

LECTURE NOTES ON GRAVITATIONAL WAVES: THEORY AND OBSERVATIONS

February 2023

CHRIS VAN DEN BROECK¹

Institute for Gravitational and Subatomic Physics
Department of Physics, Utrecht University



¹ c.f.f.vandenbroeck@uu.nl

Contents

I. Introduction	5
II. Special and general relativity	18
A. Newton's theory of gravity	18
B. Special relativity	19
C. Towards general relativity	25
III. Metrics and geodesics	29
A. Metrics: Some examples	29
B. Tensors	34
C. Metrics and coordinate transformations	36
D. Light cones and geodesics	39
IV. The Einstein field equations	45
A. The energy-momentum tensor	45
B. Tensor calculus	48
C. Geodesics revisited	50
D. Curvature and the Riemann tensor	53
E. The Einstein equations	55
V. The dynamics of spacetime	59
A. Gravity according to Newton (again)	59
B. “Linearized” general relativity	61
C. No instantaneous action at a distance!	65
D. Intermezzo: The Newtonian limit	66
E. Physical degrees of freedom in gravitational waves	69
VI. What are gravitational waves?	72
A. The effect of gravitational waves on matter	72
B. The quadrupole formula	74
C. Energy carried by gravitational waves	78
VII. The coalescence of compact binary objects	81

A.	Two point particles on fixed circular orbits	82
B.	Inspiral	84
VIII.	How are gravitational waves detected?	88
A.	The response of an interferometer to a gravitational wave	88
B.	Finding a signal in a lot of noise	91
C.	The properties of the noise	93
D.	The optimal filter	95
E.	The detection statistic	97
IX.	The inspiral of binary neutron stars and black holes: The detector point of view	108
A.	The detector response in the stationary phase approximation	108
B.	What information about the source is contained in the signal?	113
X.	Inferring the properties of the source from the signal	120
A.	How much information can be extracted from the signal?	120
B.	Gravitational wave parameter estimation in real life: The nested sampling algorithm	126
XI.	What has been detected so far? A census of black holes and neutron stars	133
A.	A census of black hole and neutron star masses	133
B.	Towards a census of black hole spins	140
XII.	Fundamental physics with gravitational waves	142
A.	The fundamental physics of gravity	142
B.	Probing the physics of black holes	147
C.	What do neutron stars look like on the inside?	150
XIII.	Gravitational waves and cosmology	155
XIV.	Gravitational wave background radiation	165
A.	Characterization of stochastic gravitational wave backgrounds	167
B.	Searching for stochastic backgrounds with a single detector	169

I. INTRODUCTION

Einstein’s theory of general relativity treats gravity not as a force, but as a side-effect of the curvature of spacetime. In this theory, the reason why the Earth stays in orbit around the Sun is not that the Sun exerts a force on the Earth, but rather that the Sun curves spacetime in its vicinity, which the Earth is compelled to follow; see Fig. 1. Paraphrasing John Archibald Wheeler, “matter tells spacetime how to curve, and spacetime tells matter how to move.”

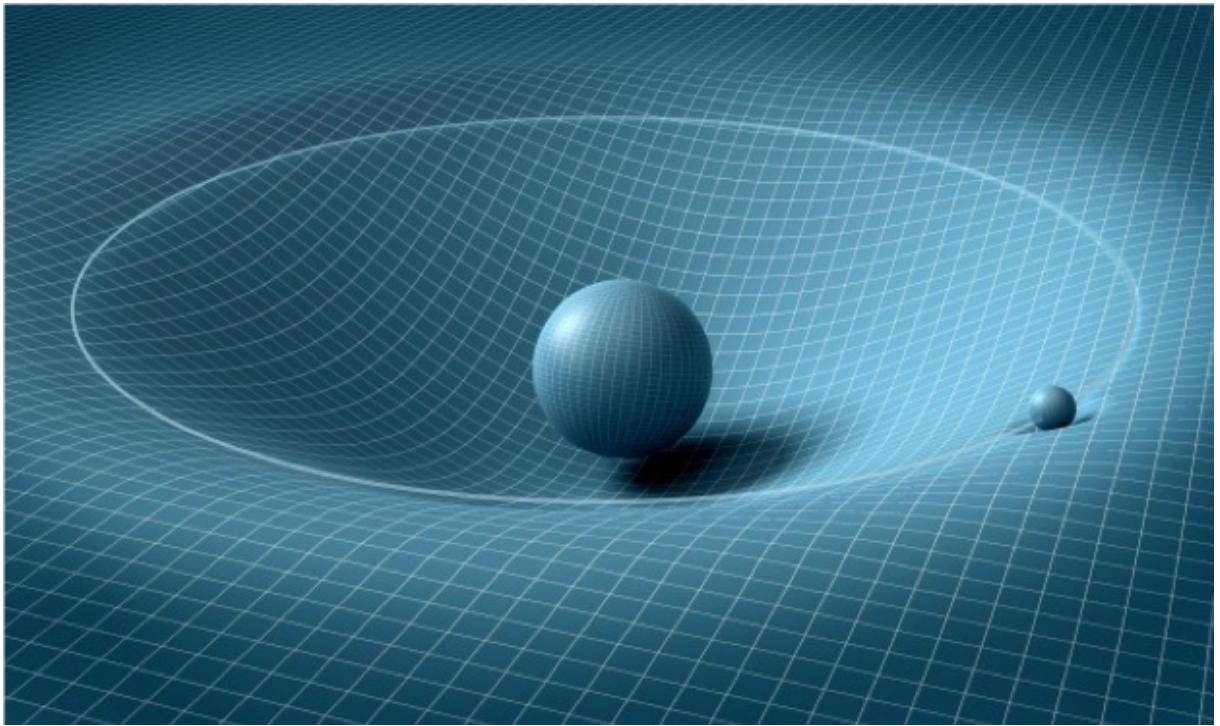


Figure 1: A “rubber membrane” analogy for the effect of spacetime curvature. The Sun curves spacetime in its vicinity, and the reason the Earth stays in orbit around it is that it is compelled to follow that curvature.

The more compact an object is – the smaller it is for a given mass – the more it curves spacetime. This is illustrated in Fig. 2. An example of a star that is more compact than the Sun is a *white dwarf*; these are objects with about one Solar mass (denoted M_{\odot}), but a radius comparable to that of the Earth. They are the end points of stars like the Sun after they have spent their nuclear fuel. These are prevented from further collapse by electron degeneracy pressure, which follows from quantum mechanics: electrons are fermions, so that no two of them are allowed to occupy the same quantum state. Bigger stars end their lives

as *neutron stars*, which typically weigh a few solar masses, but are only the size of a city. Such objects are created when the pressure becomes so high that electrons are pushed into the atomic nuclei, combining with protons to form neutrons, until for the most part only neutrons are left. Increasing the compactness even further, one gets a *black hole*: an object that has undergone complete gravitational collapse, so that all the matter has ended up in a singularity, surrounded by a horizon from which nothing, not even light, can escape.

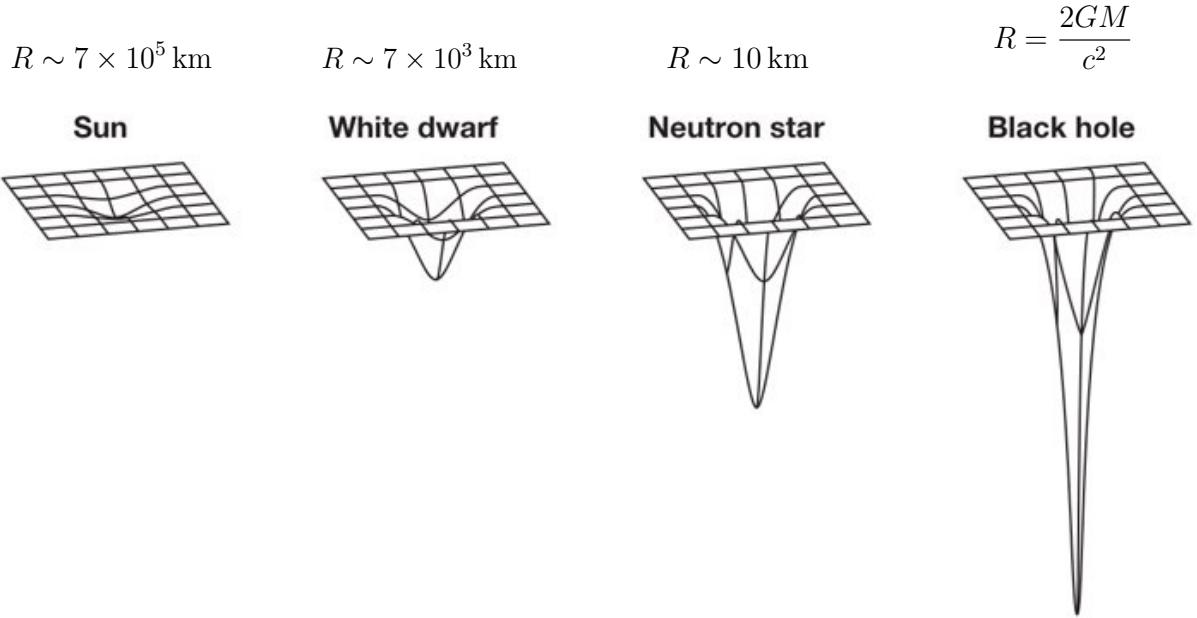


Figure 2: The more compact an object, the more it curves spacetime in its vicinity. Shown are different kinds of stars with their typical radii.

The reason Einstein started working on a theory of gravitation was a discrepancy between his theory of special relativity of 1905, and Newton's theory of gravity. As an illustration, imagine some large object hitting the Sun and knocking it out of the solar system. In Newton's theory, the gravitational force is transmitted immediately and without time delay, so that the Earth would straight away leave its current orbit. However, in special relativity, nothing can travel faster than the speed of light, not even information. Einstein's general theory of 1915 provided the solution. Should the Sun be removed from the solar system, then spacetime in the neighborhood of where it used to be must rearrange its curvature to account for this fact. This causes ripples in spacetime to spread out at the speed of light,

reaching the Earth only 8 minutes later. These ripples, called *gravitational waves*, were predicted by Einstein in 1916; see Fig. 3.

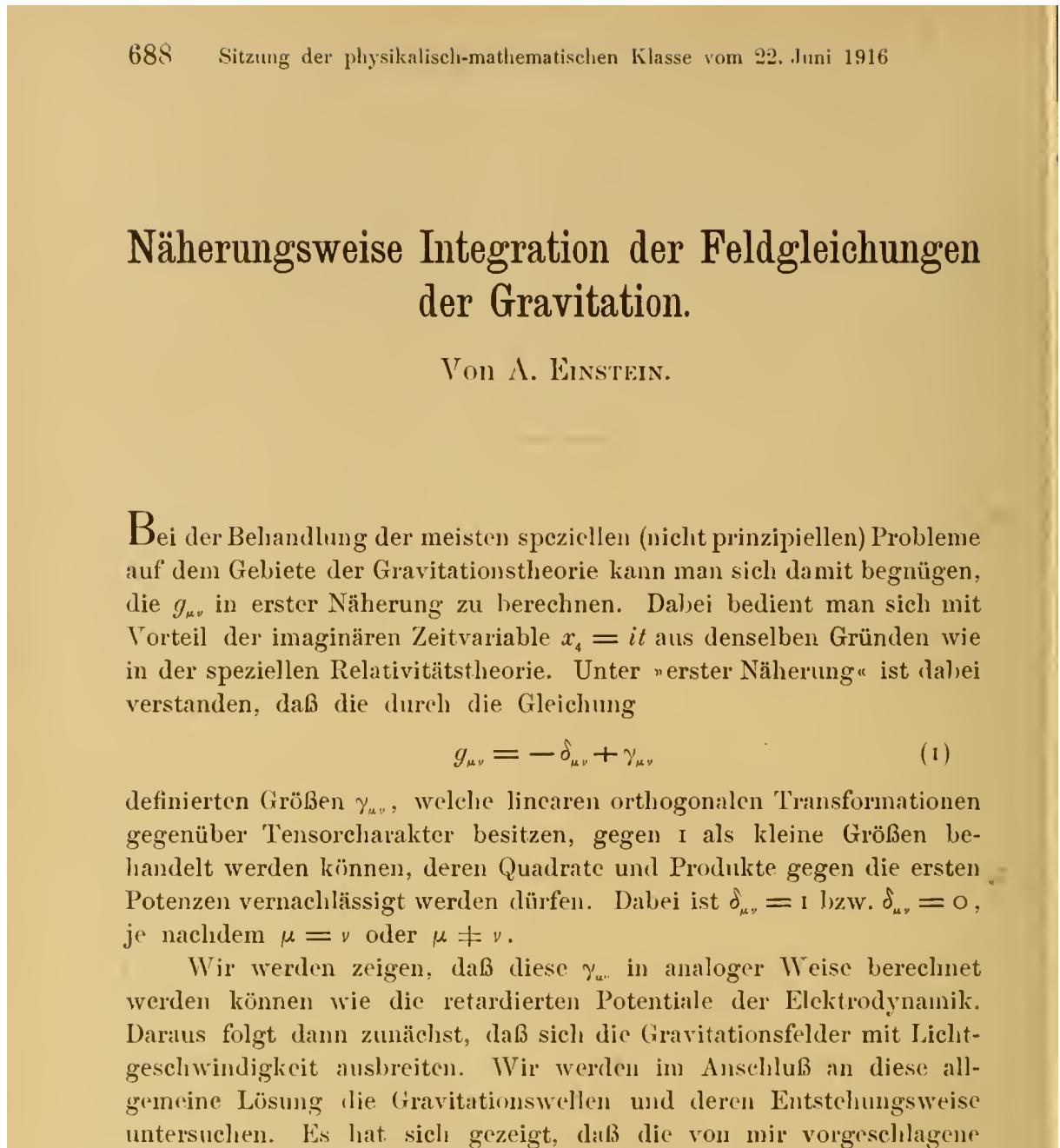


Figure 3: Einstein's 1916 paper “*Näherungsweise Integration der Feldgleichungen der Gravitation*” (*Approximative Integration of the Field Equations of Gravitation*), in which he predicted the existence of gravitational waves. In the coming lectures, we will derive all the important equations in this article!

The nature of gravitational waves is illustrated in Fig. 4. In each panel, a gravitational wave is propagating along the z axis. What a gravitational wave does is to periodically “stretch” and compress space in directions perpendicular to the propagation direction. This can happen in two ways, called the “plus” and “cross” polarizations because of the orientation of the stretching patterns. In reality, gravitational waves usually contain a mixture of these two polarizations.

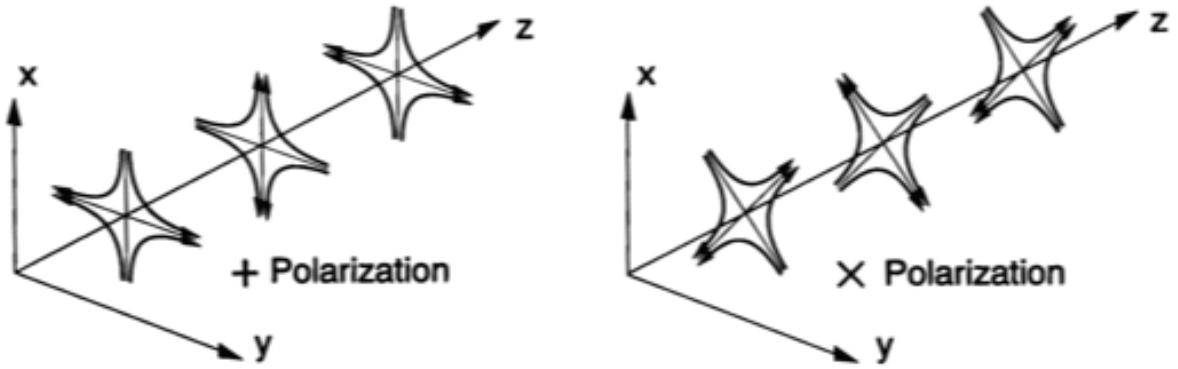


Figure 4: Gravitational waves propagating in the z direction cause a periodic “stretching” and “squeezing” of space in the xy plane. Shown are two possible stretching patterns, denoted as the “+” and “ \times ” polarizations.

Einstein calculated gravitational waves to be very weak, and concluded that in practice they would probably never be observed. In this he mis-judged both human ingenuity and the violence that the Universe can exhibit. In fact, indirect observational evidence for the existence of gravitational waves already started emerging less than 60 years after Einstein’s prediction. In 1974, radio astronomers Russell Hulse and Joseph Taylor discovered a *binary* neutron star: two neutron stars orbiting around a common center of gravity, with the very short orbital period of 7.75 hours. Subsequent measurements showed that the orbital period was decreasing over time by a tiny but nevertheless measurable amount (76.5 microseconds per year). One hypothesis was that the binary system was causing the emission of gravitational waves, which were taking away orbital energy, causing the orbits to shrink. The system was monitored over the course of decades, and the orbital decay indeed showed remarkable agreement with the prediction from general relativity under the assumption that gravitational waves were to blame (Fig. 5).

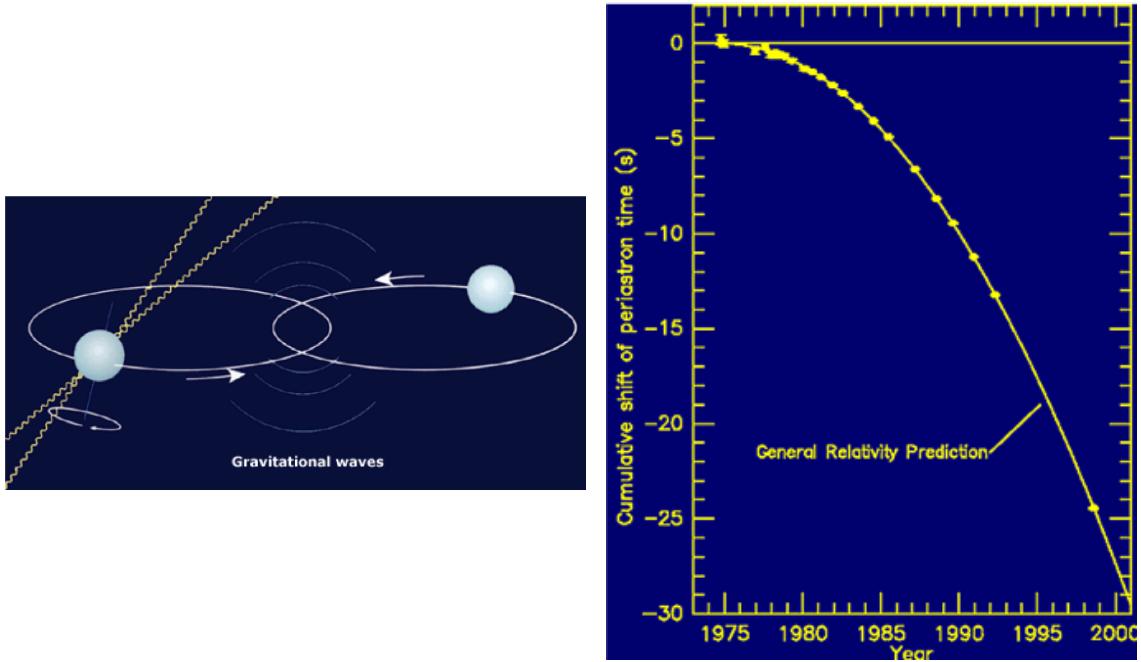


Figure 5: Left: A depiction of the binary neutron star discovered by Hulse and Taylor. One of the neutron stars can be seen as a *pulsar*, sending regular radio pulses towards Earth. From the Doppler modulation of the pulses one obtains the orbital period of the binary. Right: The orbital period of the binary is slowly decreasing, the dots being data points taken over the course of decades. The curve through the dots is not a fit to the data, but the prediction of general relativity assuming that gravitational waves are carrying energy away from the system.

Even before the discovery of Hulse and Taylor, physicists had been thinking about how to *directly* detect gravitational waves. The first attempts involved metal bars of a few meters in size. The idea was that if a gravitational wave came by, their stretching and squeezing of space would also stretch and squeeze the bar, making it vibrate. Moreover, if the oscillation frequency of the gravitational wave happened to be close to a resonant frequency of the bar, then the bar would shake more strongly, and hopefully in a detectable way; see Fig. 6. By 1969, Joseph Weber in the USA was claiming regular detections of gravitational waves, sometimes several times a day. Other researchers in Europe, Japan, and Russia were able to replicate his experimental setup, but they did not see any signals. It is now generally accepted that Weber was mistaken in the interpretation of his data.

It was recognized early on that in any case, bar detectors were not the most promising tool for gravitational wave detection. Apart from the fact that they could only find gravitational

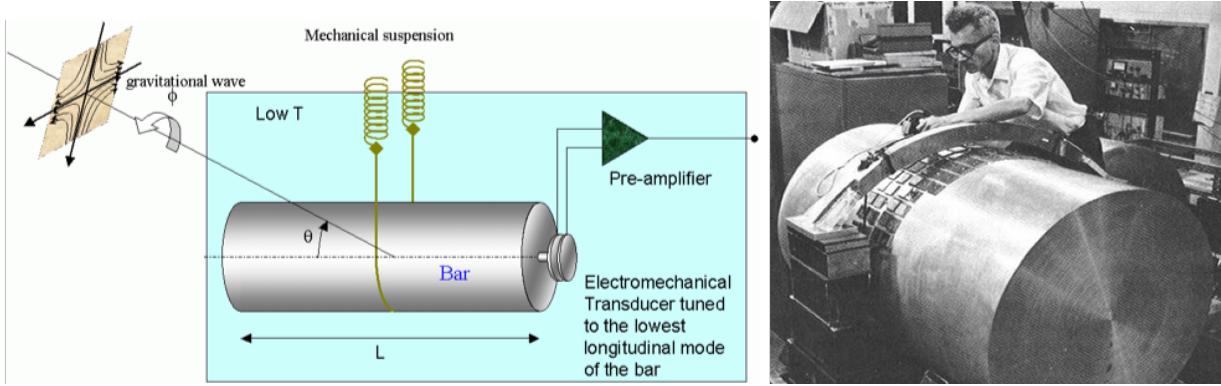


Figure 6: Left: When a gravitational wave passes through a metal bar, the stretching and squeezing of space causes the bar to vibrate. If the frequency of oscillation of the gravitational wave is close to a resonant frequency of the bar then the effect is amplified. Right: Joseph Weber at work on resonant bar detectors in his laboratory in the 1960s.

waves whose frequency was close to a resonant frequency of the bar, it was hard to see how they would be able to reach the requisite sensitivity given how tiny the spacetime vibrations were expected to be (of which more below). A different idea was the *laser interferometer*; see Fig. 7. In it, a laser beam is split in half, with each half traveling down a long tube. At the ends of the tubes there are mirrors which reflect the beams back. Things can be set up such that if nothing special is happening, upon return the beams are out of phase by half a wavelength, causing destructive interference, so that no light is seen at the output. On the other hand, if a gravitational wave passes by then periodically one arm gets stretched and the other compressed, causing a length difference between the arms. However, the laser beams still travel at the speed of light, so that there is increased travel time up and down one arm compared to the other. This causes the wave crests from one returning beam to be misaligned with the wave troughs from the other beam. (In addition, the wavelength of laser light in the two arms will have become slightly different.) This ruins the destructive interference, and flashes of light are seen to come out of the interferometer.

In 1971, Rainer Weiss made the first study of what it would take to make the idea work in practice. With much foresight, he realized that a variety of noise sources would adversely affect the operation of the device. At low frequencies (tens of Hertz) there would be seismic noise due to people walking nearby, cars driving by in the distance, or planes flying overhead. At middle frequencies (a few hundred Hz) the thermal vibrations of atoms

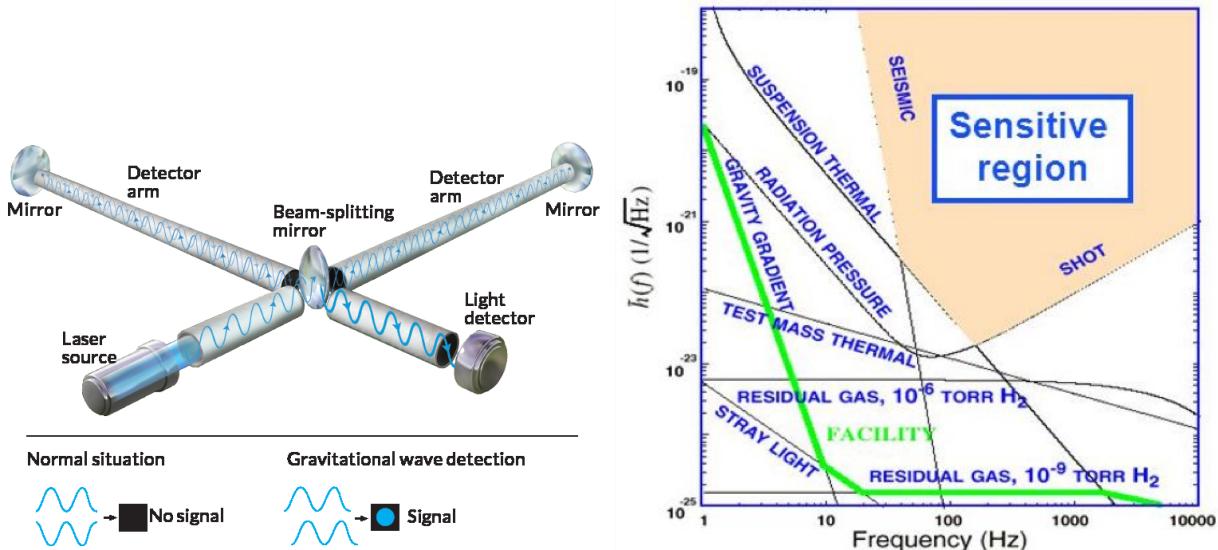


Figure 7: Left: A cartoon picture illustrating the workings of a laser interferometer (see the main text). Right: Already in the 1970s, Rainer Weiss realized what would be the main sources of noise limiting the operation of a large interferometer, at different frequencies.

in the suspending wires of the mirrors would be the main source of noise. At high frequencies, quantum mechanics would be spoiling things: laser light consists of photons, and the number of photons hitting the mirrors per unit of time will fluctuate, again causing unwanted motion of the mirrors. Weiss also understood how the various noise sources could to some extent be suppressed, probably to the level where the detection of gravitational waves would become realistic. He knew that a staggering accuracy needed to be achieved: one would need the ability to detect relative changes in interferometer arm lengths at the level of

$$\frac{\Delta L}{L} \lesssim 10^{-22}. \quad (1.1)$$

Even for arm lengths of $L = 1$ km, this implies $\Delta L \lesssim 10^{-19}$ m, or ten thousand times smaller than the size of a proton!

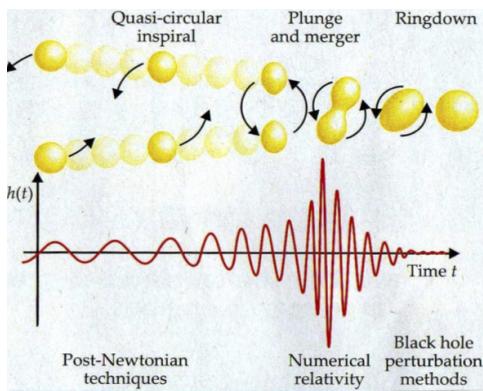
Physicists knew what accuracies were needed because calculations were being done to estimate the strengths of gravitational waves arriving on Earth from a variety of astrophysical sources, taking into account the strength of emission at the source as well as typical distances from Earth. Four types of sources were identified, which are illustrated in Fig. 8:

- Mergers of two neutron stars or black holes are expected to be relatively rare, but if one can “see” out to extragalactic distances then they can be detected; this is now

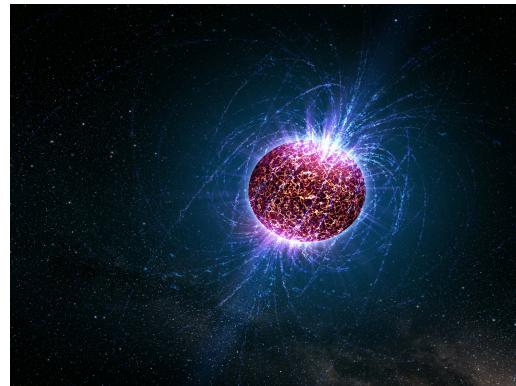
being done on a regular basis.

- Also a single neutron star can be an eligible source, if it is sufficiently asymmetric and fast-spinning, and within the Milky Way galaxy.
- Supernovae, again if they are sufficiently asymmetric and located in or near our own galaxy.
- A *primordial* background of gravitational waves, left over from right after the Big Bang.

Merging neutron stars, black holes



Fast-spinning neutron stars



Supernovae



Primordial gravitational waves

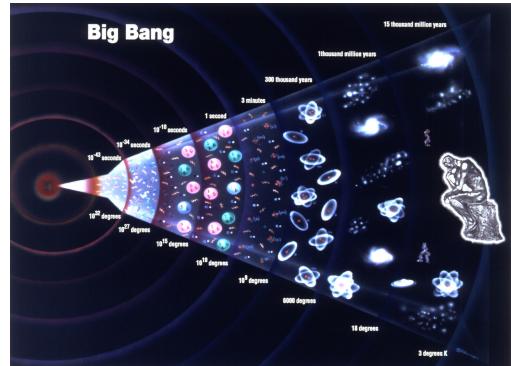


Figure 8: Different types of sources of detectable gravitational waves.

The detection of gravitational waves from the latter source would be a momentous event. Unlike the (electromagnetic) cosmic microwave background which originated a few hundred

thousand years after the Big Bang, most mechanisms for primordial gravitational waves cause them to be generated a fraction of a second after the beginning of our Universe. Should we detect such a background, we would have direct access to phenomena at energies of $10^6 - 10^7$ TeV (tera-electronvolt); for comparison, the Large Hadron Collider at CERN has access to 14 TeV.

A worldwide network of interferometric detectors is now in place, as shown in Fig. 9. Unlike in other fields of physics or astronomy, there is no competition between the different experiments: data from all the operational detectors is being shared, and analyzed jointly. This is because of the important scientific advantages of having a *network* of detectors:

- The more detectors a weak signal is seen in, the higher the confidence that it is real, as opposed to some accidental instrumental noise artefact.
- Some sources of gravitational waves, like binary neutron star mergers, also emit electromagnetic radiation. A network of at least three interferometers allows us to “triangulate” and find the approximate position of the source on the sky. In 2017 this enabled astronomers to find the electromagnetic counterpart of such a merger.
- With a network of detectors, it is possible to separately measure the “plus” and “cross” polarizations of Fig. 4. As we will see later in these lectures, among other things this leads to the remarkable ability to measure the distance to the source *directly from the gravitational wave signal*. In the near future this may have far-reaching consequences for cosmology.

The first direct detection of gravitational waves happened on 14 September 2015. The raw data are shown in Fig. 10. The nature of the event was understood using techniques that we will cover in these lectures. The signal came from the coalescence of two black holes, each about 200 kilometers wide and carrying tens of solar masses, spiraling towards each other and colliding at 60% of the speed of light, thereby emitting more than 3 solar masses worth of energy in gravitational waves. The waves then traveled for over a billion years before reaching the Earth, at which they interacted with the two LIGO detectors for only half a second, and then continued their journey through space.

Since then more detections have been made, perhaps the most notable one being the binary neutron star merger seen in 2017. Less than 2 seconds after the gravitational wave

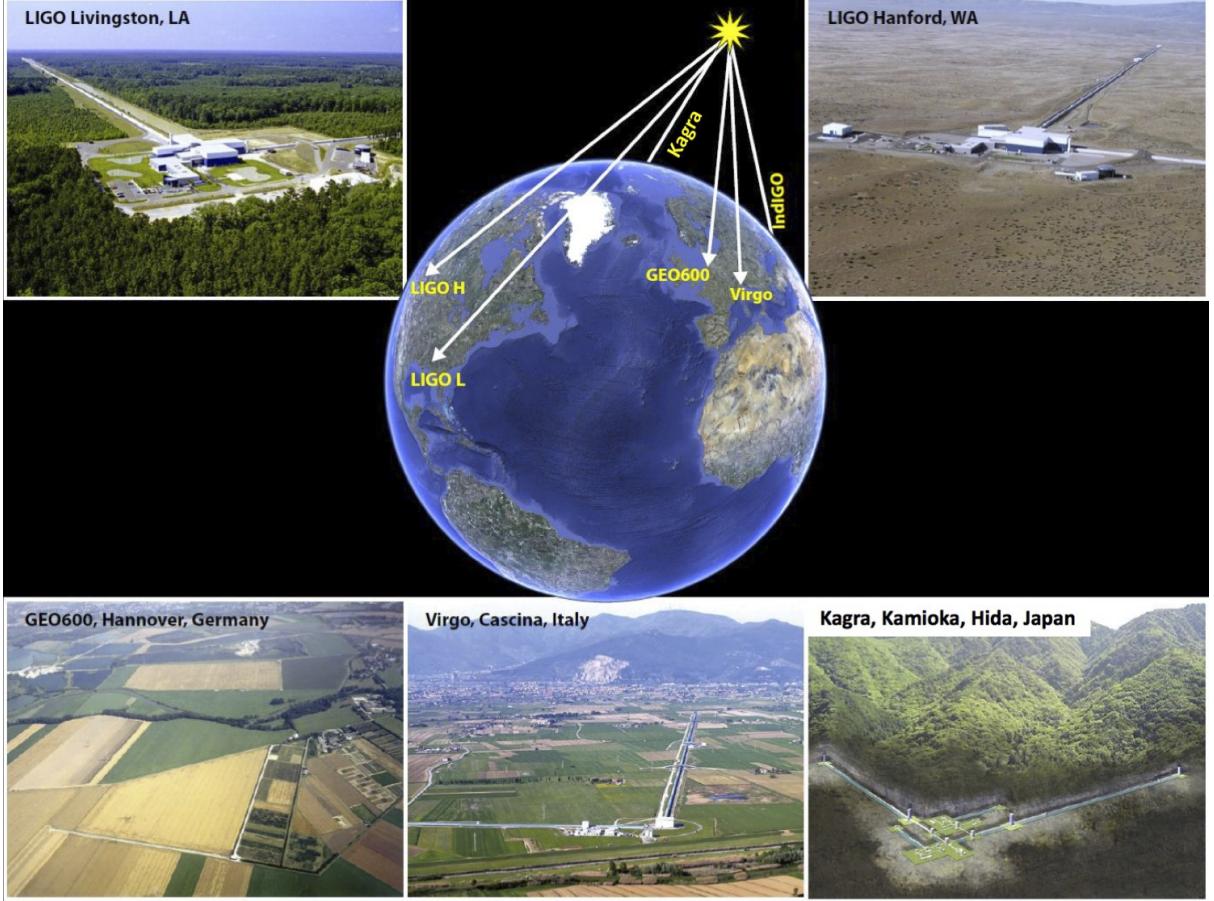


Figure 9: The global network of gravitational wave detectors. The two LIGO detectors at Livingston, Louisiana and Hanford, Washington (both with 4 km arm length), together with Virgo in Italy (3 km arms), have already been making detections. KAGRA in Japan (3 km arms) has been operational, but further upgrades are needed. A LIGO-like detector (4 km arms) will be built in India in the next few years. GEO in Germany (600 m arms) has been active for a long time now; it is not expected to make detections, but has been instrumental in the development of advanced interferometry techniques that were subsequently applied in the larger interferometers.

signal, satellites in orbit around the Earth recorded a bright flash of gamma rays. It had long been believed that short gamma ray bursts that had been seen since the late 1960s were indeed caused by collisions of two neutron stars; this had now been corroborated. Because three gravitational wave detectors (LIGO Hanford, LIGO Livingston, and Virgo) were operational, the sky position could be determined, and astronomers indeed found the optical afterglow of the event; see Fig. 11.

With the ability to directly detect gravitational waves, a new window on the Universe

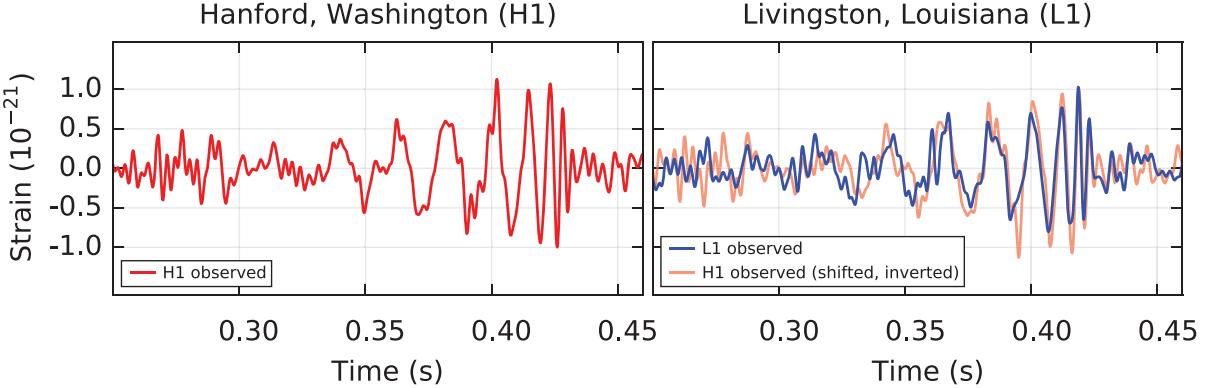


Figure 10: The first direct detection of gravitational waves, emitted by the merger of two black holes. On the left in red and on the right in blue: the signals as seen in, respectively, the LIGO Hanford and LIGO Livingston detectors. In the right panel, the Hanford signal is overlaid, with a 7 millisecond time shift as well as a sign flip to account for, respectively, the travel time of the wave from Livingston to Hanford and the different orientation of the two interferometers. (Picture from <https://arxiv.org/pdf/1602.03837.pdf>.)

has been opened, with repercussions in astronomy, astrophysics, fundamental physics, and cosmology. This wide-ranging scientific impact will be discussed extensively later in these lectures.

To better exploit this new observational field that was created, scientists are already thinking about the new steps in gravitational wave astronomy. The left panel of Fig. 12 shows LISA, the Laser Interferometer Space Antenna. This will consist of three spacecraft in orbit around the Sun at a mutual distance of about 1 million kilometers, exchanging laser beams. Because of its huge size, it will be sensitive to gravitational waves with much larger wavelengths than with ground-based detectors, or equivalently much lower frequencies, in the range $10^{-5} - 10^{-1}$ Hz. Such waves are emitted by objects that are themselves large. Whereas ground-based detectors are sensitive to coalescences of black holes with a few tens of solar masses each, LISA will pick up binary black hole mergers involving millions to billions of solar masses. The LISA mission has been approved by the European Space Agency and is scheduled for launch in 2034.

Finally, the right panel of Fig. 12 shows *Einstein Telescope*. This is envisaged to be an underground facility, to dramatically reduce the seismic noise indicated in Fig. 7. It will also be in the shape of a triangle, this time with sides of 10 km. Einstein Telescope is expected

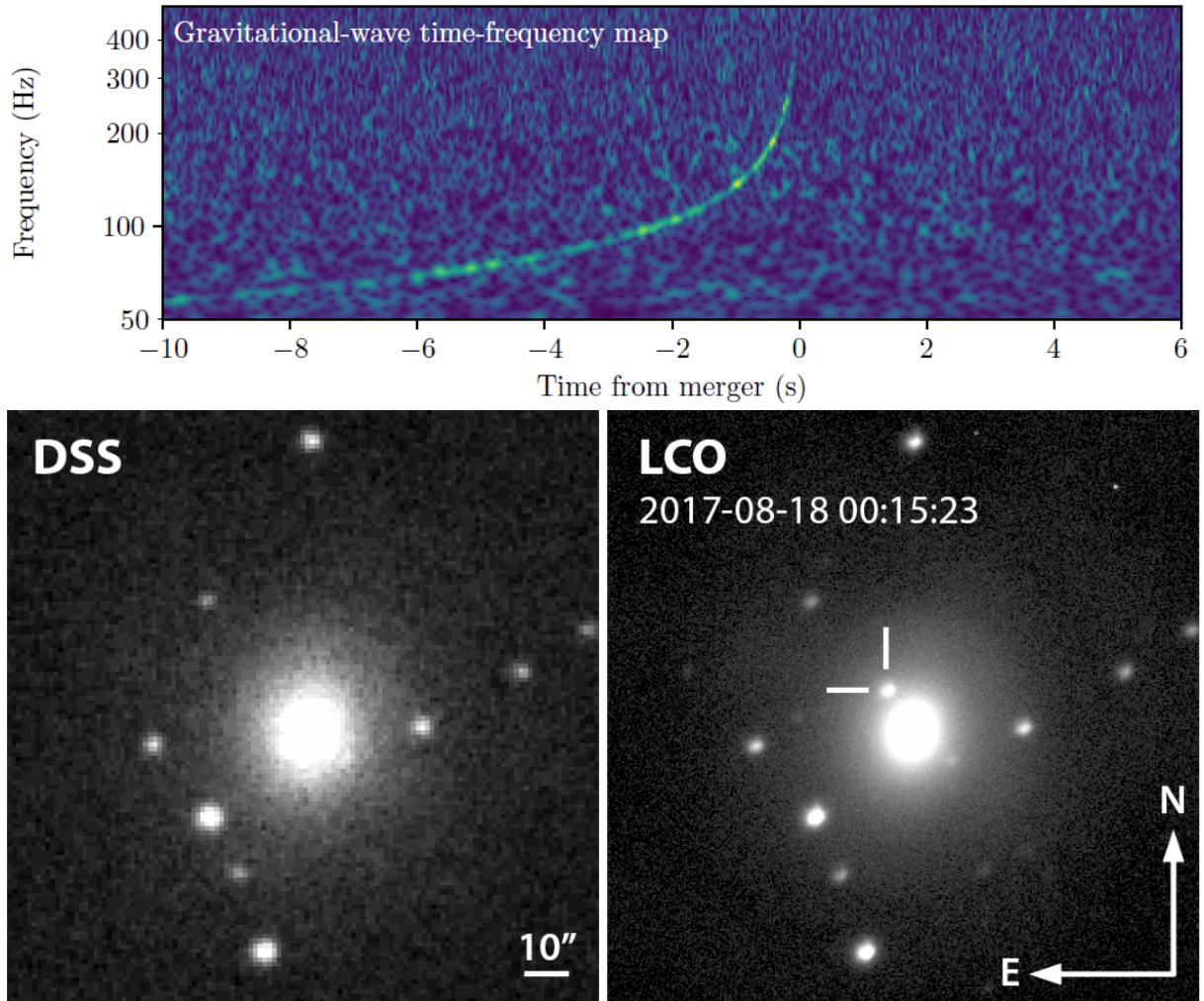


Figure 11: Top: The gravitational wave signal from the binary neutron star inspiral seen in 2017. Shown is frequency as a function of time; the signal is clearly visible in the data as a green track. Bottom: Thanks to the triangulation by the two LIGO detectors and Virgo, astronomers were able to find the optical afterglow. The left panel shows old observations of the galaxy NGC 4993; the right panel shows that the day after the gravitational wave detection, something new had appeared next to it.

to make $\mathcal{O}(10^5)$ detections of binary neutron star and binary black hole coalescences per year, up to redshifts $z > 10$, thus covering the entire cosmic history of such objects. Where Einstein Telescope is to be built has not yet been decided, but after extensive geological studies, the remaining candidate sites are on the one hand the island of Sardinia, and on the other hand the border region of the Netherlands, Belgium, and Germany. A similar ground-based project is being pursued in the USA, called *Cosmic Explorer*. The aim is to

have both Einstein Telescope and Cosmic Explorer operational on the same timescale as LISA, because of the possibility of complementary observations.

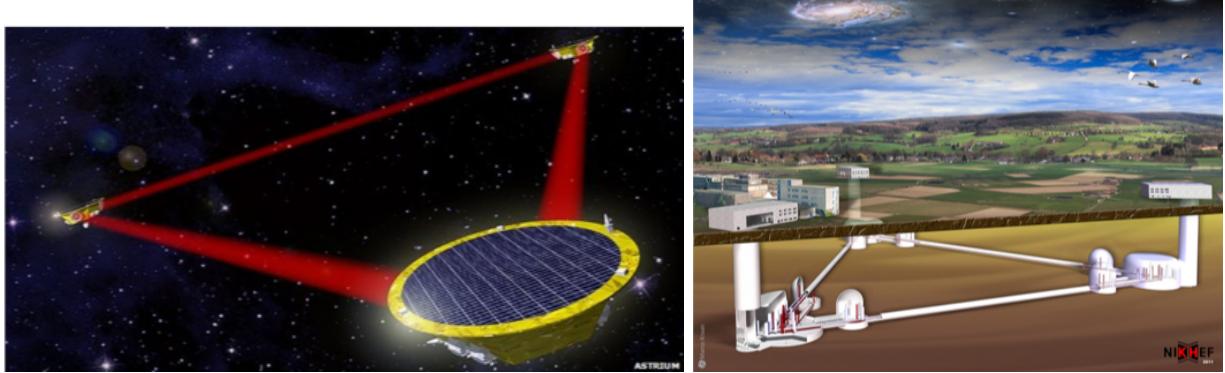


Figure 12: Future detectors are in the planning, and expected to be operational towards the middle of the next decade. Left: The space-based Laser Interferometer LISA. Right: The underground observatory Einstein Telescope, which may be built in the Dutch-Belgian-German border region.

In the next chapters we will address the following questions:

- *What are the dynamics of spacetime?*
- *What are gravitational waves?*
- *What do gravitational waves from binary neutron stars and black holes look like?*
- *How are gravitational waves detected?*
- *What kind of science do gravitational waves enable?*

II. SPECIAL AND GENERAL RELATIVITY

Since gravitational waves are a prediction of Einstein's theory of general relativity, we will need to understand some of the basic concepts that enter into it. Let us do this step-wise, first recalling Newton's theory of gravity, then introducing Einstein's description of what different observers see (special relativity), and finally motivating his theory of gravity (general relativity).

A. Newton's theory of gravity

In Newton's theory, two particles with masses m_g and M_g will act on each other with a gravitational force. If the particles are separated by a distance r , the force on m_g due to M_g is

$$\mathbf{F}_G = -\frac{Gm_g M_g}{r^2} \hat{\mathbf{e}}, \quad (2.1)$$

where G is Newton's constant, and $\hat{\mathbf{e}}$ is a unit vector pointing from M_g to m_g . The magnitude $F_G = |\mathbf{F}_G|$ of this force is then

$$F_G = \frac{Gm_g M_g}{r^2}. \quad (2.2)$$

The quantities M_g and m_g that enter Eq. (2.2) are called the *gravitational masses*, because they set the strength of the gravitational force between the particles under consideration. Given *any* force with magnitude F acting on a particle, the magnitude of the acceleration a it experiences is related to it by

$$F = m_i a, \quad (2.3)$$

where m_i is the *inertial mass*. Given the particle with gravitational mass m_g , if its inertial mass is m_i , then its acceleration due to the gravitational force exerted by M_g is

$$a_G = \frac{F_G}{m_i} = \frac{m_g G M_g}{m_i r^2}. \quad (2.4)$$

Experiments suggest that for any object, the gravitational and inertial mass are equal to each other to great accuracy: $m_g = m_i$, or $m_g/m_i = 1$, in which case the expression above simplifies to

$$a_G = \frac{G M_g}{r^2}. \quad (2.5)$$

Hence, the acceleration of an object due to gravity does not depend on the properties of that object, but only on the mass M_g that is exerting the gravitational force. This is called the *equivalence principle*.

Perhaps we have grown accustomed to gravitational and inertial mass being equal, but the equivalence principle is non-trivial. It certainly doesn't hold for electromagnetic forces; for example, given an electric field of magnitude E acting on a particle with charge q and mass m , the magnitude of the resulting force is $F = qE$, and the particle experiences an acceleration of magnitude

$$a = \frac{qE}{m}. \quad (2.6)$$

Thus, the acceleration due to an electric field *does* depend on the properties of the particle that the force acts on; in fact it explicitly depends both on its charge and its mass. Hence, Newtonian gravity presents us with a puzzle: why should it be the case that the “charge” associated with gravity – the gravitational mass m_g – is equal to the inertial mass m_i ?

As we shall see later on, the equivalence principle is a natural consequence of the theory of general relativity. However, before going there, we need to discuss the more limited theory it was built on, namely special relativity.

B. Special relativity

First recall the concept of an *inertial frame*, which is the coordinate frame of an observer who is not accelerating. Now consider two inertial frames, an “unprimed” one with associated coordinate axes x, y, z , and a “primed” one with axes x', y', z' , as shown in Fig. 13. For simplicity, we take the corresponding primed and unprimed axes to be parallel to each other, but the primed frame is moving in the x direction of the unprimed frame with a constant speed v .

In Newtonian physics, time is universal, so that in particular it flows at the same rate in the primed and unprimed frames. Again for simplicity, let us assume that at $t' = t = 0$, the origins of the two frames coincided with each other. In that case, if particle is at a position with coordinates (x, y, z) with respect to the unprimed frame, then its coordinates

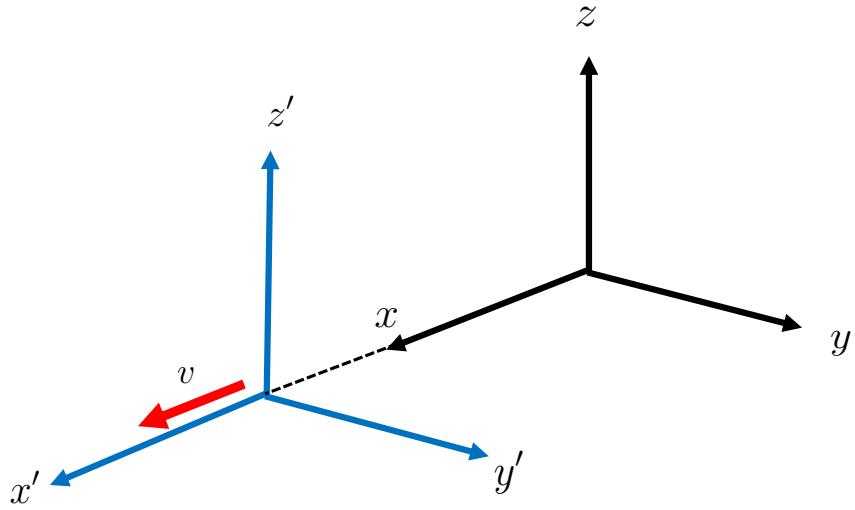


Figure 13: Two inertial frames with coordinate axes (x, y, z) and (x', y', z') , such that corresponding axes are parallel to each other, but the “primed” frame is moving in the x direction of the “unprimed” frame with a constant speed v .

with respect to the primed frame, (x', y', z') are given by

$$\begin{aligned}
 x' &= x - vt, \\
 y' &= y, \\
 z' &= z, \\
 t' &= t.
 \end{aligned} \tag{2.7}$$

This is an example of a *Galilean transformation*. From this also follows how the velocity of a particle in the primed frame, $\mathbf{u}' = (dx'/dt', dy'/dt', dz'/dt')$, is related to its velocity in the unprimed frame, $\mathbf{u} = (dx/dt, dy/dt, dz/dt)$:

$$\begin{aligned}
 u'_x &= u_x - v, \\
 u'_y &= u_y, \\
 u'_z &= u_z.
 \end{aligned} \tag{2.8}$$

Taking another time derivative, one finds that the acceleration of the particle in the two frames is the same: $\mathbf{a}' = \mathbf{a}$.

For three centuries after Galileo formulated these transformations, they were assumed to be the correct way of relating coordinates in inertial frames that are moving with respect to each other. However, towards the end of the 19th century, Einstein and others noticed something strange about the laws of electromagnetism: they don't remain unchanged under Galilean transformations. To make matters worse, an experiment conducted by Michelson and Morley in 1887 appeared to show that the speed of light is the same in all inertial frames!² The latter is inconsistent with Eqs. (2.8); if the particle under consideration is a photon, then in this example its velocities in the primed and unprimed frame are definitely not equal: $\mathbf{u}' = \mathbf{u} - v \hat{\mathbf{e}}_x \neq \mathbf{u}$, where $\hat{\mathbf{e}}_x$ is the unit vector in the x direction.

In an effort to reconcile theory and experiment, Einstein formulated the following postulates of what he called the theory of special relativity:

1. The equations describing the basic laws of physics are the same in all inertial frames.
2. The speed of light in vacuum has the same value in all inertial frames.

Especially the second of these postulates has far-reaching implications. Consider again the two inertial frames of Fig. 13. Let a pulse of light be emitted from the origins of the frames at $t' = t = 0$, when they coincided. Then in the unprimed frame, for a point on a wavefront at position (x, y, z) at a later time $t > 0$ one has

$$c^2 t^2 = x^2 + y^2 + z^2, \quad (2.9)$$

where in the right hand side we have used Pythagoras for the square of the distance from the origin, whereas the left hand side is the square of the distance traveled in a straight line by a photon on the wavefront during the time t , at the speed of light, c . Now consider the same point on the wavefront in the primed frame; then

$$c^2 t'^2 = x'^2 + y'^2 + z'^2. \quad (2.10)$$

Note that there is no prime on c , since according to the second postulate of special relativity, the speed of light is the same in all inertial frames!

² For more information on this interesting experiment, see e.g. https://en.wikipedia.org/wiki/Michelson_Morley_experiment.

Importantly, Eq. (2.9) can not be obtained from Eq. (2.10) through the Galilean transformation (2.7). Rather than trying, let us just state what transformation *does* allow one to pass from one equation to the other, namely a *Lorentz transformation*:³

$$\begin{aligned}x' &= \gamma(x - vt), \\y' &= y, \\z' &= z, \\t' &= \gamma(t - vx/c^2),\end{aligned}\tag{2.11}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}\tag{2.12}$$

is called the Lorentz factor. We leave it as an exercise to show that when substituting these into Eq. (2.10), one indeed recovers Eq. (2.9).

Before going on, we note that in the limit $v \ll c$, when the relative speed of the inertial frames is far below the speed of light, the bizarre-looking Lorentz transformation (2.11) reduces to the seemingly more intuitive Galilean transformation (2.7). Indeed, in that case $\gamma \simeq 1$, and $vx/c^2 = (v/c)(x/c) \simeq 0$. Putting intuition aside, when it comes to *e.g.* the Large Hadron Collider, where massive particles are made to move close to the speed of light, it is the Lorentz transformations, not the Galilean ones, that are required for physicists to make sense of the results of their experiments.

In relativity, it is useful to think of time and space jointly rather than separately: *spacetime*. A point in spacetime – also called a *spacetime event* – is determined by a time t and a spatial position (x, y, z) with respect to some frame, which together are written as (t, x, y, z) . Now consider two spacetime events in our unprimed frame, (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2) . The spatial distance $\Delta\ell$ between the two is as determined by Pythagoras:

$$(\Delta\ell)^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2,\tag{2.13}$$

and the time difference between the two is

$$\Delta t = t_1 - t_2.\tag{2.14}$$

³ The Lorentz transformations are named after the Dutch physicist Hendrik Antoon Lorentz, who formulated them before Einstein, based on his realization that the laws of electromagnetism do look the same after performing such transformations, rather than Galilean transformations.

What will these be in the primed frame? Using the Lorentz transformation of Eqs. (2.11),

$$\begin{aligned} (\Delta\ell')^2 &= (x'_1 - x'_2)^2 + (y'_1 - y'_2)^2 + (z'_1 - z'_2)^2 \\ &= \gamma^2 [x_1 - x_2 - v(t_1 - t_2)]^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2, \end{aligned} \quad (2.15)$$

and

$$\begin{aligned} \Delta t' &= t'_1 - t'_2 \\ &= \gamma \left[t_1 - t_2 - \frac{v}{c^2}(x_1 - x_2) \right]. \end{aligned} \quad (2.16)$$

These expressions look rather different from what we would have obtained from the Galilean transformation (2.7), namely⁴

$$\begin{aligned} (\Delta\ell')^2 &= [x_1 - x_2 - v(t_1 - t_2)]^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2, \\ \Delta t' &= t_1 - t_2. \end{aligned} \quad (2.17)$$

Of particular note is that in the Galilean Eqs. (2.17), there is no restriction on how big v can be, but there is in Eqs. (2.15) and Eqs. (2.16). Looking at Eq. (2.12) for the Lorentz factor, we see that γ diverges as $v \rightarrow c$, and for $v > c$ it is an imaginary number. Since (differences between) spatial and time coordinates must be real numbers, this means that special relativity has a built-in speed limit: *the relative speed between inertial frames must be smaller than the speed of light.*

Though the primed and unprimed observers may see very different things in terms of spatial distances and time differences, in special relativity there does exist a quantity relating spacetime events which is the same for the two observers. In particular, defining $\Delta x = x_1 - x_2$, $\Delta y = y_1 - y_2$, and $\Delta z = z_1 - z_2$, and similarly for the primed coordinates, it is easy to prove (again using the Lorentz transformation (2.11) that

$$-c^2\Delta t'^2 + \Delta x'^2 + \Delta y'^2 + \Delta z'^2 = -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2. \quad (2.18)$$

In fact, this defines a notion of *spacetime distance* Δs , defined by

$$\Delta s^2 \equiv -c^2\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2, \quad (2.19)$$

⁴ In the Galilean case one might have expected $\Delta\ell' = \Delta\ell$, but note that the two events occur at different times, and during the intervening time the primed frame will have moved with respect to the unprimed one. For two events occurring at the same time ($t_1 = t_2$), the primed and unprimed spatial distances are indeed the same.

or in infinitesimal form

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2. \quad (2.20)$$

Note that one can have $ds^2 < 0$, for instance for nearby spacetime events that are at the same position ($dx = dy = dz = 0$) but at different times $dt \neq 0$! Obviously this does not mean that special relativity forbids spacetime events from being separated in this way, since at this point Eq. (2.20) is merely a definition. How to make sense of ds , and in what way it can be interpreted as measuring distances in spacetime, are questions we will return to later; for now it suffices to know that ds^2 is a quantity that takes the same form in any inertial frame: it is invariant under Lorentz transformations.

In the literature, the expression (2.20) is variously called the *metric*, the *line element*, or just the *physical distance*. A notation that will turn out to be useful later in the is

$$(x^0, x^1, x^2, x^3) \equiv (ct, x, y, z), \quad (2.21)$$

where we stress that the superscripts 0, 1, 2, 3 in the left hand side are not meant to be powers, but just labels for the different coordinates. In this notation, for two spacetime events with infinitesimal coordinate separations (dx^0, dx^1, dx^2, dx^3) , the physical spacetime distance (2.20) between the two is

$$ds^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2. \quad (2.22)$$

Note that this can be written as

$$\mathbf{dx}^T \cdot \boldsymbol{\eta} \cdot \mathbf{dx}, \quad (2.23)$$

where the dot denotes matrix multiplication, T is the transpose, and

$$\boldsymbol{\eta} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{dx} = \begin{pmatrix} dx^0 \\ dx^1 \\ dx^2 \\ dx^3 \end{pmatrix}. \quad (2.24)$$

One often uses “index notation”, where for example dx^μ is shorthand for any of the components of \mathbf{dx} . Similarly, the components of $\boldsymbol{\eta}$ are denoted $\eta_{\mu\nu}$, with $\mu, \nu = 0, 1, 2, 3$. The spacetime distance (2.23) can then be written as

$$ds^2 = \sum_{\mu=0}^3 \sum_{\nu=0}^3 dx^\mu \eta_{\mu\nu} dx^\nu \quad (2.25)$$

or

$$ds^2 = \sum_{\mu=0}^3 \sum_{\nu=0}^3 \eta_{\mu\nu} dx^\mu dx^\nu. \quad (2.26)$$

We now introduce the *Einstein summation convention*:

Whenever an index appears twice in the same term, once “up” and once “down”, it should be considered a dummy index which is being summed over.

Hence we may write

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \quad (2.27)$$

where in this example the Einstein convention applies to both μ and ν . As we shall see later, $\eta_{\mu\nu}$ is an example of a *tensor*; in particular it is called a metric tensor. Below we will see that more general metric tensors are possible, but the specific line element associated with $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ is called the *Minkowski metric*.

C. Towards general relativity

Special relativity only considers the family of inertial frames, which are frames that are non-accelerating and at most move with constant velocity with respect to each other. General relativity also allows for accelerated frames. Remarkably, the notion of a metric tensor as introduced in the previous subsection, together with a thought experiment, led Einstein to the idea that spacetime is curved, and that this spacetime curvature is what is responsible for gravity.

Einstein first considered the situation in Fig. 14. If an observer is inside a box that is accelerated upwards by a rocket engine with acceleration \vec{g} , and he drops an apple, then with respect to him the apple will accelerate downward with acceleration $-\vec{g}$. A way to understand this is that the floor of the box “rushes upward” to meet the apple. On the other hand, if the box is sitting on the surface of the Earth and an observer inside it again drops an apple, it will also appear to be accelerating downward with an acceleration $-\vec{g}$. If in both cases the box has no windows and is sufficiently insulated against sound (so that in the first case the observer cannot hear the roaring of the rocket engine), then at least in terms of what an observer sees, there is no difference between the two situations.

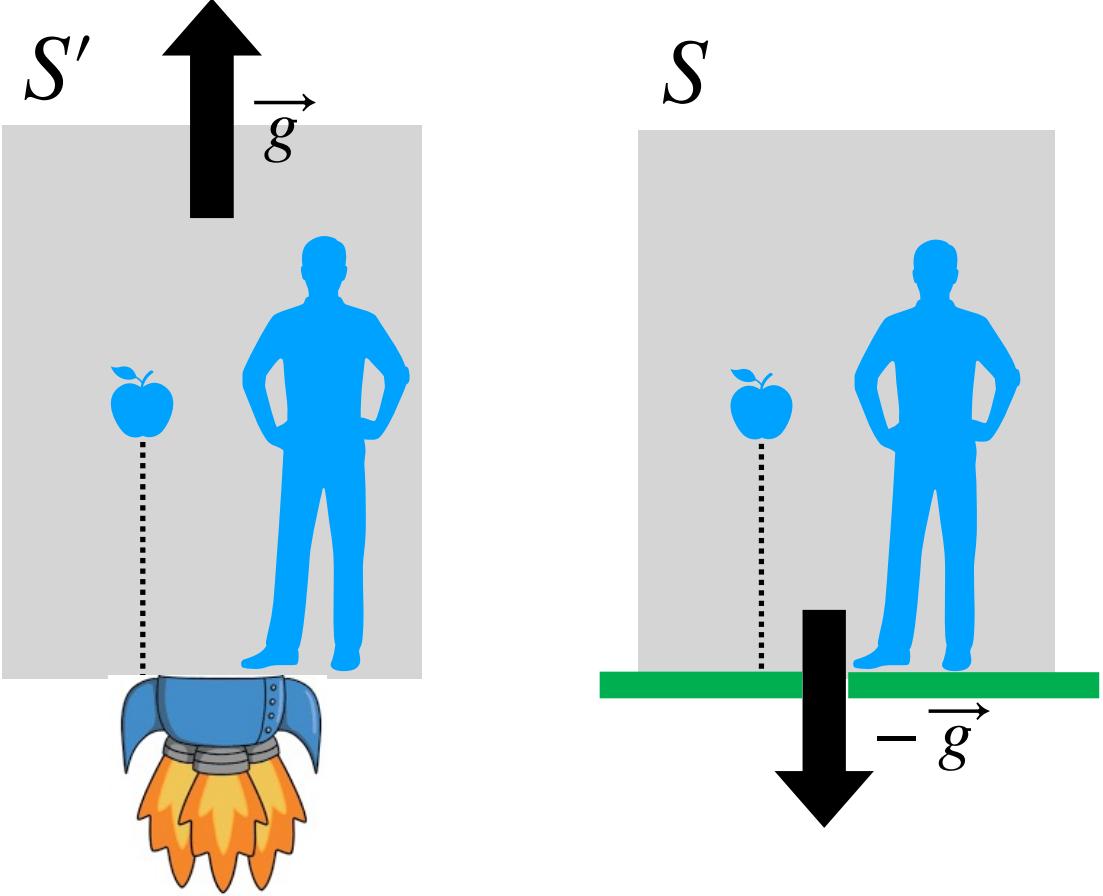


Figure 14: Left: An observer in a box that is being accelerated upwards with acceleration \vec{g} drops an apple, which with respect to him will accelerate to the floor with an acceleration $-\vec{g}$. Right: An observer in a box that is sitting on the surface of the Earth drops an apple, which with respect to him will again accelerate downwards with acceleration $-\vec{g}$.

Next, Einstein thought of the situation in Fig. 15. On the left hand side there is again the observer in a box that is accelerated by a rocket engine, but now he looks at the beam from a flashlight that is embedded in a side-wall. In this case the floor of the box can be viewed as “rushing upward” towards the photons in the light beam, so that the beam will appear to be bending towards the floor. This he again compared with the case of a box sitting on the surface of the Earth, and a light beam being emitted sideways. Einstein’s intuition now told him that also here, there should be no difference between what the observers see. In particular, gravity should be able to bend a beam of light!

Now recall that in the case of special relativity, which deals with inertial observers, photons were assumed to move on straight lines; indeed, in Eqs. (2.9) and (2.10), the left

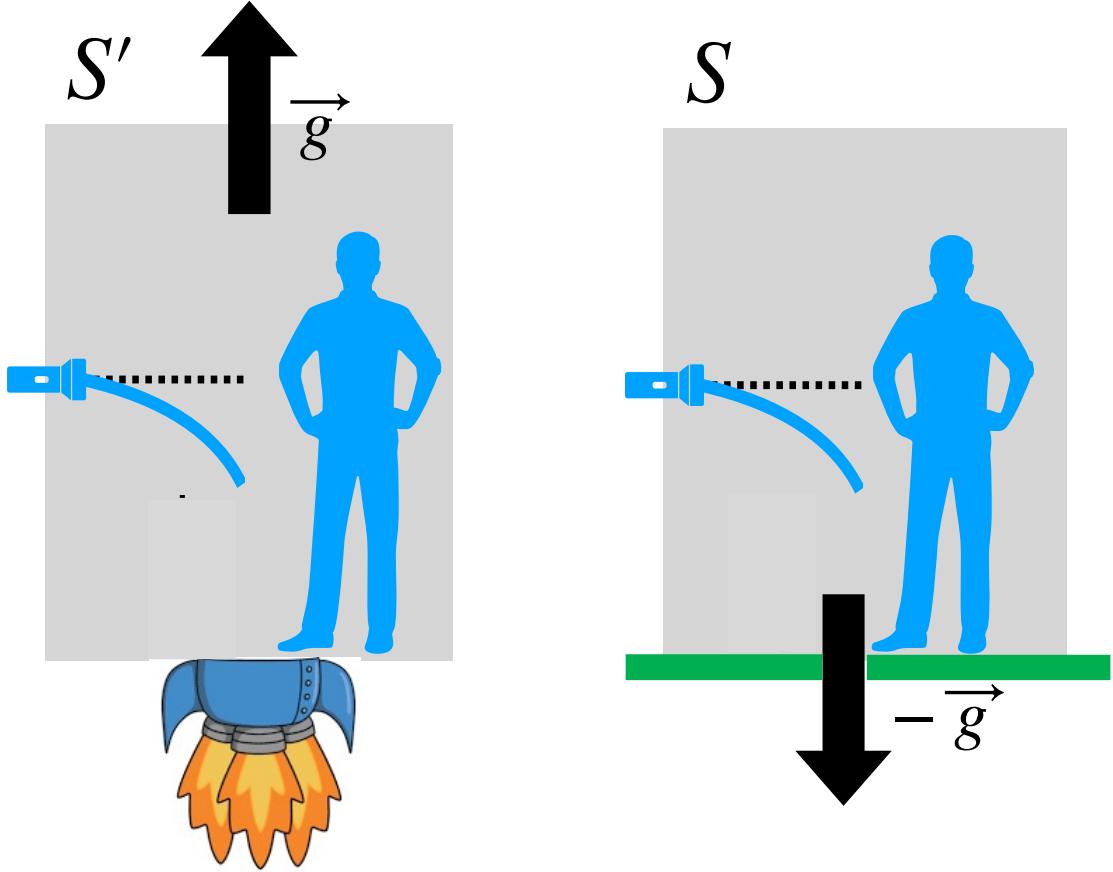


Figure 15: Left: The same situation as in the left hand panel of Fig. 14, but now a flashlight is embedded in a wall of the box. Right: The same situation as in the right hand panel of Fig. 15, but again with a flashlight in the wall.

hand sides are the (square of) the distance traveled by a photon in a time t or t' if it moves on a straight line perpendicular to an expanding wavefront. These equations led us to the Lorentz transformation (2.11); subsequently we realized that the line element $ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu$ is invariant under such transformations.

In the left hand panel of Fig. 15, which shows a non-inertial observer, light does not travel in a straight line with respect to the observer's frame, which suggests that the line element $ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu$ is not the appropriate invariant when dealing with accelerating frames. Now came an important realization: *if a light beam is equally bent in the presence of gravity, as in the right hand panel of Fig. 15, then also when gravity is acting, the line element of special relativity is no longer the quantity of interest.* Instead, Einstein posited

that gravity is associated with a more complicated line element,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (2.28)$$

with $g_{\mu\nu}$ being a different metric tensor from $\eta_{\mu\nu}$. As we shall see in more detail below, metrics different from the Minkowski metric describe curved spacetimes, and the reason why light rays get bent is that they follow the shortest paths possible, taking into account the spacetime curvature. Inspired by the terminology in geography for curves that define the shortest paths between two points on the (curved) surface of the Earth, the paths followed by photons in curved spacetime called *geodesics*. In general relativity, also matter in free fall (such as the apple in Fig. 14) follows a geodesic: “spacetime tells matter and energy how to move”.

General relativity naturally explains the equivalence principle. If particles in free fall follow geodesics, then their motion under the influence of gravity must indeed be independent of their properties.⁵ After all, what geodesics look like is only determined by the curvature of spacetime itself!

However, we will now need a prescription for what the metric tensor $g_{\mu\nu}$ should look like in the presence of a given mass-energy distribution (in Figs. 14 and 15 the mass of the Earth, but of course other arrangements of mass and energy are possible): “Matter and energy tell spacetime how to curve”, but precisely in what way?

The notion that curved spacetime metric tells free-falling particles how to move along geodesics, and the way a mass-energy distribution causes spacetime to curve, together form the cornerstones of the theory of general relativity. In the next two sections, we will make these ideas more concrete.

⁵ The particles we have in mind here are so-called “test particles”: particles that are sufficiently light that their own contribution to the curvature of spacetime can be neglected. For all practical purposes, an apple moving through the curved spacetime in the vicinity of the Earth can be considered a test particle, since almost all of the curvature will be due the mass of the Earth.

III. METRICS AND GEODESICS

A. Metrics: Some examples

Let us first consider a two-dimensional Euclidean plane, which is an example of a space that is *not* curved. In terms of the usual Cartesian coordinates, from Pythagoras we know that the physical distance (*i.e.* the metric) can be written as

$$ds^2 = dx^2 + dy^2, \quad (3.1)$$

where dx and dy are infinitesimal displacements in the x and y directions. Using the notation introduced above, the metric ds^2 can be written in terms of a metric tensor $g_{\mu\nu}$ as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (3.2)$$

although this time the summation is of course just over $\mu, \nu = 1, 2$, and $dx^1 = dx$, $dx^2 = dy$. In Cartesian coordinates, $g_{\mu\nu} = \text{diag}(1, 1)$, *i.e.* the metric tensor can be viewed as a 2×2 matrix with components

$$\mathbf{g} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.3)$$

Cartesian coordinates are but one coordinate system one can use. It is often convenient to introduce polar coordinates (r, θ) through

$$\begin{aligned} x &= r \cos \theta, \\ y &= r \sin \theta. \end{aligned} \quad (3.4)$$

It is easy to express the metric (3.1) in terms of polar coordinates. Indeed, using the chain rule,

$$\begin{aligned} dx &= \frac{\partial r}{\partial r} dr \cos \theta + r \frac{\partial \cos \theta}{\partial \theta} d\theta \\ &= \cos \theta dr - r \sin \theta d\theta, \end{aligned} \quad (3.5)$$

and a similar expression can be obtained for dy . Substituting these into the right hand side of (3.1) and simplifying, it is easy to see that one obtains

$$ds^2 = dr^2 + r^2 d\theta^2. \quad (3.6)$$

Again this can be written in terms of a metric tensor $g'_{\mu\nu}$, where we added a prime to indicate that we are using a different coordinate system from before. With $dx'^1 = dr$ and $dx'^2 = d\theta$, one reads off from Eq. (3.6) that in the new coordinates, $g'_{\mu\nu} = \text{diag}(1, r^2)$, or

$$\mathbf{g}' = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}. \quad (3.7)$$

Although Eqs. (3.3) and (3.7) look different – in the latter case the metric tensor even depends on the coordinate r – it is important to note that they are really the same tensor, just expressed in different coordinates! How metric (and other) tensors transform under general coordinate transformations will be derived later.

Metrics provide a prescription to compute physical distances. Indeed, given a curve \mathcal{C} , the distance along \mathcal{C} between points A and B on the curve is obtained by

$$\ell = \int_A^B \sqrt{ds^2} = \int_A^B \sqrt{g_{\mu\nu} dx^\mu dx^\nu}. \quad (3.8)$$

As an example, consider a straight line from the origin $r = 0$ (point A) to some $r = R$ (point B), for which $\theta = \text{const}$. Then along this line one has $d\theta = 0$, so that from Eq. (3.6) one has $ds^2 = dr^2$, and

$$\ell = \int_A^B \sqrt{ds^2} = \int_A^B \sqrt{dr^2} = \int_0^R dr = R, \quad (3.9)$$

as expected. As a further example, consider a different curve \mathcal{C} , namely a circle centered on the origin with radius R . Integrating all along the circle, this time one has $dr = 0$ and $r = R$, but $d\theta \neq 0$. Hence $ds^2 = R^2 d\theta^2$, and

$$\ell = \int_C \sqrt{ds^2} = \int_C \sqrt{R^2 d\theta^2} = \int_0^{2\pi} R d\theta = 2\pi R, \quad (3.10)$$

again as expected.

The metric of 3-dimensional Euclidean space is a straightforward generalization of Eq. (3.1):

$$ds^2 = dx^2 + dy^2 + dz^2. \quad (3.11)$$

In 3 dimensions, polar coordinates (r, θ, φ) are related to the Cartesian ones by

$$\begin{aligned} x &= r \sin \theta \cos \varphi, \\ y &= r \sin \theta \sin \varphi, \\ z &= r \cos \theta, \end{aligned} \quad (3.12)$$

and it is not difficult to show that

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.13)$$

Of course, the metric (3.13) again describes a space that is not curved. However, using it we can derive the metric of the 2-dimensional surface of a sphere, which *is* curved. Indeed, consider a spherical surface centered on the origin and with radius R ; then by setting $r = R$ and $dr = 0$ in Eq. (3.13),

$$ds^2 = R^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.14)$$

The fact that the above metric indeed *does not describe a flat surface* can be made apparent in a number of ways. For example, consider a circle on the spherical surface, centered on the “North pole” ($\theta = 0$) and determined by $\theta = \alpha$, where α is a fixed angle in radians. To beings that are stuck to the surface and moving in a direction of constant φ , $\sqrt{ds^2} = \sqrt{R^2 d\theta^2} = R d\theta$, so the physical distance they need to walk to get from the North pole to the circle is

$$\rho = \int_0^\alpha R d\theta = R\alpha. \quad (3.15)$$

On the other hand, if they walk all along the circle from $\varphi = 0$ to $\varphi = 2\pi$, then $\sqrt{ds^2} = \sqrt{R^2 \sin^2 \alpha d\varphi^2} = R \sin \alpha d\varphi$, so that they measure the circumference of the circle to be

$$C = \int_0^{2\pi} R \sin \alpha d\varphi = R \sin \alpha \int_0^{2\pi} d\varphi = 2\pi R \sin \alpha. \quad (3.16)$$

The ratio of circumference to radius is

$$\frac{C}{\rho} = 2\pi \frac{\sin \alpha}{\alpha}, \quad (3.17)$$

which in general is not equal to 2π ! If the circle is sufficiently small (*i.e.* small α), one has $\sin \alpha / \alpha \simeq 1$ and $C/\rho \simeq 2\pi$, since a sufficiently small patch of the 2-sphere will appear to be approximately flat. However, generically $C/\rho < 2\pi$. Here we see the curvature of the 2-dimensional spherical surface at work.

Now let us consider an example of a curved four-dimensional spacetime, namely that of a black hole. For a black hole with mass M ,

$$ds^2 = - \left(1 - \frac{2GM}{c^2r}\right) c^2 dt^2 + \left(1 - \frac{2GM}{c^2r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2, \quad (3.18)$$

where t is the time coordinate, and (r, θ, φ) are (polar) spatial coordinates. To simplify this expression somewhat, one can introduce the Schwarzschild radius $r_S = 2GM/c^2$:

$$ds^2 = -\left(1 - \frac{r_S}{r}\right)c^2 dt^2 + \left(1 - \frac{r_S}{r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2. \quad (3.19)$$

Now consider, at $t = 0$, a circle with coordinate radius $r = 100 r_S$, and $\theta = \pi/2$. What is its physical circumference C ? At fixed t we have $dt = 0$, and moving along the circle one also has $dr = d\theta = 0$, i.e.

$$\sqrt{ds^2} = \sqrt{r^2 \sin^2 \theta d\varphi^2} = 100 r_S \sin(\pi/2) d\varphi = 100 r_S d\varphi. \quad (3.20)$$

Therefore the physical circumference is

$$C = \int_{\text{circle}} \sqrt{ds^2} = \int_0^{2\pi} 100 r_S d\varphi = 2\pi \times 100 r_S. \quad (3.21)$$

Hence, at least in the coordinate system used here, the physical circumference of the circle happens to be related to its coordinate radius r as it would in flat spacetime. However, *the radial coordinate r does not equal physical distance in the radial direction*. First note that for $r \gg r_S$, one has

$$\left(1 - \frac{r_S}{r}\right)^{-1} \simeq 1 + \frac{r_S}{r}, \quad (3.22)$$

so that the metric (3.19) can be approximated as

$$ds^2 \simeq -\left(1 - \frac{r_S}{r}\right)c^2 dt^2 + \left(1 + \frac{r_S}{r}\right) dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2. \quad (3.23)$$

Suppose that this time we move not along a circle, but purely in the r direction so that $dt = d\theta = d\varphi = 0$. Then

$$\sqrt{ds^2} = \sqrt{\left(1 + \frac{r_S}{r}\right) dr^2} \simeq \left(1 + \frac{1}{2} \frac{r_S}{r}\right) dr, \quad (3.24)$$

where we used that $\sqrt{1+x} \simeq 1 + \frac{1}{2}x$ if $x \ll 1$. Moving *e.g.* from $r = 100 r_S$ to $r = 101 r_S$ then implies a physical distance

$$\begin{aligned} & \int_{100 r_S}^{101 r_S} \left(1 + \frac{1}{2} \frac{r_S}{r}\right) dr \\ &= 101 r_S - 100 r_S + \frac{r_S}{2} [\ln(101 r_S) - \ln(100 r_S)] \\ &= r_S \left[1 + \frac{1}{2} \ln\left(\frac{101}{100}\right)\right], \end{aligned} \quad (3.25)$$

where in the second line we have used that

$$\int \frac{1}{x} dx = \ln x + \text{const}, \quad (3.26)$$

and in the third line that $\ln(x) - \ln(y) = \ln(x/y)$. Thus, if one traverses a *coordinate* distance $101r_S - 100r_S = r_S$, the physical distance traversed is slightly larger than that, by a fraction $(1/2)\ln(101/100) \simeq 5 \times 10^{-3}$. However, the farther away one is from the black hole, the smaller the effect becomes: when going from $1000r_S$ to $1001r_S$, one has $(1/2)\ln(1001/1000) \simeq 5 \times 10^{-4}$. Indeed, at larger and larger values of r , the metric tends towards the flat metric: for $r \rightarrow \infty$,

$$ds^2 \rightarrow -c^2 dt^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2. \quad (3.27)$$

The latter is indeed just the Minkowski metric with its spatial part expressed in polar coordinates.

Let us look at one more example, this time from cosmology, to further illustrate how coordinates are but labels for spacetime events, and to compute physical distances one needs to use the metric. At sufficiently large scales, the Universe as a whole can be described using the Friedmann-Robertson-Walker (FRW) metric:⁶

$$ds^2 = -c^2 dt^2 + a^2(t) [dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2)], \quad (3.28)$$

where $a(t)$ is an increasing function of time, called the scale factor. To better understand the FRW metric, first note that at a given time t , the spatial part is proportional (by a factor $a(t)$) to that of a 3-dimensional Euclidean space:

$$d\ell^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (3.29)$$

The idea is now that the galaxies in the Universe *always remain at the same spatial coordinates*, but the *physical distance* between them keeps increasing due to the increasing function $a(t)$.

The set-up is illustrated in Fig. 16. An arbitrary galaxy is chosen to be the origin $r = 0$ of the spatial coordinate system. Other galaxies always remain at fixed coordinates, so that

⁶ Here we show the “spatially flat” Friedmann-Robertson-Walker metric, which is actually but one choice of metric to describe a Universe that is homogeneous and isotropic at large scales.

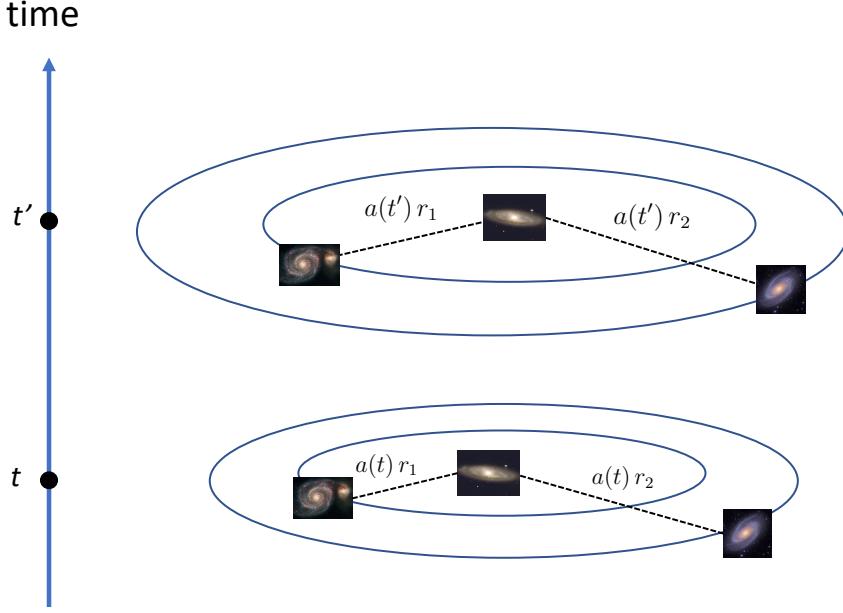


Figure 16: In a Friedmann-Robertson-Walker Universe, the spatial coordinates of galaxies, and hence their radial coordinate distances r_1, r_2, \dots from some reference galaxy, always remain the same. However, *physical* distances $a(t)r_1, a(t)r_2, \dots$ continually increase because of the increase of the scale factor $a(t)$.

in particular they always remain at the same coordinate distances r_1, r_2, \dots from the origin. However, the physical distance element in the radial direction is $\sqrt{ds^2} = a(t) dr$, so that for *e.g.* a galaxy at radial distance r_1 , the physical distance from the origin is

$$\ell_1 = \int_0^{r_1} a(t) dr = a(t) r_1. \quad (3.30)$$

Hence, at a time $t' > t$, the galaxy is at a larger distance than at time t . This is the way in which general relativity explains the expansion of the Universe!⁷

B. Tensors

In the previous section we mentioned $\eta_{\mu\nu}$ and more general $g_{\mu\nu}$ as examples of *tensors*. A tensor with two indices $\mu, \nu = 0, 1, 2, 3$ can be viewed as a 4×4 matrix. An example of

⁷ For clarity, it does not matter which galaxy is chosen to be at the origin of the coordinate system; if one chooses a different one then the picture remains qualitatively the same. The expansion of the Universe is not associated with any origin; rather, all galaxies are moving away from all others.

a tensor with only one index is dx^μ ; this can be thought of as a vector with 4 components. However, tensors can have any number of indices; for example one can have tensors $T_{\mu\nu\rho}$, which will have $4 \times 4 \times 4$ components.

In the index notation, it makes a difference whether the index is “up” or “down”. In particular, one defines $\eta^{\mu\nu}$ as the inverse of $\eta_{\mu\nu}$. In matrix notation this is $\boldsymbol{\eta}^{-1}$, and one has

$$\boldsymbol{\eta}^{-1} \cdot \boldsymbol{\eta} = \mathbf{1}, \quad (3.31)$$

with $\mathbf{1}$ the unit matrix. It so happens that $\boldsymbol{\eta}^{-1}$ has the same form as $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}^{-1} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.32)$$

In index notation,

$$\eta^{\mu\rho}\eta_{\rho\nu} = \delta_\nu^\mu \quad (3.33)$$

where the index ρ appears twice in the same term, once up and once down, and hence is being summed over following the Einstein convention. However, μ and ν are not dummy indices, and can be given definite values. The tensor δ_ν^μ is the Kronecker delta, whose components are those of the unit matrix: $\delta_\nu^\mu = 1$ if $\mu = \nu$, and $\delta_\nu^\mu = 0$ if $\mu \neq \nu$.

The metric and its inverse are used to lower or raise indices. For example, given a vector A^μ , one defines a new tensor A_μ as

$$A_\mu = \eta_{\mu\alpha} A^\alpha \quad (3.34)$$

and one has

$$A^\mu = \eta^{\mu\beta} A_\beta. \quad (3.35)$$

This is consistent, because

$$\eta^{\mu\beta} A_\beta = \eta^{\mu\beta} \eta_{\beta\alpha} A^\alpha = \delta_\alpha^\mu A^\alpha = A^\mu, \quad (3.36)$$

where in the first step we used (3.34) with the renaming $\mu \rightarrow \beta$, and in the second step we used (3.33) with $\rho \rightarrow \beta$ and $\nu \rightarrow \alpha$. One similarly raises or lowers indices on more general tensors; for example,

$$T^{\mu\nu}_\rho = \eta^{\mu\alpha} \eta^{\nu\beta} T_{\alpha\beta\rho}. \quad (3.37)$$

Note that the tensor $\eta_{\mu\nu}$ in the form (2.24) looks the same everywhere in spacetime. More generally, tensors can depend on the position in spacetime. When it is needed to indicate this explicitly, we will write *e.g.* $T_{\mu\nu\rho}(x)$, where x stands for (x^0, x^1, x^2, x^3) .

C. Metrics and coordinate transformations

The Minkowski metric is what one uses to compute distances between nearby spacetime points in special relativity, hence the name “metric”. In general relativity, other metrics $g_{\mu\nu}$ are possible, and in that case spacetime distances are computed as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \quad (3.38)$$

In the general theory, the metric usually depends on spacetime position, $g_{\mu\nu} = g_{\mu\nu}(x)$, though for brevity of notation we will not always make this explicit. By definition, metrics are symmetric tensors: $g_{\mu\nu} = g_{\nu\mu}$. Finally, we define an inverse metric $g^{\mu\nu}$ through the requirement

$$g^{\mu\rho} g_{\rho\nu} = \delta^\mu_\nu. \quad (3.39)$$

As before, the metric and its inverse are used to lower or raise indices; for example,

$$A_\mu = g_{\mu\alpha} A^\alpha, \quad T^{\mu\nu}{}_\rho = g^{\mu\alpha} g^{\nu\beta} T_{\alpha\beta\rho}, \quad (3.40)$$

and so on.

As we have seen, the Lorentz transformations preserve the spacetime distance ds^2 given by (2.22), or equivalently (2.27). General relativity allows for *general* coordinate transformations:

$$x'^\mu = x'^\mu(x), \quad (3.41)$$

where once again x is shorthand for (x^0, x^1, x^2, x^3) . An example of a coordinate transformation that is linear is a Lorentz transformation, but in general relativity we are also allowed to perform non-linear coordinate transformations. In the general theory, the spacetime distance (3.38) will be invariant under general coordinate transformations (3.41). This tells us how the metric $g_{\mu\nu}$ must transform under them. Indeed,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = g'_{\alpha\beta} dx'^\alpha dx'^\beta. \quad (3.42)$$

Note that

$$dx^\mu = \frac{\partial x^\mu}{\partial x'^\alpha} dx'^\alpha, \quad dx^\nu = \frac{\partial x^\nu}{\partial x'^\beta} dx'^\beta, \quad (3.43)$$

so that

$$\frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} g_{\mu\nu} dx'^\alpha dx'^\beta = g'_{\alpha\beta} dx'^\alpha dx'^\beta, \quad (3.44)$$

from which we read off

$$g'_{\alpha\beta} = \boxed{\frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} g_{\mu\nu}}. \quad (3.45)$$

As a special case, if the metric is the Minkowski metric, $g_{\mu\nu} = \eta_{\mu\nu}$, and if we let $x'^\mu(x)$ be a Lorentz transformation, then it is easily shown that Eq. (3.45) leads to

$$\eta'_{\alpha\beta} = \eta_{\alpha\beta}. \quad (3.46)$$

Hence Lorentz transformations do not change the Minkowski metric tensor. However, general coordinate transformations usually *do* change the metric tensor, so that $g'_{\alpha\beta} \neq g_{\alpha\beta}$! It is only the spacetime distance ds^2 which always remains the same; the metric tensor is said to transform *covariantly* as in (3.45).

As an example, consider a spatially flat Friedmann-Robertson-Walker Universe:

$$ds^2 = -c^2 dt^2 + a^2(t) [dr^2 + r^2 (d\theta^2 + \sin^2(\theta) d\phi^2)]. \quad (3.47)$$

Writing $(x^0, x^1, x^2, x^3) = (ct, r, \theta, \phi)$, the metric tensor is then

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) r^2 & 0 \\ 0 & 0 & 0 & a^2(t) r^2 \sin^2(\theta) \end{pmatrix}. \quad (3.48)$$

This can be simplified by transforming to rectilinear coordinates $(x'^0, x'^1, x'^2, x'^3) = (ct, x, y, z)$ where $x = r \sin(\theta) \cos(\phi)$, $y = r \sin(\theta) \sin(\phi)$, $z = r \cos(\theta)$. In that case

$$g'_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}, \quad (3.49)$$

so that

$$ds^2 = -c^2 dt^2 + a^2(t) [dx^2 + dy^2 + dz^2]. \quad (3.50)$$

An even simpler form of the metric is obtained by introducing a new time coordinate η such that $d\eta = dt/a(t)$, *i.e.*

$$\eta(t) = \int_0^t \frac{dt'}{a(t')}.$$
 (3.51)

Writing $a(\eta) = a(t(\eta))$, one readily finds

$$ds^2 = a^2(\eta) [-c^2 d\eta^2 + dx^2 + dy^2 + dz^2].$$
 (3.52)

How do other tensors transform under coordinate transformations? A *scalar* is a tensor without free indices, which is just a function $f(x)$ that assigns numbers to spacetime events. Hence it doesn't change – it is *invariant* – under coordinate transformations, since coordinates are mere labels for events. Next, how do tensors A_μ with a single index down transform? Since $A_\mu dx^\mu$ is a scalar,

$$A_\mu dx^\mu = A'_\alpha dx'^\alpha$$
 (3.53)

so that from (3.43) it follows that

$$A'_\alpha = \frac{\partial x^\mu}{\partial x'^\alpha} A_\mu.$$

 (3.54)

Note that for tensors A^μ and B_μ ,

$$\begin{aligned} A'^\alpha B'_\alpha &= A'^\alpha \frac{\partial x^\mu}{\partial x'^\alpha} B_\mu \\ &= A^\mu B_\mu, \end{aligned}$$
 (3.55)

where in the first line we have applied the transformation law (3.54) to B'_α , and in the second line we used that that $A^\mu B_\mu$ is a scalar and hence invariant. Since the above will be true for any B_μ , we read off that

$$A^\mu = \frac{\partial x^\mu}{\partial x'^\alpha} A'^\alpha.$$
 (3.56)

Multiplying both sides with $\partial x'^\beta / \partial x^\mu$ and summing over μ ,

$$\begin{aligned} \frac{\partial x'^\beta}{\partial x^\mu} A^\mu &= \frac{\partial x'^\beta}{\partial x^\mu} \frac{\partial x^\mu}{\partial x'^\alpha} A'^\alpha \\ &= \frac{\partial x'^\beta}{\partial x'^\alpha} A'^\alpha \\ &= \delta^\beta_\alpha A'^\alpha \\ &= A'^\beta, \end{aligned}$$
 (3.57)

where in the second line we have used the chain rule. This way we find the transformation law for vectors:

$$A'^\beta = \frac{\partial x'^\beta}{\partial x^\mu} A^\mu. \quad (3.58)$$

For a tensor with mixed indices, *e.g.* $T^{\mu\nu}_\rho$,

$$T'^{\alpha\beta}_\gamma = \frac{\partial x'^\alpha}{\partial x^\mu} \frac{\partial x'^\beta}{\partial x^\nu} \frac{\partial x^\rho}{\partial x'^\gamma} T^{\mu\nu}_\rho. \quad (3.59)$$

We end this subsection with an important property of spacetimes in general relativity, namely what is called *local flatness*. To illustrate the idea, recall the 2-dimensional spherical surface in subsection III A. Though obviously curved, at any given point a sufficiently small patch of it will resemble a piece of a flat plane, as exemplified by the behavior of sufficiently small circles on the surface. Similarly, any spacetime, however strongly curved, should resemble flat, Minkowski spacetime in a sufficiently small patch. More precisely, given a spacetime point P , there should be a coordinate transformation such that

$$\begin{aligned} g_{\mu\nu}|_P &= \eta_{\mu\nu}, \\ \frac{\partial}{\partial x^\rho} g_{\mu\nu}|_P &= 0. \end{aligned} \quad (3.60)$$

Note that generically, it will not be possible to satisfy these conditions everywhere in spacetime, using the same coordinate transformation. Should one want them to hold at some different point Q , typically some other coordinate transformation is needed, which will then no longer make the conditions hold at P . Moreover, even at a single point P , it will typically not be possible to find a coordinate transformation which also makes *second* derivatives with respect to coordinate vanish; *i.e.*, usually

$$\left. \frac{\partial^2}{\partial x^\rho \partial x^\sigma} g_{\mu\nu} \right|_P \neq 0. \quad (3.61)$$

A coordinate system that makes Eqs. (3.60) hold at some point P is called a *locally inertial frame* at that point.

D. Light cones and geodesics

The notation ds^2 for $g_{\mu\nu} dx^\mu dx^\nu$ is slightly deceptive: it suggests that the latter is non-negative, but that is not the case. First recall that in flat spacetime the distance between

nearby spacetime points is defined through Eq. (2.20). For a particle which is moving slower than the speed of light, one must have

$$dx^2 + dy^2 + dz^2 < c^2 dt^2. \quad (3.62)$$

Hence, for particles that have rest mass, $ds^2 < 0$. In the case of photons we already saw that $ds^2 = 0$. For hypothetical particles that move faster than light, one would have $ds^2 > 0$. The distinction between these three cases will be the same in general relativity, where ds^2 is defined using the metric $g_{\mu\nu}$ through (3.38).

The fact that ds^2 is invariant (*i.e.* it doesn't depend on the coordinate system used) leads to the notion of a *light cone*; see Fig. 17. Given a spacetime point P , light rays move on the surface of the light cone, defined by $ds^2 = 0$; these are called *lightlike* or *null* directions. Particles that have rest mass obey $ds^2 < 0$, so that they move “more in the time direction than in the spatial directions”; their itineraries through spacetime remain inside the light cone and are called *timelike*. Our hypothetical faster-than-light particles move in directions with $ds^2 > 0$, which are called *spacelike*.

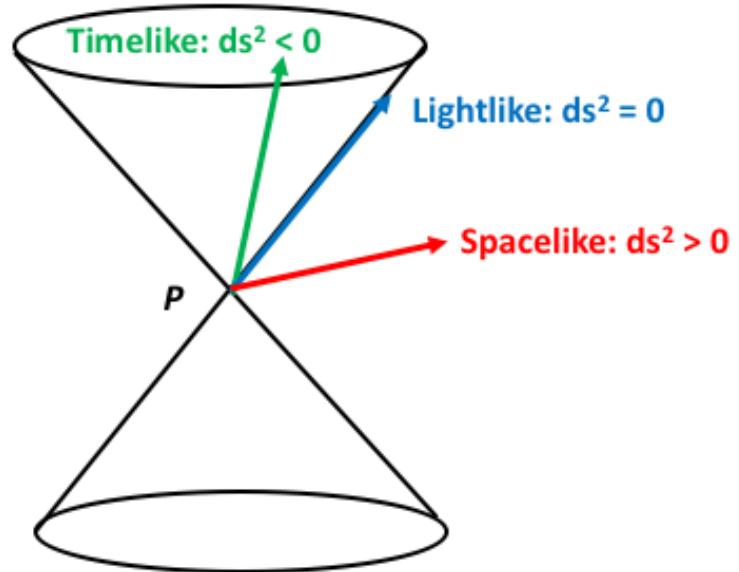


Figure 17: Light rays have $ds^2 = 0$ and define the local *light cone* at any spacetime point P ; they are said to move in *lightlike* directions. Particles that have rest mass obey $ds^2 < 0$ and move inside the light cone, in a *timelike* direction. Directions outside the light cone, with $ds^2 > 0$, are called *spacelike*.

Let us focus on particles that have rest mass, which move in a timelike direction: $ds^2 < 0$. Imagine such a particle moving from spacetime event A to spacetime event B along a path $x^\mu(\lambda)$ parameterized by λ ; see Fig. 18. Then the spacetime distance measured along the curve is

$$c\Delta\tau_{AB} = \int_A^B \sqrt{|ds^2|} = \int_A^B \sqrt{-g_{\mu\nu}dx^\mu dx^\nu} = \int_{\lambda_A}^{\lambda_B} \sqrt{-g_{\mu\nu}(x) \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda. \quad (3.63)$$

Since this is the physical distance along the curve in a timelike direction, $\Delta\tau_{AB}$ can be interpreted as the *proper time* elapsed for an observer moving through spacetime along the path $x^\mu(\lambda)$.

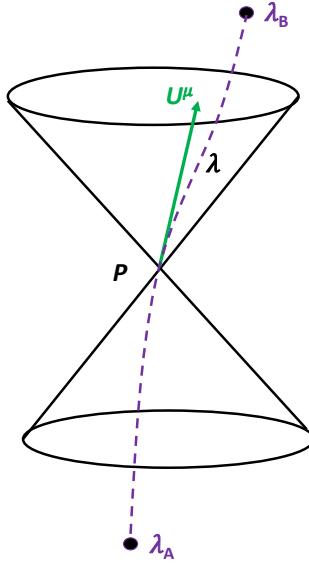


Figure 18: A timelike path $x^\mu(\lambda)$ (the dashed curve) will everywhere stay within the local light cone. The tangent vector $U^\mu = dx^\mu(\lambda)/d\lambda$ points in a timelike direction.

Fig. 18 also shows the tangent vector $U^\mu = dx^\mu(\lambda)/d\lambda$ to the curve $x^\mu(\lambda)$. Its norm, defined as $U_\mu U^\mu = g_{\alpha\beta} U^\alpha U^\beta$, is negative, again underscoring the timelike character of the path taken. Note that in the above we could have used any parameterization for our curve. In particular, instead of some arbitrary λ we could have picked the proper time τ as it increases along the curve. Moving along the curve over an infinitesimal amount of proper time $d\tau$, Eq. (3.63) implies

$$cd\tau = \sqrt{-g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} d\tau. \quad (3.64)$$

The *four-velocity*, defined as $V^\mu \equiv dx^\mu(\tau)/d\tau$, is again tangent to the curve, and from (3.64)

we find that it is normalized as

$$V_\mu V^\mu = g_{\alpha\beta} V^\alpha V^\beta = -c^2. \quad (3.65)$$

Its spatial components $V^i = dx^i/d\tau$, $i = 1, 2, 3$, can be thought of as the spatial velocity in the coordinate system x^μ with respect to the particle's proper time, while $V^0 = dx^0/d\tau = cdt/d\tau$ gives the flow of coordinate time with respect to proper time. One can also introduce a *four-momentum* $P^\mu = mV^\mu$, with m the rest mass of a particle; then $P_\mu P^\mu = -m^2c^2$.

In a flat spacetime, light cones look the same way everywhere because they are defined using the same metric $\eta_{\mu\nu}$. This is not the case in curved spacetime, where the metric $g_{\mu\nu}(x)$ may change from point to point. This can have dramatic consequences, for example in the vicinity of a black hole; see Fig. 19. Particles on the outside can escape to infinity. However, particles moving towards the black hole will eventually end up at a spacetime event where the light cone is tilted to such an extent that its future part is tangent to the horizon on the outside. At that point all timelike curves point inward. Going further inward, light cones tilt even more, in such a way that the singularity becomes the future end point of all timelike curves.

A *geodesic* is a path which minimizes (or maximizes) $\Delta\tau_{AB}$; this is the spacetime trajectory followed by a particle in free fall. This is somewhat reminiscent of classical mechanics, where the path followed is the one that extremalizes the *action*

$$S = \int_A^B L(x^\mu, \dot{x}^\mu) d\lambda \quad (3.66)$$

where the dot denotes the derivative with respect to λ , and $L(x^\mu, \dot{x}^\mu)$ is the Lagrangian. In our case the Lagrangian can be identified with

$$L(x^\mu, \dot{x}^\mu) = \sqrt{-g_{\mu\nu}(x)\dot{x}^\mu\dot{x}^\nu}. \quad (3.67)$$

Though we will not show it here, extremalizing (3.66) for the above Lagrangian leads to the *geodesic equation*

$$\frac{d^2x^\beta}{d\tau^2} + \Gamma_{\mu\nu}^\beta \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0$$

(3.68)

where

$$\Gamma_{\mu\nu}^\beta \equiv \frac{1}{2}g^{\beta\alpha}(\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}). \quad (3.69)$$

In the above, ∂_σ is shorthand for $\frac{\partial}{\partial x^\sigma}$; this is notation we will also use in subsequent sections.

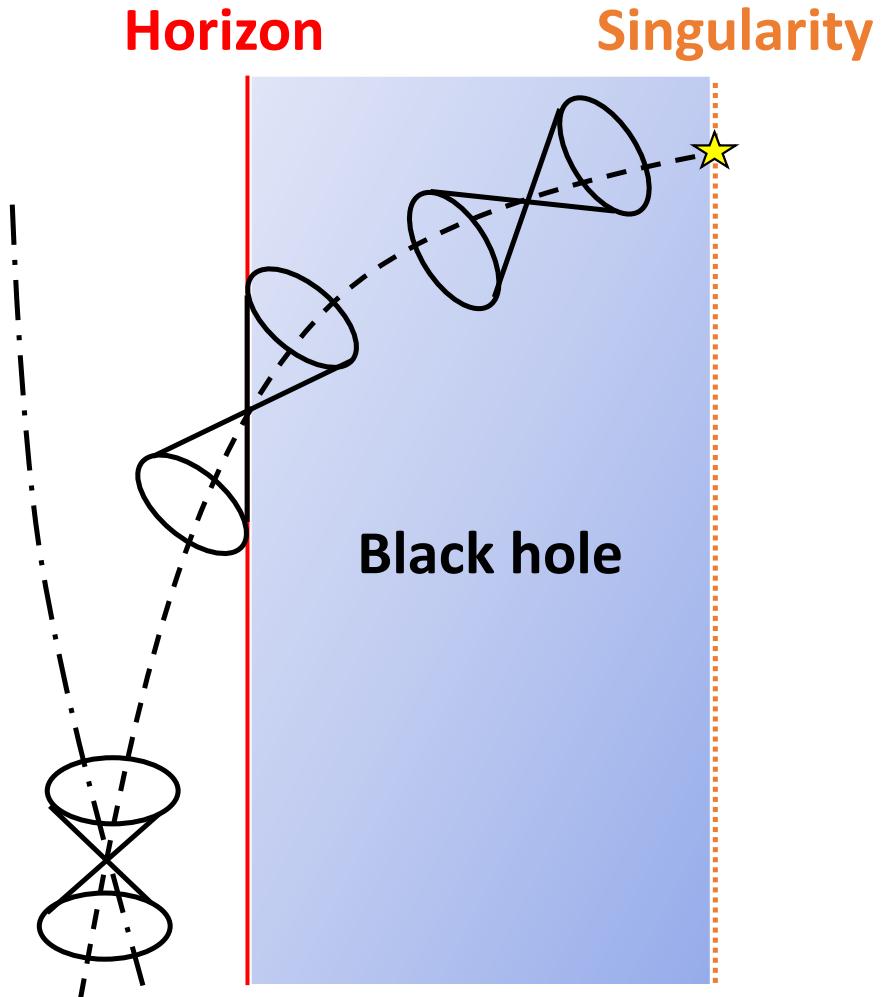


Figure 19: Particles near the horizon of a black hole. Outside the horizon, they can escape to infinity on *e.g.* the dashed-dotted curve. As one gets closer to the horizon, light cones start tilting towards it. The particle on the dashed trajectory gets too close to the horizon, and at some point reaches a spacetime event where the local light cone is tangent to the horizon as shown. After that the light cones keep tilting such that the singularity will be situated in the future of all timelike curves, so that eventually they will hit it.

The first thing to note is that in the case of the Minkowski metric, $g_{\mu\nu} = \eta_{\mu\nu}$, one has $\partial_\kappa g_{\gamma\delta} = 0$, hence $\Gamma_{\mu\nu}^\beta = 0$, so that the geodesic equation reduces to

$$\frac{d^2 x^\beta}{d\tau^2} = 0, \quad (3.70)$$

which is just the equation for a straight line. Indeed, in a spacetime that is not curved, this

is the shortest route between two spacetime points. However, in a curved spacetime this will generally not be the case, again as illustrated in Fig. 19 for the spacetime of a black hole. Since geodesics of massive particles, like all timelike curves, must stay inside the light cones, due to the tilting of the cones they can not remain straight lines.

IV. THE EINSTEIN FIELD EQUATIONS

Having gained some intuition on what it means for spacetime to be curved, and in particular how curvature affects the flow of matter and energy through spacetime, we now need to come up with a mathematical prescription for how matter/energy *causes* spacetime to curve. This is encapsulated in the *Einstein field equations*, which take the general form

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \quad (4.1)$$

The left hand side is the *Einstein tensor*, which depends on the metric $g_{\mu\nu}$ and its derivatives with respect to coordinates up to second order; it encapsulates the geometry of spacetime. The right hand side features the *energy-momentum tensor*, which tells us about the distribution of mass and energy, and the way it flows through spacetime. In this section we will provide some details, and motivate the above equations.⁸ First we will discuss the right hand side of the equations, and then come up with a proposal for what the left hand side should be in terms of the metric $g_{\mu\nu}$ and its derivatives.

A. The energy-momentum tensor

The components of the energy-momentum tensor have the following meaning:

- T^{00} is the energy density.
- $T^{0i}/c = T^{i0}/c$, $i = 1, 2, 3$, is the momentum density in the i direction.
- T^{ij} is the i th component of the force per unit area exerted across a surface with normal in the direction j .

As an example, consider a non-relativistic matter distribution whose density is $\rho(t, \mathbf{x})$. Then $T^{00} = \rho c^2$ and $T^{0i} = \rho v^i c$, with v^i the spatial velocity. The purely spatial part T^{ij} , $i, j = 1, 2, 3$ is the equivalent of the stress tensor in classical mechanics. Consider a rectangular volume element with sides dx, dy, dz . Then for example T^{xx} is the pressure on the yz

⁸ Note that the Einstein field equations can indeed only be *motivated* through intuition gained from thought experiments and observations. Like all fundamental laws of physics, the Einstein field equations constitute a postulate; they can not be *derived* from something else.

face normal to the x axis: $T^{xx} = F^x/(dydz)$. This is also equal to the momentum flux through this face, since $F^x/(dydz) = (dp^x/dt)/(dydz)$. Noting that $dp^x = \rho dx dy dz v^x$, we get $T^{xx} = \rho(dx/dt)v^x = \rho(v^x)^2$. Similarly $T^{yy} = \rho(v^y)^2$ and $T^{zz} = \rho(v^z)^2$, and the off-diagonal components are $T^{xy} = T^{yx} = \rho v^x v^y$, $T^{xz} = T^{zx} = \rho v^x v^z$, $T^{yz} = T^{zy} = \rho v^y v^z$.

Clearly the energy-momentum tensor is symmetric: $T^{\mu\nu} = T^{\nu\mu}$. Moreover, in flat space-time one has *energy-momentum conservation*:

$$\boxed{\partial_\mu T^{\mu\nu} = 0.} \quad (4.2)$$

Let us focus on this equation with $\nu = 0$:

$$\partial_0 T^{00} + \partial_i T^{i0} = 0. \quad (4.3)$$

Write $\epsilon = T^{00}$ for the energy density and $\pi^i \equiv T^{0i}/c$ for the momentum density. Since $\partial/\partial x^0 = \partial/(c\partial t)$ and $\partial_i \pi^i = \frac{\partial}{\partial x^i} \pi^i = \nabla \cdot \boldsymbol{\pi}$,

$$\frac{\partial \epsilon}{\partial t} = -\nabla \cdot \boldsymbol{\pi} c^2. \quad (4.4)$$

Consider *e.g.* π^x , the momentum density in the x direction through the yz face of a volume element. Then

$$\pi^x c^2 = \rho v^x c^2 = \frac{dm}{dxdydz} \frac{dx}{dt} c^2 = \frac{d(m c^2)}{dtdydz}, \quad (4.5)$$

which is the energy flux through the yz face, *i.e.* the energy per unit time per unit surface area. Integrating both sides of (4.4) over a volume \mathcal{V} bounded by a closed surface \mathcal{S} yields

$$\frac{dE}{dt} = - \int_{\mathcal{S}} dA \hat{\mathbf{n}} \cdot \boldsymbol{\pi} c^2, \quad (4.6)$$

with E the total energy in \mathcal{V} , and $\hat{\mathbf{n}}$ the unit normal to \mathcal{S} . Thus, Eq. (4.3) implies that the rate of change in energy inside a volume is determined by the rate at which energy escapes through the surface. This is of course nothing but conservation of energy.⁹

A simple example of an energy-momentum tensor is that of a *perfect fluid*. This is a fluid in which the force exerted across a surface is always along its normal, and the same for all orientations of the surface. Therefore the stress tensor is diagonal, with all the diagonal values equal to the pressure in the fluid, P :

$$T^{ij} = P \delta^{ij}. \quad (4.7)$$

⁹ Similarly, Eq. (4.2) with $\nu = j$ can be shown to imply a force balance equation.

Moreover, a perfect fluid is assumed to have no momentum transfer across surfaces, so that

$$T^{\mu\nu} = \text{diag}(\rho c^2, P, P, P). \quad (4.8)$$

An example of a perfect fluid is a non-relativistic perfect gas composed of particles with mass μ . The ideal gas law states

$$PV = Nk_B T, \quad (4.9)$$

with V the volume, N the number of particles in it, T the temperature, and k_B Boltzmann's constant. Writing the density as $\rho = N\mu/V$,

$$P = \frac{\rho}{\mu} k_B T. \quad (4.10)$$

According to equipartition of kinetic energy for particles with no internal degrees of freedom,

$$\frac{1}{2}\mu\langle v^2 \rangle = \frac{3}{2}k_B T, \quad (4.11)$$

where $\langle v^2 \rangle$ is the mean square velocity. Hence

$$P = \frac{\langle v^2 \rangle}{3}\rho. \quad (4.12)$$

Extrapolating to a relativistic gas, $\langle v^2 \rangle \rightarrow c^2$, so that

$$P = \frac{1}{3}\rho c^2, \quad (4.13)$$

and

$$T^{\mu\nu} = c^2 \text{diag}(\rho, \rho/3, \rho/3, \rho/3). \quad (4.14)$$

For a gas of photons, there is no notion of rest mass density ρ , but denoting the energy density as ϵ , the energy-momentum tensor becomes

$$T^{\mu\nu} = \text{diag}(\epsilon, \epsilon/3, \epsilon/3, \epsilon/3). \quad (4.15)$$

As we shall see, an energy momentum tensor that will be particularly useful to us is the one for a single non-relativistic particle with mass m at spatial position \mathbf{x}' . In that case $P = 0$, and

$$\rho(\mathbf{x}) = m \delta^3(\mathbf{x} - \mathbf{x}'). \quad (4.16)$$

Since $P = 0$, only T^{00} is non-zero, and equal to

$$T^{00} = mc^2 \delta^3(\mathbf{x} - \mathbf{x}'). \quad (4.17)$$

Note how the right hand side of Eq. (4.16) indeed has the dimensions of density: since

$$\int d^3\mathbf{x} m \delta^3(\mathbf{x} - \mathbf{x}') = m, \quad (4.18)$$

and $d^3\mathbf{x}$ has dimensions (length)³, the 3-dimensional Dirac delta function must have dimensions $1/(\text{length})^3 = 1/\text{volume}$.

B. Tensor calculus

As mentioned before, tensors can depend on the position in spacetime; for instance metrics do: $g_{\mu\nu}(x)$, where x is shorthand for (x^0, x^1, x^2, x^3) . As a more simple case, one can have a vector field $V^\mu(x)$ such that the vector is different in different spacetime points. Whenever we don't need to refer to components of a vector field, we will denote it with a boldface letter; thus, \mathbf{V} is the vector field with components V^μ .

The *inner product* between vectors \mathbf{U} and \mathbf{V} is defined as

$$\mathbf{U} \cdot \mathbf{V} \equiv g_{\mu\nu} U^\mu V^\nu. \quad (4.19)$$

Since $g_{\mu\nu} V^\nu = V_\mu$, one can write $\mathbf{U} \cdot \mathbf{V} = U^\mu V_\mu$, as before. Also as before, the norm-squared of a vector is $\|\mathbf{U}\|^2 = g_{\mu\nu} U^\mu U^\nu = U^\mu U_\mu$.

Next we define a *covariant derivative*, which relates vectors at different points in spacetime, but will also be a way in which the curvature of spacetime manifests itself. Again consider a curve $x^\mu(\lambda)$ with tangent vector $U^\mu = dx^\mu/d\lambda$, and a generic vector field \mathbf{V} . Then the covariant derivative ∇_μ at a point on the curve is defined through

$$\frac{dV^\alpha}{d\lambda} = \lim_{\Delta\lambda \rightarrow 0} \frac{V^\alpha(\lambda + \Delta\lambda) - V^\alpha(\lambda)}{\Delta\lambda} \equiv U^\mu \nabla_\mu V^\alpha, \quad (4.20)$$

and the same is taken to hold for any other curve going through the same point. Note that in flat space one would simply have $\nabla_\mu = \partial_\mu$. In curved spacetime, considering the property of local flatness, it should still be possible to find a coordinate system so that this holds in a particular point P , but as we shall see, there will generally not be a coordinate transformation that makes it hold at all points at the same time. However, suppose that $x'^\mu(x)$ is a coordinate transformation that does lead to $\nabla'_\mu = \partial/\partial x'^\mu$ at a point P . Then in terms

of the more general coordinates x^μ , and using $U'^\beta = U^\alpha \partial x'^\beta / \partial x^\alpha$ and $V'^\beta = V^\alpha \partial x'^\beta / \partial x^\alpha$,

$$\begin{aligned}
\frac{dV'^\mu}{d\lambda} &= U'^\nu \nabla'_\nu V'^\mu \\
&= U'^\nu \partial'_\nu V'^\mu \\
&= \left(U^\alpha \frac{\partial x'^\nu}{\partial x^\alpha} \right) \frac{\partial}{\partial x'^\nu} \left(V^\beta \frac{\partial x'^\mu}{\partial x^\beta} \right) \\
&= U^\alpha \frac{\partial}{\partial x^\alpha} \left(V^\beta \frac{\partial x'^\mu}{\partial x^\beta} \right) \\
&= U^\alpha \frac{\partial x'^\mu}{\partial x^\beta} \frac{\partial V^\beta}{\partial x^\alpha} + U^\alpha \frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\beta} V^\beta.
\end{aligned} \tag{4.21}$$

Multiplying both sides by $\partial x^\gamma / \partial x'^\mu$ and using $\partial x^\gamma / \partial x^\beta = \delta_\beta^\gamma$,

$$\frac{\partial x^\gamma}{\partial x'^\mu} \frac{dV'^\mu}{d\lambda} = U^\alpha \left(\frac{\partial V^\gamma}{\partial x^\alpha} + \frac{\partial x^\gamma}{\partial x'^\mu} \frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\beta} V^\beta \right), \tag{4.22}$$

or

$$\frac{dV^\gamma}{d\lambda} = U^\alpha \left(\frac{\partial V^\gamma}{\partial x^\alpha} + \frac{\partial x^\gamma}{\partial x'^\mu} \frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\beta} V^\beta \right). \tag{4.23}$$

From the definition (4.20), this leads to

$$U^\alpha \nabla_\alpha V^\gamma = U^\alpha \left(\frac{\partial V^\gamma}{\partial x^\alpha} + \frac{\partial x^\gamma}{\partial x'^\mu} \frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\beta} V^\beta \right). \tag{4.24}$$

Since we want this to hold for any vector U^α ,

$$\boxed{\nabla_\alpha V^\gamma = \partial_\alpha V^\gamma + \Gamma_{\alpha\beta}^\gamma V^\beta}, \tag{4.25}$$

where

$$\Gamma_{\alpha\beta}^\gamma \equiv \frac{\partial x^\gamma}{\partial x'^\mu} \frac{\partial^2 x'^\mu}{\partial x^\alpha \partial x^\beta} \tag{4.26}$$

are called the *connection coefficients*. As an important note for later, these are symmetric in the lower two indices: $\Gamma_{\alpha\beta}^\gamma = \Gamma_{\beta\alpha}^\gamma$.

Eq. (4.20) defines the covariant derivative acting on vectors, but it can also be made to act on more general tensors. The simplest tensor is one without any indices, *i.e.* a *function*. Given a function f , one has

$$\boxed{\nabla_\mu f = \frac{\partial}{\partial x^\mu} f}, \tag{4.27}$$

i.e. when acting on a function, the covariant derivative equals the coordinate derivative. This is because unlike the vector V'^β , there are no factors $\partial x'^\beta / \partial x^\alpha$ that complicate the transformation of a function inside a derivative.

What about the action of the covariant derivative on tensors with a single index down, *e.g.* C_α ? First, note that for an arbitrary vector field with components D^α , the combination $C_\alpha D^\alpha$ is a *function*, on which the covariant derivative must act as in Eq. (4.27). Thus,

$$\partial_\mu(C_\alpha D^\alpha) = \nabla_\mu(C_\alpha D^\alpha) = (\nabla_\mu C_\alpha) D^\alpha + C_\alpha(\partial_\mu D^\alpha + \Gamma_{\rho\mu}^\alpha D^\rho), \quad (4.28)$$

from which

$$(\partial_\mu C_\alpha) D^\alpha + C_\alpha \partial_\mu D^\alpha = (\nabla_\mu C_\alpha) D^\alpha + C_\alpha(\partial_\mu D^\alpha + \Gamma_{\rho\mu}^\alpha D^\rho). \quad (4.29)$$

This can be written as

$$(\nabla_\mu C_\alpha) D^\alpha = (\partial_\mu C_\alpha - \Gamma_{\alpha\mu}^\rho C_\rho) D^\alpha. \quad (4.30)$$

Since this holds for any vector field D^α , we find

$$\boxed{\nabla_\mu C_\alpha = \partial_\mu C_\alpha - \Gamma_{\alpha\mu}^\rho C_\rho.} \quad (4.31)$$

Finally, consistent with (4.23) and (4.31), for more general tensors, *e.g.* $T^{\alpha\beta}_\gamma$, one will have

$$\boxed{\nabla_\mu T^{\alpha\beta}_\gamma = \partial_\mu T^{\alpha\beta}_\gamma + \Gamma_{\rho\mu}^\alpha T^{\rho\beta}_\gamma + \Gamma_{\rho\mu}^\beta T^{\alpha\rho}_\gamma - \Gamma_{\gamma\mu}^\rho T^{\alpha\beta}_\rho.} \quad (4.32)$$

C. Geodesics revisited

Again consider some curve $x^\mu(\lambda)$ through spacetime parameterized by λ . Then we can define a notion of *parallel transport* of a vector along the curve. What we would mean by this in a flat space (*e.g.* the Euclidean plane) is illustrated in the left panel of Fig. (20): in each point on the curve a vector \vec{V} is defined, such that vectors at different points of the curve are all parallel to each other. Mathematically this means that everywhere along the curve one has

$$\frac{dV^\alpha}{d\lambda} = 0, \quad (4.33)$$

or

$$U^\beta \partial_\beta V^\alpha = 0, \quad (4.34)$$

with $U^\beta = dx^\beta/d\lambda$ the tangent vector to the curve at a given point.

In curved spaces, the notion of parallel transport becomes somewhat murky. This is illustrated in the right hand panel of Fig. 20. If we are on the 2-dimensional surface of a sphere, and transport a vector “as parallel to itself as possible” around a closed path, then

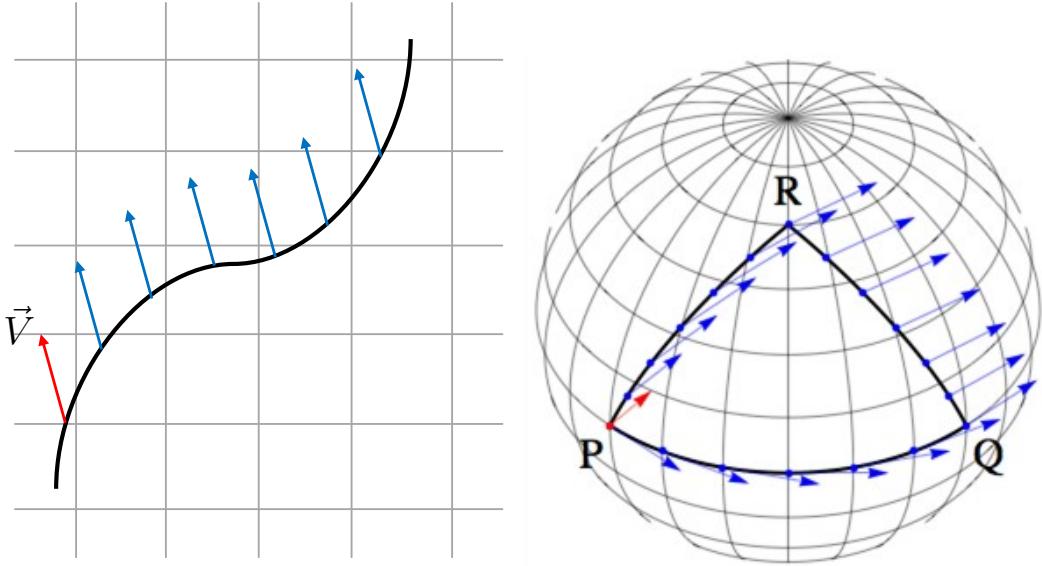


Figure 20: Left: in a flat space, such as the 2-dimensional Euclidean plane, there is a natural notion of parallel transporting a vector \vec{V} along a curve. Right: in a curved space, such as the surface of a sphere, parallel transport becomes non-trivial. For example, a vector that is transported “as parallel to itself as possible” along the closed path PRQP will end up being rotated with respect to the original vector.

generally the vector we obtain will be rotated with respect to the one we started out with. Curvature is also the reason for the appearance of the covariant derivative ∇_μ in Eq. (4.20), and in curved space or spacetime, Eq. (4.33) becomes

$$U^\beta \nabla_\beta V^\alpha = 0. \quad (4.35)$$

Though it may not be apparent from Eq. (4.20), there is actually some ambiguity in the definition of covariant derivative. To see this, consider two different vectors V^μ , W^ν being parallel transported along the same curve: $U^\alpha \nabla_\alpha V^\mu = 0 = U^\alpha \nabla_\alpha W^\nu$. Recall that we defined the inner product of two vectors as $\mathbf{V} \cdot \mathbf{W} = g_{\mu\nu} V^\mu W^\nu$. Now, when going along the curve,

$$\begin{aligned} \frac{d}{d\lambda}(\mathbf{V} \cdot \mathbf{W}) &= \frac{dg_{\mu\nu}}{d\lambda} V^\mu W^\nu + g_{\mu\nu} \frac{dV^\mu}{d\lambda} W^\nu + g_{\mu\nu} V^\mu \frac{dW^\nu}{d\lambda} \\ &= (U^\alpha \nabla_\alpha g_{\mu\nu}) V^\mu W^\nu + g_{\mu\nu} (U^\alpha \nabla_\alpha V^\mu) W^\nu + g_{\mu\nu} V^\mu (U^\alpha \nabla_\alpha W^\nu) \\ &= U^\alpha V^\mu W^\nu \nabla_\alpha g_{\mu\nu}. \end{aligned} \quad (4.36)$$

A desirable property of parallel transport will be that it preserves the angles between any

two vectors, as well as their norms. As can be seen from the above, this requires the covariant derivative to obey

$$\nabla_\alpha g_{\mu\nu} = 0. \quad (4.37)$$

This equation singles out a particular covariant derivative. To see this, first note that (4.37) implies, through the action of the covariant derivative on a tensor with two indices down,

$$\partial_\alpha g_{\mu\nu} - \Gamma_{\alpha\mu}^\rho g_{\rho\nu} - \Gamma_{\alpha\nu}^\rho g_{\mu\rho} = 0, \quad (4.38)$$

or

$$\partial_\alpha g_{\mu\nu} = \Gamma_{\alpha\mu}^\rho g_{\rho\nu} + \Gamma_{\alpha\nu}^\rho g_{\mu\rho}. \quad (4.39)$$

Simply renaming indices leads to two more equations:

$$\begin{aligned} \partial_\mu g_{\alpha\nu} &= \Gamma_{\mu\alpha}^\rho g_{\rho\nu} + \Gamma_{\mu\nu}^\rho g_{\alpha\rho}, \\ \partial_\nu g_{\alpha\mu} &= \Gamma_{\nu\alpha}^\rho g_{\rho\mu} + \Gamma_{\nu\mu}^\rho g_{\alpha\rho}. \end{aligned} \quad (4.40)$$

From this follows:

$$\begin{aligned} \partial_\mu g_{\alpha\nu} + \partial_\alpha g_{\mu\nu} - \partial_\nu g_{\alpha\mu} \\ = \Gamma_{\mu\alpha}^\rho g_{\rho\nu} + \Gamma_{\mu\nu}^\rho g_{\alpha\rho} + \Gamma_{\alpha\mu}^\rho g_{\rho\nu} + \Gamma_{\alpha\nu}^\rho g_{\mu\rho} - \Gamma_{\nu\alpha}^\rho g_{\rho\mu} - \Gamma_{\nu\mu}^\rho g_{\alpha\rho} \\ = 2\Gamma_{\alpha\mu}^\rho g_{\rho\nu}, \end{aligned} \quad (4.41)$$

where we used the symmetry of the metric, and that of $\Gamma_{\mu\nu}^\rho$, under interchange of the lower indices. This leads to

$$g^{\nu\sigma} \Gamma_{\alpha\mu}^\rho g_{\rho\nu} = \frac{1}{2} g^{\nu\sigma} (\partial_\mu g_{\alpha\nu} + \partial_\alpha g_{\mu\nu} - \partial_\nu g_{\alpha\mu}), \quad (4.42)$$

and since $g^{\nu\sigma} g_{\rho\nu} = \delta_\rho^\sigma$,

$$\Gamma_{\alpha\mu}^\sigma = \frac{1}{2} g^{\nu\sigma} (\partial_\mu g_{\alpha\nu} + \partial_\alpha g_{\mu\nu} - \partial_\nu g_{\alpha\mu}).$$

(4.43)

Thus, the condition (4.37) together with $\Gamma_{\mu\nu}^\rho = \Gamma_{\nu\mu}^\rho$ indeed singles out a particular covariant derivative, which is the one we will adopt in what follows. The resulting coefficients $\Gamma_{\mu\alpha}^\sigma$ in Eq. (4.43) are called the *Christoffel symbols*. Note that given an inertial frame at a point P , because of the second Eq. (3.60) one has $\Gamma_{\mu\alpha}^\sigma|_P = 0$. However, generally there will be no coordinate system such that the Christoffel symbols vanish simultaneously everywhere in spacetime.

Finally, consider a special kind of curve, namely one that parallel transports its own tangent vector, i.e., when considering the tangent vector in one point A and parallel transporting it to some other point B, one gets a vector that is also equal to the tangent vector at B. Mathematically,

$$U^\beta \nabla_\beta U^\alpha = 0. \quad (4.44)$$

In flat space, such a curve must be a straight line. Indeed, in that case the above equation reduces to

$$U^\beta \partial_\beta U^\alpha = \frac{dx^\beta}{d\lambda} \frac{\partial}{\partial x^\beta} \frac{dx^\alpha}{d\lambda} = \frac{d^2 x^\alpha}{d\lambda^2} = 0, \quad (4.45)$$

and the last equality does define a straight line. In curved space, where we use the covariant derivative, one instead finds

$$U^\beta \partial_\beta U^\alpha + \Gamma_{\mu\beta}^\alpha U^\mu U^\beta = 0, \quad (4.46)$$

or

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma_{\mu\beta}^\alpha \frac{dx^\mu}{d\lambda} \frac{dx^\beta}{d\lambda} = 0. \quad (4.47)$$

If we now compare Eqs. (4.47) and (4.43) with Eqs. (3.68) and (3.69), we arrive at an important realization. If we let ∇_β be the covariant derivative that preserves angles between vectors and their norms during parallel transport, then a curve that parallel transports its own tangent vector is also the shortest path between points in spacetime – i.e., it is a *geodesic!*

D. Curvature and the Riemann tensor

Having elucidated the physical meaning of our covariant derivative ∇_μ , we will now use it to further quantify the notion of spacetime curvature, which in the remainder of this section will allow us to motivate the Einstein field equations.

We start with pointing out that

$$\nabla_\mu \nabla_\nu A^\alpha - \nabla_\nu \nabla_\mu A^\alpha \neq 0. \quad (4.48)$$

We know how to calculate both terms. First of all, using Eq. (4.23),

$$\nabla_\nu A^\alpha = \partial_\nu A^\alpha + \Gamma_{\nu\rho}^\alpha A^\rho. \quad (4.49)$$

Note that this is a tensor with two free indices: an index ν which is down, and an index α which is up. Hence, when subsequently acting on this with ∇_μ , one finds

$$\begin{aligned}\nabla_\mu \nabla_\nu A^\alpha &= \nabla_\mu (\partial_\nu A^\alpha + \Gamma_{\nu\rho}^\alpha A^\rho) \\ &= \partial_\mu (\partial_\nu A^\alpha + \Gamma_{\nu\rho}^\alpha A^\rho) + \Gamma_{\mu\sigma}^\alpha (\partial_\nu A^\sigma + \Gamma_{\nu\rho}^\sigma A^\rho) - \Gamma_{\mu\nu}^\gamma (\partial_\gamma A^\alpha + \Gamma_{\gamma\rho}^\alpha A^\rho) \\ &= \partial_\mu \partial_\nu A^\alpha + (\partial_\mu \Gamma_{\nu\rho}^\alpha) A^\rho + \Gamma_{\nu\rho}^\alpha \partial_\mu A^\rho \\ &\quad + \Gamma_{\mu\sigma}^\alpha \partial_\nu A^\sigma + \Gamma_{\mu\sigma}^\alpha \Gamma_{\nu\rho}^\sigma A^\rho - \Gamma_{\mu\nu}^\gamma \partial_\gamma A^\alpha - \Gamma_{\mu\nu}^\gamma \Gamma_{\gamma\rho}^\alpha A^\rho.\end{aligned}\tag{4.50}$$

The second term in Eq. (4.48) is of course obtained from the above by interchanging μ and ν . An easy calculation shows that

$$\nabla_\mu \nabla_\nu A^\alpha - \nabla_\nu \nabla_\mu A^\alpha = R_{\rho\mu\nu}^\alpha A^\rho,\tag{4.51}$$

where

$$R_{\rho\mu\nu}^\alpha = \partial_\mu \Gamma_{\nu\rho}^\alpha - \partial_\nu \Gamma_{\mu\rho}^\alpha + \Gamma_{\mu\sigma}^\alpha \Gamma_{\nu\rho}^\sigma - \Gamma_{\nu\sigma}^\alpha \Gamma_{\mu\rho}^\sigma\tag{4.52}$$

is called the *Riemann tensor*.

The physical meaning of the Riemann tensor will become clear below, but for now we point out some basic properties, which we leave as exercises. Firstly, the Riemann tensor has the following symmetries:

$$\begin{aligned}R_{\mu\nu\alpha\beta} &= -R_{\mu\nu\beta\alpha}, \\ R_{\mu\nu\alpha\beta} &= -R_{\nu\mu\alpha\beta}, \\ R_{\mu\nu\alpha\beta} &= R_{\alpha\beta\mu\nu},\end{aligned}\tag{4.53}$$

where as always, indices are raised and lowered using the metric; for example $R_{\mu\nu\alpha\beta} = g_{\mu\rho} R_{\nu\alpha\beta}^\rho$.

Secondly, it obeys what is called the *Bianchi identity*:

$$\nabla_\mu R_{\alpha\beta\gamma\delta} + \nabla_\gamma R_{\alpha\beta\delta\mu} + \nabla_\delta R_{\alpha\beta\mu\gamma} = 0.\tag{4.54}$$

By summing over indices, the Riemann tensor defines the *Ricci tensor*

$$R_{\alpha\beta} \equiv R_{\alpha\mu\beta}^\mu\tag{4.55}$$

which in turn defines the *Ricci scalar*

$$R \equiv R_{\alpha}^{\alpha}.\tag{4.56}$$

Finally, in a local inertial frame (see Eq. (3.60)), in the definition of the Riemann tensor (3.60) only second derivatives of the metric survive, and one has

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}(\partial_\nu\partial_\rho g_{\mu\sigma} + \partial_\mu\partial_\sigma g_{\nu\rho} - \partial_\mu\partial_\rho g_{\nu\sigma} - \partial_\nu\partial_\sigma g_{\mu\rho}). \quad (4.57)$$

E. The Einstein equations

Understanding the physical meaning of the Riemann tensor will lead us to the Einstein equations. As we will see momentarily, the Riemann tensor encodes *tidal effects* experienced by particles moving through spacetime. Before elucidating this, it will be useful to think about such effects in the context of Newtonian gravity.

Imagine the trajectories of two free-falling particles in Newtonian gravity, under the influence of a gravitational potential $\phi(\mathbf{x})$. For a given particle, the acceleration is $\mathbf{a} = -\nabla\phi$. Let one particle be at spatial position \mathbf{x} and the other at $\mathbf{x} + \boldsymbol{\zeta}$, so that the separation vector between the two is $\boldsymbol{\zeta}$. For sufficiently small $\boldsymbol{\zeta}$, the acceleration of the second particle can be approximated as

$$\mathbf{a}(\mathbf{x} + \boldsymbol{\zeta}) = \mathbf{a}(\mathbf{x}) + \boldsymbol{\zeta} \cdot \nabla \mathbf{a}(\mathbf{x}) + \mathcal{O}(\boldsymbol{\zeta}^2). \quad (4.58)$$

Hence the components of the relative, or *tidal* acceleration are

$$\Delta a_i = -\boldsymbol{\zeta}^j \frac{\partial^2 \phi}{\partial x^i \partial x^j} = -\mathcal{E}_{ij} \boldsymbol{\zeta}^j, \quad (4.59)$$

where

$$\mathcal{E}_{ij} \equiv \frac{\partial^2 \phi}{\partial x^i \partial x^j} \quad (4.60)$$

is called the *tidal tensor field*.

As will be discussed in more detail below, given a matter distribution $\rho(\mathbf{x})$, the potential is determined by the Poisson equation,

$$\nabla^2 \phi = 4\pi G \rho. \quad (4.61)$$

Note that this can be written in terms of \mathcal{E}_{ij} as

$$4\pi G \rho = \nabla^2 \phi = \partial_i \partial^i \phi = \mathcal{E}_i^i. \quad (4.62)$$

Now we compare with what happens in curved spacetime. Again consider two freely falling particles, *i.e.* particles that follow geodesics. Thus, let $x^\mu(\tau)$ and $x^\mu(\tau) + \boldsymbol{\zeta}^\mu(\tau)$ be

two nearby geodesics, where we will assume the separation vector ζ^μ to be small. Then $x^\mu(\tau)$ satisfies

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma_{\alpha\beta}^\mu(x) \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0, \quad (4.63)$$

while for $x^\mu(\tau) + \zeta^\mu(\tau)$,

$$\frac{d^2(x^\mu + \zeta^\mu)}{d\tau^2} + \Gamma_{\alpha\beta}^\mu(x + \zeta) \frac{d(x^\alpha + \zeta^\alpha)}{d\tau} \frac{d(x^\beta + \zeta^\beta)}{d\tau} = 0. \quad (4.64)$$

In the second term, $\Gamma_{\alpha\beta}^\mu(x + \zeta)$ can be Taylor-expanded to

$$\Gamma_{\alpha\beta}^\mu(x + \zeta) = \Gamma_{\alpha\beta}^\mu(x) + \zeta^\sigma \partial_\sigma \Gamma_{\alpha\beta}^\mu(x) + \mathcal{O}(\zeta^2). \quad (4.65)$$

Taking the difference between these Eqs. (4.64) and (4.63) and restricting ourselves to leading order in ζ^μ and derivatives leads to

$$\frac{d^2\zeta^\mu}{d\tau^2} + 2\Gamma_{\alpha\beta}^\mu(x) \frac{dx^\alpha}{d\tau} \frac{d\zeta^\beta}{d\tau} + \zeta^\sigma \partial_\sigma \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0. \quad (4.66)$$

Now assume we are in a locally inertial frame, so that $\Gamma_{\alpha\beta}^\mu(x) = 0$. Also assume that the particles are moving much slower than the speed of light, so that the spatial components of their 4-velocities can be neglected. Then approximately

$$V^\alpha = \frac{dx^\alpha}{d\tau} \simeq (c, 0, 0, 0), \quad (4.67)$$

since $g_{\alpha\beta} = \eta_{\alpha\beta}$ in a local inertial frame, and one must have $V^\alpha V_\alpha = -c^2$. In that case Eq. (4.66) becomes

$$\frac{d^2\zeta^\mu}{d\tau^2} = -c^2 \zeta^\sigma \partial_\sigma \Gamma_{00}^\mu. \quad (4.68)$$

The right hand side of this equation can be related to the Riemann tensor defined in Eq. (4.52). Again recalling that $\Gamma_{\alpha\beta}^\mu(x) = 0$ in a locally inertial frame,

$$R^\mu_{0\sigma 0} = \partial_\sigma \Gamma_{00}^\mu - \partial_0 \Gamma_{\sigma 0}^\mu. \quad (4.69)$$

Though we will not show it here, given a timelike geodesic, it is possible to construct a frame that is locally inertial not only at a particular point, but *all along the geodesic*, and such that the time coordinate equals proper time.¹⁰ In that case $\Gamma_{\alpha\beta}^\mu = 0$ everywhere along the geodesic, and $\partial_0 \Gamma_{\sigma 0}^\mu = 0$ since $\Gamma_{\sigma 0}^\mu$ does not change when moving in the time direction.

¹⁰ Such a coordinate system is referred to as *Fermi normal coordinates*.

Writing $\Delta a^\mu = d^2\zeta^\mu/d\tau^2$ for the relative acceleration of the particles, Eq. (4.66) finally becomes

$$\Delta a_\mu = -c^2 R_{\mu 0\sigma 0} \zeta^\sigma. \quad (4.70)$$

Now compare this with the Newtonian expression Eq. (4.59). Then it will be clear that there is an analogy to be made between certain components of the Riemann tensor and the tidal tensor:

$$c^2 R_{\mu 0\sigma 0} \longleftrightarrow \mathcal{E}_{ij}, \quad (4.71)$$

and

$$c^2 R_{0\alpha 0}^\alpha = c^2 R_{00} \longleftrightarrow \mathcal{E}_i^i, \quad (4.72)$$

where R_{00} is the time-time component of the Ricci tensor (4.55). Comparing with the Poisson equation (4.62), this suggest the following expression to relate spacetime curvature (as encapsulated by R_{00}) with mass/energy density:

$$R_{00} = \frac{4\pi G}{c^2} \rho. \quad (4.73)$$

Recalling our discussion of the energy-momentum tensor and noting that $T_{00} = \rho c^2$,

$$R_{00} = \frac{4\pi G}{c^4} T_{00}. \quad (4.74)$$

This in turn led Einstein to *initially* postulate the following field equations for general relativity:

$$R_{\mu\nu} = \frac{4\pi G}{c^4} T_{\mu\nu}. \quad (4.75)$$

Unfortunately, these equations are flawed. Recall that in flat spacetime, conservation of energy-momentum was described by $\partial^\mu T_{\mu\nu} = 0$; in curved spacetime it is natural to generalize this to $\nabla^\mu T_{\mu\nu} = 0$. However, for Eq. (4.75) to make sense, one would then need $\nabla^\mu R_{\mu\nu} = 0$. As can be shown using the Bianchi identity (4.54), this in turn would imply $\nabla^\mu R = 0$, and through Eq. (4.75), $\nabla^\mu T = 0$, with $T \equiv T^\mu_\mu$. Hence T^μ_μ would have to be constant throughout the Universe, which is too strong a constraint on the properties of matter. The problem is solved if Eq. (4.75) is modified such that applying ∇^μ to the left hand side yields zero *by construction*, so that the equation will not imply further constraints on $T_{\mu\nu}$. Again from the Bianchi identity, it is possible to show that

$$\nabla^\mu \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) = 0, \quad (4.76)$$

indeed by construction. This led Einstein to the final version of what are now called the *Einstein field equations*:¹¹

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \quad (4.77)$$

The left hand side is called the *Einstein tensor*, denoted

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \quad (4.78)$$

so that

$$G_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \quad (4.79)$$

Having obtained the Einstein field equations, in the next few sections we will explore its consequences. First we will show that, apart from the equivalence principle, general relativity solves another problem with Newtonian gravity, namely that in Newton's theory, gravity acts instantaneously and without time delay over arbitrarily large distances. Related to this will be the notion of *gravitational waves*: disturbances in the spacetime curvature that propagate at the speed of light.

¹¹ Note the extra factor of 2 in the right hand side compared to Eq. (4.75). This turns out to be needed to retrieve Newtonian gravity in the limit of small spacetime curvature, as we shall see later on.

V. THE DYNAMICS OF SPACETIME

A. Gravity according to Newton (again)

Before exploring the dynamics of spacetime in general relativity, let us revisit Newton's theory of gravity one more time. In that theory, two particles with masses m and M located respectively at spatial positions \mathbf{x} and \mathbf{x}_0 act on each other with a gravitational force. In terms of the distance $r = |\mathbf{x} - \mathbf{x}_0|$, the force on m due to M is

$$\mathbf{F}_G = -\frac{GmM}{r^2} \hat{\mathbf{e}}, \quad (5.1)$$

where G is Newton's constant, and $\hat{\mathbf{e}} = (\mathbf{x} - \mathbf{x}_0)/|\mathbf{x} - \mathbf{x}_0|$ is a unit vector pointing from M to m .

Newtonian gravity can also be formulated in terms of a potential ϕ such that the force on a particle with mass m at position \mathbf{x} is

$$\mathbf{F}_G(\mathbf{x}) = -m\nabla\phi(\mathbf{x}), \quad (5.2)$$

where ∇ is the gradient with respect to \mathbf{x} . In the simple case above, the potential generated by M is

$$\phi(\mathbf{x}) = -\frac{GM}{|\mathbf{x} - \mathbf{x}_0|}, \quad (5.3)$$

and it is an easy exercise to show that using (5.2) this indeed leads to (5.1).

A more general case is a point particle m that is subject to the gravitational pull of some mass distribution with density $\rho(\mathbf{x}')$. The potential ϕ at a position \mathbf{x} then follows from the Poisson equation:

$$\nabla^2\phi(\mathbf{x}) = 4\pi G\rho(\mathbf{x}), \quad (5.4)$$

where ∇^2 is the Laplacian operator:

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (5.5)$$

Given a mass distribution ρ , the general solution to Eq. (5.4) can be obtained using the *Green's function* of the Laplacian. This is a function $G(x)$ which by definition satisfies

$$\nabla^2 G(\mathbf{x}) = \delta^3(\mathbf{x}), \quad (5.6)$$

where $\delta^3(\mathbf{x}) = \delta(x)\delta(y)\delta(z)$. Though we will not show it here, the Green's function of the Laplacian is

$$G(\mathbf{x}) = -\frac{1}{4\pi} \frac{1}{|\mathbf{x}|}. \quad (5.7)$$

For a generic mass distribution $\rho(\mathbf{x}')$ confined to a volume \mathcal{V} , the solution to the Poisson equation is given by the volume integral

$$\phi(\mathbf{x}) = 4\pi G \int_{\mathcal{V}} d^3\mathbf{x}' G(\mathbf{x} - \mathbf{x}') \rho(\mathbf{x}'). \quad (5.8)$$

Indeed, acting with the Laplacian w.r.t. \mathbf{x} on both sides, in the right hand side it will pass through the integral over \mathbf{x}' , and using Eq. (5.6) one recovers Eq. (5.4).

From Eqs. (5.8) and (5.7), we obtain

$$\phi(\mathbf{x}) = - \int_{\mathcal{V}} d^3\mathbf{x}' \frac{G\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (5.9)$$

This makes intuitive sense: an infinitesimal mass element $dM(\mathbf{x}') = \rho(\mathbf{x}') d^3\mathbf{x}'$ at a point \mathbf{x}' contributes to the potential at \mathbf{x} by an amount $-G\rho(\mathbf{x}') d^3\mathbf{x}' / |\mathbf{x} - \mathbf{x}'|$. The total potential at \mathbf{x} is obtained by summing (or integrating) over all the contributions inside the volume \mathcal{V} . Finally, in the special case that an entire mass M is concentrated at a single point \mathbf{x}_0 , one can write $\rho(\mathbf{x}') = M\delta^3(\mathbf{x}' - \mathbf{x}_0)$, and Eq. (5.9) reduces to Eq. (5.3):

$$\phi(\mathbf{x}) = -\frac{GM}{|\mathbf{x} - \mathbf{x}_0|}. \quad (5.10)$$

All of the equations above reveal an important fact about Newtonian gravity. Looking at Eq. (5.1) for the force between point particles, if we keep the position \mathbf{x} of m fixed but slightly change \mathbf{x}_0 , then immediately the force on m changes, no matter how far away m and M are from each other. Similarly, consider some bounded mass distribution $\rho(\mathbf{x}')$ which changes over time; think *e.g.* of density perturbations in a swirling gas cloud. Then the potential $\phi(\mathbf{x})$ at point \mathbf{x} will respond immediately to these time changes, irrespective how far away \mathbf{x} is from the region where the mass distribution is concentrated. *Therefore, Newtonian gravity acts instantaneously over arbitrarily large distances, without any time delay.*

Despite the huge successes of Newton's theory, this instantaneous action at a distance was considered unsatisfactory already by some of his contemporaries in the late 17th century. These tried to come up with some dynamical mechanism through which the gravitational

force would be communicated at finite speed, but without success. The issue became especially pressing after the development of special relativity (1905), which imposes a strict speed limit, the speed of light, on how fast communication of any kind can be effected.

Maxwell's theory of electromagnetism does not have this instantaneous action at a distance. Consider a localized charge/current distribution (the “source”) causing electric and magnetic fields \mathbf{E} and \mathbf{B} . From the Maxwell equations it can be shown that at a time t , the values of \mathbf{E} and \mathbf{B} at a distance D from the source depend on what the source was doing at a time $t - D/c$. The time lag, D/c , is the time needed for a signal to cross the distance D if it traveled at the speed of light: electromagnetism obeys Einstein's speed limit. \mathbf{E} and \mathbf{B} can be shown to obey a wave equation: the changes in a charge/current distribution are communicated to the rest of space by electromagnetic waves. Thus, unlike the Newtonian gravitational potential, the electromagnetic field does not just “track” its sources; it has dynamics of its own.

After special relativity was developed it was soon speculated that, just like the electromagnetic field, the gravitational field might also be dynamical. Changes in the gravitational field should propagate in a wave-like fashion, no faster than the speed of light, thus eliminating instantaneous action at a distance. A concrete mathematical implementation of these notions would have to wait for another decade, but the *general* theory of relativity of 1915 indeed incorporated all these ideas. In particular, general relativity predicts the existence of gravitational waves, as we will now show.

B. “Linearized” general relativity

Recall that the Einstein equations are

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (5.11)$$

where the Einstein tensor $G_{\mu\nu}$ is constructed from the metric $g_{\mu\nu}$ and its derivatives with respect to coordinates. An easy way to gain insight into the nature of gravitational waves will be to study them in the regime where gravitational fields are weak. In that case the spacetime metric $g_{\mu\nu}$ can be written as the flat Minkowski metric of special relativity, $\eta_{\mu\nu}$, plus a small correction due to the weak gravitational fields, $h_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1. \quad (5.12)$$

General relativity allows for any kind of coordinate transformations $x'^\mu(x)$. However, many of these spoil the above form of the metric: general coordinate transformations can make $|h_{\mu\nu}|$ arbitrarily large. Since the condition $|h_{\mu\nu}| \ll 1$ will be convenient in what follows, as a self-imposed rule we will limit ourselves to those coordinate transformations that preserve it. In particular, we only consider “small” changes in coordinates

$$x'^\mu = x^\mu + \xi^\mu, \quad (5.13)$$

where ξ^μ is a vector whose components are allowed to vary from one point in spacetime to another, but we demand that its effect be small everywhere. As we have seen, the metric transforms as

$$g'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}. \quad (5.14)$$

Since $x^\alpha = x'^\alpha - \xi^\alpha$,

$$\frac{\partial x^\rho}{\partial x'^\mu} = \delta_\mu^\rho - \partial_\mu \xi^\rho, \quad \frac{\partial x^\sigma}{\partial x'^\nu} = \delta_\nu^\sigma - \partial_\nu \xi^\sigma, \quad (5.15)$$

where we recall that ∂_μ is shorthand for partial derivatives with respect to coordinates.¹² To linear order in the small quantities $|h_{\alpha\beta}|$ and $|\partial_\alpha \xi^\beta|$, Eq. (5.14) becomes

$$\begin{aligned} g'_{\mu\nu} &= (\delta_\mu^\rho - \partial_\mu \xi^\rho)(\delta_\nu^\sigma - \partial_\nu \xi^\sigma)(\eta_{\rho\sigma} + h_{\rho\sigma}) \\ &\simeq \delta_\mu^\rho \delta_\nu^\sigma \eta_{\rho\sigma} - (\partial_\mu \xi^\rho) \delta_\nu^\sigma \eta_{\rho\sigma} - \delta_\mu^\rho (\partial_\nu \xi^\sigma) \eta_{\rho\sigma} + \delta_\mu^\rho \delta_\nu^\sigma h_{\rho\sigma} \\ &= \eta_{\mu\nu} - \eta_{\rho\nu} \partial_\mu \xi^\rho - \eta_{\mu\rho} \partial_\nu \xi^\sigma + h_{\mu\nu} \\ &= \eta_{\mu\nu} + h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu. \end{aligned} \quad (5.17)$$

Here we used *e.g.* that $\delta_\nu^\sigma \eta_{\rho\sigma} = \eta_{\rho\nu}$, and since the components of $\eta_{\rho\nu}$ are constant, $\eta_{\rho\nu} \partial_\mu \xi^\rho = \partial_\mu (\eta_{\rho\nu} \xi^\rho) = \partial_\mu (\eta_{\nu\rho} \xi^\rho) = \partial_\mu \xi_\nu$. Identifying the last three terms in (5.17) with $h'_{\mu\nu}$, we find

$$h'_{\mu\nu} = h_{\mu\nu} - (\partial_\mu \xi_\nu + \partial_\nu \xi_\mu). \quad (5.18)$$

Since the above transformations originate from coordinate transformations, they don’t change the physics of whatever situation is described by $h_{\mu\nu}$. Because of the close analogy with electromagnetism, they are called *gauge transformations*. Furthermore, to the

¹² Note that when acting on quantities like ξ^ρ , to linear order the primed partial derivatives are the same as unprimed ones:

$$\frac{\partial \xi^\rho}{\partial x'^\mu} = \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial \xi^\rho}{\partial x^\alpha} = \left(\delta_\mu^\alpha - \frac{\partial \xi^\alpha}{\partial x'^\mu} \right) \frac{\partial \xi^\rho}{\partial x^\alpha} \simeq \frac{\partial \xi^\rho}{\partial x^\mu}. \quad (5.16)$$

linear order considered here, $h_{\mu\nu}$ can be viewed as a tensor field living on a flat spacetime. Indeed, when raising or lowering indices, it makes no difference whether this is done with (the inverse of) $g_{\mu\nu}$ or with $\eta_{\mu\nu}$. To see this, first note that to leading order in $h_{\mu\nu}$, the inverse of the metric is

$$g^{\mu\nu} = \eta^{\mu\nu} - \eta^{\mu\alpha}\eta^{\nu\beta}h_{\alpha\beta}, \quad (5.19)$$

since this is the tensor such that

$$g^{\mu\rho}g_{\rho\nu} = \delta^\mu_\nu + \mathcal{O}(h^2). \quad (5.20)$$

From this it follows that, for example,

$$h^\mu_\nu = g^{\mu\alpha}h_{\alpha\nu} = \eta^{\mu\alpha}h_{\alpha\nu} + \mathcal{O}(|h|^2). \quad (5.21)$$

Next, one can “linearize” the Einstein equations, meaning that the form (5.12) of the metric is substituted into the full Einstein equations (5.11), and only terms are kept that are linear in $h_{\mu\nu}$ and its derivatives. The calculation is a lengthy one, but let us highlight the main steps. First, the linearized Riemann tensor is

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}(\partial_\nu\partial_\rho h_{\mu\sigma} + \partial_\mu\partial_\sigma h_{\nu\rho} - \partial_\mu\partial_\rho h_{\nu\sigma} - \partial_\nu\partial_\sigma h_{\mu\rho}). \quad (5.22)$$

The linearized Ricci tensor then becomes

$$\begin{aligned} R_{\alpha\beta} &= R^\mu_{\alpha\mu\beta} \\ &= \frac{1}{2}(\partial_\alpha\partial^\mu h_{\mu\beta} + \partial^\mu\partial_\beta h_{\alpha\mu} - \partial^\mu\partial_\mu h_{\alpha\beta} - \partial_\alpha\partial_\beta h^\mu_\mu), \end{aligned} \quad (5.23)$$

with $\partial^\mu = \eta^{\mu\nu}\partial_\nu$. The Ricci scalar is

$$\begin{aligned} R &= R^\alpha_\alpha \\ &= \partial^\alpha\partial^\mu h_{\mu\alpha} - \partial^\mu\partial_\mu h^\alpha_\alpha. \end{aligned} \quad (5.24)$$

The linearized Einstein tensor is

$$\begin{aligned} G_{\alpha\beta} &= R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta} \\ &= \frac{1}{2}(\partial_\alpha\partial^\mu h_{\mu\beta} + \partial^\mu\partial_\beta h_{\alpha\mu} - \partial^\mu\partial_\mu h_{\alpha\beta} - \partial_\alpha\partial_\beta h^\mu_\mu) \\ &\quad - \frac{1}{2}\eta_{\alpha\beta}(\partial^\nu\partial^\mu h_{\mu\nu} - \partial^\mu\partial_\mu h^\nu_\nu) \end{aligned} \quad (5.25)$$

To simplify notation, it will be convenient to introduce the new tensor

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \quad (5.26)$$

where

$$h = \eta^{\mu\nu}h_{\mu\nu}. \quad (5.27)$$

Note that conversely,

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}. \quad (5.28)$$

Substituting this into (5.25), and putting the linearized Einstein tensor into the left hand side of (5.11), we obtain the linearized Einstein equations:

$$\square\bar{h}_{\alpha\beta} + \eta_{\alpha\beta}\partial^\mu\partial^\nu\bar{h}_{\mu\nu} - \partial^\mu\partial_\beta\bar{h}_{\alpha\mu} - \partial^\mu\partial_\alpha\bar{h}_{\beta\mu} = -\frac{16\pi G}{c^4}T_{\alpha\beta}, \quad (5.29)$$

where

$$\begin{aligned} \square &\equiv \partial_\mu\partial^\mu \\ &= \eta^{\mu\nu}\partial_\mu\partial_\nu \\ &= -\frac{\partial^2}{(\partial x^0)^2} + \frac{\partial^2}{(\partial x^1)^2} + \frac{\partial^2}{(\partial x^2)^2} + \frac{\partial^2}{(\partial x^3)^2} \\ &= -\frac{\partial^2}{c^2\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \end{aligned} \quad (5.30)$$

is called the d'Alembertian.

Eq. (5.29) can be further simplified by making use of the gauge transformations. Using (5.26) and (5.18) it is easy to see that these transformations act on $\bar{h}_{\mu\nu}$ as

$$\boxed{\bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - (\partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial_\rho\xi^\rho)}. \quad (5.31)$$

Now, it can be shown that given a $\bar{h}_{\mu\nu}$, there always exists a gauge transformation to a $\bar{h}'_{\mu\nu}$ such that (dropping the prime)

$$\boxed{\partial^\mu\bar{h}_{\mu\nu} = 0}. \quad (5.32)$$

This condition is called the *Lorentz gauge*. With it, we see that the last three terms in the left hand side of Eq. (5.29) vanish (taking into account that partial derivatives commute, and renaming indices where needed), and we simply get

$$\boxed{\square\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}}. \quad (5.33)$$

C. No instantaneous action at a distance!

Analogously to the way the Poisson equation (5.4) of Newtonian gravity has the general solution (5.8) in terms of a mass distribution $\rho(\mathbf{x})$ and the appropriate Green's function, the linearized Einstein equations (5.33) have a general solution in terms of the energy-momentum distribution $T_{\mu\nu}$:

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = -\frac{16\pi G}{c^4} \int dt' d^3\mathbf{x}' \mathcal{G}(t - t', \mathbf{x} - \mathbf{x}') T_{\mu\nu}(t', \mathbf{x}'), \quad (5.34)$$

where the Green's function $\mathcal{G}(t, \mathbf{x})$ of the d'Alembertian satisfies

$$\square \mathcal{G}(t, \mathbf{x}) = \delta^4(t, \mathbf{x}), \quad (5.35)$$

with $\delta^4(t, \mathbf{x}) = \delta(t) \delta^3(\mathbf{x})$. It can be shown that

$$\mathcal{G}(t, \mathbf{x}) = -\frac{1}{4\pi|\mathbf{x}|} \delta\left(t - \frac{|\mathbf{x}|}{c}\right). \quad (5.36)$$

Plugging (5.36) into (5.34), we find that the most general solution to the linearized Einstein equations at an arbitrary spacetime point (t, \mathbf{x}) takes the form

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = 4\frac{G}{c^4} \int_{\mathcal{V}} d^3\mathbf{x}' \frac{T_{\mu\nu}\left(t - \frac{|\mathbf{x}-\mathbf{x}'|}{c}, \mathbf{x}'\right)}{|\mathbf{x} - \mathbf{x}'|}, \quad (5.37)$$

where we integrated over time and made use of the delta function in Eq. (5.36). The remaining integral is a spatial one, over some spatial volume \mathcal{V} in which the source $T_{\mu\nu}$ is non-zero; see Fig. 21.

We can now make an important observation: *Unlike the Newtonian potential, the value of $\bar{h}_{\mu\nu}$ at a point \mathbf{x} arbitrarily far from the source does not have instantaneous knowledge of what happens inside \mathcal{V} .* Rather, there are time lags $|\mathbf{x} - \mathbf{x}'|/c$, these being the times needed for a signal traveling at the speed of light to get from points \mathbf{x}' inside the source to the point \mathbf{x} . Just like electromagnetism, gravity does not have instantaneous action at a distance after all.

Next we note that outside the source, where $T_{\mu\nu} = 0$, Eq. (5.33) reduces to

$$\square \bar{h}_{\mu\nu} = 0, \quad (5.38)$$

or

$$\left(-\frac{\partial}{c^2 \partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) \bar{h}_{\mu\nu} = 0. \quad (5.39)$$

This is just a wave equation, for waves traveling at the speed of light. We conclude that, again like electromagnetism, the gravitational field has dynamics of its own.

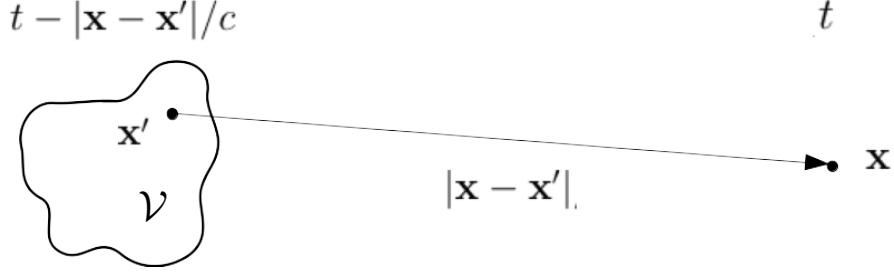


Figure 21: The gravitational field at some spatial position \mathbf{x} at a time t due to a matter distribution in some volume \mathcal{V} depends on what the distribution looked like at an *earlier* time. The time delays are equal to the times needed for signals traveling at the speed of light to get from points inside the source to the point \mathbf{x} .

D. Intermezzo: The Newtonian limit

We now know that general relativity does not allow for action at a distance, and changes in the gravitational field (embodied by $\bar{h}_{\mu\nu}$) propagate at a finite speed and according to a wave equation, very differently from what happens in the theory of Newton. Nevertheless, there should be some regime in which Newtonian gravity holds, because otherwise how could it have been so successful in describing observable phenomena for several hundred years?

Newtonian gravity should be approximately valid when (a) gravitational fields are weak, and (b) speeds are small compared to the speed of light. Note that ϕ/c^2 is dimensionless, so we will take (a) to mean $\phi/c^2 \ll 1$, and (b) $v \ll c$. First we look at the energy-momentum tensor $T^{\mu\nu}$. The energy density is $T^{00} = \rho c^2$. T^{0i} is the momentum density times the speed of light, which is due to the motion of particles in the matter distribution, and goes like $(\rho v) c$. The spatial components are pressures and shears in the matter distribution which go like ρv^2 . Hence $|T^{00}| \gg |T^{0i}| \gg |T^{ij}|$. Because of the linearized Einstein equations (5.33), this will translate to $|\bar{h}^{00}| \gg |\bar{h}^{0i}| \gg |\bar{h}^{ij}|$, and we can focus on the time-time component:

$$\square \bar{h}^{00} = -\frac{16\pi G}{c^4} \rho c^2. \quad (5.40)$$

For fields that change because matter in the source moves with speeds $v \ll c$, one has $\partial/\partial x^0 = \partial/(c\partial t) = \mathcal{O}[(v/c)\partial/\partial x]$, so that

$$\square = \nabla^2 + \mathcal{O}\left(\frac{v^2}{c^2}\nabla^2\right). \quad (5.41)$$

Hence the equation of interest is

$$\nabla^2 \bar{h}^{00} = -\frac{16\pi G}{c^2} \rho. \quad (5.42)$$

Comparing this with the Poisson equation (5.4),

$$\nabla^2 \phi = 4\pi G \rho, \quad (5.43)$$

we are led to identifying

$$\bar{h}^{00} = -\frac{4\phi}{c^2}. \quad (5.44)$$

Let us check what happens to $h_{\mu\nu}$ itself in this approximation. Eq. (5.26) can be written as

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h} \quad (5.45)$$

with $\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu}$. Since only \bar{h}_{00} is significant, one has $\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu} \simeq \eta^{00}\bar{h}_{00} = -\bar{h}_{00}$. From (5.45) and (5.44), and using $\bar{h}_{00} = \eta_{0\alpha}\eta_{0\beta}\bar{h}^{\alpha\beta} = \eta_{00}\eta_{00}\bar{h}^{00} = \bar{h}^{00}$,

$$\begin{aligned} h_{00} &= \bar{h}_{00} - \frac{1}{2}\eta_{00}(-\bar{h}_{00}) = \frac{1}{2}\bar{h}_{00} = -\frac{2\phi}{c^2}, \\ h_{11} &= \bar{h}_{11} - \frac{1}{2}\eta_{11}(-\bar{h}_{00}) = 0 + \frac{1}{2}\bar{h}_{00} = -\frac{2\phi}{c^2}, \\ h_{22} &= -\frac{2\phi}{c^2}, \\ h_{33} &= -\frac{2\phi}{c^2}. \end{aligned} \quad (5.46)$$

It is easy to see that off-diagonal elements of $h_{\mu\nu}$ are negligible. For example,

$$h_{12} = \bar{h}_{12} - \frac{1}{2}\eta_{12}(-\bar{h}_{00}) = 0, \quad (5.47)$$

because \bar{h}_{12} is negligible and $\eta_{12} = 0$. The metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ is then

$$g_{\mu\nu} = \begin{pmatrix} -\left(1 + \frac{2\phi}{c^2}\right) & 0 & 0 & 0 \\ 0 & 1 - \frac{2\phi}{c^2} & 0 & 0 \\ 0 & 0 & 1 - \frac{2\phi}{c^2} & 0 \\ 0 & 0 & 0 & 1 - \frac{2\phi}{c^2} \end{pmatrix}. \quad (5.48)$$

In terms of spacetime distances this means

$$ds^2 = -\left(1 + \frac{2\phi}{c^2}\right) c^2 dt^2 + \left(1 - \frac{2\phi}{c^2}\right) (dx^2 + dy^2 + dz^2). \quad (5.49)$$

The inverse of the metric is

$$g^{\mu\nu} = \begin{pmatrix} -\frac{1}{1+\frac{2\phi}{c^2}} & 0 & 0 & 0 \\ 0 & \frac{1}{1-\frac{2\phi}{c^2}} & 0 & 0 \\ 0 & 0 & \frac{1}{1-\frac{2\phi}{c^2}} & 0 \\ 0 & 0 & 0 & \frac{1}{1-\frac{2\phi}{c^2}} \end{pmatrix}. \quad (5.50)$$

Next we need to establish that within the approximations made, there is some notion of particles experiencing an acceleration $\mathbf{a} = -\nabla\phi$. In general relativity, particles that are only subject to the gravitational interaction are in free fall, and follow geodesics. Hence we need to look at the geodesic equation (3.68):

$$\frac{d^2x^\beta}{d\tau^2} + \Gamma_{\mu\nu}^\beta \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \quad (5.51)$$

For particles that don't move relativistically, $dx^i/d\tau \ll dx^0/d\tau$, so that to good approximation the summations over μ and ν only leave the term with $\mu = \nu = 0$:

$$\frac{d^2x^\beta}{d\tau^2} + \Gamma_{00}^\beta \frac{dx^0}{d\tau} \frac{dx^0}{d\tau} = 0. \quad (5.52)$$

Recall that the four-velocity $V^\mu = dx^\mu/d\tau$ is normalized as $V_\mu V^\mu = -c^2$. Since in our approximation $V^\mu = (dx^0/d\tau, 0, 0, 0)$, one has $dx^0/d\tau = c$, and (5.52) becomes

$$\frac{d^2x^\beta}{d\tau^2} = -c^2 \Gamma_{00}^\beta. \quad (5.53)$$

One has

$$\begin{aligned} \Gamma_{00}^\beta &= \frac{1}{2} g^{\beta\alpha} (\partial_0 g_{\alpha 0} + \partial_0 g_{0\alpha} - \partial_\alpha g_{00}) \\ &= -\frac{1}{2} g^{\beta\alpha} \partial_\alpha g_{00}, \end{aligned} \quad (5.54)$$

where we made use of the fact that time derivatives of ϕ can be neglected, so that $\partial_0 g_{\alpha 0} = \partial_0 g_{0\alpha} = 0$. From the above equation with $\beta = 0$, recalling that the inverse metric is diagonal,

$$\Gamma_{00}^0 = -\frac{1}{2} g^{0\alpha} \partial_\alpha g_{00} = -\frac{1}{2} g^{00} \partial_0 g_{00} = 0. \quad (5.55)$$

Thus, from (5.53) we get $d^2x^0/d\tau^2 = 0$. Next, setting β to each of the spatial indices in

turn,

$$\begin{aligned}\Gamma_{00}^1 &= -\frac{1}{2}g^{11}\partial_1 g_{00} = \frac{1}{1-\frac{2\phi}{c^2}}\frac{\partial_1\phi}{c^2}, \\ \Gamma_{00}^2 &= \frac{1}{1-\frac{2\phi}{c^2}}\frac{\partial_2\phi}{c^2}, \\ \Gamma_{00}^3 &= \frac{1}{1-\frac{2\phi}{c^2}}\frac{\partial_3\phi}{c^2}.\end{aligned}\tag{5.56}$$

If ϕ/c^2 and its derivatives are considered small, then to linear order one has *e.g.*

$$\frac{1}{1-\frac{2\phi}{c^2}}\frac{\partial_1\phi}{c^2} \simeq \left(1 + \frac{2\phi}{c^2}\right)\frac{\partial_1\phi}{c^2} \simeq \frac{\partial_1\phi}{c^2}.\tag{5.57}$$

Thus, the geodesic equation (5.53) reduces to

$$\frac{d^2x^i}{d\tau^2} = -\partial_i\phi\tag{5.58}$$

for $i = 1, 2, 3$. Writing $a^i = d^2x^i/d\tau^2$, this leads to

$$\mathbf{a} = -\nabla\phi,\tag{5.59}$$

as desired.

In summary, in the limit of small velocities and weak gravitational fields, we do indeed recover Newtonian gravity, as encapsulated by the Poisson equation (5.43) and acceleration (5.59) in terms of a potential ϕ , which is related to the metric perturbation $h_{\mu\nu}$ through (5.46).

E. Physical degrees of freedom in gravitational waves

Let us return to gravitational waves, and the physical information that is in them. *A priori*, $\bar{h}_{\mu\nu}$ has 10 components, as it is a symmetric 4×4 matrix. As we now argue, most of these degrees of freedom are unphysical, since they can be eliminated by using gauge transformations of the form (5.31).

In fact, we have already imposed the Lorentz gauge (5.32),

$$\partial^\mu\bar{h}_{\mu\nu} = 0,\tag{5.60}$$

which consists of four equations, since ν is a free index that runs from 0 to 3. These four constraints already reduce the number of independent components of $\bar{h}_{\mu\nu}$ to $10 - 4 = 6$.

But even after the Lorentz gauge is imposed, there is still residual gauge freedom that can be used to further reduce the number of independent components in $\bar{h}_{\mu\nu}$. Indeed, it is an easy exercise to show that the Lorentz condition (5.32) is not spoiled by transformations (5.31) with

$$\square\xi_\mu = 0. \quad (5.61)$$

In what follows, we will be interested in what happens outside the source, where $T_{\mu\nu} = 0$, so that the equations (5.38) hold, and we only consider non-stationary solutions, as is the case with gravitational waves. It can be shown that the remaining gauge transformations can then be used to impose the following additional constraints on the components of $\bar{h}_{\mu\nu}$:

$$\boxed{\bar{h}^\mu{}_\mu = 0} \quad (5.62)$$

and

$$\boxed{\bar{h}_{\mu 0} = 0.} \quad (5.63)$$

Eq. (5.62) implies

$$\bar{h}_{\mu\nu} = h_{\mu\nu}, \quad (5.64)$$

so that we are back to the original metric perturbation in Eq. (5.12). Eq. (5.63) says that we can ignore the time-time and time-space components of $h_{\mu\nu}$ and focus only on the spatial components h_{ij} .

Eq. (5.62) is a single equation (it has no free indices), so that it decreases the number of degrees of freedom from 6 to 5. Eq. (5.63) appears to consist of 4 constraints since μ runs from 0 to 3, but by the Lorentz condition one has $\partial^\mu h_{\mu 0} = 0$, hence only three of the $h_{\mu 0}$ are still independent. Thus, there are $5 - 3 = 2$ degrees of freedom left in the end.

What do these two degrees of freedom look like explicitly? Eq. (5.39) has plane wave solutions of the form

$$h_{ij} = A_{ij}(\mathbf{k}) \cos(\omega t - \mathbf{k} \cdot \mathbf{x}), \quad (5.65)$$

where $\omega = c|\mathbf{k}|$ in order to satisfy the wave equation. We are only looking at spatial components h_{ij} , since the others have been set to zero. The Lorentz condition then reduces to $\partial^i h_{ij} = 0$. This leads to

$$k^i A_{ij} = 0. \quad (5.66)$$

Define $\hat{\mathbf{n}} = \mathbf{k}/|\mathbf{k}|$, the unit vector in the direction of propagation; then

$$n^i h_{ij} = 0. \quad (5.67)$$

Thus, the non-zero components of h_{ij} are in the plane that is transverse to $\hat{\mathbf{n}}$. The tracelessness condition (5.62) reduces to

$$h^i_i = 0. \quad (5.68)$$

Without loss of generality, we can take the direction of propagation to be along the z axis, in which case $n^i = (0, 0, 1)$. To satisfy both the conditions (5.67) and (5.68), h_{ij} must be of the form

$$h_{ij}^{\text{TT}} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos [\omega(t - z/c)]. \quad (5.69)$$

The gauge we have ended up in is called the *transverse-traceless gauge*; hence the superscript TT on h_{ij}^{TT} . The two remaining degrees of freedom are the two independent components h_+ and h_\times .

More general solutions are superpositions of plane waves of the above form, moving in arbitrary directions $\hat{\mathbf{n}}$, and having different amplitudes and frequencies. These are *gravitational waves*, and they will be the subject of the rest of this course.

VI. WHAT ARE GRAVITATIONAL WAVES?

To understand what gravitational waves really are, we need to know what is their effect on matter, how they are generated, and what energy they carry. Let us address these questions in turn.

A. The effect of gravitational waves on matter

For a plane wave of the form (5.69), the spacetime metric is

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}^{\text{TT}}$$

$$= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 + h_+ \cos[\omega(t - z/c)] & h_x \cos[\omega(t - z/c)] & 0 \\ 0 & h_x \cos[\omega(t - z/c)] & 1 - h_+ \cos[\omega(t - z/c)] & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (6.1)$$

In terms of a line element $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$, this yields

$$ds^2 = -c^2 dt^2 + (1 + h_+ \cos[\omega(t - z/c)]) dx^2 + (1 - h_+ \cos[\omega(t - z/c)]) dy^2 + 2h_x \cos[\omega(t - z/c)] dxdy + dz^2. \quad (6.2)$$

First let us look at the case where $h_x = 0$ but $h_+ \neq 0$. Then

$$ds^2 = -c^2 dt^2 + (1 + h_+ \cos[\omega(t - z/c)]) dx^2 + (1 - h_+ \cos[\omega(t - z/c)]) dy^2 + dz^2. \quad (6.3)$$

Because of the cosine, in the x and y directions, physical distances become longer and shorter in a periodic fashion, in unison but with opposite signs. Hence, space gets stretched in the x direction and compressed in the y direction, and half a period later it is the other way around. This is visualized in the top part of Fig. 22, which shows a ring of test particles in the xy plane, with the gravitational wave's propagation direction (the z direction) pointing out of the sheet.

Now consider the opposite case where $h_+ = 0$ but $h_x \neq 0$. The line element is

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + 2h_x \cos[\omega(t - z/c)] dxdy + dz^2. \quad (6.4)$$

To better understand this case, it is convenient to go to a new coordinate system in which the x' and y' axes are rotated around the z axis over an angle 45° with respect to the original

x and y axes. As can easily be shown, the line element will then take the form

$$ds^2 = -c^2 dt^2 + (1 + h_x \cos[\omega(t - z/c)]) dx'^2 + (1 - h_x \cos[\omega(t - z/c)]) dy'^2 + dz^2. \quad (6.5)$$

Hence the stretching and squeezing of space is as depicted in the lower part of Fig. 22.

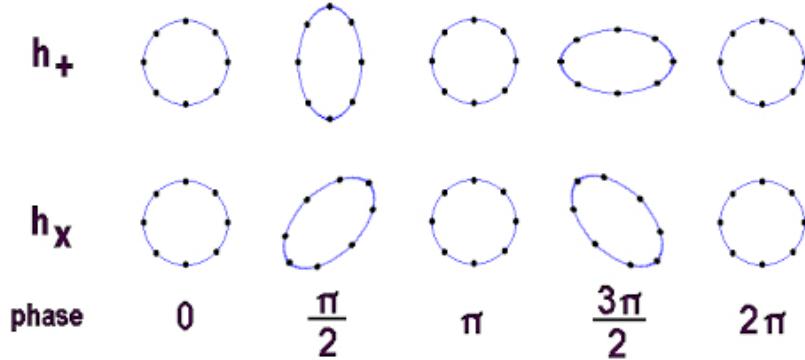


Figure 22: The deformation of a ring of test particles due to the $+$ and \times polarizations.

h_+ and h_\times are respectively called the “plus” and “cross” polarizations of a gravitational wave. Unlike in electromagnetism, where the two independent polarizations in a traveling electromagnetic wave are at right angles, the two gravitational wave polarizations are at 45° to each other, which is a consequence of the tensor character of the phenomenon. As we shall see later on, for astrophysical sources of gravitational radiation one typically has a mixture of the two polarizations, as in Eq. (6.2).

The above suggests a way to detect gravitational waves. Fig. 23 shows the principle of a *laser interferometer*. A laser beam is being shone onto a semi-transparent mirror, called a beam splitter, which causes half of the laser light to go down one long tube or “arm”, and the other half down an arm that is at right angles to it. The beams in each arm are reflected back by a mirror at the end. Things can be set up in such a way that if no gravitational wave is passing by, there is destructive interference of the two returning laser beams at the output at bottom right, so that no laser light comes out that way. If on the other hand a gravitational wave is present, then periodically one arm will be stretched and the other compressed, changing the relative distances that the laser beams have to travel through the two arms. Since the speed of light is constant, the light coming from one arm will lag behind the light from another arm, ruining the destructive interference, so that flashes of light are

observed at the output.

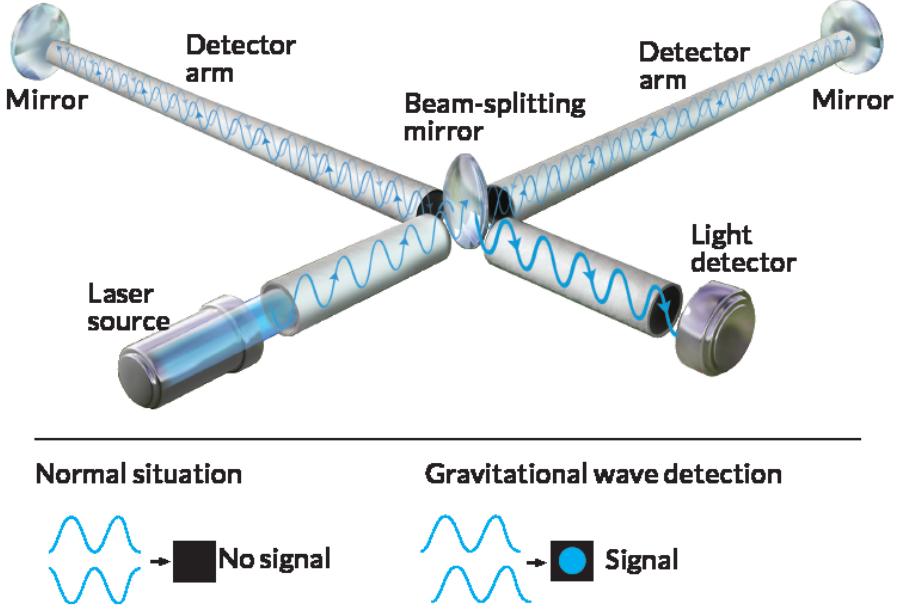


Figure 23: Illustration of the basic working principle of a laser interferometric gravitational wave detector.

There now exists a global network of such gravitational wave detectors: the two Advanced LIGO interferometers in the US (one in the state of Washington and the other in Louisiana, both with 4 km arm length), Advanced Virgo in Italy (3 km arms), KAGRA in Japan (3 km arms), and GEO600 in Germany (600 m arms). In the next few years, LIGO-India (4 km arms) will be built in the state of Maharashtra, India.

B. The quadrupole formula

The field equations of linearized gravity are (5.33)

$$\square \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}, \quad (6.6)$$

where $T_{\mu\nu}$ is the energy-momentum tensor of matter. Using the appropriate Green's function, the general solution for $\bar{h}_{\mu\nu}$ was found to be (5.37)

$$\bar{h}_{\mu\nu}(t, \mathbf{x}) = 4 \frac{G}{c^4} \int_V d^3 \mathbf{x}' \frac{T_{\mu\nu}(t - |\mathbf{x} - \mathbf{x}'|/c, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (6.7)$$

As we have seen, by making appropriate gauge choices, *outside the source* (where $T_{\mu\nu} = 0$) one can set $\bar{h}_{0\mu} = 0$ so that we can focus on the purely spatial components. The latter can

be brought into the transverse-traceless gauge, where $\bar{h}_i^i = 0$ and $n^j \bar{h}_{ij} = 0$, with $\hat{\mathbf{n}}$ a unit vector in the direction of propagation, or more generally a unit vector pointing from source to observer. In that case also $\bar{h}_{ij} = h_{ij}$.

Since outside the source there exist gauge transformations that remove the transverse and trace parts of the metric perturbation, and since gauge transformations act linearly, there must exist a linear operator $\Lambda_{ij;kl}(\hat{\mathbf{n}})$ such that

$$h_{ij}^{\text{TT}} = \Lambda_{ij;kl}(\hat{\mathbf{n}}) h^{kl} \quad (6.8)$$

is transverse-traceless. To construct this operator, first define

$$P_{ij} \equiv \delta_{ij} - n_i n_j. \quad (6.9)$$

It is easy to see that this is a projection operator, meaning that acting with it twice is the same as acting with it once:

$$P_{ik} P_{kj} = P_{ij}. \quad (6.10)$$

Also, P_{ij} projects vectors \mathbf{v} onto the plane perpendicular to $\hat{\mathbf{n}}$:

$$\hat{\mathbf{n}} \cdot (\mathbf{Pv}) = n^i P_{ij} v^j = 0. \quad (6.11)$$

From P_{ij} one then constructs $\Lambda_{ij;kl}$ itself through

$$\boxed{\Lambda_{ij;kl} \equiv P_{ik} P_{jl} - \frac{1}{2} P_{ij} P_{kl}.} \quad (6.12)$$

We leave it as an exercise to demonstrate that $h_{ij}^{\text{TT}} = \Lambda_{ij;kl} h^{kl}$ indeed satisfies

$$n^j h_{ij}^{\text{TT}} = 0 = h_i^{\text{TT}i}. \quad (6.13)$$

Using $\Lambda_{ij;kl}$, the spatial part of (6.7) can now trivially be projected to the TT gauge:

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{4G}{c^4} \Lambda_{ij;kl}(\hat{\mathbf{n}}) \int_V d^3 \mathbf{x}' \frac{T^{kl}(t - |\mathbf{x} - \mathbf{x}'|/c, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}. \quad (6.14)$$

In order to know what gravitational waves generated by a time-varying stress-energy distribution look like, we now have to perform the integral in (6.14). Generally this can not be done analytically, but good approximations can be made. Typically the observer (*i.e.* the gravitational wave detector) will be far away from the source, at a distance r (measured

from the source's center of mass) that is much larger than the source's size. Then one can write $|\mathbf{x} - \mathbf{x}'| \simeq r$, so that to good approximation

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij;kl}(\hat{\mathbf{n}}) \int_{\mathcal{V}} d^3\mathbf{x}' T^{kl} \left(t - \frac{r}{c}, \mathbf{x}' \right). \quad (6.15)$$

Now note that

$$\int d^3\mathbf{x} T^{ij} = \int d^3\mathbf{x} \delta_k^i \delta_l^j T^{kl} = \int d^3\mathbf{x} (\partial_k x^i)(\partial_l x^j) T^{kl}. \quad (6.16)$$

Using partial integration and the conservation law $\partial_\mu T^{\mu\nu} = 0$, it is not difficult to see that

$$\int d^3\mathbf{x} T^{ij} = \frac{1}{c^2} \int d^3\mathbf{x} \ddot{T}^{00} x^i x^j - \int d^3\mathbf{x} T^{ij}, \quad (6.17)$$

or

$$\int d^3\mathbf{x} T^{ij} = \frac{1}{2} \ddot{M}^{ij}, \quad (6.18)$$

where dots denote time derivatives and

$$M^{ij} \equiv \frac{1}{c^2} \int d^3\mathbf{x} T^{00} x^i x^j \quad (6.19)$$

is the *mass quadrupole moment*. Thus, within the approximations made, the gravitational wave seen by the observer is

$$h_{ij}^{\text{TT}}(t, \mathbf{x}) = \frac{1}{r} \frac{2G}{c^4} \Lambda_{ij;kl}(\hat{\mathbf{n}}) \ddot{M}^{kl}(t - r/c). \quad (6.20)$$

This is called the *quadrupole formula*.

Without loss of generality, we can take the z axis to point towards the observer. The left hand side of Eq. (6.20) then takes the form¹³

$$h_{ij}^{\text{TT}} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (6.21)$$

Now consider the right hand side of (6.20). Using the expressions (6.9) and (6.12) with $\hat{\mathbf{n}} = \hat{\mathbf{z}} = (0, 0, 1)$, one finds

$$\frac{1}{r} \frac{2G}{c^4} \begin{pmatrix} \frac{\ddot{M}^{11} - \ddot{M}^{22}}{2} & \ddot{M}^{12} & 0 \\ \ddot{M}^{12} & -\frac{\ddot{M}^{11} - \ddot{M}^{22}}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (6.22)$$

¹³ In the example of Eq. (5.69), h_+ and h_\times denoted constant amplitudes multiplying $\cos[\omega(t - z/c)]$. Henceforth our notation will be such that time dependence is included in h_+ , h_\times . This also allows for the possibility that the two polarizations come with a different time dependence from each other.

where the components are evaluated at the retarded time $t - r/c$. Setting the two sides of the matrix equation (6.20) equal to each other component by component, we find that

$$\boxed{\begin{aligned} h_+ &= \frac{1}{r c^4} (\ddot{M}^{11} - \ddot{M}^{22}), \\ h_\times &= \frac{2}{r c^4} \dot{M}^{12}. \end{aligned}} \quad (6.23)$$

The mass quadrupole moment (6.19) is one of an infinite number of quantities called *mass multipole moments*, defined as

$$M^{i_1 i_2 \dots i_N} \equiv \frac{1}{c^2} \int d^3 \mathbf{x} T^{00} x^{i_1} x^{i_2} \dots x^{i_N}. \quad (6.24)$$

What we have shown above is that, to leading order in small quantities, it is the time dependence of the quadrupole moment that causes gravitational radiation, though in reality higher-order multipole moments also contribute. But what about lower-order multipole moments? Consider the *dipole moment*

$$M^i = \frac{1}{c^2} \int d^3 \mathbf{x} T^{00} x^i. \quad (6.25)$$

For not very relativistic sources, $T_{00} = \rho c^2$ where $\rho(t, \mathbf{x})$ is the rest mass density distribution of whatever matter is present. Now specialize to the case of two point particles with masses m_1, m_2 at (possibly time dependent) positions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$, respectively. Then

$$\rho(t, \mathbf{x}) = m_1 \delta^3(\mathbf{x} - \mathbf{x}_1(t)) + m_2 \delta^3(\mathbf{x} - \mathbf{x}_2(t)), \quad (6.26)$$

and

$$\begin{aligned} M^i &= \int d^3 \mathbf{x} [m_1 \delta^3(\mathbf{x} - \mathbf{x}_1(t)) + m_2 \delta^3(\mathbf{x} - \mathbf{x}_2(t))] x^i \\ &= m_1 x_1^i(t) + m_2 x_2^i(t). \end{aligned} \quad (6.27)$$

An important point is now that *since masses are always positive*, one can always choose a coordinate system such that $M^i = 0$, namely the center-of-mass frame. Thus, there can be no gravitational dipole radiation, since the mass dipole moment can always be made to vanish just by choosing the right coordinates. This is unlike electromagnetism, where for electric charges q_1, q_2 at positions $\mathbf{x}_1(t), \mathbf{x}_2(t)$ the components of the charge dipole moment are given by

$$D^i = q_1 x_1^i(t) + q_2 x_2^i(t). \quad (6.28)$$

If q_1 and q_2 have opposite signs, then there will be not necessarily be a coordinate system that makes D^i vanish. Hence, in electromagnetism there can (and does) exist dipole radiation.

C. Energy carried by gravitational waves

So far we have been considering gravitational waves in the linearized approximation, which led to a wave equation for waves propagating on a fixed (in our case, flat) spacetime background. However, if gravitational waves carry energy, then that energy should itself cause further curvature of spacetime that is not captured by the linearized description. Therefore, in order to understand the energy in gravitational waves we need to go beyond the linear regime.

The Einstein field equations in vacuum are $G_{\alpha\beta} = 0$. Again write $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. Then the Einstein equations can be expanded as

$$0 = G_{\alpha\beta} = G_{\alpha\beta}^{(0)} + G_{\alpha\beta}^{(1)}[h_{\mu\nu}] + G_{\alpha\beta}^{(2)}[h_{\mu\nu}] + \dots, \quad (6.29)$$

where $G_{\alpha\beta}^{(0)}$ collects the terms that don't depend on $h_{\mu\nu}$ (so that $G_{\alpha\beta}^{(0)} = 0$, since it is the Einstein tensor of flat spacetime), $G_{\alpha\beta}^{(1)}$ collects the terms that are linear in $h_{\mu\nu}$ and its derivatives, $G_{\alpha\beta}^{(2)}$ the terms that are quadratic, and so on. Suppose we now formally write

$$g_{\mu\nu} = \eta_{\mu\nu} + \epsilon h_{\mu\nu}^{(1)} + \epsilon^2 h_{\mu\nu}^{(2)} + \dots, \quad (6.30)$$

where $\epsilon \ll 1$, and $h_{\mu\nu}^{(1)}, h_{\mu\nu}^{(2)}, \dots$ are $\mathcal{O}(1)$. Then Eq. (6.29) becomes

$$\begin{aligned} 0 &= G_{\alpha\beta}^{(1)}[\epsilon h_{\mu\nu}^{(1)}] + G_{\alpha\beta}^{(1)}[\epsilon^2 h_{\mu\nu}^{(2)}] + G_{\alpha\beta}^{(2)}[\epsilon h_{\mu\nu}^{(1)}] + \dots \\ &= \epsilon G_{\alpha\beta}^{(1)}[h_{\mu\nu}^{(1)}] + \epsilon^2 G_{\alpha\beta}^{(1)}[h_{\mu\nu}^{(2)}] + \epsilon^2 G_{\alpha\beta}^{(2)}[h_{\mu\nu}^{(1)}] + \dots \\ &= \epsilon G_{\alpha\beta}^{(1)}[h_{\mu\nu}^{(1)}] + \epsilon^2 \left(G_{\alpha\beta}^{(1)}[h_{\mu\nu}^{(2)}] + G_{\alpha\beta}^{(2)}[h_{\mu\nu}^{(1)}] \right) + \dots \end{aligned} \quad (6.31)$$

Since this equation must hold order by order in ϵ , one finds

$$\begin{aligned} G_{\alpha\beta}^{(1)}[\epsilon h_{\mu\nu}^{(1)}] &= 0, \\ G_{\alpha\beta}^{(1)}[\epsilon^2 h_{\mu\nu}^{(2)}] &= -G_{\alpha\beta}^{(2)}[\epsilon h_{\mu\nu}^{(1)}]. \end{aligned} \quad (6.32)$$

These equations give us the first order ($\epsilon h_{\mu\nu}^{(1)}$) and second-order ($\epsilon^2 h_{\mu\nu}^{(2)}$) corrections to the background metric $\eta_{\mu\nu}$. Note that the first equation above, which sets to zero the part of the Einstein tensor that is linear in the (derivatives of) the first-order corrections to the metric, is formally the same as the linearized Einstein equation (5.29), with $\bar{h}_{\mu\nu} = \epsilon h_{\mu\nu}^{(1)} - (1/2)\eta_{\mu\nu}\epsilon h^{(1)}$, and the right hand side set to zero. The second equation (6.32) gives the quadratic corrections to the background metric that are induced by the energy present in

the first-order corrections. Therefore it makes sense to interpret it as an energy-momentum tensor, namely the one that describes the energy-momentum carried by gravitational waves, at least to the order considered here. Thus, one writes

$$G_{\alpha\beta}^{(1)}[\epsilon^2 h_{\mu\nu}^{(2)}] = \frac{8\pi G}{c^4} t_{\alpha\beta}, \quad (6.33)$$

where by definition

$$t_{\alpha\beta} \equiv -\frac{c^4}{8\pi G} G_{\alpha\beta}^{(2)}[\epsilon h_{\mu\nu}^{(1)}]. \quad (6.34)$$

One problem with $t_{\alpha\beta}$ as defined above is that it is not gauge invariant, though one would expect this to be the case for a physically meaningful tensor defined in terms of the linear metric perturbation $\epsilon h_{\mu\nu}^{(1)}$. As it turns out, this can be solved by averaging the right hand side over at least one wavelength of the gravitational wave. Thus,

$$t_{\alpha\beta} \equiv -\frac{c^4}{8\pi G} \left\langle G_{\alpha\beta}^{(2)}[h_{\mu\nu}] \right\rangle, \quad (6.35)$$

where $h_{\mu\nu} = \epsilon h_{\mu\nu}^{(1)}$ is the linear metric perturbation. One can show that

$$\partial^\alpha t_{\alpha\beta} = 0, \quad (6.36)$$

as expected for an energy-momentum tensor.

Finally, it can be shown that in the transverse-traceless gauge, using the vacuum wave equation $\square h_{ij}^{\text{TT}} = 0$, one has

$$t_{\alpha\beta} = \frac{c^4}{32\pi G} \left\langle \partial_\alpha h_{ij}^{\text{TT}} \partial_\beta h_{ij}^{\text{TT}} \right\rangle. \quad (6.37)$$

Hence the energy density in gravitational waves is

$$\begin{aligned} t^{00} &= \frac{c^2}{32\pi G} \left\langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \right\rangle \\ &= \frac{c^2}{16\pi G} \left\langle \dot{h}_+^2 + \dot{h}_\times^2 \right\rangle. \end{aligned} \quad (6.38)$$

Similarly, the momentum density is

$$t^{0k} = -\frac{c^3}{32\pi G} \left\langle \dot{h}_{ij}^{\text{TT}} \partial^k h_{ij}^{\text{TT}} \right\rangle. \quad (6.39)$$

Consider a spatial volume \mathcal{V} bounded by a sphere \mathcal{S} of radius r . Then the gravitational wave energy in that volume is

$$E_{\mathcal{V}} = \int_{\mathcal{V}} d^3 \mathbf{x} t^{00}. \quad (6.40)$$

The energy emitted in gravitational waves is

$$\begin{aligned}\frac{dE_{\text{GW}}}{dt} &= - \int_V d^3\mathbf{x} \frac{\partial}{\partial t} t^{00} \\ &= -c \int_V d^3\mathbf{x} \partial_0 t^{00} \\ &= c \int_V d^3\mathbf{x} \partial_k t^{0k},\end{aligned}\quad (6.41)$$

where in the last line we have used $\partial_\mu t^{0\mu} = 0 = \partial_0 t^{00} + \partial_k t^{0k}$. This leads us to the energy passing through the sphere \mathcal{S} :

$$\frac{dE_{\text{GW}}}{dt} = c \int_{\mathcal{S}} dA t^{0r}, \quad (6.42)$$

with

$$t^{0r} = \frac{c^4}{32\pi G} \left\langle \partial^0 h_{ij}^{\text{TT}} \frac{\partial}{\partial r} h_{ij}^{\text{TT}} \right\rangle. \quad (6.43)$$

It is easy to see that at sufficiently large distance from the source, a gravitational wave propagating radially outward takes the general form

$$h_{ij}^{\text{TT}} = \frac{1}{r} f_{ij}(t - r/c), \quad (6.44)$$

from which

$$\frac{\partial}{\partial r} h_{ij}^{\text{TT}}(t, r) = \partial^0 h_{ij}^{\text{TT}}(t, r) + \mathcal{O}(1/r^2). \quad (6.45)$$

Hence Eq. (6.42) becomes

$$\frac{dE_{\text{GW}}}{dt} = c \int_{\mathcal{S}} dA t^{00}, \quad (6.46)$$

where the area element is $dA = r^2 d\Omega$, with r the sphere's radius and $d\Omega = \sin(\theta) d\theta d\phi$ in the usual angular coordinates (θ, ϕ) . This leads to

$$\frac{dE_{\text{GW}}}{dt} = \frac{c^3 r^2}{32\pi G} \int d\Omega \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle, \quad (6.47)$$

or in terms of the two polarizations,

$$\frac{dE_{\text{GW}}}{dt} = \frac{c^3 r^2}{16\pi G} \int d\Omega \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle.$$

(6.48)

Finally, the energy flux in gravitational waves is

$$\mathcal{F}_{\text{GW}} = \frac{dE_{\text{GW}}}{dt dA} = \frac{c^3}{16\pi G} \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle.$$

(6.49)

VII. THE COALESCENCE OF COMPACT BINARY OBJECTS

The network of Advanced LIGO and Virgo interferometers has been detecting gravitational wave signals from coalescing binary neutron stars and black holes on a regular basis. The process is illustrated in Fig. 24 in the case of black holes. To begin with, the two objects can be in wide, typically elliptical orbits around some common center of gravity. As we have seen, gravitational waves take away energy, so the orbits will shrink. At the same time (though we did not show this), gravitational waves also carry away angular momentum, causing the orbits to become more and more circular. By the time the two black holes are sufficiently close together that the frequency is high enough for the GW signal to be in the sensitive band of our detectors, one can assume that the orbits have been circularized, and we have quasi-circular *inspiral*. This goes on until some last stable orbit is reached, after which the objects *plunge* towards each other and *merge* to form a single, highly excited black hole. The latter will then undergo *ringdown*, quickly shedding its excitations until it settles into a dormant state. (In the case of neutron stars, the remnant object can be a massive, very asymmetric neutron star which sheds its excitations in a different way, and may take some time to collapse to a black hole.)

The two-body problem in general relativity has no analytic solution, so that approximations are needed to calculate the shape of the gravitational wave signal that gets emitted. To describe the early inspiral, an approximation scheme can be used, called the *post-Newtonian expansion*, which is an extension of the treatment in the previous section to include both non-linear effects and higher multipole moments. At plunge and merger one is in the strong-field regime of gravity, and the full Einstein equations can no longer be avoided; here much insight has been gained from large-scale numerical simulations, in which a discretized version of the Einstein equations is used to evolve an entire binary black hole spacetime on a computer. Finally, for the ringdown one can again employ a perturbative treatment.

Here we will apply what we have learned in the previous section, arriving at a more crude approximation of the inspiral process than what is provided by the post-Newtonian formalism, but which will nevertheless be reasonably accurate. First we will consider binary objects on fixed circular orbits, *i.e.* neglecting the shrinking of the objects due to the energy lost in gravitational waves. After that we will take energy loss into account and calculate how this causes a gravitational wave signal whose amplitude and frequency are steadily

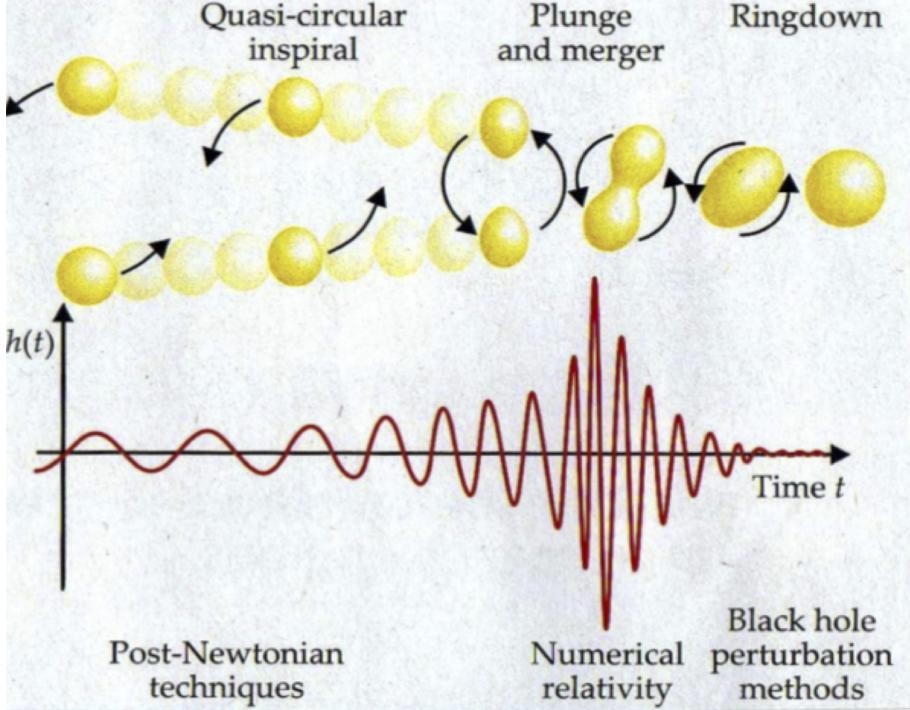


Figure 24: The inspiral-merger-ringdown process for two black holes (top), and the shape of the gravitational wave that is produced (bottom, in red).

increasing with time.

A. Two point particles on fixed circular orbits

Consider two point particles with masses m_1, m_2 moving on circular orbits, at a fixed distance R from each other. It will be convenient to use a coordinate system whose origin is the center of mass, in which case it is easy to show that the positions $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ of the particles take the form

$$\begin{aligned}\mathbf{x}_1(t) &= \frac{m_2}{m_1 + m_2} R \hat{\mathbf{e}}(t) = \frac{\mu}{m_1} R \hat{\mathbf{e}}(t), \\ \mathbf{x}_2(t) &= -\frac{m_1}{m_1 + m_2} R \hat{\mathbf{e}}(t) = -\frac{\mu}{m_2} R \hat{\mathbf{e}}(t),\end{aligned}\quad (7.1)$$

where $\hat{\mathbf{e}}(t)$ is a unit vector pointing from the center of mass to the object with mass m_1 , and $\mu = m_1 m_2 / (m_1 + m_2)$ is the *reduced mass*. The z axis of the coordinate system is chosen such that it points to the observer. The unit normal to the orbital plane is assumed to be tilted around the x axis at an angle ι (the *inclination angle*) with respect to the z axis; we

leave it as an exercise to show that then

$$\hat{\mathbf{e}}(t) = (\cos(\omega t), \cos(\iota) \sin(\omega t), \sin(\iota) \sin(\omega t)), \quad (7.2)$$

with ω the orbital angular frequency. If the two point particles are still far apart from each other and moving slowly compared to the speed of light, the quadrupole moment becomes

$$\begin{aligned} M^{ij}(t) &= \frac{1}{c^2} \int d^3\mathbf{x} T^{00}(t, \mathbf{x}) x^i x^j \\ &= \int d^3\mathbf{x} \rho(t, \mathbf{x}) x^i x^j, \end{aligned} \quad (7.3)$$

where the mass density is

$$\begin{aligned} \rho(t, \mathbf{x}) &= m_1 \delta^3(\mathbf{x} - \mathbf{x}_1(t)) + m_2 \delta^3(\mathbf{x} - \mathbf{x}_2(t)) \\ &= m_1 \delta^3 \left(\mathbf{x} - \frac{\mu}{m_1} R \hat{\mathbf{e}}(t) \right) + m_2 \delta^3 \left(\mathbf{x} + \frac{\mu}{m_2} R \hat{\mathbf{e}}(t) \right). \end{aligned} \quad (7.4)$$

Performing the integration in (7.3), we get

$$\begin{aligned} M^{ij}(t) &= \left[m_1 \frac{\mu^2}{m_1^2} R^2 + m_2 \frac{\mu^2}{m_2^2} R^2 \right] \hat{e}^i \hat{e}^j \\ &= \mu R^2 \hat{e}^i \hat{e}^j. \end{aligned} \quad (7.5)$$

This we can insert in the expressions for the gravitational wave polarizations, Eq. (6.23):

$$\begin{aligned} h_+ &= \frac{1}{r c^4} (\ddot{M}^{11} - \ddot{M}^{22}), \\ h_\times &= \frac{2}{r c^4} \ddot{M}^{12}, \end{aligned} \quad (7.6)$$

where the components of the quadrupole moment are to be evaluated at the retarded time $t_{\text{ret}} = t - r/c$. After some algebra we find

$$\begin{aligned} h_+ &= -\frac{4 G \mu R^2 \omega^2}{r c^4} \frac{1 + \cos^2(\iota)}{2} \cos(2\omega t_{\text{ret}}), \\ h_\times &= -\frac{4 G \mu R^2 \omega^2}{r c^4} \cos(\iota) \sin(2\omega t_{\text{ret}}). \end{aligned} \quad (7.7)$$

Two properties of the expressions (7.7) are of note:

- The gravitational radiation comes out with *twice* the orbital frequency. This can be understood by looking at the quadrupole moment, Eq. (7.5). After half an orbital period, $\hat{\mathbf{e}} \rightarrow -\hat{\mathbf{e}}$, and since the quadrupole moment is quadratic in $\hat{\mathbf{e}}$, $M^{ij} \rightarrow M^{ij}$. Hence the period of the quadrupole moment is half of the period of the orbital motion, so that it has twice the frequency.

- When $\iota = 90^\circ$, the observer is viewing the orbital plane “edge-on”, so that from her perspective the motion of the two objects is along a straight line. It then makes sense that in this case only h_+ is non-zero. If on the other hand $\iota = 0$ or $\iota = 180^\circ$, the orbital plane is viewed “face-on” or “face-off”, in which case the two polarizations are present with equal amplitudes.

Before going on, we note that in the amplitudes of Eqs. (7.7), ω , R , and mass are not independent quantities, which can be seen as follows. When a particle with mass m is rotating on a circle of radius d with velocity v , the centripetal force on it is $F_C = mv^2/d$. In our case, point particle 1 has $d = (\mu/m_1)R$, and $v = d\omega = (\mu/m_1)R\omega$, so that

$$F_C = m_1 \frac{[(\mu/m_1)R\omega]^2}{(\mu/m_1)R} = \mu R\omega^2. \quad (7.8)$$

On the other hand, the gravitational force between the two particles, which we recall are at distance R from one another, is given by

$$F_G = \frac{Gm_1m_2}{R^2}. \quad (7.9)$$

Setting $F_C = F_G$ yields

$$R^3 = \frac{GM}{\omega^2}, \quad (7.10)$$

where $M = m_1 + m_2$. This is in fact nothing but an instance of *Kepler's third law*. In Eqs. (7.7), we can now get rid of R in favor of mass and frequency. It is convenient to define the *chirp mass* as

$$\mathcal{M}_c \equiv \frac{(m_1m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \quad (7.11)$$

Then

$$\boxed{\begin{aligned} h_+ &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\omega}{c} \right)^{2/3} \frac{1 + \cos^2(\iota)}{2} \cos(2\omega t_{\text{ret}}), \\ h_\times &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\omega}{c} \right)^{2/3} \cos(\iota) \sin(2\omega t_{\text{ret}}). \end{aligned}} \quad (7.12)$$

B. Inspiral

Since gravitational waves are emitted which will take away energy from the system, the orbits can not remain fixed. In particular, the distance R between the point particles will decrease, and through Kepler's third law (7.10) this means that the orbital frequency ω

must increase in time. Let $\Phi(t)$ be the angle that the line connecting the two particle sweeps through as a function of time; then from the definition of angular velocity,

$$\omega(t) = \dot{\Phi}(t) \quad \text{or} \quad \Phi(t) = \int^t dt' \omega(t'). \quad (7.13)$$

For the positions of the point particles as a function of time, we can now write

$$\begin{aligned} \mathbf{x}_1(t) &= \frac{\mu}{m_1} R(t) \hat{\mathbf{e}}(t), \\ \mathbf{x}_2(t) &= -\frac{\mu}{m_2} R(t) \hat{\mathbf{e}}(t), \end{aligned} \quad (7.14)$$

where

$$\hat{\mathbf{e}}(t) = (\cos(\Phi(t)), \cos(\iota) \sin(\Phi(t)), \sin(\iota) \sin(\Phi(t))). \quad (7.15)$$

In principle we would now have to redo the calculation that led to (7.12). However, assuming “adiabatic” inspiral, meaning that the orbital decay is slow and over a single orbit the trajectories of the particles can be considered approximately circular, it will suffice to make the following substitutions in (7.12):

$$\omega \longrightarrow \omega(t_{\text{ret}}), \quad \omega t_{\text{ret}} \longrightarrow \Phi(t_{\text{ret}}), \quad (7.16)$$

so that

$$\begin{aligned} h_+ &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\omega(t_{\text{ret}})}{c} \right)^{2/3} \frac{1 + \cos^2(\iota)}{2} \cos(2\Phi(t_{\text{ret}})), \\ h_\times &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\omega(t_{\text{ret}})}{c} \right)^{2/3} \cos(\iota) \sin(2\Phi(t_{\text{ret}})). \end{aligned} \quad (7.17)$$

What remains to be done is to find an expression for the instantaneous angular frequency $\omega(t)$, which via (7.13) will also yield $\Phi(t)$ by taking an integral over time.

We do this by equating the loss in orbital energy to the power emitted in gravitational waves:

$$\frac{dE_{\text{orb}}}{dt} = -\frac{dE_{\text{GW}}}{dt} \quad (7.18)$$

The orbital energy is given by

$$E_{\text{orb}} = \frac{1}{2}mv_1^2 + \frac{1}{2}mv_2^2 - \frac{Gm_1m_2}{R}. \quad (7.19)$$

Substituting $v_1 = (\mu/m_1)R\omega_1$ and $v_2 = (\mu/m_2)R\omega_2$, and using Kepler’s third law (7.10) to eliminate R , one finds

$$E_{\text{orb}} = -\frac{1}{2}(G^2\mathcal{M}_c^5\omega^2)^{1/3}. \quad (7.20)$$

On the other hand, the gravitational wave power is (6.48)

$$\frac{dE_{\text{GW}}}{dt} = \frac{c^3 r^2}{16\pi G} \int d\Omega \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle. \quad (7.21)$$

The integral over the sphere is an integral over all inclinations ι and azimuths ϕ , so that $d\Omega = \sin(\iota) d\iota d\phi$. Furthermore, the angular brackets denote averaging over a wavelength, or equivalently a period of the wave. Plugging in the polarizations (7.17), after some work one gets

$$\frac{dE_{\text{GW}}}{dt} = \frac{32}{5} \frac{c^5}{G} \left(\frac{G\mathcal{M}_c \omega}{c^3} \right)^{10/3}. \quad (7.22)$$

Before continuing, it is convenient to define the *gravitational wave frequency* $f_{\text{gw}} = 2f_{\text{orb}}$ (so that $\omega = 2\pi f_{\text{orb}} = \pi f_{\text{gw}}$). Then the “energy balance” equation (7.18) gives us a differential equation for $f_{\text{gw}}(t)$:

$$\dot{f}_{\text{gw}}(t_{\text{ret}}) = \frac{96}{5} \pi^{8/3} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{5/3} f_{\text{gw}}^{11/3}(t_{\text{ret}}). \quad (7.23)$$

Solving this equation finally yields

$$f_{\text{gw}}(t) = \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-5/8} \left(\frac{5}{256} \frac{1}{\tau(t)} \right)^{3/8} \quad (7.24)$$

in terms of

$$\tau(t) = t_c - t. \quad (7.25)$$

The integration constant t_c is called the *coalescence time*, which is the time at which the frequency formally diverges: $t \rightarrow t_c$ implies $f_{\text{gw}} \rightarrow \infty$. From Kepler’s third law, this implies that the separation between the point particles goes to zero: $R \rightarrow 0$.

The binary objects whose inspiral causes gravitational waves that are detectable with interferometers, are neutron stars and black holes, which of course are not point particles; even a black hole with mass m has a non-zero size which is the Schwarzschild radius $R_{\text{Schw}} = 2Gm/c^2$. One might think that this finite size is what prevents the gravitational wave frequency from diverging, in that the inspiral ends when the two objects start touching each other, after which different physics takes over. It turns out that in reality, the inspiral can in fact terminate even earlier than that. To see this we would need to take recourse to full general relativity. Instead we state without proof that in the full theory, an *innermost stable circular orbit* (ISCO) is reached when the objects are at a separation $R_{\text{ISCO}} \simeq 6GM/c^2$; after that quasi-circular inspiral is no longer possible and the objects plunge towards each

other and merge together. Using Kepler's third law, it is easy to see that this corresponds to a gravitational wave frequency

$$f_{\text{gw,ISCO}} = \frac{c^3}{6^{3/2}\pi GM}. \quad (7.26)$$

Now that we have an expression for f_{gw} as a function of time, or equivalently for $\omega(t) = \pi f_{\text{gw}}(\tau(t))$, we can finally calculate $\Phi(t)$ through Eq. (7.13). Writing $\Phi_{\text{gw}} = 2\Phi$,

$$\Phi_{\text{gw}}(t) = -2 \left(\frac{5G\mathcal{M}_c}{c^3} \right)^{-5/8} \tau^{5/8}(t) + \Phi_c, \quad (7.27)$$

where Φ_c is the phase at $t = t_c$.

We now have all the ingredients to write down the polarizations h_+ and h_\times , this time taking into account the orbital evolution due to energy loss resulting from the gravitational wave emission:

$$h_+ = -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \frac{1 + \cos^2(\iota)}{2} \cos(\Phi_{\text{gw}}(t_{\text{ret}})),$$

$$h_\times = -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \cos(\iota) \sin(\Phi_{\text{gw}}(t_{\text{ret}})),$$

(7.28)

where $f_{\text{gw}}(t)$ is given by Eq. (7.24), and $\Phi_{\text{gw}}(t)$ by Eq. (7.27).

VIII. HOW ARE GRAVITATIONAL WAVES DETECTED?

We have seen that gravitational waves have the effect of periodically stretching and compressing space, and hinted at a way of utilizing the way they affect an interferometer in order to detect them. Here we will look into more detail as to how this happens, what the detector output is, and how a weak signal can be dug out of the usually very noisy data.

A. The response of an interferometer to a gravitational wave

Fig. 25 illustrates the general set-up. First we consider a Cartesian coordinate system (x, y, z) associated with the interferometer, where the x and y axes point along the two arms. Associated with a gravitational wave there is a primed coordinate system (x', y', z') , where z' is the propagation direction. In the primed coordinate system one has

$$h_{ij}'^{\text{TT}} = \begin{pmatrix} h_+ & h_x & 0 \\ h_x & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (8.1)$$

but the metric perturbation h_{ij} in the detector frame (x, y, z) will not take such a simple form. Nevertheless, first we focus on the latter.

As explained in Sec. V, when a gravitational wave passes by then periodically one arm will get longer and the other shorter. The change in the way laser beams coming from the two arms interfere with each other translates into the relative length $\Delta L/L$ between the two arms, where L is the length of each arm in the absence of gravitational waves. Now assume that the coordinate system is “anchored” to the beam splitter and end mirrors, so that the *coordinate distances* L between the beam splitter and either of the mirrors never changes, although of course the *physical distances* computed using the metric may change. One has

$$ds^2 = -c^2 dt^2 + (\delta_{ij} + h_{ij}) dx^i dx^j. \quad (8.2)$$

Now, as far as the arm along the x direction is concerned, what matters is only the induced length changes along the x direction, and similarly for the arm along the y direction. Another way of saying this is that for *e.g.* the x arm, what is important is only the change in distance between the beam splitter and the mirror in the x direction, not any change in position of the mirror in the sideways directions. Hence, when considering what happens to the x arm

we can set $dy = dz = 0$, and for the y arm, $dx = dz = 0$. Moreover, since light moves on null geodesics, $ds^2 = 0$. For the x arm we can then write

$$c^2 dt^2 = (1 + h_{xx}) dx^2, \quad (8.3)$$

and for the y arm

$$c^2 dt^2 = (1 + h_{yy}) dy^2. \quad (8.4)$$

The light travel time in the x arm from the end mirror at position $x = x_{\text{mirror}}$ to the beam splitter at $x = x_{\text{BS}}$ is then

$$\begin{aligned} \Delta t_x &= \int_{x_{\text{mirror}}}^{x_{\text{BS}}} dt = \frac{1}{c} \int_{x_{\text{mirror}}}^{x_{\text{BS}}} (1 + h_{xx})^{1/2} dx \\ &\simeq (1 + h_{xx})^{1/2} (x_{\text{BS}} - x_{\text{mirror}})/c \\ &\simeq (1 + h_{xx})^{1/2} L/c \\ &\simeq \left(1 + \frac{1}{2}h_{xx}\right) L/c. \end{aligned} \quad (8.5)$$

In the second line we have assumed that h_{xx} does not vary much either in space or in time as the laser light is making a single trip along the interferometer arm. This is the same as saying that the wavelength of the wave is much larger than the length of an arm, which will be true for signals that are in the sensitive frequency band of *e.g.* LIGO or Virgo. Similarly, for the y arm,

$$\Delta t_y \simeq \left(1 + \frac{1}{2}h_{yy}\right) L/c. \quad (8.6)$$

In the absence of a gravitational wave ($h_{xx} = h_{yy} = 0$), the travel time is simply

$$\Delta t_0 = L/c \quad (8.7)$$

in both arms. Since interferometers sense the relative difference of laser light arrival times at the beam splitter (through the mismatch between the arrival time of a wave crest from one arm and a wave trough from the other), the output of a detector is essentially

$$h(t) \equiv \frac{\Delta t_x - \Delta t_y}{\Delta t_0} = \frac{1}{2}(h_{xx} - h_{yy}).$$

(8.8)

This is called the *gravitational wave strain*.

In Eq. (8.8), h_{xx} and h_{yy} refer to the unprimed coordinate system associated with the detector. We now need to relate these to the components of the metric perturbation h'_{ij}^{TT}

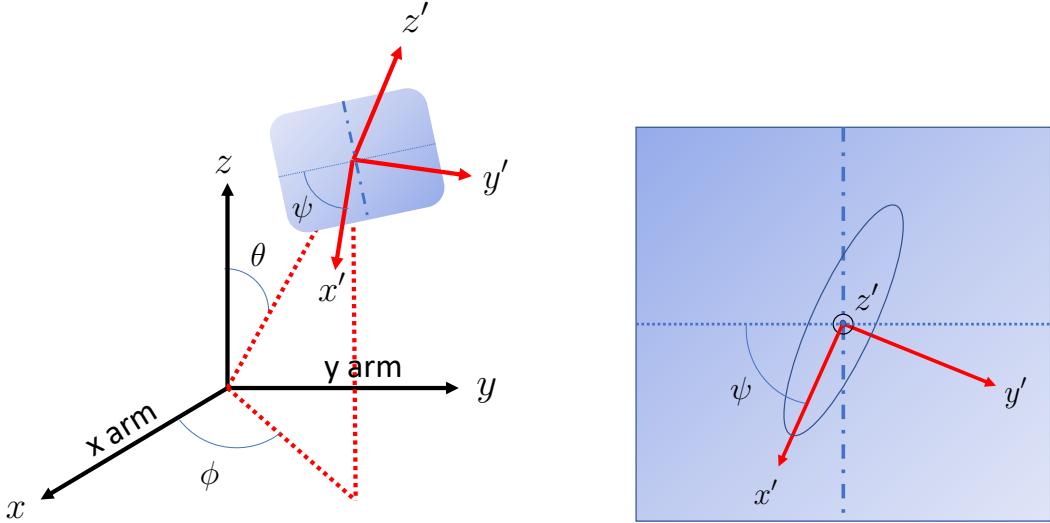


Figure 25: An illustration of the conventions used in the main text to understand the response of an interferometer (to which is attached the unprimed coordinate system (x, y, z)) to a gravitational wave coming from some arbitrary direction (to which is associated the primed coordinate system (x', y', z')). Left: The wave propagates in the direction z' , which is at angles (θ, ϕ) with respect to the unprimed coordinates. Right: The patch of the celestial sphere which the gravitational wave pierces, with the horizontal dotted line parallel to the xy plane. The z' axis points out of the sheet. In the case of binary inspiral, one chooses the x' axis to be aligned with the long axis of the projection onto the sky of an orbit (shown as an ellipse), which defines the “polarization angle” ψ .

in the primed coordinate system, which are given by (8.1). Let (θ, ϕ) give the direction of z' in a spherical coordinate system; this determines the direction of propagation.¹⁴ Then transforming from the primed to the unprimed coordinate system is effected by first rotating over an angle ψ around the z' axis to make x' parallel to the (x, y) plane, next rotating over an angle θ around the x' axis to align the z' axis with the z axis, followed by a rotation over an angle ϕ around the z axis to go from the x' axis to the x axis. The corresponding rotation matrix is

$$\mathcal{R} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (8.9)$$

¹⁴ Hence the sky position from which the gravitational wave originates is given by $(\pi - \theta, \phi + \pi)$.

In the (x, y, z) frame, the metric perturbation is

$$h_{ij} = (\mathcal{R} h'^{\text{TT}} \mathcal{R}^T)_{ij}, \quad (8.10)$$

where \mathcal{R}^T denotes the transpose of \mathcal{R} . This finally gives us the strain (8.8) in terms of the + and \times polarizations of the gravitational wave:

$$h = \frac{1}{2}(h_{xx} - h_{yy}) = F_+(\theta, \phi, \psi) h_+ + F_\times(\theta, \phi, \psi) h_\times, \quad (8.11)$$

where the *beam pattern functions* (also called antenna pattern functions) are given by

$$\begin{aligned} F_+(\theta, \phi, \psi) &= \frac{1}{2}(1 + \cos^2(\theta)) \cos(2\phi) \cos(2\psi) - \cos(\theta) \sin(2\phi) \sin(2\psi), \\ F_\times(\theta, \phi, \psi) &= \frac{1}{2}(1 + \cos^2(\theta)) \cos(2\phi) \sin(2\psi) + \cos(\theta) \sin(2\phi) \cos(2\psi). \end{aligned} \quad (8.12)$$

B. Finding a signal in a lot of noise

Interferometers are subject to a range of nuisance factors that produce noise in the output, including but not limited to seismic noise due to human activity in the region where the detector resides, thermal vibrations in the suspensions of the mirrors, and the fact that light has a particle as well as a wave character. So far we have pretended that detectors are noise-free and that the output (the strain) is solely determined by the action of gravitational waves on the interferometer arms. In reality the measured strain, $s(t)$, will be a superposition of the effect of a gravitational wave, $h(t)$, and the noise, $n(t)$:

$$s(t) = n(t) + h(t). \quad (8.13)$$

A way to check for the presence of a signal alongside detector noise is *matched filtering*. If the shape of the signal $h(t)$ is more or less known, then one can integrate it against the output and divide by the observation time T :

$$\frac{1}{T} \int_0^T s(t) h(t) dt = \frac{1}{T} \int_0^T n(t) h(t) dt + \frac{1}{T} \int_0^T h(t)^2 dt. \quad (8.14)$$

Both $n(t)$ and $h(t)$ are oscillating functions. However, the integrand of the second integral in the right hand side is positive definite. Heuristically,

$$\frac{1}{T} \int_0^T h(t)^2 dt \sim h_0^2, \quad (8.15)$$

where h_0 is the characteristic amplitude of the signal. On the other hand, the quantity $n(t) h(t)$ is oscillating, and its integral will only grow as $T^{1/2}$ for large T , as is typical of a random walk process, so that

$$\frac{1}{T} \int_0^T n(t) h(t) dt \sim \left(\frac{\tau_0}{T}\right)^{1/2} n_0 h_0, \quad (8.16)$$

where n_0 is the characteristic amplitude of $n(t)$, and τ_0 a time scale associated with $h(t)$, which roughly speaking will be the period of the gravitational wave. Thus, in the limit $T \rightarrow \infty$, the first term in the right hand side of Eq. (8.14) goes to zero, and we have “filtered out” the contribution of the noise to the output. If a signal is really present in the noise, then the second term will indicate this. If not, then (8.14) only has the noise term, and the average of $s(t) h(t)$ over the observation time will go to zero as $T \rightarrow \infty$. In principle this provides us with a simple way of determining whether the detector output contains a signal or not.

In practice it will not be possible to take the limit $T \rightarrow \infty$: the observation time will anyway be finite, and in the case of binary neutron star and black hole coalescences, the signal $h(t)$ is itself of finite duration, cutting off the time integral when it terminates. Nevertheless, the above tells us something important about the detectability of a signal. One might think that detection requires the signal to stick out above the noise ($h_0 > n_0$), but this is not the case. Instead it is enough for the second term in Eq. (8.14) to be sufficiently large compared to the typical magnitude of the first term, *i.e.*

$$h_0 \gg \left(\frac{\tau_0}{T}\right)^{1/2} n_0. \quad (8.17)$$

In the case of binary neutron star inspirals, effectively $T \sim 100$ s, and though the frequency f is monotonically increasing, we can take τ_0 to be the period at $f \sim 100$ Hz, where the detectors tend to be the most sensitive; this implies $\tau \sim 10^{-2}$ sec. Hence for this type of gravitational wave source, $(\tau_0/T)^{1/2} \sim 10^{-2}$, so that at least in principle a signal can be 100 times weaker than the noise and still be detectable. An even more extreme case is the continuous signal from a fast-spinning, slightly asymmetric pulsar. For millisecond pulsars, $\tau_0 \sim 10^{-3}$ s, so that for an observation time of a year, one has $(\tau_0/T)^{1/2} \sim 10^{-5}$. Since signals from isolated neutron stars are expected to be intrinsically weak, the corresponding gain in detectability will in fact be necessary to have a chance of finding them.

Eq. (8.14) conveys the basic idea of matched filtering, but there is more to it. In particular, the above does not make use of known properties of the detector. Folding those in will

allow us to optimize the detection process. In this context, “knowing the properties of the detector” means “knowing the properties of the detector noise”. This is what we turn to next.

C. The properties of the noise

The output of a detector gets digitized, so that in the absence of a signal, one records a time series

$$(n(t_0), n(t_1), \dots, n(t_M)), \quad (8.18)$$

where $t_{i+1} = t_i + \Delta t$ for some fixed, small Δt . Often it will be convenient to take a (discrete) Fourier transform, so as to arrive at a frequency series

$$(\tilde{n}(f_0), \tilde{n}(f_1), \dots, \tilde{n}(f_N)), \quad (8.19)$$

where $f_{i+1} = f_i + \Delta f$, with $\Delta f = 1/T$, where T is the observation time. Let us denote $\tilde{n}_i = \tilde{n}(f_i)$, so that the frequency series is

$$(\tilde{n}_0, \tilde{n}_1, \dots, \tilde{n}_N). \quad (8.20)$$

A specific set of values for the \tilde{n}_i is called a *noise realization*. Though these are random, some noise realizations will be more probable than others. At a given frequency f_i , denote the probability distribution for \tilde{n}_i as $p(\tilde{n}_i)$. As it turns out, to good approximation, this can be described by a Gaussian centered on zero:

$$p(\tilde{n}_i) \propto e^{-\frac{|\tilde{n}_i|^2}{2\sigma_i^2}}. \quad (8.21)$$

Noise that satisfies Eq. (8.21) is called *Gaussian noise*. A probability distribution allows us to calculate expectation values $\langle \cdot \rangle$. For example, in our case

$$\langle \tilde{n}_i \rangle \equiv \int \tilde{n}_i p(\tilde{n}_i) d\tilde{n}_i = 0, \quad (8.22)$$

and

$$\langle |\tilde{n}_i|^2 \rangle = \int |\tilde{n}_i|^2 p(\tilde{n}_i) d\tilde{n}_i = 2\sigma_i^2. \quad (8.23)$$

Noise that satisfies Eqs. (8.22) and (8.23) is called *stationary*.

In addition to the above, we will assume that the noise at different frequencies is uncorrelated, *i.e.* the probability distributions $p(\tilde{n}_i)$ at different frequencies are independent of

each other. In that case, the probability distribution for noise realizations, denoted $p[n]$, is just the product of the probabilities at each frequency bin:

$$p[n] = p(\tilde{n}_0, \tilde{n}_1, \dots, \tilde{n}_N) = \prod_{i=0}^N p(\tilde{n}_i), \quad (8.24)$$

so that for stationary, Gaussian noise

$$p[n] = \mathcal{N} e^{-\frac{1}{2} \sum_{i=0}^N \frac{|\tilde{n}_i|^2}{\sigma_i^2}}, \quad (8.25)$$

with \mathcal{N} some normalization factor.

For our purposes it will be convenient to approximate the discrete series $\{\tilde{n}_i\}$ by a continuous function of frequency, $\tilde{n}(f)$. We will assume that there is no correlation between noise at different frequencies, so that we can write

$$\langle \tilde{n}^*(f) \tilde{n}(f') \rangle = \frac{1}{2} \delta(f - f') S_n(f). \quad (8.26)$$

The above equation defines the function $S_n(f)$, which is called the *noise power spectral density*. Since $n(t)$ is real, $\tilde{n}(-f) = \tilde{n}^*(f)$, so that $S_n(-f) = S_n(f)$.

Note that for $f = f'$, the right hand side of Eq. (8.26) diverges. In a real experiment of finite duration T , one can take the delta function to mean

$$\delta(f = 0) \longrightarrow \left[\int_{-T/2}^{T/2} dt e^{2\pi i f t} \right]_{f=0} = T, \quad (8.27)$$

in which case Eq. (8.26) with $f = f'$ becomes

$$\langle |\tilde{n}(f)|^2 \rangle = \frac{1}{2} S_n(f) T. \quad (8.28)$$

Recalling that $\Delta f = 1/T$, and writing $2\sigma^2(f) = \langle |\tilde{n}(f)|^2 \rangle$ for the variance of the noise at frequency f , we find

$$2\sigma^2(f) \Delta f = \frac{1}{2} S_n(f). \quad (8.29)$$

This allows us to take the continuum limit of the right hand side of Eq. (8.25):

$$\begin{aligned} p[n] &= \mathcal{N} e^{-\frac{1}{2} \sum_{i=0}^N \frac{|\tilde{n}_i|^2}{\sigma_i^2}} \\ &= \mathcal{N} e^{-\frac{1}{2} \sum_{i=0}^N \frac{|\tilde{n}_i|^2}{\sigma_i^2 \Delta f} \Delta f} \\ &\rightarrow \mathcal{N} e^{-2 \int_0^\infty \frac{|\tilde{n}(f)|^2}{S_n(f)} df}. \end{aligned} \quad (8.30)$$

Finally, it is convenient to define the following inner product between arbitrary real functions $a(t)$ and $b(t)$:

$$\begin{aligned}\langle a|b \rangle &\equiv \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df \\ &= 4\Re \int_0^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df.\end{aligned}\quad (8.31)$$

In terms of $\langle \cdot | \cdot \rangle$, Eq. (8.30) can be written compactly as

$$p[n] = \mathcal{N} e^{-\frac{1}{2}\langle n|n \rangle}. \quad (8.32)$$

D. The optimal filter

In Eq. (8.14) we were integrating the expected signal $h(t)$ against the detector output $s(t)$, but this is not the optimal thing to do, since it does not make use of what we know about the detector noise. Let us instead integrate against an a priori unknown filter function $K(t)$:

$$\hat{s} = \int_{-\infty}^{\infty} dt s(t) K(t). \quad (8.33)$$

Define S to be the expected value of \hat{s} when a signal $h(t)$ is present, and let N be the root-mean-square value of \hat{s} when no signal is present:

$$S = \langle \hat{s} \rangle_h, \quad N = \langle \hat{s}^2 \rangle_{h=0}^{1/2}. \quad (8.34)$$

Now consider the ratio¹⁵

$$S/N = \frac{\langle \hat{s} \rangle_h}{\langle \hat{s}^2 \rangle_{h=0}^{1/2}}. \quad (8.35)$$

We can now ask the question: what filter function $K(t)$ maximizes S/N ? In other words, what choice of $K(t)$ will tend to maximize the filtered data \hat{s} when a signal is present (the numerator of Eq. (8.35)) but minimize it when there is only noise? Since we know the noise properties in the frequency domain (see Eq. (8.30)), it is convenient to write S and N in

¹⁵ Note that we could not simply have put $\langle \hat{s} \rangle_{h=0}$ in the denominator, because $\langle n(t) \rangle = 0$ implies $\langle \hat{s} \rangle_{h=0} = 0$. Instead, N is taken to be the square root of the variance of the noise, which is an indicator of how much the noise typically strays away from zero.

terms of the Fourier transforms of the relevant quantities:

$$\begin{aligned} S &= \int_{-\infty}^{\infty} dt \langle n(t) + h(t) \rangle K(t) \\ &= \int_{-\infty}^{\infty} dt h(t) K(t) \\ &= \int_{-\infty}^{\infty} df \tilde{h}^*(f) \tilde{K}(f), \end{aligned} \quad (8.36)$$

where in the second line we have used that $\langle n(t) \rangle = 0$ and $\langle h(t) \rangle = h(t)$, since the latter is independent of the noise, and in the last line we used Parseval's theorem. Similarly,

$$\begin{aligned} N &= \left[\left\langle \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' n(t) n(t') K(t) K(t') \right\rangle \right]^{1/2} \\ &= \left[\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' \langle n(t) n(t') \rangle K(t) K(t') \right]^{1/2} \\ &= \left[\int_{-\infty}^{\infty} df \frac{1}{2} S_n(f) |\tilde{K}(f)|^2 \right]^{1/2}, \end{aligned} \quad (8.37)$$

where in the last line we have used Eq. (8.26). Hence we arrive at

$$\begin{aligned} S/N &= \frac{\int_{-\infty}^{\infty} df \tilde{h}^*(f) \tilde{K}(f)}{\left[\int_{-\infty}^{\infty} df \frac{1}{2} S_n(f) |\tilde{K}(f)|^2 \right]^{1/2}} \\ &= \frac{2 \int_{-\infty}^{\infty} df \frac{\tilde{h}^*(f) \mathbf{K}(f)}{\frac{1}{2} S_n(f)}}{\left[2 \int_{-\infty}^{\infty} df \frac{|\mathbf{K}(f)|^2}{\frac{1}{2} S_n(f)} \right]^{1/2}}, \end{aligned} \quad (8.38)$$

where in the second line we introduced $\mathbf{K}(f) \equiv \frac{1}{2} S_n(f) \tilde{K}(f)$. The above can be written more compactly by making use of the inner product defined in Eq. (8.31):

$$S/N = \frac{\langle h | \mathbf{K} \rangle}{\langle \mathbf{K} | \mathbf{K} \rangle^{1/2}}, \quad (8.39)$$

or

$$S/N = \langle h | \hat{\mathbf{K}} \rangle. \quad (8.40)$$

Viewing h and \tilde{K} to be elements of a vector space of functions, we have found that the quantity S/N is the inner product of a given vector h and a yet-to-be-determined unit vector $\hat{\mathbf{K}}$. Clearly, S/N will be maximal when the unit vector points in the direction of h :

$$\hat{\mathbf{K}} \propto h, \quad (8.41)$$

so that also

$$\mathbf{K} \propto h. \quad (8.42)$$

This finally leads us to our desired optimal filter function:

$$\tilde{K}(f) \propto \frac{\tilde{h}(f)}{S_n(f)}.$$

(8.43)

The optimized S/N takes the form

$$\begin{aligned} S/N &= \frac{\langle h | \mathbf{K} \rangle}{\sqrt{\langle \mathbf{K} | \mathbf{K} \rangle}} \\ &= \langle h | \hat{h} \rangle, \end{aligned} \quad (8.44)$$

where $\hat{h} = h(t)/\sqrt{\langle h | h \rangle}$.

E. The detection statistic

In the previous section we motivated a choice of optimal filter by maximizing the ratio S/N , where S and N involved *expectation values* of integrals of the data $s(t)$ against the filter. We now present an alternative derivation of the optimal filter that does not deal with expectation values, and leads to a *detection statistic*, a quantity analogous to S/N but which takes as input the detector data $s(t)$ themselves. It is based on what is called *hypothesis ranking*.

Suppose as before that the interferometer data $s(t)$ consist of noise $n(t)$, and possibly a signal of known form, $h(t)$. Then we want to distinguish between two hypotheses:

1. The *null hypothesis* \mathcal{H}_0 which says that there is no signal: $s(t) = n(t)$;
2. The hypothesis \mathcal{H}_1 that there is a signal in the data: $s(t) = n(t) + h(t)$.

This is done by comparing $P(\mathcal{H}_1|s)$ with $P(\mathcal{H}_0|s)$, where $P(\mathcal{H}_1|s)$ is the probability that \mathcal{H}_1 is true *given the data* s , and similarly for $P(\mathcal{H}_0|s)$.

To compute the odds ratio we can make use of *Bayes' theorem*. Let A and B be two statements. For example, A could be the statement that the data take a particular form $s(t)$, and B the statement that the hypothesis \mathcal{H}_1 is true, but at this stage, A and B are arbitrary. First, define the *conditional probability* $P(A|B)$ as

$$P(A|B) \equiv \frac{P(A \wedge B)}{P(B)}, \quad (8.45)$$

where $P(A \wedge B)$ is the probability that A and B are both true, $P(B)$ is the probability of B by itself, and $P(A|B)$ is to be read as “the probability that A is true given B ”. Similarly,

$$P(B|A) = \frac{P(B \wedge A)}{P(A)}. \quad (8.46)$$

Since $P(A \wedge B) = P(B \wedge A)$, the above expressions straightforwardly lead to Bayes’ theorem:

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}. \quad (8.47)$$

Specializing to the case at hand, one has

$$\begin{aligned} P(\mathcal{H}_1|s) &= \frac{P(s|\mathcal{H}_1) P(\mathcal{H}_1)}{P(s)}, \\ P(\mathcal{H}_0|s) &= \frac{P(s|\mathcal{H}_0) P(\mathcal{H}_0)}{P(s)}. \end{aligned} \quad (8.48)$$

Now define the *likelihood ratio* as

$$\Lambda \equiv \frac{P(\mathcal{H}_1|s)}{P(\mathcal{H}_0|s)}. \quad (8.49)$$

If $\Lambda > 1$ then the signal hypothesis \mathcal{H}_1 is preferred, while if $\Lambda < 1$, the null hypothesis \mathcal{H}_0 is more likely. Using the Eqs. (8.48),

$$\Lambda = \frac{P(s|\mathcal{H}_1) P(\mathcal{H}_1)}{P(s|\mathcal{H}_0) P(\mathcal{H}_0)}, \quad (8.50)$$

where the common factor $P(s)$ has dropped out. The quantities $P(\mathcal{H}_1)$ and $P(\mathcal{H}_0)$ are called *prior probabilities*: They are the probabilities that respectively the hypotheses \mathcal{H}_1 and \mathcal{H}_0 are true, prior to any measurement. Often there will be no preference for the one hypothesis over the other prior to the measurement, in which case we set $P(\mathcal{H}_1) = P(\mathcal{H}_0)$, so that

$$\Lambda = \frac{P(s|\mathcal{H}_1)}{P(s|\mathcal{H}_0)}. \quad (8.51)$$

Now, we know what $P(s|\mathcal{H}_0)$ is: if the interferometer data consist of nothing but noise, then $s = n$, and $P(s|\mathcal{H}_0) = p[n]$, for which we derived the expression in Eq. (8.32). This also suggests a form for $P(s|\mathcal{H}_1)$: since in this case $s = n + h$, or $n = s - h$, we may write $P(s|\mathcal{H}_1) = p[s - h]$. Therefore,

$$\Lambda = \frac{e^{-\frac{1}{2}\langle s-h|s-h\rangle}}{e^{-\langle s|s\rangle}} = e^{\langle s|h\rangle} e^{-\frac{1}{2}\langle h|h\rangle}. \quad (8.52)$$

Though it may not be apparent from Eq. (8.52), Λ will often take on very large or very small values, so that it is convenient to work with its logarithm:

$$\log \Lambda = \langle s|h\rangle - \frac{1}{2}\langle h|h\rangle. \quad (8.53)$$

We will not know beforehand at what time a signal will be present in the data. Let $h(t) = Ag(t - t_0)$ be a signal with arrival time t_0 , and overall amplitude A ; specifically, $A = \sqrt{\langle h|h \rangle}$. The Fourier transform of $h(t)$ is

$$\begin{aligned}\tilde{h}(f) &= \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt \\ &= A \int_{-\infty}^{\infty} g(t - t_0) e^{2\pi i f t} dt \\ &= A \int_{-\infty}^{\infty} g(t') e^{2\pi i f (t' + t_0)} dt',\end{aligned}\tag{8.54}$$

where in the last line we changed the integration variable to $t' = t - t_0$. Hence

$$\begin{aligned}\tilde{h}(f) &= A e^{2\pi i f t_0} \int_{-\infty}^{\infty} g(t') e^{2\pi i f t'} dt' \\ &= A e^{2\pi i f t_0} \tilde{g}(f),\end{aligned}\tag{8.55}$$

where $\tilde{g}(f)$ is the Fourier transform of the function \tilde{g} . Note that, because of the way we defined A ,

$$\langle g(t - t_0) | g(t - t_0) \rangle = 1.\tag{8.56}$$

Now, we don't know beforehand what the amplitude A of the signal will be, but we can find out which value of A maximizes the likelihood ratio Λ , or equivalently, $\log \Lambda$ (since the logarithm is a monotonically increasing function). This value A_{\max} will then be the most likely value of the signal's amplitude. First note that

$$\begin{aligned}\log \Lambda &= A \langle s | g(t - t_0) \rangle - \frac{1}{2} A^2 \langle g(t - t_0) | g(t - t_0) \rangle \\ &= A \langle s | g(t - t_0) \rangle - \frac{1}{2} A^2.\end{aligned}\tag{8.57}$$

The maximum over A occurs where the derivative of $\log \Lambda$ with respect to A equals zero. Clearly $d \log \Lambda / dA = 0$ occurs when $A = A_{\max}$, with

$$A_{\max} = \langle s | g(t - t_0) \rangle.\tag{8.58}$$

Substituting this into Eq. (8.57), we find

$$\log \Lambda|_{A_{\max}} = \frac{1}{2} \langle s | g(t - t_0) \rangle^2.\tag{8.59}$$

Conveniently, the right hand side is independent of A . This motivates the definition of the *detection statistic* (also called the *signal-to-noise ratio*)

$$\rho(t_0) \equiv \langle s | g(t - t_0) \rangle,\tag{8.60}$$

which omits the unimportant square and factor of $1/2$. Indeed, the detection statistic is just meant to be a quantity that becomes large when a signal is strongly present – as does the log likelihood ratio. Again defining

$$\hat{h}(t) \equiv \frac{h(t)}{\sqrt{\langle h|h \rangle}}, \quad (8.61)$$

one has

$$\rho(t_0) = \langle s|\hat{h} \rangle. \quad (8.62)$$

Note that if in the above, the detector output s is replaced by its expectation value in the presence of a signal, $\langle s \rangle_h = \langle n + h \rangle = h$, or equivalently if the noise is set to zero, we obtain the maximized S/N of Eq. (8.44). This quantity is called the *optimal signal-to-noise ratio*, ρ_{opt} :

$$\begin{aligned} \rho_{\text{opt}} &= \frac{\langle h|h \rangle}{\sqrt{\langle h|h \rangle}} \\ &= \sqrt{\langle h|h \rangle}. \end{aligned} \quad (8.63)$$

Thus, the optimal signal-to-noise ratio is the value the signal-to-noise ratio would take if the detector output consisted only of a signal, without any noise. Though a somewhat abstract quantity, it can be useful to make rough estimates regarding the performance of a detector. However, for now the more useful quantity is the signal-to-noise ratio itself, since it can be used to search for gravitational waves.

Thus far we have been assuming that the particular shape of a possible signal in the data will be known beforehand, but this is not the case. Indeed, in the example of binary inspiral, we have seen that this shape is determined by the component masses (m_1, m_2) .¹⁶ Hence we will have to evaluate ρ for a large number of mass values, and see if there is some instance in time *and* particular (m_1, m_2) where the detection statistic peaks. Fig. 26 shows an example of a *template bank*: a large number of (m_1, m_2) values for which trial waveforms $h(m_1, m_2; t)$ will be filtered against the data. Note how templates are sprinkled far more densely at low masses than at high masses. This is because of the fact that low-mass signals are in band for a longer time than high-mass ones, so that they have more waveform cycles. For the same difference in masses, two low-mass waveforms will eventually get significantly out of

¹⁶ There are also other parameters that determine $h(t)$, such as distance, sky position, and inclination angle, but these only determine the *amplitude* of $h(t)$ and drop out when normalizing to $\hat{h}(t)$.

phase, and therefore have a large mismatch, but not two very high-mass waveforms. Hence, the (m_1, m_2) need to be packed more densely at low masses. The packing is done such that for neighboring points (m_1, m_2) and $(m_1 + \delta m_1, m_2 + \delta m_2)$, the difference between the corresponding waveforms is below some pre-set value ϵ in terms of the norm $\|a\| \equiv \sqrt{\langle a|a \rangle}$:

$$\|\hat{h}(m_1 + \delta m_1, m_2 + \delta m_2) - \hat{h}(m_1, m_2)\| < \epsilon. \quad (8.64)$$

In the example of Fig. 26, a maximum mismatch of $\epsilon = 0.03$ was imposed.

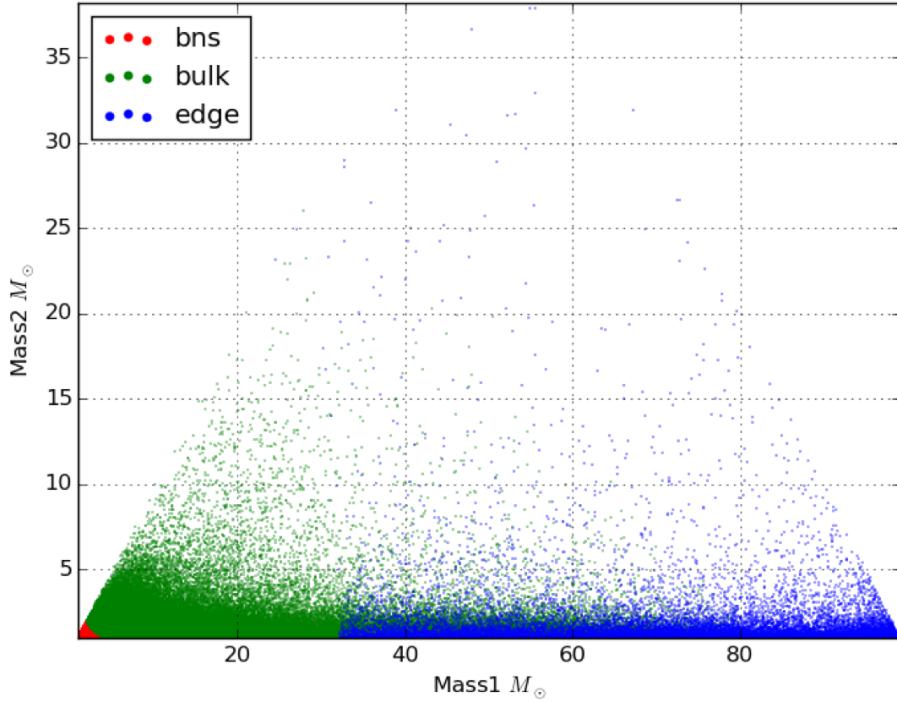


Figure 26: A typical template bank for the detection of signals from coalescing binary neutron stars and/or black holes.

Apart from being noisy, interferometric detectors also suffer from what are called *glitches*: temporary small malfunctions in some part of the instrument that can mimick gravitational waves. Since these are typically confined to a single detector in the network that currently consists of the two LIGOs and Virgo, their effect can to some extent be mitigated by only accepting candidate gravitational wave events for which the *same template* (*i.e.* the same pair (m_1, m_2)) in the template bank gave the highest value of ρ in three detectors at almost the same time (allowing for differences in arrival times due to their geographic separation); these are called *triggers*. Typically the detection statistics will still take on different values

ρ_1, ρ_2, ρ_3 in three interferometers because of the detectors having different sensitivities, and different orientations so that one may be better positioned than another for a gravitational wave coming from a particular position on the sky. What one does is to combine the detection statistics by adding them in quadrature: $\rho_{\text{comb}} = \sqrt{\rho_1^2 + \rho_2^2 + \rho_3^2}$.

However, it happens frequently that glitches in multiple detectors occur at nearly the same time and cause the same template in the template bank to give the highest value of the detection statistic. How then are we going to confidently detect gravitational waves? The answer is to compute a *probability distribution* $p(\rho_{\text{comb}})$ for glitches to cause the same template in the bank to give the highest ρ at almost the same time in three detectors. This is done as follows. For simplicity, first consider just two detectors, with outputs $s_1(t)$ and $s_2(t)$ that have been stored on a computer. Then one can “time-slide” one of the data streams relative to the other:

$$s'_2(t) \equiv s_2(t - \Delta t_{\text{shift}}), \quad (8.65)$$

where Δt_{shift} is taken to be larger than the light travel time between the detectors (which is ~ 10 ms between the two LIGOs, and ~ 27 ms between either of the LIGO detectors and Virgo). Next one looks for events for which in $s_1(t)$ and $s'_2(t)$ the same template gave the highest value of the detection statistic at nearly the same time t . Because of the artificial time shift Δt_{shift} inherent in $s'_2(t)$ being larger than the time it takes for a gravitational wave to get from detector 1 to detector 2, such events can not be of astrophysical origin; instead they must be due to glitches or other noise artifacts having occurred in detector 1 near time t and in detector 2 near time $t - \Delta t_{\text{shift}}$. Nevertheless, one can record the combined detection statistic $\rho_{\text{comb}} = \sqrt{\rho_1^2 + \rho_2^2}$ for each such event, and put these in a histogram as in Fig. 27. Normalizing and smoothening this histogram yields an estimate of the probability distribution that ρ_{comb} will follow in the case of triggers that are purely due to detector noise. This is called the *background distribution*.

In the above example we restricted ourselves to a pair of detectors, but of course the output of a third detector can also be included. Two of the three outputs can be time-slid relative to the first one by different times $\Delta t_{\text{shift},2}$ and $\Delta t_{\text{shift},3}$, always making sure that the time-sliding exceeds the light travel time between any pair of detectors. Triggers which are consistent between the resulting data streams can not be of astrophysical origin and must be due to noise artefacts. The combined detection statistics $\rho_{\text{comb}} = \sqrt{\rho_1^2 + \rho_2^2 + \rho_3^2}$ for such

events can be put in a histogram, which leads to a background distribution as in Fig. 27.

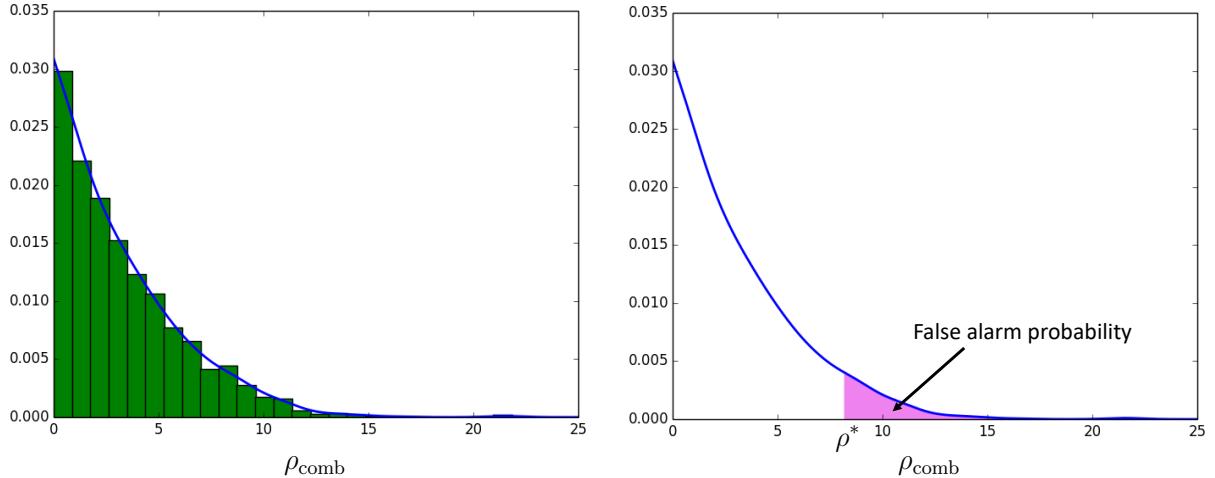


Figure 27: Left panel: An estimate of the probability distribution for the combined detection statistic ρ_{comb} due to noise artifacts in the interferometers, obtained by creating a normalized histogram (green) and approximating it by a smooth curve (blue). Right panel: To an event with $\rho_{\text{comb}} = \rho^*$, one can assign a *false alarm probability* (violet), which is the probability of obtaining a detection statistic at least this large due to random noise fluctuations.

Let us return to the case where no time sliding has been performed. Let there be a trigger with combined detection statistic $\rho_{\text{comb}} = \rho^*$, in this case possibly of astrophysical origin. Then one can compute the area underneath the background distribution corresponding to $\rho_{\text{comb}} \geq \rho^*$. Since the background distribution is a probability distribution, this area gives us the probability of obtaining a trigger with ρ_{comb} at least as large as ρ^* by coincidence, as a result of mere noise fluctuations. This is called the *false alarm probability* for the candidate detection. Finally, a false alarm probability together with the duration of an observing period T_{data} can be converted to a *false alarm rate*, usually expressed in yr^{-1} , which is the rate at which a given event will occur through random noise fluctuations.

Since the template waveforms \hat{h} are normalized, any information that resides in the amplitude of a waveform $h(t)$ gets erased. This includes distance, inclination angle, and *sky position*. However, it would be of great use if astronomers could be alerted not only of the likely detection of a gravitational wave, but also where in the sky it originated, since some sources of gravitational waves (such as mergers of binary neutron stars) also emit light. Localizing the source on the sky can be done due to the fact that a gravitational wave does

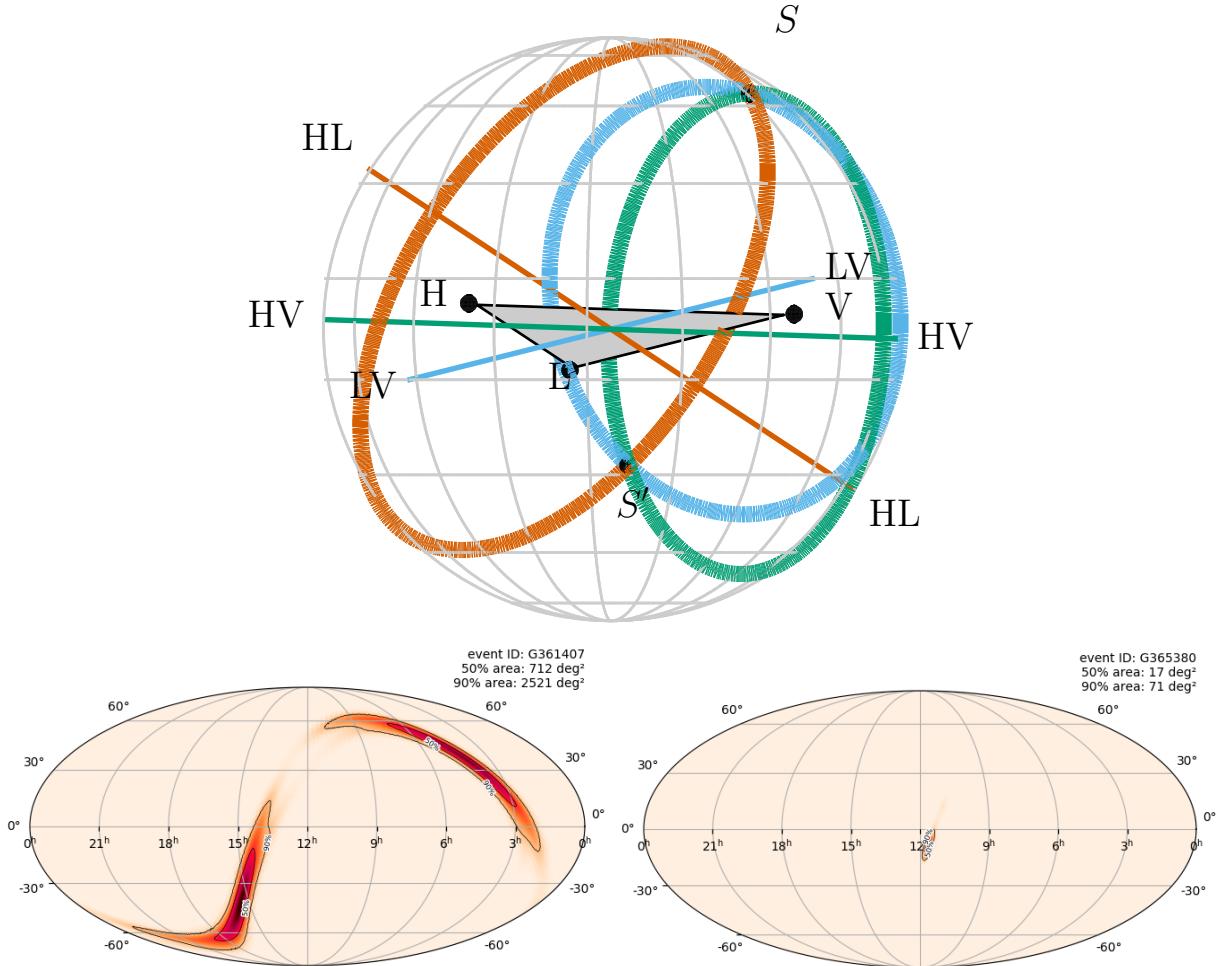


Figure 28: Top: Sky localization of a gravitational wave event with a network of detectors consisting of LIGO Hanford (H), LIGO Livingston (L), and Virgo (V), based on the differences in arrival time of a signal at the different detectors. When only two detectors are operational, the source is localized within a circular strip on the sky. With three detectors, this can be narrowed down to two distinct patches labeled S and S' . Bottom: In reality one additionally uses the differences in orientations and hence responses of the interferometers to a signal. In the two-detector case, this allows localization within large arcs on the sky (left), and in the three-detector case, a single sky patch (right).

not arrive at all the detectors at the same time. This is illustrated in Fig. 28. In the case of two interferometers being operational during the detection, the different arrival times, including measurement uncertainties, enable localization of the source in a circular strip on the celestial sphere. With three detectors working (the two LIGO interferometers, and

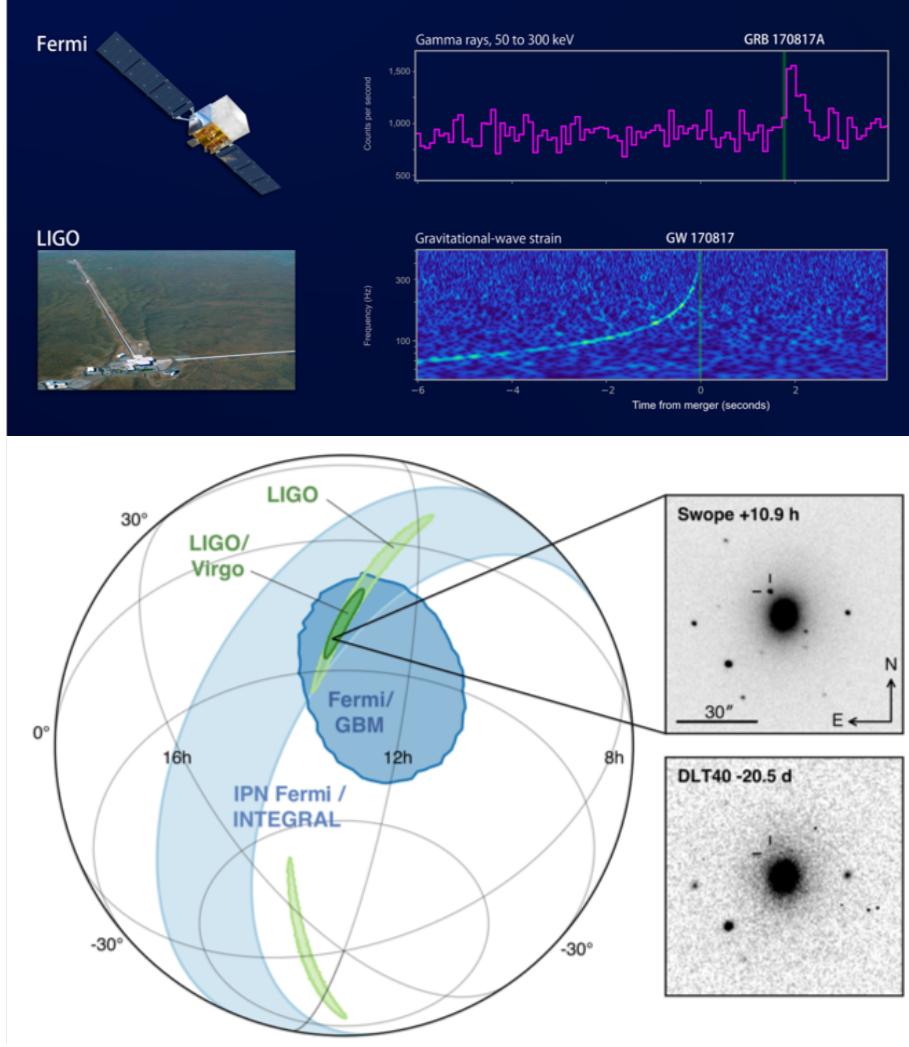


Figure 29: Top: the binary neutron star signal GW170817 was followed by a gamma ray burst (upper panel) detected by the Fermi and ANTARES satellites. The lower panel shows a diagram of frequency versus time for the LIGO-Virgo gravitational wave data; the bright curve closely follows the expression $f_{\text{gw}}(t)$ of Eq. (7.24) for a chirp mass of $\mathcal{M}_c = 1.19 M_\odot$, consistent with a binary neutron star! Bottom: With only the two LIGO interferometers, the sky localization would have been inside the two light green arcs. With the Virgo detector added, this shrinks to the dark green patch. Inside the patch the afterglow of the binary neutron star collision was discovered by optical astronomers: a new object had appeared near the galaxy NGC 4993 (boxes to the right; see the crosshairs).

Virgo), the intersections of the circular strips for any pair of detectors enable localization at one of two possible patches in the sky. In reality one also folds in the different responses of the detectors due to their different orientations together with the “loudness” of the signal in the interferometers. Indeed, if a signal has a larger detection statistic in one detector than in the others, then this is likely due to the fact that this interferometer is more conveniently oriented with respect to the sky location of the source. In the two-detector case this turns circular strips into arcs, and in the tree-detector case allows one to select a single sky patch.

Since September 2015 there have been 90 confident detections of gravitational waves, each designated as “GW” followed by the date of discovery in the format “yymmdd”:

- Most signals were from binary black holes, the first one being GW150914.
- A signal from a binary neutron star: GW170817.
- Possibly another binary neutron star: GW190425.
- Two neutron star-black hole signals, GW200105 and GW200115.

Thus, the first detection happened on 14th September 2015. GW150914 is still the “loudest” binary black hole signal recorded so far, with $\rho_{\text{comb}} = 23.6$, a false alarm probability $< 2 \times 10^{-7}$, and a false alarm rate better than 1 in 2×10^5 years. Of all the gravitational wave detections above, the loudest overall was the binary neutron star signal GW170817, with $\rho_{\text{comb}} = 32.4$; its false alarm rate was better than 1 in 8×10^4 years.

As shown in Fig. 29, GW170817 was not only observed in gravitational waves (whose frequency evolution was indeed consistent with masses in the neutron star range), but also as a gamma ray burst. Thanks to the fact that three interferometers (the two LIGOs and Virgo) were operational, the location of the source could be determined to be in a sky patch of $\sim 30 \text{ deg}^2$. This patch contained about 50 potential host galaxies, for each of which optical astronomers took pictures which they compared with archival photographs. Less than 11 hours after the gravitational wave detection, the true host galaxy, NGC 4993, was identified in this way: close to it a new light source had appeared which had not been there before. This turned out to be the *afterglow* of the binary neutron star collision, emitted by a rapidly expanding cloud of material. Indeed, the electromagnetic spectrum was monitored over the course of weeks, in wavelengths ranging from ultraviolet to infrared, and was consistent

with an abundance of heavy elements, suggestive of having originated in a neutron-rich environment.¹⁷

¹⁷ Note that heavy elements need a large number of neutrons per proton for their nuclei to stay together, so that the collision of two neutron stars is a natural way to create them. Some fraction of the neutrons from the stars will decay to protons (plus electrons and neutrinos), enabling elements to come about, but a neutron-rich environment will be in place, which is conducive to the formation of *heavy* elements in particular. It is now understood that for example most of the gold, platinum, and uranium in the Universe came about in this way.

IX. THE INSPIRAL OF BINARY NEUTRON STARS AND BLACK HOLES: THE DETECTOR POINT OF VIEW

As we shall in subsequent chapters, in a short amount of time the direct detection of gravitational waves has proven to be a game-changer in fundamental physics, astrophysics, and cosmology, with more to come in the near future. However, the amount of information that can be accessed through gravitational waves is to some extent limited by our detectors. For example, how sensitive the detectors are determines how far into the universe coalescing binaries can be observed, and, given the finite number of compact binary mergers occurring in the Universe per unit time and per unit volume, how many detections will be made during the time the detectors are operational. Additionally, the sensitivity of the detectors will determine how much information about the source of a gravitational wave can be extracted from the signal that is observed. These are things we aim to get a sense of in this chapter.

A. The detector response in the stationary phase approximation

As discussed in Chapter VIII, the response of an interferometer to a gravitational wave with polarizations $h_+(t)$, $h_\times(t)$ takes the form

$$h(t) = F_+(\theta, \phi, \psi) h_+(t) + F_\times(\theta, \phi, \psi) h_\times(t), \quad (9.1)$$

where the beam pattern functions F_+ , F_\times , given by Eqs. (8.12), depend on the sky location (θ, ϕ) and the polarization angle ψ . For the particular case of the inspiral of compact objects, we obtained approximate expressions for $h_+(t)$ and $h_\times(t)$ in Chapter VII, which for convenience we repeat here:

$$\begin{aligned} h_+(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \frac{1 + \cos^2(\iota)}{2} \cos(\Phi_{\text{gw}}(t_{\text{ret}})), \\ h_\times(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{2/3} \cos(\iota) \sin(\Phi_{\text{gw}}(t_{\text{ret}})), \end{aligned} \quad (9.2)$$

with ι the inclination of the orbital plane with respect to the line of sight, r the distance from source to observer, and $\mathcal{M}_c = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ the chirp mass. The instantaneous gravitational wave frequency is given by

$$f_{\text{gw}}(t) = \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-5/8} \left(\frac{5}{256} \frac{1}{t_c - t} \right)^{3/8}, \quad (9.3)$$

and the phase takes the form

$$\Phi_{\text{gw}}(t) = -2 \left(\frac{5G\mathcal{M}_c}{c^3} \right)^{-5/8} (t_c - t)^{5/8} + \Phi_c, \quad (9.4)$$

where t_c is the time of coalescence, and Φ_c the phase at $t = t_c$.

Eqs. (9.1)-(9.4) give us the response of a detector to gravitational waves from coalescing compact binaries, as a function of time. However, in practice, much of data analysis is performed in the Fourier domain. As we saw in Chapter VIII, given a template waveform $h(t)$, the signal-to-noise ratio is given by

$$\rho = \langle s | \hat{h} \rangle, \quad (9.5)$$

where $s(t)$ is the detector output (the detector noise, plus a possible signal), and

$$\hat{h}(t) = \frac{h(t)}{\sqrt{\langle h | h \rangle}}. \quad (9.6)$$

However, the inner product $\langle \cdot | \cdot \rangle$ is defined as (see Eq. (8.31))

$$\langle a | b \rangle = 4\Re \int_0^\infty \frac{\tilde{g}^*(f) \tilde{h}(f)}{S_n(f)} df, \quad (9.7)$$

where $S_n(f)$ is the power spectral density of the noise (essentially the variance of the noise as a function of frequency), and $\tilde{a}(f)$, $\tilde{b}(f)$ are the Fourier transforms of $a(t)$, $b(t)$, respectively. Hence, to gain insight into how the properties of a detector affect the detectability of signals as well as our ability to extract information from them, it will be convenient to have an expression for the Fourier transform of the detector response in Eq. (9.1):

$$\tilde{h}(f) = \int_{-\infty}^\infty h(t) e^{2\pi i f t} dt. \quad (9.8)$$

Unfortunately, due to the non-trivial time dependence of $f_{\text{gw}}(t)$ and $\Phi_{\text{gw}}(t)$, it is not possible to obtain an exact analytic expression for $\tilde{h}(f)$. However, one can arrive at an approximation, which for our purposes will amply suffice.

To derive an approximate expression for $\tilde{h}(f)$, first write

$$\begin{aligned} h_+(t) &= A(t) (1 + \cos^2(\iota)) \cos(\Phi_{\text{gw}}(t)), \\ h_\times(t) &= A(t) 2 \cos(\iota) \sin(\Phi_{\text{gw}}(t)), \end{aligned} \quad (9.9)$$

with

$$A(t) = -\frac{2}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}(t)}{c} \right)^{2/3}. \quad (9.10)$$

Next, by using simple trigonometry, the detector response can be written as

$$h(t) = A(t) \sqrt{F_+^2(1 + \cos^2(\iota))^2 + F_x^2 4 \cos^2(\iota)} \cos(\Phi_{\text{gw}}(t) + \varphi_0), \quad (9.11)$$

where

$$\varphi_0 = \arctan \left(\frac{-F_x 2 \cos(\iota)}{F_+(1 + \cos^2(\iota))} \right). \quad (9.12)$$

Focusing on the time dependent part of $h(t)$ that enters the right hand side of Eq. (9.8), the integral we wish to obtain an approximation for is

$$\int dt A(t) \cos(\Phi_{\text{gw}}(t) + \varphi_0) e^{2\pi i f t} = \frac{1}{2} \int dt A(t) (e^{i(\Phi_{\text{gw}}(t) + \varphi_0)} + e^{-i(\Phi_{\text{gw}}(t) + \varphi_0)}) e^{2\pi i f t}, \quad (9.13)$$

where we have used that

$$\cos(x) = \frac{1}{2} (e^{ix} + e^{-ix}). \quad (9.14)$$

Consider the first term in the right hand side of (9.13):

$$\frac{1}{2} e^{i\varphi_0} \int dt A(t) e^{i(2\pi f t + \Phi_{\text{gw}}(t))}. \quad (9.15)$$

Here $\exp[i(2\pi f t + \Phi_{\text{gw}}(t))]$ can be viewed as a unit vector in the complex plane which keeps rotating anti-clockwise as time increases, because $2\pi f t + \Phi_{\text{gw}}(t)$ is a monotonically increasing function of time. Since $\log A(t)$ varies slowly compared with $\Phi_{\text{gw}}(t)$, the contributions to the integral approximately average out over time. This is to be contrasted with the second term,

$$\frac{1}{2} e^{-i\varphi_0} \int dt A(t) e^{i(2\pi f t - \Phi_{\text{gw}}(t))}, \quad (9.16)$$

in which the unit vector $\exp[i(2\pi f t - \Phi_{\text{gw}}(t))]$ first rotates anti-clockwise, comes to a halt when the maximum of the function $2\pi f t - \Phi_{\text{gw}}(t)$ is reached, and then starts rotating clockwise as $2\pi f t - \Phi_{\text{gw}}(t)$ decreases again. The largest contribution to the integral will come from the vicinity of the *stationary point* t_s , which is the time where $2\pi f t - \Phi_{\text{gw}}(t)$ is at a maximum. Since this is the time at which the time derivative of the latter function is zero, one has

$$2\pi f = \dot{\Phi}_{\text{gw}}(t_s). \quad (9.17)$$

The exponent in (9.16) can be Taylor expanded around t_s as follows:

$$\begin{aligned} 2\pi f t - \Phi_{\text{gw}}(t) &= 2\pi f t_s - \Phi_{\text{gw}}(t_s) + 2\pi f(t - t_s) - \dot{\Phi}_{\text{gw}}(t_s)(t - t_s) - \frac{1}{2} \ddot{\Phi}_{\text{gw}}(t_s)(t - t_s)^2 + \dots \\ &= 2\pi f t_s - \Phi_{\text{gw}}(t_s) - \frac{1}{2} \ddot{\Phi}_{\text{gw}}(t_s)(t - t_s)^2 + \dots, \end{aligned} \quad (9.18)$$

where in the second line we have used Eq. (9.17). Defining a new integration variable

$$x = \sqrt{\frac{\ddot{\Phi}_{\text{gw}}(t_s)}{2}}(t - t_s) \quad (9.19)$$

and considering that $\log A(t)$ varies slowly, the expression in Eq. (9.16) approximates to

$$\frac{1}{2}A(t_s)e^{-i\varphi_0}e^{i(2\pi ft_s-\Phi_{\text{gw}}(t_s))}\left(\frac{2}{\ddot{\Phi}_{\text{gw}}(t_s)}\right)^{1/2}\int_{-\infty}^{\infty}dx e^{-ix^2}, \quad (9.20)$$

and one has

$$\int_{-\infty}^{\infty}dx e^{-ix^2} = \sqrt{\pi}e^{-i\pi/4}. \quad (9.21)$$

Finally, note that the defining equation for t_s , Eq. (9.17), is noting but the statement that for a given frequency f one has $t_s = t(f)$, where $t(f)$ is the inverse function of $f_{\text{gw}}(t)$. Returning to Eq. (9.13),

$$\int dt A(t) \cos(\Phi(t) + \varphi_0) e^{2\pi ift} \simeq \frac{\sqrt{\pi}}{2}A(t(f))e^{-i\varphi_0}\left(\frac{2}{\ddot{\Phi}_{\text{gw}}(t(f))}\right)^{1/2}e^{i\Psi(f)}, \quad (9.22)$$

where

$$\Psi(f) = 2\pi ft(f) - \Phi_{\text{gw}}(t(f)) - \frac{\pi}{4}. \quad (9.23)$$

Using that

$$t(f) = t_c - \frac{5}{256}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{-5/3}(\pi f)^{-8/3} \quad (9.24)$$

together with Eqs. (9.3), (9.4), (9.10), and recalling the angle-dependent prefactor in Eq. (9.11), we finally obtain our sought-after approximation for $\tilde{h}(f)$:

$$\begin{aligned} \tilde{h}(f) &\simeq \frac{c}{r}\sqrt{\frac{5\pi}{96}}\sqrt{F_+^2(1+\cos^2(\iota))^2+F_\times^24\cos^2(\iota)}\left(\frac{G\mathcal{M}_c}{c^3}\right)^{5/6}(\pi f)^{-7/6} \\ &\times \exp\left[i\left(2\pi ft_c - \Phi_c + \frac{3}{4}\left(\frac{8\pi G\mathcal{M}_c f}{c^3}\right)^{-5/3} - \varphi_0 - \frac{\pi}{4}\right)\right]. \end{aligned} \quad (9.25)$$

This expression is known as the *stationary phase approximation* to the Fourier transform of $h(t)$.

As a first application, let us find out to what distance we will be able to see particular kinds of sources of gravitational radiation given a particular detector. Recall that given a detector output $s(t)$, the signal-to-noise ratio is given by Eqs. (9.5) and (9.6). Assuming a template waveform $h(t)$ that exactly equals a signal in the data, and a sufficiently loud

signal so that $s(t) \simeq h(t)$, the expression for ρ leads to that of the optimal signal-to-noise ratio (SNR),

$$\rho_{\text{opt}} = \sqrt{\langle h|h \rangle}. \quad (9.26)$$

With the definition of the inner product, Eq. (9.7), one has

$$\rho_{\text{opt}} = \left[4 \int_{f_{\min}}^{f_{\max}} \frac{|\tilde{h}(f)|^2}{S_n(f)} \right]^{1/2}. \quad (9.27)$$

Here we restrict the integration range from some minimum frequency f_{\min} below which the detector has no sensitivity (in the case of LIGO or Virgo, $f_{\min} \simeq 20$ Hz), to a maximum frequency f_{\max} above which there is no signal power; for inspiral signals one can take this to be the ISCO frequency, $f_{\max} = c^3/(6^{3/2}\pi GM)$, with $M = m_1 + m_2$. Substituting our expression for $\tilde{h}(f)$ from Eq. (9.25),

$$\rho_{\text{opt}} = \left[\frac{c^2}{r^2} \frac{5\pi}{24} \pi^{-7/3} (F_+^2(1 + \cos^2(\iota))^2 + F_\times^2 4 \cos^2(\iota)) \mathcal{M}_c^{5/3} \int_{f_{\min}}^{f_{\max}} \frac{f^{-7/3}}{S_n(f)} df \right]^{1/2}. \quad (9.28)$$

Clearly the optimal SNR depends on the position in the sky relative to the detector (through the dependence of F_+ , F_\times on the sky position angles (θ, ϕ)) and on the inclination of the source with respect to the line of sight (given by ι). It is largest when the source is “face-on” ($\iota = 0$) or “face-off” ($\iota = \pi$), and when the source is located at the zenith ($\theta = 0$) or the nadir ($\theta = \pi$), in which case

$$\rho_{\text{opt,max}} = \left[\frac{c^2}{r^2} \frac{5\pi}{6} \pi^{-7/3} \mathcal{M}_c^{5/3} \int_{f_{\min}}^{f_{\max}} \frac{f^{-7/3}}{S_n(f)} df \right]^{1/2}. \quad (9.29)$$

For a source with a given chirp mass \mathcal{M}_c , it is useful to know the optimal SNR averaged over sky position and orientation. One has

$$\langle F_+^2(1 + \cos^2(\iota))^2 + F_\times^2 4 \cos^2(\iota) \rangle_{\theta, \phi, \iota, \psi} = \frac{16}{25}, \quad (9.30)$$

leading to

$$\rho_{\text{opt,ave}} = \frac{4}{5} \rho_{\text{opt,max}}. \quad (9.31)$$

Now let ρ_0 be the minimum SNR needed for confident detection of a source. Then by setting $\rho_{\text{opt,ave}} = \rho_0$ and solving for the distance r , one obtains the *angle-averaged horizon distance*:

$$d_{\text{hor,ave}} = \frac{4}{5} \left[c^2 \frac{5\pi}{6} \pi^{-7/3} \mathcal{M}_c^{5/3} \int_{f_{\min}}^{f_{\max}} \frac{f^{-7/3}}{S_n(f)} df \right]^{1/2} \rho_0^{-1}. \quad (9.32)$$

Usually one can assume $\rho_0 \simeq 8$. Note how the horizon distance depends both on the chirp mass \mathcal{M}_c and the total mass M , since the latter sets the upper limit of the integral over frequency through the expression for f_{ISCO} . In addition, it depends on the sensitivity of the detector: if the noise power spectral density $S_n(f)$ is lower then sources can be seen to higher distances. Fig. 30 shows the (square root of) the power spectral density, and the angle-averaged horizon distance for binary neutron stars during the latest observing run.

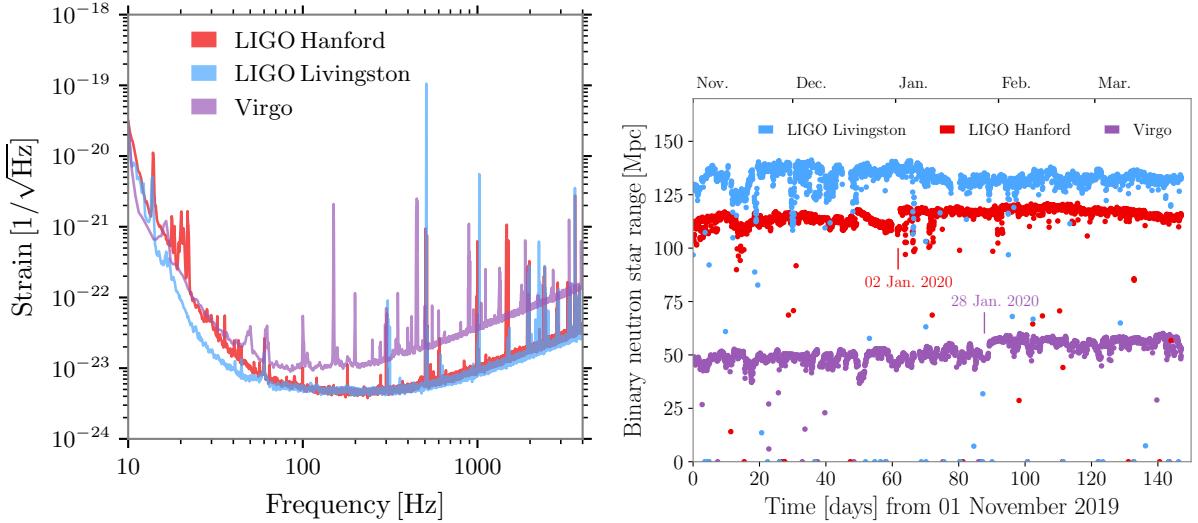


Figure 30: Left: Representative sensitivities of the two LIGO detectors and Virgo during their last observing run; here $\sqrt{S_n(f)}$ is shown as a function of frequency. Right: The angle-averaged horizon distance for a “canonical” binary neutron star inspiral with $m_1 = m_2 = 1.4 M_\odot$. Note how the detector sensitivities vary somewhat over time, mostly due to drifts in the performance of the various components comprising the interferometer. Two dates are highlighted: on 2 January 2020, LIGO Hanford underwent a minor upgrade, and similarly for Virgo on 28 January 2020.

B. What information about the source is contained in the signal?

Fig. 26 of the previous chapter illustrates how templates are laid out over parameter space (here consisting of m_1 and m_2) in such a way that the “distance” $\|\hat{h}_1 - \hat{h}_2\|$ between neighboring template waveforms $\hat{h}_1(t)$, $\hat{h}_2(t)$ does not exceed some pre-set value ϵ . Even for $\epsilon \ll 1$, in the high-mass region the templates can be placed relatively sparsely in terms of the parameters without much loss in detection efficiency. However, if a promising candidate event is found in that region (probably a binary black hole), the template bank will only offer

a very crude estimation of what the component masses (m_1, m_2) will have been. Moreover, matched filtering is done with normalized templates \hat{h} , which “erases” any information residing in the amplitude of the waveform; this includes the inclination ι of the orbital plane with respect to the line of sight, as well as the distance.¹⁸

After a detection has been made using template banks and other techniques that were introduced before, we will want to know with high accuracy what are the values of *all* the parameters that determine the observed waveform. Let us first see what these are in the case of binary black holes. Earlier we showed that the response of an interferometer to gravitational wave polarizations h_+ and h_\times is given by Eqs. (8.11), (8.12), where the beam pattern functions F_+, F_\times depend on the direction of propagation through the angles (θ, ϕ) , and on the polarization angle ψ . The direction of propagation can be related to the sky position, which is usually expressed in a coordinate system (α, δ) that is fixed to the stars rather than the (θ, ϕ) associated with a detector. Here α and δ are respectively *right ascension* and *declination*, whose definition can be found on *e.g.* Wikipedia.¹⁹ The polarization angle ψ is defined with respect to the reference frame of an interferometer as shown in Fig. 25, so that for different detectors in a network it takes on different values. However, given the value of ψ in the frame of some “preferred” detector (say, Virgo), its values for the other interferometers can be calculated from the way they are oriented with respect to the preferred one; hence there is only one independent ψ value.

For binary inspiral, in the quadrupole approximation the expressions for h_+ and h_\times are Eqs. (7.28), which together with Eqs. (7.24), (7.27), and (7.25) depend on

$$(\mathcal{M}_c, \iota, r, t_c, \Phi_c). \quad (9.33)$$

In reality there are additional parameters:

- In our derivation of h_+ and h_\times we neglected the effect of higher-order mass multipole moments, which is why we found expressions in which the component masses m_1, m_2 enter only through the combination $\mathcal{M}_c = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$. The higher-order

¹⁸ Sky position is also erased, but as explained above, this can be retrieved essentially by looking at the different arrival times of a gravitational wave at different detectors in a global network of interferometers.

¹⁹ The conversion between (θ, ϕ) and (α, δ) is not entirely trivial. Not only does it depend on the location of a detector on the Earth; because of the Earth’s rotation around its axis and around the Sun, it also depends on the arrival time of the gravitational wave. We will not go into the details here.

contributions break this degeneracy, and more accurate expressions for h_+ , h_\times do depend separately on m_1 and m_2 .

- A black hole can be spinning, with angular momentum vector \mathbf{J} . For a hole with mass m , one can define a dimensionless *spin vector* $\mathbf{S} = (c/Gm^2)\mathbf{J}$, which normally²⁰ satisfies $0 \leq |\mathbf{S}| < 1$. The spin vectors \mathbf{S}_1 , \mathbf{S}_2 in a binary black hole system will affect the inspiral; hence h_+ , h_\times depend on them. Each of the two vectors have 3 components, together introducing 6 free parameters.

Putting everything together, we find that the observed signal from a binary black hole coalescence depends on the following 15 parameters:

$$\boldsymbol{\theta} = (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \alpha, \delta, \psi, \iota, r, t_c, \Phi_c). \quad (9.34)$$

Methods based on template banks will not be adequate to arrive at high-quality measurements of these parameters. Indeed, suppose that we would define some generalization of the detection statistic ρ , now using non-normalized templates. Imagine we would lay out templates on a grid, for definiteness using 1000 grid points in each of the 15 parameter directions. Then we would need to evaluate this statistic $1000^{15} = 10^{45}$ times, which in practice is not computationally feasible. On the other hand, it could be that one is only interested in some subset of the parameters in (9.34); then one might think that it suffices to lay out a template bank for only those. However, here the problem is that any subset of (9.34) will have some amount of correlation with all the other parameters, and ignoring those correlations may lead to a biased measurement. Hence a different method is called for, which we now turn to.

In the language of the previous chapter, what we want to calculate is a probability density distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$, where:

1. \mathcal{H} is a hypothesis, *e.g.* the hypothesis that there is a binary black hole signal in the data, or a binary neutron star signal, or a gravitational wave signal from a supernova explosion.

²⁰ If $|\mathbf{S}| \geq 1$ then the black hole has no horizon, leading to a naked singularity. The *Cosmic Censorship Conjecture* states that given “reasonable” initial conditions, such objects cannot form, but this has not yet been proven mathematically.

2. d are the data, *i.e.* what we called $s(t)$ in the previous chapter.²¹

Thus, $p(\boldsymbol{\theta}|d, \mathcal{H})$ is the probability that the parameters take on particular values $\boldsymbol{\theta}$, *given* data d and a hypothesis \mathcal{H} regarding the nature of the signal. The reason why we call $p(\boldsymbol{\theta}|d, \mathcal{H})$ a probability *density* is that it will be a function which, when integrated over some volume \mathcal{V} in parameter space, tells us how much probability is contained in \mathcal{V} . Since the probability over *all* of parameter space is 1, the probability density must be normalized:

$$\int d^N\boldsymbol{\theta} p(\boldsymbol{\theta}|d, \mathcal{H}) = 1, \quad (9.35)$$

with N the number of parameters.

In the previous chapter we introduced *Bayes' theorem*, which says that for any two statements A and B ,

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)}. \quad (9.36)$$

This allows us to write

$$p(\boldsymbol{\theta}|d, \mathcal{H}) = \frac{p(d|\boldsymbol{\theta}, \mathcal{H}) p(\boldsymbol{\theta}|\mathcal{H})}{p(d|\mathcal{H})}. \quad (9.37)$$

Note how in this case, B is the statement “the parameters take on the values $\boldsymbol{\theta}$, and A is the statement “the interferometer data are $d(t)$ ”. In addition there is the hypothesis \mathcal{H} , but comparing Eqs. (9.36) and (9.37), we see that in the latter equation, it just “goes along for the ride”. Why it is nevertheless useful to have it appear explicitly in our equations will become apparent below. The following functions appear in Eq. (9.37):

- In the left hand side there is the probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$ which we wish to obtain. In the context of Bayesian statistics, it is called the *posterior probability distribution*.
- $p(d|\boldsymbol{\theta}, \mathcal{H})$ is called the *likelihood*. As we shall see below, this is something we can write an explicit expression for.
- $p(\boldsymbol{\theta}|\mathcal{H})$ is called the *prior probability distribution*. Since it doesn't involve the data d , it is the probability distribution for the parameters $\boldsymbol{\theta}$ before any measurement has occurred, but assuming the signal is of a particular kind (the hypothesis \mathcal{H}). This

²¹ We could have continued to denote the data by s , but the notation d is more in keeping with the literature on parameter estimation.

is a function we are largely free to choose ourselves. For example, if the signal is assumed to be from a binary neutron star inspiral, then it makes sense to let the prior distribution for the component masses (m_1, m_2) to be uniform on the intervals $m_i \in [1 M_\odot, 3 M_\odot]$, since on general astrophysical grounds it is highly unlikely that there exist neutron stars with masses outside this range.

- $p(d|\mathcal{H})$ is called the *evidence* for the hypothesis \mathcal{H} . Since it is independent of the parameters $\boldsymbol{\theta}$, in the context of parameter estimation $1/p(d|\mathcal{H})$ can be viewed as an unimportant overall prefactor, whose value will be set by the requirement that $p(d|\boldsymbol{\theta}, \mathcal{H})$ be normalized.

In summary, the left hand side of Eq. (9.37) is what we want to obtain; in the right hand side, the likelihood $p(d|\boldsymbol{\theta}, \mathcal{H})$ is claimed to be known; the prior probability distribution $p(\boldsymbol{\theta}|\mathcal{H})$ is a function that we may choose ourselves; and $p(d|\mathcal{H})$ is set by the requirement that the posterior probability distribution in the left hand side be normalized.

First, what is the function $p(d|\boldsymbol{\theta}, \mathcal{H})$? As we saw in the previous chapter, the data consists of noise and signal.²²

$$d(t) = n(t) + h(\boldsymbol{\theta}; t). \quad (9.38)$$

As far as the noise is concerned, we have an expression for its probability distribution:

$$p[n] = \mathcal{N} e^{-\frac{1}{2}\langle n|n \rangle}, \quad (9.39)$$

with \mathcal{N} a normalization factor. Now,

$$p(d|\boldsymbol{\theta}, \mathcal{H}) = p[d - h(\boldsymbol{\theta}; t)]. \quad (9.40)$$

Indeed, the probability for the data to be d , *given* the presence of a signal h of a particular type (the hypothesis \mathcal{H}), and *given* parameter values $\boldsymbol{\theta}$, can be identified with the probability that the noise takes the form $n = d - h(\boldsymbol{\theta}; t)$. Concretely,

$$p(d|\boldsymbol{\theta}, \mathcal{H}) = \mathcal{N} e^{-\frac{1}{2}\langle d-h(\boldsymbol{\theta})|d-h(\boldsymbol{\theta}) \rangle}. \quad (9.41)$$

Our other claim was that the value evidence $p(d|\mathcal{H})$ is set by the requirement that the posterior distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$ be normalized. To see this, write Eq. (9.37) as

$$p(\boldsymbol{\theta}|d, \mathcal{H}) p(d|\mathcal{H}) = p(d|\boldsymbol{\theta}, \mathcal{H}) p(\boldsymbol{\theta}|\mathcal{H}), \quad (9.42)$$

²² In the rest of this chapter we will simply assume that the search methodology explained in the previous chapter has indeed indicated the presence of a signal in the data, with high confidence.

and integrate both sides over all of parameter space:

$$\int d^N\theta p(\boldsymbol{\theta}|d, \mathcal{H}) p(d|\mathcal{H}) = \int d^N\theta p(d|\boldsymbol{\theta}, \mathcal{H}) p(\boldsymbol{\theta}|\mathcal{H}). \quad (9.43)$$

Note that in the left hand side, $p(d|\mathcal{H})$ does not depend on the parameters and can be taken out of the integral, so

$$\begin{aligned} \int d^N\theta p(\boldsymbol{\theta}|d, \mathcal{H}) p(d|\mathcal{H}) &= p(d|\mathcal{H}) \int d^N\theta p(\boldsymbol{\theta}|d, \mathcal{H}) \\ &= p(d|\mathcal{H}), \end{aligned} \quad (9.44)$$

where in the last line we used that the posterior probability distribution must be normalized. Setting this equal to the right hand side of (9.43), we obtain

$$p(d|\mathcal{H}) = \int d^N\theta p(d|\boldsymbol{\theta}, \mathcal{H}) p(\boldsymbol{\theta}|\mathcal{H}). \quad (9.45)$$

Now consider two competing hypotheses \mathcal{H}_1 and \mathcal{H}_0 . In the previous chapter, these were the hypotheses that there is a signal in the data, and that there is no signal in the data, respectively, but here they could be “the signal in the data is from a binary black hole inspiral” or “the signal in the data is from a binary neutron star inspiral”, etc. Then one can define an *odds ratio* (or as we called it in the previous chapter, a *likelihood ratio*)

$$\begin{aligned} O_0^1 &\equiv \frac{p(\mathcal{H}_1|d)}{p(\mathcal{H}_0|d)} \\ &= \frac{p(d|\mathcal{H}_1)}{p(d|\mathcal{H}_0)} \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}, \end{aligned} \quad (9.46)$$

where we have used Bayes’ theorem again. Recall that $p(\mathcal{H}_0)$ and $p(\mathcal{H}_1)$ were prior probabilities, and we know how to calculate the evidences $p(d|\mathcal{H}_0)$, $p(d|\mathcal{H}_1)$ through Eqs. (9.45) and (9.41). The odds ratio O_0^1 is used to rank hypotheses based on data: If $O_0^1 > 1$ then the data favor \mathcal{H}_1 , and if $O_0^1 < 1$ they favor \mathcal{H}_0 ; if $O_0^1 \simeq 1$ then the data can be considered uninformative regarding the hypotheses \mathcal{H}_0 and \mathcal{H}_1 .

Clearly, the key to measuring the probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$ parameters of the source from the available data, and ranking hypotheses about the source through odds ratios, is to calculate the likelihood function $p(d|\boldsymbol{\theta}, \mathcal{H})$. Now, while the expression (9.41) certainly allows us to calculate the likelihood at a given point $\boldsymbol{\theta}$ in parameter space, the integral defining the inner product $\langle \cdot | \cdot \rangle$ must usually be performed numerically. In order to accurately “map out” the posterior probability function $p(\boldsymbol{\theta}|d, \mathcal{H})$ (using Eq. (9.37)), this

integral needs be performed in a large number of points in parameter space. As already hinted at above, the high dimensionality of parameter space can make this computationally very challenging. There are various tricks to get around this, all of which are based on the observation that typically, the volume in parameter space where the likelihood differs considerably from zero tends to be quite small, so that computational resources can be focused on this small region. In the next chapter we will make this more concrete.

X. INFERRING THE PROPERTIES OF THE SOURCE FROM THE SIGNAL

In this Chapter we will explain how the likelihood function $p(d|\boldsymbol{\theta}, \mathcal{H})$ is obtained in practice. However, before doing so, it will be instructive to understand how much information about the source can *in principle* be extracted from the signal of a given kind, with a given detector. If only qualitative estimates are needed, for example to understand how a planned, future detector will perform depending on the noise power spectral density $S_n(f)$ that can be achieved, then a useful approximation can be arrived at through the *Fisher matrix formalism*, to which we now turn.

A. How much information can be extracted from the signal?

The posterior probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$ for the source parameters $\boldsymbol{\theta}$ is given in terms of the likelihood function, the prior probability distribution, and the evidence by Eq. (9.37). Assuming for simplicity that the prior distribution $p(\boldsymbol{\theta}|\mathcal{H})$ is chosen to be constant over some appropriate parameter ranges, the dependence on $\boldsymbol{\theta}$ only comes from the likelihood, and we can write

$$p(\boldsymbol{\theta}|d, \mathcal{H}) \propto p(d|\boldsymbol{\theta}, \mathcal{H}) = \mathcal{N} e^{-\frac{1}{2}\langle d - h(\boldsymbol{\theta}) | d - h(\boldsymbol{\theta}) \rangle}. \quad (10.1)$$

In this section we will furthermore assume that we are dealing with a loud signal that has a high signal-to-noise ratio (SNR). In that case the values of $\boldsymbol{\theta}$ at which the likelihood $p(d|\boldsymbol{\theta}, \mathcal{H})$ has its peak, called the *maximum likelihood values* $\boldsymbol{\theta}_{\text{ML}}$, will be approximately equal to the true values.²³ Now, a generic set of parameter values $\boldsymbol{\theta}$ can always be written as

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{ML}} + \delta\boldsymbol{\theta}. \quad (10.2)$$

Our goal in this section will be to estimate what are typical values for $\delta\boldsymbol{\theta}$ according to the probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$, which will then be indicative of the expected measurement uncertainties for the various parameters. We will do this by approximating this distribution by a multivariate Gaussian,

$$p(\boldsymbol{\theta}|d, \mathcal{H}) \propto e^{-\frac{1}{2}\Gamma_{IJ}\delta\theta^I\delta\theta^J}, \quad (10.3)$$

²³ In reality this will never be exactly the case; noise will cause the maximum likelihood values to be offset from the true ones. However, for sufficiently loud signals, measurement uncertainties will be small, and then our approximation will be a reasonable one.

where the repeated indices I and J are being summed over. The typical width of this function in the different parameter directions $\delta\theta^I$, $I = 1, 2, 3, \dots$ will then give us the approximate measurement uncertainties for these individual parameters. Clearly, these widths are determined by the so-called *Fisher matrix* Γ_{IJ} that appears in Eq. (10.3). Thus, our job is to figure out how to compute this matrix.

From Eq. (10.1),

$$\begin{aligned} p(\boldsymbol{\theta}|d, \mathcal{H}) &= \mathcal{N} e^{-\frac{1}{2}\langle d-h(\boldsymbol{\theta})|d-h(\boldsymbol{\theta})\rangle} \\ &= \mathcal{N} e^{-\frac{1}{2}\langle d|d\rangle + \langle d|h(\boldsymbol{\theta})\rangle - \frac{1}{2}\langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle}, \end{aligned} \quad (10.4)$$

where \mathcal{N} is some normalization factor. Taking the logarithm of both sides,

$$\begin{aligned} \log p(\boldsymbol{\theta}|d, \mathcal{H}) &= \log \mathcal{N} - \frac{1}{2}\langle d|d\rangle + \langle d|h(\boldsymbol{\theta})\rangle - \frac{1}{2}\langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle \\ &= \langle d|h(\boldsymbol{\theta})\rangle - \frac{1}{2}\langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle + \text{const}, \end{aligned} \quad (10.5)$$

where in the second line, the constant collects the terms that do not depend on $\boldsymbol{\theta}$.

Next, if the SNR is high, we can assume the width of the posterior probability distribution to be small. In that case, we can assume the values of $\delta\boldsymbol{\theta}$ in Eq. (10.2) to be small in the relevant region of parameter space, namely the region where the posterior distribution is large. It then makes sense to Taylor expand $h(\boldsymbol{\theta})$ to, say, quadratic order in the small quantities $\delta\theta^I$:

$$h(\boldsymbol{\theta}) \simeq h(\boldsymbol{\theta}_{\text{ML}}) + \left. \frac{\partial h}{\partial \theta^I} \right|_{\text{ML}} \delta\theta^I + \frac{1}{2} \left. \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right|_{\text{ML}} \delta\theta^I \delta\theta^J, \quad (10.6)$$

where once again repeated indices are being summed over. Substituting this into Eq. (10.5),

$$\begin{aligned} \log p(\boldsymbol{\theta}|d, \mathcal{H}) &= \langle d|h\rangle|_{\text{ML}} + \left\langle d \left| \frac{\partial h}{\partial \theta^I} \right. \right\rangle_{\text{ML}} \delta\theta^I + \frac{1}{2} \left\langle d \left| \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right. \right\rangle_{\text{ML}} \delta\theta^I \delta\theta^J - \frac{1}{2} \langle h|h\rangle|_{\text{ML}} \\ &\quad - \left\langle h \left| \frac{\partial h}{\partial \theta^I} \right. \right\rangle_{\text{ML}} \delta\theta^I - \frac{1}{2} \left\langle h \left| \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right. \right\rangle_{\text{ML}} \delta\theta^I \delta\theta^J - \frac{1}{2} \left\langle \frac{\partial h}{\partial \theta^I} \left| \frac{\partial h}{\partial \theta^J} \right. \right\rangle_{\text{ML}} \delta\theta^I \delta\theta^J \\ &\quad + \text{const}. \end{aligned} \quad (10.7)$$

This can be simplified by noting that since the logarithm is a monotonic function, the maximum likelihood value $\boldsymbol{\theta}_{\text{ML}}$ which maximizes $p(\delta\boldsymbol{\theta})$ will also maximize $\log p(\delta\boldsymbol{\theta})$. Thus, from Eq. (10.5),

$$\left. \frac{\partial}{\partial \theta_I} \log p(\boldsymbol{\theta}|d, \mathcal{H}) \right|_{\text{ML}} = 0 = \left\langle d \left| \frac{\partial h}{\partial \theta^I} \right. \right\rangle_{\text{ML}} - \left\langle h \left| \frac{\partial h}{\partial \theta^I} \right. \right\rangle_{\text{ML}}. \quad (10.8)$$

This leads to

$$\begin{aligned} \log p(\boldsymbol{\theta}|d, \mathcal{H}) = & \frac{1}{2} \left\langle d \left| \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J - \frac{1}{2} \left\langle h \left| \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J \\ & - \frac{1}{2} \left\langle \frac{\partial h}{\partial \theta^I} \left| \frac{\partial h}{\partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J + \text{const}, \end{aligned} \quad (10.9)$$

where we have also absorbed terms independent of $\delta \theta^I$ into the constant. Using $d = n + h(\boldsymbol{\theta})$, this further becomes

$$\log p(\boldsymbol{\theta}|d, \mathcal{H}) = \frac{1}{2} \left\langle n \left| \frac{\partial^2 h}{\partial \theta^I \partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J - \frac{1}{2} \left\langle \frac{\partial h}{\partial \theta^I} \left| \frac{\partial h}{\partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J + \text{const}. \quad (10.10)$$

Recall that we are assuming a signal with high SNR, which one can take to mean that the optimal SNR, $\rho_{\text{opt}} = \sqrt{\langle h|h \rangle}$, is high. In turn, this means that the amplitude of the signal is large. For binary inspiral and in the stationary phase approximation, we have seen that the signal takes the general form (9.25)

$$\tilde{h}(f) = \mathcal{A} f^{-7/6} e^{i\Psi(f)}, \quad (10.11)$$

so that we are assuming the constant \mathcal{A} to be large. Looking at the right hand side of Eq. (10.10), we see that the second term is proportional to \mathcal{A}^2 , while the first one is only proportional to \mathcal{A} .²⁴ Hence we can neglect the first term relative to the second term, and

$$\log p(\boldsymbol{\theta}|d, \mathcal{H}) \simeq -\frac{1}{2} \left\langle \frac{\partial h}{\partial \theta^I} \left| \frac{\partial h}{\partial \theta^J} \right. \right\rangle_{\text{ML}} \delta \theta^I \delta \theta^J + \text{const}. \quad (10.12)$$

Exponentiating both sides again, we find

$$p(\boldsymbol{\theta}|d, \mathcal{H}) \propto e^{-\frac{1}{2} \Gamma_{IJ} \delta \theta^I \delta \theta^J}. \quad (10.13)$$

This is nothing but Eq. (10.3), with the Fisher matrix Γ_{IJ} given by

$$\Gamma_{IJ} = \left\langle \frac{\partial h}{\partial \theta^I} \left| \frac{\partial h}{\partial \theta^J} \right. \right\rangle_{\text{ML}}. \quad (10.14)$$

Since a probability distribution needs to be normalized,

$$p(\boldsymbol{\theta}|d, \mathcal{H}) = \mathcal{C} e^{-\frac{1}{2} \Gamma_{IJ} \delta \theta^I \delta \theta^J}, \quad (10.15)$$

²⁴ Note that in this context we are not considering \mathcal{A} itself as being among the parameters θ^I ; it is just a scaling factor that sets the loudness of the signal.

where the normalization factor \mathcal{C} is such that

$$\int d\delta\boldsymbol{\theta} p(\boldsymbol{\theta}|d, \mathcal{H}) = 1. \quad (10.16)$$

Thus, we have arrived at the desired approximation of the posterior probability distribution as a multivariate Gaussian distribution in the $\delta\theta^I$.

As mentioned above, the width of the distribution (10.13) in the different parameter directions $\delta\theta^I$ can be viewed as the measurement uncertainties on the corresponding parameters. Characteristic widths in these directions, which we will denote by $\Delta\theta^I$, are obtained by calculating the root-mean-square expectation value of the $\delta\theta^I$ according to the probability distribution for these parameters:

$$\Delta\theta^I = \sqrt{\langle (\delta\theta^I)^2 \rangle}, \quad (10.17)$$

where

$$\langle (\delta\theta^I)^2 \rangle = \int d\delta\boldsymbol{\theta} (\delta\theta^I)^2 p(\delta\boldsymbol{\theta}), \quad (10.18)$$

with $p(\delta\boldsymbol{\theta}) = p(\boldsymbol{\theta}|d, \mathcal{H})$. Though we will not do it explicitly here, one can show that

$$\Delta\theta^I = \sqrt{\Sigma^{II}}, \quad (10.19)$$

where the *variance-covariance matrix* Σ^{IJ} is the inverse of the Fisher matrix,

$$\Sigma^{IJ} \equiv (\Gamma^{-1})^{IJ}, \quad (10.20)$$

and *no summation over indices is implied in Eq. (10.19)*.

To check that the above makes sense, let's consider the simple case where there is only one parameter θ which takes values on the real line. Then the Fisher matrix would be a 1×1 matrix Γ , and

$$\begin{aligned} p(\delta\boldsymbol{\theta}) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\Gamma\delta\theta^2} \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\delta\theta^2}{\sigma^2}}, \end{aligned} \quad (10.21)$$

where in the second line $\sigma^2 = \Gamma^{-1}$ is the variance of the Gaussian, and we have used that for a one-dimensional Gaussian, the normalization factor is $\mathcal{C} = 1/(\sigma\sqrt{2\pi})$. One has

$$\begin{aligned} \Delta\theta &= \sqrt{\langle (\delta\theta)^2 \rangle} \\ &= \sqrt{\int d\delta\theta (\delta\theta)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\delta\theta^2}{\sigma^2}}} \\ &= \sigma \\ &= \sqrt{\Gamma^{-1}}. \end{aligned} \quad (10.22)$$

This is the one-dimensional version of Eq. (10.19), as required. In one dimension, clearly $\Delta\theta$ is the one-sigma uncertainty on $\delta\theta$; in the more general case the $\Delta\theta^I$ will be the one-sigma uncertainties on the $\delta\theta^I$.

The above way of estimating measurement uncertainties is called the *Fisher matrix formalism*. Let us point out two important properties.

- Note that the integrand appearing in the inner product $\langle \cdot | \cdot \rangle$ is inversely proportional to the power spectral density $S_n(f)$ (see Eq. (8.31)), and recall from Chapter VIII that the latter is essentially the variance of the noise as a function of frequency. Thus, for a more sensitive detector (one with less noise), $S_n(f)$ will take on smaller values, which through Eq. (10.14) implies larger values of the Γ_{IJ} ; through Eq. (10.20) this leads to smaller values of Σ^{IJ} , and through Eq. (10.19), to smaller 1-sigma measurement uncertainties $\Delta\theta^I$. In other words, a more sensitive detector leads to smaller measurement uncertainties, as expected.
- For the signal from an inspiraling compact binary, the optimal SNR, $\rho_{\text{opt}} = \sqrt{\langle h|h \rangle}$, is proportional to \mathcal{A} , the overall amplitude of the waveform in Eq. (10.11). On the other hand, we already noted that the Fisher matrix elements scale like \mathcal{A}^2 . Hence, $\Sigma^{IJ} \propto 1/\mathcal{A}^2$, and $\Delta\theta^I \propto 1/\mathcal{A}$. Thus, $\Delta\theta^I \propto 1/\rho_{\text{opt}}$. This means that when a signal appears louder in the detector, measurement uncertainties will be smaller, again as one would expect.

It is instructive to apply the Fisher matrix using the approximative waveform model of Eq. (9.25), along with an assumed power spectral density $S_n(f)$. First note that

$$\mathcal{A} = \frac{c}{r} \sqrt{\frac{5\pi}{96}} \sqrt{F_+^2(1 + \cos^2(\iota))^2 + F_\times^2 4 \cos^2(\iota)} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{5/6} \pi^{-7/6}. \quad (10.23)$$

With a single detector (as opposed to a network of detectors), it will not be possible to separately measure the distance r , the angles (θ, ϕ, ψ) appearing in the beam pattern functions, and the inclination angle ι ; only the combination \mathcal{A} appears in the waveform. Secondly, the phase $\Psi(f)$ can be written as

$$\Psi(f) = 2\pi f t_c - \Phi_{\text{ref}} + \frac{3}{4} \left(\frac{8\pi G\mathcal{M}_c f}{c^3} \right)^{-5/3}, \quad (10.24)$$

where we have defined the reference phase $\Phi_{\text{ref}} = \Phi_c + \varphi_0 + \pi/4$; it is only this combination that can be measured. Third, our simplified waveform model only depends on the masses m_1 ,

m_2 through the chirp mass \mathcal{M}_c ; here too, it is only this combination that will be measurable. Thus, in this simple example, the measurable parameters are

$$\boldsymbol{\theta} = (\mathcal{M}_c, t_c, \Phi_{\text{ref}}), \quad (10.25)$$

where we reiterate that we do not include \mathcal{A} , since it is just an overall scaling of the waveform. However, note that although the parameter Φ_{ref} is not of much physical interest, this “nuisance parameter” must nevertheless be included in our Fisher matrix calculations; leaving it out would lead to incorrect values for the uncertainties $\Delta\theta^I$ on the other parameters. Typically, all the measurable parameters are to some extent correlated with each other, and will have an effect on the measurement uncertainties of all the others. Mathematically this can be seen from the fact that the variance-covariance matrix Σ^{IJ} , being the inverse of the Fisher matrix, tends to mix together all the elements of Γ_{IJ} ; thus, the quantities $\Delta\theta^I = \sqrt{\Sigma^{II}}$ will usually receive contributions also from Fisher matrix elements both of whose indices are different from I .

The Fisher matrix formalism provides a convenient way to predict how accurately one would be able to measure the parameters of an inspiraling binary with a given detector, if a model for the power spectral density $S_n(f)$ is available. More than 10 years before Advanced LIGO was switched on, an following approximate analytic expression was predicted,²⁵ which later on turned out to be reasonably close to the truth. For frequencies $f \geq 20$ Hz,

$$S_h(f) = S_0 \left[x^{-4.14} - 5x^{-2} + \frac{111(1 - x^2 + x^4/2)}{1 + x^2/2} \right], \quad (10.26)$$

with $S_0 = 10^{-49}$ Hz⁻¹ and $x = f/f_0$, where $f_0 = 215$ Hz. For $f < 20$ Hz, $S_n(f)$ can be taken to be infinite for all practical purposes.

Having made our choices for the waveform model and for the noise power spectral density, the estimation of measurement uncertainties in the Fisher matrix formalism proceeds as follows:

1. Pick values for \mathcal{A} , \mathcal{M}_c , t_c , and Φ_{ref} , which will be the “true” values in the context of the formalism. Note that with a value for \mathcal{M}_c , choosing a value for \mathcal{A} requires making a choice for the true values of the distance r and the angles $(\theta, \phi, \psi, \iota)$, even if these can not be separately measured.

²⁵ See <https://arxiv.org/pdf/gr-qc/0411146.pdf>.

2. Calculate the partial derivatives $\frac{\partial \tilde{h}}{\partial \mathcal{M}_c}(f)$, $\frac{\partial \tilde{h}}{\partial t_c}(f)$, and $\frac{\partial \tilde{h}}{\partial \Phi_{\text{ref}}}(f)$.
3. Calculate the Fisher matrix elements in Eq. (10.14), taking the upper integration limit in the calculation of inner products to be the ISCO frequency, $f_{\text{ISCO}} = c^3/(6^{3/2}\pi GM)$.
4. Invert the Fisher matrix to obtain Σ^{IJ} .
5. The square roots of the diagonal elements of the latter matrix give the measurement uncertainties $\Delta\theta^I$.

As an example, for a binary neutron star inspiral with component masses $m_1 = m_2 = 1.4 M_\odot$ and $t_c = \Phi_{\text{ref}} = 0$, with \mathcal{A} chosen such that $\rho_{\text{opt}} = 10$, one finds

$$\frac{\Delta \mathcal{M}_c}{\mathcal{M}_c} = 5.4 \times 10^{-5}, \quad \Delta t_c = 0.23 \text{ ms}, \quad \Delta \Phi_{\text{ref}} = 0.29 \text{ rad}. \quad (10.27)$$

Thus, the chirp mass can be measured to great accuracy. The parameter t_c , which can be viewed as the time of arrival of the signal at the detector, is measurable to less than a millisecond, which is considerably smaller than the light travel time between the two LIGO detectors (~ 10 ms) and either of those and the Virgo detector (~ 27 ms), enabling localization of the source in the sky (see the discussion at the end of the previous Chapter). Finally, the uncertainty on the reference phase is much smaller than 2π , so that this parameter will also be well measurable.

B. Gravitational wave parameter estimation in real life: The nested sampling algorithm

Though the Fisher matrix formalism is very useful to estimate the precision with which various parameters will be measurable for a given signal and a given detector with noise power spectral density $S_n(f)$, it can not be used to analyze real data containing a signal. Indeed, the Fisher matrix formalism assumes that the parameter values $\boldsymbol{\theta}_{\text{ML}}$, where the likelihood function $p(d|\boldsymbol{\theta}, \mathcal{H})$ has its peak, coincides with the true values $\boldsymbol{\theta}_{\text{true}}$, but in reality, noise will cause $\boldsymbol{\theta}_{\text{ML}}$ to be offset from $\boldsymbol{\theta}_{\text{true}}$. Moreover, the approximation of $p(d|\boldsymbol{\theta}, \mathcal{H})$ as a multivariate Gaussian in $\delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_{\text{ML}}$ will only be approximately true in the limit of high SNR, while in reality, the majority of our detected signals had an SNR just above the threshold of detectability. Thus, we are still faced with the problem of obtaining $p(d|\boldsymbol{\theta}, \mathcal{H})$,

or at least an excellent approximation to it, without making assumptions that are overly restrictive in a realistic data analysis scenario.

A technique that does this very efficiently is *nested sampling*, which we now explain.²⁶

Though our goal is to obtain a good approximation to the posterior probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$, with nested sampling this arises as a by-product of an approximation for the evidence integral in Eq. (9.45). Denoting the evidence $p(d|\mathcal{H})$ by Z , the likelihood $p(d|\boldsymbol{\theta}, \mathcal{H})$ by $L(\boldsymbol{\theta})$, and the prior probability distribution $p(\boldsymbol{\theta}|H)$ by $\pi(\boldsymbol{\theta})$, this integral is

$$Z = \int L(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d^N \boldsymbol{\theta}. \quad (10.28)$$

The idea of nested sampling is now to approximate this multi-dimensional integral by a *one-dimensional* integral in terms of a single scalar X , called the *prior mass*. Let $\boldsymbol{\theta}$ be a point in parameter space, whose likelihood is $L(\boldsymbol{\theta})$. Then prior mass is defined as

$$X(\boldsymbol{\theta}) = \int_{L(\boldsymbol{\theta}') > L(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}') d^N \boldsymbol{\theta}', \quad (10.29)$$

i.e. it is the integral of the prior distribution over the volume in parameter space where the likelihood is everywhere larger than the given $L(\boldsymbol{\theta})$. Note that, since the prior probability distribution $\pi(\boldsymbol{\theta})$ is normalized, one has $X \in [0, 1]$. An *element of prior mass* is defined as

$$dX(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}) d^N \boldsymbol{\theta}. \quad (10.30)$$

What nested sampling does is to replace the integral (10.28) by a one-dimensional integral

$$Z = \int_0^1 L(X) dX, \quad (10.31)$$

where $L(X)$ is the likelihood on the hypersurface that bounds the volume in parameter space on which X is defined. The way the algorithm operates is to sample the function $L(X)$ by finding points $\boldsymbol{\theta}_i$ in parameter space with progressively larger likelihoods L_k and hence progressively smaller prior mass X_k . Writing $\Delta X_k = X_k - X_{k+1}$,

$$Z \simeq \sum_{k=0}^N L_k \Delta X_k. \quad (10.32)$$

²⁶ This method was discovered by John Skilling, and published only in 2006: <https://projecteuclid.org/journals/bayesian-analysis/volume-1/issue-4/Nested-sampling-for-general-Bayesian-computation/10.1214/06-BA127.full>.

From Eq. (9.37) for the posterior distribution,

$$p(\boldsymbol{\theta}|d, \mathcal{H}) = \frac{L(\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{Z}, \quad (10.33)$$

together with Eq. (10.30), one infers the discrete version

$$p(\boldsymbol{\theta}|d, \mathcal{H}) \simeq \frac{L_k \Delta X_k}{Z}. \quad (10.34)$$

Thus, nested sampling approximates the evidence, and in the process it samples the posterior probability distribution.

The nested sampling algorithm can be outlined as follows. To initialize the algorithm, pick M points $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_M$, called *live points*, randomly from the prior distribution, and calculate their likelihoods $\tilde{L}_1, \tilde{L}_2, \dots, \tilde{L}_M$ using Eq. (9.41). The rest of the algorithm is iterative, with the k th iteration looking like this:

1. In the list of M live points, let $\boldsymbol{\theta}_i$ be the point with the lowest likelihood \tilde{L}_i . Consider a hypersurface in parameter space that goes through $\boldsymbol{\theta}_i$ and such that the likelihood is \tilde{L}_i everywhere on the hypersurface. Integrating the prior distribution over the enclosed volume we obtain a prior mass \tilde{X}_i . Rename the prior mass and likelihood to (X_k, L_k) and record them.
2. Now discard this point with lowest likelihood and pick a new point randomly from the prior distribution, but such that its likelihood is larger than that of the one we just discarded. Together with the $M - 1$ points we had remaining, we again have M live points in parameter space.
3. With this new list of M points, go back to step 2.

Thus, the algorithm records a succession of (X_k, L_k) such that the likelihoods L_k keep moving to increasingly higher values, and the X_k to lower values (because the associated prior volumes keep shrinking); see Fig. 31. In the end these can be used to approximate the evidence through Eq. (10.32) and sample the posterior distribution through Eq. (10.34).

However, step 2 in the algorithm presents us with a problem: How would one identify these surfaces of constant likelihood? Indeed, in a high-dimensional parameter space this could represent a formidable numerical problem. What one does instead is to make an “educated guess” for the value of X_k in each iteration. In statistic, this means randomly picking a value from the *probability distribution* that the variable follows.

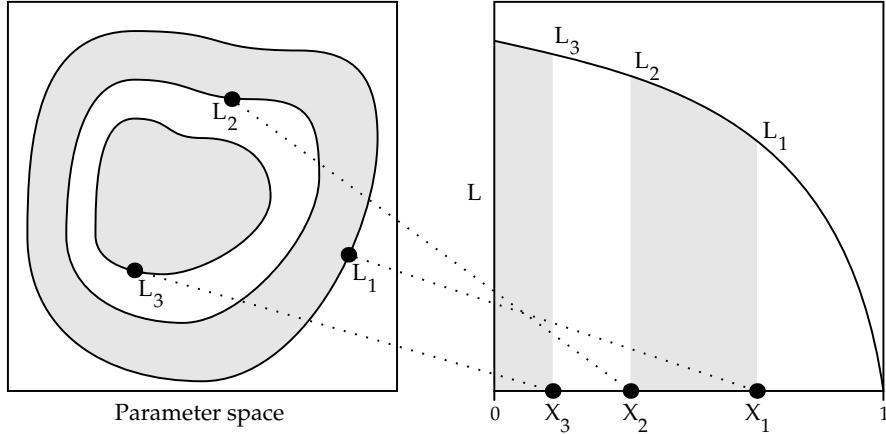


Figure 31: Left: The nested sampling algorithm produces a list of points in parameter space with progressively larger likelihoods L_k . Right: Associated with these are progressively smaller prior masses X_k . Note that the area under the curve is an approximation for the evidence Z .

Suppose we are at iteration k of the algorithm, so that the current live points have prior masses $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_M$, and the previously discarded point had prior mass X_{k-1} . Define the *shrinkage ratios*

$$\tilde{t}_1 = \frac{\tilde{X}_1}{X_{k-1}}, \quad \tilde{t}_2 = \frac{\tilde{X}_2}{X_{k-1}}, \quad \dots, \quad \tilde{t}_M = \frac{\tilde{X}_M}{X_{k-1}}. \quad (10.35)$$

Given some $s \in [0, 1]$, consider the probability that the largest shrinkage ratio, which we will denote t_k , is no greater than s . This equals the probability that *all* of the \tilde{t}_i are no greater than s :

$$\begin{aligned} P(t_k \leq s) &= \prod_{i=1}^M P(\tilde{t}_i \leq s) \\ &= \prod_{i=1}^M s \\ &= s^M, \end{aligned} \quad (10.36)$$

where in the next-to-last line we have used that the \tilde{t}_i are uniformly distributed between 0 and 1. From this we infer that the probability *density* for t_k is

$$p(t_k, M) = M t_k^{M-1}. \quad (10.37)$$

Indeed, integrating this from 0 to s should give us the probability that t_k is at most s , and this is the case:

$$\int_0^s p(t_k, M) dt_k = s^M = P(t_k \leq s). \quad (10.38)$$

Thus, to construct the X_k , in each iteration one sets

$$X_k = t_k X_{k-1}, \quad (10.39)$$

where t_k is always drawn from the same distribution,

$$p(t, M) = Mt^{M-1}. \quad (10.40)$$

An exception is $k = 0$, where it is natural to set $X_0 = 1$.

Finally, our algorithm needs a termination condition. In the k th iteration, the evidence accumulated up to that point is

$$Z^{(k)} = \sum_{j=0}^k L_j \Delta X_j. \quad (10.41)$$

Among the M live points at this iteration, let $\tilde{L}_{\max}^{(k)}$ be the largest likelihood, and $\tilde{X}_{\max}^{(k)}$ the largest prior mass. Note that these won't belong to one and the same live point, since the point with highest likelihood will have lowest prior mass, and vice versa. However, $\tilde{L}_{\max}^{(k)} \tilde{X}_{\max}^{(k)}$ can be viewed as an estimate of how much additional evidence is still to be accumulated.

The algorithm is terminated when

$$\tilde{L}_{\max}^{(k)} \tilde{X}_{\max}^{(k)} < \alpha Z^{(k)}, \quad (10.42)$$

with $\alpha \ll 1$ some arbitrarily chosen number. Thus, the algorithm terminates when the estimated amount of evidence still to be accumulated is much smaller than the evidence already accumulated.

How efficient is nested sampling? To get an idea, first note that in the k th iteration one has

$$X_k = \prod_{j=1}^k t_j, \quad (10.43)$$

where each of the t_j has been picked randomly from the distribution $p(t, M)$ of Eq. (10.40). One can show that the mean and standard deviation of $\log(t)$ are given by

$$\log(t) \sim (-1 \pm 1)/M. \quad (10.44)$$

Because of (10.43), the mean and standard deviation of $\log(X_k)$ are

$$\log(X_k) \sim (-k \pm \sqrt{k})/M. \quad (10.45)$$

Hence the mean values of the X_k themselves go like

$$\langle X_k \rangle \propto \exp(-k/M). \quad (10.46)$$

Hence, small prior masses, corresponding to high likelihoods, are reached exponentially fast, and also the errors on X_k decrease exponentially quickly.

Now that we have an algorithm to efficiently probe parameter space and approximate the posterior probability distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$, let us look at some results from a real-life example.

The distribution $p(\boldsymbol{\theta}|d, \mathcal{H})$ is a probability distribution for all of the parameters (9.34) together. However, suppose we want to focus on just a few of the parameters, for example m_1 and m_2 . Then a probability distribution for just those two parameters is arrived at by integrating out all of the others:

$$p(m_1, m_2) = \int d\boldsymbol{\lambda} p(m_1, m_2, \boldsymbol{\lambda}|d, \mathcal{H}), \quad (10.47)$$

where $\boldsymbol{\lambda}$ stands for all the parameters in (9.34) *except* m_1, m_2 . If one is only interested in m_1 in particular, then additionally m_2 can also be integrated out, arriving at a distribution that only depends on m_1 .

Fig. 32 shows the result for the binary black hole coalescence GW151226. Note how the individual distributions of m_1 and m_2 are relatively wide. On the other hand, the joint distribution $p(m_1, m_2)$ draws out a “banana” that is thin in one direction and long in another. The reason can be glimpsed from Eqs. (7.24), (7.27), and (7.28) for $h_+(t)$ and $h_\times(t)$ in terms of $f_{\text{gw}}(t)$ and $\Phi_{\text{gw}}(t)$. Within the approximations we made, all these expressions depend on the masses only through the chirp mass $\mathcal{M}_c = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$. It will then be no surprise if this combination of masses turns out to be particularly well-measurable. And indeed, $p(m_1, m_2)$ is mostly concentrated along a curve of constant chirp mass, where in this case $\mathcal{M}_c \simeq 9 M_\odot$. However, we stress again that in reality, the waveform will have corrections to the simplified expression we arrived at earlier, which depend on different combinations of m_1 and m_2 , so that the component masses are individually measurable, albeit much less accurately than \mathcal{M}_c .

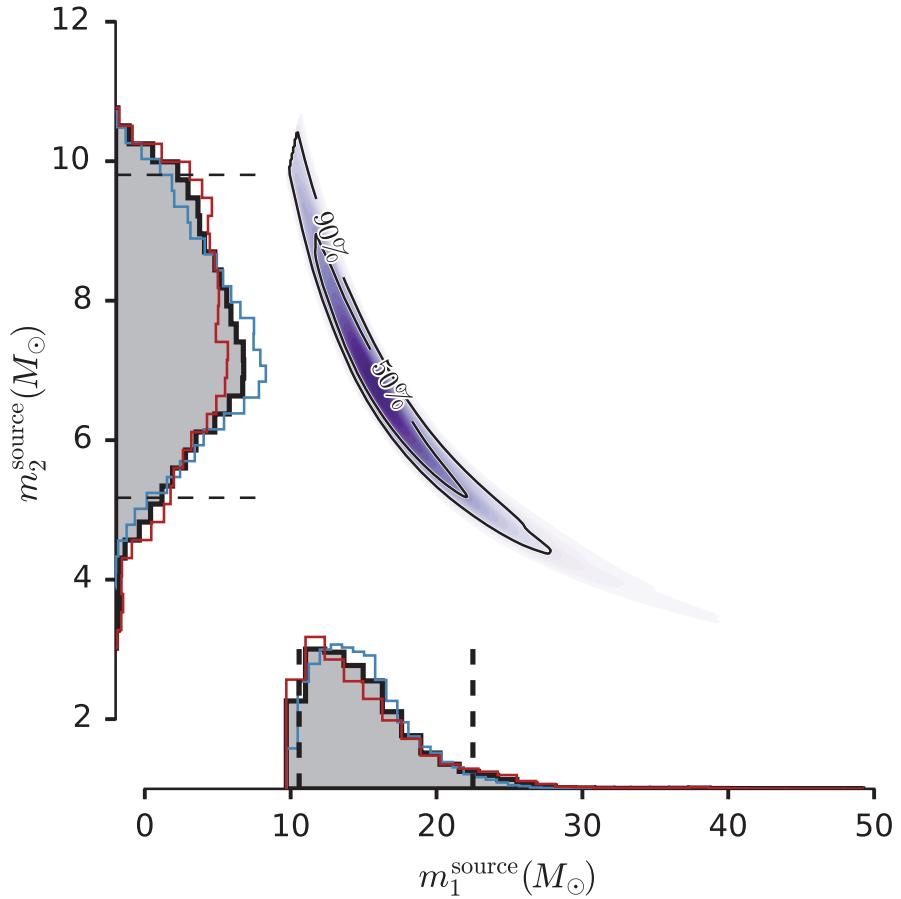


Figure 32: Probability distributions of the component masses m_1 , m_2 for the binary black hole signal GW151226. The grey distributions to the left and at bottom are for m_1 and m_2 individually; dashed lines indicate intervals containing 90% of probability. The blue shading in the middle gives an impression of the joint distribution $p(m_1, m_2)$. A darker blue indicates the region where probability is higher; also shown are contours enclosing 50% and 90% of probability. (Picture from <https://arxiv.org/pdf/1606.04855.pdf>.)

XI. WHAT HAS BEEN DETECTED SO FAR? A CENSUS OF BLACK HOLES AND NEUTRON STARS

At the time of writing, 90 detections had been made of signals from coalescing compact binaries, including binary black hole, binary neutron star, and neutron star-black hole coalescences. This has enabled us to start making a “census” of the properties of compact objects occurring in binaries in terms of their masses m_1 , m_2 and spins \mathbf{S}_1 , \mathbf{S}_2 .

A. A census of black hole and neutron star masses

Fig. 32 already showed the result of mass measurements, for the signal GW151226. In the case of binary inspiral, the component masses m_1 , m_2 enter the waveform predominantly through a particular combination of the two, namely the chirp mass \mathcal{M}_c . For this reason, the combined posterior probability distribution for (m_1, m_2) came out “banana-shaped”, with the probability distribution bunched around a curve of constant \mathcal{M}_c . Now, we have seen that the inspiral ends at a frequency $f_{\text{gw,ISCO}} \propto 1/M$, with M the total mass. The signal GW151226 originated from a relatively low-mass binary black hole, in which case the inspiral regime dominates. For more massive systems, less of the inspiral is in the detectors’ sensitive band, and the merger and ringdown becomes proportionally more important. As it turns out, the physics of the latter two regimes is dominated not so much by the chirp mass \mathcal{M}_c , but by the total mass M . The effect of this can be seen in the left panel of Fig. 33, which shows mass measurements from the binary neutron star signal GW170817 as well as the first ten binary black hole signals that were discovered. For low mass systems, the 90% probability contours are banana shaped; however, for more massive systems total mass also becomes important, which widens the probability contours.

The right panel of Fig. 33 shows the inferred properties of the remnant objects of the binary black hole mergers, *i.e.* the single black hole resulting from the coalescences of two black holes. In each case, after the ringdown had died off, a stationary, rotating black hole will have been left. The *no-hair conjecture* states that the spacetime geometry of such black holes is completely determined by just two numbers: their mass M_f and spin a_f .²⁷ Given the masses m_1 , m_2 and spins \mathbf{S}_1 , \mathbf{S}_2 of the individual black holes during the inspiral, general

²⁷ Here a_f is the magnitude of the dimensionless spin vector.

relativity allows one to calculate M_f and a_f .²⁸ Through this correspondence, probability distributions for $(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2)$ translate to probability distributions for (M_f, a_f) , and this is what is shown. Note how the masses M_f occupy a wide range, roughly between 20 and $100 M_\odot$. The values for a_f tend to lie within a range $\sim 0.5 - 0.9$. In the case of equal-mass ($m_1 = m_2$), zero-spin ($\mathbf{S}_1 = \mathbf{S}_2 = 0$) binary black hole coalescences, computer simulations show $a_f = 0.68646 \pm 0.00004$ irrespective of what the total mass is. Of course, the binary black holes that produced the detected signals will have had unequal masses and probably non-zero spins. Nevertheless, the measured m_1 and m_2 do not tend to differ dramatically, and (also as seen in simulations) in that case the angular momentum of the final black hole comes largely from the *orbital angular momentum* of the binary system during the last stages of inspiral, with relatively minor contributions from \mathbf{S}_1 and \mathbf{S}_2 . This explains the clustering of a_f around a value $\simeq 0.7$ that we see in the figure.

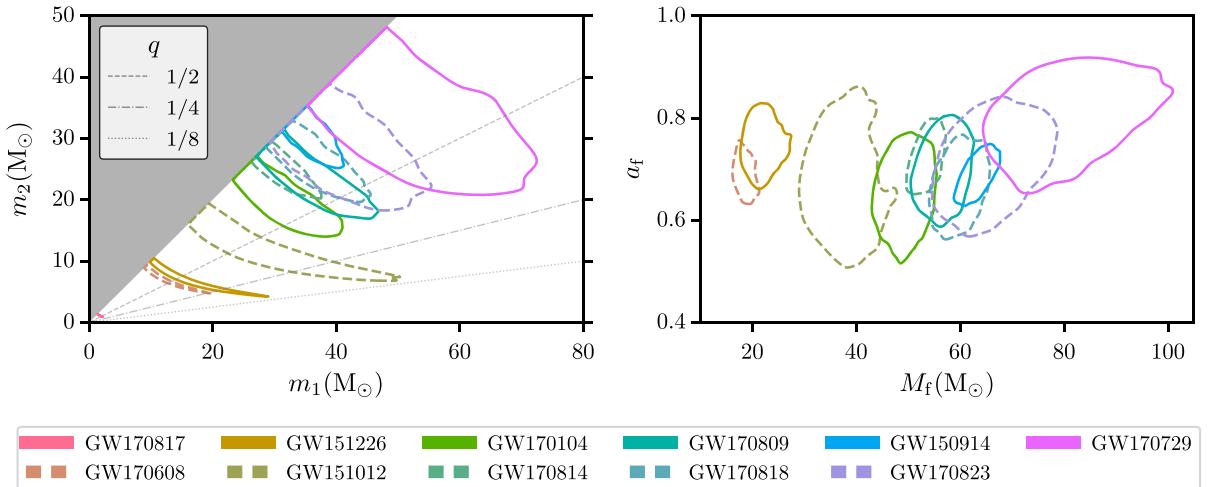


Figure 33: Left panel: Mass measurements for the binary neutron star coalescence GW170817 and for the first ten binary black hole detections. The straight lines are lines of constant mass ratio $q = m_2/m_1$, with the convention $m_2 \leq m_1$. Right panel: Inferred probability distributions for the masses M_f and spin magnitudes a_f of each of the remnant black holes resulting from the binary black hole mergers. (Picture from <https://arxiv.org/pdf/1811.12907.pdf>.)

Fig. 34 shows a glimpse of what is called the “stellar graveyard”. In particular, it shows

²⁸ This can not be done analytically, but large-scale computer simulations of coalescing binary black holes with different component masses and spins have enabled high-fidelity fits for M_f and a_f as functions of the masses and spins of the two black holes that merged.

what was known about the masses of black holes and neutron stars in binaries, up to the time of writing (April 2022). Apart from the gravitational wave observations, also indicated are black hole masses inferred from X-ray observations. The latter are black holes in orbit with an ordinary star, which causes accretion onto the black hole; the heating-up of the material in the accretion disk leads to X-ray emission, confirming that an ultra-compact object is indeed present in the binary. At the same time, Doppler modulation in the spectrum of the companion star tell us about the orbital motion, from which the black hole mass can be estimated. At lower masses, where neutron stars reside, radio observations of pulsars in binaries that are seen in our galaxy also use Doppler modulation to infer neutron star masses.

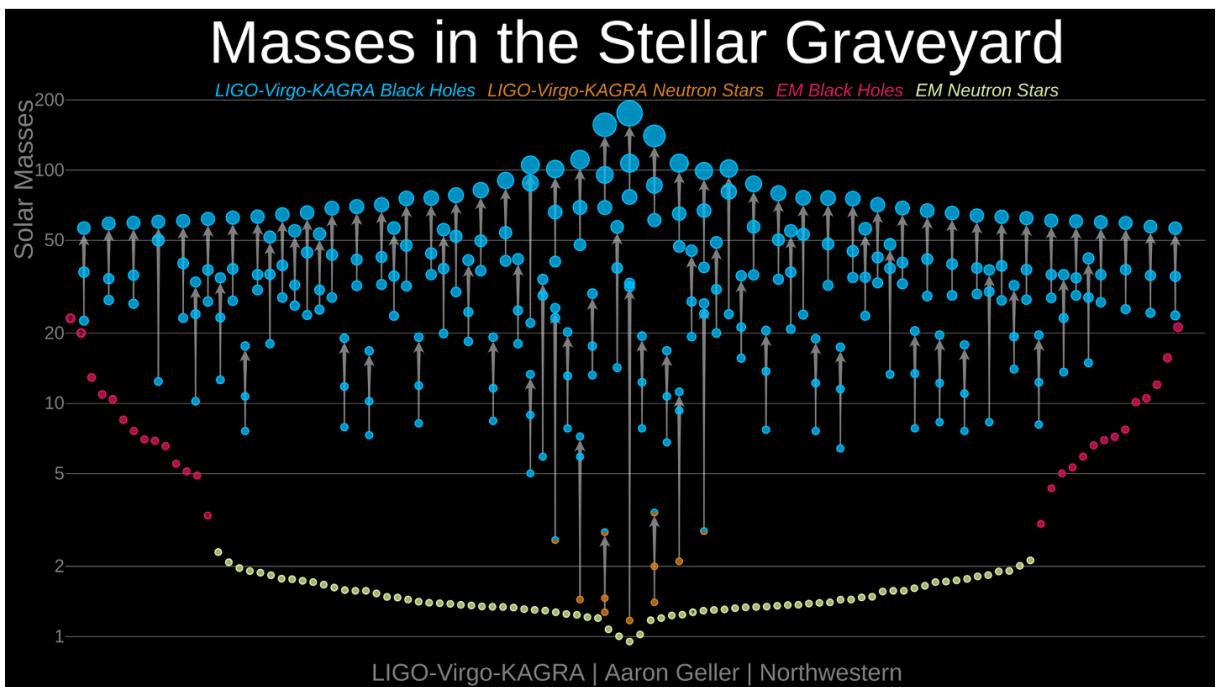


Figure 34: The “stellar graveyard”. The vertical axis indicates masses in M_{\odot} ; the horizontal axis has no meaning. Blue circles show the binary black hole mergers seen with LIGO and Virgo to date; each arrow connecting two such masses corresponds to a binary merger, with the point of the arrow indicating the mass of the remnant black hole. Purple circles are black holes seen with X-ray and other electromagnetic measurements. Yellow and orange dots are neutron stars. Note the binary neutron star coalescences in the bottom middle, which are GW190425 and GW170817, and the two neutron star-black hole coalescences, which are GW200105 and GW200115.

A few things are in evidence:

1. The black holes in gravitational wave observations go up to considerably higher masses than for X-ray binaries.
2. Generally, no black holes are seen with mass $\lesssim 5 M_\odot$, and no neutron stars with mass $\gtrsim 2 M_\odot$. There is, however, a binary coalescence for which one component is a “mystery object” with $\simeq 2.6 M_\odot$, which we will get back to below.
3. No compact objects are seen with mass $\lesssim 1 M_\odot$.

Regarding point 1, this will in part be due to observational bias: higher-mass binary black holes have stronger gravitational wave emission,²⁹ so that they can be seen out to larger distances; hence a larger volume of space is being covered, and we should detect more of them. Nevertheless, a few events are of special note, for which one or both component masses are in excess of $60 M_\odot$. There are strong physical arguments that black holes this massive can not be created directly from the collapse of an ordinary star. This is because of *pair instability*. Most of the pressure inside a star comes from photons bouncing around. Now, when a sufficiently massive star has spent most of its fuel and starts collapsing, temperature rises to such an extent that the photons acquire energies sufficient to create positron-electron pairs: $\gamma + \gamma \rightarrow e^+ + e^-$. This removes a significant fraction of photons, causing a drop in pressure inside the star. The result is a catastrophic collapse, during which the remaining nuclear fuel is spent in an extremely short time, and this in turn causes an explosion of such violence that it leaves behind no remnant – in particular, no black hole. Hence, when a black hole with mass considerably above $60 M_\odot$ is detected in a binary, then most likely it will have been the remnant from a *previous* merger of compact objects. Finally, black holes with a mass in excess of $100 M_\odot$, which is the case for the remnants of the most massive binary black hole mergers that were seen, are referred to as *intermediate mass* black holes, as they reside in a gap between black holes created directly from stars, and the supermassive black holes of millions to billions of solar masses that are known to exist in the centers of galaxies. It has been speculated that supermassive black holes form from successive mergers of smaller black holes, but this requires the existence of black holes with masses in between the two extremes. Such black holes are now being found.

²⁹ Recall that the amplitudes of h_+ and h_\times go like $\mathcal{M}_c^{5/3}$.

Point 2 refers to what is called the “mass gap”. It is reasonable that neutron stars can not be too heavy, because then they would collapse to a black hole. However, why would there not be a continuum of black hole masses starting just above $2 M_{\odot}$? If the mass gap were real, then it may be a consequence of qualitative differences between the kind of stars that make neutron stars and the kind that make black holes. Note that these speculations pertain to compact objects originating directly from the collapse of ordinary stars; the binary neutron star mergers GW170817 and GW190425 probably both led to a black hole with mass well above $2 M_{\odot}$. However, here too there is a special event, namely GW190814, for which one of the *component* masses, $2.6 M_{\odot}$, was well within the alleged mass gap. This component object is widely believed to be a black hole, but how it came about is not clear – perhaps as the result of an earlier binary neutron star merger?

Next we focus on GW190425. This event was only observed by LIGO Livingston and Virgo (the LIGO Hanford detector was down at the time), so that high-accuracy sky localization was not possible; and indeed, astronomers were not able to find an electromagnetic counterpart. The left panel of Fig. 35 shows mass measurements done in two different ways. If the component objects were neutron stars, then from general astrophysical arguments³⁰ their dimensionless spins $|\mathbf{S}_1|$, $|\mathbf{S}_2|$ will have been smaller than 0.05. Under this assumption, mass measurements put m_1 , m_2 in ranges consistent with neutron stars. But in the absence of an electromagnetic counterpart, we must leave open the possibility that one or both of the objects was a black hole. When repeating the mass measurement but allowing for a much higher spin, the posterior distributions for m_1 , m_2 become wider, and indeed m_1 may have been as large as $3 M_{\odot}$, hence potentially in the “mass gap” again!

In any case, the source of GW190425 was an unusual one. The right panel of Fig. 35 compares estimates of its *total mass* $m_{\text{tot}} = m_1 + m_2$ with total masses of binary neutron stars in our Milky Way, inferred from radio observations. Whatever assumptions one makes about spins, GW190425 is inconsistent with the total masses of galactic binary neutron stars. Future gravitational wave observations may turn up similar events, possibly with electromagnetic counterparts, so that an in-depth study can be performed. For now the results of Fig. 35 remain somewhat mysterious.

³⁰ If neutron stars spin too fast, they will fling themselves apart; it is believed that the dimensionless spin can not be much higher than ~ 0.05 . For comparison, the fastest-spinning neutron star known (from radio observations) has a dimensionless spin ~ 0.02 .

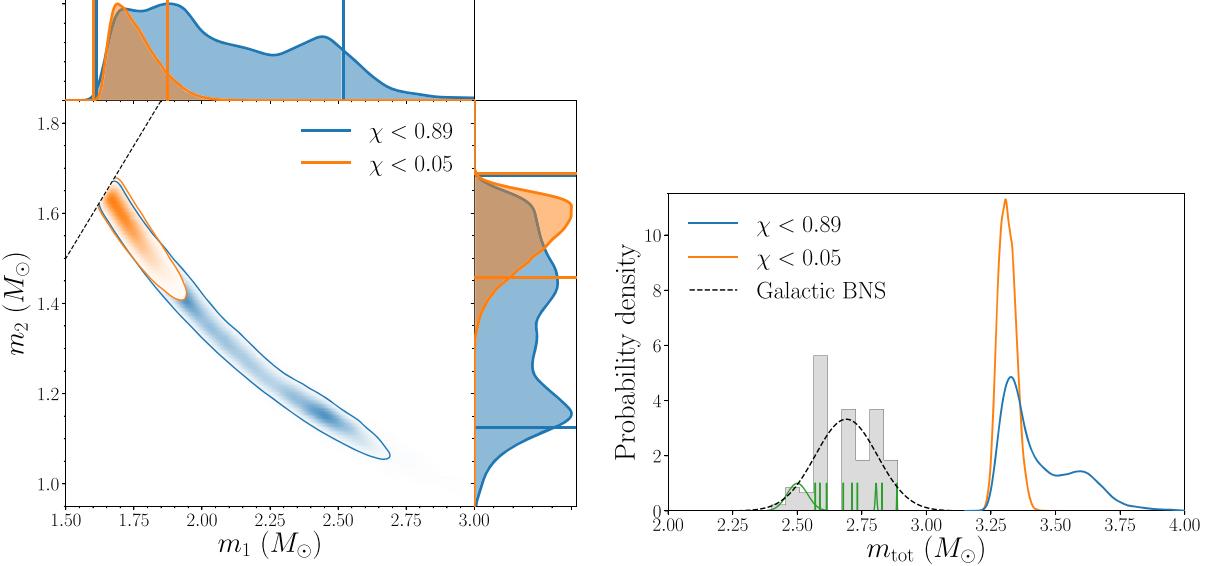


Figure 35: Left panel: mass measurements for GW190425, under the assumption that the spins were in the neutron star range (orange), or alternatively in the black hole range (blue). Right panel: a comparison of the *total mass* of GW190425 with total masses of galactic binary neutron stars from radio observations (the grey histogram). With either assumption on spins, GW190425 is inconsistent with pre-existing binary neutron star observations. (Pictures from <https://arxiv.org/pdf/2001.01761.pdf>.)

Regarding point 3, LIGO and Virgo are certainly sensitive to signals from binaries with a mass below $1 M_\odot$. The fact that so far we have not observed such objects enables us to make a statement about cosmology. There is no known astrophysical mechanism that would create such light neutron stars or black holes from ordinary stars. However, it is possible that light black holes were forged in the primordial soup that existed immediately after the Big Bang; these are called *primordial black holes*. If so, an interesting question is whether they could be among the contributors to dark matter. This is explored in Fig. 36. Given that no binaries with component objects lighter than $1 M_\odot$ are seen, one can put an upper bound on the fraction f_{PBH} of dark matter they may constitute. LIGO and Virgo have already led to a better upper bound on f_{PBH} than any other observations, with further improvements to be expected in coming years.

We stress that it is entirely possible that the black holes we are actually seeing (the ones in Fig. 34) are themselves primordial; in fact, there are models for primordial black holes

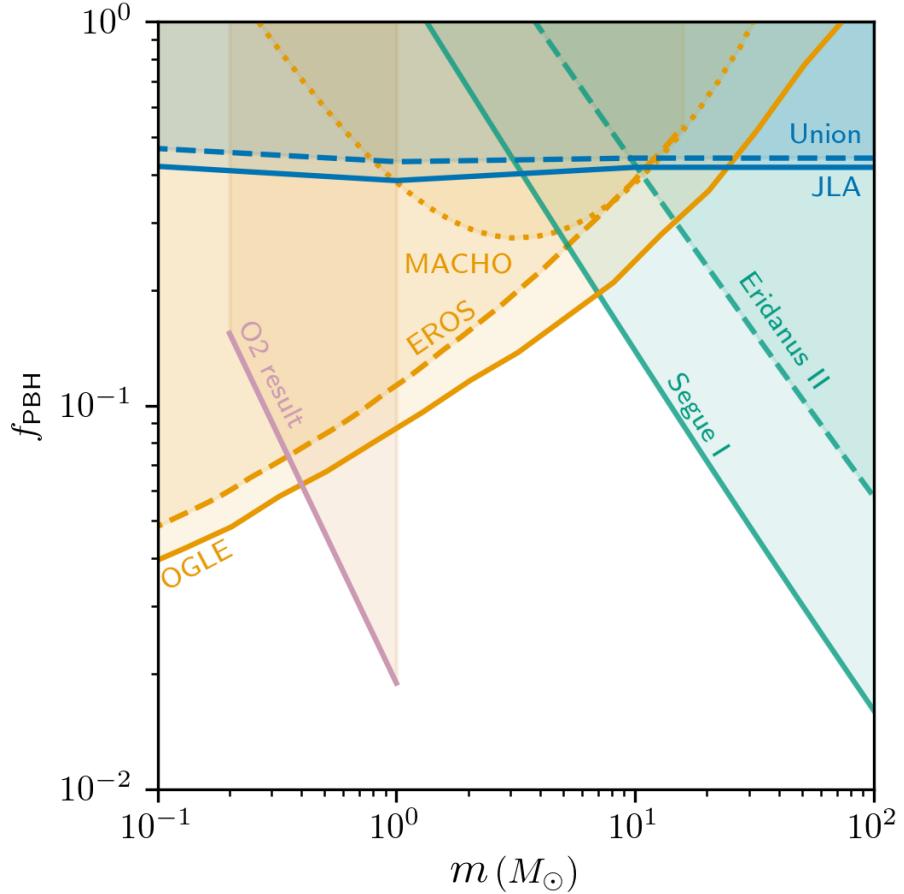


Figure 36: Upper bounds on the fraction f_{PBH} of dark matter that could be contributed by primordial black holes. “OGLE”, “EROS”, and “MACHO” refer to microlensing observations in our galaxy. In the case of binary inspiral, the horizontal axis refers to individual black hole masses with the assumption $m_1 = m_2 = m$. The shaded region marked “O2 result” is the exclusion region from the second LIGO-Virgo observing run. (Picture from <https://arxiv.org/pdf/1904.08976.pdf>.)

that prefer these higher masses. However, unlike for black holes lighter than $1 M_\odot$, there exist credible astrophysical mechanisms to create black holes of tens of solar masses from collapsing stars, so that in this case mass alone does not allow us to distinguish between formation mechanisms. In this regard we note that the future Einstein Telescope and Cosmic Explorer will be capable of observing binary inspirals out to redshifts $z \sim 20$, looking back to a time when stars were yet to form. Should a binary black hole merger be seen at such distances, then irrespective of masses, the black holes would for sure be primordial.

B. Towards a census of black hole spins

Apart from mapping out the distribution of black hole masses, an inroad is being made to do the same thing for spins. Also the latter will add a great deal to our understanding of how black holes come about.

Fig. 37 illustrates some of the physics of spins in binary black hole inspirals. For lack of time we will not be able to give mathematical details, but the basics can be summarized as follows. When one or both of the black hole spin vectors $\mathbf{S}_1, \mathbf{S}_2$ point in a different direction than the orbital angular momentum vector \mathbf{L}_{orb} , typically all three vectors will exhibit a “wobbling” motion. Since \mathbf{L}_{orb} is always perpendicular to the orbital plane, this means that this plane is no longer fixed, but itself will be “wobbling”. Now, as we have seen in Sec. VII, when the orbital plane is “face-on” with respect to the observer, the observed signal will contain both polarizations h_+ and h_\times in equal measure. If on the other hand the orbital plane is viewed “edge-on”, only h_+ is observed and the signal will be weaker. In Fig. 37 this translates into the fact that as the orbital plane wobbles about, sometimes it will be closer to “face-on” and sometimes closer to “edge-on”. The result is that the amplitude of the gravitational wave signal will no longer increase monotonically in time. Occasionally it will become larger or smaller depending on how the orientation of the orbital plane changes in time – the waveform will appear to be subject to *modulation*.

Both the orientation of the vectors $\mathbf{S}_1, \mathbf{S}_2$ with respect to \mathbf{L}_{orb} and their typical sizes can actually tell us about how binary black hole systems come into being:

- One mechanism assumes two ordinary but massive stars orbiting each other, and one after the other collapsing to a black hole. The resulting black holes will tend to have $\mathbf{S}_1, \mathbf{S}_2$ aligned with each other and with \mathbf{L}_{orb} . This is because tidal pulls between the preceding ordinary stars over many millions of years will have aligned the three vectors.
- In another mechanism, black holes form out of ordinary stars, but separately from each other. If this happens in a very dense stellar environment (*e.g.* the globular star clusters that are seen near every galaxy), then through random wandering the black holes might “find” each other and form a binary system which then undergoes coalescence. In this case $\mathbf{S}_1, \mathbf{S}_2$, and \mathbf{L}_{orb} can point in arbitrary directions with respect

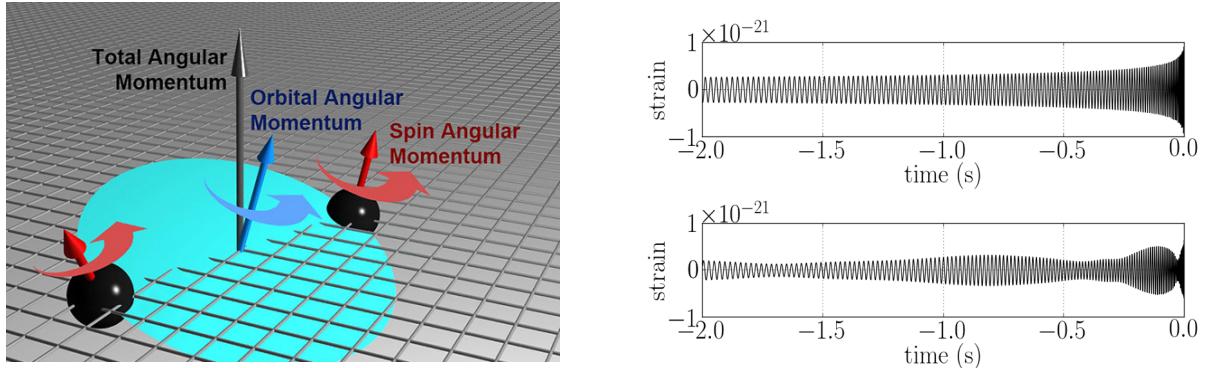


Figure 37: Left: When the spins of the black holes are misaligned with the orbital angular momentum, spin-orbit and spin-spin couplings can cause precession of the orbital plane (indicated in blue). Right: This can have a dramatic effect on the gravitational waveform that is emitted. In the absence of precession, the amplitude steadily increases (top), while with precession, there is a modulation of the wave (bottom).

to each other.

- Some mechanisms for the formation *primordial* black holes right after the Big Bang lead to black holes with zero spin. Hence, if one were to find *e.g.*, $\mathbf{S}_1 = 0$, then this would suggest a primordial origin for that black hole.

Unfortunately, gravitational wave detectors do not yet allow for accurate measurements of black hole spins. When spins are misaligned with orbital angular momentum, there will be amplitude modulation in the signal, but usually not as extreme as shown in the bottom right panel of Fig. 37, so that the modulation effect can easily get obscured by detector noise. Currently, most of the information about spins comes from the phase Φ_{gw} , where to leading order it occurs in the combinations $\mathbf{S}_1 \cdot \mathbf{L}_{\text{orb}}$ and $\mathbf{S}_2 \cdot \mathbf{L}_{\text{orb}}$. As a result, with current instruments we are mainly measuring the components of \mathbf{S}_1 , \mathbf{S}_2 along \mathbf{L}_{orb} , as illustrated in Fig. 38. To extract more information from the signals we will have to wait for the construction of better detectors.

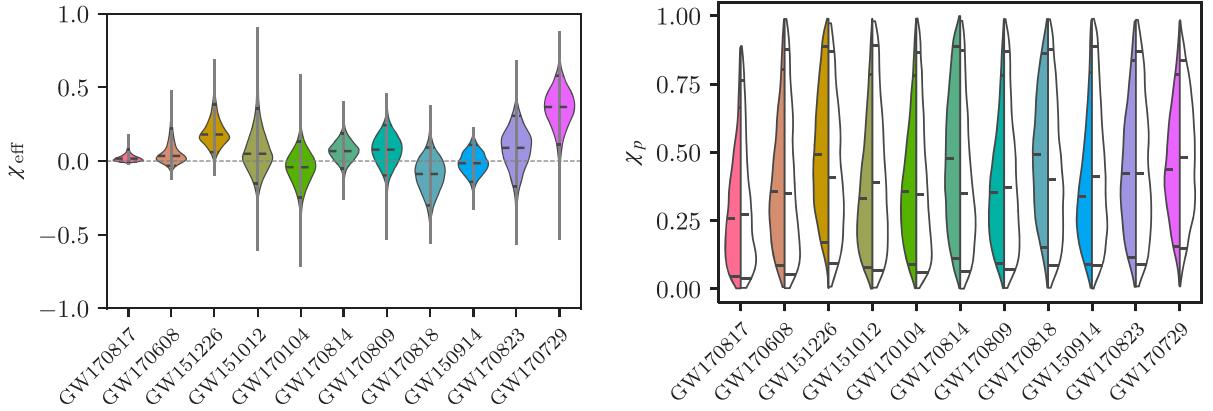


Figure 38: Left: Measurements of a particular combination of the components of the spin vectors along the direction of orbital angular momentum \mathbf{L}_{orb} , called χ_{eff} . The colored shapes are the probability distributions for χ_{eff} , rotated by 90° to easily fit them in a smaller plot. Right: Measurements of a combination of the components of the spin vectors along a direction perpendicular to \mathbf{L}_{orb} , called χ_p . In this case the probability distributions nearly fill the allowed range of the parameter, indicating that it is not well measurable. (Pictures from <https://arxiv.org/pdf/1811.12907.pdf>.)

XII. FUNDAMENTAL PHYSICS WITH GRAVITATIONAL WAVES

A. The fundamental physics of gravity

With the direct detection of gravitational waves, humankind gained its first access to the regime where general relativity really counts. Finally we are able to observationally study phenomena where spacetime curvatures are both strong and extremely dynamical. This is illustrated in Fig. 39, where a number of observations are compared with each other.

The very first success of general relativity was its explanation of the perihelion precession of Mercury, the process by which the point of closest approach of the planet to the Sun shifts over time, by $43''$ (arc-seconds) per century. More recent observations in the solar system include lunar laser ranging, whereby the orbit of the Moon around the Earth is studied by measuring the light travel time of a laser beam sent from Earth and bouncing off mirrors that were left on the Moon by Apollo astronauts. LAGEOS is a laser ranging satellite in orbit around the Earth at an altitude just below 6000 km, among other things meant to measure the Earth's shape, and thereby implicitly geographical variations of spacetime

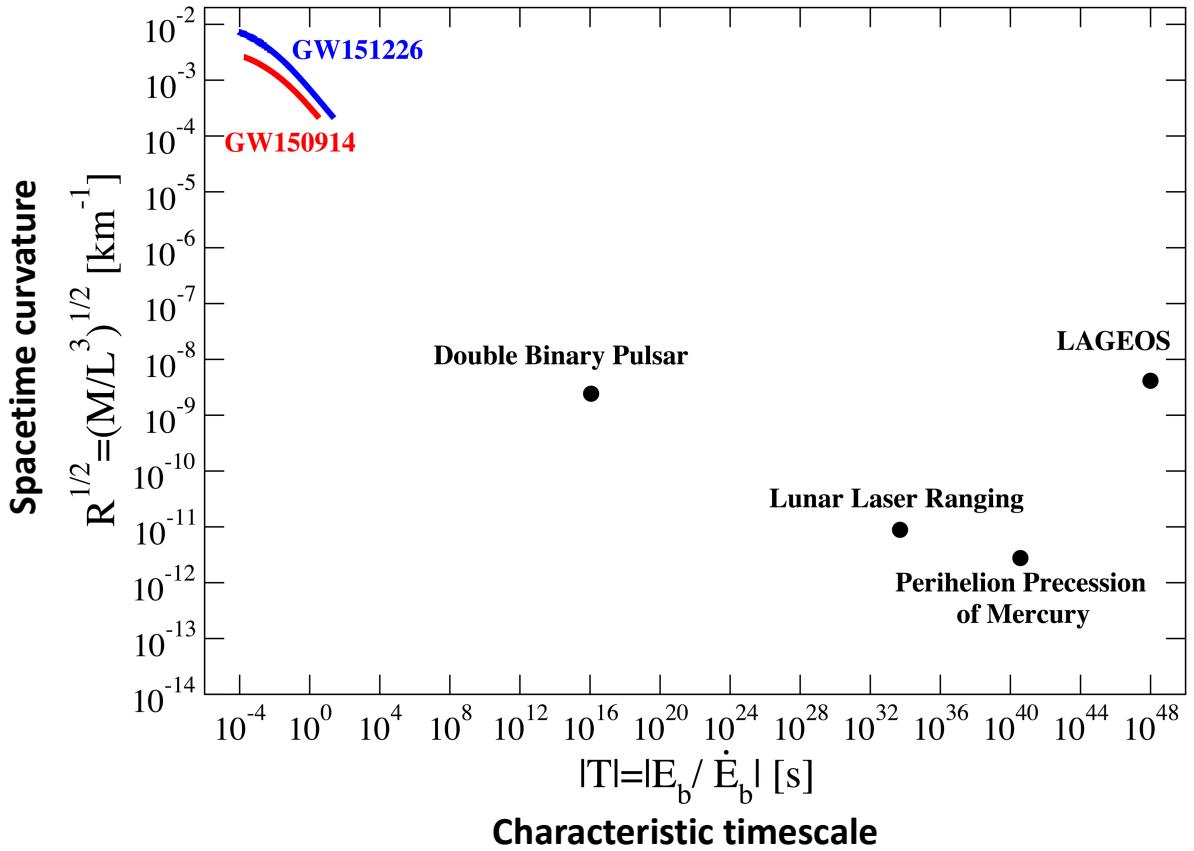


Figure 39: Various regimes of dynamical gravity, with some example observations. The horizontal axis indicates the characteristic timescale associated with an astrophysical process, while the vertical axis gives a measure for the strength of spacetime curvature involved. The solar system lives at bottom right, where spacetime curvature is weak; the long timescales show that the dynamics of spacetime are not very important there. Binary black hole mergers are all the way at top left, where spacetime curvatures are strong, and evolving over fractions of a second.

curvature near its surface. Until 2015, the most “general-relativistic” system known was the “double binary pulsar” PSR J07373039. This is a binary neutron star for which both components are seen as a pulsar (hence the “double”), orbiting around each other with a period of 2.4 hours; as with the Hulse-Taylor pulsar, the orbit slowly decays due to the emission of gravitational waves.

The binary black hole coalescences observed with LIGO and Virgo involve spacetime curvatures that are a million times stronger, evolving over timescales down to milliseconds. In addition, any matter that may initially have been around will have been swept up by the

black holes during the millions of years they existed as binaries. Hence their coalescences are pure spacetime processes. In physics it is rare that one finds processes that are this “clean”, with only one of the four interactions of nature involved, yet displaying a high level of complexity.

Indeed, the direct observation of binary black hole and neutron star mergers has enabled us to test the full non-linearity of general relativity. To understand what goes on in the regime where non-linear effects are important goes beyond the scope of these lectures, but we can at least give a flavor. The *post-Newtonian approximation* is a framework that not only goes beyond the quadrupole approximation, but also captures the non-linear aspects of general relativity in terms of perturbation theory. In particular, in this formalism all physical quantities of interest can be computed as expansions in powers of v/c , with v a characteristic velocity associated with the source. For example, the gravitational wave phase Φ_{gw} that we derived in Eq. (7.27) is really just the leading-order term in an expansion of the form

$$\boxed{\Phi_{\text{gw}}(v(t)) = \Phi_c + \left(\frac{v}{c}\right)^{-5} \sum_{k=0}^{\infty} \left[\varphi_k + \varphi_k^{(l)} \log\left(\frac{v}{c}\right) \right] \left(\frac{v}{c}\right)^k.} \quad (12.1)$$

Here the time dependence of Φ_{gw} comes about because v is monotonically increasing in time; in fact,

$$\frac{v(t)}{c} = \left(\frac{GM}{c^3} \pi f_{\text{gw}}(t) \right)^{1/3} \quad (12.2)$$

with M the total mass and $f_{\text{gw}}(t)$ the expression for the leading-order instantaneous gravitational wave frequency that we arrived at before, in Eq. (7.24). (We leave it as an exercise to show that the above expression can be understood in terms of Kepler’s third law.) Hence, the expansion parameter v/c is not constant in time. This may be unusual, but note that *at any instant in time* one has $v/c < 1$, so that it can be viewed as a “small parameter” at any point during the inspiral process.

The coefficients φ_k and $\varphi_k^{(l)}$ in Eq. (12.1) are specific functions of the masses m_1 , m_2 , and of the spin vectors \vec{S}_1 , \vec{S}_2 . If we were to restrict ourselves to the leading-order contributions (the ones with $k = 0$), then this would bring us back to the quadrupole approximation within the linearized theory. However, rich physics resides in the higher-order coefficients. For example:

- φ_3 is the lowest-order coefficient that encapsulates the fact that in full general relativity, the gravitational field is self-interacting.

- φ_3 also includes the leading-order spin-orbit interaction that leads to the precessional motions of Fig. 37.
- φ_4 is the lowest-order coefficient that includes spin-spin interactions.
- As is clear from Eq. (12.1), general relativity is incompatible with simple Taylor expansions. The lowest-order non-zero logarithmic contribution is $\varphi_5^{(l)}$.
- In the case of neutron stars, effects due to their finite size (*i.e.* the fact that they are not point particles) first enter in φ_{10} .

Since the dependences of the φ_k and $\varphi_k^{(l)}$ on m_1 , m_2 , \mathbf{S}_1 , and \mathbf{S}_2 are predictions of general relativity, one can put the theory to the test by allowing for possible deviations in these dependences. Given a coefficient φ_k , in the expression for the phase (12.1) one can replace

$$\varphi_k(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2) \longrightarrow (1 + \delta\hat{\varphi}_k) \varphi_k(m_1, m_2, \mathbf{S}_1, \mathbf{S}_2), \quad (12.3)$$

where $\delta\hat{\varphi}_k = 0$ if general relativity is correct. (The same can of course be done for the $\varphi_k^{(l)}$.) For each k , one then together measures the parameters³¹

$$\boldsymbol{\theta}' = (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \alpha, \delta, \psi, \iota, r, t_c, \Phi_c, \delta\hat{\varphi}_k). \quad (12.4)$$

and from the probability distribution $P_{\text{network}}(\boldsymbol{\theta}')$ one obtains a probability distribution $p(\delta\hat{\varphi}_k)$ for $\delta\hat{\varphi}_k$ by itself, by integrating out all the other parameters as usual.³²

The results for selected binary black hole coalescences are shown in Fig. 40, up to order $k_{\max} = 7$. Since the values $\delta\varphi_k = 0$ and $\delta\varphi_k^{(l)} = 0$ are in the support of all the probability distributions, general relativity is passing the entire battery of tests.

Many more tests of general relativity have been performed, which we won't have time to go into here; the interested reader can find the most important ones summarized in this paper: <https://arxiv.org/pdf/1903.04467.pdf>.

³¹ As before, all the parameters in the problem have to be measured together, since there will be correlations.

³² In principle one could allow for all the $\delta\hat{\varphi}_k$, $\delta\hat{\varphi}_k^{(l)}$ to be non-zero at the same time (up to some finite order k_{\max}) rather than one by one. However, this would introduce so much freedom in the waveform model that the resulting posterior density functions would be very wide and uninformative. When we allow for one kind of deviation at a time, this can be viewed as looking for a violation from general relativity that manifests itself the most strongly at some particular order.

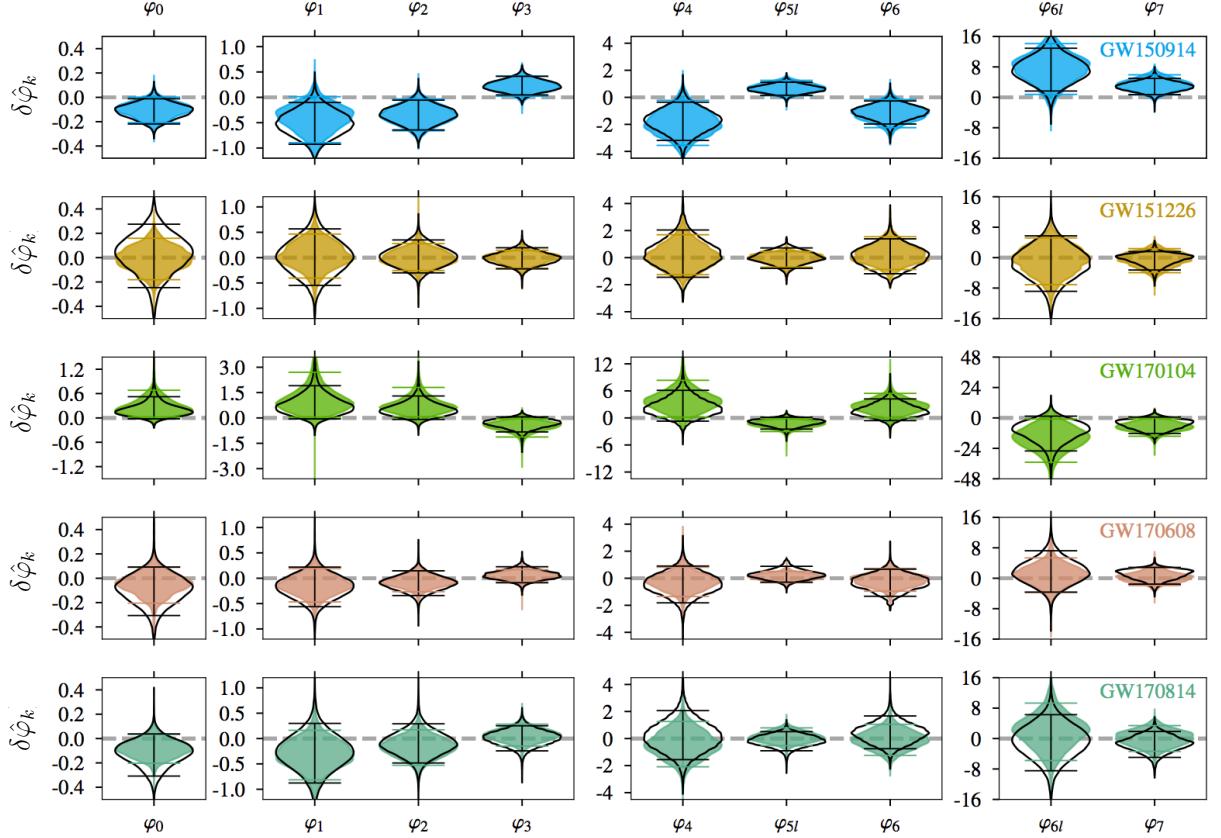


Figure 40: Measurements of possible deviations $\delta\hat{\varphi}_k$, $\delta\hat{\varphi}_k^{(l)}$ in the post-Newtonian coefficients for selected binary black hole coalescences, displayed in a similar style as that of Fig. 38. Note the different plotting ranges in the different panels; lower-order parameters tend to be more easily measurable so that they yield tighter probability distributions. The short horizontal bars on each of the distributions indicate 90% probability intervals. For all parameters, the value $\delta\hat{\varphi}_k = 0$ (or $\delta\hat{\varphi}_k^{(l)} = 0$), indicated by the thick dashed horizontal lines, is in the support of the probability distribution. Hence general relativity passes all of these tests. (Picture from <https://arxiv.org/pdf/1903.04467.pdf>.)

Before moving on, let us mention a particularly simple test of general relativity that could be performed because the binary neutron star coalescence that caused GW170817 was also seen as a gamma ray flash. As we have seen, according to general relativity the speed of gravitational waves, v_{GW} , is equal to the speed of light, c – but can this be tested observationally? The gamma rays were received 1.7 seconds after the end of the gravitational wave signal. One can now consider two extreme scenarios:

- The gamma rays were emitted at the same time as the end of the gravitational wave signal. To explain the 1.7 second delay in the arrival of gamma rays at Earth, gravitational waves would have had to travel slightly *faster* than the gamma photons.
- According to astrophysical models, gamma rays can have been emitted up to 10 seconds after the end of the gravitational wave signal. But since the gamma rays arrived at Earth only 1.7 seconds after the end of the gravitational wave, the gravitational waves would have had to travel slightly *slower* than the gamma photons.

From the gravitational wave signal as seen in three detectors, the distance to the source of GW170817 was determined to be larger than 26 Mpc with 90% confidence. Using this conservatively low distance estimate and considering the above extremes, the difference between the speed of gravitational waves and the speed of light was bounded to great accuracy as

$$-3 \times 10^{-15} \leq \frac{v_{\text{GW}} - c}{c} \leq +7 \times 10^{-16}. \quad (12.5)$$

B. Probing the physics of black holes

Until now we have mostly been focusing on the inspiral of compact objects. However, recall from Fig. 24 that after inspiral, the two objects will plunge towards each other, which in the case of two black holes leads to the formation of a single, highly excited black hole which undergoes “ringdown” as it sheds its excitations.

The gravitational wave signal from the ringdown of a newly formed black hole can be used to perform an indirect observational test of the famous *no-hair conjecture*. The latter states that the spacetime geometry of a stationary, electrically neutral black hole is completely determined by just two numbers: Its mass M and dimensionless spin a . Of course, stationary black holes can not be observed with gravitational waves, since they don’t emit any. However, a side-effect of the no-hair conjecture is that when a black hole is made to vibrate, a mixture of vibrational modes gets excited, with characteristic frequencies and damping times that

only depend on M and a . Specifically, the gravitational wave polarizations look like³³

$$h_{+,\times}(t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{n=0}^{\infty} \mathcal{A}_{\ell m n}^{+,\times} e^{-t/\tau_{\ell m n}} \cos(\omega_{\ell m n} t + \phi_{\ell m n}). \quad (12.6)$$

The amplitudes $\mathcal{A}_{\ell m n}^{+,\times}$ are set by the masses m_1, m_2 and spins $\mathbf{S}_1, \mathbf{S}_2$ of the original black holes, and the same is true of the phase offsets $\phi_{\ell m n}$ inside the cosines. However, the characteristic frequencies $\omega_{\ell m n}$ and damping times $\tau_{\ell m n}$ can only depend on the mass M_f and spin a_f of the final black hole:

$$\begin{aligned} \omega_{\ell m n} &= \omega_{\ell m n}(M_f, a_f), \\ \tau_{\ell m n} &= \tau_{\ell m n}(M_f, a_f), \end{aligned} \quad (12.7)$$

where the precise form of these functional dependences can be computed from general relativity. A way to observationally verify these relations is to make the following replacements in Eq. (12.6):

$$\begin{aligned} \omega_{\ell m n}(M_f, a_f) &\longrightarrow (1 + \delta\hat{\omega}_{\ell m n}) \omega_{\ell m n}(M_f, a_f), \\ \tau_{\ell m n}(M_f, a_f) &\longrightarrow (1 + \delta\hat{\tau}_{\ell m n}) \tau_{\ell m n}(M_f, a_f), \end{aligned} \quad (12.8)$$

where the $\delta\hat{\omega}_{\ell m n}$ and $\delta\hat{\tau}_{\ell m n}$ are additional parameters to be measured. If the measurements show consistency with $\delta\hat{\omega}_{\ell m n} = 0 = \delta\hat{\tau}_{\ell m n}$, then this can be viewed as an observational corroboration of the no-hair conjecture.

Fig. 41 shows the result of such a test using the ringdown signals of three binary black hole coalescences whose masses and SNRs were such that the ringdown signal was particularly well visible in the detectors, with a focus on the dominant vibrational mode, which is usually the one with $(\ell, m, n) = (2, 2, 0)$. We see that already with existing detectors, the no-hair conjecture is being verified with an accuracy of a few tens of percent; the future Einstein Telescope will enable tests at the sub-percent level.

Apart from the no-hair conjecture, there is also Hawking's *area increase theorem*, which states that when two black holes merge, the area of the horizon of the remnant black hole should not be smaller than the sum of the areas of the original black holes. This can be

³³ The (ℓ, m, n) are indices referring to an angular decomposition into differently shaped modes, somewhat similar to the orbitals of a hydrogen atom in quantum mechanics. However, the similarity is only a mathematical one, as the ringdown of a black hole is a purely classical phenomenon.

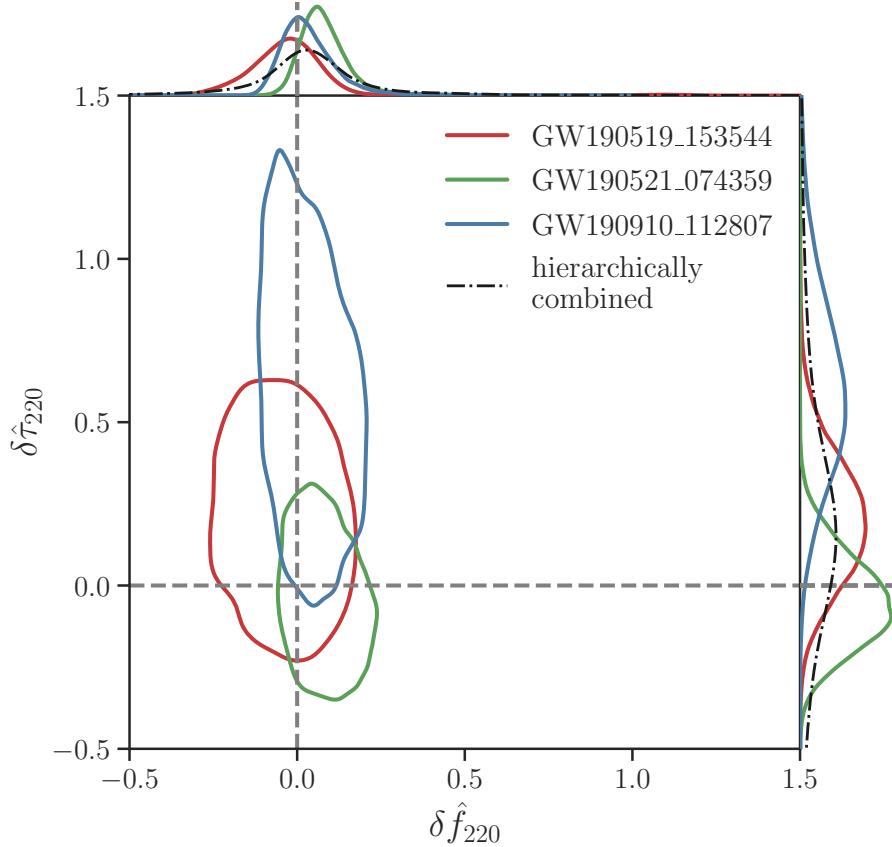


Figure 41: Measurements of deviations in the frequency and damping time of the $(\ell, m, n) = (2, 2, 0)$ mode in the ringdown signal of three binary black hole coalescences (where $\delta \hat{f}_{\ell mn} = \delta \hat{\omega}_{\ell mn} / (2\pi)$). The solid lines indicate 90% confidence contours of the posterior probability densities for $(\delta \hat{f}_{220}, \delta \hat{\tau}_{220})$, while the dashed lines indicate the values predicted by general relativity, $\delta \hat{f}_{220} = 0 = \delta \hat{\tau}_{220}$. (Picture from <https://arxiv.org/pdf/2010.14529.pdf>.)

tested with gravitational waves in the following way. From the inspiral one obtains the masses m_1 , m_2 and dimensionless spins χ_1 , χ_2 of the “ingoing” black holes. Early in the inspiral, the horizons of the ingoing black holes will have approximately looked like that of a stationary black hole. In terms of mass m and dimensionless spin χ , the horizon area of the latter is given by

$$\mathcal{A}(m, \chi) = 8\pi \left(\frac{Gm}{c^2} \right)^2 (1 + \sqrt{1 - \chi^2}). \quad (12.9)$$

On the other hand, the mass m_f and dimensionless spin χ_f of the final black hole can be measured from the ringdown signal. According to the area increase theorem,

$$\mathcal{A}(m_f, \chi_f) \geq \mathcal{A}(m_1, \chi_1) + \mathcal{A}(m_2, \chi_2). \quad (12.10)$$

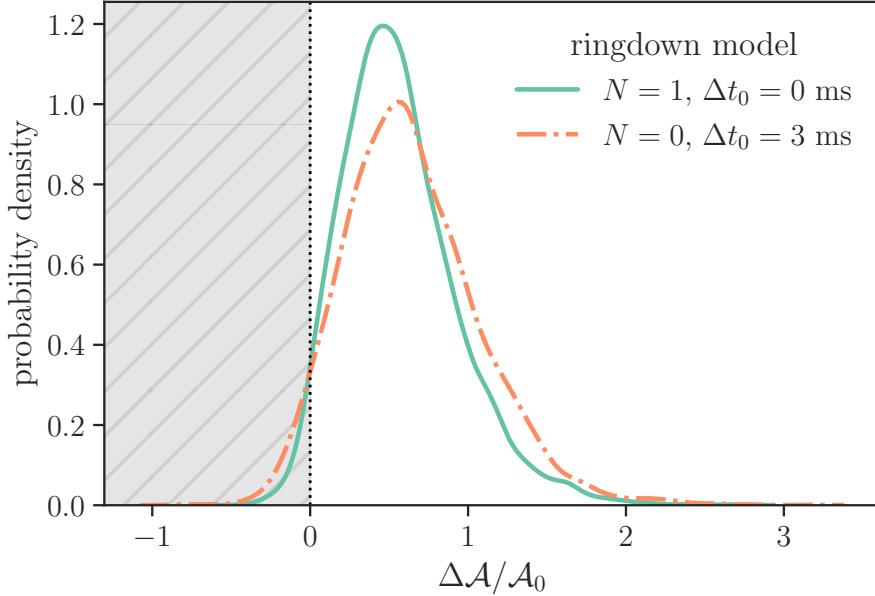


Figure 42: Posterior probability distribution for the relative change in horizon area $\Delta\mathcal{A}/\mathcal{A}_0$ during the binary black hole merger that emitted the signal GW150914. The shaded region is the one where Hawking’s area increase theorem is violated. The two curves are the posterior probabilities for measurements that make slightly different simplifying assumptions, but they are consistent with each other. In both cases, more than 95% of the probability resides to the right of the shaded region; hence the results are consistent with the theorem. (Picture from <https://arxiv.org/pdf/2012.04486.pdf>.)

Fig. 42 shows the result of such a test in terms of the fractional area increase $\Delta\mathcal{A}/\mathcal{A}_0$, where $\mathcal{A}_0 = \mathcal{A}(m_1, \chi_1) + \mathcal{A}(m_2, \chi_2)$ is the combined horizon area of the original black holes, and $\Delta\mathcal{A} = \mathcal{A}_f - \mathcal{A}_0$. The outcome is consistent with the area increase theorem with more than 95% confidence.

Thus, theorems and conjectures about black holes that were developed in the 1960s and 1970s, and which until very recently were considered to be exclusively in the domain of mathematical physics, are now being subjected to empirical tests.

C. What do neutron stars look like on the inside?

As the name suggests, neutron stars are assumed to consist mostly of neutrons. Nevertheless, to understand them, input is needed from nearly all branches of physics. The neutron

“fluid” is expected to be in a superfluid state. Some regions of neutron stars are also predicted to be superconductors for the stray protons that will still be present. From radio observations we know that neutron stars have enormously strong magnetic fields ($\sim 10^{15}$ Gauss), but where these come from is not quite understood. Fig. 43 is meant to give an impression of the complicated neutron star interior.

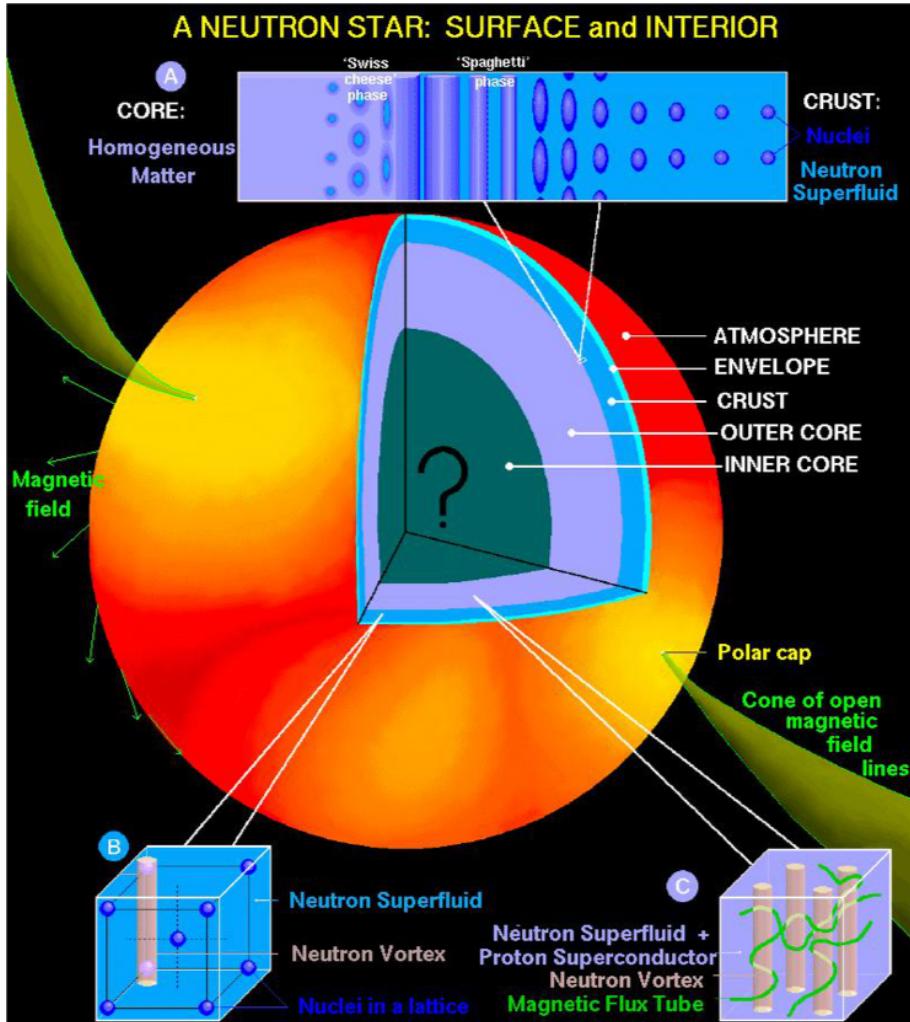


Figure 43: The interiors of neutron stars involve rich physics, making it difficult to model in detail what goes on inside.

As a result of this complexity, it turns out to be difficult to make theoretical models of neutron stars. To illustrate the seriousness of the problem, suppose one asks the question “given a neutron star of mass m , what is its radius?” Fig. 44 shows a mass-radius diagram for neutron stars, in which the curves are a selection of qualitatively different predictions coming from different theoretical models.

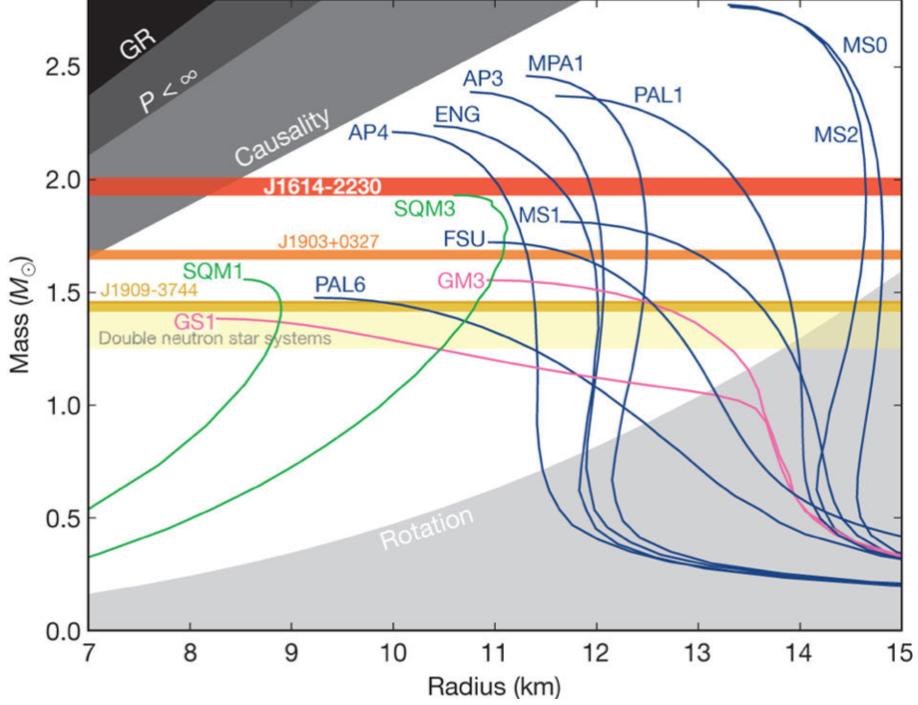


Figure 44: Mass versus radius for neutron stars. The curves illustrate the differing predictions from different theoretical models. The horizontal bars indicate mass measurements of neutron stars in binary systems using radio observations; most predicted mass-radius curves are consistent with these.

In practice, modeling neutron stars comes down to predicting their *equation of state*: pressure P as function of density ρ . Different neutron star models predict different relations $P(\rho)$, which in turn translate to different relations between mass and radius.

Gravitational wave observations of binary neutron star inspirals are expected to make major contributions towards measuring the equation of state. When two inspiraling neutron stars get close to each other, they undergo *tidal deformations*; see Fig. 45. Schematically, neutron star 2 exerts a tidal field (a tensor) $\mathcal{E}_{ij}^{(2)}$ on neutron star 1, causing the latter to develop a quadrupole moment given by

$$M_{ij}^{(1)} = -\lambda(\text{EOS}; m_1) \mathcal{E}_{ij}^{(2)}, \quad (12.11)$$

and similarly neutron star 1 causes a quadrupole deformation in neutron star 2:

$$M_{ij}^{(2)} = -\lambda(\text{EOS}; m_2) \mathcal{E}_{ij}^{(1)}. \quad (12.12)$$

The proportionality factor $\lambda(\text{EOS}, m)$ – called the *tidal deformability* – is a function of the mass m of the star that’s being deformed, and this function is determined by the equation

of state (EOS). This is as one would expect: how deformable neutron stars are depends on their internal structure.

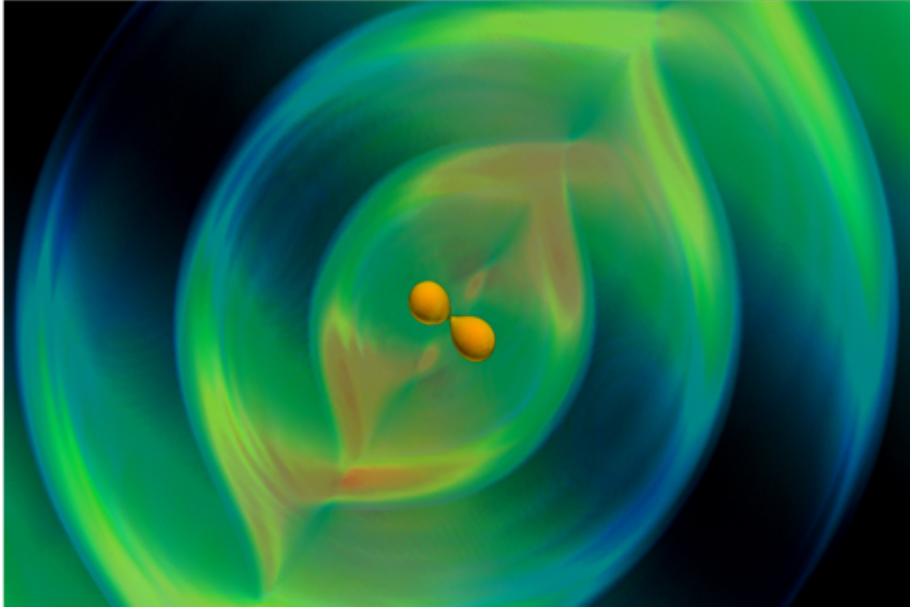


Figure 45: When two neutron stars get close to each other, they tidally deform one another, which leaves an imprint on the gravitational wave signal. How deformable neutron stars are is determined by their internal structure; hence gravitational wave observations are a way to probe the neutron star equation of state.

Thus, taking into account tidal deformations, there are three quadrupole moments involved in binary neutron star inspiral: the orbital quadrupole moment $M_{ij}^{(\text{orb})}$ associated with the orbital motion of the centers of mass of the two neutron stars, and $M_{ij}^{(1)}$, $M_{ij}^{(2)}$ associated with their deformations. Since the tidally induced “bulge” of each neutron star rotates as the neutron stars move around each other, the latter two quadrupole moments are time dependent. Hence, they cause gravitational wave emission in addition to what we computed earlier for the point particle case. The strengths of the tidal fields $\mathcal{E}_{ij}^{(2)}$, $\mathcal{E}_{ij}^{(1)}$, and hence the sizes of the quadrupoles $M_{ij}^{(1)}$, $M_{ij}^{(2)}$ will depend on how close the neutron stars are to each other, and thus (through Kepler’s third law) on f_{gw} , and eventually on v/c via Eq. (12.2). A careful calculation shows that tidal effects enter the expansion of the phase, Eq. (12.1), through the coefficient φ_{10} . This means that they only become apparent at order $(v/c)^{10}$ beyond the leading order, at first suggesting that they may not be measurable. However, it turns out that the tidal contribution to φ_{10} can be quite large: not $\mathcal{O}(1)$, but $\mathcal{O}(10^2 - 10^5)$

depending on the equation of state. If one is able to measure $M_{ij}^{(1,2)}$ and hence $\lambda(\text{EOS}, m_{1,2})$ for many different binary neutron star inspirals with different component masses m_1, m_2 , one will be mapping out the function $\lambda(\text{EOS}; m)$, and hence the equation of state $P(\rho)$.

A first measurement of tidal deformations was done with the binary neutron star signal GW170817. The tidal deformabilities $\lambda_1 = \lambda(\text{EOS}; m_1)$, $\lambda_2 = \lambda(\text{EOS}; m_2)$ are extra parameters in the measurement, the full set of parameters now being

$$\boldsymbol{\theta''} = (m_1, m_2, \mathbf{S}_1, \mathbf{S}_2, \alpha, \delta, \psi, \iota, r, t_c, \Phi_c, \lambda_1, \lambda_2). \quad (12.13)$$

Fig. 46 shows the resulting posterior probability density for (λ_1, λ_2) .³⁴ Even with this one measurement, a few of the more extreme equation of state predictions could be excluded. Generally speaking, the result favors equations of state that cause neutron stars to be more compact (smaller radius for a given mass) than what some models had predicted.

These results should be considered only the beginning; as more binary neutron star signals are detected, we can start probing the function $\lambda(\text{EOS}; m)$, and hence *measuring* the equation of state.

³⁴ More precisely, dimensionless quantities $\Lambda_i = \lambda_i/m_i^5$, $i = 1, 2$ are shown.

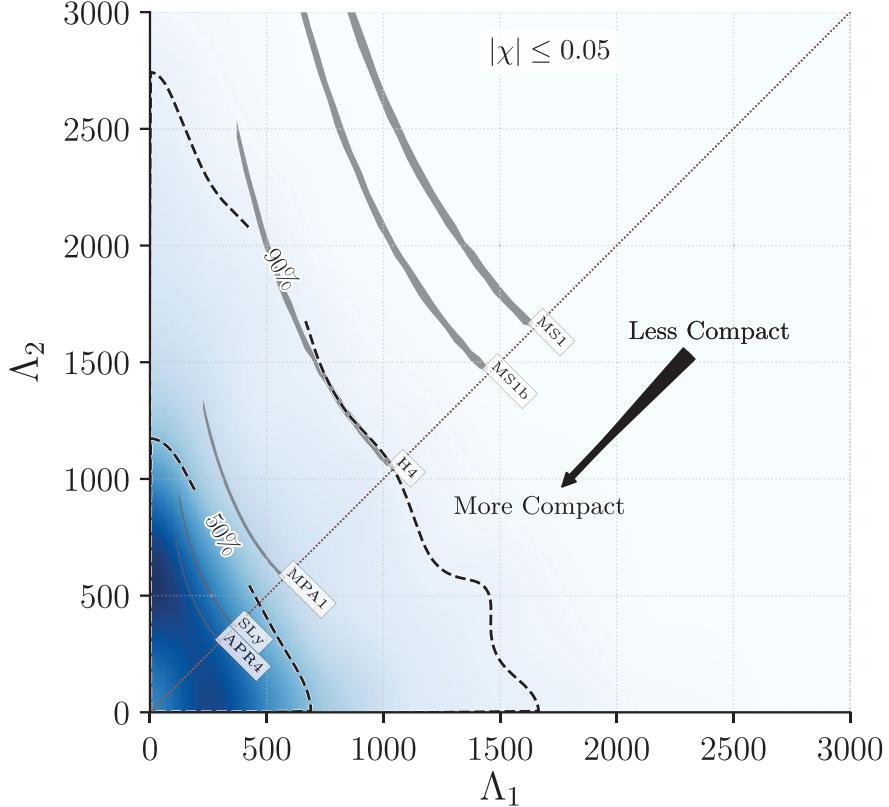


Figure 46: Measurement of neutron star tidal deformations with GW170817. The blue shading indicates the probability density, and the dashed contours enclose 50% and 90% of total probability. Overlaid are predictions from particular equations of state, some of which (the ones labeled MS1, MS1b) are already excluded with this one measurement. The result favors equations of state that lead to more compact neutron stars. (Picture from <https://arxiv.org/pdf/1710.05832.pdf>.)

XIII. GRAVITATIONAL WAVES AND COSMOLOGY

In previous discussions we have implicitly assumed that sources of gravitational waves detectable by LIGO and Virgo are sufficiently close-by that cosmological effects can be neglected. However, this is not quite the case; for example, the binary black hole signal GW170729 originated at a redshift $z \simeq 0.5$, which is not that far away in the grand scheme of things, but also not in our immediate cosmic neighborhood. Gravitational waves are affected by the expansion of the Universe, in ways we now describe.

Recall that we started our investigations into gravitational waves by assuming that for weak gravitational fields, the spacetime metric $g_{\mu\nu}$ could be written as the flat metric $\eta_{\mu\nu}$

with some small perturbation $h_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1. \quad (13.1)$$

To understand how gravitational waves propagate in a homogeneous, isotropic, expanding Universe, we can instead write

$$g_{\mu\nu} = g_{\mu\nu}^{\text{BG}} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1, \quad (13.2)$$

where this time the “background metric”, $g_{\mu\nu}^{\text{BG}}$, will be taken to be the (spatially flat) Friedmann-Robertson-Walker metric, whose line element we encountered in Eq. (3.28):

$$ds^2 = -c^2 dt^2 + a^2(t) [dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2)], \quad (13.3)$$

In principle we should now substitute the form (13.2) of the metric into the full Einstein equations, again retaining only terms linear in $h_{\mu\nu}$ and its derivatives, which would lead to an appropriately modified form of the linearized Einstein equations that we obtained earlier. Fortunately, intuition provides some shortcuts.

First note that the variable r in our expressions for gravitational waves from inspiraling binaries, Eqs. (7.28), corresponded to the spatial distance between source and observer according to the metric $\eta_{\mu\nu}$. In the case of the metric (13.3) we have to be more careful, since r is merely a coordinate. Consider a galaxy at $r = r_1$ and another galaxy at $r = r_2$ (but with the same values of θ, ϕ). Then the *coordinate distance* between the two is $r_2 - r_1$, but the *physical distance* r_{phys} is obtained from the metric. For distances in a purely radial direction we read off from (13.3) that

$$dr_{\text{phys}}^2 = a^2(t) dr^2. \quad (13.4)$$

If the first galaxy is at the origin and the second one at radial coordinate r , then

$$r_{\text{phys}}(t) = a(t) \int_0^r dr = a(t) r. \quad (13.5)$$

Now consider a source located at radial coordinate r which emit signals (electromagnetic, or gravitational waves) that travel at the speed of light, and are later received by an observer located at $r = 0$. Let a wave crest be emitted at a time t_{em} and detected by the observer at the later time t_{obs} . Setting $ds^2 = 0$ in Eq. (13.3), we get

$$c^2 dt^2 = a^2(t) dr^2 \quad (13.6)$$

so that

$$\int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{ct}{a(t)} = \int_0^r dr. \quad (13.7)$$

Now let a second wave crest be emitted at time $t_{\text{em}} + \Delta t_{\text{em}}$, and observed at time $t_{\text{obs}} + \Delta t_{\text{obs}}$.

Then

$$\int_{t_{\text{em}} + \Delta t_{\text{em}}}^{t_{\text{obs}} + \Delta t_{\text{obs}}} \frac{ct}{a(t)} = \int_0^r dr. \quad (13.8)$$

Since the right hand side in (13.7) and (13.8) is the same, the difference of the two equations gives

$$\begin{aligned} 0 &= \left(\int_{t_{\text{em}} + \Delta t_{\text{em}}}^{t_{\text{obs}} + \Delta t_{\text{obs}}} - \int_{t_{\text{em}}}^{t_{\text{obs}}} \right) \frac{ct}{a(t)} \\ &= \int_{t_{\text{obs}}}^{t_{\text{obs}} + \Delta t_{\text{obs}}} \frac{ct}{a(t)} - \int_{t_{\text{em}}}^{t_{\text{em}} + \Delta t_{\text{em}}} \frac{ct}{a(t)} \\ &\simeq \frac{c\Delta t_{\text{obs}}}{a(t_{\text{obs}})} - \frac{c\Delta t_{\text{em}}}{a(t_{\text{em}})} \end{aligned} \quad (13.9)$$

where in the last line we have assumed Δt_{em} and Δt_{obs} to be small. From this we find

$$\Delta t_{\text{obs}} = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})} \Delta t_{\text{em}}. \quad (13.10)$$

This leads us to the concept of the *redshift* z of the source due to the expansion of the Universe:

$$1 + z = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})}. \quad (13.11)$$

Eq. (13.10) implies that the time t_{obs} measured by the observer's clock and the time t_s measured by the source's clock are related by

$$dt_{\text{obs}} = (1 + z) dt_s. \quad (13.12)$$

As a consequence, any frequency measured by the observer, f_{obs} , is related to the corresponding frequency in the source frame, f_s , by

$$f_{\text{obs}} = \frac{f_s}{1 + z}, \quad (13.13)$$

and wavelengths are related by

$$\lambda_{\text{obs}} = (1 + z) \lambda_s. \quad (13.14)$$

In quantum mechanics, frequency and energy are related through Planck's constant h as $E = hf$, so that for energies

$$E_{\text{obs}} = \frac{E_s}{1 + z}. \quad (13.15)$$

Finally, let \mathcal{F} be the energy flux (energy per unit time per unit area) measured in the observer's frame, and let \mathcal{L} be the luminosity of the source, *i.e.* the energy per unit time that it radiates in its rest frame:

$$\mathcal{L} = \frac{dE_s}{dt_s}. \quad (13.16)$$

The luminosity distance d_L is then defined as

$$\boxed{\mathcal{F} = \frac{\mathcal{L}}{4\pi d_L^2}}. \quad (13.17)$$

In the absence of cosmic expansion, the energy E_{obs} measured in the observer's frame is the same as the energy E_s measured in the source frame, and $dt_{\text{obs}} = dt_s$, so

$$\frac{dE_{\text{obs}}}{dt_{\text{obs}}} = \frac{dE_s}{dt_s}. \quad (13.18)$$

When it arrives at the detector, the energy of an isotropically emitting source will have spread out over an area $A = 4\pi r^2$, so that

$$\begin{aligned} \mathcal{F} &= \frac{1}{4\pi r^2} \frac{dE_{\text{obs}}}{dt_{\text{obs}}} \\ &= \frac{1}{4\pi r^2} \frac{dE_s}{dt_s} \\ &= \frac{\mathcal{L}}{4\pi r^2}. \end{aligned} \quad (13.19)$$

Hence, in a non-expanding Universe, d_L is just equal to the distance r to the source. However, in an expanding Universe the observed energy E_{obs} is redshifted as in Eq. (13.15), and observed time intervals dt_{obs} as in Eq. (13.12). As a consequence,

$$\frac{dE_{\text{obs}}}{dt_{\text{obs}}} = \frac{1}{(1+z)^2} \frac{dE_s}{dt_s}. \quad (13.20)$$

Furthermore, using the metric (13.3) we find that at time t , the surface of a sphere with coordinate radius r is $4\pi a^2(t) r^2$, so that when the radiation arrives at the detector it is spread over an area $A = 4\pi a^2(t_{\text{obs}}) r^2$. Therefore

$$\mathcal{F} = \frac{\mathcal{L}}{4\pi (1+z)^2 a^2(t_{\text{obs}}) r^2}. \quad (13.21)$$

From this we read off that the luminosity distance is

$$\boxed{d_L = (1+z) a(t_{\text{obs}}) r.} \quad (13.22)$$

An important point about the relationship between luminosity distance and redshift is that it tells us about the expansion of the Universe. To see this, write Eq. (13.11) as

$$1 + z(t) = \frac{1}{a(t)}, \quad (13.23)$$

where for brevity we set $t_{\text{em}} = t$, and by convention, $a(t_{\text{obs}}) = 1$. Taking the time derivative:

$$\frac{dz}{dt} = -\frac{1}{a^2} \frac{da}{dt} = -\frac{1}{a} \frac{\dot{a}}{a} = -\frac{H}{a}, \quad (13.24)$$

where $H \equiv \dot{a}/a$ is called the *Hubble parameter*. From Eq. (13.7),

$$r = \int_{t_{\text{em}}}^{t_{\text{obs}}} \frac{c dt}{a(t)} = - \int_z^0 \frac{c dz}{H(z)} = c \int_0^z \frac{dz}{H(z)}, \quad (13.25)$$

and together with Eq. (13.22) for the luminosity distance this yields

$$d_L = c(1+z) \int_0^z \frac{dz}{H(z)}. \quad (13.26)$$

For small redshifts $z \ll 1$, this becomes

$$d_L \simeq \frac{c}{H_0} z, \quad (13.27)$$

where $H_0 = H(0)$ is the *Hubble constant*. Given a galaxy with redshift z , the speed with which it moves away from us due to the expansion of the Universe is $v = cz$, so that

$$v = H_0 d_L. \quad (13.28)$$

This is *Hubble's law*: the recession speed of a galaxy is proportional to how distant it is from us, and the proportionality factor is the Hubble constant, which tells us about the current expansion rate of the Universe. There are various ways of measuring the Hubble constant, but they disagree in a statistically significant way; this is the famous Hubble constant tension.³⁵

Now let us revisit the expressions for h_+ , h_\times that we derived for binary inspiral assuming a *flat* Universe, Eq. (7.28). If we are very close to the source, then the expansion of the Universe is unimportant since the scale factor $a(t)$ does not change appreciably during the propagation of the wave. In this “local zone”, physical distances can be written as

³⁵ See e.g. <https://skyandtelescope.org/astronomy-news/tension-continues-hubble-constant/>.

$r_{\text{phys}} = a(t_{\text{em}}) r$, where r is the radial coordinate and t_{em} some emission time. Hence, close to the source we can use the expressions (7.28), with 1/distance taken to be r_{phys} :

$$\begin{aligned} h_+ &= h_c(t_s^{\text{ret}}) \frac{1 + \cos^2 \iota}{2} \cos \left[2\pi \int^{t_s^{\text{ret}}}_0 dt'_s f_{\text{gw}}^{(s)}(t'_s) \right], \\ h_\times &= h_c(t_s^{\text{ret}}) \cos \iota \sin \left[2\pi \int^{t_s^{\text{ret}}}_0 dt'_s f_{\text{gw}}^{(s)}(t'_s) \right], \end{aligned} \quad (13.29)$$

where

$$h_c(t_s^{\text{ret}}) = -\frac{4}{a(t_{\text{em}}) r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(s)}(t_s^{\text{ret}})}{c} \right)^{2/3}. \quad (13.30)$$

Here t_s is the time measured by a clock near the source, and $t_s^{\text{ret}} = t_s - r_{\text{phys}}/c$ the retarded time associated with the small physical distance traveled by the wave. The frequency $f_{\text{gw}}^{(s)}$ refers to the source, and takes the form that we derived earlier (7.24):

$$f_{\text{gw}}^{(s)}(\tau_s) = \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{-5/8} \left(\frac{5}{256} \frac{1}{\tau_s} \right)^{3/8}, \quad (13.31)$$

with τ_s the time to coalescence measured by a clock at the source.

Now we want to find out what the signal will look like to a-far away observer. First, in Eq. (13.30), $4/(a(t_{\text{emis}}) r)$ one replaces $4/(a(t_{\text{em}}) r)$ by $4/(a(t_{\text{obs}}) r)$. Next, using Eqs. (13.12) and (13.13), one has

$$\int^{t_s^{\text{ret}}}_0 dt'_s f_{\text{gw}}^{(s)}(t'_s) = \int^{t_{\text{obs}}^{\text{ret}}}_0 dt'_{\text{obs}} f_{\text{gw}}^{(\text{obs})}(t'_{\text{obs}}), \quad (13.32)$$

because the redshift in dt cancels the redshift in f . Writing

$$f_{\text{gw}}^{(s)} = (1+z) f_{\text{gw}}^{(\text{obs})}, \quad (13.33)$$

Eq. (13.30) becomes

$$\begin{aligned} h_c(t_{\text{obs}}^{\text{ret}}) &= -\frac{4}{a(t_{\text{obs}}) r} (1+z)^{2/3} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3} \\ &= -\frac{4}{d_L(z)} (1+z)^{5/3} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3}, \end{aligned} \quad (13.34)$$

where in the second line we used our expression for luminosity distance, Eq. (13.22). Defining

$\mathcal{M}_c^z \equiv (1+z) \mathcal{M}_c,$

(13.35)

we finally get

$$h_c(t_{\text{obs}}^{\text{ret}}) = \frac{4}{d_L(z)} \left(\frac{G\mathcal{M}_c^z}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}})}{c} \right)^{2/3}. \quad (13.36)$$

Hence the gravitational wave amplitude takes the same form as in the absence of cosmological expansion, with the replacements $r \rightarrow d_L$ and $\mathcal{M}_c \rightarrow \mathcal{M}_c^z$. The dependence of $f_{\text{gw}}^{(\text{obs})}$ on $t_{\text{obs}}^{\text{ret}}$ is given by

$$\begin{aligned} f_{\text{gw}}^{(\text{obs})}(t_{\text{obs}}^{\text{ret}}) &= \frac{1}{1+z} f_{\text{gw}}^{(\text{s})}(t_s^{\text{ret}}) \\ &= \frac{1}{1+z} \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-5/8} \left(\frac{5}{256} \frac{1}{\tau_s} \right)^{3/8} \\ &= \frac{1}{1+z} \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-5/8} \left(\frac{5}{256} \frac{1+z}{\tau_{\text{obs}}} \right)^{3/8} \\ &= \frac{1}{\pi} \left(\frac{G\mathcal{M}_c^z}{c^3} \right)^{-5/8} \left(\frac{5}{256} \frac{1}{\tau_{\text{obs}}} \right)^{3/8}, \end{aligned} \quad (13.37)$$

where $\tau_{\text{obs}} = (1+z)\tau_s$ is the time to coalescence measured by the observer's clock. Hence also the dependence of the observed frequency on the observer's time is formally the same as in the case $z=0$, on condition that we replace $\mathcal{M}_c \rightarrow \mathcal{M}_c^z$.

Putting everything together, the gravitational wave polarizations become

$$\begin{aligned} h_+(t_{\text{obs}}^{\text{ret}}) &= h_c(t_{\text{obs}}^{\text{ret}}) \frac{1 + \cos^2 \iota}{2} \cos(\Phi(t_{\text{obs}}^{\text{ret}})), \\ h_\times(t_{\text{obs}}^{\text{ret}}) &= h_c(t_{\text{obs}}^{\text{ret}}) \cos \iota \sin(\Phi(t_{\text{obs}}^{\text{ret}})), \end{aligned} \quad (13.38)$$

where h_c is as specified in Eq. (13.36), and

$$\Phi(t_{\text{obs}}^{\text{ret}}) = 2\pi \int^{t_{\text{obs}}^{\text{ret}}} dt'_s f_{\text{gw}}^{(\text{obs})}(t'_s). \quad (13.39)$$

In summary, the signal received by an observer from a binary inspiral at redshift z is modified in the following ways compared to the case $z=0$:

1. The overall prefactor $1/r$ in the amplitude is replaced by $1/d_L(z)$.
2. The chirp mass \mathcal{M}_c is replaced by a “redshifted” chirp mass \mathcal{M}_c^z .
3. The observed frequency is redshifted with respect to the frequency in the source frame:

$$f_{\text{gw}}^{(\text{obs})} = f_{\text{gw}}^{(\text{s})}/(1+z).$$

Point 2 is in fact easy to understand. As one would expect, the evolution of frequency as a function of time is determined by the characteristic timescale in the problem, which is $G\mathcal{M}/c^3$. Because of the expansion of the universe, this timescale is redshifted, $G\mathcal{M}/c^3 \rightarrow (1+z)G\mathcal{M}/c^3$, which corresponds to the replacement $\mathcal{M}_c \rightarrow \mathcal{M}_c^z$.

Point 3 also has consequences for the observed ending frequency of a signal. We have seen that the inspiral ends at $f_{\text{gw,ISCO}} = c^3/(6^{3/2}\pi GM)$. Note that the total mass $M = \mathcal{M}_c\eta^{-3/5}$, where $\eta = m_1m_2/M^2$ is a dimensionless quantity. Hence also $M \rightarrow M^z = (1+z)M$, which means that signals from sources at higher redshift will terminate earlier in the detector frame. For example, a binary neutron star inspiral with $(m_1, m_2) = (1.4, 1.4) M_\odot$ which is very close-by has $f_{\text{gw,ISCO}} \simeq 1600$ Hz, but for the same inspiral at $z = 1$, the termination frequency in the detector will be $f_{\text{gw,ISCO}}^{(\text{obs})} \simeq 800$ Hz.

Finally, we mentioned before that when going beyond the quadrupole approximation, the waveform depends on m_1 and m_2 separately; both get redshifted to $m_1^z = (1+z)m_1$, $m_2^z = (1+z)m_2$. In an expanding Universe, the observable quantities for binary coalescence are (apart from finite size effects)

$$\boldsymbol{\theta} = (m_1^z, m_2^z, \mathbf{S}_1, \mathbf{S}_2, \alpha, \delta, \psi, \iota, d_L, t_c, \Phi_c). \quad (13.40)$$

Note that the redshift can not be obtained from a gravitational wave signal alone, because it is “hidden” inside the parameters m_1^z , m_2^z , and d_L ; there is no separate dependence on z .

An important point which we have been glossing over is how (luminosity) distance can be measured in the first place from a gravitational wave signal alone. From Fig. 28 it is intuitively clear how a three-detector network allows us to infer sky position, but how is distance extracted?

Consider the responses of three detectors that are differently oriented and located:

$$\begin{aligned} h_1 &= F_+^{(1)}(\alpha, \delta, \psi) h_+ + F_\times^{(1)}(\alpha, \delta, \psi) h_\times, \\ h_2 &= F_+^{(2)}(\alpha, \delta, \psi) h_+ + F_\times^{(2)}(\alpha, \delta, \psi) h_\times, \\ h_3 &= F_+^{(3)}(\alpha, \delta, \psi) h_+ + F_\times^{(3)}(\alpha, \delta, \psi) h_\times. \end{aligned} \quad (13.41)$$

In the above we assume that the detectors’ beam pattern functions $F_{+, \times}^{(k)}$, $k = 1, 2, 3$ have been re-expressed in terms of the sky position (α, δ) and the polarization angle ψ with respect to the frame of detector 1; see also the discussion in the previous chapter. As explained before, with three detectors operational one can infer the sky position (α, δ) with

reasonable accuracy, so let us assume this is known. Then the above are three equations for three unknowns: ψ , h_+ , and h_\times , which we can solve for. Next, the evolution of the phase as a function of time of either h_+ or h_\times , or equivalently the frequency as a function of time (Eq. (13.31)), is governed by \mathcal{M}_c^z , so the latter can be inferred. Furthermore, from Eq. (13.38), the ratio of the amplitudes of h_+ and h_\times is $(1 + \cos^2 \iota)/(2 \cos \iota)$, from which we obtain the inclination angle ι . Hence we know the common factor in the amplitudes, $h_c(t_{\text{obs}}^{\text{ret}})$. Looking at Eq. (13.36), we conclude that we can solve for d_L .

In a nutshell, the reason why we can infer the luminosity distance from the gravitational wave signal itself is that with a network of detectors we can infer the angles in the problem, and the frequency evolution of the signal gives us the chirp mass. The only other quantity appearing in the amplitude is $1/d_L$.

Thus, gravitational wave signals from binary inspirals provide us with a way to infer luminosity distance d_L . Because of the (vague) analogy of gravitational waves with sound waves, they are called *standard sirens*. If we now also separately had a way to obtain the redshift z , then with a sufficient number of sources we could make a fit of the function $d_L(z)$; using Eq. (13.27) this would lead to a measurement of H_0 . As we have seen, the value of H_0 is a hot topic: despite that it is arguably the most important constant in cosmology, existing methods significantly differ on its precise value. It would then be good to have a completely independent method for measuring it.

As we have seen, it is not possible to separately get the redshift z out of a gravitational wave signal. However, binary neutron star mergers can have detectable electromagnetic counterparts, and indeed the optical afterglow of GW170817 was found to be near the galaxy NGC 4993. By looking at the spectrum of that galaxy, the redshift was known to be $z = 0.009727$. On the other hand, the distance d_L inferred from the gravitational wave signal was $d_L = 43.8^{+2.9}_{-6.9}$ Mpc. Note the large uncertainty; as always, the detector responses h_1, h_2, h_3 were embedded in noise, which was ignored in the heuristic argument above. Now recall Eq. (13.27):

$$d_L \simeq \frac{c}{H_0} z. \quad (13.42)$$

The probability distribution for d_L inferred from the gravitational wave observations, together with the precise redshift measurement obtained via electromagnetic means, yields a probability distribution for H_0 , which is shown in Fig. 47.

Since z is known with high precision, almost all the uncertainty in the measurement of

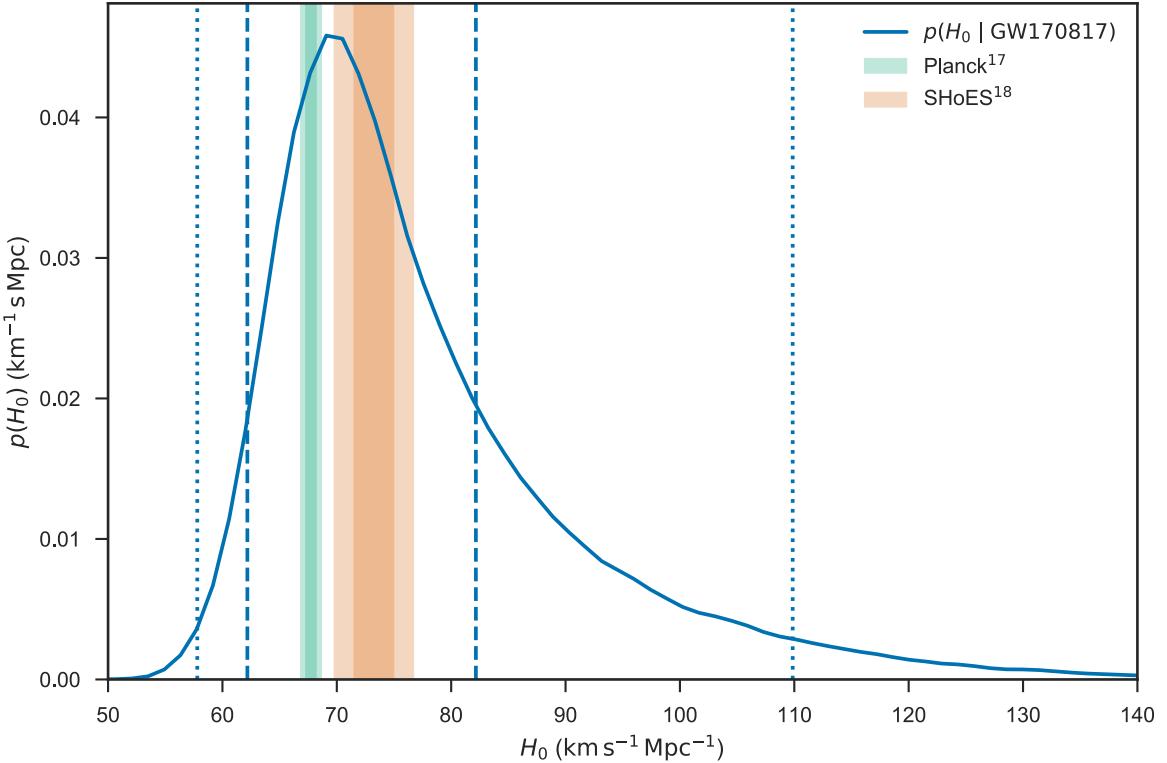


Figure 47: The probability distribution for the Hubble parameter H_0 obtained from the binary neutron star inspiral GW170817 (blue curve). The dashed and dotted vertical lines indicate 1σ and 2σ probability intervals, respectively. Also shown are the 1σ and 2σ results from Planck (green) and SHoES (orange). (Picture from <https://arxiv.org/pdf/1710.05835.pdf>.)

H_0 comes from the (considerable) width of the posterior probability density for d_L . Also future measurements of this kind should be expected to have large uncertainties in H_0 , at least *individually*. Many more gravitational wave sources with electromagnetic counterparts will be needed, but these will no doubt come about in the coming years. The simulation in Fig. 48 shows what to expect as a combined result from tens of standard siren measurements. This is the point at which gravitational wave cosmology will help resolve – or add to the problem! – of the elusive Hubble constant.

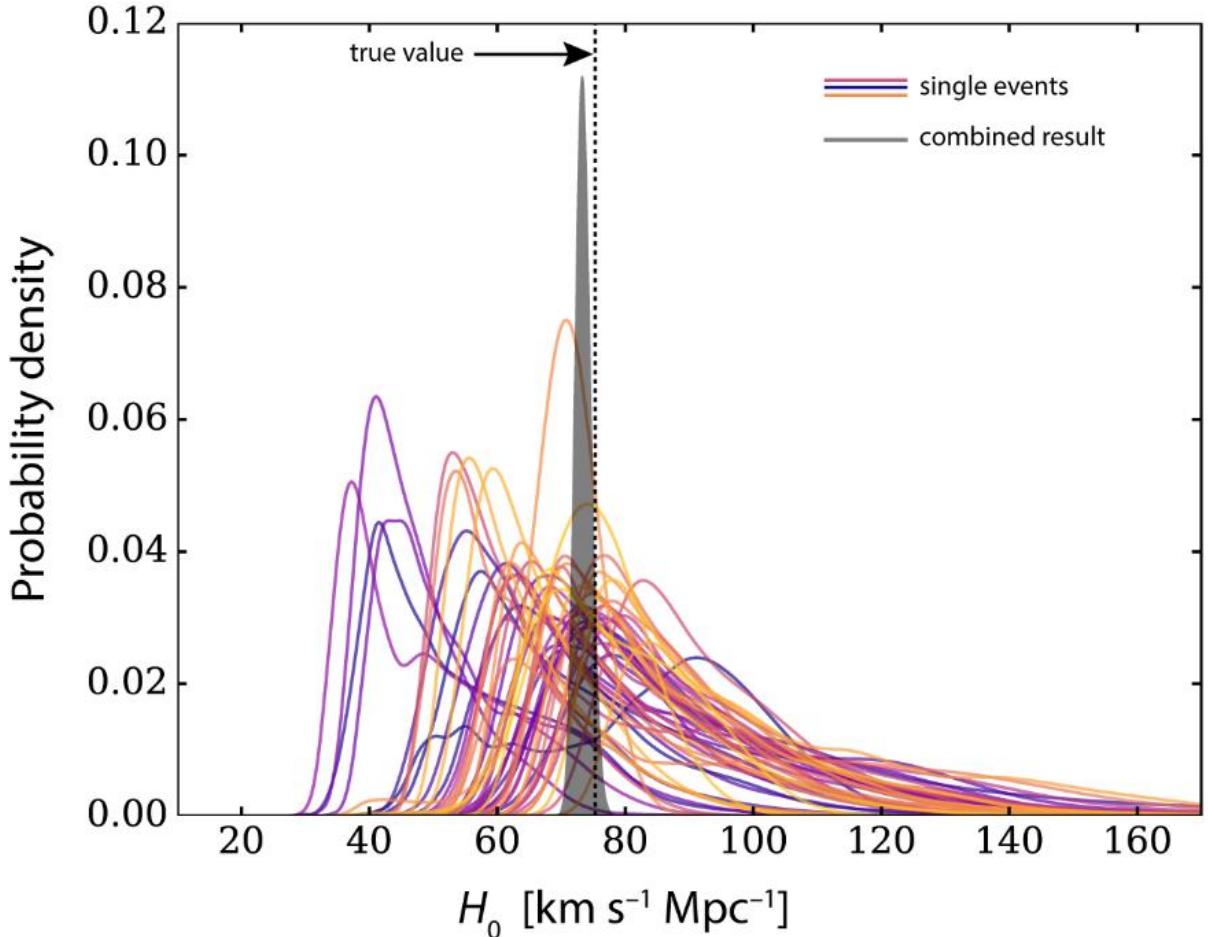


Figure 48: Probability distributions for the Hubble parameter H_0 from 51 simulated binary neutron star inspirals (colored curves). Although no single inspiral event gives an accurate measurement of H_0 , combining information from all of them leads to a value with only a few percent uncertainty (grey shaded curve). The “true” value in the simulation is indicated by the vertical dashed line. (Picture adapted from <https://arxiv.org/pdf/1802.03404.pdf>.)

XIV. GRAVITATIONAL WAVE BACKGROUND RADIATION

In 1963, Penzias and Wilson discovered the Cosmic Microwave Background (CMB), electromagnetic radiation from the time when the first atoms formed, some 300,000 years after the Big Bang. This radiation pervades the Universe, and from our perspective is coming towards us from all directions on the sky, in a very isotropic way. Gravitational radiation that is similarly coming from all directions is referred to as a *stochastic gravitational wave background*.

In fact, there are types of stochastic gravitational waves that are guaranteed to exist. For example, the pile-up of signals from all the binary black hole coalescences anywhere in the visible Universe – most of which will not be individually picked up, since the majority will be very weak – constitutes a stochastic background, which one can expect to be more or less isotropic. Also single neutron stars, i.e. ones that are not part of an inspiraling binary system, emit gravitational waves, due to the fact that they will not be perfectly symmetric around their axis of rotation, leading to a time-dependent quadrupole moment. Their continuous, sinusoidal signals also add up to a stochastic background, but this one will not be isotropic, since by far the strongest contribution will come from the relatively close-by neutron stars in the plane of the Milky Way galaxy. Another type of background, again non-isotropic, is generated by all the white dwarf binaries in the Milky Way. Most of these systems are in a stage where the distance between the two objects is still relatively large, so that the gravitational wave frequency is still nearly constant (though different for every system); nevertheless, their combined gravitational wave background is strong enough that it will constitute a source of noise for the space-based LISA observatory!

The above mentioned stochastic backgrounds are called *astrophysical*. However, there may also exist *primordial* gravitational wave backgrounds that were originally caused by processes in the early Universe, e.g. inflation, or the “phase transition” in which the first protons and neutrons were formed from quarks and gluons, once the temperature had become sufficiently low for this to happen. Most of the primordial backgrounds that have been proposed originate from a fraction of a second after the Big Bang. Though the expansion of the Universe will typically have caused their energy density at the current epoch to be exceedingly low, their spectrum may contain information about what was happening in the Universe when they were first generated, so that they could provide a direct window on processes associated with much higher energies than can be generated in particle accelerators on Earth.

Thus, it is of interest to search for stochastic gravitational wave backgrounds. In what follows we will first characterize such backgrounds and then find out how one would search for them in interferometer data, after which we will look at specific examples.

A. Characterization of stochastic gravitational wave backgrounds

A stochastic gravitational wave background can be viewed as a superposition of plane waves coming from all directions:

$$h_{ij}(t, \mathbf{x}) = \sum_{A=+, \times} \int_{-\infty}^{\infty} df \int d^2 \hat{\mathbf{n}} \tilde{h}_A(f; \hat{\mathbf{n}}) e_{ij}^A(\hat{\mathbf{n}}) e^{2\pi i f(t - \hat{\mathbf{n}} \cdot \mathbf{x}/c)}, \quad (14.1)$$

where we have introduced the *polarization tensors*

$$\begin{aligned} e_{ij}^+(\hat{\mathbf{n}}) &= \hat{\mathbf{u}}_i \hat{\mathbf{u}}_j - \hat{\mathbf{v}}_i \hat{\mathbf{v}}_j, \\ e_{ij}^\times(\hat{\mathbf{n}}) &= \hat{\mathbf{u}}_i \hat{\mathbf{v}}_j + \hat{\mathbf{v}}_i \hat{\mathbf{u}}_j, \end{aligned} \quad (14.2)$$

with $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ unit vectors perpendicular to the propagation direction \mathbf{n} of a particular plane wave. Thus, for a *given* plane wave, in a coordinate system whose (x, y, z) axes are aligned with $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{n}})$,

$$e^+(\hat{\mathbf{n}}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e^\times(\hat{\mathbf{n}}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (14.3)$$

It is easy to see that with these definitions,

$$e_{ij}^A(\hat{\mathbf{n}}) e^{A',ij}(\hat{\mathbf{n}}) = 2\delta^{AA'}. \quad (14.4)$$

When interested in the form of the radiation at the locations of detectors, in Eq. (14.1) one could set $\mathbf{x} = 0$ when considering a single detector at the origin. However, as we shall see, the detection of stochastic backgrounds with only one detector poses difficulties, so that we will consider a network of detectors; the term $-\hat{\mathbf{n}} \cdot \mathbf{x}/c$ in the exponent in Eq. (14.1) takes into account the different arrival times of the wavefronts of the plane waves at the different detector locations.

We now introduce some simplifying assumptions:

- The background is *stationary and Gaussian, with zero mean*. Like with stationary, Gaussian detector noise, this implies that there are no correlations between gravitational wave amplitudes at different frequencies: $\langle \tilde{h}_A^*(f) \tilde{h}_{A'}(f') \rangle \propto \delta(f - f')$, where $\langle \cdot \rangle$ denotes an expectation value. Moreover, $\langle \tilde{h}_A(f) \rangle = 0$.

- The background is *isotropic*: Waves coming from different directions are uncorrelated, so that $\langle \tilde{h}_A^*(f; \hat{\mathbf{n}}) \tilde{h}_{A'}(f'; \hat{\mathbf{n}}') \rangle \propto \delta^2(\hat{\mathbf{n}} - \hat{\mathbf{n}}')$. (This will not hold for the superposition of period signals from all the neutron stars in our galaxy.)
- The background is *unpolarized*, i.e. there are no correlations between the polarizations: $\langle \tilde{h}_A^*(f) \tilde{h}_{A'}(f') \rangle \propto \delta_{AA'}$.

Putting everything together,

$$\langle \tilde{h}_A^*(f; \hat{\mathbf{n}}) \tilde{h}_{A'}(f'; \hat{\mathbf{n}}') \rangle = \delta(f - f') \frac{\delta^2(\hat{\mathbf{n}} - \hat{\mathbf{n}}')}{4\pi} \delta_{AA'} \frac{1}{2} S_h(f), \quad (14.5)$$

for some function $S_h(f)$ called the *spectral density*. The latter will play a similar role as the power spectral density of the noise. Indeed, in this case the signal does share some properties with detector noise, which will lead to challenges in detecting it. Finally, the factor $1/(4\pi)$ in Eq. (14.5) was chosen such that

$$\int d^2\hat{\mathbf{n}} \int d^2\hat{\mathbf{n}}' \langle \tilde{h}_A^*(f) \tilde{h}_{A'}(f') \rangle = \delta(f - f') \delta_{AA'} \frac{1}{2} S_h(f). \quad (14.6)$$

Let us consider the *energy density* in the stochastic background. From Eq. (6.38), this is given in terms of the transverse-traceless h_{ij} as

$$\epsilon_{\text{GW}} = \frac{c^2}{32\pi G} \langle \dot{h}_{ij} \dot{h}_{ij} \rangle, \quad (14.7)$$

where the average is over at least one wavelength, or equivalently over at least one period of the wave. For all practical purposes also the $\langle \dots \rangle$ appearing in e.g. Eq. (14.5) can be taken to be an average over time, so the two notions of expectation value are equivalent.

From Eq. (14.1),

$$\dot{h}_{ij}(t, \mathbf{x}) = \sum_{A=+, \times} \int_0^\infty df \int d^2\hat{\mathbf{n}} \tilde{h}_A(f; \hat{\mathbf{n}}) e_{ij}^A(\hat{\mathbf{n}}) (2\pi if) e^{2\pi if(t - \hat{\mathbf{n}} \cdot \mathbf{x}/c)}. \quad (14.8)$$

This allows us to calculate the energy density:

$$\begin{aligned} \epsilon_{\text{GW}} &= \frac{c^2}{32\pi G} \sum_{A=+, \times} \sum_{A'=+, \times} \int_{-\infty}^\infty df \int_{-\infty}^\infty df' \int d^2\hat{\mathbf{n}} \int d^2\hat{\mathbf{n}}' e_{ij}^A(\hat{\mathbf{n}}) e_{ij}^{A'}(\hat{\mathbf{n}}') e^{2\pi i[(f' - f)t - f' \hat{\mathbf{n}}' \cdot \mathbf{x}/c + f \hat{\mathbf{n}} \cdot \mathbf{x}/c]} \\ &\quad \times (-2\pi if) (2\pi if') \langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f', \hat{\mathbf{n}}') \rangle \\ &= \frac{c^2}{16\pi G} \int_{-\infty}^\infty df (2\pi f)^2 S_h(f), \end{aligned} \quad (14.9)$$

where in the first line, in the first integral over frequency we changed the integration variable, $f \rightarrow -f$, and used that $\tilde{h}_A(-f, \hat{\mathbf{n}}) = \tilde{h}_A^*(f, \hat{\mathbf{n}})$ since $\dot{h}_{ij}(t, \mathbf{x})$ is real, and in the second line we used Eqs. (14.5) and (14.4). Hence we obtain

$$\epsilon_{\text{GW}} = \frac{c^2}{8\pi G} \int_0^\infty df (2\pi f)^2 S_h(f), \quad (14.10)$$

where we used that $S_h(-f) = S_h(f)$.

In the context of cosmology, when talking about the contribution to the total energy density of a particular matter or radiation component, it is useful to compare with the *critical density* ϵ_c , which is the total energy needed to close the Universe. Hence we define

$$\Omega_{\text{GW}} = \frac{\epsilon_{\text{GW}}}{\epsilon_c}, \quad (14.11)$$

where

$$\epsilon_c = \frac{3c^2 H_0^2}{8\pi G}, \quad (14.12)$$

with H_0 the Hubble constant. Now note that one can write

$$\epsilon_{\text{GW}} = \int_{f=0}^{f=\infty} d(\ln f) \frac{d\tilde{\epsilon}_{\text{GW}}(f)}{d\ln f}, \quad (14.13)$$

where we have defined

$$\frac{d\tilde{\epsilon}_{\text{GW}}(f)}{d\ln f} \equiv \frac{c^2}{8\pi G} (2\pi)^2 f^3 S_h(f), \quad (14.14)$$

which is referred to as the gravitational wave energy density per logarithmic frequency bin. Furthermore,

$$\Omega_{\text{GW}} = \int_{f=0}^{f=\infty} d(\ln f) \tilde{\Omega}_{\text{GW}}(f), \quad (14.15)$$

where

$$\tilde{\Omega}_{\text{GW}}(f) \equiv \frac{1}{\epsilon_c} \frac{d\tilde{\epsilon}_{\text{GW}}(f)}{d\ln f}. \quad (14.16)$$

Then from Eq. (14.10) with (14.12),

$$\tilde{\Omega}_{\text{GW}}(f) = \frac{4\pi^2}{3H_0^2} f^3 S_h(f).$$

(14.17)

B. Searching for stochastic backgrounds with a single detector

We now turn to the problem of detecting stochastic gravitational wave backgrounds. Let us first consider the case of a single detector. An immediate problem that arises is the

fact that a stochastic background looks like noise, with some power spectral density $S_h(f)$. Suppose, however, that the sources of instrumental noise in the detector can be adequately modeled, and captured by the noise power spectral density $S_n(f)$. The variance of the noise is

$$\begin{aligned}\langle n^2(t) \rangle &= \langle n^2(t=0) \rangle \\ &= \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \langle \tilde{n}^*(f) \tilde{n}(f') \rangle \\ &= \frac{1}{2} \int_{-\infty}^{\infty} df S_n(f) \\ &= \int_0^{\infty} df S_n(f),\end{aligned}\tag{14.18}$$

where we used Eq. (8.26). In order to detect a stochastic signal with a single detector, we cannot use an optimal filter. Indeed, the latter would require knowledge of the exact time evolution of the detector response $h(t)$, whereas we only have access to expectation values, as discussed in the previous section. For the stochastic background to be detectable, it will be necessary that its variance $\langle h^2(t) \rangle$ exceeds that of the noise, $\langle n^2(t) \rangle$. For each propagation direction $\hat{\mathbf{n}}$ in Eq. (14.1), the detector response is $h(t) = F_+ h_+(t) + F_\times h_\times(t)$. Averaging the expectation value $\langle h^2(t) \rangle$ over the propagation direction $\hat{\mathbf{n}}$ and the polarization angle ψ , and as before assuming an unpolarized stochastic background,

$$\begin{aligned}\int \frac{d^2 \hat{\mathbf{n}}}{4\pi} \frac{d\psi}{2\pi} \langle h^2 \rangle &= \int \frac{d^2 \hat{\mathbf{n}}}{4\pi} \frac{d\psi}{2\pi} \langle F_+^2 h_+^2 + F_\times^2 h_\times^2 \rangle \\ &= \int \frac{d^2 \hat{\mathbf{n}}}{4\pi} \frac{d\psi}{2\pi} F_+^2 \langle h_+^2 + h_\times^2 \rangle,\end{aligned}\tag{14.19}$$

where we have used that the averages of F_+^2 or F_\times^2 over $\hat{\mathbf{n}}$ and ψ are equal, as can easily be verified. Noting that $h_{ij} h^{ij} = 2(h_+^2 + h_\times^2)$ for any propagation direction, together with

$$\langle h_{ij} h^{ij} \rangle = 4 \int_0^{\infty} df S_h(f),\tag{14.20}$$

leads to

$$\langle h^2(t) \rangle = 2 \langle F_+^2 \rangle \int_0^{\infty} df S_h(f),\tag{14.21}$$

where we have extended the meaning of $\langle \dots \rangle$ to include an average over both time and $(\hat{\mathbf{n}}, \psi)$.

Given the expressions (14.18) and (14.21), a stochastic signal is detectable if the “signal-to-noise ratio” S/N is suitably large, where this time

$$\begin{aligned}(S/N)^2 &\equiv \frac{\langle h^2(t) \rangle}{\langle n^2(t) \rangle} \\ &= \frac{2\langle F_+^2 \rangle \int_0^\infty df S_h(f)}{\int_0^\infty df S_n(f)}.\end{aligned}\quad (14.22)$$

However, a stochastic signal may be stronger at some frequencies than at others, so that it will pay to look at different frequencies separately. Given an observation time T , corresponding to a frequency resolution Δf , one can also define a signal-to-noise ratios for separate frequency bins $[f_i, f_i + \Delta f]$ by

$$(S/N)_i^2 = \frac{2\langle F_+^2 \rangle S_h(f_i) \Delta f}{S_n(f_i) \Delta f}.\quad (14.23)$$

The latter reveals an interesting point about searching for stochastic backgrounds with a single detector: The integration time T drops out of the expression for the signal-to-noise ratio in given frequency bin. A longer integration time makes Δf narrower, causing there to be less noise in a given frequency bin, but also less signal. Thus, for a single detector there is no gain in long observing time; either a signal is in evidence as soon as the detector is switched on, or it is not. As we shall see below, this will be different for *networks* of detectors.

For a given frequency f and a signal-to-noise threshold $(S/N)_{\text{thr}}$, Eq. (14.23) tells us how small $S_h(f)$ can be while still being detectable:

$$[S_h(f)]_{\min} = S_n(f) \frac{(S/N)_{\text{thr}}^2}{2\langle F_+^2 \rangle}.\quad (14.24)$$

Using Eq. (14.17), this translates to a minimum density per logarithmic frequency bin that will be detectable:

$$\left[\tilde{\Omega}_{\text{GW}}(f) \right]_{\min} = \frac{4\pi^2}{3H_0^2} f^3 S_n(f) \frac{(S/N)_{\text{thr}}^2}{2\langle F_+^2 \rangle}.\quad (14.25)$$

Let us consider an example of a single detector taking data in a particular frequency range, namely the space-based LISA. The latter will be a low-frequency detector, reaching a sensitivity of $S_n^{1/2}(f) = 4 \times 10^{-21} \text{ Hz}^{-1/2}$ at a frequency of $f \simeq 10^{-3} \text{ Hz}$. Eq. (14.25) can then be written as

$$\left[h_0^2 \tilde{\Omega}_{\text{GW}}(f) \right]_{\min} \simeq 2.5 \times 10^{-13} \left(\frac{f}{10^{-3} \text{ Hz}} \right)^3 \left(\frac{S_n^{1/2}}{4 \times 10^{-21} \text{ Hz}^{-1/2}} \right)^2 \left(\frac{S/N}{5} \right)^2,\quad (14.26)$$

where we used that $\langle F_+^2 \rangle = \mathcal{O}(1)$, and we have expressed the Hubble constant as $H_0 = h_0 \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$; according to most measurements, $h_0 \simeq 0.7$.

In the case of LISA, there exists a stochastic background that is almost guaranteed to be detectable, namely the accumulation of gravitational wave signals from all the white dwarf binaries in the galaxy. For simplicity, assume that the number of white dwarf binaries that is formed per unit of time, $R = dN/dt$, is constant. Then the number of sources per frequency bin is

$$\frac{dN}{df}(f) = \frac{dN}{dt} \frac{dt}{df} = R \frac{1}{\dot{f}}. \quad (14.27)$$

Note that dN/df does depend on frequency, because the sources spend a lot more time at low frequencies (where the frequency evolves more slowly, i.e. \dot{f} is small) than at high frequencies, so that dN/df will be higher at low frequencies. The number of sources per *logarithmic* frequency bin is then

$$\frac{dN}{d \ln f}(f) = R \frac{f}{\dot{f}}, \quad (14.28)$$

and the total energy emitted in gravitational waves per unit time and per unit logarithmic frequency is

$$\frac{d^2 E}{dt d \ln f} = \frac{dE_{\text{GW}}}{dt}(f) R \frac{f}{\dot{f}}, \quad (14.29)$$

where dE_{GW}/dt is the average GW luminosity of a source at frequency f . Let $\langle r \rangle$ be the average distance to a source; then one can consider a thin shell centered on the Solar system, with radius $\langle r \rangle$ and thickness cdt . Then Eq. (14.29) can be converted into an energy density per logarithmic frequency bin:

$$\begin{aligned} \frac{d\tilde{\epsilon}_{\text{GW}}}{d \ln f} &= \frac{1}{4\pi \langle r \rangle^2 c dt d \ln f} \frac{d^2 E}{dt d \ln f}(f) \\ &= \frac{1}{4\pi \langle r \rangle^2 c} \frac{dE_{\text{GW}}}{dt}(f) R \frac{f}{\dot{f}}. \end{aligned} \quad (14.30)$$

From Eqs. (14.16) and (14.12) we obtain

$$\tilde{\Omega}_{\text{GW}}(f) = \frac{8\pi G}{3c^2 H_0^2} \frac{1}{4\pi \langle r \rangle^2 c} \frac{dE_{\text{GW}}}{dt}(f) R \frac{f}{\dot{f}}. \quad (14.31)$$

Now note that for sources whose total energy depends on frequency through a power law, $E \propto f^\alpha$, one has $f/\dot{f} = \alpha E/\dot{E}$. In that case

$$\tilde{\Omega}_{\text{GW}}(f) = \frac{8\pi G}{3c^2 H_0^2} \frac{R}{4\pi \langle r \rangle^2 c} \alpha E(f). \quad (14.32)$$

For white dwarf binaries, the total energy is the orbital energy, and from Eq. (7.20) we see that $\alpha = 2/3$. Again for simplicity, let all white dwarfs have a mass of $m_{\text{WD}} = 1 M_{\odot}$; then the chirp mass for a binary white dwarf system becomes $\mathcal{M}_c = 2^{-1/5} M_{\odot}$. An estimate for the rate R at which white dwarf binaries are formed in our galaxy is $R \simeq 1/(100 \text{ yr})$. For $\langle r \rangle$ we can use the distance between us and the center of the Milky Way (around which there will be a concentration of such binaries), which is about 10 kpc. Again taking 10^{-3} Hz as a reference frequency, this finally yields

$$h_0^2 \tilde{\Omega}_{\text{GW}}(f) \simeq 7.8 \times 10^{-10} \left(\frac{R}{1/(100 \text{ yr})} \right) \left(\frac{m_{\text{WD}}}{M_{\odot}} \right)^{5/3} \left(\frac{f}{10^{-3} \text{ Hz}} \right)^{2/3} \left(\frac{\langle r \rangle}{10 \text{ kpc}} \right)^{-2}. \quad (14.33)$$

Comparing with Eq. (14.26), we see that the combined signals from all the white dwarf binaries in the Milky Way would indeed lead to a stochastic background detectable by LISA!

In fact, when using LISA to search for signals from sources *other* than white dwarf binaries (e.g. from supermassive binary black hole inspirals), a stochastic background caused by binary white dwarfs must be viewed as a contribution to the noise power spectral density $S_n(f)$. Since this contribution will be proportional to $\tilde{\Omega}_{\text{GW}}$ and again comparing with Eq. (14.26), it might seem as if the white dwarf background will dramatically limit LISA's sensitivity at frequencies around 10^{-3} Hz . However, note that Eq. (14.33) was arrived at by simply estimating the combined energy density from all the binary white dwarf gravitational wave emission in the Milky Way. In reality, many of the signals will be loud enough to be *individually* detected as waves with slowly varying frequencies, in which case they can just be subtracted from the detector output when looking for other target sources. Nevertheless, the remaining sources will form a stochastic background that will still stick out above the purely instrumental contributions to $S_n(f)$, as illustrated in Fig. 49.

C. Searching for stochastic backgrounds with multiple detectors

As we have seen, with a single detector the only way to find a stochastic background is if it sticks out above the instrumental noise (with the assumption that the latter is sufficiently well-modeled), and there is no advantage to be had from long observation times. This is because with only one detector, a stochastic background looks like a contribution to the noise. The situation is different when considering multiple detectors: A stochastic

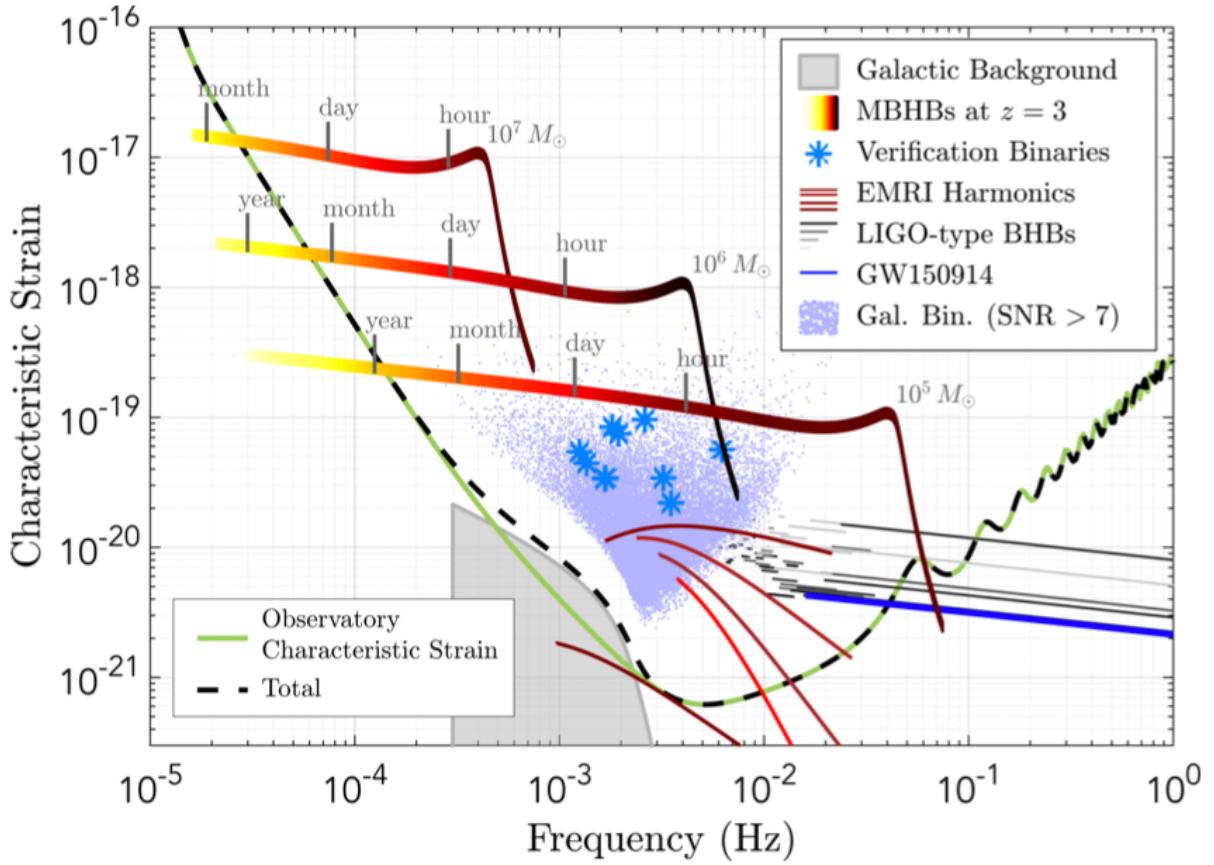


Figure 49: A plot of $\sqrt{S_n(f)}$ as a function of frequency for LISA, together with some representative gravitational wave sources. The green curve shows the contribution to $\sqrt{S_n(f)}$ from instrumental noise sources. The grey area is the contribution from unresolvable white dwarf binary signals that form a stochastic background, which combines with the instrumental noise; the resulting $\sqrt{S_n(f)}$ is indicated by the dashed black curve. The purple dots illustrate white dwarf binary signals that can easily be detected individually. In fact, a number of close-by binaries that have already been observed with ordinary telescopes will act as “verification binaries” for LISA (purple asterisks): Barring instrumental malfunctions, LISA should see their signals at the given frequencies with an already predicted loudness. (For more details about this plot, see <https://arxiv.org/pdf/1803.01944.pdf>.)

background will still appear as a contribution to the noise, but in a way that is *correlated* between detectors, while instrumental noise fluctuations in the different detectors usually are not. This distinction opens up the possibility of detecting much weaker gravitational wave background signals than would be possible with a single detector.

For simplicity we consider the case of two detectors, and write their outputs as

$$s_k(t) = n_k(t) + h_k(t), \quad k = 1, 2, \quad (14.34)$$

where $n_k(t)$ is the noise in a given detector, and $h_k(t)$ the signal as seen in that detector. Note that generally $h_1(t) \neq h_2(t)$; as will be clear from the discussion in Chapter VIII, if two detectors have different orientations of their arms, then they will respond differently to the same signal. Indeed, with the notation of Eq. (14.1),

$$h_k(t, \mathbf{x}_k) = \sum_{A=+, \times} \int_{-\infty}^{\infty} df \int d^2 \hat{\mathbf{n}} \tilde{h}_A(f, \hat{\mathbf{n}}) e^{2\pi i f(t - \hat{\mathbf{n}} \cdot \mathbf{x}_k / c)} F_k^A(\hat{\mathbf{n}}). \quad (14.35)$$

Here $F_k^A(\hat{\mathbf{n}})$, $A = +, \times$ are the antenna pattern functions for detector k given a propagation direction $\hat{\mathbf{n}}$. Since the coordinate systems attached to the two interferometers will generally be differently oriented, the angles (θ, ϕ) entering the expressions (8.12) will have different values for the same propagation direction $\hat{\mathbf{n}}$ but different detectors, since they refer to projections onto different axes; this leads to different values of F_k^A for $k = 1, 2$.³⁶ Note that in Eq. (14.35), the \mathbf{x}_k are the positions of the detectors; see the discussion below Eq. (14.4).

The benefit of having two detectors will be to look for correlations between their outputs. To this end, define

$$Y = \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' s_1(t) s_2(t') Q(t - t'), \quad (14.36)$$

where the yet to be specified filter function $Q(t - t')$ will play an analogous role to the filter $K(t)$ in Chapter VIII. Similarly to what we did in Section D of that chapter, we define

$$S = \langle Y \rangle_h, \quad N = [\langle Y^2 \rangle_{h=0} - \langle Y \rangle_{h=0}^2]^{1/2}, \quad (14.37)$$

where $\langle \dots \rangle_h$ and $\langle \dots \rangle_{h=0}$ denote expectation values with and without a signal present, and we consider the ratio

$$S/N = \frac{\langle Y \rangle_h}{[\langle Y^2 \rangle_{h=0} - \langle Y \rangle_{h=0}^2]^{1/2}}. \quad (14.38)$$

As before, our aim will be to find the filter $Q(t - t')$ that maximizes S/N .³⁷

³⁶ The polarization angles ψ will also be different between detectors. However, since it will not play much of a role in what follows, we will suppress the dependence of the F_k^A on ψ to simplify notation.

³⁷ Note the small difference between Eq. (8.34) and (14.37): in the latter, the square of the expectation value of Y is subtracted in the definition of Y . This is because we are interested in how large the departures of Y from its average value will be in the absence of the signal, and $\langle Y \rangle$ may not be zero. Previously we did not have to do this, because in Eq. (8.34), $\langle \hat{s} \rangle = 0$, since $\langle n(t) \rangle = 0$.

It will be convenient to have an expression for Y in the frequency domain. One has

$$Y = \int_{-\infty}^{\infty} df df' df'' \delta_T(f - f'') \delta_T(f' - f'') \tilde{s}_1^*(f) \tilde{s}_2(f') \tilde{Q}(f''), \quad (14.39)$$

where

$$\delta_T(f) \equiv \int_{-T/2}^{T/2} dt e^{2\pi i f t} = \frac{\sin(\pi f T)}{\pi f}, \quad (14.40)$$

and it can be shown that this becomes a Dirac delta function in the limit $fT \rightarrow \infty$. For an observation time of $T = 1 \text{ yr} \simeq 3 \times 10^7 \text{ s}$ and a frequency $f = 10 \text{ Hz}$ (which is at the low end, at least for ground-based detectors) one already has $fT \simeq 3 \times 10^8$, so in what follows we will approximate $\delta_T(f) \simeq \delta(f)$, so that

$$Y \simeq \int_{-\infty}^{\infty} df \tilde{s}_1^*(f) \tilde{s}_2(f) \tilde{Q}(f). \quad (14.41)$$

We now have to calculate S and N in Eq. (14.37), with $\tilde{s}_k(f) = \tilde{n}_k(f) + \tilde{h}_k(f)$. Note that the stochastic gravitational wave signal will be uncorrelated with the noise in either detector, so $\langle \tilde{n}_1^*(f) \tilde{h}_2(f) \rangle = \langle \tilde{h}_1^*(f) \tilde{n}_2(f) \rangle = 0$. Moreover, we will assume that there is no correlation between the noises in the two detectors, so that $\langle \tilde{n}_1^*(f) \tilde{n}_2(f) \rangle = 0$. In that case, using Eq. (14.35),

$$\begin{aligned} S &= \int_{-\infty}^{\infty} df \langle \tilde{h}_1^*(f) \tilde{h}_2(f) \rangle \tilde{Q}(f) \\ &= \int_{-\infty}^{\infty} df \sum_{A,A'} \int d\hat{\mathbf{n}} \int d\hat{\mathbf{n}}' e^{2\pi i f (\hat{\mathbf{n}} \cdot \mathbf{x}_1 - \hat{\mathbf{n}}' \cdot \mathbf{x}_2)/c} F_1^A(\hat{\mathbf{n}}) F_2^{A'}(\hat{\mathbf{n}}') \langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f, \hat{\mathbf{n}}') \rangle \tilde{Q}(f). \end{aligned} \quad (14.42)$$

Next, recall our assumptions about the stochastic gravitational wave background, namely that it is stationary and Gaussian with zero mean, isotropic, and unpolarized, which led to Eq. (14.5) for $\langle \tilde{h}_A^*(f, \hat{\mathbf{n}}) \tilde{h}_{A'}(f, \hat{\mathbf{n}}') \rangle$. Together with

$$\delta(0) \simeq \int_{-T/2}^{T/2} dt = T \quad (14.43)$$

this leads to

$$S = \frac{T}{2} \int_{-\infty}^{\infty} df S_h(f) \Gamma(f) \tilde{Q}(f), \quad (14.44)$$

where we defined the *overlap reduction function*

$$\Gamma(f) \equiv \int \frac{d\hat{\mathbf{n}}}{4\pi} \sum_A F_1^A(\hat{\mathbf{n}}) F_2^A(\hat{\mathbf{n}}) \exp \left[2\pi i f \hat{\mathbf{n}} \cdot \frac{\Delta \mathbf{x}}{c} \right],$$

(14.45)

with $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ the separation between the two detectors. Note that in this function, for $2\pi f \Delta x / c \gg 1$, or equivalently when the wavelength is much smaller than the distance between detectors, the exponential is rapidly oscillating, and the detectors will effectively experience uncorrelated signals. In fact, the overlap reduction function is maximized when

- $\Delta x = 0$, i.e. the detectors are at the same location.
- The interferometer arms are aligned, which maximizes $\int \frac{d\hat{\mathbf{n}}}{4\pi} \sum_A F_1^A(\hat{\mathbf{n}}) F_2^A(\hat{\mathbf{n}})$.

However, it is also desirable for the *noise* to be uncorrelated between detectors, which is hard to achieve if they are co-located and aligned; in particular, environmental noise (e.g. seismic vibrations caused by human activity) will be correlated. Nevertheless, the original LIGO-Hanford detector had two interferometers in the same vacuum system, respectively with 4 km and 2 km arm length, with a view on facilitating searches for stochastic backgrounds. In the transition to Advanced LIGO, the smaller detector was omitted, and both LIGO detectors now only consist of a single interferometer of 4 km arm length. Note that this still allows to search for stochastic backgrounds, at wavelengths λ satisfying $\lambda \gg \Delta x$. Since the separation between the two LIGO detectors is about 3000 km, or $\Delta x \simeq 3 \times 10^6$ m, this implies sensitivity to frequencies $f = c/\lambda \ll 10^2$ Hz.

Having obtained an expression for S , let us now focus on N . A brief calculation will show that

$$\begin{aligned} N^2 &= \langle Y^2 \rangle_{h=0} - \langle Y \rangle_{h=0}^2 \\ &= \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \tilde{Q}(f) \tilde{Q}^*(f') \\ &\quad \times [\langle \tilde{n}_1^*(f) \tilde{n}_2(f) \tilde{n}_1(f') \tilde{n}_2^*(f') \rangle - \langle \tilde{n}_1^*(f) \tilde{n}_2(f) \rangle \langle \tilde{n}_1(f') \tilde{n}_2^*(f') \rangle]. \end{aligned} \quad (14.46)$$

We will continue to assume that the noise between detectors is uncorrelated, so that $\langle \tilde{n}_1^*(f) \tilde{n}_2(f) \rangle = \langle \tilde{n}_1(f') \tilde{n}_2^*(f') \rangle = 0$. Moreover, if the noise in the two detectors is independent, then the probability distribution for n_1, n_2 has the property that $p[n_1, n_2] = p_1[n_1] p_2[n_2]$, where the two factors are the probability distributions for the individual detectors; we leave it as an exercise that this implies

$$\langle \tilde{n}_1^*(f) \tilde{n}_2(f) \tilde{n}_1(f') \tilde{n}_2^*(f') \rangle = \langle \tilde{n}_1^*(f) \tilde{n}_1(f') \rangle \langle \tilde{n}_2(f) \tilde{n}_2^*(f') \rangle. \quad (14.47)$$

Thus,

$$N^2 = \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \tilde{Q}(f) \tilde{Q}^*(f') \langle \tilde{n}_1^*(f) \tilde{n}_1(f') \rangle \langle \tilde{n}_2(f) \tilde{n}_2^*(f') \rangle. \quad (14.48)$$

For each detector $k = 1, 2$ separately, one has

$$\langle \tilde{n}_k^*(f) \tilde{n}_k(f') \rangle = \frac{1}{2} \delta(f - f') S_{n,k}(f), \quad (14.49)$$

where $S_{n,k}(f)$ are the noise power spectral densities of the two detectors; see Eq. (8.26).

Again identifying $\delta(0) = T$, we obtain

$$N^2 = \frac{T}{4} \int_{-\infty}^{\infty} df |\tilde{Q}(f)|^2 S_n^2(f), \quad (14.50)$$

where we defined $S_n(f) \equiv [S_{n,1}(f) S_{n,2}(f)]^{1/2}$.

Plugging Eqs. (14.44) and (14.50) into Eq. (14.38), we find

$$S/N = T^{1/2} \frac{\int_{-\infty}^{\infty} df S_h(f) \Gamma(f) \tilde{Q}(f)}{\left[\int_{-\infty}^{\infty} df |\tilde{Q}(f)|^2 S_n^2(f) \right]^{1/2}}. \quad (14.51)$$

We now want to find the filter function $\tilde{Q}(f)$ that maximizes S/N . To this end we define a new inner product:

$$(A|B) \equiv \int_{-\infty}^{\infty} df \tilde{A}^*(f) \tilde{B}(f) S_n^2(f). \quad (14.52)$$

Noting that $\tilde{Q}^*(f) = \tilde{Q}(-f)$, $S_h(f) = S_h(-f)$, and $\Gamma(-f)$ can be replaced by $\Gamma(f)$ by interchanging detector labels, it is easy to see that

$$S/N = T^{1/2} \frac{(Q|\Gamma S_h/S_n^2)}{(Q|Q)^{1/2}}. \quad (14.53)$$

Apart from the prefactor $T^{1/2}$, the right hand side is the inner product of a unit vector $Q/(Q|Q)^{1/2}$ with a given vector $\Gamma S_h/S_n^2$. This is maximized when the two vectors are aligned, $Q/(Q|Q)^{1/2} \propto \Gamma S_h/S_n^2$, or equivalently $Q \propto \Gamma S_h/S_n^2$. Therefore, the optimal filter is the one that satisfies

$$\boxed{\tilde{Q}(f) \propto \frac{\Gamma(f) S_h(f)}{S_n^2(f)}}, \quad (14.54)$$

and the optimal signal-to-noise ratio is

$$\boxed{\rho_{\text{opt}} = \left[2T \int_0^{\infty} df \Gamma^2(f) \frac{S_h(f)^2}{S_n(f)^2} \right]^{1/2}}, \quad (14.55)$$

This result makes intuitive sense. The overlap reduction function $\Gamma(f)$ has the information on which frequency regime contributes the most to S/N , given that a detector pair only has access to wavelengths that are much larger than the separation between detectors, and it

also folds in the detectors' relative orientations. Next, the signal-to-noise ratio must involve a comparison of the power in the signal, $S_h(f)$, with the power in the noise, $S_n(f)$, in every frequency bin. Finally, S/N grows with the square root of the observation time, as also seen in Chapter VIII.

In terms of the relative energy density of Eq. (14.17),

$$S/N = \frac{3H_0^2}{4\pi^2} \left[2T \int_0^\infty df \Gamma^2(f) \frac{\tilde{\Omega}_{\text{GW}}^2(f)}{f^6 S_{n,1}(f) S_{n,2}(f)} \right]^{1/2}. \quad (14.56)$$

Just like we did for a single detector, we can now estimate how large $\tilde{\Omega}_{\text{GW}}(f)$ needs to be in order to be detectable with a two-detector network. Let us approximate Eq. (14.55) as

$$S/N \sim (2T\Delta f)^{1/2} \Gamma \frac{S_h}{S_n}, \quad (14.57)$$

where Δf is the frequency band in which the detectors are sensitive, and Γ , S_h , and S_n are typical values of $\Gamma(f)$, $S_h(f)$, and $S_n(f)$ in the sensitive band. Let us assume that the detectors are co-located and aligned, in which case it is an easy exercise to show that $\Gamma = 2\langle F_+^2 \rangle$. Then the minimum detectable value of S_h at signal-to-noise ratio S/N is

$$[S_h]_{\min} \sim \frac{S_n}{(2T\Delta f)^{1/2}} \frac{S/N}{2\langle F_+^2 \rangle}, \quad (14.58)$$

so that

$$\left[\tilde{\Omega}_{\text{GW}} \right]_{\min} \sim \frac{4\pi^2}{3H_0^2} \frac{f^3 S_n}{(2T\Delta f)^{1/2}} \frac{S/N}{2\langle F_+^2 \rangle}, \quad (14.59)$$

where f^3 should be taken to be a typical value over the bandwidth Δf . This is to be compared with the expression for the case of a single detector, Eq. (14.25). Apart from having a different power of the somewhat arbitrary S/N threshold, we see that in the two-detector case, a factor $1/(2T\Delta f)^{1/2}$ has appeared. For the two LIGO detectors we have seen that we will only be sensitive to low frequencies within the total frequency band, so let us take the relevant range to be $f \in [20, 30]$ Hz, i.e. $\Delta f = 10$ Hz. For a one-year integration time, $1/(2T\Delta f)^{1/2} \simeq 4 \times 10^{-5}$, signifying a dramatic increase in sensitivity to stochastic backgrounds compared to the one-detector case.

Returning to the expression for the optimal filter, Eq. (14.54), we see that it expects some knowledge of $S_h(f)$, the power spectral density of the stochastic signal. As it turns out, theoretical predictions for various types of stochastic backgrounds often take the form of a power law for the relative energy density as function of frequency within the sensitive

band of a detector: $\tilde{\Omega}_{\text{GW}} \propto f^\alpha$ for a given α . Through Eq. (14.17), this means $S_h(f) \propto f^{\alpha-3}$, which when substituted into Eq. (14.54) leads to a concrete form for the optimal filter – optimized to search for a stochastic background with the given index α .

As can be seen from Eq. (14.33), for the background caused by the white dwarf binaries in the Milky Way, $\tilde{\Omega}_{\text{GW}}(f) \propto f^{2/3}$, so that in this case $\alpha = 2/3$. The latter is a source for LISA, but also for ground-based detectors there is an in principle detectable background, namely the combined gravitational radiation from all the binary black hole mergers in the visible Universe. Almost all of the individual sources are too weak to be individually detectable, but together their signals pile up to form a stochastic background, again with $\alpha = 2/3$. However, in this case the background is quite weak, and will require multi-detector correlation over observation times of years in order to become apparent. Yet another background of astrophysical origin that could be found with ground-based observatories is the one composed of the combined signals from all the neutron stars in the Galaxy that are spinning around their axis at different frequencies. A useful exercise is to show that for this background, assuming the frequencies are uniformly distributed, $\alpha = 2$.

Though finding stochastic backgrounds caused by astrophysical sources would add to our understanding of their populations, a perhaps more exciting possibility is the detection of a *primordial* background that originated in the early Universe. The process of inflation that is believed to have taken place a fraction of a second after the Big Bang will have caused such a background, but unfortunately the simplest inflationary models predict its typical strength to be well below the threshold of detectability for current and future ground-based and space-based GW observatories. However, other mechanism have been theorized that might lead to a detectable background, some of which are indicated in Fig. 50.

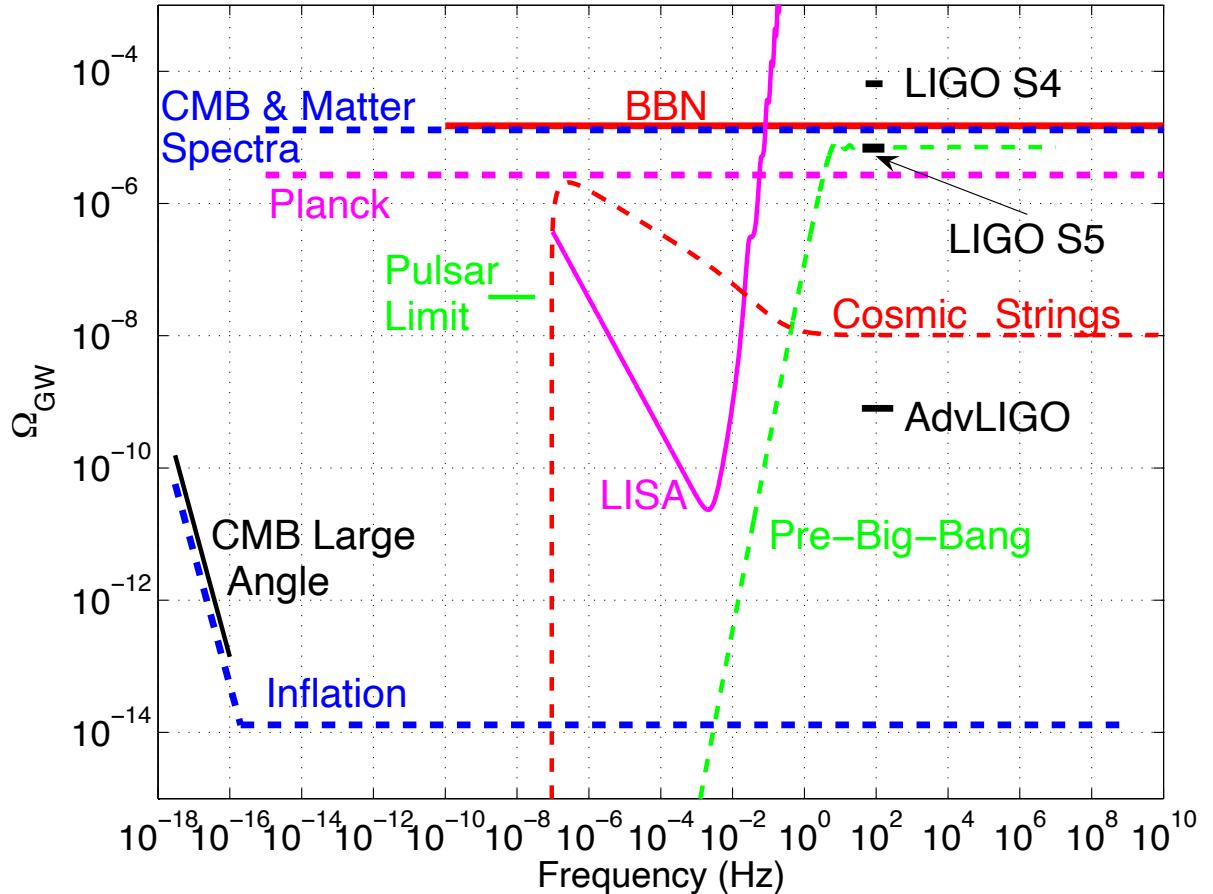


Figure 50: Various stochastic backgrounds that have been hypothesized, together with the frequency ranges and sensitivities of different detectors. The short horizontal black bars indicate two “science runs” (“S4”, “S5”) of LIGO in its initial stages, and Advanced LIGO at its design sensitivity; the purple curve is for LISA. The dashed blue line is the expected GW background from inflation. The dashed green is a possible signal from so-called Pre-Big-Bang cosmology models, and the dashed red from cosmic strings. (Figure from <https://arxiv.org/pdf/0910.5772.pdf>.)