LIGO Data Analysis: Preparations for the LIGO I Science Run

Albert Lazzarini
LIGO Laboratory at Caltech

10th Meeting of the LIGO Laboratory PAC
14 June 2001
Pasadena, CA

LIGO-G010233-00-E

Outline for this talk

- LIGO Science -- Upper limit groups
- LIGO Data Analysis System (LDAS)
  - » Hardware deployment
  - » Software integration -- mock data challenges
- Collaboration with other projects
  - » Near term data exchange, network analysis
  - » Grid computing
- Simulation
- Demonstrations
  - » end-to-end simulation
  - » LDAS pipeline for inspiral search code

LIGO I
Science Run

- Strategy
  - » Use engineering runs to commission data analysis systems
    - – 10/01: first 3X coincidence run provides scientific first-look at , e.g., h_{rms}~10^{-20}
  - » Initiate science run when good, reliable coincidence data are available and straightforward sensitivity improvements have been implemented (~ 07/02)
  - » Interleave periods of science runs with periods of sensitivity improvements
- Goal:
  - » obtain 1 year of integrated data at $h_{rms} \sim 10^{-21}$
    - – searches in coincidence with astronomical observations (eg. supernovae, gamma ray bursts)
    - – searches for sources at known positions (eg. Spinning neutron stars)
    - – stand alone searches for compact binary coalescence, periodic sources, burst sources, stochastic background and unknown sources at h ~ 10^{-21} sensitivities

Near Term: Upper Limits Groups

- Do science with engineering run data as it will be done during the science run
- Working groups originated through Internal LSC Proposals
  - » Use and test LIGO Data Analysis System (LDAS)
    - – Each group has an LDAS integration phase built into its schedule
    - – Broad cross section of people (experimentalists, theorists)
  - » Organized around four separate searches:
    Inspiral, Stochastic Background, Bursts, Continuous Waves
    - – Currently driving motivation for code development within ASIS
  - » Detector Diagnostics
    - – Integrates existing detector characterization group to work on diagnostic triggers for the searches.
  - » Publications
    - – Each group has a paper as its deliverable

Inspiral Upper Limit Group

- Technical and scientific goals:
  - Develop techniques to place upper limits on inspiral rates for binary populations with accurately modeled waveforms, i.e. non-spinning binaries with total masses between 2 and 15 Msun
  - Develop data analysis techniques to search for high mass binary systems with black holes (M_{tot} > 15 Msun)
  - Goal: 2 week observation (10^6 sec) @ $h[f] \leq 10^{-21}$ 1/Sqrt[Hz]
    R < 0.01 /hour to d < 100 kpc ***(IDEAL)***
  - Best previous: R < 0.5 /hour (Allen et al.,PRL 83 (1999) 1498)
    R < 0.59/hour (TAMA gr-qc 0012010)

REALIZED!!! **Within the Milky Way**

- Group organization:
  - Initially 26 co-proposers on T0100025-00-Z has grown to about 35 members since work started in earnest (10 Lab personnel, remainder from 11 institutions). Chaired jointly by Brady and Gonzalez.

Inspiral
Technical Approach

- Data flow pipeline
  - » The GW channel from each interferometer will be analyzed to search for binary inspiral signals and candidate events entered into the database
  - » Triggers and events in auxiliary channels will be used to provide vetoes and/or identify bad data
  - » Coincidence between interferometers using events from the database. (Coincidence in time and other parameters)
  - » Follow-up with combined coherent analysis using multiple interferometers
- Simulation to determine efficiency
  - » Inject signals from target population into data stream and push through the pipeline
- Statistical Analysis
  - » Strict records of analysis will be kept and robust methods used

Stochastic Background Upper Limit Group

- Technical and scientific goals:
  - Develop techniques to place upper limits on logarithmic frequency derivative of GW energy density in the Universe
  - Develop techniques to cross-correlate interferometers with resonant bars
  - Goal: 2 week observation (10^6 sec) ; $h[f] \leq 10^{-21}$ 1/Sqrt[Hz] @ 200 Hz
    $\Omega < 50$ @ 900 Hz - LLO - ALLEGRO
    $\Omega < 0.002$ - LLO - LHO
    IDEAL
  - Best previous: $\Omega < 60$ [bar - bar, Astone et al.,PRL 83 (1999) 1498]
    REALIZED!!!!!

- Group organization:
  - 29 co-proposers on T010017-00-Z (6 Lab personnel, remainder from 13 other institutions). Chaired jointly by Romano and Fritschel.

Stochastic Background
Technical Approach

- Data flow pipeline
  - » The GW channel from each detector will be cross-correlated to search for steady-state coherent signals, cross-power spectra archived in the database
  - » Cross-correlation with auxiliary channels to identify bad data
  - » Bar-interferometer, interferometer- interferometer correlations will be studied
- Simulation to determine efficiency
  - » Detect injected simulated coherence between detectors
- Statistical Analysis
  - » Strict records of analysis will be kept and robust methods used

Bursts Upper Limit Group

- Technical and scientific goals:
  - Search for brief transient signals - unknown waveforms, not suited to matched filter techniques.
  - Use generic techniques: excess power statistic, slope filter, and generic templates (e.g., wavelets) not assumed to be "matched"
  - Include data from other gravitational wave detectors (including Allegro resonant bar), and also other astronomical data (such as supernovae and gamma ray bursts.)
  - ***Work naturally leads to better understanding of the instruments, by focusing on time-domain transient behavior***
  - Goal : R < ~ few/wk at h_{rms} > few * 10^{-20} ***(IDEAL)***
  - Best previous (bars, ICEG) :R < 4/yr at h_{rms} > 10^{-17} ***(REALIZED!!!!!)***
- Group organization:
  - 54 proposers on T010014-00-Z (21 Lab personnel, remainder from 12 institutions). Chaired jointly by Finn and Saulson

Bursts
Technical Approach

- Data flow pipeline
  - » The GW channel from each interferometer will be analyzed to search for burst signals and events entered into the database
  - » Triggers and events in auxiliary channels used to provide vetoes and/or identify bad data
  - » Look for coincident events between interferometers using the database. (Coincidence in time, consistency of other parameters)
  - » Combined likelihood analysis using multiple interferometers
- Simulation to determine detection efficiency
  - » Injected signals from target population into data stream and push through the pipeline
- Statistical Analysis
  - » Strict records of analysis will be kept and robust methods used

Continuous Wave Upper Limit Group

- Technical and scientific goals:
  - Develop techniques to place upper limits on CW signal sources
    - Large area searches at sub-optimal sensitivity
    - Directed searches using f and t domain searches
  - Goal: 2 week observation (10^6 sec) @ $h[f] \leq 10^{-21}$ 1/Sqrt[Hz]
    $h_{min} >$ few * 10^{-24} (directed) ***(IDEAL)***
  - Best previous: $h_{min} >$ few * 10^{-21} [GEO, PRD, 47 (1993) 3106]
    $h_{min} >$ few * 10^{-24} [Explorer, Papa et. al. gr-qc 0011072
    REALIZED!!!!!
- Group organization:
  - Initially 25 co-proposers on T010019-00-Z (4 Lab personnel, remainder from 6 other institutions). Chaired jointly by Anderson and Zucker.

Continuous Wave
Technical Approach

- Data flow pipeline
  - » The GW channel from each interferometer will be analyzed for steady-state coherent signals with modulation signatures consistent with fixed sky positions
    - – Short-transform all sky search (AEI Potsdam)
    - – Known radio pulsar search
    - – Semi-coherent wide area search with spindown
  - » Use other channels to seek out periodic artifacts
- Simulation to determine efficiency
  - » Test signal injection into interferometers
- Statistical Analysis
  - » Strict records of analysis will be kept and robust methods used

LIGO Data Analysis System

LDAS Hardware

- 3-phase deployment : “just in time procurement”
  - » Phase I - 1Q2001: Hardware needed to support MDCs, engineering runs -- RAID at observatories & universities, 3 small linux clusters -- COMPLETE
  - » Phase II - 3Q2001: Hardware will be installed immediately after the “E6” engineering run to establish scientific upper limits with LIGO data -- Large linux clusters, data movers for HPSS
  - » Phase III - 4Q2001: finalize hardware configuration approximately 6 months before the beginning of the LIGO I Science Run -- Tape drive wall for HPSS, tapes for science run, large disk farm at Caltech

Software Block Diagram

LIGO-T990101-E
LAL Specification

LIGO-T990097-E
wrapperAPI Requirements

www.lam.org
LAM version 5.6.1

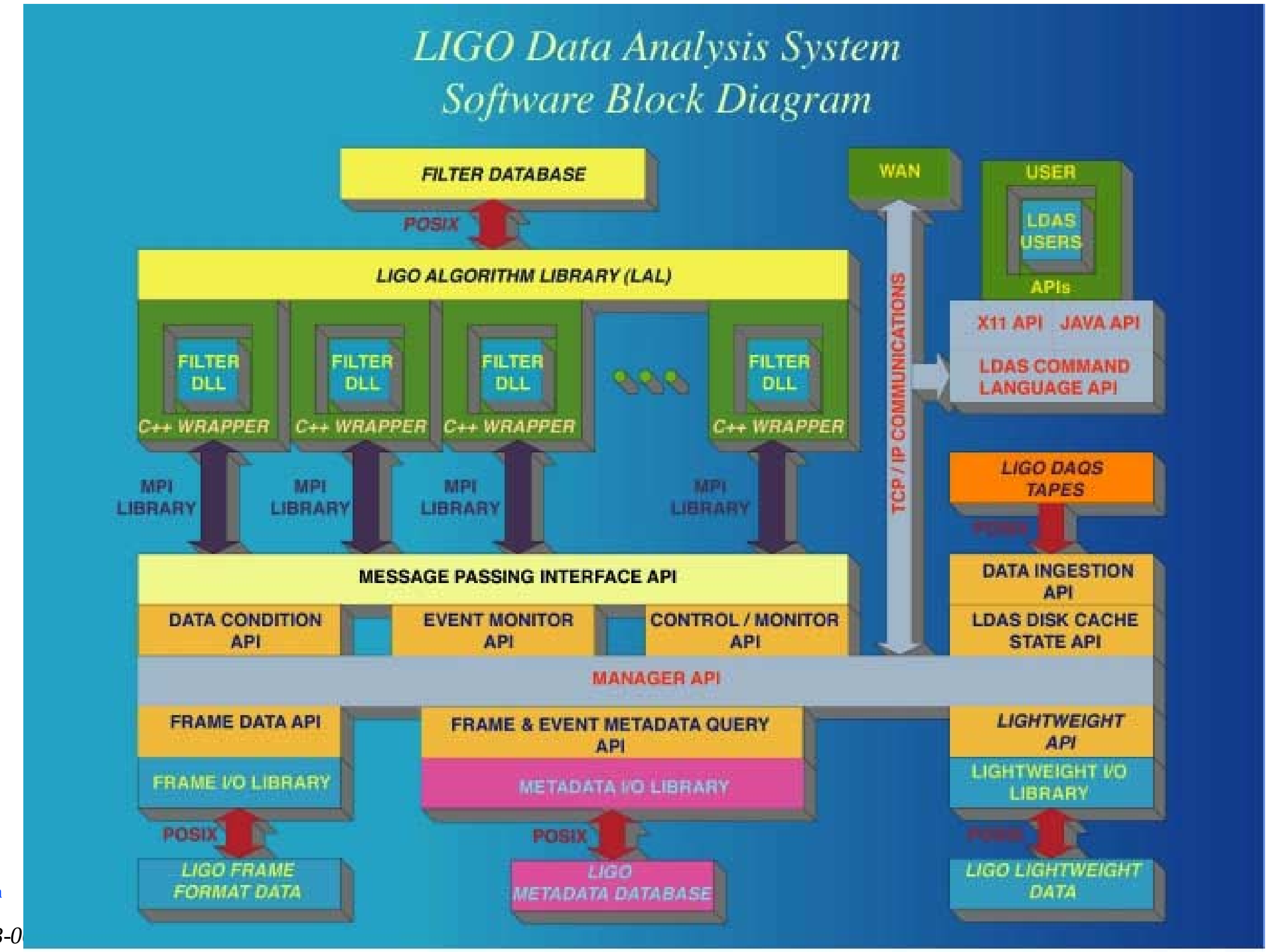
LIGO-T990086-E
mpiAPI Requirements

LIGO-T990002-E
dataConditionAPI Requirements
FFTW version 2.3.1
CLAPACK version

LIGO-T980115-E
managerAPI Requirements

LIGO-T980117-E
frameAPI Requirements

LIGO-T970130-E
LIGO-VIRGO Frame Specification
FrameCPP Version 4



GUILD
TCL/TK GUI

UNIX Sockets (TELNET)
User Commands (TCL)

LIGO-T010052-E
dataIngestionAPI Requirements
"In Development"

LIGO-T010051-E
diskCacheAPI Requirements
"In Development"

LIGO-T000026-E
controlMonitorAPI Requirements

LIGO-T990037-E
lightWeightAPI Requirements

www.apache.org
C++ XML Parser Library

LIGO-T980119-E
metaDataAPI Requirements

dbEasy Library
ODBC Level 3

LIGO-T990101-E
LDAS Database Tables

LIGO-T990023-E
LIGO-Lightweight Format

Software Deployment Schedule

- ***Mock Data Challenges form basis of LDAS Final Design Validation, Review***
- ***Organized as working groups within the Collaboration***
  - » ***COMPLETE:*** *Data Conditioning MDC (August 2000) Sam Finn*
  - » ***COMPLETE:*** *MPI MDC (January, 2001) Patrick Brady*
  - » ***COMPLETE:*** *Database MDC (January-April, 2001) Peter Shawhan*
  - » ***COMPLETE:*** *Inspiral Inchpebble (May 2001) Patrick Brady, Gabriela Gonzales*
  - » **Pending:** Joint MDCs (September 2001)
    - – Stochastic Inchpebble - Joe Romano, Peter Fritschel
    - – Bursts Inchpebble -- Sam Finn, Peter Saulson
    - – CW Inchpebble -- Stuart Anderson, Mike Zucker
  - » **Pending:** ***:*** LDAS Archive MDC (December 2001) Stuart Anderson
  - » **Pending:** Integrated Science MDC (Spring 2002)
- **Software releases**
  - » **Alpha - 0.0.XX (April 1999 - September 2001)** - *ONGOING*
  - » **Beta Release LDAS 0.1.0 (September 2001)** - after last inchpebble
  - » **First Official LDAS Release 1.0.0 (December 2001)** - after final MDC
- **Detector operations**
  - » ***Upper-Limit Engineering Run (October 2001)***
  - » First & Second Science Runs (First Half 2002)
  - » Start of LIGO I Science Run (July 2002)

LIGO-G010233-00-E

Data Condition MDC

- Tested:
  - » LDAS: CHEX, genericAPI, managerAPI, and dataConditionAPI
- Documented in “LDAS dataCondtionAPI: Mock Data Challenge #1” (185 pages!)
  - » LIGO DCC: LIGO-T000124-00-E
- Test Scripts stored in LDAS CVS Repository and are repeated weekly
- Paved the way for signal conditioning to be used in future MDCs:
  - » FFTs, Linear Filters, Mixers, Windowing, Power Spectral Densities, Cross Spectral Densities, Resampling, Statistics…
- Algorithm results verified against Matlab.
- Provided first opportunity to distribute LDAS usage model beyond LDAS and into the LSC community

MPI MDC

- Tested:
  - » LDAS: managerAPI, mpiAPI, wrapperAPI, and trivial eventMonitorAPI
  - » LSC: LAL, LALwrapper, inspiral dynamic shared object (dso), and power dso
- Documented in "MPI Mock Data Challenge" (117 pages!)
  - » LIGO DCC: LIGO-T010024-00-Z
- Test Scripts stored in LAL CVS Repository and repeated weekly
- Paved the way for inchpebble search codes used in future MDCs:
  - » Binary Inspiral, stochastic, burst, and periodic sources
- Test waveforms injected "blindly" into 75 minute data streams and found!
- Distribute parallel analysis model beyond LDAS and into the LSC community

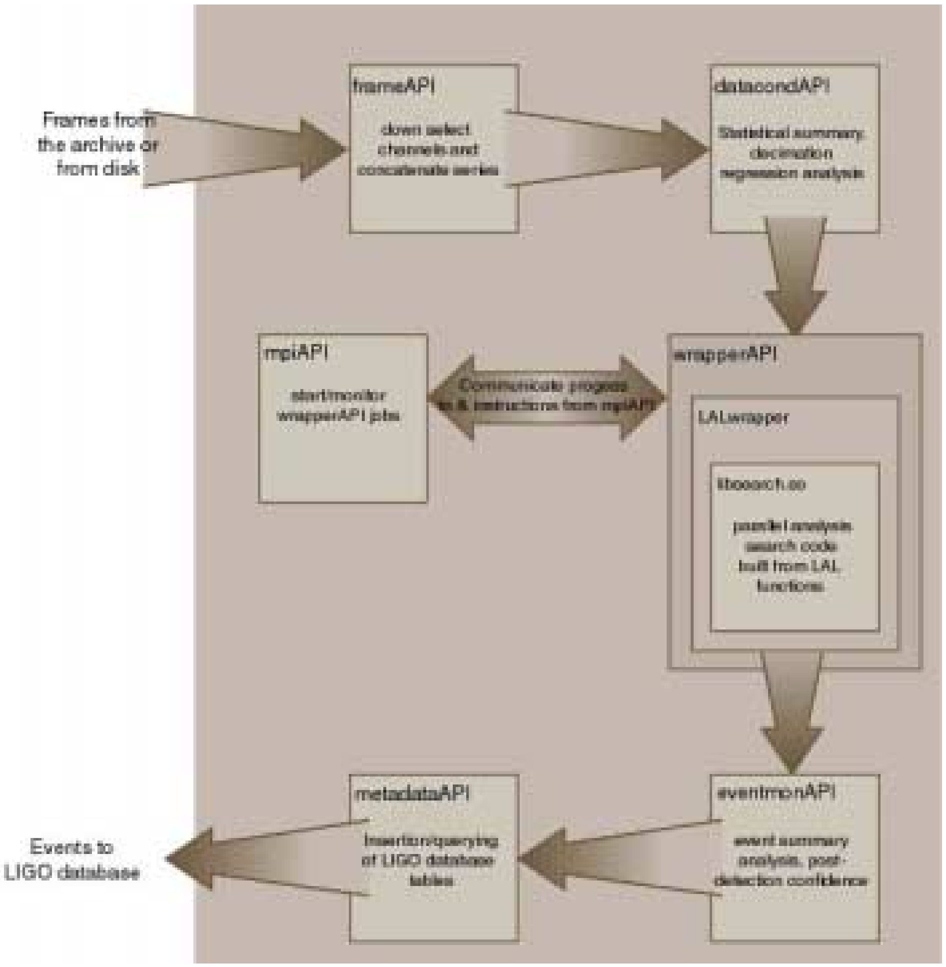
Database MDC

- Tested:
  - » LDAS: managerAPI, metaDataAPI, lightWeightAPI, Table Design, and GUILD
- Documented in "Plan for the LDAS Database Mock Data Challenge" (16 pages)
  - » LIGO DCC: LIGO-T000089-00-E
- Paved the way for inchpebble MDCs and Engineering Runs:
- Provided needed feedback in table schema to better support LIGO analyses.

Inspiral

MDC “Inchpebble”

- Completed May 16-18 2001:
  - » To implement and test a straight thru inspiral search pipeline within the LDAS environment
- Test definition:
  - » Exercise LDAS, LAL and LALWrapper
  - » Exercise search pipeline
    - – Including blind search for simulated events in data.
- Conclusions:
  - » *Templated and FCT search codes run in LDAS. Simulation package is ready in LAL*
  - » Blind search postponed



LIGO-G010233-00-E

Inspiral
Second Milestone: Full Integration

- Plan to get there
  - » Gain experience with instrument characterization tools (cross-upper-limit group task and team) using E2, E3, E4 data
  - » Software related issues:
    - – Blind signal detection in MDC was delayed due to inconsistent normalizations, units and conventions.   Being addressed by MPI, DataConditiong, LAL developers, and LDAS developers.
  - » Post-processing issues:
    - – Visualization tools for database extracted information
    - – Simulation tools improved and verified
  - » Develop science verification tests
    - – Blind signal injection/recovery

August 2001
Second Milestone

Stochastic Background

MDC Milestone

- September 2001:
  - » Implement and test a straight through cross-correlation pipeline within the LDAS environment
- Test definition:
  - » Exercise LDAS, LAL and LALWrapper
  - » Exercise search pipeline
    - – Search for simulated coherent signals in data
  - » Exercise DB insertion/retrieval
    - – Cross-spectra to be archived for easy access

Bursts
MDC Milestone

- September 2001:
  - » Implement and test a straight through burst search pipeline within the LDAS environment
- Test definition:
  - » Exercise LDAS, LAL and LALWrapper
  - » Exercise search pipeline
    - – Blind search for simulated transient signals in data
  - » Exercise DB insertion/retrieval
    - – Coordination with on-line, real-time diagnostics systems

Continuous Wave
MDC Milestone

- September 2001:
  - » Implement and test a straight through CW search pipeline within the LDAS environment
- Test definition:
  - » Exercise LDAS, LAL and LALWrapper
  - » Exercise search pipeline
    - – Blind search for simulated CW signals in data

Participation in other Collaborations

Data exchanges - Virgo/GEO/LIGO

Grid computing - GriPhyN, Virgo

LIGO-Virgo data exchange

- Background
  - » LIGO-Virgo meeting on the concept [2/10/2000]
  - » LIGO-Virgo-GEO meeting focused on PEM details [2/11/2001]
- PEM channels
  - » Sites involved: Hannover [GEO], Cascina [VIRGO], Hanford & Livingston [LIGO]
  - » Seismometers: 1/site, 256 Hz sampling, 1 kB/sec
  - » Magnetometers: 4096 Hz sampling, 8 kB /sec [compressed!]
- Data rates necessarily limited
  - » Start by June 1st 2001, exchanging few tens of kB

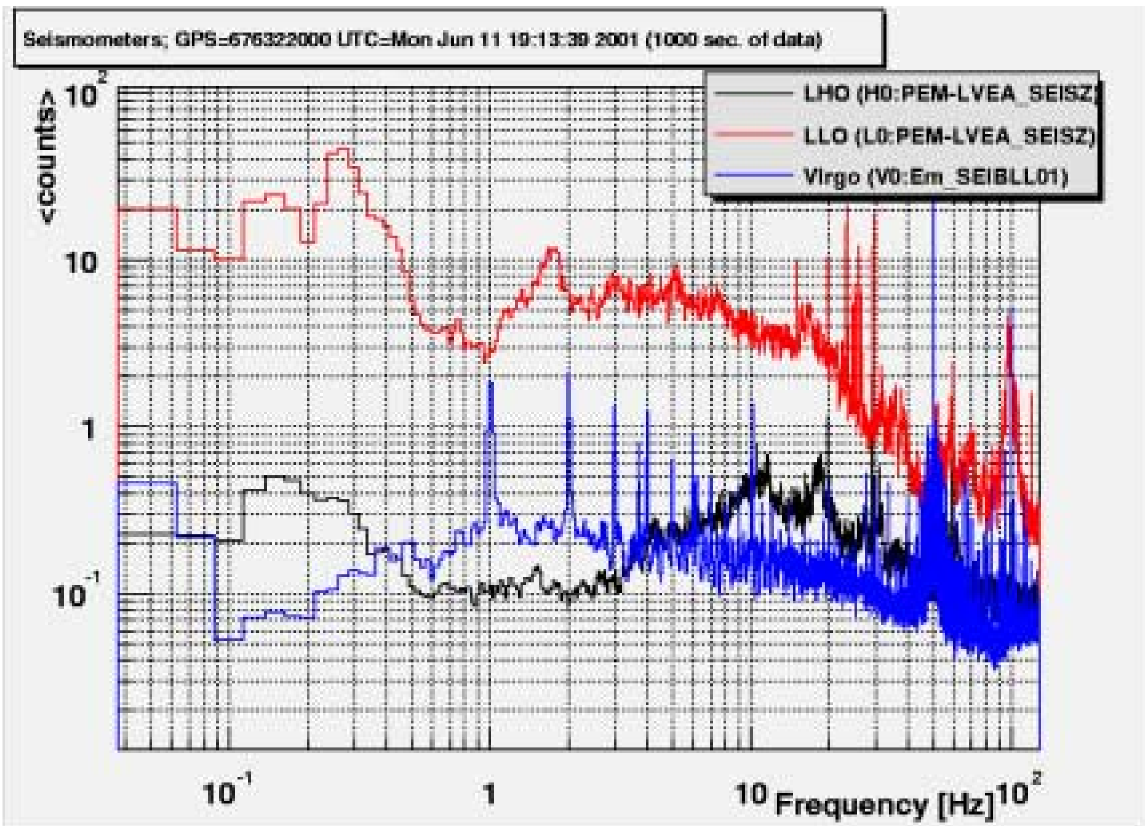
Motivations

- In the long term, we expect the observatories to do collaborative data analysis at different levels and to cooperate as a network of detectors.
- This goal requires several steps: we focus here on some of the technical ones, related to data exchange.
- Lots of data are already being acquired: in particular, environmental monitoring data.
- Learning to exchange and analyze these data shall give us the experience needed for the future physics data.

Science issues

- PEM data already contain interesting physics!
- Teleseismic events can correlate on the LIGO-VIRGO distance
  - » Events originating close to the equidistant line can correlate within the narrow (+/- 40 ms) window common with GW detection
- Electromagnetic events propagate over long distances in the ionosphere
  - » In particular the interaction of lightning and the magnetosphere causes VLF EM signals potentially affecting LIGO/VIRGO band as whistlers
- Coincidences in these data can serve in the future as vetoes for the network analysis

Example of analysis using multiple data sources



- Seismometers
- An example of what we might want to monitor continuously

Benoit Mours (Virgo)

Szabolcs Marka (LIGO)

Grid-enabled LIGO Data Analysis

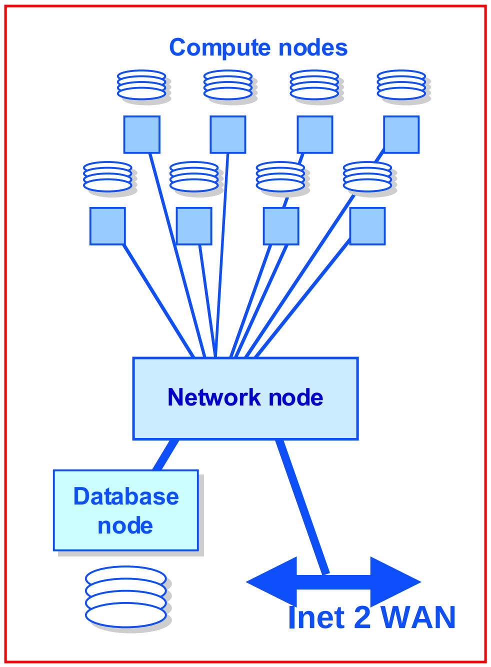
- Virtual Data
  - » Definition of a language and architecture to query LIGO databases
  - » Retrieve or generate data products transparently to user
  - » GUI interface to explore databases remotely
    - – e.g. Google for LIGO data: "LIGOVista"
- Data Replication
  - » Real-time international mirror (Virgo/GEO/LIGO)
  - » Fault-tolerant replica of LIGO data at LSC Tier 2 centers (UWM, PSU)
- Pulsar all-sky search
  - » Stage analyses requiring "infinite computing resources" onto other, e.g., NSF (but non-LIGO or LSC) resources

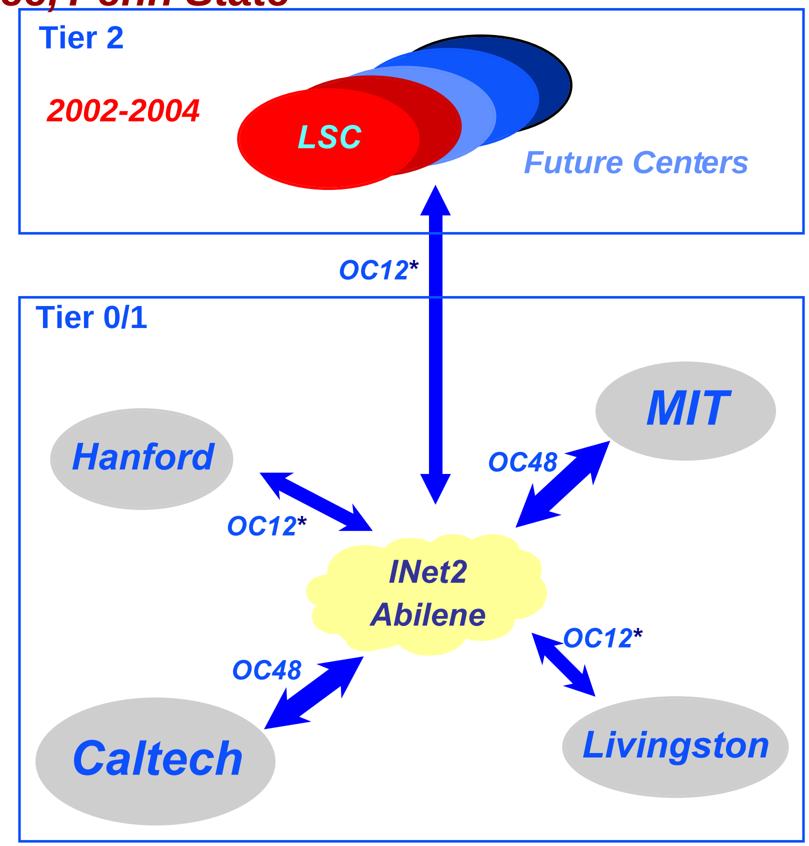
GriPhyN Tier 2 Hardware

Scenario for LSC Deployment

First two T2 centers: Milwaukee, Penn State

Grid Node:
N compute + Database + Network





* Bandwidths reflect projected capacities in ~ 4+ yrs

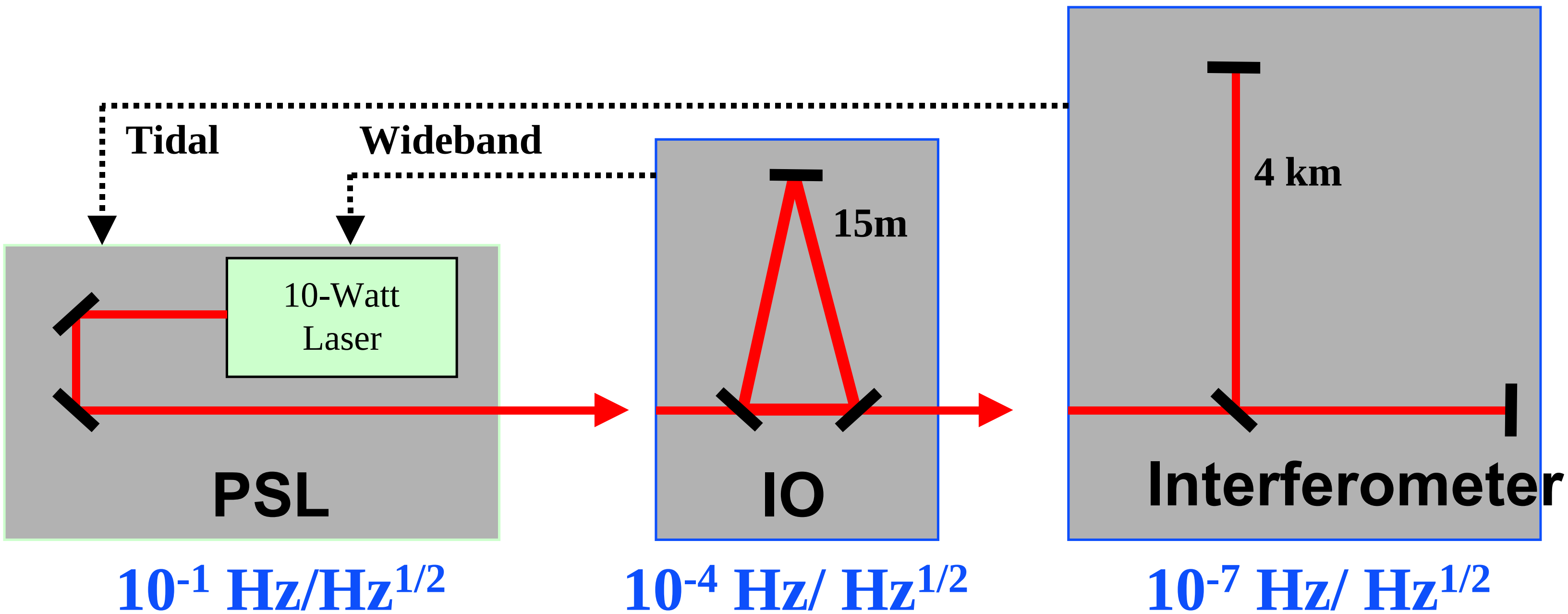
End-to-end simulation of LIGO Interferometers

LIGO-G010233-00-E

Simulation of the laser stabilization subsystem

- Deliver pre-stabilized laser light to the 15-m mode cleaner
  - Frequency fluctuations
  - In-band power fluctuations
  - Power fluctuations at 25 MHz

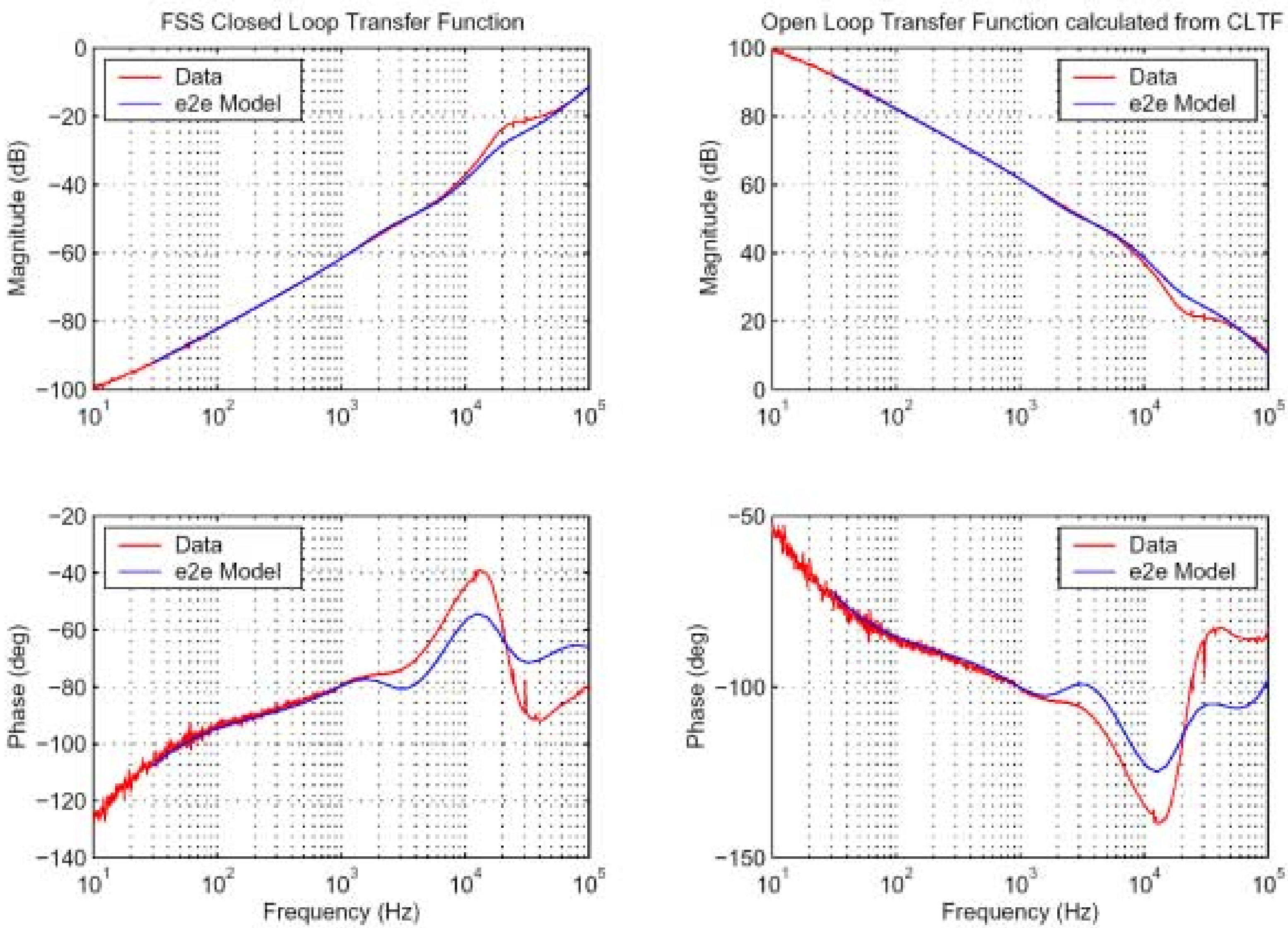
- **Provide actuator inputs for further stabilization**
  - **Wideband**
  - **Tidal**



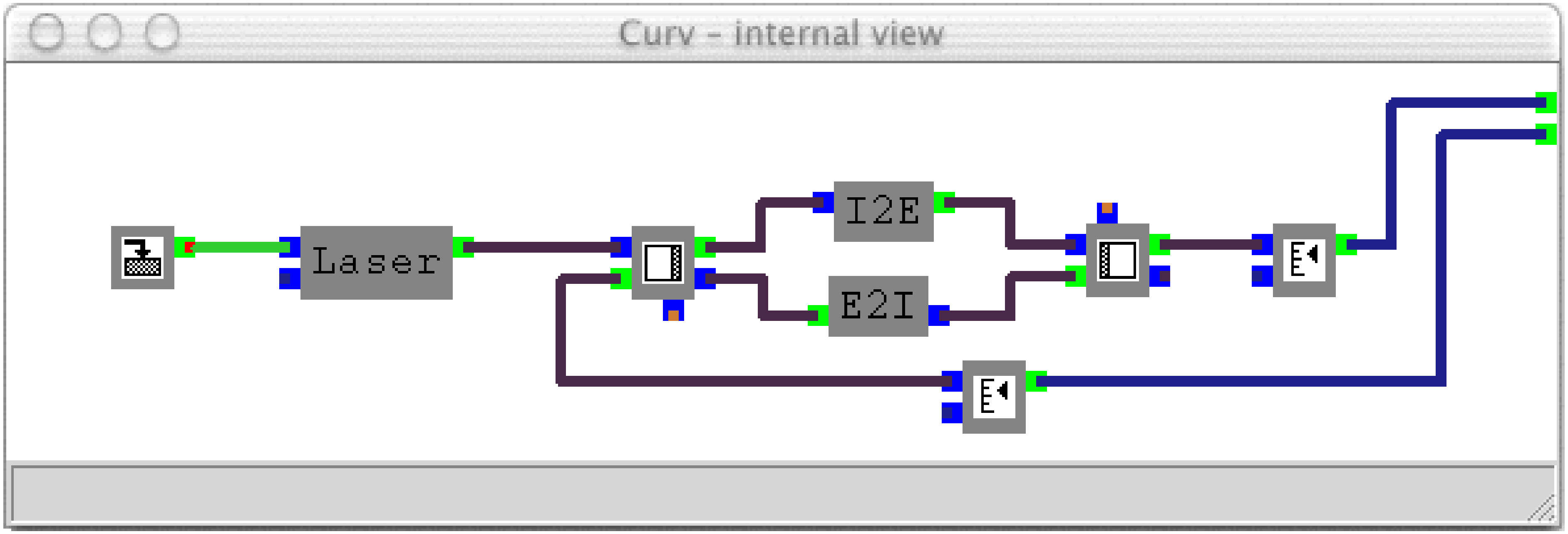
LIGO-G010233-00-E

Pre-stabilized Laser

laboratory data vs e2e simulations



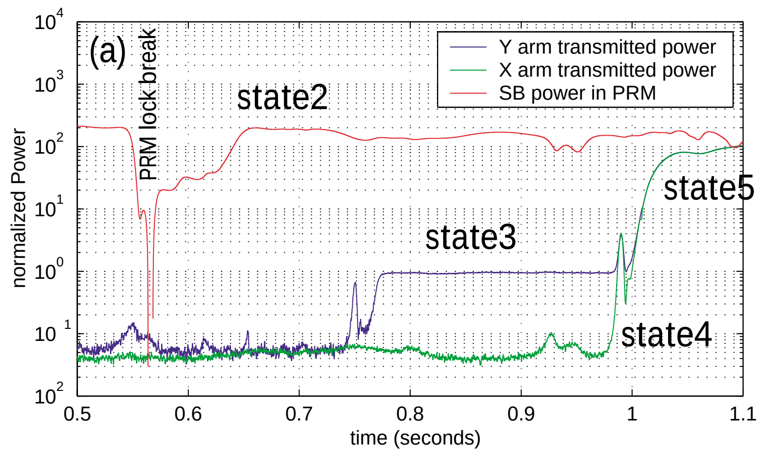
Graphical User Interface to Simulation Environment



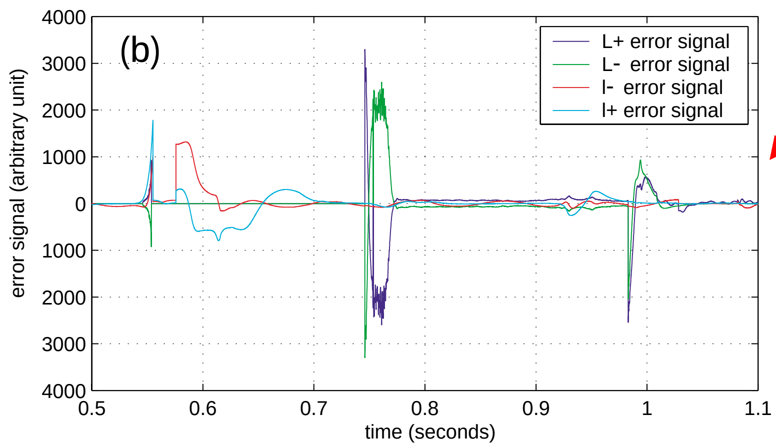
Lock acquisition of 2km at LHO

Comparison of real and simulation results

Figure 1. LHO 2k IFO data

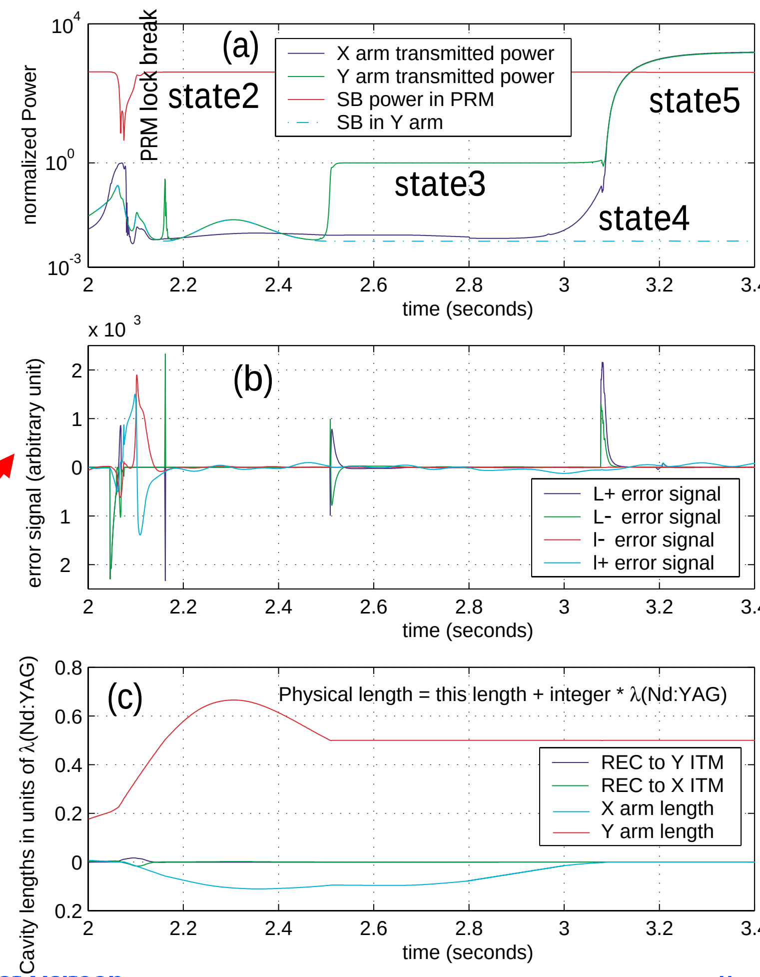


Arm powers are normalized by the power when one arm is locked.
SB power is normalized by the input SB power.



observable

Figure 2. Simulated signal



Not experimentally observable

Demonstrations of LIGO software

- Modeling of LIGO 2km interferometer at Hanford

- Use of the LDAS pipeline to perform an inspiral coalescence search

LIGO-G010233-00-E