

The Science and Detection of Gravitational Waves

Barry C. Barish

California Institute of Technology LIGO 18-34

1201 E. California Blvd.

Pasadena, CA 91125, USA

Received on 15 February, 2002

Albert Einstein predicted the existence of gravitational waves as a consequence of the general theory of relativity. In his theory, changes in the shape of concentrations of mass (or energy) warp space-time, cause distortions that propagate through the Universe at the speed of light. However, no direct detection of such waves has yet been made. A new generation of detectors promises sensitivities that will be capable of detection from a variety of catastrophic events, such as the gravitational collapse of stars or the coalescence of compact binary systems.

I Introduction

It has been 85 years since Einstein revolutionized our concept of gravity by developing a new theory where the physical consequences of gravity are based on curvature of space-time, rather than a mysterious force between massive bodies. Gravitational forces do not pull on an object, as in Newton's theory. Rather, the presence of massive objects curve space-time in the vicinity around them, and a body moving nearby has its path determined by that curvature.

This new theory immediately solved one of the most well known puzzles of the time, involving the planet Mercury. The orbit of Mercury around the sun is highly elliptical and the point of closest approach, the perihelion, rotates slightly around the sun on each traversal, making a daisy pattern after many trips. This effect is primarily due to the gravitational effects of the other planets slightly shifting the orbit. However, when these effects are calculated using Newton's laws, they account for only 574 arcseconds of the 617 arcseconds shift of the perihelion. This discrepancy represented a rare instance of a failure of Newton's theory of gravity to explain astronomical observations. Of course, this prompted many conjectures to explain the effect, for example that there undetected planets, that Venus was heavier than thought or that Mercury had a small moon. The idea that perhaps Newton's theory of gravity wasn't the whole story seemed an unlikely possibility. However, when Einstein calculated the actual shift precisely using his new theory of gravity, this provided a dramatic demonstration that he had developed a new theory that could explain more than the previous theory of Newton.

Another dramatic and natural prediction of this new theory involved the effect of gravity on the path of light rays. Since the effect of one object on another is a result of a curvature of space-time in Einstein's theory, the passage of a massless light wave past a massive body would produce the same kind of distortion on its trajectory as experienced by Mercury passing nearby the Sun. Einstein calculated and predicted a very small effect of only a couple of arcseconds for a light ray passing near the sun's surface. Nevertheless, soon after Einstein made his prediction, the challenge of seeing this effect was taken up by Sir Arthur Eddington, who led a team who went to the Southern Hemisphere to look for this predicted deflection of light coming stars behind the sun. His idea was to carefully measure the light trajectories from these stars at the time of a solar eclipse, which occurred in 1919, and allowed him to view light from these objects that grazed the sun's surface without being drowned out by light from the sun. Although by modern standards the evidence was not very convincing, Eddington reported a couple of observations that showed bending that could not be explained by Newton's laws and were consistent with the predictions of Einstein. This result was acclaimed as a dramatic proof of Einstein's new theory of gravity.

In modern astronomy with much more precise instruments, measurements of curvatures of trajectories due to gravitational effects has become a cottage industry. The general effect of bending of light rays from one object behind another is called gravitational lensing and over the past couple of decades it has become one of the most powerful new tools in astronomy with broad application at different scales. One of the more

interesting applications is the search for dark matter in the form of dark massive objects called Machos. Such dark objects are being searched for by the gravitational lensing effect as they pass between the observer and the bright object behind.

One last and crucial prediction of Einstein's new theory was the existence of gravitational waves. Newton's theory of gravity had a basic flaw in that it involved instantaneous action at a distance. In other words, there is nothing in the theory that requires any time for the message to be carried from the action to the observer. Instead, in Einstein's theory the information is carried at the speed of light by gravitational waves, just as it is for electromagnetic waves. Unfortunately, this part of the theory, while being perhaps the most crucial, was the hardest to test. At the time Einstein made his predictions there was no hope of ever measuring such weak effects and the concept was treated for many years as more of a feature of the theory than an observable experimental effect.

What Einstein showed was that time dependent gravitational fields come from the acceleration of masses and propagate away from their sources as a space-time warpage at the speed of light. Although we draw the analogy with electromagnetic radiation, the basic nature of the waves themselves is quite different. While electromagnetic radiation involves the propagation of photons through space, gravitational radiation is propagation of waves in the fabric of space-time itself. The speed of the waves is the same for electromagnetic and gravitational radiation in Einstein's theory.

Using the Minkowski metric, the information about space-time curvature is contained in the metric as an added term, $h_{\mu\nu}$. In the weak field limit, the equation can be described with linear equations. If the choice of gauge is the *transverse traceless gauge* the formulation becomes a familiar wave equation

$$(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}) h_{\mu\nu} = 0$$

The strain $h_{\mu\nu}$ takes the form of a plane wave propagating at the speed of light (c).

$$h_{\mu\nu} = h_+(t - z/c) + h_x(t - z/c)$$

A consequence of the underlying quantum mechanical description of gravity being spin 2 is that the waves have two components rotated by 45° from each other, in contrast to 90° for electromagnetic waves. It is an interesting fact that if gravitational waves are observed and the two components of the waves are decomposed in the classical experiment described in this article, they

will confirm the underlying quantum spin 2 structure of gravity.

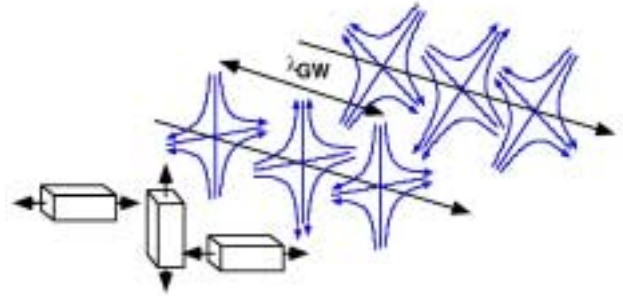


Figure 1. The propagation of gravitational waves having two polarizations.

II The evidence for gravitational waves

Indirect but convincing evidence for the existence of gravitational waves were produced from the beautiful observations of Russell Hulse and Joseph Taylor. They studied a special neutron star binary pulsar system (PSR 1913+16) that they discovered in 1974. They determined that it was a binary system because they observed a variation of the frequency with just under an 8-hour period. The pulsar in this binary system has a frequency of about 17/sec and this 'clock' allowed them to accurately determine the characteristics of the overall binary system with remarkable precision. The most important parameters for our purpose are that the two neutron stars are separated by about 10^6 miles, have masses $m_1 = 1.4$ solar masses and $m_2 = 1.36$ solar masses, and the ellipticity of the orbit is $e = 0.617$. Hulse and Taylor demonstrated that the motion of the pulsar around its companion could not be understood unless the dissipative reaction force associated with gravitational wave production were included. They found that the system radiates away energy, presumably in the form of gravitational waves, and the two neutron stars are slowly spiraling in toward one another resulting in a speeding up of the orbital period. In detail the inspiral is only 3mm/orbit so it will be more than 10^6 years before the pair of neutron stars actually coalesce.

Hulse and Taylor monitored these pulsar signals with 50-microsecond accuracy over many years. They measured the orbital speedup experimentally with an accuracy of a fraction of a percent. The value of the speedup they recorded is in complete agreement with the predictions from general relativity. Hulse and Taylor received the Nobel Prize in Physics for this work

in 1993. This impressive indirect evidence for gravitational waves gives us good reason to believe in their existence. But, as of this date, no direct detection of gravitational waves has been made.

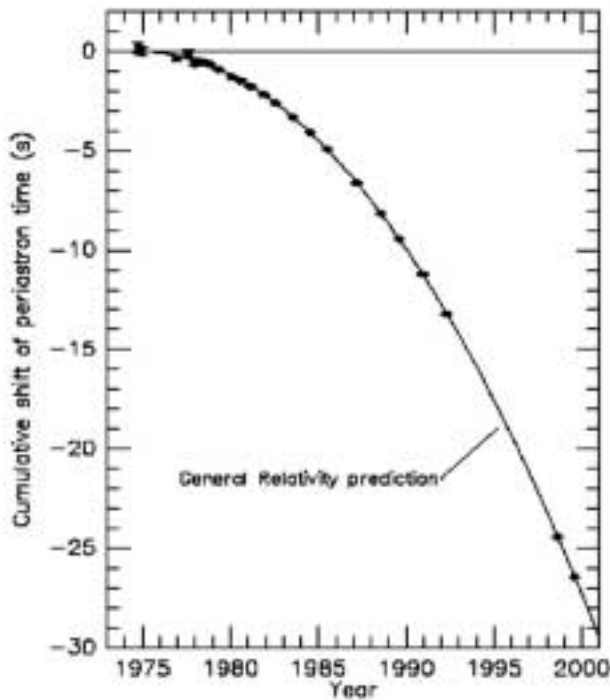


Figure 2. The evidence for gravitational wave emission from Taylor and Weinberg.

III Detection of gravitational waves

The effect of the propagating gravitational wave is to deform space in a quadrupolar form. The characteristics of the deformation are indicated in Fig. 3.

One can estimate the frequency of the emitted gravitational wave. An upper limit on the gravitational wave source frequency can be estimated from the Schwarzschild radius $2GM/c^2$ of the radiated object. We do not expect strong emission for periods shorter than the light travel time $4\pi GM/c^3$ around its circumference. From this we can estimate the maximum frequency as about $10^4 Hz$ for a solar mass object. Of course, the frequency can be much lower as illustrated by the 8 hour period of PSR1916+13, which is emitting gravitational radiation. Frequencies in the higher frequency range $1Hz < f < 10^4 Hz$ are potentially reachable using detectors on the earth's surface, while the lower frequencies require putting instruments into space. The physics goals of the terrestrial detectors and the LISA space mission are complementary, much

like different frequency bands are used in observational astronomy for electromagnetic radiation

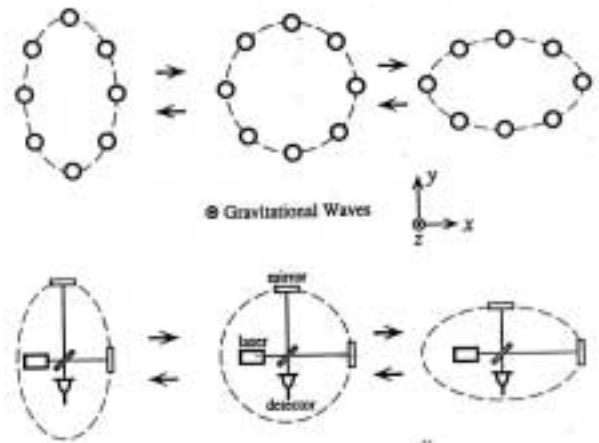


Figure 3. The effect of gravitational waves for one polarization is shown at the top on a ring of free particles. The circle alternately elongates vertically while squashing horizontally and vice versa with the frequency of the gravitational wave. The detection technique of interferometry being employed in the new generation of detectors is indicated in the lower figure. The interferometer measures the difference in distance in two perpendicular directions, which if sensitive enough could detect the passage of a gravitational wave.

The strength of a gravitational wave signal depends crucially on the quadrupole moment. We can roughly estimate how large the effect could be from astrophysical sources. If we denote the quadrupole of the mass distribution of a source by Q , a dimensional argument, along with the assumption that gravitational radiation couples to the quadrupole moment yields:

$$h \sim \frac{G\ddot{Q}}{c^4 r} \sim \frac{G(E_{kin}^{non-symm.}/c^2)}{c^2 r}$$

where G is the gravitational constant and $E_{kin}^{non-symm.}$ is the non-symmetrical part of the kinetic energy.

For the purpose of estimation, let us consider the case where one solar mass is in the form of non-symmetric kinetic energy. Then, at a distance of the Virgo cluster we estimate a strain of $h \sim 10^{-21}$. This is a good guide to the largest signals that might be observed. At larger distances or for sources with a smaller quadrupole component the signal will be weaker.

IV Long baseline suspended mass interferometry

A Michelson interferometer operating between *freely suspended* masses is ideally suited to detect the anti-symmetric (compression along one dimension and expansion along an orthogonal one) distortions of space

induced by the gravitational waves as was illustrated in Fig. 3.

The simplest configuration, a white light (equal arm) Michelson interferometer is instructive in visualizing many of the concepts. In such a system the two interferometer arms are identical in length and in the light storage time. Light brought to the beam splitter is divided evenly between the two arms of the interferometer. The light is transmitted through the splitter to reach one arm and reflected by the splitter to reach the other arm. The light traverses the arms and is returned to the splitter by the distant arm mirrors. The roles of reflection and transmission are interchanged on this return and, furthermore, due to the Fresnel laws of E & M the return reflection is accompanied by a sign reversal of the optical electric field. When the optical electric fields that have come from the two arms are recombined at the beam splitter, the beams that were treated to a reflection (transmission) followed by a transmission (reflection) emerge at the antisymmetric port of the beam splitter while those that have been treated to successive reflections (transmissions) will emerge at the symmetric port.

In a simple Michelson configuration the detector is placed at the antisymmetric port and the light source at the symmetric port. If the beam geometry is such as to have a single phase over the propagating wavefront (an idealized uniphase plane wave has this property as does the Gaussian wavefront in the lowest order spatial mode of a laser), then, providing the arms are equal in length (or their difference in length is a multiple of $1/2$ the light wavelength), the entire field at the antisymmetric port will be dark. The destructive interference over the entire beam wavefront is complete and all the light will constructively recombine at the symmetric port. The interferometer acts like a light valve sending light to the antisymmetric or symmetric port depending on the path length difference in the arms.

If the system is balanced so that no light appears at the antisymmetric port, the gravitational wave passing through the interferometer will disturb the balance and cause light to fall on the photodetector at the dark port. This is the basis of the detection of gravitational waves in a suspended mass interferometer. In order to obtain the required sensitivity, the arms of the interferometer must be long.

The amount of motion of the arms to produce an intensity change at the photodetector depends on the optical length of the arm; the longer the arm the greater is the change in length up to a length that is equal to $1/2$ the gravitational wave wavelength. Equivalently the longer the interaction of the light with the gravitational wave, up to $1/2$ the period of the gravitational

wave, the larger is the optical phase shift due to the gravitational wave and thereby the larger is the intensity change at the photodetector. The initial long baseline interferometers, besides having long arms also will fold the optical beams in the arms in optical cavities or delay lines to gain further increase in the path length or equivalently in the interaction time of the light with the gravitational wave. The initial LIGO interferometers will store the light about 50 times longer than the beam transit time in an arm. (A light storage time of about 1 millisecond.)

V The noise floor for interferometer detectors

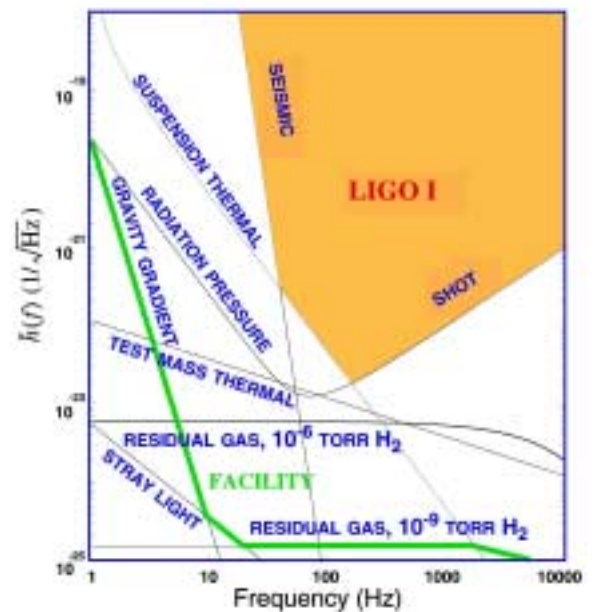


Figure 4. The limiting noise sources for the initial LIGO detectors. Note that the interferometer is limited by different sources at low frequency (eg. seismic), middle frequencies by suspension thermal noise, and at high frequencies by shot noise (or photo statistics). Lurking below are many other potential noise sources.

The success of the detector ultimately will depend on how well we one can to control the noise in the measurement of these small strains. Noise is broadly but also usefully categorized in terms of those phenomena which limit the ability to sense and register the small motions (sensing noise limits) and those that perturb the masses by causing small motions (random force noise). Eventually one reaches the ultimate limiting noise, the quantum limit, which combines the sensing noise with a random force limit. This orderly and intellectually satisfying categorization presumes that one is careful enough as experimenters in the execution of the

experiment that one has not produced less fundamental, albeit, real noise sources that are caused by faulty design or poor implementation. We have dubbed these as technical noise sources and in real life these have often been the impediments to progress. The primary noise sources for the initial detectors are illustrated in Fig. 4, where the estimated levels of the various noise sources are shown for LIGO. The other interferometers have similar curves with some difference due to the different trade-offs that have been made.

In order to control the technical noise sources, extensive use is made of two concepts. The first is the technique of modulating the signal to be detected at frequencies far above the $1/f$ noise due to the drift and gain instabilities experienced in all instruments. For example, the optical phase measurement to determine the motion of the fringe is carried out at radio frequency rather than near DC. Thereby, the low frequency amplitude noise in the laser light will not directly perturb the measurement of the fringe position. (The low frequency noise still will cause radiation pressure fluctuations on the mirrors through the asymmetries in the interferometer arms.) A second concept is to apply feedback to physical variables in the experiment to control the large excursions at low frequencies and to provide damping. The variable is measured through the control signal required to hold it stationary. Here a good example is the position of the interferometer mirrors at low frequency. The interferometer fringe is maintained at a fixed phase by holding the mirrors at fixed positions at low frequencies. Feedback forces to the mirrors effectively hold the mirrors "rigidly". In the initial LIGO interferometers the forces are provided by permanent magnet/coil combinations. The mirror motion that would have occurred is then read in the control signal required to hold the mirror.

Great care must be taken to control these technical noise sources. In order to test and understand the sensitivity and limiting noise, extensive tests have been performed with a 40 meter LIGO prototype interferometer on the Caltech campus. This interferometer essentially has all the pieces and the optical configuration used in LIGO, so represents a good place to understand noise and performance before the full-scale LIGO interferometers are in operation. The 40 m prototype device achieved a displacement sensitivity of $h \sim 10^{-19}m$, which is close to the displacement sensitivity that is required in the 4 km LIGO interferometers.

VI Astrophysical sources of gravitational waves

There are a many known astrophysical processes in the Universe that produce gravitational waves. Terrestrial interferometers, like LIGO, will search for signals from such sources in the 10Hz - 10KHz frequency band. Characteristic signals from astrophysical sources will be sought over background noise from recorded time-frequency series of the strain. Examples of such characteristic signals include the following:

VI.1 Chirp Signals

The inspiral of compact objects such as a pair of neutron stars or black holes will give radiation that will characteristically increase in both amplitude and frequency as they move toward the final coalescence of the system.

This chirp signal can be characterized in detail, giving the dependence on the masses, separation, ellipticity of the orbits, etc. A variety of search techniques, including the direct comparison with an array of templates will be used for this type of search. The waveform for the inspiral phase is well understood and has been calculated in sufficient detail for neutron star-neutron star inspiral. To Newtonian order, the inspiral gravitational waveform is given by

$$h_+(t) = \frac{2G^{\frac{5}{8}}}{c^4} (1 + \cos^2(i)) \frac{\mu}{r} (\pi M f)^{\frac{2}{3}} \cos(2\pi f t)$$

$$h_-(t) = \pm \frac{4G^{\frac{5}{8}}}{c^4} \cos(i) \frac{\mu}{r} (\pi M f)^{\frac{2}{3}} \sin(2\pi f t)$$

where the + and - polarization axes are oriented along the major and minor axes of the projection of the orbital plane on the sky, i is the angle of inclination of the orbital plane, $M = m_1 + m_2$ is the total mass, $\mu = m_1 m_2 / M$ is the reduced mass and the gravitational wave frequency f (twice the orbital frequency) evolves as

$$f(t) = \frac{1}{\pi} \left[\frac{c^3}{G} \right]^{\frac{5}{8}} \left[\frac{5}{256 \mu M^{\frac{5}{2}} (t_0 - t)} \right]^{\frac{3}{8}}$$

where t_0 is the coalescence time. This formula gives the characteristic 'chirp' signal - a periodic sinusoidal wave that increases in both amplitude and frequency as the binary system inspirals.

The Newtonian order waveforms do not provide the needed accuracy to track the phase evolution of the inspiral to a quarter of a cycle over the many thousands

of cycles that a typical inspiral will experience while sweeping through the broad band LIGO interferometers. In order to better track the phase evolution of the inspiral, first and second order corrections to the Newtonian quadrupole radiation, known as the post-Newtonian formulation, must be applied and are used to generate templates of the evolution that are compared to the data in the actual search algorithms. If such a phase evolution is tracked, it is possible to extract parametric information about the binary system such as the masses, spins, distance, ellipticity and orbital inclination. An example of the chirp form and the detailed structure expected for different detailed parameters is shown in Fig. 5.

This inspiral phase is well matched to the LIGO sensitivity band for neutron star binary systems. For heavier systems, like a system of two black holes, the final coalescence and even the ring down phases are in the LIGO frequency band. On one hand, the expected waveforms for such heavy sources in these regions are not so straightforward to parameterize, making the searches for such systems a larger challenge. Research is ongoing to better characterize such systems. On the other hand, these systems are more difficult to characterize because they probe the crucial strong field limit of general relativity, making such observations of great potential interest.

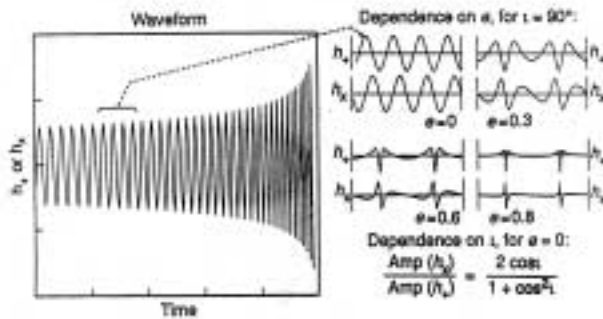


Figure 5. An example is shown of the final chirp waveforms. The amplitude and frequency increase as the system approaches coalescence. The detailed waveforms can be quite complicated as shown at the right, but enable determination of the parameters (eg. ellipticity) of the system.

VI.3 Periodic Signals

Radiation from rotating non-axisymmetric neutron stars will produce periodic signals in the detectors. The emitted gravitational wave frequency is twice the rotation frequency. For many known pulsars, the frequency falls within the LIGO sensitivity band. Searches for signals from spinning neutron stars will involve tracking the system for many cycles, taking into account the

Doppler shift for the motion of the Earth around the Sun, and including the effects of spin-down of the pulsar. Targeted searches of known pulsars and general sky searches are anticipated.

VI.4 Stochastic Signals

Signals from gravitational waves emitted in the first instants of the early universe, as far back as the Planck epoch at 10^{-43} sec, can be detected through correlation of the background signals from two or more detectors. Gravitational waves can probe earlier in the history of the Universe than any other radiation due to the very weak interaction.

Some models of the early Universe can result in detectable signals. Observations of this early Universe gravitational radiation would provide an exciting new cosmological probe.

VI.5 Burst Signals

The gravitational collapse of stars (e.g. supernovae) will lead to emission of gravitational radiation. Type I supernovae involve white dwarf stars and are not expected to yield substantial emission. However, Type II collapses can lead to strong radiation if the core collapse is sufficiently non-axisymmetric. The rate of Type II supernovae is roughly once every 30 years in our own Galaxy. This is actually a lower bound on the rate of stellar core collapses, since that rate estimate is determined from electromagnetic observations and some stellar core collapses could give only a small electromagnetic signal. The ejected mantle dominates the electromagnetic signal, while the gravitational wave signal is dominated by the dynamics of the collapsing core itself.

Numerical modeling of the dynamics of core collapse and bounce has been used to make estimates of the strength and characteristics. This is very complicated and model dependent, depending on both detailed hydrodynamic processes and the initial rotation rate of the degenerate stellar core before collapse. Estimating the event detection rate is consequently difficult and the rate may be as large as many per year with initial LIGO interferometers, or less than one per year with advanced LIGO interferometers. Probably a reasonable guess is that the initial detectors will not see far beyond our own galaxy, while an advanced detector should see out to the Virgo cluster.

The detection will require identifying burst like signals in coincidence from multiple interferometers. The detailed nature of the signal is not well known, except that it is burst like and is emitted for a short time period (milliseconds) during the actual core collapse. Various mechanisms of hang-up of this collapse have been considered and could give enhanced signatures of collapse.

Burrows et al have calculated the gravitational wave signal, taking into account the detailed hydrodynamics of the collapse itself, the typical measured recoil neutron star velocities and the radiation into neutrinos.

VII Status and plans

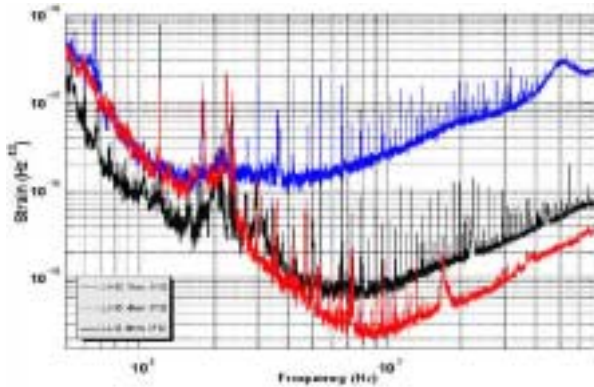


Figure 6. The sensitivity of each of the three LIGO interferometers for the first coincidence engineering run of the the LIGO interferometers.

The construction of LIGO was completed in 2000 and we are now in the process of commissioning the detectors. Since we have three interferometers to commis-

sion, we are using one of them (the 2km interferometer at Hanford) as the pathfinder and bringing the other two interferometers into operation, using the “lessons learned” on that interferometer. We now have made all three interferometers lock for significant time periods in the recombined mode and for the Hanford 2 km, we have locked for hours at a time in the final optical configuration - a power recycled Michelson interferometer with Fabry-Perot arm cavities.

Recently we took our first data-taking run of two weeks with all three interferometers operating in coincidence. The run was quite successful as we recorded in that we recorded about ~ 45 hours of data with all three interferometers locked and operating for a long stretches of time but not yet at high sensitivity (see Fig. 6). This will allow us to use this data to carry out our first end-to-end analysis of the data and prepare for the first science data run, which we expect to perform this coming summer. That run will be at reduced sensitivity, but will allow us to set new limits on several of the sources discussed above. Over the coming year, we will concentrate on sensitivity improvements that should bring us closer to the expected sensitivity required to detect astrophysical sources of gravitational waves.