Available on CMS information server **CMS IN -2010/006**

CMS Internal Note

The content of this note is intended for CMS internal use and distribution only

21 January 2010 (v3, 22 February 2010)

Anomalous HB/HE Noise at Startup: Characteristics and Rejection Algorithms

John Paul Chou, Sarah Eno, Shuichi Kunori, Seema Sharma, Jian Wang

Abstract

We describe the anomalous noise in the HCAL barrel and endcap as identified in cosmics data. We measure its rate and present a set of rejection algorithms based on timing, pulse shape, hit multiplicity, coincidence with the ECAL, and other variables that reduce the noise rate by $3-4$ orders of magnitude for noise with energy $\gtrsim 100$ GeV. We discuss the performance of all of these cuts in Monte Carlo simulation and in CRAFT '09 data, paying close attention to what assumptions may prove false at the onset of collisions. Because we believe that the overlap between 'interesting' physics events and noise events will be very low ($\lesssim 10^{-5}$), we propose an event-level rejection criteria, rather than simply identifying specific HCAL channels as noise. Although no single set of cuts will be appropriate for every use case, the cuts presented here should be applicable to the vast majority of analyses. Finally, we discuss the residual noise remaining after rejection and plans for startup.

2010/02/21
Archive Id: 1.6
Archive Date: 2009/05/05 14:08:02

Anomalous HB/HE Noise at Startup: Characteristics and Rejection Algorithms

John Paul Chou[1], Sarah Eno[2], Shuichi Kunori[4], Seema Sharma[4], and Jian Wang[3]

[1] Brown University
[2] University of Maryland
[3] Institute of High Energy Physics, Beijing
[4] Fermi National Accelerator Laboratory

Abstract

We describe the anomalous noise in the HCAL barrel and endcap as identified in cosmics data. We measure its rate and present a set of rejection algorithms based on timing, pulse shape, hit multiplicity, coincidence with the ECAL, and other variables that reduce the noise rate by $3-4$ orders of magnitude for noise with energy $\gtrsim 100$ GeV. We discuss the performance of all of these cuts in Monte Carlo simulation and in CRAFT '09 data, paying close attention to what assumptions may prove false at the onset of collisions. Because we believe that the overlap between 'interesting' physics events and noise events will be very low ($\lesssim 10^{-5}$), we propose an event-level rejection criteria, rather than simply identifying specific HCAL channels as noise. Although no single set of cuts will be appropriate for every use case, the cuts presented here should be applicable to the vast majority of analyses. Finally, we discuss the residual noise remaining after rejection and plans for startup.

1 Introduction

Anomalous HCAL noise refers to GeV−TeV scale noise observed in the detector whose origins are not due to electronics (*i.e.* pedestal) noise. The source of the anomalous noise in the HCAL barrel (HB) and endcap (HE) subdetectors is instead due to characteristics of the hybrid photodiodes (HPDs) − and the readout boxes (RBXs) that they lie within − used to convert the scintillator light into an electrical output.

In this section we present an extended summary of the issues surrounding anomalous HCAL noise. Details with respect to event reconstruction, noise versus data comparisons, and the filter performance are discussed in Sections 2, 3, and 4, respectively. Conclusions are found in Sec. 5.

1.1 Geometry and Readout

Before continuing, it's worth briefly mentioning some basics about the HCAL, particularly with respect to the HPDs and RBXsThis note is hardly an exhaustive treatment. See Refs. [1, 2] for more details. The HB/HE portion of the HCAL is a brass-scintillator sampling calorimeter. Energy depositions in the HB/HE result in scintillator light which is carried by optical fibers to HPDs that then convert the light into an electrical output. Analog-to-digital converters (ADCs) digitize the signal every 25 ns timeslice (TS). Ten consecutive TS are recorded and used for event reconstruction. The energy in a channel is determined by converting the signal in each TS into units of energy (specifically GeV) and summing the energies found in a fixed interval since the energy is not contained in a single TS. Fig. 1 shows the pulse shape over 10 TS for both testbeam and in simulation measured in fC. The conversion from fC to GeV is approximately 0.2 GeV/fC in the HB and 0.3 GeV/fC in the HE. In this note, we use a four TS-wide interval, although others have been used. Electronics (pedestal) noise is subtracted from the energy sum statistically and on a channel-by-channel basis. A channel with a reconstructed energy above a given threshold is called a 'hit.'

Each HPD contains signals from 18 different channels. In the HB, these channels correspond to 16 consecutive towers aligned with the same i_ϕ, from $i_\eta = \pm 1$ to ± 16. HB towers at $i_\eta = \pm 15$ and ± 16 have two separate depths (see Fig. 2), giving a total of 18 channels. Since the HB is segmented into 72 slices in ϕ, there are $2 \times 72 = 144$ HPDs in the HB: 72 each for the HB+ and HB−. RBXs each contain four HPDs consecutive in ϕ-slice, for a total of 36 RBXs in the HB. The HE contains the same number of HPDs/RBXs as the HB, but the geometry is more complicated due to the ϕ merging of towers and additional depth segmentation.

1.2 Noise Description

Anomalous HB/HE noise is characterized by three distinct peaks in the HPD/RBX hit multiplicity distribution. The precise origin of this noise is not well-understood. Fig. 3 shows the number of hits with energy $E > 1.5$ GeV found in RBXs from jet triggered events in CRAFT '09 run 112237. Although the rates can change with the hit reconstruction algorithm and running conditions (Sec. 2), the general three-peak structure is stable and is found throughout the HB/HE. The rates as a function of the RBX energy is also shown in Fig. 3, where the RBX energy is the sum of the energy of hits with $E > 1.5$ GeV. We categorize the peaks into three categories:

- *HPD Ion Feedback* (1−9 hits/RBX) − Anomalous noise probably related to thermal emission of electrons which ionize a molecule in the acceleration gap of the HPD. The ion liberates other electrons and causes a cascade effect resulting in a significant energy deposit in a few channels in the HPD. This noise is characterized a low hit

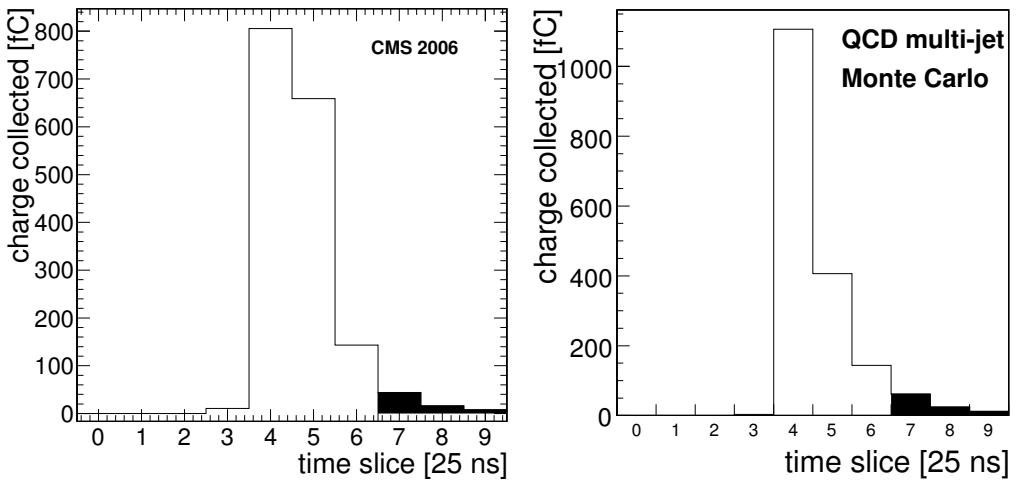

Figure 1: Left: Pulse shape over 10 timeslices from 2006 testbeam (100 GeV π^-) signal. Right: Expected pulse shape over 10 timeslices for a signal in QCD MC Simulation.

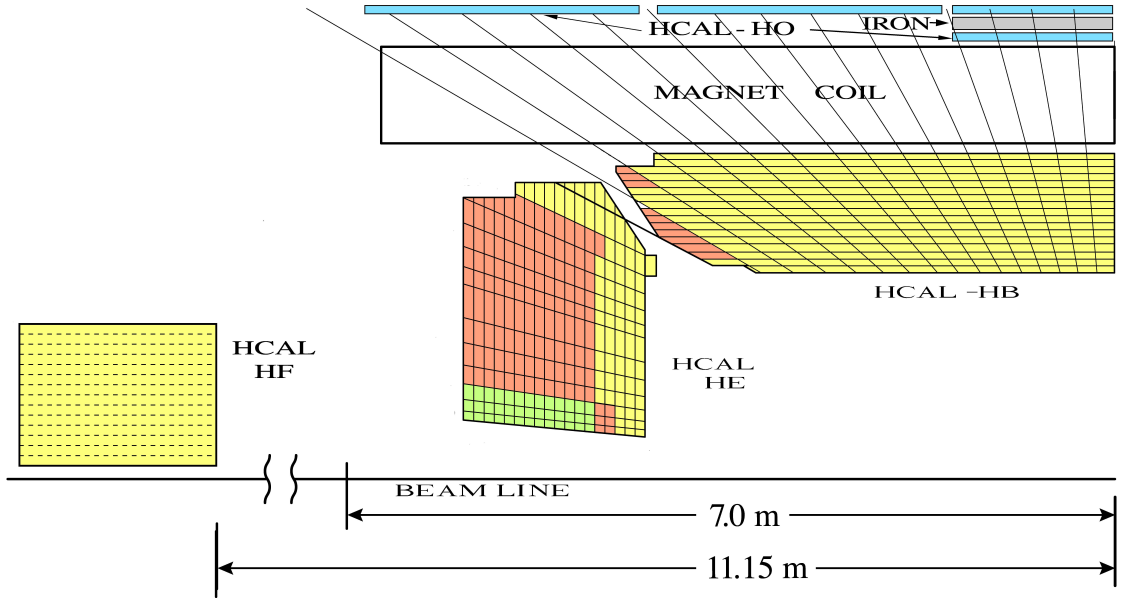

Figure 2: Illustration of the negative z-side of the HCAL. The rightmost (leftmost) tower z in the HB has an $i_\eta = -1$ (-16). The most forward tower in the HE has an $i_\eta = -29$. Note also the depth segmentation in the two most forward towers in the HB.

multiplicity in the HPD and dominates the low energy component of the noise.

- *HPD Discharge* (10−18 hits/RBX) − Noise probably related to an electrical discharge emanating from the wall of the HPD as the result of misalignment with the magnetic field. In this case, all or nearly all of the 18 channels in the HPD have hits above threshold. Note that for typical HPD discharge events, the other HPDs in the RBX do not have any hits. HPD discharge noise dominates the high energy spectrum.
- *RBX Noise* (19−72 hits/RBX) − Anomalous noise characterized by a high multiplicity of hits in all four HPDs in an RBX. Although RBX noise has the highest hit multiplicity, counter-intuitively it is not the highest energy component of anomalous noise. This also suggests a distinct production mechanism for this type of noise, although that is still speculative.

In total, the rate of RBXs with $E_{\mathrm{RBX}} > 100$ GeV is ~ 4 Hz. The occurrence of events with more than one RBX above a 20 GeV threshold is one part in $\sim 10^4$, so the RBX rate is nearly equivalent to the overall event rate, as well.

1.3 Estimated Trigger Rates

Here we estimate the noise induced trigger rate in different jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ paths. We select those events which have an energetic RBX ($E_{\mathrm{RBX}} > 5$ GeV) which we presume is due to anomalous noise. By reconstructing jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ in these samples, we can estimate how many of the events would pass the respective triggers at L1 and the HLT. Knowing that one luminosity section is approximately 93 seconds long, we can use the total number of luminosity sections to normalize the rate. Note that this only holds as long as all events in a given luminosity section is considered.

The results are summarized in the tables below. There we see that while the rate at L1 is managable, but the noise becomes a non-trivial part of the bandwidth at the HLT. Particularly at high jet E_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$, the noise rate dominates over the expected physics production rate for both the $8E29$ and $1E31$ scenarios. If we further require that beam is present in the relevant crossings, then the noise rate will be reduced by the fraction of LHC buckets that are filled. For instance, if only 10% of buckets are filled and we require that beam is present for a trigger, then the $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV trigger rate from noise will be reduced to 0.2 Hz.

8E29 Menu ($\sqrt{s}$=10 TeV)

Trigger	Total Prescale	Noise Rate (Hz)	Physics Rate (Hz)
Jet 6	×500	~ 0.15	6.21 ± 0.19
Jet 15	×25	~ 0.51	8.11 ± 0.22
Jet 30	×1	~ 7.7	17.71 ± 0.33
Jet 50	×1	~ 5.0	3.01 ± 0.13
MET 20	×2	~ 5.0	7.04 ± 0.21
MET 45	×1	~ 6.4	~ 0.2
MET 100	×1	~ 2.2	~ 0.0

Table 1: Expected rate for HCAL noise and for collision physics using the anticipated 8E29 L1+HLT menu. No assumption about a beam-presence requirement at the trigger level is given.

.

1.4 Overlap Rate

As we have seen in Sec. 1.3, the rate of noise in events collected through jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ triggers is very high, and dominates at high E_{T}. However, events collected through other triggers that do not

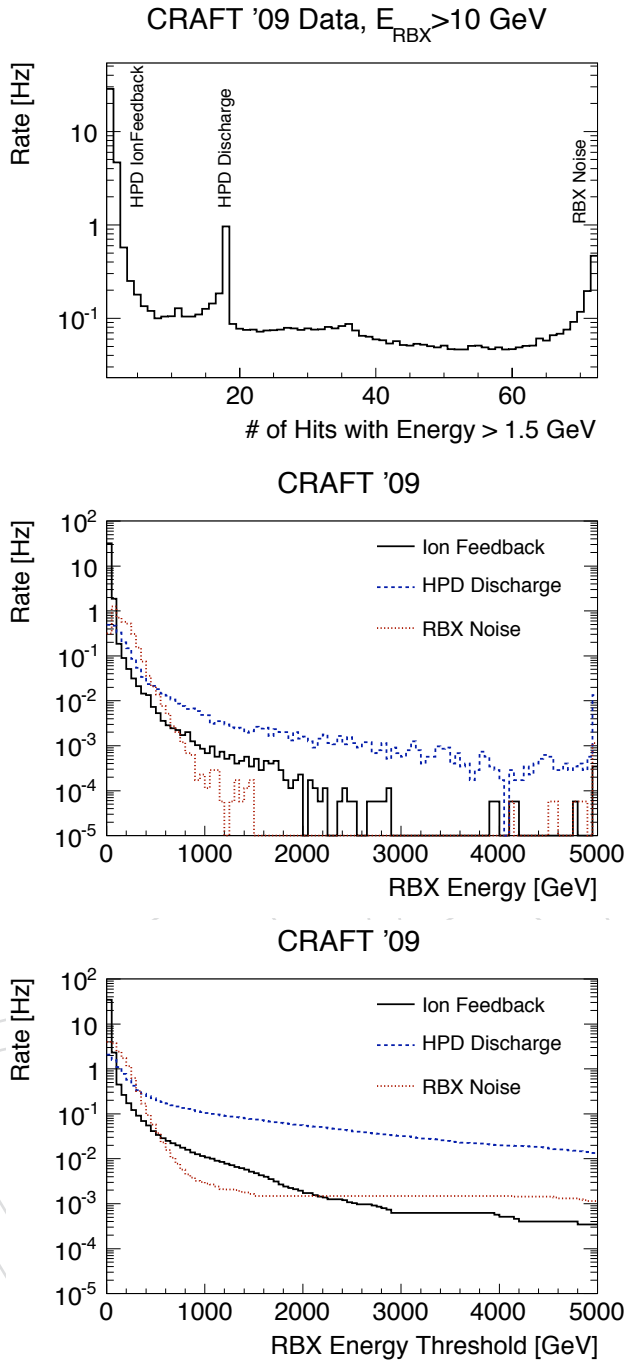

Figure 3: Top: The rate (in Hz) found in CRAFT '09 run 112237 of noise types characterized by the hit multiplicity in an RBX with $E_{\mathrm{RBX}} > 10$ GeV. Ion feedback/HPD discharge/RBX noise is categorized by $1-9/10-18/19-72$ hits in an RBX. Center: The rate in the same run as a function of the RBX energy. Bottom: The rate in the same run as a function of the RBX energy *threshold*. The total rate for noise with RBX energy $E_{\mathrm{RBX}} > 100$ GeV is ~ 4 Hz. Note that the rates may change significantly with HCAL detector running conditions such as the gains as well as the hit reconstruction algorithm.

1E31 Menu ($\sqrt{s}$ =10 TeV)			
Trigger	Prescale	Noise Rate (Hz)	Physics Rate (Hz)
Jet 15	$\times$10000	$\sim 1.3 \times 10^{-3}$	1.34 ± 0.33
Jet 30	$\times$5000	$\sim 1.5 \times 10^{-3}$	1.17 ± 0.31
Jet 50	$\times$200	$\sim 25 \times 10^{-3}$	1.84 ± 0.39
Jet 80	$\times$20	~ 0.12	1.42 ± 0.34
Jet 110	$\times$1	~ 1.3	7.11 ± 0.77
Jet 180	$\times$1	~ 0.60	1.17 ± 0.31
MET 20	$\times$2000	$\sim 5.0 \times 10^{-3}$	0.17 ± 0.12
MET 35	$\times$20	~ 0.37	0.33 ± 0.17
MET 60	$\times$1	~ 5.1	~ 0.5
MET 100	$\times$1	~ 2.2	~ 0.0

Table 2: Expected rate for HCAL noise and for collision physics using the anticipated 1E31 L1+HLT menu. No assumption about a beam-presence requirement at the trigger level is given.

.

directly use the HCAL are much less affected by this problem. As long as events are not selected based on HCAL noise, the overlap rate should remain low. Since the LHC has bunch crossings at a 40 MHz rate, we would naively guess that the total overlap rate between noise and physics is the noise rate divided by the crossing rate. The rate for events with $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV is ~ 10 Hz, so the expected overlap rate for such noise would naively be $\sim 3 \times 10^{-7}$.

This estimate assumes, however, that the noise has a negligible time structure, a fact which is not true. We use the L1 trigger to identify noise events and measure the noise rate in the CRAFT data. In Sec. 3, we see that the noise can persist for many TS. However, the trigger rules for L1 is to only accept at most one L1 trigger in 3 bunch crossings or two L1 triggers in 25 bunch crossings. Therefore, if noise spans 20 TS on average, but only causes 2 L1 accepts on average, then we will have underestimated the total overlap rate with this method by a factor of 10. Because we do not have a reliable estimate for the total length of time the average anomalous event lasts, we cannot precisely estimate the data-noise overlap rate. We do know from the data we have that noise lasts at least $4-5$ TS on average, so an additional factor of *at least* $2-3$ is necessary, and a factor of 10 is plausible.

To further complicate this issue, the noise rate is not stable with time (see Sec. 2), and it is possible noise may be stimulated (or suppressed) by collisions. Based on this we conservatively estimate that the total noise-data overlap rate for anomalous HB/HE noise with $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV is no more than 10^{-5}, noting the relevant caveats.

Because of this, we propose *event-based* rejection algorithms to remove the anomalous HB/HE noise. Although the overlap rate is small, the misidentification rate of data as noise is substantial $\sim 10^{-2} - 10^{-3}$. An event from a non-jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ trigger that is identified is noise is more likely to be real data mimicking noise than real noise itself. Masking of the towers could then result in a significant bias because a large dead zone may be introduced resulting in a substantial fake $E_{\mathrm{T}}^{\mathrm{miss}}$ signature. If only a small fraction of physics events have a high energy RBX mistakenly identified as noise, this will introduce a large $E_{\mathrm{T}}^{\mathrm{miss}}$ tail, substantially altering the event topology. Since the total number of real physics events removed is still small as a fraction of the total, we believe that event-based rejection is the safest and most effective way to eliminate this type of anomalous noise.

1.5 Rejection Algorithms

Noisy events are identified through a variety of algorithms, either based solely on the HCAL or by using coincidence with the ECAL. The variables used are defined as follows:

- *pulse shape ratio* − We measure the total amount of pedestal subtracted charge as a function of TS in all hits in an RBX (see Fig. 1). We use the ratio of the total charge in the two peak TS to the total charge in all ten as a means of quantifying the goodness of the shape. The charge in all hits are summed over. In RBXs with $E_{\mathrm{RBX}} > 50$ GeV, signals from collision physics typically contain between 70% and 90% of the total charge in the two peak TS.
- *timing* − If the reconstructed time of a hit is out of time with the rest of the event, then this suggests anomalous HCAL noise. Even if the noise itself fired the trigger, the reconstructed time and the trigger time can still be offset from each other. We use the timing in hits with $E_{\mathrm{hit}} > 25$ GeV to identify noise, since the timing resolution becomes poor at low energy.
- *ADC 0 counts* − We count the number of ADCs which read precisely 0 in any channel and in any of the 10 TS in an RBX with $E_{\mathrm{RBX}} > 10$ GeV. A high multiplicity of such '0 counts' is also indicative of noise. The probability of real data mimicking a high number of 0 counts is extremely low (see Sec. 3.3).
- *hit multiplicity* − We count the number of hits in a single HPD or in the entire RBX with $E_{\mathrm{hit}} > 1.5$ GeV. A high multiplicity of such hits indicates noise.
- *electromagnetic fraction* − We look for coincidence with the ECAL to signify the presence of collision physics. We sum the towers associated with an RBX with $E_{RBX} > 20$ GeV, and calculate the total electromagnetic and hadronic component.

Based on the studies discussed in the rest of the note, we provide the following set of suggested cuts to identify and reject anomalous HCAL noise, summarized in Table 3. These tables correspond to a set of loose filter cuts based on HCAL information only, a tighter set of filter cuts again based on HCAL information, and a "high level" set of cuts based on coincidence with the ECAL. The loose and tight filter cuts are similar to algorithms II and III, found in Ref. [3].

To quantify the efficacy of these algorithms, we define the following terms:

- *efficiency* − the fraction of events with noise that are rejected by a given algorithm;
- *inefficiency* − the fraction of events with noise that are *not* rejected by a given algorithm, that is, one minus the efficiency;
- *over-efficiency* − the fraction of non-noise (*i.e.* physics) events that are rejected by a given algorithm.

The filters described above are approximately $\sim 10^{-3}$ to $\sim 10^{-4}$ inefficient at rejecting noise in CRAFT '09 run 112237, for events with an $E_{\mathrm{RBX}} > 100$ GeV. On the other hand, they are $\sim 10^{-2}$ to $\sim 10^{-3}$ over-efficienct at rejecting QCD MC events. Every noisy event in run 112237 with an $E_{\mathrm{RBX}} > 340$ GeV can be rejected with the combined loose+high-level filter.

Of course, further noise suppression and rejection is still possible, although slightly more sophisticated methods should be used. For instance, using coincidence with tracking is another way to reject noise. However, using a jet-based, rather than a hardware-based, approach may be more fruitful here. A jet-based approach would be more sensitive to low energy anomalous noise for which the algorithms described in this note are unable to address and would be less susceptible to instances where a high E_{T} jet fluctuates to largely neutral particles, for instance.

'Loose' Noise ID cuts
pulse shape ratio *outside of* [0.70, 0.90] for $E_{\mathrm{RBX}} > 50$ GeV
hit time *outside of* [-8.0, 7.0] ns for $E_{\mathrm{hit}} > 25$ GeV
of ADC 0 counts ≥ 9 for $E_{\mathrm{RBX}} > 10$ GeV
of hits ($E_{\mathrm{hit}} > 1.5$ GeV) in one HPD ≥ 17
of hits ($E_{\mathrm{hit}} > 1.5$ GeV) in one HPD ≥ 10 when no other hits are present in the RBX
'Tight' Noise ID cuts
pulse shape ratio *outside of* [0.73, 0.88] for $E_{\mathrm{RBX}} > 50$ GeV
hit time *outside of* [-6.0, 5.0] ns for $E_{\mathrm{hit}} > 25$ GeV
of ADC 0 counts ≥ 6 for $E_{\mathrm{RBX}} > 10$ GeV
of hits ($E_{\mathrm{hit}} > 1.5$ GeV) in one HPD ≥ 13
of hits ($E_{\mathrm{hit}} > 1.5$ GeV) in one HPD ≥ 8 when no other hits are present in the RBX
of hits ($E_{\mathrm{hit}} > 1.5$ GeV) in one RBX ≥ 40
'High Level' Noise ID cuts
Electromagnetic fraction < 0.01 for $E_{\mathrm{RBX}} > 20$ GeV

Table 3: The above cuts define the three noise filters used in this note. If any condition is found to be true for a given filter, then the event should be rejected.

A jet-based approach would also aid the EMF selection, as muons which minimum ionize in the ECAL but bremsstrahlung in the HCAL could fake the EMF signature.

The cuts we suggest here have been tuned on CRAFT '09 data and QCD MC simulation. As our understanding of both noise and physics increases, we will periodically update the recommended cuts to reflect our best knowledge. For the most up to date information, please see the webpage at [4] for more details on what the precise cuts are and how the access them. For an earlier discussion of the noise, see Ref. [3].

2 Event Selection and Reconstruction

In this section, we describe the event selection and reconstruction algorithms used for the rest of this note.

We use ∼5,000,000 events collected from the "CALO" (jet/e-gamma trigger paths) dataset of CRAFT '09 run 112237. Approximately 700,000 contain an RBX with $E_{\mathrm{RBX}} > 10$ GeV. Once an RBX with at least 10 GeV is selected, these events are triggered nearly exclusively on the `L1_JET10U` trigger, where the HLT is in "passthrough" mode, that is accepting all L1 triggers. We use this run as a representative sample of the anomalous HCAL noise with the caveat that while the general characteristics of the noise are stable with time, the rates are not. Figure 4 shows the rate for the three characteristic noise types as a function of the run number and RBX energy threshold. These runs span approximately three months of real time over which the RBX noise rate fluctuates within a factor of 2. The reason for this is unknown. No other requirements on the events from this dataset are applied.

We use three different QCD datasets as a standard for comparison against the HCAL noise[1]. These samples are PYTHIA $2 \rightarrow 2$ QCD samples generated at $\sqrt{s} = 10$ TeV, with inclusive $\hat{p}_{\mathrm{T}} > 15$ GeV/c, $\hat{p}_{\mathrm{T}} > 300$ GeV/c, and $\hat{p}_{\mathrm{T}} > 3000$ GeV/c. The breadth of energy scale is useful

[1] We strongly encourage those interested to test this analysis on their favorite physics sample. We are aware that not all of the techniques employed here will work on all physics samples universally, so independent studies are necessary, but doing so here is outside the scope of this note.

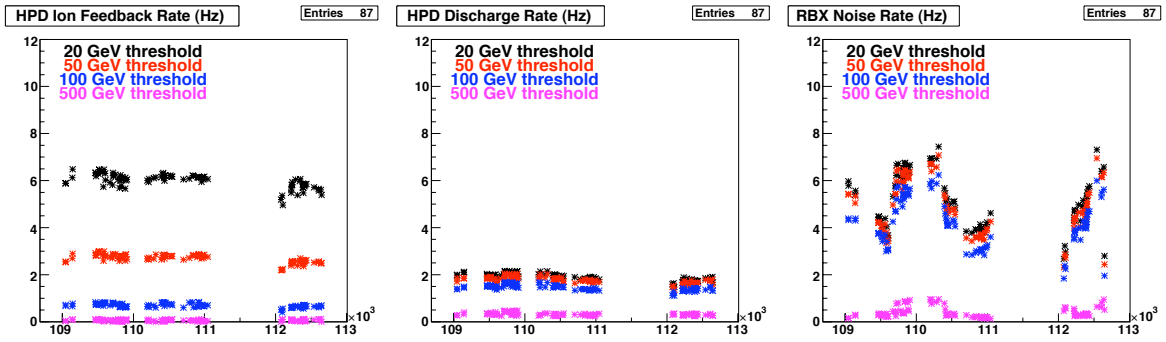

Figure 4: Rates for (left) Ionfeedback, (center) HPD discharge, and (right) RBX noise as a function of run number for different energy thresholds. The energy is calculated for the entire RBX. The runs span three months in real-time.

to compare potential biases in energy and detector occupancy. Approximately 200,000 events are taken from each sample. We use `CMSSW_3_3_4` to do the event reconstruction. No pileup, either in-time or out-of-time, is simulated in these events.

We reconstruct the energy in HB/HE channels by integrating the total pedestal subtracted charge from the ADCs over a four TS interval, beginning with the 4^{th} TS and ending with the 7^{th} TS (see Fig. 1 for the expected time structure). The total charge is converted into an energy, and if the total energy is above a 0.3−0.8 GeV threshold, then a hit is recorded. Typically in cosmics data, an 8 TS interval is used (1−9 inclusive), but in this note we use the same 4 TS interval consistently between data and MC. This assumes that the timing in data is sufficiently tuned so that the effects of an offset are minimal. Because the peak in a signal is sharply rising, a small offset could result in a substantial fraction of the total energy being left out of the charge integration. Unlike the hit energy, the timing is determined over a three TS interval with the most energetic pulse in the middle.

Channels are indexed according to their location in η, ϕ, and depth. Towers are reconstructed by summing the signal from all channels with the same i_η and i_ϕ index over all depths. In this note, the threshold for an HB (HE) hit to be included in a calotower is 0.7 (0.8) GeV energy. Different thresholds may be used in jet/$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction as well as at the HLT.

Anomalous noise can be particularly sensitive to differences in reconstruction algorithms and thresholds since the noise may differ substantially from the expected signal from collisions. For instance, the time structure for RBX noise (Sec. 3.1) is often more broad than 4 TS, so the standard reconstruction interval will not capture the full energy. Also the noise is characterized in energy scale, but physics events are typically reconstructed with E_{T} in mind. This has the positive effect that at higher $|\eta|$, noise will have a smaller overall impact.

Information regarding HPDs and RBXs are reconstructed by a geometric mapping of hits and towers to the corresponding HPD or RBX. The mapping in the HB is simple, as the HPD ϕ index is determined solely by the ϕ index of the hit or tower. The HPD and the hit index are offset, however, so that $(1) \leftrightarrow 2$, $(2) \leftrightarrow 3$, ..., $(70) \leftrightarrow 71$, $(71) \leftrightarrow 1$, $(72) \leftrightarrow 2$ is the mapping of the hit i_ϕ to/from the HPD ϕ index, regardless of the i_η or depth of the hit. This is done so that HPDs $0-3$ (and correspondingly $4-7$ and so on) are all found in the same RBX.

The mapping of the indices in the HE is more complicated. For hits with $|i_\eta| \leq 20$, the mapping is the same as the HB, but for hits with $|i_\eta| \geq 21$, the mapping alternates i_ϕ depending on the i_η and depth:

$(1,21) \leftrightarrow 1$	$(1,22) \leftrightarrow 2$	$(1,23) \leftrightarrow 1$	$(1,24) \leftrightarrow 2$	$(1,25) \leftrightarrow 1$	
$(1,26) \leftrightarrow 2$	$(1,27) \leftrightarrow 1$	$(1,28) \leftrightarrow 2$	$(1,29,1) \leftrightarrow 2$	$(1,29,2) \leftrightarrow 1$	$(1,29,3) \leftrightarrow 1$
$(3,21) \leftrightarrow 4$	$(3,22) \leftrightarrow 3$	$(3,23) \leftrightarrow 4$	$(3,24) \leftrightarrow 3$	$(3,25) \leftrightarrow 4$	
$(3,26) \leftrightarrow 3$	$(3,27) \leftrightarrow 4$	$(3,28) \leftrightarrow 3$	$(3,29,1) \leftrightarrow 3$	$(3,29,2) \leftrightarrow 4$	$(3,29,3) \leftrightarrow 3$
$(5,21) \leftrightarrow 5$	$(5,22) \leftrightarrow 6$	$(5,23) \leftrightarrow 5$	$(5,24) \leftrightarrow 6$	$(5,25) \leftrightarrow 5$	
$(5,26) \leftrightarrow 6$	$(5,27) \leftrightarrow 5$	$(5,28) \leftrightarrow 6$	$(5,29,1) \leftrightarrow 6$	$(5,29,2) \leftrightarrow 5$	$(5,29,3) \leftrightarrow 5$
$(7,21) \leftrightarrow 8$	$(7,22) \leftrightarrow 7$	$(7,23) \leftrightarrow 8$	$(7,24) \leftrightarrow 7$	$(7,25) \leftrightarrow 8$	
$(7,26) \leftrightarrow 7$	$(7,27) \leftrightarrow 8$	$(7,28) \leftrightarrow 7$	$(7,29,1) \leftrightarrow 7$	$(7,29,2) \leftrightarrow 8$	$(7,29,3) \leftrightarrow 7$

. . . . . .

where the mapping corresponds to the hit $(i_\phi, i_\eta, \text{depth}) \leftrightarrow$ HPD ϕ index. All depths should be assumed if not given. For reference, the index provided by the `reco::HcalNoiseHPD` object uses the above mapping, so that indices 0 through 71 correspond to the HB+, $72-143$ to the HB−, $144-215$ to the HE+, and $215-287$ to the HE−.

A 1.5 GeV threshold is applied to hits in an HPD/RBX. The total energy of an HPD/RBX is determined by summing over the hadronic energy of the associated`caloTowers`. In our analysis we require that RBXs contain at least 10 GeV of energy for consideration. Noise resulting in energies lower than this is beyond the scope of our analysis.

3 Comparing Noise and Data

In this section, we compare the characteristics of HCAL energy depositions from collision events in MC, and the anomalous noise found in cosmics events in data. Figure 5 shows the per event multiplicity of RBXs in QCD MC and CRAFT data. CRAFT data is composed of typically only zero or one RBXs above an energy threshold, although very rarely more RBXs may be present. QCD events, on the other hand, typically have many more RBXs present.

The energy scale of the RBXs is shown in Figure 6. It's worth noting that in this figure, unlike Figure 3, the energy of the RBXs is shown in a differential plot and normalized to unity, rather than calculating a rate. The energy scale for noise extends all the way to 25 TeV (!), although the plot shows only an overflow at $E_{\text{RBX}} = 5$ TeV. The energy in RBXs for each of the QCD samples increases with $\hat{p}_{\text{T}}$, as expected. RBXs have a minimum E_{RBX} requirement of 10 GeV, although this cannot be seen on the plot.

3.1 Pulse Shape

A big difference between collision data and the anomalous noise is the shape of the pulse. Figure 7 compares the energy distribution across TS in QCD MC and the CRAFT data. This pulse shape is reconstructed by summing the pedestal subtracted charge in fC in all hits with $E_{\text{hit}} > 1.5$ GeV in a single RBX (up to 72 channels). The plot shows the distribution of charge as a function of TS, normalized to unity to show shape. The difference between the QCD distributions is due to the pulse shape dependence on the total energy deposition; the pulse narrows with increasing energy.

The pulse shape in CRAFT data is much wider, and is found in all three major types of anomalous noise. As before, we categorize the different noise types simply by the number of hits found in an RBX. While ion feedback noise has the most 'signal-like' shape, on average it still possesses a significant leading and trailing tail, characteristic neither of collision data or of testbeam data (recall Fig. 1).

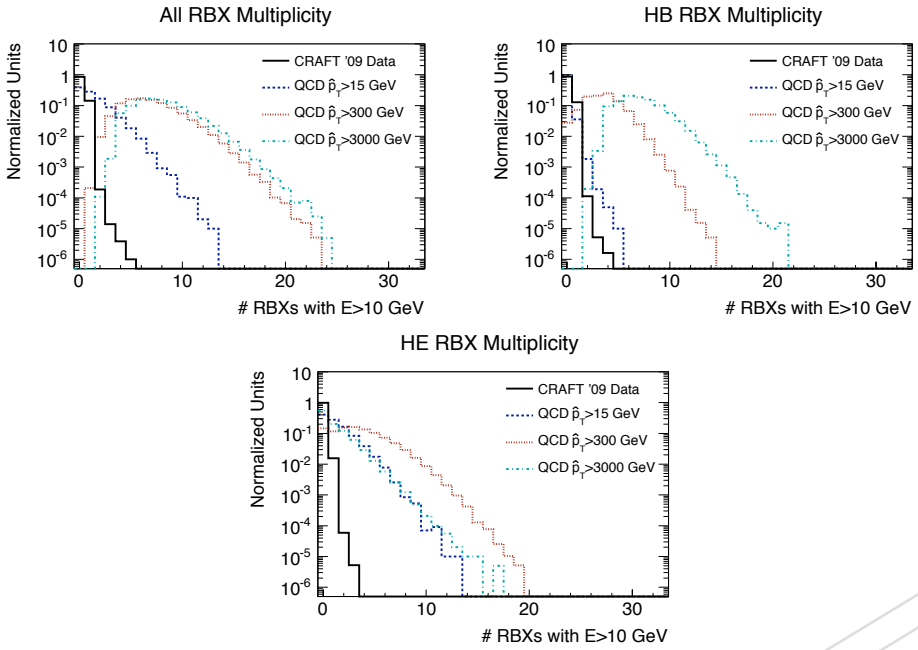

Figure 5: Multiplicity of RBXs with $E_{\mathrm{RBX}} > 10$ GeV in the HB and HE (top left), HB only (top right), and HE only (bottom). Distributions are normalized to unity to show the shape.

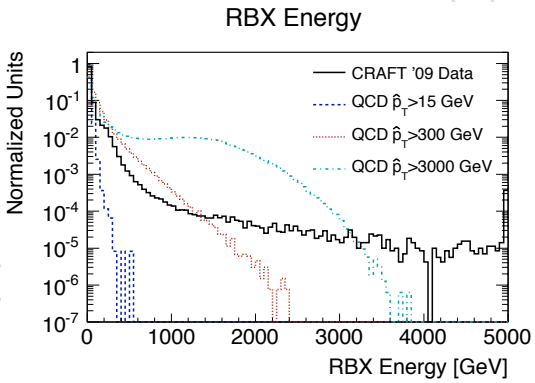

Figure 6: Energy of RBXs in CRAFT data and QCD MC. The energy of anomalous noise extends to 25 TeV in this data sample.

To separate the noise from signals, we create a pulse shape ratio variable – also called $E(2TS)/E(10TS)$ and defined earlier in Sec. 1.5 – which computes the ratio of charge found in the highest two consecutive energy TS to the total. Fig. 8 shows QCD MC for different ranges of E_{RBX} on the left. We see that the pulse shape ratio variable becomes narrower with increasing energy. By requiring that $E_{\mathrm{RBX}} > 50$ GeV, we can more effectively separate noise and signal, as low energy signals can be confused for noise. On the right of Fig. 8, we see the ratio variable for RBXs with $E_{\mathrm{RBX}} > 50$ GeV in data and MC (with tight and loose cuts shown). A clear distinction can be seen between the two. Note that a ratio exceeding unity is possible because of the charge has been pedestal subtracted, which can result in a negative charge being recorded. If a significant negative charge is recorded outside of the two peak TS, then a ratio > 1 can occur, although this occurs infrequently (less than 7% of the time in the CRAFT data sample examined here).

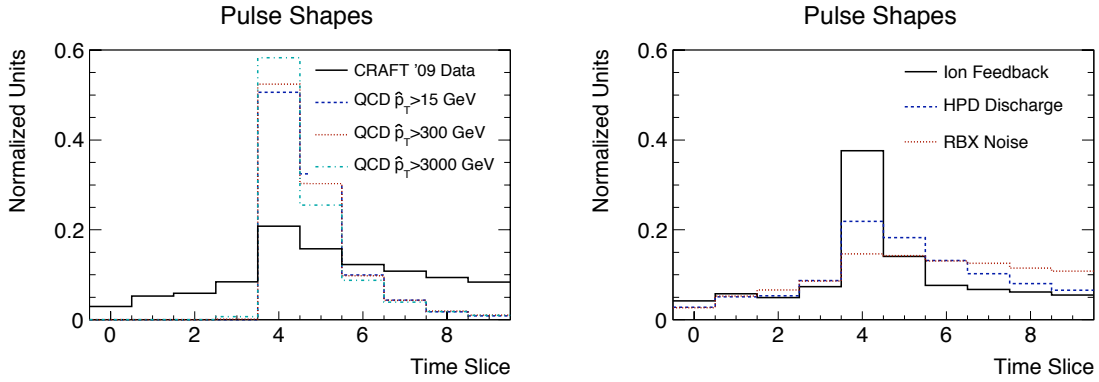

Figure 7: Pulse shape for QCD MC and noise (left). Pulse shape for different types of noise (right).

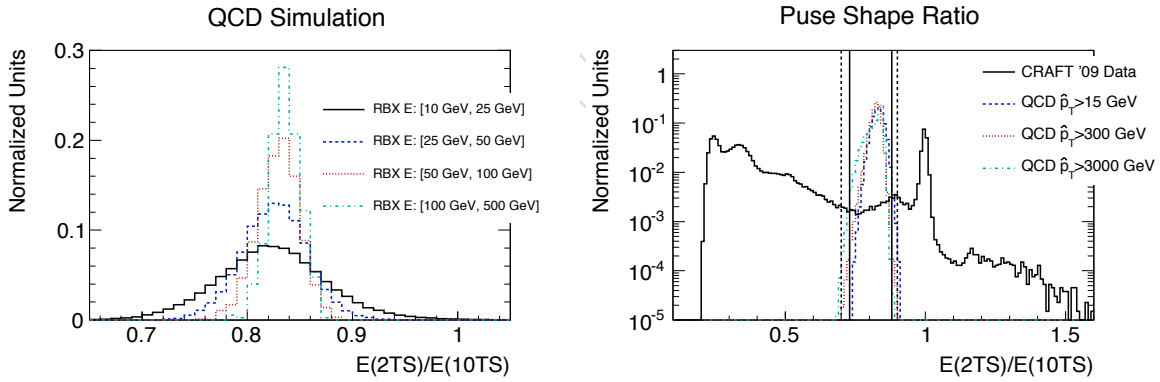

Figure 8: Pulse shape ratio variable as a function of of total energy found in an RBX in QCD MC (left). Pulse shape variable for $E_{\mathrm{RBX}} > 50$ GeV in data and MC (right). Loose (tight) cuts used by the noise filter are indicated with dotted (solid) lines. RBXs with pulse shapes outside of the indicated window are identified as noise.

Another way to view the pulse shape is as a function of the number of hits in an RBX, as seen in Fig. 9. What can be seen in this scatter plot is that the pulse shape depends strongly on the number of hits (*i.e.* the type of noise, whether ionfeedback or HPD noise, or RBX noise) associated with it. For instance, ionfeedback noise which typically consists of only a few hits very often has a $E(2S)/E(10TS)$ ratio around either 1.0 (where all the noise is deposited in two TS) or 0.3 (where the noise is uniformly distributed about all 10 TS). This averages to the pulse shape seen in Fig. 7, but the ratio distribution is really bimodal as seen in Figure 8. Similar bimodal effects can be seen in HPD discharge and RBX noise, as well, although the locations of the peaks differ. These effects are still under investigation.

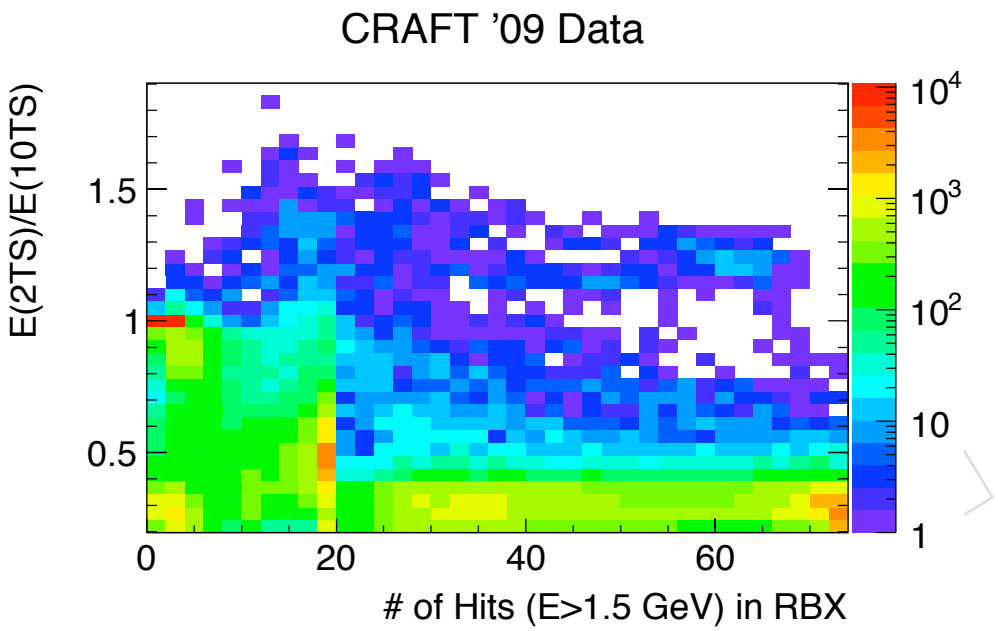

Figure 9: The pulse shape ratio variable versus the number of hits in an RBX.

It's also worth mentioning here that no pileup has been simulated in the MC. Out-of-time pileup, in particular, can effect the results shown here, adding energy to TS outside the peak, resulting in a lower pulse shape ratio. Poor timing can change the ratio, as well. For example, if too much of the signal peak finds its way into the presample, then the ratio will also be lower than expected. Pedestal drift can in principle change the ratio and so should be monitored carefully.

3.2 Timing

Timing is an important variable used to discriminate between collision signals and anomalous HCAL noise (see Ref. [5] for more on timing in CRAFT). It is, in fact, possible to use the timing relative to the trigger to identify noise, even if the noise itself was responsible for the trigger. This is due to the fact that in the HB and HE, the trigger is designed to search for a falling pulse, whereas the timing reconstruction uses a three TS window with the highest energy TS in the center to determine a time. Figure 10 in an illustration of two pulses by TS which trigger an event. In each case, the trigger comes on the 4^{th} TS, since the 5^{th} shows a falling edge. However, the reconstructed time in the pulse on the right is substantially displaced from the trigger, whereas the pulse on the left comes in time with the trigger. Because the relative timing depends on the shape of the pulse, the timing and pulse shape variables are correlated with

each other.

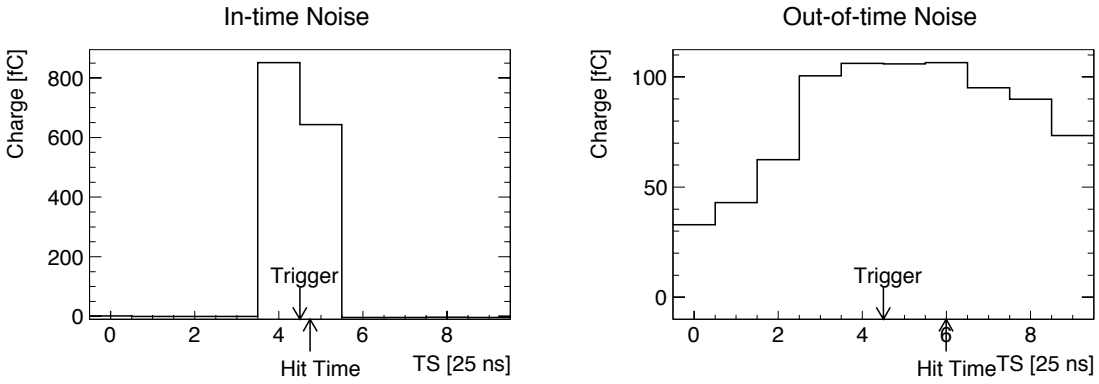

Figure 10: Timing for two pulse shapes using either the trigger or the reconstructed time.

Fig. 11 shows time for hits with $E_{\text{hit}} > 25$ GeV. Since the timing resolution improves with energy, a modest energy threshold is necessary. Noise falls characteristically outside of the signal timing window. The multiple peak structure is related to the same structures seen earlier with the pulse shape variable.

Looking to the future, it's obvious that for timing cuts to work, the timing must be properly calibrated on the order of a few ns. This has not yet been accomplished using collision data, so − until then − we suggest caution when using timing cuts. Additionally − as with the pulse shape variable − pileup can affect the timing and result in the mistaken rejection of collision signals. In the presence of significant pileup, cuts on the timing to reject whole events may need to be reconsidered.

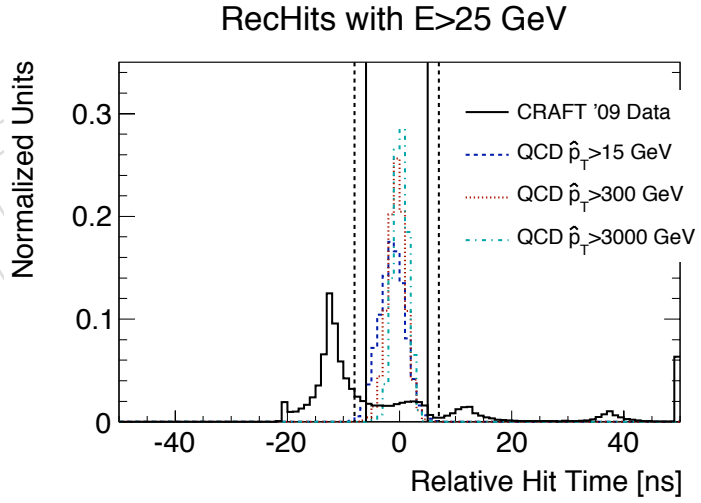

Figure 11: Timing for 25 GeV hits in QCD MC signal and CRAFT noise data. Distributions are normalized to unity, and the last bin is an overflow. The loose (tight) cuts used by the noise filter are indicated with dotted (filled) lines. Hits that fall outside the window are identified as noise.

3.3 ADC 0 Counts

Another characteristic of noise is the presence of a large number of ADC's which read identically 0. The reason for this is unknown. We calculate this variable by counting the number of ADC 0's found in any TS of any channel in an RBX without regard to the energy deposited there. Fig. 12 shows the number of ADC 0's found in an RBX in QCD MC and in CRAFT data. As many as 210 ADC 0 counts have been found in a single RBX in data, whereas the probability of finding more than 10 ADC 0's is $\sim 10^{-6}$ in QCD MC. The number of ADC 0 counts has a modest positive correlation with the energy in the RBX. The high multiplicity is found in all three types of noise and is particularly high for RBX noise, as seen on the right in Fig. 12. However, it is not seen in RBXs in which no anomalous noise is present.

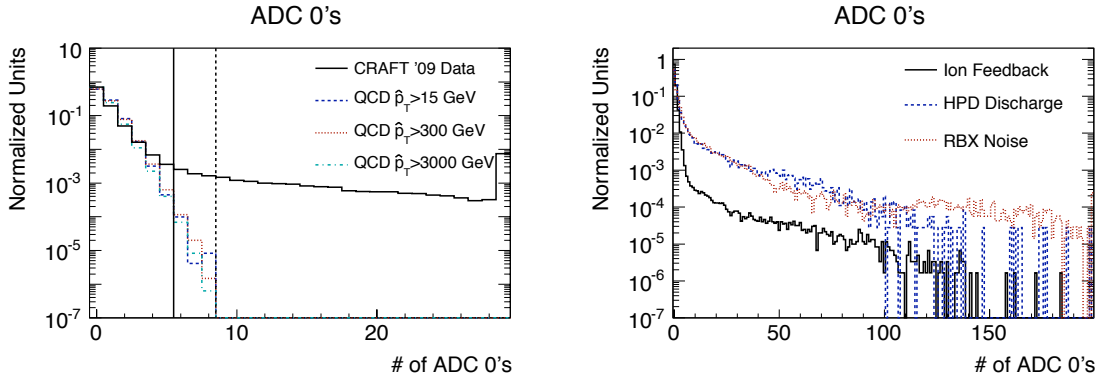

Figure 12: Number of ADC's which read 0 in a single RBX for CRAFT data versus MC (left) and for different types of noise (right). The loose (tight) cuts used by the noise filter are indicated with dotted (filled) lines. RBXs with more ADC 0's than indicated by the cuts are identified as noise.

While the cause is unknown and the discriminating power between noise and signal is small, a high multiplicity of ADC 0 counts is at least expected to be robust against misidentification of real physics signals as noise. Changes in pileup or other running conditions are not expected to have a substantial effect on the variable. Moreover, the probability of pedestal noise fluctuating down to 0 ADC counts in a single channel corresponds to $\sim 2.5\,\sigma$.

3.4 Hit Multiplicity

The hit multiplicity per HPD, per RBX, and per HPD when no other hits are found in the RBX is shown in Figure 13 (again, $E_{\text{hit}} > 1.5$ GeV). Multiple peaks at 1, 18, and 72 hits in the CRAFT data correspond to Ion feedback noise, HPD discharge noise, and RBX noise. What this makes clear is that even in the highest occupancy events the probability that an event could produce hits in nearly all 18 channels in an HPD (much less all 72 channels in an RBX) is very small. Except in the instance of significant pileup, this conclusion seems robust.

The only drawback of this variable is that much of the noise actually has a low hit multiplicity. Using pulse shape and timing cuts are still important in order to reduce the total noise rate.

3.5 Electromagnetic Fraction

Finally, we can use coincidence with the ECAL to identify anomalous noise. Here, we associate towers geometrically with the RBX and measure the total hadronic and electromagnetic energy as measured by the towers. Fig. 14 shows the fraction of RBXs with an EMF>1%, > 2%, and

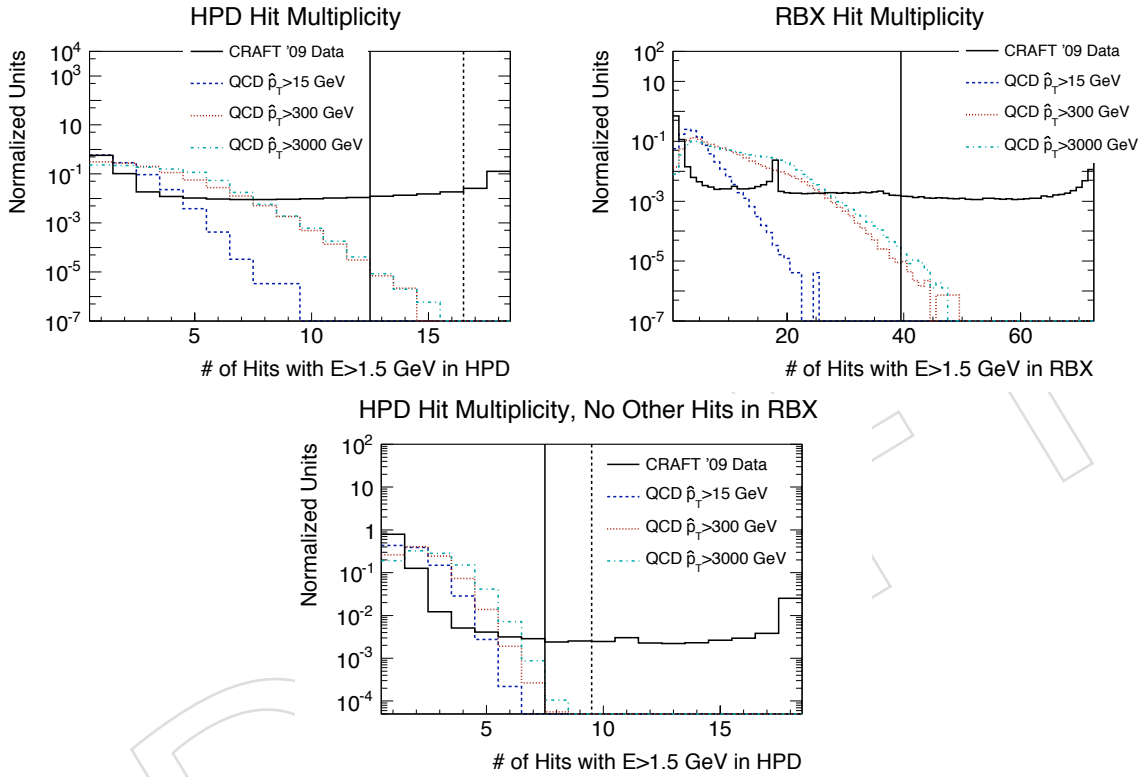

Figure 13: Number of hits ($E_{\text{hit}} > 1.5$ GeV) in a single HPD (top left), in a single RBX (top right), and in a single HPD when no other hits are found in the RBX (bottom). The loose (tight) cuts used by the noise filter are indicated with dotted (filled) lines. HPDs or RBXs with more hits than indicated by the cuts are identified as noise. There is no loose cut on the number of hits in an RBX.

$>$ 5% as a function of the energy in the RBX. For instance, we see that approximately 13% of all RBXs with more $E_{\mathrm{RBX}} > 20$ GeV in CRAFT data have an EMF which exceeds 1% (ECAL noise makes the electromagnetic fraction in CRAFT data non-0), whereas between $98 - 100\%$ of RBXs in QCD events have EMF$>$ 1%.

Fig. 14 also shows that softer QCD events have a higher rate of low EMF RBXs, and increasing the EMF cut also increases the rate that QCD events will mimic noise. However, it is also worth pointing out that for very high $\hat{p}_{\mathrm{T}}$ QCD events, a tigher EMF cut does not substantially change the over-efficiency of the cut. This suggests that an analysis dependent cut on the RBX EMF could be particularly useful in eliminating noise. A harder cut would eliminate softer physics events, but not harder physics events and could be helpful in cleaning up analyses that are particularly sensitive to anomalous HCAL noise.

We should also consider here the possibility of a substantial selection bias in the $\hat{p}_{\mathrm{T}} > 15$ GeV sample. The average energy of an RBX (after requiring a 10 GeV threshold) is only $\sim$18 GeV. RBXs which have more energy are likely to have fluctuated high, and so the EMF fluctuated low, giving the resultant over-efficiency.

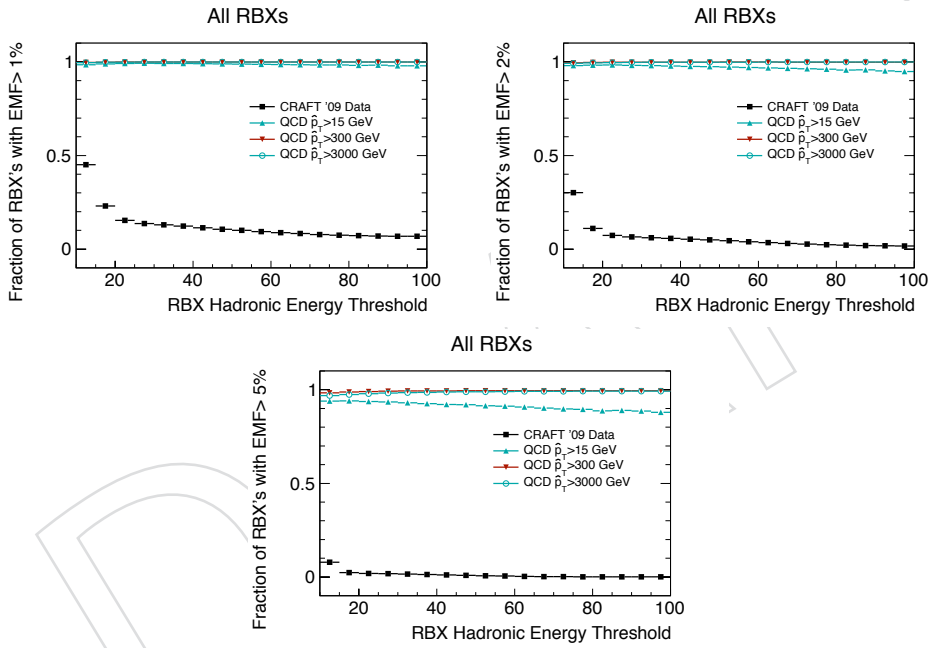

Figure 14: Fraction of RBXs with an EMF$>$ 1% (top left), $>$ 2% (top right), and $>$ 5% (bottom) as a function of a cut on total hadronic energy in the RBX. For the noise filter, we reject events if an RBX with $E_{\mathrm{RBX}} > 20$ GeV has an EMF$<$ 1%.

As per caveats, at startup when anomalous HCAL noise primarily does *not* overlap with min-bias collisions, an EMF requirement will be a very stringent requirement as it has very little dependence on a proper timing calibration, pulse shapes, or other assumptions built in to HCAL specific variables. However, as the LHC approach design luminosity, more an more noise events will coincide with some type of collision physics, even if it not *a priori* triggerable. As such overlap become more common, an EMF requirement will by itself become less powerful.

4 Noise Filter Performance

In this section we discuss the performance of the cuts we introduced in Sec. 1.5 and described in detail in Sec. 3. The cuts are introduced in sets, comprised of a loose filter which uses HCAL based variables only, a tight filter which is a tighter set of cuts than the loose filter, and a high level filter which uses coincidence with the ECAL.

Figure 15 shows the inefficiency of the three filters plus a combined loose+high level and a tight+high level filter in CRAFT '09 data, specifically run 112237. The inefficiency is plotted as a function of the maximum energy found in an RBX and as a function of a cut on the same variable. Note that this variable is calculated per-event, not per-RBX. For the combined filters, the inefficiency is 0% in CRAFT data for RBXs with $E_{\mathrm{RBX}} > 340$ GeV − that is, every event in data with an RBX exceeding 340 GeV is rejected as noise. The filter inefficiency is between 10^{-3} and 10^{-4} for events with $E_{\mathrm{RBX}} > 100$ GeV, although it increases rapidly with lower energy.

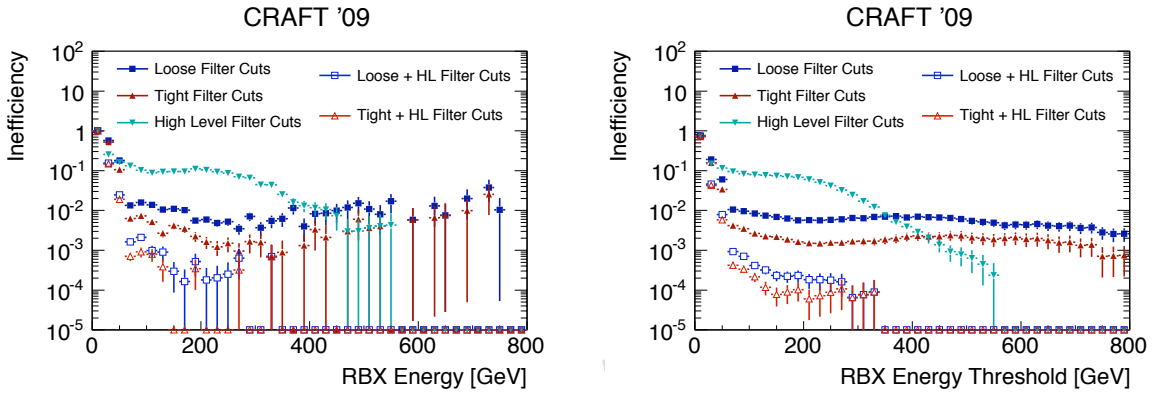

Figure 15: The inefficiency of the noise filters in CRAFT '09 data as a function of the maximum energy found in an RBX in the event (left) and as a function of the threshold (right). The filters shown are the loose, tight, and high level filters described in Sec. 1.5. Also shown are a combined loose+high level filter and a tight+high level filter.

Figs. 16, 17, and 18 show the over-efficiency for the same filters as a function of the maximum RBX energy and energy threshold in QCD MC. An energy dependence can be clearly seen in all three $\hat{p}_{\mathrm{T}}$ samples. In the QCD $\hat{p}_{\mathrm{T}} > 15$ GeV sample, the over-efficiency is larger than in the higher $\hat{p}_{\mathrm{T}}$ samples due to the EMF criteria. As discussed in Sec. 3.5, this may also be exacerbated by an selection bias whereby RBXs only contain a lot of energy if the total energy fluctuated high and if it fluctuated into hadronic, rather than electromagnetic, energy. Regardless, the average RBX energy, $\langle E_{\mathrm{RBX}} \rangle$, is less than 20 GeV in this sample, and the filters work best with RBXs exceeding 50 GeV, so it is not the best sample with which to measure a baseline.

The QCD $\hat{p}_{\mathrm{T}} > 300$ GeV sample is the least over-efficiency of the three QCD samples we study here. The over-efficiency is $\sim 0.4\%$ for the combined loose and high level filter, and $\sim 1.9\%$ for the combined tight and high level filter. These values climb with energy. The over-efficiency is nearly 3.5% for the combined tight and high level filter in the $\hat{p}_{\mathrm{T}} > 3000$ GeV sample.

5 Conclusions and Plans

We present a set of filters to reject anomalous HCAL noise. These filters use HCAL based variables − timing, pulse shapes, etc. − and coincidence with ECAL to suppress noise events with $E_{\mathrm{RBX}} > 100$ GeV by nearly four orders of magnitude with the tightest combination of

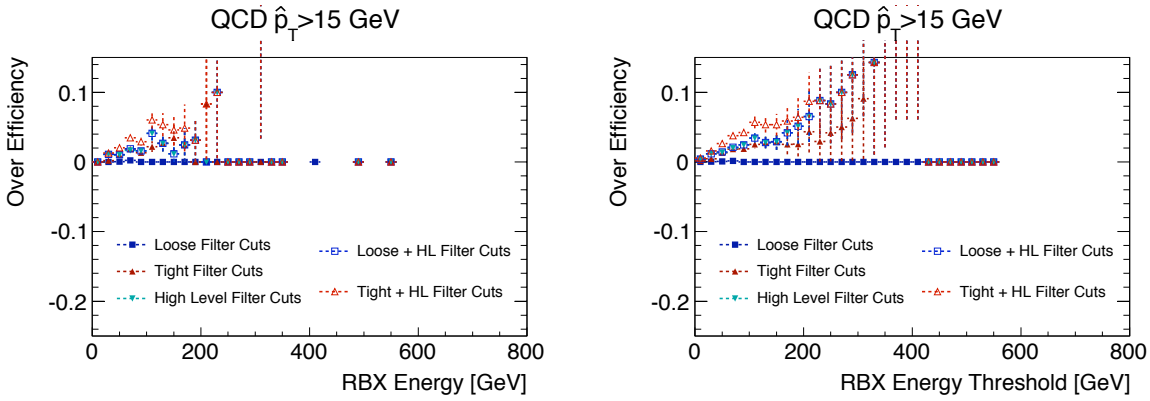

Figure 16: The over-efficiency of the noise filters in QCD $\hat{p}_T > 15$ GeV MC as a function of the maximum energy found in an RBX in the event (left) and as a function of the threshold (right). The filters shown are the loose, tight, and high level filters described in Sec. 1.5. Also shown are a combined loose+high level filter and a tight+high level filter.

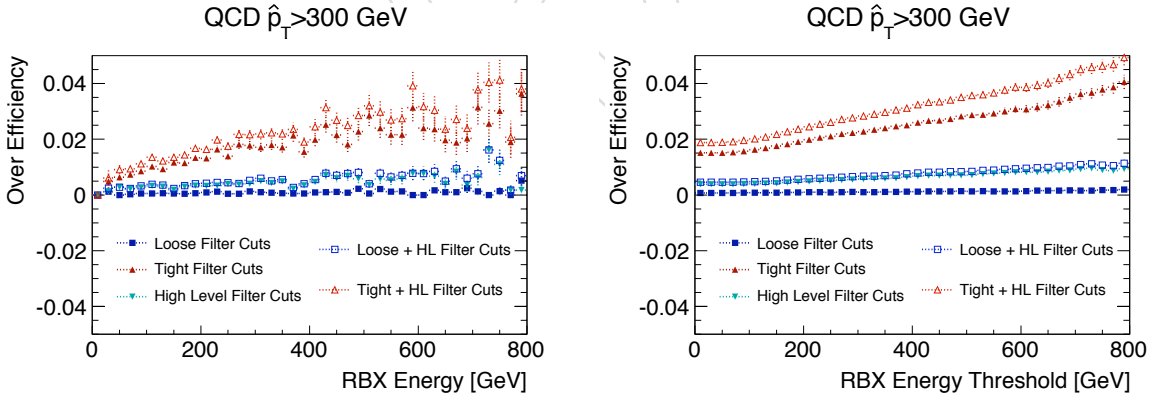

Figure 17: The over-efficiency of the noise filters in QCD $\hat{p}_T > 300$ GeV MC as a function of the maximum energy found in an RBX in the event (left) and as a function of the threshold (right). The filters shown are the loose, tight, and high level filters described in Sec. 1.5. Also shown are a combined loose+high level filter and a tight+high level filter.

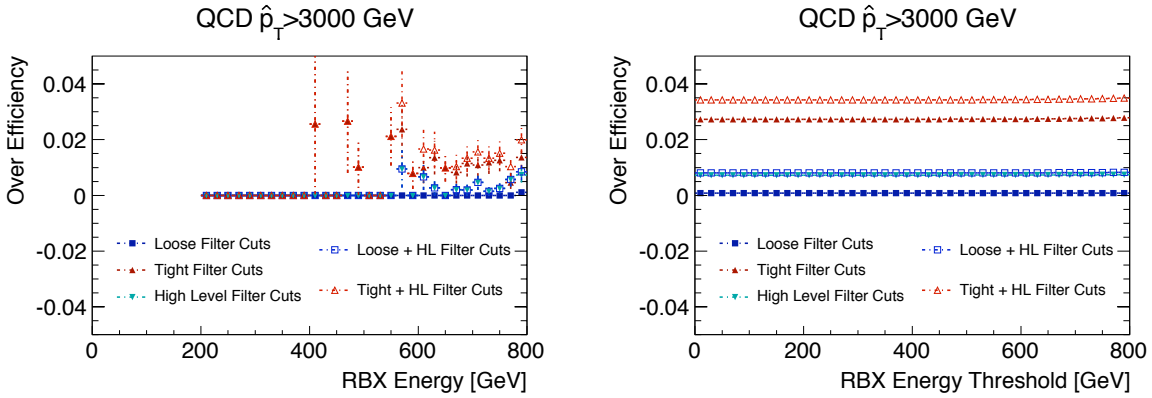

Figure 18: The over-efficiency of the noise filters in QCD $\hat{p}_T > 3000$ GeV MC as a function of the maximum energy found in an RBX in the event (left) and as a function of the threshold (right). The filters shown are the loose, tight, and high level filters described in Sec. 1.5. Also shown are a combined loose+high level filter and a tight+high level filter.

filters. In contrast, the over-efficiency in QCD MC samples is at most $\sim 3.5\%$ for events with $E_{\mathrm{RBX}} > 100$ GeV.

However, several thing yet remain to be accomplished. A more thorough study of the efficiency of the filters is needed with noise overlaid with minimum bias events. This relies on the data mixer which is only recently out of development. Another area of study is the timing and pulse shape of signals from collision data. We have thus far assumed that MC well-reproduces these distributions. Tuning of the timing and pulse shape will be an important first task once a substantial integrated luminosity of high energy collisions is recorded.

Further information including updates to the coding infrastructure will be posted at [4].

References

[1] CMS Collaboration, "CMS Physics TDR", *CERN-LHCC-2006-001* **Volume I** (2006).

[2] S. Abdullin et al., "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges", 2008. doi:10.1140/epjc/s10052-008-0573-y.335.

[3] CMS Collaboration, "Identication and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter", *arXiv:0911.4881v3; Accepted by JINST* (2009).

[4] *https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalNoiseInfoLibrary*.

[5] CMS Collaboration, "Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data", *arXiv:0911.4877v2; Accepted by JINST* (2009).