

Lectures on Astronomy, Astrophysics, and Cosmology

Luis A. Anchordoqui

*Department of Physics and Astronomy, Lehman College, City University of New York, NY 10468, USA
 Department of Physics, Graduate Center, City University of New York, 365 Fifth Avenue, NY 10016, USA
 Department of Astrophysics, American Museum of Natural History, Central Park West 79 St., NY 10024, USA*

(Dated: Spring 2016)

I. STARS AND GALAXIES

A look at the night sky provides a strong impression of a changeless universe. We know that clouds drift across the Moon, the sky rotates around the polar star, and on longer times, the Moon itself grows and shrinks and the Moon and planets move against the background of stars. Of course we know that these are merely local phenomena caused by motions within our solar system. Far beyond the planets, the stars appear motionless. Herein we are going to see that this impression of changelessness is illusory.

According to the ancient cosmological belief, the stars, except for a few that appeared to move (the planets), were fixed on a sphere beyond the last planet; see Fig. 1. The universe was self contained and we, here on Earth, were at its center. Our view of the universe dramatically changed after Galileo's first telescopic observations: we no longer place ourselves at the center and we view the universe as vastly larger [1, 2]. The distances involved are so large that we specify them in terms of the time it takes the light to travel a given distance. For example, one light second = $3 \times 10^8 \text{ m} = 300,000 \text{ km}$, one light

minute = $1.8 \times 10^7 \text{ km}$, and one light year

$$1 \text{ ly} = 9.46 \times 10^{15} \text{ m} \approx 10^{13} \text{ km}. \quad (1)$$

For specifying distances to the Sun and the Moon, we usually use meters or kilometers, but we could specify them in terms of light. The Earth-Moon distance is 384,000 km, which is 1.28 ls. The Earth-Sun distance is 150,000,000 km; this is equal to 8.3 lm. Far out in the solar system, Pluto is about $6 \times 10^9 \text{ km}$ from the Sun, or $6 \times 10^{-4} \text{ ly}$. The nearest star to us, Proxima Centauri, is about 4.2 ly away. Therefore, the nearest star is 10,000 times farther from us than the outer reach of the solar system.

On clear moonless nights, thousands of stars with varying degrees of brightness can be seen, as well as the long cloudy strip known as the Milky Way. Galileo first observed with his telescope that the Milky Way is comprised of countless numbers of individual stars. A half century later Wright suggested that the Milky Way was a flat disc of stars extending to great distances in a plane, which we call the Galaxy [3].

Our Galaxy has a diameter of 100,000 ly and a thickness of roughly 2,000 ly. It has a bulging central "nucleus" and spiral arms. Our Sun, which seems to be just another star, is located half way from the Galactic center to the edge, some 26,000 ly from the center. The Sun orbits the Galactic center approximately once every 250 million years or so, so its speed is

$$v = \frac{2\pi}{2.5 \times 10^8 \text{ yr}} \frac{26,000 \times 10^{13} \text{ km}}{3.156 \times 10^7 \text{ s/yr}} = 200 \text{ km/s}. \quad (2)$$

The total mass of all the stars in the Galaxy can be estimated using the orbital data of the Sun about the center of the Galaxy. To do so, assume that most of the mass is concentrated near the center of the Galaxy and that the Sun and the solar system (of total mass m) move in a circular orbit around the center of the Galaxy (of total mass M). Then, apply Newton's 2nd law, $F = ma$, with $a = v^2/r$ being the centripetal acceleration and the force F given by the universal law of gravitation

$$\frac{GMm}{r^2} = m \frac{v^2}{r}, \quad (3)$$

where $G = 6.674 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ [4]. All in all,

$$M = \frac{r v^2}{G} \approx 2 \times 10^{41} \text{ kg}. \quad (4)$$

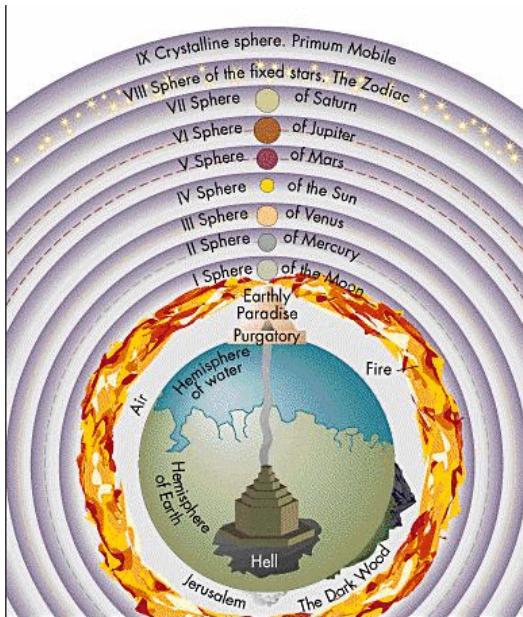


FIG. 1: Celestial spheres of ancient cosmology.

Assuming all the stars in the Galaxy are similar to our Sun ($M_{\odot} \approx 2 \times 10^{30}$ kg), we conclude that there are roughly 10^{11} stars in the Galaxy.

In addition to stars both within and outside the Milky Way, we can see with a telescope many faint cloudy patches in the sky which were once all referred to as “nebulæ” (Latin for clouds). A few of these, such as those in the constellations of Andromeda and Orion, can actually be discerned with the naked eye on a clear night. In the XVII and XVIII centuries, astronomers found that these objects were getting in the way of the search for comets. In 1781, in order to provide a convenient list of objects not to look at while hunting for comets, Messier published a celebrated catalogue [5]. Nowadays astronomers still refer to the 103 objects in this catalog by their Messier numbers, e.g., the Andromeda Nebula is M31.

Even in Messier’s time it was clear that these extended objects are not all the same. Some are star clusters, groups of stars which are so numerous that they appeared to be a cloud. Others are glowing clouds of gas or dust and it is for these that we now mainly reserve the word nebula. Most fascinating are those that belong to a third category: they often have fairly regular elliptical shapes and seem to be a great distance beyond the Galaxy. Kant seems to have been the first to suggest that these latter might be circular discs, but appear elliptical because we see them at an angle, and are faint because they are so distant [6]. At first it was not universally accepted that these objects were extragalactic (i.e. outside our Galaxy). The very large telescopes constructed in the XX century revealed that individual stars could be resolved within these extragalactic objects and that many contain spiral arms. Hubble did much of this observational work in the 1920’s using the 2.5 m telescope on Mt. Wilson near Los Angeles, California. Hubble demonstrated that these objects were indeed extragalactic because of their great distances [7]. The distance to our nearest spiral galaxy, Andromeda, is over 2 million ly, a distance 20 times greater than the diameter of our Galaxy. It seemed logical that these nebulæ must be galaxies similar to ours. Today it is thought that there are roughly 4×10^{10} galaxies in the observable universe – that is, as many galaxies as there are stars in the Galaxy.

EXERCISE 1.1 Why do we have seasons? Briefly explain the two main reasons. Use diagrams. You must talk about energy, and say more than “tilt of axis.”

EXERCISE 1.2 Suppose you are building a scale model of the nearby stars in our Galaxy. The Sun, which has a radius of 696,000 km, is represented by a tennis ball with a radius of 6 cm. On this scale, how far away would the nearest star, Proxima Centauri, be? Express your answer in km.

EXERCISE 1.3 One earth year is 365.244 (solar)

days. Humans would be much happier if there were an integral number of days in a year. Derive Newton’s form of Kepler’s third law and figure out three ways to make the length of earth’s year exactly 354.000 (solar) days. (This would also make the months be more coordinated with the phases of the moon. We could have 6 months of 29 days and 6 months of 30 days. Since the orbital period of the moon is about 29.5 days, and 12 months times 29.5 days/month = 354 days, I think this would make a great calendar!) Check your answers experimentally by actually changing the mass of the Sun and Earth, and the orbital radius of the Earth.

II. DISTANCE MEASUREMENTS

We have been talking about the vast distance of the objects in the universe. We now turn to discuss different methods to estimate these distances.

A. Stellar Parallax

One basic method to measure distances to nearby stars employs simple geometry and stellar parallax. Parallax is the apparent displacement of an object because of a change in the observer’s point of view. One way to see how this effect works is to hold your hand out in front of you and look at it with your left eye closed, then your right eye closed. Your hand will appear to move against the background. By stellar parallax we mean the apparent motion of a star against the background of more distant stars, due to Earth’s motion around the Sun; see Fig. 2. The sighting angle of a star relative to the plane of Earth’s orbit (usually indicated by θ) can be determined at two different times of the year separated by six months. Since we know the distance d from the Earth to the Sun, we can determine the distance D to the star. For example, if the angle θ of a given star is measured to be 89.99994° , the parallax angle is $p \equiv \phi = 0.00006^\circ$. From trigonometry, $\tan \phi = d/D$, and since the distance to the Sun is $d = 1.5 \times 10^8$ km the distance to the star is

$$D = \frac{d}{\tan \phi} \approx \frac{d}{\phi} = \frac{1.5 \times 10^8 \text{ km}}{1 \times 10^{-6}} = 1.5 \times 10^{14} \text{ km}, \quad (5)$$

or about 15 ly.

Distances to stars are often specified in terms of parallax angles given in seconds of arc: 1 second ($1''$) is $1/60$ of a minute ($1'$) of arc, which is $1/60$ of a degree, so $1'' = 1/3600$ of a degree. The distance is then specified in parsecs (meaning *parallax angle in seconds of arc*), where the parsec is defined as $1/\phi$ with ϕ in seconds. For example, if $\phi = 6 \times 10^{-5}^\circ$, we would say the the star is at a distance $D = 4.5$ pc.

EXERCISE 2.1 Using the definitions of parsec and light year, show that $1 \text{ pc} = 3.26 \text{ ly}$.

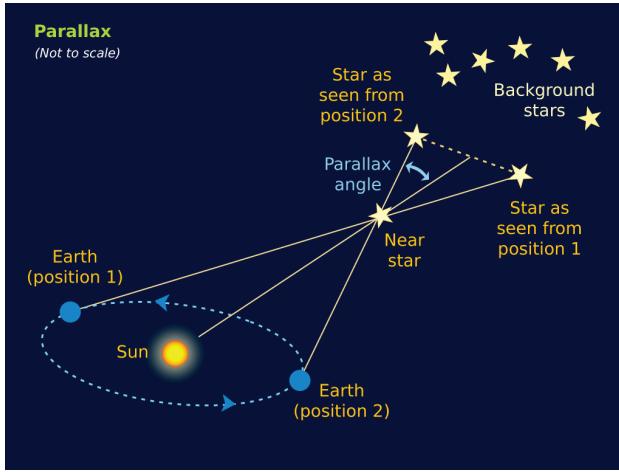


FIG. 2: The parallax method of measuring a star's distance.

The angular resolution of the Hubble Space Telescope (HST) is about 1/20 arcseconds. With HST one can measure parallaxes of about 2 milli arc seconds (e.g., 1223 Sgr). This corresponds to a distance of about 500 pc. Besides, there are stars with radio emission for which observations from the Very Long Baseline Array (VLBA) allow accurate parallax measurements beyond 500 pc. For example, parallax measurements of Sco X-1 are 0.36 ± 0.04 milli arc seconds which puts it at a distance of 2.8 kpc. Parallax can be used to determine the distance to stars as far away as about 3 kpc from Earth. Beyond that distance, parallax angles are too small to measure and more subtle techniques must be employed.

EXERCISE 2.2 One of the first people to make a very accurate measurement of the circumference of the Earth was Eratosthenes, a Greek philosopher who lived in Alexandria around 250 B.C. He was told that on a certain day during the summer (June 21) in a town called Syene, which was 4900 stadia (1 stadia = 0.16 kilometers) to the south of Alexandria, the sunlight shown directly down the well shafts so that you could see all the way to the bottom. Eratosthenes knew that the sun was never quite high enough in the sky to see the bottom of wells in Alexandria and he was able to calculate that in fact it was about 7 degrees too low. Knowing that the sun was 7 degrees lower at its highpoint in Alexandria than in Syene and assuming that the sun's rays were parallel when they hit the Earth, Eratosthenes was able to calculate the circumference of the Earth using a simple proportion: $C/4900$ stadia = 360 degrees/ 7 degrees. This gives an answer of 252,000 stadia or 40,320 km, which is very close to today's measurements of 40,030 km. Assume the Earth is flat and determine the parallax angle that can explain this phenomenon. Are the results consistent with the hypothesis that the Earth is flat?

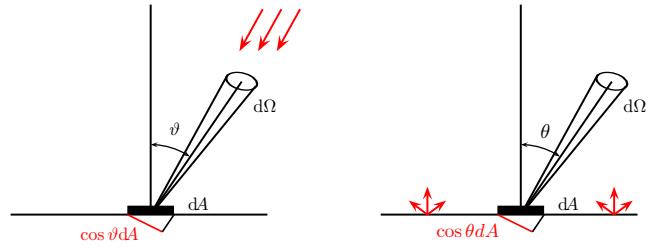


FIG. 3: **Left.** A detector with surface element dA on Earth measuring radiation coming from a direction with zenith angle ϑ (left). **Right.** An imaginary detector of area dA on the surface of a star measuring radiation emitted in the direction θ .

B. Stellar luminosity

In 1900, Planck found empirically the distribution

$$B_\nu d\nu = \frac{2h\nu^3}{c^2} \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]^{-1} d\nu \quad (6)$$

describing the amount of energy emitted into the frequency interval $[\nu, \nu + d\nu]$ and the solid angle $d\Omega$ per unit time and area by a body in thermal equilibrium [8]. The *intrinsic* (or surface) brightness B_ν depends only on the temperature T of the blackbody (apart from the natural constants k , c and h). The dimension of B_ν in the cgs system of units is

$$[B_\nu] = \frac{\text{erg}}{\text{Hz cm}^2 \text{s sr}}. \quad (7)$$

In general the amount of energy per frequency interval $[\nu, \nu + d\nu]$ and solid angle $d\Omega$ crossing the perpendicular area A_\perp per time is called the specific (or differential) intensity [9]

$$I_\nu = \frac{dE}{d\nu d\Omega dA_\perp dt}; \quad (8)$$

see Fig. 3. For the special case of the blackbody radiation, the specific intensity at the emission surface is given by the Planck distribution, $I_\nu = B_\nu$. Stars are fairly good approximations of blackbodies.

Integrating (6) over all frequencies and possible solid angles gives the emitted flux F per surface area A . The angular integral consists of the solid angle $d\Omega = \sin\theta d\theta d\phi$ and the factor $\cos\theta$ taking into account that only the perpendicular area $A_\perp = A \cos\theta$ is visible [10]. The flux emitted by a star is found to be

$$F = \pi \int_0^\infty d\nu B_\nu = \frac{2\pi}{c^2 h^3} (kT)^4 \int_0^\infty \frac{x^3 dx}{e^x - 1} = \sigma T^4, \quad (9)$$

where $x = h\nu/(kT)$,

$$\sigma = \frac{2\pi^5 k^4}{15 c^2 h^3} = 5.670 \times 10^{-5} \frac{\text{erg}}{\text{cm}^2 \text{K}^4 \text{s}} \quad (10)$$

is the Stefan-Boltzmann constant [11, 12], and where we used $\int_0^\infty x^3 [e^x - 1]^{-1} dx = \pi^4/15$.

A useful parameter for a star or galaxy is its luminosity. The total luminosity L of a star is given by the product of its surface area and the radiation emitted per area

$$L = 4\pi R^2 \sigma T^4. \quad (11)$$

Careful analyses of nearby stars have shown that the absolute luminosity for most of the stars depends on the mass: *the more massive the star, the greater the luminosity.*

Consider a thick spherical source of radius R , with constant intensity along the surface, say a star. An observer at a distance r sees the spherical source as a disk of angular radius $\vartheta_c = R/r$. Note that since the source is optically thick the observer only sees the surface of the sphere. Because the intensity is constant over the surface there is a symmetry along the φ direction such that the solid solid angle is given by $d\Omega = 2\pi \sin \vartheta d\vartheta$. By looking at Fig. 3 it is straightforward to see that the flux observed at r is given by

$$\begin{aligned} F(r) &= \int I \cos \vartheta d\Omega = 2\pi I \int_0^{\vartheta_c} \sin \vartheta \cos \vartheta d\vartheta \\ &= \pi I \cos^2 \vartheta \Big|_{\vartheta_c}^0 = \pi I \sin^2 \vartheta_c = \pi I (R/r)^2. \end{aligned} \quad (12)$$

At the surface of the star $R = r$ and we recover (9). Very far away, $r \gg R$, and (12) yields $F = \pi \vartheta_c^2 I = I \Omega_{\text{source}}$; see Appendix A. The validity of the inverse-square law $F \propto 1/r^2$ at a distance $r > R$ outside of the star relies on the assumptions that no radiation is absorbed and that relativistic effects can be neglected. The later condition requires in particular that the relative velocity of observer and source is small compared to the speed of light. All in all, the total (integrated) flux at the surface of the Earth from a given astronomical object with total luminosity L is found to be

$$F_{\text{observed @ Earth}} = \mathcal{F} = \frac{L}{4\pi d_L^2}, \quad (13)$$

where d_L is the distance to the object.

Another important parameter of a star is its surface temperature, which can be determined from the spectrum of electromagnetic frequencies it emits. The wavelength at the peak of the spectrum, λ_{max} , is related to the temperature by Wien's displacement law [13]

$$\lambda_{\text{max}} T = 2.9 \times 10^{-3} \text{ m K}. \quad (14)$$

We can now use Wien's law and the Stefan-Boltzmann equation (power output or luminosity $\propto AT^4$) to determine the temperature and the relative size of a star. Suppose that the distance from Earth to two nearby stars can be reasonably estimated, and that their apparent luminosities suggest the two stars have about the same absolute luminosity, L . The spectrum of one of the stars peaks at about 700 nm (so it is reddish). The spectrum of

the other peaks at about 350 nm (bluish). Using Wien's law, the temperature of the reddish star is $T_r \simeq 4140$ K. The temperature of the bluish star will be double because its peak wavelength is half, $T_b \simeq 8280$ K. The power radiated per unit of area from a star is proportional to the fourth power of the Kelvin temperature (11). Now the temperature of the bluish star is double that of the reddish star, so the bluish must radiate 16 times as much energy per unit area. But we are given that they have the same luminosity, so the surface area of the blue star must be 1/16 that of the red one. Since the surface area is $4\pi R^2$, we conclude that the radius of the reddish star is 4 times larger than the radius of the bluish star (and its volume 64 times larger) [14].

An important astronomical discovery, made around 1900, was that for most of the stars, the color is related to the absolute luminosity and therefore to the mass. A useful way to present this relationship is by the so-called Hertzsprung-Russell (HR) diagram [15]. On the HR diagram, the horizontal axis shows the temperature T , whereas the vertical axis the luminosity L , each star is represented by a point on the diagram shown in Fig. 4. Most of the stars fall along the diagonal band termed the main sequence. Starting at the lowest right, we find the coolest stars, reddish in color; they are the least luminous and therefore low in mass. Further up towards the left we find hotter and more luminous stars that are whitish like our Sun. Still farther up we find more massive and more luminous stars, bluish in color. There are also stars that fall outside the main sequence. Above and to the right we find extremely large stars, with high luminosity but with low (reddish) color temperature: these are called red giants. At the lower left, there are a few stars of low luminosity but with high temperature: these are white dwarfs.

Suppose that a detailed study of a certain star suggests that it most likely fits on the main sequence of the HR diagram. The observed flux is $\mathcal{F} = 1 \times 10^{-12} \text{ W m}^{-2}$, and the peak wavelength of its spectrum is $\lambda_{\text{max}} \approx 600 \text{ nm}$. We can first find the temperature using Wien's law and then estimate the absolute luminosity using the HR diagram; namely, $T \approx 4800$ K. A star on the main sequence of the HR diagram at this temperature has absolute luminosity of about $L \approx 10^{26} \text{ W}$. Then, using (13) we can estimate its distance from us, $d_L = 3 \times 10^{18} \text{ m}$ or equivalently 300 ly.

EXERCISE 2.3 About 1350 J of energy strikes the atmosphere of the Earth from the Sun per second per square meter of area at right angle to the Sun's rays. What is (i) the observed flux from the Sun \mathcal{F}_\odot and (ii) its absolute luminosity L_\odot .

EXERCISE 2.4 Estimate the angular width that our Galaxy would subtend if observed from Andromeda. Compare to the angular width of the Moon from the Earth.

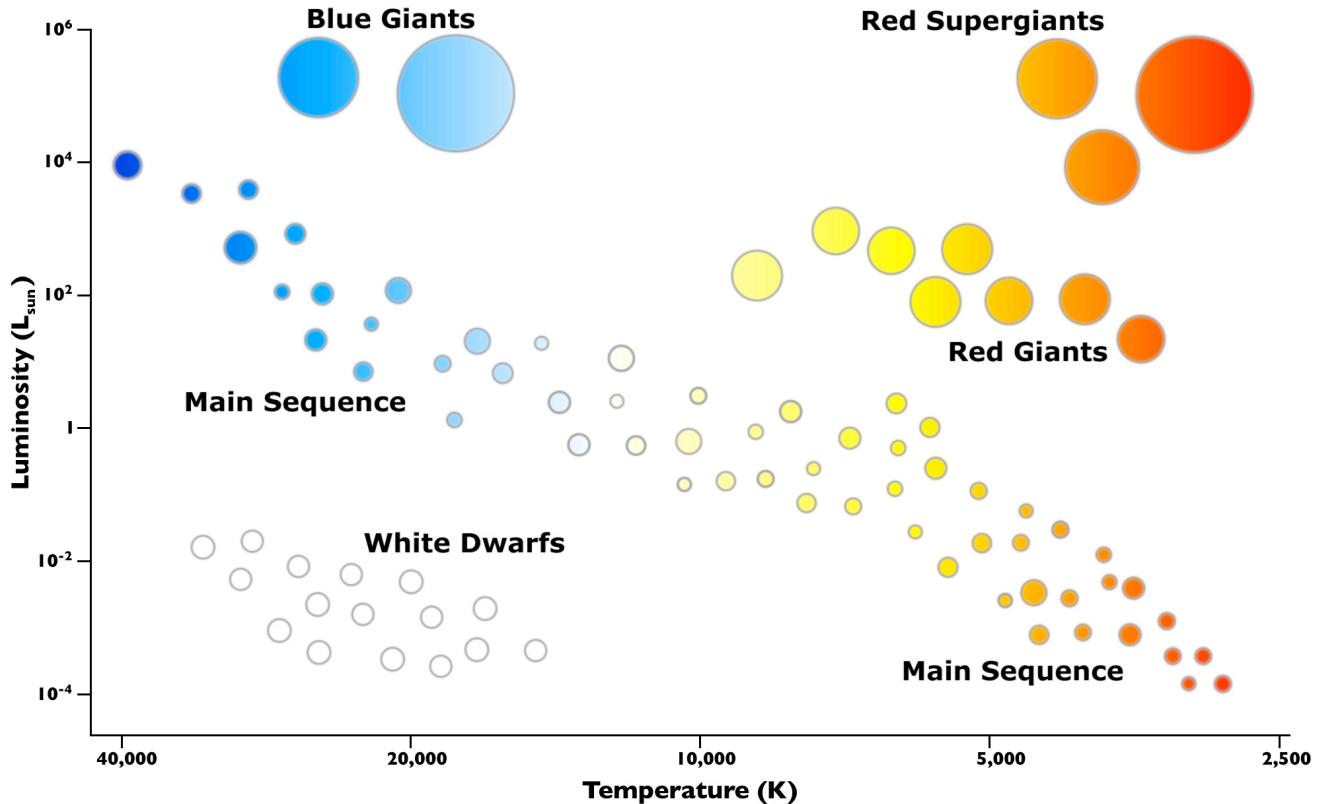


FIG. 4: HR diagram. The vertical axis depicts the inherent brightness of a star, and the horizontal axis the surface temperature increasing from right to left.

EXERCISE 2.5 The brightest star in the sky is Sirius, which is located at a distance of 2.6 pc. If a planet were orbiting Sirius at the same distance that the Earth is from the Sun, what would be the angular separation on the sky between Sirius and this hypothesized planet?

EXERCISE 2.6 Suppose the MESSENGER spacecraft, while orbiting Mercury, decided to communicate with the Cassini probe, now exploring Saturn and its moons. When Mercury is closest to Saturn in their orbits, it takes 76.3 minutes for the radio signals from Mercury to reach Saturn. A little more than half a mercurian year later, when the 2 planets are furthest apart in their orbits, it takes 82.7 minutes. (i) What is the distance between Mercury and the Sun? Give answers in both light-minutes and astronomical units. Assume that the planets have circular orbits. (ii) What is the distance between Saturn and the Sun?

EXERCISE 2.7 The photometric method to search for extrasolar planets is based on the detection of stellar brightness variations, which result from the transit of a planet across a star's disk. If a planet passes in front of a star, the star will be partially eclipsed and its light will be dimmed. Determine the reduction in the *apparent* surface brightness I when Jupiter passes in front of the

Sun.

EXERCISE 2.8 The angular resolution of a telescope (or other optical system) is a measure of the smallest details which can be seen. Because of the distorting effects of earth's atmosphere, the best angular resolution which can be achieved by optical telescopes from earth's surface is normally about 1 arcs. This is why much clearer images can be obtained from space. The angular resolution of the HST is about 0.05 arcs, and the smallest angle that can be measured accurately with HST is actually a fraction of one resolution element. Cepheid variable stars are very important distance indicators because they have large and well-known luminosities. What is the distance of a Cepheid variable star whose parallax angle is measured to be 0.005 ± 0.001 arcs?

EXERCISE 2.9 The faintest stars that can be detected with the HST have apparent brightnesses which are 4×10^{21} times fainter than the Sun. (i) How far away could a star like the Sun be, and still be detected with the HST? Express your answer in light years. (ii) How far away could a Cepheid variable with 20,000 times the luminosity of the Sun be, and still be detected with the HST? Express your answer in light years.

EXERCISE 2.10 A perfect blackbody at temperature T has the shape of an oblate ellipsoid, its surface being given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{b^2} = 1, \quad (15)$$

with $a > b$. (i) Is the luminosity of the blackbody isotropic? Why? (ii) Consider an observer at a distance d_L from the blackbody, with $d_L \gg a$. What is the direction of the observer for which the maximum amount of flux will be observed (keeping the distance d_L fixed)? Calculate what this maximum flux is. (iii) Repeat the same exercise for the direction for which the minimum flux will be observed, for fixed d_L . (iv) If the two observers who see the maximum and minimum flux from distance d_L can resolve the blackbody, what is the *apparent* brightness, I , that each one will measure? (v) Write down an expression for the total luminosity emitted by the black body as a function of a , b and T . (vi) Now, consider a galaxy with a perfectly oblate shape, which contains only a large number N of stars, and no gas or dust. To make it simple, assume that all stars have radius R and surface temperature T . Answer again the questions (i-v) for the galaxy, assuming $NR^2 \ll ab$. Are there any differences from the case of a blackbody? Explain why. (vii) Imagine that there were a very compact galaxy that did not obey the condition $NR^2 \ll ab$. Would the answer to the previous question be modified? Do you think such a galaxy could be stable?

EXERCISE 2.11 The HR diagram is usually plotted in logarithmic coordinates ($\log L$ vs. $\log T$, with the temperature increasing to the left). Derive the slope of a line of constant radius in the logarithmic HR diagram.

III. DOPPLER EFFECT

There is observational evidence that stars move at speeds ranging up to a few hundred kilometers per second, so in a year a fast moving star might travel $\sim 10^{10}$ km. This is 10^3 times less than the distance to the closest star, so their apparent position in the sky changes very slowly. For example, the relatively fast moving star known as Barnard's star is at a distance of about 56×10^{12} km; it moves across the line of sight at about 89 km/s, and in consequence its apparent position shifts (so-called “proper motion”) in one year by an angle of 0.0029 degrees. The HST has measured proper motions as low as about 1 milli arc second per year. In the radio (VLBA), relative motions can be measured to an accuracy of about 0.2 milli arc second per year. The apparent position in the sky of the more distant stars changes so slowly that their proper motion cannot be detected with even the most patient observation. However, the rate of approach or recession of a luminous body in the line of sight can be measured much more accurately than its motion at right angles to the line of sight. The technique

makes use of a familiar property of any sort of wave motion, known as Doppler effect [16].

When we observe a sound or light wave from a source at rest, the time between the arrival wave crests at our instruments is the same as the time between crests as they leave the source. However, if the source is moving away from us, the time between arrivals of successive wave crests is increased over the time between their departures from the source, because each crest has a little farther to go on its journey to us than the crest before. The time between crests is just the wavelength divided by the speed of the wave, so a wave sent out by a source moving away from us will appear to have a longer wavelength than if the source were at rest. Likewise, if the source is moving toward us, the time between arrivals of the wave crests is decreased because each successive crest has a shorter distance to go, and the waves appear to have a shorter wavelength. A nice analogy was put forward by Weinberg [17]. He compared the situation with a travelling man that has to send a letter home regularly once a week during his travels: while he is travelling away from home, each successive letter will have a little farther to go than the one before, so his letters will arrive a little more than a week apart; on the homeward leg of his journey, each successive letter will have a shorter distance to travel, so they will arrive more frequently than once a week.

The Doppler effect began to be of enormous importance to astronomy in 1968, when it was applied to the study of individual spectral lines. In 1815, Fraunhofer first realized that when light from the Sun is allowed to pass through a slit and then through a glass prism, the resulting spectrum of colors is crossed with hundreds of dark lines, each one an image of the slit [18]. The dark lines were always found at the same colors, each corresponding to a definite wavelength of light. The same dark spectral lines were also found in the same position in the spectrum of the Moon and brighter stars. It was soon realized that these dark lines are produced by the selective absorption of light of certain definite wavelengths, as light passes from the hot surface of a star through its cooler outer atmosphere. Each line is due to absorption of light by a specific chemical element, so it became possible to determine that the elements on the Sun, such as sodium, iron, magnesium, calcium, and chromium, are the same as those found on Earth.

In 1868, Sir Huggins was able to show that the dark lines in the spectra of some of the brighter stars are shifted slightly to the red or the blue from their normal position in the spectrum of the Sun [19]. He correctly interpreted this as a Doppler shift, due to the motion of the star away from or toward the Earth. For example, the wavelength of every dark line in the spectrum of the star Capella is longer than the wavelength of the corresponding dark line in the spectrum of the Sun by 0.01%, this shift to the red indicates that Capella is receding from us at 0.01% c (i.e., the radial velocity of Capella is about 30 km/s).

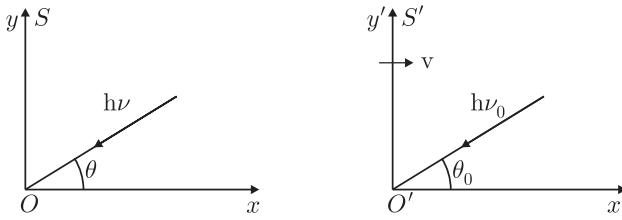


FIG. 5: A source of light waves moving to the right, relative to observers in the S frame, with velocity v . The frequency is higher for observers on the right, and lower for observers on the left.

Consider two inertial frames, S and S' , moving with relative velocity v as shown in Fig. 5. Assume a light source (e.g. a star) at rest in S' emits light of frequency ν_0 at an angle θ_0 with respect to the observer O' . Let

$$p^\mu = \left(\frac{h\nu}{c}, -\frac{h\nu}{c} \cos \theta, -\frac{h\nu}{c} \sin \theta, 0 \right) \quad (16)$$

be the momentum 4-vector for the photon as seen in S and

$$p_0^\mu = \left(\frac{h\nu_0}{c}, -\frac{h\nu_0}{c} \cos \theta_0, -\frac{h\nu_0}{c} \sin \theta_0, 0 \right) \quad (17)$$

in S' . To get the 4-momentum relation from $S' \rightarrow S$, apply the inverse Lorentz transformation [20]

$$\begin{aligned} \frac{h\nu}{c} &= \gamma \left[\frac{h\nu_0}{c} + \beta \left(-\frac{h\nu_0}{c} \cos \theta_0 \right) \right] \\ -\frac{h\nu}{c} \cos \theta &= \gamma \left(-\frac{h\nu_0}{c} \cos \theta_0 + \beta \frac{h\nu_0}{c} \right) \\ \frac{h\nu}{c} \sin \theta &= \frac{h\nu_0}{c} \sin \theta_0. \end{aligned} \quad (18)$$

The first expression gives

$$\nu = \nu_0 \gamma (1 - \beta \cos \theta_0), \quad (19)$$

which is the relativistic Doppler formula.

For observational astronomy (19) is not useful because both ν_0 and θ_0 refer to the star's frame, not that of the observer. Apply instead the direct Lorentz transformation $S \rightarrow S'$ to the photon energy to obtain

$$\nu_0 = \gamma \nu (1 + \beta \cos \theta). \quad (20)$$

This equation gives ν_0 in terms of quantities measured by the observer. It is sometimes written in terms of wavelengths: $\lambda = \lambda_0 \gamma (1 + \beta \cos \theta)$. (For details see e.g. [21].)

EXERCISE 3.1 Consider the inertial frames S and S' shown in Fig. 5. Use the inverse Lorentz transformation to show that the relation between angles is given by

$$\cos \theta = \frac{\beta - \cos \theta_0}{\beta \cos \theta_0 - 1}. \quad (21)$$

There are three special cases: (i) $\theta_0 = 0$, which gives

$$\nu = \nu_0 \sqrt{(1 - \beta)/(1 + \beta)}. \quad (22)$$

In the non-relativistic limit we have $\nu = \nu_0(1 - \beta)$. This corresponds to a source moving away from the observer. Note that $\theta = 0$. (ii) $\theta_0 = \pi$, which gives

$$\nu = \nu_0 \sqrt{(1 + \beta)/(1 - \beta)}. \quad (23)$$

Here the source is moving towards the observer. Note that $\theta = \pi$. (iii) $\theta_0 = \pi/2$, which gives

$$\nu = \nu_0 \gamma. \quad (24)$$

This last is the transverse Doppler effect – a second order relativistic effect. It can be thought of as arising from the dilation of time in the moving frame.

EXERCISE 3.2 Suppose light is emitted isotropically in a star's rest frame S' , i.e. $dN/d\Omega_0 = \varkappa$, where dN is the number of photons in the solid angle $d\Omega_0$ and \varkappa is a constant. What is the angular distribution in the inertial frame S ?

EXERCISE 3.3 Show that for $v \ll c$, the Doppler shift in wavelength is

$$\frac{\lambda' - \lambda}{\lambda} \approx \frac{v}{c}. \quad (25)$$

To avoid confusion, it should be kept in mind that λ denotes the wavelength of the light if observed near the place and time of emission, and thus presumably take the values measured when the same atomic transition occurs in terrestrial laboratories, while λ' is the wavelength of the light observed after its long journey to us. If $\lambda' - \lambda > 0$ then $\lambda' > \lambda$ and we speak of a redshift; if $\lambda' - \lambda < 0$ then $\lambda' < \lambda$, and we speak of a blueshift.

EXERCISE 3.4 Through some coincidence, the Balmer lines from single ionized helium in a distant star happen to overlap with the Balmer lines from hydrogen in the Sun. How fast is that star receding from us? [Hint: the wavelengths from single-electron energy level transitions are inversely proportional to the square of the atomic number of the nucleus.]

EXERCISE 3.5 Stellar aberration is the apparent motion of a star due to rotation of the Earth about the Sun. Consider an incoming photon from a star with 4-momentum p^μ . Let S be the Sun's frame and S' the Earth frame moving with velocity v as shown in Fig. 6. Define the angle of aberration α by $\theta' = \theta - \alpha$ and show that $\alpha \approx \beta \sin \theta$.

EXERCISE 3.6 HD 209458 is a star in the constellation Pegasus very similar to our Sun ($M = 1.1M_\odot$ and $R = 1.1R_\odot$), located at a distance of about 150 ly. In

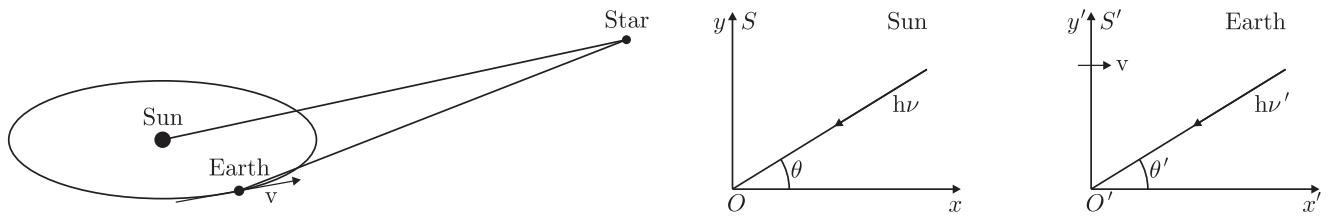


FIG. 6: Schematic representation of stellar aberration.

1999, two teams working independently discovered an extrasolar planet orbiting the star using the so-called “radial velocity planet search method” [282, 283]. Note that a star with a planet must move in its own small orbit in response to the planet’s gravity. This leads to variations in the speed with which the star moves toward or away from Earth, i.e. the variations are in the radial velocity of the star with respect to Earth. The radial velocity can be deduced from the displacement in the parent star’s spectral lines due to the Doppler shift. If a planet orbits the star, one should have a periodic change in that rate, except for the extreme case in which the plane of the orbit is perpendicular to our line of sight. Herein we assume that the motions of the Earth relative to the Sun have already been taken into account, as well as any long-term steady change of distance between the star and the sun, which appears as a median line for the periodic variation in radial velocity due to the star’s “wobble” caused by the orbiting planet. The observed Doppler shift velocity of HD 209458 is found to be $K = V \sin i = 82.7 \pm 1.3$ m/s, where $i = 87.1^\circ \pm 0.2^\circ$ is the inclination of the planet’s orbit to the line perpendicular to the line-of-sight. [284]. Soon after the discovery, separate teams were able to detect a transit of the planet across the surface of the star making it the first known transiting extrasolar planet [285, 286]. The planet received the designation HD 209458b. Because the planet transits the star, the star is dimmed by about 2% every 3.52447 ± 0.00029 days. Tests allowing for a non-circular Keplerian orbit for HD 209458 resulted in an eccentricity indistinguishable from zero: $e = 0.016 \pm 0.018$. Consider the simplest case of a nearly circular orbit and find: (i) the distance from the planet to the star; (ii) the mass m of the planet; (iii) the radius r of the planet.

IV. STELLAR EVOLUTION

The stars appear unchanging. Night after night the heavens reveal no significant variations. Indeed, on human time scales, the vast majority of stars change very little. Consequently, we cannot follow any but the tiniest part of the life cycle of any given star since they live for ages vastly greater than ours. Nonetheless, herein we will follow the process of stellar evolution from the birth to the death of a star, as we have theoretically reconstructed it.

A. Stellar nucleosynthesis

There is a general consensus that stars are born when gaseous clouds (mostly hydrogen) contract due to the pull of gravity. A huge gas cloud might fragment into numerous contracting masses, each mass centered in an area where the density is only slightly greater than at nearby points. Once such “globules” formed, gravity would cause each to contract in towards its center-of-mass. As the particles of such protostar accelerate inward, their kinetic energy increases. When the kinetic energy is sufficiently high, the Coulomb repulsion between the positive charges is not strong enough to keep hydrogen nuclei apart, and nuclear fusion can take place. In a star like our Sun, the “burning” of hydrogen occurs when four protons fuse to form a helium nucleus, with the release of γ rays, positrons and neutrinos.¹

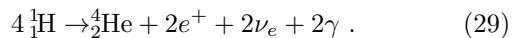
The energy output of our Sun is believed to be due principally to the following sequence of fusion reactions:



and



where the energy released for each reaction (given in parentheses) equals the difference in mass (times c^2) between the initial and final states. Such a released energy is carried off by the outgoing particles. The net effect of this sequence, which is called the *pp*-cycle, is for four protons to combine to form one ${}^4\text{He}$ nucleus, plus two positrons, two neutrinos, and two gamma rays:

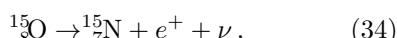
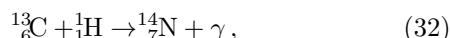
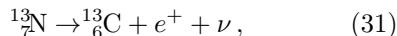
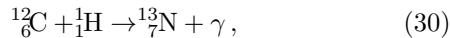


Note that it takes two of each of the first two reactions to produce the two ${}^3\text{He}$ for the third reaction. So the total

¹ The word “burn” is put in quotation marks because these high-temperature fusion reactions occur via a *nuclear* process, and must not be confused with ordinary burning in air, which is a *chemical* reaction, occurring at the *atomic* level (and at a much lower temperature).

energy released for the net reaction is 24.7 MeV. However, each of the two e^+ quickly annihilates with an electron to produce $2m_e c^2 = 1.02$ MeV; so the total energy released is 26.7 MeV. The first reaction, the formation of deuterium from two protons, has very low probability, and the infrequency of that reaction serves to limit the rate at which the Sun produces energy. These reactions require a temperature of about 10^7 K, corresponding to an average kinetic energy (kT) of 1 keV.

In more massive stars, it is more likely that the energy output comes principally from the carbon (or CNO) cycle, which comprises the following sequence of reactions:



It is easily seen that no carbon is consumed in this cycle (see first and last equations) and that the net effect is the same as the *pp* cycle. The theory of the *pp* cycle and the carbon cycle as the source of energy for the Sun and the stars was first worked out by Bethe in 1939 [22].

The fusion reactions take place primarily in the core of the star, where T is sufficiently high. (The surface temperature is of course much lower, on the order of a few thousand K.) The tremendous release of energy in these fusion reactions produces an outward pressure sufficient to halt the inward gravitational contraction; and our protostar, now really a young star, stabilizes in the main sequence.

To a good approximation the stellar structure on the main sequence can be described by a spherically symmetric system in hydrostatic equilibrium. This requires that rotation, convection, magnetic fields, and other effects that break rotational symmetry have only a minor influence on the star. This assumption is in most cases very well justified.

We denote by $M(r)$ the mass enclosed inside a sphere with radius r and density $\rho(r)$

$$M(r) = 4\pi \int_0^r dr' r'^2 \rho(r') \quad (36)$$

or

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r). \quad (37)$$

An important application of (37) is to express physical quantities not as function of the radius r but of the enclosed mass $M(r)$. This facilitates the computation of the

stellar properties as function of time, because the mass of a star remains nearly constant during its evolution, while the stellar radius can change considerably.

A radial-symmetric mass distribution $M(r)$ produces according Gauss law the same gravitational acceleration, as if it would be concentrated at the center $r = 0$. Therefore the gravitational acceleration produced by $M(r)$ is

$$g(r) = -\frac{GM(r)}{r^2}. \quad (38)$$

If the star is in equilibrium, this acceleration is balanced by a pressure gradient from the center of the star to its surface. Since pressure is defined as force per area, $P = F/A$, a pressure change along the distance dr corresponds to an increment

$$\begin{aligned} dF &= dAP - (P + dP)dA \\ &= -\underbrace{dAdP}_{\text{force}} = -\underbrace{\rho(r)dAdr}_{\text{mass}} \underbrace{a(r)}_{\text{acceleration}} \end{aligned} \quad (39)$$

of the force F produced by the pressure gradient dP . For increasing r , the gradient $dP < 0$ and the resulting force dF is positive and therefore directed outward. Hydrostatic equilibrium, $g(r) = -a(r)$, requires then

$$\frac{dP}{dr} = \rho(r)g(r) = -\frac{GM(r)}{r^2}\rho(r). \quad (40)$$

If the pressure gradient and gravity do not balance each other, the layer at position r is accelerated,

$$a(r) = \frac{GM(r)}{r^2} + \frac{1}{\rho(r)} \frac{dP}{dr}. \quad (41)$$

In general, we need an equation of state, $P = P(\rho, T, Y_i)$, that connects the pressure P with the density ρ , the (not yet) known temperature T and the chemical composition Y_i of the star. For an estimate of the central pressure $P_c = P(0)$ of a star in hydrostatic equilibrium, we integrate (40) and obtain with $P(R) \approx 0$,

$$P_c = \int_0^R \frac{dP}{dr} dr = G \int_0^M dM \frac{M}{4\pi r^4}, \quad (42)$$

where we used the continuity equation (37) to substitute $dr = dM/(4\pi r^2 \rho)$ by dM . If we replace furthermore r by the stellar radius $R \geq r$, we obtain a lower limit for the central pressure,

$$\begin{aligned} P_c &= G \int_0^M dM \frac{M}{4\pi r^4} \\ &> G \int_0^M dM \frac{M}{4\pi R^4} = \frac{M^2}{8\pi R^4}. \end{aligned} \quad (43)$$

Inserting values for the Sun, it follows

$$P_c > \frac{M^2}{8\pi R^4} = 4 \times 10^8 \text{ bar} \left(\frac{M}{M_\odot} \right)^2 \left(\frac{R_\odot}{R} \right)^4. \quad (44)$$

The value obtained integrating the hydrostatic equation using the “solar standard model” is $P_c = 2.48 \times 10^{11}$ bar, i.e. a factor 500 larger.

EXERCISE 4.1 Calculate the central pressure P_c of a star in hydrostatic equilibrium as a function of its mass M and radius R for (i) a constant mass density, $\rho(r) = \rho_0$ and (ii) a linearly decreasing mass density, $\rho(r) = \rho_c[1 - (r/R)]$.

Exactly where the star falls along the main sequence depends on its mass. The more massive the star, the further up (and to the left) it falls in the HR diagram. To reach the main sequence requires perhaps 30 million years and the star is expected to remain there 10 billion years (10^{10} yr). Although most of stars are billions of years old, there is evidence that stars are actually being born at this moment in the Eagle Nebula.

As hydrogen fuses to form helium, the helium that is formed is denser and tends to accumulate in the central core where it was formed. As the core of helium grows, hydrogen continues to fuse in a shell around it. When much of the hydrogen within the core has been consumed, the production of energy decreases at the center and is no longer sufficient to prevent the huge gravitational forces from once again causing the core to contract and heat up. The hydrogen in the shell around the core then fuses even more fiercely because of the rise in temperature, causing the outer envelope of the star to expand and to cool. The surface temperature thus reduced, produces a spectrum of light that peaks at longer wavelength (reddish). By this time the star has left the main sequence. It has become redder, and as it has grown in size, it has become more luminous. Therefore, it will have moved to the right and upward on the HR diagram. As it moves upward, it enters the red giant stage. This model then explains the origin of red giants as a natural step in stellar evolution. Our Sun, for example, has been on the main sequence for about four and a half billion years. It will probably remain there another 4 or 5 billion years. When our Sun leaves the main sequence, it is expected to grow in size (as it becomes a red giant) until it occupies all the volume out to roughly the present orbit of the planet Mercury.

If the star is like our Sun, or larger, further fusion can occur. As the star’s outer envelope expands, its core is shrinking and heating up. When the temperature reaches about 10^8 K, even helium nuclei, in spite of their greater charge and hence greater electrical repulsion, can then reach each other and undergo fusion:



Once beryllium-8 is produced a little faster than it decays (half-life is 6.7×10^{-17} s), the number of beryllium-8 nuclei in the stellar core increases to a large number. Then in its core there will be many beryllium-8 nuclei that can fuse with another helium nucleus to form carbon-12, which is stable:



The net energy release of the triple- α process is 7.273 MeV. Further fusion reactions are possible, with ${}_2^4\text{He}$ fusing with ${}_6^{12}\text{C}$ to form ${}_{16}^{16}\text{O}$. Stars spend approximately a few thousand to 1 billion years as a red giant. Eventually, the helium in the core runs out and fusion stops. Stars with $0.4M_\odot < M < 4M_\odot$ are fated to end up as spheres of carbon and oxygen. Only stars with $M > 4M_\odot$ become hot enough for fusion of carbon and oxygen to occur and higher Z elements like ${}_{10}^{20}\text{Ne}$ or ${}_{12}^{24}\text{Mg}$ can be made.

As massive ($M > 8M_\odot$) red supergiants age, they produce “onion layers” of heavier and heavier elements in their interiors. A star of this mass can contract under gravity and heat up even further, ($T = 5 \times 10^9$ K), producing nuclei as heavy as ${}_{26}^{56}\text{Fe}$ and ${}_{28}^{56}\text{Ni}$. However, the average binding energy per nucleon begins to decrease beyond the iron group of isotopes. Thus, the formation of heavy nuclei from lighter ones by fusion ends at the iron group. Further fusion would require energy, rather than release it. As a consequence, a core of iron builds up in the centers of massive supergiants.

A star’s lifetime as a giant or supergiant is shorter than its main sequence lifetime (about 1/10 as long). As the star’s core becomes hotter, and the fusion reactions powering it become less efficient, each new fusion fuel is used up in a shorter time. For example, the stages in the life of a $25M_\odot$ star are as follows:: hydrogen fusion lasts 7 million years, helium fusion lasts 500,000 years, carbon fusion lasts 600 years, neon fusion lasts 1 year, oxygen fusion lasts 6 months, and silicon fusion lasts 1 day. The star core is now pure iron. The process of creating heavier nuclei from lighter ones, or by absorption of neutrons at higher Z (more on this below) is called nucleosynthesis.

B. White dwarfs and Chandrasekhar limit

At a distance of 2.6 pc Sirius is the fifth closest stellar system to the Sun. It is the brightest star in the Earth’s night sky. Analyzing the motions of Sirius from 1833 to 1844, Bessel concluded that it had an unseen companion, with an orbital period $T \sim 50$ yr [42]. In 1862, Clark discovered this companion, Sirius B, at the time of maximal separation of the two components of the binary system (i.e. at apastron) [43]. Complementary follow up observations showed that the mass of Sirius B equals approximately that of the Sun, $M \approx M_\odot$. Sirius B’s peculiar properties were not established until the next apastron by Adams [44]. He noted that its high temperature ($T \simeq 25,000$ K) together with its small luminosity ($L = 3.84 \times 10^{26}$ W) require an extremely small radius and thus a large density. From Stefan-Boltzmann law we have

$$\frac{R}{R_\odot} = \left(\frac{L}{L_\odot} \right)^{1/2} \left(\frac{T}{T_\odot} \right)^2 \approx 10^{-2}. \quad (47)$$

Hence, the mean density of Sirius B is a factor 10^6 higher than that of the Sun; more precisely, $\rho = 2 \times 10^6 \text{ g/cm}^3$.

A lower limit for the central pressure of Sirius B follows from (44)

$$P_c > \frac{M^2}{8\pi R^4} = 4 \times 10^{16} \text{ bar}. \quad (48)$$

Assuming the pressure is dominated by an ideal gas the central temperature is found to be

$$T_c = \frac{P_c}{nk} \sim 10^2 T_{c,\odot} \approx 10^9 \text{ K}. \quad (49)$$

For such a high T_c , the temperature gradient dT/dr in Sirius B would be a factor 10^4 larger than in the Sun. This would in turn require a larger luminosity and a larger energy production rate than that of main sequence stars.

Stars like Sirius B are called white dwarfs. They have very long cooling times, because of their small surface luminosity. This type of stars is rather numerous. The mass density of main-sequence stars in the solar neighborhood is $0.04M_\odot/\text{pc}^3$ compared to $0.015M_\odot/\text{pc}^3$ in white dwarfs. The typical mass of white dwarfs lies in the range $0.4 \lesssim M/M_\odot \lesssim 1$, peaking at $0.6M_\odot$. No further fusion energy can be obtained inside a white dwarf. The star loses internal energy by radiation, decreasing in temperature and becoming dimmer until its light goes out.

For a classical gas, $P = nkT$, and thus in the limit of zero temperature, the pressure inside a star also goes to zero. How can a star be stabilized after the fusion processes and thus energy production stopped? The solution to this puzzle is that the main source of pressure in such compact stars has a different origin.

According to Pauli's exclusion principle no two fermions can occupy the same quantum state [45]. In statistical mechanics, Heisenberg's uncertainty principle $\Delta x \Delta p \geq \hbar$ [46] together with Pauli's principle imply that each phase-space volume, $\hbar^{-1} dx dp$, can only be occupied by one fermionic state.

A (relativistic or non-relativistic) particle in a box of volume L^3 collides per time interval $\Delta t = L/v_x$ once with the yz -side of the box, if the x component of its velocity is v_x . Thereby it exerts the force $F_x = \Delta p_x / \Delta t = p_x v_x / L$. The pressure produced by N particles is then $P = F/A = N p_x v_x / (LA) = n p_x v_x$. For an isotropic distribution, with $\langle v^2 \rangle = \langle v_x^2 \rangle + \langle v_y^2 \rangle + \langle v_z^2 \rangle = 3 \langle v_x^2 \rangle$, we have

$$P = \frac{1}{3} n v p. \quad (50)$$

Now, if we take $\Delta x = n^{-1/3}$ and $\Delta p \approx \hbar / \Delta x \approx \hbar n^{1/3}$, combined with the non-relativistic expression $v = p/m$, the pressure of a degenerate fermion gas is found to be

$$P \approx n v p \approx \frac{\hbar^2 n^{5/3}}{m}. \quad (51)$$

(51) implies $P \propto \rho^{5/3}$, where ρ is the density. For relativistic particles, we can obtain an estimate for the pressure inserting $v = c$,

$$P \approx n c p \approx c \hbar n^{4/3}, \quad (52)$$

which implies $P \propto \rho^{4/3}$. It may be worth noting at this juncture that (i) both the non-relativistic and the relativistic pressure laws are polytropic equations of state, $P = K \rho^\gamma$; (ii) a non-relativistic degenerate Fermi gas has the same adiabatic index ($\gamma = 5/3$) as an ideal gas, whereas a relativistic degenerate Fermi gas has the same adiabatic index ($\gamma = 4/3$) as radiation; (iii) in the non-relativistic limit the pressure is inversely proportional to the fermion mass, $P \propto 1/m$, and so for non-relativistic systems the degeneracy will first become important to electrons.

EXERCISE 4.2 Estimate the average energy of electrons in Sirius B from the equation of state for non-relativistic degenerate fermion gas,

$$P = \frac{(3\pi^2)^{2/3}}{5} \frac{\hbar^2}{m} n^{5/3}, \quad (53)$$

and calculate the Lorentz factor of the electrons. Give a short qualitative statement about the validity of the non-relativistic equation of state for white dwarfs with a density of Sirius B and beyond.

Next, we compute the pressure of a degenerate non-relativistic electron gas inside Sirius B and check if it is consistent with the lower limit for the central pressure derived in (48). The only bit of information needed is the value of n_e , which can be written in terms of the density of the star, the atomic mass of the ions making up the star, and the number of protons in the ions (assuming the star is neutral):

$$n_e = \frac{\rho}{\mu_e m_p} \quad (54)$$

where $\mu_e \equiv A/Z$ is the average number of nucleon per free electron. For metal-poor stars $\mu_e = 2$, and so from (51) we obtain

$$\begin{aligned} P &\approx \frac{\hbar^2 n_e^{5/3}}{m_e} \\ &\approx \frac{(1.05 \times 10^{27} \text{ erg s})^2}{9.11 \times 10^{-28} \text{ g}} \left(\frac{10^6 \text{ g/cm}^3}{2 \times 1.67 \times 10^{-24} \text{ g}} \right)^{5/3} \\ &\approx 10^{23} \text{ dyn/cm}^2. \end{aligned} \quad (55)$$

Since $10^6 \text{ dyn/cm}^2 = 1 \text{ bar}$, we have $P = 10^{17} \text{ bar}$, which is consistent with the lower limit derived in (48).

We can now relate the mass of the star to its radius by combining the lower limit on the central pressure $P_c \sim GM^2/R^4$ and the polytropic equation of state

$P = K\rho^{5/3} \sim K(M/R^3)^{5/3} = KM^{5/3}/R^5$. It follows that

$$\frac{GM^2}{R^4} = \frac{KM^{5/3}}{R^5}, \quad (56)$$

or equivalently

$$R = \frac{M^{(10-12)/6}}{K} = \frac{1}{KM^{1/3}}. \quad (57)$$

If the small differences in chemical composition can be neglected, then there is unique relation between the mass and the radius of white dwarfs. Since the star's radius decreases with increasing mass, there must be a maximal mass allowed.

To derive this maximal mass we first assume the pressure can be described by a non-relativistic degenerate Fermi gas. The total kinetic energy of the star is $U_{\text{kin}} = Np^2/(2m_e)$, where $n \sim N/R^3$ and $p \sim \hbar n^{1/3}$. Thus

$$U_{\text{kin}} \sim N \frac{\hbar^2 n^{2/3}}{2m_e} \sim \frac{\hbar^2 N^{(3+2)/3}}{2m_e R^2} = \frac{\hbar^2 N^{5/3}}{2m_e R^2}. \quad (58)$$

For the potential gravitational energy, we use the approximation $U_{\text{pot}} = \alpha GM^2/R$, with $\alpha = 1$. Hence

$$U(R) = U_{\text{kin}} + U_{\text{pot}} \sim \frac{\hbar^2 N^{5/3}}{2m_e R^2} - \frac{GM^2}{R}. \quad (59)$$

For small R , the positive term dominates and so there exists a stable minimum R_{\min} for each M .

However, if the Fermi gas inside the star becomes relativistic, then $U_{\text{kin}} = Ncp$, or

$$U_{\text{kin}} \sim Nc\hbar n^{1/3} \sim \frac{c\hbar N^{4/3}}{R} \quad (60)$$

and

$$U(R) = U_{\text{kin}} + U_{\text{pot}} \sim \frac{c\hbar N^{4/3}}{R} - \frac{GM^2}{R}. \quad (61)$$

Now both terms scale like $1/R$. For a fixed chemical composition, the ratio N/M remains constant. Therefore, if M is increased the negative term increases faster than the first one. This implies there exists a critical M so that U becomes negative, and can be made arbitrary small by decreasing the radius of the star: *the star collapses*. This critical mass is called Chandrasekhar mass M_{Ch} . It can be obtained by solving (61) for $U = 0$. Using $M = Nm_N$ we have $c\hbar N_{\max}^{4/3} = GN_{\max}^2 m_N^2$, or, with $m_N \simeq m_p$,

$$N_{\max} \sim \left(\frac{c\hbar}{Gm_p^2} \right)^{3/2} \sim \left(\frac{M_{\text{Pl}}}{m_p} \right)^3 \sim 2 \times 10^{57}. \quad (62)$$

This leads to

$$M_{\text{Ch}} = N_{\max} m_p \sim 1.5 M_{\odot}. \quad (63)$$

The Chandrasekhar mass derived "professionally" is found to be $M_{\text{Ch}} \simeq 1.46 M_{\odot}$ [24].

EXERCISE 4.3 Derive approximate Chandrasekhar mass limits in units of solar mass by setting the central pressures of exercise 4.1 equal to the relativistic degenerate electron pressure,

$$P = \frac{(3\pi^2)^{1/3} \hbar c}{4} n^{4/3}. \quad (64)$$

Compare the estimates with the exact limit.

The critical size can be determined by imposing two conditions: that the gas becomes relativistic, $U_{\text{kin}} \lesssim Nm_e c^2$, and $N = N_{\max}$,

$$N_{\max} m_e c^2 \gtrsim \frac{c\hbar N_{\max}^{4/3}}{R}. \quad (65)$$

This leads to

$$m_e c^2 \gtrsim \frac{c\hbar}{R} \left(\frac{c\hbar}{Gm_N^2} \right)^{1/2}, \quad (66)$$

or equivalently

$$R \gtrsim \frac{\hbar}{m_e c} \left(\frac{c\hbar}{Gm_N^2} \right)^{1/2} \sim 5 \times 10^8 \text{ cm}. \quad (67)$$

which is in agreement with the radii found for white dwarf stars.

C. Supernovae

Supernovae are massive explosions that take place at the end of a star's life cycle. They can be triggered by one of two basic mechanisms: (I) the sudden re-ignition of nuclear fusion in a degenerate star, or (II) the sudden gravitational collapse of the massive star's core.

In a type I supernova, a degenerate white dwarf accumulates sufficient material from a binary companion, either through accretion or via a merger. This material raise its core temperature to then trigger runaway nuclear fusion, completely disrupting the star. Since the white dwarf stars explode crossing the Chandrasekhar limit, $M > M_{\odot}$, the release total energy should not vary so much. Thus one may wonder if they are possible standard candles.

EXERCISE 4.4 Type Ia supernovae have been observed in some distant galaxies. They have well-known luminosities and at their peak $L_{\text{Ia}} \approx 10^{10} L_{\odot}$. Hence, we can use them as standard candles to measure the distances to very remote galaxies. How far away could a type Ia supernova be, and still be detected with HST?

In type II supernovae the core of a $M \gtrsim 8 M_{\odot}$ star undergoes sudden gravitational collapse. These stars

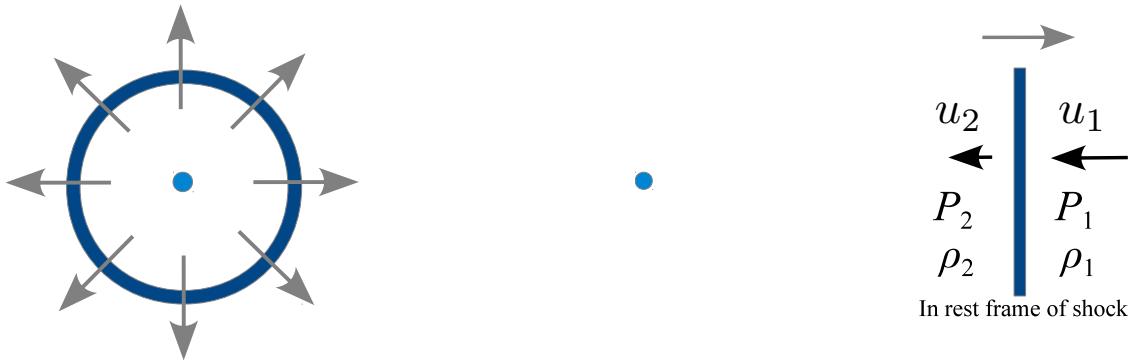
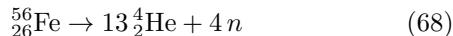


FIG. 7: **Left.** The sudden release of a large amount of energy into a background fluid of density ρ_1 creates a strong spherical shock wave, emanating from the point where the energy was released. **Right.** Jump conditions across normal shock waves. If the shock moves to the right with velocity u_{sh} , then in the rest-frame of the shock the background gas streams with velocity $u_1 = -u_{\text{sh}}$ to the left, and comes out of the shock with a higher density ρ_2 , higher pressure P_2 , and with a lower velocity u_2 . Conservation of momentum requires $P_1 + \rho_1 u_1^2 = P_2 + \rho_2 u_2^2$, see Appendix B. For the case at hand, $P_1 \ll P_2$ and so $P_2 \sim \rho_1 u_1^2$.

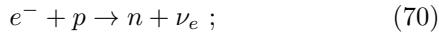
have an onion-like structure with a degenerate iron core. When the core is completely fused to iron, no further processes releasing energy are possible. Instead, high energy collisions break apart iron into helium and eventually into protons and neutrons,



and



This removes the thermal energy necessary to provide pressure support and the star collapses. When the star begins to contract the density increases and the free electrons are forced together with protons to form neutrons via inverse beta decay,



even though neutrinos do not interact easily with matter, at these extremely high densities, they exert a tremendous outward pressure. The outer layers fall inward when the iron core collapses, forming an enormously dense neutron star [25]. If $M \lesssim M_{\text{Ch}}$, then the core stops collapsing because the neutrons start getting packed too tightly. Note that M_{Ch} as derived in (63) is valid for both neutrons and electrons, since the stellar mass is in both cases given by the sum of the nucleon masses, only the main source of pressure (electrons or neutrons) differs. The critical size follows from (67) by substituting m_e with m_N ,

$$R \gtrsim \frac{\hbar}{m_N c} \left(\frac{c \hbar}{G m_N^2} \right)^{1/2} \sim 3 \times 10^5 \text{ cm}. \quad (71)$$

Since already Sirius B was difficult to detect, the question arises if and how these extremely small stars can be observed. When core density reaches nuclear density, the equation of state stiffens suddenly and the infalling material is “reflected.” Both the neutrino outburst and the

outer layers that crash into the core and rebound cause the entire star outside the core to be blown apart. The released energy goes mainly into neutrinos (99%), kinetic energy (1%); only 0.01% into photons.

Much of the modeling of supernova explosions and their remnants derives from the nuclear bomb research program. Whenever a supernova goes off a large amount of energy E is injected into the “ambient medium” of uniform density ρ_1 . In the initial phase of the expansion the impact of the external medium will be small, because the mass of the ambient medium that is overrun and taken along is still small compared with the ejecta mass. The supernova remnant is said to expand adiabatically. After some time a strong spherical shock front (a “blast wave”) expands into the ambient medium, and the mass swept up by the outwardly moving shock significantly exceeds the mass of the initial ejecta, see Fig. 7. The ram pressure, $P_2 \sim \rho_1 u_{\text{sh}}^2$, of the matter that enters the shock wave is much larger than the ambient pressure P_1 of the upstream medium, and any radiated energy is much smaller than the explosion energy E . This regime, during which the energy remains constant is known as the Sedov–Taylor phase [26–28]. The mass of the swept up material is of order $M(t) \sim \rho_1 r^3(t)$, where r is the radius of the shock. The fluid velocity behind the shock will be of order the mean radial velocity of the shock, $u_{\text{sh}}(t) \sim r(t)/t$ and so the kinetic energy is

$$E_{\text{kin}} = \frac{1}{2} M u_{\text{sh}}^2 \sim \rho_1 r^3 \frac{r^2}{t^2} = \rho_1 \frac{r^5}{t^2}. \quad (72)$$

What about the thermal energy in the bubble created by the explosion? This should be of order

$$E_{\text{therm}} = \frac{3}{2} P_2 V \sim P_2 r^3 \sim \rho_1 u_{\text{sh}}^2 r^3 \sim \rho_1 \frac{r^5}{t^2}. \quad (73)$$

This suggests that the thermal energy is of the same order as the kinetic energy, and scales in the same fashion with

time. Therefore

$$E = E_{\text{kin}} + E_{\text{therm}} \sim \rho_1 \frac{r^5}{t^2}, \quad (74)$$

yielding

$$r(t) \sim \left(\frac{Et^2}{\rho_1} \right)^{1/5}. \quad (75)$$

The expanding shock wave slows as it expands

$$u_{\text{sh}} = \frac{2}{5} \left(\frac{E}{\rho_1 t^3} \right)^{1/5} = \frac{2}{5} \left(\frac{E}{\rho_1} \right)^{1/2} r^{-3/2}. \quad (76)$$

This means that the blast wave decelerates and disappears after some time. The expanding supernova remnant then passes from its Taylor-Sedov phase to its “snowplow” phase. During the snowplow phase, the matter of the ambient interstellar medium is swept up by the expanding dense shell, just as snow is swept up by a coasting snowplow.

EXERCISE 4.5 Estimate the energy of the first detonation of a nuclear weapon (code name Trinity) from the time dependence of the radius of its shock wave. Photographs of the early stage of the explosion are shown in Fig. 8. The device was placed on the top of a tower, $h = 30$ m and the explosion took place at about 1100 m above sea level. (i) Explain the origin of the thin layer above the bright “fireball” that can be seen in the last three pictures ($t \geq 0.053$ s). Is the shock front behind or ahead of this layer? Read the radius of the shock front from the figures and plot it as a function of time after the explosion. The time and length scale are indicated in the labels of the figures. (ii) Fit (by eye or numerical regression) a line to the radius vs. time dependence of the shock front in a log-log representation, $\ln(r) = a + b \ln(t)$. Verify that b is compatible with a Sedov-Taylor expansion. Then fix b to the theoretical expectation, re-evaluate a and estimate the energy of the bomb in tons of TNT equivalent. [Hint: ignore the initial (short) phase of free expansion.]

If the final mass of a neutron star is less than M_{Ch} its subsequent evolution is thought to be similar to that of a white dwarf. In 1967, an unusual object emitting a radio signal with period $T = 1.377$ s was detected at the Mullard Radio Astronomy Observatory. By its very nature the object was called “pulsar.” Only one year later, Gold argued that pulsars are rotating neutron stars [30]. He predicted an increase on the pulsar period because of electromagnetic energy losses. The slow-down of the Crab pulsar was indeed discovered in 1969 [31].

If the mass of the neutron star is greater than M_{Ch} , then the star collapses under gravity, overcoming even the neutron exclusion principle [32]. The star eventually collapses to the point of zero volume and infinite density, creating what is known as a “singularity” [33–38]. As the

density increases, the paths of light rays emitted from the star are bent and eventually wrapped irrevocably around the star. Any emitted photon is trapped into an orbit by the intense gravitational field; it will never leave it. Because no light escapes after the star reaches this infinite density, it is called a black hole.

V. WARPING SPACETIME

A hunter is tracking a bear. Starting at his camp, he walks one mile due south. Then the bear changes direction and the hunter follows it due east. After one mile, the hunter loses the bear’s track. He turns north and walks for another mile, at which point he arrives back at his camp. What was the color of the bear?

An odd question. Not only is the color of the bear unrelated to the rest of the question, but how can the hunter walk south, east and north, and then arrive back at his camp? This certainly does not work everywhere on Earth, but it does if you start at the North pole. Therefore the color of the bear has to be white. A surprising observation is that the triangle described by the hunter’s path has two right angles in the two bottom corners, and so the sum of all three angles is greater than 180° . This implies the metric space is curved.

What is meant by a curved space? Before answering this question, we recall that our normal method of viewing the world is via Euclidean plane geometry, where the line element of the n -dimensional space is given by

$$ds^2 = \sum_{i=1}^n dx_i^2. \quad (77)$$

Non-Euclidean geometries which involve curved spaces have been independently imagined by Gauss [39], Bolyai [40], and Lobachevsky [41]. To understand the idea of a metric space herein we will greatly simplify the discussion by considering only 2-dimensional surfaces. For 2-dimensional metric spaces, the so-called first and second fundamental forms of differential geometry uniquely determine how to measure lengths, areas and angles on a surface, and how to describe the shape of a parameterized surface.

A. 2-dimensional metric spaces

The parameterization of a surface maps points (u, v) in the domain to points $\vec{\sigma}(u, v)$ in space:

$$\vec{\sigma}(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix}. \quad (78)$$

Differential geometry is the local analysis of how small changes in position (u, v) in the domain affect the position on the surface $\vec{\sigma}(u, v)$, the first derivatives $\vec{\sigma}_u(u, v)$ and $\vec{\sigma}_v(u, v)$, and the surface normal $\hat{n}(u, v)$.

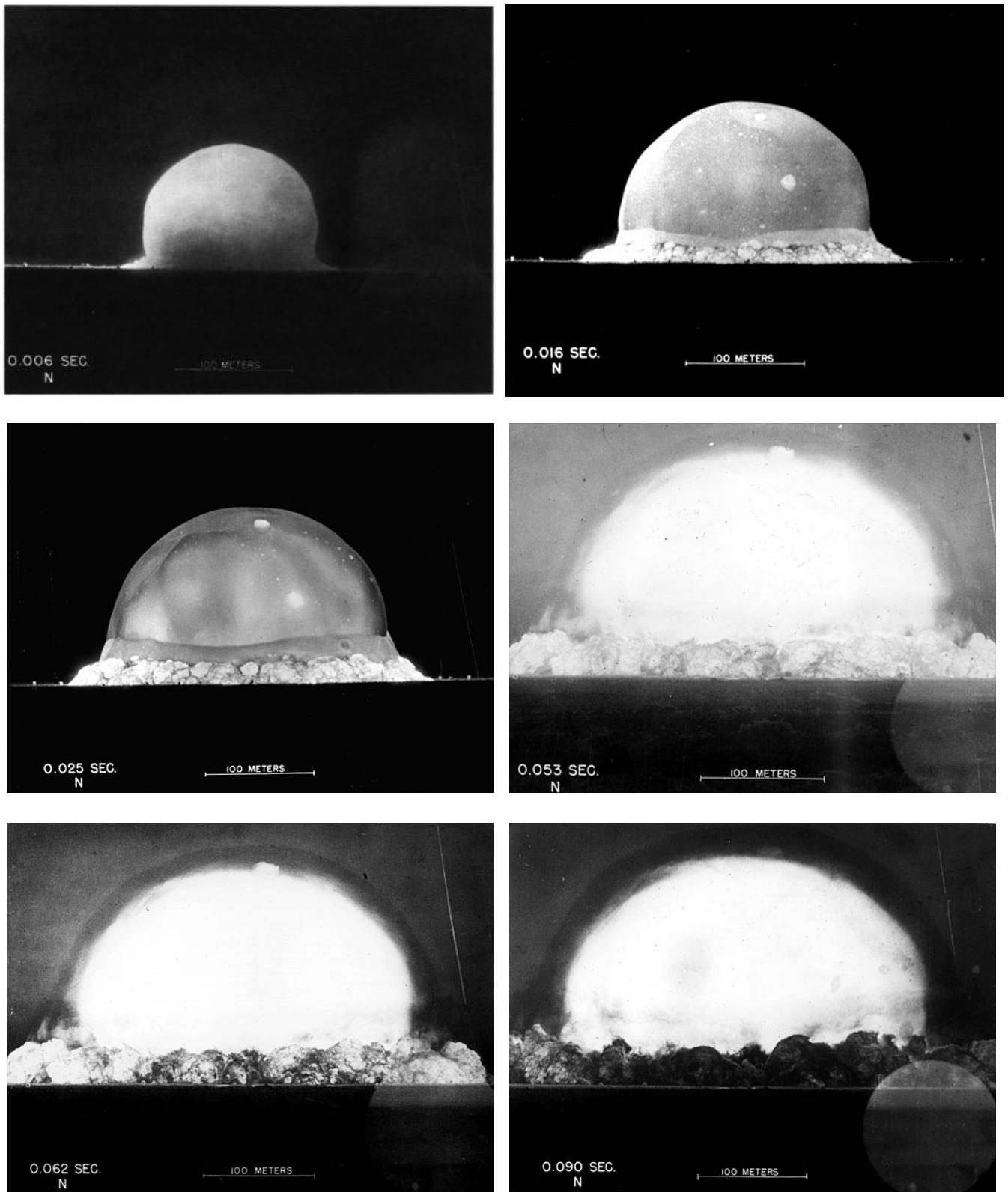


FIG. 8: Trinity test of July 16, 1945.

The first derivatives, $\vec{\sigma}_u(u, v)$ and $\vec{\sigma}_v(u, v)$, are vectors that span the tangent plane to the surface at point $\vec{\sigma}(u, v)$. The surface normal at point $\vec{\sigma}$ is defined as the unit vector normal to the tangent plane at point $\vec{\sigma}$ and is computed using the cross product of the partial derivatives of the surface parameterization,

$$\hat{n}(\vec{\sigma}) = \frac{\vec{\sigma}_u \times \vec{\sigma}_v}{\|\vec{\sigma}_u \times \vec{\sigma}_v\|}. \quad (79)$$

The tangent vectors and the surface normal define an orthogonal coordinate system at point $\vec{\sigma}(u, v)$ on the surface, which is the framework for describing the local shape of the surface.

Geometrically, $d\vec{\sigma}$ is a differential vector quantity that is tangent to the surface in the direction defined by du and dv . The first fundamental form, I , which measures the distance of neighboring points on the surface with parameters (u, v) and $(u + du, v + dv)$, is given by the inner product of $d\vec{\sigma}$ with itself

$$\begin{aligned} I &\equiv ds^2 = d\vec{\sigma} \cdot d\vec{\sigma} = (\vec{\sigma}_u du + \vec{\sigma}_v dv) \cdot (\vec{\sigma}_u du + \vec{\sigma}_v dv) \\ &= (\vec{\sigma}_u \cdot \vec{\sigma}_u) du^2 + 2(\vec{\sigma}_u \cdot \vec{\sigma}_v) du dv + (\vec{\sigma}_v \cdot \vec{\sigma}_v) dv^2 \\ &= Edu^2 + 2Fdu dv + Gdv^2, \end{aligned} \quad (80)$$

where E , F and G are the first fundamental coefficients. The coefficients have some remarkable properties. For example, they can be used to calculate the surface area. Namely, the area bounded by four vertices $\vec{\sigma}(u, v), \vec{\sigma}(u + \delta u, v), \vec{\sigma}(u, v + \delta v), \vec{\sigma}(u + \delta u, v + \delta v)$ can be expressed in terms of the first fundamental form with the assistance of Lagrange identity

$$\begin{aligned} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_i b_j - a_j b_i)^2 &= \left(\sum_{k=1}^n a_k^2 \right) \left(\sum_{k=1}^n b_k^2 \right) \\ &- \left(\sum_{k=1}^n a_k b_k \right)^2, \end{aligned} \quad (81)$$

which applies to any two sets $\{a_1, a_2, \dots, a_n\}$ and $\{b_1, b_2, \dots, b_n\}$ of real numbers. The classical area element is found to be

$$\delta A = |\vec{\sigma}_u \delta u \times \vec{\sigma}_v \delta v| = \sqrt{EG - F^2} \delta u \delta v, \quad (82)$$

or in differential form

$$dA = \sqrt{EG - F^2} du dv. \quad (83)$$

Note that the expression under the square root in (83) is precisely $|\vec{\sigma}_u \times \vec{\sigma}_v|$ and so it is strictly positive at the regular points.

The key to the second fundamental form, Π , is the unit normal vector. The second fundamental form coefficients at a given point in the parametric uv -plane are given by the projections of the second partial derivatives of $\vec{\sigma}$ at that point onto the normal vector and can be computed with the aid of the dot product as follows: $e = \vec{\sigma}_{uu} \cdot \hat{n}$,

$f = \vec{\sigma}_{uv} \cdot \hat{n}$, and $g = \vec{\sigma}_{vv} \cdot \hat{n}$. The second fundamental form,

$$\Pi = e du^2 + 2f du dv + g dv^2, \quad (84)$$

can be used to characterize the local shape of the folded surface.

The concept of curvature, while intuitive for a plane curve (the reciprocal of the radius of curvature), requires a more comprehensive definition for a surface. Through a point on a surface any number of curves may be drawn with each having a different curvature at the point. We have seen that at any point on a surface we can find \hat{n} which is at right angles to the surface; planes containing the normal vector are called normal planes. The intersection of a normal plane and the surface will form a curve called a normal section and the curvature of this curve is the normal curvature κ . For most points on most surfaces, different sections will have different curvatures; the minimum and maximum values of these are called the principal curvatures, denoted by κ_1 and κ_2 . The Gaussian curvature is defined by the product of the two principal curvatures $K = \kappa_1 \kappa_2$. It may be calculated using the first and second fundamental coefficients. At each grid point where these values are known two matrices are defined. The matrix of the first fundamental form,

$$I = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \quad (85)$$

and the matrix of the second fundamental form,

$$\Pi = \begin{pmatrix} e & f \\ f & g \end{pmatrix}. \quad (86)$$

The Gaussian curvature is given by

$$K = \frac{\det \Pi}{\det I}. \quad (87)$$

As an illustration, consider a half-cylinder of radius R oriented along the x axis. At a particular point on the surface, the scalar curvature can have different values depending on direction. In the direction of the half-cylinder's axis (parallel to the x axis), the surface has zero scalar curvature, $\kappa = 0$. This is the smallest curvature value at any point on the surface, and therefore κ_1 is in this direction. For a curve on the half-cylinder's surface parallel to the (y, z) plane, the cylinder has uniform scalar curvature. In fact this curvature is the greatest possible on the surface, so that $\kappa_2 = 1/R$ is in this direction. For a curve on the surface not in one of these directions, the scalar curvature is greater than κ_1 and less than κ_2 . The Gaussian curvature is $K = 0$.

2-dimensional metric spaces can be classified according to the Gaussian curvature into elliptic ($K > 0$), flat ($K = 0$), and hyperbolic ($K < 0$). Triangles which lie on the surface of an elliptic geometry will have a sum of angles which is greater than 180° . Triangles which lie on

the surface of an hyperbolic geometry will have a sum of angles which is less than 180° .

EXERCISE 5.1 The unit sphere can be parametrized as

$$\vec{\sigma}(u, v) = \begin{pmatrix} \cos u \sin v \\ \sin u \sin v \\ \cos v \end{pmatrix} \quad (88)$$

where $(u, v) \in [0, 2\pi] \times [0, \pi]$. (i) Find the distance of neighboring points on the surface with parameters (u, v) and $(u+du, v+dv)$, a.k.a. the line element ds^2 . (ii) Find the surface area. (iii) Find the Gaussian curvature.

EXERCISE 5.2 The tractrix is a curve with the following nice interpretation: Suppose a dog-owner takes his pet along as he goes for a walk “down” the y -axis. He starts from the origin, with his dog initially standing on the x -axis at a distance r away from the owner. Then the tractrix is the path followed by the dog if he “follows his owner unwillingly”, i.e., if he constantly pulls against the leash, keeping it tight. This means mathematically that the leash is always tangent to the path of the dog, so that the length of the tangent segment from the tractrix to the y -axis has constant length r . The tractrix has a well-known surface of revolution called the pseudosphere which, for $r = 1$, can be parametrized as

$$\vec{\sigma}(u, v) = \begin{pmatrix} \operatorname{sech} u \cos v \\ \operatorname{sech} u \sin v \\ u - \tanh u \end{pmatrix}, \quad (89)$$

with $u \in (-\infty, \infty)$ and $v \in [0, 2\pi]$. (i) Find the line element. (ii) Find the surface area. (iii) Find the Gaussian curvature.

A curve γ with parametr t on a surface $\vec{\sigma}(u, v)$ is called a *geodesic* if at every point $\gamma(t)$ the acceleration vector $\ddot{\gamma}(t)$ is either zero or parallel to its unit normal \hat{n} .

EXERCISE 5.3 Show that a geodesic $\gamma(t)$ on a surface $\vec{\sigma}$ has constant speed.

EXERCISE 5.4 A curve γ on a surface $\vec{\sigma}$ is a geodesic if and only if for any part $\gamma(t) = \vec{\sigma}(u(t), v(t))$ contained in a surface patch $\vec{\sigma}$, the following two equations are satisfied:

$$\frac{d}{dt}(E\dot{u} + F\dot{v}) = \frac{1}{2}(E_u\dot{u}^2 + 2F_u\dot{u}\dot{v} + G_u\dot{v}^2), \quad (90)$$

$$\frac{d}{dt}(F\dot{u} + G\dot{v}) = \frac{1}{2}(E_v\dot{u}^2 + 2F_v\dot{u}\dot{v} + G_v\dot{v}^2), \quad (91)$$

where $Edu^2 + 2Fdudv + Gdv^2$ is the first fundamental form of $\vec{\sigma}$. (90) and (91) are called the geodesic equations. They are nonlinear and solvable analytically on rare occasions only.

EXERCISE 5.5 Show that if γ is a geodesic on the unit sphere S^2 , then γ is part of a great circle. Consider the patch under the parametrization

$$\vec{\sigma}(\theta, \phi) = \begin{pmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ \sin \theta \end{pmatrix}. \quad (92)$$

[Hint: A great circle (a.k.a. orthodrome) of a sphere is the intersection of the sphere and a plane which passes through the center point of the sphere.]

The scalar curvature (or Ricci scalar) is the simplest curvature invariant of an n -dimensional hypersurface. To each point on the hypersurface, it assigns a single real number determined by the intrinsic geometry of the hypersurface near that point. It provides one way of measuring the degree to which the geometry determined by a given metric might differ from that of ordinary Euclidean n -space. In two dimensions, the scalar curvature is twice the Gaussian curvature, $R = 2K$, and completely characterizes the curvature of a surface. In more than two dimensions, however, the curvature of hypersurfaces involves more than one functionally independent quantity.

B. Schwarzschild metric

Consider a freely falling spacecraft in the gravitational field of a radially symmetric mass distribution with total mass M . Because the spacecraft is freely falling, no effects of gravity are felt inside. Then, the spacetime coordinates from $r \rightarrow \infty$ should be valid inside the spacecraft. Let us call these coordinates $\vec{\Sigma}_\infty(t_\infty, x_\infty, y_\infty, z_\infty)$, with x_∞ parallel and y_∞, z_∞ transversal to movement. The spacecraft has velocity v at the distance r from the mass M , measured in the coordinate system $\vec{\Sigma} = (t, r, \theta, \phi)$ in which the mass M is at rest at $r = 0$. As long as the *gravitational field is weak*, to first order approximation that the laws of special relativity hold [47], and we can use a Lorentz transformation [20] to relate $\vec{\Sigma}$ at rest and $\vec{\Sigma}_\infty$ moving with $v = \beta c$. We will define shortly what “weak” means in this context. For the moment, we presume that effects of gravity are small if the velocity of the spacecraft, which was at rest a $r \rightarrow \infty$, is still small $v \ll c$. Should this be the case, we have

$$\begin{aligned} dt_\infty &= dt \sqrt{1 - \beta^2} \\ dx_\infty &= \frac{dr}{\sqrt{1 - \beta^2}} \\ dy_\infty &= r d\theta \\ dz_\infty &= r \sin \theta d\phi. \end{aligned} \quad (93)$$

The infinitesimal distance between two spacetime events is given by the Minkowskian line element [48]

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = c^2 dt_\infty^2 - dx_\infty^2 - dy_\infty^2 - dz_\infty^2, \quad (94)$$

which, for the case at hand, becomes

$$ds^2 = (1 - \beta^2)c^2 dt^2 - \frac{dr^2}{1 - \beta^2} + r^2(d\theta^2 + \sin^2 d\phi^2). \quad (95)$$

Herein we follow the notation of [21]: Greek indices (μ, ν, \dots) run from 0 to 3 and Latin indices (i, j, \dots) from 1 to 3.

We now turn to determine β from measurable quantities of the system: M and r . Consider the energy of the spacecraft with rest mass m ,

$$(\gamma - 1)mc^2 - \frac{G\gamma mM}{r} = 0, \quad (96)$$

where the first term is the kinetic energy and the second the Newtonian expression for the potential energy. Note that here we have made the crucial assumption that gravity couples not only to the mass of the spacecraft but also to its total energy. Dividing by γmc^2 gives

$$\left(1 - \frac{1}{\gamma}\right) - \frac{GM}{rc^2} = 0. \quad (97)$$

Introducing $\alpha = GM/c^2$ we can re-write (97) as

$$\sqrt{1 - \beta^2} = 1 - \frac{\alpha}{r}, \quad (98)$$

where $\gamma = (1 - \beta^2)^{-1/2}$. (98) leads to

$$1 - \beta^2 = 1 - \frac{2\alpha}{r} + \frac{\alpha^2}{r^2} \approx 1 - \frac{2\alpha}{r}; \quad (99)$$

in the last step, we neglected the term $(\alpha/r)^2$, since we attempt only at an approximation for large distances, where gravity is still weak. Inserting this expression into (95), we obtain the metric describing the gravitational field produced by a radially symmetric mass distribution,

$$ds^2 = \left(1 - \frac{2\alpha}{r}\right) c^2 dt^2 - \left(1 - \frac{2\alpha}{r}\right)^{-1} dr^2 - r^2 d\Omega^2, \quad (100)$$

where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$. Wickedly, this agrees with the exact result found by Schwarzschild [49] by solving Einstein's vacuum field equations of general relativity [50].

As in special relativity, the line element ds^2 determines the time and spatial distance between two spacetime events. The time measured by an observer in the instantaneous rest frame, known as the proper time $d\tau$, is given by $d\tau = ds/c$ [21]. In particular, the time difference between two events at the same point is obtained by setting $dx^i = 0$. If we choose two static observers at the position r and r' , then we find with $dr = d\phi = d\theta = 0$,

$$\frac{d\tau(r)}{d\tau(r')} = \frac{\sqrt{g_{00}(r)} dt}{\sqrt{g_{00}(r')} dt} = \sqrt{\frac{g_{00}(r)}{g_{00}(r')}}. \quad (101)$$

The time intervals $d\tau(r')$ and $d\tau(r)$ are different and thus the time measured by clocks at different distances r from

the mass M will differ too. In particular, the time τ_∞ measured by an observer at infinity will pass faster than the time experienced in a gravitational field,

$$\tau_\infty = \frac{\tau(r)}{\sqrt{1 - 2\alpha/r}} < \tau(r). \quad (102)$$

Since frequencies are inversely proportional to time, the frequency or energy of a photon traveling from r to r' will be affected by the gravitational field as

$$\frac{\nu(r')}{\nu(r)} = \sqrt{\frac{1 - 2\alpha/r}{1 - 2\alpha r'}}. \quad (103)$$

Therefore, an observer at $r' \rightarrow \infty$ will receive photons, which were emitted with frequency ν by a source at position r , redshifted to frequency ν_∞ ,

$$\nu_\infty = \sqrt{1 - \frac{2GM}{rc^2}} \nu(r). \quad (104)$$

Note that the photon frequency is redshifted by the gravitational field. The size of this effect is of order Φ/c^2 , where $\Phi = -GM/r$ is the Newtonian gravitational potential. We are now in position to specify more precisely what weak gravitational fields means. As long as $|\Phi|/c^2 \ll 1$, the deviation of

$$g_{00} = 1 - \frac{2GM}{rc^2} \approx 1 - 2\frac{\Phi(r)}{c^2} \quad (105)$$

from the Minkowski value $g_{00} = 1$ is small, and Newtonian gravity is a sufficient approximation.

What is the meaning of $r = 2\alpha$? At

$$R_{\text{Sch}} = \frac{2GM}{c^2} = 3 \text{ km} \frac{M}{M_\odot}, \quad (106)$$

the Schwarzschild coordinate system (100) becomes ill-defined. However, this does not mean necessarily that at $r = R_{\text{Sch}}$ physical quantities like tidal forces become infinite. As a matter of fact, all scalar invariants are finite, e.g. $R = 0$ and $\mathcal{R} = 12R_{\text{Sch}}/r^6$. Here R is the Ricci scalar and \mathcal{R} the Kretschmann scalar [51], a quadratic scalar invariant used to find the true singularities of a spacetime. The Schwarzschild's scalar invariants can only be found by long and troublesome calculation that is beyond the scope of this course; for a comprehensive discussion see e.g. [52, 53]. Before proceeding we emphasize again that, whether or not the singularity is moved to the origin, only depends on the coordinate frame used, and has no physical significance whatsoever; see Appendix C for an example.

If the gravitating mass is concentrated inside a radius smaller than R_{Sch} then we cannot obtain any information about what is going on inside R_{Sch} , and we say $r = R_{\text{Sch}}$ defines an event horizon. An object smaller than its Schwarzschild radius, is called a black hole. In Newtonian gravity, only the enclosed mass $M(r)$ of a spherically symmetric system contributes to the gravitational potential outside r . Therefore, we conclude the

Sun is not a black hole, because for all values of r the enclosed mass is $M(r) < rc^2/(2G)$. The Schwarzschild black hole is fully characterized by its mass M . To understand this better, we consider next what happens to a photon crossing the event horizon as seen from an observer at $r \rightarrow \infty$.

Light rays are characterized by $ds^2 = 0$. Consider a light ray traveling in the radial direction, that is to say $d\phi = d\theta = 0$. The Schwarzschild metric (100) becomes

$$\frac{dr}{dt} = \left(1 - \frac{2\alpha}{r}\right)c. \quad (107)$$

As seen from far away a light ray approaching a massive star will travel slower and slower as it comes closer to the Schwarzschild radius. In fact, for an observer at infinity the signal will reach $r = R_{\text{Sch}}$ only asymptotically, for $t \rightarrow \infty$. Similarly, the communication with a freely falling spacecraft becomes impossible as it reaches $r = R_{\text{Sch}}$. A more detailed analysis shows that indeed, as seen from infinity, no signal can cross the surface at $r = R_{\text{Sch}}$. The factors $(1 - 2\alpha/r)$ in (100) control the bending of light, a phenomenon known as gravitational lensing. The first observation of light deflection was performed by noting the change in position of stars as they passed near the Sun on the celestial sphere. The observations were performed in May 1919 during a total solar eclipse, so that the stars near the Sun (at that time in the constellation Taurus) could be observed [54].

EXERCISE 5.6 In addition to the time dilation due to an object moving at a finite speed that we have learned about in special relativity, we have seen that there is an effect in general relativity, termed “gravitational redshift,” caused by gravity itself. To understand this latter effect, consider a photon escaping from the Earth’s surface to infinity. It loses energy as it climbs out of the Earth’s gravitational well. As its energy E is related to its frequency ν by Planck’s formula $E = h\nu$, its frequency must therefore also be reduced, so observers at a great distance $r \rightarrow \infty$ must see clocks on the surface ticking at a lower frequency as well. Therefore an astronaut orbiting the Earth ages differently from an astronomer sitting still far from the Earth for two reasons; the effect of gravity, and the time dilation due to motion. In this problem, you will calculate both these effects, and determine their relative importance. (i) The escape speed from an object of mass M if you are a distance r from it is given by

$$v_{\text{escape}} = \sqrt{\frac{2GM}{r}}. \quad (108)$$

That is, if you are moving this fast, you will not fall back to the object, but will escape its gravitational field entirely. Schwarzschild’s solution to Einstein’s field equations of general relativity shows that a stationary, non-moving clock at a radius $r \geq R_{\oplus}$ from the Earth will tick at a rate that is

$$\sqrt{1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{r}} = \sqrt{1 - \frac{v_{\text{escape}}^2}{c^2}} \quad (109)$$

times as fast as one located far away from the Earth (i.e. at $r \rightarrow \infty$). Note how much this expression looks like the equivalent expression from special relativity for time dilation. Here R_{\oplus} is the radius of the Earth, and M_{\oplus} is its mass. Using (109) calculate the rate at which a stationary clock at a radius r (for $r > R_{\oplus}$) will tick relative to one at the surface of the Earth. Is your rate greater or less than 1? If greater than 1, this means the high altitude clock at $r > R_{\oplus}$ ticks faster than one on the surface; if less than one, this means the high altitude clock ticks slower than one on the surface. (ii) Now consider an astronaut orbiting at $r > R_{\oplus}$. What is her orbital velocity as a function of r ? Because she is moving with respect to a stationary observer at radius r , special relativity says that her clock is ticking slower. Calculate the ratio of the rate her clock ticks to that of a stationary observer at radius r . (Note that for circular motion, the acceleration in the spaceship travelling in a circle is not zero, so the spaceship is not in a single frame of inertia.) (iii) Determine an expression for the ratio of the rate at which the orbiting astronaut’s clock ticks to a stationary clock on the surface of the Earth, as a function of the radius r at which she orbits. You may ignore the small velocity of the clock on the surface of the Earth due to the Earth’s rotation. (iv) Using $\sqrt{1-x} \approx 1-x/2 + \dots$, $(1-x)^{-1} \approx 1+x + \dots$, and $(1-x)(1-y) = 1-x-y+xy \approx 1-(x+y)$, all valid for $x \ll 1$ and $y \ll 1$, derive an expression of the form $1-\delta$ for the relative rate of a clicking clock on the surface of the Earth and the orbiting astronaut. Demonstrate that $\delta \ll 1$. (v) Calculate the radius r at which the clock of the orbiting astronaut ticks at the same rate as a stationary one on the surface of the Earth; express your result in Earth radii and kilometers. Will an astronaut orbiting at a smaller radius age more or less than one who stayed home? Thus, do astronauts on the Space Shuttle (orbiting 300 km above the Earth’s surface) age more or less than one staying home?

EXERCISE 5.7 “A full set of rules [of Brockian Ultra Cricket, as played in the higher dimensions] is so massively complicated that the only time they were all bound together in a single volume they underwent gravitational collapse and became a black hole” [55]. A quote like this is crying out for a calculation. In this problem, we will answer Adams challenge, and determine just how complicated these rules actually are. An object will collapse into a black hole when its radius is equal to the radius of a black hole of the same mass; under these conditions, the escape speed at its surface is the speed of light (which is in fact the defining characteristic of a black hole). We can rephrase the above to say that an object will collapse into a black hole when its density is equal to the density of a black hole of the same mass. (i) Derive an expression for the density of a black hole of mass M . Treat the volume of the black hole as the volume of a sphere of radius given by the Schwarzschild radius. As the mass of a black hole gets larger, does the

density grow or shrink? (ii) Determine the density of the paper making up the Cricket rule book, in units of kilograms per cubic meter. Standard paper has a surface density of 75 g per square meter, and a thickness of 0.1 mm. (iii) Calculate the mass (in solar masses), and radius (in AU) of the black hole with density equal to that of paper. (iv) How many pages long is the Brockian Ultra Cricket rule book? Assume the pages are standard size ($8.5'' \times 11''$). For calculational simplicity, treat the book as spherical (a common approximation in this kind of problem). What if the rule book were even longer than you have just calculated? Would it still collapse into a black hole?

EXERCISE 5.8 Black holes provide the ultimate laboratory for studying strong-field gravitational physics. The tides near black holes can be so extreme that a process informally called “spaghettification” occurs in which a body falling towards a black hole is strongly stretched due to the difference in gravitational force at different locations along the body (this is called a tidal effect). In the following, imagine that you are falling into a $3M_\odot$ black hole. (i) What is the Schwarzschild radius of this black hole (in km)? (ii) You are 1.5 m tall and 70 kg in mass and are falling feet first. At what distance from the black hole would the gravitational force on your feet exceed the gravitational force on your head by 10 kN? Express this distance in km and in Schwarzschild radii of the black hole. (iii) To appreciate if this amount of force is enough to “spaghettify” and kill you, imagine that you are suspended from a ceiling of your room (on Earth) with a steel plate tied to your feet. Calculate the mass of the plate (in kg) that will give you a nice tug of 10 kN (you can ignore the weight of your body here). Do you think this pull will kill you? (iv) Now consider a trip toward the supermassive black hole at the center of our Galaxy, which has an estimated mass of 4 million M_\odot . How does this change the distance at which you will be “spaghettified” by the differential gravity force of 10 kN? Express your answer in km and in Schwarzschild radii of the black hole. (v) Find the smallest mass of the black hole for which you would not die by “spaghettification” before falling within its event horizon.

C. Eddington luminosity and black hole growth

Binary X-ray sources are places to find strong black hole candidates [56, 57]. A companion star is a perfect source of infalling material for a black hole. As the matter falls or is pulled towards the black hole, it gains kinetic energy, heats up and is squeezed by tidal forces. The heating ionizes the atoms, and when the atoms reach a few million degrees Kelvin, they emit X-rays. The X-rays are sent off into space before the matter crosses the event horizon, and so we can detect this X-ray emission. Another sign of the presence of a black hole is random variation of emitted X-rays. The infalling matter that

emits X-rays does not fall into the black hole at a steady rate, but rather more sporadically, which causes an observable variation in X-ray intensity. Additionally, if the X-ray source is in a binary system, the X-rays will be periodically cut off as the source is eclipsed by the companion star.

Cygnus X-1 is one of the strongest X-ray sources we can detect from Earth [58] and the first widely thought to be a black hole, after the detection of its rapid X-ray variability [59] and the identification of its optical counterpart with the blue supergiant star HDE 226868 [60, 61]. The X-ray emission is powered mainly by accretion from the strong stellar wind from HDE 226868 [62]. While the disk of accreting matter is incredibly bright on its own, Cygnus X-1 has another source of light: a pair of jets perpendicular to the disk erupt from the black hole carrying part of the infalling material away into the interstellar space [63].

Consider a steady spherically symmetrical accretion. We assume the accreting material to be mainly hydrogen and to be fully ionized. Under these circumstances, the radiation exerts a force mainly on the free electrons through Thomson scattering, since the scattering cross section for protons is a factor $(m_e/m_p)^2$ smaller, where $m_e/m_p = 5 \times 10^{-4}$ is the ratio of the electron and proton masses [64]. If F is the radiant energy flux ($\text{erg s}^{-1}\text{cm}^{-2}$) and $\sigma_T = 6.7 \times 10^{-25} \text{ cm}^2$ is the Thomson cross section, then the outward radial force on each electron equals the rate at which it absorbs momentum,

$$F_{\text{out}} = \frac{\sigma_T F}{c}. \quad (110)$$

The attractive electrostatic Coulomb force between the electrons and protons means that as they move out the electrons drag the protons with them. In effect, the radiation pushes out electron-proton pairs against the total gravitational force

$$F_{\text{in}} = \frac{GM}{r^2}(m_p + m_e) \quad (111)$$

acting on each pair at a radial distance r from the center. If the luminosity of the accreting source is L (erg s^{-1}), we have

$$F = \frac{L}{4\pi r^2} \quad (112)$$

by spherical symmetry, so the net inward force on an electron-proton pair is

$$F_{\text{net}} = \left(GMm_p - \frac{L\sigma_T}{4\pi c} \right) \frac{1}{r^2}. \quad (113)$$

There is a limiting luminosity for which this expression vanishes, called the Eddington limit [65]

$$L_{\text{Edd}} = \frac{4\pi GMm_p}{\sigma_T} \simeq 1.3 \times 10^{38} \left(\frac{M}{M_\odot} \right) \text{ erg s}^{-1}. \quad (114)$$

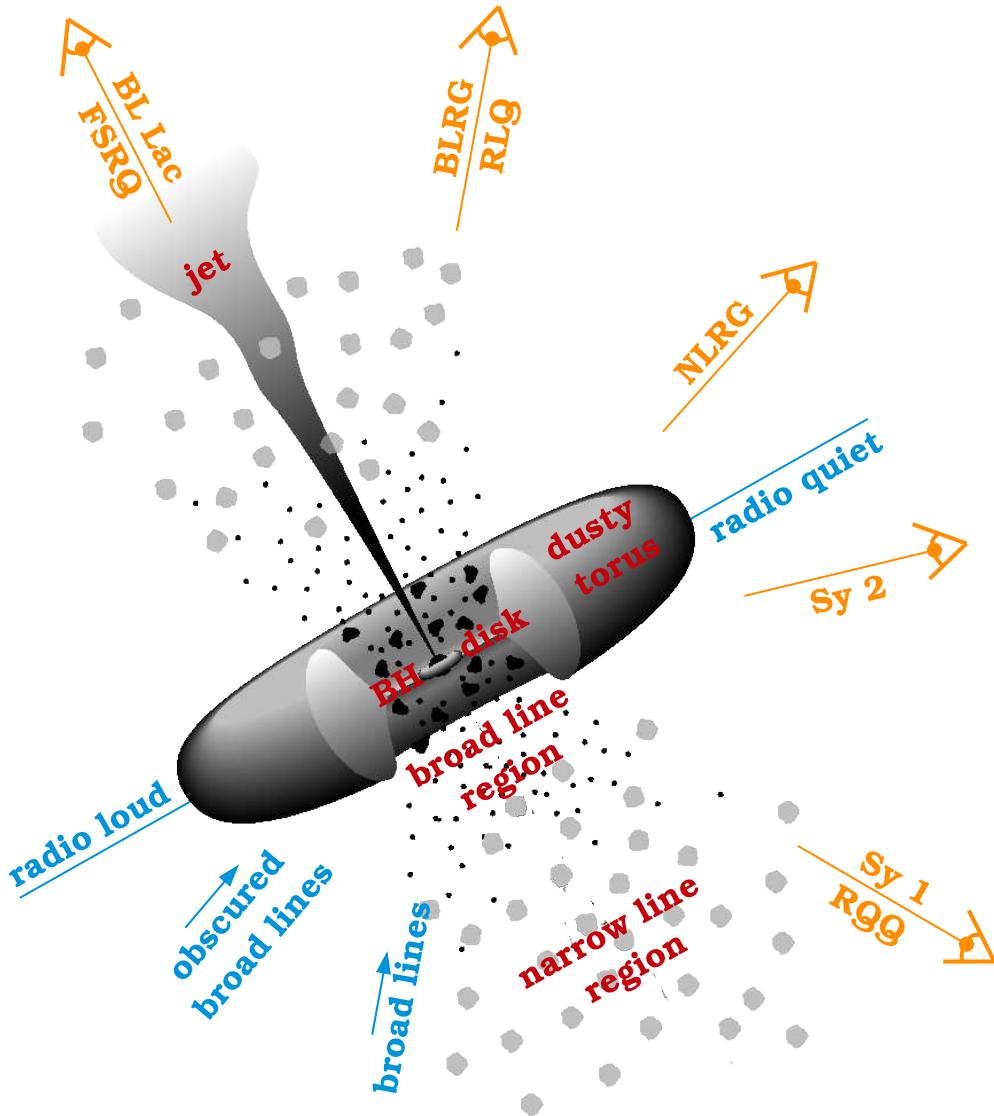


FIG. 9: Unification scheme of AGN. The acronyms for the different sub-classes of AGN are as follows: Fanaroff-Riley radio galaxies (FR I/II), narrow line radio galaxy (NLRG), broad line radio galaxy (BLRG), radio-loud quasar (RLQ), radio quiet quasar (RQQ), flat spectrum radio quasar (FSRQ), and Seyfert galaxies (Sy 1/2) [66].

At greater luminosities the outward pressure of radiation would exceed the inward gravitational attraction and accretion would be halted.

Active galactic nuclei (AGNs) are galaxies that harbor compact masses at the center exhibiting intense non-thermal emission that is often variable, which indicates small sizes (light months to light years). The luminosity of an accreting black hole is proportional to the rate at which it is gaining mass. Under favorable conditions, the accretion leads to the formation of a highly relativistic collimated jet. The formation of the jet is not well constrained, but it is thought to change from magnetic-field-dominated near the central engine to particle (electron and positron, or ions and electrons)

dominated beyond pc distances. The AGN taxonomy, controlled by the dichotomy between radio-quiet and radio-loud classes, is represented in Fig. 9. The appearance of an AGN depends crucially on the orientation of the observer with respect to the symmetry axis of the accretion disk [67]. In this scheme, the difference between radio-loud and radio-quiet AGN depends on the presence or absence of radio-emitting jets powered by the central nucleus, which in turn may be speculated to depend on: (i) black hole rotation; (ii) low power or high power, as determined by the mass-accretion rate $\dot{M}c^2/L_{\text{Edd}}$ [68].

EXERCISE 5.9 The pictures in Fig. 10 show a time sequence of radio observations of the quasar 0827+243.

The core of the quasar is the bright object at a distance of 0 ly and a fainter blob of plasma is moving away from it. (i) What is the apparent velocity of the motion of the plasma blob? (ii) Derive the apparent transverse velocity of an object ejected from a source at velocity v at an angle θ with respect to the line of sight between the source and the observer. (iii) Which angle maximizes the apparent transverse velocity? What is accordingly the minimal Lorentz-factor of the plasma blob observed in 0827+243?

VI. EXPANSION OF THE UNIVERSE

The observations that we will discuss in this section reveal that the universe is in a state of violent explosion, in which the galaxies are rushing apart at speeds approaching the speed of light. Moreover, we can extrapolate this explosion backwards in time and conclude that all the galaxies must have been much closer at the same time in the past – so close, in fact, that neither galaxies nor stars nor even atoms or atomic nuclei could have had a separate existence.

A. Olbers' paradox

The XVI century finally saw what came to be a watershed in the development of Cosmology. In 1543 Copernicus published his treatise “De Revolutionibus Orbium Celestium” (The Revolution of Celestial Spheres) where a new view of the world is presented: the heliocentric model [2].

It is hard to underestimate the importance of this work: it challenged the age long views of the way the universe worked and the preponderance of the Earth and, by extension, of human beings. The realization that we, our planet, and indeed our solar system (and even our Galaxy) are quite common in the heavens and reproduced by myriads of planetary systems, provided a sobering (though unsettling) view of the universe. All the reassurances of the cosmology of the Middle Ages were gone, and a new view of the world, less secure and comfortable, came into being. Despite these “problems” and the many critics the model attracted, the system was soon accepted by the best minds of the time such as Galileo.

The simplest and most ancient of all astronomical observations is that the sky grows dark when the Sun goes down. This fact was first noted by Kepler, who, in the XVII century, used it as evidence for a finite universe. In the XIX century, when the idea of an unending, unchanging space filled with stars like the Sun was widespread in consequence of the Copernican revolution, the question of the dark night sky became a problem. To clearly ascertain this problem, we recall that if absorption is neglected, the apparent luminosity of a star of absolute luminosity L at a distance r will be $b = L/4\pi r^2$. If the number density of such stars is a constant n , then the

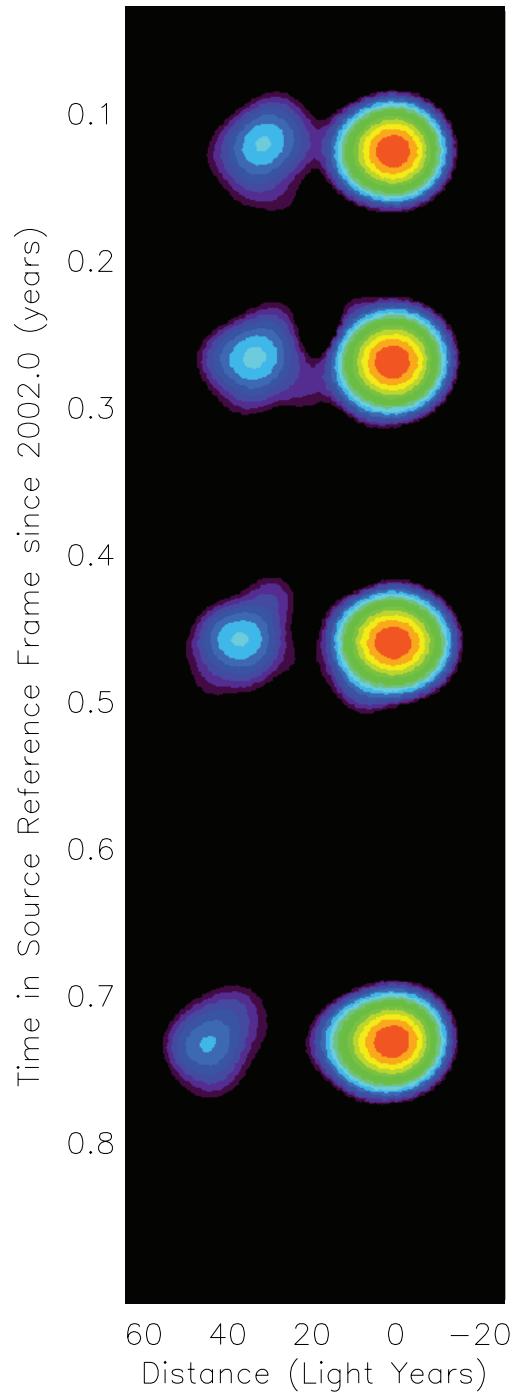


FIG. 10: Mosaic of images of 0827+243 at 22 GHz [69].

number of stars at distances r between r and $r + dr$ is $dN = 4\pi nr^2 dr$, so the total radiant energy density due to all stars is

$$\begin{aligned} \rho_s &= \int b \, dN = \int_0^\infty \left(\frac{L}{4\pi r^2} \right) 4\pi n r^2 dr \\ &= Ln \int_0^\infty dr. \end{aligned} \quad (115)$$

The integral diverges, leading to an infinite energy density of starlight!

In order to avoid this paradox, both de Chézeaux (1744) [70] and Olbers (1826) [71] postulated the existence of an interstellar medium that absorbs the light from very distant stars responsible for the divergence of the integral in (115). However, this resolution of the paradox is unsatisfactory, because in an eternal universe the temperature of the interstellar medium would have to rise until the medium was in thermal equilibrium with the starlight, in which case it would be emitting as much energy as it absorbs, and hence could not reduce the average radiant energy density. The stars themselves are of course opaque, and totally block out the light from sufficiently distant sources, but if this is the resolution of the so-called “Olbers paradox” then every line of segment must terminate at the surface of a star, so the whole sky should have a temperature equal to that at the surface of a typical star.

EXERCISE 6.1 (i) In a forest there are n trees per hectare, evenly spaced. The thickness of each trunk is D . What is the mean distance that you have an unobstructed view into the woods, i.e. the mean free path? (ii) How is this related to the Olbers paradox?

B. Hubble’s law

In 1929, Hubble discovered that the spectral lines of galaxies were shifted towards the red by an amount proportional to their distances [72]. If the redshift is due to the Doppler effect, this means that the galaxies move away from each other with velocities proportional to their separations. The importance of this observation is that it is just what we should predict according to the simplest possible picture of the flow of matter in an expanding universe.

The *redshift* parameter is defined as the traditional shift in wavelength of a photon emitted by a distant galaxy at time t_{em} and observed on Earth today

$$z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} - 1 = \frac{\nu_{\text{em}}}{\nu_{\text{obs}}} - 1, \quad (116)$$

Although measuring a galaxy’s redshift is relatively easy, and can be done with high precision, measuring its distance is difficult. Hubble knew z for nearly 50 galaxies, but had estimated distances for only 20 of them. Nevertheless, from a plot of redshift versus distance (reproduced in Fig. 11) he found the famous linear relation now known as the Hubble’s law:

$$z = \frac{H_0}{c} r, \quad (117)$$

where H_0 is a constant (now called the Hubble constant). Since in the study of Hubble all the redshift were small, $z < 0.04$, he was able to use the classical non-relativistic realtion for small velocities ($v \ll c$). From (25) the

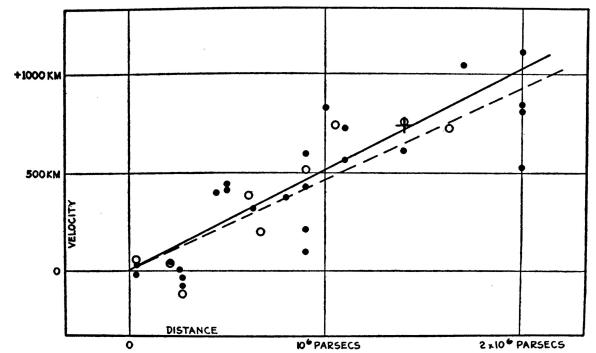


FIG. 11: Hubble’s original plot of the relation between redshift (vertical axis) and distance (horizontal axis). Note that in the vertical axis he actually plots cz rather than z , and that the units are accidentally written as km rather than km/s [72].

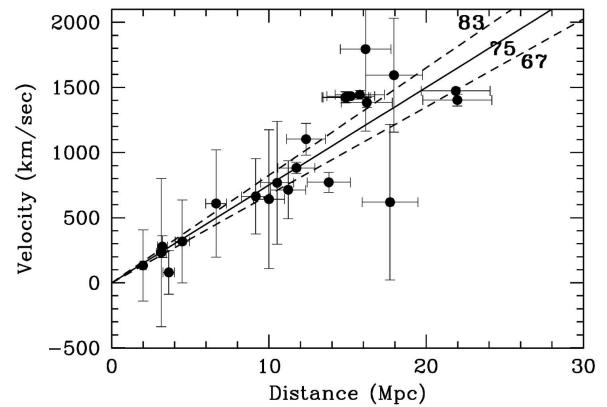


FIG. 12: A more modern version of Hubble’s plot, showing cz versus distance. In this case, the galaxy distances have been determined using Cepheid variable stars as standard candles [73].

Doppler redshift is $z \approx v/c$ and Hubble’s law takes the form

$$v = H_0 r. \quad (118)$$

Since the Hubble constant H_0 can be found by dividing velocity by distance, it is customarily written in the rather baroque units of $\text{km s}^{-1} \text{Mpc}^{-1}$. From Fig. 11 it follows that $H_0 = 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$. However, it turned out that Hubble was severely underestimating the distances to galaxies. In Fig. 12 we show a more recent determination of the Hubble constant from nearby galaxies, using HST data [73]. By combining results of different research groups, the present day Hubble expansion rate is $H_0 = 70^{+5}_{-3} \text{ km s}^{-1} \text{ Mpc}^{-1}$.

EXERCISE 6.2 The Sloan Digital Sky Survey (SDSS) is a survey that mapped positions and distances of a million galaxies using a dedicated 2.5 m telescope in New Mexico [74]. In this exercise, you will use data from this survey to calculate H_0 . In Fig. 13 we show

the spectrum of a star in our galaxy and spectra of four distant galaxies, as measured by the SDSS. For each of the galaxies, we indicate the measured brightness in units of Joules per square meter per second. Assume that each of them has the same luminosity as that of the Milky Way ($L_{\text{MW}} = 10^{11} L_{\odot}$, or $L_{\text{MW}} = 4 \times 10^{37} \text{ J/s}$). (i) Determine the distance to each of the four galaxies, using the inverse-square law relation between brightness and luminosity. Express your answers both in meters and in megaparsecs, and give two significant figures. (ii) The spectrum of each of these objects shows a pair of strong absorption lines of calcium, which have rest wavelength $\lambda_0 = 3935 \text{ \AA}$ and 3970 \AA , respectively. The wavelengths of these lines in the galaxies have been shifted to longer wavelengths (i.e., redshifted), by the expansion of the universe. As a guide, the spectrum of a star like the Sun is shown in the upper panel; the calcium lines are at zero redshift. Measure the redshift of each galaxy. That is, calculate the fractional change in wavelength of the calcium lines. [Hint: The tricky part here is to make sure you are identifying the right lines as calcium. In each case, they are a close pair; for Galaxy #2, they are the prominent absorption dips between 4100 \AA and 4200 \AA . Measure the redshift for both of the calcium lines in each galaxy (in each case the two lines should give the same redshift, of course!). Give your final redshift to two significant figures. Do the intermediate steps of the calculation without rounding; rounding too early can result in errors. (iii) Given the redshifts, calculate the velocity of recession for each galaxy, and in each case use the distances to estimate the Hubble constant, in units of kilometers per second per Megaparsec. You will not get identical results from each of the galaxies, due to measurement uncertainties (but they should all be in the same ballpark), so average the results of the four galaxies to get your final answer.

Now a point worth noting at this juncture is that galaxies do not follow Hubble's law exactly. In addition to the expansion of the universe, galaxy motions are affected by the gravity of specific, nearby structures, such as the pull of the Milky Way and Andromeda galaxies on each other. Each galaxy therefore has a peculiar velocity, where peculiar is used in the sense of "individual," or "specific to itself." Thus, the recession velocity of a galaxy is really

$$v = H_0 d + v_{\text{pec}}, \quad (119)$$

where v_{pec} is the peculiar velocity of the galaxy along the line of sight. If peculiar velocities could have any value, then this would make Hubble's law useless. However, peculiar velocities are typically only about 300 km/s, and they very rarely exceed 1000 km/s. Hubble's law therefore becomes accurate for galaxies that are far away, when $H_0 d$ is much larger than 1000 km/s. Furthermore, we can often estimate what a galaxy's peculiar velocity will be by looking at the nearby structures that will be pulling on it.

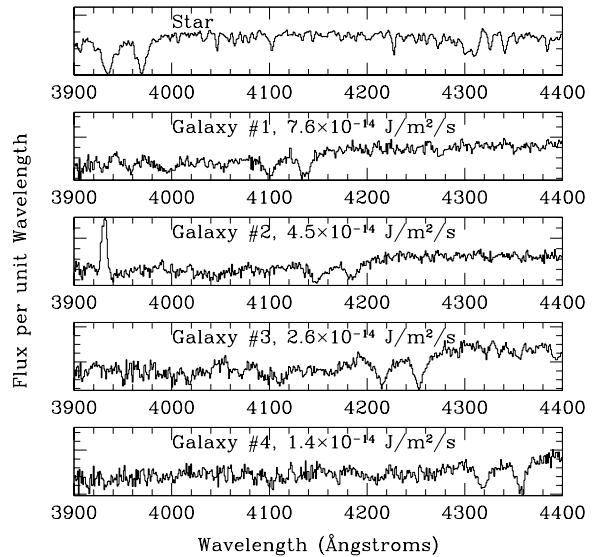


FIG. 13: Spectra measured by the SDSS.

EXERCISE 6.3 Suppose we observe two galaxies, one at a distance of 35 Mly with a radial velocity of 580 km/s, and another at a distance of 1,100 Mly with a radial velocity of 25,400 km/s. (i) Calculate the Hubble constant for each of these two observations. (ii) Which of the two calculations would you consider to be more trustworthy? Why? (iii) Estimate the peculiar velocity of the closer galaxy. (iv) If the more distant galaxy had this same peculiar velocity, how would that change your calculated value of the Hubble constant?

C. The cosmological principle

We would expect intuitively that at any given time the universe ought to look the same to observers in all typical galaxies, and in whatever direction they look. (Hereafter we will use the label "typical" to indicate galaxies that do not have any large peculiar motion of their own, but are simply carried along with the general cosmic flow of galaxies.) This hypothesis is so natural (at least since Copernicus) that it has been called the *cosmological principle* by Milne [75].

As applied to the galaxies themselves, the cosmological principle requires that an observer in a typical galaxy should see all the other galaxies moving with the same pattern of velocities, whatever typical galaxy the observer happens to be riding in. It is a direct mathematical consequence of this principle that the relative speed of any two galaxies must be proportional to the distance between them, just as found by Hubble. To see this consider three typical galaxies at positions \vec{r}_1 , \vec{r}_2 , and \vec{r}_3 . They define the triangle shown in Fig. 14, with sides of

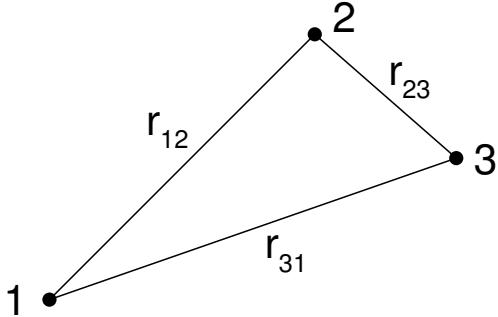


FIG. 14: A triangle defined by three galaxies in a uniformly expanding universe [76].

length

$$\begin{aligned} r_{12} &\equiv |\vec{r}_1 - \vec{r}_2| \\ r_{23} &\equiv |\vec{r}_2 - \vec{r}_3| \\ r_{31} &\equiv |\vec{r}_3 - \vec{r}_1|. \end{aligned} \quad (120)$$

In a homogeneous and uniform expanding universe the shape of the triangle is preserved as the galaxies move away from each other. Maintaining the correct relative lengths for the sides of the triangle requires an expansion law of the form

$$\begin{aligned} r_{12}(t) &= a(t) r_{12}(t_0) \\ r_{23}(t) &= a(t) r_{23}(t_0) \\ r_{31}(t) &= a(t) r_{31}(t_0), \end{aligned} \quad (121)$$

where $a(t)$ is a scale factor, which is totally independent of location or direction. The scale factor $a(t)$ tells us how the expansion (or possibly contraction) of the universe depends on time. At any time t , an observer in galaxy #1 will see the other galaxies receding with a speed

$$\begin{aligned} v_{12}(t) &= \frac{dr_{12}}{dt} = \dot{a} r_{12}(t_0) = \frac{\dot{a}}{a} r_{12}(t) \\ v_{31}(t) &= \frac{dr_{31}}{dt} = \dot{a} r_{31}(t_0) = \frac{\dot{a}}{a} r_{31}(t). \end{aligned} \quad (122)$$

You can easily demonstrate that an observer in galaxy #2 or galaxy #3 will find the same linear relation between observed recession speed and distance, with \dot{a}/a playing the role of the Hubble constant. Since this argument can be applied to any trio of galaxies, it implies that in any universe where the distribution of galaxies is undergoing homogeneous, isotropic expansion, the velocity-distance relation takes the linear form $v = Hr$, with $H = \dot{a}/a$.

If galaxies are currently moving away from each other, this implies they were closer together in the past. Consider a pair of galaxies currently separated by a distance r , with a velocity $v = H_0 r$ relative to each other. If there are no forces acting to accelerate or decelerate their relative motion, then their velocity is constant, and the time that has elapsed since they were in contact is

$$t_H = \frac{r}{v} = H_0^{-1}, \quad (123)$$

independent of the current separation r between galaxies. The time H_0^{-1} is generally referred to as the Hubble time. For $H \approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, the Hubble time is $H_0^{-1} \approx 14 \text{ Gyr}$. If the relative velocities of galaxies have been constant in the past, then one Hubble time ago, all the galaxies in the universe were crammed together into a small volume.

The observation of galaxy redshifts points naturally to a *big bang* description for the evolution of the universe. A big bang model could be broadly defined as a model in which the universe expands from an initially highly dense state to its current low-density state. The Hubble time of $\sim 14 \text{ Gyr}$ is comparable to the ages computed for the oldest known stars in the universe. This rough equivalence is reassuring. However, the age of the universe (i.e. the time elapsed since its original highly dense state) is not necessarily exactly equal to t_H . On the one hand, if gravity working on matter is the only force at work on large scales, then the attractive force of gravity will act to slow down the expansion. If this were the case, the universe was expanding more rapidly in the past than it is now, and the universe is younger than H_0^{-1} . On the other hand, if the energy density of the universe is dominated by a cosmological constant Λ (more on this later), then the dominant gravitational force is repulsive, and the universe may be older than H_0^{-1} .

The horizon distance is defined as the greatest distance a photon can travel during the age of the universe. The Hubble distance, $\mathcal{R}_H = c/H_0 \approx 4.3 \text{ Gpc}$, provides a natural distance scale. However, just as the age of the universe is roughly equal to H_0^{-1} in most big bang models, with the exact value depending on the expansion history of the universe, one horizon is roughly equal to c/H_0 , with the exact value, again, depending on the expansion history.

Before proceeding any further, two qualifications have to be attached to the cosmological principle. First, it is obviously not true on small scales – we are in a Galaxy which belongs to a small local group of other galaxies, which in turn lies near the enormous cluster of galaxies in Virgo. In fact, of the 33 galaxies in Messier's catalogue, almost half are in one small part of the sky, the constellation of Virgo. The cosmological principle, if at all valid, comes into play only when we view the universe on a scale at least as large as the distance between clusters of galaxies, or about 100 million light years. Second, in using the cosmological principle to derive the relation of proportionality between galactic velocities and distances, we suppose the usual rule for adding $v \ll c$. This, of course, was not a problem for Hubble in 1929, as none of the galaxies he studied then had a speed anywhere near the speed of light. Nevertheless, it is important to stress that when one thinks about really large distances characteristic of the universe, as a whole, one must work in a theoretical framework capable of dealing with velocities approaching the speed of light.

Note how Hubble's law ties in with Olbers' paradox. If the universe is of finite age, $t_H \sim H_0^{-1}$, then the night

sky can be dark, even if the universe is infinitely large, because light from distant galaxies has not yet had time to reach us. Galaxy surveys tell us that the luminosity density of galaxies in the local universe is

$$nL \approx 2 \times 10^8 L_\odot \text{ Mpc}^{-3}. \quad (124)$$

By terrestrial standards, the universe is not a well-lit place; this luminosity density is equivalent to a single 40 watt light bulb within a sphere 1 AU in radius. If the horizon distance is $\mathcal{R}_H \approx c/H_0$, then the total flux of light we receive from all the stars from all the galaxies within the horizon will be

$$\begin{aligned} F_{\text{gal}} &\approx nL \int_0^{\mathcal{R}_H} dr \sim nL \frac{c}{H_0} \sim 9 \times 10^{11} L_\odot \text{ Mpc}^{-2} \\ &\sim 2 \times 10^{-11} L_\odot \text{ AU}^{-2}. \end{aligned} \quad (125)$$

By the cosmological principle, this is the total flux of starlight you would expect at any randomly located spot in the universe. Comparing this to the flux we receive from the Sun,

$$F_\odot = \frac{L_\odot}{4\pi \text{ AU}^2} \approx 0.08 L_\odot \text{ AU}^{-2}, \quad (126)$$

we find that $F_{\text{gal}}/F_\odot \sim 3 \times 10^{-10}$. Thus, the total flux of starlight at a randomly selected location in the universe is less than a billionth the flux of light we receive from the Sun here on Earth. For the entire universe to be as well-lit as the Earth, it would have to be over a billion times older than it is; and you would have to keep the stars shining during all that time.

D. Friedmann-Robertson-Walker cosmologies

In 1917 Einstein presented a model of the universe based on his theory of general relativity [77]. It describes a geometrically symmetric (spherical) space with finite volume but no boundary. In accordance with the cosmological principle, the model is homogeneous and isotropic. It is also static: the volume of the space does not change. In order to obtain a static model, Einstein introduced a new repulsive force in his equations. The size of this cosmological term is given by the cosmological constant Λ . Einstein presented his model before the redshifts of the galaxies were known, and taking the universe to be static was then a reasonable assumption. When the expansion of the universe was discovered, this argument in favor of a cosmological constant vanished. Einstein himself later called it the biggest blunder of his life. Nevertheless, the most recent observations seem to indicate that a non-zero cosmological constant has to be present.

In 1922, Friedmann [78, 79] studied the cosmological solutions of Einstein equations. If $\Lambda = 0$, only evolving, expanding or contracting models of the universe are possible. The general relativistic derivation of the law of expansion for the Friedmann models will not be given

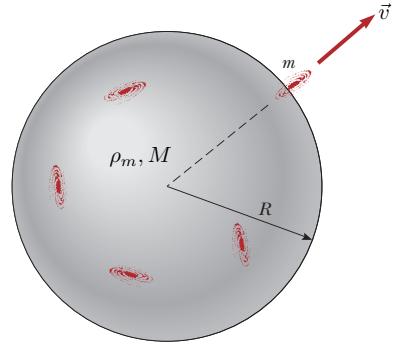


FIG. 15: Spherical region of galaxies with a larger radius than the distance between clusters of galaxies, but smaller radius than any distance characterizing the universe as a whole.

here. It is interesting that the existence of three types of models and their law of expansion can be derived from purely Newtonian considerations, with results in complete agreement with the relativistic treatment. Moreover, the essential character of the motion can be obtained from a simple energy argument, which we discuss next.

Consider a spherical region of galaxies of radius R . (For the purposes of this calculation we must take R to be larger than the distance between clusters of galaxies, but smaller than any distance characterizing the universe as a whole, as shown in Fig. 15. We also assume $\Lambda = 0$.) The mass of this sphere is its volume times the cosmic mass density,

$$M = \frac{4\pi R^3}{3} \rho_m. \quad (127)$$

We can now consider the motion of a galaxy of mass m at the edge of the spherical region. According to Hubble's law, the velocity of the galaxy is $v = HR$, and its corresponding kinetic energy

$$K = \frac{1}{2}mv^2 = \frac{1}{2}mH^2R^2. \quad (128)$$

In a spherical distribution of matter, the gravitational force on a given spherical shell depends only on the mass inside the shell. The potential energy at the edge of the sphere is

$$U = -\frac{GMm}{R} = -\frac{4\pi mR^2\rho_m G}{3}. \quad (129)$$

Hence, the total energy is

$$E = K + U = \frac{1}{2}mH^2R^2 - Gm\frac{4\pi}{3}R^2\rho_m. \quad (130)$$

which has to remain constant as the universe expands. Likewise,

$$\frac{2E}{mR^2} = H^2 - \frac{8\pi}{3}G\rho_m. \quad (131)$$

Since we assume that the universe is homogeneous, H and ρ_m cannot be functions of R . Thus, the left-hand-side of (131) cannot depend on the chosen distance R to the coordinate center. However, the value of $2E/(mR^2)$ is time-dependent, because the distance between us and the galaxy will change as the universe expands. Since the mass m of our test galaxy is arbitrary, we can choose it such that $|2E/(mc^2)| = 1$ holds at an arbitrary moment as long as $E \neq 0$. For different times, the left-hand-side scales as R^{-2} and thus we can rewrite (131) as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho_m - \frac{kc^2}{a^2R_0^2}. \quad (132)$$

Note that because E is constant, k is constant too. Actually, $k = 0, \pm 1$ is generally known as the curvature constant. Throughout the subscripted “0”’s indicate that quantities (which in general evolve with time) are to be evaluated at present epoch. Finally, we account for the equivalence of mass and energy by including not only the mass but also the energy density, $\rho = \rho_m c^2 + \dots$ and so (132) becomes

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\frac{\rho}{c^2} - \frac{kc^2}{a^2R_0^2}. \quad (133)$$

which is *Friedmann equation* (without cosmological constant) in the Newtonian limit. (133) agrees exactly with the equation derived from general relativity [52]. For $k = 0$, the value of H fixes the so-called *critical density* as

$$\rho(k=0) \equiv \rho_c = \frac{3H^2c^2}{8\pi G}. \quad (134)$$

Since we know the current value of the Hubble parameter to within 10%, we can compute the current value of the critical density to within 20%. We usually *hide* this uncertainty by introducing h ,

$$H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad (135)$$

such that

$$\begin{aligned} \rho_{c,0} &= 2.77 \times 10^{11} h^2 M_\odot / \text{Mpc}^3 \\ &= 1.88 \times 10^{-29} h^2 \text{ g/cm}^3 \\ &= 1.05 \times 10^{-5} h^2 \text{ GeV/cm}^3. \end{aligned} \quad (136)$$

Note that since $h \approx 0.70^{+0.05}_{-0.03}$ a flat universe requires an energy density of ~ 10 protons per cubic meter.

The expansion of the universe can be compared to the motion of a mass launched vertically from the surface of a celestial body. The form of the orbit depends on the initial energy. In order to compute the complete orbit, the mass of the main body and the initial velocity have to be known. In cosmology, the corresponding parameters are the mean density and the Hubble constant. On the one hand, if the density exceeds the critical density, the expansion of any spherical region will turn to a contraction

and it will collapse to a point. This corresponds to the closed Friedmann model. On the other hand, if $\rho_m < \rho_c$, the ever-expanding hyperbolic model is obtained. These three models of the universe are called the standard models. They are the simplest relativistic cosmological models for $\Lambda = 0$. Models with $\Lambda \neq 0$ are mathematically more complicated, but show the same behaviour. The simple Newtonian treatment of the expansion problem is possible because Newtonian mechanics is approximately valid in small regions of the universe. However, although the resulting equations are formally similar, the interpretation of the quantities involved is not the same as in the relativistic context. The global geometry of Friedmann models can only be understood within the general theory of relativity [52].

Next, we define the abundance Ω_i of the different players in cosmology as their energy density relative to ρ_c . For example, the dimensionless mass density parameter is found to be

$$\Omega_m = \frac{\rho_m c^2}{\rho_c} = \frac{8\pi G}{3H^2} \rho_m. \quad (137)$$

For simplicity, for the moment we will keep considering scenarios with $\Lambda = 0$, but we advance the reader that

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H^2}. \quad (138)$$

Now, what about our universe? On a large scale what is the overall curvature of the universe? Does it have positive curvature, negative curvature, or is it flat? By solving Einstein equations, Robertson [80, 81] and Walker [82], showed that the three hypersurfaces of constant curvature (the hyper-sphere, the hyper-plane, and the hyper-pseudosphere) are indeed possible geometries for a homogeneous and isotropic universe undergoing expansion. The metric they derived, independently of each other, is called the Friedmann-Robertson-Walker (FRW) metric. The line element is most generally written in the form

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{d\varrho^2}{1 - k\varrho^2/R^2} + \varrho^2 d\Omega^2 \right], \quad (139)$$

where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$. It is easily seen that the spatial component of the FRW metric consists of the spatial metric for a uniformly curved space of radius R , scaled by the square of the scale factor $a(t)$. If the universe had a positive curvature $k = 1$, then the universe would be closed, or finite in volume. This would not mean that the stars and galaxies extended out to a certain boundary, beyond which there is empty space. There is no boundary or edge in such a universe. If a particle were to move in a straight line in a particular direction, it would eventually return to the starting point – perhaps eons of time later. On the other hand, if the curvature of the space was zero $k = 0$ or negative $k = -1$, the universe would be open. It could just go on forever.

Using the substitution

$$\varrho = S_k(r) = \begin{cases} R \sin(r/R) & \text{for } k = +1 \\ r & \text{for } k = 0 \\ R \sinh(r/R) & \text{for } k = -1 \end{cases}; \quad (140)$$

the FRW line element can be rewritten as

$$ds^2 = c^2 dt^2 - a^2(t) [dr^2 + S_k^2(r) d\Omega^2]; \quad (141)$$

see Appendix D for details.

The time variable t in the FRW metric is the cosmological proper time, called the cosmic time for short, and is the time measured by an observer who sees the universe expanding uniformly around him. The spatial variables (ϱ, θ, ϕ) or (r, θ, ϕ) are called the comoving coordinates of a point in space. If the expansion of the universe is perfectly homogeneous and isotropic, the comoving coordinates of any point remain constant with time.

To describe the time evolution of the scale factor $a(t)$ we need an additional equation describing how the energy content of the universe ρ is affected by expansion. The first law of thermodynamics,

$$dU = TdS - PdV, \quad (142)$$

with $dQ = 0$ (no heat exchange to the outside, since no outside exists) becomes

$$dU = -PdV \Rightarrow \frac{dU}{dt} + P \frac{dV}{dt} = 0. \quad (143)$$

There is a caveat to the statement that the expansion of a homogeneous universe is *adiabatic*: when particles annihilate, such as electrons and positrons, this adds heat and makes the expansion temporarily non-adiabatic. This matters at some specific epochs in the very early universe.

For a sphere of comoving radius R_0 ,

$$V = \frac{4}{3}\pi R_0^3 a^3(t), \quad (144)$$

and so

$$\dot{V} = 4\pi R_0^3 a^2 \dot{a} = 3\frac{\dot{a}}{a}V. \quad (145)$$

Since $U = \rho V$,

$$\dot{U} = \dot{\rho}V + \rho\dot{V} = V \left(\dot{\rho} + 3\frac{\dot{a}}{a}\rho \right). \quad (146)$$

Substituting (145) and (146) into (143) we have

$$V \left(\dot{\rho} + 3\frac{\dot{a}}{a}\rho + 3\frac{\dot{a}}{a}P \right) = 0 \quad (147)$$

and thus

$$\dot{\rho} = -3(\rho + P) \frac{\dot{a}}{a}. \quad (148)$$

This *fluid equation* describes the evolution of energy density in an expanding universe. It tells us that the expansion decreases the energy density both by dilution and by the work required to expand a gas with pressure $P \geq 0$.

To solve this equation, we need an additional equation of state relating P and ρ . Suppose we write this in the form

$$P = w\rho. \quad (149)$$

In principle, w could change with time, but we will assume that any time derivatives of w are negligible compared to time derivatives of ρ . This is reasonable if the equation of state is determined by “microphysics” that is not directly tied to the expansion of the universe. The fluid equation then implies

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}, \quad (150)$$

with solution

$$\frac{\rho}{\rho_0} = \left(\frac{a}{a_0} \right)^{-3(1+w)}. \quad (151)$$

The pressure in a gas is determined by the thermal motion of its constituents. For non-relativistic matter (a.k.a. cosmological dust),

$$w = \frac{P}{\rho} \sim \frac{mv^2}{mc^2} \sim \frac{v}{c^2} \ll 1, \quad (152)$$

where v is the thermal velocity of particles with mass m . To a near-perfect approximation $w = 0$, implying $\rho_m \propto a^{-3}$. Light, or more generally any highly relativistic particle, has an associated pressure (radiation pressure). Pressure is defined as the momentum transfer onto a perfectly reflecting wall per unit time and per unit area. Consider an isotropic distribution of photons (or another kind of particle) moving with the speed of light. The momentum of a photon is given in terms of its energy as $p = E/c = h\nu/c$. Consider now an area element dA of the wall; the momentum transferred to it per unit time is given by the momentum transfer per photon, times the number of photons hitting the area dA per unit time. We will assume for the moment that all photons have the same frequency. If θ denotes the direction of a photon relative to the normal of the wall, the momentum component perpendicular to the wall before scattering is $p_\perp = p \cos \theta$, and after scattering $p_\perp = -p \cos \theta$; the two other momentum components are unchanged by the reflection. Thus, the momentum transfer per photon scattering is $\Delta p = 2p \cos \theta$. The number of photons scattering per unit time within the area dA is given by the number density of photons, n times the area element dA , times the thickness of the layer from which photons arrive at the wall per unit time. The latter is given by $c \cos \theta$, since only the perpendicular velocity component brings them closer to the wall. Putting these terms together, we find for the momentum transfer to the wall

per unit time per unit area the expression

$$P(\theta) = 2h\nu n \cos^2 \theta. \quad (153)$$

Averaging this expression over a half-sphere (only photons moving towards the wall can hit it) then yields

$$P = \frac{1}{3}h\nu n = \frac{1}{3}\rho. \quad (154)$$

Then for radition, $w = 1/3$, implying $\rho_{\text{rad}} \propto a^{-4}$. This behavior also follows from a simple argument: the number density of photons falls as $n \propto a^{-3}$, and the energy per photon falls as $h\nu \propto a^{-1}$ because of cosmological redshift (more on this below).

Next, we obtain an expression for the acceleration of the universe. If we multiply our standard version of the Friedmann equation by a^2 , we get

$$\dot{a}^2 = \frac{8\pi G}{3c^2} \rho a^2 - \frac{kc^2}{R_0^2}. \quad (155)$$

Take the time derivative of (155)

$$2\ddot{a}\dot{a} = \frac{8\pi G}{3c^2} (\dot{\rho}a^2 + 2\rho a\dot{a}). \quad (156)$$

divide by $2\dot{a}\dot{a}$

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3c^2} \left(\dot{\rho} \frac{a}{\dot{a}} + 2\rho \right), \quad (157)$$

and substitutte from the fluid equation

$$\dot{\rho} \frac{a}{\dot{a}} = -3(\rho + P) \quad (158)$$

to obtain the *acceleration equation*

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3c^2} (\rho + 3P). \quad (159)$$

We see that if ρ and P are positive, the expansion of the universe decelerates. Higher P produces stronger deceleration for given ρ , e.g., a radiation-dominated universe decelerates faster than a matter-dominated universe.

In the remainder of this section, we consider a flat universe, i.e., $k = 0$. It is easily seen that for non-relativistic matter, the solution to Friedmann equation (133) is given by

$$a(t) = \left(\frac{t}{t_0} \right)^{2/3} \quad \text{and} \quad \rho(t) = \frac{\rho_0}{a^3} = \frac{\rho_0 t_0^2}{t^2}, \quad (160)$$

with

$$t_0 = \frac{2}{3} \frac{1}{H_0}, \quad (161)$$

where we have used (134). Following the same steps for a bizarre universe, which is dominated today by radiation pressure, yields the solution

$$a(t) = \left(\frac{t}{t_0} \right)^{1/2} \quad \text{and} \quad \rho(t) = \frac{\rho_0}{a^4} = \frac{\rho_0 t_0^2}{t^2}. \quad (162)$$

From this simple exercise we can picture the the time evolution of the universe as follows. In the early universe all matter is relativistic and radiation pressure dominates: $a(t) \propto t^{1/2}$, $\rho_{\text{rad}} \propto t^{-2}$, and $\rho_m \propto a^{-3} \propto t^{-3/2}$. The density of radiation then falls more quickly than that of dust. On the other hand, when dust dominates: $a(t) \propto t^{2/3}$, $\rho_m \propto t^{-2}$, and $\rho_{\text{rad}} \propto a^{-4} \propto t^{8/3}$, hence dust domination increases.

EXERCISE 6.4 Using the Hubble flow $v = H_0 r$ show that the expansion of the universe changes the particle number density according to $\dot{n} = -3H_0 n$.

In closing, we discuss how to measure distances in the FRW spacetime. Consider a galaxy which is far away from us, sufficiently far away that we may ignore the small scale perturbations of spacetime and adopt the FRW line element. In an expanding universe, the distance between two objects is increasing with time. Thus, if we want to assign a spatial distance between two objects, we must specify the time t at which the distance is the correct one. Suppose that you are at the origin, and that the galaxy which you are observing is at a co-moving coordinate position (r, θ, ϕ) . We define a proper distance, as the distance between two events A and B in a reference frame for which they occur simultaneously ($t_A = t_B$). In other words, the proper distance $d_p(t)$ between two points in spacetime is equal to the length of the spatial geodesic between them when the scale factor is fixed at the value $a(t)$. The proper distance between the observer and galaxy can be found using the FRW metric at a fixed time t ,

$$ds^2 = a^2(t) [dr^2 + S_k^2(r) d\Omega^2]. \quad (163)$$

Along the spatial geodesic between the observer and galaxy, the angle (θ, ϕ) is constant, and thus

$$ds = a(t) dr. \quad (164)$$

Likewise, using spatial variables (ϱ, θ, ϕ) we have

$$ds = a(t)[1 - k(\varrho/R)^2]^{-1/2} dr \quad (165)$$

The proper distance d_p is found by integrating over the radial comoving coordinate r

$$d_p = a(t) \int_0^r dr = a(t) r, \quad (166)$$

or using (140)

$$d_p = a(t) \begin{cases} k^{-1/2} \sin^{-1}(\sqrt{k}\varrho/R) & \text{for } k = +1 \\ \varrho & \text{for } k = 0 \\ |k|^{-1/2} \sinh^{-1}(\sqrt{|k|}\varrho/R) & \text{for } k = -1 \end{cases}. \quad (167)$$

In a flat universe, the proper distance to an object is just its coordinate distance, $d_p(t) = a(t)\varrho$. Because $\sin^{-1}(x) > x$ and $\sinh^{-1}(x) < x$, in a closed universe ($k > 0$) the proper distance to an object is greater

than its coordinate distance, while in an open universe ($k < 0$) the proper distance to an object is less than its coordinate distance.

EXERCISE 6.5 A civilization that wants to conquer the universe, which is homogeneous and isotropic, and hence is described by the FRW metric, is getting ready to send out soldiers in all directions to invade all the universe out to a proper distance d_p . Every soldier leaves the galaxy where the civilization was born, and travels through the universe with its spaceship along a geodesic, out to a distance d_p from the original galaxy. At the end of the invasion, which occurs at a fixed time t , all the soldiers stand on a spherical surface at a proper distance d_p from their original galaxy. The total volume that has been invaded is the volume inside this spherical surface. What is the total volume invaded? Answer this question for the following three cases: (i) A flat metric ($k = 0$). (ii) A closed metric ($k = +1$) with radius of curvature R at the cosmic time t when the invaded volume and the proper distance d_p are measured. (iii) An open metric ($k = -1$) with radius of curvature R at the cosmic time t when the invaded volume and the proper distance d_p are measured.

EXERCISE 6.6 Consider a positively curved universe ($k = 1$), in which the sole contribution to the energy density comes from non-relativistic matter. In this case the energy density has the dependence $\rho_m = \rho_{m,0}/a^3$. (i) Write down Friedmann equation for this universe and show that the parametric solution,

$$\begin{aligned} a(\theta) &= \frac{4\pi G \rho_{m,0} R_0^2}{3c^4} (1 - \cos \theta), \\ t(\theta) &= \frac{4\pi G \rho_{m,0} R_0^3}{3c^5} (\theta - \sin \theta), \end{aligned} \quad (168)$$

satisfies the Friedmann equation. Here θ is a dimensionless parameter that runs from 0 to 2π , and R_0 is the present radius of curvature if we have normalized the scale factor at present to $a(t_0) = 1$. (ii) What is a_{\max} , the maximum possible scale factor for this universe? (iii) What is the maximum value that the physical radius of curvature (aR_0) reaches? (iv) What is the age of the universe when this maximum radius is reached? (v) What is t_{crunch} , the time at which the universe undergoes a *big crunch* (that is a recollapse to $a = 0$)?

[Hint: Recall that $\dot{a} = da/dt = da/d\theta d\theta/dt$.]

EXERCISE 6.7 Consider a positively curved universe ($k = -1$), in which the sole contribution to the energy density comes from non-relativistic matter, and so the energy density has the dependence $\rho_m = \rho_{m,0}/a^3$. (ii) Write down Friedmann equation for this universe and show that the parametric solution,

$$\begin{aligned} a(\theta) &= \frac{4\pi G \rho_{m,0} R_0^2}{3c^4} (\cosh \theta - 1), \\ t(\theta) &= \frac{4\pi G \rho_{m,0} R_0^3}{3c^5} (\sinh \theta - \theta), \end{aligned} \quad (169)$$

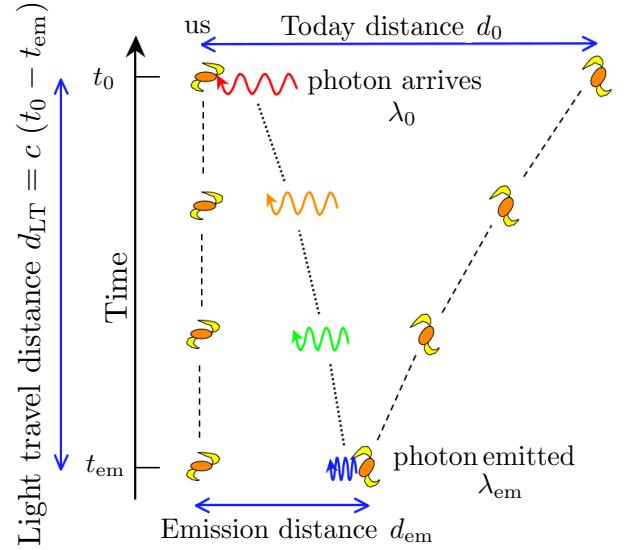


FIG. 16: Cosmological redshift.

satisfies the Friedmann equation. (ii) Compare the time dependence of the scale factor for open, closed and critical matter-dominated cosmological models in a log-log plot.

E. Age and size of the Universe

In special (and general) relativity the propagation of light is along a null geodesic ($ds = 0$). If we place the observer at the origin ($\varrho = 0$), and we choose a radial null geodesic ($d\theta = d\phi = 0$), we have

$$\frac{cdt}{a(t)} = \pm \frac{d\varrho}{[1 - k(\varrho/R)^2]^{1/2}}, \quad (170)$$

where $+$ is for the emitted light ray and the $-$ is for a received one. Imagine now that one crest of the light wave was emitted at time t_{em} at distance ϱ_{em} , and received at the origin $\varrho_0 = 0$ at t_0 , and that the next wave crest was emitted at $t_{\text{em}} + \Delta t_{\text{em}}$ and received at $t_0 + \Delta t_0$; see Fig. 16. The two waves satisfy the relations:

$$\int_{t_{\text{em}}}^{t_0} \frac{dt}{a(t)} = -\frac{1}{c} \int_{\varrho_{\text{em}}}^{\varrho_0} \frac{d\varrho}{\sqrt{1 - k(\varrho/R)^2}} \quad (171)$$

and

$$\int_{t_{\text{em}} + \Delta t_{\text{em}}}^{t_0 + \Delta t_0} \frac{dt}{a(t)} = -\frac{1}{c} \int_{\varrho_{\text{em}}}^{\varrho_0} \frac{d\varrho}{\sqrt{1 - k(\varrho/R)^2}}. \quad (172)$$

Now, subtract (171) from (172)

$$\int_{t_{\text{em}} + \Delta t_{\text{em}}}^{t_0 + \Delta t_0} \frac{dt}{a(t)} - \int_{t_{\text{em}}}^{t_0} \frac{dt}{a(t)} = 0 \quad (173)$$

and expand

$$\begin{aligned} \int_{t_{\text{em}}+\Delta t_{\text{em}}}^{t_0+\Delta t_0} \frac{dt}{a(t)} &= \int_{t_{\text{em}}}^{t_0} \frac{dt}{a(t)} + \int_{t_0}^{t_0+\Delta t_0} \frac{dt}{a(t)} \\ &- \int_{t_{\text{em}}}^{t_{\text{em}}+\Delta t_{\text{em}}} \frac{dt}{a(t)} \end{aligned} \quad (174)$$

to obtain

$$\int_{t_0}^{t_0+\Delta t_0} \frac{dt}{a(t)} = \int_{t_{\text{em}}}^{t_{\text{em}}+\Delta t_{\text{em}}} \frac{dt}{a(t)}. \quad (175)$$

Any change in $a(t)$ during the time intervals between successive wave crests can be safely neglected, so that $a(t)$ is a constant with respect to the time integration. Consequently,

$$\frac{\Delta t_{\text{em}}}{a(t_{\text{em}})} = \frac{\Delta t_0}{a(t_0)}, \quad (176)$$

or equivalently

$$\frac{\Delta t_{\text{em}}}{\Delta t_0} = \frac{a(t_{\text{em}})}{a(t_0)}. \quad (177)$$

The time interval between successive wave crests is the inverse of the frequency of the light wave, related to its wavelength by the relation $c = \lambda\nu$. Hence, from (116) the redshift is

$$z = \frac{\lambda_0}{\lambda_{\text{em}}} - 1 = \frac{a_0}{a(t_{\text{em}})} - 1; \quad (178)$$

i.e., the redshift of a galaxy expresses how much the scale factor has changed since the light was emitted.

The light detected today was emitted at some time t_{em} and, according to (178), there is a one-to-one correspondence between z and t_{em} . Therefore, the redshift z can be used instead of time t to parametrize the history of the universe. A given z corresponds to a time when our universe was $1+z$ times smaller than now.

Generally, the expressions for $a(t)$ are rather complicated and one cannot directly invert (178) to express the cosmic time $t \equiv t_{\text{em}}$ in terms of the redshift parameter z . It is useful, therefore, to derive a general integral expression for $t(z)$. Differentiating (178) we obtain

$$dz = -\frac{a_0}{a^2(t)} \dot{a}(t) dt = -(1+z)H(t)dt, \quad (179)$$

from which follows that

$$t = \int_z^\infty \frac{dz}{H(z)(1+z)}. \quad (180)$$

A constant of integration has been chosen here so that $z \rightarrow \infty$ corresponds to the initial moment of $t = 0$.

To obtain the expression for the Hubble parameter H in terms of z and the present values of H_0 and $\Omega_{m,0}$, it is

convenient to write the Friedmann equation (133) in the form

$$H^2(z) + \frac{kc^2}{a_0^2 R_0^2} (1+z)^2 = \Omega_{m,0} H_0^2 \frac{\rho_m(z)}{\rho_{m,0}}, \quad (181)$$

where the definitions in (137) and (178) have been used. At $z = 0$, this equation reduces to

$$\frac{kc^2}{a_0^2 R_0^2} = (\Omega_{m,0} - 1) H_0^2, \quad (182)$$

allowing us to express the current value of $a_0 R_0$ in a spatially curved universe ($k \neq 0$) in terms of H_0 and $\Omega_{m,0}$. Taking this into account, we obtain

$$\begin{aligned} H(z) &= H_0 \sqrt{(1 - \Omega_{m,0})(1+z)^2 + \Omega_{m,0} \rho_m(z)/\rho_{m,0}} \\ &= H_0 \sqrt{(1 - \Omega_{m,0})(1+z)^2 + \Omega_{m,0}(1+z)^3}. \end{aligned} \quad (183)$$

We can now complete our program by finding an expression for the comoving radial distance coordinate r as a function of the redshift z . Since photons travel on null geodesics of zero proper time, we see directly from the metric (141) that

$$r = - \int \frac{cdt}{a(t)} = - \int c \frac{dt}{dz} (1+z) dz = c \int \frac{dz}{H(z)}, \quad (184)$$

with $H(z)$ given by (183).

As the universe expands and ages, an observer at any point is able to see increasingly distant objects as the light from them has time to arrive, see Fig. 17. This means that, as time progresses, increasingly larger regions of the universe come into causal contact with the observer. The proper distance to the furthest observable point (the particle horizon) at time t is the “horizon distance”, $d_h(t)$.

Again we return to the FRW metric, placing an observer at the origin ($\varrho = 0$) and letting the particle horizon for this observer at time t be located at radial coordinate distance ϱ_h . This means that a photon emitted at $t = 0$ at ϱ_h will reach the observer at the origin at time t . Recalling photons move along null geodesics ($ds = 0$) and considering only radially traveling photons ($d\theta = d\phi = 0$), we find

$$\int_0^t \frac{dt'}{a(t')} = \frac{1}{c} \int_0^{\varrho_h} \frac{d\varrho}{[1 - k(\varrho/R)^2]^{1/2}}, \quad (185)$$

yielding

$$\varrho_h = \begin{cases} \sin \left[c \int_0^t dt' / a(t') \right] & \text{for } k = +1 \\ c \int_0^t dt' / a(t') & \text{for } k = 0 \\ \sinh \left[c \int_0^t dt' / a(t') \right] & \text{for } k = -1 \end{cases}. \quad (186)$$

If the scale factor evolves with time as $a(t) = t^\alpha$, with $\alpha > 1$, we can see that the time integral in (186) diverges

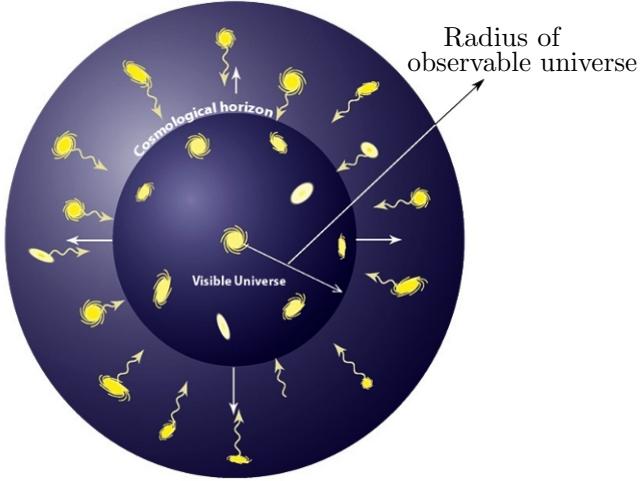


FIG. 17: Cosmological horizon.

as we approach $t = 0$. This would imply that the whole universe is in causal contact. However, $\alpha = 1/2$ and $2/3$ in the radiation and matter-dominated eras, so there is a horizon.

The proper distance from the origin to ϱ_h is given by

$$\begin{aligned} d_h(t) &= a(t) \int_0^{\varrho_h} \frac{d\varrho}{[1 - k(\varrho/R)^2]^{1/2}} \\ &= a(t) \int_0^t \frac{cdt'}{a(t')} . \end{aligned} \quad (187)$$

For $k = 0$, using (160) and (162) we obtain $d_h = 2ct$ in the radiation-dominated era, and $d_h(t) = 3ct$ in the matter-dominated era. Now, substituting (161) into (160) we have

$$a(t) = \left(\frac{3}{2} H_0 t \right)^{2/3} \quad (188)$$

and so from (178) it follows that

$$t = \frac{2}{3} \frac{1}{H_0 (1+z)^{3/2}} . \quad (189)$$

For the matter-dominated era, the proper horizon distance is

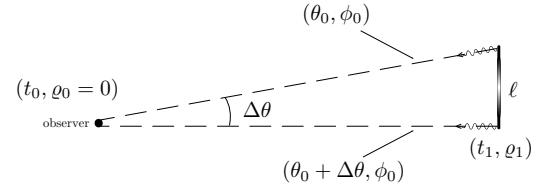
$$d_h = \frac{2c}{H_0 (1+z)^{3/2}} . \quad (190)$$

For a flat universe with $\Omega_{m,0} = 1$, we find that at present time,

$$d_{h,0} = 2c/H_0 = 1.85 \times 10^{28} h^{-1} \text{ cm} = 6 h^{-1} \text{ Gpc}. \quad (191)$$

Note that because $a_0 = 1$, we have $\varrho_{h,0} = d_{h,0}$.

EXERCISE 6.8 Consider a flat model containing only matter, with $\Omega_{m,0} = 1$, and present Hubble constant H_0 . (i) What is the comoving distance to the

FIG. 18: Extended object of given transverse size ℓ at comoving distance q_1 from the observer [83].

horizon ($z = \infty$)? (ii) What is the redshift at which the comoving distance is half that to the horizon? (iii) What is the ratio of the age of the universe at that redshift, to its present age? (iv) At which redshift did the universe have half its present age?

In closing, we show that Hubble's law is indeed an approximation for small redshift by using a Taylor expansion of $a(t)$,

$$\begin{aligned} a(t) &= a(t_0) + (t - t_0)\dot{a}(t_0) + \frac{1}{2}(t - t_0)^2\ddot{a}(t_0) + \dots \\ &= a(t_0) \left[1 + (t - t_0)H_0 - \frac{1}{2}(t - t_0)^2q_0H_0^2 + \dots \right], \end{aligned}$$

where $q_0 \equiv -\ddot{a}(t_0)a(t_0)/\dot{a}^2(t_0)$ is the deceleration parameter (it is named “deceleration” because historically, an accelerating universe was considered unlikely). If the expansion is slowing down, $\ddot{a} < 0$ and $q_0 > 0$. For not too large time-differences, we can use the Taylor expansion of $a(t)$ and write

$$1 - z \approx \frac{1}{1+z} = \frac{a(t)}{a(t_0)} \approx 1 + (t - t_0)H_0 . \quad (192)$$

Hence Hubble's law, $z = (t_0 - t)H_0 = d/cH_0$, is valid as long as $z \ll H_0(t_0 - t) \ll 1$. Deviations from its linear form arises for $z \gtrsim 1$ and can be used to determine q_0 .

F. Angular diameter and luminosity distances

The angular diameter distance to an object is defined in terms of the object's actual size, ℓ , and θ the angular size of the object as viewed from earth. Consider a light source of size ℓ at $\varrho = \varrho_1$ and $t = t_1$ subtending an angle $\Delta\theta$ at the origin ($\varrho = 0, t = t_0$) as shown in Fig. 18. The proper distance ℓ between the two ends of the object is related to $\Delta\theta$ by,

$$\Delta\theta = \frac{\ell}{a(t_1)\varrho_1} . \quad (193)$$

We now define the angular diameter distance

$$d_A = \frac{\ell}{\Delta\theta} \quad (194)$$

so that

$$d_A = a(t_1)\varrho_1 = \frac{\varrho_1}{1+z} . \quad (195)$$

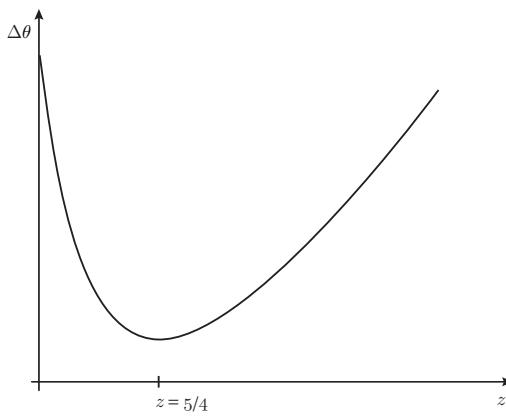


FIG. 19: For a flat universe filled with dust $d_A(z)$ has a maximum at $z = 5/4$, corresponding to the redshift at which objects of a given proper size ℓ will subtend the minimum angle $\Delta\theta$ on the sky. At redshifts $z > 5/4$ objects of a given proper size ℓ will appear bigger on the sky with increasing z [83].

In analogy with (185) we write

$$\int_0^{t_1} \frac{dt}{a(t)} = \frac{1}{c} \int_0^{\varrho_1} \frac{d\varrho}{[1 - k(\varrho/R)^2]^{1/2}}, \quad (196)$$

From an examination point of view, only proficiency in the $k = 0$ case will be expected. Hence,

$$\varrho_1 = c \int_0^{t_1} \frac{dt}{a(t)} = c \int_0^z \frac{dz}{H(z)}, \quad (197)$$

where in the last equality we used (184). Then, for a flat universe filled with dust, the angular diameter as a function of z is

$$\Delta\theta(z) = \frac{\ell H_0}{2c} \frac{(1+z)^{3/2}}{(1+z)^{1/2} - 1}. \quad (198)$$

At low redshifts ($z \ll 1$), the angular diameter decreases in inverse proportion to z , reaches a minimum at $z = 5/4$, and then scales as z for $z \gg 1$; see Fig. 19

Perhaps the most important relation for observational cosmology is that between the monochromatic flux density and luminosity. Start by assuming isotropic emission, so that the photons emitted by the source pass with a uniform flux density through any sphere surrounding the source. We can now make a shift of the origin, and consider the FRW metric as being centred on the source. However, because of homogeneity, the comoving distance between the source and the observer ϱ_1 is the same as we would calculate when we place the origin at our location. The photons from the source are therefore passing through a sphere, on which we sit, of proper surface area $4\pi a_0^2 \varrho_1^2$. However, the redshift still affects the flux density in four further ways: (i) photon energies are redshifted, reducing the flux density by a factor $1+z$; (ii) photon arrival rates are time dilated, reducing the flux density

by a further factor $1+z$; (iii) opposing this, the bandwidth $d\nu$ is reduced by a factor $1+z$, which increases the energy flux per unit bandwidth by one power of $1+z$; (iv) finally, the observed photons at frequency ν_0 were emitted at frequency $(1+z)\nu_0$. Overall, the flux density is the luminosity at frequency $(1+z)\nu_0$, divided by the total area, divided by $(1+z)$:

$$\begin{aligned} \mathcal{F}_\nu(\nu_0) &= \frac{L_\nu([1+z]\nu_0)}{4\pi a_0^2 \varrho_1^2(r)(1+z)} \\ &= \frac{L_\nu(\nu_0)}{4\pi a_0^2 \varrho_1^2(1+z)^{1+\alpha}}, \end{aligned} \quad (199)$$

where the second expression assumes a power-law spectrum $L \propto \nu^{-\alpha}$. We can integrate over ν_0 to obtain the corresponding total or bolometric formulae

$$\mathcal{F} = \frac{L}{4\pi a_0^2 \varrho_1^2(1+z)^2}. \quad (200)$$

The luminosity distance d_L is defined to satisfy the relation (13). Thus,

$$d_L = (1+z)\varrho_1 = (1+z)^2 d_A, \quad (201)$$

where we have taken $a_0 = 1$.

VII. THE FORCE AWAKENS

Independent cosmological observations have unmasked the presence of some unknown form of energy density, related to otherwise empty space, which appears to dominate the recent gravitational dynamics of the universe and yields a stage of cosmic acceleration. We still have no solid clues as to the nature of such dark energy (or perhaps more accurately dark pressure). The cosmological constant is the simplest possible form of dark energy because it is constant in both space and time, and provides a good fit to the experimental data as of today. In this section we will discuss the many observations that probes the dark energy and we will describe the generalities of the concordance model of cosmology with $\Lambda \neq 0$.

A. Supernova Cosmology

The expansion history of the cosmos can be determined using as a “standard candle” any distinguishable class of astronomical objects of known intrinsic brightness that can be identified over a wide distance range. As the light from such beacons travels to Earth through an expanding universe, the cosmic expansion stretches not only the distances between galaxy clusters, but also the very wavelengths of the photons en route. The recorded redshift and brightness of each these candles thus provide a measurement of the total integrated expansion of the universe since the time the light was emitted. A collection of such measurements, over a sufficient range of distances, would

yield an entire historical record of the universe's expansion.

Type Ia supernovae (SNe Ia) are the best cosmological yard sticks in the market. They are precise distance indicators because they have a uniform intrinsic brightness due to the similarity of the triggering white dwarf mass (i.e., $M_{\text{Ch}} = M_{\odot}$) and consequently the amount of nuclear fuel available to burn. This makes SNe Ia the best (or at least most practical) example of "standardizable candles" in the distant universe.

Before proceeding, we pause to present some notation. The apparent magnitude (m) of a celestial object is a number that is a measure of its apparent brightness as seen by an observer on Earth. The smaller the number, the brighter a star appears. The scale used to indicate magnitude originates in the Hellenistic practice of dividing stars visible to the naked eye into six magnitudes. The brightest stars in the night sky were said to be of first magnitude ($m = 1$), whereas the faintest were of sixth magnitude ($m = 6$), which is the limit of human visual perception (without the aid of a telescope). In 1856, Pogson formalized the system by defining a first magnitude star as a star that is 100 times as bright as a sixth-magnitude star, thereby establishing a logarithmic scale still in use today [84]. This implies that a star of magnitude m is $100^{1/5} \simeq 2.512$ times as bright as a star of magnitude $m + 1$. The apparent magnitude, m , in the band, x , is defined as

$$m_x - m_{x,0} = -2.5 \log_{10}(\mathcal{F}_x / \mathcal{F}_{x,0}), \quad (202)$$

where \mathcal{F}_x is the observed flux in the band x , whereas $m_{x,0}$ and $\mathcal{F}_{x,0}$ are a reference magnitude, and reference flux in the same band x , respectively. A difference in magnitudes, $\Delta m = m_1 - m_2$, can then be converted to a relative brightness as $I_2/I_1 \approx 2.512^{\Delta m}$.

In Fig. 20 we show the observed magnitude (and relative brightness) versus redshift for well-measured distant and (in the inset) nearby SNe Ia. The faintness (or distance) of the high-redshift supernovae in Fig. 20 comes as a dramatic surprise. In the (simplest) standard cosmological models described in Sec. VID, the expansion history of the cosmos is determined entirely by its mass density. The greater the density, the more the expansion is slowed by gravity. Thus, in the past, a high-mass-density universe would have been expanding much faster than it does today. So one should not have to look far back in time to especially distant (faint) supernovae to find a given integrated expansion (redshift). Conversely, in a low-mass-density universe one would have to look farther back. But there is a limit to how low the mean mass density could be. After all, we are here, and the stars and galaxies are here. All that mass surely puts a lower limit on how far—that is, to what level of faintness we must look to find a given redshift. However, the high-redshift supernovae in Fig. 20 are fainter than would be expected even for an empty cosmos.

If these data are correct, the obvious implication is that the three simplest models of cosmology introduced

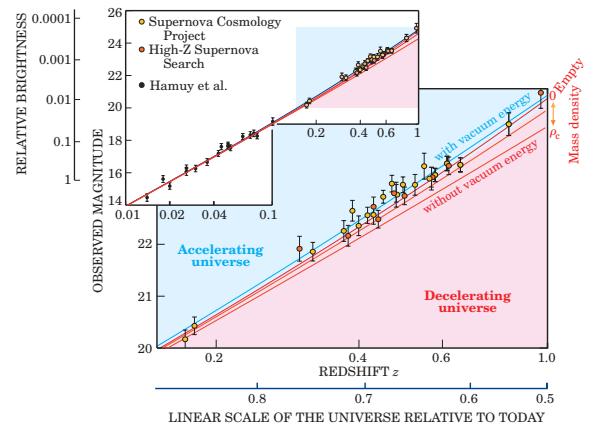


FIG. 20: Observed magnitude (and relative brightness) versus redshift is plotted for well-measured distant [85, 86] and (in the inset) nearby [87, 88] SNe Ia. For clarity, measurements at the same redshift are combined. At redshifts beyond $z = 0.1$ (distances greater than about 109 ly), the cosmological predictions (indicated by the curves) begin to diverge, depending on the assumed cosmic densities of mass and vacuum energy. The red curves represent models with zero vacuum energy and mass densities ranging from ρ_c down to zero (an empty cosmos). The best fit (blue line) assumes a mass density of about $\rho_c/3$ plus a vacuum energy density twice that large, implying an accelerating cosmic expansion [89, 90].

in Sec. VID must be too simple. The next to simplest model includes an expansionary term in the equation of motion driven by the cosmological constant Λ , which competes against gravitational collapse. The best fit to the 1998 supernova data shown in Figs. 20 and 21 implies that, in the present epoch, the vacuum energy density ρ_Λ is larger than the energy density attributable to mass ρ_m . Therefore, the cosmic expansion is now accelerating.

To accommodate SNe Ia data we must add an additional term into the Friedmann equation (133),

$$H^2 = \frac{8\pi}{3} G \frac{\rho}{c^2} - \frac{k c^2}{a^2 R_0^2} + \frac{\Lambda c^2}{3}. \quad (203)$$

The Λ term also modifies the acceleration equation (159), which becomes

$$\frac{\ddot{a}}{a} = \frac{\Lambda c^2}{3} - \frac{4\pi G}{3c^2} (\rho + 3P), \quad (204)$$

and $H(z)$ in (183) is now given by

$$H(z) = H_0 \left\{ \Omega_{m,0}(1+z)^3 + \Omega_{\text{rad},0}(1+z)^4 + \Omega_\Lambda \right. \\ \left. + (1-\Omega_0)(1+z)^2 \right\}^{1/2}, \quad (205)$$

where

$$\Omega = \Omega_m + \Omega_{\text{rad}} + \Omega_\Lambda, \quad (206)$$

and Ω_{rad} is the density fraction of relativistic matter (radiation). We might note in passing that the quantity

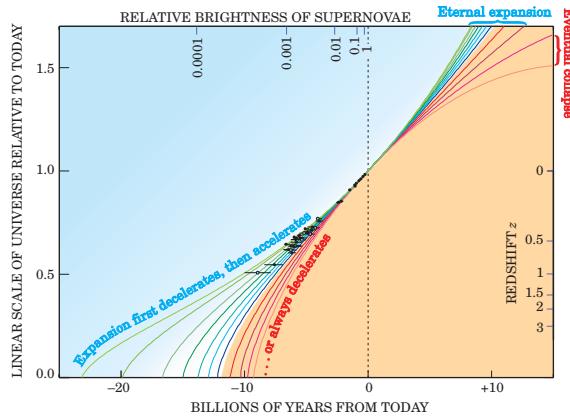


FIG. 21: The history of cosmic expansion, as measured by the high-redshift supernovae (the black data points), assuming flat cosmic geometry. The scale factor a of the universe is taken to be 1 at present, so it equals $1/(1+z)$. The curves in the blue shaded region represent cosmological models in which the accelerating effect of vacuum energy eventually overcomes the decelerating effect of the mass density. These curves assume vacuum energy densities ranging from $0.95 \rho_c$ (top curve) down to $0.4 \rho_c$. In the yellow shaded region, the curves represent models in which the cosmic expansion is always decelerating due to high mass density. They assume mass densities ranging (left to right) from $0.8 \rho_c$ up to $1.4 \rho_c$. In fact, for the last two curves, the expansion eventually halts and reverses into a cosmic collapse [89].

$kc^2/(a^2 R_0^2 H_0^2)$ is sometimes referred to as Ω_k . This usage is unfortunate, because it encourages us to think of curvature as a contribution to the energy density of the universe, which is incorrect.

EXERCISE 7.1 Imagine a class of astronomical objects that are both standard candles and standard yardsticks. In other words, we know both their luminosities L and their physical sizes ℓ . Recall that the *apparent* brightness I of an object is its flux on Earth divided by its angular area, or solid angle on the sky, i.e. $I = \mathcal{F}/\theta^2$, where θ the angular size. How does the *apparent* brightness depend on redshift for a general cosmological model, for these objects with fixed L and ℓ ?

B. Cosmic Microwave Background

The cosmic microwave background (CMB) radiation was discovered in 1964 by Penzias and Wilson, using an antenna built for satellite communication [91]. The radiation was acting as a source of excess noise (or “static”) in the radio receiver. Eventually, it became obvious that the source of noise was actually a signal that was coming from outside the Galaxy. Precise measurements were made at wavelength $\lambda = 7.35$ cm. The intensity of this radiation was found not to vary by day or night or time of the year, nor to depend on the direction to a

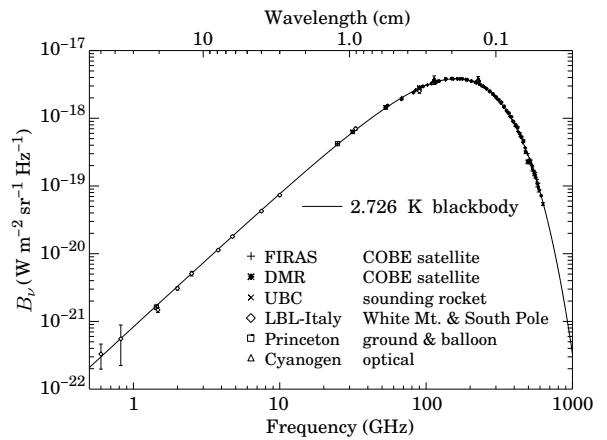


FIG. 22: The CMB blackbody spectrum as confirmed by measurements over a broad range of wavelengths [93].

precision of better than 1%. Almost immediately after its detection it was concluded that this radiation comes from the universe as a whole: a blackbody emission of hot, dense gas (temperature $T \sim 3000$ K, peak wavelength $\lambda_{\max} \sim 1000$ nm) redshifted by a factor of 1000 to $\lambda_{\max} \sim 1$ mm and $T \sim 3$ K [92]. A compilation of experimental measurements in the range $0.03 \text{ cm} \lesssim \lambda \lesssim 75 \text{ cm}$ revealed an accurate blackbody spectrum, see Fig. 22. Actually, according to the FIRAS (Far InfraRed Absolute Spectrometer) instrument aboard the COBE (*Cosmic Background Explorer*) satellite, which measured a temperature of $T_0 = 2.726 \pm 0.010$ K, the CMB is the most perfect blackbody ever seen [94].

The CMB photons we see today interacted with matter for the last time some 380 kyr after the bang. Photon decoupling occurs when the temperature has dropped to a point where there are no longer enough high energy photons to keep hydrogen ionized: ${}^1\text{H}\gamma \neq e^- p^+$. This era is known as recombination, even though the atomic constituents had never been combined prior. The ionization potential of hydrogen is 13.6 eV (i.e., $T \sim 10^5$ K), but recombination occurs at $T_{\text{rec}} \sim 3000$ K. This is because the low baryon to photon ratio, $\eta \approx 5 \times 10^{-10}$, allows the high energy tail of the Planck distribution to keep the comparatively small number of hydrogen atoms ionized until this much lower temperature.

EXERCISE 7.2 (i) For blackbody radiation, the energy density per unit frequency is given by

$$u_\nu d\nu = \frac{8\pi h\nu^3 d\nu}{c^3 [\exp(h\nu/kT) - 1]} . \quad (207)$$

Since the energy of one photon is $h\nu$, the number density of photons is given by the same expression above divided by $h\nu$. Calculate the present density of photons in the universe, knowing that the CMB temperature is $T_0 \simeq 2.726$ K. [Hint: you will find it useful to know that $\int x^2 dx/(e^x - 1) \simeq 2.404$.] (ii) If deuterium measurements require a baryon to photon

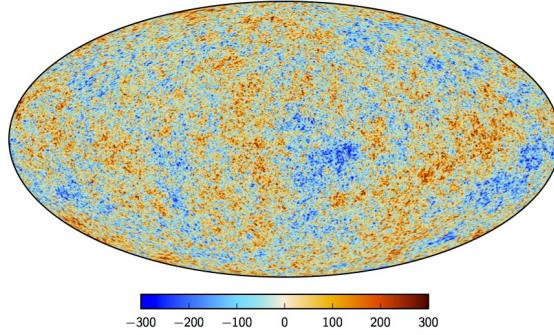


FIG. 23: The CMB over the entire sky, color-coded to represent differences in temperature from the average 2.726 K: the color scale ranges from $+300 \mu\text{K}$ (red) to $-300 \mu\text{K}$ (dark blue), representing slightly hotter and colder spots (and also variations in density.) Results are from the WMAP satellite [95] and the Planck mission [96].

ratio of $\eta = 5.5 \times 10^{-10}$, what must the current density of baryons be? (iii) Assuming that the Hubble constant is $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, calculate what Ω_b is.

Before the recombination epoch the universe was an opaque “fog” of free electrons and became transparent to photons afterwards. Therefore, when we look at the sky in any direction, we can expect to see photons that originated in the “last-scattering surface.” This hypothesis has been tested very precisely by the observed distribution of the CMB; see Fig. 23. The large photon-to-nucleon ratio implies that it is very unlikely for the CMB to be produced in astrophysical processes such as the absorption and re-emission of starlight by cold dust, or the absorption or emission by plasmas. Before the recombination epoch, Compton scattering tightly coupled photons to electrons, which in turn coupled to protons via electromagnetic interactions. As a consequence, photons and nucleons in the early universe behaved as a single “photon-nucleon fluid” in a gravitational potential well created by primeval variations in the density of matter. Outward pressure from photons, acting against the inward force of gravity, set up acoustic oscillations that propagated through the photon-nucleon fluid, exactly like sound waves in air. The frequencies of these oscillations are now seen imprinted on the CMB temperature fluctuations. Gravity caused the primordial density perturbations across the universe to grow with time. The temperature anisotropies in the CMB are interpreted as a snapshot of the early stages of this growth, which eventually resulted in the formation of galaxies [97, 98].

The full sky CMB temperature anisotropy map, as measured by the Wilkinson Microwave Anisotropy Probe (WMAP) [95] and the Planck mission [96], is shown in Fig. 23. It is convenient to expand the difference $\Delta T(\hat{n})$ between the CMB temperature observed in a direction given by the unit vector $\hat{n} = (\theta, \phi)$ and the present mean

value T_0 of the temperature in spherical harmonics

$$\Delta T(\hat{n}) \equiv T(\hat{n}) - T_0 = \sum_{l=0}^{\infty} \sum_{|m| \leq l} a_{lm} Y_{lm}, \quad (208)$$

where

$$T_0 = \frac{1}{4\pi} \int d^2\hat{n} T(\hat{n}), \quad (209)$$

$$a_{lm} = \int \Delta T(\hat{n}) Y_{lm}(\hat{n}) d\Omega, \quad (210)$$

and where Ω denotes the solid angle parametrized by the pair (θ, ϕ) . The set $\{Y_{lm}\}$ is complete and orthonormal, obeying

$$\int d\Omega Y_{l_1 m_1}(\Omega) Y_{l_2 m_2}(\Omega) = \delta_{l_1 l_2} \delta_{m_1 m_2}. \quad (211)$$

Since $\Delta T(\hat{n})$ is real, we are interested in the real-valued, orthonormal Y_{lm} ’s, defined by

$$Y_{lm}(\theta, \phi) = N(l, m) \begin{cases} P_m^l(x)(\sqrt{2} \cos(m\phi)) & m > 0 \\ P_l(x) & m = 0 \\ P_m^l(x)(\sqrt{2} \sin(m\phi)) & m < 0 \end{cases} \quad (212)$$

where

$$N(l, m) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} \quad (213)$$

is a normalization-factor,

$$P_l^m(x) = \frac{(1-x^2)^{m/2}}{2^l l!} \frac{d^{m+l}}{dx^{m+l}}(x^2-1)^l \quad (214)$$

is the associated Legendre polynomial, $P_l = P_{m=0}^l$ is the Legendre polynomial, and $x \equiv \cos\theta$; for further details see e.g. [99].

The lowest multipole is the $l = 0$ monopole, equal to the average full-sky flux and is fixed by normalization (209). The higher multipoles ($l \geq 1$) and their amplitudes a_{lm} correspond to anisotropies. A nonzero m corresponds to $2|m|$ longitudinal “slices” ($|m|$ nodal meridians). There are $l + 1 - |m|$ latitudinal “zones” ($l - |m|$ nodal latitudes). In Fig. 24 we show the partitioning described by some low multipole moments.

EXERCISE 7.3 At every point in the sky, one observes a blackbody spectrum, with temperature $T(\theta)$. The largest anisotropy is in the $l = 1$ (dipole) first spherical harmonic, with amplitude $3.355 \pm 0.008 \text{ mK}$ [101]. The dipole is interpreted to be the result of the Doppler shift caused by the solar system motion relative to the nearly isotropic blackbody field, as broadly confirmed by measurements of the radial velocities of local galaxies. Show that the motion of an observer with velocity

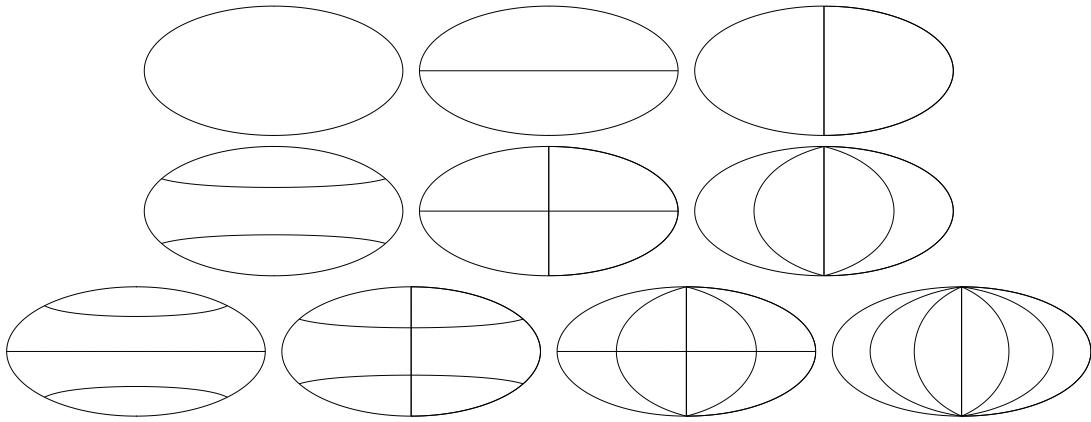


FIG. 24: Nodal lines separating excess and deficit regions of sky for various (l, m) pairs. The top row shows the $(0, 0)$ monopole, and the partition of the sky into two dipoles, $(1, 0)$ and $(1, 1)$. The middle row shows the quadrupoles $(2, 0)$, $(2, 1)$, and $(2, 2)$. The bottom row shows the $l = 3$ partitions, $(3, 0)$, $(3, 1)$, $(3, 2)$, and $(3, 3)$ [100].

$\beta = v/c$ relative to an isotropic Planckian radiation field of temperature T_0 produces a Doppler-shifted temperature pattern

$$T(\theta) \approx T_0 \left[1 + \beta \cos \theta + \frac{\beta^2}{2} \cos(2\theta) + \mathcal{O}(\beta^3) \right]. \quad (215)$$

It is easily seen that the a_{lm} coefficients are frame-dependent. Note that a simple rotation in the ϕ coordinate will change the $\sin \phi, \cos \phi$ part of the spherical harmonic for $m \neq 0$ and a rotation in the θ coordinate will change the associated Legendre polynomial part for $l \neq 0$. So only the $\ell = m = 0$ monopole coefficient is coordinate independent. To combat this problem, we use the power spectrum defined by

$$C_l \equiv \frac{1}{2l+1} \sum_{m=-l}^l a_{lm}^2. \quad (216)$$

A brief C_l initiation is provided in Fig. 26.

EXERCISE 7.4 Show that the power spectrum C_l is invariant under rotations.

To get a rough understanding of the power spectrum we can divide up the multipole representation into super-horizon and sub-horizon regions as shown in Fig. 26. The angular scale corresponding to the particle horizon size is the boundary between super- and sub-horizon scales. The size of a causally connected region on the surface of last scattering is important because it determines the size over which astrophysical processes can occur. Normal physical processes can act coherently only over sizes smaller than the particle horizon. The relative size of peaks and locations of the power spectrum gives information about cosmological parameters [102]. In Fig. 25 we show the influence of several cosmological parameters on the power spectrum. For historical reasons, the

quantity usually used in the multipole representation is

$$\Delta_T \equiv \left[\frac{l(l+1)}{2\pi} C_l \right]^{1/2}. \quad (217)$$

As an illustration, we sketch how to use the power spectrum to determine the curvature of space. At recombination the universe is already matter-dominated, so we can substitute $z_{ls} \simeq 1100$ into (190) to give an estimate of the horizon distance at the CMB epoch

$$d_{h,ls} = \frac{2c}{H_0(1+z)^{3/2}} \approx 0.23 \text{ Mpc}. \quad (218)$$

This is the linear diameter of the largest causally connected region observed for the CMB, ℓ_{ls} . Therefore, substituting (218) into (198) we find today's angular diameter of this region in the sky

$$\theta = \frac{1}{(1+z)^{1/2} - 1} = 0.03 \approx 1.8^\circ. \quad (219)$$

The reason for this “causality problem” is that the universe expands slower than light travels. Namely, as we have seen, when the age of the universe increases the part observable to us increases linearly, $\propto ct$, while the scale factor increases only with $t^{2/3}$ (or $t^{1/2}$). Thus we see more and more regions that were never in causal contact for a radiation or matter-dominated universe. We note that the sound horizon has approximately the same angular size, because of $v_s \sim c/\sqrt{3}$. The sound horizon serves as a ruler at fixed redshift z_{ls} to measure the geometry of spacetime. Moreover, the fluid of photons and nucleons performs acoustic oscillations with its fundamental frequency connected to the sound horizon plus higher harmonics. The relative size of peaks and locations then gives information about cosmological parameters. The first panel of Fig. 25 shows that, for a flat universe ($\Omega_{\text{tot}} \approx 1$, the first peak sits at $\theta \approx 1^\circ$ as we have found in our simple estimate (219)). In Fig. 25 we display a compilation of measurements of the CMB angular

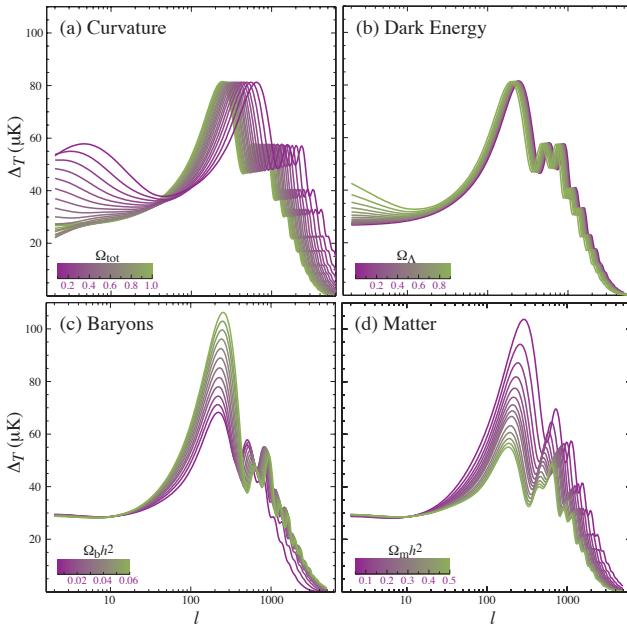


FIG. 25: Influence of several cosmological parameters on the angular power spectrum of the CMB.

power spectrum. The data agree with high significance with models when they input dark energy as providing $\approx 70\%$ of the energy in the universe, and when the total energy density ρ equals the critical density. The data also indicate that the amount of normal baryonic matter in the universe Ω_b is only 4% of the critical density. What is the other 96%?

There is a strong astrophysical evidence for a significant amount of nonluminous matter in the universe referred to as cold dark matter (CDM). For example, observations of the rotation of galaxies suggest that they rotate as they had considerably more mass than we can see [103–105]. Similarly, observations of the motions of galaxies within clusters also suggest they have considerably more mass than can be seen [106]. The most compelling evidence for CDM is that observed at the Bullet Cluster [107]. In Fig. 28 we show a composite image of the Bullet Cluster (1E 0657-558) that shows the X-ray light detected by Chandra in purple, (an image from Magellan and the Hubble space telescope of) the optical light in white and orange, and the CDM map (drwan up using data on gravitational lensing from Magellan and European Space Observatory telescopes at Paranal) in blue. Galaxy clusters contain not only the galaxies ($\sim 2\%$ of the mass), but also intergalactic plasma ($\sim 10\%$ of the mass), and (assuming the null hypothesis) CDM ($\sim 88\%$ of the mass). Over time, the gravitational attraction of all these parts naturally push all the parts to be spatially coincident. If two galaxy clusters were to collide/merge, we will observe each part of the cluster to behave differently. Galaxies will behave as collisionless particles but the plasma will experience ram pressure. Throughout the collision of two

clusters, the galaxies will then become separated from the plasma. This is seen clearly in the Bullet Cluster, which is undergoing a high-velocity (around 4500 km/s) merger, evident from the spatial distribution of the hot, X-ray emitting gas. The galaxies of both concentrations are spatially separated from the (purple) X-ray emitting plasma. The CDM clump (blue), revealed by the weak-lensing map, is coincident with the collisionless galaxies, but lies ahead of the collisional gas. As the two clusters cross, the intergalactic plasma in each cluster interacts with the plasma in the other cluster and slows down. However, the dark matter in each cluster does not interact at all, passing right through without disruption. This difference in interaction causes the CDM to sail ahead of the hot plasma, separating each cluster into two components: CDM (and collisionless galaxies) in the lead and the hot interstellar plasma lagging behind. What might this nonluminous matter in the universe be? We do not know yet. It cannot be made of ordinary (baryonic) matter, so it must consist of some other sort of elementary particle [108].

EXERCISE 7.5 We will examine galaxy rotation curves and show that they imply the existence of dark matter. (i) Recall that the orbital period \mathcal{T} is given by $\mathcal{T}^2 = 4\pi^2 a^3/GM$. Write down an expression that relates the orbital period and the orbital velocity for a circular orbit, and then write down an expression that relates the orbital velocity with the mass enclosed within R . (ii) The Sun is 8 kpc from the center of the Milky Way, and its orbital velocity is 220 km/s. Use your expression from (i) to determine roughly how much mass is contained in a sphere around the center of the Milky Way with a radius equal to 8 kpc? (iii) Assume that the Milky Way is made up of only luminous matter (stars) and that the Sun is at the edge of the galaxy (not quite true, but close). What would you predict the orbital velocity to be for a star 30 kpc from the center? and for 100 kpc? (iv) Observations show that galaxy rotation curves are flat: stars move at the same orbital velocity no matter how far they are from the center. How much mass is actually contained within a sphere of radius 30 kpc? 100 kpc? Take the orbital velocity at these radii to be the same as the orbital velocity of the Sun. (v) What do you conclude from all of this about the contents of our galaxy?

C. Λ CDM

The concordance model of cosmology predicts the evolution of a spatially flat expanding Universe filled with dark energy, dark matter, baryons, photons, and three flavors of left-handed (that is, one helicity state ν_L) neutrinos (along with their right-handed antineutrinos $\bar{\nu}_R$). The best fit to the most recent data from the Planck satellite yields the following parameters: $\Omega_{m,0} = 0.308 \pm 0.013$, $\Omega_{b,0}h^2 = 0.02234 \pm 0.00023$, $\Omega_{\text{CDM},0}h^2 = 0.1189 \pm 0.0022$, $h = 0.678 \pm 0.009$, and

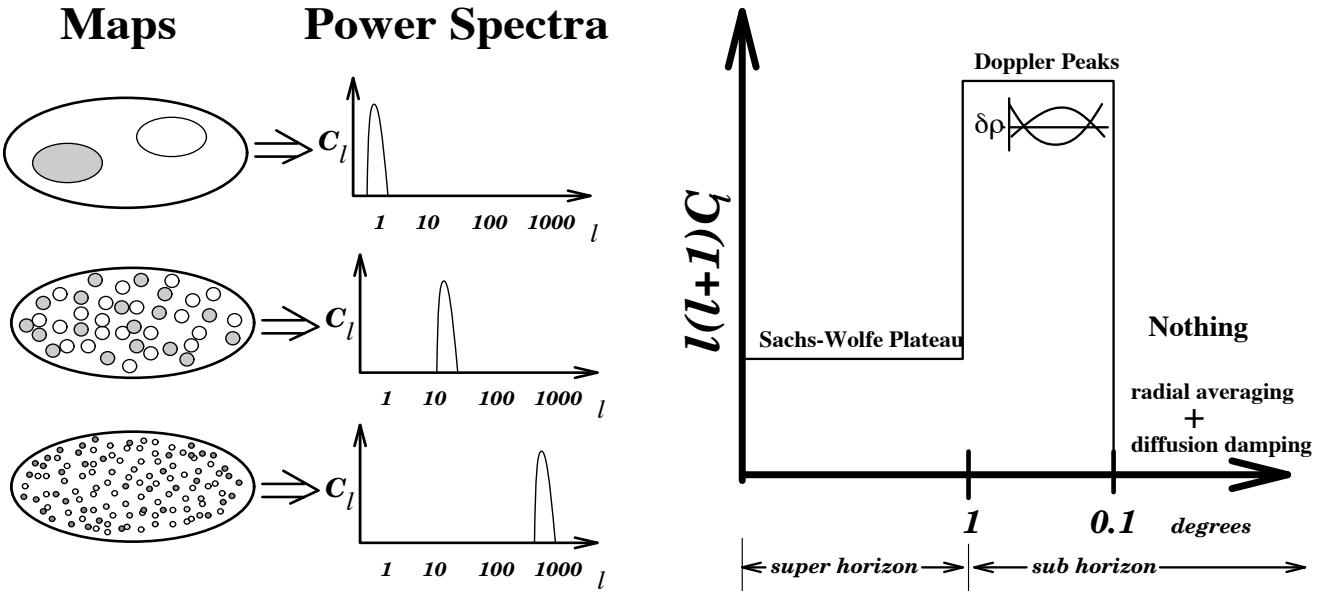


FIG. 26: **Left panel.** Illustrative sky maps and their angular power spectra. If a full-sky CMB map has only a dipole (top), its power spectrum is a delta function at $l = 1$. If a map has only temperature fluctuations on an angular scale of $\sim 7^\circ$ (middle) then all of the power is at $l \sim 10$. If all the hot and cold spots are even smaller (bottom) then the power is at high l . **Right panel** Simplified CMB power spectrum. The CMB power spectrum can be crudely divided into three regions. The Sachs-Wolfe Plateau caused by the scale independence of gravitational potential fluctuations which dominate the spectrum at large super-horizon scales. The horizon is the angular scale corresponding to ct_{ls} . The Doppler peaks on scales slightly smaller than the horizon are due to resonant acoustic oscillations. At smaller scales there is nothing because the finite thickness of the surface of last scattering averages small scale fluctuations along the line of sight. Diffusion damping (photons diffusing out of small scale fluctuations) also suppresses power on these scales [102].

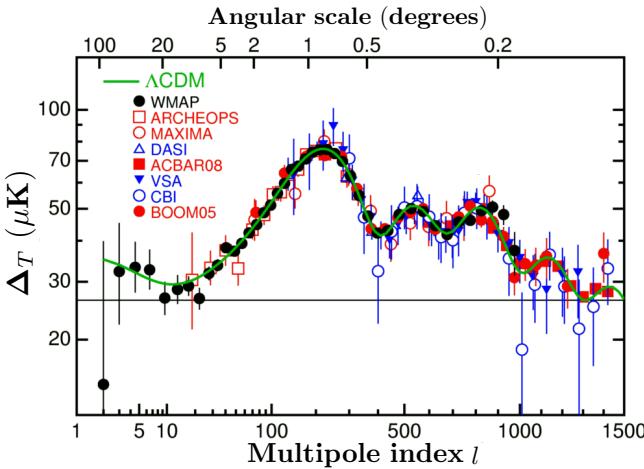


FIG. 27: A compilation of measurements of the CMB angular power spectrum spanning the region $2 \lesssim l \lesssim 1500$. The best fit of the Λ CDM model is also shown.

$1 - \Omega_0 < 0.005$ [109]. Note, however, that the data only measure accurately the acoustic scale, and the relation to underlying expansion parameters (e.g., via the angular-diameter distance) depends on the assumed cosmology, including the shape of the primordial fluctuation spectrum. Even small changes in model assumptions

can change h noticeably. Unexpectedly, the H_0 inference from Planck data deviates by more than 2σ from the previous result from the maser-cepheid-supernovae distance ladder $h = 0.738 \pm 0.024$ [110]. In what follows we will take as benchmark: $\Omega_\Lambda \simeq 0.7$, $\Omega_{m,0} \simeq 0.3$, and $1 - \Omega_0 \simeq 0$, and $h \simeq 0.7$. As shown in Fig. 29, this set of parameters is in good agreement with cosmological and astrophysical observations.

EXERCISE 7.6 The Sun is moving around the center of the Milky Way galaxy along a roughly circular orbit at radius $R = 8$ kpc, with a velocity $v = 220 \text{ km s}^{-1}$. Let us approximate the mass distribution and gravitational potential of the galaxy as spherically symmetric. (i) What is the total mass inside R ? (ii) If the mass density varies with radius as $\rho \propto r^{-2}$, then what is the density at radius R ? Express it in units of proton masses per cm^3 (the proton mass is $m_p = 1.672 \times 10^{-24} \text{ g}$). (iii) For the benchmark cosmological model, with $\Omega_\Lambda \simeq 0.7$ and $H_0 \simeq 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, what is the density of the cosmological constant (or dark energy accounting for it)? Express it in the same units as in the previous question. (iv) The dark energy is spread out uniformly in the universe and it causes a gravitational repulsion which accelerates the expansion of the universe. Do you think the cosmological constant may be strongly affecting the dynamics of stars in our galaxy?

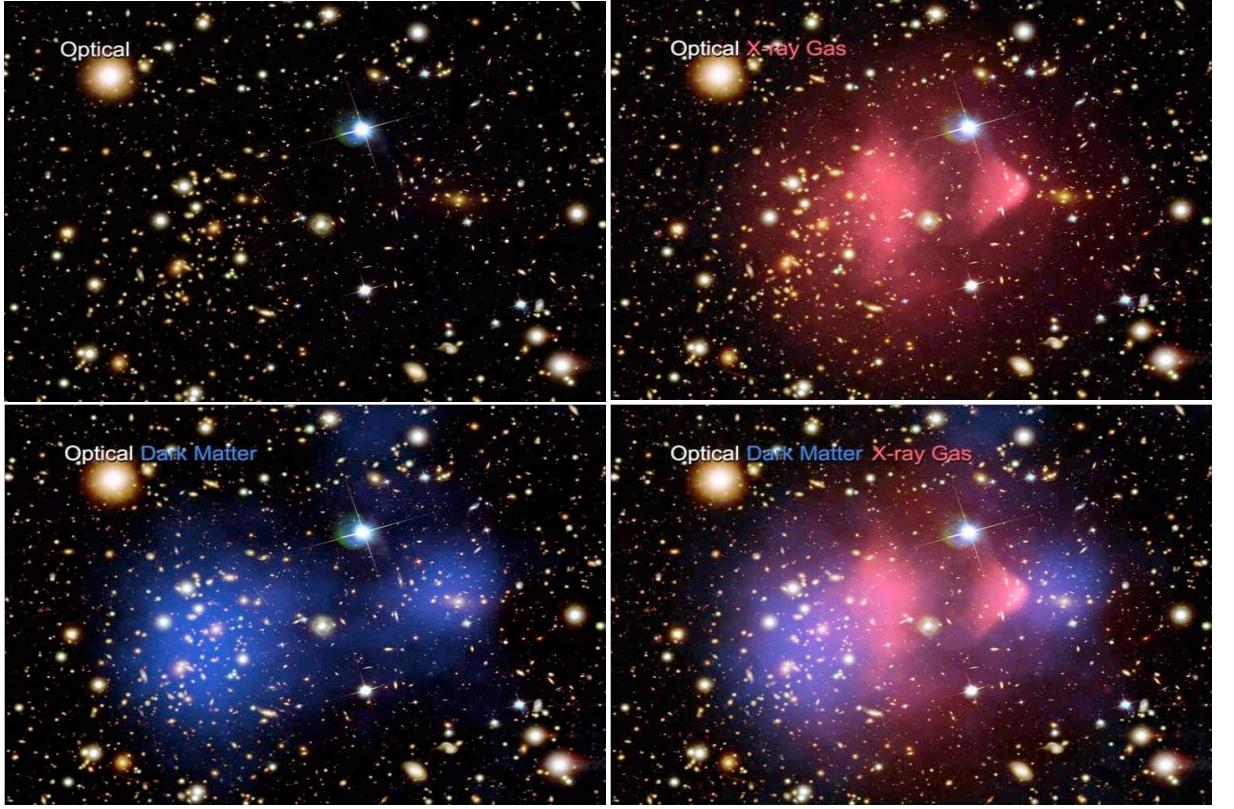


FIG. 28: The Bullet Cluster.

EXERCISE 7.7 Submillimeter Galaxies (SMGs), are extremely dusty starburst galaxies that were discovered at high redshifts ($z \sim 1$ to 3). Assume the dust emission from SMGs are well characterized by blackbodies at a single dust temperature. If the observed spectrum of a SMG peaks at $180 \mu\text{m}$, what would be its dust temperature if it is at a redshift of $z = 2$?

We now consider the benchmark model containing as its only two components pressure-less matter and a cosmological constant, $\Omega_{m,0} + \Omega_\Lambda = 1$. Hence, the curvature term in the Friedmann equation and the pressure term in the acceleration equation play no role. Multiplying the acceleration equation (204) by 2 and adding it to the Friedmann equation (203), we eliminate ρ_m ,

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = \Lambda c^2. \quad (220)$$

Next, we rewrite first the left-hand-side and then the right-hand-side as total time derivatives. Using

$$\frac{d}{dt}(a\dot{a}^2) = \dot{a}^3 + 2a\dot{a}\ddot{a} = \dot{a}a^2 \left[\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{\ddot{a}}{a} \right], \quad (221)$$

it follows that

$$\frac{d}{dt}(a\dot{a}^2) = \dot{a}a^2\Lambda c^2 = \frac{\Lambda c^2}{3} \frac{d}{dt}(a^3). \quad (222)$$

Integration is now trivial,

$$a\dot{a}^2 = \frac{\Lambda c^2}{3}a^3 + \mathcal{C}. \quad (223)$$

The integration constant, $\mathcal{C} = 8\pi G\rho_{m,0}/3$, can be determined most easily by setting $a(t_0) = 1$ and comparing (223) to the Friedmann equation (203), with $t = t_0$. Now, we introduce the new variable $x = a^{3/2}$ such that

$$\frac{da}{dt} = \frac{dx}{dt} \frac{da}{dx} = \frac{dx}{dt} \frac{2x^{-1/3}}{3}, \quad (224)$$

and (223) becomes

$$\dot{x}^2 - \frac{3}{4}\Lambda c^2x^2 + \frac{9}{4}\mathcal{C} = 0. \quad (225)$$

Using an educated guess,

$$x(t) = A \sinh(\sqrt{3\Lambda c^2}t/2), \quad (226)$$

we fix $A = \sqrt{3\mathcal{C}/\Lambda c}$. The scale factor is then

$$a(t) = A^{2/3} \sinh^{2/3}(\sqrt{3\Lambda c^2}t/2). \quad (227)$$

The time-scale of expansion is driven by $t_\Lambda = 2/\sqrt{3\Lambda c^2}$. The present age of the universe t_0 follows from the normalization condition $a(t_0) = 1$ and is given by

$$t_0 = t_\Lambda \tanh^{-1}(\sqrt{\Omega_\Lambda}). \quad (228)$$

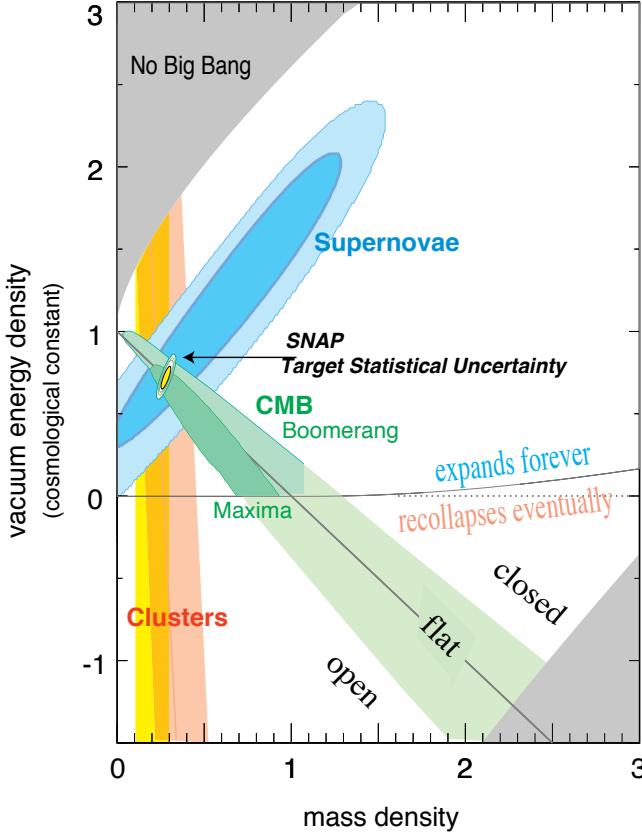


FIG. 29: Shown are three independent measurements of the cosmological parameters $(\Omega_\Lambda, \Omega_m)$. The high-redshift supernovae [111], galaxy cluster abundance [112] and the CMB [113, 114] converge nicely near $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$, as shown by the 68.3%, 95.4%, and 99.7% confidence regions. The upper-left shaded region, labeled “no big bang,” indicates bouncing cosmologies for which the universe has a turning point in its past [115]. The lower right shaded region corresponds to a universe which is younger than long-lived radioactive isotopes [116], for any value of $H_0 \geq 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Also shown is the expected confidence region allowed by the future SuperNova / Acceleration Probe (SNAP) mission [117].

The deceleration,

$$q = -\frac{\ddot{a}}{aH^2}, \quad (229)$$

is a key parameter for observational tests of the Λ CDM model. We calculate first the Hubble parameter

$$H(t) = \frac{\dot{a}}{a} = \frac{2}{3t_\Lambda} \coth(t/t_\Lambda), \quad (230)$$

and after that

$$q(t) = \frac{1}{2} [1 - 3 \tanh^2(t/t_\Lambda)]. \quad (231)$$

Note that, as expected, for $t \rightarrow 0$ we have $q = 1/2$, and for $t \rightarrow \infty$ we have $q = -1$. Perhaps more interesting is

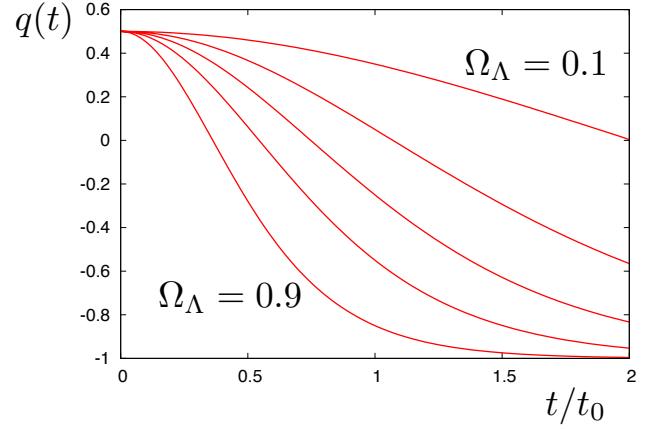


FIG. 30: The deceleration parameter q as a function of t/t_0 for the Λ CDM model for various values of $\Omega_\Lambda = 0.1, 0.3, 0.5, 0.7, 0.9$ from top to bottom.

the transition region from a decelerating to an accelerating universe. As shown in Fig. 30 for $\Omega_\Lambda = 0.7$, this transition takes place at $t \approx 0.55 t_0$. This can be easily converted to a redshift: $z_* = a(t_0)/a(t_*) - 1 \approx 0.7$. Interestingly, z_* can be directly probed by SNe Ia observations.

EXERCISE 7.8 Consider the benchmark model, with $\Omega_{m,0} \simeq 0.3$, $\Omega_\Lambda \simeq 0.7$, with flat space geometry. What was the redshift at which the universe had half its present age?

VIII. HOT THERMAL UNIVERSE

Though we can see only as far as the surface of last scattering, in recent decades a convincing theory of the origin and evolution of the *early universe* has been developed. Most of this theory is based on recent theoretical and experimental advances in elementary particle physics. Hence, before continuing our look back through time, we make a detour to overview the generalities of the standard model of particle physics.²

A. $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$

The *standard model* (SM) is our most modern attempt to answer two simple questions that have been perplexing (wo)mankind throughout the epochs: What is the Universe made of? Why is our world the way it is?

The elementary-particle model accepted today views quarks and leptons as the basic (pointlike) constituents

² You can find a more extensive but still qualitative discussion in [10]. For a more rigorous treatment see e.g. [118–121].

TABLE I: The three generations of quarks and leptons in the Standard Model.

	Fermion	Short-hand	Generation	Charge	Mass	Spin
Quarks	up	u	I		$2.3^{+0.7}_{-0.5}$ MeV	
	charm	c	II	$+\frac{2}{3}$	1.275 ± 0.025 GeV	$\frac{1}{2}$
	top	t	III		173.21 ± 0.51 GeV	
	down	d	I		$4.8^{+0.5}_{-0.3}$ MeV	
	strange	s	II	$-\frac{1}{3}$	95^{+5}_{-5} MeV	$\frac{1}{2}$
	bottom	b	III		4.18 ± 0.03 GeV	
Leptons	electron neutrino	ν_e	I		< 2 eV 95% CL	
	muon neutrino	ν_μ	II	0	< 0.19 MeV 90% CL	$\frac{1}{2}$
	tau neutrino	ν_τ	III		< 18.2 MeV 95%CL	
	electron	e	I		0.511 MeV	
	muon	μ	II	-1	105.7 MeV	$\frac{1}{2}$
	tau	τ	III		1.777 GeV	

TABLE II: The four force carriers.

Force	Boson	Short-hand	Charge	Mass	Spin
Electromagnetic	photon	γ	0	0	1
Weak	W	W^\pm	± 1	80.385 ± 0.015 GeV	1
Weak	Z	Z^0	0	91.1876 ± 0.0021 GeV	1
Strong	gluon	g	0	0	1
Gravitation	graviton	G	0	0	2

of ordinary matter. By pointlike, we understand that quarks and leptons show no evidence of internal structure at the current limit of our resolution. Presently, the world's largest microscope is the Large Hadron Collider (or LHC), a machine that collides beams of protons at a center-of-mass energy $\sqrt{s} = 13$ TeV. Remarkably, 70% of the energy carried into the collision by the protons emerges perpendicular to the incident beams. At a given transverse energy E_\perp , we may roughly estimate the LHC resolution as

$$\ell_{\text{LHC}} \approx \hbar c / E_\perp \approx 2 \times 10^{-19} \text{ TeV m} / E_\perp \\ \approx 2 \times 10^{-20} \text{ m.} \quad (232)$$

There are six quarks and six leptons, together with their antiparticles. These twelve elementary particles are all spin- $\frac{1}{2}$ and fall naturally into three families or generations. Each generation consists of two leptons with electric charges $Q = 0$ and $Q = -1$ and two quarks with $Q = +2/3$ and $Q = -1/3$. The masses of the particles increase significantly with each generation, with the possible exception of the neutrinos [122]. The properties of quarks and leptons are summarized in Table I.

Now, an understanding of how the world is put together requires a theory of how quarks and leptons interact with one another. Equivalently, it requires a theory of the fundamental forces of nature. Four such forces have been identified. They can be characterized on the basis of the following four criteria: the types of particles that experience the force, the relative strength of the force,

TABLE III: Relative force strength for protons in a nucleus.

Force	Relative Strength
Strong	1
Electromagnetic	10^{-2}
Weak	10^{-6}
Gravitational	10^{-38}

the range over which the force is effective, and the nature of the particles that mediate the force. Two of the forces, gravitation and electromagnetism, have an unlimited range; largely for this reason they are familiar to everyone. The remaining forces, which are called simply the weak force and the strong force, cannot be perceived directly because their influence extends only over a short range, no larger than the radius of an atomic nucleus. The electromagnetic force is carried by the photon, the strong force is mediated by gluons, the W and Z bosons transmit the weak force, and the quantum of the gravitational force is called the graviton. The main properties of the force carriers are summarized in Table II. A comparison of the (approximate) relative force strengths for two protons inside a nucleus is given in Table III. Though gravity is the most obvious force in daily life, on a nuclear scale it is the weakest of the four forces and its effect at the particle level can nearly always be ignored.

In the SM quarks and leptons are allotted several additive quantum numbers: electric charge Q , lepton num-

ber $L = L_e + L_\mu + L_\tau$, baryon number B , strangeness s , charmness c , bottomness b , and topness t . For each particle additive quantum number N , the corresponding antiparticle has the additive quantum number $-N$.

The additive quantum numbers Q and B are assumed to be conserved in strong, electromagnetic, and weak interactions. The lepton numbers are not involved in strong interactions, but are strictly conserved in both electromagnetic and weak interactions. The remainder, s , c , b and t are strictly conserved only in strong and electromagnetic interactions, but can undergo a change of one unit in weak interactions.

The quarks have an additional *charge* which enables them to interact strongly with one another. This *charge* is a three-fold degree of freedom which has come to be

known as color [123], and so the gauge theory describing the strong interaction has taken on the name of quantum chromodynamics (QCD) [124, 125]. Each quark flavor can have three colors usually designated red, green, and blue. The antiquarks are colored antired, antigreen, and antiblue. Each quark or antiquark carries a single unit of color or anticolor charge, respectively. The quanta of the color fields are called gluons (as they glue the quarks together). There are eight independent kinds of gluons in $SU(3)_C$, each of which carries a combination of a color charge and an anticolor charge (e.g. red-antigreen). The strong interactions between color charges are such that in nature the quarks (antiquarks) are grouped into composites collectively called hadrons [126–128]:

$$\begin{cases} q\bar{q} & \text{(quark + antiquark)} \\ qqq & \text{(three quarks)} \end{cases} \quad \begin{array}{ll} \text{mesons} & \text{integral spin} \rightarrow \text{Bose-Eisntein statistics [129, 130]} \\ \text{baryons} & \text{half-integral spin} \rightarrow \text{Fermi-Dirac statistics [131, 132]} \end{array} . \quad (233)$$

In QCD each baryon, antibaryon, or meson is colorless. However, these colorless particles may interact strongly via residual strong interactions arising from their composition of colored quarks and/or antiquarks. On the other hand the colorless leptons are assumed to be structureless in the SM and consequently do not participate in strong interactions. Before proceeding we note that aside from binding together quarks inside the hadrons, the strong force indirectly also binds protons and neutrons into atomic nuclei. Such a nuclear force is mediated by pions: spin-0 mesons with masses $m_{\pi^0} = 135.0$ MeV and $m_{\pi^\pm} = 139.6$ MeV.

Electromagnetic processes between electrically charged particles are mediated by massless neutral spin-1 photons. The interaction can be described by a local $U(1)_{\text{EM}}$ gauge theory called quantum electrodynamics (QED). The symmetry properties of QED are unquestionably appealing [133–139]. Moreover, QED has yielded results that are in agreement with experiment to an accuracy of about one part in a billion [140], which makes the theory the most accurate physical theory ever devised. It is the model for theories of the other fundamental forces and the standard by which such theories are judged.

Every quark and lepton of the SM interact weakly. The weak interaction, mediated by the massive W^+ , W^- and Z^0 vector bosons, fall into two classes: (*i*) charge-current (CC) weak interactions involving the W^+ and W^- bosons and (*ii*) neutral current (NC) weak interactions involving the Z^0 boson. The CC interactions, acting exclusively on left-handed particles and right-handed antiparticles, are described by a chiral $SU(2)_L$ local gauge theory, where the subscript L refers to left-handed

particles only.³ On the other hand, the NC interactions act on both left-handed and right-handed particles, similar to the electromagnetic interactions. In fact the SM assumes that both the Z^0 and the photon arise from a mixing of two bosons, W^0 and B^0 , via the electroweak mixing angle θ_W :

$$\begin{aligned} \gamma &= B^0 \cos \theta_W + W^0 \sin \theta_W, \\ Z^0 &= -B^0 \sin \theta_W + W^0 \cos \theta_W. \end{aligned} \quad (234)$$

The electroweak interaction is described by a local gauge theory: $SU(2)_L \otimes U(1)_Y$, where the hypercharge $U(1)_Y$ symmetry involves both left-handed and right-handed particles [141–143]. Experiment requires the masses of the weak gauge bosons W and Z to be heavy so that weak interactions are very short-ranged. The W and Z gauge bosons acquire masses through spontaneous symmetry breaking $SU(2)_L \times U(1)_Y \rightarrow U(1)_{\text{EM}}$. The breaking of the symmetry triggers the Higgs mechanism [144, 145], which gives the relative masses of the W and Z bosons in terms of the electroweak mixing angle,

$$M_W = M_Z \cos \theta_W, \quad (235)$$

while the photon remains massless. In addition, by coupling originally massless fermions to the scalar Higgs

³ A phenomenon is said to be chiral if it is not identical to its mirror image. The spin of a particle may be used to define a handedness for that particle. The chirality of a particle is right-handed if the direction of its spin is the same as the direction of its motion. It is left-handed if the directions of spin and motion are opposite.

field, it is possible to produce the observed physical fermion masses without violating the gauge invariance.

The conspicuously well-known accomplishments of the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ SM of strong and electroweak forces can be considered as the apotheosis of the gauge symmetry principle to describe particle interactions. Most spectacularly, the recent discovery [146, 147] of a new boson with scalar quantum numbers and couplings compatible with those of a SM Higgs has possibly plugged the final remaining experimental hole in the SM, cementing the theory further.

In summary, the fundamental particles can be classified into spin-1/2 fermions (6 leptons and 6 quarks), and spin-1 gauge bosons (γ , W^\pm , Z^0 , and g). The leptons have 18 degrees of freedom: each of the 3 charged leptons has 2 possible chiralities and its associated antiparticle, whereas the 3 neutrinos and antineutrinos have only one chirality (neutrinos are left-handed and antineutrinos are right-handed). The quarks have 72 degrees of freedom: each of the 6 quarks, has the associated antiparticle, three different color states, and 2 chiralities. The gauge bosons have 27 degrees of freedom: a photon has two possible polarization states, each massive gauge boson has 3, and each of the eight independent types of gluon in QCD has 2. The scalar spin-0 Higgs boson, with a mass $m_H \simeq 126$ GeV, has 1 degree of freedom.

B. Equilibrium thermodynamics

The Universe we observe had its beginning in the *big bang*, the cosmic firewall. Because the early universe was to a good approximation in thermal equilibrium, particle reactions can be modeled using the tools of thermodynamics and statistical mechanics. It will be helpful then to take a second detour and revise some concepts of statistical thermodynamics.

Consider a cubic box of volume V , and expand the fields inside into periodic waves with harmonic boundary conditions. The density of states in k -space is

$$dN = g \frac{V}{(2\pi)^3} d^3k, \quad (236)$$

where g is a degeneracy factor and k is the Fourier transform wavenumber. The equilibrium phase space distribution (or occupancy) function for a quantum state of energy E is given by the familiar Fermi-Dirac or Bose-Einstein distributions,

$$f = \frac{1}{e^{(E-\mu)/(kT)} \pm 1}, \quad (237)$$

where T is the equilibrium temperature, k is the Boltzmann constant, μ is the chemical potential (if present), and \pm corresponds to either Fermi or Bose statistics. Throughout we will consider the case $|\mu| \ll T$ and neglect all chemical potentials when computing total thermodynamic quantities. All evidence indicates that this is a

good approximation to describe particle interactions in the super-hot primeval plasma [148].

The number density of a dilute weakly-interacting gas of particles in thermal equilibrium with g internal degrees of freedom is then

$$\begin{aligned} n &= \frac{1}{V} \int f dN \\ &= g \frac{1}{(2\pi\hbar)^3} \int_0^\infty \frac{4\pi p^2 dp}{e^{E/(kT)} \pm 1} \\ &= g \frac{1}{2\pi^2 \hbar^3 c^3} \int_{mc^2}^\infty \frac{(E^2 - m^2 c^4)^{1/2}}{e^{E/(kT)} \pm 1} E dE, \end{aligned} \quad (238)$$

where in the second line we have changed to momentum space, $\vec{p} = \hbar \vec{k}$, and in the third line we used the relativistic relation $E = m^2 c^4 + p^2 c^2$. The analogous expression for the energy density is easily obtain since it is only necessary to multiply the integrand in (238) by a factor of E for the energy of each mode,

$$\rho = \frac{g}{2\pi^2 \hbar^3 c^3} \int_{m_i}^\infty \frac{(E^2 - m^2 c^4)^{1/2}}{e^{E/(kT)} \pm 1} E^2 dE. \quad (239)$$

Recalling that the pressure is the average value of the momentum transfer $\langle p \rangle^2 c^2 / E$ in a given direction, we have

$$P = \frac{g}{6\pi^2 \hbar^3 c^3} \int_{m_i}^\infty \frac{(E^2 - m^2 c^4)^{3/2}}{e^{E/(kT)} \pm 1} dE, \quad (240)$$

with the factor of $1/3$ associated with the assumed isotropy of the momentum distribution.

Let us now compute the above expressions in two asymptotic limits: relativistic and non-relativistic particles, which will be sufficient for our discussion of how the different particle species evolve in the primeval plasma. For $kT \gg mc^2$, the particles behave as if they were massless and the Bose-Einstein and Fermi-Dirac distributions reduce to

$$f(y) = \frac{1}{e^y \pm 1}, \quad (241)$$

where we have defined $y = |\vec{p}|/(kT)$. Using

$$\int_0^\infty \frac{z^{n-1}}{e^z - 1} dz = \Gamma(n) \zeta(n) \quad (242)$$

and

$$\int_0^\infty \frac{z^{n-1}}{e^z + 1} dz = \frac{1}{2^n} (2^n - 2) \Gamma(n) \zeta(n), \quad (243)$$

we obtain

$$\begin{aligned} n &= \left(\frac{kT}{c} \right)^3 \frac{4\pi g}{(2\pi\hbar)^3} \int_0^\infty \frac{y^2 dy}{e^y \pm 1} \\ &= \mathcal{A}_\pm \frac{\zeta(3)}{\pi^2} g \left(\frac{kT}{\hbar c} \right)^3, \\ \rho &= \mathcal{B}_\pm \frac{\pi^2 g}{30(\hbar c)^3} (kT)^4, \\ P &= \frac{1}{3} \rho, \end{aligned} \quad (244)$$

where $\zeta(3) \approx 1.2$ and $\mathcal{A}_- = 1$ for bosons, and $\mathcal{A}_+ = 3/4$ for fermions, $\mathcal{B}_- = 1$ for bosons and $\mathcal{B}_+ = 7/8$ for fermions.⁴

For $kT \ll mc^2$, the exponential factor dominates the denominator in both the Bose-Einstein and Fermi-Dirac distributions in (237), so that the bosonic or fermionic nature of the particles becomes indistinguishable. Furthermore, we have

$$\begin{aligned} E &= (p^2 c^2 + m^2 c^4)^{1/2} = mc^2 \left(1 + \frac{p^2}{m^2 c^2}\right)^{1/2} \\ &\simeq mc^2 + \frac{p^2}{2m}. \end{aligned} \quad (245)$$

Defining $x = |\vec{p}|/\sqrt{2mkT}$, for the number density we obtain the Boltzmann distribution

$$\begin{aligned} n &= e^{-mc^2/(kT)} (2mkT)^{3/2} \frac{4\pi g}{(2\pi\hbar)^3} \int_0^\infty e^{-x^2} x^2 dx \\ &= \frac{g}{\hbar^3} \left(\frac{mkT}{2\pi}\right)^{3/2} e^{-mc^2/(kT)}, \end{aligned} \quad (246)$$

where we have used

$$\int_0^\infty x^n e^{-x^2} dx = \frac{1}{2} \Gamma\left(\frac{1+n}{2}\right), \quad (247)$$

with $n = 2$ and $\Gamma(3/2) = \sqrt{\pi}/2$. From (245) it is easily seen that to leading order $\rho = mc^2 n$ in this case.

To obtain the associated pressure, note that to leading order $|p^2 c^2|/E \simeq |p|^2/m$, so that

$$\begin{aligned} P &\simeq e^{-mc^2/(kT)} (2mkT)^{5/2} \frac{4\pi g}{(2\pi\hbar)^3} \frac{1}{3m} \int_0^\infty x^4 e^{-x^2} dx \\ &= e^{-mc^2/(kT)} (2mkT)^{5/2} \frac{4\pi g}{(2\pi\hbar)^3} \frac{1}{3m} \frac{3\sqrt{\pi}}{8} \\ &= \frac{g}{\hbar^3} \left(\frac{mkT}{2\pi}\right)^{3/2} e^{-mc^2/(kT)} kT \\ &= nkT, \end{aligned} \quad (248)$$

where we have used $\Gamma(5/2) = 3\sqrt{\pi}/4$. Note that (248) is just but the familiar result for a non-relativistic perfect gas, $P = nkT$. Since $kT \ll mc^2$, we have $P \ll \rho$ and the pressure may be neglected for a gas of non-relativistic particles, as we had anticipated.

For a gas of non-degenerate, relativistic species, the average energy per particle is

$$\langle E \rangle = \frac{\rho}{n} = \begin{cases} \frac{\pi^4 k}{30\zeta(3)} T & \text{s} \approx 2.701 T \text{ for bosons} \\ \frac{7\pi^4 k}{180\zeta(3)} T & \text{s} \approx 3.151 T \text{ for fermions} \end{cases}, \quad (249)$$

whereas for a non-relativistic species

$$\langle E \rangle = mc^2 + \frac{3}{2}kT. \quad (250)$$

The internal energy U can be considered to be a function of two thermodynamic variables among P , V , and T . (These variables are related by the equation of state.) Let us choose V and T to be the fundamental variables. The internal energy can then be written as $U(V, T)$. Let us differentiate this function:

$$dU = \left(\frac{\partial U}{\partial V}\right)_T dV + \left(\frac{\partial U}{\partial T}\right)_V dT. \quad (251)$$

This equation can be combined with the first law (142) to give

$$TdS = \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right] dV + \left(\frac{\partial U}{\partial T}\right)_V dT. \quad (252)$$

Now, since the internal energy is a function of T and V we may therefore choose to view S as a function of T and V , and this gives rise to the differential relation

$$dS = \left(\frac{\partial S}{\partial T}\right)_V dT + \left(\frac{\partial S}{\partial V}\right)_T dV. \quad (253)$$

Substituting (253) into (252) and equating the dV and dT parts gives the familiar

$$\frac{\partial U}{\partial T} = T \frac{\partial S}{\partial T} \quad (254)$$

and

$$S = \frac{U + PV}{T}, \quad (255)$$

where we have used the relation for extensive quantities ($\partial S / \partial V = S/V$ and $\partial U / \partial V$).⁵ It is useful to define the entropy density $s = S/V$, which is thus given by

$$s = \rho + P. \quad (256)$$

For photons, we can compute all of the thermodynamic

⁴ The Gamma function is an extension of the factorial function for non-integer and complex numbers. If s is a positive integer, then $\Gamma(s) = (s-1)!$. The Riemann zeta function of a real variable s , defined by the infinite series $\zeta(s) = \sum_{n=1}^\infty 1/n^s$, converges $\forall s > 1$. Note that (243) follows from (242) using the relation $\frac{1}{e^x+1} = \frac{1}{e^x-1} - \frac{2}{e^{2x}-1}$.

⁵ Recall that an extensive property is any property that depends on the size (or extent) of the system under consideration. Take two identical samples with all properties identical and combine them into a single sample. Properties that double (e.g., energy, volume, entropy) are extensive. Properties that remain the same (e.g., temperature and pressure) are intensive.

TABLE IV: Effective numbers of degrees of freedom in SM.

Temperature	New particles	$4g_\rho(T)$
$T < m_e$	γ 's + ν 's	29
$m_e < T < m_\mu$	e^\pm	43
$m_\mu < T < m_\pi$	μ^\pm	57
$m_\pi < T < T_c^*$	π 's	69
$T_c < T < m_{\text{charm}}$	- π 's + $u, \bar{u}, d, \bar{d}, s, \bar{s}$ + gluons	247
$m_c < T < m_\tau$	c, \bar{c}	289
$m_\tau < T < m_{\text{bottom}}$	τ^\pm	303
$m_b < T < m_{W,Z}$	b, \bar{b}	345
$m_{W,Z} < T < m_{\text{Higgs}}$	W^\pm, Z	381
$m_H < T < m_{\text{top}}$	H^0	385
$m_t < T$	t, \bar{t}	427

* T_c corresponds to the confinement–deconfinement transition between quarks and hadrons.

quantities rather easily

$$\begin{aligned} n_\gamma &= \frac{2\zeta(3)}{\pi^2} \left(\frac{kT_\gamma}{\hbar c^3} \right) = 60.42 \left(\frac{kT}{\hbar c} \right)^3 \\ &= 20.28 \left(\frac{T}{K} \right)^3 \text{ photons cm}^{-3}, \\ \rho_\gamma &= \frac{\pi^2}{15} \frac{(kT_\gamma)^4}{(\hbar c)^3} = 0.66 \frac{(kT_\gamma)^4}{(\hbar c)^3}, \\ \langle E_\gamma \rangle &= \frac{\rho}{n} = 3.73 \times 10^{16} \left(\frac{T}{K} \right) \text{ erg}, \\ P_\gamma &= \frac{1}{3} \rho_\gamma, \\ s_\gamma &= \frac{4}{3} \frac{\rho_\gamma}{T_\gamma}. \end{aligned} \quad (257)$$

In the limit $kT \gg m_i c^2$, the total energy density can be conveniently expressed by

$$\begin{aligned} \rho_{\text{rad}} &= \left(\sum_B g_B + \frac{7}{8} \sum_F g_F \right) \frac{1}{(ch)^3} \frac{\pi^2}{30} (kT)^4 \\ &= \frac{1}{(\hbar c)^3} \frac{\pi^2}{30} g_\rho(T) (kT)^4, \end{aligned} \quad (258)$$

where $g_{B(F)}$ is the total number of boson (fermion) degrees of freedom and the sum runs over all boson (fermion) states with $m_i c^2 \ll kT$. The factor of 7/8 is due to the difference between the Fermi and Bose integrals. (258) defines the effective number of degrees of freedom, $g_\rho(T)$, by taking into account new particle degrees of freedom as the temperature is raised. The change in $g_\rho(T)$ (ignoring mass effects) is given in Table IV [149]. At higher temperatures, $g_\rho(T)$ will be model dependent.

EXERCISE 8.1 If in the next 10^{10} yr the volume of the universe increases by a factor of two, what then will be the temperature of the blackbody radiation?

C. The first millisecond

The history of the universe from 10^{-10} seconds to today is based on observational facts: the fundamental laws of high energy physics are well-established up to the energies reached by the LHC. Before 10^{-10} seconds, the energy of the universe exceeds 13 TeV and we lose the comfort of direct experimental guidance. The physics of that era is therefore as speculative as it is fascinating. Herein we will go back to the earliest of times – as close as possible to the *big bang* – and follow the evolution of the Universe.

It is clear that as $a \rightarrow 0$ the temperature increases without limit $T \rightarrow \infty$, but there comes a point at which the extrapolation of classical physics breaks down. This is the realm of quantum black holes, where the thermal energy of typical particles of mass m is such that their de Broglie wavelength is smaller than their Schwarzschild radius. Equating h/mc to $2Gm/c^2$ yields a characteristic mass for quantum gravity known as the Planck mass M_{Pl} .⁶ This mass scale, together with the corresponding length $\hbar/(M_{\text{Pl}}c)$ and time $\hbar/(M_{\text{Pl}}c^2)$ define the system of Planck units:

$$\begin{aligned} M_{\text{Pl}} &\equiv \sqrt{\frac{\hbar c}{G}} \simeq 10^{19} \text{ GeV}, \\ \ell_{\text{Pl}} &\equiv \sqrt{\frac{\hbar G}{c^3}} \simeq 10^{-35} \text{ m}, \\ t_{\text{Pl}} &\equiv \sqrt{\frac{\hbar G}{c^5}} \simeq 10^{-43} \text{ s}. \end{aligned} \quad (259)$$

The Planck time therefore sets the origin of time for the *classical big bang* era. It is inaccurate to extend the classical solution of Friedmann equation to $a = 0$ and conclude that the universe began in a singularity of infinite density.

At $t \sim 10^{-43}$ s, a kind of *phase transition* is thought to have occurred during which the gravitational force *condensed out* as a separate force. The symmetry of the four forces was broken, but the strong, weak, and electromagnetic forces were still unified, and there were no distinctions between quarks and leptons. This is an unimaginably short time, and predictions can be only speculative. The temperature would have been about 10^{32} K, corresponding to *particles* moving about every which way with an average kinetic energy of

$$kT \approx \frac{1.4 \times 10^{-23} \text{ J/K } 10^{32} \text{ K}}{1.6 \times 10^{-10} \text{ J/GeV}} \approx 10^{19} \text{ GeV}, \quad (260)$$

where we have ignored the factor 2/3 in our order of magnitude calculation. Very shortly thereafter, as the

⁶ Strictly speaking this is not quite the Planck mass. It is a factor of $\sqrt{\pi}$ larger. However, this heuristic derivation gives the right order of magnitude.

temperature had dropped to about 10^{28} K, there was another phase transition and the strong force condensed out at about 10^{-35} s after the bang. Now the universe was filled with a *soup* of quarks and leptons. About this time, the universe underwent an incredible exponential expansion, increasing in size by a factor of $\gtrsim 10^{26}$ in a tiny fraction of a second, perhaps $\sim 10^{-34}$ s.

As a matter of fact, the favored Λ CDM model implicitly includes the hypothesis of a very early period in which the scale factor of the universe expands exponentially: $a(t) \propto e^{Ht}$. If the interval of exponential expansion satisfies $\Delta t \gtrsim 60/H$, a small casually connected region can grow sufficiently to accommodate the observed homogeneity and isotropy [150]. To properly understand why this is so, we express the comoving horizon (185) as an integral of the comoving Hubble radius,

$$\varrho_h \equiv c \int_0^t \frac{dt'}{a(t')} = c \int_0^a \frac{da}{Ha^2} = c \int_0^a \frac{1}{aH} d\ln a. \quad (261)$$

At this stage it is important to emphasize a subtle distinction between the comoving horizon ϱ_h and the comoving Hubble radius $c/(aH)$. If particles are separated by distances greater than ϱ_h , they never could have communicated with one another; if they are separated by distances greater than $c/(aH)$, they cannot talk to each other now. This distinction is crucial for the solution to the horizon problem which relies on the following: It is possible that ϱ_h is much larger than $c/(aH)$ now, so that particles cannot communicate today but were in causal contact early on. From (261) we see that this might happen if the comoving Hubble radius in the early universe was much larger than it is now so that ϱ_h got most of its contribution from early times. Hence, we require a phase of decreasing Hubble radius, as illustrated in Fig. 31. The shrinking Hubble sphere is defined by $d(aH)^{-1}/dt < 0$. From the relation $d(aH)^{-1}/dt = -\ddot{a}/(aH)^2$ we see immediately that a shrinking comoving Hubble radius implies accelerated expansion $\ddot{a} > 0$. This explains why inflation is often defined as a period of accelerated expansion. The second time derivative of the scale factor may of course be related to the first time derivative of the Hubble parameter according to

$$\frac{\ddot{a}}{a} = H^2(1 - \epsilon), \quad (262)$$

where $\epsilon \equiv -\dot{H}/H^2$. Acceleration therefore corresponds to $\epsilon < 1$. All in all, H is approximately constant during inflation whereas a grows exponentially, and so this implies that the comoving Hubble radius decreases just as advertised. Now, consulting (159) we infer that $\ddot{a} > 0$ requires a negative pressure: $P < -\rho/3$. To see how this can be realized in various physics models see e.g. [151, 152].

EXERCISE 8.2 (203) can be rearranged to give

$$\frac{8\pi G\rho}{3c^2H^2} - \frac{kc^2}{H^2a^2R_0^2} + \frac{\Lambda c^2}{3} = 1 \quad (263)$$

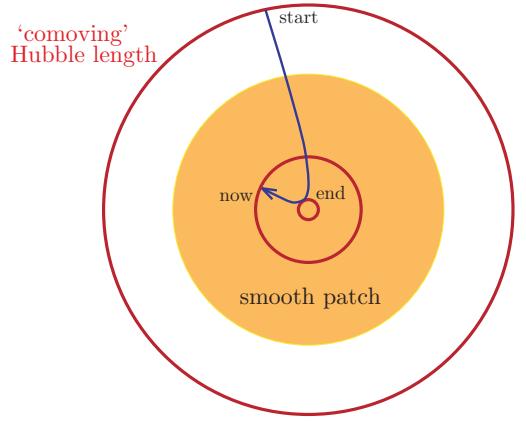


FIG. 31: Evolution of the comoving Hubble radius, $c/(aH)$, in the inflationary universe. The comoving Hubble sphere shrinks during inflation and expands after inflation. Inflation is therefore a mechanism to *zoom-in* on a smooth sub-horizon patch [151].

and so using (206) we rewrite (263) as

$$\Omega - 1 = \frac{kc^2}{a^2R_0^2H^2}. \quad (264)$$

Now let us make a tremendous approximation and assume that Friedmann equation is valid until the Planck era. From (264) we read that if the universe is perfectly flat, then $\Omega = 1$ at all times. However, if there is even a small curvature term, the time dependence of $\Omega - 1$ is quite different. In particular, for the radiation dominated era we have, $H^2 \propto \rho_{\text{rad}} \propto a^{-4}$ and $\Omega - 1 \propto a^2$, whereas during matter domination, $\rho_m \propto a^{-3}$ and $\Omega - 1 \propto a$. In both cases $\Omega - 1$ decreases going backwards in time. Since we know that $\Omega_0 - 1$ is of order unity at present, we can deduce its value at t_{PL} ,

$$\frac{|\Omega - 1|_{T=T_{\text{PL}}}}{|\Omega - 1|_{T=T_0}} \approx \left(\frac{a_{\text{PL}}^2}{a_0^2} \right) \approx \left(\frac{T_0^2}{T_{\text{PL}}^2} \right) \approx \mathcal{O}(10^{-64}). \quad (265)$$

This means that to get the correct value of $\Omega_0 - 1 \sim 1$ today, the value of $\Omega - 1$ at early times has to be fine-tuned to values amazingly close to zero, *but without being exactly zero*. This has been dubbed the *flatness problem*.⁷ Show that the inflationary hypothesis elegantly solve the flatness fine-tuning problem.

After the very brief inflationary period, the universe would have settled back into its more regular expansion. For 10^{-34} s $\lesssim t \lesssim 10^5$ yr, the universe is thought to have been dominated by radiation. This corresponds to 10^3 K $\lesssim T \lesssim 10^{27}$ K. We have seen that the equation of

⁷ A didactic explanation of the flatness fine-tuning problem is given [153].

state can be given by $w = 1/3$. If we neglect the contributions to H from Λ (this is always a good approximation for small enough a) then we find that $a \sim t^{1/2}$ and $\rho_{\text{rad}} \sim a^{-4}$. Substituting (258) into (133) we can rewrite the expansion rate as a function of the temperature in the plasma

$$\begin{aligned} H &= \left(\frac{8\pi G \rho_{\text{rad}}}{3} \right)^{1/2} = \left(\frac{8\pi^3}{90} g_\rho(T) \right)^{1/2} T^2/M_{\text{Pl}} \\ &\sim 1.66 \sqrt{g_\rho(T)} T^2/M_{\text{Pl}}, \end{aligned} \quad (266)$$

where we have adopted natural units ($\hbar = c = k = 1$). Neglecting the T -dependence of g_ρ (i.e. away from mass thresholds and phase transitions), integration of (266) yields (162) and the useful commonly used approximation

$$t \simeq \left(\frac{3M_{\text{Pl}}^2}{32\pi\rho_{\text{rad}}} \right)^{1/2} \simeq 2.42 \frac{1}{\sqrt{g_\rho}} \left(\frac{T}{\text{MeV}} \right)^{-2} \text{ s}. \quad (267)$$

At about 10^{-10} s the Higgs field spontaneously acquires a vacuum expectation value, which breaks the electroweak gauge symmetry. As a consequence, the weak force and electromagnetic force manifest with different ranges. In addition, quarks and charged leptons interacting with the Higgs field become massive. The fundamental interactions have by then taken their present forms.

By the time the universe was about a microsecond old, quarks began to condense into mesons and baryons. To see why, let us focus on the most familiar hadrons: nucleons and their antiparticles. When the average kinetic energy of particles was somewhat higher than 1 GeV, protons, neutrons, and their antiparticles were continually being created out of the energies of collisions involving photons and other particles. But just as quickly, particle and antiparticles would annihilate. Hence the process of creation and annihilation of nucleons was in equilibrium. The numbers of nucleons and antinucleons were high: roughly as many as there were electrons, positrons, or photons. But as the universe expanded and cooled, and the average kinetic energy of particles dropped below about 1 GeV, which is the minimum energy needed in a typical collision to create nucleons and antinucleons (940 MeV each), the process of nucleon creation could not continue. However, the process of annihilation could continue with antinucleons annihilating nucleons, until there were almost no nucleons left; but not quite zero!

Manned and unmanned exploration of the solar system tell us that it is made up of the same stuff as the Earth: *baryons*. Observational evidence from radio-astronomy and cosmic ray detection indicate that the Milky Way, as well as interstellar space, and distant galaxies are also made of baryons. Therefore, we can cautiously conclude that the baryon number of the observable universe is $B > 0$. This requires that the early $q\bar{q}$ plasma contained a tiny surplus of quarks. After all anti-matter annihilated with

matter, only the small surplus of matter remained

$$\eta = \frac{n_B - n_{\bar{B}}}{n_\gamma} = 5 \times 10^{-10} \frac{\text{excess baryons}}{\text{photons}}. \quad (268)$$

The tiny surplus can be explained by interactions in the early universe that were not completely symmetric with respect to an exchange of matter-antimatter, the so-called “baryogenesis” [154].

At this stage, it is worthwhile to point out that if some relativistic particles have decoupled from the photons, it is necessary to distinguish between two kinds of relativistic degrees of freedom (r.d.o.f.): those associated with the total energy density g_ρ , and those associated with the total entropy density g_s . At energies above the deconfinement transition towards the quark gluon plasma, quarks and gluons are the relevant fields for the QCD sector, such that the effective number of interacting (thermally coupled) r.d.o.f. is $g_s(T) = 61.75$. As the universe cools down below the confinement scale $\Lambda_{\text{QCD}} \sim 200$ MeV, the SM plasma transitions to a regime where mesons and baryons are the pertinent degrees of freedom. Precisely, the relevant hadrons present in this energy regime are pions and charged kaons, such that $g_s(T) = 19.25$ [155]. This significant reduction in the degrees of freedom results from the rapid annihilation or decay of more massive hadrons which may have formed during the transition. The quark-hadron crossover transition therefore corresponds to a large redistribution of entropy into the remaining degrees of freedom. To connect the temperature to an effective number of r.d.o.f. we make use of some high statistics lattice simulations of a QCD plasma in the hot phase, especially the behavior of the entropy during the changeover [156]. Concretely, the effective number of interacting r.d.o.f. in the plasma at temperature T is given by

$$g_s(T) \simeq r(T) \left(g_B + \frac{7}{8} g_F \right), \quad (269)$$

where the coefficient $r(T)$ is unity for leptons, two for photon contributions, and is the ratio $s(T)/s_{\text{SB}}$ for the quark-gluon plasma [157]. Here, $s(T)$ and s_{SB} are the entropy density and the ideal Stefan-Bolzmann limit shown in Fig 32. The entropy rise during the confinement-deconfinement changeover can be parametrized, for $150 \text{ MeV} < T < 500 \text{ MeV}$, by

$$\frac{s}{T^3} \simeq \frac{42.82}{\sqrt{392\pi}} e^{-C_1} + 18.62 \frac{C_2^2}{[e^{C_2} - 1]^2} e^{C_2}, \quad (270)$$

where $C_1 = (T_{\text{MeV}} - 151)^2 / 392$ and $C_2 = 195.1 / (T_{\text{MeV}} - 134)$. For the same energy range, we obtain

$$g_s(T) \simeq 47.5 r(T) + 19.25. \quad (271)$$

In Fig. 32 we show $g_s(T)$ as given by (271). The parametrization is in very good agreement with phenomenological estimates [158, 159].

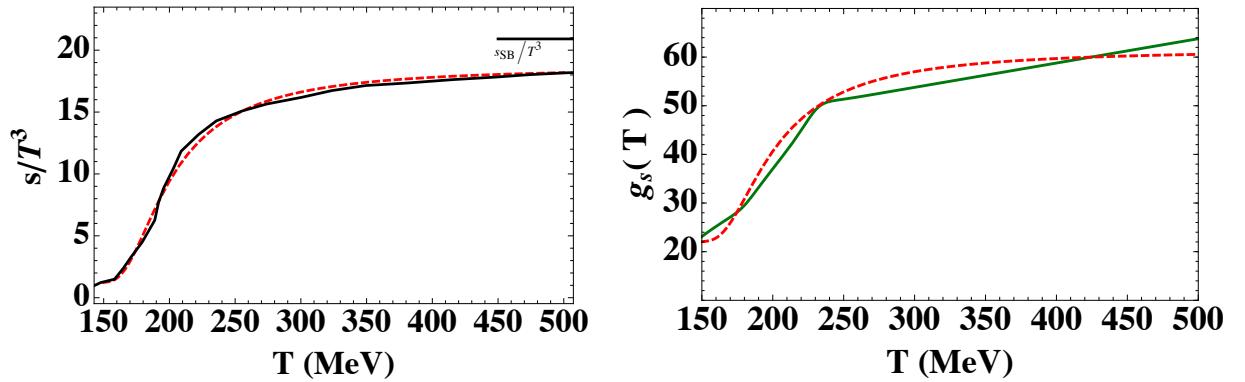


FIG. 32: **Left.** The parametrization of the entropy density given in Eq. (270) (dashed line) superposed on the result from high statistics lattice simulations [156] (solid line). **Right.** Comparison of $g_s(T)$ obtained using (271) (dashed line) and the phenomenological estimate of [158, 159] (solid line) [160].

The entropy density is dominated by the contribution of relativistic particles, so to a very good approximation

$$s = \frac{2\pi^2}{45} g_s(T) T^3. \quad (272)$$

Conservation of $S = sV$ leads to

$$\frac{d}{dt}(sa^3) = 0 \quad (273)$$

and therefore that $g_s(T)T^3a^3$ remains constant as the universe expands. As one would expect, a non-evolving system would stay at constant number or entropy density in comoving coordinates even though the number or entropy density is in fact decreasing due to the expansion of the universe. Since the quark-gluon energy density in the plasma has a similar T dependence to that of the entropy (see e.g. Fig. 7 in [156]), hereafter we simplify the discussion by taking $g = g_\rho = g_s$.

After the first millisecond has elapsed, when the majority of hadrons and anti-hadrons annihilated each other, we entered the lepton era.

D. Neutrino decoupling and BBN

After the first tenth of a second, when the temperature was about 3×10^{10} K, the universe was filled with a plasma of protons, neutrons, electrons, positrons, photons, neutrinos, and antineutrinos (p , n , γ , e^- , e^+ , ν , and $\bar{\nu}$). The baryons are of course nonrelativistic while all the other particles are relativistic. These particles are kept in thermal equilibrium by various electromagnetic and weak processes of the sort $\bar{\nu}\nu = e^+e^-$, $\nu e^- \Rightarrow \nu e^-$, $n\nu_e \Rightarrow p e^-$, $\gamma\gamma \Rightarrow e^+e^-$, $\gamma p \Rightarrow \gamma p$, etc. In complying with the precision demanded of our phenomenological approach it would be sufficient to consider that the cross section of reactions involving left-handed neutrinos, right-handed antineutrinos, and electrons is $\sigma_{\text{weak}} \sim G_F^2 E^2$, where $G_F = 1.16 \times 10^{-5}$ GeV $^{-2}$ is the

Fermi constant. If we approximate the energy E of all particle species by their temperature T , their velocity by c , and their density by $n \sim T^3$, then the interaction rate of is [148]

$$\Gamma_{\text{int},\nu}(T) \approx \langle v\sigma \rangle n_\nu \approx G_F^2 T^5. \quad (274)$$

Comparing (274) with the expansion rate (266), calculated for $g(T) = 10.75$, we see that when the temperature drops below some characteristic temperature $T_{\nu_L}^{\text{dec}}$ neutrinos *decouple*, i.e. they lose thermal contact with electrons [161–163]. The condition

$$\Gamma_{\text{int},\nu}(T_{\nu_L}^{\text{dec}}) = H(T_{\nu_L}^{\text{dec}}) \quad (275)$$

sets the decoupling temperature for left handed neutrinos: $T_{\nu_L}^{\text{dec}} \sim 1$ MeV.

The much stronger electromagnetic interaction continues to keep the protons, neutrons, electrons, positrons, and photons in equilibrium. The reaction rate per nucleon, $\Gamma_{\text{int},N} \sim T^3 \alpha^2 / m_N^2$, is larger than the expansion rate as long as

$$T > \frac{m_N^2}{\alpha^2 M_{\text{Pl}}} \sim \text{a very low temperature}, \quad (276)$$

where the non-relativistic form of the electromagnetic cross section, $\sigma \sim \alpha^2 / m_N^2$, has been obtained by dimensional analysis, with α the fine structure constant. The nucleons are thus maintained in kinetic equilibrium. The average kinetic energy per nucleon is $3T/2$. One must be careful to distinguish between kinetic equilibrium and chemical equilibrium. Reactions like $\gamma\gamma \rightarrow p\bar{p}$ have long been suppressed, as there are essentially no anti-nucleons around.

For $T > m_e \sim 0.5$ MeV $\sim 5 \times 10^9$ K, the number of electrons, positrons, and photons are comparable, $n_{e^-} \sim n_{e^+} \sim n_\gamma$. The exact ratios are of course easily supplied by inserting the appropriate “ g -factors.” Because the universe is electrically neutral, $n_{e^-} - n_{e^+} = n_p$ and so there is a slight excess of electrons over positrons.

When T drops below m_e , the process $\gamma\gamma \rightarrow e^+e^-$ is severely suppressed by the Boltzmann factor $e^{-m_e/T}$, as only very energetic photons in the “tail-end” of the Bose distribution can participate. Thus positrons and electrons annihilate rapidly via $e^+e^- \rightarrow \gamma\gamma$ and are not replenished (leaving a small number of electrons $n_{e^-} \sim n_p \sim 5 \times 10^{-10} n_\gamma$). As long as thermal equilibrium was preserved, the total entropy remained fixed. We have seen that $sa^3 \propto g(T)T^3a^3 = \text{constant}$. For $T \gtrsim m_e$, the particles in thermal equilibrium with the photons include the photon ($g_\gamma = 2$) and e^\pm pairs ($g_{e^\pm} = 4$). The effective total number of particle species before annihilation is $g_{\text{before}} = 11/2$. On the other hand, after the annihilation of electrons and positrons, the only remaining abundant particles in equilibrium are photons. Hence the effective number of particle species is $g_{\text{after}} = 2$. It follows from the conservation of entropy that

$$\frac{11}{2} (T_\gamma a)^3 \Big|_{\text{before}} = 2 (T_\gamma a)^3 \Big|_{\text{after}} . \quad (277)$$

That is, the heat produced by the annihilation of electrons and positrons increases the quantity $T_\gamma a$ by a factor of

$$\frac{(T_\gamma a)|_{\text{after}}}{(T_\gamma a)|_{\text{before}}} = \left(\frac{11}{4}\right)^{1/3} \simeq 1.4 . \quad (278)$$

The energy density stored in relativistic species is customarily given in terms of the so-called “effective number of neutrino species,” N_{eff} , through the relation

$$\rho_{\text{rad}} = \left[1 + \frac{7}{8} \left(\frac{4}{11} \right)^{4/3} N_{\text{eff}} \right] \rho_\gamma , \quad (281)$$

and so

$$N_{\text{eff}} \equiv \left(\frac{\rho_{\text{rad}} - \rho_\gamma}{\rho_\nu} \right) \simeq \frac{8}{7} \sum_B' \frac{g_B}{2} \left(\frac{T_B}{T_\nu} \right)^4 + \sum_F' \frac{g_F}{2} \left(\frac{T_F}{T_\nu} \right)^4 , \quad (282)$$

where ρ_ν denotes the energy density of a single species of massless neutrinos, $T_{B(F)}$ is the effective temperature of boson (fermion) species, and the primes indicate that electrons and photons are excluded from the sums [164, 164]. The normalization of N_{eff} is such that it gives $N_{\text{eff}} = 3$ for three families of massless left-handed standard model neutrinos. For most practical purposes, it is accurate enough to consider that neutrinos freeze-out completely at about 1 MeV. However, as the temperature dropped below this value, neutrinos were still interacting with the electromagnetic plasma and hence received a tiny portion of the entropy from pair annihilations. The non-instantaneous neutrino decoupling gives a correction to the normalization $N_{\text{eff}} = 3.046$ [166–169].

Near 1 MeV, the CC weak interactions,

$$n\nu_e \rightleftharpoons pe^- , \quad ne^+ \rightleftharpoons p\bar{\nu}_e , \quad n \rightleftharpoons pe^-\bar{\nu}_e \quad (283)$$

guarantee neutron-proton chemical equilibrium. Defining λ_{np} as the summed rate of the reactions which convert neutrons to protons,

$$\lambda_{np} = \lambda(n\nu_e \rightarrow pe^-) + \lambda(ne^+ \rightarrow p\bar{\nu}_e) + \lambda(n \rightarrow pe^-\bar{\nu}_e) , \quad (284)$$

the rate λ_{pn} for the reverse reactions which convert protons to neutrons is given by detailed balance:

$$\lambda_{pn} = \lambda_{np} e^{-\Delta m/T(t)} , \quad (285)$$

Before the annihilation of electrons and positrons, the neutrino temperature T_ν is the same as the photon temperature T_γ . But from then on, T_ν simply dropped like a^{-1} , so for all subsequent times, $T_\nu a$ equals the value before annihilation,

$$(T_\nu a)|_{\text{after}} = (T_\nu a)|_{\text{before}} = (T_\gamma a)|_{\text{before}} . \quad (279)$$

We conclude therefore that after the annihilation process is over, the photon temperature is higher than the neutrino temperature by a factor of

$$\left(\frac{T_\gamma}{T_\nu} \right) \Big|_{\text{after}} = \frac{(T_\gamma a)|_{\text{after}}}{(T_\nu a)|_{\text{after}}} \simeq 1.4 . \quad (280)$$

Therefore, even though out of thermal equilibrium, the neutrinos and antineutrinos make an important contribution to the energy density.

EXERCISE 8.3 By assuming that neutrinos saturate the dark matter density derive an upper bound on the neutrino mass.

where $\Delta m \equiv m_n - m_p = 1.293$ MeV. The evolution of the fractional neutron abundance $X_{n/N} \equiv n_n/n_N$ is described by the balance equation

$$\frac{dX_{n/N}(t)}{dt} = \lambda_{pn}(t)[1 - X_{n/N}(t)] - \lambda_{np}(t)X_{n/N}(t) , \quad (286)$$

where n_N is the total nucleon density at this time, $n_N = n_n + n_p$. The equilibrium solution is obtained by setting $dX_{n/N}(t)/dt = 0$:

$$X_{n/N}^{\text{eq}}(t) = \frac{\lambda_{pn}(t)}{\lambda_{pn}(t) + \lambda_{np}(t)} = \left[1 + e^{\Delta m/T(t)}\right]^{-1} . \quad (287)$$

The neutron abundance tracks its value in equilibrium until the inelastic neutron-proton scattering rate decreases sufficiently so as to become comparable to the Hubble expansion rate. At this point the neutrons freeze-out, that is they go out of *chemical* equilibrium. The neutron abundance at the freeze-out temperature $T_{n/N}^{\text{FO}} = 0.75$ MeV can be approximated by its equilibrium value (287),

$$X_{n/N}(T_{n/N}^{\text{FO}}) \simeq X_{n/N}^{\text{eq}}(T_{n/N}^{\text{FO}}) = \left[1 + e^{\Delta m/T_{n/N}^{\text{FO}}}\right]^{-1} . \quad (288)$$

Since the ratio $\Delta m/T_{n/N}^{\text{FO}}$ is of $\mathcal{O}(1)$, a substantial fraction of neutrons survive when chemical equilibrium between neutrons and protons is broken.

At this time, the photon temperature is already below the deuterium binding energy $\Delta_D \simeq 2.2$ MeV, thus one would expect sizable amounts of D to be formed via $n p \rightarrow D \gamma$ process. However, the large photon-nucleon density ratio η^{-1} delays deuterium synthesis until the photo-dissociation process becomes ineffective (deuterium *bottleneck*). Defining the onset of nucleosynthesis by the criterion

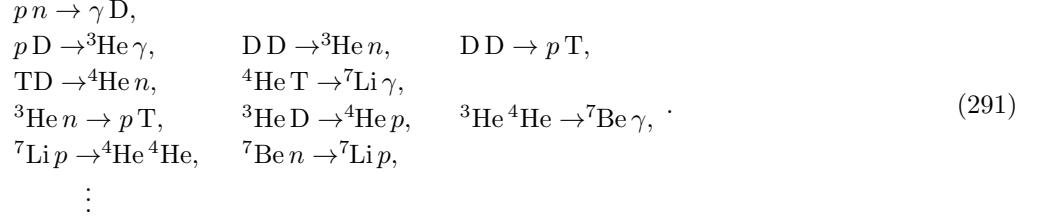
$$e^{\Delta_D/T_{\text{BBN}}} \eta \sim 1 , \quad (289)$$

we obtain $T_{\text{BBN}} \approx 89$ keV. Note that (289) ensures that below T_{BBN} the high energy tail in the photon distribution, with energy larger than Δ_D , has been sufficiently diluted by the expansion. At this epoch, $N(T) = 3.36$, hence the time-temperature relationship (267) dictates that big bang nucleosynthesis (BBN) begins at

$$t_{\text{BBN}} \simeq 167 \text{ s} \approx 180 \text{ s} , \quad (290)$$

as widely popularized by Weinberg [17].

Once D starts forming, a whole nuclear process network sets in [170, 171]. When the temperature dropped below ~ 80 keV, the universe has cooled sufficiently that the cosmic nuclear reactor can begin in earnest, building the lightest nuclides through the following sequence of two-body reactions



By this time the neutron abundance surviving at freeze-out has been depleted by β -decay to

$$X_{n/N}(T_{\text{BBN}}) \simeq X_{n/N}(T_{n/N}^{\text{FO}}) e^{-t_{\text{BBN}}/\tau_n} , \quad (292)$$

where $\tau_n \simeq 887$ s is the neutron lifetime. Nearly *all* of these surviving neutrons are captured in ${}^4\text{He}$ because of its large binding energy ($\Delta_{{}^4\text{He}} = 28.3$ MeV) via the reactions listed in (291). Heavier nuclei do not form in any significant quantity both because of the absence of stable nuclei with $A=5$ or 8 , which impedes nucleosynthesis via $n {}^4\text{He}$, $p {}^4\text{He}$ or ${}^4\text{He} {}^4\text{He}$ reactions, and because of the large Coulomb barrier for reactions such as ${}^4\text{He} T \rightarrow {}^7\text{Li} \gamma$ and ${}^3\text{He} {}^4\text{He} \rightarrow {}^7\text{Be} \gamma$. By the time the temperature has dropped below ~ 30 keV, a time comparable to the neutron lifetime, the average thermal energy of the nuclides and nucleons is too small to overcome the Coulomb barriers; any remaining free neutrons decay, and BBN ceases. The resulting mass fraction of helium, conventionally referred to Y_p , is simply given by

$$Y_p \simeq 2X_{n/N}(t_{\text{BBN}}) = 0.251 , \quad (293)$$

where the subscript p denotes primordial. The above calculation demonstrates how the synthesized helium abundance depends on the physical parameters. After a bit of algebra, (293) can be rweritten as [170]

$$Y_p \simeq 0.251 + 0.014 \Delta N_\nu^{\text{eff}} + 0.0002 \Delta \tau_n + 0.009 \ln \left(\frac{\eta}{5 \times 10^{-10}} \right). \quad (294)$$

In summary, primordial nucleosynthesis has a single adjustable parameter: the baryon density. Observations that led to the determination of primordial abundance of D , ^3He and ^7Li can determine η . The internal consistency of BBN can then be checked by comparing the abundances of the other nuclides, predicted using this same value of η , with observed abundances. Interestingly, in contrast to the other light nuclides, the BBN-predicted primordial abundance of ^4He is very insensitive to the baryon density parameter. Rather, the ^4He mass fraction depends on the neutron-to-proton ratio at BBN because virtually all neutrons available at that time are incorporated into ^4He . Therefore, while D , ^3He , and ^7Li are potential baryometers, ^4He provides a potential chronometer.

EXERCISE 8.4 Suppose that the difference in rest energy of the neutron and proton were 0.1293 MeV, instead of 1.293 MeV, with all other physical parameters unchanged. Estimate the maximum possible mass fraction in ^4He , assuming that all available neutrons are incorporated into ^4He nuclei.

EXERCISE 8.5 A fascinating bit of cosmological history is that of Gamow's prediction of the CMB in the late 1940s [172–174]. Unfortunately, his prediction was premature; by the time the CMB was actually discovered, his prediction had fallen into obscurity. This problem reproduces Gamow's line of argument. Gamow knew that nucleosynthesis must have taken place at a temperature $T_{\text{BBN}} \approx 10^9$ K. He also knew that the universe must currently be $t_0 \sim 10^{10}$ years old. He then assumed that the universe was flat and radiation dominated, even at the present time. (i) With these assumptions, what was the energy density of the universe at the time of nucleosynthesis? (ii) What was the Hubble parameter at the time of nucleosynthesis? (c) What was the age of the universe at BBN? (iv) Given the present age, what should the present temperature of the CMB be? (v) If we then assume that the universe changed from being radiation dominated to matter dominated at a redshift $z_{\text{eq}} > 0$, will this increase or decrease the CMB temperature, for fixed values of T_{BBN} and t_0 ?

The observationally-inferred primordial fractions of baryonic mass in ^4He ($Y_p = 0.2472 \pm 0.0012$, $Y_p = 0.2516 \pm 0.0011$, $Y_p = 0.2477 \pm 0.0029$, and $Y_p = 0.240 \pm 0.006$) [175–177] have been constantly favoring $N_\nu^{\text{eff}} \lesssim 3$ [178]. Unexpectedly, two recent independent

studies yield Y_p values somewhat higher than previous estimates: $Y_p = 0.2565 \pm 0.001(\text{stat}) \pm 0.005(\text{syst})$ and $Y_p = 0.2561 \pm 0.011$ [179–181]. For $\tau_n = 885.4 \pm 0.9$ s and $\tau_n = 878.5 \pm 0.8$ s, the updated effective number of light neutrino species is reported as $N_{\text{eff}} = 3.68_{-0.70}^{+0.80}$ (2σ) and $N_{\text{eff}} = 3.80_{-0.70}^{+0.80}$ (2σ), respectively. The most recent estimate of Y_p yields $N_{\text{eff}} = 3.58 \pm 0.25(68\% \text{CL}), \pm 0.40(95.4\% \text{CL}), \pm (99\% \text{CL})$. This entails that a non-standard value of N_{eff} is preferred at the 99% CL, implying the possible existence of additional types of neutrino species [182].

EXERCISE 8.6 We have seen that the best multi-parameter fit of Planck data yields a Hubble constant which deviates by more than 2σ from the value obtained with the HST. The impact of the Planck h estimate is particularly important in the determination of N_{eff} . Combining observations of CMB data the Planck Collaboration reported $N_{\text{eff}} = 3.15 \pm 0.23$ [109]. However, if the value of h is not allowed to float in the fit, but instead is frozen to the value determined from the maser-cephied-supernovae distance ladder, the Planck CMB data then gives $N_{\text{eff}} = 3.62 \pm 0.25$, which suggests new neutrino-like physics (at around the 2.3σ level) [183]. The hints of extra relativistic degrees of freedom at BBN and CMB epochs can be explained, e.g., by means of the right-handed partners of the three, left-handed, SM neutrinos. In particular, milli-weak interactions of these Dirac states may allow the ν_R 's to decouple much earlier, at a higher temperature, than their left-handed counterparts [184]. Determine the minimum decoupling temperature of the right-handed neutrinos which is consistent with Planck data at the 1σ level.

E. Quantum black holes

As we have seen, black holes are the evolutionary endpoints of massive stars that undergo a supernova explosion leaving behind a fairly massive burned out stellar remnant. With no outward forces to oppose gravitational forces, the remnant will collapse in on itself.

The density to which the matter must be squeezed scales as the inverse square of the mass. For example, the Sun would have to be compressed to a radius of 3 km (about four millionths its present size) to become a black hole. For the Earth to meet the same fate, one would need to squeeze it into a radius of 9 mm, about a billionth its present size. Actually, the density of a solar mass black hole ($\sim 10^{19}$ kg/m³) is about the high-

est that can be created through gravitational collapse. A body lighter than the Sun resists collapse because it becomes stabilized by repulsive quantum forces between subatomic particles.

However, stellar collapse is not the only way to form black holes. The known laws of physics allow matter densities up to the so-called Planck value 10^{97} kg/m^3 , the density at which the force of gravity becomes so strong that quantum mechanical fluctuations can break down the fabric of spacetime, creating a black hole with a radius $\sim 10^{-35} \text{ m}$ and a mass of 10^{-8} kg . This is the lightest black hole that can be produced according to the conventional description of gravity. It is more massive but much smaller in size than a proton.

The high densities of the early universe were a prerequisite for the formation of primordial black holes but did not guarantee it. For a region to stop expanding and collapse to a black hole, it must have been denser than average, so the density fluctuations were also necessary. As we have seen, such fluctuations existed, at least on large scales, or else structures such as galaxies and clusters of galaxies would never have coalesced. For primordial black holes to form, these fluctuations must have been stronger on smaller scales than on large ones, which is possible though not inevitable. Even in the absence of fluctuations, holes might have formed spontaneously at various cosmological phase transitions – for example, when the universe ended its early period of accelerated expansion, known as inflation, or at the nuclear density epoch, when particles such as protons condensed out of the soup of their constituent quarks.

The realization that black holes could be so small prompted Stephen Hawking to consider quantum effects, and in 1974 his studies lead to the famous conclusion that black holes not only swallow particles but also spit them out [185, 186]. The strong gravitational fields around the black hole induce spontaneous creation of pairs near the event horizon. While the particle with positive energy can escape to infinity, the one with negative energy has to tunnel through the horizon into the black hole where there are particle states with negative energy with respect to infinity.⁸ As the black holes radiate, they lose mass and so will eventually evaporate completely and disappear. The evaporation is generally regarded as being thermal in character,⁹ with a temperature inversely proportional to its mass M_{BH} ,

$$T_{\text{BH}} = \frac{1}{8\pi GM_{\text{BH}}} = \frac{1}{4\pi r_s}, \quad (295)$$

and an entropy $S = 2\pi M_{\text{BH}} r_s$, where r_s is the

Schwarzschild radius and we have set $c = 1$. Note that for a solar mass black hole, the temperature is around 10^{-6} K , which is completely negligible in today's universe. But for black holes of 10^{12} kg the temperature is about 10^{12} K hot enough to emit both massless particles, such as γ -rays, and massive ones, such as electrons and positrons.

The black hole, however, produces an effective potential barrier in the neighborhood of the horizon that backscatters part of the outgoing radiation, modifying the blackbody spectrum. The black hole absorption cross section, σ_s (a.k.a. the greybody factor), depends upon the spin of the emitted particles s , their energy Q , and the mass of the black hole [191]. At high frequencies ($Qr_s \gg 1$) the greybody factor for each kind of particle must approach the geometrical optics limit. The integrated power emission is reasonably well approximated taking such a high energy limit. Thus, for illustrative simplicity, in what follows we adopt the geometric optics approximation, where the black hole acts as a perfect absorber of a slightly larger radius, with emitting area given by [191]

$$A = 27\pi r_s^2. \quad (296)$$

Within this framework, we can conveniently write the greybody factor as a dimensionless constant normalized to the black hole surface area seen by the SM fields $\Gamma_s = \sigma_s/A_4$, such that $\Gamma_{s=0} = 1$, $\Gamma_{s=1/2} \approx 2/3$, and $\Gamma_{s=1} \approx 1/4$.

All in all, a black hole emits particles with initial total energy between $(Q, Q + dQ)$ at a rate

$$\frac{d\dot{N}_i}{dQ} = \frac{\sigma_s}{8\pi^2} Q^2 \left[\exp\left(\frac{Q}{T_{\text{BH}}}\right) - (-1)^{2s} \right]^{-1} \quad (297)$$

per degree of particle freedom i . The change of variables $u = Q/T$, brings Eq. (297) into a more familiar form,

$$\dot{N}_i = \frac{27\Gamma_s T_{\text{BH}}}{128\pi^3} \int \frac{u^2}{e^u - (-1)^{2s}} du. \quad (298)$$

This expression can be easily integrated using (242) and (243), and yields

$$\dot{N}_i = \mathcal{A}_{\pm} \frac{27\Gamma_s}{128\pi^3} \Gamma(3) \zeta(3) T_{\text{BH}}. \quad (299)$$

Therefore, the black hole emission rate is found to be

$$\dot{N}_i \approx 7.8 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (300)$$

$$\dot{N}_i \approx 3.8 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (301)$$

$$\dot{N}_i \approx 1.9 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (302)$$

⁸ One can alternatively think of the emitted particles as coming from the singularity inside the black hole, tunneling out through the event horizon to infinity [187].

⁹ Indeed both the average number [185, 186] and the probability distribution of the number [188–190] of outgoing particles in each mode obey a thermal spectrum.

for particles with $s = 0, 1/2, 1$, respectively.

At any given time, the rate of decrease in the black hole mass is just the total power radiated

$$\frac{d\dot{M}_{\text{BH}}}{dQ} = - \sum_i g_i \frac{\sigma_s}{8\pi^2} \frac{Q^3}{e^{Q/T_{\text{BH}}} - (-1)^{2s}}, \quad (303)$$

where g_i is the number of internal degrees of freedom of particle species i . A straightforward calculation gives

$$\dot{M}_{\text{BH}} = - \sum_i g_i \mathcal{B}_\pm \frac{27\Gamma_s}{128\pi^3} \Gamma(4) \zeta(4) T_{\text{BH}}^2. \quad (304)$$

Assuming that the effective high energy theory contains approximately the same number of modes as the SM (i.e., $g_{s=1/2} = 90$, and $g_{s=1} = 27$), we find

$$\frac{dM_{\text{BH}}}{dt} = 8.3 \times 10^{73} \text{ GeV}^4 \frac{1}{M_{\text{BH}}^2}. \quad (305)$$

Ignoring thresholds, i.e., assuming that the mass of the black hole evolves according to (305) during the entire process of evaporation, we can obtain an estimate for the lifetime of the black hole,

$$\tau_{\text{BH}} = 1.2 \times 10^{-74} \text{ GeV}^{-4} \int M_{\text{BH}}^2 dM_{\text{BH}}. \quad (306)$$

Using $\hbar = 6.58 \times 10^{-25}$ GeV s, (306) can then be re-written as

$$\begin{aligned} \tau_{\text{BH}} &\simeq 2.6 \times 10^{-99} (M_{\text{BH}}/\text{GeV})^3 \text{ s} \\ &\simeq 1.6 \times 10^{-26} (M_{\text{BH}}/\text{kg})^3 \text{ yr}. \end{aligned} \quad (307)$$

This implies that for a solar mass black hole, the lifetime is unobservably long 10^{64} yr, but for a 10^{12} kg one, it is $\sim 1.5 \times 10^{10}$ yr, about the present age of the universe. Therefore, any primordial black hole of this mass would be completing its evaporation and exploding right now.

The questions raised by primordial black holes motivate an empirical search for them. Most of the mass of these black holes would go into gamma rays (quarks and gluons would hadronize mostly into pions which in turn would decay to γ -rays and neutrinos), with an energy spectrum that peaks around 100 MeV. In 1976, Hawking and Don Page realized that γ -ray background observations place strong upper limits on the number of such black holes [192]. Specifically, by looking at the observed γ -ray spectrum, they set an upper limit of $10^4/\text{pc}^3$ on the density of these black holes with masses near 5×10^{11} kg. Even if primordial black holes never actually formed, thinking about them has led to remarkable physical insights because they linked three previously disparate areas of physics: general relativity, quantum theory, and thermodynamics [193].

EXERCISE 8.7 Very recently, it has become evident that a promising route towards reconciling the apparent mismatch of the fundamental scales of particle

physics and gravity is to modify the short distance behavior of gravity at scales much larger than the Planck length. Such modification can be most simply achieved by introducing extra dimensions (generally thought to be curled-up) in the sub-millimeter range [194]. In the canonical example, spacetime is a direct product of ordinary 4-dimensional spacetime and a (flat) spatial n -torus with circumferences of length $2\pi r_i$ ($i = 1, \dots, n$), generally of common linear size $r_i = r_c$. The SM fields cannot propagate freely in the extra dimensions without conflict with observations. This is avoided by trapping the fields to a 3-dimensional *brane-world*. Applying Gauss' law at $r \ll r_c$ and $r \gg r_c$, it is easily seen that the effective Planck scale is related to the fundamental scale of gravity M_* simply by a volume factor,

$$\begin{aligned} r_c &= \left(\frac{M_{\text{Pl}}}{M_*} \right)^{2/n} \frac{1}{M_*} \\ &= 2.0 \times 10^{-17} \left(\frac{\text{TeV}}{M_*} \right) \left(\frac{M_{\text{Pl}}}{M_*} \right)^{2/n} \text{ cm}, \end{aligned} \quad (308)$$

so that M_* can range from $\sim \text{TeV}$ to 10^{19} GeV, for $r_c \leq 1 \text{ mm}$ and $n \geq 2$. If nature gracefully picked a sufficiently low-scale gravity, the first evidence for it would likely be the observation of microscopic black holes produced in particle collisions [195]. Although the black hole production cross section, $\mathcal{O}(M_W^{-1})$, is about 5 orders of magnitude smaller than QCD cross sections, $\mathcal{O}(\Lambda_{\text{QCD}}^{-1})$, it was proposed that such black holes could be produced copiously at the LHC [196, 197] and in cosmic ray collisions [198, 199], and that these spectacular events could be easily filtered out of the QCD background. To a first approximation it is reasonable to assume that the evaporation process is dominated by the large number of SM brane modes [200, 201]. Therefore, the emission rate per degree of particle freedom i of particles of spin s with initial total energy between $(Q, Q + dQ)$ can be approximated by (297). The characteristic temperature of a $4 + n$ -dimensional black hole is [202]

$$T_{\text{BH}} = \frac{n+1}{4\pi r_s}, \quad (309)$$

where

$$r_s = \frac{1}{M_*} \left[\frac{M_{\text{BH}}}{M_*} \frac{2^n \pi^{(n-3)/2} \Gamma(\frac{n+3}{2})}{n+2} \right]^{1/(1+n)}, \quad (310)$$

is the Schwarzschild radius [203]. As in the conventional 4-dimensional case, we can conveniently rewrite the greybody factor as a dimensionless constant, $\Gamma_s = \sigma_s / A_{4+n}$, normalized to the black hole surface area

$$A_{4+n} = 4\pi \left(\frac{n+3}{2} \right)^{2/(n+1)} \frac{n+3}{n+1} r_s^2 \quad (311)$$

seen by the SM fields [200]. The upper limit on the accretion rate for a $4 + n$ -dimensional black hole is

$$\frac{dM}{dt} \Big|_{\text{accr}} \approx \pi \left(\frac{n+3}{2} \right)^{2/(n+1)} \frac{n+3}{n+1} r_s^2 \epsilon, \quad (312)$$

where ϵ is the nearby quark-gluon (or parton) energy density [204]. The highest earthly value of energy density of partonic matter is the one created at the LHC, $\epsilon_{\text{LHC}} < 500 \text{ GeV/fm}^3$. Consider the case with $n = 6$, which is well motivated by string theory [205]. (i) Show that the black holes that could be produced at the LHC (or in any foreseeable accelerator built on Earth) would evaporate much too quickly to swallow the partons nearby. (ii) Determine the black hole lifetime. [Hint: For $n = 6$, you can evaluate the numerical results of [206] at $\langle Q \rangle$ and normalize the cross sections results to the capture area $A_{4 \subset 4+n}$ to obtain $\Gamma_{s=1/2} \approx 0.33$ and $\Gamma_{s=1} \approx 0.34$.]

IX. MULTI-MESSENGER ASTRONOMY

For biological reasons our perception of the Universe is based on the observation of photons, most trivially by staring at the night-sky with our bare eyes. Conventional astronomy covers many orders of magnitude in photon wavelengths, from 10^4 cm radio-waves to 10^{-14} cm gamma rays of GeV energy. This 60 octave span in photon frequency allows for a dramatic expansion of our observational capacity beyond the approximately one octave perceivable by the human eye.

The γ -ray sky has been monitored since 1968. The pioneering observations by the third Orbiting Solar Observatory (OSO-3) provided the first γ -ray sky map, with 621 events detected above 50 MeV [207]. In addition, these observations revealed the existence of an isotropic emission. The presence of an isotropic diffuse γ -ray background (IGRB) has been confirmed by the Small Astronomy Satellite 2 (SAS-2) [208] and the Energetic Gamma Ray Experiment Telescope (EGRET) on board of the Compton Gamma Ray Observatory (CGRO) [209, 210]. Very recently, the Fermi-LAT has released a new measurement of the IGRB spectrum from 100 MeV to 820 GeV at Galactic latitude $|b| > 20^\circ$ [211]; see Fig. 33. The LAT has also measured the extragalactic γ -ray background (EGB), which is the sum of the IGRB and the flux from detected sources. For the first time a deviation from a power-law shape in the high-energy part of the EGB and IGRB has been observed as an exponential cut off with a break energy of about $E_\gamma = 280 \text{ GeV}$. The origin of the IGRB is not yet fully understood. This leaves intriguing puzzles for the next generation of GeV γ ray instruments to uncover [214]. What happens at higher energies?

Above a few 100 GeV the universe becomes opaque to the propagation of γ rays, because of e^+e^- production on the radiation fields permeating the universe; see Fig. 34. The pairs synchrotron radiate on the extragalactic magnetic field before annihilation and so the photon flux is significantly depleted. Moreover, the charged particles also suffer deflections on the \vec{B} -field camouflaging the exact location of the sources. In other words, the injection photon spectrum is significantly modified *en route* to Earth. This modification becomes dramatic at around

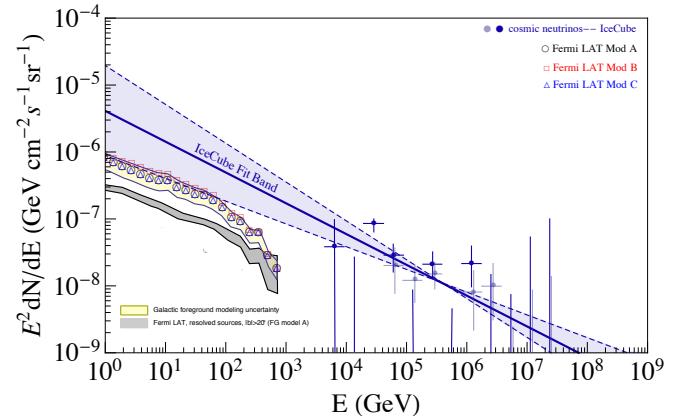


FIG. 33: The open symbols represent the total extragalactic γ -ray background for different foreground (FG) models as reported by the Fermi Collaboration [211]. For details on the modeling of the diffuse Galactic foreground emission in the benchmark FG models A, B and C, see [211]. The cumulative intensity from resolved Fermi-LAT sources at latitudes $|b| > 20^\circ$ is indicated by a (grey) band. The solid symbols indicate the neutrino flux reported by the IceCube Collaboration [212]. The best fit to the data (extrapolated down to lower energies), $\Phi(E_\nu) = 2.06^{+0.4}_{-0.3} \times 10^{-18} (E_\nu/10^5 \text{ GeV})^{-2.46 \pm 0.12} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$, is also shown for comparison [213].

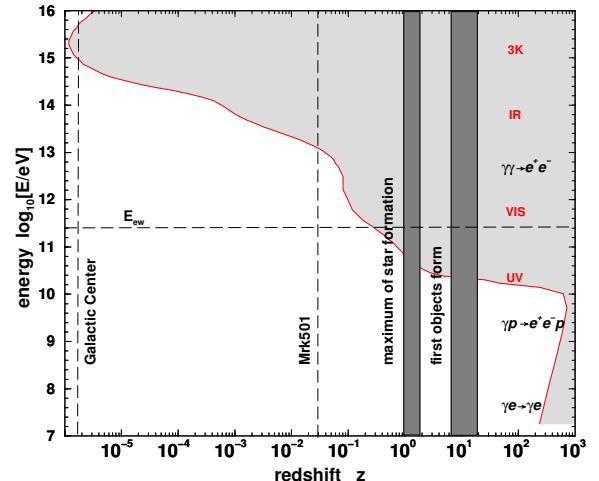


FIG. 34: Mean interaction length for photons on the ultraviolet (UV), visible (VIS), infrared (IR), and microwave (3K) backgrounds. The electroweak scale is indicated by a dashed line. The redshifts of the star formation epoch and the famous γ -ray source Markarian 501 are also indicated [215].

10^6 GeV where interaction with the CMB dominates and the photon mean free path is smaller than the Galactic radius.

Therefore, to study the high energy behavior of distance sources we need new messengers. Nowadays the best candidates to probe the high energy universe are cosmic rays, neutrinos, and gravitational waves. Of course

in doing multi-messenger astronomy one has to face new challenges. It is this that we now turn to study.

A. Cosmic rays

In 1912 Hess carried out a series of pioneering balloon flights during which he measured the levels of ionizing radiation as high as 5 km above the Earth's surface [216]. His discovery of increased radiation at high altitude revealed that we are bombarded by ionizing particles from above. These cosmic ray particles are now known to consist primarily of protons, helium, carbon, nitrogen and other heavy ions up to iron.

Below 10^5 GeV the flux of particles is sufficiently large that individual nuclei can be studied by detectors carried aloft in balloons or satellites. From such direct experiments we know the relative abundances and the energy spectra of a variety of atomic nuclei, protons, electrons and positrons as well as the intensity, energy and spatial distribution of X -rays and γ -rays. Measurements of energy and isotropy showed conclusively that one obvious source, the Sun, is not the main source. Only below 100 MeV kinetic energy or so, where the solar wind shields protons coming from outside the solar system, does the Sun dominate the observed proton flux. Spacecraft missions far out into the solar system, well away from the confusing effects of the Earth's atmosphere and magnetosphere, confirm that the abundances around 1 GeV are strikingly similar to those found in the ordinary material of the solar system. Exceptions are the overabundance of elements like lithium, beryllium, and boron, originating from the spallation of heavier nuclei in the interstellar medium.

EXERCISE 9.1 Consider a simple model of cosmic rays in the Galaxy (height $H \ll$ radius R) in which the net diffusion of cosmic rays is mainly perpendicular to the Galactic disk. In this case the density of cosmic rays depends only on the vertical coordinate z and follows the diffusion equation

$$\frac{\partial n}{\partial t} = D \frac{\partial^2 n}{\partial z^2} + Q(z, t), \quad (313)$$

where $D = \beta c \lambda / 3$ is the diffusion coefficient, λ is the mean free path, and the source term is given by $Q(z, t)$. Use the approximation $Q(z, t) = Q_0 \delta(z)$ to describe a time-independent concentration of stars close to $z = 0$, $\delta(z)$ is the Dirac delta function (see Appendix E). (i) Find the steady-state solution to the diffusion equation given a vanishing cosmic ray density at the edges of the Galaxy, $n(z = +H) = n(z = -H) = 0$. (ii) Calculate the cosmic-ray column density

$$N = \int_{-H}^{+H} n(z) dz \quad (314)$$

and determine the average residence time τ_{res} from $N = Q_0 \tau_{\text{res}}$. What is the mean free path for $H = 500$ pc

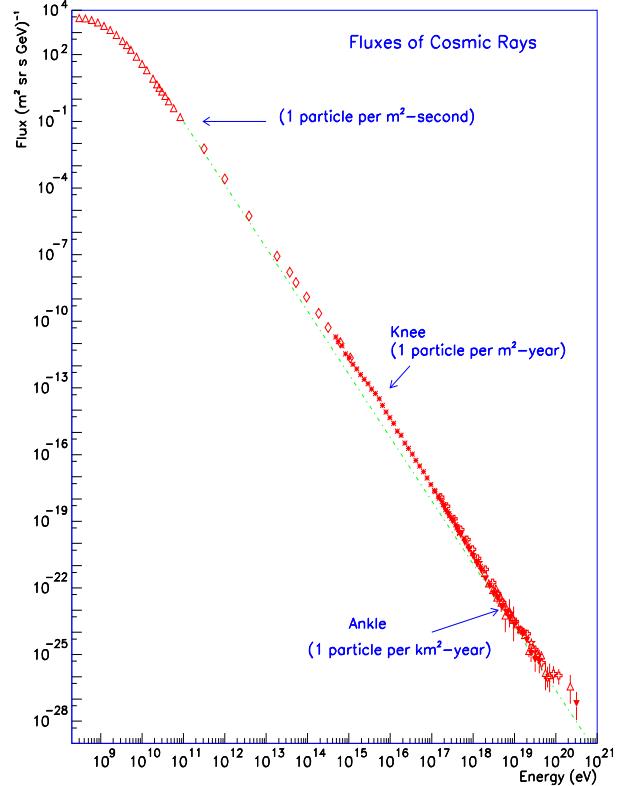


FIG. 35: Compilation of measurements of the differential energy spectrum of cosmic rays. The dotted line shows an E^{-3} power-law for comparison. Approximate integral fluxes (per steradian) are also shown.

and $\tau_{\text{res}} = 10^7$ yr?

Above 10^5 GeV, the flux becomes so low that only ground-based experiments with large apertures and long exposure times can hope to acquire a significant number of events. Such experiments exploit the atmosphere as a giant calorimeter. The incident cosmic radiation interacts with the atomic nuclei of air molecules and produces extensive air showers which spread out over large areas. Already in 1938, Auger concluded from the size of extensive air showers that the spectrum extends up to and perhaps beyond 10^6 GeV [217, 218]. Nowadays substantial progress has been made in measuring the extraordinarily low flux (~ 1 event $\text{km}^{-2} \text{ yr}^{-1}$) above 10^{10} GeV. Continuously running experiments using both arrays of particle detectors on the ground and/or fluorescence detectors which track the cascade through the atmosphere, have detected events with primary particle energies somewhat above 10^{11} GeV [219].

The Pierre Auger Observatory employs the two detection methods [220]. It consists of an array of about 1,600 water Cherenkov surface detectors (SD) deployed over a triangular grid of 1.5 km spacing and covering an area of 3,000 km^2 [221]. A SD event is formed when

at least 3 non-aligned stations selected by the local station trigger are in spatial and temporal coincidence. The ground array is overlooked by 24 fluorescence telescopes, grouped in four sites, making up the fluorescence detector (FD) [222]. The FD observes the longitudinal development of the shower in the atmosphere by detecting the fluorescence light emitted by excited nitrogen molecules and Cherenkov light induced by shower particles in air. The two detection methods have different strengths, and together allow for large statistics data samples and unrivaled control over systematic uncertainties.

The FD provides a calorimetric measurement of the primary particle energy, only weakly dependent on theoretical models. The most common strategy to determine the nature of the primary cosmic ray is to study the longitudinal shower profile of the electromagnetic component in the atmosphere. The slant depth is the amount of atmosphere penetrated by a cosmic ray shower at a given point in its development, and is customarily denoted by the symbol X . The value of X is calculated by integrating the density of air from the point of entry of the air shower at the top of the atmosphere, along the trajectory of the shower, to the point in question. The depth of the shower maximum X_{\max} is the position of the maximum of energy deposition per atmospheric slant depth of an extensive air shower. Lighter primaries penetrate the atmosphere deeper than heavier primaries. In addition, due to the larger number of nucleons and the larger cross section, the event-by-event fluctuations of X_{\max} should be smaller for heavier nuclei. Therefore, the first two moments of the X_{\max} distribution, which are the mean $\langle X_{\max} \rangle$ and standard deviation $\sigma(X_{\max})$ provide good discriminators between different primary cosmic rays; for details see e.g. [223].

The mechanism(s) responsible for imparting an energy of more than one Joule to a single elementary particle continues to present a major enigma to high energy physics [224]. It is reasonable to assume that, in order to accelerate a proton to energy E in a magnetic field B , the size R of the accelerator must encompass the gyro radius of the particle: $R > R_{\text{gyro}} \sim E/B$, i.e. the accelerating magnetic field must contain the particle's orbit. By dimensional analysis, this condition yields a maximum energy $E \sim \gamma BR$. The γ -factor has been included to allow for the possibility that we may not be at rest in the frame of the cosmic accelerator, resulting in the observation of boosted particle energies. Opportunity for particle acceleration to the highest energies is limited to dense regions where exceptional gravitational forces create relativistic particle flows. All speculations involve collapsed objects and we can therefore replace R by the Schwarzschild radius $R \sim GM/c^2$ to obtain $E < \gamma BM$.

At this point a reality check is in order. Such a dimensional analysis applies to the Fermilab accelerator: 10 kilogauss fields over several kilometers (covered with a repetition rate of 10^5 revolutions per second) yield 1 TeV. The argument holds because, with optimized design and perfect alignment of magnets, the accelerator reaches ef-

ficiencies matching the dimensional limit. It is highly questionable that nature can achieve this feat.

Given the microgauss magnetic field of our galaxy, no structures are large or massive enough to reach the energies of the highest energy cosmic rays. Dimensional analysis therefore limits their sources to extragalactic objects. A common speculation is that there may be relatively nearby active galactic nuclei powered by a billion solar mass black holes. With kilo-Gauss fields we reach 10^{11} GeV. The jets (blazars) emitted by the central black hole could reach similar energies in accelerating sub-structures boosted in our direction by a γ -factor of 10, possibly higher.

EXERCISE 9.2 (i) Derive the magnetic field strength needed to hold a charge on a circular orbit of radius R given its momentum p . Assume that the magnetic field is uniform, that the motion of the particle is perpendicular to the magnetic field and let $\beta \sim 1$. (ii) Given the circumference (~ 26.659 km) of the LHC, determine the uniform magnetic field strength needed to keep 7 TeV protons in orbit. (iii) Using this magnetic field strength find what would be the required size needed for an LHC-like accelerator to launch particles to cosmic-ray energies $\sim 10^{11}$ GeV. Compare this with the orbits in the solar system and estimate the cost of the accelerator. (iv) Which of the following astrophysical objects are able to keep ultrahigh energy cosmic rays in orbit? Neutron stars ($R \sim 10^{-13}$ pc, $B \sim 10^{12}$ G), AGN jets ($R \sim 1$ kpc, $B \sim 10^{-5}$ G), supernova remnants ($R = 1$ pc $B \sim 10^{-4}$ G). Consider protons and iron nuclei of energy 10^{11} GeV and that you are at rest in the frame of the cosmic accelerator.

The almost structureless power law spectrum spans many decades of energy, 10^1 GeV $< E < 10^{11}$ GeV. A close examination of Fig. 35 reveals three major features: (i) the steepening of the spectrum dubbed the *knee* centered at $10^{6.6}$ GeV [225]; (ii) a pronounced hardening of the spectrum at about $10^{9.6}$ GeV, the so-called *ankle* feature [226]; (iii) a cutoff around $10^{10.6}$ GeV [227, 228]. Three additional more subtle features have been recently spotted between the knee and the ankle: a hardening of the spectrum at around $10^{7.3}$ GeV [229, 230] followed by two softenings at $\approx 10^{7.9}$ GeV [229, 230] and $\approx 10^{8.5}$ GeV [231, 232]. The latter is traditionally referred to as the *second knee*.

The variations of the spectral index reflect various aspects of cosmic ray production, source distribution, and propagation. The first and second knee have unequivocal explanations, as reflecting the maximum energy of Galactic magnetic confinement or acceleration capability of the sources, both of which grow linearly in the charge Z of the nucleus; the first knee being where protons drop out and the second knee where the highest- Z Galactic cosmic rays drop out. As the energy increases above the second knee to the ankle, the nuclear composition switches from heavy to light [233] whereas the cosmic ray

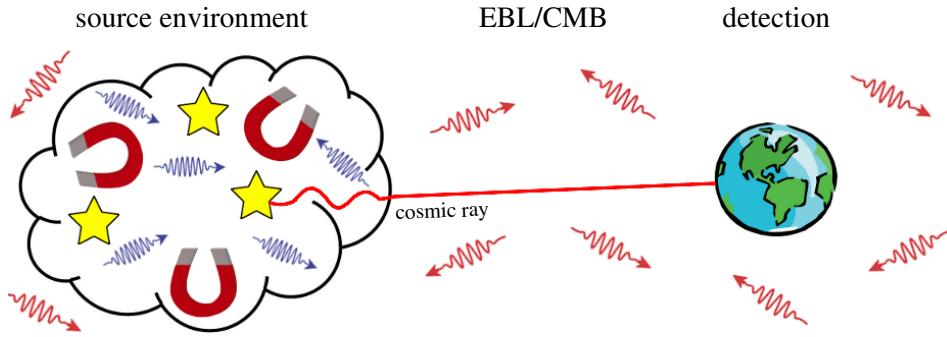


FIG. 36: Sources (yellow stars) inject cosmic ray nuclei with a power law in energy into a surrounding region of radiation and turbulent magnetic fields. After propagation through this local environment and intergalactic space, these cosmic rays and their spallation products are detected at Earth. The photon energies in the source environment are characteristically of much higher energy than in the extragalactic background light (EBL) [243].

arrival directions are isotropic to high accuracy throughout the entire range [234–236]. Lastly, as the energy increases above the ankle, not only does the spectrum harden significantly, but the composition gradually becomes heavier (interpreting the data using conventional extrapolations of accelerator-constrained particle physics models) [237, 238].

The observed evolution in the extragalactic cosmic ray composition and spectral index presents a complex puzzle. A pure proton composition might be compatible with the observed spectrum of extragalactic cosmic rays [239] when allowance is made for experimental uncertainties in the energy scale and the fact that the real local source distribution is not homogeneous and continuous [240] (although the sharpness of the ankle is difficult to accommodate). However, a pure proton composition is incompatible with the X_{\max} and $\sigma(X_{\max})$ distributions reported by the Auger Collaboration [237, 238] unless current extrapolations of particle physics are incorrect. On the other hand, models which fit the spectrum and composition at highest energies, predict a deep gap between the end of the Galactic cosmic rays and the onset of the extragalactic cosmic rays. Models can be devised to fill this gap: fine-tuning is required to position this new population so as to just fit and fill the gap [241, 242], unless we consider interactions in the region surrounding the accelerator as illustrated in Fig. 36.

The discovery of a suppression above $10^{10.6}$ GeV was first reported by the HiRes and Auger collaborations [227, 228] and later confirmed by the Telescope Array Collaboration [244]; by now the significance is well in excess of 20σ compared to a continuous power law extrapolation beyond the ankle feature [245]. This suppression is consistent with the Greisen-Zatsepin-Kuzmin (GZK) prediction that interactions with cosmic background photons will rapidly degrade cosmic ray energies [246, 247]. Intriguingly, however, there are also indications that the source of the suppression may be more complex than originally anticipated. The trend toward heavier composition above the ankle could reflect the endpoint of cosmic accel-

eration, with heavier nuclei dominating the composition near the end of the spectrum, which coincidentally falls off near the expected GZK cutoff region [248]. If this were the case, the suppression would constitute an imprint of the accelerator characteristics rather than energy loss in transit. It is also possible that a mixed or heavy composition is emitted from the sources, and photodisintegration of nuclei and other GZK energy losses suppress the flux.

The main reason why this impressive set of data fails to reveal the origin of the particles is undoubtedly that their directions have been scrambled by the microgauss Galactic magnetic fields. However, above 10^{10} GeV proton astronomy could still be possible because the arrival directions of electrically charged cosmic rays are no longer scrambled by the ambient magnetic field of our own Galaxy. Protons point back to their sources with an accuracy determined by their gyroradius in the intergalactic magnetic field B ,

$$\theta \simeq \frac{d}{R_{\text{gyro}}} = \frac{dB}{E}, \quad (315)$$

where d is the distance to the source. Scaled to units relevant to the problem,

$$\frac{\theta}{0.1^\circ} \simeq \frac{(d/\text{Mpc})(B/\text{nG})}{E/10^{11.5} \text{ GeV}}. \quad (316)$$

Speculations on the strength for the inter-galactic magnetic field range from 10^{-7} to 10^{-9} G. For the distance to a nearby galaxy at 100 Mpc, the resolution may therefore be anywhere from sub-degree to nonexistent. Moreover, neutrons with energy $\gtrsim 10^9$ GeV have a boosted $c\tau_n$ sufficiently large to serve as Galactic messengers [249].¹⁰ The decay mean free path of a neutron is $c\gamma_n \bar{\tau}_n = 9.15(E_n/10^9 \text{ GeV})$ kpc, the lifetime being

¹⁰ Neutron astronomy from the nearby radio galaxy Centaurus A may also be possible [250].

boosted from its rest-frame value, $\bar{\tau}_n = 886$ s, to its lab value by $\gamma_n = E_n/m_n$. It is therefore reasonable to expect that the arrival directions of the very highest energy cosmic rays may provide information on the location of their sources.¹¹

B. Cosmic neutrinos

For a deep, sharply focused examination of the universe a telescope is needed which can observe a particle that is not much affected by the gas, dust, and swirling magnetic fields it passes on its journey. The neutrino is the best candidate. As we have seen, neutrinos constitute much of the total number of elementary particles in the universe, and these neutral, weakly-interacting particles come to us almost without any disruption straight from their sources, traveling at very close to the speed of light. A (low energy) neutrino in flight would not notice a barrier of lead fifty light years thick. When we are able to see outwards in neutrino light we will no doubt receive a wondrous new view of the universe.

EXERCISE 9.3 In 1987, the astronomical world was electrified with the news of a supernova exploding in the Large Magellanic Cloud, a dwarf galaxy companion to the Milky Way, at a distance of 150,000 ly. It was the nearest supernova to have gone off in 400 yr, and was studied in great detail. Its luminosity was enormous; the explosion released as much visible light energy in a few weeks as the Sun will emit in its entire lifetime of 10^{10} yr. It was easily visible to the naked eye from the Southern hemisphere. However, models of the mechanisms taking place in the supernovae predict that the visible light represents only 1% of the total energy of the supernova; there is 100 times more energy emitted in the form of neutrinos, in a blast lasting only a few seconds. (i) Calculate the total amount of energy emitted by the supernova in neutrinos. Express your answer in Joules. (ii) Each neutrino has an energy of roughly $\langle E_\nu \rangle \sim 1.5 \times 10^{-12}$ J. Calculate how many neutrinos are emitted by the supernova. (This is an easy calculation, but will give you a very large number). (iii) Kamiokande is one of the largest neutrino detectors. In 1987, it consisted of 2.140 kton of water (it has since been expanded). Calculate how many electron neutrinos should have been detected by Kamiokande if the detection efficiency at $\langle E_{\nu_e} \rangle$ is about 60% [252].

We have seen that MeV neutrinos are produced by nuclear reaction chains in the central core of stars. Moving up in energy, neutrinos would also be inevitably produced in many of the most luminous and energetic objects in the universe. Whatever the source, the ma-

chinery which accelerates cosmic rays will inevitably also produce neutrinos, guaranteeing that high energy neutrinos surely arrive to us from the cosmos.

Neutrino detectors must be generally placed deep underground, or in water, in order to escape the backgrounds caused by the inescapable rain of cosmic rays upon the atmosphere. These cosmic rays produce many muons which penetrate deeply into the earth, in even the deepest mines, but of course with ever-decreasing numbers with depth. Hence the first attempts at high energy neutrino astronomy have been initiated underwater and under ice [253].

The IceCube facility is located near the Amundsen-Scott station below the surface of the Antarctic ice sheet at the geographic South Pole [254]. The main part of the detector is the *InIce* array, which covers a cubic kilometer of Antarctic glacial ice instrumented with digital optical modules (DOMs) that detect Cherenkov light [255]. The DOMs are attached to km-long supply and read-out cables called strings. Each string carries 60 DOMs spaced evenly along 1 km. The full baseline design of 86 strings was completed in December 2010. In addition to the *InIce* array, IceCube also possesses an air shower array called *IceTop* which comprises 80 stations, each of which consists of two tanks of water-ice instrumented with 2 DOMs to detect Cherenkov light [256]. The hybrid observations of air showers in the *InIce* and *IceTop* arrays have mutual benefits, namely significant air shower background rejection (for neutrino studies) and an improved air shower muon detection (for cosmic ray studies).

In 2012, the IceCube Collaboration famously announced an observation of two ~ 1 PeV neutrinos discovered in a search for the nearly guaranteed cosmogenic neutrinos (which are expected to be produced as secondaries in the GZK chain reaction) [257]. The search technique was later refined to extend the neutrino sensitivity to lower energies [258, 259], resulting in the discovery of an additional 26 neutrino candidates with energies between 50 TeV and 2 PeV, constituting a 4.1σ excess for the combined 28 events compared to expectations from neutrino and muon backgrounds generated in Earth's atmosphere [260]. Interpretation of these results, however, does not appear to be entirely straightforward. For instance, if one makes the common assumption of an unbroken E_ν^{-2} neutrino energy spectrum, then one expects to observe about 8-9 events with higher energies than the two highest energy events observed thus far. The compatibility between IceCube observations and the hypothesis of an unbroken power-law spectrum requires a rather steep spectrum, $\Phi(E_\nu) \propto E^{-2.3}$ [261]. Very recently, the IceCube results have been updated [262–264]. At the time of writing, 54 events have been reported in four years of IceCube data taking (1347 days between 2010 – 2014). The data are consistent with expectations for equal fluxes of all three neutrino flavors and with isotropic arrival directions [265]. The best-fit power law is $E_\nu^2 \Phi(E_\nu) = 2.2 \pm 0.7 \times 10^{-8} (E_\nu/100 \text{ TeV})^{-0.58} \text{ GeV cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ and

¹¹ For a more extensive discussion of this subject see e.g. [251].

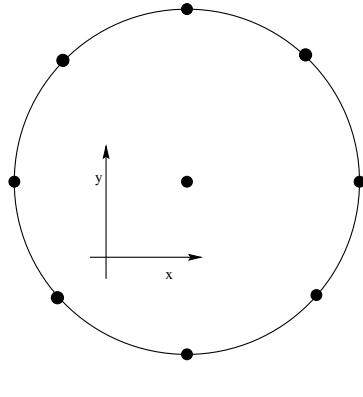


FIG. 37: Initial configuration of test particles on a circle of radius L before a gravitational wave hits them.

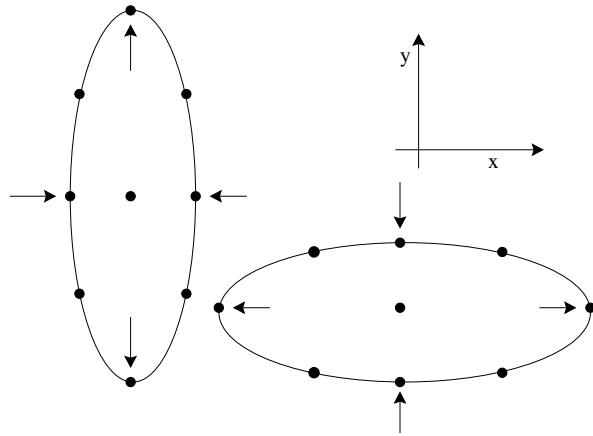


FIG. 38: The effect of a plus-polarized gravitational wave on a ring of particles. The amplitude shown in the figure is roughly $h = 0.5$. Gravitational waves passing through the Earth are many billion billion times weaker than this.

rejects a purely atmospheric explanation at 5.7σ . There are numerous proposed explanations for the origin of Ice-Cubes events [266]. However, considerably more data are yet required before the final verdict can be given.

C. Gravitational waves

Ever since Newton in the XVII century, we have learned that gravity is a force that acts immediately on an object. In Einstein theory of general relativity, however, gravity is not a *force* at all, but a curvature in space [50]. In other words, the presence of a very massive body does not affect probed objects directly; it warps the space around it first and then the objects move in the curved space. Inherit from such a redefinition of gravity is the concept of gravitational waves: as massive bodies move around, disturbances in the curvature of spacetime can spread outward, much like a pebble tossed into a pond will cause waves to ripple outward from the source. Prop-

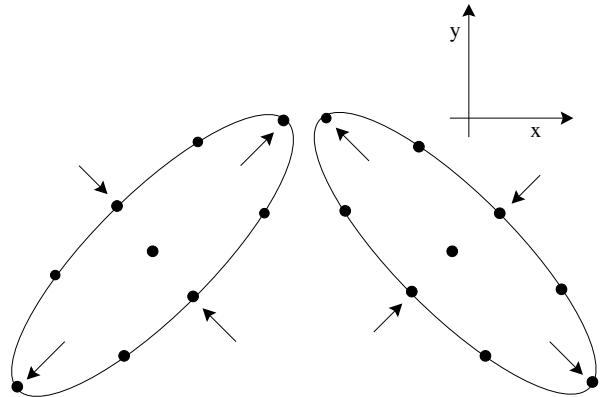


FIG. 39: The effect of cross-polarized gravitational waves on a ring of particles.

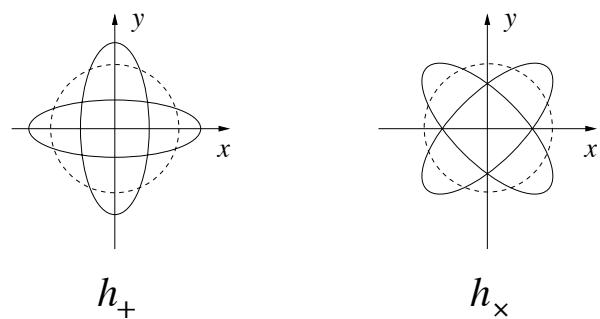


FIG. 40: Two linearly independent polarizations of a gravitational wave are illustrated by displaying their effect on a ring of free particles arrayed in a plane perpendicular to the direction of the wave. The figure shows the distortions in the original circle that the wave produces if it carries the plus-polarization or the cross-polarization. In general relativity there are only 2 independent polarizations. The ones shown here are orthogonal to each other and the polarizations are transverse to the direction of the wave.

agating at (or near) the speed of light, these disturbances do not travel *through* spacetime as such – the fabric of spacetime itself is oscillating!

The simplest example of a strong source of gravitational waves is a spinning neutron star with a small mountain on its surface. The mountain's mass will cause curvature of the spacetime. Its movement will stir up spacetime, much like a paddle stirring up water. The waves will spread out through the universe at the speed of light, never stopping or slowing down.

As these waves pass a distant observer, that observer will find spacetime distorted in a very particular way: distances between objects will increase and decrease rhythmically as the wave passes. To visualize this effect, consider a perfectly flat region of spacetime with a group of motionless test particles lying in a plane, as shown in Fig. 37. When a weak gravitational wave arrives, passing through the particles along a line perpendicular to the ring of radius L , the test particles will oscillate in a

cruciform manner, as indicated in Figs. 38 and 39. The area enclosed by the test particles does not change, and there is no motion along the direction of propagation. The principal axes of the ellipse become $L + \Delta L$ and $L - \Delta L$. The amplitude of the wave, which measures the fraction of stretching or squeezing, is $h = \Delta L/L$. Of course the size of this effect will go down the farther the observer is from the source. Namely, $h \propto d^{-1}$, where d is the source distance. Any gravitational waves expected to be seen on Earth will be quite small, $h \sim 10^{-20}$.

The frequency, wavelength, and speed of a gravitational wave are related through $\lambda = c\nu$. The polarization of a gravitational wave is just like polarization of a light wave, except that the polarizations of a gravitational wave are at 45° , as opposed to 90° . In other words, the effect of a *cross*-polarized gravitational wave (h_x) on test particles would be basically the same as a wave with

plus-polarization (h_+), but rotated by 45° . The different polarizations are summarized in Fig. 40.

In general terms, gravitational waves are radiated by very massive objects whose motion involves acceleration, provided that the motion is not perfectly spherically symmetric (like a spinning, expanding or contracting sphere) or cylindrically symmetric (like a spinning disk). For example, two objects orbiting each other in a quasi-Keplerian planar orbit will radiate [267, 268]. The power given off by a binary system of masses M_1 and M_2 separated a distance R is [269]

$$P = -\frac{32}{\pi} \frac{G^4}{c^5} \frac{(M_1 M_2)^2 (M_1 + M_2)}{R^5}. \quad (317)$$

For the Earth-Sun system R is very large and M_1 and M_2 are relatively very small, yielding

$$P = -\frac{32}{\pi} \frac{(6.7 \times 10^{-11} \frac{\text{m}^3}{\text{kg s}^2})^4}{(3 \times 10^8 \text{ m/s})^5} \frac{(6 \times 10^{24} \text{ kg} 2 \times 10^{30} \text{ kg})^2 (6 \times 10^{24} \text{ kg} + 2 \times 10^{30} \text{ kg})}{(1.5 \times 10^{11} \text{ m})^5} = 313 \text{ W}. \quad (318)$$

Thus, the total power radiated by the Earth-Sun system in the form of gravitational waves is truly tiny compared to the total electromagnetic radiation given off by the Sun, which is about 3.86×10^{26} W. The energy of the gravitational waves comes out of the kinetic energy of the Earth's orbit. This slow radiation from the Earth-Sun system could, in principle, steal enough energy to drop the Earth into the Sun. Note however that the kinetic energy of the Earth orbiting the Sun is about 2.7×10^{33} J. As the gravitational radiation is given off, it takes about 300 J/s away from the orbit. At this rate, it would take many billion times more than the current age of the universe for the Earth to fall into the Sun.

Although the power radiated by the Earth-Sun system is minuscule, we can point to other sources for which the radiation should be substantial. One important example is the pair of stars (one of which is a pulsar) discovered by Hulse and Taylor [270]. The characteristics of the orbit of this binary system can be deduced from the Doppler shifting of radio signals given off by the pulsar. Each of the stars has a mass about $1.4 M_\odot$. Also, their orbit is about 75 times smaller than the distance between the Earth and Sun, which means the distance between the two stars is just a few times larger than the diameter of our own Sun. This combination of greater masses and smaller separation means that the energy given off by the Hulse-Taylor binary will be far greater than the energy given off by the Earth-Sun system, roughly 10^{22} times as much.

The information about the orbit can be used to predict just how much energy (and angular momentum) should be given off in the form of gravitational waves. As the

energy is carried off, the orbit will change; the stars will draw closer to each other. This effect of drawing closer is called an inspiral, and it can be observed in the pulsar's signals. The measurements on this system were carried out over several decades, and it was shown that the changes predicted by gravitational radiation in general relativity matched the observations very well, providing the first experimental evidence for gravitational waves.

Inspirals are very important sources of gravitational waves. Any time two compact objects (white dwarfs, neutron stars, or black holes) come close to each other, they send out intense gravitational waves. As the objects come closer and closer to each other (that is, as R becomes smaller and smaller), the gravitational waves become more and more intense. At some point these waves should become so intense that they can be directly detected by their effect on objects on the Earth. This direct detection is the goal of several large experiments around the world.

The great challenge of this type of detection, though, is the extraordinarily small effect the waves would produce on a detector. The amplitude of any wave will fall off as the inverse of the distance from the source. Thus, even waves from extreme systems like merging binary black holes die out to very small amplitude by the time they reach the Earth. For example, the amplitude of waves given off by the Hulse-Taylor binary as seen on Earth would be roughly $h \approx 10^{-26}$. However, some gravitational waves passing the Earth could have somewhat larger amplitudes, $h \approx 10^{-20}$ [267, 268]. For an object 1 m in length, this means that its ends would move by 10^{-20} m relative to each other. This distance is about a

billionth of the width of a typical atom.

A simple device to detect this motion is the laser interferometer, with separate masses placed many hundreds of meters to several kilometers apart acting as two ends of a bar. Ground-based interferometers are now operating, and taking data. The most sensitive is the Laser Interferometer Gravitational Wave Observatory (LIGO) [271]. This is actually a set of three devices: one in Livingston, Louisiana; the other two (essentially on top of each other) in Hanford, Washington. Each consists of two light storage arms which are 2 to 4 km in length. These are at 90° angles to each other, and consist of large vacuum tubes running the entire 4 kilometers. A passing gravitational wave will then slightly stretch one arm as it shortens the other. This is precisely the motion to which an interferometer is most sensitive.

On September 14, 2015 at 09:50:45 UTC gravitational waves were detected by both of the twin LIGO detectors [272]. The waves originated in the collision and merger of two black holes (with 29 and 36 M_{\odot}) approximately 400 Mpc from Earth. About 3 times the mass of the sun was converted into gravitational waves in a fraction of a second, with a peak power output about 50 times that of the whole visible universe. This detection inaugurates a new era of astronomy in which gravitational waves are tools for studying the most mysterious and exotic objects in the universe.

EXERCISE 9.4 (i) Estimate the power radiated in gravitational waves by a neutron star of $M_{\star} = 1.4M_{\odot}$ orbiting a black hole of $M_{\text{BH}} = 20M_{\odot}$, assuming the orbital radius is $R = 6GM_{\text{BH}}/c^2$. (ii) If the kinetic energy of the neutron star orbiting the black hole is about 7×10^{47} J, how much time will it take the neutron star to fall into the black hole?

D. Looking ahead

The recent observation of a diffuse astrophysical flux of high energy neutrinos and the direct detection of gravitational waves represents the *first light* in the nascent field of multimessenger astronomy. The search for correlations in the different data sample has already started [273–276]. Thus far, there are no excesses beyond randomly expected.

An in-depth exploration of the neutrino universe requires a next-generation IceCube detector. IceCube-

Gen2 is based upon the robust design of the current detector [277]. The goal for this new observatory is to deliver statistically significant samples of very high energy astrophysical neutrinos, in the $10^6 \text{ GeV} \lesssim E_{\nu} \lesssim 10^9 \text{ GeV}$ range, and yield hundreds of neutrinos across all flavors at energies above 100 TeV. This will enable detailed spectral studies, significant point source detections, and new discoveries. Companion experiments in the deep Mediterranean are moving into construction phase.

Resolving the fundamental questions of UHECR composition and origins, and investigating particle physics above accelerator energies, will require both enhanced experimental techniques implemented at the existing observatories, as well as a significant increase in exposure to catch the exceedingly rare highest energy events. In the very near future the upgrade of the Pierre Auger Observatory, named *Auger Prime*, will allow: (i) a precise reconstruction of mass dependent energy spectrum; (ii) the identification of primaries, event-by-event, up to the highest energies; (iii) a systematic study of arrival direction(s) of an enhanced proton data sample [278].

Even before we know the results from Auger Prime it seems clear that still larger aperture observatories with much better energy and X_{max} resolution will be called for in order to measure the spectra and composition distribution of individual sources. It is inspiring to note that some 5 million UHECRs above about 5.5×10^{10} GeV strike the Earth's atmosphere each year, from which we currently collect only about 50 or so with present observatories. In this sense, there exists some 5 orders of magnitude room for improvement! It may well be that the best hope to make inroads in this area is to take the search for UHECR sources into space from which a huge volume of atmosphere can be viewed using the fluorescence technique. To this end several path finder efforts are underway to develop the requisite technologies. For example, in 2017 a NASA/CNES supported mission to fly a *super-pressure* stratospheric balloon with a fluorescence detector will take place. Such balloons can fly for hundreds of days, and may observe the first air showers from above. Eventually these technologies may lead to a permanently orbiting satellite to detect UHECRs. An optimist might even imagine an eventual constellation of satellites to tap the remaining 5 orders of magnitude of UHECR “luminosity”, accessing naturally occurring particle beams at energies far in excess available to terrestrial colliders with an event rate opening up a new window on beyond-the-standard-model phenomena.

-
- [1] G. Galilei, *Sidereus Nuncius*, (T. Baglioni, Republic of Venice, 1610).
 - [2] N. Copernicus, *De revolutionibus orbium coelestium*, (1543). Reprinted on *On the Shoulders of Giants: The Great Works of Physics and Astronomy*, (Ed. S. Hawking, Running Press, Philadelphia, 2002) ISBN 0-7624-1348-4; p.7.
 - [3] T. Wright, *An Original Theory of New Hypothesis of the Universe*, (H. Chapelle, London, 1750).
 - [4] I. Newton, *PhilosophiNaturalis Principia Mathematica*, (1687). Reprinted on *On the Shoulders of Giants: The Great Works of Physics and Astronomy*, (Ed. S. Hawking, Running Press, Philadelphia, 2002) ISBN 0-7624-1348-4; p.733.

- [5] C. Messier, *Catalogue des Nébuleuses & des amas d'Étoiles*, 1781. K. G. Jones, *Messier's nebulae and star clusters*, (Cambridge University Press, 1991) ISBN 0-521-37079-5.
- [6] I. Kant, *Allgemeine Naturgeschichte und Theorie des Himmels*, (Germany, 1755).
- [7] E. Hubble, *The Realm of Nebulae*, (Yale University Press, New Haven, 1936; reprinted by Dover Publications, Inc., New York, 1958).
- [8] M. Planck, Verh. d. deutsch. phys. Ges. **2**, 202 (1900); Verh. d. deutsch. phys. Ges. **2**, 237 (1900); Annalen Phys. **4**, 553 (1901).
- [9] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*, (John Wiley & Sons, Massachusetts, 1979) ISBN 978-0-471-82759-7.
- [10] L. A. Anchordoqui, arXiv:1512.04361 [physics.pop-ph].
- [11] J. Stefan, Wiener Ber. **79**, 391 (1879).
- [12] L. Boltzmann, Annalen Phys. **22**, 291 (1884).
- [13] W. Wien, Annalen Phys. **52**, 132 (1894).
- [14] For further details see e.g., H. Karttunen, P. Kröger, H. Oja, M. Poutanen, K. J. Donner, *Fundamental Astronomy*, (4th Edition, Springer-Verlag Berlin Heidelberg New York, 2003).
- [15] E. Hertzsprung, Astron. Nachr. **196**, 201 (1913); H. N. Russell, Science **37**, 651 (1913).
- [16] C. Doppler, Abh. Königl. Böhm. Ges. Wiss. **2**, 465 (1843).
- [17] S. Weinberg, *The First Three Minutes: A Modern View of the Origin of the Universe*, (BasicBooks, New York, 1993) ISBN 0-465-02437-8.
- [18] J. Fraunhofer, *Determination of the refractive and color-dispersing power of different types of glass, in relation to the improvement of achromatic telescopes*, Memoirs of the Royal Academy of Sciences in Munich **5**, 193 (1814-1815); see especially pages 202-205 and the plate following page 226.
- [19] W. Huggins, Philos. Trans. Roy. Soc. London **158**, 529 (1968) doi:10.1098/rstl.1868.0022
- [20] H. A. Lorentz, Proc. R. Neth. Acad. Arts Sci. **6**, 809 (1904).
- [21] L. A. Anchordoqui, arXiv:1509.08868 [physics.pop-ph].
- [22] H. A. Bethe, Phys. Rev. **55**, 434 (1939).
- [23] S. Chandrasekhar, Mon. Not. Roy. Astron. Soc. **95**, 207 (1935).
- [24] M. S. Longair, (Cambridge University Press, UK, 2011) ISBN 978-0-521-75618-1.
- [25] J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374 (1939).
- [26] L. I. Sedov, J. App. Math. Mech. **10**, 241 (1946).
- [27] G. I. Taylor Proc. Roy. Soc. **201**, 159 (1950) doi:10.1098/rspa.1950.0049.
- [28] G. I. Taylor, Proc. Roy. Soc. **201**, 175 (1950) doi:10.1098/rspa.1950.0050.
- [29] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins Nature **217**, 709 (1968) doi:10.1038/217709a0.
- [30] T. Gold, Nature **218**, 731 (1968) doi:10.1038/218731a0.
- [31] P. E. Boynton, E. J. Groth III, R. B. Partridge, and D. T. Wilkinson, Astrophys. J. **157** L 197 (1969).
- [32] J. R. Oppenheimer and H. Snyder, Phys. Rev. **56**, 455 (1939).
- [33] R. Penrose, Phys. Rev. Lett. **14**, 57 (1965).
- [34] S. Hawking, Phys. Rev. Lett. **15**, 689 (1965).
- [35] S. Hawking, Proc. Roy. Soc. Lond. A **294**, 511 (1966).
- [36] S. Hawking, Proc. Roy. Soc. Lond. A **295**, 490 (1966);
- [37] S. Hawking, Proc. Roy. Soc. Lond. A **300**, 187 (1967).
- [38] S. W. Hawking and R. Penrose, Proc. Roy. Soc. Lond. A **314**, 529 (1970).
- [39] K. F. Gauss, *General investigations of curved surfaces of 1827 and 1825*, (C. S. Robinson & Co., University Press Princeton, N. J., 1902).
- [40] J. Bolyai, *Appendix: Explaining the absolute true of space*, published as an appendix to the essay by his father F. Bolyai *An attempt to introduce youth to the fundamentals of pure science, elementary and advanced, by a clear and proper method* (Maros Vásárhely, Transilvania, 1832).
- [41] N. I Lobachevsky, Kasanski Vestnik (Kazan Messenger), Feb-Mar, 178 (1829); April, 228 (1829); Nov-Dec, 227 (1829); Mar-Apr, 251 (1830); Jul-Aug 571 (1830).
- [42] F. W. Bessel; communicated by J. F. W. Herschel Mon. Not. Roy. Astron. Soc. **6**, 136 (1844).
- [43] A. Clark, communicated by T. H. Safford, *The observed motions of the companion of Sirius* (Cambridge: Welch, Bigelow, and Company, MA, 1863).
- [44] W. S. Adams, Publications of the Astronomical Society of the Pacific **27**, 236 (1915) doi:10.1086/122440.
- [45] W. Pauli, Z. Phys. **31**:765 (1925).
- [46] W. Heisenberg, Z. Phys. **43**, 172 (1927).
- [47] A. Einstein, Annalen Phys. **17**, 891 (1905) [Annalen Phys. **14**, 194 (2005)].
- [48] H. Minkowski, Physikalische Zeitschrift **10**, 104 (1909).
- [49] K. Schwarzschild, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1916**, 189 (1916) [physics/9905030].
- [50] A. Einstein, Annalen Phys. **49**, 769 (1916) [Annalen Phys. **14**, 517 (2005)]. doi:10.1002/andp.200590044
- [51] E. Kretschmann, Annalen Phys. **53**, 575 (1917).
- [52] S. Weinberg, *Gravitation and Cosmology* (John Wiley & Sons, New York, 1972) ISBN 0-471-92567-5
- [53] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, (W. H. Freeman, San Francisco, 1973) ISBN 978-0-7167-0344-0.
- [54] F. W. Dyson, A. S. Eddington, and C. Davidson, Phil. Trans. Roy. Soc. **220A**, 291 (1920).
- [55] D. Adams, *Hitchhiker's Guide to the Galaxy: Life, the Universe and Everything*, (Harmony Books, NY, 1982) ISBN 0-345-39182-9; see chapter 17.
- [56] A. G. W. Cameron, Nature **229**, 178 (1971);
- [57] R. E. Wilson, Astrophys. J. **170**, 529 (1971).
- [58] R. Giacconi, P. Gorenstein, H. Gursky, J. R. Waters, Astrophys. J. **148**, L119 (1967)
- [59] M. Oda, P. Gorenstein, H. Gursky, E. Kellogg, E. Schreier, H. Tananbaum, R. Giacconi, Astrophys. J. **166**, L1 (1971).
- [60] B. L. Webster and P. Murdin, Nature, **235**, 37 (1972).
- [61] C. T. Bolton, Nature **235**, 271 (1972).
- [62] J. A. Pettersson, Astrophys. J. **224**, 625 (1978).
- [63] A. M. Stirling, R. E. Spencer, C. de la Force, M. A. Garrett, R. P. Fender and R. N. Ogley, Mon. Not. Roy. Astron. Soc. **327**, 1273 (2001) doi:10.1046/j.1365-8711.2001.04821.x [astro-ph/0107192].
- [64] J. J. Thomson, *Conduction of electricity through gases* (Cambridge University Press, Cambridge, 1906).
- [65] A. S. Eddington, *The internal constitution of the stars* (Cambridge University Press, Cambridge, 1926).
- [66] J. Biteau, PhD thesis, 2013, pastel-00822242.

- [67] C. M. Urry and P. Padovani, *Publ. Astron. Soc. Pac.* **107**, 803 (1995) doi:10.1086/133630 [astro-ph/9506063].
- [68] C. D. Dermer and B. Giebels, arXiv:1602.06592 [astro-ph.HE].
- [69] B. G. Piner, D. Bhattacharjee, P. G. Edwards and D. L. Jones, *Astrophys. J.* **640**, 196 (2006) doi:10.1086/500006 [astro-ph/0511664].
- [70] J. P. L. de Cheseaux, *Traité de la Comète* (Lausanne, 1774), pp. 223 ff; reprinted in *The Bowl of Night*, by F. P. Dickson (MIT Press, Cambridge, 1968) Appendix II.
- [71] H. W. M. Olbers, *Bode's Jahrbuch*, 111 (1826); reprinted by Dickson, *op. cit.*, Appendix I.
- [72] E. Hubble, *Proc. Nat. Acad. Sci.* **15**, 168 (1929).
- [73] W. L. Freedman *et al.* [HST Collaboration], *Astrophys. J.* **553**, 47 (2001) doi:10.1086/320638 [astro-ph/0012376].
- [74] <http://www.sdss.org/iow/archive.html>
- [75] E. A. Milne, *Z. Astrophysik* **6**, 1 (1933).
[76]
- [76] B. Ryden, *Introduction to cosmology*, (Addison-Wesley, San Francisco, USA, 2003) ISBN 978-0805389128
- [77] A. Einstein, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1917**, 142 (1917).
- [78] A. Friedmann, *Z. Phys.* **10**, 377 (1922).
- [79] A. Friedmann, *Z. Phys.* **21**, 326 (1924).
- [80] H. P. Robertson, *Astrophys. J.* **82**, 284 (1935).
- [81] H. P. Robertson, *Astrophys. J.* **83**, 187, 257 (1936).
- [82] A. G. Walker, *Proc. Lond. Math. Soc. (2)*, **42** 90 (1936).
- [83] V. Mukhanov, *Physical Foundations of Cosmology*, (Cambridge University Press, UK, 2005) ISBN: 978-0-521-56398-7
- [84] N. Pogson, *Mon. Not. Roy. Astron. Soc.* **17**, 12 (1856).
- [85] A. G. Riess *et al.* [Supernova Search Team Collaboration], *Astron. J.* **116**, 1009 (1998) doi:10.1086/300499 [astro-ph/9805201].
- [86] S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], *Astrophys. J.* **517**, 565 (1999) doi:10.1086/307221 [astro-ph/9812133].
- [87] M. Hamuy *et al.*, *Astron. J.* **106**, 2392 (1993).
- [88] M. Hamuy, M. M. Phillips, J. Maza, N. B. Suntzeff, R. A. Schommer and R. Aviles, *Astron. J.* **109**, 1 (1995). doi:10.1086/117251
- [89] S. Perlmutter, *Phys. Today*, April 2003.
- [90] N. A. Bahcall, J. P. Ostriker, S. Perlmutter and P. J. Steinhardt, *Science* **284**, 1481 (1999) doi:10.1126/science.284.5419.1481 [astro-ph/9906463].
- [91] A. A. Penzias and R. W. Wilson, *Astrophys. J.* **142**, 419 (1965).
- [92] R. H. Dicke, P. J. E. Peebles, P. G. Roll and D. T. Wilkinson, *Astrophys. J.* **142**, 414 (1965). doi:10.1086/148306
- [93] G. F. Smoot, astro-ph/9705101.
- [94] J. C. Mather *et al.*, *Astrophys. J.* **420**, 439 (1994).
- [95] C. L. Bennett *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **208**, 20 (2013) [arXiv:1212.5225 [astro-ph.CO]].
- [96] R. Adam *et al.* [Planck Collaboration], arXiv:1502.01582 [astro-ph.CO].
- [97] S. Weinberg, *Cosmology*, (Oxford University Press, UK, 2008) ISBN 978-0-19-852682-7.
- [98] S. Dodelson, *Modern Cosmology*, (Academic Press, Elsevier, Amsterdam, 2003) ISBN 978-0-12-219141-1.
- [99] L. A. Anchordoqui and T. C. Paul, *Mathematical models of physics problems* (Nova, New York, 2013) ISBN 978-1-62618-600-2.
- [100] P. B. Denton, L. A. Anchordoqui, A. A. Berlind, M. Richardson, and T. J. Weiler (for the JEM-EUSO Collaboration), *J. Phys. Conf. Ser.* **531**, 012004 (2014) doi:10.1088/1742-6596/531/1/012004 [arXiv:1401.5757 [astro-ph.IM]].
- [101] G. Hinshaw *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **180**, 225 (2009) doi:10.1088/0067-0049/180/2/225 [arXiv:0803.0732 [astro-ph]].
- [102] C. H. Lineweaver, *ASP Conf. Ser.* **126**, 185 (1997) [astro-ph/9702042].
- [103] V. C. Rubin and W. K. Ford, Jr., *Astrophys. J.* **159**, 379 (1970). doi:10.1086/150317
- [104] V. C. Rubin, N. Thonnard and W. K. Ford, Jr., *Astrophys. J.* **238**, 471 (1980). doi:10.1086/158003
- [105] V. C. Rubin, D. Burstein, W. K. Ford, Jr. and N. Thonnard, *Astrophys. J.* **289**, 81 (1985). doi:10.1086/162866
- [106] F. Zwicky, *Helv. Phys. Acta* **6**, 110 (1933).
- [107] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones and D. Zaritsky, *Astrophys. J.* **648**, L109 (2006) doi:10.1086/508162 [astro-ph/0608407].
- [108] J. L. Feng, *Ann. Rev. Astron. Astrophys.* **48**, 495 (2010) doi:10.1146/annurev-astro-082708-101659 [arXiv:1003.0904 [astro-ph.CO]].
- [109] P. A. R. Ade *et al.* [Planck Collaboration], arXiv:1502.01589 [astro-ph.CO].
- [110] A. G. Riess *et al.*, *Astrophys. J.* **730**, 119 (2011) Erratum: [*Astrophys. J.* **732**, 129 (2011)] doi:10.1088/0004-637X/732/2/129, 10.1088/0004-637X/730/2/119 [arXiv:1103.2976 [astro-ph.CO]].
- [111] R. A. Knop *et al.* [Supernova Cosmology Project Collaboration], *Astrophys. J.* **598**, 102 (2003) [arXiv:astro-ph/0309368].
- [112] S. W. Allen, R. W. Schmidt and A. C. Fabian, *Mon. Not. Roy. Astron. Soc.* **334**, L11 (2002) [arXiv:astro-ph/0205007].
- [113] A. E. Lange *et al.* [Boomerang Collaboration], *Phys. Rev. D* **63**, 042001 (2001) [arXiv:astro-ph/0005004].
- [114] A. Balbi *et al.*, *Astrophys. J.* **545**, L1 (2000) [Erratum-ibid. **558**, L145 (2001)] [arXiv:astro-ph/0005124].
- [115] S. M. Carroll, W. H. Press and E. L. Turner, *Ann. Rev. Astron. Astrophys.* **30**, 499 (1992).
- [116] B. S. Meyer and D. N. Schramm, *Astrophys. J.* **311**, 406 (1986).
- [117] G. Aldering *et al.* [SNAP Collaboration], arXiv:astro-ph/0209550.
- [118] F. Halzen and A. D. Martin, *Quarks and leptons: An introductory course In modern particle physics*, (John Wiley & Sons, New York, 1984) ISBN 0-471-88741-2
- [119] V. D. Barger and R. J. N. Phillips, *Collider physics*, *Front. Phys.* **71**, 1 (1991) ISBN 0-201-14945-1
- [120] C. Quigg, *Gauge Theories of the strong, weak, and electromagnetic interactions*, *Front. Phys.* **56**, 1 (1983) ISBN 978-0805360202
- [121] L. Anchordoqui and F. Halzen, arXiv:0906.1271 [physics.ed-ph].
- [122] K. A. Olive *et al.* [Particle Data Group Collaboration], *Chin. Phys. C* **38**, 090001 (2014). doi:10.1088/1674-1137/38/9/090001
- [123] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Phys. Lett. B* **47**, 365 (1973).
- [124] D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973).

- [125] H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
- [126] M. Gell-Mann, CTSL-20, TID-12608.
- [127] Y. Ne'eman, Nucl. Phys. **26**, 222 (1961).
- [128] M. Gell-Mann, Phys. Lett. **8**, 214 (1964).
- [129] S. N. Bose, Z. Phys. **26**, 178 (1924).
- [130] A. Einstein, [Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **22**, 261 (1924); **1**, 3 (1925); **3**, 18 (1925)].
- [131] E. Fermi, Rend. Lincei **3**, 145 (1926); Z. Phys. **36**, 902 (1926).
- [132] P. A. M. Dirac, Proc. R. Soc. Lond. Ser. A **112**, 661 (1926).
- [133] J. S. Schwinger, Phys. Rev. **74**, 1439 (1948).
- [134] J. S. Schwinger, Phys. Rev. **75**, 651 (1948). doi:10.1103/PhysRev.75.651
- [135] S. Tomonaga, Prog. Theor. Phys. **1**, 27 (1946).
- [136] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
- [137] F. J. Dyson, Phys. Rev. **75**, 486 (1949).
- [138] F. J. Dyson, Phys. Rev. **75**, 1736 (1949).
- [139] R. P. Feynman, Phys. Rev. **80**, 440 (1950). doi:10.1103/PhysRev.80.440
- [140] J. S. Schwinger, Phys. Rev. **73**, 416 (1948).
- [141] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).
- [142] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
- [143] A. Salam, Conf. Proc. C **680519**, 367 (1968).
- [144] P. W. Higgs, Phys. Rev. Lett. **13**, 508 (1964).
- [145] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).
- [146] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **710**, 49 (2012) [arXiv:1202.1408 [hep-ex]].
- [147] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **710**, 26 (2012) [arXiv:1202.1488 [hep-ex]].
- [148] E. W. Kolb and M. S. Turner, *The Early Universe*, Front. Phys. **69**, 1 (1990). ISBN 0-201-11603-0
- [149] K. A. Olive, CERN Yellow Report CERN-2010-002, 149-196 [arXiv:1005.3955 [hep-ph]].
- [150] A. H. Guth, Phys. Rev. D **23**, 347 (1981). doi:10.1103/PhysRevD.23.347
- [151] D. Baumann, doi:10.1142/9789814327183_0010 arXiv:0907.5424 [hep-th].
- [152] A. Riotto, hep-ph/0210162.
- [153] C. H. Lineweaver, astro-ph/0305179.
- [154] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5**, 32 (1967) [JETP Lett. **5**, 24 (1967)] [Sov. Phys. Usp. **34**, 392 (1991)] [Usp. Fiz. Nauk **161**, 61 (1991)]. doi:10.1070/PU1991v03n05ABEH002497
- [155] C. Brust, D. E. Kaplan and M. T. Walters, JHEP **1312**, 058 (2013) doi:10.1007/JHEP12(2013)058 [arXiv:1303.5379 [hep-ph]].
- [156] A. Bazavov *et al.*, Phys. Rev. D **80**, 014504 (2009) doi:10.1103/PhysRevD.80.014504 [arXiv:0903.4379 [hep-lat]].
- [157] L. A. Anchordoqui and H. Goldberg, Phys. Rev. Lett. **108**, 081805 (2012) doi:10.1103/PhysRevLett.108.081805 [arXiv:1111.7264 [hep-ph]].
- [158] M. Laine and Y. Schroder, Phys. Rev. D **73**, 085009 (2006) doi:10.1103/PhysRevD.73.085009 [hep-ph/0603048].
- [159] G. Steigman, B. Dasgupta and J. F. Beacom, Phys. Rev. D **86**, 023506 (2012) doi:10.1103/PhysRevD.86.023506 [arXiv:1204.3622 [hep-ph]].
- [160] L. A. Anchordoqui, H. Goldberg and B. Vlcek, arXiv:1305.0146 [astro-ph.CO].
- [161] R. A. Alpher, J. W. Follin and R. C. Herman, Phys. Rev. **92**, 1347 (1953). doi:10.1103/PhysRev.92.1347
- [162] Ya. B. Zel'dovich, Adv. Astron. Astrophys. **3**, 241 (1965).
- [163] Ya. B. Zel'dovich, Sov. Phys. Usp. **9**, 602 (1967).
- [164] G. Steigman, D. N. Schramm and J. E. Gunn, Phys. Lett. B **66**, 202 (1977). doi:10.1016/0370-2693(77)90176-9
- [165] G. Steigman, K. A. Olive, D. N. Schramm and M. S. Turner, Phys. Lett. B **176**, 33 (1986). doi:10.1016/0370-2693(86)90920-2
- [166] D. A. Dicus, E. W. Kolb, A. M. Gleeson, E. C. G. Sudarshan, V. L. Teplitz and M. S. Turner, Phys. Rev. D **26**, 2694 (1982). doi:10.1103/PhysRevD.26.2694
- [167] S. Dodelson and M. S. Turner, Phys. Rev. D **46**, 3372 (1992). doi:10.1103/PhysRevD.46.3372
- [168] G. Mangano, G. Miele, S. Pastor and M. Peloso, Phys. Lett. B **534**, 8 (2002) doi:10.1016/S0370-2693(02)01622-2 [astro-ph/0111408].
- [169] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P. D. Serpico, Nucl. Phys. B **729**, 221 (2005) doi:10.1016/j.nuclphysb.2005.09.041 [hep-ph/0506164].
- [170] S. Sarkar, Rept. Prog. Phys. **59**, 1493 (1996) doi:10.1088/0034-4885/59/12/001 [hep-ph/9602260].
- [171] K. A. Olive, G. Steigman and T. P. Walker, Phys. Rept. **333**, 389 (2000) doi:10.1016/S0370-1573(00)00031-4 [astro-ph/9905320].
- [172] G. Gamow, Phys. Rev. **70**, 572 (1946). doi:10.1103/PhysRev.7.0572
- [173] R. A. Alpher, H. Bethe and G. Gamow, Phys. Rev. **73**, 803 (1948). doi:10.1103/PhysRev.73.803
- [174] G. Gamow, Rev. Mod. Phys. **21**, 367 (1949). doi:10.1103/RevModPhys.21.367
- [175] Y. I. Izotov, T. X. Thuan and G. Stasinska, Astrophys. J. **662**, 15 (2007) doi:10.1086/513601 [astro-ph/0702072 [ASTRO-PH]].
- [176] M. Peimbert, V. Luridiana and A. Peimbert, Astrophys. J. **666**, 636 (2007) doi:10.1086/520571 [astro-ph/0701580].
- [177] G. Steigman, Ann. Rev. Nucl. Part. Sci. **57**, 463 (2007) doi:10.1146/annurev.nucl.56.080805.140437 [arXiv:0712.1100 [astro-ph]].
- [178] V. Simha and G. Steigman, JCAP **0806**, 016 (2008) doi:10.1088/1475-7516/2008/06/016 [arXiv:0803.3465 [astro-ph]].
- [179] Y. I. Izotov and T. X. Thuan, Astrophys. J. **710**, L67 (2010) doi:10.1088/2041-8205/710/1/L67 [arXiv:1001.4440 [astro-ph.CO]].
- [180] E. Aver, K. A. Olive and E. D. Skillman, JCAP **1103**, 043 (2011) doi:10.1088/1475-7516/2011/03/043 [arXiv:1012.2385 [astro-ph.CO]].
- [181] E. Aver, K. A. Olive and E. D. Skillman, JCAP **1005**, 003 (2010) doi:10.1088/1475-7516/2010/05/003 [arXiv:1001.5218 [astro-ph.CO]].
- [182] Y. I. Izotov, T. X. Thuan and N. G. Guseva, Mon. Not. Roy. Astron. Soc. **445**, no. 1, 778 (2014) doi:10.1093/mnras/stu1771 [arXiv:1408.6953 [astro-ph.CO]].
- [183] P. A. R. Ade *et al.* [Planck Collaboration], Astron. Astrophys. **571**, A16 (2014) doi:10.1051/0004-6361/201321591 [arXiv:1303.5076 [astro-ph.CO]].
- [184] L. A. Anchordoqui, H. Goldberg and G. Steigman, Phys. Lett. B **718**, 1162 (2013) doi:10.1016/j.physletb.2012.12.019 [arXiv:1211.0186 [hep-ph]].

- [185] S. W. Hawking, Nature **248**, 30 (1974).
- [186] S. W. Hawking, Commun. Math. Phys. **43** (1975) 199.
- [187] J. B. Hartle and S. W. Hawking, Phys. Rev. D **13** (1976) 2188.
- [188] L. Parker, Phys. Rev. D **12**, 1519 (1975).
- [189] R. M. Wald, Commun. Math. Phys. **45**, 9 (1975).
- [190] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
- [191] D. N. Page, Phys. Rev. D **13** (1976) 198.
- [192] D. N. Page and S. W. Hawking, Astrophys. J. **206**, 1 (1976).
- [193] S. W. Hawking, Phys. Rev. D **13**, 191 (1976).
- [194] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **429**, 263 (1998) doi:10.1016/S0370-2693(98)00466-3 [hep-ph/9803315].
- [195] T. Banks and W. Fischler, hep-th/9906038.
- [196] S. Dimopoulos and G. L. Landsberg, Phys. Rev. Lett. **87**, 161602 (2001) doi:10.1103/PhysRevLett.87.161602 [hep-ph/0106295].
- [197] S. B. Giddings and S. D. Thomas, Phys. Rev. D **65**, 056010 (2002) doi:10.1103/PhysRevD.65.056010 [hep-ph/0106219].
- [198] J. L. Feng and A. D. Shapere, Phys. Rev. Lett. **88**, 021303 (2002) doi:10.1103/PhysRevLett.88.021303 [hep-ph/0109106].
- [199] L. A. Anchordoqui, J. L. Feng, H. Goldberg and A. D. Shapere, Phys. Rev. D **65**, 124027 (2002) doi:10.1103/PhysRevD.65.124027 [hep-ph/0112247].
- [200] R. Emparan, G. T. Horowitz and R. C. Myers, Phys. Rev. Lett. **85**, 499 (2000) doi:10.1103/PhysRevLett.85.499 [hep-th/0003118].
- [201] L. A. Anchordoqui, J. L. Feng, H. Goldberg and A. D. Shapere, Phys. Lett. B **594**, 363 (2004) doi:10.1016/j.physletb.2004.05.051 [hep-ph/0311365].
- [202] L. Anchordoqui and H. Goldberg, Phys. Rev. D **67**, 064010 (2003) doi:10.1103/PhysRevD.67.064010 [hep-ph/0209337].
- [203] R. C. Myers and M. J. Perry, Annals Phys. **172**, 304 (1986). doi:10.1016/0003-4916(86)90186-7
- [204] A. Chamblin, F. Cooper and G. C. Nayak, Phys. Rev. D **69**, 065010 (2004) doi:10.1103/PhysRevD.69.065010 [hep-ph/0301239].
- [205] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **436**, 257 (1998) doi:10.1016/S0370-2693(98)00860-0 [hep-ph/9804398].
- [206] C. M. Harris and P. Kanti, JHEP **0310**, 014 (2003) doi:10.1088/1126-6708/2003/10/014 [hep-ph/0309054].
- [207] W. L. Kraushaar, G. W. Clark, G. P. Garmire, R. Borken, P. Higbie, and C. Leong, and T. Thorsos, Astrophys. J. **177**, 341 (1972).
- [208] C. E. Fichtel, R. C. Hartman, D. A. Kniffen, D. J. Thomson, H. Ogelman, M. E. Ozel, T. Turner, and G. F. Bignami Astrophys. J. **198**, 163 (1975).
- [209] P. Sreekumar *et al.* [EGRET Collaboration], Astrophys. J. **494**, 523 (1998) doi:10.1086/305222 [astro-ph/9709257].
- [210] R. C. Hartman *et al.* [EGRET Collaboration], Astrophys. J. Suppl. **123**, 79 (1999).
- [211] M. Ackermann *et al.* [Fermi-LAT Collaboration], Astrophys. J. **799**, 86 (2015) doi:10.1088/0004-637X/799/1/86 [arXiv:1410.3696 [astro-ph.HE]].
- [212] M. G. Aartsen *et al.* [IceCube Collaboration], Phys. Rev. D **91**, no. 2, 022001 (2015) doi:10.1103/PhysRevD.91.022001 [arXiv:1410.1749 [astro-ph.HE]].
- [213] L. A. Anchordoqui, H. Goldberg, T. C. Paul, L. H. M. da Silva and B. J. Vlcek, Phys. Rev. D **90**, no. 12, 123010 (2014) doi:10.1103/PhysRevD.90.123010 [arXiv:1410.0348 [astro-ph.HE]].
- [214] For further details see e.g., F. A. Aharonian, *Very high energy cosmic gamma radiation: A critical window on the extreme universe*, (Singapore: World Scientific Publishing, 2004) ISBN 981-02-4573-4.
- [215] J. G. Learned and K. Mannheim, Ann. Rev. Nucl. Part. Sci. **50**, 679 (2000).
- [216] V. F. Hess, Phys. Z. **13**, 1804 (1912).
- [217] P. Auger, R. Maze, T. Grivet-Meyer, Comptes Rendus **206**, 1721 (1938).
- [218] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, Robley, and A. Fréon, Rev. Mod. Phys. **11**, 288 (1939).
- [219] D. J. Bird *et al.*, Astrophys. J. **441**, 144 (1995).
- [220] A. Aab *et al.* [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **798**, 172 (2015) doi:10.1016/j.nima.2015.06.058 [arXiv:1502.01323 [astro-ph.IM]].
- [221] J. Abraham *et al.* [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **613**, 29 (2010) doi:10.1016/j.nima.2009.11.018 [arXiv:1111.6764 [astro-ph.IM]].
- [222] J. Abraham *et al.* [Pierre Auger Collaboration], Nucl. Instrum. Meth. A **620**, 227 (2010) doi:10.1016/j.nima.2010.04.023 [arXiv:0907.4282 [astro-ph.IM]].
- [223] L. Anchordoqui, M. T. Dova, A. G. Mariazzi, T. McCauley, T. C. Paul, S. Reucroft and J. Swain, Annals Phys. **314**, 145 (2004) doi:10.1016/j.aop.2004.07.003 [hep-ph/0407020].
- [224] D. F. Torres and L. A. Anchordoqui, Rept. Prog. Phys. **67**, 1663 (2004) doi:10.1088/0034-4885/67/9/R03 [astro-ph/0402371].
- [225] T. Antoni *et al.* [KASCADE Collaboration], Astropart. Phys. **24**, 1 (2005) doi:10.1016/j.astropartphys.2005.04.001 [astro-ph/0505413].
- [226] D. J. Bird *et al.* [HiRes Collaboration], Phys. Rev. Lett. **71**, 3401 (1993). doi:10.1103/PhysRevLett.71.3401
- [227] R. U. Abbasi *et al.* [HiRes Collaboration], Phys. Rev. Lett. **100**, 101101 (2008) doi:10.1103/PhysRevLett.100.101101 [astro-ph/0703099].
- [228] J. Abraham *et al.* [Pierre Auger Collaboration], Phys. Rev. Lett. **101**, 061101 (2008) doi:10.1103/PhysRevLett.101.061101 [arXiv:0806.4302 [astro-ph]].
- [229] W. D. Apel *et al.*, Astropart. Phys. **36**, 183 (2012). doi:10.1016/j.astropartphys.2012.05.023
- [230] M. G. Aartsen *et al.* [IceCube Collaboration], Phys. Rev. D **88**, no. 4, 042004 (2013) doi:10.1103/PhysRevD.88.042004 [arXiv:1307.3795 [astro-ph.HE]].
- [231] T. Abu-Zayyad *et al.* [HiRes-MIA Collaboration], Astrophys. J. **557**, 686 (2001) doi:10.1086/322240 [astro-ph/0010652].
- [232] D. R. Bergman and J. W. Belz, J. Phys. G **34**, R359 (2007) doi:10.1088/0954-3899/34/10/R01 [arXiv:0704.3721 [astro-ph]].
- [233] K. H. Kampert and M. Unger, Astropart. Phys. **35**, 660 (2012) doi:10.1016/j.astropartphys.2012.02.004 [arXiv:1201.0018 [astro-ph.HE]].

- [234] P. Abreu *et al.* [Pierre Auger Collaboration], Astropart. Phys. **34**, 627 (2011) doi:10.1016/j.astropartphys.2010.12.007 [arXiv:1103.2721 [astro-ph.HE]].
- [235] P. Abreu *et al.* [Pierre Auger Collaboration], Astrophys. J. Suppl. **203** (2012) 34 doi:10.1088/0067-0049/203/2/34 [arXiv:1210.3736 [astro-ph.HE]].
- [236] A. Aab *et al.* [Pierre Auger Collaboration], Astrophys. J. **802**, no. 2, 111 (2015) doi:10.1088/0004-637X/802/2/111 [arXiv:1411.6953 [astro-ph.HE]].
- [237] A. Aab *et al.* [Pierre Auger Collaboration], Phys. Rev. D **90**, no. 12, 122005 (2014) doi:10.1103/PhysRevD.90.122005 [arXiv:1409.4809 [astro-ph.HE]].
- [238] A. Aab *et al.* [Pierre Auger Collaboration], Phys. Rev. D **90**, no. 12, 122006 (2014) doi:10.1103/PhysRevD.90.122006 [arXiv:1409.5083 [astro-ph.HE]].
- [239] V. Berezinsky, A. Z. Gazizov and S. I. Grigorieva, Phys. Rev. D **74**, 043005 (2006) doi:10.1103/PhysRevD.74.043005 [hep-ph/0204357].
- [240] M. Ahlers, L. A. Anchordoqui and A. M. Taylor, Phys. Rev. D **87**, no. 2, 023004 (2013) doi:10.1103/PhysRevD.87.023004 [arXiv:1209.5427 [astro-ph.HE]].
- [241] R. Aloisio, V. Berezinsky and P. Blasi, JCAP **1410**, no. 10, 020 (2014) doi:10.1088/1475-7516/2014/10/020 [arXiv:1312.7459 [astro-ph.HE]].
- [242] G. Giacinti, M. Kachelrie and D. V. Semikoz, Phys. Rev. D **91**, no. 8, 083009 (2015) doi:10.1103/PhysRevD.91.083009 [arXiv:1502.01608 [astro-ph.HE]].
- [243] M. Unger, G. R. Farrar and L. A. Anchordoqui, Phys. Rev. D **92**, no. 12, 123001 (2015) doi:10.1103/PhysRevD.92.123001 [arXiv:1505.02153 [astro-ph.HE]].
- [244] T. Abu-Zayyad *et al.* [Telescope Array Collaboration], Astrophys. J. **768**, L1 (2013) doi:10.1088/2041-8205/768/1/L1 [arXiv:1205.5067 [astro-ph.HE]].
- [245] J. Abraham *et al.* [Pierre Auger Collaboration], Phys. Lett. B **685**, 239 (2010) doi:10.1016/j.physletb.2010.02.013 [arXiv:1002.1975 [astro-ph.HE]].
- [246] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966). doi:10.1103/PhysRevLett.16.748
- [247] G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4**, 78 (1966) [Pisma Zh. Eksp. Teor. Fiz. **4**, 114 (1966)].
- [248] R. Aloisio, V. Berezinsky and A. Gazizov, Astropart. Phys. **34**, 620 (2011) doi:10.1016/j.astropartphys.2010.12.008 [arXiv:0907.5194 [astro-ph.HE]].
- [249] L. A. Anchordoqui, H. Goldberg, F. Halzen and T. J. Weiler, Phys. Lett. B **593**, 42 (2004) doi:10.1016/j.physletb.2004.04.054 [astro-ph/0311002].
- [250] L. A. Anchordoqui, H. Goldberg and T. J. Weiler, Phys. Rev. Lett. **87**, 081101 (2001) doi:10.1103/PhysRevLett.87.081101 [astro-ph/0103043].
- [251] L. A. Anchordoqui, doi:10.5170/CERN-2013-003.303 arXiv:1104.0509 [hep-ph].
- [252] D. N. Schramm, Comments Nucl. Part. Phys. **17**, no. 5, 239 (1987).
- [253] L. A. Anchordoqui and T. Montaruli, Ann. Rev. Nucl. Part. Sci. **60**, 129 (2010) doi:10.1146/annurev.nucl.012809.104551 [arXiv:0912.1035 [astro-ph.HE]].
- [254] A. Achterberg *et al.* [IceCube Collaboration], Astropart. Phys. **26**, 155 (2006) doi:10.1016/j.astropartphys.2006.06.007 [astro-ph/0604450].
- [255] R. Abbasi *et al.* [IceCube Collaboration], Nucl. Instrum. Meth. A **601**, 294 (2009) doi:10.1016/j.nima.2009.01.001 [arXiv:0810.4930 [physics.ins-det]].
- [256] R. Abbasi *et al.* [IceCube Collaboration], Nucl. Instrum. Meth. A **700**, 188 (2013) doi:10.1016/j.nima.2012.10.067 [arXiv:1207.6326 [astro-ph.IM]].
- [257] M. G. Aartsen *et al.* [IceCube Collaboration], Phys. Rev. Lett. **111**, 021103 (2013) doi:10.1103/PhysRevLett.111.021103 [arXiv:1304.5356 [astro-ph.HE]].
- [258] S. Schonert, T. K. Gaisser, E. Resconi and O. Schulz, Phys. Rev. D **79**, 043009 (2009) doi:10.1103/PhysRevD.79.043009 [arXiv:0812.4308 [astro-ph]].
- [259] T. K. Gaisser, K. Jero, A. Karle and J. van Santen, Phys. Rev. D **90**, no. 2, 023009 (2014) doi:10.1103/PhysRevD.90.023009 [arXiv:1405.0525 [astro-ph.HE]].
- [260] M. G. Aartsen *et al.* [IceCube Collaboration], Science **342**, 1242856 (2013) doi:10.1126/science.1242856 [arXiv:1311.5238 [astro-ph.HE]].
- [261] L. A. Anchordoqui, H. Goldberg, M. H. Lynch, A. V. Olinto, T. C. Paul and T. J. Weiler, Phys. Rev. D **89**, no. 8, 083003 (2014) doi:10.1103/PhysRevD.89.083003 [arXiv:1306.5021 [astro-ph.HE]].
- [262] M. G. Aartsen *et al.* [IceCube Collaboration], Phys. Rev. Lett. **113**, 101101 (2014) doi:10.1103/PhysRevLett.113.101101 [arXiv:1405.5303 [astro-ph.HE]].
- [263] M. G. Aartsen *et al.* [IceCube Collaboration], Astrophys. J. **809**, no. 1, 98 (2015) doi:10.1088/0004-637X/809/1/98 [arXiv:1507.03991 [astro-ph.HE]].
- [264] M. G. Aartsen *et al.* [IceCube Collaboration], arXiv:1510.05223 [astro-ph.HE].
- [265] M. G. Aartsen *et al.* [IceCube Collaboration], Phys. Rev. Lett. **114**, no. 17, 171102 (2015) doi:10.1103/PhysRevLett.114.171102 [arXiv:1502.03376 [astro-ph.HE]].
- [266] L. A. Anchordoqui *et al.*, JHEAp **1-2**, 1 (2014) doi:10.1016/j.jheap.2014.01.001 [arXiv:1312.6587 [astro-ph.HE]].
- [267] K. S. Thorne, Rev. Mod. Phys. **52**, 285 (1980). doi:10.1103/RevModPhys.52.285
- [268] K. S. Thorne, Rev. Mod. Phys. **52**, 299 (1980). doi:10.1103/RevModPhys.52.299
- [269] P. C. Peters and J. Mathews, Phys. Rev. **131**, 435 (1963). doi:10.1103/PhysRev.131.435
- [270] R. A. Hulse and J. H. Taylor, Astrophys. J. **195**, L51 (1975).
- [271] A. Abramovici *et al.*, Science **256**, 325 (1992).
- [272] B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], Phys. Rev. Lett. **116**, no. 6, 061102 (2016) doi:10.1103/PhysRevLett.116.061102 [arXiv:1602.03837 [gr-qc]].
- [273] M. G. Aartsen *et al.* [IceCube and LIGO Scien-

- tific and VIRGO Collaborations], Phys. Rev. D **90**, no. 10, 102002 (2014) doi:10.1103/PhysRevD.90.102002 [arXiv:1407.1042 [astro-ph.HE]].
- [274] M. G. Aartsen *et al.* [IceCube and Pierre Auger and Telescope Array Collaborations], JCAP **1601**, no. 01, 037 (2016) doi:10.1088/1475-7516/2016/01/037 [arXiv:1511.09408 [astro-ph.HE]].
- [275] M. Ackermann *et al.* [Fermi-LAT Collaboration], arXiv:1602.04488 [astro-ph.HE].
- [276] S. Adrian-Martinez *et al.* [ANTARES and Ice-Cube and LIGO Scientific and Virgo Collaborations], arXiv:1602.05411 [astro-ph.HE].
- [277] M. G. Aartsen *et al.* [IceCube Collaboration], arXiv:1412.5106 [astro-ph.HE].
- [278] A. Aab *et al.* [Pierre Auger Collaboration], arXiv:1604.03637 [astro-ph.IM].
- [279] R. Osserman, Am. Math. Monthly **108**, 813 (2001).
- [280] J. Kepler, *Harmonices Mundi* (Johann Planck, Linz, Austria, 1619)
- [281] J. Bradley Phil. Trans. **35**, 637, (1727 - 1728) doi:10.1098/rstl.1727.0064.
- [282] T. Mazeh *et al.*, Astrophys. J. **532**, L55 (2000) doi:10.1086/312558 [astro-ph/0001284].
- [283] D. Queloz, A. Eggenberger, M. Mayor, C. Perrier, J. L. Beuzit, D. Naef, J. P. Sivan and S. Udry, Astron. Astrophys. **359**, L13 (2000) [astro-ph/0006213].
- [284] R. A. Wittenmyer *et al.*, Astrophys. J. **632**, 1157 (2005) doi:10.1086/433176 [astro-ph/0504579].
- [285] G. W. Henry, G. W. Marcy, R. P. Butler and S. S. Vogt, Astrophys. J. **529**, L41 (2000). doi:10.1086/312458
- [286] D. Charbonneau, T. M. Brown, D. W. Latham and M. Mayor, Astrophys. J. **529**, L45 (2000) doi:10.1086/312457 [astro-ph/9911436].
- [287] M. Unger, *High energy astrophysics*, lectures given at New York University, 2015.
- [288] S. B. Giddings and M. L. Mangano, Phys. Rev. D **78**, 035009 (2008) doi:10.1103/PhysRevD.78.035009 [arXiv:0806.3381 [hep-ph]].
- [289] S. B. Giddings and M. L. Mangano, arXiv:0808.4087 [hep-ph].
- [290] A. M. Hillas, Ann. Rev. Astron. Astrophys. **22**, 425 (1984). doi:10.1146/annurev.aa.22.090184.002233
- [291] R. M. Bionta *et al.*, Phys. Rev. Lett. **58**, 1494 (1987).
- [292] K. Hirata *et al.* [KAMIOKANDE-II Collaboration], Phys. Rev. Lett. **58**, 1490 (1987);
- [293] J. N. Bahcall, A. Dar and T. Piran, Nature **326**, 135 (1987). doi:10.1038/326135a0
- [294] M. D. Kruskal, Phys. Rev. **119**, 1743 (1960). doi:10.1103/PhysRev.119.1743
- [295] G. 't Hooft, *Introduction to the theory of black holes*, lectures given at Utrecht University, 2009.
- [296] A. Einstein and N. Rosen, Phys. Rev. **48**, 73 (1935). doi:10.1103/PhysRev.48.73
- [297] R. W. Fuller and J. A. Wheeler, Phys. Rev. **128**, 919 (1962). doi:10.1103/PhysRev.128.919

Answers and Comments on the Exercises

1.1 The cycle of the seasons occurs because the Earth's axis of rotation is inclined 23.5° from the perpendicular to its orbital plane. Although precession moves the axis very slowly, the axis is essentially fixed in space. As Earth orbits the Sun, its axis remains pointing in the same direction in space. Seasonal temperature depends on the amount of heat (energy) we receive from the Sun. Two factors are most responsible for the change in this amount. Firstly, and most importantly, the Sun stands higher in the sky at noon on a summer day and shines almost straight down, providing more energy per unit area. On a winter day, however, the noon sun is lower and hence it spreads out and we receive less energy per unit area; see Fig. 41. In other words, the amount of total energy coming towards the Earth from the Sun being almost the same throughout the year, during summer it is intercepted by a smaller area on the ground (Sun shining at a more direct angle), thereby increasing the energy per unit area. During winter, the same amount of energy is spread out over a larger area (Sun shining at a more inclined angle), and therefore the energy per unit area is decreased.

Secondly, the summer Sun is above the horizon for more hours each day than the winter Sun: summer days are longer, and winter days are shorter. Because the Sun is above our horizon longer in summer, we receive more energy each day. The earth, especially the oceans and the atmosphere, retains this heat somewhat even during the night, thereby creating a generally warmer temperature (otherwise, the nights would be equally cold throughout the year). In fact, this storage of the energy also causes the slight delay in the seasons – in the northern hemisphere, July/August are warmer than 21 June (the summer solstice), and January/February are colder than 21 December (the winter solstice).

1.2 Using (1) we have that $4.2 \text{ ly} = 3.97 \times 10^{16} \text{ m}$. The actual radius of the Sun is $R_\odot = 696,000 \text{ km} = 6.96 \times 10^8 \text{ m}$ and the radius of the Sun in our model is $r = 6 \text{ cm} = 6 \times 10^{-2} \text{ m}$. We solve the problem by equating the ratio between the actual solar radius and the model solar radius to the ratio between the actual distance of Proxima Centauri D to the model distance of Proxima Centauri d . It follows that $d = Dr/R_\odot = 3430 \text{ km}$.

1.3 The standard treatment in elementary calculus books of the derivation of Kepler's empirically deduced laws of planetary motion from Newton's law of gravity and his second law of motion is flawed for at least three important

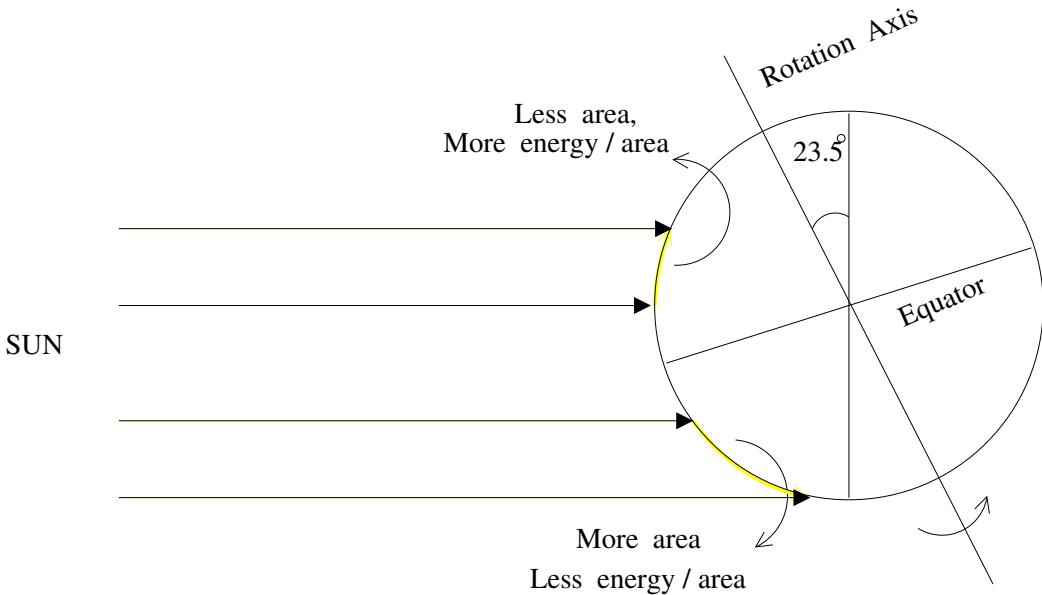


FIG. 41: Sunlight strikes a given area of the Earth at different angles as we progress through the seasons since the direction of Earth's rotation axis is essentially fixed in space.

reasons.¹² First, Newton's laws are used to derive a differential equation for the displacement vector from the Sun to a planet; say the Earth. Then it is shown that the displacement vector lies in a plane, and if the base point is translated to the origin, the endpoint traces out an ellipse. This is said to confirm Kepler's first law, that the planets orbit the Sun in an elliptical path, with the Sun at one focus. However, an alert student may notice that the identical argument for the displacement vector in the opposite direction would show that the Sun orbits the Earth in an ellipse, which, it turns out, is very close to a circle with the Earth at the center. That would seem to provide aid and comfort to the Church's rejection of Galileo's claim that his heliocentric view had more validity than their geocentric one. Second, by placing the Sun at the origin, the impression is given that either the Sun is fixed, or else, that one may choose coordinates attached to a moving body, inertial or not. However, Newton's equations in their usual form hold precisely in inertial coordinates. Furthermore, it is an immediate consequence of Newton's laws that the center-of-mass of the two bodies can serve as the origin of an inertial coordinate system. One then finds that both the Earth and the Sun describe elliptical paths with a focus at the center-of-mass of the pair. Third, by considering only the displacement vector, the treatment misses an opportunity to present one of the greatest recent triumphs of Newton's derivation of Kepler's laws: the discovery of extrasolar planets during the past decade, none of which has been observed directly, but whose existence and orbital parameters have been deduced by detecting and analyzing the motion of the stars that they orbit. Two other considerations make the correct way of treating the subject particularly desirable. First, the current treatment for the displacement vector could be left as is, and only a little algebra is needed to show that the same vector referred to the center-of-mass is a constant scalar multiple of the one referred to the Sun. Second, it connects with an exciting area of current research in astronomy where the methods of elementary calculus have direct application. The purpose of this exercise is to spell out the slight modifications to the standard treatment of Kepler's laws that are needed to obtain the additional and accurate information.

Let $\vec{r}(t)$ be a (3-dimensional) vector function of the parameter t , and assume that it satisfies the differential equation

$$\ddot{\vec{r}}(t) = -c\vec{r}(t)/r^3 \quad (319)$$

where c is a positive constant and r is the magnitude of the vector \vec{r} . Then the following statements hold, and are generally proved in calculus books under the general rubric of Newton's laws imply Kepler's laws [280]:

1. If \vec{r} is viewed as a position vector whose base point is at the origin, then its endpoint lies on a plane curve that is either along a ray through the origin, or else a conic section with focus at the origin. In the latter case, if the solution curve is bounded, it must be an ellipse. (Kepler's First law.)

¹² Our discussion follows closely that given in [279]

2. The vector $\vec{r}(t)$ sweeps out equal areas in equal times; that is, in equal Δt intervals. (Kepler's second law.)
3. If T is the period, the Δt -interval in which $\vec{r}(t)$ traverses the ellipse once, and if a is the semi-major axis of the ellipse, then T and a are related by the equation

$$T^2/a^3 = d, \quad (320)$$

where $d = 4\pi/c$.

In the two-body problem, one asks for the motion determined by the gravitational force between two bodies, assuming one can neglect all other forces, including those of other more distant bodies whose effect on these two would be secondary. Let us denote the two bodies as the Earth and the Sun, and let \vec{r} be the displacement vector from the Sun to the Earth. Then, the vector function $\vec{r}(t)$, where t is time, does indeed satisfy (319) for a suitable value of the constant c , but not the equation that one finds in calculus books. Namely, if $M_{\oplus} = 5.974 \times 10^{24}$ kg is the mass of the Earth, and $M_{\odot} = 1.989 \times 10^{30}$ kg is the mass of the Sun, then the gravitational force exerted by the Sun on the Earth is $-GM_{\oplus}M_{\odot}\vec{r}/r^3$, where G is the gravitational constant. The claim is then made that by Newton's second law: force equals mass times acceleration, one can equate the gravitational force with the quantity $M_{\oplus}\ddot{\vec{r}}(t)$, so that the mass of the Earth cancels out, and one arrives at (319) with the value of the constant c being $c = GM_{\oplus}$. Furthermore, if we apply the same reasoning to all the other planets circling the Sun, we get the same value of the constant d in (320), as asserted by Kepler's third law. The flaw in this argument is that in order to use the expression $\ddot{\vec{r}}(t)$ for the acceleration, we must be using an inertial frame for our coordinate system. For example, if the Sun happened to be accelerating at exactly the same rate as the Earth at a given instant, then the acceleration of the displacement vector $\vec{r}(t)$ would be zero, even though the force was unchanged. So the implicit assumption being made is that the Sun is fixed. One could of course point out that assumption explicitly, together with a remark that it is close enough to being true that the resulting conclusions are very close to being correct.

Let \vec{X} and \vec{Y} denote the position vectors of the Earth and Sun, respectively, in some inertial coordinate system. If M_{\oplus} is the mass of the Earth, and M_{\odot} the mass of the Sun, then the center-of-mass C of the pair is defined by the equation

$$(M_{\oplus} + M_{\odot})\vec{C} = M_{\oplus}\vec{X} + M_{\odot}\vec{Y}. \quad (321)$$

Let $\vec{r} = \vec{X} - \vec{Y}$ be the displacement vector from the Sun to the Earth. Then Newton's law of gravity says that the gravitational force \vec{F} exerted by the Sun on the Earth is given by $\vec{F} = -GM_{\oplus}M_{\odot}\vec{r}/r^3$, while Newton's second law asserts that $\vec{F} = M_{\oplus}\ddot{\vec{X}}(t)$. Equating these two expressions gives

$$\ddot{\vec{X}}(t) = -GM_{\odot}\vec{r}/r^3. \quad (322)$$

In exactly the same way, the gravitational force of the Earth on the Sun gives

$$\ddot{\vec{Y}}(t) = GM_{\oplus}\vec{r}/r^3. \quad (323)$$

From equations (322) and (323) we see using (321) that $\ddot{\vec{C}}(t) = \vec{0}$. Hence $\vec{C}(t) = t\vec{a} + \vec{b}$ for constant vectors \vec{a} , \vec{b} , and $\vec{C}(t)$ can be used for the origin of a new inertial coordinate system. It also follows by subtracting (322) from (323) that

$$\ddot{\vec{r}}(t) = -G(M_{\oplus} + M_{\odot})\vec{r}/r^3; \quad (324)$$

in other words, (319) holds for $\vec{r}(t)$, with the constant $c = G(M_{\oplus} + M_{\odot})$. We can also write this as $c = M_{\odot}G(1 + \mu)$, with $\mu = M_{\oplus}/M_{\odot}$. For the Earth-Sun pair, one has μ approximately equal to 3×10^{-6} . Even for Jupiter, μ is approximately 10^{-3} so that using the value $c = M_{\odot}G$ in (319) should not have too great an effect on the solution.

Next, let $\vec{x} = \vec{X} - \vec{C}$ and $\vec{y} = \vec{Y} - \vec{C}$ be the position vectors of the Earth and Sun referred to their common center-of-mass. Then $M_{\oplus}\vec{x} + M_{\odot}\vec{y} = \vec{0}$ and $\vec{x} - \vec{y} = \vec{r}$ so that

$$(M_{\oplus} + M_{\odot})\vec{x} = M_{\odot}\vec{r} \quad \text{and} \quad (M_{\oplus} + M_{\odot})\vec{y} = -M_{\oplus}\vec{r}. \quad (325)$$

It follows that the vectors \vec{x} , \vec{y} , and \vec{r} are collinear and that the orbit of the Earth is just a contraction of the orbit of the displacement vector between the Sun and Earth. In fact, we can write these equations as $\vec{x} = \vec{r}/(1 + \mu)$ and $\vec{y} = -\mu\vec{r}/(1 + \mu)$ which implies, in particular, that $\vec{y} = -\mu\vec{x}$.

Hence the Earth and Sun both describe an ellipse with one focus at their common center-of-mass.¹³ Further, both orbits have the same eccentricity and the same period. But since μ is so small, the orbit of the Sun is equally small, and that of the Earth very close to that of the displacement vector \vec{r} . That is the reason that one gets close to the correct answer by assuming that the Sun is fixed, and also the reason that neither Galileo nor the Church was exactly right in asserting that the Sun or the Earth, respectively, was fixed, although Galileo was a lot closer to the truth. However, as minimal as the motion of the Sun or of any other star may be in response to the pull of a planet orbiting around it, it is precisely that motion that has led to the conclusion that there are many planets orbiting other stars, even though none of the planets themselves has ever been directly observed.

What would the mass of the Sun have to be changed to, if nothing else were changed? From (320) we have $T^2 G(M_\oplus + M_\odot^{\text{new}}) = 4\pi^2 a^3$, and so $M_\odot^{\text{new}} = 1.065 M_\odot$, where we have taken $a = 1.496 \times 10^{11} \text{ m} = 1 \text{ AU}$. What would the mass of the Earth have to be changed to, if nothing else were changed? From (320) we have $T^2 G(M_\oplus^{\text{new}} + M_\odot) = 4\pi^2 a^3$ and so $M_\oplus^{\text{new}} = 2.164 \times 10^4 M_\oplus$. What would the orbital radius of the Earth have to be changed to, if nothing else were changed? From (320) we have $T^2 G(M_\oplus + M_\odot) = 4\pi^2 (a^{\text{new}})^3$ and so $a^{\text{new}} = 0.979 \text{ AU}$.

2.1. The parsec is the distance D when the angle ϕ is 1 second of arc. Hence $\phi = 4.848 \times 10^{-6}$, and since $\tan \phi = d/D$, we obtain $D = 3.086 \times 10^{16} \text{ m} = 3.26 \text{ ly}$. ($d = 1.5 \times 10^8 \text{ km}$ is the distance to the Sun.)

2.2 The Earth radius is $R_\oplus \simeq 6378 \text{ km}$ and we consider the two measurements separated 784 km in latitude. Then the two cities are separated by 7° in latitude. If we interpret the 7° separation in terms of stellar parallax then the distance to the Sun would be $d \approx 784 \text{ km} / \tan 7^\circ \sim 6,385 \text{ km}$. If we interpret night and day as the Sun revolving around a flat disk, then the Sun will crash into the Earth.

2.3.(i) The solar constant is $\mathcal{F}_\odot = 1.3 \times 10^3 \text{ W/m}^2$. (ii) The absolute luminosity is $L_\odot = 3.7 \times 10^{26} \text{ W}$.

2.4 The angular width is the inverse tangent of the diameter of our Galaxy divided by the distance to Andromeda,

$$\phi = \arctan \left[\frac{\text{Galaxy diameter}}{\text{Distance to Andromeda}} \right] \approx 2.4^\circ. \quad (326)$$

For the Moon we obtain,

$$\phi = \arctan \left[\frac{\text{Moon diameter}}{\text{Distance to Moon}} \right] \approx 0.52^\circ. \quad (327)$$

Therefore, the galaxy width is about 4.5 times the Moon width.

2.5 In astronomy, the sizes of objects in the sky are often given in terms of their angular diameter θ as seen from Earth, rather than their actual sizes d . For a given observer on Earth, the distance to the object D and d are related by $\tan \theta = d/D$. Since these angular diameters are often small, we can use the small angle approximation which will give us $\tan \theta \approx \theta$. Now, since θ is in units of radians, we often need to convert our measurements into units of degrees. One radian is $180/\pi \approx 57.3^\circ$. So we can rewrite our small angle approximation as $d/D = \theta/57.3^\circ$. When dealing with astronomically distant objects, where angle sizes are extremely small, it is often more practical to present our angles in terms of arcs, thus $\theta = 206,265 d/D$. Since Sirius is far away from the Earth, we assume that the hypothesized planet would be observed very close to Sirius, subtending small angle θ . In this regime we can use the small angle formula to obtain $\theta = 0.38 \text{ arcs}$, where we have taken $d = 1 \text{ AU}$.

2.6 Let D_1 be the distance between Mercury and Saturn when they are closest to each other and D_2 the distance when they are further apart. (i) Then the distance from Mercury to the Sun is $D_{\text{MS}} = (D_2 - D_1)/2 = 3.2 \text{ lm} \simeq 0.385 \text{ AU}$. (ii) The distance between Saturn and the Sun is $D_{\text{SS}} = D_1 + D_{\text{MS}} = 79.5 \text{ lm} \simeq 9.57 \text{ AU}$. NASA's MESSENGER spacecraft slammed into the surface of Mercury on the 30 April 2015 bringing a groundbreaking mission to a dramatic end. The probe crashed at 3:26 pm red-sox time (1926 GMT), gouging a new crater into Mercury's heavily pockmarked surface. This violent demise was inevitable for MESSENGER, which had been orbiting Mercury

¹³ Because the mass of the Sun is about 1 million times the mass of the Earth, the center-of-mass of the Earth-Sun system is pretty close to the center of the Sun, in fact it lies within the body of the Sun. However, the center of the Sun does orbit this center-of-mass point; or, it would if the Earth were the only planet in the Solar System. In fact, the Sun orbits a point which is dominated by the center-of-mass in the Jupiter-Sun system, but you get the idea.

since March 2011 and had run out of fuel. The 10-foot-wide (3 meters) spacecraft was traveling about 8,750 mph (14,080 km/h) at the time of impact, and it likely created a smoking hole in the ground about 52 feet (16 m) wide in Mercury's northern terrain. No observers or instruments witnessed the crash, which occurred on the opposite side of Mercury from Earth. Cassini is the fourth space probe to visit Saturn and the first to enter orbit, and its mission is ongoing as of 2016. It has studied the planet and its many natural satellites since arriving there in 2004.

2.7 The observed luminosity from the Sun when it is not eclipsed is $\pi R_\odot^2 \sigma T_\odot^4$. When Jupiter passes in front of the Sun, it blocks an area of size πr_J^2 , and the observed luminosity decreases to $\pi(R_\odot^2 - r_J^2)\sigma T_\odot^4$, where $r_J = 71,492$ km. The fractional decrease in the apparent surface brightness is

$$\frac{\Delta I}{I_\odot} = \frac{\pi(R_\odot^2 - r_J^2)\sigma T_\odot^4}{\pi R_\odot^2 \sigma T_\odot^4} - 1 = -\frac{r_J^2}{R_\odot^2} \approx -0.01, \quad (328)$$

The eclipse only reduces the brightness by about 1%.

2.8 The range of distances consistent with the measured parallax angle is $1/0.006$ pc $< D < 1/0.004$ pc, or equivalently 167 pc $< D < 250$ pc.

2.9 The faintest stars that can be detected with the HST have apparent brightnesses which are 4×10^{21} fainter than the Sun. This implies that the HST apparent brightness lower threshold is $I_{\text{th}} = 2.5 \times 10^{-22} I_\odot$. For a star \star like the Sun, $L_\star = L_\odot$. Hence using (13) and the ratio method the distance to the star is found to be $d_L = d\sqrt{I_\odot/I_{\text{th}}} = 6.3 \times 10^{10}$ AU = 306 kpc = 10^6 ly, where $d = 1$ AU is the Earth-Sun distance and 1 pc = 206265 AU = 3.262 ly. (ii) For Cepheids, $L_C = 2 \times 10^4 L_\odot$. Because we want to calculate the limiting distance for observing this object we take $I_C = I_{\text{th}}$ and so $d_L = d\sqrt{(L_C I_\odot)/(L_\odot I_{\text{th}})} = 8.94 \times 10^{12}$ AU = 43.4 Mpc = 1.41×10^8 ly.

2.10 (i) The luminosity is not isotropic, because the solid angle subtended by the blackbody from different directions will be different, and the surface brightness should be constant, so each observer will see a different flux depending on the direction. (ii) The maximum flux will be seen along the z -axis, because that is where the subtended solid angle is greatest. The solid angle is $\Omega = \pi(a/d_L)^2$, because the blackbody appears as a circular disk of angular radius $a/d_L \ll 1$. In addition, using $a \ll d_L$, the flux is simply $F = \int I d\Omega$ (because we can approximate $\cos \theta \simeq 1$, where θ is the angle between each point and the center of the observed object, whenever the object observed has an angular size much smaller than 1 radian). Using $I = B = (\sigma T^4/\pi)$ at the surface of the blackbody, $F = I\Omega = \sigma T^4(a/d_L)^2$. (iii) The minimum flux will be seen along any direction along the equator, where the subtended solid angle will be smallest. Now, the projected image of the ellipsoid is an ellipse, with solid angle $\Omega = \pi(ab/d_L^2)$, so just as before, $F = I\Omega = \sigma T^4(ab/d_L^2)$. (iv) Everyone sees the same *apparent* surface brightness, $I = \sigma T^4/\pi$. (v) The total luminosity is the area times σT^4 . All we need is to find the area of the ellipsoid, which can be done for example by dividing the ellipsoid into thin rings of radius x parallel to the $x-y$ plane. For the area A we find:

$$A = 2 \times 2\pi \int_0^a r dr \left[1 + \left(\frac{dz}{dr} \right)^2 \right]^{1/2}, \quad (329)$$

where $z = b(1 - r^2/a^2)^{1/2}$, and

$$L = (\sigma T^4) 4\pi \int_0^a dr r \left[\frac{a^2 - x^2(1 - q^2)}{a^2 - x^2} \right]^{1/2}, \quad (330)$$

where $q = b/a$. If you have the patience, the integral can actually be solved. (vi) The galaxy is different because it is made of individual stars, and each star is spherical (or even if they are oblate because they are rotating, they should have their spin axes randomly oriented and uncorrelated). The condition $NR^2 \ll ab$ guarantees that stars do not block each other, so the flux observed is the sum of the flux from each star and the luminosity is isotropic. This is true also for any optically thin gas, where each atom emits isotropically. Therefore, the flux is the same for all observers along different directions:

$$F = \frac{L}{4\pi d_L^2}, \quad (331)$$

where L is the total luminosity of the galaxy, and the *apparent* surface brightness is different. If all the luminosity is contained within the oblate ellipsoid, the average apparent surface brightness within the projected ellipsoid as seen

by the observer on the z -axis is

$$I = F/\Omega = \frac{L}{4\pi^2 a^2}, \quad (332)$$

and for the observer on the equator,

$$I = F/\Omega = \frac{L}{4\pi^2 ab}. \quad (333)$$

The apparent surface brightness is higher when the galaxy is seen edge-on because the surface density of stars is greater, since the line of sight crosses a greater pathlength through the galaxy. (*vii*) The answer would be modified because stars would block each other so some stars would be occulted. This would be a really compact galaxy since its mean surface brightness would be similar to that of the Sun, and the stellar atmospheres would be actually heated by the other stars in the galaxy significantly. The galaxy would not be stable because every star would on average have a physical collision with another star once every orbit (a good recipe for making a big black hole and some fireworks, more on this below).

2.11 Taking the \log_{10} of (11) we have

$$\log_{10} L = \log_{10}(4\pi R^2 \sigma T^4) = \log_{10}(4\pi\sigma) + \log_{10} R^2 + \log_{10} T^4 = \log_{10}(4\pi\sigma) + 2\log_{10} R + 4\log_{10} T. \quad (334)$$

If R is constant then $\log_{10}(4\pi\sigma) + 2\log_{10} R = \text{constant}$. Then, for constant R , the slope of $\log_{10} L$ vs. $\log_{10} T$ plot is 4, but since in the HR diagram T is plotted increasing to the left the slope is -4 .

3.1 From the inverse Lorentz transformation (19) we get

$$\tan \theta = \frac{\sin \theta_0}{\gamma(\beta - \cos \theta_0)}. \quad (335)$$

Using the identity $1 + \tan^2 \theta = \sec^2 \theta$ it is straightforward to obtain (21).

3.2 The distribution in the S system is given by

$$\frac{dN}{d\Omega} = \frac{dN}{d\Omega_0} \frac{d\Omega_0}{d\Omega}, \quad (336)$$

so we need $d\cos \theta_0 / d\cos \theta$. Invert (21) to obtain

$$\cos \theta_0 = \frac{\beta + \cos \theta}{\beta \cos \theta + 1} \quad \text{and so} \quad \frac{dN}{d\Omega} = \frac{\varkappa(1 - \beta^2)}{(1 + \beta \cos \theta)^2}. \quad (337)$$

This is for a source moving away from O . To get the result for motion towards O , just replace $\beta \rightarrow -\beta$.

3.3 To show that for $v \ll c$ the Doppler shift in wavelength is approximately v/c we use the binomial expansion:

$$\lambda' = \lambda (1 + v/c)^{1/2} (1 - v/c)^{-1/2} \approx \lambda \left[1 + \frac{v}{2c} + \mathcal{O}\left(\frac{v^2}{c^2}\right) \right] \left[1 - \left(-\frac{v}{2c}\right) + \mathcal{O}\left(\frac{v^2}{c^2}\right) \right] \approx \lambda [1 + v/c + \mathcal{O}(v^2/c^2)], \quad (338)$$

and so $\Delta\lambda/\lambda \approx v/c$.

3.4 The wavelengths from single electron energy level transitions are inversely proportional to the square of the atomic number of the nucleus. Therefore, the lines from singly-ionized helium are usually one fourth the wavelength of the corresponding hydrogen lines. Because of their redshift, the lines have 4 times their usual wavelength (i.e., $\lambda' = 4\lambda$) and so $4\lambda = \lambda \sqrt{(1 + v/c)/(1 - v/c)} \Rightarrow v = 0.88 c$.

3.5 Let

$$p^\mu = \left(\frac{h\nu}{c}, -\frac{h\nu}{c} \cos \theta, -\frac{h\nu}{c} \sin \theta, 0 \right) \quad (339)$$

be the momentum 4-vector for the photon as seen in S and

$$p'^\mu = \left(\frac{h\nu'}{c}, -\frac{h\nu'}{c} \cos \theta', -\frac{h\nu'}{c} \sin \theta', 0 \right) \quad (340)$$

in S' . Do the direct Lorentz transformation from $S \rightarrow S'$ to get $\nu' \cos \theta' = \nu \gamma(\cos \theta + \beta)$, $\nu' \sin \theta' = \nu \sin \theta$, and $\nu' = \nu \gamma(1 + \beta \cos \theta)$. Use the last relation to eliminate ν and ν' giving

$$\cos \theta' = \left(\frac{\beta + \cos \theta}{1 + \beta \cos \theta} \right) \quad \text{and} \quad \sin \theta' = \frac{\sin \theta}{\gamma(1 + \beta \cos \theta)}. \quad (341)$$

For small β , this gives $\cos \theta' \approx \cos \theta + \beta \sin^2 \theta$. Now, using $\theta' = \theta - \alpha$, since α is small we have $\cos(\theta - \alpha) = \cos \theta + \alpha \sin \theta$, and so $\alpha \approx \beta \sin \theta$. This is in agreement with the data of Bradley [281], a result which caused problems for the æther theory of electromagnetic waves (for details see e.g. [21]).

3.6 The universality of Newton's law of gravitation tells us that all the equations and conclusions derived for the Sun and Earth interaction also hold for any system consisting of a star and a single planet orbiting around it. In particular, the period T of the planet is the same as the period of the star: the time it takes each of them to complete one orbit around their common center-of-mass. The period is the easiest parameter to determine, since precisely what is detected is a periodic motion of the star. (At least this is true in theory; in practice, determining the period from a finite set of observations can prove tricky.) The other quantity that we are usually able to estimate fairly accurately is the mass M of the star, based on the spectral type and luminosity of the star. From (320) it follows that the ratio T^2/a^3 is not exactly the same for all planets, but is very close to $4\pi^2/(GM)$, since the ratio m/M is, almost by definition of a planet, very close to zero. (i) As a consequence, the first fact about the unseen planet that we can infer immediately from the knowledge of M and T is its distance from the star, $a = [(GMT^2)/(4\pi^2)]^{1/3} = 0.046$ AU. (ii) In the simplest case of a nearly circular orbit, the planet describes a circle of radius a at constant velocity v , and the star describes a circle with constant velocity V , both orbits having period T . Then $vT = 2\pi a$, and since T determines a , by Kepler's third law, we also have the velocity v of the planet. Then, from (325) we conclude that $vm = VM$ so that we could determine the mass m of the planet, if we knew the value of V . If the plane of the orbit contained our line of sight, then V would simply be the maximal radial velocity. In general, if i denotes the angle of inclination between the normal to the plane of the orbit and our line of sight to the star, then the maximal radial velocity would be $K = V \sin i$, and hence we can deduce the quantity $m \sin i = KM/v$. Using the measured value of i we get $m = 0.63M_J$, where $M_J = 1.898 \times 10^{27}$ kg is the mass of Jupiter. (iii) From (328) it follows that $r = \sqrt{0.02R^2} \approx 1.5r_J$.

4.1 The equation of hydrostatic support is (40). For a constant density, we can set $M(r) = 4\rho_0\pi r^3/3$. Separation of variables and integration from the center to the surface (where $P = P_s = 0$) yields

$$\int_{P_c}^{P_s} dP = 0 - P_c = - \int_0^R \frac{4}{3} \frac{G\rho_0\pi r^3 \rho_0}{r^2} dr = \frac{4}{3}\pi G\rho_0^2 \int_0^R r dr = -\frac{4}{6}\pi G\rho_0^2 R^2. \quad (342)$$

Substituting ρ_0 by $M/V = M/(4\pi R^3/3)$ we obtain the result

$$P_c = \frac{3}{8\pi} G \frac{M^2}{R^4}. \quad (343)$$

(ii) The mass within a radius r is

$$M(r) = \int_0^r \rho(x) 4\pi x^2 dx = 4\pi \rho_c \left(\frac{1}{3}r^3 - \frac{1}{4}\frac{r^4}{R} \right). \quad (344)$$

For $r = R$ the total mass of the star is found to be $M = \pi \rho_c R^3/3$ and therefore we can express the central density in terms of M and R

$$\rho_c = \frac{3}{\pi} \frac{M}{R^3}, \quad \text{with which} \quad M(r) = 12M \left(\frac{1}{3}\xi^3 - \frac{1}{4}\xi^4 \right), \quad (345)$$

where $\xi = r/R$ is the scaled radius. The integral of the equation of hydrostatic support is then

$$P_c = 12M \frac{3}{\pi} \frac{M^2}{R^4} G \int_0^1 \frac{(\frac{1}{3}\xi^3 - \frac{1}{4}\xi^4)(1-\xi)}{\xi^2} d\xi = \frac{5}{4\pi} G \frac{M^2}{R^4}. \quad (346)$$

4.2 From (50) we have

$$P = \frac{1}{3}nmv^2 = \frac{2}{3}n\langle E \rangle, \quad (347)$$

where n is the particle density. For a non-relativistic degenerate electron gas we have

$$P = \frac{2}{3} \frac{\rho}{\mu_e m_p} \langle E \rangle = \frac{1}{20} \left(\frac{3}{\pi} \right)^{2/3} \frac{h^2}{m_e} \left(\frac{\rho}{\mu_e m_p} \right)^{5/3}, \quad (348)$$

where ρ is the mass density. Equating the two relations and solving for $\langle E \rangle$ gives

$$\langle E \rangle = \frac{3}{40} \left(\frac{3}{\pi} \right)^{2/3} \frac{h^2}{m_e} \left(\frac{\rho}{\mu_e m_p} \right)^{2/3}. \quad (349)$$

Using the numerical value of the density of Sirius B we obtain: $\langle E \rangle = 155.27$ keV. This corresponds to a Lorentz factor $\gamma = 1.30$ and $\beta = 0.64$. The electrons are thus mildly relativistic. The non-relativistic approximation agrees with the full relativistic result to an accuracy of 20% (note that the derivation of the equation of state uses the electron momentum). For larger densities the non-relativistic equation of state is surely not appropriate and we need to use the relativistic one.

4.3 By setting $P_c = P$ we obtain

$$\alpha \frac{G M^2}{\pi R^4} = \frac{1}{8} \left(\frac{3}{\pi} \right)^{1/3} hc \left(\frac{\rho}{\mu_e m_p} \right)^{4/3} = \frac{1}{8} \left(\frac{3}{\pi} \right)^{1/3} hc \left(\frac{3}{4\pi\mu_e m_p} \right)^{4/3} \frac{M^{4/3}}{R^4}, \quad (350)$$

with $\alpha = 3/8$ and $5/4$ for constant and linear density, respectively. Solving for M yields

$$M_{\text{Ch}} = \alpha^{-3/2} \frac{9}{256\pi} \sqrt{\frac{3}{2}} \left(\frac{hc}{G} \right)^{3/2} \left(\frac{1}{\mu_e m_p} \right)^2. \quad (351)$$

This evaluates numerically to $M_{\text{Ch}}^{\text{const}} = 0.44M_{\odot}$ and $M_{\text{Ch}}^{\text{linear}} = 0.07M_{\odot}$. For a constant density, the value is about a factor of 3 smaller than the exact result.

4.4 A type Ia supernova explosion is 10 billion times more luminous than the Sun (for a few days). Using the result of exercise 2.9 we write $D = d\sqrt{(L_{\text{Ia}} b_{\odot})}/(L_{\odot} b_{\text{th}}) = 6.32 \times 10^{15}$ AU = 3.07 Gpc = 10^{11} ly, where b_{th} is the limiting brightness for detection with HST.

4.5 (i) The radius of the blast wave can be read off the figures taking into account the height of the tower. Note that the shock wave is not at the border of the fireball, but at the end of the compression layer that is growing with time (seen as a faint layer in e.g. the figure at 0.053 s). Using a ruler we estimate the numbers given in Table V, with an estimated precision of 6 m (corresponding to the 1/16th inch sub-division of the ruler). (ii) The Sedov-Taylor expansion is described by (75). In a log-log plot this corresponds to a line $\log_{10} r = a + b \log_{10} t$, with $a = 1/5 \log_{10}(E/\rho_1)$ and $b = 2/5$. The fitted slope, $b = 0.42$, is indeed close to the value expected for the Sedov-Taylor phase of $2/5 = 0.4$ (least-square fitting of a power law with error bars yields 0.40 ± 0.02 , i.e. perfect compatibility within one standard deviation). The best value of a is the one minimizing the squared distances to the data points:

$$\frac{d}{da} \left\{ \sum_{i=1}^N [\log_{10}(r_i) - a - b \log_{10}(t_i)] \right\}^2 = -2 \sum_{i=1}^N [\log_{10}(r_i) - a - b \log_{10}(t_i)] \quad (352)$$

leading to

$$a = \frac{1}{N} \left[\sum \log_{10}(r_i) - b \sum \log_{10}(t_i) \right] = 2.77 \quad (353)$$

The results of the fit are shown in Fig. 42. Solving a for E yields $E = 10^{5a} \rho_1$. The density of air at 1,100 m above sea level is about 1.1 kg/m^3 and 1 ton TNT equivalent is $4.184 \times 10^9 \text{ J}$. Therefore, $E = 7.78 \times 10^{13} \text{ J} = 18,595 \text{ ton TNT}$. The official yield estimate of Trinity is 16,800 ton TNT < E < 23,700 ton TNT [28]. Note that the Sedov-Taylor

TABLE V: Expansion of the shock front as a function of time.

time (s)	shock radius (m)
0.006	75
0.016	108
0.025	138
0.053	200
0.062	206
0.090	218

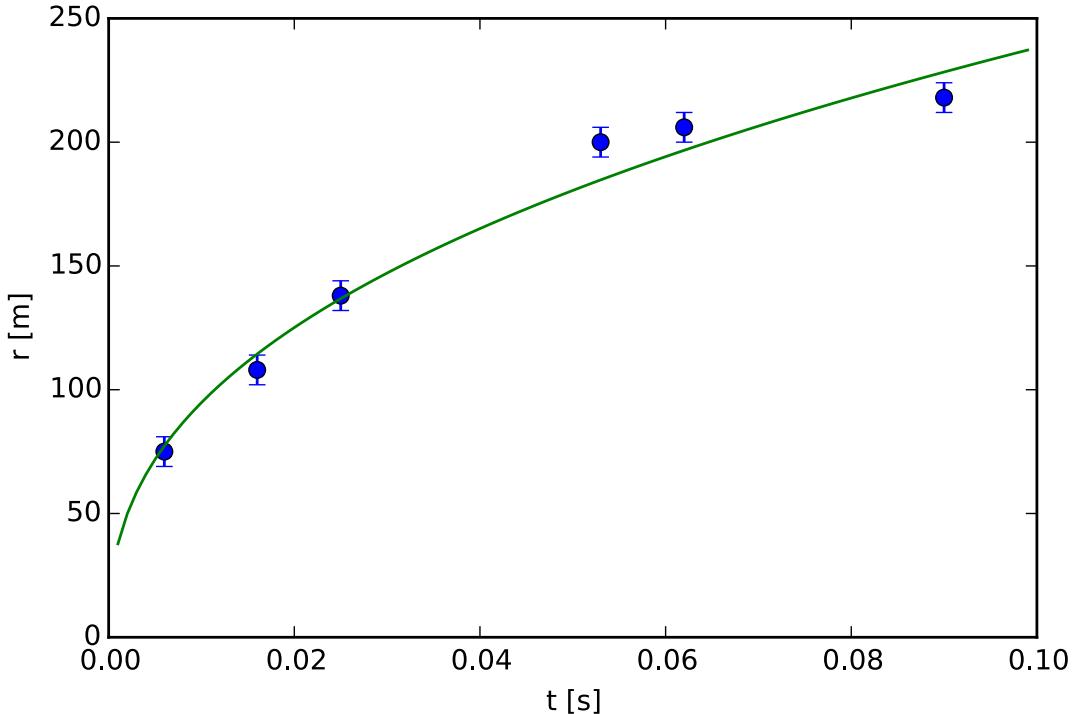


FIG. 42: Expansion of the shock front as a function of time [287].

expression was derived on the basis of a spherical explosion, whereas in this case the blastwave expands hemispherically. Following the original reasoning by Taylor [28], we have implicitly assumed “...that it may be justifiable to assume that most of the energy associated with the part of the blast wave which strikes the ground is absorbed there.”

5.1 (i) Differentiating $\vec{\sigma}(u, v)$ with respect to u and v yields

$$\frac{\partial \vec{\sigma}}{\partial u}(u, v) \equiv \vec{\sigma}_u(u, v) = \begin{pmatrix} -\sin u \sin v \\ \cos u \sin v \\ 0 \end{pmatrix} \quad \text{and} \quad \frac{\partial \vec{\sigma}}{\partial v}(u, v) \equiv \vec{\sigma}_v(u, v) = \begin{pmatrix} \cos u \cos v \\ \sin u \cos v \\ -\sin v \end{pmatrix}. \quad (354)$$

The coefficients of the first fundamental form may be found by taking the dot product of the partial derivatives

$$E = \vec{\sigma}_u \cdot \vec{\sigma}_u = \sin^2 v, \quad F = \vec{\sigma}_u \cdot \vec{\sigma}_v = 0, \quad G = \vec{\sigma}_v \cdot \vec{\sigma}_v = 1. \quad (355)$$

The line element may be expressed in terms of the coefficients of the first fundamental form as $ds^2 = \sin^2 v du^2 + dv^2$. (ii) The surface area is given by

$$A = \int_0^\pi \int_0^{2\pi} \sqrt{EG - F^2} du dv = \int_0^\pi \int_0^{2\pi} \sin v du dv = 2\pi(-\cos v) \Big|_0^\pi = 4\pi. \quad (356)$$

The coefficients of the second fundamental form are

$$e = \vec{\sigma}_{uu} \cdot \hat{n} = \sin^2 v, \quad f = \vec{\sigma}_{uv} \cdot \hat{n} = 0, \quad g = \vec{\sigma}_{vv} \cdot \hat{n} = 1. \quad (357)$$

(iii) The Gaussian curvature is

$$K = \frac{\det \text{II}}{\det \text{I}} = \frac{eg - f^2}{EG - F^2} = 1. \quad (358)$$

5.2 (i) The coefficients of the first fundamental form are

$$E = \tanh^2 u, \quad F = 0, \quad G = \operatorname{sech}^2 u. \quad (359)$$

The line element is $ds^2 = \tanh^2 u du^2 + \operatorname{sech}^2 u dv^2$. (ii) The surface area is

$$A = 2 \int_0^{2\pi} \int_0^\infty \operatorname{sech} u \tanh u du dv = 4\pi \quad (360)$$

which is exactly that of the sphere. (iii) The coefficients of the second fundamental form are

$$e = -\operatorname{sech} u \tanh u, \quad f = 0, \quad g = \operatorname{sech} u \tanh u. \quad (361)$$

The Gaussian curvature is $K = -1$.

5.3 We have

$$\frac{d}{dt} \|\dot{\gamma}\|^2 = \frac{d}{dt} (\dot{\gamma} \cdot \dot{\gamma}) = 2\ddot{\gamma} \cdot \dot{\gamma}. \quad (362)$$

Since γ is geodesic, $\ddot{\gamma}$ is perpendicular to the tangent plane which contains $\dot{\gamma}$. Hence, $\ddot{\gamma} \cdot \dot{\gamma} = 0$. Subsequently, $d\|\dot{\gamma}\|^2/dt = 0$. Therefore, the speed $\|\dot{\gamma}\|$ is constant.

5.4 The tangent plane is spanned by $\vec{\sigma}_u$ and $\vec{\sigma}_v$. By definition the curve γ is a geodesic if $\ddot{\gamma} \cdot \vec{\sigma}_u = \ddot{\gamma} \cdot \vec{\sigma}_v = 0$. Since $\dot{\gamma} = \dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v$, it follows that $\ddot{\gamma} \cdot \vec{\sigma}_u = 0$ becomes

$$\left[\frac{d}{dt} (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \right] \cdot \vec{\sigma}_u = 0. \quad (363)$$

We rewrite the left hand side of the above equation:

$$\begin{aligned} \left[\frac{d}{dt} (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \right] \cdot \vec{\sigma}_u &= \frac{d}{dt} [(\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \cdot \vec{\sigma}_u] - (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \cdot \frac{d\vec{\sigma}_u}{dt} \\ &= \frac{d}{dt} (E\dot{u} + F\dot{v}) - (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \cdot (\dot{u}\vec{\sigma}_{uu} + \dot{v}\vec{\sigma}_{uv}) \\ &= \frac{d}{dt} (E\dot{u} + F\dot{v}) - [\dot{u}^2 (\vec{\sigma}_u \cdot \vec{\sigma}_{uu}) + \dot{u}\dot{v} (\vec{\sigma}_u \cdot \vec{\sigma}_{uv} + \vec{\sigma}_v \cdot \vec{\sigma}_{uu}) + \dot{v}^2 (\vec{\sigma}_v \cdot \vec{\sigma}_{uv})]. \end{aligned} \quad (364)$$

We have that

$$\vec{\sigma}_u \cdot \vec{\sigma}_{uu} = \frac{1}{2} \frac{\partial}{\partial u} (\vec{\sigma}_u \cdot \vec{\sigma}_u) = \frac{1}{2} E_u, \quad \vec{\sigma}_v \cdot \vec{\sigma}_{uv} = \frac{1}{2} G_u, \quad \vec{\sigma}_u \cdot \vec{\sigma}_{uv} + \vec{\sigma}_v \cdot \vec{\sigma}_{uu} = F_u. \quad (365)$$

Substituting them into (364), we obtain

$$\left[\frac{d}{dt} (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \right] \cdot \vec{\sigma}_u = \frac{d}{dt} (E\dot{u} + F\dot{v}) - \frac{1}{2} (E_u \dot{u}^2 + 2F_u \dot{u}\dot{v} + G_u \dot{v}^2). \quad (366)$$

This establishes the first differential equation (90). Similarly, (91) can be established from

$$\left[\frac{d}{dt} (\dot{u}\vec{\sigma}_u + \dot{v}\vec{\sigma}_v) \right] \cdot \vec{\sigma}_v = 0. \quad (367)$$

5.5 For the parametrization in (92) the first fundamental form is found to be $ds^2 = d\theta^2 + \cos^2 \theta d\phi^2$, with $E = 1$, $F = 0$, and $G = \cos^2 \theta$. We restrict to unit-speed curves $\gamma(t) = \vec{\sigma}(\theta(t), \phi(t))$, so that

$$E\dot{\theta}^2 + 2F\dot{\theta}\dot{\phi} + G\dot{\phi}^2 = \dot{\theta}^2 + \dot{\phi}^2 \cos^2 \theta = 1. \quad (368)$$

If γ is a geodesic, then (91) is satisfied. Here (91) reduces to

$$\frac{d}{dt}(\dot{\phi} \cos^2 \theta) = 0, \quad \text{or equivalently} \quad \dot{\phi} \cos^2 \theta = \zeta, \quad (369)$$

where ζ is a constant. There are two cases: (i) $\zeta = 0$; then $\dot{\phi} = 0$. In this case, ϕ is constant and γ is part of a meridian. (ii) $\zeta \neq 0$. Substituting (369) into the unit-speed condition (368), we have

$$\dot{\theta}^2 = 1 - \frac{\zeta^2}{\cos^2 \theta}. \quad (370)$$

Combining the above with (369), along the geodesic it holds that

$$\left(\frac{d\phi}{d\theta} \right)^2 = \frac{\dot{\phi}^2}{\dot{\theta}^2} = \frac{1}{\cos^2 \theta (\cos^2 \theta / \zeta^2 - 1)}. \quad (371)$$

Integrate the derivative $d\phi/d\theta$:

$$\phi - \phi_0 = \pm \int \frac{d\theta}{\cos \theta \sqrt{\cos^2 \theta / \zeta^2 - 1}}, \quad (372)$$

where ϕ_0 is a constant. The substitution $u = \tan \theta$ yields

$$\phi - \phi_0 = \pm \int \frac{du}{\sqrt{\zeta^{-2} - 1 - u^2}} = \sin^{-1} \left(\frac{u}{\sqrt{\zeta^{-2} - 1}} \right), \quad (373)$$

which leads to

$$\tan \theta = \pm \sin(\phi - \phi_0) \sqrt{\zeta^{-2} - 1}. \quad (374)$$

Multiply both sides of the above equation by $\cos \theta$:

$$\sin \theta = \pm \sqrt{\zeta^{-2} - 1} (\cos \phi_0 \cos \theta \sin \phi - \sin \phi_0 \cos \theta \cos \phi). \quad (375)$$

Since $\vec{\sigma}(\theta, \phi) = (xy, z)$, we have

$$z = \mp (\sin \phi_0 \sqrt{\zeta^{-2} - 1}) x \pm (\cos \phi_0 \sqrt{\zeta^{-2} - 1}) y. \quad (376)$$

Clearly, $z = 0$ when $x = y = 0$. Therefore, γ is contained in the intersection of S^2 with a plane through the center of the sphere. Hence it is part of a great circle.

5.6 (109) gives the rate at which a clock at radius r ticks relative to one infinitely far away. Here we are asked to compare the rate of a clock at radius r relative to one at the radius of the Earth (i.e., at the Earth's surface). We can think about this by considering the rate of each of these clocks relative to a distant clock. That is, the clock on the Earth's surface has a rate a factor

$$\sqrt{1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{R_{\oplus}}} \quad (377)$$

slower than the distant clock, while the clock at radius r has a rate

$$\sqrt{1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{r}}. \quad (378)$$

Note that both these expressions are less than one, but because $r > R_{\oplus}$, the stationary clock at radius r ticks faster than that at the Earth's surface. Indeed, the relative rate of the two is just the ratio of these two expressions

$$\sqrt{\left(1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{r}\right) \left(1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{R_{\oplus}}\right)^{-1}}. \quad (379)$$

Again, the expression in (379) is greater than 1. (ii) Circular motion at speed v at a radius r gives rise to an acceleration v^2/r , which we know is due to gravity. Thus if the astronaut has a mass m , Newton's second law yields

$$\frac{GM_{\oplus}m}{r^2} = m \frac{v^2}{r}, \quad \text{or solving for } v \text{ gives } v = \sqrt{\frac{GM_{\oplus}}{r}}. \quad (380)$$

The time dilation in special relativity is due to the by now familiar factor $(1 - v^2/c^2)^{1/2}$, which gives

$$\sqrt{1 - \frac{GM_{\oplus}}{rc^2}}. \quad (381)$$

Note again how similar this looks to the expression above for time dilation due to gravity. Again, this is the rate that an orbiting clock at radius r ticks relative to a stationary clock at the same radius. (iii) In part (379), we calculated the ratio of rates of stationary clocks at radius r and R_{\oplus} (due to general relativity), while in part (381), we calculated the ratio of the rates of an orbiting clock at radius r to a stationary clock at the same radius. Therefore, the ratio of the rate of an orbiting clock at radius r to a stationary one on the ground is simply the product of these two results; namely,

$$\sqrt{\left(1 - \frac{2GM_{\oplus}}{rc^2}\right) \left(1 - \frac{2GM_{\oplus}}{R_{\oplus}c^2}\right)^{-1} \left(1 - \frac{GM_{\oplus}}{rc^2}\right)}. \quad (382)$$

(iv) We now simplify (382). We will do this in pieces, starting from (379) we can write:

$$\sqrt{\left(1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{r}\right) \left(1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{R_{\oplus}}\right)^{-1}} \approx \sqrt{\left(1 - \frac{1}{c^2} \frac{2GM_{\oplus}}{r}\right) \left(1 + \frac{1}{c^2} \frac{2GM_{\oplus}}{R_{\oplus}} + \dots\right)}. \quad (383)$$

We then use $(1 - x)(1 - y) \approx 1 - (x + y)$ to re-write (383) as

$$\sqrt{1 - \frac{GM_{\oplus}}{c^2} \left(\frac{2}{r} - \frac{2}{R_{\oplus}}\right)}. \quad (384)$$

This then gets multiplied by (381), yielding

$$\sqrt{1 - \frac{GM_{\oplus}}{c^2} \left(\frac{3}{r} - \frac{2}{R_{\oplus}}\right)} \approx 1 - \frac{GM_{\oplus}}{2c^2} \left(\frac{3}{r} - \frac{2}{R_{\oplus}}\right). \quad (385)$$

However, we do need to justify the use of the various approximations we have made. We dealt with a variety of expressions of the form $1 - x$; in every case x is of the form $(GM_{\oplus})/(rc^2)$. The smallest r we considered, and therefore the largest the expression $(GM_{\oplus})/(rc^2)$, is $r = R_{\oplus}$. So we plug in numbers at $r = R_{\oplus}$ to obtain

$$\frac{GM_{\oplus}}{R_{\oplus}c^2} = \frac{2}{3} \frac{10^{-10} \text{ m}^3 \text{ s}^{-2} \text{ kg}^{-1} \times 6 \times 10^{24} \text{ kg}}{6.4 \times 10^6 \text{ m} \times (3 \times 10^8 \text{ m/s})^2} \approx 7 \times 10^{-9}, \quad (386)$$

which is indeed a number much much smaller than 1. (v) We are asked to find the radius at which (385) is equal to unity. This clearly holds when $3/r - 2/R_{\oplus} = 0$, or $r = 1.5R_{\oplus}$. Given the radius of the Earth is 6,400 km, the critical radius $r = 1.5R_{\oplus}$ is at a distance of 9,600 km from the Earth's center, or 3,200 km above the Earth's surface. Now, (385) is less than 1 for $r < 1.5R_{\oplus}$, and so astronauts on the space shuttle age less than those staying home.

5.7 (i) The Schwarzschild radius of a black hole of mass M is $R_{\text{Sch}} = 2GM/c^2$. The volume of a sphere of this radius is just the familiar $4\pi R_{\text{Sch}}^3$. The density is the mass divided by the volume, giving:

$$\rho_{\text{BH}} = M \times \left[\frac{4}{3}\pi \left(\frac{2GM}{c^2} \right)^3 \right]^{-1} = \frac{3c^6}{32\pi G^3 M^2}. \quad (387)$$

The more massive the black hole, the smaller the density. Thus there is a mass at which the black hole has the density of paper, which is what we are trying to figure out. (ii) The density is the mass per unit volume. If we can figure out the volume of a square meter of paper (whose mass we know, 75 g), we can calculate its density. The volume of a piece of paper is its area times its thickness. The thickness is 0.1 mm, or 10^{-4} m, and so the volume of a square meter of paper is 10^{-4} m³. Therefore, the density of paper is

$$\rho = \frac{7.5 \times 10^{-2} \text{ kg}}{10^{-4} \text{ m}^3} = 7.5 \times 10^2 \text{ kg/m}^3, \quad (388)$$

similar to (but slightly less than) the density of water (remember, paper is made of wood, and wood floats in water). (iii) Here we equate (387) with (388) and solve for the mass

$$M = \sqrt{\frac{3c^6}{32\pi G^3\rho}} \approx 3 \times 10^{38} \text{ kg}, \quad (389)$$

where we have made all the usual approximations of $\pi = 3$, $32 = 10$, and so on. We need to express this in solar masses, so we divide by $M_\odot = 2 \times 10^{30}$ kg to obtain $M \approx 1.5 \times 10^8 M_\odot$. A black hole 150 million times the mass of the Sun has the same density as a piece of paper. We know the Hitchhiker's Guide to the Galaxy is science fiction, but do such incredibly massive black holes actually exist? Indeed they do: the cores of massive galaxies (including our own Milky Way) do contain such enormous black holes. Actually, the most massive such black hole known to exist is in the core of a particularly luminous galaxy known as M87, with a mass of 3 billion solar masses. We still have to calculate the Schwarzschild radius of a black hole. We could plug into the formula for a Schwarzschild radius and calculate away, but here we outline a simpler approach. We know the Schwarzschild radius is proportional to the mass of a black hole, and we happen to remember that a M_\odot mass black hole has a Schwarzschild radius of 3 km, (106). So a 150 million M_\odot black hole has a Schwarzschild radius 150 million times larger, or 4.5×10^8 km. We are asked to express this in terms of AU; 1 AU = 1.5×10^8 km, so the Schwarzschild radius of such a black hole is 3 AU. (iv) We know the entire mass of the black hole. If we can calculate the mass of a single piece of paper, the ratio of the two gives the total number of pages. So we now turn to calculate the mass of a single piece of paper. We know that a square meter of paper has a mass of 75 g. How many square meters is a standard-size sheet? One inch is 2.5 cm = 2.5×10^{-2} m. So $8.5 \times 11 \text{ inch}^2 \approx 100 \text{ inch}^2 \approx 6 \times 10^{-2} \text{ m}^{-2}$. Thus the mass is

$$\text{Mass of a piece of paper} = 7.5 \times 10^{-2} \text{ kg/m}^2 \times 6 \times 10^{-2} \text{ m}^2 \approx 5 \times 10^{-3} \text{ kg}. \quad (390)$$

That is, a piece of paper weighs about 5 g. We divide this into the mass we calculated above:

$$\text{Number of sheets of paper} = \frac{\text{Mass of rule book}}{\text{Mass per page}} = \frac{3 \times 10^{38} \text{ kg}}{5 \times 10^{-3} \text{ kg/page}} = 6 \times 10^{40} \text{ pages}. \quad (391)$$

Strictly speaking, if the rule book is printed on both sides of the page, we should multiply the above result by a factor of two. That is one seriously long set of rules! Finally, note that because the density of a more massive black hole is smaller (387), the mass and number of pages of the Brockian Ultra Cricket rule book as given by (391) is really just a lower limit. That is, if the rule book were even larger than what we have just calculated, it would still collapse into a black hole.

5.8 (i) The Schwarzschild radius of a $3M_\odot$ black hole is $R_{\text{Sch}} = 2GM/c^2 = 9$ km. If you remember that for $1 M_\odot$ black hole the Schwarzschild radius is 3 km, you can scale from there. (ii) Using the Newton's law of gravitation, we can write the difference in gravitational forces acting on two bodies of mass m which are located at distances r_1 and r_2 from the massive body of mass M

$$\delta F \equiv F_1 - F_2 = \frac{GmM}{r_1^2} - \frac{GmM}{r_2^2} = GmM \left(\frac{1}{r_1^2} - \frac{1}{r_2^2} \right). \quad (392)$$

We are interested in the difference in gravitational forces in two locations that are close to each other, since the height of the person falling into the black hole is small compared to the Schwarzschild radius. We take $r_2 = r_1 + \Delta$, where $\Delta \ll r_1$. Now, we simplify (392); dropping the subscript 1 we have

$$\delta F = GmM \left[\frac{1}{r^2} - \frac{1}{(r + \Delta)^2} \right] = GmM \frac{(r + \Delta)^2 - r^2}{r^2(r + \Delta)^2} = GmM \frac{r^2 + 2r\Delta + \Delta^2 - r^2}{r^2(r + \Delta)^2} \approx GmM \frac{2\Delta}{r^3}. \quad (393)$$

In obtaining this expression, we used the approximations $r + \Delta = r(1 + \Delta/r) \approx r$ and $2r\Delta + \Delta^2 = 2r\Delta[1 + \Delta/(2r)] \approx 2r\Delta$, because $\Delta \ll r$. Next, we use (393) to find the distance r_{crit} from the black hole where the relative stretching force between your head and your legs is equal to some critical force

$$\delta F_{\text{crit}} = GmM \frac{2\Delta}{r_{\text{crit}}^3} \quad \text{and so} \quad r_{\text{crit}} = \left(\frac{2GmM\Delta}{\delta F_{\text{crit}}} \right)^{1/3}. \quad (394)$$

Finally, we can plug in the numbers. The mass of the black hole is $M = 3M_{\oplus} = 6 \times 10^{30}$ kg. The mass of the body is $m = 70$ kg, $\delta F_{\text{crit}} = 10$ kN. The critical radius is then, $r_{\text{crit}} \approx 2000$ km or recalling that the Schwarzschild radius for $3M_{\odot}$ black hole is $R_{\text{Sch}} \sim 10$ km, we have $r_{\text{crit}} \sim 200R_{\text{Sch}}$. Note, that in this case a significant amount of stretching occurs already relatively far from the black hole. (iii) The force with which the metal plate is pulling on you is given by the Newton's second law, $F = m_{\text{sp}}g$, where m_{sp} is the mass of the steel plate, and $g \sim 10$ m/s² is the gravitational acceleration on the Earth. Thus $m_{\text{sp}} = 10$ kN/(10 m/s²) = 1,000 kg, or 1 ton. If you still have hard time imagining how much weight 10 kN is, this is the weight of a typical car. So, imagine attaching a car to your feet: not pleasant. Most likely this is enough to kill or at the very least severely disable a person. (iv) We can apply the formula for r_{crit} from (394). Now the mass of the black hole is 1.3×10^6 times larger, so the radius increases by $(1.3 \times 10^6)^{1/3} \approx 100$ times. The answer is then $r_{\text{crit}} = 2 \times 10^5$ km. In terms of the Schwarzschild radii, remember that R_{Sch} is linearly proportional to the mass. For 4 million solar mass black hole, the Schwarzschild radius is then $1.3 \times 10^6 \times 10$ km ≈ 107 km and so $r_{\text{crit}} \approx 2 \times 10^{-2} R_{\text{Sch}}$. Since $r_{\text{crit}} < R_{\text{Sch}}$, the "spaghettification" happens inside the Schwarzschild radius. (v) As we saw in (iv), the radius at which the tidal force reaches 10 kN grows as the third root of the mass of the black hole, but the Schwarzschild radius grows linearly with the mass of the hole. In part (iii) for the $3M_{\odot}$ hole the critical radius was outside R_{Sch} , while in part (iv) for $4 \times 10^6 M_{\odot}$ hole the critical radius was inside. Thus, there should be a minimum mass of the hole, at which $r_{\text{crit}} = R_{\text{Sch}}$, i.e., we can just barely pass through the horizon before getting fatally stretched. We find this mass setting $r_{\text{crit}} = R_{\text{Sch}}$, which leads to

$$\left(\frac{2GmM_{\min}\Delta}{\delta F_{\text{crit}}} \right)^{1/3} = \frac{2GM_{\min}}{c^2}, \quad (395)$$

and so

$$M_{\min} = \left(\frac{c^3}{2G} \right) \left(\frac{m\Delta}{\delta F_{\text{crit}}} \right)^{1/2} = 2 \times 10^{34} \text{ kg} = 10^4 M_{\odot}. \quad (396)$$

So, if you fall into a $10^4 M_{\odot}$ black hole, you will be killed right as you go through the horizon. If the black hole is more massive, then you can go through the horizon while still alive, and enjoy the sights! Sadly, you will not have much time to enjoy the view anyways, because you will be crushed by the singularity in 0.01 s seconds for this $10^4 M_{\odot}$ black hole. This time is proportional to mass of the hole.

5.9 (i) From Fig. 10 estimate an initial position of about 32 ly at a time 2002.12 and 45 ly at 2002.73. The apparent velocity is therefore $v_{\text{app}} \approx 13$ ly/0.61 yr = 21 ly/yr = $21c$, which is in agreement with the value $v_{\text{app}} = (25.6 \pm 4.4)c$ given in [69] reporting this measurement. The apparent velocity of the blob is thus highly superluminal. (ii) The light emitted at point A at time $t_{i,1}$ will reach the observer located at a distance d_1 at time $t_1 = d_1/c$; see Fig. 43. The blob travels with "true" velocity v from A to B a distance H which takes a time

$$\Delta t_{A \rightarrow B} = t_{i,2} - t_{i,1} = \frac{1}{v} \frac{L}{\sin \theta}, \quad (397)$$

where the only hypothesis is that the signal travels at the speed of light c . The remaining distance to the observer is $d_2 = d_1 - L/\tan \theta$ and therefore the light from position B will arrive at

$$t_2 = \frac{1}{v} \frac{L}{\sin \theta} + \frac{1}{c} \left(d_1 - \frac{L}{\tan \theta} \right). \quad (398)$$

The time difference between the signals from A and B is

$$\Delta t = t_2 - t_1 = \frac{1}{v} \frac{L}{\sin \theta} + \frac{1}{c} \frac{L}{\tan \theta} = L \left(\frac{c - v \cos \theta}{vc \sin \theta} \right) = L \left(\frac{1 - \beta \cos \theta}{v \sin \theta} \right), \quad (399)$$

where $\beta = v/c$, and the apparent transverse velocity is therefore

$$\beta_{\text{app}} = \frac{1}{c} \frac{L}{\Delta t} = \frac{\beta \sin \theta}{1 - \beta \cos \theta}. \quad (400)$$

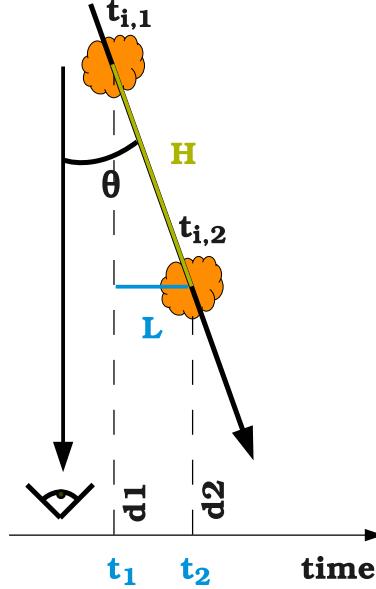


FIG. 43: The situation in exercise 5.9.

(iii) Using $x = \tan \theta/2$, (400) can be re-written with standard trigonometry as $\beta_{\text{app}} = 2\beta x/[(1 - \beta) + (1 + \beta)x^2]$ so that after a bit of algebra it follows that

$$v_{\text{app}} = kc \Leftrightarrow [(1 + \beta)x - \beta/k]^2 = (\beta/k - \gamma^{-1})(\beta/k + \gamma^{-1}), \quad (401)$$

where $\gamma = (1 - \beta^2)^{-1/2}$ is the Lorentz factor. The left hand term is positive and the equation in x admits at least one solution if $\beta/k \geq \gamma^{-1}$. Superluminal motion $v_{\text{app}} \geq c$ (i.e. $k \geq 1$) is then possible as long as $\gamma\beta \geq \beta_{\text{app}}$. A direct consequence of the previous equation is $\gamma \geq \beta_{\text{app}}$, which proves that, even for moderate superluminal motions, the true velocity is relativistic. The angle that maximizes the apparent transverse velocity can be found by differentiating v_{app} and solving for θ_{\max} , we have

$$\frac{dv_{\text{app}}}{d\theta} = \frac{\beta \cos \theta}{1 - \beta \cos \theta} - \frac{\beta^2 \sin^2 \theta}{(1 - \beta \cos \theta)^2} = 0 \Rightarrow \cos \theta_{\max} = \beta. \quad (402)$$

The maximum apparent transverse velocity is therefore

$$\beta_{\text{app}}^{\max} = \frac{\beta \sin \theta_{\max}}{1 - \beta \cos \theta_{\max}} = \frac{\beta \sqrt{1 - \cos^2 \theta_{\max}}}{1 - \beta \cos \theta_{\max}} = \frac{\beta \sqrt{1 - \beta^2}}{1 - \beta^2} = \beta\gamma. \quad (403)$$

Therefore,

$$\beta_{\text{app}}^{\max} = \beta\gamma = \gamma \sqrt{1 - 1/\gamma^2} \Rightarrow \gamma = \sqrt{1 + \beta_{\text{app}}^2}, \quad (404)$$

that is to say the plasma blob moves with a highly relativistic Lorentz factor of at least ~ 21 .

6.1 (i) Imagine a circle with radius x around the observer. A fraction $s(x)$, $0 \leq s(x) \leq 1$, is covered by trees. Then we'll move a distance dx outward, and draw another circle. There are $2\pi n x dx$ trees growing in the annulus limited by these two circles. They hide a distance $2\pi x n D dx$, or a fraction $n D dx$ of the perimeter of the circle. Since a fraction $s(x)$ was already hidden, the contribution is only $[1 - s(x)]n D dx$. We get

$$s(x + dx) = s(x) + [1 - s(x)] n D dx, \quad (405)$$

which gives a differential equation for s :

$$\frac{ds(x)}{dx} = [1 - s(x)] n D. \quad (406)$$

This is a separable equation which can be integrated:

$$\int_0^s \frac{ds}{1-s} = \int_0^x n D dx . \quad (407)$$

This yields the solution

$$s(x) = 1 - e^{-nDx} . \quad (408)$$

This is the probability that in a random direction we can see at most to a distance x . This function s is a cumulative probability distribution. It is as if we have compressed the 2-dimensional forest into an imaginary 1-dimensional structure, with a characteristic mean free path. The corresponding probability density is its derivative ds/dx . The mean free path λ is the expectation of this distribution

$$\lambda = \int_0^\infty x \frac{ds(x)}{dx} dx = \frac{1}{nD} . \quad (409)$$

For example, if there are 2000 trees per hectare, and each trunk is 10 cm thick, we can see to a distance of 50 m, on average. (ii) The result can be easily generalized into 3 dimensions. Assume there are n stars per unit volume, and each has a diameter D and a surface $A = \pi D^2$ perpendicular to the line of sight. Then we have

$$s(x) = 1 - e^{-nAx} , \quad (410)$$

where $\lambda = (nA)^{-1}$. For example, if there were one sun per cubic parsec, the mean free path would be 1.6×10^4 pc. If the universe were infinite old and infinite in size, the line of sight would eventually meet a stellar surface in any direction, although we could see very far indeed.

6.2 (i) The relation between luminosity, distance, and brightness is given by the inverse-square law, namely

$$\text{brightness} = \frac{\text{luminosity}}{4\pi \text{distance}^2} . \quad (411)$$

Here we are given the luminosity of each galaxy (the four are the same, namely 4×10^{37} J/s), and the brightness, in units of Joules/meters²/second. Solving for the distance gives

$$\text{distance} = \left(\frac{\text{luminosity}}{4\pi \text{brightness}} \right)^{1/2} , \quad (412)$$

and so we find: galaxy #1, distance = 6.5×10^{24} m = 210 Mpc; galaxy #2, distance = 8.4×10^{24} m = 270 Mpc; galaxy #3, distance = 1.1×10^{25} m = 360 Mpc; galaxy #4, distance = 1.5×10^{25} m = 490 Mpc. (ii) The redshift is given by $z = (\lambda - \lambda_0)/\lambda_0$, where $\lambda_0 = 3935 \text{ \AA}$ and 3970 \AA for the two calcium lines. The measured wavelengths for each of the two lines in each of the galaxies, the corresponding redshift from each of the lines, and the average redshift are given in Table VI. (iii) The redshift is equal to the velocity of recession divided by the speed of light. So we can calculate the velocity of recession as the redshift times the speed of light. The Hubble constant is given in Table VII. The four galaxies give consistent values of the Hubble constant, at about 60 km/s/Mpc. Not identical to the modern value of 70 km/s/Mpc, but close. That seemed quite straightforward; so why is there such controversy over the exact value of the Hubble constant? The difficult point is getting an independent measurement of the luminosity of each galaxy. The problem stated that each of the galaxies has the same luminosity of the Milky Way. That is only approximately true; the numbers were adjusted somewhat to make this come out with a reasonable value for H_0 .

6.3 (i) In the “local Universe” approximation where we pretend that cosmological redshifts are Doppler shifts and it is a good approximation to pretend that galaxies which are getting more distant from us due to the expansion of space are flying away from us at a given velocity, we can use the Hubble’s law. Then for the closer galaxy, we get $H_0 = 580 \text{ km/s}/35 \text{ Mly} = 17 \text{ km s}^{-1} \text{ Mly}^{-1}$. For the farther galaxy, we get $H_0 = 25,400 \text{ km/s}/1,100 \text{ Mly} = 23 \text{ km s}^{-1} \text{ Mly}^{-1}$. (ii) The calculation from the more distant galaxy. Reason: peculiar velocities are always a few hundred km/s. They are random, so they could be anywhere from minus a few hundred to plus a few hundred. Potentially, this could be a large fraction of the 580 km/s of the closer galaxy. However, it will be a small fraction of the 25,400 km/s of the more distant galaxy. It is noteworthy that closer galaxies tend to be brighter and therefore the measurements are less likely to suffer from observational errors. While this is true, the exercise presents numbers to (at least approximately) equivalent significant figures in both cases. It

TABLE VI: Redshifts of the four galaxies in exercise 6.2

Galaxy	λ (Å) first line	λ (Å) second line	z first line	z second line	z average
1	4100	4135	0.042	0.041	0.042
2	4145	4185	0.053	0.054	0.053
3	4215	4255	0.071	0.072	0.071
4	4318	4360	0.097	0.098	0.098

TABLE VII: Determination of the Hubble constant.

Galaxy	Redshift	Velocity (km/s)	Distance (Mpc)	H_0 (km/s/Mpc)
1	0.042	12600	210	60
2	0.053	15900	270	59
3	0.071	21300	360	59
4	0.098	29400	490	59

may well have taken a lot more telescope time and effort to get the numbers on the more distant galaxy, but you have them. The peculiar velocity issue, however, is an intrinsic effect that perfect observations cannot get around. We will always have to deal with galaxies moving about in the universe even if we have amazing data. (iii) We use the value of H_0 derived from the distant galaxy to calculate the receding v for the closer galaxy, $v = H_0 d = 805$ km/s, which implies the peculiar velocity of the nearby galaxy is $v_{\text{pec}} = -220$ km/s, that is 220 km/s toward us. (iv) Assuming that 220 km/s of the 25,400 km/s we observed for the more distant galaxy were due to peculiar velocity, from (119) we have $H_0 = 25,620$ km/s/1,100 Mly = 23.3 km/s/Mly. Note that 1,100 Mly only has two significant figures and so the difference is smaller than the precision of our measurement. This is a specific illustration of why, given that all galaxies will tend to have peculiar velocities of a few hundred km/s, more distant galaxies give you a more reliable estimate of the Hubble constant.

6.4 The Hubble flow $v = H_0 r$ induces the flux vn through the surface $4\pi r^2$ of a sphere with radius r , and thus $\dot{N} = 4\pi r^2 vn$. These particles escape from the sphere containing the total number of particles $N = Vn$. Hence $\dot{N} = -4\pi r^2 vn = 4\pi r^3 \dot{n}/3$, or $\dot{n} = -3vn/r = -3H_0 n$.

6.5 At the final time of the invasion, t , the invaders are at proper distance d_p , and therefore comoving distance $r = d_p/a(t)$; see (166). (i) For a flat space, the proper volume is obviously the usual one in Euclidean geometry,

$$V = \frac{4\pi}{3} d_p^3. \quad (413)$$

(ii) In a closed universe, and if R is the comoving radius of curvature, the proper area of a sphere at comoving coordinate r is $4\pi a^2(t) R^2 \sin^2(r/R)$, and the proper distance between two spheres at r and $r + dr$ is just $a(t)dr$, as obtained from the FRW metric. Therefore, the proper volume of each spherical shell between r and $r + dr$ is $4\pi a^3(t) R^2 \sin^2(r/R) dr$, and the proper volume of the sphere is

$$V = 4\pi a^3(t) R^2 \int_0^r \sin^2(r/R) dr = 4\pi a^3(t) R^3 \int_0^{r/R} \sin^2(r/R) d(r/R) = 4\pi a^3 R^3 \left[\frac{d_p}{2aR} - \frac{\sin(2d_p/aR)}{4} \right]. \quad (414)$$

(iii) In an open universe, the calculation is just like for the closed universe but with the substitution $\sin(r/R)$ by $\sinh(r/R)$,

$$V = 4\pi a^3(t) R^3 \int_0^{r/R} \sinh^2(r/R) d(r/R) = 4\pi a^3 R^3 \left(\frac{\sinh(2d_p/aR)}{4} - \frac{d_p}{2aR} \right). \quad (415)$$

6.6 For the case when the universe contains only matter with negligible pressure, the energy density changes as $\rho_m(t) = \rho_{m,0}/a^3(t)$. Multiplying (133) by $a^2(t)$, we have

$$(\dot{a})^2 = \frac{8\pi G \rho_{m,0}}{3c^2 a} - \frac{c^2}{R_0^2}. \quad (416)$$

Now, using $\dot{a} = (da/dt) = (da/d\theta)(d\theta/dt)$, we find the left-hand-side of (416) is

$$(\dot{a})^2 = \frac{c^2}{R_0^2} \frac{\sin^2 \theta}{(1 - \cos \theta)^2} = \frac{c^2}{R_0^2} \frac{1 + \cos \theta}{1 - \cos \theta}, \quad (417)$$

where the last equality follows from $\sin^2 \theta = 1 - \cos^2 \theta = (1 - \cos \theta)(1 + \cos \theta)$, and the right-hand-side of (416) is

$$\frac{8\pi G \rho_{m,0}}{3c^2 a} - \frac{c^2}{R_0^2} = \frac{c^2}{R_0^2} \left(\frac{2}{1 - \cos \theta} - 1 \right) = \frac{c^2}{R_0^2} \frac{1 + \cos \theta}{1 - \cos \theta}. \quad (418)$$

So the two sides of (416) are indeed equal, confirming that this parametric solution given as $a(\theta)$ and $t(\theta)$ is indeed a solution of Friedmann's equation. (ii) The maximum value of a occurs at $\theta = \pi$, and is

$$a_{\max} = \frac{8\pi G \rho_{m,0} R_0^2}{3c^4}. \quad (419)$$

(iii) Correspondingly, the maximum value of the proper radius of curvature is

$$a_{\max} R_0 = \frac{8\pi G \rho_{m,0} R_0^3}{3c^4}. \quad (420)$$

(iv) The age of the universe at $\theta = \pi$ is

$$t_{\max} = \frac{4\pi^2 G \rho_{m,0} R_0^3}{3c^5}. \quad (421)$$

(v) The *big crunch* happens when $\theta = 2\pi$, and we then have

$$t_{\text{crunch}} = \frac{8\pi G \rho_{m,0} R_0^3}{3c^5}. \quad (422)$$

6.7 Multiplying (133) by $a^2(t)$, we have

$$(\dot{a})^2 = \frac{8\pi G \rho_{m,0}}{3c^2 a} + \frac{c^2}{R_0^2}. \quad (423)$$

Now, using the relations

$$\sin(ix) = \frac{e^{i(ix)} - e^{-i(ix)}}{2i} = \frac{e^{-x} - e^x}{2i} = -\frac{e^x - e^{-x}}{2i} = -\frac{1}{i} \sinh x = i \sinh x \quad (424)$$

$$\cos(ix) = \frac{e^{i(ix)} + e^{-i(ix)}}{2} = \frac{e^{-x} + e^x}{2} = \cosh x, \quad (425)$$

$$\cosh^2 - \sinh^2 x = \cos^2(ix) - [\sin(ix)/i]^2 = \cos^2(ix) + \sin^2(ix) = 1 \quad (426)$$

we can rewrite (417) as

$$(\dot{a})^2 = \frac{c^2}{R_0^2} \frac{\sinh^2 \theta}{(\cosh \theta - 1)^2} = \frac{c^2}{R_0^2} \frac{\cosh \theta + 1}{\cosh \theta - 1}, \quad (427)$$

and the right-hand-side of (423) is

$$\frac{8\pi G \rho_{m,0}}{3c^2 a} + \frac{c^2}{R_0^2} = \frac{c^2}{R_0^2} \left(\frac{2}{\cosh \theta - 1} + 1 \right) = \frac{c^2}{R_0^2} \frac{\cosh \theta + 1}{\cosh \theta - 1}. \quad (428)$$

The two sides of (416) are indeed equal, confirming that this parametric solution given as $a(\theta)$ and $t(\theta)$ is indeed a solution of Friedmann's equation. (ii) A comparison of the scale factors in (160), (168), and (169) corresponding to

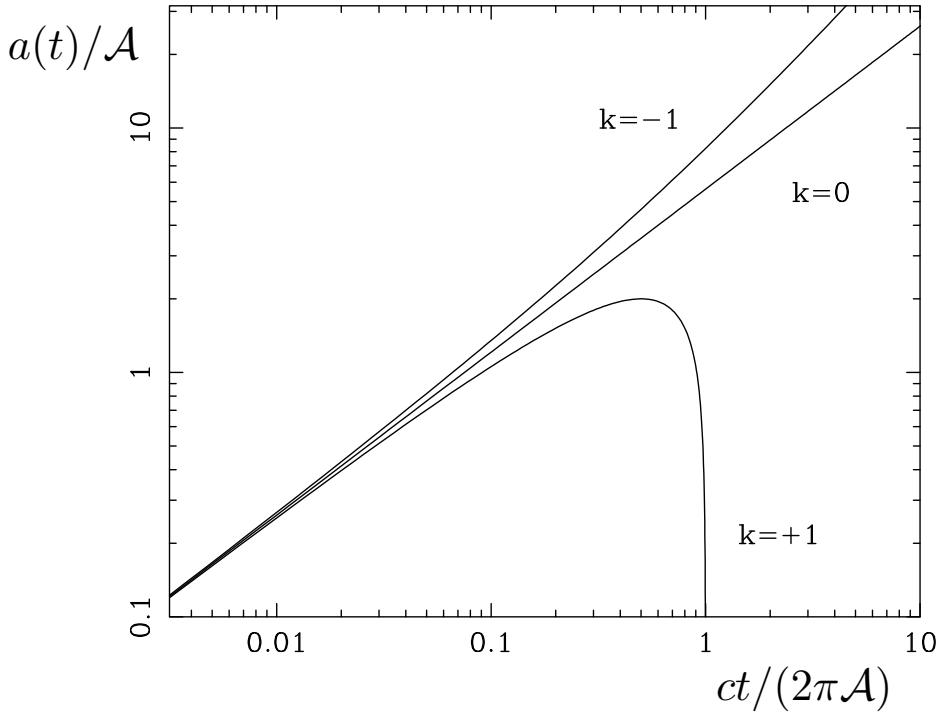


FIG. 44: The time dependence of the scale factor for open, closed and critical matter-dominated cosmological models. The upper line corresponds to $k = -1$, the middle line to the flat $k = 0$ model, and the lowest line to the recollapsing closed $k = +1$ universe. The log scale is designed to bring out the early-time behaviour, although it obscures the fact that the closed model is a symmetric cycloid on a linear plot of a against t . We have set $\mathcal{A} = 4\pi G\rho_{m,0}R_0^3/(3c^2)$.

solutions with $k = 0$, $k = 1$, and $k = -1$, respectively is exhibited in Fig. 44.

6.8 (i) For the model with $\Omega_{m,0} = 1$, the comoving distance is

$$r = c \int_0^z \frac{dz}{H(z)} = \frac{c}{H_0} \int_0^z \frac{dz}{(1+z)^{3/2}} = \frac{2c}{H_0} \left(1 - \frac{1}{\sqrt{1+z}} \right). \quad (429)$$

The comoving distance to the horizon, at $a = 0$ or $z = \infty$, is $r = 2c/H_0$. (ii) For this model, half the comoving distance to the horizon is $r = c/H_0$, and the redshift at which the comoving distance has this value is obtained as:

$$1 - \frac{1}{\sqrt{1+z}} = \frac{1}{2} \Rightarrow z = 3. \quad (430)$$

(iii) For this same model, and at $z = 3$, the age of the universe is obtained from

$$t(z) = \int_z^\infty \frac{dz}{(1+z) H(z)} = \frac{2}{3} \frac{1}{H_0} \frac{1}{(1+z)^{3/2}}. \quad (431)$$

The present age of the universe is of course $t_0 = 2/(3H_0)$, and so the ratio of the age at $z = 3$ to its present age is just

$$\frac{t(z=3)}{t_0} = \frac{1}{(1+z)^{3/2}} = \frac{1}{8}. \quad (432)$$

(iv) From the same equation as above, we find

$$\frac{t(z)}{t_0} = \frac{1}{(1+z)^{3/2}} = \frac{1}{2} \Rightarrow z = 2^{2/3} - 1 = 0.5874. \quad (433)$$

Note that all these equations are of course valid only for the specific model that is flat and contains only matter, with $\Omega_{m,0} = 1$.

7.1 The flux of one of these objects is $\mathcal{F} = L/(4\pi d_L^2)$, and its angular size is $\theta = \ell/d_A$. Hence, the apparent surface brightness is $I \propto \mathcal{F}/\theta^2$, or

$$I = \text{constant} \times \frac{\mathcal{F}}{\theta^2} = \text{constant} \times \frac{d_A^2}{d_L^2} = \text{constant} \times (1+z)^{-4}, \quad (434)$$

where in the last equality we have used (201). Note that L , ℓ , and 4π are constants and so can be absorbed in the constant of proportionality. Therefore, the apparent surface brightness I will always decrease with redshift as $(1+z)^{-4}$ compared to the *intrinsic* surface brightness B , without any dependence on the cosmological model.

7.2 (i) The number density of photons per unit frequency is equal to the energy density per unit frequency divided by $h\nu$, or

$$n_\nu d\nu = \frac{8\pi\nu^2 d\nu}{c^3 \{ \exp[h\nu/(kT)] - 1 \}}. \quad (435)$$

The total number density is found by integrating over frequency, which gives

$$n = \frac{8\pi}{c^3} \int \frac{\nu^2 d\nu}{\exp[h\nu/(kT)] - 1}. \quad (436)$$

Substituting $x = h\nu/(kT)$, we find

$$n = \frac{8\pi}{c^3} \left(\frac{kT}{h} \right)^3 \int \frac{x^2 dx}{e^x - 1}. \quad (437)$$

For $T_0 = 2.725$, we find 410.4 cm^{-3} . (ii) The current density of baryons must then be

$$n_b = 5.5 \times 10^{-10} 410.4 \text{ cm}^{-3} = 2.25 \times 10^{-7} \text{ cm}^{-3}. \quad (438)$$

(iii) Every baryon weighs approximately like the mass of a proton (this is not exact because, for example, the helium nuclei weigh a little less than 4 protons because of the helium nucleus binding energy, but the difference is rather small). So the density of baryons is $n_b m_p = 3.78 \times 10^{-31} \text{ g/cm}^3$. The critical density is $3H_0^2/(8\pi G) = 9.2 \times 10^{-30} \text{ g/cm}^3$, so $\Omega_b = 0.041$.

7.3 To analyze the measurement of our own galaxy through the CMB, it is useful to consider the density $N_\gamma(\vec{p})$ of photons in phase space, defined by specifying that there are $N_\gamma(\vec{p}) d^3 p$ photons of each polarization (right or left circularly polarized) per unit spatial volume in a momentum-space volume $d^3 p$ centered at \vec{p} . Since $|\vec{p}| = h\nu/c$ and $4\pi h^3 \nu^2 d\nu/c^3$ is the momentum-space volume between frequencies ν and $d\nu$, (435) gives

$$N_\gamma(\vec{p}) = \frac{1}{2} \frac{n_T(c|\vec{p}|)/h}{4\pi h^3 \nu^2/c^3} = \frac{1}{h^3} \frac{1}{\exp[|\vec{p}|c/(kT)] - 1}, \quad (439)$$

where n_T is the number density of photons in equilibrium with matter at temperature T at photon frequency between ν and $\nu + d\nu$, and the factor $1/2$ takes account of the fact that n_T includes both possible polarization states. This is of course the density that would be measured by an observer at rest in the radiation background. The phase space volume is Lorentz invariant, and the number of photons is also Lorentz invariant, so N_γ is a scalar, in the sense that a Lorentz transformation to a coordinate system moving with respect to the radiation background that takes \vec{p} to \vec{p}' also takes N_γ to N'_γ , where

$$N'_\gamma(\vec{p}') = N_\gamma(\vec{p}). \quad (440)$$

If the Earth is moving in the x -direction with a velocity (in units of c) of β , and we take \vec{p} to be the photon momentum in the frame at rest in the CMB and \vec{p}' to be the photon momentum measured on Earth, then from (19) it follows that

$$|\vec{p}'| = \gamma(1 - \beta \cos \theta)|\vec{p}| \quad (441)$$

where θ is the angle between p and the x -axis. Thus

$$N'_\gamma(\vec{p}') = \frac{1}{h^3} \frac{1}{\exp[|\vec{p}'|c/(kT')] - 1}, \quad (442)$$

where the temperature is a function of the angle between the direction of the photon and the Earth's velocity

$$T = T' \gamma (1 - \beta \cos \theta). \quad (443)$$

This means that the temperature $T(\theta)$ observed in the direction θ , is given in terms of the average temperature T_0 by

$$\begin{aligned} T(\theta) &= T_0 \frac{\sqrt{1 - \beta^2}}{1 - \beta \cos \theta} = T_0 (1 - \beta^2)^{1/2} (1 - \beta \cos \theta)^{-1} \approx T_0 (1 - \beta^2/2 + \dots) (1 + \beta \cos \theta + \beta^2 \cos^2 \theta + \dots) \\ &\approx T_0 [1 + \beta \cos \theta + \beta^2 (\cos^2 \theta - 1/2) + \dots]. \end{aligned} \quad (444)$$

Using the trigonometric relation $\cos^2 x = (1 + \cos 2x)/2$ we obtain (215). The motion of the observer (us) gives rise to both a dipole and other, higher order corrections. The observed dipole anisotropy implies that [97]

$$\vec{v}_\odot - \vec{v}_{\text{CMB}} = 370 \pm 10 \text{ km/s} \quad \text{towards } \phi = 267.7 \pm 0.8^\circ, \quad \theta = 48.2 \pm 0.5^\circ, \quad (445)$$

where θ is the colatitude (polar angle) and it is in the range $0 \leq \theta \leq \pi$ and ϕ is the longitude (azimuth) and it is in the range $0 \leq \phi \leq 2\pi$. Allowing for the Sun's motion in the Galaxy and the motion of the Galaxy within the Local Group, this implies that the Local Group is moving with

$$\vec{v}_{\text{LG}} - \vec{v}_{\text{CMB}} \approx 600 \text{ km/s} \quad \text{towards } \phi = 268, \quad \theta = 27^\circ. \quad (446)$$

This "peculiar" motion is subtracted from the measured CMB radiation, after which the intrinsic anisotropy is isolated (Fig. 23), and revealed to be about few parts in 10^5 . Even though minuscule, these primordial perturbations provided seeds for the structure of the Universe.

7.4 To be written.

7.5 (i) The circumference of a circle of radius a is $2\pi a$, so the orbital speed is the circumference divided by period:

$$v = \frac{2\pi a}{T} = \frac{2\pi a}{\sqrt{4\pi^2 a^3/(GM)}} = \sqrt{\frac{GM}{a}}. \quad (447)$$

(ii) According to the Birkoff's theorem, the orbit about a mass distributed within a sphere is the same as if the mass is all concentrated in the center of the sphere. So, we can use the velocity formula derived in (i), and invert it to obtain the mass enclosed by an orbit:

$$M(a) = \frac{a}{G} v^2. \quad (448)$$

The enclosed mass at 8 kpc is then $M(8 \text{ kpc}) \approx 2 \times 10^{41} \text{ kg} = 10^{11} M_\odot$. (iii) If the mass enclosed by the orbit stays at $10^{11} M_\odot$ as the radius increases, which follows from the fact that the Sun is at the edge of the luminous galaxy, then at different radii the velocity given by (448) will decrease with square root of the distance. At 30 kpc, it is $v \approx 110 \text{ km/s}$. At 100 kpc, we have $v \approx 60 \text{ km/s}$. (iv) Let's look again at (448), which says that if the orbital velocity stays the same, the mass enclosed will increase linearly with a as the radius of the orbit grows. We already calculated the mass enclosed by 8 kpc orbit in part (ii). So, at 30 kpc, the mass enclosed will be 30 kpc/8 kpc times larger, or $3.8 \times 10^{11} M_\odot$. At 100 kpc, the mass enclosed will be 100 kpc/8 kpc times larger, or $1.3 \times 10^{12} M_\odot$. (v) We see that the mass of the gravitating matter is increasing linearly with radius, and exceeds by more than factor of 10 the mass of the luminous matter (e.g. stars and gas). We thus infer that the outer halo of the galaxy is dominated by invisible dark matter.

7.6(i) The total mass inside R is obtained from $GM(R)/R = v_c^2(R)$. The answer can of course be found by substituting for the value of G and everything else in your favorite system of units, and if you are lucky not to make any mistake you may even get the right answer. Often, it is faster and safer to work it out using proportionality comparing to an example that you know and love. What could this example be but the Earth moving around the Sun? For the Earth, with M_\odot and an orbit of 1 AU the velocity is 30 kms^{-1} (if you did not know how fast the Earth moves around the Sun, this is a good number to remember). So, the mass inside radius R is

$$M(R) = 8 \times 10^{10} M_\odot \quad (449)$$

(ii) If the density at R is ρ_0 , then the density at any other radius r is $\rho_0(R/r)^2$, so the mass inside R is

$$M(R) = 4\pi \int_0^R dr r^2 \rho_0 \left(\frac{R}{r}\right)^2 = 4\pi \rho_0 R^3. \quad (450)$$

Hence the density at R is

$$\rho_0 = \frac{M(R)}{4\pi R^3} = 0.51 m_p \text{ cm}^{-3}. \quad (451)$$

The result is most easily computed remembering that the solar mass contains 1.19×10^{57} proton masses (another useful number to remember), and a parsec is 3.086×10^{16} m. (iii) The density is

$$\rho_\Lambda = \Omega_\Lambda \frac{3H_0^2}{8\pi G} = 5.5 \times 10^{-6} \Omega_\Lambda m_p \text{ cm}^{-3} = 4 \times 10^{-6} m_p \text{ cm}^{-3}. \quad (452)$$

(iv) Because the dark energy is spread out uniformly, whereas the dark matter and baryonic matter are highly concentrated in the inner parts of galaxies, the density of dark energy is very small compared to the density of matter inside the radius of the solar orbit in the Milky Way. The dark energy therefore must have a tiny dynamical effect.

7.7 The relation between the emitted T_{em} and the observed T_{obs} is

$$T_{\text{em}} = T_{\text{obs}}(1+z) = \frac{2.9 \times 10^{-3} \text{ mK}}{\lambda_{\text{max}}^{\text{obs}}} (1+z), \quad (453)$$

where in the last equality we used Wien's displacement law [13]. For $\lambda_{\text{max}}^{\text{obs}} = 180 \mu\text{m}$ and $z = 2$, we have $T_{\text{em}} \approx 48 \text{ K}$. If we did not account for redshift, we would have thought the galaxy was only at 16 K.

7.8 In the benchmark model, at the present moment, the ratio of the vacuum energy density to the energy density in matter is

$$\frac{\rho_\Lambda}{\rho_{m,0}} = \frac{\Omega_\Lambda}{\Omega_{m,0}} \approx 2.3. \quad (454)$$

In the past, however, when the scale factor was smaller, the ratio of densities was

$$\frac{\rho_\Lambda}{\rho_m(a)} = \frac{\rho_\Lambda}{\rho_{m,0}/a^3} = \frac{\rho_\Lambda}{\rho_{m,0}} a^3. \quad (455)$$

If the universe has been expanding from an initial very dense state, at some moment in the past, the energy density of matter and Λ must have been equal. This moment of matter- Λ equality occurred when

$$a_{m\Lambda}^3 = \frac{\Omega_{m,0}}{\Omega_\Lambda} = \frac{\Omega_{m,0}}{1 - \Omega_{m,0}} \approx 0.43. \quad (456)$$

where we have used the normalization $a_0 = 1$ for the present. Next, we generalize (180) to write the age of the universe at any redshift z , for a flat model with matter and a cosmological constant,

$$t(z) = \frac{1}{H_0} \int_z^\infty \frac{dz}{(1+z)\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_\Lambda}} = \frac{1}{H_0\sqrt{\Omega_\Lambda}} \int_z^\infty \frac{dz}{(1+z)\sqrt{1 + (\Omega_{m,0}/\Omega_\Lambda)(1+z)^3}}. \quad (457)$$

With the change of variables $y = \sqrt{1 + (\Omega_{m,0}/\Omega_\Lambda)(1+z)^3}$ we find $2ydy = 3(y^2 - 1)dz/(1+z)$ and so

$$t = \frac{2}{3H_0\sqrt{\Omega_\Lambda}} \int_y^\infty \frac{dy}{y^2 - 1}. \quad (458)$$

The integral can be solved analytically as:

$$-\int \frac{dy}{y^2 - 1} = \frac{1}{2} \ln \left[\frac{y+1}{y-1} \right], \quad (459)$$

which yields for t

$$\begin{aligned} t &= \frac{2}{3H_0\sqrt{\Omega_\Lambda}} \ln \left(\frac{1}{\sqrt{y^2 - 1}} + \frac{y}{\sqrt{y^2 - 1}} \right) \Big|_{\infty}^{\sqrt{1+\Omega_{m,0}(1+z)^3}/\Omega_\Lambda} \\ &= \frac{2}{3H_0\sqrt{\Omega_\Lambda}} \ln \left(\sqrt{\frac{\Omega_\Lambda}{\Omega_{m,0}(1+z)^3}} + \sqrt{1 + \frac{\Omega_\Lambda}{\Omega_{m,0}(1+z)^3}} \right). \end{aligned} \quad (460)$$

Using (178) and (456) we can rewrite (460) as

$$H_0 t = \frac{2}{3\sqrt{1-\Omega_{m,0}}} \ln \left[(a/a_{m\Lambda})^{3/2} + \sqrt{1+(a/a_{m\Lambda})^3} \right] = \frac{2}{3\sqrt{1-\Omega_{m,0}}} \ln A, \quad (461)$$

and so

$$H_0 t_0 = \frac{2}{3\sqrt{1-\Omega_{m,0}}} \ln A_0, \quad (462)$$

where we have defined

$$A_0 = a_{m\Lambda}^{-3/2} + \sqrt{1+a_{m\Lambda}^{-3}} \simeq 3.35. \quad (463)$$

Now, we want to find the value of a for which $t = t_0/2$. This implies

$$\ln A = \frac{1}{2} \ln A_0, \quad (464)$$

where

$$A = x + \sqrt{1+x^2} \quad \text{with} \quad x = \left(\frac{a}{a_{m\Lambda}} \right)^{3/2}. \quad (465)$$

(464) implies $A = \sqrt{A_0}$ and so

$$x + \sqrt{1+x^2} = \sqrt{A_0} \quad \Rightarrow \quad 1+x^2 = (\sqrt{A_0}-x)^2 \quad \Rightarrow \quad x = \frac{A_0-1}{2\sqrt{A_0}} \simeq 0.64, \quad (466)$$

yielding

$$a = a_{m\Lambda} x^{2/3} = 0.56 \quad \text{and} \quad z = \frac{1}{a} - 1 = 0.78. \quad (467)$$

We see that the redshift at which the age of the universe was half the present age is larger in this benchmark model than in the model with $\Omega_{m,0} = 1$, see (433). This is because in the benchmark model, which contains vacuum energy, the universe has started to accelerate recently, roughly since the epoch at $a = a_{m\Lambda}$. The universe took a longer time to expand to $a = 0.56$ and then it picked up speed again in its expansion up to the present $a_0 = 1$.

8.1 We have seen in (257) that $s_\gamma \propto T^3$, so that $S_\gamma \propto VT^3$. For a reversible adiabatic expansion, the entropy of the (non-interacting) CMB photons remains unchanged. Hence, when V doubles, T will decrease by a factor $(2)^{-1/3}$. So after 10^{10} yr the average temperature of the blackbody will become $\langle T \rangle = 2.2$ K.

8.2 Since during inflation the Hubble rate is constant

$$\Omega - 1 = \frac{kc^2}{a^2 H^2 R_0^2} \propto a^{-2}. \quad (468)$$

On the other hand, (265) suggests that to reproduce today's observed value $\Omega_0 - 1 \sim 1$ the initial value at the beginning of the radiation-dominated phase must be $|\Omega - 1| \sim 10^{-54}$. Since we identify the beginning of the radiation-dominated phase with the end of inflation we require

$$|\Omega - 1|_{t=t_f} \sim 10^{-54}. \quad (469)$$

During inflation

$$\frac{|\Omega - 1|_{t=t_f}}{|\Omega - 1|_{t=t_i}} = \left(\frac{a_i}{a_f} \right)^2 = e^{-2H\Delta t}. \quad (470)$$

Taking $|\Omega - 1|_{t=t_i}$ of order unity, it is enough to require that $\Delta t \gtrsim 60/H$ to solve the flatness problem. Thus, inflation ameliorates the fine-tuning problem, by explaining a tiny number $\mathcal{O}(10^{-54})$ with a number $\mathcal{O}(60)$.

8.3 The effective number of neutrinos and antineutrinos is $g_{\nu_L} = 6$ and the temperature of the cosmic neutrino background is $T_\nu = 0.7 T_\gamma \approx 1.9$ K. Now, from (244) we have

$$n_\nu = \frac{3\zeta(3)}{4\pi^2} g_{\nu_L} \left(\frac{kT_\nu}{\hbar c}\right)^3 \approx 45.63 \left(\frac{T_\nu}{K}\right)^3 \approx 313 \text{ neutrinos/cm}^3. \quad (471)$$

If neutrinos saturate the dark matter density the upper bound on the neutrino mass is then

$$m_\nu < \frac{0.26 \rho_c}{n_\nu} \approx \frac{2.6 \times 10^{-27} \text{ kg/m}^3}{3.13 \times 10^8 \text{ neutrino/m}^3} \approx 10^{-35} \frac{\text{kg}}{\text{neutrino}} \times \frac{9.38 \times 10^8 \text{ eV}/c^2}{1.67 \times 10^{-27} \text{ kg}} \sim 5.6 \text{ eV}/c^2. \quad (472)$$

where we have used $m_p = 938 \text{ MeV}/c^2 = 1.67 \times 10^{-27} \text{ kg}$ to obtain the result in natural units.

8.4 If we change the difference between the proton and neutron mass to $\Delta m = 0.129 \text{ MeV}$ while all other parameters remain the same, then the time of freeze-out of the neutron abundance occurs at the same temperature $T_{n/N}^{\text{FO}} = 0.75 \text{ MeV}$. Therefore, the neutron abundance freezes out at $n_n/n_p = e^{-0.1293/0.75} = 0.84$. If there were no neutron decays and all neutrons combined to form helium, the maximum primordial ${}^4\text{He}$ abundance would then be

$$Y_{\text{p}}^{\text{max}} = \frac{2n_n}{n_n + n_p} = 0.91. \quad (473)$$

Note that neutrons would in fact not decay, they would be stable because the difference with the mass of the proton would be less than the mass of the electron. It would be rather unfortunate if the neutron had a mass so close to the proton mass: almost all the matter in the universe would have turned to helium in the beginning of the universe, and main-sequence stars would not live very long with the very small amount of hydrogen they would have left. The Sun would live for less than 1 billion years and the planet Earth would not have had enough time to sustain life on it for us to be here now.

8.5 (i) From (257), the energy density of photons at the time of BBN was

$$\rho_{\gamma,\text{BBN}} = 0.66 \frac{(kT_{\text{BBN}})^4}{(\hbar c)^3} = 7.56 \times 10^{20} \text{ J/m}^3, \quad (474)$$

where $T_{\text{BBN}} \approx 10^9 \text{ K}$. Note that in reality we should also account for the neutrinos, but Gamow did not know much about the three families of neutrinos and their interactions. A better estimate of the energy density at BBN goes as follows. The effective number of neutrinos and antineutrinos is 6, or 3 times the effective number of species of photons. On the other hand, the fourth power of the neutrinos temperature is less than the fourth power of the photon temperature by a factor of $3^{-4/3}$. Thus the ratio of the energy density of neutrinos and antineutrinos to that of photons is

$$\rho_\nu/\rho_\gamma = 3^{-4/3} \cdot 3 = 0.7. \quad (475)$$

Hence the total energy density after electron positron annihilation is

$$\rho_{\text{BBN}} \simeq \rho_{\nu,\text{BBN}} + \rho_{\gamma,\text{BBN}} = 1.7\rho_{\gamma,\text{BBN}} \simeq 1.3 \times 10^{21} \text{ J/m}^3. \quad (476)$$

(ii) Since the universe was radiation dominated, the critical density at BBN had to be equal to this radiation density, so

$$\frac{3c^2 H^2}{8\pi G} = \rho_{\text{BBN}} \quad (477)$$

This gives a Hubble parameter at the time of BBN in Gamow's radiation dominated universe of $H = 2.17 \times 10^{-3} \text{ s}^{-1}$. (iii) Readjusting (162), the time for BBN is found to be

$$t_{\text{BBN,G}} = \frac{1}{2H} = 231 \text{ s}. \quad (478)$$

(iv) For a present age $t_0 \approx 10^{10} \text{ yr}$, the temperature is given by

$$\frac{3c^2 H_0^2}{8\pi G} = \frac{3c^2}{32\pi G t_0^2} = 0.66 \frac{(kT_{0,\text{G}})^4}{(\hbar c)^3} \quad (479)$$

which gives $T_{0,G} = 27$ K. Note that actually this temperature just depends on t_0 and the assumption of a flat, radiation-dominated universe, but it does not depend on T_{BBN} . (v) If the universe changed from being radiation dominated to matter dominated at some redshift z_{eq} , then at the present time the matter density is greater than the radiation density by a factor $1 + z_{\text{eq}}$; so $\rho_{\text{rad}} = \rho_m/(1 + z_{\text{eq}})$. In a flat universe with *only matter and radiation*, the total density has to be equal to the critical density, therefore $\rho_{\text{rad}} + \rho_m = \rho_{\text{rad}}(2 + z_{\text{eq}}) = \rho_c$. So,

$$\rho_{\text{rad}} = \frac{3H_0^2}{8\pi G} \frac{1}{2 + z_{\text{eq}}} = 0.66 \frac{(kT_0)^4}{(\hbar c)^3}, \quad (480)$$

and the radiation temperature is smaller by a factor $(2 + z_{\text{eq}})^{-1/4}$.

8.6 The effective number of neutrino species contributing to r.d.o.f. can be written as

$$N_{\text{eff}} = 3 \left[1 + \left(\frac{T_{\nu_R}}{T_{\nu_L}} \right)^4 \right]. \quad (481)$$

Taking into account the isentropic heating of the rest of the plasma between ν_R decoupling temperature $T_{\nu_R}^{\text{dec}}$ and the end of the reheating phase,

$$\delta N_\nu = 3 \left(\frac{g(T_{\nu_L}^{\text{dec}})}{g(T_{\nu_R}^{\text{dec}})} \right)^{4/3}, \quad (482)$$

where $T_{\nu_L}^{\text{dec}}$ is the temperature at the end of the reheating phase (when left-handed neutrinos decouple), and we have taken $N_{\text{eff}} = 3 + \delta N_\nu$. To be consistent with Planck data at 1σ we require $N_{\text{eff}} < 3.68$. We take $g(T_{\nu_L}^{\text{dec}}) = 10.75$ reflecting $(e_L^- + e_R^+ + e_R^- + e_L^+ \nu_{eL} + \bar{\nu}_{eR} + \nu_{\mu L} + \bar{\nu}_{\mu R} + \nu_{\tau L} + \bar{\nu}_{\tau R} + \gamma_L + \gamma_R)$. From (482) the allowable range is $g(T_{\nu_R}^{\text{dec}}) > 33$. This is achieved for $r(T_{\nu_R}^{\text{dec}}) > 0.29$. Using (270) this can be translated into a decoupling temperature: $T_{\nu_R}^{\text{dec}} > 185$ MeV.

8.7 At a given time, the rate of decrease in the BH mass is just the total power radiated

$$\frac{d\dot{M}_{\text{BH}}}{dQ} = - \sum_i g_i \frac{\sigma_s}{8\pi^2} Q^3 \left[\exp \left(\frac{Q}{T_{\text{BH}}} \right) - (-1)^{2s} \right]^{-1}. \quad (483)$$

Integration of (483) leads to

$$\dot{M}_{\text{BH}} = - \sum_i g_i \mathcal{B}_\pm \frac{\Gamma_s}{8\pi^2} \Gamma(4) \zeta(4) T_{\text{BH}}^4 A_{4 \subset 4+n}. \quad (484)$$

The net change of the BH mass is therefore

$$\frac{dM_{\text{BH}}}{dt} = \frac{dM_{\text{BH}}}{dt} \Big|_{\text{accr}} + \frac{dM_{\text{BH}}}{dt} \Big|_{\text{evap}}. \quad (485)$$

Substituting $M_{\text{BH}} \sim \sqrt{s}$ into (484), where \sqrt{s} is the center-of-mass energy of the constituents of the protons (quarks and gluons), a rather lengthy but straightforward calculation shows that $dM/dt > 0 \Leftrightarrow \epsilon > 10^{10}$ GeV/fm³. Note that the energy desnity of partonic matter produce at the LHC is more than 7 orders of magnitude smaller. (ii) Since the ratio of degrees of freedom for gauge bosons, quarks and leptons is 29:72:18 (the Higgs boson is not included), from (484) we obtain a rough estimate of the mean lifetime,

$$\tau_{\text{BH}} \approx 1.67 \times 10^{-27} \left(\frac{M_{\text{BH}}}{M_*} \right)^{9/7} \left(\frac{\text{TeV}}{M_*} \right) \text{s}. \quad (486)$$

then (486) indicates that black holes that could be produced at the LHC would evaporate instantaneously into visible quanta. For further thoughts on this subject [288, 289].

9.1 (i) For a steady state, $\partial n/\partial t = 0$ and

$$\frac{\partial^2 n}{\partial z^2} = - \frac{Q_0}{D} \delta(z). \quad (487)$$

Integration yields

$$\frac{\partial n}{\partial z} = A - \frac{Q_0}{D} \Theta(z), \quad (488)$$

where A is an integration constant and $\Theta(z)$ the Heaviside step function (see Appendix E). A second integration leads to

$$n(z) = B + Az - \frac{Q_0}{D} z \Theta(z), \quad (489)$$

where B is an integration constant. From $n(-H) = 0$ we can conclude that $B = AH$ and so

$$n(z) = AH + Az - \frac{Q_0}{D} z \Theta(z). \quad (490)$$

On the other hand, $n(+H) = 0$ yields

$$2AH - \frac{Q_0}{D} H = 0 \quad (491)$$

or

$$A = \frac{1}{2} \frac{Q_0}{D}. \quad (492)$$

Then the particle density in the range $-H \leq z \leq H$ is

$$n(z) = \frac{1}{2} \frac{Q_0}{D} (H + z) - \frac{Q_0}{D} z \Theta(z), \quad (493)$$

which can be rewritten as

$$n(z) = \frac{1}{2} \frac{Q_0}{D} (H - |z|) \quad (494)$$

(ii) The column density is

$$N = \int_{-H}^{+H} n(z) dz = 2 \int_0^H \frac{1}{2} \frac{Q_0}{D} (H - z) dz = \frac{Q_0 H^2}{2D}. \quad (495)$$

Using $N = Q_0 \tau_{\text{res}}$ we have

$$\tau_{\text{res}} = \frac{H^2}{2D} \Rightarrow D = \frac{H^2}{2\tau_{\text{res}}}. \quad (496)$$

Using $D\beta c\lambda/3$ the mean free path is

$$\lambda = \frac{3}{2} \frac{H^2}{\beta c \tau_{\text{res}}}. \quad (497)$$

For $H = 500$ pc, $\tau_{\text{res}} = 10^7$ yr and $\beta \sim 1$ the mean free path is about 0.1 pc.

9.2 (i) The equation of motion is

$$\frac{d\vec{p}}{dt} = \vec{F} = \frac{d}{dt}(\gamma m \vec{v}) = Ze(\vec{v} \times \vec{B}) \quad (498)$$

where e is the elementary charge and Z is the charge number. The acceleration in a magnetic field is always perpendicular to the velocity, $\vec{v} \perp \dot{\vec{v}} = \vec{a}$ and hence $\dot{\gamma} = 0$. Therefore

$$\gamma m \dot{\vec{v}} = Ze(\vec{v} \times \vec{B}). \quad (499)$$

For $\vec{v} \perp \vec{B}$ we can write down the component-wise differential equations which read as

$$\dot{v}_x = \frac{Ze}{\gamma m} v_y B \quad \text{and} \quad \dot{v}_y = -\frac{Ze}{\gamma m} v_x B. \quad (500)$$

The solution is

$$v_x = v \sin \left(\frac{ZeB}{\gamma m} t \right) \quad \text{and} \quad v_y = v \cos \left(\frac{ZeB}{\gamma m} t \right), \quad (501)$$

which leads to

$$x = -\frac{v\gamma m}{ZeB} \cos \left(\frac{ZeB}{\gamma m} t \right) \quad \text{and} \quad y = \frac{v\gamma m}{ZeB} \sin \left(\frac{ZeB}{\gamma m} t \right). \quad (502)$$

The radius is therefore

$$R = \sqrt{x^2 + y^2} = \frac{v\gamma m}{ZeB} \approx \frac{c\gamma m}{ZeB}, \quad (503)$$

where in the last step we set $v \approx c$. (ii) For a given radius R , the magnetic field strength can thus be expressed as

$$B = \frac{c\gamma m}{ZeR}. \quad (504)$$

For $R \simeq 27 \text{ km}/(2\pi)$ and $c\gamma m_p = E/c$ we can calculate the average magnetic field at the LHC, $B_{\text{LHC}} \approx 5.43 \text{ T}$. Note that in reality the particles in a collider are not in a uniform magnetic field, but the collider ring is composed of alternating sections for bending, accelerating and focussing the particles. Therefore the actual magnetic field strengths needed are slightly larger than the ones calculated above. At the LHC, the magnets produce a field of 8.7 Tesla. (iii) Useful formulae for the radius of a particle in a magnetic field can be obtained by introducing $E = \gamma mc^2$ and evaluating the numerical constants, which gives the rule of thumb for particle physics detectors

$$R = 3.3 \text{ m} \frac{E/(\text{GeV})}{Z(B/T)}, \quad (505)$$

and the rule of thumb for cosmic ray acceleration (sometimes called the *Hillas criterion* [290])

$$R = 1.1 \text{ kpc} \frac{(E/\text{GeV})}{Z(B/\mu\text{G})} \quad (506)$$

The radius of a collider, with the average magnetic field of the LHC, that is expected to launch particles to 10^{11} GeV would be $6 \times 10^{10} \text{ m}$. This radius is comparable to the Sun-Mercury distance, which is $5.76 \times 10^{10} \text{ m}$ (see exercise 2.6). Hence, such a collider would be priceless! (iv) The maximum attainable energies in the given astrophysical objects are: for neutron stars, $E_{\text{max}}^p \sim 10^{11} \text{ GeV}$ and $E_{\text{max}}^{^{56}\text{Fe}} \sim 2.6 \times 10^{12} \text{ GeV}$; for AGN jets, $E_{\text{max}}^p \sim 10^{10} \text{ GeV}$ and $E_{\text{max}}^{^{56}\text{Fe}} \sim 2.6 \times 10^{11} \text{ GeV}$; for supernova remnants $E_{\text{max}}^p \sim 10^7 \text{ GeV}$ and $E_{\text{max}}^{^{56}\text{Fe}} \sim 2.6 \times 10^8 \text{ GeV}$.

9.3 (i) We are told that the energy emitted by the supernova in visible light is equal to that emitted by the Sun in 10^{10} yr . We can look up the luminosity of the Sun (energy emitted per second), and simply multiply by the 10^{10} yr ,

$$\text{Total energy emitted in visible light} = 4 \times 10^{26} \text{ J/s} \times 10^{10} \text{ yr} \times \frac{3 \times 10^7 \text{ s}}{1 \text{ yr}} \approx 10^{44} \text{ J}. \quad (507)$$

The energy associated with the neutrinos is 100 times larger still than that, namely 10^{46} J . (ii) If each neutrino has an energy of $\langle E_\nu \rangle \sim 1.5 \times 10^{-12} \text{ J}$, the total number of neutrinos emitted by the star is $\sim 7 \times 10^{57}$. (iii) These neutrinos are emitted essentially all at once, and thereafter, travelling at the speed of light, they expand into a huge spherical shell of ever-increasing radius. Thus, by the time they impinge on the Earth, they are spread out over a spherical shell of radius 150,000 ly. The number density on the shell is:

$$\frac{\text{Number}}{\text{Surface Area}} = \frac{7 \times 10^{57} \text{ neutrinos}}{4} \pi (1.5 \times 10^5 \text{ ly} \times 10^{16} \text{ m/ly}) \approx 2.5 \times 10^{14} \text{ neutrinos/m}^2. \quad (508)$$

That is, every square meter on the Earth's surface was peppered with 250 trillion neutrinos from the supernova. The detector has 2.14 kton of water ($N_{\text{target}} \sim 1.28 \times 10^{33}$ free target nucleons) and so using the average cross section for weak interactions we have

$$\frac{1}{3} \frac{\text{Number}}{\text{Surface Area}} \sigma_{\text{weak}} N_{\text{target}} = \frac{1}{3} \cdot 2.5 \times 10^{14} \cdot 5 \times 10^{-48} \left(\frac{E_\nu}{\text{MeV}} \right)^2 1.28 \times 10^{33} \sim 46 \text{ electron neutrinos}. \quad (509)$$

When the discovery of the supernova was first announced, Bahcall, Dar, and Piran (BDP) immediately realized the possibility that Kamiokande could have detected the neutrinos from it. They locked themselves in their office, took the phone off the hook, did essentially the calculation that you have just done, and sent a paper off to Nature, all within 24 hours. They wanted to make a prediction about the neutrinos, untainted by any news that the neutrinos actually were found. Indeed, a few days later, the news of IMB [291] and Kamiokande [292] detection came out. The two detectors in deep mines recorded a total of 19 neutrino interactions over a span of 13 seconds. The BDP paper was published on 1987 March 12 (the supernova itself went off on February 23), and has the following understated but triumphal final sentence: “Note added in proof: Since this paper was received on 2 March, the neutrino burst was found by the Kamiokande experimental group, with properties generally consistent with the calculated expectations” [293]. Making a rough correction for the 60% efficiency reduces the expected number of events to 28, within about a factor of 2 of the actual detection.

9.4 Substituting M_\star and M_{BH} in (317) we obtain $P = 2.4 \times 10^{47}$ W. At this emission rate the neutron star will fall into the black hole in $t \approx 2.9$ s.

Appendix A: Geometry of radiation

The solid angle Ω is the two dimensional analog of the conventional one dimensional angle ϑ . Just as the angle ϑ is defined as the distance along a circle divided by the radius of that circle, so the solid angle Ω is analogously defined as the area on the surface of a sphere divided by the radius squared of that sphere. The units for ϑ and Ω are radians (r) and steradians (sr), respectively; although it should be noted that both of these measures of angle have no actual dimensions. Since the total surface area of a sphere of radius R is $4\pi R^2$, the total solid angle in one sphere is 4π sr.

Consider a differential area dA on the surface of a sphere, in the form of a thin ring centered about the symmetry axis. This ring can be thought of as the intersection of the spherical surface with two cones, one of half-angle ϑ , and other of half-angle $\vartheta + d\vartheta$. The width of this ring is $Rd\vartheta$, and the radius of the ring is $R \sin \vartheta$. The differential solid angle is then

$$d\Omega = \frac{dA}{R^2} = \frac{(2\pi R \sin \vartheta)(Rd\vartheta)}{R^2} = 2\pi \sin \vartheta d\vartheta. \quad (\text{A1})$$

The solid angle inside a cone of half-angle ϑ_c can be determined by integrating

$$\begin{aligned} \Omega &= \int d\Omega = \int_0^\pi 2\pi \sin \vartheta d\vartheta = -2\pi \cos \vartheta \Big|_0^{\vartheta_c} \\ &= 2\pi(1 - \cos \vartheta_c). \end{aligned} \quad (\text{A2})$$

It is often of interest to consider the small-angle approximation, where $\vartheta_c \ll 1$. In this limit, $\cos \vartheta_c \approx 1 - \vartheta_c^2/2$. Therefore, the solid angle of a cone with small half-angle ϑ_c is $\Omega \approx \pi\vartheta_c^2$.

Appendix B: Conservation of mass and momentum

Consider a fluid with local density $\rho(t, x, y, z)$ and local velocity $\vec{u}(t, x, y, z)$. Consider a control volume V (not necessarily small, not necessarily rectangular) which has boundary S . The total mass in this volume is

$$M = \int \rho dV. \quad (\text{B1})$$

The rate-of-change of this mass is just

$$\frac{\partial M}{\partial t} = \int \frac{\partial \rho}{\partial t} dV. \quad (\text{B2})$$

The only way such change can occur is by stuff flowing across the boundary, so

$$\frac{\partial M}{\partial t} = \int \rho \vec{u} \cdot d\vec{S}. \quad (\text{B3})$$

We can change the surface integral into a volume integral using Green's theorem, to obtain

$$\frac{\partial M}{\partial t} = - \int \vec{\nabla} \cdot (\rho \vec{u}) dV. \quad (\text{B4})$$

Note that (B2) and (B4) must be equal no matter what volume V we choose, so the integrals must be pointwise equal. This gives us an expression for the local conservation of mass

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{u}) = 0, \quad (\text{B5})$$

which is sometimes called *continuity equation*.

We can go through the same process for momentum instead of mass. We use Π to represent momentum, to avoid conflict with P which represents pressure. The total momentum in the control volume is:

$$\Pi_i = \int \rho u_i dV, \quad (\text{B6})$$

where the index i runs over the three components of the momentum. The rate-of-change thereof is just

$$\frac{\partial \Pi_i}{\partial t} = \int \frac{\partial(\rho u_i)}{\partial t} dV. \quad (\text{B7})$$

We (temporarily) assume that there are no applied forces (i.e. no gravity) and no pressure (e.g. a fluid of non-interacting dust particles). We also assume viscous forces are negligible. Then, the only way a momentum-change can occur is by momentum flowing across the boundary,

$$\frac{\partial \Pi_i}{\partial t} = \int (\rho u_i) \vec{u} \cdot d\vec{S} = \int (\rho u_i u^j d_j S). \quad (\text{B8})$$

We are expressing dot products using the Einstein summation convention, i.e. implied summation over repeated dummy indices, such as j in the previous expression. We can change the surface integral into a volume integral using Green's theorem, to obtain

$$\frac{\partial \Pi_i}{\partial t} = - \int \nabla_j (\rho u_i u^j) dV. \quad (\text{B9})$$

Note that (B7) and (B9) must be equal no matter what volume V we choose, so the integrands must be pointwise equal. This gives us an expression for the local conservation of momentum,

$$\frac{\partial \Pi_i}{\partial t} = \frac{\partial(\rho u_i)}{\partial t} = -\nabla_j (\rho u_i u^j). \quad (\text{B10})$$

We can understand this equation as follows: each component of the momentum-density ρu_i (for each i separately) obeys a local conservation law. There are strong parallels between (B5) and (B10). Note that the ∇_j operator on the right-hand-side is differentiating two velocities (u_i and u_j) only one of which undergoes dot-product summation (namely summation over j). Using vector component notation (such as $\nabla_j u^j$) is a bit less elegant than using pure vector notation (such as $\vec{\nabla} \cdot \vec{u}$) but in this case it makes things clearer.

We now consider the effect of pressure. It contributes a force on the particles in the control volume, namely

$$F_i = \int P d_i S = - \int \nabla_i P dV. \quad (\text{B11})$$

A uniform gravitational field contributes another force, namely

$$F_i = \int \rho g_i dV. \quad (\text{B12})$$

These forces contribute to changing the momentum, by the second law of motion:

$$\frac{\Pi'_i}{dt} = F_i. \quad (\text{B13})$$

Note the tricky notation: we write d/dt rather than $\partial/\partial t$, and Π' rather than Π , to remind ourselves that the three laws of motion apply to particles, not to the control volume itself. The rate-of-change of Π , the momentum in the control volume, contains the Newtonian contributions, (B11) and (B12) via (B13), plus the flow contributions (B10).

Combining all the contributions we obtain the main result, Euler's equation of motion:

$$\frac{\partial(\rho u_i)}{\partial t} + \nabla_j(\rho u_i u^j) = -\nabla_i P + \rho g_i. \quad (\text{B14})$$

One sometimes encounters other ways of expressing the same equation of motion. Rather than emphasizing the momentum, we might want to emphasize the velocity. This is not a conserved quantity, but sometimes it is easier to visualize and/or easier to measure. If we expand the left-hand-side we have

$$\rho \frac{\partial u_i}{\partial t} + u_i \frac{\partial \rho}{\partial t} + u_i \nabla_j(\rho u^j) + \rho u_j \nabla_j u_i = \rho g_i - \nabla_i P, \quad (\text{B15})$$

where the second and third terms cancel because of conservation of mass (B5), leaving us with

$$\rho \frac{\partial u_i}{\partial t} + \rho u_j \nabla_j u_i = -\nabla_i P + \rho g_i. \quad (\text{B16})$$

Converting from component notation to vector notation, we obtain

$$\rho \frac{\partial \vec{u}}{\partial t} + \rho (\vec{u} \cdot \vec{\nabla}) \vec{u} = -\vec{\nabla} P + \rho \vec{g}. \quad (\text{B17})$$

If we now consider a plane-parallel ($\partial/\partial y = 0$, $\partial/\partial z = 0$, $\partial/\partial x = d/dx$) steady-state ($\partial/\partial t = 0$) flow and we ignore gravity, (B5) and (B17) become

$$\frac{d}{dx}(\rho u) = 0, \quad (\text{B18})$$

and

$$u \frac{du}{dx} = -\frac{1}{\rho} \frac{dP}{dx}; \quad (\text{B19})$$

respectively. (B18) immediately gives

$$\rho u = \text{constant} \rightarrow \rho_1 u_1 = \rho_2 u_2. \quad (\text{B20})$$

Using

$$\begin{aligned} \frac{d}{dx}(\rho u^2) &= 2\rho u \frac{du}{dx} + u^2 \frac{d\rho}{dx} \\ &= \rho u \frac{du}{dx} + u \left(\rho \frac{du}{dx} + u \frac{d\rho}{dx} \right) \\ &= \rho u \frac{du}{dx} + u \frac{d}{dx}(\rho u) \\ &= \rho u \frac{du}{dx} \end{aligned} \quad (\text{B21})$$

(B19) can be rewritten as

$$\rho u \frac{du}{dx} + \frac{dP}{dx} = \frac{d}{dx}(\rho u^2 + P) = 0. \quad (\text{B22})$$

This leads to

$$\rho u^2 + P = \text{constant} \rightarrow \rho_1 u_1^2 + P_1 = \rho_2 u_2^2 + P_2. \quad (\text{B23})$$

Appendix C: Kruskal coordinates

One elegant coordinate substitution is the replacement of r and t by the Kruskal coordinates x and y , which are defined by the following two equations [294]

$$xy = \left(\frac{r}{2M} - 1 \right) e^{r/(2M)} \quad (\text{C1})$$

and

$$x/y = e^{t/(2M)}. \quad (\text{C2})$$

The angular coordinates θ and ϕ are kept the same. Hereafter, we adopt geometrodynamic units $G = c = 1$. By taking the ln of (C1) and (C2), and partially differentiating with respect to x and y , we read off

$$\frac{dx}{x} + \frac{dy}{y} = \frac{dr}{r - 2M} + \frac{dr}{2M} = \frac{dr}{2M(1 - 2M/r)} \quad (\text{C3})$$

and

$$\frac{dx}{x} - \frac{dy}{y} = \frac{dt}{2M}. \quad (\text{C4})$$

The Schwarzschild metric (100) is now given by

$$\begin{aligned} ds^2 &= -16M^2 \left(1 - \frac{2M}{r} \right) \frac{dx}{x} \frac{dy}{y} - r^2 d\Omega^2 \\ &= -\frac{32M^3}{r} e^{-r/(2M)} dx dy - r^2 d\Omega^2. \end{aligned} \quad (\text{C5})$$

Note that, in the last expression, the zero and the pole at $r = 2M$ have cancelled out. The function $r(x, y)$ can be obtained by inverting the algebraic expression (C1) and is regular in the entire region $xy > -1$. In particular, nothing special seems to happen on the two lines $x = 0$ and $y = 0$. Apparently, there is no physical singularity or curvature singularity at $r \rightarrow 2M$. We do notice that the

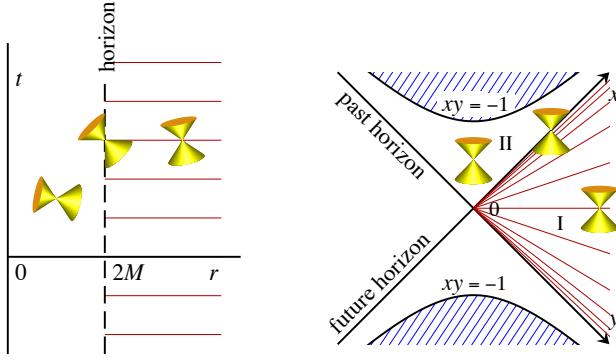


FIG. 45: **Left.** The black hole in the Schwarzschild coordinates (t, r) . The event horizon is at $r = 2M$. **Right.** Kruskal coordinates. Here, the coordinates of the horizon are at $x = 0$ and at $y = 0$. The orientation of the local lightcones is indicated. Thin red lines are the time = constant lines in the physical part of spacetime [295].

line $x = 0$, θ and ϕ both constant, is lightlike, since two neighboring points on that line obey $dx = d\theta = d\phi = 0$, and this implies that $ds^2 = 0$, regardless the value of dy . Likewise, the line $y = 0$ is lightlike. Indeed, we can also read off from the original expression (100) that if $r = 2M$, the lines with constant θ and ϕ are lightlike, as $ds^2 = 0$ regardless the value of dt . The line $y = 0$ is called the future horizon and the line $x = 0$ is the past horizon.

An other important point to highlight is that (C2) attaches a real value for the time t when x and y both have the same sign, such as is the case in the region marked I in Fig. 45, but if $xy < 0$, as in region II, the coordinate t gets an imaginary part. This means that region II is not part of our universe. Actually, t does not serve as a time coordinate there, but as a space coordinate, since there, dt^2 enters with a negative sign in the metric (100). r is then the time coordinate.

Even if we restrict ourselves to the regions where t is real, we find that, in general, every point (r, t) in the physical region of spacetime is mapped onto two points in the (x, y) plane: the points (x, y) and $(-x, -y)$ are mapped onto the same point (r, r) . This leads to the picture of a black hole being a “wormhole” connecting our universe to another universe, or perhaps another region of the spacetime of our universe [296]. However, there are no timelike or light like paths connecting these two universes [297]. If this is a wormhole at all, it is a purely spacelike one.

Appendix D: Geometry of S^3 and H^3

Herein we provide a geometric interpretation of the hyper-sphere S^3 and the hyperbolic hyper-plane H^3 . The explanation given herein will build upon the content of the exquisite book by Kolb and Turner [148].

We begin by studing the familiar two dimensional

surfaces. To visualize the two sphere it is convenient to introduce an extra fictitious spatial dimension and to embed this two-dimensional curve space in a three-dimensional Euclidean space with cartesian coordinates x_1, x_2, x_3 . The equation of the two sphere S^2 of radius R is

$$x_1^2 + x_2^2 + x_3^2 = R^2. \quad (D1)$$

The line element in the three-dimensional Euclidean space is

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2. \quad (D2)$$

If x_3 is taken as the fictitious third spatial coordinate, it can be eliminated from ds^2 by the use of (D1)

$$ds^2 = dx_1^2 + dx_2^2 + \frac{(x_1 dx_1 + x_2 dx_2)^2}{R^2 - x_1^2 - x_2^2}. \quad (D3)$$

Now, introduce the coordinates ϱ and θ defined in terms of x_1 and x_2 by

$$x_1 = \varrho \cos \theta \quad \text{and} \quad x_2 = \varrho \sin \theta. \quad (D4)$$

Physically, ϱ and θ correspond to polar coordinates in the x_3 -plane; $x_3^2 = R^2 - \varrho^2$. In terms of the new coordinates, (D3) becomes

$$ds^2 = \frac{R^2 d\varrho^2}{R^2 - \varrho^2} + \varrho^2 d\theta^2. \quad (D5)$$

Note the similarity between this metric and the spatial hypersurface with $k = 1$ in (139).

Another convenient coordinate system for the two sphere is that specified by the usual polar and azimuthal angles (θ, ϕ) of spherical coordinates, related to x_i by

$$x_1 = R \sin \theta \cos \phi, \quad x_2 = R \sin \theta \sin \phi, \quad x_3 = R \cos \theta. \quad (D6)$$

In terms of these coordinates, (D2) becomes

$$ds^2 = R^2 [d\theta^2 + \sin^2 \theta d\phi^2]. \quad (D7)$$

This form makes manifest the fact that the space is the two sphere of radius R .

The equivalent formulas for a space of constant negative curvature can be obtained with the replacement $R \rightarrow iR$ in (D1). The metric corresponding to the form of (D5) for the negative curvature case is

$$ds^2 = \frac{R^2 d\varrho^2}{R^2 + \varrho^2} + \varrho^2 d\theta^2. \quad (D8)$$

and the metric in the form corresponding to (D7) is

$$ds^2 = R^2 [d\theta + \sinh^2 \theta d\phi^2]. \quad (D9)$$

The embedding of the hyperbolic plane H^2 in an Euclidean space requires three fictitious extra dimensions, and such an embedding is of little use in visualizing the

geometry. While H^2 cannot be globally embedded in \mathbb{R}^3 , it can be partially represented by the pseudosphere (89).

The generalization of the two-dimensional models discussed above to three spatial dimensions is trivial. For the three sphere S^3 a fictitious fourth spatial dimension is introduced and in cartesian coordinates the three sphere is defined by $R^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$. The spatial metric of a four-dimensional Euclidean space is $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$. The fictitious coordinate can be removed to give

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + \frac{(x_1 dx_1 + x_2 dx_2 + x_3 dx_3)^2}{R^2 - x_1^2 - x_2^2 - x_3^2}. \quad (\text{D10})$$

In terms of coordinates $x_1 = \varrho \sin \theta \cos \phi$, $x_2 = \varrho \sin \theta \sin \phi$, $x_3 = \varrho \cos \theta$, the metric is given by the spatial part of (139) with $k = 1$. In terms of a coordinate system that employs the 3 angular coordinates (χ, θ, ϕ) of a four-dimensional spherical coordinate system, $x_1 = R \sin \chi \sin \theta \cos \phi$, $x_2 = R \sin \chi \sin \theta \sin \phi$, $x_3 = R \sin \chi \cos \theta$, $x_4 = R \cos \chi$, the metric is given by

$$ds^2 = R^2 [d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2)]. \quad (\text{D11})$$

The substitution $\chi = r/R$ leads to the spatial part of (141) with $k = 1$.

As in the two-dimensional example, the three-dimensional open model is obtained by the replacement $R \rightarrow iR$, which gives the metric in the form (139) with $k = -1$, or in the form (141) with $\sin \chi \rightarrow \sinh \chi$. Again the space is unbounded and R sets the curvature scale. Embedding H^3 in an Euclidean space requires four fictitious extra dimensions.

Appendix E: Dirac Delta Function

Dirac's delta function is defined by the following property

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases}, \quad (\text{E1})$$

with

$$\int_{t_1}^{t_2} dt \delta(t) = 1 \quad (\text{E2})$$

if $0 \in [t_1, t_2]$ (and zero otherwise). It is "infinitely peaked" at $t = 0$, with the total area of unity. You can view this function as a limit of Gaussian

$$\delta(t) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi} \sigma} e^{-t^2/(2\sigma^2)}, \quad (\text{E3})$$

or a Lorentzian

$$\delta(t) = \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \frac{\epsilon}{t^2 + \epsilon^2}. \quad (\text{E4})$$

The important property of the delta function is the following relation

$$\int dt f(t) \delta(t) = f(0), \quad (\text{E5})$$

which is valid for any *test function* $f(t)$ that is bounded and differentiable to any order, and which vanishes outside a finite range. This is easy to see. First of all, $\delta(t)$ vanishes everywhere except $t = 0$. Therefore, it does not matter what values the function $f(t)$ takes except at $t = 0$. You can then say $f(t) \delta(t) = f(0) \delta(t)$. Then $f(0)$ can be pulled outside the integral because it does not depend on t , and you obtain the right-hand-side. This equation can easily be generalized to

$$\int dt f(t) \delta(t - t_0) = f(t_0). \quad (\text{E6})$$

Mathematically, the delta function is not a function, because it is "too singular." Instead, it is said to be a "distribution." It is a generalized idea of functions, but can be used only inside integrals. In fact, $\int dt \delta(t)$ can be regarded as an "operator" which pulls the value of a *test function* at zero. Put it this way, it sounds perfectly legitimate and well-defined. But as long as it is understood that the delta function is eventually integrated, we can use it as if it is a function.

The step (Heaviside) function,

$$\Theta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}, \quad (\text{E7})$$

is the "primitive" (at least in symbolyc form) of the $\delta(x)$. Equivalently, $\Theta'(x)$, has the symbolic limit $\delta(x)$, as we show next. For any given *test function* $f(x)$, integration by parts leads to

$$\begin{aligned} \int_{-\infty}^{+\infty} \Theta'(x) f(x) dx &= - \int_{-\infty}^{+\infty} \Theta(x) f'(x) dx \\ &= - \int_0^{\infty} f'(x) dx = f(0); \end{aligned} \quad (\text{E8})$$

therefore $\Theta'(x) = \delta(x)$.