# LSC Data Analysis White Paper
# Draft V

Stuart Anderson, Keith Riles, Albert Lazzarini,
Tom Nash, Alan Weinstein, Alan Wiseman (Chair) [*]

December 19, 2001

RCS Revision: 1.18 Date: 2001/12/19 10:19:50 UTC — Compiled: 2001/12/19

## Contents

---

[*]Drafts I-IV were written by Bruce Allen, Sam Finn, Tom Prince, Keith Riles, Rainer Weiss

# 1 Executive Summary

This paper outlines the LIGO Scientific Collaboration plan for analyzing LIGO data. All aspects of this plan – and the organization of this document – flow naturally from the overarching goal of the project: **to use gravitational-wave interferometers to test relativistic gravitation and observe the universe.** In support of this goal, this document spells out the science that is attainable with LIGO detectors, an organizational structure that is capable of sustaining a long-term data analysis effort, a plan for validating and maintaining the data analysis software, a usage model that insures a broad group of collaborators can contribute to the analysis effort, and both a bold and a practical look to the future.

Although much of this document is devoted to laying out a baseline plan for analyzing the data, there are several forward-looking themes that run throughout: How can the collaboration effectively utilize additional computational resources as they become available? How do we organize a scalable and adaptable analysis effort? What new research directions should the collaboration pursue in the long term? The answer to these question begins in Section 2 where we lay out an expansive set of science goals, many requiring vast computational resources. The organizational structure laid down in Section 3 is designed to be scalable in such a way as to embrace new technology such as "grid" computing. Chapter 4 provides a set of guidelines to ensure that software development is manageable. The usage model in Chapter 6 is designed to expand the number of researchers brought to bare on the research problems. In Chapter 7 we take an ambitious look forward to examine what possible research directions we should pursue in the future.

The basic outline of the document is as follows:

- In Chapter 2 we set ambitious but realistic scientific goals. The underlying themes are:

  - Testing Relativistic Gravity
  - Gravitational Wave Astronomy. Specifically, we will search for all anticipated types of sources:
    * Inspiraling compact binaries
    * Continuous waves, e.g. from pulsars
    * Stochastic background
    * Bursts of limited time-duration. Perhaps unmodeled, or poorly understood sources.

- In Chapter 3 we outline the organizational structure for the data analysis effort. It will be organized along both functional lines (i.e. astrophysics and detector characterization) and along analysis topics (e.g. an analysis group devoted to detecting, or setting an upper limit on, the strength of the stochastic background of gravitational waves).

  - Analysis groups (e.g. the current "Upper Limits Groups") are formed within the collaboration by an internal "proposal-driven" process. Details of this process are laid out in Section 3.1.
  - Any scientific conclusion drawn from LIGO data will require a detailed understanding of the instrument. Section 3.2 lays out the LSC **plans for characterizing the detectors** (e.g. In broad measure, these efforts will be coordinated by the Detector Characterization Data Analysis subgroup. Their efforts will focus on:

* Estimation of the detector noise level
* Calibration of the instrument and the response function
* Sensitivity of the detector to the environment
* Generation of triggers and vetoes from ancillary and environmental measurements
* Coordinating the on-site LSC participation in the engineering and science data runs.

- Detecting a signal will also require detailed understanding of the astrophysical source that produced it. Section 3.3 lays out the LSC plans for for developing algorithms and software to search for astrophysical signals. These efforts will be coordinated by the Astrophysical Source Identification and Signature [ASIS] Data Analysis subcommittee. Primarily these efforts are focused on:

    * Addressing broad theoretical and astrophysical questions that are relevant to data analysis.
    * Educating the collaboration on ways to use grid computing for astrophysical searches.
    * Continued efforts on algorithm development and implementation in support of the primary search types: inspiral, continuous waves, stochastic background and unmodeled bursts.

• The integrity of our scientific results depends squarely on the integrity of the software that produced them. In Chapter 4 we outline our plans for maintaining and testing our search code. The cornerstones of the software policy are:

- The collaboration software will be publicly available in order to maximize the number of users and testers.

- The Software will be tested and validated in mock data challenges.

- We will require the use of uniform, agreed-upon data products at the beginning and the end of each analysis pipeline to facilitate comparing results from different analyses.

- A stable software environment will be maintained and controlled by the Software Change Control Board.

• LIGO will accumulate about 500 TBytes of raw data during the first two years of operation. With this volume of data it will be impossible for every researcher to have his or her own copy. In Chapter 5 we lay out a hierarchical array of "reduced data sets", each bringing the data set down to a smaller, more manageable volume.

• The collaboration is made up of researchers from all types of institutions: state universities, liberal arts colleges, national labs, international institutes. In Chapter 6 we outline a usage model that is designed to incorporate a large – and scalable – group of scientists from all types of institutions into the data analysis effort. This is done by laying out a hierarchical system of (Tier1, Tier2 and Tier3) centers. These centers are designed around the GriPhyN (Grid Physics Network) model. The underlying principle is to bring the maximum number of researchers in close contact with LIGO data.

- Chapter 7 takes a look to the future: first with an eye to the far reaching possibilities of gravitational wave research, and then with a more practical view of what we can and should do to position the collaboration to achieve these far reaching goals.

# 2   The Science Goals

## 2.0   Overview

The science goals of the LIGO Science Collaboration are **to test relativistic gravity**, and **to develop and exploit gravitational wave detection as an astronomical probe, both by itself and in conjunction with other astronomical observations.** Accomplishing these goals will require a major effort to understand the sources of the gravitational waves and to understand the detectors.

   In planning for LIGO I data analysis, we make several assumptions:

1. There are no known gravitational wave sources whose "best-guess" rates and strengths are sufficiently large that we can be sure of detections during the first several years of LIGO operation.

2. There are great uncertainties associated with either or both the rates and strengths of all conjectured sources.

3. LIGO, GEO and VIRGO will extend our sensitivity to gravitational wave sources in a new frequency regime by two to three decades in amplitude and bandwidth.

Consequently, the LIGO I data analysis strategy is opportunistic, emphasizing breadth over depth (*i.e.,* range of "covered" sources over in-depth focus on a single source). In particular, we will develop strategies that pay close attention to detection of entirely unanticipated – serendipitous – sources. Our strategies also recognize the current theoretical bias that the signals may be too weak for a detection, and therefore our data analysis approach is also geared toward placing upper limits on signal strengths and in the event rates. However, we it will be sufficiently flexible to recognize and permit the characterization of unexpectedly strong signals.

## 2.1   Testing Relativity

The existence of gravitational radiation is not a unique property of general relativity; nevertheless, general relativity makes several unambiguous predictions about the character of gravitational radiation. In the event of a detection with high signal to noise ration, these predictions can be tested.

**Black holes and strong-field gravity.**   The radiation associated with the violent formation of a black hole reflects the detailed nature of strong-field gravity,and is, in fact, a unique probe of this fundamental phenomenon. In general relativity, the late-time radiation is a superposition of several damped normal modes. In general relativity, the frequency and damping constant each mode (overtone) is uniquely determined by the by the final black hole's mass and spin, and thus an observation of any single overtone gives a measurement of the black hole mass and spin. Thus, within the context of general relativity, an observation any additional overtones should yield the same mass and spin: any inconsistency is evidence of non-Einsteinian strong-field gravity.

**Spin character of the radiation field.**   General relativity makes a specific prediction for the polarizations of the gravitational wave field. Multiple interferometers can detect this polarization as well as components associated with other relativistic theories of gravity (scalar, vector, non-metric tensor). By using the radiation from long-lived (*e.g.,* CW) sources it is possible to distinguish between different polarization components and thereby set limits on alternate gravitational theories.

**Gravitational wave propagation speed.**   In general relativity gravitational radiation travels at the speed of light. The measurement of burst gravitational-wave sources associated with distant astronomical events (*e.g.,* supernovae or gamma-ray bursts) also observed by electromagnetic channels can be exploited to limit a difference between the actual propagation speed and the speed of light. (This can also be characterized as a measurement of the mass of the graviton.)

## 2.2   Gravitational Wave Astronomy

The gravitational-wave "sky" is entirely unexplored. Since many prospective gravitational wave sources have no corresponding electromagnetic signature (*e.g.,* black hole interactions), there are good reasons to believe that the gravitational-wave sky will be substantially different from the electromagnetic one. Mapping the gravitational-wave sky will provide an understanding of the universe in a way that electromagnetic observations cannot. Being a new field of astrophysics it is quite likely that gravitational wave observations will uncover new classes of sources not anticipated in our current thinking, hence data analysis strategies need to be broad based and flexible.

Detectable gravitational wave signals of limited time duration will most likely involve stellar mass compact objects undergoing relativistic motion. Observed gravitational wave signals can tell us about the characteristics of underlying sources while their statistics can tell us about the broader character of the source population and can be used as markers for cosmological measurements.

Some gravitational-wave signals will be accompanied by an electromagnetic, neutrino or cosmic ray signal. For example, core-collapse supernovae are strong electromagnetic and neutrino sources. Still other electromagnetic sources *may* have a substantial gravitational radiation component: examples include pulsars, quasi-periodic oscillators and low-mass x-ray binaries, nascent neutron stars in the year following their birth in a supernova explosion, and gamma-ray bursts. For these sources, multi-channel (electromagnetic, neutrino, particle and gravitational) observations of the signals will provide important information regarding the physics of the underlying sources and, in some cases, may be the only way to differentiate between different source models.

LSC analysis goals are guided by guided by consideration of the the following sources:

**Compact binary inspiral:**   We will measure (or place an upper limit) on the rate of compact binary inspirals. In the event of a strong signal(s) from a binary neutron star, we will be able to study the supernuclear equation of state of the matter comprising the star. If the signal arises from a binary black hole, we will the strong-field predictions of general relativity.

**Gravitational waves and gamma-ray bursts:**   We plan to study gravitational-wave data that is coincident with gamma-ray bursts, and thus set upper limits on the in-band gravitational wave power associated with gamma-ray bursts.

**Stellar collapse and black hole formation:**   We will search for stellar mass black hole formation. If the radiation associated with the formation of a black hole (i.e., the ring down) is observed, the black hole mass and angular momentum will be quantified and, to the extent possible, general relativistic predictions tested.

**Supernovae:**   We will search for the gravitational arising from core-collapse supernovae or place upper limits on the gravitational-wave power radiated in-band. For sufficiently strong signals, an analysis goal is to provide early-warning to astronomical observatories, allowing those observatories to capture the early part of the supernova light curve. Should radiation from core-collapse supernovae be observed, it will be used together with neutrino observations to test theories of supernova dynamics.

**Nascent neutron stars:**   We will search for neutron stars formed in supernovae. New-born neutron stars are rapidly rotating and may have a gravitational-radiation driven instability that carries away the bulk of the angular momentum during the first year following birth. The greatest contribution to the signal occurs in the last several weeks before cooling of the neutron star damps-out the instability. An LSC analysis goal is to be prepared to search for this radiation, testing this conjecture and possibly characterizing the evolution of the supernova remnant.

**General gravitational wave bursts:**   We will search for bursts whose source or detailed character (i.e., waveform) is not known in advance. Such bursts might arise during compact binary coalescence (following inspiral but before the black hole ringdown), during "optically silent" stellar core collapse (failed supernovae); however, other, unimagined sources might also be responsible for observable bursts. The analysis of the data from multiple detectors is essential for this type of investigation and and many others, such as a stochastic background.

**Pulsars and rapidly rotating neutron stars:**   We will observe (or set limits) on the power radiated by known, young pulsars and by previously unidentified rapidly rotating neutron stars at specified, fixed locations in the sky. Should gravitational radiation associated with a pulsar be observed, it will be used to determine the ellipticity of the neutron star and characterize the stress supported by its crust. A longer range goal is to develop the techniques to observe or set limits on the power radiated by unknown rapidly rotating neutron stars throughout the entire sky (i.e. an unbiased, all-sky, broad band search for periodic sources).

**QPOs and LMXBs:**   We will search for gravitational wave power radiated by certain quasi-periodic oscillators and low-mass x-ray binary systems, either bounding or setting upper limits on the radiated power.

**Stochastic Signals:**   We will search for the presence of a cosmological stochastic gravitational wave signal, either bounding or setting an upper limit on the in-band signal power. In addition to stochastic background from cosmological sources, there could be a background from many, distant, unresolved sources.

# 3   LSC Data Analysis Activities

## 3.0   Overview

Our goal in organizing the LSC data analysis activities is to satisfy two competing criteria. On the one hand, the collaboration must be able to sustain long-lead time research and development activities. On the other hand, the structure should should be flexible, so that we can accommodate new ideas and new people. To meet these two competing criteria, we will organize the effort with both permanent structures organized along functional lines to meet our long term needs (detector characterization and astrophysics), and a a more fluid set of "analysis" groups organized around particular searches. For example, a group might focus on the stochastic background.

The permanent data analysis structures will be the *Detector Characterization* subgroup [Section 3.2] and the *Astrophysical Source Identification and Signature* (ASIS) subgroup [Section 3.3]. Currently, there are four analysis groups. These are commonly referred to as the "Upper Limits Groups", as their central goal is to use engineering data to place an upper-limits on source strengths and event rates. In Section 3.1.1, we also describe a procedure to formally recognize other working groups engaged in specific software development tasks.

The analysis groups will be fluid. Groups of LSC members with an idea can submit an internal proposal to initiate their effort. These proposals will be reviewed by a committee within the LSC. The proposal criteria are spelled out below, but the intent is simple: to insure that all data analysis efforts that require significant Lab or LSC resources (or significant data) have broad support and participation within the collaboration.

In this paper we give a fairly detailed outline of the Detector Characterization and Astrophysical Source Identification and Signature activities; however we only give a short description of the purpose of the analysis groups and the mechanisms by which they form. Details about the activities of the analysis groups are (by design) fluid, and are given in each group's proposal. These these can be accessed from the LSC website at http://www.ligo.caltech.edu/LIGO_web/lsc/lsc.html.

## 3.1   Analysis Groups

In order to maintain an open, innovative program of data analysis within the collaboration, we have established a simple "proposal-driven" process for members to initiate efforts to tackle specific scientific problems with LIGO data. Any member (or group of members) may submit a proposal.

Proposals should be submitted to the LSC spokesperson. A committee consisting the Spokesperson, the Lab Directorate, the LSC Software Coordinator and the data analysis subgroup chairs will review the proposals.[1] Although this process is more formal than typical interactions among collaborators, it is not intended to be either a time consuming or a laborious process.

The rationale behind this proposal process is as follows:

1. This process is simple mechanism to allow an influx of new people and new ideas into the data analysis effort.

---

[1]Some proposals may have far reaching impact on the collaboration; therefore, at the discretion of the Spokesperson, the review committee may be enlarged, e.g. the entire LSC Executive Committee.

2. The proposals will be announced to the entire collaboration (e.g. by email and posted on the LSC web page) to encourage participation of people with a range of expertise. Specifically, we want a healthy mix of theorists and experimentalists in all analysis groups.

3. Laboratory and LSC resources (read: people) are often stretched thin. The proposal process allows an early assessment of whether or not a given activity can be supported. Simply put, we do not want collaborators embarking on research efforts that cannot be completed because of insufficient Lab and LSC resources.

4. This process helps insures that all rules governing access to LIGO to data can be followed.

A comprehensive list of proposal review criteria can be found on the LSC web page, however those criteria are derived from the following list. Proposals should address the following items.

- the scientific problem to be addressed

- the computational and analysis methods to be used

- the logistics to carry out the analysis:

  - an estimate of the laboratory resources required
  - an estimate of the LSC resources required
  - the division of responsibility between the proposers
  - students assigned to the effort
  - an estimated schedule for completion

- an outline of the publication(s) that are expected to arise from the analysis

  Publications resulting from the analysis will be reviewed and authorized by the entire collaboration as described in the LSC Publications Policy.

### 3.1.1 Working groups

From time to time, it has been necessary to form *ad hoc* working groups to accomplish software development tasks. Here we present a *chartering* process (similar to the proposal process described above) that can be used to "formalize" such efforts. The purpose of establishing this procedure is to insure that the members of these groups receive the proper recognition and support for their efforts, and to insure that all members of the collaboration have an opportunity to participate.

In order to formalize the activities of any working group, the organizers should write a charter. Although these groups will be organized around software tasks and not specific scientific goals, these charters should should be written along the lines of proposals for the Analysis Groups described above. In particular, the charter should briefly (a few pages at most!) address the following points:

- the description of the work to be done

- the importance of the specific task to be accomplished

- the logistical plan to carry out the task, including:

  - an estimate of the laboratory resources required

  - an estimate of the LSC resources required

  - the division of responsibility between the proposers

  - students assigned to the effort

  - an estimated schedule for completion

These charters should be submitted to the LSC spokesperson. A committee consisting the Spokesperson, the Lab Directorate, the LSC Software Coordinator and the data analysis subgroup chairs will review them.

## 3.2 Detector Characterization Subgroup

### 3.2.1 Introduction

Data analysis requires a systematic understanding and characterization of the detectors: its response function, noise behavior and sensitivity to the environment, and the correlated noise between detectors. The confidence associated with source detection or upper limits for detection depends on detector performance characteristics, including: power spectra, the probability distribution of the detector output, stationarity of the noise, line noise sources, and the statistics of transients. Detector characterization is also critical to improving the detector's performance and in designing new detectors.

Detector characterization involves both invasive (*e.g.*, stimulus-response) and passive (*e.g.*, monitoring) techniques and is carried out at several levels. The Global Diagnostic System (GDS) is closest to the detector, monitoring all data channels on-line and before archiving. GDS establishes rudimentary performance diagnostics during commissioning and has the unique ability to stimulate the detector and measure its transfer functions between different input and output ports.

The second level is represented by the Data Monitor Tool (DMT) which operates off-line and monitors the detector and environmental sensors in real-time using dedicated workstations at the observatories. The DMT's primary function is to update the LIGO meta-database with information on interferometer performance and identified instrumental/environmental transients. Selected transient types (triggers) also cause alarm messages to be sent to the control room, and real-time displays provide updated measures of interferometer performance. In addition, the DMT provides customized data sets for particular detector investigations.

The third level is in-depth, offline (and often off-site) analysis, which includes detailed performance characterization, transient analysis and statistics and trend analysis, as well as earches for correlations amongst detectors. An associated activity is instrument and noise modeling in which an End-to-End model of the detector, built up from its various sub systems, is driven with both astrophysical signals and the observed noise. This is one of the principal Monte Carlo simulation tools to establish the confidence of a detection.

Although much of the detector characterization is carried out at the observatories using the full data set, the algorithm development and testing takes place at many locations in the collaboration. It is sometimes necessary to perform more refined characterization in periodic "reruns" over the archived data at a central repository (Caltech) or over standard reduced data sets at the sites. It has also proven useful to carry out detector characterization at LSC members, institutions, using customized reduced data sets. It is important that all LSC groups have a means of receiving these reduced data sets, a requirement that affects data storage formats and network bandwidths, as described in the chapter below on the Usage Model.

The Detector Characterization working group within the LSC is dedicated to implementing the tasks outlined above. Its main goal is to provide the tools for delivering the highest quality, best understood data and characterization of the interferometer noise, to enable the highest sensitivity for gravitational wave searches. This working group provides scientific support both to LIGO operations and to astrophysical analysis. Many of the group's scientists are also members of existing Upper Limits analysis groups examining engineering run data and will participate in analysis groups formed during the science runs. We expect this dual membership to provide a synergistic interplay of commissioning/characterization and astrophysics analysis that will lead to detector

characterization that is more refined and better suited to the needs of particular gravitational wave searches.

During 1999 and 2000 the Detector Characterization group focused on development of software tools. During the 2001-2002 commissioning phase it is focusing on particular detector investigations tied to engineering runs. In the longer term it will apply its tools and the lessons learned during the engineering runs to the LDAS pipelined data analysis and gravitational wave search effort.

### 3.2.2 The 2001-2002 Commissioning Activities

More than a dozen investigation teams have carried out various studies of engineering run data and have reported at monthly teleconferences and at LSC meetings. It is expected that similar investigations, but of increasing depth as interferometer sensitivity improves, will be carried out during the science runs. Group members have also manned 24-hour scientific monitoring shifts during engineering runs to ensure the quality of the data and to carry out specialized studies[2]. These scientific monitoring shifts will continue in science runs. The commissioning of monitoring software at the sites and participation in engineering runs has required and will continue to require significant travel support from the NSF for LSC institutes active in detector characterization. In the following sections, we discuss in more detail the current and future work on

- Online diagnostics

- Offline performance characterization

- Offline transient analysis

- Data set simulation

- Engineering run activities

### 3.2.3 Online Diagnostics / Environmental Monitoring

Online diagnostics allow a rapid measure of data quality and verification of the instrument's current state, information that can be fed back to the control room and recorded for later use in offline analysis. In addition, diagnostics include invasive measurements, such as applying known waveforms at different inputs to the interferometers (*e.g.*, swept-sine transfer functions) and changing the state of the interferometers (*e.g.*, measurement of optical loss in arms via single-arm-lock visibility). Most of the initial work in online diagnostics is being carried out as part of instrument installation & commissioning. This work is extensive, requiring low-level software for hardware control (*e.g.*, control of D/A converters via VME reflective memory modules), medium-level software for implementing specific algorithms (*e.g.*, stimulus-response) and high-level software for control and display of diagnostics results.

---

[2]Detailed information on software tasks and developers, on engineering run investigations, and on scientific monitoring during engineering runs can be found on the Detector Characterization web site: `http://www-mhp.physics.lsa.umich.edu/∼keithr/lscdc/home.html`.

### 3.2.4  Offline Performance Characterization

The goal of offline performance characterization is primarily to establish average noise properties of the system, identify correlations between signals and to gain statistics on recurring transient phenomena, especially, those with a small duty cycle. This information will be used in gravitational wave searches to remove well understood stationary noise and to populate a database for vetoing putative astrophysical burst signals. Ongoing studies include the influence and reduction of narrow spectral peaks in the data such as

- Electrical mains contamination (60 Hz & harmonics)

- Suspension fiber violin modes

- Internal mirror resonances

- Isolation stack normal modes.

A particularly interesting study is the variation of the amplitude and frequency of these narrow features as a means of enhancing their removal from the data. To understand the rms instrument noise, studies of the broad band seismic and thermal noise are being carried out. Techniques are being developed to identify and remove non-Rayleigh spectral components in the data such as wandering oscillators.

It is also necessary to describe the operating state of the instrument. Examples of ongoing studies include: the operation of the servos (*e.g.*, full/partial/poor lock), linear interchannel correlations (including frequency dependence), and non-linear cross couplings. It is also desirable to provide immediate measures of astrophysical sensitivity, *e.g.*, summary metrics such as strain sensitivity at particular representative frequencies, maximum viewing distance for an inspiral standard "candle", and the rate of single-IFO transients matching astrophysical templates.

The above measurements of stationary or quasi-stationary behavior rely primarily upon analysis tools in the frequency domain,such as: power spectra, band-limited rms, matched filters and principal value decomposition. More general methods using time-frequency analysis (e.g. wavelet analysis) have also been developed and their advantages & disadvantages are under evaluation.

### 3.2.5  Offline Transient Analysis

It is necessary to identify and record transient signals in the gravitational-wave and other channels due purely to the instrument or to its terrestrial environment. Identifying such waveforms prevents possible confusion with astrophysical burst sources, but more important, allows for correction of the data and may provide diagnosis of curable problems.

Examples of anticipated transients include a large variety of instrumental and environmental impulses such as:

- Internal relaxation of suspension wires

- Dust particles dropping through the beam

- Flickering optical modes

- Ringdown of violin modes after lock acquisition

- Onset of servo instability or of out-of-band line excitation

- Onset of analog or digital saturation in the controls system

- Data acquisition malfunctions

- Lightning and wind gusts.

Some of these may be recognized immediately in the dark port signal. Others require correlation with one or more instrumental or environmental channels. Detection methods for transients include sudden increases in band-limited RMS, matched filters, threshold triggers on time-domain or frequency-domain amplitude and more general time-frequency analysis (*e.g.*, wavelet analysis). An event catalog of known transient types is under development and will continue to evolve, as experience is gained.

### 3.2.6  Data Set Simulation

Simulation includes both near-term phenomenological modeling to test monitoring algorithms and far-term *a priori* detailed, component-based modeling for comparison with actual instrument response. The former includes modeling of random noise, lines (*e.g.*, violin modes) and other parameterized waveforms and allows superposition of these waveforms. The latter falls under the heading of the ongoing LIGO End-to-End modeling and attempts a bottoms-up model of full interferometer response in the time or frequency domain. The End-to-End Model is meant to simulate LIGO optics, servo control loops, suspensions, ambient environmental noise, time delays, misalignments, thermal lensing, and other effects. It includes a user-friendly graphical user interface and data visualization tools. One of the functions of the End-to-End model will be to test the recovery of astrophysical waveforms injected into the simulated data stream.

### 3.2.7  Engineering Runs and Other Site Activities

Detector characterization at the sites by LSC members has increased dramatically in the last year and will continue to increase as science running begins. Short engineering runs starting in 2000 have provided a valuable opportunity to prepare for science running in 24-hour operation. Shifts are manned by both interferometer operators (site staff) and by scientific monitors (LSC scientists). To prepare a pool of operators and scientists for full-time running in 2002, engineering run monitoring shifts have been double or triply manned by experts and trainees. In normal science running in the future, we expect one operator and one scientist on duty at all times.

Detector investigations have been carried out during each engineering run and with its recorded data afterward. Topics of study have included:

- Seismic noise

- Line noise

- Frequency noise propagation

- Linear correlations between the GW and environmental channels

- Inter-site environmental correlations

- Environmental disturbances

- Data stationarity/stability

- Angular fluctuations

- Tidal modeling

- Causes of lock loss

- Timing precision

- Data integrity

- Data merging.

The results have been used to guide commissioning and to help prepare for upcoming astrophysical searches.

In addition to the periodic large gatherings of LSC members at the sites for engineering runs, there is a small but steady stream of LSC visitors to the site who carry out detector characterization tasks. These range from installation and commissioning of environmental monitors to commissioning, upgrading and tuning of DMT monitors. Making the DMT monitors steadily more useful in real-time to operators and scientists on duty in the control room will continue as an ongoing effort and requires regular site visits by monitor developers.

### 3.2.8 Looking Ahead

An active detector characterization effort will continue to be essential for the foreseeable future. As the sensitivity of the interferometers improves, the character of the noise will change. The importance of certain noise sources will decrease as interferometers are tuned, while other sources will become more visible and limiting as the overall noise floor lowers, requiring further investigation and removal. Hence it will be necessary not only to maintain the infrastructure of already-developed characterization tools, but also the means to enhance those tools, as needed, to characterize an evolving detector. We expect that many of the tools developed for monitoring and describing the data online will evolve with time and better understanding into data correction algorithms applied within the data conditioning API of the LDAS data pipeline. Similarly, the size and content of reduced data sets (see Chapter 5) will evolve with time, as both the instrumental noise and our understanding of it improve. Also, should a putative gravitational wave source be seen by one of the astrophysical analysis groups, the Detector Characterization group will be called upon to concentrate all resources upon determining whether a terrestrial artifact can explain the signal. For these reasons we expect the Detector Characterization group to continue as an essential component of the LSC analysis effort throughout the science runs and once more in the Advanced LIGO era.

## 3.3 Astrophysical Source Identification and Signature (ASIS) Subgroup

### 3.3.0 Introduction and Overview

The mission of the Astrophysical Source Identification and Signature (ASIS) subgroup of the LSC follows naturally from a simple observation: In order to find signals buried in a noisy LIGO data stream, the collaboration will need search software that makes optimum use of our astrophysical understanding of the sources of the gravitational radiation. To meet this need, the ASIS subgroup will work at the interface between theoretical astrophysical research and the practical development of search code to analyze LIGO data. The primary mission of the ASIS efforts is to ensure that there is vigorous, on-going astrophysical research program within the collaboration that is focused on understanding potential sources of gravitational waves and to insure that the fruits of this research are swiftly and properly implemented in our gravitational-wave search software.

Since the formation of this subgroup in the Spring of 1998, ASIS efforts have focused on issues near this interface: the development of search algorithms and code for inspiral searches, unmodeled-burst searches, continuous waves searches and stochastic background searches. These effort exploited the current astrophysical knowledge of the sources, and the results now form the core of the data analysis being conducted by the current (Upper-Limits) analysis groups. [See section 3.1 for a discussion of the analysis groups.] Going forward, the ASIS research will continue in in this manner, by laying the ground work for the next generation of targeted analysis groups.

**Common Software Activities:** Unlike the analysis groups that focus on a single source type, ASIS is a forum with interest and knowledge in all types of searches. This "cross disciplinary" aspect of ASIS makes this the appropriate place to organize and/or advertise software activities that are common to all the analysis groups. An example of this is the need for a common set of software tools for performing Monte Carlo simulations to test the efficiency of analysis pipelines.

**Scientific Validation of Search Algorithms:** Most of the software used by the analysis groups will be validated through Mock Data Challenges as described below; however there is a separate issue of the scientific (specifically astrophysical) validity of the search methods being used. Although ultimately the entire collaboration has a voice in assessing such validity, the early discussions and debates should make use of the broad astrophysical interest of the people involved in ASIS activities.

**Grid Computing:** Grid computing is another activity that may be utilized in several types of astrophysical searches, because, for many sources, a thorough, deep search requires huge computing resources. In the future, these resources will likely come from the grid computing environment. Therefore, to ensure that the collaboration is prepared to make use of grid facilities, ASIS will begin an effort to educate the collaboration on this new paradigm. These efforts should include porting existing algorithms to the grid environment, as well as studying what new types of searches and algorithms are best suited for grid computing.

**Theoretical Questions:** Although much of the data analysis will be carried out in the individual analysis groups under the proposal driven mechanism described in Section 3.1, there is a tremendous need within the collaboration to organize long-lead-time, and exploratory data analysis efforts. In particular the ASIS subgroup will organize efforts to address theoretical questions that are important in searching for astrophysical sources, but are not being aggressively pursued by the broad theoretical community. These open questions are addressed Section 3.3.1.

### 3.3.1 Search-Specific Areas of Research and Development

Through the various analysis groups that are now operating, there are now vigorous efforts underway to search for inspirals, stochastic background, unmodeled bursts and continuous waves; however there is still development work that must be done to improve these searches. The work described below is intended to extend the work being carried out in the analysis groups.

**3.3.1.1 Inspiral/Merger/Ringdown Signals** Coalescing binary systems can produce both known and unknown waveforms. The parts of the waveform arising from the merger phase cannot presently be calculated; techniques to search for these signals are described in the section on unmodeled sources. Matched filters may be used for the known waveforms, including:

- Inspiral of systems with masses of a few $M_\odot$ (visible in the sensitive band below $\sim 300$ Hz for $\sim 90$ sec).

- The characteristic ring-down after formation of a black hole horizon (exponentially damped sinusoids with $2 \lesssim Q \lesssim 10$.) Since such waveforms could also be produced by other sources such as stellar core collapse, this search must be independent of the inspiral one.

Filtering methods to search for the inspiral and ringdown signals at a single site are well understood. Searches of this sort have already been carried out on data from prototype instruments, so the work required is primarily development. For a reasonable range of masses the search can be carried out on-line.

**Templates** for the expected gravitational waveforms are the main theoretical input to the inspiral and ringdown detection process. The literature contains time-domain template approximations that are sufficiently accurate for detection work, but potentially better methods of approximation have been proposed. The efficiency with which templates can be computed determines whether templates are computed once and used many times, or computed as needed. Efficient means of computing the templates can greatly reduce the computational demands of this search technique. These require development.

Templates vary significantly depending on the source characteristics (for example, binary masses, spins, and orbital eccentricity); consequently, the detector output must be correlated against many templates to detect a signal. Template spacing in parameter space depends on the detector's performance: templates and their spacing will need to be recomputed if the detector noise power spectrum changes shape significantly during the time-scale of the data segments being filtered. Practical ways of determining when this is necessary, and of re-locating the templates, need to be developed.

**Hierarchical searches** should be the most computationally efficient means of filtering the detector data through the bank of filters. The first pass uses a large, coarsely spaced grid of filters, identifying segments of data passing a low SNR threshold. A second pass uses a smaller, finely spaced grid of filtersderived near the region of interest, and a higher SNR threshold. Studies assuming Gaussian detector noise have derived optimal values for the two thresholds, and predict computational gains in the range from 5 to 30 compared to a one-pass filtering scheme. A flexible implementation of this method and the experimental determination of optimal thresholds for real instrument noise are now needed. Additional study of correlations between nearby filters, and of methods for constructing robust rather than optimal filter banks would be useful.

**Discriminators** are statistical tools which help distinguish between large filter outputs arising from instrument artifacts and those arising from potential gravitational wave sources. In this way they reduce the sizes of final event lists. Specialized $\chi^2$ statistics developed for analysis of interferometric data and for resonant mass detectors have proved useful in reducing false alarm rates. Discriminators which test if the postulated waveform is consistent with the frequency and time distribution of a signal in a given filter and with the registration of the signal across the filter bank need further development and characterization.

**Coincident event lists** are produced by filtering the data streams from different interferometers and comparing event lists, and then selecting those which match certain criteria. These include arrival time differences consistent with the light travel time, best fit source parameter differences smaller than some threshold, SNR ratios within certain bounds, and so on. While somewhat less sensitive than optimal filtering (or maximum likelihood analysis) of all signal streams simultaneously, it yields greater confidence. The criteria for combining and comparing these event lists still need to be determined.

**Combined searches** use output from different filter banks or lists of metadata to look for signals coming from all three stages (inspiral, blind search, ringdown) of binary coalescence. This can be done at either the single or multidetector level. The tools for such a search need to be developed.

The final stage in a search will probably be the use of multidetector statistics from a 2- or $N$-detector data stream to estimate the likelihood that a source is present. The scientific work on these methods is complete, and only implementation work remains.

### Establishing detection confidence.

Methods of establishing confidence include the detection of the ringdown associated with black hole formation juxtaposed after an inspiral waveform, and simultaneous observation of the signal in two or more detectors but not in the various environmental and instrument monitoring channels. Unfortunately there is only a small range of masses for which both the inspiral and ringdown signals can be observed with significant SNRs. It may also be possible to observe the harmonic structure (overtones) of these signals of black hole formation. Establishing confidence for ringdown signals will require a thorough understanding of the instrument, since such signals can easily arise from from electrical and mechanical control loops.

### Upper limits.

The effective volume of space surveyed for binary inspiral by LIGO varies as the $5/2$ power of the system mass up to a combined mass of approximately 25 M$_\odot$. For NS/NS binaries, the volume (at the expected sensitivity of the Initial LIGO interferometers) corresponds to a sphere of $\approx 15$ Mpc radius which includes the Virgo cluster of galaxies. Better modeling of this dependence of source number as a function of radius in our cosmological neighborhood for $R \lesssim 50$ Mpc is required. Once an analysis pipeline is operating, it can be thoroughly characterized using Monte Carlo simulations. In this way the most efficient operating point can be determined for setting upper limits on the rate.

### 3.3.1.2 Unmodeled Sources

There are many sources for which waveforms are not calculated, including supernovae, and the merger phase of binary coalescence. Since sources for which waveforms *are* accurately predicted probably do not have rates/amplitudes large enough to see with LIGO I, a substantial effort to search for sources with generic characteristics is desirable. Here, matched filtering cannot be used and more general techniques are needed. These methods may also be useful for identifying periods of unusual instrument behavior, and should be carried out on-line. In some cases (for example, supernovae) it is desirable to identify the source location quickly enough to alert electromagnetic observatories, so some analysis must be in real time on-site. Detection confidence for unmodeled sources relies heavily on near-simultaneous detection by multiple interferometers, and/or by other EM or neutrino detectors, and therefore the development of real-time $N$-detector techniques is crucial for this purpose.

In general, knowledge gained from numerical and analytical studies of poorly understood signals such as the neutron star or black hole merger waveform makes it possible to construct more efficient and sensitive search techniques, as well as expanding our knowledge of the relevant astrophysical sources.

It may also be possible to detect unmodeled sources using statistical correlation techniques, for example using gamma ray bursts or other triggers to identify short time windows in which a significant gravitational wave flux may be present. These correlation techniques require further development. They are low bandwidth but will be carried out offline due to the need for external astrophysical trigger data.

**Pulse matching techniques** use a bank of filters designed to look for generic pulses with $\lesssim 20$ cycles. Typically the set of filters consists of a Gaussian and (say 20) derivatives of it, similar to a wavelet analysis Since the time-scale is not known, Gaussian pulses of different widths are required. The techniques used to generate banks of optimal filters can be applied here to construct an efficient bank of such filters. Time domain thresholding is a variation of this method, which looks looks for signal amplitudes exceeding a certain threshold in the whitened data stream.

**Time-frequency** methods locate statistically-significant excesses of power in particular frequency bands. The best-studied method was developed to search for line-like features in the T/F plane. A related technique uses short FFTs to monitor energy in particular frequency intervals. There are currently efforts underway to port these analyses to the LDAS (LIGO Data Analysis System) environment.

**Power-monitoring** is a variation on this technique, which looks for excess power in the outputs of a set of filters designed to cover specific frequency ranges and durations. A good example is supernovae. Their waveforms can probably never be accurately characterized, since they probably depend sensitively upon initial conditions. Despite this uncertainty, numerical simulations suggest that the radiation power spectrum is a power-law, with $|\tilde{h}(f)|^2 \propto f^{-2}$, between 10 Hz and 1 KHz.

**Correlation techniques** look for unusual correlations between the outputs of two or more detectors, and correlation between other types of signals, such as gamma-ray and neutrino bursts. They can be applied to event lists generated using the above methods, or to a simultaneous data stream. Special filters could be developed for coincident detection of supernovae and other source types.

**Establishing detection confidence**

Until environmental and detector noise-burst artifacts are completely understood, the only way of establishing detection confidence for unmodeled signals is through correlation with other detectors (gravitational, neutrino, and electromagnetic) and by veto from the environment and instrument monitoring channels.

**Upper limits.**

A method for setting upper-limits on in-band signal strength for the trigger population is being developed.

### 3.3.1.3   Continuous Wave (CW) and Pulsar Signals

Rapidly rotating neutron stars are the most likely sources of continuous gravitational waves in the observable band. The signal from a CW source will be nearly sinusoidal at twice the rotational frequency of the underlying neutron star (plus weaker harmonic and sub-harmonic components). The signal amplitude from these sources will be sufficiently weak that observations over periods of months or years are required to accumulate enough signal power to detect the source or to set astrophysically interesting upper limits. During this period, the frequency and phase of the detected signal will change due to the diurnal and annual motion of the Earth and also due to evolution of the source. Variations arising from the motion of the Earth depend on the source position on the sky; slow variations arising from source evolution may be observable electromagnetically for some sources.

The computational complexity of a CW signal search varies dramatically depending on the amount of prior knowledge about the source parameters. If the position and intrinsic spin evolution are unknown, the search entails looking through a discretized parameter space with a huge number of mesh points. Since such searches are computer limited, there is a premium on the development of efficient algorithms. When the source position is known (a *directed search*) a search to the limit of instrument sensitivity is possible. For an unbiased (all-sky, broad band) search, instrument-limited sensitivity requires more computing power than is practical, because the signals are modulated by the earth's motion, and have unknown intrinsic frequency drifts.

**Directed searches for known phase pulsars** may be carried out using folding, in which the time-series is added together with a time shift equal to the period of oscillation. This technique is widely used to search for radio pulsars. Some further development may be required to produce the optimal SNR if the instrumental noise levels are drifting with time. The search in a known direction for **pulsars of unknown phase** is more difficult, but should be feasible if the intrinsic frequency drift of the source is not too large.

Searches for unknown pulsars require substantial computation. Since an all-sky search at the instrumental limit of sensitivity is not currently possible, the goal is to make the most sensitive search constrained by the available computational power. The most efficient known techniques are a two- or three-stage **FFT-based stack-slide or Hough-transform hierarchical** search. The methods have similar computational efficiency for Gaussian detector noise, but they may have different performance for digging signals out from non-Gaussian instrumental noise. These methods share many common features and work is underway to implement both of them within a single code.

**Robust algorithms** are specialized methods capable of searching for waves from poorly modeled sources (e.g., accreting x-ray binaries, r-modes in nascent neutron stars). Methods are also needed to search for emission from wobbling neutron stars, where significant energy is present in sidebands of the main "carrier" signal. Searches for pulsars in binary systems should also be possible, but algorithms don't yet exist.

**Discrimination techniques** will be needed as a way of verifying that signals which are found are gravitational in origin and not instrumental. These techniques should be capable of identifying wandering oscillators, and should also test for amplitude modulation consistent with the time-dependent detector response. These methods do not yet exist.

**Multiple interferometer** search techniques for both the detection and the confirmation stages of discovery do not yet exist.

### 3.3.1.4 Stochastic Background Detection

Stochastic backgrounds are signals produced by many weak incoherent sources. They are non-deterministic and can only be characterized statistically. Such signals can arise from early-universe processes (analogous to the electromagnetic CBR) and from present-day phenomena. They give rise to a (probably stationary and Gaussian) signal which is correlated between the two detectors. It will have the same spectrum in each detector, and is differentiated from detector noise by its inter-detector correlation, which depends in a known way on the signal spectrum and the detector separation and orientation. The greatest risk is that similar correlations may be produced by the (electromagnetic) environment.

Stochastic signals are expected to be quite weak compared to the intrinsic noise of an individual LIGO interferometer; consequently, detecting or placing a limit on a stochastic gravitational wave signal will require long observation periods over a bandwidth a few times the inverse light travel time between the interferometers.

Detection of a stochastic background signal requires fairly simple analysis of long stretches of data. This is well-suited to off-line analysis. Two detection techniques have been extensively studied, one based on combining cross-correlations of pairs of detectors, and the other based on a likelihood formed from $N$-detector data.

**Correlation statistic analysis** combines the data streams from pairs of detectors in an optimal fashion and has been shown to perform as expected with Gaussian detector noise. Additional work is needed to design tests to search for similar correlations between environmental channels at the different sites.

**Robust correlation statistics.** Correlation analysis appears to be badly affected by non-stationary and non-Gaussian detector noise. Recent work indicates that more robust methods which carry out a form of limiting should give about the same performance in the case where the noise is Gaussian, and are optimal or near-optimal in the non-Gaussian case.

**Maximum likelihood techniques** are an alternative to the correlation statistic analysis. In principle they are the most sensitive search technique, but in practice, if there are many unknown parameters (i.e. the detector's noise spectrum at every frequency) in which to maximize the likelihood function, they may not perform well. Further work is needed to determine the utility of this technique.

**Establishing detection confidence.**

Since stochastic background detection requires at least two detectors finding a signal with two detectors is not enough to establish confidence. Terrestrial effects, particularly correlated electro-magnetic noise at the two sites, can mimic a gravitational stochastic background signal. LIGO can place an upper bound on the amplitude of a stochastic gravitational wave signal, but it will be extremely difficult to assert confident detection. This will probably require including the data from additional detectors in the analyses. Many tests may prove useful as diagnostics: including correlation between nearby resonant-mass detectors and the LIGO interferometers, studies of the correlation matrix between gravitational strain and electromagnetic signals at the sites and the correlation analysis of the 4km and 2km interferometers at the same site.

# 4   Software Development Guidelines

## 4.0   Overview

One of the obvious goals of the LSC is to address scientific questions and write scientific papers. The integrity of our scientific results will rest squarely on the integrity of our software; therefore the underlying goal for our software "policy" is to insure the validity of the software and thus guarantee the accuracy of our scientific results.
To accomplish this goal, we rely on four main points:

1. An open development model. The code will kept readily available so that anyone can download it, test it, use it, and debug it.

2. Formal mock data challenges to validate the code. All software will used for formal scientific analysis must be validated in a mock data challenge. This applies to all types of software (LDAS, DMT, LAL, DTT). The goal of the mock data challenges is to test the entire data analysis pipeline under (increasingly) realistic conditions.

3. Uniform data formats for the input data and output results.[3] By requiring researchers begin there analysis with data in an agreed upon form (e.g. FRAME data) and end with an agreed upon format for output results (e.g. database table entries), it will make it easier to independently confirm results.

4. Stable software environment. As the collaboration and the Lab come to agreement on software specifications, datatype descriptions, mathematical conventions, and development environments, these will be controlled by the Software Configuration Control Board.

## 4.1   Software Policy

LSC science analysis pipelines will be implemented from modular software components that are validated and controlled as part of the LIGO/LSC Analysis Library (LAL). All LAL software will conform to a standard that has been defined jointly by LIGO Laboratory and the LSC (ref. LIGO-T9900030). LAL software is archived in a LAL CVS repository maintained by the LSC.

Analysis software, whether used locally or in a wide area network distributed computing environment (grid computing), will also conform to this standard, unless it is intended for personal stand-alone use without interaction with other LIGO/LSC software components. Individuals carrying out scientific analysis as part of the LSC must use search algorithms and code that has been validated as part of the LAL repository (e.g., they must use LAL libraries for any numerical analysis.) Input and output data will be in standard LIGO formats (frames, LIGO-LightWeight, XML, ilwd (internal LIGO lightweight data). All code integrated into the wide area distributed environment must undergo the same rigor and test methodology described below that is applied to the integration of LDAS with LAL search algorithms.

The LSC Software Coordinator has the principal responsibility for managing the LSC software development effort. Verification and validation of LAL components will take place at three levels:

---

[3]This document does not address the issues about access to data. Such issues are governed by the memoranda of understanding between the LIGO Lab and the individual research institutions.

(i) compliance to standards, (ii) piecewise component tests and (iii) integrated tests of the analysis pipeline through Mock-Data Challenges.

Software, database definitions, and other data representation standards, once adopted, shall be controlled by a Software Configuration Control Board (SCCB) that is chaired by the Software Coordinator. (See LIGO document T010050-00-Z for details.) In the case of changes to a standard that has been adopted by the larger GWIC community (e.g., the frame data format), the proposed changes must be approved by the GWIC representatives from the LSC.

## 4.2 The LIGO/LSC Analysis Library

The LAL configuration is managed by the LSC Software Coordinator, who coordinates regular releases of the LAL library with and between LDAS releases. Major releases will be scheduled to coincide with major LDAS releases 0.0.X throughout the development phase 1998 - 2001; 0.1.X for the initial Upper Limit Run; 1.0.X for the LIGO I Science Run itself. These will test LDAS functionality and support the development and testing of analysis pipelines. Intermediate releases will take place quarterly to correct bugs and provide incremental increases in functionality and performance.

All LIGO data analyses involve filtering operations — either linear or non-linear — on time series consisting of weak signals in the presence of additive noise. These analyses can all be described as compositions of "atomic" operations on a small number of rigidly structured data types. Typical atomic operations include linear algebra and filtering, signal processing methods and descriptive statistics; typical data types are time series, frequency spectra and linear filter transfer functions. LAL consists of these atomic operations acting on these structured data types.

All LAL software development will conform to style specified in T990030, which describes coding rules, documentation standards, software diagnostic and test requirements.

We expect that LAL will evolve and grow with accrued data analysis experience. Changes to LAL will be authorized by the Software Configuration Control Board (the SCCB, introduced above) whose members are appointed by the LSC and the LIGO Lab. Proposed changes will be weighed for relevance, impact on existing systems and resource, and benefits offered. See LIGO document T010050-00-Z for a definition of the SCCB charter and scope.

# 5   Data Products: Reduced Data Sets and Artifacts

## 5.0   Introduction

While operating, LIGO will collect data at a rate of approximately 9 MB/s during the first two years of operation. While all of these data will be archived in a central repository, most are expected to be of ephemeral value, being useful only for near-term instrument diagnosis or characterization. Few LSC scientists are expected work with significant quantities of this raw detector data.

   Therefore, three datasets of progressively reduced size and correspondingly increased scientific information density will be created and archived. Most analysis activities are expected to access one of these reduced data products in either FRAME or LIGO Lightweight XML format.

## 5.1   Archival and Reduced Data Sets

The LIGO data are acquired as a collection of several thousand channels at rates up to 16 KHz. Subsequent to acquisition the LIGO data stream will be reduced in volume through three successive stages. At each step some channels will be discarded, reduced in resolution (either dynamic range or bandwidth), or combined into new summary channels. This process of reduction is expected to be repeated a few times as our understanding of the instrument and what set of channels are important continues to improve. Here we identify four data sets, corresponding to the raw data and the product of each stage or data reduction:

**Level 0: Full IFO Data Stream.**   Level 0 data will be available on-site for a minimum of 1 week in FRAME format. After this time they will be available from the central LIGO archive in FRAME format for a minimum lifetime of 1 year.

**Level 1: Archived Reduced Data Set.**   Level 1 data consist of all important IFO and PEM data channels stored in FRAME format, together with regression, whitening, calibration, and instrument state data. Like Level 0 data these data will be used principally for detector diagnostic studies. The Level 1 data set will be approximately 10% of the full data, corresponding to approximately 50 TB of data during the first two years of operation, and will be generated and stored at the central LIGO data center indefinitely.

**Level 2: IFO Strain plus Data Quality Channels.**   For more detailed science analyses a further reduced data set containing basic IFO strain data plus a variety of quality channels will be provided. Quality channels will include calibration, whitening, and regression coefficients, as well as the most important auxiliary IFO and PEM channels. The total Level 2 data will be about 1% of the full data, or approximately 5 TB of data during the first two years of operation.

**Level 3: Whitened GW Strain Data.**   Level 3 data will consist of the best estimate of the (whitened) GW strain. The reported strain will be as free as possible from instrumental artifacts and reduced to approximately 1 kHz bandwidth. The Level 3 data set will include all the relevant whitening filter coefficients, regression and calibration information used in its production from the

Level 2 data. At a nominal 2 kHz sampling rate, a 2 year data stream from the three interferometers will be approximately 500 GB of data.

## 5.2   Metadata and Event Data

Most LIGO data will be selected for analysis on the basis of some distinguishing characteristic, e.g., coincidence in time with an astrophysical event, period of high seismic activity, or anomalous behavior of a control system. The LDAS system includes a database system for searching and making queries on summary information. The following types of information will be available from the database:

**Frame Data Information.**   This includes tables of locations of sets of frames, as well as statistics and spectra derived from sets of frames.

**Trigger, Veto, and Instrumental Artifacts.**   This includes information about the triggers and vetoes generated by Global Diagnostics System (GDS) and Data Monitoring Tool (DMT) filters, including information about the filters themselves. It also includes astrophysical search triggers, such as those generated by the binary inspiral, ringdown, burst, and periodic source analyses performed by LDAS.

**Non-LIGO Generated Event Information.**   The database will include environmental and astrophysical information from sources outside of LIGO, e.g., seismic alerts from external monitoring networks, electromagnetic storms, $\gamma$-ray burst events, neutrino events, UVOIR (UV, optical, IR) events such as supernovae, and events generated by other GW detectors.

## 5.3   Software Verification and Validation

Software verification tests the behavior of individual components. LSC component software verification involves documentation, component tests, and run-time diagnostics. Documentation describes in detail what the component is supposed to do, how it is supposed to do it, error conditions and how they are handled, and accuracy requirements or guarantees. Each LAL software component will include documented test code which tests the component for fault tolerance, accuracy and correctness of implementation as described in the documentation. Finally, each component is required to return at run time a status structure, which reports on the component's current functioning and provides diagnostic information in the event of an error condition. All these components — the documentation, the test suite, and software status reporting and error handling — are the responsibility of the LSC member(s) who supply the software component It is important for the quality of the integrated final product that all software modules be tested thoroughly at the module level by individuals who are not themselves the code developers. A thorough procedure of independent validation at the software component level is necessary to preclude the appearance for the first time of accumulated errors downstream at the integration level.

Software Validation tests the software components to be integrated into analysis pipelines in order to ensure that they can perform the analyses described in the science goals (Section 1 of this

document) with the requisite speed on the target hardware platform (i.e., the on-site and off-site LIGO Beowulfs.

Software system integration is tested at several levels. The LAL has a hierarchical, modular design, with increasingly sophisticated analyses built upon a base of more primitive library calls: e.g., power spectrum estimation by Welch's method involves sub-division of a time series into sequential overlapping components, the generation and application of a window function, discrete Fourier transform of the windowed sub-sequence, term-by-term modulus of the DFT results, and summing and normalizing the resulting frequency series. Each of these operations is a low-level library function that must properly integrate to compute successfully a power spectrum estimate

At higher levels, system integration, performance and analysis goals are tested through "Mock-Data Challenges" (MDCs). In a MDC, data of known character (e.g., noise of known statistical properties possibly superposed with a signal of known character) is passed through the system, whose response is observed and compared to the expected response. MDCs of increasing sophistication are carried out first on sub-systems and finally on the full system in different configurations.

System integration and performance testing will involve a single LSC/LDAS team that both generates test data and characterizes the system's performance. End-to-End tests of an analysis pipeline will be carried-out single-blind by two teams: one team generates data, which may include signals, and a second team analyses the data and reports back the conclusions. The two teams operate independently, with only the data (but no details of its character) passing between them. The system's ability to handle the analysis goals will be verified statistically by comparing the conclusions reached by the second team with the known character of the input data, generated independently by the first team.

These final MDCs require the ability to generate data streams with the statistical character of LIGO data. This characterization comes from the LSC detector characterization effort, described above, and involves the LIGO End-to-End modeling effort.

MDCs will be performed on an incremental basis. MDCs will be coordinated with each of LAL and LDAS major release; additionally, there will be MDCs in between major releases, continually testing the software in different configurations. MDCs are organized by the Software Coordinator in collaboration with the LIGO Laboratory LDAS team.

# 6  Computational Infrastructure and the Usage Model

## 6.1  Introduction and Overview

The computational infrastructure required for data analysis is determined by the emerging LIGO/LSC user/usage model. The model is based on a hierarchically arranged infrastructure of computational and storage resources. Three tiers of infrastructure are envisioned and each has a specific role. The tiers include

**Tier 1: LIGO Laboratory.**  The LIGO Laboratory sites at Caltech, Hanford, and Livingston constitute the distributed Tier 1 Center for LIGO.

**Tier 2: LSC Institution Sites.**  Eventually there will be between 3 and 5 Tier 2 centers established at LSC institutions, in addition to Caltech and MIT. The Tier 2 centers will be operated by a single LSC Institution subject to LSC priorities and accessible to all LSC institutions. Typically, a Tier 2 center will support institutions that are relatively nearby in network space.

**Tier 2: LIGO Laboratory.**  In addition to the archive and production system at Caltech, there are secondary scale systems (e.g., LDAS-dev and LDAS-test) that are of a Tier 2 scale. The MIT system is also similarly scaled. MIT will serve as an East Coast data mirror for key reduced data sets. In order to support Laboratory-based data analysis efforts, a part of the role of the MIT and the ancillary systems at Caltech will be to provide the effective Tier 2 support for Laboratory scientists and engineers.

**Tier 3: University research group resources.**  Individual research groups will have stand-alone computational hardware available to them for local autonomous use. The distinction between Tier 2 and Tier 3 is basically one of scale and the fact that Tier 3 resources are dedicated to the local group, while all or most of resources at a Tier 2 center are subject to LSC scheduling.

Three broad categories of usage are also defined:

- Local Processing/Local Data/Low-bandwidth WAN. This type of usage involves workstation-based analysis and analysis development activities using local data files. Typical activities will involve requesting small (1-10 MB) data files from the archive over the net (e.g., T1, T3, or DS3), or larger ones (1-100 GB) via tape, and analysis using programs running on local workstations. The analysis environment may or may not involve the LDAS software environment. It is expected that a large fraction of the LSC software development and instrument characterization will fall under this model. This mode of operation will typically involve local Tier 2 or 3 resources.

- Remote Processing/Remote Data/Low-bandwidth WAN. This model describes development and analysis using significant LSC resources accessed via the net through a browser or X-window interface. A typical example would be LSC scientists connecting from their home institution to an LDAS system at one of the Tier 1 or Tier 2 Centers. Analysis will take place principally within the LDAS software environment. Code validation, Monte Carlo analyses, as well as a large fraction of the computational intensive science analysis are expected to

fall under this model. This mode of operation will be particularly intensive during the commissioning phase for the LSC components of the LDAS software, particularly during the first 1-2 years of engineering and science data runs. During this period, the distributed LSC scientific community will be continuously improving their algorithms. Much (but not all) of this activity will require access to increasingly extensive data sets and issues and problems (like false triggers) become more subtle and rare. During this period, it is essential to provide access to data sets at the Tier 1 and 2 centers through X windows, with sufficient bandwidth and low enough latency that the user is not continuously aware that (s)he is a continent away. The computing capacity at the Tier 1 and 2 centers must be adequate to support the anticipated usage during this phase. The number of simultaneous X Window sessions that must be supported by the Tier 1 and 2 Centers in order to accommodate the LSC computing needs is at present not well defined. Experience during the engineering runs and during the early science run will be used to determine this more fully. However, our preliminary expectation is that on average during daytime operations, 10 such sessions will be active, with peak periods requiring the accommodation of as many as 25 sessions.

- Local Processing/Remote Data/High-bandwidth WAN. This usage model encompasses analysis on a local workstation or supercomputer using remote data files provided via high-bandwidth (OC-3 or greater) from the LIGO archive. Usage under this model is not expected initially; however, it is expected to play an increasingly large role in the future as high-bandwidth network connections and increasingly powerful local computing resource become more common. This mode of use can be supported by Tier 1, Tier 2 (or even Tier 3) resources, depending on the local resources.

## 6.2  Role of the Tier 1 Center

### Observatories

Operation of the interferometers and storage of Level 1 data is the highest priority activity at the IFO sites. In addition, local pipeline analyses will be operated to continuously monitor the strain channel data streams for a class of astrophysical waveforms having relevance for real-time detection. The on-site computing infrastructure is oriented toward local-access, and local processing with access from off-site controlled and given a lower priority. Three LANs will be supported: CDS/GDS, LDAS and general computing.

### The LIGO Data Center at Caltech

Caltech will house the LIGO data archive. Its principal roles are to provide access to archival data and support detailed science analysis on the combined multiple-interferometer data set. The reduction of Level 1 data to Levels 2 and 3 will be performed as data are ingested into the archive Remote user support will include searching the archive and selecting archival data for analysis. Analyses may be carried-out on the LIGO/Caltech workstations or Beowulf clusters, or transferred to a remote site via network or tape. The LIGO/Caltech LDAS is designed to provide support for five simultaneous high-bandwidth users, assuming a mix of tape and disk data transfers.

**LSC Tier 2 Centers**

For the foreseeable future, LIGO Laboratory's Tier 1 Center will likely not be able to provide all the computational resources that will be required to support the numerous research programs being undertaken by the LSC. For this reason, LIGO Laboratory and several LSC institutions have begun to develop resources that will eventually become Tier 2 Centers for the collaboration. At the time of this writing, the exact number and locations of the LSC Tier 2 institutions has not been defined. Much of this deployment will be carried out as part of the NSF's GriPhyN Project and the LSC Tier 2 Centers will constitute a portion of the US wide area network distributed computing grid ,which GriPhyN and other DOE and NSF funded programs are developing. The LSC Tier 2 Centers will number between 3 and 5 in addition to the Tier 2 centers at MIT and Caltech.

The Tier 2 Centers at MIT and Caltech will provide computational resources for Laboratory scientific staff to carry out their research that does not require the analysis pipelines. This includes exploratory R&D on algorithms, analysis techniques, and detector characterization.

MIT will be equipped with a Beowulf cluster for software development and local data analysis. MIT will act as a mirror for the Level 2 data product, in which case it will support use in the Remote Processing/Remote Data/Low-bandwidth WAN mode using the LDAS software environment.

LIGO Laboratory and the LSC will work together to define and develop the prototypical Tier 2 Center for the collaboration. Subsequent centers will be replicated from the prototype, with allowances for specific configuration details as they may be warranted. The designated Tier 2 Center LSC host institutions will be selected according to a set of criteria that are aimed at maximizing the accessibility and utility of these centers for LIGO scientific research and data analysis across the entire collaboration. Although the specific criteria have not yet been set down, it is expected that they will include, as a minimum, the following elements. The home institutions must:

- Be acceptable to the NSF, if the Tier 2 center funding support is provided through this agency;

- Have a dedicated PI who has demonstrated expertise and interest in distributed computing and who will devote a significant fraction of her/his time to the Tier 2 development and operations effort. The PI must be involved in grid-related research or other relevant activities.

- Tier2 facilities will need to have resources such as high speed LAN and WAN connections, support staff and possibly existing hardware such as Pentium Linux processors and disk storage. Much of these basic resources should be preexisting and leverage existing institutional infrastructure, because the institution should have had some experience at managing a facility of this type.

- Tier2 facilities must be located in a geographic location with suitable available network connections.

## 6.3   Role for Tier2 Centers

Tier 2 centers for the LSC shall provide mirrors for critical datasets that are useful to a broad segment of the LSC. These will include, e.g., Level 3 data and, to the greatest extent possible, Level 2 or subsets thereof. In order for these to be truly accessible to the LSC community, adequate

bandwidth dedicated to LIGO (ultimately OC 12) must be available at the Tier 2 Centers in order to ensure that no local bottlenecks exist in the distribution of LIGO data.

Tier 2 centers also represent significant computational capacity that is intended to augment the Tier 1 capacity. It is envisioned that the scale of each Tier 2 center will be comparable to the LIGO Tier 1 Observatory Site components. Unlike the Tier 1 Center, the Tier 2 center resources shall be available for exploratory analysis of datasets or analysis that is not in the mainstream of LSC searches.

In both these regards (data distribution and mirroring, and computation) the Tier 2 centers serve to offload the demand of resources at the Tier 1 Center in those instances where the analysis or data requests can be handled at the Tier 2 Center.

### Access to Tier 2 Center resources

Tier 2 Centers serve a role for the entire LSC. In this regard, the resources invested by NSF in these centers must be managed with a stewardship for all LSC member institutions. The operation of the Tier 2 Centers for LSC science will be the responsibility of the host institution. Access to the resources shall be provided by a mechanism within the LSC that ensures equitable distribution of bandwidth and CPU time to all LSC members, One mechanism that will be introduced is a computational resources allocation committee within the LSC that entertains proposals and requests and then distributes time and bandwidth for LSC common resources (Tier 1 and 2) according to the requests.

## 6.4   Infrastructure requirements

### 6.4.1   Tier1 Center

### LIGO Laboratory - Observatories

The observatories will have the capability of providing (approximately) a 28-day look back period for data acquired locally [4]. These will be available on a disk farm so that there will be a minimum overhead to access data.

The observatories will be running continuous pipeline analyses on PC Linux clusters. It is estimate that each interferometer will require O[20 GFLOPS] of computational capacity in order to process data at the same rate at which it is acquired. .

The metadata storage capacity for the observatories will accommodate approximately 500GB per interferometer.

These capabilities will be extended during the LIGO I Science Run as experience and resources will permit.

It is envisioned that, by the time of the LIGO I Science Run, the LIGO Laboratory WAN connecting the two observatories with Caltech and MIT shall be able to support OC3 bandwidth. This is sufficient to enable the data acquired at the observatories to be streamed to the archive.

---

[4]This look-back time will depend on the other (non primary) data that needs to be recorded

**LIGO Laboratory - LIGO Data Center at Caltech**

The archive at LIGO/Caltech will have the capability of providing access to at least 360 TB of data (which exceeds the presently envisioned volume of data from the entire LIGO I Science Run). Depending on which data are requested, access time will vary from delays consistent with disk I/O and Internet access to delays consistent with the retrieval of archived tape data from the large robotic silo.

The LIGO archive will be augmented with an extensible disk farm designed to accommodate the most commonly accessed LIGO data (e.g., Level 3 and Level 2). The farm will be extended throughout the LIGO I Science Run as data growth dictates and as resources permit.

LIGO/Caltech will have available a number of PC Linux clusters for different purposes. Software development and algorithm development will be supported on two small-scale clusters. Multiple interferometer pipeline analysis will be performed on a dedicated cluster whose scale is approximately equal to the aggregate capacity of both observatories.

The metadata storage capacity for the Caltech will accommodate approximately 100GB of relational database storage.

These capabilities may be scaled during the LIGO I Science Run as experience and resources will permit.

It is envisioned that, by the time of the LIGO I Science Run, Caltech LIGO Laboratory will be connected through university infrastructure to the Internet with an OC48 bandwidth.

LIGO/MIT will have data storage resources sufficient to provide data mirroring for the Level 3 data set. There will also be a PC Linux cluster of sufficient size to enable algorithm development and data analysis on a sufficient scale to support the research program of the MIT LIGO science staff.

### 6.4.2   Tier 2 Centers

The configuration of the Tier 2 Centers will be defined through a prototype R&D phase in association with the GriPhyN Program and similar projects. The centers are expected to have the following characteristics:

- Between 64 and 128 of the latest generation Linux/Intel processors, each with 512-1024 MB RAM, 72 GB or greater disks, 100BT or faster connections to a switch

- Initially have an OC3 and ultimately an OC12 connection to the national internet2 infrastructure;

- At least 20 TB of disk storage and accompanying high throughput data servers;

- A small AIT-2 or equivalent technology robotic tape unit for creating datasets for distribution to the LSC;

  This system configuration is expected to support a standard software environment, consisting of

- The LDAS software environment, which is supported only on Intel/Linux and Sun/Solaris systems;

- * a DB2 client for database access; and

- other TBD software, as this becomes identified and defined.

All computing and data storage infrastructure used for LIGO data analysis are to be accessible to the entire LSC as communal resources, with access modes described under remote-usage models described above. In addition, the Laboratory will provide resources for its science staff at MIT, Caltech, and the observatories, who are participating in analysis and detector-based R&D.

### 6.4.3 LSC-wide Support for Computational Resources

The infrastructure described above constitutes a formidable array of resources that will become available across the collaboration. In order to guarantee open access and efficient usage of these resources, it is necessary that the collaboration as a whole develop a mechanism of support in the form of a distributed help desk system. All institutional members of the LSC will share the burden of operating this network of infrastructure and resources. These terms of participation in overall operations will be defined in the MOU each institution has with LIGO Laboratory.

# 7   Looking to the Future

## 7.0   Introduction

The near-term LSC data analysis program ensures that within human- and computational-resource limitations, we can carry out reasonably sensitive searches for the primary categories of expected sources. The most pressing need is to provide support for the on-going activities, so that during the commissioning phase of the LIGO detectors the data analysis systems can be tested, debugged, and optimized.

The prospect of making our first direct detection of gravitational waves is exciting; however the ultimate goal is to make gravitational wave detections routine: to make gravitational wave observations a standard part of astrophysical measurements. Eventually, when detections have been made, our analysis program will transform into a study of the nature of the signals and the properties of their sources.

In this section we begin with a far-reaching, bold look at the exciting science that gravitational wave detectors may bring us. As there are no no guarantees which avenues will pay off in the long-run, these longer-term activities should develop naturally out of the LSC's nearer-term research efforts. Therefore we also we also present a practical plan for the near future: a plan flexible enough to position the collaboration for a smooth transition to more ambitious goals as they arise. Shaping the future in this way will require a timely, well-placed investment in computing hardware, and the support and training of additional scientists.

## 7.1   A Bold Look to the Future

The LIGO detector (and similar international detectors) will give us a view of the gravitational-wave sky far clearer than any detectors previously built. Even in its initial configuration, the LIGO detector will be roughly two orders of magnitude[5] more sensitive than its 40m prototype, and thus will probe roughly one million times as much volume. Another factor of 10 improvement[5] in an advanced configuration of the detector brings a further thousand-fold increase in volume into view. This leap in detector sensitivity motivates us to develop search algorithms, computational facilities and human resources that can fully exploit these instruments.

With these powerful new detectors, we look forward to the day when gravitational wave detections are common place and we can do statistical analysis of the results. But we should not lose sight of the fact that the first direct detection of a signal will, on its own, be a truly profound result: a result confirming a prediction of one of the fundamental theories of physics: the general theory of relativity. Since we have little idea what type of event will produce the first detectable signal, it is important to maintain a vigorous effort to understand all types of sources and develop algorithms to dig deep in the noise to find them.

1. **Mapping the Early Universe.** Just as the cosmic microwave background radiation is a picture of the universe when it is about 100,000 years old, the early universe has also left an imprint in the form of a stochastic background of gravitational waves. The gravitational-wave signals will give us a picture of the Universe when it was less than a second old (about

---

[5]These numbers are given for illustrative puroses only. See the official LIGO documents for the exact numbers.

$10^{-25}$ seconds). The prospect of viewing the universe at such an early epoch – an epoch extremely difficult to view by other means – is not only incentive to develop the most sensitive detectors, but also to develop optimized search algorithms. Developing such algorithms will require broadening the gravitational-wave data analysis community to include more scientists with expertise in the physics of the early universe.

2. **The Inverse Problem: Mapping the Spacetime of a Black Hole Coalescence.** When two spinning black holes coalesce, the spacetime undergoes highly dynamic, swirling contortions. The time series of the gravitational-wave amplitude from such an event contains an imprint of this violent motion. Solving the "inverse problem" by disentangling the information stored in the waveform will give us a picture of spacetime in a regime not visible by any other means. We will be witnessing a pure display of gravitational physics in its most non linear regime.

   Our ability to decipher the signal and map the spacetime contortions that produced it will depend on the amount of information about the coalescence that we are able to record and coherently analyze. The full inversion will certainly require data from many detectors around the world, thus giving us strong incentive to cultivate international collaborations. It will also require special multidetector algorithms, thus motivating a continued effort in algorithm development.

   In order to solve this "inverse" problem for a black hole coalescence, we will also need a better theoretical and numerical understanding of the the nature of black hole coalescence than we have today. This will require vast computing resources and a tight coupling of the numerical relativity community to the gravitational wave data analysis effort.

3. **The Inverse Problem: The Unexpected and the Unknown.** As more sensitive detectors increase the volume of the space we are able to see by a factor of a million, one of the sources we are most likely to detect is *the unexpected source*, or *the unknown source*. Finding such signals certainly places a high demand on our effort to develop unbiased search methods, but it also places a high premium on our ability to decipher such signals: to solve the inverse problem with little information about the nature of source. This will require us to reconstruct the warpage of spacetime that produced the signal from the measured data stream. Such analysis may stretch the bounds of our theoretical understanding of gravitational astrophysics by requiring us to find new solutions to the gravitational field equations for new types of compact gravitating bodies. This analysis will also benefit from other signal analysis disciplines that search for unknown signals.

4. **Measuring the Nuclear Equation of State.** A binary neutron star system spiraling toward coalescence is the prototypical source for LIGO data analysis: a well understood signal, a promising event rate, and detectable by straight-forward use of matched filtering. However, as the two stars near each other and begin to distort or tear apart, the late-time signal will be imprinted with information about the the nuclear equation of state. These equation of state measurements (or, more precisely, the measurement of the mass-to-radius ratio of the star) may allow us to rule out proposed equations of state. These observations will give us a glimpse of the nuclear matter not only at high energy, but at high density as well. As these

signals will occur at the high-frequency end of the LIGO sensitivity curve, they will require special algorithms and additional computational resources to pull the signal from the noise.

5. **Measuring the Hubble Constant.** As the detectors become more sensitive and we transition into a study of the nature and the statistics of the sources, one potential astronomical observation is a measurement of the Hubble constant. As two neutron stars spiral together, both the overall amplitude and the frequency sweep encode information about the redshift factor. When enough observations are made that statistical inferences can be drawn, the Hubble constant can be measured. Unlike electromagnetic measurements of the Hubble constant, this purely gravitational observation will not be biased by attenuation of the signal as it passes through opaque material.

6. **Multi-Messenger Astronomy.** In the early stages of LIGO observations, it will be useful to correlate (perhaps after the fact) the detector output with other astronomical observations, such as supernova, neutrinos or gamma-ray bursts. However, after we have detected signals and understood the sources, LIGO (and similar gravitational wave detectors) should be able to participate in real-time observations of the sky. Gravitational wave detectors have an advantage over electromagnetic observations in that they do not need to be aimed: they see all the sky, all the time. This means that gravitational wave detectors can form an "early warning" system for events such as supernova. The ability for LIGO to participate in the area can be greatly enhanced if the data from several gravitational wave detectors can be coherently analyzed in real-time. This ability will require very fast computer networking capability between the sites.

## 7.2   A Practical Look to the Future

The possibilities outlined in Section 7.1 above are very exciting, yet it is unclear which lines of research will bear fruit. Therefore, it is important to take a *practical look* to the future and embark on a methodical, sustainable research program which enhances – and flows naturally from – the immediate data analysis needs and yet puts the collaboration in a position to seize dramatic opportunities when may come along. The areas where we need to make an investment are obvious: software, computing hardware, computer networking, and people.

### 7.2.1   Enhancments to the Collaboration

- **Cross Disciplinary Outreach.** All of the items mentioned in Section 7.1 can profit greatly from researchers with broad expertise. For example, as our computationally intensive analyses (such as the inverse problem) migrate to the grid environment, the collaboration would benefit from interaction and collaboration with computer scientists. Interpreting measurements of the Hubble constant would benefit from interactions with astronomers. Our efforts to devise algorithms to decipher signals from unknown events can profit from researchers in other disciplines with data analysis problems similar to ours, e.g. engineering, speech analysis and oceanography. However, to seize these opportunities when the observations are made, such cross disciplinary collaborations should be developed now.

- **International Outreach.** The LSC is natural forum for international collaboration, especially in light of recent data-sharing agreements made by LIGO and GEO. For example, many of the items mentioned in the Section 7.1 will require concurrent data streams from many detectors. It will also require those familiar with each instrument to help glean as much information as possible from their detector output. This necessitates a strong international ties.

- **The LIGO Vistors Program.** One very practical way to support the growth and broaden the base of knowledge in the collaboration is the LIGO Visitors Program. We strongly endorse it. This program has proved to be an effective way of reaching out for expertise from broad a range of scientists and engineers from all over the world.

### 7.2.2 Improvements in Detection Algorithms

Because LIGO measures the amplitude of the gravitational wave, even small increases in sensitivity result in significant changes in event rate. For example, a 25% improvement in sensitivity can increase the event rate by a factor of 2 or make a corresponding change in an upper limit. The sensitivity can be improved not only via advances in detector configuration, but also by improved detection algorithms.

- **Extended searches.** Development of advanced algorithms for binary inspiral and periodic sources will open more of the gravitational wave sky in this branch of the research which is both software and hardware limited. To take advantage of improved detector sensitivity at low frequency, new inspiral search algorithms will be required that accommodate much longer-lived signals, and better computing facilities will be needed to handle the larger template banks.

- **Modeling of astrophysical sources.** Research into predicting gravitational waveforms of astrophysical sources will continue to play a critical role in the design of search filters. For example, the collaboration should consider expanding its scope to include numerical efforts to determine the waveforms from colliding black holes with spin angular momentum. When addressing questions about the nuclear equation of state during the late stages of neutron star coalescence, the collaboration would benefit from close ties with researchers with experience in numerical relativistic hydrodynamics.

- **Improved visualization techniques.** Automated pattern recognition has been developed for speech recognition and oceanographic research may provide new methods to diagnose the detectors as well as to search for unmodeled gravitational wave sources. The methods used in these fields should be incorporated into our analysis effort.

### 7.2.3 Improvements in Computer Networking

Because LIGO's data rates are fixed at around 9 Mbytes/sec, and the speed of the national and international networking infrastructure continues to improve exponentially, easy access to LIGO data should become available in the long term. This will enhance our ability do coherent real-time analysis of signal from many sites. The collaboration should act now to actively grow its participation in these efforts.

- **GriPhyN and iVDGL**

  The GriPhyN (Grid Physics Network) Project and the iVDGL (international Virtual Data Grid) Project are two promising avenues into grid computing and grid data management. The LSC should actively participate in these and similar projects.

  The tools being developed by these projects will help us not only obtain the necessary computer cycles to do our larger data analysis problem but also with porting data across the network swiftly and robustly.

- **Dedicated operations linking LIGO interferometers with other major gravitational wave projects worldwide.**

  Research is required to learn how to optimally utilize the emerging available network of interferometers and bars world-wide in order to realize the greatest possible science potential promised by these machines. Coordinated operations of the international network of gravitational wave detectors (GEO, VIRGO, TAMA, ACIGA, bar detectors) is needed in order to localize sources and to gain polarization information on the observed sources. Work is required on the definition, design and implementation of long range data mirroring, data exchange and data merging techniques for gravitational wave data analysis.

- **Multimessenger Astronomy: Trigger and Event Data Exchanges.**

  Work on improved networks will also enhance the ability of the gravitational wave detectors to provide a trigger to other astrophysical observations after an impulsive event has been detected. A model for this is the Supernova Neutrino Early Warning System (SNEWS) which has been set up to provide alerts if neutrino bursts associated with supernovae are detected.

### 7.2.4   Improvements in Computer Hardware

Improvements in computer hardware will enhance the effectiveness of the LSC data analysis activities. The rapidly-decreasing price of commodity computer hardware and the concurrent development of very cost-effective parallel computing architectures such as Beowulf systems should make it feasible for different LSC groups to make timely and effective contributions to the overall computing infrastructure needed to analyze LIGO data. These efforts will benefit from development efforts in other fields to create software and hardware configurations that can handle these enormous data sets.

- **Broader band inspiral binary systems.** Searches for inspiraling binary systems over a wider range of system masses and spins would be enabled by faster computation. The amount of computation power required grows as a rapid power of the lower-mass limit of the search: currently LIGO's data analysis facilities are scoped to carry out a search down to 1 solar mass (10 Gflops). A search for objects to a lower mass limit (0.1 solar mass) would require $\approx 1$ Tflop.

- **Unbiased search for periodic sources** $\approx 1$ Tflop computer could carry out an all sky searches for CW/pulsar signals to within about a factor of three of the limit of instrument

sensitivity. Additional computational power would make it possible to approach the instrument sensitivity, and also consider larger ranges of spin-down parameters.