DRAFT
CMS Physics Analysis Summary

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

2010/06/26
Head Id: 11612
Archive Id: 11611:11612
Archive Date: 2010/06/27
Archive Tag: trunk

HCAL Commissioning in Proton Collisions at sqrts=7 TeV

The CMS Collaboration

DRAFT

Abstract

We present results on the commissioning and performance of the CMS hadron calorimeters in proton collisions at a center of mass energy of 7 TeV at the Large Hadron Collider. The hadron calorimeter of CMS is divided into sub-systems covering a wide range of pseudo-rapidity utilizing different technologies and electronics. Anomalous background signals, which had been previously observed in data collected in test beam running, have been characterized and studied in collision data. Methods to identify and remove these anomalous signals have been developed and their performance is presented. The hadronic calorimeters are used to trigger the experiment on energy clusters and the trigger performance is discussed. Methods to calibrate the calorimeters using cosmic muons, beam splash events (where the LHC beam is targeted on upstream collimators), and collision data are presented.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: Frank Chlebana
PDFTitle: HCAL Commissioning in proton Collisions at sqrts=7 TeV
PDFSubject: CMS
PDFKeywords: CMS, physics, software, computing

Please also verify that the abstract does not use any user defined symbols

DRAFT

1 Introduction

We present results on the commissioning and performance of the CMS hadron calorimeters (HCAL) in proton collisions at a center of mass energy of 7 TeV at the Large Hadron Collider. The hadron calorimeter of CMS is divided into sub-systems covering a wide range of pseudo-rapidity utilizing different technologies and electronics. The HCAL is used to measure hadronic energy deposits and helps determine the missing transverse energy resulting from neutrinos or exotic particles and plays an essential role in searches for new physics. An overview of the HCAL detectors is presented in Section 2. HCAL modules were exposed to test beams in order to establish the energy response to 50 GeV pions. Results from a few channels are then extrapolated to the entire calorimeter. This pre-calibration does not take into account any additional material of the fully assembled CMS detector. Methods to calibrate the calorimeters using cosmic muons, beam splash events (where the LHC beam is targeted on upstream collimators), and collision data are presented in Section 3.

Continuous monitoring is required in order to ensure a high efficiency of data collection that is of a quality suitable for physics analysis. Details of the monitoring tools are presented in Section 6. Output from the monitoring is used to certify the data so that it can be included in physics analysis. The data certification procedure is presented in Section 7.

Anomalous background signals have previously been observed during the exposure of modules to test beams and early commissioning. Algorithms to identify and remove these anomalous signals have been developed and tested with collision data. Details are presented in Section 8. The algorithms exploit the properties of the anomalous signals, such as energy isolation and timing. In order to take full advantage of these properties it is essential to have a good understanding of the pulse shape and well aligned channels.

The hadronic calorimeters are used to trigger the experiment on energy clusters and missing E_T. An overview of the HCAL trigger is presented in Section 5.

2 HCAL Description

The CMS calorimeter is composed of an inner electromagnetic calorimeter (ECAL) surrounded by a hadronic calorimeter (HCAL) enclosed within a solenoid operating at 3.8 Tesla. The ECAL consists of a barrel region (EB) covering the pseudorapidity range of $1.5 < |\eta|$ and an endcap section (EE) covering $1.4 < |\eta| < 3.0$. The HCAL barrel (HB) covers the region $\eta < 1.3$ and consists of 36 azimuthal wedges assembled into two half-barrels (HB+ and HB-). The Hadronic endcap calorimeter (HE) covers the pseudo rapidity range $1.3 < |\eta| < 3$. The hadronic forward calorimeter (HF) extends the coverage $3.0 < |\eta| < 5.0$. Additional scintillators (HO) are located outside of the solenoid and act as "tail catchers" effectively increasing the thickness of the calorimeter in the central pseudo rapidity region. Figure 1 is an elevation view of the CMS detector showing the HCAL components with lines of constant pseudo rapidity overlayed.

The HB is a sampling calorimeter consisting of alternating plates of brass absorber and scintillator tiles embedded with wavelength shifting (WLS) fibers. The WLS fibers are spliced to clear fibers which direct the light to an optical decoder unit (ODU). The ODU arranges the fibers into readout towers covering an area of 0.087×0.087 in $\eta - \phi$ and direct the light from each tower to seperate channels of a hybrid photo diode (HPD) which can operate in a high magnetic field. Each HPD has 18 channels of readout.

Analog-to-digital converters (ADCs) digitize the signals from the calorimeter for readout. Charge is integrated over each 25 ns time sample seperately and a total of ten time samples are recorded.

DRAFT

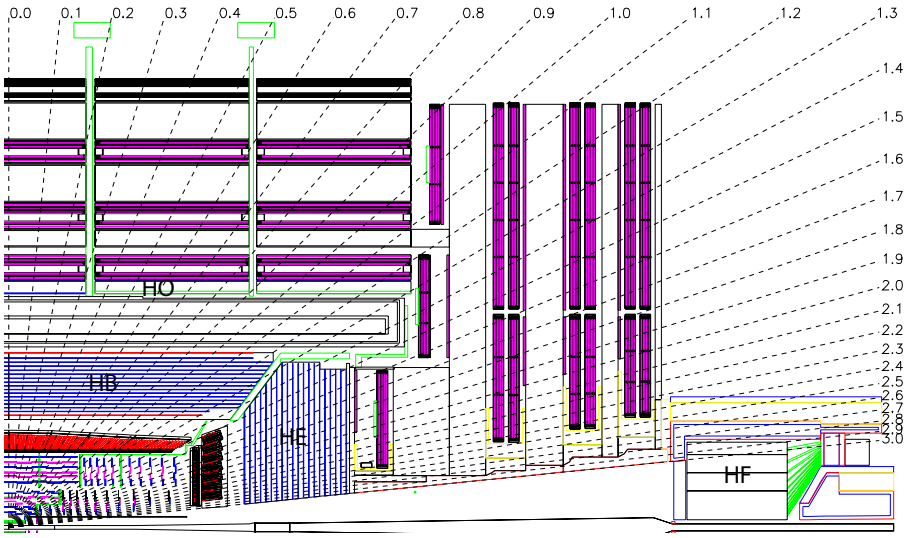

Figure 1: An elevation view of the CMS detector showing the HCAL subsystems (HB, HE, HO, and HF). Lines of constant pseudo-rapidity are shown as the dashed lines in the figure.

Signals from 4 HPDs in HB/HE (or 24 PMTs in HF) are digitized within a single read-out box (RBX). A more detailed description of the CMS calorimetry system can be found in [1]. Discussions of the design and performance of the HB, HE, and HF sub-detectors may be found in [2]-[3].

The HF sub-detector is a Cerenkov light detector made up of quartz fibers embedded within a 165-cm-long steel absorber. There are two types of fibers within the HF: "long" fibers which span the length of HF, and "short" fibers which begin 22 cm into the detector. Differences between signals read out from the long and short fibers can be used to distinguish between electrons or photons and hadrons. Photomultiplier tubes (PMTs) connected to the fibers via light guides convert detected light to electrical signals.

Energy is reconstructed from the digitized charge measurements. In HB, HE, and HO 4 time samples are summed together. The signal in HF is very fast and has a width of about 10ns which is mostly contained within one 25ns time sample. Individual channels are combined into a projective tower structure which combines information from ECAL and HCAL. In order to save disk space, the raw digitized data is dropped from the event record and most physics objects which use calorimeter information are then derived from the CaloTower objects.

3 Calibration

Calorimeter modules were exposed to test beams of pions with known energies in order to obtain the measured response at severeal energies. The energy scale for 50 GeV pions was then extrapolated to the remaining channels of the entire calorimeter using Co60 radioactive sources. Details of the test beam analysis and the wiresourcing can be found in Ref. [4]. This procedure does not take into account any addition dead material of the fully assembled CMS detector.

The calibration of HCAL was further improved using cosmic muons and so called "beam splash" events. Beam splash refers to data collected when the proton beam was intentionally steered into collimators upsteam of the CMS detector in order to produce a spray of muons. Both methods provided valuable information that allowed us to equalize the detector response

DRAFT

in the +/- sides of HB, and derive relative channel-to-channel corrections in HB and HE. Splash events were also used in the identification of a few channels that had a relativly large deviation in the response from adjacent cells. More details can be found in Ref. [5].

An example of a ϕ-profile of the energy deposited in HB from splash events is shown in the left plot of Figure 2. We expect smooth energy distribution and miscallibrated channels are identified as having energy measurement that deviate significantly from the trend in their vicinity. The same arguments are also valid for the energy distributions in z which is shown in the right plot of Figure 2. Combining the information from the ϕ- and z-distributions allows us to equalize the response over the entire HB with the exception of cells near the outer boundaries. A similar procedure was applied to HE.

Figure 2: The left plot shows the ϕ-profile of the energy deposition in HB by splash events while the right plot shows the z-profile.

These techniques combined with the source calibration provide our initial measure of the absolute energy response of the HCAL sub-detectors and is referred to as the precalibration condition. Precalibration does not fully account for all the dead material that particles orignating from the interaction point will traverse. Additional calibration methods using collision data are needed. It is also necessary to continuously monitor the response of HCAL to compensate for aging effects.

A calibration workflow was developed that includes several techniques covering HB, HE and HF. It was targeted at early data calibrations that could be performed with tens of pb^{-1}. The workflow includes the following steps; first an azimuthal symmetry correction is applied to equalize the response in HB, HE, and HF within rings of constant η; next, an absolute energy correction is determined using isolated charged particles for HB and part of HE; finally, the absolute response is extended to the forward region using dijet and photon-jet events. The calibration steps are not completely independent and the results of each step are used in the subsequent one. These corrections are derived with respect to the "precalibrated" conditions.

Due to the complex structure of the hadronic calorimeter, its large coverage and different overlapping regions with other detector systems used for the calibration it is necessary to use multiple techniques and data samples. Additional complications arise from the non-linearity of the HCAL energy response and the relatively large lateral size of hadronic showers and it is not possible to set an absolute scale that is valid for all energies of incident hadrons. We define the target absolute scale to correspond to E_{had}/p_{trk}=1 for charged hadrons with momentum 50 GeV that do not interact in the electromagnetic calorimeter where (E_{had} is the measured energy in a tower cluster described later in the text. The criteria for this choice is that the energy is in a region where the calorimeter response as a function of energy is slowly changing and it can be set and tested directly.

The first step of HCAL calibration with collisions data is to equalize the response in ϕ for each η ring. The procedure takes advantage of the azimuthal symmetry of the detector and the corresponding ϕ-symmetric energy deposition from minimum bias (MinBias) events. The intercalibration is performed by comparing the average energy deposition in a calorimeter cell to the mean of the average energy distributions in the entire η-ring (cells with $i\eta$=const). One of the main challenges is the large channel to channel noise fluctuations and relatively small signal in HB and HE.

Isolated charged hadrons are used to set the absolute energy scale of the hadronic calorimeter in the region with tracking coverage. Tracks with momenta in the range $40 < p < 60$ GeV are

selected to correspond with the energy scale set in the test beam. Requiring that the track is isolated prevents overlapping energy depositions from other particles. Since charged hadrons can often deposit a substantial part of their energy in the electromagnetic calorimeter (ECAL), we select only tracks that behave like minimum ionizing particles (MIPs) in the ECAL by requiring their energy deposition in ECAL to be below a threshold.

The tracks are selected based on their quality, isolation from charged particles in the tracker, isolation from neutral particles in ECAL, and their energy deposition in ECAL. Energy in HCAL is collected in a cluster defined by a cone around the direction of the track at the HCAL surface. The selections are optimize to ensure good energy containment in the cluster, low energy contamination from neighboring tracks, and sufficiently high selection efficiency under realistic data-taking scenarios. Our current estimation is that we need 5-10 pb^{-1} of data collected using a dedicated trigger to obtain η-dependent response corrections in HB and HE.

Di-jet balancing and photon-jet balancing were proposed for calibrating HF and the HE region not covered by the tracking system. This calibration procedure is performed after the phi symmetry corrections are available and after HB is fully calibrated. The procedures are similar to the ones used for jet energy scale determination. The difference is that the derived corrections should equalize the response on a cell-by-cell basis. This poses a significant challenge and the procedures are still under development.

Here we listed only the basic set approaches that are going to used with early data. There are a number of other techniques that are under development and showing promising results.

3.1 First Results with Collision Data

The full calibration of HCAL with collision data requires the collection of large samples of events. Until we have sufficient statistics we can use the existing data to test the procedures and to get a first look at the HCAL response by using tracks with a lower momenta range. Samples with $p_{trk} > 5$ GeV tracks were selected from MinBias events to check the relative response of different detector systems and the quality of the MC simulation.

[description of the early studies, at this time put only the plots that we want to show]

Figure 3 shows the distribution of E/p_{trk} for isolated tracks with a momentum rage of $9 < p < 11$ GeV.

[responses in +/- sides, ratios]

4 HCAL Timing

4.1 Timing in HB and HE

Timing and synchronization methods for the HCAL barrel and endcap from test beams and beam splash events were previously reported in [12]. With the advent of LHC collisions at CMS came the first opportunity to assess the effectiveness of this prior campaign for timing and synchronization of collisional signals. The difficulty in synchronizing the barrel and endcap channels with collision data is the reduced number of high-energy, well-timed events as $|\eta|$ goes to zero.

A study was undertaken to assess the performance of front-end channel phase settings in the barrel and endcap that were derived from so-called "beam splash" events in late 2009. Collision events were selected by requiring a high-quality reconstructed primary vertex for the event. Gaussian fits were applied to each channel, and the means of the fits are plotted. The resulting

DRAFT

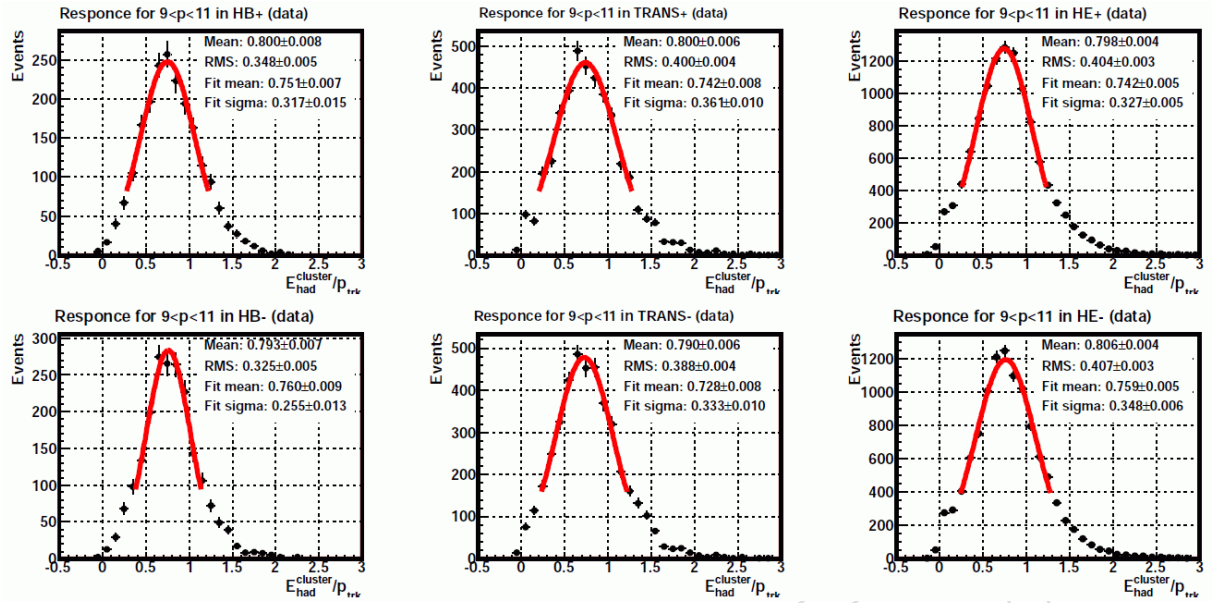

Figure 3: Example response profiles for tracks with momenta $9.0 < p_{trk} < 11.0$ GeV in the positive and negative sides of HB, HE, and the "transition" region. The solid lines represent the results of Gaussian fits.

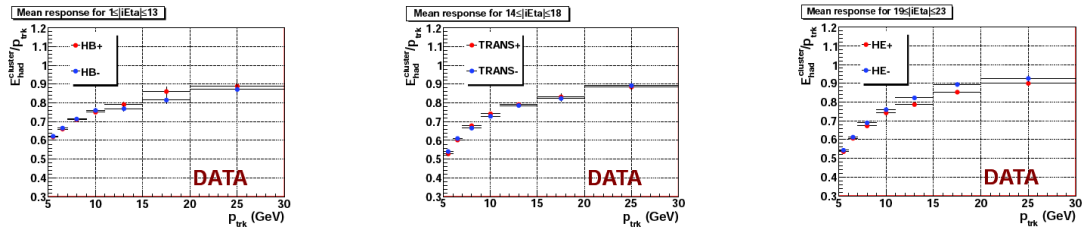

Figure 4: Comparison of the response in the positive and negative sides of the hadron calorimeter. The left, right, and middle column show the results for tracks in HB, HE, and the "transition" region as described in the text. The top row of plots show the measured response as a function of track momentum, while the lower plots display the ratios of the response in the positive side to the response in the negative side.

DRAFT

htbp

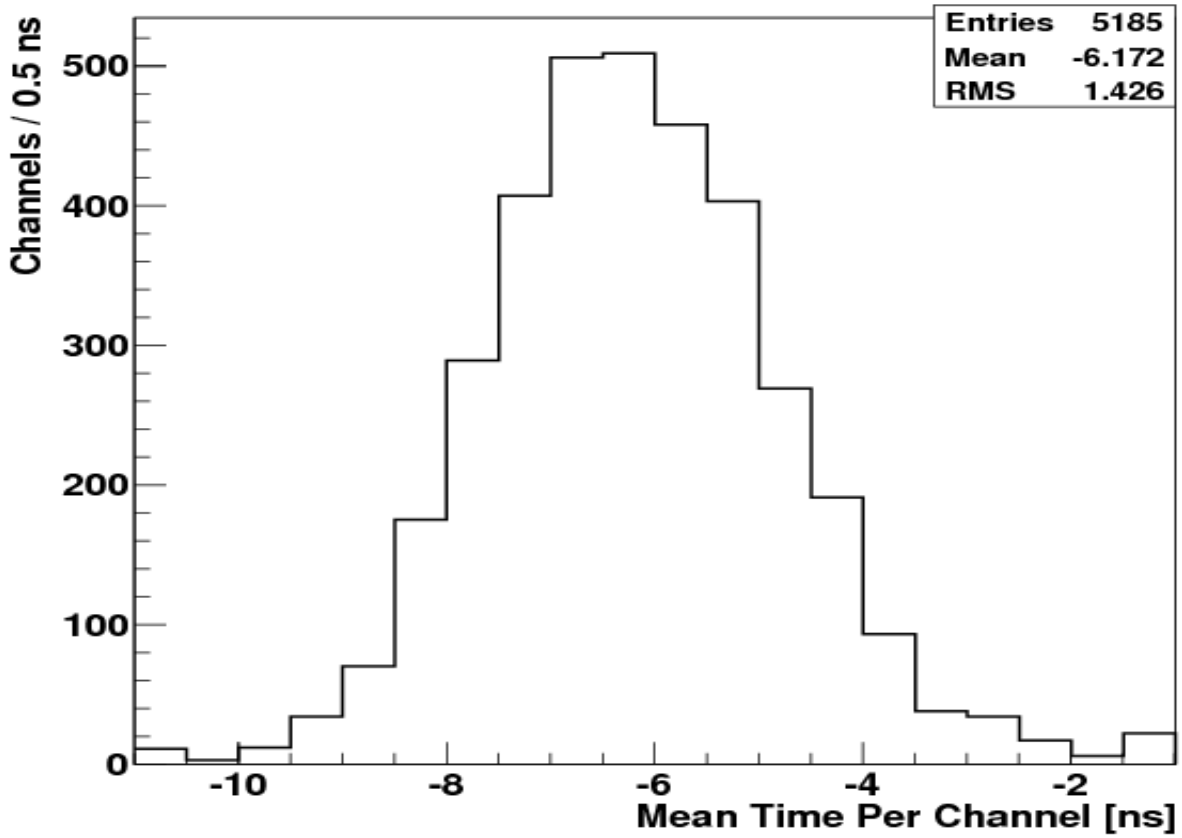

Figure 5: The distribution of per-channel timing means in the HCAL barrel and endcap, showing the quality of front-end channel synchronization. A 12 GeV minimum cell energy threshold was applied.

htbp

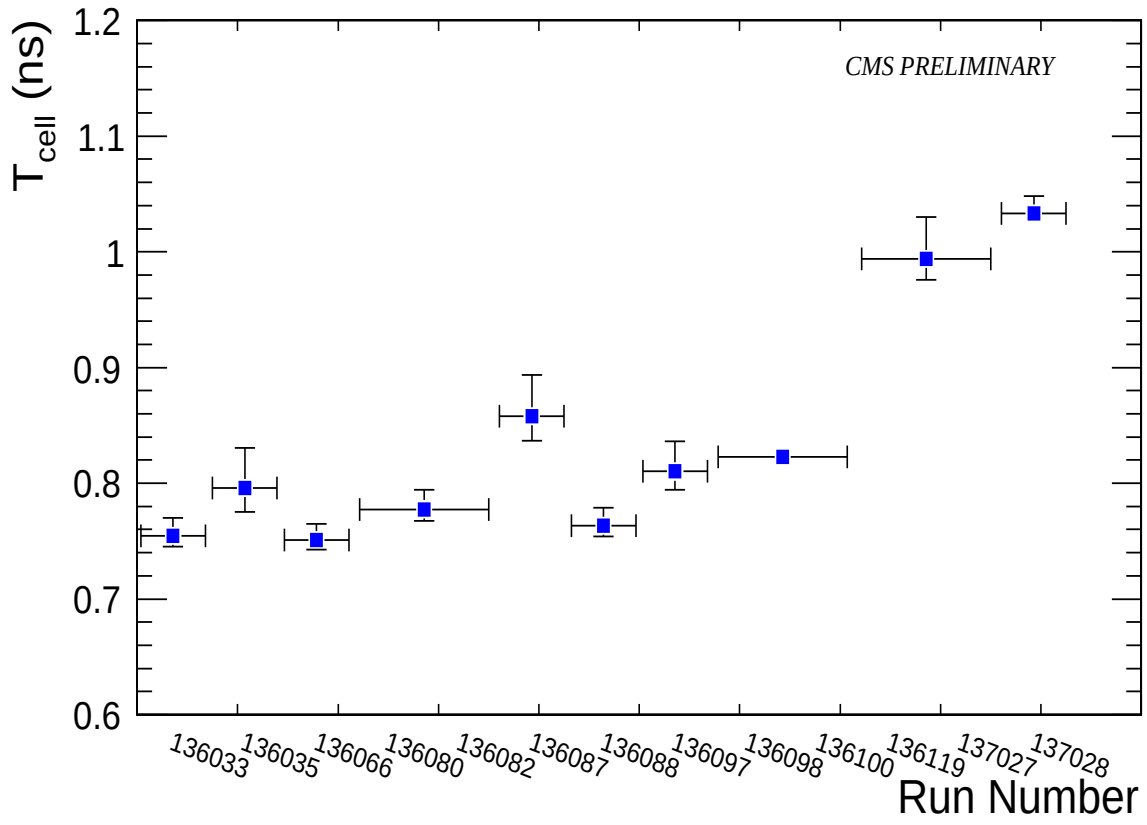

Figure 6: Timing stability in the barrel and endcap combined, as a function of run number. A 20 GeV minimum cell energy threshold was applied.

htbp

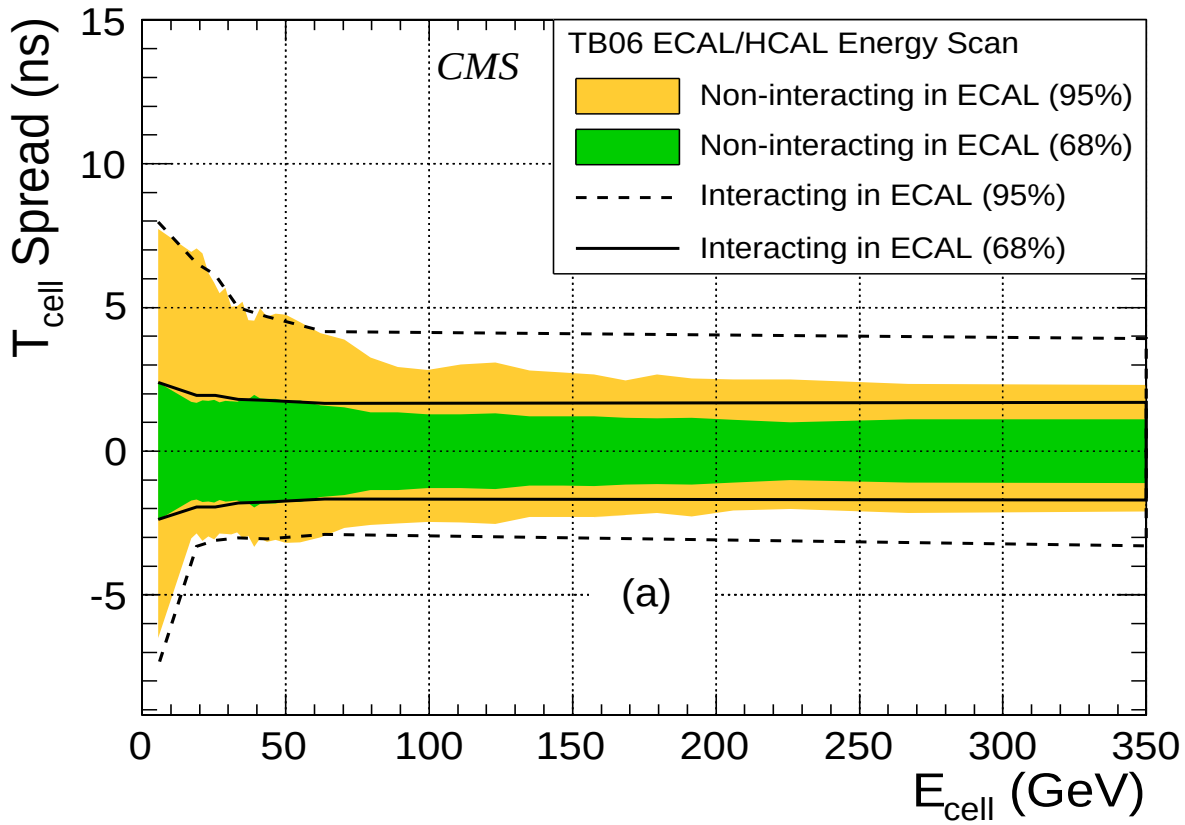

Figure 7: *NB: PLOT TO BE UPDATED*. Timing resolution as a function of energy measured during test beam runs in 2006, showing the consistency of time reconstruction for particles that begin showering in the ECAL (lines) and those that do not (areas).

htbp

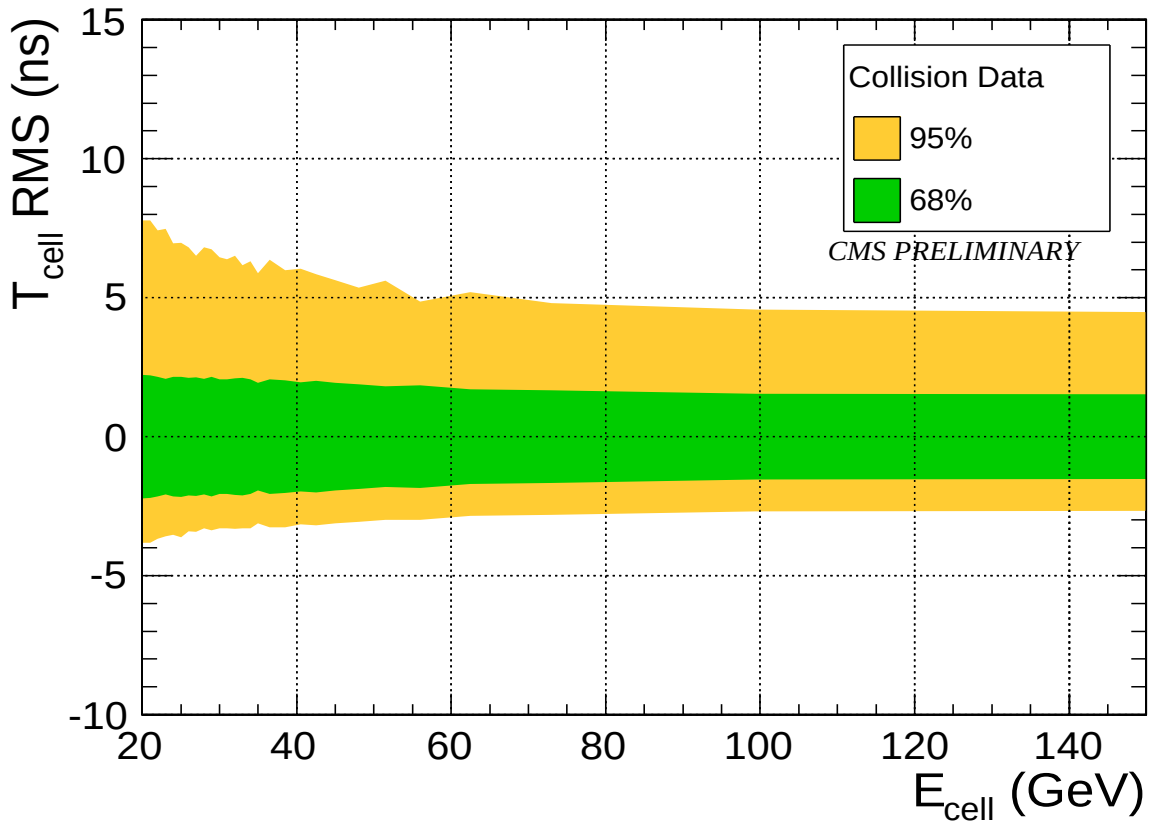

Figure 8: Timing resolution as a function of energy measured during 3.5 TeV/beam collision runs in 2010. A 20 GeV minimum cell energy threshold was applied.

distribution of mean channel timing is shown in Figure 5. These corrections are then applied offline to each calorimetric timing measurement in order to improve the channel alignment and to define a nominal "t=0" for physics analysis, as evidenced by the Figures 6 and 8. Efforts to improve the channel alignment are continuing, particular as more data are collected.

Figure 6 exhibits the stability of timing in the barrel and endcap over several of the data runs taken so far. Figures 7 and 8 compare the timing resolutions as a function of cell energy as measured in test beams and in collision data. The additional spread in collision data is attributed to ion feedback and residual channel misalignment.

4.2 Timing in HF

In contrast to the barrel and endcap, the arrival time of signals from the HCAL Forward detector cells are much more difficult measure. This is because, due to the high expected occupancy, the forward detector was designed to yield faster signals. Depending on the phase of the sample clock with respect to LHC collisions, well over 90% of the signal energy can be contained in a single time sample; in fact, the forward detector channels were phase aligned to accomplish exactly this result. This phase alignment is described below. The purpose was to avoid the generation of"pre-triggers" from the forward channels.

As long as the trigger timing is synchronized with the rest of CMS, there would be little else to motivate the measurement of HF signal times more precisely than a time sample. The exception comes from PMT window interactions. Since all channels are calibrated for fiber light amplitudes, these interactions generate artificially high signal energies, thereby inducing fake missing transverse energy. The use of precise timing represented one potential handle for the rejection of this noise.

At the same time the arrival of collisions at the LHC represented the first opportunity to time-align all of the HF channels in-situ. Other portions of the HCAL had been aligned previously with so-called beam splash events, but the position and geometry of the HF cells, as well as the high flux of muons hitting the PMT windows and fiber bundles, prohibited the use of splash data for such a purpose with HF.

An HF scan campaign was therefore undertaken at the start of 7 TeV beam commissioning, to time-align the channels and to investigate ways to leverage timing as a means of PMT interaction rejection. Since the precise timing depends on the relative energy sharing between adjacent time samples, it would be necessary to find the per-channel sample clock phase settings that would place the nominal signal at the boundary between two adjacent time samples, represented by an even 50-50 energy sharing.

Data were acquired in intervals of 15-20 minutes, time enough to collect sufficient statistics at each phase setting. Then data acquisition was paused to load new settings to the front-end electronics. Online Data Quality Monitoring was used to estimate when optimal energy sharing was achieved for most channels. The data were then analyzed offline to determine for each channel at which phase setting the 50-50 sharing crossover occurred. This point represents the point in phase space with maximally sensitive time measurement capability. Setting all channels at their respective maximal sensitivity simultaneously synchronizes the channels.

One question to be answered by the phase scan was whether the magnitude of the jitter in signal times and PMT interaction times is too large to resolve them effectively. The next step was therefore to apply the alignment settings and acquire data at this setting with maximal timing sensitivity, to measure the signal jitter. It was determined that the signal jittered with a spread of ± 2 ns, which is small enough to adequately resolve PMT interactions in time.

DRAFT

htbp

Figure 9: Timing distributions in HF for both collision events and PMT interactions.

htbp

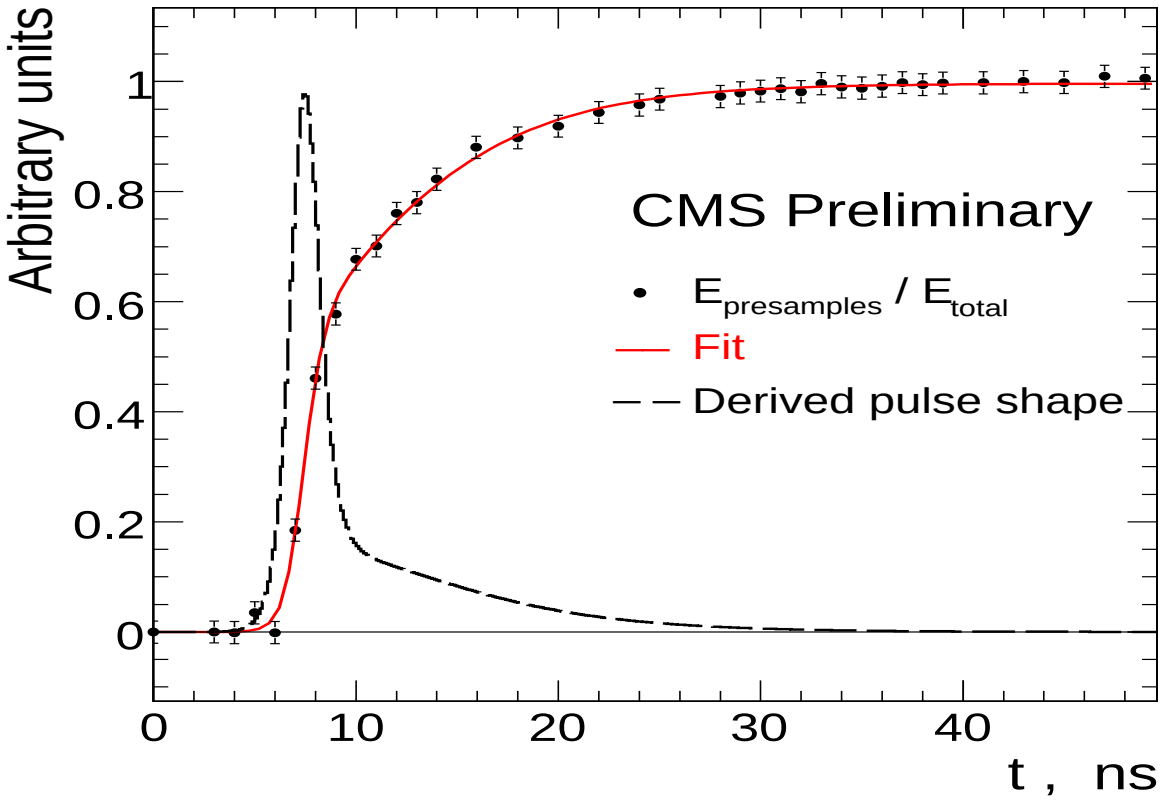

Figure 10: Pulse shape reconstruction in HF as a result of time scan measurements taken during the first 3.5 TeV collision runs. The data points represent the average fraction of the total energy within the ten-sample window that is contained in the first n samples, where n=4 for the first 25 ns and n=5 in the second 25 ns, as a function of front-end phase setting. The fit curve then represents the integrated pulse shape, and the dashed curve represents the differentiated pulse shape.

Since 50-50 sharing also represented an unacceptable perturbation of the trigger timing, the next step was to apply a global shift to the settings so that the signal for all channels occurred, with sufficient safety margin, just after the time sample boundary, while simultaneously remaining aligned. PMT window interactions are known to occur earlier in time, since the light from particle showers propagate in the quartz fibers at reduced speed. Setting the global phase in this way preserved the trigger while also affording the potential capability to detect and reject PMT interactions.

This intensive timing and synchronization effort for HF also provided the opportunity to assess the effectiveness of the precise time reconstruction algorithm for collisional signals in HF. This algorithm starts in the same way as that reported for the barrel and endcap in [12], with a first

order time estimate from a center-of-gravity technique using the three samples centered on the peak,

$$\text{Weighted peak bin} = \left[\frac{(p-1)A_{p-1} + pA_p + (p+1)A_{p+1}}{A_{p-1} + A_p + A_{p+1}}\right] \times C \quad , \tag{1}$$

where A_i represents the amplitude of an arbitrary time sample i, and p is the value of i such that A_p is maximum over the set of samples. In the case of multiple samples with the same amplitude, the earliest one is picked. The constant C is an amplitude-independent normalization constant that rescales the first order estimate to a range from zero to one. The weighted peak bin is then used to determine a second order correction that compensates for the asymmetry of the pulse shape, yielding the phase of the signal within the peak time sample.

The second order correction derives from a functional form of the pulse that was fit to data; at the start of the scan campaign the pulse was derived from test beam data taken in 2004. Since the actual front-end phase setting represented an independent truth time, it was possible to re-derive the pulse shape seen with collisions (Fig. 10). The correction function derived from the new fit was then used to investigate other timing effects and uses; one such additional effect was the determination of energy dependence on timing at low energies. Efforts are ongoing to optimize the algorithm for HF and to determine the limitations of the timing algorithm across all phase settings.

5 HCAL Trigger Performance

5.1 HCAL Triggers

The HCAL trigger is based on the HCAL calorimeter towers. The calorimeter trigger towers in HB/HE are typically comprised of physical calorimeter towers ganged together in depth. The exception to this rule are the $\Delta\phi = 10^\circ$ HE calorimeter towers, which are split in two trigger towers in ϕ, each of which is assigned one-half of the calorimeter tower energy, and the HF trigger towers.

The first 20 trigger towers in HB/HE up have constant $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. While the azimuthal granularity stays the same in the inner part of HE, the *eta* size starts increasing starting with tower 21 and reaches 0.35 for the last HE trigger tower 28. Altogether there are $17 \times 72 = 1224$ trigger towers in each half of the HB, and $11 \times 72 = 792$ trigger towers in each HE.

Each HF trigger tower is formed from 6 physical HF towers, with the long and short fibers ganged together. Thus, each HF trigger tower consists of 12 individual calorimeter inputs. There are 13 η rings of calorimeter towers in the HF mapped to 4 trigger towers in eta (29-32). The outermost 11 of them have ϕ segmentation of 0.175 radian, and for the two innermost one the segmentation is twice as coarse. The three outermost trigger towers in eta combine two adjacent towers in three adjacent eta rings. The innermost trigger tower contains two adjacent ϕ towers in the rings 10 and 11, and one tower each from rings 12 and 13. Thus, all the HF trigger towers have the same azimuthal size of 0.349 radian and the same $\eta = 0.5$ size. Altogether, each HF contains 72 trigger towers.

The HCAL trigger primitives are formed for each trigger tower by combining the information from individual calorimeter towers. Since the HCAL QIE's use 7-bit non-linear ADC's with several ranges, the energy reported by the QIE is first linearized as a 10-bit word, via a

DRAFT

dedicated look-up table (LUT). The linearization depends on several parameters, such as gain, conversion from the collected charge in fC to the ADC counts, and the pedestal. Since in general these parameters vary from one calorimeter channel to another, each input channel is linearized with an individual LUT, automatically generated from the database containing all the above information. These LUT's essentially remove the channel-to-channel variations and make linearized output to look the same for each group of calorimeter towers using the same type of LUT. In the HB/HE the linearized output for each calorimeter tower is proportional to the energy registered in this tower. This is the optimum linearization, as it allows to suppress noise (constant in energy) uniformly over the large η range spanned by the calorimeter. Since the conversion from energy to transverse energy, E_T, the quantity typically used in the trigger, depends on the position of the readout channel in the calorimeter, it is impossible to achieve both high resolution and large dynamic range in E_T with the single set of LUT's. Thus, the HB/HE is divided in three ranges in *eta*, each served by a different LUT: towers 1-20 in η have the LUT that linearizes energy with the highest granularity, with the least significant bit (LSB) corresponding to $\sim$ 0.2 GeV; towers 21-26 have twice as coarse granularity and expanded dynamic range in energy; towers 27-28 have 5 times as coarse granularity as in HB and even more expanded dynamic range. as a result, we achieve the individual readout channel saturation above $E_T = 64$ GeV over the entire calorimeter and as the same time keep the LSB below the typical noise level in most of the HB/HE.

Since the pulse in the HB/HE spans more than one time crossing, the trigger primitive (TP) is more complicated than a simple sum of several linearized calorimeter readout channel energies. One also must perform a temporal sum over the two adjacent bunch crossings, which contain about 95% of the calorimeter pulse energy, and look for peaks in the time sequence, indicating a signal. The peak-finding is done by requiring the running sum of energy in a certain bunch crossing and the one after it to exceed that in the previous one and to be at least as large as in the subsequent bunch crossing. If the condition is satisfied, a non-zero trigger primitive is generated for the bunch crossing corresponding to the leading edge of the pulse; otherwise it is set to zero. Consequently, most of the TP's in each event are identical to zero, which allows for efficient trigger data compression.

In addition to the TP generation in HF, a fine-grain bit is also generated if at least one of the six sums over the transverse energies in the long and short fibers of the same HF tower in any HF trigger tower is greater or equal to a certain threshold (currently set to 0.125 GeV). This bit is used for minbias trigger generation.

5.2 HCAL Trigger Primitive Validation

Since HCAL trigger is fully digital, it is possible to emulate its performance precisely. In order to validate the trigger, we run the emulation on the trigger data and compare its results with the actual trigger primitives generated. If everything works fine, the two must agree exactly. The comparison of the emulator output and actual TP's for laser calibration runs are shown in Fig. 11. As expected, we have perfect agreement between the hardware performance and the emulation.

It is somewhat harder to validate TP's in collision data, as it is zero-suppressed (ZS). In case of zero suppression the information available for the trigger emulator misses towers with energies below the threshold. The comparison of the emulator output and actual TP's in collision data are shown in Fig. 12. As expected, the correlation is not perfect due to ZS, but still very good.

DRAFT

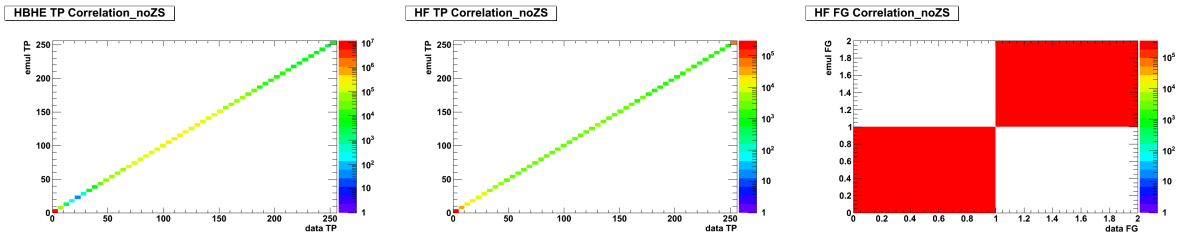

Figure 11: Results of the emulation vs. generated trigger primitives in: left — HB/HE TP's; center — HF TP's; right — HF FG bit. The values on the x-axis correspond to results of emulation; the values on the y-axis are TP's generated by the hardware. Results are based o n calibration runs 136658 (HB/HE) and 136869 (HF).

DRAFT

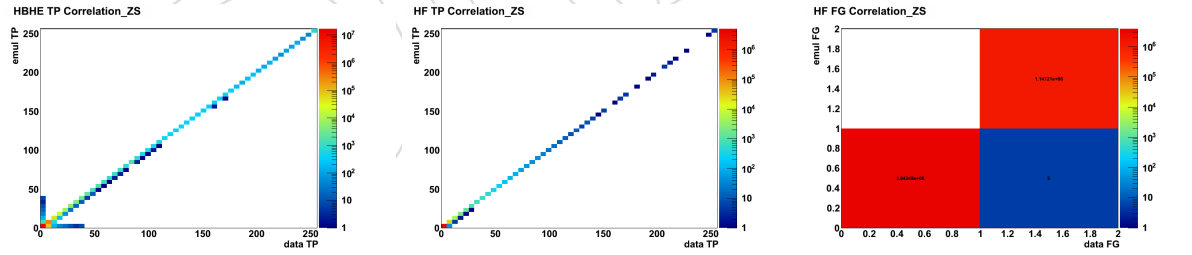

Figure 12: Results of the emulation vs. generated trigger primitives in: left — HB/HE TP's; center — HF TP's; right — HF FG bit. The values on the x-axis correspond to results of emulation; the values on the y-axis are TP's generated by the hardware. Results are based o n 7 TeV collision run 137027 with zero-suppression enabled.

6 Detector and Data Quality Monitoring

Data to monitor the performance of Hcal are collected under different conditions. Special runs are taken between stores when no beam is circulating in order to monitor noise rates and detector response. A fraction of the collision data is collected in real time and monitored as part of the Online DQM. Additional online data is read out during orbit gaps in ongoing runs to monitor detector stability. Finally, as finished runs are processed and reconstruction, offline DQM is run on the results in order to determine the Data Certification status for the runs.

6.1 Data Quality Monitoring Tasks

The Hcal DQM software provides a number of checks on data quality for each channel in the detector. Any channel that fails even one of these tests is considered problematic, and the overall Data Certification status value is lowered accordingly for each such channel found. Testing modules that are available within the Hcal DQM include:

- Raw Data Test – this performs checks on the format of the raw data received from Hcal. Corrupted data from the Hcal front ends are detected by this test.
- Digi Test – this checks the integrity of the digitized outputs ("digis") constructed from the raw data.
- Dead Cell Test – this checks for channels that never or rarely report valid digis or reconstructed hits (RecHits).
- Hot Cell Test – this checks for channels that report otherwise valid data, but which record larger-than-expected energies in a significant fraction of events.
- Trigger Primitive Test – this compares the Hcal trigger primitives reported by the trigger framework against expected primitives as emulated from Hcal input data.
- HF Luminosity Test – this performs additional data quality tests for those channels in HF used for luminosity monitoring.
- Pedestal Test – this compares the measured background from electronics noise fluctuations for each Hcal channel during a run to the expected level of noise.
- Laser Test – this checks the response of Hcal channels to input signals of known energy, as provided by the laser calibration system.

Each of the above tests produces a set of four histograms, indicating the number and rate of problem channels in each Hcal depth (see Figure 13). The results from these histograms are combined over all tests to form an overall "reportSummaryMap" indicating the overall health of the Hcal system. This map is used for quick monitoring of the Hcal system by DQM shifters and for a first determination of the data certification status for the run.

Additional diagnostic histograms that don't feed the reportSummaryMap are also produced by the DQM. These include plots of the energy and timing distributions of reconstructed Hcal hits, monitoring plots of noise rates during the run, and checks on the zero suppression algorithms used by Hcal. Some plots are trigger dependent, filling only when a Minimum Bias or Hcal coincidence trigger fires, and are thus useful indicators of the overall beam status.

The Hcal DQM is set up within a flexible framework, so that additional plots can be easily added upon request. However, constraints on the speed, memory usage, and size of the DQM output limits the total number of histograms that may be added to the monitoring system.

DRAFT

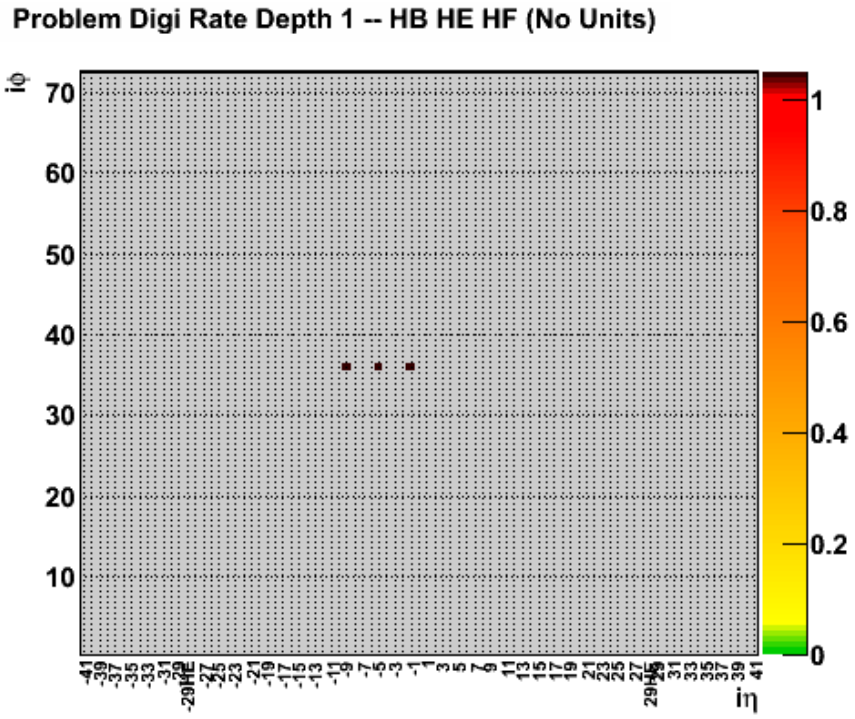

Figure 13: Problem channels detected by the Hcal DQM Digi Monitor task in run 137022. The depth 1 plot shows three known dead HB channels from a broken fiber.

6.2 Online Data Quality Monitoring

The online Data Quality monitoring system receives a prescaled set of data from the "all" stream and from the calibration stream. (Calibration stream data includes data from pedestal and laser events recorded during orbit gaps.) The Hcal DQM runs on this subset of events, and produces the monitoring and diagnostic histograms described above. Shifters are able to observe these plots in real time, and problems observed in the DQM are immediately reported to Hcal experts. This allows to quick responses to sudden changes in the detector during the run (e.g., RBXes that stop sending data, channels that suddenly appear noisy, etc.).

Because the online DQM stream only receives a subset of all calibration events, many laser pulse calibration events are not seen by the standard Hcal DQM. It was found that the rate of laser events received in the DQM stream would not be large enough to provide meaningful real-time monitoring of the Hcal detector response. Increasing the rate of laser events into the online stream would mean increasing the number of laser pulses beyond the allowed pulse budget and decreasing the fraction of online DQM events from actual collision events. Neither of these outcomes was desirable, so instead a separate data stream was set up that accepts only unprescaled Hcal calibration events.

This calibration data stream feeds a duplicate online Hcal DQM process, referred to as "HcalCalib". This process monitors the stability of each channel's pedestal and laser response during global runs. The pedestal and laser responses may also be checked during special "local" runs when Hcal is not included in global data taking. The HcalCalib DQM runs in a standalone mode during these local runs, and produces the same monitoring plots that are produced during global runs. The HcalCalib DQM also produces additional monitoring plots from local LED calibration runs.

6.3 Offline Data Quality Monitoring

The offline DQM runs the same Hcal DQM framework as the online DQM. However, a few tests that are run in the online DQM are disabled in offline DQM. The offline DQM does not have access to emulated Hcal trigger primitives or to Hcal calibration data. Thus, the monitoring of trigger primitives, pedestals, and laser data is not performed on offline data.

Offline DQM runs both on the "express" stream data and the full dataset as they become available. Extra monitoring of Hcal noise that is not done in online DQM is performed on these data due to the higher statistics available in offline DQM. The express stream contains a larger fraction of high-energy jet triggers than either the online stream or the all stream, and thus contained a biased contribution from Hcal HPD/RBX noise. Care must be taken when using the express stream for data certification not to mark the Hcal subdetectors as bad

7 Data Certification

Determining the usable fraction of data is essential for any experiment. For large scale experiments a well-defined data certification process is needed that combines information from each sub-detector and physics object group. Since a stable run could take hours, it is important to have the certification information at a fine grained level. Data certification identifies impermanent or new problems in hardware, software and conditions. The goal is to have the significant fraction of the certification performed in an automated way to be able to cope with very long periods of data taking and to be able to identify the *bad* portions of each dataset quickly. This should also allow the user to select and use only the *good* parts of the data in an easy fashion. Certification needs to be able to determine 'good' portion of each dataset at the lumi-section level [1] (i.e. 23 seconds). There are several sources of information that feeds data certification and each piece of information must be judged and the quality of the data has to be determined. *DQM (Data Quality Monitoring)*, *run-registry*, automated run certification information from *DCS (Detector Slow Controls)*, *DAQ (Data Acquisition)* and *DQM*, and information from operation and detector performance experts need to be taken into account in this process. The details of the CMS central *DQM*, *run-registry* and certification can be found in the references [6–8]. The overall HCAL data certication workflow is shown in Figure 14 .

Run-registry is a database that contains information on all *significant* runs that might be used by physics analyses. The information kept in the *run-registry database* is accessible through a web interface which also allows to enter quality flags and production of *certified* run and lumi-section lists. Run registry contains separate quality flags obtained from manual and automated certification for each data set. Manual quality flags of the *run-registry* are provided by the central online shifter at P5 and the offline shifter at the CMS Center at CERN Meyrin site. The automatic run certication combines the results from DCS, DAQ, and DQM in which each component is assessed independently from each other. The results of DQM are summarized in the so-called Report Summary Values which are formed from the hot cells, dead cells, data format errors, bad digis, pedestal and laser monitors. The report summary values show the goodness of each subdetector of HCAL; HB, HE, HF, HO, HO0, HO12. HCAL is considered *good* if the fraction of cells without problems is greater than a predetermined value. The DAQ, DCS, and DQM, and the certication results combing these three are written to run-registry and to DQM root files. The DAQ information at this stage is irrelevant to HCAL certification since it is duplicated in DQM summary values.

[1] It is also important to note that some portion of data is filtered at the event or channel level during reconstruction (see Section 7).

DRAFT

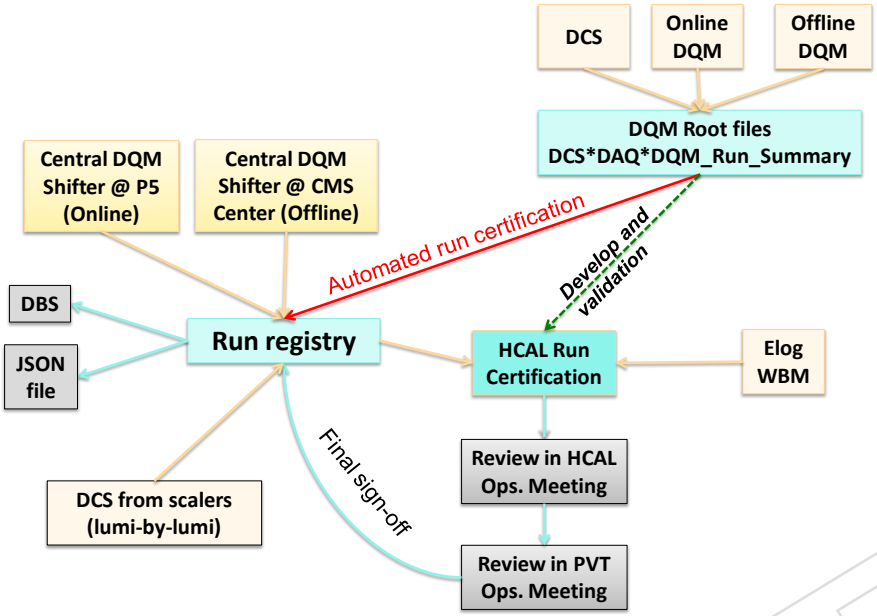

Figure 14: Overview of DQM and manual and automated run certfication framework.

DCS information could not be used extensively until now, however, with the recent developments, it started to become a part of the data certication. Run registry currently contains a list of bad lumi sections for hardware partitions of each sub-detector. Any hardware error reported from the DCS system causes the corresponding lumi section to be *bad* for the corresponding sub-detector partition. Also, a procedure is setup to automatically record high-voltage values that deviate beyond preset tolerances, correlate them with corresponding run numbers and lumi sections, and write them to database tables. This information is transferred to the offline database by automatically every hour. During the harvesting stage, offline DQM will read these values from the offline database, and combine the decisions for each subdetector of HCAL with report summary values and DAQ values and produce a single certification value for each subdetector per lumi section.

7.1 HCAL Run Certification Procedure

Status values are taken from the HCAL DQM results, and reflect the fraction of channels in each subdetector that report valid data. The following checks must all be passed for an HCAL channel to be considered valid:

- Reconstruction checks: Raw data from the channel must be able to be *unpacked* into *reconstructed hits* without encountering errors in timing, size, or hardware performance;
- Dead cell check: *Reconstructed hits* from the channel must be present in at least a minimum fraction of overall events;
- Hot cell check: The number of *reconstructed hits* above a threshold energy from the channel must be less than a given fraction of total events, and the average energy seen in 'quiet' (orbit gap) events must be within tolerances.

There are three categories of flags which are used as quality flags; *good, bad, excluded*. The

HCAL quality flag is set to 'GOOD' if the following conditions for the four HCAL subdetectors (HB,HE,HO0,HF) are met: HB status $>$ 0.98, HE status $>$ 0.98, HO0 status $>$ 0.95, and HF status $>$ 0.98. For some special cases, a run is still usable if its quality flag is not set to *good*. For example, if a run is marked *bad* due to HF problems, that run would still be usable by analyses that do not use HF or depend on missing transverse energy. Automatic quality flags allow separate judgments to be made and stored for each HCAL sub-detector. When HCAL is not in readout, the run is marked as *excluded*.

GOOD status criteria might change in time. Certification criteria are entered manually in the central *HcalDataValidation twiki page*[10] and run-registry database since the overall criteria changes very rarely. However, if the criteria needs to be changed more frequently, then a better bookkeeping is needed.

7.1.1 Manual Certification

Significant runs are created and the corresponding quality flags are entered into the run registry database by the central online DQM shifter who inspect DQM histograms and communicate with the shift leader, DCS, DAQ and sub-detector shifters and experts. The runs entered in the run registry is also checked by offline DQM shifters before individual subdetector experts check the flags. Following the instructions provided by each sub-group, the shifter enters the quality flags along with her/his observations and specific comments. Manual section of *Run-registry* allows only one quality flag per sub-detector for each run unlike the case in the automatic certification which is described in the next sub-section. The quality flags are entered in the run registry database separately for online and multiple offline datasets including re-reconstructed datasets.

Run-registry database also keeps a list of bad lumi sections for each partition of each subdetector using the DCS information obtained from scalers. This will be discussed in more detail in the next section. As part of the certification procedure the *bad* lumi-section lists of HCAL in run registry are also checked.

For certification of HCAL data, the summary values for online and offline datasets are also checked. More thorough checks are made for the runs that have summary values below the threshold values for each HCAL subdetector. When more detailed analyses are needed, the observations are communicated to other HCAL performance group to investigate the problem.

To get more complete information for each run, the information provided by the HCAL shifter is also checked. For this purpose, a web interface through the web-based monitoring tool, WBM [11] is used to access the information that is logged into a database (*OMDS*) by the HCAL shifter. The *comments* that are stored in the database by the shifter are particularly important since they provide first hand information from the control room that could not have been obtained automatically.

Combining all the certification information described above, the HCAL certification team determines the goodness of each run taken during one week of data taking for each dataset and checks the consistency of this decision with the HCAL quality flags in run-registry.

Using the manually run registry flags and the lumi-by-lumi flags in run registry, HCAL is found to be 98.5% good in the run range 132440-136297.

7.1.2 Automatic Certification

Automatic certification is performed in the offline DQM including re-reconstructed data. Automatic certification combines DQM, DCS, and DAQ summary values for each HCAL component

DRAFT

(HB,HE,HO0,HF) for each lumi section. HCAL DQM produces a simple count of bad channels using the combined summary value at the end of each lumi section using dead cell lumi-by-lumi, gidi, and raw data format histograms.The produced are normalized by the number of events and number of channels at the end of each lumi section. Automatic certification is still in development and it is planned to be completed in a few months.

7.1.3 Weekly Sign-off of Certification Results

The quality flags and additional checks and findings by the HCAL certification group are presented to the HCAL operations community every week to be able to commission the certification procedure and to understand any remaining issues. Then, changes needed for run registry quality flags and DCS lumi-by-lumi flags are communicated to the central DQM team. The central DQM team collects certification information from each sub-detector, and then these are *signed-off* in the weekly *Physics Validation Team (PVT)* meetings. After the sign-off of the quality flags, a so-called 'JSON' file is produced for each dataset which lets analyzers to select only *good* lumi-sections in their analyses. The quality flags are also stored in the offline condition database and also the dataset bookkeeping system (DBS) [9].

8 Anomalous Signals in HCAL

Anomalous signals in HCAL have been observed while calorimeter modules were exposed to test beams and during early commissioning with comic data [13]. Large reconstructed energies were observed resulting from electronics noise as well as beam related effects. Figure 15 shows the distribution of the number of channels within an RBX having energy above 1.5 GeV and requiring that the total energy of the channels within the RBX is greater than 10 GeV. The events were collected by triggering on energy deposits in the calorimeter and occur randomly. This type of anomalous signal is collectively referred to as RBX/HPD noise. Several clear categories of noise can be observed. Large signals are occasionally observed in the HCAL HPDs even when no light is incident on the photo-cathodes. These signals are thought to originated by a thermally emitted electron which ionizes the gas or surface molecule in the acceleration gap of the HPD which in turn is accelerated back to the cathode liberating additional electrons. Ion feedback typically affects 1 HPD channel. Dielectric flashover from the wall of the HPD can produce large signals in up to 18 channels of the HPD. Electronics noise from the RBX can produce large signals in up to 72 channels. the source of this type of noise is currently not understood andis under investigation. Energy in the HB and HE is reconstructed from 4 time samples of 24 ns and the chance of an overlap of the random RBX/HPD noise within this 100ns window is small.

The rate of noise is continuously monitored and is seen to be relatively constant. Ion feedback and HPD discharge appear relatively constant while RBX noise is seen to exhibit a greater variation.

Anomalous signals were observed in the HF when modules were exposed to test beams. The dominant source of the anomalous signals is due to a charged particle producing Cerenkov light in the thick glass window of the HF PMT. A contribution is also observed from charged particles traversing the fiber bundle behind HF.

The anomalous signals have distinct properties which can be exploited and filters developed to remove the energy from these events. A summary of the filters that have been developed is provided in the following sections. There are two approaches to filter noise. The energy reconstructed in channels having anomalous signals can be removed during the event processing

DRAFT

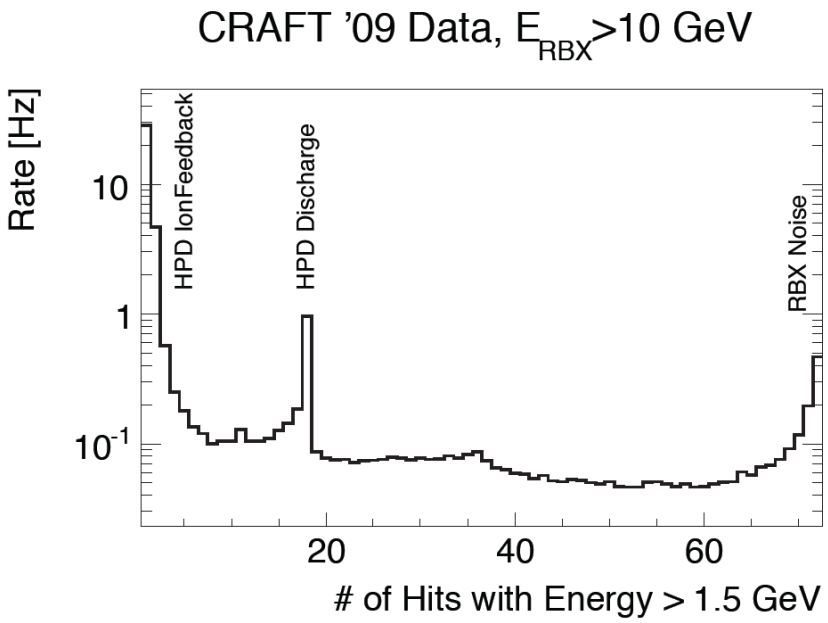

Figure 15: Number of channels within a RBX above threshold.

so that the flagged rechits do not contribute to the reconstructed physics objects. Since rechit filtering affects all users it is important to remove only well identified noise. As the event complexity increases it will become more difficult to use properties such as isolation to identify anomalous signals and more sophisticated filtering can be applied during the offline analysis level.

A flexible software framework was developed to provide the ability to remove rechits during data reconstruction, apply event filtering by the user, and to apply rechit cleaning by the user after the data has been processed. The energy for the calorimeter cells is reconstructed from the digitized data (digis) which is recorded in 10 time samples each having a 25 ns width. The digitized data is reconstructed as energy (rechit) and saved in the RECO (reconstructed) data format. The RecHit is a data structure containing energy, time, and flags which indicate the quality of the reconstructed quantities. Thresholds are applied to the rechit energies and the rechits are combined into a calorimeter tower structure having an electromanetic and hadronic component. Physics objects such as jets and missing E_T are reconstructed started from the CaloTower. Filter can be applied to remove known sources of noise at the stage that the CaloTower is created producing what is referred to as "cleaned CaloTower". One important limitation is that the full digi information is not available in the RECO data format. The implication is that filters that use digi information, such as the pulse shape, cannot be applied by the user on RECO data.

Additional filtering can be applied at the analysis level on top of the default filtering to create a custom cleaned CaloTower. This requires that any physics objects that are derived from CaloTower be recreated using the custom cleaned CaloTower. This flexibility allows users to develop and test filters which can then be applied by default during a subsequent reprocessing. It also allows users to take advantage of the most up to date cleaning without having to wait for the data to be reprocessed. Different analysis may require more aggressive cleaning which also can be applied at this stage.

The HCAL DPG has developed a set of baseline cleaning algorithms that are applied during the processing stage. Table **??** summarizes the algorithms and the details are provided in the following sections.

8.1 HF Noise

8.1.1 Introduction

Commissioning studies performed with data from the past test beams, cosmic runs, and early LHC beam runs have identified uncharacteristic noise (i.e. noise not produced solely from expected fluctuations in the electronics) in the forward hadronic calorimeter (HF) [13]. Such anomalous signals are caused by relativistic charged particles directly impinging upon the window of an HF photomultiplier tube (PMT), generating Cerenkov light, and thereby producing an abnormally large apparent energy signal for the single HF channel associated with that PMT. Due to the nature of the signal generation, the energy spectrum of such noise is relatively well defined, with a peak at an energy of $E \approx 100$ GeV and pronounced tails at higher energy values. Since the HF detector occupies the forward region of CMS, the bulk of the transverse energy (E_{T}) spectrum of such noise is mostly constrained to relatively low values, e.g., an energy of 100 GeV corresponds to $E_{\mathrm{T}} = E/\cosh(\eta)$ of ≈ 10 (1.3) GeV at $\eta = 3$ (5). Nevertheless, such anomalous signals can occasionally produce fake jets with high p_{T} and large fake in the event. The application of the HF noise cleaning algorithms, described in this section, is therefore recommended for any physics analysis which employs energy deposits reconstructed in HF and/or . A fully comprehensive description on the HF noise cleaning algorithms and the complete set of results on their performance can be found at [ADD HERE NOTE NUMBER OF HF DETECTOR NOTE].

8.1.2 Handles to identify anomalous signals

PMT hits are typically characterized by a large apparent energy in a long (short) fiber and very little or no energy in the short (long) fiber in the same HF tower; in addition they are usually more "isolated" than real energy deposits, i.e. very little or no energy is observed in the surrounding HF towers. The topology of PMT window hits can be compared against expected longitudinal and lateral shower profiles in HF. PMT hits are typically characterized by a large apparent energy in a short fiber and very little or no energy in the long fiber in the same HF tower, as shown in Figure 16. Therefore, a simple ratio

$$R = \frac{E_{\mathrm{L}} - E_{\mathrm{S}}}{E_{\mathrm{L}} + E_{\mathrm{S}}}, \tag{2}$$

where E_{L} and E_{S} are the respective energies of the long and short fiber RecHits in a given HF tower, can be used to identify the short fiber PMT hits. The PMT hits in short fibers will have $R \approx -1$.

The pulse shape of the PMT hits, originated by the Cerenkov light produced in the PMT window, is expected to be almost fully contained within a single 25-ns time window, as are the pulses of signals generated by particles showering in the HF absorber. However, the PMT hit signal is expected to arrive earlier in time (by several ns) with respect to the signal from real energy deposits. This is due to the fact that Cerenkov light produced in the HF fibers by particle showers requires additional time to reach the PMTs (quartz fibers have an high index of refraction, $n = 1.458$, and varying length that increases with increasing η).

Figure 17 is a scatter plot of the ratio R, defined as $R = (E_L - E_S)/(E_L + E_S)$, vs the reconstructed time of both long and short fibers with energy E_L and E_S, respectively, above 90 GeV,

DRAFT

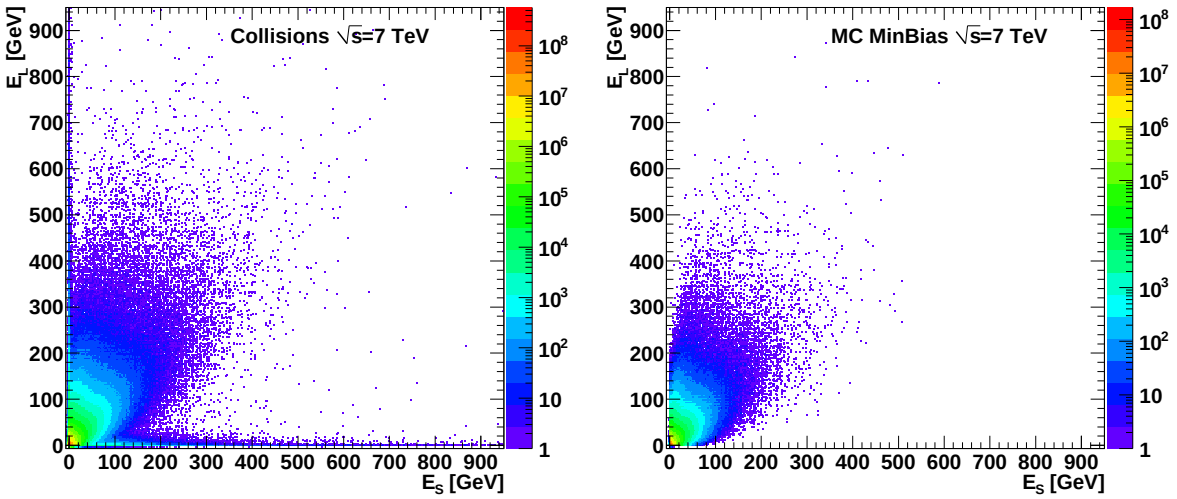

Figure 16: Long fiber RecHit energy vs. short fiber RecHit energy from the same HF tower in 7 TeV collision data (left) and 7 TeV MinBias Monte Carlo simulation (right). High energy entries close to either of the two axes only present in collision data (left plot) are the HF PMT window hits.

from a sample of 7 TeV collision data. High energy deposits that have R value close to -1 or 1 (corresponding to energy present only in short or long fibers, respectively) and/or that are early in time (approximately $t < -10$ ns), can be potentially flagged as PMT hits. Figure 17 indicates that also late hits (approximately $t > 10$ ns) are present in collision data. The exact source of late hits is not yet fully understood. However it can be noted that most of late hits have also topological properties consistent with some type of anomalous signal (R close to -1 or 1).

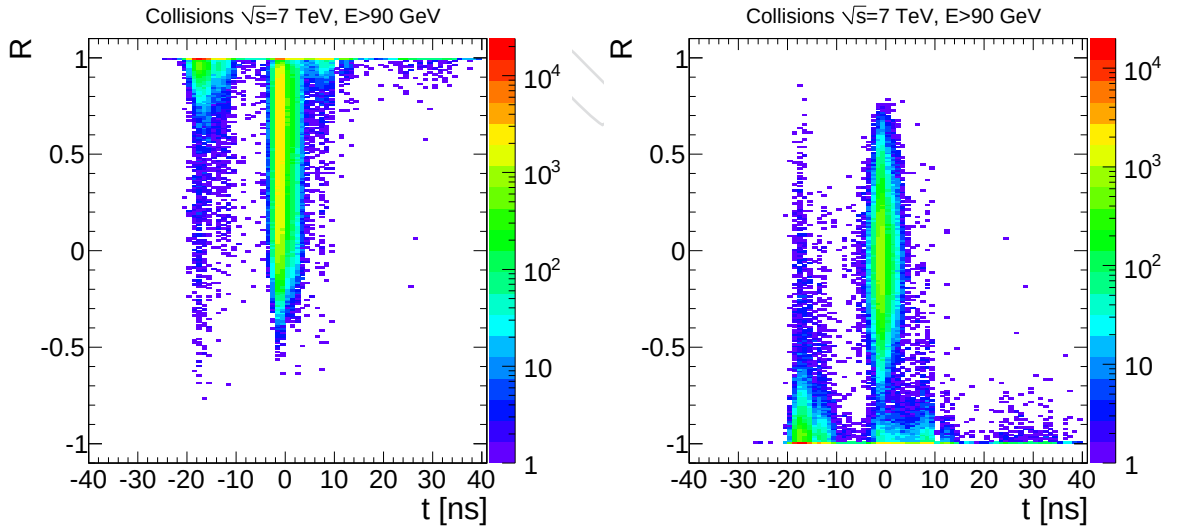

Figure 17: R ratio vs. reconstructed time for long (left) and short (right) fiber reconstructed hits with $E > 90$ GeV in 7 TeV collision data.

8.1.3 Description of HF noise cleaning algorithms

Based on these considerations, the main handles to identify the HF PMT window hits (and anomalous HF signals in general) are the topology of their energy depositions in η/ϕ/depth space and the pulse shape/timing of the signals.

The topology of PMT window hits can be compared to the expected longitudinal and lateral shower profiles in HF. Two different algorithms are used for short and long fibers, respectively:

- the Polynomial Energy Threshold (PET) algorithm flags a reconstructed hit in a short fiber as a PMT hit if its energy, E_S, is i) above some energy threshold, and ii) very large compared to the long fiber energy, E_L, in the same HF tower. The energy threshold is a polynomial function of η, ranging from about 35 GeV (at $\eta = 3$) to 50 GeV (at $\eta = 5$). If this threshold is passed, and the ratio $R = (E_L - E_S)/(E_L + E_S) < 0.8$, the cell is identified as a PMT hit;
- the S9S1 algorithm, employed for long fibers, allows to identify the PMT hits by comparing the energy in a long fiber to the sum of energy in 9 of its neighbors (4 long fibers and 4 short fibers in the adjacent HF towers, plus the short fiber of the same HF tower). The isolation variable "S9S1" is defined as a ratio between the energy sum of the 9 neighbors, S9, and the energy of the long fiber cell under consideration, S1. If this ratio is smaller than a threshold, defined as a function of long fiber energy, the cell is flagged as a PMT hit.

The S9S1 algorithm is found to be less efficient at identifying PMT hits (mostly at $E_T < 10$ GeV) with respect to PET algorithm. The use of the S9S1 algorithm is motivated by the need of making the noise cleaning in long fibers safe for real energy deposits coming from isolated photons and electrons. In fact, high energy photons and electrons hitting HF can occasionally deposit large fraction of their energy in the long fibers and very little energy in short fibers at a given (η,ϕ) location, thus faking the PET signature of an isolated PMT hit. While PET algorithm can incorrectly flag such events, the S9S1 algorithm is safe since it takes into account that electrons and photons will deposit some energy in the adjacent cells as well. The HF ring at the smallest η ($i\eta = 29$, i.e. $\eta = 3$) represents an exception, since it's partially shielded by the HE material, and it receives very little energy coming from the interaction point. For this reason, the PET algorithm can be safely used to identify PMT hits occurring in the $i\eta = 29$ ring.

In addition to the topological cuts, the pulse shape/timing information of the signals are used to identify PMT hits in both long and short fibers. In order to employ the expected timing difference between regular signals and PMT hits, a precise phase alignment in HF is required. A phase scan in HF was performed in April 2010 with early 7 TeV collision data and the timing phase was adjusted in the hardware settings so that the PMT hits show up in an earlier time sample than the normal energy deposits. More details on the HF phase scan are discussed in Section 4.2.

As a result of the timing adjustment, real signals are expected to peak in time slice 4 for all channels in HF. The fast response time of the PMTs means that the bulk of the charge for real HF signal pulses will be recorded in this single time slice. PMT window hits differ from real signals in that they generally arrive earlier or later than real signals. Even window hits arriving within the expected signal time range typically have broader distributions than real signals. The shape of the HF pulse can thus be used to distinguish real signals from PMT window hits. For each HF hit reporting an energy of at least 40 GeV, the charge deposited in time slice 4 ($TS4$) is compared to the total charge in the larger time range between time slices 3 and 6 ($\sum TS3-6$).

An HF hit is flagged as noise if the following condition is met:

$$\frac{TS4}{\sum TS3-6} <= 0.93 - exp(-0.38275 - 0.012667 * E), \quad (3)$$

where E is the energy of the hit. This cut flags both early and late hits, as well as in-time hits that report significant charge outside the expected signal time range.

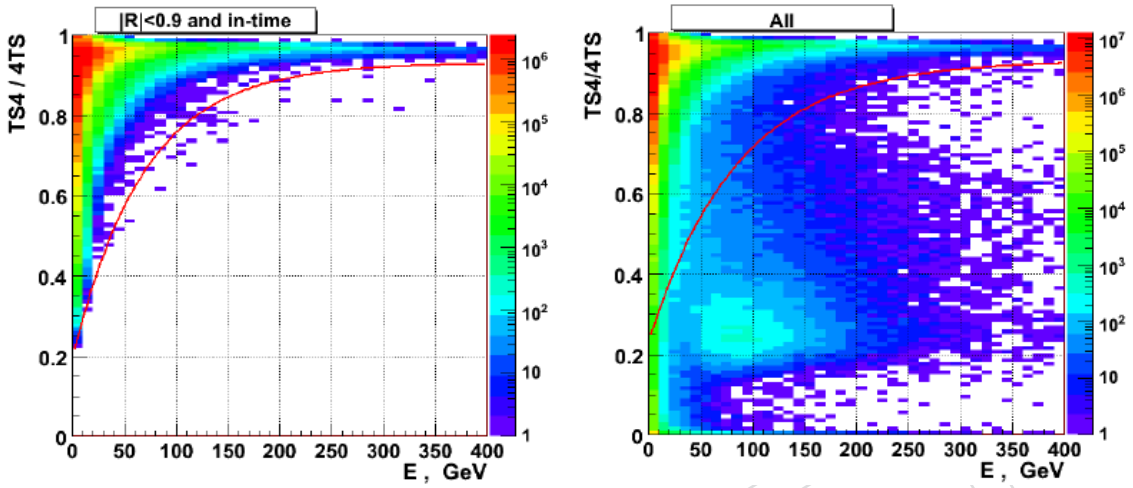

Figure 18: Q ratio ("TS4/4TS") vs Energy for (left) in-time RecHits with $|R| < 0.9$ and (right) all RecHits . The line indicates the contour used to separate good energy deposits from anomalous signals.

Figure 19 shows the reconstructed time for all energy deposits in long and short fibers above 40 GeV, and for those flagged by topological and pulse shape cleaning algorithms, in 7 TeV collision data. It should be noted that i) the use of pulse shape cleaning improves significantly the efficiency in identifying PMT hits occurring in long fibers, ii) some energy deposits in long and short fibers are flagged by the topological cuts but they are reconstructed in-time and not flagged by the pulse shape algorithm. Therefore, the combination of pulse shape and topological cleaning provides a set of complementary criteria to identify anomalous signals in HF.

HF channels flagged by either the topological cuts (PET, S9S1) or the pulse shape algorithm are removed from the reconstruction of high level objects, such as jets and . It is relevant to point out that i) the reconstructed time/pulse shape for HF energy deposits from collision data is not optimally reproduced by the current simulation, ii) the HF anomalous signals are not included in the current simulation. For these reasons, the pulse shape cleaning algorithm is currently applied only to collision data. The topological cleaning is instead applied to both data and simulation.

8.1.4 Performance of HF noise cleaning algorithms

The topological cleaning (PET, S9S1) was applied to Monte Carlo samples of different physics processes (Minimum Bias, QCD multi-jets in different ranges of jet p_T, isolated photons in the HF acceptance in different p_T ranges) to evaluate the efficiency of the flagging algorithms on the simulated energy reconstructed in individual HF channels, as well as the impact on jets reconstructed in HF and MET. The probability to incorrectly flag a real energy deposit in HF with $E_T > 5$ GeV as an anomalous hit, is estimated to be less than 10^{-3} for the physics processes considered; this value is considered small enough to allow the application of this cleaning by default in the standard CMS reconstruction.

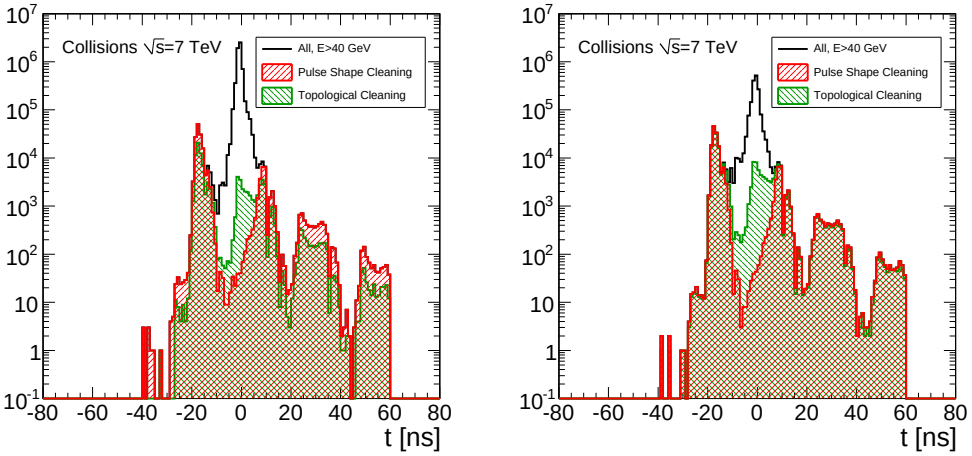

Figure 19: TO BE UPDATED WITH NEW PLOTS. Reconstructed time for energy deposits above 40 GeV in long (left) and short (right) HF fibers in 7 TeV collision data. Reconstructed hits flagged by topological and pulse shape cleaning algorithms are also shown.

The performance of the topological and pulse shape cleaning algorithms has been studied on a sample of 7 TeV pp collision data triggered by minimum bias triggers and jet/met triggers. An offline event selection to identify good collision candidates is applied, including the requirement of i) two proton bunches crossing in the center of CMS, ii) a good reconstructed primary vertex, and iii) the removal of data collected during periods with known detector problems. Beam related background events (including beam halo and beam scraping events) are removed by appropriate filters that use trigger and tracker information, respectively. Events with HPD/RBX noise in HCAL barrel and endcaps are removed from the analysis as described in Section 8.2. Finally, anomalous noisy channels observed in the ECAL barrel (so called "ECAL spikes") are flagged and removed from the jet and MET reconstruction using the recipe provided by the ECAL group.

Figure 20 shows the calorimeter-based MET before and after applying the topological and pulse shape cleaning algorithms in HF. It is observed that the noise cleaning algorithms strongly reduces the tails in MET distribution caused by HF anomalous noise. Figure 21 shows the cleaned calorimeter-based MET in data compared with a Monte Carlo simulation of Minimum Bias events; the good agreement between data and simulation confirms that the cleaning algorithms remove efficiently the HF noise, without affecting the real energy deposits from collisions. Some events due to residual HF noise are still visible in the tails of the MET. They have been visually inspected and classified as i) double-hits in the same HF tower, ii) multi-hits affecting several channels typically in the same ϕ strip, and iii) PMT hits embedded inside a jet. Such residual anomalous hits, not flagged by the existing algorithms and producing high MET (MET$>$ 45 GeV), occur at a very small rate of approximately 10^{-4} with respect to the total rate of identified anomalous hits in HF. Nevertheless, new criteria to identify this kind of noise are currently under study and will be presented in future updates of this analysis.

This section concludes with a couple of remarks on the characteristics of the PMT hits. One of the properties noticed in collision data is that the rate of HF PMT hits per event roughly scales linearly with the amount of energy deposited in HF, excluding the energy of the PMT hits themselves. This is shown in the left plot of Figure 22 for 7 TeV collision data, where PMT hits were identified using both topological and pulse shape algorithms. Based on this linear dependence and the mean $\sum E$ in HF coming from Monte Carlo simulation of any type of process, one can roughly predict the rate of PMT hits in such events. It can also be noticed

DRAFT

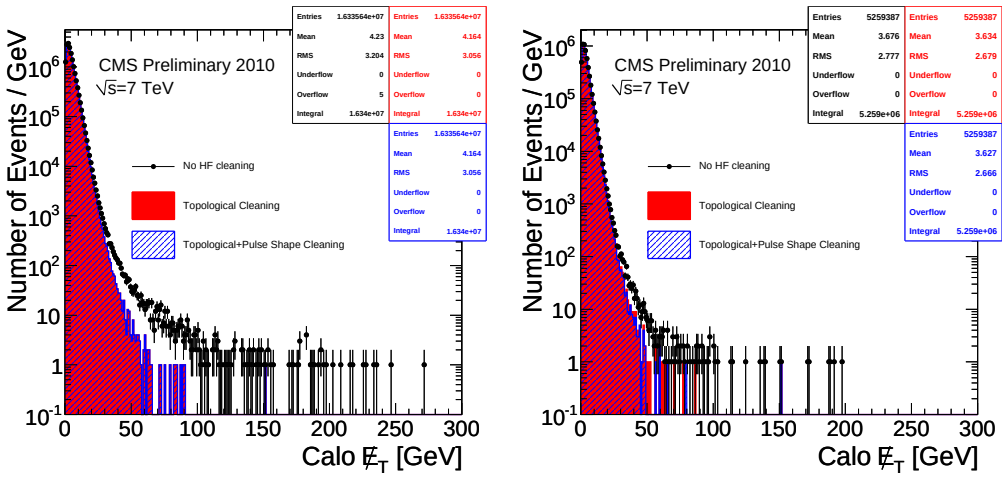

Figure 20: TO BE UPDATED WITH NEW PLOTS. Calorimeter-based MET in $\sqrt{s} = 7$ TeV pp collision data before and after applying topological and pulse shape cleaning in HF. The offline selection for good collision events, described in the text, as well as the ECAL barrel noise cleaning and the event filter for HPD/RBX noise are applied. The plot on the left represents the whole statistics of the data-sample analyzed; the right plot includes only events passing the minimum bias triggers. The number of events in the right plot is smaller than the left plot because the minimum bias triggers were prescaled.

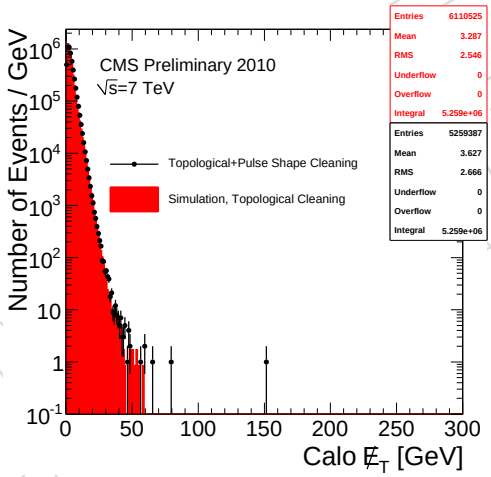

Figure 21: TO BE UPDATED WITH NEW PLOTS. Calorimeter-based MET in $\sqrt{s} = 7$ TeV pp collision data after applying topological and pulse shape cleaning in HF, compared with a Monte Carlo simulation of Minimum Bias events. The offline selection for good collision events, described in the text, as well as the ECAL barrel noise cleaning and the event filter for HPD/RBX noise are applied. Only events passing minimum bias triggers are considered for the comparison.

676 that, despite of the fact that HF PMT hits most frequently occur as single hits, cases of multiple
677 hits in a single event have also been identified, as shown in the right plot of Figure 22.

678 8.2 RBX/HPD Noise Event Filter

679 Event displays showing RBX and single HPD noise are shown in Figure 23. Thereare 18chan-
680 nels in aHPD and within a subdetector (HB or HE) towers having the same *iphi* are read out
681 by a single HPD. FourHPDs areconnectedtoone readout box (RBX) through 4 readout modules

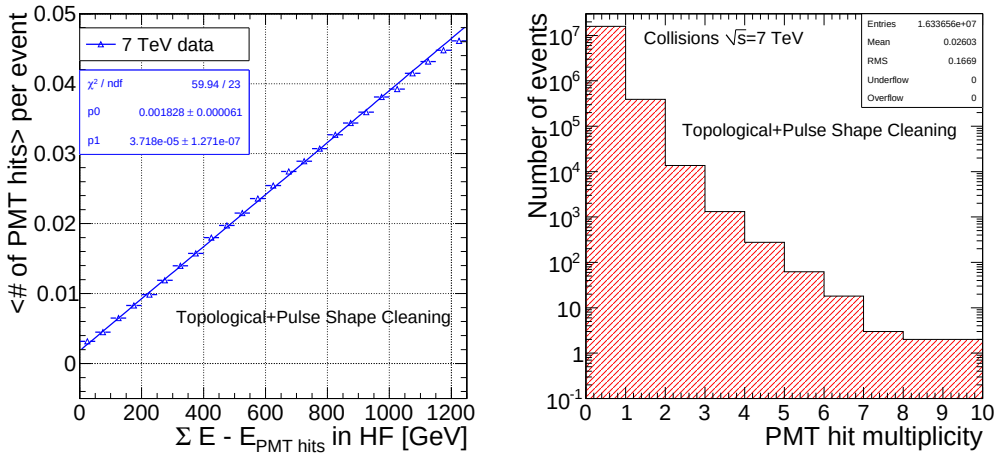

Figure 22: TO BE UPDATED WITH NEW PLOTS. Left plot: Average number of HF PMT hits per event as a function of $\sum E$ in HF, excluding the energy of the PMT hits, in 7 TeV collision data where PMT hits were identified using both topological and pulse shape algorithms. Right plot: HF PMT hit multiplicity in 7 TeV collision data obtained using both topological and pulse shape algorithms.

(RM).

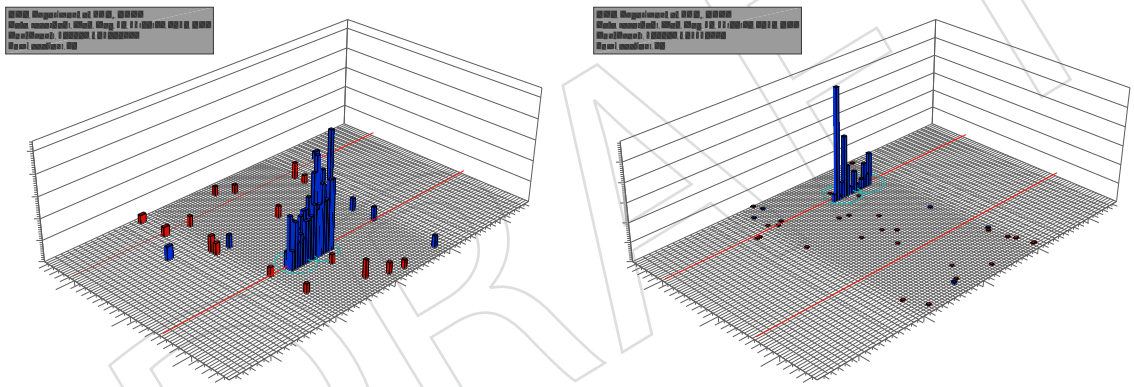

Figure 23: Characteristic pattern of energy deposited due to RBX noise (left) and HPD discharge (right). RBX noise can affect up to 72 channels of 4 HPDs being readout by the same Readout Box. HPD noise can effect up to 18 channels.

Some properties of RBX/HPD noise which can be used to identify noise includes the following.

- Large signals are observed for channels associated with either the RBX or HPD. This produces a distinct pattern of energy in eta, phi which is isolated and does not have any significant surrounding energy. The noise is random and the overlap with a collision is rare ($\sim 10^{-5}$). Typically there is little or no other activity from the other detector components.
- Since the noise is random, it is not in time with a physics collision. The rechit time or the pulse shape will be an effective handle to supress the noise.
- We observe a low rate of events that have an ADC=0. This is a characteristic of so called "Flash" noise where a large pulse will undershoot the pedestal resulting in

ADC=0.

Since the overlap of RBX and HPD noise with a collision event is rare it is possible to reject events identified as having RBX/HPD noise. The RBX/HPD noise event filter discussed here is recommended by the CMS HCAL Detector Performance Group for the summer 2010 analyses of pp collision data delivered by the LHC with 25 ns bunch crossing and at $\sqrt{s} = 7$ TeV.

The event filter removes events based on the following variables and requirements:

• **HPD Hit Multiplicity:** Events are identified as noise if there are more than 17 channels in an HPD having $E > 1.5$ GeV. Different settings for the threshold values (0.5-10 GeV) and number of HPD hits (14-18) were tried and the chosen settings were found to provide good noise rejection while maintaining a high signal efficiency.

• **Hot HPD:** The event is flagged as noise if a single HPD has $N_{HPD}^{hits} \geq 10$ with $E > 1.5$ GeV and there are no hits in the other HPDs within the same RBX (independent of the RBX energy). HPDs that satisfying the second condition are referred to as a "hot HPD". If there is at least one hot HPD in the event the event should be removed.

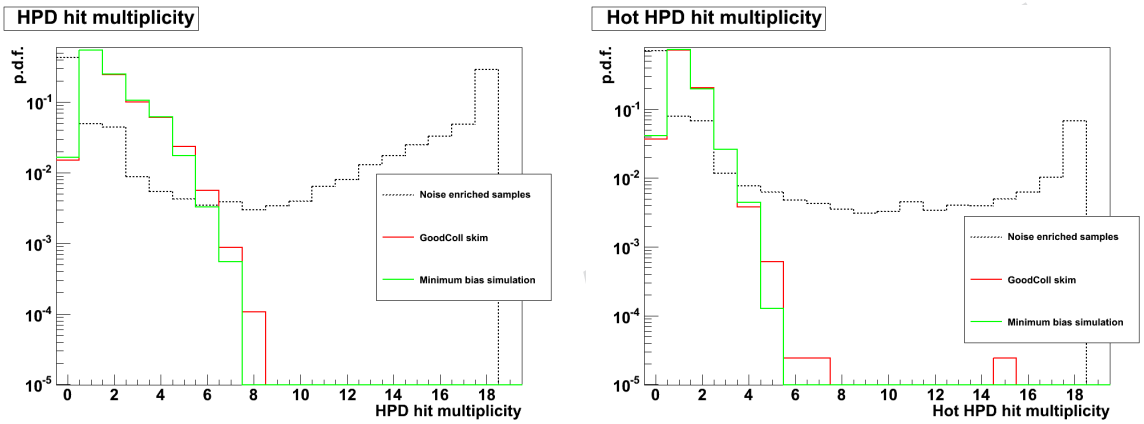

Figure 24: HPD Hit Multiplicity and Hot HPD hit multiplicity in all the CMS collected and Monte Carlo Simulated samples studied

Figure 24 shows the N_{HPD}^{hits}, N_{hotHPD}^{hits} for the default filter threshold and the after the baseline selection has been applied.

• **E2/E10:** The ratio E2/E10 is constructed for all channels associated with the RBX that have $E > 1.5$ GeV. E2 is the maximum of the sum of two consecutive time samples within the 10 TS used to digitize the data. E10 is the sum of the charge for all ten TS for all channels associated with the RBX that pass the requirement of $E > 1.5$ GeV. The E2/E10 distribution for $E_{RBX} >$50 GeV in the samples studied after the baseline selection is shown in Figure 25.

The event filter discards events as "RBX noise" if E2/E10$>$ 0.96 or E2/E10$<$ 0.7 and $E_{RBX} >$ 50 GeV. We observe that the MC simulation does not reproduce the pulse shape of the data and work is ongoing to improve the simulation.

• **ADC0:** One class of RBX/HPD noise that is observed is referred to as "RBX Flash Noise". Large pulses are observed that will undershoot the pedestal resulting in ADCs having 0 counts. A subset of these events are identified by the "HPD Hit Multiplicity" requirement, however there are events with lower channel energy that survive. The remaining noise events can be filtered using the a cut on the number of channels within a RBX observed to have ADC=0 (ADC0) This type of noise is not modeled in the Monte Carlo Simulation. The $N_{ADCZero}$ distribution af-

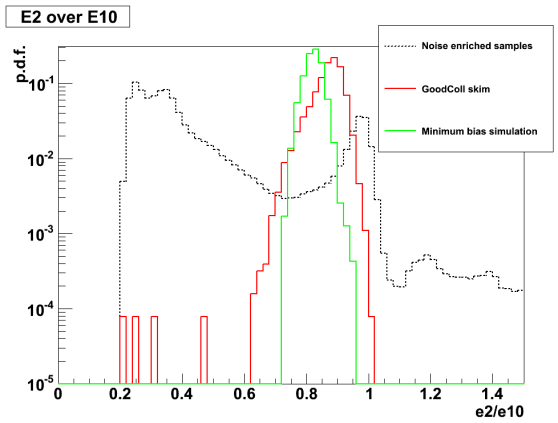

Figure 25: E2/E10 (for E_{RBX} >50 GeV) for the noise enriched data, good collision data, and several different MC samples.

ter the baseline selection is shown in Figure 27.

Figure 26: Pulse shape for channel with ADC=0.

Events are removed if $N_{ADCZero} > 10$ in an RBX with $E > 10GeV$.

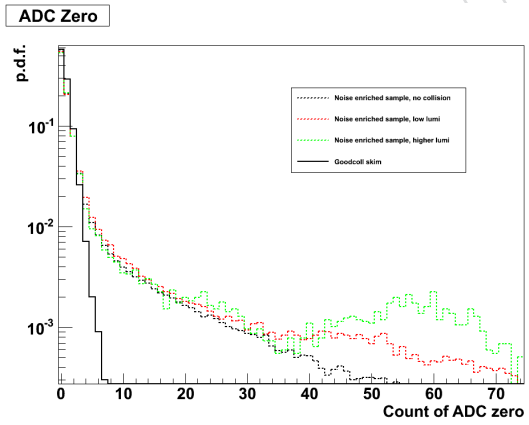

Figure 27: $N_{ADCZero}$ in all the CMS collected data samples

Events are dropped as noisy if $N_{ADCZero} > 10$ in an RBX with $E > 10$ GeV.

Figure 28 shows the rechit time distribution for the high MET sample, and a pure noise sample. The RBX/HPD noise is seen to be distributed across the enture 4 TS window as expected since it occurs randomly while the time of energy from collisions is grouped near 0. Sufficient collision data is needed to adjust the channel phase and a good understanding of the pulse shape is needed. Once this work is done we expect to have more sophisicated algorithms that use time or use the chi2 from the fit to the pulse shape in order to identify anomalous signals.

8.3 Performance of the RBX/HPD Noise Event Filter

In Table **??** we show the rejection efficiency of each of the requirements of the filter normalized to the baseline selection (i.e L1SingleJet10 trigger and at least one RBX with $E_{RBX} > 50$

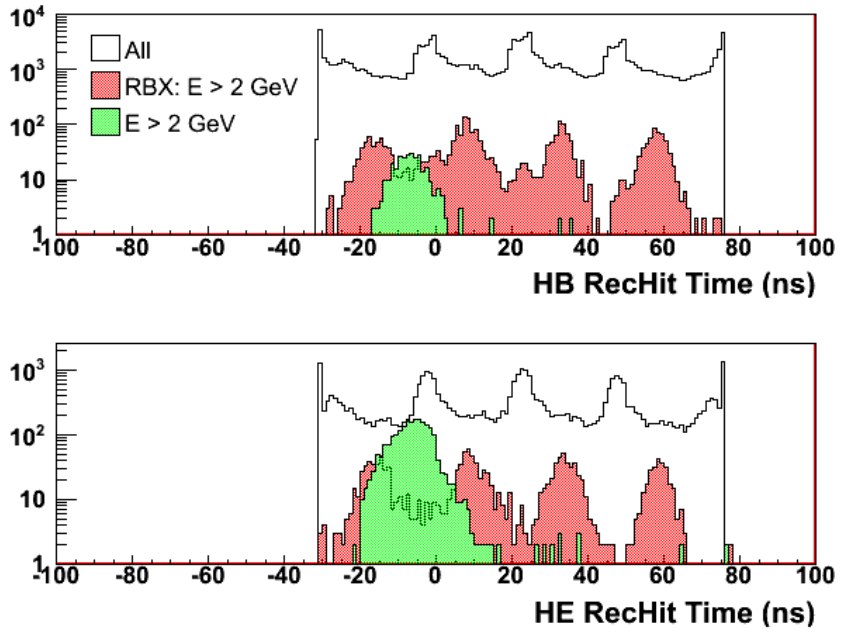

Figure 28: The rechit time for events that remain in the high MET tail (MET $>$ 45 GeV) compared with a pure noise sample. The RBX/HPD noise is random and occurs uniformly over the entire 4 TS window, while energy deposits from collision data is grouped near 0.

GeV). In the "pure" noise runs (absent beam) the maximum variation is on the HPD hit multiplicity requirement while the application of the entire filter has the same rejection efficiency within a couple of percent. The results from the collision data runs and the same baseline (i.e `L1SingleJet10U_noBPTX` trigger and at least one RBX with $E_{RBX} > 50$ GeV) are also compatible overall with the pure noise runs results. For the collision data (gated with the BPTX) 1% of the events are rejected almost exclusively from the E2/E10 requirement of the filter (always normalized to the baseline).Additional studies need to be performed to determine if these events are misidentified events as noise and/or noise events from other sources. The corresponding events rejected in the Minimum Bias MC sample are 0.1% from the same requirement. Finally less than 0.5% of signal top MC events (with and without pileup) and SUSY events are rejected also single handidly from the E2/E10(RBX(50)) requirement of the filter.

We have also studied the filter using only the `L1SingleJet10U` as baseline and the relative rejection rates are understood.

8.4 HCAL Noise Filters at the HLT

Anomalous HB/HE noise with energy $>$ 20 GeV occurs at a rate as high as 20 Hz, as measured in 2009 CRAFT data, although the exact rate has dependencies on detector conditions which change with time. This rate comprises a substantial fraction of the total HLT bandwidth. Table 1 shows the expected trigger rate due to physics events and due to HB/HE noise for common jet and missing E_T HLT paths. The menu assumes an instantaneous luminosity of 8×10^{29} cm^{-2} s^{-1}, the noise rate is extrapolated from 2009 CRAFT data, and the physics rate is extrapolated from lower instantaneous luminosity collision runs from May 2010.

Although the total rate of HB/HE noise is high compared to the HLT bandwidth, it is still quite low with respect to the crossing rate (40 MHz). As the presence of this type of noise

$8E29$ Menu ($\sqrt{s} = 7$ TeV)

Path Name	Total Prescale	Physics Rate (Hz)	HB/HE Noise Rate (Hz)	Noise Rate After Filtering (H
HLT_Jet15U	20	5.7 ± 1.2	~ 0.6	~ 0.3
HLT_Jet30U	1	11.5 ± 1.7	~ 7.7	~ 2.5
HLT_Jet50U	1	1.7 ± 0.6	~ 5.0	~ 0.9
HLT_MET45	1	~ 0.5	~ 6.4	~ 1.2
HLT_MET100	1	~ 0.0	~ 2.2	~ 0.3

Table 1: Expected rate for events triggered by physics and events triggered by HB/HE noise in different jet and missing E_T HLT paths in the $8E29$ menu for pp collisions at $\sqrt{s} = 7$ TeV. Note that the rate is dominated by noise in the HLT_Jet50U, HLT_MET45, and HLT_MET100 paths. The last column of the table presents the expected HB/HE noise rate after the filtering procedure described in this section.

$E > 50$ GeV RBX with:	$E2/E10$ outside the range [0.65, 0.98], or ≥ 10 ADC 0 counts, or contains an HPD with ≥ 17 hits, or contains an HPD with ≥ 10 hits and no other hits in the RBX
and the RBX has an EMF $< 2\%$	
and ≤ 2 RBXs with E> 10 GeV in the event	

Table 2: Summary of conditions for rejecting an event as HB/HE noise at the HLT. The $E > 50$ GeV RBX must be noisy, have a low electromagnetic fraction, and be only one of two energetic RBXs in the event. These requirements have a high efficiency for identifying noisy events, while maintain a very low misidentification rate.

is predominantly a stochastic process, the probability of noise overlap with a collision that would otherwise have fired the trigger is very low (measured to be at most a part in 10^5). Therefore, in the intial phases of CMS detector commissioning when LHC buckets are mostly empty, requiring that a trigger is coincident with the presence of beam substantially reduces the impact of anomalous noise at the HLT. As LHC buckets continue to be filled, this condition will be degraded and beam coincidence will no longer be sufficient to suppress HB/HE noise. A different approach will be necessary.

We can use the same techniques to reject noise described in previous sections to reject events that contain noise at the HLT. Because rejected events at the HLT are unrecoverable, we modify the conditions of the algorithm to suppress the rate of noise misidentification to levels better than one part per 10^4. Much like before, we look for high energy RBXs which present the characteristics of anomalous noise, but we also require that the RBX has an electromagnetic fraction less than 2%. This requirement ensures that the RBX is unlikely to be due to a high energy jet that has a pulse shape that fluctuates in a uncharacteristic manner. Moreover, we require that there are at most two RBXs with $E > 10$ GeV. Since noise events will typically overlap with minimum bias collisions or no collision at all, this requirement minimizes the likelihood that noise is coincident with anything but soft collisions. Table 2 summarizes the requirements for event rejection at the HLT.

By applying these requirements as a filter to events that pass the jet and missing E_T paths at the HLT, we find a very high efficiency for noise identification. The expected rates from

DRAFT

2009 CRAFT data after filtering are shown in the last column of Table 1. The overall rate of HLT triggers due to HB/HE noise is reduced approximately 65%. The same filter can be applied to events in MC to test the misidentification rate. We find that out of 5,000 simulated $t\bar{t}$ events, none were identified as noise by this filter. Preliminary tests of the misidentification rate from events collected directly by the HLT suggest a similarly low rate. While studies of this procedure are ongoing, we believe that this process will be crucial to mitigating the effects of anomalous HB/HE noise at the HLT.

8.5 Simulation of HCAL Noise

Figure 29 shows a GEANT4 simulation of the production points of particles hitting PMT glass windows in 7TeV minimum bias collisions. Fifty nine percent of those are muons (black dots) from decay in flight of pions and kaons before HF along the path from primary vertex toward PMTs. Thriry five percent are electrons (blue) produced in material just before PMTs. Most of those electrons are traced to leakage of hadronic showers in HF. Six percent are pions (red) and kaons (green) from hadron showers in HF.

Figure 30 shows the number of the HF PMT hits as a function of energy in HF in 7TeV collision data. PMT hits were identified using the algorithm described below and were removed from the energy calculation. The number of anomalous increases linearly with the HF energy. GEANT4 simulation reproduces the trend, but shows some deficiency. Tuning of GEANT4 simulation is in progress.

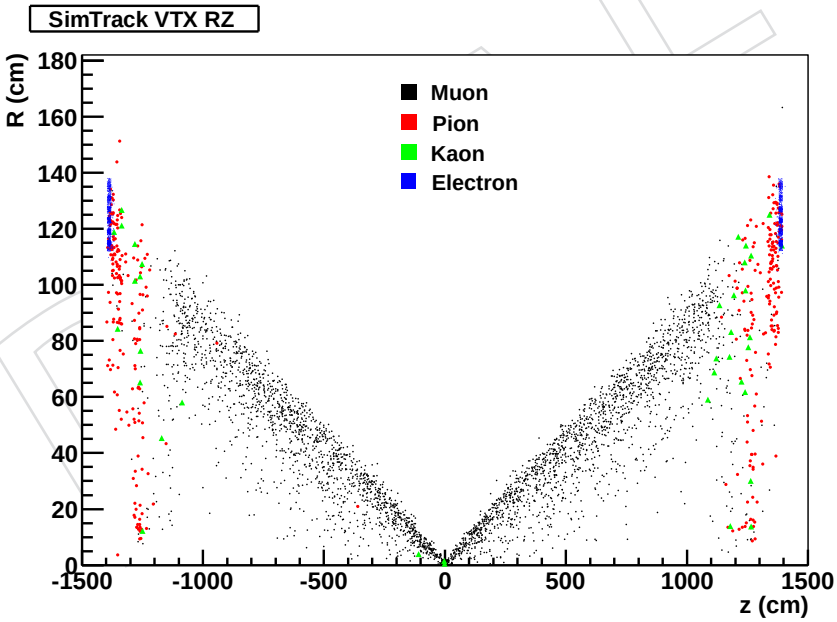

Figure 29:

9 Operating Experience

10 Summary and Outlook

The HCAL is an esential component of the CMS detector. Searches for new physics using a missing E_T signature will rely heavily on a good understanding of the HCAL performance.

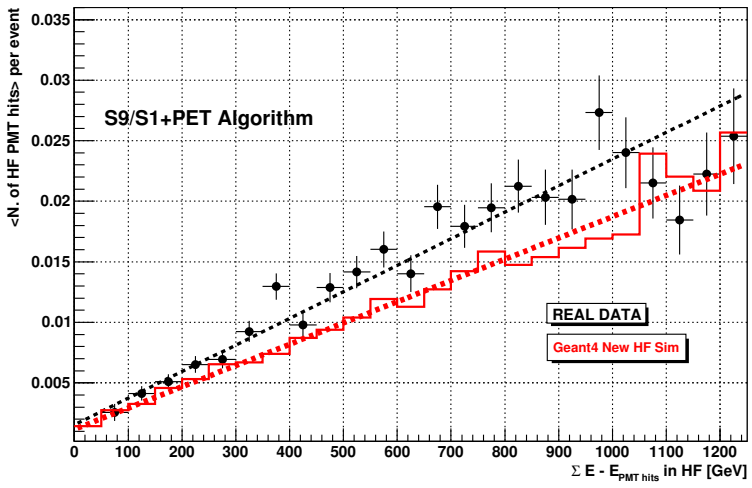

Figure 30:

The CMS HCAL has been precalibrated in the test beam in order to reproduce the response of 50 GeV pions. Additional tuning has been performed using cosmic and beam splash events in order to smooth out any channel by channel variation. This precalibration condition does not take into account additional material of CMS once it is fully assembled. A strategy to determine the HCAL response using collision data has been developed. The in situ calibration requires non zero supressed data in order to apply relative corrections in phi. Isolated charged particles are then used to set the energy scale and determine the response as a function of eta. Special triggers have been defined in order to collect sufficient samples of events for the calibration. The calibration procedure is being tested using low momenta tracks and we have observed some discrepancies in the MC response. As we collect more data we will be able to examine the Data/MC agreement over a wide range of variables.

The development of extensive monitoring tools allows for the early detection of problems helping to ensure we collect good data with a high efficiency. HCAL monitoring shifts and a prompt

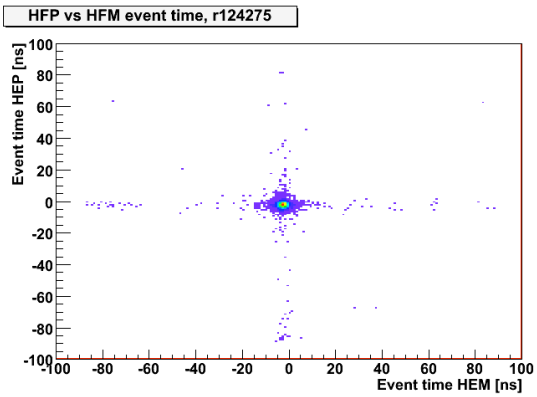

Figure 31: The HF time.

feedback team are available to quickly respond to problems and assist in tracking down the underlying source. Special procedures have been developed to respond to hardware faults. For example a hot trigger tower resulting in a high trigger rate can easily be masked. The detector conditions are continually monitored and any necessay changes to the reconstruction conditions can be updated during offline processing. In some cases a team of experts are assembled in a Working Group to study problems and develop solutions. This arrangement has been working well but it does rely extensively on a few key people. It will be necessary to developed a deeper pool of experts to provide continuous coverage for the life of the experiement. The monitoring tools have to be efficient at detecting causes of problems so that quick action can be taken. we expect that the monitoring tools will require continuous development and adaption to detect new problems.

The data is certified as "good for physics" by a special team. Results from online and offline Data Quality Monitoring are used as the primary input to the decision. Additional information is collected from entries by the shift crew using the "shifter checklist", Logs, and a list of known problems. This collection of information is time consuming and work is ongoing to automate the collection of the information and to present it in an easy to interpret format. Ultimately the procedure should be fully automated.

Several sources of anomalous signals are identified and algorithms to remove the associated energy have been implemented. The main sources of anomalous events in HCAL come from HF PMT window interactions associated with collisions and a low rate of random RBX/HPD noise. The HF is a Cerenkov calorimter with a very fast response allowing the possibility to separate early ariving particles impacting the PMT window from the light signals generated in the bulk of HF. Collision data was used to adjust the phase of the HF so that energy deposits from collisions are recorded at the same time throughout HF.

Cleaning algorithms are applied during event reconstruction so that all users have the same starting point. A flexible software framework for the treatment of anomalous signals allows one to apply the most up to date cleaning algorithms and to test new ideas. More aggressive cleanning can be applied at the user analysis level. RBX/HPD noise is random and the overlap with collision events is rare. It is easier to to remove entire events identifed as having high channel multiplicity RBX/HPD noise than to introduce a large hole in the detector. Since some exotic physics may mimic this signal we do not apply this level of cleaning by default.

As the event occupancey increases, cleaning algorithms that utilize energy isolation will become less effective. Algorithms using time or the pulse shape will have to be further developed.

DRAFT

847 This will require a more precise understanding of the pulse shape as measured on the detector
848 and also an accurate simulation. Hardware upgrade options that would reduce the effect of
849 the PMT window interactions have been proposed and under evaluation. Improvements to
850 the simulation of noise are being developed.

DRAFT

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[2] S. Abdullin et al., "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges", Eur. Phys. J. C55 (2008) 159171. doi:10.1140/epjc/s10052-008-0573-y.

[3] S. Abdullin et al., "Design, performance, and calibration of the CMS forward calorimeter wedges", Eur. Phys. J. C53 (2008) 139166. doi:10.1140/epjc/s10052-007-0459-4.

[4] S. Abdullin et al., "The CMS barrel calorimeter response to particle beams from 2 to 350 GeV/c", Eur. Phys. J. C, 60, 359, (2009).

[5]

[6] L. Tuura, A. Meyer, I. Segoni, G. Della Ricca, 'CMS data quality monitoring: systems and experiences', Journal of Physics: Conference Series **219 (2010) 072020**

[7] L. Tuura, G. Eulisse, A. Meyer, 'CMS data quality monitoring web services', Journal of Physics: Conference Series **219 (2010) 072055**

[8] The CMS Collaboration, 'CMS Data Processing Workflows during an Extended Cosmic Ray Run CMS Paper CFT-09-007, arXiv:0911.4842v1 [physics:ins-det]

[9] Afaq A, Dolgbert A, Guo Y, Jones C, Kosyakov S, Kuznetsov V, Lueking L, Riley D, Sekhri V, 2007, Proc. CHEP07, Computing in High Energy Physics, The CMS Dataset Bookkeeping Service (Victoria B.C., Canada)

[10] https://twiki.cern.ch/twiki/bin/view/CMS/HcalDataValidation

[11] Badgett W, 2007, Proc. CHEP07, Computing in High Energy Physics, CMS Online Web-Based Monitoring and Remote Operations (Victoria B.C., Canada)

[12] The CMS Collaboration, "Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data", CMS Paper CFT-09-018, arXiv:0911.4877v2 [physics:ins-det]

[13] S. Chatrchyan *et al.* [CMS Collaboration], "Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter," JINST **5** (2010) T03014 [arXiv:0911.4881 [physics.ins-det]].

DRAFT