Determining the Hubble constant from gravitational wave observations

Bernard F. Schutz

Department of Applied Mathematics and Astronomy, University College Cardiff, PO Box 78, Cardiff CF1 1XL, UK

I report here how gravitational wave observations can be used to determine the Hubble constant, H_0. The nearly monochromatic gravitational waves emitted by the decaying orbit of an ultra-compact, two-neutron-star binary system just before the stars coalesce are very likely to be detected by the kilometre-sized interferometric gravitational wave antennas now being designed[1-4]. The signal is easily identified and contains enough information to determine the absolute distance to the binary, independently of any assumptions about the masses of the stars. Ten events out to 100 Mpc may suffice to measure the Hubble constant to 3% accuracy.

The signal from a system of two $1M_\odot$ stars (where $M_\odot$ is the mass of the Sun) will sweep from 100 Hz to 1 kHz in ~3 s. There might be three events per year out to 100 Mpc, and if the detectors achieve their current design sensitivity, such events will be detectable with a signal-to-noise ratio of 30. To determine the distance, the signal has to be observed by a worldwide network of three, and preferably four, detectors. By measuring both the response of the detectors and the delays between the arrival times of the signal at different detectors, the network should be able to locate the source in an error box of ~36 square degrees. There is some chance that the coalescence event will be optically identifiable (I.D. Novikov, personal communication); otherwise, clustering of galaxies provides a statistical method that will still yield H_0 after remarkably few events. Here I give only a brief discussion; full details will be published elsewhere.

Several detectors being developed in the United States and Europe[2-4] will take the form of interferometers with arm lengths 1-4 km, observing bandwidths 10^2-10^4 Hz and r.m.s. noise levels at 100 Hz of $<10^{-24}$ strain $\mathrm{Hz}^{-1/2}$. Within 10 years we may expect that there will be four or five such detectors in operation in America and Europe, with typical separations of 6,500 km. [It is possible that bar detectors could contribute to these observations. However, because of their narrow bandwidth, their detection of coalescing binaries requires quite different methods, which have not been studied. I shall therefore concentrate on interferometric detectors.]

Although there are many possible sources of gravitational waves, the most promising for detection by these instruments seems to be the coalescence of binary neutron stars, as will happen to the binary pulsar PSR1913+16 in $\sim 10^8$ yr. The gravitational waves from these sources before coalescence can be predicted very reliably (K. S. Thorne, personal communication). As an orbit decays through the emission of gravitational radiation, its eccentricity is reduced, so we need only consider systems with circular orbits[5]. Consider a binary at a distance $100r_{100}$ Mpc, with total mass $m_T M_\odot$ and reduced mass $\mu M_\odot$, emitting waves at frequency $100f_{100}$ Hz (twice its orbital frequency). The standard 'quadrupole formula' of general relativity[6,7] shows that the waves will have amplitude (r.m.s.-averaged over detector and source orientations)

$$\langle h\rangle = 1\times 10^{-23} m_T^{2/3}\mu f_{100}^{2/3} r_{100}^{-1} \quad (1)$$

and that their frequency will change on a timescale

$$\tau = f/\dot{f} = 7.8 m_T^{-2/3}\mu^{-1} f_{100}^{-8/3}\ \mathrm{s} \quad (2)$$

Two $1.4M_\odot$ neutron stars will coalesce[1] when $f \simeq 10^3$ Hz. By using matched filters to analyse the data[5], the noise can effectively be limited to a bandwidth of $\sim\tau^{-1}$. This will enable the detectors to see binary neutron star sources at 100 Hz at a distance of 100 Mpc, with a mean signal-to-noise ratio (SNR) of >30. An observation will therefore determine τ and h to perhaps 3%. The key to our method is that the stars' masses enter equations (1) and (2) in exactly the same way, so that

$$r_{100} = 7.8 f_{100}^{-2}(\langle h_{23}\rangle\tau)^{-1} \quad (3)$$

where $\langle h_{23}\rangle = \langle h\rangle \times 10^{23}$, independently of the masses of the stars.

This result is not quite so strong as it seems, as equation (1) gives the r.m.s. value of h averaged over orientations, whereas the value of h inferred from the network's observations will depend on the binary system's orientation and position relative to the detectors as well as its distance. However, these can be determined from the observations: as I show below, provided that three or more detectors register the same event, they can determine the location on the sky and the degree of elliptical polarization of the wave. (In general relativity, gravitational waves are transverse and have only two independent polarizations[6-7].) Now, the radiation emitted by the binary along its angular momentum axis is circularly polarized, whereas that in the equatorial plane is linearly polarized. The degree of eliptical polarization therefore determines the inclination of the orbit to the line of sight, which enables us to solve for r_{100} in terms of the observed h. Equation (3) also depends on being able to model the system as two newtonian point masses. As we shall see below, tidal and relativistic corrections are negligible in the range of orbital parameters we require.

Being able to determine r directly from the observations is remarkable in itself, but it is only really useful if the source of the event can be identified. For this an accurate position is required. Because this accuracy is crucial for the determination of the Hubble constant, I will discuss it in some detail.

Each detector has quadrupolar linear polarization, so it is not highly directional; however, the differences in arrival time of a wave at different detectors can be used to triangulate the position. Between any two detectors with separation d, a wave travelling at an angle θ to the line joining the detectors will arrive at the second detector with a delay $\Delta t = d\cos\theta/c$ relative to the first, where c is the speed of light. For $d \simeq 6.5\times 10^3$ km, we have $|\Delta t| \leq 22$ ms. As the two detectors will generally not have the same polarization, there will be a further effective time delay due to the wave's elliptical polarization. Such a polarization can be regarded as a superposition of the two independent linear polarizations defined by the detectors, with a phase shift between them. This sphase shift means that differently polarized detectors record the wave train with extra time delays of up to one period (+10 ms for a 100-Hz signal). The two independent time delays measured among three detectors and the three measured amplitudes are sufficient to determine the waves' five unknowns: arrival directions (two), amplitudes of the different polarizations (two), and phase lag of the polarizations (one).

The precision with which the source's position and polarization can be measured depends on the two sorts of errors: the accuracy with which the arrival time of the wave at a detector (and hence the time delays) can be determined, and the accuracy with which the amplitude of the detector's response can be measured. In what follows, I will assume that $m_T^{2/3}\mu = 1$ (for example, two stars of $\sim 1.1M_\odot$) to illustrate the situation. We shall see that the timing accuracy is typically 1% of the maximum timing range (from −22 to +22 ms), and the amplitude error is ~3%. When only three detectors see an event, there are actually two error boxes of size $\sim 10°\times 10°$, which may be too large for our purposes. I will therefore consider events detected in four instruments. The seven data overdetermine the five unknowns, and this redundancy offers us the opportunity to reduce the effective amplitude noise (it also allows a test of Einstein's polarization predictions). In this way, three timing measurements at ±1% and one amplitude measurement with effective error ±3% can be used to locate the source. This suggests that a positional error of ±3° is not unreasonable, giving an error box of 6°×6°. Before discussing how this can be used to identify the source, it is necessary to describe how the arrival time can be measured to 0.5 ms.

Suppose the data are extracted from the noise by convolution with a 'template' waveform (D. Dewey, in preparation) (one way of implementing a matched filter technique). Then the maximum value of the convolution marks the time at which the signal best matches the beginning of the template. This is the 'arrival' of the wave. If the signal-to-noise ratio is 30 then this convolution will have 3% fluctuations. The arrival time can then be determined only to within an error ε, which is the time shift that reduces the noise-free convolution by 3%. I have performed these convolutions for a template spanning 100-200 Hz with a variety of values of $f/\dot{f}$, and I find $\varepsilon \approx 0.5$ ms. (If the signal-to-noise ratio doubles, this value decreases by $\sqrt{2}$.)

The error box is roughly the size of a Schmidt plate. If such events are optically visible, then a search of the error box may identify the galaxy, whose redshift will determine H_0 to a few per cent (limited by the 'random' velocities of galaxies and the distance accuracy). This accuracy improves with the number of events, N, as $N^{1/2}$. If optical identifications are not possible, then the following statistical method, based on galaxy clustering, should still work.

If we accept that H_0 is less than some H_{max} (say 120 km s^{-1} Mpc^{-1}), then the error box can be surveyed for bright galaxies (up to ~15 mag) with velocities below $H_{max}r$, where we consider at first only events with $r < 100$ Mpc. Existing surveys (see ref. 8) show that galaxies with velocities <12,000 km s^{-1} cluster strongly, with ~1 cluster per square degree, and that bright galacies are good tracers of these clusters. In our error box, therefore, the source ought to be in one of ~36 clusters. Each cluster redshift gives a candidate value for H_0. As r is known to 3%, we can divide the range of H_0 into 30 bins. Each observed coalescence event produces one 'correct' H_0 and 35 spurious ones, which are distributed randomly among the 30 bins with probability $(H_0^2/3H_{max}^3)\mathrm{d}H_0$ that a value lies between H_0 and $H_0+\mathrm{d}H_0$. After N events, the bin at $H_0 = 120$ therefore accumulates the largest number of 'randoms', three times the mean number per bin: $3.5N \pm (3.5N)^{1/2}$. In the worst case, if the true H_0 is 120, we will need $N \gtrsim 2(3.5N)^{1/2}$ to see the true values above the fluctuations in the randoms, or $N \gtrsim 14$. If H_0 is really near 60, then we should see a good peak after only three or four events. These values of N will give H_0 to 3% or better. These numbers are only illustrative, as there are many variables which affect them: for example, the typical masses of the coalescing stars, the actual sensitivities achieved by the detectors, and optical effects such as obscuration by the Galaxy. But they show that even the statistical method looks very promising, given a sufficient event rate.

What, then, is the event rate? Estimates[9] based on the pulsar birth rate and the fact that the binary pulsar is the only compact-object binary system known with a lifetime $<10^{10}$ yr suggest that there will be ~3 events per yr out to 100 Mpc. At this rate, the Hubble constant would be determined in 1-10 yr. But the event rate is highly uncertain, at least by a factor of 10. Therefore it is possible that this method would work only in the long term, some decades after the interferometers begin working. The question of the event rate deserves more attention from astrophysicists.

A small event rate can be offset to some extent by improving the positional accuracy of each event. Obviously, increasing the detectors' sensitivity will reduce the size of the error box. A dramatic improvement can also be achieved by adding a fifth detector to the network. An extra detector in Asia would provide a longer baseline for the timing accuracy. This would improve the position by perhaps a factor of 2-3, and probably speed up the determination of H_0 by a larger factor if the statistical method needs to be used. If the statistical method is used, it will not be helped by including events from further away (events are visible to ~0.5 Gpc at moderate SNR). This is because degrading the SNR enlarges the error box and increases the confusion with randoms.

Finally, how secure is the model? When two $1.4M_\odot$ neutron stars are emitting 100-Hz waves, their separation will be $a \sim 160$ km, or ~10-15 stellar radii. Tidal effects on the orbital period will be negligible: angular momentum transfers are proportional to $(R/a)^5$, where R is the radius of either star. Eardley and Clark[1] considered the tidal mass transfer, which is not important below 200 Hz unless the smaller star has a mass $\lesssim 0.3M_\odot$. It seems very likely, therefore, that tidal effects will not seriously contaminate the sample. Post-newtonian gravity introduces corrections of ~1% to the orbital period of these systems; moreover, the rate at which they radiate gravitational waves also has corrections of this order from both post-newtonian and octupole contributions. These last have not been accurately calculated, but it is unlikely that they will limit the coherence of the 'template' with the true signal in a way that degrades the accuracy of the timing measurement by much more than the few per cent we have already allowed for. The great attraction of this method is, therefore, the simplicity of the model. Even if the event rate is low, the value of H_0 obtained in this way should, in the long run, be less troubled by systematic errors than that from other methods. All it needs is the continued development of large-scale gravitational wave detectors.

For my discussion of clustering I received invaluable assistance from Steve Phillipps. I also acknowledge helpful conversations with Brandon Carter, Mike Disney, Jim Hough, Kip Thorne, Massimo Tinto and Harry Ward.

Received 6 May; accepted 17 July 1986.

1. Clark, J. P. A. & Eardley, D. M. *Astrophys. J.* **215**, 311-322 (1977).
2. Drever, R. W. P. & Weiss, R. preprint, California Institute of Technology (Pasadena, 1986).
3. Hough, J. *et al. British Long Baseline Gravitational Wave Observatory* GWD/RAL/86-001 (Rutherford-Appleton Laboratory, Chilton, 1986).
4. Maischberger, K. *et al.* Max-Planck-Institut für Quantenoptik Publ. No. MPQ 96 (Garching, 1985).
5. Peters, P. C. *Phys. Rev.* **136**, 1224-1232 (1964).
6. Schutz, B. F. *A First Course in General Relativity* (Cambridge University Press, 1985).
7. Misner, C. W., Thorne, K. S. & Wheeler, J. A. *Gravitation* (Freeman, San Francisco, 1973).
8. Kirshner, R. P., Oemler, A., Jr., Schechter, P. L. & Shechtman, S. A. *Astr. J.* **88**, 1285-1300 (1983).
9. Clark, J. P. A., van den Heuvel, E. P. J. & Sutantyo, W. *Astr. Astrophys.* **72**, 120-128 (1979).