LIGO Data Analysis System (LDAS)

A. Lazzarini

LDAS Software

J.K. Blackburn

LIGO WAN

A. Lazzarini

LSC Meeting
LIGO Hanford Observatory (LHO)
12 - 14 March 1998

LIGO-G980124-00-E

LIGO
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Data Analysis System for LIGO I
Status

1. Design Requirements Review completed 12/97

- Provide on-line analysis at the observatories; data distribution from on-line cache -- 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.

Data Analysis System for LIGO I
Status

2. Testing & prototyping

- Target 40m prototype -- identify & confront practical issues
- Collaborative involvement
  - CACR/Caltech, Michigan, Northwestern, Wisconsin
- Data distribution
  - Hardware: RAID/UNIX Server configuration
  - Software: Web-based data retrieval/conditioning/distribution/display
  - Data Model definition
    - Size: what is needed?
    - Media: how soon?
    - Cache: how often?
- Compute server
  - Hardware: PC/Linux (Alpha/Linux?) Fibrechannel/Ethernet cluster -- BEOWULF
    - 10+ GFLOPS @ observatories (on-line);
    - 3 x observatory @ Caltech (off-line);
    - Performance:cost ~ 5X - 10X shared memory parallel processors;
  - Software:
    - MPI distributed processing
    - Benchmark of inspiral searches - optimal Wiener filter
    - Evolution of GRASP elements to LIGO filters for LDAS

Data Analysis System for LIGO I
Status

3. Data type definition

- Frame format for raw data (time series)
  - coordinated effort with VIRGO
  - structured/highly generalized/extensible
  - API in C/C++ being developed for I/O
  - interface to MATLAB available, being tested

- Lightweight (LW) format
  - defined implementation
    - SDF (ASCII) - standard developed/used at CACR (J. Salmon author)
    - netCDF (binary) implementation -- commercial standard
  - suitable for single/few channels; spectra; reduced datasets, ...
  - implementable/manipulable by individual researchers - less comprehensive, easier to code on one's own.
  - interfaces to MATLAB/IDL/...

Data Analysis System for LIGO I
Status

3. Data type definition (cont.)

- Metadata -- “Data about data”
  - developed (preliminary, non-comprehensive) list
    - machine state vector/configuration/operational modes
    - calibrations/triggers/vetoes/...
    - operator logs -- electronic notebook
    - non-LIGO (collateral) data/links -- seismic reports; weather EM storm activities; other astrophysics - GRBs/ν
    - heterogeneous: series, vectors, files, text, binary, images,...
  - will likely be distributed across LIGO Laboratory/LSC
  - still need to define environment -- cost.

- LIGO Event Data
  - pending definition
  - archive of “events” discovered in LIGO data -- anomalies & true
  - time series excerpts -- raw data, striped across many channels for brief epochs containing the events
  - collateral data -- environmental, other GW detectors, other astrophysical observatories...
  - parametric descriptions
  - templates, algorithms,...

Data Analysis System for LIGO I
Status

4. Software system design

Data analysis -- scope: LSC + LIGO Laboratory

- Data analysis flows - sizing of requirements:
- Data analysis software prototyping - GRASP code
- Data usage model

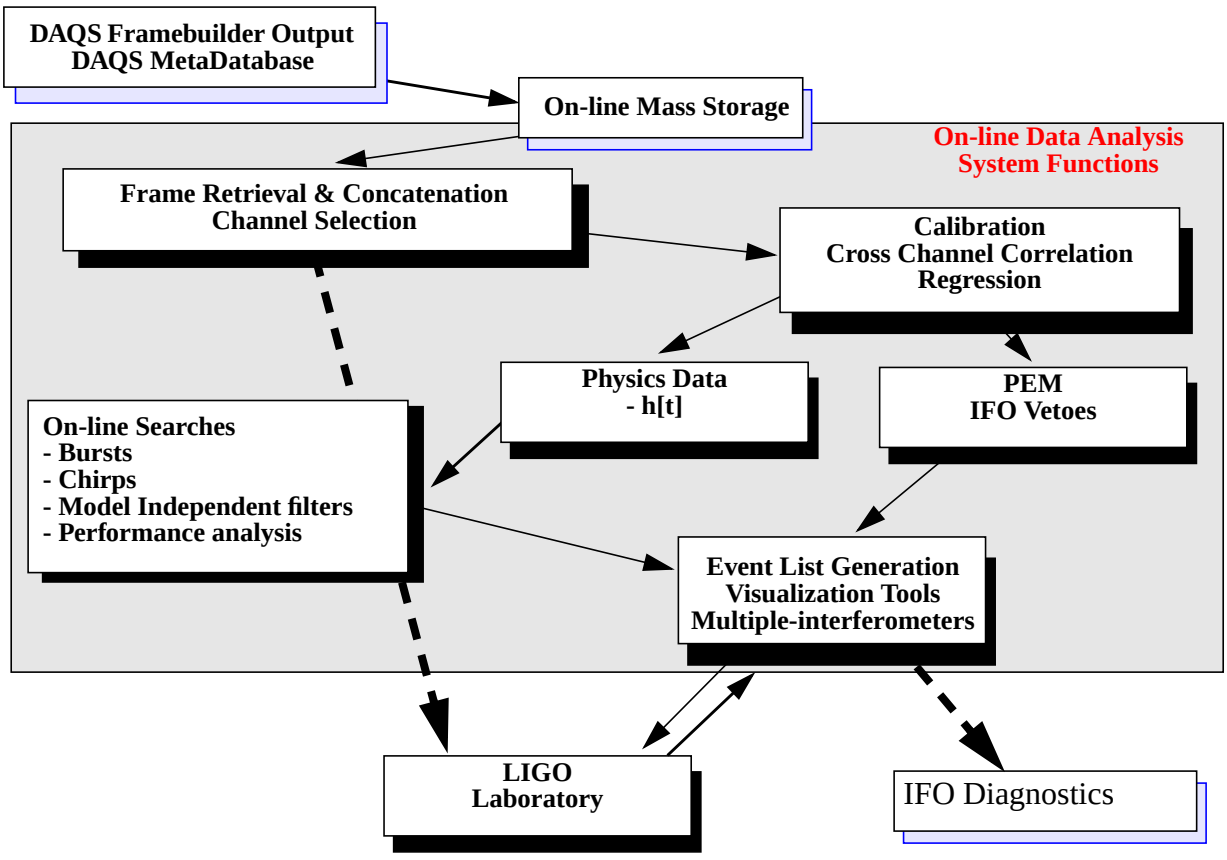
Data management -- scope: LIGO Laboratory

- Design and definition of architecture & components:
- Data distribution & access
  - Storage systems & archives
  - Data transmission & downloading
- Metadata creation/archival/retrieval
- API design/development
  - Data ingestion (incorporation of new/recent data)
  - LDAS command language
  - Interprocess communications -- LDAS distributed data analysis manager
  - Disk cache management
  - Access to data libraries - Frame/LW/Metadata/Event
  - Filtering/MPI/Conditioning
  - Control/Reporting

LIGO Data Analysis System (LDAS)
Design

- Two LDAS components
  - Observatory LDAS (on-line)
    - Two systems, one for Hanford, and one for Livingston
    - Hanford system handles 2 interferometers
  - Caltech LDAS (off-line)
    - Collaborative arrangement with CACR
      - Dedicated LIGO hardware within CACR on scale of observatory systems
      - Database archive
      - Strategic use of other CACR facilities as available
    - Transparent access for off-line analysis of archived data
      - LIGO Laboratory
      - LIGO Scientific Collaboration
- Wide area network (WAN) to enable inter-site communications
  - University scientific and engineering support to Observatories
  - Access to archive database
  - Access to real-time data from observatories
  - Inter-observatory event sharing

LIGO Data Analysis System (LDAS)
On-line Functions



LDAS Flow Block Diagram: On-line Analysis

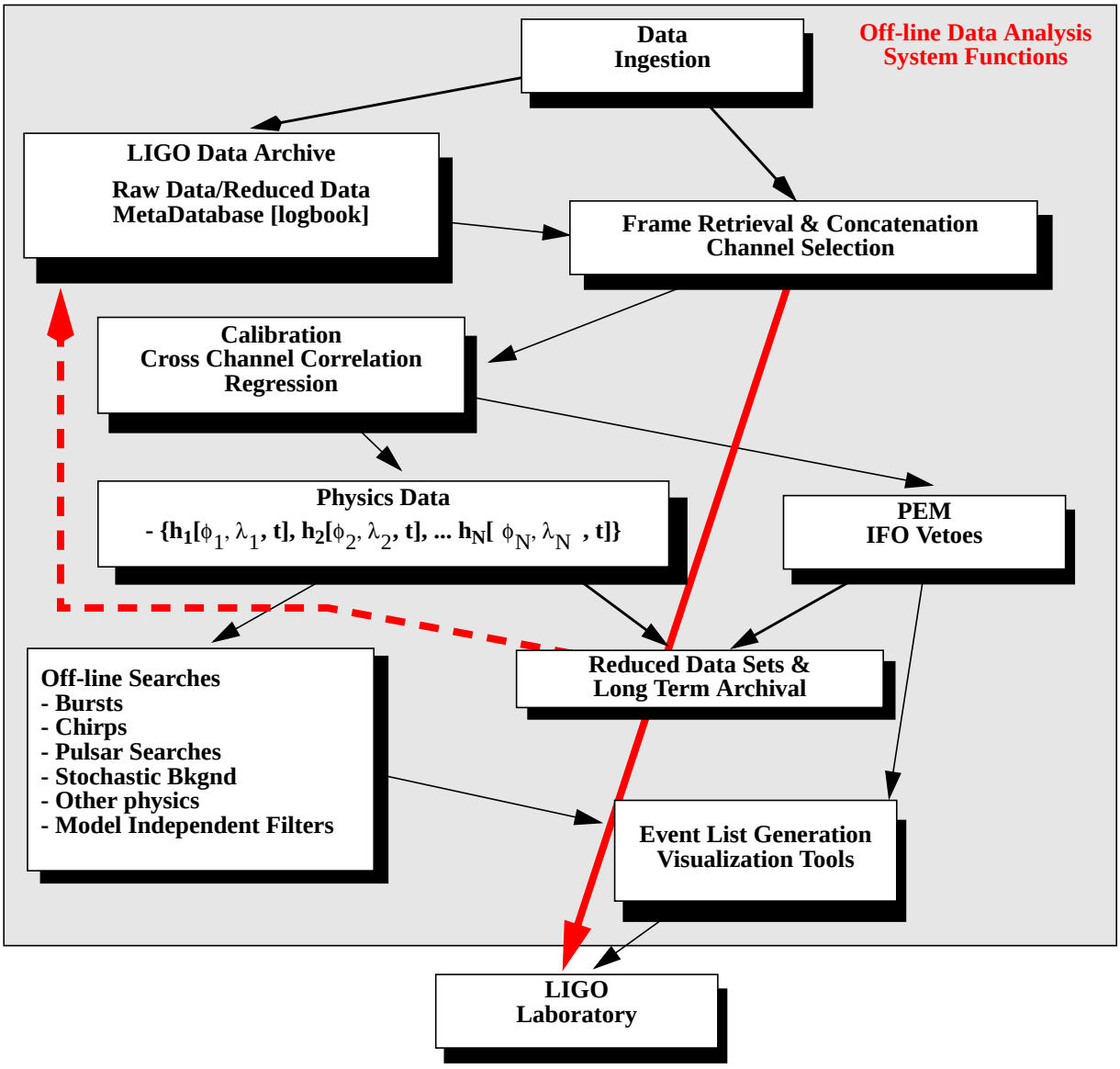
DAQS: Data AcQuisition System

LDAS: LIGO Data Analysis System

PEM: Physical Environment Monitors

IFO: InterFerOmeter

LIGO Data Analysis System (LDAS)
Off-Line Functions



LDAS Flow Block Diagram: Off-line Analysis

DAQS: Data AcQuisition System

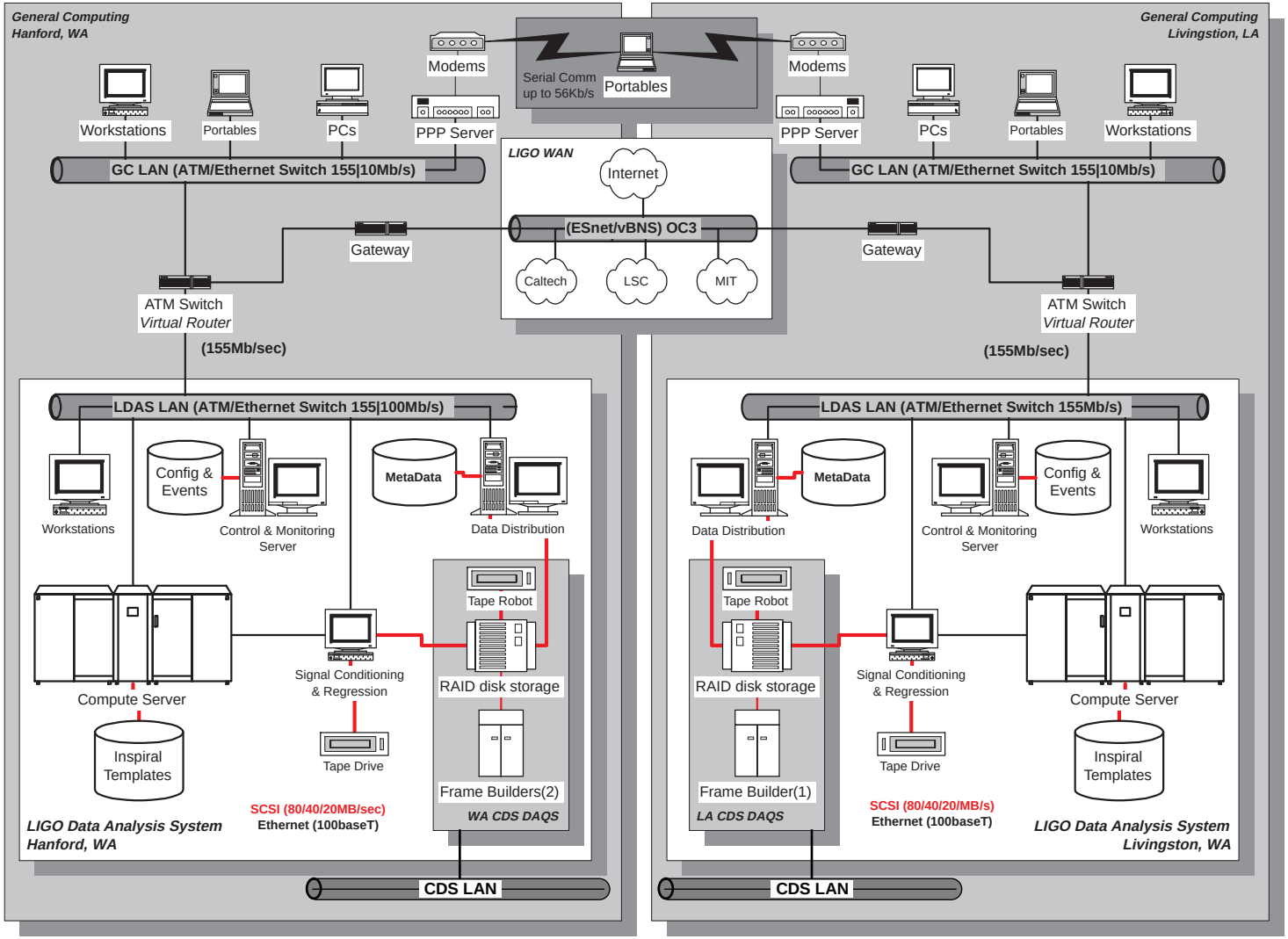
LDAS: LIGO Data Analysis System

PEM: Physical Environment Monitors

IFO: InterFerOmeter

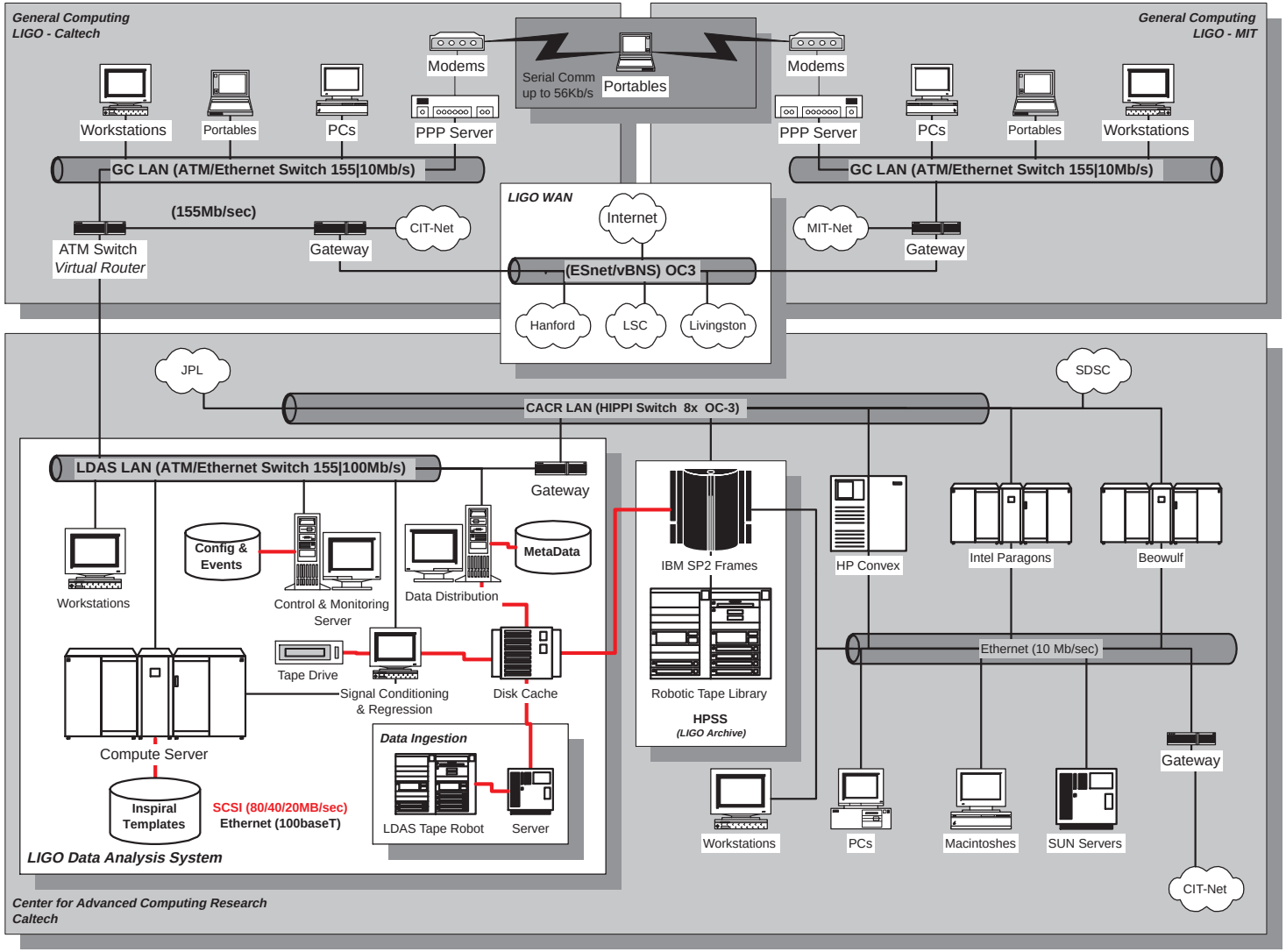
LIGO Data Analysis System

On-line architecture



LIGO Data Analysis System

Off-line Architecture

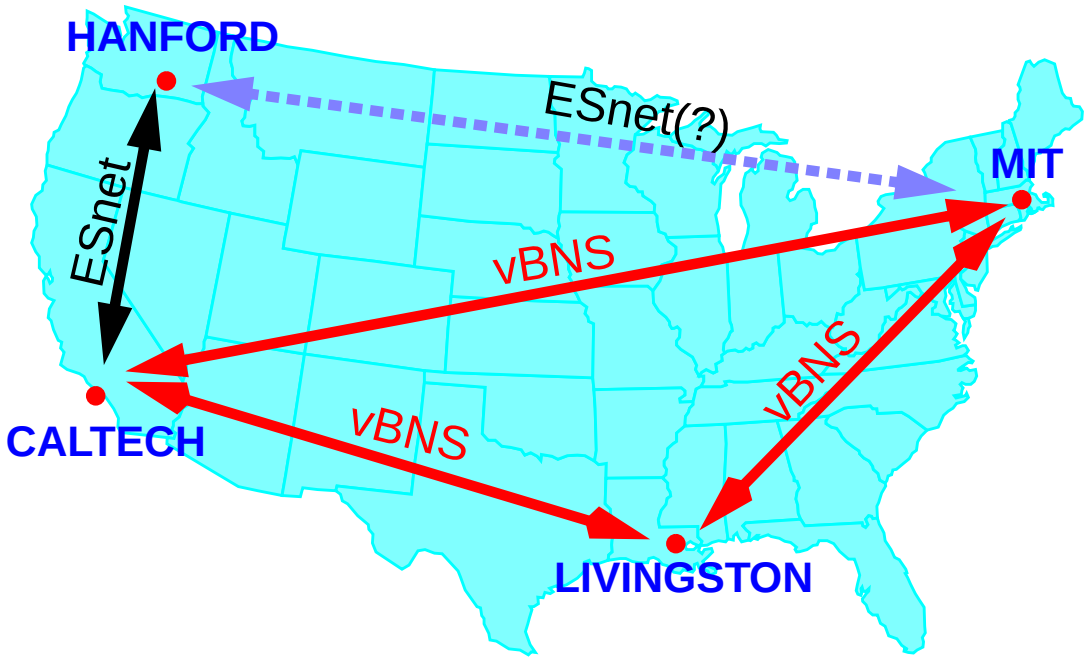


LIGO Data Analysis System
To-Do List

- Design & definition
  - Data -- channel lists/frame contents/types of frames/...
  - Metadata -- contents/environment
  - Algorithms -- hierarchical searches/periodic searches/f-t processing/wavelets/...
  - Events -- definition
  - LDAS architecture -- complete design/definition
- Development & prototyping
  - LDAS command language syntax
  - Scripting language implementation -- interprocess control & communication
  - Data distribution -- 40m implementation
  - Compute server -- BEOWULF cluster; integrate ~8 node cluster
  - Algorithms -- same as above
  - Benchmarks -- algorithms; data distribution;...
  - Visualization tools
    - Applets
    - Plug-ins
    - AP Interfaces to commercial/extant products
      - Matlab, IDL, Triana(GEO), ...
  - Data transmission -- quantify WAN performance/ limitations

LIGO Wide Area Network

Plan



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
  - Final MOU complete: awaiting signatures at NSF, DOE
  - Proceeding to implement initial (T1) capability; requested up to 4 x T1 BW (cost is an issue).
  - SOW/PO with PNNL & Lockheed-Martin to procure switching & routing equipment almost complete
  - Cross-over between ESnet and vBNS takes place at CACR/Caltech-HEP
  - MIT may be added later as a separate action
- Exploring with PNNL (EMSL) and WSU/ Richland possibility of a consortium to propose to NSF a direct vBNS hookup in Tri-Cities area
  - Follows model in place at Livingston
  - EMSL needs better connectivity to Seattle, Caltech, SDSC

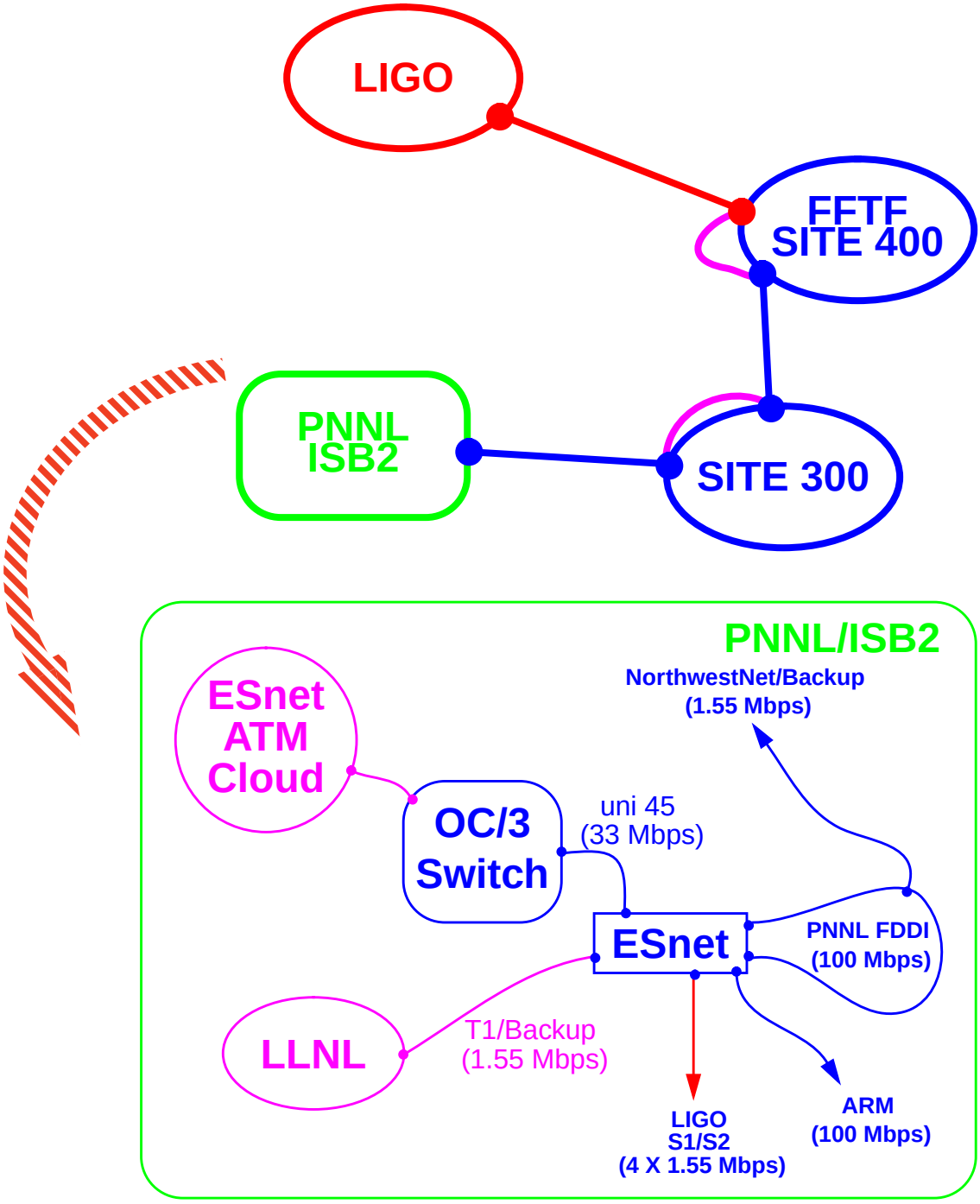
LIGO Wide Area Network

Status

- T1 link to Livingston Observatory in place
  - LSU provides gateway service
  - Recent proposal by LSU to NSF for vBNS connection includes LIGO access at Livingston
  - FO link from observatory to campus via Bell South switch near Livingston

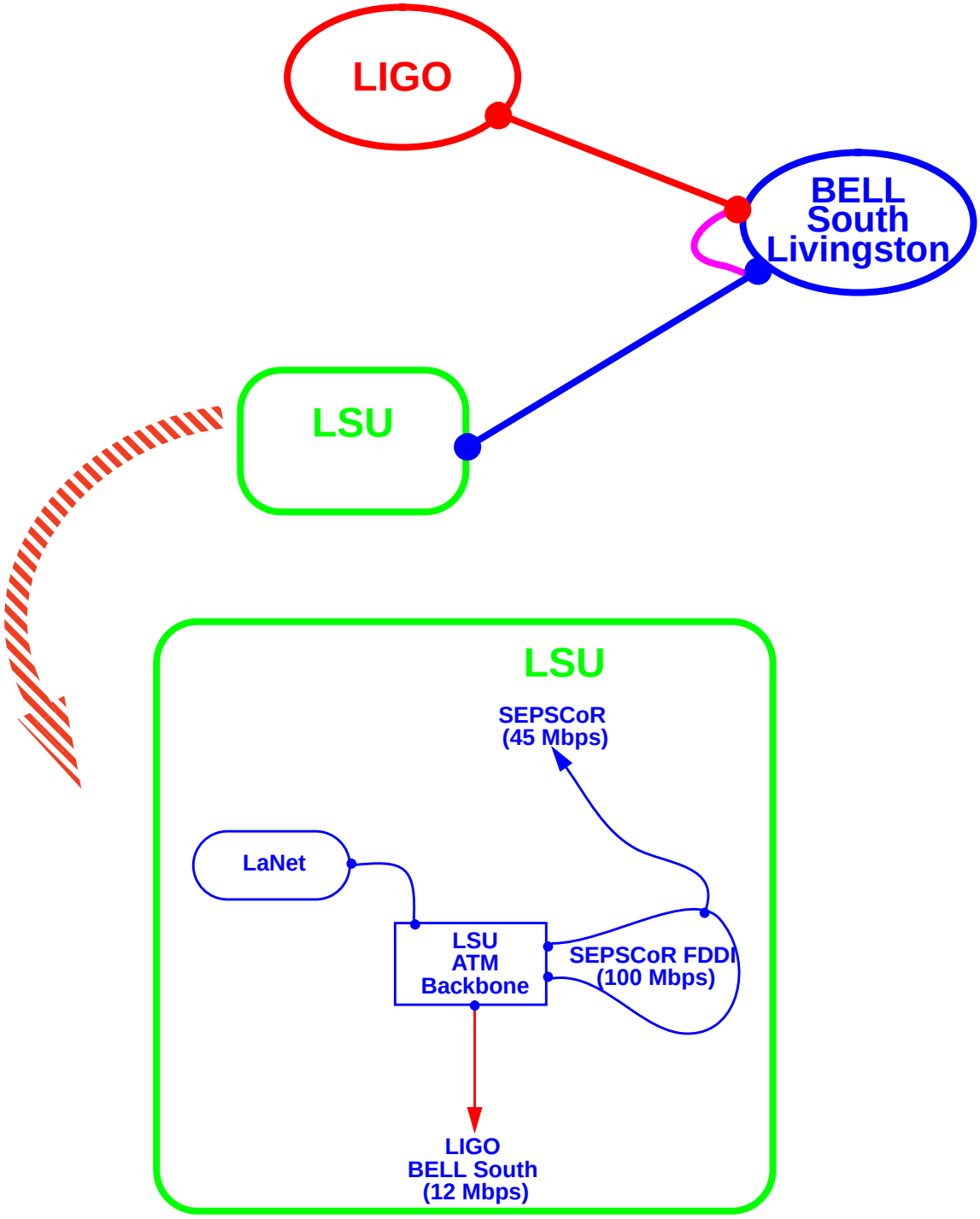
LIGO Hanford WAN

Link to ESnet



LIGO Livingston WAN

Link to LSU/vBNS



LDAS Development
Timeline

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

Detector Milestone:	Date	LDAS Need
›› Data Acquisition System, 2km:	9/98	Min. data dist.
›› PSL/Input Optics	2/99	“
›› Vertex Michelson, first light	7/99	Full data dist.
›› 2km operational	6/00	On-line system

- 4 km interferometers staggered in time by 3 & 6 mos.
- Staged installation at CACR of off-line system in period 6/99 - 12/01