3.5 Reconstruction and analysis environment

The raw data taken online are saved on IBM data cartridges. These raw data then go through a standard offline reconstruction immediately after data taking. The reconstructed events are also saved on IBM data cartridges. The raw data size is about 100kbyte per event and reconstruction adds about 60kbyte per event(it took a few thousand 250Mbyte IBM data cartridges to hold the raw data of the 6.8 million events taken during the 1993 running period). For easy and fast access to data, all reconstructed events are passed through offline DST(data summary tape) filters provided by the physics groups. For the 1993 data, about 1.5 million events were selected by the DST filters and copied to ZARAH disks in reduced data structure(event size 35kbyte). ZARAH[49,51] is a farm of SGI work stations for offline reconstruction and physics analysis.

The standard offline reconstruction program is called ZEPHYR.[18,49] Both real data and Monte Carlo events can be reconstructed by ZEPHYR. Reconstruction is done in three phases: (1) reconstruction within individual components, (2) global cluster finding and global matching between tracks and cluster, and (3) particle identification. In Phase 1 reconstruction, quantities specific to readout hardware such as charges on PMTs, hit wire positions and drift time, are converted to physics quantities such as energy deposits, particle arrival time in calorimeter cells, trajectory parameters and momentum of the tracks in the CTD. Constants sets used in the offline reconstruction, such as calibration constants, description of the detector geometry and magnetic field and data taking conditions, are included in a temporal data base. Correct version of the constants is selected according to an event time stamp. In Phase 1 reconstruction of the CAL, clusters and jets are also reconstructed. However, there is no unique algorithm for cluster and jet finding. In Phase 2 reconstruction, global matching identifies which track and cluster are associated with one particle. If a hadron or a jet of particles deposits energy in both the CAL and the BAC(backing calorimeter), global cluster finding in the CAL and the BAC is then required. Global matching is not a trivial task since there are gaps between the different detector components. Extrapolation of a CTD track into the surface of the CAL must take into account the magnetic field bending effect in case of low momentum tracks. Particle identification is performed during Phase 3 reconstruction. This is a very complex task which combines measurements from the CTD, CAL, HES(silicon pad hadron electron separator) and the TRD(transitional radiation detector). Not all of these components were completely implemented in 1993. The response to electron and hadrons in these components are still under study. In particular, $J/\psi \rightarrow e^+e^-$ events have contributed to our understanding of the low energy electron identification.

Reconstructed data are the starting point of the analysis by users. The DST data on the ZARAH disks(MINIDST data) suffice for most purposes. The ZEUS analysis environment has been developed over several years (see Ref. 49 for an introduction to the ZEUS software). User analysis codes are written in Fortran77 subroutines. Most of the routine activities, such as event reading and writing, reading of the ZEUS detector constants, some initialization and termination procedures, are handled by a skeleton program called EAZE. ZEUS data uses the ADAMO[52] format which can be accessed with the CERN TAP library subroutines. Software from CERN such as HBOOK[53] and PAW[54] are used extensively for data analysis and presentation. Graphic display of the ZEUS detector and events is provided by LAZE[49] and GAZE.[49] The PHANTOM[49] library contains subroutines provided by the ZEUS users, extending from (x, y, Q^2) reconstruction, cluster and jet finding, track-cluster matching, particle identification, to structure function calculation.

The ZEUS software are transportable and have been installed in off-site institutes. Time consuming Monte Carlo event generation tasks are now shared by off-site institutes as well. Off-site users can access to data on ZARAH and DESY IBM computer as easily as on-site users.