Method for HCAL Data Overlay

Upgrade Workshop,
November 20, 2008

E. Berry[1], C. Tully[1], R. Wilkinson[2], M. Hildreth[3]

(1) Princeton University,
(2) California Institute of Technology,
(3) University of Notre Dame

Outline

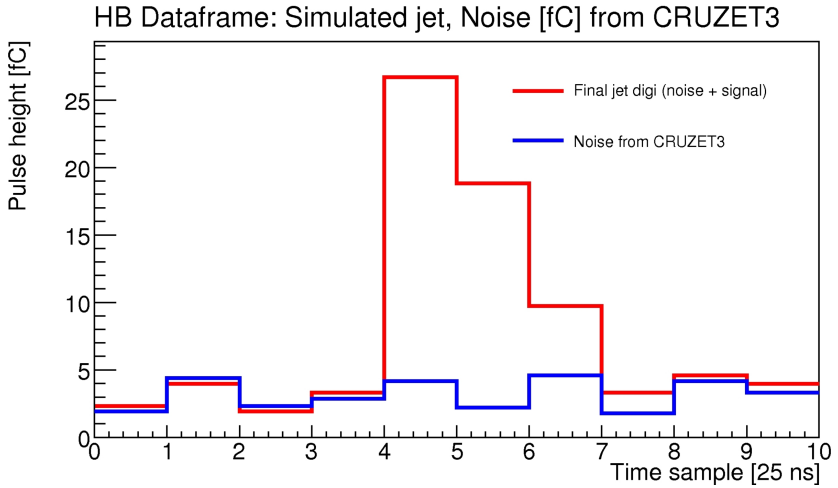
- Why is data overlay needed in HCAL?
- Two methods for noise modeling
  - Simulated digis
  - Data Mixer digis
- Results
- Possible future application for pileup simulation
- Plans for the future

All of this is discussed in CMS IN-2008/040

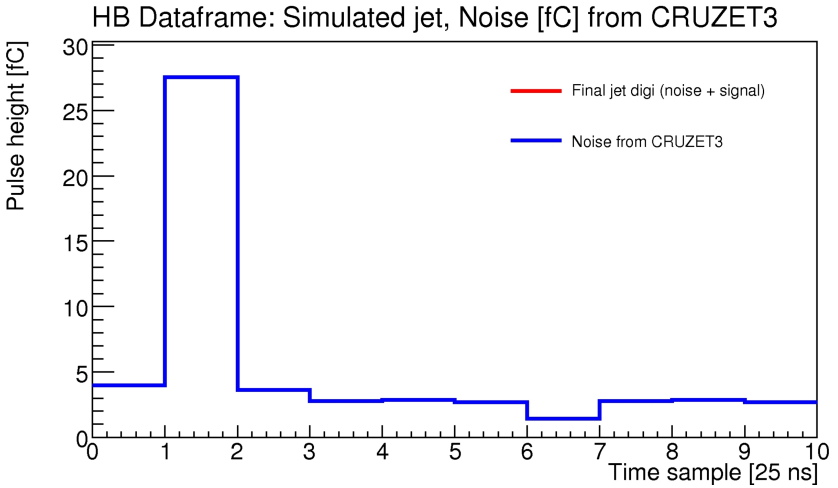
Why Data Overlay Needed in HCAL? Not perfectly understood noise

Plots below are graphical representations of HB digis
Both digis below are created with data overlay

- **Blue** corresponds to measured noise
- **Red** corresponds to overlaid simulated signal



Signal pulse due to a simulated jet
Noise is well-behaved

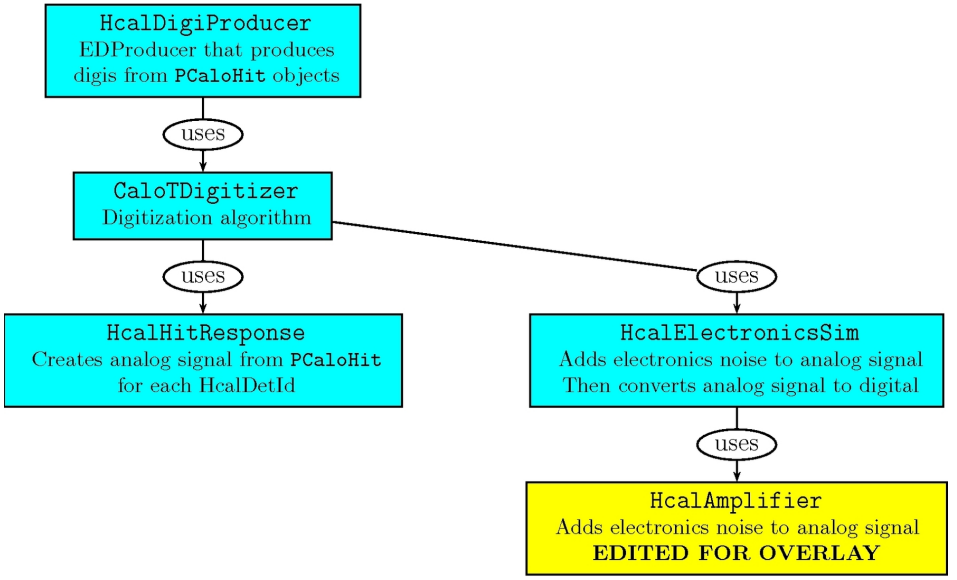


No signal simulated for this cell
Noise has **large, short-lived pulse**

Pulses of the type shown in the left-hand plot are not perfectly understood and cannot currently be modeled accurately.
Many pulses are related to HPD ion feedback and field-related HPD discharges

Normal HCAL Digi Simulation

- Create analog signal (PE's) from PcaloHit energy, according to pulse shape
- Convert PE signal to fC
  - Add cap-ID dependent pedestal & noise.
  - Uses values from the conditions database
  - HcalAmplifier makes empty simulated noise frames and adds pulses
- Convert to ADC (QIE nonlinear compression)



Highlighted step is the only step modified to make "data-mixer digits"

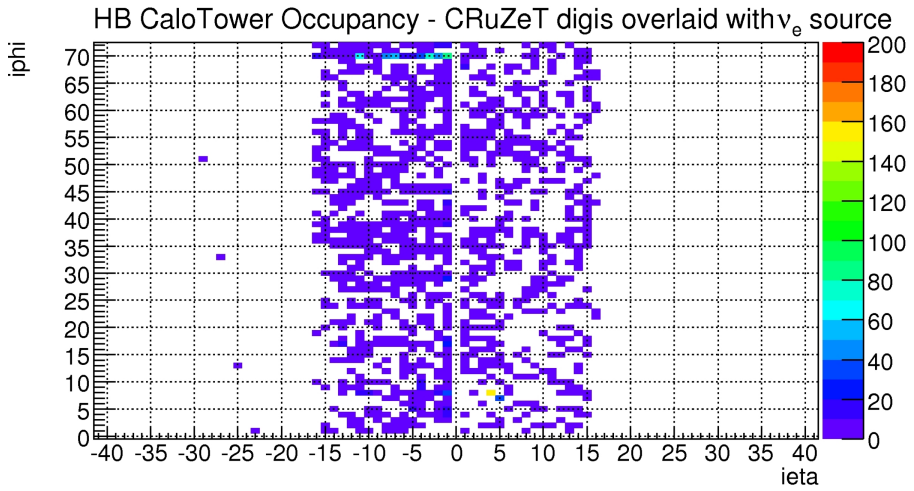
This method could also be used for ECAL, due to commonality of CMS calorimetry code

Why this method of Data Overlay?

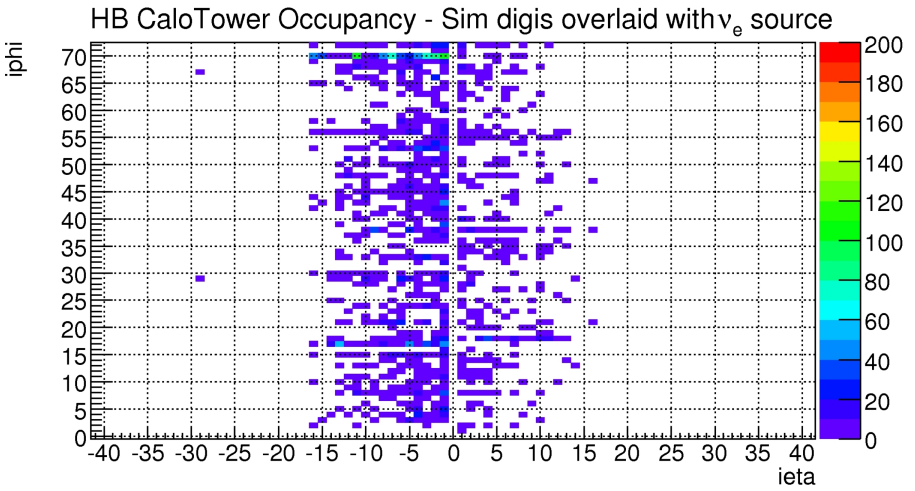
- Two advantages:
  - Fast: dataframes are taken directly from data (no “empty”/intermediate dataframes are needed)
  - Accurate: No additional quantizations of analog signal are performed
- Drawbacks of mixing dataframes as opposed to creating data-mixer dataframes
  - Slower: requires an additional step of simulation which created zero noise simulation dataframes
  - Less accurate: Additional non-linear quantization of signal-only dataframes (integer effect at 1 fC level) will not be accurate enough to do the minbias phi-uniformity calibration and other analyses that require accurate low-level signal modeling

Results: CaloTower Occupancy

Data Overlay



Normal Simulation

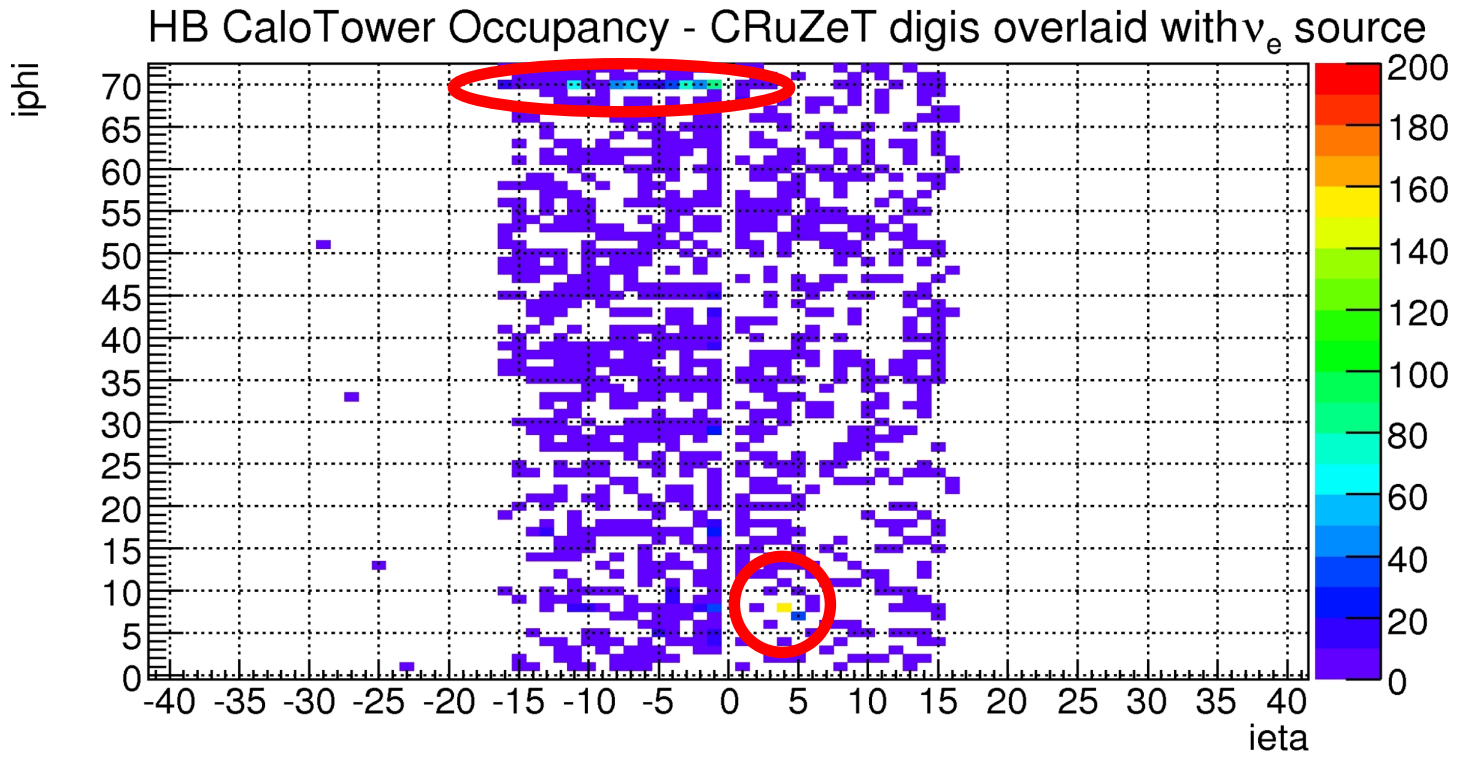


Compare data overlay vs. simulated noise + pedestal using CaloTower occupancy

This is useful, because CaloTowers are used to reconstruct physics objects, including jets.

Above plots show occupancy of CaloTowers with at least one constituent in the HB

Results: Hot Areas

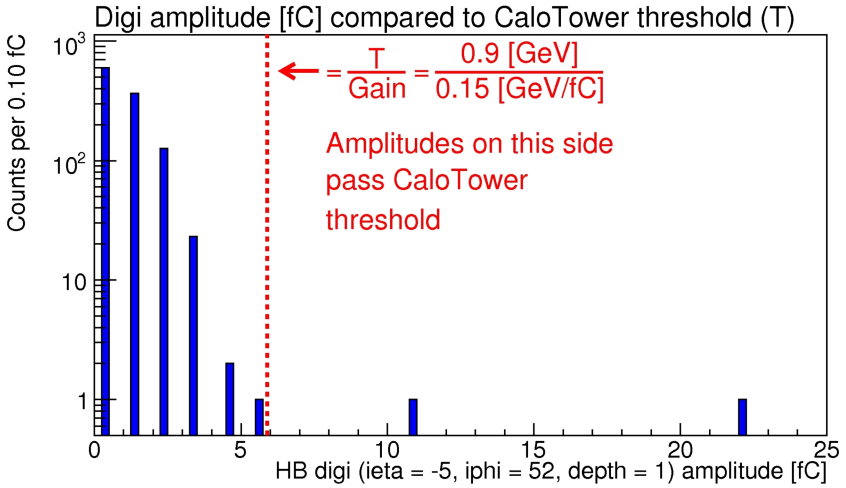


iphi strip ... seen with BOTH data overlay and simulated noise (high gain setting)

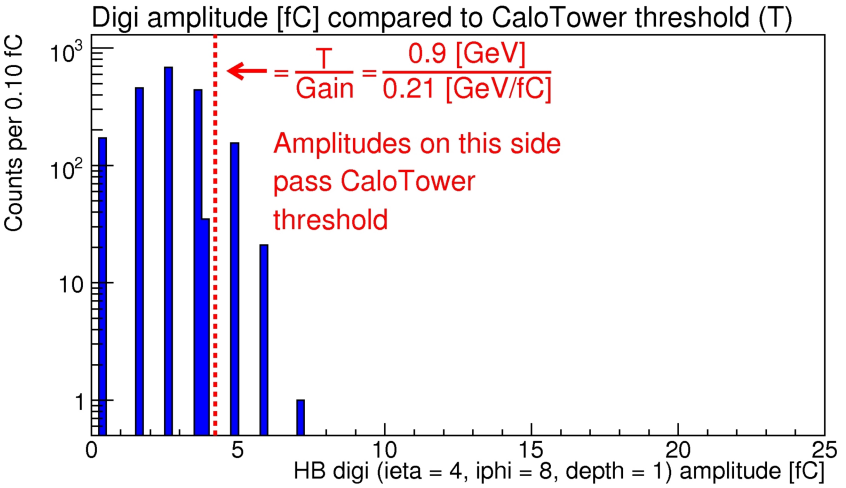
ieta = 4, iphi = 8 ... seen with data overlay ONLY (noisy cell)

Other effects: cosmics, asynchronous pulses

Results: Noisy Cells from data overlay initiate many more jets than simulated noise



Cell with two short-lived pulses

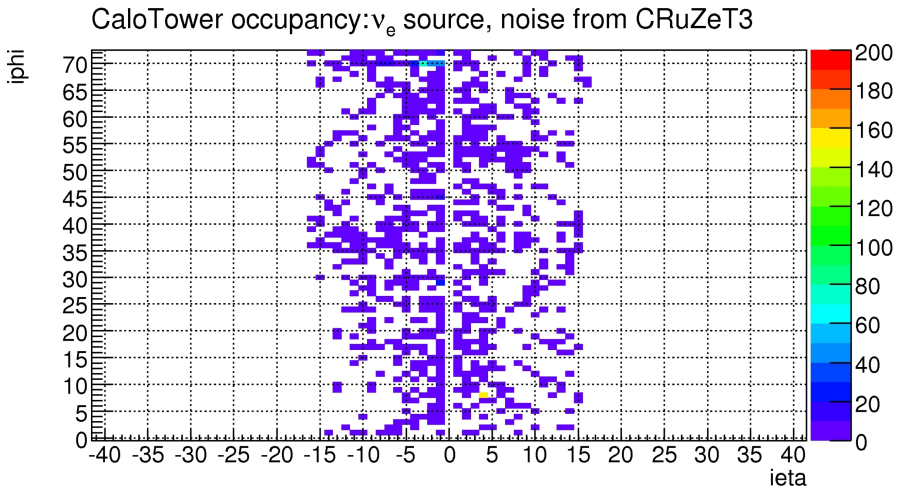


Generally noisy cell

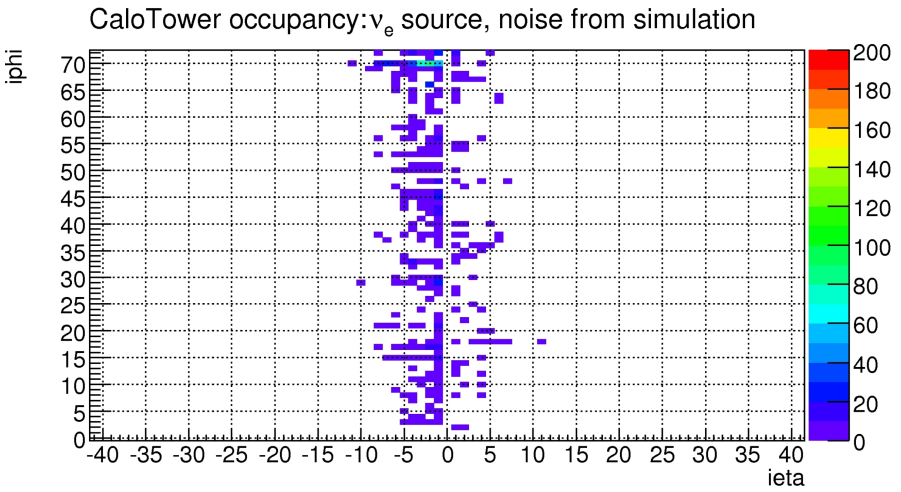
As in the last slide, both plots show amplitude histograms for digis from 2000 CRUZET events overlaid with a neutrino source.

Results: Reco Jet Occupancy

Data Overlay



Normal Simulation



Occupancy of CaloTowers that make up final reconstructed jet candidates.
The effects of the hot cell at ieta = 4, iphi = 8 and the hot iphi strip are clear.

These distributions show further evidence of the effects of readout electronics noise on reconstructed physics objects.

Application for Pileup Simulation

- Modeling pileup at the digi level is CPU and memory intensive
  - ~400 events/crossing for several successive crossings
- Could be circumvented using data overlay
  - Store a large sample of “pileup” digis
  - Add pileup digis to simulation digis just as global run noise is added for overlay
  - Pileup digis for different instantaneous lumi's could be reused for any high lumi with no additional resource usage (as compared to zero lumi)

Plans for the future

- Mike Hildreth is coordinating with Rick Wilkinson to incorporate this method into ProdAgent as a tool for general use (this meeting)
- HCAL needs to decide the what data samples are needed for data overlay
  - Zero bias taken over a long run
  - Zero bias interspersed throughout all data-taking (part of CRAFT)
  - Enriched using HCAL technical triggers (also part of CRAFT)
    - Enriched samples need special treatment for rates and timing biases
    - Needed to full live time for low rate anomalous pulse triggering
- We need to make data overlay resource light and hence part of the normal MC production from now on