View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by CERN Document Server

THE H1 TRIGGER SYSTEM

Pál Ribarics
Max-Planck-Institut für Physik, München, Germany
Central Research Institut of Physics, Budapest, Hungary

Abstract

To deal with the low ep physics cross section at HERA, large proton and electron beam currents are required, which give rise to large machine background - typically 10^5 times larger than the rate coming from physics. This background situation, the short bunch time intervall of 96 ns and the request for low deadtime of the readout system result in a new challenge for a collider experiment: a centrally clocked fully pipelined front-end system keeps the detector information stored during the first level trigger calculations. The central trigger decision is distributed again to the subdetectors to stop the pipelines. This scheme makes the first level trigger completely deadtime free. Together with pipelining a 4 level trigger scheme is used in order to obtain a rate acceptable for the data logging on the storage media (5 Hz). Each level reduces significantly the rate of event candidates and allows more time for a more sophisticated decision at the subsequent level. At L2 a new technology, artificial neural networks are implemented.

1 Introduction: the H1 experiment

HERA is a large proton-electron collider in Hamburg. Nearly 4000 magnets are used to collide 30 GeV electrons with 820 GeV protons in two separated rings 7 km in circumference. The main dipole and quadrupole magnets of the proton ring are all superconducting. H1 is one of two large experiments currently exploiting this unique facility, involving over 300 scientists from more than 30 institutes worldwide. The detector system allows to study the deep inelastic scattering in a new kinematical domain, achieving 2 orders of magnitude increase in Q^2 to 98400 GeV^2.
The detector, itself, consists of a central and a forward tracking system, each containing different layers of drift chambers and trigger proportional chambers. The liquid argon cryostat surrounds the trackers. It houses the lead absorber plates and readout gaps of the electromagnetic section, which are followed by the steel plates of the hadronic section with their readout gaps. A superconducting cylindrical coil with a diameter of 6 m provides the analysing field of 1.15 T. The iron return yoke of the magnet is laminated and filled with limited streamer tubes. The small fraction of hadronic energy leaking out of the back of the calorimeter is registered here and muon tracks are identified. Muon identification further benefits from additional chambers inside and outside of the iron. Stiff muon tracks in forward direction are analysed in a supplementary toroidal magnet sandwiched between drift chambers. The remaining holes in the acceptance of the liquid argon calorimeter are closed with warm calorimeters, a Si-Cu plug at very forward angles, a Pb-scintillator calorimeter backed by a tail catcher (part of the muon system) in backward direction and lastly an electron tagger at $z = -33$ m from the interaction

Figure 1: The H1 detector

point. The tagger marks the energy of an electron with very small scattering angle inducing a photoproduction event and, taken in coincidence with a corresponding photon detector at $z = -103$ m upstream from the interaction point, monitors the luminosity by the bremsstrahlung process. Two scintillator walls in backward direction are installed to recognize background produced by the proton beam upstream of the H1 detector.

2 Trigger requirements

The purpose of the trigger system is to select interesting ep collision events and to reject background events. To deal with the relatively low ep physics cross section, large proton and electron accelerator beam currents are required, which is only possible by running in a multibunch mode. In HERA 220 p and e bunches (design values) circulate and cross each other in every 96 ns. These beams give rise to three types of background: synchrotron radiation from the electron beam, proton gas interaction in the beam pipe vacuum and stray protons, which produce particle showers by hitting the material close to the beam area. The rate of the last two is about 100 kHz which exceeds by far the physical rate. (100 Hz for photoproduction, 3 Hz for neutral current interactions for $Q^2 > 10GeV^2$). The data logging rate can not be larger than about 5 Hz, a limitation coming from the available mass storage.

This background situation, the short bunch time intervall of 96 ns and the request for low deadtime of the readout system result in a new challenge for a collider experiment: a centrally clocked fully pipelined front-end system keeps the detector information stored. During all this time the first level trigger calculations take place, their result are transported to the central trigger decision logic and the global decision is distributed again to the subdetectors to stop the pipelines. Of course the trigger calculation and decision logic has to be built in a pipelined architecture such that there is a trigger decision for each

bunch crossing separately. In such a system the first level trigger is completely deadtime free.

Most of the many subdetectors of H1 produce trigger information reflecting directly basic physics quantities. However, to allow decisions of increasing complexity, a multilevel trigger concept is being used: Following the deadtime free level 1 trigger there are two levels of synchronous trigger systems (level 2 and 3) which operate during the primary deadtime of the front-end readout and one asynchronous event filter system (level 4) consisting of a fast processor farm. This later has access to the full event information and allows online event reconstruction.

The unique feature, which distinguishes the ep events from most of the background, is their vertex from the nominal fiducial volume of the ep interaction region. Consequently we use the track origin information in several different ways: The time of flight walls give us information whether there are tracks coming from upstream by comparing the arrival time with the accelerator clock phase. The central jet chamber measures the distance of closest approach (DCA) of single tracks in the plane perpendicular to the beam axis and allows a global fit to the event origin in this plane. The central and forward multiwire proportional chambers allow a fast estimation of the position of the vertex along the beam axis.

However there is still background originating from beam gas interaction in the nominal ep interaction region or from secondary interactions in the beam pipe and the inner detector regions faking an event origin inside the fiducial volume. Further requirements on the event selection are needed, depending on the event classes looked at.

First of all hard scattering events have higher total transverse energy. Therefore both the liquid argon calorimeter and the backward electromagnetic calorimeter (BEMC) deliver information about the observed energy deposition. The missing total transverse energy of the liquid argon signal is used to identify charged current deep inelastic events, while the requirement of some electromagnetic but no hadronic energy, deposited in a given position of the liquid argon calorimeter, spots a scattered electron from a neutral current event.

There are two further event classes, for which we require special conditions: Events with an electron scattered under small angle into the electron tagger of the luminosity system (low Q^2 photoproduction) and events with muons detected in the instrumented iron or forward muon system indicating a heavy quark or exotic physics candidate. Since these signatures mark rather uniquely an ep event, the general requirements on the vertex determination and the calorimetric energy can be somewhat relaxed here.

All conditions mentioned so far have been applied already in the first level trigger. However for photoproduction and heavy quark physics, where the scattered electron remains in the beam pipe and no muon is observed in the final state with sufficient energy, triggering becomes more difficult: Here we can make use of the event topology, since proton beam induced background has a more forward oriented kinematics compared to ep events. This was done so far only in the offline analysis.

3 Front-end pipelines

The time intervall between two consecutive bunch crossings of 96 ns is used as the time unit (1 BC) in the following. The time needed to run trigger signals even through a few circuits performing simple logical calculations is usually longer than that. Moreover the large size of the experiment and the electronic trailor attached to it introduces cable delays of several BC. Finally certain detectors have a long detector response time which means that the information of these detectors is only available some BC after the event (liquid argon calorimeter 13 BC due to long integration time of the preamplifiers, central drift chamber 11 BC due to a maximum drifttime of 1 μs). Of course such a long response times can only be tolerated because due to a relatively low event rate (compared to a pp

collider) the probability for an interaction per bunch crossing is small (of order 10^{-3}).

The final L1 trigger decision (called L1keep signal) is available centrally 24 BC after the real ep event time. Further time is needed to distribute this signal to stop the various subdetector pipelines. The total pipeline length varies between 27 and 35 BC (depending on the subdetector) and turned out to be in some cases just long enough to operate the system. For future system designs we would advise to increase this pipeline length to gain more flexibility in the timing of such a system or - even better - to perform signal processing and zero suppression before entering the pipelines and store the information dynamically.

The chosen concept of a pipelined front-end system also avoids huge amount of analog cable delays and allows to reconstruct offline the history of the event over several BC for timing studies and to identify pile up.

H1 uses four different types of pipelines

- Fast random access memory (RAM) is used to store the digitised information of the drift chambers as well as of the liquid argon calorimeter (LAr) for trigger purposes. At L1keep time the overwriting of the RAMs is stopped to save the information for the readout process.
- Digital shift registers are used to store the single bit information, generated by threshold discriminators in the instrumented iron system, the multiwire proportional chambers, the driftchamber trigger branch, the backward electromagnetic calorimeter (BEMC) trigger branch and the two scintillator systems.
- Analog delay lines are used to store the pulseheight of the BEMC.
- Signal pulseshaping of the LAr is adjusted such, that the signal's maximum occurs at L1keep time. The same type of sample and hold and digitisation is used as in the BEMC case.

The timing of the synchronisation step and the analog to digital conversion clocks is critical. The information produced needs to be uniquely attached to the bunch crosssing the event originated from, such that the trigger calculations based on all channels within a subsystem and also systemwide are derived from the same bunchcrossing.

4 Multi level trigger

It is impossible to get the final trigger decision in $2.2\,\mu s$ with a rate acceptable for the data logging on the storage media (5 Hz). Therefore together with pipelining a multi-level trigger and buffering scheme is used. For H1 the final trigger decision is supposed to be done in 4 different levels (L1-L4, see Fig. 2). The higher the level, the more complex and the more time-consuming is the process. The task of each level is to reduce significantly the rate of event candidates and to allow more time for a more sophisticated decision at the subsequent level.

In 1994, only two trigger levels were used: L1 and L4. Operation with only two levels was possible due to the fact that the HERA machine was operated at $\sim 7\%$ of the design luminosity. An L1 output rate of $\sim 50\,Hz$ safely matched the L4 input limit rate.

The final trigger decision at H1 will be done on the following trigger levels:

- **L1** - An acceptable output trigger rate is 1 kHz for an expected total interaction rate of $\sim$ 50-100 kHz, i.e. the required reduction factor is $\sim$ 50-100. These numbers are valid for the full HERA and H1 trigger performance.
  The dead time free first level trigger, due to the $2.2\,\mu s$ decision time, must be based on hardwired algorithms and combines only subsets of the full information available from all subdetectors. An important property is the big flexibility in combining different trigger elements from the same subdetector.

Figure 2: The overview of the H1 trigger

- **L2** - hardware trigger with dead time, starting with L1keep. The task of L2 is to reduce the input rate of 1 kHz to about 200 Hz. The L2 decision is taken at fixed time $\approx 20\,\mu s$. The trigger elements for level 2 will be based on the same information as is used in L1, but now more time is available to combine trigger elements from different detector parts.

- **L3** - software trigger, starting in parallel with L2 further reduces the rate of triggered events to maximum 50 Hz. A dedicated μ-processor will compute the L3 decision in $800\,\mu s$ on the basis of the more complex matching of the information from different detector components. L3reject stops the readout and restarts the pipelines.

- **L4** - software filter. The aim of this level is further reduction of the data volume before it is sent to the final storage media at the DESY computer center. The calculations are performed by the processor farm on the full event data, asynchronously with the rest of the trigger. Therefore this level is also called L4 filter farm. The aim is a reduction of the final data logging rate to $\sim 5Hz$.

5 The L1 trigger

The L1 system consists of different trigger systems each based on the information of a certain subdetector. The outputs of these systems are called **trigger elements**. These trigger elements are fed to a central trigger logic where they are combined to various **subtriggers**. Each single subtrigger suffices to produce a Level 1 **trigger decision** (L1keep signal) to stop the pipelines and prepare the event readout.

5.1 Vertex position oriented trigger systems

The geometrical origin of the event is the main handle to suppress background at a HERA experiment. Vertices which lie far outside the nominal ep interaction region identify uniquely background events. These trigger elements are therefore used for almost all subtriggers as veto, with the exception of the higher threshold triggers of the calorimeters.

5.1.1 The backward time-of-flight system

Beam wall and beam gas events originating from the proton upstream direction produce showers, which mostly run through both scintillator walls behind the BEMC. A background (BG) and an interaction (IA) timing window define for each scintillator of each wall whether the hits belong to particles originating from the upstream region or from the nominal interaction region. The ToF-BG trigger element is the simplest and most effective background rejection criterium and is therefore applied to most of the physics subtriggers as a veto condition.

5.1.2 The $z-$vertex trigger

The central and the first forward proportional chambers are used to estimate the event vertex position along the beam axis ($z-$axis). A particle originating from the beam passes four layers of chambers.

The first step of the vertex estimator, the so called rayfinder, needs therefore to combine four cathode pad signals which lie on a straight line into an object called ray. In the plane perpendicular to the beam a 16 fold segmentation (ϕ - sectors) is used, such that the rays of each segment are treated separately. A histogram with 16 bins along z is filled according to the $z-$coordinate of the origin of each ray. The rays which were formed by the correct combinations of pads all enter in the same bin and form a significant peak above the background entries, which originate from rays from wrong combinations of pads and are therefore randomly distributed. Events which have the vertex far outside of the nominal interaction region do not develop significant peaks, in this case the histogram contains only the background from accidental rays.

From this histogram various trigger elements are derived. First of all the zVTX-t0 trigger element is activated, if there is at least one entry in the histogram. This information is used as an indication, that there is at least some activity in the central region of H1 and also for bunch crossing identification. Then peak significance analysis is performed and the trigger elements are activated, if the histogram peak exceeds a given significance threshold. This histogram analysis is fully programmable, such that the meaning of the trigger elements can easily be changed.

The rayfinder is based on a custom designed gate array (1.5 μm CMOS technology). For the final histogram bulding and the peak analysis programmable logic cell arrays (XILINX) and a 22 bit lookup table realised with 4 Mbyte of fast static RAM are being used.

5.1.3 The forward ray trigger

The cathode pad signals of the forward and central proportional chambers are fed into a logic, which finds rays originating from the nominal interaction region and pointing in the forward direction. A ray here is a set of impacts on three or four chambers, compatible with a track coming from the interaction region. These rays are counted and a trigger element is activated if there is at least one road found. Furthermore certain topology conditions in 16 ϕ-sectors can be recognised, e.g. if the rays lay all in two back to back sectors a special trigger element is activated. This system is realised by a total of 320 RAMs, which are used as hierachically organised lookup tables.

5.1.4 Big rays

The rays found by the forward ray trigger and the z-vertex trigger are combined to 224 'regions of interest' called big rays having the same geometrical properties as the 'big towers' of the liquid argon calorimeter (see later) with which they are put into coincidence.

5.1.5 The central jet chamber trigger

This trigger finds tracks in the central jet chamber (CJC), which have a distance of closest approach of less than 2 cm from the nominal beam axis and therefore suppresses beamwall events as well as synchrotron radiation background.

In a first step the signals from the CJC are digitised by a threshold comparator and synchronised to the HERA clock of 10.4 MHz. This way the drifttime information is kept with an accuracy of 96 ns or about 5 mm of position resolution in general.

In a second step the hits are serially clocked into shiftregisters. Track masks are defined according to their position in driftspace and their curvature in the magnetic field. A total of 10'000 different such masks are now applied to the parallel outputs of the shiftregisters to mark the active roads. Tracks with low or high transverse momentum can be distinguished as well as the sign of the low momentum tracks. The number of roads found in each of the 15 ϕ-segments and in the two momentum bins for each sign are counted separately as 3 bit numbers.

In the final step these track counts are further processed to generate trigger elements. Two different thresholds on the total number of tracks can be used simultanously. In addition a topological analysis in the x-y plane is performed. For instance a track activity opposite in ϕ can be recognised. Most of the digital logic is programmed into about 1200 programmable logic cell arrays (XILINX).

5.1.6 The $z-$chamber trigger

The $z-$chamber trigger uses the signals of the driftchambers in a similar way as the CJC trigger, utilizing the high spatial resolution obtained from the drift chambers. Signals are stored in shift register pipelines. Their parallel outputs are fed into coincidence circuits used as lookup tables for all possible tracks coming either out of the interaction region (vertex tracks) or from the proton beam region (background tracks).

The vertex tracks are entered into a 96 bin vertex histogram, a resolution of 5 mm for the vertex reconstruction is achieved. The background tracks are summed up per drift cell and form the "background histogram". This histogram is analysed by a neural net chip. The shiftregisters and the lookup tables are realized by 1060 logic cells arrays (XILINX). This trigger is still under development.

5.2 Calorimetric triggers

The calorimeter triggers have to cope with a wide spectrum of trigger observables, from narrow, localized energy depositions (e.g. electrons) to global energy sums such as transverse or missing transverse energy.

5.2.1 The liquid argon calorimeter trigger

The liquid argon trigger system is designed to calculate the energy deposited in various topological parts of the calorimeter as well as the total energy and other global energy sums which can be weighted by position-dependent weighting factors.

The realisation of this system contains an analog and a digital part. In the analog part the signals from the calorimetric stacks are split from the readout chain at the preamplifier input already and are separately amplified, shaped to a pulse width of about 600 ns FWHM and added to Trigger Towers (TT). The TT's are approximately pointing to the vertex and are segmented in 23 bins in θ and in ≤ 32 bins in ϕ. While the electromagnetic and hadronic signals are still separated in the TT's, the sum of the two is fed into an analog discriminator turning both signals off if the level is below an adjustable threshold, determined by the electronic noise. The same signal is used to determine the exact time (called t_0) of the signal. The total number of all active t_0 signals is available as a trigger element.

Depending on the θ region, either one, two or four TT's are summed up to 240 big towers (BT), providing finer granularity in the forward direction. The electromagnetic and hadronic signals of each BT are then digitised separately by ADCs running at the speed of the HERA clock. The digital outputs are calibrated by a RAM lookup table and two threshold discriminators are used to look for a potential electron signature in each BT, which is defined by high electromagnetic and low hadronic energy in the respective sections of the tower. Another discriminator lookup table marks all BT's to be transferred to the higher trigger levels. The electromagnetic and hadronic parts of each BT are summed up and the total BT energies are then available for further processing. A threshold is set on the total BT signal put into coincidence with the Big Rays derived from the MWPC triggers, and the number of these towers is counted, discriminated and provided as a trigger element to the central trigger logic.

The total BT energy is next fed into a set of lookup tables producing the weighted energy of this big tower for the various global sums (missing transverse energy, forward energy etc.) For the summing of the weighted BT energies, custom specific gate arrays are used, all summing being done in 8 bit accuracy.

In the last step further RAM-based lookup tables are used to encode the various global and topological sums into two-bit threshold functions provided as trigger elements to the central trigger logic.

5.2.2 The BEMC single electron trigger

The purpose of the BEMC single electron trigger is to identify scattered electrons from DIS processes. The basic concept of this trigger is to provide cluster recognition and to place energy thresholds on the sum of all energy clusters in the BEMC.

Analog signals are first added to form stack sums representing a high granularity trigger element. A cluster identification module then detects the cluster seeds and assigns neighbouring stacks to define clusters. Two trigger elements reflect the cluster multiplicities above certain thresholds. The energy of all clusters is then summed up and thresholds can be placed on this sum activating the respective trigger elements. Finally the cluster energy and the total energy sum are digitised into eight-bit numbers to be used for correlations with other quantities at the central trigger logic.

5.3 Muon triggers

Inclusion of efficient muon triggers covering a large solid angle, substantially increases the physics potential of H1. Muon triggering and measurement is for many processes complementary to electron triggering and allows a comparison between channels involving intrinsically the same physics, but with different systematic effects. A second important asset is the possibility of cross-calibration of the other parts of the H1 detector. This can be done by cosmic or beam-halo muons, and muons from the physics channels. Both the instrumented iron system and the forward muon spectrometer deliver level 1 trigger information, as described below.

5.3.1 The instrumented iron muon trigger

The instrumented iron system is logically divided into 4 subdetectors. Each subdetector consists of 16 modules, 5 of the 12 chambers of each module have their wire signals made available to the level 1 trigger. The "OR" of 16 wires of these signals is called a profile and all profiles of one chamber are again ORed together to form a single plane signal. Any condition on the 5 plane signals of one module can be requested by means of RAM lookup tables (e.g. a 3 out of 5 condition of the chamber planes) for each module independently. An additional signal from the first plane detects, when there is more than one hit in the plane indicating the hits rather originating from a cluster tail of the calorimeter than a single muon.

The (maximum eight different) outputs of each of the 64 modules are then fed into a central muon trigger logic which is organised in RAM lookup tables again. So far only a counting of the number of muons found in each subdetector has been loaded into the RAMs and two trigger elements per subdetector were used: exactly one muon candidate and more than one muon candidate.

5.3.2 The forward muon trigger

The trigger deals with each octant of the forward muon chambers separately. The track candidates found in each octant are allocated to eight regions at different polar angles to the beam. The 8-bit hit patterns from all eight octants are fed into a RAM based lookup table which counts the number of muon candidates and allows programmable topological correlations to be made . Eight bits of trigger information are then sent to the central trigger as trigger elements.

5.3.3 Triggers derived from the luminosity system

The luminosity system runs with an independent data acquisition and triggering system. However the trigger signals, derived from the detectors of this system, are available also to the main trigger system: Independent thresholds can be set on the electron energy, the photon energy and the calibrated sum of the two. Together with the signals of the veto counter in front of the photon detector this information is fed into a lookup table to form logical combinations. So far mainly the electron signal was used to tag photoproduction events.

5.4 Central trigger L1 decision

The information generated by the subdetector triggersystems described above is represented in a total of 128 bits, which are connected to the central trigger logic. Here all trigger elements are fed into a pipeline (realised as double ported RAM based circular buffers), which allows to adjust the delays of all incoming signals to the proper bunch crossing. Furthermore the information stored in these trigger element pipelines is recorded at readout time, allowing to study the time evolution some bunchcrossings before and after the actual event took place.

The trigger elements are logically combined to generate a level 1 trigger signal. Up to 128 different subtriggers are formed by applying coincidence and threshold requirements. (Lookup tables are used to form 16 fold coincidences of arbitrary logic expressions from up to 11 predefined input bits). A trigger description language (TDL) has been developed to keep up with the ever changing demands for new subtriggers and to properly log the logic and the status of the triggers loaded. The subtriggers are assigned to a given physics event class (physics trigger), to experimental data needed e.g. for measuring the efficiency of a given detector (monitor trigger) or to cosmic ray events for calibration purposes (cosmics trigger).

The rate of each subtrigger is counted separately and can be prescaled if needed. The final L1keep signal is then given by the logical OR of all subtriggers after prescaling and is distributed to the front-end electronics of all subsystems to stop the pipelines. At this point the primary deadtime begins. Of course all this logic works in a pipelined way as all the subdetector triggersystems described above, clocked by the HERA clock, and delivering a trigger decision every 96 ns.

6 Intermediate trigger levels (L2-L3)

The two intermediate trigger levels L2 and L3 operate during primary deadtime of the readout and are therefore called synchronous. The calculations which are performed in these systems and the decision criterias applied depend on the subtrigger derived in the L1 system, which acts in this way as a rough event classification.

As we have seen in the previous chapter, deadtime starts after the L1 trigger system has given a positive decision. The L2 trigger system now can evaluate complex decisions based on more detailed information. After a fixed time of typically 20 μs the decision of the L2 trigger defines whether a fast reject should happen or whether the event is to be treated further. For the L2 decision processsors various hardware solutions are under construction including a complex topological correlator and a neural network approach to exploit the correlations between the trigger quantities from the various subsystems in a multidimensional space. The massively parallel decision algorithm of these systems makes them ideally suited for fast trigger applications.

Only if the event is accepted by the L2, the bigger readout operations like zero-suppressing of the drift chamber digital signals and the calorimeter analog to digital conversion and DSP processing are started. During this time the trigger L3 system based on an AM 29000 RISC processor performs further calculations. The level 3 decision is available after typically some hundred μs (max. 800 μs), in case of a reject the readout operations are aborted and the experiment is alive again after a few μsec.

The calculations of both L2 and L3 triggers are based on the same information prepared by the trigger L1 systems described in the previous section. Topological and other complex correlations between these values are the main applications for these intermediate trigger systems.

6.1 The L2 topological classifier

The selection is based on the topological analysis of the events. L1 variables, tracks and calorimeter energy depositions are projected onto 16 $\times$ 16 boolean matrices defined in the Θ and Φ coordinates. From these matrices variables are derived which characterize the global event shape. The topological analysis is performed on the following trigger elements: big towers, MWPC big rays, R-Φ trigger (track candidates with a precise Φ determination from the drift chambers) R-Z trigger (track candidates with a precise Z determination), μ counters. The projections are grouped into the following projection families:

1. C_n (n=0 to 7): big towers with an electromagnetic energy $E_e \geq 2^n$ or a hadronic energy $E_h \geq 2^n$, the energy is given in FADC counts.

2. Cv_n (n=0 to 7) big towers with the same energy limits as before, validated by big rays (a big tower is retained only if a big ray points to it).

3. hadronic big rays where the information given by the electromagnetic part of the calorimeter is ignored, or electromagnetic big rays vetoed by their associated hadronic big towers with smaller energies (these may or may not be validated by tracker informations).

4. projections using only tracker or μ informations.

For a projection family using calorimeter information one energy index is determined: the maximal energy threshold giving a non-empty projection. If this projection is too simple (only one small cluster of big towers), a secondary energy index is determined: the maximal threshold giving a more complicated projection.

Any useful projection can be described by the following parameters: the number of small clusters (a small cluster is a set of adjacent cells of the 16*16 boolean matrix which can be included in a 2*2 square), the presence of big clusters, the presence of neighbour small clusters: i.e. small clusters with a narrow free space between them (only 1 free line). We determine the Θ-pattern and the Φ-pattern as well, which are projections of the boolean matrix on the Θ and Φ axis. These values are combined to define four topology indices for each projection. To each projection an energy index is associated, for pure tracker families it is taken from an associated calorimeter family or from a L1 global energy sum (total or transverse energy). One can analyse up to 120 different projections per event, each determination takes only 100 ns.

The topological cuts are performed on the energy index - topology index 2 dimensional plots which are derived from the above Boolean matrices. Before data taking these plots are filled with pure background data from pilot bunches (proton or electron bunches with no colliding partner) populating well defined areas on them. During data taking the distance of the analysed event to the background border is calculated for each topology index. An event is accepted if its distance to background is large enough.

The hardware has the volume of a Triple-Europe crate. It contains 15 specialized cards (Fig. 3):

- Two receiver cards, which are not shown in the figure, receive the L1 information, store it and dispatch it on 8 buses on the crate in less than 5 μs.

- Ten acquisition cards (ACQ) store the informations, the regions corresponding to each card are shown in Fig. 3.

- One topology card. After the acquisition the Command-Decision card sends the different projection modes to the ACQ cards which send parts of the boolean 16*16 matrix to the topology card. The topology card computes the topology indices, and send them back to the Command-Decision card.

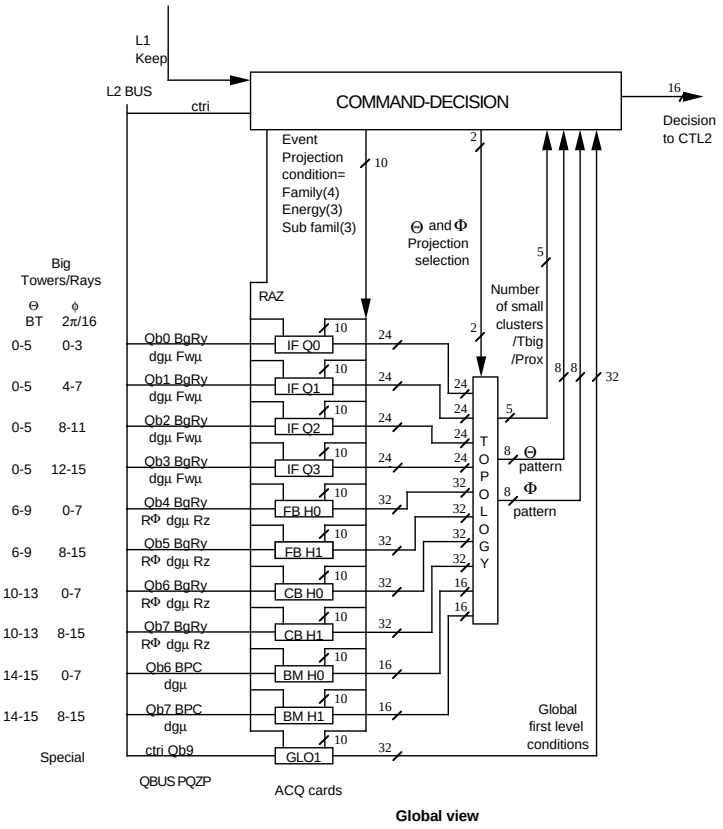

Figure 3: Global view of the topological trigger

- One Command-Decision card, which contains a microprocessor (a DSP 21010 from Analog Devices with a cycle time of 30 ns). It controlls the other cards and computes the distance to the background border for each topology index. These distances are added to compute the total distance to the background border. Actually, there are 16 independant sums (called "machines"). At the end of the treatment, the DSP compares these 16 sums to 16 predefined thresholds, and sends its decision as a 16 bit word to the level 2 Central Trigger Logic. An event should be kept if one of these 16 bits is true. The 16 machines may allow the shift crew to downscale one of the machines, if a change of beam, or of the detector state, gives a too large rate of accepted events for this machine.

- An 11th ACQ card has a special function: it determines indices coming from some first level informations.

There are 3 phases for the treatment of an event after the acquisition phase. During the first phase, the DSP, together with the ACQ and topology cards, determines the energy indices of the calorimeter projection families. This phase takes from 4 to 6 μs. After that the DSP computes, together with the ACQ and topology cards, the distances to the background border for the different topology indices. It adds them to the total distances for the different machines. This phase takes also from 4 to 6 μs. During the third phase the DSP compares the total distances for the 16 machines to the minimum distances required for an event acceptation. This phase takes 1.6 μs.

The 11 ACQ cards are identical: they differ only through the programming of their ROMs and PALs and are realized in printed circuits. The topology and Command-Decision cards are unique and are realized in multiwire technology.

6.2 The L2 neural network trigger

In the L2 neural network trigger the L1 trigger information is viewed as a pattern or feature vector which is analysed by a standard feed-forward neural network. Offline, the network has to learn to separate the patterns of physics reactions from those of the background by the usual training methods like backpropagation. In this way a complete pattern recognition is performed in the space of the L1 trigger quantities and the correlations among the various trigger quantities are exploited. Detailed investigations within the H1 experiment using Monte-Carlo data have shown that feed-forward networks trained on the first level trigger information are indeed able to obtain the necessary reduction in background rate while keeping high efficiency for the physics reactions.
Since the network's decision is taken at a hardware level, one is forced to very high speed computations and neural networks with their inherent massive parallelism are ideally suited for this task. Recently fast digital neural VLSI hardware has become available and the realisation of networks using the digital trigger information from the H1 experiment can now be attempted.

6.2.1 Neural networks

The basic processing element in a neural network is a neuron or node. The node j receives signals I_{jk} from a number n of input channels. The net input to that node is the weighted sum of these signals

$$N_j = \Theta_j + \sum_k W_{jk} I_{jk} \tag{1}$$

where the thresholds Θ_j (also called bias) and the connection strengths W_{jk} are node-associated parameters. The final output of the node is assumed to be a simple function of N_j, called the transfer or threshold function $f(N_j)$. A standard form for this function is a sigmoid.

$$O_j = f(N_j) = \frac{1}{(1 + e^{-N_j})} \tag{2}$$

With a given threshold on the output a single node performs a simple linear discrimination and defines a separation plane in the input space.

Given the elementary nodes a full neural network is defined by specifying the total number of nodes and the linking among them.

For event classification normally a restricted topology is used (Fig. 4), the so called feed-forward networks. All nodes are arranged into distinct layers. The bottom (input) layer has one node for each component of the input vector. The top (output) layer has a single node. The classification is given by the output of this node (e.g. 1 for physics and 0 for background). The connectivity is feed-forward and complete: A node in a given layer receives input only from nodes in the next lower layer, each node in a given layer sends its output to all nodes in the next higher layer.

In a three-layer network (Fig. 4), containing one hidden layer, each hidden node corresponds to a plane in the multidimensional input space and the output node builds up a volume - not necessarily closed - from them. In this sense we can say that a neural network is the generalization of the linear discriminant classifier.

6.2.2 Trained networks

The neural network classification functions defined above have a number of free parameters, the thresholds Θ_j and connection weights W_{jk} for each non-input node. To determine these parameter values a simple prescription - the so called backpropagation - exists, given training sets from Monte Carlo for which the desired output of the network is known. (For the background training sets real data can be used as well.)

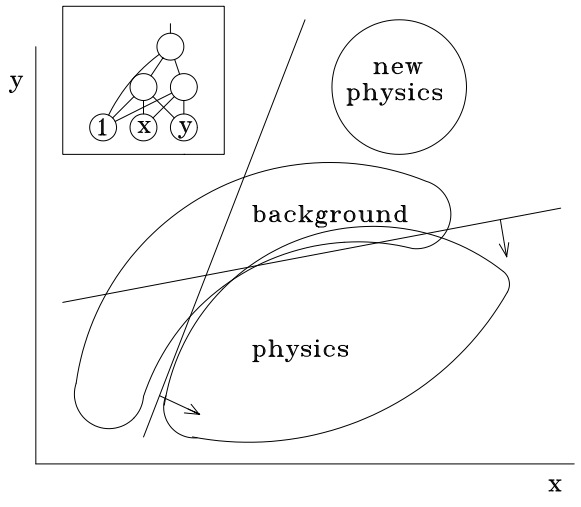

Figure 4: Classification with one hidden layer.

Back propagation minimizes the global classification error. Most frequently a least mean square error measure is used

$$E_{global} = \sum_{events} (O_{out}^{obtained} - O_{out}^{desired})^2 = \sum_{events} E_p$$

where the sum is over all events in the combined training sets and for each event the contribution to the global error is simply the squared difference between actual ($O_{out}^{obtained}$) and target ($O_{out}^{desired}$) network output for that event.
The corrections to the network weights W_{jk} associated with a particular event are done by steepest descent steps :

$$\Delta_p W_{jk} = -\eta \frac{\partial E_p}{\partial W_{jk}}$$

where the parameter η is called the learning rate. The process is repeated until the difference between all output nodes and all patterns is within some tolerance.

6.2.3 Background Encapsulation

A major drawback using backpropagation is that the various networks will only efficiently tag those physics reactions which have been offered for training to the networks. In this spirit, "new physics" (Fig. 4) may be discarded or only inefficiently selected. This difficulty can be overcome by arranging the separation planes in a way as to completely *encapsulate* the background. This procedure does not rely on any physics input, it only rests on the topological properties of the background. A straightforward algorithm to enclose the volume of background has been proposed by us and has been intensively investigated. Depending on the point density of the background events in trigger space, several encapsulating volumes could be necessary (e. g. when the background space is fragmented). Although this encapsulation might now seem to be the only necessary type of net to use, one should consider that physics "close" to the background will not efficiently be triggered with such a scheme. Since specific physics channels have not been considered in the encapsulation algorithm, the corresponding boundaries are not optimized for physics. The background encapsulation is a safety channel to retain events from unexpected physics processes.

6.2.4 Committee of networks

It is not expected that the various physics reactions require identical trigger quantitites to be distinguished from the background. Therefore it seems reasonable to study the different physics reactions and develop nets to select these reactions by using the relevant trigger informations. This results in networks which are smaller and easier to handle. Our investigations have shown that *small* nets trained for *specific* physics reactions, working all in parallel, are more efficient in comparison to a single larger net trained on all physics reactions simultaneously. Most importantly, putting these nets to a real trigger application, the degree of modularity is very helpful when a new trigger for a new kind of physics reaction is to be implemented.
The final algorithm for arriving at a trigger decision using the outlined neural network architecture is the following.

1. build an OR from all physics selectors: OR(phys)
2. build an OR from all background rejectors: OR (back)
3. reject the event, unless
   - OR(phys) is true, or
   - OR(phys) is false, but OR(back) is also false

This latter condition is a signal for potential new physics.

6.2.5 Hardware Realisation

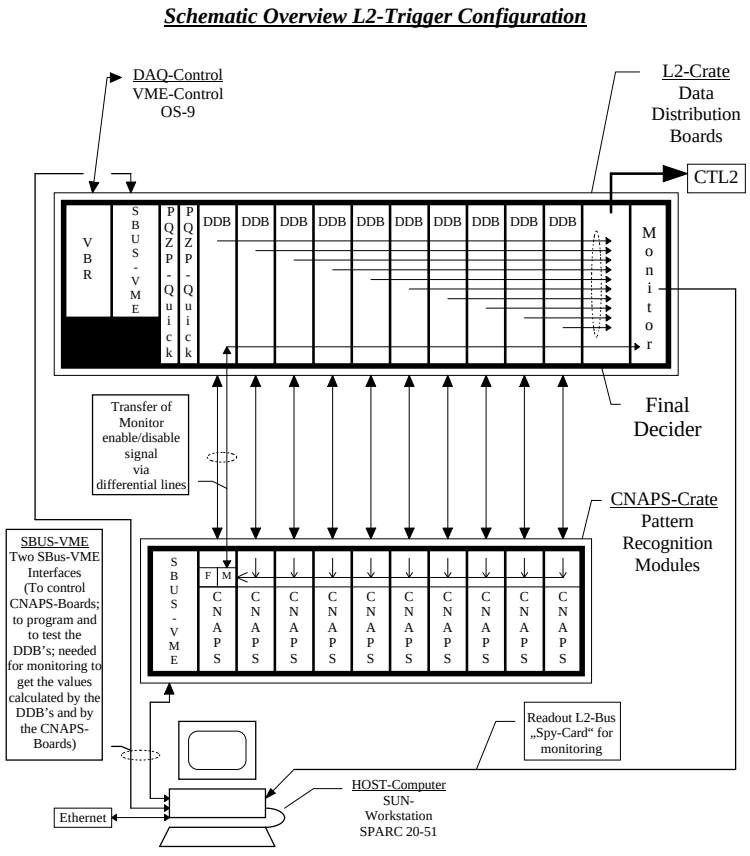

Figure 5: Overview of the L2 neural network trigger hardware

A digital VLSI neural network chip, the CNAPS chip from Adaptive Solutions (USA) has become available recently, which is fast enough to meet the 20 μs time constraint

for L2. Adaptive Solutions Inc. has in its program a VME board built around the chip, together with the necessary software. The hardware realisation of the neural network level 2 trigger for H1 is depicted in Fig. 5: The CNAPS board - can model a neural net with 64 inputs, 64 hidden nodes, and 1 output node. A VME crate houses one board for each physics reaction or for a background region. Each CNAPS board has a companion data distribution board (DDB) which supplies exactly the foreseen trigger information for the neural net modelled on the board. Both the DDB and the CNAPS boards are loaded from a host computer via micro-processor controller boards.

A 20 MHz L2 data bus, divided in 8 parallel groups, each 16 bits wide, provides the trigger information from the various subdetectors to the levels 2 and 3. During the time of data transfer the DDB selects precisely the information from the bus which is needed for their companion CNAPS boards.The DDB is able to perform more complex operations as well (8 bit sums, selective summing, bit counting) on the L2 data and creates more significant 8-bit input values from bit patterns. For instance a jet finder algorithm will be implemented which will provide the 3 most energetic jets using the big tower granularity.

After 5 μs the data are loaded onto the CNAPS boards and the data boxes give start signals to the CNAPS boards which present their outputs back to the data boxes after another 10 μs . These in turn supply their signal to a Final Decider. Rate limiting measures, such as prescaling or bypassing of trigger decisions, can be taken in the Final Decider.

6.2.6 The CNAPS/VME Pattern Recognition Board

CNAPS (Connected Network of Adaptive ProcessorS) is a hardware concept based on SIMD technology (Fig. 6). A matrix of max. 256 processing nodes (PN) is driven by a

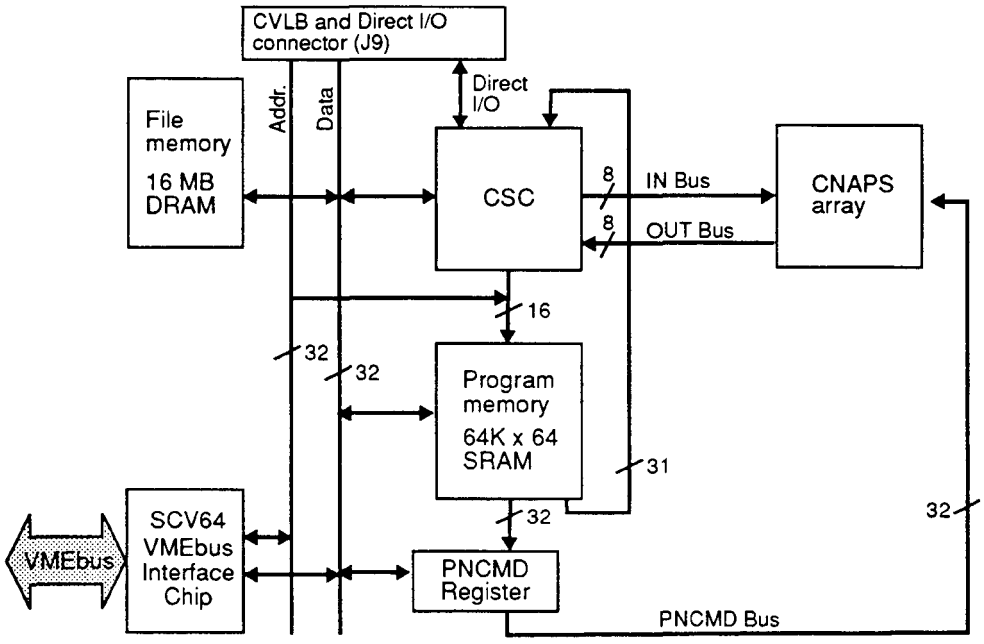

Figure 6: The CNAPS/VME block diagram

sequencer (CSC) through a common bus (PNCMD). The board has local file memory and a VME interface. Direct IO channels (8 bits wide) are used to input the data from the DDBs. The architecture of the CNAPS processor nodes is shown on Fig. 7. They have a local memory of 4 kBytes and a complete fixed-point arithmetic unit. The elements of the input vectors are broadcasted on the IN bus which allows parallel processing by more PNs. Each PN implements one neuron. The results are accumulated locally and may be

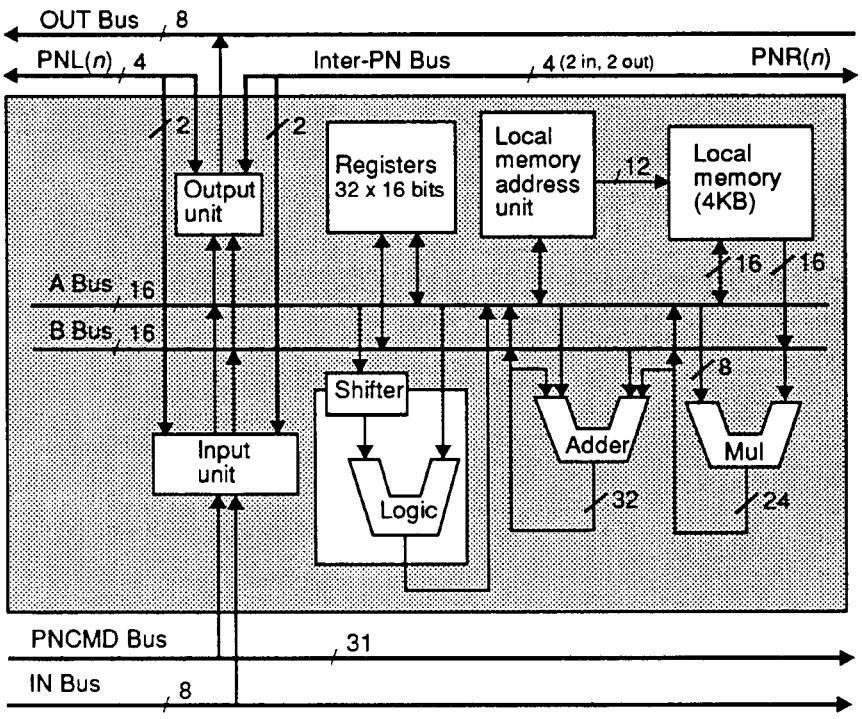

Figure 7: CNAPS/VME processor node architecture

read out serially on the OUT bus. A feed-back of the results onto the IN bus through the CSC allows the evaluation of multilayer nets. In one clock cycle one multiply-accumulate operation can be performed. With this architecture the computation time is proportional to the number of neurons instead of being proportional to the number of weights. With 20.8 MHz clock speed our maximum net (64 x 64 x 1) will need 8.9 μs to be evaluated in one CNAPS board which corresponds to the computing power of $\sim$ 100 IBM RISC workstations.

7 The L4 filter farm

The level 4 filter farm is an asynchronous software trigger based on fast mips R3000 processor boards. It is integrated into the central data acquisition system (Fig. 8) and has the raw data of the full event available as a basis for its decision making algorithms. This allows for online trigger selections with the full intrinsic detector resolution. The processor boards run in parallel. Each board (33 in 1994) processes one event completely until a decision is reached. The typical maximum input rate is 50 Hz. Since this system works asynchronous to the primary trigger system, there is no further deadtime involved as long as the L3 accept rate stays safely below 50 Hz.

In order to reach a decision in the shortest possible time, the L4 algorithm is split into various logical modules, which are run only if a quantity calculated by the respective module is needed to reach this decision. The L4 modules use either fast algorithms designed specifically for the filter farm, or contain parts of the standard offline reconstruction program. The average execution time is 500 msec and a rejection factor of $\sim$ 6 can be obtained. The execution of the modules is controlled by a steering bank containing text in a steering language written explicitly for this purpose. The final decision is based on statements containing logical combinations of numerical or logical values. Execution of the statement is terminated and the next statement is executed as soon as a subcondition is false. It is possible to run any statement in test mode without influence on the actual

Figure 8: Overview of the data acquisition

decision. This allows the evaluation of the effect of new statements with high statistics prior to activation and the flagging of particularly interesting events, e.g. for the online event display. This scheme allows for high flexibility without changes in the program code and facilitates book keeping as the steering bank is automatically stored in the H1 database.

The filter farm is not only used for event filtering purposes, but it is also well suited for monitoring and calibration. The reconstruction modules fill numerous monitor histograms which can be inspected online. Warning messages can also be sent to the central control console, informing the shift crew immediately of potential problems. Calibration data are sent to the data base for immediate use by the online reconstruction process.

8 Conclusions

In 1994 we could run H1 with only L1 and L4 without cutting significantly into physics acceptance and with acceptable performance concerning rates and deadtimes. However at design luminosity (15-20 times higher than the actual one) we will have to tighten the requirements on the events in our L2 and L3 trigger systems which are under development. But it looks still possible to trigger with high acceptance for physics events, perhaps with

the exception of the heavy quark production of which a large fraction of events have not enough transverse energy for the calorimetric triggers and no or only low energy electrons or muons in the final state. Therefore they have to be triggered in the first level by central tracking information alone resulting in a high beam gas background from the nominal ep interaction region. We will have to use topological criterias in the level 4 or even in level 2 or 3 to recognize these events.

9 Acknowledgements

The trigger system of the H1 experiment has been put together by a large number of people working with a lot of dedication. Their support and achievements are gratefully acknowledged here. My special thanks go to my colleagues in the neural network trigger group. We had a nice time and a lot of fun at working out the ideas and the implementation of this trigger.

References

[1] The H1 detector at HERA, DESY 93-103

[2] Proceedings of the Third International Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics
October 4-8,1993, Oberammergau, Germany

[3] A topological level 2 trigger, H1-06/91-181