LIGO Data Analysis and Simulation Systems

Development Status

A.Lazzarini

NSF Spring Review
27 April 1999
LIGO Livingston Observatory
Livingston, LA

LIGO-G990027-00-E

CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PB_G3_HD:lazz:LIGO:Presentations:NSF_9904:NSF_99_04_v2.fm55

Outline of this talk

- Responses to NSF Recommendations from the Fall Review
- E2E Status
  - Overview of End-to-End simulation system [E2E]
  - Software development status
  - E2E implementation strategy
  - Development schedule
- LDAS Status
  - Overview of LIGO Data Analysis System [LDAS]
  - Software development status
  - Operational plan for archive at CACR
  - LDAS implementation strategy
  - Development schedule
  - Procurement plan

LIGO Data Analysis System (LDAS)
Recommendations *(from 10/98 review)*

1. The LSC and LIGO should organize beginning-to-end infrastructure and data analysis software validation tests including embedded signals (Mock Data Challenges, MDCs)

- 40m prototype analysis [discussed by B. Allen] began sequence of MDCs.
  - Validation of paradigm for a certain class of signals
  - Used REAL data (vis-a-vis simulated data)
  - The analysis DID NOT exercise the LDAS system
  - Does not exercise the number of signals LIGO has
- LIGO and LSC will continue with sequence of MDCs targeting specific subsets of LDAS SW releases - α, $\alpha 2$, β, release 1, ...
  - Incremental testing allows debugging components, user feedback, etc.
  - End-to-end test is culmination of MDC sequence
  - => No [big] unexpected surprises
- The top-level schedule presented later identifies a proposed the sequence of MDCs
  - Details of MDCs to be worked out jointly with LSC

LIGO Data Analysis System (LDAS)
Requirements *(from 10/98 review)*

2. The LIGO Director and the LSC spokesman must find a way to provide validated physics analysis software for the collaboration and exercise it through the MDC.

- LIGO and LSC have begun to address issue:
  - LIGO and LSC data analysis chairs instituted standing monthly meeting (~8 attendees):
    - Agenda is developed;
    - Minutes are circulated;
    - Issues common to both groups aired and discussed;
    - Actions are assigned.
  - Last LIGO-LSC meeting focused on software development.
    - Consensus on how to proceed.
    - Priority-ordered list of needed software developed
    - First ~ 3 components to be developed as examples for a style guide.
  - DRAFT specification and style guide for code development drafted, reviewed.
    - Accepted as a starting specification.
  - Timeline for software deliveries to be developed

LIGO Data Analysis System (LDAS)
Requirements *(from 10/98 review)*

3. The LSC spokesman should assure that the constraints inherent in the computing model adopted by LIGO Laboratory is well understood by the LSC.

- All-day LIGO-LSC information exchange on 26 February 1999.
- The LDAS Preliminary Design Review [PDR] held 11 March 1999.
  - The review panel comprised 50% LSC representatives.
- Recent teleconference held by LSC spokesman to review possible LSC issues with LDAS
  - No “show stoppers” identified
  - LIGO & LSC committed to working together to resolve issues as they appear.

LIGO Data Analysis System (LDAS)
Requirements *(from 10/98 review)*

4. Given the distributed and broad nature of the LIGO Scientific Community, access to data for analysis will require either a central computing facility open to all participants or a highly distributed system. Operations plan for the LIGO central facility and CACR's operations plans for the support of the facility should be presented at the next review.

- Resources for LIGO/LSC use at CACR either OWNED by or FULLY DEDICATED to LIGO; will be under LIGO/LSC control.
- Resources limited in sense that more analyses could be performed than the computational power being provided can support. Addressed in two ways:

1 Key analyses organized by LSC in a team-based approach.
- Groups of researchers with common interests will work together.
- Analyses will be proposal-driven.
- Statement of interest and analysis approach submitted to LSC for prior approval.
- LIGO resources provided to approved analyses. Allocation on priority basis, determined by LSC.

2 Researchers may request reduced data sets:
- Analysis on other resources
  - home institutions or
  - via other grants (e.g., independent access to supercomputing resources).
- Approval by LSC still required.

LIGO Data Analysis System (LDAS)
Requirements *(from 10/98 review)*

4. Given the distributed and broad nature of the LIGO...

- LIGO Laboratory provides the initial resources;
  - LIGO Laboratory cannot become "full service provider" to LSC: we do not have sufficient resources.
  - LSC participation with in-kind [FTE, materials, ...] support needed:
    - Post-docs, GRAs for certain activities, etc.

5. The LIGO leadership should aggressively seek to fill the four budgeted software positions and should be flexible about hiring conditions (salaries, telecommuting, consultants, contractors) in order to find qualified staff in a timely manner.

- LIGO has been working on this aggressively since the last review:
- To be discussed below

6. Because simulations are very important, we recommend that status and plans for simulation software, including the end-to-end simulation support package, be an additional focus of the next meeting of this panel.

- Overview presented next
- Details available tomorrow.

LIGO Data Analysis System (LDAS)
Requirements *(from 10/98 review)*

7. LIGO Laboratory should evaluate the need for expanded software manpower in the operations phase to support data analysis for the LSC. This evaluation should be completed on a time scale that will allow for an additional budget request.

- Issue noted,
  - Consider carefully valid “in-scope” functions vs.
  - "service" functions to be shared with LSC.
- LIGO positions under consideration:
  - LIGO-wide SW librarian
  - LIGO FTE support/maintenance for LIGO resources at CACR.
  - Observatory system administrators for analysis hardware
- LSC positions needed:
  - SW coordinator
  - Data analysis coordinator
- Issue will be addressed in the renewal proposal for LIGO Laboratory operations for the period beginning with LIGO I.

Simulation:

End to End Model [E2E]

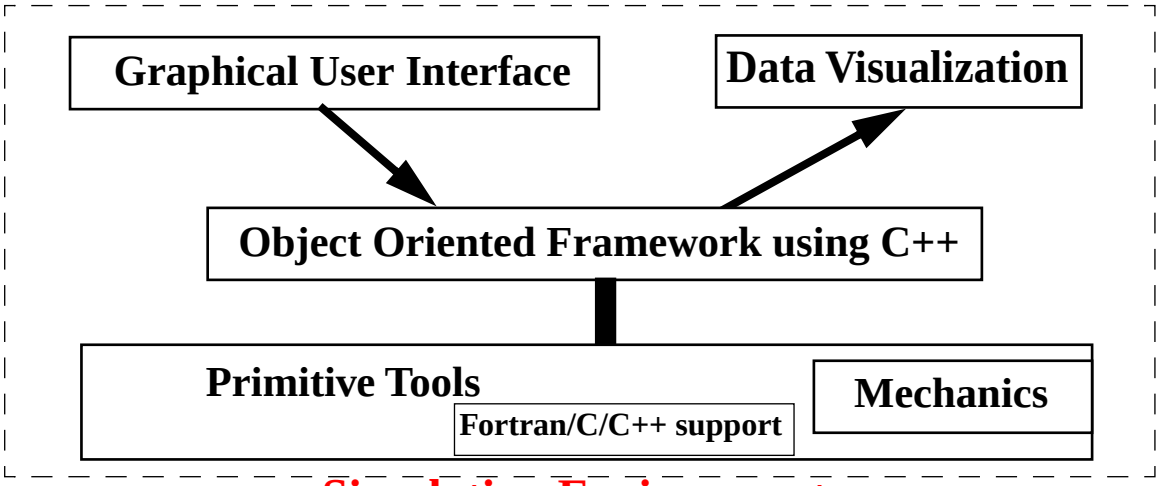
End to End model
Overview

- End to End simulation framework is ready
  - Written in C++
  - Built-in tools for GW detector simulation
    - laser, mirror, digital filter, etc.
  - Optimized for time domain simulation
  - Time domain modal model implemented
  - Alfi: Graphical User Interface completed
  - E2E can replace existing [limited] detector modeling tool for servo design
- Examples of simulation capabilities:
  - Thermal noise spectrum of LIGO suspensions/mirrors
  - 40m prototype IFO modeling
  - Cavity lock acquisition process
  - Dual recycling configuration studies [advanced R&D]
- Development is collaborative:
  - Caltech, Univ. of Florida, Penn State, Univ. of Pisa
  - MOU attachments for cooperative simulation development with international partners:
    - TAMA
    - GEO
    - VIRGO

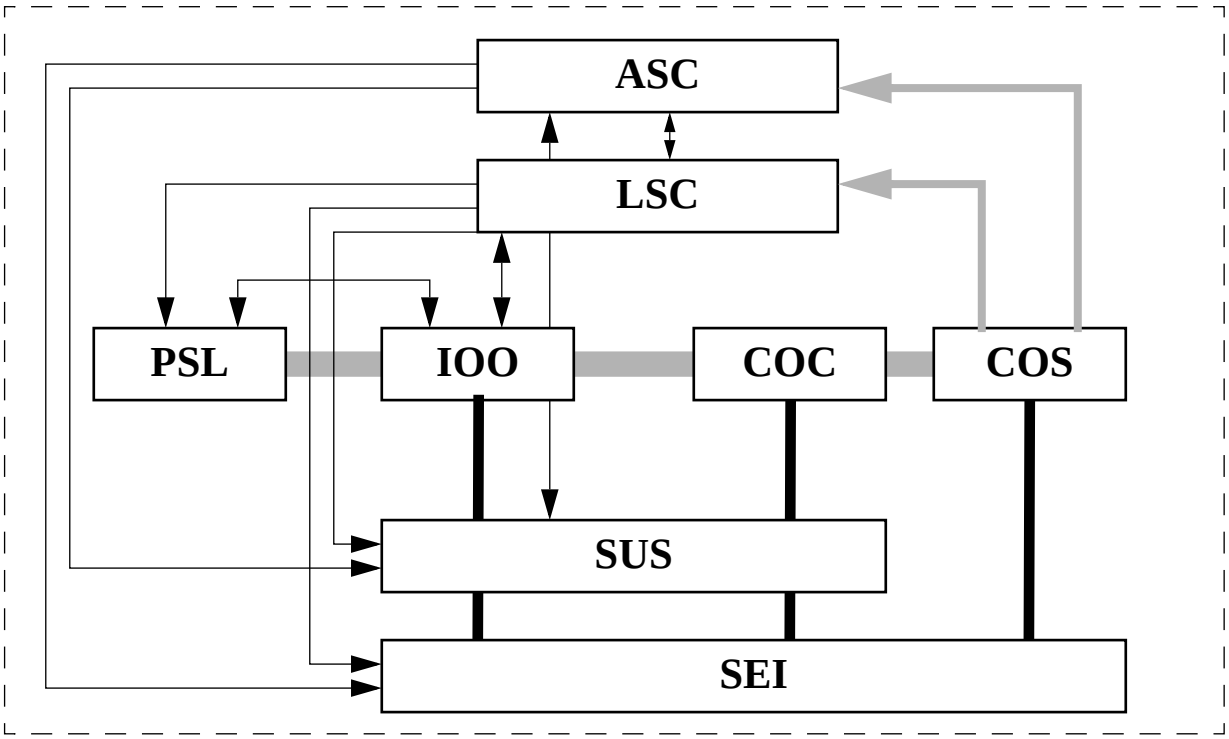
Development Team

- Continued E2E framework and optics modeling improvement:
  - B.Bhawal, M.Evans, H.Yamamoto @ Caltech
- Mechanical systems modeling framework:
  - S.Mohanty / Penn State, G.Cella / Univ. Pisa
- Integrated subsystem models
  - PSL: R.Savage / Hanford, P.King / Caltech, B.Bhawal / Caltech
  - IO: S.Klimenko / Florida
  - LSC/ASC, Lock Acquisition: P.Fritschel / MIT, M.Evans / Caltech
  - SUS/SEI: G.Cella / Pisa, H.Yamamoto / CIT, + ...
- 40m model development
  - A.Weinstein / Caltech
- GUI / Alfi improvement
  - E.Maros / Caltech

E2E Model
Organization



Simulation Environment



Subsystem Implementation

Schedule
Subsystem development

1 PSL subsystem [Savage, King, Bhawal]
   - 10% complete
   - initial implementation due 7/99

2 IO subsystem optics [Klimenko]
   - 60% complete
   - initial implementation due 7/99

3 Simple suspended 3D mirror assy. based on matlab model [Mohanty, Yamamoto]
   - scheduled for: 5/99 - 7/99

4 Simple seismic isolation model & integration with suspended mirror (#3).
   - planned for 6/99 - 7/99.

5 Integration of (1)-(4)
   - planned for 7/99 - 8/99

6 LSC/ASC for 2km arm [Fritschel, Evans]
   - planned for 5/99 - 9/99

7 SEI/SUS of COC with ground motion correlation & integration with control (#6) [Cella, Yamamoto]
   - planned for 8/99 - 9/99

- Comparison of model with 2km FP cavity: 9/99

Schedule

Planned improvements to model

1 Modular mechanical system model [Cella]

- 30% complete
- basic version to be delivered by 7/99

2 Integration of (#1) in E2E environment

- 6/99 - 7/99

3 Improvements of the modular model [Bhawal]

- continuing, after 7/99

4 Parallelization of E2E, fast simulation of in-lock LIGO, etc. planned after 9/99 [Yamamoto, Evans]

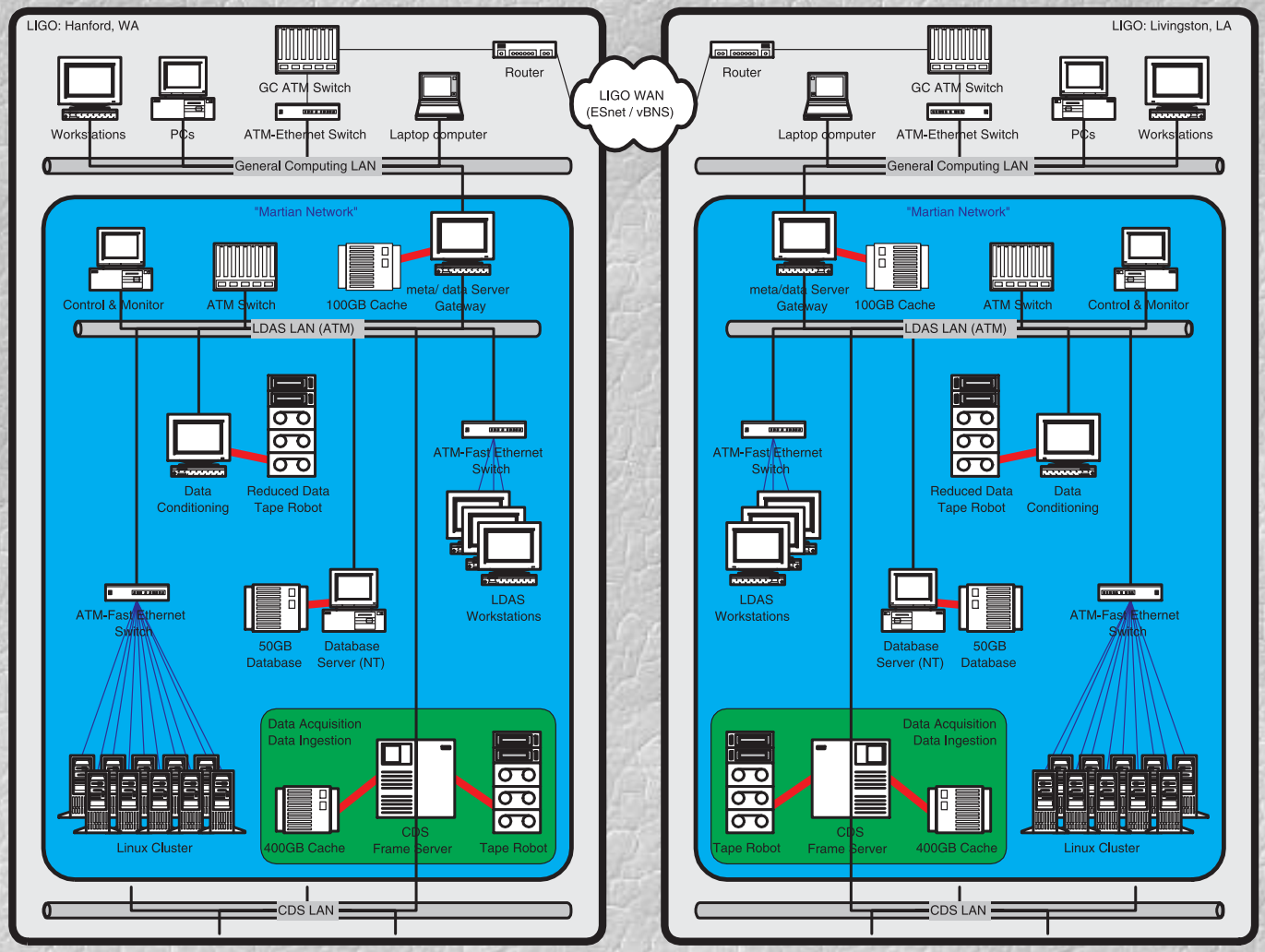
LIGO Installation Major Milestones
E2E Model Deployment

Detector Milestone		E2E Model	
LHO 2km Start	7/98		
PSL + Input Optics	3/99 - 9/99	PSL + IO Suspended TM Simple seismic	7/99 5/99 - 7/99 5/99 - 7/99
Single arm cavity tests	8/99 - 10/99	Integrated cavity w/o ctrl LSC/ASC Cavity w/ctrl	8/99 9/99 9/99
Power Recycled Vertex Michelson complete	11/99 - 3/00		
LHO 2km IFO complete: $h[f] < 10^{-20}$	3/00 - 11/00		

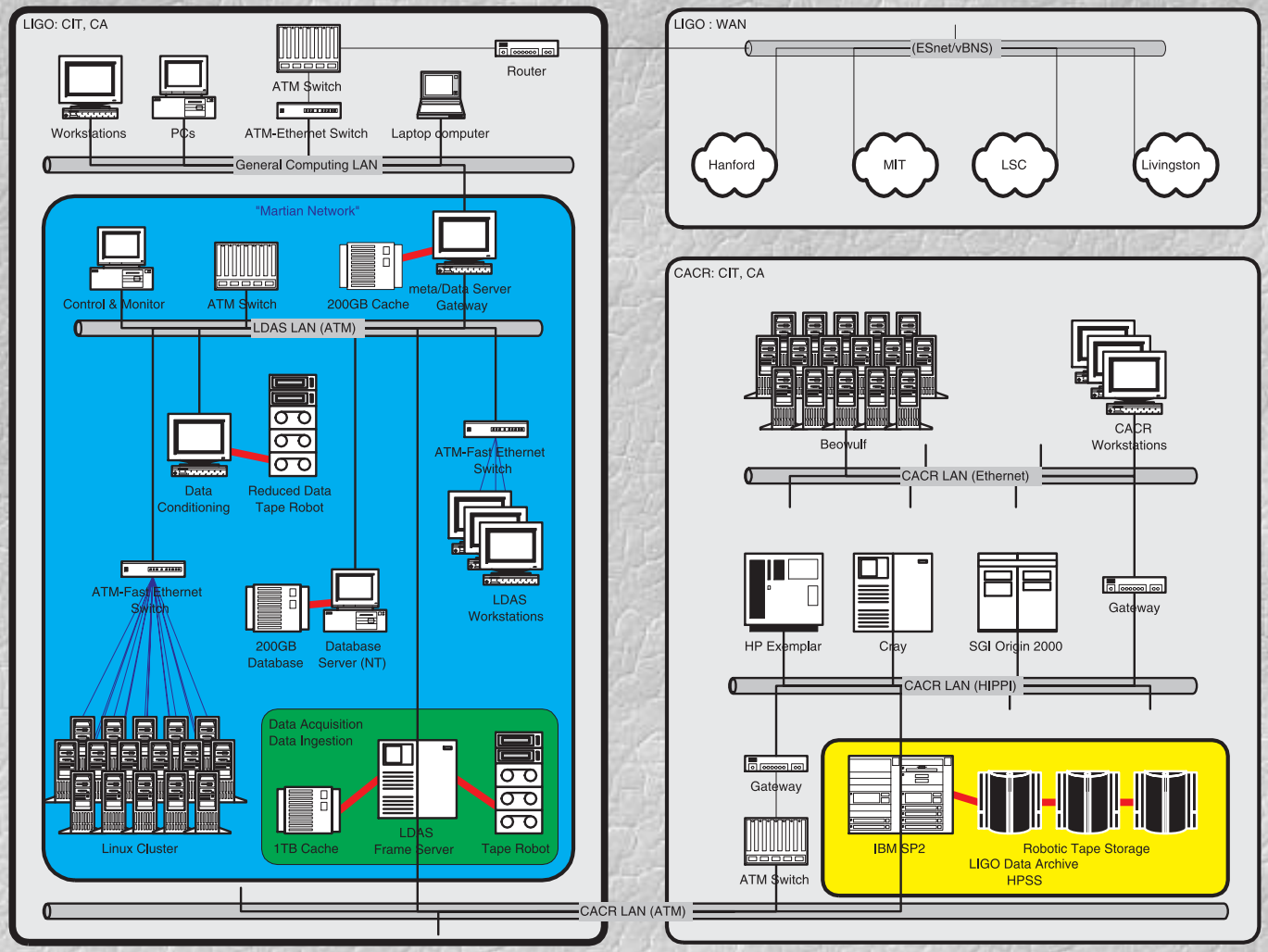
LIGO Data Analysis System

[LDAS]

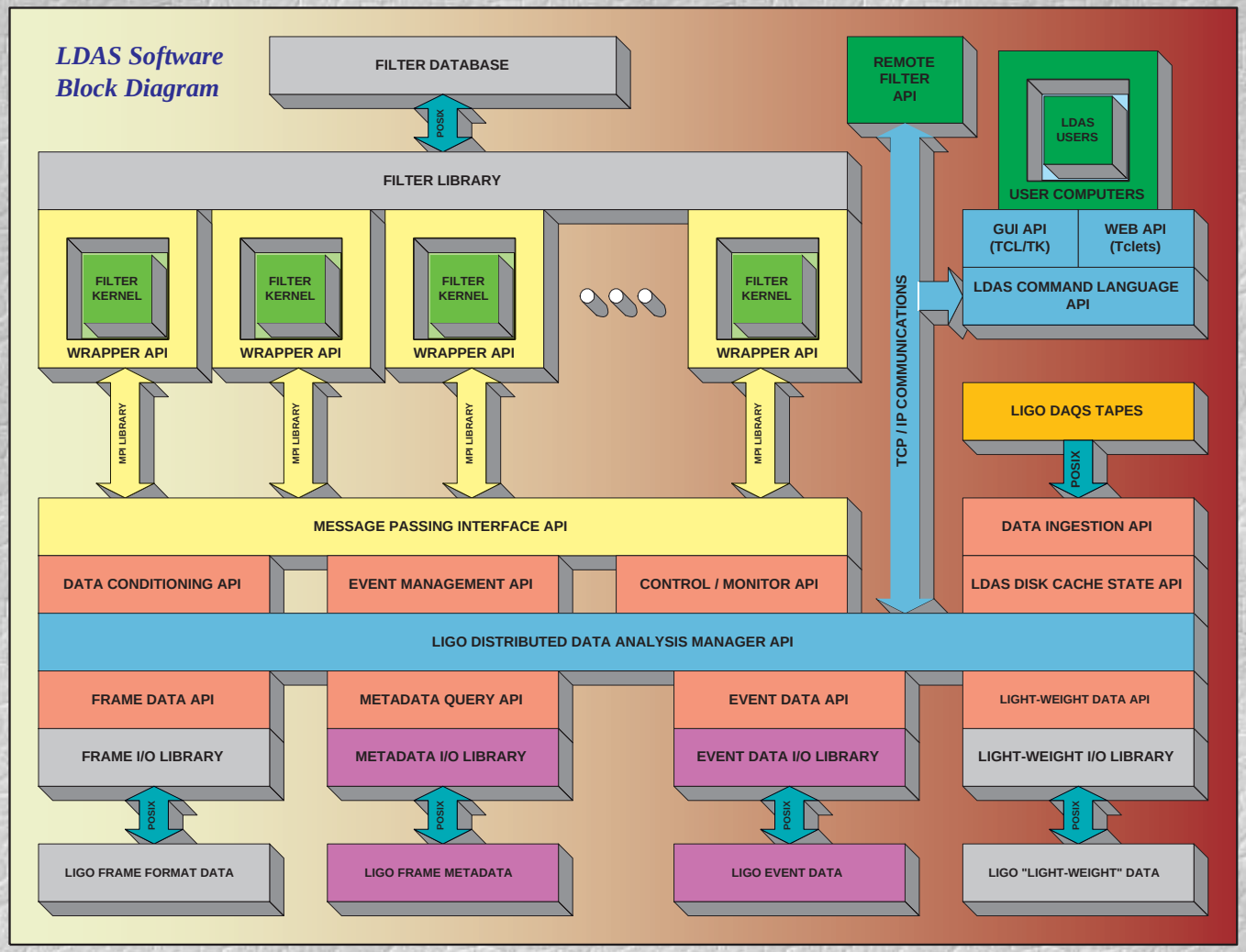
Hardware Architecture(online):



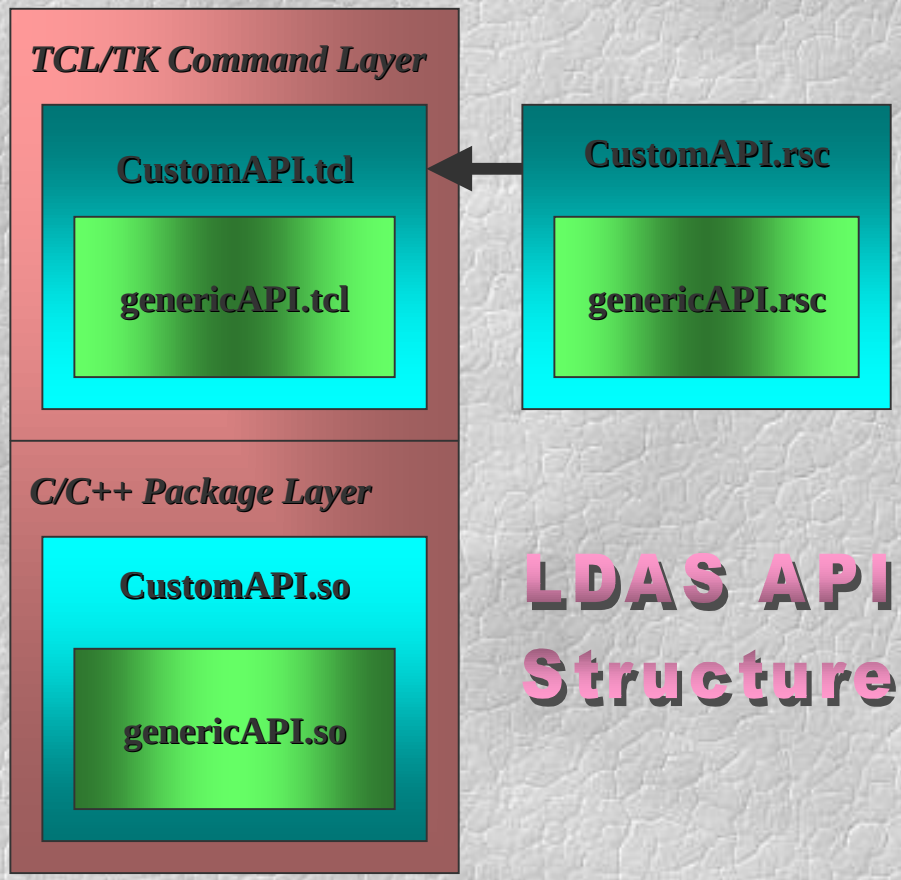
Hardware Architecture(offline):



Software Architecture(block):



Layered LDAS API Design:



LDAS API Structure

LDAS API's:

1. Two Layers:
   - TCL/TK
   - C/C++ *(extends TCL Language)*
   - SWIG Unifies Layers
2. GenericAPI *(core)* Module:
   - Communications
     - ✓ *TCL <-> C++*
     - ✓ *API <-> API*
   - Common TCL proc's:
     - ✓ *Help*
     - ✓ *Logging*
     - ✓ *Command Socket Management*
     - ✓ *Resource Management*
   - Common C/C++ methods:
     - ✓ *Data Socket Management*
     - ✓ *Internal Data Management*
     - ✓ *Class Save & Restore*
3. *Custom (specialization) Module*

LDAS Development Status
- General -

- LDAS Preliminary Design Review held 11 March 1999
  - Review panel consisted of future “customers”:
    - Chaired by detector deputy group leader
      - Panel report is in process of being written
    - LIGO Scientific Collaboration chairs of the data analysis groups
    - LIGO Observatory staff scientists
- Software programming is nearly fully staffed:
  - Promoted a contract programmer to LIGO staff position
  - Database programmer (hired)
  - Second Tcl programmer (hired)
  - Senior C++ programmer (will be hired through JPL)
  - Junior C++/MPI programmer (will be hired through JPL)
- MOU with CACR under revision
  - Identifies CACR commitments to LIGO/LSC:
    - Dedicated access to HPSS space for tapes, networks
    - Floor space for LIGO hardware (HPSS, Beowulf?)
    - Staff support for HPSS, parallel machine architectures, scientific computing
    - Licenses, H/W pricing structures
  - Identifies LIGO commitments to CACR:
    - FTE support for LIGO-provided and controlled hardware
  - LIGO will augment existing HPSS infrastructure for its own use.

LDAS Development Status
- Software -

- Software development:
  - Phased installation plan for LDAS developed
    - Software development schedule organized around 4 phases
  - α Release will consist of a subset of the full LDAS
    - **managerAPI: 100% complete**
      - receives requests from userAPI and schedules necessary scripts and schedules.
      - individual meta-scripts used to carry out requests must be developed
    - **genericAPI: 99% complete**
      - provides all the common communication components to the LDAS APIs as reusable code. This included the OO Internal Light Weight Data Format.
      - usage has identified occasional bugs to fix
    - **frameAPI: ~ 80% complete**
      - carries out Frame File I/O, translates frame format data into internal light-weight format data and visa versa. Can select, extract, and concatenate data from frames.
      - New OOP Framecpp library completed with full support for exception handling

LDAS Development Status
- Software -

- **metaDataAPI: ~ 50% complete**
  - Interfaces LDAS with commercial database server.
  - ODBC C++ layer client complete
  - Working DB2 Tables designed and implemented
  - Integration of genericAPI.so and ODBC.so shared objects complete.
  - API still needs TCL command super structures.
- **lightWeightAPI: started, ~ 20%**
  - Carries out XML based language I/O. Fully validating DTD C parser has been identified.
  - Formal requirements pending
  - High level XML DTD objects have been drafted
  - Work on wrapping C parser expected to begin soon after new staff arrive
- **dataConditionAPI: started, < 20%**
  - Responsible for basic signal conditioning filters.
  - Formal requirements complete.
  - Working with LSC to develop Analysis Library with needed algorithms for filters.

LDAS Development Status
- Software -

- **userAPI: started, ~ 5%**
  - A subset of the full LDAS userAPI will be developed for the alpha release of the software which supports the alpha release LDAS APIs.
  - Will provide user connectivity with the LDAS managerAPI and be capable of handling the data types supported in the alpha release
- **Relational database design proceeding: ~ 75%**
- Frame metadata descriptors, tables defined and extant 40m data being ingested for testing
- LIGO event definitions prototyped:
  - Astrophysical events for chirps being ingested from 40m prototype data analysis (Allen et. al.)
  - Diagnostics events/triggers tables defined with collaboration from the Global Diagnostics System design team
    - 40m dataset will be filtered by GDS triggers to create prototype database for querying.
  - Testing will exercise LDAS API components

LDAS Development Status
- Software -

- **LIGO/LSC Analysis Library [LLAL]: started, <5%**
- Specification in draft form is being reviewed by LIGO, LSC
- First ~3 critical algorithms to be developed in next months
  - Serve as prototypes/templates for style guide.
  - Testing and validation will be within the LDAS framework
- **Electronic logbook [iLog] developed and installed for LIGO: 100% complete**
  - Adopted FNAL's Elog and modified the source [Perl] to accommodate LIGO needs
  - Installed at both observatories for use in:
    - Detector installation/integration logging
    - Control & Data Systems logging
    - Facilities/bakeout logging
  - iLog will become the LIGO operations logbook environment
  - key components data logs will be ingested into the metadatabase for long-term archival and correlation with data analysis

LDAS Development Status
- Hardware -

- Cluster [parallel] computing:
  - Installed, debugged prototype 8-node/16CPU Linux cluster from Alta Tech.
  - Developed new tools, ported existing tools from CACR to maintain a stable Linux configuration that will inter-operate with the standard LIGO network and software tools:
    - NFS to LIGO servers,
    - NIS+ user accounts,
    - automatic O/S installation on new nodes, ...
  - Performed preliminary benchmarks:
    - Disk I/O: 90Mbyte/s
    - Network: 60Mbyte/s
    - Small FFT: 2.1Gflop (16×2^{13})
      - Note: due to 512kB cache no penalty for using both CPUs per node.
    - Large FFT: 700Mflop (16×2^{23})
      - Note: this is 7% slower than using only 1 CPU/node and multiplying by 2 due to limited memory bandwidth.

LDAS Development Status
- Hardware -

- Next: set up a real-time template search testbed to test:
  - stability of various versions of Linux and MPI
  - continuous transmission of 40m IFO frames to the prototype cluster for template searching.
- Two algorithms identified for testing extremely large (> 1Gpoint) distributed FFTs on cluster for coherent pulsar searches.

- High speed network development:
  - The ATM connection upgraded from OC-3 (155Mbit/s) to OC-12 (622Mbit/s).
    - Initial benchmarking of the CACR High Performance Gateway Node (HPGN):
      - ATM to SP2 backplane packets, HiPPI, Fast Ethernet
      - ~300Mbit/s for most routes
      - As high as 480Mbit/s for tuned packet size HiPPI to HiPPI transfers.

LDAS Development Status
- Hardware -

- Mass storage research.
  - Testing and benchmarking of the recently upgraded CACR HPSS, see
    - **http://pcbunn.cithep.caltech.edu/presentations/briefing/index.htm** for details of upgrade
    - Transfer rates up to 13MB/s for Sun workstation using parallel ftp over OC12 through HPGN to two separate SP2 data mover nodes.
    - Developed software to automatically and robustly archive 40m data into the CACR HPSS. Initial test ran successfully for a mock 24 hour data run.
    - Developed a baseline proposal for CACR/LIGO MOU covering the long term mass storage collaboration.
    - Developed affordable model for an on-site archive to hold weeks of frame data to maintain duplicate copies during tape transfer to the main LIGO archive, and to facilitate unexpected on-site data analysis needs.
    - Developed a detailed hardware implementation concept for main LIGO data archive
      - based on recently upgraded CACR installation
      - allows detailed prototyping before commitment to procurement of hardware.

LDAS Implementation Strategy

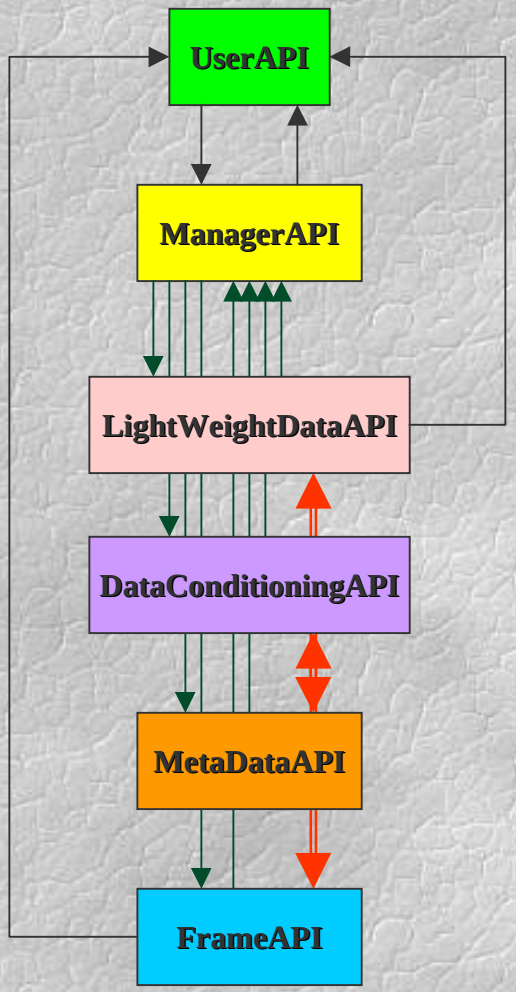
- Focus on initial installation for 2km IFO at Hanford
- Initial α release (partial LDAS) 6/99
  - data distribution/archival
  - limited data conditioning
  - metadata search/archival
  - no MPI
  - limited user base to test system
  - Supports initial commissioning of PSL + IO
- α2 release (full LDAS) 10/99:
  - 8 - 16 node beowulf for data analysis (diagnostics)
  - software update, includes MPI
- Full LDAS β release at 12/00
  - full user interfaces
  - includes HPSS archiving capability
- First official release in 01/02
  - concurrent with LIGO I science run

LDAS Development Strategy

- Full off-line system availability contemporaneous with full LDAS release
  - large linux cluster
  - HPSS + data servers
  - data conditioning
  - data ingestion
- Replicate design at each stage of development as needed for other interferometers.

First Software Deliverable:

- **α-release Summer 99**
  1. LDAS Components to be included:
     - GenericAPI
     - FrameAPI
     - Meta(Event)dataAPI
     - LightWeightDataAPI
     - DataConditioningAPI
     - ManagerAPI
     - Simple UserAPI
  2. Some "Quick Look" Capabilities
  3. Support of Site Installation Activities
  4. Provide User Base for Testing and Debugging System



LDAS Deployment & Mock Data Challenges [MDCs]

- System tests are staged in 4 phases.
  - Incremental MDCs will be matched to the system capabilities at each stage.
- Initial α release
  - Mock-data/benchmarking runs:
    - bandwidth tests:
      - user loads vs. time-of-service
    - software verification of software:
      - filtering routines
      - data re-sampling
      - metadata queries
- α2 release
  - Possible mock-data/benchmarking runs:
    - Repeat a subset of α test, plus:
      - MPI-based analysis/filtering:
      - bandwidth/stress testing of system

LDAS Deployment & Mock Data Challenges [MDCs]

- Full β release
  - Mock data runs:
    - Bandwidth/stress testing of systems
      - Verification of user interfaces
      - HPSS access/retrieval times
      - End-to-end GW search tests using templates/filter-based pipeline analysis with PC linux cluster(s)
- First official release
  - Mock data runs:
    - GW search tests using templates/filter-based pipeline analysis with PC linux cluster(s)

LIGO Installation Major Milestones
LDAS Deployment

Detector Milestone	LDAS S/W	LDAS H/W
LHO 2km Start 7/98		
PSL + Input Optics 3/99 - 9/99	1st α Release 6/99	Data distribution, LHO, 5/99
Single arm cavity tests 8/99 - 10/99	2nd α Release 10/99	
Power Recycled Vertex Michelson complete 11/99 - 3/00		
LHO 2km IFO complete: h[f] < 10^{-20} 3/00 - 11/00		
LLO 4km Start 1/99		
PSL + Input Optics 4/99 - 12/99		Data distribution, LLO 9/99
Single arm cavity tests [?]		
Power Recycled Vertex Michelson complete 2/00 - 5/00		PC cluster, LLO 2/00 - 6/01
LLO 4km IFO complete: h[f] < 10^{-20} 5/00 -2/01	Full β Release 12/00	Local archive, LLO 8/00

LIGO Installation Major Milestones
LDAS Deployment

Detector Milestone	LDAS S/W	LDAS H/W
LHO 4km Start 7/98		
PSL + Input Optics 8/99 - 3/00		
Single arm cavity tests [?]		PC cluster, LHO 11/99 - 5/01
Power Recycled Vertex Michelson complete 3/00 - 7/00		Local archive, LHO 5/00
LHO 4km IFO complete: h[f] < 10^{-20} 7/00 - 2/01		
Design sensitivity: h[f] < 10^{-21} 1/02	1st Official Release 12/01	CACR: HPSS archive 1/01 - 6/01 Data distribution PC cluster 10/00 - 6/01
First science run: 3X operation 02/02	LDAS Available	

List of acronyms

Acronym	Description
Alfi	LIGO's simulation package graphical front end
API	Application Programming Interface
ASC	Alignment Sensing & Control subsystem
ATM	Asynchronous Transfer Mode (data transmission protocol for networks)
B, kB, MB GB, TB	Byte (8 bits), kiloByte, megaByte, gigaByte, teraByte [2^n bytes for n = 0,10,20,30,40]
Beowulf	Class of multiprocessor parallel architecture comprised of many identical commercial PCs or workstations networked to function as a parallel machine
C,C++	Programming languages
CACR	Center for Advanced Computing Research
COC	Core optics Components subsystem
COS	Core Optics Support subsystem
CPU	Central Processing Unit, a computer (same as node)
DB2	IBM's relational database management system
DTD	Document Type Definition, dictionary of valid data for XML parsers
E2E	End-to-End LIGO Simulation Environment
E-log, iLog	Electronic notebooks for logging data
FFT	Fast Fourier Transform
FNAL	Fermi National Accelerator Laboratory
FTE	Full-Time Equivalent
GDS	Global Diagnostics System (for the detector)
GEO	British-German Interferometer Program
GRA	Graduate Research Assistant
GUI	Graphical User Interface
HiPPI	High Performance Peripheral Interface (network for supercomputing systems)
HPGN	High Performance Gateway Node
HPSS	High Performance Storage System, a robotic system for mass data archival
HW, H/W	Hardware
IFO	Interferometer
IO	Input Optics System
I/O	Input/output

List of acronyms

Acronym	Description
JPL	NASA Jet Propulsion Laboratory
linux, UNIX	Computer operating systems (written in C)
LDAS	LIGO Data Analysis System
LHO	LIGO Hanford Observatory
LLO	LIGO Livingston Laboratory
LSC	LIGO Scientific Collaboration, also Length Sensing and Control
MBit/s	Data Transmission bandwidth, megabits per second
MC	Mode Cleaner
MDC	Mock Data Challenge
MOU	Memorandum of Understanding
MPI	Message Passing Interface, protocol for parallelized computation
NSF	National Science Foundation
OC3, OC12	Optical Carrier bandwidth specifications 155 and 622 MBit/s
ODBC	Open Database Convention
OOP	Object oriented programming (design)
OS, O/S	Operating system (for a computer)
PC	Personal computer (or workstation)
PERL	A scripting and command language for UNIX systems
PDR	Preliminary Design Review
PSL	Prestabilized Laser
SEI	Seismic Isolation subsystem
SUS	Suspension subsystem
SW, S/W	Software
TAMA	Japanese Interferometer Project
Tcl/Tk	Tool Command Language, a scripting language
XML	Extensible Markup Language
VIRGO	French-Italian Interferometer Project