LIGO Scientific Collaboration
ITR 2003 Advisory Committee Meeting

K. Blackburn[1], P. Brady[2], S. Finn[3], E. Katsavounidis[1*],
S. Koranda[2], A. Lazzarini[1],A. Wiseman[2], J. Zweizig[1]

[1] LIGO Laboratory,[2] Penn State University
[3] University of Wisconsin, Milwaukee
*presenting

Argonne National Laboratory
May 10, 2004

Observing the Universe

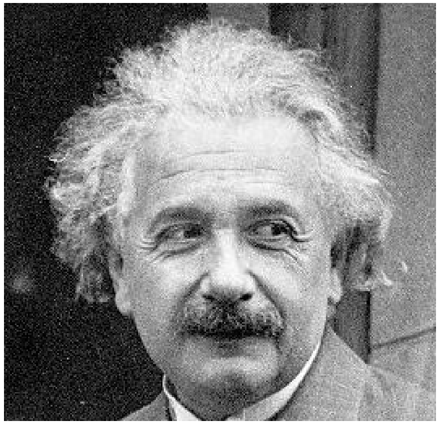
- Electromagnetic waves
- Particles: neutrinos, cosmic rays
- Gravitational waves

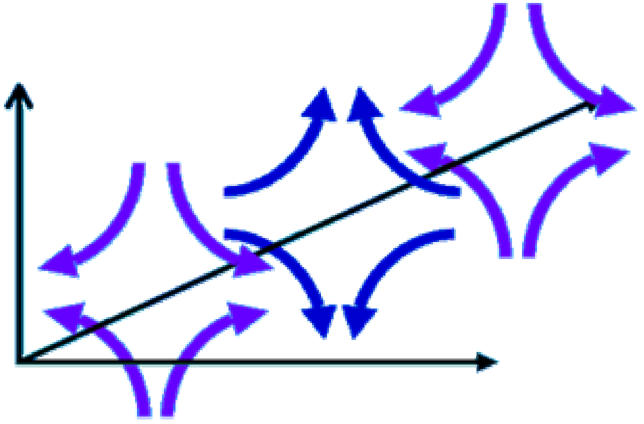
Gravitational Waves

- Distortions arising as a consequence of Einstein’s general relativity

$$\nabla^2\psi = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}$$

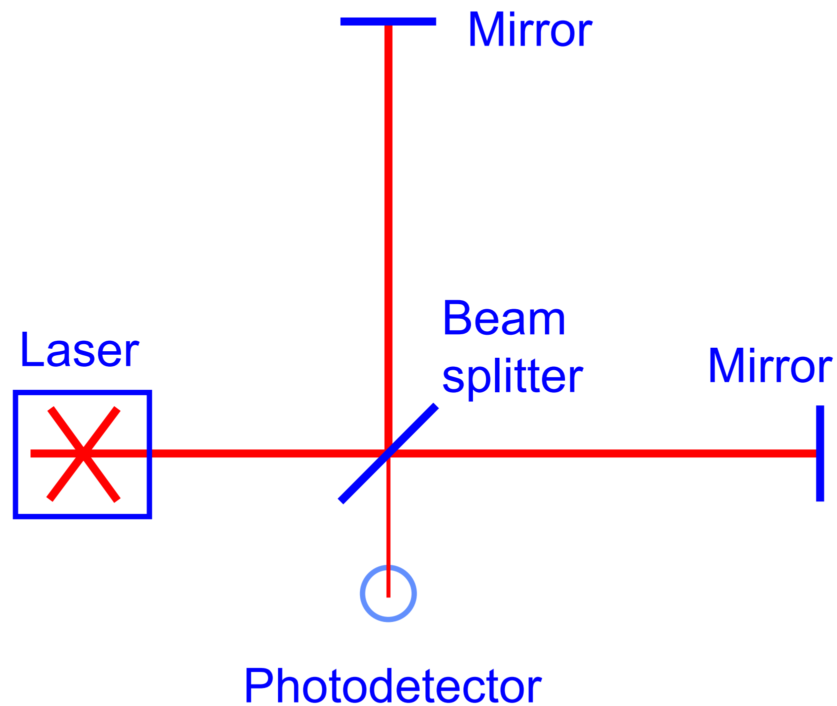
- They are emitted when matter moves or changes its configuration
- Waves propagate at the speed of light
- They distort space itself: stretching one direction and squeezing the perpendicular in the first half period and vice versa in the second half





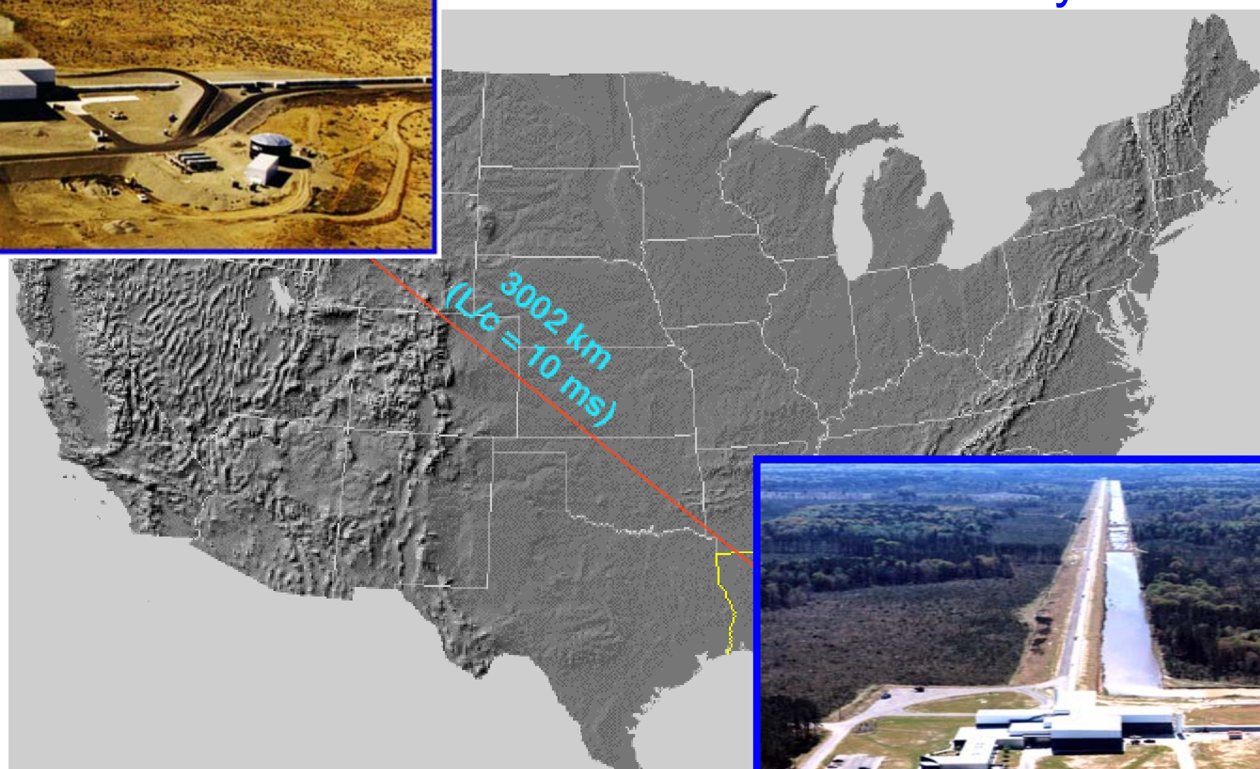
The challenge for LIGO

- Michelson interferometer (1881): the intensity of light at the photodetector depends on the length difference between the two perpendicular arms
- As gravitational waves pass, they change the distance between the mirrors.
- Expected signals:
  - » Rare, need to look a large volume of the universe
  - » Weak, need to be able to see typical signal at earth at the $h=\delta L/L \sim 10^{-21}$!!



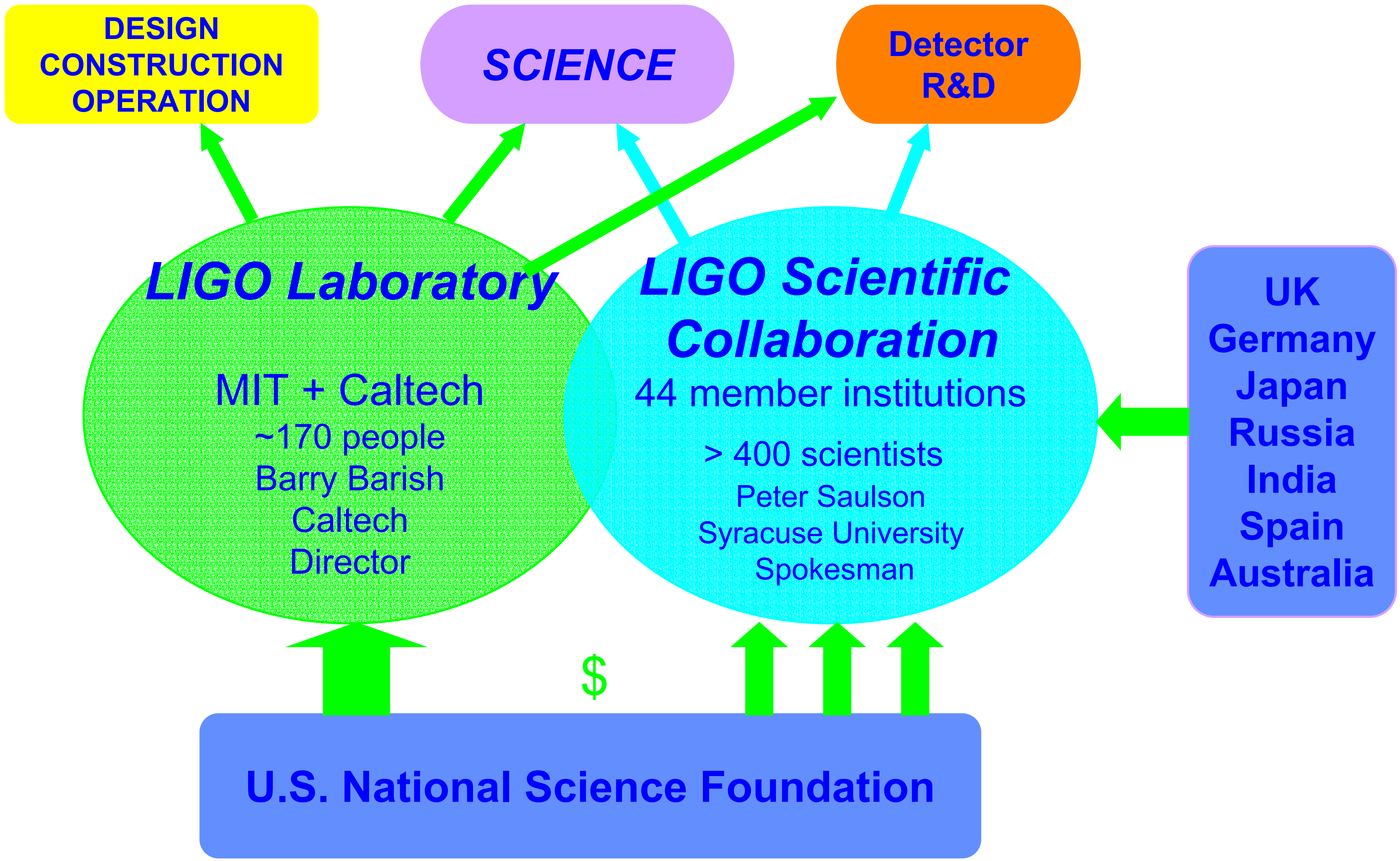
Laser Interferometer Gravitational wave Observatory (LIGO)

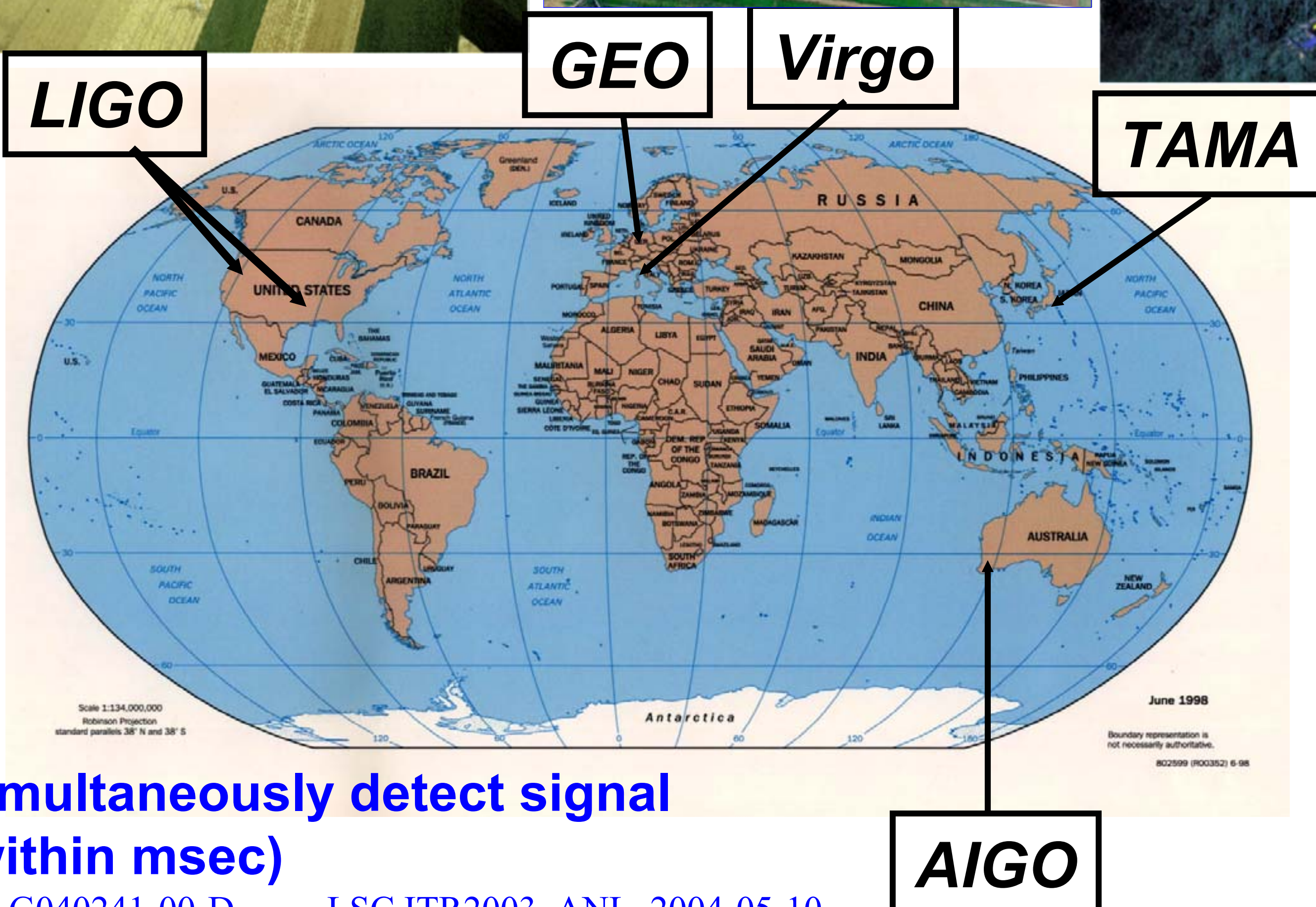
LIGO Hanford Observatory



LIGO Livingston Observatory

LIGO Organization & Support





detection confidence

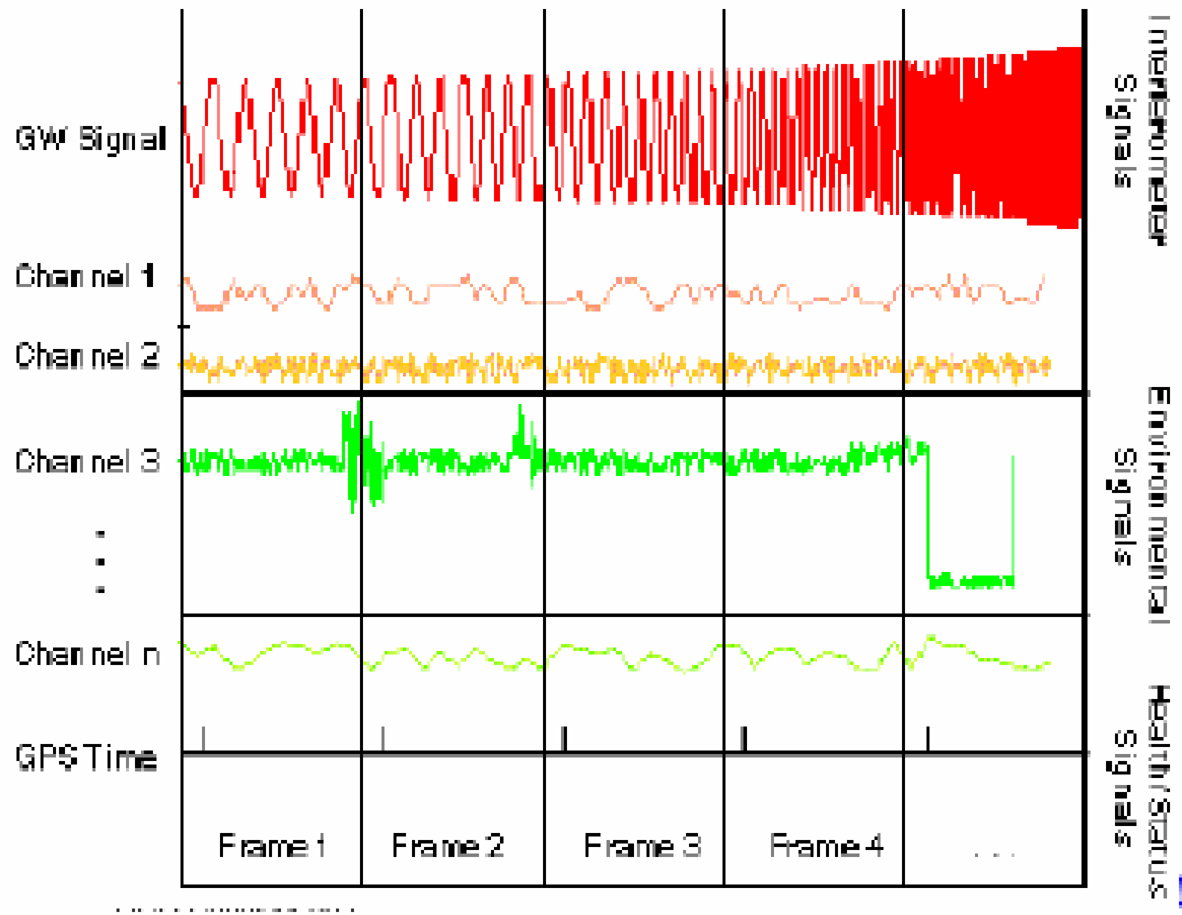
locate the sources

decompose the polarization of gravitational waves

Simultaneously detect signal (within msec)

What are the LIGO Data?

- Continuous time series (1Hz, 128Hz … 16kHz), all written in an internationally agreed data format: frames
- Data are comprised of:
  - » Gravitational Wave channel: 96 KB/s (16 bits x 16KS/s x 3 IFOs)
  - » Physical Environment Monitors (seismometers, accelerometers, magnetometers, microphones etc): ~ 1.2 MB/s
  - » Internal Engineering Monitors (sensing, housekeeping, status etc): ~7.2 MB/s



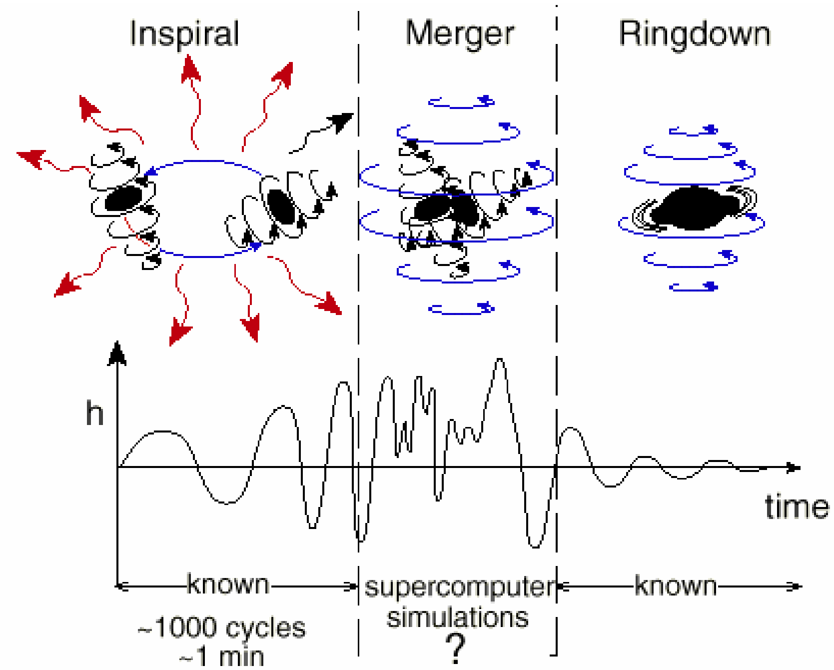
Data Issues

- Multiple data products beyond raw data
  - » Environmental vetoes, astrophysical events, summaries, channel trend files, spectra, links to metadata tables
- LIGO writes **1TB/day** when in 24/7 operation
- Data organization needs to:
  - » Archive and distribute data promptly and efficiently
  - » Guarantee security and integrity
  - » Be transparent among sites, among users

- **These challenges are well suited for the Grid paradigm.**

Realizing the science potential of LIGO

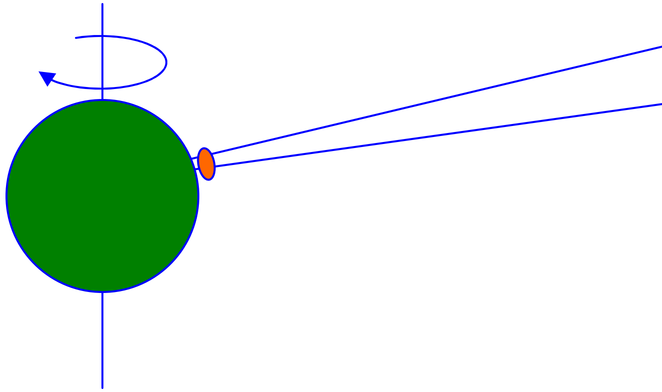
- Inspiral of compact binary systems (P. Brady-UWM, G. Gonzalez-LSU)
  - Two neutron stars, two black holes, or one of each
  - Template-based optimal match filtering
  - Templates depend on masses and spins of the binary members as well as on their eccentricity and orientation; more than **10^5 templates** are needed thus making it a **x100GFlop computation problem**



- Stochastic background (J. Romano-Cardiff, P. Fritschel-MIT)
  - » Random type of radiation described by its spectrum
  - » Use optimally filtered cross-correlation of detector pairs: look for correlated noise, **few GFlops problem**

LIGO Realizing the science potential of LIGO

- Burst sources (E. Katsavounidis-MIT, S. Whitcomb-Caltech)
  - » Short duration transients, inherently powerful, accompanying cosmic catastrophes
  - » Generic time-frequency search methods, **few GFlops problem**
- Periodic waves (M. Landry-LHO, M. Papa-AEI)
  - » Pulsars: steady emitters of electromagnetic waves (radio!)
  - » Signals from such stars depend on sky location (Doppler shifts) and their orientation
  - » An all-sky blind search is **1000GFlop+ computing problem**
- Revealing the full science content of LIGO data is a computationally and data intensive challenge
- **These analyses are ideally suited for distributed grid-based computing**



Summary

- The gravitational wave sky is unkown; sources are expected to be weak and rare
- Optimal signal processing is needed over a huge parameter space
- Mining the full science content of LIGO data is a computationally and data intensive challenge

Analyzing the LIGO data

When will you start searching for these waves?

LIGO Data Taking

- LIGO construction is **completed**
- Interleaved commissioning and engineering/science runs started in 2000 and generated over **200TB** of data
- Science runs
  - » S1 : August 23 – September 9, 2002 (17 days) – analysis complete
  - » S2 : February 14 – April 14, 2003 (59 days) – analysis in progress
  - » S3 : October 31, 2003 – January 9, 2004 (70 days)
  - » S4 : Planned to begin in late 2004
- 24/7 operations expected to commence in 2005, i.e., a steady state **1TB/day** data and analysis challenge is imminent
- Detector commissioning and characterization continuous
- **LIGO's data challenge is NOT an issue of tomorrow, it is a problem of TODAY**

Analyzing the LIGO data

What software is available and where is it?

LSC Data Analysis Tools

- » Mirroring of LIGO data across compute sites -- **Lightweight Data Replicator (LDR)**
- » Real-time data quality and monitoring; off-line use for veto analysis -- **Data Monitor Tool (DMT)**
- » Parallel (MPI) analysis with clusters -- **LIGO Data Analysis System (LDAS)**
- » Autonomous analysis on clusters -- **Condor** (batch job manager)
- » Work-station post-processing of results
  - – Matlab (graphical/analytical analysis)
  - – ROOT (high energy physics analysis environment)
- » Software Libraries: LAL, LALapps, DMT, Frame, FFTW… more

Analyzing the LIGO data

What hardware is available and where is it?

LIGO Laboratory Hardware
Tier 1 and Tier 2 Centers

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TIER-1 CENTER

	CPU (GHz)	Disk (TB)	Tape (TB)	Network()
CIT	1150	30	500	GigE
LHO	763	14	140	OC3
LLO	391	7	140	OC3
MIT	244	9	-	Fast
TOTAL	2548	60	780	

Archive ALL data under SAM-QFS system

Several SUN SMP nodes

LIGO Scientific Collaboration (LSC) Tier 2 Sites

- 889 GHz CPU
- 34 TB RAID 5 storage
- OC-12 (622 Mbps) to Abilene

LIGO Scientific Collaboration (LSC) Tier 2 Sites

UNIVERSITY of WISCONSIN UWMILWAUKEE

- 296 GHz CPU
- 64 TB storage (commodity IDE) for data
- OC-12 (622 Mbps) to Abilene

Other Analysis Sites

Max-Planck-Institut
für Gravitationsphysik
Albert-Einstein-Institut

- 670 GHz CPU
- 40 TB storage (commodity IDE)
- Fast Ethernet to G-WiN (to GEANT to Abilene)

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

- 272 GHz CPU
- 18 TB storage (RAID 5 and commodity IDE)
- GigE to SuperJANET (to GEANT to Abilene)

THE UNIVERSITY OF BIRMINGHAM

- 508 GHz CPU
- 18 TB storage (RAID5 and commodity IDE)
- 10 Gbps to SuperJANET (to GEANT to Abilene)

Funding

- LIGO Tier 1 and 2 centers (NSF funding)
- UWM and PSU Tier 2 centers (NSF MRI, iVDGL, University)
- Limited FTE support for operations
- Alliances with private sector (Samsung, DELL, SUN, Infinicom)
- **ITR2003** -- $3M over 4 years (start September 2003) mostly for FTE support for (US) LSC Tier 1 and 2 centers
  - » Lightweight Data Replicator (LDR)
  - » Data Monitoring Tool (DMT)
  - » Tcl/Tk interface to Globus
- **Extend LIGO data analysis systems to the GRID**

Lightweight Data Replicator (LDR)
Scott Koranda

- Grid-based data mirroring service for LIGO data
- From prototype to full deployment across LIGO and its international partners
- Research addresses the ability to access, manage, and communicate information on a large, international scale.

Data Monitoring Tool (DMT) John Zweizig

- LIGO data analysis software includes a ROOT-based package intended for near real-time detector performance monitoring software suite .
- Mission critical package for delivering the LIGO science – need to use for purposes way beyond real-time detector investigations
- Proposal aims to Grid-enabling this key software component so that the DMT can run on all LIGO Tier centers. Address job control and resource manager, data management and security.
- Research goals: Grid-enabled class of tools for a distributed computational environment. Successful implementation would lead a new paradigm for distributed data analysis in near real-time.

TclGlobus
Kent Blackburn

- Effective integration of existing LIGO software infrastructure with the VDT necessitates that a new SWIG-wrapped version of the Globus package be developed using Tcl as the interface language.
- This research specifically addresses the utilization of Grid tools to ensure secure access to scientific databases and Grid-registered computational infrastructure.
- A Tcl interface to the Globus toolkit has a broad application and would benefit the LIGO and the larger astronomy community.

In a nutshell…

- LIGO is opening a new frontier in observational astrophysics: it is now observing, acquiring science data and is full analysis production
- Realizing the full science content of LIGO data is computationally and data intensive
- Hardware resources for data analysis are geographically distributed over 9 sites and 2 continents (and growing…)
- People resources for data analysis are geographically distributed
- Emerging Grid Computing technology helps put data + hardware + people together for more science

Today's schedule

- 09:30-12:00 Presentations
  - » 09:30-09:50 Erik Katsavounidis: Introduction
  - » 09:50-10:20 Scott Koranda: LDR
  - » 10:20-10:50 John Zweizig: DMT
  - » 10:50-11:10 coffee break
  - » 11:10-11:40 Kent Blackburn: TclGlobus
  - » 11:40-12: 00 Scott Koranda: Summaries
- 12:00-13:00 Lunch break
- 13:00-16:30 Open discussions