DOE/ER/03130--9

DOE/ER/03130--9

DE90 013330

A DATA ACQUISITION ARCHITECTURE FOR THE SSC*

Richard Partridge
Brown University, Providence, RI 02912

ABSTRACT

An SSC data acquisition architecture applicable to high-p_T detectors is described. The architecture is based upon a small set of design principles that were chosen to simplify communication between data acquisition elements while providing the required level of flexibility and performance. The architecture features an integrated system for data collection, event building, and communication with a large processing farm. The interface to the front end electronics system is also discussed. A set of design parameters is given for a data acquisition system that should meet the needs of high-p_T detectors at the SSC.

1. INTRODUCTION

The motivation for this work is to lay out a simple model for an SSC data acquisition architecture that, while needing further details, provides an integrated and consistent approach to the problem of data acquisition at the SSC. No attempt has been made to survey the wide variety of architectural options available; rather, a specific data acquisition architecture is developed based on a particular set of goals, assumptions, and design principles.

2. GOALS

The overall goal is to design a robust data acquisition architecture capable of meeting the needs of high-p_T detectors at the SSC. More specifically, the architecture must possess:

- Sufficient bandwidth to handle the expected data rate with minimal deadtime and substantial margin for increasing luminosity and the hardware trigger rates.
- The ability to operate in inaccessible areas having high levels of radiation with low power dissipation.
- A scalable architecture that can be used for small-scale test beam and electronic system tests as well as the full experiment.
- Facilities for testing and partitioning the data acquisition system to simplify commissioning of the detector.
- Built in redundancy, fault detection, and a high degree of fault tolerance.
- Well defined upgrade paths to meet evolving physics and detector requirements.

* Contributed Paper to Computing in High Energy Physics '90.

3. ASSUMPTIONS

In designing a data acquisition architecture, it is necessary to make a variety of assumptions regarding data rates, event size, trigger architecture, front end electronics, and event processing requirements. These assumptions are summarized in the table below.

Data Acquisition Assumptions	
Interaction Rate	10^8 Hz
Level 1 Trigger Rate	$10^4 - 10^5$ Hz
Level 2 Trigger Rate	$10^2 - 10^4$ Hz
Level 3 Trigger Rate	$10 - 100$ Hz
Number of Front End IC's	10^5
Number of Level 3 Processor Nodes	5000
Level 3 Processing Power	1 TIPS
Average Event Size	1 MB
Data Acquisition Capacity	10 GB/s

A three level trigger system is assumed. A synchronous Level 1 trigger reduces the event rate to $10^4 - 10^5$ Hz with minimal latency. The Level 2 trigger provides an additional rejection factor of $10 - 100$ rejection with $\approx 50\ \mu s$ latency. Following a Level 2 trigger, the data acquisition system must move the event from the $\approx 10^5$ front ends to one or more computers in the Level 3 processing farm where final event selection is made. Level 3 event selection will require the sort of event reconstruction algorithms normally associated with "offline" processing to reduce the rate of accepted events to $10 - 100$ Hz. The amount of processing time spent per event depends in part on the rejection achieved in Levels 1 and 2, but $100 - 1000$ Vax-seconds is probably a reasonable estimate. A 1 TIPS (10^6 MIPS) processing farm composed of 5000 processors, each providing 200 MIPS, would then be capable of handling $10^3 - 10^4$ events/second. It is likely that the Level 3 processors will be the dominant cost in the data acquisition system, limiting the capacity of the data acquisition system to $\approx 10^4$ Hz. This figure is in good agreement with the worst case Level 2 trigger rate given above and will be taken as the maximum event rate. Thus, a 10 GB/s data acquisition capacity is required assuming an average event size of 1 MB.

4. DESIGN PRINCIPLES

Design of the data acquisition architecture was guided by a small set of design principles. The selection of this particular set of design principles was largely

aimed at simplifying communication between the various data acquisition components while providing the robust and flexible data acquisition environment that will be required. A short summary of each design principle is given below.

- Only one Level 1 trigger will occur within the resolving time of the slowest detector element. A small amount of deadtime is preferable to the complications involved in untangling two triggers that contain some of the same data.
- The number of control and readout signals between the data acquisition and front end electronics will be minimized. Eliminating unnecessary signals will simplify front end interconnection, reduce noise, and lower power dissipation.
- Each front end and data collection node will output one and only one data packet for each hardware trigger. The receiver of the data packet will then be able to detect when all data for an event is collected by keeping track of the data packets it receives.
- The data for an event will have a fixed sequential ordering. A fixed data order will aid data acquisition diagnostics and may be useful to pattern recognition algorithms that find calorimeter clusters and track segments.
- An event can be routed to any set of available processors in the Level 3 farm. Providing a flexible processing environment will maximize the opportunity for implementing sophisticated event selection algorithms in the Level 3 trigger and allow events to be routed to specialized processing nodes for calibration, monitoring, and software testing.

The above set of design principles is certainly not unique, nor is it obviously the best choice for the SSC. There is also room for considerable architectural variation consistent with these principles. Design principles and architectural variations must be examined within the context of a particular set of detector requirements and the results of computer modeling and simulation studies. The goal of this paper is to provide a specific example of design principles and data acquisition architecture that can serve as a starting point for a more complete study of data acquisition design.

5. FRONT END ELECTRONICS INTERFACE

Past experiments have relied on standardized buses (CAMAC, FASTBUS, VME, etc.) to provide the interface between the electronics and data acquisition systems. Adopting such standards simplifies the integration of the electronics and data acquisition systems and provides a well defined specification for communication between the systems. With the highly integrated electronics systems envisioned for the SSC, front end electronics chips mounted on the detector must interface directly to the data acquisition system. In effect, the role previously played by standardized buses must now be replaced by a standardized way of interfacing to front end chips.

There are many possible ways to interface with the front end electronics. Rather than explore all possibilities, a minimalist approach is taken that reduces the number of connections between the front end and data collection systems to the absolute minimum. This approach is driven by the need to minimize the power required for output drivers, reduce the pin count on front end chips, and simplify the interconnection with the front end. Only three signals are needed to provide a fully functional communication path:

- Level 1 accept signal. This is an NRZ indication to accept or reject an event in the synchronous Level 1 buffers. Maintaining the same state indicates acceptance, a change of state indicates rejection of the event. This signal is used to set the timing and synchronization of the entire detector, with the 60 MHz crossing clock derived from this signal using a phase-locked loop. Systems that require very precise timing, such as drift chamber readout, may require a separate crossing clock signal if adequate phase stability cannot be maintained by the phase-lock loop.

- Serial data input. Message packets are delivered to the chip using this line. Each chip must be able to decode at least 3 different message packets: Level 2 accept/reject, initiate data readout, and synchronize on next Level 1 trigger. Additional types of message packets may be useful for downloading calibration constants and implementing detector specific functions. Messages to the chip must be kept short to minimize Level 2 latency. A 24 bit message length is probably adequate and would add no more than 4 μs to the Level 2 latency assuming a 15 Mbit/s data rate.

- Serial data output. Data packets are sent on this line in response to event readout requests and query requests for detector specific information. Event data packets will include a short header, the data for an event (if any), and error information. This line operates at speeds up to 60 Mbit/s when driven by the crossing clock, giving an output bandwidth sufficient for front end chips with up to $\approx$ 200 bytes of output data. Chips with low data rates can divide the crossing clock and operate at lower frequency.

While the goal of minimizing the number of signals is clearly met, a number of other advantages result from this approach. The serial input path can identify which Level 1 trigger is being referenced in Level 2 accept/reject and initiate data readout messages. This makes it easy for a chip to identify and recover from loss of synchronization and buffer overflow conditions that would otherwise be difficult to handle. Resynchronization is accomplished by sending all systems a synchronize message followed by a Level 1 trigger. The output serial line is not shared with other front end chips and provides a higher aggregate bandwidth into the data collection chip than would be achieved with a parallel bus shared by many front ends. Avoiding shared data buses also helps prevent a single front end failure from taking a major segment of the detector down with it. Finally, there is no need to download a chip ID to each front end since the data collection chip can provide the required addressing functions

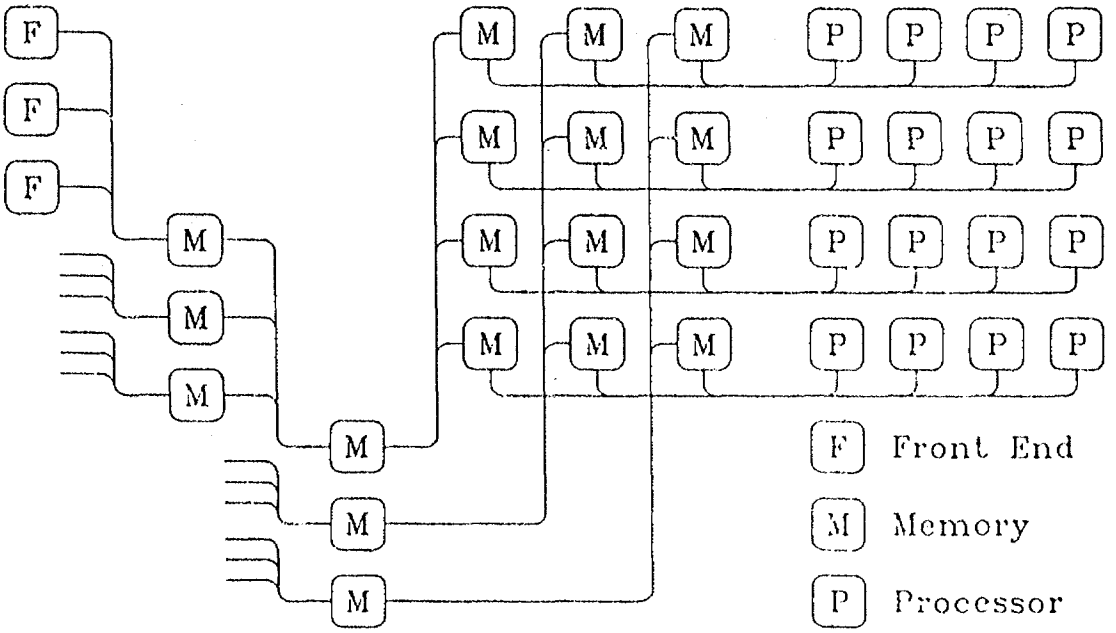

Figure 1. Data Acquisition Architecture.

6. AN INTEGRATED DATA ACQUISITION ARCHITECTURE

The traditional distinction between data collection and event building is somewhat arbitrary since both have the same goal: collecting data from various sources, organizing the data into complete events, and passing the data on to processing elements for further data selection and compression. This section outlines an integrated data acquisition architecture that uses essentially the same elements for both data collection and event building tasks.

The data acquisition architecture is strongly influenced by the design principle establishing the output of each data acquisition element to be a single data packet for each event. An elegant way to implement this requirement is to first collect all the incoming data packets for a given event using the input port of a dual port memory. When the data acquisition element has received packets from each input element, it forms a single outgoing data packet that includes all of the packets received. At this point, the element is free to arbitrate for access to the output bus and send the data on to the next stage using the output port of the dual port memory. The element also sends a request for the next event (if one is pending) to its data sources to begin the process of collecting the next event.

Figure 1 shows a block diagram of a data acquisition architecture that makes extensive use of the dual port memories described above to provide both data collection and event building functions. As is clear from the figure, the difference between data collection and event building elements is in the pattern of

connection, not in their function. During the data collection phase, multiple input streams are concentrated onto a single output stream. At the point where further concentration of data is not possible due to the limited bandwidth of the connections, a parallel connection scheme is employed that uses multiple output streams to connect to the Level 3 processor farm.

A high performance data bus is used for transferring data between the event builder and the processor farm. Industry support for the data bus is important to ensure a wide selection of processing elements that can be directly connected to the event builder. Processors in the farm can then access the event data as part of their address space, with the processor's data cache acting to minimize multiple accesses of the same piece of data. This approach provides seamless integration of the event builder and processor farm and allows a great deal of flexibility in implementing multiprocessing strategies that use several processors to analyze an event.

The data acquisition architecture has a number of desirable features:

- No global synchronization of the data collection elements is required. Each data collection element acts to move data toward the processor farm as fast as possible without waiting for neighboring elements to finish with an event.
- The data collection system provides automatic buffering of the data flow and allows a fixed data order to be maintained.
- Redundancy is easily provided by incorporating a second, redundant data collection system. Note that the event builder is essentially a highly redundant data collection system.
- An event can be routed to an arbitrary set of processing nodes. The event stored in the event builder becomes part of the processor farm address space, eliminating the need to move the data and allowing multiprocessing of an event to occur.
- The architecture is scalable to a wide variety of data rates. Upgrades can be performed by adding additional data links, processor buses, and processing nodes.
- Partitioning is easily accomplished by grouping sets of data links with sets of processor buses. Testing can be accomplished by downloading test data into the data collection elements.

7. SCALING LAWS AND DESIGN PARAMETERS

It is useful to estimate some of the parameters in the architecture needed to satisfy the design goals and assumptions described above. A variety of scaling laws are also developed for the architecture.

The data collection system is relatively straight forward. Assuming that 10 front end chips are connected to each data collection node, $\approx 10^4$ data collection nodes are required. If a front end chip can output a maximum of 200 bytes of data

for an event, each data collection node requires a minimum of 2 KB of memory storage. Integration of the front end data collection node, including memory and control circuits, onto a single chip using a rad-hard process is one of the major R&D efforts required in this architecture. Further levels of data collection require $\approx 10^3$ additional data collection nodes. These higher levels of data collection are likely to require somewhat different control circuitry, use off-chip memory for data storage, and be able to buffer several events to maintain a high average bandwidth. The additional control circuitry required for the event builder will probably require a third version of the control chip.

The transition between data collection nodes and event builder nodes occurs when the limited bandwidth of the data links prevents further merging of data from different detector elements onto a single data link. At this point, a parallel approach is taken so that the bandwidth is divided among a number of parallel data streams. The most important parameters setting the scale of the event builder are the speeds of the data links, both from the data collection system and to the processing farm.

The aggregate bandwidth of N_{FO} fiber optic data links, each having a bandwidth B_{FO}, is given by

$$B_{IN} = U_{FO} \cdot N_{FO} \cdot B_{FO} \tag{1}$$

where U_{FO} is the effective utilization of the theoretical bandwidth allowing for fluctuations in the data rate. Currently available technology can support 100 MB/s data rates on a single optical fiber; using this data rate and assuming a value of .5 for the utilization factor gives 200 input data links for an aggregate bandwidth of 10 GB/s.

The outputs of the event builder are high speed communication links to the processing elements. The Scalable Coherent Interconnect (SCI) is being developed to provide high speed point to point connections for future generations of multiprocessor architectures. The 1 GB/s bandwidth of SCI and its potential for widespread industry support make it an attractive candidate for the connection between the event builder and processor farm. The aggregate bandwidth of N_{SCI} SCI links, each having a bandwidth of B_{SCI}, is given by

$$B_{OUT} = U_{SCI} \cdot N_{SCI} \cdot B_{SCI} \tag{2}$$

where U_{SCI} is the utilization factor. A 10 GB/s aggregate bandwidth can be achieved using 40 SCI links with a utilization factor of .25.

The total number of nodes in the event builder is then given by

$$N_{EB} = \frac{B_{IN} \cdot B_{OUT}}{U_{FO} \cdot U_{SCI} \cdot B_{FO} \cdot B_{SCI}}. \tag{3}$$

For the above assumptions, 8000 event builder nodes are required. It is assumed that an event builder node consists of a control chip, input and output interface

chips, and a static RAM memory module. With ≈ 20 such nodes packaged on a printed circuit board, the event builder requires 10 event builder boards per SCI link, each serving ≈ 125 processors.

The total amount of memory required by the event builder is given by

$$M_{EB} = \frac{N_P \cdot \bar{M}}{U_M} \tag{4}$$

where $\bar{M}$ is the average event size, N_P is the total number of Level 3 processors, and U_M is the average utilization of the available memory. For a 1 MB average event size, 5000 processors, and assuming $U_M = .25$ yields an estimate of 20 GB of memory. While 20 GB of fast memory would be prohibitively expensive at current prices, prices should drop drastically over the next few years as computing technology becomes driven increasingly by memory speed rather than CPU speed. Note that the amount of memory used by the event builder is a fraction of the memory needed to equip 5000 processors.

The design parameters are summarized in the following table.

Data Acquisition Design Parameters	
Data Collection Chips	10^4
Fiber Optic Data Links	200
Fiber Optic Bandwidth	100 MB/s
SCI Links	40
SCI Bandwidth	1 GB/s
Processors per SCI Link	125
Number of Event Builder Nodes	8000
Event Builder Memory	20 GB

8. CONCLUSIONS

A data acquisition architecture applicable to high-p_T SSC detectors is described that satisfies a specific set of goals, assumptions, and design principles. The interface to the front end electronics is accomplished using only three signals. Dual port memories are used to form an integrated system of data collection, event building, and processor communication. A set of scaling laws is obtained and design parameters are estimated for a data acquisition capacity of 10 GB/s. Further work needs to be done to develop computer models and perform simulation studies using this architecture that will test the ability of the architecture to achieve the level of performance required for an SSC detector.

DISCLAIMER

- END -

DATE FILMED

07 / 24 / 90