COMPUTING AND DATA HANDLING REQUIREMENTS FOR SSC AND LHC EXPERIMENTS*

A. J. Lankford
Stanford Linear Accelerator Center,
Stanford University, Stanford, *CA 94309*

ABSTRACT

A number of issues for computing and data handling in the online environment at future high-luminosity, high-energy colliders, such as the Superconducting Super Collider (SSC) and Large Hadron Collider (LHC), are outlined. Requirements for trigger processing, data acquisition, and online processing are discussed. Some aspects of possible solutions are sketched.

INTRODUCTION

At the Superconducting Super Collider (SSC) and Large Hadron Collider (LHC), very high energy interactions of colliding protons are expected to occur at a rate of approximately 100 MHz (SSC) to 1 GHz (LHC). These interactions will be studied by high energy physics experiments consisting of more than a million electronic channels each. Sophisticated trigger processing systems will be needed to select rare interactions of physics interest. High-performance data acquisition systems will be needed to move large quantities of data from detector elements through trigger processors and to mass storage. Extensive online processing will be required to filter the number of interactions and the amount of data per interaction down to a rate of approximately ten to one thousand interesting events per second and to an overall data rate which is compatible with mass storage techniques and future offline computing capacity. The architecture of the online processors, the efficient high-speed transfer of data among the processors, and the effective management of processing resources and software will be crucial.

The requirements for triggering, data acquisition, and online processing will be greatly increased from the current generation of experiments. Table I outlines the change in scale in some of the parameters of interest between the present CDF experiment at the Fermilab Collider and a typical SSC experiment. This paper attempts to sketch the requirements for online systems, highlighting some of the new issues. Following a sketch of an overall architecture, issues for trigger processing, data acquisition, and online processing are discussed. The paper concludes with a summary of some of the issues in the design of the large online systems necessary to address the requirements for experimentation at the SSC and LHC.

Table I Comparison of an SSC Detector to the CDF Detector at FNAL

Parameter	CDF	SSC
Collision energy (TeV)	2	40
Luminosity ($cm^{-2} sec^{-1}$)	2×10^{30}	10^{33}
Crossing interval (nsec)	3500	16
Inelastic cross section (mb)	40	100
Total rate (Hz)	8×10^4	10^8
<# interactions/crossing>	0.3	1.6
Channel count	10^5	10^6
Silicon vertex detector	1991	yes
Event size (Mbyte)	0.15	1
Events to tape (Hz)	1	10-1000

In focusing on data handling and processing in the online environment, this paper overlooks the tremendous challenge of handling, processing, and analyzing the large sets of data which these future experiments will generate. Another talk at this conference' considers some of the offline issues.

SYSTEM OVERVIEW

The overall architecture of detector readout systems will be determined by the architecture of the trigger. The data flow through the trigger and data acquisition systems of a representative high-P_T SSC experiment is shown in Fig. 1. As in current systems, the trigger will select event candidates in a series of stages, or levels, which are progressively more complex and more time-consuming. Each level, by reducing the rate of event candidates, will afford the subsequent level more processing time and reduce the bandwidth required for data transmission. Our model in Fig. 1 shows three trigger levels.

In this model, data from individual detector elements is buffered in front-end electronics while Levels 1 and 2 reduce the initial rate of 10^8. Only the data required by each level of the trigger is transported to the trigger processors while all the data is buffered. After selection by the Level 2 trigger, all the relevant data from an interaction of interest is moved to processors in a Level 3 "farm." The data of interest, or event size, will be between 0.2 and 1 Mbyte. In order to reduce the data from approximately ten million detector elements to less than a Mbyte, considerable processing power must be provided to suppress data from elements without signals and to compress data and filter data from hit elements.

The algorithms which select event candidates at each level of the trigger will determine both the data bandwidth required for input into the trigger processors and the data rate between stages of the data acquisition. Within a given experiment, a certain amount of flexibility will be available with respect to choosing
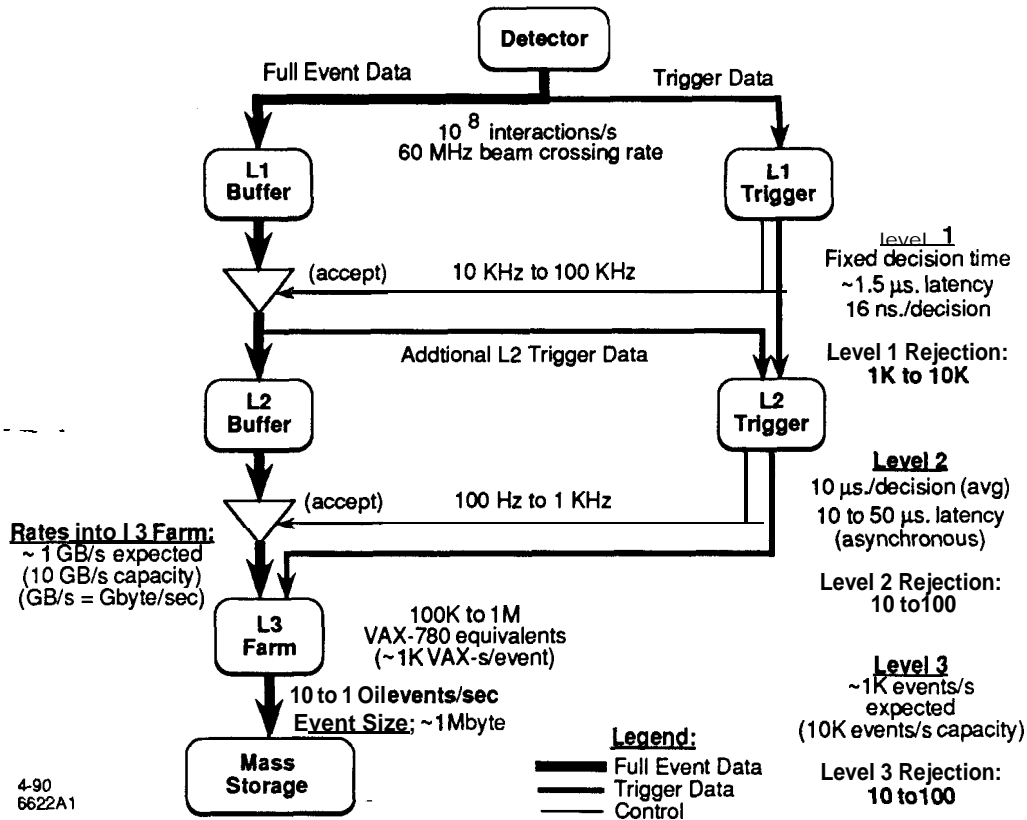
2

Detector

Full Event Data Trigger Data

10^8 interactions/s
60 MHz beam crossing rate

L1 Buffer   **L1 Trigger**

(accept) 10 KHz to 100 KHz

Level 1
Fixed decision time
~1.5 µs. latency
16 ns./decision

Addtional L2 Trigger Data

Level 1 Rejection:
1K to 10K

L2 Buffer   **L2 Trigger**

(accept) 100 Hz to 1 KHz

Level 2
10 µs./decision (avg)
10 to 50 µs. latency
(asynchronous)

Level 2 Rejection:
10 to 100

Rates into L3 Farm:
~ 1 GB/s expected
(10 GB/s capacity)
(GB/s = Gbyte/sec)

L3 Farm

100K to 1M
VAX-780 equivalents
(~1K VAX-s/event)

Level 3
~1K events/s
expected
(10K events/s capacity)

10 to 1 Oil events/sec
Event Size: ~1 Mbyte

Legend:
▬ Full Event Data
━ Trigger Data
─ Control

Level 3 Rejection:
10 to 100

Mass Storage

4-90
6622A1

Fig. 1: Model of data flow through trigger and data acquisition systems for a high-P_T experiment.

at which trigger level to deploy selection criteria; however, the algorithms are detector and physics dependent. For instance, the trigger criteria and the final rate of interesting events are quite different for experiments studying high-P_T phenomena and those studying decays of beauty. Thus, a high degree of inter-play exists between the capabilities of the trigger and of the data acquisition at each level in the system.

TRIGGER PROCESSING

Trigger processing is perhaps the most exciting technical challenge at future colliders. It is crucial for extracting the physics signals which we seek to study from extremely high rates of complex background events. In fact, unprecedented interaction rates will require the full power of offline physics analysis techniques to be available in the trigger for event filtering. Consequently, the trigger interacts broadly with both physics goals and detector design.

Trigger processors at future colliders must contain algorithms which iden-tify, count, and measure the quanta which characterize the physics at high en-ergies: jets, muons, electrons, photons, and weakly interacting particles, such as neutrinos, which leave missing E_T. The trigger processors must also combine

requirements on these quanta and on event topology in order to select event candidates.

Trigger Levels

At future colliders, even the first stage of trigger decision cannot be made during the interval between bunch crossings. Consequently, every detector signal from every bunch crossing must be buffered until the Level 1 trigger decision is complete, and the Level 1 trigger must complete a trigger decision each 16 nsec in order to keep pace with the rate of bunch crossings. The Level 1 processing time must be minimized in order to reduce the number of bunch crossings for which data will be buffered. Decision times of about 1 μsec are generally discussed in light of the propagation times to and from the trigger on a large detector (about $1/2\,\mu$sec) and the need to form some global event quantities such as missing E_T. A fully pipelined hardware processor which exploits extensive parallelism in order to reduce latency will address these requirements. Its pipelined architecture suggests that this processor will have a fixed decision time, which is also convenient for the architecture of the signal buffers. A subset of all detector signals will be provided to the Level 1 processor on data paths which are separate from the paths used for data acquisition. The Level 1 trigger will provide rejections of between 10^3 and 10^4.

Between 10^4 and 10^5 event candidates per second remain at the input to the Level 2 trigger, affording it 10–100 μsec on average per decision. Thus, its processing must be prompt; however, the additional decision time available allows iterative processing, such as sequential processing of track candidates. Additional time also allows event candidates to be directed to independent processors for processing in parallel. In this way, the Level 2 trigger can exploit "event parallelism" in the processor farm sense, as well as "parallelism within an event" as used by Level 1. With or without the use of event parallelism, microprocessors embedded within the Level 2 architecture may play a significant role in the Level 2 trigger selection. The Level 2 processor will still operate only on a subset of all detector data transported on a separate data path, including the data used by Level 1 and the output of Level 1.

The iterative nature of Level 2 suggests that its decision time will be variable, in the range of tens of microseconds; however, for the convenience of the architecture of the front-end signal buffering, the Level 2 trigger processor will preserve the order of event candidates, performing resequencing if trigger decisions complete out of order. Rejections of about 10^2 are expected for Level 2.

The rate of event candidates into the Level 3 trigger is then between 10^2 and 10^3, a rate which is sufficiently low to allow transport of data from all parts of the detector and to accommodate a farm of microprocessors as the Level 3 trigger processor. In fact, rates into Level 3 higher than 10^4 may be feasible. The full event, with the full detector resolution, consequently is available, as are the power and flexibility of general-purpose, high-level language programmable CPUs. Rejections of between 10 and 10^2 are expected from Level 3, resulting in a final rate of event candidates of a few tens per second.
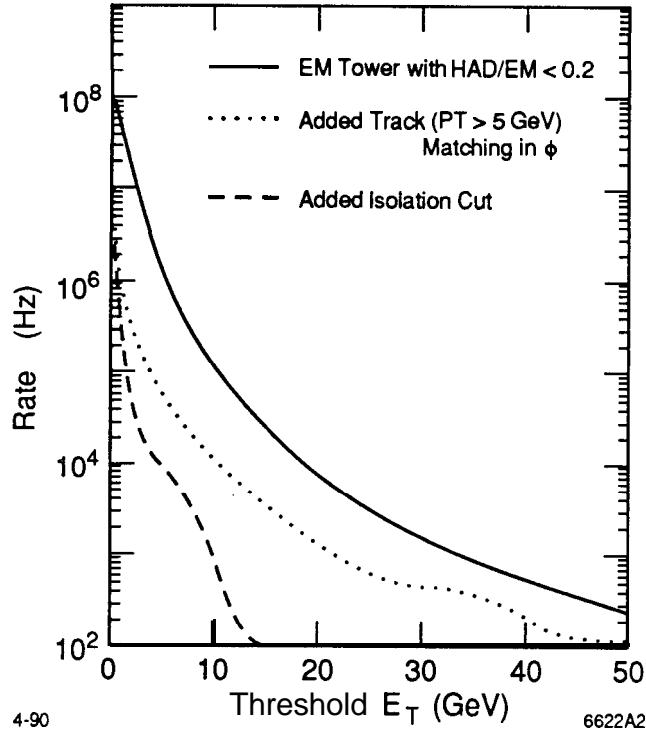
Fig. 2: Trigger rates vs. energy threshold for electron selection criteria.

Trigger Example: An Inclusive Electron Trigger

One of the many triggers of interest at the SSC and LHC, an inclusive electron trigger, illustrates the general nature of event selection criteria, and hence of trigger processing, which might be used.

At Level 1 the energy deposit in the electromagnetic section of a calorimeter tower of size approximately $\Delta\phi$ x $\Delta\eta$ = 0.2 x 0.15 will be required to be above threshold, probably in the range 20 to 40 GeV. The energy in the hadronic section will also be required to be less than some fraction (approximately 20%) of the energy in the electromagnetic section. These criteria have been studied by Sakai of KEK using a simple calorimetric model with fast shower simulation of QCD events generated by ISAJET. The resulting rate is shown by the solid curve in Fig. 2 as a function of energy threshold. For example, rejection greater than 10^4 (i.e., rate less than 10^4) is achieved for thresholds above 20 GeV.

At Level 1 or 2, the presence of a stiff track segment with $P_T > 5$ GeV pointing towards the trigger cell in ϕ (i.e., with no z requirement) will also be required. This criteria will reduce the rate by about another factor of 10, as shown by the dotted curve in Fig. 2.

At Level 2 the trigger cell will also be required to be isolated. That is, the energy in nearest neighbor cells, electromagnetic and hadronic, will be required to be less than about 20% of the energy in the trigger cell. Rejection greater than 10^6 will then be achieved for all energies greater than about 12 GeV, as shown by the dashed curve in Fig. 2.

5

At Level 3 further rejection will be achieved via selection based on a combination of longitudinal and lateral shower profiles, track matching in space and in momentum, conversion rejection by tracking and dE/dx, and tighter isolation cuts to reject heavy quark (c and b) decays.

Some General Trigger Processing Issues

The bandwidth required to transport data to prompt trigger processors for 60 MHz bunch crossings is quite high, even for subsets of the detector data. For instance, 5000 calorimeter sums of 2 bytes each require a bandwidth of 600 Gbytes/sec.

Most trigger quantities are topologically localized on the detector. For instance, the detector signals which characterize an electron originate in a small region of solid angle. Consequently, much trigger processing could be done locally, which would ease the data bandwidth problem.

Power dissipation of trigger processors, and of drivers which transmit data to the trigger, may limit the amount of trigger processing on various parts of the detector, or it may limit the amount of data which is available to the trigger. For instance, transmission of all hit wire information from a central drift chamber to a remote trigger processor may be problematic, as may be local processing of all hit wires into track segments.

The trigger designer will often have a choice between exploiting event parallelism or parallelism within an event. Event parallelism is exploited by processors working in parallel on separate events, as in a microcomputer farm; whereas, parallelism within an event is exploited by parallel processors working on separate portions, such as different regions of solid angle, of the same event.

The trigger latency, even for deadtimeless triggers, is important in that it affects the design of front-end electronics. In the simplest solutions, it affects the amount of buffering, and possibly the architecture of the buffers, in the front-end. In some solutions, such as "smart" pixels, the effect on occupancies, ambiguities, and resets is profound. The Level 1 latency is at least half a microsecond, which is the propagation time of signals to and from a central trigger processor.

Processing must be provided such that each detector entity which provides a trigger, e.g., each calorimetric trigger tower, can identify the bunch crossing being triggered upon. Positive crossing identification is possible even for detector components which do not have single crossing response times. For instance, the time of arrival of liquid ionization calorimeter signals can be derived from the zero-crossing of their predictable pulse shape. Time resolution in the 1-2 nsec range should be achievable for 10 GeV electrons and 50-100 GeV jets in liquid argon calorimeters. In drift chambers, correlations in drift times between nearby offset layers allow untangling of the drift time from the time origin of the ionization.

Processing to identify or disentangle multiple interactions during the resolving time of the detector will also be needed.

The questions of "How selective should the trigger be?" and "How many events should be written to tape?". are closely related to physics goals. However, there exist tradeoffs between recorded event size and number of events recorded,

as well as in applying processing power to reducing one or the other. Both reductions are forms of data filtering.

Calorimeter Trigger Processing

Calorimetric energy is basic to the identification of physics quanta, quarks (i.e., jets), electrons, and neutrinos (via missing E_T). The scalar sum of E_T of particles from an interaction is related to the $\sqrt{\hat{s}}$ of the parton interaction. Processing for calorimeter triggers requires minimal pattern recognition and is naturally implemented in low-level triggers.

Calorimeter trigger processors must calculate global energy-related quantities, such as ΣE_t and E_t, and identify local energy deposition, such as electromagnetic showers and jets. Use of digital processing for these triggers is likely to continue to increase.

A variety of clustering algorithms are now in use for low-level triggers on jets. These include energy clustering about a seed tower as done by CDF, energy summing in overlapped fixed cones as done by UA1, energy clustering in detector subregions with special treatment of edge effects as done by Zeus, and identifying a seed tower only as done by DO.

In order to avoid a separate trigger bias, the trigger processor should achieve the required level of rejection using the same jet algorithm, or a subset of it, as is used offline for physics analysis. For ease of theoretical interpretation, most experiments now seem to prefer a jet algorithm which defines a jet as energy flow within a fixed cone about a jet axis. The cone size, however, varies. with the physics being studied.

What is the ideal prompt calorimeter trigger? Perhaps it would be provided by a massively parallel architecture in which a single, simple processor corresponding to each tower investigates the hypothesis that its tower is the center of an energy cluster (for several fixed apertures), with all towers being processed in parallel, and perhaps even employing the full granularity. A second level of logic could arbitrate overlapping clusters. This trigger processor implements an offline algorithm with the full resolution of the offline analysis. On the other hand, a much less ambitious solution may also provide the required level of rejection without introducing trigger biases.

Any future prompt calorimeter trigger processor will more fully exploit the segmentation, calibration, and resolution of the calorimeter than in the past. In fact, few selection criteria may remain for use by the higher-level processors. Higher-level processors may be limited to refinement of electron identification and further selection and combination of criteria which are formed by the prompt logic.

Processing for Tracking Triggers

Tracking of charged particles by the trigger is instrumental to selection of electron and muon candidates. For electrons, the presence of a stiff charged track directed towards an electromagnetic shower reduces photon and π^0 backgrounds.

In addition, tracking can link information from transition radiation detectors to showers and can provide an **E/p** check to help reject chance overlap of a charged track with a shower produced by a photon. Identification of track segments, rather than full track reconstruction and momentum measurement, may be sufficient for any of these tasks at Levels 1 and 2; however, full track reconstruction will be needed at higher levels.

Requirements for high-& muons depend on the detector configuration. In the central region of a detector with iron absorber, sufficient rejection is provided by demanding the presence of a penetrating track segment in the muon system which points back to the interaction vertex, where a cut on the angle of the segment in the bend plane provides a P_T cut. At smaller angles, below about i5 degrees, or in a detector with air-core toroids, a sharper P_T cut, in the range of 10–15 GeV is needed. This will require use of drift time information and track reconstruction even at Level 1.

Beauty physics places a premium on track finding by prompt triggers since the transverse momenta of particles from B decay are not sufficiently large for calorimeter triggers. On the other hand, relatively stiff tracks, in the few GeV range, do arise from the B mass and **PT.** A prompt trigger which selects events with at least one track with $P_T > 3$ GeV or at least two tracks with $P_T > 2$ GeV may provide an enhancement in B events of about a factor of 50. For this purpose, it may be possible to define a track as a segment at the outer radius of the tracking system whose P_T is measured by linking the segment to the interaction vertex. At higher levels, full track reconstruction and precise vertexing will be required for B physics.

In considering the processing required by tracking triggers, recall the basic steps in tracking algorithms for typical chambers. The first step is track segment finding within a localized region of the chamber. Then segments are linked into tracks, essentially by clustering in curvature-angle $(\rho - \phi)$ space. Tracks are then fit. Finally, vertexing, another clustering task, is performed. Although the fitting step is computationally intense, the other steps are characterized by local pattern recognition and clustering which could be performed by parallel or otherwise novel processors.

Some Level 2 Processing Techniques Under Investigation

A number of techniques for trigger processing at Level 2, which requires execution times between 10 and 100 μsec, are under investigation. In fact, the preponderance of R&D on triggers is directed at Level 2. Clustering algorithms, primarily for calorimetry, are being worked upon. Neural networks, which are discussed in talks at this conference,[2,3] are very topical. They may be well matched to some of the problems of local pattern recognition, such as track segment finding and energy clustering. Image processing is being actively studied by LAA and CERN groups. Commercial image processors are promising for pattern recognition on the Level 2 time scale, and may be particularly suited to two-dimensional detectors, such as calorimeters, pad chambers, and pixel detectors. Data driven pipelined processors which are evolutions of the one used in FNAL E690 could

provide tremendous computational power to algorithms which can be defined well and which do not need the full flexibility of high-level language programmable processors. The computational power of these processors can be applied to pattern recognition problems as well as to performing calculations such as track fits. Custom content addressable memories for pattern matching are being developed, offering more patterns than possible with standard memory look-up techniques. Fine-grained parallelism, as provided by the Associative String Processor (ASP) or by the Connection Machine, is also under investigation. These processors may provide a way of matching the granularity of the processing to the granularity of the detector.

Special-Purpose vs. General-Purpose Trigger Processors

Special-purpose processors, such as traditional hardwired triggers, and general-purpose microprocessor farms often seem in competition as trigger processors. In fact, both types of processors have roles in the trigger. Special-purpose processors are necessary for speed at the first levels of prompt triggers, and can be designed to be programmable with respect to important parameters. General-purpose processors are required for flexibility at the last level of event selection. Furthermore, the distinction between special-purpose and general-purpose will fade as DSP and RISC cores are embedded in custom circuits and as custom coprocessors are attached to general-purpose CPUs. The crucial issues in choosing technologies are "How much processing power is required?" and "How much flexibility is needed?" Physics-goals and detector design will determine the technology requirements.

DATA ACQUISITION

As outlined in Fig. **1** for a high-P_T detector, the data acquisition system must buffer signals from all detector elements and from many interactions while Level 1 and 2 trigger processing occurs. Then it must collect and transmit the data at rates of 1 to 10 Gbytes/sec to an online processor farm. In order to achieve these bandwidths, parallel data links must be used, and a routing mechanism, referred to here as a parallel event builder, must enable the data from the parallel links to be directed as complete events to parallel processors. Finally, between 10 to 100 events, each of about 1 Mbyte, must be recorded per second.

The distinctive features of the data acquisition system will be front-end electronics based on custom VLSI, high-speed data collection and data transmission using fiber optics, parallel event building, massive processor farms, and large amounts of data placed onto mass storage. Separate control and data paths are likely. Processors will be intensively used for triggering, calibration, data compression, and monitoring tasks, which will lead to an increased dependence on software.

Integrated Front-End Electronics

Many data acquisition functions which have traditionally been executed in the counting house, such as digitization, multiplexing, and buffering, will be
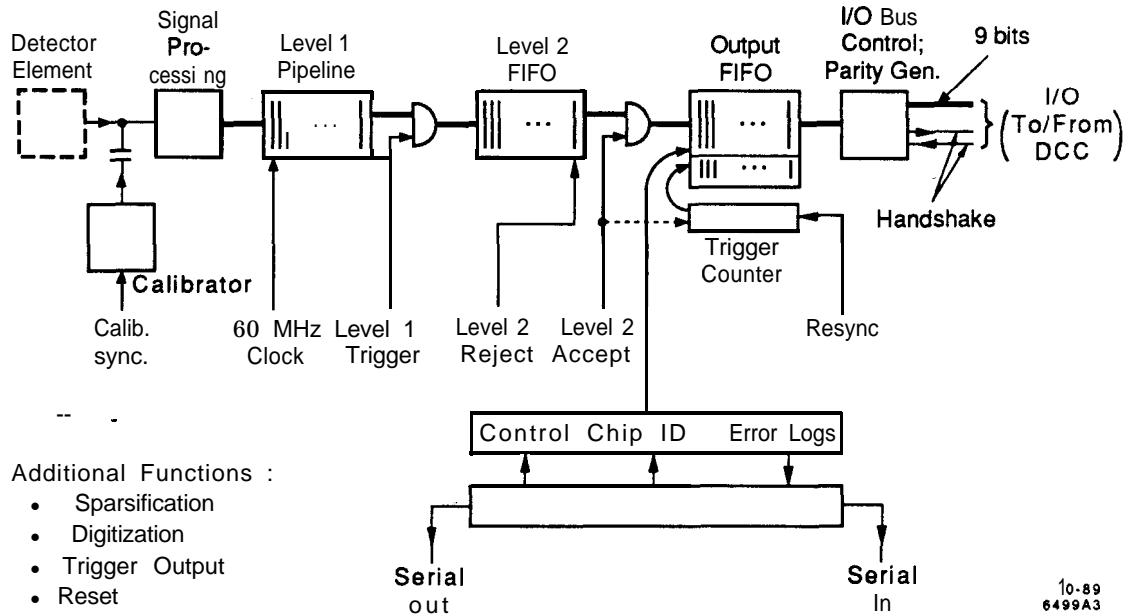
Detector Element Signal Processing Level 1 Pipeline Level 2 FIFO Output FIFO I/O Bus Control; Parity Gen. 9 bits I/O (To/From DCC)

Handshake

Trigger Counter

Calibrator

Calib. sync. 60 MHz Clock Level 1 Trigger Level 2 Reject Level 2 Accept Resync

Additional Functions :
• Sparsification
• Digitization
• Trigger Output
• Reset

Control Chip ID Error Logs

Serial out Serial In

10-89
6499A3

Fig. 3:   Functional architecture of front-end electronics systems.

performed in highly integrated, detector-mounted electronics at the SSC and LHC. Silicon microstrip detectors and the SLD detector have recently pioneered detector-mounted custom VLSI and hybrid circuits; however, the trend must be carried much further in the future. A number of--motivations, including'improved analog performance, increased immunity to RF pickup, density of connections, limited cable space, cost effectiveness, space efficiency, and reliability, are joined by the compelling needs of reduced power dissipation and of increased functionality (e.g., multiple event buffering, integrated trigger solutions, and simultaneous read in and readout). Solutions discussed for the SSC normally include the entire functionality shown in Fig. 3 for several readout channels, including control logic, on one or two custom chips. A separate talk at this conference[4] discusses development of these chips. For an SSC detector, these chips will replace both the boxes of detector-mounted amplifiers and the crates of remote FASTBUS TDC modules found in today's large detectors, as well as the hundreds of long cable interconnections. Further detector-mounted multiplexing and data preprocessing will replace today's crate-level scanners and segment interconnects.

Note that the functional architecture shown in Fig. 3 is the "logical" architecture, not necessarily the physical architecture, of the front-end electronics. The electronics of all detector components are expected to have similar architectures, enabling a common control and readout scheme for the entire detector. A possible scheme for this control and readout is discussed[5] at this conference.

Data Collection

Data from as many as several hundred thousand front-end chips, each with data rates of roughly hundreds of Kbytes/sec, must be multiplexed onto a man-

10

ageable number-perhaps 100 to 1000—of high-speed data channels which provide an aggregate data rate of several to 100 Gbytes/sec. A hierarchical solution to data collection, starting with groupings of nearby detector channels and proceeding towards large groupings of all the data from one region of solid angle, is appropriate. The entire data collection process, reducing the number of data paths to the few hundred to be input to the parallel event builder, will occur within and on the detector.

The most ambitious solutions to the problem of data collection, those aimed at the highest achievable rates of data transfer, are data driven. At each step in the data collection process, every data source is pushing data into intermediate buffers as the data becomes available. Data collectors then gather the data from the buffers at the highest possible rate and push the data into the next stage of buffers. The bandwidth of all data links can be used to full efficiency. The data is transmitted with appropriate event and channel tags; however, packets of data do not necessarily correspond to individual events. The process of event building is therefore to a large extent decoupled from the data collection and transmission. In these data-driven schemes, control is minimized as data is moved along a series of simple data-transmission links. Control occurs on paths separate from the data paths. Operation of such a system should be easy to verify and troubleshoot, since verification and fault identification will be amenable to a series of communications tests, which in fact could be performed by simple expert systems.

Data Transmission

Transmission of data to each stage of data collection will occur via links of technology appropriate to the bandwidth required at that stage. Data collected from the front-end chips, where bandwidths are low, will be transported via copper buses on detector-mounted printed circuit boards. At the other end of the data collection process, the perhaps hundreds of long links carrying the data from all parts of the detector to the parallel event builder in the control room will be high-speed fiberoptic links. The speed and number of links at that stage will be determined by practical considerations, such as the cost and size of the switching network in the parallel event builder. The transition from high-speed copper links to fiberoptic links of modest speed will occur at some intermediate stage.

The principal advantage offered by fiberoptic transmission is that of high bandwidth, particularly over distances longer than several meters. Fiber optics promise performance that makes data acquisition of Gbytes per second feasible. Fiberoptic transmission also offers the important advantages of immunity to electromagnetic interference and low transmission losses. In addition, if used within the detector, they offer advantages in size and mass over copper cables.

The fiberoptic needs of the computer industry are driving technology to increased performance and decreased cost for links similar to those needed for SSC data acquisition. Industry is currently well advanced in developing integrated gallium arsenide electronics for complete fiberoptic systems in the 1 to 2 Gbits/sec/link range. These links should be quite accessible for use in SSC experiments.

Parallel Event Builder

The parallel event builder addresses the bandwidth bottleneck arising in traditional event builders, where data all passes through one path. In **a** parallel event builder, a number of input data paths from the detector are connected to a number of output data paths to the processors, and all the data paths can be active simultaneously to maintain the aggregate bandwidth. The number of input and output data paths need not be equal; however, if bandwidth is nearly optimized then the numbers are naturally the same.

Several schemes for parallel event builders have been discussed. These schemes generally utilize a matrix of buffer/router nodes or utilize switching networks.- The schemes have many similarities, particularly the need for extensive buffering to smooth out event-to-event fluctuations in amounts of data on each link and the need to balance the average data rates on each data path. These needs arise from the fact that the bandwidth will be limited by the longest event fragment of each event if the buffers are insufficient or by the slowest data path if rates are not balanced. A talk at this conference[6] discusses options and issues in parallel event building.

Mass Storage

The required bandwidth for recording selected events will be in the 10–100 Mbytes/sec range. This bandwidth can be provided by parallel output data streams. In fact, parallel data streams may also be desirable in order to record different event types on separate drives. It has been suggested that helical scan magnetic tape technology developed for the commercial broadcast industry will provide storage media of sufficiently high density (200 Gbytes/cassette) for the expected large data samples at the same time as providing drives in the 15-30 Mbytes/sec range.

An alternative to directly recording the output event stream at the site of the experiment is to transmit the data via a high-speed link to the site of the offline computing. At the offline site the data can be recorded by a robotic data archiving system which is shared by offline computing. Fiberoptic systems with the necessary bandwidth for the high-speed link now exist and will be commonplace in advance of SSC operation. Standard protocols for such links are now being developed.

ONLINE PROCESSING

Two categories of parallel processing exist in a large high energy physics experiment. A processor farm performs data processing and event selection on data from the entire detector, with each processor executing the same program on a separate event. Other processors, distributed throughout the architecture of the online system, preprocess streams of data from portions of the detector and control and monitor detector components.

Online Processor Farm Requirements

The highest level of processing for event selection will generally occur in a farm of many microprocessors which may be characterized by its input and output bandwidths, its processing power, and its software environment.

The required input bandwidth to the farm is dependent upon the physics goals of the experiment and upon the deployment of trigger selection criteria between low-level trigger processors and Level 3. The aggregate bandwidths most often discussed range from 10 to 100 Gbytes/sec. The 10 Gbytes/sec rate arises from a conservatively designed data acquisition system for a high-P_T experiment with a prompt trigger rejection of 10^4, i.e., 10^4 events/sec x 1 Mbyte/event = 10 Gbytes/sec. Clearly, an experiment with a prompt trigger rejection of $10^5 - 10^6$ would-require less input bandwidth. On the other hand, a B-physics experiment operating at $\mathcal{L} = 10^{32}$ cm^{-2} sec^{-1} with a prompt rejection of about 10^2 would require input bandwidth of 100 Gbytes/sec. These bandwidths to the parallel processors can be provided by parallel data links.

The required output bandwidth from the farm to mass storage is between 10 and 100 Mbytes/sec, based upon writing 10–100 events/sec at 1 Mbyte/event or 1000 events/sec at 100 Kbytes/event. Parallel output data links can be used.

The aggregate processing power of the farm is usually described as being between 10^5 and 10^6 MIPs. These estimates are loosely based upon needing approximately 100 sec on a 1 MIP machine to perform final event selection with rejection of approximately 10^2.

The architecture of the farm must allow execution of background tasks to the event selection process. Such tasks include testing of new trigger code in parallel with the execution of standard code, verification of event selection processing, and detector performance monitoring. This requirement demands the ability to share events or data among processors.

At least three options exist for the implementation of the farm using commercial products. Commercial microprocessors could be implemented on custom processor boards, the approach chosen by ACP. Commercial single board computers could be implemented as processing nodes, the approach chosen by DO. Finally, a commercial multiprocessor system could be implemented in order to provide the entire farm, an option which may be made possible by the growing interest of industry in large-scale application of parallel processing for general scientific computing problems. Intel, for instance, is developing multiprocessor systems with thousands of loosely-coupled RISC-based nodes which utilize message passing in a two-dimensional mesh. The INTEL Touchstone Project funded by **DARPA** targets a system providing approximately 10^5 MIPs with 2048 i860 processors by 1991. Such systems of high-performance nodes require i/o bandwidths comparable to the needs of high energy physics. They may also provide the required software environment, as well as connections to host/control processors and to workstations. The mesh architecture, which interconnects each note to its four nearest neighbors at high bandwidth (128 Mbytes/sec/connection), may offer processing alternatives to the traditional HEP one event per one processor. For instance, the processors may be mapped onto the topology of the detector for certain processing tasks.

An open architecture is another often-mentioned requirement of the farm. A truly open architecture would allow one to exploit the most cost-effective microprocessor at the time of system implementation, instead of at the time of system design. This point of view is reinforced by the tendency to employ as much computing power as is available and by the frequent need to expand computing power.

Distributed Processing and Control

Although the largest-scale use of commercial processors will be in the processor farm, they will be used extensively for other functions throughout the data acquisition architecture. Processing functions will largely be of the same nature as in current experiments; however, the amount of processing will substantially increase. More than in the past, standard microprocessors will be found embedded in special-purpose low-level trigger processors, in data preprocessors, and in detached control processors. Commercial processors will continue to serve as hosts for the system as a whole and for each detector subsystem. Workstations will be used to interface physicists to the online system, to control and monitor the detector and its performance, and as powerful online graphics machines.

Some Software Requirements for Online Processors

The software environment provided by the farm is of critical importance. The farm must execute large programs written for offline processors, which implies that the farm processors must have high-quality compilers compatible with those used offline. The farm must provide a code development environment which facilitates production and initial debugging of new code, or be compatible with such an environment on another machine. It must also offer adequate tools for in *situ* debugging of code during operation, i.e., debugging of code executing on any node in a multiprocessor system and debugging of interprocessor communications. Code running on such a powerful machine will require new levels of reliability because of the tremendous number of instructions being executed per second. In addition, the operating system must provide tools for data transfer to and from processors and for control and monitoring of processors. In short, the farm must provide a software environment as comfortable as that provided by today's popular minicomputers.

The above requirements are also necessary for many of the processors distributed throughout the data acquisition architecture.

CONCLUSION: SOME ISSUES IN DESIGN OF LARGE SYSTEMS

The systems necessary to address requirements of triggering, data acquisition, and online processing for experiments at the SSC and LHC will be substantially larger and more complex than the corresponding systems in existing experiments. Consequently, new **issues** arise in the design of these large systems.

Functional modeling (i.e., behavioral simulation) of the overall system, including the trigger, data acquisition, and processing, will be necessary to study

system performance with respect to many parameters and to verify system design. The overall system can be modeled at a high level. Mixed-level simulation will be needed to simulate components at various levels of detail in the context of the overall system design. Tools for mixed analog and digital simulation of the demanding front-end electronics would be extremely useful.

The overall design must not allow system complexity to scale with the number of detector channels. Readout solutions should be integrated across detector components. Control mechanisms should be simple.

The applicability of commercial developments and of emerging technologies must be monitored for performance and cost advantages. The overall system architecture should permit the exploitation of technical advances which occur during the development of the experiment, and even during its operational phase. Issues of reliability, redundancy, and in some cases radiation tolerance will require additional engineering techniques and skills. Finally, the verifiability and maintainabilty of very large systems must be considered throughout design.

ACKNOWLEDGMENT

REFERENCES

1. P. Liebold and B. Scipioni, "SSCL Computing Requirements for. Physics and Detector Simulation," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.
2. B. Denby, "Neural Networks for High Energy Physics," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.
3. D. Cutts, "Applications of Neural Networks in High Energy Physics," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.
4. H. F. W. Sadrozinski, "Detector and Front-End Integration," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.
5. R. Partridge, "A Data Acquisition Architecture for the SSC," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.
6. E. Barsotti, "New Approaches to Event Building," *Proc. 8th Conf. on Computing in High Energy Physics,* Santa Fe, NM, 1990.