L. O'Neill
3.8.78

JADE - Computer Note No. 6

# Proposal
for
## Management of Calibration Data

This note essentially summarizes the discussion at the JADE offline software meeting of July 25, 1978. "Calibration data" as used in this note will refer, as usual, to conversion factors between raw data and physically interesting quantities, for example between ADC channel numbers and energies in MeV, but it will also refer more generally to any data describing the state of the JADE - detector or of PETRA at a given time, for example to survey data on the positions of detector elements or to data on the currents and polarization of the stored beams.

It is clear that a very large number of data describing the state of the experimental apparatus as a function of time and covering all times at which data have been taken must be available to the various analysis programs, since we must assume that events from the entire course of the experiment will continue to be reanalyzed, by more refined programs and for different reasons, long after they have first been processed by the offline software. The management of the necessary calibration data will be complicated by at least two distinct effects : First, some of the physical quantities of interest, photomultiplier gains, ADC pedestals, drift velocities and so on, will actually change with time, second, our estimate of what some of these quantities were at a given time will improve, or at least change, at later times, for example when tracks have been used to estimate wire positions or to check the z coordinate measurements by the inner drift chamber. Thus calibration data will have to be updated not only in the sense of being added to, but also in the sense of being retroactively revised.

The following table shows the various kinds of calibration data for which need has been anticipated. Also shown are the rates at which significant changes are likely to occur, according to current best estimates, and when the data will be available with various degrees of accuracy. Unavoidably the information is still inexact and the definitions vague in many cases. The term "data reduction" refers to the stage at which a given set of events passes through the first offline analysis and is split from the most obvious background events. The term "advanced data analysis" refers to a time when, for example, particle identification is being attempted via dE/dx measurements or when tracks are being fitted for the ultimate in momentum resolution.

Not noted in the table are the numbers of parameters needed to specify each set of data at any one given time. These are often not yet known very precisely. However it is clear that the total number of parameters will be dominated by the ~ 30 000 analog memory gains and parameters describing the ADCs for the inner detector. It is estimated that a total of ~ 40 000 parameters will be required to specify the state of the apparatus at any one time.

Most of these data will be available in raw form from special calibration runs during which the flash lamps for the leadglass counters will be pulsed, and/or the test pulsers for the drift chamber electronics operated, etc. For technical reasons concerning the operation of the flash lamps for the leadglass counters, it does not appear feasible to have special calibration events occur continuously and be transmitted one by one, interspersed among the normal events, to the IBM. However the calibration run data will be sent as usual over the hardware link to the IBM and written to the normal raw data tapes, once a day or so.

The calibration data so transmitted may be in the form of individual events, in which case the actual calibration parameters will have to be derived at the IBM, or it may consist of the parameters themselves, already derived at the NORD. In either case the group responsible for each hardware system requiring calibration data is responsible for providing programs to extract the necessary parameters from the raw calibration data. No one in the online or offline software group is developing such programs. The rest of this note concerns only the administration of calibration data assumed to be available in reduced form.

The proposal made here envisages the creation and continuous maintenance of two disk files : one, a cyclic buffer, would contain smooth fits to the time dependence of the various calibration parameters as best they were known, but it would cover only a short interval around the time at which



| Type of Data | Time for significant change to occur | First guess available | Known as needed for data reduction | Known with final accuracy |
|---|---|---|---|---|
| Survey | ~ months | before run | before run | advanced data analysis |
| Drift velocities | ~ 1/2 day | before run | ? | ? |
| Leadglass P.M. gains, ADC params. i.e. pedestals, linearity | ~ 1/2 day | before run | after each calibration run | ? |
| Drift chambers (inner + muon) Analog Memory Gains ADC Pedestals | ~ days | before run | after each calibration run | advanced data analysis |
| Drift chambers TDC Pedestals | ~ days | before run | after each calibration run | after each calibration run |
| ADCs and TDCs for TOF Counters | ~ days | before run | after each calibration run | advanced data analysis |
| Overall JADE Status (Parts not working or not there) | ? | after each calibration run | after each calibration run | ? |
| PETRA conditions | ~ minutes | continuously | continuously | continuously |

the events currently passing through the first data reduction phase were recorded. Old data would be erased from this file and new data added to it at frequent intervals. These data would be used by the program making the first data reduction to write epilogues to the events it accepted. The epilogues would contain reduced data, such as lead-glass energies in MeV, based upon the available calibration data. The old data taken off the cyclic buffer would be written for permanent storage on tape. The rationale for bringing in preliminary calibration data at the data reduction stage will become clear below.

The second disk file would describe the long term time development of each calibration parameter for the course of the whole experiment. The data on this file would be in the form of corrections, as they later become known, to the calibration data used at the first data reduction stage. The time development file could be organized as follows : The variation of each parameter could be fitted with a quadratic polynomial within each of a sufficiently large number of time bins. The sizes of the bins could be tailored to the actual rate of variation of each parameter, and bin bounderies could be adjusted to match abrupt discontinuities. Except where discontinuities occur, the quadratic fits in adjacent bins and their first derivatives could be made to match at the bin bounderies (spline fit). For each parameter and each time bin, four constants would then have to be stored on disk : the three constants specifying the quadratic polynomial, and one giving the time at which the bin in question ends. The entire history of calibration parameter number one would be written first on the disk, then the entire history of parameter two, etc. with time changing in the "inner loop" and parameter number in the "outer loop". A program requiring new calibration data would have to read the entire file, and rewind it if new data were required more than once per job.

This organization has the clear disadvantage of requiring a great deal of input/output. It also requires that when new time bins are added the entire file must be rewritten. This however must be done in any case when calibration data are retroactively revised. The enormous advantage of the proposed organization is that it requires writing to disk only as much information as is really needed. Parameters which are stable can be described in one time bin. In the alternative approach, that of making the parameter number the "inner loop" variable, time bin size would have to be set by the most rapidly varying parameter and there would be no way at all of matching abrupt discontinuities to time bin bounderies parameter



by parameter. If the most unstable parameter could be described by three constants in each of 100 time bins (probably a very optimistic hypothesis) 40 000 parameters would require 40 000 x 300 = 12 million constants, or 24 million bytes, or 1850 tracks. This would be an enormous, probably prohibitively large file, and it would be mainly occupied with repeated data on stable parameters.

Even with the suggested organization the time development file may tend to become very large, depending on how many of the calibration parameters vary with time and on how rapidly they vary. It may be extremely useful, in keeping this file to manageable size, if most of the time variation of most of the calibration parameters can already be factored out on the data tapes. Many parameters, e.g. leadglass photomultiplier gains, may be known as well or nearly as well as they will ever be known as soon as the standard calibration runs are taken. On the other hand it is certainly not possible to have smooth fits to the current calibration data ready for the dump job at the IBM. A logical time at which to bring in current calibration data may be when the first data reduction is done, when obvious background events are rejected. This first reduction will be done by batch jobs running asynchronously with, and some time behind, data acquisition. These circumstances make it clear why the procedure described above for bringing in preliminary calibration data at the first reduction phase is proposed. It also becomes clear, in view of the proposed spline fit scheme for the long term time development file, that the cyclic buffer file must already have smooth fits to the time variation of the calibration parameters for the interval at interest, and not just the most recent set of values. Otherwise the corrected pulse heights etc. written into the tape epilogues will have a "step function" time dependence and the long term time development file will have to have a new time bin for each step. On the other hand no great harm is done if a few of the parameters used in writing the epilogues are very bad, since the development file can correct just these. The corrections on this latter file could be additive rather than multiplicative in case, for example, a few gains were multiplied by zero due to dead pulser channels. Alternatively a program could easily be written to keep unreasonable values from being written to the cyclic buffer file and to replace them with default values.

The long term time development file should probably be split into several disk data sets, each corresponding to a major block of running time. A new time development data set could be started after long shutdowns, major changes in the apparatus, changes in the running mode of PETRA etc. Many analysis jobs will run on events within one such block of time and they will need less disk space and use less input/output time if the large time development file is split up.

In summary this note suggests the establishment of two disk files for handling calibration data : (1) a cyclic buffer containing fits to the current calibration data needed to write the tape epilogues described above. (2) a large file describing the long term time development of each calibration parameter. The parameters on this file will be for application to corrected data written into epilogues on the first and later generations of reduced data tapes. These corrected data will already contain the calibration parameters as best they were known when the first data reduction job ran. The time development disk file will then only have to parameterize residual corrections. If these are small, smoothly varying in time and applicable only to a small fraction of the 40 000 calibration parameters, this disk file can be manageably sized. The following chart shows how the suggested scheme would fit into the general flow of information from the NORD to the offline analysis jobs. This chart supplements and to some extent modifies the outline on p.2 of the note by W. Bartel, First draft for an IBM data organisation system, dated 26.4.1978.

L. O'Neill will begin to work on testing the feasibility of managing the disk files described above or some generally similar system. Members of the JADE-Collaboration are strongly encouraged to consider how hardware and programming problems would interact with such a system, and to make the appropriate suggestions.

One point bears reemphasis : The above discussion concerns only the administration of calibration data assumed to be available. No one in either the online or offline software group is working on programs to derive the calibration parameters from the raw test run data. The groups responsible for the various hardware systems should consider themselves responsible for providing such programs for whatever calibration parameters their system will require. This applies expressly to those groups, if any, which hope to run such programs on the NORD 10/50, as well as to those which prefer to run at the IBM. Clearly if a tape epilogue system such as the one outlined above is to be adopted, these programs must be ready for the start of the data run.