Journal of Physics: Conference Series **119** (2008) 022030 doi:10.1088/1742-6596/119/2/022030

# Alignment data streams for the ATLAS Inner Detector

**B Pinto[1], A Amorim[1], P Pereira[1], M Elsing[2], R Hawkings[2], J Schieck[3], S Garcia[4], A Schaffer[5], H Ma[6] and A Anjos[7]**

**On behalf of the ATLAS Collaboration**

[1] SIM and University of Lisbon, Portugal
[2] CERN/ATLAS, Geneva 23, Switzerland
[3] Max-Planck Institut fr Physik, Munich, Germany
[4] Instituto de Fsica Corpuscular, Valencia, Spain
[5] LAL, Univ Paris-Sud, IN2P3/CNRS, Orsay, France
[6] Physics Department, Brookhaven National Laboratory, Upton, USA
[7] University of Wisconsin, Madison, USA

E-mail: Belmiro.Pinto@fc.ul.pt

**Abstract.** The ATLAS experiment uses a complex trigger strategy to be able to reduce the Event Filter rate output, down to a level that allows the storage and processing of these data. These concepts are described in the ATLAS Computing Model which embraces Grid paradigm. The output coming from the Event Filter consists of four main streams: physical stream, express stream, calibration stream, and diagnostic stream. The calibration stream will be transferred to the Tier-0 facilities that will provide the prompt reconstruction of this stream with a minimum latency of 8 hours, producing calibration constants of sufficient quality to allow a first-pass processing. The Inner Detector community is developing and testing an independent common calibration stream selected at the Event Filter after track reconstruction. It is composed of raw data, in byte-stream format, contained in Readout Buffers (ROBs) with hit information of the selected tracks, and it will be used to derive and update a set of calibration and alignment constants. This option was selected because it makes use of the Byte Stream Converter infrastructure and possibly gives better bandwidth usage and storage optimization. Processing is done using specialized algorithms running in the Athena framework in dedicated Tier-0 resources, and the alignment constants will be stored and distributed using the COOL conditions database infrastructure. This work is addressing in particular the alignment requirements, the needs for track and hit selection, and the performance issues.

## 1. The Atlas Computing Model

### *1.1. Input parameters and High-Level Trigger (HLT) CPU requirements*

The Large Hadron Collider (LHC) proton bunches will cross at a frequency of 40 MHz. On average about 23 inelastic proton-proton collisions will be produced at each bunch crossing. The level-1 trigger (LVL1) is responsible for the first level of event selection, reducing the initial event rate to less than 75 kHz. The HLT must reduce the event rate further down to O(100) Hz. Each selected event will have a total size of ∼1.5 MB.

The system is designed for a maximum LVL1 rate of 100 kHz in order to ensure that in case ATLAS decides to run at this LVL1 rate, the HLT/DAQ system will be able to handle it.

To estimate the size requirements of the HLT farms and the number of Sub-Farm Inputs (SFIs), we assume the use of PCs with 8 GHz dual-CPUs for the level-2 trigger (LVL2) and Event Filter (EF) processors, and computer servers with 8 GHz single-CPUs for the SFIs. With a LVL1 rate of 100 kHz, about 500 dual-CPU machines are needed for LVL2, and approximately 50–100 SFIs would be required, for an assumed input bandwidth per SFI of ∼60 MB/s. For the Event Filter, about 1500 dual-CPU machines, providing an output rate of ∼200 Hz and ∼320 MB/s, would be needed for a ∼3 kHz event-building rate and an average processing time of one second per event.

### *1.2. Grid concept and latency*

Grid implies a high degree of decentralization. A Tier-0 facility, at CERN, will be responsible for archivation and distribution of primary RAW data received from the EF, and will also provide the prompt reconstruction of the calibration and express streams, first-pass processing of the primary event stream and final distribution of the derived datasets (Event Summary Data (ESD), primary Analysis Object Data (AOD) and Tag data (TAG) sets) to the Tier-1 facilities.

The Tier-0 streaming baseline model includes four basic streams coming from the EF:

- the primary physics stream containing all physics events;
- an express stream containing a subset of events (∼5% of the full data);
- the calibration stream;
- the diagnostic stream with pathological events.

The calibration stream will produce calibrations of sufficient quality to allow a useful first-pass processing of the main stream with minimum latency. At 1.5 MB per event, each Sub-Farm Output (SFO) at 4Hz fills a 2 GB file with ∼1250 events every 5 minutes. This defines the minimum latency to start the processing of any stream. The latency of the primary stream is defined by the necessary time to have calibration, alignment, and other conditions data available on the Tier-0 processors.

The current goal is to be able to reconstruct the express and calibration streams within 8 hours and the primary data stream reconstruction ("prompt" reconstruction) beginning in 24 hours. It was not yet demonstrated that this goal will be achievable.

## 2. Detector readout parameters distribution

The ATLAS Inner Detector (ID), which tracks charged particles, consists of three sub-detectors: Pixels, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT). The Pixels sub-detector consists of semiconductor detectors with pixel readout. It is divided into two endcaps, an innermost barrel 'B-layer', and two outer barrel layers. The SCT sub-detector is built from silicon micro strip detectors. It is sub-divided into two endcaps and a barrel part. The TRT sub-detector is a tracking detector built out of straw tubes and a radiator. It's role is to identify highly-relativistic particles through transition radiation.

Regarding the distribution of the Readout Drivers (RODs), the following information was obtained from private communications with Andreas Korn (Pixel), John Hill (SCT), and Mike Hance (TRT), per detector per partition:

- Pixel: The B-Layer will have 1 ROD for 6/7 modules, e.g only a half stave. At Layer1/Disk 12/13 modules are connected to a ROD. At Layer2 26 modules are connected to a ROD.
  - Each ROD connects to one ROB.
  - In total we have 132 RODs: 24 Disk RODs, 44 B-Layer RODs, 38 L1-Layer RODs, 26 L2-Layer RODs.

- SCT: Up to 48 SCT modules are connected to each ROD (in the barrel this is exact, in the endcaps some RODs are more sparsely populated due to the more complex geometry). A set of up to 6 modules constitute a Minimum Unit of Readout (MUR).
  - There is a 1-to-1 mapping of ROD to ROB.
  - In total we have 90 ROD's: 44 Barrel and 46 Endcap ROD's.
- TRT: The barrel is divided into 32 "stacks", each side (A or C) contains 1642 straws. Each stack-side (1642 straws) is read out by 1 ROD, which we also call "logical ROD". Each endcap is also divided into 32 parts: there are 3840 straws, which takes up two full RODs (table 1).

| Sector | ECA | Barrel A | Barrel C | ECC |
|---|---|---|---|---|
| ROD's | 2 | 1 | 1 | 2 |
| Straws | 2×1920 | 1642 | 1642 | 2×1920 |

**Table 1.** Number of TRT RODs per stack.

  - There is a 1:1 mapping between RODs and ROBs.
  - In total we have 192 RODs: 64 in the barrel and 128 in the Endcap.

## 3. Inner Detector calibration stream

This stream will allow the calculation and update of a set of calibration and alignment constants, after every fill, processing the stream to accumulate residuals and other histogram quantities before prompt reconstruction. To accomplish the current goal we need to respect the following CPU processing requirements [4]:

- Pixel and SCT: 50 KSI2k [3] for derivation of silicon alignment constants;
- TRT: estimated as 20 KSI2k for derivation of TRT alignment constants;

The constants are then verified by re-reconstruction on an independent part of the calibration stream within 12 hours.

Because it is not feasible to process offline the whole primary physics stream, and furthermore due to the large data volume, needed CPU resources and time restrictions (requiring one constants set every 24h), we investigated a model where we write custom byte stream files with raw data of the tracks suitable for alignment as found in the EF. This would make use of the Byte Stream converter infrastructure and possibly give us a better bandwidth usage.

### *3.1. High-Level Trigger and Read Out Buffers data*

ATLAS decided to define the boundary between the detector readout and the data acquisition to be at the input of the ROBs. The LVL1 trigger identifies regions in the detector, so-called Regions Of Interest (RoIs), where relevant signals are marked. As described in [4], the RoI Builder (RoIB) combines the RoI information from the various parts of the LVL1 trigger and feeds it into the LVL2 Supervisor (L2SV). The L2SV allocates an event to one of the computing nodes in the LVL2 farm and is responsible for the transfers of the RoI information for the event to the allocated processor.This means that requested data fragments from selected ROBs are served to the LVL2 trigger element of the HLT system. These RoIs are then used to seed the LVL2 algorithms. This enables the algorithms to select precisely the region of the detector in which the interesting features reside and therefore the ROBs from which to request the data for analysis.

If the event is accepted, the Event data fragments for LVL2-accepted events are built, on the initiation of the Data Flow Manager (DFM), from the ROBs across a switched Ethernet

network, into a complete event by one of the Sub-Farm Interfaces (SFIs). The SFIs then serve the complete events to the second element of the HLT system, the EF.

The EF receives fully built events from the SFI, and so the complete set of the data is locally available for analysis.

All the selection processing for a given event is done in a single processor of the EF processor farm. Events not selected by the EF are deleted, and those accepted are passed to the SFO for transfer to mass storage.

It should be noted that all the data for a given event are stored in the ROBs during the LVL2 processing and, for selected events, until the event building process is completed.

Most of the element interconnection in the Data Flow system is done using standard Gigabit Ethernet network and switching technology.

*3.2. Computing model for alignment data*

Due to the described ROB properties (late RoI event information survival at the EF) and the need to feed different alignment algorithms with the same stream, it was decided to investigate the possibility to make a list of ROB's, per event, with hit information of each selected track and to use it to eliminate ROBFragment's from the EventFragment. The remaining data will be written to bytestream files.

To accomplish this, a software package named InDetCalibStream (figure 1) has been developed which runs in the ATLAS framework, ATHENA, and whose major steps for each event are:

- Basic track selection.
- getROBList method: This method makes a list of selected ROBs starting from track particles and returns the status code of the method. It iterates through track particles and their surfaces (which holds various properties defining a track on a particular surface). For each surface checks if there is simultaneously a measurement and track parameters, and in the affirmative case gets the RIO_onTrack (ROT) identifier. This class contains the Reconstruction Input Objects (RIOs) according to a track. Using the ROT identifier, and for Pixel and SCT, get the wafer identifier from which it's possible to know the ROB identifier. For TRT use the ROT to get the detector element, and through it a vector of ROBs. The method checks if the ROB is on the list, and adds/stores it if it is not.
- Gets the event using the method getEvent of the ROB Data Provider Service which takes one full event fragment per event to decode it to ROB fragments.
- robSelector method: processes the retrieved event in a "parallel" stream. This method makes a memory ByteStream from a user provided list of ROBs.
  - Iterates through event fragment, sub-detector fragment, ROS fragment and ROB fragment, and only creates new fragments for the ID sub-detectors (testId() method).
  - Use the list to eliminate ROBs without useful information.
  - Appends each of the selected objects to each other in hierarchical order.
  - Uses the copy method of the Event Format package to perform a memcpy like operation of the "new" event.
- Finally, after decoding and selecting uses the putData method, from DataWriter algorithm of EventStorage package, to write a single block of data, the "new" event, into a bytestream format file (figure 2).

*3.2.1. Track selection and calibration events from LVL2 (partial event building).* In order to be considered the track must have surfaces that contain simultaneously hits and parameters that are grouped together by the same surface they exist on.

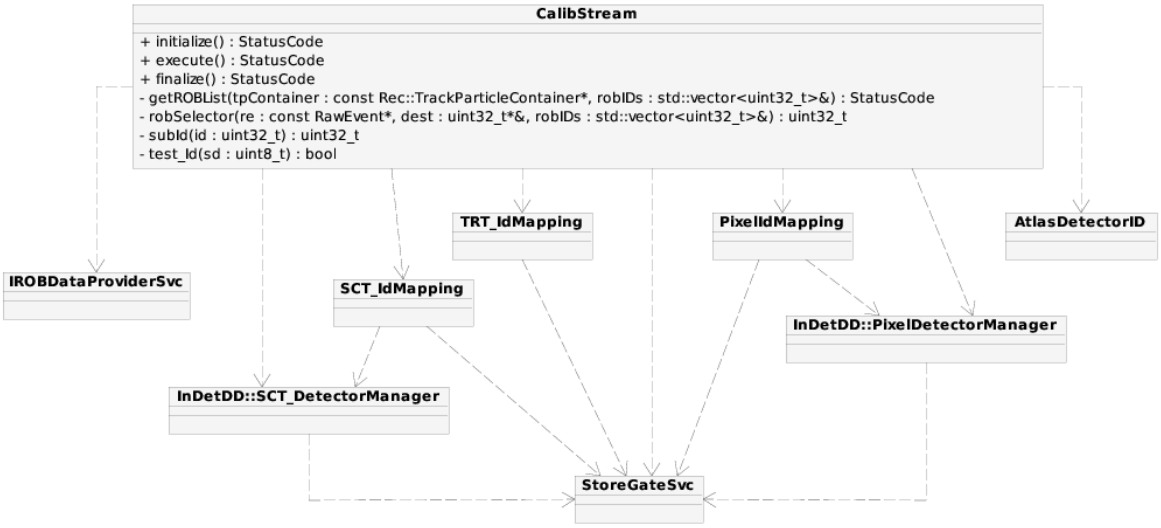


**Figure 1.** CalibStream class diagram. The arrows represent the classes' dependencies.

It is foreseen, but not yet decided, that this algorithm must process tracks from calibration events selected by LVL2. These events are flagged as physics events in the LVL1. The LVL2 Supervisor (L2SV) assigns these events to an LVL2 Processing Unit (L2PU) which decides if the quality of the events is enought for calibration (these events can be accepted or rejected for physics). The SFI will build the events and pass them to the EF where they are processed by a specialized Processing Task (PT) dedicated to calibration events. At the SFO the stream selection depends on the decision of EF calibration algorithms. The full event remains in the Event Filter until the SFO has confirmed the reception of the event to guarantee recoverability.

*3.2.2. Raw event format.* A full event is a collection of sub-detector fragments, and each sub-detector fragment is a collection of Readout Sub-system (ROS) fragments. Each ROS fragment is a collection of ROB fragments. There is a one-to-one correspondence between a ROD fragment and a ROB fragment. Each fragment, except the ROD fragment, has a header which contains the entire event formatting information. For ROD fragments, hardware considerations have led to the combination of a header and a trailer, however, the general principles are similar and it is the combination of the header and the trailer which provides the event formatting information.

All header, trailer, status and data elements are 32-bit integers. The Status element contains information about the status of the data within the fragment [5].

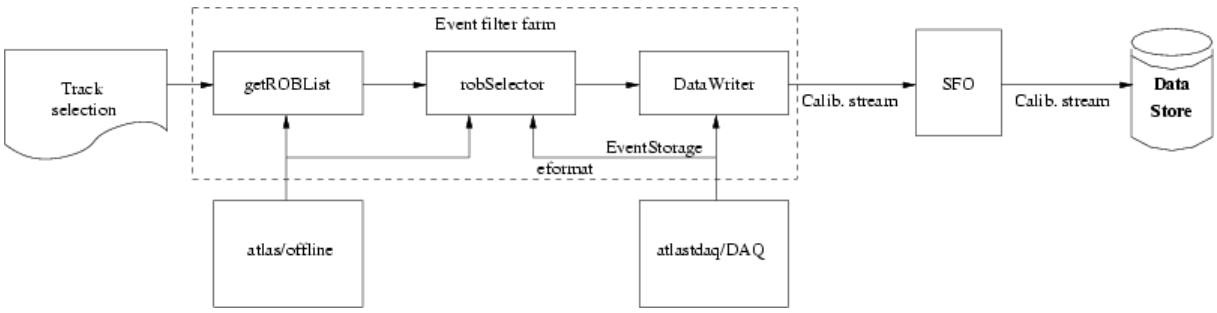


**Figure 2.** Online/offline software.

*3.3. Software test and conclusions.*
All the files used were produced with ATHENA version 11.0.4 and the test was done with 12.0.3.

Related to the detector readout parameters, and in order to have an idea of the number of ROBs involved in the passage of a particle through the ID, the test involved the use of a simulated file of single muons. For the file considered one concludes that in average we need the information contained in 7 ROBs to define a track (figure 3). This number represents ≈2% and 1% of all the number of ROBs of the ID and ATLAS detector respectively. If the predicted size of an event is 1.5 MB [2], than it is expected that to select one track per event, assuming equal distribution of data in the 960 RODs [1] of the ATLAS detector, we will have about 10KB per track. This number can be reduced if several tracks overlay on the same ROBs.

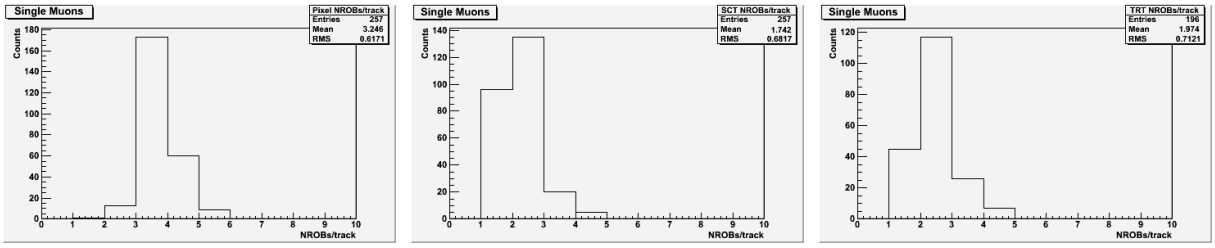


**Figure 3.** Number of ROBs per track for each ID sub-detector (Pixel, SCT and TRT). Input file of 300 generated events of single Muons with $p_T = 50$GeV and $\eta < 3.0$.

The performance and size comparing tests used a simulation file with 200 minimum bias events with an average of 10 tracks per event crossing the ID. The results and first conclusions are summarized in table 2.

From the table we can infer that the processing of an ID calibration stream designed with ROB selection will permit to have much more track information, using the same size of bandwith and storage, and the additional advantage of, with a proper track selection, getting a file with only good alignable tracks. These extra amount of tracks will permit to increase statistics, and the alignment contants quality. Finally, with these characteristics, the ROB selection will also help to respect the CPU calibration processing requirements (about 70 machines [4]).

| File | Size (MB/evt) | Rec. time (KSI2K-sec) | Time/evt (KSI2K-sec) |
|---|---|---|---|
| Original file | 0.2 | 1696 | 8 |
| Calib. Stream | 0.04 | 718 | 4 |

**Table 2.** Performance and size tests: the calibration stream is about 80% smaller and needs half the time to be reconstructed.

## 4. Acnowledgement

The authors gratefully acknowledge the support from their respective institutes. This work is partially funded by the Portuguese Foundation for Science and Technology (FCT) grant number POCI/FP/81940/2007.

## References

[1] ATLAS collaboration 2003 ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report, ATLAS TDR-016, CERN/LHCC/2003-022.
[2] ATLAS collaboration 2005 ATLAS Computing Technical Design Report, ATLAS TDR-017, CERN-LHCC-2005-022.

[3] Renshall H Change of Batch Time Units from NCU to KSI2K. `http://computingcourier.web.cern.ch/ComputingCourier/CNL_2004Nov/Batch-Time-Units.doc`
[4] Hawkings R and Gianotti F. 2005 ATLAS detector calibration model — preliminary subdetector requirements, CERN-ATL-GEN-INT-2005-001.
[5] Bee C, Francis D, Mapelli L, Mclaren R, Mornacchi G, Petersen J and Wickens F 2006 The raw event format in the ATLAS Trigger&DAQ, ATL-D-ES-0019.