

GLOBAL ARCHITECTURE FOR THE ATLAS DAQ AND TRIGGER

Part I - Functional Model

Livio P Mapelli
CERN

Abstract

This note is not meant to be a one-time document but rather a working review document which will be kept up-to-date with the evolution of the work in the ATLAS Trigger/DAQ working group, and associated R&Ds, and the modifications suggested by simulation results and technology investigations. The up-to-date version will be maintained in WWW. Hopefully it should turn into a requirement and preliminary design document, if this Trigger/DAQ scheme will be proven viable.

(First draft: June 1, 1994, L Mapelli, CERN & LBL).

1 Introduction

Global studies of the trigger rates and data volumes of LHC experiments have indicated that unprecedented challenges will be put on the Trigger and Data Acquisition system (T/DAQ). Present architecture and technologies are either inadequate or prohibitively expensive and new solutions must be studied in order to meet the required performance in a cost-effective manner. In such a demanding environment, a top-down design of the T/DAQ for ATLAS can only be approached with an accurate understanding of the experimental requirements, which, in turn, necessitates a detailed evaluation of physics processes and the definition of the detector configuration.

Although neither of the two points has reached the necessary level of maturity for the final design, the present understanding of the fundamental requirements of the LHC triggering and data acquisition is sufficient for outlining the global functional features of the T/DAQ architecture and for indicating which technologies to evaluate.

This note is intended to develop a functional model for a global Trigger and Data Acquisition system on the base of the scheme originally proposed to the ATLAS Collaboration in April 1993 [1]. Taking maximum advantage of the nature of the physics processes and in order to meet the demanding requirements of event rates and data volumes, the scheme is developed in a multi-level trigger structure driving a hierarchical data acquisition system. Although the scheme is based on a clearly defined architecture, minimum assumptions are made on technological solutions, in such a way as to maintain:

- a high level of generality to help the evaluation of system requirements and functional properties
- a suitable level of abstraction to drive the design of a behavioural model for system simulations
- a modular structure to support the investigation of the impact of different technological solutions on system elements and on the global architecture.

The preparatory work prior to the design of the ATLAS Trigger/DAQ architecture can be viewed at three levels:

1. Part I - Functional Model
2. Part II - System Parameters and Design Issues
3. Part III - R&D Workplan.

In its present form, this note covers Part I and is organised in the following Sections:

- Requirements and assumptions on page 3
- Triggering and Data Flow on page 9
- Global architecture: a functional view on page 10
- Trigger/DAQ architecture elements on page 12
- Appendix - List of abbreviations on page 18.

Part I

Functional Model

2 Requirements and assumptions

Relevant requirements are:

- event rates as a function of physics processes and energy thresholds (Section 2.1 on page 4)
- trigger rejections as a function of increasingly sophisticated selections (Section 2.2 on page 5)
- data volumes as a function of detector types and configurations and their front-end read-out schemes (Section 2.3 on page 6).

Only global average values can be specified today, insufficient for a detailed design, but adequate for defining basic assumptions and for an overall analysis of the functional model presented here. Such a model is based on a three-level trigger architecture, representing the classical splitting of the functions required from a Data Acquisition and Trigger (T/DAQ) system of a high rate HEP experiment:

1. *Level 1 trigger (T1)*: must operate the selection of potentially interesting events on the base of specific particle signatures in certain detectors at the maximum rate. Its fundamental requirement is speed, currently assumed to be obtainable only with hard-wired algorithms. The only programmability is at the parameter level, the algorithms being 'hard-coded' in ASICs.
2. *Level 2 trigger (T2)*: is where all the rejection obtainable with "local" analysis is achieved. "Local" means that access only to specific areas of the detectors is required. Matching of multi-detector signatures and overall event topology analysis are required here, currently assumed to be obtainable with a suitable combination of special and general purpose processors. Full programmability of algorithms, not requiring full event data, is foreseen at this trigger level.
3. *Level 3 trigger (T3)*: adds to the previous two levels the rejection obtainable with single event, as opposed to statistical, analysis on the whole of the detector data. Final reconstruction algorithms, considered "off-line" in today's culture, have to be run here, on the highest performance and most cost-effective processor farm architecture.

The remaining of this section describes and justifies the fundamental assumptions and the consequent requirements. A pictorial summary of them is given at the end of the section (Figure 1).

2.1 Event rates.

The ATLAS Level-1 trigger is based only on calorimetry for electrons/photons, jets and missing transverse momentum triggers and on muon trigger chambers for muon triggers. Although no detailed study has ever been attempted for a Level 1 tracking trigger, it is generally recognised that there are compelling arguments against a tracking T1, such as:

- at a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the occupancy in tracking detectors is so high that a simple, reduced granularity trigger would have no effect
- the extraction of signals from the central region at 40 MHz would constitute a very big problem for power dissipation, cooling, cabling, etc.
- triggering on calorimetry allows single (double) electron thresholds of 30 (20) GeV. There is not much point in lowering them, because real W and Z events would saturate the DAQ and would have to be rejected at higher trigger levels.

Arguments in favour of a tracking T1 exist, the strongest one being related to B-physics and the possibility of triggering on non-isolated electrons. Although more studies must be done to definitely prove that the T/DAQ architecture presently under study would cope with no problem with B events (for reason that will become apparent later, since no Region of Interests can be used from T1 to drive the T2), current investigations are directed to the following trigger scheme:

- at T1, we rely on the excellent muon detection of ATLAS and trigger on a single muon at the lowest possible momentum (6 GeV p_t , giving a rate of 8 kHz at $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ and for the full muon trigger acceptance of $|\eta| < 2.2$)
- at T2, refine the muon signature (gaining a factor 2 in rate) and search for other B-decay signatures, such as a second low momentum muon or electron, or an e^+e^- pair at very low momenta, taking advantage of the electron identification capabilities of the TRT down to 1 GeV p_t
- other signatures will be used at T2, for an estimated aggregate rate of less than 1 kHz.

Detailed analyses of prompt trigger rates for the dominating processes were done already for the LoI [2]. More accurate evaluations of T1 rates, complemented by T2 analyses, have been performed for the preparation of the Technical Proposal [3]. The results are summarised in Table 1.

Table 1 T1 rates at $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$

TRIGGER	RATE
1 or more isolated em cluster with $p_T > 30 \text{ GeV}$	20 kHz
2 or more isolated em clusters, each with $p_T > 20 \text{ GeV}$	4 kHz
1 or more muon candidate with $p_T > 20 \text{ GeV}$	4 kHz

Table 1 T1 rates at $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$

TRIGGER	RATE
2 or more muon candidates, each with $p_T > 6 \text{ GeV}$	1 kHz
1 or more jets with $p_T > 150 \text{ GeV}$	3 kHz
Large E_T^{miss}	1 kHz
Other prescaled triggers	5 kHz
TOTAL	$< \sim 38 \text{ kHz}$

The aggregate T1 rate being estimated at about 38 kHz, the T/DAQ working group define as

Requirement: the ATLAS T/DAQ system must operate at T1 rates of up to 100kHz.

All evaluations in this document are done on this assumption.

2.2 Trigger rejections.

2.2.1 Level 2.

For the preparation of the Technical Proposal, a large effort has been invested to study T2 algorithms for electron and muon based channels. The rejection of fake electrons, the dominating background from T1, is strongly based on the association of a charged track to the corresponding calorimeter cluster and preshower information. Algorithms and rejections depend critically on the inner tracker and preshower configurations, whose design, although advanced, it is not final in all the features relevant to T2 usage.

Waiting for the finalisation of the ATLAS Inner Detector and of the results of T2 tracking algorithms, I base the evaluations in this document on the

Assumption: the T2 rejection will be at least 100:1 for the dominating electron/photon T1 rate.

This assumption, supported by many independent studies on different inner detector configurations [4], is conservative and leads to the following

Requirement: the ATLAS T/DAQ system must operate at T2 rates of up to 1 kHz.

It is important to stress that the entire T/DAQ architecture is strongly based on the nature and performance of the Level 2 trigger. Needless to say that the better

we understand the T2 algorithms and the more cost-effective the full architecture will be.

2.2.2 Level 3

The present trigger performance studies do not adequately cover yet the level 3. It is currently assumed that, after the event building, the T3 will have the capability of performing detailed reconstruction using complicated algorithms similar to those used off-line (i.e. after data recording). Although it is expected that the use of rejection at T3 will increase with the knowledge of the apparatus and of the LHC physics, it would be a dangerous over-simplification to consider the Level 3 trigger as the complete substitute of today's off-line filtering. Even in a scenario in which enough computer power is available to run the full off-line reconstruction algorithms at 1 kHz, the non-immediate availability of final detector calibration constants and the need of refining reconstruction algorithms will certainly impose the use of more conservative (i.e. looser) selection cuts than the ones used after the data have been safely recorded. In other words, ATLAS will aim at using the same code for T3 selection and off-line analysis, but the values of the selection parameters may be different.

The relevant parameter at this level is reduction of the data volume to be recorded rather than event rejection. Not all trigger types require full detector data for final analysis. These observations lead to the following

Requirement: the ATLAS T3 system must achieve a data-flow reduction of $\sim x10$

Requirement: the ATLAS T3 system must allow selective data and event-summary recording for different trigger types.

2.3 Data volumes.

The present definition of the ATLAS detector configurations and their read-out schemes does not allow any better evaluation of data throughput than an "order-of-magnitude" one, at least for most sub-detectors. Although one might argue that, the DAQ system needing to be scalable and modular, we should be able to proceed towards the design of an architecture, this is only partially true. Not only realistic requirements from peak and average throughputs have a direct impact on architectural solutions in many parts of the system (T2/DAQ interface, event building, bus/links distribution, etc.), but also data distributions in the read-out network are necessary input to the simulation programs. Simulations based on behavioural models of the system are an indispensable support of the ATLAS T/DAQ design.

The evaluation of the data throughput used in this document is very rough and is only meant to

- illustrate the order-of-magnitude dimension of the ATLAS T/DAQ problem

- help to outline the basic features and parameters of the functional model proposed here for architecture studies.

Since no final specification of read-out schemes and technology has yet been done, it is necessary to make assumptions on when, how and in what form the detector information will leave the detector and enter the DAQ system. This functional model is based on the following assumptions:

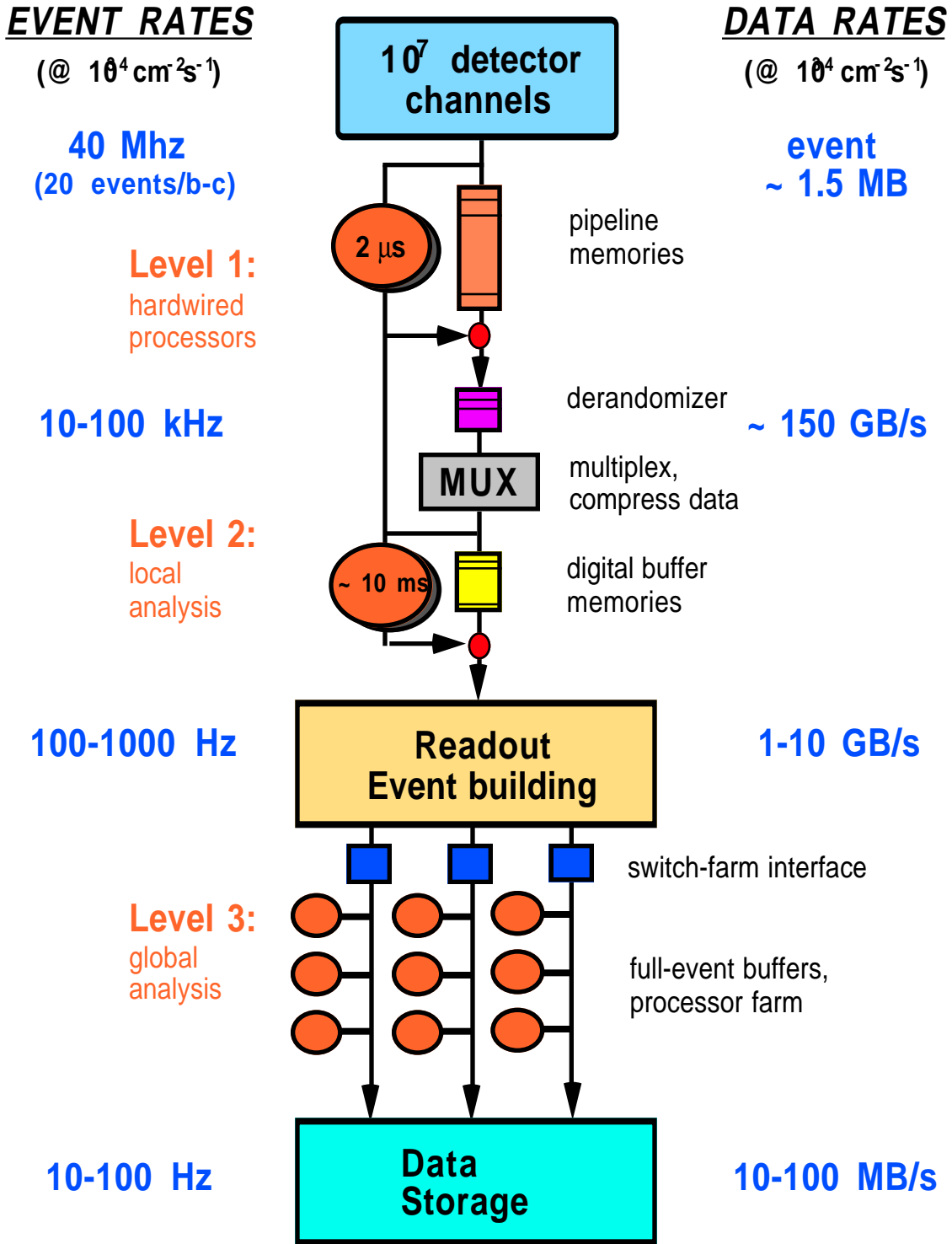
data of all the detector elements is stored locally (on detector analog or digital front-end pipelines) for the T1 latency

data of ALL detector elements are digitised, zero-suppressed and transferred from the front-end electronics to off-detector memory buffers via optical links of some technology, possibly, but not necessarily, the same for all the sub-detectors

at least from the receiver part of the optical links onwards (DAQ side), the DAQ architecture and components are the same for all the sub-detectors.

Detailed evaluation of data throughput from each detector based on the above assumptions is presently being finalised and will constitute the ultimate requirement specification of every element of the architecture [5]. Here, global average estimates are summarised in Figure 1 as indication of the overall requirements for the three-level T/DAQ system discussed in this note.

Figure 1 The ATLAS three-level trigger architecture.



3 Triggering and Data Flow

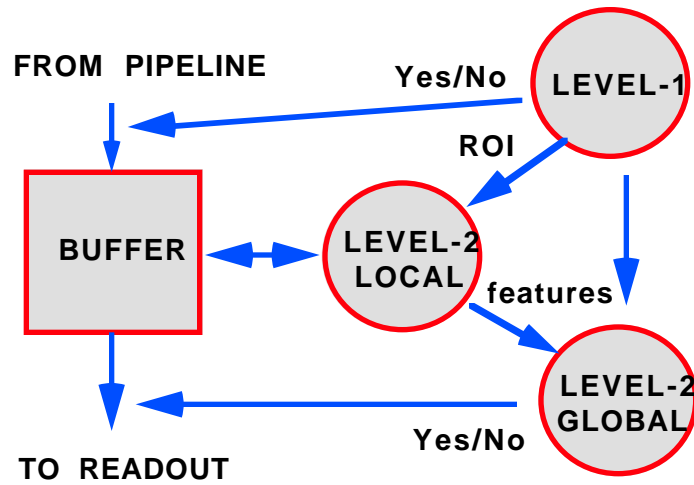
This note is intended to introduce an overall functional description of the multi-level triggering sequence and of the flow of data in the system, leaving design issues, as they are presently under investigation, to another document. Design consideration must include software aspects, today object of a dedicated DRDC project [6]. For the sake of clarity, though, sometimes the text is based on terms and concepts which strictly speaking would be more appropriate for an implementation description and might suggest design choices. This should not be taken to be the case, but rather than making the text unnecessarily heavy with a more formally correct, but less clear, jargon, I prefer to ask the reader to consider only the functional aspects of items used in the remaining of this document.

3.1 Triggering

The scheme presented here depends entirely on the possibility of driving the Level-2 trigger with information from the T1 processor, such as to avoid the need of the huge bandwidth and processing power required to handle the data at the T1 accept rate. The triggering sequence foresees:

1. the T1 system, essentially a separate box in this scheme, while processing the T1 data, keeps track of all the energy clusters and muon candidates above a given thresholds (Region-of-Interests, RoIs). The RoI thresholds are programmable and, in general, lower than the T1 thresholds.
2. for every T1 accepts, the RoI coordinates are passed to the T2 system to enable the analysis only of the data associated with potentially interesting candidates.
3. the T2 system is based on a "Local-Global" scheme (see Figure 2) [7], where the T2 processing is separated into at least two stages:
 - a set, or farm, of processors analyse in parallel "local" data from each sub-detector and each RoI. The parallelism can be defined at the sub-detector level or at the RoI level or both. This operation is sometimes referred to as "feature extraction" [8].
 - the results of the local analysis are transferred to another set, or farm, of processors which perform "global" topological analysis at the event level and issues a T2 accept/reject message.
4. the full event data of each T2 accept is assembled and transferred to a processor farm which perform full event analysis for T3 selection.

Figure 2 Functional view of the “Local-Global” T2 operation.



3.2 Data Flow

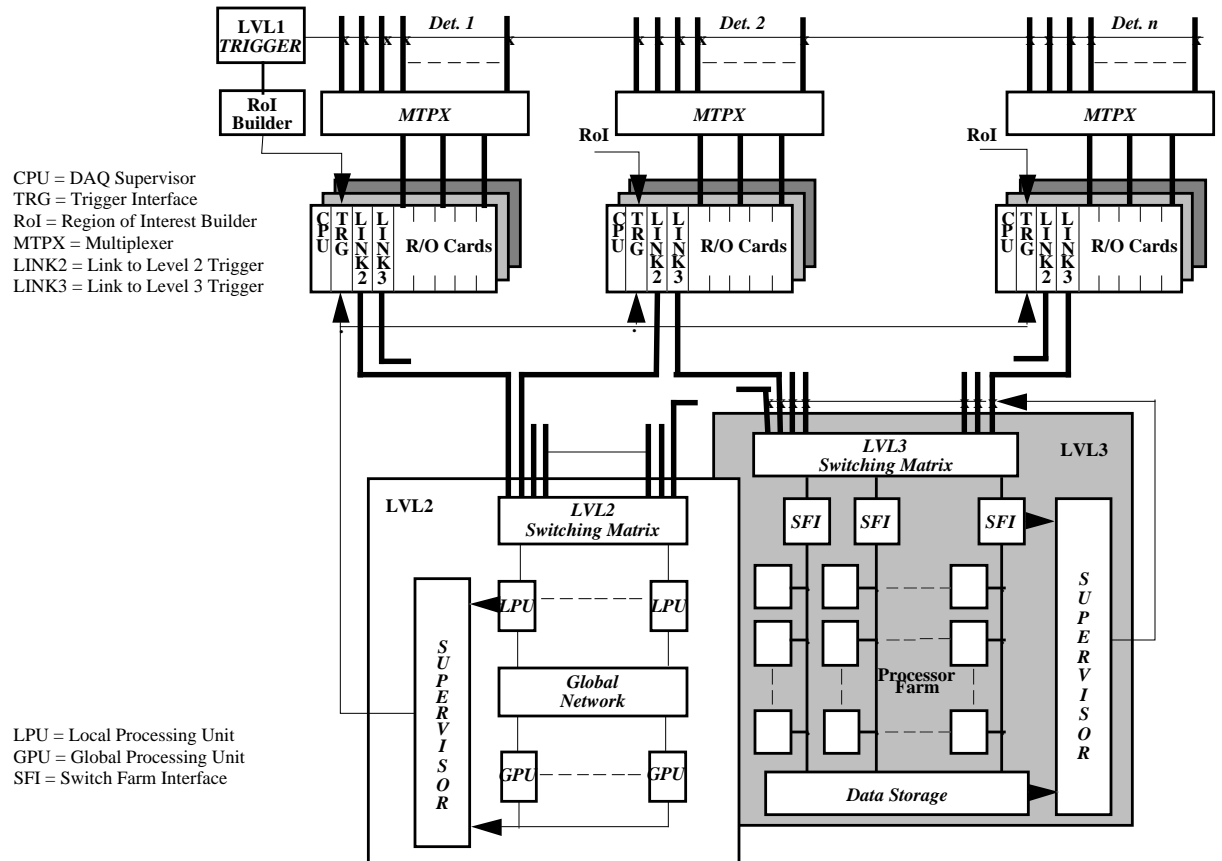
The read-out and the flow of the data are organised in the following sequence:

1. the signals of each detector electronics channel is stored locally in on-detector pipelines for each bunch crossing and for the duration of the T1 processing (T1 latency, today estimated at 2 μ s). The pipelines can be analog or digital, fixed-size or virtual
2. the fully digitised raw data from all the front-end pipelines are transferred to off-detector buffers via optical links
3. only RoI data from the detectors required by the L2 trigger algorithms are accessed by the L2 system from the memory buffers. The bulk of the data remains in the buffers for the length of the T2 latency, which is therefore responsible to a large extent to the buffer sizes (the part due to the event building latency being small since it only intervenes for the order 1% T2 accepted events)
4. the full detector data for each T2 accepted event are transferred from the memories to the event builder and T3 system for the final selection before recording.

4 Global architecture: a functional view

A generic high level model of the global Trigger/DAQ architecture based on the functionality described is in the scheme of Figure 3, as it was proposed to the ATLAS Collaboration [1] and approved as global baseline architecture.

Figure 3 Global Trigger/DAQ architecture.



One recognises the following elements:

- Level-1 trigger (T1)
- Region-of-Interest (RoI) builder
- on-detector front-end electronics
- data links and a multiplexer stage of the front-end read-out (MTPX)
- DAQ (or Read-out) crates with
 - read-out cards containing the T2 buffers
 - T1 interface
 - DAQ supervisor
 - data link(s) to the T2 system (T2 link)
 - data link(s) to the T3 system (T3 Link)
- Level-2 trigger system (T2) with
 - T2 switching fabric
 - T2 local processors (T2L)

- T2 local-global network
- T2 global processor (T2G)
- T2 supervisor
- Level-3 trigger system (T3) with
 - T3 switching fabric
 - T3 processor farm
 - T3 supervisor
 - data storage.

The functionality of each of the architecture element in this model is described in the following section 5.

5 Trigger/DAQ architecture elements

5.1 L1 trigger and Region-of-Interest

The L1 trigger system is viewed here as a highly independent processor, linked to the rest of the T/DAQ architecture via a few control lines. A detailed discussion of T1 is out of the scope of this paper (see for example ref. [9]) and I will concentrate only on the function of the control links: the T1 accept/reject and the RoI building.

5.1.1 T1 accept/reject.

When at least one of the calorimetry or muon based algorithms hardwired in the T1 processors is satisfied, the T1 system issues a global accept which is distributed to all the front-end electronics channels in the experiment [10]. This initiates the read-out process of all the channels containing data corresponding to that trigger and their transfer off the detector electronics to the buffers in the read-out cards. The correct association of a trigger accept to the data from the corresponding bunch crossing is guaranteed, in the absence of malfunctioning, by the synchronisation given by a fixed latency of the L1 trigger processing and signal distribution.

In ATLAS, we aim at a very short latency to minimise cost, size and power consumption of the front end pipelines. The current estimate, assuming that enough room will be available in a radiation protected area close to the detector (75+75 m of cabling), is 750 ns for signal distribution, 875 ns for the calorimeter trigger processing (the muon trigger is slightly faster) and 125 ns for the central trigger logic, for a total of 1750 ns. This time does not include the contribution of the signal distribution within the FE electronics, which is not expected to exceed a few bunch crossings in the slowest case. The ATLAS T/DAQ group has, therefore, defined 2 μ s as latency of the T1 trigger, leaving some redundancy to absorb the present evaluation uncertainty, at least if no major changes are introduced in the present scheme.

5.1.2 RoI building

The entire T2/DAQ architecture proposed here is based on the technical feasibility of the exploitation of the nature of particle identification - a non negligible rejection against background being achievable with local analysis based on a fraction of the event data.

In processing the Level-1 data, the T1 system keeps track of all the em and jet E_T clusters and muon candidates that defines the RoIs for T2. The RoI thresholds are programmable and, in general, lower than the thresholds used in the T1 decision. It is the task of the T1 complex to assemble the coordinates of each RoI, together with other T1 information, and broadcast them to the Trigger Interface modules in the DAQ Crates for the selection of RoI related data and transfer to the T2 system.

5.2 FE read-out, data links and multiplexers

The ATLAS detectors have a variety of requirements for their FE read-out systems. No final requirements have been specified yet neither at the detector nor at the DAQ system level for the functionality and the protocol of the front-end read-out and the transfer of data into the off-detector memory buffers in the read-out cards. It is clear that read-out and data transfer protocols can be only partially specified in a technology independent form.

The nature and technology of the FE read-out electronics, as well as the read-out protocol, are largely irrelevant to the Trigger/DAQ architecture presented here. One should not underestimate, though, the obvious advantages of a well defined common protocol and it is the aim of the ATLAS Trigger/DAQ team to achieve the maximum level of technology and protocol uniformity throughout the different detectors.

The detector specific FE electronics is out of the scope of this note, which is meant to deal with the Trigger/DAQ side of the read-out process. FE read-out here is just the primary source of digitised data.

Qualitative considerations suggest already different overall requirements for the read-out when seen from the detector or the DAQ side. In general, considerations on fault tolerance, geometrical constraints, power consumption and similar might suggest that, at least for some detectors, a high number of low bandwidth fibres is to be preferred to a smaller number of high bandwidth links. On the other hand, feasibility, cost, data load balancing and system complexity issues might suggest exactly the contrary at the DAQ/RO-cards side. Furthermore, although every detector will use optical links to transfer the T1 accepted data, it must be assumed that FE link technology will be detector specific, while there are no technical constraints imposing inhomogeneity on the DAQ side, from the RO cards onwards. On the contrary, it is crucial for a system of such size and complexity, to achieve the highest degree of uniformity.

The present scheme, therefore, foresees a multiplexer stage at the interface between the different FE technologies and the uniform DAQ side. Although no technical specification has yet been made, some possible solutions are outlined in Section 5.3.1 on page 15.

5.3 DAQ crate.

“Crates” in this architecture model have to be seen at this stage as functional units - a convenient way to house the data storage at the interface from the front-end to the rest of the Trigger/DAQ system. Although the realisation of the architecture presented here must have a real crate unit, at this stage we are not concerned about implementation aspects. This note, therefore, simply describes the functionality of this part of the architecture assuming the presence of a DAQ crate of some technology containing the following modules: a set of read-out cards, at least one processor for DAQ supervision, a trigger interface module, one or more data links to T2 and one or more data links to T3.

- *Read-out card*: receives the data from the front-end after each T1 and stores them for the length of the T2 latency. It has 4 functional elements:
 - *optical receiver*: is the input path for the data from the front-end
 - *buffer memory*: stores the data inputted via the optical receiver, allows access for RoI data transfer to the T2 system, allows access to T2 accepted data for transfer to the T3 event builder. Since it is here that the data are stored during the T2 selection process, for simplicity this buffer can be referred to as “T2 buffer”, although it has to be kept in mind that it serves the more general purpose of DAQ and event building buffer.
 - *on board processor*: for data formatting, data reduction, local monitoring, memory management and RoI data selection on request from the Trigger Interface Module. Although it is very likely that the necessary performance will require that some of the above functions, in particular the last two, will have to be implemented in hardware, it is convenient to keep some flexibility during the initial conceptual design. They will, therefore, be studied as provided by the on board CPU.
 - *“bus” interface*: to output the T2 and T3 data onto the “bus(es)” for the transfer to the T2 and T3 link respectively. Depending on throughput requirements, the “bus(es)” might be partitioned. In this case, each partition must be associated to its own T2 and T3 links.
- *Trigger Interface*: converts RoI coordinates into address of the read-out card memory where the corresponding data are stored, for the transfer to the T2 link(s). The same, or a separate but equivalent, module converts the event ID of each T2 accepted event into memory address for data transfer to the T3 link(s).
- *DAQ Supervisor*: overall local control and monitoring of the DAQ Crate, running full size, multi-task Operating System.
- *T2 Link(s)*: transfers RoI data to T2 switching fabric. One or more links per crate (up to one per each read-out card) depending on required throughput.

- *T3 Link(s)*: transfers T2 accepted data to the T3 switching fabric. One or more links per crate (up to one per each read-out card) depending on required throughput.

5.3.1 Read-out Card

The read-out card, with the T2/DAQ buffer, is the fundamental building block of the entire ATLAS T/DAQ architecture. Although this document intends to address primarily functional issues, one cannot avoid to address some implementation aspects of the read-out card in order to correctly emphasise the crucial role of the functions of this element, whose proper design is vital for the entire system.

- *Optical receiver*. In order to keep the system at an affordable size, high bandwidth is required at the input of the read-out card. Links at 1 Gb/s are currently assumed, for the transfer of the 150 GB/s of front-end data at the maximum T1 rate of 100 kHz. Once geometrical constraints and bandwidth-utilisation efficiency are taken into account, the total number of read-out cards is evaluated at about 2000 housed in about 150-200 crates. As already discussed (Section 5.2 on page 13), this global DAQ requirement imposes the existence of a multiplexer stage as interface to the detector specific front-end electronics. Before proceeding to the specification of the interface, more understanding of detector specific front-end requirements is necessary. Some possible solutions can be outlined, though:
 - a multiplexer board with several optical receivers of detector specific technology as input and one 1 Gb/s optical transmitter as output, housed in an intermediate station between the detector and the DAQ crate
 - or, customised RO cards with different optical receiver technologies for the different detectors (for low bandwidth link technologies, there would be many input links per RO card)
 - or, several, customised optical receiver boards sitting in the DAQ crate, electrically coupled to the RO card.
 In all cases, whatever solutions is adopted, the interface will have to house also detector specific processing electronics (ADCs, DSPs, filters, zero-suppression, etc.) which cannot be installed on the detector but is required before sending data to the T2 buffers in the RO cards.
- *T2 Buffer*. A multi-port memory is required for the high rate input/output (I/O). At least two fast ports are needed, one for data input at 1Gb/s and one to output the data to the T2 and T3 systems. Although only a small fraction of the input data are output, the second memory port needs to be fast in order to keep a short latency. A third, slower, port is required for monitoring. The size of event fragments (i.e. the part of the event stored within a single T2 buffer) is 1-10 kB, depending on the detector and the occupancy. The number of events to be accommodated in the buffer, and the consequent memory size per card, are determined by the combined latency of T2 data transfer, T2 algorithm processing and T3 event building. Current estimates give a combined average latency

of 1-10 ms, corresponding to a few hundred event fragments in the buffer. However, it is essential to have a safety margin and allowance must be made for tails in the latency distribution.

- “*Bus Interface*”. No technology has yet been specified for the interface between the memory buffers and the T2 and T3 switching systems. The use of the term “bus” is to indicate that the two orders of magnitude reduction of the T2 and T3 I/O requirements allows to combine data from many RO cards, so as to reduce the number of links, and therefore the size, of the T2 and T3 switching systems, which are reduced to about 200 from the total of about 2000 RO cards.

5.4 Level-2 Trigger System

The Level-2 Trigger provides extra rejection to the T1 rate before full event building is done, on the base of local analysis of a fraction of the event data and a partial set of sub-detectors. It has 5 elements: switching fabric, local processors, global processor, local-global network and supervisor.

- *Switching Fabric*. For each RoI defined by the T1 system, the corresponding data from different sub-detectors have to be combined in order to run the T2 algorithms. A switching fabric capable of connecting all possible sources (DAQ Crates) to all possible destinations (T2 Local Processors) provides parallel “RoI building”. More studies of T2 algorithms will determine the best sequence of local operations (detector-by-detector analysis, RoI detector matching). The switching fabric scheme proposed here has the necessary flexibility to accommodate either combination, and even drastic changes required by evolution of algorithms. Although full connectivity (all sources to all destinations) is available, it is unlikely that it will be necessary: some level of geographical assignment of DAQ Crates to T2L would probably be possible, with some advantage in shorter latency.
- *Local Processors (T2L)*. Their function is to analyse the data from each RoI and each relevant detector to extract the quantities required by the T2 algorithms. The analysis proceeds in parallel for each RoI and possibly for each subdetector in order to minimise the processing latency, which determines the size of the RO card memories (T2 buffers). A detailed knowledge of the algorithms and of the detectors involved is necessary in order to specify the kind of parallelism and the sequence of the processing. Nevertheless, the global nature of the algorithms suggests a sequence of the kind:
 - 1 - analysis of individual sub-detectors RoI data (sub-detector parallelism), followed by
 - 2 - analysis of detector matching for each RoI (RoI parallelism), followed by
 - 3 - transfer of results to the global processors.
- *Global Processors (T2G)*. Their function is to receive the results of the local processing and combine them to give a global T2 accept-reject. All the T2L processing results for a given event have to be transferred to one T2G. In order to minimise the T2 latency, several global processor will analyse events in parallel. The global decision is distributed to the DAQ crates for memory flushing (reject) or data transfer to the T3 system (accept).

- *Local-global network.* Since only summary data from the local analysis have to be transferred to the global processors, bandwidth requirements on the local-global network are expected to be relatively light (order of 100 MB/s).
- *T2 Supervisor.* It controls all operations of the above T2 elements.

The nature of the processing required by the T2 triggering might suggest a slightly more complicated architecture than this two-level local-global scheme. Given the present T1 algorithms, the first step of T2 processing for photon/electron based triggers, for example, would be to validate the calorimeter energy deposition as compatible with an electromagnetic shower, i.e. analysis of the energy cluster with better calibration constants and full calorimeter granularity. This independently from, and prior to, any other detector analysis (tracking or preshower). The analysis of full granularity calorimeter data gives a one-order-of-magnitude rejection. A scheme supporting a two step local analysis of this kind (calo first, followed by track and preshower matching) would give two major advantages:

- lower T1 rate for tracking and preshower analysis
- more accurate definition of the RoI for tracking data transfer.

The combination of the two gives a substantial reduction in the bandwidth required at all levels of the T2 data I/O (DAQ crate, T2 links and switching fabric).

5.5 Level-3 Trigger System

For those events which satisfy at least one T2 algorithm, the full data of all the sub-detectors must be transferred from the RO card memories to the memory of one processor of the T3 farm. This requirement is based on the assumption that the T2 system will provide all the rejection obtainable with local analysis of RoI objects, such that further rejection can only be achieved only with in-depth analysis of the fully reconstructed events. The T3 system consists of 4 elements: event builder, switch-farm interfaces, processor farm, data storage and T3 supervisor.

- *Event Builder (EB).* A switching fabric which connects all possible sources (DAQ Crates) to all possible destinations (T3 Processors) provides parallel event building. In the architecture presented here, the EB switching fabric has the same functionality of the T2 switching fabric, with the only exception that here all input ports must be connected to all the output ones. Although there would be considerable practical advantages in using the same technology for the two fabrics, that will depend on the bandwidth and switching speed requirements.
- *Switch-Farm Interface.* This element in the architecture is required at each output port of the EB to assemble data into full events and distribute them to the first available processors in their sub-farm (see below).
- *T3 Processor farm.* It has the function of minimising the amount of data to be permanently recorded by:
 - rejecting uninteresting events on the base of full event analysis
 - formatting and reducing the total amount of data of an event.

An organisation of T3 processors into a set of sub-farms (many processors attached to each EB output port and served by one SFI) might be convenient, depending on the optimization of I/O vs processing time [11]. Enough CPU power must be present in the T3 farm to reject all events which are not required by statistical physics analysis. Several events must be processed in parallel. Differently from the T2 system, processing latency is not an issue. Therefore, parallel processing is not required at the single event level, but at the level of one event per processor. Nevertheless, parallel processors might be used in case that would constitute a more cost-effective solution.

- *Data Storage*. Whatever will be the available technology to best suit the ATLAS Data Model for recording and analysis.
- *T3 Supervisor*. It controls all operations of the above T3 elements.

6 Appendix - List of abbreviations

- DAQ = Data Acquisition
- T/DAQ = Trigger and DAQ
- T1 = Level 1 trigger
- T2 = Level 2 trigger
- T2L = Local Level 2 trigger
- T2G = Global Level 2 trigger
- RoI = Region of Interest
- T3 = Level 3 trigger
- EB = Event Builder
- RO = Read-Out
- Gb/s = Gigabit/second
- GB/s = GigaByte/second

Acknowledgements

The ideas developed in this note have substantially benefited from the work of, and discussions with, many colleagues in the ATLAS Trigger/DAQ working group and in the RD13 project.

References

- [1] L Mapelli, ATLAS DAQ Requirement Model, ATLAS Meeting Minutes DAQ-TR-148, (1993).
- [2] ATLAS Letter of Intent, CERN/LHCC/92-4, LHCC/12 (1992).
- [3] J Bystricky et al., ATLAS Internal Note DAQ-NO-28 (1994).
L Nisati et al., ATLAS Internal Note DAQ-NO-38 (1994).
- [4] D Froidevaux et al., ATLAS Internal Note INDET-NO-17 (1992)
R Hawkings et al., ATLAS Internal Note INDET-NO-52 (1994).
- [5] L Mapelli, Global Architecture for the ATLAS DAQ and Trigger, Part II - System parameters and Design issues, ATLAS DAQ Note in preparation.
- [6] RD13 Status Report, CERN/DRDC 94-24.
- [7] D Crosetto et al., in Proc. LHC Workshop, Aachen, 1990, CERN 90-10, p. 145.

- [8] RD11 Status Report, CERN/DRDC 94-20.
- [9] RD27 Status Report, CERN/DRDC 93-32
 - C Bohm et al., ATLAS Internal Note DAQ-NO-32 (1994)
 - R Cardarelli et al., ATLAS Internal Note DAQ-NO-39 (1994).
- [10] BG Taylor, Timing trigger and control distribution for LHC experiments, Working Review Document, CERN/ECP-RA.
- [11] G Mornacchi, ATLAS Internal Note DAQ-NO-23 (1994).

