Data compaction system for E821 (g-2) experiment

A.G. Chertovskikh [a,1], P.Cushman [b], A.V.Klishin [a] Yu.I.Merzlyakov [a], I.V.Sorokin [a], Ch.Timmermans [b],

[a] *Budker Institute of Nuclear Physics, Novosibirsk, 630090, Russia*

[b] *University of Minnesota, Minneapolis, MN55455, USA*

The E821 (g-2) experiment is due to start in 1997. It's second phase, including muon injection will be producing a huge flow of data demanding on-line lossless data compaction before storing it on magnetic media. A hardware/software data compaction system is described including software algorithms and parallel data handling multicomputer VME system.

Key words: Algorithms; communication; data acquisition systems; event building; prallelization.

1 Introduction

The goal of AGS E821 experiment [1] is a precision measurement of the muon (g-2) value to 0.35 ppm which would be an improvement by a factor of 20 compared to the present experimental accuracy. For this purpose a unique, high precision superconducting storage ring for 3 GeV/c muons has been created at the Brookhaven National Laboratory, USA. 24 detector stations distributed at equal intervals around the ring are used for fixing muon decay electrons being a main source of experimental information. Data from the detector's stations are taken by the E821 data acquisition system (DAQ) by means of two main types of electronic devices for fixing time of arrival of an electron and a shape of a pulse it produces. The devices are called MTDC (a 27-channel multi-hit time-to-digital converter with at least count of 1.25 ns) [2], and WFD (waveform digitizer) consisting of four channels of 200 MHz, 8 bit, flash A/D converter [3].

The task of the E821 data compression (compaction) comes from the fact of a future heavy data flow during a second phase of the experiment called muon injection evaluated as 5.4 MB/s. According to estimations, a data compression factor of 2 to 3 will be demanded to meet requirements of recording data to magnetic media.

2 Hardware

The DAQ consists of 6 front end VME crates having 2 MTDC and 2 WFD slave units each, with the MVME167 (or MXI-controller) master. It is assumed the master extracts data from the units and passes it to VME transputer modules TTVME [4] to form a data structures in a specific compacted format. The transputer units used are in fact prototypes to be probably upgraded in the future. The data structures via a fast VME interconnect (MXI-bus) are collected in a central VME crate to be recorded to magnetic media.

The compaction system (see Fig. 1) can be divided into several steps or stages of data processing:

[1] E-mail: A.G.Chertovskikh@inp.nsk.su, Fax:+7(3832)352163

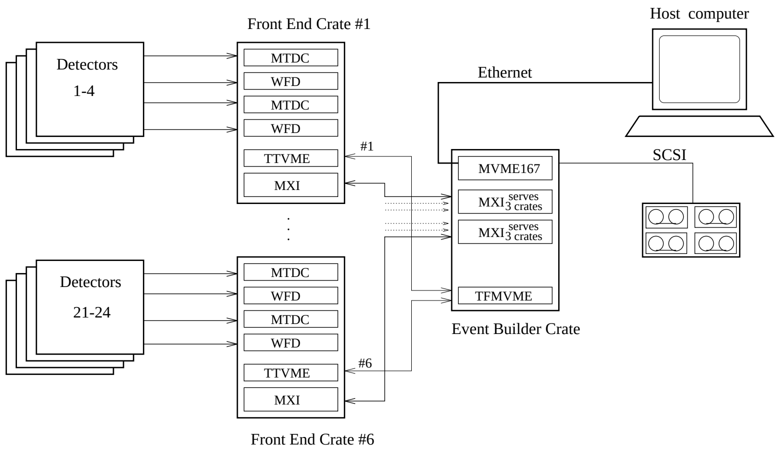

Fig. 1. The Data Aquisition System Architecture

(i) The readout step which is responsible for collecting information from the MTDC-WFD system and passing it to the transputer network.
(ii) Data processing by the transputer network which receives the data from local VME-masters and compresses information using the described below algorithms.
(iii) The VME master which records compressed events to magnetic medium and passes some events to remote computers for real-time analysis.

The TFMVME is a control and configure processor module for VME - based transputer farm. It has A16 D8(OE) slave interface and two levels interrupter from VME bus. The module has T425 25Mhz transputer [4] with 2MB RAM and C004 transputer link switch for configuration and programming transputer network. Transputer network may consist of TTVME modules on the same VME crate or a combination of transputer modules from other systems (VME,CAMAC, PC, etc.). All the transputer modules are packaged in 6U high Eurocard format and requires one standard VME slot.

3 MTDC and WFD data compaction algorithms

There can be two different approaches to reducing experimental data size: (1) online data handling resulting in getting rid of 'uninteresting' information not required in a future offline processing, and (2) lossless data compaction. Typically the first way seems hardly acceptable at least during an initial period of experiment, as the risk of filtering out data that can be considered useful later is high enough. Such an approach was proposed in [5], and de-facto rejected. General purpose methods of lossless data compaction are known for years and used quite widely [6]. Using these methods for compressing simulated (g-2) data gave a compression factor of only 1.3.

Data-dependent compression methods in case of experimental data seem more promising. The main idea is the same as in general purpose methods — getting rid of redundant information, but this redundancy is specific to the way the MTDC and WFD work as they cannot produce highly optimized and compact data at frequencies they operate. Thus, compaction algorithms were developed to reduce data size without losses. Besides, some hardware support for front end VME crate masters to unload them from compaction computations was considered. All results were obtained using simulated (g-2) data being supposedly closer to future E821 data than test beam information we have had.

A basic approach to the compaction is extracting data of the same type from the common data stream, compressing and putting them back to the stream. The MTDC produces uniform 32-bit words, and the WFD generates data of three types: 32-bit ADC pulse sampling results (4 value bytes), 16-bit time values for this ADC data and 4-bit ADC discriminator values. Thus, there are four different compaction algorithms corresponding to these types.

The MTDC data stream consists of groups of words describing an arriving electron. The compression is achieved by means of transformation of the stream to optimal record structures eliminating redundant information. Each record corresponds to the group. The record keeps only one full time value, all other time values are converted to short offsets. All pair of words describing leading and trailing edges of a pulse are transformed to just a channel number and a short pulse width. Error field and special ones are rejected if they contain no significant values.

Numerous consistency checks are made 'on the fly' allowing to improve the quality of data. Unbalanced pairs of words, too wide pulses, MTDC hardware errors and some more situations are marked, and intact input stream group is passed to output stream as a special record type. According to conducted measurements, compression factor for the MTDC data is 2.27 to 2.42 depending on input group lengths in different files.

The WFD ADC values describe a pulse shape produced by an arriving electron. The main idea was keeping deviations of these ADC values from some template assuming they would be smaller than the values themselves. Choosing such a template led to using the pulse as a template for itself. Deviations are computed as differences between two sequential ADC values. As the difference can be either negative or positive, a sign bit is also kept. Non-zero bits of the difference determine a bit width of the deviation. Consequent deviations not exceeding predefined number of bits are combined in records.

The presented WFD ADC data compression method efficiency depends on a 'quality' of pulses. Smooth pulses give good compression factors, noisy ones with a lot of distortions can lead to even 'data expansion' instead of compression. For the simulated data containing mostly 'good' pulses the compression factor is 1.43 to 1.56.

The WFD time values are just 16-bit counters, one for each four ADC samples. As one pulse is described by several 'fours' of samples, time values are just an initial value monotonously increased by 4 for each 'four' of samples. Quite naturally, time values are encoded to records corresponding to a pulse each containing initial time and number of eliminated consequent time counters. Records are packed to the output stream similarly to ADC ones. The compression factor for this type of data is 2.38 to 2.71.

The WFD discriminator values describe crossing ADC thresholds with leading and trailing edges of arriving electron pulses. The values are not changed inside the pulse, i.e. between leading and trailing edges. This allows to use a well-known 'run-length' compression method to create output records consisting of a repetitive discriminator value and a number of such values. The data are compressed by 1.54 to 1.60.

4 Conclusion

Summing up all presented above, one can compute an integral compression factor taking into account that the MTDC data, WFD ADC samples, WFD times, and WFD discriminators are distributed in the simulated data files as 2:2:1:1 approximately. This would give the factor slightly less than 2. Thus, the E821 lossless data compaction system is developed including hardware and software elements.

References

[1] *A New Precision Measurement of the Muon (g-2) value of the level of 0.35 ppm.* BNL (Design Report),1994

[2] E. Hazen and G. Varner, *A User's Guide to the g-2 Multi-hit TDC (MTDC).* Boston University. Document Release 1.0 - for Rev A PCBs. January 25, 1995.

[3] William E. Earle, *VMEbus Specification for the MACRO Waveform Digitizer (WFD).* Boston University. February 12, 1993.

[4] G.A.Aksenov et al., *Transputer based data acquisition system for the CMD-2 detector.* Nucl.Instr.Meth. A 379 (1996) 550-551.

[5] Bill Worstell, *Correlation and Reduction of WFD and MTDC Data – a Proposed Architecture.* Boston University, E821, Detector and DAQ working meetings, October 1, 1993.

[6] Steve Apiki, *Lossless Data Compression.* BYTE, March 1991,pp. 309-314, 386-387.