PRAMANA
— journal of
physics

Vol. 63, No. 4
October 2004
pp. 717–730

Data analysis techniques for gravitational wave observations

S V DHURANDHAR
Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India
E-mail: sdh@iucaa.ernet.in

Abstract. Astrophysical sources of gravitational waves fall broadly into three categories: (i) transient and bursts, (ii) periodic or continuous wave and (iii) stochastic. Each type of source requires a different type of data analysis strategy. In this talk various data analysis strategies will be reviewed. Optimal filtering is used for extracting binary inspirals; Fourier transforms over Doppler shifted time intervals are computed for long duration periodic sources; optimally weighted cross-correlations for stochastic background. Some recent schemes which efficiently search for inspirals will be described. The performance of some of these techniques on real data obtained will be discussed. Finally, some results on cancellation of systematic noises in laser interferometric space antenna (LISA) will be presented and future directions indicated.

Keywords. Data analysis; gravitational waves; signal extraction.

PACS Nos 04.80.Nn; 07.05.Kf; 95.55.Ym; 97.80.-d

1. Introduction

In the last one or two years several gravitational wave laser interferometric detectors world-wide are completing construction and are near to attaining sensitivities required to carry out serious science from gravitational wave (GW) observations of astrophysical sources. Great strides have been taken by experimentalists in improving the sensitivity of these GW detectors. In order that the detectors attain their requisite sensitivities, technology must be driven to its limits; this includes, the technology of lasers, laser-optics, vacuum systems, seismic isolation, electronics etc. Even with these enormous efforts, the GW signal is still very weak compared to the noise in the detectors. Sophisticated data analysis techniques are required to extract the GW signal from the noise and thus efficient data analysis is an important component in the experiment of GW detection.

In this talk I will review several important data analysis techniques used in GW data analysis. This talk is not meant to be comprehensive, but on the other hand, my goal here will be to bring out the flavour of the data analysis techniques that are being used and which will be applied in future on detector data.

2. GW signals and detector sensitivities

It is important to classify the GW signals by their various properties because the data analysis strategies depend on the type of source. GW signals can be broadly classified into those with parametrisable wave-forms and unmodelled sources such as bursts and transients. The first category of signals can be subclassified into signals that are deterministic and stochastic. An important type of source which falls into this subclass is the inspiraling compact binary. Two neutron stars or black holes spiral together in a bound orbit as they emit GW and fall towards each other. Just before they merge, perhaps for few tens of seconds or so, they emit a powerful burst of GW which has a characterstic wave-form depending on their masses and other kinematical parameters. Such a source is modelled on the Hulse–Taylor binary pulsar which emits a burst of GW radiation just before it merges. The typical GW strain – typical component of the metric perturbation – is denoted by h and for the inspiraling binary is given by

$$h \sim 2.5 \times 10^{-23} \left(\frac{\mathcal{M}}{\mathcal{M}_\odot}\right)^{5/3} \left(\frac{r}{100\ \mathrm{Mpc}}\right)^{-1} \left(\frac{f}{100\ \mathrm{Hz}}\right)^{2/3}. \tag{1}$$

Here $\mathcal{M}$ is called the chirp-mass and is given by $\mathcal{M} = \mu M^{2/3}$ where μ and M are the reduced and total mass of the system respectively, r is the distance to the source – it is given in terms of a large distance of 100 Mpc because such events are deemed to be rare and a large volume is required to obtain a reasonable event rate, and f is the typical instantaneous frequency of GW – it is taken to be 100 Hz because the detectors are likely to possess high sensitivity at this frequency.

Another type of deterministic source is the periodic or continuous wave source which continuously emits GW such as an asymmetric rotating neutron star or pulsar. The extent of asymmetry characterised by the parameter ϵ determines the strength of the waves. The typical h is given by

$$h \sim 1.9 \times 10^{-25} \left(\frac{I}{10^{45}\ \mathrm{g\,cm^2}}\right) \left(\frac{r}{10\ \mathrm{kpc}}\right)^{-1} \left(\frac{f}{500\ \mathrm{Hz}}\right)^{2} \left(\frac{\epsilon}{10^{-5}}\right), \tag{2}$$

where I is the moment of inertia of neutron star, r is the distance to the source and f is the GW frequency.

The stochastic GW source arises from numerous independent, unresolved sources which produce a signal that can only be characterised by its statistical properties. The strength of the source is characterised by a quantity denoted by $\Omega_{\mathrm{GW}}(f)$ which is defined as the energy–density of GW per unit logarithmic frequency interval divided by the critical energy density ρ_{crit} required to close the universe. The corresponding Fourier transform of the GW strain $\tilde{h}(f)$ for the frequency bandwidth $\Delta f = f$ is given by

$$\tilde{h}(f) \sim 10^{-26} \left(\frac{\Omega_{\mathrm{GW}}(f)}{10^{-12}}\right)^{1/2} \left(\frac{f}{10\ \mathrm{Hz}}\right)^{-3/2} \mathrm{Hz}^{-1/2}. \tag{3}$$

If one takes a glance at the sensitivities of GW detectors, say the LIGO detectors [1] of US, the design sensitivity of the initial LIGO [2] falls in the range of $h \sim 10^{-22}$ –

10^{-23} which will improve by an order of magnitude or so for the enhanced detector. But even then the signal strength is below or far below this raw sensitivity. Data analysis, therefore, plays a vital role in gravitational wave detection where the GW signal must be extracted from the detector noise.

The data analysis technique depends on the type of source. I will consider four types of sources, namely, (i) binary inspirals, (ii) periodic signals, (iii) stochastic background and (iv) bursts among unmodelled sources.

3. Binary inspirals

3.1 *Matched filtering*

It is well-known (see [3]) that the inspiral wave-form can be well-modelled in terms of post-Newtonian (PN) expansions. Resummation techniques such as Padé approximants and effective one body methods can be applied to extend the validity of the PN expansion [4], so that the wave-form is modelled adequately. The appropriate data analysis technique when one knows the wave-form well is that of matched filtering. It is the optimal strategy under the circumstances: it yields the maximum signal-to-noise ratio (SNR) among all linear filters, independent of the noise characterstics and secondly it is optimal in the Neyman–Pearson sense – if the noise is additive and Gaussian, it yields maximum detection probability for a given false-alarm rate. Let $x(t)$ be the output data from the detector, where t denotes the time, then the matched filter output $c(\tau)$ at time-lag τ is

$$c(\tau) = \int x(t)q(t+\tau)\mathrm{d}t. \tag{4}$$

Here $q(t)$ is the matched filter which for stationary noise (noise characterstics independent of absolute time) is given conveniently in the Fourier domain by

$$\tilde{q}(f) = \frac{\tilde{h}(f)}{S_h(f)}, \tag{5}$$

where $h(t)$ is the expected GW signal and $S_h(f)$ is the power spectral density (PSD) of the noise. The matched filtering operation has been shown in figure 1.

3.2 *Searching for inspirals: Spinless case*

However, one does not have a single signal to search for, but a family of signals. For inspirals, the family is parametrised by the amplitude, the time of arrival t_a, the initial phase – the phase at t_a, the two individual masses and the two individual spins of the stars comprising the binary. The data analysis strategy is to use the matched filtering technique. The parameter space is densely covered with a bank of templates, so that the mismatch between the signal and some template in the bank is minimal. Then one maximises the matched filter output over the template bank and compares the maximum with a pre-assigned threshold. The threshold is set

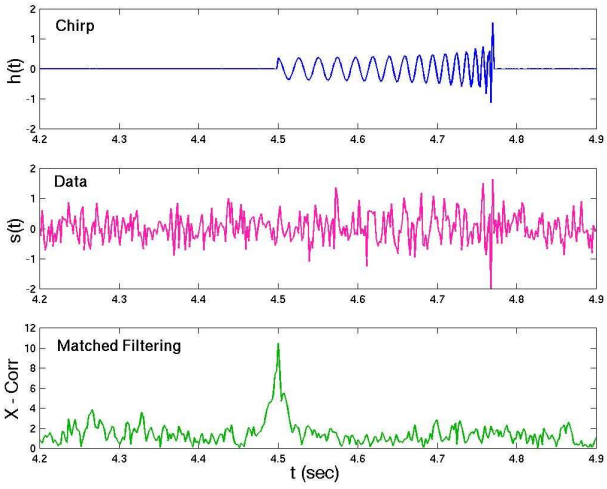

Figure 1. Matched filtering: The top part of the figure shows just the inspiral signal, the middle part shows the signal embedded in the noise, while the bottom part of the figure shows the peak when the signal is 'condensed' by the matched filter. Note that the peak occurs at the arrival time of the signal.

by the false alarm rate (events due to noise only) that one is prepared to tolerate, which in turn, is decided by the expected event rate.

However, it has been shown in [5] that one does not need templates in all the parameters, but only in those parameters which decide the shape of the wave-form or the dynamics of the system; in this case the masses and the spins. Such parameters have been called intrinsic, while the other kinematical parameters have been termed extrinsic [6]. Here I will discuss the spinless case and the wave-form given up to 2 PN order. I will later qualify the discussion with some remarks when spin is included in the analysis.

The amplitude can be easily handled by using normalised templates: one can scan over t_a by using the fast Fourier transform (FFT) which evaluates the integral in eq. (4) for all values of the time-lag τ, cheaply. The initial phase is maximised over by quadratures – one evaluates $c(\tau)$ for only two values of the inital phase 0 and $\pi/2$ and maximises the result by squaring and adding the two filtered outputs. One however, needs templates in the intrinsic parameters, in this case, the two masses m_1 and m_2.

A more convenient set of parameters than the masses are the chirp times:

$$\tau_0 = \frac{5}{256\pi\eta f_a}(\pi M f_a)^{-5/3}, \quad \tau_3 = \frac{1}{8\eta f_a}(\pi M f_a)^{-2/3}, \tag{6}$$

where f_a is a fiducial frequency (we use units in which $c = G = 1$) and $\eta = \mu/M$. The templates are arranged so that the maximum mismatch allowed is fixed at 3% corresponding to a maximum of 10% loss in SNR. These parameters are used because the template spacings are almost uniform in them. One can put a natural

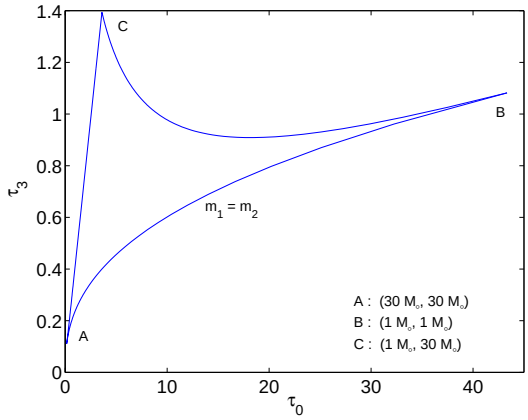

Figure 2. Parameter space in τ_0 and τ_3 (fiducial frequency $f_a = 40$ Hz).

metric on the parameter space which gives the distance between templates [6,7]. It turns out that this metric is almost independent of τ_0 and τ_3 and therefore these parameters can be looked upon as Cartesian coordinates on the manifold of signals. Shown in figure 2 is the parameter space in (τ_0, τ_3) for $f_a = 40$ Hz and masses in the range $1M_\odot \leq m_1, m_2 \leq 30M_\odot$.

With the maximum mismatch of 3% and taking the LIGO I (initial LIGO) PSD curve, the number of templates required is $\sim 10^4$. The wave-form is cut off at $\sim$1 kHz and the sampling rate is $\sim$2 kHz. To perform the search on-line, an on-line speed of 3.3 GFlops is required.

The threshold is set by the false alarm rate and the noise statistics. Assuming Gaussian noise, each of the correlations for initial phases 0 and $\pi/2$ is Gaussian distributed. The correlation c maximised over initial phase (summing the two correlations in quadrature and taking square root) is distributed as Rayleigh $R(c) = c\exp(-c^2/2)$ in the absence of the signal. The threshold ζ is determined by the condition $\int_\zeta^\infty R(c)\ \mathrm{d}c = P_\mathrm{F}$ where P_F is the false alarm probability. Assumimg a false alarm rate of 1/yr, $P_\mathrm{F} \sim 10^{-14}$ gives $\zeta \sim 8.2$. Detection is announced if $c > \zeta$. In order to have reasonably high detection probability the statistic c must significantly be above the threshold. A detection probability more than 95% implies $c > 8.9$ [8].

The on-line speed requirement goes up even more if the lower mass limit of the mass range is reduced and/or the bandwidth is increased as could be the case with advanced detectors. If the lower mass limit is reduced to $0.2M_\odot$ and the detector bandwidth lies in the range of 10 Hz to 1 kHz, then the on-line speed required can escalate to 300 GFlops. Clearly a more efficient search strategy is called for – the hierarchical search.

3.3 *Hierachical search*

Hierarchical search reduces the on-line speed requirement by performing the search efficiently with less computing cost. It also frees up CPU which allows for a search

over more parameters. Because one is mainly sifting through noise, an efficient strategy can be designed which reduces the effort involved in the search – this is the hierarchical search method. I describe below a two-step hierarchy:

1. First stage is the trigger stage: Use a lower threshold so that the template bank is coarse and fewer templates are required to scan the parameter space. However, the false alarm rate is high.
2. Second stage follows up the false alarms by a fine search around the triggers: Because such events are few, the fine search is just limited around these events and one does not have to search the whole parameter space with the fine bank.

The total cost of the search is then the sum of costs incurred in the two stages. The trigger stage threshold is then optimised for minimum total cost. In [8] the hierarchy was implemented over two parameters, namely, the masses. In [9] the idea of extending the hierarchy also over t_a was first given and then fully implemented in [10]. With initial LIGO PSD the time decimation gave an increased factor of 4 over [8]. The total gain factor in computational cost over the flat search, taking into account detailed boundary effects, etc. was about 60; that is, the on-line speed required would be about 50 MFlops as compared to 3 GFlops for the flat search.

3.4 *Searches for inspirals with a detector network and effects of spin*

In this subsection I briefly discuss searches with a network of laser interferometric detectors and also effects of spin.

Two types of searches can be envisaged: (a) the coincidence search, where individual thresholds are set up on the data from each detector and then event lists compared; (b) the coherent search, where the phase information is crucially used in constructing the statistic.

The coincidence approach involves separately matched filtering the signal in each detector, applying separate thresholds to each detector filtered output and then preparing event lists for each detector [11]. The event lists are matched by the criterion that the events lie in a certain neighbourhood – a 'window' in the parameter space. A detection is registered if the events lie within this window, that is they give results consistent with a true GW signal. However, it is a difficult task to determine the window because of the noise present in the data.

In the coherent search, on the other hand, a *single* statistic is constructed from the data of all the detectors and compared with a single threshold, so that the network is effectively a single synthesised detector. In [12] it was shown that just as in a single detector search, one only needs a template bank in the intrinsic parameters – the two masses. One also has to search over the directions, but this can be easily achieved by linearly combining the filtered outputs with appropriate time-delays. The computational costs soar up $\sim\times 10^3$ for the LIGO–VIRGO network [1,13] over the single detector search. For detectors with similar individual sensitivities, the network sensitivity increases as the square root of the number of detectors.

When one or both the stars are spinning rapidly, the spin plays an important role in the detection strategy, because ignoring this effect can result in not detecting the source at all. The spin couples to the orbital angular momentun (spin–orbit

coupling) which makes the orbital plane precess. The precession in turn modulates the wave-form. The problem is extremely complex and the wave-form involves a lot of parameters. Recently, detection template families (DTF) were proposed [14] which have few physical parameters and model adequately the modulated wave-form – the average maximum scalar product between the two families is better than 97%. In differential geometric language, the DTF manifold is 'close' to the manifold of actual wave-forms but has favourable properties amenable for data analysis. For single-spin binaries, that is, only one component star (black hole) with significant spin, the templates are in just three parameters: the two masses and the spin. It is shown that for LIGO I PSD and mass range $7M_\odot < m_1 < 12M_\odot$ and $1M_\odot < m_2 < 3M_\odot$, at an average mismatch of 97% one requires about 76,000 templates.

Clearly hierarchical methods must be used to cut computational costs. Efforts are in progress in this direction.

4. Periodic sources

Rotating neutron stars (NS) are one of the important sources of GW for the ground-based as well as space-based detectors. Since the waves are emitted continuously, the source is termed also as a continuous gravitational wave (CGW) source. The NS must be non-axisymmetric in order that it radiates gravitationally. Several mechanisms have been envisaged which can give rise to non-axisymmetry in NS. These have been detailed in [15]. Here we briefly mention the mechanisms. The measure of non-axisymmetry is denoted by ϵ and the characterstic amplitude h of the CGW source is given in eq. (2). Some of the mechanisms producing non-axisymmetry are: (i) large non-axial magnetic fields which produce asymmetry due to the magnetic pressure, (ii) the Chandrasekhar–Friedmann–Schutz (CFS) instability, (iii) accretion of hot material onto the neutron star surface – the induced quadrupole moment is directly related to the accretion rate which can be copius, etc.

CGW emitters pose one of the most computationally intensive problems in GW data analysis. In fact, the weakness of the expected signal requires very long observation times, of the order of a year (or possibly more) in order to accumulate enough signal-to-noise ratio (SNR) for ensuring detection. During this time a monochromatic signal, as measured in the source reference frame, is Doppler modulated by the motion of the detector carried by the spinning Earth orbiting the Sun. The emitted energy is spread over $\simeq 2\times 10^6\,(T/10^7\,\mathrm{s})^2\,(f/1\,\mathrm{kHz})$ frequency bins of width $\Delta f = 1/T$, where T is the observation time. In order to recover the whole power in one frequency bin, one has to 'correct' the recorded data stream for each possible source position in the sky; the Doppler shift in frequency is $\Delta f = (\mathbf{n}\cdot\mathbf{v})f_0/c$, where f_0 is the intrinsic frequency of the source, $\mathbf{n}$ the unit vector in the direction of the source and $\mathbf{v}$ the relative velocity of the source with respect to the detector. The problem is made worse, if the intrinsic frequency of the source changes, say due to spin-down. Then the power is spread over $3\times 10^6\,(\tau/10^3\,\mathrm{yr})^{-1}\,(T/10^7\,\mathrm{s})^2\,(f/1\,\mathrm{kHz})$ bins, where $\tau = f/\dot{f}$ is the spin-down age of the NS. Indeed, one then needs to correct also for this effect, searching through one or more of the spin-down

parameters. It is clear that searches for CGW are limited by the available computational resources [15,16].

Due to the large computational burden, algorithms investigated so far have been restricted to isolated NS, i.e., NS which are at rest or in uniform motion with respect to the barycentre. Targeted searches are possible with the currently available computer resources, where parameters of the source such as its position and/or frequency are known from other astronomical observations. Then one searches only within the allowed parameter window. In fact in [17] it has been shown that targeted searches for GW are possible for binary radio pulsars where the search is performed within the error bars.

Another type of search which has received a lot of attention recently is of an isolated NS but whose GW frequency and location in the sky are unknown. This is called the 'all sky all frequency' search [16]. Even here the computational costs are formidable [15]. Even ignoring the spin-down parameters, the search amounts to scanning over Doppler corrections corresponding to large number of directions in the sky called 'patches'. The number of patches $N_{\rm patch}$ and the number of operations $N_{\rm ops}$ associated with the search are given by

$$N_{\rm patch} \sim 10^{10} \left(\frac{f_0}{500\ {\rm Hz}}\right)^2 \left(\frac{T}{100\ {\rm days}}\right)^5,$$
$$N_{\rm ops} \sim N_{\rm patch} \times 3N \log_2 N \sim 10^{22}. \tag{7}$$

Here $N \sim 10^{10}$ is the number of points in the data train of 100 days sampled at the Nyquist frequency of a kHz.

These large costs result because the parameter space is huge. For an year's coherent integration, during which the detector moves along with the Earth, the effective size of the GW synthesised telescope is $D = 2$ AU or $\sim 3 \times 10^8$ km. The resolution of the telescope for a kHz wave ($\lambda = 300$ km) is $\lambda/D \sim 10^{-6}$ rad. Thus $N_{\rm patch} = 4\pi(D/\lambda)^2 \sim 10^{13}$.

Again hierarchical searches prove to be useful. I will mention two such searches [18], namely, (1) the Hough transform and (2) the stack and slide search.

Because the computational requirement is overwhelming, the philosophy here is to assume fixed available computing power at one's disposal and then gauge the performance of the method by its improvement in sensitivity over a flat search. Both the methods alternate between coherent and incoherent stages where short-term Fourier transforms are taken. In the case of the Hough transform, one looks for patterns in peaks in the time–frequency plane – patterns that must result from actual parameter values. One then constructs a histogram in the parameter space (transform from time–frequency plane to parameter space – the Hough transform) and look for a peak. Then one does a full time coherent search in the vicinity of the peak. In case of the stack and slide search, short term power spectra are computed and then summed by appropriately sliding them for changes in frequency due to Doppler effects or spin-down. Again candidate events are selected and a full coherent search is performed. This method gives a gain between 2 and 4 in sensitivity. A similar gain can be achieved with the Hough transform method.

5. Stochastic sources

Stochastic background of GW is produced by a large number of weak, independent, unresolved sources of GW. The radiation can be characterised only statistically. Such type of radiation could be produced in the early universe immediately after the Big Bang and could bring us information about the universe at this epoch. It could also result from unresolved binary stars in our galaxy or outside our galaxy. However, since the focus here is on data analysis, only these aspects will be discussed here.

We will assume that the stochastic background is isotropic, stationary, Gaussian and does not prefer any polarisation. Then the stochastic background can be characterised in terms of its spectrum $\Omega_{\rm GW}(f)$, where

$$\Omega_{\rm GW}(f) = \frac{1}{\rho_{\rm crit}} \frac{{\rm d}\rho_{\rm GW}}{{\rm d}\ln(f)}. \tag{8}$$

Here ${\rm d}\rho_{\rm GW}$ is the energy–density of GW contained in the logarithmic frequency interval ${\rm d}\ln(f)$ and $\rho_{\rm crit}$ is the energy–density required to close the universe, i.e., $\sim 10^{-29}$ g/cc.

In general the data from each detector $x_i(t),\ i = 1, 2$ will be given by

$$x_1(t) = h_1(t) + n_1(t), \quad x_2(t) = h_2(t) + n_2(t), \tag{9}$$

where $h_i(t), n_i(t),\ i = 1, 2$ are the signals and noises respectively in detectors 1 and 2. We will consider here two detectors only. The signal statistic is defined as the correlation [19]

$$C = \int_0^T {\rm d}t \int_0^T {\rm d}t' x_1(t) x_2(t') Q(t, t'), \tag{10}$$

where $Q(t, t') = Q(t - t')$ is a filter function which optimises the SNR and T is the observation time. The simplest case occurs when the detectors are coincident and coaligned in which case the filter function is unity. The mean value of C is $\mu = \langle C \rangle$. The variance of C is given by

$$\sigma^2 = \langle C^2 \rangle - \langle C \rangle^2 \sim \frac{T}{4} \int_{-\infty}^{\infty} {\rm d}f P_1(|f|) P_2(|f|) |\tilde{Q}(f)|^2, \tag{11}$$

where $P_i(|f|),\ i = 1, 2$ are the one-sided PSDs of the noise in the two detectors and $\tilde{Q}$ is the Fourier transform of the filter function. It is assumed that the noises in the detectors are stationary, Gaussian and statistically independent of each other. The SNR is then just given by the ratio μ/σ. The optimal SNR, by choosing the filter function Q optimally, is given by

$$\text{SNR} \sim \frac{3H_0^2}{10\pi^2} \sqrt{T} \left[\int_{-\infty}^{\infty} {\rm d}f \frac{\gamma(|f|)^2 \Omega_{\rm GW}^2(|f|)}{f^6 P_1(|f|) P_2(|f|)} \right]^{1/2}, \tag{12}$$

where H_0 is the present Hubble constant and $\gamma(f)$ the so-called overlap reduction function. The overlap reduction function arises because in general the detectors

are neither coincident nor co-aligned; it incorporates the reduction in the SNR because the time delay between non-coincident detectors and the orientations of the detectors is not identical. The two effects mean that the signal in both detectors is not the same and so the overlap expressed in eq. (10) is only partial. The $\gamma(f)$ depends on the distance between the detectors (in units of wavelength of the GW) and the antenna pattern functions of the two detectors. If the detectors are more or less co-aligned then $\gamma(f)$ is close to unity in the limit of low frequency. For coincident, co-aligned detectors $\gamma(f)$ is unity. For the LIGO detector pair, the arms are rotated by almost 90° and thus $\gamma(f)$ is negative, i.e., ~ -0.89.

The statistical analysis now proceeds almost exactly on the lines of inspiraling binaries discussed earlier, except that here the two hypotheses are (i) signal absent (ii) signal present but with unknown value of $\mu > 0$. The following results are obtained for the 4 km, LIGO pair of interferometers. If $\Omega_{\rm GW}(f)$ is considered to be a constant in the model, say, Ω_0 then one just turns eq. (12) around to obtain a limit on the detectable Ω_0. For an observation time of 4 months, at a false alarm rate of 5% and detection probability of 95%, $\Omega_0 \sim 10^{-5}$–10^{-6} for the initial LIGO detector pair while $\Omega_0 \sim 10^{-10}$–10^{-11} for the advanced LIGO detector pair.

6. Unmodelled sources

Examples of unmodelled sources are supernovae, hypernovae, binary mergers, ring-downs of binary black holes etc. The importance of these sources is that a large amount of GW is expected to be emitted by them and should be among the brightest of those that we can observe. However, their physics is so complex (supernovae, hypernovae) or it is non-linear (binary mergers) that it is extremely difficult to obtain wave-forms. Therefore, the technique of matched filtering cannot be succesfully employed here and alternative signal detection methods must be sought.

In case of binary mergers it is possible to make crude estimates of the duration and the frequency bands of the GW signal. In such cases time–frequency methods are best suited for performing a blind search. I discuss the excess power method [20], where one computes the total power within the expected time–frequency window and searches over all start times. Detection is announced if there is more power than one expects from noise alone. The excess power statistics $\mathcal{E}$ is defined as

$$\mathcal{E} = 4 \sum_{k_1 \le k \le k_2} \frac{|\tilde{h}_k|^2}{P_k}, \tag{13}$$

where $[k_1, k_2]$ is the frequency band, $\tilde{h}_k$ is the kth Fourier component of the signal and P_k is the corresponding component of the one-sided PSD of the detector noise. If the relevant time–frequency window is characterised by the time interval Δt and the frequency band Δf then the volume of the window in the time–frequency plane is $V = \Delta t \Delta f$. V is the number of frequency components $k_2 - k_1$. If the noise is stationary, Gaussian then $\mathcal{E}$ follows a χ^2 distribution with $2V$ degrees of freedom. If a signal is present, then the distribution will be non-central χ^2 with the non-centrality parameter being the signal power. Usually V is large and because of the central limit theorem the distributions are nearly Gaussian and it is easy to do the

statistics. The signal is detectable if $\mathcal{E} - 2V > \zeta$ where $\zeta = (\text{few}) \times \sqrt{4V}$ is the pre-assigned threshold.

However, this method cannot distinguish non-Gaussian bursts of noise from the signal. In fact turning the argument around it can be used to analyse non-Gaussian components of the noise. In order to distinguish non-Gaussianity from a genuine GW signal, one must use a network of detectors. In this context, a coherent burst detection statistics has been proposed in [21], where the data from different detectors is linearly combined with appropriate time delays with scale factors involving antenna pattern functions. The main fact used is that the signal must lie in the plane spanned by the vectors $h_+(t)$ and $h_\times(t)$ called the 'polarisation' plane. Only two search parameters are needed in addition to the source direction which are chosen to be the ratio of lengths of the vectors $h_{+,\times}(t)$ and the angle between them. This strategy dramatically narrows down the search.

An important consideration that is missing from all these methods is that no source model based on general physics has been considered. I believe that investigations in these directions need to be carried out in order to further narrow down the search for unmodelled sources.

7. Dealing with real data

Real data are now available from the LIGO, TAMA and GEO detectors and although the requisite sensitivity has not been reached, it gives us the opportunity to test the performance of the algorithms and the codes. Codes have already been developed based on the algorithms mentioned above. It is good to learn that the codes are working and yielding sensible results.

Real detector noise is neither stationary nor Gaussian. However, the algorithms have been developed which assume the stationarity and the Gaussianity of noise. One therefore needs to adapt the algorithms to the real world. This is achieved by applying vetos to distinguish the non-Gaussian noise from the signal. Several vetos have been designed such as excess noise level veto and instrumental vetos. The excess noise level veto discards the data segment if the noise in that segment exceeds a certain threshold. The instrumental vetos look for correlations between other channels (e.g. environmental) with the GW channel. However, for inspirals an ingenious veto has been proposed by Allen *et al* [22] based on time–frequency analysis. The veto uses the fact that the power in the inspiral signal approximately falls off as $f^{-7/3}$ irrespective of the parameters. One divides the frequency domain into p sub-bands so that the signal has equal power in each sub-band and then calculates $\chi^2 = \sum_{k=1}^{p} (\rho_k - \rho/p)^2$, where ρ is the full SNR and ρ_k is the SNR in the sub-band k. χ^2 is small if the noise is Gaussian with or without the chirp signal; on the other hand, a large value of χ^2 is taken to indicate non-Gaussianity. One then uses an appropriate threshold to decide detection. Other vetos can also be designed for this purpose which use the phase information directly – a veto could be constructed to follow the ambiguity function. However, the catch is that the computation of the veto should not need too much computational effort, otherwise the efficacy of the data analysis method would be compromised.

Although at this early stage of the experiment, no detection can be announced so far, we can place upper limits on the event rates. I will not mention here the actual numbers because more appropriate numbers from more recent data have been quoted elsewhere in this conference. The point however, is that with the recent data, upper limits are becoming astrophysically interesting.

8. Laser interferometric space antenna (LISA)

Ground-based detectors will operate in the high frequency range of GW of $\sim$10 Hz to a few kHz. A natural limit occurs on decreasing the lower frequency cut-off of 10 Hz because it is not practical to increase the arm-lengths on ground and also because of the gravity gradient noise which is difficult to eliminate below 10 Hz. Thus, the ground-based interferometers will not be sensitive below the limiting frequency of 10 Hz. But on the other hand, there exist in the cosmos, interesting astrophysical GW sources which emit GW below this frequency such as the galactic binaries, massive and super-massive black hole binaries etc. If we wish to observe these sources, we needto go to lower frequencies. The solution is to build an interferometer in space, where such noises will be absent and allow the detection of GW in the low frequency regime. LISA – laser interferometric space antenna – is a proposed ESA–NASA mission which will use coherent laser beams exchanged between three identical spacecrafts forming a giant (almost) equilateral triangle of side 5×10^6 km to observe and detect low frequency cosmic GW [23]. The groundbaseddetectors and LISA complement each other in the observation of GW in an essential way, analogous to the optical, radio, X-ray, γ-ray etc., observations for the electromagnetic waves.

An important data analysis issue for LISA is the cancellation of laser frequency noise in the data. The noise in the data is due to (i) frequency fluctuations of the lasers used in transmission and reception – laser frequency noise, (ii) fluctuations due to non-inertial motions of the spacecraft, (iii) beam-pointing fluctuations and shot noise. Cancellation of laser frequency noise in interferometers is crucial for attaining the requisite sensitivity. Raw laser noise is several orders of magnitude above the other noises and thus it is essential to bring it down to the level of other noises such as shot, acceleration, etc. Since it is impossible to maintain equal distances between spacecrafts, laser noise cancellation must be achieved by appropriately combining the six beams with appropriate time delays. It has been shown in several recent papers that such combinations are possible [24]. The actual procedure can easily be understood in terms of properly defined time-delay operators that act on the one-way Doppler measurements. A rigorous formalism has been given involving the algebra of the time-delay operators, based on the theory of rings and modules and computational commutative algebra. The space of all possible interferometric combinations cancelling the laser frequency noise is a module over the polynomial ring in which the time-delay operators play the role of the indeterminates. The module, in the literature, is called the Module of Syzygies [25]. The module is generated from four generators, so that any data combination cancelling the laser frequency noise is simply a linear combination of these generators. This is the mathematical structure underlying time-delay interferometry in LISA.

However, the above analysis holds for a stationary LISA. Efforts are now in progress to take into account a moving LISA with changing arm-lengths.

9. Summary

Data analysis is an important aspect in the gravitational wave observation experiment. Different types of sources need different data analysis methods and strategies. Because the signal is weak, sophisticated data analysis is required for designing computationally efficient algorithms. Codes have been developed based on the algorithms and are now being tested and adapted to real data. Finally the problem of cancellation of laser frequency noise in LISA has been discussed which is important if LISA is to attain its goal sensitivity. In the design of the algorithms as well as in the time-delay interferometry problem for LISA, the future directions have been indicated.

References

[1] A Abramovici, W E Althouse, R W P Dever, Y Gursel, S Kanwamura, F J Raab, D Schoemaker, L Sievers, R E Spero, K S Thorne, R E Vogt, R Weiss, S E Whitcomb and Z E Zucker, *Science* **256**, 325 (1992)
[2] http://www.ligo.caltech.edu/∼lazz/distribution/LSC-Data
[3] L Blanchet, T Damour, B R Iyer, C M Will and A G Wiseman, *Phys. Rev. Lett.* **74**, 3515 (1995)
L Blanchet, T Damour and B R Iyer, *Phys. Rev.* **D51**, 5360 (1995)
T Damour, B R Iyer and B S Sathyaprakash, *Phys. Rev.* **D63**, 044023 (2001); **D66**, 027502 (2002)
[4] T Damour, B R Iyer and B S Sathyaprakash, *Phys. Rev.* **D57**, 885 (1998)
A Buonnano and T Damour, *Phys. Rev.* **D59**, 084006 (1999); **D62**, 064015 (2000)
T Damour, B R Iyer, P Jaranowski and B S Sathyaprakash, *Phys. Rev.* **D67**, 064028 (2003)
[5] B S Sathyaprakash and S V Dhurandhar, *Phys. Rev.* **D44**, 3819 (1991)
S V Dhurandhar and B S Sathyaprakash, *Phys. Rev.* **D49**, 1707 (1994)
[6] B Owen, *Phys. Rev.* **D53**, 6749 (1996)
[7] R Balasubramanian, B S Sathyaprakash and S V Dhurandhar, *Phys. Rev.* **D53**, 3033 (1996)
[8] S Mohanty and S V Dhurandhar, *Phys. Rev.* **D54**, 7108 (1996)
[9] T Tanaka and H Tagoshi, *Phys. Rev.* **D62**, 082001 (2001)
[10] A Sengupta, S V Dhurandhar and A Lazzarini, *Phys. Rev.* **D67**, 082004 (2003)
[11] A Pai, S Bose and S V Dhurandhar, *Class. Quantum Gravit.* **19**, 1477 (2002)
[12] A Pai, S V Dhurandhar and S Bose, *Phys. Rev.* **D64**, 042004 (2001)
S Finn, *Phys. Rev.* **D63**, 102001 (2001)
[13] C Bradaschia *et al*, *Nucl. Instrum. Methods* **A289**, 518 (1990)
[14] A Buonanno, Y Chen and M Vallisneri, *Phys. Rev.* **D67**, 104025 (2003)
Y Pan, A Buonanno, Y Chen and M Vallisneri, gr-qc/0310034
[15] P R Brady, T Creighton, C Cutler and B F Schutz, *Phys. Rev.* **D57**, 2101 (1998)
[16] B F Schutz, in *The detection of gravitational waves* edited by D Blair (Cambridge University Press, Cambridge, England, 1989) pp. 406–427

[17] S V Dhurandhar and A Vecchio, *Phys. Rev.* **D63**, 122001 (2001)
[18] M A Papa, B F Schutz and A M Sintes in *Gravitational waves: A challenge to theoretical astrophysics*, ICTP Lecture Notes Series (2000)
P Brady and T Creighton, *Phys. Rev.* **D61**, 082001 (2000)
[19] B Allen and J Romano, *Phys. Rev.* **D59**, 102001 (1999)
[20] W Anderson, P Brady, J Creighton and E Flanagan, *Phys. Rev.* **D63**, 042003 (2001)
[21] J Sylvestre, *Phys. Rev.* **D68**, 102005 (2003)
[22] B Allen *et al*, *Phys. Rev. Lett.* **83**, 1498 (1999)
[23] P Bender *et al*, *System and technology study report ESA-SCI (2000)*, 11 (2000)
[24] M Tinto, F B Estabrook and J W Armstrong, *Phys. Rev.* **D65**, 082003 (2002)
[25] S V Dhurandhar, K R Nayak and J-Y Vinet, *Phys. Rev.* **D65**, 102002 (2002)