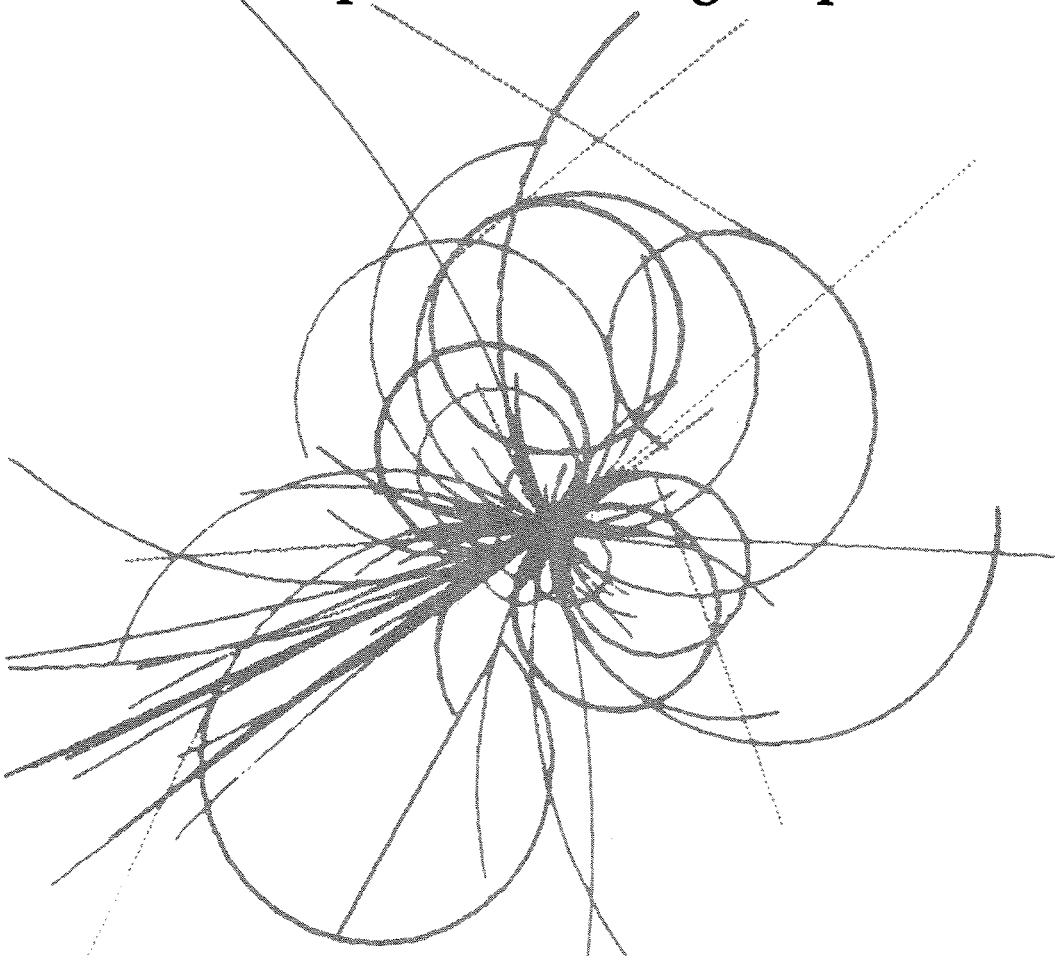
SSC-SR-1039

The Superconducting Super Collider



Proceedings of the Workshop on Triggering and Data Acquisition for Experiments at the Supercollider

R. Donaldson, Editor
SSC Central Design Group

April 1989

SSC-SR-1039

PROCEEDINGS OF THE WORKSHOP ON TRIGGERING AND DATA ACQUISITION FOR EXPERIMENTS AT THE SUPERCOLLIDER*

R. Donaldson, Editor
SSC Central Design Group

April 1989

*Workshop held at the University of Toronto, Ontario, Canada, 16–19 January 1989.

Table of Contents

Triggering Requirements for SSC Physics*

M. G. D. Gilchriese
SSC Central Design Group†
c/o Lawrence Berkeley Laboratory, Berkeley, California 94720 USA

Abstract

Some aspects of triggering requirements for high P_T physics processes at the Superconducting Super Collider (SSC) are described. A very wide range of trigger types will be required to enable detection of the large number of potential physics signatures possible at the SSC. Although in many cases trigger rates are not now well understood, it is possible to conclude that the ability to trigger on transverse energy, number and energy of jets, number and energy of leptons (electrons and muons), missing energy and combinations of these will be required. An SSC trigger system must be both highly flexible and redundant to ensure reliable detection of many new physics processes at the SSC.

Introduction

The intent of this note is to give a few examples of potential high P_T physics signatures at the SSC and their related triggering requirements. The reader should be warned that not enough work has been done to quantify trigger rates (from background processes) so that precise energy thresholds or multiplicities for triggering may be given. Since it is readily apparent that any trigger system must have the ability to vary thresholds, multiplicity and other criteria, this lack of precise information about rates is of little consequence at the present stage of design of triggering systems for the SSC. I will not discuss triggering for B physics experiments. This issue has been covered in some detail in previous workshops.[1]

General Comments

It is important to remember that there is a very high probability that there will be a gradual approach to the design luminosity (10^{33} cm^{-2}sec^{-1}) of the SSC, just as has been the case in most previous colliders. This implies, for example, that energy or other thresholds of interest will change with time. In addition it will be important to be able to trigger on known physics processes, particularly during the initial operation of the SSC. Production cross sections for known processes (e.g., W production or jet production) will be of intrinsic interest in the new energy domain explored by the SSC. Furthermore, such processes will often be backgrounds to new physics and hence must be understood in detail. Known processes (e.g., $Z \rightarrow ee$ or $\mu\mu$)

---

*Talk given at the Workshop on Triggering and Data Acquisition for SSC Experiments, Toronto, January 16–19, 1989, to appear in the Proceedings.

†Operated by the Universities Research Association, Inc., for the Department of Energy.

must also be used as in-place calibration sources for tracking systems, calorimetry or muon identifiers and hence a trigger system must be designed to include these types of reactions. Finally, there is the obvious point that any trigger system must be able to allow substantial prescaling of most if not all types of triggers. Given the enormous interaction rate (10^8 Hz) at the design luminosity, substantial amounts of data for basic cross section measurements, for determination of general event properties and for calibration of detector response can be obtained with large prescaling factors.

New Physics Signatures and Trigger Requirements

In the subsequent paragraphs I will describe some examples of new physics signatures and the related trigger requirements. The trigger requirements will range from those that are easy to provide to rather complicated topologies. This list is by no means exhaustive nor are the various thresholds or other criteria given very well understood. Nevertheless it is possible from these examples to obtain most of the general requirements for an SSC trigger system.

Quark Compositeness

One of the basic experiments to be done at the SSC is to search for evidence of deviations from point-like behaviour by quarks. Depending upon the compositeness scale there are a number of possible signatures that may be observed at the SSC. The most convincing evidence for quark substructure would be the observation of the decay of excited quark states.[2] Possible decay modes for an excited quark, Q^*, are

$$Q^* \rightarrow qg \rightarrow \text{jet} - \text{jet}$$

$$Q^* \rightarrow qZ \rightarrow \text{jet} - Z$$

$$Q^* \rightarrow q\gamma \rightarrow \text{jet} - \gamma$$

A plot of the jet – Z invariant mass distribution is shown in Fig. 1 as an example of a possible signature.

If the compositeness scale is larger, then the observable signature would be a deviation from QCD expectations in the jet transverse energy (E_T) distribution in the TeV range.[3] In both this case and the case of excited quarks, a simple trigger on large E_T, observed in central calorimetry ($|y| < 2.5$–3), is probably sufficient. The trigger rate versus the minimum value of E_T is shown in Fig. 2.[4] If indeed there is a relatively low mass Q^* below a TeV or so, then one may need a "Z-trigger" to restrict the rate at the design luminosity.

New Gauge Bosons

A simple high P_T electron or muon trigger will be sufficient to trigger on new gauge bosons, W' or Z', via their decay modes[5]

$$W' \rightarrow e \quad \text{or} \quad \mu + \text{neutrino}$$

$$Z' \rightarrow ee \quad \text{or} \quad \mu\mu$$

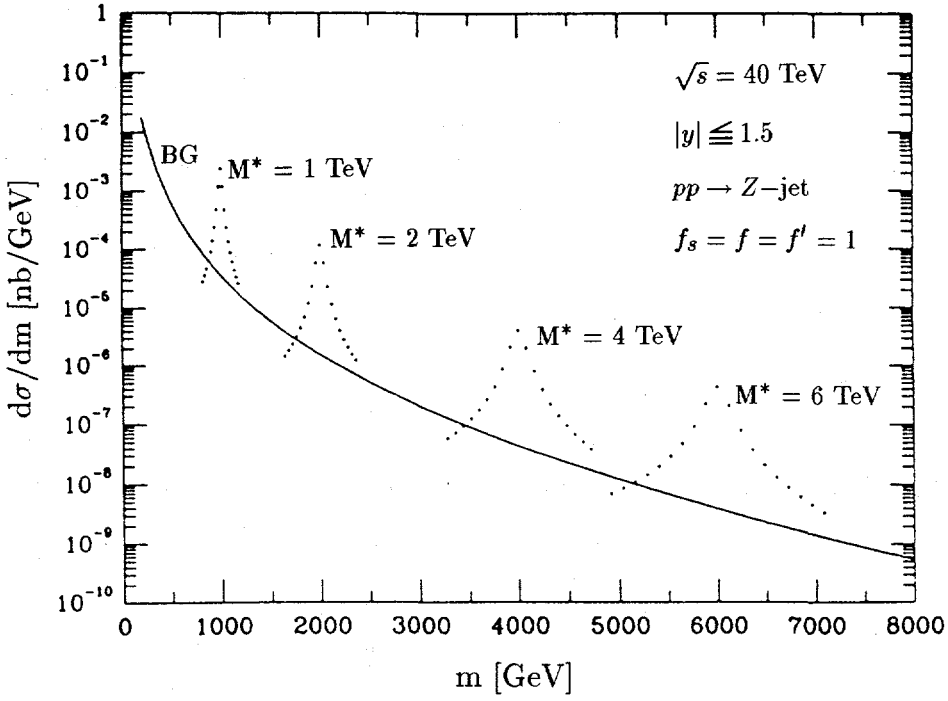


Fig. 1. The invariant mass distribution $d\sigma/dm$ of excited quarks produced in pp collisions at the SSC versus invariant mass for various values of the excited quark mass, M^*. The solid curve represents the Z + jet background. Taken from Ref. 1.

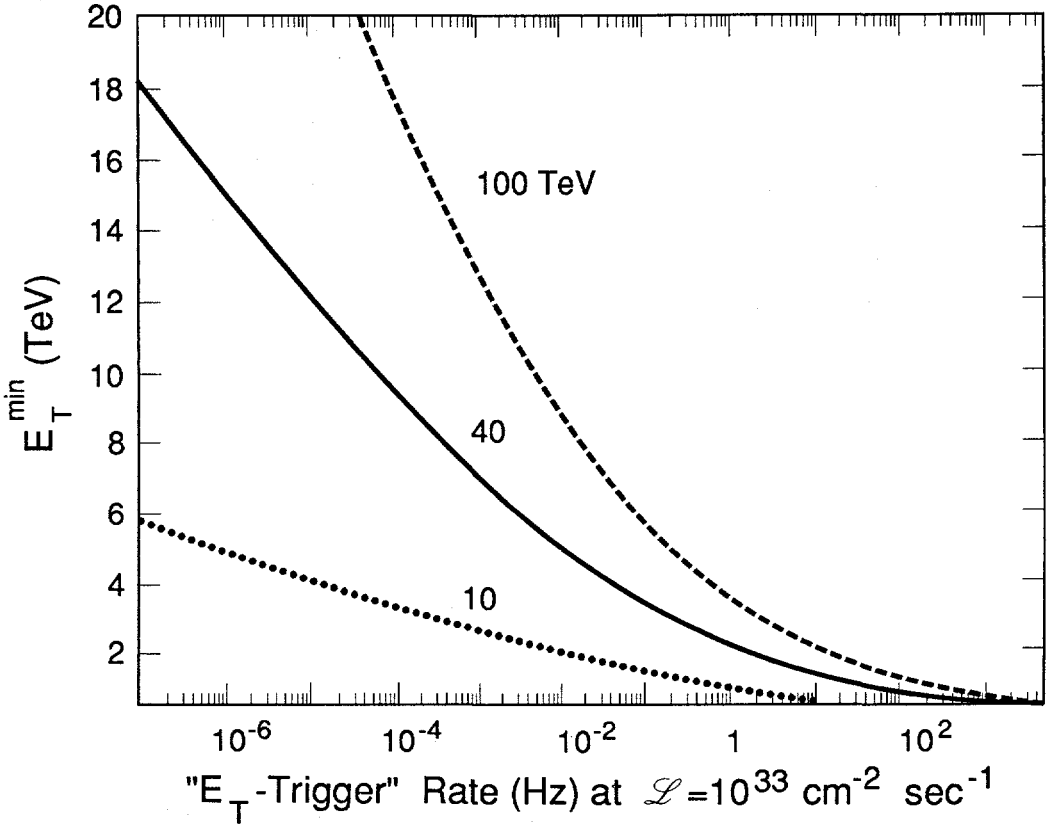


Fig. 2. The minimum transverse energy versus trigger rate for pp colliders at 10, 40 (SSC) and 100 TeV. Taken from Ref. 4.

At the design luminosity of the SSC, a minimum P_T of about 100–150 GeV will be required to reduce the single muon trigger rate to about 1 Hz—see Fig. 3.[6] For electrons the trigger rate is more complicated, since to some extent single jets will appear as an electron, depending upon the sophistication of the electron trigger. On the other hand, the electron rate will not include contributions from π and K decay or punchthrough as will the muon rate. Since the P_T spectrum falls rapidly with increasing P_T, it is my guess that the "electron" threshold will not be too much greater than the muon one. This has not been verified by an explicit simulation.

Heavy Higgs Boson

By heavy Higgs boson, I imply a Higgs boson with mass greater than the ZZ threshold. For masses below about 600 GeV, the decay mode of primary interest will be $H \rightarrow ZZ$, at least one $Z \rightarrow ee$ or $\mu\mu$.[7] For the mode in which both Z's decay to ee or $\mu\mu$, the signal will be a peak in the ZZ invariant mass (Fig. 4). The lepton P_T spectrum from this decay for a Higgs mass of 400 GeV is shown in Fig. 5. Looking at this spectrum it is clear that a single lepton trigger would require too low a threshold for a reasonable rate. Thus the simplest trigger to detect the heavy Higgs in this mode would be a dilepton trigger, with variable P_T cuts on each electron or

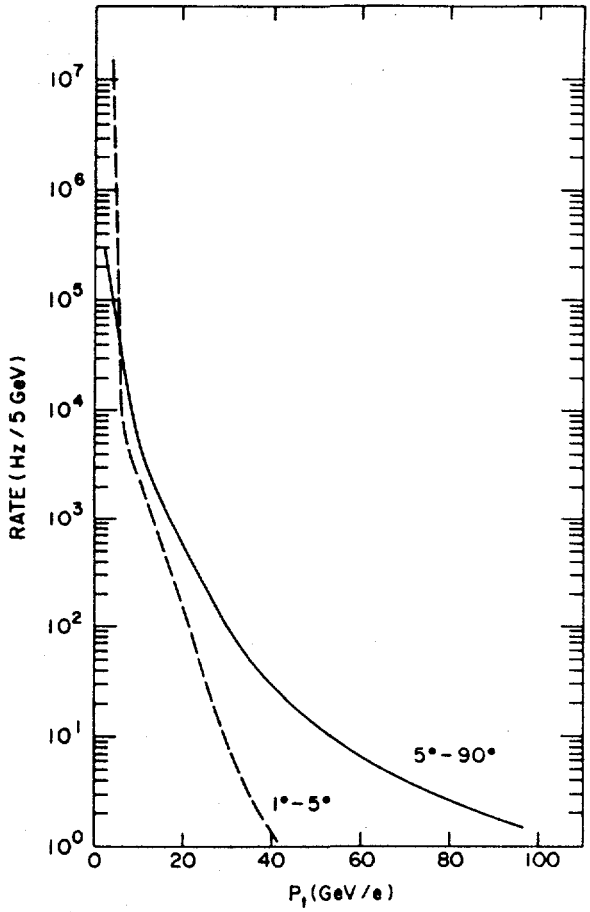


Fig. 3. The rate of particles (muons) exiting outer iron shielding in a typical detector versus the transverse momentum of the particle for two different angular regions. Taken from Ref. 6.

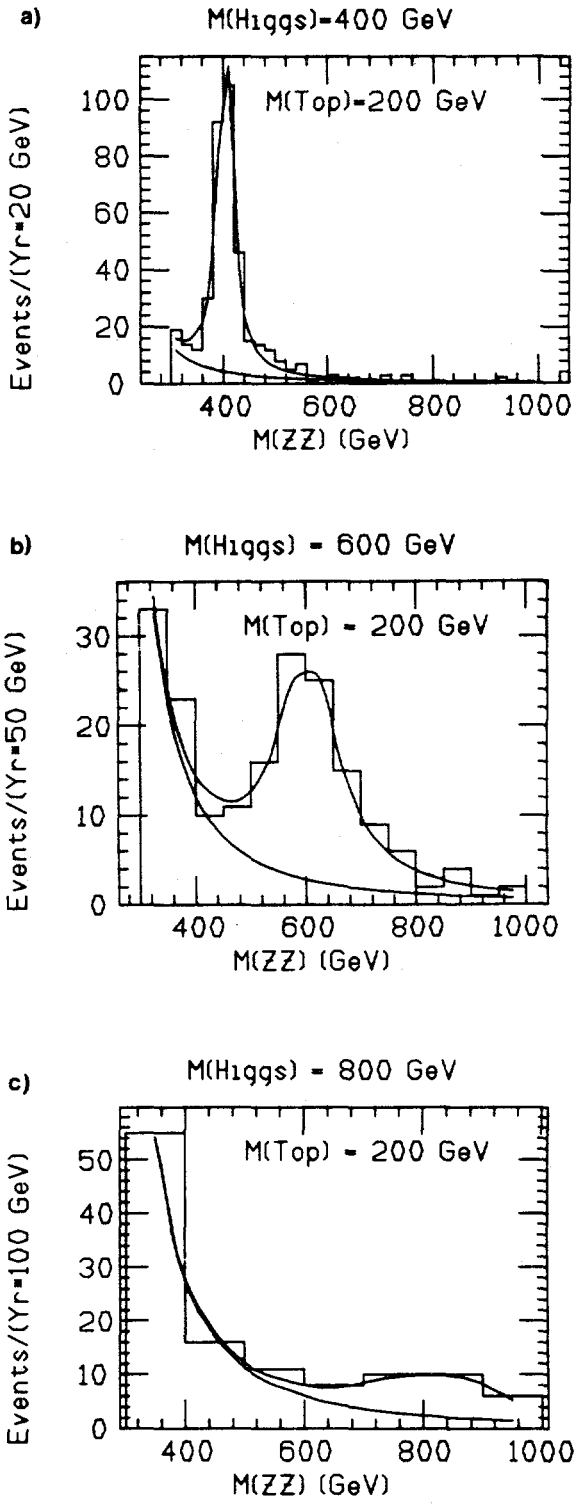


Fig. 4. The ZZ invariant mass from Higgs decay to ZZ, both Z's $\rightarrow ee$ or $\mu\mu$ and from background production of Z pairs for different masses of the Higgs boson and a top quark mass of 200 GeV. Taken from Ref. 7.

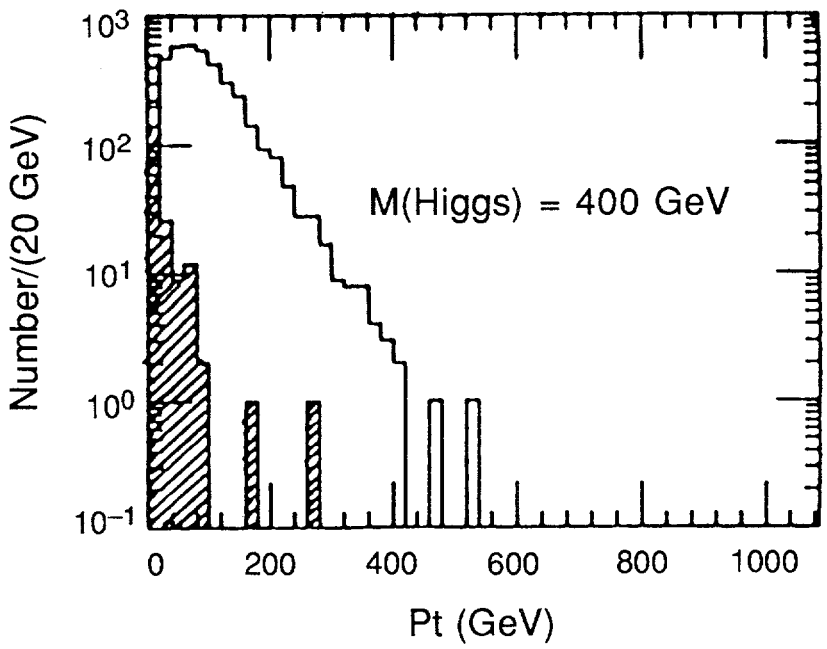


Fig. 5. The P_T distribution of leptons from Higgs events for $M_H = 400$ GeV. The hatched bins indicate the contribution from non-Z sources. Taken from Ref. 7.

muon. Estimates do not exist on this rate but with sufficiently low P_T cuts the rate may be too high at the design luminosity. In this case a "Z trigger" would be required with a variable cut on the P_T of the Z.

For Higgs masses above about 600 GeV it may be interesting to trigger on the $H \rightarrow W^+ W^-, W \rightarrow (e \text{ or } \mu)\, \upsilon$ and $W \rightarrow$ jets decay mode. In this case the trigger could be an isolated electron or muon with some minimum P_T (a reasonable guess would be about 50 GeV) in coincidence with missing E_T (also of about 50 GeV), perhaps along with some jet E_T as well. The rate for this type of trigger is not known.

A Lighter Higgs

If the Higgs is lighter than twice the Z mass, it will be copiously produced at the SSC. A possible, but difficult, mode for discovery is $H \rightarrow b\bar{b}$ and subsequent fragmentation of the b jets.[8] Simple production of the Higgs is overwhelmed by background so people have suggested using production of the Higgs in association with a W via

$$pp \rightarrow H\, W + X,\ W \rightarrow (e \text{ or } \mu) + \upsilon \quad \text{and} \quad H \rightarrow b\bar{b}, \rightarrow \text{jets}$$

This process produces relatively low P_T electrons and muons, jets and missing energy. The charged lepton E_T and missing E_T spectra (from the W decay) for a Higgs mass of 100 GeV are shown in Fig. 6. A possible trigger scenario that has been studied (see Ref. 8) would include

- missing $E_T > 40$ GeV
- an isolated electron or muon with $P_T > 25$ GeV

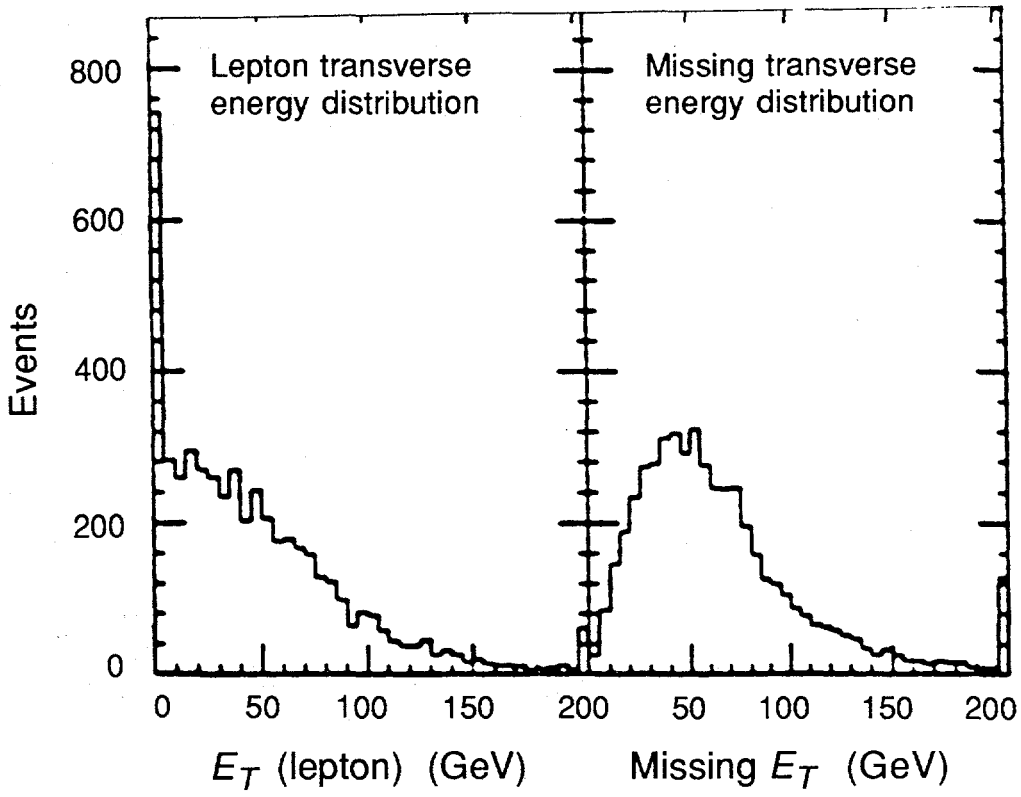


Fig. 6. The charged lepton E_T distribution and missing E_T distribution from W decays in Higgs events as described in the text. Taken from Ref. 8.

- multi-jets, each with some minimum P_T, say 10 GeV and
- possibly a lepton associated with at least one jet with a P_T relative to the jet axis > 1 GeV

The last criterion will be particularly challenging to implement.

New Heavy Quark

If additional quark families exist, they will be copiously produced at the SSC—see Fig. 7.[9] There are many possible decay scenarios for new heavy quarks. For illustrative purposes I have chosen the one described in Ref. 9 and illustrated in Fig. 8. I will not describe in any detail the analysis chain needed to attempt to identify and reconstruct the D—see Ref. 9. For the example of a D mass of 500 GeV, a possible trigger scenario might need to include the combination of

- two isolated leptons with $P_T > 40$ GeV
- missing $E_T > 100$ GeV and
- either total jet $E_T > 75$ GeV or better yet multi-jets, each with some minimum P_T, say 20 GeV

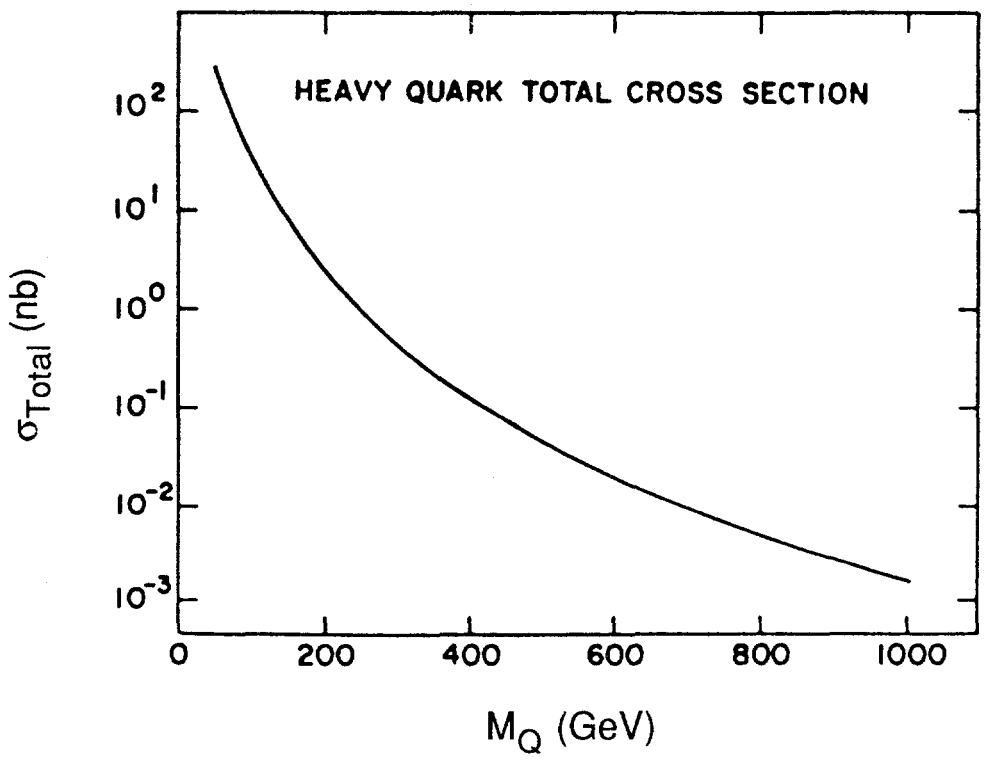


Fig. 7. The total production cross section for new down-like quark pair production at the SSC. Taken from Ref. 9.

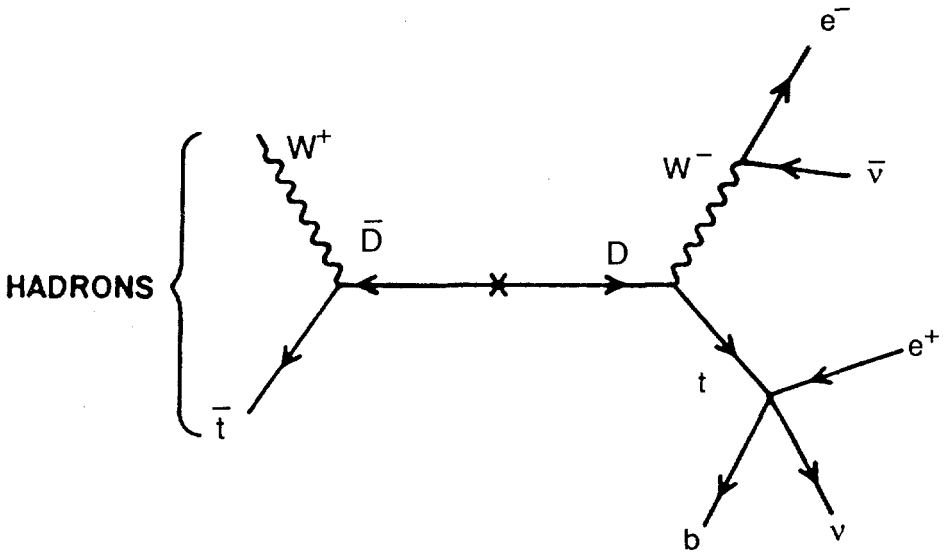


Fig. 8. A diagram for the two lepton decay of a new down-like quark pair, $D\bar{D}$. Taken from Ref. 9.

It may be that just the combination of the first two is sufficient for this mass and above. However, lower masses will be of interest and hence thresholds will need to be lower for lower masses. In this case the full dilepton, missing E_T and jet trigger may be needed. No estimates exist of the trigger rate for such cuts.

Conclusions

Although much more work remains to be done to quantify trigger rates and to refine detection strategies for new physics at the SSC, it is my contention that the basic criteria are sufficiently well known to provide guidance in the design of a trigger system for high P_T physics at the SSC. Given the very large variety of new physics that is possible, flexibility in using and combining

- variable E_T
- jets with variable E_T for each jet and number of jets
- variable missing E_T
- variable energy and number of electrons and muon, including isolation criteria and
- variable energy and number of photons, including isolation criteria

will be the keys to a successful trigger system for an SSC experiment. The ability to vary thresholds and multiplicity criteria in a continuous and selective way will also be crucial to a successful system.

References

1 See the Proceedings of the Workshop on High Sensitivity Beauty Physics at Fermilab, November, 1987.

2 U. Baur, I. Hinchliffe and D. Zeppenfeld, *Excited Quark Production at Hadron Colliders*, From Colliders to Super Colliders, V. Barger and F. Halzen, eds., World Scientific 1987.

3 See, for example, V. Barnes et al., *Compositeness and QCD at the SSC*, in the Proceedings of the Workshop on Experiments, Detectors and Experimental Areas for the Supercollider, R. Donaldson and M. Gilchriese, eds., World Scientific, 1987.

4 C. Quigg, *Opening the High-Energy Frontier*, SSC-199 (1988).

5 See J. S. Whitaker and N. Deshpande, *Physics Parameters for New W's and Z's*, in the Proceedings of the Workshop on Experiments, Detectors and Experimental Areas for the Supercollider, R. Donaldson and M. Gilchriese, eds., World Scientific, 1987.

6 D. Carlsmith et al., *SSC Muon Detector Group Report*, Proceedings of the Summer Study on the Physics of the Superconducting Super Collider, Snowmass, Colorado (1986).

[7] R. N. Cahn et al., *Detecting the Heavy Higgs Boson at the SSC*, in the Proceedings of the Workshop on Experiments, Detectors and Experimental Areas for the Supercollider, R. Donaldson and M. Gilchriese, eds., World Scientific, 1987. I say at least one $Z \rightarrow ee$ or $\mu\mu$ rather than both because it may be useful to detect the mode in which one Z decays to neutrinos.

[8] J. Brau, K. Pitts and L. Price, *Detection of Higgs Bosons Decaying to Bottom Quarks at the Superconducting Super Collider*, OREXP-88-1201, Dec. 1988.

[9] S. Dawson et al., *Heavy Quark Production at the SSC*, in the Proceedings of the Workshop on Experiments, Detectors and Experimental Areas for the Supercollider, R. Donaldson and M. Gilchriese, eds., World Scientific, 1987.

THE CDF LEVEL3 TRIGGER

T. Carroll, U. Joshi
Fermilab

P. Auchincloss, T. Devlin, B. Flaugher, P. Hu, Terry Watts
Rutgers University

K. Ragan
University of Pennsylvania

Abstract

CDF is currently taking data at a luminosity of 10^{30} cm^{-2} sec^{-1} using a four level event filtering scheme. The fourth level, LEVEL3, uses ACP (Fermilab's Advanced Computer Program) designed 32 bit VME based parallel processors[1] capable of executing algorithms written in FORTRAN. LEVEL3 currently rejects about 50% of the events.

Introduction

The CDF detector at FERMILAB has approximately 100,000 channels of electronics, ADC and TDC, to be read out. At the current luminosity of 10^{30} cm^{-2} sec^{-1}, the interaction rate is about 50 KHz. The event size into LEVEL3 is about 150 Kb in and 100 Kb out. CDF has designed a four level trigger system to bring this rate down to a manageable 1 Hz out of LEVEL3 onto tape. The CDF data acquisition and triggering systems are mostly based on FASTBUS interfaced to a VAX cluster[2]. LEVEL3 is, however, VME based and interfaced to both FASTBUS and a micro-VAX. The LEVEL3 design specification called for a 100 node ACP system each with 2 Mb of memory, an input rate of 100 events/sec (event size of 100 Kb), an output rate of 1 event/sec, and an average execution time of 1 sec/node. LEVEL3 is currently taking data with a 58 nodes ACP system each with 6 Mb of memory, an input rate of approximately 3 events/sec, and an average execution time of 8 sec/node.

Level3 within the CDF DAQ system[2]

The LEVEL3 processor farm uses 32 bit ACP processors in VME. Each processor has a Motorola 68020 cpu, a 68881 floating point coprocessor, and 6 Mb of memory (DRAM). The processor farm is interfaced to the FASTBUS network using a FBBC (FASTBUS to BRANCH BUS controller), a BRANCH BUS, and a BVI (BRANCH BUS to VME interface) (figures1 and 2)[3]. The FBBC is a FASTBUS slave and a BRANCH BUS master while the BVI is a BRANCH BUS slave and a VME master. BRANCH BUS transfers use a non-handshake pipelined protocol with a maximum tested transfer rate of 20 Mb/sec. Multiple VME crates are daisy chained on the BRANCH BUS. Each processor functions as an independent unit and processes one event at a time.

The farm is managed by a process called the FARM STEWARD which executes in a dedicated MICRO-VAX which is interfaced to FASTBUS with a QPI (QBUS to FASTBUS interface) and to the farm with a QBBC (QBUS to BRANCH BUS interface) (figure 1). The FARM STEWARD among other things does the following:

(i) downloads algorithms and datbase information into the nodes.
(ii) communicates with the rest of the DAQ system.
(iii) starts up each node after an event has been sent.
(iv) checks to see if a node is done.

Synchronization of dataflow is carried out by the TRIGGER SUPERVISOR, using dedicated control lines and FASTBUS messages, and the BUFFER MANAGER, using FASTBUS messages (figure 3). Data flow in and out of LEVEL3 is synchronized by exchange of FASTBUS messages between the BUFFER MANAGER on one hand and the EVENT BUILDER, FARM STEWARD, and the BUFFER MULTIPLEXER on the other (figure 3).

In the present system the L0/L1/L2 triggers filter the event rate from about 50 KHz to about 3 Hz using fast output signals. The L0 decision is made within 2.5 micro seconds - well within the 3.5 micro seconds between crossings in the 6 bunch mode of operation - and causes no dead time. The L1 decision is made within 7 micro seconds and causes a dead time of about 13.5% at this luminosity while the L2 decision is made in about 20 micro seconds and causes about 6% dead time. Once an event passes L2, data from the front end is digitized and read out by the scanners. The EVENT BUILDER, a set of FASTBUS modules, then reads this data from the scanners, reformats it in standard CDF YBOS[3] format, and pushes the event into an available LEVEL3 node - YBOS is a memory management scheme used in CDF. The EVENT BUILDER, however, does not reformat the LeCroy 1879 TDC data from the tracking chambers. This is done in LEVEL3. This should probably be the method adopted for online event building as it reduces the overhead in managing the message traffic.

Level3 algorithms

Each processor is capable of executing algorithms written in standard FORTRAN 77. The programs are compiled and linked in the processor itself using a compiler and a linker written by ABSOFT. LEVEL3 is basically a software filter that has incorporated into it ideas and concepts derived from offline analysis. The LEVEL3 environment is very similar to that existing offline, and the algorithms are authored by physicists involved in offline analysis. LEVEL3 has access to data from the entire detector (figure 3) - something L1 and L2 don't - so that it is possible to "fine tune" the event selection process in various ways by, for instance, employing more accurate clustering techniques used offline rather than the coarser one used online, having online track reconstruction and using that information for better particle identification, and cleaning the data sample of various kind of noise.

LEVEL3 algorithms, as in offline analysis, uses the YBOS memory management scheme to access event data. An intermediate buffer, a common block, is used for input and output of data from LEVEL3 (figure 4). The EVENT BUILDER writes the event data to this buffer and the BUFFER MULTIPLEXER reads it out. For each event data is copied from the buffer to another location in memory, the local YBOS array(s), followed by the execution of the algorithms. Calibration and Geometry information are copied into their respective YBOS arrays at initialization and maintained throughout the run.

LEVEL3 is organized as a series of independent filters. Each filter module accesses data in the YBOS array(s)[3] and executes. Which filters are to be executed is determined by what triggers are defined in the trigger table downloaded in the nodes at initialization time. A sample LEVEL3 trigger table is shown in figure 5. The filters that have been incorporated into LEVEL3 and used for rejecting events are:

(1) Missing ET

A missing ET threshold of 15c GeV in LEVEL3 reduces thr trigger rate by a factor of 3. Cleanup is a major contributor to this reduction.

(2) Central muon

A muon hit in the central region is matched to a track reconstructed online. This reduces the trigger rate by a factor of 8.

(3) Central electron

This filter requires a cluster with greater than 12 GeV EM_ET with a 6 GeV PT track pointing to the cluster. The HAD/EM ratio for the cluster is required to be less than 0.125, and the cluster requires a matching strip chamber hit. This reduces the trigger rate by a factor of 2.

(4) Central photon

This requires a EM cluster with ET greater than 23 GeV and an isolation cut of 0.4. Cleanup is a major factor, and the trigger rate is reduced by a factor of 10.

(5) Diphoton

This requires that each of the photons have an EM_ET greater than 10 GeV. Cleanup is again a major factor, and reduces the trigger rate by a factor of 2.

(6) Forward muon

In this trigger 5 out of 6 muon chambers in the forward region require hits, and these hits should all point back to the interaction region. Trigger rate in this case is reduced by a factor of 4.

The above filters in LEVEL3 are responsible for rejecting 50% of the data. The input rate into LEVEL3 is on the average 2.6 ub and the output 1.3 ub. Algorithms under study but not yet used in rejecting events
include among others:

(i) central photon filter with a lower threshold
(ii) central electron filter with a lower ET threshold.
(iii) tau filter
(iv) jet filter
(v) plug photon filter

Studies indicate that these algorithms could increase the rejection factor
to about 75%.

The times to execute the above mentioned filter algorithms are very small - of the order of 1 sec or less in total. In addition to running the filter algorithms, LEVEL3 performs a host of other time consuming tasks. They include:

(i) verifying the validity of the YBOS banks in the data written by the EVENT BUILDER. This is done for every event.

(ii) reformatting and compacting LeCroy 1879 TDC data. Portions of it are written in assembler, mainly the time consuming simple loops. The factor gained in speed over the FORTRAN version is about 10. This takes less than 200 msec to execute and is done for every event.

(iii) all tracks with PT greater than 4 GeV are reconstructed. This takes about 4 sec.

(iv) calorimetry and cleanup. This module packs the entire calorimetry data in CDF tower geometry format. Pug EM spikes, "Texas Towers", single PMT noise, and cable pickup are removed. This mod ule takes about 4 sec. to execute.

The LEVEL3 algorithms on average take about 8 sec to execute, most of the time being used for calorimetry, clean up, and tracking.

Conclusion

LEVEL3 has been an integral part of the CDF data acquisition system during the '88 run, and I must say that it has indeed performed remarkably well given the complex environment it has to operate in. There is plenty of room available to enhance its performance which is far from its design goal of handling 100 events/sec with a 99% rejection factor. For LEVEL3 to be able to handle a data rate of 100 Hz, it is quite important that the EVENT BUILDER data rate be increased from its present 5 Hz. Work towards that is in progress, and it seems that a rate of 10 Hz may be attainable in the near future. However, most important of all is the rejection factor of LEVEL3 which is only 50% . Although a 75% rejection factor seems a possibility in the not too distant future, an even higher rejection factor is necessary to be able to deal with a higher data input rate unless the data rate out of LEVEL3 is also increased. This becomes quite

a difficult task given that only 30% of the data out of LEVEL3 is rejected during offline analysis. LEVEL3 at the present takes about 8 secs/event to analyze an event. To handle a higher rate, this time should be decreased substantially - for a 58 node system the execution time has to be brought down to about 5.5 secs for a data rate of about 10 Hz. Except for code used to reformat TDC data, most of the code in LEVEL3 is unoptimized and very little attention has been given to that since it has not been necessary so far. However, this will not remain true in the not too distant future.

References

[1] H. Areti et al, Proc. 23rd Int. Conf. on High Energy Physics, Berkeley, California(1986).

[2] E. Barsotti et al, Fastbus Data Acquisition for CDF, Nucl. Instr. and Meth., A269 (1988), 82-92

[3] T. Carroll et al, LEVEL3 System at CDF, Nucl. Instr. and Meth., A263 (1988), 199-205

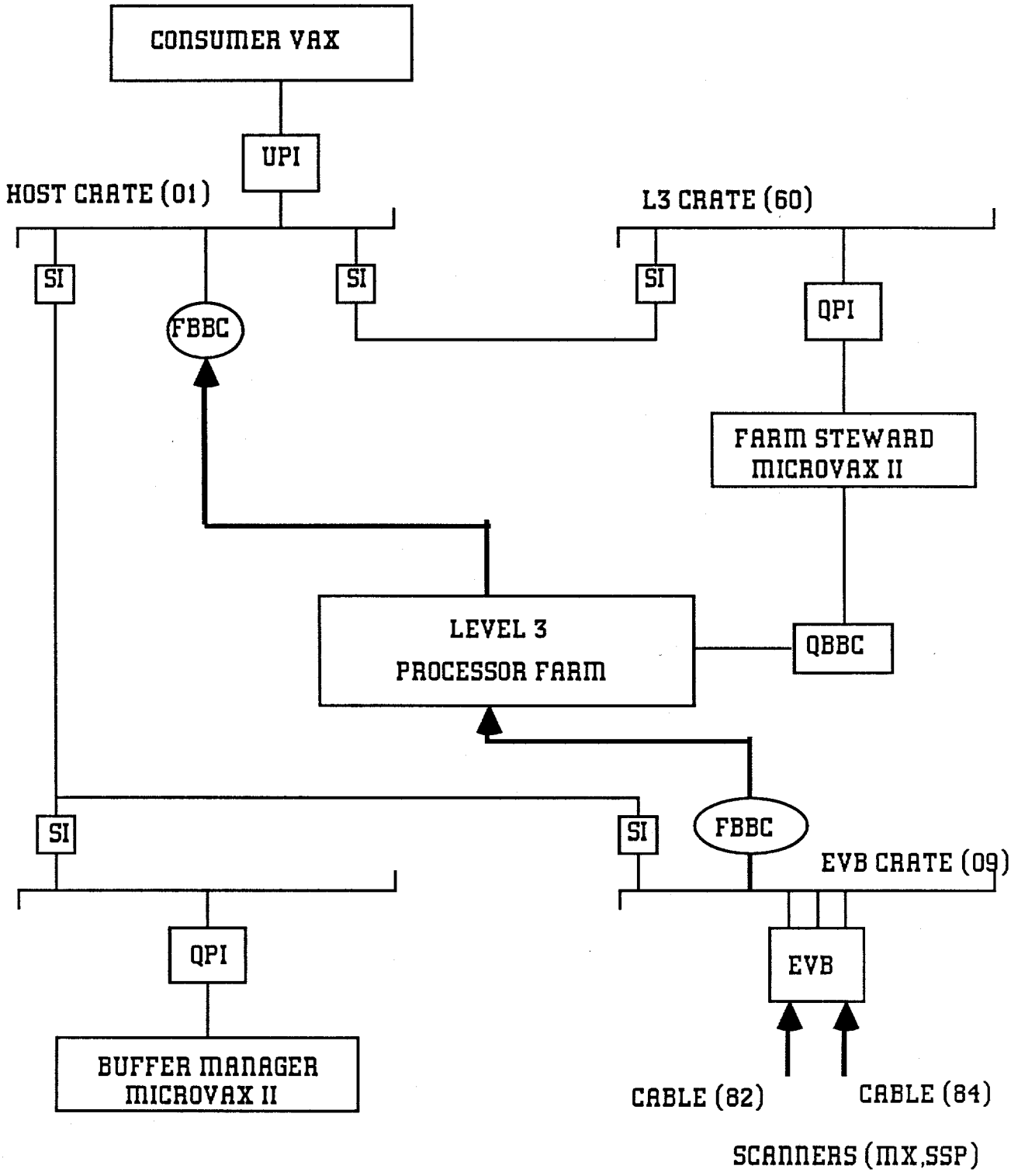
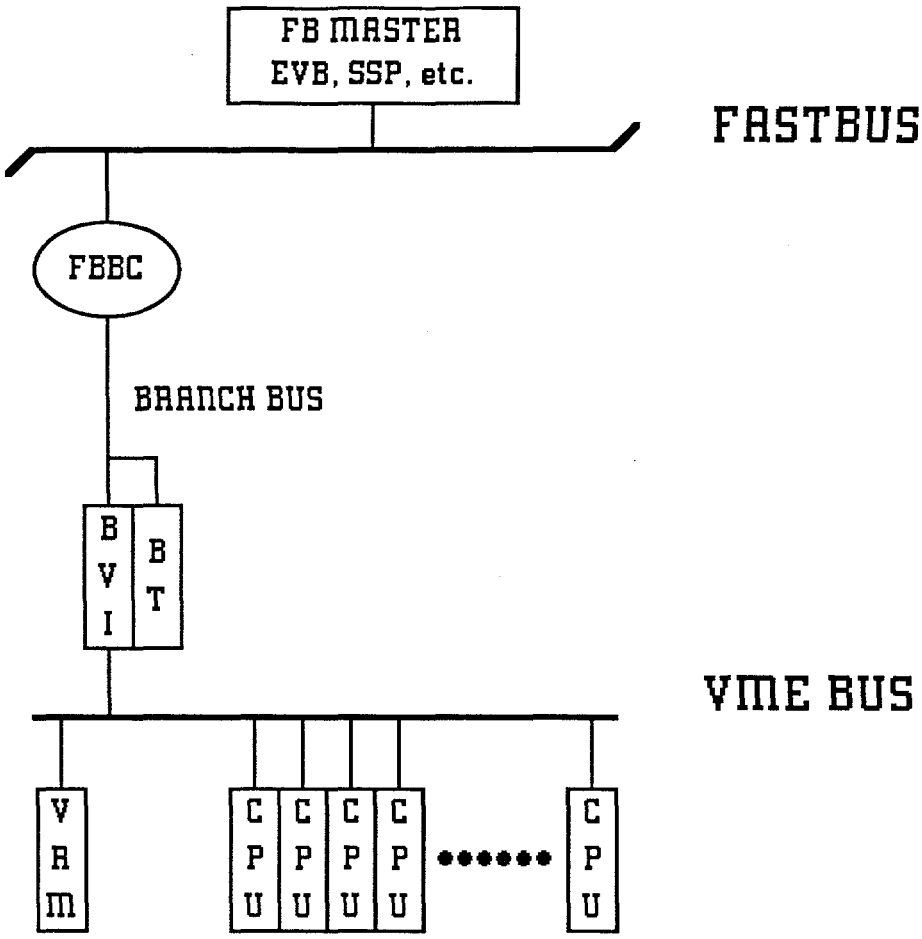

CONSUMER VAX
UPI
HOST CRATE (01)
L3 CRATE (60)
SI
SI
SI
SI
FBBC
QPI
FARM STEWARD
MICROVAX II
LEVEL 3
PROCESSOR FARM
QBBC
SI
SI
FBBC
EVB CRATE (09)
QPI
EVB
BUFFER MANAGER
MICROVAX II
CABLE (82)
CABLE (84)
SCANNERS (MX,SSP)

FIGURE 1



FBBC - FASTBUS BRANCH BUS CONTROLLER
VRM - VME BUS RESOURCE MODULE
BVI - BRANCH BUS VME INTERFACE
BT - BRANCH BUS TERMINATOR

Fig. 2: VME and BRANCH BUS

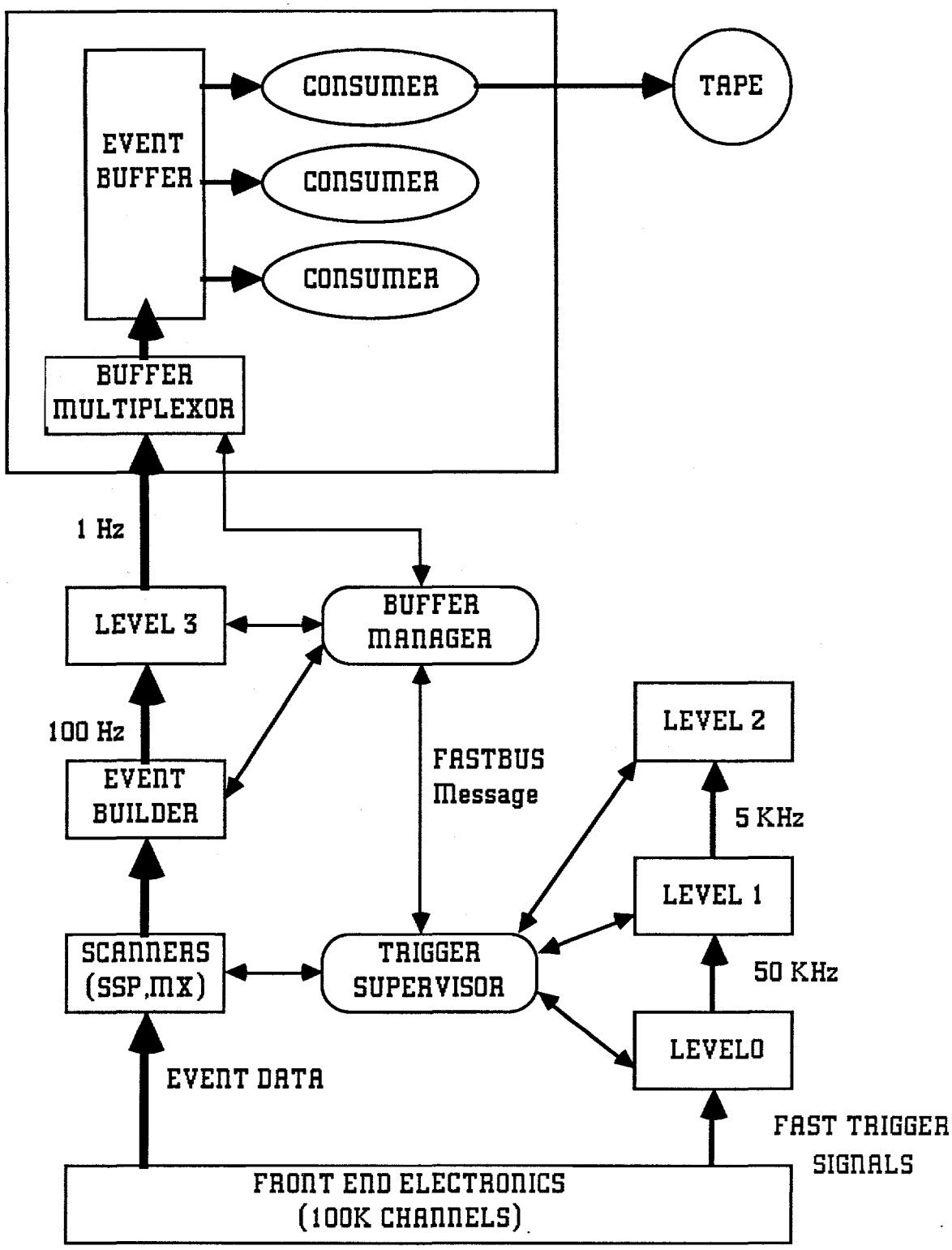


Fig. 3: DAQ PIPELINE

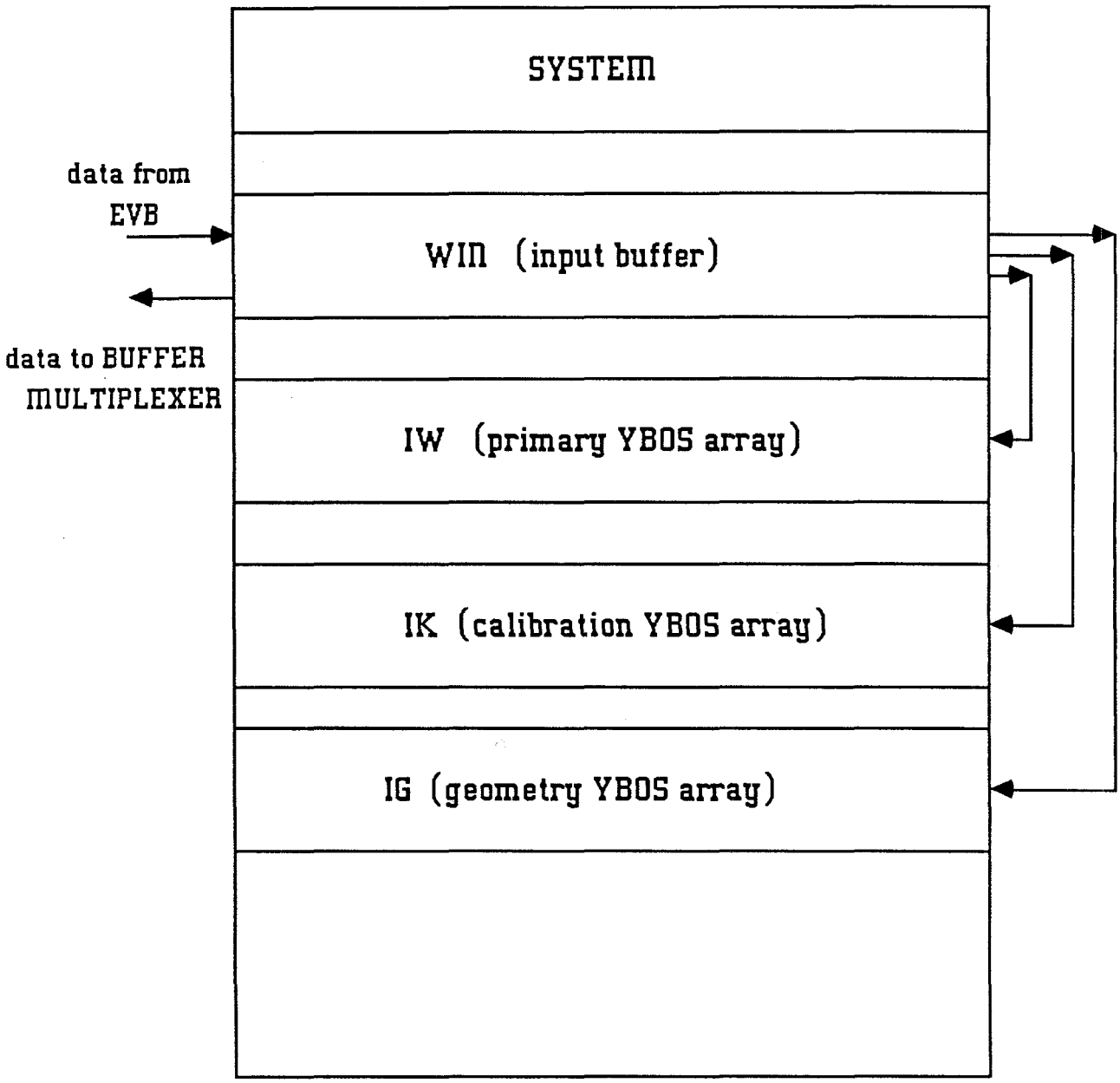

SYSTEM
data from
EVB
WIN (input buffer)
data to BUFFER
MULTIPLEXER
IW (primary YBOS array)
IK (calibration YBOS array)
IG (geometry YBOS array)

figure 4

```
TRIGGER_MISSING_ET_15_V6
    PREREQUISITE MISSING_ET_25_TEX_NOT_FWD          ! demand LEVEL2
                                                      MET
    EXECUTE CALORIMETRY
        PARAMETER TALK_TO
            PEM
            OFF
        END_TALK
    EXECUTE NCABLE                                  ! cleanup
    EXECUTE FHA_CABLE                               ! cleanup
    EXECUTE FILT_GAS                                ! cleanup
    EXECUTE GETL3                                   ! compute MET in L3
    SELECT GETFLT
        PARAMETER TALK_TO
            MISSET  15.0
        END_TALK
```

figure 5

SOME LESSONS FROM THE DØ TRIGGER DESIGN RELEVANT TO THE SSC

Maris Abolins, Daniel Edmunds, Philippe Laurens and Bo Pi
Department of Physics and Astronomy
Michigan State University, East Lansing, MI 48824

ABSTRACT

We discuss features of the Trigger Design for the DØ experiment at the Fermilab Tevatron Collider that have applications to designs for triggers at the SSC.

INTRODUCTION

We have spent the last four years designing and building major parts of the trigger and data acquisition system for the DØ experiment at the Tevatron proton-antiproton collider. Our primary responsibility has been for the First Level Trigger Framework and the First Level Calorimeter Trigger, shown in Figure 1 which gives an overview of the DØ triggering and data acquisition system.[1,2,3,4,5] In designing the First Level Calorimeter Trigger we have had to face and overcome many of the problems that will confront us at the SSC. In the narrative below we start with a brief overview of the features of the DØ trigger emphasizing those relevant to the SSC environment, discuss specific items of particular importance and finally offer some conclusions based on our experience.

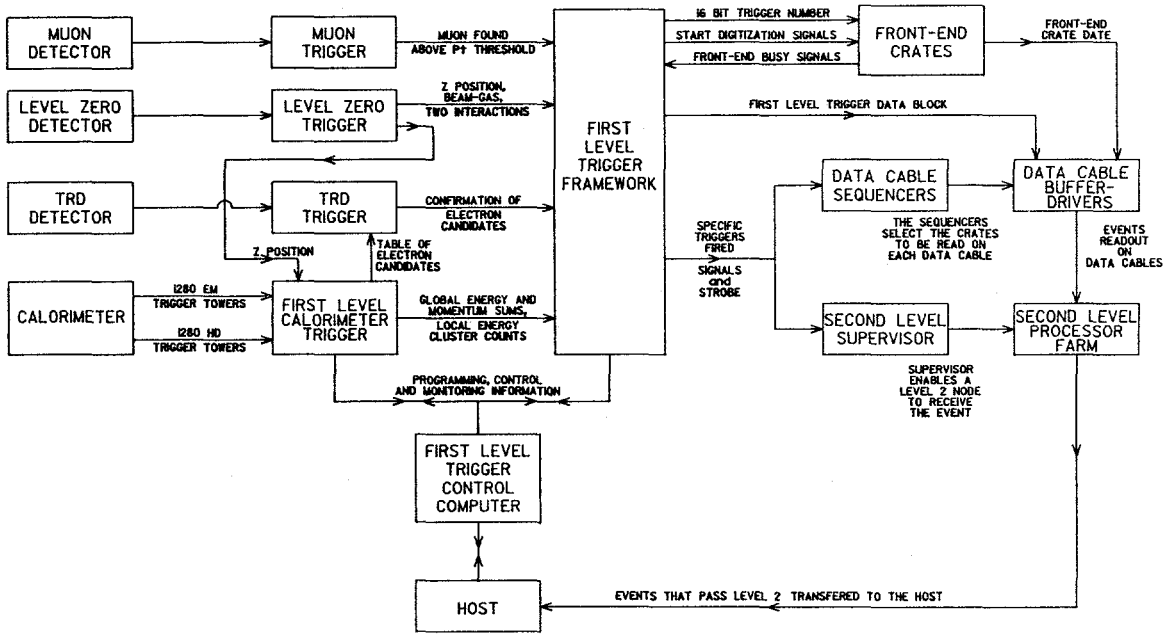


Figure 1. A block diagram of the DØ triggering and data acquisition system.

THE FIRST LEVEL CALORIMETER TRIGGER FOR DØ

During the design phase for the First Level Calorimeter Trigger for DØ, a number of proposals to raise the luminosity of the collider were put forth by the Fermilab management. As there was some uncertainty as to which of these would ultimately be implemented, it seemed prudent to design a trigger able to function in a deadtime-less mode under the most extreme of these scenarios. This led to a fully pipe-lined trigger capable of operating with as many as 144 bunches of protons (and antiprotons) in the machine with a resulting inter-bunch spacing as low as 132 ns. This trigger system can take input from as many as 32 distinct detector front-end sections and provides for the definition of 32 separate "specific triggers". A specific trigger can be either a test input trigger or a trigger defined by the detector elements, such as the transverse energy depositions in any of the 1,280 electromagnetic or hadronic towers that comprise the DØ calorimeter. Any combination of the 32 specific triggers can be in operation at the same time and each specific trigger can result in the digitization of any combination of the 32 separate detector front-end sections. In addition, this trigger system incorporates a number of features to aid in monitoring of the experiment and in the installation and commissioning of the detectors and the main acquisition system. Complete events that pass this first level trigger are sent to the Level 2 trigger system and only upon passing the trigger decisions there are they recorded by the host computer.

IMPORTANT DESIGN CONSIDERATIONS

Flexibility

Since it is difficult to anticipate in detail all of the functions that a trigger system may need to have at a new accelerator such as the SSC, it is important to design in some flexibility that will permit the experimenters to react to unexpected events. This could be the need for a different kind of trigger, the need to add specialized processors, etc. This is more easily accomplished if some "hooks" are built into the design from the beginning. For example, in the DØ First Level Trigger we have made the 9 bit total transverse energy signals from all 1,280 trigger towers easily accessible for possible future use by more sophisticated processors.

Understanding trigger acceptances and efficiencies

An obvious point to make is that trigger acceptance and efficiency must be understood in order to extract meaningful cross-section measurements. Care must be taken that all trigger conditions during a run can be modeled, that they are well documented and that the effects of the details of the bunch structure are properly taken into account. This can become particularly complicated if the bunches in the machine are not uniformly distributed but, are arranged in super-bunches with large spacings between them. The dependence of the trigger on the instantaneous rate as well as on the history of previous crossings must be carefully analyzed and understood.

Testing the trigger

Because the SSC trigger system will be large, complicated and will probably be pipe-lined with many events resident in it at any given time, it is not sufficient to test just the individual electronics cards but rather it is necessary to be able to make complete system tests under simulated running conditions. In the DØ Trigger this is accomplished by loading one or two events into the "front-end" of the trigger and then cycling the trigger at full speed for one or two cycles. Both the final results and all of the intermediate results are then available to the trigger control computer. A number of system-level problems were found in this way. These were problems that did not show up during the debugging and production testing of the individual cards and typically took much longer to understand and solve than the relatively simple problems associated with individual cards. To facilitate debugging of the system, all registers in the DØ trigger system, that are used to control the trigger operation, can be read and written into by means of a control computer

Monitoring the trigger

It is essential that the SSC trigger system be able to monitor its own operation. This is true for a number of reasons: the trigger is the last point in the experiment that sees all of the beam crossings (including those that do not pass), the trigger must provide the operators with deadtime information and the trigger must provide information about the integrated luminosity to which each of the physics triggers has been exposed. To meet these goals, the monitoring functions of the trigger system must be designed in from the beginning and not added on later. Trigger monitoring must be incorporated into the other monitoring functions of the experiment including the archiving of monitoring data for later retrieval during data analysis.

The DØ trigger system includes a large number of scalers for monitoring its operation. The contents of some of these scaler are read out with the event data stream and all of them are read out by the experiment monitoring system. The scalers that are part of the event data stream are those that are most important for the off line physics analysis. These include scalers for each type of event indicating the number of beam crossings to which it has been exposed and the number of times it has fired. In addition there is a matrix of scalers covering each type of event and each source of dead time.

Events from nearby crossings

Because of the close spacing in time of the SSC beam crossings, with most current calorimetry techniques there will be overlap of information from different beam crossings in at least some of the detector elements. If the trigger information has time resolution shorter than the crossing time, then it can help "un-tangle" this overlap in later off-line analysis. Because the trigger system is built with a pipeline architecture it can store copies of input data from the detector elements in FIFO memories. Depending on the details of the triggering and data acquisition systems, it may be necessary to read out detector information from crossings both before and after the one resulting in a trigger.

The DØ trigger incorporates a double buffered memory system that is eight crossings deep. Information from a selected range of history is read out in the event data stream. The geometric resolution of this data, in pseudo-rapidity and azimuthal angle, is twice as coarse as that of the main acquisition system of the calorimeter.

Accelerator parameters

The SSC Trigger system should be able to take advantage of some of the details of the arrangement of the bunch batches in the SSC[6]. In the SSC there are a number of gaps between the batches of bunches where there are no beam crossings. Each batch is 17.5 μs long and contains 1,092 bunches. There is a 0.15 μs gap between each of the 5 batches. In addition there is one gap of 3.1 μs (the abort gap) between the 15th batch and the 1st batch. These periods during which there are no beam crossings could be used by an SSC trigger system for such activities as monitoring or moving data from a pipeline into the acquisition system. It is important to note that because of these gaps the trigger system does not need to process a continuous stream of crossings but rather it has frequent periods where it can accommodate other activity.

In some of the higher luminosity scenarios for the Tevatron the DØ trigger and data acquisition will take advantage of the gaps between batch of beam crossings. In all cases the trigger will fully analyze all of the data from each beam crossing and maintain a history record of adjacent crossings in double buffered FIFO memory. The acquisition system will use each of the gaps to obtain a fresh baseline sample. In order to make use of these batch to batch gaps one needs to understand early in the design cycle of the trigger system the details of the beam structure. The trigger design must be flexible enough to be able to accommodate possible changes in the beam structure as the machine construction itself advances.

Connection with front-end systems

The connection between the front-end systems and the trigger system in an SSC experiment is more complicated than in a conventional experiment. Because of this many of the design optimizations will need to be done in a global way taking into consideration the characteristics and costs of both the front end system and the trigger system. The most obvious example of this is the relationship between the length of time required to form the trigger decision and the number of data storage elements that are required for each detector channel.

Detector partitioning and multi-stream operation

An SSC trigger - data acquisition system will face a problem found in many HEP experiments, namely that the recording of data from an interesting beam crossing is blocked because just one of the detector front-end electronics systems is still busy processing data from a previous trigger. To help minimize this problem an SSC trigger can incorporate a programmable matrix relating trigger types to the detector front-end sections that are read out in response to each of these trigger types. In this way an individual trigger type need only be restricted by the busy status of the front-end sections that are associated with it. This flexible mapping of trigger types onto detector front-end sections will be especially useful during installation and commissioning.

We have designed the DØ trigger to incorporate such a mapping and to make separate decisions about the physics merits of a beam crossing and the readiness of the acquisition system to proceed with each of the different trigger types. The following items are analyzed for each of 32 trigger types in making the decision about the readiness of the acquisition system:

1) Are the front-end electronics systems of all the detector elements that will be read out with this event type ready to proceed or are they currently buffering data from previous triggers?

2) Is there a processor node available in the second level trigger processor farm in the queue of nodes that receives this type of event?

3) Has the acquisition system requested that this type of event be prescaled?

4) Has the acquisition system disabled the data stream for this type of event?

5) Has this type of event been programmed to be "auto-disabled" (the acquisition system must explicitly request each time it wishes to receive an event of this type)?

Digital vs. analog operation

An SSC trigger should be an all digital system incorporating as the first stage an ADC for each input followed by a pipeline architecture system. By design the skew in the signals arriving at the trigger can be taken out in the first stages of the pipeline. The pipeline needs to be different from that of the main acquisition system to minimize the cost.

In the early stages of design much of the DØ trigger was an analog system. This was abandoned after it became clear that a digital system offered better noise immunity, greater flexibility and ease of operation at high beam crossing rates. It would be a very difficult problem to design and maintain a 60 MHz analog trigger system for the SSC.

Trigger location

A major design question for an SSC trigger is the choice of the side of the shielding wall on which to place the trigger. The advantage of shorter cable runs obtained by putting the trigger on the accelerator side of the shield wall may be overcome by the disadvantage of limited access. It is probably unrealistic to contemplate placing a system as complicated as an SSC trigger in an area that is inaccessible for long periods of time. The failure of a few channels of electronics can probably be tolerated by most well-designed experiments but, the failure of even a portion of a trigger system would almost surely incapacitate the experiment.

By careful physical placement of cable runs and the trigger system components the time penalty for putting the system on the "people" side of the shield wall could be kept to something like 600 ns or about 40 beam crossings. There may be other reasons that will require the pipeline length to be this long or longer so this may not be the overriding consideration.

Our choice for DØ was to place the trigger electronics outside the shielding wall.

Monte-Carlo simulation

Starting with the earliest stages, the design of a trigger for the SSC will need to be guided by a Monte Carlo simulation of detector response and trigger operation. Most likely the "full official" detector simulation Monte Carlo will not be available when the trigger designers start their work and thus they will be faced with creating their own detector simulation. It will be important to understand how detailed their detector simulation needs to be in order to guide their design of the trigger system. Most likely not all of the detector parameters will be fixed at this time which will also contribute to uncertainties in the simulation results.

With the benefit of hind sight provided by comparing our early Monte Carlo simulation[7] of the detector and trigger operation to the current full GEANT[8] detector simulation and full trigger model, we have found only small differences in those topics of concern to the trigger[9,10,11]. The

early detector simulation used ISAJET[12] events and a bulk parametrization of the detector with simple geometry. It emphasized flexibility and speed and did not attempt to simulate features of the detector that were not directly relevant to the early trigger design trade-offs.

CONCLUSIONS

We have designed and built a trigger for the DØ experiment that can operate without incurring deadtime in a collider where successive collisions are separated by as little as 132 ns. The basic clock frequency that we use is 65 ns which is within a factor of four of the crossing frequency at the SSC. We note that ten years ago this would have been extremely expensive and difficult to do. Given the rate of development of commercial fast electronics it seems reasonable to assume that extending the design parameters a factor of four should be possible in the ten years that it will take to build the SSC. It is our conclusion that a speeded-up DØ design could be used as a starting point for the design and development of a trigger suitable for an SSC detector.

REFERENCES

[1]DØ Design Report. Fermilab Publication, 1985.

[2]A High Luminosity Trigger Design for the Tevatron Collider Experiment in DØ, Maris Abolins, Daniel Edmunds Philippe Laurens, James Linnemann, and Bo Pi, Proceedings of the IEEE Nuclear Science Symposium, November 1988, to be published.

[3]DØ Trigger Framework, Maris Abolins, Daniel Edmunds and James Linnemann, DØ Note 328, February 24, 1986.

[4]The Level One Framework: DØ Note 328 Revised, Maris Abolins, Daniel Edmunds, Philippe Laurens, James Linnemann, and Bo Pi, DØ Note 705, May 26,1988.

[5]The Level One Calorimeter Trigger for DØ, Maris Abolins, Daniel Edmunds, Philippe Laurens, James Linnemann, and Bo Pi, DØ Note 706, October 25,1988.

[6]Conceptual Design of the Superconducting Super Collider, SSC Central Design Group, March 1986

[7]DØ Calorimeter Simulator Version II, January 1985 written by S. Linn.

[8]GEANT, a Monte Carlo program developed at CERN for the simulation of particle interactions in detectors.

[9] Level One Trigger Studies for DØ, Maris Abolins, Daniel Edmunds and James Linnemann, DØ Note 259, September 12,1985.

[10]Single Tower Trigger Studies, Maris Abolins, Daniel Edmunds and Bo Pi, DØ Note 456.

[11]On the Size of Hadronic Trigger Clusters, Maris Abolins, Daniel Edmunds and Bo Pi, DØ Note 457.

[12]ISAJET, a QCD-inspired Monte Carlo event generation program written by F. Paige and S. Protopopescu of Brookhaven National Laboratory.

AMY TRIGGER SYSTEM

Yoshihide Sakai

National Laboratory for High Energy Physics
Tsukuba, Ibaraki, Japan 305

ABSTRACT

A trigger system of the AMY detector at TRISTAN e^+e^- collider is described briefly. The system uses simple track segment and shower cluster counting scheme to classify events to be triggered. It has been operating successfully since 1987.

INTRODUCTION

The AMY detector,[1] one of the experiments at the TRISTAN e^+e^- storage ring of the Japan National Laboratory for High Energy Physics (KEK), is a general purpose detector that is characterized by a very high magnetic field (3 Tesla) and its compact size. The detector is optimized for lepton and photon identification and also for future micro-beta operation. The detector has been operating successfully since January 1987.As shown in Fig. 1, the AMY detector consists of Inner Tracking Chamber (ITC), Central Drift Chamber (CDC), and electromagnetic Shower Counters (SHC) inside a 3 Tesla solenoidal magnet coil. The coil is surrounded by a steel flux return yoke followed by a drift chamber/scintillation

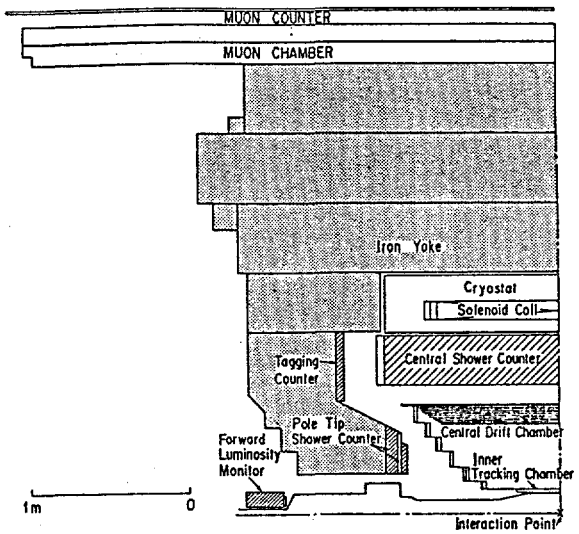


Fig. 1 The AMY Detector

counter muon detection system. In the endcap region are a Luminosity monitor (LUM), a Pole Tip Shower Counter (PTC), and a Tagging Counter (TC).

The main components of AMY trigger system are SHC (energy triggers), ITC, and CDC (track triggers). They are designed more or less independent and give redundant triggers so that trigger efficiency for given type of events can be calculated. This also gives the capability to take data with sufficient trigger efficiency even if one of the components has a problem. Besides central components, endcap detectors also provide the triggers for luminosity measurement and electron tagging. The AMY trigger system was developed mainly by groups of Univ. of South Carolina, Rutgers, and U.C.Davis. In the following, CDC and SHC triggers are described briefly.

CDC TRIGGER

AMY CDC is a cylindrical multi-wire drift chamber with hexagonal cell structure and consists of 6 disks as indicated in Fig. 1. Each disk contains super-layers of axial wires (4 layers) and stereo wires (3 layers) except disk 1 which has 5 layers of axial wires. The trigger system uses hit information of axial layers only.

The philosophy of the CDC trigger is to count the number of "radial" track segments in each disk and then classify them into event types (Multi-track, 3-4 track, 2 track event etc..). No correlation between disks is used at present stage. The system consists of 3 stages: track segment finder, track segment counting, and classification.

Track Segment Finder

Each super-layer (4 wire layers) is divided into small sections with 14 wires as

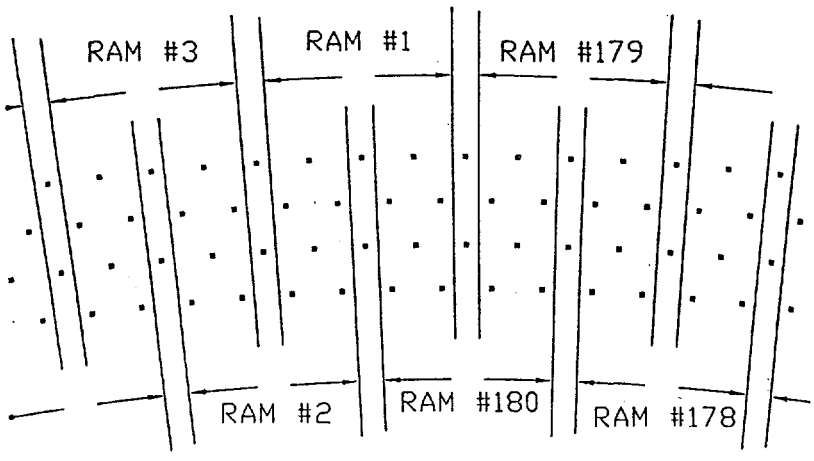


Fig. 2 Mapping of the track segment finder sections in one super-layer

shown in Fig. 2. Each section has 6 overlapping wires with neighboring sections. The hit information of each section is fed into a 16K $\times$ 1 bit RAM whose memory is preloaded so that it gives output "1" only if the section has at least one track pattern corresponding to "radial" track. The requirement of track to be "radial" effectively filters out the events with only low transverse momentum tracks which are likely background events. RAM's are implemented in the personality cards which sit on the back of the Fastbus TAC modules.[2] A TAC module receives wire signals from CDC and stores time information for digitization and also gives hit patterns through auxiliary connectors. The patterns for RAM are preloaded through TAC module using "test pattern" function.

Track Segment Counting

Track segment counting part is constructed using LRS CAMAC ECLine logic units.[3] Outputs of each section (RAM) of track segment finder are fed into LRS 4532 32-input Majority Logic Units (MALU) which give analog majority outputs. Analog majority outputs are summed by daisy chaining within one super-layer and fed into LRS 4504 Quad 4-bit FADC's which provide the digitized number of track segments in one super-layer. Due to the limited number of inputs in next stage, 4-bit numbers (0-15 or more) are reduced into 3-bit numbers (0-7 or more) using LRS 4508 Dual 8-In/8-Out Programmable Logic Units.

Event Type Classification

The numbers of track segments for super-layers 2 to 6 (total 15 bits) are fed into two LRS 2372 Memory Look Up (MLU) modules which give the event type class (4-bits, class 0-15) according to patterns of numbers of track segments in each super-layer. The event types are; multi-track (5 or more tracks), 3-4 track, two track, single/bachelor V (one track growing to two tracks) track events. Same type of events are classified different classes depending on the quality, which are used differently with different combination to other detector components in the final global trigger processing. Information of inner most super-layer (2 bits) is passed separately.

SHC TRIGGER

AMY Shower Counter consists of 20 alternating layers of lead and gas proportional tubes made from resistive plastic. Signals from both anode wires and cathode pads are read out with different ganging. Cathode pads have fine segmentation (total 11K channels) and not used in trigger. Anode wires are ganged into 48 in ϕ within each layer. Among 20 layers, signals from outer 4 layers in same ϕ tower are summed and used to trigger minimum ionizing particles. Sums of inner 16 layers are used to trigger EM showers.

The philosophy and firmware for SHC trigger is quite similar to CDC trigger.

48 summed signals each for outer 4 layers (MIN.I) and inner 16 layers are fed into Phillips CAMAC 16-channel latching discriminators which provide the analog majority output proportional to number of outputs. Signals from inner layers are split into 2 sets of discriminators which have different threshold levels (High and Low energy thresholds). Majority outputs of each 48 inputs are fed into LRS 4504 FADC and 2372 MLU which counts number of clusters for MIN.I, High, and Low energy thresholds and then classify events just in the same way as CDC trigger case.

Total sum of all inner 16 layers is also formed and used for total energy trigger separately.

FINAL TRIGGERS AND SUMMARY

Outputs from each component are combined together and final triggers are formed. Beside the information described above (6 bits of CDC and 5 bit of SHC), information from ITC (2 bits) and Endcap detector (4 bits) are also available. This stage is constructed also using several LRS 2372 MLU modules. The final triggers are formed so that they are more or less dominated by one detector component in order to keep redundancy. However, combining with other detectors are essential to reduce the trigger rate keeping good efficiency, especially, for example two track triggers which requires CDC, ITC and SHC MIN.I information.

The processing time is about 1.2 μsec total (600 nsec to wait CDC signal + 600 nsec for logic processing) and fast enough comparing with beam crossing interval of TRISTAN machine (5 μsec for 2 $\times$ 2 bunches). The trigger system has been operating with the rate of 1 - 3 Hz for luminosity of 1 - 10 $\times$ 10^{30} cm^2/sec.

AMY trigger system described above corresponds to 1st level (prompt) trigger and higher level triggers are not necessary at present condition. For SSC, certainly this system is too simple to apply though the idea based on might be similar even for SSC case.

The author wishes to thank C.Rosenfeld and S.Lusin, who have been deeply involved in AMY trigger system, for giving detailed and helpful information about the system.

REFERENCES

1) 'Major Detectors in Elementary Particle Physics', Particle data Group, LBL-91, 1985; H.Sagawa *et al.*, Phys. Rev. Lett. **60**, 93 (1988).

2) M.Ikeno *et al.*, IEEE Trans. on Nucl. Sci. Vol. **33**, 779 (1986)

3) 'ECLine Programmable Trigger Product Summary' LeCroy Corporation

THE ZEUS CALORIMETER FIRST LEVEL TRIGGER

Wesley H. Smith
Physics Department, University of Wisconsin, Madison, WI. 53706 USA

ABSTRACT

The design of the Zeus Detector Calorimeter First Level Trigger is presented. The Zeus detector is being built for operation at HERA, a new storage ring that will provide collisions between 820 GeV protons and 30 GeV electrons in 1990. The calorimeter is made of depleted uranium plates and plastic scintillator read out by wavelength shifter bars into 12,864 photomultiplier tubes. These signals are combined into 974 trigger towers with separate electromagnetic and hadronic sums. The calorimeter first level trigger is pipelined with a decision provided 5 μsec after each beam crossing, occurring every 96 nsec. The trigger determines the total energy, the total transverse energy, the missing energy, and the energy and number of isolated electrons and muons. It also provides information on the number and energy of clusters. The trigger rate needs to be held to 1 kHz against a rate of proton-beam gas interactions of approximately 500 kHz. The summed trigger tower pulseheights are digitized by flash ADC's. The digital values are linearized, stored and used for sums and pattern tests.

INTRODUCTION

The Zeus detector will measure[1] nucleon structure functions in neutral and charged current reactions at Q^2 up to 30,000 GeV^2/c^2. This is more than two orders-of-magnitude higher than previously obtained. This will enable a study of electron and quark substructure, the properties of the electroweak current, and QCD interactions over distances smaller by more than a factor of ten. If these measurements agree with the Standard Model, then one can measure the couplings between fundamental constituents. If these measurements do not agree, then this may reveal new forms of interactions due to internal structure of quarks or leptons, new higher-mass W and Z particles, or unexpected properties of the weak, electromagnetic,or strong forces. The Zeus experiment, with its high resolution on jet energy, is designed to precisely measure deviations of structure functions from the Standard Model.

The Zeus detector will search[2] for leptoquarks, heavy leptons, heavy quarks, Majorana neutrinos and SUSY particles at masses up to 180 GeV/c^2. The mass range for these searches is about a factor of two higher than the range of the experiments at LEP and SLC, but not necessarily as clean. The mass range of the Zeus experiment is lower than the experiments at the SppS and Tevatron, but is significantly cleaner. What is more important is that the collisions of leptons with quarks or gluons are unique and different from the collisions of quarks and gluons with quarks and gluons, or the collisions of leptons with leptons. Each advance in the energy scale of lepton-nucleon scattering has brought exciting discoveries, such as the discovery of scaling violation.

The layout of the Zeus detector[3] is shown in Figure 1. The essential elements are a central track detector (CTD) plus a transition radiation detector (TRD) and planar chambers (FTD, RTD) within a thin magnetic solenoid (SOLENOID), an electromagnetic (EMC) and a hadron calorimeter (HAC) surrounding the coil over the full solid angle, a backing calorimeter (BAC), barrel and rear muon detector (MU), and a forward muon spectrometer.

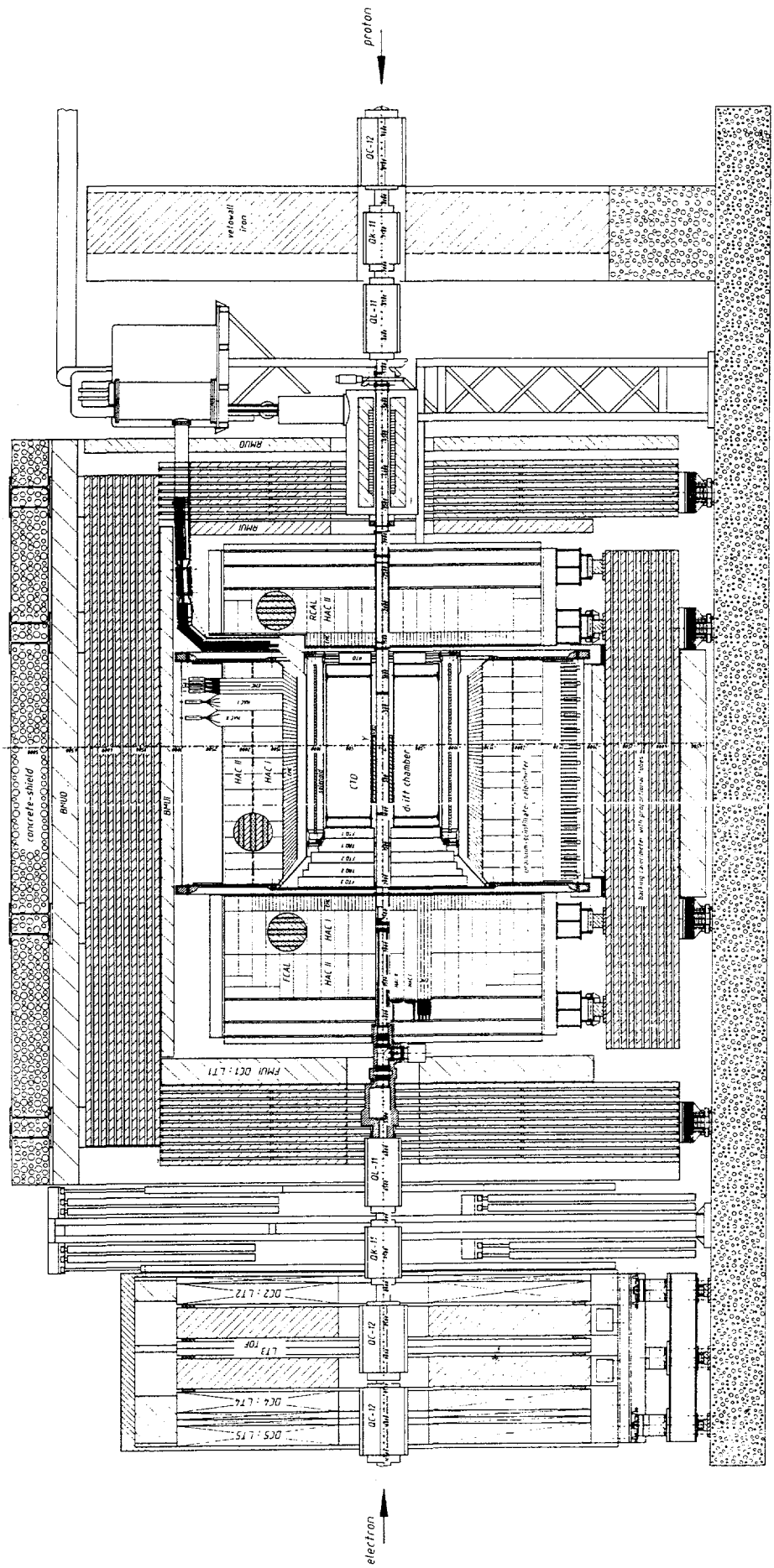


Figure 1. *Section of the Zeus Detector along the beam.*

The Zeus calorimeter consists of depleted uranium plates interleaved with plastic scintillator in order to achieve compensation and the best possible energy resolution for hadrons. The scintillator plates form towers which are read out on two sides with wavelength shifter bars, light guides and photomultipliers. Tests performed at CERN during 1986 demonstrated[4] that the design resolution ($35\%/\sqrt{E}$), heretofore not achieved in a large detector, can be met.

The layout of the Zeus calorimeter is shown in Figure 2. The solid angle coverage is 99.8% of the forward hemisphere and 98.0% of the backward hemisphere. The calorimeter is divided into three main components: the forward calorimeter (FCAL) covering polar angles between θ = 2.3° and 39.4°, the barrel calorimeter (BCAL) covering angles between θ = 39.4° and 130.4°, and the rear calorimeter (RCAL) covering angles between θ = 130.4° and 172.0°. Each of the sections is divided longitudinally into hadronic and electromagnetic sections. The electromagnetic part is 1 λ in depth and is read out in 20 cm x 5 cm transverse sections. The hadronic part varies in depth between 3 λ and 6 λ and is read out in 20 cm square transverse sections with one longitudinal division. Each of the 6,432 calorimeter sections is read out by two photomultipliers, one connected to the wavelength shifter on each of two sides of the section.

TRIGGER SYSTEM OVERVIEW

The calorimeter trigger presents a challenge that is new to colliding detectors. This difficulty is not unique to the Zeus detector, but is found in other detectors planned for the SSC as well. The time between beam crossings at HERA is 96 nsec. This is too short a time for the calorimeter to read out the data and provide a trigger decision. The data are therefore kept in a pipeline and the first level trigger decision is postponed until 5 μsec after the crossing. In order to avoid deadtime, the trigger electronics itself must be pipelined: every process in the trigger must be repeated every 96 nsec.

The Zeus calorimeter trigger must detect charged and neutral current processes. In these events the current jet(s) and lepton emerge on opposite sides of the beam axis, balancing each other in transverse momentum. The debris of the proton is emitted forward in a narrow cone (~10 mrad). Neutral current events are characterized by an electron and jet(s) with balanced transverse momentum (p_T). Charged current events contain jets and missing p_T. In addition, exotic processes would be distinguished by the presence of lepton(s) and jet(s) with missing p_T. The trigger rate from conventional physics is expected to be ~3 Hz. The most serious background is from proton beam-gas interactions in the 70 m long straight section of HERA upstream of the detector. Assuming a flux of 10^{18} protons/sec and a beampipe vacuum of 3×10^{-9} torr, this interaction rate is approximately 500 kHz. The first level trigger rate needs to be held to 1 kHz.

The calorimeter trigger is based on several quantities. These are the total energy deposited in the calorimeter, the sum of transverse energy, missing p_T, and the number and energies of jets and isolated electrons. The calculations required for the calorimeter trigger include summing all of the pulseheights recorded in the photomultipliers every 96 ns. In addition, calculation of the transverse energy and missing p_T requires algebraically summing pulseheights multiplied by geometric factors. The detection of jets requires preservation of individual tower information that would be lost in a global sum. The detection of an electron requires evidence of electromagnetic energy. This is done by comparing energy deposited in the first interaction length of the calorimeter with that deposited later on a tower by tower basis.

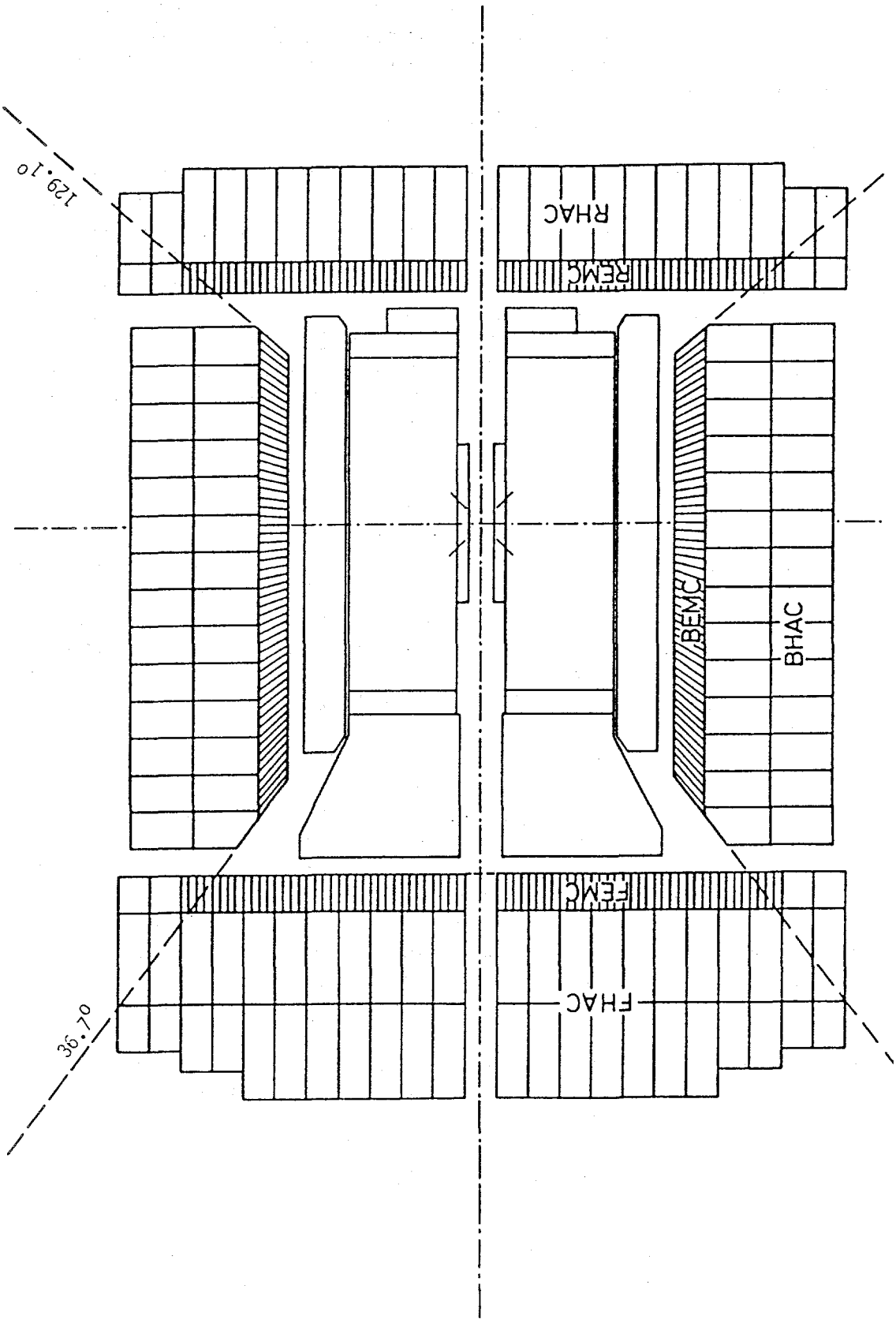


Figure 2. *The Zeus High Resolution Calorimeter.*

TRIGGER ORGANIZATION

The photomultiplier signals are combined into analog supertower sums. Due to the geometry of the detector, these supertowers vary considerably in dimensions. For the sake of conceptual explanation, the ideal supertower described has the most common geometry. This ideal supertower has a cross sectional area of 20 cm square longitudinally divided into EMC and HAC sections. The EMC section is divided transversely into four 5 cm x 20 cm sections, while the HAC section is divided into two 20 cm square longitudinal sections. Each section is read out through a wavelength shifter on both the right and left sides. Therefore, two phototubes read out each section. The most common variation on this geometry is a supertower with an undivided front HAC section read by 2 phototubes in place of the 4-part EMC section read by 8 tubes. A side view of an ideal supertower is shown in Figure 3.

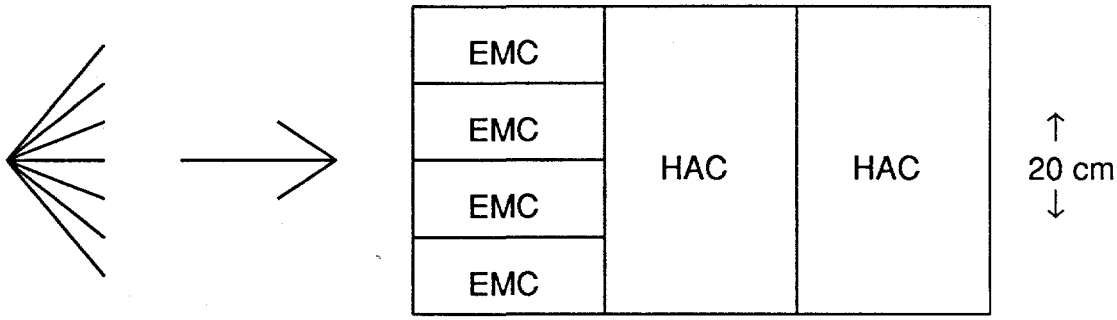


Figure 3. *A particle from an interaction entering an ideal supertower.*

Generally, there will be 4 EMC sections, 5 cm x 20 cm apiece, read out by 2 PMT's each, for a total of 8 PMT's, which are summed for the supertower EMC output. These are followed by 2 HAC sections, each 20 cm square, read out by 2 PMT's each, for a total of 4 PMT's, which are summed for the supertower HAC output. There are 460 supertowers in the Forward Calorimeter (FCAL), of which 264 have EMC sections. The remaining 196 FCAL supertowers have a HAC section in place of the EMC section and its pulseheight is included in the HAC sum. There are 448 supertowers in the Barrel Calorimeter (BCAL), all of which have EMC sections. There are 452 supertowers in the RCAL, of which 262 have EMC sections. The 1360 total calorimeter supertowers each provide a HAC sum and 974 of the supertowers provide an EMC sum. The tower geometry is set by the positions of the EMC supertower sections. These are projected back to HAC sections which are then summed to match the EMC projection. The result is a HAC sum for every EMC sum, yielding 1948 analog sums. These are then digitized by two 8-bit flash, ADC's (FADC) with a two different gains, to cover the dynamic range properly, and multiplexed to registers. Each FADC is matched to a memory lookup table that stores the linearized, pedestal-subtracted pulseheight for each FADC response placed in the registers.

The structure of the calorimeter trigger is severely constrained by the necessity to repeat each step and to pass processed information to the next step every 96 nsec. The trigger must first form the total energy, transverse energy and missing p_T sums. Given an EMC or HAC supertower sum of energy E_i located at polar angle θ_i and azimuthal angle ϕ_i, the sums may be formed from the quantities E_i (total energy), $E_i\sin\theta_i$ (transverse energy) and the components $E_i\sin\theta_i\cos\phi_i$ (missing E_x) and $E_i\sin\theta_i\sin\phi_i$ (missing E_y). The digitized HAC and EMC pulseheight from each of the 974 supertowers is stored in a register, multiplied by a lookup

table containing these four geometric factors and injected into a ten stage ($974 < 1024=2^{10}$) summing network.

The formation of jet and isolated electron triggers in a pipelined fashion is considerably more complicated than making sums. The trigger design uses tables to make local tests on the amount of energy in an individual supertower against six jet energy thresholds. In addition, supertowers are tested for minimum ionizing pulseheight and "quiet" pulseheight (less than five times background noise), and overflows. The ratio of HAC to EMC energy is also tested. The results of these tests are encoded in a series of bits and are passed forward with the energy sums to a search table that looks for matches with desired patterns. Regional energy sums and tests are also made at this stage. The final information from the tests and sums is forwarded to the global first level trigger, which compares the calorimeter information with that of the other detector elements in making the final trigger decision.

The Calorimeter First Level Trigger system begins with the left and right supertower sums from the Front End Cards. Each set of left and right sums is combined into a single sum on the Trigger Sum Card. These analog sums are then sent to the Trigger Encoder Cards in the electronics house. The Trigger Encoder Cards digitize, linearize, sum and test these signals and send results to Adder Cards in the same crate. The Adder Cards continue the summation process and search for patterns among the test results. The Adder Cards forward their results to the Trigger Processor Crate, which makes the final calculations with this information to forward to the Global First Level Trigger. This logical organization is illustrated in Figure 4.

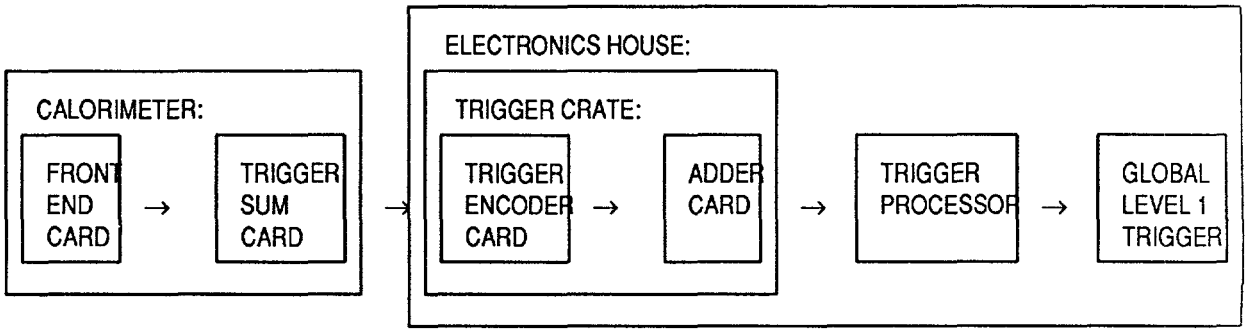


Figure 4. *Logical flow and location of Calorimeter First Level Trigger.*

ANALOG TRIGGER SUMS

The front end analog electronics is divided into cards servicing 12 photomultipliers connected to either the right or left sides of 2 supertowers. (This becomes 4 supertowers in the case where the supertowers have front HAC sections with only 2 tubes in place of EMC sections that have 8 tubes). The Trigger Sum Cards are located adjacent to the front end cards. Each front end card sums all of the hadronic and electromagnetic phototube signals in each supertower and outputs a total hadronic and electromagnetic pulseheight for either the left or right side of the supertower. Since every HAC supertower section is assigned to an EMC supertower section, there are additional sums of two HAC sections performed at the front end of the Trigger Encoder Card for the cases where two HAC supertower sections are assigned to one EMC supertower section.

The Trigger Sum Card adds two sets of left and right supertower sums. It integrates and clips the two sums at 96 nsec in a unity gain stage. The Trigger Sum Card contains an interface which connects to a command cable. This cable carries commands that cause the Trigger Sum

Card to turn off either the left or right input sum. The purpose of this is to compensate for the failure of a left or right signal. If, for example, the right sum of a particular supertower section were very noisy, it would be shut off, while the Trigger Encoder Card would double the gain of the supertower section to compensate for the loss of pulseheight from the right sum. Under normal conditions, all sums are turned on.

DIGITAL TRIGGER ENCODING

The function of the Trigger Encoder Card is to receive and digitize the analog trigger sum signals. The card produces a linearized energy value for each of these signals. It tests these energy values against six different thresholds. It tests for a "quiet" tower, i.e. a value of energy consistent with noise, for a minimum ionizing energy level, and for overflows (energy beyond the input dynamic range). It also tests the ratio of EMC to HAC energy to see if the energy deposition is consistent with an electromagnetic shower. It forwards the results of these tests to an Adder Card that resides in the same crate. Finally, it multiplies the energy values by geometric factors to compute total Energy, E_T, E_X, and E_Y for each supertower HAC and EMC, and sums up these EMC and HAC energies separately for transmission to the Adder Card.

The organization of the front end of the Trigger Encoder Card is shown in Figure 5. The signal is received by 2 amplifiers with high and low gain. These two amplifiers each drive a FADC which digitizes the signal. These digital results are then used as input to two memory lookup tables that produce linearized energies in the various ranges described below.

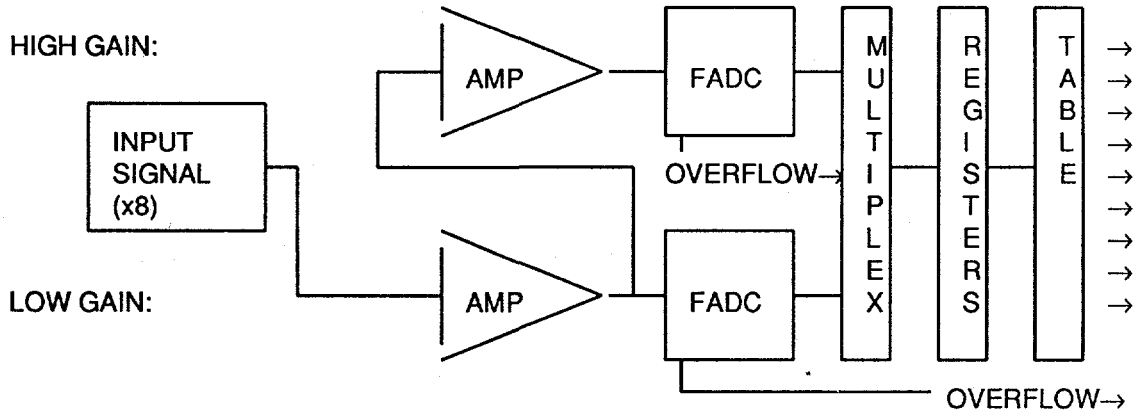


Figure 5. *Organization of the Trigger Encoder Card front end.*

The EMC and HAC sums from 4 supertowers are digitized by each Trigger Encoder Card. In those cases where two HAC sums are grouped with an EMC sum, the card digitizes the HAC sums in pairs from up to 8 supertowers. Each of the channels on the Trigger Encoder Card has 2 input connectors. These are received separately and combined at the input to a single amplifier. This allows the combination of 2 separate HAC supertower sums before amplification. Each input signal (single or combined) is brought through one amplifier for the low gain, and through a second for the high gain. The amplifiers employed are video amplifiers with 2 nsec propagation delays.

After amplification, each signal is digitized by an 8-bit FADC. The overflow bit of the high gain channel FADC drives the select on a multiplexer. If the overflow bit is low, the value of the high FADC is written into the register. If the overflow bit is high, the value of the low FADC is

written into the register. In either case, the value of the overflow bit is also written into the register. This creates 9 bits of data, including the 8 bits of FADC data (either high or low) and the high channel overflow bit, which are written into the register. The low gain channel overflow bit is also brought out to the test circuitry via a path independent of this data.

The contents of the register form an address for a 2K×8 memory organized into 4 pages with 512 8-bit values apiece. The memory puts out the values selected from its first three pages, one value every 32 nsec. The first page of the memory contains energies on a scale corresponding to the choice of 8-bit FADC output for whichever of the two FADC's was written into the register. This choice is indicated by the high gain channel overflow bit saved in the register. These energies are corrected for pedestal, gain and nonlinearities.

The second page of the memory contains a set of energy values which are fully corrected as for the first page, but are all on the same linear scale for both high and low channels. The high gain channel is placed on this same scale, where it covers the overlap portion of the range. The third page of the memory also contains a set of fully corrected energy values which are multiplied by different factors for HAC and EMC sections. The EMC sections, whether high or low gain channel is written, are placed on the same scale. The HAC section values are placed in the same scale as the EMC sections, but have had their complements taken (yielding negative values) and are multiplied by a factor K, such that if for a particular supertower, $E_{EMC} - K \bullet E_{HAC}$ is positive, the energy deposition is deemed electromagnetic.

TRIGGER SUMS AND TESTS

There are two parallel TTL adder trees employed on the Trigger Encoder Card instead of a single ECL tree in order to save power and space. Each of these trees accepts four input values every 48 nsec and outputs their sum 96 nsec later. The adder tree starts with the contents of the registers from the second pages of the 8 input memories. Each of these values forms an address presented to two 2K×8 memories clocked at 48 nsec. Both memories serve as lookup tables that effectively multiply their input values by geometric and scale factors to convert E_i into E_i, $E_i\sin\theta_i$, $E_i\sin\theta_i\sin\phi_i$, and $E_i\sin\theta_i\cos\phi_i$ on the same scale for the whole calorimeter. These values are combined separately for EMC and HAC sections to make the total energy, E_T, E_X, and E_Y sums.

Each of these 2 memories has its 2 pages summed with the values from the adjacent supertower section of the same type (HAC or EMC). The 9-bit results are summed with the contents of the sum of the other 2 supertowers. The 10-bit results are shifted into a TTL register, translated to ECL and sent to the Adder Card over 10 bus lines on the backplane at a 12 nsec rate. The structure of one of the two adder trees on each Trigger Encoder Card is shown in Figure 6.

The test encoding circuitry makes use of the first 3 pages of the front end memory. The tests made are "quiet", minimum ionizing, electromagnetic energy, overflow of the low gain channel, and six different energy thresholds. The first step is to sum the two supertower section (HAC and EMC) front end memory results separately by page. The section results contained in registers are summed in 8-bit parallel adders clocked at 32 nsec. The tests are made by a 2K×8 memory that serves as a lookup table. The results of the tests are assembled in six bits as described below.

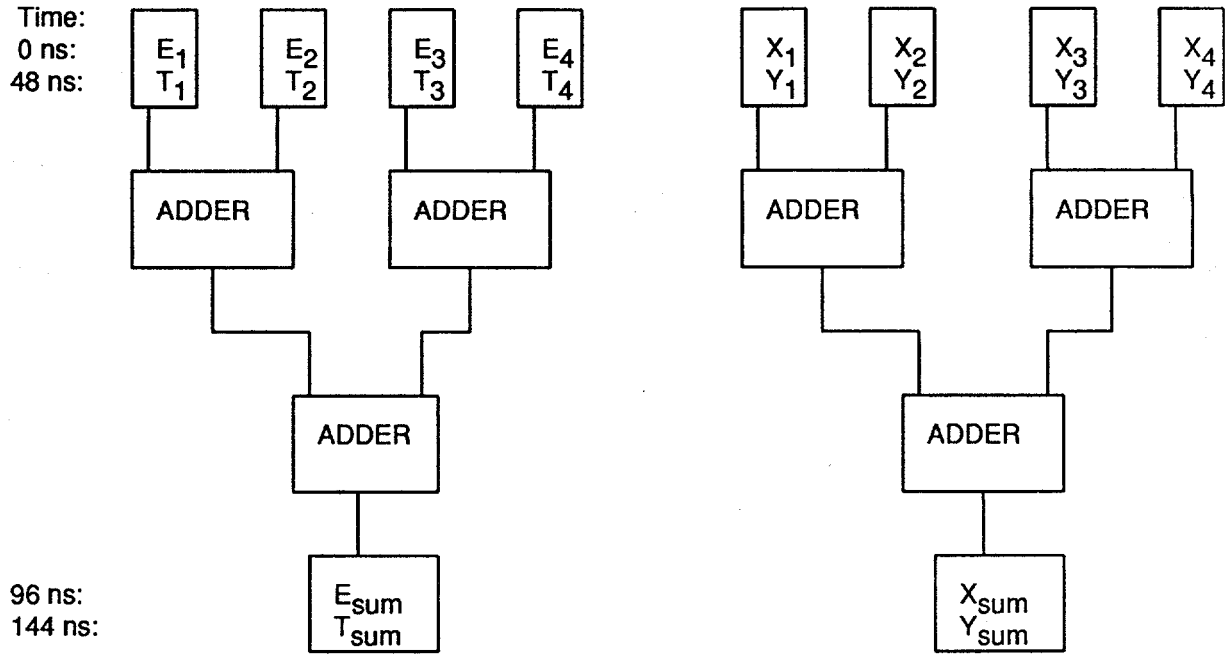


Figure 6. *Half (either HAC or EMC part) of the adder tree network on each Trigger Encoder Card. E_i represents the total energy in the supertower section, T_i is $E_i \sin\theta_i$ (transverse energy), X_i is $E_i \sin\theta_i \cos\phi_i$ (E_x) and Y_i is $E_i \sin\theta_i \sin\phi_i$ (E_y). The elapsed time is shown on the left.*

The test memory uses 3 pages. The first page of the test memory performs the "quiet" tower and minimum ionizing tests. The values presented to it as an address are a sum of the EMC and HAC channels, along with a ninth bit which is an OR of the carry bit from the adder, the EMC high gain channel overflow and the HAC channel high gain overflow. If the ninth bit is set, both tests are flunked and the elements are zero. The page elements are set up so each address has one bit assigned for each test.

The second page of the test memory makes the six energy threshold tests. Each element has an address that corresponds to a sum of the energies in the HAC and EMC supertower sections. The sums are made from the second pages of the front end memories, which are on the same linear scale regardless of whether a high or low gain channel was used. Each memory address contains the threshold test results in 3 bits which cover the 7 possible outcomes (pass none through pass all 6 -- 000 through 110). In addition, all addresses which correspond to the ninth bit set (OR of the low gain channel overflow bits) have all three threshold test bits set to 1.

The third page of the test memory makes the electromagnetic energy test. Since the two sections being summed in front of the test memory are HAC and EMC sections, the values summed are the energy in the EMC section and -K times the energy in the HAC section. Here a positive result means that the hadronic energy was less than 1/K of the electromagnetic energy. The elements of this page contain 1 bit to indicate whether the test is passed or not.

The results of the 3 test pages are contained in 6 bits accumulated in a TTL register after 96 nsec. The bits are the "quiet" test bit, the minimum ionizing test bit, the three energy threshold test bits, and the electromagnetic test bit. After translation to ECL, the 24 bits of test data from the 4 supertowers (6 bits apiece) are sent out over 3 bus lines on the VME backplane at a 12 nsec rate.

TRIGGER ENCODER DATA FLOW

Each Trigger Encoder card is assigned one of 8 sets of 14 lines out of the 88 user defined lines on the J3 bus, or the 56 user defined lines on the J2 bus. The ECL drivers send the 13 bits of data every 12 nsec as shown in Table 1. Each 13 bit word includes one of the 8 4-supertower sums (total Energy, E_T, E_X, and E_Y for each supertower HAC and EMC) in 10 bits and 3 out of the 6 total test bits for each supertower. In addition, the Trigger Encoder Card puts a bit on a 14th line that toggles from high to low with each word that is placed every 12 nsec on the bus. This toggling is used by the Adder Card to verify that the Trigger Encoder Cards are sending data.

Table 1. *Trigger Encoder Card output bit structure on a 13-bit wide bus, clocked at 12 nsec so that the data is contained in 8 words sent out in 96 nsec. The energy sums of total energy for HAC and EMC, transverse energy for HAC and EMC, and the x and y components of energy for HAC and EMC are summed over 4 supertowers and contained in 10 bits of each of the eight words as shown. TIJ stands for the Jth test bit from supertower I. The time is given in nsec.*

Word	Time	Bt1	Bt2	Bt3	Bt4	Bt5	Bt6	Bt7	Bt8	Bt9	Bt10	Bt11	Bt12	Bt13
1	0	←	-	-	-	EMC	TOTAL SUM		-	-	→	T11	T12	T13
2	12	←	-	-	-	HAC	TOTAL SUM		-	-	→	T14	T15	T16
3	24	←	-	-	-	EMC	TRANS SUM		-	-	→	T21	T22	T23
4	36	←	-	-	-	HAC	TRANS SUM		-	-	→	T24	T25	T26
5	48	←	-	-	-	EMC	X SUM		-	-	→	T31	T32	T33
6	60	←	-	-	-	HAC	X SUM		-	-	→	T34	T35	T36
7	72	←	-	-	-	EMC	Y SUM		-	-	→	T41	T42	T43
8	84	←	-	-	-	HAC	Y SUM		-	-	→	T44	T45	T46

ADDER CARD

There are two Adder Cards, which are identical in all respects and reside in the middle of the VME crate. Each is connected to half of the split J2 and J3 backplanes. They perform several functions, including the summation of total Energy, E_T, E_X, and E_Y. In addition, the Adder Card handles the clock distribution, recognition of isolated muons and electrons, counting of supertowers exceeding the different thresholds and computation of energies summed over a subset of the supertowers sent to it.

The adder tree is continued from the Trigger Encoder Cards onto the Adder Cards. Every 12 nsec, the summing section of the Adder Card strips off the lowest order 10 bits from the 13 bits of data reaching it from each of the 8 Trigger Encoder Cards it is connected to on its half backplane. These bits are directly injected one of 8 inputs to a 3-stage ECL parallel adder network. Four of these adders combine the 10 bits of data in groups of two and produce four sums each with 11 bits of dynamic range. The outputs of the adders are fed directly through subsequent sets of adders that keep 12 bits of dynamic range, truncating the lowest order bit. These sums are delayed in a FIFO clocked at 24 nsec until the results of the pattern tests are complete so that all results of the Adder Card from one crossing are contained in the same 96-nsec time space, subdivided into 4 words within that time space.

The two identical Adder Cards perform the operations above in parallel and feed results to another 12-bit adder. They also feed the same result to a connector at the front of the card. The cable connecting the two Adder Cards is bidirectional. During VME initialization a control register on each Adder Card is written to designate one as the "Master" and the other as the

"Slave". (These definitions have nothing to do with their VME operation). The data from the "Slave" is sent to the other side of the adder on the "Master". Thus the last Adder on the "Master" combines the data of the two Adder Cards and this last sum is fed out the front of the Adder Card on a cable to the Trigger Processor Crate.

PATTERN LOGIC

The pattern logic of the Adder Card uses the upper 3 bits of the data arriving every 12 nsec from each of the 8 Trigger Encoder Cards to which it is connected. The Trigger Encoder Cards cover 4 supertowers apiece, each of which contributes 6 bits of test information. The Adder Card accumulates this information during the 96 nsec time span and shifts it forward to pattern logic RAM's. After 96 nsec, each Adder Card has accumulated information from 32 supertowers. The two Adder Cards combine this information to search for patterns in a 64-supertower region.

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

E	E	E	E	E	E	E	E
E	C	C	C	C	C	C	E
E	C	C	C	C	C	C	E
E	C	C	C	C	C	C	E
E	C	C	C	C	C	C	E
E	C	C	C	C	C	C	E
E	C	C	C	C	C	C	E
E	E	E	E	E	E	E	E

Figure 7. *Assignment of supertowers in a 64 supertower region as either edge (E), or contained (C).*

The pattern logic sums up the number of supertowers exceeding each of the thresholds and not the one above it. It sums up supertowers passing no threshold and those with overflows. It also performs sums in subregions of the 64-supertower region by computing the energy from the threshold test information. These regions and sums are programmable. One set of regions is shown in Figure 7. The 64 supertowers are assigned to either the contained region (36 supertowers) or the edge region (28 supertowers). The number of thresholds exceeded in each of these regions is then used to calculate a contained energy and an edge energy. Another set of regions involves grouping the supertowers that are close to the beam-pipe into a beam-gas region, and the other supertowers into a non-beam-gas region. A third regional division matches calorimeter regions with the regions of another detector element such as the Central Tracking Detector (CTD).

The pattern logic also searches for a single or group of up to 4 supertowers with their electromagnetic or minimum ionizing test bits set that are completely surrounded by "quiet" cells. These are counted as contained isolated electrons or contained isolated muons. The

pattern logic searches for groups of supertowers that would satisfy these conditions except that one or more of the electromagnetic or minimum ionizing supertowers is next to the 64-supertower region boundary. In these cases, isolation on one of the sides cannot be proved without checking the adjacent 64-supertower region. These are counted as edge isolated electrons or edge isolated muons, and are categorized by which of the four edges they touch. Examples of edge and contained isolated muon patterns are shown in Figure 8.

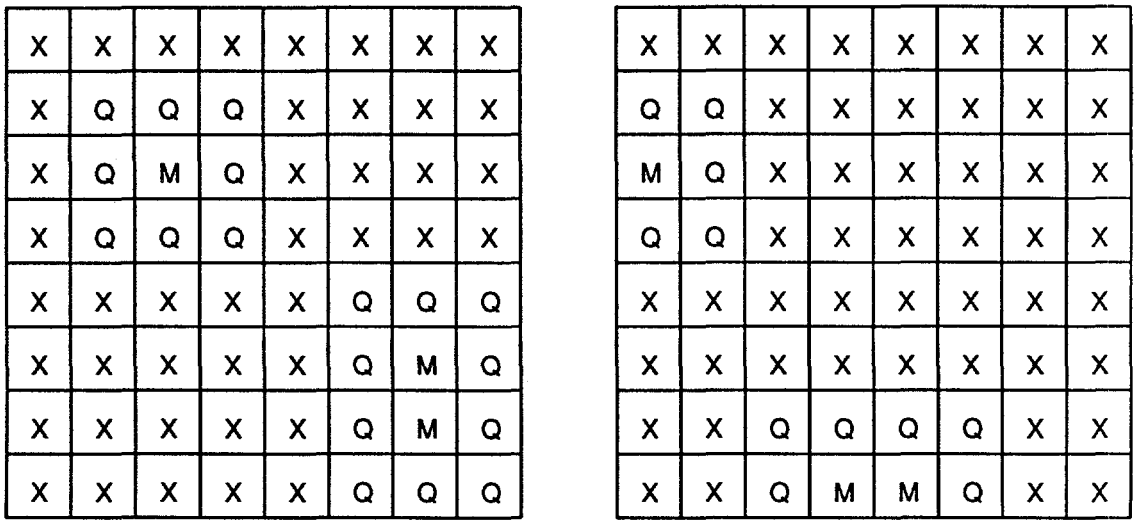


Figure 8. *Examples of two patterns accepted as contained isolated muons (left figure) and two patterns accepted as edge isolated muons (right figure). supertowers are labelled as quiet (Q), minimum ionizing (M), or do not care (X). The same patterns satisfy the isolated electron test if the same supertowers passing the minimum ionizing test, would pass the electromagnetic test instead.*

The two Adder Cards together are able to sum the energy in up to 8 different subregions of the 64 supertowers. These sums are performed on the data coming into each Adder Card without consideration of the data being received by its partner. The Adder Cards make these sums independently. The two Adder Cards send partial sums to each other and perform the final summation to make up the entire subregion. This mode of operation permits the creation of a single design for the Adder Card.

TRIGGER CRATE BUS AND CARD STRUCTURE

The Trigger Crate contains two groups of 8 Trigger Encoder Cards connected to a single Adder Card. The Trigger Encoder Cards communicate with the Adder Card over the J2 and J3 VME backplanes. The J2 and J3 backplanes have all lines bussed together and are split between the 2 adder cards as is shown in Figure 10. In this sense, the J2 and J3 backplanes are non-standard since they have the user-defined lines connected to each other with the exception of a break between slots 11 and 12. In addition, these backplanes do not extend into the first slot, where the Crate Processor board is located. The crate also has a VME service module in the 21st slot. This module contains an IEEE interface used for voltage monitoring. The J1 backplane is operated in the standard manner, yielding 24 address lines and 16 data lines.

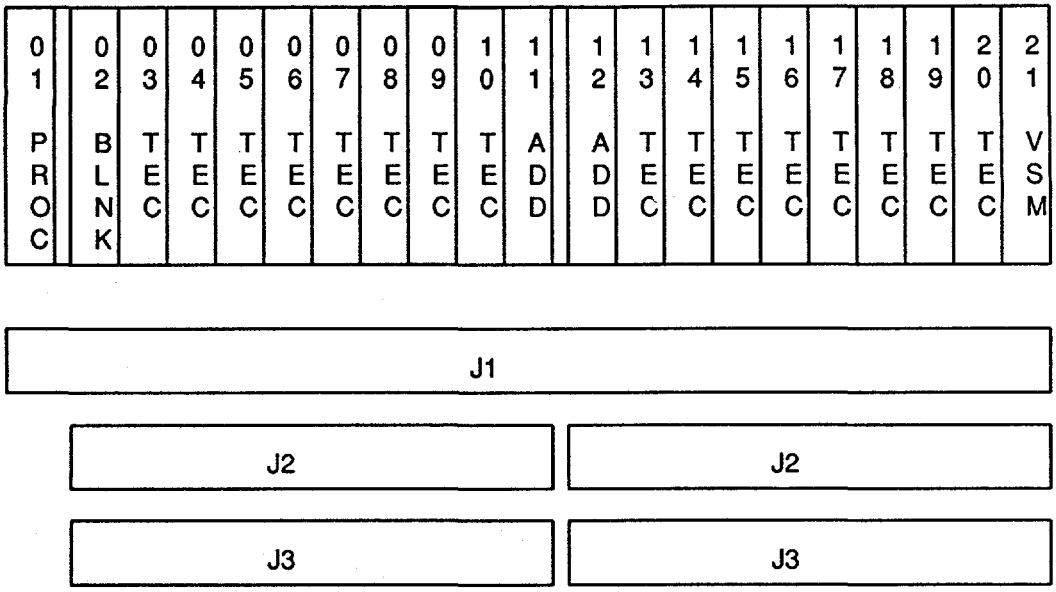


Figure 10. *Assignment of modules to slots in trigger VME crates with backplane divisions shown below. Trigger Encoder cards are labelled TEC, Adder Cards are labelled ADD, the VME Service Module is labelled VSM, there is a blank slot labelled BLNK, and the Crate Processor Board is labelled PROC. The J1 bus is is connected to all 21 slots. The J2 and J3 busses are broken between slots 11 and 12 and do not continue to slot 1. All user-defined lines on the continuous sections of the J2 and J3 busses are linked together.*

TRIGGER PROCESSOR

The Trigger Processor Crate does the final trigger calculations with data from the "Master" Adder Cards and ships these results to the Global First Level Trigger for the final trigger decision. The 16 "Master" Adder Cards send data from each of their crates to the Trigger Processor Crate with 24 nsec clocking. The first set of information includes the number of contained isolated electrons, edge isolated electrons, contained isolated muons and edge isolated muons. The "Master" Adder Cards also send the number of supertowers that exceeded each (or no) threshold, but not the next higher threshold, and the number that overflowed their low gain channel.

A second set of information contains the energy in a specific subregion of the 64-supertower region as calculated from the supertowers passing the thresholds. There are 8 total subregion supertower sums available. The initially planned subregions are the contained, edge, beam-gas, and non beam-gas regions. Additional subregions are defined for matching trigger subregions in other components, such as the Central Tracking Detector. The third set of information contains the EMC and HAC 64-supertower sums of total energy and transverse energy, E_x and E_y

The Trigger Processor produces a global status word, regional status word, beam-gas information word, a beam-gas probability score, and a word containing the number of isolated muons and number of isolated electrons. The global status word contains the number of electronic overflows, the number of corrected electronic overflows, the number of isolated electronic overflows and the number of corrected isolated electronic overflows. These electronic

overflows are considered corrected for if the trigger system compensated for their inclusion in the global sum and regional pattern information. Correction for a non-isolated overflow means inclusion of the maximum possible energy, and correction for an isolated overflow means inclusion of zero energy from this supertower. The regional status word indicates whether each of the 16 regions had an electronic overflow.

The beam-gas information word contains the sum of the energy contained in the beam-pipe region of the FCAL and the RCAL and the ratio of HAC energy to EMC energy in the RCAL. This ratio is expected to be high for beam-gas supertowers entering the back of the RCAL. The beam-gas information also has a global BCAL EMC sum, and an indication whether any supertower had an energy over the maximum kinematically possible. The beam-gas score represents the result of an algorithm that uses the information summarized in the beam-gas information word and additional information to create a score representing the probability that the energy observed is due to beam-gas interactions.

The number of isolated electrons and muons found in the detector is the sum of the contained isolated electrons and muons and the verified edge isolated electrons and muons. Edge isolated electrons and muons are verified by checking the edge region of the adjacent 64-supertower region for another edge isolated isolated electron or muon, or all "quiet". In the case where another isolated electron or muon is found, the two neighboring edge isolated electrons or muons are combined and counted as a single contained isolated electron or muon. In the case where the edge region adjoining an edge isolated electron or muon is "quiet", the edge isolated electron or muon is counted as a single contained isolated muon or electron. In the case where the adjacent edge region is neither "quiet" nor contains an isolated electron or muon, the edge isolated electron or muon is rejected.

The Trigger Processor also sends the global total, EMC and HAC energy sums, and cluster information. This includes the total energy and transverse energy for the HAC sums, EMC sums and HAC + EMC sums. A word is sent that contains the number of total and EMC clusters (~jets) passing 3 sets of thresholds. Further information contains E_x and E_y for the HAC sums, EMC sums and HAC + EMC sums. The total missing energy $(E_x^2 + E_y^2)$ and the total missing electromagnetic energy are also directly calculated. Information is provided on the regional number of isolated electrons and isolated muons (verified edge depositions are reported in both regions), amount of total energy transverse energy, amount of energy in the beam-gas subregion of the region (i.e. part closest to the beam-pipe), and the amount of electromagnetic energy.

GLOBAL FIRST LEVEL TRIGGER

The calculations of the Trigger Processor take 2 μsec to complete. The results from the Trigger Processor are shipped to the Global First Level Trigger (GFLT) that combines different detector elements for a final trigger decision in 5 μsec. A typical example is to require an $E_T > 10$ GeV in the CAL accompanied by a Central Tracking Detector (CTD) multiplicity > 5 tracks pointing at the interaction point. The GFLT is also pipelined, accepting data and sending a decision every 96 nsec. Due to propagation delays in receiving the data and sending the trigger, the GFLT calculation time is also 2 μsec. After the issuing of the GFLT, the event is analyzed by the Second Level Trigger. This system can handle a rate of 1 kHz.

The GFLT can also use regional information from the detector elements such as the CTD and the CAL to make its decisions. In a specific region, the CTD provides the multiplicity of tracks, whether the tracks were well measured, and whether the tracks pointed at the interaction point. A track that is not well measured only intersects a single superlayer of the

CTD and hence its pointing at or away from the interaction point is not well known. The total CTD regional information is combined with the CAL regional information including the number of isolated electrons and muons, and the electromagnetic energy. The result is the number of isolated muons and electrons gated on CTD information indicating a track pointing at the vertex. A second correlation combines the regional CTD information mentioned above with the CAL regional information on the sum of transverse energy, beam-gas subregion energy and total energy to produce a transverse energy sum gated on CTD and CAL beam-gas indicators. A third correlation combines the regional CTD information with the CAL regional information on the number of supertowers passing the sixth, fifth and fourth energy thresholds. This constitutes a cluster trigger that is associated with the interaction point.

CONCLUSIONS

The Zeus Calorimeter First Level Trigger system is designed to calculate energy sums, search for energy clusters, isolated muons and electrons, and summarize regional information for correlation with other detector elements. It completes these tasks in a pipelined fashion in 2 μsec, accepting new data and shipping out results once every 96 nsec. It uses digital logic with memory lookup tables to provide programmable flexibility for changing calibration, thresholds and even the types of tests made. The results of its calculations are shipped to the Global First Level Trigger (GFLT). The GFLT must reduce the initial beam-gas rate from 500 kHz to 1 kHz. The GFLT is also pipelined, reaching a decision in 2 μsec after receipt of the component data. The First Level Trigger system reads in a new event every 96 nsec and returns a trigger to the detector components for that event 5 μsec after it occurred. The capabilities of this design are close to those required of an SSC detector First Level Trigger. The Zeus Calorimeter First Level Trigger provides a useful proving ground for the techniques that must be managed to construct such a trigger for an SSC detector.

ACKNOWLEGEMENTS

Many members of the Zeus Collaboration are involved in the Calorimeter First Level Trigger. Joe Lackey of the University of Wisconsin is responsible for much of the electronic design. The original concept reflects the input of Bill Sippach of Columbia University. Among other members of the Zeus Collaboration making significant contributions are Allen Caldwell, John Dawson, Matthew Jaworski, Robert Stanek, and Richard Talaga.

REFERENCES

1. G. Ingleman et al., in *Proceedings of the HERA Workshop*, R.E. Peccei ed., (Hamburg, Germany, Oct. 12-14, 1987).

2. W.H. Smith et al., in *Proceedings of the Snowmass Summer Study on Physics in the 1990's*, F. Gilman, ed. (Snowmass, CO., July, 1988).

3. G. Wolf et al., The Zeus Detector: Status Report (DESY-HERA-PRC-02, Sept., 1987); G. Wolf et al., The Zeus Detector: Technical Proposal (DESY-HERA-ZEUS-1, Mar. 1986).

4. M.J. Catenesi et al., Nucl. Instr. Meth. **A260**, 43, 1987.

DATA ACQUISITION FOR THE ZEUS CENTRAL TRACKING DETECTOR

S.Quinton
Rutherford Appleton Laboratory, Chilton, OX11 0QX, UK

ABSTRACT

The Zeus experiment is being installed on the Hera electron-proton collider being built at the Desy laboratory in Hamburg. The high beam crossover rate of the Hera machine will provide experience in data acquisition and triggering relevant to the SSC environment. This paper describes the Transputer based data acquisition for the Zeus Central Tracking Detector, and outlines some proposed development work on the use of parallel processing techniques in this field.

INTRODUCTION

The development and construction of the Central Tracking Detector (CTD) of the Zeus experiment and its associated electronics are the responsibility of a British consortium. Those directly involved in the data acquisition system described here are the Rutherford Appleton Laboratory, Oxford University and University College, London.

The detector is a cylindrical wire drift chamber 2.4m long and 1.6m in diameter. 4608 sense wires are arranged in nine superlayers each containing eight sense wire layers. Measurements are made of drift time (the R-ø position), energy deposition (dE/dx), and position along the wire axis (Z by time difference).

The short bunch spacing in the Hera machine of 96nS imposes the requirement to pipeline the incoming data to cope with a reasonable first level trigger decision time and with a chamber drift time (500ns) longer than the bunch spacing. The high input data rates are reduced both by trigger selection and processing of the data.

FRONT END DATA ACQUISITION

The most demanding part of the detector readout from the point of view of data rates is the R-ø and dE/dx measurement. The front end of that system is shown in figure 1.

The analogue signals from each wire are digitised in parallel by 8 bit 100MHz flash ADCs. The total data rate from the outputs of these FADCs is:

$$(4608 \text{ wires}) \times (100\text{MHz}) = 0.46\text{Tbytes/s}$$

Data from each FADC is written into fast pipeline memory. The continuously running pipeline is 5µs long to cover the first level trigger decision time. Table 1 shows the parameters for the different trigger levels.

Table 1. Trigger Parameters

Level	Type	Process Time	Decision Rate	Event Reduction Factor
1	hardware	5µs	10MHz	10^4
2	software	10ms	1kHz	10
3	software	secs.	100Hz	30

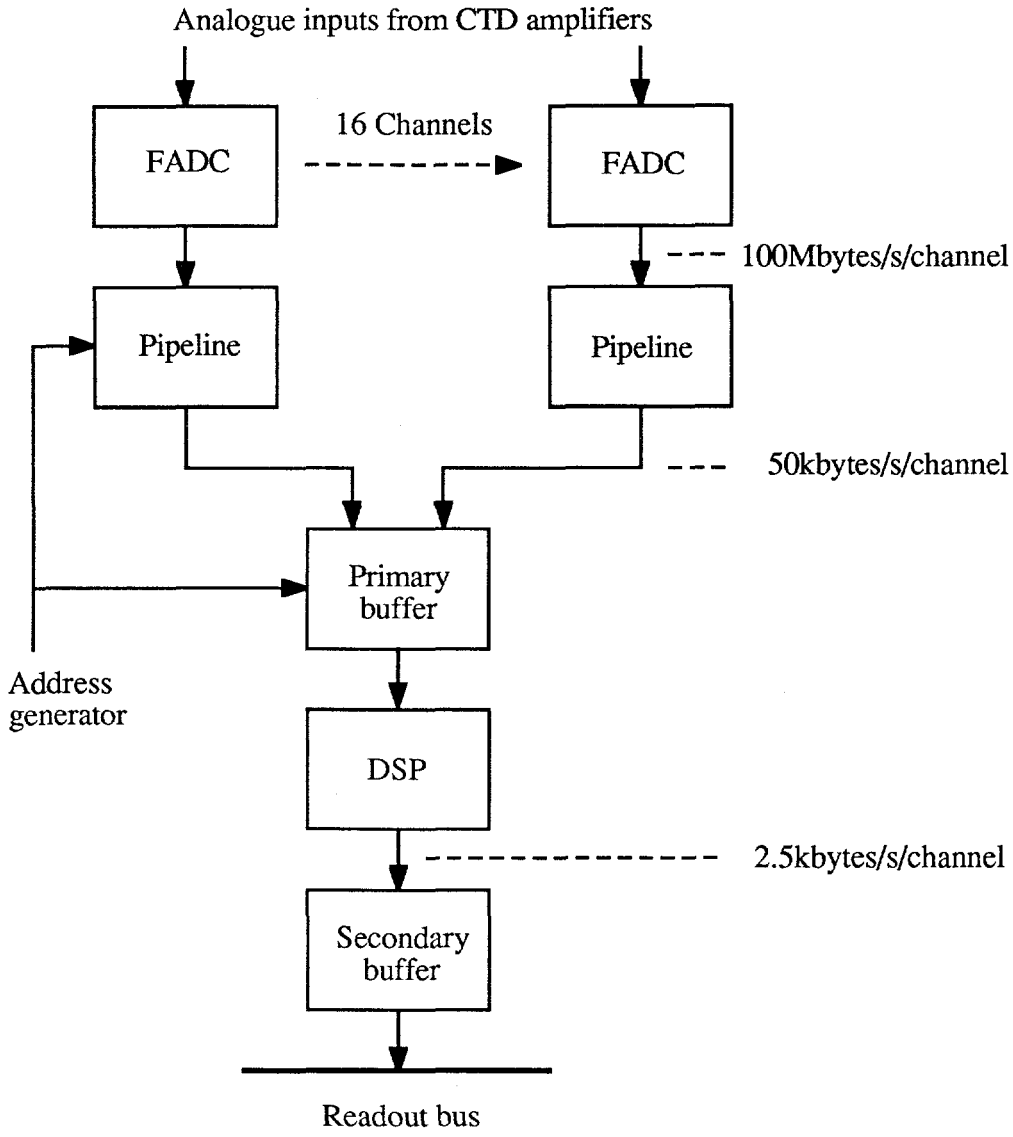


Fig 1. Front end data acquisition.

On a positive first level trigger the pipeline clock is stopped, the address is set to to point to the pipeline position corresponding to the bunch crossing, and 50 bytes of data corresponding to the 500ns drift time of the chamber is transferred to the primary buffer. This transfer is at high speed to minimise the deadtime incurred while the pipeline is stopped. The average overall data rate after a 1kHz first level trigger is:

(4608 wires) x (50 bytes) x (1kHz) = 230Mbytes/s

Event data in the primary buffer is operated on by a DSP, one per 16 channel R-ø card, which performs zero suppression and extracts the drift time and pulse energy. These results are written into a secondary buffer. The maximum DSP processing time is 1ms, corresponding to the mean

spacing between first level triggers. Queuing theory leads to a primary buffer depth of four events as being the point at which further increases in storage give rapidly diminishing returns in terms of reducing deadtime. Monte Carlo simulations indicate that for a typical event hits can be expected on 10% of the CTD wires. For single hits the DSP will produce 12 bytes of data describing the pulse. For the 25% of hit wires with two or more hits the DSP transfers, in addition to pulse analysis data, the raw pipeline data for offline analysis. So the overall data rate into the secondary buffers is:

((4608 wires) x (10% hit) x (12 bytes) x (1kHz)) +
((4608 wires) x (10% hit) x (25% double hit) x (50 bytes) x (1kHz)) = 11.3Mbytes/s

The R-ø readout is distributed over 16 crates, each containing 18 cards of 16 channels each, giving an average crate data readout rate of about 0.7Mbytes/s.

CRATE READOUT

The crate readout system is shown in figure 2. A readout controller is used which is based on a T212 Transputer and the readout bus is essentially the buffered Transputer memory expansion bus. Readout is triggered by all the DSPs indicating that they have finished processing a particular event, then Transputer block move instructions are used to transfer the data from the modules into a dual port memory on the readout controller card. On a positive second level trigger the higher levels of the readout system access this data through the Transputer links. The second port of the dual port memory is accessed by the T800 floating point Transputers which perform the track

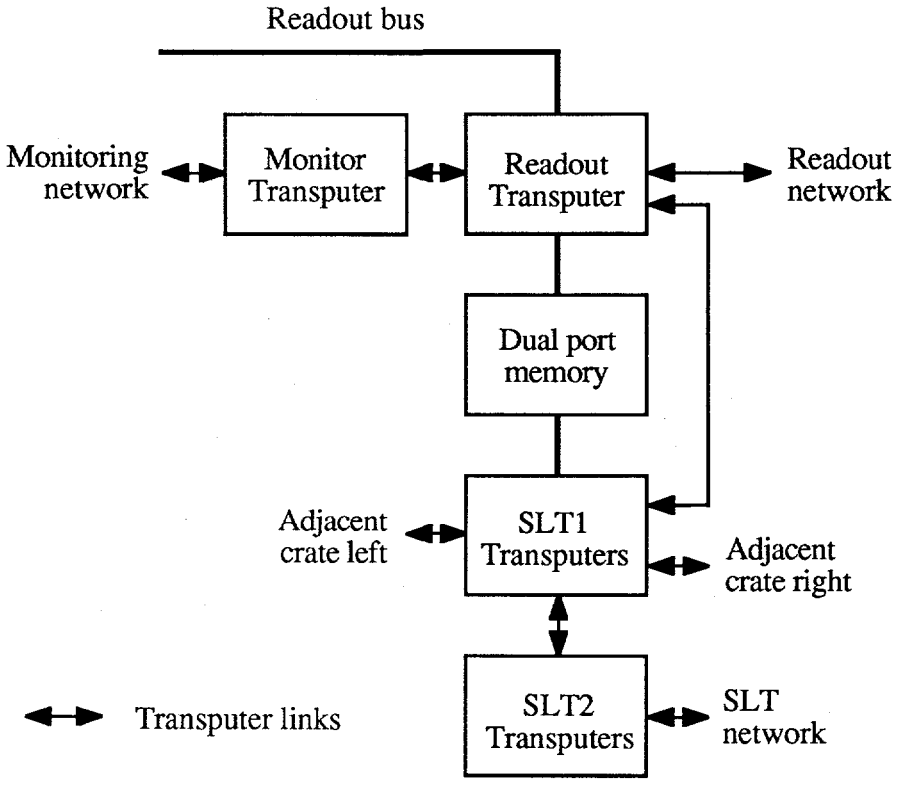


Fig 2. Crate readout

finding algorithms for the second level trigger. The dual port memory can be accessed at full Transputer speed for second level trigger processing whereas wait states would have to be used if the data were directly accessed on the readout cards using the slower readout bus.

It was the requirements of the second level trigger processing that provided the initial driving force behind the choice of Transputers as the processing and communication elements. The Transputer is a high performance processor in its own right but with the major additional advantage of its hardware and software support for parallel processing. Hardware support is in the form of the four 20Mhz bidirectional serial links on each chip, a simple and effective way of providing the necessary communication between the second level trigger processors for track finding across crate boundaries. In addition the practical upper limit on link transfer speeds of 1.7Mbytes/s is more than sufficient for the crate data readout rate after second level trigger of 70kbytes/s. The Occam software fully supports these links and synchronises parallel processes by allowing the programmer to create processes which run on receipt of data from a link connection or output data down links. The on chip multitasking scheduler provides fast switching between the different processes running on a particular Transputer.

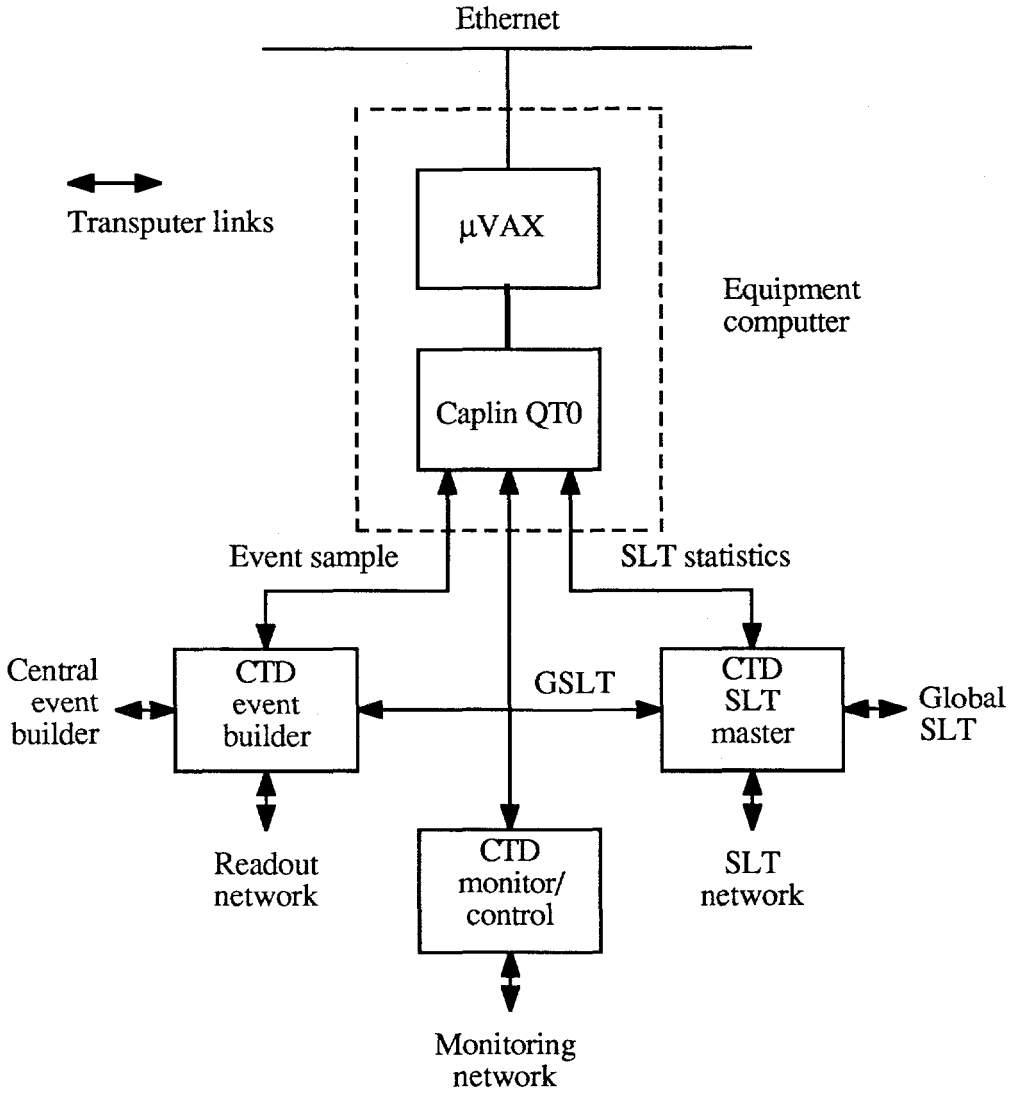


Fig 3. Detector readout

DETECTOR READOUT

The final design of the higher levels of the data acquisition and second level trigger is not yet completed but figure 3 shows the major components. The CTD event builder and second level trigger master will both be Transputer based and may be implemented on standard VME boards. The event builder transfers data from the crate readout controllers through the Transputer links, the interconnection between them being either a chain or tree structure. The total CTD data rate of 1.2Mbytes/s will require either one or two Transputer links to the central data acquisition event builder. Similarly second level trigger results from the individual readout controllers are collated by the second level trigger master and passed on to the global second level trigger.

A Caplin QT0 Transputer board provides the connection between the Qbus of the μVAX equipment computer and the Transputer networks. As well as event sampling this provides a route, through the Transputer link structure, to download data acquisition control information, such as DSP programs, and to run tests on the readout system. Additional fast hardware connections will be required to distribute first and second level trigger results to the front end cards and readout controllers.

Software support in the final system will be provided on the μVAX with Occam, parallel C and parallel Fortran available. During development, commissioning and repair IBM PCs will provide similar support as portable, stand alone machines.

FUTURE DEVELOPMENTS

One idea considered during the development of this architecture was to equip each front end card with a Transputer, doing away completely with the concept of a readout bus and using the Transputer links for the transfer of all readout, control and test data. This idea was rejected for a number of subjective reasons, not least being the short time scale for completion of the system - mid 1990. It was felt that further research and development work was needed to give confidence in running such large Transputer networks (several hundred processors) in a data acquisition system and to fully understand the problems of latency and event synchronisation. It is intended to pursue these topics in a separate research programme which will include system modelling and simulation. The results of this work will be highly relevant to data acquisition on future planned accelerators such as the SSC.

TRIGGERING AND DATA ACQUISITION ASPECTS OF SSC TRACKING*

GAIL G. HANSON, BOGDAN B. NICZYPORUK† AND ANDREA P. T. PALOUNEK‡
Stanford Linear Accelerator Center, Stanford University,
Stanford, California 94309, USA

ABSTRACT

Possible conceptual designs for wire chamber tracking systems which meet the requirements for radiation damage and rates in the SSC environment are discussed. Computer simulation studies of tracking in such systems are presented. Results of some preliminary pattern recognition studies are given. Implications for data acquisition and triggering are examined.

1. INTRODUCTION

The primary motivation for the SSC is the expectation that it will lead to new discoveries, such as Higgs bosons, supersymmetric particles, heavy W's or Z's, new heavy fermions, or composite particles with masses in the TeV region. Such particles would be produced in the central rapidity region, that is, over $\pm$ 3 units of rapidity, and would decay to high-p_T electrons, muons, or jets, often with large missing transverse energy (E_T) due to undetectable neutrinos. In order to fully investigate the physics opportunities in this regime, a general-purpose detector which includes charged particle tracking is needed. Some of the most important functions of charged particle tracking include:

1. Identification of electrons.
2. Separation of multiple interactions within the same bunch crossing.
3. Matching electrons, muons, and jets to the correct vertex.
4. Electron charge sign determination.
5. Improving e/π separation.
6. Identification of secondary vertices.
7. Identification of τ leptons.

* Work supported by the Department of Energy, contract DE–AC03–76SF00515.

† Permanent Address: CEBAF, Newport News, Virginia 23606.

‡ Present Address: Lawrence Berkeley Laboratory, Berkeley, California 94720.

8. Invariant mass or momentum cuts.
9. Improving the missing E_T measurement and verifying calorimeter data.
10. Establishing the credibility of new physics and providing redundancy.

Many of these functions require momentum measurement in a magnetic field.

Tracking at the SSC at the full design luminosity of 10^{33} cm^{-2}s^{-1} is expected to be a difficult problem. The limitations imposed by rates and radiation damage are severe. However, the dominant constraint is the combination of occupancy and double-hit resolution. Single events from new physics at the SSC have many (several hundred) charged particle tracks and are further complicated by curling tracks in a magnetic field, photon conversions, hits from events from out-of-time bunch crossings, and multiple interactions within the same bunch crossing. These problems can probably be solved, but at the cost of mechanical complexity and many signal channels.[1] We report here on a computer simulation study of tracking in complex SSC events and discuss some implications for triggering and data acquisition.

2. TRACKING SYSTEM REQUIREMENTS

2.1 The SSC Environment

The design luminosity, $\mathcal{L}$, of the SSC is 10^{33} cm^{-2}s^{-1} with an energy of 40 TeV in the center of mass. The inelastic cross section, σ, at 40 TeV is expected to be about 100 mb, which gives 10^8 interactions per second at the design luminosity. The bunch separation is 4.8 m, so the time between bunch crossings, t_B, is 16 ns, which leads to an average number of interactions per bunch crossing, n_I, of 1.6 at the design luminosity. Most of these interactions are minimum bias events or low-p_T hard scattering processes in which particle production is expected to be uniform in rapidity. The average charged particle multiplicity per unit of rapidity, n_c, is expected to be 7.5 over the rapidity range $|\eta| < 6$.[2] Figure 1 (from Ref. 2) shows the resulting charged particle flux and annual dose as a function of perpendicular distance from the beam for standard SSC operating conditions.

2.2. Rates and Radiation Damage

Radiation damage and rate limitations impose severe constraints on charged particle tracking detectors at the SSC, as described in several references.[1,3] These constraints are summarized here since they are necessary considerations for the design of any SSC tracking system.

A tracking system for the SSC is assumed to be made up of wires or other detectors running (nearly) parallel to the beam line. The width, w, of a cell is

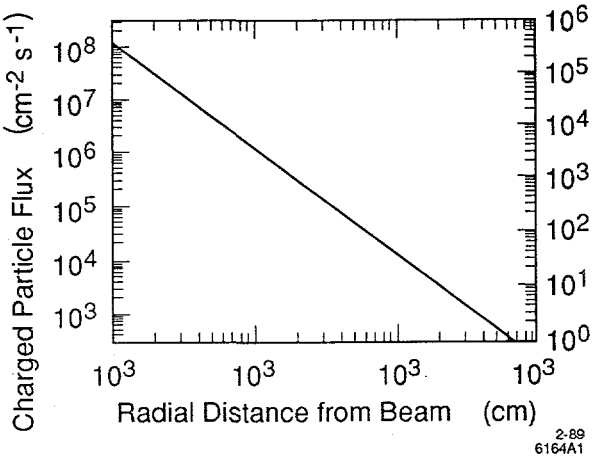


Fig. 1. The charged particle flux and annual dose as a function of perpendicular distance from the beam under standard SSC operating conditions (from Ref. 2).

assumed to be equal to the height, h, and the drift distance, d, is half the cell width. The ionization rate, α, in the gas is assumed to be 100 electrons/cm. The gas gain, G, is assumed to be 2×10^4, which is rather low.

The flux of particles per unit length (ℓ) of wire in a cell at radius r is given by

$$\frac{d^2n}{d\ell\ dt} = \frac{n_c\, w\, \sigma\, \mathcal{L} \sin\theta}{2\pi r^2}, \tag{1}$$

where θ is the angle relative to the beam direction. The ionization produced by a charged particle at angle θ is $h\alpha/\sin\theta$, so the ionization per unit length of wire is independent of θ. Thus the current draw per wire, I, for a layer of wires of length L at radius r is given by

$$I = \frac{n_c\, w\, h\, \sigma\, \mathcal{L}\, G\, e\, \alpha\, L}{2\pi r^2}, \tag{2}$$

where e is the electron charge. A layer of 4 mm wide cells at a radius of 50 cm covering $|\eta| < 1.5$ ($L = 213$ cm) will draw 0.52 μA/wire. The limit of acceptable current draw before breakdown will occur is about 1 μA/wire.

Wire chamber lifetimes are measured in deposited charge per unit length of wire before a decrease in gain occurs due to the buildup of material on the wires. For the above example, the collected charge over a chamber lifetime of five years (5×10^7 s) would be 0.12 C/cm. Chamber lifetimes of 1.0 C/cm have been measured under

very clean laboratory conditions.[4] For the purposes of a realistic experiment, it is probably best to assume a chamber lifetime about an order of magnitude below this.

Changes in gain for wire chambers have been observed at the level of 10^5 particles/cm-s at a gas gain of $\sim 4 \times 10^5$ due to space charge buildup.[5] The particle flux is given by Eq. 1. For the above example, the flux would be 1.9×10^4 particles/cm-s at $\theta = 90°$ where the flux is maximum. Since the gas gain must be much smaller than 4×10^5 because of current draw and lifetime considerations, space charge should not be an important limitation.

The hit rate per wire, R, for SSC tracking chambers is quite large and is given by

$$R = \frac{n_c \, \eta_{max} \, \sigma \, \mathcal{L} \, w}{\pi \, r} \tag{3}$$

for chambers covering $|\eta| < \eta_{max}$. For the above example, the hit rate per wire would be 2.9 Mhz. Existing electronics can probably handle rates of $\sim$ 10 Mhz.

A very serious limitation for tracking systems at the SSC is occupancy. Since the time between bunch crossings at the SSC is shorter than the resolving time of a typical drift chamber cell, the cell is sensitive to several bunch crossings. The occupancy, O, is given by

$$O = \frac{2 \, n_c \, \eta_{max} \, n_I \, n_B \, d}{\pi \, r}, \tag{4}$$

where n_B is the number of bunch crossings during the resolving time of the cell. n_B is given by

$$n_B = 1 + \mathrm{int}\,(t_R/t_B)\big[\, 2 - t_B/t_R - (t_B/t_R)\,\mathrm{int}\,(t_R/t_B) \,\big], \tag{5}$$

where t_R is the resolving time of the cell, d/v_D, for drift velocity v_D, and $\mathrm{int}\,(x)$ is the largest integer $\leq x$. Actually, n_B is very close to $t_R/t_B = d/(v_D \, t_B)$. A 4 mm wide cell (2 mm drift) has a resolving time of 40 ns for a typical drift velocity of 50 μm/ns and is therefore sensitive to 2.6 bunch crossings. A layer of such cells at a radius of 50 cm and covering a rapidity range $|\eta| < 1.5$ would have an occupancy of 12% per cell. It is guessed that an occupancy of $\sim$ 10% is reasonable, but a realistic answer depends on the effects on pattern recognition and track finding, which are discussed in more detail in Section 3. The real limitation to occupancy is due to the double-hit resolution because of the loss of information. A faster gas,[6] such as mixtures of CF_4 with a saturated drift velocity of 125 μm/ns, would improve the situation considerably by reducing the occupancy for a given cell width. Silicon

microstrip or pixel devices have very small "cells" and so avoid the problem of high occupancy, but they must either be able to operate in the higher radiation environment at small radius or be built into a very large silicon tracking system.

The rates given above are based only on particles produced in an interaction and must be increased by a factor of 2–4 because of curling tracks in a magnetic field, converted photons, and albedo particles leaking out of the front face of the calorimeter. Regardless of pattern recognition considerations, the effects on current draw and chamber lifetime must be carefully considered in the design of any SSC tracking system based on wire chambers.

2.3. Tracking System Considerations

2.3.1. Cell Size and Shape. For the reasons discussed in the previous Section, cell widths are constrained to a few mm. Straw tube chambers are a natural candidate for a small cell design. Construction possibilities for a tracking system made of straw tubes are discussed in Refs. 1 and 7. A tracking system based on jet cells, such as shown in Fig. 2, might also be considered. However, since the jet cell would be only a few mm wide and would be tilted so that the electron drift trajectories in a large magnetic field would be perpendicular to the sense wire planes, the advantages of a jet cell, e.g., uniform drift electric field over most of the cell and long track segments in one cell, are lost. In addition, there would be large forces on the endplates due to the tension of the large number of field wires. Also, as in any open wire geometry, it would be difficult to devise a mechanism to support the long sense wires in order to maintain electrostatic stability. On the other hand, straw tubes provide the possibility of supports for the wires.

The straws are typically made of aluminized polyester film (Mylar) or polycarbonate (Lexan) with wall thicknesses of about 30 μm. Several layers of straw tubes can be glued together to form superlayers which would be rigid, mechanically stable structures. Within each superlayer the layers can be staggered by half the cell width in order to allow hits from out-of-time bunch crossings to be rejected and resolve left-right ambiguities, as illustrated in Fig. 3. By dividing the chamber into superlayers, locally identifiable track segments can be obtained at the pattern recognition stage. The track segments can then be linked to form tracks. There must be a sufficient number of layers in the superlayers to provide redundancy.

2.3.2. z-Reconstruction. The wires or other sensing elements are assumed to run parallel, or nearly parallel, to the beam direction, or z-axis. The three conventional methods for measuring the coordinate along a wire are charge division, small-angle stereo, and cathode strips (or pads) running perpendicular to the wires. A fourth, less conventional, method is the time-difference method which probably has similar resolution to charge division, but may be worth further consideration.

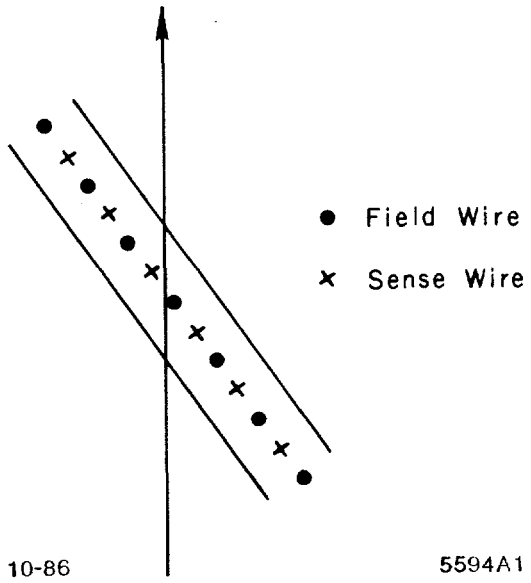


Fig. 2. Narrow tilted jet cell with radial track.

Charge division, at best (high gas gain $\sim 10^5$), gives z-coordinate resolution of about 1% of the length of the wire. Since the wires in an SSC tracking system would be quite long (3–6 m) in order to cover the required rapidity range, the resolution would be only 3–6 cm. Since low gas gain is needed to reduce current draw and increase chamber lifetime, the resolution in an SSC tracking system would be even worse. Also, charge division requires electronics readout at both ends of the wire which increases the complexity of a system with a large number of wires. For these reasons charge division does not appear to be a practical method for measuring the z-coordinate in an SSC tracking system.

Small-angle stereo ($\sim 3°$) wires typically give z-coordinate resolution of a few mm (the drift distance resolution divided by the stereo angle). The same electronics for time measurement can be used for all wires. In a system of superlayers of straw tubes, every other superlayer might be small-angle stereo. However, in complex SSC events it may be difficult to associate the hits on stereo wires with the correct tracks.

Azimuthal cathode strips or pads can give a z resolution of better than 1 mm. They might be included on the outer surfaces of the superlayers to aid in bunch assignment and reducing stereo ambiguities. However, they present added electrical and mechanical difficulties, as well as increasing the number of readout channels.

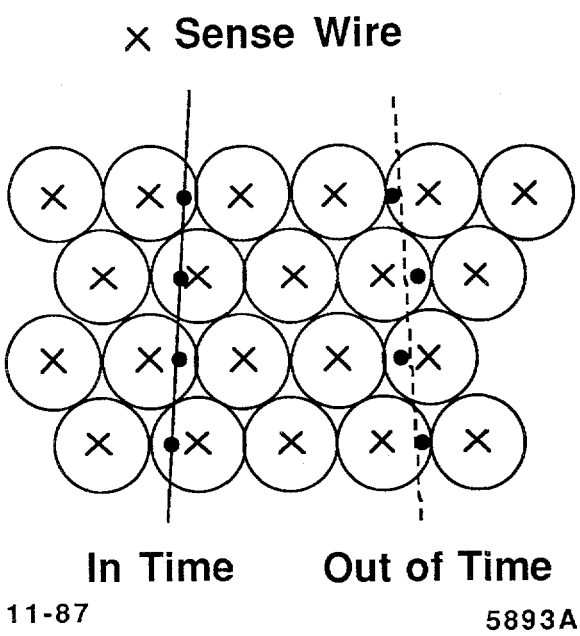


Fig. 3. Layers of straw tubes in a superlayer with every other layer staggered by the straw tube radius. A single in-time track will appear as a series of hits on the wires on alternate sides of the track. The left-right ambiguity is easily resolved locally. A track from an out-of-time bunch crossing will produce hits which are displaced from possible tracks by at least 16 ns in drift time.

2.3.3. Momentum Measurement. At the 1987 Berkeley Workshop[8] an examination of the requirements for momentum resolution based on the physics led to the criterion that the sign of the charge for electrons should be measured for $p_T \leq 0.5$–1.0 TeV/c. The momentum resolution is given by[9]

$$\frac{\sigma_{p_T}}{{p_T}^2} = \sqrt{\frac{720}{1+5/N}} \left(\frac{\sigma_x}{0.3\, q\, B\, D^2 \sqrt{N}} \right) \quad , \tag{6}$$

where p_T is the transverse momentum of the particle in GeV/c, q is the charge in units of the electron charge, σ_x is the spatial resolution in m, B is the magnetic field in Tesla, D is the track length in m, and N is the number of measurements, assumed to be equally spaced. Momentum resolution of $\sim 30\%$ is needed for charge sign determination. As an example, charge sign determination for $p_T \lesssim 450$ GeV/c could be obtained with a spatial resolution of 150 μm, 2 Tesla magnetic field, track length of 1 m, and 100 measurements.

2.3.4. Electronics Considerations. As discussed above, wire chambers for SSC tracking are required to have a small-cell design. Straw tube chambers seem to be particularly suitable. Since these chambers must have low gas gain, preamplifiers need to have low noise. Small-cell wire chambers probably have no multihit capability within a single event since the width of the pulses is approximately equal to the maximum drift time across the cell, so careful pulse shaping and digitization are probably not useful. A pole-zero filter to suppress the $1/t$ tail from pulses from previous bunch crossings is needed. Fast leading-edge timing using a threshold, double-threshold, or constant-fraction discriminator is dictated. The resolution for drift-time measurement should be $\lesssim 500$ ps, which corresponds to 25–60 μm depending on the drift velocity. All of the electronics – preamplifiers, pulse shapers, discriminators, time measurement electronics, track processors to find track segments, and digital or analog pipelining – should be located on the tracking detector in order to reduce the number of cables and processing time. The implication is that the electronics must have low power dissipation as well as radiation hardness. Electronics for cathode strips or pads is also needed.

2.3.5. Example of an SSC Tracking System. A large solenoid detector based on more-or-less "conventional" technology was discussed at the 1987 Berkeley Workshop.[10] Calorimetry and tracking were located inside a large superconducting solenoid with 2 Tesla field. A schematic view of the Large Solenoid Detector is shown in Fig. 4.

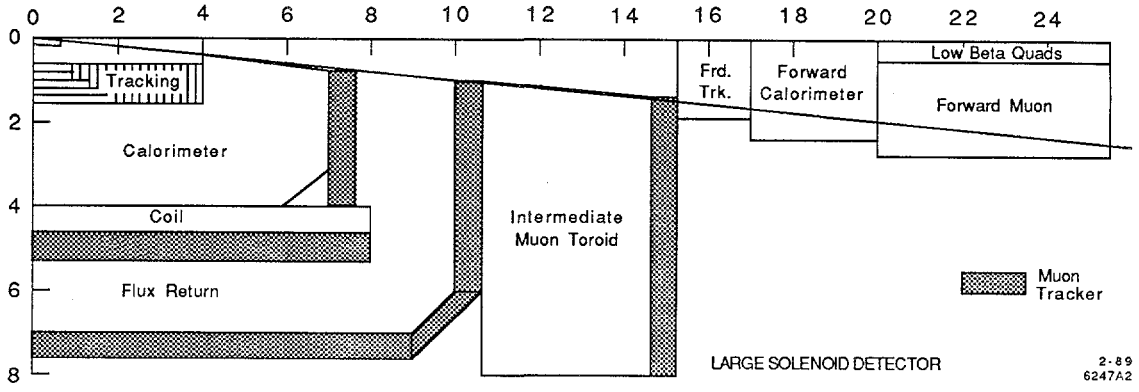


Fig. 4. Schematic view of the Large Solenoid Detector from the 1987 Berkeley Workshop.[10]

Table 1. Summary of Large Solenoid Detector Central Tracking System (from Ref. 10)

Superlayer Number	Inner Radius (cm)	Module Thickness (cm)	Half Length (cm)	Straw Diameter (mm)	Rapidity Range	Cell Occupancy (%)
1	40	2.7	85.2	3.92	1.50	12.1
2	48	2.7	85.2	3.92	1.34	9.1
3	56	2.7	119.0	3.92	1.50	8.8
4	64	2.7	119.0	3.92	1.38	7.0
5	72	4.1	119.0	5.89	1.28	13.0
6	80	4.2	170.0	6.04	1.50	14.5
7	88	4.2	170.0	6.17	1.41	12.9
8	96	4.3	170.0	6.28	1.34	11.6
9	104	4.4	170.0	6.38	1.27	10.5
10	112	4.5	238.5	6.47	1.50	11.9
11	120	4.5	238.5	6.55	1.44	10.9
12	128	4.6	238.5	6.61	1.38	10.0
13	136	4.6	238.5	6.68	1.33	9.3
14	144	4.6	238.5	6.73	1.28	8.5
15	152	4.7	238.5	6.78	1.23	7.9

The tracking detector design for the Large Solenoid Detector was divided into central tracking ($|\eta| \lesssim 1.2$) and intermediate tracking ($1.2 \lesssim |\eta| < 2.5$). The central tracking system was assumed to be built of straw tubes of radii from 2 to 3.5 mm parallel or nearly parallel to the beam direction. The straws are assumed to be at atmospheric pressure. Eight layers of straws are glued together to form superlayers. Within each superlayer the layers are staggered by half the cell width, as illustrated in Fig. 3. Every other superlayer is small-angle stereo ($\sim 3°$) in order to measure the coordinate along the wire. Azimuthal cathode pads or strips are included on the outer surfaces of the superlayers. The central tracking system extends radially from 40 cm to 160 cm with 15 superlayers in all. Only the superlayers at radii greater than 50 cm are expected to be operable at the full design luminosity. Assuming a spatial resolution of 150 μm, the momentum resolution which can be obtained with such a system is $0.54 p_T$ (TeV/c) using only wires at radii larger than 50 cm. If the particles are constrained to come from the interaction region, the momentum resolution would improve to $0.26 p_T$. The total number of cells is 122,368. The total number of radiation lengths is 8% for a particle traversing the central tracking chambers at 90°. The Large Solenoid Detector central tracking system geometry is summarized in Table 1.

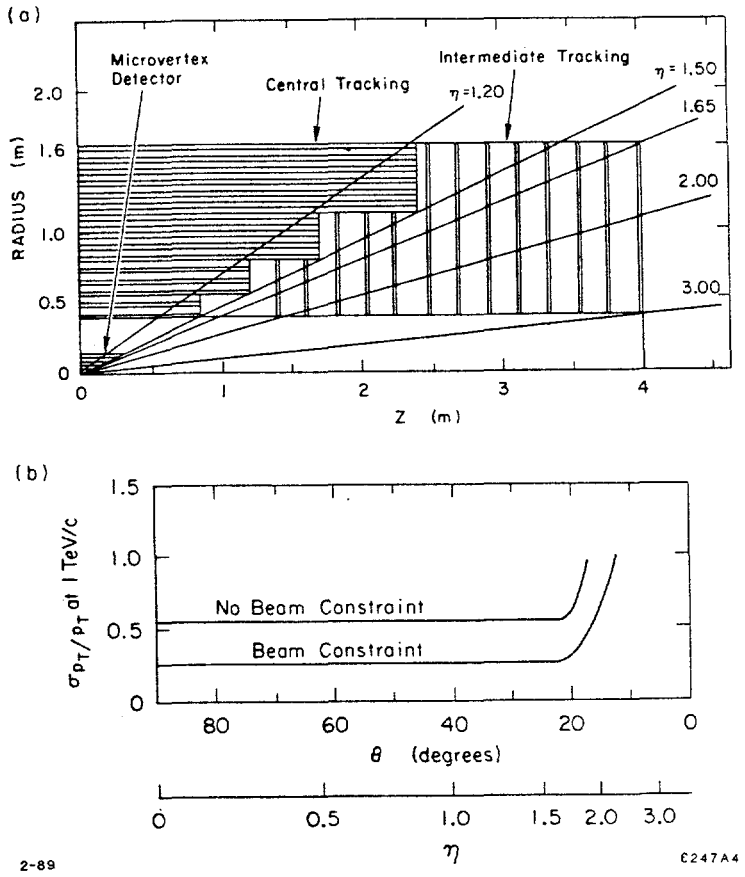


Fig. 5. (a) Schematic view of central and intermediate tracking systems in the Large Solenoid Detector. (b) Momentum resolution as a function of polar angle and rapidity in the Large Solenoid Detector for the 13 superlayers at radii > 50 cm in the central tracking system and including intermediate tracking (from Ref. 10).

In order to provide momentum measurement for $1.2 \lesssim |\eta| < 2.5$, the Large Solenoid Detector included tracking in the intermediate region to take over where the central tracking ends. Two options were considered: planes of parallel wires and radial chambers. The options for intermediate tracking have not yet been worked out in as much detail as the central tracking.

The central and intermediate tracking systems for the Large Solenoid Detector are shown in Fig. 5(a), and the momentum resolution as a function of polar angle and rapidity is shown in Fig. 5(b).

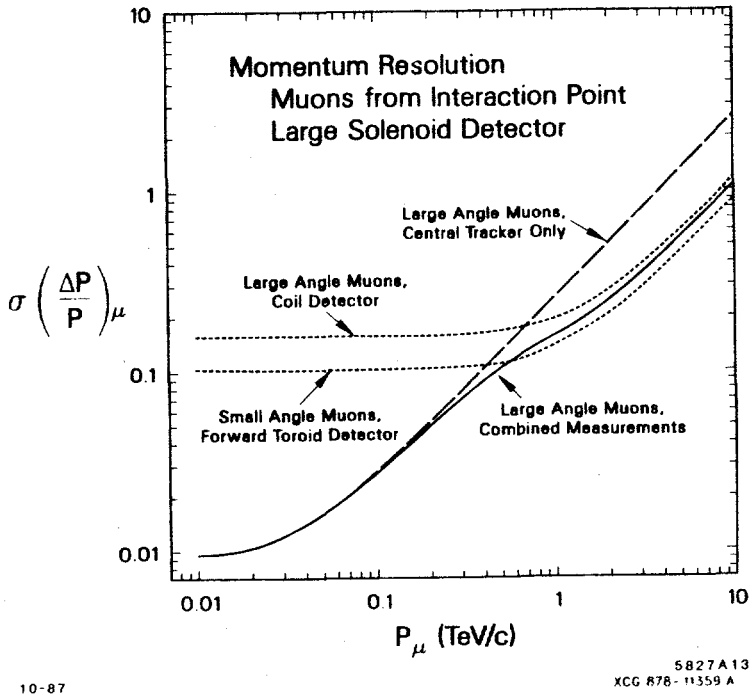


Fig. 6. Momentum resolution for measurement of large-angle and small-angle muons in the Large Solenoid Detector (from Ref. 10).

2.3.6. Muon Identification and Momentum Measurement. Another very important function of charged particle tracking in an SSC detector is muon identification and momentum measurement. The muon detection system in the Large Solenoid Detector[10] integrates muon momentum measurements with the central tracking detector, the large solenoid magnetic flux return yoke, and the calorimetry. Muons at small angles are measured by means of conventional magnetized iron toroids placed around the beam pipe. The relationship of the muon detection elements is shown in Fig. 4. In this conceptual design the spatial precision of points along the orbit in the bending direction is assumed to be 50 μm. The momentum resolution obtainable for both large-angle and small-angle muons is shown in Fig. 6. In both cases the momentum resolution is dominated by multiple Coulomb scattering below about 500 GeV/c for measurements made outside the calorimeter or magnet iron. The momentum resolution is improved considerably

for large-angle muons, especially for momenta below about 500 GeV/c, by combining measurements made in the central tracking system with those in the muon detection system. The momentum resolution remains at about the 10% level for momenta up to about 600 GeV/c in both the large-angle and small-angle systems.

How good does the muon momentum resolution need to be? Momentum resolution requirements for detecting the heavy Higgs boson in the mode $Higgs \rightarrow Z^0Z^0 \rightarrow 4\ell^{\pm}$ have been examined.[11] Figure 7 shows the di-lepton invariant mass for several energy or momentum resolution functions for a 400 GeV/c^2 Higgs decaying into $ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$: (a) $\sigma_E/E = 0.15/\sqrt{E}$ (E in GeV), (b) $\sigma_{p_T}/p_T = 0.54p_T$ (p_T in TeV/c), (c) $\sigma_E/E = 10\%$, and (d) $\sigma_E/E = 20\%$. The mass resolution in Fig. 7 (a) is quite good – a cut of $\pm$ 5 GeV/c^2 on the mass around the mass of the Z (M_Z) could be used. For Figs. 7 (b) and (c) one could still use $\Delta M_Z < 10$ GeV/c^2. However, some signal would be lost with this cut for the case of 20% resolution. The effects of momentum resolution on the Z-pair mass were then examined. Figure 8 (a)–(e) shows the Z-pair mass distribution along with the background from ZZ continuum production for a 400 GeV/c^2 Higgs boson for perfect detection and for the four energy or momentum resolutions shown in Fig. 7. We see that an energy resolution of $0.15/\sqrt{E}$ gives a Higgs peak which is indistinguishable from perfect detection because of the intrinsic width of the Higgs. The cases $0.54P_T$ and 10% energy resolution give somewhat broader but still observable peaks at the Higgs mass. However, 20% energy resolution broadens the Higgs peak so much that it is not distinguishable from background. The conclusion was that a momentum resolution of no worse than 10% was needed to find a heavy Higgs signal for a 400 GeV/c^2 Higgs.

2.3.7. Summary of Large Solenoid Detector Parameters. The design parameters of the Large Solenoid Detector are summarized in Table 2.

3. TRACKING SIMULATION

3.1. Simulation of a Central Tracking System for the SSC

The SSC central tracking system design used for this simulation was based on that for the Large Solenoid Detector[10], described in Section 2.3.5, although it is quite general and can be used for any system of cylindrically oriented sensing elements. All parameters of the detector, such as number of superlayers, number of layers in each superlayer, minimum and maximum radius and length of each superlayer, and azimuthal spacing between sense wires can be specified independently. The parameters used are as shown in Table 1, except that we included only the outer thirteen superlayers. We used a solenoidal magnetic field of 2 Tesla. The spatial resolution was taken to be 150 μm. So far, we have simulated only axial wires, that is, wires parallel to the cylinder axis.

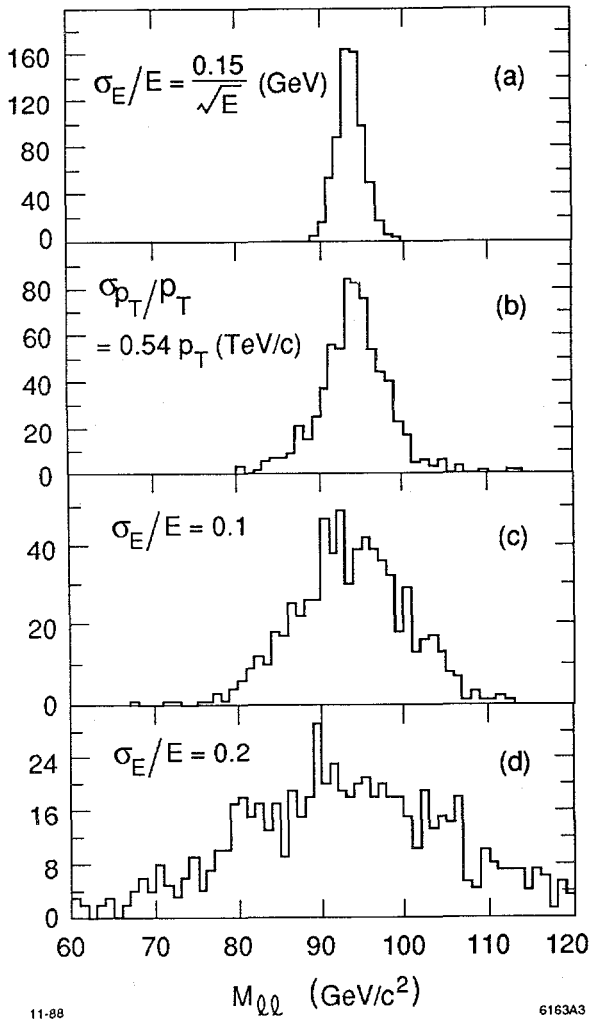


Fig. 7. Di-lepton invariant mass distributions for various energy or momentum resolutions for leptons from Higgs bosons of 400 GeV/c^2 mass decaying into $ZZ \to 4\ell^{\pm}$.

We used ISAJET[12] to generate events from interesting physics processes, such as high-p_T two-jet events or Higgs boson production, and from inelastic scattering background, for which we used minimum bias events. We used the GEANT3[13] general-purpose detector simulation package running on the SLAC IBM 3081 to simulate the interactions of the particles with the detector.

Using GEANT, the particles interact in the 8% of a radiation length of material due to straw tube walls, wires, and gas (the material was assumed to be distributed uniformly throughout the tracking volume), including photon conversion and multiple Coulomb scattering. The digitizations consist of a wire number

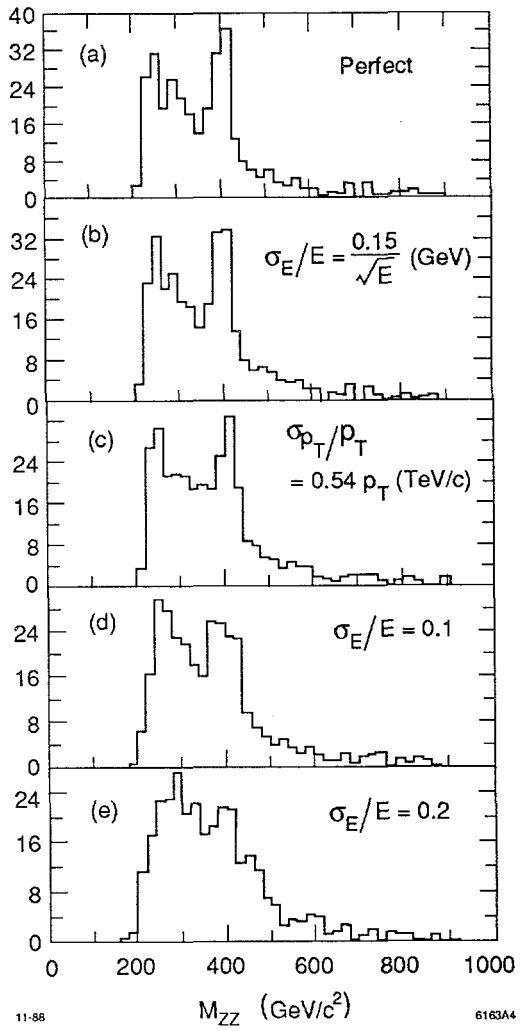


Fig. 8. Z-pair mass distributions for various energy or momentum resolutions for $H \rightarrow ZZ \rightarrow 4\ell^{\pm}$ for a 400 GeV/c^2 Higgs boson, along with the background from continuum ZZ production.

and a drift time, calculated from the distance of closest approach of a track to a wire using a drift velocity of 50 μm/ns, for each track in each layer. Background from inelastic scatterings in the same and out-of-time bunch crossings is included by superimposing the digitizations from minimum bias events. The number of bunch crossings is determined by the resolving time of the straw tube cell. At each bunch crossing the number of events to be included is determined from a Poisson distribution with a mean of 1.6 interactions per bunch crossing. Drift times from background events are then corrected for the time difference between the bunch crossing of the background event and the bunch crossing of the event of interest.

Table 2. Summary of Large Solenoid Detector Design Parameters (from Ref. 10)

SOLENOID COIL	
Inner diameter	8.2 meters
Length	16 meters
Central field	2 Tesla
Weight (including flux return)	16,450 metric tons
CENTRAL TRACKING	
Inner radius	0.40 meters
Outer radius	1.6 meters
Number of superlayers	15
Number of cells	122,368
$\|\eta\|$ coverage	< 1.2
INTERMEDIATE TRACKING (OPTIONS A & B)	
$\|\eta\|$ coverage	$1.2 < \|\eta\| < 3.0$
z position	$\|z\| < 4.0$ meters
Total number of chambers	26 (A) or 18 (B)
Total anode wires	128,000 (A) or 172,800 (B)
Total cathode pad channels	500,000 (A) or 293,760 (B)
ELECTROMAGNETIC CALORIMETER	
Depth	25 X_0
Transverse segmentation	$(\Delta\eta \times \Delta\phi)$
$\|\eta\| < 2.0$	$.02 \times .02$
$2.0 < \|\eta\| < 4.5$	$.03 \times .03$
$4.5 < \|\eta\| < 5.5$	$.03 \times .03$ to $.08 \times .08$
Longitudinal segmentation	6 X_0, 8 X_0, 11 X_0
Total number of towers	104,000
Total number of electronics channels	312,000
Weight	
Central	200 metric tons
Forward	35 metric tons
HADRONIC CALORIMETER	
Depth	10–12 λ
Transverse segmentation	$(\Delta\eta \times \Delta\phi)$
$\|\eta\| < 2.0$	$.06 \times .06$
$2.0 < \|\eta\| < 4.5$	$.06 \times .06$
$4.5 < \|\eta\| < 5.5$	$.06 \times .06$ to $.08 \times .08$
Longitudinal segmentation	2 segments
Total number of towers	19,100
Total number of electronics channels	37,200
Weight	
Central	4800 metric tons
Forward	965 metric tons
MUON SYSTEM	
Total number of electronic channels	$\sim$ 100,000
Weight of toroids	13,000 metric tons

The double-hit resolution is equal to the cell width, that is, only the earliest hit on a wire is kept. The simulation program is described in more detail in Ref. 14.

3.2. Results of the Simulation

We used the simulation described above to study tracking in SSC events. First we examined high-p_T ($p_T > 1$ TeV/c) two-jet events. Figure 9(a) shows such an event in the detector described in Section 2.3.5. Figure 9(b) shows an enlargement of the same event in the outer two superlayers in the area of the dense jet. Figure 9(c) shows the earliest hits in the cells for the tracks shown in Fig. 9(b). Hits from background events and converted photons are not shown in Fig. 9.

We can make the following observations, which still need to be quantified with high-statistics studies:

1. Although these events have very dense jets which seem at first to be impossible to resolve, when viewed on the scale of the wire spacings most of the hits appear to lie on identifiable tracks with a 2-Tesla magnetic field, particularly in the outer superlayers.
2. Eight layers in a superlayer is probably close to the optimum number because two tracks which are as close as the wire spacings produce hits only on every other layer because of the staggering. Some of these hits may be lost due to nearby curling tracks or background hits. Three tracks within the wire spacing distance would not be resolvable.
3. Although a 2-Tesla magnetic field produces curling tracks which obscure the high-p_T tracks to some extent, particularly in the inner superlayers, the effect in the outer superlayers is to spread out the tracks and, of course, remove the low-p_T tracks from consideration.

We next turned our attention to events from Higgs boson production, $pp \rightarrow HX$, with the Higgs decaying to Z^0Z^0 and both Z^0's decaying to e^+e^- or $\mu^+\mu^-$. We used a Higgs mass of 400 GeV/c^2. Such events allowed us to focus on the measurement of the high-p_T particles from the Higgs decay. Leptons from heavy Higgs decay typically have $p_T > 20$ GeV/c. Any large solid angle SSC detector must be able to measure such events. Also, these events are not as trivial to deal with as might have been naïvely guessed. There are many tracks from the underlying event and from the particles recoiling against the Higgs boson, even before adding the hits from background interactions. For these events we used the full simulation as described in the previous Section. An example of a Higgs event in the simulated central tracking system is shown in Fig. 10. We generated ~ 200 such events.

The fully-simulated events, including adding digitizations from minimum bias background events and removing digitizations within the double-hit resolution, had

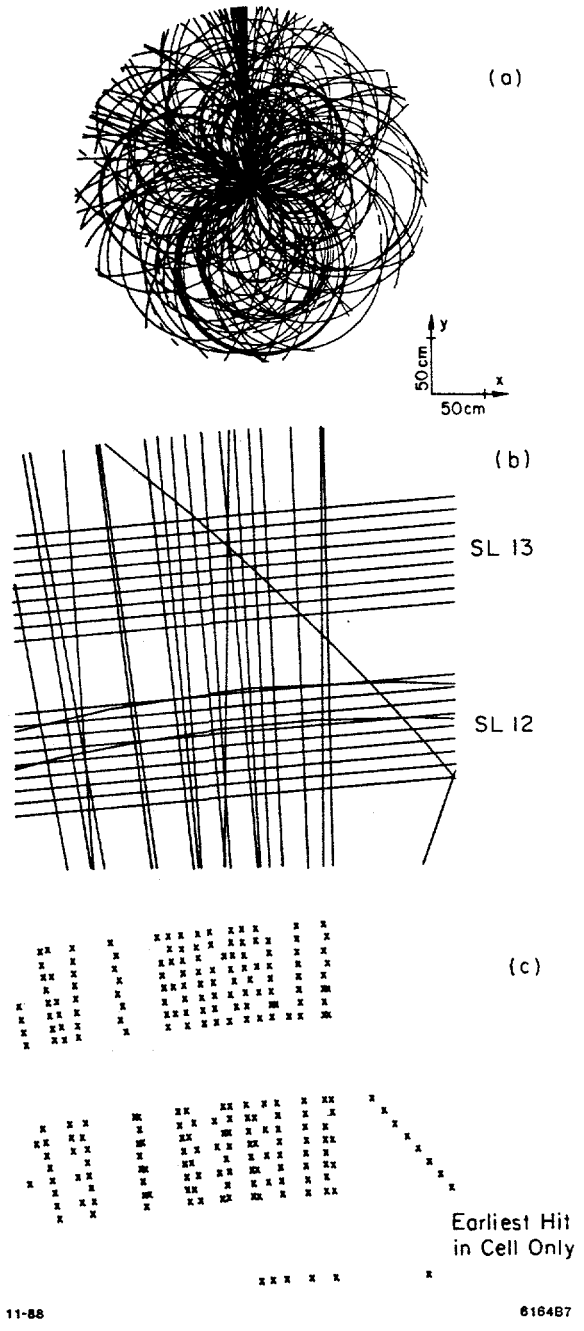


Fig. 9. (a) Two-jet event from ISAJET with $p_T > 1$ TeV/c in a 2-Tesla magnetic field in a detector of the geometry of the Large Solenoid Detector. There are 223 particles with $p_T > 200$ MeV/c and $|\eta| < 1.5$. Converted photons and background from minimum bias events are not shown. (b) Enlargement of the event in the outer two superlayers in the area of the dense jet at the top of the detector. (c) Earliest hit in each cell for the tracks shown in (b).

12,000 – 30,000 digitizations, as shown in Fig. 11(a). The fraction of digitizations from the minimum bias background events is shown in Fig. 11(b). On the average 57% of the digitizations were due to background events. For all tracks $(11.6 \pm 0.7)\%$ of the digitizations were lost because of the double-hit resolution, and the loss was about the same in all superlayers. For the leptons from the Higgs decay an average of $(7.3 \pm 0.6)\%$ of the digitizations were lost with the worst losses being in the inner superlayers.

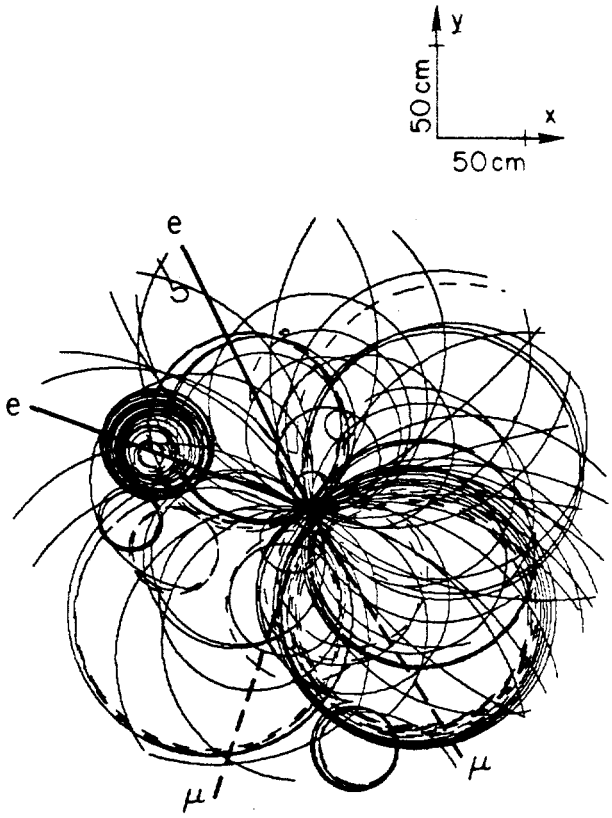


Fig. 10. Example of a Higgs event in the simulated central tracking system. The leptons from the Higgs decay are indicated by the heavier lines. Converted photons and other interactions with the material are included.

3.3. Pattern Recognition

We began working on pattern recognition algorithms in order to examine our original design goals of finding track segments in superlayers and removing hits

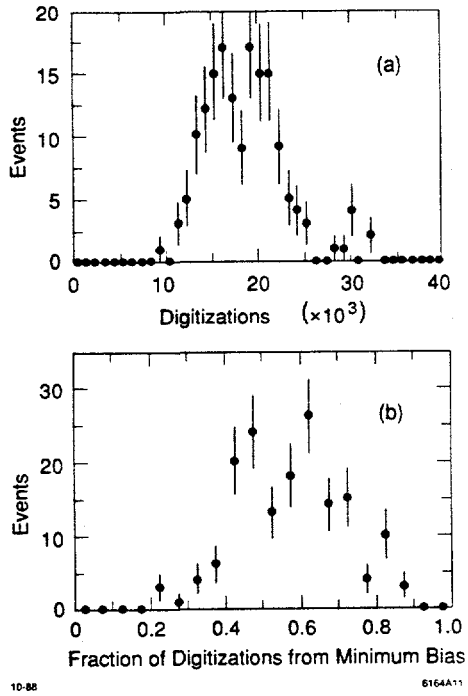


Fig. 11. (a) Total number of digitizations in Higgs events, including digitizations from minimum bias background events. (b) Fraction of digitizations from the minimum bias background events. The mean is 0.572 ± 0.011.

from out-of-time bunch crossings. We also wanted to make the algorithm simple with the hope of using it in the trigger. The algorithm for finding track segments was the following:

1. In each superlayer we identified "roads" containing hits. There are two parameters which can be varied: the width of the road and the number of hits required on the road. We used a width of five wires and required three or more hits out of eight possible. For isolated tracks one could require more hits; however, if two tracks are close together, as in Fig. 9, they will produce hits only on alternate layers and if one is lost due to the double-hit resolution there will be only three hits. The road requirement discriminates against low-p_T tracks.

2. We required that at least one of the hits be in a layer with the opposite wire stagger from the others so that the left-right ambiguities could be resolved and hits from out-of-time bunch crossings rejected.

3. We required that the hits be consistent with a straight line to within an error and in the process resolved the left-right ambiguities. Of course, the tracks approximate straight lines only locally within the superlayer, and the spatial resolution must also be taken into account.

Figure 12(a) shows all of the digitizations[15] for the event shown in Fig. 10, including those from minimum bias background events. Figure 12(b) shows only those digitizations which are included in segments. Keeping only those digitizations which form segments cleans up the events considerably. Figure 12(c) shows the tracks from the original event in the outer five superlayers in the region around the muon at the lower right. Figure 12(d) shows all of the digitizations in the event in the enlarged region (the digitizations are displayed at the locations of the hit wires). Finally, Fig. 12(e) shows only those digitizations which form track segments; here, the left-right ambiguities have been resolved, the drift times have been converted to distances, and the digitizations are displayed at the positions of closest approach of the tracks to the wires. One can clearly identify the muon track, and most of the extra hits have been removed.

Next, we applied our segment-finding algorithm to the e and μ tracks from Higgs boson decays. We defined two classes of segments: a "good" segment was one with at least five hits from a lepton track and no other hits, and an "OK" segment was one with at least five hits from the lepton track and one hit from another track. The effects of hits from other tracks remain to be studied; we plan to compare measured momenta with produced momenta in future work. With these definitions, we counted the number of segments found for each lepton track.

The distribution of the number of good segments for the e's and μ's in the Higgs events is shown in Fig. 13(a). The corresponding distribution of total (good or OK) segments is shown in Fig. 13(b). We see that the lepton tracks from Higgs decay have an average of about 8 good segments and 10 total segments out of 13 possible. Typically 30–50% of segments were good in the inner superlayers, increasing to almost 80% for the outer superlayers. When OK segments are counted as well, 50–60% of segments are accepted for inner superlayers and over 80% for outer superlayers.

3.4. Future Work

We are planning to continue our tracking simulation studies using the software we have developed. Future work will include simulation of small-angle stereo wires and cathode pads or strips for reconstruction of the direction along the wires; linking of segments, both axial and stereo, to form tracks; studying how much additional information is needed from cathode pads or strips to link the stereo segments properly; a more realistic simulation of electron drift in small-cell or straw tube

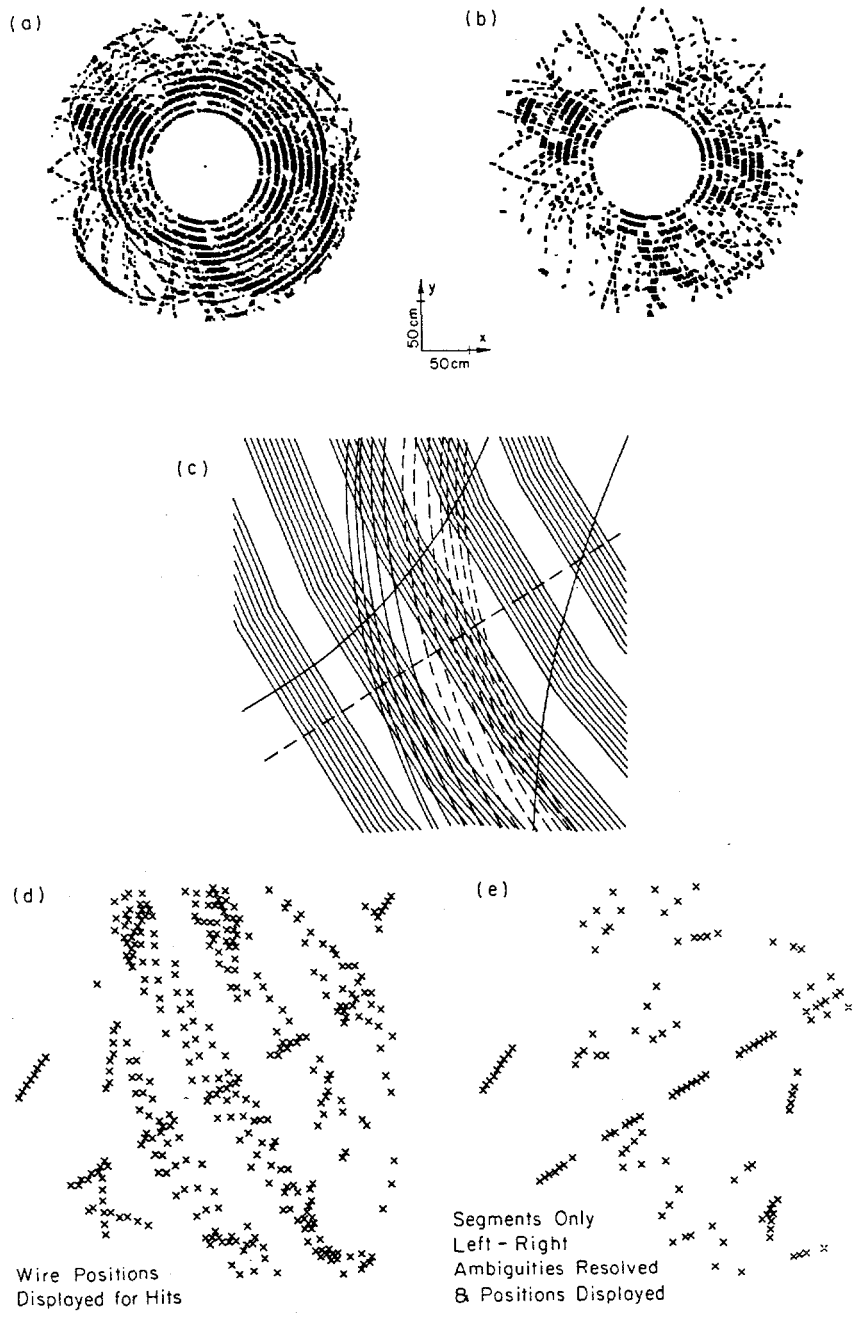


Fig. 12. (a) All of the digitizations for the Higgs event shown in Fig. 10, including those from minimum bias background events. (b) Digitizations for this event which are included in track segments, as defined in the text. (c) Tracks from the original event in an enlarged region in the outer five superlayers in the region around the muon at the lower right. (d) All of the digitizations in the event in the enlarged region of (c) (the digitizations are displayed at the locations of the hit wires). (e) Only those digitizations which form track segments in the enlarged region. Here, the left-right ambiguities have been resolved, the drift times have been converted to distances, and the digitizations are displayed at the positions of closest approach of the tracks to the wires.

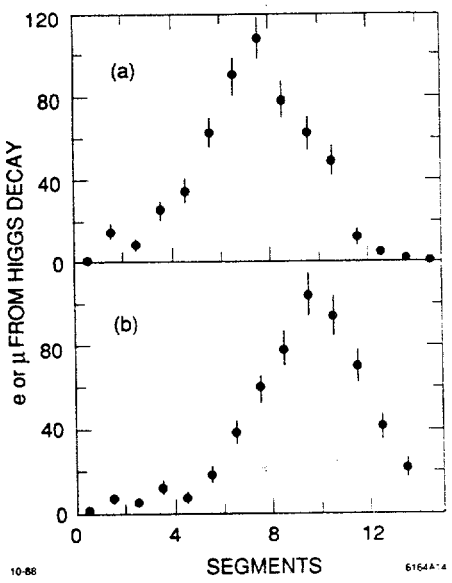


Fig. 13. (a) Distribution of the number of good segments out of 13 possible for the e's and μ's from the Higgs decays. (b) Distribution of the number of total segments (good or OK) for the leptons from the Higgs decay.

drift chambers, including the effects of $\mathbf{E} \times \mathbf{B}$; and conceptual design and simulation of intermediate tracking, as described briefly in Section 2.3.5. In addition, we will study tracking for different physics processes, such as new heavy fermions, supersymmetric particles, and high-p_T two-jet events, and begin to develop a realistic design for a tracking system for a complete SSC detector, including other detector components.

4. IMPLICATIONS FOR TRIGGERING AND DATA ACQUISITION

We have shown that an SSC tracking system design based on a pattern recognition strategy of finding track segments in superlayers appears to provide a powerful means of finding tracks in complex SSC events, even in an environment of multiple events from several bunch crossings. So far, detailed simulations have verified the concepts developed over several years for SSC tracking detectors. An algorithm for finding track segments such as that described here could be used in the trigger for high-p_T tracks. Depending on the effects on the physics analyses, we might envision making this requirement at the processor level, reading out only the hits that form track segments or even just the segments themselves.

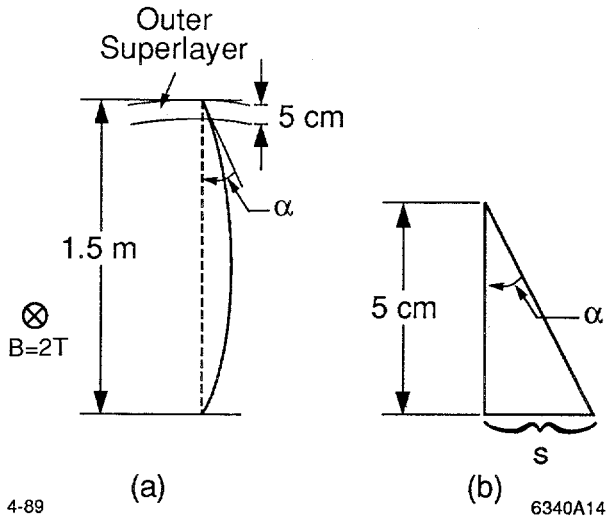


Fig. 14. (a) Illustration of the angle α between a radial line and the tangent to a track at the exit point of the track from the outer superlayer. (b) Illustration of the displacement s of a track from a radial line across the depth of a superlayer.

Track segments in superlayers can be characterized as local straight line segments since the sagitta over a superlayer is too small to measure for a relatively high p_T track. The two parameters characterizing a line segment could be the slope relative to a radial line at the track segment and the position given by the azimuthal angle. The slope is a measurement of the p_T of the track. The geometry for an outer superlayer is illustrated in Fig. 14. The radial extent of an outer superlayer is about 5 cm at a radius of about 1.5 m. The angle α is the angle between the tangent to the track at the outer superlayer and a radial line at the exit point of the track from the superlayer. The distance s is the displacement of the track from the radial line over the width of the superlayer. Table 3 shows the relationship between the angle α and the displacement s for various p_T values for a solenoidal magnetic field of 2 Tesla.

On the basis of these dimensions, one could imagine requiring $p_T > 50$ GeV/c in the trigger, but drift times are needed in the segment finding. An added complication is the need to have an estimate of the position of the track along the length of the wire since the propagation time is about 16 ns. Once track segments are found, they can be matched to clusters in the calorimeter, for example.

The triggering and data acquisition systems should be designed in such a way

Table 3. Angle Between Track Segment and Radial Line and Displacement Across Outer Superlayer for Tracks at Various p_T Values

p_T (GeV/c)	$\sin\alpha$	s (mm)
10	0.045	2.25
50	0.009	0.45
100	0.005	0.23
500	0.001	0.05

that they can evolve to greater sophistication as more is learned about the detector and background conditions. At first, we would need to read out all of the data and the track segments, presumably at lower than design luminosity, in order to find the constants for the time-distance relations in the straw tubes and the geometric positioning, optimize the tracking code, and check the segment-finding algorithms. Then later we might read out only the hits which make up the track segments along with the track segments. Finally we might gain enough experience to read out only the segments. However, we would always need to have the capability to read out all of the hits for at least some of the events in order to update constants for the time-distance relation and check the segment-finding algorithms.

Although a great deal of work remains to be done, we are optimistic that an SSC tracking system based on finding local track segments, whether in straw tubes or in silicon, will enable us to explore the new physics which awaits us in the SSC regime.

ACKNOWLEDGMENTS

We would like to thank members of the SLD Collaboration at SLAC, especially D. Aston, for their help in getting GEANT running on the SLAC IBM computer and for providing the interface to GEANT graphics. We also thank the organizers of the Workshop, particularly R. Orr, for providing such a stimulating environment. We gratefully acknowledge the support of the Department of Energy Program for Generic Detector Research and Development for the SSC.

REFERENCES

1. D. G. Cassel, G. G. Hanson *et al.*, "Report of the Central Tracking Group," *Proceedings of the 1986 Summer Study on the Physics of the Superconducting Supercollider*, edited by R. Donaldson and J. Marx, Snowmass, CO, 1986, p. 377.

2. *Radiation Levels in the SSC Interaction Regions*, Task Force Report, edited by D. E. Groom, SSC–SR–1033, SSC Central Design Group, June, 1988. For relativistic particles, rapidity and pseudorapidity can be used interchangeably. The pseudorapidity variable is given by $\eta = -ln\ (tan\ \theta/2)$ where θ is the angle relative to the beam direction.

3. *Report of the Task Force on Detector R&D for the Superconducting Super Collider*, SSC-SR-1021, SSC Central Design Group, June, 1986, pp. 44-60; M.G.D. Gilchriese, *Proceedings of the 1984 Summer Study on the Design and Utilization of the Superconducting Super Collider*, edited by R. Donaldson and J. G. Morfin, Snowmass, CO, 1984, p. 607; G. Hanson and D. Meyer, *ibid.*, p. 585.

4. J. Va'vra, *Proceedings of the Workshop on Radiation Damage to Wire Chambers*, edited by J. Kadyk, Lawrence Berkeley Laboratory, Berkeley, CA, 1986, p. 263.

5. A. H. Walenta, Nucl. Instr. and Meth. **217**, 65 (1983).

6. R. Thun, "Prospects for Wire Chambers at High Luminosity," to be published in *Proceedings of the 1988 DPF Summer Study on High Energy Physics in the 1990's*, Snowmass, CO, June 27–July 15, 1988.

7. R. DeSalvo, "A Proposal for an SSC Central Tracking Detector," CLNS 87/52.

8. R. J. Cashmore, S. Ozaki, and G. Trilling, "Summary and Comparison of High p_T Detector Concepts," *Proceedings of the Workshop on Experiments, Detectors, and Experimental Areas for the Supercollider*, edited by R. Donaldson and M.G.D. Gilchriese, Berkeley, CA, 1987, p. 301.

9. R. L. Gluckstern, Nucl. Instr. and Meth. **24**, 381 (1963).

10. G. G. Hanson, S. Mori, L. G. Pondrom, H. H. Williams *et al.*, "Report of the Large Solenoid Detector Group," *Proceedings of the Workshop on Experiments, Detectors, and Experimental Areas for the Supercollider*, edited by R. Donaldson and M.G.D. Gilchriese, Berkeley, CA, 1987, p. 340.

11. I. Hinchliffe and E. M. Wang, "Effect of Lepton Energy Resolution on Higgs Searches at the SSC," to be published in *Proceedings of the 1988 DPF Summer Study on High Energy Physics in the 1990's*, Snowmass, CO, June 27–July 15, 1988; E. M. Wang, G. G. Hanson *et al.*, "Higgs $\rightarrow$ Four Leptons at the SSC," *ibid.*.

12. F. E. Paige and S. D. Protopopescu, "ISAJET 5.30: A Monte Carlo Event Generator for pp and $\bar{p}p$ Interactions," in *Proceedings of the 1986 Summer*

Study on the Physics of the Superconducting Supercollider, edited by R. Donaldson and J. Marx, Snowmass, CO, 1986, p. 320. (The current version of ISAJET is 6.12.)

13. R. Brun, F. Bruyant, and A. C. McPherson, *GEANT3 User's Guide*, CERN DD/EE/84–1.
14. A. P. T. Palounek, "Simulating a Central Drift Chamber for a Large Solenoid Detector at the SSC," SLAC–PUB–4787.
15. The GEANT3 graphics does not display digitizations, only hits. To display digitizations, they must be converted to spatial coordinates and stored in the hit bank.

A SIMULATION OF DATA ACQUISITION SYSTEM FOR SSC EXPERIMENTS

Y. Watase and H. Ikeda

National Laboratory for High Energy Physics (KEK)

1-1 Oho, Tsukuba, Ibaraki, 305 Japan

ABSTRACT

A simulation on some parts of the data acquisition system was performed using a general purpose simulation language GPSS. Several results of the simulation are discussed for the data acquisition systen for the SSC experiment.

INTRODUCTION

Simulation of a system using computer is usual technique for an analysis of conceivable problems prior to any real design and construction. This process provides various opportunity to evaluate the system performance by changing design parameters and system configuration. In the case of the data acquisition system, there has been rare case to be studied by some simulation program because of the simplicity of the system functions and data stream[1]. But, in the case of the SSC experiments, the system becomes complicated not only in the physical or hardware structure, but also in the time progress. The system simulation helps the designer considerably.

A feature of the SSC experiments is its extremely high event rate due to the 62 MHz beam crossing rate, large cross section of the pp interaction as well as the high luminosity if it is achieved. The second feature is its large data size coming from the high multiplicity in an event, and also from the large number of signal channels in order to resolve the narrow and high multiplicity jet topology. Because of the above high frequency of the beam crossing, pipelined system is required in the front-end. At the back-end, event filtering by powerfull computer farm is absolutely necessary. The typical numbers of the electronics channels in a large 4 π detector amount about 800,000. The anticipated trigger rates are respectively, 100 kHz and 1 kHz for the first and the second level trigger. The final event rate to be stored in the magnetic tape is assumed to be order of 1 Hz.

The aim of the system simulation is to find a deadlock, bottleneck and dead-time within the data acquisition system in the stage of the system design, which in turn provides optimization of design parameters of the system, such as depth of a FIFO buffer,bandwidth of the data transfer, required power of micro-computers. At the same time, the simulation gives understanding of the relevant system and also a tool for common understanding of the system in the large collaboration of wide spreaded groups. In the various stages of the system design and debugging, the simulation can be used.

After a brief introduction of the general purpose simulation language:GPSS, some examples of the system simulation for the data acquisition are described. The present results were obtained from the simulation using FACOM GPSS/X : product of FUJITSU Ltd..

SIMULATION LANGUAGE: GPSS

The GPSS is a general purpose simulation language originally coded by IBM which can deal with an event oriented descrete system, such as a traffic control system, a ticketing center, and a flow control in the production line. Men, cars, and parts in these problems are expressed by TRANSACTIONs which flow in the system. Each TRANSACTION possesses attributed parameters to be assigned and referred during the simulation process. A system is described in GPSS with various functional Blocks such as FACILITY Block, STORAGE Block, ADVANCE(delay) Block, flow Control Block (ex. gate, traffic signal) and Statistical Block for system evaluation.

The functions of the GPSS is explained by a simple simulation for a ticket booth. The TRANSACTION (indicated by TRN hereafter) corresponds to a customer who wants tickets with different options. The TRN's are generated in the GENERATE Block at a certain frequency. Then, the needs of the customer; number of tickets and options are assigned to the parameters of TRN at the ASSIGN Block. The TRN proceeds to seize the FACILITY Block of the ticket booth. If the booth is occupied by a prior customer, the TRN is queued for waiting. When the booth is released by the prior TRN, the current TRN seizes the booth facility for the service time which is expressed by the ADVANCE Block. Then, the TRN releases the facility and advances to the next step. Finally the TRN's disappear from the system at the TERMINATE Block. The Fig. 1 shows the block diagram of this simulation procedure. Each characteristic shape of the block is specified by the GPSS simulation system. By inserting the statistical Block in front of the SEIZE Block followed by DEPART Block, the queuing information such as the length of the queue and the waiting time are obtained.

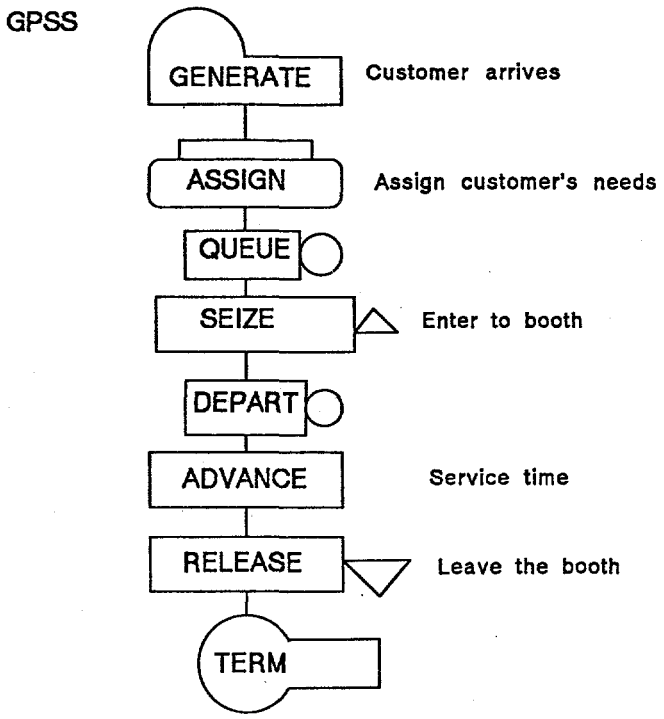


Fig. 1: A Block diagram of GPSS

MODEL OF DATA ACQUISITION SYSTEM IN GPSS

The performance of a data acquisition system can be evaluated by using a system simulation program that is one of the system design methodology. In the following, a modeling of the data acquisition system by GPSS is described.

The event data flow in the system can be expressed by a behavior of TRN in GPSS. The TRN is generated and some features of the event might be assigned to the parameters attributed to the TRN. These are such as particle multiplicity, momentum vectors of each particle, detector signals simulated, and others related information. The event rate is defined by an operand of the GENERATE Block. The time interval distribution is assigned by the functional name; ex. Poisson distribution, normal distribution, and exponential one. These procedure is illustrated in Fig. 2 for a conceptual data acquisition sytem at SSC. As shown in the figure, the TRN generated is split into the individual detector elements. Each data fragment in the detector element is processed along with the system description. The data is stored in the first level buffer which provides a pipeline to wait for the first level trigger decision. After the trigger, the data is transferred to the second level buffer. These

buffers are expressed as STORAGE Block in GPSS. The second level trigger transfers a specific event:TRN to the ADC, TDC or other digitization process. The particular information depending on the detector is additionally assigned to the TRN parameters. One of the possible parameters is a size of digitized data for individual event. Data transfer via backplane bus or serial transmission line can be expressed with FACILITY to be shared by TRN's. Pieces of an event data flowing the different system of the detector elements are synchronized by a MATCH Block for the coherent data transfer toward the event builder. Trigger logics and processor farm can be implemented in this simulation with real program tasks written by FORTRAN which runs cocurrently. The used CPU time by this task can be assigned as a delay time in the ADVANCE Block after scaling to the anticipated CPU power in the relevant system.

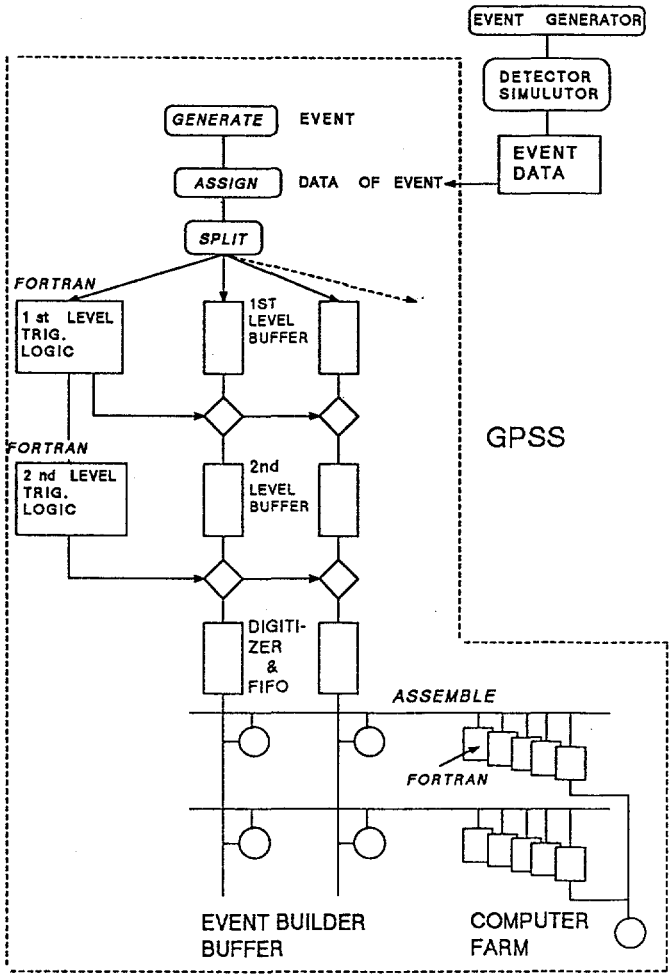


Fig. 2: Schematic diagram of a data acquisition system in GPSS

In order to show an example of GPSS programing, a simple system consisted of an ADC and FIFO buffer followed by a data transmission line is examined. The block diagram is illustrated in Fig.3. The ADC converts the arriving event signal into digital data within a fixed time of 30 μsec. The data are stored temporally in a FIFO with the depth of 10

and transmitted to the next stage through a data transmission line. The required time for the transmission is assumed as 50 μsec. The list of the GPSS program is shown in Fig. 4. The system evaluation is carried out by changing the arrival rate of the input signal. The rate is given by a exponential distribution with the mean of 2 kHz - 20 kHz. The result is shown in the Fig. 5. The number of FIFO depth used and the dead time percentage of the ADC are plotted. This result shows that the system functions marginally below the frequency of 5 kHz.

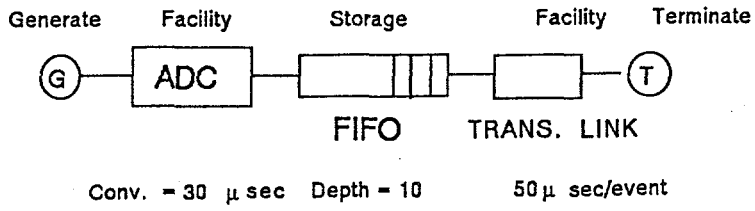


Fig. 3: Block diagram of ADC system

```
        FUNCTION    10,EXP,1    :event gen. function
FIFO    STORAGE     10          :define depth of FIFO
        GENERATE    FN1         :generate EVENT
        QUEUE       ADCQ        :QUEUE for ADC
        SEIZE       ADC         :get ADC
        DEPART      ADCQ
        ENTER       FIFO        :get FIFO
        ADVANCE     30          :ADC conv. time (30us)
        RELEASE     ADC         :conv. end
        SEIZE       TRNS        :readout start
        ADVANCE     50          :readout time(50us)
        LEAVE       FIFO        :remove from FIFO
        RELEASE     TRNS        :end of readout
        TERMINATE               :remove the event
```

Fig. 4: List of GPSS program

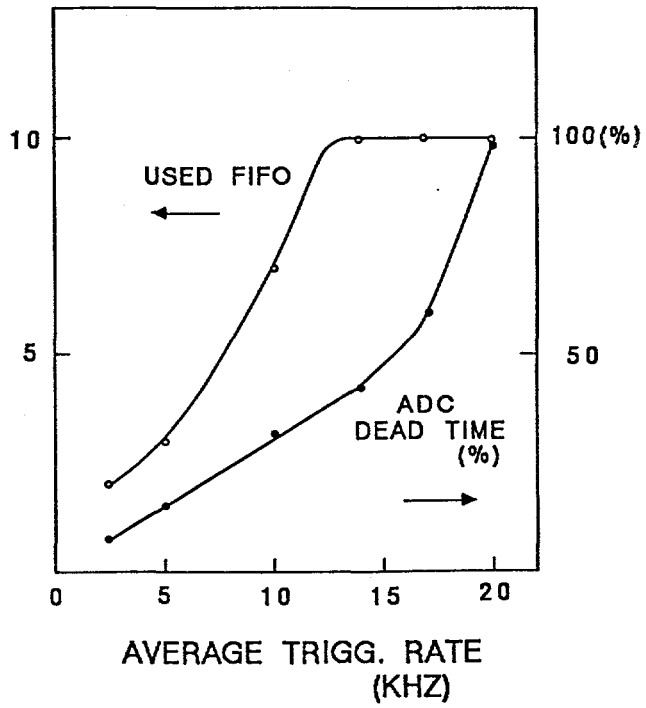


Fig. 5: The dead time percentage of ADC and the number of FIFO used

SIMULATION OF DATA ACQUISITION SYSTEM FOR SSC

The typical data acquisition system for SSC is shown in Fig. 6. The system contains the level 1 and 2 buffers to accomodate the delay of the first and second level trigger decision about 1 μsec and 10 μsec, respectively. In order to avoid the dead time, a pipelined trigger logics are needed as well as the above buffer memories. The second level trigger initiates the conversion of analog data to the digital which are buffered and read out into the dual port memory arrays. In this step, all data from various detector elements are built into an complete event data.

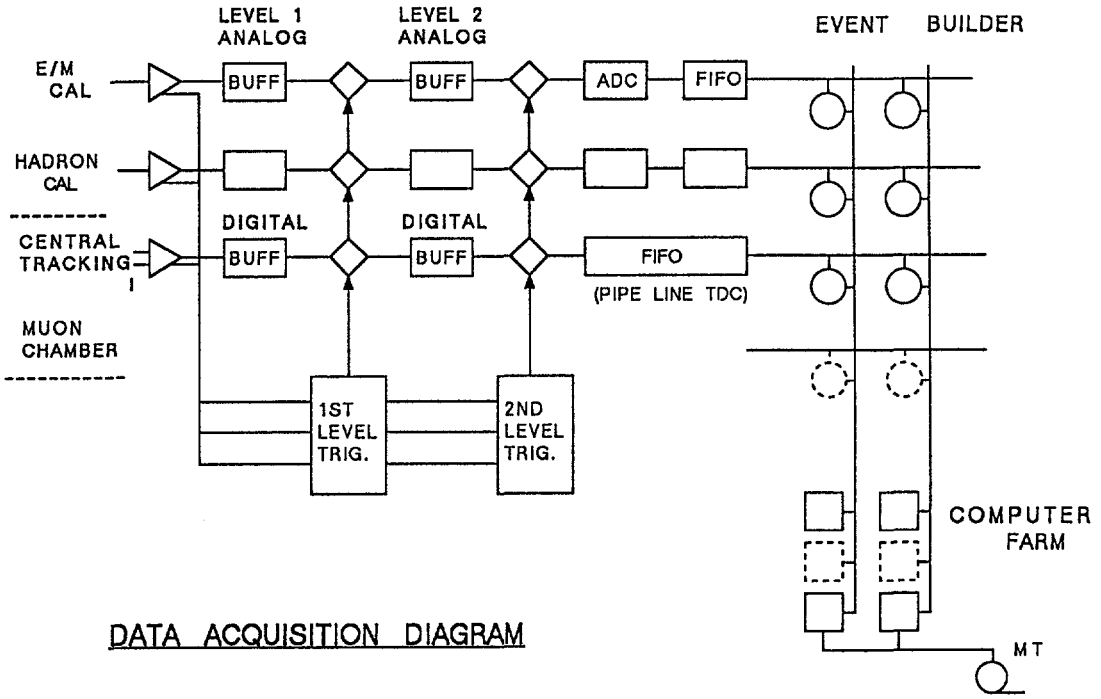


Fig. 6: Schematic diagram of the data acquisition system at SSC

Since only one unit of clock is available in the GPSS, a simulation must be carried out by a time unit of nanosecond if the system ranges from the frontend to the computer farm. That limits a time range of the simulation. However,in the simulation, certain time range is required, because the stream of TRN(event data) becomes steady state throughout the whole system after a certain relaxation time. Because of this limitation, the system after the second level trigger has been described in this paper.

The conditions and default parameters of the model are listed below.

Event generation	: Generate events following a time interval distribution of the exponential function with the mean of 1 kHz.
Feature of an event	: A number of particles within an event are ASSIGNed to the TRN by a Gaussian distribution with the mean: 200 and the sigma: 20.
Detectors	: Two major detectors of EM calorimeter and Central tracking detector are treated.
ADC system	: Conversion time of the ADC is 30 μsec which is effective time for several channels above threshold within a board. The data size is 0.1 kBytes/particle.
TDC system	: Pipe lined TDC with full digital fashion The data size is 0.4 kB/particle.
Data transfer line from the detector	: Effective speed is assumed as 100 MB/s
BUS of the computer farm	: The effective speed of data transfer is assumed as 100 MB/s
Event building	: Event building by a dualport memory buffer array. It contains 10 rows of buffer.
No. of CPU node in the computer farm	: 40 nodes per branch, 400 nodes in total.
CPU time for the processing	: CPU time depends on the number of particles to be reconstructed and processed. The default value is 1 msec/particle which correspond to 300 VAX's if the CPU time of CDF third level trigger processor is used as a standard for scaling[2].

The several results of the system simulation are discussed below. For the EM calorimeter section, a block diagram is shown in Fig. 7. The analog buffer,ADC and FIFO are operated

in parallel in the whole calorimeter readout system. The data transmission from the front-end system to the event builder is modeled with a data transmission link with the effective speed of 100 MB/s as the default value. The whole data from the calorimeter shares this transmission link. The data size from the ADC system is calculated from the number of particles in the particular event as 0.1 kB/particle. The average data size of the calorimeter per event is about 20 kB in the present model.

The use rate of the transmission line and the required depth of the FIFO buffer are shown in Fig. 8 as a function of the speed of transmission link. The other system parameters were fixed at the default values. If the transmission link is slow, the event data are stayed in the FIFO buffer, that requires the large capacity of FIFO. If the depth of the FIFO is designed as 20, the speed of the transmission link must be higher than 30 MB/sec as seen in Fig. 8 which corresponds to a optical fiber serial link of 300 M bits/s. The use rate of the link is 60 %.

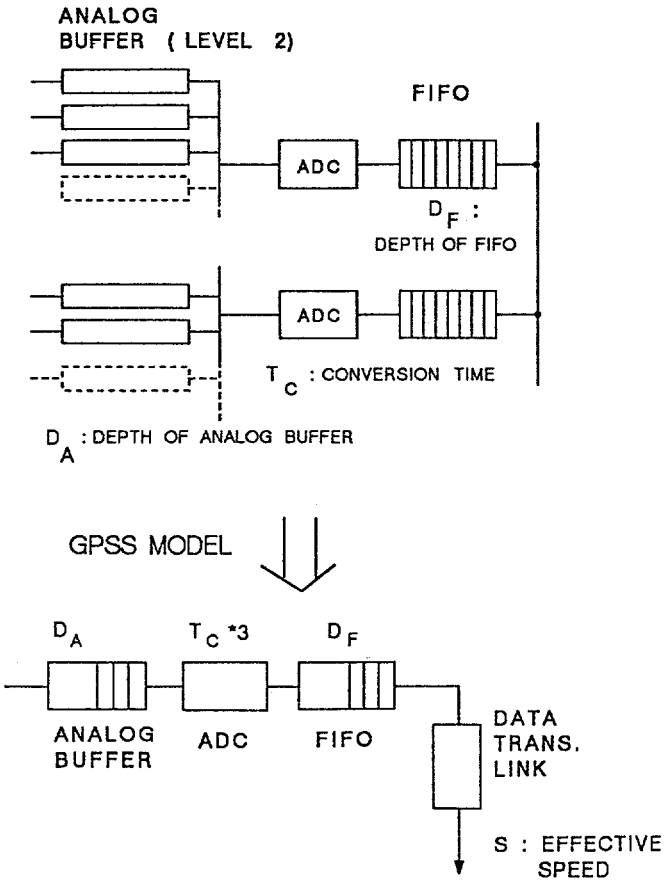


Fig. 7: GPSS model of the calorimeter section

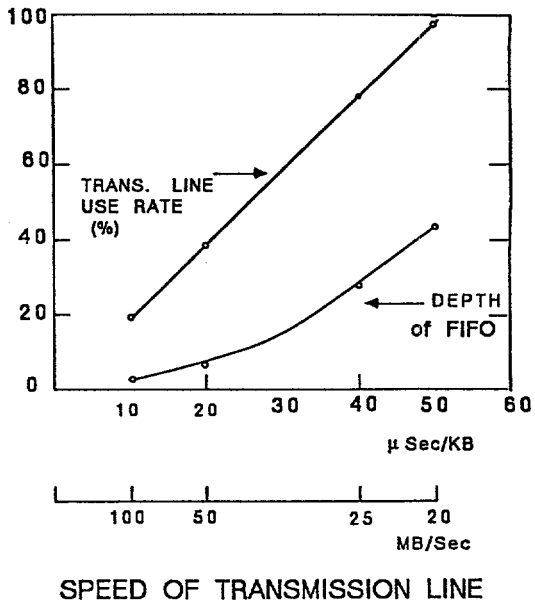


Fig. 8: Simulated results on the required FIFO depth and the use rate of the transmission line as a function of the speed

The second example of the simulation deals with the event building section where dualport memory array and computer farm are connected via the high speed BUS as shown in Fig. 9. Data of an event from the various detector elements are distributed and stored in a column of the buffer memory array. These data are transferred to one of the CPU in the farm through the BUS. The data size of an event is calculated from the number of the particles. The average size is about 250 kB per event. The buffer memory array has 10 rows as the default. The BUS speed is firstly examined by looking into the required depth of the FIFO for the various speed of the BUS. Fig. 10 shows results on the BUS occupation rate and the number of required depth of FIFO. In the lower speed than 30 MB/s, the BUS is full of the bandwidth, which cause the data over flow on the FIFO.

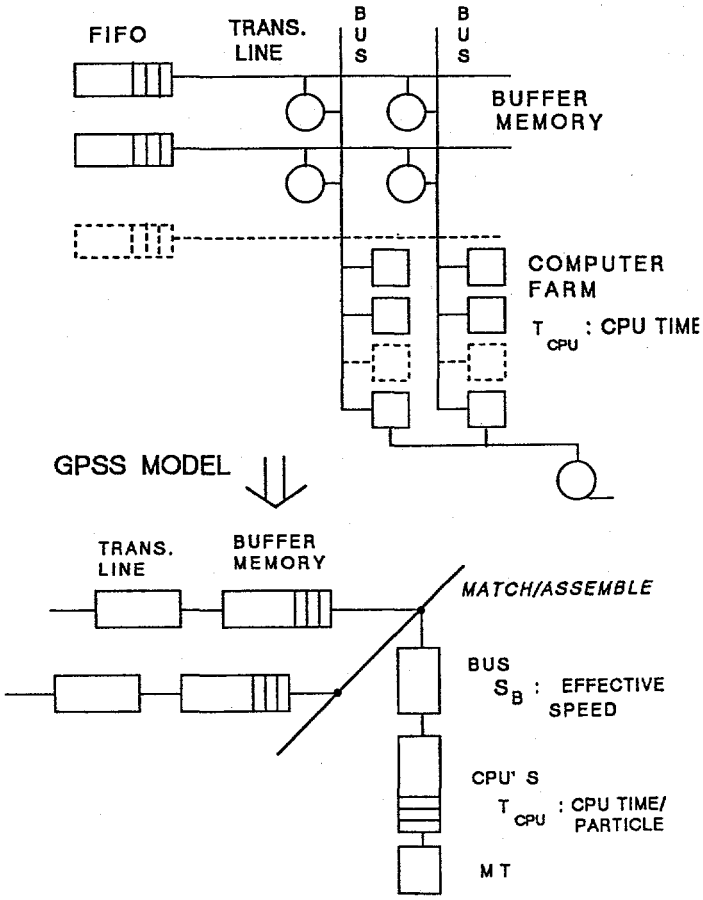


Fig. 9: GPSS model of the event builder section

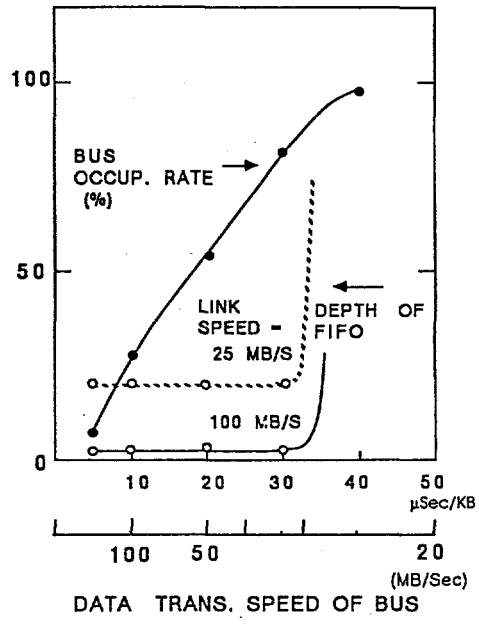


Fig. 10: Required depth of FIFO and the occupancy of BUS as a function the transfer speed of BUS

The similar over flow on the FIFO can be seen by decreasing the number of rows of the buffer memory as shown in Fig. 11. For the BUS speed of 100 MB/s, it occurs at the number of rows less than 8, and less than 24 for the BUS speed of 10 MB/s.

In order to examine the CPU speed of the computer farm in the final event filtering, the required number of CPU nodes are evaluated as a function of the speed of the CPU. The CPU speed is measured by a CPU time per particle normalized on some realistic data. In the present case, a speed of 1 msec/particle is supposed to correspond to the order of 300 VAX's equivalent[2]. But this number, of course, depends on the programing and the level of online filtering. The result of the simulation is shown in Fig. 12. The number of CPU in a branch which is connected to a column of buffer memory array is plotted as a function of its speed. These CPU's in a branch must accomodate the data processing for the event rate of 100 Hz because of sharing the 1 kHz event rate by the 10 rows of the buffer memory array (10 branches are availabale). If a CPU with the power of 1 msec per particle reconstruction (300 VAX's equivalent) is available, 25 CPU nodes are required for a branch. The total number of CPU nodes is 250 in this case.

Since the results shown above are based on tentative modeling and system parameters for the data acquisition and the detectors, the discussions described are some examples for a system analysis and design. The system simulation must be built and refined based on a realistic design of the experiment.

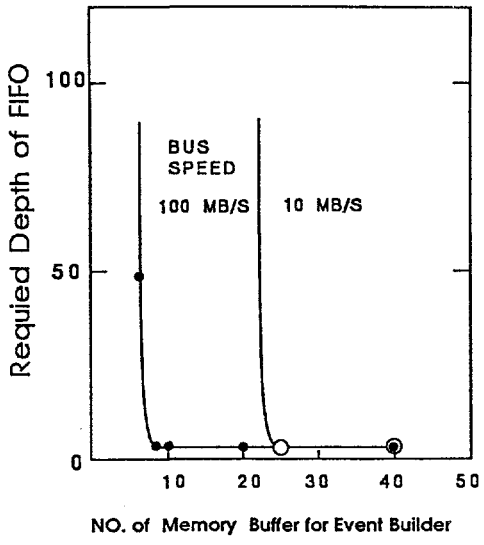


Fig. 11: Required depth of FIFO as a function of the number of memory buffer stages

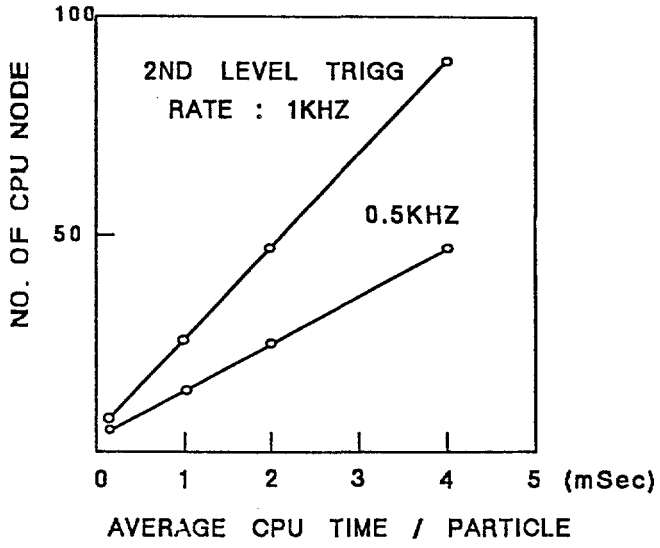


Fig. 12: Number of CPU nodes required for the farm as a function of the CPU speed (msec/particle)

CONCLUSIONS

A system simulation helps us greatly to design the system and understand its performance in a realistic conditions. In this paper, introductory examples are shown, but the full analysis must be done in the course of the detetor design in future. There might exist more powerful system simulation tools on a modern workstation which will provide interactive use under the user-friendly environments. The system simulation for the SSC experiment has to be built on such a platform to which various collaborators can access easily.

ACKNOWLEDGEMENTS

Authors wish to express their thanks to Prof. S. Mori for his encouragemnet and support, and also to Mr. H. Ishida of FHL for his advices on the program GPSS/X.

REFERENCES

1. D. Cutts and C. van Ingen, Proc. of Workshop on Triggering,Data acquisition and Computing for High Energy/High Luminosity Hadron-Hadron Colliders, Fermilab, Nov. 11-14,1985,p 216

2. Private communication on the experience at CDF third level trigger.

VALIDATING NON-PROMPT TRIGGERS IN CDF LEVEL 3

T.L.Watts
Rutgers University
February 16, 1989

ABSTRACT

Apart from normal techniques of debugging and testing offline software, some special problems are discussed which arise when software is buried in an online farm of microprocessors. Solutions should be considered at system design time.

DESCRIPTION OF RELEVANT ASPECTS OF THE CDF LEVEL 3 TRIGGER

The current farm of microprocessors is based on the Fermilab ACP design with 55 nodes each containing the MC68020 processor with 6 MB of memory. The host computer, the "Farm Steward", is a microvax II now in the B0 local Vax cluster.

Modules of software see almost the same environment as during offline processing and analysis, so migration is easy from offline to level 3.

The CDF "Analysis-Control" package is used to set the flow chart of the level 3 processing at the beginning of each run. The executable image is only relinked when code inside analysis modules changes. A trigger table text file for each run sets this flow chart and supplies parameters to the analysis modules to control trigger decisions (thresholds, etc).

The responsibility for creation of analysis modules belongs to the CDF physics groups, i.e. to individuals associated with analysing the data. There is loose central control of the creation of analysis modules.

TESTING AND DEBUGGING AFTER THE CREATOR DELIVERS A WORKING ANALYSIS MODULE

Similarity to Offline Testing. Problems of testing, debugging, & validation are very like those that occur in the creation of an offline production package, except that the trigger collection of analysis modules is smaller. However, the execution of the image takes place much more remotely, since it is buried in a farm of microprocessors which are in the data acquisition pipeline and often being used by others to test other parts of the system, or even at later stages to take data.

Transferring from Vax to ACP. Algorithm authors (creators of analysis modules) usually tested and debugged code in an offline environment. This caused problems in transferring to a different compiler and operating system in the ACP environment but not as many as might be expected because of the coding rules and pre-compiler processing adopted by CDF.

Stand-Alone Farm Steward. We used a "Stand-alone Farm Steward" and farm for the major step of assembling analysis modules into an executable image and for finding problems associated with the move from the offline Vax environment

to the ACP environment. This separate farm was independent of the online pipeline. The nodes in the farm used a node image which was identical to the online node images. The host vax simulated the data acquisition pipeline by reading events from a data file. Usually only 1-2 nodes were activated in the farm and printout from them appeared on the host interactive terminal. The node interactive debugger was used when appropriate. This separate farm was very important since the online farm was almost never available for testing and debugging.

Test Run. The next step in validating a level 3 executable image was to make a test data taking run of 4000-5000 events. The data was placed in a disk file and all algorithm authors were asked to inspect their trigger results and output banks and compare them against their standard offline code.

Tagging Only Running. After this, the level 3 executable image was placed in production data taking, but the new analysis modules and triggers were only allowed to mark events in the data stream without rejecting them. After further inspection of the data, event rejection using new analysis modules and triggers was turned on.

Status Display. During data acquisition, an online level 3 display driven by the host farm steward showed the status of each node. This sufficed to detect gross problems with node execution times or crashing nodes. Another display showed trigger statistics continually updated for the complete event stream.

IMPROVEMENTS

Better Simulation of Online Environment. The test bed Stand-alone Farm Steward host process was not placed in an environment enough like online, so that problems with reading external files such as data bases or with interacting with the external environment were sometimes not found until a new executable image was tried online during data taking. This obviously wasted luminosity. In addition, the data files used for testing in the stand-alone farm were not restored to their original state as they would have been seen by level 3 nodes online (level 3, as well as rejecting events, also processed events, and so output events were different from input). This problem was not critical in current CDF running but could become severe in future CDF data taking and in SSC experiments. The restoration of the data stream could conceivably be not trivial and might involve the taking of special runs to store data for this purpose.

Access to Print Statements from Nodes while Online. Bugs and troubles showed up at all levels of testing, and obviously great effort was made to catch them before software was used to take data. However, inevitable troubles showed up only after data taking was started where time spent debugging cost luminosity directly. This searching for troubles while online was difficult because most debugging tools were disabled or too time consuming. Eventually a useful method of monitoring the execution of algorithm modules in nodes was evolved whereby the Fortran print buffer (PRINT statements) for each event was added to the data stream for each event. This buffer was then dumped offline. As useful perhaps might have been the scrolling of one node's Fortran print statements on an experimenter's screen.

Access to Hung Nodes while Online. Occasional nodes (about 1 during 10000 input triggers to level 3) hang up in an exception state during data taking. During separate stand-alone tests these nodes could be inspected individually,

but during data taking it wasted luminosity to debug. An automatic way of saving the state of the node for later inspection would be very helpful here. A presumption about the cause of these node hangups is that they are related to corruption of the input data, so plans to save the input event are under consideration. For designers of new systems it might be more general to save the state of the complete node image.

Statistics in Each Event. CDF added run-accumulated trigger statistics to every event for monitoring purposes down stream. In a farm of microprocessors, only some central element like the farm steward can collect these statistics, so that extra transactions outside the normal flow of data were needed to transmit the array of data to each node to add to the event record. New experiments with more parallelism probably need at an early design stage to allow for monitoring of a large fraction of the event stream during data taking.

Pipelines of Farms while Debugging. In a new level 3 executable image, a change in an analysis module, or a new analysis module, may subtly overwrite into other long established modules. Authors of long established modules were reluctant to validate their triggers yet again when a new executable image was made to add another person's new code. Thus validation became tedious and got skipped sometimes. A possible way around this for future experiments (suggested by A.Mukherjee of CDF) might be to break up a farm temporarily during development to be several pipelined farms each with a different collection of software analysis modules. Events and processing results would be passed along the pipeline, and additions of new code in one farm would be more decoupled from other collections of analysis modules in other farms. This obviously is a radical modification of architecture and needs some careful thought.

Volatility of Offline Data Bases. The use of data bases in online triggers presents problems in later validation if the information in the data base can be modified after data taking for a particular run is finished. If the data base is a standard offline data base for instance, such modification is likely to happen because offline analysis will tune up equipment parameters, gains, thresholds, etc. Mechanisms can obviously be set up for later trigger simulation and validation with the data base information restored to its state at the time the particular run was taken. But too much complexity seems likely to creep in to the validation technique and should be avoided if possible. Particularly desirable is the option of writing algorithms to use only data base information which is very stable, or keeping a separate online data base for volatile information such as hot and dead calorimeter channels.

Online Verification. Because of the use of a dynamic memory management system (Ybos) for the L3 modules, the L3 code relies heavily on the pointers to the data 'banks'. Pointer errors may corrupt event data, calibration data, or code in such a way that subsequent events will be improperly handled, and errors may be history dependent. For this reason we plan to implement a verification system which will permit an event to be analysed on more than one node and will bit-to-bit compare the intermediate 'banks' and the final result. Errors will be signalled to the host and will result in the disabling of both the original node and the verifying node.

ACKNOWLEDGEMENTS

The remarks here are based on the experience of many people involved in the installation of the level 3 triggers, particularly P.Auchincloss, T.Carroll, T.Devlin, B.Flaugher, U.Joshi, P.Hu, K.Ragan.

OPTICAL DATA TRANSMISSION AT THE SUPERCONDUCTING SUPER COLLIDER

Branko Leskovar
Electronics Engineering Department
Lawrence Berkeley Laboratory
One Cyclotron Road
Berkeley, California 94720 U.S.A.

ABSTRACT

Digital and analog data transmissions via fiber optics for the Superconducting Super Collider have been investigated. The state of the art of optical transmitters, low loss fiber waveguides, receivers and associated electronics components are reviewed and summarized. Emphasis is placed on the effects of the radiation environment on the performance of an optical data transmission system components. Also, the performance of candidate components of the wide band digital and analog transmission systems intended for deployment in the Superconducting Super Collider Detector is discussed.

INTRODUCTION

The concept of guided lightwave communication along optical fibers has stimulated a major new technology over the past two decades. This technology profoundly impacts communication and instrumentation systems as well as computer interconnections and system architecture. Fiber optic links provide several major advantages over conventional electronic systems. These include immunity to electromagnetic interference, and low transmission losses for very high data rates. It also makes possible thinner and lighter cables and has a strong potential for long data transmission link capabilities extending to the gigahertz region.

The emergence of optical communication using fibers was made possible by the parallel development of low loss fibers, heterojunction lasers, light-emitting diodes (which emit in spectral regions of low fiber loss), and sensitive photon detectors. The technology of optical fiber communication systems is advancing at a very rapid rate. As the short optical wavelength multimode fiber systems are being field-proven and used commercially, the technology is progressing towards single-mode fiber systems in the long-wavelength region. For example, significant advances have been made in the fabrication of low-loss and low-dispersion optical fibers. Losses of approximately 0.20 dB/km at 1300 nm have been achieved for single mode fibers with minimum dispersion wavelengths near 1300 nm. Furthermore, the development of optical sources, and optical receivers, for long wavelength applications is also advancing rapidly. Several experimental transmission systems capable of operating at a 4 Gbit/s rate over a distance of 155 km and 16 Gbit/s over 8 km have been reported.[1,2]

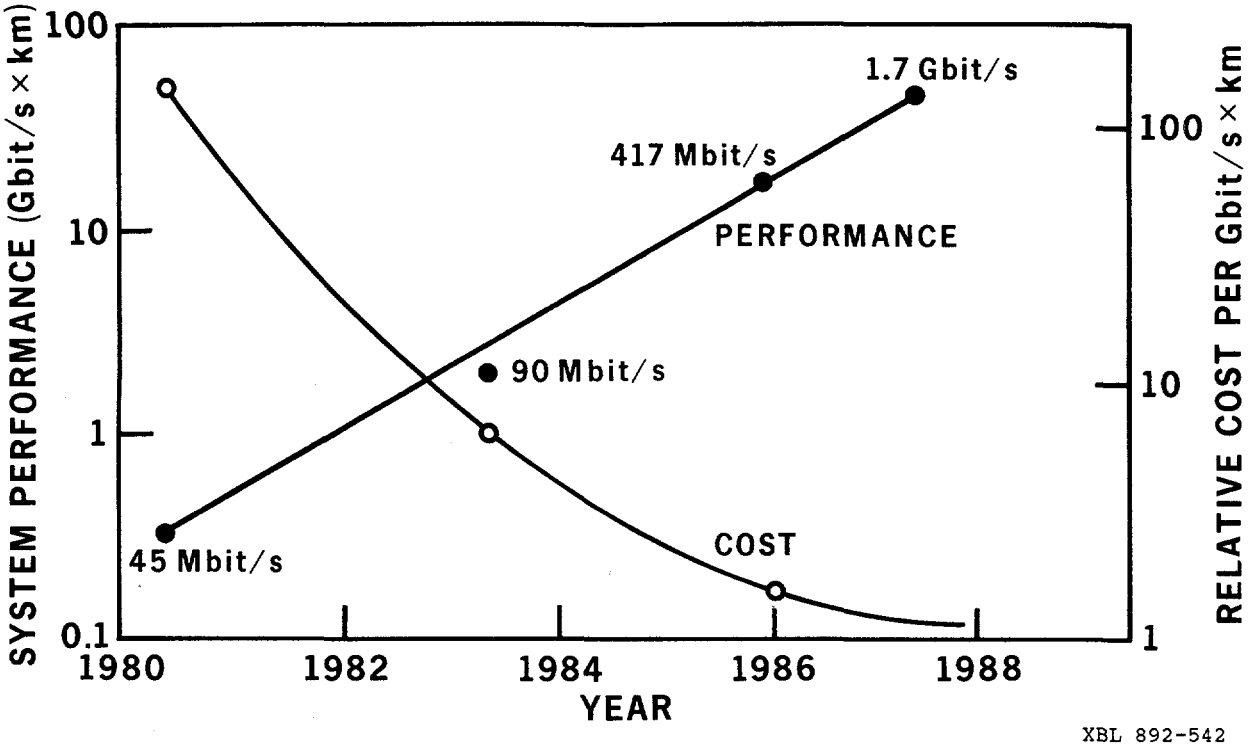


Fig. 1. Performance and cost of fiber optics communication systems.

Although these impressive results were obtained under highly optimized experimental conditions they do give an indication of future capabilities. It should be pointed out that at the present time practical high data rate optical transmission systems are operating between 45 Mbit/s and 1.7 Gbit/s. Figure 1 shows the communication system performance in the Gbit/s region and relative cost per Gbit/s × km as a function of time.[3] There has been approximately a tenfold improvement in the system performance every three years and a corresponding decrease in the cost per Gbit/s × km.

In addition to the requirements placed on the optoelectronic components of these systems, considerable attention was paid to the associated logic circuit families with switching speeds in the microwave region. Such capability is necessary for multiplexing and demultiplexing functions. These functions have been implemented using both silicon and gallium arsenide (GaAs) technology. Devices based on GaAs technology have become available recently for applications in practical systems[4] with a data rate capability of 1.5 Gbit/s.

Although the major thrust for development of high data rate fiber optic systems has been for long distance communication links, the local data communication needs have given a new impetus for the development of advanced system components. The recent systems for local communication, such as computer interconnections, rf distributions in phased array radars and instrumentation for basic and applied research, require a high data rate capability.

SSC DETECTOR SYSTEM CONSIDERATIONS

The Superconducting Super Collider Detector System will contain calorimetry, particle tracking, electron and muon identification subsystems, each involving typically 100,000 to 200,000 channels of readout electronics.

The overall detector system should be capable of operating with luminosity corresponding to an interaction rate of the order of 10^8 events per second. The required dynamic range is determined by the maximum energy that must be measured without saturation and by the precision that is required at low energy. Typically, the dynamic range will be of the order of 8 bits for the particle tracking chamber and 14 bits for the calorimeter subsystems. Furthermore, the nonlinerity should be less than 1% over the operating range for some subsystems. The high interaction rate, requires adequate time response capability of the readout electronics. In addition to wide dynamic range and fast time response, the readout electronics should have a power dissipation as low as possible, typically less than 50 mW/channel. This low power dissipation would make it possible to locate the entire front end readout electronics directly on or very close to the detector elements thus preserving the hermeticity of the calorimetry.

In general, for the SSC Detector System, if the signal transmission from the detector elements to the remote signal processing electronics and data acquisition subsystems were done by conventional cables it would present an extremely difficult packaging problem and would require an enormous space allocation. This in turn would compromise performance because of electromagnetic interference and signal loss as well as reduced maintainability and reliability of the complete detection system. To reduce the cable subsystem size and to simplify the overall system architecture it will be necessary to multiplex readout electronics. Using digital and analog fiber optics transmission links and radiation-hardened components will lead to further simplification and size reduction.[7-12] Furthermore, because the cost of electronics subsystems for SSC detectors will be a major part of the cost of the total detector system, the application of multiplexing digital and analog fiber optics transmission links in front end electronics, triggering and data acquisition will significantly reduce this cost.

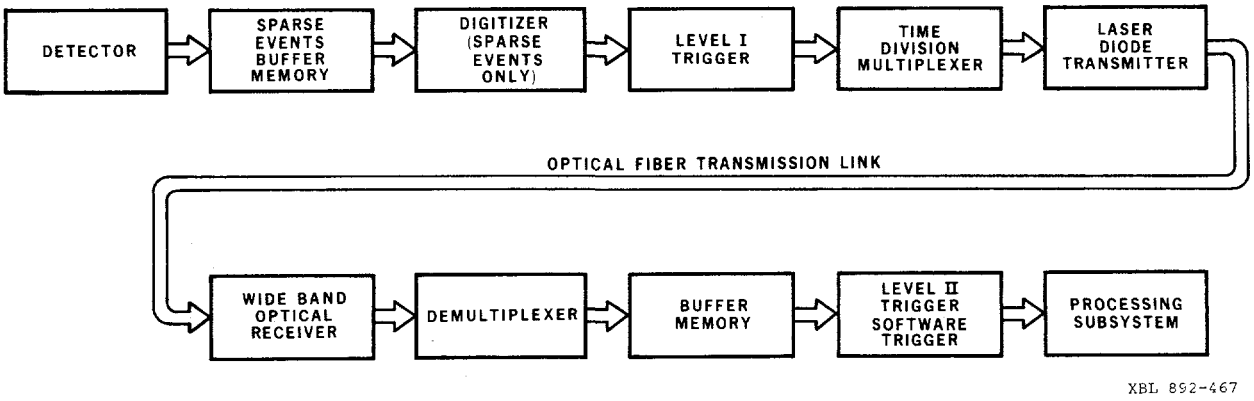


Fig. 2. Simplified block diagram for optical fiber transmission link in SSC detector system where signals are multiplexed after the Level I Trigger.

An example of the use of fiber optic transmission links is given in Fig. 2 where signals are multiplexed onto the fiber optic waveguide after the Level I Trigger.[11] Only sparse data are digitized from all signals emanating from the detector and stored in the buffer memory. After the Level I Trigger has selected the potential events of interest, the data are passed on to the time division multiplexer. The time division multiplexer converts the data from parallel bits to serial bit form for transmission over fiber optic waveguide. After reception of the optical signals, the demultiplexer converts the serial signals back to parallel bits for storage in another buffer memory. The data are then passed through the Level II Trigger, the Software Trigger and on to the Processing Subsystem.

Similarly, another example of the use of fiber optics transmission links is given in Fig. 3 where detector signals are multiplexed before the Level I Trigger.

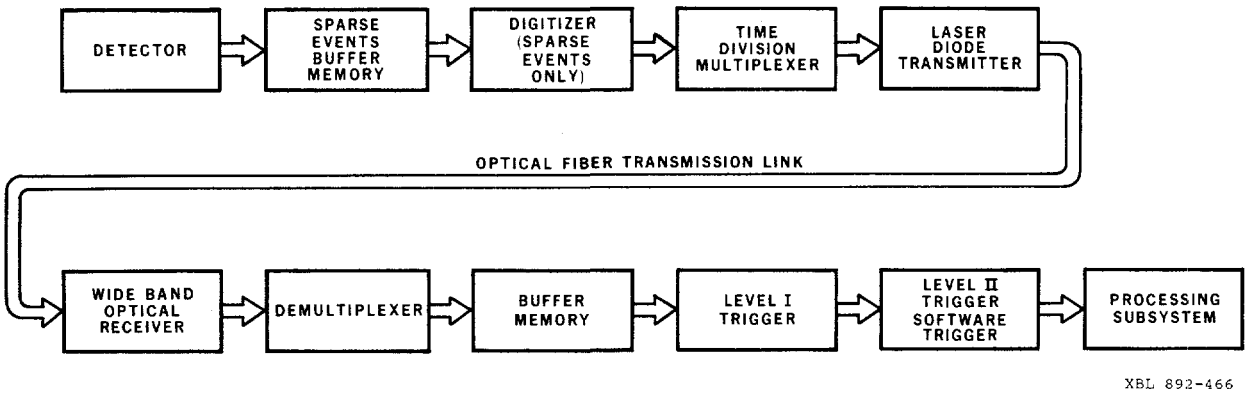


Fig. 3. Simplified block diagram for optical fiber transmission in SSC detector system where signals are multiplexed before the Level I Trigger.

Another application of fiber optic transmission links is with the use of highly segmented detector subsystems in SSC detectors which require a large number of signal channels even with innovative schemes for sparse data scan logic. For example, high resolution pixel devices might be used for vertex detector which would require many signal channels to be read out from deep within the SSC detector. Highly multiplexed signal channels onto just a few high-speed fiber optic systems could be employed here. Multiplexing would alleviate the problem of the large number of wires which would otherwise be required for parallel dataways operating at tens of Mbit/s speeds.

The hermeticity of SSC detectors will be significantly improved by the use of fiber optics because of their lower space requirements as indicated in the above examples. Furthermore, the immunity to noise pickup and the low mass of fiber optic cables are additional advantages.

Similarly, an application of optical fibers appears attractive for a number of other data transmission and communication tasks in various SSC accelerator systems.

For example data base for the collider beam monitoring subsystems and control system must accommodate approximately 62,000 monitoring and control points.[6] Each of these points has a number of words in the data base for its description and specification of properties. Monitoring and control points in the data base are divided among the various control subsystems. The control system of the collider consists of a host computer cluster, eight sector computers, two cluster computers, five injector subsystem computers, and approximately 400 distributed front-end processors. These computers and processors are connected together by local networks and a major long haul network of approximately 80 km in length. All the necessary communications from the central control facility to subsystems around the main accelerator ring will be accomplished by a ring information network which could profitably use optical fiber links instead of broad band coaxial cables. However, analog and digital transmission links and associated electronics components which will be used in the SSC detector and primary beam tunnel will be required to withstand exposure to the nuclear radiation background. Presently preliminary existing radiation background estimates for the dose rate and neutron fluence are 10^2-5×10^6 Gy/year and 10^{12}-10^{13}n/cm^2/year, respectively.[12] These values of background radiation are high enough to cause an increase of the transmission loss in optical fibers and measurable degradation of operating characteristics of optical transmitters, receivers and associated electronics.

The author has investigated the feasibility of designing and developing high speed digital and analog data transmission systems to meet demand of various SSC detector subsystems. Furthermore, a short review of radiation damage in optical fibers, optical transmitters and receivers as well as associated electronics components and subassemblies will be given.

TIME DIVISION MULTIPLEXER

A time division multiplexer scheme using parallel to serial data conversion is shown in Figure 4. This multiplexer scheme demonstrates the principle by which data can be prepared for transmission over optical fiber cables.[10] By appropriately controlling the Parallel in/Serial Out lines for the two shift registers (SRA and SRB), SRA is parallel loaded from the Digital Buffer Register A and then serially read out. In the meantime while SRA is being read out, SRB is parallel loaded with data from Digital Buffer Register B and awaits its turn to be read out. Then while SRB is being read out, SRA is again parallel loaded with new data. SRA now awaits its turn to be read out serially. This procedure is repeated until all of the data from the buffer registers are transmitted. Control signals A and B steer the serial data through their respective gates, and the data are combined in the OR gate for presentation to the optical transmitter. In this manner relatively slow digital operations handle the data to and from the digital buffer registers while extremely high serial data rates are handled by the optical fiber transmission system.

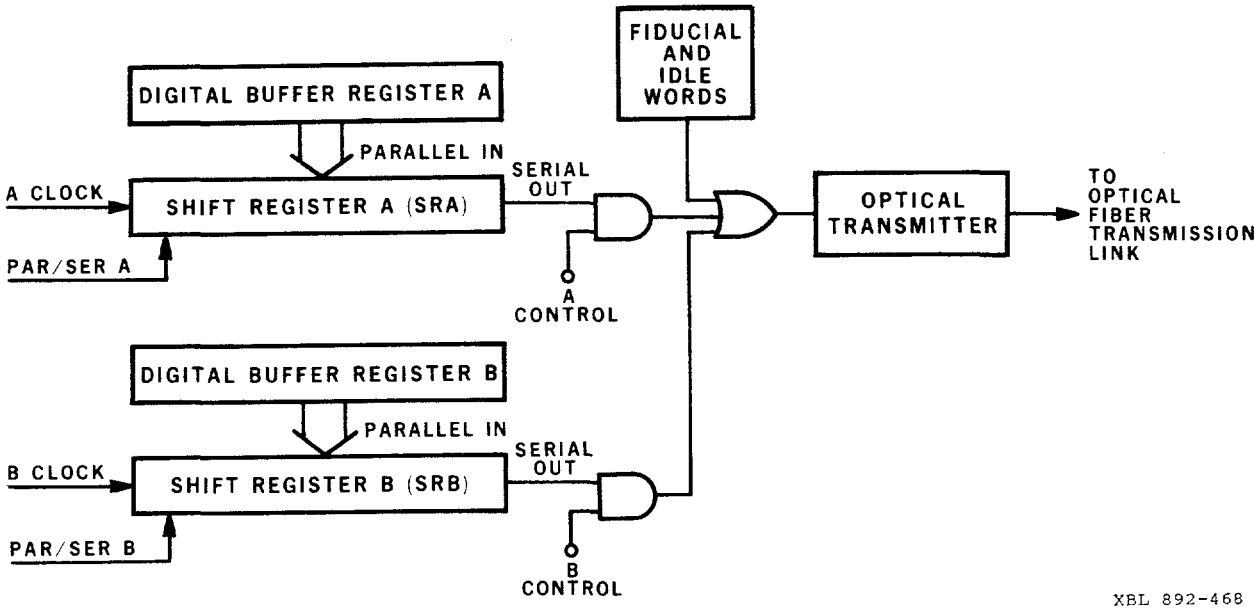


Fig. 4. Time division multiplexer.

In order to preserve the high data rate capabilities of an optical fiber system, light emitting diode or laser diode transmitters are coupled to single-mode optical fibers. Non-return to zero (NRZ) digital coding format for transmission of data is proposed for wide band data transmission system because bandwidth requirements are effectively one-half that of a return to zero (RZ) pulse code format. The NRZ pulse code format, because of its lower bandwidth requirements, will also contribute to lower bit error rates. Other pulse code formats such as bi-phase, amplitude modulation, frequency modulation and phase modulation are too complex or requiring more bandwidth than NRZ. In conjunction with NRZ pulse coding, 4 bit/5 bit or 5 bit/6 bit encoding/decoding could be used to further improve reliability.

The idle and fiducial words which are OR'ed with the serial data will be described. Their functions are helpful in the operation of the optical fiber system.

Idle words composed of some arbitrary alternating bit pattern are transmitted at the beginning of any new transmission signal. Moreover, idle words are transmitted whenever gaps occur in the transmission signal and loss of synchronization would result at the receiver. More detailed information about data synchronization will be given later in the description of the receiver system.

Fiducial words are transmitted to alert the receiver system that the words immediately following are real data. In addition, fiducial words are transmitted periodically to pull the receiver system back into step in the event that data have become garbled and unrecoverable. Parity bits and other error detecting schemes will also be employed to enhance reliability.

LIGHT SOURCES FOR OPTICAL DATA TRANSMISSION SYSTEMS

Light emitting diodes and semiconductor lasers are the most frequently employed as light sources in optical systems.[8-10] Light emitting diodes (LEDs) offer the advantages of simple fabrication and operation as well as low cost, high reliability and good linearity and small temperature dependence of the light output. Semiconductor index-guided injection laser diodes offer high output power level, efficiency and bit rate modulation capability as well as extremely narrow spectra and excellent mode stability of the emitted light.

Two basic LED structures are widely used for optical sources in the second generation transmission systems operating at wavelengths of about 1300 nm. These are the surface-emitting and the edge-emitting structures. Both structures use an n-type InP substrate over which four epitaxial layers are grown by liquid-phase or vapor phase epitaxy. The first is an n-InP buffer layer, followed by the light emitting InGaAsP region with the composition chosen to emit at approximately 1300 nm. The light emitting region is followed by a p-type InP cladding layer and a p-InGaAsP contact layer. For the surface emitting structure the light is emitted perpendicular to the grown layer. In the edge limiting structure the light is emitted along the phase of the epitaxially grown layer. The surface emitting LED and edge emitting LED is used as a light source for multimode and single mode fiber systems, respectively.

A surface emitting InGaAsP LED operating at 1300 nm wavelength, having a monolithically form lens for efficient optical coupling, can launch 50-200 μW of light into a multimode fiber and up to 50 μW into single mode fiber. The amount of injected optical power depends on the numerical aperture of the fiber and its core diameter. Although the total available power from an edge emitting LED is smaller than that of surface emitting LED, the higher coupling efficiency to optical fiber compensates for it. Furthermore, edge emitting LED's can be conveniently packaged in an array creating a number of individual addressable elements. At present, an array consisting of 12 LEDs with center-to-center spacing of 250 μm is available. The array can be aligned to a 12-optical fiber ribbon using a silicon block.

In digital data transmission conventionally designed light emitting diodes are mostly used for moderate speed applications, up to approximately 100 Mbit/s. For higher speeds LEDs having double heterostructures and high-doped active layers are combined with monolitic GaAs integrated circuit drivers. A bit rate of approximately 400 Mbit/s non-return-to-zero (NRZ) pulse transmission have been demonstrated.[13] The incident light power has been about 15 μW into 50 μm-core graded index fiber when a micro-ball lens and spherical ended fiber were used.

A peak electroluminescence wavelength of a surface emitting LED output spectrum is primarily determined by the bandgap composition of the active layer. The peak wavelength value increases with ambient temperature shifting by 0.6 nm/^{0}C for InGaAsP surface emitting diode. The spectral width of the same LED, emitting at 1300 nm, is approximately 100 nm, FWHM. As a result of self absorption along the length of the active layer the spectral width of an edge emitting LEDs is approximately 50 nm. The spectral width of LED should be as small as possible because the rms pulse broadening in the transmission link is directly proportional to the light source spectral linewidth. This results in a limitation on the bandwidth-length product which may be obtained using a particular optical source and fiber.

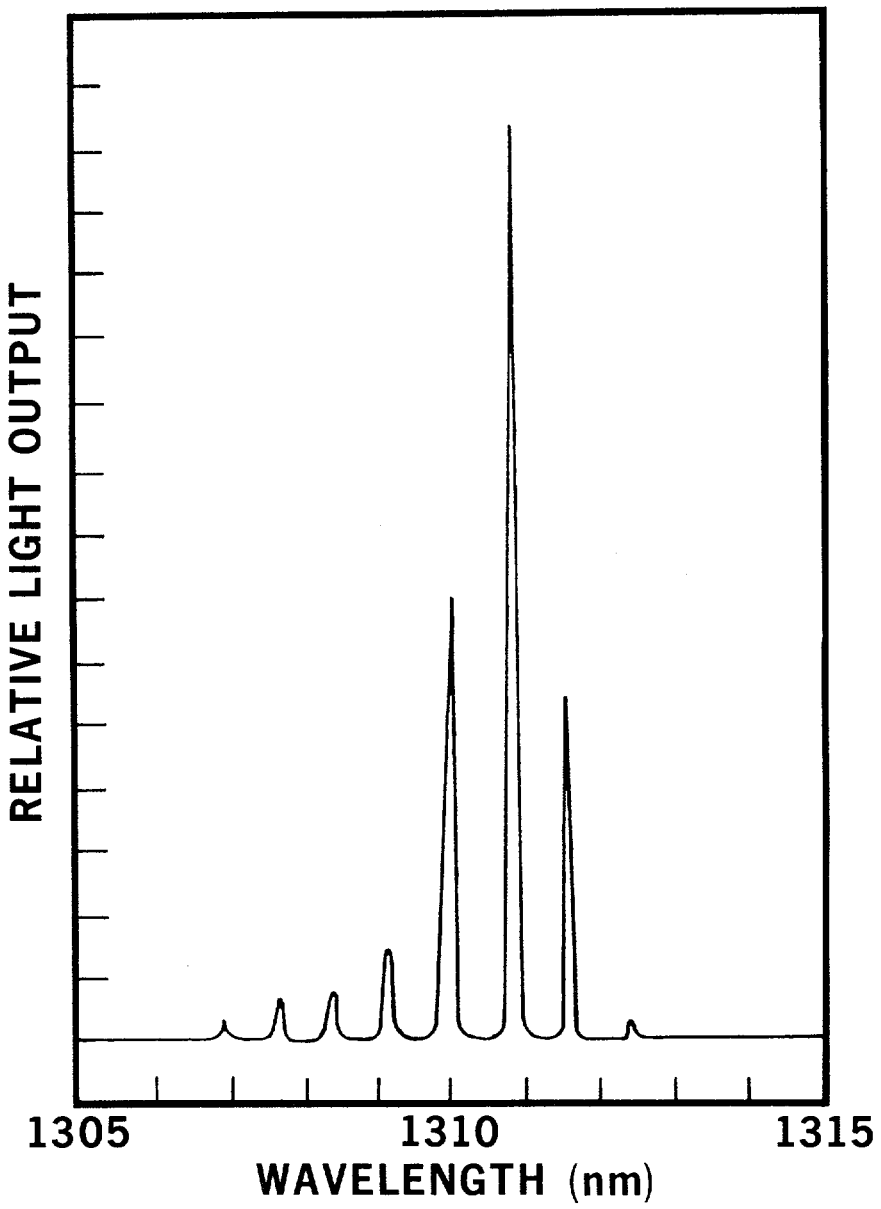


Fig. 5. Typical spectrum of the laser diode transmitter output.

Analog data transmission by fiber optics in SSC detector systems is highly desirable because it will result in significant reduction of electromagnetic interference and ground loop currents. It will also allow a multiplexing of analog data by means of specially developed hybrid electronics which can be mounted directly on detectors. This will lead to a substantial reduction of the cable plant. However, the operating conditions of the LED, in optical

transmitter should be optimized to reduce the device nonlinearity (harmonic and intermodulation products) and temperature dependance to an acceptable level. Accurate nonlinearity measurements performed on several LEDs showed that the output light intensity depends on the driving current in a nonlinear but predictable way. Therefore, the LED intrinsic nonlinearity will be reduced by the local nonlinearity compensation using predistortion.

Semiconductor lasers are the most frequently employed as light sources in optical broadband systems because of their compactness and high level of efficiency in comparison to light emitting diodes. The semiconductor compounds, indium gallium arsenide phosphide $In_{1-x}Ga_xAs_yP_{1-y}$ has emission wavelengths between 920 and 1620 nm, depending on its composition, making it particularly suitable for 1300 nm zero dispersion wavelength of an optical silica fiber.

Various types of semiconductor lasers, their structure and characteristics as well as pertinent optical processes, such as absorption, spontaneous emission and stimulated emission are extensively treated in the literature.[8,14] For the wide band data transmission system of particular interest are laser structures that employ a variation in the real-refractive index along the junction plane to form an optical waveguide. These index-guided injection InGaAsP lasers are almost free from the light output nonlinearities because of excellent mode stability. The index-guided laser with a thermoelectric cooler, temperature sensor unit power monitor and associated circuitry are packaged in a single module.[8] The integral thermoelectric cooler maintains the laser and power monitoring photodiode temperatures at 25°C over an ambient temperature range of -40 °C to + 65°C. Also, an internal InGaAsP PIN photodiode, mounted directly behind the laser diode functions as a power detector. The photodiode monitors the emission from the rear facet of the laser and controls the optical power level sent into the fiber.

Using an index-guided injection InGaAsP laser the optical transmitter has the following characteristics: emission wavelength = 1300 nm, maximum data rate capability = 1.7 Gbit/s, CW optical power = 2.5 mW (+4dBm), spectral width = 4 nm, rise and fall time smaller than 100 ps, RMS spectral width smaller than 3 nm, laser threshold current = 20 mA, laser forward current for maximum optical output = 40 mA, laser forward voltage at output power of 1 mW = 1.3 V, and life greater than 10^5 hours. A typical optical spectrum of the output light is shown in Fig. 5.

OPTICAL SOURCES IN A RADIATION ENVIRONMENT

The physical mechanism which causes radiation-induced degradation of the light output from LEDs is that nonradiative recombination centers are introduced which compete with radiative centers for excess carriers.[11,15] This results in a decrease in minority carrier lifetimes. These various centers, such as unintentionally added impurities, dislocations, growth-induced lattice defects and radiation induced lattice defects can act as sites for non-radiative recombination events producing heat rather than light.

Irridiation test performed on InGaAsP LEDs, with γ-rays operating at 1300 nm, showed that no significant degradation of parameters can be observed[16] with total doses of less than 10^5 Gy. The light output power decreased by 5% from its initial value upon an irradiation dose of 10^6 Gy. It was also estimated, that the light output power decreases to 50% of its initial value for the total dose of 2×10^7 Gy.

The normalized light output characteristics as a function of the neutron fluence for various LEDs under constant current operating conditions are shown in Fig. 6. The data are shown for the following devices: Plessey InGaAsP-High Radiance LED, Laser Diodes Laboratories GaAlAs IRE-160-High Radiance LED, Texas Instruments GaAlAs LED, Radio Corporation of America InGaAs LED, Hewlett Packard GaAs 4120 LED, and Texas Instrument GaAs: Si, Si LED. The high radiance (HR) devices, show the smallest sensitivity to radiation. These LEDs have very small source and junction areas, so that the injected minority carrier current density is large even at moderate current levels. Consequently, it can be expected that the radiative recombination rate is enhanced at typical operating conditions. These devices can provide sufficient light output for many applications even after neutron fluences in excess of 2×10^{14} n/cm^2.

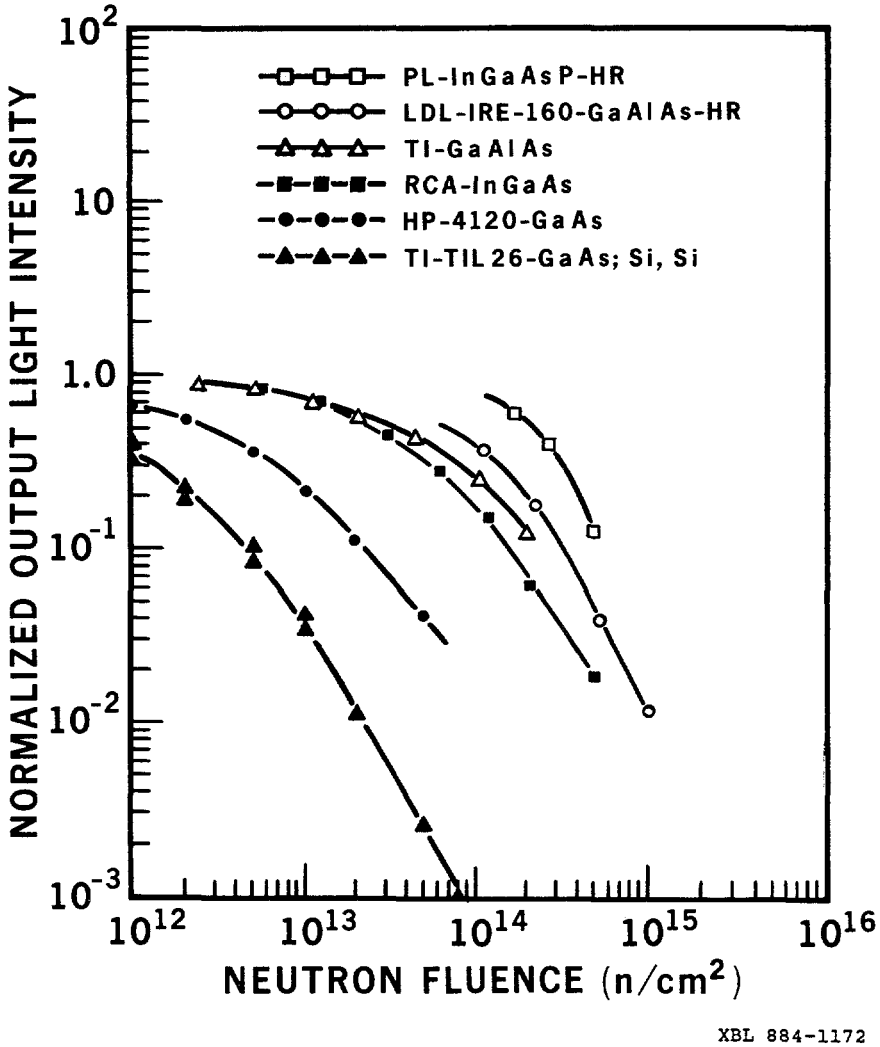


Fig. 6. Normalized light output as a function of the neutron fluence for various LEDs under constant current operating conditions.

Devices with GaAs:Si,Si junction are significantly more sensitive on radiation. The minority carrier lifetime in these LEDs is very long, typically 200-400 ns. Consequently, the output light decreases for approximately 10^3 when compared with typical GaAlAs LEDs at neutron fluence of 10^{14} n/cm^2, Fig. 6. The GaAs:Si,Si devices were originally developed for high light output.

More recent studies of the effect of neutron irradiation on LEDs fabricated from strained-layer superlattice structures in the GaAs/GaAsP configuration show that there is no light output degradation until a fluence of approximately 3×10^{14}n/cm^2 is exceeded.[15]

Radiation induced degradation of the light output from the semiconductor laser diode is caused by a reduction of minority carrier lifetime resulting from displacement damage. Total light output of a GaAs laser diode as a function of current density with neutron fluence as a parameter is shown in Fig. 7 In the subthreshold region 1 of the characteristics, the laser

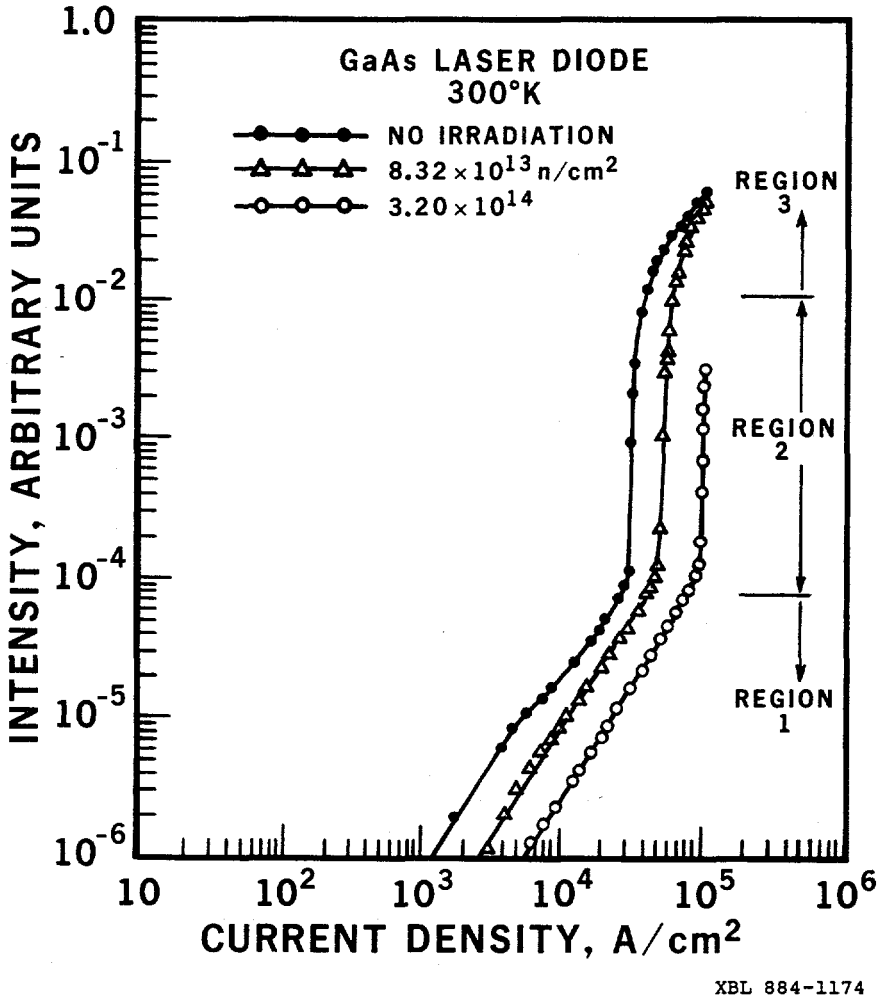


Fig. 7. Light output as a function of current density with neutron fluence as parameter for GaAs laser diode.

behaves like a light emitting diode. The light output, at constant current, decreases with neutron fluence at about the same rate as for LEDs. At the beginning of lasing action (region 2) the neutron irradiation causes a stronger decrease in light output. Finally, when the device is deeply in the lasing action (region 3), irradiation does not have a significant effect until the increase in the threshold current density prevents the laser from reaching region 3. Minority carrier lifetimes are of order 1 to 10 ps in GaAs junctions under intense stimulated emission. The lifetimes are of order 1-10 ns in sub-threshold region. Therefore, a much larger concentration of radiation induced defects is required to influence the radiative recombination rate in lasing region. Consequently, for a radiation environment a semiconductor laser diode should be selected with a low threshold current and a very high maximum operating current. Recently developed double heterostructure GaAs laser diodes are still capable of lasing action after a neutron fluence in excess of $2\times10^{14} n/cm^2$.

TRANSMISSION CHARACTERISTICS OF OPTICAL FIBERS

The basic transmission mechanisms of the various types of optical fiber waveguide have been discussed elsewhere.[8,17] Basically, the most important transmission characteristics of optical fiber are attenuation and bandwidth.

The attenuation or transmission loss, of optical fibers is influenced by the material composition, the fiber preparation technique and the waveguide structure. In particular, the attenuation is determined by material optical absorption, material scattering, macrobending and microbending losses, mode coupling radiation losses and losses due to leaky modes. Furthermore, there are also losses at the input and output connectors. In Fig. 8 the typical attenuation spectrum is shown for some state-of-the-art single-mode optical fibers.[8,17] The attenuation characteristics are shown for three major fiber preparation processes: (a) inside vapor deposition, (b) vapor-phase axial deposition and (c) outside vapor deposition.

In practical optical fibers, prepared by above mentioned processes, a major cause of signal attenuation below a wavelength of about 1500 nm is Rayleigh scattering resulting from excitation of small irregularities (on the order of $\lambda/10$), in the medium by the propagating electromagnetic wave. It is strongly wavelength dependent. The attenuation coefficient per unit length L_S can be expressed as $L_S = A/\lambda^4$, where A is a constant which depends on the refractive index of the medium, the average photoelastic coefficient, the isothermal compressibility and the temperature at which the glass can reach a state of thermal equilibrium and is closely related to the annealing temperature. This λ^{-4} dependence results in considerable attenuation at shorter wavelengths.

In addition to the Rayleigh scattering, attenuation in optical fiber can be caused by Mie scattering from large defects and by Brillouin scattering from the thermally driven density fluctuations that are present in fibers at room temperature. Furthermore, at high power levels, nonlinear loss mechanism such as stimulated Raman and Brillouin scattering can further increase the attenuation and limit the dynamic range of the fiber. Another major extrinsic loss mechanism in fibers is caused by absorption of light in negative hydroxyl ions

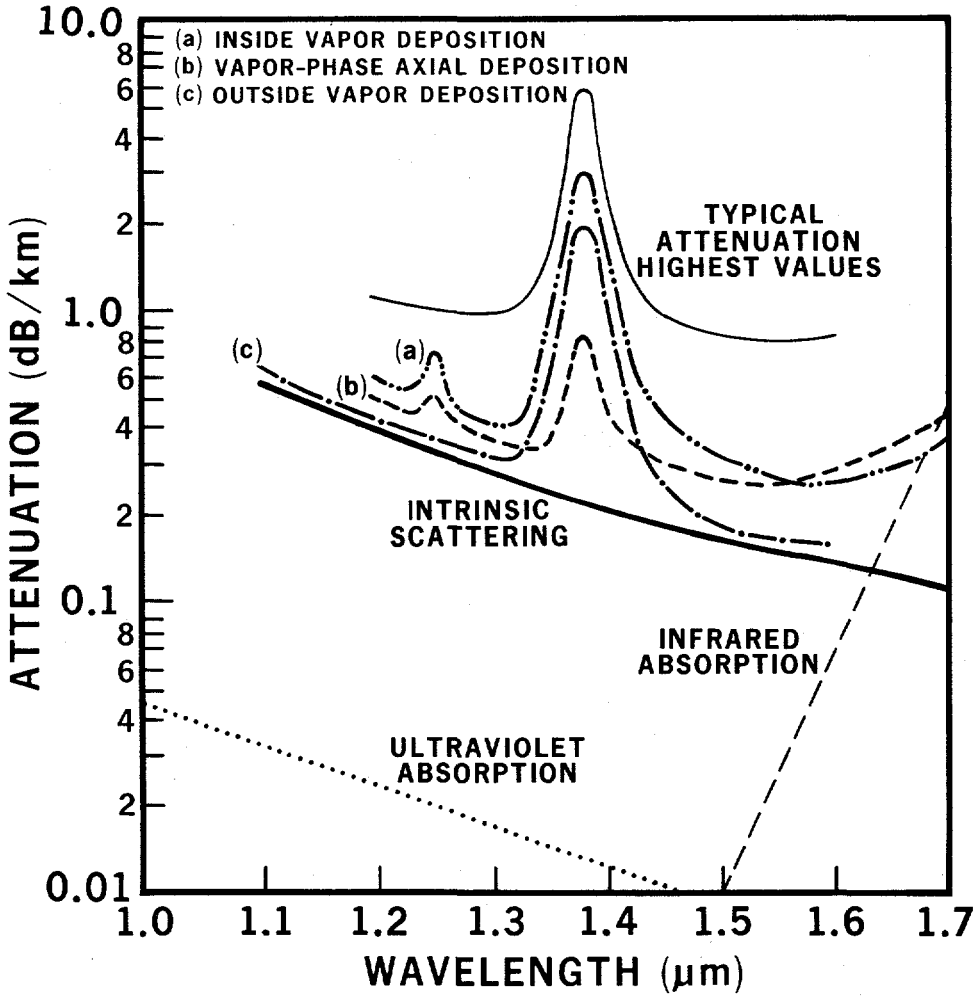


Fig. 8. Attenuation as a function of transmission wavelength for single mode optical fibers.

(OH) trapped during fiber processing. This produces sharp absorption peaks at 1250 nm and 1390 nm and smaller peaks at 750 nm and 970 nm. Since these peaks are present to some degrees in even the highest quality fiber, transmission at these wavelength is usually avoided. The ultraviolet absorption loss is also shown in Fig. 8 . This loss is due to the stimulation of electron transition within the silica glass by higher energy excitations. Its value was estimated by extrapolating the data from GeO_2-doped silica glass. Furthermore, in Fig. 8 the infrared absorption loss is shown. This loss was obtained by extrapolation from the loss characteristics of GeO_2-doped silica-core fiber and optical data for pure glass at long wavelengths where strong absorption bands are caused from the interaction of photons with molecular vibrations within the glass.

It appears from Fig. 8 that three major manufacturing processes are giving approximately the same attenuation levels. In general, the single-mode attenuation level is intrinsically slightly lower than that for multimode attenuation because of lower core dopant concentrations.[8] In the figure the typical attenuation highest level is also shown indicating tolerances of the fiber manufacturing process.

The information carrying capacity or bandwidth of a fiber is inversely related to its total dispersion. The total dispersion consists of mode, material and waveguide dispersions. Mode dispersion in multimode fibers is dependent only on the fiber, while the material and waveguide dispersions depend on the width of the wavelength spectrum from the light source.

Mode dispersion occurs in multimode fibers because different modes travel different effective distances through the fiber. This causes light pulse to spread out temporarily as it travels along the fiber.[8] For multimode graded index fiber this effect limits the bandwidth to approximately 1 GHz per kilometer of length, or 1 GHz $\times$ km.

Multimode dispersion in a data transmission system can be completely avoided by using a single-mode fiber which carries light only in a single-wavelength mode. In such a fiber no dispersion between modes can exist and very large bandwidths are possible. However, material and waveguide dispersions limit the bandwidth of a single-mode fiber. Waveguide dispersion occurs because light in a single-mode fiber is not confined completely to the core. About 20% of the light travels in the cladding adjacent to a step-index core. The refractive index of the cladding is lower than that of core and so the light travels somewhat faster than in core. The wavelength dispersion is wavelength-dependent, although the change in wavelength dispersion with wavelength is smaller than that of material dispersion. Furthermore, the material and waveguide dispersion can have different signs and thus completely cancel each other out. In conventional germanium-doped silica fibers this "zero-dispersion" wavelength is near 1300 nm.

In Fig. 9 the material dispersion of silica, curve (a), and the waveguide dispersion of step-index fibers having core diameters of 3.5 and 11 μm are presented as a function of the wavelength, curves (b) and (c). The combined effect of material and waveguide dispersion is also shown, curves (d) and (e). Reducing the core diameter has two effects. The zero dispersion point is shifted to longer wavelengths and the gradient of the dispersion in the vicinity of zero gradient point is reduced.

In addition to the dispersion components considered above for single-mode fibers, there are other higher order effects which impose limitations on the maximum bandwidth. These together with secondary effects, such as birefringence which arises from ellipticity or mechanical stress in the fiber core, give a fundamental lower limit to pulse spreading of between 3.5 and 5 ps/km $\times$ nm. The theoretical bit rate for an optical channel is estimated to be between 40 and 57 Gbit/s if an injection laser with rms spectral linewidth of 1 nm is used as a pulsed light source and there is a small amount of intersymbol interference in the channel.[8]

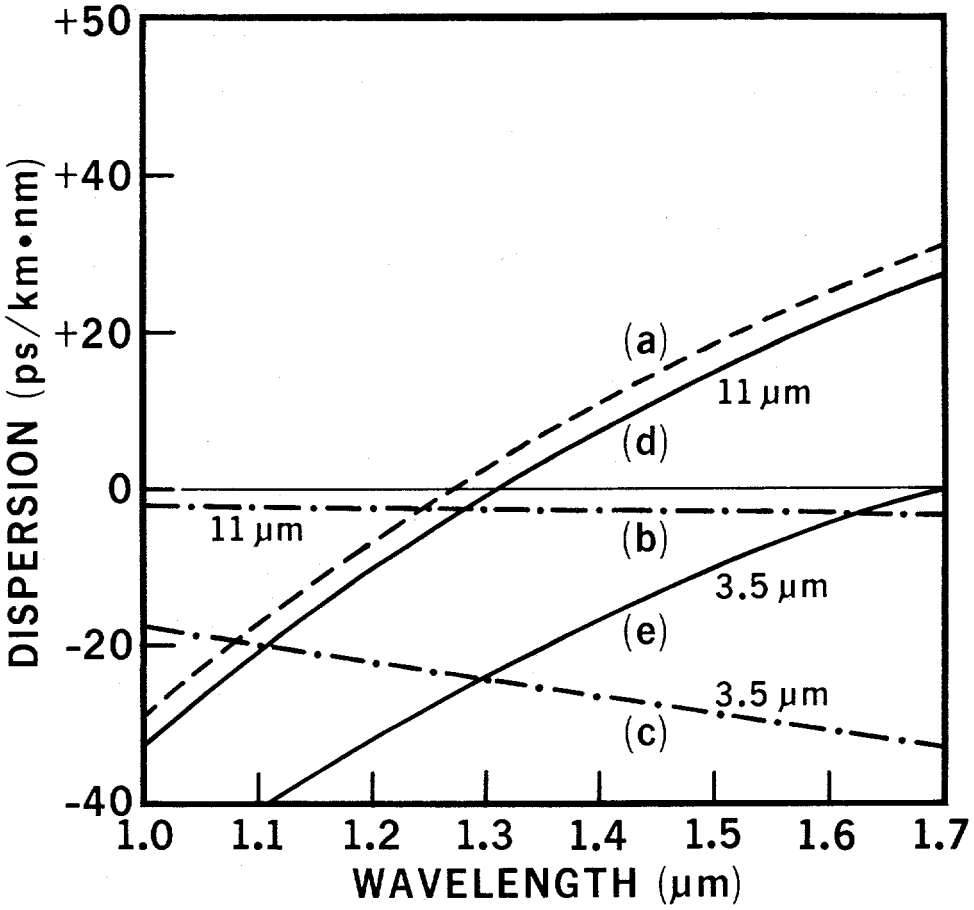


Fig. 9. Dispersion curves of Step-index single mode fibers.

RADIATION-INDUCED ATTENUATION OF OPTICAL FIBERS

The optical properties of fiber waveguide are degraded by exposure to nuclear radiation, primarily through the generation of color centers in the fiber core. Color centers are formed by radiolytic electrons and holes which are trapped on defects that either exist in fiber prior to irradiation or are created by the exposure. These centers cause the optical attenuation which can be significantly greater than the intrinsic fiber loss. In addition to the radiation-induced absorption light is generated in fibers during pulsed irradiation by photoexcitation of the color centers or by the Cerenkov process.

Radiation-induced attenuation consists of a permanent and metastable components. The permanent component lasts for a long period of time after initial exposure. The metastable component consists of a transient part which decays by 10 db/km in less 1 s after pulse irradiation and a component which decays after 10 s after irradiation. The detailed behavior of the induced absorption depends on a number of factors such as the fiber parameters (fiber structure, core and cladding composition fabrication and dopands), radiation parameters (total dose, dose rate, time after irradiation, and energy, nature and history of the radiation), and system parameters (operational wavelength, light intensity and temperature).

The radiation-induced attenuation of fiber initially increases linearly with increasing dose under steady state irradiation as it is shown in Fig. 10. However, at higher doses, the loss characteristic shows saturation due to the recovery processes that occur simultaneously

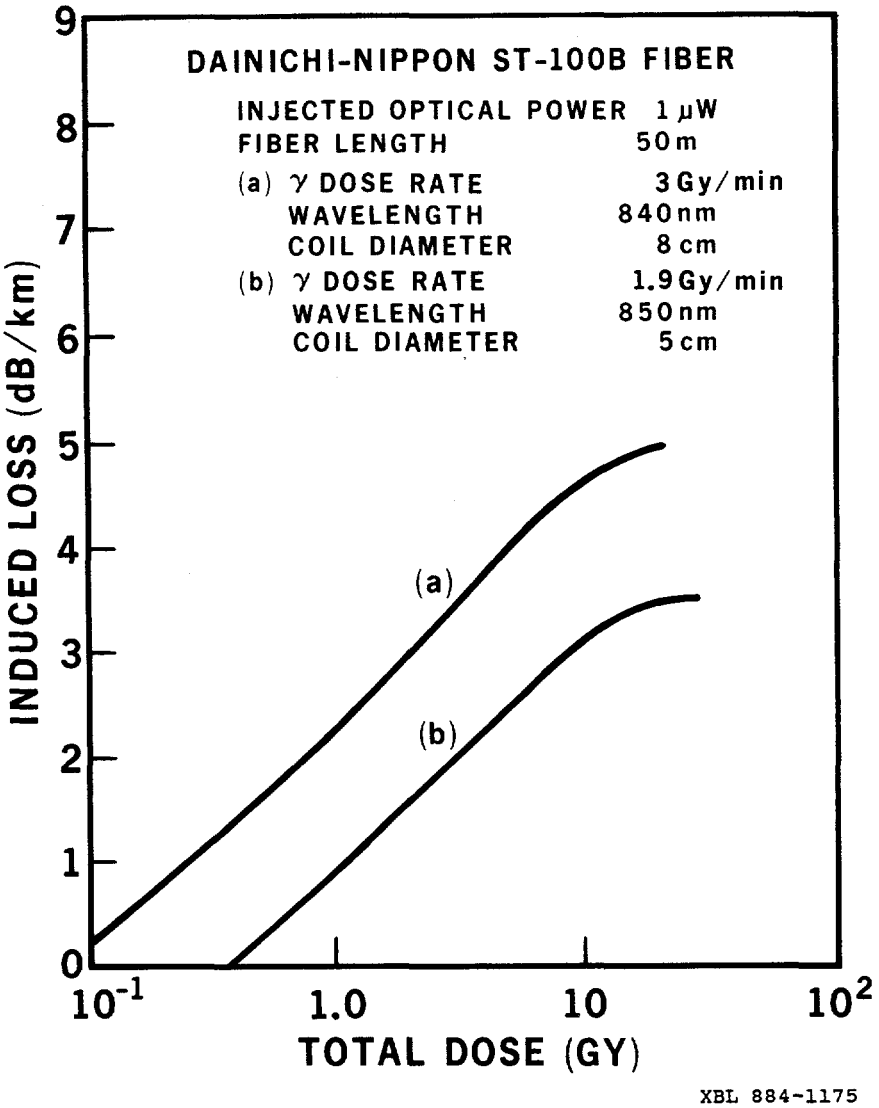


Fig. 10. Radiation-induced attenuation in Dainichi-Nippon St-100 B fiber for 1.9 and 3 Gy/min γ dose rate.

with the fiber darkening. The level of saturation depends upon fiber, radiation and system parameters. In multimode polymer clad silica fibers, having the high OH content core, such as Suprasil,[11] manufactured in the late 1970's, saturation levels were near 70 dB/km at a total dose of 10^2 Gy and operating wavelength of 820 nm. At doses higher than 10^2 Gy the induced loss decreased with increasing dose due to the radiation and photobleaching of color centers causing the absorption loss. Also, at doses larger than 10^4 Gy the fiber loss increased drastically because of embrittlement of the polymer. In multimode pure silica core fibers with florine doped cladding manufactured in the middle 1980's saturation levels are near 5 dB/km at dose of 10^2 Gy.[11,18]

Radiation induced attenuation for multimode Dainichi-Nippon St-100B fiber is shown in Fig. 10 using a LED injected optical power of 1 μW and a fiber length of 50 m. This fiber has a SiO_2 100 μm-diameter core, with a fluorine/boron doped SiO_2 140 μm-diameter cladding. The OH content and intrinsic attenuation is 5-10 ppm and 6.7 dB/km, respectively. The wavelengths of the injected optical signals were 840 and 850 nm. The fiber showed a radiation induced attenuation of 4.6 ± 0.27 dB/km at 30 Gy total dose with a γ-rays dose rate of 3 Gy/min.[11]

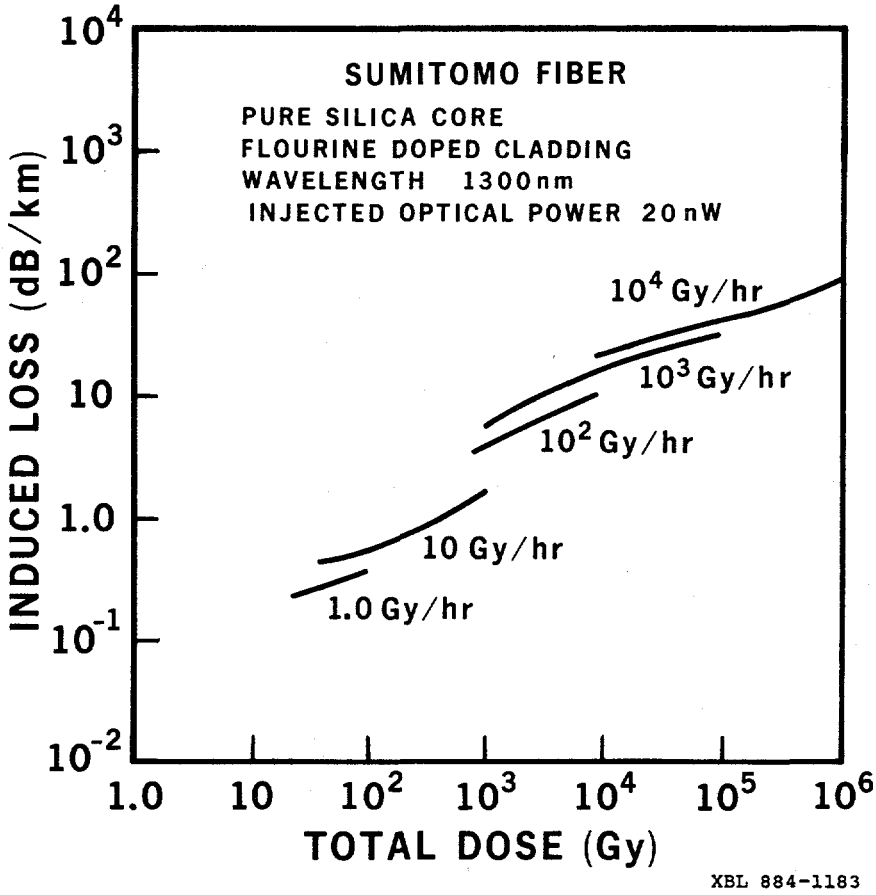


Fig. 11. Radiation-induced attenuation as a function of total dose with dose rate as parameter for single mode Sumitomo fiber having pure silica core.

Radiation-induced attenuation as a function of total dose with dose rate as parameters for a single mode Sumitomo fiber having a pure silica core and a fluorine doped cladding is shown in Fig. 11.[19] This fiber was fabricated by the vapor axial deposition process. The fiber

core diameter and cladding diameter is 8.5 μm and 125 μm, respectively. Depending upon the grade of the fiber, the preirradiation attenuation varies from 0.4 to 0.7 dB/km. Dispersion is smaller than 3.5 ps/nm/km for the range from 1285 to 1330 nm.

Measurements were performed using ^{60}C source and changing the dose rate from 1.0 to 10^3 Gy/hour. The injected optical power level was 20 nW at a wavelength of 1300 nm. It can be seen from Fig. 11 that the radiation-induced attenuation is 20 dB at total dose of 7×10^4 Gy and dose rate of 10^3 Gy/hour.

Similar measurements were performed on the fiber waveguide with the same geometry as given above but with germanium doped silica core and pure silica cladding. Under identical measuring conditions this fiber showed a radiation-induced attenuation of 20 dB at a total dose of 1.5×10^4 Gy and dose rate of 10^2 Gy/hour.

Furthermore, both fiber waveguides were irradiated with 14 MeV neutrons. The neutron induced attenuation of 20 dB was obtained after approximately 0.6 hour of irradiation for germanium doped core fiber. The attenuation increased up to 70 dB/km after 3.5 hours of irradiation (neutron fluence of 1.5×10^{14} n/cm^2). However for pure silica core fiber the induced attenuation reached approximately 10 db/km after the same irradiation time. Experimental results have clearly showed that a pure silica core fiber has significantly lower radiation sensitivity than a germanium doped silica core fiber for both γ-ray and neutron irradiation.

OPTICAL RECEIVER CONSIDERATIONS

The optical fiber receiver design requires careful attention for data transmission systems having a high-bit-rate capability. The receiver performance characteristics, such as sensitivity, dynamic range, bit rate transparency, bit pattern independency and acquisition time for functions of both the photodetector and receiver preamplifier configuration as well as operating bit rate.[20]

A simplified block diagram of a typical digital optical receiver is shown in Fig. 12. The input optical signal is detected by a photodetector. The photodetector is followed by a low-noise preamplifier, an automatic gain control amplifier and a shaping filter. The regenerator samples and detects the signal and regenerates the original data that is being transmitted. The shaping filter minimizes noise and intersymbol interference at the regenerator input. The clock pulses required for sampling are recovered by the timing extraction circuitry which is a typically phase-locked-loop.

After the data are reconstituted by the data recovery circuit they are applied to the shift register along with a clock signal. When the shift register is filled with appropriate data, the data are transferred to the buffer registers via a load command. The act of translating data from serial to parallel demultiplexes the data. Flags will be raised whenever data are

improperly transmitted or received. The data from the buffer registers are placed on the dataway for further signal processing. The most important components in determining receiver performance are the photodetector and low-noise preamplifier.

Photodetector Characteristics

The most useful photodetectors for optical fiber systems are the PIN and avalanche photodiode. The APD is more sensitive than the PIN diode because of its internal avalanche gain. For the wavelength region around 850 nm photodiodes are fabricated from germanium or several of the III-V components (InGaAsP, GaAlAsSb). At present long-wavelength transmission systems have achieved better performance with III-V compound APD's than

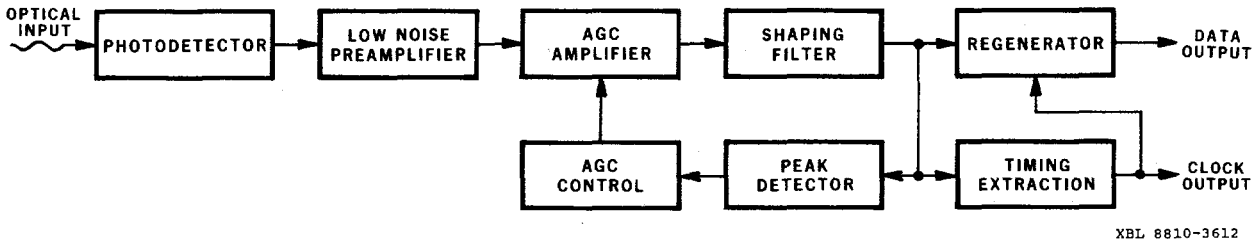


Fig. 12. Simplified block diagram of a typical digital optical receiver.

with Ge APD's. Ge APD's have relatively higher dark current, unfavorable ratio of ionization coefficients and low absorption coefficient compared with III-V compounds APD's at these wavelengths.[20,21] The APD structure which has achieved very good performance to date has separate absorption and multiplication regions (SAM-APD's).[22]

The ADP's of this type utilize InP in the multiplication region and a lattice-matched epitaxial layer of InGaAs in the absorbing region. The SAM APD's were developed to eliminate the tunneling component of the dark current in InGaAs APD's. In this structure the p-n junction, and thus the high field region, is located in InP where tunneling is insignificant and absorption occurs in an adjacent layer of InGaAs. To improve the frequency response of the SAM APD's single (or more) InGaAsP transition layer(s) is interposed between the InP multiplication region and the InGaAs absorption region. The transition layer(s) grades stepwise the bandgap energy between that of InP and InGaAs. The net effect of adding the transition layer(s), is to reduce the valence band discontinuity which results in the elimination of most of the charge accumulation at the heterojunction interface. Such structures are called SAGM-APD, where G denotes the presence of one or more grading layers.

The state-of-the art performance of photodetectors for long wavelengths (1000 - 1600 nm), such as InGaAs APD, InGaAs PIN and InP/InGaAsP/InGaAs SAGM-APD devices, have a quantum efficiency of approximately 80%, response time of 100 ps and a diode capacitance less than 1 pF.[20-22] The SAGM-APD device with optimized parameters exhibits excellent performance, such as dark current of 2-50 nA, ionization ratio of 0.35, and avalanche gain-bandwidth product of 70 GHz.

Digital Receiver Sensitivity

As receiver bandwidth is increased to accommodate higher transmission bit rates its sensitivity decreases mostly because the preamplifier noise increases rapidly with bit rate. The receiver sensitivity is defined as the minimum optical power level required at the receiver input so that it will operate reliably with a bit error rate less than a desired value. In this case it is given as the average optical power $\overline{P}$ required for a bit error rate, (BER), of 10^{-9}.

Bit error rate is defined as the ratio of bits incorrectly identified to the total number of bits transmitted. The receiver sensitivity is also often specified in terms of the average optical power $\eta\overline{P}$ which is detected by the photodetector, where η is the quantum efficiency of the photodetector. Until recently, PIN photodiodes fabricated from InGaAsP alloys grown epitaxially on high quality InP substrates were regarded, as most suitable for use in the low-noise optical receivers of long wavelength communication systems. However, recently, at high bit rates, reverse biased, avalanche photodiodes exhibit high-speed response in the absence of large dark currents.

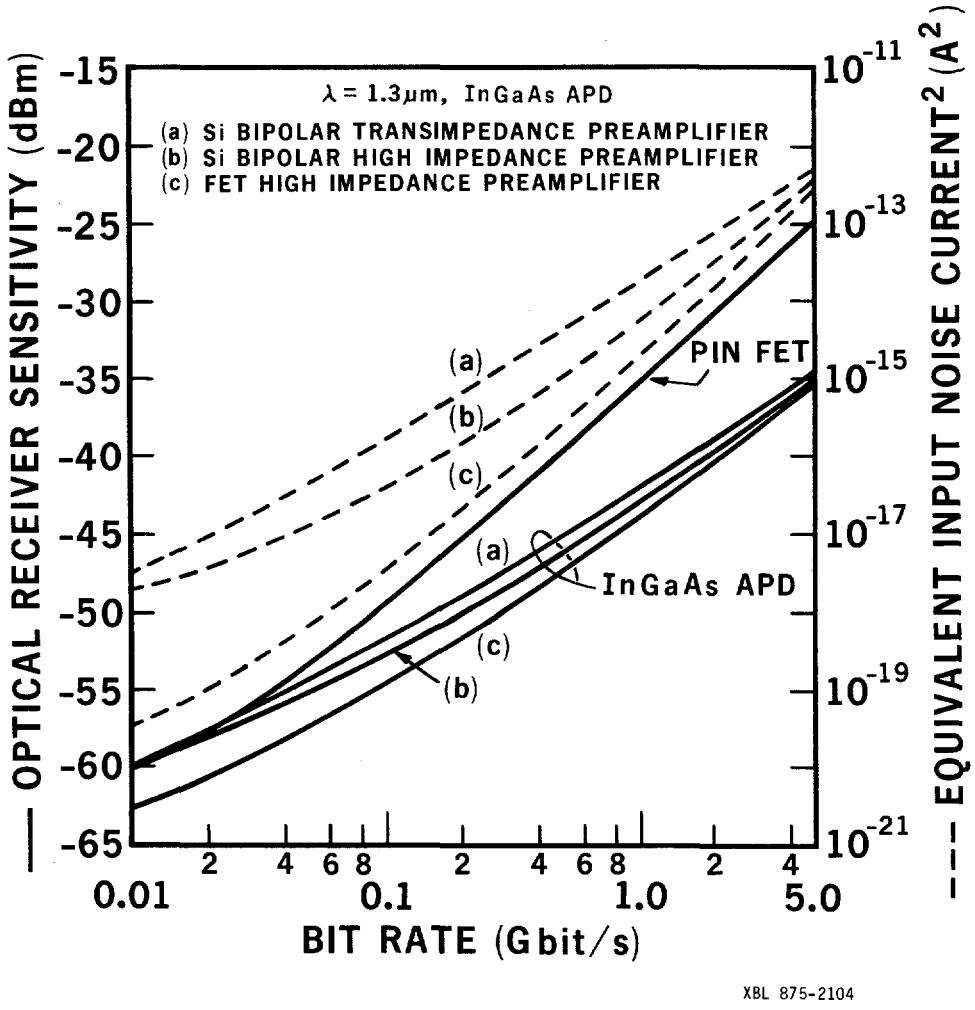


Fig. 13. Optical receiver sensitivity and equivalent input noise power versus operating bit rate.

Depending on their configuration, preamplifiers for optical receiver are classified into two types: high impedance and transimpedance designs. The high impedance preamplifiers offer the lowest noise level and hence the highest detection sensitivity. However, the frequency response is limited by the RC time constant at the input making necessary an equalizer following the preamplifier to extend the receiver bandwidth. Furthermore, this design has a limited dynamic range due to the high-input load resistance. The transimpedance amplifiers have a large dynamic range and bandwidth due to their negative feedback. However, because of the thermal noise of the feedback resistor the preamplifier noise level is higher and sensitivity is lower than that of a high-impedance design.

Because a detailed analysis of the optical receiver sensitivity have been discussed elsewhere,[17,20] only results (Fig. 13) will be given here.

It can be seen from this figure that the FET high impedance preamplifier provides the lowest noise level. In general, the advantage of the FET high impedance design over the bipolar transistor high impedance and transimpedance designs is reduced as the operating bit rate increases. As a matter of fact the FET noise power increases at a faster rate than the bipolar transistor noise power with an increase of the operating bit rate. Thus above 5 Gbit/s the bipolar transistor application can result in a smaller noise power level. The exact crossover point is dependent on particular characteristics of the photodiode and receiver preamplifier components.

Also included in Fig 13 are calculated values of the optical receiver sensitivity for three preamplifier designs using InGaAs avalanche photodiode and PIN photodiode-FET preamplifier. It can be seen from the figure that if PIN photodiode is used, a low-noise FET preamplifier is necessary to achieve high sensitivity. Furthermore, it can be seen that InGaAs APD's offer receiver sensitivity which is approximately 5-10 db better than PIN FET receivers at high operating bit rates.

Germanium avalanche photodiodes can be also used in receiver designs. The sensitivity of a Ge APD receiver is more dependent on temperature because of dominant noise effect of their dark current. Furthermore, Ge APD receiver sensitivity is 5-10 dB lower than that of a receiver of the same design using InGaAs APD.

The dependence of the receiver sensitivity on the preamplifier noise current is also given in Fig. 13. For PIN photodiodes the receiver sensitivity is proportional to $\langle i_n \rangle^{1/2}$. For InGaAs APD with negligible dark current the receiver sensitivity is approximately proportional to $\langle i_n \rangle^{1/4}$. Also from the same expressions the optimum avalanche gain can be calculated as a function of bit rate for various photodetector and preamplifier designs. In general, the optimum avalanche gain is such that the photodetector shot noise is comparable to receiver preamplifier noise. Thus the optimum avalanche gain is higher for a photodetector having lower dark current and lower excess noise. Similarly, optimum gain is lower for a preamplifier having lower noise level. Typically, the optimum avalanche gain values are from 10 to 35 for InGaAs APD. For the SAGM APD's the optimum gain is approximately 12 at a bit rate of 5 Gbit/s.

At very high bit rates, the limiting gain-bandwidth product, f_{GB}, of an avalanche photodiode can prevent operation at the optimum gain causing a decrease of the receiver sensitivity. Calculations have been made in Ref. 20 for a 1300 nm APD with f_{GB} values of 20 and 50 GHz assuming negligible intersymbol interference and a preamplifier using a

GaAs FET having f_T = 20 GHz, C_d = 0.5 pF and I_{GATE} = 100 nA. Calculations have shown that at 5 Gbit/s data rate the sensitivity decrease is 1 and 4 dB for photodiode with f_{GB} = 50 GHz and 20 GHz, respectively.

The dark current of a photodetector degrades the receiver sensitivity, as shown in Fig. 13, depending upon the operating bit rate. Our studies have shown that for a PIN FET preamplifier, having a photodetector dark current of 10 nA, the receiver sensitivity will be reduced by 0.5 and 1.0 dB from values given in Fig. 13 for bit rates of 10 Mbit/s and 100 Mbit/s, respectively. For an InGaAs ADP's with dark current of 6 nA the receiver sensitivity will be reduced by 0.5 dB at 1 Gbit/s.

Very high speed avalanche photodiodes and high-sensitivity optical receivers will have considerable impact on future Gbit/s SSC detector data transmission systems. Results of experimental receiver sensitivities as reported in literature for bit rates up to 8 Gbit/s are summarized in Ref. 20. In general, the receiver sensitivities are still approximately 20 dB above the quantum limit due to the lack of low noise long-wavelength APD's and imperfections of the receiver electronic circuitry. The highest experimental sensitivities have been obtained with InGaAs APD's and high impedance GaAs FET preamplifiers. These best results are still lower than theoretical sensitivities, calculated for an APD receivers by 3 dB at 1 Gbit/s and 8 dB at 8 Gbit/s. The larger difference at the higher bit rate reflects both the limited APD gain-bandwidth product and greater difficulties achieving ideal operation of the electronic circuitry at very high bit rates. Electronic components and subassemblies for bit rates near 1 Gbit/s, such as preamplifiers, decision circuits, multiplexers, and demultiplexers have very good performance. However the same components at 4 and 8 Gbit/s are still in the experimental stage and they will require considerable development before the performance obtained at lower data rates can be achieved. Limited capabilities of 8 Gbit/s electronics components and circuitry causes inter-symbol interference and imperfect receiver equilization.

The 420 Mbit/s and 1 Gbit/s receivers have demonstrated sensitivity of -46.2 dBm and -42.1 dB, respectively.[23] These receivers employed APDs with 18 GHz gain-bandwidth product, and quantum efficiencies η = 95%. Results obtained are within 4-6 dB of the best results obtained for receiver sensitivity with receivers employing Si APD's.

The measured sensitivity of the 2 and 4 Gbit/s receivers were -36.6 and -31.2 dBm, respectively.[23] These receivers used SAGM-APD with a gain-bandwidth product of 18 GHz, and a quantum efficiency η = 71%. The preamplifier employed high impedance GaAs FET.

The 8 Gbit/s receiver with -25.8 dBm sensitivity employed a SAGM-APD with a gain bandwidth product of 60 GHz, quantum efficiency η = 62% and a high impedance GaAs preamplifier.

OPTICAL RECEIVER IN RADIATION ENVIRONMENT

High sensitivity optical receivers will be required for use in some SSC networks particularly those in high data rate long haul link. In radiation environment in which the optical data transmission systems will operate, the photodetector of the receiver can be the limiting component. To achieve the required high performance the best possible detector and receiver configuration should be selected.

The same physical process that make photodetector sensitive for optical radiation also make most detectors sensitive to ionizing radiation. However, the ionizing radiation generates electron-hole pairs uniformly throughout the semiconductor material of the photodiode while optical radiation generates carriers only in the active region of the photodiode pn junction. Consequently low ionizing radiation sensitivity can be achieved by (1) reducing the volume of the active region and at the same time keeping the responsivity of the device to optical signals high, and (2) reducing the volume of the optically nonactive regions of the photodiode. Also, it is beneficial to reduce the collection at the photodiode junction of the ionization radiation-induced carriers from optically non-active regions. The first task can be accomplished by fabricating the photodiode from III-V semiconductors, such as GaAs which has a large absorption coefficient at the wavelength of the optical radiation. In this case the photodiode active region is very thin, typically several tenths of a micron, offering a very small volume to ionizing radiation. The second task can be accomplished by a heterostructure configuration of the photodiode.[15] In such configuration additional barrier layers are introduced and the geometry of the active region is precisely defined preventing collection of carriers from optically non-active regions of the device.

Various radiation studies were performed on these radiation hardened devices, measuring the increase of leakage current caused by ionization and neutron irradiation.[15] For comparison purposes a figure of merit was introduced, M_{PD}, defined as the ratio of the photodiode signal current per unit of incident optical flux to the ionization induced current per unit of dose rate. The figure of merit M_{PD} = 40-70 $\times 10^{-10}$ Gy/optical photon for a GaAs photodiode in a heterostructure configuration. For comparison M_{PD} = 0.5-2.0$\times 10^{-10}$ Gy/optical photon for a typical Si photodiode. This and other data have revealed that double heterostructure AlGaAs/GaAs devices are far superior to Si radiation hardened photodiodes. GaAs devices were able to operate up to 10^6 Gy/s, a level several orders of magnitude above the capability of Si PIN photodiodes.

It can be concluded from the measurements of neutron irradiation effects on photodiodes[15] that the device leakage current increases by a factor of 10 in the AlGaAs/GaAs photodiode and factor of 10^3 in the Si devices after exposure to a neutron fluence of 7×10^{14}n/cm^2. At this neutron fluence there is no optical responsivity degradation of AlGaAs/GaAs photodiode while Si device responsivity decreased to 60% of its preirradiation level.

Similarly, in InGaAs photodiodes, intended for data transmission links operating at wavelengths of 1300 nm, no degradation of optical and electrical characteristics were observed up to 10^6 Gy dose, with an exception of some increase of leakage current. The leakage current increases up to a factor of 6 from the pre-irradiation value when the total radiation dose is 10^6 Gy.

ELECTRONIC CIRCUITS IN RADIATION ENVIRONMENT

Associated electronic circuits and subassemblies of digital and analog data transmission systems will suffer a measurable degradation of their operating characteristics when exposed to the SSC nuclear radiation background. Various damaging mechanisms, such as ionization effects and atom-displacement phenomenon, are responsible for the device degradation. In general, it is assumed that the ionization effects and displacement phenomenon are independent processes, so that damage characterization can be done separately.

Nearly all important device parameters, such as current gain, transconductance, cut-off frequency, speed, breakdown voltage, noise figure, power output and resistance are degraded by the respective irradiation to some degree. The estimated ranges of total ionizing dose and neutron damage susceptibility of various silicon and gallium arsenide devices are shown in Fig. 14 and 15, respectively.

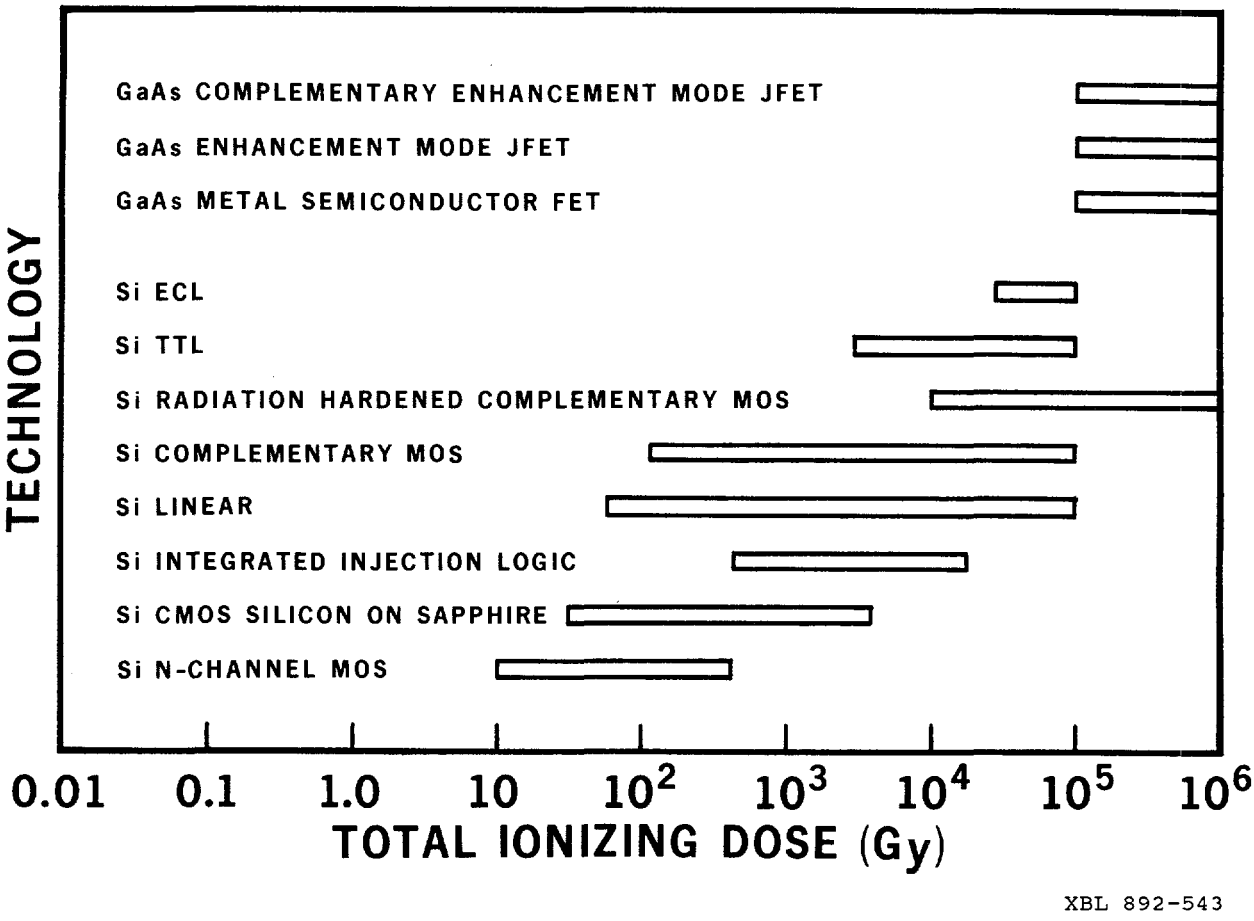


Fig. 14. Estimated ranges of total dose damage susceptibility for silicon and gallium arsenide devices.

The long term ionization damage for semiconductor devices depends strongly on the electrical bias condition during exposure and after exposure, measurement time after exposure and the ionization radiation energy spectrum at the device. Furthermore, the damage susceptibility is extremely sensitive to processing temperature and chemistry. Very small processing changes can introduce an order of magnitude change in total dose damage susceptibility. Also the damage level depends upon the dose rate at which the total dose is delivered. In ionization radiation environment a variety of technologies provide the total dose damage susceptibility of 10^4-10^6 Gy. These include radiation hardened Si CMOS, bipolar and GaAs FET technologies.

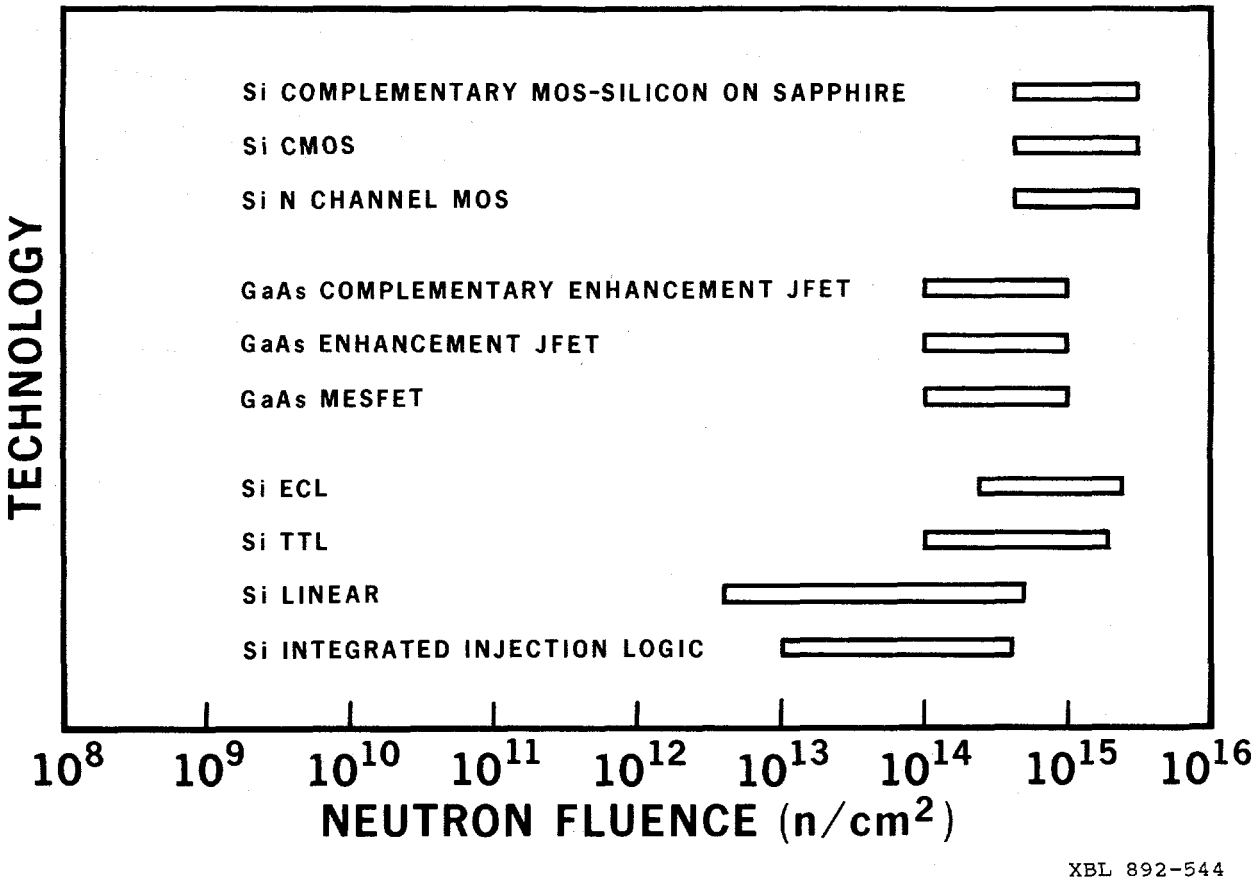


Fig. 15. Estimated ranges of neutron damage susceptibility for silicon and gallium arsenide devices.

In neutron environments FET technologies have 1 to 3 orders of magnitude smaller damage susceptibility than bipolar technologies. Also, compound semiconductors can additionally have yet another order of magnitude smaller susceptibility on neutron irradiation. Therefore, Si, C MOSFET, JFET and GaAs FET devices are suitable for applications where a high neutron irradiation levels are expected.

A considerable effort in being made in industry to develop radiation hardened semiconductor electronic components and circuits. The hardening techniques include selection of technology, special process procedures, device layout and circuit design, circuit computer simulation and selection of optimum operating conditions to achieve the desired value of radiation tolerance.[27]

CONCLUSIONS

Using available digital IC's, electro-optical and opto-electrical transducers and fiber optic cable, the design and implementation of a digital optical wide band data transmission system with approximately 1 Gbit/s data rate capability can be accomplished. This limit is at present determined by the characteristics of electronic circuits which will be used for multiplexing and demultiplexing functions. The transmission and reception of data over single mode fiber optic cables at 1300 nm wavelength will be realized with very low bit error rates (BER) of 10^{-9}.

Analog data transmission in the optical wavelength regions of 850 and 1300 nm can be also accomplished over a very wide bandwidth. However, the operating conditions of light emitting and laser diodes must be optimized to reduce the device harmonic and intermodulation products as well as their temperature sensitivity to an acceptable level.

Optical sources, fiber waveguides, optical receivers and associated electronics circuits will suffer a measurable degradation of their operating characteristics when exposed to radiation. However, by using existing radiation hardened components and choosing an appropriate operating wavelength, digital and analog fiber optics transmission links can be designed to function during the exposure to ionizing radiation and neutron fluence expected in the SSC environment. Systems design will then include adequate optical power margins to maintain the signal-to-noise ratio necessary for reliable operation. It is essential to evaluate the expected incremental optical power loss in various SSC data transmission systems.

Accurate evaluation of the expected optical power loss in data transmission systems involves specific data on the amount and spatial distribution of the total dose, dose rate, required bandwidth and the environmental temperature in the SSC detector system and primary beam tunnel. The total dose and dose rate data are particularly important because the net radiation-induced attenuation in a fiber waveguide strongly depends on the competing processes of color center formation and recovery. In general, for smaller dose rate the induced attenuation is smaller, providing that the fiber recovers in the time scale of the exposure. Also, the attenuation can be significantly reduced by photobleaching effects, increasing the injected optical power levels and by higher environmental temperatures. In especially critical radiation environments the optical data link should be operated at longer wavelength, such as 1300 or 1500 nm, where the induced attenuation is smaller in comparison with that at shorter operating wavelengths.

It will be necessary to evaluate contemporary low radiation sensitivity fibers made by several manufacturers at long wavelengths for total doses larger than 10^2 Gy and dose rates typical for various SSC subsystems.

ACKNOWLEDGEMENTS

This work was performed as part of the program of the Electronics Research and Development Group, Electronics Engineering Department of the Lawrence Berkeley Laboratory, University of California, Berkeley. The work was partially supported by the U.S. Department of Energy under Contract Number DE-AC03-76SF00098. Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

15. C.E. Barnes, The Effects of Radiation on Optoelectronic Devices, Proceedings of SPIE-Fiber Optics in Adverse Environments III, Vol. 721, pp. 18-25, 1986.

16. H. Okuda, T. Fujitani et. al., Radiation Effects on InGaAsP/InP DH LEDs ($\lambda_p = 1.3\ \mu m$), 46th Meeting of the Japan Society of Applied Physics, 3a-N-1, p. 209, 1985.

17. Optical Fiber Telecommunications II, S.E. Miller, I.P. Kaminow, Eds., Academic Press, Inc., 1988.

18. Report of the Nuclear Effects Task Group, NATO Panel IV, Research Study Group 12, P.B. Lyons (Ed.) Los Alamos National Laboratory Publication LA-CP-87-158, 1987.

19. Y. Chigusa, M. Watanabe, M. Kyoto, M. Ooe, T. Matsubara, S. Okamoto, T. Yamamoto, T. Iida and K. Sumita, γ-ray and Neutron Irradiation Characteristics of Pure Silica Core Single Mode Fiber and its Lifetime Estimation, IEEE Transactions on Nuclear Science, Vol. 35, No. 1, pp. 894-897, February 1988.

20. B. Leskovar, Optical Receivers for Wide Band Data Transmission Systems, LBL-25119, April 5, 1988.

21. T. Kaneda, Silicon and Germanium Avalanche Photodiodes, Semiconductor and Semimetals, 22, Pard D, W.T. Tsang, Ed., pp. 247-328, Academic Press, 1985.

22. J.C. Campbell, W.T. Tsang, G.J. Qua and J.E. Bowers, In/InGaAsP/InGaAs Avalanche Photodiode with 70 GHz Gain-Bandwidth Product, App. Phys. Lett., 51, (18), pp. 1454-1456, 2 Nov. 1987.

23. B.L. Kasper, and J.C. Campbell, Multigigabit-per Second Avalanche Photodiode Lightwave Receivers, Journal of Lightwave Technology, LT-5, No. 10, pp. 1351-1364, 1987.

24. B.L. Kasper, J.C. Campbell, J.R. Talman, A.H. Gnauck, J.E. Bowers and W.S. Holden, An APD/FET Optical Receiver Operating at 8 Gbit/s, Journal of Lightwave Technology, LT-5, pp. 344-347, 1987.

25. K. Sleger, S. Mack, C. Scott and F. Buot, Compound Semiconductor Digital Integrated Circuits, Microwave Journal, Vol. 29, No. 8, pp. 85-95, 1986.

26. J.E. Gover and T.A. Fischer, Radiation-Hardened Microelectronics for Accelerators, IEEE Transactions on Nuclear Science, Vol. 35, No. 1, pp. 160-165, 1988.

27. S.E. Kerns and B.D. Schafer, The Design of Radiation-Hardened ICs for Space: A Compendium of Approaches, Proceedings of the IEEE Transactions on Nuclear Science, Vol. 76, No. 11, pp. 1470-1509, 1988.

REFERENCES

1. R.S. Tucker, S.K. Korotky, G. Eisenstein, L.L. Buhl, J.J. Veselka, G. Raybon, B.L. Kasper, A.H. Gnauck, and R.C. Alferness, 16 Gbit/s Optical Time-Division-Multiplexed Transmission System Experiment, Technical Digest of Optical Fiber Communication Conference, Paper THB-2, Optical Society of America, 1988.

2. S. Fujita, N. Henmi, I. Takano, M. Yamaguchi, T. Torikai, T. Suzaki, S. Takano, H. Ishihara and M. Shikada, Technical Digest of Optical Fiber Communication Conference, pp. 276-279, Optical Society of America, 1988.

3. R.J. Sanferrare, Terrestrial Lightwave Systems, AT&T Technical Journal, 66, No. 1, pp. 95-107, 1987.

4. H.H. Williams, Detectors for the SSC: Summary Report, Proceedings of the 1986 Summer Study on the Physics of the Superconducting Supercollider, R. Donaldson and J. Marx, Eds., June 23-July 11, pp. 327-349, 1986.

5. T.J. Devlin, A. Lankford, and H.H. Williams, ibid, pp. 439-454.

6. SSC Central Design Group, Conceptual Design of the Superconducting Super Collider, SSC-SR-2020, March 1986.

7. B. Leskovar, M. Nakamura and B.T. Turko, Optical Wide Band Data Transmission System, Lawrence Berkeley Laboratory Report, LBL-23113, March 15, 1987.

8. B. Leskovar, M. Nakamura, and B.T. Turko, Wide Band Data Transmission System Using Optical Fibers, IEEE Transactions on Nuclear Science, Vol. 35, No. 1, pp. 334-341, February 1988.

9. M. Nakamura, B. Leskovar and B.T. Turko, Signal Processing in Wide Band Data Transmission System, IEEE Transactions on Nuclear Science, Vol. 35, No. 1, pp. 197-204, February 1988.

10. B. Leskovar, M. Nakamura, F.A. Kirsten, and A.R. Clark, Study of Fiber Optics for Application to SSC Detector Systems, LBL-PUB-5216, August 18, 1988.

11. B. Leskovar, Radiation Effects on Optical Data Transmission Systems, Lawrence Berkeley Laboratory Report, LBL-25062, April 15, 1988.

12. SSC Central Design Group, Radiation Effects at the SSC, M.D.G. Gilchriese, Ed., SSC-SR-1035, p. 74, June 1988.

13. T. Suzuki, T. Ebata, K. Fukuda, N. Hirakata, K. Yoshida, S. Hayashi, H. Takada, T. Sugana, High-Speed 1.3 μm LED Transmitter Using GaAs Driver IC, Journal of Lightwave Technology, Vol. LT-4, No. 7, pp. 790-794, July 1986.

14. C.H. Henry, Spectral Properties of Semiconductor Lasers. Semiconductors and Semimetals, Vol. 22, Part B, W.T. Tsang, Ed., pp. 153-204, Academic Press, 1985.

Time Measurement System at the SSC

Yasuo Arai
KEK, National Laboratory for High Energy Physics, Ibaraki 305 JAPAN

ABSTRACT

A proposal of time measurement system at the SSC experiment is described. An example of a possible scheme for central tracking chambers is shown. Designs of a preamp/shaper/discri chip and a time digitizer chip are described. A method to distribute system clock and power/cooling problems are also discussed.

INTRODUCTION

Time measurement of nano second resolution has been important technique in High-Energy Physics experiments. This is also true at the SSC experiment. Several detectors, such as the central tracking and muon detectors, will need a high-accuracy (< 1 ns) timing measurement system to get track information from charged particles. However, at the SSC, there are many difficulties in the time measurement because the beam crossing time (16 ns) is less than the detector response time (~100 ns) and the trigger decision time (~1 μs). In addition, the number of channels to be read is very large ($\geq 10^6$ channels) and the front-end electronics must work in a high radiation environment. These conditions require new technology for the time measurement system, which should include radiation-hard, low-power, intelligent and highly-multiplexed devices.

A schematic diagram of a proposed front-end system is shown in Fig.1. Following the preamplifier, which is optimized for a particular detector capacitance, a shaping amplifier determines the bandwidth and the pulse shape. A discriminator is included to generate a timing signal. A preamp/shaper/discriminator chip is being developed by using bipolar technology to get the best performance while keeping the power consumption as low as possible. A time digitizer chip (TMC) is also being developed by using CMOS technology. The TMC has a time resolution better than 1 ns and dissipates little power compared with a shift-register. These improvements result from a new digital time measurement scheme and a high technology CMOS process. The readout control chip has a second level buffer and readout sequencers which manages the buffer read/write operation. Data sent out from the detector through coax or optical fiber cables will be multiplexed only after receiving the second level trigger signal. There is another approach[1] for time digitization. It is a mixed analog and digital system, which use a time to voltage converter, analog buffers and a on-chip A/D converter. This system may have better resolution than the TMC under bench-test conditions, but the circuit is heavily relied on the matching of capacitors and current sources. So it may be hard to use in large scale applications in a high radiation environment.

Since we need 1 μsec to make the 1st level decision whether or not to save the data, it is necessary to install buffer memories to keep all data for 1 μsec. If we have a two level trigger scheme, we must also keep the data in the second level buffer after receiving the level 1 trigger. When the level 2 trigger comes, the data must be sent outside the detector in ~10 μsec (We expect a combined rejection ratio of 10^4 for level 1 and level 2 trigger). For the case that only one level trigger is used with a rejection ratio of 10^3, we don't need to have the second level buffer although we may still need some FIFO buffer. But in this case we must send all the data in ~1 μs.

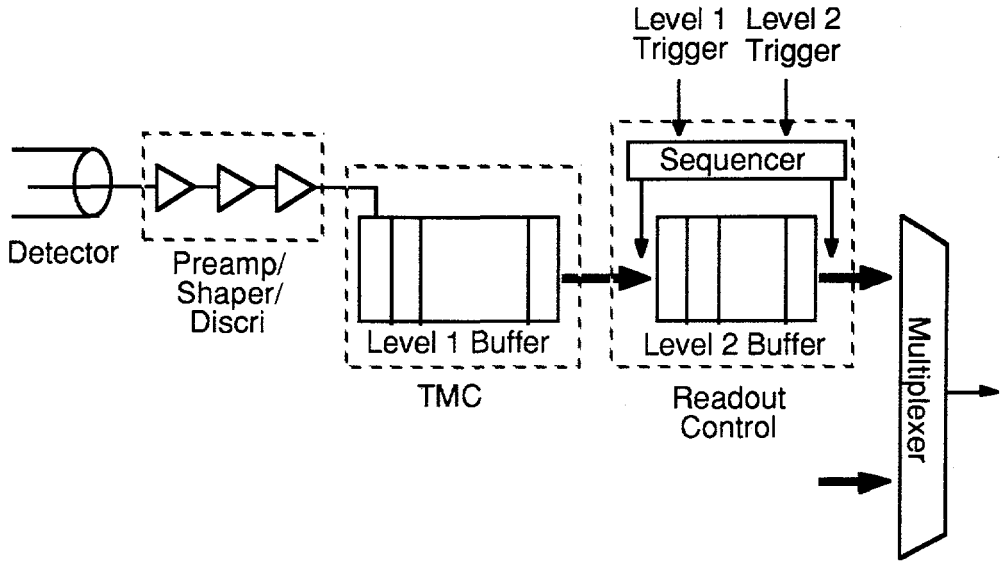


Fig. 1 A schematic diagram of the front-end electronics for time measurement system.

PREAMP/SHAPER/DISCRI CHIP

A monolithic preamp/shaper/discriminator chip is being developed at KEK. A test chip of preamplifiers has already been designed and tested[2]. A new 64 channel chip[3] is now being developed for use in Si strip detectors. The main features of this chip are listed in Table 1 and the mask layout of the chip is shown in Fig. 2.The amplifier head is a current sensitive preamplifier, followed by a low pass filter that determines the noise frequency bandwidth. Finally the signals are fed into a discriminator. Total power dissipation for a preamp-shaper-discriminator chain is estimated to be 3.3 mW/ch. The process technology for fabrication is a super self-aligned bipolar technology (SST) developed by NTT LSI Laboratories. Another advantage of the SST process, in addition to its high current gain (120) and high cut-off frequency (f_T = 17 GHz), is its radiation hardness (up to 10^6 rads and 10^{13} neutron / cm2)[4]. So the chip will be able to work at the vertex detector. Although the chip is designed to be used in digital readout of Si strip detectors, almost the same design can be applied to time measurements.

Table 1 Characteristics of the preamp/shaper/discriminator chip.

* No. of Channel	64 ch
* Chip Size	6.82 x 4.9 mm2
* Shaping Time	T_M = 15 nsec
* Reference Voltage Stability	±0.1 mV/°C
* Comparator Input Voltage	16 mV/25000 electron
* Noise Value	1000 electron (5 - 10 pF)
* Power Consumption	3.3 mW/ch

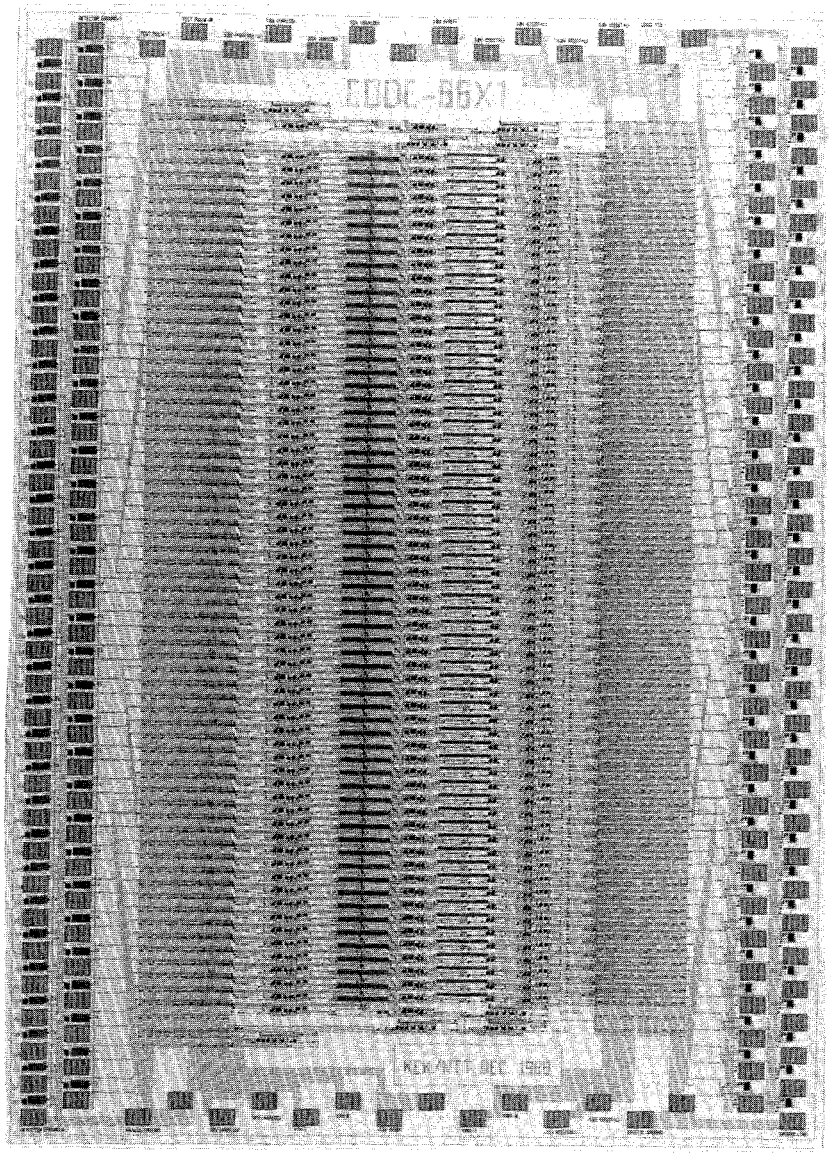


Fig. 2 Layout of the 64 channel preamp/shaper/discri chip.

TMC CHIP

A new CMOS TDC chip, the Time Memory Cell (TMC)[5,6,7] is now being developed to use in high rate experiments. The chip has multi-hit capability and the first-level buffer is inside the chip. Through use of CMOS, VLSI sub-micron lithography and a new circuit scheme, the TMC is a very low power device, compared with a shift register. The high accuracy of the TMC is accomplished by referencing the delay time of the buffer stage to an external clock.

A test chip has been developed and tested. The chip has 21 bit x 5 column basic cells and a reference cell as shown in Fig.3. The basic element of the TMC is composed of a memory and a buffer. Input data are recorded in memories where the write pulse is delayed by an interval $\Delta\tau$ in each buffer. Although the uniformity of the delay time of a buffer is fairly good within a chip, the delay time may vary by more than 20% because of instability in silicon process, the power supply voltage or temperature variations. To keep the delay time constant, a feedback circuit is

implemented. The circuit measures the total delay time of a delay line (buffers in series), compares it with an external clock and makes corrections to the delay times. In the case a shift-register is used for the TDC, we need a 1 GHz clock to attain 1 ns resolution. However, in the TMC, the required clock frequency is lowered by a factor of the length of delay line. Furthermore, the data are not moved once written to memory, so the power consumption of the circuit is low.

The length of the TMC reference cell is 20 bits; the last bit of the other column is used as a overlap bit between columns. The external clock pulses enter the ring counter which, in turn, selects a column or the reference cell. The ring counter generate a pulse to the column with width one half of the clock period. The contents of the memory in a column are read out in parallel through the sense amps by asserting one of the Read lines.

The circuit diagram of a basic cell (a memory and buffer) of the TMC is shown in Fig. 4. Each CMOS static memory is composed of two cross-coupled inverters (M1-4). We selected a static cell because it may be more stable to use in a high radiation environment. Each buffer is composed of two inverters (M9-12) and a transistor (M13). The gate voltage (Vg) of M13 controls the rise time of the first inverter. The pulse shape is restored by the second inverter, but the pulse width is changed. As the input on the data line is latched in the memory at the falling edge of the pulse at the second inverter, the write timing between each cell is changed by Vg.

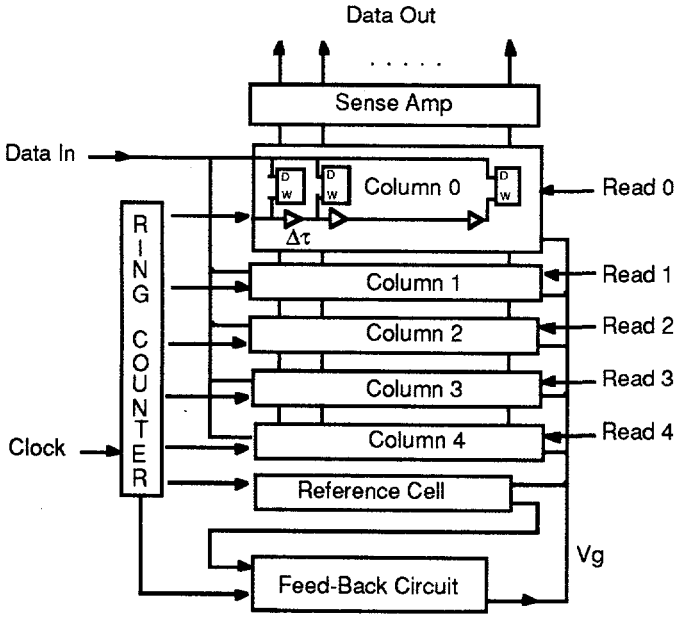


Fig. 3 Block diagram of the TMC test chip.

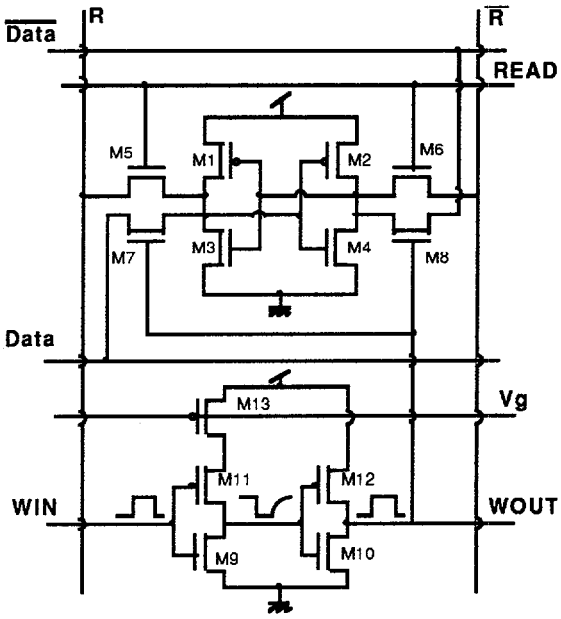


Fig. 4 Basic Cell of the TMC.

Fig.5 shows the schematic of the feedback circuit. When an external clock sets two flip-flops (F1 and F2) at the falling edge, capacitors C1 and C2 will begin to charge. The charging of C1 will be stopped by the falling edge of the pulse at the end of the delay line. The charging of C2 will cease at the next falling edge of the clock pulse. Hence, the voltage difference between C1 and C2 is proportional to the time difference between the delay line and one clock period. If the delay time is less than the clock period, C3 will charge during a store period. This will increase the delay time of the delay line. If the delay time is longer than the one clock period, C3 will discharge and the delay time will be reduced.

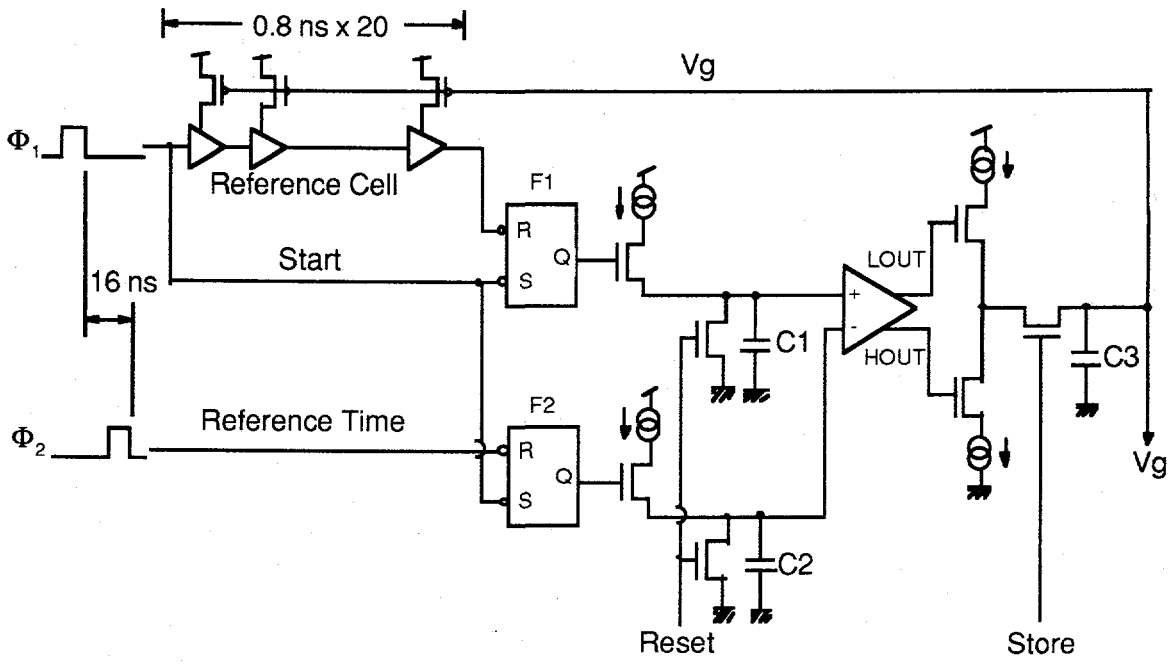


Fig. 5 TMC Feedback Circuit.

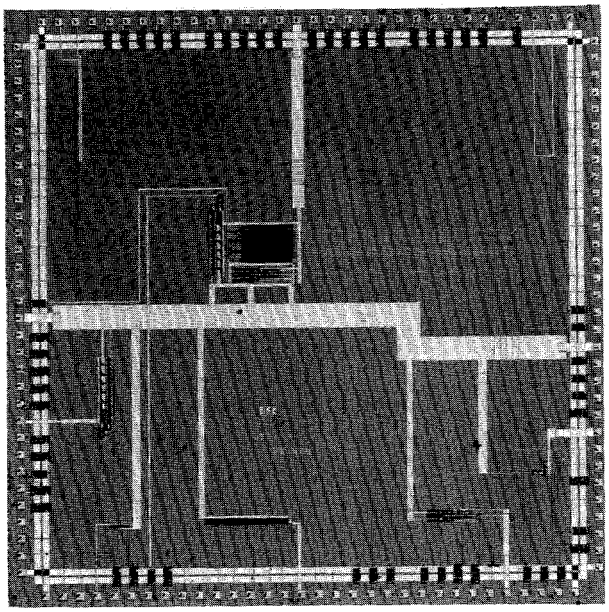


Fig. 6 Microphotograph of the TMC chip.

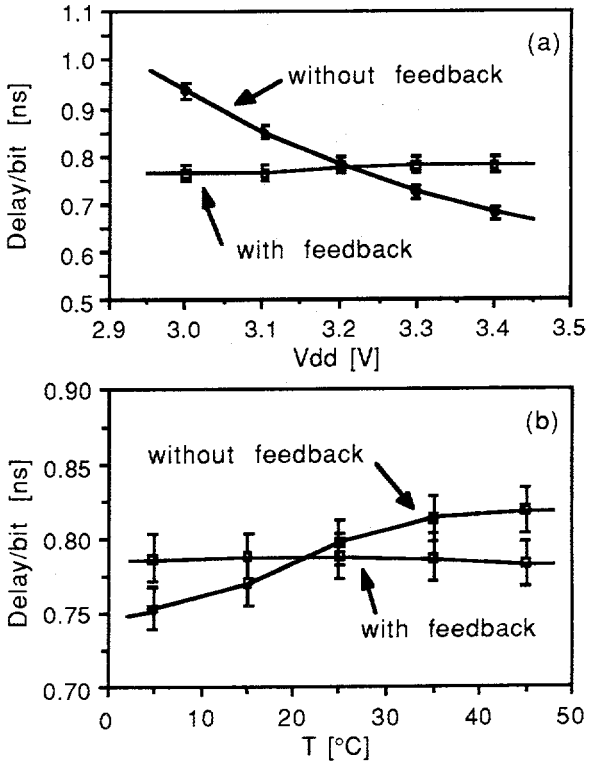


Fig. 7 (a) Supply voltage dependence of the delay element. (b) Temperature dependence of the delay element.

The test chip, shown in Fig. 6, measures 5 by 5 mm. The chip was operated at a supply voltage of 3.2 V, and a clock frequency of 60 MHz. Power consumption for the circuit shown in Fig. 3 (including the I/O buffer) is about 20 mW. The delay time distribution for each buffer is 0.77 ± 0.08 ns. Integral linearity of the chip is better than 0.5 ns and can be improved more in the next version. The supply voltage (Vdd) and temperature dependence of the delay are shown in Fig.7. As seen in the figure, the feedback circuit worked excellently in compensating the delay shift due to the power supply voltage and temperature variation .

New TMC chip, which has 1 ns x 1024 bit buffer and 2 to 4 channels in a chip, is being developed.

READOUT SCHEME

Assuming the central tracking detector of the large solenoid detector designed at the Berkeley workshop[8], an example of the readout scheme is described below. Fig.8 shows the front-end electronics for the time measurement system. These electronics will be composed of a set of 3 chips; a preamp/shaper/discriminator chip, a TMC chip, and a readout control chip. Both the preamp and the TMC chip are assumed to have 8 channel in a chip. The readout control chip will be developed by using a gate array or a standard cell. We need 64 columns in the TMC to implement first level buffer of 1 μs storage. As the straw tube used in the detector can not distinguish two hits in a single event and the maximum drift time will be more than 20 ns, the TMC does not need to have multihit capability within a column. So we can reduce the data size by encoding a column of data. We expect the TMC will have the resolution of 0.5 ns/bit and 32 bit column length, then these 32 bit data will be encoded into 5 bit by using a priority encoding logic. Write select pointer selects the column to be written, and incremented by the beam crossing signal. Read select pointer points a column to be read at the 1st level trigger.

The readout control chip has a second level buffer, sequencers and parallel to serial converter. The readout sequencer for the level 1 trigger generates addresses of the column to be read when it receives the level 1 trigger. This chip must have a register recording the number of columns to be read [N], because the maximum drift time in the wire chamber is expected to be longer than the beam crossing interval. The sequencer set a trigger bit at the first column of the data. This marks trigger point. Second level buffer will be controlled by the level 2 readout sequencer which has 3 pointers (write, read and trigger pointer) and emulate FIFO buffer by using these pointers. The write pointer points the next write position The read pointer points next readout position, and the trigger pointer points next trigger position. The trigger pointer is needed because we may have several triggers within the detector response time.

When the readout control chip receives the level 2 trigger, the data is sent out through parallel to serial converter. As the longest drift time of the straw tube is about 40 ns, we must to read at least 3 columns for each trigger. If the trigger decision circuit has some ambiguity in determining the bunch crossing time, we need to read more. So we will have 120 bit (8 ch x 5 bit x 3 column) of the data and the start column address (6 bit). By multiplexing 64 channels, we have to send out 1k byte data from the chip.Assuming the transmission line speed of 100 M bit/sec, we can send 1k byte data in 10 μs. If we have only one level trigger before sending the data from the detector, we may need to use 1G bit/sec transmission line to send the data in 1 μs. In this scheme, we have only 2000 cables from the detector to a counting room. We may be able to suppress zeros and encoding the data at the output of the second level buffer. Although the occupancy ratio of a wire is about 10 %, we also need to send address of each data at this time. Then the data size will be reduced to only one fifth of above scheme.

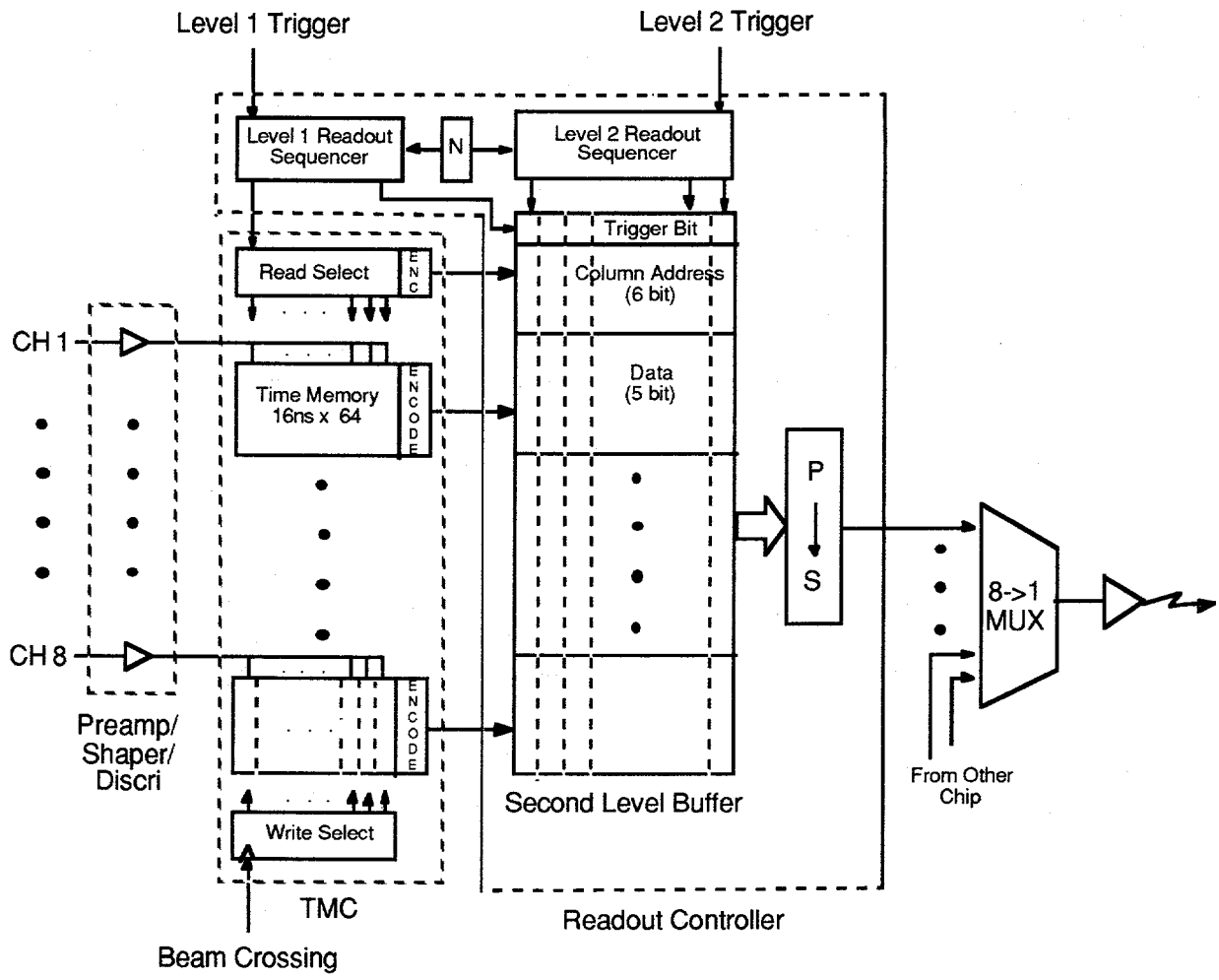


Fig. 8 Readout scheme of the time measurement system.

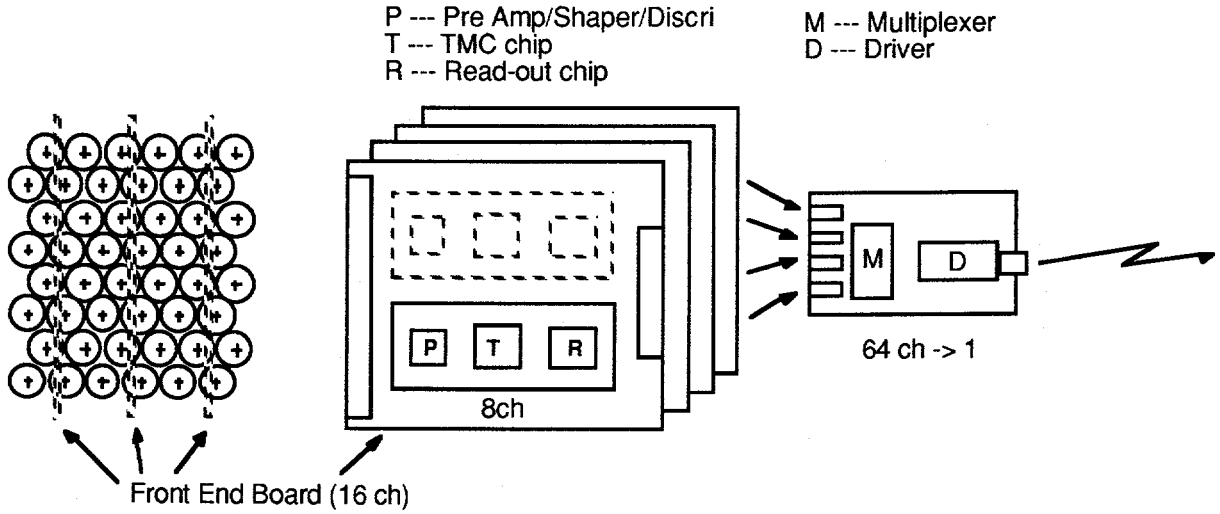


Fig. 9 Mounting of the front-end electronics to the 8 layer straw tube.

Fig. 9 illustrates the front-end boards and mounting to the straw tube chamber. The preamp chip, the TMC chip and the readout chip will be molded in a hybrid IC form. The front-end board mounts these hybrids on both sides, then 16 channels are managed by the board. The board interval for each superlayer is shown in table 2. To multiplex and transmit the data of 64 channels, the outputs of 4 board are connected to a multiplexer/driver board. This board will be placed between the superlayer or behind the preamp board.

POWER CONSUMPTION AND COOLING

In Table 2, several parameters for the the central tracking detector[8] which use straw tube chamber are summarized. The detector has 15 superlayers and each superlayer has 8 layers of staggered straw tubes. Total number of channels is about 1.2×10^5. Power consumption of the chip set (8 ch) is assumed to be 500 mW (25 mW, 250 mW and 225 mW for preamp, TMC and readout chip respectively). Then each board dissipates 1W, and the total power will be 7.7 kW. We also need 2000 multiplexers/driver cards to send the data out. Assuming the power dissipation of 0.5 W for this card, we have additional 1 kW power source. Then we have 8.7 kW power in total at the central tracking detector.

The power density is calculated assuming the front-end board size of 5 cm by 5 cm. The figures can be compared with that of the FASTBUS standard (30 mW/cm^3 for 75 W board). The values are similar to the FASTBUS except for inner 4 layers. To remove 8.7 kW heat by air, we need about 50 m^3/min air flow when the temperature rise of 10°C is admitted. This requires 50 cm diameter duct of 5 m/sec air velocity. This will make large crack in the detector. If we use water cooling system, we need 14 l/min flow of water for $\Delta T=10°C$, and need 5 cm diameter tube of 12 cm/sec water velocity. We can admit temperature difference of 10 °C in the system, because the system is digital and very stable to the temperature difference.

Table 2 Parameters for central tracking chamber of the straw tube.

Superlayer	Inner Radius (cm)	Straw Diameter (mm)	Board Interval (mm)	# of Channel	# of Board	# of Cable	Power Density (W/cm^3)	Radiation Level (krad/5year)
1	40	3.92	7.84	5120	320	80	51	125
2	48	3.92	7.84	6144	384	96	51	87
3	56	3.92	7.84	7168	448	112	51	64
4	64	3.92	7.84	8192	512	128	51	49
5	72	5.89	11.78	6144	384	96	34	39
6	80	6.04	12.08	6656	416	104	33	31
7	88	6.17	12.34	7168	448	112	32	26
8	96	6.28	12.56	7680	480	120	32	22
9	104	6.38	12.76	8192	512	128	31	18
10	112	6.47	12.94	8704	544	136	31	16
11	120	6.55	13.10	9216	576	144	31	14
12	128	6.61	13.22	9728	608	152	30	12
13	136	6.68	13.36	10240	640	160	30	11
14	144	6.73	13.46	10752	672	168	30	10
15	152	6.78	13.56	11264	704	176	29	9
Total				122368	7648	1912		

The radiation level at the position of the front-end electronics are shown in the last column of the table 2. Number of charged particle per unit rapidity 7 is used for the calculation. The bipolar process which is used at the preamp chip is known as radiation hard for both charged particle/γ-rays and neutron around the level expected at the SSC. Although the radiation hardness of the CMOS process used at the TMC is not yet tested, CMOS is known to be hard for neutron irradiation. Several CMOS process is also known as hard for charged particle/γ-rays of more than 100 krad. Radiation damage tests of the CMOS process for γ-rays and neutrons are now planned. There is no radiation problem for the muon chamber readout.

SYSTEM CLOCK DISTRIBUTION

It is not an easy task to distribute the beam crossing signal (62.5 MHz) to all the detector elements. Such a high frequency clock may induce noises in the system. Besides we must be careful to the system synchronization in a large system. The noise power of the clock is proportional to the capacitance of transmission line, the amplitude square and the frequency. The capacitance is determined by the cable length and the impedance. The signal amplitude may be reduced as much as possible (< 3 V). The frequency can be lowered by dividing the clock, and regenerate locally by using a Phase Lock Loop (PLL) technique. Fig. 10 shows a scheme of the system clock distribution. By dividing the frequency of 62.5 MHz by 64, we get about 1 MHz clock. This does not mean the bandwidth of the transmission line can be reduced, because we still need accurate phase. However this 1 MHz clock is easy to use and the noise power can be reduced very much. This clock is used as a reference to regenerate 62.5 MHz clock at the local system. In addition, the rising/falling edge can be used for synchronization of the system. For example, the edge can be used to synchronize the bunch number counter or reset circuits which need periodical clear. The PLL circuit is easily implemented by using LSI technologies.

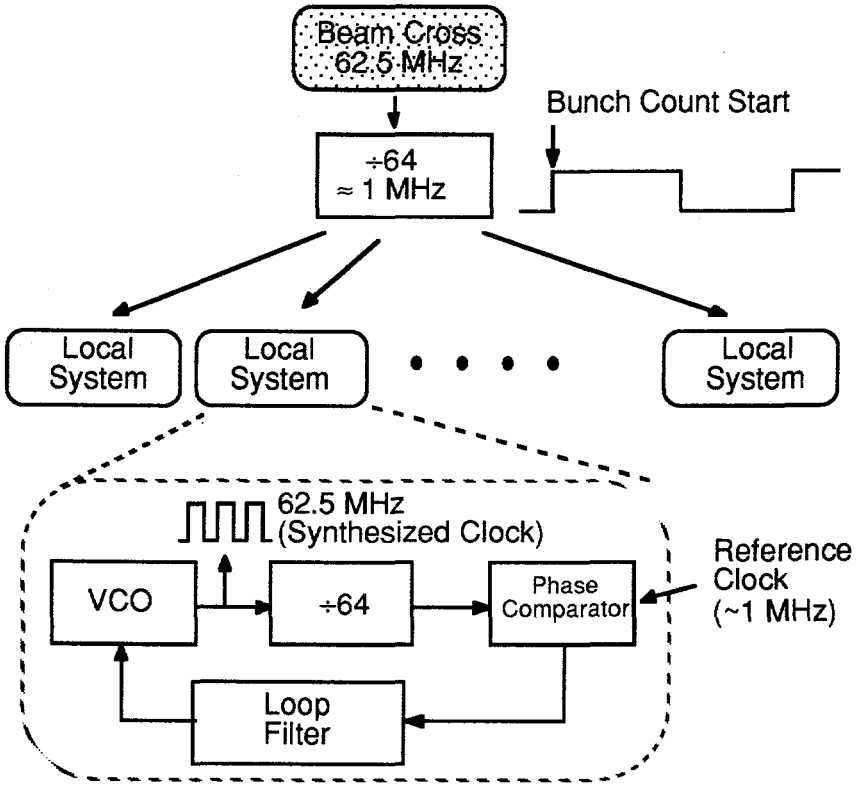


Fig. 10 An example of the system clock distribution by using frequency synthesizer.

SUMMARY

A possible scheme of the time measurement system for the SSC experiments is described. Key components of this scheme are the bipolar preamp/shaper/discri chip and the CMOS time digitizing chip (TMC). The prototypes of these chips have been tested and the results have already been reported[2,6,7]. It is possible to achieve time resolution of better than 1 ns, double pulse resolution of 16 ns, level 1 storage for 1 μs, and enough capacity for level 2 storage. The total power dissipation will be less than 60 mW/ch and can be cooled by water. Most of the system is digitally operated and has closed loop-back circuit, so it will work very stable for a long period. The loop-back circuit can also be used for the system clock distribution. The reasonably slow system clock frequency allows us to make easy synchronization of electronics of very large scale detector system.

ACKNOWLEDGEMENTS

I am grateful to Y. Watase, T. Ohsugi, H. Ikeda and T. Kondo for their continuing advice and support to the work.

REFERENCES

1 L. Callewaert et al., "Front end and signal processing electronics for detectors at high luminosity colliders", IEEE Trans. Nucl. Sci. Vol. 36, No. 1(1988).

2 H. Ikeda et al., "Monolithic Preamplifier with Bipolar SST for Silicon Strip Readout", IEEE Trans. Nucl. Sci. Vol. 36, No. 1(1989). KEK Preprint 88-71

3 H. Ikeda, "64 Channel Bipolar Amplifier for Silicon Strip Readout", Report for Workshop on Solid State Detector, Hiroshima University, Dec. 1988.

4 N. Ujiie, Talk at Workshop on Solid State Detector, Hiroshima University, Dec. 1988.

5 Y. Arai and T. Ohsugi, "An Idea of Deadtimeless Readout System by using Time Memory Cell", Proceedings of the 1986 Summer Study on the Physics of the Superconducting Super Collider, p455. KEK Preprint 86-64(1986).

6 Y. Arai and T. Ohsugi, "TMC: A Low-Power Time to Digital Converter LSI", IEEE NS Symposium, Oct. 1987. KEK Preprint 87-113(1987).

7 Y. Arai and T. Baba, "A CMOS Time to Digital Converter VLSI for High-Energy Physics", 1988 Symposium on VLSI Circuits, Aug. 1988/Tokyo, IEEE CAT.No. 88, TH 0227-9, p121.

8 "Report of the Large Solenoid Detector Group", Proceedings of the Workshop on Experiments, Detectors, and Experimental Areas for the Supercollider, p340, Berkeley, July 1987.

NEURAL NETWORKS, DO, AND THE SSC†

C. Barter, D. Cutts, J. S. Hoftun, R. A. Partridge, A. T. Sornborger
Physics Department, Brown University, Providence, RI 02912 USA

C. T. Johnson and R. T. Zeller
ZRL, Cranston, RI, 02905 USA

ABSTRACT

We outline several exploratory studies involving neural network simulations applied to pattern recognition in high energy physics. We describe the D0 data acquisition system and a natural means by which algorithms derived from neural networks techniques may be incorporated into recently developed hardware associated with the D0 MicroVAX farm nodes. Such applications to the event filtering needed by SSC detectors look interesting.

INTRODUCTION

This talk, at our workshop for triggering at the SSC, is very much of a working paper describing some activities underway at Brown studying the use of neural network techniques for event filtering. My descriptions are of work unfinished with results (if any) which are very preliminary — yet it seems appropriate to present activities at this workshop which may point in new and potentially important directions.

I shall describe two exercises involving the use of neural networks for pattern recognition. Before doing so, let me give some defense for our spending time with the "games" involving our PC-based network simulator. A proper answer to the question: "Why neural networks?" has been given by Bruce Denby in the previous talk, and in several publications.[1] Here you will find discussion of these networks, and why they learn. In my talk let me simply repeat what little I know of the "lore", which is that conventional Artificial Intelligence programs are built using tree–structured logic much like our usual filter programs. In these systems, like our analysis code, the larger and more complex the data, the slower and more difficult the AI recognition process. However, neural networks (such as our own) seem to work better with more data, not worse. Somehow these structures are truly "Parallel Distributed Processing" — the title of the basic textbook[2] of neural networks. Since we in high energy physics have learned with computer farms that parallel processing greatly enhances event filtering, perhaps through neural networks we can similarly enhance algorithm performance.

THE D0 DATA ACQUISITION SYSTEM

The immediate stimulus of this work is its potential application to the D0 experiment at the Fermilab 2 TeV$p\bar{p}$ collider. We are responsible for the D0 data acquisition system,[3,4,5] which is based on a MicroVAX "farm" and is the site of D0's Level-2 (software) trigger. Because of a concern that the filtering performance of this system be maximized, we are looking at various methods to supplement each node's capabilities. As I will outline briefly, a new generation of "multiport" memories provides a special function port which is appropriate for special purpose devices which support neural network-derived algorithms.

† This work supported in part by the U.S. Department of Energy.

Data Flow at D0

The basic feature of the D0 data collection is a parallel, high speed data flow directly from 100 VME digitization crates to memories associated with a single, selected Level-2 filter node. Digitization of an event ($p\bar{p}$ interaction) is started when a hardware (Level-1) trigger, using separate, fast-summed outputs, is satisfied. Each VME ADC or FADC crate is equipped with a VME buffer board which has an external port coupling to a 40 MByte/sec data cable. The total of 8 data cables, allocated by detector type to groups of VME digitization crates, are connected in turn to memories in the Level-2 nodes. The 8 data cables have an aggregate throughput of 320 MBytes/sec for the event (typically 250 KBytes) into the filter processor's memory, from which it is NOT moved: the data management structure (ZEBRA) is built around it. The 50 or more Level-2 MicroVAXes, each analyzing a separate event, must in aggregate handle a 400 Hz input event rate and select an expected 1-2/sec interaction events to be recorded.

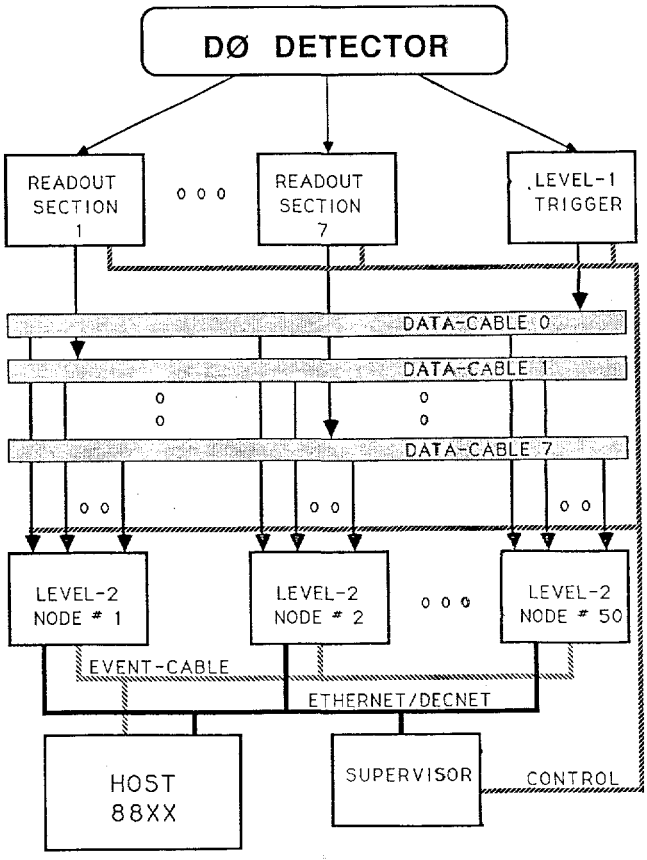


Fig. 1 Block diagram of the DØ data acquisition system.

Multiport Memory for the Level-2 MicroVAXes

The Level-2 nodes will make use of a new generation of memory boards which has been developed by Zeller Research.[6] In distinction to the first generation of dualport memories, these boards are truly "multiported": in addition to the external data cable port and the Q-Bus port, there is a direct memory connection to the processor, a high speed output port (for sending events to the host), and a special function port. Hardware-assisted algorithms such as those derived from neural network studies would access the data through this last port. Figure 2 describes the population of each Level-2 node, while Figure 3 shows as an example a multiport memory board, with one specific array processor implementation. Because of the extreme demands on Level-2 for real time filter analysis, such hardware-assisted algorithms could be very important.

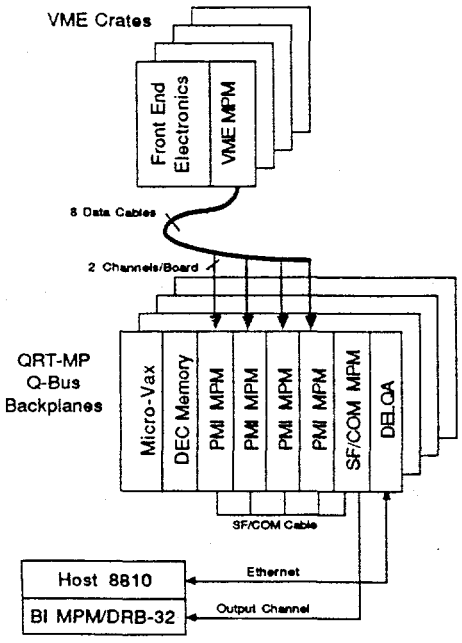


Fig. 2 Schematic of next generation DØ Level–2 node.

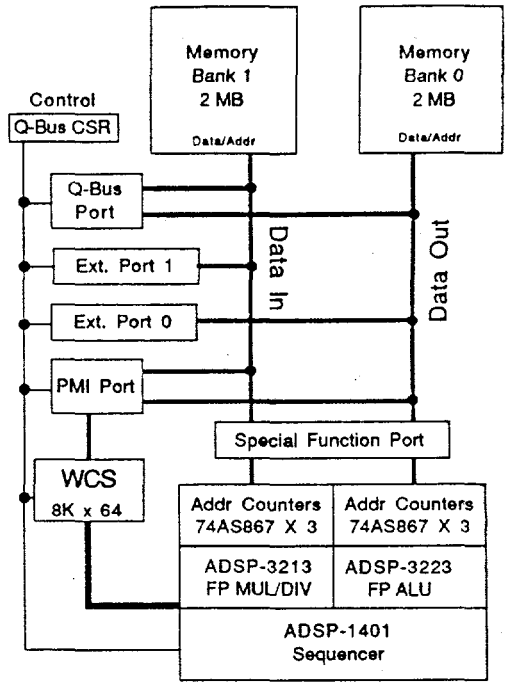


Fig. 3 Block diagram of multiport memory with floating point coprocessor.

EXERCISES WITH NEURAL NETWORK SIMULATIONS

Network Simulation Tools

To explore the usefulness of neural networks for pattern recognition in high energy physics, one tests the ability of some computer-based neural model to learn on sample data. A great convenience in this effort is a software package that allows one to create various simulations and perform the "training and recall" which tests the model. One of us (A.T.S.) attended the Neural Network Conference in Boston last September, and from the various products demonstrated chose the simulation software "Professional-II".[7] This package is graphic/mouse-based and very user-friendly; with it a newcomer to the field can jump in and quickly exercise variants of canned networks, while the expert has a very broad range of options and building tools. For the beginner the variety of network architectures, parameters — weights, transfer functions, learning rules, and so forth — is impressive; the package serves as a user-training tool as the user attempts to train some modeled network on physics data.

We ran the Professional-II simulation package on IBM-PCs (it is also avaliable for some workstations). First installed on a PS/2 model 50, the package trained the small network described below on 10,000 events, in roughly 10 minutes. After this work, we moved the package to a PS/2 model 70 where the much larger network described in the following section processed 40,000 events in a few hours.

The procedure we have adopted is to use the simulation package to study various designs and training procedures — to evaluate the ability of different neural networks to recognize specific features of high energy physics events. As will be obvious from the examples below, this work is very much in progress (rather than completed). Our immediate goal is to find designs which prove sufficiently effective to implement in the Level-2 hardware described above, while a longer term aim is to develop algorithms appropriate for SSC detectors.

Exercise 1: electron/photon separation in E734

We first studied the ability of a neural network to distinguish between 1 GeV electrons and photons as was needed for an earlier experiment which involved some of us (D.C., J.S.H.). This project was E734 at Brookhaven, a study of neutrino-electron elastic scattering at the AGS. This rare process, fundamentally important in the study of weak interactions, has major backgrounds due to the fact that a forward-going photon (produced in neutrino charged-current π^0 production) is difficult to separate from the isolated electron from $\nu - e$ scattering. The E734 detector was designed specifically to reduce this and related backgrounds, and has a superb signal to noise compared to its competitors. The basic idea was to build the detector out of fully active elements with long radiation lengths: planes of liquid scintillator interspersed with (x,y) planes of aluminum proportional drift tubes. Repeated samplings of the dE/dx close to the interaction vertex (within the first radiation length) then provided a good means to separate forward electromagnetic showers produced by photons from those produced by electrons.

In E734 our ability to separate electron and photon showers, based on recognizing the energy depositions of 1 vs. 2 electrons, was limited by a combination of poor resolution, Landau tails on the single electron dE/dx distribution, and inefficiencies of the PDTs. We

tried various methods of analyzing the data and settled on the following algorithm: for each x,y pair of PDT's at the front of the shower we picked the lower of the dE/dx values (or if only one was available, that value) — and then averaged these values for a parameter on which the selection was made. To retain most of the electron sample we placed the cut on this parameter such that 90 % of Monte Carlo electrons, or test beam electrons, were passed; the cut then eliminated 50 percent of the photons. For the experiment this effectiveness was fine, as other criteria such as the extreme forward kinematic peak of true $\nu_\mu + e \rightarrow \nu_\mu + e$ events combined with our good angular resolution enabled us to have an excellent signal/noise of 3/1 for this rare process.[8,9]

For our first study we started with data that modeled the dE/dx distributions observed by E734 for electrons and photon showers (as, in a test beam) — including broad resolution (60 %), Landau tails, and detector inefficiencies. Using the simulator described above, we created the 3 layer (input, hidden, output) feed-forward network shown in Figure 4. In this network there are 6 "processing elements" (PE's) to which the input data for each event (6 dE/dx values) are presented. Each of these PE's is "connected" to all 3 PEs in the middle layer, which are each in turn connected to both output PEs. This particular network is thus described by the 6*3+3*2 = 24 "weights" (one per connection) as well as the details of the "transfer function" which specifies the output of a PE in terms of its inputs. For example, the output of the second layer elements, represented as a vector V, is given by

$$V_i = \zeta(\sum_j W_{ij} U_j) \tag{1}$$

where each W_{ij} is a weight assigned to the connection between PE_i and PE_j, U_j is the output of first layer PE_j, and ζ is the Sigmoid function that maps with a "S-curve" smoothly to the output, between 0 and 1 (see References 1,2).

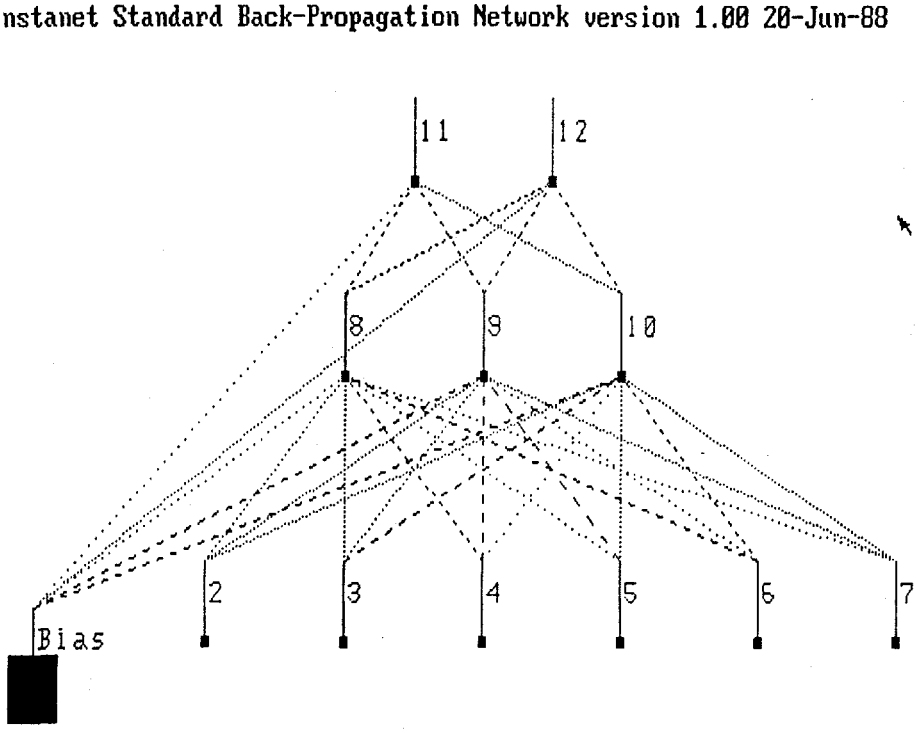


Fig. 4. Three layer feed-forward network used to study electron/photon separation.

To train the network, we presented a file of events to the simulation package, which included the appropriate outputs from the two 3rd layer PEs (1,0 for electron data and 0,1 for photons). Using the "back-propagation" technique[1,2] the network's weights were adjusted after each event to be more consistent with the desired output. After training the network on 1000 events (randomly mixed electrons and photons), run 4 times, we tested the degree of learning by studying its response to 100 new events. Its response to one of these events is shown in Figure 5; here the large and dark (colored) boxes represent highly excited PEs, and the network's output layer correctly indicates a recognition of a photon.

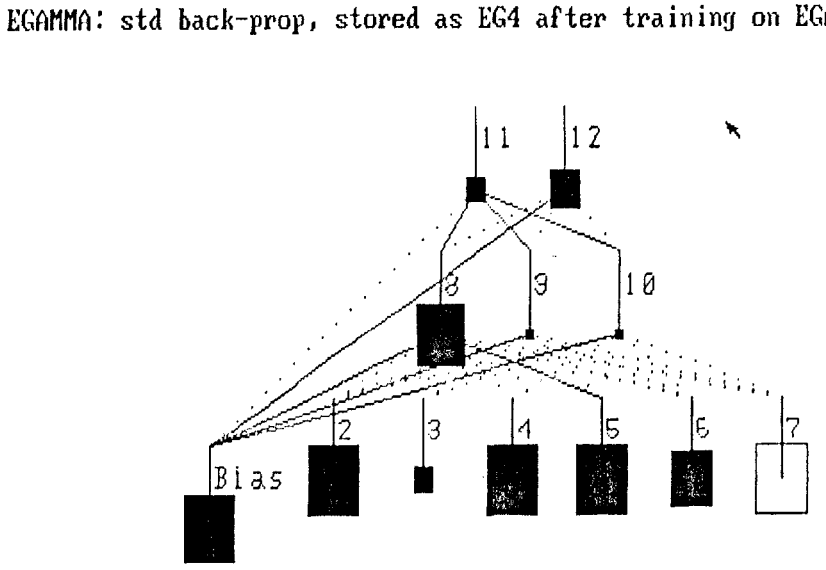


Fig. 5 Response of the trained network to a typical event.

The two output values of this network are constrained to sum to be 1, so we can plot the value of, say, the first of these PE excitations on a scale 0 to 1, for both electrons and photons in the independent 100-event sample. The results of this recognition test are shown in Figures 6 and 7, where there appears a very clear separation between the two intermixed samples. In principle one could take the weight matrices acquired by the trained network and create a fast electron identifier using matrix multiplier chips operating on the input data. The performance of this neural network derived algorithm is similiar to that actually used in the E734 analyses, though because of the different data samples a direct comparison isn't possible. It is clear, however, that the simple network described here learned to distinguish electrons from photons.

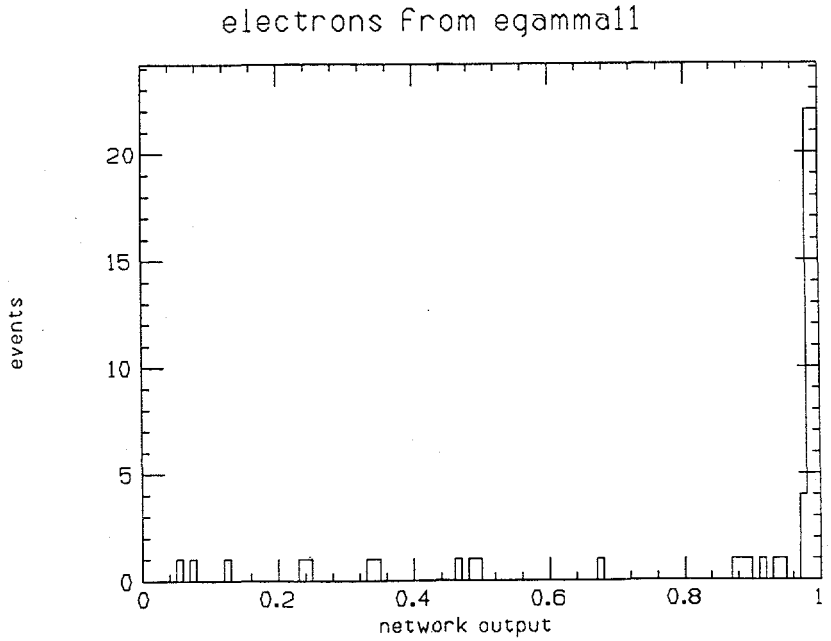


Fig. 6 Response of the trained network to electron events.

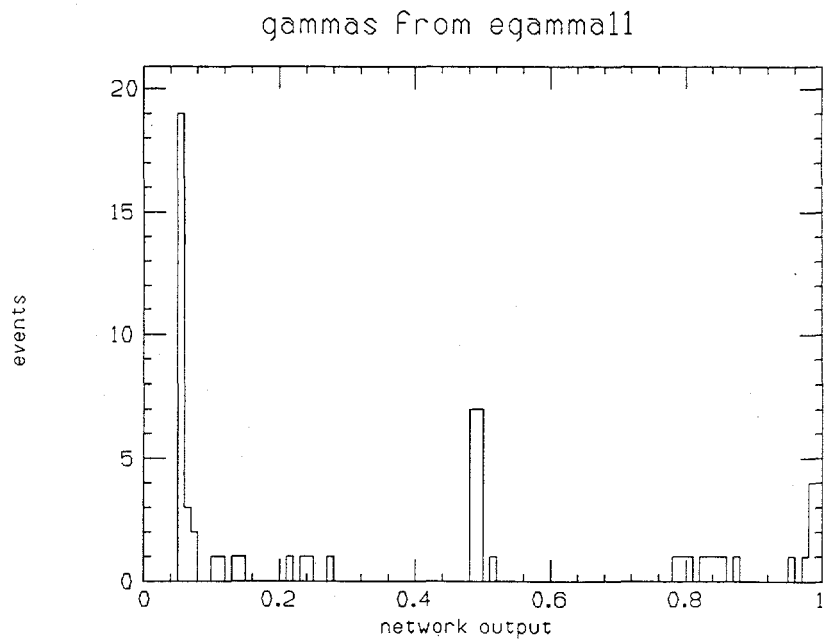


Fig. 7 Response of the trained network to photon events.

Exercise 2 : Higgs' photon / hadron shower separation for the SSC

As a second exercise we looked for an event recognition problem more relevant to those we will face in D0. It happened that there was a readily avaliable data set generated for a Snowmass study of Higgs $\rightarrow \gamma\gamma$ at the 40 TeV energy of the SSC.[10] The problem addressed in Ref. 10 by Barter and Partridge is to separate photon showers from the decay

$$Higgs \rightarrow \gamma\gamma \tag{2}$$

from background showers in the predominant 2-jet events. For this study ISAJET events were collected by a model calorimeter with bins of 0.1 in $\delta\phi$ and $\delta\eta$. As shown in Figure 8, the signal for the Higgs $\rightarrow 2\gamma$ decay is 2 isolated high E_t deposits. The problem is that at some low rate (perhaps larger than the Higgs $\rightarrow 2\gamma$ decay rate of 10^{-4}) the 2-jet events also can give this signal.

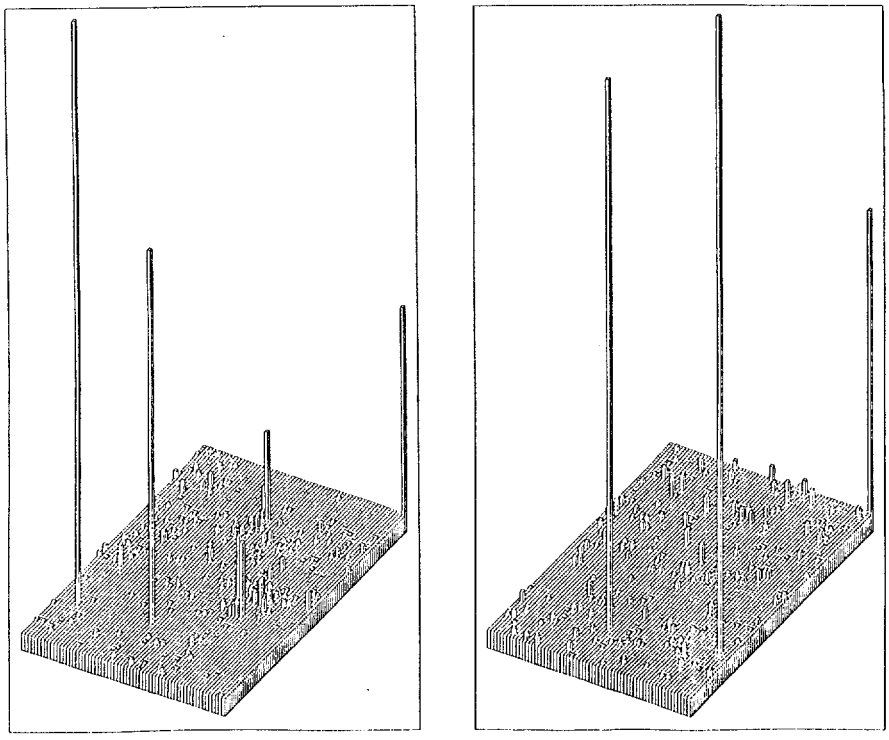


Fig. 8 Lego plot showing E_t distribution for two typical Higgs $\rightarrow \gamma\gamma$ events. To show the vertical scale, 25 GeV was entered in the right rear corner bin.

The algorithm used in Reference 10 to distinguish the Higgs events considered, beyond the presence of two high Pt showers, the population of bins nearby the peaks : the very high energy photons showed less grass around the main trunks. This recognition problem seemed appropriate for a neural network. We used the same data sets but rewrote the showers into arrays of calorimeter energy deposits, in the 0.1 bins, of size 11 x 11, centered about the peak of each shower (jet or Higgs photon). Each such shower was tagged as to its type, as needed during network training. Using the network simulator described above, we built the array shown in Figure 9.

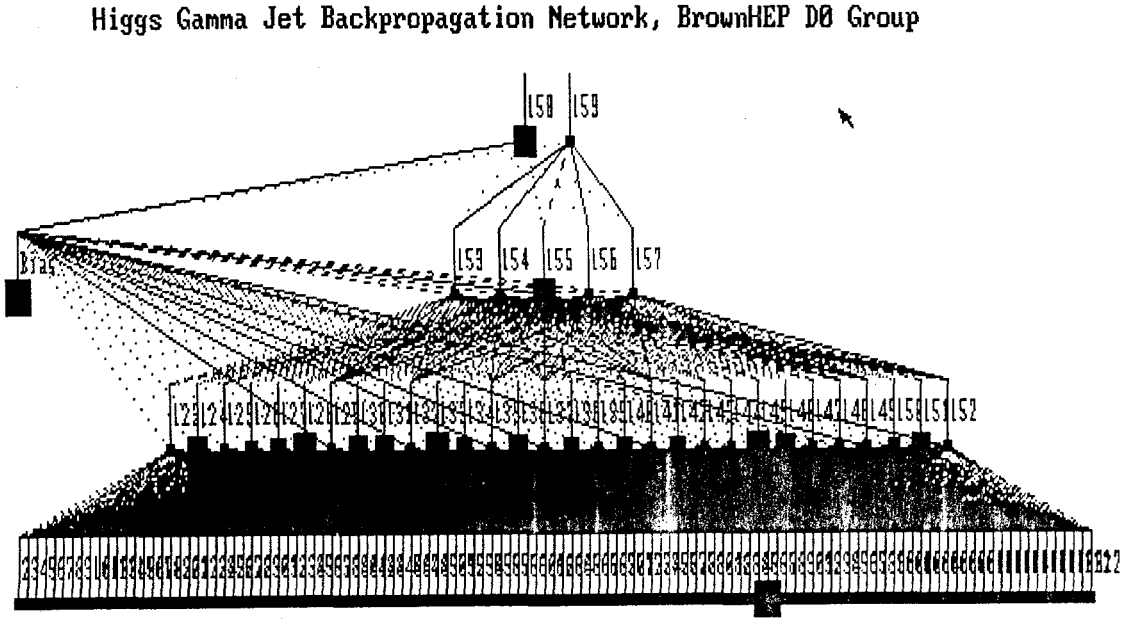


Fig. 9 Response of the trained network to a typical Higgs $\rightarrow \gamma\gamma$ shower.

The neural network shown above has 4 layers: an input layer of 121 processing elements (the 11 x 11 calorimeter energy deposits), a hidden layer of 30 PEs, a second hidden layer of 5 PEs, and an output layer of 2 PEs — which during training were presented with the desired output of (1,0) for a Higgs-derived photon shower and (0,1) for a background shower. Figures 10 and 11 show the results on test data for the output of the first PE in layer 4, after training on an admixture of Higgs and background events.

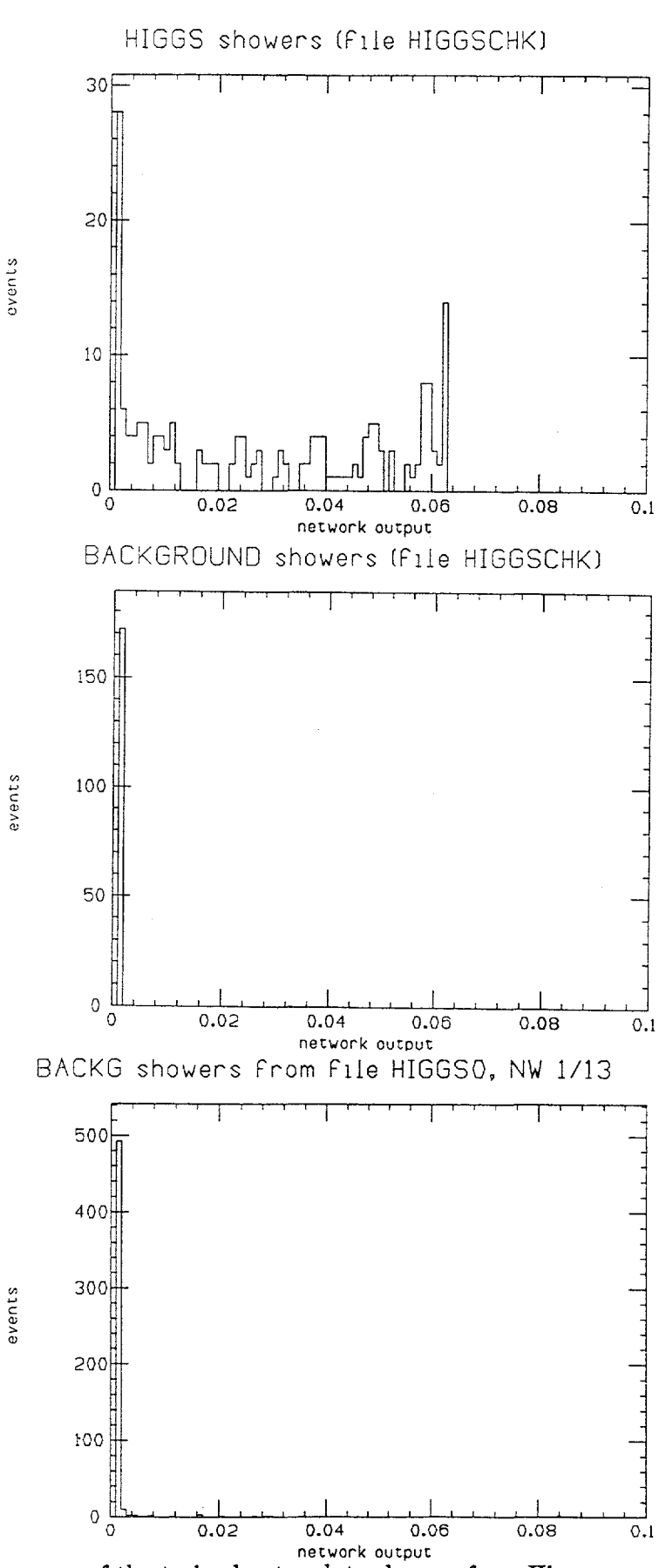


Fig. 10 Response of the trained network to showers from Higgs $\rightarrow \gamma\gamma$ photons and to backgrounds. The network was trained with equal numbers of events containing Higgs showers and background events.

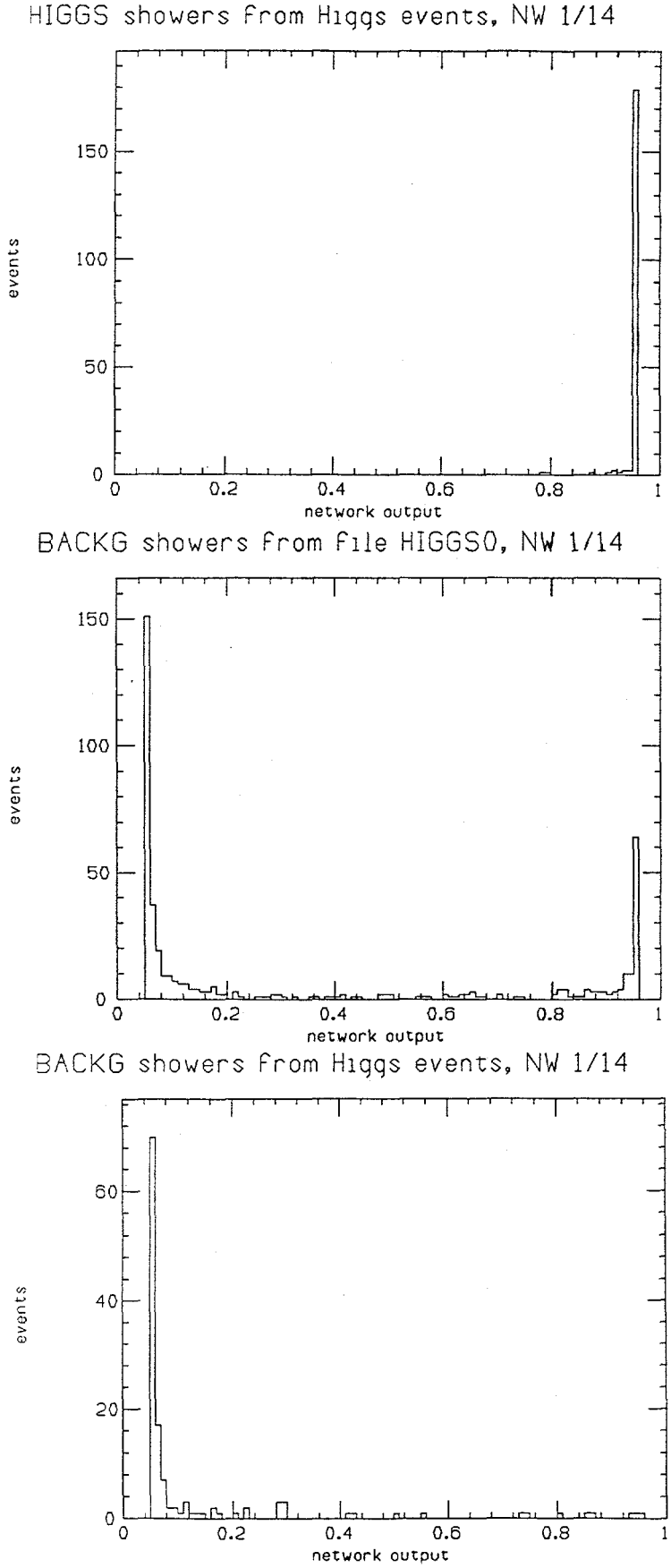


Fig. 11 Response of the trained network to showers from Higgs $\rightarrow \gamma\gamma$ photons and to backgrounds. The network was trained with equal numbers of Higgs showers and background showers.

The network clearly shows an ability to distinguish the showers. Its performance with this first test structure seems roughly comparable to that obtained in Reference 10, but a direct comparison can't be made at present because the algorithm used before selected events (first requiring two high P_t showers) while we have studied recognition of the individual showers. We are continuing these studies with the goal of optimizing the network architecture and deriving performance numbers that we can relate to those from alternate recognition schemes.

You must have noticed the difference between Figures 10 and 11: although in both figures the Higgs data is clustered on the right, and the background, to the left, there is a marked change both in the range of the outputs (note the scale) and on the degree of separation of response to the two shower types. This striking change is due to the use of a different training sample for the two networks. The first set, in Figure 10, used a combination of Higgs and 2-jet events in equal proportions; but, since each Higgs event contained 2 or 3 background showers in addition to the two photon showers, the Higgs-related vs. backgound ratio was more like 2/1. The second set of results, in Figure 11, come from a network trained on data with roughly equal numbers of Higgs showers and background, and shows an improved ability to distinguish them. Why are the details of training important? That should not be a surprise — the neural network can only learn what it has been shown; and if there are particular characteristics it must recognise, the training must include them in sufficient amount. Are we somehow cheating, by tuning the training procedure to obtain specific results? I don't believe so, any more than one is cheating when one sets a discriminator threshold after looking at a range of pulses on a scope. After all, the network after training is then frozen, and can be replaced by a few matrices and imbedded in hardware. We should use any methods possible to perfect those weights.

CONCLUSIONS

Actually, as promised at the beginning of the talk, there are no conclusions yet. Nonetheless, I can make a few comments based on the experience with neural networks described here. We have demonstrated, to ourselves at least, that neural networks definitely can learn to recognize features of high energy physics data. We are sufficiently encouraged in fact that we recognize the need to be more serious. To convince others we need to make hardnosed comparisons of the performance of the neural network with standard algorithms. We need to become more professional in applying this technology, exploring different network architectures and training sequences, and perhaps even obtaining some understanding about the operation of these systems. Finally, we are excited about the possibilities of going beyond the simulator and encoding a network into D0, using the special funtion ports of the MicroVAXes' multiport memory described above. For D0 as for the SSC, efficient algorithms are crucially important; perhaps for the SSC detectors and for D0, neural networks will have a role.

REFERENCES

[1]B. Denby, Neural Networks and Cellular Automata in Experimental High Energy Physics, Comp. Phys. Commun. **49** (1988) p. 429 - 448, and presentation, this workshop.

[2]D. E. Rumelhart et. al., Parallel Distributed Processing (MIT Press, Cambridge, 1986).

[3] D. Cutts et al., "The MicroVAX-Based Data Acquisition System for D0", Conference on Real-time Computer Applications in Nuclear, Particle and Plasma Physics, IEEE Transactions on Nuclear Science, Vol. **NS-34** (August 1987).

[4] D. Cutts et al., "Data Acquisition Hardware for the D0 MicroVAX Farm", International Conference on the Impact of Digital Microprocessors on Particle Physics, Trieste, Italy (World Scientific, Singapore, 1988).

[5] D. Cutts et al., "A Microprocessor Farm Architecture for High Speed Data Acquisition and Analyses", IEEE Nuclear Science Symposium (November 1988) to be published.

[6]Zeller Reserach Ltd., 8 Rushton Drive, Cranston, RI 02905. See reference 5 for details on the multiport memory.

[7]Available from NeuralWare Inc., 103 Buckskin Court, Sewickley, PA 15143. All the networks studied here were generated by NeuralWorks Professional II (tm) serial number NW2B01-11343, Copyright (c) 1987, 1988 by NeuralWare, Inc.

[8]Ahrens, et al., "A Massive, Fine-Grained Detector for the Elastic Reactions Induced by Neutrinos in the GeV Region", Nuclear Instruments and Methods A 254, 515 (1987).

[9]Abe et al., "Measurement of the Weak Neural Current Coupling Constants of the Electron and Limits on the Electromagnetic Properties of the Muon Neutrino", Phys. Rev. Lett. **58**, 636 (1987).

[10]Barter et al., Proceedings of the Summer Study on High Energy Physics in the 1990's (Snowmass, 1988), to be published.

SSC/BCD Data Acquisition System Proposal

E. Barsotti, M. Bowden & C. Swoboda
Fermilab
Batavia, Illinois

(Presented at the Toronto January 1989 SSC Workshop on Triggering and Data Acquisition for Experiments at the Superconducting Super Collider)

Data acquisition systems for colliding experiments in the United States are limited in bandwidth to small numbers of MegaBytes of data into online processing farms (i.e., filters). For example, the following summarizes data rates for systems beginning with the Colliding Detector at Fermilab (CDF) through the SSC (MB/s equals MegaBytes/second; GB/s equals GigaBytes/second):

	Rates Into Processing Farm		
Detector	Current	Initial Design Goal	Processing Farm Use
CDF	1-2 MB/s	10 MB/s	Filter (i.e., online reconstruction)
D0	---	40 MB/s	Filter
BCD	---	50 GB/s	Trigger & Filter
SSC	---	100 GB/s	Filter

Note the very large increase in data rate requirements between D0 and BCD. This is due in part to the 2.5 MHz BCD interaction rate and also to the fact that special and hard to maintain combinatorial logic is not being proposed for the usual '2nd level trigger system'. Because of the increasing availability of very high-performance industry-supported processors, it is being proposed that the BCD do only prompt triggers in the order of a microsecond with 'special' hardware. All other triggers and all filters would be done in online processor farms. Doing this will provide a more flexible, more powerful and much simpler data acquisition system. However, this architecture does inherently require higher data rates into processor farms. The SSC with its 100 MHz interaction rate and even with first and second level triggers systems near the detector still requires data rates from the detector of 100 GB/s. The triggers might reduce the event rate to an online filter (i.e., processing farm) to between 10 and 100 Khz.

The proposed new data acquisition system architecture takes event fragments off a detector over fiber optics and to a parallel event building switch. ~~as shown in Figure 1.~~ The parallel event building switch concept, taken from the telephone communications industry, along with expected technology improvements in fiber-optic data transmission speeds over the next few years, should allow data acquisition system rates to increase dramatically and exceed those rates needed for the SSC. The report briefly describes the switch architecture and fiber optics for a SSC data acquisition system.

The technology dealing with sending data over fiber optic cable (i.e., light sources and receivers, fiber optic cable and support integrated circuits) is one of the fastest growing areas in the world today. Today an optical link transmitter/receiver pair transmitting at 125 Megabits per second over 500 meters costs $200. Five years ago such a link was in its experimental stages, was used by only a select few, and cost

ten thousand dollars. Today, a Gigabit per second link costs ten thousand dollars and is in its experimental stages. By SSC production quantity ordering time, this link should be very cost effective allowing the event-building switch data acquisition system to exceed needed SSC data rates.

The switch architecture data acquisition system uses fiber optics extensively to transmit digital data from the detector. Thus, the three new ideas in the data acquisition system are:

1. Fiber optics for digital signal data transmission.
2. Barrel Switch Event Builder.
3. Industry-supported processors for triggers and filters.

The parallel event building switch will also be capable of modular expansion of its numbers of inputs and outputs and thus be capable of adapting to future higher fiber optic data transmission speeds and higher experiment data rate requirements with no fundamental changes to the data acquisition system architecture.

Switch Architecture Data Acquisition System Description

The following along with Figure 1 briefly describes the data acquisition system architecture which uses the Parallel Event Builder Switch. Data is clocked out from the detectors at the interaction rate and stored in local buffers pending a decision by simple first level trigger hardware or in the case of the SSC pending a decision by both the first and second level triggers. An accept from the trigger(s) allows the data to be gathered into a local output stream, otherwise it falls out the end of the buffers. The detector is partitioned so that the data rate for each stream is well matched to the capacity of a standard data channel. FIFO buffers keep the data rate entering the Parallel Event Builder constant. For colliding beam detectors, where data is produced continuously, these buffers need not be large. For fixed-target experiments, where the duty cycle of the beam is low, large buffers can be used to maintain the data rate between beam cycles. The cost of memory for very large buffers should be weighed against the cost of an acquisition system with higher processing throughput.

The core of this parallel system is the Parallel Event Builder Switch. A typical example of its implementation uses an N input, N output barrel shifter with a very limited number of input to output combinations, namely N, as opposed to a crossbar switch which is capable of N! possible switch configurations. Because all of the Parallel Event Builder's interconnect possibilities leave N connections intact, the output bandwidth is always equal to the input bandwidth. Let us examine a simple case, that of a four input, four output switch. Consider a detector that produces four data streams, A, B, C and D, each with a series of equal size event fragments.

As illustrated in Figure 2a, data passes through the system in fixed-length packets with each input channel delayed by one packet time slot relative to the adjacent channel. With the switch control set to 0, the first data packet (1A)

passes directly through the switch along with three empty packets. The switch control is then incremented by one (Figure 2b) and packets 1B and 2A are transmitted. During the next time slot (Figure 2c), packets 1C, 2B and 3A are transmitted. After one rotation of the switch control, the system reaches a steady-state condition as shown in Figures 2e and 2f. Parallel event fragments are converted to assembled event streams with no loss of bandwidth.

The same principle can be extended to any N X N switch. After one rotation of the switch control, all the fragments from event E are delivered to a single output port (E modulo N).

Figure 2 is an idealized version of the barrel shifter switch operation. Because event fragments are not all equal length, segregating event data by packet in a real system would not be very efficient. Some provision must be made which allows a single event fragment to span several packets or allows fragments of several events to occupy a single packet.

To eliminate any such correlation between event and packet boundaries, the hardware actually maintains N "logical" FIFO buffers for each input and output channel instead of the single buffer shown. Data placed in one logical buffer of a transmitter appears in the corresponding logical buffer of a receiver, independent of event boundaries. Viewed externally, the system appears as shown in Figure 3. As in a standard telephone switching system, the N^2 exhaustive interconnects are actually implemented using 2N serial channels and a time-division multiplexed switch. But whereas a telephone switching system must cope with a completely random bi-directional traffic pattern, the case for data acquisition is far less complicated. A simple barrel shifter replaces the crossbar network and control is reduced to a single packet clock.

Data appearing in the receiver is essentially a memory image of the complete event at the detector. Some minor formatting may take place in the receiver or as a preprocessing step in the higher level trigger software.

The Barrel Shifting Event Builder Switch performs the task of event building with no significant loss in bandwidth. N streams of event fragments are "switched" into N streams of assembled events. In practice some bandwidth will be lost to a non-zero interpacket switching time and any long-term fluctuations in average event fragment size. The first effect can be minimized by sending several events' worth of data between switching, the second by increased buffering in the transmitters and receivers.

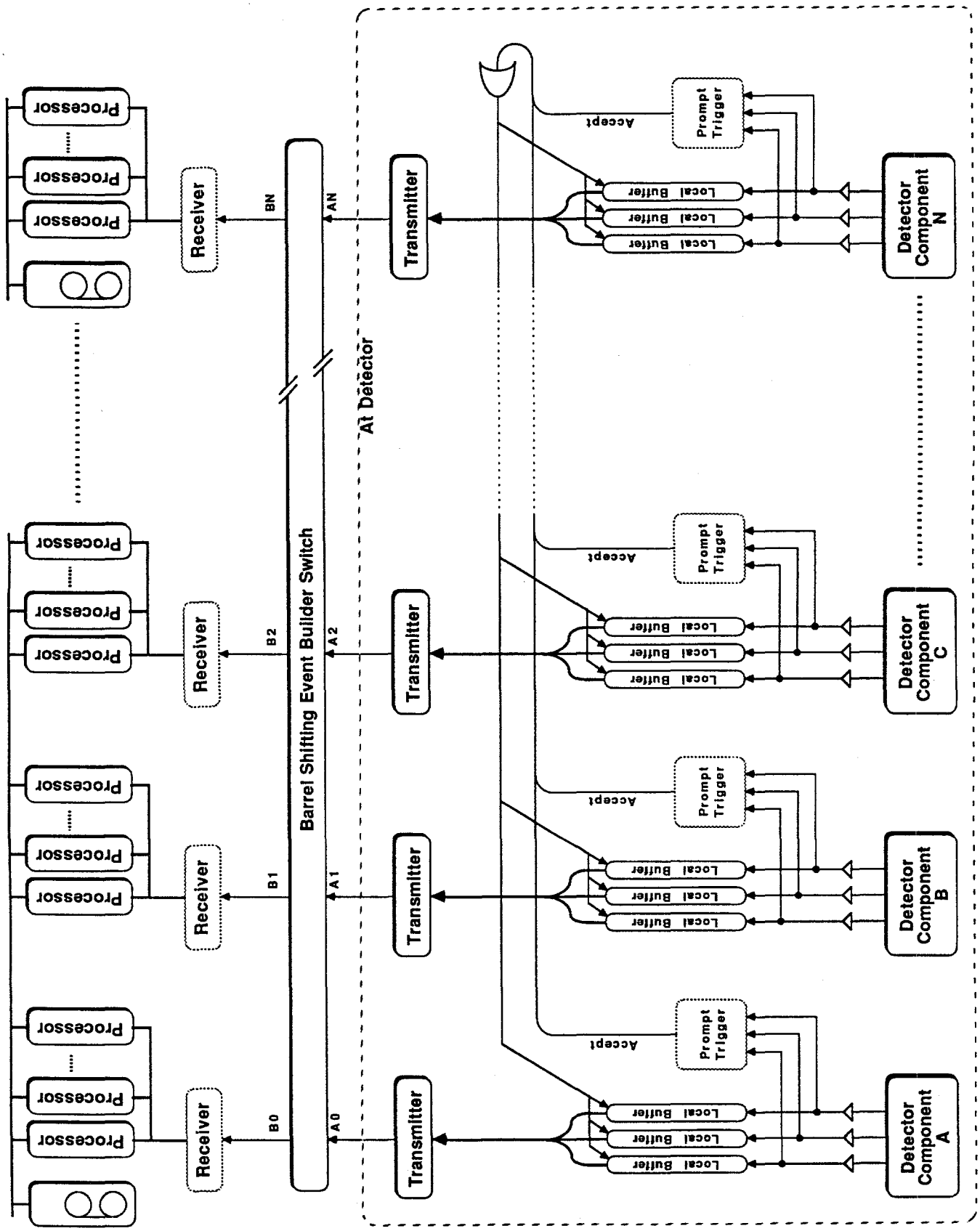
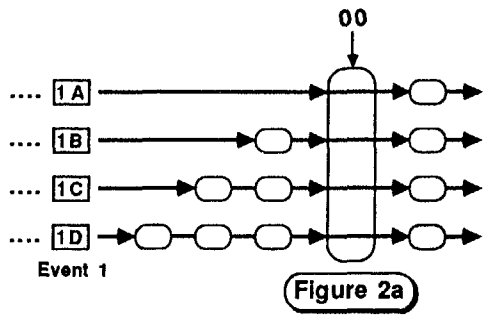
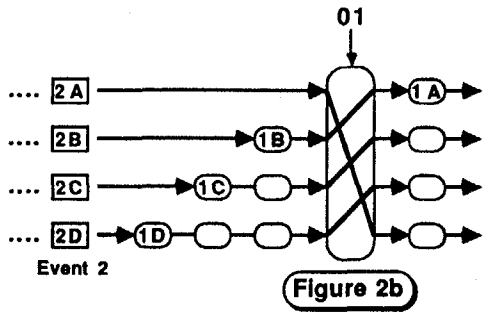
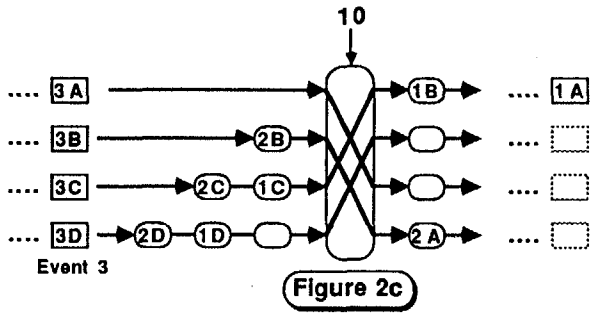
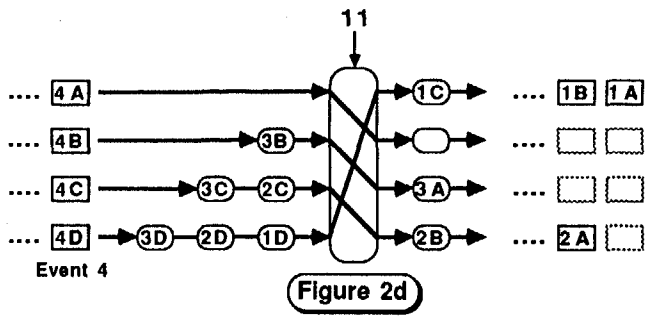


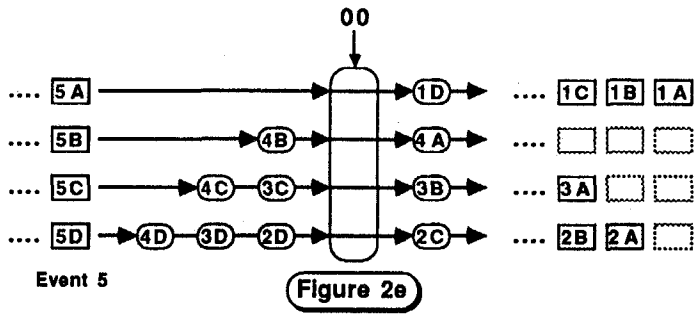
Figure 1 System Architecture

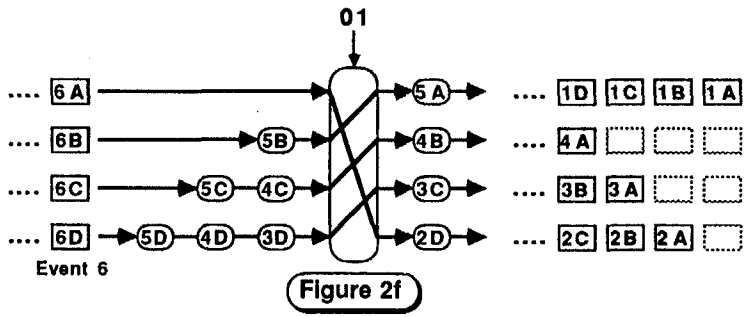












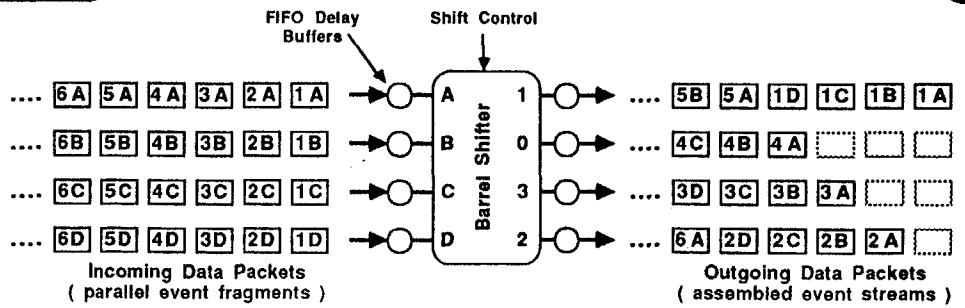


Figure 2 High Bandwidth Event Builder Using Simple TDM Switch

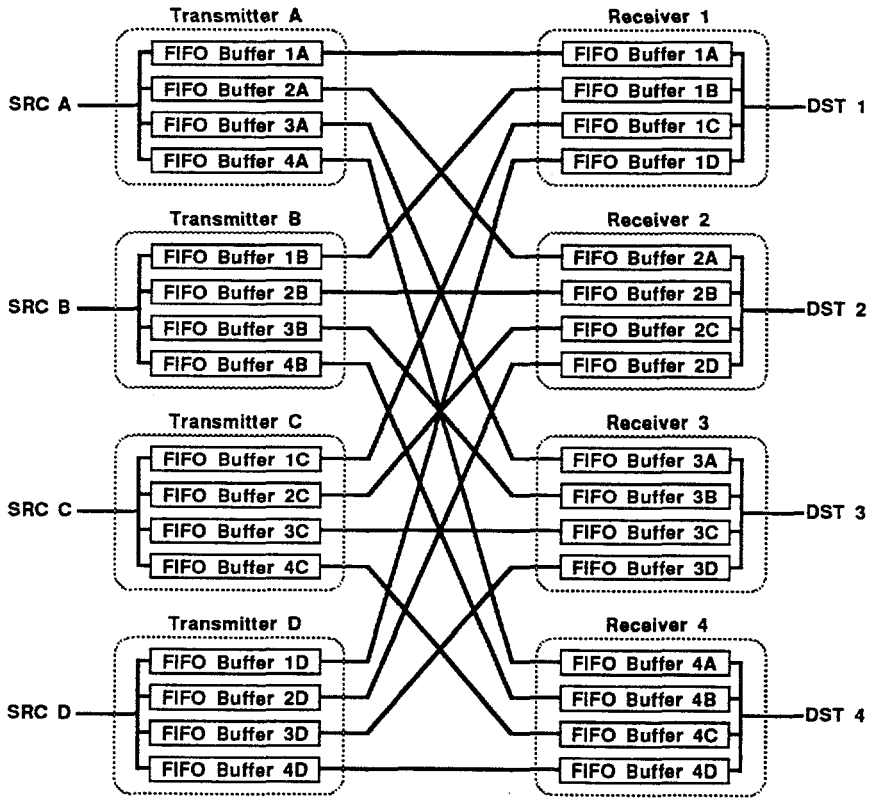


Figure 3
"Logical" Switch Operation

Microprocessors and Other Processors for Triggering and Filtering at the SSC

Irwin Gaines
Fermilab Advanced Computer Program, Batavia, IL 60510

INTRODUCTION

The rapid increase in processing power available in commercial integrated circuits presents the high energy physics community with important opportunities for SSC era data acquisition systems. The processors that will be available in the late 90's will allow enormous amounts of computing power to be utilized on-line and will permit commercial high level language programmable devices to be used for tasks previously performed by home-brew, hard-wired, or microcoded devices.

I will describe this processor revolution, in particular with respect to the new RISC (reduced instruction set computer) microprocessors now becoming available. These processors are already commercially available with processing power of 20 VAX 11/780 equivalents per chip, and a number of different manufacturers expect chips of 100 VAX power by the early 1990's. I will also discuss the plans the Fermilab ACP group has to exploit one such RISC chip for both off-line and on-line use. Finally, I will mention digital signal processors (DSPs) and other more specialized chips that offer even greater amounts of processing power with only slightly less convenience.

RISC PROCESSORS

Reduced Instruction Set (RISC) microprocessors have gone from an academic research project to practically a computing industry standard in a short period of time. Every leading semiconductor manufacturer and computer vendor have RISC projects underway. The current generation of RISC processors has already surpassed the more common CISC (Complex Instruction Set) architecture (typified by the Motorola 68032 and the Intel 80386) in performance, and shows signs of surpassing mainframe performance as well. Figure 1, showing one vendor's projections of the computing power available in different architectures, illustrates these ideas.

What is RISC, and why do these processors have such high performance? The principle of RISC is to keep the instruction set of the processor as simple as possible, so that all instructions can be executed in a single clock cycle. This is in contrast to the prevailing design philosophy of the 60's and 70's, where instruction sets were filled with an enormous variety of instruction types and addressing modes, supposedly to make it easier to write

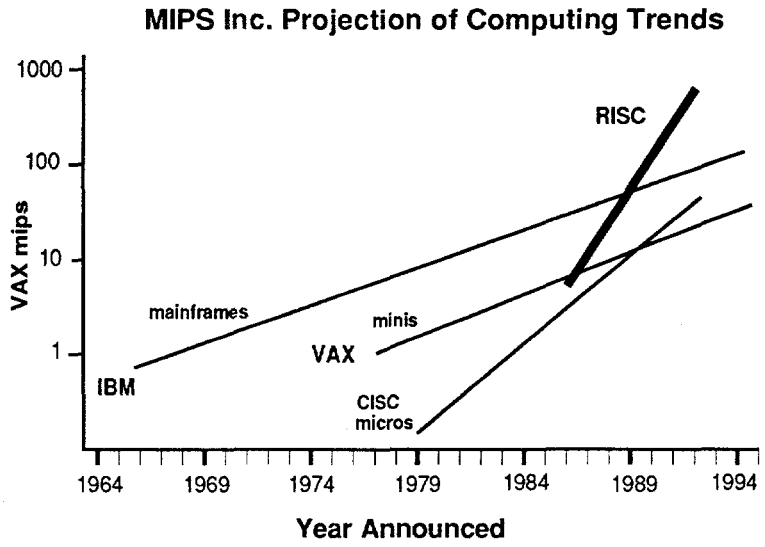


Figure 1. Processing power available in different architectures.

compilers for high level languages. However, study of the instructions used by compilers indicated that only a small fraction of the complex instruction sets were frequently used, and that the complexity led to a large loss in performance, even when performing the simplest instructions. Eliminating many of these instructions results in an architectural simplicity allowing the remainder of instructions to be executed in one processor clock cycle. Infrequently performed complex operations are done in software rather than in hardware, so that there will be no performance penalty for the vast majority of simpler operations.

More specifically, the time to perform any given computing task depends on the product of three factors: the number of instruction needed to do the task; the number of clock cycles needed for each instruction; and the amount of real time required for each clock cycle. RISC processors win by giving an enormous reduction in the 2nd component (cycles/instruction)—while a VAX 11/780 averages 10.6 cycles per instruction and the Motorola 68020 averages 6.3 cycles, a modern RISC processor like the MIPS R3000 requires only 1.25 cycles per instruction (and future RISC processors expect to lower this average to 1.0 or below).[1] Moreover, the simplicity of the architecture allows the RISC processors to run at higher clock speeds than CISCs (as well as allowing implementation in faster technologies like ECL or GaAs that are not suitable for CISC architectures). This gain is accompanied by an increase in the required number of instructions, but only by 20-50%, leaving the RISCs with a large overall performance gain.

A variety of architectural techniques allow RISC processors to achieve their goal of single cycle instruction execution. Typically there is a relatively small number of instructions and addressing modes, and a fixed instruction format. The RISC designs are often a load/store architecture, with large register sets and no memory-to-memory instructions. Control logic is usually hard-wired, with none of the microcoded control typical of

minicomputers and CISC processors. Much more of a burden is put on the compilers, with sophisticated optimizations required to achieve full performance in high level languages. (This is why the currently leading RISC implementations are those where significant effort was put into compilers right from the very beginning). In a sense, the RISC philosophy shares the complexity of processor design between the chip architects and the software writers, rather than putting all the burden on the hardware design as in a CISC processor.

The advantages of RISC architecture are by now widely recognized, and there are a large number of RISC processors now commercially available. These include the MIPS R2000 and R3000, the SUN SPARC, the Motorola 88000, the AMD 29000, the Intergraph (formerly Fairchild) Clipper, and the Intel 80960. Moreover there are proprietary RISC chips in use in systems from IBM (the RT personal computer), Apollo (the PRISM family of workstations), and Hewlett Packard. Table 1 summarizes some of the features and measured performance for the current generation of RISC processors.

Furthermore, all of the leading RISC manufacturers have announced plans for higher speed versions of their chips that will meet the performance projections shown in figure 1. To cite some examples:

1) The Intergraph Clipper is available in 50 MHz (14 MIPS) versions now, with 20 MIPS expected in March 1989 and 60 MIPS ECL versions in 1990;
2) Data General is designing a 100 MIPS ECL 5 chip version of the Motorola 88000 (instruction processor, memory management unit, cache controller, system controller, and system bus interface) expected by 1991;
3) Sun has licensed the SPARC technology to LSI Logic, Fujitsu, Bipolar Integrated Technology, and Cypress Semiconductor. All are working on higher speed implementations, with Fujitsu having 25 MHz (15 MIPS) parts available now, 33 MHz in 1989 and 40 MIPS in 1990, LSI with 40-50 MIPS in biCMOS in 1990, and BIT with 40 MIPS ECL in 1989;
4) MIPS has licensed the R3000 technology to LSI Logic, Integrated Device Technology, Performance Semiconductor, Siemens, and NEC. 33 MHz (25 MIPS) parts are available now, with IDT, for example, projecting 40 MHz in 1990, 60 MHz in 91 and 100 MHz by 92.

Clearly it is not overly optimistic to expect individual processors with between 50 and 100 VAX equivalents in performance well before the turn-on of the SSC.

Finally, the simplicity of the RISC designs has led to successful implementations in Gallium Arsenide, where limitations on the number of gates has prevented CISC implementations.[2] Two examples (each of which has the goal of building a 200 MHz or 150 MIPS processor by 1992) are:

1) a TI/CDC collaboration (CDC did the chip architecture while TI did the GaAs implementation). This chip has 12,895 gates, 6 pipeline stages, has already run at 68 MHz, and is expected to draw 1W at 200 MHz; and

Table 1. RISC Chips Features Comparison

	R3000	SPARC	CLIPPER	29000	88000
VLSI Chip Count CPU + FPU + MMU + Cache Control	2 CPU/MMU/ Cache + FPU	6 CPU + MMU + Cache + FPU(3 chips)	3 CPU/FPU + ICache/MMU + DCache/MMU	3 CPU/MMU + FPU + Cache Control	3 CPU/FPU + 2 Cache/ MMU/RAM
Bus Structure	"Harvard" 2 Address/Data (MUXed)	von Neumman 1 Address/Data	"Harvard" 2 Address/Data (MUXed)	"Harvard" 2 Address/Data (MUXed)	"Harvard" 2 Address/Data
Cache Support	64K ICache 64K DCache	No Direct Cache Support	8K ICache 8K DCache	No Direct Cache Support	Special Cache RAM w/MMU
MMU Support	Onchip, 64 TLB CAM	No Direct MMU Support	Onchip, 2 x 64 TLB	Onchip 64 TLB	On Cache Chips
Registers	32 General Purpose 16 x 64-bit Floating Point	128 8 x 32 Windows	64 Gerneral Purpose 8 Floating Point	192 General Purpose 3 Floating Point	32 General Purpose/ Floating Point
Clock Rate	25 Mhz	16 Mhz	33 Mhz	25 Mhz	20 Mhz
Measured Performance	**20 VaxMips**	**10 VaxMips**	**5 VaxMips**	**14 - 17 VaxMips**	**14 - 18 VaxMips**

2) an MDAC (McDonnell Douglas Astronautics Corp.) chip, with 23,178 transistors, a 5 stage pipeline, working 60 MHz versions, and 4-6W expected at 200 MHz.

Both of these chips will require GaAs cache memory to allow 1 memory access every 5 nanosecond cycle, and will require extremely sophisticated compilers to keep pipelines full even in the presence of branches. Most likely, the cost of the chips will prevent them from being in widespread use, but they will be available, for example, for on-line applications requiring the utmost in processing power from a single chip.

RISC AND THE ACP

As an example of what this new RISC technology makes possible for both off-line and on-line applications, I will briefly describe the Fermilab Advanced Computer Program (ACP) group's Second Generation Multiprocessor Project.[3] Their original parallel processing system was based on the Motorola 68020, with several systems of more than 100 processors in use. The Second Generation System uses the MIPS R3000 RISC chip set to provide an increase in processing power per board of more than a factor of 20.

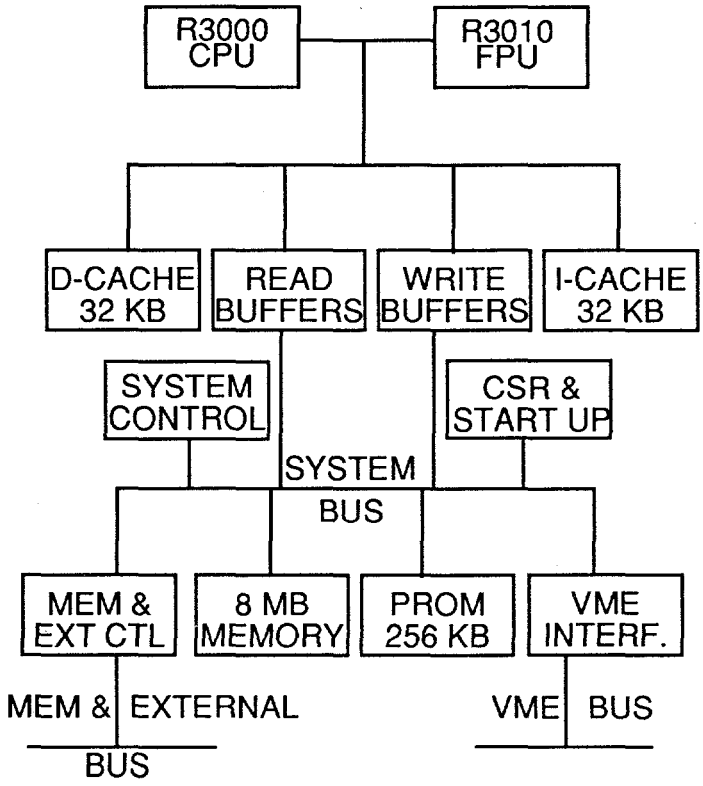


Figure 2. The ACP MIPS Processor Block Diagram.

The MIPS chip set was chosen after benchmarking several potential RISC processors. The R3000 was selected primarily because of its superb FORTRAN compiler, which now offers full VMS FORTRAN compatibility and highly sophisticated optimization. The MIPS compiler is an in-house product, and the compiler writers were involved even in the architectural design of the chip, which has given MIPS a lead over other RISC vendors who have up till now relied on outside third parties to provide compilers. The R3000 performs at 15 VAX 11/780 equivalents on a variety of high energy physics reconstruction programs (or roughly 20 times the original 68020's). Note that DEC has also selected the MIPS chips for use in their new line of high performance UNIX workstations (the DECstation 3100).

A VME processor board using the R3000 set has been designed (see figure 2 for a block diagram of the processor board). The board features a 25 MHz R3000 CPU and 25 MHz R3010 Floating Point Unit, 32 KB each of both instruction and data cache, 8 MB of on-board memory with parity, 256 KB of EPROM, a serial port, a full VME master/slave interface, and a memory expansion interface allowing expansion up to 32 MB of memory. Prototypes of this board will be available this spring, and commercial availability should follow shortly.

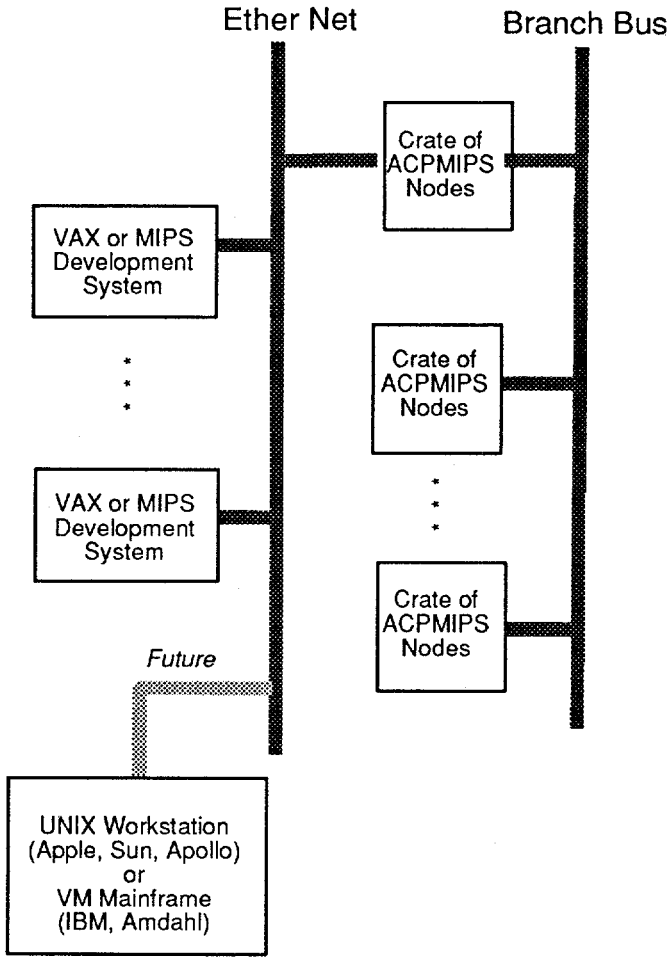


Figure 3. Second generation ACP multiprocessor systems allow any VMS or UNIX processor or workstation to take part.

The board runs the full UNIX operating system, including support for virtual memory, disk and tape I/O, interprocess communication and multi-tasking. A full suite of program development tools (compilers, linkers, librarian, and debugger) also runs on the board. Besides this vastly improved programming environment, the existence of a full-blown operating system also allows more complicated forms of multiprocessing than was allowed in first generation ACP systems, where the processors were all slaves to a dedicated VAX host.

The software support is thus provided in a layered fashion.[4] Underneath is UNIX, a commercial standard operating system. Next are networking protocols, again using the commercial standards of NFS and TCP/IP. Finally, a set of higher level service routines is provided to:

1) transfer blocks of data between processes;
2) call a remote subroutine on another process;
3) send a datagram message to another process; or
4) place and remove processes to and from queues.

Most users will only use these service routines, remaining ignorant of any of the operating system details, but more sophisticated users will also have the full resources of UNIX and the networking protocols available. This reliance on standards also allows any other processors or workstations running UNIX or VMS to be included as full members of a second generation mutiprocessing system, providing graphics or other capabilities not present in the VME processors, as shown in figure 3.

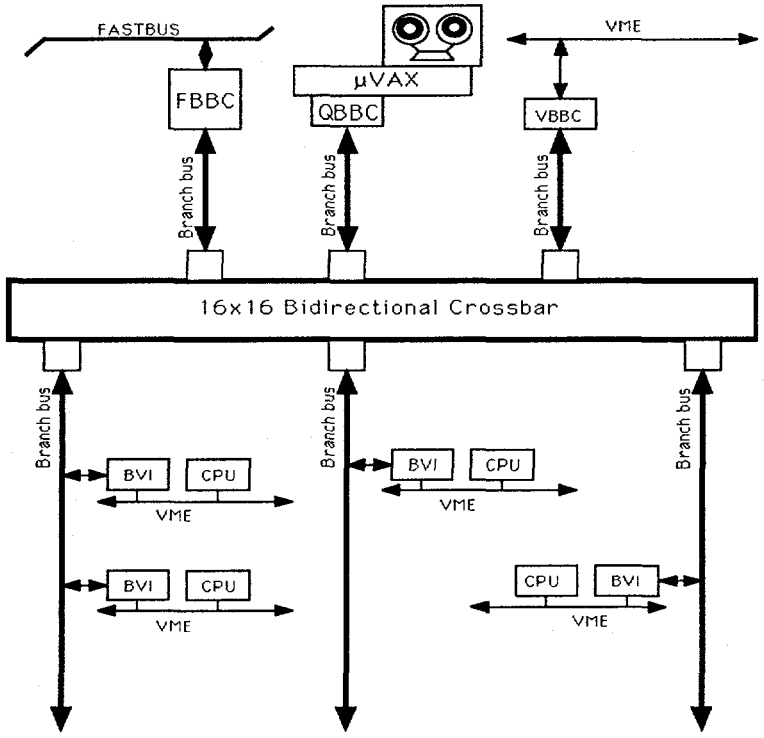


Figure 4. The bus switch used to interconnect data acquistion buses with VME processor farms.

On-line use of these powerful processors is aided by the existence of several interface modules. The ACP Branch Bus (a high-speed parallel data bus allowing transfers at 20 MB/sec) permits VME and other data acquisition buses to be interconnected. FASTBUS (FBBC), VME (VBBC) and Q-Bus (QBBC) interfaces to the Branch Bus exist, as does the Bus Switch, a full 16x16 crossbar switch allowing arbitrary interconnection of Branch Buses. These modules can be combined to provide high performance connections between DA systems and farms of processors, as shown in figure 4. In addition, the memory expansion (XBus) bus on the MIPS processor boards can be used to provide direct access from a DA bus into processor memory (see figure 5), much as the D0 data acquisition system does with microVAXes.

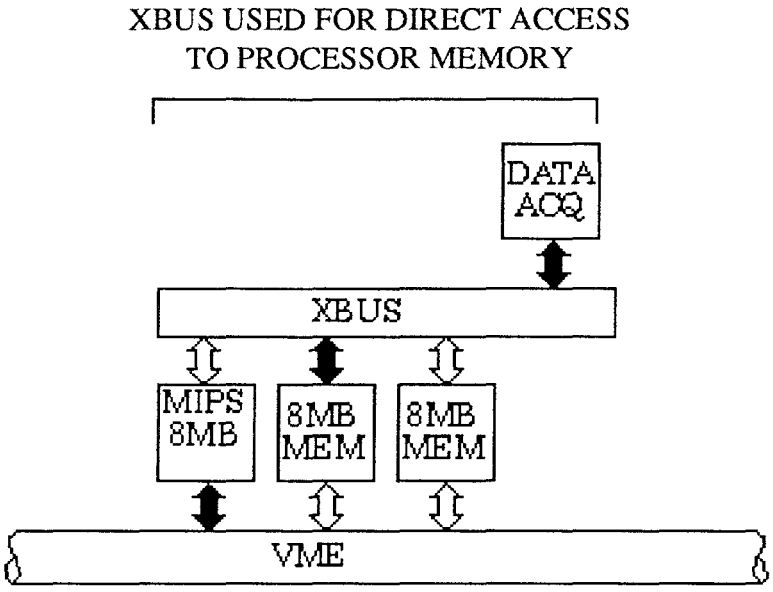


Figure 5. The memory extension bus (XBus) allows direct access into the memory of the ACP MIPS processor board.

DSPS AND OTHER PROCESSORS

Finally, mention should be made of digital signal processors (DSPs). These chips have traditionally only been used for special purpose applications such as dedicated trigger processors, and have not been suitable for more general purpose use due to limited memory space, no floating point, and lack of high level languages and good program development tools.

The current generation of DSPs remedies most of these deficiencies. These new chips typically have full IEEE floating point, large memory address spaces, and speeds of up to 100 MFLOPs. They are supported by operating systems offering high level languages and good program development tools. Examples of such chips[5] are:

1) the Texas Instruments TMS320C30, TI's third generation DSP, now available in a 16 MHz version offering 16 MIPs and 33 MFLOPs of performance. Despite the more than 700,000 transistors on the chip, it sells for under $100 in quantity. It also supports the SPOX operating system ,which provides math libraries, memory management and I/O services together with a small real-time kernel;
2) the Motorola DSP96002, available in 27 MHz offering 13.5 MIPs and 40.5 MFLOPs, with hardware hooks to support parallel processing applications;
3) the United Technologies UT69532 IQMAC (In-phase Quadrature Multiplier Accumulator), a highly pipelined chip with 3 ALUs and 2 multipliers, offering 100 MFLOPs of performance at 20 MHz.

While not yet appropriate for general purpose processing farms (the RISC processors have the important advantage of allowing code development on workstations or minicomputers using the same chip sets and software tools as the processor farms), these DSPs are highly suited for front end processing and triggering applications. The high level software tools should make these "special purpose" processors much more accessible to the average physicist than ever before, when their coding was normally restricted to a small group of experts.

As a last indication of things to come, I would like to mention Intel's N10 chip, details of which were revealed at the recent ISSCC (International Solid State Circuits Conference). This million transistor chip, originally planned as a co-processor for the forthcoming Intel 80486, instead combines a general purpose RISC core, a high-speed double precision floating point unit, and a 3-D graphics processor. A 50 MHz version of the chip allows 150 MIPs and 100 MFLOPs of performance. No pricing or availability is yet known for this superchip, but it indicates the kind of technology we can expect from industry to be used at the SSC. (On February 27, 1989, Intel formally announced availability of 33 MHz versions of this chip, now known as the Intel 80860, for under $1000.)

CONCLUSIONS

Industry will provide us with extremely powerful high level language processors for both filtering (level 3) and triggering (level 1-2) applications. We should resist as much as possible the temptation to build hard-wired, non-programmable or microcoded devices. Processor farms of 10^5 VAX equivalents are likely for both off-line and on-line applications: 100-1000 processor boards with 100-1000 VAX power on each board (possibly using multiple processors on each board).

The real challenge will not be in providing the processing power, but rather in insuring that the extraordinarily powerful arrays of processors are actually doing what we want them to do. It is not to early to start developing tools for program specification and

verification. Without these tools, we will be unable to enjoy the full benefits that processor technology can supply.

REFERENCES

1 R. Wilson, Computer Design, "RISC Architectures Take on Heavyweight Applications", p. 59-79, May 15, 1988.

2 B. Cushman,VLSI Systems Design,"GaAs Technology meets RISC Architectures", p. 68-77, September 1988.

3 T. Nash, H. Areti, R. Atac, J. Biel, A. Cook, J. Deppe, M. Edel, M. Fischler, I. Gaines, R. Hance, D. Husby, M. Isely, M. Miranda, E. Paiva, T. Pham, and T. Zmuda. "High Performance Parallel Computers for Science: New Developments at the Fermilab Advanced Computer Program", to be published in proc. of the Workshop on Computational Atomic and Nuclear Physics at One Gigaflop, Oak Ridge, TN, April 14-16, 1988.

4 J. Biel, "Second Generation ACP Multprocessor System: System Specification Document", July 25, 1988 and revisions; Fermilab Advanced Computer Program internal document; unpublished. Availability of ACP publications may be determined by reference to the file at the HEPNET location: fnacp::acpdoc_root:[docs]doclist.doc

5 IEEE Computer Society, Micro, Digital Signal Processors,vol. 8, no. 6, December 1988.

Summary Talk: Data Acquisition, Event Building, and On-Line Processing

Irwin Gaines
Fermilab Advanced Computer Program, Batavia, IL 60510

Our subgroup of working group 4 concerned itself with general architectural issues for SSC data acquisition. Fiber optic buses were described in two talks, and software and project management were discussed in a separate subgroup. These topics are covered in separate papers. A number of different issues were considered, with certain natural differences of opinion, and a list of projects needing further R&D is given at the end of this paper. The major conclusion, however, can be simply stated and was agreed to unanimously: that the SSC data acquisition systems can achieve much higher data rates than those assumed at previous workshops. On-line processor farms of 10^5 VAX equivalents (throughout this paper we will use VAX 11/780 units as a performance standard) are quite conceivable, and data recording rates of between 100 and 1000 Hz will be feasible.

ARCHITECTURE AND RATES

The architectural framework for the subgroup's discussions is shown in fig. 1. We assumed that data would be buffered on (or near) the detector in systems discussed by the front-end working group, and that a prompt trigger together with some data processing at the front end would reduce both the event rate and event size before the actual data acquisition system began to deal with the events. Our discussions were confined to the area below the dotted line in the figure, where a stream of digital data emerges from the front end systems.

The components of the DA system that we considered include the non-prompt triggers, data and control buses, event builders, high level language processing farms, and data recording ("tape"). Also, it is important to point out that when we give a rate capability for a particular component of the data acquisition system, that does not imply that we are required to run the system at that rate. In particular, it will be necessary to make choices, guided by physics, as to where the biggest payoff comes from making improvements in the rate capability. It might cost the same amount of dollars to double the "tape" writing speed as to increase the power in the processor farm by 20%, and we will need physics judgement to decide where to put our resources. Our discussions were primarily aimed at identifying the "brick walls" in the rate capabilities, beyond which we would require significantly new designs or unanticipated technological breakthroughs.

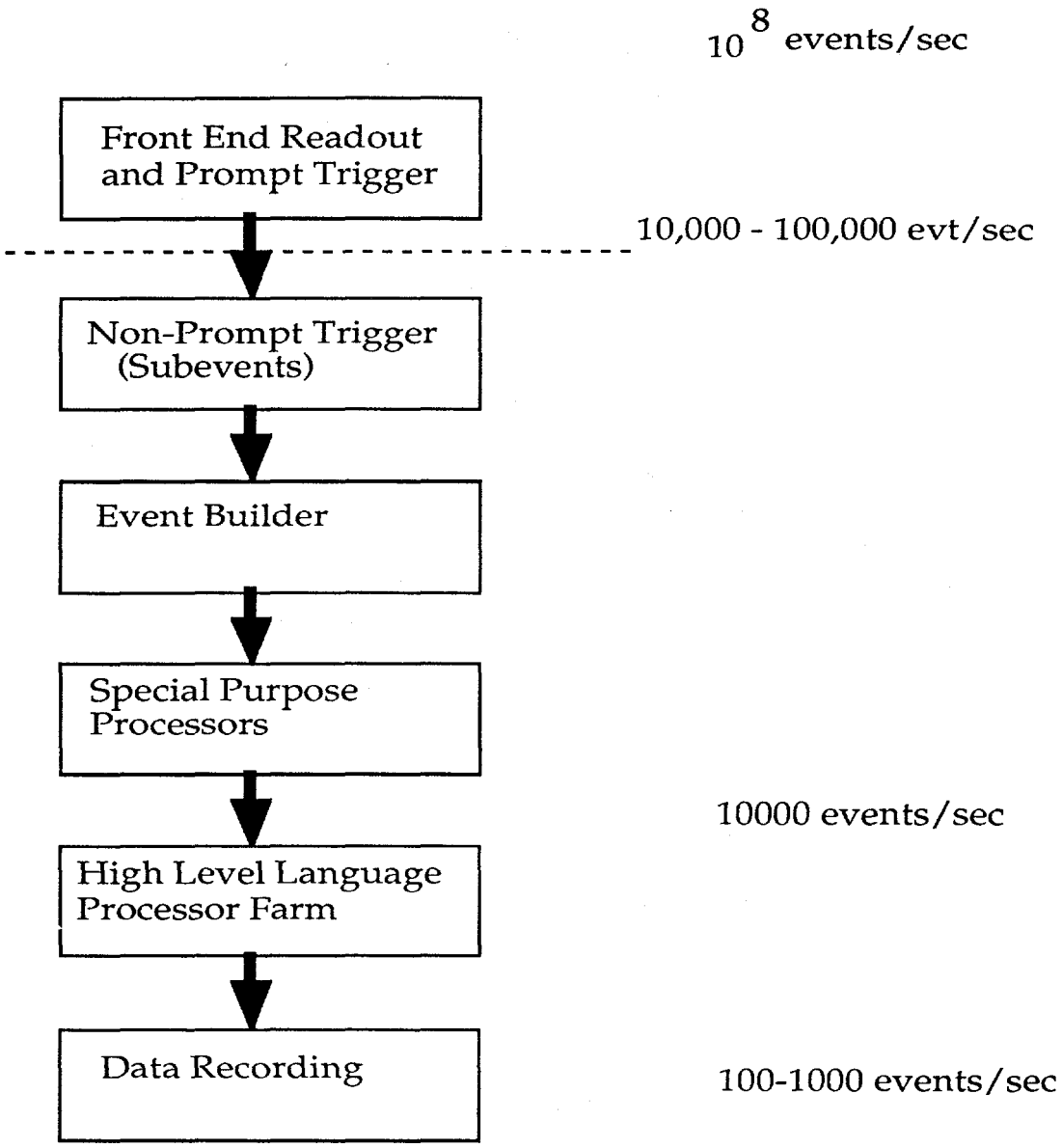


Figure 1. Block diagram of data acquisition architecture.

With that in mind, we can summarize the rate capabilities as shown in the figure. We can take between 10^4 and 10^5 events/sec out of the front-end, with an event size of between 10^5 and 10^6 bytes. Special purpose non-prompt triggers could deal with these rates. The high level language processor farm can deal with 10^4 events/sec. Data can be recorded at somewhere between 100 and 1000 Hz, depending on how much data reduction is done in the DA system. The remainder of this paper will justify these numbers and summarize our discussions on several architectural issues.

EVENT BUILDING

Discussion of event builders considered the difference between "classical" event builders and newer ideas of using a switching network as an event builder (see fig. 2). It was generally agreed that the traditional designs, where one (or a few) event builders act as a funnel through which all the data must pass, were unacceptable for SSC rates. Schemes based on switching networks, on the other hand, have the advantage of being scalable to match the required rates and of being able to keep all the input and output pipelines busy at the same time. Mark Bowden presented one design based on a barrel shifter; other more complicated designs are also possible. Note that events can be processed and rejected both in the input pipelines (where the trigger processors work on sub-events) and in the output pipelines (where the processors can work on the entire event).

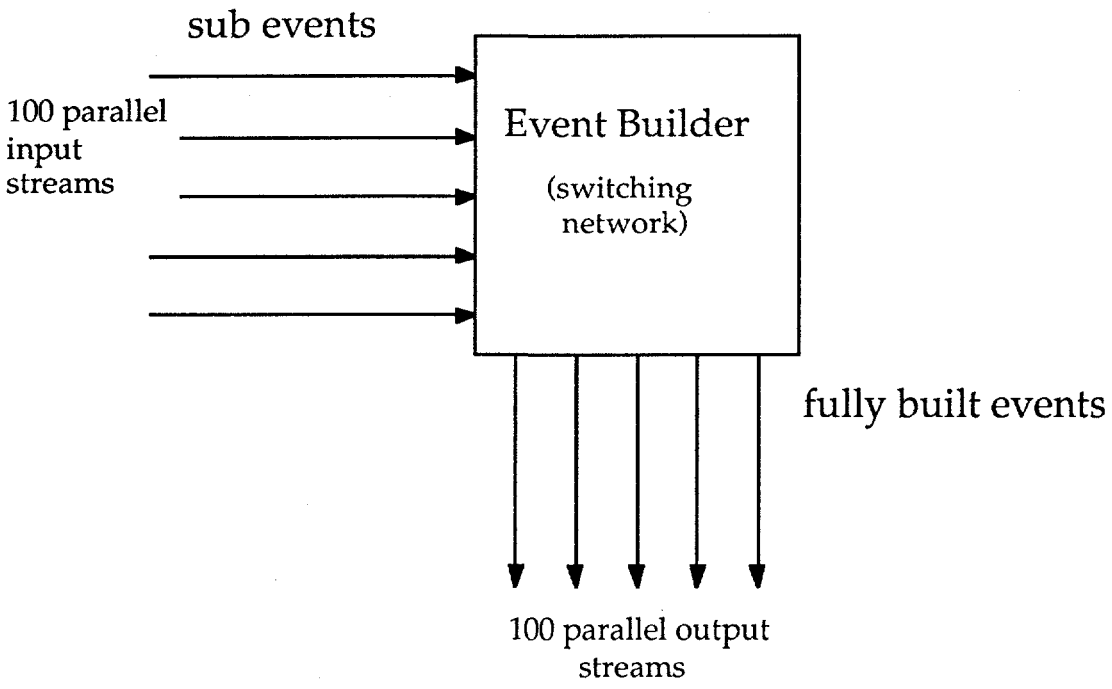


Figure 2. Event builder based on switching network.

At the highest rates contemplated, of 10^5 events/sec of size 1 MB each out of the front end, the input rate to the event builder will be 100 GBytes/sec. Assuming we have fiber optic data links capable of 1 GByte/sec, this would require a parallelism of 100 into and out of the event builder, which seems quite feasible.

TO BUS, OR NOT TO BUS

Discussion of buses centered not around which bus standard was more or less suitable to the SSC, but instead on whether or not we need a bus at all. There was general agreement on the need to separate out data and control paths, and on the need for some sort of diagnostic and control bus connecting most of the DA system. However, data transmission may likely be handled with a standard for a high-speed data link rather than a bus. Optical fiber technology appears capable of providing us with such a link that could work at rates of up to 1 GByte/sec. The separate control bus can run at much lower speeds.

DATA DRIVEN ARCHITECTURES

We discussed the relevance of data flow/data driven architectures to SSC DA systems. These ideas seem well matched to the needs for pipelines that are clearly evident in the front-end systems. Cited as potential advantages of such data driven architectures were simplicity, modularity, flexibility, and speed (since centralized control is unnecessary). Issues that were raised but not fully answered include the problems of decentralized control (maintaining synchronization across parallel pipelines, insuring no data loss between pipeline stages when the next stage is busy, etc.) and understanding how to propagate a fast reject signal through a pipeline.

SPECIAL VS. GENERAL PURPOSE PROCESSORS

Despite the enormous power that will be available in the form of general purpose processors, it was felt that there will continue to be a place for more specialized designs, for such tasks as data manipulations (pedestal suppression and rescaling) as well as segment and track finding and jet cluster finding. The potential utility of such special purpose devices is illustrated by two examples from CDF: cluster finding takes roughly a millisec per cluster on the VAX, but the special purpose hardware finds clusters at 200 nanosec per cluster, giving a performance of 5000 VAX equivalents; while a tracking task (finding all tracks with more than 3.5 GeV of transverse momentum) which takes over 100 millisec on the VAX is done in 2-4 microsec with limited resolution and 20-30 microsec at full resolution in the special purpose hardware, again giving a performance of better than 5000 VAX equivalents.

The input rate to the special purpose non-prompt triggers is expected to be between 10^4 and 10^5 events per second, depending on how much filtering is done by prompt triggers at the front end. The high-level language processor farm will be capable of accepting 10^4 events/sec. Thus, at the highest rates we will require at least a factor of 10 rejection from special purpose devices; while at the lowest rates we may simply perform data compression and preprocessing at this stage. Even at these lower rates, however, we may choose to reject events at this stage so as to put less of a burden on the processor farm.

Regardless of the amount of rejection needed from special purpose devices, as much use as possible should be made of commercial programmable devices, including DSPs and ASICs with RISC processor cores, to avoid the proliferation of processors that are only understood by a few experts.

HIGH LEVEL LANGUAGE PROCESSOR FARMS

The availability of powerful RISC processors allows us to consider farms of order 10^5 VAX equivalents. This will permit 10 VAX seconds of processing per event at an input rate of 10^4 events/sec, or more processing if there is additional reduction at the special purpose processor stage. Such a system could be constructed out of 500 processor boards, each with 4 50 MIP processors. The price of the processor boards should be under $5000 by the mid 1990's, making the total cost of the farm of order $2.5 million. Note that each processor board need only process 20 events/sec, so the data bandwidth needed into each board is only 20 MBytes/sec. This stage needs to provide a further event rejection of between 10 and 100 to match the "tape" writing capabilities.

EVENT COMPACTION

It would be useful if the size of the events could be substantially reduced before "tape" writing so as to reduce the required bandwidth for data recording. One might even hope to write out only data summary tapes (DSTs) rather than raw data given the large amounts of on-line processing power that will be available. Two factors mitigate against this.

First, it is unlikely that final calibration constants will be known to sufficient accuracy as the data is being taken to allow the raw data to be thrown away. At best, an automated production line could hope to determine constants within a few hours after the data is taken, allowing the DSTs to be produced shortly after the raw data is recorded. However, this will not relieve the DA system from the need for recording all the raw data, even if just temporarily.

More importantly, it appears that the event size on the DSTs is in fact larger than the size of the raw event! CDF, for example, has raw event sizes averaging 120 kBytes, while the DST event size is 250 kBytes. Only at the mini-DST stage is the event size reduced (to about 25kB/event). Thus, even if we had up-to-date calibrations it is unlikely that we can reduce the event size by additional on-line processing.

Nevertheless, the raw event size of 1 MByte at the SSC includes significant contributions from noise and from out-of-time events. On-line processing in the farm might be able to eliminate much of this non-event data, reducing the demands on the "tape" system. More detailed detector simulation (as well as greater understanding of the readout schemes) is required to know how much of a help this will be.

"TAPE' WRITING SPEEDS AND OFF-LINE ANALYSIS

Thus, we need to contemplate recording events of close to 1 MByte in size. Even so, it was generally agreed that the "tape" system can easily cope with 100 events/sec (100 MB/sec), with the possibilities of going another factor of 10 faster if necessary. Certainly we will be writing in parallel, with multiple streams of events of different trigger types. A parallel 100 MB/sec system could be built with today's technology for a cost of well under $1 million, and we can expect to do significantly better by the time the SSC turns on. It is still too early to decide the exact media we will record data on; high density tape cartridges, 8 mm video cartridges, and optical disks are all candidates. We will almost certainly need a "juke-box" like technology whatever the media is to minimize required user/operator intervention.

Before we consider recording 100-1000 events/sec, we should make sure that we will have off-line computing resources available to analyze the events. This is not a problem, however. A processor farm of the same scale as described above (10^5 VAX equivalents) can deliver 1000 VAX seconds per event at 100 Hz data taking rates or 100 VAX seconds at 1000 Hz, which should be sufficient.

OPEN QUESTIONS NEEDING R & D

Finally, we list some issues which were not resolved but were identified as needing additional work:

1) Simulation of DA systems--we will need sophisticated simulations of data and control flow at an early stage of the design of SSC data acquisition systems.
2) Fault tolerance and redundancy--It was generally felt that we do not need totally redundant readout paths for all data, but we do need to understand carefully just how much can go wrong before we no longer are taking useful data. Simulation will help a lot here.
3) Use of expert systems for fault diagnosis and debugging.
4) Playback of (Monte Carlo or real) data through the DA system.
5) Buses--what should be the standard for high speed data links (presumably some form of fiber optics);
   what should be used for the diagnostic/control network;
   what use can be made of data driven communications protocols;
   are any of the emerging new bus standards (Futurebus, SCI, etc) or any existing bus standards of any relevance to SSC.
6) We need a standard suite of benchmarks for processor evaluation.
7) Can we use high performance RISC processor cores in an application specific integrated circuit (ASIC); how will we interconnect such powerful special processors.

The LAA Real-time Benchmarks

R.K.Bock, W.Krischer, S.Lone[1]
CERN, Geneva, Switzerland

Introduction

In the context of the LAA detector development program a subgroup *Real Time Data Processing* has tackled the problem of intelligent triggering. The main goal of this group is to show how fast digital devices, implemented as custom-made or commercial processors, can execute some basic algorithms, and how they can be embedded in the data flow between detector readout components and fully programmable commercial processors, which are expected to be the final data processing filter in real time.

This work has to be done without making specific assumptions about detectors, using the most likely accelerator parameters of future high-energy hadron colliders, like a bunch separation of no more than nanoseconds (for the SSC 16 nsec), and luminosities of 10^{33} $cm^{-2}s^{-1}$. Even assuming a first-level analogue trigger with a reduction factor of 10^3, the average decision time at the next (post-digitization) stage is still of the order of 10 μsec. Currently available standard data busses and processors can not be extrapolated to such performance numbers; their basic architectures, however, may well serve as models for processors that are custom-built for speed.

We have therefore, as part of our studies, defined a set of algorithms which should cover the most basic trigger concepts of future colliders. We intend to run these 'benchmarks' on a variety of architectures, to obtain conclusions about their suitability for this kind of application.

Benchmarking for real-time algorithms

When evaluating computer architectures for specific applications, average performance figures given in MIPS or MFLOPS give little indication. For critical algorithms, the only significant benchmark is to run the actual algorithms on actual data, and thus measure performance under a precise definition.

For typical HEP analysis programs and algorithms, used on main frame computers, constraints like program invariance (portability) are of higher significance than optimal execution time. In the real-time case, the computer architecture is more complex, and adaptation of the code to the specific architecture is likely to be necessary. In our case of second-level trigger, the execution time may even determine whether or not an architecture can be used for a certain algorithm. *Optimization to architectures* is therefore part of this particular set of benchmarks.

---

[1] Presented by R.K.Bock at the 'Workshop on Triggering and Data Acquisition at the SSC', Toronto, January 1989

In such circumstances, the ideal description would specify the actual problem and the desired result only, leaving entirely open the algorithm to solve the problem. This would leave room for creativity and flexibility in tailoring to the strong points of an architecture.

With such a general definition, however, comparisons become difficult if not entirely meaningless. We have therefore attempted in our benchmark description to give beyond the *problem* also a *method* in sufficient detail, but without imposing a fixed *algorithm*. Although we indicate algorithms, as precise problem descriptions, we understand them as no more than that. They are not the simplest or fastest algorithms, but they have been used with success earlier. Our formulation may be slightly biased towards image processing, as we presently believe the specific parallelisms built into commercial image processors to be particularly suitable for some of our tasks.

Problem definitions

We are concerned with *representative* tasks in real-time data acquisition systems such as they could be used for triggering detectors using fine-grain and local information, or for data compaction. Our ultimate goal is to find algorithms and architectures suitable for 'second level' triggers, with decision times of few μsec. They are thought to be critical components in the data acquisition of detectors around future high-intensity hadron colliders.

We want to adapt our algorithms to a variety of commercial architectures. Although practical implementations of the future triggers will more likely be in semi-custom made electronics integrated into the data acquisition system, successful commercial architectures may be copied with few changes into the electronics design. For each problem-area, there is a set of data to verify and to measure the algorithm performance. The data are partly from simulations and partly from real experiments. The described algorithms have been coded in Fortran, but this should be seen as accidental: they have not been optimized under any aspect with respect to execution time. The supplied code is meant only to constitute a precise description of the method. At least some adaption to the architecture under study is definitely necessary.

As of now, the most exciting physics at future hadron colliders like the SSC seems to be the search for missing basic elements of the Standard Model, and the exploration of the physics beyond. Most prominently, the search is for the Higgs boson, heavy sequential quarks and leptons, additional gauge bosons, supersymmetric particles, and leptoquarks. Their signatures can be abstracted to very few significant elements: The recognition of *high-p_t charged leptons or jets*, or of *escaping transverse energy*, possibly with constraints (mass) between them. All features are comparatively central, i.e. in a rapidity range ± 4. (This list explicitly leaves out a very demanding trigger area, that of b-physics, which concentrates on high-precision tracking and vertex finding, and can not be confined to the central rapidity region).

Charged leptons are single charged particles with a signature in an electromagnetic calorimeter or muon chamber (we ignore tau leptons), possibly supplemented by track identification information. Missing transverse energy is obtained by summing energy vectors over all cells of a hermetically closed hadronic calorimeter. Jets are defined by local clusters of energy (e.g. peaks with some criterion of isolation), again in a hadronic calorimeter. Supporting information on jets from charged tracks is not considered relevant at this level. We can assume as generic detector elements calorimeter cells (to trigger on and distinguish electrons from QCD jets), muon chambers, limited tracking, possibly some track identification devices like TRDs or even RICHes. The detectors are expected to be 'fine-grained', i.e. with a large number of channels. Some local intelligence, with a large degree of parallelism, is thought necessary to identify the *significant* signals in a sea of trivial cells. Beyond the role of identifying relevant signatures, and of making them available to a higher level trigger logic, substan-

tial data compaction may also be expected from such intelligent devices. They may be relevant in alleviating serious future problems with data path bandwidth, if we can assume that a large class of events can be recorded without the finest detail of detector signals.

Digital signal processors play an increasingly important role in such 'smart' detectors. Today, they mostly operate on single channels. Large arrays of detector cells do, however, provide information that can be likened to 'images', and can be subjected to methods of conventional image processing. An image in this sense is nothing more than a two-dimensional array of integer values referred to as intensities or 'greyvalues'.

Due to the limited transfer rates for data through the detector electronics, a simple technique of thresholding data and only transferring values above the threshold together with their addresses has been used extensively and may be expected to be used in the future. These data are often referred to as '{x, y, greyvalue} lists'. We use the same technique in our benchmarks, in order to reduce the data transmission problem. We have no intention of thereby biasing future overall system designs.

The following tasks constitute a representative selection from the phase of triggering and data acquisition:

1. *GENERIC PEAK FINDING.*

   A subtask of several algorithms is to distinguish fluctuations of intensities in an image above a background-level that may vary over the image. An algorithm often used, called a peak finder, uses a fixed neighbourhood around each pixel. The average of the nearest neighbours is compared to the centre pixel. In addition, the average of the next nearest neighbours is compared to the average of the nearest neighbours. The actual data give the thresholds for the two comparisons.

   The input image is an N by M array of 6 to 12 bit pixels, e.g. 512 by 512, 8 bits. Alternatively, only the greyvalues above a threshold are transferred. The result should be an x,y list of addresses.

2. *GENERIC PATTERN RECOGNITION BY PARAMETERIZATION.*

   In many cases a detector allows to observe quantities from which physics parameters can be derived by a straightforward parameterization. In RICH detectors, for instance, particles leave a hit pattern whose shape depends on the angle with which the particle hits the detector and the velocity of the particle. The shapes are curved lines, like conic sections. The patterns can be precalculated and parameterized by polyniomials, if one assumes tracks come from the interaction point. Converting the positions of hits (pixel positions) into a measured quantity requires then a polynomial evaluation or table lookup and peak finding in the resulting histogram, or more simply, calculating the mean of the accumulated table lookups.

   The input might be a part of a detector and the address of the window and the result the velocity β of the particle causing the pattern.

3. *CALORIMETER CLUSTER ANALYSIS.*

   The particles deposit all their energy in the calorimeters. Usually they enclose the interaction area hermetically and are divided up in segments (pixels) according to the wanted spatial resolution. The information may be used to identify single particles or 'jets'. Electrons give a narrow (high intensity) peak in the calorimeter, spread over very few segments only. Jets cause energy-depositions from several particles hitting the calorimeter within a comparatively small area. Both give rise to an intensity-distribution within a confined area of the detector,

of a statistically predictable form. The number of segments involved depends on the energies and the calorimeter's cell size. Jets may, of course, include several separate local peaks (e.g. from gluon bremsstrahlung). The details of windows to use and of the test statistics to calculate, must remain open; for our benchmarks, they have to be taken as examples.

When the calorimeter data is used for triggering, we can assume the first level trigger to divide the total calorimeter region into fixed subregions (windows). The energy in each window is determined in parallel by, say, an analogue add. The window with the highest total energy, plus a number of the surrounding pixels, is submitted to the second level trigger which calculates a set of parameters giving the probability that there is a jet or an electron (or pile-up energy) in the window. Two algorithms of different complexity are described to parameterize the area around the highest intensity inside such a window.

4. *CALORIMETER TOTAL TRANSVERSE ENERGY CALCULATION.*

Finding particles that escape detection like neutrinos is done by forming a vectorial sum of all elements in a calorimeter, to calculate the 'missing energy'. This means projecting the measured energy into the transverse plane and summing up the resulting 2-vectors. With no escaping energy, the sum should be close to zero as the original colliding particles have no transverse energy.

5. *TRACK FINDING.*

The principle of track detection can be quite different according to the detector used, and the wanted resolution and accuracy. At the lower levels of triggering, the data may be assumed to be simple hitpatterns, i.e. a binary input image not taking into account the information giving higher accuracy. The tracks can be straight or curved, crossing each other, be close to each other and be non-contiguous. We make the assumption that track finding is done only in windows defined externally, and that only high-energy tracks from the interaction point are considered relevant, i.e. tracks with little or no curvature. This can be used, for example, for certifying high-p_T charged leptons in a tracking device. Curved tracks or tracks from secondary vertices are not part of the present benchmarks.

The input image for a fast trigger can be assumed to be only a part of the total detector, e.g. maximally 128 by 128 pixels; the spatial resolution with respect to closeness of tracks and angle between two tracks depend on the actual detector. The result of the applied algorithm should be the number of tracks identified in the image, and the number of points found per track.

Two algorithms for track finding are proposed: The Hough-transform to do general straight-line detection, and a grey level histogramming method with a mask defining the possible tracks as a general track finder for tracks with known origin.

Description of methods

In the following is a description of the methods chosen as solutions to the above described tasks. These are expected to be implemented on the architectures to be benchmarked, tailoring the algorithm to the architecture to get the best possible performance.

- ***Generic peak finder***: The described algorithm is a combination of a generic peak finder and an operation specific to these data: The generic part says that a pixel should be bigger than the average of the neighbours and in addition there is the data specific requirement that there should be a local maximum. To be a local maximum, a pixel (C) must be equal to or greater than the maximum value of the pixels (N) in a 3 by 3 neighbourhood. Further, for each pixel (C) in the image, the average of the four nearest neighbours (I) and the eight next nearest neighbours (O) as defined in the following figure are calculated. A peak is then defined as a centre pixel greater than the average of the nearest neighbours by a given threshold, if also the average of the nearest neighbours is greater than the average of the next nearest neighbours by another threshold and the centre pixel is a local maximum. If these three conditions are true, the centre pixel is set to one, otherwise to zero.

```
  O
 OIO          NNN
OICIO         NCN
 OIO          NNN
  O
```

$$C - \frac{1}{4}\sum_{i=1}^{4} I > C1 \qquad C \geq \max(N_1 \ldots N_8)$$

$$\frac{1}{4}\sum_{i=1}^{4} I - \frac{1}{8}\sum_{i=1}^{8} O > C2$$

- ***Generic pattern recognition***: The values of all possible β's of the RICH-detector in function of the coordinate pairs taken with respect to the impact point are precalculated and stored in a table. The impact point of the particle on the detector defines the detector window and the part of the table to be used. For each hit in the input window, the content of the β-table is entered into a sum. The resulting average β is then the accumulated sum divided by the number of non-zero points (the benchmark ignores background-problems, i.e. a more sophisticated histogram analysis).

- ***Calorimeter cluster-analysis, simple jet/electron finder***: The first part of this algorithm is to find the pixel with the highest intensity, the centre pixel (C). Then three parameters are calculated according to the pixels in a fixed neighbourhood around the centre pixel.

```
00000
0IIIO
0ICIO
0IIIO
00000
```

The three parameters are defined as follows:

$$E = \sum_{j=1}^{16} O + \sum_{i=1}^{8} I + C$$

$$I/C = (\frac{1}{8}\sum_{i=1}^{8} I) / C$$

$$O/C = (\frac{1}{16}\sum_{j=1}^{16} O) / C$$

Jets, electrons or the background (pile-up)can then be distinguished by different thresholds on E, I/C and O/C.

- ***Calorimeter cluster-analysis, variable neighbourhood***: In the above analysis, a fixed neighbourhood around the peak-value was used. In this algorithm, a number of centre pixels (c) should first be found by considering all neighbours of the absolute peak (C) above a given threshold as centre pixels. Expand the 'centre pixel region' until none of the neighbours are above the threshold. We have now an irregular region of centre pixels. Calculate their average greyvalue 'cav'. Then find the average greyvalue of the 8-connected neighbours (i) of the centre pixels 'iav'. The neighbourhood may in principle extend to the end of the window. Then the average of the next nearest neighbours (o) should be found: 'oav'. Like the previous algorithm, the sum of the centre pixels and the ratios iav/cav and oav/iav define jets, electrons or background.

```
ooooo
oiiiooo
oiciiioo      C - original peak value
oiiCciio      c - expanded centre pixels
ooicccio      i - nearest neighbours
 oiiiiio      o - next nearest neighbours
 ooooooo
```

- ***Calorimeter total transverse energy calculation***: The transverse energy is found by projecting cell energy vectors (vector direction from cell position with respect to interaction point, vector length from measured energy deposition) onto a plane perpendicular to the beam direction. The x and y components are then summed separately over all cells. The total missing transverse energy is then given by $\sqrt{(\Sigma Ex^2 + \Sigma Ey^2)}$, and can be thresholded to identify escaping neutrinos. The cell energy E is projected onto the XY-plane (transverse energy) by:

$$E_T = E * sin(\theta)$$

and the X and Y components are found by:

$$E_X = E_T * cos(\phi)$$

$$E_Y = E_T * sin(\phi)$$

- ***Track finding by Hough-transform***: In the Hough-transform an input image given in coordinate-system (x,y) is transformed into the Hough-parameter space (ρ,θ) according to the formula:

$$\rho = x\cos(\theta) + y\sin(\theta)$$

ρ is the distance of a straight line from the origin of the coordinate system, while θ is the angle of the normal to the line with respect to the x-axis. The size of the parameter space, i.e. the number of distances and angles, might vary from application to application, but in this test it is assumed they are of the same dimensions as the input image. In a serial fashion, this means going through all the pixels or the x,y list of the input image θ times. Once the transformation has been done, tracks are identified as peaks in the new image. The peaks might not be single pixel values, but clusters of pixels. Identifying tracks lying close or crossing each other at a small angle might demand complicated cluster-analysis, or an iterative way of finding the individual tracks. A brute force method could be for instance to redo a Hough-transformation without the coordinates of the found track. It is therefore considered most important to be able to do the Hough-transformation fast, and the analysis is not a part of the benchmark.

Track finding by grey level histograming: The input is a list of (x,y) pairs in the range 0 to 255 (8 bits). For each entry in the input list the predefined sector number the (x,y) pair belongs to is taken from a table, and the sector hits are histogrammed. Peaks in the greyvalue histogram identify directly the tracks. The tracks in the benchmark data are straight lines from the centre of the image, and the table defines the angle ϕ.

Description of data

The data to be used as input for the benchmarks, exist as ASCII-formatted sequential files, written with FORTRAN format statements. Some contain a complete image (i.e. not zero-suppressed) in hexadecimal format, others are images in list format, i.e. x,y addresses and a greyvalue in decimal format, with pixels containing zero not included.

There are 7 files containing images and lists and in the following is a short description of each.

Peak finder data: a complete greylevel image of 512 by 512 bytes.

Generic pattern recognition data: For this task there are two input files, containing the input image and the lookup table with precalculated β's. Both images have the dimension 64 by 64. The data are for a simulated liquid radiator RICH with pad readout, the table derived from a fluctuation-free training sample.

Cluster finding and transverse energy data: For the calorimeter analysis, data from the ISAJET-program, used for LHC simulations, is supplied. The selected events are QCD jets with $E_T \geq 500GeV$. The original data have been run through a calorimeter simulation, resulting in an image fed into the 1st level trigger. The detector is assumed to cover a rapidity range of -4.0 to $+4.0$ and 2π in azimuth with a resolution of 512 in both directions, giving an image of 512 by 512 10-bit numbers. Each hadron particle reaching the calorimeter within its range, results in an energy-deposition which is spread to the neighbouring pixels according to a Gaussian function where 95% of integral is contained in 6cm radius ($\sigma = 3$cm). Electrons are spread with a width 1/5 of the hadronic width. Longitudinal segmentation has been ignored. Only non-zero and thresholded x,y lists have been produced. There are two files, containing several events of different characteristics: high-p_T jets only (pure QCD) for one, and jets with high-p_T electrons and neutrinos, resulting from a high-p_TW.

Hough-transform data: The Hough-transform is to be done on a x,y list of data. The address range in x,y is 0 to 127. Other parameters e.g. the bin widths in ρ and θ are defined as DATA statements in the benchmark routine.

Grey level histogram data: This algorithm is to be run on a list of x,y's from an image of 256 by 256 pixels. The tracks originate from the centre of the image.

The intended benchmark results

Timing and modularity.

In the examples it is indicated which part of the algorithm is interesting from the benchmark's point of view. The rest is input and setting up of data and not regarded as a part necessary to provide timing for (embedding is a vital but separate problem). The timing for benchmark components should be extracted whenever possible, especially for commonly recurring operations.

Numerical results.

The numerical results of the coded benchmarks exist on ASCII files. When running a benchmark on an architecture, these should compare within a reasonable limit taking into account the different precision used by the hardware. A more precise definition of 'reasonable' in the form of a fixed number is not given, but will have to be evaluated from one instance to the other.

I/O.

The I/O system of an architecture is a very important feature which becomes even more important as the processing power increases. It is of particular interest in our applications which demand high throughput ***and*** low latency, that the processing power and I/O-bandwidth ratio must be balanced. If processing and I/O can not overlap, some I/O-time will add to the total execution time. Although it is not directly included in the benchmark to provide timing for I/O, it should be documented that the architectures under study have the capabilities to achieve the necessary I/O-bandwidth.

Target architectures

We are presently running this set of benchmarks on four different specific parallel architectures. In addition, tests on general-purpose processors like the SPARC chip or transputers, are being done, in C and/or Fortran.

Data Cube

Modules taken from the MaxVideo family of image processors are being combined differently for the various algorithms. They communicate data through a 'Maxbus' with present bandwidth of 10Mbyte/s, and control signals over VME. The standard host is a Sun workstation. An internal memory card allows to access subregions in a larger picture preserving full pixel speed. DataCube is a company in Peabody, Mass.

GAPP

The 'Geometric Arithmetic Parallel Processor' (GAPP SCG72) of NCR is a SIMD machine, with a basic cell of 72 mesh-connected 1-bit processors. Each processor contains 128 bits of RAM, 4 1-bit registers, a full adder/subtracter, and connects to the 4 nearest neighbours. Interprocessor communication runs at 10MHz. The chip contains 20 control lines, 7 address lines, and 13 lines for the instruction. Chips can be cascaded into larger arrays, connecting on all four sides. The benchmarks on the GAPP chips are executed for LAA by Laben SPA., Milano.

ASP

The 'Associative String Processor' is a novel parallel architecture, made of basic building blocks which are dynamically reconfigurable. Substrings of 'Associative Processing Elements' are connected through a Communications Network, all elements sharing bit-parallel data, control and activity busses, and a single feedback line. Processing elements can be selected by matching with the data bus, hence they can operate as an associative memory. Data can be in the range of 32-128 bits, activity registers may be 4-8 bits. Intended for wafer scale implementations, fault-tolerance is built into the system from the outset. The benchmarks with a specific implementation of ASP-s, called TRAX-1, is executed in collaboration with the heavy ion experiment NA35 and Aspex Microsystems, Uxbridge.

DAP 500

The 'Distributed Array of Processors' (DAP 500) is commercialized as a two-dimensional mesh of 32-by-32 1-bit processors. Each processor has a 32 Kbit deep memory, there is a single master control unit with its own memory, and a SCSI or VME connection to a host computer (Sun). The DAP can be programmed in Fortran. The system is produced by AMT Ltd., Reading. Benchmarking for the DAP is being done using the available software simulation.

OBJECT-ORIENTED SYSTEM BUILDING FOR THE SSC

G.A. Ludgate
TRIUMF, 4004 Wesbrook Mall, Vancouver, B.C., Canada V6T 2A6

ABSTRACT

The concepts and terminology of object-oriented system development are presented assuming the reader is familiar with the Structured Analysis and Structured Design (SA/SD) methodology. An improvement to SA/SD, known as Object Oriented Analysis/Object Oriented Design (OOA/OOD) is described and the steps in such an undertaking explained. Object Oriented Programming (OOP) is briefly introduced before SA/SD and OOA/OOD are compared. A recommendation to the SSC Central Design Group concludes the paper.

INTRODUCTION

Physicists building large detectors have, at last, over the last few years ventured away from a purely FORTRAN mind-set in their approach to software system building and have tried the somewhat newer technology of Structured Analysis and Structured Design (SA/SD). In this paper an advance on SA/SD is introduced that simplifies some of the areas addressed by traditional SA/SD, provides more guidance to physicists when specifying systems to be built and creates more robust systems. The approach is called Object-Oriented Analysis/Object Oriented Design (OOA/OOD). It is just starting to gain ground and acceptability in the mainstream of software engineering, due in part to the introduction and heavy support for Ada in North America.

OOA/OOD is an extension of SA/SD. The SA/SD tools of Entity Relationship Diagrams (ERD), State Transition Diagrams (STD), Control Flow Diagrams (CFD), Data Flow Diagrams (DFD) and Network Diagrams are used unchanged in OOA/OOD. What is new though, is the emphasis on structuring a system being built around the objects mentioned in the description of the problem - the so called problem statement. This requires a variation in the sequence of using the traditional SA/SD tools when carrying out OOA/OOD.

Object-oriented specifications more closely resemble the structure of the problem domain, by construction. They are, therefore, simpler to understand and to change. The addition or removal of an object from the problem statement usually leads to the addition or removal of those parts of the specification model that represent that object alone. Other parts remain untouched.

While contemporary Computer Aided Software Engineering (CASE) tools can be used to represent OOA/OOD models, these CASE tools do not yet provide any global design-rule checking specific to this approach, nor do they provide assistance in performing any of the elementary steps comprising the methodology.

This paper will assume the reader has a working knowledge of SA/SD. For the new student of software engineering references are made in the paper to text books covering the topics in detail. No attempt is made to refer to the original development literature.

THE OBJECT-ORIENTED PARADIGM

Humans employ an object-oriented approach to life. By this we mean that as part of trying to understand their surrounding, humans classify uniquely identifiable objects into groups with similar properties and behaviors. These groups, called "classes" in object-oriented programming, are familiar to all e.g. doctors, electricians, physicists etc. We will continue to use the term "class" in this paper as it is somewhat traditional in object-oriented literature. Each object is said to be an "instance" of a class.

When humans need a particular form of assistance they make use of one or more members of one or more classes of experts for advice. They send the expert a message (a written message, a phone call or even a face to face conversation) specifying the problem and the expert replies with the solution. The human does not access the expert's resources (e.g. personal library, case history files, brain-memory etc.) directly during the interaction, nor are they really aware of how the solution was arrived at; but take the validity of the solution on faith and proceed to use it. In turn, they may be a member of one or more classes and, like-wise, be asked for solutions to problems. If all existing members of a problem solving class were replaced by a new, and possibly vastly improved, membership the original questions should still be answered the same. Similarly, if a new class of problem solvers came into being all humans would not restructure their entire lives. Rather, only those that needed the services of the newly created class would do so and the remainder would be unaffected in their day-to-day interactions. Could we not build systems that operate in a similar manner, embodying autonomy and independence of the parts of the whole?

We already do! Cars are one of the classes that are built in an object-oriented manner. The nuts and bolts objects can be replaced without causing major restructuring of the original car. A new and more powerful engine could replace the original engine and one would expect the same services from the car, though it may provide such services much faster. Similarly, microchip manufacturers provide pin-compatible versions of more powerful microprocessors to replace older versions of a microprocessor. After the old microprocessor is replaced, one expects the same services from the existing computer software. New services provided by the chip may go unused until the functionality of the old software is enhanced and requires the new services of the microchip to support the implementation.

In software engineering at most HEP labs the above object-oriented approach is limited to the use of software library components as part of an application program. Improvements to the subroutines in the library are realized as soon as the program is re-linked and executed. In general though, major changes to the services provided by any program involves ad-hoc patches to the software or a major restructuring of the original. One of the causes of these drastic changes can be traced to fact that the structure of most software is not related to the structure of the problem it solves, but is strongly influenced by the structures provided in the programming languages used to solve the problem. The structure of these languages, are in turn, dictated by the underlying structure of the computers on which the programs are expected to run (the traditional Von Neumann architecture). The languages adopted by most scientists provide mechanisms that allow you to:

- Specify a number of statement to be executed sequentially,
- Select the next statement to execute based on a the truth of a condition,
- Repeat a sequence of statements,
- Group the above constructs into higher level constructs called "subroutines or functions",

- "call" lower level subroutines from a higher level one.

As these are normally the only available mechanisms provided by the language, programmers are forced into a style of coding in which a "main" program solves its problems by calling upon the services of lower level subroutines, which in their turn call even lower level subroutines. This has led to the popularity of "Top Down Decomposition" for program design and the resulting hierarchy of subroutines "calls" are normally depicted on a Structure Chart.

This style of programming is very suitable for a certain class of problems in which the outputs of the program are a function of its inputs; but even here, as there is a very large number of ways of specifying even the simplest functions there is no *a priori* manner of deriving the structure of the program from a knowledge of its inputs and outputs. A simple modification to the required inputs and outputs may, therefore, result in changes to code in many parts of the program structure.

THE PATH TO SOFTWARE ENGINEERING

The earliest attempts to improve programmer productivity resulted in the guidelines for Structured Programming[1,2]. Proponents of Structured Programming are, in effect, saying that one program is better than another program if it obeys the guidelines of Structured Programming regardless of the problem it is solving. These guidelines often lead to an efficiency loss in solving a given problem due to the increase in the number of programming statements demanded by the structured programming guidelines. This decrease in efficiency was usually irrelevant as the capacity of computers increased rapidly during those early years. Structured Programming improved the comprehensibility of programs and therein lay its usefulness and success.

As the size of application programs increased the next improvement to appear was Structured Design[3,4]. Structured Design addressed the problems involved in grouping language statements together to form subroutines. It gave a set of guidelines that allowed programmers to determine whether a particular grouping of statements into a set of related subroutines was appropriate or not. In addition its guidelines allowed programmers to decide for themselves whether one grouping was "better" than another. As before there was a loss of efficiency in grouping statements into subroutines rather than solving the problem by writing one, very large program; but again the loss was deemed acceptable due to the perceived improvement in understandability and maintainability of the resulting program.

Finally Structured Analysis[5,6,7] appeared to guide developers automating existing systems and creating new systems composed of many related programs (though not necessarily concurrent). Coupled with Structured Design, SA/SD is now the backbone of most contemporary applications development methodologies in the business world. CASE tools provide for the creation of the graphical SA/SD models and automate a lot of the work involved in building the specification of a new system. In recent years extensions to SA/SD have appeared[8] that address the problems of programmers constructing applications that are dominated by control actions/reactions. These extensions, usually called "real-time extensions", have been added to most CASE tools. With the availability of CASE tools the chores of manually constructing and keeping up-to-date many graphical models and data dictionary entries have been alleviated.

Systems developers, using SA/SD, have gained in improved system maintainability and comprehension due, in part, to the creation and availability of a high level model of the new system before coding takes place. This high level model is the specification of the system-to-be-built and, as such, can be shown to users for validation of their requirements during early phases in the development cycle. Higher levels of productivity have been reported[9] along with lower levels of errors in the delivered system.

The creation of software based systems has now evolved into an engineering discipline[10]. The creators are known as software engineers and one of the multiplicity of tools in their toolkits is SA/SD. For many systems now being built is no longer the case that efficiency is the primary driving force in determining their structure; rather structures based on other criteria are now considered acceptable (that is the system's structure is no longer solution dominated but problem dominated). A discipline has matured into an engineering discipline when multiple solutions exist for most problems. Engineers of the discipline use guidelines such as cost, performance, size and maintainability to determine the "best" solution to the problem at hand. Product quality must be measurable and assured at every stage in the construction of the solution. The rapidly decreasing cost of computing power compared to the rapidly rising costs of staff has helped to bring this change about.

INFORMATION (OBJECT) MODELLING

The concept of objects, as described above, re-occurs throughout computing. During the same period that SA/SD was becoming popular, research in database technology led to the development of relational databases[11,12] and the use of graphical models, such as the Entity Relationship Diagram (ERD), to depict the work[7].

On an ERD a set of "entities" with similar properties are represented by a named rectangle. "Entities" are real world objects or concepts and a set of them is equivalent to a class of objects, as described above. Similarly, a set of "relationships" is represented by a named diamond linked to one or more "entities" by lines. "Relationships" depict the possible associations that can occur between the real world objects connected to the diamond by the lines. As an example one could imagen an ERD of a system constructed for the Registrar of Births, Deaths and Marriages (Fig. 1) on which a MAN entity and a WOMAN entity appeared as named boxes. The relevant relationships between these two objects could be MARRIED, DIVORCED, FATHER_OF, MOTHER_OF etc. The ERD essentially captures the "meaning" of a given portion of the real world.

An ERD can also be used to depict both the state of a portion of the real world and the state of a database that, by design and implementation, stores a symbolic description of that same part of the real world. If, at all times the state of the database can be kept isomorphic with the state of that portion of the real world under examination, then enquires made of the database will provide the same answers as enquires made of the real world; the hope of database creators being that the former enquiry can be serviced faster than the latter, and at less cost.

In a relational database each class of objects is represented by a table (Fig.1). Each row of the table represents and describes a separate object (an instance of the class). The columns of the table store the attributes of the object. Each object must have one attribute, called the "primary key", that uniquely identifies the object in the table. Similarly "relationships" could be represented by tables. Each row of the table would represent an instance of an association between one or more objects. If there were no entries associating objects at a given time then no relationship is said to exist between the objects at that time. The attributes of the relationship describe the association and are, therefore, the primary keys of the several objects associated by that instance of the relationship. In practise, for efficiency, the columns of a relationship are merged with those

of the several associated objects in a real database.

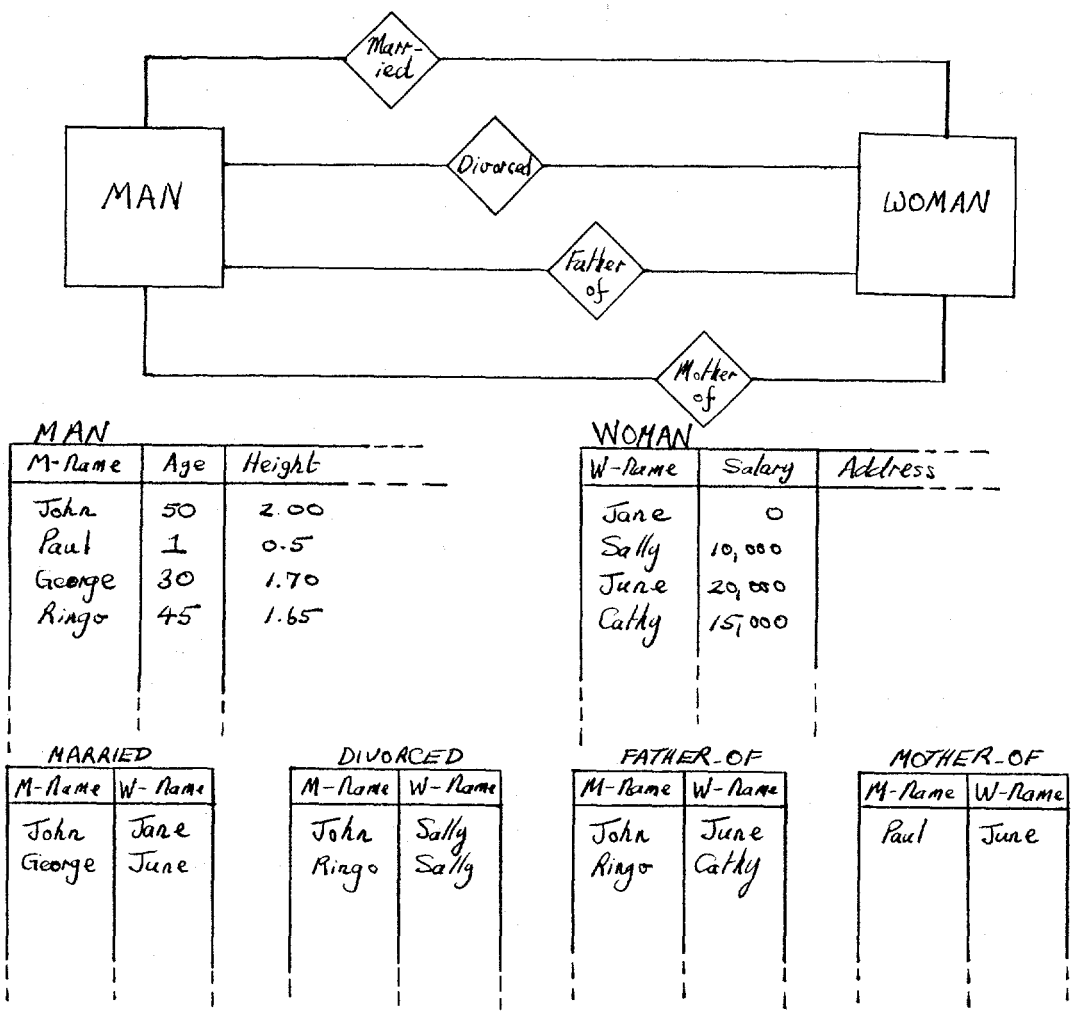


Fig. 1. ERD fragment and the corresponding relational tables in an example system constructed for the Registrar of Births, Deaths and Marriages

The ERD is central to the topic of object-oriented system building. Relationships noted on the ERD have a structuring effect throughout the systems development life-cycle. It is an implementation choice whether the ERD represents a database in the final application system, or in-memory tables. This does not imply that arrays will appear everywhere in object-oriented programs, a table comprising one line and 3 attributes may be implemented as 3 distinct program variables. One must just realize WHAT it is a programmer is trying to represent and exercise good judgement in using programming language constructs that do not obscure the mapping between real-world entities, the analysis models and the final language constructs.

Two relationships are of special importance to the object-oriented approach, namely "IS_A" and "IS_PART_OF". Two objects, say A and B, are associated by the "IS_A" relationship whenever A IS_A B; that is A has all of the properties of B but not the converse (e.g. MAN IS_A HUMAN,

HUMAN IS_A VERTEBRATE etc.). Similarly two objects, say X and Y, are associated by the "IS_PART_OF" relationship whenever X IS_PART_OF Y; that is Y has X as a component of itself (e.g. WHEEL IS_PART_OF AXLE, AXLE IS_PART_OF CAR etc.). The IS_A relationship is very important for object-oriented programming where it forms the basis for state and behavioral inheritance[13,14]; while the IS_PART_OF relationship is important for OOA/OOD where it forms the basis for hierarchically partitioning the final requirements model - usually called "leveling up the Network Diagram" in SA/SD parlance. Both relationships induce local hierarchical structuring on classes appearing on an ERD.

OBJECT-ORIENTED ANALYSIS

The final requirements model produced by OOA (the Network Diagram) is event-partitioned similar to the modern approach to SA/SD[7,8]; however, unlike SA/SD an "event generator" technique is provided utilizing the ERD rather than forcing users to guess the events to which the new system must respond. Following the discovery of all events the structure of the system-to-be-built is derived in a near mechanical manner from the models constructed during the analysis phase.

The first step in carrying out an object-oriented analysis of a problem statement is to prepare an ERD embodying all of the objects that are part of the problem and the proposed solution. One should imagen that the real world objects are going to be represented in the system being built by tables, as described above, and that these tables will be kept up-to-date by the interaction of the new system with the real world. Attributes are identified for each object and one is chosen as the primary key (the object identifier). Only attributes relevant to the problem at hand are chosen. Later stages will highlight missing attributes and call for redundant attributes to be removed if the analyst is unsure. When in doubt include the attribute.

On any of the tables representing classes of objects only four basic operations need to be carried out to maintain the table state and report on its contents, namely:

- The creation of a new object; a new instance of the class,
- The reporting of the values of one or more attributes,
- The updating of one or more attributes of an object,
- The deletion of an object; removal of an instance of the class.

By systematically itemizing the names of all real-world events that cause the creation, reporting, updating and deletion of objects in every table a full list of events is generated to which the system must respond. The system's responses to these events will be discussed below.

A similar process must be carried out for the "relationship" tables but now only creation and deletion are relevant.

The last of the events the system is to respond to are again derived from the ERD by considering the possibility of discrete attributes having life-cycles. A trivial example of this is provided by a DOOR (Fig. 2) in which the attribute STATE can only take values from the set {OPEN, CLOSING, CLOSED, OPENING }; but is restricted in its ability to move from one value of STATE to another by the laws of physics. The values of STATE are permitted to change following a life-cycle modelled by a State Transition Diagram (STD) (a finite state machine). On an STD named boxes represent named behavioral-states. That is to say, the behavior of the system is always the same when it is in that state. Directed lines joining the boxes represent allowed

transitions. The condition that causes a transition is written adjacent to the line. In SA/SD and OOA/OOD events cause an STD to transition from one state to another. It is possible to specify responses to the event by writing them below a line drawn under the event name.

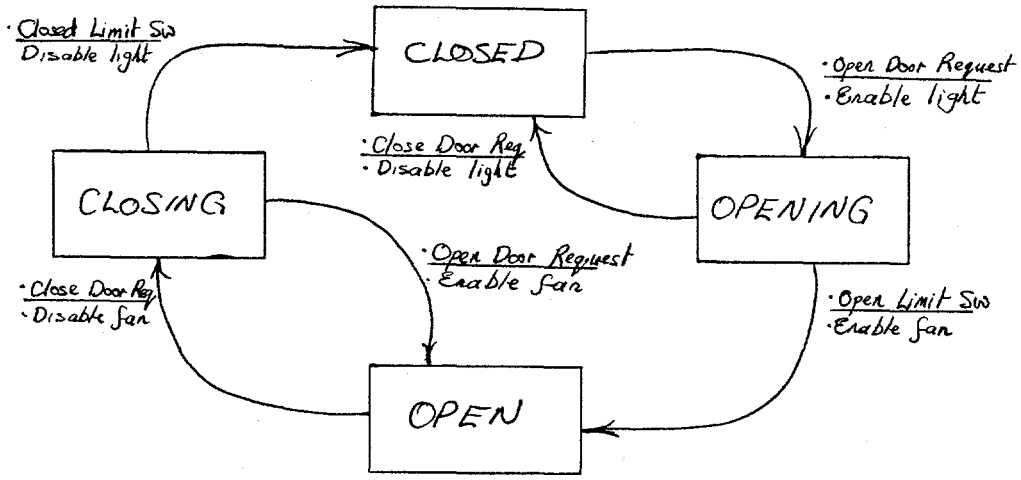


Fig. 2. STD of a DOOR object's STATE attribute.

For consistency the ERD should contain an "entity" representing the system-being-built and show its relationship to other entities. A list of attributes may also be prepared, some of which may be discrete and require STDs to describe their life-cycles. If the new system must have one or more life-cycles in its externally visible modes of behavior then one or more of these attributes must represent the modes of behavior and an STD for the attribute must be "designed" to match the behavior of the system as expected by the users. Thus no object, including the system being built, is treated differently in OOA/OOD.

The STDs discovered during data modeling affect the behavior of the system being built. They are represented on the final system specification as Control Transformations - dashed bubbles - that control the execution of Data Transformation bubbles[8].

Having "generated" the maximum number of events the new system is to respond to, the event list must be sorted by event name to reveal the complete picture of creates, reports, updates and deletes required for each event. Care must be taken in establishing whether an event has one or more aliases during this step. Responses to each event are added to the sorted table. These responses may take the form of changes in the state of objects, processes being triggered to generate data, processes being enabled to produce or consume data and processes being disabled from performing such functions. A description of each process identified may be required if the computation is complex.

Having completed the analysis phase the requirements model can now be derived from the models described (briefly, informally and incompletely) above. The final requirement model is a

graphical model known as the Network Diagram. The Network Diagram symbolically represents the entire system-to-be-built.

The first step in creating the Network Diagram is to represent the data to be stored by the new system. A data-store is created on the Network Diagram for each entity on the ERD. It is normally represented by two horizontal lines with the entity name written between the lines. Second, a data transformation bubble is drawn on the Network Diagram for each event recognizer process and all of its required response processes. The bubbles that use a particular data-store must be drawn close to that store and the data flows to and from the stores must be shown. Third, control transformation processes are added to the model and all control flows representing "trigger control", "enable/disable control" and "events flows" must be drawn according to the control requirements shown on all STDs. Finally all data-flow and control-flow inputs from and outputs to the real world are connected to the appropriate data transformations and control transformations.

At this point it may become apparent that there are problems with the flows into and out of the Network Diagram:

- Some input data-flows and/or control-flows from the real-world are unused on the Network Diagram; an indication that either one or more entities and relationship were omitted from the initial ERD, or that the inputs are not required,
- Some required input data-flows and/or control-flows are not available to the Network Diagram; an indication that either too many entities and relationship were placed on the initial ERD, or no provision was made for obtaining the data needed by the new system to accomplish its purpose,
- Some output data-flows and/or control-flows provided are unused by the real-world; an indication that an event response is wrong or too many entities and relationships were added to the initial ERD,
- Some output data-flows and/or control-flows required by the real-world are not available; an indication that an event response is wrong or entities and relationships were omitted from the initial ERD.

A refinement step can now be applied to the derived models by iterating the steps described above to remove mis-match problems with input and output flows.

The remaining step to be taken in completing the requirements model is to package the model for presentation and review. Again, at this step, the object-oriented approach provides guidance: the set of data transformations and control transformations bubbles that appear to be highly connected with a given data-store are all identified as a single unit, on the Network Diagram, called a Dynamic Object. This unit is the requirements model representation of the real world class of objects bearing the same name as the data-store inside the Dynamic Object. The components of the Dynamic Object are re-drawn on a separate Network Diagram and replaced on the original by a single data transformation bubble having a name closely linked to the name of the data-store removed. When this partitioning step has been completed on the Network Diagram it is repeated again and again until a single bubble remains - the Context Diagram - representing the new system-to-be-built object. This step of grouping of objects into higher level object is guided by the "IS_PART_OF" hierarchical relationship on the original ERD.

Having now covered all of the steps in modelling a new system, one can now see why the initial ERD should contain only "relevant" objects and relationships.

OBJECT-ORIENTED DESIGN

In traditional SA/SD parlance a "design" step follows the "Analysis" step. This step is often broken into two phases namely:

- System Design; wherein the functionality of the requirements model is partitioned into a set of one or more processes executing on a set of one or more interconnected processors, and
- Software Design; wherein the structure of the individual processes is decided in terms of a programming language.

In moving toward "design", the abstraction level at which the designer works decreases compared to the level used during requirements modelling. Processors and processes now "appear" and should be added to an enlarged Design-ERD. With these entities now represented it is possible to re-do some of the earlier analysis steps and investigate the events that cause process creation, process deletion etc. The sets of Dynamic Objects partitioned during "design" into processes now satisfy the "IS_PART_OF" relationship with the containing processes. Similarly processes executing on a processor satisfy the "IS_PART_OF" relationship with the processor. These relationship may be added to the Design-ERD for completeness.

During "design" the proper interface of the new system to the real world must replace the "essential flow" interface of the requirements model. The interface must be detailed e.g. ASCII terminal, windowing system etc. and only at this stage can the User's Guide be prepared in detail.

The user's interface is modelled by an elaboration of those data transformation bubbles that have inputs from or output to the real world.

OBJECT-ORIENTED PROGRAMMING

A full description of Object Oriented Programming is beyond the scope of this paper but the flavour of the subject will be briefly presented.

Object-oriented languages support the dynamic creation and deletion of objects, and the passing of messages and data between objects as the model of computation. One of the earliest object-oriented languages was Smalltalk. In the Smalltalk language there is no concept of "data type" that figures so prominently in Pascal, Modula II, Ada etc. When a new class of objects is defined the programmer must create the new class as a sub-class of an existing class. This placement of the new class implies that the new class and the old class are related by the IS_A relationship:

New class IS_A Existing class

The new class immediately inherits all of the properties and behavior of its superclass. As object-behavior is related to an objects ability to understand messages sent to it, this means that instances of the new class will automatically respond to every message already defined for the superclass - and no new code has even been written yet. The possibility for improving programmer productivity by this mechanism alone has not been lost on the software engineering community and is one of the driving forces behind attempts to provide tools and languages to support object-oriented methodologies from analysis through to coding.

There are now many languages commercially available purporting to be object-oriented, for example C++, Objective-C, various dialects of LISP, Simula, Eiffel, Trellis/Owl, Modula II and Ada.

Not all of these languages support all of the features of Smalltalk-like object-oriented computing but then Smalltalk is normally an interpreted language and runs an order of magnitude slower than the other compiled languages.

Possibly the most important of these languages in the next decade will be Ada; in part due to its heavy support by the U.S. Dept. of Defense for mission critical software. Ada does not support the "inheritance" of properties nor behavior. Thus, for each new class of objects defined, software must be written for every "message" that the object responds to (passed from object to object by traditional subroutine call). Programming support environments for this language will abound and compilers will be available for all common computers, micro-processors and micro-controllers.

identical

COMPARISON OF SA/SD AND OOA/OOD

The analysis tools of SA/SD and OOA/OOD are the same. The emphasis, order and manner in which they are used is different. The ERD is given prominence in OOA/OOD and is applied first to "objectify" the problem. It is the primary analysis model discussed with users of the system being built. The intent of the model is to capture as much of the meaning as possible about the problem under study in terms of the objects appearing in the problem and the relationships between them.

In early SA/SD[5] the system-to-be-built was decomposed into data flow diagrams, starting from the Context Diagram. This approach suffered from two major drawbacks:

- It completely ignored the wealth of data available about the "information environment" of the new system,
- The decomposition of the Context Diagram into data flow diagrams was not deterministic. If N programmers were given the task then at least N different decompositions would result: showing that there are many ways to represent the set of functions that describe the services to be provided by the new system.

These deficiencies were later corrected[8].

Control Transformations incorporating STDs are identified, in OOA/OOD, as the discrete dynamic behavior of an object on the ERD. In SA/SD they are introduced in a somewhat *ad hoc* manner to provide the required element of control demanded by the users. The end result is the same for either methodology.

Upward leveling of the Network Diagram, in SA/SD is based on grouping related functions together. In OOA/OOD it is based on object composition - the "IS_PART_OF" relationship rather than the "IS_FUNCTIONALLY_COMPOSED_OF" relationship. Both are hierarchical relationships but system building experience has shown that objects taking part in a particular problem are far more stable than the services expected from any new system. It seems to be the rule that the functionality of a system is expanded or modified far more often than the objects change in the problem statement. As an example one would note that a compact disc player is completely different animal from the wind-up record players of yester-year, but they solve the same problem, namely to convert stored music into audible music for humans to hear.

CONCLUSION

The object-oriented approach to systems building has been briefly described above. The orientation of this paper has been toward the construction of software based systems but, in general, the methodology can be used to model the behavior of any stimulus/response system being built, be it destined for implementation in software, hardware or peopleware. It is a general systems development methodology.

The field is just coming into its own but the potential programmer productivity gain from the availability of artificially intelligent CASE tools for OOA/OOD that understand the English language, an optimizing Smalltalk compiler and object-oriented databases cannot be discounted.

The SSC Central Design Group should study the object-oriented technologies available now and those projected to be available inside of the next 10 years, with the same open mindedness that is afforded the study of RISC architectures and other high technology items in various engineering disciplines. While no comprehensive text books exist for OOA/OOD the concepts missing from the text books referenced above can be read in[15,16,17].

REFERENCES

1. O.J. Dahl, E.W. Dijkstra and C.A.R. Hoare, Structured Programming (Academic Press, 1972).
2. R.C. Linger, H.D. Mills and B.I. Witt, Structured Programming: Theory and Practise (Addison-Wesley, 1979).
3. E. Yourdon and L.L Constantine, Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design (Prentice-Hall, 1979).
4. M. Page-Jones, The Practical Guide to Structured Systems Design (Yourdon Press, 1988).
5. T. De Marco, Structured Analysis and System Specification (Prentice-Hall, 1979).
6. L.Peters, Advanced Structured Analysis and Design (Prentice-Hall,1987).
7. E. Yourdon, Modern Structured Analysis (Yourdon Press, 1989).
8. P.T. Ward and S.J. Mellor, Structured Development for Real-Time Systems (Yourdon Press, 1985).
9. C. Jones, Programming Productivity (McGraw-Hill, 1986).
10. R.S. Pressman, Software Engineering: A Practitioner's Approach (McGraw-Hill, 1987).
11. C.J. Date, An Introduction to Database Systems (Addison-Wesley, 1986).
12. J. Martin, Managing the Data-Base Environment (Prentice-Hall, 1983).
13. B.J. Cox, Object Oriented Programming (Addison-Wesley, 1987).
14. B. Meyer, Object-oriented Software Construction (Prentice-Hall, 1988).
15. S. Shlaer and S.J. Mellor, Object-oriented Systems Analysis: Modeling the World in Data (Yourdon Press, 1988).
16. R.J. Abbott, An Integrated Approach to Software Development (John Wiley, 1986).
17. H. Abelson, G.J. Sussman and J. Sussman, Structure and Interpretation of Computer Programs (The MIT Press, 1985).

Workshop on Triggering and Data Acquisition for Experiments at the Superconducting Super Collider

University of Toronto, January 16 – 19 1989

Working Group 4-B Summary

SOFTWARE and PROJECT MANAGEMENT

Prepared by E.M.Rimmer, DD Division, CERN, 1211 GENEVE 23, Switzerland

The conclusions and recommendations of Working Group 4 – B are divided into three sections : (1) the overall project management framework; (2) the software system design environment; (3) some proposals for technical studies. None of the sections contains an exhaustive check-list; the group sought to highlight basic areas in which muddles or failures would seriously detract from the efficacity of financial and personnel investment in the production of software for an SSC experiment, and possibly even jeopardise the success of such a project.

It should be noted that the group had in mind the control, monitoring, communication and data collection packages required to support a multi-processor data acquisition system attached to a large and complex detector, and not the embedded physics algorithms for event filtering and data compaction. It could be argued however, that the production and maintenance of physics code might also benefit from an improved design and management approach.

Project Management Framework

It was agreed that the software component of an SSC experiment should be considered on a par with the detector and read-out hardware. A description of the software system should appear already in the letter of intent, and a basic design should form part of the formal experiment proposal. The integrity of this design and the ability of the proposers to build and run the software should be assessed and vetted periodically, as is presently done for complex detectors. The proposers should be required to demonstrate that sufficient resources, both material and personnel, are foreseen for the various stages of software design and implementation to reach a working system of adequate performance and reliability. The software design, planning and status documents should be refined with time, and progress should be scrutinised by the usual vetting bodies.

We suggest that the proposers should define and set up an organisational framework for software production within the experiment itself. This should consist of a co-ordinating group responsible for preparing a coherent plan for the entire software system of the experiment, plus an outline of the temporal evolution of the individual packages, with critical path analyses, milestones and so on. Some questions which should be addressed by this group are :

- *funding*; purchase of software development hardware?; purchase of design and development tools?; purchase of commercial packages?; fees for educational programs?

- *manpower*; distributed among remote sites or concentrated at the host laboratory, in which proportions?; how to attract and hire top quality software engineers?; contract out system design?; contract out code writing?; how to ensure continuity of effort and experience?

- *discipline*; responsibilities for software design, production and maintenance?; internal monitoring and vetting procedures?; quality control of software packages?; authority to impose conventions, standards and working practices?

Software System Design Environment

It is obvious (at least to Working Group 4-B !) that correctly functioning software for a SSC experiment cannot be produced by the methods in use at some of today's experiments. In fact, it is admitted that these methods are already inadequate and result in system incoherence which severely detracts from reliability and maintainablity. 'Down-time' (that is, the fraction of data taking lost during production runs because of system malfunctions) is sometimes much worse than the intrinsic hardware' dead-time', which is usually regarded as unacceptable if it exceeds more than a few per cent.

We believe that design methodologies already exist which greatly improve the quality and coherence of software systems and packages. These methodologies are in widespread use in commerial software houses and in the engineering divisions of many of our laboratories; however, they have not yet been adopted whole heartedly by the HEP physics community. We fully realise that available Computer Aided Software Engineering CASE tools often leave much to be desired, and that their inadequacies have contributed to their rejection in many instances. However, this is a field in which much progress in being made, and, on the time-scale of the SSC, a useable tool kit could and should be part of the basic equipment of all programmers in an experiment.

We point out a paramount need for the widest possible education and experience in the use of these methods. The SSC community will require real experts to participate in assessment of methodologies, evaluation of tool kits, organisation of formal design teams and so on. These jobs cannot be done in a climate of considerable ignorance, and in some cases even antagonism, which, we regret to observe, is roughly the situation today.

3. Some Proposals for Technical Studies

Having attended presentations by other Working Sub-Groups, we list below a selection of areas in which further technical studies are required, by inter-disciplinary collaborations, to elucidate design choices which will impact system software. These studies should take the form of creating scenarios and 'mini-designs', evaluation and appraisal excercises, setting up pilot projects, and so on, as appropriate to the area in question. The manpower, money and time required for a serious approach to these studies should not be underestimated.

- separate control and data paths for the read-out system?
- data-driven or instruction-driven read-out control?
- diagnostics in both the laboratory and production environments?
- system simulation and performance prediction?
- system validation and testability?
- use of expert systems/artificial intelligence?
- distributed processing techniques?

- information management and the use of databases?
- human interfaces to the experiment?
- the use of graphics?
- the choice of operating systems and programming languages?

Footnotes

We conclude with a few random comments, which we felt should be generally borne in mind when preparing for experiments at the SSC.

- The economic benefits of standards and conventions (whether in hardware or software) cannot be overemphasised; their use should be agreed and followed.
- Can the SSC community learn from (the mistakes of?) other large-scale projects such as NASA, CERN/LEP, the DOD, ???
- ***For obvious reasons, many physicists will be willing to work in a SSC experiment under almost any conditions. However, highly qualified and highly able technical personnel can certainly find job satisfaction elsewhere, not to mention better salaries and career prospects!***

  A real challenge for the SSC is to attract the best of these engineers and to keep some of them around to see the project through to completion. We note, with some apprehension, that HEP is no longer seen to be 'where it is at' by many computer science graduates and post-docs. If the field doesn't attract top talent, the achievements therein will be correspondingly second-rate.