THE ALICE DATA ACQUISITION SYSTEM

J.NELSON, O.VILLALOBOS BAILLIE

University of Birmingham

E.DENES, A.HOLBA, G.RUBIN, L.SZEDREI

KFKI-RMKI, Budapest

T.KISS, Z.MEGGYESI

Technical University of Budapest

W.BOZZOLI, R.DIVIA, G.HARANGOZO, R.McLAREN, H.G.RITTER, E.VAN DER BIJ, P.VANDE VYVRE, A.VASCOTTO

CERN, Geneva

R.BROCKMANN, B.KOLB, M.PURSCHKE

GSI, Darmstadt

M.BELDISHEVSKI, M.BELLATO, G.MARON

INFN, Legnaro

B.KVAMME, B.SKAALI, B.WU

University of Oslo

H.BEKER

Universita "La Sapienza" and Sezione INFN, Roma

(The ALICE Collaboration)

ALICE (A Large Ion Collider Experiment) is the future LHC experiment dedicated to study the physics of strongly interacting matters. The trigger and data-acquisition system of ALICE will have to process different kinds of events: short runs of heavy-ions events resulting in huge volume of data and long runs with pp events generating less data. These two types of run result in different, and sometimes conflicting requirements and require that the computing system is flexible enough to run in different modes of operation and to switch easily from one mode to another. This paper presents the initial system design, the results of the simulation work, the prototyping which will be started. The current situation of the permanent storage industry is also mentioned due to the particular importance of this technology for the ALICE data acquisition system.

1

1 Introduction

ALICE (A Large Ion Collider Experiment) [1] is the future experiment dedicated to study the physics of strongly interacting matters. This study will be made with the nucleus-nucleus interactions at LHC energies. The experiment will be designed to cope with the highest particle multiplicities anticipated for Pb-Pb reactions. Collisions of lower mass ions are a mean of varying the energy density. Running with protons is foreseen as part of the experimental program to provide reference data. These different running modes produce very different requirements for the DAQ system. We will review these requirements and their influence on the design of the DAQ system.

2 Experiment Requirements

The trigger and data-acquisition system of ALICE will have to manage three different kind of processing:

- The first consists in managing the acquisition and reduction of the huge volume of data produced in Pb-Pb interactions. The experiment will produce a low rate (50 Hz) of very large events (between 25 and 50 MByte) (see Table 1) resulting in an unprecedented volume of data, of a maximum of 2.5 GByte/s. A large computing facility will preprocess the events with the goal of reducing the volume of data recorded onto permanent storage to a maximum of 1.25 GByte/s. The heavy-ion run will last 4 weeks per year and will require a maximum of bandwidth to permanent storage.

Table 1: Event sizes of the sub-detectors for Pb-Pb events

Detector	Subdector or Option	Number of channels	Minimum event size (MByte)	Maximum event size (MByte)
ITS	Si Pixel	6.6×10^6	0.057	0.057
	Si Drift	485×10^3	0.029	0.290
	Si Strips	2.8×10^6	0.142	0.142
	MSGC	1.4×10^6	0.175	0.175
PID	RICH	2.0×10^6		0.400
	TOF	170×10^3	0.017	
TPC		500×10^3	28.000	45.000
Others		$\times 10^6$	$\approx 1.$	$\approx 1.$
Total		13.9×10^6	$\approx 30.$	$\approx 47.$

- The second concerns an extension of the physics covered by ALICE to include the measurement of high-mass lepton pairs. This would result in an additional class of events with higher event rate being superimposed on the previous one. For these events, the TPC which is producing 90 % of the data for Pb-Pb interactions, will not be read out but several other detectors will be readout

2

Time Projection Chamber   Inner Tracking System   Particle Identification   Electro Magnetic Calorimeter   Forward Multiplicity Counters Zero-Degree Calorimeters

FER   FER   FER   FER   FER   FER

DDL

2-4 KHz.   TDL

L1 Trigger

FEDC   FEDC   FEDC   FEDC   FEDC   FEDC

LDC   LDC   LDC   LDC   LDC   LDC

EBL:

2500 MBytes/sec. Pb-Pb run
20 MBytes/sec.   p-p run

Switch

EDM ← L2 Trigger

50 Hz+1000Hz Pb-Pb run
1000 Hz   p-p run

GDC   GDC   GDC   GDC

FER: Front-End Readout   RDL Readout Data Link
FEDC: Front-End Digital Crate   DDL Detector Data Link
LDC: Local Data Concentrator   EBL Event Buidling Link
GDC: Global Data Collector   TDL Trigger Distribution Link
EDM: Event Destination Manager

1250 MBytes/sec. Heavy Ion
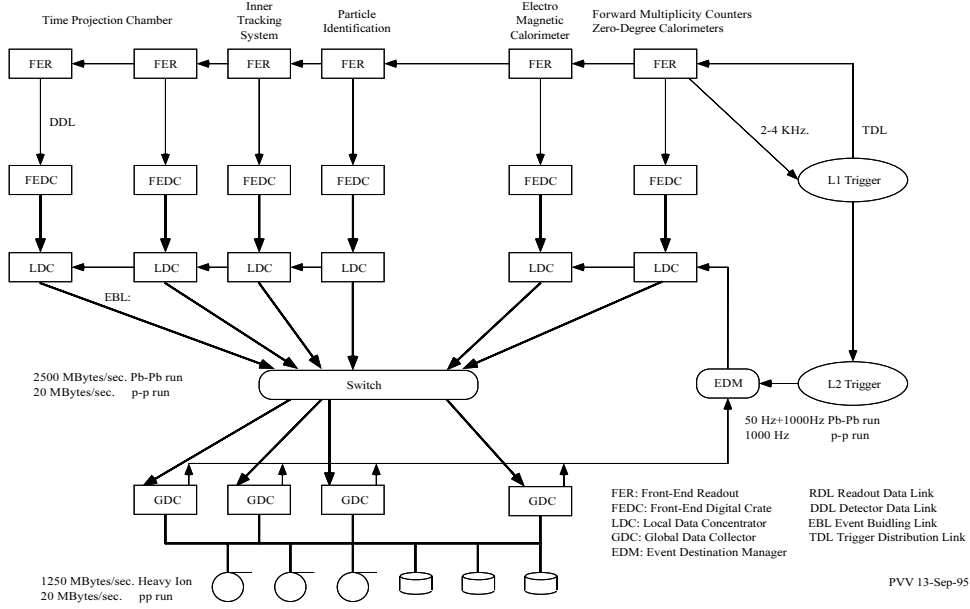20 MBytes/sec.   pp run

PVV 13-Sep-95

Figure 1: Full Daq architecture

for both types of trigger. The events should be stored on separate permanent storage volumes depending on the corresponding trigger type.

- Finally, the ALICE experiment will take data with heavy-ion beams for only 4 weeks in a year. The rest of the running period will be used to acquire data produced in pp interactions, which result in two orders of magnitude less data (20 MByte/s). During this period of the year, the computing infrastructure will be used partly for data acquisition and partly for data analysis.

These different, and sometimes conflicting requirements require the computing system to be flexible and easily reconfigured.

2.1 DAQ System

2.2 Global architecture

The functions of the DAQ and control system are to transport the data from the front-end electronics to the computing room, to assemble data into full events, to format and condense the information and to record it onto permanent storage. The DAQ architecture based on a network of high-speed links linking all the data sources and the data destinations through a switch seems the most promising (see Fig. 1). This architecture provides the required flexibility and scalability required to run in very different modes. Several technologies (ATM, Fibre Channel, HiPPI, SCI) are or will soon be available to implement a switching network to connect the data sources and the data destinations with an aggregate bandwidth of up to 2.5

3

GByte/s. The event building protocol works as follows. For each event selected, the trigger sends a message to the Event Destination Manager (EDM). The EDM selects the Global Data Collector (GDC) which will build the next event and sends then a message containing the trigger information and the destination address to all the Local Data Concentrators (LDC). Every LDC performs the readout of the data in the same Front-End Digital Crate (FEDC), formats these data and sends them to the destination GDC selected for the current event. The complete event is transferred fragment by fragment to the GDC. This creates some contention between all the LDC trying to access the same GDC destination. The event building and distribution protocol is in charge of optimizing the traffic and of reducing these contentions.

2.3 Simulation

A flexible simulation program has been designed to simulate the current requirements and to evaluate the relative merits of the possible architectures proposed for the data acquisition system and the different technologies existing to implement it [2]. This program has been used extensively to simulate the proposed architecture with the current technologies [3]. One of the results of the simulation has been to determine the performances needed by the individual components of the DAQ system to match the global system requirements.

2.4 Prototyping

We will verify during the prototyping if the performances defined during the simulation can be achieved with commercially available components. We will start with the heart of the system which is the event-building and distribution system [4,5]. A prototype of the ALICE event-building system will be assembled using VME processors and workstations interconnected via a switch.

3 Permanent Data Storage

The problem of the permanent data storage for ALICE is that it must combined:

- very high bandwidth per device to avoid an unacceptable duplication of devices to achieve the desired bandwidth of around 1.25 GByte/s.

- and very high capacity of data per volume to avoid a multiplication of the number of tapes needed to store the 1000 TByte written during one run of heavy-ion.

These two characteristics combined limit for the time being our choice to a limited set of technologies and devices. The current tape storage market can be divided in 3 categories [6]. Some existing technologies of each of these categoreis are reviewed in the Table 1. Technical solutions exist today in the high-end segment of the market. However, they are still very expensive to achieve the desired bandwidth and storage capacity.

New technologies are emerging now in the laboratories [7,8,9]. It is however too soon to predict a time for the products to appear on the market.

Table 2: Storage Technologies Summary

Technology	Band. MB/s	Device Cost SF/D	Band. Cost SF/MB/s	Cartr. Unc. GB/Car.	Cartr. Cost SF/Car.	Media Cost SF/GB
Low End						
Exabyte 8500	0.5	3.4	6.0	5.0	10.8	2.2
DEC DLT 4000	1.5	9.0	6.0	20.0	40.0	2.0
Midrange						
IBM MagStar	9.0	50	5.5	10.0	65	6.5
STK Redwood	11.0			25.0		
High-end						
SONY	32.0	300	9.4	100.0	250	2.5
Interferometrics	256.0	420	1.7	800.0	2880	3.6

4 Conclusion

We have presented the needs of the ALICE experiment. A general architecture of a DAQ system has been presented. The components needed to implement the event building and distribution system should be available from the industry. This will be verified by carrying on the simulation and by building a prototype of the final system. The storage is much more problematic, in particular from a financial point of view. The evolution of the magnetic tape and the possible emergence of new technologies will be followed carefully.

References

1. N.Antoniou et al., ALICE Letter of Intent, 1st March 1993, CERN/LHCC 93-16.
2. B. Wu et al. ALSIM, Reference Manual and User's Manual for ALICE DAQ Modeling and Simulation, ALICE Internal Note 94-34.
3. Simulation of the ALICE DAQ system, ALICE Internal Note 95-26.
4. E.Denes et al., ALICE Data Acquisition System - Program of R&D work, ALICE Internal Note 95-03.
5. G.Harangozo and G.Rubin, Event Building schemes for Fibre Channel switching networks.
6. R.I. Patel et al., The emerging role of the flexible media, Proceedings of the IEEE, Apr 93, Vol 81, No4.
7. J.F. Heanue et al., Volume Holographic Storage and Retrieval of Digital Data Science, Science Vol. 265 Aug 1994 pp749-752.
8. J.P. Castera and J.C. Lehureau, High-Data-Rate, High-Density Magnetic Tape Recorder using Matrix Writing Head and Kerr reading head, Proceedings of the IEEE Symposium on Mass Storage System Annecy, France (1994).
9. Technical Directions for the 21st century, Chapter 2 Storage Technology Trends, NASA (March 1993).