LIGO Data Analysis and Simulation Summary

7th Meeting of the LIGO PAC

Cambridge, Massachusetts

18 November 1999

Albert Lazzarini

LIGO Laboratory Caltech

LIGO-G990122-00-E

E2E Development Status

- Schedule is ~ 4 months behind plan
  - Discovered an architectural issue in simulation engine- e.g. difficulty in dealing with servo loops
  - Miscellaneous code improvements along the way
  - Debugging of code elements more difficult than expected
  - Scientist and Programmer shortage
    - Lost one programmer;
- Goal for next 6 months
  - Catch up to hardware configuration - 2k FP or short Michelson
  - Improve mechanical simulation (add realism)
  - Greater LSC participation

LDAS Development Status

- Implementation is behind plan
  - » $\alpha 1$, $\alpha 2$ releases will be combined to a single α release, ~6 month delay
  - » β released still planned for end of 2000
- Reasons
  - » Staffing ramp-up slower than expected
    - – Learning curve
    - – Programming staff retention (lost 3, looking for 2)
  - » Development of infrastructure is a front-loading activity
    - – Extensive core API module development
    - – C++ frame library development (1/2 m-yr)
    - – C++ code standardization to ANSI specification (egcs -> gnu 2.95.1 transition took ~ 2 months)
  - » Early investment of time to develop tools to leverage lean staff'
    - – Developed www page online resources to assist new programmers
    - – Extensive use of cfengine scripting utility to enforce OS installation configuration control across remote sites from Caltech

LDAS Development Status

- Reasons
  - » Early investment of time to develop tools to leverage lean staff'
    - – Extensive use of autoconfig/automake unix utilities to compile, build code releases
      - ♦ Difficult learning curve to implement well, correctly
      - ♦ Long term labor saver to permit automated code builds, releases, along with automated documentation generation (html)
      - ♦ Permits uniform code environment to be enforced across all LIGO sites, remotely managed from Caltech
- Solutions
  - » Add 2 senior programmers
    - – Access to existing LIGO programmers coming off other work
    - – Offloading existing programmer from system administration load
    - – New hire
  - » Development of new modules reuses much of code foundation

Laboratory - LSC Coordination

- LSC White Paper on Data Analysis
  - » Co-authored by Tom Prince
  - » Draft released 27 October 1999, pending LSC review, adoption

- Draft specification for algorithm library software drafted, released for review
  - » LLAL: LIGO/LSC Algorithm Library
  - » C code implementation
  - » Interface to on-site & off-site analysis pipelines
  - » Test of specification through prototype code development

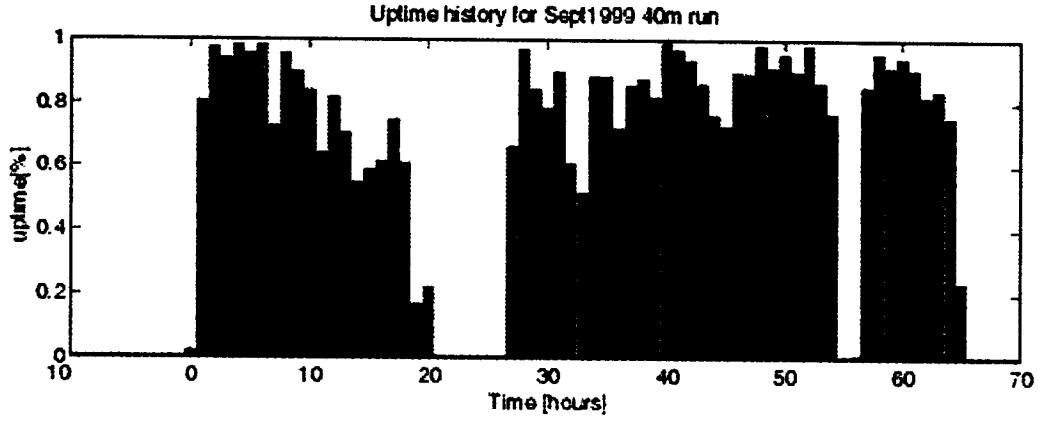
Laboratory - LSC Coordination

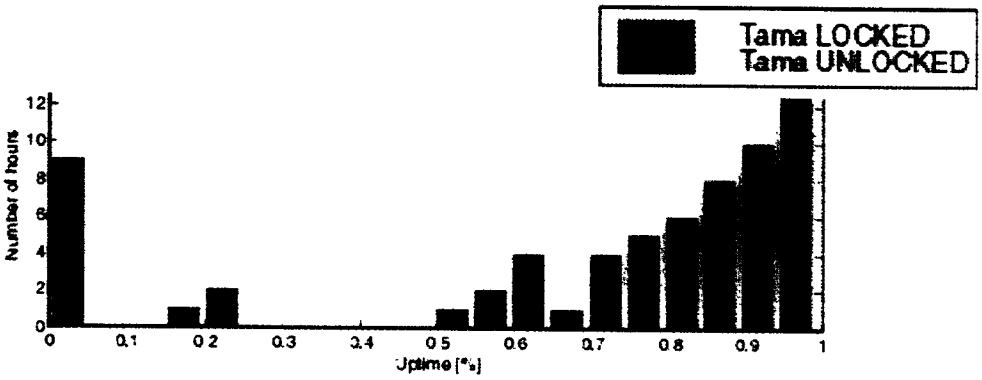
- LSC Software Coordinator
  - » Role identified in white paper, no candidate identified
  - » Interim role filled by committee appointed by LSC Spokesman, LIGO Director
    - – S. Anderson (Chair, LIGO Lab), L. S. Finn (PSU), M. A. Papa (GEO/AEI Potsdam) , T. Prince (LIGO Lab), A. Wiseman (UWM)
    - – Charter:
      - ♦ Review/approve LLAL S/W specification
      - ♦ Coordinate SW development schedule with Lab.
      - ♦ Define code acceptance, validation criteria
      - ♦ Coordinate code archive with Laboratory
- International data analysis working group
  - » Reports to GWIC regarding coordination of data analysis among major international programs
  - » Joint Lab, LSC representation
  - » 1st meeting scheduled for Rome on 4 Dec 1999

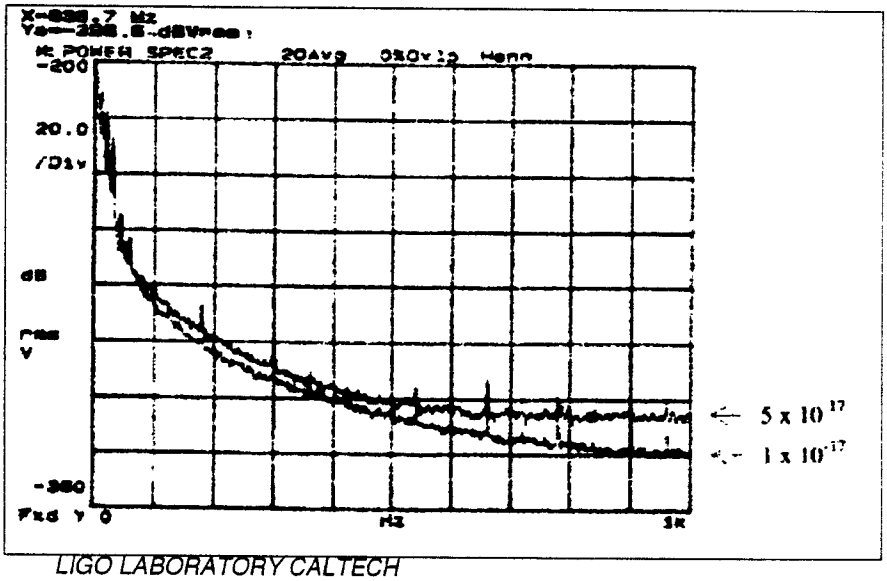
Laboratory - LSC Coordination

- Mock data challenges
  - » LSC-led effort to organize a simulated 2-interferometer analysis of 40m data (split run into two segments, shift second in time)
    - – Reprocess data set for binary chirps
    - – Develop event list
    - – Use LDAS metadata tables to log events, perform queries, sorts
    - – Pending announcement of proposed activity to LSC
      - ♦ Open to interested participants
  - » 40m + TAMA coincidence run
    - – 3-day run coincided with TAMA run;
    - – Coordinated at Caltech by University of Michigan and LIGO
      - ♦ Data transmission/storage to HPSS (900GB, 99.9+% availability; no lost data)

40m - TAMA Run Summary







CACR Collaboration with LIGO

- Mass data storage technology for LIGO I
  - » Use of HPSS for LIGO prototype tests
    - – Performance evaluation - bandwidth, reliability, sizing for LIGO I
    - – Direct (real time) archival of 40m datastream
    - – LSC access to existing 40m databases
  - » Use of existing HPSS infrastructure for LIGO I
    - – Space in CACR StorageTek robotic silo (100+TB, 2000 cassettes*)
    - – Site license
    - – Direct connection to Calren2, LIGO Caltech WAN
    - – Enhancements by LIGO Laboratory for LIGO/LSC specific use
      - ♦ Robotic arm upgrade (?)
      - ♦ 4 tape heads (StorageTek 9840 tape drives
      - ♦ SP2 nodes to control tape heads
      - ♦ High Perf. Gateway Node for communication
      - ♦ Intermediate disk cache (SSA, ~400+GB)
  - » 1 LIGO FTE to support LIGO HPSS operations

CACR Collaboration with LIGO

- Cluster computing
  - » Access to CACR expertise in prototyping beowulf (linux cluster) at LIGO
  - » Site of LIGO I production beowulf under review (CACR vs. LIGO Lab)

- Supercomputing
  - » LSC access to CACR supercomputers
  - » LIGO Laboratory provides interface for LSC members to CACR
    - – User accounts (includes HPSS databases)
    - – Block time allocation
- Software development of user (www)interfaces
  - » Prototype of GUIs to view frame, xml data
  - » Write data to Matlab files for analysis

LDAS Procurement Approach

- All items commercial, off-the-shelf
  - » Multiple vendor quotes, catalog items, etc.
- Flexible, extensible
- Tests on prototypes, loaners, etc.

- Networking/switching technology
  - » ATM (OC3/OC12) using LANE
    - – LAN
    - – Server-server connections
  - » Ethernet (100BT/1000BT)
    - – Server connections to LAN
    - – PC cluster interconnections

LDAS Procurement Approach

- CPUs & workstations
  - » SUN
    - – For licensed applications, databases
      - ♦ HPSS (v 4.2)
      - ♦ DB2 (IBM)
    - – LDAS APIs
  - » Intel Pentium
    - – Linux
    - – LDAS APIs

- Mass storage technology
  - » Robotic tape systems
    - – Large scale system for Caltech archive
      - ♦ HPSS (IBM or SUN)
      - ♦ 100+TB in campus StorageTek Silo
      - ♦ 4 9840 drives

LDAS Procurement Approach

- Mass storage technology
  - » Robotic tape systems
    - – Medium volume system at each site for near term look-back (~20TB)
      - ♦ StorageTek Timberwolf with 9840 drives compatible with HPSS at Caltech
    - – Small volume system for (few) tapes for experimenters at each site
      - ♦ AIT-2 (SONY/Cybernetics) technology
      - ♦ 10 - 20 cassette robot
  - » Disk systems
    - – Serverless RAID system (e.g., Network Appliances 760) for on-site data cache
      - ♦ ~1 TB for Hanford
      - ♦ Robust
      - ♦ High performance throughput
      - ♦ Does not need server
    - – SSA (IBM) intermediate disk cache for HPSS (400+ GB)