15. Data Acquisition System

15.1 Introduction

The primary task for the readout and data acquisition system (DAQ) is to concentrate the data from the large number of APD channels into a single stream that can be analyzed and archived. The DAQ also provides for an intermediate buffering location where the data can be held until it is determined that the data should be recorded or rejected. Online trigger processors will be used to analyze the data stream to correlate data with similar time stamps and to look for clusters of hits indicating an interesting event. Additional functionality for dealing with flow control, monitoring, system operations and alarms is also included.

15.2 Technical Design Criteria

The NOνA front end electronics (described in Chapter 14) operates in un-triggered mode with data continuously being digitized, time-stamped, pedestal subtracted and zero-suppressed. There is no spill trigger required at the front-end. The data must be buffered for up to 20 seconds awaiting arrival of a spill trigger message. While the data is in the cache it can be searched for data satisfying a software trigger condition, and analyzed for calibration and monitoring purposes. A spill signal is required to arrive within the buffering time so that the spill time can be correlated with the time-stamped data to determine if the hits occurred in or out of spill. There is no additional selection of in-spill data. All hits that occur in a 30 μs window centered on the 11 μs spill are recorded for further processing. The data selected in this manner corresponds to approximately 190 GB per year.

Randomly selected data for calibration and monitoring will be collected off-spill at a rate that is approximately 100 times higher than the in-spill rate. This corresponds to about 18 TB/yr of raw data. The overall data rate is driven by cosmic ray muons that occur at a rate of approximately 200 kHz with a modest overburden. Monte Carlo simulations predict approximately 200 hits per cosmic ray muon. This corresponds to a total hit rate of 40 MHz and a total data rate of about 0.5 GB/s. Table 15.1 summarizes the channel count and rates in the NOνA Far Detector and in the Near Detector where the rate is much lower at the underground location.

	Far Detector	Near Detector
Front end boxes	14,136	497
Front end boxes per plane	12	2.5
APD channels per box	32	32
Total channels	452,352	15,904
Average Noise hit rate per channel	< 30 Hz	< 30 Hz
Total Noise hit rate	14 MHz	500 kHz
Bytes per hit (channel ID, TDC, ADC, status)	10	10
Muon rate	200 kHz	5 Hz
Average Hit channels/muon	200 (multiple hits/plane)	50
Muon Hit Rate	40 MHz	250 Hz
Total hit rate	54 MHz	500 kHz
Average hits rate per channel	120 Hz	50 per hour
Digitizing rate	2 MHz	8 MHz
Total data rate	0.5 GB/s	5 MB/s
Average occupancy	1.2×10^{-4}	3×10^{-5}

Table 15.1 Channel count and rates in the NOνA Detectors.

15.3 System Architecture

NOνA has only one type of detector and one type of readout, making the NOνA DAQ conceptually simple compared to typical high energy physics experiments. The near detector is slightly different, in that the electronics will be required to accommodate the high burst rate of hits during a spill. However, the detector modules will be significantly smaller, and the detector will be located underground, so that the average data rate is much lower for the near detector. This means that the DAQ components can be the same and still accommodate the high hit rate during a spill.

The front-end electronics are described in detail in Chapter 14. In summary, the front-end boards digitize signals from the APD arrays. The digitized data are then input into a Field Programmable Gate Array (FPGA) that applies zero suppression, timestamps and buffers the data before serialization and transmission to the DAQ. The FPGA also provides control and monitoring of the front end electronics and APDs and the external interface to the DAQ using a custom protocol over inexpensive CAT5 cabling.

The DAQ system is composed of four main components, a timing and command distribution system, an array of Data Concentrator Modules (DCM), a Gbit Ethernet network, and a buffer farm. Timing synchronization and command packets for the data concentrator modules come from the timing and command system consisting of one Master Timer Unit and ten Timing Distribution Units that act as a separate network backbone for these signals. Digitized data from up to 64 front end boards are routed through CAT5 cables to a custom DCM. The DCMs send data to the buffer farm, and eventual storage over the Gbit Ethernet network. The buffer farm and DCMs are under the control of a Run Control Computer.

15.4 Timing and Command Distribution System

The Global Timing and Command Distribution system provides a timing infrastructure and distributes commands from the run control computer to the DAQ and the detector. Timing information and actions are carried by special timing packets on high speed serial links. Timing packets have deterministic timing and the system is self-compensating for propagation delays. This means that any action initiated by timing commands to all DCMs are coherent across the detector with an accuracy of better than one 163.84 MHz clock cycle. The Master Timer Unit (MTU) contains a Common View Global Positioning System (GPS) trained clock oscillator with a time-stamp counter and generates timing command packets at a fixed repetition rate which are distributed in a serial loop as shown in Fig. 15.1. The bidirectional high speed links facilitate compensation for the propagation time of the system as described below. All non-timing controls command packets to the system, usually reads & writes, are queued and interlaced between the timing packets at a lower priority.

The ten Timing Distribution Units (TDU) make up the timing backbone of the detector with the Master Timer Box at the near end as shown in Fig. 15.1. The self compensation of the TDU backbone is based on the round trip time-of-flight (TOF) of a timing command through each TDU to the far end of the cable loop and back. The reference point for all of the timing is the far endpoint of the backbone loop as shown in Fig. 15.2. The measurement is made in each TDU independently and the TOF count is related to the TDU position along the backbone. The TOF count starts on the first arrival of the timing packet and ends on the arrival of the return packet as in Fig. 15.3. This time-of-flight measurement is divided by two to give the specific TOF for each TDU across the entire backbone. A synchronization trigger output is generated by each TDU after receiving a new timing command by counting down the TOF/2 measured from the previous timing command. All TDUs perform the same timing measurement which allows each TDU to delay the trigger signal to be synchronized to the far end of the backbone. This calibrates the propagation delay out of the TDU system. The cables between TDUs can be any length without affecting the timing synchronization. The synchronized trigger signal generated at all TDUs, causes the timing command to be issued on all of the DCM links along the backbone simultaneously. The MTU can also synchronize to the far end reference point in the same way and compensate the GPS clock time for the cable propagation.

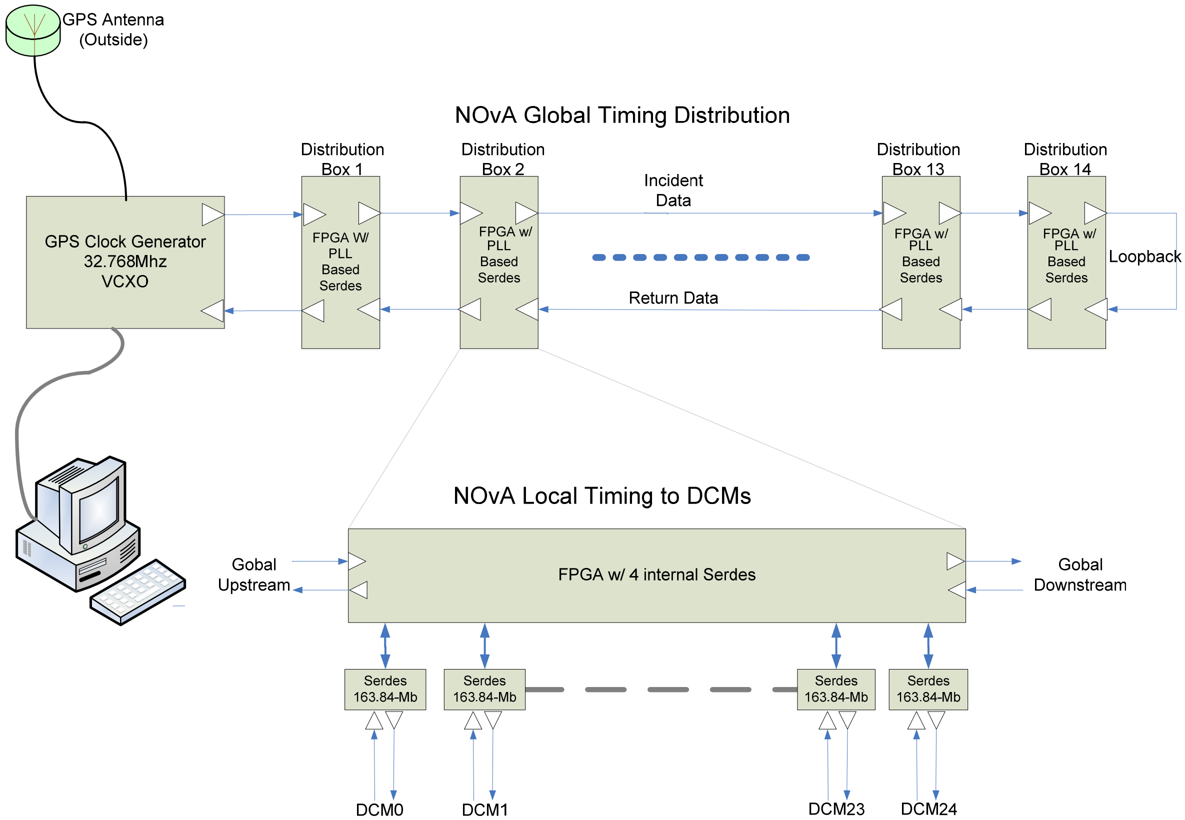

Fig. 15.1: Diagram of Global and Local Timing Distribution System

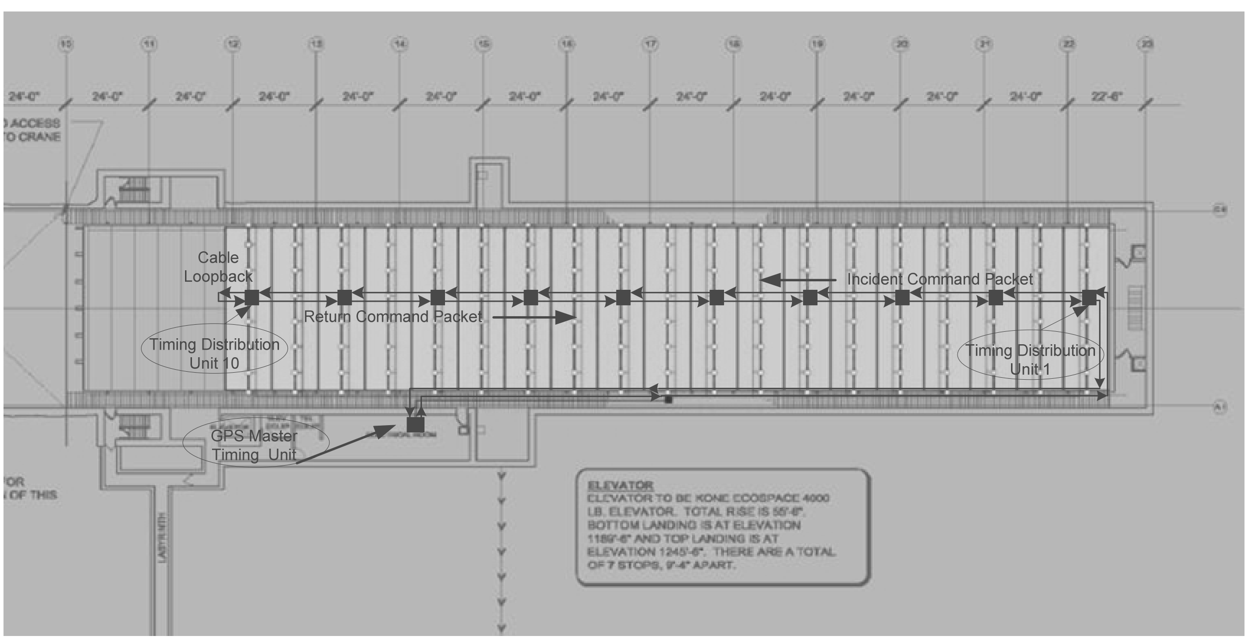

Fig. 15.2: NOνA detector cabling for timing and command distribution system

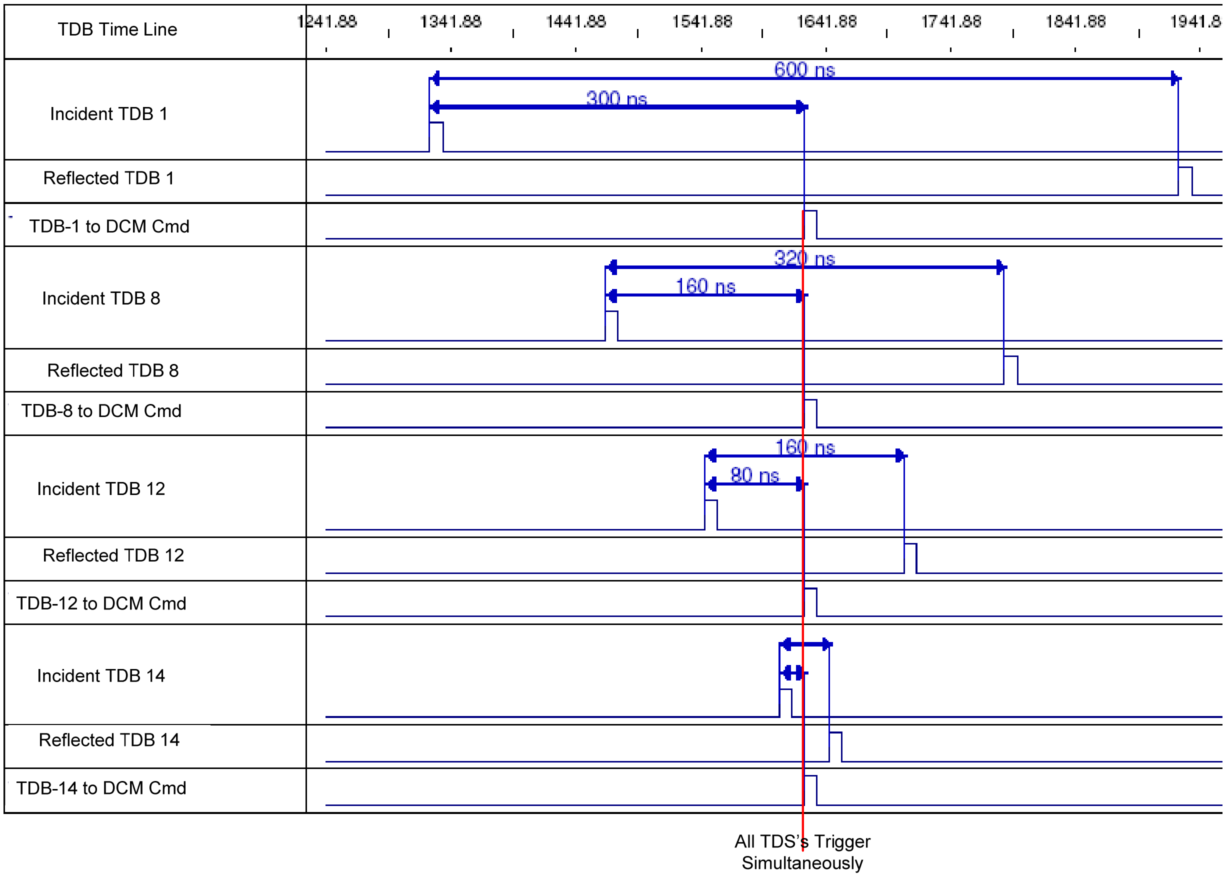

Fig. 15.3: Command Time-of-Flight and Synchronized Trigger at TDBs

Compensation of the propagation delay on each of the TDU to DCM links is performed similarly to the TOF/2 measurement of the detector backbone except there are multiple loops so the longest TOF/2 needs to be determined and distributed such that it is used in all of the timing calculations. Rather than determining the longest TOF/2 and broadcasting it, an alternative method is to predetermine a delay parameter that is slightly greater than the longest actual measurement of all of the DCM links. The actual measured TDU to DCM round trip TOF is subtracted from this delay parameter in the timing calculation for each DCM port. Long cables will have short added delays and short cables will have long added delays such that all the DCMs receive timing signals at the same time. The TDU still compensates for cable lengths differences but the reference point is artificially longer than the longest link cable. An advantage to this method is that communicating the TOF parameters between TDBs about the TDU to DCM links is unnecessary and the single predetermined delay is simply hard coded in the firmware.

15.5 Data Concentrator Module (DCM)

The NOνA Data Concentrator block diagram is shown in **Error! Reference source not found.**. The main functions of the Data Concentrator boards are to collect serial data from up to 64 Front-End Boards (FEBs), combine and packetize this data for transmission over Gigabit Ethernet to Buffer Nodes, and finally to fan out timing and control information from the timing system to the FEBs.

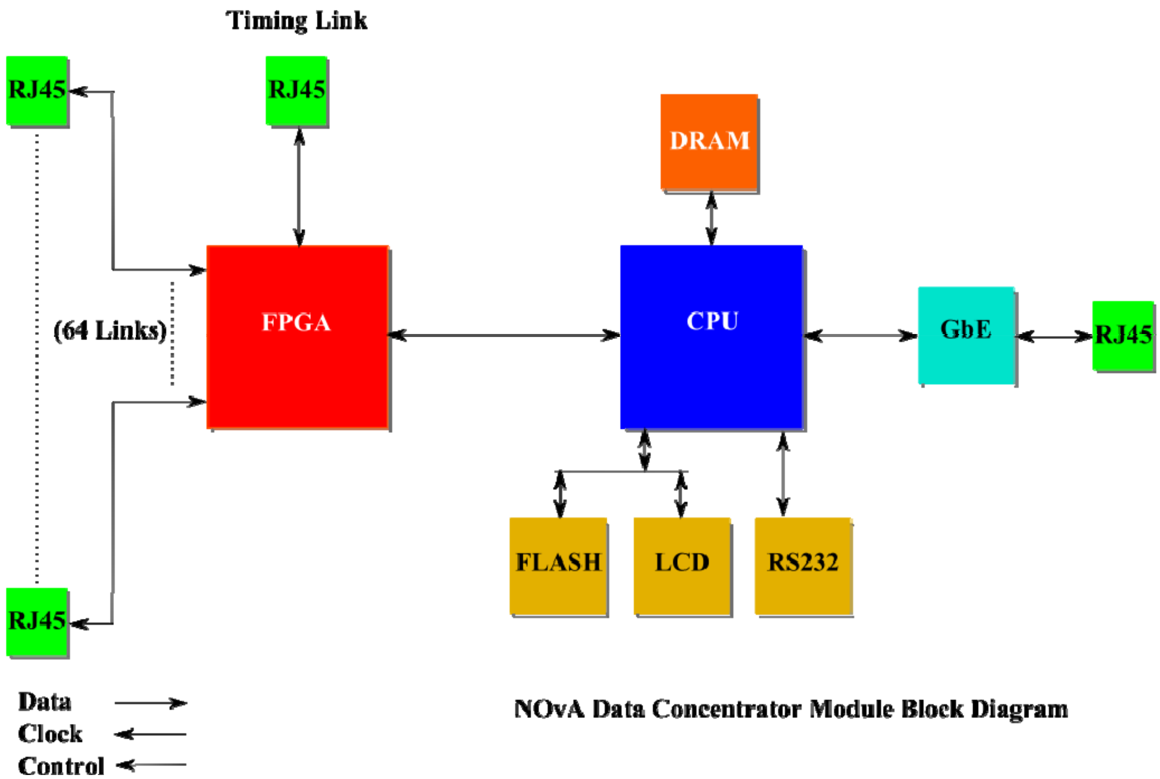

Fig. 15.4: Block Diagram of Data Concentrator

The DCM mainly consists of a midsized FPGA, an embedded Power PC processor running embedded Linux, and numerous connectors, as shown in the prototype design in Figure 15.5. On power-up, the processor runs a bootloader from onboard flash memory. The bootloader initializes the Gigabit Ethernet port and connects to a boot server to download its embedded Linux image. Once Linux is booted, a connection to the server is once again established so that the FPGA firmware, DAQ application software, and DCM configuration data can be downloaded. The processor then configures the FPGA and starts the DAQ application code. Among the DCM configuration data is the buffer node destination table. This table contains a list of available destination buffer nodes and the timeslices each node is assigned to. Other configuration information will include configuration data for FEBs, as well as any other configuration data necessary for operation.

DCMs connect to FEBs with CAT5 cables through RJ-45 jacks. Of the four pars of wires in an FEB to DCM cable, one is used for serial data transmission from the FEB to the DCM, one is used for system clock transmission from the DCM to the FEB, one is used for serial Command transmission from the DCM to the FEB, and the fourth is used for transmission of a sync pulse from the DCM to the FEB. Data from the FEB consists of a header and hit information for a given timeslice. The FPGA on the DCM combines the hit information for all 64 FEBs in each timeslice into a data block. The data is then read from the FPGA by the embedded Power PC processor. Application code running on the processor packetizes the data into Ethernet packets and transmits these packets to a buffer node through its Gigabit Ethernet port.

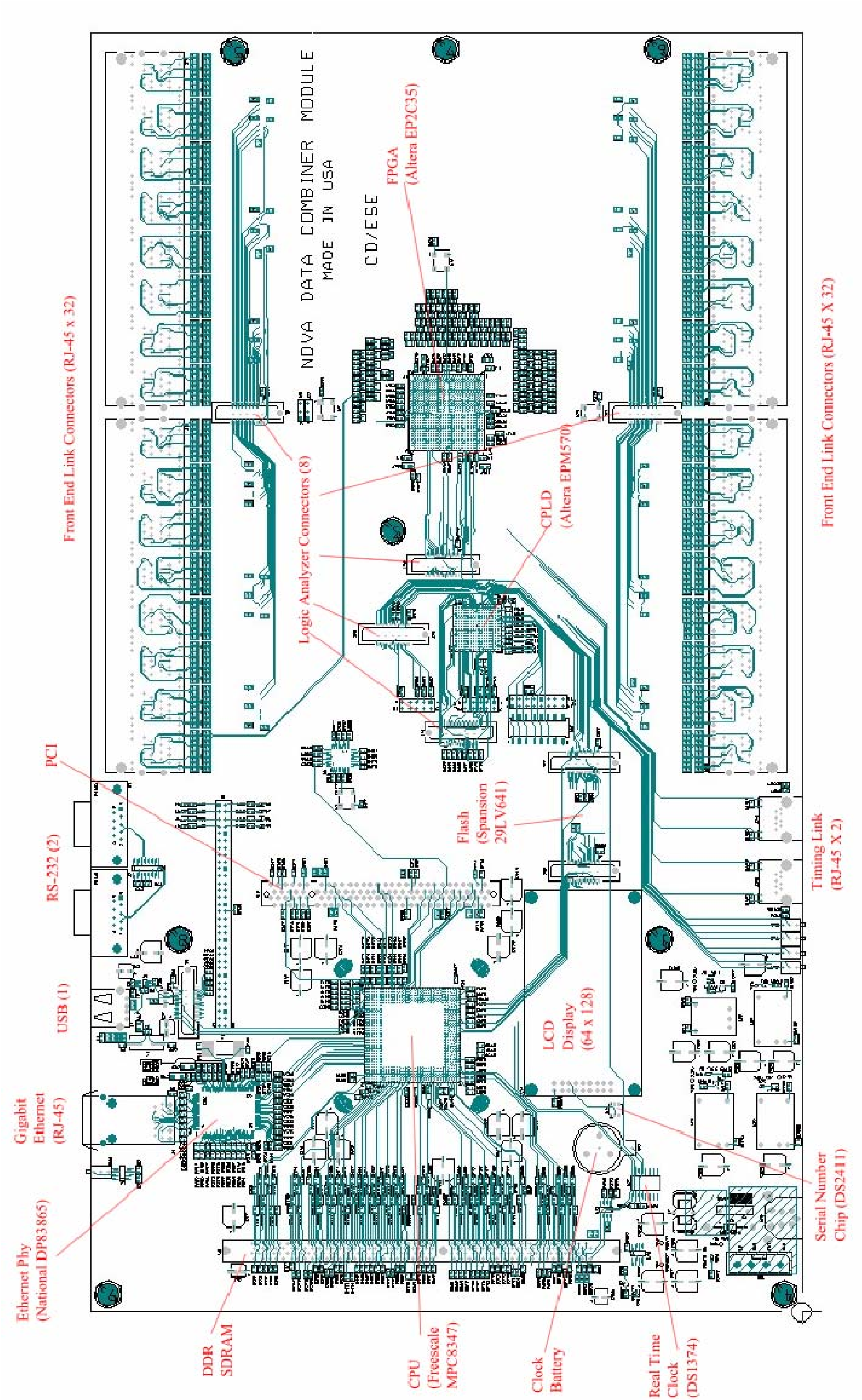

Fig. 15.5: Layout of Prototype Data Concentrator Module

DCMs are connected to the buffer nodes through a Gigabit Ethernet switch array. To make the most effective use of network bandwidth through the switch array, DCMs are each programmed to transmit data to different buffer nodes in a round robin rotation, with each DCM starting with a different buffer node. Timing information from the timing system is used to divide time into slices such that data packets are transmitted to a different buffer node in each slice. In this way, DCMs stay synchronized in their rotation, and no two DCMs ever transmit to the same buffer node at once. Buffer nodes are receiving data from only one DCM at a time, and all buffer nodes are receiving data on all time slices. Because there are more DCMs than buffer nodes, each DCM will have null time slices where it will be idle part of the time. This ensures that collisions on the network are minimized. The buffer nodes collect data from all DCMs for different events. Event data files are stored locally until they can be copied to Fermilab for permanent storage.

The DCM also connects to the timing system with a CAT5 cable. Clock, Command, and Sync information are encoded in a serial bitstream on one of the pairs of wires in the timing system connection and transmitted to the DCM. A second pair of wires on the cable is used to transmit clock information back to the timing system to allow a round-trip time measurement to occur. Once the timing system measures the round-trip delay time to each DCM, it can insert delays as necessary to synchronize all DCMs. Clock, Command, and Sync are decoded by the DCM and transmitted to each FEB.

15.6 Gigabit Ethernet links and Data Buffer Processor Farm

The buffer farm nodes serve as a large data caching system for the detector with search capabilities. The buffer size must be large enough to hold 20 seconds of data at the design rate for the system (twice the nominal rate). The data is buffered in these nodes until a trigger is received from the global trigger system. This trigger is a time window (GPS based), and the buffer nodes use the time window to search for all data within that frame of time and route it to a logging process. All data for one time frame will be buffered in the same buffer node. The buffer node will need to aggregate the data from each of the DCMs for the time frame and insure data from all DCMs is seen. This is a form of event building where data for a large time frame is built as a single event. For any frame seen by a buffer node, data must be received from each DCM. This provides a system integrity check and also allows more performant searching when a trigger is received.

While all the data from a given time frame will be contained in one farm node, an actual 30μs trigger, which defines an event, may span these large time frame boundaries and thus be split across two farm nodes. Each farm node will need to determine if it has any data for the trigger window in its memory buffer and send it to the logging process. It might also be sent to an event monitor server which could cache events for event displays. Any problems sending would raise an error. In addition to buffering the data, the data may need to be made available to a processor framework while waiting for a trigger.

The buffer farm nodes will be connected to the Data Combiner Modules through a set of interconnected Gigabit Ethernet switches. The switches are interconnected to allow a connection from every DCM to every buffer farm node over a Gigabit interface. The physical layout of the cabling on the detector is shown in Fig. 15.6, while the connection to the buffer farm and control room is shown in Fig. 15.7.

To determine the number of switches, it is assumed that full bandwidth through the switch is needed, and we want complete flexibility for data routing, in that any DCM can talk to any buffer node. That means each switch module needs as many inputs as outputs on each switch module on the side of the switch with the fewest modules, and a three stage switch is needed. The buffer node side of the switch has 152 buffers and the DCM side has 228 DCMs, so the buffer side is the side that determines the maximum data rate through the switch. For full bandwidth through a 48 port switch module 24 ports are used as inputs and 24 are used as outputs. Since 152 / 24 ~ 7,

Fig. 15.6: Top view (left) and Side view (right) of DAQ Cable routing on the NOvA detector.

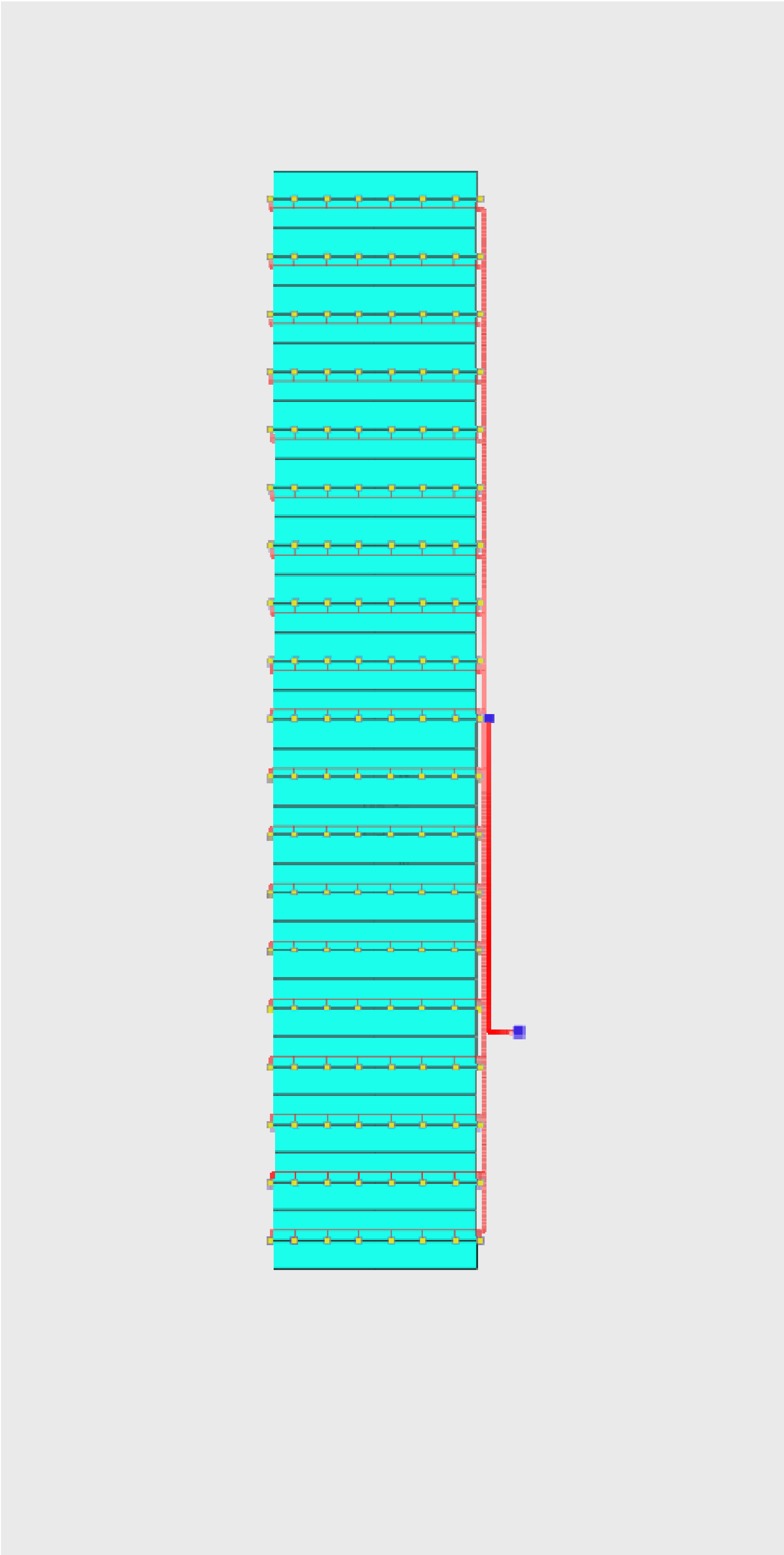

Fig. 15.7: Cable routing to buffer farm in the NOνA Detector Hall at Ash River.

there will be eight output switch modules with some empty ports. An even number of switch modules is desirable to ensure even bandwidth distribution across the switch modules. The intermediate switch stage connects directly to the output stage, so there needs to be eight switch modules here also. The input switch stage needs enough ports to connect 228 DCMs to 152 intermediate switch ports. A total of 380 ports are needed to connect the 228 DCM inputs and 152 buffer farm outputs. The minimum number of 48 port switches that could to this is 10. Connecting these 10 switches to 8 intermediate switches does not map very well with regard to the number of links per intermediate switch connecting to input switches. Some would have three and some would have two connections. This could cause problems with the time-slicing algorithm, which prevents collisions and blocking. Using 12 input switches works much better. There will be 21 input connections from the DCMs and 16 output connections, two between each input switch and each intermediate switch. That gives a switch module count for the DAQ of 28. More switches are needed for the control network. All of the buffer nodes need to connect to run control through the control network. Four 48 port switches would have enough ports to connect to the buffer nodes, but the switches need to connect to each other and to the run control machines, so this will take about five more switches, bringing the total to 33.

15.7 DAQ Application Tasks

There are several tasks that the DAQ Application is required to perform, control of the DCM elements, management of the buffer farm, trigger distribution, data logging, system monitoring. These functions and their relationships are shown schematically in
Fig. 15.8, and covered in some detail in the subsequent sections.

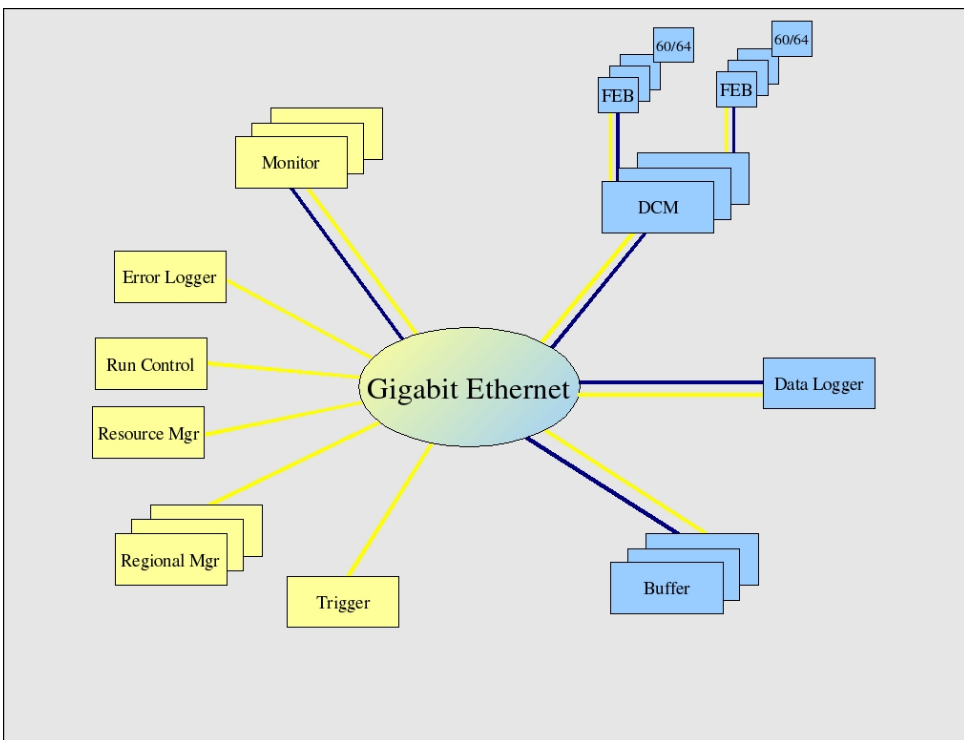

Fig. 15.8: Block diagram of DAQ system showing interconnects through Gigabit Ethernet, and custom protocol from the DCMs to the FEBs. Detector data flow is found on the blue links and control and error messages are passed on the yellow links.

15.7.1 *Run Control System*

The run control system is responsible for configuring and controlling the data acquisition system. For debugging and commissioning purposes of the large detector of essentially one type of detector sub-system, the run control will support partitioning of the resources. There will need to be one central resource manager that tracks assignments of DCMs and farm nodes to partitions. The resource manager will provide the run control client the necessary information for using only the hardware and applications reserved for its partition.

Run control will interact with the configuration database to provide appropriate information to processes on how to configure themselves, download or receive calibration parameters, etc. It controls when runs start, stop, pause, resume, etc. In addition, it is responsible for saving all relevant information on the configuration of the run, including partition information, to a run history database for offline access. Any changes in the configuration would require a new run to be started. Runs could extend over the course of several days.

15.7.2 *Resource Manager*

The resource manager is responsible for managing the configuration and resources of the system. It serves as the master process for allocating hardware to partitions and providing the necessary configuration information for configuration of the partitions. The resource manager provides an interface to the configuration database for run control. Run control will request a partition and configuration data for the partition. The resource manager will use the configuration database to fulfill the request if possible and mark the resources as in use for the given partition. The resource manager and run control instances can communicate directly for command and reply messages. For error messages the resource manager contacts the error logger directly.

In order to reduce the number of connections between applications and increase the rate capacity for control and error messages, the system will provide a set of regional managers for both the control and error messages. There will be three classes of regional managers: DCM regional managers; buffer farm regional managers; and generic regional managers (for the rest of the applications).

The data combiner module regional manager is the main interface for control and error messaging to the processes on the DCMs. The regional manager is responsible for the processes in its region and reporting back to run control and the error logger on their behalf. To support partitioning of the data acquisition system, there should be a separate DCM regional manager for each partition. This level of segregation will reduce the impact of problems in one partition from affecting other partitions.

The buffer farm regional manager is the main interface for control and error messaging to the processes on the buffer farm nodes. The regional manager is responsible for the processes in its region and reporting back to run control and the error logger on their behalf. To support partitioning of the data acquisition system, there should be a separate buffer farm regional manager for each partition. This level of segregation will reduce the impact of problems in one partition from affecting other partitions.

The generic regional manager is the main interface for control and error messages to the rest of the data acquisition system for several applications. These applications are: the monitor server; the event monitor; the DAQ monitor; the logger; and the global trigger. The generic regional monitor provides a segregated region for applications that are not part of the primary error messaging or control system. This reduces the number of direct connections to run control and to the error logger and allows for a more distributed system.

15.7.3 *Global Trigger System*

The global trigger system is responsible for receiving a beam spill signal, or other triggering condition. The beam spill signal will be generated at Fermilab in response to the signal firing the kicker magnet. The actual time will be logged and transmitted to the far detector global trigger via the internet. Upon receipt of the beam spill time the global trigger system will then generate the time windows based on the signal and send that signal to the buffering processes on each of the buffer farm nodes. Since the data acquisition system will support partitioning, the global trigger will need to confine the trigger signal to the appropriate partition or partitions. The global trigger manager interfaces to the data acquisition system through the generic regional manager, however the trigger signal may bypass this and go directly to the buffer regional manager to expedite delivery if necessary. In the absence of actual spill triggers the system will generate random triggers to allow for tracking of calibration and monitoring of the detector.

The reliability of the internet link for transmission of the trigger signal will be evaluated initially during installation and continuously thereafter. The MINOS detector has been operating in this way and has found that with a sufficient buffer time on the order of 10 seconds, the reliability is 99%. Since the data rate is so large for this large surface detector it is unrealistic to buffer the data to disk for later searching. If there is a reliability problem an additional type of link could be implemented using other technologies.

15.7.4 Error Logger

The error logger will be responsible for routing error messages to various places and logging them for later analysis. The error logger should act as a gate keeper to the run control clients with respect to error messages. It is the responsibility of this part of the system to determine which messages need to be seen by run control.

15.7.5 Monitoring System

Interactive monitoring of the system will be accomplished through a set of monitoring display applications which connect to a monitor server. The DAQ monitor system provides a window into the health of the DAQ system and could be used to spot trouble before it reaches the point of raising alarms and sending error messages. This is especially true for monitoring performance related aspects of the system where behavior might show signs of degradation well before a component fails. The event monitor allows for a statistical analysis and display of data flowing through the detector. This allows one to see effects that one would not be able to spot on an event by event basis. The event display allows data to be displayed graphically. Some subset of the data will be routed to event displays when they are running to allow monitoring of the system on a trigger by trigger basis. This provides complementary information to statistical monitoring of data flowing through the system.

15.7.6 Data Loggers

Triggered data is sent from the buffer farm nodes to the logger system which is responsible writing data to local files and ultimately insuring it is safely stored in a more permanent mass storage system. The data arrives at the loggers in a mix of complete and partial events. It will need to determine whether the event is complete or not, and if not build the complete event. Failure to build an event which was incomplete should be an error. Once a complete event is in hand, the logger needs to determine the appropriate file for writing the event and write the event to that file. The logger should determine when to close a file (size, timeout) and initiate a request to move the file to the local disk cache area. The logger needs to handle multiple streams. At a minimum there is spill data and calibration data and the two need to be written to separate files. The local disk cache area is where the data sits until it is transferred to permanent storage. This serves as a buffer when the permanent storage system is unavailable. It can also serve as a buffer area to allow grouping of transfer requests so that permanent storage transfers can be done efficiently. The far site must be able to buffer data for up to one week to allow for possible extended outages of the network link to Fermilab.

The system will also consist of a set of databases for configuration, monitoring and keeping track of run conditions for later analysis. The data acquisition system will support the concept of partitioning. The configuration database will store information on what hardware is available for a run, which partition is currently assigned that hardware, routing tables for DCMs to buffer farm nodes, and other parameters used by the various processes in the data acquisition system. This database is used by the resource manager to manage available resources and fulfill resource requests from the run control processes for each current partition. Configuration information will need to be tagged and referenced in the run history database. The calibration database is a repository for APD channel settings for the detector. The database will need the ability to group settings into production sets that are downloaded to the detector. Because the calibration settings are expected to change over the lifetime of detector operations, the database and infrastructure must support multiple calibration sets and allow for easy association of the sets with the runs in which they were used. Thus some form of tagging scheme will be needed to mark which set individual data elements belong, and that tag needs to be stored in the run history database on a run by run basis.

The run history database stores run dependent information that documents the condition of the detector, quality of data as determined at the time data was taken, and indications of type and amount of data. The information must be sufficient to reconstruct the state of the system during any run period for offline analysis purposes. From the perspective of the data acquisition system the run history database is mainly a repository for writing.

The monitor database is a transient database for storing monitor data on the state and performance of the data acquisition system. The aim is to allow access to information on how the system was performing for debugging and diagnostic purposes. Several months of data is desirable as that allows for an examination of trends and the ability to look for hints of when performance or reliability might have started to degrade. The ability to compare how the system looked when it was behaving well to when there is a problem is a great asset when investigating a problem.

15.8 Supernova Detection

While not part of the scope of the NOνA project, it is possible to detect the neutrinos from supernova events in the NOνA far detector. A supernova explosion at a distance of 10 kpc will result in about 9000 neutrino interactions in the NOνA far detector. Electrons at these energies will deposit about 15 MeV of energy per cell. The majority of neutrinos from such a supernova explosion will be between 20 to 40 MeV and will result in coincident hits in adjacent cells with energy deposits greater than 0.5 MIPs.

Supernova detection requires front end electronics with low deadtime and a free running DAQ system that does not rely on a beam trigger signal from Fermilab. The NOνA front end electronics and DAQ system already meet these requirements. The search for supernova would have to be done in real time requiring the DAQ system to be able to handle the full 0.6 GB/s rate described in Table 14.1. The modular, parallel design of the NOνA DAQ makes this possible as well. Supernova detection would require up to 20 seconds of data buffering, compatible with what is necessary for the neutrino oscillation program. Additional online processing and software would also be required. While the cost of these additions is relatively small, they are not included in the NOνA base program and would have to be funded separately.

15.9 Detector Controls and Monitoring

The large scale nature and staged installation schedule of NOνA, mandates a detector control system and detector monitoring systems capable of controlling and monitoring the front end detector readouts and the environmental and infrastructure systems simultaneously in both production physics data taking and installation and commissioning modes. The detector controls and detector monitoring have been designed to share a common interface and communications infrastructure to minimize development times and reduce runtime support needs. The systems are designed to be multi-fault tolerant due to the necessity of having periods of prolonged unattended operation of the detectors.

15.9.1 *Detector Controls*

The detector controls systems (DCS) are modeled off of a client/server architecture with physical hardware access hidden by a protocol independent hardware abstraction layer (HAL). The main control systems communicate via the HAL's protocol specifications with an Input/Output Control daemon (IOC) which interfaces with the physical hardware and contains the device specific data acquisition mechanisms for accessing the specialized hardware. The client applications included in the DCS suite provide the production and commissioning human interface facilities, as well as the run time status indicators, and logging facilities.

The NOvA controls and monitoring needs are broken down by modular subsystem and scaled by the total number of units required for the detector size. In the base configuration, the DCS systems include the power and bias voltages for the front end detector electronics, the data concentrator module's run time configurations and status reporting, the water cooling and pump system's operational data, general environmental monitoring and operations data, and the status and run time parameter sets for the front end APD boards. The high channel count of FEBs compared to all other monitoring and control data, allows us to divide the system and consider the FEB parameters separately in terms of system load and redundancy. The subsystem monitoring requirements are listed in

Table 15.2. In all, the NOvA far detector DCS system is estimated to require periodic monitoring and reporting of approximate 420k operational parameters.

System	Report Values	Channels	Total Monitoring
Low Voltage	6	81	486
High Voltage	2	162	324
DCMs	~20	324	6480
Water Cooling	~16	144	2304
Env. And Opps.	128	8-12	~1536
FEBs (via DCMs)	10-20	20,088	400k
TOTAL			11,130 410k w/FEBs

Table 15.2: Detector controls and monitoring estimated channel count

Infrastructure

The DCS software model is based upon the infrastructure provide by the EPICS software libraries developed by Argonne National Lab and used extensively in other modern experimental setups. The EPICS packages run on a modern Linux based PC to provide a robust communications protocol for interfacing client and server applications using the abstracted idea of universal channel identification and structured data records. Internal representations of the detector state data are maintained by the EPICS infrastructure's databases and handling routines. Data reporting is handled through different push/pull modes of the EPICS calls which allow for individual and block channel reporting. Reporting can be initiated either upon client request, at regular intervals via server push, or upon state changes as would be desired for reporting anomies or alarm conditions.

The costs associated with the free EPICS infrastructure libraries are restricted to the software engineer resources that are needed to develop new IOC hardware drivers for the unique systems present in the NOvA detector, as is the case for the FEB and DCM status modules. Device drivers for most of the commercial hardware employed by NOvA already have had EPICS drivers developed for them by other D.O.E. funded experiments and laboratories.

The base design of the EPICS readout server system hardware requirements is predicated on load projections extrapolated from similar EPICS control system implementations currently in use for the DØ experiment's central fiber tracker and from benchmark tests done at Argonne National Lab. The base design distributes the required hardware access load of the core subsystems over three physical linux PC style computers. The core subsystems handled by these machines include the monitoring and alarm requirements of the environmental systems, voltage controls, water cooling, data concentrator modules, and additional miscellaneous systems. Processing of the operations data for the front end boards is distributed over an additional 9 computers in order to keep the expected processing loads balanced and provide single fault tolerance via fail-over redundancy on a mirror server.

Each of the DCS servers are designed around the current generation of dual 3GHz, dual core AMD processor PCs running a modern Linux operating system. The servers are required to have a minimum of 4 gigabytes of system memory to allow for full in core memory manipulation of the EPICs channel databases as well as the full memory footprint of each of the IOC processes. The servers are equipped with two gigabit over copper Ethernet controllers to permit them to isolate traffic across the DAQNet and DCSNet lines. The dual dual-core processor configuration allows for effectively 4 simultaneous processor stacks, permitting more efficient data manipulation of the large EPICS channel databases concurrently with performing read/write operations to the hardware interfaces. The configuration also allows for the promotion of two of the processors to the real time scheduling queue for correct handling of the hardware access without affecting system responsiveness to outside interrupts. The servers are housed in a 1U form factor, 19 inch rack mountable case which include the appropriate cooling and power supplies. Local disk on each server is limited to storing the operating system and EPICS control systems and local EPICS channel databases.

The DCS infrastructure relies on access to four types of high speed networks to acquire and distribute the control and monitoring data. The requirements for the networks are summarized in

Table 15.3. The DAQNet is shared with the normal data acquisition system and is primarily used for access to the DCMs and associated FEB controls. Bandwidth usage on the DAQNet is confined to the DCM to DCS IOC server communications. Client queries, database and logging traffic are partitioned off by placing the monitoring client to EPICS channel server traffic on a dedicated gigabit network. This partitioning of the system as shown in

Fig. 15.9, eliminates the possibility of user level queries and periodic monitoring spikes from adversely impacting the available bandwidth on the DAQNet and creating deadtime in the acquisition systems.

The physical extent and topology of the DCSNet switches and interconnects at the far site are centered on NOvA control room utilizing a high reliability 48 port gigabit switch as the master router for the network. Lines to the detector hall for the DCS subsystems to handle monitoring and control of the voltage, cooling and environmental systems are branched off from the master switch and run through fiber channels down to second level switches in the detector hall as shown in

Fig. 15.10: DCS voltage controls, network infrastructure Figure 15.9. The master router also connects in to the 802.11g wireless network that services the detector hall and operations facility. The wireless access provides the ability for mobile laptops to access the detector controls from the detector assembly area and act as local control panels for many of the front end devices that lack local control panels. Access to the DCS network will be controlled through a master control server that will act as the gateway and firewall between the outside world and the internal control and monitoring systems. Additional access to the network from Fermilab will be

negotiated using VPN style tunneling to provide greater access to key systems that need to be mirrored in the control center in Wilson hall.

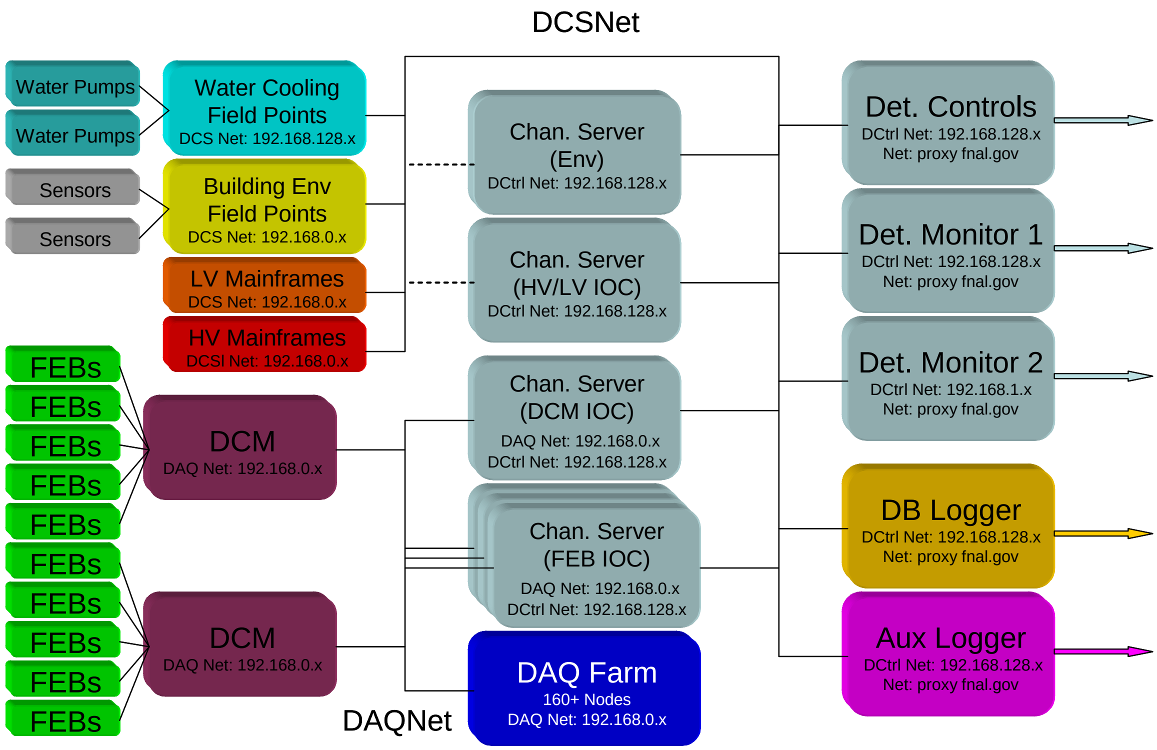

Fig. 15.9: DAQNet and DCSNet topology for controls and monitoring

Network	Bandwidth	Location	Topology	DCS Nodes	Purpose
DAQNet	GHz	Detector	gigabit Ethernet	10	DAQ, FEB, DCM Comm.
DCSNet	GHz	Control Room	gigabit Ethernet	15	EPICS/Logging Client to IOC Server Comm.
DCSWireless	54Mbit	Detector Hall	802.11g	2	Detector Local configuration, testing, commissioning via EPICS clients
FNAL Net	Mbit	Control Room	TCP/IP Ethernet	1	Offsite access, Monitoring from Wilson hall, etc…

Table 15.3: Detector controls and monitoring network infrastructure

System Readout

The detector controls system is required to access hardware that is scattered throughout the detector hall. In order to economize on readout infrastructure and cabling, a combination of the existing gigabit DAQNet Ethernet wiring is used in conjunction with a moderate number of branches off of the DCSNet gigabit lines to provide wired coverage throughout the detector hall to all devices that support conventional 10/100/1000 BaseT Ethernet style connections. Devices that do not support control over TCP/IP Ethernet, have been chosen to allow for control over the RS485 serial protocol. The RS485 wiring is designed to allow for 32 device chains which can span the length of the detector and tap off at each of the cooling system pump stations. Devices not local to either the RS485 or gigabit Ethernet lines will be connected either by dedicated lines or bridged in through the wireless system.

Front End Boards

Communications with the front end boards housing the APD units is designed to occur via intermediate communication with the DCMs. Each DCM is designed to receive data packets from the FEBs with header bytes that allow for flagging of the packet as either data, calibration, status or alarm. The packets from the front end boards are filtered on their header type and then are buffered locally in the DCM to await a polling cycle. The packets can be broadcast immediately to an IOC when an alarm condition necessitates immediate acknowledgement of a state change. The data buffers on the DCM that are dedicated to the FEB status and monitoring data are treated for the purpose of the IOC server as the access point for acquiring the channel data. Control of specific registers and operational parameters for the front end boards are similar passed to the front ends via the DCM, and transmitted over the local serial bus between the two objects. Processing of the command directives is handled through pushing the command set onto the DCM run time environments and allowing for an acknowledgement packet to be passed by up from the FEB to verify the transaction.

Data Concentrator Modules

The data concentrator modules each communicate with the DCS over the standard DAQ gigabit Ethernet. Status and monitoring packets are constructed in the DCM and stored in a dedicated buffer area which can then be polled by the IOC on the main DCS servers. This model reduces the overhead associated with the DCM needing to store the EPICS IOC channel database records, and limits the network bandwidth that is used by the controls systems on the DAQnet.

Each DCM runs a small lightweight status client which performs the access to the local registers and queries the operational parameters of interest for the DCS system. The data from the DCM daemon writes the information into a packed buffer in shared memory which can then be transmitted and interpreted by the EPICS IOC. Control parameters for the DCM are similarly passed through the shared memory segment and notification to the local DCM clients is given through a standard semaphore call.

The goal in designing the DCM control interface is to minimize the workload and memory footprint that is needed on front end system as well as the network bandwidth that is required between the front end and control room systems.

High Voltage Mainframes

The CAEN Sys1527 high voltage mainframes that are used to bias the frontend APD units are controlled through an Ethernet (TCP/IP) interface option board that is available from CAEN. The interface board provides access to a telnet session on the local mainframe and from there to a SCADA control layer. Driver support for the Sys1527 exists for the EPICS software infrastructure, making development of the experiment specific IOC servers easier than for comparable high voltage systems which do not support modern communications protocols.

The CAEN mainframes for the far detector will be located on the DCSNet gigabit lines. The proximity of the high voltage mainframe and patch panel racks to the DCS main switch can be optimized to satisfy either building space or total cabling costs. Each mainframe requires a single 10 megabit connection to the main switch resulting in the need to run only two branches off the DCSNet to connect the detector left side and detector right side mainframes.

Low Voltage Supplies

The Wiener low voltage supplies that provide the 3.3V and 24V power to the DCMs and FEBs support control over Ethernet (TCP/IP). The Wiener units will be connected to the DCS network in the same manner as the CAEN high voltage supplies. Due to the large number of low voltage pods that are required to support the full NOvA detector, the Wiener voltage supplies will be distributed in clusters, each in close proximity to a 24 port gigabit Ethernet switch, which are in turn connected to the DCSNet main switch to the control room. This minimizes the total cabling requirements for full control of the low voltage systems and allows for geographic distribution of the low voltage clusters across the detector geometry as shown in Fig. 15.10.

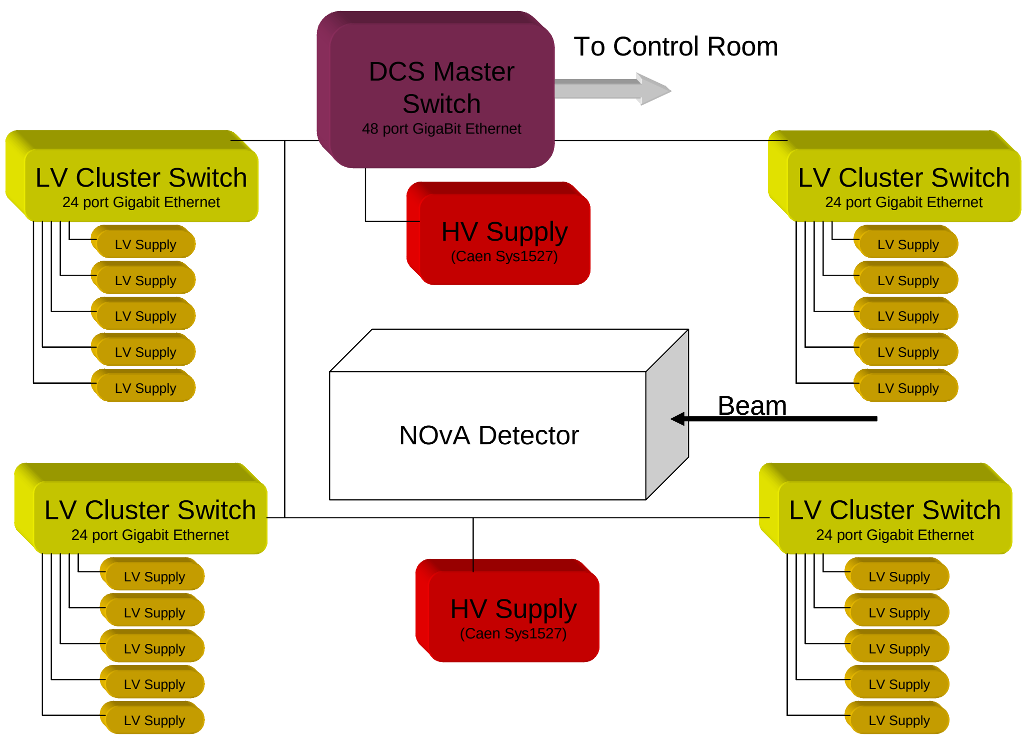

Fig. 15.10: DCS voltage controls, network infrastructure

Environmental Systems

Environmental systems are monitored and controlled by a series of National Instruments compact field point stations which provide monitoring for analog currents and voltages, digital inputs, as well as pods for driving RTD sensors. Discrete output disables for system control are also provided by the field point modules. Each field point can be equipped to handle a maximum of 128 discrete channels using a full population of 8 pods, each with 16 analog or digital channels. The RTD driver pods contain a maximum of 8 channels, which reduce the total monitoring capacity of each fully populated backplane requiring temperature sensors of this type.

The field point stations are interfaced to the DCS through standard Ethernet and will be addressable on the DCSNet branch lines that run down the dectector. The field point stations are 19 inch rack-mountable and will be mounted in the same rack or in close proximity to a rack with one of the secondary network switches. The field points are connected to these switches which then allow a full path back to the mater control stations.

The physical size of the experimental detector mandates that environmental monitoring and control stations be distributed through out the far site with coverage adequate to cover all portions of the sensitive detector. This results in a series of staggered monitoring stations, each covering a zoned area of the detector. The environmental systems zones are divided into detector left and right with 8 zones evenly spaced along the 293 foot long detector at 36 foot intervals. In this configuration the longest cable run distance between a sensor point a the edge of the monitoring zone and the monitoring station is 55ft assuming the stations housed in racks on the central side catwalks.

Water Cooling Systems

The water cooling systems uses the same 8 module compact field point units used for the environmental stations outfitted with 32 analog inputs, 16 digital counter channels and driving 16 RTDs. The interface to the field points is the standard TCP/IP Ethernet allowing for them to be wired into the DCSNet through the secondary switches. The higher channel density per station allows each unit to monitor 4 pump stations.

15.9.2 Detector Monitoring

Control Rooms

The NOνA experiment will require both a far site and FNAL control facility to handle the monitoring and control needs of the detector during the initial construction phase as well as in the general operations phase. Additional provisions must be made for remote monitoring and control by experimenters that are not stationed at either FNAL or the far site full time. The facilities must allow them to access their controls and calibration systems with the intent of allowing experts to remotely diagnose and recover from anomalous system behavior. See Figure 15.11.

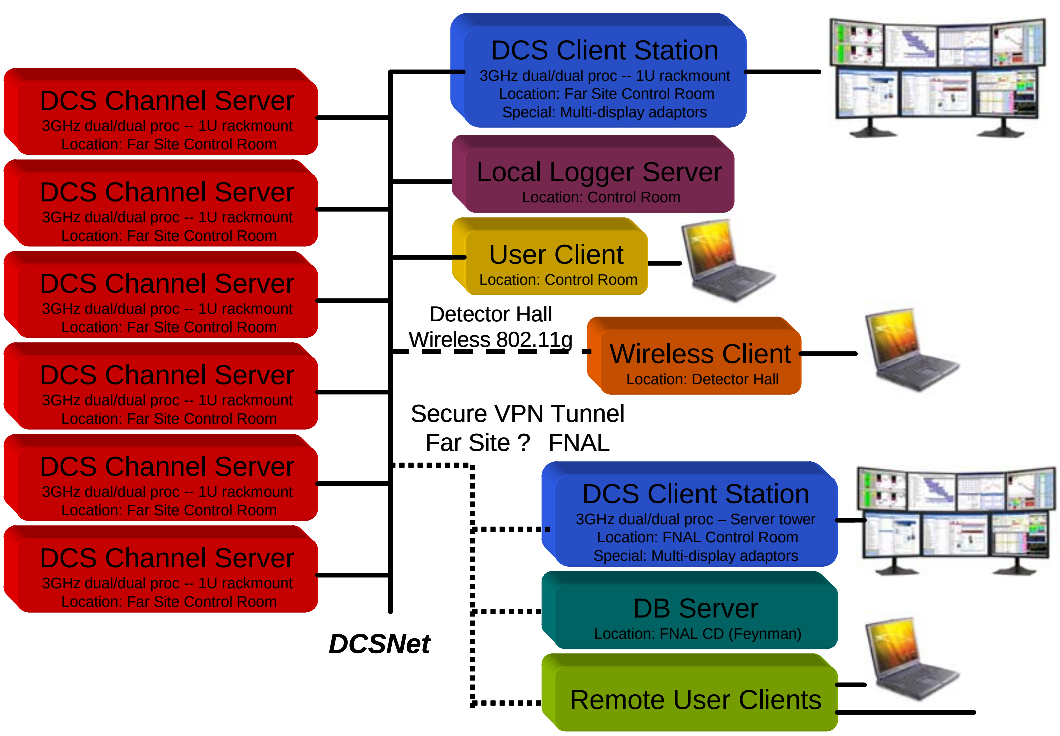

Fig. 15.11: DCSNet overview of attached clients and servers

The far site control room and the Wilson hall control room are designed to provide identical functionality and access to all NOvA detector systems. This level of redundancy is provided by placing two identical DCS Client Workstations on the DCSNet network with the remote station connected via a VPN style tunnel. Each station runs an identical set of monitoring and controls clients which broadcast to and communicate with the EPICs based DCS channel servers on the DCSNet subnet. Monitoring and alarm traffic can be broadcast and displayed in both the far site and FNAL control rooms simultaneously without conflict or contention for network bandwidth. Control systems are similarly run on both systems but with a synchronization daemon which negotiates control locks and mandatory update queries between the two systems. This ensures that one and only one of the control stations can have write level hardware access to a system and that all changes to the operational parameters of that systems are forced to be updated on the sister station before further set point operations are permitted.

Fig. 15.12 shows the schematic breakdown of one such lock-to-modify transaction between two sister DCS Client stations.

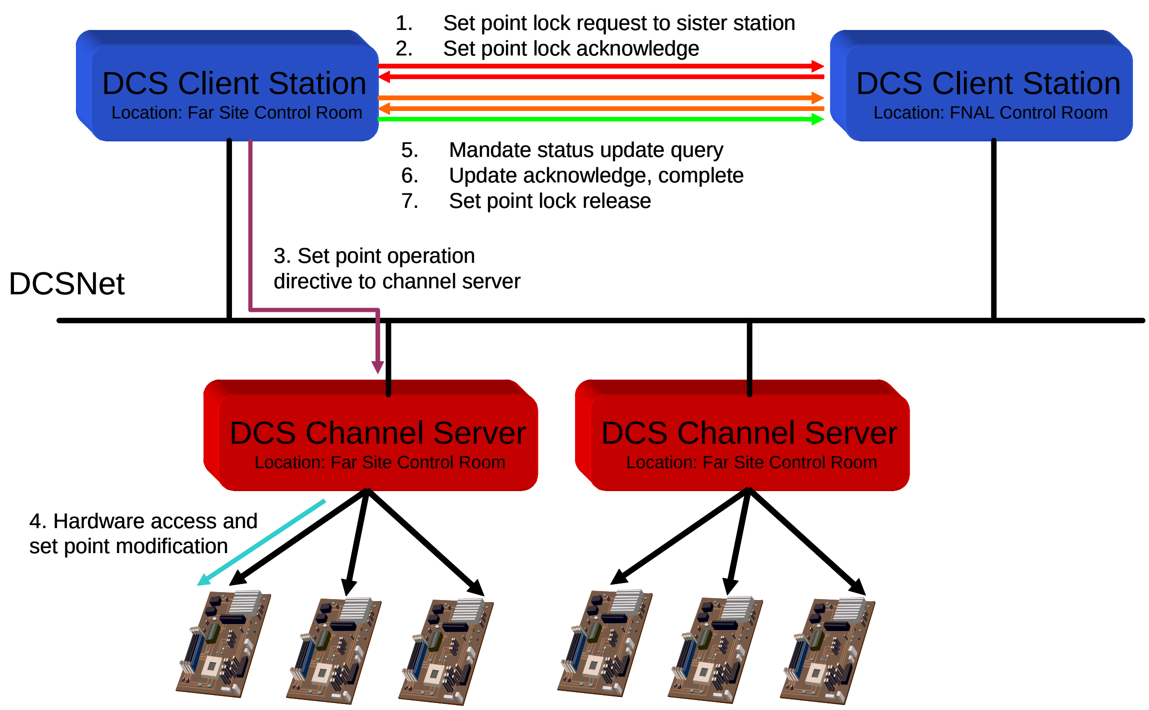

Fig. 15.12: DCS client locking scheme

Off site access to the controls systems are handled in a similar manner. Each client that wishes to modify operational parameters of the detector subsystems is registered into the "current clients" list on the controls master. The client list is subsequently accessed for control operations and locks are established across peer machines in the same manner as the lock-to-modify scheme for the main control rooms.

Access to hardware controls and monitoring from the detector floor is provided in the same manner as for offsite clients with the exception that access is made through the DCSWireless network that is installed in the far site detector building. The DCSWireless infrastructure is based on a series of general 802.11g wireless access points spread throughout the far site detector hall, control rooms and assembly areas. The 802.11g specification operating at 2.4GHz, limits effective transmission and reception range to a maximum of 30 meters of unobstructed, line of sight pathway between client and access point. The control room and assembly area facility's projected reception rings are shown in

Fig. 15.13. These radii assume the installation of two access points located in the building ceiling above the electrical room/shop and the primary control room. In the experimental hall, the detector is not transparent to the carrier frequencies that standard 802.11 wireless operate at, as a result the reception between clients and the access points is effectively limited to true line of sight. Coverage of the north and south catwalk areas will need to be provided by access points placed 50m apart on the middle catwalk areas. The placement of the access points is shown in

Fig. 15.14 indicating their reduced coverage blocks due to attenuation in the detector material. Access points in the experimental hall ceiling area will provide full coverage to the top areas of the detector. The three access points shown in

Fig. 15.14 are spaced at 50m intervals, offset from the catwalk access points by 10m Access points for the detector hall have their reception radius obstructed by the detector, as a result access points will need to be placed in the north and south catwalk areas at 50 meter intervals, and above the detector in the rafter area at 50 meter intervals, staggered by 10 meters from the side catwalk access points with the third access point extending out into the assembly bay, providing full unobstructed coverage to that location.

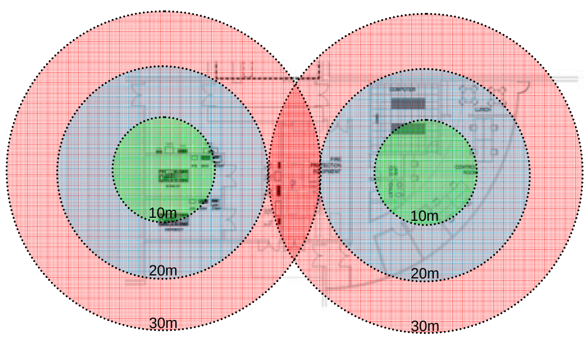

Fig. 15.13: Wireless network coverage radii (meters) for operations center Need new figure to match latest building design.

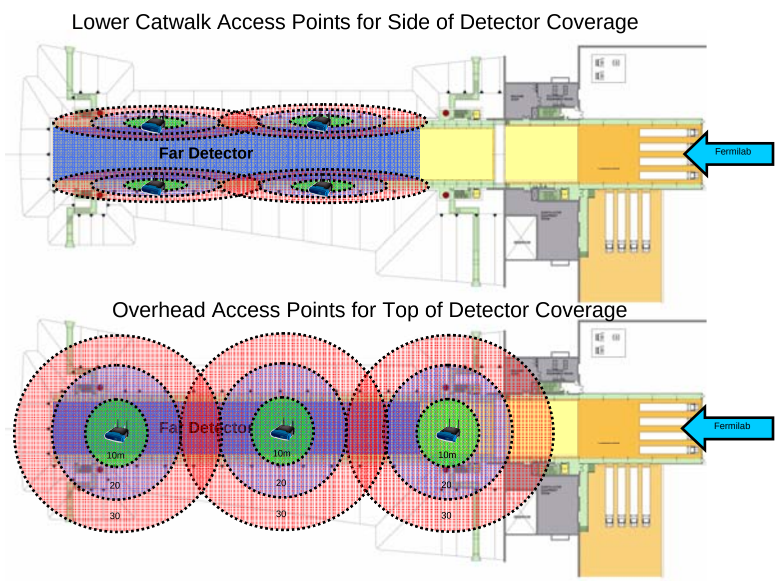

Fig. 15.14: Wireless network coverage for the Far Detector Hall.

DCS client hardware is expect to conform to a standard setup of Linux based PCs and laptops, configured to run the required set of monitoring and control software. The primary DCS control room station is intended to be a 6U rack-mountable multi-processor Linux based PC equipped with graphics adaptors capable of driving a large format multi-display console system. The primary controls station is intended to efficiently display critical monitoring for all subsystems of the detectors in such a manner that a single operator can effectively manage the run time needs of the detector system. Display panels are allocated for voltage controls, DCM controls and operations, FEB status and operations, environmental and rack monitoring, water cooling systems. Data quality displays and histogramming are not included in the DCS operations displays. The DCS station for the Wilson hall control room is a similar multi-processor linux based PC housed in a standard server case, capable of driving a similar multi-display console setup.

At the far site, a set of standardized laptop stations will be combined with the full wireless coverage of the detector hall, to grant experimenters local access to the DAQ and front end infrastructure systems for the purposes of installing, calibrating and trouble shooting hardware. These mobile stations are intended to replace the local controls and display systems that would otherwise be required during the installation and operations phases of the experiment.

The expected breakdown of computing needs for the DCS system are shown in Table 15.4.

System	Units	Type	Memory	Location	Notes
DCS Chan Server (Voltage, DCM, Water, Env.)	3	+3GHz 4-core PC	4 Gb	Far Site	1U rack mount, RS485 interface card
DCS Chan Server (FEB)	8	+3GHz 4-core PC	4 Gb	Far Site	1U rack mount
DCS Chan Server (Hot spare)	1	+3GHz 4-core PC	4 Gb	Far Site	1U rack mount, RS485 interface card
DCS Client Master	1	+3GHz 4-core PC	8 Gb	Far Site Ctrl. Room	6U rack mount, Multi display adapters
DCS Client Master	1	+3GHz 4-core PC	8 Gb	Wilson Hall	Server tower, Multi display adapters
DCS Local Logger	1	+3GHz 4-core PC	4 Gb	Far Site	6U rack mount, RAID-5 storage array (2-3 Tb)
DCS Mobil Station	4	1-2 GHz low power	1 Gb	Detector Hall	Wireless equipped laptops for detector maint.

Table 15.4: DCS computing hardware requirements

15.9.3 Local DCS Logging

The DCS and monitoring system will require a data logging facility to keep a short term record of detector status and operational parameters for use in system performance and diagnose trouble or degradation in subsystems. The logging system will need to be able to handle the full DCS monitoring load which is estimated to be over 420,000 parameters. Based upon the DCS load, the logging system is designed to handle 512,000 parameters, thereby allowing each FEB to monitor a maximum of 25 operational values. The logging record for each parameter is expected to include the time stamp corresponding to the polling time at which the information was extracted from the front end system, an identification flag indicating the system, channel and parameter being reported, the value of the operational parameter and a status word indicating the assigned state value of the channel (e.g. on/off, nominal, high/low, critical, tripped, etc…). The record for local logging will be packed into a 64 byte structure appropriate for output to fixed storage. A single logging dump of this type is expected to consume a maximum of 262 million bits. Burst transmission of the logging data across the Gigabit Ethernet DCSNet network could be accomplished in under 100ms when provisions for network latency and 80% network utilization are included, as such the available network bandwidth is not impacted by nor does place any limits on the logging systems. The total data volume recorded during normal operations relies on the frequency at which the system is logged. The data volumes for different monitoring cycles are tabulated in

Table 16.5 along with their impact on long term storage.

Channels	**Logging Interval**	**Storage/Hour**	**Storage/Month**	**Storage/Year**
512,000	1 min	1.97 Gb	1.46 Tb	17.55 Tb
512,000	5 min	393 Mb	293 Gb	3.51 Tb
512,000	10 min	196 Mb	146 Gb	1.75 Tb
All ops (20k) + half FEBs	1min/5min staggered	273 Mb	203 Gb	2.44 Tb

Table 16.5: DCS expected local logging rates

The local DCS logging station is designed to comply with the storage needs of housing one month of the full channel count 1min polling interval data or the requirements of the a full year of either the 10min cycle or 1min/5min readout schemes. The data can be stored locally in a convenient run indexed/linked Ntuple/RootTuple format or other packed format to allow for quick parsing of the relevant data. Logging data can be expired or written to permanent storage at regular intervals. In the 1 minute polling scheme this expiration time would be on the order of 31 days with nightly rollovers.

The local logging facilities are not intended to act as a full database for physics analysis, and as such do not include provisions for database overhead or long term data access. The facilities do however include provisions for a temporary storage bank to house critical information that may need to be buffered if the connection to the experiment database is lost for an extended period. In this scenario the local logger would assume the duties of the fall back database until such time as connections were restored and the information could be synchronized with the master database at FNAL. This redundancy should be able to handle 2-3 days of operations given the current disk capacity specifications.

15.10 Racks and Infrastructure

15.10.1 DCS Detector Hall Racks

The DCS monitoring systems require limited rack space in the detector hall to mount the distributed I/O modules, sensors and network routers. Each DCS field point station will require 6U of rack space to mount the 8 module back plane. Miscellaneous environmental sensors, terminal blocks and instrumentation that feed into the field point units are expected to require another 6U rack space when mounted on an appropriate panel. The 24 port network switches are rack mountable with a 1U form factor. Wireless access points are wall mountable and will not require any dedicated rack space. Since there is a 1-to-1 correspondence between the environmental monitoring units and the 24 port network switches, we expect that each station set will take up 13U of space which can be placed in the same area as the DCM and power distribution nodes. In all the 16 DCS stations will comprise 208U of space, the equivalent of approximately 5 standard 42U racks.

Race space for the water cooling systems is similarly expected to take 6U of rack room. If the high channel counter systems are used this will be the equivalent of 3.5 standard 42U racks.

15.10.2 DCS Control/Electronics Room Racks

The DCS systems will require additional space in the electronics/control rooms to house the computers used to run the EPICS servers. Each DCS channel server is contained in a 1U rack mount enclosure, with the logger and main controls machines housed in 6U enclosures. Additional 3U of room is required for the 48 port main switch, 1U for the fiber interconnect switch, 4U for the network VPN gateway/router, and 4 U is required for the Keyboard/Video/Mouse (KVM) switch that allows local console access to each of the channel servers.

In total the computing systems for the DCS are expected to take up 36U of rack space which will need to be housed in a cabinet that allows for force air cooling. This rack should be placed in the electronics room with cabling run to the control room facility to drive the output displays and provide user keyboard and mouse control of the master systems.