# 16
# Data processing, analysis, and storage for interferometric antennas

BERNARD F. SCHUTZ

## 16.1 Introduction

Laser-interferometric gravitational wave antennas face one of the most formidable data handling problems in all of physics. The problem is compounded of several parts: the data will be taken at reasonably high data rates (of the order of 20 kHz of 16 bit data); they may be accompanied by twice as much 'housekeeping' data to ensure that the system is working appropriately; the data will be collected 24 hours a day for many years; the data need to be searched in real time for a variety of rare, weak events of short duration (one second or less); the data need to be searched for pulsar signals; the data from two or more detectors should be cross-correlated with each other; and the data need to be archived in searchable form in case later information makes a re-analysis desirable. One detector might generate 400 Mbytes of data each hour. Even using optical discs or digital magnetic tapes with a capacity of 3 Gbytes, a network of four interferometers would generate almost 5000 discs or tapes per year. The gathering, exchange, analysis, and storage of these data will require international agreements on standards and protocols. The object of all of this effort will of course be to make astronomical observations. Because the detectors are nearly omni-directional, a network of at least three and preferably more detectors will be necessary to reconstruct a gravitational wave event completely, from which the astronomical information can be inferred.

In this chapter I will discuss the mathematical techniques for analysing the data and reconstructing the waves, the technical problems of handling the data, and the possibilities for international cooperation, as they appear in mid-1989. This discussion can only be a snapshot in time, and a personal one at that. The subject is one that can be expected to develop considerably in the next decade. I will orient the discussion toward ground-based interferometers, with the sensitivity and spectral range expected of the instruments that are planned to be built in the next decade. Much of the discussion naturally is equally applicable to present prototypes, but it is important to look ahead towards future detectors so that their data problems can be anticipated in their design. A large part of the section on data analysis also applies to space-based interferometers or to the analysis of ranging data for interplanetary spacecraft, although in these cases the volume of data is much lower because they operate as low-frequency detectors. I will also assume that the interferometers will operate with a bandwidth greater than that of the signal, even when they are configured in a resonant mode. In the extreme narrow-banding case, in which the detectors have a bandwidth smaller than that of the waves, the data analysis problem resembles that for bar detectors, as discussed by Pallottino and Pizzella in chapter 10.

### 16.1.1 Signals to look for

The likely sources of gravitational radiation are described by David Blair in part I of this book. If a source is strong enough to stand out above the noise in the time-series of data coming off the machine, then simple threshold-crossing criteria can be used to isolate candidate events. If the event is too weak to be seen immediately, it may still be picked up by pattern-matching techniques, but the sensitivity to such events will depend upon how much information we have about the expected waveform. At the present time, we have little idea of what waveform to expect from bursts of radiation from gravitational collapse (supernovae or electromagnetically quiet collapses), so their detectability depends upon their being strong enough to stand up above the broad-band noise. (Future detailed numerical calculations of gravitational collapse may change this, of course.) On the other hand, we have detailed predictions for the waveforms from binary coalescence and from continuous-wave sources such as pulsars; these can be extracted from noisy data by various techniques, such as matched filtering. Pulsars with a known position may be found from the output of a single detector by sampling techniques. An all-sky search for unknown pulsars will be performed at a sensitivity that will ultimately be limited by the available computing power. Cross-correlation techniques between detectors can search for a stochastic background of radiation and detect weak, unpredicted signals.

## 16.2 Analysis of the data from individual detectors

Bursts and continuous-wave signals can in principle be detected by looking at the output of one instrument. Of course, one must have coincident observations of the same waves in different detectors, for several reasons: to increase one's confidence that the event is real, to improve the signal-to-noise ratio of the detection, and to gain extra information with which to reconstruct the wave. The simplest detection strategy splits into two parts: first find the events in single detectors, then correlate them between detectors. In most cases this is likely to work, but in some cases it will only be possible to detect signals in the first place by cross-correlating the output of different detectors. In this section I will address the problem of finding candidate events in single detectors. Cross-correlation will be treated later.

### 16.2.1 Finding broad-band bursts

A broad-band burst is an event whose energy is spread across the whole of the bandwidth of the detector, which I will take to be something like 100–5000 Hz (although considerable efforts are now being devoted to techniques for extending the bandwidth down to 40 Hz or less). To be detected it has to compete against all of the detector's noise, and the only way to identify it is to see it across a pre-determined amplitude threshold in the time-series of data coming from the detector. The main burst of radiation from stellar core collapse may be like this. Numerical simulations of axisymmetric collapse (Evans, 1986; Piran and Stark, 1986) reveal, among other things, that after the main burst there is – at least if a black hole is formed – a 'ringdown phase' in which the radiation is dominated by the fundamental quasi-normal mode of the black hole. This phase lends itself to some degree of pattern-recognition, such as that which I will describe for coalescing binaries in the next section. But it is unlikely that ringdown radiation will substantially improve the signal-to-noise ratio of a collapse burst, since it is damped out very quickly. Some simplified models of non-axisymmetric collapse (e.g. Ipser and Managan, 1984) suggest that if angular momentum dominates and non-axisymmetric instabilities deform the collapsing object into a tumbling tri-axial shape, then a considerable part of the radiation will come out at a single slowly changing frequency. If future three-dimensional numerical simulations of collapse bear this out, then this would also be a candidate for pattern-matching. But one must bear in mind that even if we have good predictions of waveforms from simulations, there will be an intrinsic uncertainty due to our complete lack of knowledge of the initial conditions we might expect in a collapse, particularly regarding the angular momentum of the core. So it is not yet clear whether collapses will ever be easier to see than the time–series threshold criteria described next would indicate.

#### (i) Simple threshold criteria

The idea of setting thresholds is to exclude 'false alarms' – apparent events that are generated by the detector noise. Thresholds are set at a level which will guarantee that any collection of events above the threshold will be free from contamination from false alarms at some level. The 'guarantee' is of course only statistical, and it relies on understanding the noise characteristics of the detector. I will assume here that the noise is Gaussian and white over the observing bandwidth.

This should be a good first approximation, but there are at least two important refinements: first, detector noise is frequency-dependent, and when we consider coalescing binaries this will be important; and second, we must allow for unmodelled sources of noise that will occasionally produce large-amplitude 'events' in individual detectors.

This latter noise can be eliminated by demanding coincident observations in other detectors, provided we assume that it is independent of noise in the other detectors and that it is not Gaussian, in particular that there are fewer low-amplitude noise events for a given number of large-amplitude ones than we would expect of a Gaussian distribution. This implies that the cross-correlated noise between detectors will be dominated by the Gaussian component. These assumptions are usually made in data analysis, but it is important to check them as far as possible in a given set of data.

**Thresholds for single detectors** Assuming that the noise amplitude $n$ in any sampled point has a Gaussian distribution with zero mean and standard deviation $\sigma$, the probability that its absolute value will exceed a threshold $T$ (an event that we call a 'false alarm' relative to the threshold $T$) is

$$p(|n|>T)=\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{\sigma}\int_T^{\infty}\mathrm{e}^{-n^2/2\sigma^2}\,\mathrm{d}n=\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{\sigma}{T}-\frac{\sigma^3}{T^3}+\cdots\right)\mathrm{e}^{-T^2/2\sigma^2}. \quad (16.1)$$

In the asymptotic approximation given by the second equality, the first term gives 10% accuracy for $T>3.2\sigma$, and the first two terms give similar accuracy for $T>2.5\sigma$. If we want the expected number of false alarms to be one in $N_{\text{obs}}$ data points, then we must choose $T$ such that

$$p(|n|>T)=1/N_{\text{obs}}. \quad (16.2)$$

This is a straightforward transcendental equation to solve. For example, if we imagine looking for supernova bursts of a typical duration of 1 ms, then we might be sampling the noise in the output effectively 1000 times per second. (If we want to reconstruct the waveform we might use the data at its raw sampled rate, say 4 kHz; but this would require a larger signal-to-noise ratio than simple detection, for which we could use the data sampled at or averaged over 1 ms intervals.) If we wish no more than one false alarm per year, then we must choose $T=6.6\sigma$.

**Thresholds for multiple detectors** If we have two detectors, with independent noise but located on the same site, then we can dig deeper into the noise by accepting only *coincidences,* which occur when both detectors simultaneously cross their respective thresholds $T_1$ and $T_2$. Given noise levels $\sigma_1$ and $\sigma_2$, respectively, the criterion for the threshold is

$$p(|n|>T_1)p(|n|>T_2)=1/N_{\text{obs}}. \quad (16.3)$$

For two identical detectors ($\sigma_1=\sigma_2$), each making 1000 observations per second, the threshold $T$ needs to be set at only $4.5\sigma_1$ to give one false alarm per year. Similarly, three identical detectors on the same site require $T=3.6\sigma$ and four can be set at $T=3.0\sigma$. The improvement from two to four detectors is a factor of 1.5 in sensitivity, or a factor of three in the volume of space that can be surveyed, and hence a similar improvement in the expected event rate. This favourable cost/benefit ratio – in this case, a factor of three improvement in event rate for a

Table 16.1. *Thresholds (in units of σ) for various arrays and false-alarm probabilities.*

| Number of detectors | False-alarm probability $1/3\times10^{10}$ | $1/1.5\times10^{12}$ | $1/6\times10^{12}$ | $1/3\times10^{14}$ |
|---|---|---|---|---|
| 1 | 6.63 | 7.19 | 7.37 | 7.88 |
| 2 | 4.53 | 4.93 | 5.06 | 5.43 |
| 3 | 3.59 | 3.92 | 4.03 | 4.33 |
| 4 | 3.03 | 3.31 | 3.41 | 3.67 |

factor of two increase in expenditure – is characteristic of networks of gravitational wave detectors, and indeed of any astronomical detector network whose sensitivity is limited by internal noise uncorrelated between instruments. In table 16.1 appropriate thresholds for a number of possible computer arrays and interesting false-alarm probabilities are given. (The last two columns are relevant to coalescing binaries, as discussed later.) The detectors are assumed to be identical. Notice that the thresholds are relatively insensitive to the false-alarm probability, since we are far out on the Gaussian tail. Thresholds are given in units of $\sigma$, the r.m.s. noise amplitude.

**(ii) Threshold criteria with time delays**

I have qualified the discussion of multiple detectors so far by demanding that they be on the same site; the reason is that if they are separated, then allowing for the possible time delay between the arrival of a true signal in different detectors opens up a larger window of time in which noise can masquerade as signal. Suppose that two detectors are separated by such a distance that the maximum time delay between them is $W$ measurement intervals. (For example, Glasgow and California are separated by about 25 ms, which we take to be effectively ±25 measurement intervals for collapse events. This gives a total window size of 50 measurements.) Then in equation (16.3), the appropriate probability to use on the right-hand side is $1/N_{obs}/W$, since each possible 'event' in one detector must be compared with $W$ possible coincident ones in the other.

In table 16.1, the second and fourth columns of thresholds correspond to false-alarm probabilities that are one-fiftieth of the first and third columns, respectively. For two identical detectors, this 'typical' window $W = 50$ raises the threshold $T$ from $4.53\sigma$ to $4.93\sigma$. This is a 9% decrease in sensitivity, or a 29% decrease in the volume of space that can be surveyed.

For three detectors, the situation begins to get more complex: as we will see later, if three detectors see an event that lasts considerably longer than their resolution time, there is a self-consistency check which may be used to reject spurious coincidences. (The check is that three detectors can determine the direction to the source, which must of course remain constant during the event.) For four detectors, even a few resolution times are enough to apply a self-consistency check. In principle, the quantitative effect of these corrections will depend on the signal-to-noise ratio of the event, since strong events can be checked for consistency more rigorously than weak events. But the level of the threshold in turn will determine the minimum signal-to-noise ratio. A full study of this problem has not yet been made, and can probably only be undertaken in the light of a more thorough investigation of the signal-reconstruction problem (see section 16.5).

### 16.2.2 Extracting coalescing binary signals

Coalescing binaries are good examples of the type of signal that will probably only be seen by applying pattern-matching techniques: the raw amplitude from even the nearest likely source will be below the level of broad-band noise in the detector. Nevertheless, the signal is so predictable that interferometers should be able to see such systems ten times or more as distant as collapsed sources. We will see that the signal depends on two parameters, so when we discuss the coincidence problem from the point of view of pattern-matching, we will have to consider the added uncertainty caused by this.

#### (i) The coalescing binary waveform

The amplitude of the radiation from a coalescing binary depends on the masses of the stars and the frequency $f$ of the radiation, which together determine how far apart the stars are. It is usual to assume that the stars are in circular orbits. This is a safe assumption if the binary system has existed in its present form long enough for its orbit to have shrunk substantially, since the timescale for the loss of eccentricity, $e/\dot{e}$, is 2/3 of the similar timescale for the decrease of the semimajor axis $a$. If the binary has only recently been formed, e.g. by tidal capture in a dense star cluster, then more general waveforms can be expected. This complication will not be treated here.

**Amplitude** The model assumes point particles in a Newtonian orbit, with energy dissipation due to quadrupolar gravitational radiation reaction; corrections to this are discussed briefly below. The radiation amplitude when the radiation frequency is $f$ is given by the function:

$$A_h(f) = 2.6 \times 10^{-23}\left(\frac{\mathcal{M}}{M_\odot}\right)^{5/3}\left(\frac{f}{100\ \mathrm{Hz}}\right)^{2/3}\left(\frac{100\ \mathrm{Mpc}}{r}\right), \tag{16.4}$$

where $\mathcal{M}$ is what I shall call the *mass parameter* of the binary system, defined for a system consisting of stars of masses $m_1$ and $m_2$ by the equation

$$\mathcal{M} = m_1^{3/5} m_2^{3/5}/(m_1 + m_2)^{1/5}, \tag{16.5}$$

or equivalently by the more transparent formula,

$$\mathcal{M}^{5/3} = \mu M_{\mathrm{T}}^{2/3}, \tag{16.6}$$

where $\mu$ is the usual reduced mass and $M_T$ the total mass of the system. A system consisting of two $1.4M_\odot$ stars has $\mathcal{M} = 1.22M_\odot$.

The numerical value of $A_h(f)$ is actually the *maximum* observable value of the amplitude which one obtains when the system is viewed down the axis of its angular momentum. One must insert angular factors in front of the expression to get the wave amplitude in other directions. If one averages over these angular factors *and* over the angular factors that describe the antenna pattern of an interferometer, one obtains an effective *mean amplitude* only 2/5 of the maximum (Krolak, 1989; Thorne, 1987).

**Frequency** The binary's orbital period changes as gravitational waves extract energy from the system. The frequency of the radiation is twice the orbital frequency, and its rate of change is

$$\frac{\mathrm{d}f}{\mathrm{d}t} = 13\left(\frac{\mathcal{M}}{M_\odot}\right)^{5/3}\left(\frac{f}{100\ \mathrm{Hz}}\right)^{11/3}\ \mathrm{Hz\,s^{-1}}. \tag{16.7}$$

The maximum wave amplitude we expect, therefore, has the time-dependence

$$h_{\max}(t) = A_h[f(t)]\cos\left(2\pi\int_{t_a}^{t} f(t')\,\mathrm{d}t' + \Phi\right), \tag{16.8}$$

where $t_a$ is an arbitrarily defined 'arrival time', at which the signal reaches the frequency $f_a$, and $\Phi$ is the signal's phase at time $t_a$. This depends on where in their orbits the stars are when the frequency reaches $f_a$. The amplitude increases slowly with the frequency-dependence of $A_h$.

Doing the frequency integral explicitly gives

$$f(t) = 100\ \mathrm{Hz}\times\left[\left(\frac{f_a}{100\ \mathrm{Hz}}\right)^{-8/3} - 0.33\left(\frac{\mathcal{M}}{M_\odot}\right)^{5/3}\left(\frac{t-t_a}{1\ \mathrm{s}}\right)\right]^{-3/8}. \tag{16.9}$$

The phase integral is then

$$2\pi\int_{t_a}^{t} f(t')\,\mathrm{d}t' = 3000\left(\frac{\mathcal{M}}{M_\odot}\right)^{-5/3}$$
$$\times\left\{\left(\frac{f_a}{100\ \mathrm{Hz}}\right)^{-5/3} - \left[\left(\frac{f_a}{100\ \mathrm{Hz}}\right)^{-8/3} - 0.33\left(\frac{\mathcal{M}}{M_\odot}\right)^{5/3}\left(\frac{t-t_a}{1\ \mathrm{s}}\right)\right]^{5/8}\right\}. \tag{16.10}$$

Putting this into equation (16.8) for $h_{\max}(t)$ gives the desired formula, which we will use in the next section.

Notice that coalescence in the two-point-particle model occurs when $f = \infty$. For a system whose radiation is at frequency $f$, the remaining lifetime until this occurs is

$$T_{\mathrm{coal}}(f) = 3.0\left(\frac{\mathcal{M}}{M_\odot}\right)^{-5/3}\left(\frac{f}{100\ \mathrm{Hz}}\right)^{-8/3}\ \mathrm{s}. \tag{16.11}$$

This is 3/8 of the formal timescale $f/\dot{f}$ deducible from equation (16.7). Of course, for realistic stars the Newtonian point-particle approximation breaks down before this time, but if the stars are neutron stars or solar-mass black holes, corrections need be made only in the last second or less. Corrections due to post-Newtonian effects are the first to become important in this case, followed by tidal and mass-transfer effects. These have been considered in detail by Krolak (1989) and Krolak and Schutz (1987). If at least one of the stars is a white dwarf, tidal corrections will become important when $T_{\rm coal}$ is still 1000 years or so, and $f$ is tens of millihertz; the system would only be observable from space (Evans, Iben and Smarr, 1987).

**Fourier transform of the coalescing binary signal** We shall need below not only the waveform $h(f)$, but also its Fourier transform. We shall denote the Fourier transform of any function $g(t)$ by $\tilde{g}(f)$, given by

$$\tilde{g}(f) = \int_{-\infty}^{\infty} g(t)\mathrm{e}^{-2\pi i f t}\,\mathrm{d}t. \tag{16.12}$$

Provided that the frequency of the coalescing binary signal is changing relatively slowly (i.e., that $T_{\rm coal} \gg 1/f$), the method of stationary phase can be used to approximate the transform of $h_{\max}(t)$, $\tilde{h}_{\max}(f)$ (Dhurandhar, Schutz and Watkins, 1990; Thorne, 1987). We shall only need its magnitude,

$$|\tilde{h}_{\max}(f)| \approx 3.7 \times 10^{-24}\left(\frac{\mathcal{M}}{M_{\odot}}\right)^{5/6}\left(\frac{f}{100\ \mathrm{Hz}}\right)^{-7/6}\left(\frac{100\ \mathrm{Mpc}}{r}\right)\mathrm{Hz}^{-1}. \tag{16.13}$$

This gives good agreement with the results of some numerical integrations performed by Schutz (1986). We shall use it in the following sections.

### (ii) The mathematics of matched filtering: finding the signal

Matched filtering is a linear pattern-matching technique designed to extract signals from noise. For references on the theory outlined in this and subsequent sections, the reader may consult a number of books on signal analysis, such as Srinath and Rajasekaran (1979).

**Describing the noise** To use matched filtering we have first to define some properties of the noise, $n(t)$. We expect that $n(t)$ will be a random variable, and we use angle brackets $\langle\ \rangle$ to denote expectation values of functions of this noise. It is usually more convenient to deal with the noise as a function of frequency, as described by its Fourier transform $\tilde{n}(f)$. We shall assume that the noise has zero mean,

$$\langle n(t)\rangle = \langle \tilde{n}(f)\rangle = 0.$$

We shall also assume that the noise is *stationary*, i.e. that its statistical properties are independent of time. Then the *spectral density* of (amplitude) noise $S(f)$ is defined by the equation

$$\langle \tilde{n}(f)\tilde{n}^*(f')\rangle = S(f)\delta(f-f'), \tag{16.14}$$

where a * denotes complex conjugation. This says two things: (i) the noise at different frequencies is uncorrelated; and (ii) the autocorrelation of the noise at a single frequency has variances $S(f)$, apart from the normalization provided by the delta function, which arises essentially because our formalism assumes that the noise stream is infinite in duration. (Texts on signal processing often define $S(f)$ in terms of a normalized Fourier transform of the autocorrelation function of a discretely sampled time-series of noise $n_j(t)$. The continuous limit of this definition is equivalent to ours.) Since $n(t)$ is real, $S(f)$ is real and an even function of $f$.

**Noise in an interferometer** *White* noise has a constant spectrum, which means that $S(f)$ is independent of $f$. Interferometers have many sources of noise, as described in chapter 11 by W. Winkler in this volume or by Thorne (1987). In this treatment we will consider only two: shot noise, which limits the sensitivity of a detector at most frequencies; and seismic noise, which is idealized as a 'barrier' that makes a lower cutoff on the sensitivity of the detector at a frequency $f_s$.

The shot noise is intrinsically white (that is, as a noise on the photodetector), but — depending on the configuration of the detector — the detector's sensitivity to gravitational waves depends on frequency, so the relevant noise is the photon white noise divided by the frequency response of the detector (called its *transfer function*). We denote this 'gravitational wave' spectral density by $S_h(f)$. I will assume that the detector is in the standard recycling configuration, so that (allowing for the seismic cutoff) we have

$$S_h(f) = \begin{cases} \frac{1}{2}\sigma_f^2(f_k)[1 + (f/f_k)^2] & \text{for} \quad f > f_s, \\ \infty & \text{for} \quad f < f_s \end{cases} \tag{16.15}$$

Here $f_k$ is the so-called 'knee' frequency, which may be chosen by the experimenter when recycling is implemented, and $\sigma_f(f_k)$ is the standard deviation of the frequency-domain noise at $f_k$.

In the usual discussions of source strength vs. detector noise (e.g. Thorne, 1987), what is taken to be the detector noise as a function of frequency $f$ is $\sigma_f(f)$, not $[S_h(f)]^{1/2}$, because it is assumed in those discussions that the knee frequency $f_k$ will be optimized by the experimenter for the particular range of frequencies being studied, so that $\sigma_f$ is representative of the noise that the experimenter would encounter. Later in this section we will see that the optimum value of $f_k$ for observing coalescing binaries is $1.44 f_s$.

**The matched filtering theorem** Now, the fundamental theorem we need in order to extract the signal from the noise is the matched filtering theorem. If we have a signal $h(t)$ buried in noise $n(t)$, so that the output of our detector is

$$o(t) = h(t) + n(t),$$

and if the Fourier transform of the signal is $\tilde{h}(f)$, then any stationary, linear operation on the output can be expressed as a correlation with a *filter* $q(t)$:

$$c(t) = (o \circ q)(t)$$

$$= \int_{-\infty}^{\infty} o(t')q(t'+t)\,\mathrm{d}t' \tag{16.16}$$

$$= \int_{-\infty}^{\infty} \tilde{o}(f)\tilde{q}^*(f)\mathrm{e}^{2\pi i f t}\,\mathrm{d}f \tag{16.17}$$

The expectation value of the output $c(t)$ of the filter is the filter's signal,

$$\langle c(t)\rangle = (h \circ q)(t). \tag{16.18}$$

The noise that passes through the filter is Gaussian if $n(t)$ is Gaussian, and its variance is

$$\langle [c(t) - \langle c(t)\rangle]^2\rangle = \int_{-\infty}^{\infty} S(f)\,|\tilde{q}(f)|^2\,\mathrm{d}f. \tag{16.19}$$

This gives a 'raw' signal-to-noise ratio of

$$\frac{S}{N}(t) = \frac{(h \circ q)(t)}{\left[\int_{-\infty}^{\infty} S(f)\,|\tilde{q}(f)|^2\,\mathrm{d}f\right]^{1/2}}. \tag{16.20}$$

The idea of matching the filter to the signal comes from finding the filter $q(t)$ that maximizes this signal-to-noise ratio. It is not difficult to show that the optimal choice of filter for detecting the signal $h(t)$ is

$$\tilde{q}(f) = k\tilde{h}(f)/S_h(f), \tag{16.21}$$

where $k$ is any constant. With this filter, if the output contains a signal, then $c(t)$ will reach a maximum at a time $t$ that corresponds to the time in the output stream at which the signal reaches the point $t' = 0$ in the waveform $h(t')$. Of course, noise will distort the form of $c(t)$, but the expected amplitude signal-to-noise ratio $S/N$ in $c(t)$ (ratio of maximum value to the standard deviation of the noise) is given by the key equation

$$\left(\frac{S}{N}\right)^2_{\mathrm{opt}} = 2\int_0^{\infty} \frac{|\tilde{h}(f)|^2}{S_h(f)}\,\mathrm{d}f. \tag{16.22}$$

This is the largest $S/N$ achievable with a linear filter. Moreover, given a waveform $h(t)$ that one wants to look for, and given a seismic cutoff frequency $f_s$, one can ask what value of the knee frequency $f_k$ one should take in $S_h(f)$ in equation (16.22) to maximize $S/N$. For coalescing binaries, one can use the explicit expression for $\tilde{h}(f)$ given in equation (16.13) to show that this value, as mentioned earlier, is (Krolak, 1989; Thorne, 1987)

$$(f_k)_{\mathrm{opt}} = 1.44 f_s.$$

**Thresholds for the detection of coalescing binaries** Naturally, in a real experiment one does not know if a signal is present or not. One then uses the size of $S/N$ to decide on the likelihood of the correlation being the result of noise. A widely used criterion is the Neyman–Pearson test of significance (Davis, 1989), based on the *likelihood ratio*, defined as the ratio of the probability that the signal is present to the probability that the signal is absent (false alarm). If the noise is Gaussian, then the Neyman–Pearson 'best' criterion is just to calculate the chance of false alarm in the matched filter given by equation (16.21), exactly as described in section 16.2.1(i) with $x/\sigma$ replaced by $S/N$.

Searches for coalescing binaries can therefore be carried out by applying threshold criteria to the correlations produced by filtering. The false-alarm probabilities for detecting a coalescing binary have to be calculated with some care, however, because we must allow for the fact that we have in general to apply many independent filters, for different values of the mass parameter $\mathcal{M}$, and this increases the chance of a false alarm. I will consider the necessary corrections in section 16.2.2(iii) below.

**Determining the time-of-arrival of the signal** It is important for gravitational wave experiments that, by filtering the data stream, one not only determines the presence of a signal, but one also fixes its 'time-of-arrival', defined as the time $t_{\mathrm{arr}}$ at which the signal reaches the $t' = 0$ point in the filter $h(t')$. The standard deviation in the measurement of $t_{\mathrm{arr}}$ is $\delta t_{\mathrm{arr}}$, which is given by an equation similar to equation (16.22) (Dhurandhar, Schutz and Watkins, 1990; Srinath and Rajasekaran, 1979):

$$\frac{1}{\delta t_{\mathrm{arr}}^2} = 2\int_0^\infty \frac{|\tilde{\dot{h}}(f)|^2}{S_h(f)}\,\mathrm{d}f = 8\pi^2 \int_0^\infty \frac{f^2\,|\tilde{h}(f)|^2}{S_h(f)}\,\mathrm{d}f, \tag{16.23}$$

where $\tilde{\dot{h}}(f)$ is the Fourier transform of the time derivative of $h(t)$. If either the signal or the detector's sensitivity is narrow-band about a frequency $f_0$, then a reasonable approximation to equation (16.23) is

$$\delta t_{\mathrm{arr}} = \frac{1}{2\pi f_0}\frac{1}{S/N}, \tag{16.24}$$

where $S/N$ is the optimum signal-to-noise ratio as computed from equation (16.22). This is a good approximation as long as $S/N$ is reasonably large compared to unity. If we use equation (16.13) for $\tilde{h}(f)$ then it is not hard to show that, for coalescing binaries (Dhurandhar, Schutz and Watkins, 1990)

$$\delta \tau_{\mathrm{arr}} = 0.84\left(\frac{100\ \mathrm{Hz}}{f_s}\right)\frac{1}{S/N}\ \mathrm{ms}. \tag{16.25}$$

For example, if the signal-to-noise ratio is 7 (the smallest for detection by a single detector) and the seismic limit is 100 Hz, then the timing accuracy would be 0.1 ms. If the signal-to-noise is as high as 30, which could occur a few times per year (see below), then the signal could be timed to 30 $\mu$s. Considering that the time it takes the wave to travel from one detector to another will typically be 15–20 ms, this timing accuracy would translate into good directional information. I will explain below how this can be done.

However, in practice it will turn out that these numbers are too optimistic, perhaps by a factor of two. The reason is that one needs to determine other parameters as well from the signal, such as the mass parameter $\mathcal{M}$ and the phase. The errors in these parameters correlate, with the result that $\delta t_{\rm arr}$ is affected by, for example, $\delta\mathcal{M}$. Schutz (1986) has shown numerically that a small change in the mass parameter can masquerade as a displacement in the time-of-arrival of the signal. This effect will have to be quantified before realistic estimates of the timing accuracy can be made.

Another serious source of error in timing has been stressed by Alberto Lobo (private communication). As is apparent in the calculations of Schutz (1986), when a waveform has a frequency that changes only slowly with time, there can be an ambiguity in the identification of the peak in the correlation that gives the correct time-of-arrival. This is because a shift of the filter by one cycle relative to the waveform will not degrade the correlation much if the frequency is roughly constant. Our timing accuracy formula gives in some sense the width of the correlation peak, but the spacing between peaks is much larger, of order $1/f_0$ for coalescing binaries. Unless the signal-to-noise ratio is high enough to permit reliable discrimination between peaks, this may be the dominant timing error. It is possible that cross-correlation between detectors will still be able to give correct time delays, as in section 16.4.2 below, but this remains to be investigated.

It may seem paradoxical that, if detector physicists succeed in lowering the seismic barrier to, say, 50 Hz, the arrival-time-resolution given by equation (16.24) appears to get worse as $f_s^{-1}$! This is not a real worsening, of course: the increase in $S/N$ due to the lower seismic cutoff (gaining as $f_s^{-7/6}$ if $f_k$ remains optimized to $f_s$) more than compensates the $1/f_s$ factor, and the timing accuracy improves.

**Implications for the sampling rate** In practice, one only samples the data stream at a finite rate, not continuously. It is clear from equation (16.22) that one must sample at least as fast as is required to determine $\tilde{h}(f)$ at all frequencies that contribute significantly to the integral for the optimum signal-to-noise ratio: at least twice as fast as the largest required frequency in $\tilde{h}(f)$. For the coalescing binary, whose transform is given approximately by equation (16.13), the power spectrum $|\tilde{h}(f)|^2$ falls off as $f^{-7/3}$, and the recycling shot noise multiplies a further factor of $f^{-2}$ into this. Thus, when $f$ rises to, say, four times $f_s$, the integrand in equation (16.22) will have fallen off to about 0.005 of its value at $f_s$. Truncating the integration here should be enough to guarantee that the filter comes within 1% of the optimum signal-to-noise ratio. This would require a sampling rate of $8f_s$, or 800 Hz if we take $f_s = 100$ Hz.

Similar but more stringent requirements apply if one wants good timing. If the sampling rate is smaller than twice the largest frequency at which the integrand in equation (16.23) contributes significantly, then in the numerical calculation the arrival time accuracy will be worse than optimum. This is an important lesson: *in choosing one's sampling speed one should ensure that one can get good accuracy in equation* (16.23), *whose integrand falls off less rapidly with frequency than that of equation* (16.22). If one does sample at an adequate rate, then it is possible to determine the time-of-arrival of a signal to much greater precision than the sampling time, provided the signal-to-noise ratio is much greater than unity. (See, for example, the numerical experiments reported by Gursel and Tinto, 1989.) For a coalescing binary, taking timing accuracy into account does not significantly increase the sampling rate over that required for a good signal-to-noise ratio.

**Determining the parameters of the waveform** Naively, one might expect that by performing filtering of the incoming data stream with many independent filters, one would just identify the filter that gives the best correlation with the signal and then infer the mass parameter, phase, amplitude, and time-of-arrival from that. It is possible to do better than this, however, using these values as a starting point. This is called non-linear filtering, and there are many possible ways to proceed. For our problem, one of the most attractive is the Kallianpur–Striebel (KS) filter, described by Davis (1989). Rather than reproduce Davis's clear discussion of this method, I will simply refer the reader to his article and to the M.Sc. thesis of Pasetti (1987), which is the first attempt to design a numerical system capable of detecting coalescing binary signals and estimating their parameters. Pasetti gives listings of his computer programs and tests them on simulated data.

### (iii) Threshold criteria for filtered signals

**Number of filters needed** When searching a data stream for coalescing binary signals, we cannot presume ahead of time that we know what the mass parameter $\mathcal{M}$ will be: not all neutron stars may have mass $1.4M_\odot$, and some binaries may contain black holes of mass 15 or $20M_\odot$. We therefore will have to filter the data with a family of filters with $\mathcal{M}$ running through the range, say, $0.25$–$30M_\odot$.

How many filters should there be? This question has not yet received enough study. The calculations of Dhurandhar, Schutz and Watkins (1991) show that two filters with mass parameters differing by a few per cent have significantly reduced correlation, so the filters in the family should not be more widely spaced than this. However, it is not known whether they should be more closely spaced, to avoid missing weak signals. If we take successive filters to have mass parameters that increase by 1% at each step, then we need about 500 filters to span the range (0.25, 30) in $\mathcal{M}$.

However, there is also another parameter in the filter, equation (16.8): the phase $\Phi$, about which I have so far said little. When the wave arrives at the detector with frequency $f_s$, so that it is just becoming detectable, its phase may be anything: this depends on the binary's history. Filters with different phases must therefore be used. Inspection of equation (16.8) reveals that the phase is a constant within the cosine term for the duration of the signal. It follows that only two filters with different phases will suffice to determine the phase and amplitude of the signal on the assumption of a given mass parameter. For convenience one might choose $\Phi = 0$ and $\Phi = \pi/2$. This increases the number of filters to about 1000. In section 16.2.2(v) we will look at the computing demands that this filtering makes on the data analysis system. In the present section we shall consider the signal-to-noise implications.

**Effective sampling rate** First it will be necessary to establish what the filtering equivalent of the sampling rate is, so that we can calculate the probability of, say, one false alarm per year. In our original calculation of the false-alarm probability, the sampling rate told us how many independent data points there were per year, on the assumption of white noise, which meant that each data point was statistically independent, no matter how rapidly samples were taken. In the present case, the output of the filter is the correlation given in equation (16.16). It has noise in it, but the noise is no longer white, having been filtered. The key number that we want here is the 'decorrelation time', defined as the time interval $\tau_s$ between successive applications of the filter that will ensure that the outputs of the two filters are statistically independent. The analogue here of the sampling rate in the burst problem is $1/\tau_s$, which I will call the *effective sampling rate*. This is the rate at which successive independent data points arrive from each filter.

To develop a criterion for statistical independence, we consider the autocorrelation function of the filter output when the detector output $o(t)$ is pure noise $n(t)$:

$$a(\tau) = \int_{-\infty}^{\infty} c(t)c(t+\tau)\,\mathrm{d}t. \tag{16.26}$$

We shall take the decorrelation time to be the time $\tau_s$ such that $a(\tau)$ is small for all $\tau > \tau_s$. We can learn what this is by noting that it is not hard to show that the Fourier transform of $a(\tau)$ is, when the optimal filter given in equation (16.21) is used,

$$\tilde{a}(f) = \frac{|\tilde{h}(f)|^2}{S_h(f)}. \tag{16.27}$$

For coalescing binaries, we have already discussed some of the properties of this function in section 16.2.2(ii). It is strongly peaked near $f_s$, and in particular the seismic barrier cuts it off rapidly below $f_s$. It follows that for times $\tau \gg 1/f_s$ the autocorrelation function is nearly zero: the effective sampling rate is about $f_s$. To play it safe, we will work with a rate twice this large, or an effective sampling time of 0.005 s. This gives effectively $6 \times 10^9$ samples – statistically independent filter outputs – per year.

**Thresholds for coalescing binary filters** Now, assuming that the noise is Gaussian, the calculation of the false-alarm probability for any size network looks similar to our earlier one in section 16.2.1(ii). What we have to allow for is that there will be some 1000 independent filters, each of which could give a false alarm. Of course, the false alarm occurs only if each detector registers an event in the *same* filter, so it is like doing 1000 independent experiments with no filter at all and a sampling time of 0.005 s, or one experiment with no filter and a sampling time of $5 \times 10^{-6}$ s. This increases the number of points by a factor of 200 over the number we used in section 16.2.1(i), but this factor makes only a modest difference in the level of the thresholds. For example, for one false alarm per year, and no correction for time-delay windows, the thresholds are: for one detector, 7.4; for two, 5.1; for three, 4.0; and for four, 3.4. For example, the three-detector threshold is 12% higher than for unfiltered data taken at 1 kHz. For further details see table 16.1.

These figures should not be taken as graven in stone: they illustrate the consequences of a particular set of assumptions. A better calculation of the noise properties of the filters is needed, and in any case one will have to ensure that the detector noise really obeys the statistics we have assumed.

### (iv) Two ways of looking at the improvement matched filtering brings

The discussion of matched filtering so far has been fairly technical, with the emphasis on making reliable and precise estimates of the achievable signal-to-noise ratios and timing accuracy. In this section I will change the approach and try to develop approximate but instructive ways of looking at the business of matched filtering. The idea is to understand how matched filtering improves the sensitivity of an interferometer beyond its sensitivity to wide-band bursts. We will look at two points of view: comparing the sensitivity of the detector to broad-band and narrow-band signals that have either (i) the same amplitude or (ii) the same total energy.

**Improving the visibility of signals of a given amplitude** Let us consider two signals of the same amplitude $h$, one of which is a broad-band burst of radiation centred at $f_0$ and the other of which is a relatively narrow-band signal with $n$ cycles at roughly the frequency $f_1$. The signals are observed with different recycling detectors optimized at their respective frequencies, $f_0$ and $f_1$, possibly contained in the same detector system, as is envisioned in some present designs. The broad-band signal has

$$\begin{aligned}\left(\frac{S}{N}\right)^2_{\mathrm{bb}} &= 2\int_0^{\infty} \frac{|\tilde{h}(f)|^2}{S_h(f)}\,\mathrm{d}f. \\ &\approx \frac{2}{\sigma_f^2(f_0)}\int_0^{\epsilon} |\tilde{h}(f)|^2\,\mathrm{d}f \\ &\approx \frac{1}{\sigma_f^2(f_0)}\int_{-\epsilon}^{\epsilon} |h(t)|^2\,\mathrm{d}t. \end{aligned} \qquad (16.28)$$

Now, the integrand in equation (16.28) for a burst lasts typically only for a time $1/f_0$, so we have

$$\left(\frac{S}{N}\right)_{\rm bb} \approx \frac{h}{\sigma_f(f_0)f_0^{1/2}}. \tag{16.29}$$

For the narrow-band signal, we obtain again equation (16.28), but with $f_0$ replaced by $f_1$. Now, however, the signal lasts $n$ cycles, a time $n/f_1$. This leads immediately to

$$\left(\frac{S}{N}\right)_{\rm nb} \approx \frac{hn^{1/2}}{\sigma_f(f_1)f_1^{1/2}}. \tag{16.30}$$

Comparing equations (16.29) and (16.30), we see that a narrow-band signal has an advantage of $n^{1/2}$ over a burst of the same amplitude and frequency, provided we have enough understanding of the signal to use matched filtering*.

For the coalescing binary one may approximate $n$ by $f^2/\dot{f}$, and this can be large (of order 200). Coalescing binaries gain further when compared to supernova bursts because of their lower frequency: because $\sigma_f$ depends on $f$ as $f^{1/2}$, there is a further gain of a factor of $f_0/f_1$, which can be 7 or so. Therefore, a coalescing binary signal might have something like 100 times the $S/N$ of a supernova burst *of the same amplitude*! This exaggerates somewhat the advantage that coalescing binaries have as a potential source of gravitational waves, since their intrinsic amplitudes may be smaller than those from supernovae, but it does show why they are such interesting sources.

**Improving the visibility of signals of a given energy** The other way of looking at filtering is in terms of energy. This is very instructive, because it shows 'why' matched filtering works. We have just seen that a narrow-band signal with $n$ cycles has a higher $S/N$ than a broad-band burst of one cycle that has the same amplitude and frequency, by a factor of $n^{1/2}$. But the *energy* in the narrow-band signal is $n$ times that in the burst. This is because the energy flux in a gravitational wave is

$$\mathscr{F}_{\rm gw} \approx \frac{4c^3}{\pi G} h^2 f^2, \tag{16.31}$$

and thus the total energy $E$ in a signal passing through a detector during the time $n/f$ that the burst lasts is given by the proportionality

$$E \propto h^2 f^2 (n/f) = nfh^2.$$

If we solve this expression for $nh^2$ and put it into equation (16.30), we find

$$\frac{S}{N} \propto \frac{E^{1/2}}{f\sigma_f(f)}. \tag{16.32}$$

* For this reason, plots of burst sensitivity for broadband detectors, such as one finds in Thorne (1987), typically plot the *effective amplitude* $hn^{1/2}$ of a signal, rather than just $h$. This allows one to compare supernova bursts and coalescing binary signals on the same graph.

Since this is independent of $n$, it applies to broad-band and narrow-band signals equally. It shows that if two signals send the same total energy through an interferometric detector, and if they have the same frequency, then they will have the same signal-to-noise ratio, again provided we have enough information to do the matched filtering where necessary.

This provides a somewhat more realistic comparison of coalescing binaries and supernovae, since a coalescing binary radiates a substantial amount of energy in gravitational waves, of the order of $0.01M_{\odot}$. This is similar to the energy one might expect from a moderate to strong gravitational collapse. The advantage that coalescing binaries have is that they emit their energy at a lower frequency. The factor of $f\sigma_f \propto f^{3/2}$ in equation (16.32) gives them an advantage of a factor of roughly 20 over a collapse generating the same energy at the same distance. If laser interferometric detectors achieve a broad-band sensitivity of $10^{-22}$, as current designs suggest will be possible, then they will be able to see moderate supernovae as far away as 50 Mpc. This volume includes several starburst galaxies, where the supernova rate may be much higher than average. They will therefore also be able to see coalescing binaries at distances approaching 1 Gpc.

### (v) The technology of real-time filtering

**Basic requirements** In this section I will discuss the technical feasibility of performing matched filtering on a data stream in 'real time', i.e. keeping up with the data as it comes out of a detector. Since coalescing binaries seem to make the most stringent demands, I will take them as fixing the requirements of the computing system. We have seen that we need a data stream sampled at a rate of about 1 kHz in order to obtain the best $S/N$ and timing information, so I will use this data rate to discover the minimum requirements. It is likely that the actual sampling rates used in the experiments will be much higher, but they can easily be filtered down to 1 kHz before being analysed. If the seismic cutoff is 100 Hz, then the duration of the signal, at least until tidal or post-Newtonian effects become important, will be less than 2 s in almost all cases. This means that a filter need have no more than about 2000 2-byte data points.

The quickest way of doing the correlations necessary for filtering is to use fast Fourier transforms (FFTs) to transform the filter and signal, multiply the signal transform by the complex conjugate of the filter transform, and invert the product to find the correlation. The correlation can then be tested for places where it exceeds pre-set thresholds, and the resulting candidate events can be subjected to further analysis later. This further analysis might involve: finding the best value of the mass parameter and phase parameter; filtering with filters matched to the post-Newtonian waveform to find other parameters that could determine the individual masses of the stars; looking for unmodelled effects, such as tides or mass transfer; looking for the final burst of gravitational radiation as the two stars coalesce; and of course processing lists of these events for comparison with the outputs of other detectors. Since the number of significant events is likely to be relatively small, the most demanding aspect of this scenario is likely to be the initial correlation with 1000 coalescing binary filters.

**Discrete correlations** One way the processing might be done is as follows. The discrete correlation between a data set containing the $N$ values $\{d_j, j = 0, \ldots, N-1\}$ and a filter containing the $N$ values $\{h_k, k = 0, \ldots, N-1\}$ is usually given by the *circular* correlation formula:

$$c_k = (d \circ h)_k = \sum_{j=0}^{N-1} d_j h_{j+k}, \qquad k = 0, \ldots, N-1, \tag{16.33}$$

where we extend the filter by making it periodic:

$$h_{j+N} = h_j \,\forall j.$$

The circular correlation formula has a danger, because the data set and filter are not really periodic. In practice, this means that we should make the data set much longer than the (non-zero part of the) filter, so that only when the filter is 'split' between the beginning and the end of the data set does the circular correlation give the wrong answer. Thus, even if each filter requires only $N_h \leq 2000$ points, it is more efficient to split the data set up into segments of length $N \gg N_h$ points, and to use a filter which has formally the same length, but the first $N - N_h$ of whose elements are zero. (I am grateful to Harry Ward for stressing the need to pay attention to this point.) The 'padding' by zeros ensures that the periodicity of $h$ corrupts only the last $N_h$ elements of the correlation. This can be rectified by forgetting these elements and beginning the next data segment $N_h$ elements before the end of the previous one: this overlap ensures that the first $N_h$ elements of the next correlation replace the corrupt elements of the previous one with correct values. Since this procedure involves filtering some parts of the data set twice, it is desirable to make it a small fraction of the set, namely to make $N_h$ small compared to $N$. This efficiency consideration is, however, balanced by the extra numerical work required to calculate long correlations, increasing as $\ln N$. This arises as follows.

**Correlation by FFT** The fastest way to do long correlations on a general-purpose computer is to use Fourier transforms (or related Hartley transforms). For a discrete data set $\{d_j, j = 1, \ldots, N-1\}$ the discrete (circular) Fourier transform (DFT) is the set $\{\tilde{d}_k, k = 1, \ldots, N-1\}$ given by

$$\tilde{d}_k = \sum_{j=0}^{N-1} d_j \mathrm{e}^{-2\pi i jk/N}, \tag{16.34}$$

whose inverse is

$$d_j = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{d}_k \mathrm{e}^{2\pi i jk/N}. \tag{16.35}$$

Then the discrete version of the convolution theorem equation (16.17) is as follows. Given the (circular) correlation $\{c_j\}$ of two sets $\{d_j\}$ and $\{h_j\}$ as in equation (16.33), its DFT is

$$\tilde{c}_k = (\tilde{d}_k)^* \tilde{h}_k, \tag{16.36}$$

where an asterisk denotes complex conjugation.

Fast Fourier transform (FFT) algorithms may require typically $3N \log_2 N$ real floating-point operations (additions and multiplications) to compute the transform of a set of $N$ real elements, provided $N$ is an integer power of two (which can usually be arranged). (I neglect overheads due to integer arithmetic concerned with the index manipulations in such routines and, possibly significantly, memory access overheads.) To compute the correlation of two such sets, then, would require three transforms – two to produce $\tilde{d}_k$ and $\tilde{h}_k$ and a third to invert the product $\tilde{c}_k$ – and the multiplication of the two original transforms, giving a total of $9N \log_2 N + 4N$ real floating-point operations. This is to be compared with the $2N^2 - N$ operations required to calculate the correlation directly from equation (16.33). As long as $N \geq 16$ it will be quicker to use FFTs.

In practice, one would compute once and store the DFT of all $M$ filters, so that in real time the data would have to be transformed only once, and then $M$ products of data and filter calculated and inverse-transformed. This would require $3N(M+1)\log_2 N + 4NM$ floating-point operations.

**Optimal length of a data set** We must now remind ourselves that in order to achieve the economies of the FFT algorithm, we must use the circular correlation, which has an extra cost associated with the overlaps we are required to take in successive data sets. For a given filter length (say $N_f < N$ non-zero points in the filter time-series), we can reduce the fractional size of these overlaps by making $N$ larger, but this increases the cost of the FFT logarithmically in $N$. Is there an optimum ratio $N_f/N$? The total cost of analysing a data set containing a very large number $N_{tot} \gg N$ of elements, split up into segments of length $N$ is

$$N_{\text{fl-pt ops}} = \frac{N_{tot}}{N - N_f}[3N(M+1)\log_2 N + 4NM].$$

We want to minimize this with respect to variations in $N$ holding $N_f$ and $M$ (the number of filters) fixed. It is more convenient to introduce the varible $x = N_f/N$, which measures the fractional overlap of successive data sets. In terms of $x$ the expression is:

$$N_{\text{fl-pt ops}}(x) = \frac{N_{tot}}{1-x}\left[3(M+1)\log_2 \frac{N_f}{x} + 4M\right]. \tag{16.37}$$

As long as the number of filters $M$ is large, the optimum $x$ will be independent of $M$: it will depend only on $N_f$, the 'true' length of the filter. This is illustrated in table 16.2, which gives $x$ and $N_{\text{fl-pt ops}}/N_{tot}M$, the number of floating-point

Table 16.2. *The consequences of various strategies for applying filters of 'true' length $N_f$, padded out with zeros to a length $N$, to very long data sets. See text, especially equation (16.37), for details.*

| $N_f$ | $N$ | $x$ | $N_{\text{fl-pt ops}}/N_{\text{tot}}M$ |
|---|---|---|---|
| 1000 | $2^{11}$ | 0.488 | 72 |
| 1000 | $2^{12}$ | 0.244 | 53 |
| 1000 | $2^{13}$ | 0.122 | 49 |
| 1000 | $2^{14}$ | 0.061 | 49 |
| 1000 | $2^{15}$ | 0.031 | 50 |
| 2000 | $2^{12}$ | 0.488 | 78 |
| 2000 | $2^{13}$ | 0.244 | 57 |
| 2000 | $2^{14}$ | 0.122 | 52 |
| 2000 | $2^{15}$ | 0.061 | 52 |
| 2000 | $2^{16}$ | 0.031 | 54 |

operations per data point per filter, as required by various strategies, always taking $N_{\text{tot}}$ to be an integer power of two.

If we take $N_f$ to be 2000, then the optimum $x$ is 0.057; if $N_f = 1000$ then the best $x$ is 0.061. But the minimum in $N_{\text{fl-pt ops}}$ is a flat one, and one can increase the value of $x$ quite a bit without compromising speed. This is important, because each stored filter transform must contain $N$ points, so the larger we make $x$, the smaller will be our core memory requirements. From this it is clear that choosing an overlap between successive data sets of around 25% gives a CPU demand that is only slightly higher than optimum and reduces storage requirements to a minimum.

**Demands on computing power** Based on these calculations, and assuming a data rate of 1000 2-byte samples per second with a 2 s filter length ($N_f = 2000$), it follows that doing 1000 filters in real time requires a computer capable of 60 Mflops (where 1 Mflop is $10^6$ floating-point operations per second), and storage for 1000 filters, each of length 16 kbytes. This is within the capabilities of present-day inexpensive (<$100k) workstations with add-on array-processors, or of stand-alone arrays of transputers or other fast microprocessors. In five years it should be trivial.

There are many possible ways to speed up the calculation if CPU rates are a problem. It may be that special-purpose digital-signal-processing chips would be faster than general-purpose microprocessors for this problem. It might be possible to do the calculation in block-integer format rather than floating-point, with filters that consist of crude steps rather than accurate representations of the waveform (Dewey, 1986). These should be analysed further. Another possible CPU-saver is described in the next section.

### (vi) Smith's interpolation method for coalescing binaries

**A different way of looking at coalescing binary signals** An alternative strategy for coalescing binaries has been proposed and implemented by Smith (1987). This interesting idea is based upon the following observation: if two coalescing systems of different mass parameters happen to have the same time of coalescence, then their signals' frequencies will remain strictly proportional to one another right up to the moment of coalescence. This follows from the fact that $df/dt$ is proportional to a power of $f$, so that, as remarked after equation (16.11), there is a constant $\alpha$ independent of the masses such that $T_{\text{coal}} = \alpha f/\dot{f}$. If two signals with present frequencies $f_1$ and $f_2$ have the same $T_{\text{coal}}$, then it follows that

$$\frac{df_1}{df_2} = \frac{\dot{f}_1}{\dot{f}_2} = \frac{f_1}{f_2}.$$

Since if their times to coalescence are equal at one time then they are necessarily equal for all later times, this equation can be integrated to give $f_1/f_2 = \text{const}$.

Now suppose that the data stream is sampled at constant increments of the phase of signal 1, i.e. it is sampled at a rate that accelerates with the frequency $f_1$. Then if a Fourier transform is performed on the sampled points, the signal will appear just as pure sinusoid, allowing it to be identified without sophisticated filtering. Moreover, and this is the key point, every other signal with the same time to coalescence will have been sampled at constant increments of its phase as well, since its frequency has been a constant times the first signal's frequency. So signals from any binary coalescing at the same time, no matter what its mass parameter, will be exposed by the single Fourier transform. Thus, one Fourier transform would seem to have done the work of all 1000 filters!

**How much work is required?** The situation is not quite that good, however, because a signal with a different coalescence time will not be visible in the transform of the points sampled in the manner just described. Therefore, data must be sampled over again at the increasing rate *ending at each possible time of coalescence of the binary*. If this is done, then every possible signal will be picked up.

One way of implementing this method would be to sample the detector output at a constant rate (e.g. 1000 Hz) and then interpolate to form the data sets that are given to the FFT routine. (Livas, 1987, used this method to search for pulsars in a particular direction.) If we compare this interpolation method with the filtering described earlier, one trades the work of doing 1000 Fourier transforms on a stretch of data for the work of interpolating many times. The actual comparison depends on the number of operations required by the interpolation algorithm, but in general Smith's method with interpolation becomes more attractive as the number of filters one must use increases.

**Stroboscopic sampling** Another way of implementing Smith's method – and the way she herself used – would be to sample the detector output very fast, say at

10 kHz, and then to extract a data set at a slower rate (perhaps 500–1000 Hz) by selecting from the sampled points those points closest in time to the places one ideally would wish to sample. This is a far faster procedure than interpolating, and it seems to me that it would not necessarily be less accurate than a simple interpolation algorithm. I will call this *stroboscopic sampling*; we will meet it again when we discuss searches for pulsars. I do not know of any detailed theoretical analysis of it; in particular, one would like to understand what it does to the noise background. One also has to be careful about aliasing problems. The idea, at least in astronomy, seems to go back to Horowitz (1969), who devised it for optical searches for pulsars.

**Comparison with matched filtering** It may well be that for 1000 filters Smith's method will be more efficient than filtering. However, it has at least two significant disadvantages over filtering:

(1) It is restricted only to looking for the Newtonian coalescing binary signal: even any corrections (such as for post-Newtonian effects) will have to be searched for by filtering the sampled data sets, and the sets are essentially useless in searches for other kinds of signals that we may wish to filter from the data.
(2) Signals with the same coalescence time but different mass parameters will enter the observing window (say, $f > 100$ Hz) at different times, and this presents a possible problem that was first pointed out by Harry Ward. If one decides to break the data stream into sets of length, say, 2–3 s, appropriate to coalescing $1.4M_\odot$ neutron stars starting at 100 Hz, then the set will be much too long for a signal from a binary system of two $14M_\odot$ black holes that will coalesce at the same time. The black hole system will have frequency 24 Hz when the data set begins, and will be buried in the low-frequency detector noise. When the data are transformed, this noise will be included in the transform, and the signal-to-noise ratio will accordingly be reduced. The matched filtering method does not suffer from this drawback, since it filters out the low-frequency noise. It might be possible to avoid this problem by pre-filtering the data stream before it is sampled or interpolated, removing the low-frequency noise (and signal).

Given our present uncertainties about sources, my own prejudice is to use filtering because of its inherent flexibility; but Smith's method may become important if filtering places too great demands on the computing system.

### 16.2.3 Looking for pulsars and other fixed-frequency sources

#### (i) Why the data-analysis problem is difficult

There are many possible sources of gravitational radiation that essentially radiate at a fixed frequency. Pulsars, unstable accreting neutron stars (the Wagoner mechanism), and the possible long-term spindown of a newly formed neutron star are examples. In some cases, such as nearby known pulsars, we will know ahead of time the frequency to look for and the position of the source. But most continuous sources may have unknown frequencies; indeed they will only be discovered through their gravitational waves. I will first discuss the detection problem for sources of known frequency, and then consider searches for unknown sources. Throughout this discussion, the word 'pulsar' will stand for any continuous source with a stable frequency. The most complete discussion of this problem of which I am aware is the Ph.D. thesis of Livas (1987).

If we were on an observing platform that had a fixed velocity relative to the stars, and therefore to any pulsar we might be looking for, then finding the signal would be just a matter of taking the Fourier transform of the data and looking for a peak at the known frequency. This is a special case of matched filtering, since the Fourier integral is the same as the correlation integral in equation (16.17) with the filter equal to a sinusoid with the frequency of the incoming wave. Therefore, the signal-to-noise ratio for an observation that lasts a time $T_{obs}$ would increase as $T_{obs}^{1/2}$, just as in equation (16.30). However, the Earth rotates on its axis and moves about the Sun and Moon, and these motions would Doppler-spread the frequency and reduce its visibility against the noise.

How long do we have to look at a source before it becomes necessary to correct for the Earth's motion? If we consider only the Earth's rotation for the moment, then in a time $T_{obs}$ the detector's velocity relative to the source changes by an amount $\Delta v = \Omega_{\oplus}^2 R_{\oplus} T_{obs}$, where $R_{\oplus}$ is the Earth's radius and $\Omega_{\oplus}$ its angular velocity of rotation. In a source of frequency $f$, this produces a change $\Delta f_{Dop} = vf/c$. But the frequency resolution of an observation is $\Delta f_{obs} = 2/T_{obs}$. The Doppler effect begins to be important if $\Delta f_{Dop} = \Delta f_{obs}$. Solving this for $T_{obs}$ gives $T_{max}$, the maximum uncorrected observing time:

$$T_{max} = \left(\frac{2c}{\Omega_{\oplus}^2 f R_{\oplus}}\right)^{1/2} \approx 70\left(\frac{f}{1\ \mathrm{kHz}}\right)^{-1/2} \mathrm{min}. \tag{16.38}$$

Using the same formula for the effects of the Earth's orbit around the Sun gives a time roughly 2.8 times as long. The Earth's motion about the Earth–Moon barycentre also has a significant effect. Since any serious observation is likely to last days or longer, the Doppler effects of all these motions must be removed, even in searches for very low-frequency signals (10 Hz).

**(ii) Angular resolution of a pulsar observation**

The Doppler corrections one has to apply depend on the location of the source in the sky. Since the spin axis of the Earth is not parallel to orbital angular momentum vectors of its motion about the Sun or Moon, there is no symmetry in the Doppler problem, and every location on the sky needs its own correction.

It is of interest to ask how close two points on the sky may be in order to have the same correction; this is the same as asking what the angular resolution of an observation might be. Let us first imagine for simplicity that our detector participates in only one rotational motion, with angular velocity $\Omega$ and radius $R$. If two sources are separated on the sky by an angle $\Delta\theta$ (in either azimuth or altitude), then the difference between the Doppler corrections for the two sources depends on the *difference* between the changes in the detector's velocities relative to the two sources. For small $\Delta\theta$ this is $\Delta v = \Delta\theta\Omega^2 R T_{obs}$. Its maximum value is $2\Omega R\Delta\theta$. Using this velocity change, the argument is otherwise identical to that given in the previous section, provided that we keep $\Delta v$ no larger than $2\Omega R$. The result is that

$$\Delta\theta = T_{max}^2 \max\left(\frac{\Omega^2}{4}, \frac{1}{T_{obs}^2}\right). \tag{16.39}$$

The dependence of this expression on $T_{obs}$ will be significant when we come to discuss all-sky searches for pulsars in section 16.2.2(v) below, so it is well to remind ourselves how it comes about. There are two factors of $T_{obs}$ because, as $T_{obs}$ increases, (i) our frequency resolution increases, so we are more sensitive to the Doppler effect; and (ii) the Doppler velocity change over the observing period becomes larger.

When looking at a source with a frequency of 1 kHz, then for the Earth's rotation, and an observation lasting longer than half a day, this gives

$$\Delta\theta_{rot} = 0.02\left(\frac{f}{1\ \mathrm{kHz}}\right)^{-1} \mathrm{rad}, \tag{16.40}$$

which is about half a degree for a millisecond pulsar. The Earth's motion about the Earth–Moon barycentre can have a greater effect, falling to a minimum of 0.002 rad at two weeks. But this is swamped by the effect of the Earth's motion about the Sun, which gives

$$\Delta\theta_{orbit} = 1\times 10^{-6}\left(\frac{f}{1\ \mathrm{kHz}}\right)^{-1}\left(\frac{T_{obs}}{10^7\ \mathrm{s}}\right)^{-2} \mathrm{rad}, \qquad \text{for } T_{obs} < 1\times 10^7\ \mathrm{s}. \tag{16.41}$$

This reaches a minimum of about 0.2 arcsec for a millisecond pulsar observed for four months. Even at two weeks this motion gives a resolution of $2\times 10^{-5}$ rad, much finer than the Earth–Moon motion gives. So the orbital motion of the Earth always dominates the Earth–Moon motion. But it does not dominate the Earth's rotation for short times: up to about 20 hours the limit is given by equation (16.40).

For observations longer than about a day, the Earth's orbital motion therefore affords the better angular resolution, but it also makes the most stringent demands on applying the corrections. In particular, uncertainties in the position of the pulsar being searched, for orbital motion of the pulsar in a binary system, proper motion of the pulsar (e.g., a transverse velocity of $150\ \mathrm{km\ s^{-1}}$ at 100 pc), or unpredicted changes in the period (anything larger than an accumulated fractional change $\Delta f/f$ of $10^{-10}(f/1\ \mathrm{kHz})^{-1}$) will all require special techniques to compensate for the way they spread the frequency out over more than the frequency resolution of the observation.

**(iii) The technology of performing long Fourier transforms**

We shall see that there are several different strategies one can adopt to search for pulsars, whether known ahead of time or not, but all of them can involve performing Fourier transforms of large data sets. It will help us compare the efficiencies of different strategies if we first look at how this might be done.

If one imagines that the observation lasts $10^7$ s with a sampling rate of 1 kHz, then one must perform an FFT with roughly $10^{10}$ data points. This requires roughly $3N \log_2 N$ operations for $N = 2^{34} = 1.7 \times 10^{10}$. This evaluates to $1.7 \times 10^{12}$ operations per FFT. Given the 50 Mflops computer we required earlier for filtering for coalescing binaries, this would take about 10 hours. This is not unreasonable: over 200 FFTs could be computed in the time it took to do the observation.

The real difficulty with this is the memory requirement: FFT algorithms require access to the whole data set at once. To achieve these processing speeds, the whole data set would have to be held in fast memory, all 20 Gbytes of it. Unless there is a revolution in fast memory technology, it does not seem likely that this will be possible, at least not at an affordable level. One could imagine being able to store the data on a couple of 10-Gbyte read/write optical discs, and then using a mass-store-FFT algorithm, which uses clever paging of data in and out of store. This would still be very slow, but its exact speed would depend on the computer system.

One method of calculating the Fourier transform would be to split the data set up into $M$ chunks of length $L$, each chunk being small enough to fit into core. by performing FFTs on data sets of length $L$ it is possible to calculate the contribution of each subset to the total transform. It is not hard to show that the work needed to construct the full transform from these individual sets is about $M$ times the work needed to do it as a single set (see, e.g., Hocking, 1989). With a memory limit of 200 Mbytes and a machine capable of 50 Mflops, it might be possible to do one or two Fourier transforms in the time it takes to do the observation. With the same memory in a machine capable of 1 Gflop, one could do 40 Fourier transforms in the same time. These are big numbers for memory and performance, but they may be within reach of the interferometer projects by the time they go on-line. The numbers become even more tractable if we are looking for a pulsar under 100 Hz: with a data rate of only 100 Hz, say, the work for a given number $M$ of subsets goes down by a factor of about 11. It is clear that it is possible to trade-off memory against CPU speed; the technology of the time will dictate how this trade-off is to be made.

If it proves impossible to compute the full transform exactly, there are approximate methods available, such as to subdivide the full set into $M$ subsets as above, but then only to compute the power spectrum of each subset and to add the power spectra together. This reduces the frequency resolution by a factor of $M$, with a proportionate decrease in the spatial resolution and in the number of different positions that an observation might need to search. It also reduces the signal-to-noise ratio of the observation. it is likely that techniques developed for radio pulsar searches (Lyne, 1989) will be useful here as well.

**(iv) Detecting known pulsars**

The earliest example of using a wide-band detector to search for a known pulsar is the experiment of Hough *et al.* (1983), which set an upper limit of $h < 8 \times 10^{-21}$ on radiation from the millisecond pulsar, PSR 1937 + 214. Future interferometers could better this limit by many orders of magnitude, but they will have to do long observing runs (some $10^7$ s) to achieve maximum sensitivity. The analysis of the vast amount of data such experiments will generate poses greater problems for analysis than those we addressed for coalescing binaries.

Let us assume that we know the location and frequency of a pulsar, and we wish to detect its radiation. We need to make a correction for the Doppler effects from the known position, or from several contiguous positions if the position is not known accurately enough ahead of time. One might be tempted to approach this problem by filtering, as for coalescing binaries. But because of the computational demands, this is not the best method. Much better is a numerical version of the standard radio technique called *heterodyning**, followed by stroboscopic sampling.

**Difficulties with filtering for pulsars** Let us consider first why filtering is unsuitable. In this context a filter is just a sinusoidal signal Doppler-shifted to give the expected arrival time of any phase at our detector. If only one rotational motion of our detectors were present, and if the observation were to last several rotation periods, then only points separated in the polar direction would need separate filters: points separated in azimuth have waveforms that are simply shifted in time relative to one another, and so correlating the data in time with only one filter would take care of all such points. This might be useful even for a pulsar of known position, since it might not be known to the accuracy of equations (16.40) and (16.41).

However, our detectors participate in at least *three* rotational motions about different centres, and the observations will probably last only a fraction of a period of the most demanding motion, the solar orbital one. This means that filters lose one of their principal advantages: searching whole data sets for similar signals arriving at different times.

Filtering requires that at least three FFTs of long data sets must be performed: of the filter, of the sampled data, and of their product to find the correlation.

* I am indebted to Jim Hough and Harry Ward for suggesting this method. The details in this section are based on conversations with them and with Norman MacKenzie, Tim Niebauer, and Roland Schilling.

Even for a well-known source, there will have to be several filters, because the phase of the wave as it arrives will not be predictable, nor will its polarization. The phase of the wave depends on exactly where the radiating 'lump' on the pulsar is. A given detector will respond to the two independent polarizations differently as it moves in orbit around the Sun; the polarization will generally be elliptical, but the proportion of the two independent polarizations and the orientation of the spin axis are unknown. Each of these variables must be filtered for, and each filter needs two more FFTs (the data set needs to be transformed only once). If the source's position and/or frequency are not known accurately, then even more filters will be required, each adding two further FFTs. Given the problems we saw we might have with FFTs, this could be a costly procedure.

**Heterodyne detection** Suppose the frequency of the pulsar is $f_p$ in the barycentric frame (Solar System rest frame). Then Doppler effects of the Earth's motion plus uncertainties in the pulsar's frequency and its rate of change will require us to look in a narrow range of frequencies $(f_0, f_0 + \Delta f)$ containing $f_p$. The idea underlying heterodyning is that if the data contain a sinusoidal signal of frequency $f$,

$$s(t) = \sin(ft + \phi),$$

where $\phi$ is a possible phase, then if we *multiply* the signal by a 'carrier' sinusoid of frequency $f_c$ in the bandwidth, the result can be written as

$$\sin(f_c t)s(t) = \frac{1}{2}\cos[(f - f_c)t + \phi] + \frac{1}{2}\cos[(f + f_c)t + \phi].$$

We may choose $f_c$ so that the difference frequency $f - f_c$ is within a bandwidth $\Delta f$ about zero, and yet it contains all the information (amplitude and phase) of the original signal. By filtering the resultant data set down to that bandwidth about the origin, and then re-sampling it at its (now much lower) Nyquist frequency, one can produce a data set containing many fewer points that will still contain all the information in the original band of frequencies. This set will be easier to apply Fourier transforms to than the original.

The saving in size is of order $\Delta f / f$, or $1 \times 10^{-4}$ for the Doppler broadening due to the Earth's orbital motion. This would reduce the typical data set discussed in the previous section down from $10^{10}$ points to $10^6$. This is of a size that can reasonably be handled on our 50 Mflops computer: an FFT can be done in a matter of seconds, so that complicated filtering and searches for signals become practical without expensive computing machinery.

When one looks at the details of how to implement heterodyning, one has to worry about how the noise is affected and how the procedure can be done with minimum cost. Much more work needs to be done on this question, but two possible implementations might be as follows. The first step in both is to filter the data stream with a narrow band-pass filter that allows only the required bandwidth through. This is to ensure that subsequent steps do not introduce noise (or signals) from other regions of the spectrum into our bandwidth.

In the first implementation, the next step would be to multiply by the heterodyne carrier with frequency $f_c = f_0$, i.e. at the lower edge of the bandwidth. This will ensure that noise from outside the bandwidth is not heterodyned. This allows the band-pass filter to be imperfect, as it must be if it is not to involve prohibitive amounts of computing: it will perhaps need to fall off by a factor of ten within a distance of $\Delta f/2$ of the edges of the band. Then a low-pass filter needs to be applied to get rid of the *sum* frequencies $f_c + f$. The resulting data set is still running at the rate of 10 kHz or so, but all we want is a narrow band, perhaps less than 1 Hz, about zero frequency. By *stroboscopically sampling* (defined earlier) this set at a rate equal to the appropriate Nyquist frequency ($2\Delta f$) in the barycentric frame for signals arriving from the pulsar's direction, one can produce a data set that is at once small and Doppler-corrected. This sampling involves accepting only one point in every $10^4$ or so.

The alternative implementation, which might be even faster, is based on a suggestion of Norman Mackenzie. This is to apply stroboscopic sampling (at a slow rate $f_s$ near the Nyquist rate) directly to the data set after it has been put through the band-pass filter but before heterodyning. This may be thought of as heterodyning by aliasing: what appears in the low-frequency spectrum of the sampled data set is the aliased signal. The aliasing condition is that an original frequency $f$ will appear in the sampled set at a frequency $f - nf_s$, where $n$ is an integer. By choosing $n$ and $f_s$ appropriately, it should be possible to alias the required range of frequencies into a range near zero, without introducing extraneous noise. If the sampling is done at a rate equal to the phase arrival rate for a constant frequency in the barycentric frame at the pulsar's position, it will make all the necessary Doppler corrections automatically. Because this is potentially a very fast method, it deserves more study.

Further refinements can be made. For instance, in the first heterodyning implementation, one should multiply independently by two carriers 90 degrees apart in phase, and then add the resultant difference signals with a similar 90 degree phase shift. This reinforces the signal but adds the two independent quadratures of noise together incoherently, so that the noise is reduced by $\sqrt{2}$ relative to the signal.

Moreover, once a 'slow' data set (near zero frequency) is produced, it may still be necessary to do quite a lot of work on it to extract a pulsar signal. One will have to correct for uncertainties in the pulsar position (and hence in the stroboscopic sampling rate), for changes in the pulsar's intrinsic frequency during the observation period, for possible proper motion or binary motion effects, for the changing orientation of the detector relative to the pulsar direction and so on.

However, regardless of which of the two types of heterodyning implementations turns out to be best, the general principle is clear: if we are only interested in a bandwidth $\Delta f$ about a frequency $f_p$, then we should be able to deal with a data set sampled at an effective rate $2\Delta f$ rather than $2f_p$. The resultant savings in computing effort make it possible to contemplate on-line searches for a few selected pulsars with computing resources that are no larger than are needed for filtering for coalescing binaries.

**(v) Searching for unknown pulsars**

One of the most interesting and important observations that interferometers could make is to discover old nearby pulsars or other continuous wave sources. There may be thousands of spinning neutron stars – old dead pulsars – for each currently active one. The nearest may be only tens of parsecs away. But we would have to conduct an all-sky, all-frequency search to find them. We shall see in this section that the sensitivity we can achieve in such a search is limited by computer technology.

The central problem is the number of independent points on the sky that have to be searched. As we saw in equation (16.39), the angular resolution increases as the square of the observing time, so the number of patches on the sky increases as the fourth power. For observations longer than 20 hours, equation (16.41) implies

$$N_{\text{patches}} = 4\pi/(\Delta\theta)^2 = 1.3 \times 10^{13}\left(\frac{f}{1\ \text{kHz}}\right)^2\left(\frac{T_{\text{obs}}}{10^7\ \text{s}}\right)^4. \qquad (16.42)$$

We will now look at what seems to me to be the most efficient method of searching these patches.

**The barycentric Fourier transform** The signal from a simple pulsar (i.e. one that does not have added complications like a binary orbit, a rapid spindown, or a large proper motion) would stand out as a strong peak if we were to compute its Fourier transform with respect to the time-of-arrival of the waves at the barycentre of the Solar System, which we take to be a convenient inertial frame. In this section I shall look at the relationship between this transform and the raw-data transform with respect to time at the detector, which relationship depends on the direction we assume for the pulsar. I also look at the relationship between the barycentric transforms of the same signal on two different assumptions for the pulsar position.

We shall need some notation. Let $t_d$ be the time that a given part of the pulsar signal arrives at the detector. Let $t_b(\theta, \phi, t_d)$ be the time that the same signal would arrive at the barycentre if it comes from a pulsar at angular position $(\theta, \phi)$. Let $s_d(t_d)$ be the signal itself at the detector and $s_b(t_b)$ the signal at the barycentre. Note that

$$s_b[t_b(\theta, \phi, t_d)] = s_d(t_d),$$

by definition. The relation between the two timescales is given by

$$t_b = t_d + k(\theta, \phi, t_d), \qquad (16.43)$$

where the function $k$ is slowly varying in time for our problem,

$$\left|\frac{\partial k}{\partial t_{\mathrm{d}}}\right| \ll 1$$

due to the slow velocities that the Earth participates in. The inverse of equation (16.43) is

$$t_{\mathrm{d}} = t_{\mathrm{b}} + g(\theta, \phi, t_{\mathrm{b}}) \tag{16.44}$$

Again the derivative of $g$ is small. From the definition it is evident that

$$g(\theta, \phi, t_{\mathrm{b}}) = -k[\theta, \phi, t_{\mathrm{b}} + g(\theta, \phi, t_{\mathrm{b}})]. \tag{16.45}$$

The exact forms of the functions $g$ and $k$ are complicated, but they need not concern us here.

Now we wish to find the relation between the Fourier transform of $s_{\mathrm{b}}$ and that of $s_{\mathrm{a}}$ with respect to their respective local times. For a given set of detector data, we have

$$\tilde{s}_{\mathrm{b}}(f_{\mathrm{b}}, \theta, \phi) = \int_{-\infty}^{\infty} s_{\mathrm{b}}[t_{\mathrm{b}}(\theta, \phi)]\mathrm{e}^{-2\pi i f_{\mathrm{b}} t_{\mathrm{b}}}\, \mathrm{d}t_{\mathrm{b}},$$

$$= \int_{-\infty}^{\infty} s_{\mathrm{d}}(t_{\mathrm{d}})\mathrm{e}^{-2\pi i f_{\mathrm{b}} t_{\mathrm{b}}}\, \mathrm{d}t_{\mathrm{b}}, \tag{16.46}$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \tilde{s}_{\mathrm{d}}(f_{\mathrm{d}})\mathrm{e}^{2\pi i f_{\mathrm{d}} t_{\mathrm{d}}}\, \mathrm{d}f_{\mathrm{d}}\right]\mathrm{e}^{-2\pi i f_{\mathrm{b}} t_{\mathrm{b}}}\, \mathrm{d}t_{\mathrm{b}},$$

$$= \int_{-\infty}^{\infty} \tilde{s}_{\mathrm{d}}(f_{\mathrm{d}})m(\theta, \phi, f_{\mathrm{d}}, f_{\mathrm{b}})\, \mathrm{d}f_{\mathrm{d}}, \tag{16.47}$$

where we define

$$m(\theta, \phi, f_{\mathrm{d}}, f_{\mathrm{b}}) = \int_{-\infty}^{\infty} \mathrm{e}^{2\pi i f_{\mathrm{d}} t_{\mathrm{d}}(t_{\mathrm{b}})}\, \mathrm{e}^{-2\pi i f_{\mathrm{b}} t_{\mathrm{b}}}\, \mathrm{d}t_{\mathrm{b}}. \tag{16.48}$$

The inverse of this relation is obtained by a simple permutation of indices:

$$\tilde{s}_{\mathrm{d}}(f_{\mathrm{d}}) = \int_{-\infty}^{\infty} \tilde{s}_{\mathrm{b}}(f_{\mathrm{b}})n(\theta, \phi, f_{\mathrm{d}}, f_{\mathrm{b}})\, \mathrm{d}f_{\mathrm{b}}, \tag{16.49}$$

where the kernel here is

$$n(\theta, \phi, f_{\mathrm{d}}, f_{\mathrm{b}}) = \int_{-\infty}^{\infty} \mathrm{e}^{2\pi i f_{\mathrm{b}} t_{\mathrm{b}}(t_{\mathrm{d}})}\, \mathrm{e}^{-2\pi i f_{\mathrm{d}} t_{\mathrm{d}}}\, \mathrm{d}t_{\mathrm{d}}. \tag{16.50}$$

These equations allow us to find the barycentric transform from the detector transform, and vice versa. In principle, by applying equation (16.47) to the Fourier transform of the detector data one produces a transform in which the signal from a pulsar at a given position should stand out much more strongly. In practice, if one only wants to do this for a few cases, it is much more efficient to use stroboscopic sampling, which effectively computes equation (16.46) by selecting the appropriate values of the integrand. However, when searching the whole sky for pulsars this would involve more work than the method of the next section.

**Barycentric transforms for nearby locations** If one has computed the barycentric transform $\tilde{s}_b$ for some location on the sky, the quickest way to find the transform for a nearby location is to find a direct transformation of $\tilde{s}_b$, rather than to start again with $s_d$ or $\tilde{s}_d$. In this manner one can compute $\tilde{s}_b$ for one location and then 'step' around the sky from there. We derive in this section the appropriate equations.

Consider two locations $(\theta, \phi)$ and $(\theta', \phi')$. We want $\tilde{s}_b$ at $(\theta', \phi')$ in terms of that at $(\theta, \phi)$. From equations (16.47) and (16.49) we have

$$\begin{aligned}\tilde{s}_b(f'_b, \theta', \phi') &= \int_{-\infty}^{\infty} \tilde{s}_d(f_d) m(\theta', \phi', f_d, f'_b)\, df_d, \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{s}_b(f_b, \theta, \phi) n(\theta, \phi, f_d, f_b) m(\theta', \phi', f_d, f'_b)\, df_d\, df_b, \\ &= \int_{-\infty}^{\infty} \tilde{s}_b(f_b, \theta, \phi) q(\theta', \phi', f'_b; \theta, \phi, f_b)\, df_b, \end{aligned} \qquad (16.51)$$

where we define the 'stepping' kernel $q$ by

$$q(\theta', \phi', f'; \theta, \phi, f) = \int_{-\infty}^{\infty} m(\theta', \phi', f'', f') n(\theta, \phi, f'', f)\, df''. \qquad (16.52)$$

If $(\theta', \phi')$ is close to $(\theta, \phi)$, then the kernel $q$ should be sharply peaked in frequency near $f = f'$. In fact, it is easy to show from the inverse properties that

$$q(\theta, \phi, f'; \theta, \phi, f) = \delta(f - f') \quad \forall \theta, \phi.$$

The peaking of this function is in fact the mathematically precise way of doing the calculation we did roughly earlier, namely seeing how many independent patches on the sky one would have to search. Two angles are independent if $q$ is wider than the frequency resolution of the observation.

**The stepping method** The way to do an all-sky search uses in fact the converse of the last statement. In order to convert the barycentric Fourier transform for a source at one position to that at another, one must do the integral given in equation (16.51). If the two positions are adjacent patches on the sky, then by definition the function $q$ will be only (at least on average) two frequency bins wide, so that one can produce the barycentric transform for the second patch from that for the first by a calculation taking of order $N$ operations, where $N$ is the number of data points. This can represent a significant saving over doing stroboscopic sampling and an FFT for each patch. This is particularly true for large data sets that exceed the core memory capacity of the computer, because FFT algorithms on such data sets will be very much slower. The present method does not suffer from this drawback because – after the first barycentric transform has been computed – it does not require the whole transform to be held in memory at once. I shall refer to this method as *stepping around the sky*.

**Depth of a search as a function of computing power** We can now assemble what we know and make an assessment of the computing power required to make a search of a given sensitivity, at least by the method of stepping described here. From equation (16.42) the number of patches on the sky is

$$N_{\text{patches}} = 1.3 \times 10^{13}\left(\frac{f}{1\ \text{kHz}}\right)^2\left(\frac{T_{\text{obs}}}{10^7\ \text{s}}\right)^4.$$

The data set will have a length

$$N_{\text{pts}} = 2 \times 10^{10}\left(\frac{f}{1\ \text{kHz}}\right)\left(\frac{T_{\text{obs}}}{10^7\ \text{s}}\right) \text{ points},$$

provided we interpret $f$ as the highest observable frequency, so we sample at a rate $2f$. If the stepping operation between adjacent patches requires ten real floating-point operations per data point, then we need to perform

$$N_{\text{fl-pt ops}} = 2.5 \times 10^{24}\left(\frac{f}{1\ \text{kHz}}\right)^3\left(\frac{T_{\text{obs}}}{10^7\ \text{s}}\right)^5$$

floating-point operations to search the whole sky.

In order to do repeatable searches, it must be possible to analyse the data in roughly the time it takes to take it. If the computer speed is called $\mathcal{S}$, measured in floating-point operations performed per second, then the time to perform $N_{\text{fl-pt ops}}$ operations is $N_{\text{fl-pt ops}}/\mathcal{S}$ s. Ignoring overheads due to other factors, we therefore find that the time to analyse the data is

$$T_{\text{anal}} = 2.5 \times 10^{16}\left(\frac{f}{1\ \text{kHz}}\right)^3\left(\frac{T_{\text{obs}}}{10^7\ \text{s}}\right)^5\left(\frac{\mathcal{S}}{100\ \text{Mflops}}\right)^{-1} \text{s}.$$

By equating $T_{\text{anal}}$ and $T_{\text{obs}}$, we obtain the maximum observation time allowed by a computer of a given speed:

$$T_{\text{max}} = 4.4 \times 10^4\left(\frac{f}{1\ \text{kHz}}\right)^{-3/4}\left(\frac{\mathcal{S}}{100\ \text{Mflops}}\right)^{1/4} \text{s}. \tag{16.53}$$

This is about 12 hours for a 100 Mflops computer analysing data for millisecond pulsars (up to 1 kHz). If we lower our sights and try to search for pulsars under 100 Hz (still very interesting), we can run for about three days. Another improvement comes from making a narrow-band search. This is attractive anyway, since narrow-banding enhances the detector's sensitivity in the bandwidth. In a narrow-band search one would use heterodyning to reduce the size of the data set. For a bandwidth $B$, the analogue of equation (16.53) is

$$T_{\max} = 2.1 \times 10^5 \left(\frac{f}{1\ \mathrm{kHz}}\right)^{-1/2} \left(\frac{B}{2\ \mathrm{Hz}}\right)^{-1/4} \left(\frac{\mathcal{S}}{100\ \mathrm{Mflops}}\right)^{1/4} \mathrm{s}. \tag{16.54}$$

This is better, but still permits only about 2.4 days of observing in a narrow bandwidth at 1 kHz.

The actual figures given here may change with the invention of more efficient algorithms, but what is not likely to change is that the minimum number of operations per patch on the sky scales linearly with the number of data points. This means in turn that the permissible observation time will grow only as the fourth root of the computer speed. Even worse, since the sensitivity one can reach in $h$ scales as the square root of the observation time, the limits on $h$ will scale as the *eighth* root of the computer speed! Changing from a desktop computer capable of 0.1 Mflops to a supercomputer capable of 10 Gflops improves one's limits on $h$ by only a factor of four.

This is the central problem of the all-sky search for pulsars: it is quite possible to run detectors for several months gathering data, and this will probably be done to search for known pulsars, but computing power limits any all-sky, all-frequency search for unknown pulsars to periods of the order of days.

## 16.3 Combining lists of candidate events from different detectors

Until now I have kept the discussion to the analysis of one detector's data, but it is clear that for the best signal-to-noise ratio and for the extraction of complete astrophysical information, detectors must operate in coincidence. I will consider in this section the simplest method of coordinated observation: exchanging lists of events detected in individual detectors. I have elsewhere (Schutz, 1989) called this the 'threshold mode' of network data analysis, because each detector's criterion for an 'event' is that its amplitude crosses a pre-set threshold.

### 16.3.1 Threshold mode of data analysis

We have seen in section 16.2.1(ii) how the thresholds can be determined. Once events have been identified by the on-line computer – either in the time-series of data directly or by filtering – it is important that the data from these events be brought together and analysed as quickly as possible. If the event is a supernova, we have considerably less than a day before it might become bright enough to be seen optically, and optical astronomers need to be told of it as quickly as possible. If the event is a coalescing binary, there may be even more urgency: the absence of an envelope around a neutron star means that any radiation emitted may come out with much less delay than in a supernova. Since we know so little about what such events look like, it would be valuable to have optical telescopes and orbiting X-ray telescopes observe the region of the event as quickly as possible.

The rapid exchange of data is certainly possible: with modern computer networks, it would be easy to arrange that the on-line computers could automatically circulate lists of events and associated data periodically, such as every hour. We should bear in mind that, if the threshold is set so that a network would have a four-way false alarm only once per year at a data rate of 1 kHz, then each detector will see a spurious noise-generated event three times per second! It will be impossible to distinguish the real events from the spurious until the lists of events from the various detectors are compared. The initial lists need not contain much data, so links over the usual data networks will be fast enough at this stage.

What sort of data must be exchanged? If the event is seen in a filter, the list should include the amplitude of the event, the parameters of the best-fit filter, and an agreed measure of the time the signal arrived at the detector (such as when a coalescing binary signal reached some fiducial frequency, e.g. 100 Hz). It will probably also be necessary to include calibration data, as the sensitivity of interferometers will probably change from time to time. If the signal has a high signal-to-noise ratio, then it may be desirable to include other information, such as its correlation amplitude with other filters, or even the raw unfiltered data containing the signal. The feasibility of this will depend upon the bandwidth of available communication channels.

If the event is a broad-band burst seen in the time-series, then it will be even more important to exchange the raw data, along with timing and calibration information. If raw-data exchange is impossible, then at least some description of the event will be needed, such as when it first crossed the threshold, when it reaches its maximum, and when it went below threshold.

Once likely coincidences among detectors have been identified, it will then be useful to request the on-line computers to send out more detailed information about the selected candidate coincidences. Since these requests will be rarer, it will not overburden the communications networks to exchange raw data and more complete calibration information for the times in question. If the events then still seem significant, they should be broadcast to other astronomers and analysed more thoroughly at leisure.

### 16.3.2 Deciding that a gravitational wave has been detected

The question that underlies all of the present article is, how do we decide that a gravitational wave has actually arrived? Various of our topics, such as the construction and use of filters and the setting of appropriate thresholds, are important components of such a decision. What we want to stress in this section is that the laser interferometer community must make sure that it has well-defined criteria for accepting a gravitational wave event as real, and a well-defined procedure for modifying and updating these criteria, *before* it begins observing in earnest.

The first detection of a gravitational wave will be such a momentous event that – if it occurs in an interferometer network – those who operate the network should leave no room for doubt that the event was well above the threshold expected of known noise sources during the time of observation. If criteria are established ahead of time, there can be no question that they have been 'adapted' to the data; conversely, if criteria for gravitational wave detection are formulated after looking at the data, there is always doubt that the events that are then identified really have the significance that might be claimed for them.

In this connection, one should not naively believe that because an unexpected event has a signal-to-noise ratio that would give it a small probability $p$ of arising by chance, then that automatically means that the probability of its being real is correspondingly high. It is very hard to make an accurate calculation of $p$, since it involves not only the modelled noise but also unmodelled noise and even the circumstances of an experiment. Some Bayesian-type criterion, which involves an *a priori* estimate of the probability that the candidate event would be real, should also be used in such circumstances.

This is not to say that there should be no criteria for accepting unexpected events or unpredicted waveforms. Provided the signal-to-noise ratio of such events is high enough and they have been processed in the same way as all other data, there should be no problem accepting them. But when the signal-to-noise ratio is relatively small and/or the data have been processed in a way that had not been agreed ahead of time, there is considerable danger of accepting false alarms as real.

What can and should be done, however, if unusual events with marginal signal-to-noise ratio are seen, is that new criteria can be adopted to look for them in subsequent data. If they continue to turn up – or if re-analysis of archived data show them – then they can be accepted as real. Similarly, if new theoretical models of gravitational wave sources are evolved, they can be incorporated into the criteria. But the community should not claim detections before this second stage of verification. In particular, if there are marginal and unexpected one-off events apparently associated with rare astronomical phenomena, then it may not be possible to call them real until they have been seen again, however long that may take.

## 16.4 Using cross-correlation to discover unpredicted sources

The threshold mode of analysis is unsuitable for some sources, such as continuous waves or weak events that we have not predicted well enough ahead of time to construct filters for. In these cases, the 'correlation mode' is appropriate: using cross-correlations between the data streams of different detectors.

Cross-correlation has its own problems, however: its signal-to-noise relations are rather different from filtering, and the different polarizations of different detectors mean that signals in two different detectors from the same gravitational wave may not exactly correlate. In the next section I will give a general discussion of cross-correlation, addressing the behaviour of noise and assuming that the two data streams contain the same signal. One solution to the problem of polarization has been given by Gursel and Tinto (1989). Their approach will be discussed in section 16.4.2.

### 16.4.1 The mathematics of cross-correlation: enhancing unexpected signals

It is useful to think of cross-correlation as the use of one data stream as a filter to find things in the other data stream. Thus, if the first stream contains a signal that hasn't been predicted, one can still find it in the second. If we adopt this point of view, then we must face two important differences between matched filtering and cross-correlation as a means of enhancing signal-to-noise ratios. These are:

(1) The 'filter' is *noisy*. In fact, in the case of most interest, the signal is below the broad-band noise and the power in the filter is dominated by the noise. If we really had an instrument with an infinite bandwidth, then the noise power would be infinite and we would never see the signal. In practice, we will see below that we must filter the data down to a finite bandwidth before performing the correlation in order to achieve an acceptable signal-to-noise ratio.

(2) The 'filter' also contains the signal we wish to find, of course, but the amplitude of this part of the filter is not known *a priori*: it is the amplitude of the incoming signal. This means that if the incoming signal is reduced by half, the response of the filter to it will go down by a factor of *four*. We shall see that this leads to the biggest difference between matched filtering and cross-correlation when they are applied to long wavetrains: the enhancement of signal-to-noise in cross-correlation increases only as the fourth root of the observing time or the number of cycles in the signal, not as the square root we found in equation (16.30).

If we have two data streams $o_1$ and $o_2$ containing the same signal $h$ but independent noise amplitudes $n_1$ and $n_2$,

$$o_1(t) = h(t) + n_1(t), \qquad o_2(t) = h(t) + n_2(t), \tag{16.55}$$

their cross-correlation is

$$o_1 \circ o_2 = h \circ h + n_1 \circ h + h \circ n_2 + n_1 \circ n_2. \tag{16.56}$$

The 'signal' is the expectation of this (averaged over both noise amplitudes), which is just $h \circ h$. The variance of the correlation, however, is a problem. The final term contributes

$$\begin{aligned}\langle |n_1 \circ n_2|^2 \rangle &= \left\langle \int \tilde{n}_1(f)\tilde{n}_1^*(f')\tilde{n}_2^*(f)\tilde{n}_2(f')\mathrm{e}^{2\pi\mathrm{i}(f-f')t}\,\mathrm{d}f\,\mathrm{d}f' \right\rangle \\ &= \int S_1(f)S_2(f)\delta(f-f')\delta(f-f')\mathrm{e}^{2\pi\mathrm{i}(f-f')t}\,\mathrm{d}f\,\mathrm{d}f'.\end{aligned}$$

The presence of *two* delta functions in the integrand makes this expression infinite: if we allow all the noise in the detectors to be cross-correlated, then the variance of the correlation will swamp the signal. The solution is (i) to *filter* the output down to a suitable bandwidth $B$ before correlating, and (ii) to perform the correlation only over a finite stretch of data lasting a time $T$. If we use a superscript F to denote the filtered version of a quantity, then the analogue of $n_1 \circ n_2$ is

$$I_{12}(t) = \int_0^T n_1^{\mathrm{F}}(t') n_2^{\mathrm{F}*}(t' + t)\, \mathrm{d}t'. \tag{16.57}$$

Its variance is

$$\langle |I_{12}(t)|^2 \rangle = \int_0^T \int_0^T \langle n_1^{\mathrm{F}}(y) n_1^{\mathrm{F}*}(y') n_2^{\mathrm{F}}(y + t) n_2^{\mathrm{F}*}(y' + t) \rangle \, \mathrm{d}y \, \mathrm{d}y'. \tag{16.58}$$

The key to evaluating this is the expectation

$$\langle n_1^{\mathrm{F}}(t) n_1^{\mathrm{F}}(t') \rangle = 2 \int_{f_1}^{f_2} S_1(f) \cos[2\pi f(t - t')]\, \mathrm{d}f, \tag{16.59}$$

where $f_1$ and $f_2$ are the lower and upper limits of the filtered frequency band ($f_2 = f_1 + B$), and where the factor of two arises because negative frequencies must be included in the filtered data as well as positive ones. It is a straightforward calculation to show that, assuming for simplicity that $S_i(f)$ has the constant value $\sigma_{if}^2$ over the bandwidth, then for the most important case $2\pi f_1 T \gg 1$ and $2\pi BT \gg 1$,

$$\langle |I_{12}(t)|^2 \rangle \approx 2\sigma_{1f}^2 \sigma_{2f}^2 BT. \tag{16.60}$$

This part of the noise is proportional to the bandwidth of the data and the duration of the correlation. The duration will usually be chosen so that the above conditions on $B$ and $T$ are satisfied, for otherwise the experiment would be too brief to detect any signal that fits within the bandwidth $B$. The remaining contributions to the variance of the cross-correlation come from the second and third terms of equation (16.56) (strictly, from their filtered and finite-time analogues). These are just like equation (16.19), and add to equation (16.60) a term equal to $(\sigma_{1f}^2 + \sigma_{2f}^2) \int_0^T |h^{\mathrm{F}}(t)|^2 \, \mathrm{d}t$.

The case of most interest to us is where the 'raw' signal $h^{\mathrm{F}}(t)$ is smaller than the time-series noise in the bandwidth $B$ in each detector, $n_i^{\mathrm{F}}(t)$. Then the variance is dominated by equation (16.60) and we have the following expression for the signal-to-noise ratio of the cross-correlation:

$$\frac{\text{correlation signal}}{\text{correlation noise}} = \frac{\int_0^T |h^{\mathrm{F}}(t)|^2 \, \mathrm{d}t}{|2\sigma_{1f}^2 \sigma_{2f}^2 BT|^{1/2}}. \tag{16.61}$$

This has considerable resemblance to the filtering signal-to-noise ratio given in equation (16.20), and this justifies and makes precise our notion that cross-correlation can be thought of as using a noisy data stream as the filter. To convert equation (16.20) into equation (16.61), we must (i) replace the filter in the numerator with the signal $h^{F}$ that is in the noisy 'filter', and (ii) replace the filter power in the denominator with the noise power of the noise filter, since we have assumed this power is the largest contributor to the noise.

However, equation (16.61) does not give us the signal-to-noise ratio for the gravitational wave signal, since its numerator is proportional to the *square* of the wave amplitude. This is the effect that we noted at the beginning of this section, that the 'filter' amplitude is proportional to the signal amplitude. A better measure of the amplitude signal-to-noise ratio is the square root of the expression in equation (16.61):

$$\frac{S}{N}=\frac{\left[\int_{0}^{T}|h^{F}(t)|^{2}\,\mathrm{d}t\right]^{1/2}}{[2\sigma_{1f}^{2}\sigma_{2f}^{2}BT]^{1/4}}. \tag{16.62}$$

There are two cases to consider here: long wavetrains and short pulses.

**(i) Long wavetrains**

The best signal-to-noise is achieved if we match the observation time $T$ to the duration of the signal or, in the case of pulsars, make $T$ as long as possible. Let us assume for simplicity that the two detectors have the same noise amplitude, and let us denote by $R$ the 'raw' signal-to-noise ratio of the signal (its amplitude relative to the full detector noise in the bandwidth $B$),

$$R=\frac{h}{(2B\sigma_{f}^{2})^{1/2}}.$$

Then we find

$$\frac{S}{N}\approx\left(\frac{1}{2}BT\right)^{1/4}R. \tag{16.63}$$

The signal-to-noise ratio increases only as the fourth root of the observation time. If we are looking at, say, the spindown of a newly formed pulsar, lasting 1 s, and we filter to a bandwidth of 1 kHz because we don't know where to look for the signal, then the enhancement factor $(BT/2)^{1/4}$ is about five: short wavetrains are improved, but not dramatically. If we are looking at a pulsar, again in a broad-band search with 1 kHz bandwidth, but in an observation lasting $10^{7}$ s, then the enhancement of signal-to-noise is a factor of about 250. This enhancement could be achieved by the $T_{\mathrm{obs}}^{1/2}$ effect in a single-detector observation lasting only three minutes, for which the data could be trivially analysed. If the single detector is narrow-banded, the time would be even less. Therefore, cross-correlation is not a good way of finding pulsars.

There are other differences between filtering and cross-correlation. Since for signals below the broad-band noise ($R<1$), we do not know where the signal is in the data stream used as a filter, it follows that we cannot determine the time-of-arrival of the signal from the correlation, apart from a relatively crude determination based upon the presence or absence of correlations between given data sets of length $T$. The correlation also does not tell us the waveform and therefore it cannot determine the true amplitude of the signal. It can, however, determine the time-delays between the arrival of brief events at different detectors.

**(ii) Short pulses**

Here one would set the bandwidth $B$ equal to that of the pulse; if the pulse has duration roughly $T = 1/B$, and if again the two detectors have the same noise amplitude, then equation (16.62) gives a signal-to-noise ratio that is a factor of roughly $2^{1/4} \approx 1.2$ smaller than the optimum that filtering can achieve. For $TB \approx 1$ our approximations are breaking down, but it is reasonable that using this noisy filter would reduce the signal-to-noise by a factor of order two. Since in this case filtering does not enhance the signal-to-noise ratio, neither does cross-correlation: if a pulse is too weak to be seen above the broad-band (bandwidth $B$) noise in one detector, if will not be found by cross-correlation.

### 16.4.2 Cross-correlating differently polarized detectors

A more sophisticated approach to correlation has been devised by Gursel and Tinto (1989) in their approach to the signal-reconstruction problem, which I will describe in detail in section 16.5 below. It works if there are at least three detectors in the network. I shall neglect noise for simplicity in describing the method. If we let $\theta$ and $\phi$ be the angles describing the position of the source on the sky and we use $\alpha_i$, $\beta_i$, and $\chi_i$ to represent the latitude, longitude, and orientation of the $i$th detector, respectively, and if we have some definition of polarization of the waves so that we can describe any wave by its amplitudes $h_+$ and $h_\times$, then the response $r = \delta l/l$ of the $i$th detector is a function of the form

$$r_i(t) = E_{+i}(\theta, \phi, \alpha_i, \beta_i, \chi_i)h_+[t - \tau_i(\theta, \phi)] + E_{\times i}(\theta, \phi, \alpha_i, \beta_i, \chi_i)h_\times[t - \tau_i(\theta, \phi)], \quad (16.64)$$

where $\tau_i(\theta, \phi)$ is the time-delay between receiving a wave coming from the direction $(\theta, \phi)$ at some standard location and at the position of the detector. We shall define the 'standard location' by setting $\tau_1 = 0$. We need not be concerned here with the precise form of the functions $E_{+i}$, $E_{\times i}$, and $\tau_i$, nor with the exact definitions of the various angles.

The response equations of the first two detectors may be solved for $h_+$ and $h_\times$ and substituted into the response equation for the third to predict its response, for an assumed direction to the source. Let this prediction be $r_{3\text{-pred}}$:

$$r_{3\text{-pred}}(t) = -[D_{23}r_1(t - \tau_3) + D_{31}r_2(t + \tau_2 - \tau_3)]/D_{12}, \quad (16.65)$$

where $D_{ij}$ is the determinant

$$D_{ij} = E_{+i}E_{\times j} - E_{\times i}E_{+j}.$$

If there were no noise in the detectors, then for some choice of angles $\theta$ and $\phi$ there would be exact agreement between $p_{3\text{-pred}}$ and the actual data from detector 3, $r_{3\text{-obs}}$. Given the noise, the best one can do is to find the angles that minimize the squared difference $d(\theta, \phi)$ between the predicted and observed responses during the interval of observation:

$$d(\theta, \phi) = \int_0^T |r_{3\text{-obs}}(t) - r_{3\text{-pred}}(t)|^2 \, dt. \tag{16.66}$$

Hidden in the integral for $d$ are the correlation integrals we began with, e.g. $\int r_3(t) r_1(t - \tau_3) \, dt$. These will normally be the most time-consuming part of the computation of $d$ for various angles, and should usually be done by FFTs. Once the correlations have been computed for all possible time-delays, they may be used to find the minimum of $d$ over all angles; this will determine the position of the source. Notice that if the noise is small, this information can then be substituted back into equation (16.64) for the first two detectors to find $h_+(t)$ and $h_\times(t)$. This reconstructs the signal. But if the source is weaker than the noise, then this substitution will give mostly noise.

The information we have gained about the unpredicted source, even if it is weak, is that it is there: its position is known and its arrival time can be determined roughly by restricting the time-interval over which the correlation integrals are done and finding the interval during which one gets significant correlations.This is enough to alert other astronomers to look for something in the source's position.

The paper by Gursel and Tinto (1989) contains a more sophisticated treatment of the noise than we have described here, allowing for different detectors to have different levels of noise, and constructing almost optimal filters for the signals that weight given detector responses according to where in their antenna pattern the signal seems to be coming from. They also give the results of extensive simulations and estimate the signal-to-noise ratio that will be required to give good predicitons. This paper is an important advance towards a robust solution of the reconstruction problem.

### 16.4.3 Using cross-correlation to search for a stochastic background

Another very important observation that interferometers will make is to find or set limits upon a background of radiation. This is much easier to do than finding discrete sources of continuous radiation, because there is no direction-finding or frequency-searching to do. This problem has been discussed in detail by Michelson (1987).

The most sensitive search for a background would be with two detectors on the same site, with the same polarization. Current plans for some installations envision more than one interferometer in one vacuum system, which would permit a correlation search. One would have to take care that common external sources of noise are excluded, especially seismic and other ground disturbances, but if this can be done then the two detectors should respond identically to any random waves coming in, and should therefore have the maximum possible correlation for these waves. The correlation can be calculated either by direct multiplication of the sampled data points ($2N$ operations per time delay between the two data sets) or by Fourier transform methods as in section 16.2.3(iii) above. We are only interested in the zero-time-delay value of the correlation, but in order to test the reality of the observed correlation, one would have to compute points at other time delays, where the correlation is expected to fall off. (How rapidly it falls off with increasing time delay depends on the spectrum of the background.) The choice of technique – direct multiplication or Fourier transform – will depend on the number of time-delays one wishes to compute and the capacity of one's computer.

If separated detectors are used, the essential physical point is that two separated detectors will still respond to waves in the same way if the waves have a wavelength $\lambda$ much longer than the separation between the detectors. Conversely, if the separation between detectors is greater than $\lambda/2\pi$, there is a significant loss of correlation. It is important as well to try to orient the detectors as nearly as possible in the same polarization state. In order to perform a search at 100 Hz, the maximum separation one would like to have is 500 km. This may be achievable within Europe, but it seems most unlikely that detectors in the USA will be built this close together. The data analysis is exactly the same as for two detectors on the same site.

## 16.5 Reconstructing the signal

The inverse problem is the problem of how to reconstruct the gravitational wave from the observations made by a network of detectors. A single detector produces limited information about the wave; in particular, on its own it cannot give directional information and therefore it cannot say what the intrinsic amplitude is. With three detectors, however, one can reconstruct the wave entirely. In the last two or three years there has been considerable progress in understanding the inverse problem: see Boulanger, le Denmant and Tourrenc (1988), Dhurandhar and Tinto (1988), Gursel and Tinto (1989), and Tinto and Dhurandhar (1989). I will summarize the main ideas as I understand them at present but this is an area in which much more development is likely soon. My thinking in this section has been shaped by conversations with Massimo Tinto and Kip Thorne.

### 16.5.1 Single bursts seen in several detectors

#### (i) Unfiltered signals

A gravitational wave is described by two constants – the position angles of its source, $(\theta, \phi)$ – and two functions of time – the amplitudes of the two independent polarizations $h_+(t)$ and $h_\times(t)$. Simple counting arguments give us an idea of how much we can learn from any given number of detectors. I will assume here that we do not have an *a priori* model (filter) for the signal. For signals that stand out above the broad-band noise:

- A single detector gives its response $r(t)$ and nothing else. Nothing exact can be said about the waves unless non-gravitational data can be used, as from optical or neutrino detections of the same event.
- Two detectors yield two responses and one approximate time-delay between the arrival of the wave in one detector and its arrival in the other. Two functions of time and one constant should not be enough to solve the problem, and indeed they are not. The time-delay is only an approximate one, because the two detectors will generally be responding to different linear combinations of $h_+(t)$ and $h_\times(t)$, so there will not be a perfect match between the responses of the two detectors, from which the time-delay must be inferred. The time-delay will confine the source to an error-band about a circle on the sky in a plane perpendicular to the line joining the detectors. The antenna patterns of the detectors can then be used to make some places on this circle more likely than others, but the unknown polarization of the wave will not allow great precision here. If the location of the source can be determined by other means, and if noise is not too large, then the two responses can determine the two amplitudes of the waves.
- Three detectors cross the threshold into precision astronomy, at least when the signals stand out against the broad-band noise. Here we have three functions of time (the responses) and two constants (the time-delays) as data, and this should suffice. As described in section 16.4 above, correlations among the three detectors can pin down the location of the source and, if noise is not too important, the time-dependent amplitudes as well. In this case, there is redundant information in the data that effectively test Einstein's predictions about the polarization of gravitational waves: the waveforms constructed from any pair of detectors should agree with those from the other two pairs to within noise fluctuations.

**(ii) Filtered signals**

If noise is so important that filtering is necessary, there is a completely different way of doing the counting. A given filter yields only constants as outputs, such as the maximum value of the correlation and the time the signal arrives (i.e. when it best matches the filter). It does not give useful functions of time. We can only assume that the signal's waveform matches the 'best' filter, so instead of two unknown time-dependent amplitudes we will have the response of the filter, the time-of-arrival, and a certain number of parameter constants that distinguish the observed waveform from others in its family.

Let us concentrate on coalescing binaries. The signal from a coalescing binary is an elliptically polarized, roughly sinusoidal waveform. The filters form a two-parameter family, characterized by the mass parameter $\mathcal{M}$ and the phase of the signal $\Phi$, as in equation (16.8). The parameters we want to deduce are: the amplitude $h$ of the signal, the ellipticity $e$ of its polarization ellipse (one minus the ratio of the minor and major axes), an orientation angle $\psi$ of the ellipse on the sky, and the binary's mass parameter $\mathcal{M}$. From these data we can not only determine the distance to the system, but also the inclination angle of the binary orbit to the line of sight (from $e$) and the orientation of the orbital plane on the sky ($\psi$).

The mass parameter $\mathcal{M}$ will be determined independently in each detector, and of course they will all agree if the event is real. Each detector in addition contributes the response of the filter, the phase parameter, and the time-of-arrival; these data must be used to deduce the five constants $\{\theta, \phi, h, e, \psi\}$. Here is how various numbers of detectors can use their data*:

- One detector does not have enough data, so it can only make average statements about the amplitude.
- Two detectors provide four useful data: two responses, one phase difference, and one time-delay. (Only the *differences* between the phases and times-of-arrival matter: the phase and time-of-arrival at the first detector are functions of the history of the source.) If the two detectors were identically polarized, the phase difference would necessarily be zero. A non-zero phase difference arises because the two principal polarizations in an elliptically polarized wave are 90° out of phase, so if the detectors respond to different combinations of these two polarizations, they will have different phases. With four data chasing five unknowns, the solution will presumably be a one-dimensional curve on the sky, but the problem has not yet been studied from this perspective.
- Three detectors have seven data: three responses, two phase differences and two time-delays. The two time-delays are sufficient to place the source at either of the intersections of two circles on the sky. For either location, the three responses determine $h$, $e$, and $\psi$. Presumably the phase differences would be consistent only with one of these positions, thereby solving the problem uniquely and incidentally providing the phase differences as a test of general relativity's model for the polarization of gravitational waves.

## 16.6 Data storage and exchange

Although the amount of data generated by a four-detector network will be huge, I would argue strongly that our present ignorance of gravitational wave sources

* This discussion is very different from previous ones I have given, e.g. Schutz (1989). In these I had not yet appreciated the importance of being able to determine the phase parameter independently of the time-of-arrival. This extra information makes it possible to solve the inverse problem with fewer detectors than I had previously believed.

makes it important that the data should be archived in a form that is relatively unprocessed, and kept for as long a time as possible, certainly for several years. It may be that new and unexpected sources of gravitational waves will be found, which will make it desirable to go over old data and re-filter it. It may also be that new classes of events will be discovered by their electromagnetic radiation, possibly with some considerable delay after the event would have produced gravitational waves, and a retrospective search would be desirable. In any case, we have already seen that it will be important to exchange essentially raw data between sites for cross-correlation searches for unknown events. Once exchanged, it is presumably already in a form in which it can be stored.

### 16.6.1 Storage requirements

We have seen in the introduction that a network could generate 5000 optical discs or videotapes per year. Data compression techniques and especially the discarding of most of the housekeeping data at times when it merely indicated that the detector was working satisfactorily could reduce this substantially, perhaps by as much as a factor of four. The cost of the storage media is not necessarily trivial. While videotapes are inexpensive, optical discs of large capacity could cost $250k at present prices (which will, hopefully, come down). Added to this is the cost of providing a suitable storage room, personnel to supervise the store, and equipment to make access to the data easy.

### 16.6.2 Exchanges of data among sites

We have already seen how important it will be to cross-correlate the raw data streams. At a data rate of some 100 kbytes per second, or even at 30 kbytes per second if the data volume is reduced as described above, one would have difficulty using standard international data networks. But these networks are being constantly upgraded, and so in five years the situation may be considerably different: it may be possible, at reasonable cost, to exchange short high-bandwidth bursts of data regularly via optical-fibre-to-satellite-to-optical-fibre routes. Alternatively, a cheaper solution might be to exchange optical discs or videotapes physically, accepting the inevitable delay. If lists of filtered events were exchanged on electronic data networks, then there may be less urgency about exchanging the full data sets.

#### (i) Protocols, analysis and archiving

It will be clear from our discussion that exchanging and jointly analysing data will require careful planning and coordination among all the groups. Discussions to this end are in a rudimentary stage now, but could soon be formalized more. Besides decisions on compatible hardware, software, data formats and modes of exchange, there are a number of 'political' questions that need to be resolved before observations begin. We are dealing with data that the groups involved have spent literally decades of their scientific careers to be in a position to obtain, and the scientific importance of actual observations of gravitational waves will be momentous. Questions of fairness and proprietary rights to the data could be a source of considerable friction if they are not clearly decided ahead of time. A model for some of these decisions could be the protocols adopted by the GRAVNET network of bar antennas, described elsewhere in this volume. Other models might be international VLBI, or large particle-physics collaborations.

Some of the questions that need to be addressed are:

- how much data needs to be exchanged;
- what groups have the right to see and analyse the data of other groups and what form of acknowledgement they need to give when they use it;
- what powers of veto groups have over the use of their data, for example in publications by other people;
- how long the proprietary veto would last before the data become 'public domain' (the funding agencies will presumably apply pressure to allow ready access to the data by other scientists after some reasonable interval of time);
- how long the data need to be archived.

Given the volume of data and the logistical complications of multi-way exchanges of it, it may be attractive to establish one or more joint data analysis and archiving centres. These could be particularly attractive as sites for any large computers dedicated to the pulsar-search problem. These would collect the data and store it, and perform the cross-correlations that can only be done with the full data sets on hand.

## 16.7 Conclusions

In this review I have set out what I understand about the data analysis problem as of September, 1989. Evidently, the field is covered very non-uniformly: coalescing binaries have received much more attention than pulsars or stochastic sources so far, and protocols for data exchange are something mainly for the future.

Nevertheless, it is clear that questions of the type we have discussed here will influence in an important way decisions about the detectors: how many there will be, where they will be located, what their orientations will be, what weights one should apply to the various important parameters affecting their sensitivity (e.g., length, seismic isolation, laser power) when deciding how to apportion limited budgets to attain the maximum sensitivity. Other questions that I have not addressed will also be important, particularly choosing the particular recycling configuration most suitable to searching for a given class of sources.

From the present perspective, it seems very likely that in ten years or so a number of large-scale interferometric detectors will be operating with a broadband sensitivity approaching $10^{-22}$. The data should contain plenty of coalescing binaries and at least a few supernovae; but the most exciting thing that we can look forward to is the unexpected: will this sensitivity suffice to discover completely unanticipated sources? The best way to ensure that it does is to make sure that our data-analysis algorithms and data-exchange protocols are adequate to the task: given the enormous efforts being made by the hardware groups to develop the detectors, and the considerable amount of money that will be required to build them, it is important that development of the data-analysis tools not be left too late. Solutions to data-analysis problems must be developed in parallel with detector technology.

## Acknowledgements

It is a pleasure to acknowledge the helpfulness of many colleagues, from whom I have learned much about data analysis and many of whom have contributed crucial insights into the problem. They include Mark Davis, Ron Drever, Jim Hough, Andrzej Krolak, Sean Lawrence, Alberto Lobo, Norman MacKenzie, Brian Meers, Eduardo Nahmad-Achar, Tim Niebauer, Norna Robertson, Roland Schilling, Kip Thorne, Massimo Tinto, Philippe Tourrenc, Harry Ward, Rai Weiss, and Sheryl Smith Woodward. I am grateful to Drs Hough, Robertson, Tinto, and Ware for correcting errors in earlier drafts of this chapter. Any remaining errors are, of course, my own.

## References

Boulanger, J. L., le Denmant, G. and Tourrenc, Ph. (1988). *Phys. Lett. A* **126,** 213–18.

Davis, M. H. A. (1989). In *Gravitational Wave Data Analysis*, ed. B. F. Schutz, pp. 73–94, Kluwer, Dordrecht.

Dewey, D. (1986). In *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity*, ed. R. Ruffini, p. 581, Elsevier, Amsterdam.

Dhurandhar, S. V. and Tinto, M. (1988). *Mon. Not. R. Astr. Soc.* **234,** 663–76.

Dhurandhar, S. V., Schutz, B. F. and Watkins, W. J. (1991). In preparation.

Evans, C. R. (1986). In *Dynamical Spacetimes and Numerical Relativity*, ed. J. M. Centrella, pp. 3–39, Cambridge University Press.

Evans, C. R., Iben, Jr., I. and Smarr, L. (1987). *Astrophys. J.* **323,** 129–39.

Gursel, Y. and Tinto, M. (1989). *Phys. Rev. D* **40,** 3884–938.

Hocking, W. K. (1989). *Computers in Physics*, Jan/Feb 1989 issue, pp. 59–65.

Horowitz, P. (1969). *Rev. Sci. Instrum.* **40,** 369–70.

Hough, J., Drever, R. W. P., Ward, F., Munley, A. J., Newton, G. P., Meers, B. J., Hoggan, S. and Kerr, G. A. (1983). *Nature* **303,** 216.

Ipser, J. R. and Managan, R. A. (1984). *Astrophys. J.* **282,** 287.

Krolak, A. (1989). In *Gravitational Wave Data Analysis*, ed. B. F. Schutz, pp. 59–69, Kluwer, Dordrecht.

Krolak, A. and Schutz, B. F. (1987). *Gen. Rel. Gravit.* **19,** 1163–71.

Livas, J. C. (1987). Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

Lyne, A. (1989). In *Gravitational Wave Data Analysis*, ed. B. F. Schutz, pp. 95–103, Kluwer, Dordrecht.

Michelson, P. F. (1987). *Mon. Not. R. Astr. Soc.* **227,** 933–41.

Pasetti, A. (1987). M.Sc. thesis, Imperial College, London.

Piran, T. and Stark, R. F. (1986). In *Dynamical Spacetimes and Numerical Relativity*, ed. J. M. Centrella, pp. 40–73, Cambridge University Press.

Schutz, B. F. (1986). In *Gravitational Collapse and Relativity*, H. Sato and T. Nakamura, pp. 350–68, World Scientic, Singapore.

Schutz, B. F. (1989). In *Gravitational Wave Data Analysis*, ed. B. F. Schutz, pp. 315–26, Kluwer, Dordrecht.

Smith, S. (1987). *Phys. Rev. D* **36,** 2901–4.

Srinath, M. D. and Rajasekaran, P. K. (1979). *An Introduction to Statistical Signal Processing with Applications*, Wiley, New York.

Thorne, K. S. (1987). In *300 Years of Gravitation*, eds. S. W. Hawking and W. Israel, pp. 330–458, Cambridge University Press.

Tinto, M. and Dhurandhar, S. V. (1989). *Mon. Not. R. Astr. Soc.* **236,** 621–7.