# New Developments in H1 Computing

C. Grab [*†], G.Eckerlin[†], R.Gerhards[†], C.Kleinwort[†], F.Niebergall[‡]

[*] *Institute of Particle Physics, ETH- Zürich, Switzerland,*
[†] *DESY Laboratory, Hamburg, Germany*
[‡] *II. Inst. of Particle Physics, U. of Hamburg, Germany*

**Abstract.** The H1 experiment has been successfully operating since 1992 at the positron proton collider HERA at the DESY laboratory in Hamburg. From an average input rate of order 10 MHz, good physics events are filtered out in various stages of trigger levels and finally logged to tape at an average rate of about 1 MByte/s yielding of order 5 TByte of RAW data in annual tape volumes. The data are being in part reconstructed already online, and are fully reconstructed offline within just a few hours. The offline computing system thus keeps up with the data taking producing also of order 5 TByte of fully reconstructed data per year. A large portion of data selected according to physics interest priorities are kept on disk for fast access to analyses, being done presently via a batch processing facility on SGI Challenge computers. In this contribution only offline aspects of H1 computing will be covered.

## INTRODUCTION

The main constraints for the overall data handling and computing needs in H1 are to the short bunch separation in the HERA machine of 90 ns, in conjunction with the relatively large cross sections of actual physical processes. The dominant process is photoproduction ($\gamma p$) with a cross section of around 20$\mu$b, which results in an approximate rate of 200 Hz of real physics events when assuming a luminosity of $10^{31} cm^{-2} s^{-1}$. Hence, at an integrated luminosity of 5 $pb^{-1}$ these conditions translate into a total yearly data volume of about $10^8$ physics events, that are successfully handled and thus provide the basic input for H1 physics analyses and publications.

## TRIGGER AND DATA LOGGING

To cope with such high event rates, a dedicated triggering system was developped, that allows to reduce the rate successively in various stages from the input rate of 10 MHz to a reasonably writing-to-tape rate of typically 5 Hz. This was achieved by two trigger stages fully implemented in hardware (level-1 and level-2), augmented by a dedicated software trigger (level-4) running on a PowerPC-farm. The level-1 trigger is deadtime free, employs event pipelining and is based only on subdetector information. It accepts events at a rate of 1-2 kHz. The level-2 trigger exploits the correlations between various subdetectors and can initiate the actual event readout. It is based on neural networks and on topological correlators. The fourth trigger level consists of a farm of some 30 PowerPC processors (mounted on VME-boards) which perform asynchronous calculations on the event input stream. This L4-farm reconstructs and classifies the events according to physics interests, performs calibration and data formatting tasks. The data are then shipped over an FDDI link ($\sim$ 5 MB/s) from the experiment's site to the computing centre site, where the tape robots reside. A job based on PVM (parallel virtual machine) software controls the data transfer, and ensures the proper logging of the raw data onto permanent storage tapes, which happens typically at a rate of $\sim$ 0.5 MB/s. Yet another event processing job, running offline on a large SGI computer, fully reconstructs all events, writes the reconstructed data information onto permanent storage and produces a compressed data format (DST) that resides on disk. This latter data set provides the basic input for all physics analyses work. Calibration and monitoring information is collected, stored in databases and used for data quality checks.

The final data logging rate is thus below 10 Hz. The data taking concept then foresees that even with increasing luminosity, H1 will write to tape at a fixed output rate, thus putting higher demands on the level-4 filter farm. For the long run, H1 plans to merge the functionality of the present level-4 filter farm and the offline reconstruction (L5) into a single operational unit, e.g. a PC-farm.

## OVERVIEW OF THE OFFLINE COMPUTING MODEL

The H1 computing model at present relies on using dedicated servers for the various computing task. They are interconnected by a fast network (HiPPi, FDDi and Fast-Ethernet).

A batch environment has been set up on two SGI Challenge XL computers with a total of about 150 CPU units (one unit corresponds to one R4400 (150 MHz) processor). Presently the NQS batch system is employed, however it is planned to soon move on to LSF.

The first offline reconstruction at present is also performed on one of these SGIs. However, in the short term H1 plans to replace this step by a dedicated PC-farm. H1 has already set up and commissioned such a PC-farm, operated under the LINUX OS, by arranging 10 PCs within a cluster. All nodes are presently dual-PentiumII processors with 128 MB of memory, interconnected by Fast-Ethernet and served by a dedicated IO-node. Monte Carlo production and partial data-reconstruction already is being performed on this PC-farm.

The interactive work is carried by some 30 work-group server (WGS) UNIX workstations (mostly SUN Ultras and SGI Indys). Each WGS serves some 5-10 users, and provides the local computing facilities such as N-tuple manipulation, program development, submission of batch jobs, and all the standard everyday's needs for communications (email, netscape, etc.). Large data transfers from and to the batch servers is discouraged, but not prohibited.

Desktop computing traditionally was achieved by X-terminals connected to the WGS. However, presently all X-terminals are complemented and successively replaced by desktop PCs, operated under the LINUX OS. Only for very special applications the usage of Win-NT on the desktop is tolerated. In the offline environment H1 - in collaboration with DESY ZDV - supports the following OS: Solaris, Irix, Linux, AIX, HP-UX.

All user home directories and all standard application software are served centrally through the usage of the AFS filesystem, which has proved to be very convenient and in general very reliable.

The disk space consists of a total of 1.5 TB served by the two batch and a separate 4-processor SGI Challenge computer, of which about 80% is used to hold compressed data for fast access of the batch processes. Local WGS provide additional user space of order of 10-20 GB each.

The overall H1 computing environment is shown in fig.1.

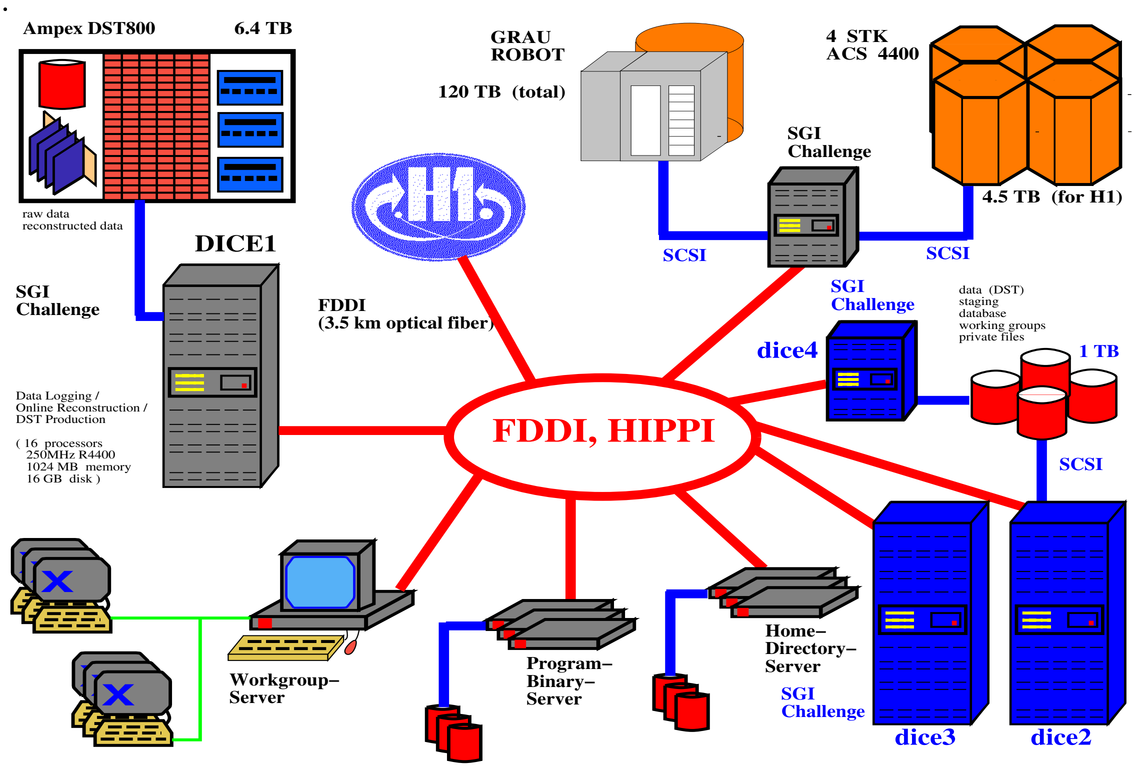


**FIGURE 1.** Overview of H1 computing environment

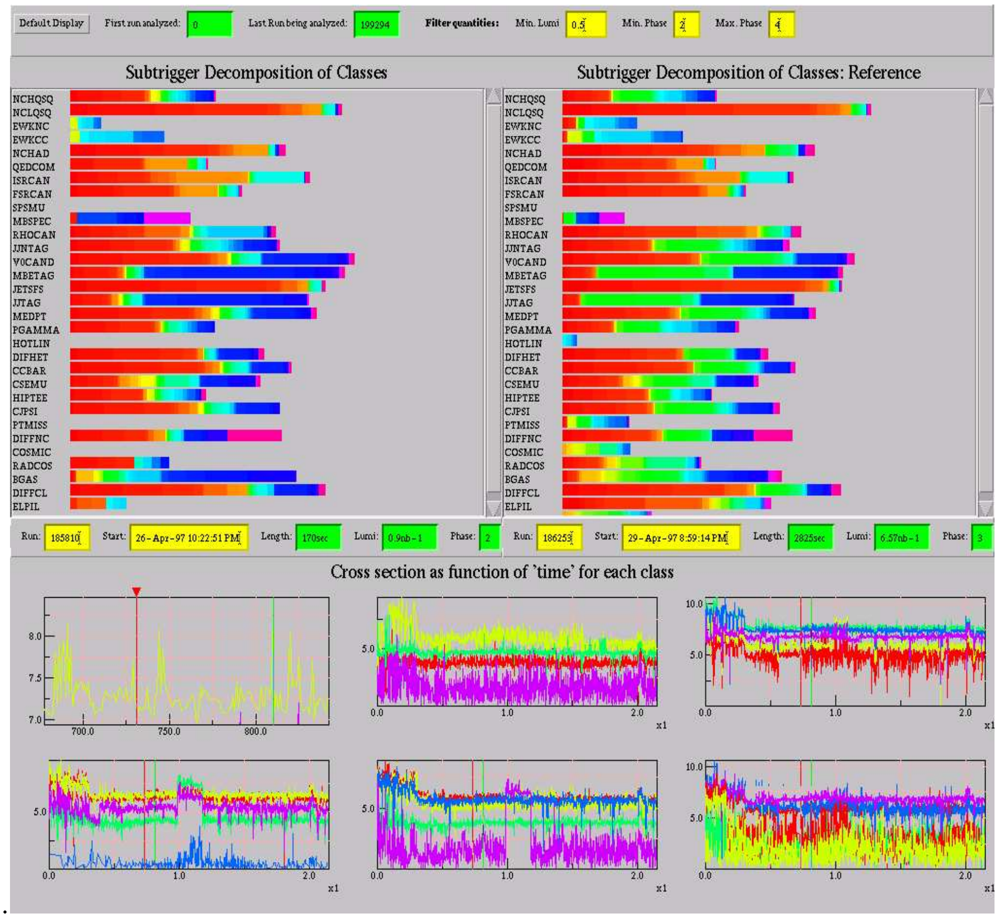


**FIGURE 2.** Example display of the L5-monitoring tool.

# DATA STORAGE AND ACCESS

Many data need to be centrally available and fast accessible by most batch programs and of course all simulation and reconstruction programs.

This includes all detector related data such as geometry, calibration and monitoring information, but also bookkeeping of simulation jobs. Furthermore, personell and equipment information has to be easily accessible and manageable. H1 traditionally stored all in a purpose built database (NDB), but lately H1 migrated most of the data over to a commercial product, Oracle-DB. Its current size is about 1.3 GB, with a growth rate of typically 200 to 300 MB per annum. The H1-specific database NDB is however still used for special cases, such as e.g. satellite installations at outside institutions or regular backups. In the future H1 plans to exploit the increased functionalities of modern database technology (see below).

## NEW DEVELOPMENTS

Various tools are being developped within H1 which make use of new software technologies. However, being a running experiment (taking data *now*), a migration from *OLD* to *NEW* concepts is difficult since the experiment cannot just be interrupted, but rather backward compatibility must always be provided with the existing legacy code. The reconstruction, simulation and most analysis codes exist in Fortran and H1 cannot afford to re-write all; hence languages must be mixed.

A different perspective exists for monitoring and diagnostic tools. They being written newly do not have to comply with legacies, but are now mostly written in modern languages, such as Java or C++; they of course also profit from high-level scripting tools (e.g. Perl).

One point one definitely should not underestimate is the educational issue. Giving an opportunity to new students to acquire skills in the use of modern software-techniques greatly enhances the attractivity of HEP experiments by providing a perfect training ground for later (industry) jobs.

### Event OO-database

An example of an effort to profit from the modern software-developments in the commercial world is the move to use an object-oriented database, in this case it is Objectivity-DB. The compressed data volume on disk themselves are presently stored within a purpose-built serial data format (BOS and FPACK). Additional tools provide for fast data access to a subset of events (=logical records) within a file in the form of so-called index-files. The record information stored in index-file contains the record name, run number, event number, a classification word and the position of the (logical) record in the file. The classification word provides a rather limited number of boolean flags, which are assigned to the events at reconstruction time.

This functionality will be greatly enhanced and much more generalized with the introduction of a new event-tag database, similar to the one in development and in operation in other experiments (ZEUS, STAR, Ceres) and foreseen for the LHC experiments (CMS and Atlas). H1 presently is developping the essential data model according to which the event-tag DB should be structured. The most crucial part being addressed is the design and implementation of the appropriate physics variables. If chosen cleverly, these then will provide the large diversity and functionality for very sophisticated data selection inquiries on the database level, prior to actually accessing the real event data on disk. The information introduced in the OO-DB in the first step includes kinematics and topological variables of event, major final state objects and control variables (trigger, filter).

This additional selective capabilities should in total result in both a higher efficient data selection process for batch jobs, as well as reduce the actual access to the data directly on disk by about an order of magnitude. At a later stage it is planned to add more sophisticated relations beetween objects into Objectivity.

## Examples of using new software-techniques

*JAVA Online monitoring of event classes* : A special tool has been developped to ensure a very high level of data quality by providing real-time measurements of cross sections of all the important physics event classes. It has been completely written in Java following the Model-View-Controller (MVC) design pattern; it accesses the data through network from a mySQL-DB, uses JDBC and the free Graph-package. It's a highly interactive and very physicist-friendly tool with a modern GUI. An example of such a display is given in fig.2.

*Online monitoring of hardware components* : To online monitor the hardware status of the essential subdetector components and their performance, a set of relevant distributions are online filled within the level-4 selection process. A special tool called *KooL* was written in C++ to provide an interactive, user-friendly GUI to visualize these distributions, to manipulate them and to automatically compare them with appropriate reference distributions. A display example is shown in fig.3

*Software design tools:* In the development of the overall reconstruction code of one of the new H1 subdetectors (VLQ), various modern techniques play a role. First of all, the C++ written software had to be fully integrated into the standard H1 Fortran-based reconstruction environment, communicating with and controlling data access to the Oracle database. For the software-development modern tools for software-design and documentation, such as Rational Rose in particular, were applied.

## Safety issues

Lately security issues have intruded the quiet life of high-energy physicists. H1 has therefore also introduced safety measures to better secure and shield the experiment towards unwanted access. Measures taken include (i) introduction of port filtering at the main DESY gateway for all H1 subnets, to better separate the internet from the intranet; (ii) allowance of only secure access to the internal site via secure protocols. Only one computer remains openly accessible to outside, but is very closely monitored; (iii) H1 actively participates in developing DESY-wide solutions, in close collaboration with the DESY computing division ZDV.

## SUMMARY

H1 is successfully handling large data sets and supplying the infrastructure for fast physics turn around. However, H1 is a *running* experiment and has to cope with the legacy of the past, i.e. it must maintain continuity and keep backward compatibility. Nevertheless, H1 started to move step by step into applying more modern techniques, such as event storage in OODB, and using modern software tools for diagnostic, monitoring and analyses to provide highest quality data for excellent physics.

The major challenges for H1's future are high capacity data storage and fast data access in view of the high-luminosity upgrade that HERA is undergoing; high-bandwidth and reliable networks are of paramount importance.

Finally, the required intellectual capacity in terms of qualified software personpower to satisfy the very complex software needs in such experiments is huge and tends to be underestimated. In the long term, future HEP-experiments have to increase their commitment in terms of software personpower in order to garantee the successful operation of HEP-experiments with ever-increasing technological and intellectual challenges.

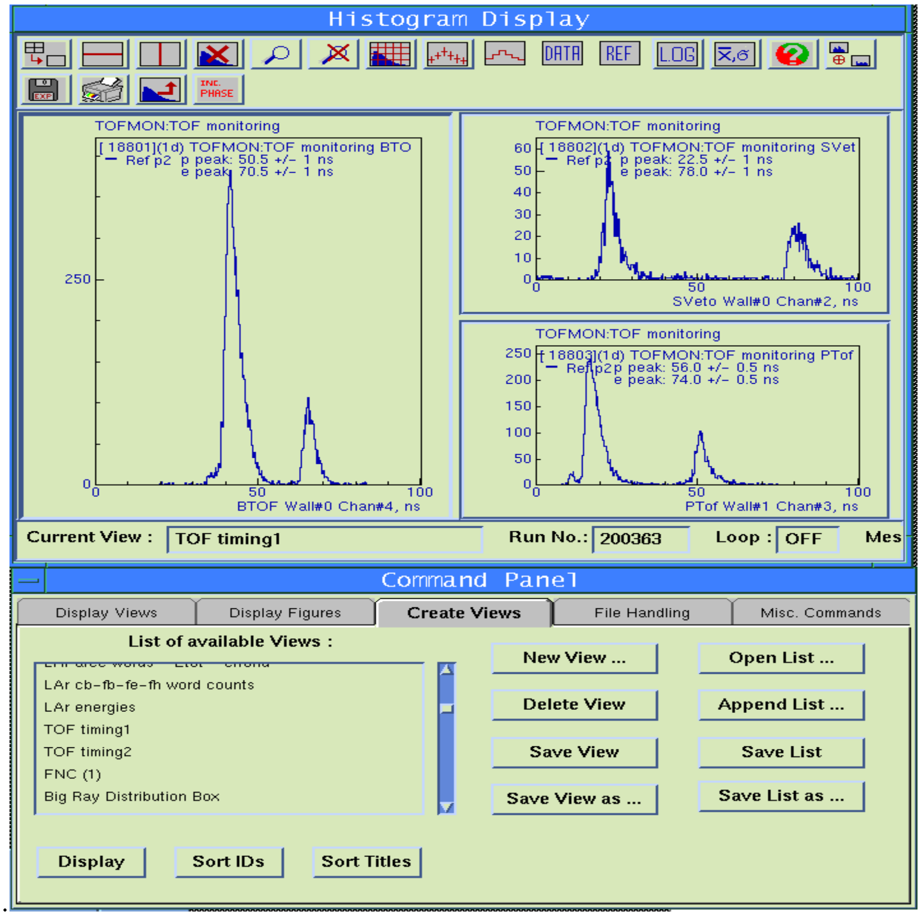


**FIGURE 3.** Example display of the visualisation tool KooL.

## Acknowledgments

The authors gratefully acknowledge the support of many colleagues of the DESY computer centre and all the support from their H1 colleagues. Finally, we express our thanks to the organizers of this conference for setting up this stimulating meeting and choosing such an exciting location.