Data Analysis,

Simulation

and Collaborative Activities

Status

Albert Lazzarini

PAC Meeting
16,17 November 1998

LIGO Laboratory at Caltech

Pasadena, CA

Outline

- LIGO Data Analysis System (LDAS)
  - Data types and products
  - Database needs
  - Software development & prototyping
  - Hardware prototyping
- Simulation environment
  - End-to-end model
- Collaborative activities
  - VIRGO
  - GEO
  - TAMA

LDAS Status Update

- Software design essentially complete
  - PDR scheduled for January.
  - Software components specification for Application Programmer Interfaces (APIs) complete
  - Lightweight data type specification released
  - Selection of database management system is pending
- Completed a GenericAPI template, from which specific APIs will be extended
- Prototyping activities under way in several important areas:
  - Software module development
  - Data distribution using web tools
  - Interprocess communications, data transmission
  - Data flow for (directed) pulsar searches
- Hardware conceptual design complete
  - On-line systems at observatories
  - Off-line system at data repository (CACR/Caltech)
  - Wide area network for inter-site connectivity
- Staff has been increased
  - ✓ 1 senior scientist
  - ✓ 3 programmers on staff;
  - 1 offer out;
  - seeking additional 2 programmers

LIGO Data Types/Products
Frame format data

- The full [raw] detector datastream will be acquired and recorded as data frames.
  - Format for data frames has been unified with VIRGO in anticipation of being able to share software (now) and data (at some future date)
  - Other major interferometer projects have adopted standard
    - GEO
    - TAMA

- Frame Class Library (C++ implementation)
  - Implements Frame Format Specification
  - Progress to date:
    - v1.01
    - Documentation available on web (http://docuserv.ligo.caltech.edu/~wmajid/fcl/index.html):
    - HowTo's, Sources
    - Compatible and interfaced with CERN's ROOT package
    - Fcl to LigoLW(XML) translation module completed
  - Planned work (next quarter):
    - Implement interface to Matlab
    - Provide additional UNIX shell tools
    - Develop Tcl level Frame API for incorporating Fcl into LDAS architecture

LIGO Data Types/Products

Lightweight Data Format -- XML

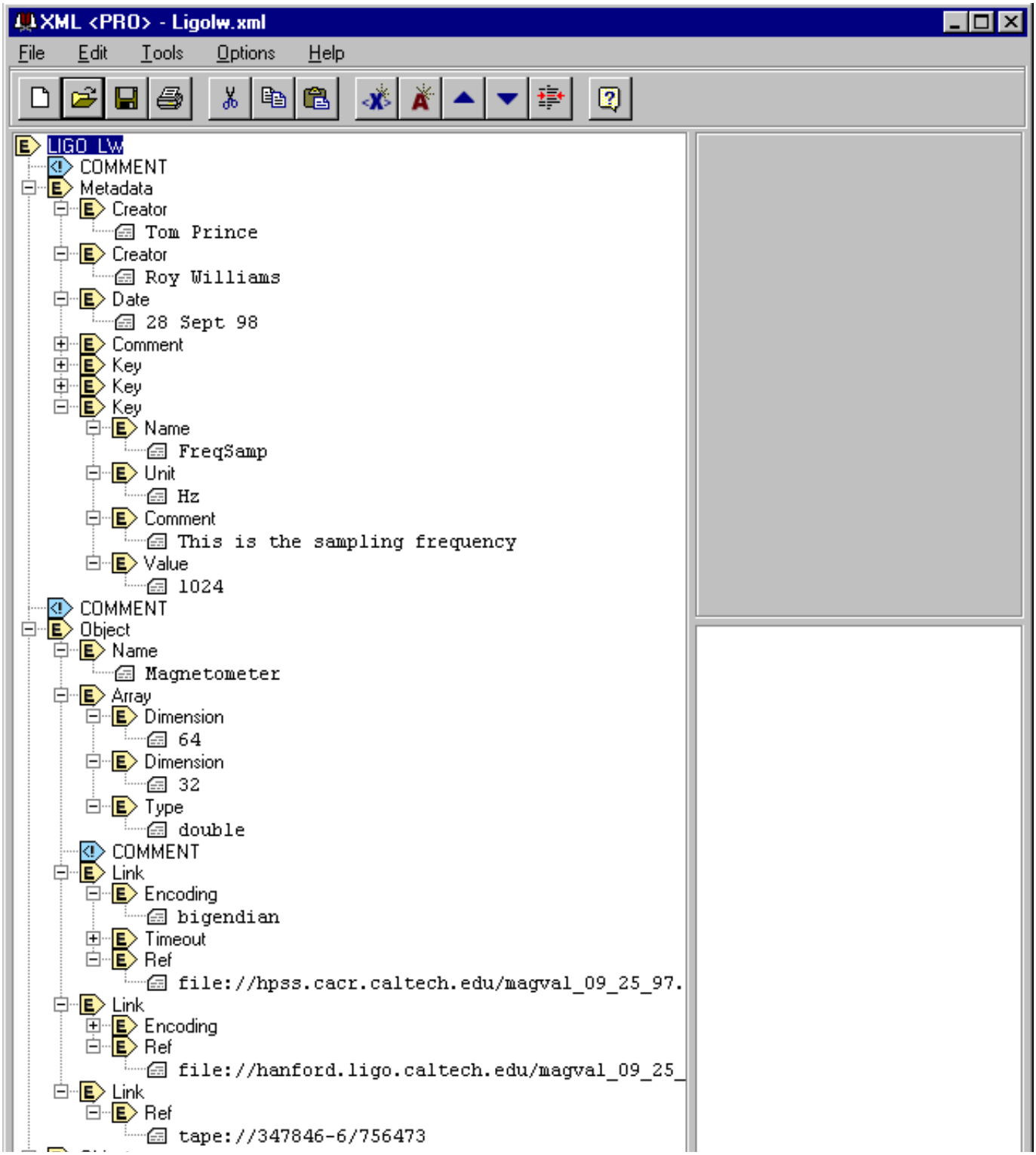
- LigoLW is based on XML to anticipate web-distribution, network distributed processing
  - Metadata: tags, keywords, elements, attributes
  - Data: encoded binary; ASCII; raw binary(?); other objects;...
- Lightweight format complements frames:
  - interprocess data communications (@ socket level)
  - easily readable/parsable format for end users
    - quick-look products, single channels
    - spectra
    - plots
    - events
    - metadata
  - estimated data volume: ~< 600 GB/yr reduced data;
    ~ 135 MB/yr metadata
- Reduced, processed, or otherwise non-frame data will be recorded in LIGO-standardized lightweight data format (LigoLW)
  - Metadata (data about data: frame catalog indices, operator logs, textual data, etc.)
  - Event data [event specification still TBD]
  - Spectra, time series snapshots, intermediate analyses performed with commercial/public-domain tools (MATLAB, Mathematica, ROOT, Triana, ...)

LIGO Data Types/Products

Lightweight Data Format -- XML

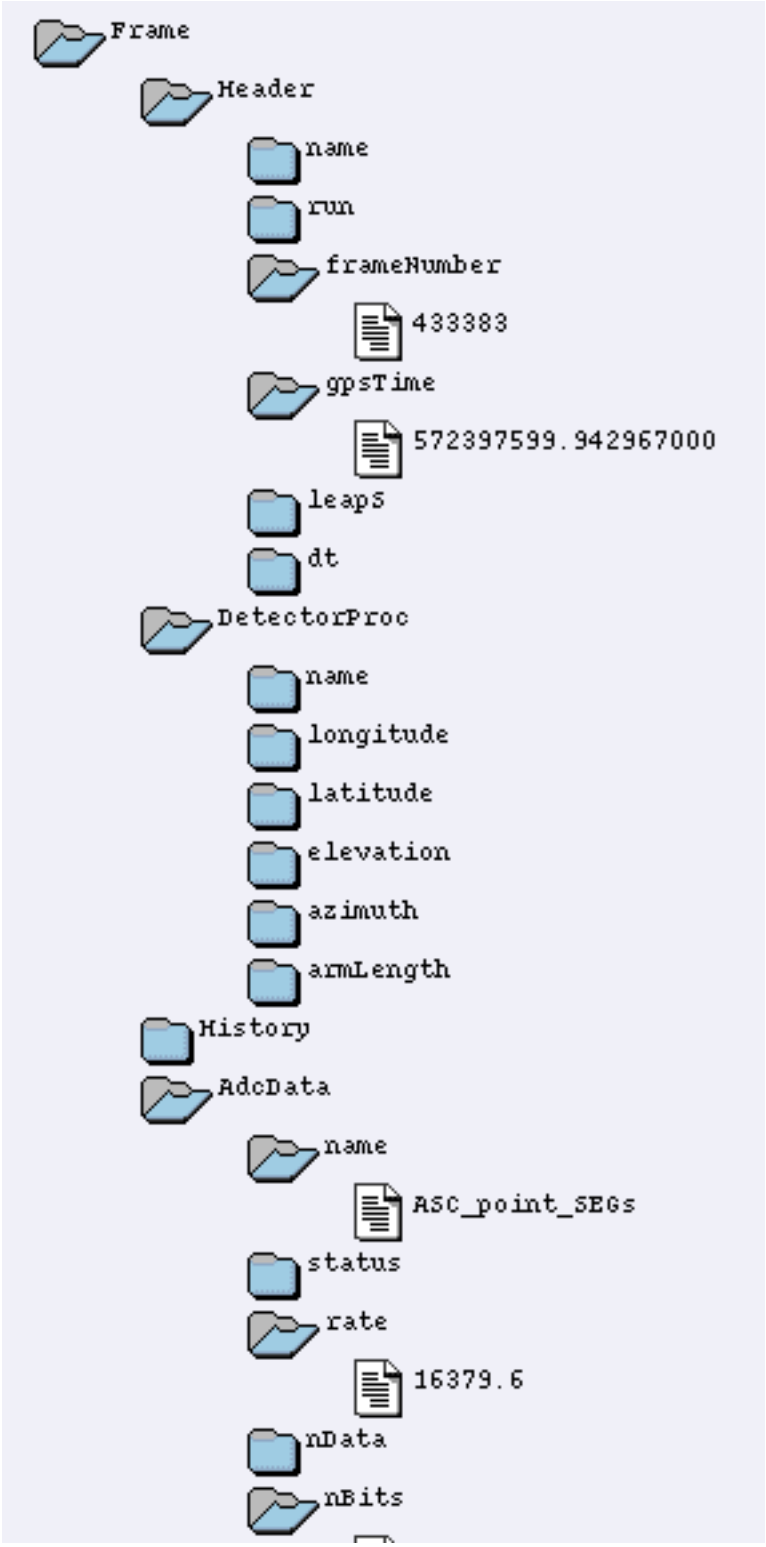
- Status:
  - ›› Specification released - draft DTD
    - – LIGO-defined data objects with defaults enable simple utilization *[http://www.cacr.caltech.edu/ligo/ligolw]*
      - – tables (ntuplets: points in a hyperspace)
      - – arrays (indexed elements of data)
        - • matrix
        - • vector, time-series, power-spectrum, ...
  - ›› First implementations
    - – Directed pulsar search results from 40m dataset - summer student project
    - – Socket-to-socket and Tcl-C++ LDAS interprocess communications prototyped
  - ›› Parser built to extract metadata from frames and to create LigoLW metadata catalog
  - ›› Revise specification over next quarter as experience indicates
    - ✓ Inputs from detector team (diagnostics) received, being incorporated

LigoLW
Parsed LW data object



Frame - LigoLW Conversion

Parsed frame object [XML <-> Fcl]



LigoLW
Example -- Metadata

```
<?xml version="1.0"?>
<!DOCTYPE LIGO_LW SYSTEM "Ligolw.dtd">
<LIGO_LW>
<!-- First the Metadata ------------------------------ -->
 <Metadata>
  <Creator>Tom Prince</Creator>
  <Creator>Roy Williams</Creator>
  <Date>28 Sept 98</Date>
  <Comment>LIGO power spectrum of 32 magnetometers at 64 frequencies</Comment>
  <Key>
   <Name>LIGOType</Name>
   <Comment>The Ligo data type is defined here...</Comment>
   <Value>Power Spectrum</Value>
  </Key>
  <Key>
   <Name>StartDate</Name>
   <Comment>Can't remember exactly but this date is close!</Comment>
   <Value>03/21/97</Value>
  </Key>
  <Key>
   <Name>FreqSamp</Name>
   <Unit>Hz</Unit>
   <Comment>This is the sampling frequency</Comment>
   <Value>1024</Value>
  </Key>
 </Metadata>
```

LigoLW
Example -- Data

```
<!-- Now for the Data objects ------------------------ -->
  <Object>
    <Name>Magnetometer</Name>
    <Array>
      <Dimension>64</Dimension>
      <Dimension>32</Dimension>
      <Type>double</Type>
    </Array>
<!-- This Array is at Cacr, Hanford, and on a tape -->
    <Link>
      <Encoding>bigendian</Encoding>
      <Timeout>600</Timeout>
      <Ref>file://hpss.cacr.caltech.edu/magval_09_25_97.bin</Ref>
    </Link>
    <Link>
      <Encoding>base64</Encoding>
      <Ref>file://hanford.ligo.caltech.edu/magval_09_25_97.bin</Ref>
    </Link>
    <Link>
      <Ref>tape://347846-6/756473</Ref>
    </Link>
  </Object>
   <Object>
    <Name>Magscale</Name>
    <Array><Dimension>32</Dimension></Array>
<!-- Embedded data -->
    <Data>
      1.28374 1.23453 1.94847 2.148474 2.39484 2.84746 3.10928 4.92827
      5.28374 5.23453 5.94847 6.148474 6.39484 6.84746 7.10928 8.92827
      9.28374 9.23453 9.94847 10.18474 10.3984 10.8446 11.1928 12.9827
      13.2874 13.2453 13.9847 14.18474 14.3984 14.8446 15.1928 16.9827
    </Data>
  </Object>
  <Object>
    <Name>Magoffset</Name>
    <Comment>This is the magnetic offset</Comment>
    <Array><Dimension>32</Dimension></Array>
<!-- Data follows from the end of the previous Object in the same stream -->
    <Follows/>
  </Object>
</LIGO_LW>
```

Database Management Systems
DBMS

- LIGO has four data types that need to be managed:
  - raw, framed data -- HPSS or equivalent network file system
  - lightweight data -- HPSS or database management system (DBMS)
  - events (as they are generated, cataloged) -- DBMS
  - metadata -- DBMS
    - catalogs & indices
    - operator logs
    - trends and high-level descriptions of detector performance

- Still in process of deciding DBMS for LIGO
  - Held workshop 22,23 October with consultants from CERN, SDSC, CACR, Astronomy(IPAC/CIT) to review LIGO needs
  - Choices being considered:
    - relational [deemed sufficient for LIGO needs]
      - ORACLE (CIT license for campus MIS)
      - DB2 (CACR HESC license for Caltech and related site activities -- platform specific)
    - object-oriented DBMS [deemed unneeded for LIGO needs]
      - Objectivity
  - Issues: Buy-in costs; operational costs; upgrades if we start too low; metadata only vs (metadata+data); ...
  - Selection by next design review

Database Management Systems
DBMS prototyping activities

- Ongoing BT Bakeout activity is generating 4 disparate DBs
  - temperature, current, pressure data along BT
    - Microsoft ACCESS DB
  - residual gas data from RGA
    - Proprietary SW from RGA vendor: spreadsheet compatible
    - partial pressure vs time scans
    - mass spectra
    - calibrations
  - weather station/environmental data: proprietary SW with station vendor: spreadsheet compatible
  - operator logs
    - text (ASCII) files
- Data arrive weekly as (~80MB files)
  - 700 channels x 10000 rows
  - data are ingested (transformed) and metadata produced for indexing into archive -- 3 hour ingestion process on NT server @ CACR
- Need to make data available at future dates for intercomparisons as bakeout progresses
  - DB will eventually grow to ~ 2GB, indexed by timestamp
  - Metadata+data co-located in DB

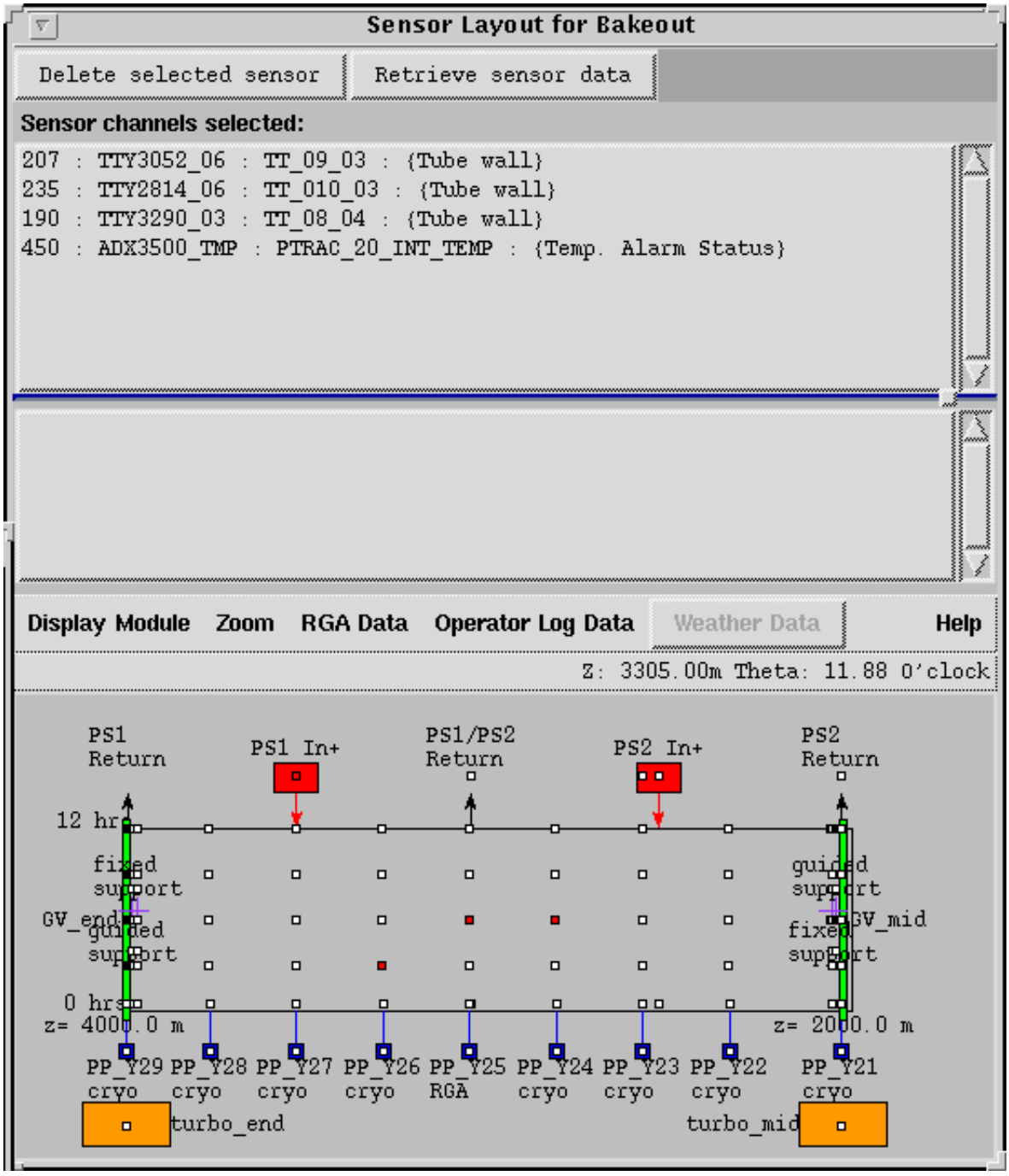
Database Management Systems

DBMS prototyping activities

- Developed a GUI using the LDAS model to allow web-based (via browser-plus-plugins) access to DBs
- Data server independent of GUI
- Communications through query standard protocol issued by browser
- Presently temperature data are available
- To do:
  - Integrate remaining datasets -- RGA data, weather data, operator logs
  - Return LigoLW objects [XML]

Database distribution tools

BT Bakeout Data Distribution Prototype



Data Analysis System for LIGO I
Software Design

LDAS SOFTWARE DESIGN FEATURES -- LAYERED DESIGN

Languages:

- ANSI C++
- ANSI C for wrappers to C, FORTRAN and TCL
- TCL (Tool Control Language) for control of resources.processes
- TK for Graphical User Interfaces
- Tclets (TCL/TK plug-ins) for web browser connectivity
- *TBD* database for data/metadata

Communications:

- TCL layer sockets to communicate commands and messages between processes
- C++ socket class library to communicate data between processes
- MPI (Message Passing Interface) for numerically intense parallel [scientific] computing.

Libraries:

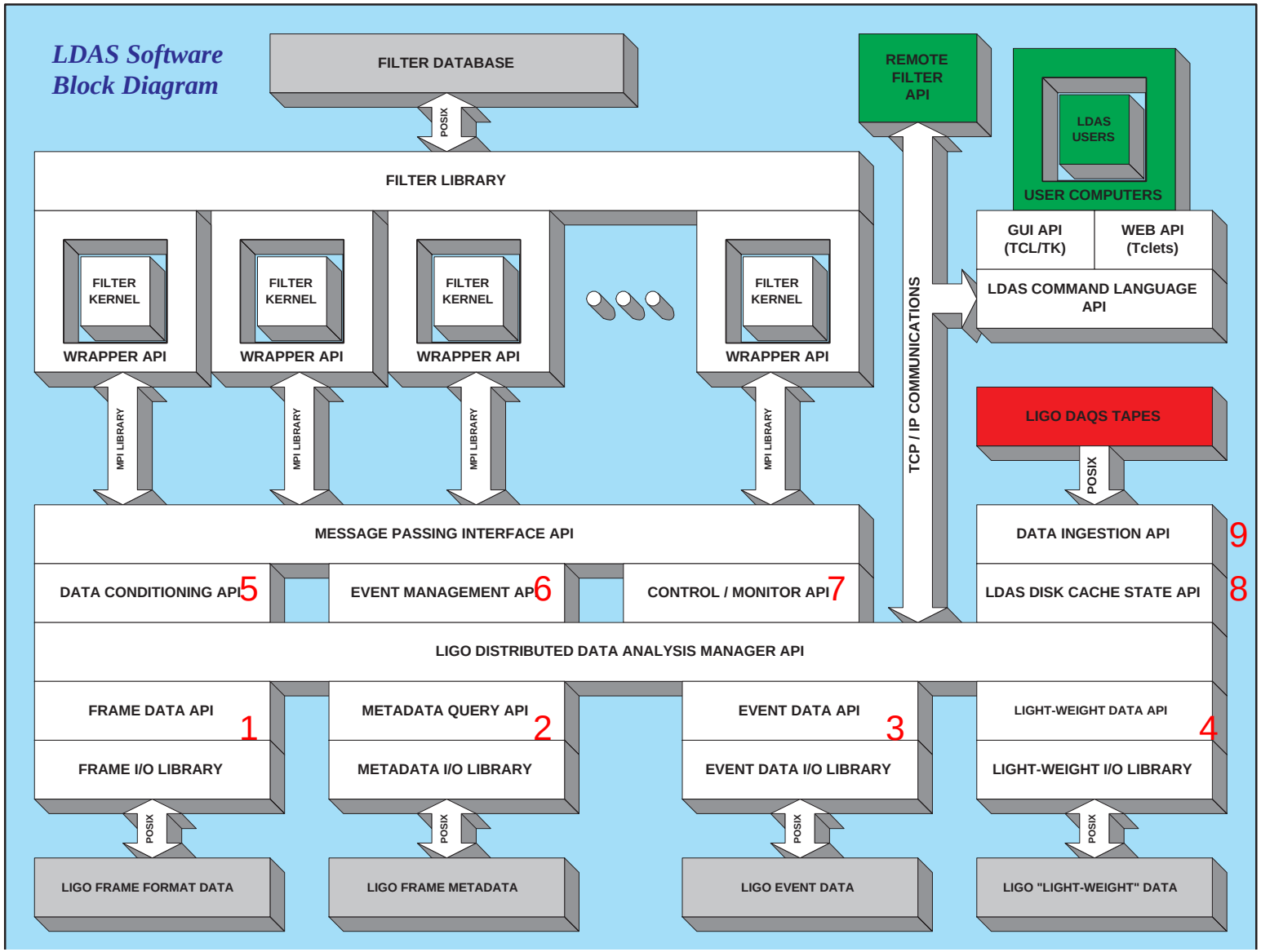
- Shared C++ Class Libraries, numerical libraries and I/O libraries on supporting platforms for efficient use of hardware resources

Data Analysis System for LIGO I
Software Design

›› Application Programmer Interfaces (APIs)

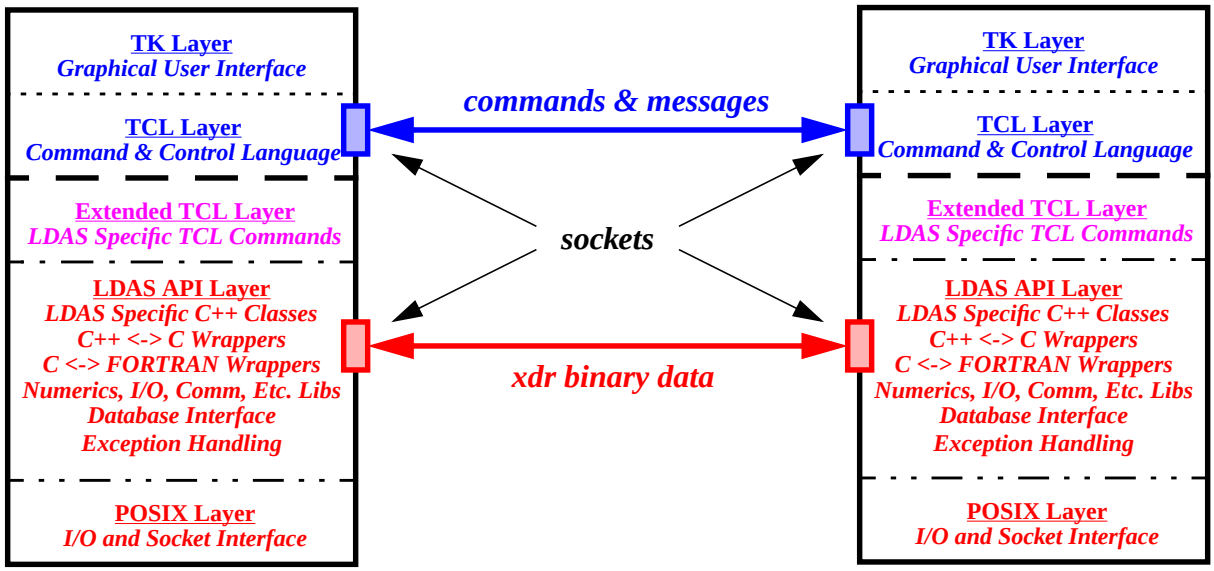
1. Frame Data
   - Manipulates framed data; I/O; channel extraction/ insertion; concatenation; ....
2. Metadata API
   - Interacts with the DB environment; data entry/extraction; data searches/sorts/queries.
3. Event Data API
   - Updates event lists; classifies events; searches on events;
4. LigoLW Data
   - Frame->LigoLW translator; data object extraction/ insertion.
5. Data Conditioning API
   - Data pre-processing; calibration; filtering; regression; computation either done using filter kernels or within this API (depends on complexity);
6. Event Management API
   - Receives output from the MPI based filter kernels; reports events; displays; ...
7. Control & Monitoring API
   - LDAS configuration, monitoring, exception handling, resource allocation; user interaction;
8. Disk Cache API
   - Stages data from archive/large disk farm to intermediate cache for efficient retrieval; queues data requests.
9. Data Ingestion API
   - Incorporates new data into archive; filter; reduce; compress.

LIGO Data Analysis System
Software Design



LIGO Data Analysis System
Software Design

APIs "*TWO-LEVEL*" SOCKET COMMUNICATIONS



Yellow boxes below indicate option to use SCSI

API	**FW**	**MPI**	**DC**	**EM**	**CM**	**FD**	**MD**	**ED**	**SDF**	**FK**	**COM**	**DI**	**DCS**	**MAN**	**RF**
FW		mpi								inherit					
MPI	mpi	mpi	socket	socket	socket					inherit					
DC		socket			socket	socket	socket		socket					socket	socket
EM		socket		socket	socket			socket	socket					socket	socket
CM		socket	socket	socket		socket	socket	socket	socket					socket	socket
FD			socket		socket				socket				socket	socket	
MD			socket		socket		socket	socket	socket				socket	socket	
ED				socket	socket		socket	socket	socket					socket	
SDF			socket	socket	socket	socket	socket	socket					socket	socket	socket
FK	inherit	inherit													
COM														socket	
DI													socket		
DCS						socket	socket		socket			socket		socket	
MAN			socket	socket	socket	socket	socket	socket	socket		socket		socket		socket
RF			socket	socket	socket				socket					socket	

Data Analysis System for LIGO I
Software Development - Status

- GenericAPI:
  - Basis for all other APIs. Initial investment in prototyping and design will allow rapid diversification into specific APIs by extension of the generic class.
  - Complete

The first APIs to be developed will support Detector Installation milestones for the first interferometer

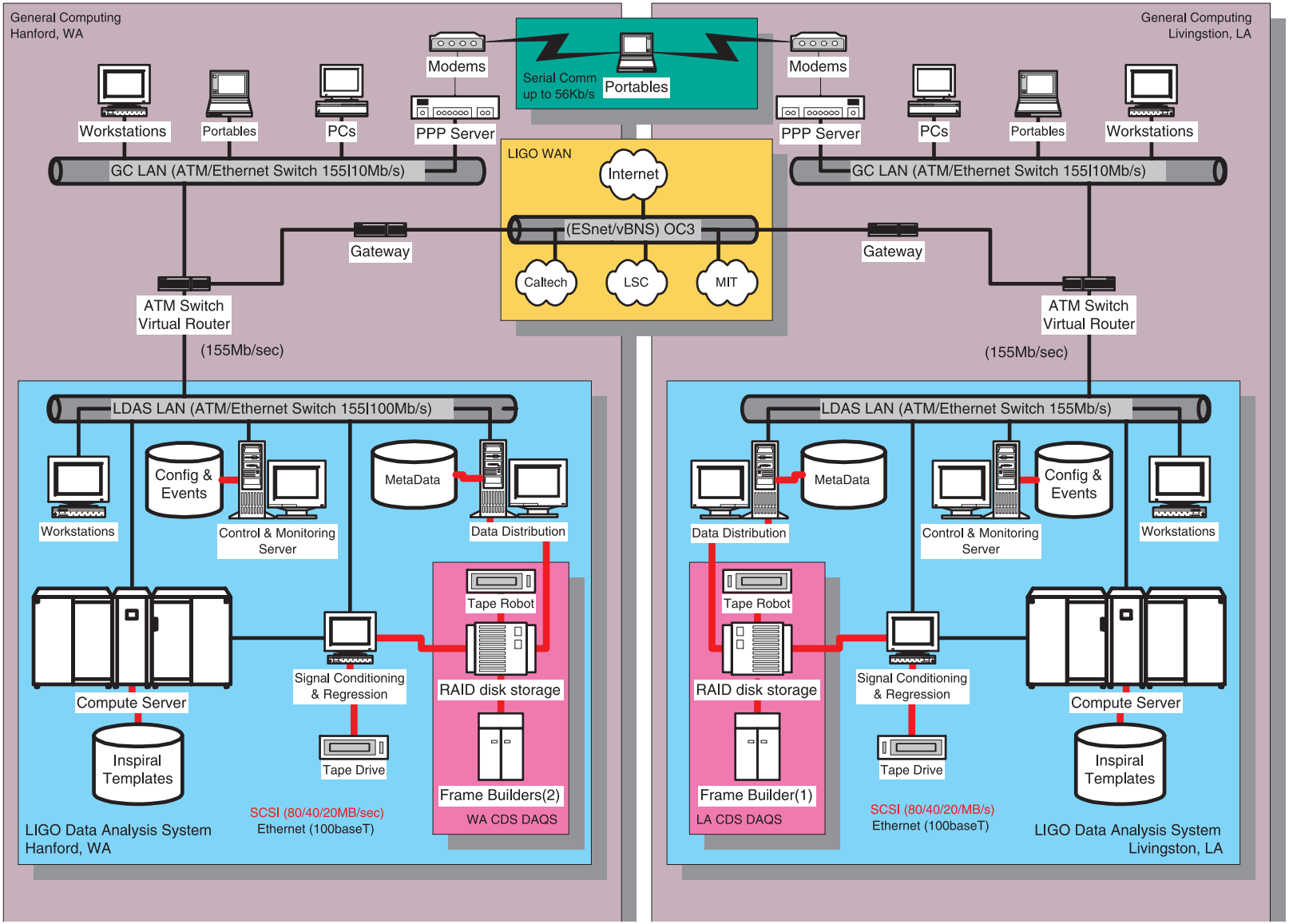
- Under development:
  - Fcl I/O Library: complete
  - Fcl Specification: complete
  - FrameAPI
  - ManagerAPI
  - DataConditioningAPI
  - UserAPI

- Reviews:
  - Develop prototype UserAPI by end of Dec.
  - Hold Preliminary Design Review in January to support first deliveries to Hanford.

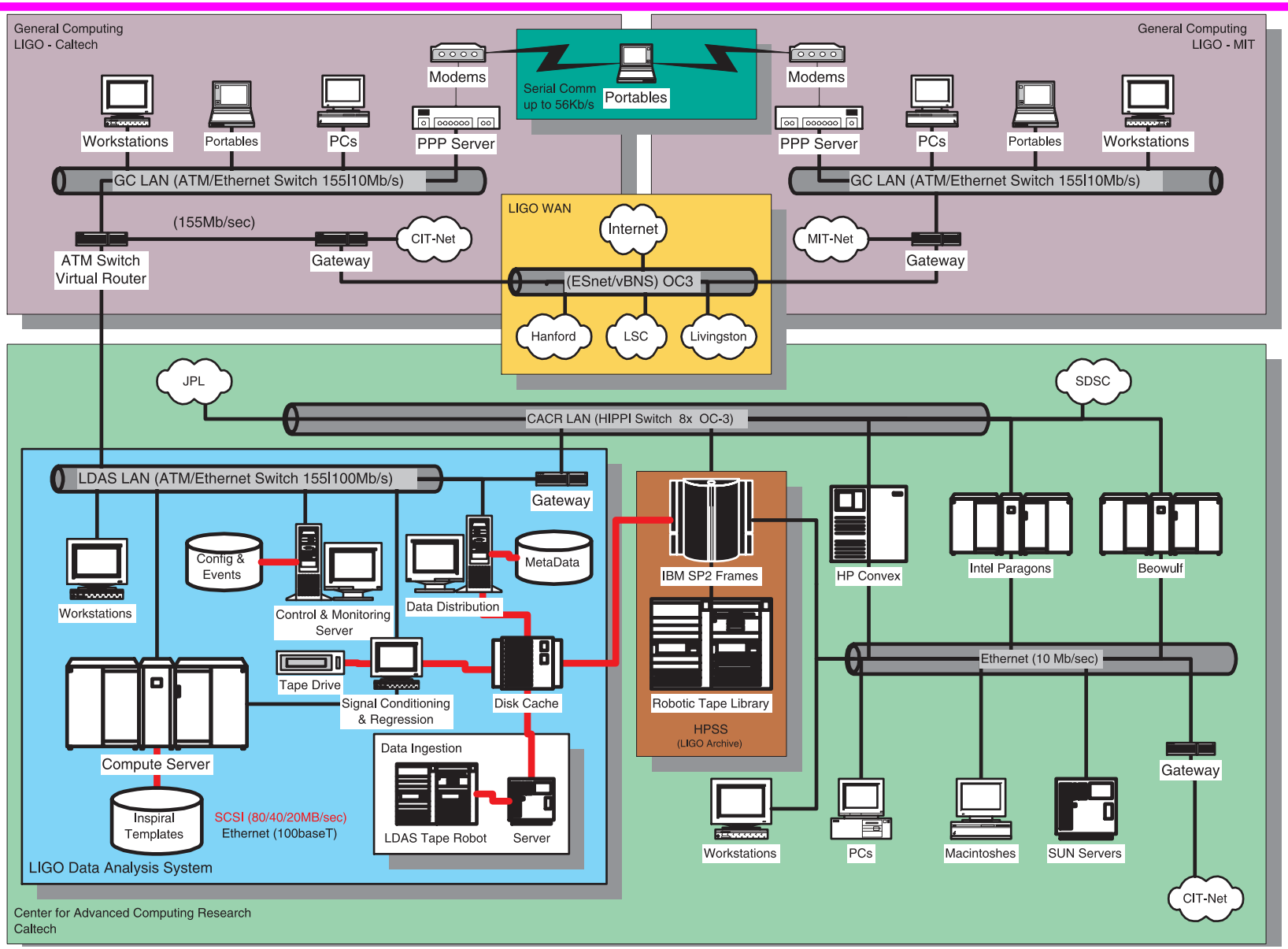
Hardware Status

- Implementing prototype data server for Hanford at LIGO Caltech
  - LAN implementation
  - Transmission of data from 40m prototype
  - Data server
  - Data conditioning machine
  - Software installed on corresponding hardware platform per on-line LDAS design.
  - Linux cluster installation at LIGO
- Data analysis:
  - Beowulf and MPI has been demonstrated on LIGO-scalable data flows for inspiral detection
    - Joint effort with CACR (Paragon) & Univ. Wisc. (PC/ linux)
    - 8 node/16CPU integrated cluster in house at LIGO - being set up.
  - Directed pulsar search prototype code has been implemented using 40m data and CACR machines.
- Data archival technology choice will be deferred as late as possible (2001) => working through CACR
  - Optical tape technology replacement for magnetic media in HPSS (LOTS)
  - 1TB/cassette (same form factor as present IBM robot cassettes)
  - ~$250/cassette ($0.25/GB)
  - Optical heads replace magnetic tape heads in same cabinetry.

LIGO Data Analysis System
On-line architecture

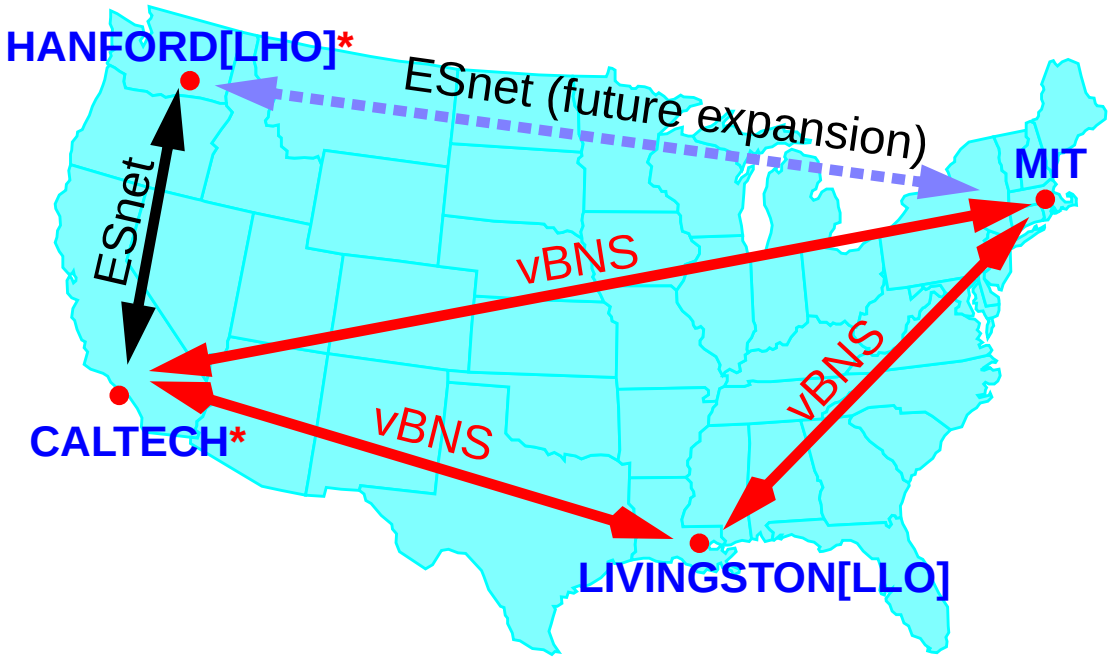


LIGO Data Analysis System
Off-line Architecture



LIGO Wide Area Network

WAN Topology



***LHO will be accessible to LSC member institutions via Caltech**

WAN/LAN Connectivity among LIGO Laboratory Sites

Site	*Livingston, LA*	*Hanford, WA*	*MIT*	*Caltech*
Caltech	vBNS(OC3)	ESnet (4 X T1) <-> vBNS(OC3)	vBNS(OC3)	OC3/ATM 100BT
MIT	vBNS(OC3)	MIT<->Caltech<->Hanford ESnet (4 X T1) <-> vBNS(OC3)	100BT OC3/ATM(?)	
Hanford, WA	ESnet (4 X T1) <-> vBNS(OC3)	OC3/ATM 100BT		
Livingston, LA	OC3/ATM 100BT			

LIGO Wide Area Network

Status

- LIGO proposed & drafted an MOU between NSF/DOE to provide access to ESnet at Hanford
- Authorized in October 1998
- Implementing initial (T1) capability; requested up to 4 x T1 BW (cost is an issue).
  - Routing: LHO-PNNL-SDSC-CACR-LIGO/Caltech
  - In Progress:
    - DNS in process of being turned over to Caltech ligo-wa.caltech.edu
    - T1 connectivity tested and working
    - Move workstations over to new IP addresses
    - Setup E-mail and Web services
  - Planned:
    - MOU covers 4 T1 connections -- may take advantage of contingency.
    - WSU/Pullman (~ 100km NE) awarded an NSF grant to establish a vBNS hook-up
    - UW/Seattle (~350 km W) has vBNS at present
    - PNNL is investigating future high speed connections via Seattle -- LIGO will participate if costs are acceptable.
- MIT may be added later as a separate addendum to MOU

- LHO will be accessible to LSC member institutions via Caltech (ESnet gateway)

LIGO Wide Area Network

Status

- T1 link to Livingston Observatory is in place
  - LSU awarded vBNS access in latest round of NSF awards -- includes LIGO access at Livingston
  - LSU provides gateway service
    - Caltech providing DNS services ligo-la.caltech.edu
    - E-mail and Web services in process of being setup (last week)
  - Planned:
    - Finalize hardware logistics with LSC
    - Install main server
    - Establish modem services and contingency plan
    - Establish OC3 Connectivity in the next 1-2 years depending on fiber availability (present connection is Cu)
      - LIGO will have to install FO lines from Livingston to the Observatory
      - Upgrade the routing equipment to accommodate new connectivity
- LLO will be directly accessible by LSC member institutions via internet (vBNS)

LDAS Development
Timeline

- Highest priority: staged implementation of on-line systems to support detector testing:

Detector Milestone:	Date	LDAS Need
PSL/Input Optics	4/99	Min. data dist.
Vertex Michelson, “first light”	9/99	Full data dist.
2km operational	8/00	On-line system

- Staged installation of off-line system at Caltech in period 9/99 - 12/01

Simulation environment -- End to end modeling

End-to-End (E2E) Simulation

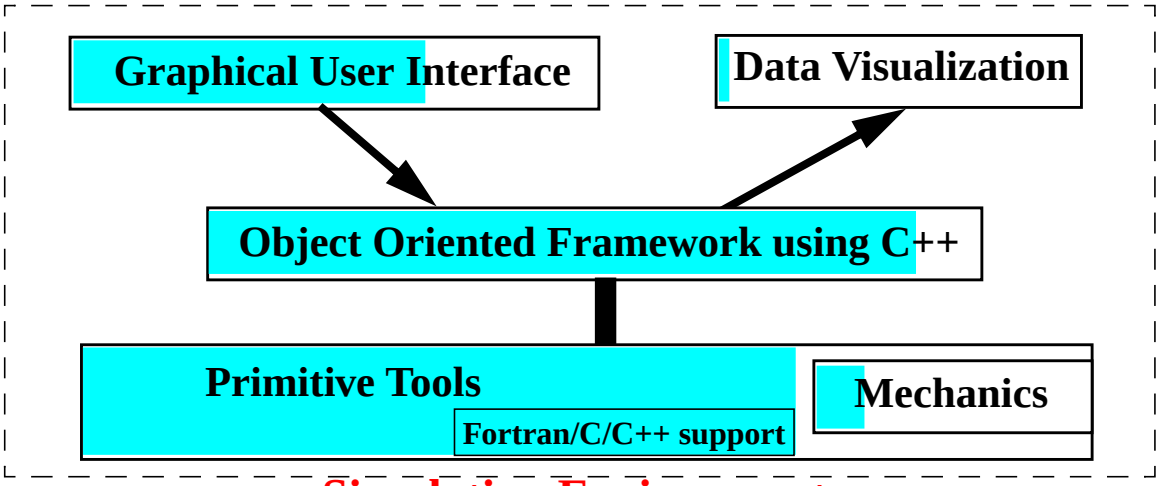
- Time domain simulation of LIGO interferometer output(s)
- Object Oriented structure using C++
  - Modular and expandable
  - Support for plug-ins using FORTRAN/C/C++
- No low level language (i.e., C++) needed to use
  - Easy to use high level language
  - GUI
- “Toolbox” Primitives
  - mirror - reflection, transmission, tilt, ...
  - field propagator - time delay, Guoy phase, ...
  - modulator and demodulator - arbitrary number of sidebands, ...
  - digital filter - models servos, electronics, linearized response, ...
  - mechanical components
    - test mass, beam, clamp...

Support to Detector Installation
E2E simulation

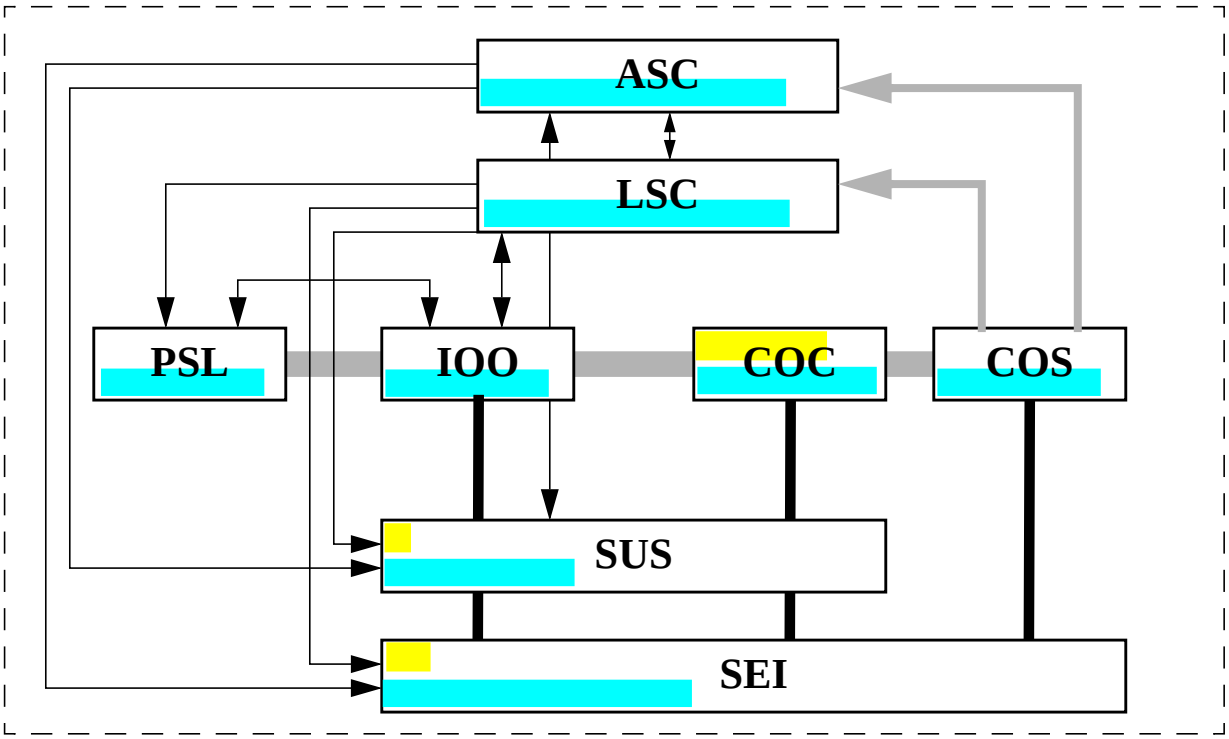
- Plan
  - When the vertex Michelson at Hanford is available, E2E will provide the minimum set of subsystems so that semi-quantitative comparison of performance can be made.
- Construct the simulation models to map into the real hardware
  - Implement phenomenological models for those parts which cannot be simulated using primitives.
  - PSL - 4/99
  - IOO - 7/99 [with UFl]
  - SUS/SEI - 7/99
- Collaborative participation [e.g., U. Fl.] to develop LIGO physics modules using available toolbox primitives
- Simulation team will participate in shakedown of hardware alongside detector subsystem teams

Status Overview

E2E simulation



Simulation Environment

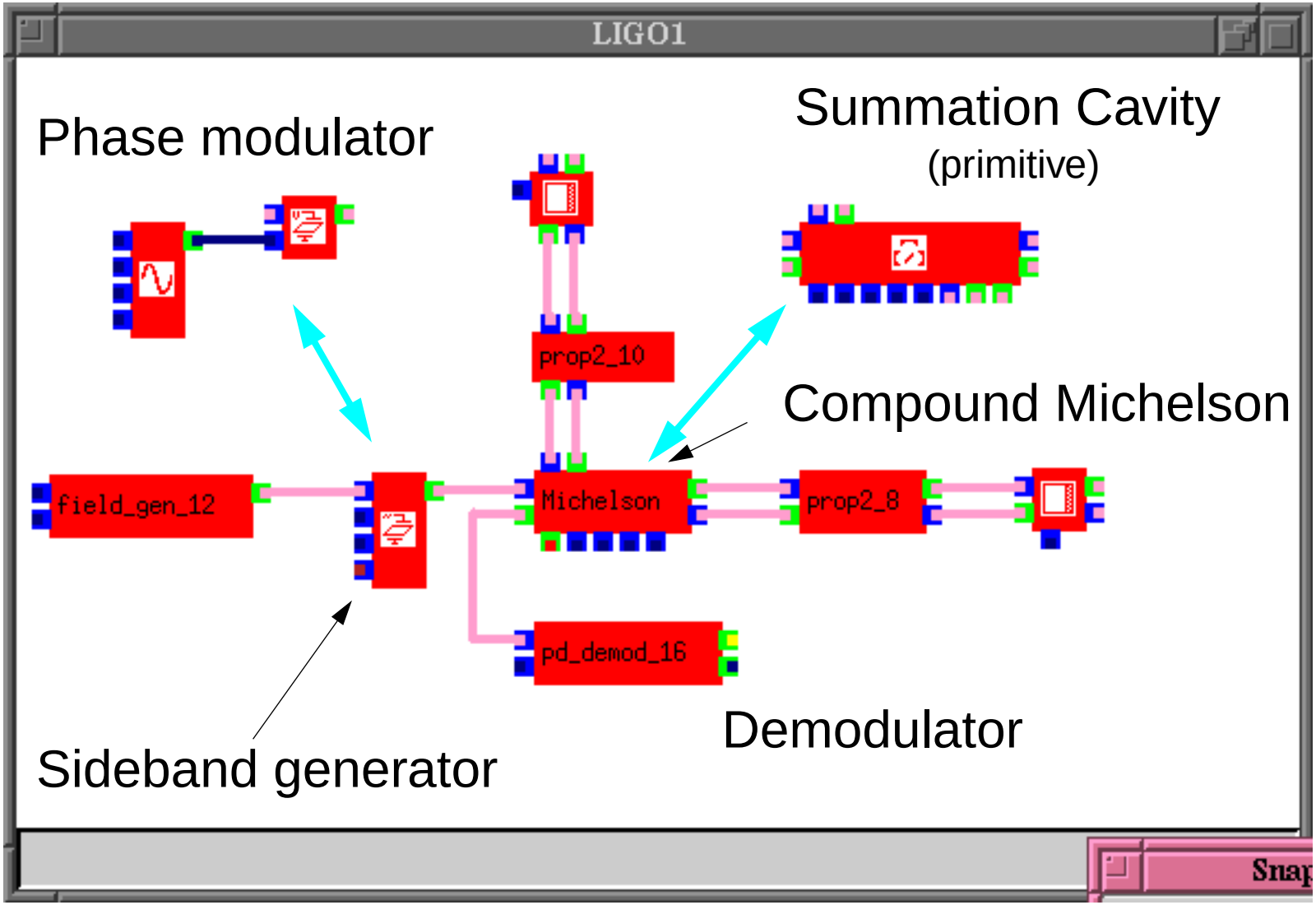


Subsystem Implementation

Primitive tools completed

Explicit construction completed

Example
E2E simulation



Status
E2E simulation

- Single mode time domain model
  - Improved capabilities
  - Validation alost completed
- Modal model in time domain
  - Field expanded by finite Hermite-Gaussian modes
    - Mirror tilt and displacement
    - Mode mismatching
    - Thermal lensing
  - Implementation in progress
    - validation for FP case done
- Primitives (toolbox components) complete except for mechanical subsystems

Status
E2E simulation

- Mechanics module development
  - Any simple linearized model can be built using Digital Filter
  - A more detailed simulation needs physical model implementation
  - S. Mohanty - Penn. State Univ. (visitor 1997/1998)
    - Formulation of dynamics of a mechanical structure
    - Self-consistent inclusion of thermal noise sources
    - Explicit formula for a single pendulum derived
  - G. Cella - Pisa Univ. (formerly with VIRGO)
    - Author of simulation program of mechanics model for VIRGO
    - C++ based, modular and expandable
    - Similar syntax as e2e -- easily adapted to e2e environment
- E2E incorporation of mechanics models
    - Integrate framework of Cella into e2e framework
    - Include dynamics and thermal noise formulation developed by Mohanty as appropriate
    - Use the same GUI as LIGO e2e
  - Short term implementation strategy
    - Implement single pendulum model by Mohanty
    - Validate dynamics of model
    - Use for simulating a simple SEI model
    - Validate modular model of Cella

International collaborative activities in data analysis and simulation

Collaboration in Data Analysis and Simulation

- VIRGO
  - VIRGO set up web site for frame library in C [FrameL]
  - LIGO has completed a C++ object class library implementation [Fcl]: http://docuserv.ligo.caltech.edu/~wmajid/fcl/index.html
  - C++ implementation is continually tested by both VIRGO and LIGO; all changes agreed upon mutually.
  - VIRGO is developing a quick-look analysis environment based on ROOT (VEGA). VEGA uses Fcl to access interferometer data.
  - VEGA tested and installed at LIGO -- candidate for on-line diagnostics tool.
  - Design of the numerical filters library development has begun in C++ as a joint effort; algorithm class library [Acl]
    - Prototype implementation has been done by VIRGO.
    - Work starting on identifying all methods LIGO and VIRGO want to employee.
  - G. Cella, Univ. Pisa (formerly with VIRGO) has joined in developing simulation components for mechanical seismic isolation assemblies.
  - Weekly teleconferences with VIRGO/Annecy

Collaboration in Data Analysis and Simulation

- GEO
  - Active participation in LSC groups recently formed to tackle data analysis issues:
    - ASIS [sources and signatures]
    - DCSA [statistical techniques; confidence assessment]
  - Dr. Ian Taylor installed GEO's quick-look package TRIANA at LIGO during a previous visit; evaluated at 40m.
  - Dr. Ian taylor presently visiting for 6 months to study and understand LIGO's VME-based data acquisition system. GEO is planning to adopt the design for their acquisition system. 40m prototype scale is adequate for GEO.] as-is.
  - LIGO contributed its frequency-domain small-signal interferometer model (Twiddle) to a GEO-sponsored working group within LSC to develop advanced interferometer simulation tools. Dr. Hiro Yamamoto attended workshop at Garching in late spring 1998.

Collaboration in Data Analysis and Simulation

- TAMA
  - TAMA is using FrameL C library for its data acquisition system.
  - Dr. Hiro Yamamoto installed and demonstrated LIGO simulation testbed, E2E during recent visit.

LDAS Reduced Data and Metadata

SOURCE	Data	Data Types	Basis of size estimate: #Parameters #Bins #Pixels #Samples	#Bytes/Unit	#/Hr	LW Data Volume/Year [GB]	MetaData Volume/Year [MB]
LIGO - Interferometer	Machine state vector	String[XML]	2048	1	10	0.0	90
		Binary	128	1	10	0.01	0.0
	Operator Logs	Strings	20480	1	20	0.0	180
		Graphics[JPEG]	32768	1	10	2.9	89.8
	Diagnostics	Video	4096	1	60	2.2	538.6
		Spectra/Fast Scopes	2048	2	20	0.7	179.5
		Calibrations - Spectra	2048	4	10	0.7	89.8
		Calibrations - Coeffi-cients	4096	1	10	0.4	89.8
		Calibrations - Matrices	2048	4	10	0.7	89.8
		Triggers/Discrete Logic	128	2	60	0.1	538.6
	Frame Data Catalog	String[XML]	1024	1	3600	0.0	64630.0
LIGO - Environment [PEM]	Facilities state vector	String[XML]	512	1	10	0.0	134.6
	Seismometers	Spectra	1024	2	60	1.1	538.6
	Magnetometers	Spectra	1024	2	60	1.1	538.6
	Tiltmeters	Time Series@0.1 Hz Stored 1/Hr	16	1	360	0.1	9.0
	Acoustic Sensors	Spectra	8192	2	60	8.6	538.6
	Diagnostics - Calibrations	Matrices/coefficients	2048	1	0.41666667	0.01	3.7
	Diagnostics - Triggers	String[XML]: Model parameters	1024	1	0.41666667	0.004	7.5
		Discrete logic	128	2	60	0.1	538.6

LDAS Reduced Data and Metadata

SOURCE	Data	Data Types	Basis of size estimate #Parameters #Bins #Pixels #Samples	#Bytes/Unit	#/Hr	LW Data Volume/Year [GB]	MetaData Volume/Year [MB]
Non-LIGO	Seismic	String[XML]	512	1	10	0.0	89.8
	Electromagnetic storms	String[XML]	256	1	100	0.2	897.6
	Astrophysics - GRBs	String[XML]	256	1	0.04	0.0	0.4
	Astrophysics - neutrinos	String[XML]	256	1	0.00	0.0	0.0
	Astrophysics - visible	String[XML]	256	1	0.00011408	0.0	0.0
	Astrophysics - gravitational	String[XML]	2048	1	10	0.2	89.8
LDAS Events	Event Lists	String[XML]	2048	1	3600	64.6	32315.0
		Images/Graphics[GIF]	8192	2	3600	517.0	32315.0
				Total Database [GB]	== >	**600.8**	**134.5**