H1oo
A Centralised Analysis Framework for the H1 Experiment

Michael Steder, DESY

H1oo revises previous analysis paradigms in various respects. It provides analysis-ready data rather than mere collections of functions and algorithms. The H1 collaboration has continuously improved a **common, extendable and re-usable framework** in which the best expert knowledge and standardised physics algorithms are accessible for all users. All data files for testing and analysing are produced centrally in a semi-automatic procedure, the production includes the application of the **latest alignment and calibration constants** derived from the experts of the particular subdetector.

This effort substantially enhanced the physics capabilities of official H1 software and - at the same time - reduced the turn-around time of physics analyses. For **transparent exchange** of algorithms between different working groups and portability of code between the different stages of data production and physics analysis, it is vital to have a **homogeneous framework** in which one common programming language and coding convention is used.

Raw Data and MC

HERA I : 0.3 billion events
HERA II : 0.8 billion events

MC events are passed through the detector simulation to mimic data.

Reconstructed Files

POT	200 kB/event
total	200 TB
DST	15 kB/event
total	15 TB

After reconstruction the events are stored in DSTs (Data Summary Tapes), which serve as input for the production of H1oo analysis files.

H1oo Analysis Files

HAT	½ kB/event
total	0.2 TB
µODS	3 kB/event
total	1.5 TB
ODS	15 kB/event
	not stored

Simulation and Reconstruction (DST 7)

The **raw data** recorded with the H1 detector at HERA were written to POT (Physics On Tape) files, containing among other things wire hits, channel numbers and cell energies as well as a first reconstruction of cells and tracks. In a next step the events are fully reconstructed with the H1 reconstruction software h1rec and information relevant for analysis is copied to the DST (Data Summary Tape) files.
Simulated physics events are generated using various Monte Carlo (MC) programs and passed to the Geant 3 based H1 detector simulation. After the simulation step, the MC events are reconstructed using the same reconstruction software as for data.

Over the years, the **H1 reconstruction software** has been improved continuously. With the latest data reprocessing '**DST 7**' the H1 software has reached its final and best precision. Tracks in the central tracking detector are now searched for using a broken line fit treating the transition region between inner and outer jet chamber as thick scatterer, the silicon tracking detectors are included in the vertex reconstruction and dE/dx information is used as mass hypothesis during track reconstruction. The dead material description in the MC simulation has been tuned using the output of a nuclear interaction (NI) finder. The performance of the nuclear interaction finder is impressively shown in figure 1. After all the z-vertex resolution is heavily improved, a systematic uncertainty of 1% per track and of less then 1 mrad on Θ_{Trk} is achieved for the central tracking detector, allowing to employ the **best possible precision** for the final H1 publications.

Figure 1: Detector map as obtained from the NI finder

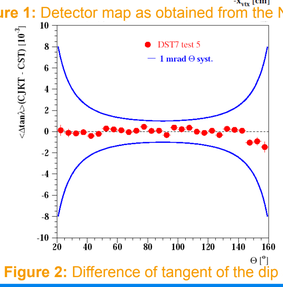

Figure 2: Difference of tangent of the dip angle λ

Production of H1oo Analysis Files

Data Structure

The H1oo data is organized in a **three-layer structure**, corresponding to different RooT trees. The data stored in the different layers are produced by so-called filling code, which is organized in modular units called finders. The algorithms and the data are organized in about 600 classes in around 50 packages all inherited from TObject (RooT) and implemented in C++. With this H1oo provides a **collaboration wide standard** of doing event and particle reconstruction and selection.

The base layer, **ODS** (Object Data Store), is an RooT-format interface to the full information stored in the DST files.
The middle layer, **µODS** (micro ODS), consists of particle level four-vector information and contains all information needed by physics analyses.
The uppermost layer, **HAT** (H1 Analysis Tag), contains event level information like particle numbers and event kinematics for fast selection of events.

Cluster Separation, Alignment, and Calibration

To improve the software compensation of the LAr calorimeter even more, a neural network has been trained to improve the separation of electromagnetic and hadronic clusters. This allows an **almost perfect reconstruction of the electromagnetic fraction** as shown in figure 3. Applying the cluster separation yields a better starting point for the subsequent calibration, since the measured energies are already close to the true ones.

In the latest H1oo release, a new calibration scheme has been implemented **unifying different calibration methods** and making the application to data and MC fully transparent for all users and very easy to access.

After calibration the electron energy uncertainty is well below 0.5% and a jet energy uncertainty of less than 1% has been achieved (fig. 4). In order to achieve such a precise jet energy reconstruction, a **new hadronic calibration package** has been developed. In a first step, all clusters are calibrated, and in a second step, all clusters inside jets are fine-tuned as function of Θ_{jet} and η_{jet} using an unbinned χ^2 method in order to avoid steps at bin edges.

Particle Finders and Physics Algorithms

During the production of the analysis files, particle finders run on the reconstruction output creating objects for identified particles (e.g. electrons, muons) and composed particles (e.g. D*, Jets, J/ψ, K^0) in the µODS files. The created objects allow an fast, efficient, and user-friendly access to all relevant information of the particular particle. The **particle finders unify the knowledge of all experts** of the collaboration and are subject of regular enhancements.

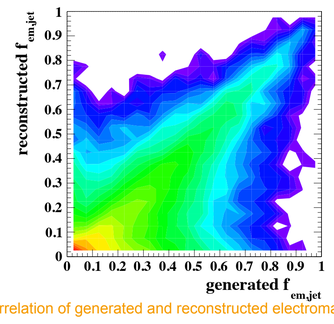

Figure 3: Correlation of generated and reconstructed electromagnetic fraction.

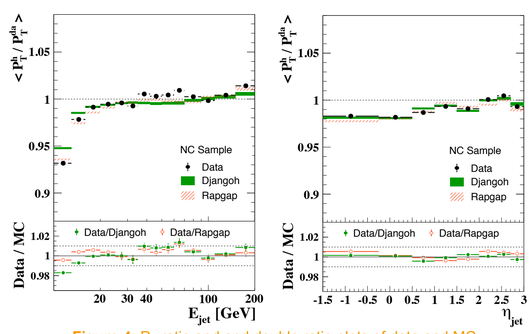

Figure 4: P_T ratio and and double ratio plots of data and MC as function of jet energy and jet rapidity.

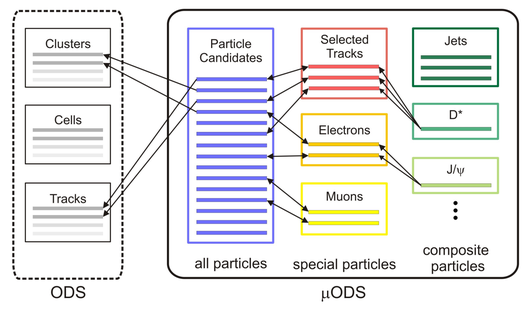

Figure 5: Schematic view of the ODS and µODS structure.

Analysis Level Software

Equality in diversity

The common software development in the H1 collaboration does not end at analysis level, though all users are can write their own code to access data and MC events in the HAT and µODS files. **Many aspects are similar** or even identical for all high energy physics analyses - e.g. event selection, filling and binning of histograms. Within H1oo exists a framework which **decomposes all these aspects of physics analyses** into dedicated classes like a 'histogram manager' or an 'event selector'. This makes it easily possible to migrate (parts) of one analysis into another one and to reuse existing code.

H1Calculator

Another H1oo package, the H1Calculator, **provides access to many event and particle quantities**. Its modular design makes it easy to extend the H1Calculator functionality by adding new classes. The H1Calculator is implemented as singleton to ensure self-consistency for all access to HAT/µODS variables or to variables composed of these. This is in particular important for systematic studies, where for example a shift in the electron energy results in a corresponding change of the total calorimetric transverse momentum.

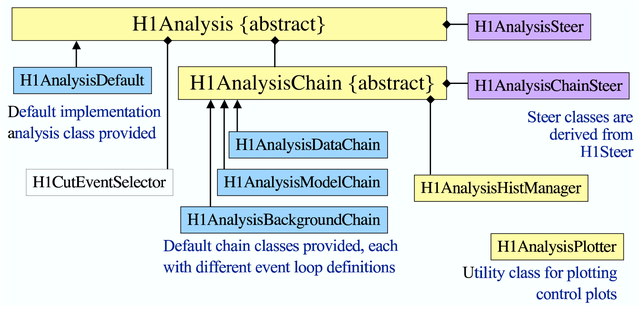

Figure 6: Main classes and their relation in the H1Cuts package

MC Production on the Grid

After final data reprocessing (DST 7), **2.8 billion MC events** for physics analyses have been simulated and reconstructed by the H1 MC team within few month. This number could only be achieved by a very fast and efficient production scheme employing the Grid infrastructure in Europe and Russia. In addition, the powerful H1 batch farm plays an important role for the production of the numerous small requests with less than 10^6 events, since the overhead on the batch system to produce these small requests is small compared to the grid.

All MC requests are registered in a central data base and the DST and H1oo files are produced centrally by the MC team. This scheme allows to produce up to **0.5 billion MC events per month**.

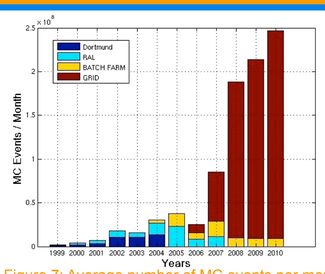

Figure 7: Average number of MC events per month

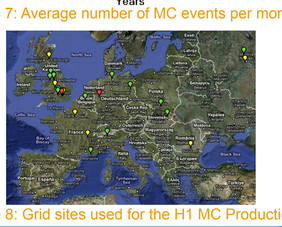

Figure 8: Grid sites used for the H1 MC Production

Data Preservation

The H1 software team is very active in the field of data preservation within the DPHEP initiative. H1 is aiming for 'level 4' of the Data Preservation Models: preserving reconstruction and simulation software as well as basic level data to **allow full data analyses in the future**.

In this context the full H1 software chain will be put into a validation scheme suggested by DESY-IT, in which the software is automatically recompiled on a regular basis. The reconstruction and H1oo file production is run and an automatised validation script checks all links of the software chain.

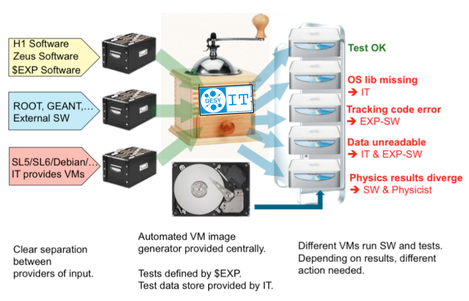

Figure 9: Illustration of the data preservation infrastructure as suggested by DESY-IT.