On the Challenge of Keeping ATLAS Tile Calorimeter Raw Data

V. K. TSISKARIDZE (on behalf of the ATLAS Collaboration)

*Physikalisches Institut, Albert-Ludwigs-Universität Freiburg,
3, Hermann-Herder-Str., Freiburg, 79104, Germany*

*Physics Faculty, Iv. Javakhishvili Tbilisi State University,
1 Chavchavadze ave., Tbilisi, 0128, Georgia*

E-mail: vakhtang.tsiskaridze@cern.ch

The Tile Calorimeter (TileCal) for the ATLAS experiment at the CERN Large Hadron Collider (LHC) is currently taking data with proton-proton collisions. The TileCal read-out system was initially designed to reconstruct the data in real-time and to store for each channel the signal amplitude, time and quality factor at the required high rate. This approach implied discarding 80% of the raw data that correspond to noise or small signals. Practical experience operating in this scheme with increasing rate have led to several modifications and understanding that some kind of data compression is helpful during data processing and storing.

An alternate approach is to use online reconstruction for Level 2 triggering only and to implement a data flow lossless compression scheme for further offline analysis. A new version of the lossless compression algorithm is proposed which allows to both save the complete raw data and to feed the trigger with the reconstructed signal amplitude and time. It does not increase the data flow as compared to the existing approach and the size of the data fragments transmitted is more stable. We will describe the lossless compression algorithm as a possible upgrade of the Tile data acquisition and highlight some details of the implementation. We will report on its testing and validation and on the overall performance measured on high rate tests, calibration and $\sqrt{s} = 7$ TeV proton-proton collisions runs.

Keywords: ATLAS; TileCal; LAr; Lossless Compression; Frag5; DSP.

1. Introduction

Data processing in ATLAS Tile Calorimeter^{1,2} consists of online and offline phases. Online processing is effectuated in the fixed-point arithmetic Digital Signal Processors (DSP). Operation environment limits output bandwidth to 400 (32 bits) words and the processing time to 10 μ s assuming the ATLAS Level 1 trigger rate of 100 kHz. The initial approach comprised

providing reconstructed Amplitude and Time (using Reco fragment³) for High Level Trigger (HLT) as well as storing the complete raw data for up to 8 (out of 45) selected channels for further offline analysis. An appropriate threshold is set to select the channels for which the complete raw data is stored (indicated as Frag1 fragment³). Practical experience operating with this scheme has shown that the constant increase of the collision rate requires a frequent threshold tuning. This indicated that some kind of data compression would be highly desirable during data processing and storing.

An alternate approach is to use online reconstruction for Level 2 triggering only and to implement a lossless compression scheme to record all the raw data. Saving all the raw data has several advantages. It makes possible full offline reprocessing as well as debugging and validation; may help to increase performance for physical quantities used in the analysis (jets, missing transverse energy) and enforce efforts on small signals like those produced by muons traversing the calorimeter. Saving the raw data increases the possibility to cope with Minimum Bias pile-up or unforeseen problems. Furthermore, low signals may appear helpful for exotic searches. Keeping all the raw data allows analyzing background noise and inter-channel dependencies to increase precision of the measurement. It simplifies data collection as there is no need of complicated estimates and threshold adjustments. It has already inspired various optimizations in the current working scheme. Shortly, with all the raw data the offline processing is always open to further improvements.

2. Lossless Compression

The initial idea for lossless compression was to use the fact that pulses in different channels are soundly correlated. Thus the real amount of information to store would be much less than it looks at the first sight and compression may help to improve the performance. The first version of lossless data compression comprised the pedestal compression only and soon was replaced by a more complicated and powerful compression scheme presented on CHEP09 (Praga, March 2009).⁴ According to this scheme the channels are processed in an appropriate order and the differences between samples in consecutive channels are recorded. The method substantially used the geometry of the Calorimeter as it was sensitive to the channels ordering used during the compression. It needed *no information* about the signal pulse shape and proved to be able to compress piled-up and other non-standard signals from *all* TileCal channels fitting within the bandwidth

constraint.

While the proposed algorithm was able to pack all the raw data and fit within the bandwidth limitations, it could not fit within the tight time constraints imposed by the Level 1 trigger rate of 100 kHz. To be competitive with the existing processing scheme meeting the following requirements was considered mandatory for new compression tool called later Frag5:

- compression formats should be simple enough to enable software fitting within the DSP Level 1 trigger time constraints;
- reconstructed magnitudes should be easily accessible for HLT (no unpacking should be needed);
- the energy should be reconstructed with the same precision as in the currently used fragment;
- a reliable theoretical model should exist to evaluate the algorithm performance under various circumstances;
- in case of "good" signal in *all* or almost all channels the compression should be able to fit within the bandwidth limit;
- it should be possible to compress effectively both piled-up and "unexpected" signals.

To improve DSP performance all possible precalculations were moved outside the loops, Optimal Filtering constants⁵ were rearranged to tune them for DSP commands format, specific DSP commands were used to increase the performance, such as built-in support for rounding. DSP uses a robust Software-Pipelined Loop mechanism that can significantly speed-up loops without branching. This appeared an important resource: eliminating 'if' statements from the loops and replacing them by arithmetical operations allowed to twice speed-up the code. Besides the other benefits, these tricks allowed to include energy sums Sum(Et, Ez, E) calculation into the DSP.

To simplify calculations in lossless compression tool it was decided to use some *standard*, previously measured pulse shape as a reference and store differences between samples and reference shape rather than differences between consecutive samples. Frag5, the third version of the lossless compression algorithm, fully benefits of these improvements. The closer recorded signal to the standard pulse shape assumed by the algorithm, the higher the compression of the data. Worst case analysis shows that:

- any kind of data in all channels can be recorded at 72 kHz rate at least;
- "reasonable" signals (including pile-up) with energy deposits in all channels, even significantly out of time with respect to the reference signal (± 25 ns), can be stored at 80 kHz rate;

- the standard pulse shape in all channels, even if somewhat spread in time, can be recorded at 95 kHz rate.

3. Implementation: Pros and Cons

By the end of February 2011 a fully functional version of Frag5 was successfully tested demonstrating the feasibility of the digits lossless compression approach. It was discussed within the TileCal community and later briefly reported at TIPP11 conference³ as TileCal R&D project.

Since it was clear that the increasing luminosity and energy of the LHC would pose ever more challenging conditions to the signal reconstruction, it was therefore considered very important to be ready with realistic upgrades of the Tile data acquisition that can support lossless compression schemes.

Even lacking urgent motivations (like clear physics cases or operational bottleneck in the current model) to abandon the current scheme, continuation of studies and validation tests of the lossless compression scheme was encouraged aiming at a realistic goal to have it implemented in the system.

Recent physics runs demonstrated that the current scheme requires considerable modification. No single fixed threshold could be set to select required limited amount of channels for recording under conditions of ever increasing collision rate. It was decided to change current (Frag1) format and to implement some level of data compression similar to that of Frag5. Nevertheless, new Frag1 format is still dropping a signal below the predefined threshold while it is compressing all sufficiently high pulses using two formats for small and for large signals. Quality Factor (QF) calculations are also ceased for dropped signals as they are time consuming. This means discarding potentially valuable information about the signals below the threshold. In parallel with these changes, additional layer was added to Frag5 to ensure particular handling of weak signals.

Here is a brief comparative study of Frag5 and Frag1 approaches:

Reconstruction: Frag5 uses exactly the same standard code for online amplitude reconstruction as the current scheme, which is already validated.⁵

Size: Frag5 reduces fragment size by about 10% compared to Frag1 (for Threshold = 6 ADC counts) and shows more stability in fragment size. The same remains valid comparing the new versions of Frag5 and Frag1.

Payload: Frag5 reduces data network payload and improves its stability.

Scaling: Frag1 performance is scaling with the increasing luminosity, while Frag5 is much less affected by this factor.

Example 1: Laser calibration run (Fig. 1, signal ~ 70 ADC counts, time jittering (i.e. phase variation) ~ 25 ns): Frag1 will be forced to raise the threshold up to 70 ADC counts or drop the rate down to 55 kHz (for both versions), while Frag5 has recorded all raw data at 93 kHz rate. For the TileCal Long and Extended Barrels² two peaks appeared because some Tile modules² were temporarily switched off.

Example 2: Maximal average bandwidth load (Fig. 2) for TileCal Long Barrel (the average number of inelastic interactions per bunch crossing $\mu = 15.5$, bandwidth threshold for Frag1 is 6 ADC counts, Frag5 stores all raw data). For each of 32 Tile Read-Out Driver (ROD)⁵ fragments produced by 16 Long Barrel RODs, the moving average of fragment sizes over 16 consecutive events is calculated. Among these 32 numbers corresponding to the ROD fragments the maximal is selected and the histogram created.

Fig. 1. Laser calibration run

Fig. 2. Bandwidth load

The bandwidth limit for 100 kHz Level 1 trigger rate is 398 words.

To help ensure smooth incorporation of lossless compression scheme into present and operational system and to avoid possible (if any) impact on data taking during the transition, it was proposed to make Frag5 interchangeable on-the-fly with the current scheme (Reco+Frag1). This approach called Twin Mount Framework (TMF) was successfully implemented into the system. It increases the safety, stability and overall performance of the system:

- has no impact on current data taking;
- provides the opportunity of unobstructed development and validation of lossless compression, as well as Frag1, with a realistic goal to have lossless compression safely implemented in the system.

High Rate tests have shown, that installation of Twin Mount Framework does not affect the performance, i.e. Reco+Frag1 works at the same rate both with and without TMF.

The TileCal operation experience and the outcome, as well as the evolution of currently used strategy indicate that using lossless compression is

very likely in the nearest future. Whether it happens by directly installing Frag5, which seems preferable, or gradually importing its solutions into Frag1, the mission of the first lossless compression tool will be fulfilled. Moreover, the similar approach may appear useful in the ATLAS Liquid Argon Calorimeter (LAr) which has very similar data acquisition scheme and uses the same Optimal Filtering approach for data reconstruction.⁶ The LAr data exceed significantly those of the TileCal and take up almost a half of the ATLAS event size, thus compression here may appear particularly attractive. Lossless compression approach may be applied to the LAr to study the possibility of storing *all* the raw data without increasing the currently used data-flow and with a minor (if any) increase of the currently used capacity. Should it happen we will have managed to record all the raw data for ATLAS Calorimetry (LAr and TileCal). This will be done without installing additional hardware to upgrade the subdetectors.

4. Conclusions.

A lossless compression tool (Frag5) is a fully functional software able to store all the TileCal raw data at 100 KHz rate fitting both within bandwidth and time limitations of the DSP. Some details of implementation have been highlighted and results of testing presented. Comparative study with respect to the existing approach is performed considering Frag5 as possible upgrade of current data reconstruction and storing strategy. Evolution and current state of both systems have been traced indicating the importance of compression schemes in the ATLAS Calorimetry data processing.

References

1. The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003
2. *ATLAS Tile Calorimeter: Technical Design Report*, CERN-LHCC-98-015
3. A. Valero and the ATLAS Tile Calorimeter, *Implementation and performance of the signal reconstruction in the ATLAS Hadronic Tile Calorimeter*, <http://cdsweb.cern.ch/record/1357024>
4. V. Tsiskaridze and the ATLAS Tile Calorimeter, *Lossless compression of ATLAS Tile Calorimeter raw data*, 2010 J. Phys.: Conf. Ser. 219 022046
5. G. Usai and the ATLAS Tile Calorimeter, *Signal Reconstruction of the ATLAS Hadronic Tile Calorimeter: implementation and performance*, 2011 J. Phys.: Conf. Ser. 293 012056
6. The Liquid Argon Back End Electronics collaboration, *ATLAS liquid argon calorimeter back end electronics*, 2007 JINST 2 P06002