Chapter 8

Trigger, data acquisition, and controls

2008 JINST 3 S08003

8.1 Introduction to event selection and data acquisition

As described in chapter 1, the trigger consists of three levels of event selection: Level-1 (L1), Level-2 (L2), and event filter. The L2 and event filter together form the High-Level Trigger (HLT). The L1 trigger is implemented using custom-made electronics, while the HLT is almost entirely based on commercially available computers and networking hardware. A block diagram of the trigger and data acquisition systems is shown in figure 8.1.

The L1 trigger searches for signatures from high-p_T muons, electrons/photons, jets, and τ-leptons decaying into hadrons. It also selects events with large missing transverse energy (E_T^{miss}) and large total transverse energy. The L1 trigger uses reduced-granularity information from a subset of detectors: the Resistive Plate Chambers (RPC) and Thin-Gap Chambers (TGC) for high-p_T muons, and all the calorimeter sub-systems for electromagnetic clusters, jets, τ-leptons, E_T^{miss}, and large total transverse energy. The maximum L1 accept rate which the detector readout systems can handle is 75 kHz (upgradeable to 100 kHz), and the L1 decision must reach the front-end electronics within 2.5 μs after the bunch-crossing with which it is associated.

The L2 trigger is seeded by Regions-of-Interest (RoI's). These are regions of the detector where the L1 trigger has identified possible trigger objects within the event. The L2 trigger uses RoI information on coordinates, energy, and type of signatures to limit the amount of data which must be transferred from the detector readout. The L2 trigger reduces the event rate to below 3.5 kHz, with an average event processing time of approximately 40 ms.

The event filter uses offline analysis procedures on fully-built events to further select events down to a rate which can be recorded for subsequent offline analysis. It reduces the event rate to approximately 200 Hz, with an average event processing time of order four seconds.

The HLT algorithms use the full granularity and precision of calorimeter and muon chamber data, as well as the data from the inner detector, to refine the trigger selections. Better information on energy deposition improves the threshold cuts, while track reconstruction in the inner detector significantly enhances the particle identification (for example distinguishing between electrons and photons). The event selection at both L1 and L2 primarily uses inclusive criteria, for example high-E_T objects above defined thresholds. One exception is the L2 selection of events containing the decay of a B-hadron, which requires the reconstruction of exclusive decays into particles with low momentum.

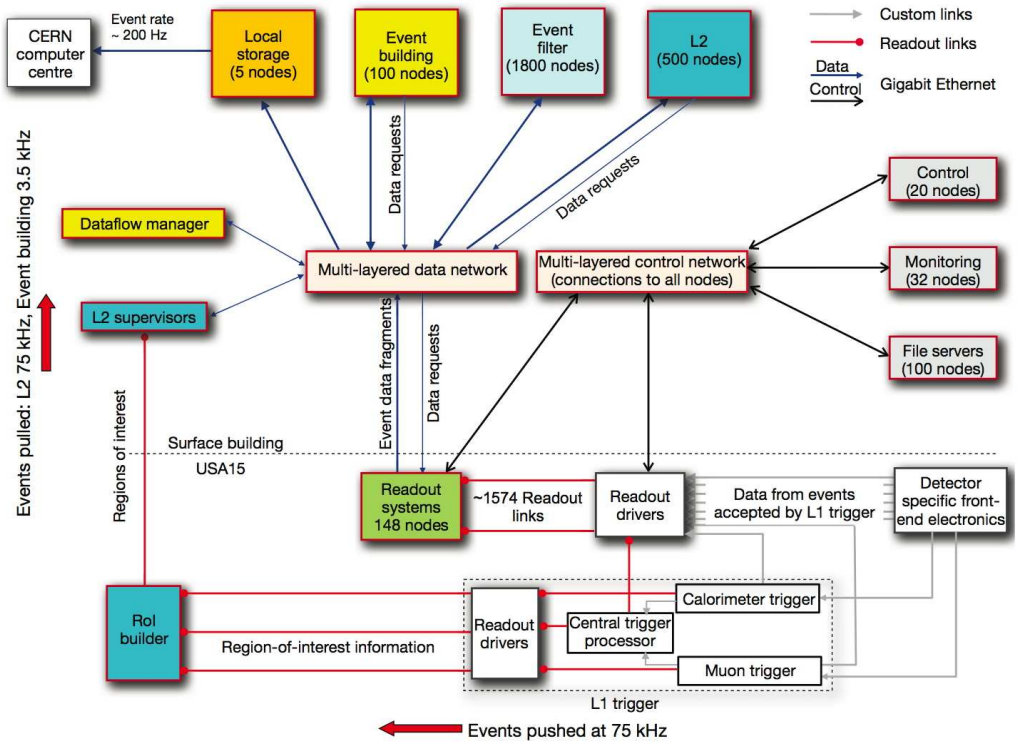

Figure 8.1: Block diagram of the ATLAS trigger and data acquisition systems (see sections 8.2 and 8.3 for further details).

The data acquisition system (DAQ) receives and buffers the event data from the detector-specific readout electronics at the L1 trigger rate. The data transmission is performed over point-to-point Readout Links (ROL's). It transmits to the L2 trigger any data requested by the trigger (typically the data corresponding to RoI's) and, for those events fulfilling the L2 selection criteria, event-building is performed. The assembled events are then moved by the data acquisition system to the event filter, and the events selected there are moved to permanent event storage.

In addition to controlling movement of data down the trigger selection chain, the data acquisition system also provides for the configuration, control and monitoring of the ATLAS detector during data-taking. Supervision of the detector hardware (gas systems, power-supply voltages, etc.) is provided by the Detector Control System (DCS).

Section 8.2 presents the design, algorithms, and implementation of the L1 trigger. The HLT and data acquisition system are described in section 8.3, which gives an overview of the flow of events through the system, a brief description of the main system components, and the performance expected for initial operations. The implementation and capabilities of the DAQ/HLT are presented in section 8.4. Finally, the detector control system is described in section 8.5.

2008 JINST 3 S08003

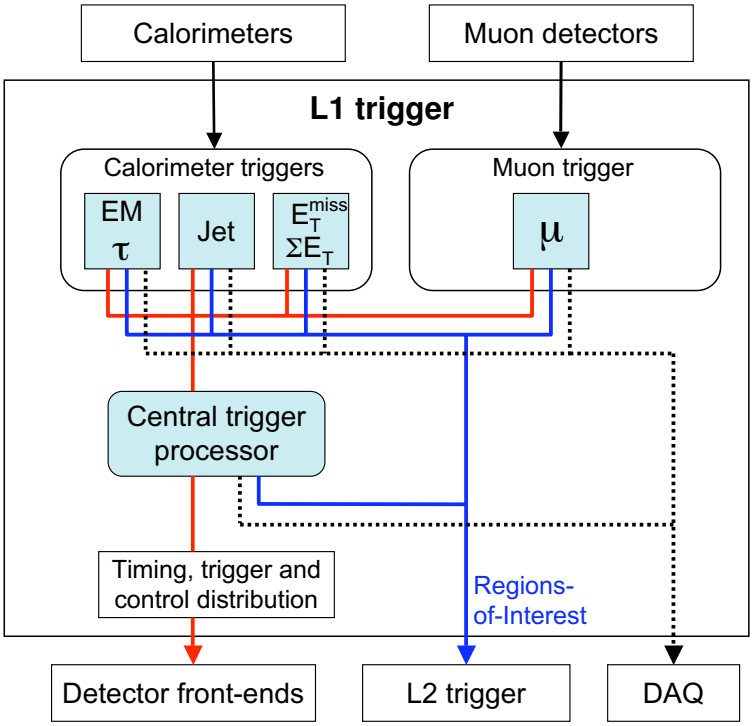

Figure 8.2: Block diagram of the L1 trigger. The overall L1 accept decision is made by the central trigger processor, taking input from calorimeter and muon trigger results. The paths to the detector front-ends, L2 trigger, and data acquisition system are shown from left to right in red, blue and black, respectively.

8.2 The L1 trigger

The flow of the L1 trigger is shown in figure 8.2. It performs the initial event selection based on information from the calorimeters and muon detectors. The calorimeter selection is based on information from all the calorimeters (electromagnetic and hadronic; barrel, end-cap and forward). The L1 Calorimeter Trigger (L1Calo) aims to identify high-E_T objects such as electrons and photons, jets, and τ-leptons decaying into hadrons, as well as events with large E_T^{miss} and large total transverse energy. A trigger on the scalar sum of jet transverse energies is also available. For the electron/photon and τ triggers, isolation can be required. Isolation implies that the energetic particle must have a minimum angular separation from any significant energy deposit in the same trigger. The information for each bunch-crossing used in the L1 trigger decision is the multiplicity of hits for 4 to 16 programmable E_T thresholds per object type.

The L1 muon trigger is based on signals in the muon trigger chambers: RPC's in the barrel and TGC's in the end-caps. The trigger searches for patterns of hits consistent with high-p_T muons originating from the interaction region. The logic provides six independently-programmable p_T thresholds. The information for each bunch-crossing used in the L1 trigger decision is the multiplicity of muons for each of the p_T thresholds. Muons are not double-counted across the different thresholds.

2008 JINST 3 S08003

The overall L1 accept decision is made by the Central Trigger Processor (CTP), which combines the information for different object types. Trigger menus can be programmed with up to 256 distinct items, each item being a combination of requirements on the input data. The trigger decision, together with the 40.08 MHz clock and other signals, is distributed to the detector front-end and readout systems via the Timing, Trigger and Control (TTC) system, using an optical-broadcast network.

While the L1 trigger decision is based only on the multiplicity of trigger objects (or flags indicating which thresholds were passed, for global quantities), information about the geometric location of trigger objects is retained in the muon and calorimeter trigger processors. Upon the event being accepted by the L1 trigger, this information is sent as RoI's to the L2 trigger (see section 8.3.6), where it is used to seed the selection performed by the HLT.

An essential function of the L1 trigger is unambiguous identification of the bunch-crossing of interest. The very short (25 ns) bunch-crossing interval makes this a challenging task. In the case of the muon trigger, the physical size of the muon spectrometer implies times-of-flight exceeding the bunch-crossing interval. For the calorimeter trigger, a serious complication is that the width of the calorimeter signals extends over many (typically four) bunch-crossings.

While the trigger decision is being formed, the information for all detector channels has to be retained in pipeline memories. These memories are contained in custom electronics placed on or near the detector, where often radiation levels are high and access is difficult. In the interest of cost and reliability, it is desirable to keep the pipeline length as short as possible. The L1 latency, which is the time from the proton-proton collision until the L1 trigger decision, must therefore be kept as short as possible. The design of the trigger and front-end systems requires the L1 latency to be less than 2.5 μs, with a target latency of 2.0 μs, leaving 0.5 μs contingency. About 1 μs of this time is accounted for by cable-propagation delays alone. To achieve this aim, the L1 trigger is implemented as a system of purpose-built hardware processors, which are described in more detail below.

8.2.1 Calorimeter trigger

L1Calo [227] is a pipelined digital system designed to work with about 7000 analogue trigger towers of reduced granularity (0.1×0.1 in $\Delta\eta \times \Delta\phi$ in most parts, but larger at higher $|\eta|$) from the electromagnetic and hadronic calorimeters. It sends the results for each LHC bunch-crossing to the CTP approximately 1.5 μs after the event occurs, resulting in a total latency for the L1Calo chain of about 2.1 μs, well within the allowed envelope.

The L1Calo system is located off-detector in the service cavern USA15. Its architecture, shown in figure 8.3, consists of three main sub-systems. The pre-processor digitises the analogue input signals, then uses a digital filter to associate them with specific bunch-crossings. It uses a look-up table to produce the transverse-energy values used for the trigger algorithms. The data are then transmitted to both the Cluster Processor (CP) and Jet/Energy-sum Processor (JEP) subsystems in parallel. The CP sub-system identifies electron/photon and τ-lepton candidates with E_T above the corresponding programmable threshold and satisfying, if required, certain isolation criteria. The JEP receives jet trigger elements, which are 0.2×0.2 sums in $\Delta\eta \times \Delta\phi$, and uses these to identify jets and to produce global sums of scalar and missing transverse energy. Both processors count the multiplicities of the different types of trigger objects. The CP and JEP send these

2008 JINST 3 S08003

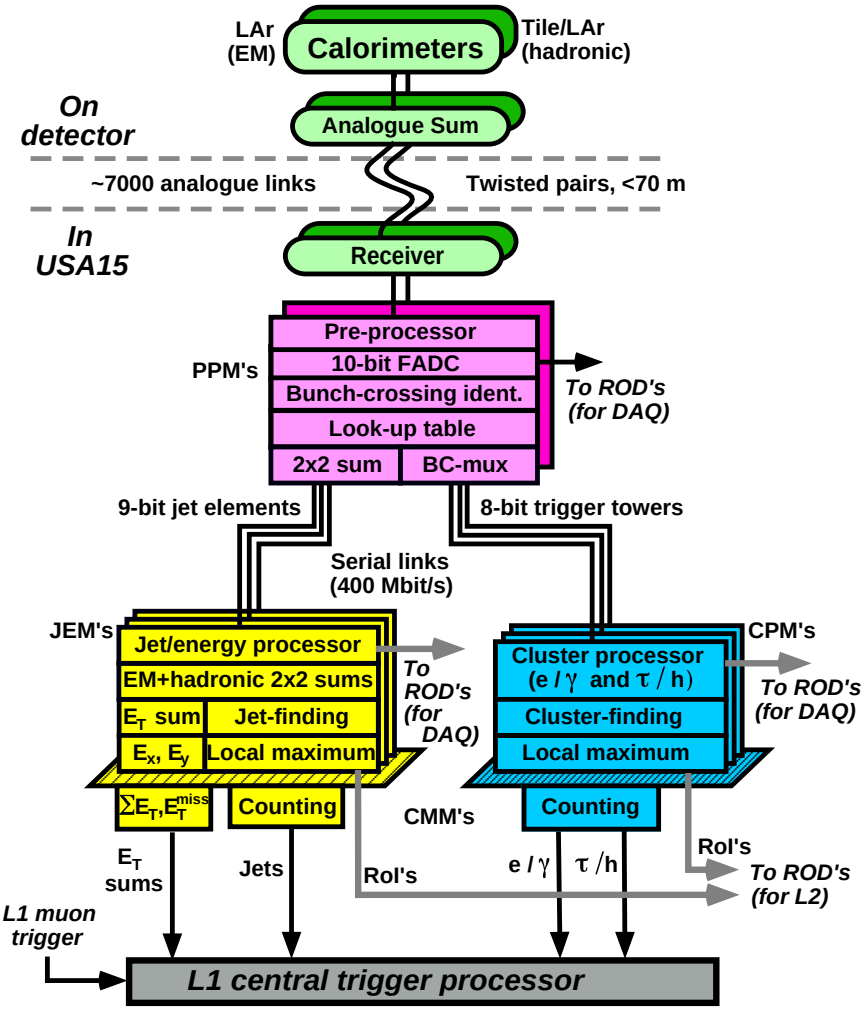

Figure 8.3: Architecture of the L1 calorimeter trigger. Analogue data from the calorimeters are digitised and associated with the correct bunch-crossing in the pre-processor and then sent to two algorithmic processors, the jet/energy-sum processor and the cluster processor. The resulting hit counts and energy sums are sent to the central trigger processor.

feature multiplicities, as well as transverse-energy threshold information, to the CTP for every bunch-crossing.

When there is a L1 Accept (L1A) decision from the CTP, the stored data from the L1Calo subsystems are read out to the data acquisition system: this includes input data, intermediate calculations and trigger results in order to allow full monitoring and verification of the L1 trigger functionality. These data can also provide useful diagnostics for the LHC machine (see section 9.10) and the ATLAS sub-detectors. The types and positions of jet, τ-lepton and electromagnetic cluster candidates are also collected and sent to the RoI builder (see section 8.3.6) for use by the L2 trigger.

The L1Calo architecture is relatively compact, with a minimal number of crates and cable links. This helps in reducing the latency. Some of the hardware modules were designed to fulfil several different roles in the system, in order to reduce hardware costs and design efforts, as well as to reduce the number of spares required.

2008 JINST 3 S08003

8.2.1.1 The analogue front-end

Analogue signals from trigger towers in all the calorimeters are sent to the USA15 cavern on 16-way twisted-pair cables. These cables are specially routed to minimise their length, and hence the trigger latency; they range from about 30 m to 70 m in length. Liquid-argon electromagnetic calorimeter signals are converted from energy to transverse energy in the tower builder boards located on the detector, but all hadronic calorimeter signals are transmitted proportional to energy. All the trigger-tower signals arrive at 64-channel receiver modules. The main function of the receiver modules is to adjust the gains, in order to provide transverse energy rather than energy for hadronic calorimeter signals, and to compensate for differences in energy calibration and signal attenuation in the long cables. The receivers reshape the signals, and contain linear, variable-gain amplifiers controlled by DAC's. Receiver outputs are sent as differential signals on short twisted-pair cables to the pre-processor. A further function of the receivers is to monitor a small, programmable selection of analogue input signals.

8.2.1.2 The pre-processor

The pre-processor consists of eight 9U VMEbus crates. Four crates process electromagnetic trigger towers and four process hadronic towers. Each crate contains 16 Pre-Processor Modules (PPM's) which each receive four analogue cables on the front panel and process 64 analogue input signals. The granularity of these signals is reduced compared to the full calorimeter data. This is done by analogue summing at the detector of variable numbers of calorimeter cells, ranging from a few up to 60. The main signal processing is performed by 16 multi-chip modules, each of which processes four trigger towers. Ten-bit Flash ADC's (FADC's) digitise the signals with a sampling frequency of 40.08 MHz. Fine adjustment of the timing of each digitisation strobe is performed by a four-channel ASIC, which provides programmable delays in steps of 1 ns across the 25 ns LHC clock period. The digitised values are then sent to a custom pre-processor ASIC.

The pre-processor ASIC synchronises the timing of the four inputs, to compensate for different times-of-flight and signal path-lengths. It then assigns signals to the correct bunch-crossing, as detailed below. A look-up table is used to carry out pedestal subtraction, apply a noise threshold, and do a final transverse-energy calibration, resulting in 8-bit trigger-tower energies. Finally, it performs bunch-crossing multiplexing (see below) for the CP and it sums the four values into 0.2×0.2 jet elements (2×2 sum) for the JEP. Two 10-bit low-voltage differential signalling (LVDS) serialisers operating at 400 Mbit/s transmit the processed trigger-tower data to the CP, while a third serialiser sends the summed 9-bit jet elements to the JEP.

The pre-processor ASIC reads out data to the data acquisition system upon receiving a L1A signal. The readout data are taken from pipeline memories at two stages: the raw digitised values from the FADC's, and the 8-bit processed trigger-tower data from the look-up tables. Data from the bunch-crossing of interest, as well as a programmable number of bunch-crossings around it (typically up to five in all), allow monitoring of pulse shapes at the FADC's, and of the bunch-crossing identification and energy calibration at the look-up table outputs. These readout data are serialised and sent to the data acquisition readout over an optical fibre. In addition, various rates based on the input signals are monitored and histogrammed automatically in the ASIC and are read out by VMEbus.

2008 JINST 3 S08003

Bunch-crossing identification. The analogue pulses from the electromagnetic and hadronic calorimeters have widths of several bunch-crossings, so it is essential that trigger-tower signals are associated with the correct bunch-crossing. This is a crucial requirement not only for normal-size pulses, but also for saturated pulses (above about 250 GeV) and for pulses as small as possible (down to 2–3 GeV, just above the noise level). The pre-processor ASIC is capable of identifying a signal's bunch-crossing using three different methods, which provides ample redundancy for consistency checks during commissioning.

For normal, unsaturated signals, a digital pipelined finite-impulse-response filter processes five consecutive FADC samples. A subsequent peak-finder attributes the maximum value of this sum to the corresponding bunch-crossing. The working range of the method spans from small trigger signals (energy depositions of a few GeV) up to the near-saturation level of around 250 GeV.

For saturated signals, two consecutive samples are compared to a low and a high threshold, making use of the finite peaking-time (approximately 50 ns) of an analogue input signal. Thus, detection of a leading edge allows attribution of the virtual peak to a specific bunch-crossing. This method is valid from around 200 GeV up to the maximum energy range of the calorimeters.

A third method uses comparators with programmable thresholds on the analogue input signals to present a rising-edge signal to the multi-chip modules. Given the known peaking time, bunch-crossing identification can be performed using an appropriate programmed delay in the pre-processor ASIC. The validity of this method begins well above the comparator threshold and extends up to the full energy range. There is thus a large overlap with the two previous methods, allowing consistency checks between the methods to be performed.

The finite-impulse-response filter output is presented to the look-up table to extract a calibrated E_T value for the trigger tower. If the bunch-crossing identification criteria are met, this value is sent to the CP outputs. In the case of saturation, the tower is assigned the maximum 8-bit value of 255 GeV. For the JEP outputs, any 0.2×0.2-sum jet element which contains a saturated trigger tower, or which has a 9-bit sum in overflow, is assigned the maximum 9-bit value of 512 GeV. A tower or jet element with a maximum value is understood to be saturated by the CP and/or JEP sub-systems. The trigger menu will be set up so that any event where a saturation condition occurs will produce a L1A signal, and the RoI's sent to the L2 trigger will be flagged by saturation bits.

Bunch-crossing multiplexing. The data from the pre-processor modules consist of four 8-bit trigger towers per multi-chip module to the CP, and one 9-bit 0.2×0.2-sum jet element to the JEP. To economise on the number of links needed, it was noted that the bunch-crossing identification algorithm is essentially a peak-finding scheme. This means that an occupied bunch-crossing will always be followed by an empty (zero) one. This allows two trigger towers being sent to the CP to share a single serial link. Trigger towers are paired at the pre-processor ASIC output stage, and the scheme is called Bunch-Crossing Multiplexing (BC-mux). By using it, data transmission to the CP sub-system is achieved with only two links per multi-chip module instead of four. However, for the JEP, where a sum of four towers is transmitted, this cannot be done.

Output signal fan-out and pre-compensation. The high-speed serial outputs to the CP and JEP are fanned out in order to provide the trigger algorithms with overlapping data between detector quadrants in azimuth. The data pass through the back-plane to 11 m long shielded parallel-pair

2008 JINST 3 S08003

cables. RC pre-compensation is done to improve signal-driving capabilities, since observed signal attenuation and distortion from the cables may compromise data integrity. Bit-error rates of less than 10^{-14} have been achieved.

8.2.1.3 The cluster and jet/energy-sum processors

The CP and JEP sub-systems share many architectural features and some common hardware. The jet algorithm in the JEP and the electron/photon and τ cluster algorithms in the CP both perform feature searches in overlapping, sliding windows. Therefore, a large amount of data duplication between processor modules is required, and this is done as follows. Both sub-systems divide the calorimeters into four azimuthal quadrants, with each processor module within a quadrant covering a slice in pseudorapidity and 90° in azimuth. Overlapping data from neighbouring azimuthal quadrants are provided by duplicated serial links from the pre-processor. Within each quadrant, modules only need to share input data with their nearest neighbours, over short (roughly 2 cm) point-to-point back-plane links. This architecture minimises the number of cable links from the pre-processor, and the back-plane fan-out is simplified.

The CP is a four-crate system, with 14 Cluster Processor Modules (CPM's) in each crate covering one calorimeter quadrant. The JEP is contained in two crates, each containing eight Jet/Energy Modules (JEM's) from two opposing quadrants in azimuth (16 JEM's total). Results from the processor modules are brought to two Common Merger Modules (CMM's) in each crate: these sum the data to produce crate-level results. The CMM's also perform the system-level summation of data from the different crates, and transmit the final results to the CTP.

The electron/photon and τ triggers extend out to $|\eta| = 2.5$, which is the fiducial limit for precision measurements with the inner detector and electromagnetic calorimetry. The jet trigger extends out to $|\eta| = 3.2$. The E_T^{miss} and total transverse-energy triggers include the forward calorimetry, in particular to provide adequate E_T^{miss} performance, which means that they extend to $|\eta| = 4.9$. This also allows the FCal to be used for forward-jet triggers.

8.2.1.4 The cluster processor module

The electron/photon trigger algorithm [228], shown in figure 8.4, identifies 2×2 clusters of trigger towers in which at least one of the four possible two-tower sums (1×2 or 2×1) of nearest-neighbour electromagnetic towers exceeds a pre-defined threshold. Isolation-veto thresholds are set for the 12-tower surrounding ring in the electromagnetic calorimeter, as well as for the 2×2 hadronic-tower core sum behind the cluster and the 12-tower hadronic ring around it. All these thresholds are programmable.

The τ algorithm uses the same basic elements to select narrow hadronic jets. Each of the four possible two-tower sums of nearest-neighbour electromagnetic towers is added to the 2×2 hadronic-tower core sum directly behind, and the result is compared to a pre-defined threshold. Isolation veto thresholds are set separately for each of the surrounding 12-tower rings in both the electromagnetic and hadronic calorimeters.

The isolation thresholds for both algorithms are absolute values, rather than ratios of isolation energy to cluster energy. This simpler approach was chosen on the basis of studies, which showed

2008 JINST 3 S08003

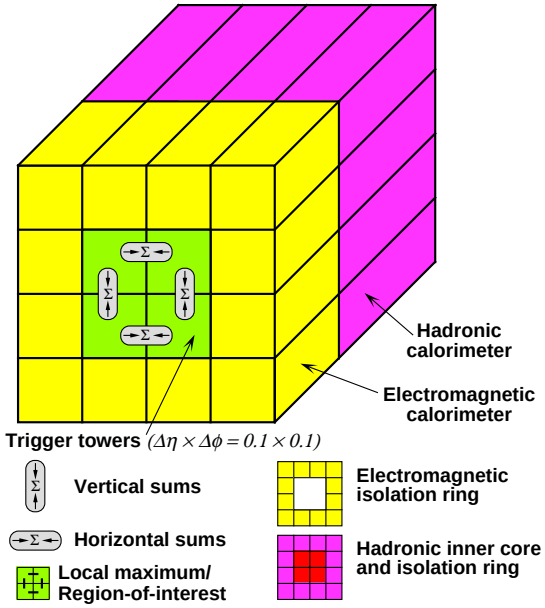

Figure 8.4: Electron/photon and τ trigger algorithms, as described in the text.

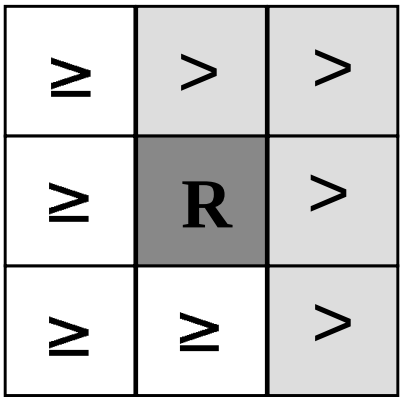

Figure 8.5: E_T local-maximum test for a cluster/RoI candidate. The η-axis runs from left to right, and the ϕ-axis from bottom to top. The symbol R refers to the candidate 2×2 region being tested.

that the expected isolation sums are relatively insensitive to shower energies. In practice, high-energy clusters will generally have looser isolation criteria to maximise the efficiency for possible low-rate exotic signal processes, while lower-energy clusters will have stricter isolation criteria in order to minimise the rates at the expense of a limited loss of signal.

These algorithms are run over all possible 4×4 windows, which means that the windows overlap and slide by steps of 0.1 in both η and ϕ. This implies that an electron/photon or τ cluster can satisfy the algorithm in two or more neighbouring windows. Multiple-counting of clusters is avoided by requiring the sum of the four central electromagnetic plus the sum of the four central hadronic towers to be a local maximum with respect to its eight nearest overlapping neighbours. In order to avoid problems in comparing digital sums with identical values, four of the eight comparisons are 'greater than' while the other four are 'greater than or equal to', as shown in figure 8.5. The location of this 2×2 local maximum also defines the coordinates of the electron/photon or τ RoI.

The CPM identifies and counts clusters satisfying sets of threshold and isolation criteria. Eight threshold sets are reserved for electron/photon triggers, while eight further threshold sets can each be used for either electron/photon or τ triggers.

Each CPM receives and deserialises input data on 80 LVDS cables from the pre-processor modules, brought in to the rear of the module through back-plane connectors. The data are then shared between neighbouring modules via the back-plane, and finally fanned out to eight CP FPGA's, which perform the clustering algorithms. The serialiser FPGA's also store the input data in pipelines for eventual readout to the data acquisition system upon reception of a L1A signal.

2008 JINST 3 S08003

The eight CP FPGA's each service eight overlapping 4×4 windows. Pipelines implemented in each one of them save output data for readout to the data acquisition system, and also save cluster types and coordinates for readout as RoI's to the L2 trigger. Two hit-multiplicity FPGA's collect and then sum the 3-bit cluster multiplicities from the CP FPGA's, for reporting to the crate-level merging of CP results. These multiplicities are transmitted via the back-plane. If more than seven instances of a cluster type are identified (a very rare occurrence, given that the mean occupancy is less than one), the multiplicity is reported as seven. Two additional FPGA's collect input data from the serialiser FPGA's, RoI data from the CP FPGA's, and output data from the hit-multiplicity FPGA's upon reception of a L1A signal, and transmit them to readout driver modules serving the data acquisition system and the L2 trigger on two optical fibres from the front panel of the module.

8.2.1.5 The jet/energy module

The Jet/Energy Module (JEM) works with jet elements which are the sums of 2×2 trigger towers in the electromagnetic calorimeters added to 2×2 trigger towers in the hadronic calorimeters. The jet algorithm identifies E_T sums within overlapping windows consisting of 2×2, 3×3, or 4×4 jet elements, corresponding to window sizes of 0.4, 0.6, or 0.8 in η and ϕ, as shown in figure 8.6. These sums are then compared to pre-defined jet energy thresholds. Multiple-counting of jet candidates is avoided by requiring the window to surround 2×2 jet elements whose sum is a local maximum, with the same definition as for electron/photon and τ clusters. The location of this 2×2 local maximum also defines the coordinates of the jet RoI. Eight independent combinations of jet E_T threshold and window size are available for trigger menus.

The energy-summation algorithm produces sums of E_T, E_x and E_y, and uses the system-level sums of these to report on four total-E_T and eight E_T^{miss} thresholds to the CTP.

Each JEM receives and deserialises data from 88 LVDS links, corresponding to 44 jet elements for both the electromagnetic and hadronic calorimeters. Four input FPGA's receive the data, sum the electromagnetic and hadronic parts of each jet element to 10-bit values, and send these sums to the main processor FPGA's on the same and neighbouring modules. Pipelines in each input FPGA save input data for readout to the data acquisition system upon reception of a L1A signal.

The jet and energy-summation algorithms are implemented in two large main-processor FPGA's per JEM. The main processors are also responsible for reporting results to the crate-level merging, as well as pipelining of data acquisition and RoI information for readout. The jet output of each JEM is a data stream consisting of eight 3-bit jet multiplicities. The energy output is also a data stream containing the values of E_x and E_y, each compressed from 12 bits to an 8-bit (6-bit mantissa plus two multiplier bits) quad-linear scale (a data-compression technique that multiplies the mantissa by 1, 4, 16, or 64).

A single readout-controller FPGA collects input data from the input FPGA's, and output and RoI data from the main processor FPGA's, for readout to readout driver modules serving data acquisition and the L2 trigger on two optical fibres from the front panel of the module.

2008 JINST 3 S08003

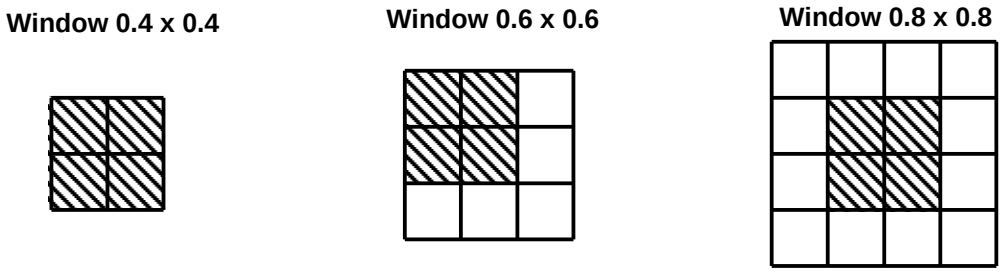

Figure 8.6: Jet trigger algorithms, based on 0.2×0.2 jet elements and showing RoI's (shaded). In the 0.6×0.6 case there are four possible windows containing a given RoI. In the 0.8×0.8 case the RoI is required to be in the centre position, in order to avoid the possibility of two jets per window.

8.2.1.6 The common merger module

Two modules in each CP and JEP crate carry out crate-level merging of results received from the crate's processor modules. In the CP crates, each merger module is responsible for calculating 3-bit cluster multiplicities for eight of the 16 electron/photon and τ cluster definitions. In the JEP crates, one merger module produces 3-bit multiplicities for the eight jet definitions, while the other produces sums of E_T, E_x and E_y. Each Common Merger Module (CMM) receives data from the crate's 14 CPM's or 16 JEM's, over point-to-point links on the crate back-plane.

The CMM carries out all of these merging functions by using different firmware versions. Each CMM receives up to 400 bits of data per bunch-crossing from the crate's CPM's or JEM's. A large FPGA performs crate-level merging. Parallel LVDS cable links between the sub-system crates bring all crate-level results to one CMM of each type, which is designated as the system-merger CMM. A second FPGA on the CMM carries out the system-level merging.

At the system level, the CMM carries out the logic to provide global trigger results. Three-bit overall multiplicities for each of the electron/photon, τ, and jet thresholds are formed and sent to the CTP. The overall sums of E_x and E_y are applied together as the address to a look-up table. In one operation this works out whether the resulting vector sum, i.e. missing E_T, is above or below eight programmable missing-E_T thresholds and codes the result in an eight-bit word. For total scalar E_T, the global sum is compared to four threshold values. Finally, a rough approximation of the total E_T in jets, based on the numbers of jets passing each of the eight jet thresholds, is compared to four threshold values.

All of these calorimeter trigger results are passed to the CTP by cable. As with other processor modules, input and output data on each CMM are stored in FIFO's and read out to the data acquisition system over an optical fibre. RoI data on the missing and total E_T values are sent to L2.

8.2.1.7 The processor back-plane

The CP and JEP use a common, custom processor back-plane. It has 16 CPM/JEM positions flanked by two CMM positions. At the left it has a slot for a commercial VMEbus CPU. At the right is a slot for a timing control module, which interfaces to the TTC (e.g. to distribute clock signals)

2008 JINST 3 S08003

and to the detector control system, which uses CANbus to monitor voltages and temperatures on all trigger modules.

The processor back-plane is a monolithic printed-circuit board of 9U height. It is populated almost entirely with 2 mm hard-metric connectors, with 1148 signal and ground pins in each JEM/CPM and CMM position. There are point-to-point links between neighbouring processor modules for input data fan-in/fan-out. Connections are provided from each CPM or JEM to the two CMM's at the right and left of the processor modules. To conserve pins, a non-standard VMEbus with the minimum possible number of pins (43 signals with 16 data bits and 24 address bits) is used.

The LVDS serial-input and merger-interconnect cables are connected to the rear of the processor back-plane and passed through it to the modules in front. This results in a system with fewer cables on the front panels of the modules and, as a consequence, hopefully fewer recabling errors and less cable damage over the lifetime of the experiment.

8.2.1.8 The readout driver

The trigger system has two separate readout systems. Input, output, and some intermediate data from each module are read out to the data acquisition system and at the same time the CP and JEP sub-systems report feature types and coordinates as RoI data to the L2 trigger.

The readout system has been designed to handle one bunch-crossing of RoI data and up to five bunch-crossings of data acquisition data per event at a L1A rate of up to 75 kHz. A common approach has been adopted in all L1Calo sub-systems for data acquisition and RoI readout.

On each module to be read out, readout FIFO's on each processor FPGA or ASIC are read out as serial streams to a readout controller FPGA for timing alignment. This passes the serial streams in parallel to the inputs of a G-link (high-speed serial link) transmitter, which transmits them serially at 800 Mbit/s over optical fibres to a Readout Driver (ROD).

A common ROD module is used by both the data acquisition and RoI readout sub-systems to gather and report data from the pre-processor modules, CPM's, JEM's, and CMM's, using different firmware configurations for different readout tasks and modules. The ROD is a 9U-module residing in a standard VME64x crate. It has 18 G-link receivers, which pass their parallel outputs to the FPGA's for data compression, zero suppression, and some data monitoring. The ROD also contains four S-link transmitters on a rear-transition module for passing compressed event data to the data acquisition and RoI readout buffers. The S-link interface specification defines the signals and protocol for the readout links; it does not define the hardware implementation. Routing of data to the different outputs is carried out by a switch controller FPGA, whose settings depend on the type and source of data being read out. In addition, a further large FPGA provides monitoring capability on a sample of readout data.

8.2.2 Muon trigger

The L1 muon trigger is based on dedicated finely segmented detectors (the RPC's in the barrel and the TGC's in the end-caps, as described in detail in section 6.6) with a sufficient timing accuracy to provide unambiguous identification of the bunch-crossing containing the muon candidate.

2008 JINST 3 S08003

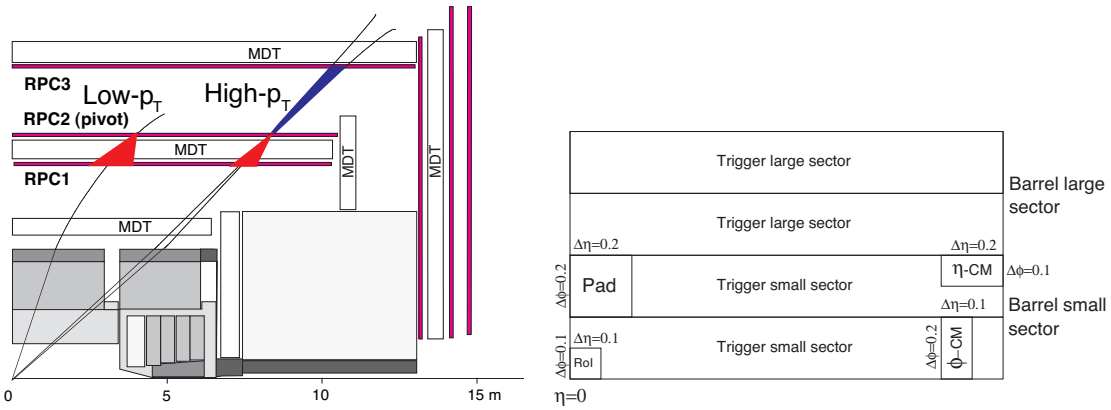

Figure 8.7: Schema (left) and segmentation (right) of the L1 muon barrel trigger. Left: The RPC's are arranged in three stations: RPC1, RPC2, and RPC3. Also shown are the low-p_T and high-p_T roads. See text for details. Right: areas covered by η and ϕ coincidence-matrix (CM) boards, by an RoI, by a Pad logic board, and by sector logic boards.

The trigger in both the barrel and the end-cap regions is based on three trigger stations each. The basic principle of the algorithm is to require a coincidence of hits in the different trigger stations within a road, which tracks the path of a muon from the interaction point through the detector. The width of the road is related to the p_T threshold to be applied. A system of programmable coincidence logic allows concurrent operation with a total of six thresholds, three associated with the low-p_T trigger (threshold range approximately 6–9 GeV) and three associated with the high-p_T trigger (threshold range approximately 9–35 GeV). The trigger signals from the barrel and the muon end-cap trigger are combined into one set of six threshold multiplicities for each bunch-crossing in the muon to CTP interface, before being passed on to the CTP itself.

8.2.2.1 Muon barrel trigger

Trigger signals. The muon trigger for the barrel regions ($|\eta| < 1.05$) makes use of dedicated RPC detectors. The RPC is a gaseous detector providing a typical space-time resolution of 1 cm $\times$ 1 ns and a rate capability of about 1 kHz/cm^2. As shown on the left side of figure 8.7, the RPC's are arranged in three stations. The two Barrel Middle (BM) stations, RPC1 and RPC2, are arranged on either side of the Monitored Drift Tube (MDT) BM stations at approximately 7.5 m radial distance from the interaction point (see chapter 6). The RPC3 Barrel Outer (BO) station, mounted on the inside (large sectors) or outside (small sectors) of the MDT BO stations, is located at a radial distance of about 10 m. Each station is made of one RPC doublet, i.e. two independent detector layers, each measuring η and ϕ. Both planes are used in the trigger. The η-strips are parallel to the MDT wires and provide the bending view of the trigger detector. The ϕ-strips are orthogonal to the MDT wires and provide the second coordinate measurement. These strips are also needed for the pattern recognition. The RPC's are organised in several modules, and their dimensions have been chosen to match those of the corresponding MDT chambers. In most stations the RPC's are composed of two units along the beam direction. To avoid dead areas between adjacent units, the active zones of neighbouring RPC's are partially overlapped in η.

2008 JINST 3 S08003

Trigger algorithm. The trigger algorithm operates in the following way: if a track hit is generated in the second RPC doublet (the pivot plane), a search for a corresponding hit is made in the first RPC doublet, within a road whose centre is defined by the line of conjunction of the hit in the pivot plane with the interaction point. The width of the road is a function of the desired cut on p_T: the smaller the road, the higher the cut on p_T. The system is designed so that three such low-p_T thresholds in each projection can be applied simultaneously. The algorithm is performed in both the η and the ϕ projections to reduce accidental triggers from low-energy particles in the cavern. A 3-out-of-4 coincidence of the four layers of the two doublets is required, which ensures excellent rejection of fake tracks from noise hits and greatly improves the stability of the trigger performance in the presence of small chamber inefficiencies.

The high-p_T algorithm makes use of the low-p_T trigger built from hits in RPC1 and RPC2, and of the information generated in the RPC3 station. The algorithm operates in a similar way to the low-p_T one. The centre of the road is determined in the same way as for the low-p_T trigger, and in addition to the low-p_T trigger pattern result, 1-out-of-2 possible hits of the RPC3 doublet is required. As with the low-p_T trigger, three p_T thresholds operate simultaneously, resulting in a total of six thresholds reported to the central trigger logic for each event. For both low and high-p_T triggers, trigger information in η and ϕ is combined to form RoI's to be sent to the L2 trigger.

System implementation. The trigger scheme for the barrel muon trigger is implemented in custom-built electronics, mounted either directly on the RPC detectors or located outside the main experimental cavern. A schema of the trigger signal and readout chain is shown in figure 8.8. Signals from the RPC detectors are processed in Amplifier-Shaper-Discriminator (ASD) boards (shown as triangles in figure 8.8) attached to the chambers at the end of the RPC strips. In the low-p_T trigger, for each of the η and the ϕ projections the RPC signals of the RPC1 and RPC2 doublets are sent to a coincidence matrix board containing a coincidence matrix chip. This chip performs most of the functions of the trigger algorithm and of the readout. At this stage the signals are aligned, the coincidence operations are performed, and the three p_T thresholds are applied. The coincidence matrix board produces an output pattern containing the low-p_T trigger results for each pair of RPC doublets in the η and ϕ projections. The information of the two adjacent coincidence matrix boards in the η projection, and similarly in the ϕ projection, are combined in the low-p_T Pad Logic board (low-p_T Pad in figure 8.8) board. The four low-p_T coincidence matrix boards and the corresponding Pad board are mounted on top of the RPC2 detector stations. The low-p_T Pad board generates the low-p_T trigger result and the associated RoI information. This information is transmitted to the corresponding high-p_T Pad board, which collects the overall results for low-p_T and high-p_T.

In the high-p_T trigger, for each of the η and ϕ projections the signals from the RPC3 doublet, and the corresponding pattern result of the low-p_T trigger, are sent, via dedicated LVDS links, to a coincidence matrix board very similar to the one used in the low-p_T trigger. This board contains the same coincidence matrix chip as the low-p_T board, programmed for the high-p_T algorithm. The high-p_T board produces an output pattern containing the high-p_T trigger results for a given RPC doublet in the η and ϕ projection. The information of two adjacent coincidence matrix boards in the η projection, and similarly in the ϕ projection, are combined in the high-p_T Pad logic board

2008 JINST 3 S08003

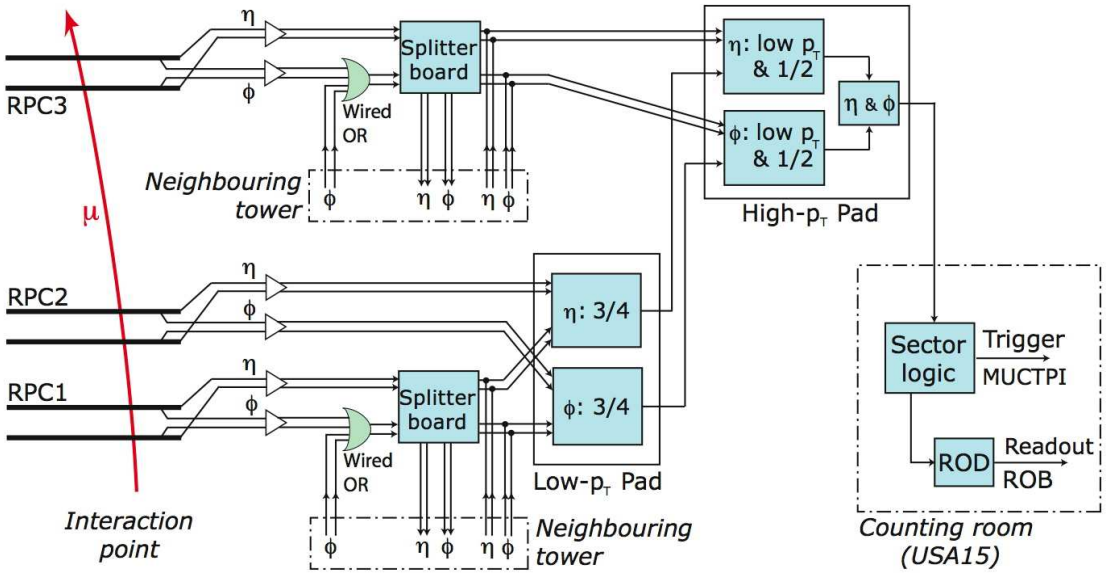

Figure 8.8: Schema of the trigger signal and readout chain of the L1 barrel muon trigger.

(high-p_T Pad in figure 8.8). The four high-p_T coincidence matrix boards and the corresponding Pad board are mounted on top of the RPC3 detector.

The high-p_T Pad board combines the low-p_T and high-p_T trigger results. The combined information for each bunch-crossing is sent via optical links to sector logic boards located in the USA15 counting room. Each sector logic board receives inputs from seven (six) low-p_T (high-p_T) Pad boards, combining and encoding the trigger results of one trigger sector. The sector logic board sends the trigger data for each bunch-crossing to the Muon to Central Trigger Processor Interface (MUCTPI, see section 8.2.2.3), located in the USA15 counting room.

For events which are selected by the L1 trigger, data are read out from both the low-p_T and the high-p_T Pad boards. These data include the RPC strip pattern and some additional information used in the L2 trigger. The readout data for events accepted by the L1 trigger are sent asynchronously to ROD's located in the USA15 counting room and from there to Readout Buffers (ROB's). The data links for the readout data are independent of the ones used to transfer partial trigger results to the sector logic boards.

System segmentation and latency. From the trigger point of view the barrel is divided into two halves, $\eta < 0$ and $\eta > 0$, and within each half-barrel 32 logically identical sectors are defined. The correspondence between these logical sectors and physical chambers is indicated in the diagram on the right of figure 8.7. The barrel large chambers and the barrel small chambers of both middle and outer RPC stations are each logically divided in two in azimuth to produce two large sectors and two small sectors per half-barrel octant. Inside a sector, the trigger is segmented in Pads and RoI's.

A large sector contains seven Pad regions, while a small sector contains six Pad regions. The region covered by a Pad is about 0.2×0.2 in $\Delta\eta \times \Delta\phi$. Inside the Pad the trigger is segmented into RoI's. A RoI is a region given by the overlap of an η coincidence-matrix and a ϕ coincidence-

2008 JINST 3 S08003

matrix. The dimensions of the RoI's are about 0.1×0.1 in $\Delta\eta \times \Delta\phi$. The total number of Pads is $7 \times 2 \times 32$ for the large sectors and $6 \times 2 \times 32$ for the small ones, giving 832 Pads altogether. Since one Pad covers fours RoI's, the total number of RoI's is 3328.

To avoid losing efficiency due to uncovered regions in the trigger system, different parts of the system overlap. However, this overlap can cause double-counting of muon candidates. In the barrel trigger system, overlap is treated and solved at three different levels. Within a Pad region the Pad logic removes double-counting of tracks between the four RoI's of the region. In addition, if it is found that a trigger was generated in a zone of overlap with another Pad region, this trigger is flagged as 'border' trigger and any overlap will be solved later on. The sector logic then prevents double-counting of triggers within a sector. Triggers generated in zones of overlap between different sectors are flagged by the sector logic and sent to the MUCTPI, which prevents double-counting between sectors.

The latency of the muon barrel trigger is about 2.1 μs, well within the allowed envelope.

8.2.2.2 Muon end-cap trigger

Trigger signals. The muon trigger for the end-cap regions is based on signals provided by TGC detectors. The time resolution is not as good as for RPC's, but good enough to provide an efficiency greater than 99% for bunch-crossing identification for the 25 ns gate of ATLAS. Crucial for the end-cap region of ATLAS is their larger rate capability of more than 20 kHz/cm^2. The TGC's are arranged in nine layers of gas volumes grouped into four planes in z (see figure 8.9 left, and also section 6.8). The TGC inner station (I) at $|z| \sim 7$ m consists of one plane of doublet units. At $|z| \sim 14$ m seven layers are arranged in one plane of triplet chambers (M1, closest to the interaction point) and two planes of doublet chambers (M2, M3). The doublet forming the plane farthest from the interaction point in each end-cap (M3) is referred to as the pivot plane, and its chamber layout and electronics are arranged such that, to a good approximation, there are no overlaps or holes in this plane. For triggering, the TGC's cover a pseudorapidity range $1.05 < |\eta| < 2.4$, except for the innermost plane which covers a range $1.05 < |\eta| < 1.92$. Each trigger plane consists of a wheel of eight octants of chambers symmetric in ϕ. Each octant is divided radially into the 'forward region' and the 'end-cap region'. Anode wires of TGC's are arranged in the azimuthal direction and provide signals for R information, while readout strips orthogonal to these wires provide signals for ϕ information. Both wire and strip signals are used for the muon trigger. Signals from two wire-planes and two strip-planes are read out from the doublet chambers, and signals of three wire-planes but only two strip-planes are read out from the triplet chambers. Anode wires are grouped and fed to a common readout channel for input to the trigger electronics, resulting in wire-group widths in the range between 10.8 mm and 55.8 mm. Wire groups are staggered by half a wire group between the two planes of a doublet station, and by one third of a wire group between each of the planes of a triplet station. Each chamber has 32 radial strips, and thus the width of a strip is 4 mrad (8 mrad for the forward region). Strips are also staggered by half a strip-width between the two strip-planes in a triplet or a doublet chamber.

Trigger algorithm. The scheme of the L1 muon end-cap trigger is shown on the left hand side of figure 8.9. The trigger algorithm extrapolates pivot-plane hits to the interaction point, to construct

2008 JINST 3 S08003

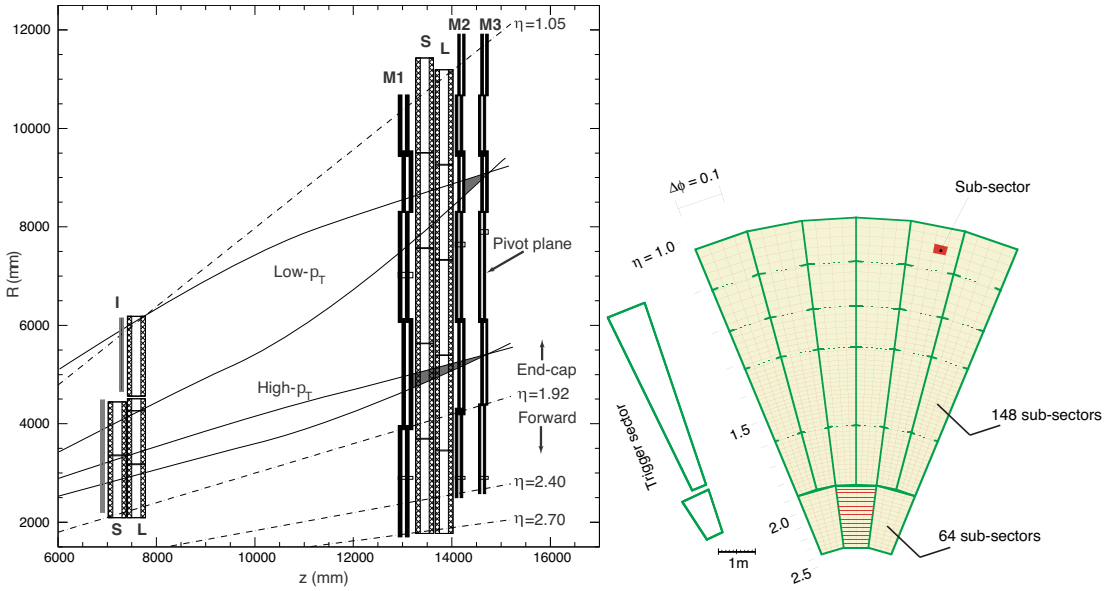

Figure 8.9: Schema (left) and segmentation (right) of the L1 muon end-cap trigger. See text for details.

roads following the apparent infinite-momentum path of the track. Deviations from this path of hits in the trigger planes closer to the interaction point are related to the momentum of the track. Coincidence signals are generated independently for R and ϕ. A 3-out-of-4 coincidence is required for the doublet pair planes of M2 and M3, for both wires and strips, a 2-out-of-3 coincidence for the triplet wire planes, and 1-out-of-2 possible hits for the triplet strip planes. The final trigger decision in the muon end-cap system is done by merging the results of the $R-\phi$ coincidence and the information from the EI/FI chambers in the inner station (see section 6.8.1). As the $\eta-\phi$ coverage of the EI/FI chambers is limited, the coincidence requirements depend on the trigger region, in order to keep a uniform efficiency in the end-cap region. Six sets of windows are constructed around the infinite-momentum path, corresponding to three different high-p_T and three different low-p_T thresholds. Trigger signals from both doublets and the triplet are involved in identifying the high-p_T candidates, while in case of the low-p_T candidates the triplet station may be omitted to retain high efficiency, given the geometry and magnetic field configuration of a specific region.

System implementation. The trigger scheme outlined above is implemented in purpose-built electronics, partly mounted on and near the TGC chambers, and partly located in the USA15 counting room. A schema of the trigger signal and readout chain is shown in figure 8.10. The wire and strip signals emerging from the TGC's are fed into ASD boards physically attached to the edge of a TGC and enclosed inside the TGC electrical shielding. Each ASD board handles 16 channels. From the ASD boards signals are routed to the so-called PS-boards (patch panel and slave), which integrate several functions in one unit. Each PS-board receives signals from up to 20 ASD's. First the signals are routed to a patch-panel section, which also receives timing signals from the TTC system. Signal alignment and bunch-crossing identification (BCID) is performed

2008 JINST 3 S08003

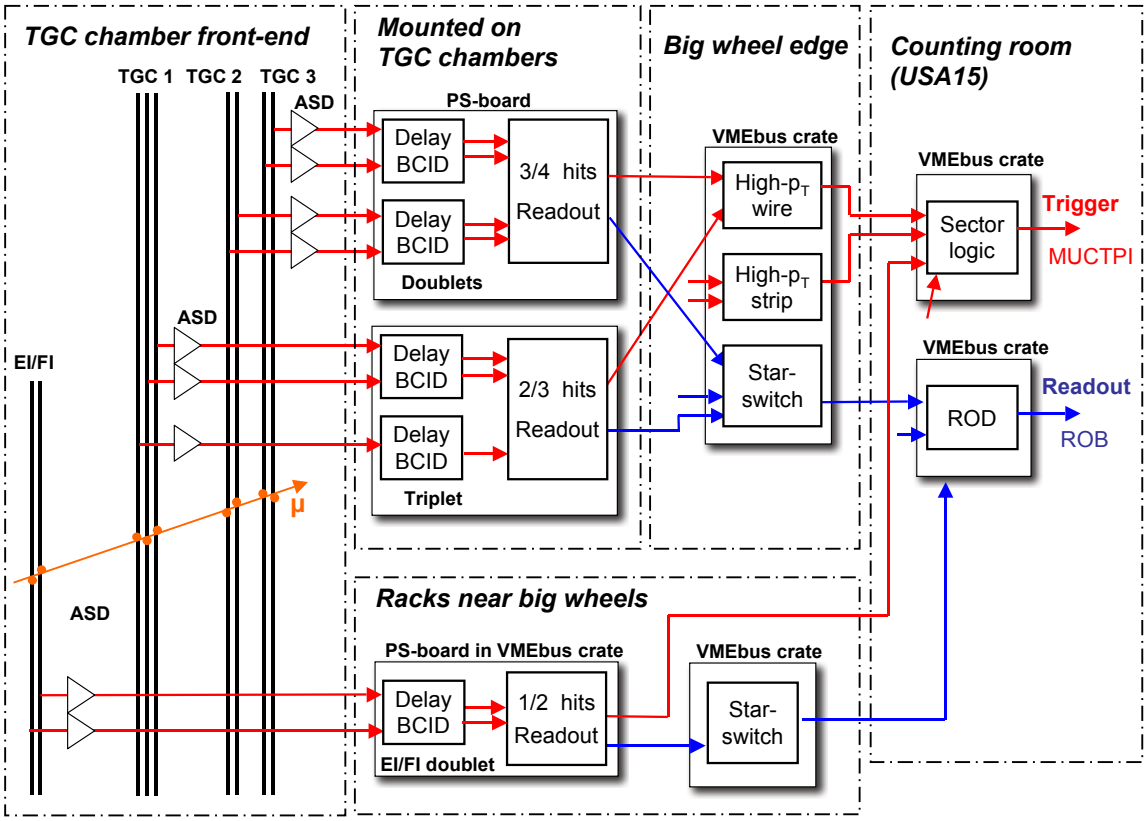

Figure 8.10: Schema of the trigger signal and readout chain of the L1 muon end-cap trigger. See text for details.

at this stage, and physical overlaps of TGC chambers are handled. In addition, detector control system and other control and monitoring signals are routed to the other parts of the electronics mounted on the chambers. The aligned signals are passed to the so-called slave section, where the coincidence conditions are applied and readout functions are performed. The PS-boards are placed on the accessible outer surfaces of the TGC wheels: the electronics for the two doublets are mounted on the outside of the outer doublet wheel M3 and those for the triplets on the inner surface of the triplet wheel M1. The EI/FI PS-boards are installed in racks located near the EI/FI chambers. Signals from the doublet and triplet slave boards are combined to identify high-p_T track candidates in coincidence boards combining all three trigger planes (M1, M2, M3), so-called high-p_T boards, located in dedicated mini-racks around the outer rim of the triplet wheel. Wire (R-coordinate) and strip (ϕ-coordinate) information is still treated separately at this point. Signals from high-p_T boards are sent to sector logic boards containing an $R-\phi$ coincidence unit and a track selector to select the highest-p_T coincidences. The sector logic also receives directly the signals from the EI/FI slave boards and can incorporate them into the trigger logic. The sector logic boards are located in the USA15 counting room. The resulting trigger information for 72 separate trigger sectors per side is sent to the MUCTPI.

2008 JINST 3 S08003

Full-information data sets are read out through the data acquisition system in parallel with the primary trigger logic. For readout purposes the slave boards of one or more trigger sectors are grouped into local data acquisition blocks. Each slave board is connected to a so-called star switch, which manages the data collection for a local data acquisition block. From the star switch, the data are passed on to the ROD's located in the USA15 counting room, and from there to ROB's.

System segmentation and latency. The trigger-sector segmentation of one pivot-plane octant is shown in figure 8.9 (right). The pivot plane is divided into two regions, end-cap ($|\eta| < 1.92$) and forward ($|\eta| > 1.92$). Each octant of the end-cap region is divided into six trigger sectors in ϕ, where a trigger sector is a logical unit which is treated independently in the trigger. Trigger sectors are constructed to be projective with respect to the interaction point, and therefore may cross chamber boundaries (see figure 8.9, left). Each octant of the forward region is divided into three trigger sectors. There are hence 48 end-cap trigger sectors and 24 forward trigger sectors per end-cap of TGC detectors. Each trigger sector consists of independent sub-sectors corresponding to eight channels of wire groups and eight channels of readout strips, 148 for each end-cap trigger sector and 64 for each forward trigger sector. The trigger sub-sectors correspond to the RoI's sent to the L2 trigger for events accepted by the L1 trigger.

The latency of the muon end-cap trigger is about 2.1 μs, well within the allowed envelope.

8.2.2.3 Muon to central trigger processor interface

Functional overview. The results from the muon barrel and end-cap trigger processors which form the input to the Muon to Central Trigger Processor Interface (MUCTPI) provide information on up to two muon-track candidates per muon trigger sector. The information includes the position and p_T threshold passed by the track candidates. The MUCTPI combines the information from all the sectors and calculates total multiplicity values for each of the six p_T thresholds. These multiplicity values are sent to the CTP for each bunch-crossing. For each sector either all muon candidates may be taken into account, or only the candidate with the highest p_T per sector. In forming the multiplicity sums, care has to be taken to avoid double-counting of muon candidates in regions where trigger chambers overlap. As described above, many cases of overlaps are resolved within the barrel and end-cap muon trigger processors. The remaining overlaps to be treated by the MUCTPI are those in ϕ direction between neighbouring barrel trigger sectors, and between barrel and end-cap trigger sectors. The maximum overall multiplicity is seven candidates. Larger multiplicities will appear as a multiplicity of seven.

Additional functions of the MUCTPI are to provide data to the L2 trigger and to the data acquisition system for events selected at L1. The L2 trigger is sent a subset of all muon candidate information which form the RoI's for the L2 processing. The muon RoI's sent to L2 are ordered according to decreasing p_T. The data acquisition receives a more complete set of information, including in addition the computed multiplicity values.

System implementation and latency. The MUCTPI is divided into a number of building blocks which are housed in one 9U VMEbus crate. The different functionalities of the MUCTPI are

2008 JINST 3 S08003

implemented in three types of VMEbus modules which are connected to each other via an active back-plane, and controlled by a commercial CPU unit acting as VMEbus master.

A total of 16 octant input boards each receive data corresponding to an octant in the azimuthal direction and half the detector in the η direction. They form muon-candidate multiplicities for this region, correctly taking into account the overlap zones between barrel sectors, and between barrel and end-cap sectors. There is no overlap between muon trigger sectors associated with different octant boards. The interface board to the CTP collects the multiplicity sums for the six p_T thresholds over the custom back-plane described below. The sums are transmitted to the CTP for each bunch-crossing. The interface board is also responsible for distributing time-critical control signals to the rest of the MUCTPI system. The readout driver of the system sends candidate information to the data acquisition and the L2 trigger for each accepted event. All modules are connected via a custom-built back-plane. It contains two components: an active part forms the total candidate multiplicities by adding the multiplicities of the input boards and a passive part contains a bus system to transfer data to the readout driver of the system on receipt of a L1A.

The latency of MUCTPI is included in the latency numbers for the barrel and end-cap muon trigger systems quoted above, to which it contributes 0.2 μs.

8.2.3 Central trigger processor

8.2.3.1 Functional overview

The Central Trigger Processor (CTP) [229] receives trigger information from the calorimeter and muon trigger processors, which consists of multiplicities for electrons/photons, τ-leptons, jets, and muons, and of flags indicating which thresholds were passed for total and missing transverse energy, and for total jet transverse energy. Additional inputs are provided for special triggers such as a filled-bunch trigger based on beam-pickup monitors, and a minimum-bias trigger based on scintillation counters. Up to 372 signals can be connected to the input boards of the CTP; however, only up to 160 can be transmitted internally. The selection of the signals used from all signals available at the input boards is programmable. The currently foreseen input signals listed in table 8.1 sum up to 150 bits and will therefore all be available in parallel.

In the next step the CTP uses look-up tables to form trigger conditions from the input signals. Such a condition could be, for example, that the multiplicity of a particular muon threshold has exceeded one, i.e. at least two muons in this event have passed this threshold. For such an event, this trigger condition would be set to *true*. Further trigger conditions are derived from internally generated trigger signals: two random triggers, two pre-scaled clocks, and eight triggers for programmable groups of bunch-crossings. The maximum number of trigger conditions at any one time is 256.

The trigger conditions are combined to form up to 256 trigger items, where every trigger condition may contribute to every trigger item. An example for a trigger item would be that the following conditions have been fulfilled: at least two muons have passed a particular threshold, and at least one jet has passed a particular threshold. Furthermore each trigger item has a mask, a priority (for the dead-time generated by the CTP), and a pre-scaling factor (between 1 and 2^{24}). The L1A signal generated by the CTP is the logical OR of all trigger items.

2008 JINST 3 S08003

Table 8.1: Trigger inputs to the CTP of the L1 trigger. The number of bits implies the maximum multiplicity which can be encoded, i.e. up to seven for three bits. Multiplicities larger than this value will be set to the possible maximum, in this case seven.

Cable origin	**Number of bits**	**Trigger information**
Muon processor	6 thresholds $\times$ 3 bits	muon multiplicities
Cluster processor 1	8 thresholds $\times$ 3 bits	electron/photon multiplicities
Cluster processor 2	8 thresholds $\times$ 3 bits	electron/photon or τ multiplicities
Jet/energy processor 1	8 thresholds $\times$ 3 bits	jet multiplicities
	4 bits	total jet transverse energy
Jet/energy processor 2	2$\times$4 thresholds $\times$ 2 bits	forward-jet multiplicities for each side,
Jet/energy processor 3	4 $\times$ 1 bit	total transverse-energy sum,
	8 $\times$ 1 bit	missing transverse-energy sum
CTP calibration	28 bits	up to 28 input bits for additional trigger inputs from beam pick-ups (see section 9.10), beam condition monitors (see section 3.4.1), luminosity detectors (see sections 7.1 and 7.2), zero-degree calorimeters (see section 7.3), and others.

The CTP provides an eight-bit trigger-type word with each L1A signal. This indicates the type of trigger, and can be used to select options in the event data processing in the front-end electronics and readout chain. The CTP sends, upon reception of each L1A signal, information about the trigger decision for all trigger items to the L2 trigger (RoI builder) and the data acquisition (ROS). Part of the readout data of the CTP is the number of the current luminosity block. A luminosity block is the shortest time interval for which the integrated luminosity, corrected for dead-time and pre-scale effects, can be determined. In case of detector failures, data can be rejected from the boundary of the last luminosity block known to be unaffected, and the interval should therefore be as small as possible to avoid unnecessary data loss. On the other hand, each luminosity block should contain enough data such that the uncertainty of the luminosity determination is limited by systematic effects, not by the available statistics in the interval. For ATLAS this interval will be on the order of minutes. A luminosity block transition is initiated by the CTP, which will momentarily pause the generation of triggers, increment the luminosity block number in a register located on the CTP decision module, and release the trigger again. From this location the number is included in the readout data for each event. At each transition a set of scalers is read out from the CTP and stored, marked with the luminosity block number, in a database. These scalers keep track of the number of triggers generated by the trigger logic, the number of triggers surviving the pre-scale veto, and the number of triggers surviving the dead-time veto. The values of these counters are needed to later derive the corresponding corrections of the luminosity value associated with each luminosity block. For monitoring purposes the CTP provides bunch-by-bunch scalers of inputs and, integrated over all bunches, scalers of trigger inputs and trigger items before and after pre-scaling.

2008 JINST 3 S08003

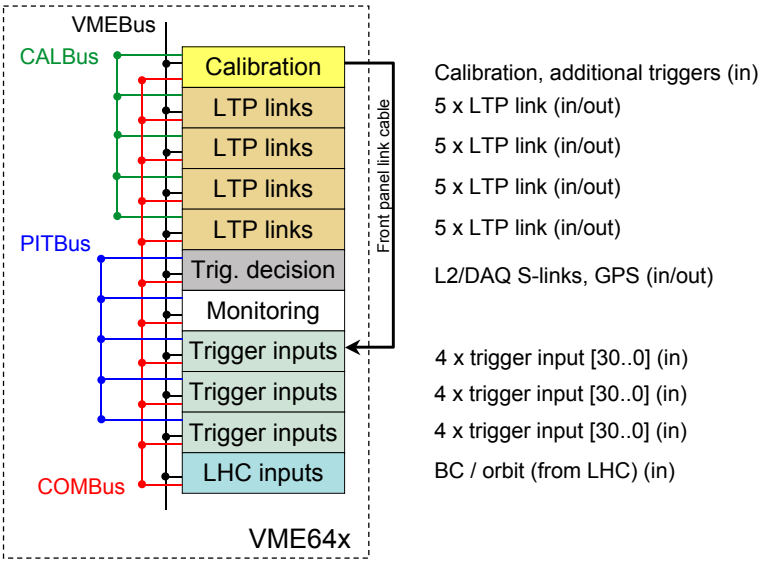

Figure 8.11: Layout of the VMEbus crate for the central trigger processor of the L1 trigger. The calibration module has the further function of receiving additional trigger signals, which are transmitted to one of the input module connectors via a front panel cable. The Local Trigger Processor (LTP) links are the connection to the individual sub-detector systems.

In addition to its function in the selection chain, the CTP is also the timing master of the detector. The clock signal synchronised to the LHC beams arrives at the CTP, and is distributed from here together with the L1A and other timing signals to all other sub-systems.

8.2.3.2 System implementation and latency

The CTP consists of six types of modules which are housed in a single 9U VMEbus crate, as shown in figure 8.11. Internal communication between the controller CPU and the modules, and between the modules proceeds by bus systems implemented on the back-planes of the crate. In addition to VMEbus, the CTP modules use custom busses for the synchronised and aligned trigger inputs (PITbus, where PIT = pattern in time), for the common timing and trigger signals (COMbus), and for the sub-detector calibration requests (CALbus). These extra busses are implemented on a custom-built backplane installed in the CTP crate.

Six different module types are employed in the CTP system. The timing signals from the LHC are received by the machine interface module (designated LHC inputs in figure 8.11), which can also generate these signals internally for stand-alone running. This board also controls and monitors the internal and external busy signals, as for example the busy signal transmitted from a sub-detector in case of overload on its data acquisition system. The module sends the timing signals to the COMbus, thereby making them available to all of the other modules in the CTP.

The trigger input modules receive trigger inputs from the muon and calorimeter trigger processors and other sources. The input boards select and route the trigger inputs to the PITbus, after

2008 JINST 3 S08003

synchronising them to the clock signal and aligning them with respect to the bunch-crossing. Three boards with four connectors of 31 trigger input signals each allow for a total of 372 input signals to be connected, of which up to 160 can be made available on the PITbus at any given time.

The trigger decision module receives the trigger inputs from the PITbus. It combines them and additional internal triggers using several look-up tables to form up to 256 trigger conditions. A typical example is a 3-to-4 look-up table, with the three input bits encoding a threshold multiplicity between 0 and 7, and the four output bits enabled if the multiplicity exceeds 0, 1, 4, 6, respectively. The first bit in this list may then be used as the trigger condition that one or more objects fulfilled the corresponding trigger threshold. In a further step the trigger conditions are combined using content-addressable memories to form up to 256 trigger items. Any of the up to 256 trigger conditions may participate in any of the up to 256 trigger items. The trigger masks, pre-scales, and dead-time generation following the forming of the trigger items are implemented in this module. The trigger results are transmitted to the COMbus. The trigger decision module also acts as the readout driver of the system, sending information to the L2 trigger and the data acquisition for each accepted event.

The monitoring module receives the 160 trigger inputs from the PITbus and monitors their behaviour on a bunch-by-bunch basis. The frequency of signals on each input line can be monitored and histogrammed, and can be retrieved via VMEbus. Trigger signals encoding multiplicities are decoded before they are monitored.

The output module (labelled LTP links in figure 8.11) receives the timing and trigger signals from the COMbus and fans them out to the sub-detectors. The module receives back from the sub-systems the busy signals, which are sent to the COMbus, and 3-bit calibration trigger requests, which are routed to the CALbus.

The calibration module time-multiplexes the calibration requests on the CALbus and sends them via a front-panel cable to one of the input modules. The calibration module also has front-panel inputs for beam pick-up monitors, minimum-bias scintillators, and test triggers.

The latency of the CTP is contained in the latency numbers for the barrel and end-cap muon trigger systems and the calorimeter trigger system quoted above, to which it contributes 100 ns.

On the sub-detector side, the timing signals are received by the Local Trigger Processor LTP [230], which acts as an interface between the CTP and the timing distribution system of each sub-detector. During stand-alone data taking of a sub-detector the LTP can generate all timing signals locally and also provides inputs for locally generated triggers. The LTP is fully programmable and can therefore act as a switch between the global and locally generated signals without the need for re-cabling. The timing distribution system of a sub-detector may be partitioned into several parts, each with its own LTP. In this case LTP's can be daisy-chained in order to save output ports on the CTP and the associated cabling. The LTP is complemented by an interface module which provides an additional input and output port, such that interconnections of sub-detectors are possible without removing the link to the CTP. Both the LTP and the interface module are implemented as 6U VMEbus boards.

From the LTP the timing signals are distributed to the detector front-end electronics using the Timing, Trigger and Control system (TTC). The ATLAS TTC system is based on the optical fan-out system developed within the framework of RD12 [231]. Clock and orbit signals synchronous to the LHC beams arrive at the machine interface module of the CTP after passing through the

2008 JINST 3 S08003

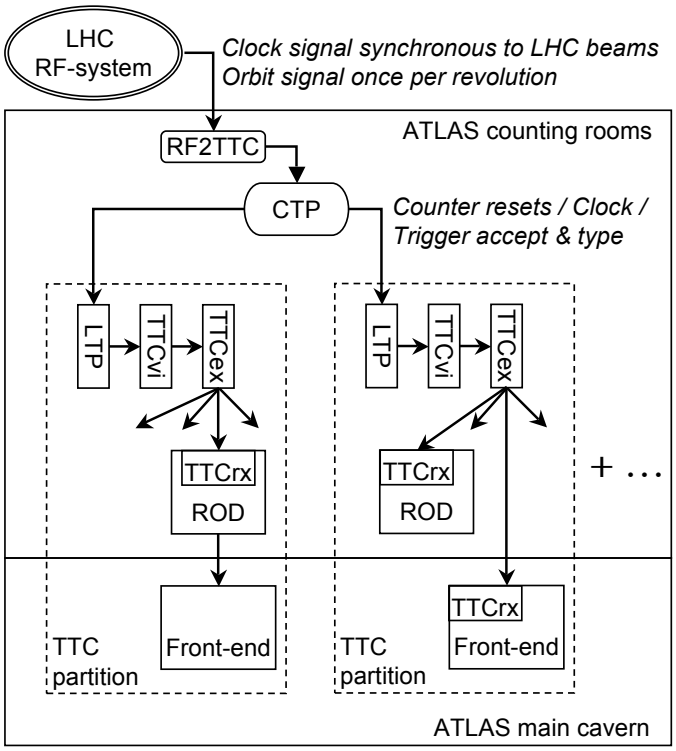

Figure 8.12: Schema of the distribution of timing signals from the LHC radio-frequency system to ATLAS and within the experiment. Here ROD (Readout Driver) more generally denotes the readout electronics in the counting rooms which receive the timing signals, while front-end denotes electronics mounted on the detector components in the main cavern.

RF2TTC (Radio-Frequency to TTC) interface module shown in figure 8.12. In the RF2TTC, the signals are cleaned and delays may be applied to account for any drift in the signal phase. From the CTP the signals are transmitted to the LTP's together with detector-specific timing and control signals like the L1A or the event counter reset signal. From the LTP onwards TTC components are available again to serialise the signals (TTCvi) and transmit them (TTCex) via optical fibres to the detector front-end electronics, where TTC receiver chips (TTCrx) decode the transmitted information and make it available as electrical signals for further use. The implementation and use of the TTC system is sub-system specific. As an example the muon trigger systems use TTC standard components to transmit the timing signals all the way to the electronics mounted on the chambers, while in case of the inner tracking detector a custom-built distribution system is used to transmit the signals from the counting rooms to the cavern.

2008 JINST 3 S08003

8.3 Data acquisition system and high-level trigger

8.3.1 Overview

As explained in section 8.1, the main components of the data acquisition system/High-Level Trigger (DAQ/HLT) are: readout, L2 trigger, event-building, event filter, configuration, control and monitoring. An overview of the event selection performed by the HLT is given in section 10.9. Here the movement of data from the detectors to the HLT and subsequently to mass storage is described. The main features of each component are described below.

A block diagram of the DAQ/HLT is shown in figure 8.1. The movement of events from the detector to mass storage commences with the selection of events by the L1 trigger. During the latency of the L1 trigger selection, up to 2.5 μs, the event data are buffered in memories located within the detector-specific front-end electronics. On selection by the L1 trigger the event data is transferred to the DAQ/HLT system over 1574 Readout Links (ROL's), having first transited through the detector-specific ROD's. The 1574 event fragments are received into the 1574 Readout Buffers (ROB's) contained in the Readout System (ROS) units where they are temporarily stored and provided, on request, to the subsequent stages of the DAQ/HLT system.

For every selected event, the L1 trigger sub-systems (calorimeter, muon, and CTP) also provide the RoI information on eight ROL's, a dedicated data path, to the RoI builder where it is assembled into a single data structure and forwarded to one of the L2 supervisor (L2SV). As its name suggests, the L2SV marshals the events within the L2 trigger. It receives the RoI's, assigns each event to one of the L2 trigger's processing units (L2PU's) for analysis, and receives the result of the L2PU's analysis.

Using the RoI information, requests for event data are made to the appropriate ROS's. The sequence of data requests is determined by the type of RoI identified by the L1 trigger and the configuration of the L2 trigger processing, i.e. the order of items in the trigger menu and the order of the algorithms per trigger item. The result, accept or reject, of the analysis is returned to the L2SV which subsequently forwards it to the DataFlow Manager (DFM). In addition to sending the result of its analysis to the L2SV, an L2PU also sends a summary of the analysis which it has performed to a L2 trigger-specific ROS.

The DFM marshals the events during the event-building. For those events which were found not to fulfil any of the L2 selection criteria, the DFM informs all the ROS's to expunge the associated event data from their respective ROB's. Each event which has been selected by the L2 trigger is assigned by the DFM to an event-building node (called SFI). The SFI collects the event data from the ROS's and builds a single event-data structure, the event. An SFI can build more than one event at a time and the requests to the ROS's for their data are dispatched according to an algorithm which ensures the quantity of data being received by the SFI does not exceed its available input bandwidth. The full event structure is sent to the event filter for further analysis. On completing the building of an event an SFI notifies the DFM, which subsequently informs all the ROS's to expunge the associated event data from their respective ROB's.

The event filter, in addition to the selection, classifies the selected events according to a predetermined set of event streams and the result of this classification is added to the event structure. Selected events are subsequently sent to the output nodes (SFO's) of the DAQ/HLT system. Con-

2008 JINST 3 S08003

versely, those events not fulfilling any of the event filter selection criteria are expunged from the system. The events received by an SFO are stored in its local file system according to the classification performed by the event filter. The event files are subsequently transferred to CERN's central data-recording facility.

8.3.2 Control

The overall control of the experiment covers the control and monitoring of the operational parameters of the detectors and experiment infrastructure, as well as the coordination of all detector, trigger and data acquisition software and hardware associated with data-taking. This functionality is provided by two independent, complementary and interacting systems: the data acquisition control system, and the Detector Control System (DCS). The former is charged with controlling the hardware and software elements of the detectors and the DAQ/HLT needed for data-taking, while the DCS handles the control of the detector equipment and related infrastructure. The DCS is described in section 8.5.

The DAQ/HLT system and detector systems are composed of a large number of distributed hardware and software components which in a coordinated manner provide for the data-taking functionality of the overall system. Likewise, their control and configuration is based on a distributed control system. The control system has two basic components: the process manager and the run control.

On each computer a process management daemon waits for commands to launch or interrupt processes. On the reception of such commands it interrogates the access manager and the resource manager to ascertain whether the requested operation is permitted. It is a task of the access manager to indicate whether the requester is authorised to perform the operation, while it is a task of the resource manager to check that the resources are available to perform the operation.

A hierarchical tree of run controllers, which follows the functional de-composition into systems and sub-systems of the detector, steers the data acquisition by starting and stopping processes and by carrying all data-taking elements through a finite state machine, which ensures that all parts of ATLAS are in a coherent state. As with the handling of commands to the process manager, run control commands also have to be authorised by the access manager. In addition to implementing a global finite state machine and managing the lifetime of processes, the run controllers are further customised according to the sub-system for which they are in charge. One example of a customised controller is the root controller, the starting point of the run control tree, which retrieves the run number from the run number service before starting any new run and drives luminosity block changes during the data-taking.

Another fundamental aspect of the control is the diagnostic and error recovery system. Several aspects of it are integrated into the run control processes. Errors raised by any data-taking node enter the error reporting system and can elicit an appropriate reaction. The diagnostic system can launch a set of tests to understand the origin of the reported problem and the recovery system can then take corrective actions. These aspects of the control have been implemented using an expert system. In order to allow for analysis of errors a posteriori, all error and information messages are archived in a database via the log service. A user interface is provided for efficient searching of messages.

2008 JINST 3 S08003

Computer management and monitoring. Access to the experiment's local area network, hence all computers located at the experimental site, is gained via an application gateway. User accounts and passwords are stored in a central directory, which is used to authenticate all the users for logging into the computers associated with the experiment. In addition, the same directory holds the configuration of various system services, such as the servers of the auto-mounted directories, and the user roles and policies required by the role-based access control scheme adopted.

All PC's and single board computers boot over the network using the pre-boot execution environment. The kernel and boot image files for the nodes are provided by a system of Local File Servers (LFS's) each of which serve approximately thirty clients. The initial boot image is a reduced version of Scientific Linux CERN (SLC), which provides the minimum set of binaries and libraries to operate a node. The remaining non-essential parts of the operating system, e.g. the X-window environment, are then loaded via the networked file system from the LFS's. The LFS's also provide, again via the networked file system, the ATLAS software and disk space to the nodes they serve.

A Central File Server (CFS) holds the master copy of the ATLAS software, which is distributed to other CFS's and to the LFS on a daily basis or on request. The unique installation of the operating system and the ATLAS software in conjunction with booting over the network, ensures the uniformity of the software throughout the computer cluster.

The disks served by the LFS are primarily used by the disk-less nodes as scratch space and for storage of detector specific software. Another use of the LFS is for running various central services related to the configuration and control of the trigger and data acquisition systems, thus providing a hierarchical structure for these services to provide the required scaling. Examples are the information server and remote database server (see section 8.3.3).

All nodes are monitored by a customised version of Nagios, an application which monitors the health of the cluster by performing checks of various node services, e.g. disk utilisation, at regular intervals. The results of these checks are displayed by Nagios on a web page and recorded in a database for subsequent retrieval and analysis. Based on the collected data, Nagios allows the graphical tracking and analysis of the health of the cluster or an individual node. It is also configured to set alarms, send notifications via emails and SMS (Short Message Service) messages and, for some services, perform recovery operations, for example the restarting of a network interface.

8.3.3 Configuration

The description of the hardware and software required for data-taking are maintained in configuration databases. A configuration is organised as a tree of linked segments according to the hierarchy describing the DAQ/HLT and detector systems. A segment defines a well defined sub-set of the hardware, software, and their associated parameters. The organisation of the data is described by common object-oriented database schemas which may be extended to describe the properties of specific hardware and software.

To support concurrent access to the configuration data by thousands of applications, and to notify control applications of changes to the configuration data during run time, remote database servers are used to ensure that access times to the configuration data do not scale with the number

2008 JINST 3 S08003

of deployed applications. Additional servers are also deployed to cache the results of queries to the relational databases, e.g. the conditions database.

Trigger configuration. At any point in time the complete trigger chain needs to be consistently configured. For L1, i.e. the CTP, a trigger menu comprising up to 256 trigger items which should cause an event to be selected, see section 8.2.3, is defined. Moreover the calorimeter and muon trigger systems have to be configured such that they deliver the information required by the trigger menu (multiplicities for trigger thresholds).

To ensure a coherent and consistent configuration of the L1, L2, and event filter, all components of the trigger are configured using an integrated system. It makes the configuration parameters available to all systems participating in the L1 trigger decision, and to all nodes forming the HLT farms.

The trigger configuration system itself contains a central relational database which stores the configuration data, tools to populate the database and ensure its consistency, the functionality to archive the configuration data and interfaces to extract the data in preparation for data taking or other purposes, e.g. simulation and/or data analysis. A detailed description of the system is given in [232].

Partitioning. Partitioning refers to the ability to operate subsets of the ATLAS detector in parallel and disjointly, thus facilitating the concurrent commissioning and operation of subsets of the detector. Once two or more partitions have been commissioned they may then be operated together as a single partition. Partitions are in this way combined into a fully integrated and operational detector.

A partition maps to a TTC partition, therefore defining the subset of detector components within a partition (see section 8.2.3.2 and also table 8.4). In addition, the static point-to-point connections between the detector ROD's and the ROS's uniquely associates a set of ROS's to a partition. Other components of the DAQ/HLT (i.e. event-building nodes, event filter nodes and SFO's) are connected by multi-layered networks and can therefore be assigned to a partition as required by the operations to be performed. The management of resources, e.g. event-building nodes, between partitions is achieved by the resource manager.

The RoI builder drives the input to the L2 trigger and, from an operational perspective, can only be operated as a single unit. Thus the L2 trigger cannot be partitioned. Analogously, the CTP can only be operated as a single unit, therefore the complete L1 trigger may not be operated in more than a single partition. However, to facilitate calibration and checking of L1Calo input signals, it is possible to operate a partition which consists of the L1Calo and the LAr and/or tile calorimeters but independent of the CTP. Similarly, the L1 muon trigger can operate, for example the RPC's with the MDT's, as a separate partition without the CTP.

8.3.4 Monitoring and information distribution

The monitoring component provides the framework for the routing of operational data and their analysis. Operational data ranges from physics event data, to histograms and the values of parameters. The routing of operational data is performed by the information, on-line histogramming

2008 JINST 3 S08003

and event monitoring services. The information service provides the distribution of the values of simple variables. Consumers of the information are able to subscribe to notifications of changes to one or more information items. It also provides a means for any application to send commands to any of the information providers, specifically for the control of information flow, e.g. an application may ask a particular provider of information to increase the frequency at which it publishes a particular piece of information. Complementing the exchange of the values of simple variables, the message reporting system transports messages among trigger and data acquisition applications. Messages may be used to report debug information, warnings or error conditions. The message reporting system allows association of qualifiers and parameters to a message. Moreover, receivers of messages are able to subscribe to the service to be notified about incoming messages and apply filtering criteria.

The On-line Histogramming Service (OHS) extends the functionality of the information service to histograms, in particular raw and ROOT histograms. Within the DAQ/HLT there are many instances of the same application, e.g. L2PU's, active at any one time producing histograms. Via the OHS, a gatherer application sums histograms of the same type and in turn publishes, via the OHS, the resulting histograms. The visual presentation of histograms is based on ROOT and Qt, and allows for the presentation of reference histograms, fitting, zooming and the sending of commands to histogram providers.

The event filter processing application is based on the off-line computing framework. The substitution of the selection algorithms with a monitoring or calibration algorithm allows for monitoring and/or calibration tasks based on the offline computing framework to operate on-line, receiving events from the SFI's. It is also possible to configure these applications to receive events from the event monitoring service. The latter provides a framework to enable the sampling and distribution of event data as they flow through the DAQ/HLT system. Monitoring applications are able to request event fragments according to the values of elements in the event fragment, e.g. trigger and/or sub-detector types, from a specific sampling point, e.g. a particular ROS (part of an event) or SFI (a complete event). Examples of monitoring applications using this service are the event dump and event display.

To complement the viewing and analysis of histograms by an operator, a data quality monitoring framework provides the automatic comparison of recently acquired data to reference data (e.g. reference histograms), statistical checks and alarm generation. More specifically, user-supplied algorithms and/or reference data are used to automatically analyse the large quantities of monitoring data, and generate alarms when deviations from the specified criteria occur.

8.3.5 Readout system

As described in section 8.3.1, the Readout System (ROS) receives event data from the detector ROD's via 1574 ROL's. All of the ROL's have the same design and implementation, based on the S-link interface. It allows for the transmission of 32-bit data at 40.08 MHz, i.e. up to 160 Mbyte/s, and implements flow control and error detection [233]. ROB's are the buffers located at the receiving end of the ROL's, there being one ROL associated to one ROB. Three ROB's are physically implemented on a module called a ROBIN, and up to six ROBIN's can be located in a ROS, which is implemented on a server-class PC. The ROS provides the multiplexing of up to 18 ROL's to the

2008 JINST 3 S08003

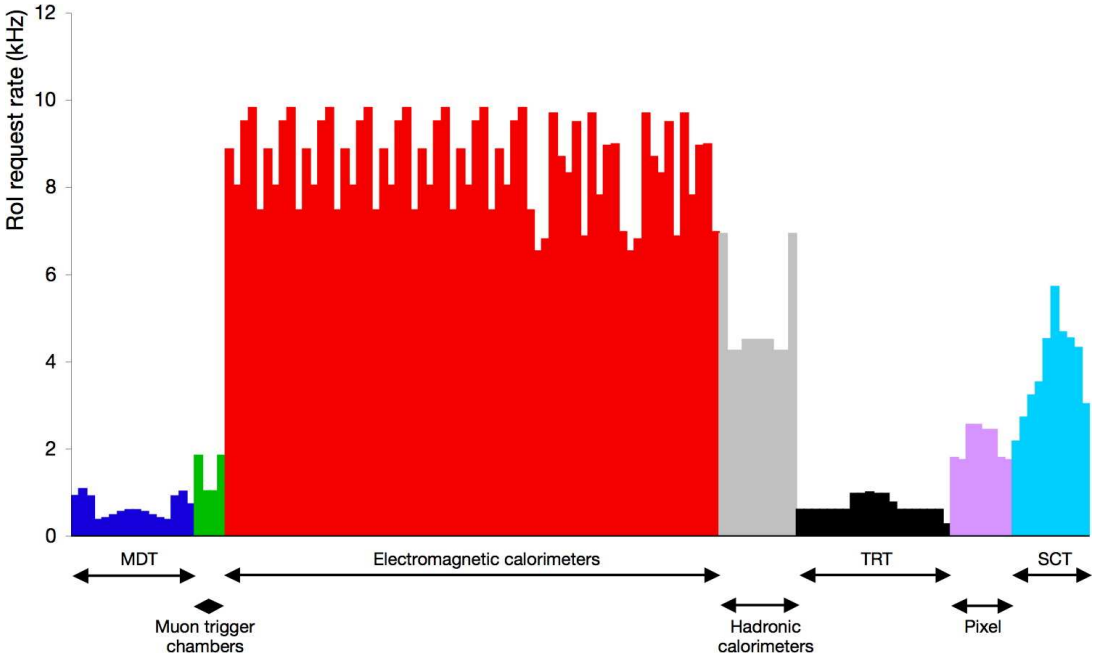

Figure 8.13: Expected average RoI request rate per ROS for a luminosity of 10^{33} cm^{-2} s^{-1}.

subsequent components of the DAQ/HLT, i.e. L2 trigger and event-building, reducing the number of connections by approximately an order of magnitude.

A request by an L2PU for data involves, on average, one or two ROB's per ROS, whereas the requests for data from the event-building nodes concern the event data from all the ROB's of a ROS. In either case, the ROS replies to the requester with a single data structure. At the L1 trigger rate of 75 kHz, and an average of 1 kbyte received per ROL, the ROS is able to concurrently service up to approximately 20 kHz of data requests from the L2 trigger, up to 3.5 kHz of requests from event-building nodes, and expunge events on request from the DFM. The rate of data requests received by a specific ROS depends on the $\eta - \phi$ region of the data it receives over the ROL's and from which detector it receives data. For example, a ROS which receives data from the liquid-argon calorimeter barrel region is solicited for data more frequently than a ROS associated with the barrel MDT's. The expected average rate of RoI requests as a function of 135 ROS's, which participate to the L2 trigger is shown in figure 8.13 (see section 10.9.3 for examples of initial trigger menus). Figure 8.14 shows the expected maximum L1 trigger accept rate sustainable by the ROS which is most solicited for RoI data by the L2

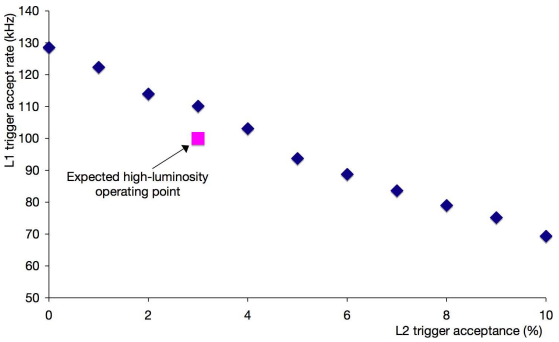

Figure 8.14: The maximum sustainable L1 trigger accept rate as a function of the L2 trigger acceptance for the ROS which is most solicited for RoI data by the L2 trigger. Also shown is the expected operating point at high luminosity.

2008 JINST 3 S08003

trigger, for different values of the L2 trigger's acceptance, i.e. the event-building rate. Also shown is the expected operating point at high luminosity.

ROBIN. The ROBIN component provides the temporary buffering of the individual event fragments produced by the ROD's for the duration of the L2 trigger decision and, for approximately 3% of the events, for the duration of the event-building process. In addition, it services requests for data at up to rates of approximately 20 kHz. As a consequence of the rates which have to be supported, the ROBIN is a custom-designed and built PCI-X mezzanine [234]. All functions related to the receiving and buffering of event fragments are realised in an FPGA. A PowerPC is used to implement the functions of memory management, servicing of data requests, control and operational monitoring.

8.3.6 L2 trigger

The L2 trigger is achieved by the combined functionality of the RoI builder, L2SV, L2PU and L2 trigger-specific ROS (pseudo-ROS). The RoI builder receives the RoI information from the different sources within the L1 trigger on eight input ROL's and merges them into a single data structure. It is thus at the boundary between the L1 and L2 trigger systems and operates at the L1 trigger rate (see next section). The single data structure containing the RoI data is transmitted by the RoI builder over one of the output ROL's to the L2SV's. As described in section 8.3.1, L2SV's marshal the events through the L2 trigger.

The principal component of the L2 trigger is the L2 processing farm, where the event selection is executed. The system is designed to provide an event rejection factor of about 30, with an average throughput per farm node of about 200 Hz, using (but not exclusively, see section 10.9.4.5) only the data located in the RoI's, i.e. 1–2% of the full data of an event. The number of L2PU applications performing the physics selection per node is configurable. On the hardware currently deployed (see section 8.4) there are eight L2PU's per node, and one L2PU per processing core of the node, thus the average event processing time per L2PU should be less than 40 ms.

The transmission of a summary of the L2 trigger's selection is achieved by the deployment of a pseudo-ROS. At the end of its event analysis the L2PU sends to the pseudo-ROS information which summarises the results of its analysis. Subsequently, the pseudo-ROS participates in event-building like any other ROS within the system, its event data being the L2 trigger's summary analysis. In this way, the results of the L2 trigger's analysis are built into the final event and subsequently used by the event filter to seed its selection.

The failure of one or more L2PU's during run time does not incur system down time. The system continues to operate at a reduced rate while the failed application, the L2PU, can be restarted under the supervision of the run control.

Steering of the event selection. The HLT starts from the RoI's delivered by the L1 trigger and applies trigger decisions in a series of steps, each refining existing information by acquiring additional data from increasingly more detectors. A list of physics signatures (trigger chains), implemented event reconstruction (feature extraction) and selection algorithms are used to build signature and sequence tables for all HLT steps. Feature extraction algorithms typically request detector data

2008 JINST 3 S08003

from within the RoI and attempt to identify features, e.g. a track or a calorimeter cluster. Subsequently, a hypothesis algorithm determines whether the identified feature meets the criteria (such as a shower shape, track-cluster match or E_T threshold) necessary to continue. Each signature is tested in this way. The decision to reject the event or continue is based on the validity of signatures, taking into account pre-scale and pass-through factors. Thus events can be rejected early after an intermediate step if no signatures remain viable. In this manner the full data set associated with the RoI is transferred only for those events which fulfil the complete L2 trigger selection criteria, i.e. the amount of data transferred between the ROS's and the L2 trigger is minimised for those events which are rejected. The stepwise and seeded processing of events in the HLT is controlled by the steering.

The steering runs within the L2 and event filter processing tasks. It implements two of the key architectural and event-selection strategies of the trigger and data acquisition systems: RoI-based reconstruction and step-wise selection. Both are designed to reduce processing time and L2 network bandwidth. The steering takes the static configuration described in section 8.3.3 and applies it to the dynamic event conditions (which RoI's are active and the status of their reconstruction) in order to determine which algorithms should be run on which RoI's and in which order, and ultimately to decide whether the event has fulfilled the criteria for acceptance.

Counters, maintained by the steering, for each trigger chain and step are made available by the online monitoring system, see section 8.3.4, enabling the real time monitoring of the trigger rates. The steering also provides a mechanism for the selection algorithms to publish parameters necessary for monitoring the quality of the event selection.

Region-of-interest builder. The RoI builder [235] is one of only three custom-built components within the DAQ/HLT system. It is a 9U VMEbus system composed of an input stage, an assembly stage, and a single-board computer for the purpose of configuration, control and operational monitoring. The input stage consists of three input cards which each receive and buffer, over three ROL's, the RoI data from three of the eight L1 trigger sources, namely: four CP ROD's, two JEP ROD's, the MUCTPI, and the CTP.

The eight RoI fragments are subsequently routed over a custom back-plane to builder cards in the assembly stage, where they are assembled into a single data structure (RoI record). The assignment of each event to a specific builder card for assembly is based on a token-passing mechanism between the builder cards. Each builder card has four output ROL's which are used to transfer the assembled RoI records to up to four L2SV's according to a round-robin algorithm.

Detector calibration using RoI data. The calibration of the muon MDT chambers requires large data samples within a well-defined time window to establish the relationship between the drift path and measured time as a function of time. This measurement has to be made from the data of the MDT's alone using candidate tracks, and is based on an iterative procedure starting from a preliminary set of constants.

The L1 single-muon trigger rate at a luminosity of 10^{33} cm^{-2} s^{-1} for a threshold of 6 GeV is approximately 25 kHz. For these candidate events the first step of the L2 trigger selection is the reconstruction of tracks in the muon system. To facilitate the calibration of the MDT's, each L2PU can be configured to additionally write, to a pre-defined buffer, the data of the candidate tracks and

2008 JINST 3 S08003

the results of its analysis, i.e. the RoI information and the results of track fits. Subsequently the data are transferred from this buffer to a L2-wide calibration server which stores the data to disk prior to sending it to a remote calibration farm for processing.

8.3.7 Event-building

The event-building functionality is provided by the DFM, ROS's and SFI's [236]. The SFI is the application which collects the event data from the ROS's and assembles the event as a single formatted data structure. An SFI is configured with a randomised list of the ROS's within the system, which is used to define the order in which data requests are sent to the ROS's. This results in the randomisation of the traffic pattern in the underlying network and hence improved network performance. To meet the rate requirements a number of SFI's work in parallel, each instance building a number of events concurrently. Each SFI informs the DFM of its readiness to receive events, and the DFM allocates events to the SFI's so as to ensure that the load is balanced across all available SFI's.

The default behaviour of the SFI is to collect all the event data associated with a given event into a single formatted data structure. However, a subset of the events accepted by the L2 trigger are for the purposes of detector calibration and do not necessitate the collection of all the event data. For this type of event, the SFI is capable of collecting a subset of the available event data. The subset is identified by the L2 trigger and communicated to the SFI via the DFM. The subset can range from a few ROB's to a whole sub-detector.

If a requested ROS data fragment is not received within a configurable time budget, the outstanding data fragment can be requested again. Only if several consecutive requests are not fulfilled does the SFI abandon the inclusion of the missing data and assemble an incomplete event. After an event has been moved to the event filter the SFI marks its buffers for re-use. If, for whatever reason, the buffers of the SFI become full, the SFI informs the DFM, i.e. exerts back-pressure, which subsequently suspends the allocation of events to the specific SFI until the SFI indicates it is again available.

The event-building system is designed to function even in case of failure of one or more SFI nodes. In this situation, the DFM ceases to assign events to the failed SFI's. Once the failed nodes become available again, they can be re-integrated into the event-building system without the system incurring down time.

8.3.8 Event filter

The event filter is a processing farm; on each processing node a configurable number of independent processing tasks receive and process events. Unlike the L2 trigger, these tasks are based on standard ATLAS event reconstruction and analysis applications. The steering of the event selection is the same as L2, as described in 8.3.6. For those events passing the selection criteria, a subset of the data generated during the event analysis is appended to the event data structure, enabling subsequent offline analysis to be seeded by the results from the event filter. An integral part of the selection process is the classification of the events according to the ATLAS physics streams, see section 8.3.9. To this end, for those events which fulfil the selection criteria, a tag is added to the event data structure identifying into which physics stream the event has been classified.

2008 JINST 3 S08003

Table 8.2: Overlap (Hz) between the data streams at a luminosity of 10^{33} cm^{-2} s^{-1}.

Stream	e	μ	Jet	γ	E_T^{miss} & τ	B-physics
e	31 ± 7.9	0.0056 ± 0.00058	$0.00053 \pm 6.2 \times 10^{-5}$	1.2 ± 0.4	1.4 ± 0.035	$(1.3 \pm 1.3) \times 10^{-5}$
μ	–	34 ± 8.7	0.021 ± 0.015	0.0028 ± 0.002	0.22 ± 0.022	0.076 ± 0.0043
Jet	–	–	38 ± 5.9	0.48 ± 0.4	0.71 ± 0.4	0 ± 0
γ	–	–	–	22 ± 5.7	0.22 ± 0.073	0 ± 0
E_T^{miss} & τ	–	–	–	–	32 ± 7.9	$(15 \pm 6.4) \times 10^{-6}$
B-physics	–	–	–	–	–	9.5 ± 5.5

The failure of one or more event-filter processing tasks or of a complete node during run-time does not provoke any system down-time. The system continues to operate at a reduced rate while the failed application, or node, can be restarted under the supervision of the run control. To ensure that no events are lost during such failures, each event on arrival in the event filter is written to a memory mapped file. On the restart of the failed application or of the node itself, an attempt can be made to re-analyse the event or accept the event without analysis.

8.3.9 Event output

The main functionality of the event-filter output nodes (SFO's) is to receive events which have passed the event filter selection criteria, interface the DAQ/HLT to CERN's central data-recording facility, and de-couple the data-taking process from possible variations in the central data-recording service.

The SFO maintains, locally, a set of files into which it records events at a peak event rate of up to 400 Hz. In the eventuality of a prolonged failure in the transmission of data to CERN's central data recording service, there is sufficient local storage capacity to buffer all events locally for up to 24 hours. Under normal operating conditions, this storage capacity is only partially used. The set of files maps to the ATLAS-defined data streams: electrons, muons, jets, photons, E_T^{miss} and τ-leptons, and B-physics. Each event is recorded in one or more files according to the stream classification made by the event-filter processing task. Table 8.2 shows the rates for each of the data streams and in the off-diagonal elements, the rates of the overlaps between them.

In addition to the data streams mentioned above, a subset of the events is also written to calibration streams and an express stream. The express stream is a subset of the events selected by the event filter and fulfil additional criteria which select the events as being useful for monitoring the quality of the data and the detector. The calibration stream provides the minimum amount of information needed for detector calibration, possibly at a rate higher than the data streams provide. These events will only contain a subset of the event data.

8.4 Implementation and capabilities of the DAQ/HLT

Most of the DAQ/HLT functionality is implemented on commodity, rack-mountable, server-class PC's. The PC's run Scientific Linux CERN and are interconnected by multi-layer gigabit-Ethernet networks, one for control functionality and another for data movement. The majority of PC's have similar specifications (e.g. two CPU sockets, two gigabit-Ethernet connections, support for

2008 JINST 3 S08003

Table 8.3: The main data-acquisition system components to be deployed for initial operation: the readout system (ROS), the event-building node (SFI), the dataflow manager (DFM), the L2 supervisor (L2SV), the high-level trigger (HLT) and the event filter output nodes (SFO).

Component	Number of nodes	Number of racks	Number of CPU's/node	Memory (Gbyte)	Type of CPU
ROS	145	16	1	0.512	3.4 GHz Irwindale
SFI	48	3	2	2	2.6 GHz Opteron 252
DFM	12	1			
L2SV	10				
HLT	1116	36	8	8	Xeon E5320 1.86 GHz
SFO	6	2	2	4	Xeon E5130 2.0 GHz
Monitoring	32	4	4	8	Xeon E5160 3.0 GHz
Operations	20		2	4	Xeon E5130 2.0 GHz

IPMIv2.0), and differ only by the number and type of CPU's implemented and the amount of memory. The main features per component and the number of nodes deployed for initial operations in 2008 are given in table 8.3. A few components, the RoI builder, ROL and ROBIN, are, however, implemented in custom hardware.

The ROS PC's are installed in standard ATLAS 52U racks, while all other PC's are installed in standard 47U or 52U server racks. The number of racks (for initial operation) for each component type is given in table 8.3. In addition to the PC's, each rack also contains a local file server and two gigabit-Ethernet switches. The latter form part of the multi-layered gigabit-Ethernet network which implements the control and data networks. Each rack is also equipped with a water-cooled heat-exchanger, designed for the horizontal airflow within a rack, which provides up to 9.5 kW of cooling power. The number of 1U PC's per rack is typically just over thirty, constrained by cooling power, power distribution (particularly in-rush current) and weight limits.

For initial operations, the DAQ/HLT system will be fully configured in the area of configuration, control and monitoring functionality. The operations PC's are used to provide the various central services for configuring and controlling the trigger and data acquisition systems (e.g. run control, error logging). The monitoring PC's are used to monitor the system and sampled event data.

The initial system will also support full detector readout, over the 1574 point-to-point ROL's, into the ROS's at L1 trigger rates up to 75 kHz. The number of ROD's, ROL's and ROS's per detector TTC partition are given in table 8.4. Also given in this table is the expected size of event data per L1 trigger for each part of the detector for a luminosity of 10^{34} cm^{-2} s^{-1}.

As described in section 8.3.7, the event-building functionality is performed by a set of SFI's and scales linearly with the number of SFI's, each SFI contributing 60 Hz and approximately 90 Mbyte/s to the total event-building rate and aggregate bandwidth. For initial operations, forty-eight SFI's are deployed allowing a sustained event-building rate of approximately 2.0 kHz, for an average event size of approximately 1.3 Mbyte.

2008 JINST 3 S08003

Table 8.4: Numbers of readout drivers (ROD's), readout links (ROL's) and readout systems (ROS's) per detector TTC partition, as well as expected data size per L1A signal for a luminosity of 10^{34} cm^{-2} s^{-1}.

TTC Partition			Number of ROD's	Number of ROL's	Number of ROS's	Data per L1A signal (kbyte)
Inner detector	Pixel	Layer 0	44	44	4	60
		Disks	24	24	2	
		Layers 1–2	64	64	6	
	SCT	End-cap A	24	24	2	110
		End-cap C	24	24	2	
		Barrel A	22	22	2	
		Barrel C	22	22	2	
	TRT	End-cap A	64	64	6	307
		End-cap C	64	64	6	
		Barrel A	32	32	3	
		Barrel C	32	32	3	
Calorimetry	Tile	Barrel A	8	16	2	48
		Barrel C	8	16	2	
		Extended barrel A	8	16	2	
		Extended barrel C	8	16	2	
	LAr	EM barrel A	56	224	20	576
		EM barrel C	56	224	20	
		EM end-cap A	35	138	12	
		EM end-cap C	35	138	12	
		HEC	6	24	2	
		FCal	4	14	2	
Muon spectrometer	MDT	Barrel A	50	50	4	154
		Barrel C	50	50	4	
		End-cap A	52	52	4	
		End-cap C	52	52	4	
	CSC	End-cap A	8	8	1	10
		End-cap C	8	8	1	
L1	Calorimeter	CP	4	8	1	28 (can be varied)
		JEP	2	8	1	
		PP	8	32	3	
	Muon RPC	Barrel A	16	16	2	12
		Barrel C	16	16	2	
	Muon TGC	End-cap A	12	12	1	6
		End-cap C	12	12	1	
	MUCTPI		1	1	1	0.1
	CTP		1	1	1	0.2
		Total	932	1574	145	1311

In addition to the features given in table 8.3, the PC's for the SFO functionality are each equipped with three RAID controllers each managing eight 500 Gbyte SATA II disks. The three sets of disks are operated as a circular buffer: while events are being written to the event streams of one set of disks, a second set of disks is used to send data to CERN's central data recording service. In this configuration, a single RAID controller does not perform both writing and reading operations simultaneously, thus maximising the throughput of an SFO. The deployed set of SFO's fulfil the final design specifications: a sustained output bandwidth of 300 Mbyte/s and a peak rate of 600 Mbyte/s. Thus for an average event size of 1.3 Mbyte, this gives a sustained event rate of 200 Hz.

2008 JINST 3 S08003

Of the 36 HLT racks available for initial operations, eight racks (248 nodes) are dedicated to the event filter selection while the remaining twenty-eight racks (868 nodes) can be configured to perform either the L2 trigger or event filter selection, i.e. the amount of computing power apportioned to the L2 trigger and or the event filter will be adjusted according to the data-taking conditions. The baseline apportioning of these nodes envisages nine racks (279 nodes) for the L2 trigger and twenty-seven racks (837 nodes) for the event filter. In the DAQ/HLT Technical Design Report [237], the algorithm processing times and rejection rates were based on single-core processors with an expected clock speed of about 8 GHz, giving processing times per event of order 10 ms at L2 and 1 s at the event filter. These figures correspond to approximately 40 ms and 4 s respectively per core on today's quad-core processors operating at a clock speed of 2 GHz. Measurements with simulated raw data show that the processing times per event at L2 and event filter are consistent with the available computing resources for acceptable trigger rates and with a representative mixture of simulated events passing the L1 trigger. Therefore the system deployed for initial operations should be able to handle an initial L1 trigger rate of approximately 40 kHz, about half of the final design specification.

8.5 Detector control system

In order to enable coherent and safe operation of the ATLAS detector, a Detector Control System (DCS) has been defined and implemented [238]. The DCS puts the detector hardware into selected operational conditions, continuously monitors and archives its run-time parameters, and performs automatically corrective actions if necessary. Furthermore, DCS provides a human interface for the full control of ATLAS and its sub-detectors. Figure 8.15 shows the general system architecture consisting of two parts: the front-end systems and the back-end control.

The front-end consists of the hardware components used to control the detector, ranging from simple sensors to complex devices such as software controlled power supplies. In order to minimise the effort of integration of devices into the DCS and to achieve a homogeneous system, a small set of commercial devices, such as crates or power supplies, has been selected as standard. The readout of these devices is normally done using the industry-standard protocol OPC.

Due to the special conditions in the experiment cavern, strong magnetic field and ionising radiation, a general-purpose I/O concentrator, the Embedded Local Monitor Board (ELMB) [239], has been developed. An ELMB comprises a multiplexed ADC (64 channels with 16 bit resolution), 24 digital I/O lines and a serial bus SPI to drive external devices. The ELMB can be configured for various types of sensors. A micro-controller pre-processes the data (e.g. calibration, threshold detection) before they are transferred via CANbus to the back-end. The ELMB is designed and tested to be radiation tolerant to a level of about 1 Gy/y and can hence also be placed inside the detector at places shielded by the calorimeter. The ELMB is used in two ways: either directly embedded in the detector electronics, or attached to a general-purpose motherboard to which sensors can be connected. In total about 5000 ELMB's are installed and they are controlled by an OPC server using the CANopen protocol.

The back-end is organised in 3 layers: the Local Control Stations (LCS) for process control of subsystems, the Sub-detector Control Stations (SCS) for high-level control of a sub-detector allowing stand-alone operation, and the Global Control Stations (GCS) with human interfaces in

2008 JINST 3 S08003

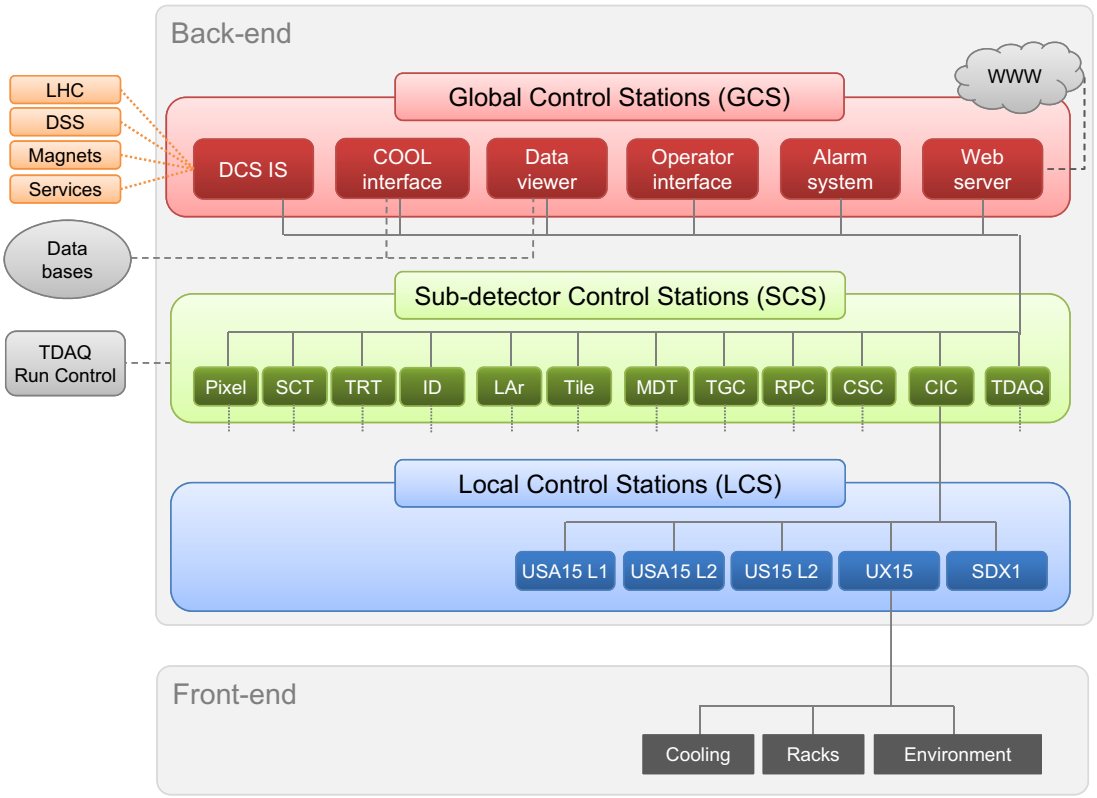

Figure 8.15: Architecture of the DCS.

the ATLAS control room for the overall operation. Each station is a PC running the commercial controls software PVSS-II [240], which provides the necessary supervisory functions such as data analysis with the possibility of triggering the execution of pre-defined procedures, raising alarms, data visualisation, and archiving. This supervisory control and data acquisition software package (SCADA) has been chosen for all LHC experiments in the frame of the Joint Controls Project [241], which provides a set of tools on top of PVSS-II and also software components for the standardised devices. In total, the back-end consists of more than 150 PC's, connected as a distributed system for which PVSS-II handles inter-process communication via the local area network. The full back-end hierarchy down to the level of individual devices is represented by a distributed finite state machine allowing for the standardised operation and error handling in each functional layer.

The LCS are connected to the front-end of a specific sub-system and read, process, and archive the respective data. They execute the commands issued from the SCS and GCS layers and additionally allow the implementation of closed-loop control.

The SCS enable full stand-alone operation of a sub-detector by means of the finite state machine and provide a user interface to control the different subsystems. The SCS also handles the synchronisation with the data acquisition system.

The GCS in the top layer provide all functions needed in the ATLAS control room to operate the complete detector. The operator interface shows the detector status and allows to navigate in the tree-like finite state machine hierarchy and to execute high level commands which are propagated

2008 JINST 3 S08003

to the layers below. A data viewer provides selection and plotting of all data available in the DCS. The alarm system collects and displays all parameters which are outside of pre-defined ranges, classified in three severity levels: Warning, Error, and Fatal. An information server (DCS IS) provides an interface to external control systems such as the sub-detector gas systems, the liquid-argon and helium cryogenics systems, the ATLAS magnets, and the electricity distribution for the detector. The data collected from those systems is made available inside the distributed DCS to the individual sub-detectors. Summarised status information is openly available via the World Wide Web.

In addition to the individual sub-detector control stations, one SCS is dedicated to Common Infrastructure Control (CIC) to supervise the common environment and services of the detector. Each of the five geographical zones defined for the CIC (three electronics rooms underground, the cavern of the experiment, and the trigger and data acquisition systems computer rooms) is controlled by an LCS. In each zone, the environmental parameters are monitored, the operational parameters of the electronics racks are supervised, and the electricity distribution is controlled. Furthermore, the CIC includes a large network of I/O points in the experimental cavern, consisting of about 100 ELMB's. It reads the data of radiation monitors, the movement sensors of the safety system to look for personnel inside the ATLAS area (see section 9.9), and some 200 temperature sensors positioned at the support structure of the barrel toroid. Additional readout capacity for future system extensions is available.

The DCS comprises a set of common software tools and packages, used by sub-detector controls and the CIC. A configuration database stores the settings needed for the different operational modes of the detector. All status information and measured data can be transferred to the ATLAS-wide conditions database COOL. Another package allows the synchronisation and information exchange between the DCS and the data acquisition system.

2008 JINST 3 S08003