LIGO-T880009-

# 05

List of 3/1/89

# Control Systems and Data Analysis for Large Gravitational Detectors

A. Jeffries, P. Linsay, J. Livas

June 9, 1988

DRAFT DRAFT DRAFT

## Contents

# 1 Overview

Efficient operation of the LIGO receivers will require a control and data acquisition system of considerable complexity. We attempt here to outline a number of conceptual design issues to be considered.

The computational requirements in the analysis of gravitational wave data need be understood prior to system installation. The sheer quantity of signal data may require preprocessing and the installation of specific hardware.

# 2 Guiding Philosophies

We have attempted generate an operational philosophy to guide a first order design:

- Use existing electronic hardware standards and commercially available circuit modules. Minimize custom hardware.
- Clear and rapidly accessible graphical display of equipment status from various locations is desired.
- The control system should be as flexible and automatic as possible.
- Minimize the need for direct and local human contact of detector.
- In event of power failure the system should protect itself. Automatic alignment procedures, self-calibration, automatic time scheduling for different experiments should be considered.
- Future full scale remote operation from universities would be desired.
- Full data archive is necessary. Should it be done at the site or university?
- Adopt a standard data format. Retain flexibility for data analysis.
- Software update control for site computers.
- Expandable system for multiple detector capability

# 3 Hardware Issues

1. Bandwidth questions:
    - between end and mid to central stations
    - between central and control stations
    - between site to site
    - between university to site or university
2. Number of channels and sample rate.
3. Local small digital servos. Mass positions
4. Clock timing and data stamping
    - What are short term frequency stability requirements
    - What are long term frequency stability requirements
    - What are absolute and site to site timing requirements
5. items to monitor and flag:
    - GW signal from interferometers
    - GW signal from mid station interferometers
    - pressure along tubes and in stations
    - temperature at various points of tubes and tanks
    - laser parameters:
        - Output power
        - Cooling temperature
        - Frequency stability
        - Power spectrum of intensity fluctuations
    - seismic activity
    - A.C. power fluctuations
    - various valve positions

   - state of other site or other observatories
   - mass positions, mirror alignments

6. Computer/Manual control of:
   - beam and optical alignments
   - laser power and beam control. Safety issues.
   - stepper motor mass positions
   - Vacuum valves

7. Spare parts and backup computer policy

# 4 Software Issues

A number of different programs are envisioned:

1. GW signal data acquisition
2. "Quick look" analysis to exam quality of data
3. Ability to cross reference housekeeping and signal data streams
4. Data base system or catalog data for later analysis

# 5 Data Analysis

## 5.1 Introduction

The analysis of the data from a gravitational wave antenna has two distinct goals. The first goal is merely to detect the presence or absense of an event. Once an event is detected, astrophysical characterization can begin, and the analysis should be able to extract the location of the source, its complete waveform, and any polarization information.

Gravitational wave sources fall into three overlapping categories, each with different data gathering and analysis requirements.

1. Impulsive sources: supernovae, binary star coalescences; characterized by a short period signal, typically a damped sine wave, $\sim 1$ millisec in length, corresponding Fourier spectrum, broad about 1 KHz wide, centered on principle frequency of event

2. Periodic sources: pulsars, binary stars; characterized by a coherent signal at a single frequency. Time series very long, Fourier spectrum a spike at a well defined frequency and harmonics.

3. Stochastic source: remnant from the Big Bang, superposition of point sources; signal(=noise) remains after eliminating all other sources of noise from data. Need to cross correlate several antennas to see this. Hard to do since local noise has to be well understood. Need very long time series since signal to noise only improves as $T^{1/4}$. Won't discuss in detail because data collecting requirements not different from impulsive or periodic sources. Computation capabilities and systematics will limit sensitivity. Specific algorithms for stochastic analysis have not yet been worked out.

## 5.2 Outline

For impulsive and periodic sources we will discuss:

- algorithms
- data ratesand storage requirements
- computation requirements

It is assumed that a commercially available 16 bit A/D will be employed for the data collection, regardless of the type of source. The existence of a precision timing system incorporating overlapping cesium and rubidium time standards and synchronized between sites and to universal time via the Global Positioning System (GPS) or something similar is also assumed. Timing information should be recorded as part of the signal data stream, and it would be desirable to synchronize and trigger the A/D system with signals derived directly from the time standards.

## 5.3 Impulsive Sources

Once the intrinsic sensitivity of the antennas has been improved as much as possible, the best way to improve detectability for impulsive searches is to use additional information in the form of a predicted waveform or the output of another antenna. Using the output of another antenna would require cross-correlating short time series (about 50 millisecond to accomodate the time delay between antennas) from two or more antennas. The output of a single antenna could also be cross correlated with a prediction of the expected waveform or a set of templates approximating a class of waveforms. Cross correlation with a predicted waveform is known as matched filtering and results in the highest signal to noise ratio of the three methods, but suffers from the drawback that the prediction must be good. A set of templates loses only slightly in signal to noise over that of a matched filter, but can be used to search for a broad class of possibilities. Cross correlation with the output of another antenna would be expected to have the lowest signal to noise ratio of the three methods because both of the inputs contain noise (a template or a prediction are "perfect" signals), but would be less likely to miss an unexpected signal. Efficient computational techniques for performing cross correlation (the FFT) are well developed, as are statistical techniques for interpreting the results. 3?

A matched filter or template analysis on the output of a single antenna can only yield a detection of a waveform as received; the direction to the source, the polarization information, and the complete waveform cannot be extracted without the use of more than one antenna.

Cross correlating the outputs of two antennas is complicated because spatially separated antennas will have different responses to the same signal; the waves will be incident on the different antennas in different directions. The extraction of the gravitational waveforms from the separate responses is known as the "Inverse Problem" and needs further development. Massimo Tinto is studying this problem and the problem of the detection of sources and the extraction of information with a global network of antennas.

The data rates for impulsive sources are rather modest and are dictated by the choice of the operating bandwidth of a large antenna, and the Nyquist sampling theorem. For an assumed 2 kHz operating bandwidth, the sampling rate would be 4 kilosamples per second. The storage requirements are not so modest, at least 700 Megabytes per day at this sampling rate. This will increase if seismic monitors, etc., are also sampled and stored at comparable rates as it will if more than one antenna is operating at each site. The cross corelation can be done by use of array processors, and if a data link is available (9600 Baud is sufficient), the calculation can be done in real time since the time series will be rather short (200 data points for a 50 msec time sample at 4 kHz). A typical AP can do a 1024 FFT in 2 msec so that one could contemplate cross correlating each data burst with 25 templates in real time before reading the next burst into the AP's private memory. A bank of , say, 10 AP's could be at work simultaneously analysing overlapping data streams. Storage of the data for later retrieval will be most efficient on optical storage discs although the new VCR style tape drives may do just as well. It is the unanimous consensus of this committee that **ALL** data should be integrated with selected portions of the housekeeping data system and archived until signals are found and understood. We want to be able to go back and look if, for example, the neutrino detectors see a signal.

Techniques for cross correlating the output of a single antenna with a set of templates were developed by Dan Dewey at MIT for his PhD thesis. A summary of his analysis of the June 1985 data run for impulsive sources: It was done by using a template search since no other interferometers were operating at the time, making a cross correlation search impossible. Examination of the literature, revealed that most signal waveform predictions could be approximated by a damped sine wave, or even by a few cycles of a pure sine wave. A set of three level (+1,0,-1) square waves, not neccessarily a complete orthonormal set, was designed which covered the frequency spectrum of the data and made an acceptable approximation to a number of half cycles of a sine wave. A three level square wave is a good estimate for a sine wave and is much more efficient computationally. Since the input data is Gaussian noise, and cross correlation is a linear process, the output is also Gaussian. Consequently, the pulse height distribution made by cross correlating the data with the templates should be Gaussian distributed if the data is only noise. One then searches for events by looking for deviations from a normal distribution. For the data runs analyzed, there was no simultaneous monitoring of local environmental disturbances such as traffic and trains which are known to cause spurious signals, so the deviations from a normal distribution that were observed could not with confidence be attributed to signals of astrophysical origin.

Data analysis techniques to search for the signals from binary star coalescences were developed by Cheryl Smith at Caltech for her PhD thesis. Binary star coalescences have some of the attributes of both impulsive and periodic sources, but the signals can be detected with the highest signal to noise ratio in the last few milliseconds before impact and are thus more like impulses. The expected signal is a sinusoid with center frequency linearly increasing with time (termed a frequency "chirp") up to some maximum frequency which depends on the structure of the source. The signals can be detected by using a template to adjust the data sampled at equally spaced intervals in time so that it is sampled equally spaced in phase, and then performing an FFT to look for the pure starting frequency.

## 5.4 Periodic Sources

A general search for periodic signals presents different problems from those encountered in looking for impulsive sources. The algorithm is quite simple: one uses the FFT and looks for a statistically significant narrow peak. In principle, it is possible to improve the signal to noise ratio arbitrarily since the signal to noise improves as the square root of the integration time at least until the signal bandwidth exceeds the integration bandwidth. The broadening of the signal due to the Doppler shift induced by the earth' s rotation and orbital motion is the major problem, and there is an additional broadening from the amplitude modulation which occurs as a source sweeps through the non-isotropic response lobes of the antenna. It is necessary to remove the Doppler shift in order to integrate a periodic signal for times longer than 1/2 hour. It is possible, for long enough integration times, to both detect a source and extract all of the relevant astrophyics from the output of a single antenna because the passage of time allows the detector to look at the source from different directions.

There are Three types of analysis problems: i) a search for a known frequency signal from a specific source, ii) a search in a specific direction for sources of unknown period, and iii) an all sky search for signals of unknown period. The first is straightforward: on must continuously adjust the sampling rate to keep track of the known phase of object under scrutiny. Adjustment of the sampling rate compensates for the Doppler shift of the signal. This is described in the thesis of Mark Hereld at CalTech to search for fast pulsar. A simple Fourier transform of the data is all that is required in order to look for the signal. The frequency resolution of this method is limited by the window function describing the gaps in the data. One should seek to be on the air as continuously as possible so that there are few or no holes in the window to degrade the resolution.

An all sky survey requires sampling data in an unbiased way and then undoing the Doppler shift in the analysis. The idea is simple. One would like to do in software what is done in hardware for a known source, that is, sample at constant intervals of phase rather than constant intervals in time. The FFT of a signal sampled at constant phase intervals has its Doppler shift removed and is as narrow as it can be. The Doppler shift causes the frequency of a signal to vary with time as

$$f(t) = f_0 \cdot \left(1 + \frac{v}{c} \sin(\theta) \cos(\omega t + \phi)\right)$$

and hence the phase to vary as

$$\Phi(t) = f_0 \cdot \left(t + \frac{v}{c\omega} \sin(\theta) \cos(\omega t + \phi)\right)$$

Here $\theta$(polar angle) and $\phi$(azimuth) fix the position of the source in the sky, v/c is the earth's orbital speed and $\omega$ is $2\pi$/year. The complication due to the earth's rotation is not shown here but a similar term is added. By choosing sampling times $t_0, t_1, t_2, \ldots$ at nonuniform intervals, the step size in $\Phi$(t) can be kept constant, i.e., by sampling at nonuniform intervals in time one can sample at uniform intervals in phase. Notice that the correction is frequency independent. All frequencies can be searched for at once given a particular direction on the sky. However, only one spot in the sky can be searched at one time due to the $\theta$ and $\phi$ dependence. Furthermore, the pixel angular size decreases linearly with the integration time, requiring more searches. We probably have to pick and choose points of interest. The algorithm involves computing sampling times $t_0, t_1, \ldots$ then linearly interpolating data between samples taken at uniform intervals in time. Since we would be interpolating, we estimate a need to oversample by at least a factor of ten to prevent introducing spurious modulation. This means sampling at least at 40 KHz for a 2 KHz bandwidth interferometer. The data storage is at least 7 Gb per day. These requirements may be relaxed by performing some of the calculations in parallel, using special purpose hardware. This possibility is discussed in more detail at the end of this section. A million second integration requires a $4 \times 10^9$ point FFT. This is quite formidable, probably not cpu limited on a Cray or equivalent but i/o and memory bound due to shuffling of data on and off disk even with very efficient data shuffling FFT's available. The minimum arithmetic precision required for these large transforms may be estimated by requiring that the size of a step in the phase factor, $2\pi/N$, be resolvable. N is the number of data points and is the product of the sampling rate and the integration time. The number of bits needed to resolve a single step in the phase factor is given by

$$b = [\log_2(N/2\pi)] + 1.$$

The square brackets denote the greatest integer function. Table 1 is a tabulation of this factor for several integration times and a 4 kHz sampling rate.

**Table 1**

| Integration time | N | b |
|---|---|---|
| 1 year = $3.2 \times 10^7$ secs | $2^{37}$ | 35 |
| 1 mo. = $2.4 \times 10^6$ secs | $2^{33}$ | 31 |
| 1 week = $6.1 \times 10^5$ secs | $2^{31}$ | 29 |
| 1 day = $8.7 \times 10^4$ secs | $2^{28}$ | 26 |

Once the Fourier transform is complete, one must scan the spectrum with a peak finding routine to make distributions and select candidates for possible sources. In our experience, 15 minutes of data sampled at 20 KHz required 12 minutes of (CPU time) analysis on a Cray 2 with 256 Megawords of storage. Approximately 4 minutes were spent reading in the data and converting it to the proper format, and the FFT itself required only 71 seconds. The remainder was spent in performing the Doppler correction and scanning for peaks.

This technique was developed at M.I.T. by Jeff Livas for his PhD thesis.

We do not recommend removing the Doppler shift in software for the periodic survey for the LIGO except for developmental work and for experiments in which maximum flexibility is needed. Instead we recommend that individual points in the sky be chosen and the Doppler shift removed in hardware by sweeping the sampling frequency appropriately and in parallel for several different directions. Special purpose hardware can be built to do the FFT, there is probably no need for our own personal CRAY. The data storage and manipulation requirements are a big problem and must be carefully considered in the design of any special purpose hardware.

An alternative to using an FFT is to coadd the time series with a given period. With a long averaging time this will also find a signal. It has the advantage that all harmonics of the signal contribute simultaneously. It has the disadvantage that it only can pick out one frequency component at a time, one must keep many different coadded time series each with a different period. This technique is probably most useful once a periodic source is tentatively identified from the Fourier spectrum. In fact, once a source is tentatively identified, there are a wealth of very sophisticated signal processing techniques which may be brought to bear on signals of approximately known form.