LIGO DATA PROCESSING

NSF REVIEW

April 15th, 1997

James Kent Blackburn

LIGO Integration Group

OUTLINE

1) Plans for 40 Meter Data Analysis

a) Data Analysis background
b) On-line Analysis
c) Center for Advanced Computing Research (Off-line)
d) Gravitational Radiation Analysis & Simulation Package

2) Refined Data Flow Models for Initial LIGO Data Analysis

a) Estimated Data Rates & Channel Counts
b) Data Flow Diagram for Binary Inspiral Search
c) Data Flow Diagram for Periodic Source Search
d) Computational Requirements associated with b) & c)

Plans for 40 Meter Data Analysis

Data Analysis Background Info:

- *46 hours of data collected in November, 1994 of which 88.5% of the data is in lock*
- *13.5 hours of this data previously analyzed as part of thesis research*
- *49 Newtonian templates between 1.2 and 1.6 solar masses used as filters*
- *noise characterized using pulse height histograms*
- *data contained significant non-Gaussian events (some later determined to be software artifacts)*

+ *reanalysis of data using current understanding to achieve higher signal sensitivity has begun*

On-line Analysis:

LIGO prototype data acquisition system installed at the 40 meter

- *This system will collect data from up to 32 fast channels and 128 slow channels*
- *The data will be stored in the pre-release Frame structure planned for LIGO and VIRGO*
- *DADisp is being used for quick look data display and simple signal processing*

High Speed ATM access to data over wide area

- *ATM backbone connecting 40 meter lab with LIGO general computing in place*
- *ATM connectivity to the Center for Advanced Computing Research in the works*

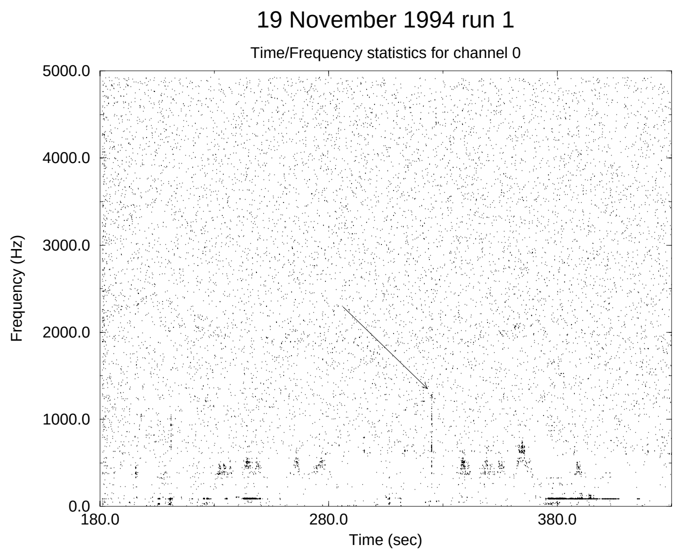
New algorithms and filtering techniques for characterizing the noise

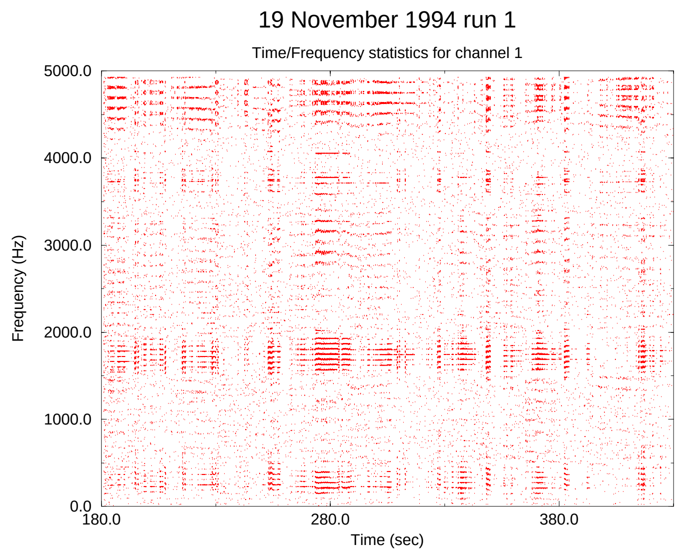
- *time-frequency "carpets" of the running power spectrum*
- *multi-taper methods for tracking narrow lines*
- *2PN binary inspiral template for optimal filtering*

Plans for 40 Meter Data Analysis

On-line Analysis:

Time-Frequency Analysis: *provides detector diagnostics & non-Gaussian noise characterization*



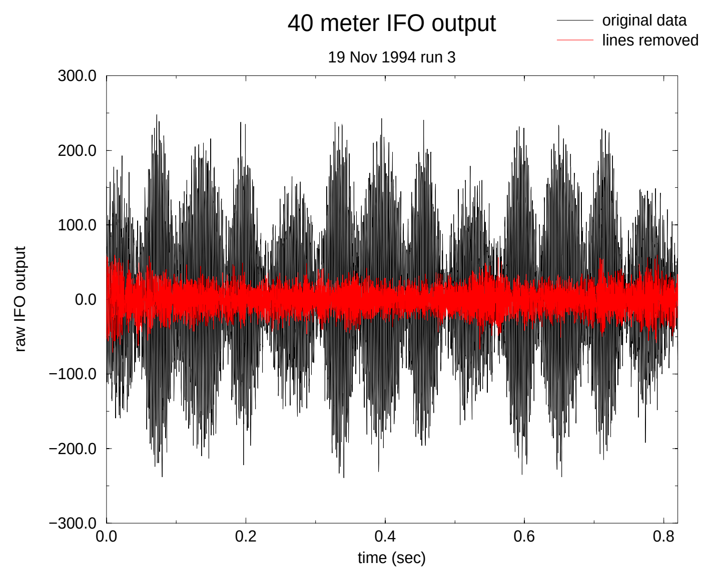


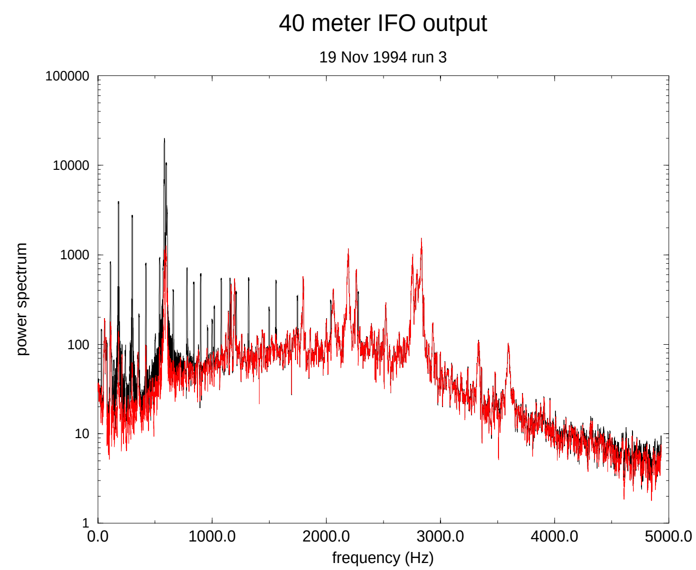
The left figure is a time frequency plot of the gravitational wave output channel of the 40 meter interferometer. The graph was generated by calculating an exponentially weighted running power spectrum and comparing to the current average power spectrum. If the difference between the two exceeds a user defined threshold, then the associated time-frequency pixel is set. The arrow locates a non-Gaussian event in the detector which sounds like a “drip” to the ear and is believed to be off-axis modes in the optical cavities. The right figure is the magnetometer channel during the same time interval. (*figures generated with GRASP*)

Plans for 40 Meter Data Analysis

On-line Analysis:

Multi-Taper Methods: *better estimators of power spectrum & tracking/removal of narrow lines*



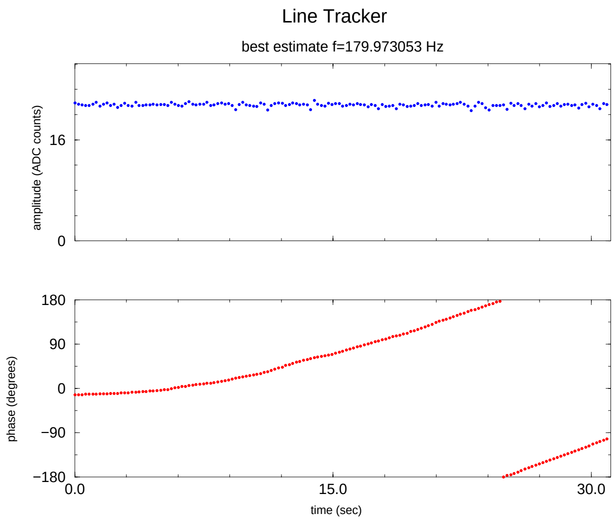


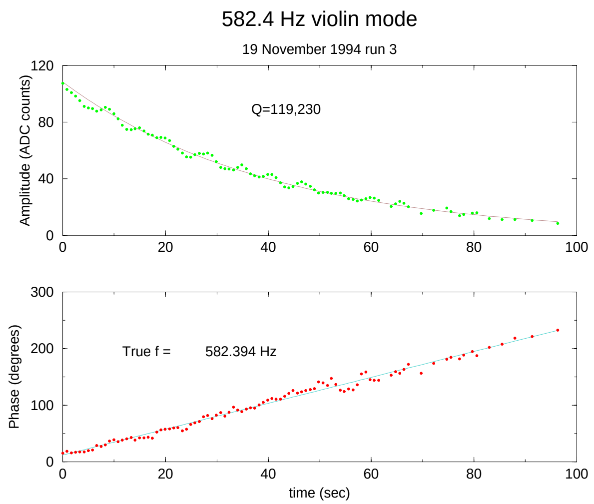
The raw (40 meter) data on the left is processed using the multi-taper method to remove 39 narrow lines from the signal in real time. The figure on the right shows the difference between the raw power spectrum and the power spectrum after the lines are removed. A simulation in which a binary inspiral signal was injected into the raw data showed a 30% improvement in the signal to noise ratio, equivalent to an 80% increase in the observable universe in the 40 meter data. (*figures generated with GRASP*)

Plans for 40 Meter Data Analysis

On-line Analysis:

Multi-Taper Methods: *diagnostic tool for monitoring narrow lines*



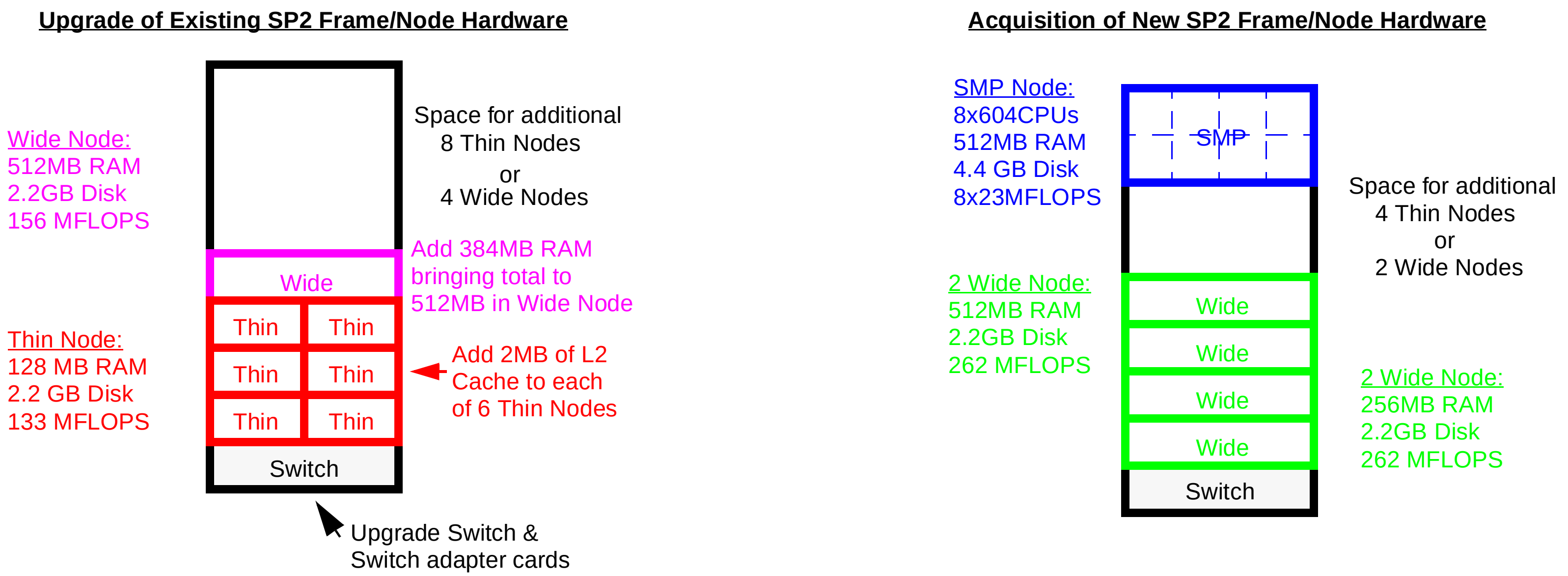


The multi-taper method can also be used for on-line detector diagnostic. The figures above illustrate the ability of the method to track harmonics of line frequencies and narrow mechanical resonances such as the violin mode at 582.6 Hz. Using an Ultra1 SPARCstation, the method was able to keep up with the input of the November 1994 data making a real-time implementation of the method viable. (*figures generated with GRASP*)

Plans for 40 Meter Data Analysis

Center for Advanced Computing Research (Off-line):

IBM SP2 Scalable Parallel Computer Hardware (SUR Award)



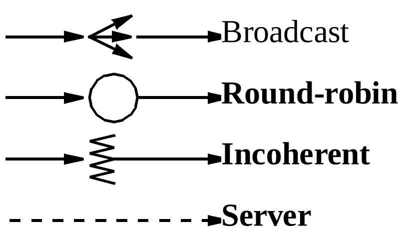
New hardware arrived the first week of January, 1997
Current configuration consists of 12 nodes capable of 2,186 MFLOPS
System presently operational and in verification phase
Minimal mass in binary search expected for one LIGO data stream: 1.34 solar masses
Preliminary 40 meter data analysis for binary inspiral has begun using GRASP
Other resources at the CACR include a 15 TB tape robot and a 512 node Paragon

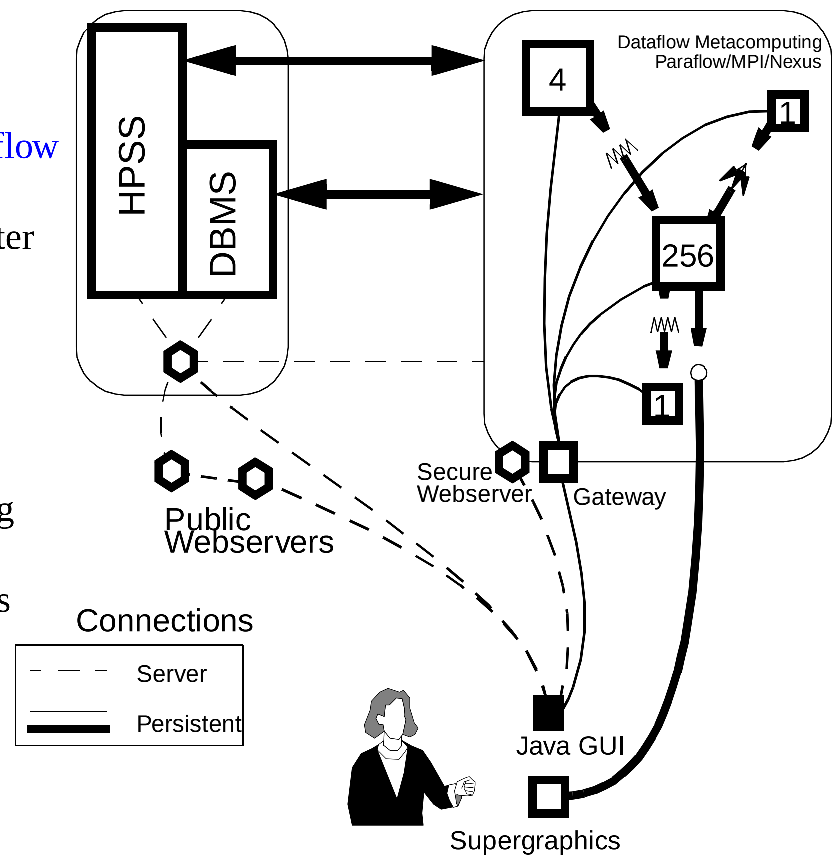
Plans for 40 Meter Data Analysis

Center for Advanced Computing Research (Off-line):

PARAFLOW: A Dataflow Distributed Data-Computing System

- Developed by Roy Williams & Bruce Sears, CACR
- Prototype based on MPICH, including Users Guide, available on web: http://www.cacr.caltech.edu/~roy/paraflow
- Combines heterogeneous computing resources together into a flexible efficient data-mining metacomputer
- Three levels of parallelism provided:
    - functional parallelism in dataflow paradigm
    - each service can be a parallel program
    - each node of the service can be multithreaded
- easily integrated with web-computing
- meant to easily provide a bridge between public browsing of data to collaborative high performance computing
- services are connected by channels of the following types





Plans for 40 Meter Data Analysis

Gravitational Radiation Analysis & Simulation Package (GRASP):

USERS MANUAL

GRASP: a data analysis package for gravitational wave detection

Bruce Allen*
Department of Physics
University of Wisconsin - Milwaukee
PO Box 413
Milwaukee WI 53201, USA

March 31, 1997

Abstract

GRASP (**G**ravitational **R**adiation **A**nalysis & **S**imulation **P**ackage) is a public-domain software tool-kit designed for analysis and simulation of data from gravitational wave detectors. This users manual describes the use and features of this package.

GRASP RELEASE 1.0
manual version 1.0.2

*ballen@dirac.phys.uwm.edu

1

- Developed by Bruce Allen
- Software Toolkit with Library and Examples
- User's Manual currently over 200 pages
- Binary Inspiral Searches implemented
- Calculates 2PN templates/spacing based on $S_N(f)$
- Optimal filtering uses MPI to distribute CPU load
- Stochastic Background searches implemented
- Diagnostic tools: time-freq., multi-taper
- Plans for Periodic and Supernovae searches
- Primary testbed has been Nov. 94 40 meter
- Handles data calibration
- I/O to data based on Frames implemented
- Software developed in C language
- Under CVS software version/release management

Plans for 40 Meter Data Analysis

Gravitational Radiation Analysis & Simulation Package (GRASP):

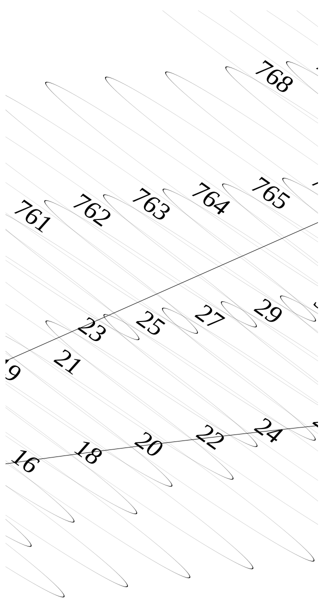
Binary Inspiral Template Spacing

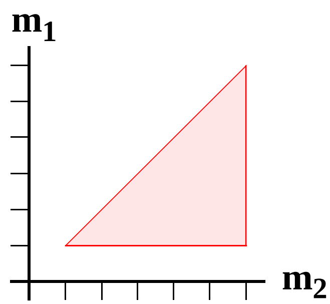
$$\tau_0 = \frac{5}{256} \cdot \left(\frac{G \cdot M}{c^3}\right)^{-5/3} \cdot \eta^{-1} \cdot (\pi \cdot f_o)^{-8/3}$$

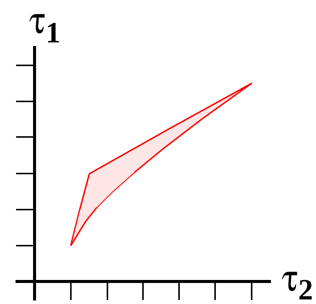
$$\tau_1 = \frac{5}{192} \cdot \left(\frac{c^3}{G \cdot \eta \cdot M}\right) \cdot \left(\frac{743}{336} + \frac{11}{4} \cdot \eta\right) \cdot (\pi \cdot f_o)^{-2}$$

$$M \equiv m_1 + m_2$$

$$\eta \equiv \frac{m_1 \cdot m_2}{(m_1 + m_2)^2}$$





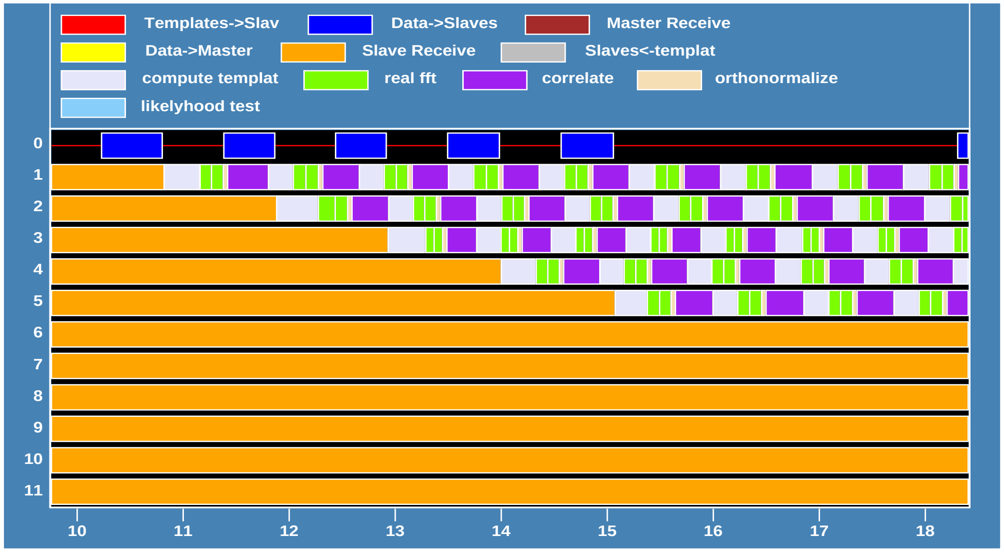


- templates correspond to chirp signals in the parameter space of (m_1, m_2)
- templates shaped like ellipses in the coordinates of (τ_1, τ_2)
- template size/orientation determined by shape of detector noise
- template location determined by desired ambiguity (~3% in figure to the left)
- the Nov. 94 40 meter data needs 403 templates to span $0.8 <= (m_1, m_2) <= 50$ solar masses

Plans for 40 Meter Data Analysis

Gravitational Radiation Analysis & Simulation Package (GRASP):

Binary Inspiral Optimal Filtering using MPI



- optimal filtering code uses the Message Passing Interface (MPI/MPE) libraries from MPICH
- MPI provides portability, code runs on LIGO network of SUNs, Intel Paragon and IBM SP2 at the CACR
- GRASP model has each node analyzing different stretches of date with full template bank used on each node
- templates can either be stored and read back or recomputed as needed
- using optimized FFT on Paragon, the analysis of 1 template requires 0.8 seconds for 6 seconds of 40 meter data
- spanning 1.2 -> 1.6 solar mass, templates for the Nov. 1994 40 meter data requires 66 templates
- 5 hours of 40 meter data can be filtered through the 66 template in 10.3 minutes using 256 nodes on the Paragon

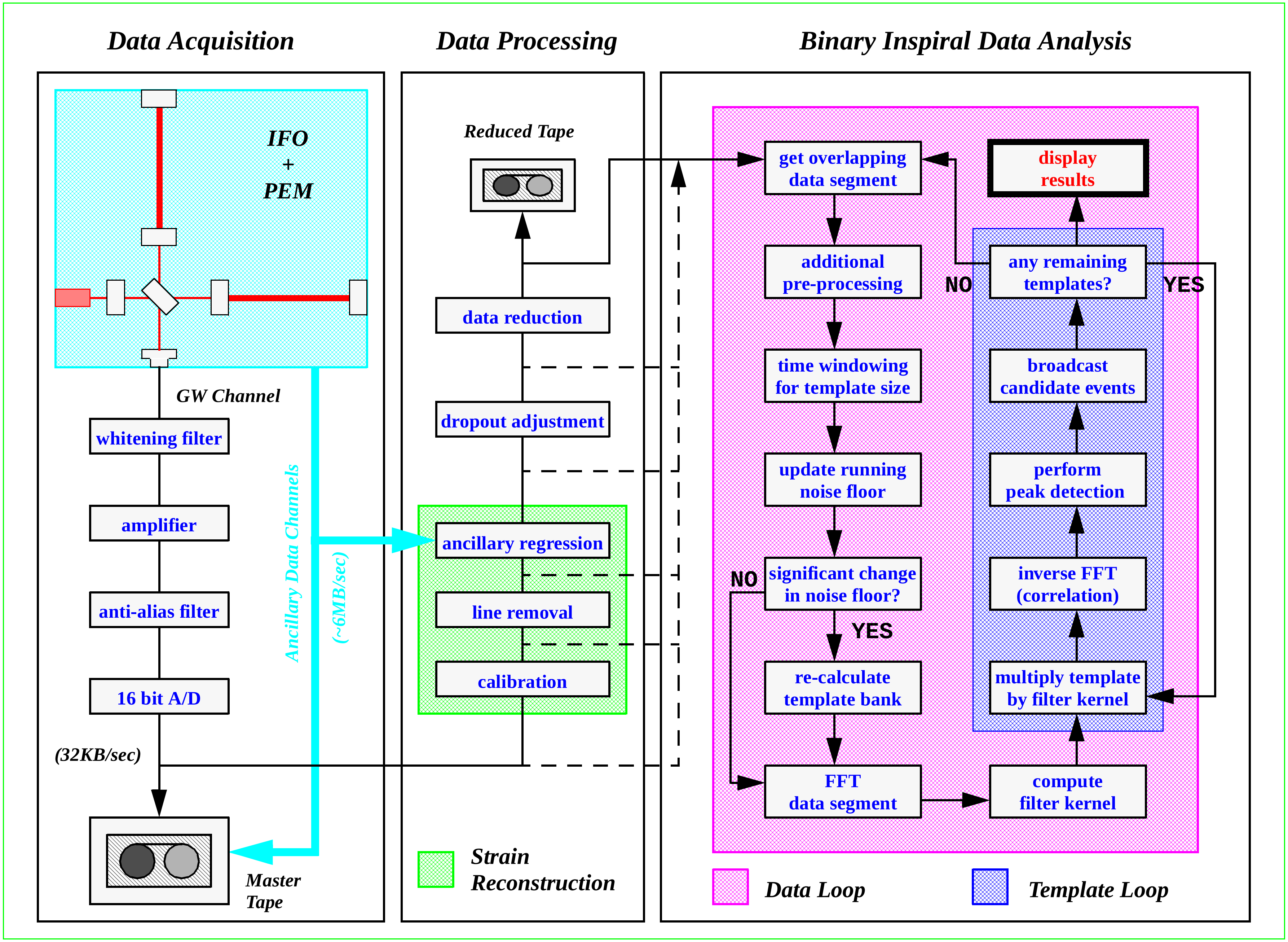
Refined Data Flow Models for Initial LIGO Data Analysis

Estimated Data Rates & Channel Counts:

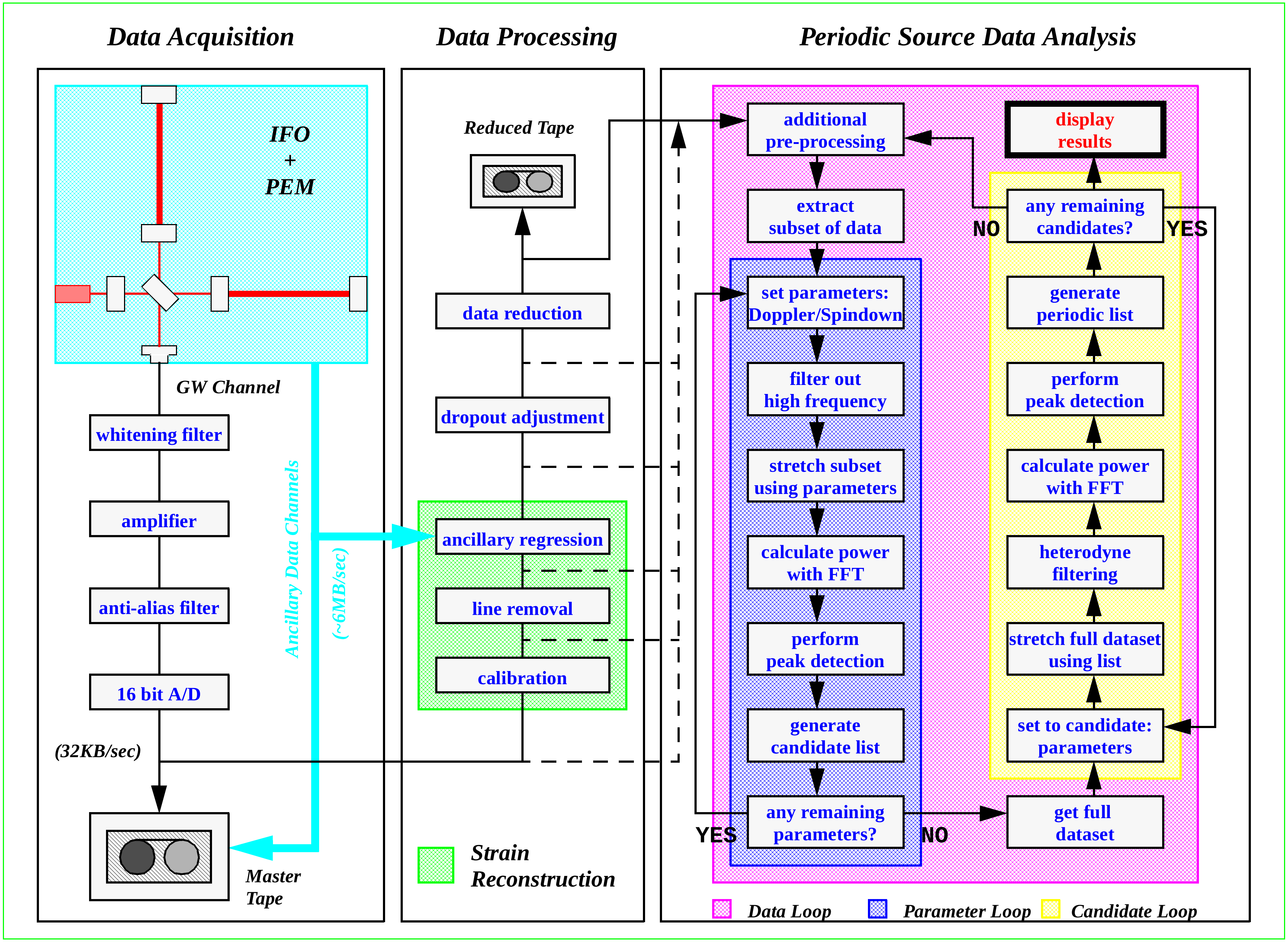
Estimates of LIGO IFO and PEM channel counts and sampling rates

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. This results in a yearly data rate of $5.02 \times 10^{+14}$ bytes for three interferometers. If 50 gigabyte tapes were used to store the data, 10,000 tapes would be needed to record the raw data (***no copies***).



Data Acquisition
Data Processing
Binary Inspiral Data Analysis
IFO + PEM
GW Channel
whitening filter
amplifier
anti-alias filter
16 bit A/D
(32KB/sec)
Master Tape
Ancillary Data Channels
(~6MB/sec)
Reduced Tape
data reduction
dropout adjustment
ancillary regression
line removal
calibration
Strain Reconstruction
get overlapping data segment
additional pre-processing
time windowing for template size
update running noise floor
significant change in noise floor?
NO
YES
re-calculate template bank
FFT data segment
compute filter kernel
multiply template by filter kernel
inverse FFT (correlation)
perform peak detection
broadcast candidate events
any remaining templates?
NO
YES
display results
Data Loop
Template Loop

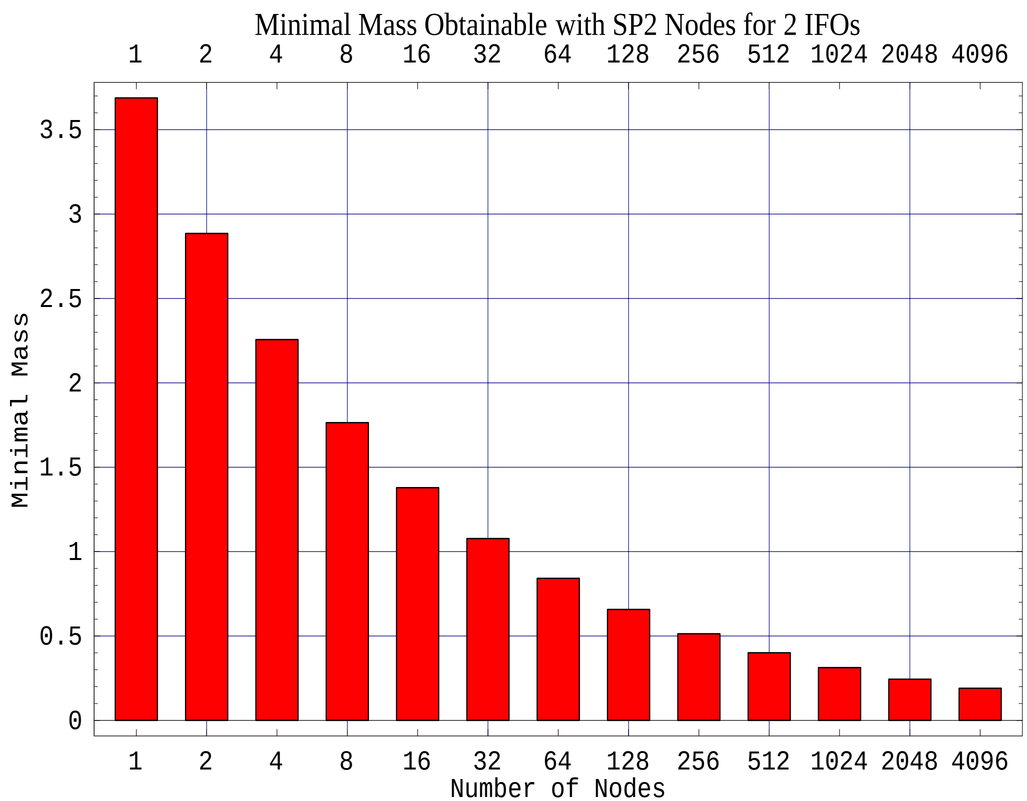


Data Acquisition
Data Processing
Periodic Source Data Analysis
IFO
+
PEM
GW Channel
whitening filter
amplifier
anti-alias filter
16 bit A/D
(32KB/sec)
Master
Tape
Ancillary Data Channels
(~6MB/sec)
Reduced Tape
data reduction
dropout adjustment
ancillary regression
line removal
calibration
Strain
Reconstruction
additional
pre-processing
extract
subset of data
set parameters:
Doppler/Spindown
filter out
high frequency
stretch subset
using parameters
calculate power
with FFT
perform
peak detection
generate
candidate list
any remaining
parameters?
YES
NO
get full
dataset
set to candidate:
parameters
stretch full dataset
using list
heterodyne
filtering
calculate power
with FFT
perform
peak detection
generate
periodic list
any remaining
candidates?
NO
YES
display
results
Data Loop
Parameter Loop
Candidate Loop

Refined Data Flow Models for Initial LIGO Data Analysis

Estimated Computing Requirements for Binary Inspiral Search:

Compute Model for Hanford Binary Search based on IBM SP2 Hardware



Model Assumptions:

- 2 IFO data streams
- 2km IFO <=> 4km IFO
- 235 MFLOPS per NODE
- Optimized FFTs 70% of above
- 256 MB RAM per NODE
- 32 MB RAM used by OS/CODE
- 2GB HD per NODE
- $N_{templates}$ = 7592(0.1/loss) $(M_{min})^{-2.7}$
- FFT ~ $5N\log_2(N)$ “complex”
- 14 MB/s Xfer between HD & NODES
- GW frequency < 2KHz
- Templates stored on HD
- ~ $90K per NODE list price

Model Predictions:

- Analysis is CPU limited!
- FFT & IFFT drive CPU usage!
- optimal padding => $36xN_{template}$
- 16 NODES => 1.4 M_{sun} => $1.4M
- 40 NODES => 1.0 M_{sun} => $3.6M
- CPUs = 7186 $(M_{min})^{-2.806}$ MFLOPS

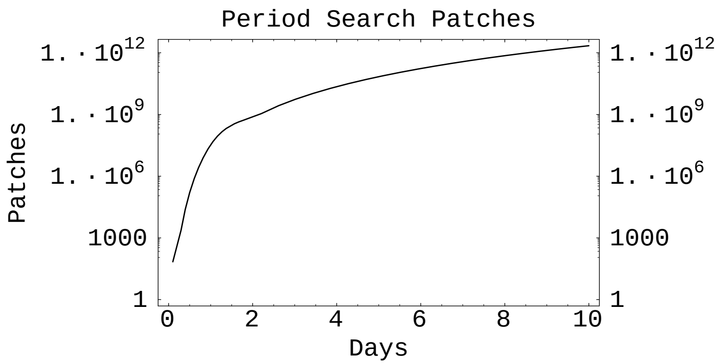
An Excel spreadsheet was developed based on the data processing flow for the binary inspiral search, the template density for expected initial LIGO sensitivity and the performance specifications for IBM SP2 hardware. Using the spreadsheet, the basic hardware configuration is easily modified and the resulting performance analyzed.

Refined Data Flow Models for Initial LIGO Data Analysis

Estimated Computing Requirements for Periodic Source Search:

The number of sets of parameters or "patches" needed to search the sky with 0, 1, or 2 spin down parameters is

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



where μ_{max} is the mismatch between signal and

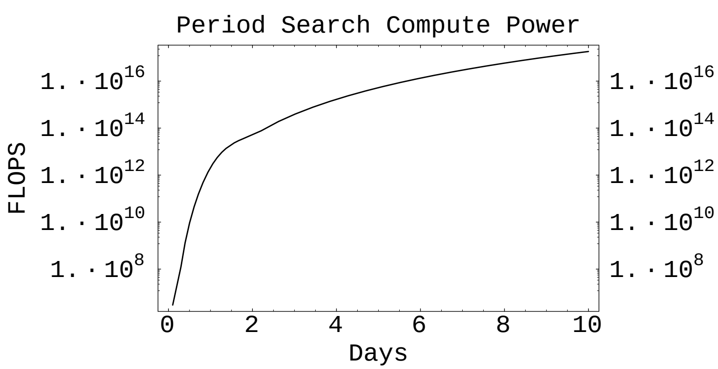
$$F_0(T) = 6.9\times10^3\cdot T^2 + 3.0\cdot T^5$$

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

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

patch and T is the integration time in days. The computer power needed to keep up with the data is

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



With a teraflops computer analysis one can do becomes:

- 18 days coherently searched for gravitational wave with f<200 Hz and minimal spindown ages $\tau > 1000$ years
- 0.8 days coherently searched for gravitational wave with f<1kHz and minimal spindown ages as low as 40 years
- Directed searches at known supernova remnants, galactic center, etc. increase observation times by order 10 (only).