LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY

- LIGO –

CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type	LIGO-M000040-01-E	09 March 2000
Technical Note	**VIRGO-MOM-DIR-7000-112**	

Minutes of meeting between VIRGO and LIGO to discuss joint data analysis requirements to coherently process multiple interferometer data for binary inspiral detection

A. Lazzarini, L.S.Finn (LIGO & LSC)
G. Cella, A. Giazotto, B. Mours, A. Vicere' (VIRGO)

Distribution of this draft: LIGO, LSC, VIRGO
This is an internal working note
of the LIGO Project.

VIRGO CNRS/INFN
Traversa H di Via Macerata
56021 S. Stefano a Macerata
Cascina (PI), Italy
Phone (39) 050 752 521 Fax (39) 050 752 550

California Institute of Technology
LIGO Laboratory - MS 18-34
Pasadena CA 91125
Phone (626) 395-212
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Laboratory - MS 16NW-145
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

www: http://www.ligo.caltech.edu/
www: http://www.virgo.infn.it/

A meeting was held at Caltech on 12, 13 February 2000 to review the data analysis computational requirements of multi-detector data analysis for binary inspiral coalescence events using multiple interferometers. In attendance were the LIGO and VIRGO members of the GWIC working group on this subject. This meeting was a follow-up to discussions of the full committee that began during the last GWDAW meeting in Rome on December 3 – 5 1999.

At this meeting, we reviewed the single-detector pipeline architecture that is common to both LIGO and VIRGO. We identified how this pipeline would need to be modified to accommodate a coherent multiple interferometer analysis involving all available detectors of comparable sensitivity. We evaluated how the calculation of a fully coherent search could be organized to minimize the resources required for this effort and determined how those costs would scale with the number of detectors involved.

The scope of a fully coherent network analysis scales linearly with the number of detectors involved in the network, reducing in the case of a single detector to the understood cost of a single detector search (Table 1 summarizes these conclusions). Correspondingly, it exceeds the capabilities required to handle individual detectors. The computational cost scales *only* linearly with the number of detectors, however, making a fully coherent network analysis a realizable opportunity. GWIC Working Group representatives from VIRGO and LIGO will recommend to their respective management that they investigate increasing the scope of the analysis resources to exploit this scientific opportunity.

In the remainder of this memorandum we describe the organization and computational cost of a fully coherent network analysis and describe a pipeline for processing and staging the data from each participating detector through the pipeline.

Computational Cost:

In evaluating the computational requirements for a multi-detector analysis we began with three requirements:

(i) The analysis is fully coherent and parallelized using data exchanged among all participating detectors.

(ii) The searches should take place in near-real time (correspondingly, *reduced* data must be exchanged in near real time, via internet or dedicated phone line).

(iii) As additional instruments with comparable sensitivity and bandwidth come on line, they must be simply incorporated in the analysis (modulo an increase in the computing resources dedicated to that task).

The fully coherent analysis is formulated as an extension of the standard template-based optimal Wiener filter scheme for a single detector, as described in a formalism developed by L.S. Finn. In the simplest case (i.e., no correlated noise) the total likelihood over the detector network is just the product over the likelihoods evaluated for the individual detectors, evaluated at a common set of signal parameters (e.g., chirp mass, sky location, etc.). If there is significant correlated noise between detectors, the situation is more complicated. Describing the collected data from the network by a vector-valued time series ***h*,** a particular template corresponding to a source can be written as the vector-valued time series ***T*** and the noise characteristics of the network by the matrix-valued cross-spectral $\boldsymbol{S_h(f)}$. Inter-detector correlated noise appears as off-diagonal components of $\boldsymbol{S_h}$. The application of the template to the data can then be written as

$$\xi(t) = \int df\, h(f)^{\dagger} \cdot S(f)^{-1} \cdot T(f) e^{2\pi i f t}$$

Now, there exists a unitary transformation ***U(f)*** that diagonalizes $\boldsymbol{S_h}$, so that we can write

$$\vec{h}^{\dagger}\mathbf{S}^{-1}\vec{T} \quad \vec{h}^{\dagger}\left[U^{\dagger}U\right]\mathbf{S}^{-1}\left[U^{\dagger}U\right]\vec{T} \quad \left[U\vec{h}\right]^{\dagger}\left[U\mathbf{S}^{-1}U^{\dagger}\right]\left[U\vec{T}\right]$$

The transformation ***U*** will not depend on time as long as the noise is reasonably stationary. We can pre-process the data and the templates, using ***U***, so that the equivalent noise cross-spectral density is diagonal.

Write ***h*** for ***Uh*** and $\underline{\boldsymbol{S}}_{\boldsymbol{h}}$ for the transformed $\boldsymbol{S_h}$. Since $\underline{\boldsymbol{S}}_{\boldsymbol{h}}$ is diagonal the likelihood of the multi-detector network can be written as the product of the likelihoods of N_{det} individual “likelihoods”, corresponding to the N_{det} pseudo-detectors whose data are the row time-series of ***h***, evaluated at a common set of parameters describing the source and its orientation with respect to the network. The templates for these pseudo-detectors are just the projections of the templates for the real detectors projected onto the eigenvectors of the cross-spectral density $\boldsymbol{S_h}$.

The computational time involved in a matched filter search is dominated by the correlation of the templates with the data. Reducing the number of correlations reduces the computational cost in direct proportion. To determine the templates that need to be applied to each of the pseudo-detectors focus first on the characterization of a template for a single, real detector k:

$$T_k\left(t \mid \Omega,\Omega',\sigma,\vec{\theta}\right) = \alpha_k(\Omega,\sigma)\beta(\Omega',\sigma)\tau\left(t + \Delta_k(\Omega) \mid \sigma,\vec{\theta}\right)/r\,,$$

where Ω (Ω’) is the position of the source (Earth) in the sky of the detector (source), σ is the polarization index (corresponding, e.g., to + or x), $\Delta_k(\Omega)$ is the time delay corresponding to the arrival of a fixed wavefront from the source at the Earth’s barycenter and the detector’s location, $\vec{\theta}$ is a vector of other, non-orientation/position astrophysical parameters that characterize the source (e.g., binary component masses, etc.), α_k describes the detector's sensitivity to radiation of given polarization originating at the source, and β describes the radiation amplitude in the given direction and polarization, relative to the source.

The templates against which the data must be correlated are thus just the τ: the other terms correspond to either a time-translation of the filter output or an adjustment of its amplitude. Furthermore, the τ themselves are independent of the detector position or orientation (up to a polarization rotation). Correspondingly, the templates to be applied to the pseudo-detectors are just linear combinations of the filters τ with appropriate time translations. These depend only on the orientation of the source and the detectors and can be reconstructed from our knowledge of **U** and the filter output corresponding to the template τ applied to each pseudo-detector.

Thus, each pseudo-detector need be correlated with a set of templates no greater in number than the number of templates involved in a single detector search. The output of the network (i.e., the value of the multi-detector likelihood) for a fixed source in a given location can be reconstructed from the output of the single detector templates for fixed astrophysical source parameters θ applied to the individual pseudo-detectors, with the appropriate linear combinations and time translations.

For a maximum likelihood search, the entire likelihood need never be held in memory at any one time. Correspondingly, the calculation can be organized on a Beowulf or supercluster system, with each node responsible for searching over Ω and Ω' for different values of θ. The computational cost is thus the cost of a single detector search times the number of detectors, plus an additional cost corresponding to the combinatorics of linear combinations and time translations.

Pipeline Architecture:

Three assumptions were important in considering the design of the pipeline architecture:

(i) Redundant and independent analysis of commonly shared data sets should be carried out. The reasons for at least two independent analyses include the ability to use algorithmically dissimilar approaches that provide for validation of scientific conclusions.

(ii) Individuals from all participating projects should be encouraged to join any one of the independent network. This strategy ensures that individuals with detailed knowledge of the instrumental characteristics of all detectors will be represented in all efforts.

(iii) We explicitly assume that all data involved in any candidate detections identified by the network analysis will be subjected to intense scrutiny and validation by the detector teams

All assumptions are regarded as crucial to obtaining reliable analysis results. The second and third assumptions particularly simplify the pipeline schema and reduce the amount of data that must be transferred and processed in the fully coherent search described above.

With these assumptions, the existing LIGO or VIRGO pipeline schema forms a baseline that can be extended to represent a coherent analysis pipeline as follows:

(i) Signal conditioning and preprocessing for individual interferometers proceeds as before. Each detector team is responsible for producing their best estimate of (a suitably whitened) h(t). This involves (a) line removal, (b) dropout correction, (c) calibration to strain, (d) bandwidth limited filtering and decimation, (e) linear regression against other interferometer channels to produce a best estimate of strain, etc. Only the reduced h(t) (suitably whitened) is involved in the later stage of the network analysis.

(ii) The detector network noise covariance is calculated using the vector of h[f] data from all interferometers. The inverse covariance matrix is diagonalized by finding the corresponding unitary transformation, *U*, which is applied to form the input data stream to the parallel pipeline (i.e., the pseudo-detector time series).

(iii) The parallel pipeline also uses *U* to calculate the rotated templates. The parallel analysis over pseudo-templates proceeds as with the single interferometer search except for a few changes in details. These include how polarization states are tracked and how the network-correlated time-of-arrival offsets are computed and used.

(iv) Candidate detections correspond to maximizing the network likelihood over source position, inclination, wavefront arrival time and intrinsic astrophysical source parameters.

Table 1: Principal computational costs for a binary inspiral search. An analysis involving the three km-scale projects (GEO, LIGO , VIRGO) would have $N_{IFO} = 5$.

Step in analysis	Relative Cost	Comment
1. Single detector parallelized template analysis	C_o	C_0 ~ 20 – 30 GFLOPS (sustained power) for searches down to 1 $M_\odot$ NOTE: C_0 is the dominant computational cost for this analysis.
2. Data sharing, transport & storage	C_1 NOTE: This is an annual budget item and NOT a one-time cost.	C_1 ~ \$100k/yr to maintain an open phone line between LIGO and VIRGO on a 7 x 24 basis. This should be an upper limit to the cost to share data in near-real time. Suitably decimated and reduced datasets are estimated to require ~ 32 kB/s per IFO.
3. Calculation of network covariance matrix continuously in near real time	$(N_{det})^2 C_2$	C_2 is the cost to calculate the noise denominator in a single interferometer analysis.
4. Data selection, event list sort from single interferometer searches	C_3	C_3 corresponds to a maximization of SNR for a single interferometer by searching over templates and times during the epoch of the FFT.
5. Calculation of strain and template pseudo-detector vectors.	$2N_{det}C_4$	This is a new step in the process for multiple interferometers.
6. Correlation of templates with pseudo-detector data.	$N_{det} C_0$	Linear scaling with number of interferometers.
7. Post detection maximization and search over extrinsic parameters.	$N_{\Delta\Omega} C_3$	$N_{\Delta\Omega}$ is the number of source positions that need to be tested.