Update on the HCAL code in ORCA, and Toy Simulations of the Readout Response

S. Abdullin\textsuperscript{a)}, S. Enos\textsuperscript{b)},

\textit{Experimental High Energy Physics Group, Department of Physics, University of Maryland, Regents Dr., College Park, MD 20742, USA}

J. Elias\textsuperscript{c)},

\textit{Fermilab, P.O. Box 500, Batavia, IL 60510-0500, USA}

Abstract

We describe a major update to the HCAL code in ORCA which makes the simulation of the HCAL response much more realistic. The new code includes a simulation of the photo statistics of the HPD, integration and quantization of the signal by the QIE, and separate simulations for the HB/HE and the HF. We also describe studies done using a stand-alone version of the code (outside of the ORCA + Objectivity Database framework) that facilitates quantitative estimates of the HCAL readout response. We find that if we use the energy recorded during only one time bucket, we get somewhat better bunch crossing identification efficiency than if we use two time buckets. The energy resolution for fairly low signal is also slightly better with one time bucket collection. In the Appendix we consider the case with the long signal shape, mainly comparing energy collection in 2 time buckets and in 3 ones.

\textsuperscript{a)} On leave of absence from ITEP, Moscow, Russia. Email: abdullin@mail.cern.ch

\textsuperscript{b)} Email: eno@physics.umd.edu

\textsuperscript{c)} Email: elias@fnal.gov
1 Introduction

We may start from the top of the “tree” : Universe (Nature) and High-Energy Physics. Then go down to the LHC [1] as the fore-front of science and CMS [2] as one of the main detectors at the LHC. The next sub-branch of the three is the Hadron Calorimeter (HCAL) [3] intended to catch up hadrons emitted from pp-collisions thus providing a hermetic detector for energy and missing energy measurements.

It is essential that any detector under construction have code that provides a realistic simulation of the detector response. In CMS, the code that provides both the simulation of the electronics, the trigger, and higher-level object reconstruction is called ORCA [4]. A realistic description of the front-end electronics is an important part of a proper HCAL simulation, and is necessary for designing trigger tables, evaluating trigger rates and evaluating the physics potential of the CMS detector.

The currently available HCAL readout simulation in ORCA is far from realistic. The code was written by members of the ECAL group and thus the code and its functionality is very similar to that of the ECAL (Electromagnetic Calorimeter), which, despite the fact that the calorimeters do have some common features, is sometimes quite wrong, and sometimes just not quite correct.

The purpose of this Note is to describe an update to the HCAL part of ORCA code, and some inevitable changes this update requires in the code that is common to both ECAL and HCAL. In addition, some more general updates in calorimetry that improve the performance of the code are described.

The Note is subdivided as follows. The second Section briefly describes the current HCAL code in ORCA, with an emphasis on the problematic or missing features, and outlines the algorithms used in the new code. The third Section describes some problems encountered when the first variants of our updated code were tested on a single pion sample, problems that led us to write a stand-alone (outside of the ORCA framework) code, (a “toy MC”). The next, fourth Section, gives a short description of the toy MC features and functionality. Then, the sixth Section describes the main results obtained with the toy MC for pure noise, noise + signal, noise + pileup + signal and shows some attempts to reduce the noise using a shape-based filter. In the seventh Section we describe the technical details of the changes in the existing classes and implementation of new classes and methods in ORCA. The conclusions follow in the eighth Section. Finally, plans for a future (ORCA-based analysis) are uncovered and acknowledgements are given.

In the Appendix we put the results of our previous study where we considered possible long shape signal. The sequence of sections more or less follows that of the main part of the Note, skipping introduction and some descriptive sections. The results of this study can be compared with those obtained in the main body of the present Note.

2 Current (Obsolete) HCAL Simulation in ORCA Code and Outline of the Requirements for the New Code

A brief description of the current HCAL implementation can be found elsewhere [5]. Here we emphasize the issues which are not quite correct or completely incorrect and thus require an update. This is not an attempt to criticize what was done in the past 2-3 years but rather a list of the critical issues which needed improvement.

A sketch of the current HCAL readout simulation is shown in Fig. 1. There are 3 main components which contribute to the signal shape : i) the scintillator and wave-length shifter; ii) the HPD (Hybrid Photo Detector) signal shape; and iii) the preamplifier. Formulae which parameterize the response of each component, along with the corresponding time constants, are shown in Fig. 2. The resulting convoluted shape has a peak position at \(\approx 16 \text{ ns}\) after the energy is deposited in the calorimeter, see sharp curve in Fig. 3. In the same Fig. we plot also an old shape with much bigger peak time of 32 ns. In the main body of this note we consider the signal as having the shape with peak time of 16 ns. In the Appendix we show the results of our previous study obtained with the long signal shape (peak position at 32 ns).

The current HCAL readout simulation assumes that delta-function-like signals from the sensitive volumes of the calorimeter are stretched according to a universal signal shape by the front-end electronics, before the input of the sampling ADC (Amplitude-to-Digit Converter). This sampling ADC measures the amplitude of the signal every 25 ns (which is the LHC bunch-crossing time cycle), thus providing multiple measurements of the same amplitude, convoluted with the shape. Each single measurement thus reflects a fixed fraction of amplitude of the “delta-function” signal, fixed by the universal signal shape. The evaluation of the signal and pedestal is a linear combination of N measured amplitudes slicing the signal shape at 25 ns intervals. The filled circles in Fig. 1
Figure 1: Sketch of the current HCAL readout simulation in ORCA.

- **Scintillator + wave-length shifter**
  \[ f_d(t) = e^{-t/\tau_d}, \quad \tau_d = 10 \text{ ns} \]

- **HPD**
  \[ f_{HPD}(t) = 1.0 + \left( \frac{t}{\tau_{HPD}} \right), \quad \tau_{HPD} = 12 \text{ ns} \]

- **Preamplifier**
  \[ f_p(t) = t \cdot e^{-t/\tau_p}, \quad \tau_p = 5 \text{ ns} \]

Figure 2: Signal shape time components.

Figure 3: Shape of the HCAL response to the instantaneous deposit of energy.
represent the N amplitude measurements of the signal (solid line) in the presence of pileup coming from each bunch crossing (dashed line). The pedestal measurement (“presamples”) occurs in time buckets 1-3, while the signal is measured in time buckets 4-8. For more details see e.g. [6].

In reality the HCAL has quite a different scheme for signal evaluation. The region in Fig. 1 that is circled and marked as “absent” shows 2 main issues for the simulation of the HCAL. The first one is the QIE (Q stands for charge, I - for integration, E - for range encoding), which takes the signal that comes from preamplifier and possesses some extended shape and integrates it within each 25 ns long time bucket. The rectangles show the effect of the integration, and represent the signal shape at the input of ADC. The second issue is that the current version of the code effectively contains no ADC (quantization of the signal as a part of the QIE). This is also true for the current “realistic” version of the ECAL code. The code does contain some integer multiplication, but since it is by $10^5$, it preserves typically 3-5 decimal digits and thus gives much better precision than actual ADC. This current conversion gives an LSB (Least Significant Bit) that depends on the size of the signal: from a fraction of an MeV for small signals to maximum of a few tens of MeV (for maximal signals). For most of the signals it is really irrelevant.

![QIE integration](image)

**Figure 4:** Sketch of actual time buckets QIE integration.

As shown in Fig. 1 the ADC output contains 2 bits for the range (exponent) and only 5 bits for mantissa. Two additional bits are used for capacitor identification, and thus do not need to be simulated. A realistic, detailed specification of the ADC table is given in Fig. 5. As seen from the table and inscription beneath it, the real value of the LSB is about 300 MeV. The entire dynamic range of electronics is subdivided into 4 main ranges with 5 sub-ranges inside each range. The size of variable-size LSB is set to closely follow the energy resolution ($\sim 1/\sqrt{E}$).

Another important issue missed in the HCAL code so far (and not shown in Fig. 1 either) is the effect of the HPD, and its resulting contribution to the resolution due to photo statistics. This effect can play an important role in the energy resolution as will be shown in Section 5, as the light yield from the HCAL is fairly modest.

The current code also uses the same parameters for the HB/HE and the HF. In fact, they are treated as one detector in the code. However, as the HB/HE (HCAL Barrel and Endcaps) and the HF (HCAL Forward calorimeter) have quite different natures and signal formation mechanisms, they naturally have different parameters for the simulation and should be separated at some point in the code. The HF has a much lower light yield than the HB/HE, photo multipliers are used in HF while HPD’s are used in the rest of HCAL. They have different gains, and hence different ADC bin sizes and noise levels. Besides, the HF has so short a signal (a few ns peak time) that it can be comfortably integrated in a single time bucket.

Noise is an extremely important issue in all of these studies. In the current version of the HCAL code, the noise is assumed to be 0.6 MeV multiplied by the readout scale factor. The readout scale factors are different for different calorimeters (HB,HE, HF) and depth segments (1st, 2nd or 3rd), and range from 72-220, so that the noise varies from 43 to 132 MeV per amplitude measurement. The most recent measurements from currently existing electronics show Gaussian noise with an RMS of $\sigma=200$ MeV per time bucket per channel. This noise is uncorrelated between different time buckets, and thus the more time buckets used in the signal evaluation, the bigger the noise “contamination”. Taken in combination with the effective noise from “photo statistics”, the total noise is much larger than the default value in the “current” simulation.

In addition, several minor things are missing in the current HCAL-ORCA code. One of these is the GEANT hit proper time. As will be shown in Section 4, taking into account the jitter induced by this time variation does not have a significant effect for reasonably high-energy signals.

In the current HCAL readout simulation the main quantity is energy, either deposited or reconstructed. Noise is
# HCAL QIE Transfer Function

## QIE Data Word

<table>
<thead>
<tr>
<th>MSB</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>LSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1</td>
<td>1</td>
<td>-1.5</td>
<td>32</td>
<td>5</td>
<td>□</td>
<td>64</td>
</tr>
<tr>
<td>2 1</td>
<td>0.5</td>
<td>34</td>
<td>5</td>
<td>62.5</td>
<td>66</td>
<td>25</td>
</tr>
<tr>
<td>3 1</td>
<td>1.5</td>
<td>35</td>
<td>5</td>
<td>67.5</td>
<td>67</td>
<td>25</td>
</tr>
<tr>
<td>4 1</td>
<td>2.5</td>
<td>36</td>
<td>5</td>
<td>72.5</td>
<td>68</td>
<td>25</td>
</tr>
<tr>
<td>5 1</td>
<td>3.5</td>
<td>37</td>
<td>5</td>
<td>77.5</td>
<td>69</td>
<td>25</td>
</tr>
<tr>
<td>6 1</td>
<td>4.5</td>
<td>38</td>
<td>5</td>
<td>82.5</td>
<td>70</td>
<td>25</td>
</tr>
<tr>
<td>7 1</td>
<td>5.5</td>
<td>39</td>
<td>5</td>
<td>87.5</td>
<td>71</td>
<td>25</td>
</tr>
<tr>
<td>8 1</td>
<td>6.5</td>
<td>40</td>
<td>5</td>
<td>92.5</td>
<td>72</td>
<td>25</td>
</tr>
<tr>
<td>9 1</td>
<td>7.5</td>
<td>41</td>
<td>5</td>
<td>97.5</td>
<td>73</td>
<td>25</td>
</tr>
<tr>
<td>10 1</td>
<td>8.5</td>
<td>42</td>
<td>5</td>
<td>102.5</td>
<td>74</td>
<td>25</td>
</tr>
<tr>
<td>11 1</td>
<td>9.5</td>
<td>43</td>
<td>5</td>
<td>107.5</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>12 1</td>
<td>10.5</td>
<td>44</td>
<td>5</td>
<td>112.5</td>
<td>76</td>
<td>25</td>
</tr>
<tr>
<td>13 1</td>
<td>11.5</td>
<td>45</td>
<td>5</td>
<td>117.5</td>
<td>77</td>
<td>25</td>
</tr>
<tr>
<td>14 1</td>
<td>12.5</td>
<td>46</td>
<td>5</td>
<td>122.5</td>
<td>78</td>
<td>25</td>
</tr>
<tr>
<td>15 1</td>
<td>13.5</td>
<td>47</td>
<td>5</td>
<td>127.5</td>
<td>79</td>
<td>25</td>
</tr>
<tr>
<td>16 2</td>
<td>14.5</td>
<td>48</td>
<td>10</td>
<td>132.5</td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>17 2</td>
<td>16.5</td>
<td>49</td>
<td>10</td>
<td>142.5</td>
<td>81</td>
<td>50</td>
</tr>
<tr>
<td>18 2</td>
<td>18.5</td>
<td>50</td>
<td>10</td>
<td>152.5</td>
<td>82</td>
<td>50</td>
</tr>
<tr>
<td>19 2</td>
<td>20.5</td>
<td>51</td>
<td>10</td>
<td>162.5</td>
<td>83</td>
<td>50</td>
</tr>
<tr>
<td>20 2</td>
<td>22.5</td>
<td>52</td>
<td>10</td>
<td>172.5</td>
<td>84</td>
<td>50</td>
</tr>
<tr>
<td>21 2</td>
<td>24.5</td>
<td>53</td>
<td>10</td>
<td>182.5</td>
<td>85</td>
<td>50</td>
</tr>
<tr>
<td>22 2</td>
<td>26.5</td>
<td>54</td>
<td>10</td>
<td>192.5</td>
<td>86</td>
<td>50</td>
</tr>
<tr>
<td>23 3</td>
<td>28.5</td>
<td>55</td>
<td>15</td>
<td>202.5</td>
<td>87</td>
<td>75</td>
</tr>
<tr>
<td>24 3</td>
<td>31.5</td>
<td>56</td>
<td>15</td>
<td>217.5</td>
<td>88</td>
<td>75</td>
</tr>
<tr>
<td>25 3</td>
<td>34.5</td>
<td>57</td>
<td>15</td>
<td>232.5</td>
<td>89</td>
<td>75</td>
</tr>
<tr>
<td>26 3</td>
<td>37.5</td>
<td>58</td>
<td>15</td>
<td>247.5</td>
<td>90</td>
<td>75</td>
</tr>
<tr>
<td>27 4</td>
<td>40.5</td>
<td>59</td>
<td>20</td>
<td>262.5</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td>28 4</td>
<td>44.5</td>
<td>60</td>
<td>20</td>
<td>282.5</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>29 4</td>
<td>48.5</td>
<td>61</td>
<td>20</td>
<td>302.5</td>
<td>93</td>
<td>100</td>
</tr>
<tr>
<td>30 5</td>
<td>52.5</td>
<td>62</td>
<td>25</td>
<td>322.5</td>
<td>94</td>
<td>125</td>
</tr>
<tr>
<td>31 5</td>
<td>57.5</td>
<td>63</td>
<td>25</td>
<td>347.5</td>
<td>95</td>
<td>125</td>
</tr>
</tbody>
</table>

Bin Size and Charge are in fC, for HB 1fC ≈ 0.3Gev
Charge - minimum charge for the bin (low threshold)
□ - These codes never occur, if charge is less than 62.5 fC we see code 31 and never see codes 32,33,64,65,96,97
If there is an occasional fluctuation at the range #1 we might see code 0.
In this table assumption is that pedestal is in bin 2 (code 1), which is probably close to what the ASIC developer is trying to achieve.

Figure 5: ADC table of collected charge ↔ code correspondence.
simulated in terms of energy. On the other hand, in electronics simulation another gauge is commonly used - the photo electron (pe). This reflects the nature of the energy-to-light conversion, the light collection and amplification. To properly simulate the true nature of the noise, it is therefore necessary to change the internal (“hidden”) data processing to operate with photo electrons. The energy → pe conversion should be performed at the very beginning of the GEANT hits treatment before the QIE integration and converted back (pe → energy) in the DAQ path (Data AcQuisition) and in the TPG path (Trigger Primitive Generator).

Summarizing what is discussed in this Section, we list the new features implemented in our new version of HCAL readout simulation in ORCA code:

- realistic noise
- HPD photo statistics effect
- signal integration in 25 ns time buckets
- ADC quantization of each time bucket measurement
- HF readout simulation separated from that of HB/HE
- the signal expressed in terms of photo electrons

The technical details of the code implementation in ORCA are given in Section 6.

3 “Straightforward” Implementation Problems

An initial version of these changes was added to the HCAL code in ORCA and tested with a sample of single pions with \( p_T = 30 \text{ GeV} \). This “beta” version of the code contains 3 depth layers in the HCAL barrel (including the tail catcher) and 2 depth layers in the endcap. The weight pattern used when calculating the level 2 (L2 or “off-line”) energy in an individual depth segment as a weighted sum of the energy measurements in each time bucket was: \(-2/3, -2/3, -2/3, 1, 1\). The “long shape” signal parameterization was used in the calculations showed in this Section. This preliminary study showed an excessive contribution to the signal from noise, especially in the barrel. This is not surprising as noise is generated in terms of energy and hence falls down with pseudorapidity increase (as \( \sin \theta \)). Fig. 6 shows the result. The points show the energies of all the jets found in these events by an iterative cone algorithm jet finder with \( R=0.5 \) and a seed cut of 1 GeV. Since many L2 towers (EcalPlusHcalTower objects) have noise-originated energy above the modest seed threshold, many fake jets appear in the scatter plot, mainly as a bump in the center. The crosses (profile histogram) denote the average of the reconstructed jet \( E_T \) as a function of \( \eta \) after a much harder seed cut of 5 GeV. The latter eliminates all fake jets (the bump) but the effect of the noise can still be seen as a bump-like slope in the line of crosses.

Fig. 7 shows a scatter plot of the same pion sample shown in Fig. 6, except produced using the old HCAL in ORCA. Here both the scatter plot and the profile histogram (crosses) were obtained with an IterativeConeAlgorithm with \( R=0.5 \) and a seed cut of 1 GeV (they are just different representations of the same objects). There are no fake jets here even with the relatively low seed cut. And, there is no evidence of any noise contribution to the jet energy in the very center of the barrel. (The signal is higher in the endcap region because of the nonlinearity of the calorimeter, and the fact that pions with a fixed \( E_T \) have larger \( E \) in the endcap.) The overall shape of the distribution is well understood when produced using old HCAL code.

Figs. 8 and 9 further illustrate the situation with the new higher level of noise and with the signal integration. Figure 8 shows the probability that a tower (the sum of all depth segments of the HCAL) is occupied versus a threshold for the old and for the new code. \(^1\) (In Fig. 8, the occupancy seems to be independent of the threshold value for energies less than about 0.3 GeV. This is an artifact of a default threshold that was applied when the sample was first created, an HcalRecHit cut of 0.3 GeV.) Fig. 9 shows an average energy collected in 100 HCAL towers (this corresponds to the cone size of \( R=0.5 \) in \( \eta - \varphi \) space) versus the value of a zero-suppression energy cut that is applied to every single tower. This plot illustrates how the zero-suppression cut impacts on the

\(^1\) The results labeled “new code” were obtained with 2 signal + 3 pedestal buckets collection mode (see explanation in Section 4). The terms “single” and “double” readout denote cases with a single readout per HCAL tower or 2 separate readouts corresponding to the depth layers (0 and 1) respectively. The HCAL noise is taken to be Gaussian with \( \sigma=200 \text{ MeV} \) in each time bucket, and uncorrelated between time buckets.
Figure 6: Reconstructed transverse energy for single pions ($p_T = 30$ GeV) from an initial updated version of the HCAL code. The scatter plot corresponds to an Iterative Cone Jetfinder ($R=0.5$) with a seed cut=1 GeV, the crosses denotes the same algorithm with the seed cut=5 GeV. See more explanation in the text.

Figure 7: Same as Fig. 6, except from the old version of the HCAL code in ORCA. Both scatter plot and crosses correspond to a seed cut of 1 GeV.

Figure 8: Single and double readout occupancy for old and new simulations. 2 signal + 3 pedestal buckets signal collection mode is used in “new” simulation.

Figure 9: Noise collected in 100 HCAL towers for single and double readout as a function of zero-suppression cut on an tower.
reconstructed jet energy. In case of single/double readout, when this cut is applied, a bias of to the jet energy of ~ 18/26 GeV is introduced. This should also give an increase to the noise contribution to the reconstructed jet energy and a corresponding worsening of the resolution. But to which extent? And, how can we optimize the algorithms used to convert time samples to energy to minimize this effect? Thus, the problem of the HCAL update in ORCA becomes not so much a code update (at least at the relatively modest level of sophistication of our new code) but more in the optimization of the signal evaluation parameters. The HCAL group does not yet have optimised/fixed/committed/algorithms for the TPG and DAQ paths, so such parameters as number of signal and pedestal buckets, and Bunch Crossing IDentification (BCID) algorithm needed to be studied. In addition, we expect the optimum algorithm/parameters to depend on luminosity.

So, in parallel with re-writing the code, there was a need to optimize what was put inside it. Such a multi-parameter study is quite difficult to perform within the ORCA framework, as ORCA is a optimized to be a rather powerful processing tool, not an instrument for detailed investigations. We needed a kind of “test bed” for the ORCA code, a stand-alone code, which was easy to use an manipulate, “to simulate a simulator”, i.e. to mimic the HCAL readout simulation in ORCA. The next Section describes briefly this compact and fast code.

4 Stand-alone Code for Toy Simulation of the HCAL Readout Response

The code simulates a trigger tower response (1 or 2 readouts) for a fixed energy “injected” into the tower. The term “injected” means that a fixed energy is assumed to be deposited in HCAL tower connected to a single readout channel. This energy can be 0, in which case only noise will matter. Practically all of ORCA’s functionality is reproduced by this toy MC:

- parameterized and tabulated shape of the signal
- photo statistics effect
- QIE integration
- ADC quantization
- baseline position in any selected ADC channel
- variable number of time frames (buckets) analogous to ORCA’s TimeSample and DataFrame masses
- Gaussian noise injection in all time frames (uncorrelated between different time frames)
- variable number of signal and noise buckets with appropriate weights for signal extraction (e.g. -2/3,-2/3,-
  2/3, 1, 1 for 3 pedestal + 2 signal buckets)
- adjustable time phase for various collection modes

The shape, photo statistics, QIE integration, and ADC quantization coincide exactly with what was implemented in ORCA.

We used the toy MC to tune the time phase for optimal signal collection for various number of time buckets. Fig. 10 shows a comparison of the signal fraction collected in 1 and 2 time buckets as a function of the adjustable time phase. A similar comparison is presented in Fig. 11 for 3 versus 2 time buckets. In Figs. 10 and 11, each filled circle corresponds to a certain value of time phase, with a step-size of 1 ns. There are 25 circles and thus the plot corresponds to a full turn round from 0 to 24 ns counterclockwise. It is clearly seen that with 2 time buckets collection mode, one is able to get 99-99.5 % of the entire signal within ~ 5 ns, while the 3 buckets contain the signal entirely. The one time bucket collection mode has a region of stability of about 5 ns where the fraction of the signal collected is as high as 83-83.5 % (the rightmost side of the curve in Fig. 10). Though in the latter case the question of the robustness of the single time bucket collection mode against variation of the signal shape and signal-to-noise ratio when compared to 2 time bucket collection needs to be considered. In the rest of this note, when we use 1- or 2-buckets collection modes, we correspondingly correct for the fraction of the signal collected ~ 83 %, 99 % respectively.

The only thing missing in the toy MC compared with the “full simulation” is the GEANT Hit time jitter due to variation of the signal arrival time in the sensitive volume. This should not be a big effect if the time phase is adjusted properly as we see from Figs. 10 and 11.
Figure 10: Maximal fraction of the signal collected in 2 time buckets versus that collected in 1 time bucket. Every single filled circle corresponds to a value of the time phase ranging from 0 to 24 ns.

Figure 11: Same as Fig. 10, except for the signal collected in 2 time buckets vs that collected in 3 buckets.

Figure 12: Probability distribution to get a “true energy” in a single readout from a single minimum bias events for 4 different pseudorapidities.
In addition, the code has an option to simulate the contribution to the time frames from pileup interactions. This
was done as follows: We processed with ORCA_4.5.0 a single minimum bias event sample to get the probability
 distributions versus the energy reconstructed in a single readout for a few pseudorapidity values. Noise was
switched off and the RecHit energy cut was set to zero to get “true energy”. These data are then used to simulate
pileup using a Poisson distribution for the number of overlaid interactions, with an average of $\langle 17.3 \rangle$ minimum
bias events per bunch crossing (time bucket). Fig. 12 shows the probability distributions from the single mini-
mum bias events for 4 different values of pseudorapidity. It’s worth noticing here that we use the value of 2.8 as
everywhere as a shortcut of the actual pseudorapidity position (2.825) of the last HCAL tile in the CMSIM 122
geometry description.

5 Results of Toy Simulation

In the following we simulate the noise using a Gaussian distribution with $\sigma_{\text{noise}}=200$ MeV, which is equivalent of
2 pe. We assume 10 pe/GeV of deposited energy for the entire HB/HE. This is a slightly conservative assumption,
as the readouts closer to the HB/HE conjugation ($\eta \approx 1.5$) may have a value which approaches 15 pe/GeV and this
will soften the photo statistics effect in this region.

The LSB (ADC count value at the low end of its scale) is taken to be 300 MeV or 3 pe.

The simulation models a single readout (depth segment) per HCAL tower, unless it explicitly states otherwise in
the text or in Figure. “Double readout” means separate readouts on each layer (layer 0 and 1).

The baseline position (or mean value of the noise) is smeared randomly (flat random) within a given fixed ADC
channel for each new series of time buckets but is the same within the series. (Each series is used to get a single
energy extraction for a concrete “trigger” signal.) This models the “slow” noise component and this component
(unlike “fast” one described by $\sigma_{\text{noise}}$) is modeled by a flat, not a gaussian distribution (and thus never fluctuates
outside the chosen ADC channel).

Tab. 1 illustrates the baseline position in the ADC table and other ADC table attributes.

<table>
<thead>
<tr>
<th>Channel No</th>
<th>ADC code</th>
<th>Bin Energy Range (MeV)</th>
<th>Code $\rightarrow$ Energy (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-450 to -150</td>
<td>-300</td>
</tr>
<tr>
<td>2 *</td>
<td>1</td>
<td>-150 to 150</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>150 to 450</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>450 to 750</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>750 to 1050</td>
<td>900</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1050 to 1350</td>
<td>1200</td>
</tr>
</tbody>
</table>

* Default baseline position channel

5.1 Noise

In this Subsection we consider the response of the HCAL readout to noise without any signal, just noise-induced
“energy”.

We assume the “floating” baseline position is measured as a running average of a series of 64 digitized measure-
ments taken for instance in the abort gap of the accelerator, which is probably close to how the “precise” baseline
would be implemented in practice. In the following we use the term “presample” to denote a time bucket preceding
the signal collection bucket(s).

Fig. 13 shows the reconstructed energy distributions for various baseline positions when the signal is collected in
one time bucket. Fig. 14 is similar to Fig. 13, except the baseline is estimated from 3 presamples. In both Figs. 13
and 14 the difference between the “no ADC” case and all the other cases is not very significant, except for the case with the baseline in the 1st ADC channel.

Another (“integrated”) representation of a single readout response to noise in the form of readout occupancy versus energy cut-off, is shown in Figs. 15 and 16 respectively. Here we apply an additional selection compared to Figs. 14, 13, the “Bunch Crossing IDentification” (BCID), which is applied to the signal after the energy evaluation. In ORCA the procedure used for Bunch Crossing IDentification (BCID) is quite simple [7]:

- signal is evaluated with weights corresponding to the “true time”
- signal is also evaluated with the weights pattern shifted by +1,-1 time bucket
- if the signal is maximal (“≥”) with zero shift, then BCID is OK

We use BCID in all that follows when we calculate occupancy rates.

As we will discuss in Subsection 5.2 (Fig. 26), there is a non-linearity in the energy evaluation when the baseline is set in the 1st ADC channel. In what follows (from Figs. 15,16 on), we apply corrections to compensate this kind of non-linearity in all occupancy plots.

One sees in Figs. 15, 16 that there is some additional reduction of the noise if the baseline is set in the very first ADC channel. Effectively it means that noise is not allowed to fluctuate to the “left” - to the negative values and thus it is more unlikely that there will be negative values, with respect to 2 signal buckets in pedestal buckets (the pedestal buckets enter the signal evaluation sum with negative weights).

If one fixes the default position of the baseline in the 2d ADC channel and looks at the various signal collection modes, one sees that the 2-buckets collection mode gathers more noise that the 1-bucket one, as seen in Figs. 17 and 18.

For completeness we plot 3-D distributions of the noise occupancy for a single readout as a function of both pseudorapidity and $E_T$ cutoff in Figs. 19 and 20. Figs. 21 and 22 present some selected slices from Figs. 19 and 20 at $\eta = 0,1,2,3$. The curves show just the angular suppression of the noise, whose value is defined in terms of energy, not $E_T$. 
Figure 15: Another representation of what is plotted in Fig. 13, now in terms of readout occupancy versus reconstructed energy cut-off.

Figure 16: Same as Fig. 15, except the pedestal is estimated from 3 presamples.

Figure 17: Same as Fig. 13, except the baseline set in the 2d ADC channel, for various signal collection modes.

Figure 18: Same as Fig. 15, except the baseline is set in the 2d ADC channel, for various signal collection modes (as in Fig. 17).
Figure 19: HCAL readout occupancy due to pure noise versus $E_T$ cutoff and pseudorapidity for signal evaluation with 1 signal + 3 pedestal buckets.

Figure 20: Same as Fig. 19, except the signal evaluation is done with 1 signal + pedestal measured as a running average of 64 buckets (see details in the text).

Figure 21: HCAL readout occupancy due to pure noise versus $E_T$ cutoff when the signal evaluation is done with with 1 signal + 3 pedestal buckets.

Figure 22: Same as Fig. 21, except the signal evaluation is done with 1 signal + pedestal measured as a running average of 64 buckets (see details in the text).
5.2 Signal + Noise

Now let us see what happens when a fixed-energy signal is injected in a single readout (more precisely, in a tower connected to a single readout). The choice of parameters is the standard set, as discussed in the preamble to Section 5.

![Figure 23: Resolution of the readout response as a function of the number of pedestal buckets for injected energy of 1 GeV and for both (1 and 2 buckets) signal collection modes. The left column of symbols (circles, squares and triangles) denotes the 1-bucket collection mode, while the right one (stars, diamonds and inverted triangles) denotes the 2-buckets collection mode.](image1)

![Figure 24: Same as Fig. 23, except for 10 GeV signal.](image2)

Figs. 23, 24 shows the decomposition of various factors affecting the resolution for injected energies of 1 and 10 GeV. The 0 pedestal buckets bin is a short-hand for the baseline measured as a running average of 64 buckets. One may notice that the photo statistics effect plays a decisive role for relatively high energies, see its contribution marked by arrow in Fig. 24. The 2-buckets collection mode looks worse for relatively low (not very much higher than the noise level, hence a few GeV) energies, while in case of a 10 GeV signal the difference between the different collection modes is not so pronounced, as the photo statistics dominates over the noise contribution. This is in agreement with the conclusion made in the previous Subsection on pure noise considerations, that the 2-buckets mode is worse than the 1-bucket one. There is practically no difference between full resolution curves for the case of 1 signal + 3 pedestal buckets and one of 1 signal + 64 pedestal buckets, so only the latter one is plotted in Fig. 24.

![A slightly different representation of similar data is shown in Fig. 25, where the baseline in the 6th ADC channel is considered together with the case of the 1st channel for the signal of 1 GeV. From the previous Subsection we know that the case of baseline in the 6th ADC channel is practically coincide with the baseline in the second ADC channel, see Figs. 13–16. Here we should emphasize that the baseline in the 1st ADC channel causes some non-linearity, as in this case a systematic shift of the reconstructed signal energy is observed, as shown in Fig. 26 both for the 1 (circles) and 2 (squares) buckets collection mode for pedestal measured from 64 running buckets. A very similar non-linearity observed for any number of presamples if the baseline is set in the first ADC bin. As we already mentioned in a previous Subsection, we apply corrections for this nonlinearity whenever we evaluate the readout occupancy. There is just a tiny difference from the point of view of RMS/mean between the case when the baseline is in the](image3)
6th ADC channel (the noise can fluctuate freely) and in the 1st one in Figs. 25 for 1 GeV signal. This difference becomes smaller with energy increase, as photo statistics starts to play bigger role. For the energy of 10 GeV, as was already mentioned, photo statistics dominates and baseline position is not important from the point of view of the resolution.

One again sees in these Figs. that the 1 signal bucket collection is slightly preferred to the 2 signal buckets mode and that good knowledge of the baseline (0 pedestal buckets bin) does not improve the resolution much.

![Graph showing RMS vs Mean for different signal collection modes and baseline positions.](image1)

**Figure 25:** Same as Fig. 23, except the case where the baseline is in the 1st ADC channel is compared to the one in the 6th ADC channel.

![Graph showing average reconstructed energy vs injected energy for 1-bucket and 2-bucket collection modes.](image2)

**Figure 26:** Average reconstructed energy versus injected energy for the 1-bucket collection mode (circles) and 2-buckets one (squares) for baseline put in the 1st ADC channel. Dashed line shows an ideal linearity case.

### 5.3 Bunch Crossing Identification without Pileup

Fig. 27 shows the BCID efficiency as a function of the injected $E_T$ for various signal collection modes and for 4 pseudorapidity regions. Here we consider only positive values of the signal, since a zero-suppression cut is applied after BCID.

There is a purely geometrical factor ($1/\cosh(\eta)$) which defines the noise suppression with $\eta$, so the BCID is more efficient for higher pseudorapidities at a given reconstructed $E_T$. In the central HCAL region the turn-on curves are fairly slow, especially for the 2 signal buckets mode.

### 5.4 Noise Shape Filter (?)

A simple straightforward idea inevitably comes to mind: is it possible to suppress the noise using the difference between the pronounced shape of the signal and random nature of noise?

Suppose we tune the time phase in such a way that the signal shape provides approximately equal content in two adjacent buckets, collecting $\approx 97\%$ of the entire signal. Or we may tune the time phase to use 1-time bucket collection mode with $\approx 83\%$ of the signal in one time bucket. The baseline is smeared in the 2d ADC channel and pedestal is estimated as the average value of 64 digitized measurements. At the first glance the signal formatted in such a way should be easy to distinguish from random noise. Let us try to quantitatively estimate the filtering effect for 2 types of filter:
Figure 27: The BCID efficiency as a function of the injected signal $E_T$ for various signal collection modes and for 4 pseudorapidity regions.

- Filter 1 for 1-bucket collection:
  - the content of the signal bucket $\geq 1$ ADC count (code=2)
  - content of both 1 preceding and 1 subsequent bucket $< \text{that of the signal bucket}$.

- Filter 2 for 2-buckets collection:
  - both signal buckets (B1 and B2) $\geq 1$ ADC count (code=2)
  - $\text{abs}(\max(B1,B2) - \min((B1,B2)) \leq 3$ ADC counts

Fig. 28 show the results of the application of the Filter 1 in terms of the fraction of events that passed the filter as a function of injected energy for 1-bucket collection mode (circles). An example of how Filter 1 affects the pure noise distribution is shown in the thumbnail plot in the lower part of the Fig. 28. It is clearly seen (where the arrow points) that noise fraction surviving the Filter 1 is $\sim 12\%$. Looking at the thumbnail plot one can conclude that a similar noise suppression can be achieved by application of some simple cut on the reconstructed energy. This is shown by curve of squares in Fig. 28 which denotes a fraction of events passed the cut of 0.3 GeV imposed on $E_{rec}$ (don’t mix it with the injected energy). One can see that this simple cut is less damaging for the signal than the shape-based Filter 1.

The same calculations for 2-buckets collection mode and Filter 2 are shown in Fig. 29. Here the noise is suppressed slightly further than in a previous plot (just due to the choice of the harder filter algorithm), down to $\sim 8\%$. And again, applying energy cut of 0.45 GeV on the reconstructed energy (slightly harder than in case of 1-bucket collection mode) one can achieve the same noise suppression as Filter 2 does but loosing less the signal.
Of course, one can make the filter tighter to suppress the pure noise down to a percent level but the loss of the signal increases accordingly. So it does not look profitable to apply a special shape-based noise filter instead of just a simple zero-suppression cut.

The conclusion for the shape-based noise filtering can be formulated as follows: the high noise level in the HCAL plus the relatively big LSB hampers such a filtering.

Figure 28: Turn-on curves for 1-bucket collection mode with Filter 1 (circles) and simple cut on reconstructed energy (squares) as a function of injected energy. The superimposed thumbnail plot in the lower part of the Figure shows how the distribution produced by pure noise changes when Filter 1 is applied.

Figure 29: Same as Fig. 28, except for 2-buckets collection with Filter 2.

5.5 Noise + Pileup

As was mentioned in Section 4, our toy MC is able to simulate the contribution from pileup to a series of time buckets by using the probability P(E) for a single minimum bias event to contain an energy above a threshold, see Fig. 12. Here and in the following we consider as a pileup a sum of contributions from minimum bias events produced during one time bunch crossing with Poisson average of 17.3 events. Using this option, first we calculated distributions of the response from a tower (using a single readout) to noise and pileup for various choices of the parameters for signal evaluation, Figs. 30 - 37. The occupancy plots, Figs. 38 - 41, are obtained with applied BCID algorithm and corrected non-linearity in case of the baseline set in the 1st ADC channel.

In Fig. 30 one can see that there is no significant difference between noise only (circles) and noise + pileup (squares). There is just a small (< 1%) tail on the right side of the noise + pileup distributions caused by pileup. A similar tendency might be observed in Fig. 31, where pedestal is estimated from 3 presamples. These 2 plots shows that noise plays a more important role than pileup to the shape of the signal in the central region. We also see here both effects observed earlier in Subsection 5.1: collecting the signal in 1 bucket is better than in 2 (signal/background is better) and that setting the baseline in the 1st ADC channel gives a narrower distribution than in other channels (in the absence of the signal).

Figs. 32 and 33 show similar plots for η = 1, and are pretty much the same as those for η=0: pileup is small compared to noise. Some angular suppression of the width of distributions is observed compared to η=0 case.

The situation changes quite drastically at higher η=2 and 2.8, as shown in Figs. 34 – 37. The noise now is (much) smaller than the pileup energy, and the total effective width of the distribution becomes bigger than that for low η values. As a consequence of all this, the difference between the results obtained when the pedestal is estimated using 64 buckets and 3 presamples disappears if we take into account only positive E_T values (as if a
Figure 30: The response of a single readout to noise + pileup at $\eta=0$ for various collection modes and when the pedestal is estimated from 64 digitized time buckets.

Figure 31: Same as Fig. 30, except the pedestal is estimated from 3 presamples.

Figure 32: Same as Fig. 30, except for $\eta=1$.

Figure 33: Same as Fig. 32, except the pedestal is estimated from 3 presamples.
Figure 34: Same as Fig. 30, except for η=2.

Figure 35: Same as Fig. 34, except the pedestal is estimated from 3 presamples.

Figure 36: Same as Fig. 30, except for η=2.8.

Figure 37: Same as Fig. 36, except the pedestal is estimated from 3 presamples.
Figure 38: Another representation of what is plotted in Figs. 30 and 31, in terms of occupancy versus $E_T$ cutoff.

Figure 39: Same as Fig. 38, except for $\eta=1$.

Figure 40: Same as Fig. 38, except for $\eta=2$.

Figure 41: Same as Fig. 38, except for $\eta=2.8$. 
zero-suppression is done).

Let us again take a look at the distributions of occupancy versus the $E_T$ cutoff for the same $\eta=0, 1, 2, 2.8$ regions, Figs. 38 – 41. One may see that from this point of view, 1-bucket mode doesn’t look significantly better than 2-buckets one.

5.6 Bunch Crossing Identification with Pileup

In this Subsection we consider BCID for the case of noise + pileup. Again, only positive reconstructed energies are considered (as if zero-suppressed). Fig. 42 presents the results of the calculation (cf. to Fig. 27). Now the 1-bucket collection mode slightly edges out the 2-buckets one at low pseudorapidities, while at high ones there is no big difference among them. This needs to be confirmed by a full ORCA simulation.

![Figure 42: Same as Fig. 27, except for noise + pileup case.](image)

6 Update Description

Our ORCA code update defines a few new classes and a few new features (changed public methods and new public/private members) in the existing classes.

The list of changed Calorimetry packages, with some brief comments, is as follows:

**HcalRealistic**
• **Class** **HcalRealisticShapeHF** added

• **Class** **HcalRealisticReadout** changed
  - public methods added: ShapeHF, SigmaElecNoiseHF, PeakTimeHF, PeakTimeAdjust, PeakTimeAdjustHF, GeVtoPE, PEtoGeV
  - public methods changed: SigmaElecNoise, HcalRealisticNoisifier

• **Class** **HcalRealisticCoder** changed
  - public methods added: GetCodeFADC, GetEnergyFADC
  - public methods changed: convert

The Class HcalRealisticShapeHF is added so that the HF can have a different (short-shaped) signal than the HB/HE (both of which have longer, similar shapes). As both HF and HB/HE cells are defined in a common class (HcalBase), our new code distinguishes HF cells from HB/HE cells using rather primitive

```cpp
if (subdetector == "HF")
```

... methods. For the same reason, we needed to create public methods SigmaElecNoiseHF, PeakTimeHF, PeakTimeAdjustHF for HF. We created a handle in .orcac for switching on and off the QIE (integration and quantization) simulation so we could switch between the old/new simulation without relinking. When the QIE is switched off, SigmaElecNoise/SigmaElecNoiseHF returns the old noise settings (0.0006 GeV in terms of ooHit energy) and when on, it returns the “new” settings. The new noise settings are expressed in terms of photoelectrons (2 pe for HB/HE and 0.125 pe for HF). Because both HB/HE and HF are handled by the common base class HcalBase, we’ve had to write some kludged code so that they have different signal shapes. For the HF, ShapeHF creates an artificial short-time constant shape which is basically a delta function, and then tabulates it using the same 1ns step size as is used for the HB/HE (even though the entire HF pulse thus fits in a few steps). PeakTimeHF (like PeakTime for HB/HE) returns the time between the deposition of the energy in the calorimeter and the time when the collected signal has its maximum value, which is a few ns for the HF. PeakTimeAdjust returns an offset which is used to adjust the timing of the signal relative to the bin edges of the 25 ns time buckets to get a desirable signal collection in the time buckets, e.g. maximum content in the 2 or 3 time buckets starting with the 5-th time bucket out of the 10 used in the simulation (from 0 to 9 in C++). For the HF this method returns a time phase that puts the signal in the middle of the 5-th time bucket. GeVtoPE and PEtoGEV convert ooHit energy (GEANT hit’s, wrapped in C++) to photoelectrons. The conversion factors are SimpleConfigurables, and can be set in .orcac. There is one constant for each depth layer in HB/HE. In HF ooHits are expressed in terms of pe, so there is no need for a conversion. In HcalRealisticCoder::convert, which does the conversions, TimeSample → DataFrame and back : DataFrame → TimeSample, two new methods GetCodeFADC and GetEnergyFADC make conversions between the number of photoelectrons and the resulting binary code using the ADC table.

**HcalRealisticRec**

• **Class** **HcalRealisticAnalyser** changed
  - public methods changed: evalAmplitude, computeWeights

The Public methods in class HcalRealisticRec changed because of the transition from ECAL-like amplitude evaluation to a simplified HCAL integrated signal evaluation. Besides, the scale factors formerly presented in this class are moved to CellReadout (see description of this class below) to be accessible from various classes.

**HcalTrigPrim**

• **Class** **HcalTrigAnalyser** added
  - public methods added: evalAmplitude, computeWeights

• **Class** **HcalTrigPrimFormatter** changed
  - public methods changed: constructor, reconstruct
In the old version of the code, all the trigger primitives (both ECAL and HCAL ones) were evaluated similarly, and there was no HcalTrigAnalyser that would allow you to process the HCAL items separately. Class HcalTrigPrimFormatter changed to use the scale factors from CellReadout instead its own ones earlier and to use some separation between QIE / no-QIE cases.

**HcalBarrel**

- Class **HcalBase** changed
  - public methods changed: `upDate`, `MakeID`
  - public methods added: `HitReWeight`

The main changes in HcalBase consist of: i) assigning the ooHits from layer 0 to layer 1 in both HB and HE (MakeID method) and ii) complementary to that - re-weighting the ooHit energy coming from layer 0 by a factor of `scale(layer 0)/scale(layer 1)` so that when the layer 1 scale factor is applied later, layer 0 will get the correct weighting. The merging of layer 0 and layer 1 reflects a recent decision of the HCAL community to have a single readout in HB/HE (excluding the HO layer - the tail catcher). The re-weighting of the “sub-layers” (former layers 0 and 1) is assumed to be possible optically before common electronics mix the signal.

**CaloHit**

- Class **CaloHitFormatter** changed
  - public methods changed: `fill`

The CaloHitFormatter::fill method is the place where HcalBase::HitReWeight is used when GEANT hits are converted to ooHits.

**CaloTrigRec**

- Class **CaloTriRecSetup** changed

The change consists of the replacement of the private member

```
// CaloTrigAnalyser hcalAnalyser; with
HcalTrigAnalyser hcalAnalyser;  // see comments for HcalTrigAnalyser above.
```

**CaloReadout**

- Class **CaloFrontEndResponse** changed
  - public methods added: `HcalReadoutTreatment`
  - public members changed: `samplecache`
  - public methods changed: `getCachereplaced with getSampleCache, constructor, lazyUp-Date`

- Class **CellReadout** changed
  - public methods added: `Calibration`
  - public methods changed: `constructor`

- Class **CaloScaler** changed
  - public methods changed: `Scale`

- Class **CaloDataFrame** changed

Class CaloTimeSample changed


CaloFrontEndResponse::lazyUpdate is the main steering method for converting ooHits to TimeSamples and contains the readout simulation (shape, integration, ADC etc.). In order to separate the HCAL treatment from the ECAL one, and to hide those details specific to HCAL, we added the HcalReadoutTreatment method. The ECAL part of the code is thus untouched.

The CellReadout class had to be changed to accommodate the scale factors for HCAL. These factors were placed earlier in both in HcalRealisticAnalyser and HcalTrigPrimFormatter. Class CaloScaler is changed accordingly, and assigned a scale factor = 1 by default instead former 0 in order not to touch ECAL as ECAL doesn’t contain it’s own scale factor(s) and a default value is used.

The CaloDataFrame and CaloTimeSample classes change are not necessary in the scope of the HCAL readout simulation and can be treated as optional. See an explanation at the end of the current Section.

CaloRecHit

Class CaloRecHitFormatter changed

public methods changed: reconstruct

The change is optional, see more details at the end of this Section.

CaloDetector

Class CellID changed

public methods added: getUniqueNum, setUniqueNum

Class CellProperties changed

public methods added: getNum, setNum

Most of the changes create a fixed size vector (with sequential cell indexing) of CaloTimeSample objects for each particular instance (“HCAL”, “EBRY”, “EFRY”, ...) of CaloBase which can be addressed directly via the reference operator “[]”. This vector has a fixed size equal to the maximal size of each CaloBase instance and is created once per run/job and then is re-filled with noise and ooHits for each new (trigger) event. It is intended to replace current map<CellID, CaloTimeSample,...> and avoids time-consuming searches in the map for new or existing members and the creation and deletion of the map for each CaloBase in each event. It brings a sizeable CPU time economy, especially for mixed trigger event + pileup, though at the price of an additional (constant size) memory demand.

Another option implemented in the code is the evaluation and transport through the data flow of 2 additional doubles for each readout cell: i) the trigger event energy and ii) the in-time-only energy (both after multiplication by the appropriate scale factor). These data represent the energy deposited in the sensitive volume corresponding to a given readout. These quantities are brought to CaloRecHitFormatter::reconstruct and are saved with the reconstructed energy of RecHit. To make these 2 quantities persistent (like RecHit quantities) would require an update of schema (an Objectivity Database description of objects to be stored). This is not yet done, and this makes the changes implemented in classes CaloDataFrame, CaloTimeSample, CaloRecHitFormatter and the use of filling methods (get/set) for these 2 quantities in CaloReadout::lazyUpDate useless (for a while).

A few SimpleConfigurable objects were added so the user can access the newly implemented parameters and keys via their .orcarc file:
<table>
<thead>
<tr>
<th>Symbolic name</th>
<th>Type</th>
<th>Default value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hcal:DoADC</td>
<td>bool</td>
<td>1</td>
<td>QIE simulation switch</td>
</tr>
<tr>
<td>Hcal:TakeGeantHitTime</td>
<td>bool</td>
<td>1</td>
<td>GEANT Hit time switch</td>
</tr>
<tr>
<td>Hcal:FADCcalibration</td>
<td>double</td>
<td>3 pe</td>
<td>LSB for HB/HE</td>
</tr>
<tr>
<td>Hcal:FADCcalibrationHF</td>
<td>double</td>
<td>0.43 pe</td>
<td>LSB for HF</td>
</tr>
<tr>
<td>Hcal:noiseHF</td>
<td>double</td>
<td>0.125 pe</td>
<td>$\sigma_{\text{noise}}$ for HF</td>
</tr>
<tr>
<td>Hcal:enoise</td>
<td>double</td>
<td>0.0006 GeV</td>
<td>old $\sigma_{\text{noise}}$ (for DoADC=0)</td>
</tr>
<tr>
<td>Hcal:CollectionFactor</td>
<td>double</td>
<td>1.20</td>
<td>RecHit correction for 1 bucket collection</td>
</tr>
<tr>
<td>Hcal:TimePhaseAdjust</td>
<td>double</td>
<td>10 ns</td>
<td>time phase for signal collection in HB/HE</td>
</tr>
<tr>
<td>Hcal:TimePhaseAdjustHF</td>
<td>double</td>
<td>8 ns</td>
<td>time phase for signal collection in HF</td>
</tr>
<tr>
<td>Hcal:photostatHB1</td>
<td>double</td>
<td>0.00085 GeV/pe</td>
<td>GEANT Hit energy per photo electron</td>
</tr>
<tr>
<td>Hcal:photostatHB2</td>
<td>double</td>
<td>0.00068 GeV/pe</td>
<td>idem</td>
</tr>
<tr>
<td>Hcal:photostatHB3</td>
<td>double</td>
<td>0.00045 GeV/pe</td>
<td>idem</td>
</tr>
<tr>
<td>Hcal:photostatHB4</td>
<td>double</td>
<td>0.00067 GeV/pe</td>
<td>idem</td>
</tr>
<tr>
<td>Hcal:photostatHE1</td>
<td>double</td>
<td>0.00064 GeV/pe</td>
<td>idem</td>
</tr>
<tr>
<td>Hcal:photostatHE2</td>
<td>double</td>
<td>0.00045 GeV/pe</td>
<td>idem</td>
</tr>
<tr>
<td>Hcal:photostatHE3</td>
<td>double</td>
<td>0.00045 GeV/pe</td>
<td>idem</td>
</tr>
</tbody>
</table>

7 Conclusion

We have updated the HCAL code in ORCA. The main improved points are:

- front-end simulation
  - photo statistics
  - signal integration
  - ADC quantization
  - HF splitting from HB/HE
  - layer 0 and 1 merging in one readout with re-weighting
- Trigger primitive generation
- Reconstructed Hit formatting

Also some improvements/modifications of the general calorimetry code are proposed.

A test bed stand alone code (toy MC) containing almost all ORCA functionality is written to have a quick guess about various parameters impact on the signal evaluation and the noise reduction.

Main results obtained with toy MC simulation in case without pileup:

- properly tuned time phase enables one to collect the signal in 1-2 time buckets with fairly stable (within ~1% of the entire signal) content within the time range of 5-6 ns, which helps to reduce to a minimum effect of the hit arrival time jitter
- 1 signal buckets collection mode looks better that 2-buckets one due to better signal/noise ratio and BCID efficiency
- baseline position in the very first ADC channel helps to suppress the noise but makes slightly worse a small signal resolution.
- from the point of view of the signal resolution, no significant difference found between case of fairly precise knowledge of the baseline and a realistic case of pedestal estimate from 3-4 presamples, mainly due to photo statistics effect

Main results obtained with toy MC simulation in case pileup is taken into account are the following:
• the situation in HB (noise > pileup) is similar to no-pileup case
• in HE 1-bucket mode still better than 2-buckets one; baseline position is irrelevant; pedestal estimate with “infinite” 64 buckets is not better than with just 3 presamples
• a simple BCID turn-on is quicker for “parallel” 64-buckets pedestal estimate than for 3-presamples one
• 1-bucket collection mode generally looks as a preferred one

8 Future plans
• to test with ORCA the results obtained with toy MC using di-jet events from simple QCD and from decays of $Z'$ with low and high masses
• to study an impact of zero-suppression cut on the jet energy resolution
• to make the code public (expected to be published in ORCA in the fall of the year, hopefully in October)

9 Acknowledgements
We would like to thank Serguei Los, Tullio Grassi, Drew Baden, Shuichi Kunori and Dan Green for advices and fruitful discussions.

References
10 Introduction

In the Appendix we consider the case of the relatively long signal in contrast to the short signal case discussed in the main part of this Note.

11 Parameters of the Long Signal

- **Scintillator + wave-length shifter**
  
  \[ f_{s}(t) = \exp(-t/\tau_s), \quad \tau_s = 11 \text{ ns} \]

- **HPD**
  
  \[ f_{HPD}(t) = 1.0 + (t/\tau_{HPD}), \quad \tau_{HPD} = 10 \text{ ns} \]

- **Preamplifier**
  
  \[ f_{p}(t) = t \cdot \exp(-t/\tau_p), \quad \tau_p = 25 \text{ ns} \]

Figure 43: Signal shape time components.

Figure 44: Shape of the HCAL response to the instantaneous deposit of energy.

Figure 45: Maximal fraction of the signal collected in 2 time buckets versus that collected in 1 time bucket. Every single filled circle corresponds to a value of the time phase ranging from 0 to 24 ns.

Figure 46: Same as Fig. 45, except for the signal collected in 2 time buckets vs that collected in 3 buckets.

Formulae which parameterize the response of each component, along with the corresponding time constants, are shown in Fig. 43. The resulting convoluted shape has a peak position at 32 ns after the energy is deposited in the calorimeter, see Fig. 44. We used the toy MC to tune the time phase for optimal signal collection for various number of time buckets. Fig. 45 shows a comparison of the signal fraction collected in 1 and 2 time buckets as a function of the adjustable time phase. A similar comparison is presented in Fig. 46 for 3 versus 2 time buckets.
In Figs. 45 and 46, each filled circle corresponds to a certain value of time phase, with a step-size of 1 ns. There are 25 circles and thus the plot corresponds to a full turn round from 0 to 24 ns counterclockwise. It is clearly seen that with 2 time buckets collection mode, one is able to get 90-91% of the entire signal within ≈ 6 ns, while the 3-buckets collection mode has about 12 ns time of phase range in which the variation of the collected signal does not exceed 1% (the upper part of the curve in Fig. 46). The one time bucket collection mode has a region of stability of about 7-8 ns where the fraction of the signal collected is as high as 58-59% (the rightmost side of the curve in Fig. 45). In the rest of this note, when we use 1-, 2- or 3-buckets collection modes, we correspondingly correct for the fraction of the signal collected ∼ 58%, 90% and 99% respectively.

12 Results of Toy Simulation

Here we use the same set of parameters and definitions as in Section 5.

12.1 Noise

In this Subsection we consider the response of the HCAL readout to noise without any signal, just noise-induced “energy”.

We assume the baseline position is precisely known (we have been told this is possible) even though it “floats” within the chosen ADC channel. We use the term “presample” to denote a time bucket preceding the signal collection buckets.

Fig. 47 shows the reconstructed energy distributions for various baseline positions. The quantization effect due to the ADC is smeared out, because the “floating” baseline/pedestal is assumed to be known and is subtracted precisely. Fig. 48 is similar to Fig. 47, except the baseline is estimated from 3 presamples. The quantization effect is now fairly pronounced as both the signal and the subtracted pedestal are quantized and the resulting shapes are thus no longer smooth. In both Figs. 47 and 48 the difference between the “no ADC” case and all the other cases is not very significant, except for the case of 2 signal + 3 pedestal buckets with the baseline in the 1st ADC channel.

Another (“integrated”) representation of a single readout response to noise in the form of readout occupancy versus energy cut-off, is shown in Figs. 49 and 50 respectively. We use BCID in all that follows when we calculate occupancy rates.

As we will discuss in Subsection 12.2 (Fig. 60), there is a non-linearity in the energy evaluation when the baseline is set in the 1st ADC channel. In what follows (from Figs. 49,50 on), we apply corrections to compensate this kind
Figure 49: Another representation of what is plotted in Fig. 47, now in terms of readout occupancy versus reconstructed energy cut-off.

Figure 50: Same as Fig. 49, except the pedestal is not known precisely and is estimated from 3 presamples.

Figure 51: Same as Fig. 47, except the baseline set in the 2d ADC channel, for various signal collection modes.

Figure 52: Same as Fig. 49, except the baseline is set in the 2d ADC channel, for various signal collection modes (as in Fig. 51).
Figure 53: HCAL readout occupancy due to pure noise versus $E_T$ cutoff and pseudorapidity for signal evaluation with 2 signal + 3 pedestal buckets.

Figure 54: Same as Fig. 53, except the signal evaluation is done with 2 signal + 0 pedestal buckets (see details in the text).

Figure 55: HCAL readout occupancy due to pure noise versus $E_T$ cutoff when the signal evaluation is done with 2 signal + 3 pedestal buckets.

Figure 56: Same as Fig. 55, except the signal evaluation is done with 2 signal + 0 pedestal buckets (see details in the text).
of non-linearity in all occupancy plots.

One sees in Figs. 50 that for a simple signal evaluation pattern (2 signal + 3 pedestal buckets) there is some additional reduction of the noise if the baseline is set in the very first ADC channel. Effectively it means that noise is not allowed to fluctuate to the “left” - to the negative values and thus it is more unlikely that there will be negative values, with respect to 2 signal buckets in 3 pedestal buckets (the pedestal buckets enter the signal evaluation sum with negative weights). This is not the case if pedestal is evaluated precisely, see Fig.49.

If one fixes the default position of the baseline in the 2d ADC channel and looks at the various signal collection modes, one sees that the 3-buckets collection mode gathers more noise that the 2-buckets one, as seen in Figs. 51 and 52.

For completeness we plot 3-D distributions of the noise occupancy for a single readout as a function of both pseudorapidity and \( E_T \) cutoff in Figs. 53 and 54. In Fig.54 the results are calculated with the pedestal estimated as an average value of 20 digitized measurements, which is probably close to how the “precise” baseline would be implemented in practice. Figs. 55 and 56 present some selected slices from Figs. 53 and 54 at \( \eta =0,1,2,3 \). The curves show just the angular suppression of the noise, whose value is defined in terms of energy, not \( E_T \).

### 12.2 Signal + Noise

Now let us see what happens when a fixed-energy signal is injected in a single readout (more precisely, in a tower connected to a single readout). The choice of parameters is the standard set, as discussed in the preamble to Section 5.

![Figure 57](image)

Figure 57: Resolution of the readout response as a function of the number of pedestal buckets for various injected energies and for various signal collection modes. The filled symbols (circles, squares and triangles) denote the 2-buckets collection mode, while the open symbols denote the 3-buckets collection mode. The circles stand for Gaussian noise plus QIE integration, squares : as circles plus ADC quantization, triangles : as squares plus photo statistics effect.

Fig. 57 shows the decomposition of various factors affecting the resolution for injected energies of 1, 3, and 10 GeV. In this particular simulation the baseline is set fixed in the middle of the 6th ADC channel (from the previous Subsection we know that the difference between having the baseline in 2d and 6th channel is insignificant). One may notice that the photo statistics effect plays a decisive role for relatively high energies (here 10 GeV plot, rightmost one of three). The 3-buckets collection mode looks worse for relatively low (not very much higher than the noise level, hence a few GeV) energies, while in case of a 10 GeV signal the difference between the different collection modes is not drastic, as the photo statistics dominates over the noise contribution. This is in agreement with the conclusion made in the previous Subsection on pure noise considerations, that the 3-buckets mode is worse than the 2-buckets one. A slightly different representation of similar data is shown in Figs. 58, 59, where “0” pedestal buckets bin is added and putting the baseline in the 1st ADC channel case is considered. Here we
should emphasize that the latter causes some non-linearity, as in this case a systematic shift of the reconstructed signal energy is observed, as shown in Fig. 60 both for the 2 signal + 3 pedestal buckets collection mode (circles) and 2 signal + 0 pedestal (i.e. when the position of the baseline is known precisely) buckets (squares). As we already mentioned in a previous Subsection, we apply corrections for this nonlinearity whenever we evaluate the readout occupancy when the baseline is in the 1st channel.

Figure 58: Same as Fig. 57, except the “0” pedestal buckets analysis is added and the case of where the baseline in the 1st ADC channel is plotted.

Figure 59: Same as Fig. 58, except for 10 GeV.

Figure 60: Average reconstructed energy versus injected energy for the 2 signal + 0 pedestal buckets collection mode (squares) and for 2 signal + 3 pedestal buckets collection mode (circles) when the baseline is set in the 1st ADC channel. Dashed line shows an ideal linearity case.

There is just a tiny difference from the point of view of RMS/mean between the case when the baseline is in the 6th ADC channel (the noise can fluctuate freely) and in the 1st one (remember that 0 pedestal buckets is short-hand for ideal knowledge of the baseline in both Figs. 58 and 59). One again sees in these Figs. that the 2 signal buckets collection is slightly preferred to the 3 signal buckets mode and that ideal knowledge of the baseline (0 pedestal buckets bin) does not improve the resolution much. (again, the practical implementation of “ideal” knowledge of the baseline would be to take quite a large number of (in time to the signal measurement) digitized pedestal
12.3 Bunch Crossing Identification without Pileup

Fig. 61 shows the BCID efficiency as a function of the injected $E_T$ for various signal collection modes and for 4 pseudorapidity regions. Here we consider only positive values of the signal, since a zero-suppression cut is applied after BCID.

There is a purely geometrical factor $(1/\cosh(\eta))$ which defines the noise suppression with $\eta$, so the BCID is more efficient for higher pseudorapidities at a given reconstructed $E_T$. In the central HCAL region the turn-on curves are fairly slow, especially for the 3 signal buckets mode.

12.4 Noise Shape Filter (?)

A simple straightforward idea inevitably comes to mind: is it possible to suppress the noise using the difference between the pronounced shape of the signal and random nature of noise?

Suppose we tune the time phase in such a way that the signal shape provides approximately equal content in two adjacent buckets, collecting $\approx 90\%$ of the entire signal. The baseline is smeared in the 2d ADC channel and pedestal is estimated as the average value of 64 digitized measurements. At the first glance the signal formatted in such a way should be easy to distinguish from random noise. Let us try to quantitatively estimate the filtering effect for a few types of filters:
Filter 1:
- both signal buckets (B1 and B2) \( \geq 1 \) ADC count (code)
- \( \text{abs}(\text{max}(B1, B2) - \text{min}(B1, B2)) \leq 3 \) ADC counts

Filter 2:
- \( B1 \text{.and.}B2 \geq 1 \) ADC count
- \( B1 \geq 0.5*B2, \ B1 \leq 2*B2 \)

Filter 3:
- \( B1 \text{.and.}B2 \geq 2 \) ADC counts
- 2 presample and 1 postsample \( \leq \text{min}(B1, B2) \)

Fig. 62 shows the results of the application of these filters in terms of the fraction of events that passed the filters as a function of injected energy. An example of how Filter 1 affects the pure noise distribution is shown in the thumbnail plot in the lower part of the Fig. 62. It is clearly seen (where the arrow points) that noise fraction surviving the Filter 1 is \( \sim 7-8\% \).

On the other hand, by just applying a simple zero-suppression cut at 0.5 GeV one can achieve the same suppression factor, as shown in Fig. 63. Of course, one can make the filter tighter, e.g. as Filter 3 which suppresses the pure noise down to about 2 \%, as seen in Fig. 62, but the loss of the signal increases accordingly. So it does not look profitable to apply a special shape-based noise filter instead of just a simple zero-suppression cut.

Let us see if the noise filter can help improve the BCID, compared to the standard one described in Subsection 5.1. Here we slightly tighten the criteria: we use “>” instead of “\( \geq \)” in the comparison statement. The result of the calculation is presented in Fig. 64. As one might expect, BCID with a zero-suppression cut of 0.5 GeV give results that are practically identical to those from BCID + Filter 1.

The conclusion for the shape-based noise filtering can be formulated as follows: the high noise level in the HCAL plus the relatively big LSB hampers such a filtering.
As was mentioned in Section 4, our toy MC is able to simulate the contribution from pileup to a series of time buckets by using the probability $P(E)$ for a single minimum bias event to contain an energy above a threshold, see Fig. 12. Using this option, we calculated distributions of the response from a tower (using a single readout) to noise and pileup for various choices of the parameters for signal evaluation, Figs. 65 – 72. The occupancy plots, Figs. 66 - 72, are obtained with applied BCID algorithm and corrected non-linearity in case of the baseline set in the 1st ADC channel.

In Fig. 65 one can see that there is no significant difference between noise only (circles) and noise + pileup (squares). There is just a small ($<1\%$) tail on the right side of the noise + pileup distributions caused by pileup.

A similar tendency might be observed in Fig. 66, where pedestal is estimated from 3 presamples. In the latter case the shape of the distribution is affected by the ADC digitization, and it is wider because the pedestal is not as well estimated. In fact, here is rater line spectra with filled bins sometime interlaced with empty ones (triangles)- this reflects the ADC quantization effect : subtracted pedestal (unlike in Fig. 65) is quantized as it’s measured in 3 pedestal buckets. Similar effect can be seen in Fig. 48. But again, this plot shows that noise plays a more important role than pileup.

Figs. 67 and 68 show similar plots for $\eta = 1$, and are pretty much the same as those for $\eta=0$ : pileup is small compared to noise. Some angular suppression of the width of distributions is observed compared to $\eta=0$ case.

The situation changes quite drastically at higher $\eta=2$ and 2.8, as shown in Figs. 69 – 72. The noise now is (much) smaller than the pileup energy, and the total effective width of the distribution becomes bigger than that for low $\eta$ values. As a consequence of all this, the difference between the results obtained when the pedestal is estimated using 64 buckets and 3 presamples disappears if we take into account only positive $E_T$ values (as if a zero-suppression is done).

Let us again take a look at the distributions of occupancy versus the $E_T$ cutoff for the same $\eta=0$, 1, 2, 2.8 regions, Figs. 73 – 76 which show as well the distributions from the 1 signal bucket collection mode. One may see that the latter generally does not look better than 2 signal buckets collection mode. It’s worth recalling here that with an appropriate time phase tuning, 1 bucket contains $\approx 59\%$ of the entire signal, while 2 buckets - $91\%$, so the former requires bigger correction factor to be applied. Here we see again (as earlier in Subsection 12.1) that 3-buckets collection mode is slightly worse than 2-buckets one.
Figure 65: The response of a single readout to noise + pileup at $\eta=0$ for various collection modes and when the pedestal is estimated from 64 digitized time buckets.

Figure 66: Same as Fig. 65, except the pedestal is estimated from 3 presamples.

Figure 67: Same as Fig. 65, except for $\eta=1$.

Figure 68: Same as Fig. 67, except the pedestal is estimated from 3 presamples.
Figure 69: Same as Fig. 65, except for $\eta=2$.

Figure 70: Same as Fig. 69, except the pedestal is estimated from 3 presamples.

Figure 71: Same as Fig. 65, except for $\eta=2.8$.

Figure 72: Same as Fig. 71, except the pedestal is estimated from 3 presamples.
Figure 73: Another representation of what is plotted in Figs. 65 and 66, in terms of occupancy versus $E_T$ cutoff. 1 signal + 3 pedestal buckets collection mode is shown in addition by red “ice flurries”.

Figure 74: Same as Fig. 73, except for $\eta=1$.

Figure 75: Same as Fig. 73, except for $\eta=2$.

Figure 76: Same as Fig. 73, except for $\eta=2.8$.
12.6 Bunch Crossing Identification with Pileup

In this Subsection we consider BCID for the case of noise + pileup. Again, only positive reconstructed energies are considered (as if zero-suppressed). Fig. 77 presents the results of the calculation (cf. to Fig. 61). Now the 1 signal bucket collection mode slightly edges out the other ones. This needs to be confirmed by a full ORCA simulation.

Figure 77: Same as Fig. 61, except for noise + pileup case.

13 Conclusion

Main results obtained with toy MC simulation in case without pileup:

- properly tuned time phase enables one to collect the signal in 1-3 time buckets with fairly stable (within $\sim 1\%$ of the entire signal) content within the time range of 6-12 ns, which helps to reduce to a minimum effect of the hit arrival time jitter

- 2 signal buckets collection mode looks better than 3-buckets one due to better signal/noise ratio and BCID

- baseline position in the very first ADC channel helps to slightly suppress the noise, but deteriorates the small signal resolution

- from the point of view of the signal resolution, no significant difference found between ideal case of precise knowledge of the baseline and a realistic case of pedestal estimate from 3-4 presamples

Main results obtained with toy MC simulation in case pileup is taken into account are the following:
• the situation in HB (noise > pile up) is similar to no-pile up case

• in HE 2-buckets mode still better than 3-buckets; baseline position is irrelevant; pedestal estimate with “infinite” 64 buckets is not better than with just 3 presamples

• a simple BCID turn-on is quicker for “parallel” 64-buckets pedestal estimate.

• though 1-bucket collection mode generally is not better than other ones for noise reduction, still it looks slightly better for BCID