An Evaluation of PAW

J.J. Bunn   *CERN, Geneva, Switzerland*

Abstract

This short paper describes the Web-based questionnaire that was made available for users of the PAW (Physics Analysis Workstation) program. Over three hundred replies to the questionnaire have been received, and they are summarised with the help of data from PAW session monitoring. Following this, some requirements for a second generation tool that would replace PAW are proposed.

Keywords: Data analysis visualisation PAW

1   Introduction

In late 1995, users of PAW were invited to fill out a lengthy Web-based questionnaire *http://wwwcn.cern.ch/asd/paw/question.html*. The questionnaire contained sections that addressed the purpose for which PAW was being used, how often it was used, on what operating system, the extent of satisfaction with the tool, the favoured data formats for analysis, and so on. In addition, the users were asked to provide details such as home institute, experiment name, and where PAW was being used. The questionnaire is still available, and is linked to via the PAW Web page *http://wwwcn.cern.ch/asd/paw/*.

Since that time, over 300 replies have been received. These replies are stored in a text file, which has been processed into statistical information in the form of histograms, using a sophisticated PAW macro. A summary of this statistical information, together with data obtained from PAW session log records is made (also linked via the PAW Web page), and conclusions drawn.

2   Typical use of PAW

PAW is typically used to analyse experimental HEP data stored as Ntuples on disk. Ntuples may be thought of as a single table database. There are two flavours of nTuple, those that have the data stored with one row after the after, and those which have the data stored with one column of data after the other. Row-wise Ntuples are the original form. The newer column-wise Ntuples were introduced to speed up data access for typical queries (which tend to require data from one or a few columns only), and are heavily used on the PIAF systems. The PIAF systems run a special version of PAW that has been parallelised, and the PAW user can connect to this system using an IP connection from his/her workstation.

The survey results show that the majority of users are still using row-wise Ntuples. These are typically attached by PAW, and a selection function applied to a range of columns in the nTuple. The selection function is often a Fortran routine that encodes an algorithm on the column variables. Such routines are interpreted by the COMIS program, which is incorporated in PAW. In some cases, the native Fortran compiler may be used instead of COMIS to dynamically link the user's Fortran function with PAW. Rows of the nTuple which pass the selection function are used to fill histograms, which then show a distribution of the function value, or some combination of the column variables. The histogram appears in the PAW graphics window. At this point, the user will often decide to fit the distribution in the histogram

to a theoretical prediction, which is also encoded as a Fortran function. For this, the MINUIT program (also incorporated into PAW) is used, and the results of the fit are shown as a χ^2 value in the PAW text window, and as a curve in the PAW graphics window.

Once the user is satisfied with the plot, he/she will annotate it ready for publication, maybe using a mixture of fonts, super- and sub-scripts, and so on. At this point the plot will typically be stored as a PostScript file.

In contrast to the above pattern of use, approximately one quarter of all PAW "sessions" take place in batch, where the program is run without keyboard input, and under the control of a KUIP macro.

Summarising, users of PAW make heavy use of macros, interpreted/compiled Fortran selection functions and fitting.

3   Suitability to Tasks

Users were asked to rate PAW for ease of use, reliability, functionality, and so on. In general the functionality of PAW was highly rated, but there were poorer results for reliability and documentation. Despite a considerable publicity campaign when it was introduced, the Motif version of PAW (called PAW++) was still little used (around 10% of respondents), and the reason for this is suspected to be due to both a poorly designed user interface, and, for proficient users, its having few advantages over the command line mode version.

Considerable independent use of the PAW component packages HBOOK (for histogramming), MINUIT (for fitting) and COMIS (for Fortran function interpretation) is made. This is not surprising, as these ubiquitous packages existed before the PAW system integrated them into its fabric.

The incorporation of packages such as these, places the PAW system simply without competition in terms of functionality for HEP data analysis.

4   A Replacement for PAW

By examining how PAW is used, and looking at the users' evaluations of it, some requirements for a replacement tool can be drawn up. For example, from the replies to the questionnaire, the *functionality* of PAW appears sufficient for today's physicists needs. Any new tool should thus target equal or increased functionality.

So why would one want to replace PAW ? The answer to this question lies in the intrinsic limitations of the data model used. This model is instantiated by Ntuples, which are database-like tables. The table view of HEP data is insufficiently rich to fit with the new object paridgm that is gaining considerable footing in HEP data analysis. Reading and writing Ntuples is achieved using CERN-written data access routines, which are known to be buggy and occasionally inefficient. The Ntuples themselves are limited in size, a limit which is known to be too low for LHC data. The PAW system itself is hard to maintain, partly because of its size, and partly because of a lack of coherence between its component packages. It relies little on industry standards, most noticeably for its graphics and I/O parts. Finally, it is hard to customize to individual needs: one either takes the whole system or nothing at all.

A new tool would offer equivalent or better functionality to PAW in the areas of:

- Object representations of the data

- The convenience and speed of access to user data

- The manipulation of the data, with user-defined functions and so on

- Easy automation of data manipulations

A new tool should be more modular, robust, maintainable and modern:

- With more user freedom to pick the required analysis components

- With less home-grown code

- With standard layers for graphics and access to an OO database

- With more emphasis on ease of use for beginners

It is proposed that any new tool should target Windows/NT as the primary OS environment.

Acknowledgements