SLAC-PUB-12976

THE ILC GLOBAL CONTROL SYSTEM*

J. Carwardine, N. Arnold, F. Lenkszus, C. Saunders (Argonne), K. Rehlich, S. Simrock (DESY),
B. Banerjee, B. Chase, E. Gottschalk, P. Joireman, P. Kasley, S. Lackey, P. McBride, V. Pavlicek,
J. Patrick, M. Votava, S. Wolbers (Fermilab), K. Furukawa, S. Michizono (KEK),
R.S. Larsen, R. Downing (SLAC)

Abstract

The scale and performance parameters of the ILC require new thinking in regards to control system design. This design work has begun quite early in comparison to most accelerator projects, with the goal of uniquely high overall accelerator availability. Among the design challenges are high control system availability, precision timing and rf phase reference distribution, standardizing of interfaces, operability, and maintainability. We present the current state of the design and take a prospective look at ongoing research and development projects.

INTRODUCTION

The International Linear Collider (ILC) is a 200- to 500-GeV center-of-mass high-luminosity linear electron-positron collider, based on 1.3-GHz superconducting radio frequency accelerating cavities [1]. The machine operates at a pulse repetition rate of 5-Hz, with each 1-ms beam pulse comprising ~3000 bunches.

The control system overall design is evolving as details of the accelerator technical design are developed. We present a snapshot of the design as it is described in the ILC Reference Design Report (RDR) [2].

The control system reference design in the ILC RDR serves several purposes: to establish a functional and physical model for costing purposes; to establish a starting point for further engineering design and development efforts; and to communicate our vision of the control system to the technical groups. The functional and physical models reflect current requirements to the extent they are known or can be inferred.

The physical model of the control system is intended to be technologically agnostic. Where the reference design describes technologies such as Ethernet and ATCA electronics platforms, they are included as representative technologies for architectural and costing purposes, and are not intended to reflect actual design decisions.

ILC ACCELERATOR OVERVIEW

The layout of the ILC accelerator, as described in the RDR comprises the following major accelerator systems:

- Polarized electron source based on a photocathode DC gun, and undulator-based positron source.
- 5-GeV electron and positron damping rings, with 6.7-km circumference, housed in a common tunnel at the center of the facility.
- Beam transport from damping rings to main linacs, followed by bunch compression systems.
- Two 11-km-long 250-GeV linacs, each comprising ~8,000 SCRF accelerating cavities fed from ~320 rf klystrons.
- A 4.5-km beam delivery system, bringing the two beams into collision at a single interaction point.

CONTROL SYSTEM REQUIREMENTS

General requirements for the ILC Global Control System are largely similar to those of integrated control systems for other large-scale accelerators, and include control and monitoring of accelerator technical systems, online diagnostics, data archiving, and machine configuration save/restore. Additionally, several features of the ILC accelerator will push the control system implementation beyond the basis of experience for accelerator control systems. These are summarized below.

Scalability

Initial estimates suggest the control system could be required to interface with 100,000 network-attached devices and several million control points, with the overall accelerator complex spanning some 31 km.

Availability

Accelerator availability goals drive much of the control system design. As will be described later, the control system itself must be highly available, and it should provide functionality to help minimize overall accelerator downtime.

Machine Automation and Feedback

The ILC accelerators comprise large numbers of complex technical systems, including some 640 rf klystrons driving some 16,000 SCRF cavities. A high degree of automation will be required in order to manage routine operations tasks such as machine startup, tuning, cavity conditioning, and accelerator optimization.

ILC operations will rely extensively on beam-based feedback to meet performance goals. Feedback systems will stabilize the electron and positron trajectories throughout the machine, correct for emittance variations, and provide measurement and correction of dispersion in the main linac. Two timescales of beam-based feedback are anticipated, namely pulse-to-pulse feedback at the 5-Hz pulse repetition rate, and intratrain feedback within the macropulse containing nominally 3000 bunches spaced at 330-ns intervals. Dedicated systems will be needed for the

intratrain feedback, while the 5-Hz feedback loops will be implemented through the global control system.

Precision timing and synchronization

Precision timing & synchronization is required by many of the accelerator technical systems, including injection & extraction fast kickers and the ~640 RF systems in the main linac. The reference design for the timing and RF phase reference distribution system is described elsewhere [3].

CONTROL SYSTEM FUNCTIONAL MODEL

We have defined a three-tiered control system functional model [4], shown in Fig. 1, which introduces an intervening 'Services' tier between the Client and Front-end tiers of older two-tier control system models.
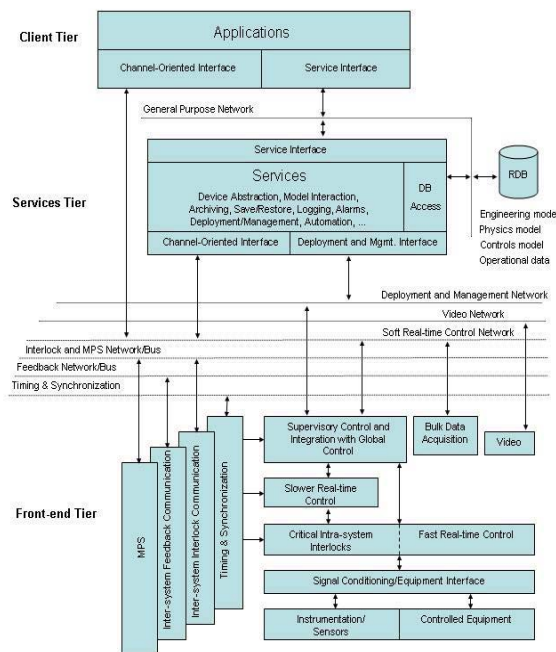

Figure 1: Controls systems functional model.

The functions of each tier are summarized below [5].

Client Tier

The Client tier consists of applications with which people directly interact. Applications will range from engineering-oriented control consoles to high-level physics control applications to system configuration management applications. Engineer-oriented consoles are focused on the operation of the underlying accelerator equipment. High-level physics applications will require a blend of services that combine data from the Front-end tier and supporting data from the relational database in the context of high-level device abstractions (e.g., magnets, BPMs).

Services Tier

From the perspective of a user of the client tier, the Services tier is largely invisible. The goal of the Services tier is to manage the execution of logic in the problem domain, and leave the problems of user interaction and graphical presentation of data and status to the Client tier. The Services tier provides services that coordinate many activities while providing a set of well-defined non-graphical interfaces. Device abstractions such as magnets and BPMs that incorporate engineering, physics, and control models are represented in this tier. An intrinsic component of the Services tier is an online relational database, which makes it possible to relate high-level machine parameters with low-level equipment settings in a standard and centralized way. This centralization of control provides many benefits in terms of coordination, security, automation, optimization, and conflict avoidance. For example, a parameter save/restore service can prevent two client applications from simultaneously attempting to restore a common subset of operational parameters.

Some candidate services are:
- Abstraction of devices and physics parameters
- Access to online engineering and physics models
- Dynamical feedback control
- Archiving of operational data
- Save/Restore
- Logging of control system status
- Alarm handling
- Deployment and management of installed software
- Machine automation

Front-End Tier

The Front-end tier provides access to the field I/O and underlying dedicated fast feedback systems. This tier is configured and managed by the Services tier, but can run autonomously. For example, the Services tier may configure a feedback loop in the Front-end tier, but the loop itself runs without direct involvement. The primary abstraction in this tier is a channel, or process variable, roughly equivalent to a single I/O point.

AVAILABILITY CONSIDERATIONS

Without a doubt, the most significant consideration for the control system reference design is the role the control system must play in delivering high accelerator up-time at peak performance, so as to deliver the design integrated luminosity. In broad terms, accelerator availability is affected by:
- Mean time between accelerator downtimes,
- Mean time to recover from a downtime event,
- Efficient startup after a maintenance period.

In addition to delivering intrinsic availability, the control system should provide functionality that supports:

- Rapid diagnosis and recovery from machine downtime events,
- Compensation for technical equipment failures via system-level reconfiguration, e.g., rebalance rf systems or reconfigure feedback loops,
- Rapid diagnosis of accelerator performance issues,
- Automation of routine processes, such as rf cavity conditioning.

The complexity and scale of the ILC machine places unprecedented reliance on the Global Control System to support operation. The Global Control System has been assigned a downtime budget of 1% (99% available) over a 5000 hrs/year operating schedule. Several factors make the availability goal particularly challenging:

- The control system must meet its availability goals from the onset of operations, so the designers cannot rely on an approach of continual improvement based on operational experience.
- Extensive reliance on machine automation and feedback make it difficult to implement a policy where technical equipment can continue to function when a control system failure occurs.
- In the pursuit of ever-increasing luminosity, the control system (and accelerator as a whole) will be subjected to frequent configuration changes.
- Meeting an overall availability of 99% requires that many individual control system components have availability approaching 99.999% (5-nines).
- There are many pitfalls to managing, testing, and integrating a high-availability control system that is developed across multiple sites and regions.

There are many well-known techniques that can be utilized to mitigate the various failure modes and effects[6]. These vary from simple, inexpensive administrative procedures to complex, costly, redundant components with automatic failover. Some techniques improve component reliability, while others reduce component time-to-repair. It is important to make clear that the intent is to selectively apply techniques where needed, based on analysis of both the failure modes and effects, and the cost benefit from deploying the techniques. A detailed failure modes and effects analysis of the control system is complicated by the fact that there is not a one-to-one correlation between control system downtime and accelerator downtime.

The control system reference design incorporates some high availability techniques explicitly, while other techniques are either implied or have yet to be fully analyzed in the context of control system design. Explicitly included are:

- Automation – the Services tier of the functional model and distributed computing nodes of the physical model are designed to support automation.
- Redundancy for commercial off-the-shelf (COTS) components – most of the commodity computing components such as servers, databases, networking switches, and file servers, all of which can be purchased in high-availability, redundant configurations.
- Extensive monitoring and diagnostics – the physical model implements an "out-of-band" monitoring subsystem and technical equipment Diagnostic Interlock Layer (DIL), collectively providing unprecedented levels of resource monitoring of the control system and of technical components.
- Configuration management – remote loading of code and configuration to servers, switches, FPGAs, etc. shall be possible via a deployment and management interface function, and implemented over the out-of-band monitoring network.
- Front-end electronics chassis with capability for hot-swap of electronics cards, redundant communications (inter- and intra-chassis), redundant power, and remote resource (shelf) management. The reference design shows front-end electronics chassis that meet the ATCA Standard [7]. This technology is a representative example of the breadth of high-availability functionality that will be expected of the electronics platform selection at project start time. There is also growing interest in ATCA [8] by the physics community.

CONTROL SYSTEM PHYSICAL MODEL

Figure 2 shows the control system physical model with the main components and features described below.

Controls Networks and Computing Services

Conventional computing services dedicated to the controls system will include storage arrays, file servers, and compute nodes. The overall philosophy is to develop an architecture that can meet the requirements, while leveraging the cost savings and rapid evolutionary advancements of COTS components [9].

Data collection, issuing and acting on set points, and pulse-to-pulse feedback algorithms are all synchronized to the pulse repetition rate. The controls network must provide adequate response and determinism to support this pulse-to-pulse synchronous operation. An initial assessment indicates that anticipated future commodity computing equipment would be more than adequate to meet network bandwidth and latency requirements.

Dedicated compute nodes associated with each backbone network switch will run localized control system services for monitoring, data reduction, and implementation of feedback algorithms.

CONTROLS FRONT END

The controls system model front end contains the following three main elements:

1U Switch: Aggregates the many Ethernet-controlled devices in a rack or neighborhood of racks. Some of these devices will speak the controls protocol natively, while others will have proprietary protocols that must be
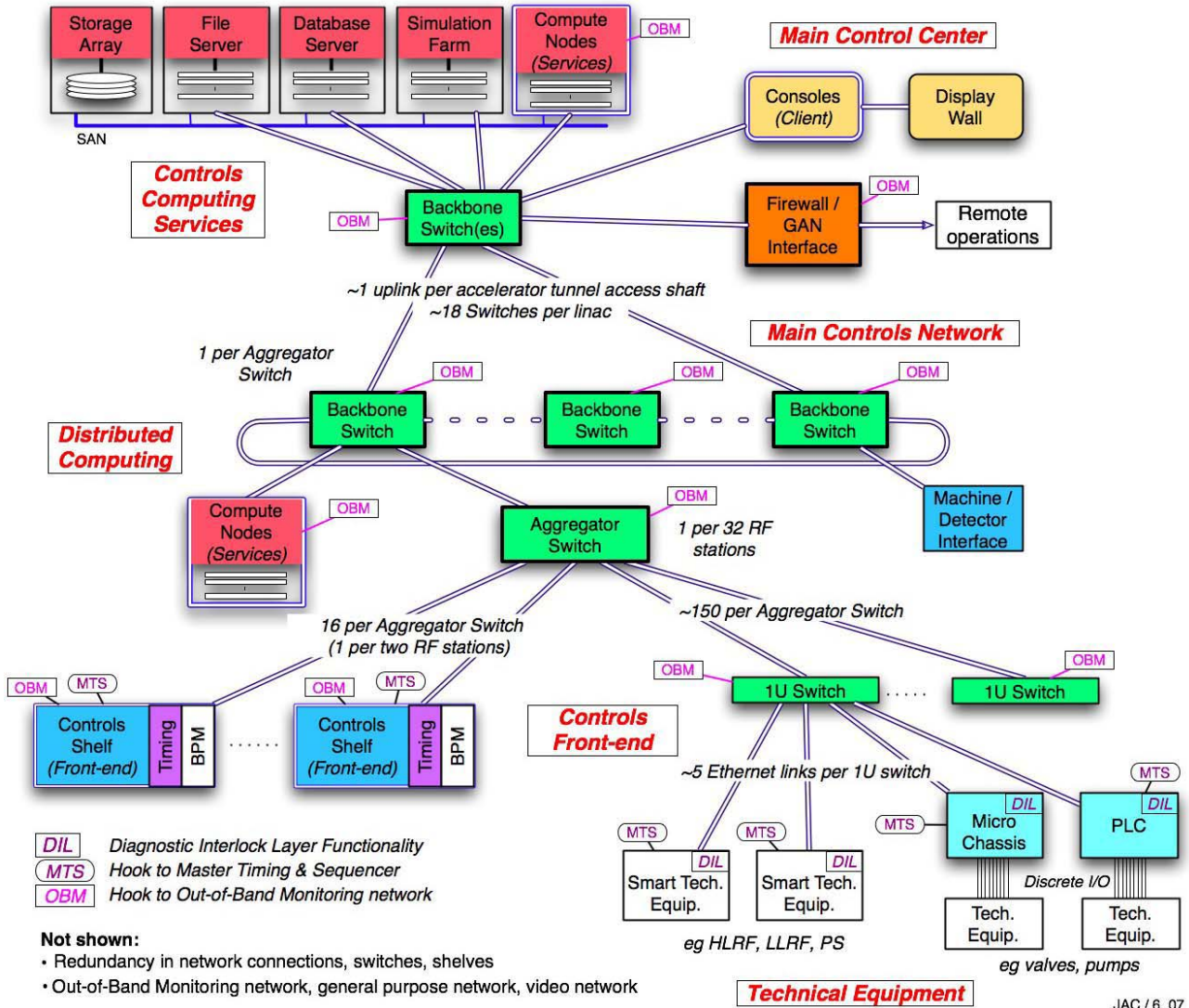
Figure 2: Global Controls Systems physical model as applied to the Main Linac.

interfaced to the control system. It is assumed these 1U switches will reside in many of the technical equipment racks.

Controls Shelf: Consists of an electronics chassis, power supplies, shelf manager, backplane switch cards, CPUs, timing cards, and instrumentation cards (mainly BPMs). The controls shelf serves several purposes: (1) hosts controls protocol gateways, reverse gateways, and name servers to manage the connections required for clients to acquire controls data, (2) runs the core control system software for managing the various Ethernet device communication protocols, including managing any instrumentation (BPM) cards in the same shelf; and (3) performs data reduction, for example, so that full-bandwidth rf/BPM waveforms need not be sent northbound in the control system.

Aggregation Switch: Aggregates network connections from the 1U switches and controls shelves and allows flexible formation of virtual local area networks (VLANs), as needed.

Technical Equipment Interface: It has been common practice at accelerator facilities for the control system to accommodate a wide variety of interfaces and protocols, leaving the choice of interface largely up to the technical system groups. The large scale of the ILC accelerator facility means that following this same approach would almost certainly make the controls task unmanageable, so we anticipate following an approach of specifying a limited number of interface options.

Initial front-end component counts for the accelerator complex are summarized in Table 1.

Table 1: Summary of Controls Equipment

Controls Equipment	Counts
1U Switch	8356
Controls Shelf	1195
Aggregation Switch	71
Controls network backbone switch	126

Feedback Architecture

As noted earlier, many of the beam-based feedback algorithms required for ILC will apply corrections at the relatively low machine pulse rate (nominally 5 Hz). This low correction rate and the distributed nature of many of the monitors and actuators make it desirable to use the integrated control system infrastructure for these feedback systems. This has the inherent advantage that dedicated interfaces are not required for equipment involved in feedback loops, making all equipment potentially available for use within synchronous feedback loops. Feedback algorithms will be implemented as services running in both distributed and centralized compute nodes.

Remote Access / Remote Control

It is becoming increasingly common for accelerator-based user facilities to provide means for accelerator physicists and technical experts to remotely access machine parameters for troubleshooting and machine tuning purposes. This requirement for remote access will be accentuated on the ILC because of the likelihood that expert personnel are distributed worldwide. There is ongoing development of remote access systems in both the accelerator and detector communities [10].

ENGINEERING DESIGN AND R&D

The ILC Global Design Effort is now entering into an engineering design phase intended to answer important R&D questions and to mature the technical design and value estimate. In support of authoring the controls chapter of the Engineering Design Report, the team will be acquiring further controls requirements, refining the architecture, and conducting a more precise value estimate.

Informing the entire process will be a collection of R&D work packages, mostly focused on the topic of high availability. It is crucial to gain experience with the tools and techniques applicable to particular failure modes in order to make value-based judgments of cost versus benefit. Research on high-availability controls falls into four broad categories: controls system failure mode analysis, high-availability electronics platforms, high-availability integrated control systems, and controls systems as tools for implementing high availability at a system level.

A strong component of the R&D plan will be to leverage ongoing and planned activities at beam facilities worldwide to focus activities on specific requirements and to gain field experience.

High-availability electronics platform research is currently focused on ATCA, which encompasses a number of hardware and software standards. High-availability integrated control system research involves investigation of a number of techniques such as conflict avoidance, controller redundancy and failover, model-based resource monitoring, and model-based configuration management.

Research on controls systems as a tool for system-level high availability focuses on the issues of fault detection (e.g., identifying a failed beam position monitor), automated diagnosis, and adaptive control.

Since we do not yet have a concrete design with which to conduct a formal Failure Modes and Effects Analysis (FMEA), the goal is instead to collect failure modes and downtime data from existing facilities using more conventional controls system designs. Such data can be used to enhance an existing availability simulation [11], which in turn can provide guidance as to priorities for improving aspects of our design.

REFERENCES

[1] http://www.linearcollider.org/cms/
[2] International Linear Collider Reference Design Report, ILC-REPORT-2007-001, 2007.
[3] F. Lenkszus et al, 'Timing and RF Phase Reference for ILC Reference Design,' ILC-NOTE-2007-025, 2007
[4] C. Saunders et al., Control System Architecture Model for ILC Reference Design, ILC-NOTE-2007-022, 2007.
[5] A. Gotz, D. Schmidt, M. Clausen, "Middleware in Accelerator and Telescope Control Systems," Proc. of ICALEPCS 2003, Gyeongju, Korea, Oct. 13-17, 2003, p. 322 (2003), http://www.jacow.org.
[6] E. Marcus, H. Stern, *Blueprints for High Availability*, Second Edition, (Wiley Publishing Inc.:2003).
[7] http://www.picmg.org
[8] R.W. Downing, R.S. Larsen, "High Availability Instrumentation Packaging Standards for the ILC and Detectors," SLAC-PUB-12208,
[9] M. Votava et al, "Commodity Computing Architecture for ILC Reference Design" ILC-NOTE-2007-026, 2007
[10] S. Karstensen et al, "Remote control and monitoring of Accelerators and Detectors in a global facility (GAN/GDN)," IEEE NPSS Real-Time Conference 2007, Fermilab, Il.
[11] N. Phinney et al, "Reliability Simulations for a Linear Collider," SLAC-PUB-10505 (2004).