LIGO Data Analysis and Simulation

Overview and Status

Albert Lazzarini

NSF Fall Review
28 October 1998

LIGO Hanford Observatory
Hanford, WA

Outline

- Datastream description
  - Datastream characteristics
  - Data analysis challenges
- Data types & data products
  - Framed data
  - Lightweight format
  - Metadata
  - Events
- LIGO Data Analysis System (LDAS)
  - Requirements
  - Software
  - Databases
  - Hardware
  - WAN
- Simulation environment
  - End-to-End (E2E)

LIGO Datastream Characteristics

- LIGO datastream consists of continuous broadband signals
  - Audio frequency (16384 samples/s, 16 bit) digitization & acquisition of key channels *(lower sample rates for ancillary channels)*
  - LIGO detection band: 40 Hz < f < 3 kHz
- No directionality
  - Require signal processing to deduce source locations
    - modulation of CW sources due to Earth motion
    - Time delay between coincident responses along 3000km baseline

LIGO Datastream Characteristics (cont.)

- LIGO signals expected to be at limits of detectability
  - ›› Instantaneous SNR ~ 10^{-4}(strong chirp; pulsars) (weakest EM pulsars: SNR ~ 5 x 10^{-3}, require T_{int}~ hours to detect)
  - ›› Need to integrate over entire (most) of the waveform to generate detectable SNRs (~ 10).
  - ›› Requires coherent detection & signal processing
  - ›› False alarms: validation of an “event” requires ability to preclude all other (terrestrial) interpretations -- vetoes

[Ref. LIGO Sources charts, A1,2 at end of talk]

- LIGO acquisition data rates are high
  - ›› Many parallel channels of instrumentation to monitor instrument behavior, environment, anthropogenic disturbances, ...
  - ›› GW channel: 100kB/s for three interferometers (IFOs) -- 16384 samples/s @ 2bytes = 32kB/s per IFO
  - ›› Data acquisition is ~ 10 MB/s for 3 IFOs
    - 573 channels *per* IFO (only 1 is GW channel)
    - 610 channels on physical environment monitors (PEM) (at each site)
  - ›› Science (astrophysics) channels constitute as little as ~ [100 kB/s]/[10MB/s] ~ 1%

LIGO Data Analysis
Challenges

- Techniques are those for detecting the possible presence of weak signals embedded in noise (radar, sonar, pulsar searches, ...):
  - Continuous processing of interferometer output
  - Parallelization
  - Frequency-domain spectral analysis (spectral cross-correlation)
    - Optimal matched filtering; [1-3] x 10^4 physics-based templates; 90+% of CPU time spent of Fourier transformations.
    - Frequency-time analyses (spectrograms, Wigner-Ville distributions, etc.); pattern/ridge detection (2-D)
  - Wavelet analysis; novelty detection; phenomenology
  - Kalman filtering to remove instrumental signatures -- data conditioning

  [Ref. LIGO Data Flow chart, A3 at end of talk]

- Data archival & Distribution
  - volume reduction by 10X => reduction/veto algorithms
  - access to archived data => database engine/network tools
  - 100% processing => computational power
    - LIGO inspiral search requires ~300 kFLOP/Byte of data
    - This is higher than typical processing requirements
      - radar/sonar ~ 100 FLOP/Byte (many fewer templates, shorter durations!)
      - directed EM pulsar searches ~ 1 kFLOP/Byte
    - Compare: blind EM pulsar searches ~ 100 -1000 kFLOP/Byte

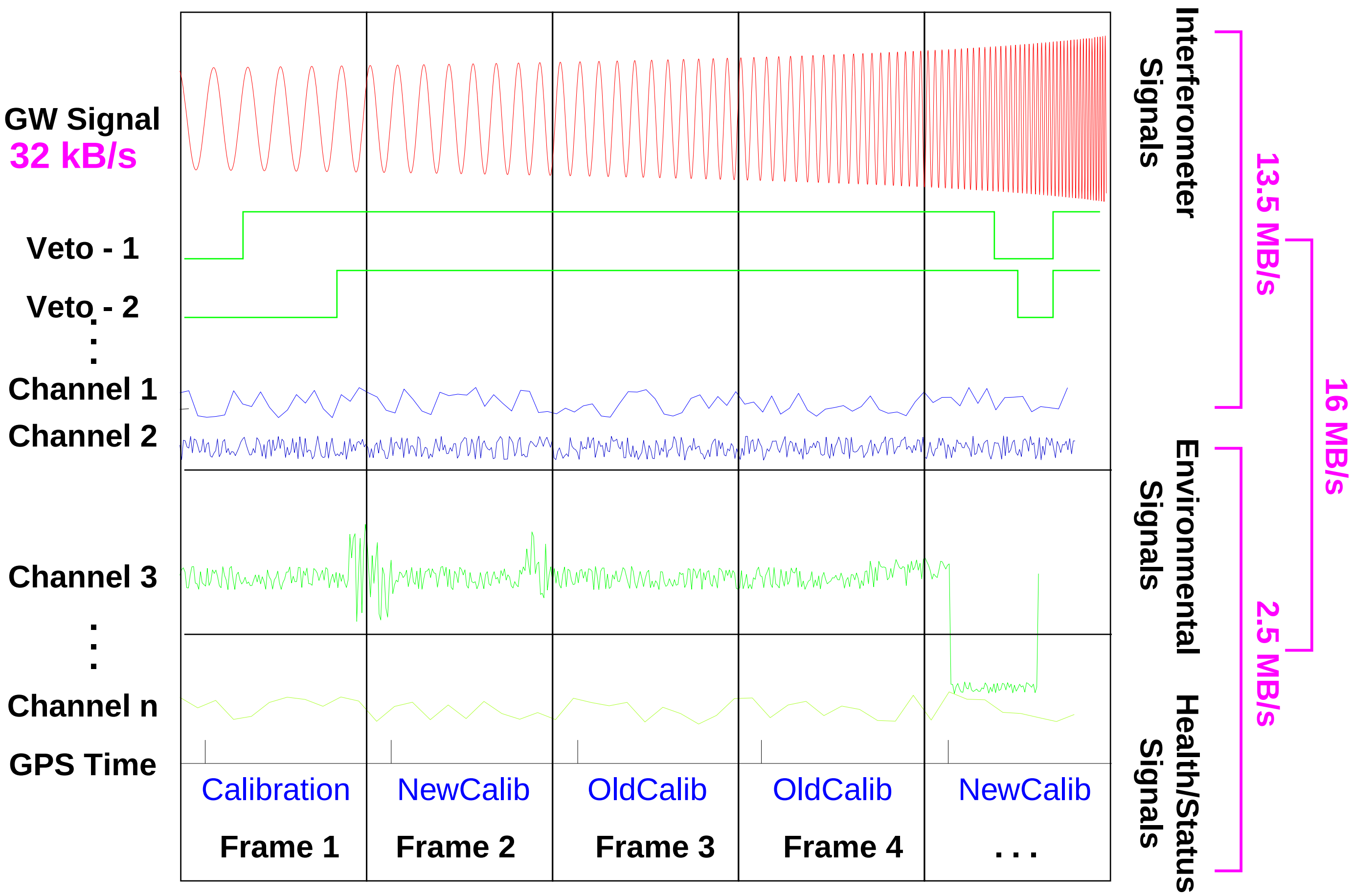
LIGO Data Products/Types

- 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 released 98.10.16
    - 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

LIGO Datastream

Frame Design

[Ref. LIGO Frame Format chart, B1 at end of talk]



- Frame is (structured) self-contained snapshot of data for a period of time
  - GW channel & ancillary IFO channels
  - Environmental monitoring (veto) channels
  - Facilities/Vacuum health & status
  - Hierarchical organization of data reflects IFO subsystems for more efficient veto utilization
- Full datastream could be ~ 300TB/yr
  - Plan to reduce (and compress) to ~ 50 TB/yr

Lightweight Data Format
XML

- Reduced, processed, or otherwise non-frame data will be recorded in a LIGO-standard 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 snippets, intermediate analyses performed with commercial/public-domain tools (MATLAB, Mathematica, ROOT, Triana, ...)
  - ›› 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; ...
- Need a lightweight format to complement 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

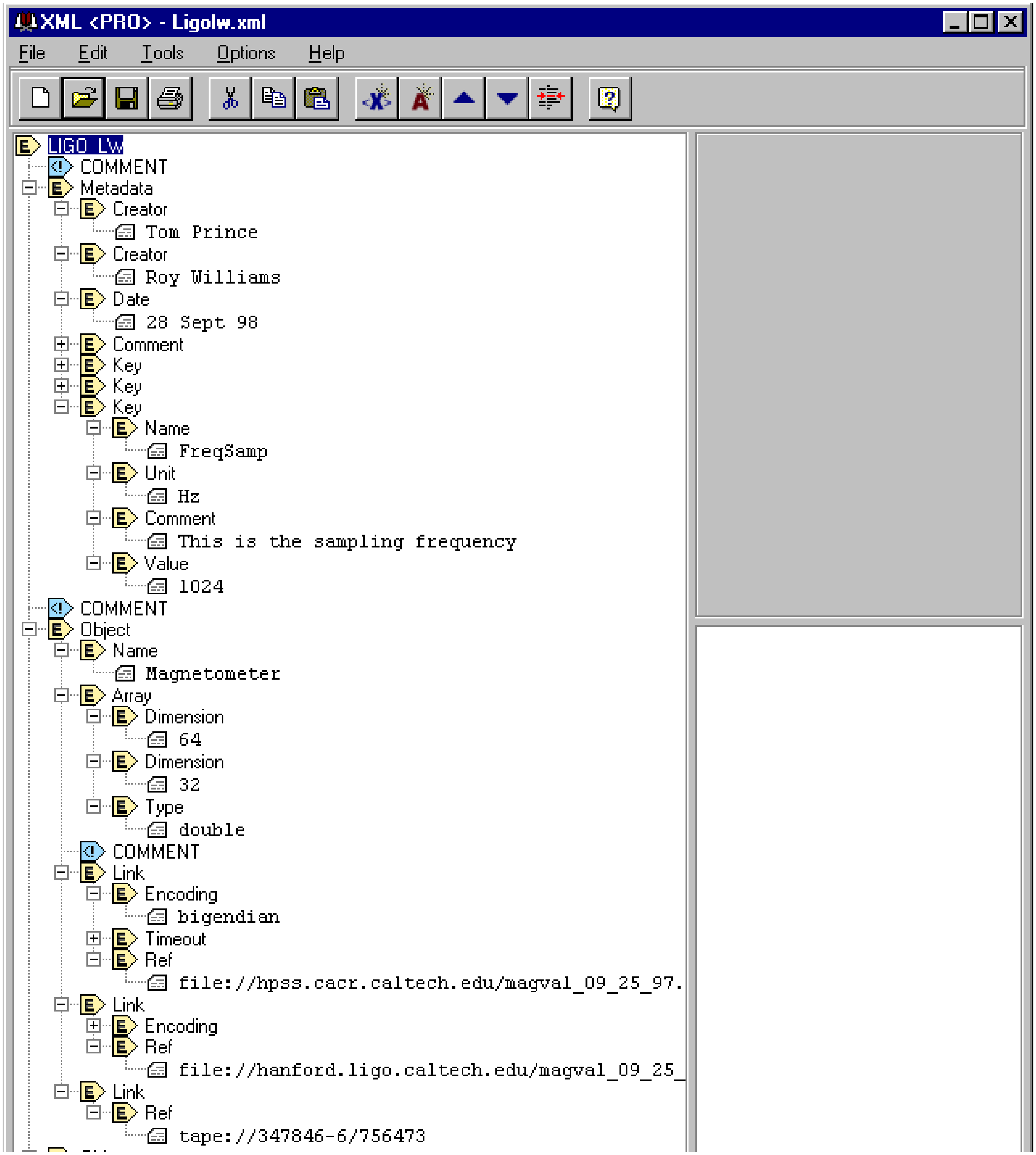
[Ref. LIGO Reduced & Meta Data Estimate charts, C1,2 at end of talk]

Lightweight Data Format
XML

- Status:
  - ›› Specification released
    - – 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 3 months as experience dictates

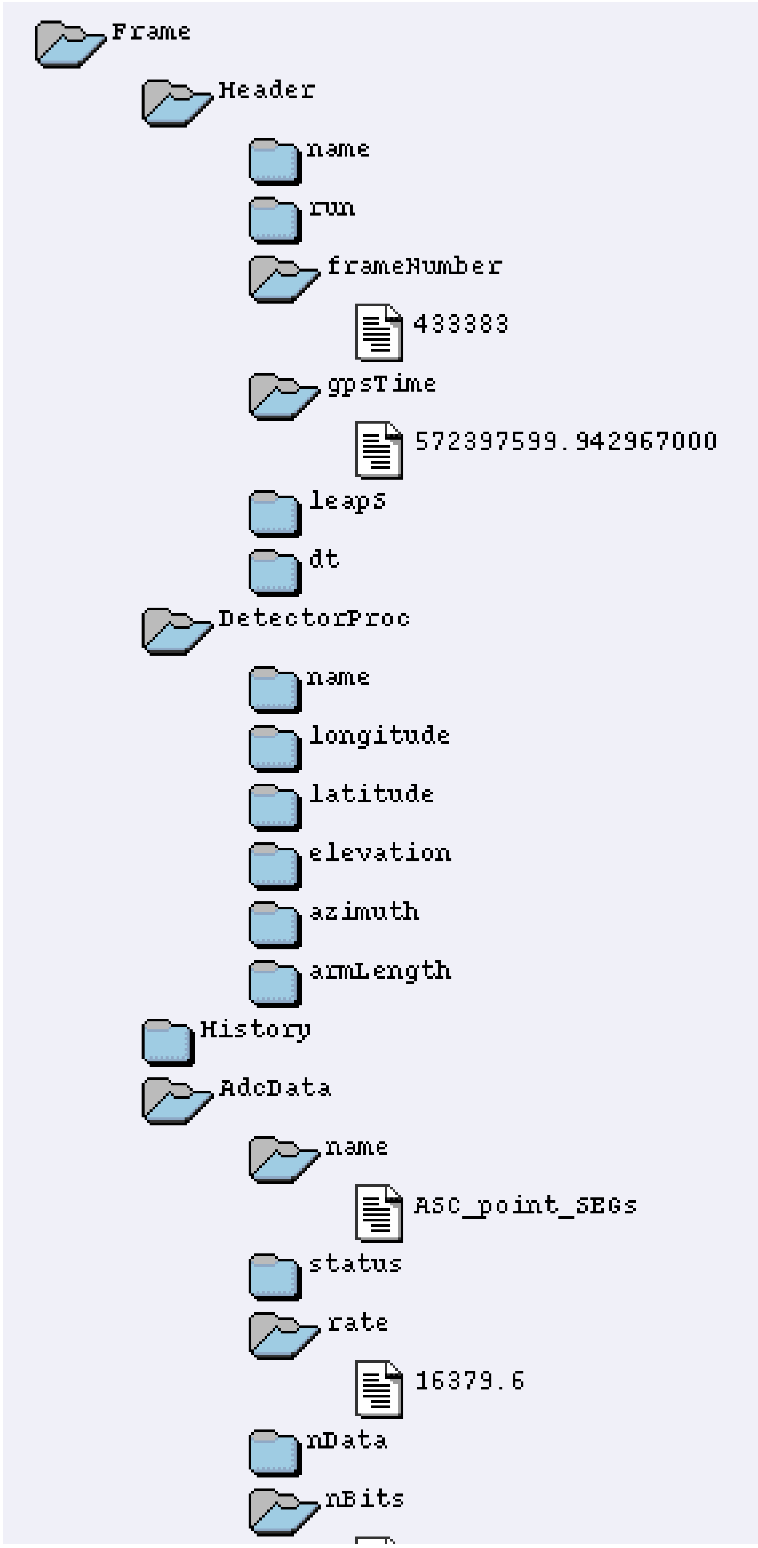
[Ref. LigoLW XML example, D1,2 at end of talk]

LigoLW
Parsed LW data object



Frame - LigoLW Conversion

Parsed frame object [XML <-> Fcl]



LIGO Data Analysis System

LDAS Primary Requirements

EXPLOIT THE GW CHANNEL TO THE MAXIMUM EXTENT POSSIBLE TO DETECT GRAVITATIONAL WAVES FROM ASTROPHYSICAL PROCESSES

- Provide on-line analysis capability at the observatories; data distribution from on-line data cache -- astrophysics, diagnostics.
- Process and reduce the raw LIGO datasets at the off-line center to prepare the data for archival storage and retrieval.
- Provide computational and storage resources for off-line analysis using the archived data
- Provide a flexible design which can be reconfigured to reflect new analysis or computational requirements as they evolve.
- Provide access to LIGO data from all LIGO Laboratory sites and also from member institutions of the LIGO Scientific Collaboration for the LIGO I search.

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 at CACR of off-line system in period 9/99 - 12/01

LDAS Summary

- Software design complete -- design requirements review held 12/97
- Software components specification for Application Programmer Interfaces (APIs) complete
- Work under way to develop Generic (template) API, from which specific APIs may 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 configuration definition complete
  - On-line systems at observatories
  - Off-line system at data repository (CACR/Caltech)
  - Wide area network for inter-site connectivity

Data Analysis System for LIGO I
Software Design

SOFTWARE DESIGN MUST SUPPORT A DISTRIBUTED NETWORK-BASED COMPUTING ENVIRONMENT

- Software Specific Requirements:
  - Portability:
    - Portable Operating System Interface compliant (POSIX) on Unix Platforms
    - ANSI Languages Compliant Code (C++ Standard, 11/14/97! - http://www.research.att.com/~bs/iso_release.html)
  - Extensible:
    - Object Oriented Programming Techniques in C++
    - Modular, Reusable Code Units elsewhere
    - Distributed Computing based on MPI
  - Maintainability:
    - Source Code Management using Concurrent Version System (CVS configured in client-server mode using CVSH)
    - Expressly Coded in Object Oriented C++ Language whenever possible
    - Keep It Simple Style (KISS) Guidelines for Coding Constructs
  - Flexibility:
    - Object Oriented Design (C++)
    - Modular Libraries (C, C++, others: e.g Fortran...)
    - Centralized Server-Client(s) paradigm for program control
    - Remaining infrastructure based on Standard Libraries (STL)

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

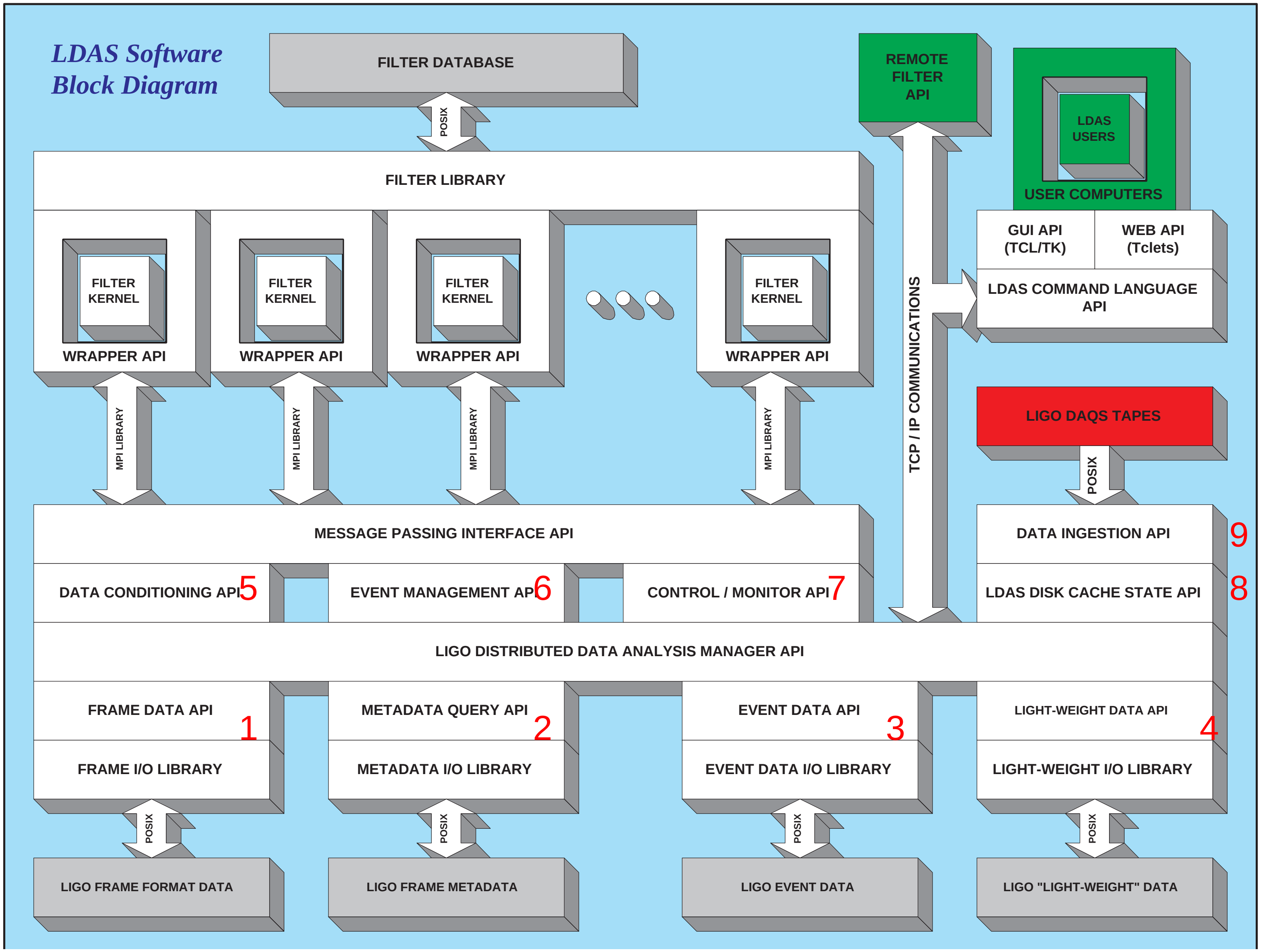
[Ref. LDAS Software Design chart, E1 at end of talk]

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



Data Analysis System for LIGO I
Software Development - Status

- GenericAPI [15 October status]:
  - Basis for all other APIs. Initial investment in prototyping and design will allow rapid diversification into specific APIs by extension of the generic class.
  - 90% complete

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

- FrameAPI [15 October status]:
  - Fcl I/O Library: complete
  - Fcl Specification: complete
  - FrameAPI 5% complete, expect mid-Nov completion date
- ManagerAPI:
  - FrameAPI 5% complete, expect mid-late Nov completion date
- DataConditioningAPI:
  - DataCondAPI 5% complete, expect mid-late Dec completion date
- Preliminary Design Review:
  - Develop prototype UserAPI by end of Dec.
  - Hold Preliminary Design Review Late December/Early January

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
  - » Recent workshop (22,23 October) with representatives from CERN, SDSC, CACR, Astronomy(IPAC/CIT) to review LIGO needs, compare with other programs from HEP, Astronomy
  - » Choices being considered:
    - – relational [deemed sufficient for LIGO needs]
      - – ORACLE (CIT license for campus MIS)
      - – PostgresSQL (INFORMIX precursor; public domain - 'free')
      - – miniSQL (similar to above)
    - – object-oriented DBMS
      - – Objectivity
  - » Issues: Buy-in costs; operational costs; upgrades if we start too low; metadata only vs (metadata+data); ...
  - » Plan to have a decision by PDR

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

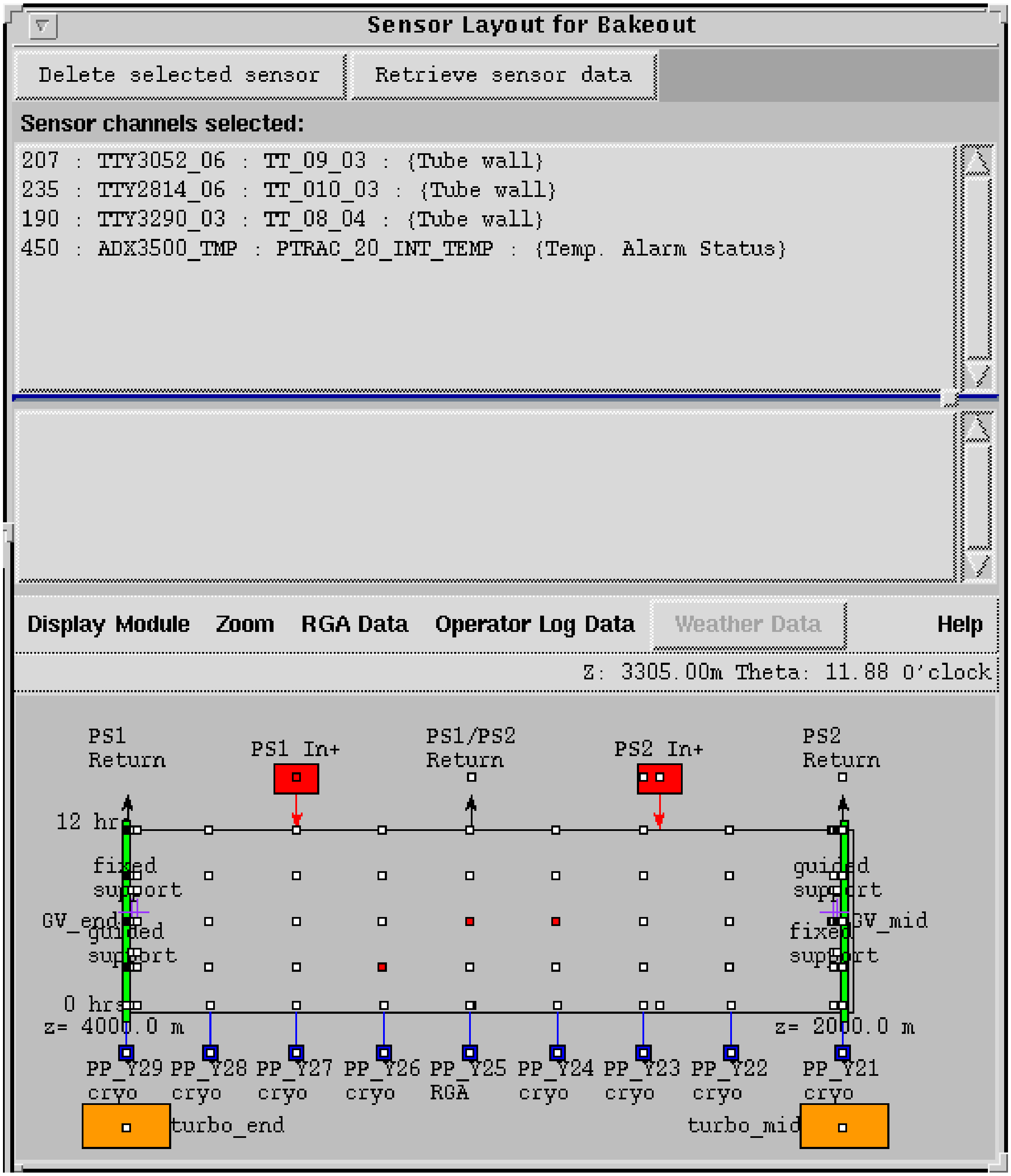
Database Management Systems

DBMS prototyping activities

- Developed a GUI using the tools being developed for LDAS to allow web-based (via browser-plus-plugins) access to DBs
- Data server independent of GUI
- Communications through query standard protocol
- Present: temperature data now available
- Next:
  - remaining datasets -- RGA data, weather data,operator logs
  - return LigoLW objects [XML]

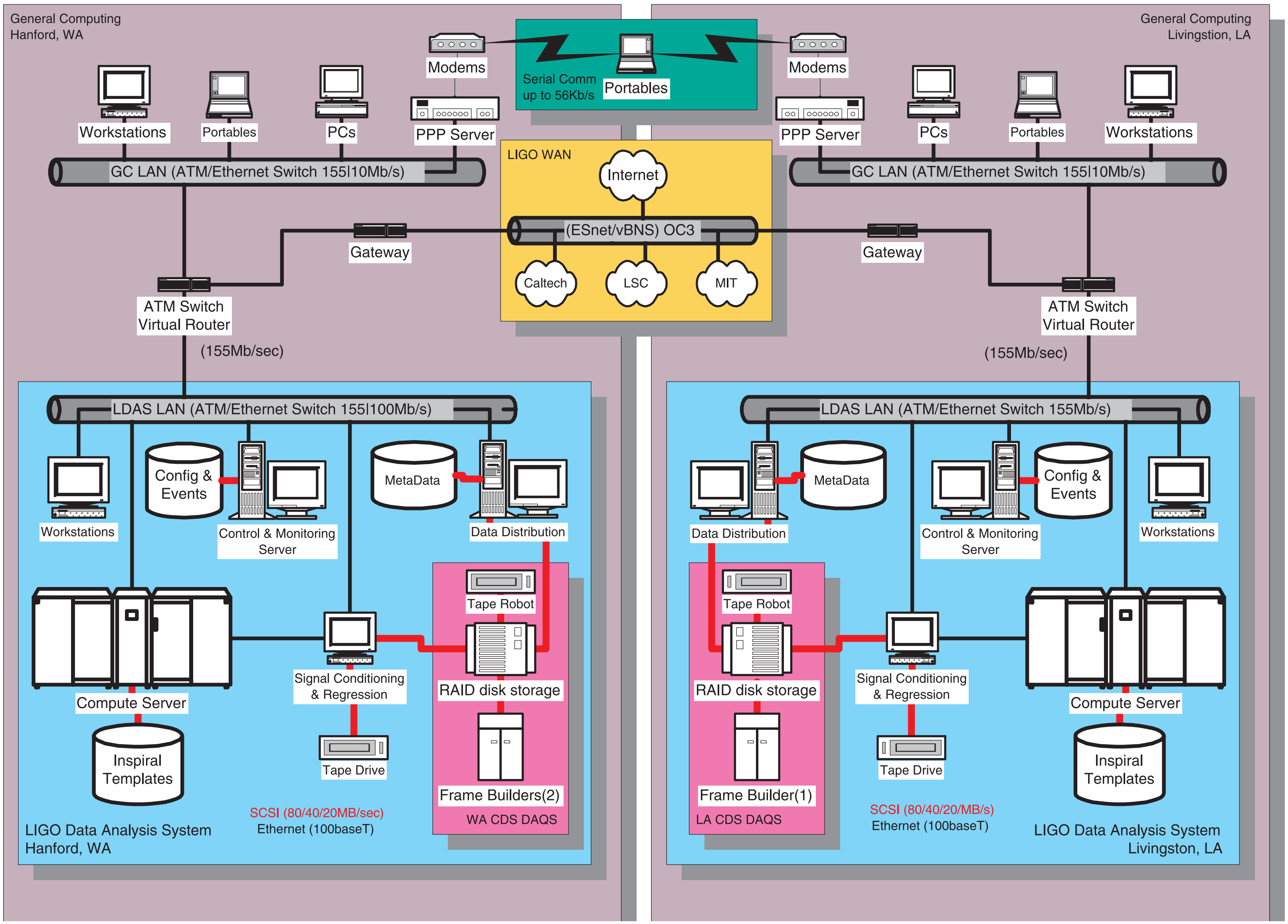
Database distribution tools

BT Bakeout Data Distribution Prototype



LIGO Data Analysis System
On-line architecture

[Ref. LDAS Off-line Hardware chart, F1, Hardware Spec. charts, F2-5 at end of talk]

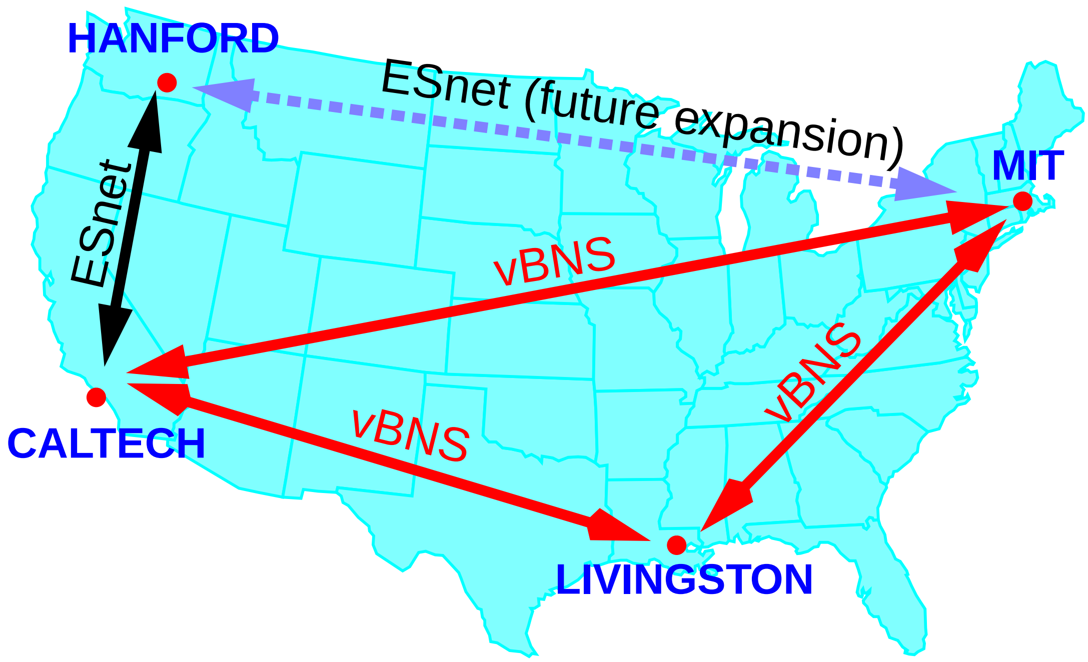


Hardware Status

- Network based data transmission at LIGO BW demonstrated
  - ATM<->ATM is adequate (6 - 15 MB/s; depends on platform, TCP/IP vs UDP).
  - 100BT<->100BT (needed for MPI, Beowulf cluster inter-node communications) is acceptable (>4.5 MB/s)
  - HIPPI<->HIPPI (with supercomputer mainframes) is superior (>30MB/s)
- Beowulf and MPI has been implemented on LIGO-scalable data flows for inspiral detection
  - Joint effort with CACR (Paragon) & Univ. Wisc. (PC/linux)
  - 8 node/16CPU integrated cluster ordered, due end October.
- Directed pulsar search prototype code has been developed using 40m data and CACR machines.
- Data archival technology choice deferred as late as possible (2001) - letting CACR lead way
  - Exploring 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 Wide Area Network

WAN Topology



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 as of 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

[Ref. LDAS WAN Toplogy.charts, G1,2 at end of talk]

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

End-to-End (e2e) Simulation

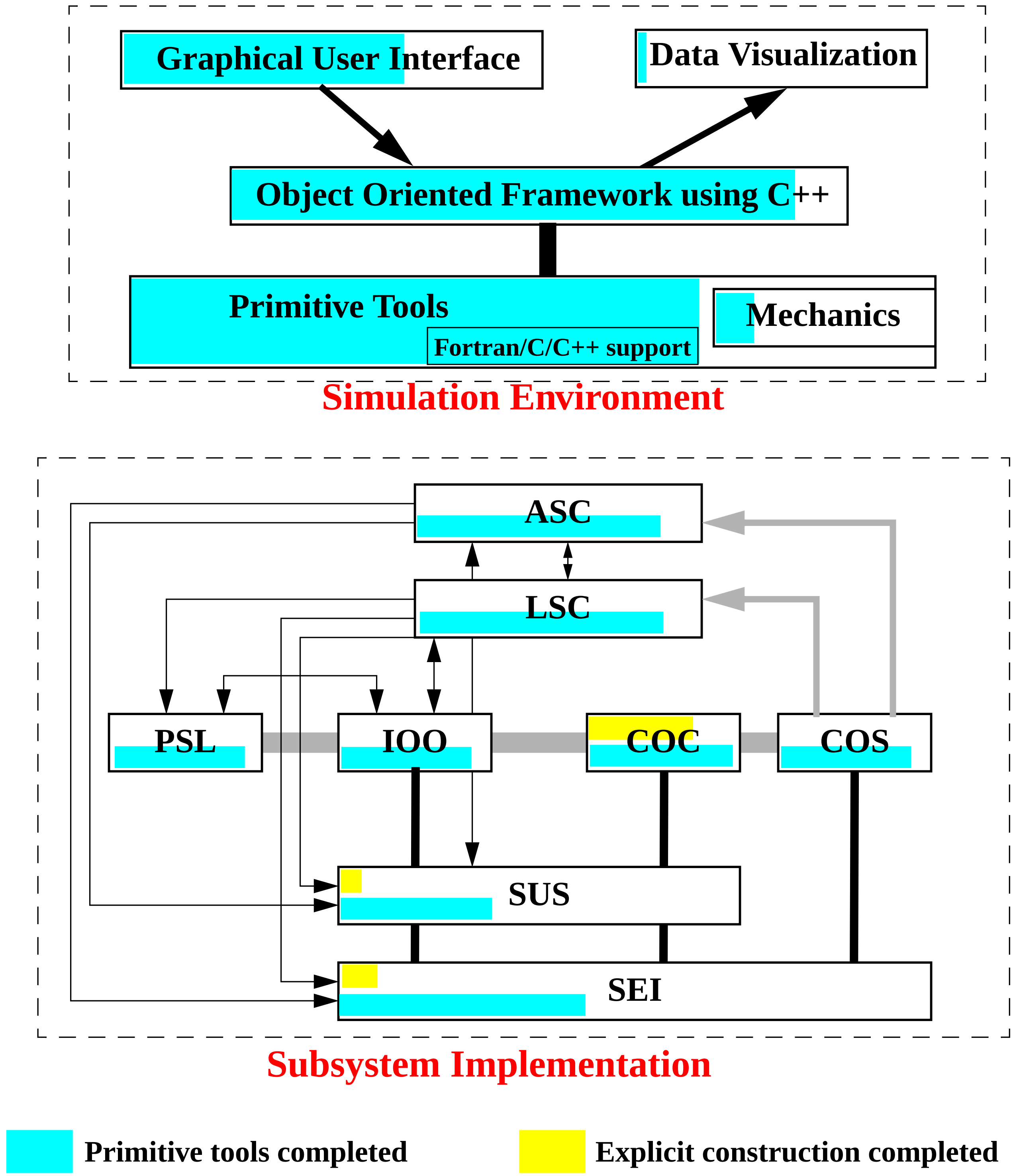
- Time domain simulation of LIGO interferometer output(s)
- Object Oriented structure using C++
  - Modular and expandable
  - Support for plug-ins using FORTRAN/C/C++
- No low level language (i.e., C++) needed to use
  - Easy to use high level language
  - GUI
- "Toolbox" Primitives
  - mirror - reflection, 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

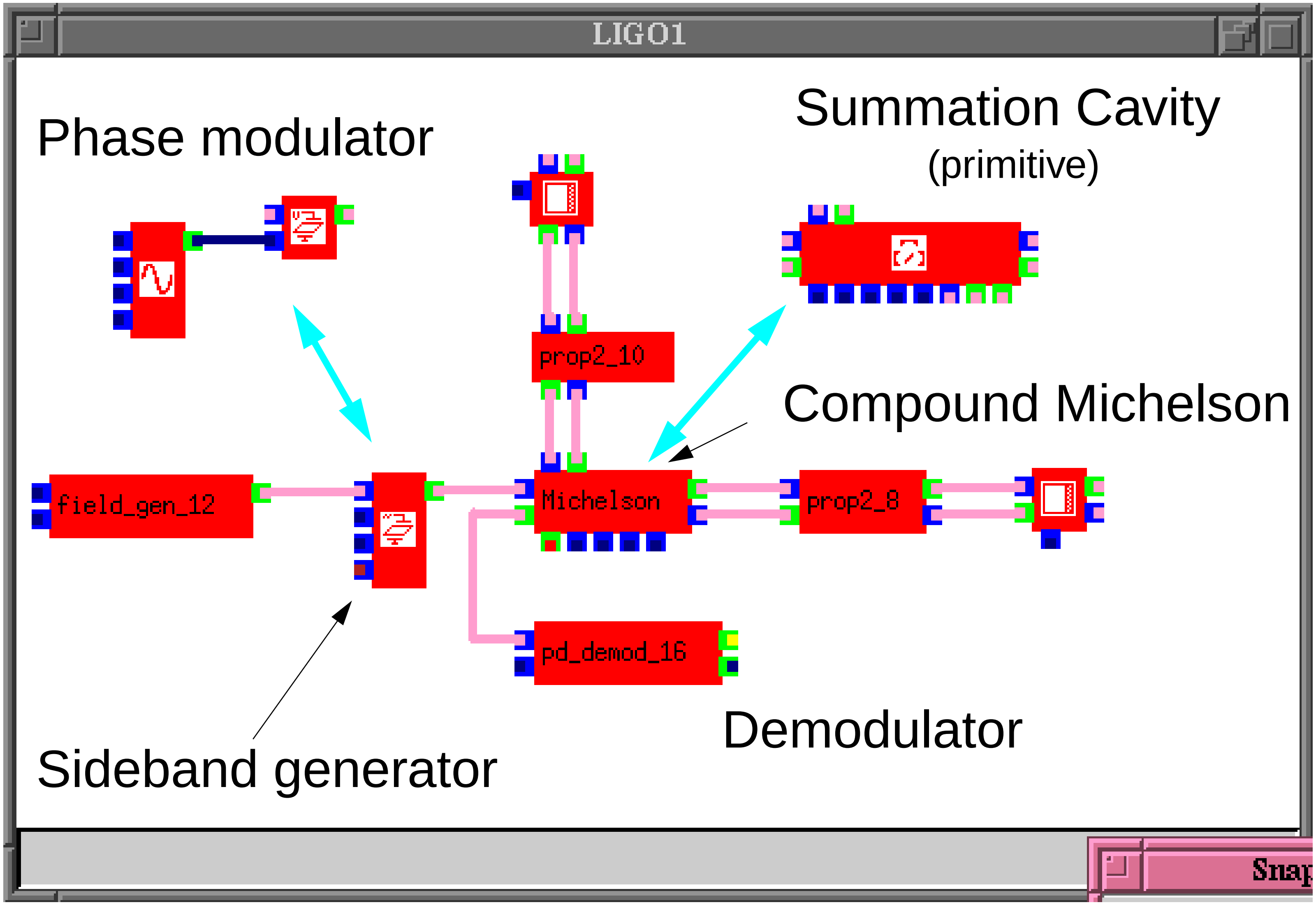
- Plan
  - » When the vertex Michelson at Hanford is available, E2E will provide the minimum set of subsystems so that semi-quantitative comparison of performance can be made.
- Construct the simulation models to map into the real hardware
  - » Implement phenomenological models for those parts which 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



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

Initial LIGO Sources - A1

Table 1: Initial LIGO Sources and Estimated Analysis Capability Requirements

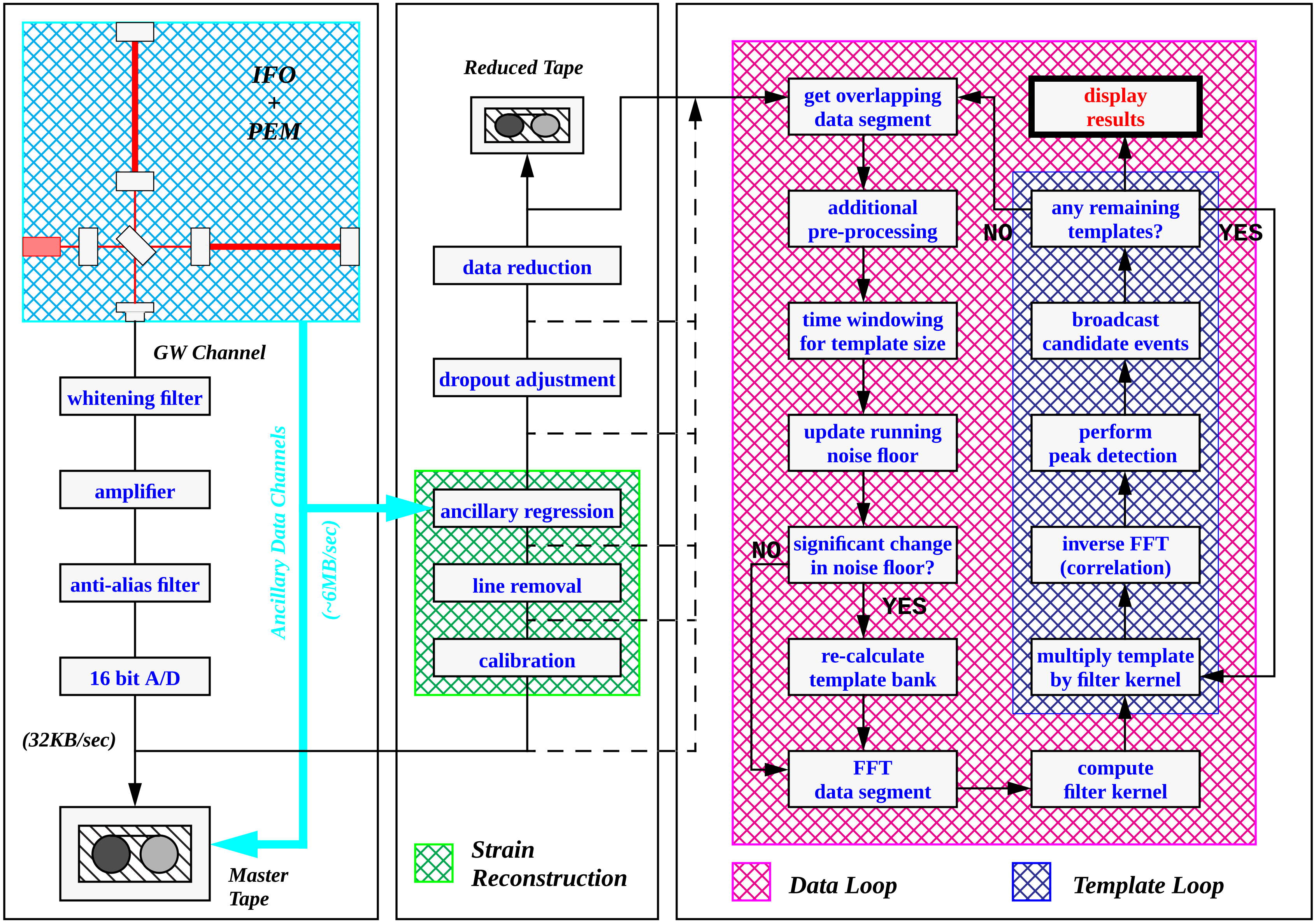
	Sources	Initial LIGO Performance Estimate	Data Analysis Requirements: CPU	Data Analysis Requirements: Storage	Data Analysis Requirements: Comments
Burst Signals $\Delta T < 1s$	Supernovae & Accretion-induced collapse of white dwarfs	$\Re_0 \sim 2-3/\,yr$ @ 15 Mpc If sufficiently asymmetric; however, ΔE_{GW} expected to be significantly less than $10^{-7} M_{solar}$	Minimal	Minimal Need PEM/houskeeping data for veto	• On-line analysis desirable for correlation with other astrophysics: EW • visible/radio/γ • ν Gravity • VIRGO/GEO • Resonant bars • Waveforms unknown • 2x/3x IFO correlation of events
	BH/BH Collisions	$\Re_0 \sim 1/\,yr(?)$ @ 500 Mpc;			
Chirped Waveform $10s < \Delta T < 1000s$	NS/NS Inspirals	$\Re_0 \sim 3/\,yr$ (?) @ 23 Mpc; for $M_{NS} \sim M_{solar}$ $\Delta T \sim 36 \times T_{inspiral} = 360s$	~ 7.2 GFLOPS (WA)	Templates/Data ~5 GB /~24 MB	• On-line analysis appears feasible down to $\sim 1\ M_{solar}$ • 1x/2x/3x correlations feasible depending on SNR. • Coalescence event may generate correlated (EW) signals as above. • PEM/housekeeping needed for vetoing • Template matching (Wiener filtering) or wavelet analysis in f-t domain.
	BH/BH & NS/BH Inspirals	$\Re_0 \sim 1/\,yr$ @ 150 Mpc; for $M_{BH} \sim 3M_{solar}$ $\Delta T \sim 36 \times T_{inspiral} = 60\ s$	~ 330 MFLOPS (WA)	~ 41 MB /~4 MB	

Initial LIGO Sources - A2

Table 2: Initial LIGO Sources and Estimated Analysis Capability Requirements

	Sources	Initial LIGO Performance Estimate	Data Analysis Requirements		
			CPU	**Storage**	**Comments**
Periodic Signal $\Delta T \sim 10^6$ - 10^7 s	Pulsars with mass asymmetry $\frac{S}{N} \approx 1.5\left(\frac{\varepsilon}{10^{-6}}\right)\left(\frac{10\text{kpc}}{r}\right)\left(\frac{1\text{ms}}{P}\right)^{\frac{5}{2}}\sqrt{\frac{T_{int}}{1\text{month}}}$ $\tau_{spindown} \sim 830\text{yr}\left(\frac{f_{rot}}{1\text{kHz}}\right)^{-4}\left(\frac{\varepsilon}{10^{-6}}\right)^{-2}$	$\frac{S}{N} \approx 8$ $\varepsilon = 10^{-5}$ r=10 kpc P=1 ms $T_{int} = 10^7\text{s}$	Only directed searches feasible for nearby sources	10 GB for 10^6s (GW waveform)	• Off-line analysis • Detection less sensitive to non-Gaussian noise; more sensitive to calibration drifts. • Detection techniques as for pulsars -- narrow line sources with modulated frequency. • Correlations among interferometers may be performed (if needed) after detection. • A 4π sr. search requires decomposition of the sky into a very large number of pixels. Exact number is sensitive details of stacking.
Broadband Signals $\Delta T \sim 10^6$ - 10^7 s	Stochastic Background $\zeta \equiv \frac{\Omega_g}{\Omega_0}$	$\zeta \geq 3\times10^{-6}$ $40\text{Hz} < f < 300\text{Hz}$ $T_{int} = 10^7\text{sec}$	Minimal requirements -- analysis may done on single workstations; study of systematic correlated noise effects may require significantly more processing.		• Off-line analysis • Requires multiple interferometers to be correlated

LIGO Data Flow (Model) -- A3

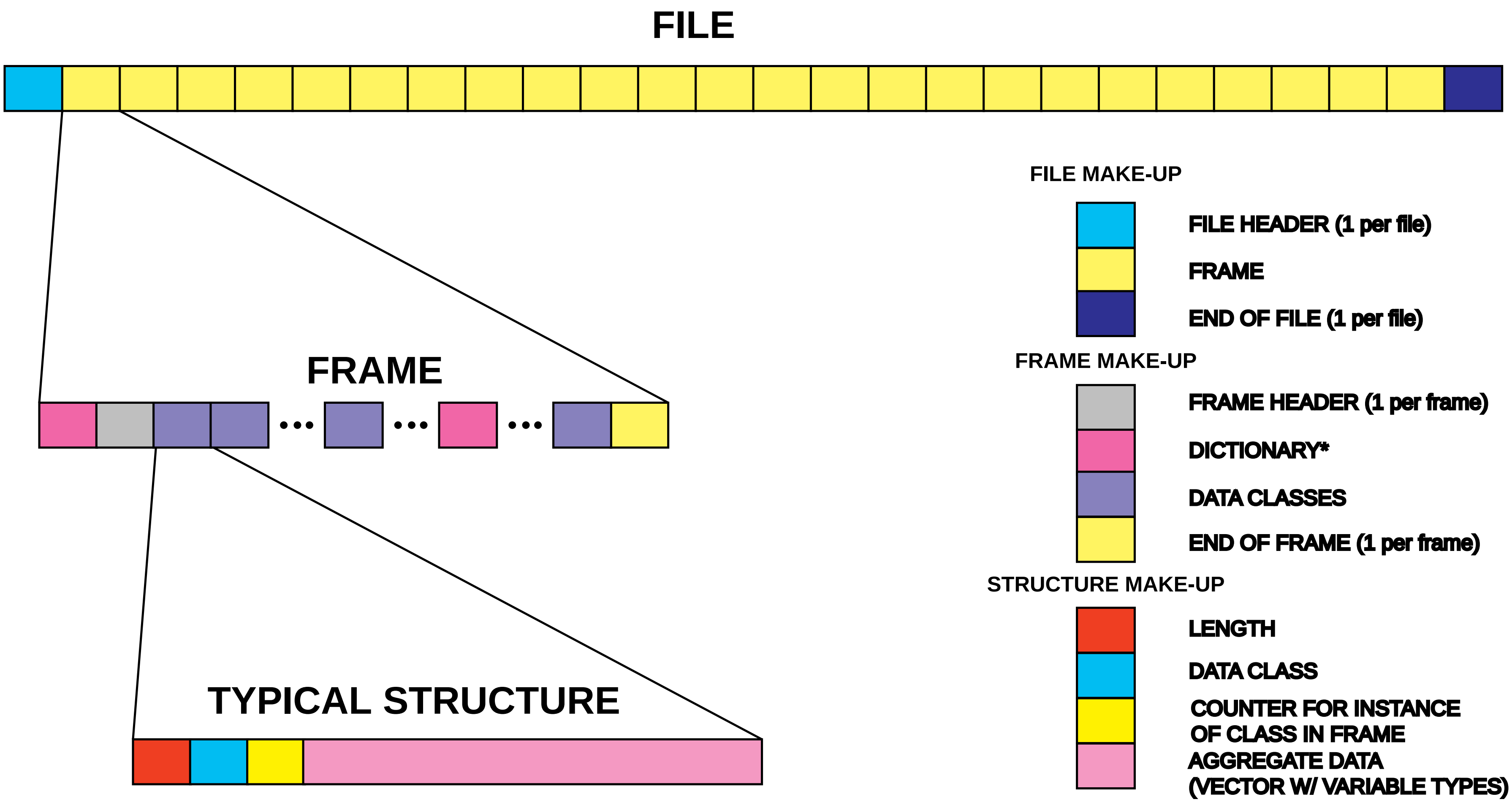


LIGO

LIGO-G980118-00-E

Frame Format Implementation - B1

Frame Composition



*** Dictionary structure behavior is unique in that:**
1. It preceeds header for first frame of file;
2. Dictionary is built up incrementally as addititional structures are incorporated into frame
3. It is valid for entire file (persistent)

LDAS Reduced Data and Metadata - C1

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 - C2							
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**

LigoLW - D1

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>
```

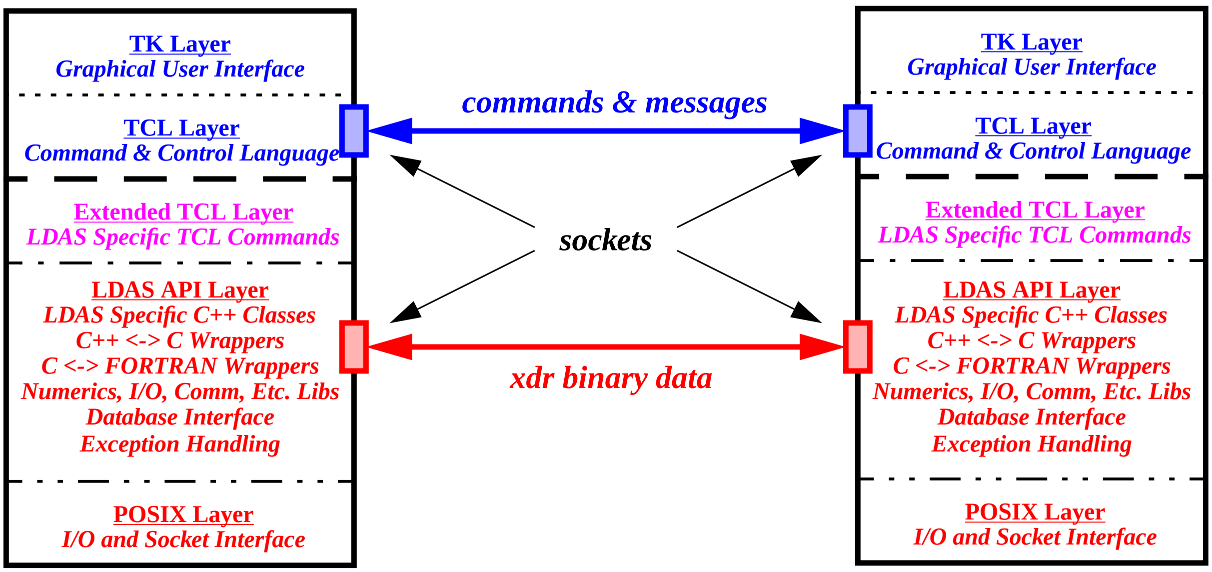
LigoLW - D2

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>
```

LIGO Data Analysis System
Software Design - E1

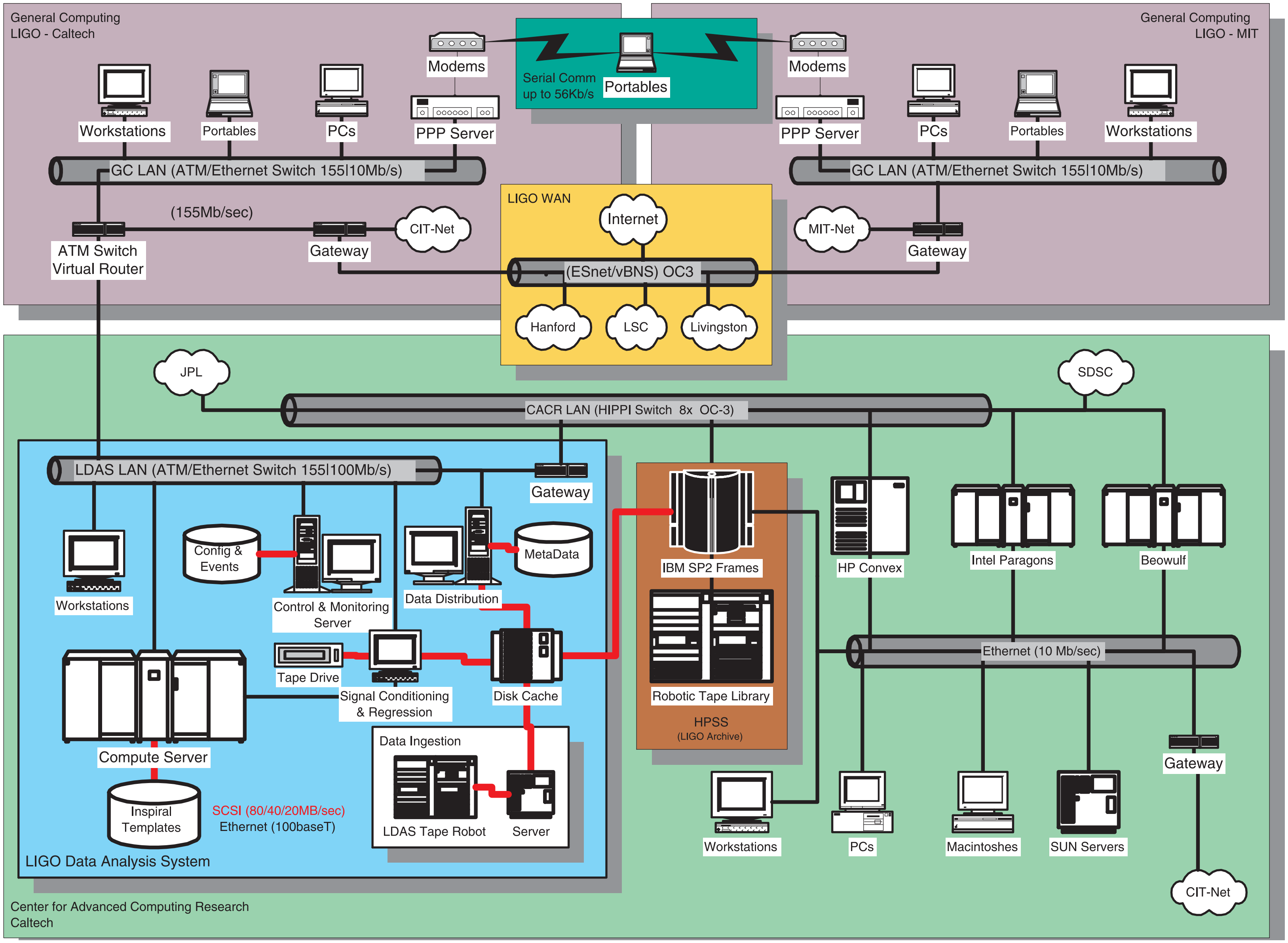
APIs “*TWO-LEVEL*” SOCKET COMMUNICATIONS



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	

LIGO Data Analysis System
Off-line Architecture - F1



LIGO Data Analysis System Hardware Elements - F2

- Provides distributed computing and archival across the four laboratory sites

Hanford: Operations for 2 interferometers

Component	Specification	Cost (K$)
Data distribution system, On-line		
Servers	6ea @ 600MHz	
Disk system	RAID, Ultrawide/fast SCSI 4 ports; shared w/CDS	
	500GB	
Metadata storage system	50GB, Ultrawide/fast SCSI	
Computational engines, On-line		
Signal conditioning, regression engines	4ea @ 600MHz	
Compute server (BEOWULF system, 2 X 10 GFLOPS)	64 nodes @ 600MHz ea, local disk space + RAM	
Control & monitoring	2ea @ 600MHz	
Networking		
Networking switches/routers)	100BT/OC3(ATM)	
ESnet access, hardware required	4 x T1	
Off-line analysis		
Off-line analysis stations	10ea @ 600MHz	
SW & Peripherals	Licenses/printers/plotters/ tape drives/scanners/local SCSI disk/...	
Hanford, total estimated		**$925**

LIGO Data Analysis System Hardware Elements - F3

Livingston: Operations for 1 interferometer

Component	Specification	Cost (K$)
Data distribution system, On-line		
Servers	3ea @ 600MHz	
Disk system	RAID, Ultrawide/fast SCSI 4 ports; shared w/ CDS	
	375	
Metadata storage system	50GB, Ultrawide/fast SCSI	
Computational engines, On-line		
Signal conditioning, regression engines	2ea @ 600MHz	
Compute server (BEOWULF system, 2 X 10 GFLOPS)	32 nodes @ 600MHz ea, local disk space + RAM	
Control & monitoring	1ea @ 600MHz	
Networking		
Networking switches/routers)	100BT/OC3(ATM)	
vBNS access, hardware required		
Off-line analysis		
Off-line analysis stations	7ea @ 600MHz	
SW & Peripherals	Licenses/printers/plotters/ tape drives/scanners/local SCSI disk/...	
Livingston, total estimated		**$545**

LIGO Data Analysis System Hardware Elements - F4

MIT: Data analysis

Component	Specification	Cost (K$)
Off-line analysis		
Off-line analysis stations	10ea @ 600MHz	
Local disk cache	400 GB	
SW & Peripherals	Licenses/printers/plotters/ tape drives/scanners/local SCSI disk/...	
Networking		
Networking switches/routers	100BT/OC3(ATM)	
vBNS hardware/hookup	OC3/ATM	
MIT, total estimated		**$410**

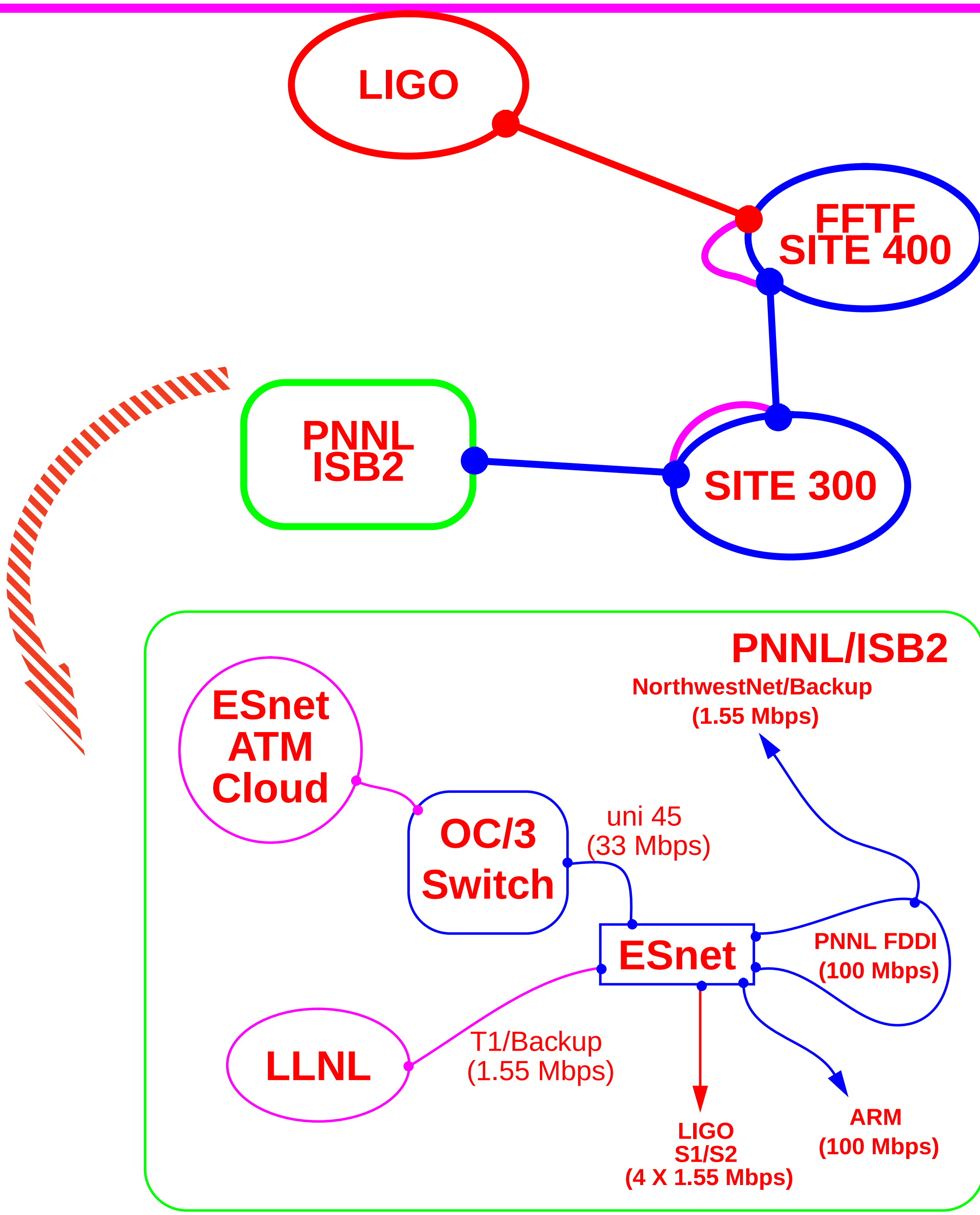
LIGO Data Analysis System Hardware Elements - F5

Caltech: Off-line operations & data analysis

Component	Specification	Cost (K$)
Data distribution system, Off-line		
Tape robots	8ea	
Cabinets	4ea	
Disk Cache	4000 GB, RAID	
Servers	10ea @ 600MHz w/extra memory kits/ATM	
Computational engines, Off-line		
Signal conditioning, regression engines	4ea @ 600MHz	
Compute server (BEOWULF system)	96 nodes @ 600MHz ea, local disk space + RAM	
Post-processing workstations	4ea @ 600MHz	
Networking switches for compute server	100BT/OC3(ATM)	
Networking		
vBNS access, hardware required		
Networking, LAN	100BT/OC3(ATM)	
Off-line analysis		
Off-line analysis stations	15ea @ 600MHz	
SW & Peripherals	Licenses/printers/plotters/ tape drives/scanners/local SCSI disk/...	
Caltech, Off-line total estimated		**$2504**

LIGO Hanford WAN Topology - G1

Link to ESnet



LIGO Livingston WAN Topology - G2

Link to LSU/vBNS