electron (`NoPV`). Two electrons above 45 GeV.
10. Multi Jet (`NoPV`). Five jets above 50 GeV.
11. Multi Photon (`NoPV`). Three photons above 30 GeV.
12. Top Electron (`NoPV`). b-jet PT min 50 GeV, b-jet discriminator 1.7, one b-jet. `HtJetThreshold` 30, `HtMin` 150. At least one electrons above 80 GeV.
13. Top Muon (`NoPV`). Same as the previous path.
14. Top Hadron (`NoPV`). b-jet requirement same as before. `HtMin` become 200.

2.2.3.2 Event cleaning in exotica hotline

1. Basic filters (with/without PV).
   (a) Primary vertex. n.d.o.f. at least 4, max z deviation 25, max d0 deviation 2. `offlinePrimaryVertices` collection
   (b) Physics declared.
   (c) Scraping. 10 tracks above 0.2.
2. `ak5CaloJetL2L3`. `JetMETCorrections.Configuration.DefaultJEC_cff`.
3. `HBHENoiseFilter`. Same as the one in `CommonTools.RecoAlgos.HBHENoiseFilter_cfi` except that the `EMF` requirement is added. The version in `CMSSW_3_8_2` is confirmed to be the same as the ICHEP one.

2.2.4 Reflection

Orz. Not much done today.... Sunday isn't really productive.

The ecal spike cleaning seems to be turned on by default before the hotline. Maybe we want to somehow read from the uncleaned version of it, and coincide spike signature with physics signature. But for the time being I'll reverse the Hcal filter only.

2.2.5 Goals for next work day

1. Make the hotline work!
2. Ecal spikes vs. hotline?
3. Read EXO-10-002-001 and compile a list of comments.
4. Bug Artur for DQM code
5. Bug Maurizio/Matthias for fitting strategy

2.3 6271 Log (October 4, 2010)

2.3.1 Goals

1. Make the hotline work!
2. Ecal spikes vs. hotline?
3. Read EXO-10-002-001 and compile a list of comments.
4. Bug Artur for DQM code
5. Bug Maurizio/Matthias for fitting strategy

2.3.2 Summary List

1. First attempt at running exotica hotline code. Local running is fine. For CAF access or to run cronjobs, I'll have to wait a bit.
2. Quickly read through EXO-10-002 and made a small list of comments.
3. Got the Hcal DQM instructions from Artur.
4. Start getting a signal shape sample.

2.3.3 Making exotica hotline run as is

Following instructions on the twiki `https://twiki.cern.ch/twiki/bin/viewauth/CMS/ExoticaHotline`.

1. First step is to run it locally and see if it works.

   ```
   scramv1 project CMSSW CMSSW_3_8_2
   cd CMSSW_3_8_2/src
   cvs co -r V00-02-02 -d HotlineSkimCode/RecoSkim \\
      UserCode/ExoticaHotLine/src/HotlineSkimCode/RecoSkim
   scramv1 build
   cd HotlineSkimCode/RecoSkim/test
   eval `scramv1 runtime -sh`
   cmsRun hotlineSkim_AllFilters_cfg.py
   ```

2. Successfully built. `hotlineTopHadron_cfi.py` is missing! Copy from the checked out version from last week. This example file read SUSY LM1 relval samples. Now it seems to be working after copying the configuration file fragment.
3. Disk quota exceeded. :(Change to output to tmp directory.
4. Done. The output is a edm format file with some events in it. Meow. `cmsShow` is cool.
5. Next thing to try is to setup cron job.

```
cd /afs/cern.ch/user/t/tulika/scratch2/ExoticaHotLine/cronJob
python exoHotLineCron.py Run2010B ExpressPhysics \\
   /afs/cern.ch/user/t/tulika/scratch2/ExoticaHotLine/CMSSW_3_8_2/src
```

6. Permission denied!!! Oops.

7. Maria is asking Luca Malgeri for permissions on CAF.

2.3.4 Reading of CMS paper draft EXO-10-002-001

Information on the version read:

```
2010/09/21
Head Id: 17377
Archive Id: 17381
Archive Date: 2010/09/20
Archive Tag: trunk
```

Overall it is a good paper, though it will be good to elaborate a bit on details. Whether to add more detail to paper is another thing, but I would like to know how the analysis is done. Main questions/comments below:

1. Line 26. Why choose positive sign?

2. Line 54. Cone size of 0.7 is huge, especially since the eta range is only 1.4 units or so for each region. Is there any particular reason smaller cone sizes are not used?

3. Line 80-83. Elaborate a bit more on this? The reference didn't say much either.

4. Line 92. How is the 20% determined?

5. Figure 1. The choice of division points for mass is perplexing....

6. Figure 1. How much systematics is expected on pythia prediction? Is it okay to plot the band too?

7. Line 100. What is ρ_j exactly?

8. Line 108-109. Is the overall shift in units of ratio? Could the choice of constant shift be justified?

9. Line 129-134. Where does the 10x amplification come from?

Cosmetics left for other people to pick. I'm not good at English.

2.3.5 Hcal noise classification

1. The most important thing is to get a sample of signal pulse shape, and also noises from a couple other runs. Let me do that now.

2. Copied code from `Work` to `Workspace` on `lxplus`. The code is in `lxplus:scratch0/Workspace/HcalNoise/ExamplePulseShapes/6271`. Batch submitted for signal pulses. Input: DigiTree, Output: QM histograms and PulseShape tree.

3. The run time for individual job is about 20 minutes, as estimated by the test run.
4. Encountered "No space left on device" error. Since the files are too large (more than 2 GB), so I copied them to local tmp folder before running....Orz....
5. For now, let's leave it there and see how bad it is. All I need is some samples of pulses.
6. No. Change it back to rfio access. At least I will have some useful pulses this way.
7. Cloned a copy of the analyzer code to `pccit28:Workspace/HcalNoise/PulseShapeVariableBrainStorm/6271`.
8. There are repeating entries due to castor error. After 2GB is reached, trees won't be read and the content of the branches appear repeating.
9. For now make a comparison of ADC to previous entry, and if it is the same, drop the event.

2.3.6 Checking status of the ZJets candle note

1. First I need to find the svn repository....can't find it.
2. Maybe it's easier to bug people for it.

2.3.7 Reflection

Start from the afternoon, I was getting dizzy and lowered the work efficiency. How can I improve this?

2.3.8 Goals for next work day

1. Install DQM from Artur
2. Fit!
3. Finish up Hcal signal sample shape

2.4 6272 Log (October 5, 2010)

2.4.1 Goals

1. Install DQM from Artur
2. Fit!
3. Finish up Hcal signal sample shape

2.4.2 Summary List

1. Hcal DQM basic code checkout. The GUI is not installed yet.
2. Created signal pulse shape sample.
3. Compared signal and noise pulse shapes with the brainstorm ideas.
4. Met with Artur to briefly talk about noiseline ideas
5. Encorporated JP's isolation filter into the noiseline package
6. The float-all version of the candle fit works.

2.4.3 Signal pulse shape for classification studies

Comparing ADC collection might take a lot of time. Changed to event number/run number. The result looks fine from a test run.

2.4.4 Hcal DQM installation

Following the instructions in the twiki page https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalDQM, I checked out the three packages in CMSSW version 3_8_2.

```
cvs co -r V14-00-16      DQM/HcalMonitorClient
cvs co -r V14-00-10      DQM/HcalMonitorModule
cvs co -r V14-00-27      DQM/HcalMonitorTasks
scramv1 build
```

Then proceed into running the test python configuration file. The input file is missing but after relacing the input list with a root file of `RAW` format, it runs fine. The root file used is

```
/store/data/Run2010B/MinimumBias/RAW/v1/000/146/804/
   E89586EA-A0CA-DF11-BBD1-001617DBD472.root.
```

After the job is done, there is a root file produced in the temp directory. Inside there are a lot of root files and pieces of information. Does this mean that the installation was successful?

Now onto the GUI. The twiki is at https://twiki.cern.ch/twiki/bin/view/CMS/DQMTest.

....Looks complicated. Wait until Artur comes to see what is what.

2.4.5 Meeting with Artur on CMS noiseline

We agreed on trying the following

1. Put in the noise filters (standard one, JP isolation filter) to the noiseline
2. Are there spike in $|i\eta|$ 85?
3. Double-spike algorithm
4. Try to run on the runs from last week where the bad channel is unmasked accidentally. And see if it can be picked out.

2.4.6 Adding basic filters to the noiseline

Checked out the JP isolation filter and it compiled fine. However, the boolean value that was used to be there is gone. So I copied the boolean part from his code (earlier version) and merged it into the current version. One can now run the python configuration file and use the result from the reflagging.

2.4.7 Vecbos Z fitting

Updated the fit script. New things:

1. The confusing yield plot is not produced anymore. Instead, a table (plot) is produced which includes the yields.
2. Initial guess of signal yields is estimated using the total number of events and slope 0.2.
3. No top level models that contain signal/background. Each jet bin is a model by itself and included in the final pdf separately. This is to simplify the final yield calculation.

Note. If the initial guess is too far from reality, the errors will go crazy. Sometimes they will be incredibly small (1e-8 level), or more often they will be larger than the fitted yields. Currently I fix only α_L. The result of fit for calojet (30 GeV) and PF jet (30 GeV) are shown in figure 2.2 and 2.3.

Description of the fit implemented.

1. Each jet bin is to have its own function. Cruijff for signal, and exponential for background. The function for each jet bin is multiplied by a `RooSameAs` or `RooAtLeast` to constrain it in the designated bin.
2. The bins are exclusive, except the last one.
3. Signal yields are declared to be inclusive, and the number assigned to the signal function in each jet bin is a `RooFormulaVar` which is just the subtraction of the two relavent inclusive numbers (except the last one).
4. Initial guess of inclusive signal yields is the total number of events, with slope 0.2.
5. The cruijff function is constrained to be exactly the same in all jet bins.
6. The α_L value is fixed, and all others are left floating.

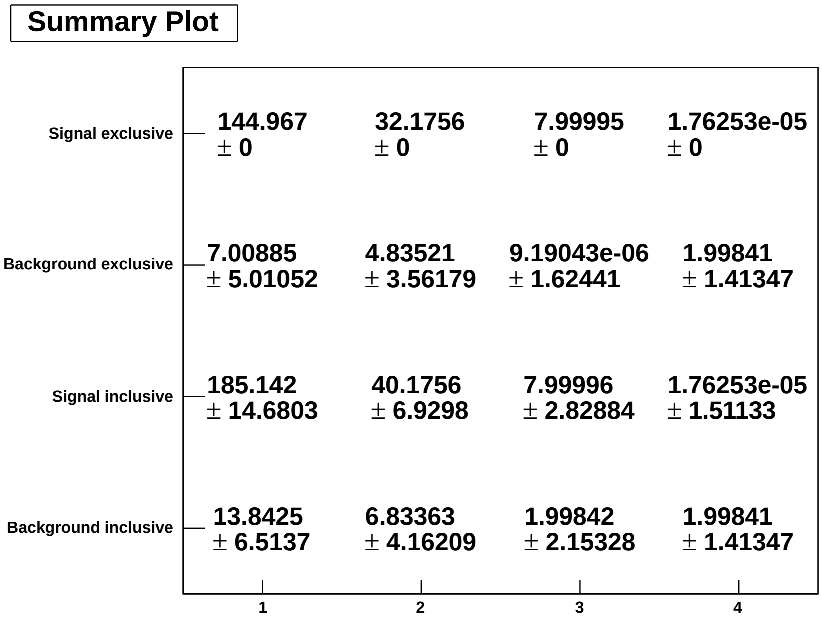


Figure 2.2: Result of float-all-signal-yield fit of Calo jets with 30 GeV. No restriction on relative yield is applied. The error on inclusive yield is from roofit directly. The fit takes care of the error. The error on exclusive signal yield is not defined for `RooFormulaVar`, so it shows zero. The out-of-the-box background yields are the exclusive ones, and the inclusive yield errors are added in quadrature. (α_L is fixed to 0.485)

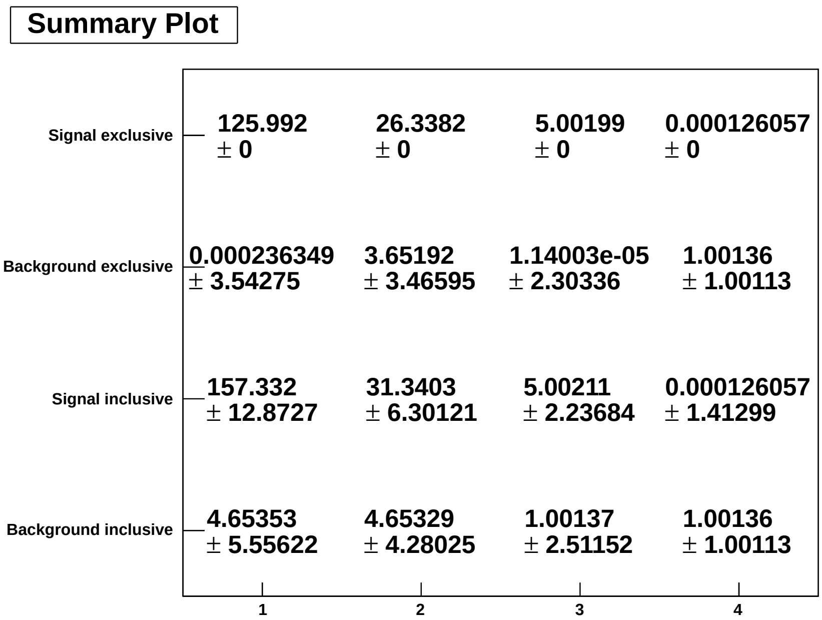


Figure 2.3: Result of float-all-signal-yield fit of PF jets (30 GeV). Refer to the calojet one (figure 2.2) for more information.

2.4.8 Reflection

Not enough focus on vecbos.... The DQM work can be slower.

2.4.9 Goals for next work day

1. Read and understand the double spike algorithm
2. Make sure if anyone is doing the vecbos PDMu dataset
3. Make a set of loose cuts to remove obvious noise, and skim through pulse shapes of the rest
4. Port the HCAL ideal pulse shape (as in `CMSSW`) to `RooFit`
5. Vecbos fitting strategy?
6. To-do chart for ZJets candle note for fast execution once target data ($\sim 50pb^{-1}$) arrives

2.5 6273 Log (October 6, 2010)

2.5.1 Goals

1. Read and understand the double spike algorithm
2. Make sure if anyone is doing the vecbos PDMu dataset
3. Make a set of loose cuts to remove obvious noise, and skim through pulse shapes of the rest
4. Port the HCAL ideal pulse shape (as in `CMSSW`) to `RooFit`
5. Vecbos fitting strategy?
6. To-do chart for ZJets candle note for fast execution once target data ($\sim 50pb^{-1}$) arrives

2.5.2 Summary List

1. Made a checklist and sent to Maurizio for comment

2.5.3 Vecbos to-do chart

1. Data, numbers after each cut. (Ready)
2. Data, tight isolation to get alphaL. (Ready)
3. Data, simultaneous fit to get the right error on signal inclusive yields without enforcing linear constraint. (Ready)
4. Data, fit with the linear constraint. To be compared with the one without enforcing constraint. (Coming soon)
5. Data, get the ratio plot from the result of the fit. (Need the script to take the fit output and make the final plot)
6. Data, repeat the fit 4 times for JEC uncertainty. (Not yet done...coming soon)
7. Data, anti-isolation selection. (Ready)
8. Data, splot to get Z PT, leading-jet PT spectrum, etc. (???)
9. MC, numbers after each cut, as well as numbers in each jet bin. (Ready)
10. MC, demonstration that the isolation does not have a large impact on alphaL. (Ready)
11. MC, anti-muon selection & QCD normal selection. (Ready...though QCD sample is small)
12. MC, justify that an isolation is needed with PF stuff. (Ready)
13. MC, $Z \to \mu\mu$ PT shape vs. $Z \to \nu\nu$ MET shape comparison. (Not yet done)

14. MC, efficiencies in different jet bin. (Not ready)

15. Toy MC, on the functional form itself. Generate sample with cruijff, fit to check the spread. (OK, repeatable within a day)

16. Toy MC, if the parameters are off, how much bias will there be. Show that only alphaL matters. (OK, repeatable within a day)

17. Show that the error we get from the simultaneous fit is reasonable (???)

18. Efficiency of the trigger (???)

Reply from him:

1. There is a tool to calculate splots in `RooStat`

2. Not much to be done in trigger efficiency

2.5.4 Reflection

An extremely slow day....

2.5.5 Goals for next work day

1. Read and understand the double spike algorithm

2. Make sure if anyone is doing the vecbos PDMu dataset

3. Make a set of loose cuts to remove obvious noise, and skim through pulse shapes of the rest

4. Port the HCAL ideal pulse shape (as in `CMSSW`) to `RooFit`

2.6 6274 Log (October 7, 2010)

2.6.1 Goals

1. Read and understand the double spike algorithm
2. Make sure if anyone is doing the vecbos PDMu dataset
3. Make a set of loose cuts to remove obvious noise, and skim through pulse shapes of the rest
4. Port the HCAL ideal pulse shape (as in `CMSSW`) to `RooFit`
5. Make a simple timetable for documenting candle analysis

2.6.2 Summary List

1. Script to do JES uncertainty fits
2. Ideal hcal pulse shape ported to `RooFit`

2.6.3 Setting up scripts to fit for JES uncertainty

Copied the fitting script from two days ago, and applied five times for both calo jet and PF jet. The script is done so once data come, I can just press button and it will do the fitting. For current data ($2.66pb^{-1}$), the result is shown in figures 2.4 and 2.5. The correction uncertainties are assumed to be

1. 10% overall scale for calojet
2. 5% overall scale for PF jet
3. 2% $\times$ $|\eta|$ for both types of jets.

2.6.4 Port ideal HCAL pulse shape to `RooFit` and include in the brainstorming plots

Copied out the `HcalPulseShapes.cc` and `HcalPulseShapes.h` from `CMSSW382`. The files are not dependent on root libraries or `CMSSW`, and I hacked the files to get the ideal shape. The ported pulse shape is shown in figure **??**.

2.6.5 Reflection

No work is done after going home....not good....

2.6.6 Goals for next work day

1. Make a list of things that need to be documented for the ZJet candle analysis
2. Check what Artur needs in terms of vecbos ntuples
3. Try out Hcal DQM test bench on lxplus.
4. Double-spike algorithm.

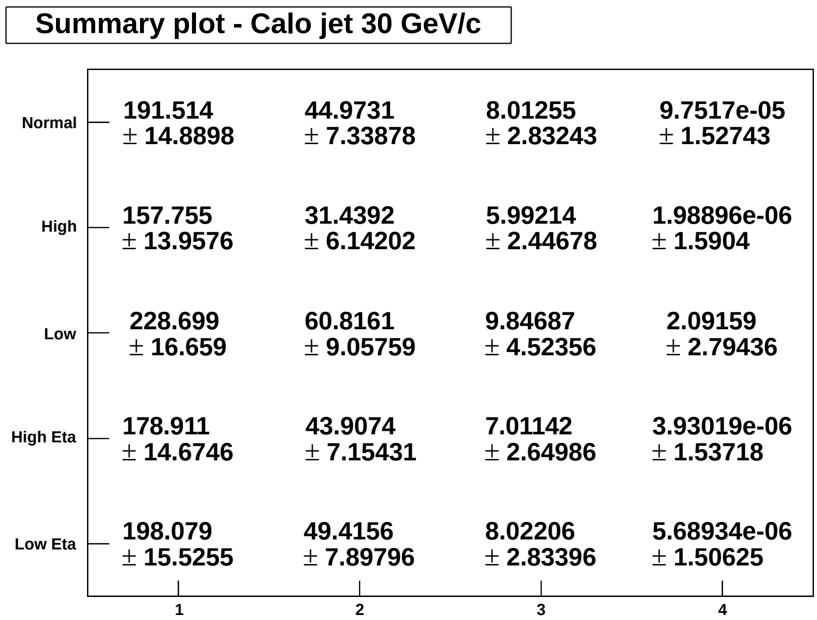


Figure 2.4: Jet energy correction uncertainty estimation for calojet with threshold 30 GeV/c.

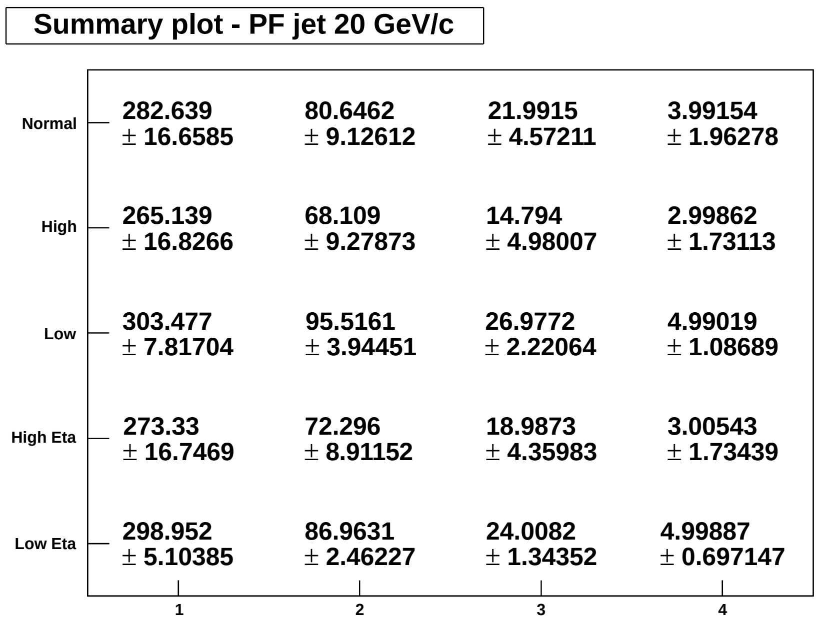


Figure 2.5: Jet energy correction uncertainty estimation for particle-flow jet with threshold 20 GeV/c.

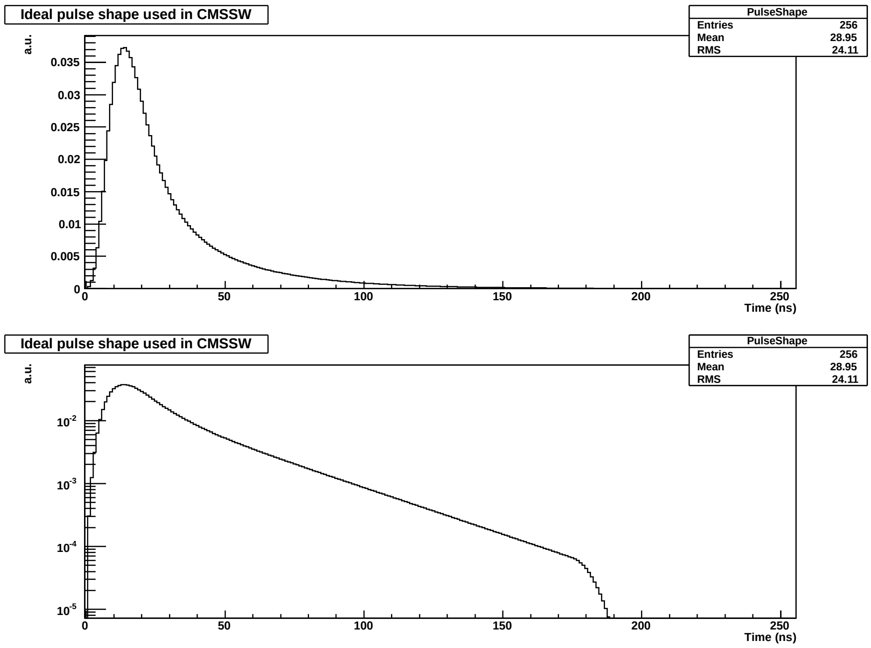


Figure 2.6: Extracted hcal ideal pulse shape from `CMSSW` version `3_8_2`. The two panels are the same histogram, except that one is in log scale.

2.7 6275 Log (October 8, 2010)

2.7.1 Goals

1. Make a list of things that need to be documented for the ZJet candle analysis
2. Check what Artur needs in terms of vecbos ntuples. Check out new ntuple dumper code and test run.
3. Try out Hcal DQM test bench on lxplus.
4. Double-spike algorithm.

2.7.2 Summary List

1. Checked out vecbos ntuple code (and compiled)
2. Custom-made Hcal pulse shape fitting working - preliminary results on double-chi2

2.7.3 Checking out new vecbos ntuple producer code

Followed the instruction in http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/UserCode/HiggsAnalysis/HiggsToWW2e/doc/README?revision=1.57, except that the version is changed to CMSSW 3_8_2.

```
scramv1 project CMSSW CMSSW_3_9_0_pre1
cd CMSSW_3_9_0_pre1/src
eval `scramv1 runtime -sh`

cvs co -r V04-3_9_X -d HiggsAnalysis/HiggsToWW2e UserCode/HiggsAnalysis/HiggsToWW2e
cvs co -r V00-3_5_X -d Configuration/ElectronIdentification \\
   UserCode/emanuele/Configuration/ElectronIdentification
addpkg RecoEgamma/ElectronIdentification
cvs co -r V02-3_6_0 -d MyAnalysis/IsolationTools \\
   UserCode/emanuele/MyAnalysis/IsolationTools

cp /afs/cern.ch/user/e/emanuele/public/4Likelihood/PDFsSQLite/CMSSW_3_2_X/electronIdLikeli
   HiggsAnalysis/HiggsToWW2e/test/python
cd RecoEgamma/ElectronIdentification/python/
rm likelihoodPdfsDB_cfi.py
wget http://emanuele.web.cern.ch/emanuele/ElectronID/patchLikelihoodDb/likelihoodPdfsDB_cf
cd -

scramv1 build
```

It builds fine.

2.7.4 Hcal pulse shape fitting

The long process from yesterday with `RooFit` finally ended after a few hours, and that's only a small percentage of the total events. And.....it doesn't look good. Signal distribution and noise distribution look the same. The black box named `RooFit` is too complicated, and it's hard to guess what really is happening. One iteration takes hours so it's bad.

So I tried to mathematically calculate the minimum and implement the result. Suppose T_i is the ten pedestal-subtracted charges, and F_i is the pulse shape, integrated into 25ns time slices. Then the only free parameter to fit to is α, the overall scale. So

$$\begin{aligned}
Error^2 &= \sum_{i=0}^{9}(T_i - \alpha F_i)^2 \\
&= \sum_{i=0}^{9} T_i^2 - 2\alpha \sum_{i=0}^{9} T_i F_i + \alpha^2 \sum_{i=0}^{9} F_i^2 \\
\frac{d(Error^2)}{d\alpha} &= -2\sum_{i=0}^{9} T_i F_i + 2\alpha \sum_{i=0}^{9} F_i^2 = 0 \\
\alpha &= \frac{\sum_{i=0}^{9} T_i F_i}{\sum_{i=0}^{9} F_i^2} \\
\min(Error^2) &= \sum_{i=0}^{9} T_i^2 - \frac{\left(\sum_{i=0}^{9} T_i F_i\right)^2}{\sum_{i=0}^{9} F_i^2}
\end{aligned}$$

Now all I need to do is to scan through different possible offsets (in units of ns), calculate the above quantity, and find the minimum. This works pretty well. Now the whole thing finishes within a couple minutes. And the result makes sense. One possible upgrade is to go down to sub-ns after the minimum region is identified....but let's not worry about that now. The fit result is shown in figures 2.7, 2.8, 2.9 and 2.10. Preliminary (just plotted) results on test statistics are shown in figures 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17 and 2.18.

As a next step I should upgrade all $Error^2$ to χ^2. I don't even need the conversion factor from charge to number of photoelectron! All that matters is the square-root dependence, and any additional factor will show up as shifts in the test variable. Plus, the factor will cancel away when I take the ratio.... This is for the weekend~

2.7.5 Artur's request of vecbos ntuples

1. `/JetMET/Run2010A-Sep17ReReco_v2/RECO` (Run Range 141956-144114)
2. `/Jet/Run2010B-PromptReco-v2/RECO` (Run Range 146240-147370)

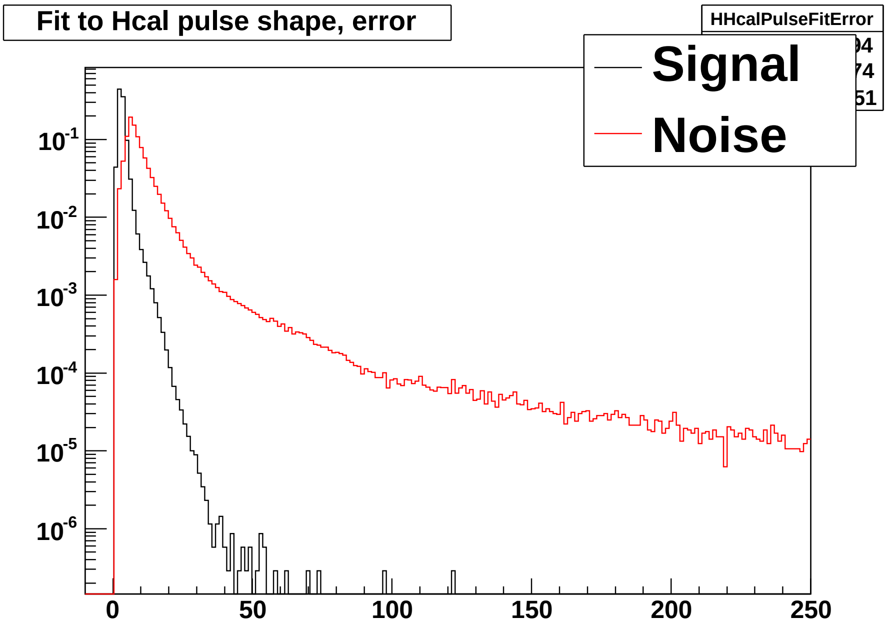


Figure 2.7: Error from fit to nominal shape using the above custom method.

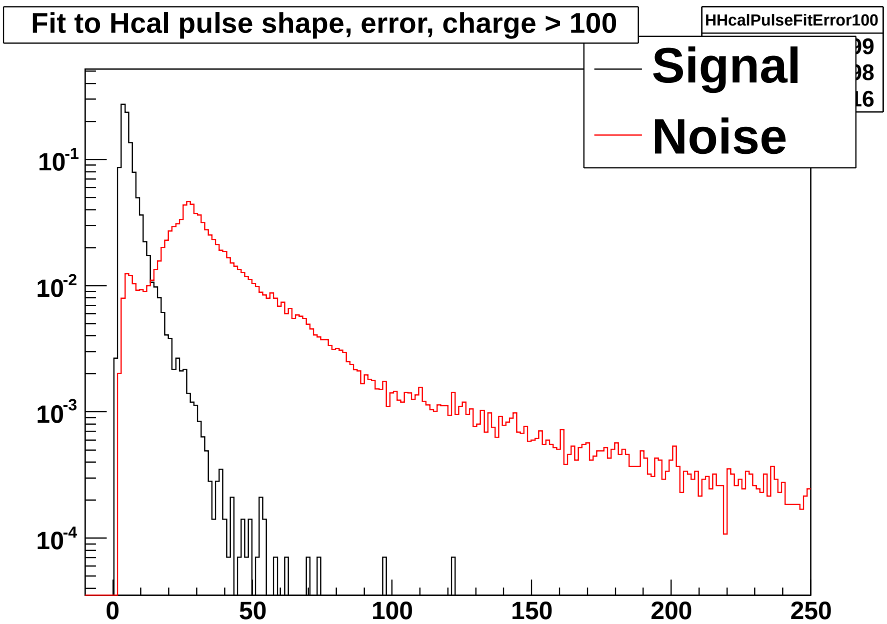


Figure 2.8: Error from fit for pulse with charge at least 100 fC

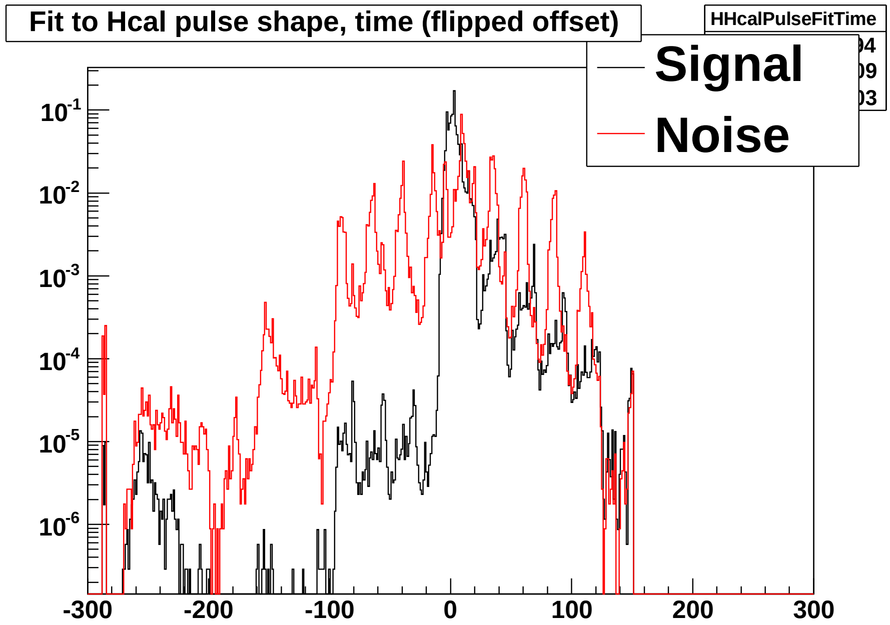


Figure 2.9: Peak time from fit to nominal shape using the above custom method.

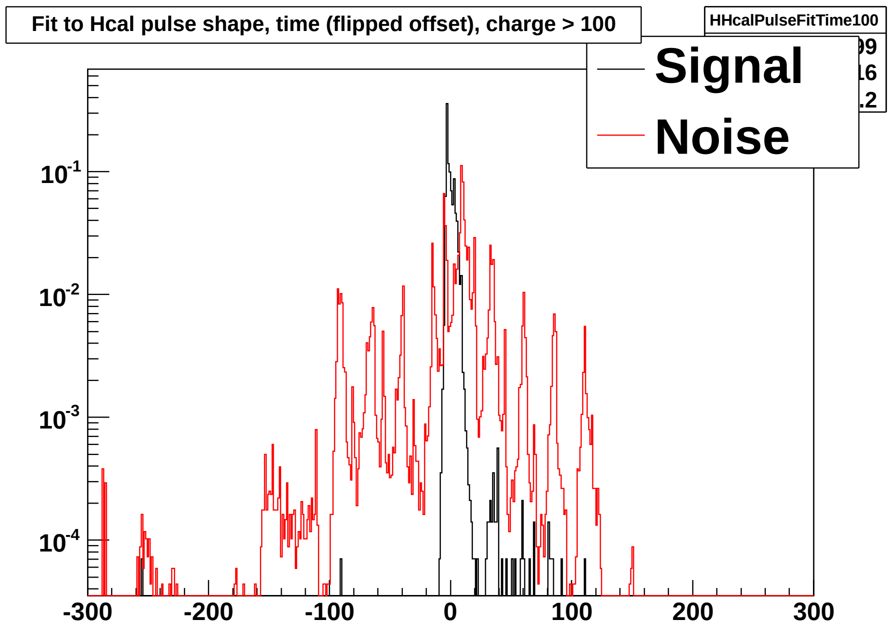


Figure 2.10: Peak time from fit for pulse with charge at least 100 fC

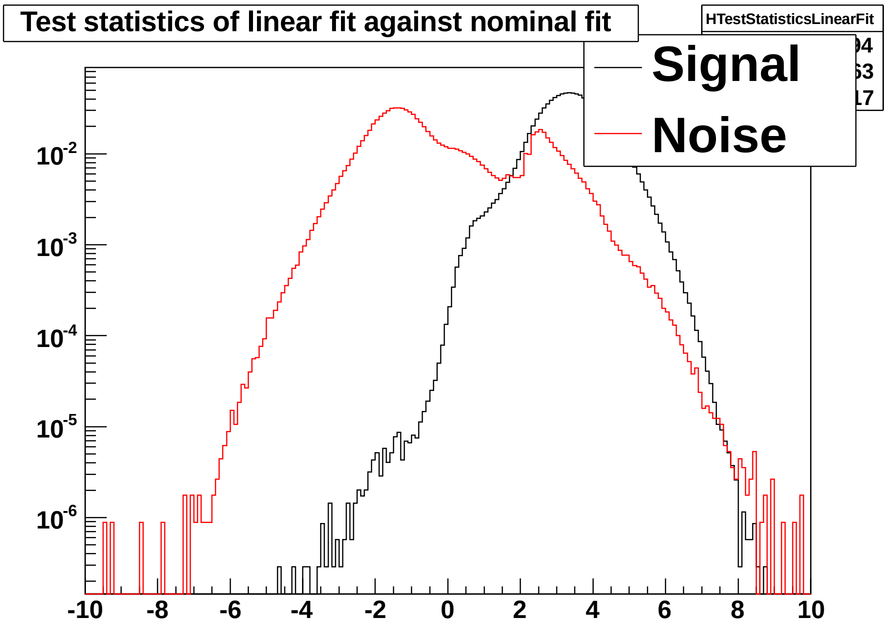


Figure 2.11: Test statistics on linear fit.

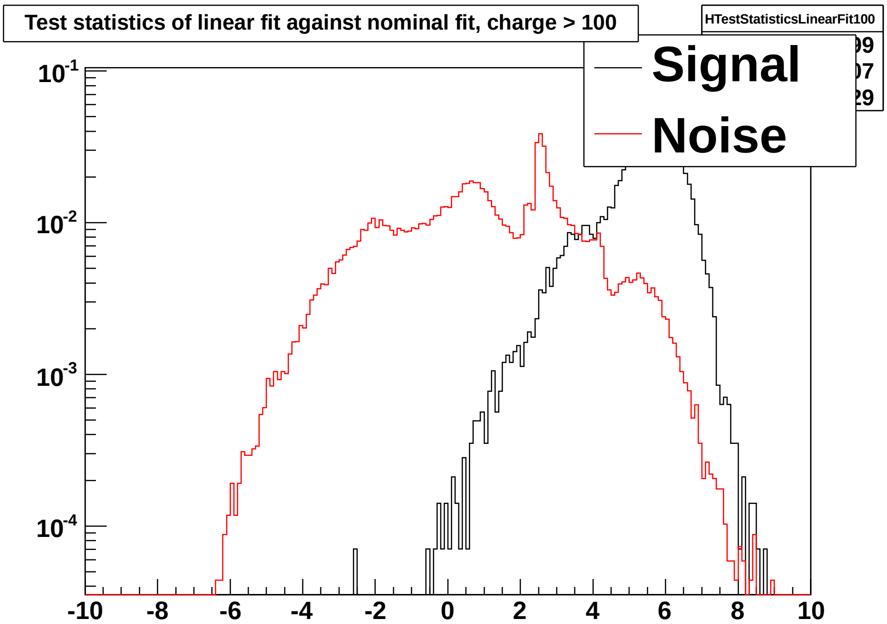


Figure 2.12: Test statistics on linear fit, charge >100 fC.

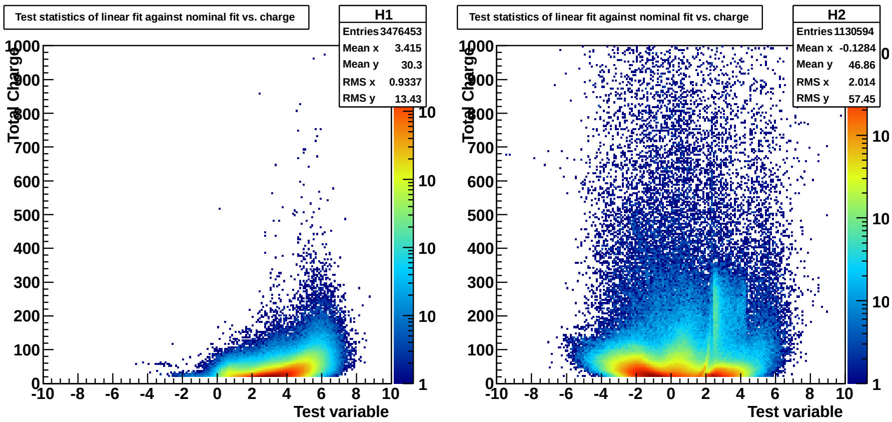


Figure 2.13: Test statistics on linear fit. Left: signal, right: noise

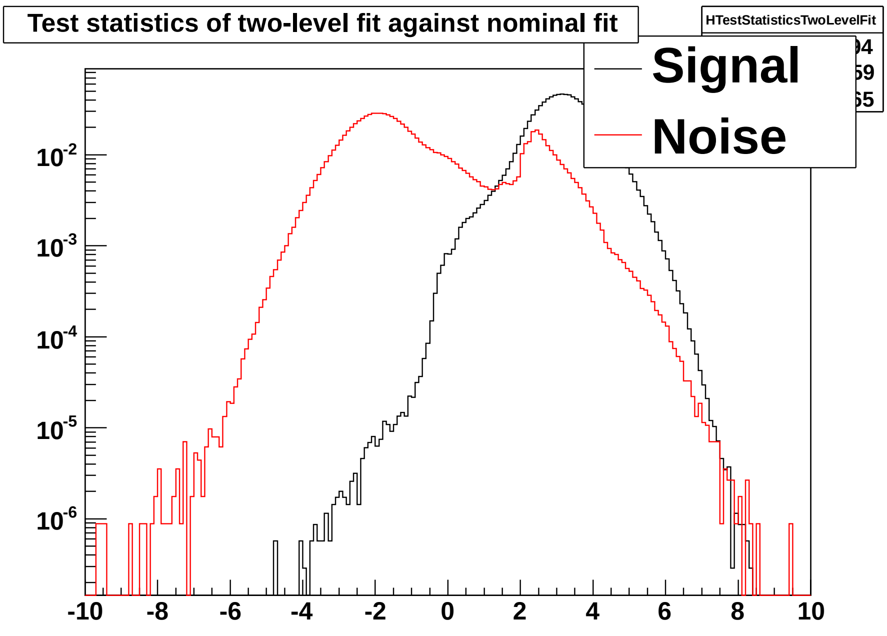


Figure 2.14: Test statistics on two-level fit.

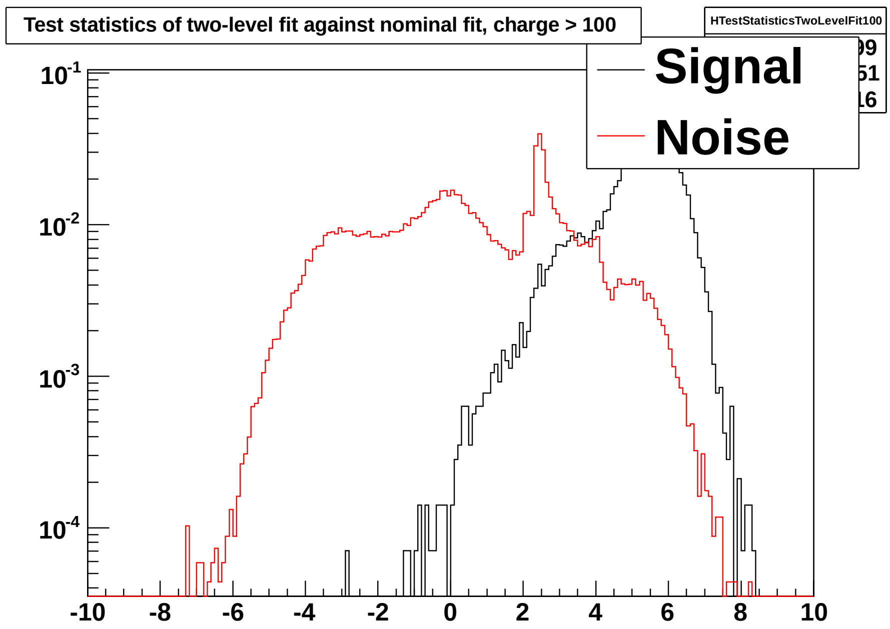


Figure 2.15: Test statistics on two-level fit, charge >100 fC.

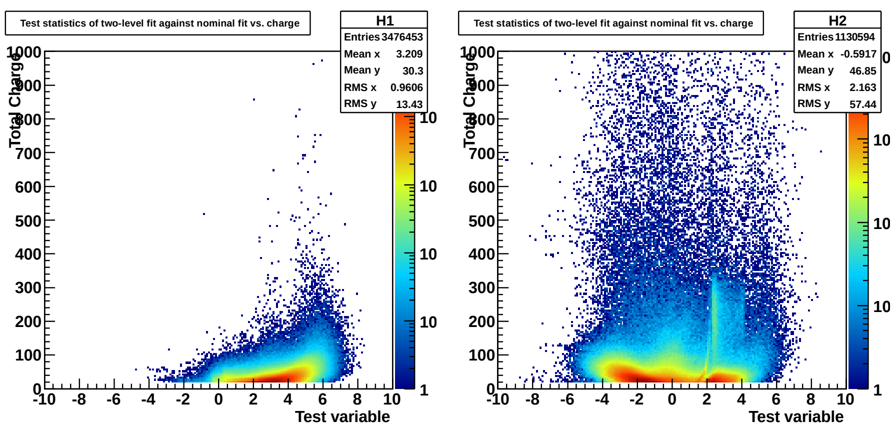


Figure 2.16: Test statistics on two-level fit. Left: signal, right: noise

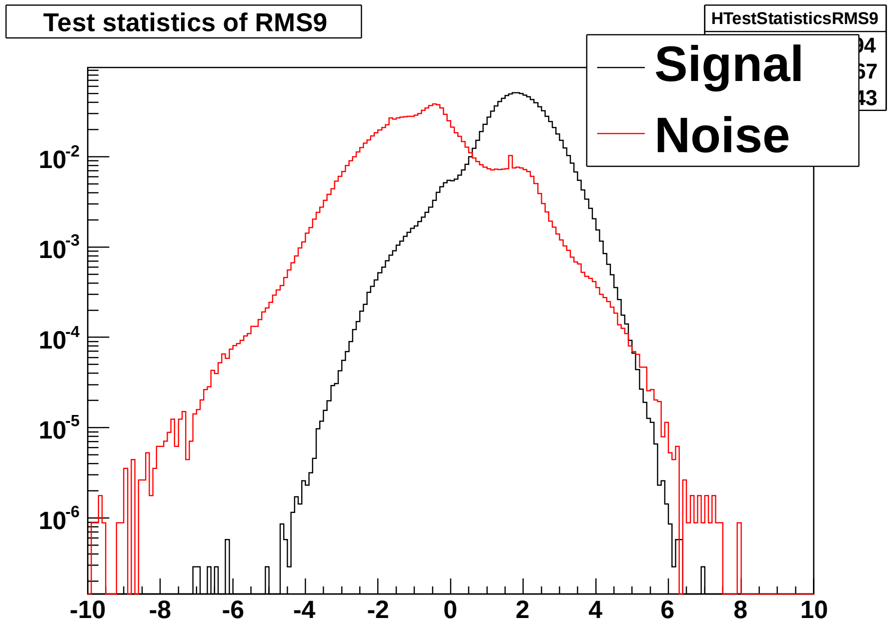


Figure 2.17: Test statistics on RMS9.

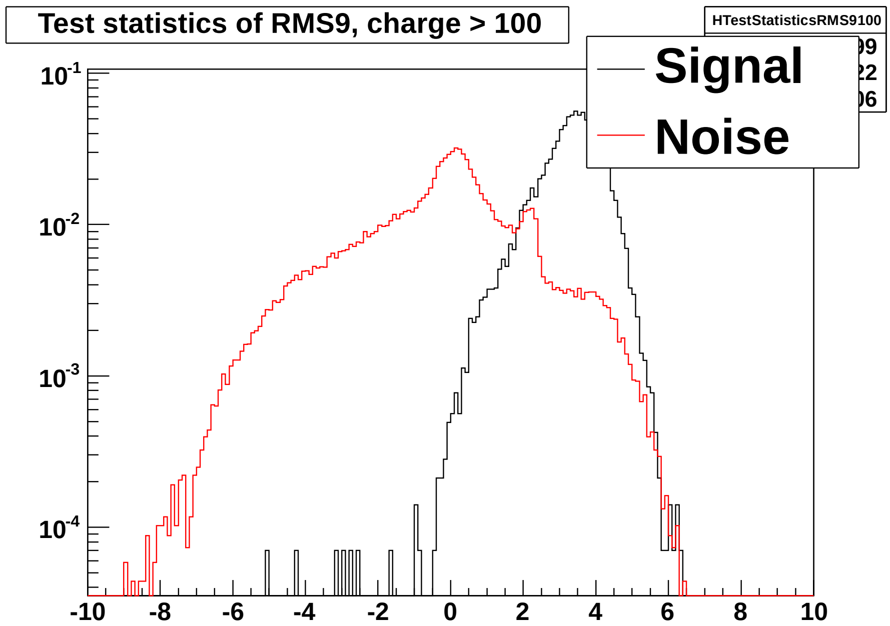


Figure 2.18: Test statistics on RMS9, charge >100 fC.

3. `/METFwd/Run2010B-PromptReco-v2/RECO` (Run Range 146240-147370)

Will begin running during the weekend, and hopefully they will finish early next week.

2.7.6 Reflection

2.7.7 Goals for next work day

1. Submit vecbos ntuple jobs
2. Make a to-write list for Z candle
3. Do one round of subtraction and look at remaining noise
4. Try out Hcal DQM GUI on lxplus

2.8 6267 Log (October 10, 2010)

2.8.1 Goals

1. Submit vecbos jobs!

2.8.2 Summary List

1. Submitted vecbos jobs - with a lot of resubmit....

2.8.3 Vecbos ntuple dumper job submission

From the checkout in CMSSW_3_8_2, the template did not work out of the box. I had to add in btagging-related configuration files and sequences

```
process.load("HiggsAnalysis.HiggsToWW2e.btagProducerSequence_cff")
process.load("HiggsAnalysis.HiggsToWW2e.btagPFJetsProducerSequence_cff")
process.load("HiggsAnalysis.HiggsToWW2e.btagJPTJetsProducerSequence_cff")
...
   process.newBtaggingSequence *
   process.newPFJetBtaggingSequence *
   process.newJPTJetBtaggingSequence *
...
```

to make it work. Also I have to go into HiggsAnalysis.HiggsToWW2e.jetProducerSequence_cff and remove the line which includes JetMETCorrections.Configuration.JetCorrectionCondDB_cff. There is also a problem on castor. One needs to change permission to 775 to be able to stage out to the directory:

```
rfchmod 775 /castor/cern.ch/user/....
```

Now it seems to be working.....Hopefully nothing else go wrong.

2.8.4 Reflection

Playing with crab/castor is really a waste of time.....

2.8.5 Goals for next work day

1. Submit vecbos ntuple jobs
2. Make a to-write list for Z candle
3. Do one round of subtraction and look at remaining noise
4. Try out Hcal DQM GUI on lxplus

2.9 6311 Log (October 11, 2010)

2.9.1 Goals

1. Submit vecbos ntuple jobs
2. Make a to-write list for Z candle
3. Do one round of subtraction and look at remaining noise
4. Try out Hcal DQM GUI on lxplus

2.9.2 Summary List

1. Yay

2.9.3 Vecbos ntuple jobs

Submitted jobs for `Jet` and `METFwd` datasets. Wait until afternoon, if things are looking good then I will submit the `JetMET` dataset. This one is a lot larger than the other two newer datasets.

2.9.4 Update errors to χ^2 instead of $error^2$

Just replace all error calculation parts by χ^2, ie., divide the deviations by the charge of that time slice. Let's see if the result will be improved.

The linear one seems better! Shown in figures 2.19 and 2.20. The small group of pulses in signal distribution that is smaller than zero needs more attention. If they were noise, that will be really great.

2.9.5 Vecbos candle note

Let's form the items that need to be written here...

1. Introduction
2. Data & MC samples
3. ECAL spike cleaning
4. Jet flavors.
5. Selections. Muon, Z. Isolation
6. Signal extraction (fit strategy)
7. Toy studies: shape, parameters
8. Toy studies: show that the errors are reasonable (!)
9. JES uncertainty
10. Anti-muon control sample

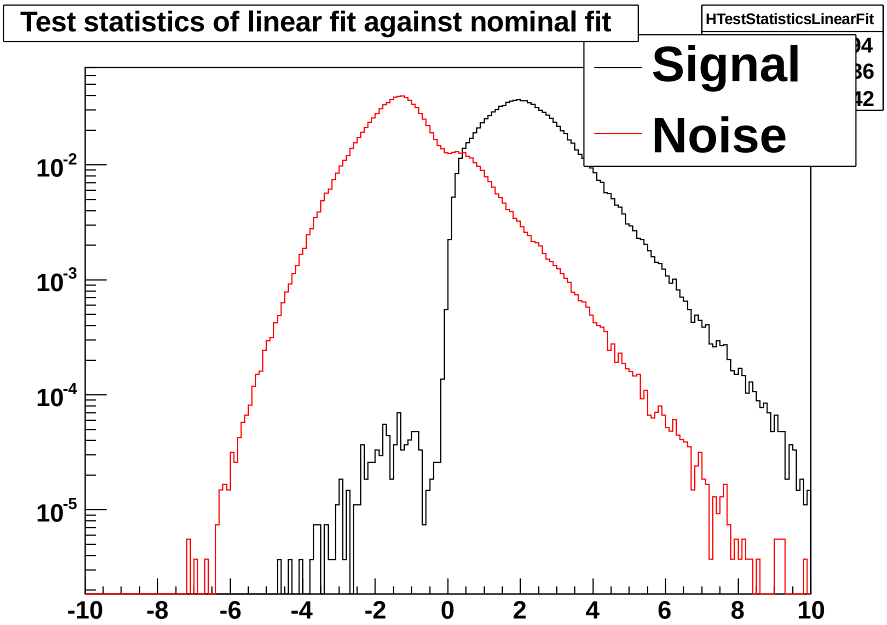


Figure 2.19: Test statistics on linear fit.

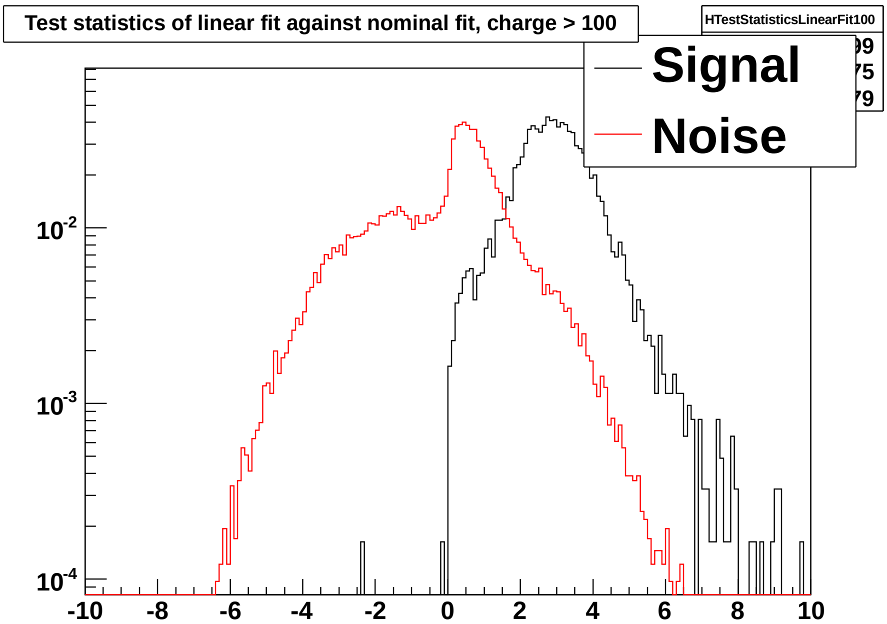


Figure 2.20: Test statistics on linear fit, charge >100 fC.

2.9.6 Installing (temporary) Hcal DQM GUI on lxplus

Following instructions on https://twiki.cern.ch/twiki/bin/view/CMS/DQMTest....and the DQM GUI is there. `=___+`

After starting the GUI and the collector, run on `hcal_dqm_sourceclient-file_cfg.py` (with a few parameters changed). While the job is running, the GUI responds and I can see the plots. But once it's done running, the DQM plots are not there anymore....

Now I should ask what Artur wants to be done for the DQM.

As a start, the following can be done relatively quickly according to Artur:

1. Understand the basic histograms (especially the ones in the summary page)
2. Adding code from Shuichi
3. Check if there is any other useful parameters that could be incorporated into the DQM.

2.9.7 Reflection

2.9.8 Goals for next work day

1. Write Z candle note - have all the individual paragraphs ready by end of Wednesday!
2. Read Hcal DQM code
3. Double-spike....?
4. Setup script that takes the output of the Z candle fit and calculate ratio

2.10 6312 Log (October 12, 2010)

2.10.1 Goals

1. Write Z candle note - have all the individual paragraphs ready by end of Wednesday!
2. Read Hcal DQM code
3. Double-spike....?
4. Setup script that takes the output of the Z candle fit and calculate ratio

2.10.2 Summary List

1. Started writing down Z candle note
2. Took some time checking the `Jet` and `JetMET` vecbos ntuple production status.

2.10.3 Vecbos jobs status

Most of the jobs are aborted by site `T2_BE_IIHE`... Remeber to add it to blacklist in the future!!!

`METFwd` is done, `Jet` will be done soon, and as for `JetMET` I still need to fight the the bad T2.

2.10.4 Reflection

Oops....

2.10.5 Goals for next work day

1. Continue jotting down Z candle draft fragments.
2. Make sure the jobs are running
3. ....

2.11 6314 Log (October 14, 2010)

2.11.1 Goals

1. Keep writing down Z candle note fragments
2. Meet with Maria to learn more about W fit
3. Start leptoquark jobs
4. Move the toy fit code to the computer and start rerunning
5. Resubmit the failed jobs
6. Keep reading Hcal DQM code

2.11.2 Summary List

1. Yay

2.11.3 DQM code!

An analyzer (`HcalMonitorModule`) keeps filling histograms and keeping them in memory. The histograms (embedded in `MonitorElement`) are "booked" through `DQMStore` service and the ownership of the histogram memory is the service....whatever it is. There are a list of "clients" that correspond to subfolders....and same story with the `DQMStore` and histogram booking. The entrance module for the clients is `HcalMonitorClient`. To make sure that I know what's going on, let me introduce a new subfolder called "HcalFHeadMonitor" where I can put my own histograms. Good. Successfully installed the new subfolder. See figure 2.21. Try to add a histogram. It showed up wouldn't fill at all. Let's play with it tomorrow.

2.11.4 Meeting about WJet simple fit

This is the simple fit before Lukas finished with his b-tag complicated fit strategy studies.

Ilaria and somebody are going to do the fit, and I am charged to create the datasets by the end of the week.

Will showed me his code structure on making datasets.

2.11.5 Reflection

Meow.

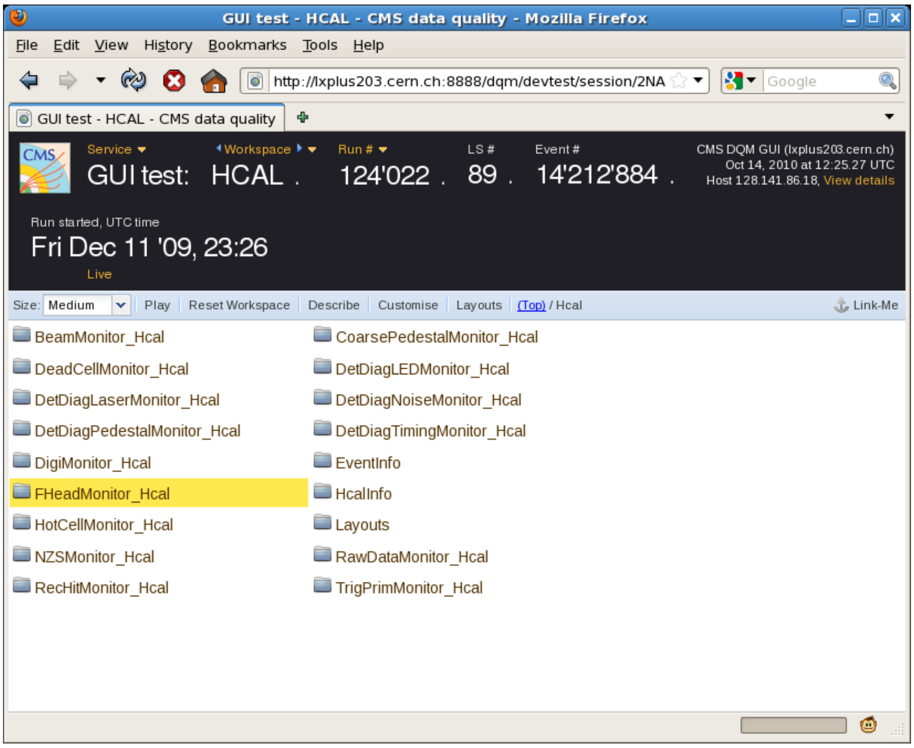


Figure 2.21: Successfully installed FHead subfolder in Hcal DQM test bench

2.12 6315 Log (October 15, 2010)

2.12.1 Summary List

1. Spent a lot of time debugging the WJet dataset code....

2.12.2 Note on WJet dataset code debugging

I keep getting "glibc: memory corruption" errors. It happens when my `MuonCandidate` has all the fields in. If I remove any of them, then the code runs fine. I ended up removing the combined isolation since people won't be using it.

2.12.3 Reflection

Debugging has taken a lot of time. Maybe I haven't written program for too long. As a result it was not a productive day.

2.13 6321 Log (October 18, 2010)

2.13.1 Goals

1. Vecbos candle data exercise: start after lunch!
2. Hcal DQM: finish the implementation of task module
3. Hcal noise: look at correlations between the three test variables
4. Hcal noise: have a preliminary version of presentation

2.13.2 Summary List

1. Sent out the `RooDataSets` to people.
2. Looked at the correlations between test variables
3. Fixed a fatal error in the two-level fit. Now the chi2 is always positive, and the result look a lot more like the linear fit
4. Made a preliminary version of the presentation.
5. Finished implementation of the FHead task module in private Hcal test bench. The histogram fills!!!
6. Added parabolic fit into the brainstorm reservoir.

2.13.3 Note on progress during Sunday

1. Started the leptoquark jobs. Expect them to be finished soon.
2. Don't know how to run Will's code. I will ask him later, but for the moment I'll write my own `RooDataSet` converter.
3. Successfully exported `RooDataSets` for W, Z, $t\bar{t}$, single t (s, t, tW), QCD (ppMuX).

2.13.4 Hcal noise test statistics correlation

Correlation between linear fit vs. two-level fit, as well as linear fit vs. RMS-series are plotted. The results between linear and two-level fits are shown in figures 2.22 and 2.23. Results between linear and combinations of RMS8/9 and max charge/total charge are shown in figures 2.24 - 2.31.

As we can see, the two level fit is highly correlated with linear fit, and we might not be able to gain much from this. The RMS-series, however, picks out the spike ones and is a more natural handle to pick out those than the linear fit. While linear fit could cut out the spike ones (if we squeeze the envelope enough....and also make it dependent on total charge), the RMS-series can be complementary. Overall the max charge make a better denominator, since the signal cluster is higher. Comparing RMS8/Max (2.25) and RMS9/Max (figure 2.27), The signal is more concentrated in RMS9/Max, but there is a clearer noise region in RMS8/Max. And RMS8/Max seems to have better separation between the signal cluster position and the lower wing of noise.

The scatter plot of (test statistics of) RMS8/Max vs. RMS9/Max (figures 2.32 and 2.33) show that RMS8/Max seems to be a better one to use.

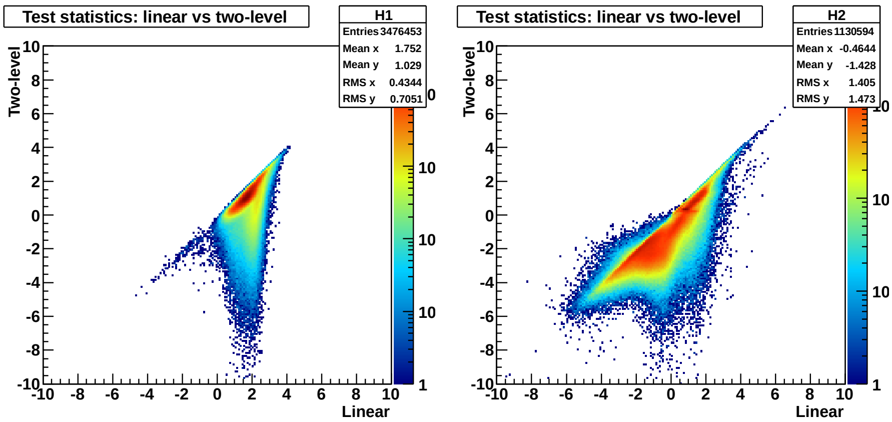


Figure 2.22: Test statistics of linear fit vs. two-level fit. Left: signal, right: noise

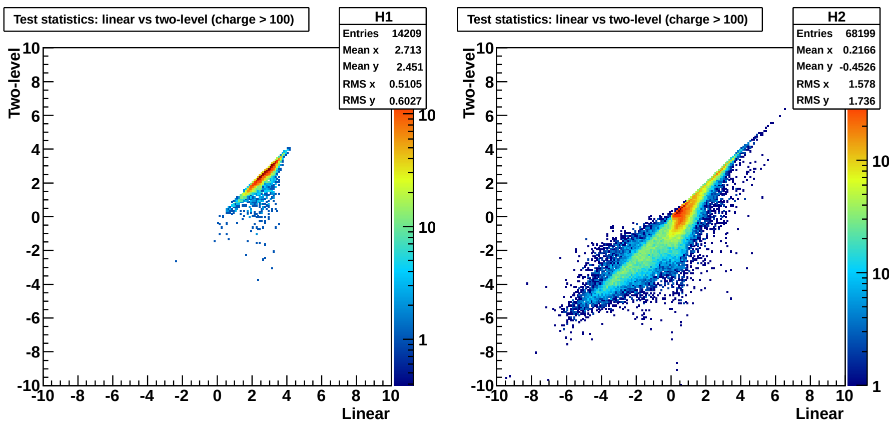


Figure 2.23: Test statistics of linear fit vs. two-level fit, charge >100. Left: signal, right: noise

2.13.5 Parabolic fit

The parabolic fit seems to be only a bit better than linear fit. I recall seeing a lot of small-hill-like shapes: rise up slowly for a few time slices, and then fall back down. The fact that the improvement is not drastic means that the humps are still pretty efficiently picked out by the linear hypothesis. The chi2 can be seen in figure 2.34, test statistics in figure 2.35 and the correlation between parabolic fit and linear fit is shown in figure 2.36 and 2.37.

2.13.6 Vecbos candle data exercise

Start at 12:34 computer time. Goal is to produce all plots/tables used in the notes that are related to data as soon as possible. Hopefully I can get all of them done in one afternoon. While it is running, let me compile a list of plots/tables/numbers to produce.

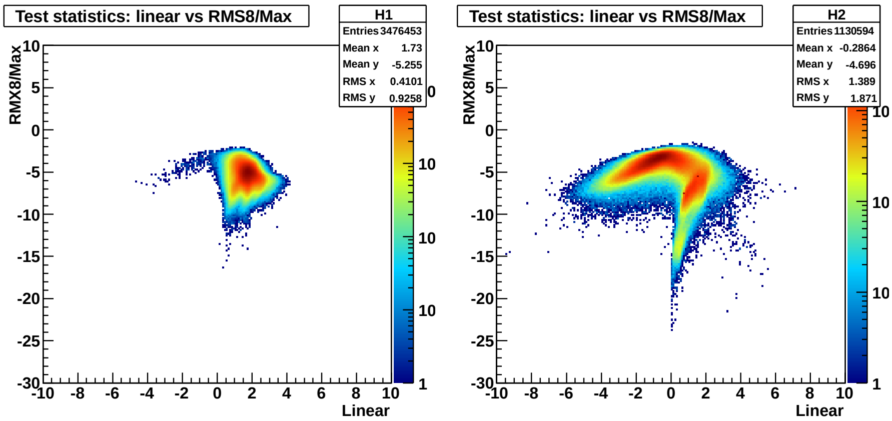


Figure 2.24: Test statistics of linear fit vs. RMS8/max charge. Left: signal, right: noise

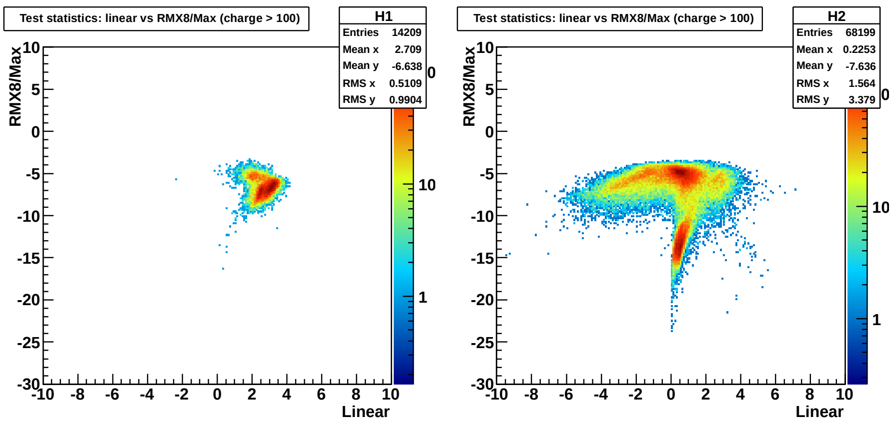


Figure 2.25: Test statistics of linear fit vs. RMS8/max charge, charge >100. Left: signal, right: noise

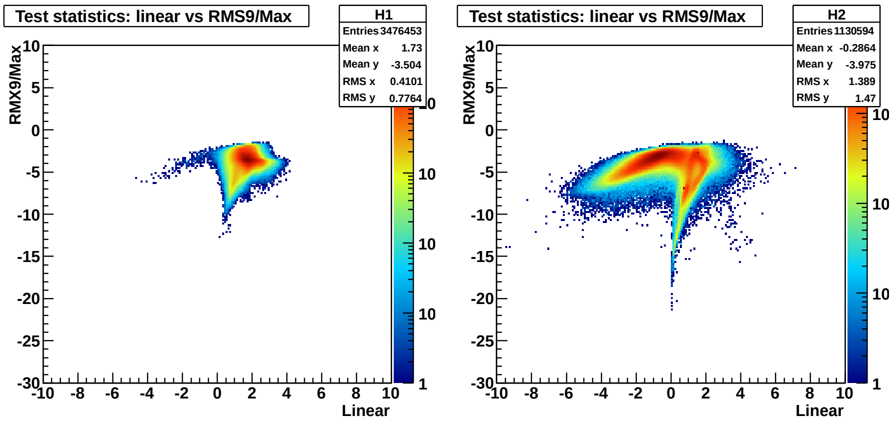


Figure 2.26: Test statistics of linear fit vs. RMS9/max charge. Left: signal, right: noise

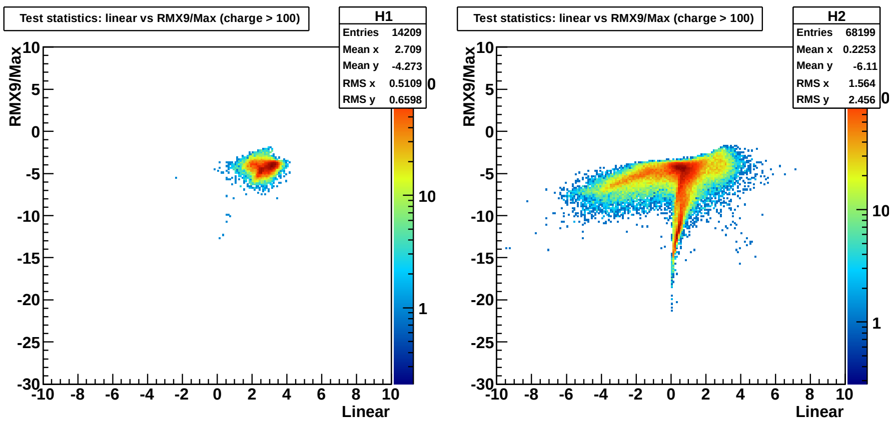


Figure 2.27: Test statistics of linear fit vs. RMS9/max charge, charge >100. Left: signal, right: noise

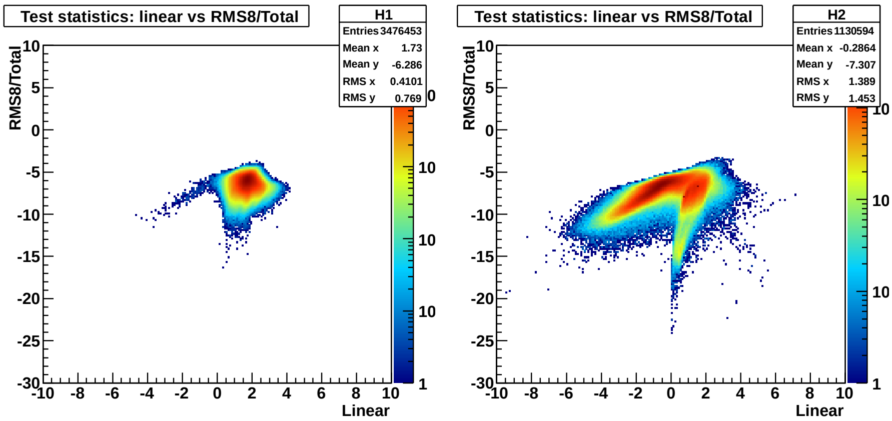


Figure 2.28: Test statistics of linear fit vs. RMS8/total charge. Left: signal, right: noise

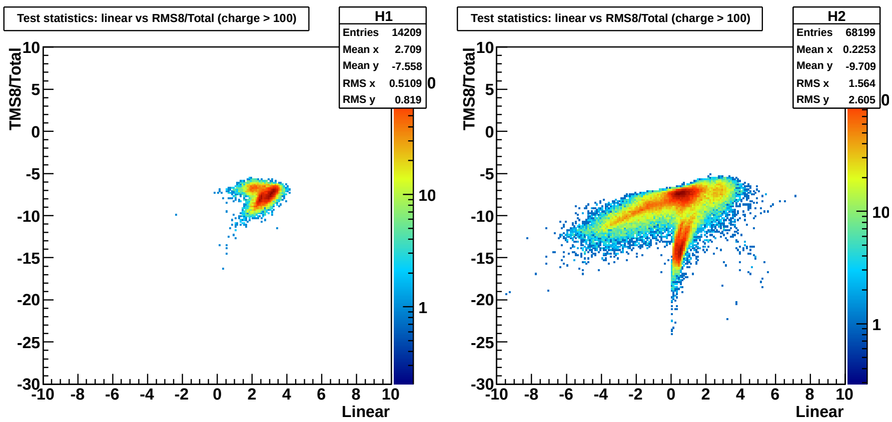


Figure 2.29: Test statistics of linear fit vs. RMS8/total charge, charge >100. Left: signal, right: noise

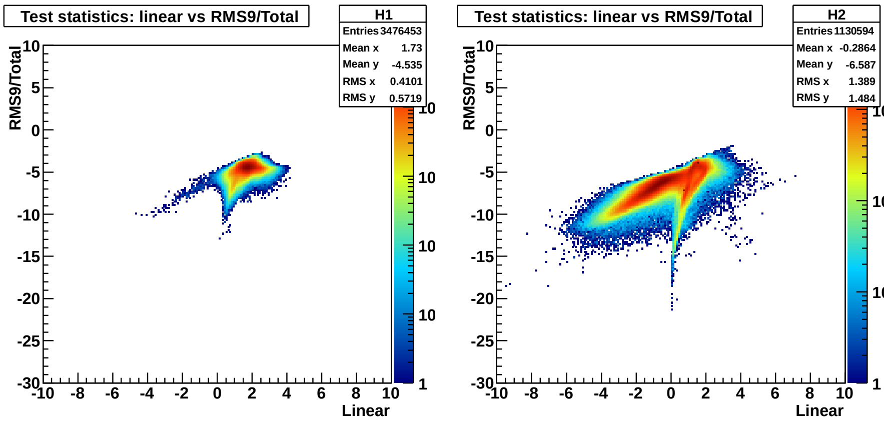


Figure 2.30: Test statistics of linear fit vs. RMS9/total charge. Left: signal, right: noise

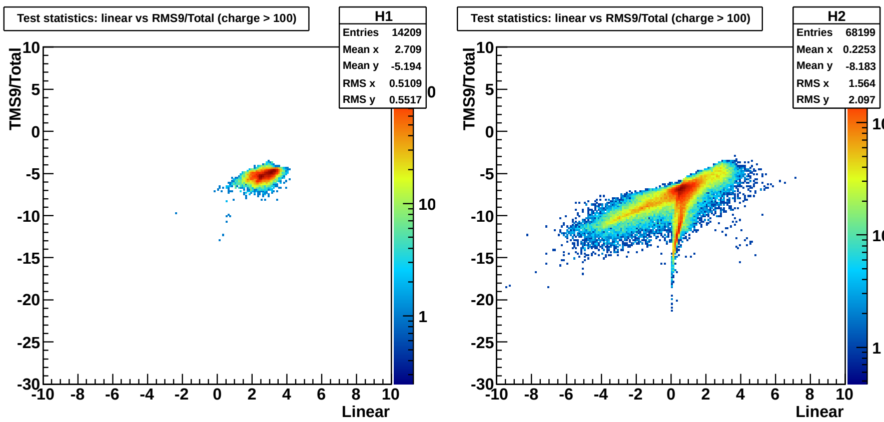


Figure 2.31: Test statistics of linear fit vs. RMS9/total charge, charge >100. Left: signal, right: noise

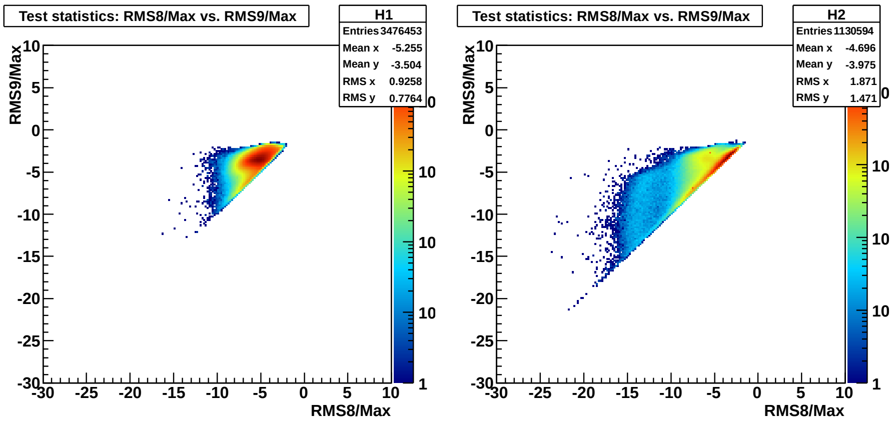


Figure 2.32: Test statistics of RMS9/max and RMS8/max. Left: signal, right: noise

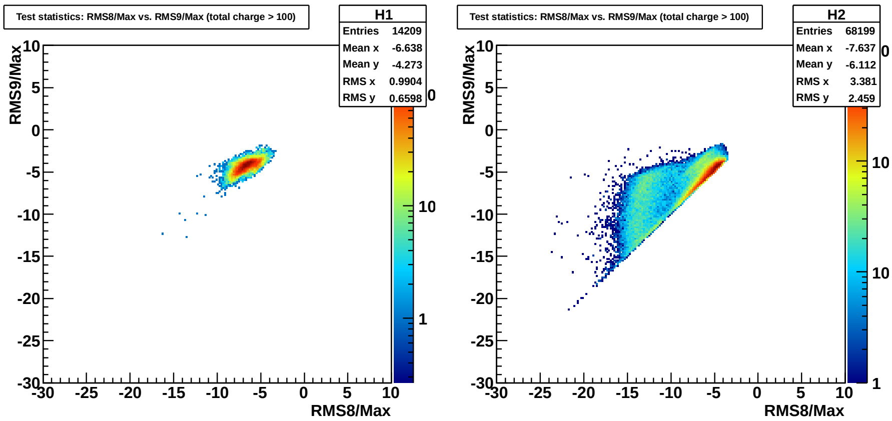


Figure 2.33: Test statistics of RMS9/max and RMS8/max, charge >100. Left: signal, right: noise

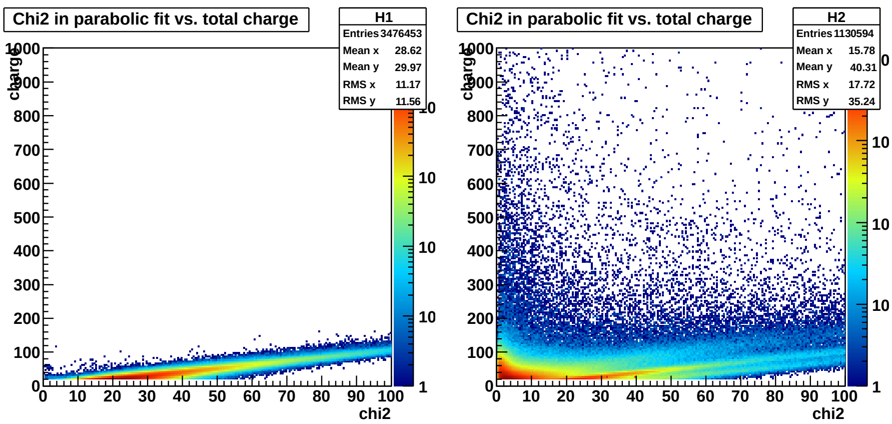


Figure 2.34: Chi2 of parabolic fit. Left: signal, right: noise

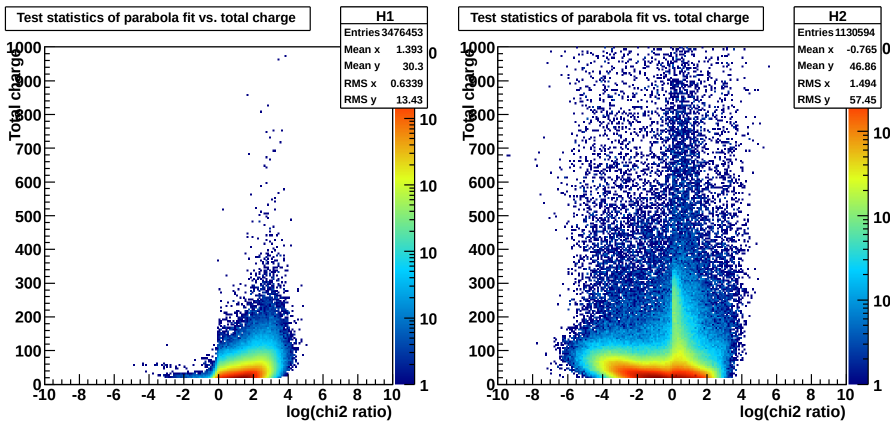


Figure 2.35: Test statistics of parabolic fit. Left: signal, right: noise

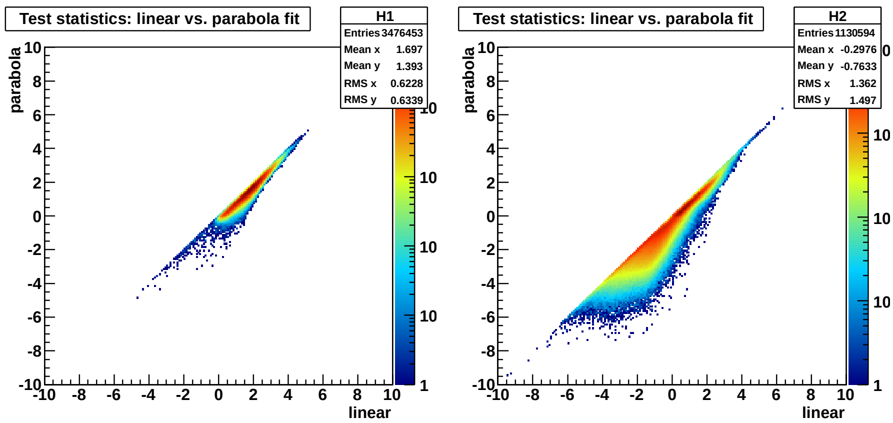


Figure 2.36: Test statistics of parabolic fit and linear fit. Left: signal, right: noise

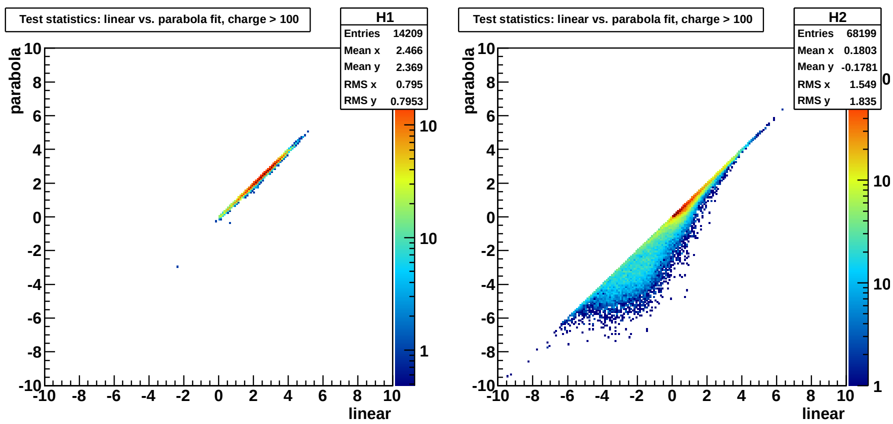


Figure 2.37: Test statistics of parabolic fit and linear fit, charge >100. Left: signal, right: noise

1. Fit with tight isolation and extract α_L. Compare with normal isolation cut.

2. Simultaneous fit

3. Data extracted yields

4. Get the ratio from the yields

5. Redo fit with different jet threshold and produce JES uncertainty estimate

6. Anti-isolation to produce QCD control shape

7. Check basic plots to see if everything looks roughly okay.

Now it's 14:27. I'm still submitting jobs. :(Splitting jobs in smaller segments makes it a lot faster. 14:43. Jobs done. Finished submitting job at 14:31. It takes about 20 minutes to submit all jobs, so....the total turnaround time on my side is about 30 minutes for around 3/pb. Copying takes a while though... 15:18. Finished copying.

The fits to MC signal sample takes a lot of time. Don't redo it again when data come. Todo item #1: make a script to output the fit result tables directly. For now let me proceed. Signal inclusive fits finish in a flash. Todo item #2: make a script to output the fit result lables. The JES fits are also really fast. Todo item #3: ditto. Todo item #4: setup script to produce the ratio.

Uhm....all the scripts run fine.... So I only need to write scripts to generate tables in latex format and I'm done with data part of the note (also to get the ratio). Tomorrow let me do a MC rush to generate all MC plots/tables.

2.13.7 Hcal DQM progress

Finished implementing the Hcal DQM task analyzer. Also changed a bit in the configuration file to make it run....now it fills plots! A dummy histogram is now in.

2.13.8 Reflection

Not enough time! But things are finally moving.

2.13.9 Goals for next work day

1. Z candle MC rush

2. Finish up vecbos jobs

3. Include Shuichi's code into the Hcal DQM test bench

4. Include double-chi2 variables into the Hcal DQM test bench

5. Finish up Hcal noise presentation

2.14 6322 Log (October 19, 2010)

2.14.1 Goals

1. Z candle MC plots - recreate all plots and tables (and identify things that are missing)
2. Implement Shuichi's code into the DQM structure
3. Make the digi tree/pulse shape tree of those picked events
4. Finish up Hcal noise presentation. Add a couple slides on parabolic fit.
5. Finish up vecbos jobs
6. Investigate what the small group of pulses are in signal events where linear fit is better
7. Lower priority: Introduce double-chi2 variables into Hcal DQM structure

2.14.2 Summary List

1. Finished up Hcal noise presentation draft. Sent to Maria and Artur for comments.
2. Made a skim of the 35 events from last Friday.
3. Code from Shuichi's student is only available inside FNAL. :(
4. Investigated the small group of pulses in the linear test variable.

2.14.3 Comments on the hcal noise presentation

1. Check if the chi2 distribution looks like chi2
2. Try to make the RMS8/Max into a fit.
3. Event selection. Use other parts of detector to make sure there is less signal contamination.
4. Give the test variables names.

2.14.4 The small group of pulses in the signal population where linear test variable is negative

All of them look like noise. There is no peak anywhere in the 10 time slices. I haven't looked at the geometrical distribution, but it should be pretty safe to cut out.

2.14.5 Reflection

???

2.14.6 Goals for next work day

1. Create MC plots for Z candle
2. Submit MultiJet jobs
3. Try out low-N fit

2.15 6323 Log (October 20, 2010)

2.15.1 Goals

1. Create MC plots for Z candle
2. Submit MultiJet jobs
3. Try out low-N fit

2.15.2 Summary List

1. Created all MC plots for Z candle note, as well as the tables. Now the only thing left is the toys.
2. The low-N fit doesn't work well.
3. Updated data plots up to....11-ish /pb

2.15.3 Reflection

Making all the plots and tables actually isn't as bad as I had thought.

2.15.4 Goals for next work day

1. Re-scan through high energy pulse shapes in the MET-mismatch events
2. Update Hcal noise plots to new name
3. Revisit low-N fit
4. Move Z candle toy MC code to desktop computer and start running
5. Start MultiJet jobs
6. Finish up old jobs that Artur asked and send out noifications

Table 2.2: I will fill this in later today.....

S	A	C	E

2.16 6324 Log (October 21, 2010)

2.16.1 Goals

1. Re-scan through high energy pulse shapes in the MET-mismatch events
2. Update Hcal noise plots to new name
3. Revisit low-N fit
4. Move Z candle toy MC code to desktop computer and start running
5. Start MultiJet jobs
6. Finish up old jobs that Artur asked and send out noifications

2.16.2 Summary List

1. Updated hcal noise plots to call the test variable "lambda"
2. Scanned through all of the MET-mismatch events one by one
3. Revisited LowN fits
4. Presented at Hcal noise WG meeting
5. Checked duplicates of currently-finished vecbos jobs
6. There are a few runs that were skipped by the launching script. Submitted those.

2.16.3 Scan through MET events from Dinko last Friday

2.16.4 Revisit LowN fits

There are a couple stupid error that makes the result yesterday horrible. After fixing them, the result looks much more reasonable. Shown below in figures 2.38, 2.39, 2.40 and 2.41.

The seperation power is not as great as RMS8/Max, though a clear group can be seen in the chi2 vs. charge plot. It will probably need a couple more twist (include somehow the information of height in the excluded two times slices) before being useful and replace RMS8/Max.

2.16.5 Reflection

2.16.6 Goals for next work day

1. Put codes from Shuichi's student and also the Jason variable and pulse shape fitting in Hcal DQM.
2. Check Matthias' note and see what I need to put in.

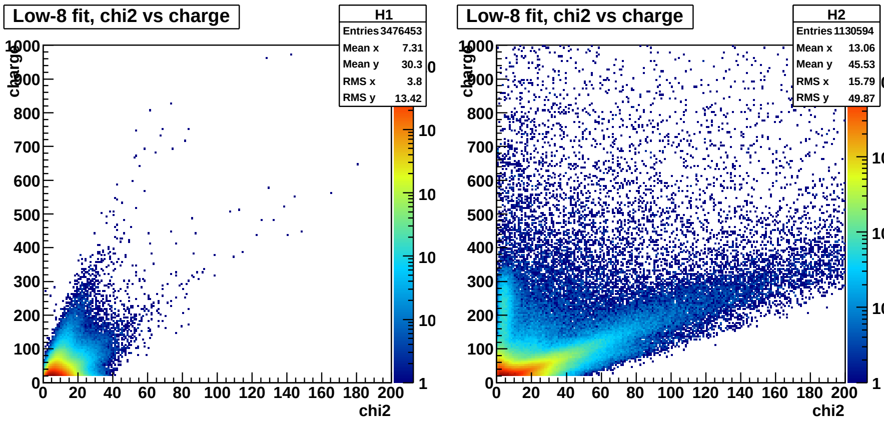


Figure 2.38: Low-8 fit chi2 vs. charge

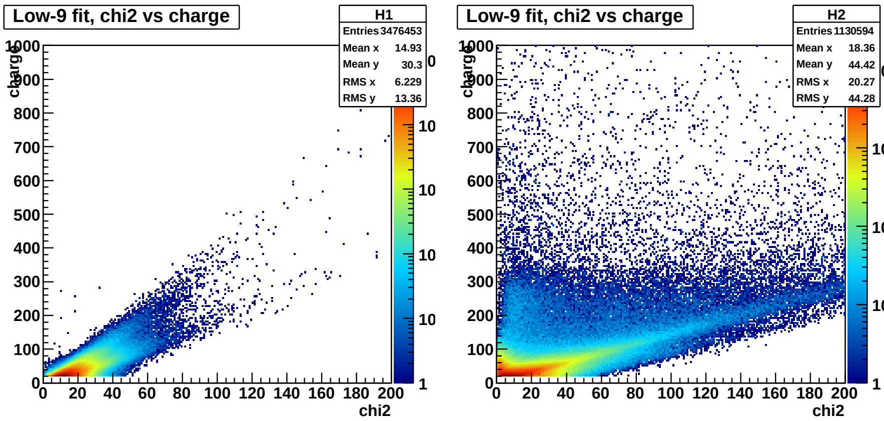


Figure 2.39: Low-9 fit chi2 vs. charge

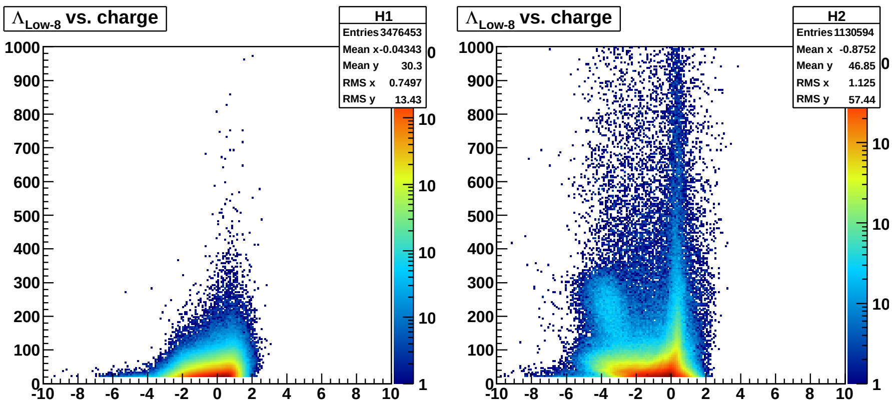


Figure 2.40: Low-8 fit test statistics vs. charge

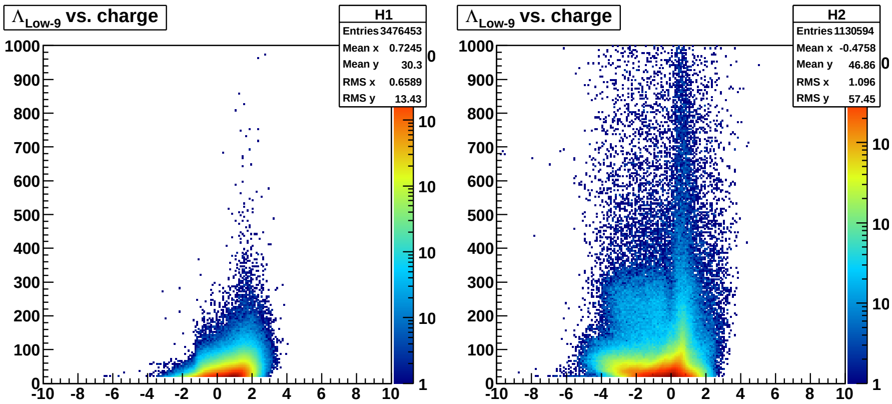


Figure 2.41: Low-9 fit test statistics vs. charge

3. Make a plan of rerunning Hcal digi trees and include some more information.

2.17 6325 Log (October 22, 2010)

2.17.1 Goals

1. Put codes from Shuichi's student and also the Jason variable and pulse shape fitting in Hcal DQM.
2. Check Matthias' note and see what I need to put in.
3. Make a plan of rerunning Hcal digi trees and include some more information.

2.17.2 Summary List

1. Implemented the test variables (and Jason's TS4TS5 variable) into offline DQM testbench.
2. Putting part of Shuichi et al.'s code into the DQM testbench.
3. Added an ad-hoc version of JP's isolation filter into the DQM testbench. Everything becomes so slow....

2.17.3 Check code from Shuichi et al

Starting from `HcalNoiseMon/NoiseAnalyzer/src/HcalNoiseInfoAnalyzer.cc`. Clean collision events if any of these are satified:

1. IC5 calo jet. Reject if any jet with 8 GeV/c, $|\eta| < 2.95$, EM fraction >0.05
2. Any track (from the `generalTracks` collection) above 0.1 GeV/c

Get CaloMET collection, L1 trigger and fill some histogram. Loop over `HcalNoiseRBXCollection` collection. Consider a certain RBX only if number of rechits >0 and the total energy >10 GeV (threshold 1.5 GeV). Consider HPD only if there is one hit above 5 GeV. Noise category:

1. Category 1 (RBX): MaxHPD/RBX ≤ 0.98, E2/E10 > 0.8, RBX hit $= 10$. (bug!?)
2. Category 2 (RBX): MaxHPD/RBX ≤ 0.98, E2/E10 < 0.33.
3. Category 3 (RBX): MaxHPD/RBX ≤ 0.98, $0.33 \leq$ E2/E10 < 0.8.
4. Category 4 (RBX): MaxHPD/RBX ≤ 0.98, E2/E10 > 0.8, RBX hit > 10.
5. Category 5 (RBX): MaxHPD/RBX ≤ 0.98, E2/E10 > 0.8, RBX hit < 10.
6. Category 6: undefined.
7. Category 7 (HPD, HPD discharge): MaxHPD/RBX > 0.98, nHitsHighest (in highest HPD) ≥ 9.
8. Category 8 (HPD, HPD ion feedback): MaxHPD/RBX > 0.98, nHitsHighest (in highest HPD) < 9.

Then it's a lot of histogram filling. There are literally a few hundred histograms. What should I put in? Let's choose a few to start with

1. Max zero, total zero, E2/E10 trio, HPD/RBX hit count
2. Hcal noise category summary plot (count of each type of noise, with respect to RBX)

2.17.4 Notes on Hcal DQM implementation

It's possible to have sub-directories. We do not need to declare the directories beforehand. The only thing to do is to set the subdirectory before booking histograms. They will be automatically created.

2.17.5 Reflection

2.17.6 Goals for next work day

1. Eat lunch
2. Eat dinner
3. Find the red herring

2.18 6331 Log (October 25, 2010)

2.18.1 Goals

1. Make a plan for Hcal pulse shape tree and start running
2. Identify things missing in the Vecbos note
3. Finalize vecbos ntuples that are done
4. Read the institutional review paper

2.18.2 Summary List

1. Finalized vecbos ntuples, and checked in lists.
2. Running a few jobs checking for duplicates in other datasets
3. Read through BPH-10-002 and made some comments.

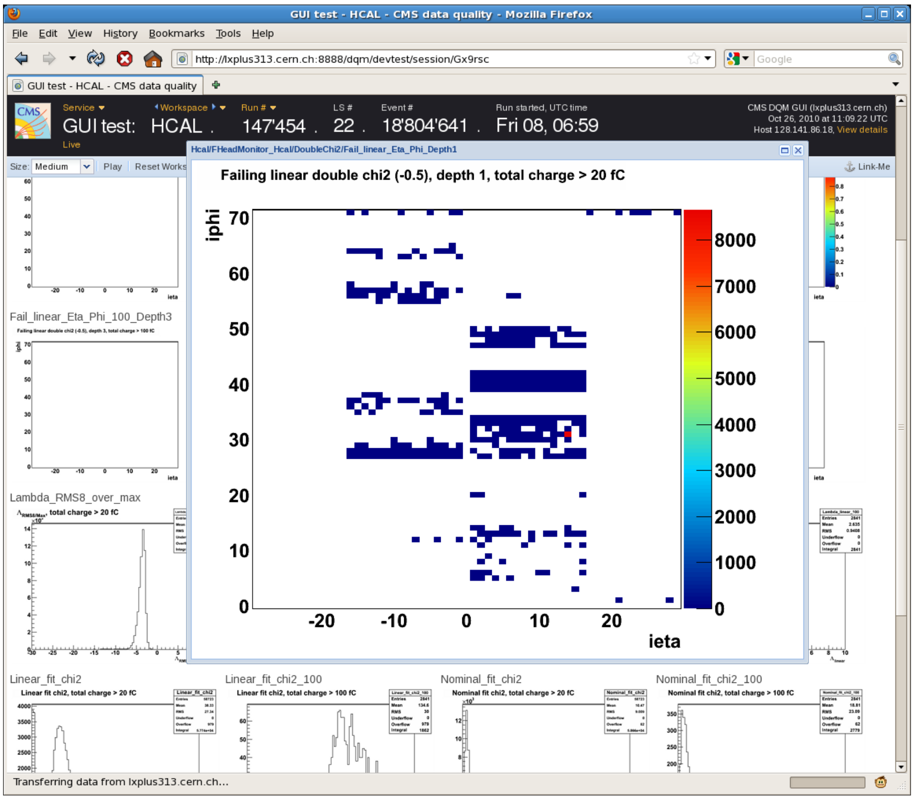


Figure 2.42: Linear double chi2 picks out the famous hot cell

2.19 6332 Log (October 26, 2010)

2.19.1 Goals

1. Make a plan for Hcal sample rerunning.
2. Finish up vecbos note - aim for Thursday, first draft
3. Make a plan for noiseline deployment

2.19.2 Summary List

1. The duplicate-checking jobs are done. Lots of duplicate events found. Some missing files spotted.
2. Added ieta-iphi plots to Hcal DQM testbench for the two discriminators. **Linear double-chi2 picks out the famous (14, 31) cell.** See plot 2.42.
3. Made a short noise DQM presentation at the group meeting.

Threshold (fC)	1	3	5	10	20	100	1-5	10-20
Number of pulses	2736251	720865	276245	94951	27272	1307	2460006	67679
Rough ratio to next	3.8x	2.6x	2.9x	3.5x	20.9x	N/A	N/A	N/A

Table 2.3: Relative number of pulses for different total charge threshold cut

2.19.3 Planning for Hcal noise studies

Things I need to check:

1. Different total charge threshold?
2. Event cleaning. What to use for the "cleaning" of noise sample.
3. Check on MC pulse shape.
4. Correlation with other rechit filters. How?
5. Other subtypes of noise?
6. Check signal distribution channel by channel.
7. Run on different types of runs.
8. Geometrical locations of each type of noise.
9. Look at observables, ex. MET distribution in signal events.
10. (Education purpose) Figure out how to estimate ratiation level.

2.19.4 Hcal noise double-chi2: total charge threshold

The relative number of cells that will be included with different total charge threshold cut is shown in table 2.3. The ntuple size for run 135175 (20fC threshold) signal is 637MB, and for noise it is 230MB. The total size is then 867MB. If we go to 10 fC threshold, it will be roughly 3GB. Let's now say that the size of dataset grows linearly with luminosity (worst case scenario). The integrated luminosity of run 135175 is 0.4/nb, so it will be 7.5GB per inverse nb, 7.5TB per inverse pb. So....O(TB) space needed. Still okay.

Interesting note: for pulses that are too small, signal hypothesis will always work better than linear or RMS/Max, in the presense of pedestal fluctuation. For pulses within 1-5 fC, this is always true. 10-20 fC, a little bit of discrimination power remains. In other words, 10fC is a good threshold for now.

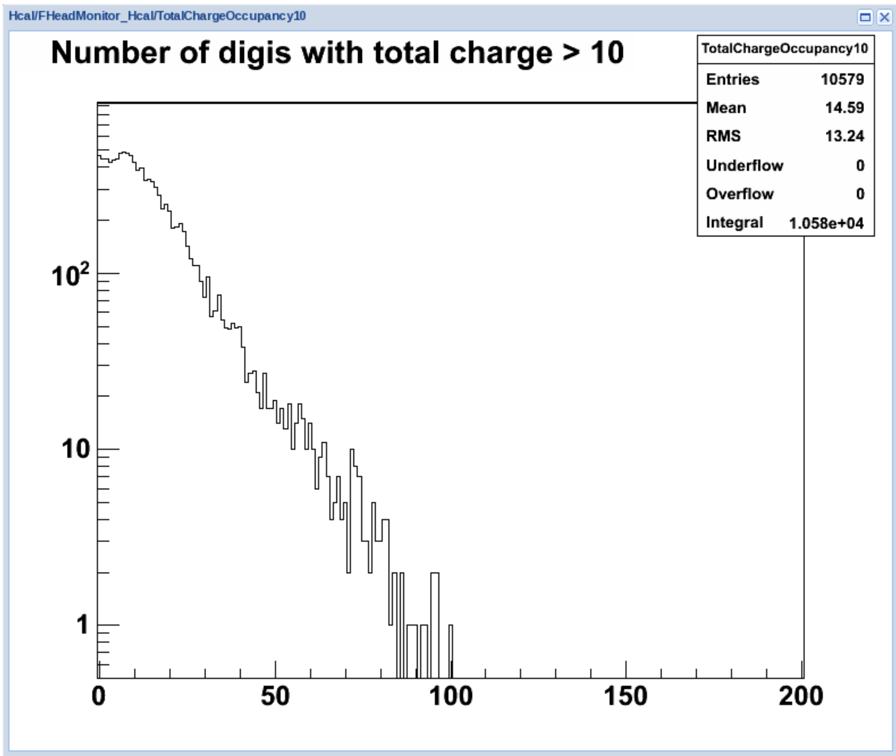


Figure 2.43: Estimation of upper limit for number of pulses in an event.

2.20 6333 Log (October 27, 2010)

2.20.1 Goals

1. By the end of the day, have a working version of hcal tree and start running
2. Copy the toy MC code to desktop and start running

2.20.2 Summary List

1. Implemented the new hcal pulse tree that has MET information and also event cleaning variables.
2. First version of basic plots from this new tree

2.20.3 Number of pulses above threshold estimation

Test run on DQM on express physics stream on a few recent runs. The result is similar across runs, and one example is shown in figure 2.43. From the slope and number of events there, assuming it's an exponential drop, by the time we reach 10^6 events, the expected number of event line will cross the 1-event level at 190-200 counts. Therefore setting the limit to 250-300 should be enough for our purpose.

2.20.4 Implementing new Hcal pulse shape tree

2.20.4.1 Specifications of the double-chi2 tree (6333 version)

The new tree will include the following:

1. Event coordinate: event number, run number, lumi section, bunch crossing, orbit (?), time
2. All pulses above 10 fC: charges, pedestals, energy, ieta, iphi, depth
3. Calorimeter energies: ET vector, sum E, sum —ET— of HB, HE, HF, EE, EB. We should be able to reconstruct MET from these later.
4. Also MET from the nominal calomet object.
5. Triggers: technical, L1, HLT
6. Event cleaning related: number of good tracks, total pt (both scalar and vector sum) of good tracks, number of good primary vertices.

The "good track" is selected by these requirements:

1. PT within 0.5-500 GeV/c
2. $\chi^2/ndof < 20$
3. $|\eta| < 2.4$
4. Valid hit at least 6
5. No requirement on compatibility to PV, since this is going to be used as event selection, and no serious use of the tracks is intended.

2.20.4.2 Time and space estimation of the tree

Tried run on 200 events from the `RAW` Jet dataset (run 148058) interactively on lxplus.

Time usage is

```
real    11m18.164s
user    11m1.374s
sys     0m4.953s
```

and the output root file has size of 745 kB. So it is about 3.3 kB per event, and it takes on average 3.39 seconds every event. The single root file I copied out from run 148058 (`Jet` dataset, `RAW`) has about 20000 events. This translates to 67800 seconds, or 18 hours. There should be plenty of time before next Thursday for them to run. 20000 events will have file size of roughly 73 MB.

Let's say we start the production on two/three runs, evaluate/debug what we need for one or two days, and rerun for real for more runs in preparation for Thursday update.

Started running on `Jet` dataset and `HcalHPDNoise` stream, run 147454, 147754 and 148058. Expect roughly a day for all to finish.

2.21 6334 Log (October 28, 2010)

2.21.1 Goals

1. R&D on double-chi2 tree.
2. With muon veto, look at remaining noise.

2.21.2 Summary List

1. Added muon counts to the tree
2. Tested the idea of adding a off-time penalty to the nominal chi2 fitting
3. Checked late valid-looking pulse.

2.21.3 Testing out modified χ^2 - adding timing and fit parameter information

2.21.3.1 Adding timing offset from the nominal fit

From the offset distribution of signal (`Jet` dataset, 100 fC threshold), the peak is at offset -109. Therefore we can add an additional term to penalize the nominal fit from going too far. The modified χ^2 is

$$\chi^2 = \left(\sum_{i=1}^{10} \left(\frac{F(i) - TS_i}{\sigma_{TS_i}} \right)^2 \right) e^{+\frac{(t- < t >)^2}{2\sigma^2}},$$

where $F(i)$ is the ideal pulse, TS_i is the pulse, and t the offset, $< t >$ average (expected) offset. When the width is set to 25 ns, no too much has changed in both signal and noise. Noise sample moves more to the negative Λ'_{linear} side than signal. Still not usable. Try to rerun with 12.5 ns. The noise sample moves even more....but it seems that they approach $\Lambda'_{linear} = 0$. Same story as before, let's try to look at them and see if we can design anything new for them.

2.21.3.2 Pulses where $\Lambda'_{linear} - \Lambda_{linear} > 1$

They all share the same characteristic. Most of them are good-looking pulses, but one TS late. Not sure what is causing this. If the tracking looked only in TS4, then it's normal that there isn't a PV around. In that case something probably pre-fired. Let's look at a few event display.

....No, usually there isn't much around in these events. What is going on?

Chapter 3

November 2010

3.1 6341 Log (November 1, 2010)

3.1.1 Goals

1. Give a presentation at the Hcal operations meeting
2. Check MC simulation pulses
3. Go to Zurich!

3.1.2 MC pulse shape behavior

The `RelValTTbar` sample finishes very fast (within a couple hours). There are 9000 events in them. Everything (above 20 fC) passed the linear discriminator if we cut at -0.5. For the RMS8/Max, if we cut at -11, there is a pulse (roughly 5 GeV) that will fail. I decided that I won't care about it now. If the threshold is lowered to 10 fC, 11 pulses fail the linear discriminator, all below 2 GeV. For RMS8/Max, still, only one pulse fails.

3.2 6344 Log (November 4, 2010)

3.2.1 Goals

1. Check out the plot macros from Matthias
2. Make a presentation at the noise WG meeting
3. Implement rate plots in noise DQM, as well as TS2TS4 variable
4. Side project kick-off

3.2.2 Summary List

1. Reported in the hcal noise WG.

3.2.3 Progress in the pass two days

I was in Zurich for the run coordination workshop. Things went slow since the internet connection isn't great.

3.2.4 50ns brainstorm session

What are we going to do for 50 ns bunch spacing?

1. Collision identification. We need to know if there are any collisions in TS 0, 2, 6, 8. If no collision, safe to proceed with simple scenario. In sub-percent level, there will be overlap. What to do then?

3.2.5 Artur's request for noise study after the meeting

1. MET distribution after each one applied. `METFwd` dataset?
2. What to do with rechit reflagger?
3. TS 3456
4. Chi2 distributions

3.2.6 Reflection

3.3 6345 Log (November 5, 2010)

3.3.1 Goals

1. Try out 4TS algorithms (?)
2. Finalize MET distributions
3. Side-project kick-off
4. Start running on `METFwd` dataset and see how things are
5. Eavesdropping in the "Charaterization of new physics at the LHC II" workshop

3.3.2 Summary List

1. Moving the pulse shape brainstorm structure to t3-susy so that I can use the batch system

3.3.3 Reflection

3.3.4 Goals for next work day

1. Eat lunch
2. Eat dinner
3. Find the red herring

3.4 6351 Log (November 8, 2010)

3.4.1 Goals

1. Finalize flagger code
2. Have fun with Hcal noise fitting
3. Side project: prepare code structure for particle gun studies

3.4.2 Summary List

1. Implemented fit flagger into CMSSW reco structure

3.4.3 Flagger implementation

Added class `HBHEPulseShapeFlagSetter` in package `RecoLocalCalo/HcalRecProducers`, files `HBHEPulseShapeFlag.cc` (and `.h`), in `CMSSW` version `3_9_1`. Pulse shape is directly from `HcalPulseShape` class, without extra input text file. Things compile. Now I should add the flag fields in the double-chi2 tree, and check performance. Now that the reconstruction is in `CMSSW_3_9_1`, reflagger for JP filter is probably not necessary, and the bit position is also different. Let's write out the full aux word for now.

ps. it should be the `flag` word

3.5 6352 Log (November 9, 2010)

3.5.1 Goals

1.

3.5.2 Summary List

1. Flag validation. Some residual remains.

2. Started running on HI data run 150431, `HIAllPhysics` stream.

3. Trying out reverse fit. Doesn't work well. Signal spread too much, and can go negative (fit better to reversed shape....)

3.5.3 Flag validation

Debugged a bit on flag setter. Some points:

1. "TS4 / Slope ¿ limit" is not the same as "TS4 ¿ limit * slope". There could be a sign problem when TS4 is negative. We would like to cut this kind of pulse out (TS4 ¡ 0) in any case, so..... maybe a protection system could help.

2. Towers (-1, 36, 1), (-5, 36, 1) and (-9, 36, 1) are inconsistent between flagger and offline fit across all noise types. The source doesn't look simple to find out, though it would be useful to double-check what is going on here. Also observed are some small energy pulses with apparent large charge in TS5. This is also something to understand.

3. All other channels are consistent.

3.5.4 First look at HI run 150431

Keep all events after lumisection 151, according to the twiki page `https://twiki.cern.ch/twiki/bin/view/CMS/HIData2010RunInformation`. See figures 3.1, 3.2, 3.3, 3.4 and 3.5 for more information on pulses failing each condition. Overall it looks like the "noise purity" is higher in this run compared to proton runs. Flat noise is roughly the same geometrically The distribution of different discriminants are shown in figure 3.6 for reference. Noise distribution is pretty different.... the late-dropping ones are worth investigating. HBHE MET distribution before and after cleaning is shown in figure 3.7.

ps. the "signal" here is partial statistics on run 147454 and 147754, after JP cleaning. The HI run is also partial statistics, no selection applied. Noise runs include 147454, 147754, 148058.

3.5.5 TS3 investigation

Time slice 3 itself cannot be used, but TS3 divided by total charge can be used to cut out a few percent of noise after round 1. It's too far from the fitting framework, it might be better to find a fit that includes this behavior.

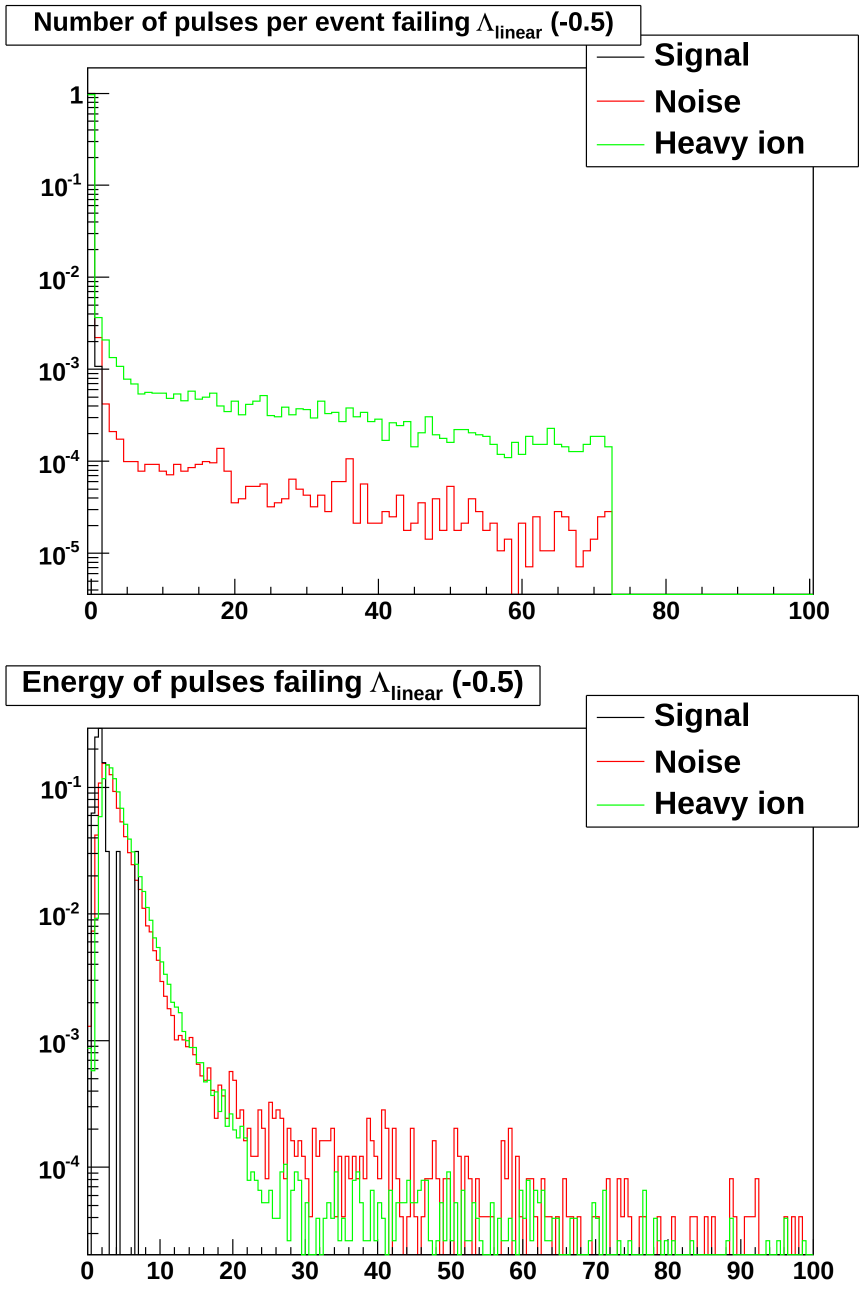


Figure 3.1: Pulses failing linear discriminant

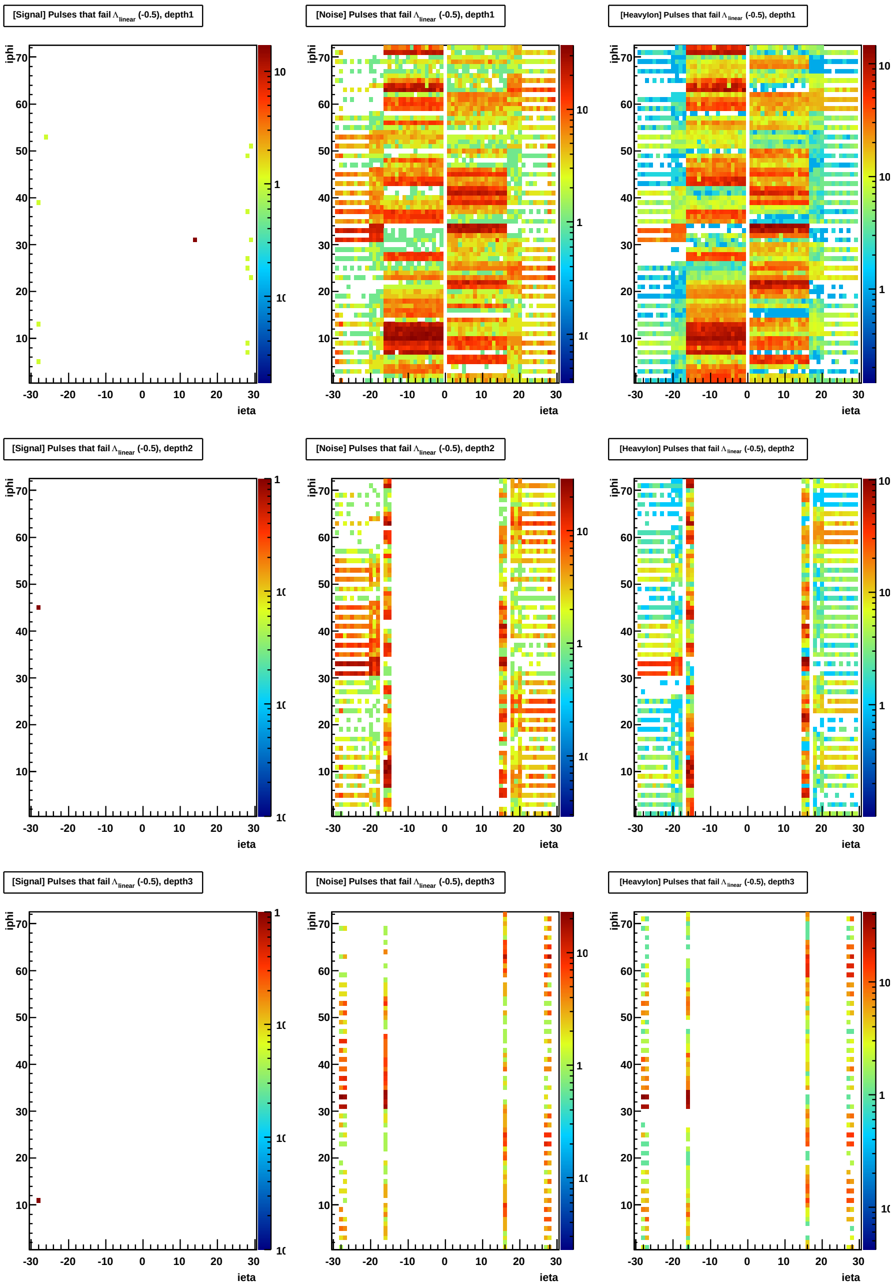
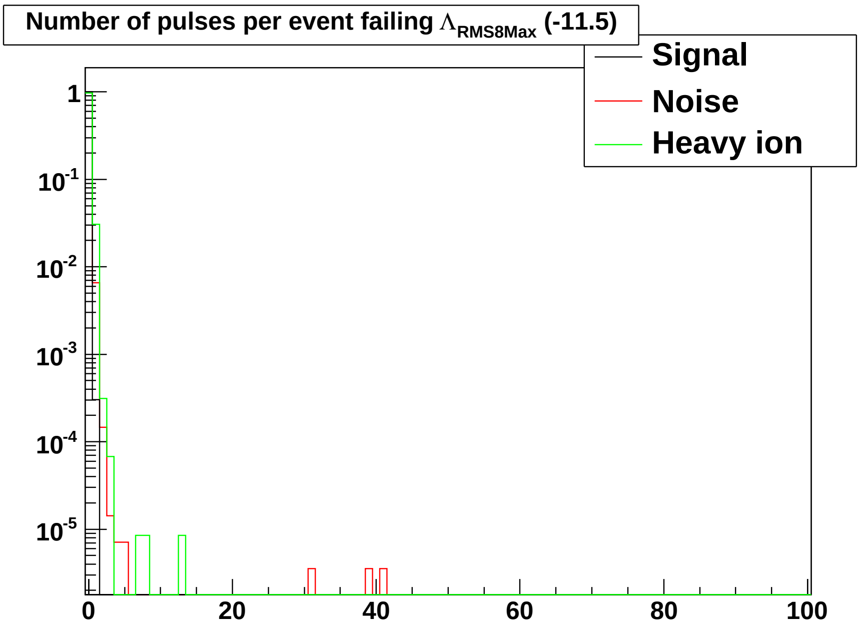


Figure 3.2: Pulses failing linear discriminant



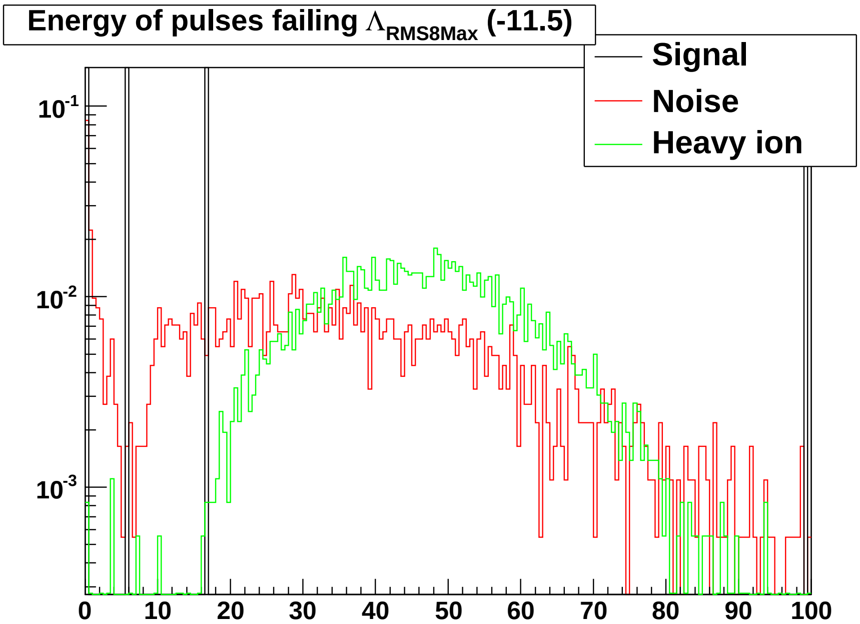


Figure 3.3: Pulses failing RMS8/Max discriminant

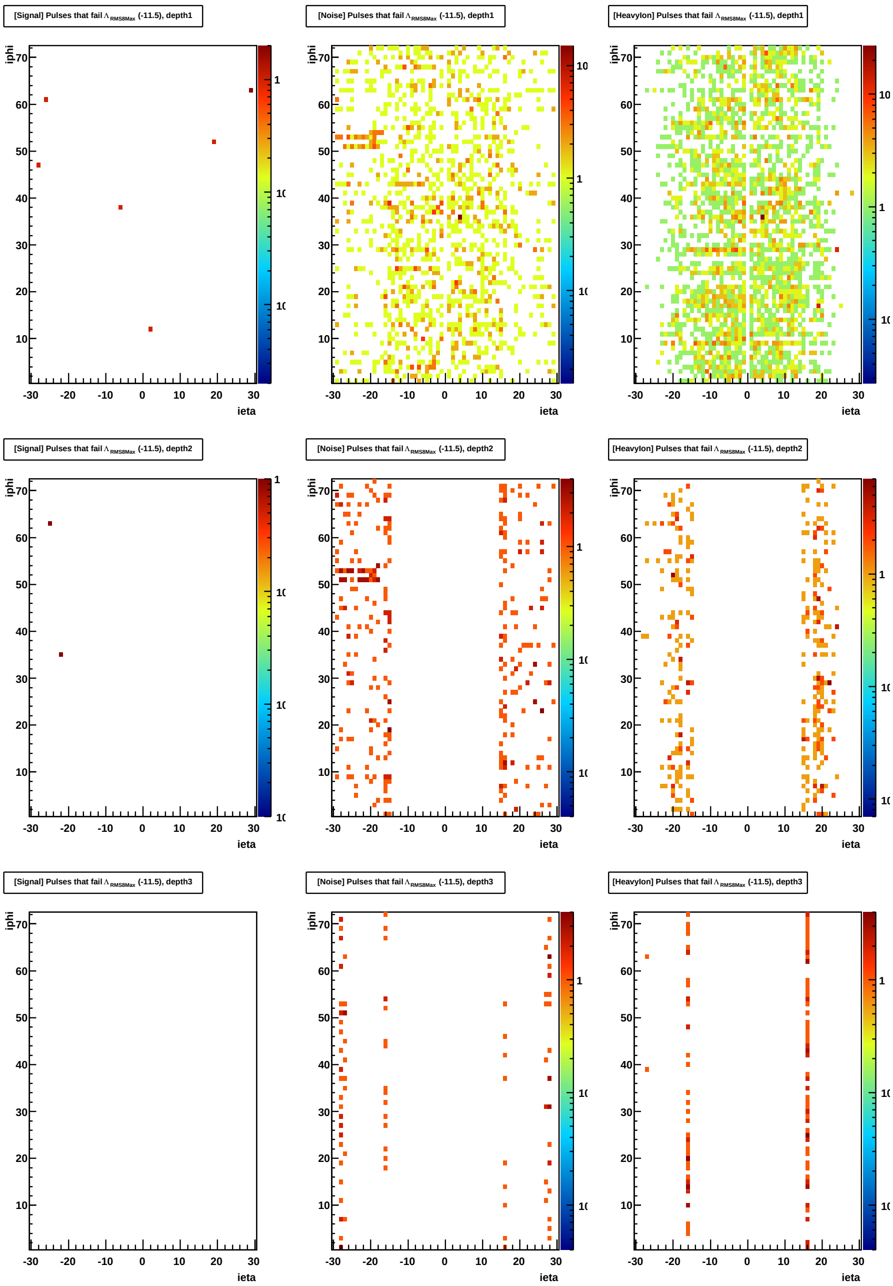


Figure 3.4: Pulses failing RMS8/Max discriminant

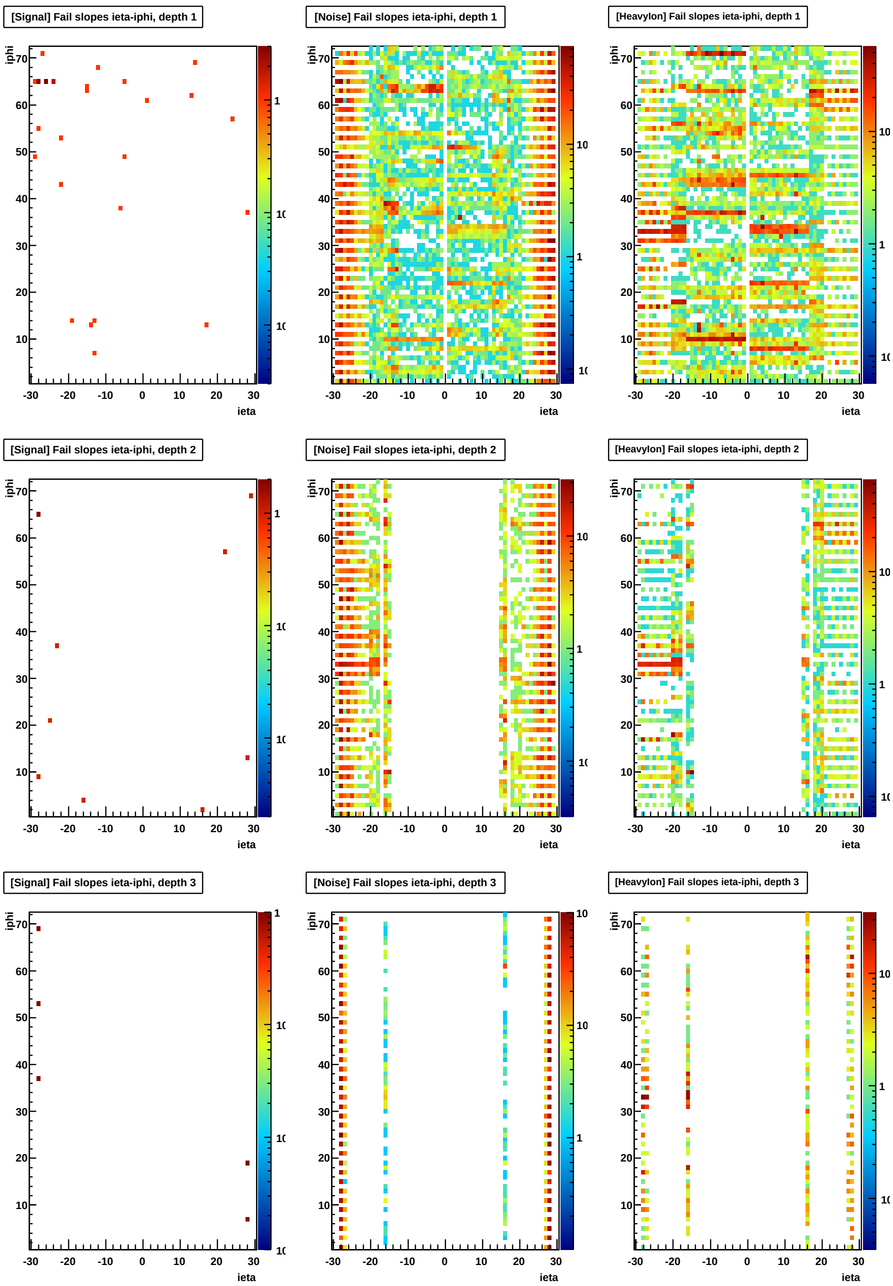


Figure 3.5: Pulses failing triangle fit slope

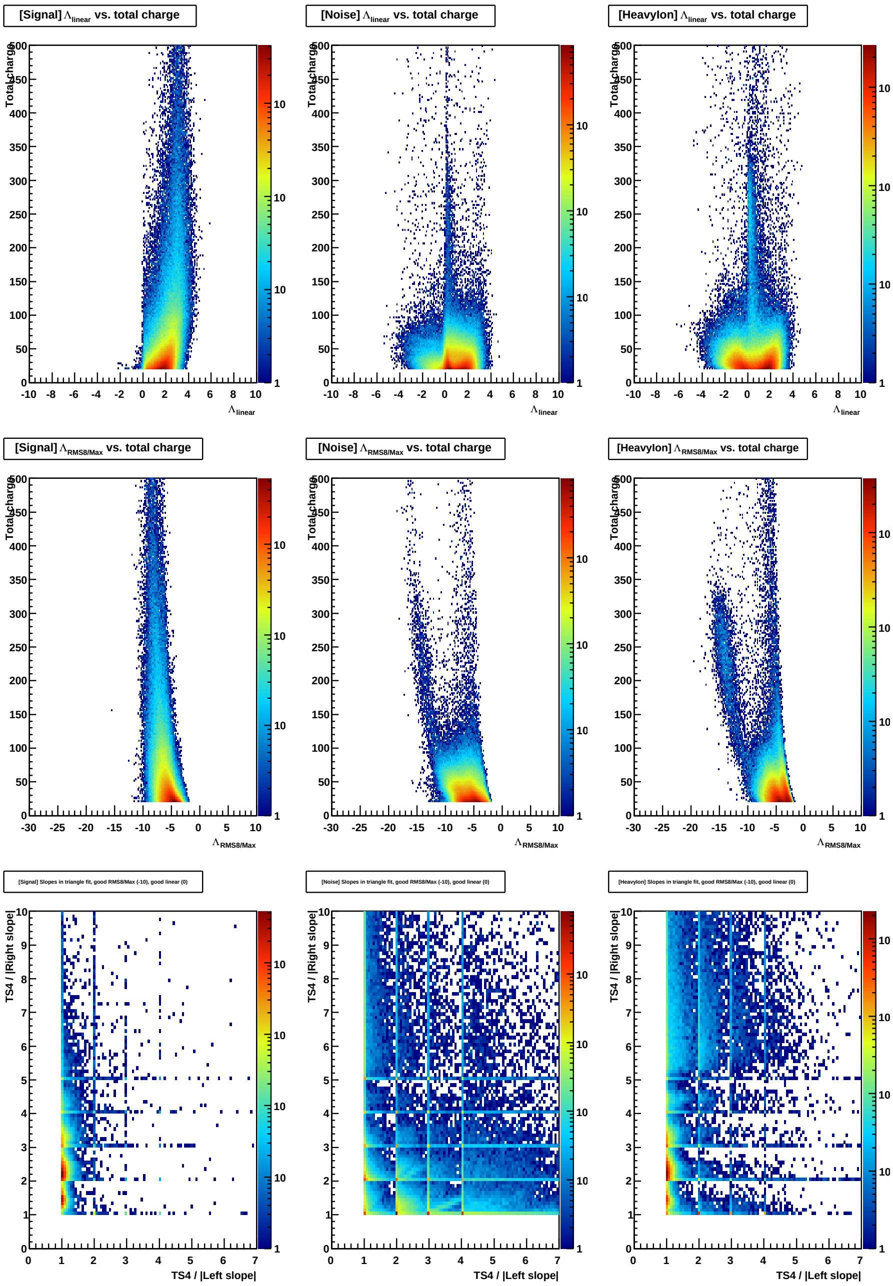


Figure 3.6: Distribution of main discriminants

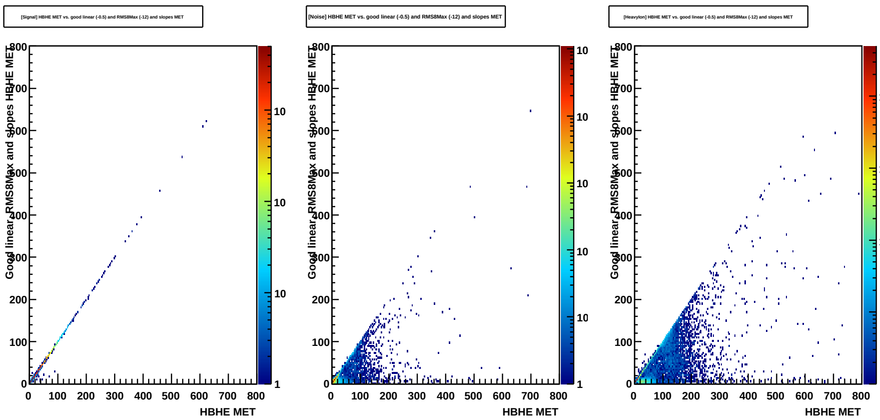


Figure 3.7: HBHE MET distributions

3.5.6 Reflection

3.5.7 Goals for next work day

1. Eat lunch
2. Eat dinner
3. Find the red herring