LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO –
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type Report	LIGO-M050462-01-E	14 March 2006

Final Report for the period 2002-2006
NSF-DST Grant#INT-0138459

Albert Lazzarini

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 NW17-161
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

www: http://www.ligo.caltech.edu/

Extracting gravitational wave signals of inspiraling binary stars from laser interferometric data.

TABLE OF CONTENTS

LIST OF FIGURES

Extracting gravitational wave signals of inspiraling binary stars from laser interferometric data.

Project number: DST-NSF/RPO-077/2001

Period of coverage for the report: March 2002 to March 2006

Principal Investigators: A. Lazzarini (US) & S. Dhurandhar (India)

Co-investigator: S. Finn (US)

1. OUTREACH AND EDUCATIONAL ACTIVITIES:

1.1. India to US

During each year of the grant, one or more trips were taken to the U.S. by our Indian collaborators. These trips entailed a number of activities:

- Visits to Caltech and other LIGO Laboratory sites, including the observatories. Visits included presentation of seminars on ongoing collaborative research activities.
- Visits to The Pennsylvania State University to continue collaboration with the grant co-Investigator.
- Attendance at LIGO Scientific Collaboration meetings. This included presentation of talks to collaboration working groups on the ongoing research.
- Performing science monitoring duties during the LIGO observational science runs;
- Attending grid technology computing workshops.

In addition to the PI (Prof. S. Dhurandhar), several students participated in the program. Dr. A. Sengupta made several visits to the U.S. He is presently a postdoctoral scholar at the LIGO Scientific Collaboration member institution, The University of Cardiff, UK. Mr. S. Mitra joined the collaboration as a new graduate student during the second half of the award period. He is completing his doctoral research now and he is expected to accept a postdoctoral appointment at one of the LIGO Scientific Collaboration institutions in the U.S.

1.2. US to India

During each year of the grant, U.S. collaborators took one or more trips to India. These trips enabled us to:

- Attend annual Indian Association of General Relativity and Gravitation meetings and to give invited talks on LIGO.
- Continue our collaboration with the larger team of IUCAA members of the LIGO Scientific Collaboration.
- Review the status of the research program, begin development of publication drafts, and plan for the next year's program of research.

In addition to the PI and co-Investigator, two young faculty and scientists participated in the exchange program. Dr. Peter Shawhan from LIGO Laboratory at Caltech made a visit during the early phase of the program to help orient the IUCAA group on grid computing methods for data analysis. Prof. S. Bose of Washington State University took the final trip. He joined the collaborative research program at a later stage when a new research project was initiated that involved his expertise in cross-correlation techniques in data analysis.

2. HIGHLIGHTS AND RESULTS OF THE RESEARCH PROGRAM

2.1. Hierarchical Search for Inspiraling Binaries

It is important to reduce the computational cost of the search, because the usual flat search is quite expensive even in the simple case when spins are ignored. The method becomes vital if the number of parameters is increased, e.g., when spins are taken into account.

1. In the initial analysis of the hierarchical search for gravitational wave signals from inspiraling binaries, the saving in the computational cost was roughly estimated to be over 100 as compared to the usual flat search for the LIGO I noise floor spectrum. This factor can be deduced considering the ratios of the areas in stages I and II of the search. The hierarchy was over the two masses and the time of arrival of the signal; first time over three parameters.
2. Taking into consideration the boundaries and rotation of templates of the deemed parameter space, the realistic cost gain factor was estimated between 60X - 70X.
3. A hierarchical search code was developed in the LDAS (LIGO Data Analysis Software) environment and implemented on real data obtained from the first and the second science run S1 and S2 respectively. Because it was real data, the noise had non-stationary and non-Gaussian features. The code nevertheless gave a factor from 6 to 10 in spite of it not being optimized.

This research formed the basis of the doctoral thesis for Dr. A. Sengupta.

2.2. Interpolated Search for Binary Inspirals

The search exploits the correlations between the filtered output to extract the signals with the same efficiency but with fewer templates. Chebyshev interpolation is used because of its minimax property. This type of search is vital if the number of parameters is high as in a multidetector search in which the computational costs are prohibitive.

1. The cost saving factor is shown to be 3.5 per dimension when the minimal match criterion is used.
2. The cost saving factor is about 30 % per dimension if the ROC curves are used which is a more fair comparison which involves comparing false alarm and dismissal properties for the two type of searches.
3. Hierarchical search strategies can be used in conjunction with the interpolated search especially in the trigger (1st) stage to boost the cost gain factor

2.3. All Sky Map of the Stochastic Gravitational Wave Background

The stochastic background of gravitational waves is a promising avenue for exploring astrophysics as well cosmology. In addition to a nearly isotropic relic background from the early universe, stochastic backgrounds can arise in the confusion limit of astrophysical sources. The strong anisotropy in the GW background of the latter would be an invaluable discriminator between the underlying sources.

During the last year of the program, a new research effort was initiated in this area.

The research described in this and the previous section forms the basis of the doctoral thesis for Mr. S. Mitra.

3. DETAILS OF THE RESEARCH

3.1. Introduction:

Inspiraling compact binaries consisting of neutron stars or black holes have been of great interest to the gravitational wave community and particularly to the data analysts for almost fifteen years. Their popularity among data analysts stems from the fact that they can be modeled cleanly, in that, the compact objects can be approximated as point masses; their gravitational waveforms can be obtained to adequate degree of accuracy by computing post-Newtonian terms to high order. The GW signal emitted by the binaries is a `chirp' and the waveform is well suited for broadband detectors such as the laser interferometric detectors. Because the waveform can be modeled sufficiently accurately, matched filtering techniques can be employed to optimally extract the signal from the interferometric data. In recent years, the discovery of several such systems in our galaxy has led to the prediction of a higher event rate for the inspiral-merger events, which has made the search for inspiraling binaries even more relevant.

In the past, several strategies have been devised in searching for inspiral signals from interferometric data. The most straightforward method is the flat search, which involves scanning the parameter space with a dense bank of templates. Although this type of search is rigorous, it entails huge computational costs. The online computational speed required is of the order of a few hundred GFLOPS (1 GFLOPS = 10^9 FLoating point OPerations per Second) when searching in the individual mass range from 0.2 – 30 solar masses in a frequency range from 10 Hz to a few kHz at a minimal match of 0.97, even when spins are ignored. If spins are taken into consideration, the computational cost can sore a factor ~ 100X greater than the spinless case. Thus, there is a compelling reason to devise efficient search strategies in order to reduce the computational burden; one must reduce the computational cost *without sacrificing performance*.

With this goal in mind, this project investigated two strategies:

- Hierarchical search techniques;
- Interpolation techniques.

3.2. Hierarchical Search for Inspirals:

The essential idea is to carry out the search in two stages:

(i) Trigger stage: the parameter space is coarsely sampled with a low threshold;

(ii) The first threshold crossings are followed up with a fine search with a high threshold.

The total gain factor was estimated to be about a 100X over the flat search. However, it was found that because the parameter space tapers near the low mass end, the tiling is inefficient which reduces the gain factor between 60X – 70X for the LIGO I noise floor spectrum. This is still very useful.

In the first paper entitled *Extended hierarchical search (EHS) algorithm for the detection of gravitational waves from inspiraling compact binaries, A. Sengupta, S. V. Dhurandhar, A. Lazzarini & T. Prince, Class. Quant. Grav. **19**, 1507 (2002),* the analytics of the extended hierarchical search was presented. The two-step hierarchical search first searches with a fewer, but coarser set of templates and then the candidate events are followed up with a fine search. This method tends to reduce the computations for the search. By extending the hierarchy over the time-of-arrival parameter, we gain an *extra* factor of 4 when the hierarchy is performed only over the masses of the binaries.

The following panel of graphs in Figure 1 summarizes our findings. Refer to the caption for details.

The second part of the work was on the implementation of the hierarchical search. The main goal here was to design an algorithm to place the templates in the parameter space for the first stage of the hierarchical search. This is a tiling problem where one must fully cover the parameter space without leaving any `holes'. The algorithm cuts out rectangular tiles within the contours and the entire parameter space is covered by these rectangles. The algorithm takes into consideration the rotation and varying sizes of the rectangles across the parameter space. A paper on this work has been published: *Faster implementation of the hierarchical search algorithm for detection of gravitational waves from inspiraling compact binaries, A. Sengupta, S. V. Dhurandhar & A. Lazzarini, Phys. Rev. D 67, 082004 (2003)*. It is available at http://xxx.lanl.gov/gr-qc/0301025 .

The focus of the research was to design an algorithm to place the templates in the parameter space for the first stage of the hierarchical search. This is a tiling problem where one must fully cover the parameter space without leaving any `holes'. The algorithm cuts out rectangular tiles within the contours and the entire parameter space is covered by these rectangles. The algorithm takes into consideration the rotation and varying sizes of the rectangles across the parameter space. This must be achieved in real time if the power spectral density of noise (PSD) changes frequently. Therefore, an algorithm, which places templates efficiently and rapidly, is crucial.

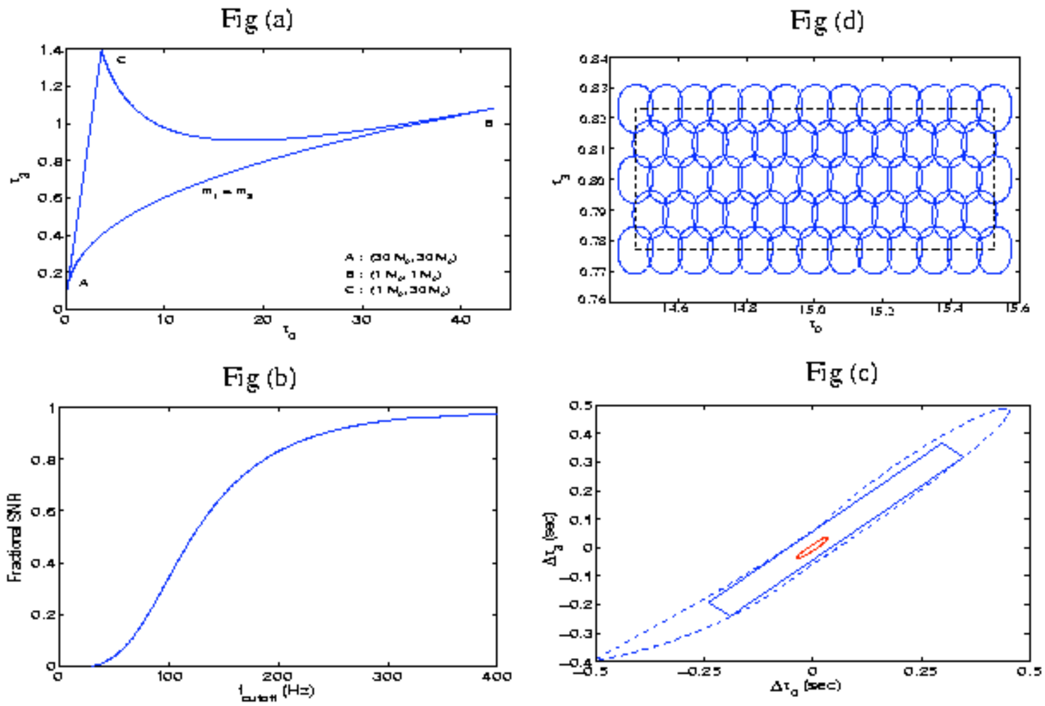


Figure 1: (a) depicts the parameter space in the two chirp times; (b) shows that even if one cuts off the signal at 256 Hz which is quarter of the usual sampling frequency of 2048 Hz, there is enough power – about 92 % - to carry out the hierarchy over the two chirp times. (c) and (d) compare the size of the templates in stages I and II in the two masses. (d) shows that several - about 50 or so - second stage templates fit into first stage templates. Sampling in the first stage at quarter rate gives an additional factor 4X. The overall factor however, suffers a reduction since the parameter space tapers at the low mass end making the tiling in the first stage inefficient.

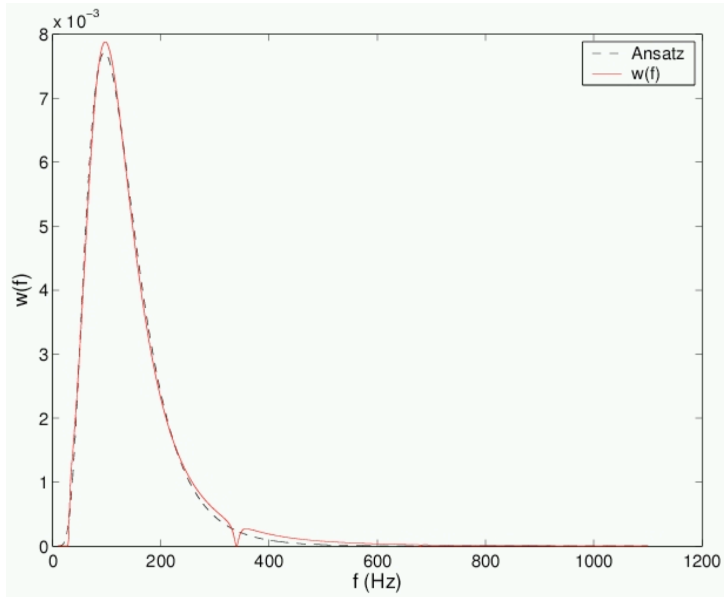


Figure 2: A log-normal distribution (dashed line) is fitted to w(f) = f - 7/3 Sh (f) shown in red. The sharp peak in w(f) is responsible for making the ambiguity function almost singular at its peak. This has implications for the Chebyschev interpolation in the part II of the project.

3.2.1. Fast placement of templates:

The template placement algorithm has been accelerated using the stationary phase approximation to obtain an approximate analytic expression for the ambiguity function even for large mismatches. The part $f^{-7/3} S_h(f) = w(f)$ is modeled as a log-normal distribution (shown in the figure) while the signal phase is expanded to the second order. The integral is performed using the saddle point approximation.

This approximation allows us to obtain the orientations of the rectangles analytically, so that the length and the width of the template can be obtained by solving for the contour boundary numerically in just four directions instead of all around the contour. This makes the template placement fast – seconds as compared to hours.

3.2.2. The clustering problem:

When a signal is present in the data, several templates are triggered in the bank. The problem is of determining the signal parameters from this cloud of events, namely that of condensing the cloud. We are exploring an algorithm based on graph theory, whereby choosing a connected

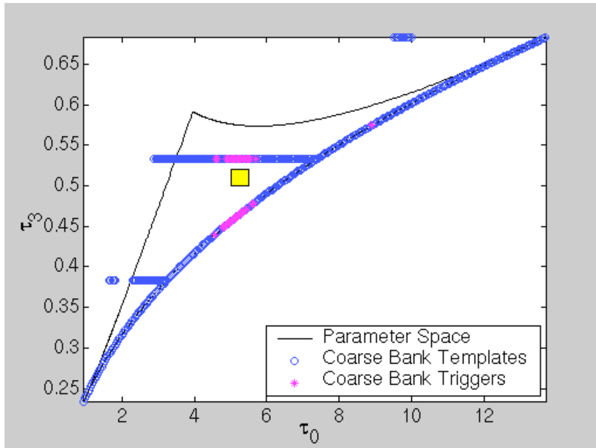


Figure 3: The figure shows the templates that are triggered (pink) while the templates not fired (blue). The actual signal lies in the yellow box. As is evident most triggers lie close to the signal, although a few outliers are present. Our spanning tree algorithm picks out a connected component of adjacent templates, the one containing the template with the highest SNR. Thus, the outliers are eliminated. The signal parameters are computed from those of the templates lying in the connected component.

component of the triggered events as explained in Figure 3 eliminates outliers.

3.2.3. Testing the code on real data:

The code for the hierarchical search is ready. The algorithm has been implemented on a stand alone machine and implemented in the LDAS environment. The performance of the code on real data from the S1/S2 runs was the final goal.

To test the validity of the code a signal of SNR 10 was injected in 1000 realizations of Gaussian noise having typical S2 noise PSD. The parameters of the injected signal were randomly chosen from the parameter space. The signal then was extracted using out code. The results are shown in Figure 4.

3.2.4. The EHS pipeline:

The EHS (Extended Hierarchical Search) pipeline for data analysis of gravitational wave data has been developed on LDAS (LIGO Data Analysis Software), which is a custom development environment created by the gravity-wave community to meet the requirements of different data analysis groups. The LDAS has been created to detect and characterize gravitational waves from astrophysical sources. In addition, LDAS has the following other functions like raw data archival, database management functions for raw data, descriptors of diagnostic triggers and astrophysical events data, and metadata distribution services. The LDAS consists of dedicated hardware and follows a computer centre model. Machines are on a private network and the access is via Internet through a single gateway machine. Remote job submission and result retrieval are possible via gateway. It has a socket based job submission protocol and no Unix login by users is necessary. Access requires LDAS login and password. At present LDAS systems exist at many places - systems are available at Caltech, MIT, and a few institutions that are a part of the LIGO Science Collaboration (LSC). LDAS systems are also available at both the LIGO detector sites at Hanford and Livingston.

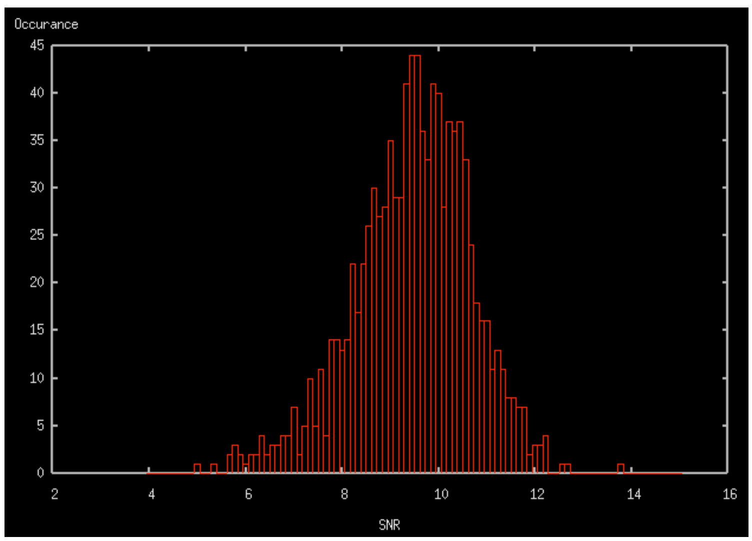


Figure 4: A signal with SNR 10 and randomly chosen parameters is extracted using our code from simulated data having a typical S2 noise PSD. Most of the extracted events cluster around the SNR of 10. Very few events appear with SNR less than ~ 8.2 the threshold. The results are consistent with the detection probability envisaged.

The illustration in Figure 5 shows how the data once ingested in the pipeline is analyzed inside the DSO. We begin at the point where gravitational wave data from a data server is fetched or 'queried' by the driver script. Strictly speaking, this is not, a part of the DSO itself. The query is specified inside the driver script along with some basic data conditioning algorithm (like fetching the PSD of the data, fetching the response function over the query time etc) that is then executed by the LDAS components before being passed on to the DSO in the internal light weight data (ILWD) format. Two streams of data and their corresponding PSD is fetched - one that is sampled at 1024 Hz is used in the coarse bank level of the hierarchy, and the other sampled at 4096 Hz is used for matched filtering in the second stage of the EHS. Here we pause to remind the reader that the AS_Q channel has a maximum sampling rate of 16 kHz, and the downsampling algorithms are available as LDAS commands.

Once the data are available, we perform additional conditioning of the data inside the DSO itself. This includes (a) extraction of the response function from the reference calibration spectrum and

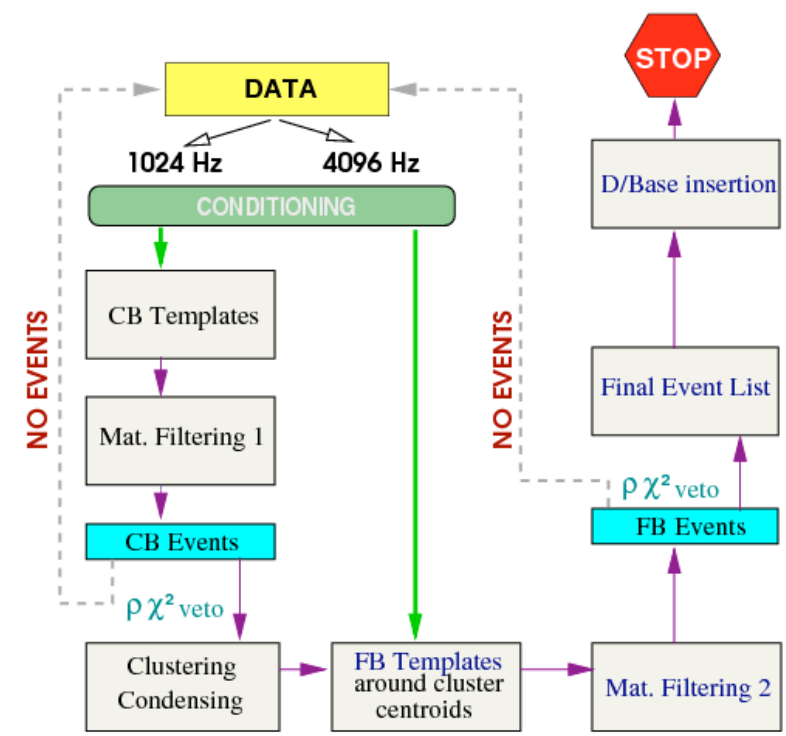


Figure 5: LDAS data analysis pipeline implementation.

time dependent coefficients (b) Calibrating the data using the response function thus obtained and (c) if performing a Monte-Carlo study, we also need to inject a software signal in both the data streams.

All the nodes participating in the search perform the above procedure. At the end of DSO specific data conditioning, the slave nodes are requested to use the PSD of the 1024 Hz data stream and generate the appropriate coarse template bank covering the entire deemed parameter space that spans a specified mass range. The slaves report to the master node, which in turn partitions the job of matched filtering equally between the slave nodes. Communication between the participating nodes is made using standard Message Passing Interface (MPI) commands. The slaves then execute the matched filtering of the coarsely sampled data against the apportioned part of the template bank and records the events. After the completion of every 25 % of the job, the slaves return the control to the master node, which in turn updates a parameter via which the LDAS manager is made aware of the progress of the entire job. When the slave returns the control, it saves a snapshot of its state so that it 'remembers' the position where it had left off, and starts from that point in the next call. In concrete terms, if node #2 has been assigned the job of reporting the matched filtering against 100 templates, it returns control first after finishing 1 - 25 templates and saves the state, so that when the job is restarted on this node begins at the 26-th template, and so on.

Once any node finishes processing all the allotted coarse bank templates, it communicates all the results to the master node. After the master node has collected results from all the slave nodes, it uses the coarse bank thresholds to identify an initial list of coarse bank triggers. If no triggers are generated, the job is terminated and the control goes back all the way to the driver script which restarts another pipeline job using a new set of data. On the other hand, if there are surviving triggers at this stage, then a clustering module is called from the master node which decided which one of these triggers need to be followed up by the next level of hierarchy. The clustering module returns 'cluster centroids' or seed points around which the second layer of templates need to be constructed.

The master node now communicates the list of seed points to all the slaves - each of which (after receiving the list) call the second stage template bank generation module. The master node does not know a priori how many fine bank templates have been generated. The slaves now report this information. The master node now again partitions the fine bank templates between the slave nodes. The fine bank correlations are now carried out by the slaves in a similar fashion as described for the coarse bank level matched filtering.

It may so happen that the number of fine bank templates are less than the total number of slaves participating in the search (this is unlikely in the coarse bank) - in which case the master node should decide which one of the slaves to 'rest'. The 'rested' slaves are instructed to return control for the last time.

After the fine bank matched filtering stage, the slaves return the second stage of trigger events. We retain all the trigger events at the end of this stage that are above a specified SNR threshold - in other words, no clustering is done by the DSO itself on the fine bank events. We have chosen this strategy in order to retain the flexibility of post processing the triggers after they are recorded from the second stage of the EHS pipeline. Note also that the choice of first stage thresholds is very crucial as the clustering and event identification is done by the software on-the-fly. Thus, it is good idea to keep the coarse bank thresholds a little loose - so that we do not miss any true gravitational wave event from being passed on to the next level of the hierarchy.

The final set of pipeline triggers are recorded by the master node, after which all the participating nodes (master and slaves) are rested by returning control.

3.2.5. Computational performance of EHS:

Figure 6 shows the results of the timing tests of the EHS in the two different modes where we plot the time taken to complete the pipeline job for a single data chunk of 256 sec as a function of the number of CPUs used in the search. Note that in order to thwart the effect of likely variation in the timing due to varying load on the LDAS and the peculiarities of the data that is being analyzed, we have plotted the average time over 4 runs using four different data samples each of 256 sec. The numbers are dependent on many parameters - for instance, they change quite a bit with the thresholds, the minimal match levels used for template placement etc (and of course with the number of CPUs). This is just a representative plot.

If the parallel program is very efficiently designed, it should scale as the inverse of the number of CPUs. However, in any parallel program, there are communication overheads involved in transfer of data or information between CPUs that participate in the execution of the parallel program. In addition, parts of the parallel program are inherently 'serial' because of which the true scaling is not 1/n (where n is the number of processors used in the job). In the top panel of figure, we plot the ideal scaling law as broken lines (for the flat as well as full EHS search modes), as a visual aid to appreciate the departure of the true (observed) scaling from the idealized situation.

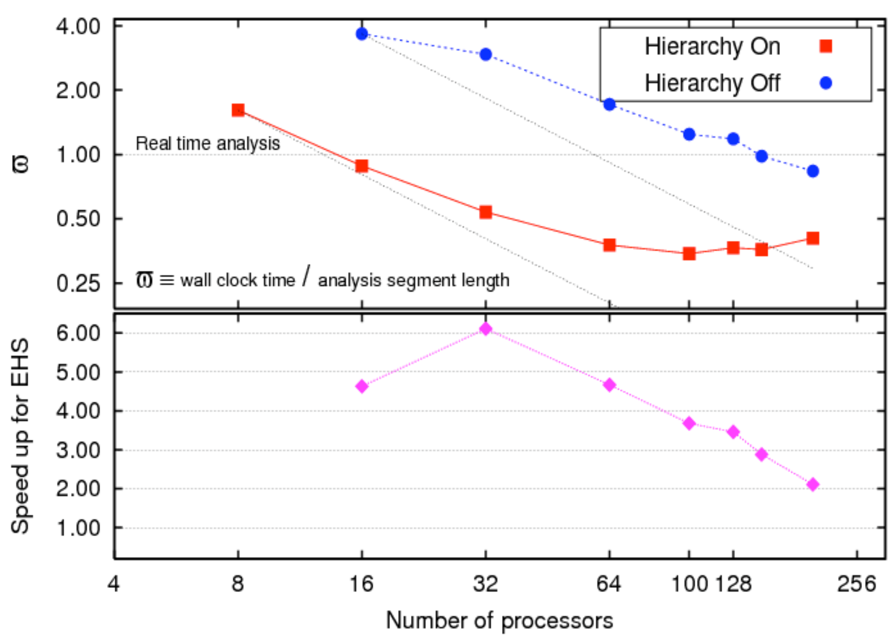


Figure 6: Performance metrics of the algorithm.

One immediately notices that for the case of EHS with hierarchy flags turned on, there is a critical number of CPUs above which the wall clock time starts increasing as the inter-processor communication over heads start to show up. However, for the other case where the hierarchy flag has been turned off, the scaling law holds over a large number of CPUs. This is understood from the fact that the flat search is an embarrassingly parallel code - where (in the master-slave design of code), the template bank is divided equally by the master node between the slave (worker) nodes. There is hardly any need for inter-processor communication as once the slave nodes are assigned (or made aware of) their share of the template bank; they can report the result of the analysis after completing the full subset of the pipeline job. We pause to note that in the LDAS system this is not strictly true - the slaves have to send heartbeat (progress) messages to the master node once in every 300 sec, otherwise they are stopped by the LDAS manager. This is a trivial example of program check pointing so that the LDAS manger can make an inventory of the available resources at any given time. However, this message can be done using 'non-blocking' MPI communications where the slave node does not wait for the master node to acknowledge the heartbeat message and carries on with its share of job immediately after sending the message.

The computational speed-up factor is also shown in the bottom panel of Figure. For the particular choice of run parameters used, we conclude that if the search were carried out over 100 CPUs, it would be 4X faster than the flat search. In terms of the comparison of the number of CPUs required to perform an on-line search, we find that if the hierarchical mode is 'on', a factor 10X less number of CPUs are required. Note that these factors are 'unoptimized' as we have not tuned the pipeline over mass parameters to get best possible performance. If the pipeline is optimized, the gain in time and CPU may be larger than those quoted above. Also, note that the speed-up factor is vastly reduced. This is due to the fact the number of seed points that are passed on from the first stage to the second stage of the hierarchy are much greater than 1 (as is assumed in the table). This is a 'surprise' that we get when faced with real data from interferometers, where the number of triggers are exceedingly large. This problem may be alleviated to some extent if the code is optimized over the mass hierarchy. This also underlines the importance of clustering tools in the EHS algorithm. With the use of smarter clustering algorithms, it may perhaps be possible to reduce the number of seed events that followed up in the fine bank level of the hierarchical search. This will increase the computational gain factor by a large margin.

3.3. Chebyshev Interpolation:

This project was motivated by the following observation: It has been noticed that the filtered output - the match between the templates and the data - from the fine bank in the flat search is statistically correlated, if the templates lie close to each other in the parameter space. When the minimal match is as high as 0.97, the templates are so closely spaced in the parameter space, that the filtered output is significantly correlated. Thus, one is oversampling the parameter space; the same information should be obtainable from fewer samples; that is, coarser sampling should suffice for constructing the full match; that is, a kind of Nyquist theorem should be invoked.

For the sake of simplicity, we will consider the spinless case and address this question pertaining only to Newtonian waveform which is characterized by only one intrinsic parameter: the chirp time τ_0. Thus, we are essentially dealing with a one dimensional parameter space. It may be noted that this is the more relevant parameter for analysis, as very large number of templates are required in the τ_0 dimension than in the τ_3 dimension, more than a hundred times more for the usual noise curves such as the LIGO noise curves. Our goal is to explore the possibility of coarsely resampling the match, which is the surrogate of the likelihood ratio, with the primary goal of reducing the computational costs and secondly, also for the sake of principle, of doing the problem more cleanly. This is important for future investigations that would benefit from a cleaner basic analysis. Here, we employ *Chebyshev interpolation*, to construct the match at all points in the parameter space from its sampled values. We expect that fewer templates will be needed to construct the full match than the number of templates required in the usual flat search in order to obtain the same level of performance.

We have opted for Chebyshev interpolation over other types of interpolation for the following reason: we desire that the function interpolating the match differ from the match by less than a pre-assigned value of the mismatch. This was also the criterion in constructing the template bank for the flat search. Now, it is known from the literature that, for a given continuous function defined on an interval (the match in our case) and for a given degree of the polynomial, there exists a polynomial whose largest difference between it and the function is smallest among all polynomials. Such a polynomial is called the *minimax* polynomial. The minimax polynomial is however hard to compute; but, it can be shown that the *Chebyshev interpolating polynomial* is often just as good. We use the Chebyshev interpolating polynomial to construct the match to the desired level of accuracy. The match is constructed from its samples evaluated at the roots of the Chebyshev polynomial. However, we find that the match is not so well behaved and possesses a cusp-like structure in the neighborhood of its peak, when the signal lies close to the template. We remedy this by placing a template at the maximum of the interpolating polynomial, thus avoiding the need of high degree interpolating polynomials for attaining the requisite match. In order to compare fairly the two approaches, we construct the false dismissal versus false alarm curves – these are the so-called Receiver Operating Characteristic (ROC) curves. We find that the Chebyshev interpolation method performs better - that is, the false dismissal is less for the Chebyshev method than the flat search for a given false alarm.

Specifically, we consider a Newtonian chirp with upper frequency cut-off at 512 Hz and the Science Requirements Document (SRD) noise curve for LIGO-I. We consider a smaller region of the parameter space in τ_0 of length 4 sec. The full range of τ_0 reaches up to ~ 43 sec when the lower limit on inspiraling masses is 1 solar mass and the fiducial frequency is 40 Hz; the binary takes about 43 seconds to coalesce from the epoch when the instantaneous frequency of the chirp signal is 40 Hz. We plot the false dismissal versus false alarm curves for various numbers of templates from 30 to 133 that correspond to minimal matches from 0.78 to 0.97 respectively. We find that the interpolated search requires about 75 % number of templates compared to that of the flat search in order to achieve the same level of performance. On the other hand, if we just consider the minimal match criterion, the interpolated search requires a factor of 3.5 less number of templates from that required in the flat search. Specifically for the minimal match of 0.97 we need just 37 + 1 templates in order to achieve the desired match, while the flat search requires 133.

3.3.1. The interpolated search procedure

In the flat search, we had obtained the match between the signal and the templates at discrete points in the parameter space. In the interpolated search, the goal is to obtain the match *everywhere* in the parameter space by interpolating between the discrete set of values of the match determined by the templates. However, we found that the Chebyshev interpolation is not so effective because the match or the ambiguity function has a cusp like structure near the maximum. Because of this, we need very high degree polynomials, thus a large number of templates to attain the desired match. This is a general feature that arises because the seismic noise rises steeply at low frequencies - the seismic `wall' - combined with the fact the signal power drops off also steeply as $f^{-7/3}$. These features of the signal and the noise produce a sharp peak in the differential SNR, namely, as a function of frequency, which then leads to a cusp like structure for the ambiguity function near its maximum.

We can remedy this situation by the following procedure:

(i) Fix the degree n of the Chebyshev polynomial;

(ii) Compute the correlations of the templates placed at the roots of the Chebyshev polynomial;

(iii) Compute the Chebyshev interpolating polynomial;

(iv) Find the maximum of the polynomial and place a template at this τ_0 and compute the correlation.

The last step removes the cusp and we are able to get much better match with the above procedure.

3.3.2. : Performance: minimal Match criterion

We compare the performance of the interpolated search with the flat search in terms of their respective efficiencies in detecting the inspiral chirp signal from detector noise.

The usual method to compare the template placement efficiencies of two algorithms uses the Minimal Match (MM) criterion. In this method, templates are placed in such a way that the MM is greater than a certain pre-decided value (which is often 0.97). The algorithm that requires a smaller number of templates to satisfy the criterion is the better one. Further the ratio of the minimum number of templates required, by the two methods under consideration, to satisfy the criterion gives us essentially the gain or the speed up factor in the computational costs, where we have ignored overheads.

We restrict ourselves to a 4 second range in the τ_0 space. The panel of four graphs in Figure 7 summarizes the results.

Comparing, the flat search requires about 133 templates, to achieve the minimal match of 0.97 while just using 37 +1 templates suffice for the interpolated search. Thus with the interpolated search the cost saving factor is about 3.5 with the minimum match criterion.

3.3.3. False Alarm versus False Dismissal criterion

We compare the performance by plotting the False Alarm (FA) versus False Dismissal (FD) probability curves. Since in real experiments the signal is buried inside relatively high amplitude noise, a data analysis pipeline will miss several such signals and many times, it may claim wrong detections, so in short, there will be false dismissals and false alarms. Both of them should be minimized in order to have a reliable algorithm in searching for gravitational wave signals. Therefore, an algorithm performs better than its competitor if it gives less false dismissal for a given false alarm for the same computational cost.

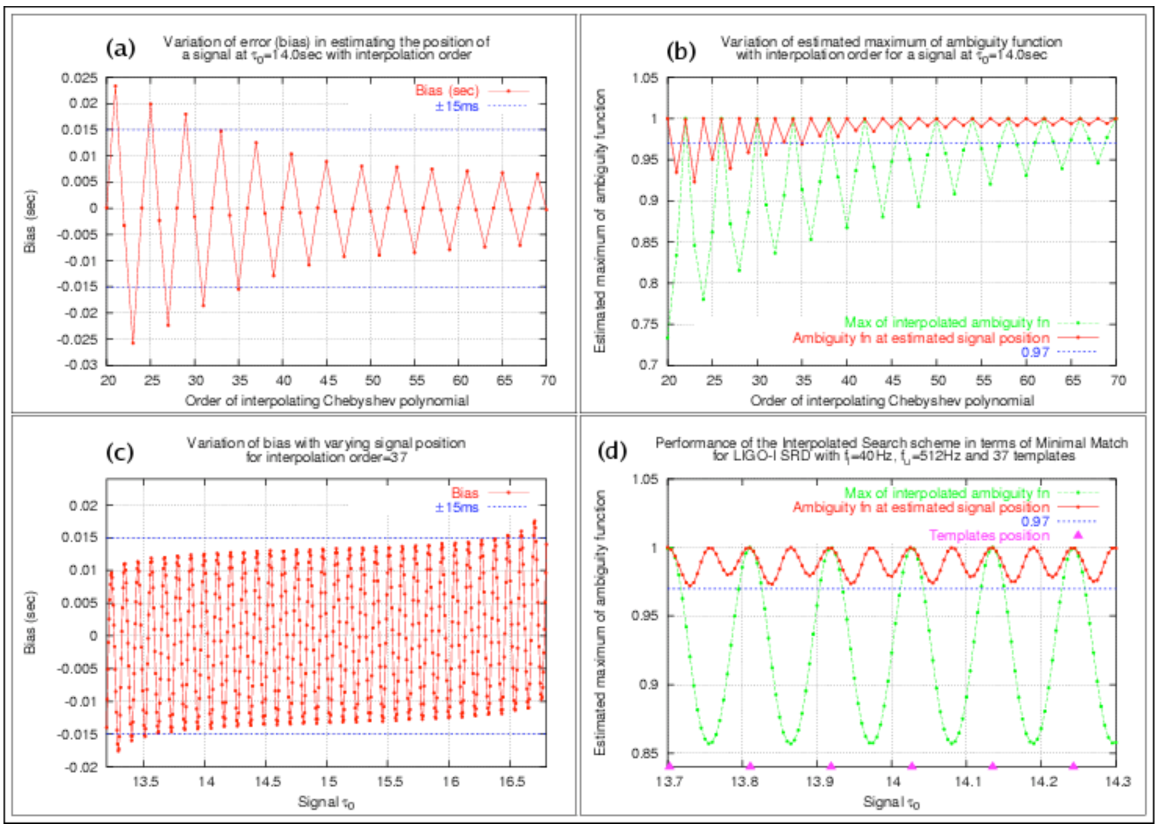


Figure 7: In general, the maximum of the interpolating polynomial is displaced from the true signal position. This is shown in (a) how the bias changes as a function of the degree of the interpolating polynomial, where the signal parameter has been chosen at a non-special position. In (b) the mismatch is shown as function of the degree of the interpolating polynomial. Both figures imply that the degree of the polynomial must be greater than 35 in order to obtain a match of better than 0.97. In Fig. (c) and (d) we fix the degree of the polynomial to be 37. (c) shows the bias while (d) shows the match as the signal parameter τ_0 is varied. The solid (red) curves correspond to placing a template at the maximum of the interpolating polynomial while the dotted (green) curves correspond to the maximum of the interpolating polynomial. As is evident, the solid curves show significantly improved match over using just the maximum of the interpolating polynomial. One observes in (d) that the match is better than 0.97 for the solid (red) curve.

To plot the FA versus FD curves we perform numerical simulations. The detailed procedure is as follows:

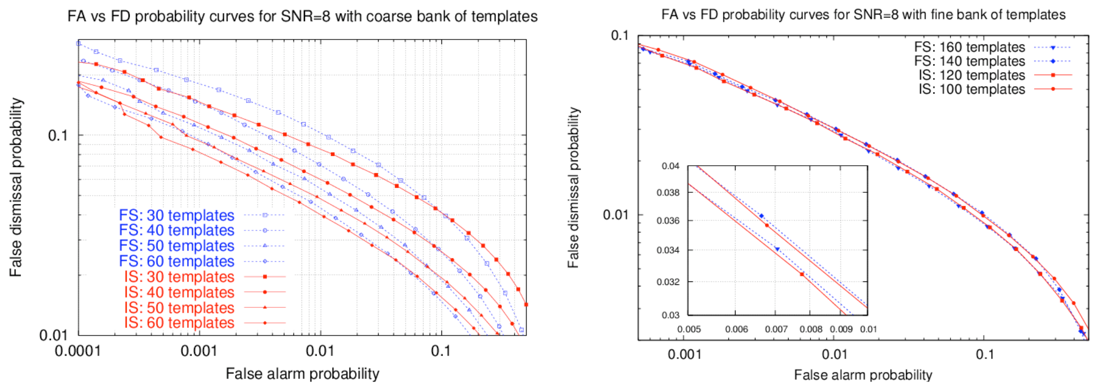
False Alarm probability: We generate 32 sec data trains sampled at 1024 Hz containing colored Gaussian noise where the PSD is that of LIGO-I SRD noise curve. We first generate white Gaussian noise directly in the Fourier domain using the Mersenne Twister Pseudo Random Number Generator. We then color it by multiplying each Fourier component by square root of the PSD.

We apply both the algorithms - the flat search and the interpolated search on these realizations. We count the number of events above a fixed threshold and divide by the total number of realizations in order to obtain the false alarm probability. Typically, we use about 50,000 realizations, which give reliable results for the FA probability up to ~ few × 10^{-4}.

False Dismissal probability: In order to compute the false dismissal probabilities we need to inject signals into the data trains. Due to limitations on the available computational capacity, we can not inject sufficient number of signals uniformly all over the parameter space. Rather we assume the periodicity of the detection probability over the parameter space with the period equal to one template separation. Accordingly, we inject 10-20 signals of SNR = 8 uniformly between two successive templates one at a time. This injection scheme implicitly assumes a uniform distribution of signals in τ_0. For each signal injection, we then count the number of crossings above a threshold and divide by the total number of realizations to get the detection probability for that placement of the signal. This detection probability is a function of τ_0. The detection probability is then averaged over all the injected signals to get the average detection probability over the parameter space between two adjacent templates. Then the periodicity implies that this is the required average detection probability over the full parameter space. The average false dismissal probability equals *1 – (average detection probability)*. This procedure is carried out for the flat search and the interpolated search to obtain the required FD versus FA curves.

The panels in Figure 8 show the results of our simulations. The panels show the false alarm versus false dismissal curves for the flat search (a, dotted curves) and the interpolated search (b, solid curves). Figure 8a shows the curves for relatively smaller matches - MM for the flat search lies in the range 0.78-0.87 – with relatively smaller number of templates as are used in the coarse bank in the first stage of the hierarchical search. The number of templates for each search is chosen to be 30, 40, 50, and 60. The curves for both searches `move downward' in the diagram with increasing number of templates. For a given number of templates, the solid curves are lower – less false dismissal for the same false alarm – than the dotted curves in the regime of low false alarm showing that the interpolated search performs better than the flat search at low false alarm. We have used typically 50,000 realizations for each of the simulations. Therefore, the curves show significant numerical error below a false alarm of few times 10^{-4}. Near a false alarm probability of 10^{-3}, we can see that only 30 interpolated search templates perform as well as 40 flat search templates.

Figure 8b shows that a similar gain factor is achieved even for high matches of ~ 0.98 which are usually chosen for the fine search. We require just 120 templates for the interpolated search as compared to 160 for the flat search or about 100 templates as compared to 140 (the zoomed inlet



(a) (b)

shows this even more clearly). Therefore, the gain in the cost factor is at least 30%. Another way of looking at this is that the interpolated search loses about 20% less signals than the flat search using same number of templates for a false alarm of 10^{-3}. We also conjecture by looking at the trend of the curves that the interpolated search may perform better at lower false alarms, which correspond to the astrophysically realistic regime considering the current estimates of the event rates.

To summarize, we obtain a gain factor of about 30 percent over the flat search. This factor is expected to improve with increasing dimensionality of the parameter space when more parameters are included such as both masses and spin parameters. The other evident application is that of the efficient implementation of the hierarchical search. The interpolation technique can be used in the first stage to obtain the triggers, which could then be followed up with a fine search. Since the major component of the cost in the hierarchical search arises from the first stage, the gain factor in such a search could be boosted substantially to make the search even more efficient.

3.4. Work in Progress: All Sky Map of the Stochastic Gravitational Wave Background

The stochastic background of gravitational waves is a promising avenue for exploring astrophysics as well cosmology. In addition to a nearly isotropic relic background from the early universe, stochastic backgrounds can arise in the confusion limit of astrophysical sources. The strong anisotropy in the GW background of the latter would be an invaluable discriminator between the underlying sources.

The detection and mapping of the anisotropy in stochastic GW background bears strong semblance to the analysis of the cosmic microwave background (CMB) anisotropy and polarization which is also a stochastic field statistically described in terms of its correlation properties. The first goal of the IUCAA group will be to identify the key features that would allow us to adapt appropriate techniques from CMB analysis.

Raw sky map of the GW background is the signal convolved with antennae pattern. The IUCAA group will explore and compare the various image reconstruction techniques for the application to the GW background. The IUCAA group will identify and develop an efficient method to reconstruct the true sky map. Bayesian image reconstruction techniques will be explored. The measured stochastic GW background would have components arising from various astrophysical sources. The spectral signature of the sources would allow separation of the components. The IUCAA group is focusing on developing efficient methods related to optimal filters and their analysis. We are quite sure that the time dependent optimal filter for a certain direction (what is being used in the current analysis) maximizes the SNR over short correlation intervals if GWs are coming only from the pointing direction. This assumption becomes self-consistent if we could prove that the same filter significantly washes out the effect of the other points on the sky. Maps produced assuming a point source are a promising and required first step in developing a complete formalism (this is the subject of an MIT Ph.D. thesis). However, the gravitational wave foreground sky could be composed of multiple sources having comparable strength. In this case, then, an effective means to deconvolve the instrument response from the raw map is required. Now, we have an analytic derivation of the point spread function, we would like to check it numerically and compare with single point source analysis. We are exploring the stationary phase methods in order to understand the point spread function. We plan to explore inversion methods, like Maximum Entropy/Bayesian. A typical plot of the point spread function is shown

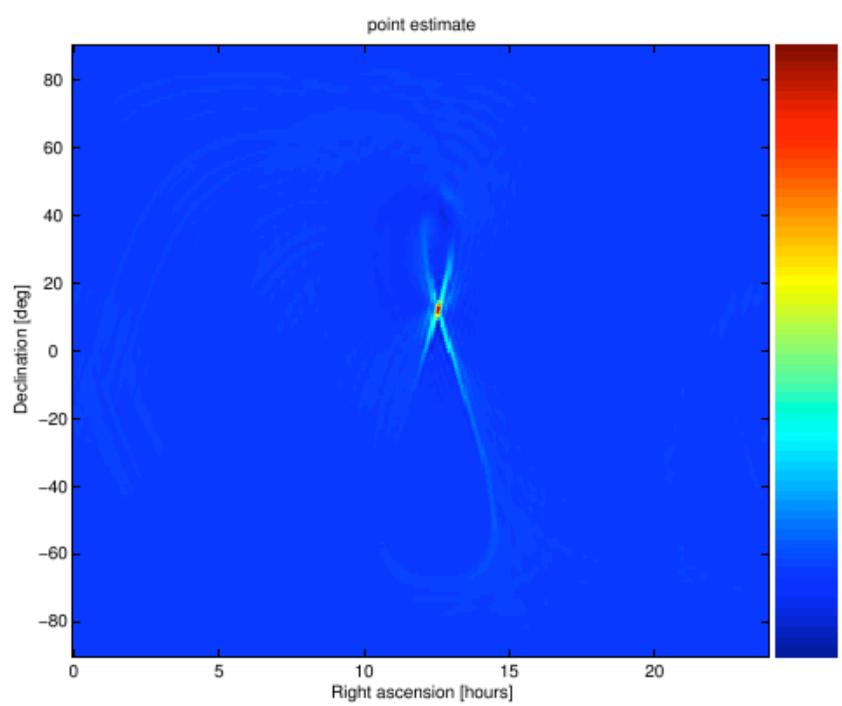


Figure 9.

in Figure 9.

We plan to perform simulations and injections of stochastic signals to test the developed detection strategy. These will be conducted both on simulated data (in software) and in hardware. The merit of hardware injections is that they are an "end-to-end" test, which means that they test the entire search pipeline starting from the swinging of the end-test masses / mirrors in the interferometer (to mimic the effect of an impinging GW) to recovering the injected signals by running the search code on the acquired data.

U.S. collaborators will work with IUCAA on researching the fundamental aspects of the formalism required for creating a sky-map of the astrophysical / cosmological gravitational wave background arising from a host of unmodeled or unresolved sources while using the full potential of the operating network of detectors by coherently combining their data. This work was initiated during the last year of the project and we plan to pursue it further.

4. PAPERS PUBLISHED/SUBMITTED FROM 2002-2005:

(i) *Extended Hierarchical Search (EHS) algorithm for the detection of gravitational waves from inspiraling binaries.*

Sengupta, S. V. Dhurandhar, A. Lazzarini and T. Prince, Class. Quant. Grav. **19,** 1507 (2002).

(ii) *Faster implementation of the hierarchical search algorithm for detection of gravitational waves from inspiraling compact binaries,*

A. Sengupta, S. V. Dhurandhar & A. Lazzarini, Phys. Rev. D **67**, 082004 (2003).

(iii) *Improving the efficiency of the detection of gravitational wave signals from inspiraling compact binaries: Chebyshev interpolation*

S. Mitra, S. V. Dhurandhar and L. S. Finn (to be submitted to Phys. Rev. D).