8 Off-line data handling and simulation

This section concentrates on the off-line data handling, quasi-online data reconstruction, event simulation and the H1 analysis environment.

8.1 Off-line computing

An overview of the H1 offline computing environment is given in Figure 26. The H1 experiment is using two multiprocessor SGI Challenge series computers for all computing purposes [72]. They are connected to the DESY computing infrastructure via fast UltraNet, HIPPI, and FDDI links as well as Ethernet. Fast access to data both on disks and tapes is provided. The disk space amounts to about 500 GB connected via SCSI interfaces. The access to the data stored on disk allows many users to analyze the data simultaneously, thus offering an efficient environment for physics analysis. Major tape service is provided by four Storagetek Automatic Cartridge Systems with a total capacity of about 35 TB, directly accessible from the SGI Challenge using the Open Storage Manager software. Mass data are also kept on an Ampex TeraStore mass storage system, equipped with 3 helical scan D2 recorders and connected to the SGI Challenge via SCSI interfaces. The total capacity of the Ampex system amounts to 6 TB. Multi-user access to data on tape is also offered by an efficient staging mechanism. Export of data is performed via 3490 and Exabytes cartridges.

During data taking, one of the SGI computers is dedicated to the data logging and online reconstruction tasks. The analysis of data is performed on the second SGI Challenge computer and partly on the IBM mainframe. Several institutes have also installed RISC processor based workstations at DESY for analysis purposes, connected to the local network facilities.

Care has been taken from the beginning that all offline processing of the data and simulation was portable between DESY and external institutes. Meanwhile many of these institutes have installed the complete offline program chain and participate in simulation and data analysis. More than half of the detailed detector simulated data were produced in external institutes. Furthermore, most institutes are connected to DESY via fast network links.

H1 uses the dynamic memory management package BOS [73] for its software. A general input/output package, FPACK [74], was written in order to have a simple and unique system for all data transfer in the experiment. It contains automatic wordformat conversions between different machine representations (IBM,VAX,DEC,IEEE), includes record selection options and supports event directory based fast access to data. A database package was developed and is used to store run dependent data and Monte Carlo information. Further, the code manager CMZ [75] is used for the H1 software packages. A set of simple but solid rules, complemented by the necessary subdivision of responsibilities and the imposing of discipline, has created a successful and efficient environment for code development in this large collaboration. Details on code management with CMZ in H1 can be found in [75].

An important aspect of the software is the concept of modularity [76]. Modules are self-contained sets of routines with clear I/O interfaces. They take care of their own initialization upon first call. Modules communicate to each other only via BOS banks. A larger program, such as the reconstruction program, consists of a simple series of module calls. Purpose written utility software performs automatic book-keeping of used input and created output BOS banks and guarantees the internal consistency of the data after reprocessing of modules. The dynamic memory management package BOS includes modularity oriented functions to support this scheme.

H1 also uses an entity relationship model as a basis for data structures of all event data. A data management tool DATMAN [77] provides an easy and userfriendly access to the data.

8.2 Data reconstruction and reduction

The H1 data are reconstructed quasi-online [70] on an 18 processor SGI Challenge computer. The reconstruction task runs in parallel up to 25 identical and independent processes, whereby a set of shared memory areas and semaphores is used to share data among the processes. The raw and reconstructed data are stored on D2 tapes on a dedicated tape robot. During normal operation, the data taking proceeds with an average rate of about 500 KB/s, leading to an expected yearly data volume of about 5 TB. A part of the event information is stored on disk (DST) for physics analysis for the full data taking period.

The reconstruction of an event takes on average 1.5 s on a SGI Challenge processor. Thus the reconstruction task can cope with the data logging rate of 10 Hz of the experiment. In all, the reconstructed data become available to the users typically a few hours after the raw data are recorded by the experiment. The amount of recorded data by experiments at HERA is large. During the 1994 data taking period about 7×10^7 triggers were recorded by H1. The total integrated luminosity in that period was about 3.5 pb^{-1}. Off-line event classification and background rejection filters, based on reconstructed energies and tracks, reduce the amount of background presently by a factor 3 to 5 during reconstruction. Background rejection filters include cosmic muon and muon-halo filters, recognition of coherent noise patterns and filters for beam-wall and beam-gas collisions originating from outside the detector.

8.3 Event simulation

A complete detector simulation program has been assembled within the GEANT [78] framework. The geometry of the full detector and the beamline within ± 100 m around the interaction region is implemented with two different levels of detail, called fine and coarse granularities. For the fine granularity the longitudinal structure of all calorimeter stacks is implemented layer by layer, whereas for the coarse geometry a calorimeter stack is implemented as a block of properly mixed homogeneous material with no longitudinal structure. For the tracking detectors separate volumes are implemented for each active cell (wire) for the detailed geometry, whereas for the coarse geometry an entire gas volume is treated as one volume. For accurate detector response, particularly in the detailed geometry option, the tracking cutoff parameters for the kinetic energies have to be set as low as 1 MeV. This leads to simulation times of the order of 200 s for a typical low Q^2 deep inelastic event on a SGI Challenge processor. Three strategies were followed to tackle this large CPU time consumption for event simulation:

- The time consuming part of the actual tracking of particles through the geometry was strictly separated from the digitization part of the detector response. Run dependent detector effects and the actually achieved resolutions can easily be adjusted by reprocessing the digitization part only, using only a fraction of a second per event. A similar scheme was used for the simulation of the trigger response.

- A book-keeping is made of all energy depositions in the detector, both for visible and invisible (nuclear breakup, neutrinos, slow neutrons) energies, and both in active and dead detector regions [79]. This allows to reconstruct the calorimeter response after simulation with simple response functions, which are basically sums over the smeared true energy depositions. The absolute energy scale can thus be reconstructed correctly, which gives an important handle for testing the reconstruction programs and also to understand the behaviour of a non-compensating calorimeter.

- A fast but accurate energy shower parametrization [80, 81] was developed. This so-called H1FAST mode was included into the GEANT framework. The main idea is sketched in

Figure 27. The time consuming part of the full shower development in the detailed geometry is replaced by a shower parametrization in the coarse geometry leading to reduction of the CPU time consumption by a factor 10. Showers which cross crack boundaries are simulated in detail, in order to keep accuracy.

8.4 Physics analysis

The visual aspect of the physics analysis is mainly handled by a purpose written general system for graphics applications, baptized LOOK [73]. The aim is to have one –and only one– graphics package for all graphical applications: event display, histogramming, analysis, etc. LOOK can be considered as a layer between the user application and low level graphics functions, for which at present GKS functions are used. LOOK organizes the management of the display and e.g. hardcopy devices. It further contains a histogram package and a powerful command processor and is portable to many platforms. The H1 event display program, H1ED, is an application based on LOOK. Figure 28 shows an example of an event display and histogram analysis combined in one session. LOOK can be used for interactive analysis, e.g. n-tuple analysis of data. Also PAW ([82]) is used for interactive analysis.

The program H1PHAN is a tool to access event data and simplify physics analysis programs. The data are filled in internal event buffers called Q-vectors, which are accessible in simple Fortran DO-loops. Thus the user does not need to know the –sometimes complicated– underlying bank structure of the data. H1PHAN further contains tools for particle identification, secondary vertex fitting, jet finding and analysis, determining kinematic variables of deep inelastic event candidates, etc.