A General Scheme for Optimization of Trigger Rates in an Experiment with Limited Bandwidth

H.-C. Schultz-Coulon[1], J. Coughlan[2], E. Elsen[1], T. Nicholls[1] and H. Rick[3]
[1]Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[2]Particle Physics Division, CLRC Rutherford Appleton Laboratory, Oxfordshire,UK
[3]CERN, Geneva, Switzerland

Abstract

Modern, general purpose high energy physics experiments are designed to record the events of physics processes which greatly differ in rate. Both rare signatures with trigger rates of order 1 Hz and minimum bias interactions with rates sometimes approaching 1 MHz have to be triggered at the same time. The typical means of accommodating these requirements within a limited bandwidth is to downscale individual trigger channels appropriately. This paper describes a novel procedure for finding an optimized set of downscale factors that automatically adapts to the beam conditions. A first application at the H1 experiment at HERA showed that the number of events recorded for specific channels could be significantly increased.

I. INTRODUCTION

Most modern colliders for particle physics experiments produce a large total interaction rate which exceeds the readout and data storage capabilities of the detectors involved, necessitating the use of a sophisticated trigger system. Such a system must keep rare or hitherto unobserved processes with maximum possible efficiency, while recording only a fraction of those processes with high rates. A stable trigger rate and good background rejection often require complicated trigger conditions with moderate efficiencies, which are mostly biased towards certain topologies or kinematic regions and therefore difficult to control with sufficient precision. On the other hand, trigger conditions with high efficiencies and small bias are susceptible to varying backgrounds, resulting in large and often unstable output rates.

A scheme for optimizing trigger rates is presented which is implemented in the H1 trigger system and allows the use of unbiased, variable rate trigger conditions without compromising overall data taking efficiency. The procedure is described and first results on its performance are reported.

II. THE H1 DETECTOR AND TRIGGER SYSTEM

At the particle collider HERA at DESY, which collides electron and proton beams at energies of 27 and 820 GeV[1] respectively, two colliding beam experiments — H1 and ZEUS — have been running successfully since 1992, allowing the study of a wide range of physics topics in a previously unexplored regime.

The H1 detector, described in detail elsewhere [1], consists of a number of complex, highly-segmented subdetectors designed to provide both complementary and redundant measurements of various aspects of high energy electron-proton collisions. The large number of channels in the H1 detector ($\sim$270,000) leads to a raw event data size of approximately 3 Mbytes. After compression of data at the front end, 50 – 100 Kbytes of data per event are logged to tape.

Data of interest are selected by the H1 trigger system, which is divided into four levels of which three — the first (L1) and second (L2) level trigger and the asynchronous online filtering (L4) — have been routinely operated during the past years. The L1 system is phase-locked to the HERA accelerator clock signal of 10.4 MHz and provides a trigger decision for each bunch crossing after 2.3 μs without causing deadtime. The subdetector systems feed data into front-end pipelines and generate fast information concerning the general properties of the event. This is encoded in Boolean decisions known as trigger elements which are sent to the L1 logic of the central trigger [2]. A decision is made on the basis of 128 logical combinations of these trigger elements, called subtriggers. After a positive L1 decision all front-end pipelines are frozen and deadtime begins. The L1 is then validated by the second level trigger [3] based on the output of two independent systems [4, 5] which operate in more detail on the information from the subdetectors during the primary deadtime. An L2 decision requires 20 μs leading to typical input rates of 1 kHz. A further validation by the L3 system, based on software algorithms running on a single RISC processor, has been tentatively used in previous years but is hampered by the restriction on the input rate due to the rather long decision time. It is therefore not currently used for physics selection. Subdetector data are read out asynchronously by the central data acquisition [6] over an optical fibre ring and the full event is then assembled. The use of sufficient buffering prevents this stage from contributing to the deadtime, providing that the total bandwidth is not exceeded. A further level (L4) of triggering operates asynchronously to reduce the data logging rate and volume. This takes the form of a farm of RISC processors [7], each running a reduced version of the offline reconstruction code and operating on a full event in approximately 100 ms. The output of the L4 farm is logged to tape at a typical rate of 5 – 10 events per second.

III. ADJUSTMENT OF TRIGGER RATES

In order to allow the operation of different subtrigger conditions in parallel each of the 128 individual subtriggers generated at L1 can be downscaled by a pre-settable factor. In case of downscaling of a subtrigger i by a factor d_i, only one

[1]Since summer 1998 the proton energy is 920 GeV.

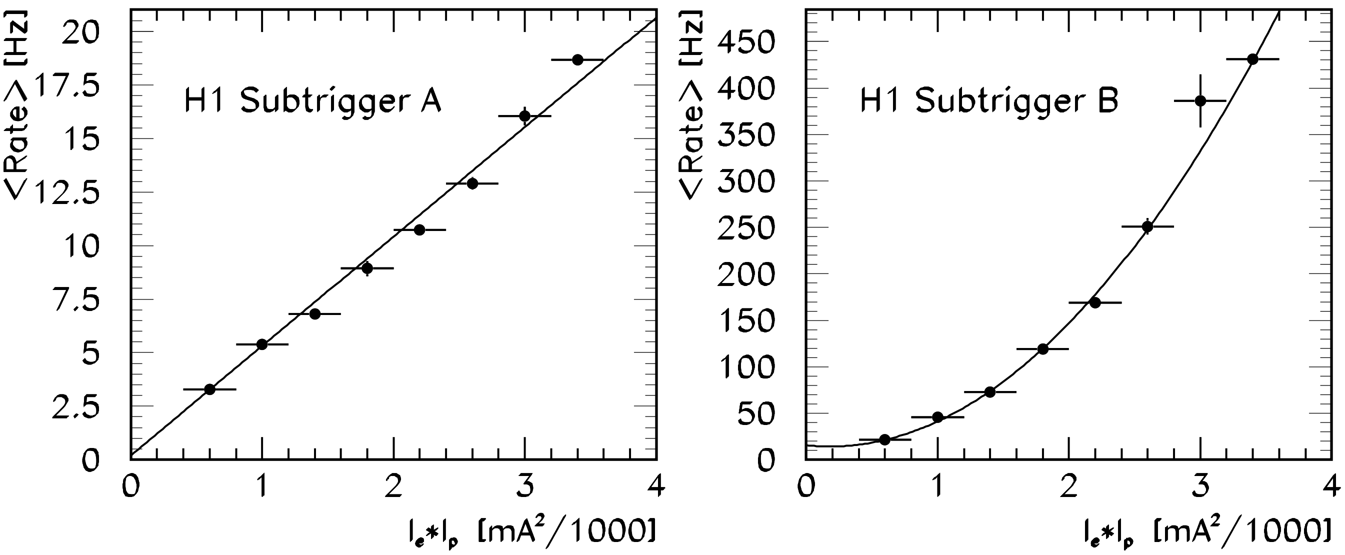

Figure 1: Examples from H1 of two subtrigger conditions with different background dependence on luminosity $\mathcal{L} \sim I_e \cdot I_p$, where I_e and I_p are the electron and proton beam currents respectively.

out of d_i events satisfying its triggering condition leads to a positive L1 decision. This downscaling allows the storage of a realistic amount of high rate physics processes for analysis as well as monitoring and calibration purposes, without exceeding overall bandwidth limitations given by the front-end readout time, the L4 computing power and the maximum data logging rate. During the first years of operating H1 these downscale factors were defined in advance. However, since collider operating conditions often differ from those assumed, this choice was not always optimal. Ad hoc manual corrections to the downscale factors often led to unsatisfactory results. Therefore an automatic optimization procedure was developed which is described below.

A. *Nature of the Problem*

At an electron-proton collider the lifetime of the electron beam defines the duration of a luminosity fill; at HERA a typical fill lasts for more than 10 hours during which time the luminosity decreases by a factor of 3 or more. A bandwidth-limited data acquisition system therefore requires a frequent adjustment of trigger rates in order to accommodate the maximum possible yield for those processes that otherwise would remain downscaled. At H1 several sets of downscale factors were predefined in the past and applied in turn as luminosity decreased. The selection of appropriate sets of downscale factors is a non-trivial task since, in addition to the time evolution of luminosity, the influence of background conditions, the relative importance of different subtriggers and possible overlap between trigger channels have to be taken into account. The transition between predefined downscale sets is therefore arbitrary and might not be optimal for all trigger conditions involved.

The impact of beam conditions on the behaviour of subtriggers is illustrated in Figure 1 which shows two subtriggers used by H1 for the selection of DIS and photoproduction events. They exhibit markedly different dependences on luminosity; the rate of subtrigger A rises linearly with luminosity $\mathcal{L}$, while that of B varies with $\mathcal{L}^2$. The latter behaviour is typical of a trigger which suffers from pile-up or the effects of accidental coincidences. Such a trigger therefore does not produce the best physics yield at high luminosity and, in an experiment with limited bandwidth where triggers are operated concurrently, should preferably be run towards the end of a fill.

An automatic online determination of optimal downscale factors can therefore clearly improve the utilization of the available bandwidth. From the above considerations, such a procedure requires the following:

- Knowledge of the dependence of individual trigger rates on beam conditions.
- Estimation of the beam quality from measured trigger rates.
- A strategy defining the relative priorities of subtriggers based on physics goals and the desired dependence on beam conditions.

Such a procedure was developed and implemented by the H1 experiment.

B. *Optimization Procedure*

The priorities between trigger channels are defined using a set of explicit rules. According to these rules the downscale factors are calculated online by an iterative analysis which assigns weights to a sample of triggered events. The weights are determined using the algorithm described in section V in the context of cross-section calculation. This procedure leads to an optimal output rate and allows a fast reaction to changing beam and background conditions.

The rules are defined in a strategy whose syntax allows subtriggers to be grouped in a nested structure. For every group the maximum output rate can be either fixed or expressed in terms of relative fractions of the available total output rate. In order to allow for time-dependent downscaling which takes beam and background conditions into account, the relative weighting of subtriggers can be defined as a function of the measured trigger rates; this is discussed in section IV. Finally the assignment of fixed downscale factors is also possible in order to guarantee full luminosity for subtriggers dedicated to rare processes.

The flexibility of the scheme is illustrated in the following simple example:

```
Total rate: 40 Hz
{
  High Q2 Process     : df 1   { s1 }
  DIS Process         :  0.5   { s2 }
  Photoproduction     :  0.5
  {
     stable   :  0.9           { s3 }
     unbiased :  0.1           { s4 }
  }
  Monitoring          : 6 Hz   { s5 }
}
```

Four different groups of processes are considered — named *High Q^2 Process*, *DIS Process*, *Photoproduction* and *Monitoring* according to the performance example given in the next section — to be selected by the subtriggers `s1` to `s5`. It is assumed that `s1` produces a stable and sufficiently low output rate such that the corresponding downscale factor can be safely fixed to 1 by `df 1`. For monitoring purposes 6 Hz are reserved for `s5`. The remainder of the 40 Hz total output rate is equally divided among the *High Q^2* and the *DIS Process* groups. The weights within the *Photoproduction* group emphasize subtrigger `s3`, which takes 90% of the assigned rate. The 10% remaining for `s4` enable, for instance, unbiased cross checks of this analysis channel.

In case the allowed output rate for a group is not fully consumed, the remaining bandwidth is filled up by assigning additional rate to other subtriggers within the same group. The given strategy therefore allows the maximum proportion of events to be recorded by highly efficient, unbiased subtriggers without compromising the stability of data taking.

C. Performance Example

A strategy similar to that described above was first used by H1 at the end of the 1997 data taking period. The application of the new scheme during this time proved to be very successful and led to a clear gain in the number of recorded events for specific channels.

The aim of this special running period was the acquisition of special minimum bias data in order to improve earlier measurements of photoproduction processes [8, 9] and to measure the the longitudinal proton structure function F_L [10] in deep-inelastic scattering (DIS). Furthermore, the guaranteed acquisition of rare or potential new processes at high Q^2 remained paramount. These considerations led to the definition of four subtrigger classes as a basis for a strategy to optimize downscale factors:

- High Q^2 and other triggers dedicated to rare signatures.
- DIS triggers for measurement of F_L.
- Photoproduction triggers.
- Monitor triggers.

All subtriggers in the high-Q^2 group remained unprescaled during the full data taking period. The relative weighting of the DIS and photoproduction groups was defined such that the photoproduction triggers opened up only when the rate of the DIS group became saturated. The rate of the monitor triggers was fixed to a few Hz.

The strategy was used to evaluate all downscale factors once per hour. The result for a typical luminosity fill is depicted in Figure 2a. While the rates of the high-Q^2 and monitor classes stayed rather constant, the evolution of the other classes shows the expected behaviour with a rate increase for photoproduction triggers at the end of the fill. Figure 2b shows the same picture for the trivial scenario where all downscale factors were fixed at the beginning of data taking and never subsequently changed.

During the whole period where the automatic rate optimization was used about 170 nb^{-1} of data were taken for the high-Q^2 class. The event yield for the DIS processes was a factor of 5–8 higher than possible in the scenario with fixed downscale factors. Similar improvements were achieved for the photoproduction triggers, reaching factors of 20 for certain channels. Due to this experience it was decided to use this scheme in 1998 during normal data taking. A first testing period again demonstrated the ability of the new scheme to allow smooth data taking even in the face of rapidly changing background conditions.

IV. Optimizing Bandwidth Distribution

As discussed earlier, an important component of the optimization procedure is the ability to modulate the weights assigned to subtriggers as a function of the beam conditions. The benefit of this can be illustrated using a simple example where *subtrigger A* and *subtrigger B* are considered to be the only trigger conditions active. It is assumed that the background fraction for subtrigger A is constant, while for subtrigger B the triggered background rises quadratically with luminosity $\mathcal{L}$, behaviours already demonstrated with the example H1 triggers shown in figure 1.

Given a maximum allowed output rate of 40 Hz — approximately matching H1 conditions — and an electron beam lifetime of about 10 hours, two different scenarios are considered. In the first case, the maximum rate is equally split between subtriggers A and B, given by the following trivial strategy:

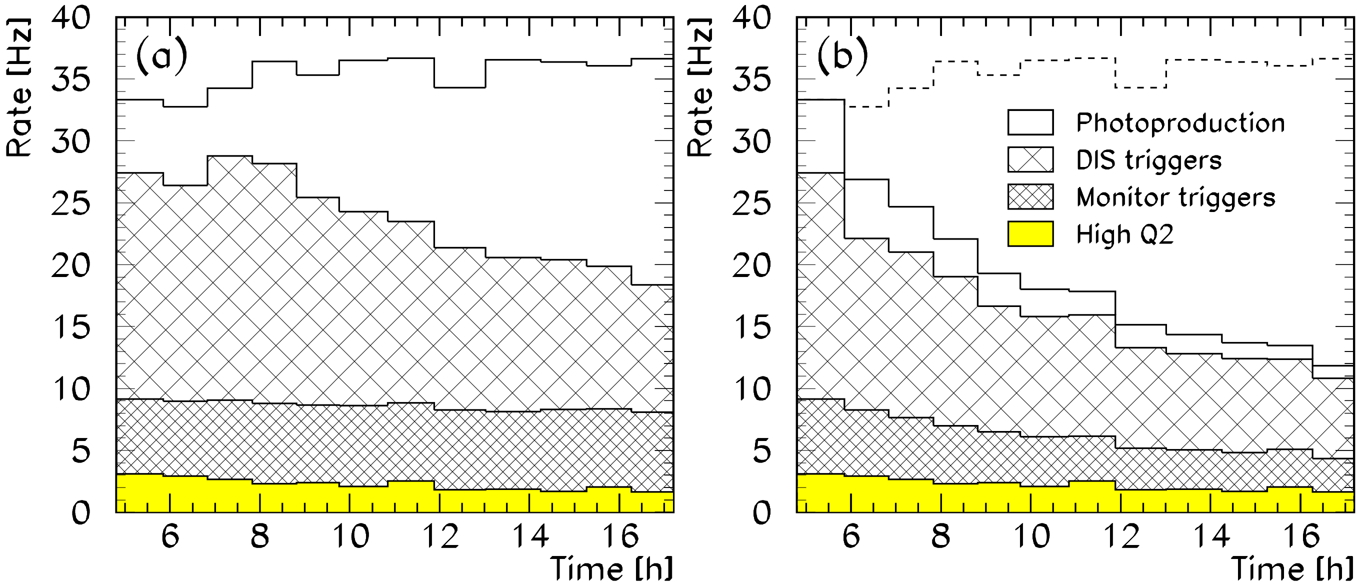

Figure 2: Output rate as a function of time for different event classes defined in the text for (a) a typical luminosity fill using the new optimization scheme and (b) without time dependent adaptation of downscaling. The total output rate from (a) is shown as dashed line for comparison.

```
Total rate: 40 Hz
{
  subtrigger A: 20 Hz { s1 }
  subtrigger B: 20 Hz { s2 }
}
```

In the other case an optimized strategy is used where the downscale factors are distributed according to a relative weight, determined as a function of the rate of subtrigger A:

```
Total rate: 40 Hz
{
  subtrigger A: 1.0 { s1 }
  subtrigger B: k/rate(s1)^2 { s2 }
}
```

The parameter k is chosen for the purpose of direct comparison such that the total luminosity taken for subtrigger A remains constant. The choice of functional form of the weight is a compromise between relative yield of the two subtriggers and ensuring that, in the event of shorter luminosity fills, some data is acquired for subtrigger B.

Figure 3 shows the time evolution of luminosity (3a), the corresponding background fractions (3b) and the integrated luminosity obtained for the two subtriggers in both scenarios (3c). The gain achieved in the optimized case is clearly visible and exhibits an increased yield for subtrigger B of more than 20%.

This demonstrates the additional benefit of a sophisticated scheme for optimizing output rates beyond a simple adjustment of downscale factors with time using fixed ratios between subtrigger rates. Nevertheless, the improvement depends greatly on the luminosity and background conditions and their evolution during a fill. These parameters have to be taken into account when choosing a relative weighting between subtrigger rates. The development of an optimal rate evolution and weighting strategy is therefore non-trivial and requires substantial operating experience.

V. Prediction of Rates and Cross Section Calculation

Central to the algorithm described here is the ability to predict trigger rates for any given set of downscale factors. With several redundant trigger conditions for a given event class the overlap between subtriggers can be considerable. For calculating cumulative rates, it is then not sufficient to add individual subtrigger rates.

A rate prediction which takes into account overlaps between subtriggers can be achieved by reweighting a sample of the triggered events. For a given event j, the probability P_{jk} to be accepted after downscaling by at least one out of N_s subtriggers can be calculated as

$$P_{jk} = 1 - \prod_{i=1}^{N_\mathrm{s}} \left(1 - \frac{r_{ij}}{d_{ik}}\right) \qquad (1)$$

when r_{ij} is the Boolean decision for subtrigger i in the event j and d_{ik} is the downscale factor of subtrigger i. The index k identifies the active sets of downscale factors. The ratio of the probability $P_{jk'}$ evaluated with a new set of downscale factors $d_{ik'}$ to the probability P_{jk} obtained with the active downscale factors is assigned as a weight to each event. Summing these weights over a sample of N_e events and dividing by the active time T during which the events were triggered gives an estimate of the cumulative rate $\nu_{k'}$ of subtriggers $1 \ldots N_\mathrm{s}$ which is to be

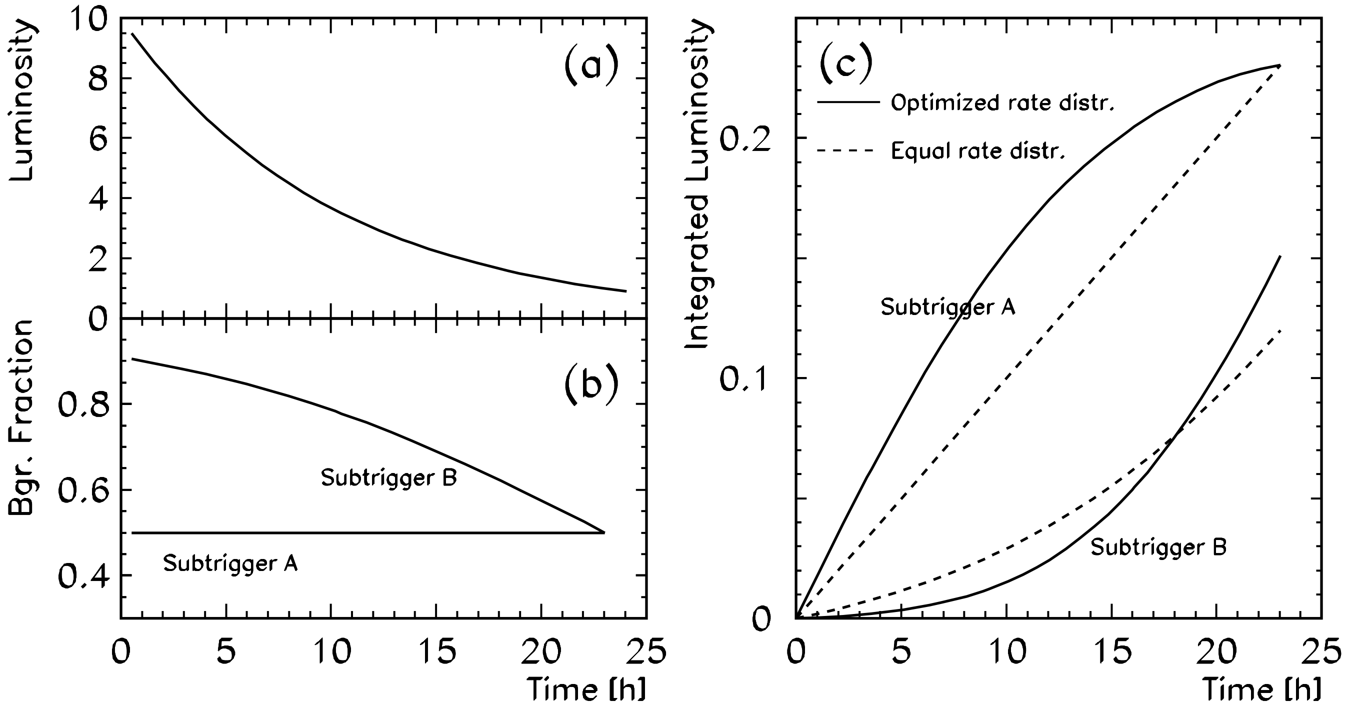

Figure 3: An example of the performance gain due to non-trivial rate adjustment of two subtriggers A and B. (a) Luminosity (arbitrary units) as a function of time; (b) assumed background fraction for each subtrigger; (c) integrated luminosity (arbitrary units) for A and B in the two cases considered.

expected when changing the downscale factors from d_{ik} to $d_{ik'}$:

$$\nu_{k'} = \frac{1}{T} \sum_{j=1}^{N_e} \frac{P_{jk'}}{P_{jk}} \qquad (2)$$

In this way, a reoptimization of the trigger rates for the current running conditions can be performed by iteratively adjusting the downscale factors, calculating the predicted rates and comparing them with the defined rules.

An additional aspect of using varying sets of downscale factors has to be considered and care has to be taken when determining cross sections from recorded data. In principle formula (1) can also be used in offline data analyses in order to calculate from a selected event sample the event rate before downscaling. One would then use the inverse probability, $1/P_{jk}$, as a weight for an event j which was triggered while the downscale factor set k was active. However, with variable downscale factors this can lead to large variations in the event weights. Events which are triggered by the same subset of subtriggers can obtain very different weights, depending on the set of downscale factors that was active. Mixing events with different weights always implies a loss of statistical significance.

This problem can be avoided to a large extent by averaging the weights over as long time periods as possible. Under the assumption that the definition of individual subtriggers and their efficiencies for the selected event class remains unchanged during a time period in which N_d different sets of downscaling factors have been active, a raw cross section σ^{raw} can be calculated from a set of N_e selected events by

$$\sigma^{\mathrm{raw}} = \frac{1}{\varepsilon} \sum_{j=1}^{N_\mathrm{e}} \sum_{k=1}^{N_\mathrm{d}} \left[\mathcal{L}_k \left(1 - \prod_{i=1}^{N_\mathrm{s}} \left(1 - \frac{r_{ij}}{d_{ik}} \right) \right) \right]^{-1} . \qquad (3)$$

Here the integrated luminosity $\mathcal{L}_k$ which has been recorded while the downscale factor set k was active, and the combined trigger efficiency ε of all subtriggers before downscaling have been used in the calculation.

VI. Conclusion

A procedure for optimizing trigger rates has been developed and implemented within the H1 experiment. It was used for the first time towards the end of the 1997 during a special data taking period and its application proved to be very efficient, as it allowed the acquisition of a large amount of unbiased data over a short time interval, giving rise to high precision measurements in several physics channels. Although individual subtriggers showed a large rate variation, the new scheme ensured smooth running with a stable total output rate. Consequently the deadtime remained small and the full sensitivity to rare or potentially unobserved processes was maintained throughout the whole period. A first application during standard data taking after the 1998 startup showed similar results.

In summary, the concept of this new scheme for optimizing trigger rates has proven to be highly successful and the experience gained with this approach will be beneficial for the future running of H1 as well as the next generation of high rate experiments.

VII. References

[1] I. Abt et al., “The H1 detector at HERA”, *Nucl. Instr. and Meth.*, vol. A386, pp. 310-396, 1997.

[2] F. Sefkow, E. Elsen, H. Krehbiel, U. Straumann and J. Coughlan, “Experience with the first level trigger of H1”, *IEEE Transactions on Nuclear Science*, vol. 42 pp. 900-904, 1995.

[3] T. Nicholls, M. Charlet, J. Coughlan, E. Elsen, D. Hoffmann, H. Krehbiel, H.-C. Schultz-Coulon, J. Schütt and F. Sefkow, “Concept, Design and Performance of the Second Level Triggers of the H1 Detector”, *IEEE Transactions on Nuclear Science*, vol. 45 pp. 810-816, 1998.

[4] J. Köhne et al., “Realisation of a second level neural network trigger for the H1 experiment at HERA”, *Proceedings of AIHENP 96, Lausanne*, World Scientific, 1996.

[5] J.C. Bizot, G. Nowak and J. Turnau, “Hardware study for a topological level 2 trigger”, *H1 Internal Note H1-09/92-240*, unpublished, 1992.

[6] W. Haynes, “Experiences at HERA with the H1 data acquisition system”, *Proceedings of the Int. Conference on Computing in High Energy Physics 1992, Annecy, France*, pp. 151-161, 1992.

[7] A. Campbell, “A RISC multiprocessor event trigger for the data acquisition system of the H1 experiment at HERA”, *IEEE Transactions on Nuclear Science*, vol. 39, pp. 255-258, 1992.

[8] S. Aid et al., “Measurement of the total photon-proton cross-section and its decomposition at 200 GeV center-of-mass energy” *Z. Phys. C*, vol 69, pp. 27-38, 1995.

[9] C. Adloff et al. (The H1 Collaboration), “Diffraction dissociation in photoproduction at HERA”, *Z. Phys. C*, vol. 74, pp. 221-236, 1997.

[10] M. Klein, “Precision Measurement of the Inclusive Deep Inelastic e-p Scattering Cross Section at Low Q^2 at HERA”, *Paper presented to the 29th International Conference on High-Energy Physics, Vancouver, Canada*, July 1998.