LSC Data Analysis White Paper
Draft III

Bruce Allen, Sam Finn, Tom Prince, Keith Riles, Rainer Weiss

July 6, 1999

Contents

1 Executive Summary

The LSC Data Analysis White Paper provides an overview of the data analysis program planned for the LIGO I data runs and outlines a baseline strategy for the future. The White Paper will be updated yearly as the results of the research become available.

The scientific program of the LSC is to test relativistic gravitation and to open the field of gravitational wave astrophysics. The initial effort is designed to understand the detector and execute searches for astrophysical sources of all types: impulsive, periodic and a stochastic background. The analysis is designed to make detections as well as to set upper limits.

The analysis program has the basic components:

- calculation of parametrized source waveforms from the astrophysics
- characterization of the detector: calibration, sensitivity to the environment
- estimation of the detector noise
- design of optimal filters and efficient detection algorithms
- modeling of the detector and the source to establish detection efficiency and errors
- generation of triggers and vetoes from ancillary and environmental measurements
- tests for confidence using multiple LIGO (and worldwide) detector coincidence

The data analysis will be carried out by collaborating groups within the LSC that propose to take on specific scientific projects. A communal LSC software development and test program will be organized and guided by the LSC Software Coordinator. The LSC Spokesperson will work with the LSC Data Analysis Committee composed of the Laboratory Directorate, the Software Coordinator and the LSC committee chairs to organize the data analysis program.

The large quantity of data collected by the LIGO detector, defined as all three interferometers and the ancillary instrument and environmental channels, needs to be reduced to intellectually manageable levels. A function of the LSC is is to design useful reduced data sets for various scientific and technical purposes. One classification of reduced data sets is described in the White Paper.

The LSC Software development is being carried out at many of the collaborating institutions. The ability for software generated by different groups to be compatible rests on setting guidelines as described in the LIGO/LSC Specification and Style Guide which establishes software structure and testing standards. A LIGO/LSC Analysis Library is being developed which consists of both elemental and more complex modules to be used in the analysis. The library and tests of its programs are maintained and organized by the LSC Software Coordinator.

The overall data analysis pipelines will be tested in "Mock-Data Challenges" carried out by teams of LSC members organized by the LSC Software Coordinator. The "Mock-Data Challenges" are coordinated with major releases of LIGO Data Analysis System (LDAS) and LSC software.

The LIGO Laboratory plan for computational infra-structure and network utilization assumes

- on-line and real time analysis will take place at the Hanford and Livingston sites.
- sufficient computing capability exists at the sites to carry out impulsive searches to establish triggers.
- the major tape archive will be at Caltech which supports a limited number of external users
- reduced data sets will be directly available by network to remote LSC users
- a user is expected to have a minimum hardware capability.

Advances in the data analysis program for the longer range include

- New detection software and source modeling to improve the sensitivity of the searches and, thereby, the event rate and depth of the search in the universe.
- Software development and hardware changes (evolution) to allow a larger search range of the parameters for inspiral sources and ultimately an unprejudiced all-sky search for periodic sources.
- Utilization of improved blind-search techniques to help in the detection of unmodeled sources.

The anticipated advances in network speed and in storage technology allow contemplation of

- Full transmission of and fast access to the entire LIGO data set.
- The on-line coupling of the world-wide gravitational-wave network.

2 Science Overview

Introduction

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.

Neither of these goals can be accomplished without a major effort to understand the detectors.

In planning for LIGO I data analysis, we assume that

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). Particular attention is paid to the detection of serendipitous sources (*i.e.,* sources entirely unanticipated). Recognizing current theoretical prejudice, the data analysis approach is capable of placing upper limits on signal strengths or rates in the event of non-detection; however, it is also sufficiently flexible to recognize and permit the characterization of strong, serendipitous signals.

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, which can be tested by observations with LIGO and other gravitational wave detectors providing there are high signal to noise detections.

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. In general relativity, the late-time radiation is a superposition of several damped normal modes, whose frequency and damping constants are determined entirely by the final black hole's mass and spin. Observation of any single overtone gives, in the context of general relativity, a measurement of the black hole mass and spin. Observation of a pair or more overtones must yield the same masses and spins: 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. LIGO 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.)

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.

Discrete gravitational wave signals detectable by LIGO 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 a 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 organized by source:

Compact binary inspiral: to measure or place an upper limit on the rate of compact binary inspiral, and to characterize the source of detected binary inspiral radiation. With strong signals to test strong field dynamics and, if neutron stars, to study the supernuclear equation of state of the matter comprising the star.

Gravitational waves and gamma-ray bursts: to measure or set limits on the in-band gravitational wave power associated with gamma-ray bursts.

Black hole formation: to observe stellar mass black hole formation, or set limits on its rate as a function of the black hole mass and energy radiated gravitationally. If the radiation associated with the formation of a black hole is observed, the black hole mass and angular momentum will be quantified and, to the extent possible, general relativistic predictions tested.

Supernovae: to observe the gravitational radiation 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: to search for neutron stars formed in supernovae. The stars are born 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 S/N 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: to 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.

Pulsars and rapidly rotating neutron stars: to observe or set limits on the power radiated by known, young pulsars and by previously unidentified rapidly rotating neutron stars at certain, 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 sky, the so called unprejudiced search for periodic sources.

QPOs and LMXBs: to 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: to search for the presence of a cosmological stochastic gravitational wave signal, either bounding or setting an upper limit on the in-band signal power.

3 Detector Characterization

Introduction

Data analysis requires a systematic understanding and characterization of the detector: its response function, noise behavior and sensitivity to the environment. The confidence associated with source detection or upper limits for detection depend on detector performance characteristics, including: power spectra, the probability distribution of the detector output, stationarity of the noise 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 online and before archiving. GDS will establish rudimentary performance diagnostics and will have 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) will also cause alarm messages to be sent to the control room. It should be noted that many offline monitoring tasks are expected to migrate upstream to online diagnostics, as experience and confidence in the algorithms increase. The DMT is also likely to provide the first level of data reduction.

The third level is offline monitoring which includes detailed performance characterization, transient analysis and statistics and trend analysis. 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 tools to carry out Monte Carlo calculations of the system to establish the confidence of a detection.

Detector characterization will be carried out mainly at the observatories, using the full data set. The algorithm development and testing will take place at many locations in the collaboration, as shown in Table 1. In addition, it may be necessary to carry out more refined characterization in periodic "reruns" over the archived, reduced data at Caltech, and it will be useful to carry out limited detector characterization at LSC member's 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.

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

Detector Characterization

Task Category	Priority	Institutions
Online Diagnostics/Measurements	1, 2, 3	Caltech, MIT
Offline Monitoring Infrastructure	1, 2	Caltech, MIT
Environment Monitor (hardware)	1, 2, 3, R	Caltech, LaTech, LSU, MIT, Uoreg, PSU
Line Identification	1	AEI, LSU, Umich, PSU, UWM
Instrument Correlations	1	AEI, Caltech, MIT, UWM
Environmental Correlations	1, 2, 3, R	LaTech, LSU, Umich, Uoreg, PSU, Syr, UWM
IFO State Identification	1, 2, 3	Caltech, Umich, MIT, PSU
IFO/IFO Correlation	3	
Transient Id/Analysis (instr)	1, 2, 3, R	UFL, Umich, PSU, UWM
Transient Id/Analysis (envir)	2, 3	Caltech, LSU, MIT, Uoreg
Time/Frequency Analysis	1, 2	Caltech, MIT
Phenomenological Modeling	1, 2	Caltech, MIT, PSU
End-To-End Modeling	1, 2	Caltech, UFL, MIT, Pisa, PSU
Customized data sets	1,2	Uoreg

Table 1: The program for the development of detector characterization algorithms and software is prioritized in the table by the key:

1 – needed at start of 2-km commissioning (10/99),
2 – needed during 2-km commissioning (5/00),
3 – needed by six months before science run (6/01), and
R – research area for advanced LIGO.

A much more detailed version of the table with task breakdowns, estimated FTE-months requirements and the names of individual scientists responsible for the effort can be found at
http://www-mhp.physics.lsa.umich.edu/ keithr/lscdc/tasktables.html

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.

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. Studies that will be needed include the influence and reduction of narrow spectral peaks in the data such as electrical line contamination (60 Hz & harmonics), suspension fiber violin modes, internal mirror resonances and 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 will be carried out. Techniques need to be 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 such 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 generally, they rely upon time-frequency analysis.

Offline Transient Analysis

It is necessary to identify and record transients 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: suspension wire relaxation, 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). As experience with the interferometers grows, it should be possible to classify the vast majority of the transients via an event catalog.

Data Set Simulation

Simulation includes both near-term phenomenological modeling to test monitoring algorithms and far-term *a priori* Monte Carlo 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.

4 Astrophysical Source Detection

Overview

Each type of astrophysical search will have a data analysis pipeline, whose input is data and diagnostic information from the detector(s), and whose output is a list of potential source candidates. Each stage of the pipeline makes cuts and selections, passing smaller amounts of data to the next stage. Some of the events which pass the cuts and selections will also be recorded in a 'metadata' database .

The different data analysis pipelines will have many common elements, particularly at the input end, where measures of data quality and detector performance are most important. The later stages contain more specialized discriminators, which make cuts and selections based on how well the signals match the posited sources. The design and characterization of these filter pipelines and discriminators is an optimization problem (for example minimizing false dismissal rates for a given false alarm rate). This will be done using Monte Carlo simulations on a mixture of real and simulated data.

In general, the most effective means of gaining detection confidence is the observation of a signal in two or more independent detectors. While the beam patterns, polarization sensitivities, and frequency response of the non-LIGO detectors differ significantly, the LSC hopes to work with them to gain increased confidence and sensitivity.

The near-term program in source detection is divided into four main categories: inspiral, uncharacterized, CW, and stochastic background. The goal is to have basic searches in place and working when the instruments reach a 10^{-20} strain sensitivity in November 2000. The CW and stochastic background searches will be carried out offsite; the other two searches will be distributed between onsite and offsite. During the following year, while the instrument sensitivity reaches the design goal of 10^{-21}, the focus will be on testing, characterizing, and improving the methods.

This program is summarized in a set of four tables, which prioritize the necessary work and divide it into tasks. One or more research groups will work on each task. In each case, one group will be identified as having the ultimate responsibility to ensure that the task is completed on schedule, and to coordinate other groups participating in the development.

The development and implementation work will be carried out at the individual LSC sites, using the resources available at those sites. When the development has

Inspiral Source Searches

	Hierarchical	Coincident Event List	Discriminator	Combined Inspiral, Merger, Ringdown	Noise Power Spectrum Changes	Multi-Detector Analysis	Template Generation
Priority	1	1	1	2	2	1	1
FTE (Code+Test)	6+3	2+2	1+4	1+1	2+3	3+4	2+2
FTE (Science)		2		2	1		
AEI							X
Cardiff	X						X
Caltech			X				X
Cornell			X				
Guelph							X
Penn State	X	X				X	
TAMA	X						
UWM	X	X	X				X

Table 2: Tasks and group assignments for inspiral source searches. FTE's are shown in person-months.
Priority 1 tasks are essential and must be completed by November 2000.
Priority 2 tasks are useful, and should be completed by November 2001.

reached the stage where it needs to be run on CACR/LIGO hardware, the group will, through the LSC, obtain access to the necessary resources for development, testing, and production.

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 Section 4. 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 ~ 300 Hz for ~ 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. The plan is shown in Table 2.

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 filters near the region of interest, and a higher SNR threshold. Studies assume that the detector noise is Gaussian, derive 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 of 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 χ^2 statistics developed for analysis of interferometric data and for resonant mass detectors have proved useful in reducing false alarm rates. Discriminators which see 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 (automatically) comparing event lists produced by a filtering process at different sites , and selecting those which match certain criteria. These include arrival time differences less than 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 mass of approximately 25 $M_{\odot}$. For NS/NS binaries, the volume corresponds to a sphere of ≈ 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.

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

Unmodeled Source Searches

	Time/ Frequency	Power Monitoring	Time Domain Thresholding	Pulse Matching	Two-site Correlation
Priority	1	2	2	1	1
FTE (Code+Test)	3+1	1+1	1+1	2+2	1+1
AEI	X				
Caltech	X				
Cornell		X			
LAPP		X	X	X	
UWM	X	X		X	
Penn State					X
Italy (1)	X				
Poland	X				
Italy (2)	X				

Table 3: Tasks/group assignments for unmodeled source searches. FTE's shown in person-months. Priority 1 tasks are essential and must be completed by November 2000. Priority 2 tasks are useful, and should be completed by November 2001.

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. 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.

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. The near term program for detection of unmodeled sources is shown in Table 3.

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. This method needs to be ported to the LDAS environment. A related technique uses short FFTs to monitor energy in particular frequency intervals.

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. 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 has been developed.

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.

CW Source Searches

	Directed known phase	Directed unknown phase	FFT stack/slide or Hough Hierarchical	Robust Algorithms	Discriminators	Multiple Detector Analysis
Priority	1	3	1	3	1	2
FTE (Code+Test)	2+2		6+6		TBD	
FTE (Science)		TBD	1	TBD		1
AEI			X	X		
Cardiff	X					
Caltech	X		X			
LAPP						
Michigan	X	X				
Penn State						X
Stanford				X		X
UWM			X			

Table 4: Tasks and group assignments for CW (pulsar) source searches. FTE's in person-months.
Priority 1 tasks are essential and must be completed by November 2000.
Priority 2 tasks are useful, and should be completed by November 2001.
Priority 3 tasks are research.

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 unprejudiced 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. The near-term program is shown in Table 4.

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 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.

Establishing detection confidence.

CW gravitational wave signals will become apparent only after long integration times, so techniques may be needed to discriminate these from instrumental artifacts. These techniques should be capable of identifying wandering oscillators, and should also test for amplitude modulation consistent with the time-dependent detector response.

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.

Stochastic Background Searches

	Correlation Statistic	Robust	Maximum Likelihood	Non Gaussian Sources
Priority	1a	1a	1b	3
FTE (Code+Test)	2+2	1+2	3	
FTE (Science)		3	1	
AEI	X			
Cardiff				
Cornell	X	X	X	
LAPP				
Penn State	X		X	X
UT-Brownsville	X	X	X	X
UWM	X	X		X

Table 5: Task/group assignments for stochastic-background searches. FTE's in person-months.
Priority 1 tasks are essential and must be completed by November 2000.
Priority 2 tasks are useful, and should be completed by November 2001.

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. The near-term program is shown in Table 5.

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 it may not perform as well. In addition, existing work indicates that in the weak-signal limit of interest, these methods are equivalent to the simpler correlation statistic. Further work is needed to determine if these techniques are needed.

Establishing detection confidence.

Since stochastic background detection requires a pair of detectors, finding a signal with two detectors is not enough to establish confidence. Terrestrial effects, particularly correlated electromagnetic 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 another baseline. 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.

5 Data Products: Reduced Data Sets and Events

Few, if any, LSC scientists will want to work with large quantities of raw, unprocessed, uncalibrated data. As in most data-intensive projects, careful thought will be given in LIGO to the design of the data reduction procedures which produce "science data sets".

LIGO science data will be accessed through the LIGO Data Analysis System (LDAS). The data sets will be archived either in Frame Format or as a relational data base. LSC scientists can request data in Frame Format, or as LIGO Light-Weight XML format.

Archival and Reduced Data Sets

At 15 MB/s, the full 2 year LIGO data stream will yield about 1 Petabyte of data. With the planned duty cycle of 50%, about 500 Tbytes of raw data is anticipated. Four data levels can be specified:

Level 3: Whitened GW Strain Data. This is the best estimate of the whitened GW strain as a function of time, as free as possible from instrumental signatures. At a nominal 1 kHz sampling rate, a 2 year, 90% duty cycle data stream yields $\sim$55 Gigasamples, or about 200 GB uncompressed and perhaps 100 GB compressed. Level 3 data consists of IFO strain data plus the calibration, regression, and whitening coefficients needed to construct a best estimate of the whitened GW strain versus time. For three interferometers the total size of the Level 3 data set should be about 1 TB.

Level 2: IFO Strain plus Data Quality Channels. For more detailed scientific analysis, a 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 set will be about 1% of the full data stream, or about 10 TB in size.

Level 1: Archived Reduced Data Set. The lowest level archived data set will consist of all important IFO and PEM data channels stored in time-striped Frame format. In addition, data conditioning info will be archived including the usual regression, whitening, and calibration data. The total Level 1 data set will be about 10% of the full data stream or about 100 TB in size.

Level 0: Full IFO Data Stream. The full data stream will be available in frame format for about 16 hours after acquisition, but will not be archived. The processing from Level 0 to Level 1 is done continuously and in general there should be no need to keep Level 0 data.

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 (called "metadata"). 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 Events.** This includes information about the triggers and vetoes generated by the Global Diagnostics System (GDS) and Data Monitoring Tool (DMT) filters and information about the filter programs themselves. It also includes "astrophysical events" such as those generated by the binary inspiral, ringdown, burst, and periodic source analyses performed by LDAS. Events are considered a particularly interesting form of LIGO data and can be delivered to the user in LIGO Light-Weight format.
- **Non-LIGO-Generated Event Information.** There will be numerous sources of "events" besides those generated by analysis of IFO and PEM data. These include seismic alerts from external monitoring networks, electromagnetic storms, γ-ray burst events, neutrino events, UVOIR (UV, optical, IR) events such as supernovae, and events generated by other GW detectors.

- **CDS and LDAS Log Information.** Information normally kept in logs will be available electronically via the database.

6 LSC Organization of Data Analysis

The obligations that LSC members have made to the data analysis effort and the rights to access of the LIGO I data are defined in the Memoranda of Understanding between the LIGO Laboratory and the LSC member's institution. Broadly, rights to the LIGO I data are gained by making a substantial and recognized contribution to LIGO 1 construction, commissioning and or software development. The LIGO Scientific Collaboration is responsible for the data analysis, validation and scientific interpretation of the LIGO data. The intent is that LSC members will organize group efforts to study specific scientific problems with the LIGO data. The spokesperson of the collaboration will coordinate the data analysis effort. Analysis efforts will be initiated by proposal made to the collaboration by individuals or a group of members or organized by the spokesperson. The proposals will include:

- 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
  - the division of responsibility between the proposers
  - students assigned to the effort and PhD theses expected
  - an estimated schedule for completion
- an outline of the publication(s) that are expected to arise from the analysis

The LSC software development, test and maintenance will be organized by the LSC Software Coordinator. The functions of the Coordinator include:

- define and manage the software development across the LSC
- maintain the LIGO/LSC analysis library
- chair the LSC software change control board

The LSC Spokesperson will work with a committee consisting of the Laboratory Directorate, the LSC committee chairs and the LSC Software Coordinator in evaluating proposals and in receiving advice to guide the LSC data analysis program. Publications resulting from the analysis will be reviewed and authorized by the entire collaboration as described in the LSC Publications Policy.

7 LSC Software Development

Overview

The LSC science analysis pipeline will be implemented from modular components that are validated and controlled as part of the LIGO/LSC Analysis Library (LLAL). All delivered software will conform to a standard that has been defined jointly by the LIGO Lab and the LSC.[1]

The LSC Software Coordinator has the principal responsibility for defining and managing the LSC software development effort. Verification and validation of LLAL components will take place at three levels: *(i)* compliance to standards, *(ii)* piecewise component tests and *(iii)* integrated tests of the analysis pipeline through Mock-Data Challenges.

The LIGO/LSC Analysis Library

The LLAL configuration is managed by the LSC Software Coordinator, who coordinates regular releases of the LLAL library with and between LDAS releases. Major releases will be scheduled to coincide with major LDAS releases (α-1 in Q2'99, α-2 in Q4'99, β in Q4'00 and V1.0 in Q4'01). 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 nonlinear — 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. LLAL consists of these atomic operations acting on these structured data types.

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

We expect that LLAL will evolve and grow with accrued data analysis experience. Changes to LLAL will be authorized by a Change Control Board 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.

[1] LIGO-T9900030.

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 LLAL 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.

Software Validation test that the software components can be integrated into analysis pipelines that can perform that analyses described in the science goals (§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 LLAL 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 LLAL 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.

8 A Usage Model for the LSC

The near-term computational infra-structure for the LSC depends upon the usage model adopted by the LSC to fulfill its computing needs. Here we define the usage model for the LSC through the period of the LIGO I science run.

LIGO has at least 3 distinct types of sites, each with its own functions and computing usage model.

- **IFO Lab Sites** (Hanford and Livingston)
- **Non-IFO Lab Sites** (Caltech and MIT)
- **Non-Lab Sites** (about one dozen)

The usage model for each site depends both on technical issues such as available network bandwidth and on the primary functions expected to be performed at the site. For convenience, we define 3 broad categories of usage. In the following categories, the local/remote perspective is taken from the perspective of the user siting at their terminal at the specified site. Bandwidth refers to the Wide Area Network (WAN) connection from the specified site to one of the LIGO data repositories.

- **(Local Processing/Local Data/Low-Bandwidth WAN)**: Workstation-based development and analysis using local data files typically obtained by ftp or tape transfer. A typical activity will be to request a small data file ($\approx$ 1-10 MB) from the archive and then analyze the data using programs running on the local workstation. The analysis and development activity may or may not use the LDAS software environment. Another example that falls under this usage model is local analysis of a moderately large data file ($\approx$ 1-10 GB) using data from a tape generated at the archive. It is envisioned that a large fraction of the code development and instrument characterization work of LSC scientists will fall under this usage model.

- **(Remote Processing/Remote Data/Low-Bandwidth WAN)**: Development and analysis on significant LSC resources via remote network access using an X-term or browser interface. A typical example is an LSC scientist connecting to the LDAS system at the Caltech archive site and performing an analysis on a multi-Gigabyte data set using the Beowulf Linux cluster. It is envisioned that code validation will be accomplished under this usage model as well as a large fraction of the computation intensive science analysis. The analysis will typically use the LDAS environment.

- **(Local Processing/Remote Data/High-Bandwidth WAN)**: Data analysis on a local workstation or supercomputer using remote data files provided via high-bandwidth networks (OC1 or greater) by the LIGO archive. This is not expected to be a common usage mode during LIGO I and there is currently no requirement for high-speed WAN connection between the LSC sites.

Usage Models at the LSC Sites

- **IFO Sites.** Because the highest priority of the IFO sites is the operation of the interferometers and the corresponding pipeline data analysis, the computing systems at the IFO sites will be oriented towards local access, with limitations on remote access. The CDS/GDS/LDAS systems will have their own LANs, not readily accessible from the outside. Remote access to the CDS/GDS/LDAS systems at the sites will be restricted by password and limited primarily to terminal interactions and small ftp transfers.

  Each IFO site will also have a General Computing LAN with an environment similar to that at Caltech and MIT Non-IFO sites. The General Computing LAN will be readily accessible from the outside, but may have limitations in the near-term because of network bandwidth (currently 1-3xT1).

- **Non-IFO Lab Sites, Caltech.** The primary functions of the Caltech site are to provide access to archival data and to perform detailed science analysis on the combined data from the 3 LIGO interferometers. Remote LSC users will be able search the LIGO database and select various types of science data for processing. The data may be processed using the Caltech workstations and Beowulf cluster connected via high-speed LAN to the archive, or by requesting a data transfer via tape or ftp. The LDAS system at Caltech is currently scoped to provide support to 5 high-bandwidth users simultaneously, assuming a mix of tape and disk data transfers. The system will support users at other sites who wish to use a high-speed WAN connection to their local computers, but the bandwidth for data delivery will depend on the connectivity between Caltech and the remote site.

- **Non-IFO Lab Sites, MIT.** The detailed usage model for the MIT site is TBD. Currently, MIT will have a Beowulf cluster for development and data analysis. If MIT acts as a mirror site for some portion of the archived data (e.g. the 1% strain + quality channels data), then it will likely also support usage by the LSC in the (Remote Processor/Remote Data/Low-Bandwidth WAN) mode, using the LDAS environment.

- **Non-Lab Sites.** The primary usage modes for the non-Lab sites will be in the (Local Processing/Local Data/Low-Bandwidth WAN) and (Remote Processing/Remote Data/Low-Bandwidth WAN) modes. It should be noted that local LAN bandwidth at the Non-Lab site may be quite high, but high-bandwidth connection to the Lab sites is not a requirement at this time. The primary requirement is to provide efficient remote access to data and LSC computational resources necessary to perform code validation and analysis of data, as well as an efficient local environment to perform code development.

  Some non-Lab sites may obtain or have access to significant computer resources for LIGO analysis. In this case, the resources should be open for use by remote LSC users who should be able to access the resources in the (Remote Processor/Remote Data/Low-Bandwidth WAN) mode.

Infrastructure Requirements

The usage model adopted by the LSC for LIGO I requires certain network, computing, and storage capabilities as an infrastructure for LSC data analysis. In addition, the infrastructure will require personnel support for maintenance and operation. The level of system and service support required will depend crucially on the usage model adopted.

We assume the following capabilities for the LIGO Laboratory computing resources:

- Computational capability: A minimum of 20 Gflops at each of the sites for pipeline processing and 40 Gflops at the Caltech archive site for science data analysis. The actual capability is expected to grow by a factor of 2-3 by the end of the science data run due to continued improvements in processor speed.

- Network bandwidth: Greater than OC3 from the Caltech archival site.

- Storage Capacity: 100 TB tape robot at the Caltech archival site + 1 TB disk storage. The disk storage should grow by at least a factor of 2 by the end of the LIGO I science run.

- Software and Application Environments: LDAS implemented on standard workstation configuration and on Beowulf Linux clusters.

The primary usage modes for LSC computing an non-Lab sites have been described above. The LDAS system has been designed to provide significant data analysis capability to LSC users in the (Remote Processing/Remote Data/Low-Bandwidth WAN) mode using typical network connections and a web browser or X-term interface. It has also been designed so that LSC users can select specified data and transmit it to the user via ftp or tape transfer. The minimum infrastructure requirements for non-Lab LSC sites are therefore the following:

- A standard LSC workstation configuration with the following approximate characteristics: 0.5 Gflop processor, 50 Gbyte disk, standard network connections, and a tape drive.

- A standard LSC software environment for the workstation which the LDAS environment, DB2 client software, and other software.

A requirement of the standard LSC workstation configuration is to efficiently support the remote processing usage mode by making it transparent to the user whether the computing resources being accessed are located locally or remotely. The standard configuration will also support efficient local processing and code development on the workstation through adoption of standards for tape drive, data formats, and LIGO software tools.

Some non-Lab sites may acquire or have access to significant computational resources. These resources should be accessible to LSC researchers and provide support for remote usage by others in the LSC.

9 Long Range Program and Anticipated Needs

The near-term research program ensures that within the limitations of the available manpower and computing resources, LIGO can carry out reasonably sensitive searches for the primary categories of expected sources. The most pressing need is to begin these activities early, so that during the commissioning phase of the LIGO detectors, the data analysis systems can be tested, debugged, and optimized.

In the longer term, LIGO's program will evolve toward increasing detection sensitivity and bandwidth and in the ability to widen the scope of the search. Eventually, when detections are made, the program will transform into a study of the nature of the signals and the properties of their sources.

Elements in a long range program are both in the intellectual development of improved understanding and software and in the exploitation of the improvements in the hardware.

1. **Development of improved detection algorithms**

    - *Improved sensitivity* Because the 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 through improved algorithms can increase the event rate by a factor of 2 or make a corresponding change in an upper limit.
    - *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. A relevant study is the influence on the data analysis of the improvements at low frequencies being projected for LIGO II which will extend the search at low frequencies by about a decade, to approximately 10Hz.
    - *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. Two examples are: a program to bound the waveforms of the recently hypothesized r-mode sources and NS-BH and BH-BH systems and the completion of the program to determine the waveforms from colliding black holes with orbital and spin angular momentum.
    - *The inverse problem* Research is required in the development of the computational techniques to fully utilize the dynamical information in the gravitational wave time series in a high signal to noise observation. The detected gravitational waves signals are field amplitudes rather than intensities and retain detailed information of the dynamics at the source. The full inversion will most likely require both position and polarization information determined from detections at multiple sites.

- *Improved visualization techniques* Automated pattern recognition as 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.

2. **Improved hardware**

   - *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 ≈ 1 Tflop.
   - *Unprejudiced search for periodic sources* ≈ 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.

These longer-term activities should develop naturally out of the LSC's near-term research program but will require a greater concentration of effort in software and theoretical development. A well placed investment is in the support of additional scientists interested in the software and data analysis of gravitational wave astrophysics.

Improvements in computer hardware and the bandwidth of communications networks 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. In common with some of the data from other fields, (most notably, high-energy physics) much of LIGO data has an *event independence* which allows the data to be efficiently analyzed in parallel. This suggests that the databases and tools which are used or might be developed for these fields have substantial overlap with GW detection.

Because LIGO's data rates are fixed at around 15 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.

But the next five years are crucial ones, and during this time the LSC needs to make a continued effort to improve access to the data and resources. For example, by the end of the first science run it may be possible to put all the LIGO data onto spinning media, and make it available anywhere within the US, at reasonable cost.

These improvements in networking and facilities will enable another critical step in the field's evolution by the full use of the international network of gravitational wave detectors (GEO, VIRGO, TAMA, ACIGA, bar detectors) to gain position and polarization information on the observed sources. 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 Network (SNNET) which has been set up to provide alerts if neutrino bursts associated with supernovae are detected.

We strongly endorse the LIGO visitor's program. This has proved to be an effective way of reaching out for expertise and assistance from the scientific and engineering community. Data analysis problems comparable to those encountered in gravitational wave detection occur in several research areas such as speech analysis, oceanography and other branches of observational astrophysics. The visitor's program is an effective way to bring individuals who have developed particular methods and abilities into close contact with the LIGO detectors and data.

It is our expectation that gravitational wave observations will become a standard part of astrophysical measurements in the next decade and add new and complementary insight into the nature of the universe. The most promising direction in which the field will develop is not easy to predict. It is, however well known, that those best prepared will be most likely to discover something new and enduring.