Use of Commercial Gigabit Data Switches for SSC and LHC Event Builders*

William H. Greiman, Stewart C. Loken and Charles P. McParland

Lawrence Berkeley Laboratory
Berkeley, CA 94720 USA

Abstract

Existing event building techniques will be inadequate for detectors at the SSC and LHC. These detectors will require event builders with 100-1000 times the throughput of current event builders. Data acquisition systems for these detectors must be designed to allow incremental upgrades that will increase performance by at least ten fold. This paper presents details of a scalable parallel data acquisition system architecture that meets requirements for detectors at SSC and LHC. An implementation of the proposed architecture is described. This implementation is based on commercial gigabit Fiber Channel data switches. A prototype test bed has been constructed.

Introduction

The high trigger rate and large event size in future high-energy physics experiments will require new techniques for event readout ("event building"). A high degree of parallelism will be required to achieve this level of performance.

The event builder has become the bottleneck in current data acquisition systems. The performance of current event builders is limited by the interconnection network that is used to multiplex data between sources and destinations. The interconnects used in current event builders, shared buses and multiport memory, can not provide the performance required for SSC and LHC.

Gigabit Network Technology

During the last few years standards for networks with several orders of magnitude greater performance than today's Ethernet have been developed. Examples include Asynchronous Transfer Mode (ATM) [1], Fibre Channel [2], and Synchronous Optical NETwork (SONET) [3]. All of these standards specify data links with a bandwidth of at least 1 Gigabit/sec. These standards also specify that the network fabric must be able to provide full bandwidth for simultaneous communication between all possible pairs of nodes. This means that the network fabric must be equivalent to a full cross bar switch.

The first implementations of these standards are now available. A switch that supports the ANSI standard Fiber Channel protocol over 1 gigabit links has been delivered by Ancor Communications. This switch has been designed to be modular and scalable in range of 16-2000 ports. A 2000 port switch could handle up to 100 Gigabytes/sec of user data. Similar switches will be available from other companies during the next year.

---

* This work was supported by the Director, Office of Energy Research, Scientific Computing Staff, of the US Department of Energy under contract No. DE-AC03-76SF00098

Generic Architecture

A common architecture has evolved for the primary flow of data from the detector to processing and recording elements in high performance data acquisition systems. At the highest level of abstraction, almost all proposed high performance data acquisition systems have the classic data flow architecture shown in figure 1 [4]. A scalable fully parallel interconnection network is one of the key components required to implement this architecture. Commercial switches provide an alternative to the custom interconnects that have been used in high performance data acquisition systems. They have the advantages of supporting standard protocols, being supplied by multiple vendors and are scalable or upgradable without additional engineering cost.

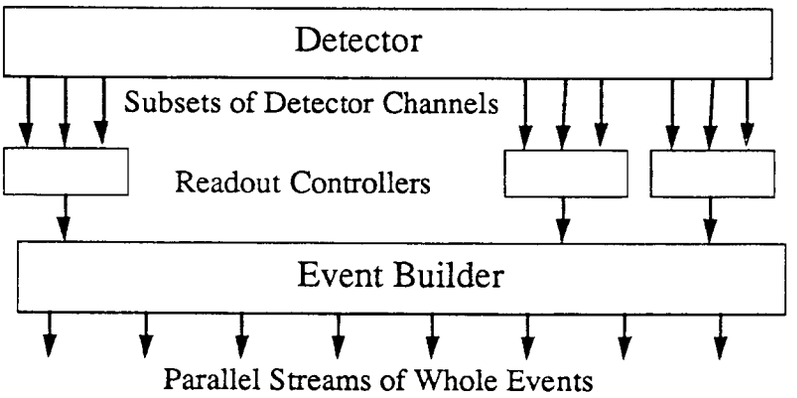

Figure 1

A Fibre Channel Based Event Builder for SDC

SDC requires an event builder that accepts data from about 400 readout sources and distributes whole events to 500-1000 processors in a level 3 farm. The initial performance requirement is 1000 events per second for one megabyte events. The SDC data acqusition system must be based on a scalable architecture that will allow a factor of ten increase in performance.

Figure 2 shows an a scalable parallel event builder suitable for SDC. This architecture has three stages, subevent builders, data switch and full event builders. All data paths are assumed to be 1 Gigabit/sec Fibre Channel links with a capacity of 100 MB/sec of user data. Ten processors are required in each stage to achieve the one GB/sec required by SDC. Each stage can be scaled by increasing the number of processors.

The first stage consists of processors that receive data from a subset of readout sources over Fibre Channel links and assemble subevents. This stage is required to reduce the transaction rate and/or memory requirements in the event builder stage. The subevent builders are not required for smaller detectors such as STAR at RHIC. In this case readout sources can go directly to the switch.

The second stage is a Fibre Channel switch that distributes subevents to appropriate event builder processors. Current Fibre Channel switches, such as supplied by Ancor

Communications, are designed to support over 2000 ports at full bandwidth. This exceeds the SDC scalability requirement by a factor of ten.

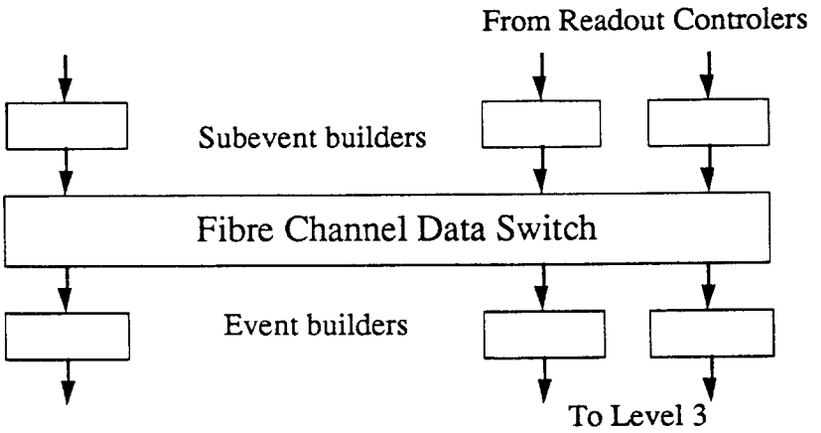

Figure 2

Processors that perform full event assembly form the third stage. All subevent fragments for a given event are directed to one of the event builder nodes through the switch. Event builder nodes assembles these fragments into complete events and distributes them to the level three processors over Fibre Channel links.

Features and Constraints

The benefits of an interconnect based on commercial networks are not free. The key disadvantages are constraints imposed on the architecture by network hardware and protocols. These constraints require that some functions be implemented by techniques that are not commonly used in data acquisition systems.

The flow control and error recovery of the network protocol can be used to implement a simple data flow architecture. Distributed flow control is provided by a sliding window algorithm. This allows high link utilization without special hardware elements to implement control functions. Readout synchronization in the trigger system and front-end electronics is required to achieve this benefit. Hardware trigger inhibit must be asserted if any readout controller lacks sufficient buffering, due to flow control, to readout it's portion of an event. This prevents partial readout of an event.

The network protocols and interfaces have been designed for efficient transmission of large messages and modest transaction rates. A message size of about one Mbyte is required for efficient use of the network. This will require multiple event fragments to be sent in each message. This has the advantage of decoupling the real time event rate from the rest of the data acquisition system at the readout controllers. Sufficient memory must be provided to buffer large messages.

Event assembly can be performed by DMA engines on communication controllers to reduce CPU and bus usage. Data formats and algorithms must be designed to take advantage of these features of the network.

Distributed load balancing can be achieved by simple algorithms. The full connectivity of the data switch allows flexible load balancing algorithms without changes to the hardware configuration.

Gigabit Test Bed

A prototype test bed has been constructed. This prototype is based on 16 channel Fibre Channel data switch modules and VME based processors and communication controllers. This test bed is being used to determine the performance a Fiber Channel based event builder.

Performance tests include switch throughput, switch overhead, VME controller performance, error rates and bus usage. The results of tests performed with this prototype will be used to predict the cost and performance of a full scale event builder.

Summary

Gigabit networking technology is now becoming available with sufficient bandwidth to serve as the high performance interconnect in data acquisition systems for experiments at SSC and LHC. The advantages of using commercial technology are well known. There are also risks and constraints in the use of commercial technology. Constraints that are imposed on the system architecture by network protocols and performance characteristics are a key disadvantage. The biggest risk is that some unforeseen aspect of the technology will cause it to be unusable.

A test bed has been constructed at LBL to evaluate the use of gigabit networking technology in high performance data acquisition systems. Prototype event builders will be constructed and performance and functional tests will be performed. The purpose of these tests is to evaluate use of this technology for SDC at SSC and STAR at RHIC.

References

1] T1S1.5/92-002R3, " Broadband ISDN ATM Aspects -- ATM Layer Functionality and Specification", ANSI Draft Standard, May 1992.

[2] FC-P/92-001R3.0, "Fibre Channel Physical and Signaling Interface (PC-PH)", Working Draft June 16, 1992.

[3] TA-NWT-000253, "Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria", Bellcore Issue 6, Sept. 1990.

[4] Vicky White, "Future Data Acquisition Architectures", *Proc 8th Conf. on Computing in High Energy Physics*, p 65-81, Santa Fe, NM, 1990.