New Data Analysis Environment in H1

U. Berthon[2] *,T.Benisch*[1] *,R. Gerhards*[1] *C. Grab*[3], *T. Hadig*[4], *P. v. Mechelen*[5]

[1] DESY, Hamburg(Germany)
[2] LPNHE Ecole Polytechnique,IN2P3-CNRS, Palaiseau (France)
[3] Institut für Teilchenphysik, ETH, Zuerich (Switzerland)
[4] I. Physikalisches Institut der RWTH, Aachen (Germany)
[5] Universitaire Instelling Antwerpen, Antwerpen (Belgium)

Abstract

In view of the expected increased luminosity which will be translated into an increased level of high level filtering for physics analysis, the H1 Collaboration is pursuing a major revision of its computing environment. H1 has adopted ROOT as the tool of choice for a compact framework to do the H1 data I/O, to perform physics analysis, and to interactively display events. We present the general strategy for physics analysis pursued by the H1 experiment and related implementation issues.

Keywords: n-tuple analysis,object-oriented,analysis framework,event display,ROOT,H1

1 Introduction

H1 is an experiment at the electron proton collider HERA at DESY (Germany) which started collecting data in 1992. Thus, essential pillars of the analysis software architecture as well as the basic data model were established nearly ten years ago. The year 2000 will be marked by a long shutdown to allow for a substantial upgrade of the HERA collider and H1 detector. The H1 Collaboration has decided to profit from this new phase of HERA and develop a new analysis framework to be used after the shutdown. The current analysis is performed in ”independant sequential steps”. The goal of the redesign is to provide a more unified access to data of various level of abstraction, to incorporate a new data storage optimized for efficient access to data, and to provide physics analysis tools as well as user-friendly interactive capabilities for event display in a unique framework [1]. By adopting ROOT [2] for data storage as well as analysis and graphics, H1 hopes to benefit from mature and promising object oriented technologies.

2 The current Analysis and its Drawbacks

The data are currently written in a platform independent format provided by the Fortran package FPACK especially tailored to the original H1 needs. For reasons of data storage and management, the “data model” was separated into two increasing levels of abstraction: the POTs (Production Output Tapes, approx 200 kB per event) combine the event reconstruction output and raw data, while the DSTs (Data Summary Tapes, approx. 15 kB per event) provide a summary of data, sufficient for most of Physics analysis as well as event selection. Due to the technology used, events must always be read and written completely. This results in very long delays both for analysis jobs (approx. one week to process one year’s data) and rewriting of the DSTs (several times a year, lasting a couple of weeks each time).

Another important shortcoming is the multiplicity of separated frameworks implied in doing analysis: H1PHAN, the H1 Physics ANalysis software together with hbook is used for the first step (event selection and preparation of n-tuples), PAW for the data analysis, and the homemade LOOK graphics framework for event display. This breaking of the analysis process in separate steps makes the physicists work clumsy and inefficient, and maintenance difficult. Moreover the outdated design and especially the old-fashioned user interfaces are a contributing factor in the lack of user-friendliness.

3 The new Analysis Environment

The architecture of the new analysis environment is shown in Figure 1.

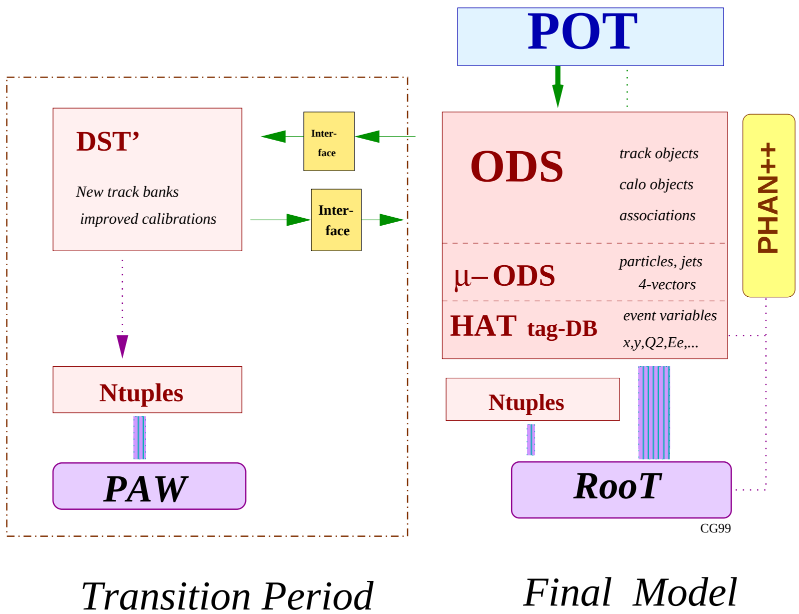

Figure 1: Architecture of new environment

3.1 The new Data Storage

One aspect of the revolution in analysis techniques in H1 is the redesign of the data storage both in terms of segmentation and universality of data access [3]. Two additional layers will be added on top of the presently existing DST (becoming ODS=Object Data Store) and POT layers: first of all a "micro Object Data Store" layer (μODS) at the 4-vector level. On top of these layers we will put a HAT (H1 Analysis Tag), containing approximately 100 kinematical Physics variables for fast event selection. From POT via ODS, μODS to HAT the level of abstraction as well as access frequencies increase.
In order to improve efficiency of access to data, the underlying data storage technique will be changed. The I/O package chosen for ODS, μODS and HAT is ROOT, since it meets our needs in several respects:

- high I/O efficiency
- possibility of writing logically coherent data onto several separate streams
- possible access to parts of the data only

3.2 The new Analysis Framework

For the other aspect, the integrated analysis framework, ROOT also proved to be a good candidate. It has very complete data analysis capabilities and provides a graphical user interface and graphics classes which are well suited for event display. Using ROOT obliges us to rewrite the physics analysis algorithms from H1PHAN in C++. The new and improved package for physics analysis, H1PHAN++, will fulfill two tasks: on one hand it will be used for the filling of the μODS and HAT and on the other hand it will enable the analysis of the data on all levels. The transition from H1PHAN to H1PHAN++ can be done gradually, using C-wrappers for the parts of the FORTRAN code not yet rewritten in C++.

The redesign will have several advantages:

- Standardisation of certain algorithms by combining different analysis methods into one coherent package. This is of course highly recommended in an experiment that has been running for nearly 10 years and where a lot of alternatives exist.
- Lowering of the threshold for newcomers in physics analysis by providing a complete and consistent framework. Technical complexity will remain hidden until necessary in the user's analysis.
- Improvement and sophistication of algorithms by taking the new memory and CPU capabilities into account.

The event display (Figure 2) is being rewritten using the same object-oriented technology, and becomes as H1RED (H1 ROOT Event Display) an integrated part of the analysis framework. It is derived from the Alice event display [4] in what GUI and 3D graphics is concerned. For the 2D part, the former display h1ed will be completely integrated, but the same interactive capabilities that exist in the 3D part will be added: for example immediate access to the underlying Physics information, or sliders providing interactive cuts on Physics variables.

3.3 Backwards Compatibility Issues

Since H1 is a running experiment, a big effort has to be made to ensure complete backwards compatibility. As can be seen from the architectural scheme, forward and backward translation from the (old) DSTs to the (new) ODS is guaranteed, allowing users to continue using their Fortran analysis code in the future. But since data are written directly in ROOT format, such a way of working, though possible, will be penalizing in speed and is discouraged.

4 Expected Benefits from the new Scheme

We expect an important speed improvement due to the redesign of the datastorage, and especially due to the fact that

- event selection can be done at the HAT-level
- most of the analysis can be done at the HAT or μODS level

The framework will profit from modern GUI techniques in general (menus, multiple windows,...). Specific ROOT features making life easy are:

- the unicity of the language C++ as programming, macro and interactive language
- the RTTI introduced by the dictionary, giving you transparent and possibly visual access to the complete H1 event structure, removing the need to know detailed location and naming of the event data

The choice of the same framework for both critical parts, data storage and data analysis, allows to introduce a dialogue between event selection, analysis and offline graphics. An example of such a dialogue is the possibility to interactively choose an event to be looked at in the display by

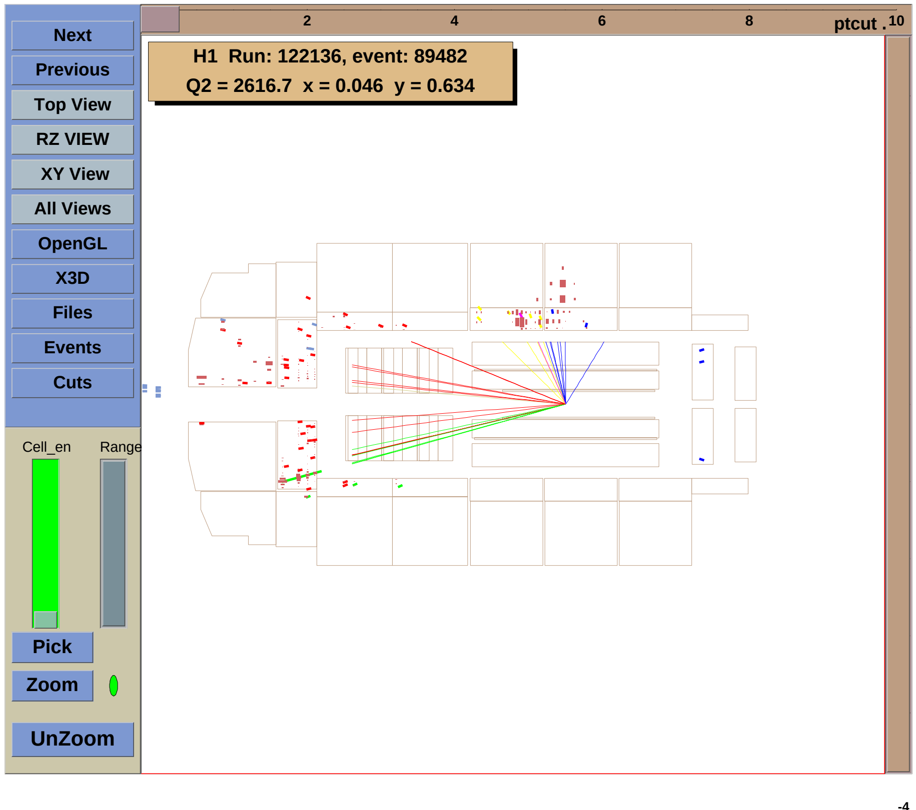

Figure 2: Interactive Jet Analysis in H1RED Event Display

selecting a certain region in a scatterplot or a 1-dimensional histogram. The integration of analysis software and event display into the same framework makes the event display able to immediately reflect the results of a piece of interactive analysis (Figure 2).

5 Conclusions

The new H1 analysis environment is a thorough reform of the system used until now. It will correct the serious shortcomings of the current environment presents. We expect the benefits to be important enough to overcome possible reluctance of H1 physicists in learning new languages and techniques.

6 Acknowledgements

I would like to thank Y. Sirois for many very helpful discussions.

References

1 C.Grab et al.: New Developments in H1 Computing. International Conference on Computing in High Energy Physics 98 (CHEP98), Chicago, Illinois, August 31 - September 4,1998.

2 Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also `http://root.cern.ch/`

3 T. Benisch et al.: New Data Storage Model for H1. Submitted to this conference.

4 `http://alisoft.cern.ch/offline/`