Chapter 11

Front-End Electronics and DAQ

11.1 Overview

The CDF electronics systems must be substantially altered to handle run-II accelerator conditions. The increased instantaneous luminosity requires a similar increase in data transfer rates. However it is the reduced separation between accelerator bunches that has the greatest impact, necessitating a new architecture for the readout system.

Figure 11.1 shows the functional block diagram of the readout electronics. To accommodate a 132 nsec bunch-crossing time and a 5.5 μsec decision time for the first trigger level, all front-end electronics are fully pipelined, with on-board buffering for 42 beam crossings. Data from the calorimeters, the central tracking chamber, and the muon detectors are sent to the Level-1 trigger system, which determines whether a $\bar{p}p$ collision is sufficiently interesting to hold the data for the Level-2 trigger hardware. The Level-1 trigger is a synchronous system with a decision reaching each front-end card at the end of the 42-crossing pipeline. Upon a Level-1 trigger accept, the data on each front-end card are transferred to one of four local Level-2 buffers. The second trigger level is an asynchronous system with an average decision time of 20 μsec.

A Level-2 trigger accept flags an event for readout. Data are collected in DAQ buffers and then transferred via a network switch to a Level-3 CPU node, where the complete event is assembled, analyzed, and, if accepted, written out to permanent storage. These events can also be viewed by online monitoring programs running on other workstations.

With the new system architecture, both the Level-1 and Level-2 accept rates will be an order of magnitude larger than in run I. For example, with a 40 KHz accept rate from Level 1 and a 300 Hz rate out of Level 2, the system deadtime will be $< 10\%$. The expected rate of events written to mass storage is 30-50 Hz.

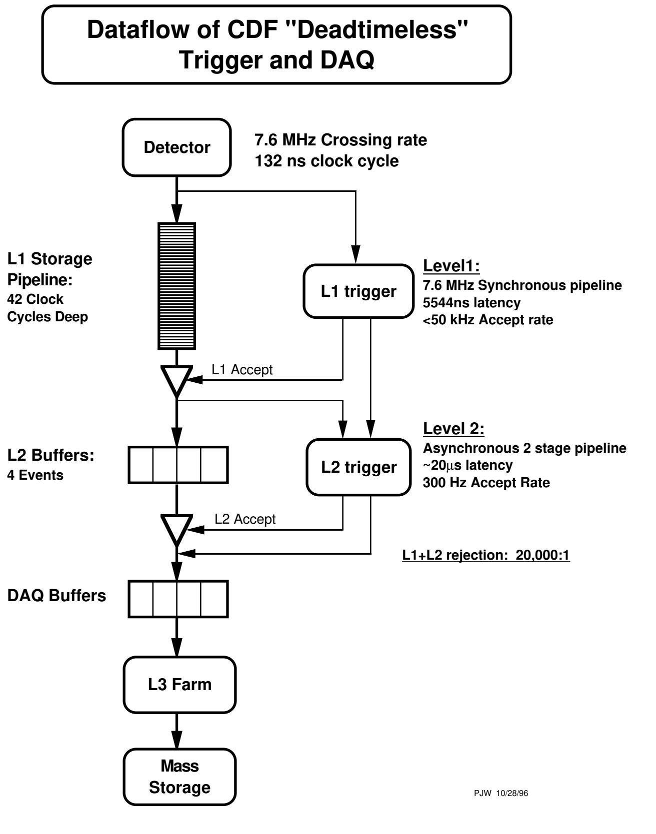

Figure 11.1: The run-II readout functional block diagram.

In this chapter, we begin with a detailed description with the front end electronics for the calorimeter, central outer tracker, and muon systems. We then describe the data acquisition system and the online computing. The readout of the silicon systems is detailed in Chapter 5 and the trigger is described in further detail in Chapter 12.

11.2 Calorimeter Front-End

The installation of new scintillator-based calorimeters in the plug region allows a common design to be used for much of the front-end electronics for the central and plug calorimeters and their associated preradiators and shower-maximum detectors. The calorimeter ADC readout is based on the QIE (Charge Integrating and Encoding) chip, a custom multi-ranging circuit developed at Fermilab for the KTEV experiment. A VME based front-end board carries the QIE, a commercial ADC, and FPGA based circuits to perform pipeling, buffering, and the creation of tranverse-energy sums for the trigger. A simplified version of the QIE design is used for all shower maximum and preshower detectors.

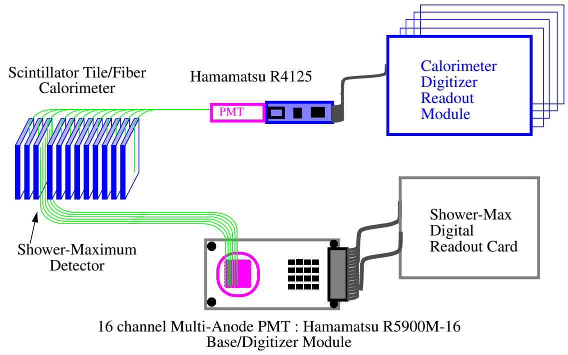

Figure 11.2: Plug-calorimeter readout.

11.2.1 Calorimeter ADC: The QIE

The readout of the calorimeters is done by integrating the charge coming from the individual calorimeter photomultiplier tubes (Fig. 11.2). The anticipated maximum signals from the PMT's will be greater than the present electronics were designed for. Estimates of the maximum signals for a collider energy of 2 TeV and an integrated luminosity of 2 fb^{-1} have been made for all of the calorimeters[1]. The conclusion is that the maximum signal expected is approximately a factor of two larger than that seen in the data up to now. The increase in maximum signal size requires one bit of dynamic range more than the present 16 bits. In addition, of course, the minimum time between bunch crossings will be reduced from 3.5 μsec to 132 nsec. We felt a multi-range solution was needed to solve the inherent difficulties of fast digitization of signals with large dynamic range. Because the signals from the PMTs can be as long as 80 nsec, there is not time to integrate, read out, and reset the integrating capacitor before the next bunch crossing, so the system also must be pipelined. These requirements fit reasonably well with an ASIC that was originally conceived for the SDC experiment at the SSC, and later adapted for use with the KTEV CsI calorimeter. The QIE (Charge Integrating and Encoding) custom integrated circuit[2] is used with a commercial ADC. The QIE and ADC are both mounted on a small front-end module, which is mounted in turn on an adc/memory (ADMEM) VME board. Each ADMEM board holds 20 of the front-end modules, provides the Level-1 trigger with transverse-energy sums, and also provides level-2 storage.

The QIE is the heart of the front-end module (Fig. 11.3). It has five major sections: a current splitter, a gated integrator/switch, a comparator, an encoder, and an analog multiplexer. The processing of signals within the chip is done in a four-stage pipeline. In the first step, the current splitter divides the incoming PMT current into 8 binary-weighted current outputs. The outputs are applied through a current switch to eight identical 2.6 pf integrating capacitors for a period of 132 nsec. The voltage integrated on the capacitor that receives half the charge (I/2 range) is the largest, with the voltage decreasing on the other capacitors down to a minimum at the last capacitor (I/256 range). At the end of the first step, the current switches to the next set of eight integrating capacitors. The capacitor voltages are then applied to a set of comparators to determine which capacitor voltage should be sent to the ADC for digitization. The selected range is encoded as a three-bit number, which is sent to the output for use as the exponent of a floating point number(the ADC value forms the mantissa). In the third step, the comparator outputs are used to control a multiplexer that selects one of the capacitor voltages to be digitized by the ADC. Following this, the capacitors are reset.

Tests were initially done using a version of the QIE that was the final preproduction version of the KTEV chip, fabricated using the Orbit 2.0μ process, along with an 8-bit Analog Devices AD9002 FADC. That chip has a full-scale charge input of 500 pC, and was not optimised for our power supply levels or clock speeds. The design was then modified to CDF specifications, which include the following: full-scale in-

Figure 11.3: Simplified schematic of the QIE ASIC

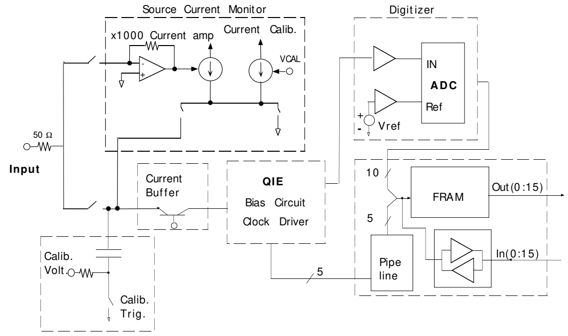

Figure 11.4: Simplified schematic of the front-end module.

put charge of 1300 pC, maximum instantaneous current of 50 mA, linearity better than .2% or 10 fC, whichever is larger, and temperature stability greater than .2% (or 10 fC, whichever is larger) per degree Celsius.

A schematic of the front-end module is shown in Fig. 11.4. It includes a current buffer, a charge-injection circuit for calibration, a radioactive-source current-calibration circuit, the QIE, a 10-bit ADC to digitize the voltage presented by the QIE, and digital memory for converting the floating-point QIE+ADC output to a dual-range linear 16-bit number. The input impedance of the QIE is not a constant 50Ω, which is not a problem in its original application where it was located very close to the PMT base. However in our application, the QIE will be some tens of nanoseconds from the PMT base, so the input impedance needs to be well controlled to minimize reflections back down the cable and subsequent mismeasurement of the signal. The design of the QIE is such that it is intrinsically difficult to control the input impedance to the level required, so an external, low power-consumption circuit has been designed for use between the PMT base and QIE. The performance of prototype circuits in conjunction with the QIE has been excellent, maintaining the required linearity, with a power consumption no greater than 50 mW.

The QIE has 32 integrating capacitors, 8 ranges for each of 4 pipeline depths. One possible source of error in the charge measurement is miscalibration of those capacitors. Several circuits have been designed and tested to inject a known charge into the QIE for calibration. The basis of the current design is capacitive discharge using a precision voltage source. The circuit must be quite linear and capable of injecting near full-scale charge inputs(≈ 1000 pC) in less than 132 ns. Initial tests were capable of measuring the middle ranges to the required accuracy, but large QIE input impedances for small inputs, and nonlinearities for large pulses made testing the circuit challenging for the low and high ranges(Fig. 11.5). The final circuit development will be done after the prototype CDF QIE is in hand.

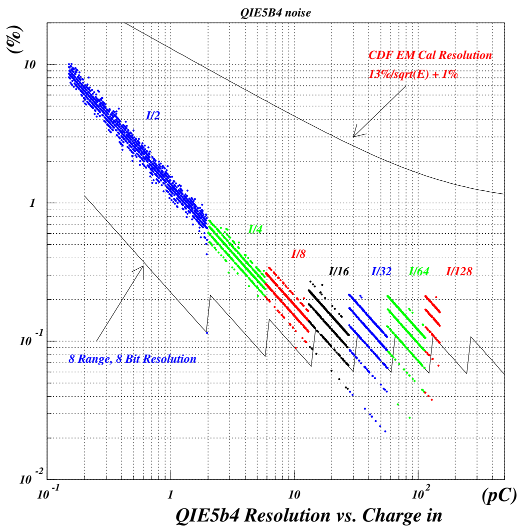

Figure 11.5: Noise performance of the prototype QIE5b4 ASIC compared with the calorimeter resolution and the ideal curve for an 8-range, 8-bit ADC system.

The other calibration circuit is needed to measure the DC current from the PMTs when a radioactive source is passed in front of the calorimeter scintillator. Both the plug and central calorimeters adjust tube gains based on this information, and both plan to use this information to carry calibrations from the test beam to the collision hall. The source strength is such that the typical current from a PMT is about 100 nA. A 1% measurement of the current is desired. The circuit uses a high-gain current amplifier and the QIE to read out the source current. A 100 nA current integrates only 13 fC of charge in 132 ns, and only half of that goes to the I/2 range. Using a 10-bit ADC gives 5.2 fC per count in the I/2 range, so a gain of several hundred will be needed to achieve the required accuracy.

There is also digital memory on the front-end module. The output of the QIE contains a 2-bit integrating-capacitor ID and a 3-bit range ID. With the 10-bit ADC output, that makes 15 bits per channel per bunch crossing. These bits will be converted to a linearized 15-bit number, plus a 1-bit range, in a 1-Mbyte Flash RAM. The calibration constants for doing the conversion will be determined using the charge-injection circuit described above.

The 16-bit output of the flash RAM is sent off the front-end module to a Xilinx field-programmable gate array (FPGA), which is also mounted on the ADMEM Boards. The Xilinx sums the digitized data into trigger towers and passes that data to another flash RAM which converts E into E_T and delivers a 10-bit number to the Level-1 trigger. The digitized data are also stored in the Xilinx for a period of 42 bunch crossings(about 5.5 μsec). That is the time required for the Level-1 trigger decision to reach the front-end crates. A Level-1 accept causes the data to be written into one of four Level-2 buffers, and a Level-1 reject causes the data at the end of the pipeline to be discarded. A Level-2 accept will cause the VMEbus Readout Controller(VRC) to do a block read of the data in each ADMEM card.

11.2.2 Shower-max detectors and preradiators

The shower-maximum and preshower detectors will also use the QIE technology. A simplified ASIC has been developed and tested that has lower resolution, an integral 5-bit flash ADC, and Level-1 and Level-2 storage on chip. The circuit has 8 ranges and sufficient matching of the integrating capacitors that channel-by-channel calibration constants will not be needed (Fig. 11.6). The chip uses the Orbit 1.2μ double-poly, double-metal process. The power dissipation per channel is about 100 mW. Rather than mounting the chip on a VME board, we will construct readout boards containing 32 to 80 channels that are connected to a VME shower-max readout board via flat cables, as shown in Fig. 11.7. The plug-detector outputs can go directly to the shower max QIE, but the gas-based central shower-max detector outputs must first go to a preamplifier circuit, described in the next section.

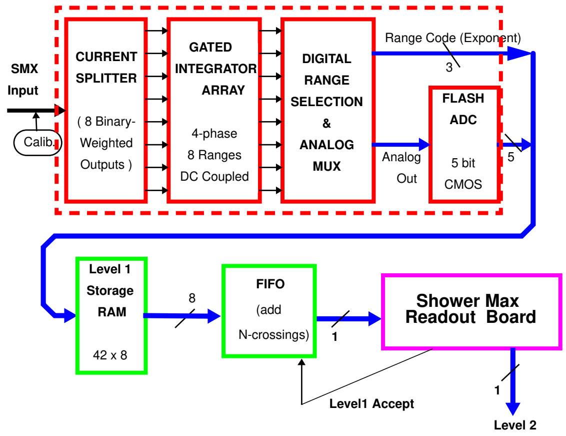

Figure 11.6: Schematic showing the general layout of the shower-max digitizer boards.

11.2.2.1 Central shower-max preamplifiers

The central-calorimeter shower-maximum proportional chambers measure the charge deposition on orthogonal strips and wires. In the run-I configuration, the Rabbit electronics matched the gain of the strip and wire signals, giving the same dynamic range to both. In the upgraded run-II configuration, we plan to use a preamplifier circuit to match the gains in the two views. As described in the previous section, separately-housed front-end cards will contain both the preamplifier circuit and the shower-max QIE. The digitized output from these cards will be multiplexed and sent to the DAQ electronics and to the trigger as shown in Fig. 11.7. We plan to mount the front-end card at the rear of the detector wedge as close as possible to where the cables emerge.

One of the key criteria for the selection of a preamplifier is the intrinsic noise of the circuit. Since the strip signal is sensitive to digital traffic, we want

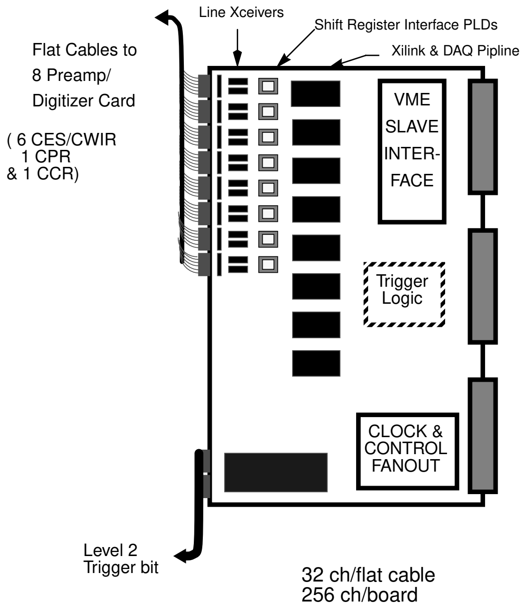

Figure 11.7: Schematic showing the general layout of the shower-max VME readout-boards.

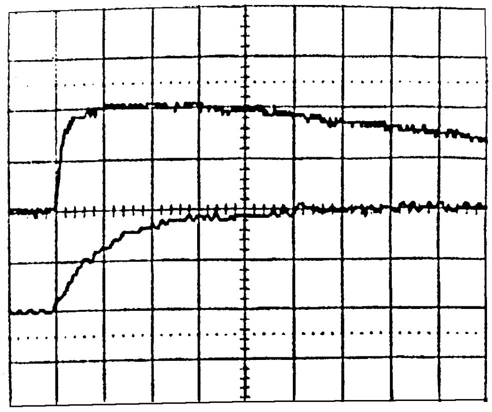

Figure 11.8: Time dependence of the amplified central-strip-chamber wire signal as observed. The top trace is 1μs per division, the bottom trace 200 ns per division.

our preamplifier circuit to introduce minimal additional noise. Recently, noise studies performed on the shower-max detector in the B0 collision hall showed that digital-traffic noise could be reduced by a factor 2-5 by using a different detector grounding scheme. This could impact where we mount the front-end cards, giving us more flexibility.

Another feature of the central strip and wire signals is the long signal rise time (τ_{rise}(wires)=530 ns).(Fig. 11.8) We plan to digitize the signals every 132 ns, but sum the signals from several consecutive windows on the VME readout board.

11.2.2.2 Central preradiator (CPR) and crack preamplifiers (CCR)

The central preradiator detector and central-crack chambers will also use the front-end cards containing both a preamplifier circuit and the shower-max QIE. Although the preradiator cabling is shielded twisted pair, the crack chambers have unshielded cable which will need to be replaced. We may add gain before the cable run (at the front of the wedges) for both the CPR and CCR; the preamp for the run-I VTX tracking chamber is a possible candidate.

11.2.3 Calorimeter TDC's

CDF has historically utilized timing on the hadronic calorimeter, and the need for this information continues into Run-II [3]. Timing requirements on hadronic energy deposition are important for the removal of cosmic-ray background. The cosmic-ray background for Run II will be comparable to that of Run I given the ADC and TDC gate widths and the luminosity. In addition to suppressing background, hadron-TDC information can be used to search for energy deposition from exotic heavy stable charged particles.

The baseline conceptual design of the CDF Run II Hadron-TDC upgrade is to pick off and discriminate the phototube dynode signals for the central, wall, and plug hadron calorimeters, and send the discriminated outputs to 96-channel pipelined TDC boards of the same kind used in the muon and tracking systems. The TDC design is described in the next section.

Although CDF has not used or considered timing on the EM channels in the past, we are now beginning to appreciate the importance of cosmic ray backgrounds to low rate signals from exotic processes involving photons. SUSY signatures in $ee\gamma\gamma\not{E}_T$, and $b\gamma\not{E}_T$, discussed in Sec. 2.4 are timely examples. The cost and schedule issues associated with TDC's on the EM calorimeter channels are under consideration, but will not be discussed further here.

	COT	Muon	Had	JMC96
Multi Hit Capability	yes	yes	yes	yes
Number of channels	30,240	8,773	1,200	40,213 total
Leading and trailing edge	yes	yes	yes	yes
Timing resolution	1 nsec	2 nsec	1 nsec	1 nsec
Max. drift time	100 nsec	2000 nsec	-	2048 nsec

Table 11.1: TDC requirements and specification

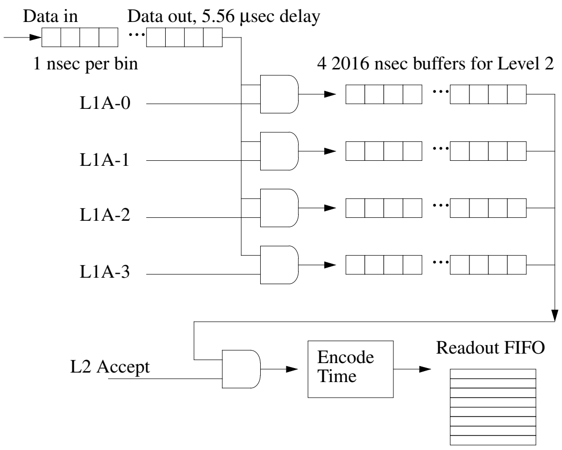

Figure 11.9: Schematic diagram of JMC96 chip.

11.3 Pipelined TDCs

The hadron calorimeters, scintillators, muon chambers, and the central drift chamber all require TDC's for recording signal times. The requirements for these systems, and the specification for the TDC are listed in Table 11.1. While there is some difference in the requirements for the three systems, there is enough commonality that a single solution is viable. In addition the TDC must obey the DAQ protocol of a 5.3 microsecond Level 1 buffer followed by four Level 2 buffers. The combination of these requirements were not available in any commercial module, and a custom circuit has been designed at the University of Michigan. The design includes the custom integrated circuit, 96 channel VME boards, and a calibration system.

The TDC chips, the JMC96, incorporate all of the features of the TDC required for the capture of the signals, the Level 1 and Level 2 buffering, and encoding data for readout. A functional diagram for the chip is shown in Fig. 11.9. The signal to be digitized is brought into the chip and is transmitted down a 48 stage delay cell line. Each delay cell has 4 transistors, and the delay per stage is controlled by a voltage. The overall delay as determined from this voltage is derived from a phase locked loop which is in turn driven by an external crystal controlled oscillator. Thus the absolute timing accuracy is determined by the accuracy of the crystal, which can easily be made to 100 parts per million. Once every 48 nanoseconds the content of the delay line is copied to a 48 bit register. This register then contains the history of the input signal for the last 48 nanoseconds. The register is then written to a static on-chip memory which is operated in a FIFO mode. The FIFO is large enough to hold 5.2 microseconds of data.

When a Level 1 accept is generated the data coming out of the FIFO is stored in one of four Level 2 buffers. Which buffer to use is specified in the Level 1 accept message. There are four level 2 buffers in the JMC96. Each buffer has forty-two 48 bit words, so that the level 2 buffers store up to 2 microseconds of data. The DAQ protocol specifies that Level 1 accepts may be generated on consecutive crossings. If this happens, then the output of the Level 1 TDC buffer will be written to multiple Level 2 buffers with a 132 nanosecond time offset. The minimum time between Level 1 accepts for the TDC chip is 96 nanoseconds. The Level 1 accept signal is recorded in a second 48 stage delay cell line, and the content of this line is written to one of four registers.

When a Level 2 accept is generated the data stored in the Level 2 buffers is digitized. The content of the Level 1 accept buffer is first read in order to calculate a t0. The Level 2 buffer is then scanned and the time of each transition from 0 to 1 or 1 to 0 is calculated. The time of the level 1 accept is then subtracted, resulting in a 12 bit word - 11 bits indicate the time within the 2 microsecond window, and the 12'th bit

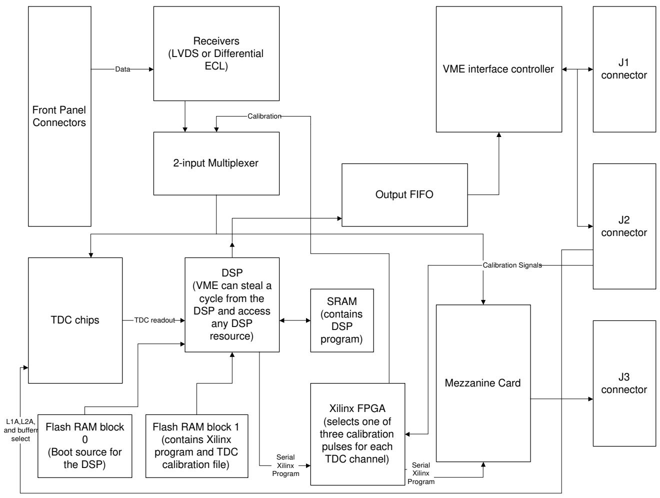

Figure 11.10: Block diagram of the 96 channel TDC boards

indicates a leading or trailing edge. The 12 bit word is then written into an on-chip 16 word FIFO.

The TDC board contains 96 JMC96 chips and the associated circuitry needed to receive and process the data. A block diagram of the board is shown in Fig. 11.10. The receiver is either ECL or LVDS. The board is manufactured with sockets for both ECL and LVDS receiver chips. It is most likely that the muon and scintillator systems will use ECL and the COT will use LVDS. This allows the boards to be manufactured in the same way and the final determination of the input voltage level is determined by which chips are plugged in. After the signals are received they are input to a 2-input multiplexor. The second input of the multiplexer comes from the onboard calibration circuit which can be programmed to produce up to three different calibration pulses and distribute them to any combination of the 96 channels. This calibration system will allow testing of not only the TDC board but also of muon or XFT trigger systems that get input from the TDC board.

A DSP on the TDC board is used for reading data out of the TDC chips and storing it in a memory. After a level 2 accept all 96 TDC chips encode the data in the level 2 buffer and store the hit times in a on-chip FIFO. The DSP queries the status of each TDC chip, and if it has data, reads it out. The DSP will then try to match leading and trailing edges of a pulse and convert the two edges into a single word that has the pulse width, the time of the leading edge, and the channel number. Several cuts can be made on the hits by the DSP. The pulses can be required to have a minimum pulse width before they will be readout. This can be used to reduce the number of noise hits and reduce the size of the event record. A limit on the maximum drift time can be set. The TDC records 2 microseconds of data, which is much longer than the 100 nanoseconds drift time of the COT. A cut on the maximum time removes hits from previous crossing. A cut can be placed on the maximum number of hits per channel in order to suppress noisy channels, and a channel can be marked a bad or dead and excluded from the readout entirely. When all channels for an event have been readout the header word, a word count, and the timing data are written to a FIFO on the TDC's. The data is read out of the other end of the FIFO through VME.

The COT TDC's will be mounted on the detector. Fifteen VME crates are needed to house the 315 TDC cards. The signals from the COT are transmitted from the chamber-mounted ASD's over small coax. There may have to be a transition card to a lower loss cable before the signals are sent to the TDC.

The TDC board has connectors to support a mezzanine card. The mezzanine card will be used for processing or transmitting the timing signals to trigger circuits. Different systems will have different mezzanine cards. The COT mezzanine card has circuits to identify the prompt and delayed hits for each cell and multiplex the hits to connectors on the backplane. The signals will then be transmitted by the existing Ansley cable to the trigger room for use by the XFT track processor.

11.4 Muon-Detector TDCs

The muon system share the TDC design with the COT, as described above. The muon TDC's reside on the first floor counting room. Signals are transmitted from the detector via the existing Ansley cable. The TDC receives the signals and records the hit times. A mezzanine card mounts on the TDC board which generates a copy of the muon signals to be transmitted to the muon trigger. A total of about 91 TDC modules are required for the muon systems.

11.5 Silicon Front-End Electronics

The design of the front-end and DAQ electronics for SVX II is intimately connected with the detector's mechanical design because of the need to minimize the amount of material between the beamline and the COT. Consequently, the SVXII and ISL electronics is described along with the SVXII detector in Chapter 5.

11.6 Data Acquisition System

11.6.1 Overview

The Data Acquisition System is responsible for collecting data fragments from front-end electronics systems for events satisfying the Level-2 trigger and sending them to the Level-3 trigger subsystem. There complete events are built and more sophisticated algorithms classify events and determine whether they should be saved. Events passing the Level-3 trigger criteria are sent to mass storage. The Online Computing System, described below, provides processes that control and monitor the functioning of the detector and data acquisition system.

A block diagram of the system is shown in Fig. 11.11. The basic architecture is very similar to that used successfully in Run Ib [7]. Front-end and trigger electronics are housed in VME crates replacing the FASTBUS and RABBIT crates used in the original detector. Timing signals associated with the beam crossing are distributed to each crate by the Master-Clock subsystem. Trigger decision information is distributed by the Trigger-System-Interface subsystem. Commercial processors read data from modules in their local crate and deliver it to the Event-Building subsystem. This system concentrates the data and delivers it to the Level-3 trigger subsystem through a commercial network switch. The Level-3 trigger is a "farm" of parallel processors, each fully analyzing a single event. The Data-Logging subsystem delivers events to mass storage and also to online monitoring processes to verify that the detector, trigger, and data acquisition system are functioning correctly. While the architecture is similar to the run-Ib system, many individual components will be upgraded or replaced. This is necessary to deal with the new VME-based front-end electronics, and to provide increased throughput required for the higher luminosity now expected in run II. And in some cases commercial products in the run-Ib system are no longer supported by the manufacturer and must be replaced.

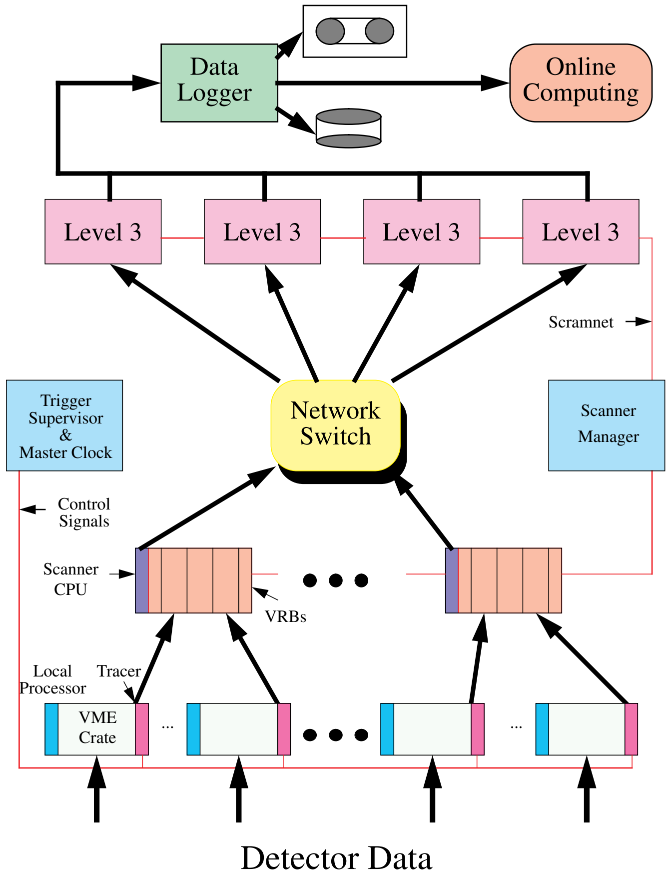

Figure 11.11: A schematic of the CDF Data Acquisition system, showing data flow from the front-end and trigger VME crates to the Online Computing system.

11.6.2 Front-End Crates and Controllers

All front-end and trigger electronics are packaged as VME modules and housed in standard 21-slot commercial VME crates. These modules process detector signals and make the data available to the data acquisition system via the VME bus. In addition to the front-end modules, each front-end and trigger crate will contain a commercial VME-based processor for event readout, and a "TRigger And Clock + Event Readout" (Tracer) module. The Tracer receives serial fast control messages such as Level-1 and Level-2 Accept from the Trigger Supervisor and fans them out on dedicated lines on the backplane. It also provides the interface to the serial data link to the Event-Builder subsystem.

11.6.2.1 VME Crates and Modules

Modules and crates will conform to VME standards including the VME64 extensions (ANSI/IEEE 1014-1987 and ANSI/VITA 1-1994, VITA 1.1-199x) [8]. Among these extensions are 64-bit transfers, additional ground and user-defined backplane pins, better EMI protection, geographical addressing, and board keying. They will also conform to the proposed VME International Physics Association (VIPA) standard (VME-P) [9] that defines additional specifications for high-energy-physics applications. These include crate and card mechanics for 9U x 400 mm systems, including rear transition cards as well as additional power required for analog or other custom applications.

Of the approximately 120 crates in the system, half each are located on the detector and in the counting room. User-defined pins on the standard J2 backplane will be used to distribute fast control signals such as Level-1 Accept, Level-2 Accept, and the crossing clock to all modules in a crate. Most systems will employ a custom J3 backplane. In some cases this will be used to receive signals from the detector via a 9U x 120mm rear transition card. Another use of J3 will be communication of trigger information to adjacent boards where required. The newly defined J0 connector will be implemented and used primarily to provide additional power.

In addition to following the VSO standards, front-end cards are required to present a common interface to the data-acquisition system. They receive fast timing and control signals via specific backplane pins. On receipt of a Level-1 accept, they must store the data for the appropriate crossing into the specified one of the four Level-2 buffers. Then on receipt of a Level-2 accept, this data must be readable over VME by a single D32 block transfer. Data from each module will include a header word containing a module ID and event number for use in data integrity checks higher up in the system.

All crates are specified to have 300 amp +5V power supplies. Additional power is system dependent. Some systems require separate supplies for sensitive analog circuits, others additional voltages for ECL or 3.3V applications.

Crates both on the detector and in the counting rooms will be individually cooled by fans and air-water heat exchangers. These will be configured to provide cooling to rear transition cards as well as the main crate. The cooling system will be designed to support a maximum power dissipation of 2 KW per crate. Most crates will be well under this limit. Racks in the counting rooms will contain 2 VME crates each. Protection circuits in each rack will automatically shut off power in case of fan failure, overtemperature, or detection of smoke or water.

11.6.2.2 Local Processor

Each crate will contain a local VME-based processor. On receipt of a Level-2 accept, this processor will read data from the front-end modules and send it via the Tracer module to the Event-Building network. Also, the processor provides the interface between the host control programs and the front-end modules for initialization and diagnostics.

The event data readout may be accomplished in one of two ways. In the direct mode, the front-end data are transmitted to the Tracer as they are read from the modules as described in more detail below. In the second mode, this Tracer "spy" mode is disabled, and data are first read into the processor. The processor then applies calibration corrections and removes channels where there are no hits. The compressed data are then written back to the Tracer data link. For the calorimeter crates, a significant reduction in data size may be accomplished by removing the empty channels. Benchmarks have shown this compression can be accomplished within the allotted time.

The processor will also provide the only external path into the crate for downloading and monitoring operations; there will be no other intercrate network. Commands and data are sent from the host to the local processor via ethernet. The required VME operations are done, and the results communicated back to the host. Commands may be single VME operations, or more complex operations such as an entire begin-run initialization sequence.

Currently the processor is specified to be a Motorola MVME-162 This is the choice due to both its excellent performance as a VME master and its modest cost. The processors will run the VxWorks operating system [10]. VxWorks provides a multitasking operating system with priority-based scheduling and fast interrupt response. It also provides excellent networking including BSD sockets, remote file access via nfs, and remote logins. Code is developed and compiled on a host unix system, then downloaded to the processor. Since the processor form factor is 6U x 160mm, a 6U-9U adapter card will be used to install the modules in the crates.

11.6.2.3 Tracer Module

The Tracer module provides the interface between the front-end modules and the Trigger Supervisor and Master Clock for fast control signals. It also provides the data interface between the front-end modules and the Event Building Network.

Serial control messages from the Trigger Supervisor (Level-1 Accept, Level-2 Accept and Reject, Control) are received and decoded by the Tracer. In addition, timing signals from the Master Clock (Beam Crossing, Bunch 0, Abort Gap, and Clock) are also received by the Tracer. All these signals are distributed via the J2 backplane to all modules in the crate.

To provide maximum efficiency for event readout, the Tracer contains a "spy" function that enables it to clock event data from the backplane as the VME Readout Controller reads it. The Tracer stores the event data in a local FIFO. A sequencer on the Tracer will then read the FIFO, serialize it, and then transmit the data to the Event Building Network. Should the higher level not be able to accept data (due to its internal buffers being full) a BUSY signal will be transmitted and inhibit data transfer until the condition clears.

Finally, the Tracer provides a means by which to issue a crate reset. The VMEbus signal SYSRESET* will be pulsed low on the backplane if a reset pulse is detected from the CDF Reset Crate. The Reset Crate contains a set of IndustryPak output registers that drive RS-422 signals to a front-panel connection on each Tracer. A signal is sent from the Reset Crate to a Tracer if a reset is desired. This capability is required in case a local processor board hangs.

11.6.3 Trigger System Interface

The Trigger-Interface subsystem provides the interface between the hardware trigger (Levels 1 and 2) and the data-acquisition system. Trigger decisions are received by this subsystem and forwarded to the front-end electronics. Then, when a Level-2 trigger has been successfully processed, the Event Building Network is informed and manages readout of the event into the Level-3 trigger. Components of this subsystem include a set of Trigger-Supervisor modules, one per partition, and a set of crosspoint modules that link signals from the Trigger Supervisors to the front-end crates in the appropriate partition.

11.6.3.1 Protocol

The global-trigger-decision subsystem will send results of the Level-1 trigger to the Trigger Supervisor each beam crossing. In addition, when a Level-2 decision is available, this is also sent to the Trigger Supervisor. The information is forwarded to the front-end crates in the form of serial messages.

Each beam crossing, two frames of 9 bits each are transmitted to each Tracer card, which decodes the messages and drives the appropriate lines on the backplane. The first frame contains the results of the Level-1 decision for the crossing that is emerging from the front-end pipeline. If the decision is positive, the data is stored in one of four Level-2 buffers as indicated by another field in the frame. Otherwise data from the crossing is discarded.

The second 9-bit frame may be either a Level-2 decision, a Calibration enable, or Control message. The Level-2 decision word indicates which is the pertinent buffer of the four, as well as a reject or accept indication. If the event is accepted, the data is moved from the Level-2 buffer to a buffer awaiting readout by the DAQ system. The Calibration Enable message causes generation of a calibration event by previously enabled pulsers in the crate. The Control message is used to halt, reset, and start the Level-1 pipeline so that all crates are synchronized. It also includes an encoded field that may be used to alter the default readout, for example in the case of calibration events.

Information is returned from the Tracer to the Trigger Supervisor via four (nonserial) cables. These include Done, which is deasserted on receipt of Level-2 Accept and reasserted when the front end is ready to accept further triggers. Error is asserted if some fatal error is encountered. In response, the Trigger Supervisor will halt and reset the Level-1 pipeline. Should there be no place to store data from a Level-2 accept due to a backup from higher levels in the data-acquisition system, then the Wait signal is asserted. Level 1 Done is issued following a Level-2 reject to indicate the specified buffer is ready for another Level-1 accept. This is required only in the SVX system where there can be a delay of many beam crossings before a capacitor cell may be reused following a Level-1 accept.

The system will be clocked every 132 ns even when there are only bunches present every 396 ns, as expected during the first part of Run II. Level-1 rejects will be issued on empty crossings. When the situation arises that there are no free Level-2 buffers, Level-1 rejects are issued, thus incurring dead time. This can occur due to the time required for Level-2 trigger processing and data-acquisition system processing of Level-2 accepts. The Trigger Supervisor will keep track of the live fraction and provide a Live Gate to external scaler modules to enable proper computation of the live integrated luminosity.

11.6.3.2 Trigger Supervisor

The Trigger Supervisor will be a custom 9U x 400 mm VME module. A series of state machines will manage the Level-1, Level-2, and readout sequences. These will be implemented using Xilinx gate arrays. A custom J3 backplane will connect the Trigger Supervisors to the Crosspoint modules. Should an error condition be encountered such as failure to receive an expected message from the trigger or front end, status information will be stored in the module. A task in the local crate processor will note the condition and inform the Run-Control program.

11.6.3.3 Crosspoints

Two different types of crosspoint modules route signals between the Trigger Supervisors and front-end crates. For the serial data, custom VME modules based on the Triquint TQ8016 chip are used. There are 16 channels per board. Each output is attached to a single front-end crate, and may be individually programmed to select data from any Trigger Supervisor as input. The output messages are transmitted via fiber-optic cables. In addition, each channel has a programmable delay to account for differences in propagation time to different detector components. For the signals returning from the front-end crates (Done, Busy, Error, Level-1 Done), the use of commercial latches is under consideration.

11.6.4 Master Clock

The Master Clock subsystem is responsible for distributing timing signals to the front-end electronics and trigger. Four signals are distributed to all systems: Beam Crossing, Bunch 0, Abort Gap signifying the current crossing is in the accelerator abort gap and so there is no collision, and Clock (every 132 nsec). The Beam-Crossing signal is used to gate front-end ADCs and TDCs and should be stable with respect to the beam crossing to better than 1 ns. The Bunch-0 signal is used for synchronizing the front-end pipelines following a reset operation. As it indicates there is no beam collision, the Abort Gap signal is used by the SVX system to reset the front-end amplifiers to baseline. This information is also used to occasionally trigger pulsers to generate calibration triggers during a data-taking run. In addition, the SVX will use a 53 MHz clock derived from the Tevatron RF system to clock serial data links.

A variation on the Master Clock system developed for the D0 experiment [11] will be used. This consists of three module types. The first, the Phase Coherent Clock (PCC) module phase locks a 53 MHz oscillator to an RF signal from the Tevatron. This, together with a marker signal, is input to the Master Sequencer module. The sequencer module drives a number of output signals whose state at each of the 1113 RF buckets is determined by a lookup memory. Groups of four of these outputs will be configured as the base set of timing signals for each detector system. These signals are fed to Selector/Fanout modules where they are distributed to the various crates within a subsystem. These modules can add a delay in nanosecond increments for each channel to fine tune the timing for each channel.

Past experience has shown medium-term drifts in the Tevatron input signals relative to the beam crossing time of several ns. This drift was measured using precisely-timed information from the detector beam counters, and manual adjustments made to compensate for the drifts. For the Run-II system, a more automated feedback mechanism will be developed.

11.6.5 Event Building Network

The Event Building Network takes data from the front-end crates and delivers it to the Level-3 trigger where the fragments are assembled into complete events. First the data are received by VME Readout Boards (VRBs) distributed among 11 VME crates. A processor in the VRB crates ("Scanner CPU") reads data from the VRBs in that crate and sends it via a data switch to the Level-3 subsystem. Data flow is controlled by a Scanner-Manager task running in a separate processor. It communicates with the Level-3 processors and Scanner CPUs over a separate control network so as to not interfere with the event data traffic.

11.6.5.1 VME Readout Boards (VRB)

Rather than going directly into the data switch, the serial data from the front-end crates are received by custom VME Readout Boards (VRBs). The amount of data per event ranges from a few hundred bytes to a few kilobytes over the different crates in the system. This extra VRB layer combines data from different front-end crates so as to approximately equalize the data size on each switch link. It also reduces the number of switch ports required by over a factor of 10. This greatly reduces the cost and complexity of the switch as well as improves switch utilization.

The VRB module contains a number of serial data inputs together with a set of dual ported memory buffers for each. An event fragment received via the data link is placed in one of the buffers. It is held in the buffer, available for later asynchronous readout via the VME slave interface. Buffers are managed in a FIFO manner, input of one buffer and VME readout of another may occur simultaneously. If a channel cannot accept more data, a BUSY signal is transmitted back to to the Tracer module to halt flow of data. This will also be transmitted to the Trigger Supervisor. Data from all channels for a given event may be read by a single VME64 block-read operation to minimize the overhead.

For all systems except the SVX, slower 256 Mbps TAXI links will be used rather than the 1 Gbps GLinks used by the SVX. This is because data are only transmitted on every Level-2 accept rather than every Level-1 accept. Also, the slower links allow additional multiplexing to give 10 rather than 5 data links per board. The data links are implemented on rear transition cards, allowing the core part of the module to be the same for all systems.

11.6.5.2 Scanner CPU

The Scanner CPU is a commercial VME-based processor that reads all fragments for an event from all VRBs in its local crate and sends it to the Level-3 trigger via the Network Switch. In addition to a high bandwidth VME64 interface, these processors must also support an interface to the network switch. The MVME 162s used in the front-end crates only have a VME interface. In this application the switch interfaces would also have to be VME slaves, meaning the data would have to travel over the limited bandwidth VME bus twice. To avoid this situation, current R&D has focused on Power-PC-based processors with Peripheral Component Interconnect (PCI) buses [12]. PCI is a synchronous 32-bit 33-MHz bus with a total bandwidth of 132 MB/sec. The network interface connects directly to the PCI bus, hence the system throughput is not limited by the VME bus. Evaluation of these and other modules will continue. The Scanner CPUs will also run the VxWorks operating system.

11.6.6 Network Switch

Data collected by the Scanner CPUs is sent to the Level-3 system through the Network Switch. Data flow is managed such that event fragments are flowing in parallel from all Scanner CPUs to all Level-3 systems simultaneously. The proposed configuration in Table 11.2 shows that each link must sustain an average of about 5 MB/sec for 300 Hz operation. This is the throughput that must be sustained including all software-driver and data-flow-control overheads and hardware-link inefficiencies. This is usually less than the peak data rate on the physical link. About 45 MB/sec is the required aggregate bandwidth for 300 Hz operation.

The Ultranet system used in Run Ib is no longer supported by the manufacturer. It also does not have sufficient performance to deal with the luminosities now expected in Run II; hence it must be replaced. A number of other commercial options developed for telecommunications or computer-networking applications exist and are under study. These include Asynchronous Transfer Mode (ATM) [13], FiberChannel (FCS) [14] and HIPPI [15] among others.

Currently a detailed evaluation of ATM technology for this application is in progress at Fermilab. In ATM, data blocks are broken into cells of 48 bytes plus a 5-byte header for transmission. A test system with an 8-port ATM switch from Fore Systems [16] has been constructed. Data links are 155 Mbps. Eight Power-PC processors with PCI-based ATM interfaces are used as data sources and sinks. When faster 622 Mbps links are available, they will be evaluated as well. Evaluation of other technologies at other labs [17] will be monitored and serve as input to the final technology choice.

System	Crates	Total Bytes	Switch Links
Calorimeter Total	44	30544 (6800)	2(1)
Central Calorimeter	24	16704(2600)	
Endwall Calorimeter	4	2128 (500)	
Endplug Calorimeter	16	11712(3700)	
Central Tracking	18	27000	2
Muon System Total	9	2000	1
SVX II	3	40000	3
ISL	2	16000	1
Trigger	42	30000	2
Total	120	148204(122800)	11(10)

Table 11.2: Summary of the data acquisition configuration. For the estimated data size, a luminosity of $2 \times 10^{32}/cm^2/sec$ is assumed. Quantities in parentheses are following data compression by the front-end processors.

11.6.6.1 Event Data Flow Control

Data flow between the Scanner CPUs and Level-3 system is controlled by the Scanner Manager. This is a task running on an independent VME-based processor. Communication with the Scanner CPUs and Level-3 processors is over an independent reflective-memory network.

After the front-end processors have completed readout of an event, the Trigger Supervisor passes the Level-2 accept message to the Scanner Manager. Level-3 processors, when they become active, make their presence known to the Scanner Manager. The Scanner Manager then directs the Scanner CPUs to read the events from the VRBs and send the fragments to a specific processor in the Level-3 system. This is done in such a manner so that multiple Scanner CPUs are not sending to the same Level-3 port at the same time so as to improve switch utilization. For ATM, this can be accomplished by making use of the rate division capabilities built into the hardware. For other switch types, messages between the scanner manager and scanner CPUs can be used. While this centralized strategy results in more message traffic, it allows better control of the traffic distribution among the switch ports and Level-3 systems. In the Level-3 system, events are built by dedicated tasks then queued to the algorithm task.

The Scramnet reflective-memory network [18] used in Run Ib will continue to be used for message transmission. This network provides a very low overhead communication path that does not interfere with data traffic. Each Scanner CPU and Level-3 system in the Event-Building network contains a Scramnet module. Short control messages are written to the memory by the local processors, and automatically transmitted ("reflected") to the analogous memory modules in all other crates. This transfer incurs no overhead beyond that of a simple VME single-word write, plus approximately 1 microsecond per network node.

11.6.7 Level 3 Trigger

The Level-3 trigger subsystem receives event fragments from the Scanner CPUs and builds complete events into the appropriate data structures for analysis. The events are then passed to a trigger algorithm that categorizes the event and makes the decision as to whether it is written to mass storage. Events destined for mass storage and also other event categories requested by online monitor programs are passed on to the Data-Logger subsystem. During the event building process, the event integrity is checked and complaints are issued about mismatched fragments or other problems. The same algorithm is run on all processors. This method does not require careful tuning of the event distribution to maximize processor utilization. Also events satisfying multiple Level-2 triggers require no special handling as would be the case if events were allocated to processors based on trigger type.

The Level-3 algorithms take advantage of the full detector information and improved resolution not available to the lower trigger levels. This includes full

3-dimensional track reconstruction and tight matching of tracks to calorimeter and muon-system information. Results from the lower levels are used to drive the algorithms. For example, if the Level-2 trigger indicates only the presence of an electron candidate, only those portions relevant to verifying it are invoked. In previous runs the overall rejection factor has been approximately 80%. For Run II, use of SVX information and targeting triggers at more specific physics analyses will be used to increase the rejection.

The overall algorithm structure is based on the offline analysis package. This is a modular system consisting of a number of general reconstruction modules and separate filter modules for specific triggers. Algorithms are provided by various detector and physics groups and integrated into a final package by the Level-3 group. As the offline system is expected to evolve for Run II, this evolution will be carried over into the Level-3 algorithm structure. This includes evolution in both the general framework and the programming languages supported.

The system operated in Run I contains four Silicon Graphics Challenge XL systems each with 8 processors and 4 VME bus interfaces. The baseline plan is to replace the existing processors with the fastest available at the time they must be purchased. This takes advantage of the large investment in the Challenge-system classes. In addition to raw processing capacity, the input I/O capability is very important. Assuming a 50% increase in processing per event compared to Run Ib, approximately 45,000 MIPS will be required to meet the 300 Hz input rate goal. It is expected that the price/performance of these processors will improve such that this upgrade can be supported within the current budget. Of course as reconstruction programs are written for new detectors and trigger algorithms developed this estimate must be continually evaluated.

Due to the extended time scale for Run II, support for the Challenge platform by Silicon Graphics could be an issue, as they will likely be superceded by new systems in the next few years. The marketplace will be continually monitored during this time. If a more cost-effective solution than upgrading the Challenges is found, that could be adopted instead. Note that in addition to the raw processing speed requirements, any alternate solution must be configurable with of order 16 high-bandwidth interfaces to the data switch as well as links to the Scramnet network. Any alternate platform must also be supported by the CDF offline system in order to develop trigger algorithms.

11.6.8 Data Logging

Events passing the Level-3 trigger are delivered to mass storage by the Data-Logging subsystem. Also a sample of events are sent to monitoring processes in the online computing system. The default strategy is to send all events to a dedicated data-logging/monitoring machine. The expected event rate is around 30-50 Hz corresponding to 7.5 - 12.5 MB/sec. This corresponds to a trigger cross section as high as 250 nb at 2×10^{32} luminosity. An alternative would be to log events directly from the Level-3 processor systems. However this method makes calculation of the integrated luminosity for a run more difficult in case of a crash or other failure in a part of the Level-3 system. This is because events may not be evenly distributed across all systems; hence the fraction of luminosity seen by a particular part is not easily determined.

In Run Ib, the same Ultranet network used for event building was also used for transmission of accepted events to the data logger. Although the output data volume from Level 3 is much smaller than the input, it still disrupts the traffic in the switch, making it more difficult to minimize contention. Hence a separate network for accepted events will be implemented. Since the required aggregate bandwidth is much less than for the Event-Building network, a less costly loop topology rather than a switch topology should be sufficient. Several commercial possibilities exist. A dedicated FDDI ring may barely have sufficient bandwidth. Another possibility is FibreChannel Arbitrated Loop [19] This logically is similar to FDDI but supports much greater data rates. Other possibilities includes Serial Storage Architecture (SSA) [20] and FireWire [21]. The commercial marketplace will be tracked and a technology choice made closer to the start of the run. A FibreChannel test-bed is begin constructed at Fermilab to evaluate its use for data logging.

Also no decision has been made on the mass storage medium. For reliability reasons it is very desirable to migrate from the 8mm tape used in the past to a different technology. New technologies that would meet the requirements include Digital Linear Tape (DLT) [22], and the RedWood tape [23]. For operational convenience, the possibility of sending raw data directly to the computing center will be considered. Depending on the future cost of disk storage, it may be feasible to leave the data on disk until the offline production analysis has run. Then the production output and raw data may be written together to mass storage.

11.6.9 Partitioning

A key element of the system is the concept of partitioning. This is the ability to operate different parts of the data-acquisition system in parallel to facilitate debugging. No recabling is ever required to accomplish this. The system is designed to support a maximum of eight partitions. Software in the host control programs controls allocation of resources, ensuring that multiple partitions do not access the same front-end crates.

The lowest-level granularity is a single front-end VME crate. This makes it possible to repair a faulty crate within a detector system while continuing operation of the remainder of that detector. An exception is the main physics trigger, which may not be divided among multiple partitions. However simple triggers appropriate for calibrations may be routed by a multiplexor to any partition.

A Trigger Supervisor is assigned to each partition. Crosspoint modules route the control messages from the partition's Trigger Supervisor to the selected set of front-end crates. Software in the Event Building Network and Level-3 trigger supports multiple partitions. A specific portion of the Level-3 system is allocated to each partition. This method is chosen rather than sending events from any partition to any Level-3 processor and routing the output based on partition. Different Level-3 algorithms may thus be used in different partitions; also throughput is not affected by a different partition generating a very high trigger rate.

11.7 Online Computing System

The Online Computing System is responsible for configuration, calibration, monitoring, and diagnostics for the data-acquisition system and front-end and trigger electronics. It also provides monitoring of a sample of the data stream to insure that the detector is functioning correctly. And finally it provides for control and monitoring of the detector systems such as high voltages and VME crate power. Each of these subsystems provides an operator interface, as well as automated error detection and reporting.

Data-acquisition control and monitoring is provided by unix workstations connected via a local-area network to all embedded processors in the system. Event monitoring programs also run on unix workstations, receiving data either directly from the data logger or via the local network. Slow control and monitoring of the detector systems is provided by the ALARMS subsystem. This is tightly integrated with the accelerator control system (ACNET). It is distinct from the data-acquisition control system in both hardware and software, though interfaces are provided to transfer information between the two.

A diagram of the logical architecture is shown in Fig. 11.12. Shown are the Run Control, Event Monitoring, and ALARMS subsystems and their relationships to the Data Acquisition and Offline systems.

11.7.1 Run Control

The Run-Control subsystem provides the following functions:

- Primary user interface to the data-acquisition system. This includes starting and stopping runs, viewing the system status, etc.
- Initialization of all front-end and trigger hardware.
- Configuration of the data-acquisition system.
- Automatic startup of event monitor tasks.
- Any additional procedures required to perform calibrations.
- Monitoring of system performance parameters.
- Reporting of any errors in any part of the system and taking action to recover if appropriate.

Each partition runs a separate copy of the Run-Control program. A partition manager allows the user to select parts of the detector while preventing multiple partitions from accessing the same pieces. For initialization, configuration information such as the correspondence between physical modules and logical detector components, hardware trigger tables, calibration constants etc. is retrieved from various databases. A finite-state-machine model is applied to the components of the system and the system as a

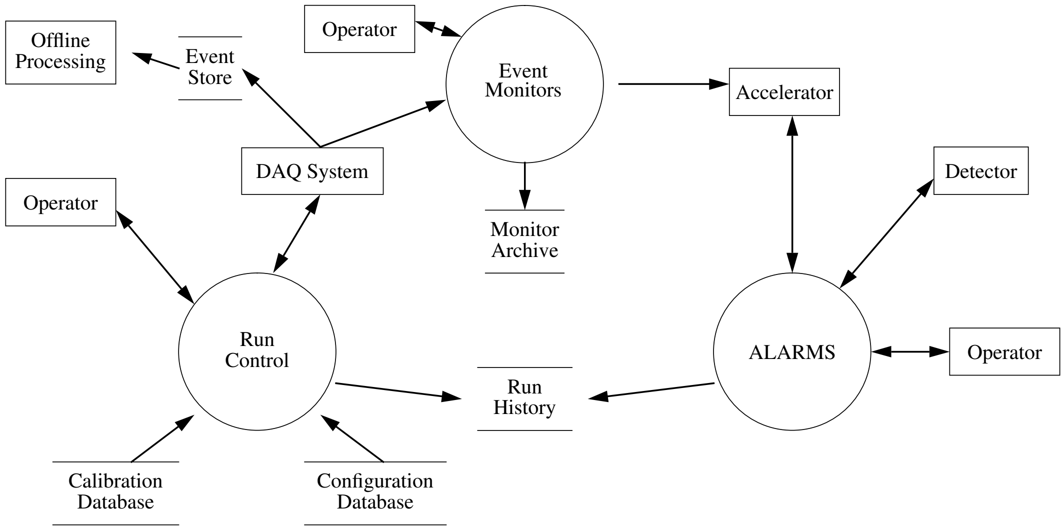

Figure 11.12: Logical model of the Online Computing system.

whole. A state manager communicates transition requests to all components and collates the replies into an overall system state transition. Error messages from all components are received by a centralized message server and displayed. Should data taking be stopped due to some error, the operator is notified and a recovery procedure may be attempted.

The command-line-driven VMS-based program used in the past will be replaced by a unix version with an X-window-based graphical user interface. It will be location independent (with appropriate security) to allow for remote diagnosis of system problems. The overall design will follow an object-oriented model; this maps well to components of the data-acquisition system. A modular design will be followed to allow easy integration of various detector components.

Higher level code will be written in an object-oriented language. Although C++ is perhaps the most obvious candidate, scripting languages including object-oriented extensions to tcl/tk [24] such as itcl [25] should give adequate performance for most high-level functions while being more straightforward to program. In addition the tk toolkit provides an easier way of generating graphical user interfaces than Motif or related C++ toolkits. Also under evaluation is the Java[26] language recently developed by Sun Microsystems. This is an object-oriented language that lacks many of the difficulties and pitfalls associated with C++. It includes class libraries for GUI generation and several networking protocols. Although currently not very mature, it is rapidly developing and shows some promise for this application. Lower-level code including all code residing in VME-based processors will be written in C. Fortran will not be used within the Run-Control system.

The calibration database developed for the offline system will also be used for storage of online calibration data. Other databases are required for various types of configuration information, and run summary information. No decision has yet been made on the implementation of any of these. Commercial relational and object-oriented databases will be evaluated. The dart information server (dis) [27] developed at Fermilab will be considered for configuration information. For some configuration information, flat files may be adequate.

Communication with the large number of front-end processors is an issue. For random access (diagnostics, monitoring), TCP/IP will be used. However state transitions such as begin-run initialization must be performed in parallel to minimize the time required. Hence some multicast protocol will be necessary. One existing package that may meet these requirements is the dart multicast services (dms) [27]. Other possibilities that will be evaluated are SNMP, CORBA event services, and the Network Data Delivery Service (NDDS) [28].

All source code will be maintained in a code management system such as the Concurrent Version System (CVS). As the requirements for code distribution are more relaxed than for the offline system, and the number of developers is much fewer, packages other than CVS that provide more functionality will be considered. Versioning will be done for all component packages and the Run-Control package as a whole. An Integrated Developer Environment (IDE) will be used for development of at least the core portion of the system. These generally include a graphical debugger, editor, interface to the code management system, class browser, and automatic documentation extraction in an integrated package. CASE tools which aid in development of the overall system object model are also under evaluation.

11.7.2 Event Monitoring

A sample of events, which may or may not be destined for mass storage, is sent to monitoring processes from the data-logging system. The trigger type can be used by a monitor task to select a specific class of events, such as minimum-bias triggers. These processes look at per channel occupancies, energy flow, trigger rates, the trigger decision, etc. to assess whether the detector is functioning properly. Comparisons may be made to a set of template information. If discrepancies are found, then an alarm is signaled to the operator.

Also some events are used to provide course calibration constants for several systems such as drift velocities for the central tracking chamber. As these constants are determined, they are fed back for use by the Level-3 trigger algorithms. One such monitor task will reconstruct vertices using the SVX and relay information on the beam position to the accelerator control system.

The VMS based programs used for Run Ib will be converted to unix as well as be modified for the upgraded trigger and detector components. Those tasks requiring the largest rates will run on the same (multiprocessor) system as the data logger. Other monitors receive events via the local area network (FDDI, ethernet).

The general structure of the monitors follows that of offline analysis programs. This allows them to run either online, receiving data from Level 3, or offline, receiving data from disk or tape, without any changes to the code. Other offline standards regarding programming languages, histogramming packages, database access etc. will also be followed by these programs.

Most monitor programs will consist of an analysis portion that runs in a background mode storing histograms and other information in shared memory. Operators will then use a display program based on PAW, Histoscope or a successor to view this information. Summary information from each run may be archived. Programs that analyze data for calibration runs will be structured in a similar manner.

11.7.3 Hardware Configuration

A block diagram of the proposed hardware configuration is shown in Fig. 11.13. A series of unix workstations will be used for Run Control and Event Monitor functions. Workstations may be used interchangeably for either. Additional workstations will be used for program development. X-window terminals will be used for auxiliary displays as well as for program development. A central file server provides storage for code, databases, monitor archive, and program development facilities for remote users. A local area network consisting of ethernet and FDDI interconnects the file server, workstations, and processors in the Data-Acquisition system. There will be a high speed connection to the offline computing systems at the Feynmann Computer Center. Routers to isolate portions of the network from external traffic and switches to improve performance will be used as needed.

The computing system will be configured for reliability so as to be able to operate in case of the failure of single components. Workstations and displays are interchangeable so the failure of one does not affect operation. Disk files critical to operation will be stored in RAID arrays. These will be attachable to multiple CPUs so the files should always be accessible. Operation will be possible without a general network link outside of the experimental area. Internal networking will be configured to minimize the single points of failure.

All code will be as platform independent as practical. However there will be some limitations. Workstations used for Event Monitor tasks must be supported by the offline system. Third-party support particularly in the area of VxWorks host software also restricts the range of allowable platforms. Currently supported platforms for Run-Control related functions include Silicon Graphics (IRIX) and Sun (Solaris). It is expected that these choices will evolve. The progress of Windows-NT will be followed and this platform may be supported if deemed appropriate.

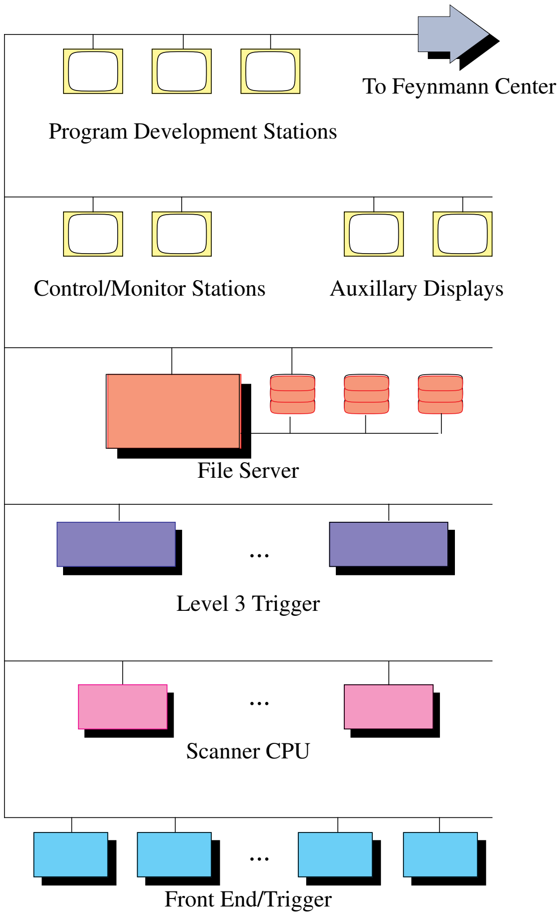

Figure 11.13: A schematic of the Online Computing system. In the actual implementation, the local-area network will be segmented using routers and switches that are not shown.

11.7.4 CDF ALARMS

CDF contains about 20 different detector systems. The Alarms, Limits and Remote Monitoring System (ALARMS), provides the interface between the on-shift personnel and control and monitoring of the detector systems. CDF ALARMS is a fully integrated node on the accelerator-wide network, ACNET, and allows the monitoring of the accelerator status at the beginning and during beam stores as well. ALARMS data is not part of the normal event data stream. There are three primary functions of the CDF ALARMS system: high-voltage control, online monitoring, and data logging. A brief comment on each is given below:

- CDF ALARMS is the primary mode of High Voltage Control and online status display for detector systems. The High Voltage Control function can be used to set individual detector channels, segments of a certain detector system, or an entire detector system.

- CDF ALARMS provides online monitoring and display of the HV status (on, off, or standby), the status of low-voltage power, temperatures, and the detector systems.

- Data-logging and time plots for luminosity and various detector functions are also plotted on the CDF ALARMS consoles.

The CDF control room contains three ACNET Consoles (Fig 11.14) and two dedicated monitors for displaying detector status. The Consoles are presently VAX workstations running the VAX/VMS operating system. Any X-station can run the ALARMS program and display the various plots pertaining to detector status. The heart of the monitoring system are two “i386” Front-End processors that continuously monitor the detector from the first floor of the B0 assembly building. The ACNET consoles communicate with the Front-End processors via ethernet, and the Front-End processors communicate with the Central Database and Operations VAX via ethernet as well. They also communicate with a series of IBM-compatible PCs that control high voltage for all wire-chamber systems.

CDF displays three types of Alarms: normal, ignored, and severe. Severe alarms require immediate action and are triggered if crucial channels are out of tolerance. Data taking is stopped and cannot resume until the problem is fixed; severe alarms are announced by DECTALK as they appear. Normal alarms indicate a channel that may have drifted out of tolerance, but data taking can continue. However if a large number of normal alarms appear, then a severe alarm is set, and data taking is inhibited. Ignored alarms allow the shift personnel to be reminded that a number of problems do exist but that these can be fixed later. Normal alarms can be ignored by clicking on the alarm. Severe alarms cannot be ignored.

Information may be exchanged between the ALARMS and data-acquisition control systems in several ways. At begin run, the ALARMS system accesses the daq-system configuration database to

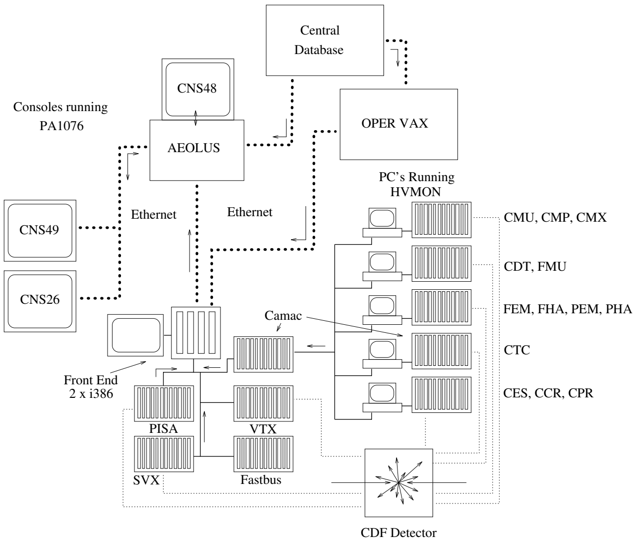

Figure 11.14: A schematic of the CDF ALARMS system as configured for Run Ib showing the consoles, Front End, and control of the detector systems.

record high-voltage settings and other detector information. Alarms encountered during the run are similarly recorded. In a few cases some ALARMS hardware is duplicated in the daq system allowing information such as the solenoid magnetic field to appear in the data stream for each event. Should a high-voltage trip condition be encountered, a signal is provided to the trigger system to inhibit the generation of Level-1 accepts. And finally, a software interface is available from the Computing Division to allow a DAQ process to exchange information with ACNET. This will be used, for example, to send to ACNET the beam position as measured by SVX tracks reconstructed online.

The CDF ALARMS system has been running successfully now for several years. No major upgrades are planned for ALARMS. Substantial work will be required to install the new detector systems into ALARMS for Run-II. The collaboration will continue to use the system through Run II.

Bibliography

[1] P. Chang, V. Papadimitriou, "Specifications for Calorimeter Front-end Electronics", CDF note CDF/DOC/CALORIMETRY/CDFR/3646.

[2] R.J.Yarema et. al., *A Fast, Wide Range Charge Integrator and Encoder ASIC for Photomultiplier Tubes*, IEEE Trans. on Nuclear Science, **Vol 4** (August 1993), pp. 750-752.

R.J.Yarema et. al., *A High Speed, Wide Dynamic Range Digitizer Circuit for Photomultiplier Tubes*, Presented at the 6'th Pisa Meeting on Advanced Detectors, (22-28 May, 1994) La Biodola, Isola d'Elba, Italy.

[3] V. Papadimitriou, "Calorimeter TDCs for Run II and Beyond", CDF note 2209.

[4] Many operational details are given in the description of the earlier SVX2 chip: "A Beginner's Guide to the SVX2," Ray Yarema, *et. al.,* FERMILAB-TM-1892, 1994.

[5] The current SVX3 specifications are contained in "Device Specifications for the SVX3 Chip Set," Raimund Strohmer and Walter Knopf, http://www-ese.fnal.gov/eseproj/svx/svx3/svx3spec.html.

[6] S. R. Amendolia et al., "IFT Upgrade Project: Readout System Overview"

[7] J. Patrick *et. al.*, "The CDF Ultranet Based Data Acquisition System", Proc. Conf. Comp. High Energy Phys. (1994).

[8] VME International Standards Organization, Scottsdale, AZ

[9] VME-P Committee, VITA 1.X-199X.

[10] Wind River Systems, Alameda, CA.

[11] C. Rotolo *et. al.* Proc. IEEE Nucl. Sci. Symp. (1991).

[12] PCI standard v2.1, PCI Special Interest Group and PCI Mezzanine Card (PCM) standard, IEEE 1386.

[13] ATM Standard, Internation Telegraph and Telephone Consultive Committee.

[14] ANSI Standard X3T9.3/91-071, Fibre Channel Association, Austin TX.

[15] HIPPI, ANSI Standard X3T9.3/91-005 and Serial HIPPI, Rev 2.2, Serial HIPPI Implementors Group.

[16] Fore Systems Inc., Pittsburgh PA.

[17] Proc. First Int. Data Acquisition Conf., Fermilab (1994).

[18] Systran Inc., Dayton OH.

[19] FC-AL, ANSI X3T11, Fibre Channel Association, Austin TX.

[20] International Business Machines, Kingston, NY.

[21] Apple Computer, Cupertino, CA.

[22] Quantum Corp., Milpitas CA.

[23] StorageTek, Boulder CO.

[24] J. Ousterhout, "Tcl and the Tk Toolkit", Addison Wesley.

[25] M. McLennan, "The New itcl", Proc. 1995 Tcl/Tk Workshop.

[26] Sun Microsystems Inc., Sunnyvale, CA.

[27] G. Oleyik *et. al.*, "Fermilab DART Run Control", Proc. IEEE Real Time Conference (1995).

[28] Real-time Innovations, Sunnyvale, CA.