Gravitational wave astronomy: needle in a haystack

Neil J. Cornish

Department of Physics, Montana State University, Bozeman, MT 59717, USA

A worldwide array of highly sensitive ground-based interferometers stands poised to usher in a new era in astronomy with the first direct detection of gravitational waves. The data from these instruments will provide a unique perspective on extreme astrophysical objects, such as neutron stars and black holes, and will allow us to test Einstein’s theory of gravity in the strong field, dynamical regime. To fully realize these goals, we need to solve some challenging problems in signal processing and inference, such as finding rare and weak signals that are buried in non-stationary and non-Gaussian instrument noise, dealing with high-dimensional model spaces, and locating what are often extremely tight concentrations of posterior mass within the prior volume. Gravitational wave detection using space-based detectors and pulsar timing arrays bring with them the additional challenge of having to isolate individual signals that overlap one another in both time and frequency. Promising solutions to these problems will be discussed, along with some of the challenges that remain.

1. Introduction

According to Einstein’s general theory of relativity, the merger of two black holes should emit more power in gravitational waves than the light from all the stars in all the galaxies in the Universe. But, despite the technological advances in the century since Einstein described his theory, we are yet to observe the most energetic phenomena in the Universe. If we are lucky, the first direct detection of gravitational waves may occur by the centennial anniversary of Einstein’s 1916 paper on general relativity [1], or, failing that, by the anniversary of his 1918 paper on gravitational waves, in which he corrected several errors in the original description of the phenomena [2]. The reason for this optimism is that
the kilometre-scale Laser Interferometer Gravitational Observatory (LIGO) instruments in the USA [3] and the Virgo instrument in Italy [4] will soon be back on line after undergoing major upgrades that are designed to increase their sensitivity by a factor of 10, and the volume of space explored by a factor of 1000 [5,6]. Even with fairly conservative estimates for the astrophysical event rates [7], we should be seeing multiple detections each year once the detectors reach design sensitivity.

The great French mathematician Pierre-Simon Laplace anticipated several of the key elements in gravitational wave astronomy, most notably the gravitational collapse of stars to form black holes [8], and the effects on a binary orbit owing to a finite propagation speed for the gravitational force—though the specific form for the force law that he considered violates Lorentz invariance and leads to the prediction that binary systems will absorb rather than emit gravitational waves [9]. But Laplace would probably be puzzled (and perhaps a little dismayed) by the way the data are currently being analysed, where talk abounds of ‘detection statistics’ and ‘false alarm and false dismissal probabilities’. There are several reasons why this frequentist approach to signal analysis has held sway in gravitational wave astronomy (for a review of this approach, see Jaranowski & Królak [10]). Some of the reasons can be traced to the historical development of the field, and some are related to the computational challenges of implementing the Bayes–Laplace approach to inference. However, I will argue that the main reason that the frequentist approach continues to prevail can be traced to the difficulty in specifying a suitable form for the likelihood, combined with an inflexible application of the Bayesian approach. Just as Monte Carlo studies are used to tune and calibrate the frequentist detection statistics, we need to take a more experimental approach to defining the likelihood. Over the past few years, there has been a significant increase in the number of papers applying Bayesian inference to gravitational wave astronomy (e.g. [11,12] and references therein), and it is possible to sense a sea change in the community. In my description of gravitational wave signal analysis, I will follow a Bayesian approach, with occasional diversions to connect with the traditional description in terms of matched filtering, $\chi^2$ statistics and signal-to-noise (SNR) ratios.

I will begin in §2 with a brief description of gravitational waves and methods for their detection, and outline some of the questions we hope to answer with these observations. This is followed in §3 by an overview of gravitational wave signal analysis, and in §4 I discuss some of the challenges in defining a suitable likelihood function. Models for the gravitational wave signals are described in §5. Detection strategies and techniques for mapping out the posterior distributions are covered in §6. The problem of defining a likelihood function is revisited in §7, and a number of promising new approaches are discussed. The challenge of disentangling the signals from multiple sources is reviewed in §8. I close with a summary in §9.

In describing gravitational wave signal analysis, I have endeavoured to use a minimum of jargon in the hope that the approaches we are using may prove useful elsewhere, and in the hope that practitioners from other fields can suggest useful techniques that we may not be aware of. When suitably abstracted, all signal analysis problems share a common core, and while the parts of a particular analysis that prove the most challenging will differ from problem to problem, the same sticking points have a habit of showing up in many different fields of study.

2. Gravitational wave astronomy

The key properties of gravitational waves can be derived from the linearized Einstein equations. In this description, the space–time metric is decomposed into a slowly varying background $g_{\mu\nu}$ and a small, rapidly varying perturbation $h_{\mu\nu} \ll 1$ [13]. In what follows I will suppress the tensorial indices on the dimensionless gravitational perturbation $h$. The linearized Einstein equations tell us that gravitational waves travel at the speed of light and come in two transverse polarization states. The physical nature of the waves can be seen by computing the Riemann curvature tensor from the perturbed metric, which reveals that gravitational waves manifest as a time-varying, quadrupolar tidal field that alternately squeezes then stretches space along
orthogonal directions that lie in the plane perpendicular to the direction of propagation. The fractional change $\Delta L/L$ in the distance between two free masses is proportional to the gravitational wave strain $h$ up to geometrical factors of order unity (at least in the limit where $L$ is small compared with the gravitational wavelength). The two polarization states produce distortion patterns that are rotated by $45^\circ$ with respect to each other, and are referred to as the ‘plus’ (+) and ‘cross’ (×) polarizations. Conservation of mass and linear momentum forbids the emission of monopole or dipole radiation, so, to leading order, it is the time-varying quadrupole moment of a mass distribution that is responsible for gravitational wave generation [2].

Simple back-of-the-envelope calculations can be used to arrive at surprisingly accurate estimates for the strength, frequency, luminosity and duration of gravitational wave signals for self-gravitating systems [14]. The quantities that enter these expressions are the internal and external gravitational field strengths and the light crossing time of the source. The dimensionless gravitational potential is given by

$$\Phi = \frac{GM}{c^2 r}, \quad (2.1)$$

where $M$ is the mass-energy involved in generating the gravitational waves and $r$ is either the distance to the source $D$ (for the external potential) or the size of the source $R$ (for the internal potential). The light crossing time is simply $T = R/c$. The waves have amplitude

$$h \sim \left( \frac{GM}{c^2 D} \right) \left( \frac{GM}{c^2 R} \right), \quad (2.2)$$

frequency

$$f \sim \frac{1}{T} \left( \frac{GM}{c^2 R} \right)^{1/2}, \quad (2.3)$$

evolution time scale

$$\frac{f}{f_0} \sim T \left( \frac{GM}{c^2 R} \right)^{-3} \quad (2.4)$$

and luminosity

$$\mathcal{L} \sim \frac{c^5}{G} \left( \frac{GM}{c^2 R} \right)^5. \quad (2.5)$$

The quantity $\mathcal{L}_* = c^5/G = 3.6 \times 10^{59}$ erg s$^{-1}$ sets the maximum luminosity of a gravitational wave source in general relativity as the potential $GM/c^2 R$ saturates at around unity for a black hole. For binary systems, equation (2.3) is just Kepler’s Third Law in disguise. Plugging in some numbers for a source that may occur with a reasonable event rate, such as the merger of a binary system of stellar remnant black holes at the distance of the Coma cluster, we find that the amplitude of the waves reaching the Earth is a paltry $h \sim 10^{-22}$ despite an energy release of over $10^{54}$ erg in the final second of the merger.

The small amplitude of the waves can be attributed to the extreme stiffness of space–time. Drawing an analogy between Einstein’s equations and Hooke’s Law for elastic solids implies that space–time has a very large ‘metrical elasticity’ $c^4/8\pi G$ [15], and a correspondingly large specific impedance $c^2/G = 4.5 \times 10^{38}$ g s$^{-1}$. This is vastly larger than the specific impedance of materials such as aluminium (approx. $10^{10}$ g s$^{-1}$) which are used to make resonant bar detectors. The difficulty in detecting gravitational waves with bar detectors can be seen as an impedance matching problem. In contrast, the dense nuclear material that makes up a neutron star has a specific impedance of around $10^{30}$ g s$^{-1}$, making the vibration modes of these objects promising sources of gravitational waves.

While the weak coupling of gravitational waves to matter poses a challenge for detection, it also makes them the ultimate form of ‘X-ray vision’ for seeing deep inside the most extreme environments in the Universe. When the iron core of a massive star collapses in the first stage
of a supernova explosion, the density is so great that even neutrinos are trapped and the only signals that are able to escape and provide information about the dynamics are gravitational waves. Gravitational waves are also the only signals that can reach us from the first fraction of a second after the Big Bang—electromagnetic signals cannot penetrate beyond the recombination era some 400 000 years after the Big Bang.

There are two basic approaches to detecting gravitational waves. The first to be developed were acoustic transducers that convert the strain exerted by the time-varying tidal gravitational field into the vibrations of a mechanical system. Modern versions of these detectors achieve impressive sensitivities, but in a limited frequency band set by the natural vibrational frequency of the detector. The second approach looks for changes in the light travel time between two free or partially constrained masses. One novel implementation of this approach uses the radio pulses from rapidly rotating neutron stars as a very stable time reference, and looks for small variations in the arrival times as a signature of gravitational waves imparting a time delay or advance. It has been predicted that binaries composed of supermassive black holes with masses $10^7$–$10^9$ times larger than our Sun, orbiting with periods of the order of years, would impart periodic variations in the pulse arrival time of order a few nanoseconds. Using a large collection of millisecond pulsars to form a pulsar timing array, this level of variation should be detectable within the next decade [16]. The same idea can be implemented on a smaller scale using laser beams as a time reference. The kilometre-scale ground-based LIGO and Virgo detectors operate in this way, as will the much larger space-based interferometers that are currently being planned [17,18]. The LIGO and Virgo detectors demonstrated incredible levels of sensitivity in their initial science runs, operating at strain sensitivities of $h \sim 10^{-21}$ in the frequency band 100–400 Hz over coherence times of approximately 0.1 s. This equates to tracking the centre of mass motion of the mirrors to less than one-thousandth of the width of a proton. In contrast to acoustic detectors, these light-travel time detectors have broad band sensitivity that spans several decades in frequency. As shown graphically in figure 1, these detectors span the spectrum between $10^{-9}$ and $10^5$ Hz: pulsar timing arrays cover the band between $10^{-9}$ and $10^{-6}$ Hz; million kilometre-scale space-based detectors, such as the Laser Interferometer Space Antenna, cover the band between $10^{-5}$ and $10^{-1}$ Hz; and the advanced LIGO and Virgo detectors cover the band between 10 and $10^5$ Hz.

Regardless of the detection method or the frequency range, the detector output is a time series that encodes the time-varying gravitational wave strain $h(t)$.

The goals of gravitational wave astronomy extend far beyond making the first direct detection, and we are already seeing important astrophysical results coming from the bounds that can be derived from non-detections [19–21]. There are currently large uncertainties about the event rates for black hole and neutron star mergers [7], and establishing these rates will tell us a great deal about the poorly understood processes by which these systems form. If we are lucky enough to record a nearby core-collapse supernova explosion the insight into the internal dynamics of the
explosion mechanism will be a boon for efforts to numerically simulate these events. The list of applications goes on and on, but, if history is any guide, many of the most important results probably have not been anticipated.

In addition to providing a powerful new tool for astronomers, gravitational wave detectors can also be used to perform precision tests of Einstein’s theory of gravity. At present, there are many basic predictions that await testing, such as gravitational waves travelling at the speed of light and coming in two transverse polarization states. In alternative theories of gravity, the waves can travel at different speeds, and there are as many as six polarization states, some of which are longitudinal. Measuring the full range of possible polarizations requires time-of-flight measurements along a minimum of six independent paths [22], which is not currently possible with the existing ground-based interferometers, but can be done using pulsar timing arrays [23,24]. Existing non-gravitational wave tests probe the static, weak field regime with gravitational potentials in the range $\Phi \sim 10^{-8}–10^{-6}$ and orbital velocities in the range $v/c \sim 10^{-4}–10^{-3}$. It is easy to construct candidate theories that are consistent with all existing data, and yet differ significantly from general relativity in the final stages of a black hole merger where $\Phi$ and $v/c$ approach unity [25].

3. Gravitational wave signal analysis

The read-out from a gravitational wave detector is a time series of some quantity that can be related to the gravitational wave strain. Common examples of the quantity being measured include phase differences, frequency shifts, time delays and displacements. For LIGO, the response is derived from an error point in the differential arm length servo control loop. In any case, the output can be converted into a time series for the dimensionless strain $s(t)$ registered by the detector, which includes contributions from a gravitational wave signal $h(t)$ and instrument noise $n(t)$,

$$s(t) = R(t, \tau) \star h(\tau) + n(t).$$  \hspace{1cm} (3.1)

The detector response function $R(t, \tau)$ depends on the polarization angle and sky location, and includes contributions from the different times $\tau$ that the wavefronts encounter the reference masses in the detector. These delays are important for space-based detectors and pulsar timing arrays, but are generally negligible for individual ground-based detectors as the light crossing time for the detector is much less than the period of the waves they seek to detect. The delays are important for a worldwide network of gravitational wave detectors, where we use a standard reference for $t$, such as Greenwich Mean Time, and the times $\tau$ record when the waves arrive at each detector. With three or more detectors in the array, it is possible to triangulate where the signals came from. Additional directional information is also encoded in the antenna patterns.

Very often we have multiple data streams $s_i(t)$, which can be stacked together to form a single data vector,

$$s = R \cdot h + n.$$  \hspace{1cm} (3.2)

The individual elements of the data vector are then the time samples in a particular detector. Alternatively, we may choose to transform to some alternative representation where the elements might be Fourier or wavelet amplitudes. The response matrix $R$ is a known function that can be computed for any polarization and sky location. Our task is to infer the properties of the gravitational wave signal $h$ from the data $s$, and to assess our degree of confidence in the detection by comparing the evidence of models that include gravitational wave signals with models that only describe the instrument noise. In a Bayesian approach to this inference problem, we need to fully specify models for the gravitational wave signals and the instrument noise. We also need to provide priors on the quantities that enter these models, and define a likelihood function. Once this is done, we can attack the purely mechanical task of computing the posterior distribution for our model parameters and the model evidence using techniques such as Markov chain Monte Carlo (MCMC) and nested sampling algorithms. A second level of analysis can then follow that uses the information gleaned about individual systems to constrain astrophysical models for the
sources (e.g. the mechanisms at play in a core-collapse supernova explosion), and models for the source populations (e.g. which stellar evolution pathways lead to the formation of neutron star or black hole binaries). Studies of this second level of analysis are just getting underway [26,27].

4. Likelihood function and models for the instrument noise

There are two basic approaches to the problem of extracting information about the gravitational wave signals. The first of these works directly with the raw data, while the second deals with quantities that can be derived from the raw data. In either case, it is necessary to define a likelihood function, and this turns out to be a surprisingly challenging task. In a recent article, Skilling writes ‘The instrument acquiring the data can usually be calibrated with known inputs to find out how often it produces specific outputs, which effectively fixes the likelihood to any desired precision. If there remain any unknown calibration parameters in the likelihood, they can be incorporated as extra parameters to be determined, leading to extra computation but no difficulty in principle’ [28, p. 11]. While fine in theory, the calibration of the likelihood turns out to be extremely challenging in practice for gravitational wave detectors, such as LIGO and Virgo. There are literally dozens of people devoted to characterizing and calibrating these instruments. The difficulty is that the signals we are looking for are rare and weak, so we need to be able to map the likelihood far into the tail of the noise distribution. Compounding the problem is the fact that the behaviour of the detectors changes over time owing to environmental changes and modifications that are made during the weekly maintenance activities.

To understand the challenges involved in defining the likelihood, consider a direct approach that works with the raw data. Here we need a model \( M \) for the gravitational wave signals \( h \) that may be present in the data. We can then subtract this model from the data to form up the residual \( r = s - R \cdot h \). If our model matches the gravitational wave signal contained in the data then the residual should be consistent with pure instrument noise \( n \). Thus, the likelihood function \( p(s|h,M) \) is nothing other than the joint noise distribution \( p_n(n) \),

\[
p(s|h,M) = p_n(r). \tag{4.1}
\]

In other words, the likelihood is the noise model. A typical stretch of data from one of the LIGO/Virgo instruments can be approximately modelled as stationary, coloured and normally distributed noise defined by the expectation values

\[
E[n(f)] = 0,
\]

\[
E[n(f)n^*(f')] = \frac{T}{2} \delta_{ff'}S_n(f), \tag{4.2}
\]

where \( S_n(f) \) is the one-sided noise spectral density at frequency \( f \) and \( T \) is the observation time. The joint noise distribution in this case factors into a product, and the likelihood becomes

\[
p(s|h,M) = \prod_f \frac{1}{2\pi TS_n(f)} \exp \left( -\frac{r(f)r^*(f)}{TS_n(f)} \right)
\]

\[
= C e^{-\langle r^2 \rangle / 2} = C e^{-\chi^2/2}, \tag{4.3}
\]

where in the last line the product of the exponentials has been converted to a sum of exponents to define the noise-weighted inner product

\[
\langle a | b \rangle = \frac{2}{T} \sum_f \frac{a(f)b^*(f) + a^*(f)b(f)}{S_n(f)}. \tag{4.4}
\]

This is nothing other than the standard \( \chi^2 \) goodness of fit first introduced by Gauss. The likelihood for a network of detectors with independent noise realizations is simply the product of the individual likelihoods. If correlations are present they can be dealt with by introducing a noise correlation matrix.
In the frequentist approach, the Neyman–Pearson criterion is used to identify detection candidates by minimizing the false dismissal probability for a given false alarm probability. In stationary Gaussian noise the optimal decision statistic for the Neyman–Pearson test is the likelihood ratio \[ \Lambda = \frac{p(s|h,M)}{p(s|0,M)} = \exp \left( -(s|\mathbf{R} \cdot \mathbf{h}) + \frac{1}{2} (\mathbf{R} \cdot \mathbf{h}|\mathbf{R} \cdot \mathbf{h}) \right). \] (4.5)

The space of possible observations is partitioned by setting a threshold \( \Lambda_0 \) that yields the desired false alarm probability. The threshold can be computed analytically for the simple noise model considered here. Maximizing \( \ln \Lambda \) with respect to the amplitude of the signal \( A = (\mathbf{R} \cdot \mathbf{h}|\mathbf{R} \cdot \mathbf{h})^{1/2} \) yields the detection statistic \[ \rho = \frac{(s|\mathbf{R} \cdot \mathbf{h})}{(\mathbf{R} \cdot \mathbf{h}|\mathbf{R} \cdot \mathbf{h})^{1/2}}. \] (4.6)

The maximum value of \( \rho \) is referred to as the amplitude SNR ratio of the best fit signal \( \mathbf{h} \). The SNR can be computed on a per detector basis, and for the full detector network. The quantity \( \rho \) is referred to as the Wiener matched filter statistic, and it can be shown to be the optimal statistic in stationary Gaussian noise for signals with known amplitude. It is the reason why almost every book, talk and paper on gravitational wave astronomy makes frequent references to ‘matched filtering’, and cites the power of this technique in lifting signals that are buried beneath the instrument noise. Unfortunately, none of the assumptions used to motivate this approach pertain in practice, and the real-world analysis is far more complex.

**Figure 2.** A scalogram showing examples of non-stationary features in data from one of the LIGO detectors (adapted from [29]). (Online version in colour.)

---

**Figure 2** shows a time–frequency map of a small segment of data from the LIGO detector in Hanford, WA, where a loud transient feature or ‘glitch’ stands out above the surrounding noise. Studies have shown that these instrument glitches follow a power-law distribution in amplitude, with a greater prevalence of low-amplitude glitches. In relatively clean stretches of data with no obvious glitches, the noise looks Gaussian, but sections with glitches develop heavier tails. It is clear, however, that glitches do more than fatten the tails of the distribution, they also introduce correlations across the time–frequency plane. Ideally, we would like to follow Skilling’s advice and fully characterize this more complicated noise model to arrive at a realistic description of the likelihood. Short of this ideal, I will describe some of the methods that are being used to account for the non-Gaussianity and non-stationarity of the data.
In the traditional approach, statistics like $\rho$ and $\chi^2$ are adopted, and rather than using theoretical distributions for these quantities, Monte Carlo techniques are used to estimate the probability distributions in pure instrument noise and when signals are present. Since we do not know in advance whether a given segment of data contains a gravitational wave, unphysical time shifts are introduced between different instruments. This ensures that any signals that might be present will fail the coincidence test that is used to decide whether the signals seen in each detector are consistent within the light travel times between the detectors. The behaviour when signals are present is determined by adding simulated signals to the data. The analysis also makes use of the thousands of monitors attached to the detectors, such as seismometers, microphones and ammeters. Sections of data where correlations are detected between the gravitational wave channel and the environmental channels are vetoed and discarded.

In recent years, the traditional approach has been extended to include a large number of quantities that can be measured from the raw data. These include the parameters $\lambda$ that define a particular gravitational wave model $h(\lambda)$ that yields the highest SNR in each detector, as well as the individual SNRs and the standard $\rho$, $\Lambda$ and $\chi^2$ statistics. The distributions of these parameters for the signal-plus-noise model and noise-only model are then estimated using the time slide and signal injection procedure described above. The samples are then fed into a multi-dimensional classifier to produce a decision tree that assigns a probability that an event is consistent with noise [30], or are used to develop joint likelihoods that feed into a Bayesian model selection analysis [31]. Both of these approaches are better able to account for the non-Gaussianity of the instrument noise than the traditional approach.

More recently, a new approach for working with the raw data has been put forward that includes a sophisticated noise model that explicitly accounts for glitches in the data by introducing a glitch model $g$ [12]. The noise is then split into two components: $n = n_R + g$, where the glitches are modelled as localized concentrations of energy in the time–frequency plane, while the ‘regular’ component $n_R$ is assumed to be uncorrelated in frequency and well approximated by a Gaussian or some other similar simple distribution. The BayesWave algorithm for implementing this approach will be described in a subsequent section.

### 5. Signal models

Typical analyses focus on a single class of gravitational wave signals. Examples include the inspiral signal from a neutron star binary, the merger signal from two colliding black holes, and stochastic signals from processes in the early Universe. The more tightly focused the search, and the more prescriptive the signal model, the deeper we can dig into the noise. On the other hand, we also want to be open to finding unexpected signals about which little is known.

Within the confines of general relativity, the most general signal model is one in which the waves come in two transverse polarizations and travel at the speed of light. The parameters in this model are the discrete time samples $h_-(t)$ and $h_+(t)$. Using three or more independent and non-aligned detectors, it is possible to infer the presence of bright gravitational wave signals that stand out above the noise in each detector [32,33]. In order to detect weaker signals, we need more restrictive models for the signals and the instrument noise. For example, we could demand a degree of smoothness in the signal model by parametrizing the signal in terms of control points of a cubic spline. This reduces the number of parameters in the signal model, and allows us to find signals that are slightly below the noise level. Similar gains can be achieved by adopting a time–frequency decomposition of the data such as a discrete wavelet basis and demanding that the signal is clustered in time and frequency. The model can be further narrowed by imposing restrictions on the frequency range and development in time. For example, the signal from a black hole binary should increase in amplitude and frequency as the system nears merger, so we can narrow our signal model to focus on chirp-like signals. Following this line of development, we eventually arrive at the so-called template-based analyses where approximate solutions to Einstein’s equations are used to predict the waveforms that are produced by various types of astrophysical sources. Rather than using discrete time samples or wavelet amplitudes, these
models describe the signals $h(t, \lambda)$ in terms of a relatively small number of physical parameters $\lambda$ that describe the source. One of the most important examples are the templates used to describe the inspiral and merger of two black holes, which are derived using a combination of analytic and numerical approximations, and yield waveforms that depend on just 17 parameters. The list of parameters includes the masses and spins of the black holes, the distance from the detector and the sky location.

Template-based searches allow us to detect signals that are buried in the instrument noise. The explanation for this is usually given in terms of matched filtering, and is attributed to a coherent build-up of power over many wave cycles in the cross-correlation of the signal and the template. While coherent integration over many cycles is important, the same coherent build-up occurs in models without templates [32]. The real reason why template-based models outperform less prescriptive signal models is that they involve far fewer parameters, and thus yield much higher evidence values. A template-based search for the inspiral signal from a binary neutron star may detect signals with integrated signal-to-noise SNR $\sim 7$ that builds up over $N \sim 10^4$ cycles. This corresponds to an average per-cycle or raw signal-to-noise $S/N = \text{SNR}/\sqrt{N}$ of around 0.07. In contrast, the most general signal model described above requires $S/N > 1$, and integrated signal-to-noise SNR $> 100$.

Another advantage of template-based models is that they allow us to extract detailed information about the sources. Posterior distributions for $p(h|s)$ become posterior distributions for the model parameters $p(\lambda|s)$, and we can quote confidence intervals for quantities, such as the masses, distance to the source and the sky location. The sky location is very important for performing coordinated observations with electromagnetic telescopes. Connecting with physical parameters is more difficult in models that work directly with the waveforms, but it is possible to produce posterior distributions for quantities, such as the duration, peak frequency, time–frequency volume and degree of polarization.

Rather than trying to infer $h(t)$ directly from the data, some analysis techniques use cross-correlations between data streams. The idea is that the signal components combine coherently while the instrument noise contributions average to zero. This approach is used to search for stochastic signals and long duration signals of unknown morphology [34]. By introducing appropriate time delays in the cross-correlation, the data from widely separated instruments can be used to perform a directional search [35,36].

6. Detection and characterization

The question of whether a signal has been detected or not is central to gravitational wave astronomy. In a Bayesian setting, we can attempt to answer this question by computing the odds ratio between competing hypotheses. Even without a good detection candidate, it is possible to constrain event rates and the astrophysical models used to predict these rates.

Physicists have a natural attraction to the idea that there is a number we can calculate that tells us which model better describes the data. It is easy to forget that the output of this calculation is only as good as its inputs. The choice of likelihood function can have a dramatic effect on the evidence and the odds ratio. What might seem like an extreme outlier when using a Gaussian distribution may be unsurprising using a Cauchy distribution. It is also important to be clear about what hypotheses are being tested. We are not making statements about whether all possible gravitational wave signals have been detected, but, rather, we are comparing some simplified noise model with some subset of possible signal models.

The detection problem is further complicated by the ‘needle-in-a-haystack’ challenge of locating weak signals in a vast search space. For example, to yield a good fit to a binary black hole merger signal in LIGO/Virgo data requires a template with a merger time that matches that of the signal to within a few milliseconds in a dataset that spans tens of billions of milliseconds. Other quantities, such as the ‘chirp mass’ (defined by $M = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$, where $m_1$ and $m_2$ are the individual component masses), also need to match to within tiny fractions of their prior range. Put another way, the region $\Delta V$ that contains 99 per cent of the posterior
mass occupies a minute fraction of the total prior volume $V$. Ratios in excess of $10^{40}$ are not unheard of. Standard implementations of MCMC and nested sampling algorithms have no hope of finding these regions. Instead, the analysis proceeds in two steps, starting with a search phase that attempts to locate the modes of the posterior, followed by a characterization phase where the regions of high posterior weight are mapped and the evidence is calculated.

During the search phase anything goes, and we are free to maximize over the parameters used to describe the templates. The time of arrival can be maximized using a standard Fourier transform trick, as can the initial phase. The amplitude or distance to the source can be solved for by analytically maximizing the posterior. In some cases, it is possible to maximize over several more parameters. These maximization techniques can dramatically shrink the search space. When few parameters remain to be searched over it becomes computationally feasible to methodically work through a grid of templates that have been laid out so as to ensure some minimum overlap for signals with parameters that lie between the grid points. Most of the current LIGO/Virgo searches are performed using template grids. When the remaining search space has more than three or four dimensions, it becomes necessary to use other techniques, such as hierarchical grids or semi-stochastic searches based on genetic algorithms or greedy non-Markovian variants of MCMC algorithms [37]. The posterior distributions for these more complex waveforms are often multi-modal, narrowly concentrated and highly curved when plotted in terms of the physical parameters that define the waveforms. The efficiency of the search algorithms is highly dependent on the parametrization that is used, and physical insight can help guide the selection of parametrizations where the posterior distribution takes a simpler form. In many cases, it is possible to use the functional form of the templates to predict where the secondary modes will be located. All it takes is for the search to find a secondary, tertiary or even a quaternary mode for the full posterior structure to be revealed [37].

The next task is to map out the posterior distribution and compute the model evidence. This is done using a combination of MCMC and nested sampling algorithms. Multi-modality and strong correlations between parameters make these distributions difficult to explore. This has led to the development of what can best be described as ‘everything but the kitchen sink’ MCMC algorithms [37] that combine a wide variety of techniques, such as parallel tempering, differential evolution, delayed rejection, mode hopping proposals and local diagonalization of the covariance matrix. The last technique can be applied to samples taken from the history of the chain (a procedure that can be shown to be asymptotically Markovian), or by computing the Fisher information matrix. Parallel tempering is very effective in helping the chains move between local modes since the hot chains with ‘temperature’ $T > 1$ explore a flatter-likelihood landscape with $p(s|h)_T = p(s|h)^{1/T}$. As an added bonus, having chains at different temperatures allows the evidence to be computed via thermodynamic integration, though attaining accurate estimates using this procedure can be computationally expensive. Nested sampling has become the preferred method for computing the evidence, especially since the general purpose MULTINEST software package has been made publicly available [38]. As with the MCMC routines, MULTINEST needs help to locate the tightly concentrated modes from within the vast prior volume. One solution is to run MULTINEST on small sub-volumes that contain the vast majority of the posterior mass and sum the contributions to the evidence. These sub-volumes can be identified from the MCMC chains. To ensure that the majority of the posterior mass is enclosed, it is a good idea to use the hot ($T \sim 4 \rightarrow 9$) MCMC chains, as these explore regions of lower likelihood.

7. Likelihood redux

The posterior distributions and odds ratios are highly dependent on the choice of the likelihood function. As Sivia & Rawlings [39] have pointed out, people are ‘led to regard the prior as subjective, whereas the likelihood is seen as objective. In reality, each is just a conditional PDF and, as such, they are on par with each other’. At least a poor choice of prior can be overwhelmed by sufficiently informative data, whereas the bias introduced by a poor choice of likelihood grows with the information content.
But how do we know if we have made a bad choice for the likelihood function? One way is to look for systematic bias in parameter recovery using an ensemble of signal injections. For example, do the injected signal parameters lie within the 90 per cent confidence intervals 90 per cent of the time? This frequentist style of testing can also be used to compare Monte Carlo studies of false-positive rates with what would be expected based on the odds ratio. It has even been suggested that we simply accept that our likelihood function is wrong, and treat the Bayes factors as a frequentist detection statistic to be calibrated using time slides and signal injections [40]. But this does not address the issue of bias in the posterior distributions [41].

Several authors have put forward likelihood functions that allow for non-Gaussianity and non-stationarity in the data [42,43]. The BayesWave algorithm introduced earlier takes a different approach, and rather than building in the effect of instrument glitches, it seeks to model individual glitches and remove them from the data [12]. The glitch model $g$ uses a discrete wavelet basis, with the wavelet amplitudes as the model parameters. The number of active (non-zero) pixels in the model is a free parameter that is determined from the data. The likelihood function assumes an independent Gaussian distribution for the residual $r = s - R \cdot h - g$ in each wavelet pixel. The variance of the residual in each frequency band over an approximately 1000 s time interval is a free parameter to be determined from the data. Allowing the noise level to float in this way lets us model slow drifts in the instrument sensitivity over time. The BayesWave algorithm is implemented using a reversible-jump MCMC (RJMCMC) routine that allows the number of active glitch pixels to vary. Relatively quiet stretches of data light up very few glitch pixels, while loud instrument glitches can light up hundreds of glitch pixels. The signal model can use waveform templates for targeted searches or wavelet amplitudes for a more general search. Wavelet amplitudes do not work for a single detector as there is nothing to distinguish between the signal and glitch models. With a network of detectors, it is more parsimonious to assign coincident excess power to the signal model than to the glitch model. The glitch identification can be enhanced by using physically motivated priors, such as assigning a higher probability to glitch pixels that have active neighbours.

The complexity of the BayesWave noise model does pose a challenge when trying to compute the model evidence. Since the number, location and amplitude of the active pixels are all variable, we are in fact dealing with a very large collection of distinct noise models that have to be marginalized over. Nested sampling is ill suited to this task, and it is more natural to extend the RJMCMC model space to include the signal model. The Bayes factor between the noise-only and signal-plus-noise model can be estimated from the number of iterations that are spent exploring each model. The challenge is finding proposals that allow the chains to jump between models. An un-informed jump proposal has little or no chance of landing in the small $\Delta V$ volume where the signal posterior has significant weight. Our solution is to do two pilot runs, one with the signal model and one without, and use the posterior distributions from these chains as proposal distributions for the trans-dimensional jumps in the combined run with both models [12]. The distributions can be stored in a sparse matrix representation of a multi-dimensional histogram with only a small number of significant cells, or in a K-dimensional data tree [44] with uniform occupancy and variable cell size.

Tests have shown that the residuals produced by the BayesWave algorithm pass all the standard tests of Gaussianity. All significant non-Gaussian features are assigned to either the signal or glitch models, and the signal parameters show none of the systematic biases that occur with a simple Gaussian likelihood.

8. The cocktail party problem

A new set of challenges awaits us in the next phase of gravitational wave astronomy, where we will have to contend with the signal-rich data streams from space-based detectors and third-generation ground-based detectors such as the Einstein telescope. In the near term, the data from pulsar timing arrays will pose their own challenges. The main difference from the LIGO/Virgo analysis is the long dwell time of the signals, which results in many signals being in-band at the
Figure 3. A simulation of the strain spectral density recorded by a future space-based gravitational wave observatory [45] broken out into the individual contributions from galactic sources (blue), instrument noise (red), massive black holes and EMRIs (shown in a multitude of colours). The black line denotes the spectrum of combined detector output. PSD, power spectral density; TDI X, time-delay-interferometry X-channel; SMBH, supermassive black hole. (Online version in colour.)

same time. A space-based detector operating in the milli-hertz regime will record the signals from tens of millions of short-period white dwarf binaries in our Galaxy, in addition to the signals from dozens of massive black hole mergers and possibly hundreds of capture signals from stellar-mass black holes or neutron stars being swallowed by a massive black hole (also called extreme mass ratio inspirals or EMRIs). Figure 3 shows a simulated power spectrum for 2 years of data from the proposed Laser Interferometer Space Antenna (LISA) mission that has been broken out into some of the individual contributions. The simulation includes a complete Galaxy realization, but only a small fraction of the expected number of black hole merger and capture signals. The signals from unresolved sources are the dominant source of ‘noise’ for the LISA mission concept. From this cacophony, we need to be able to pick out and interpret individual signals, a situation akin to trying to follow a conversation at a loud cocktail party.

The signals overlap in both time and frequency, and while having accurate template families for the signals helps, a serial analysis is not possible owing to correlations between the templates. We are faced with the challenge of having to simultaneously extract a vast number of signals, not knowing in advance how many signals are resolvable, or what their parameters might be. Again it is natural to separate the analysis into search and characterization stages. The search can start with the brightest signals and gradually work down to the weakest signals while updating the global solution along the way [46,47]. For the characterization stage, it is natural to consider an RJMCMC implementation that generalizes the methods used to identify an unknown number of sinusoids in noisy data [48,49]. A realistic implementation of this approach has yet to be demonstrated.
9. Summary

With the first direct detection of gravitational waves just around the corner, a large group of researchers is racing the clock to develop robust analysis techniques to maximize the science yield from the first detections. The application of Bayesian inference is becoming more widespread and established, but the complicated nature of the instrumental noise is a challenging problem that still needs to be addressed. Future detectors promise to deliver data that are so rich in signals that the unresolved components will become the dominant source of noise. After a century of waiting for the first detection, an over-abundance of signals will be a good problem to have.

The author is grateful for the comments and suggestions from Albrecht Rudiger, Ilya Mandel, Ray Frey, Ben Owen and Alberto Vecchio. This work was support by NSF award 0855407 and NASA grant NNX10AH15G.

References


