# LIGO Poster Presentation

## *International Conference on Computing in High Energy Physics*

**Berlin, Germany**

**April 7-11, 1997**

***James Kent Blackburn***

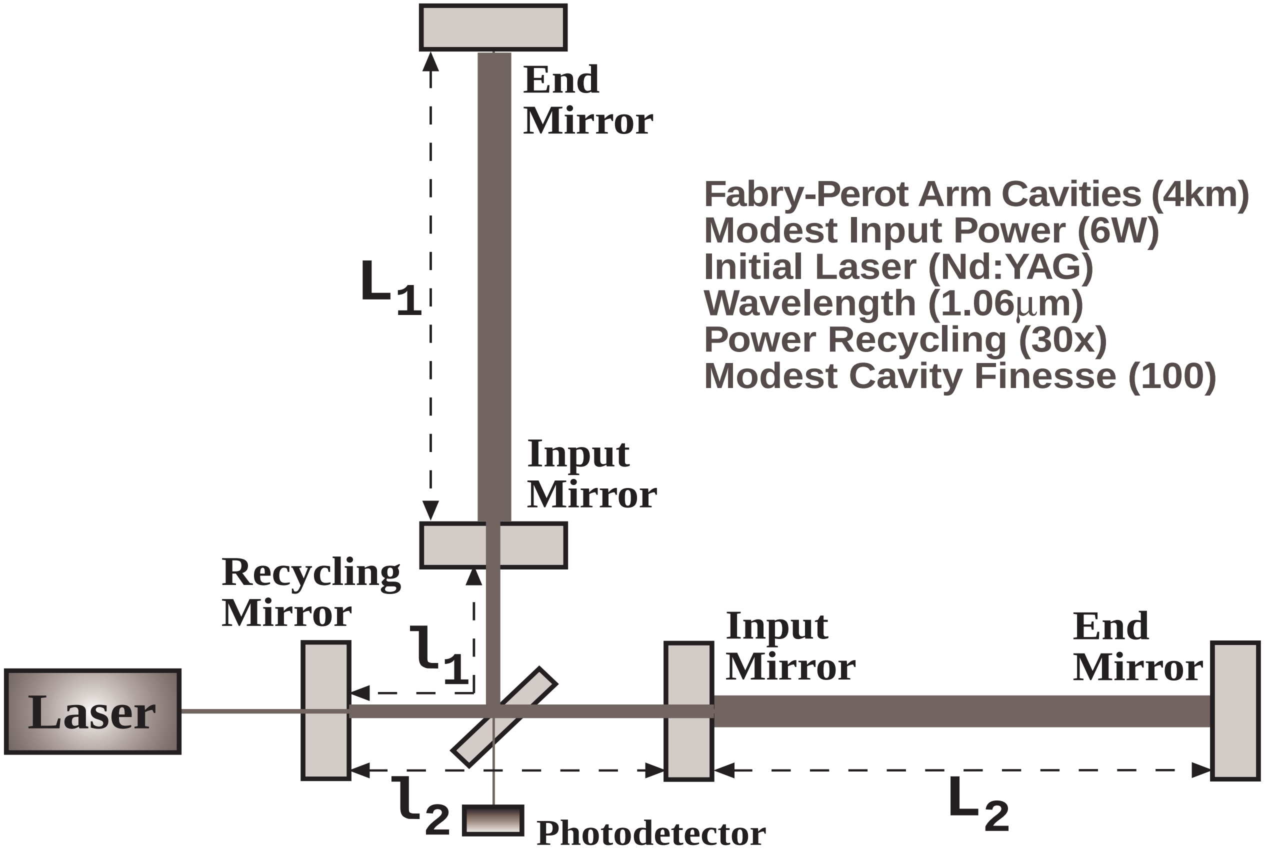
# Modeling LIGO Data Analysis

James Kent Blackburn

***California Institute of Technology***
***Physics, Math, Astronomy Division***
***LIGO Project, Mail Code 51-33***
***Pasadena, CA 91125 U.S.A.***

The Laser Interferometer Gravitational Wave Observatory (LIGO) will search for direct evidence of gravitational waves emitted by astrophysical sources in accord with Einstein's general theory of relativity. State of the art laser interferometers located in Hanford, Washington and Livingston Parish, Louisiana will unambiguously measure the infinitesimal displacements of inertially isolated test masses which convey the signature of these gravitational waves. The initial commissioning of LIGO will consist of three interferometers operating in coincidence to remove spurious terrestrial sources of noise. These initial LIGO interferometers will search for gravitational wave signatures with very low event rates and low detection signal to noise ratios out to distances as great as 300 million light years. Data will be collected continuously from the three interferometers at rates as high as 16 megabytes per second. Data analysis for LIGO ranges from the extremely simple for the case of intense short duration supernova bursts which will rely on site to site communications to share data used in coincidence, to state of the art parallel and distributed computing utilizing several hundred nodes to detect the chirp signals from neutron star/black hole binary systems, to petaflop computers of the future needed by the all-sky periodic source surveys.

# Initial LIGO Interferometer Configuration



Fabry-Perot Arm Cavities (4km)
Modest Input Power (6W)
Initial Laser (Nd:YAG)
Wavelength (1.06μm)
Power Recycling (30x)
Modest Cavity Finesse (100)

Initially LIGO will be configured as a Michelson interferometer with Fabry-Perot arm cavities and a recycling cavity. The interferometers are being designed to detect RMS displacement motions on the order of $10^{-18}$ meters. To achieve this level of sensitivity, high power stabilized Nd:YAG lasers will provide input to each recycling mirror. All optical components contributing the sensitivity of the instrument will be suspended as pendula and isolated seismically to reduce coupling to sources of thermal noise and ground motion. The optical path lengths of the interferometer will be maintained by a servo-system, keeping the interferometer phase locked on a particular dark fringe at the output port.

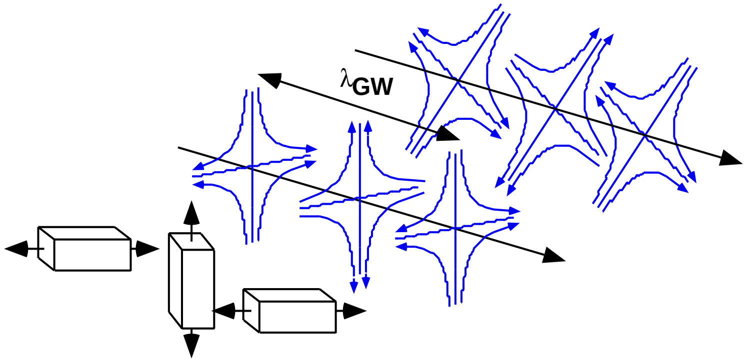
# LIGO Data Channels and Rates

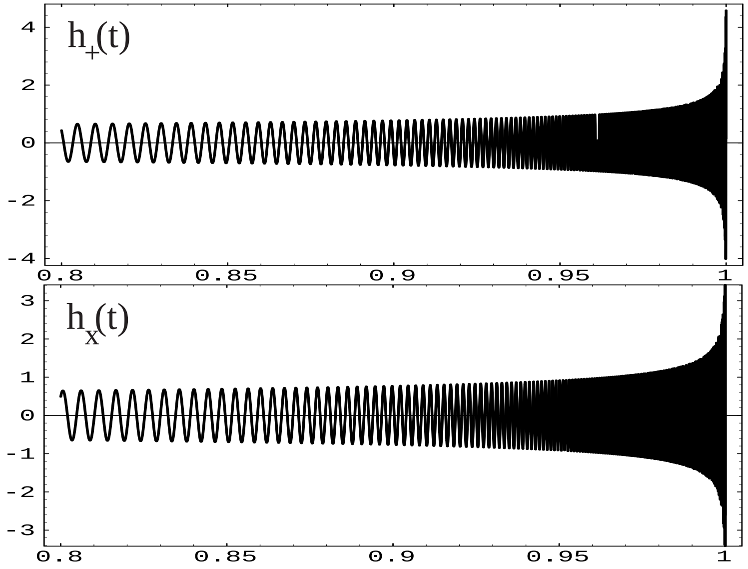
**Estimates of LIGO IFO and PEM channel counts and sampling rates**

| System | 2 Hz | 256 Hz | 2048 KHz | 16384 KHz | Total (KHz) |
|---|---|---|---|---|---|
| Suspension | 120 | 90 | 30 | 60 | 300 |
| Prestablized Laser | 20 | 10 | 5 | 8 | 43 |
| Mode Cleaner | 30 | 20 | 10 | 20 | 80 |
| Injection Optics | 20 | 15 | 5 | 10 | 50 |
| IFO Readout | 20 | 15 | 0 | 30 | 65 |
| Auto Alignment | 20 | 15 | 0 | 0 | 35 |
| Channels/IFO | 230 | 165 | 50 | 128 | 573 |
| KBytes/sec/IFO | 0.9 | 84.5 | 204.8 | 4194.3 | 4484.5 |
| Auxiliary | 0 | 200 | 10 | 30 | 240 |
| Housekeeping | 300 | 50 | 20 | 0 | 370 |
| Channels/site | 300 | 250 | 30 | 30 | 610 |
| KBytes/sec/site | 1.2 | 128 | 122.9 | 983.0 | 1235.1 |

LIGO data will be collected from roughly 3600 channels at sampling rates ranging from 2 to 16384 samples per second. These channels will compliment the gravitational wave output channel with physical environmental monitoring, housekeeping information and controls monitors. The total data collected from each interferometer is estimated at over 5 megabytes per second. LIGO will collect data continuously throughout the year in order to record the raw gravitational events. This results in a yearly data rate of $5.02\text{x}10^{+14}$ bytes for all three interferometers. If 50 gigabyte tapes were used to store the data, ten thousand tapes would be needed to record the raw data from the three interferometers. The data will be packed into units called frames which carry a complete representation of the data for one second. These frames will be the bases for data archival, processing and analysis.

# Gravitational Waves





When compact massive objects such as neutron stars and black holes experience an acceleration as in the case of a supernova or the inspiral of a compact binary system, the geometry of space-time experiences a dynamic change which propagates at the speed of light in the form of gravitational waves. The gravitational waves transverse space-time producing a cyclic elongation and contraction of the bodies in the plane perpendicular to the propagation direction (figure on left). Like electro-magnetic waves, the gravitational waves can be represented by to orthogonal polarizations, $h_x$ (h-cross) and $h_+$ (h-plus). The figure on the right shows the waveforms for these two polarizations for the final 200 milliseconds of a binary system of two 10 solar mass black holes with an inclination angle of $30^o$ and at a distance of 10 megaparsecs is shown above. The vertical axis is the strain on space-time in units of $10^{-20}$.

# Binary Inspiral Data Analysis

The inspiral and coalescence of compact binary systems composed on neutron stars and black holes are among the most promising sources of gravitational waves for LIGO to detect. The sensitivity of the initial LIGO interferometers will all an inspiral of 1.4 solar mass neutron stars to be detected out to 20 megaparsecs. A pair of 10 solar mass black holes could be detected at roughly 100 megaparsecs. Detection of this type of source is greatly enhanced by the fact that the waveforms are known to a very high precision from the second order post-Newtonian approximation(2PN). This allows for the use of Wiener optimal filtering (*matched filtering*) techniques on the frequency representation of the interferometer data $\boldsymbol{h_{obs}(f)}$ using the equation

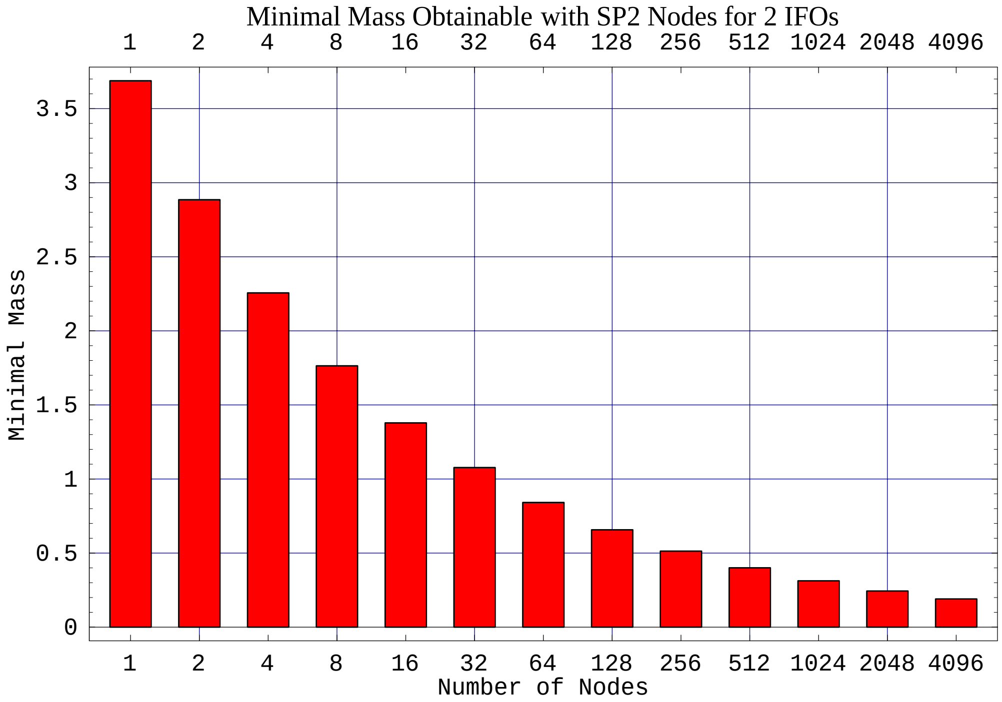
$$S = \int_{-\infty}^{\infty} \frac{h_{obs}(f) \cdot T^{*}(f) \cdot e^{-i2\pi f t_o}}{S_h(f)} df$$

Templates ($\boldsymbol{T(f)}$) parameterized by the two masses in the system are constructed using the 2PN waveforms (spin is currently neglected). Because of the detector’s noise ($\boldsymbol{S_h(f)}$), it is not possible to say with absolute certainty that a particular inspiral epic was seen in the data using the matching parameters. However, it is statistically possible to state that an inspiral of sufficiently large amplitude occurs, it will be seen with a sufficiently high probability (say 90%) if the template used to detect it lies in a sufficiently nearby region of the parameter space for the masses. The spacing of these templates in the parameter space of masses is a function of the interferometer noise floor. For the initial LIGO interferometers the number of templates(***N***) needed for detection with specific statistical probability for event ***loss*** is approximately given by

$$N = 7592 \cdot \left(\frac{loss}{0.1}\right)^{-1} \cdot (Mmin)^{-2.7}$$

where ***Mmin*** is the mass of the smaller member if the binary system in solar mass units.

# Compute Model for Binary Inspiral Analysis



Computer Configuration:

- 70% of 235 Megaflops/Node
- 256 MB Ram/Node
- 2 GB disks/Node
- 14 MB/sec I/O Xfer Rate
- FFT ~5Nlog$_2$(N) “complex”
- $90K/Node for IBM SP2

Model Predictions:

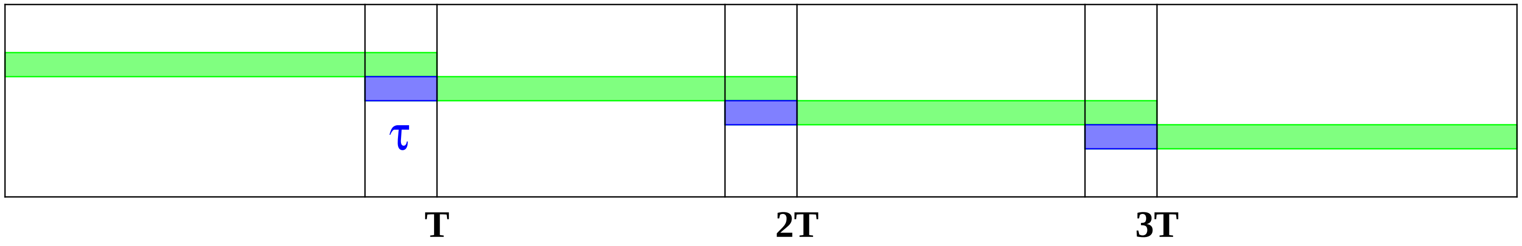
- CPU Limited Analysis
- Dominated by FFT & IFFT

The detailed computational needs for the two interferometers at the Hanford, Washington Site have been modeled. The model is based using an Excel spreadsheet program which takes into account all floating point operations associated with the flow diagram. It calculates the total number of template swaps carried out at in each node based on the available memory and includes the time to move templates and data between the nodes and out of storage. The megaflops (***P***) needed to reach a given

$$P = 7186 \cdot (Mmin)^{-2.806}$$

***Mmin*** from the model fits the equation shown above. A total of 16 nodes are needed to carry out analysis down to the Chandresekhar limit of 1.4 solar masses at an estimated cost of $1.4 million using IBM SP2 hardware. To reach a minimum mass of 1 solar mass would require nearly 40 nodes.

# Marching through the Binary Inspiral Data

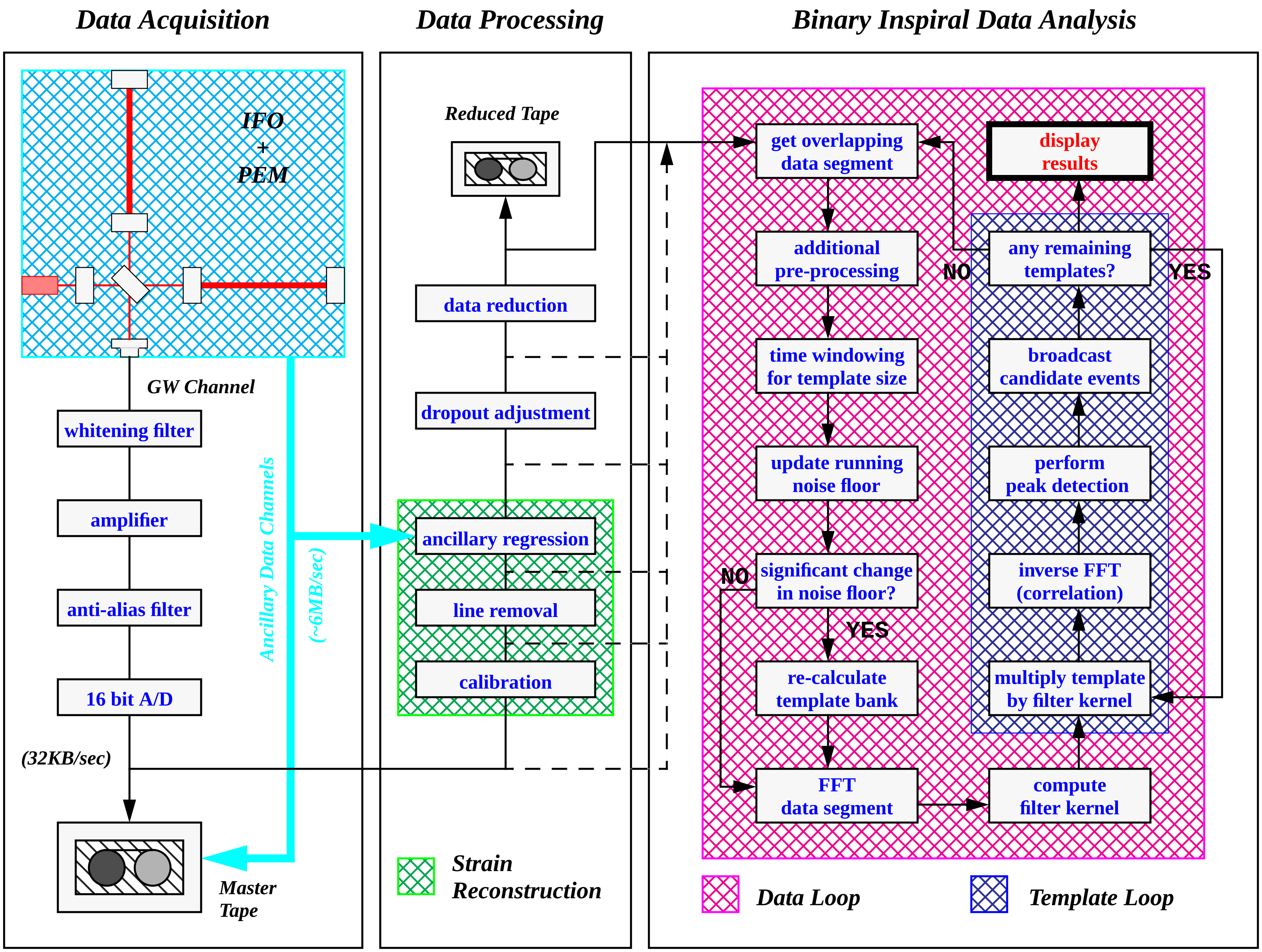


τ - optimal template length; amount of data reused in each pass

**T** - amount of data analyzed in each pass; **36** times longer than τ

The underlying principle to the binary inspiral analysis shown in the flow diagram is the correlation of the gravitational wave data channel with the templates. This is carried out by using the correlation theorem, appropriately weighed for instrument noise, i.e., matched filtering. The correlation of two discretely sampled functions, one of which goes on indefinitely and the other which is of fixed length is dealt with by zero padding of the shorter function (the templates in this case). To be able to see a binary inspiral event that begins at one of the edges, that data must be reused. Taking these two effects together introduces a competition between doing long inverse FFTs and analyzing the most data in a single pass. The lowest order approach to how to optimize these competing processes suggests a ratio between T and t of somewhere between 5 and 15. Using the data analysis model for the binary inspiral, this ratio is actually optimized to give the largest data throughput on the parallel hardware and the result suggests that the actual ratio is **36**. The minimum in this optimization is very shallow and the actual value is sensitive to the hardware configuration but has little impact on the performance for ratios within 10 to 15 of this value.

# Data Flow Diagram for Binary Inspiral



# Periodic Source Data Analysis

The LIGO frequency band will open a window into periodic sources of gravitational waves from rapidly rotating neutron starts (pulsars). Currently there are more than 700 known pulsars, all within the galactic distances. Using sensitivity arguments for radio astronomy observations this implies that the galaxy is populated with more than 100,000 active pulsars. Gravitational wave emission from pulsars is expected to be very weak, but not out of the question for detection by LIGO.

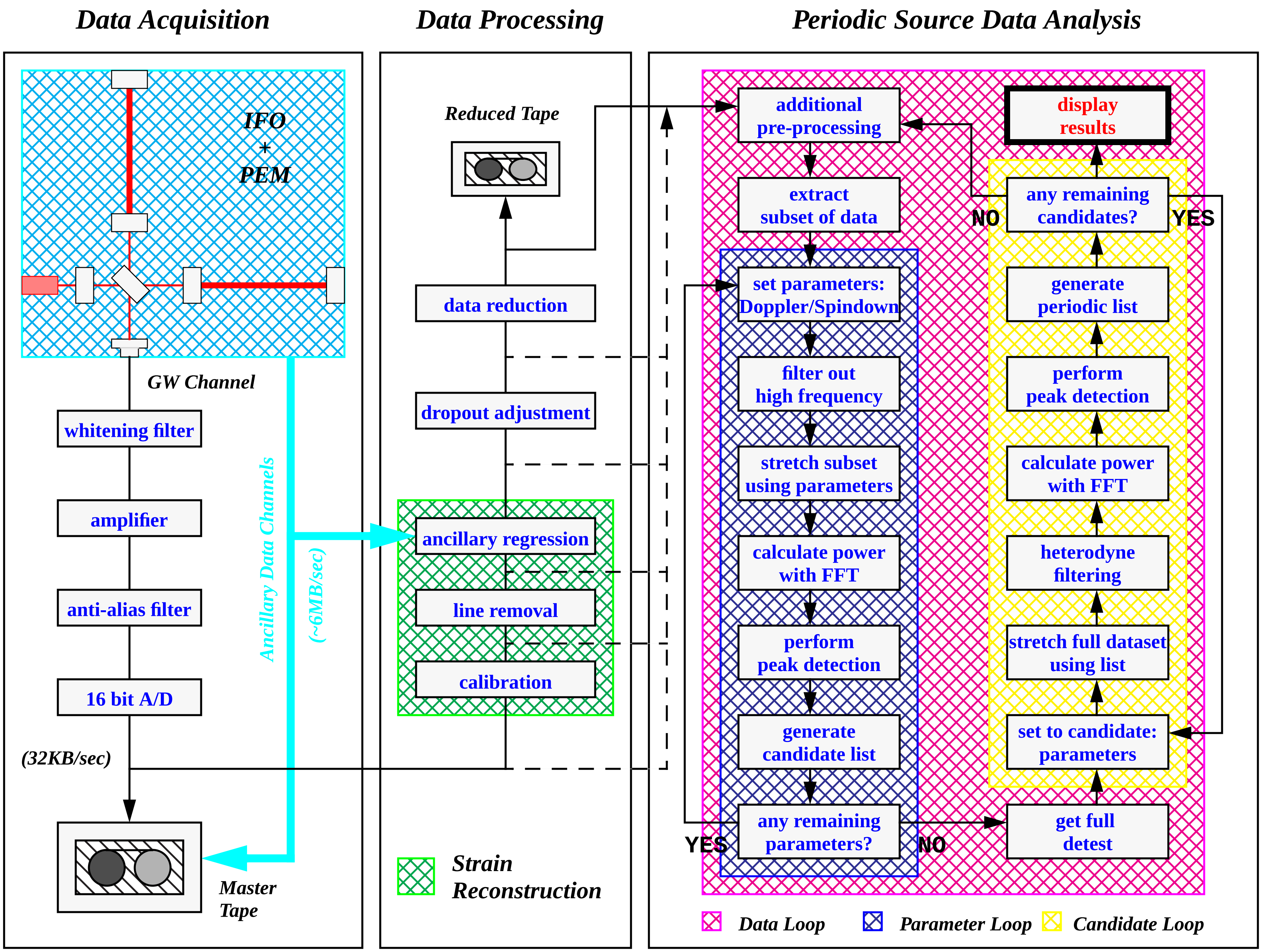
The detection strategy for pulsars or any other periodic source of gravitational waves consists of building up power in frequency bins by Fourier transforming long stretches of data. By doing such, the power in monochromatic signals will grow as the square root of the integration time. Expected signal strengths suggest that weeks to months of data may need to be analysis to bring the signal out of the noise. This would require FFTs of $10^{+10}$ data points for source frequencies of order 1000 Hz. A single such FFT would require about 1 second on a teraflops computer.

The detection problem is complicated by several factors related to the way the signal is received:

- Complex motions of Earth bound detectors lead to significant Doppler shifts in frequency
- Energy loss through radiation mechanisms result in frequency spindown of the sources
- Pulsars may be members of a binary system having rapidly changing proper motion
- Large proper velocities my carry pulsars across more than one resolution element, “~arcsecond”
- Frequency glitches triggered by instabilities in the internal structure of the pulsar routinely occur.

Of these the first two can be addressed with straight forward methods. The third requires an additional 5 parameters to model making the parameter space too large to search. The fourth can be neglected if observation times are kept well under one year and frequencies are under 1 kHz. Based on these a significant about of study and modeling have gone into the correcting the detected signal for Doppler shifts and spindown. The other complications are neglected due to their inherent difficulties.

# Data Flow Diagram for Periodic Sources



# Compute Model for Periodic Sources

Using a technique based on differential geometry similar to the technique used to determine the number of templates needed in the binary inspiral data analysis, the number of sets of parameters or “patches” required to search the whole sky with 0, 1, or 2 spin down parameters is given by

$$N_{patches} \cong Max\left\{\left[\left(\frac{f_o}{1kHz}\right)^{s+2}\cdot\left(\frac{40yrs}{\tau}\right)^{s\frac{(s+1)}{2}}\cdot\left(\frac{0.3}{\mu_{max}}\right)^{\frac{(s+2)}{2}}\right]\cdot F_s(T)\right\}$$

where $\mu_{max}$ is the mismatch between signal and patch and $T$ is the integration time in days and the

$$F_0(T) = 6.9\times10^3\cdot T^2 + 3.0\cdot T^5$$

$$F_1(T) = \frac{1.9\times10^8\cdot T^8 + 5.0\times10^4\cdot T^{11}}{4.7 + T^6}$$

$$F_2(T) = \frac{2.2\times10^7\cdot T^{14}}{56.0 + T^9}$$

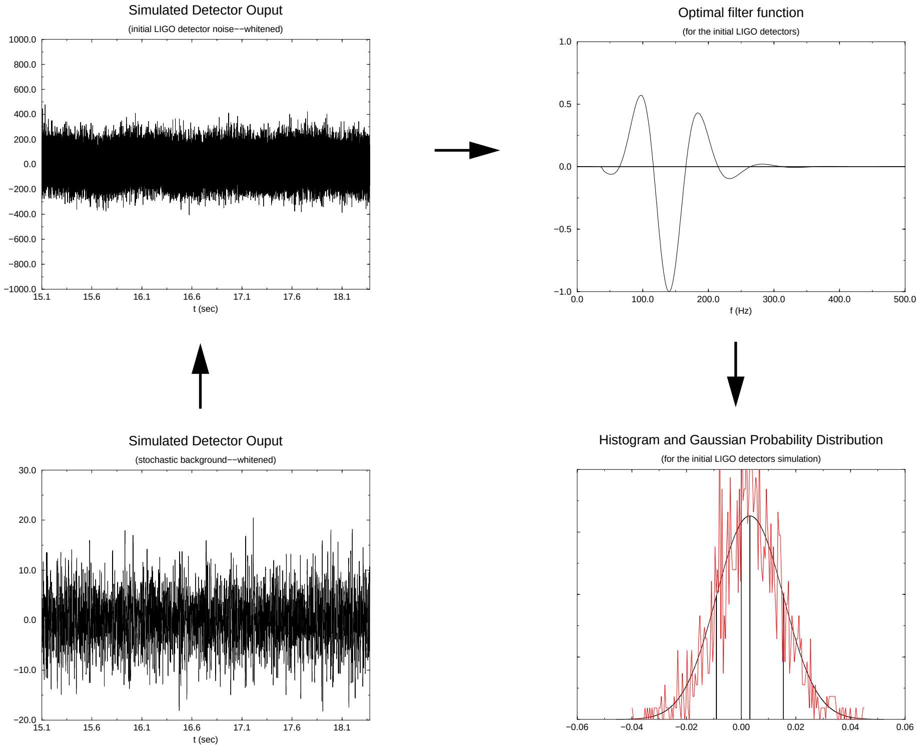
functions F are given above. The computational requirements are driven by the number of patches and the floating point operations for the FFT (neglecting the correct for the Doppler shift and thresholding). Taken together, the computer power needed to keep up with the data for one detector is

$$P_{pulsar} = 6\cdot f_{max}\cdot N_{patches}\cdot\left[\log_2(2\cdot f_{max}\cdot T) + \frac{1}{2}\right]$$

When values consistent with current understanding of pulsars are used in these expressions, the compute requirement exceeds $10^{+15}$ FLOPS! With a teraflops computer analysis one can do becomes:

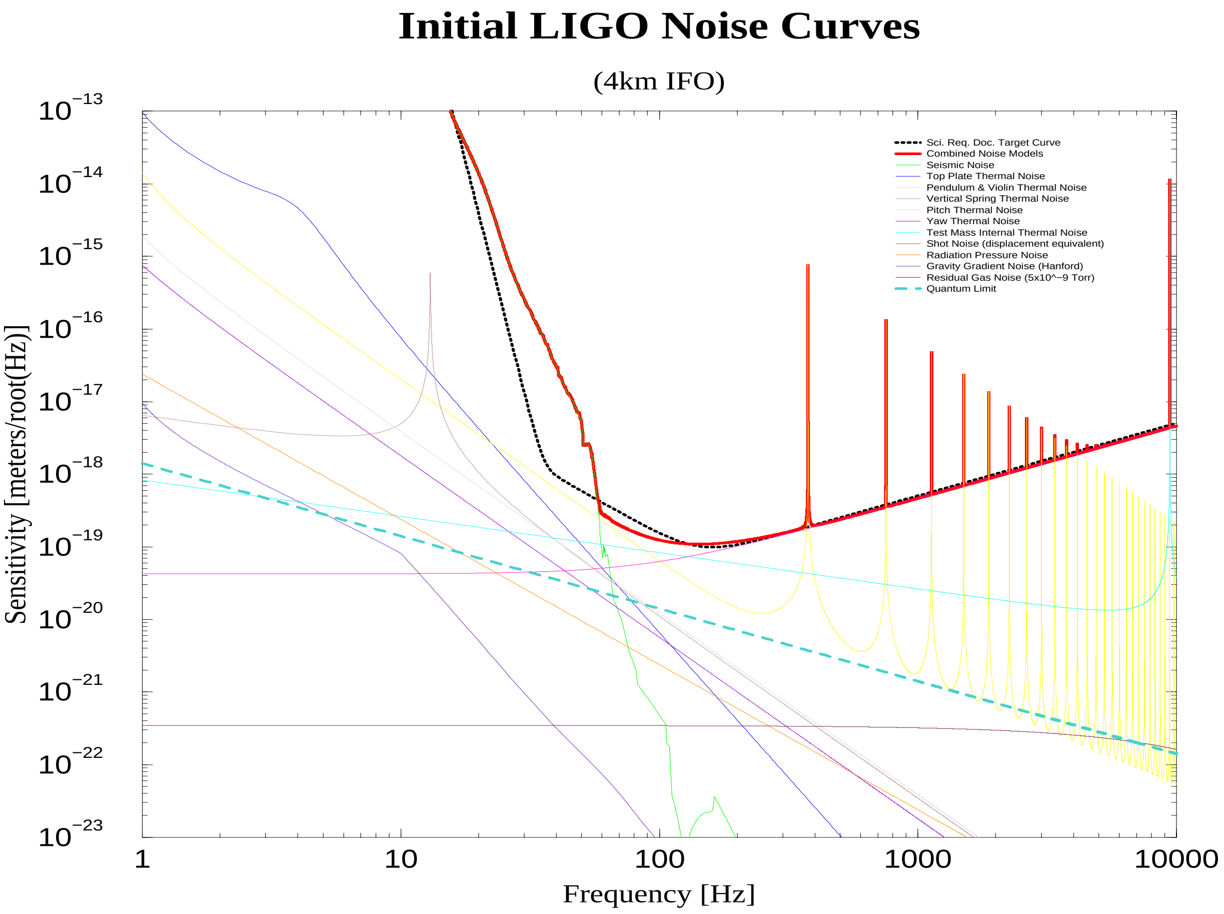
- 18 days of data could be coherently searched for gravitational wave frequencies f < 200 Hz and minimal spindown ages $\tau > 1000$ years (reasonable data lengths for expected signal strengths)
- 0.8 days of data could be coherently searched for gravitational wave frequencies f < 1kHz and minimal spindown ages as low as 40 years (the more interesting case)
- Directed searches at known supernova remnants, galactic center, etc. increase observation times by ~10 (only).

# Stochastic Background Simulation



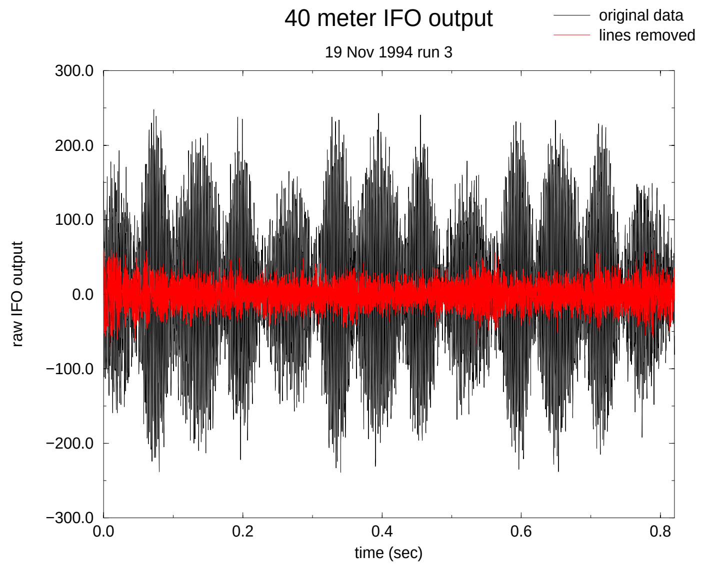
Bruce Allen and Joe Romano have carried out studies of stochastic back-ground analysis of LIGO data. The method is computationally simple as LIGO data analysis goes, and involves the correlation of data from two different interferometers. An optimal filter is constructed which accounts for the differences in arrival times and orientations for the two interferometers. This filter is shown on the top right for the LIGO Hanford, WA and Livingston, LA detector pair. The analysis combines simulated stochastic signals with simulated initial LIGO detector noise through the optimal filter to calculate the signal to noise ratio from the stochastic background. The simulations shown here finds an (SNR) of 7.937 corresponding to an energy density $\Omega_o$ of $1.955 \times 10^{-4}$ with a 95% detection confidence. The bottom right figure shows the measured cross correlation signal values against the Gaussian fit.

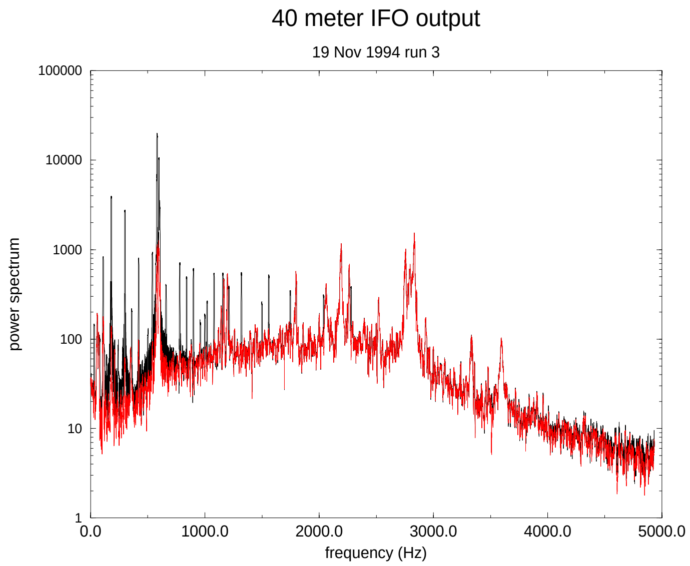
# Modeled LIGO Noise Floor



The sensitivity of LIGO to gravitational waves is limited by the noise sources in the detector. The signal output from the initial LIGO interferometers will span four orders of magnitude over the frequency band of interest as a result of the colored noise floor. Gravitational waves from anticipated sources will only occasionally have signal to noise ratios detectable above this noise floor.

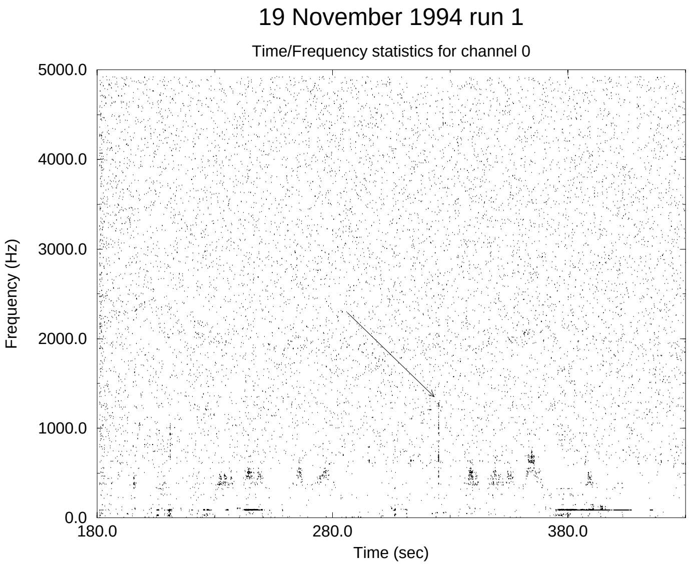
# Line Removal using Multi-Taper Methods

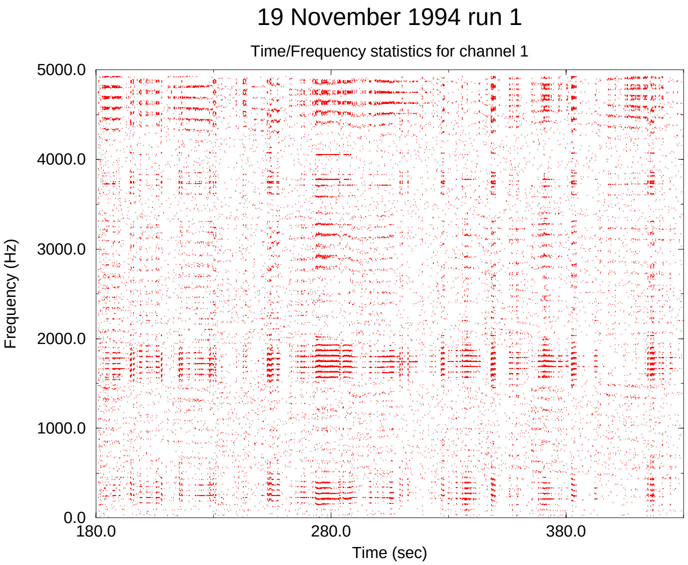




LIGO data will possess a multitude of narrow line resonances found in the interferometer's noise floor. These lines are highly predictable and thus easily removed from the signal. The figures above illustrate the ability of the multi-taper method for removing these types of resonances. The figure on the left shows the raw signal of the 40 meter proto-type interferometer at Caltech in black. The signal is processed using the multi-taper method found in the GRASP software package developed by Bruce Allen. The results are shown in red. On the right is the frequency representation of the signal showing the narrow lines (39 in total) which were removed. These lines are associated with harmonics of the 60 Hertz line frequencies and thermal noise resonances in the suspension system. A simulation where by a binary inspiral signal was added to real 40 meter data demonstrated that this method improves signal to noise by 30%, equivalent to an 80% increase in the volume of the universe observable by the interferometer.

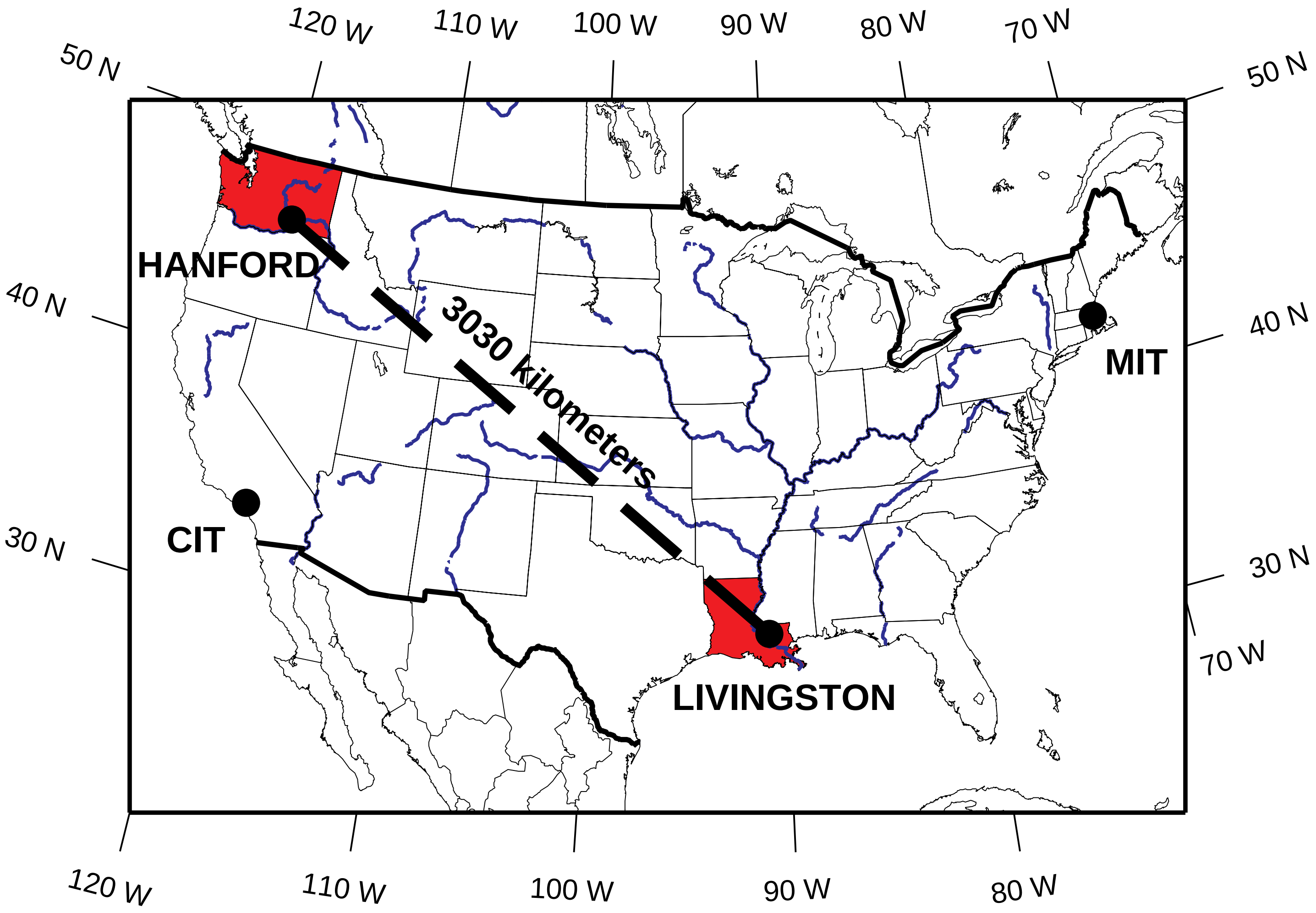
# Time Frequency Analysis Methods





The figures above demonstrate an application of time-frequency analysis methods, used here for diagnostics. This particular method uses an auto-regressive averaging technique to derive a mean power spectrum with an exponential-decay time constant of 10 seconds. The current power spectrum is calculated and compared to the mean spectrum. If the difference exceeds a specified threshold, the time-frequency pixel is highlighted. The figure on the left is of the gravitational wave signal on the 40 meter proto-type interferometer. The arrow indicates the appearance of an instrumental signature. The figure on the right is for the magnetometer, one of the ancillary channels acquired at the 40 meter lab. These time-frequency methods have proven very useful in instrument diagnostics and are expected to be important to data analysis. Other time-frequency methods currently being explored include Gabor transforms and wavelets analyses.

# Primary Sites for LIGO Wide Area Network



LIGO networks will communicate data between the two sites in Hanford, Washington and Livingston Parish, Louisiana, as well as the joint developing institutions of California Institute of Technology and Massachusetts Institute of Technology. LIGO is currently studying technologies and options for implementing this wide area network.