CHAPTER XI

DATA ACQUISITION AND PROCESSING

(should be read next to Chapter VI)

Data acquisition and data processing, Summary report of the Working Group
(D. Linglin, P. Charpentier, S. Cittolin, A. Clark, M. Demoulin, D. Jacobs, L. Mapelli, B. Nielsen, J.-P. Porte and A. Rothenberg)

Presented by D. Linglin

DATA ACQUISITION AND DATA PROCESSING

Working Group summary report

1) INTRODUCTION

The major task of the DACQ/processing group was to examine the possibility of Data Acquisition systems which could sustain very high transfer rates along with a maximum of high level trigger intelligence and flexibility. In addition, it had to investigate practical ways of processing the large amounts of data that will finally come out of detectors at the LHC (Juratron).

This chapter should be regarded as a continuation of the summary provided by the trigger group. There were extensive discussions with that group, and our group assumed as input their estimates of level-1 and level-2 trigger rates.
The reader can refer to their report for an overview of the various trigger signatures, possible problems, and expected rates.

Our approach was the following : first, we examined the trends in the present large detectors, especially those installed at colliders, and then we tried to define by extrapolation a scheme that could reasonably operate at higher energies, with higher event rates, sizes and complexity, in 10-15 years from now.

This short report is divided into 3 general parts :

Section 2 gives an overview of our proposal for a DACQ system.
Section 3 describes in more detail the various components of such a system : Level 2, Data Acquisition Bus, Level 3, recording media.
Section 4 deals shortly with the problems of data processing and analysis.

2) DACQ system - Overview of a proposal

"The UPSTREAM MOVE" :

In the past two decades, one has observed a general move to instal computer "intelligence" as early as possible in the data taking system, in parallel with growing detector complexity and event rates, and also in parallel with the rapid development of the electronics and computer industries. Not so long ago, recorded data were only looked at in off-line computer centres (eg. the so-called "Bicycle On-Line" or BOL activities). We observe now that more and more decision tasks, monitoring or calibration tasks, and even partial reconstruction tasks are performed locally, either near the detector or in the control room, on the on-line computer(s) of the experiment or in dedicated microprocessors. This "upstream move" will certainly continue in the years ahead, given its many advantages (for example avoiding BOL or large numbers of recorded tapes, improving response time, etc..).

FLEXIBILITY :

On the other hand, flexibility is an absolute must. Large 4π detectors at LHC, built to observe physics in a new energy domain, must be ready to adapt to many scenarios, whether it be the existence of surprising event topologies, the demand of increasing luminosity, etc..

Everyone knows the few basic triggers on "elementary" constituents that one plans to deal with. For example :

- Jet trigger (quark or gluon),
- Localised EM shower (electron or photon, depending upon charged track trigger),
- Missing Et (neutrino, ...),
- Penetrating particle (Muon),
- other triggers such as Total (transverse) Energy.

However other triggers may be needed (eg. new types of particles) and, above all, any type of combination of the above triggers is a-priori desirable. Hence implementation of trigger algorithms resulting from the observation of new topologies, or from an improving understanding of the data must be easy to do. A flexible trigger solution is to have the same software running off-line (when developing new algorithms) and on-line (in the high level trigger processors).

Also, the CPU capacity of the high level triggers must be flexible, to cope with increasing luminosity and/or level-2 trigger rates.

PROPOSAL :

Our proposal, which solves these trends reasonably well, is based on 3 main ideas :

- A high speed DACQ Bus (between the detector area and the control room, that is also between level-2 and level-3 triggers, a distance of ≈50 to 100m.)
- A single system with very large CPU capacity (up to 1000 processors, each one equivalent in speed and memory to a present large mainframe). This system is used for high level ("level 3") trigger decisions.
- Data storage and processing mainly at the experiment.

3) DACQ system - Description

Let us now describe in more detail a possible scheme (Figs 1 & 2) and its main consequences :

LEVEL 2 TRIGGER :

The so-called level-2 trigger has been described in the report of the trigger group. It can use digitized data from the calorimeters and the muon chambers, and also from central tracking chambers. Transition radiation detector signals seem to be too slow to be used at this level. However, at level 2, the events remain split into several branches and the full event information is not available to a single processor.

Talking about the level-2 DACQ system (the readout), it is proposed to hold there, in sequence, the digitizers (10-100 μs), FIFO multi event buffer memories (to derandomize the event arrival time), data formatting tasks, reduction tasks, and finally data concentration tasks to put together the many parallel event pieces into a small number of branches that form the DACQ bus.
Details of the digitization of the calorimeter cell pulse heights are described in the report of the trigger group, where they considered a level-2 trigger description based on calorimeter signals alone.

Canonical numbers usually quoted are a maximum rate of 1 KHz, for an event size of 1 Mbyte at the exit of level 2, to enter the DACQ bus.

DATA ACQUISITION BUS :

This bus should be able to sustain a rate of 1 Gbyte/sec (1KHz ⊗ 1Mbyte) over a distance of 50-100m, that is from the detector area to the control room. Presumably this

will only be feasible with several (between, say, 10 and 25) parallel branches and possibly with optical fibres.

For comparison, VME can reach ≈10 Mbytes/s.
FASTBUS can theoretically reach a maximum transfer rate of ≈20 to 40 Mbytes/s over short distances. The 1-Gbyte/s quoted above represents 25 parallel branches with a speed not so different from today's maximum value for Fastbus, apart from the larger distance involved and N-branch coordination problems.
Although the goal of 1 Gbyte/s does not seem unrealistic, we feel that Research and Development will be desired in this field.

LEVEL 3 TRIGGER :

The event information is still in N separate pieces when it arrives in the control room at the end of the DACQ bus. Only here, at level 3, does a single processor has access to the full event information, ready for recording.

It is proposed to install at this stage a "stack" made of a large (50→1000) number of processor units (as for the 3081E emulator of today), as shown on fig. 2. Each unit of this stack has a typical CPU speed of 10 Mflops with 10-16 Mbytes of central memory. This is roughly the speed and memory sizes of the large computers we are using today in our computer centres. We assume that the computer industry will be able to deliver such processors in a volume equivalent to one (or a few) CAMAC units, at a price of one to a few KSF. This means one rack could hold 5 crates with 5 to 20 processors each, plus its (optical disk) recording unit. A 1000 processor system would then be accomodated in 10 to 40 racks, at a price of a few MSF.

Each incoming event selects the first unit available and, depending upon the bits set by lower level triggers, starts one of the fast filter programs. This can be, for instance, a refined level-2 trigger, with the final calibration constants, or an elaborate jet finding algorithm, with for example an improved Et cut or a multi-jet effective mass selection.
If the event passes the test, one starts a second, more elaborate, selection program. Thanks to CPU power, this can even be the reconstruction of tracks from the central detector for additional rejection power.
Other selection programs may follow, increasingly elaborate as the remaining events decrease, with consequently more CPU time available per event.
Possibly, selected events can be fully reconstructed before being recorded.
With enough memory, each unit can hold all the filter and reconstruction programs and play the role of "several-in-one" high level triggers.

Moreover, the scheme allows :
- a flexible number of microprocessor units, to match increasing luminosity, level 2 rates or decreasing costs per unit,
- an easy implementation of new algorithms (the development of which depends mainly on the off-line analysis of previous data). It is very important that algorithms run with the most up-to-date information and calibration constants.

RECORDING :

The best choice foreseen as a recording medium seems to be the optical disk (although magtapes may have not yet given their last word). 12" optical disks are now arriving on the market, with reasonable prices, although one must admit that none has been delivered yet to customers. Advertisements already propose 2-Gbytes capacity disks (1 Gbyte per side), that is more than 10 6250bpi tapes, with typical writing speeds of 0.4 Mbytes/s, only limited by laser power. Capacity and recording speeds should increase, and prices should decrease, in the next few years. With rather cheap disk systems available in 10 years from now, one can choose between a good "juke-box", or a disk pack system, or 5-20 independent disk drives as shown on fig. 2.

Although it would be possible to record rates as high as 10-50 Hz (eg. with 20 independent disk drives), we feel that one should aim at a standard rate of ≈1 Hz, with

peak values of ≈5 Hz. Otherwise, the high level trigger programs in the stack would not be fully efficient.
For an average 1 Hz trigger rate of events, with a typical 1 Mbyte length, we would need to change a 1 Gbyte side of a disk every 15-20'.

4) DATA PROCESSING AND ANALYSIS

PROCESSING :

Since the full experimental data base is available in the control room, and having in addition a large CPU capacity there, the reconstruction and data reduction tasks should be made with the multiprocessor stack. This can be performed off-line or even on-line when one has enough confidence in the programs. This would ease all the administrative aspects such as bookkeeping, calibration constant base, etc.., and also the present tape handling bottleneck, which presently afflict large experiments.
Also, the random access facility of the optical disks must ease many of the data processing activities.

With the large filtering power of the stack, every recorded event should be interesting enough for further analysis. Each 1000-hours period of useful data taking should then yield, for 1 Hz average recording speed, 3.6 10^6 events of 1 Mbyte each, stored for instance on 360 10-Gbyte disks. Assuming 20 to 50" CPU time to fully process an event in any given processor of the stack, this means between 20 and 1000 hours for 50→1000 processor units. Such computing power allows for several complete processings of the same events, whenever wanted.

ANALYSIS :

Analysis (and analysis development) should be done on private workstations or on large mainframes because of the niceties which are not available with the multiprocessor stack.
To provide data information to any external laboratory, disks can be copied and shipped.
Individual events can also flow from the control room through inter-computer networks.
However, for bulk analysis, the best scheme would be to connect private workstations to the on-line computer ("supervisor") and from there, use the stack and the data base.

5) CONCLUSION

- The above scheme follows the present trends and is flexible enough to adapt to many scenarios.
- The computer industry should deliver by LHC turn-on time, at a reasonable cost, all the elements.
- Some R&D however might be needed on fast data bus development.

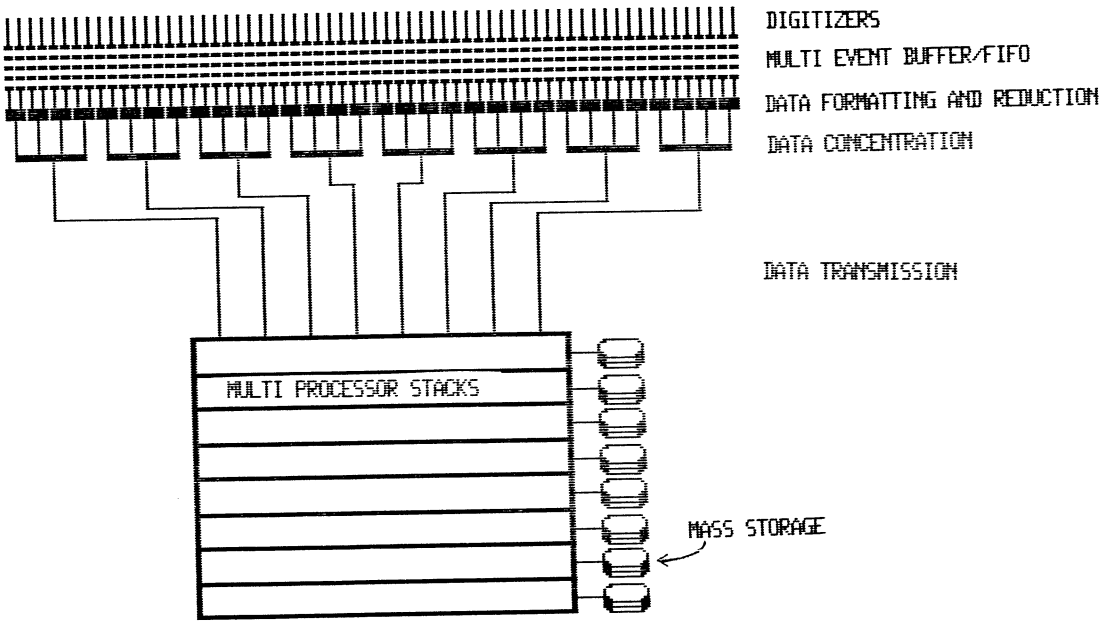
DIGITIZERS
MULTI EVENT BUFFER/FIFO
DATA FORMATTING AND REDUCTION
DATA CONCENTRATION
DATA TRANSMISSION
MULTI PROCESSOR STACKS
MASS STORAGE

Fig. 1

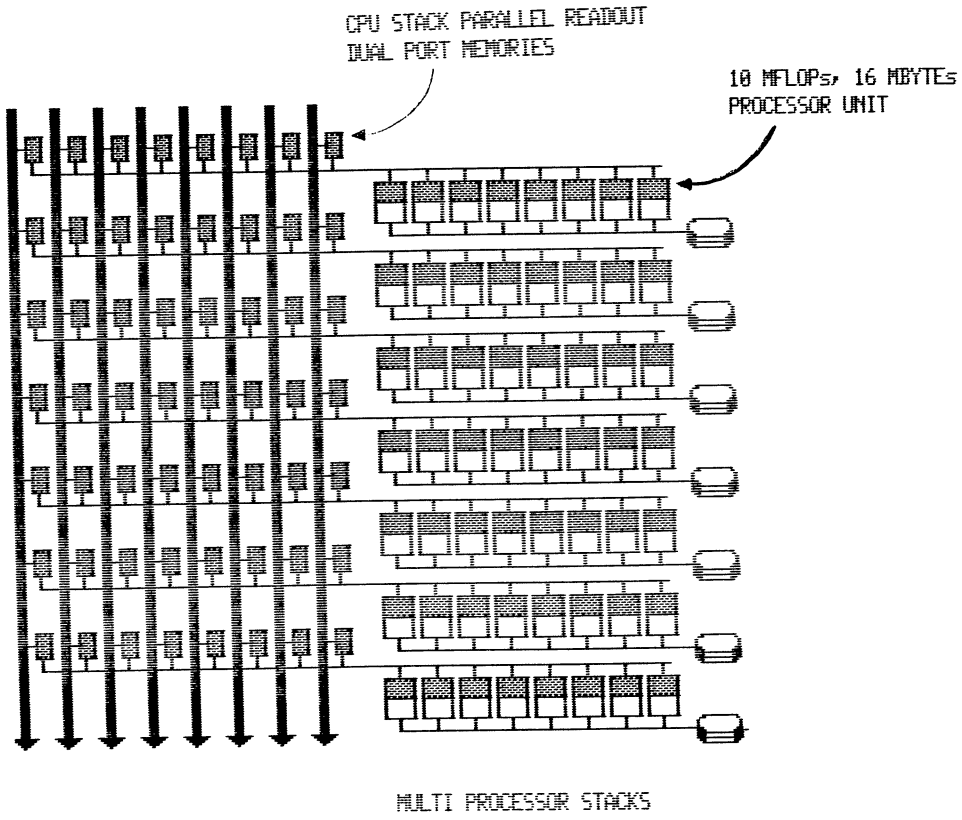
CPU STACK PARALLEL READOUT
DUAL PORT MEMORIES
10 MFLOPs, 16 MBYTEs
PROCESSOR UNIT
MULTI PROCESSOR STACKS

Fig. 2