LIGO Science and GriPhyN

Albert Lazzarini
LIGO Laboratory, Caltech
lazz@ligo.caltech.edu

GriPhyN NSF Project Review
29-30 January 2003
Chicago

LIGO- G030002-02-E

LIGO-GriPhyN Collaborators

- Caltech -- LIGO applications
  - Staurt Anderson (LIGO)
  - Masha Barnes (LIGO)
  - Kent Blackburn (LIGO)
  - Philip Ehrens (LIGO)
  - Albert Lazzarini (LIGO)
  - Greg Mendell (LIGO-Hanford)
  - Peter Shawhan (LIGO)
  - ***Roy Williams* (CACR)***
- UW Milwaukee/LIGO -- LIGO applications
  - Bruce Allen
  - ***Scott Koranda (also NCSA)***
- GEO6000 - AEI/Berlin LIGO applications
  - Maria Alessandra Papa
- Balearic Islands University Spain - LIGO applications
  - Alicia Sintes
- UT Brownsville/LIGO -- Outreach
  - Joe Romano
  - ***Manuella Campanelli***
- ISI/USC -- CS development
  - ***James Blythe***
  - ***Ewa Deelman***
  - ***Yolanda Gil***
  - ***Carl Kesselman***
  - ***Gaurang Mehta***
  - ***Gurmeet Singh***
  - ***Karan Vahi***

GriPhyN supported FTEs
Off-project

LIGO: Laser Interferometric Gravitational-wave Observatory

Mission: Gravitational Wave Searches ($h_{rms} \sim 10^{-21}$)

- Merger of neutron star, black hole (NS-NS, NS-BH, BH-BH) systems
  - Signal is minutes-long and characteristic in shape
- BH birth from supernova, starquakes in NS
  - Short, unknown signal profile
- Periodic sources (from rotating asymmetric compact stars -- NS)
  - Very faint, requires highly directed search
- Primordial gravitational wave background
  - GW analog of the cosmic microwave background
  - Requires cross-correlation between pairs of detectors

LIGO Data Growth

- LIGO I Engineering Runs: since 1999 and continuing
  - 35 TB and growing
- LIGO I Science Runs: interspersed with engineering runs
  - 2002
    - S1: 17 TB (raw) (Aug-Sep 2002) ***<- COMPLETED***
  - 2003 - 2004
    - S2: 45 TB (raw) (Feb - Apr 2003) ***<- NEXT MONTH***!
    - S3: 135 TB (raw) (Nov 2003 - May 2004) ***<- THIS FALL***
  - 2005 - 2007
    - Nearly continuous operation @ 270 TB/yr
- Advanced LIGO Upgrade: operational by 2009
  - 1-2 PB/yr

Motivations, Goals, Requirements
Example data processing challenges where grids can enable LIGO science

- Blind all-sky search for periodic sources
  - Transform data for every sky direction, frequency, d^n/dt^n [frequency], (n=1,2,3..)
  - >>Petaflop problem if the full scientific content of the data is to be exploited - we do not know how to do this
    - -> suboptimal methods deployed on large scale clusters using grid-enabled interfaces
- Global gravitational wave detector array
  - Establish a network of interferometers
  - Coincidence analysis using a phased array and coherent signal processing
    - Phased array introduces new parameters into the analysis -- increases dimensionality of the search 10X - 100X
    - Sky position sensitivity to searches
    - Wave polarization
    - Cross-spectral correlations of noise
  - Virgo (France/Italy), GEO (Germany/UK), TAMA (Japan)

Motivations, ***Goals***, Requirements

Services from a Grid

- Distributed Computing Power
  - Dedicated "burst" use of extensive resources
  - Enable CPU-limited analyses as background jobs
    - "GW-Search@GriPhyN" project
  - Challenge: making analysis codes portable within a grid environment
    - kernel dependence, compiler dependence, …

- Virtual Data
  - Tracking data through efficient catalogs
  - Accessing reduced data, data replication, data mirrors
  - Data discovery, data transformations

Motivations, Goals, ***Requirements***

Integration with LDAS

- For LIGO I, the infrastructure code is 95%+ complete. Current activities:
  - Maintenance, debugging
  - Upgrades
  - Optimization
- Integration of infrastructure code with grid must be at the API level

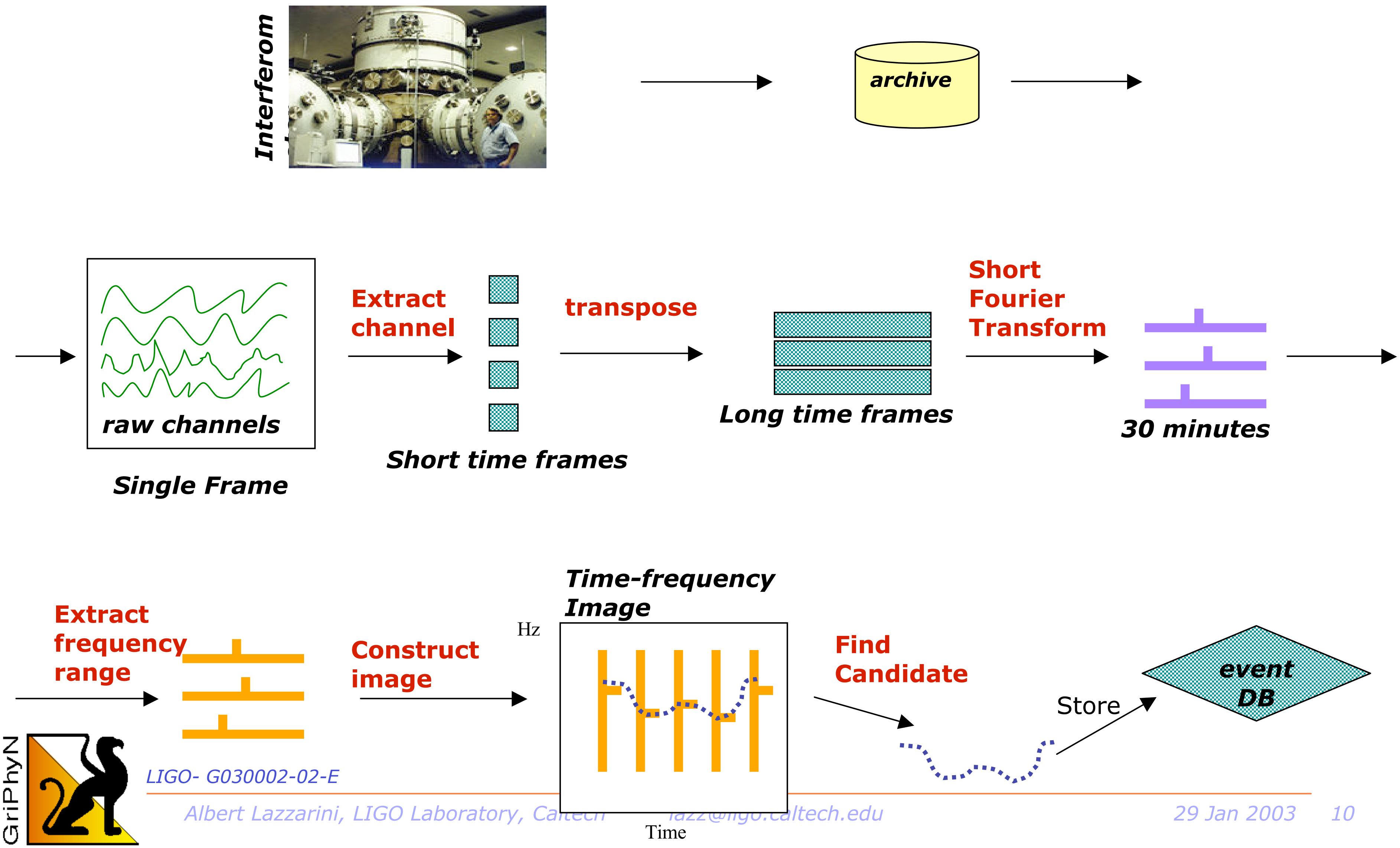
Approach and CS / Experiment Engagement

- LIGO Scientific Collaboration
  - New API development, integration with existing system (LDAS)
  - Identify, specify, build a suitable set of APIs to stage LIGO data analysis on remote grid cluster using grid-enabled protocols for data management, resource access, etc.
  - Specify, build an automated data transfer, data discovery utility (LIGO Data Replicator, LDR) for staging data from Tier 1 to Tier 2 centers
- ISI/USC
  - Build virtual data cataloging utilities (Chimera)
  - Build job workflow manager (Pegasus, DAGMan)
    - abstract workflow in scientific context -> concrete DAG

Progress During Years 1 & 2

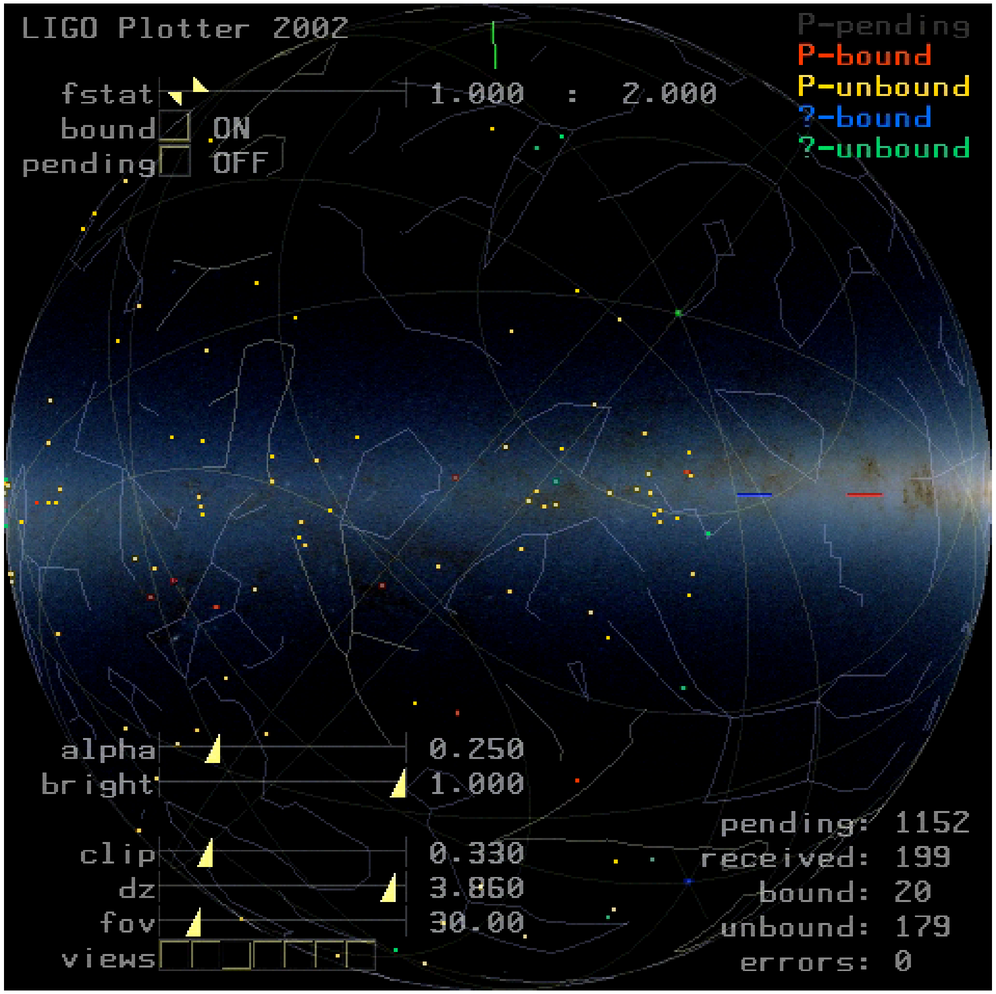
- Plan for Year 1: Virtual data model for LIGO
  - Information model for LIGO data cataloging approach
  - Access model for LIGO data across the grid
- Accomplished in Year 1:
  - ✓ Developed transaction catalog design, replica catalog
  - ✓ Virtual data language to represent LIGO transformations
  - ✓ Extracted data subsets from LIGO archive using Globus toolkit interface to LIGO data environment
  - ✓ Culminated in SC01 demonstration/prototype
- Plan for Year 2: Data streaming through caches
  - Replicate LIGO data subsets across the grid automatically
- Accomplished in Year 2:
  - ✓ LDR: LIGO Data Replicator to mirror data between sites used after S1 to transfer data from Caltech (Tier 1) to UW-Milwaukee (Tier 2) -- built on Globus Replica Location Service (RLS), GridFTP, GRAM, and pyGlobus(LBNL)
  - ✓ Demonstrated succcesful staging of LIGO (LDAS) data analysis tasks onto grid resources using Chimera, Pegasus, and Globus toolkit
    - ✓ Originally planned to commence in Year 3
  - ✓ Culminated in SC02 demonstration/prototype

Search for periodic sources: SC02



Prototype directed periodic source search at SC 2002

- ***The pulsar search conducted at SC 2002***
  - Used LIGO data collected during the first scientific run of the instrument
  - Targeted a set of 1000 locations of known pulsar as well as random locations in the sky
  - Results of the analysis are available via LDAS (LIGO Data Analysis System) DB queries
  - performed using LDAS and compute and storage resources at Caltech, University of Southern California, University of Wisconsin Milwaukee.
- ***During SC 2002 demo***
  - Over 58 directions in the sky searched
  - Total of
    - 330 tasks
    - 469 data transfers
    - 330 output files produced.
  - The total runtime:11.4 hr
- ***To date***
  - 185 directions in the sky searched
  - Total of
    - 975 tasks
    - 1365 data transfers
    - 975 output files
  - Total runtime: 96.8 hr



Plans

1. Take SC02 prototype and develop a robust utility for LIGO scientists
   - Periodic sources search needs to be rendered useful for a larger number of scientists.
   - A second class of search has immediate need for this
     - > Search for a gravitational wave stochastic background
       - Many autonomous jobs need to be run over the same data to obtain on-source/off-source analyses using time shift and FFT techniques.

2. Additional API development is needed to enable a better (i.e., more seamless) integration between LDAS (LIGO Data Analysis System) and the Globus Toolkit.
   - The specific area of effort that will be targeted is interoperation of the authentication schemas

3. Extend the LIGO Data Replication (LDR) service to iVDGL-based interchange of data among international partners
   - Work with GEO, possibly Virgo to implement LDR on both sides
     - > GEO is committed to EU grid participation
     - > Virgo may cease grid integration activities due to manpower shortage

Issue - people, people, people

- Extremely demanding and difficult on LIGO Laboratory and Collaboration partners to be busy developing at the beta level of the analysis infrastructure (LDAS) and scientific search codes (LAL -- LIGO Algorithm Library) while we are also in the midst of integrating LDAS with grid tools for the purpose conducting prototype tests and building up collaboration Tier II centers

For More Information

- LIGO
  - http://www.ligo.caltech.edu
    - LIGO Laboratory home page
  - http://www.ligo.org
    - Collaboration home page
  - http://www.ldas-sw.ligo.caltech.edu
    - LIGO Data and Computing Group home page