

Lectures on Astronomy, Astrophysics, and Cosmology

Luis A. Anchordoqui

Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

(Dated: Spring 2007)

These lecture notes were prepared as an aid to students enrolled in (UG) Astronomy 320, in the 2007 Spring semester, at UW-Milwaukee. They are for study purposes only.

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.

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. 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, 1 light second = $1 \text{ s} \times 3 \times 10^8 \text{ m/s} = 3 \times 10^8 \text{ m} = 300,000 \text{ km}$,

1 light minute = $18 \times 10^6 \text{ km}$, and 1 light year

$$\begin{aligned} 1 \text{ ly} &= 2.998 \times 10^8 \text{ m/s} \times 3.156 \times 10^7 \text{ s/yr} \\ &= 9.46 \times 10^{15} \text{ m} \\ &\approx 10^{13} \text{ km}. \end{aligned} \quad (1)$$

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 $1.5 \times 10^{11} \text{ m}$ or 150,000,000 km; this is equal to 8.3 light minutes. 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 4.3 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 (about 1750) Thomas Wright suggested that the Milky Way was a flat disc of stars extending to great distances in a plane, which we call the Galaxy (Greek for "milky way").

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

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

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 Law, $F = ma$, with $a = v^2/r$ being the centripetal acceleration and F being the Universal Law of Gravitation:

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

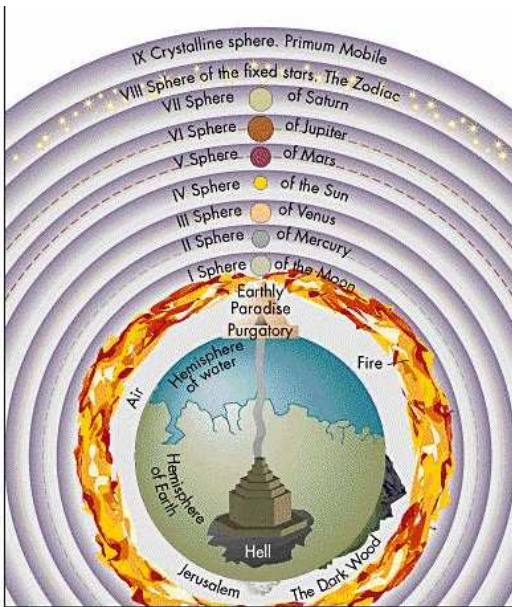


FIG. 1: Celestial spheres of ancient cosmology.

where $G = 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ is Newton's constant. All in all,

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

Assuming all the stars in the Galaxy are similar to our Sun ($M_{\odot} = 2 \times 10^{30} \text{ kg}$), 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 “nebulae” (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, Charles Messier published a celebrated catalogue. 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. Immanuel Kant (about 1755) 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. 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. Edwin Hubble (1889-1953) 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 [1]. 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 nebulae 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 [2].

II. DISTANCE MEASUREMENTS BY PARALLAX

Last class we have been talking about the vast distance of the objects in the universe. Today, we will discuss different methods to estimate these distances. One basic

method employs simple geometry to measure the parallax of a star. By parallax we mean the apparent motion of a star against the background of more distant stars, due to Earth's motion around the Sun. The sighting angle of a star relative to the plane of Earth's orbit (usually indicated by θ) can be determined at different times of the year. 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 $d = 1.5 \times 10^8 \text{ km}$ 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 $\phi = 6 \times 10^{-5}''$, we would say the star is at a distance $D = 4.5 \text{ pc}$. It is easily seen that

$$1 \text{ pc} = 3.26 \text{ ly} = 3.08 \times 10^{16} \text{ m} . \quad (6)$$

Parallax can be used to determine the distance to stars as far away as about 100 ly ($\approx 30 \text{ pc}$) from Earth. Beyond that distance, parallax angles are too small to measure and more subtle techniques must be employed.

III. LUMINOSITY AND BRIGHTNESS

A useful parameter for a star or galaxy is its luminosity (or “absolute luminosity”), L , by which we mean the total power radiated in watts. Also important is the apparent brightness, l , defined as the power crossing unit area at the Earth perpendicular to the path of light. Given that energy is conserved and ignoring any absorption in space, the total emitted power L when it reaches a distance D from the star will be spread over a sphere of surface area $4\pi D^2$. If D is the distance from the star to Earth, then

$$L = 4\pi D^2 l. \quad (7)$$

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.*

IV. SURFACE TEMPERATURE

Another important parameter of a star is its surface temperature, which can be determined from the spectrum of electromagnetic frequencies it emits.

The rate at which an object radiates energy has been found to be proportional to the fourth power of the Kelvin temperature T and to the area A of the emitting object, i.e., $L \propto AT^4$. At normal temperatures (≈ 300 K) we are not aware of this electromagnetic radiation because of its low intensity. At higher temperatures, there is sufficient infrared radiation that we can feel heat if we are close to the object. At still higher temperatures (on the order of 1000 K), objects actually glow, such as a red-hot electric stove burner. At temperatures above 2000 K, objects glow with a yellow or whitish color, such as the filament of a lightbulb.

Planck's law of blackbody radiation predicts the spectral intensity of electromagnetic radiation at all wavelengths from a blackbody at temperature T . The spectral radiance or brightness (i.e., the energy per unit time per unit surface area per unit solid angle per unit frequency ν) is

$$I(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1}, \quad (8)$$

where $h = 6.626 \times 10^{-34}$ J s and $k = 1.38 \times 10^{-23}$ J K $^{-1}$ are the Planck and Boltzmann constants and c is the speed of light. The law is sometimes written in terms of the spectral energy density

$$u(\nu, T) = \frac{4\pi}{c} I(\nu, T), \quad (9)$$

which has units of energy per unit volume per unit frequency. Integrated over frequency, this expression yields the total energy density. Using the relation $\lambda = c/\nu$, the spectral energy density can also be expressed as a function of the wavelength,

$$u(\lambda, T) = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}. \quad (10)$$

Stars are fairly good approximations of blackbodies. As one can see in Fig. 2, the 5500 K curve, corresponding to the temperature of the Sun, peaks in the visible part of the spectrum ($400 \text{ nm} < \lambda < 750 \text{ nm}$). For lower temperatures the total radiation drops considerably and the peak occurs at longer wavelengths. It is found experimentally that the wavelength at the peak of the spectrum, λ_p , is related to the Kelvin temperature by

$$\lambda_p T = 2.9 \times 10^{-3} \text{ m K}. \quad (11)$$

This is known as Wien's law.

We can now use Wien's law and the Steffan-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 measured apparent brightnesses 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

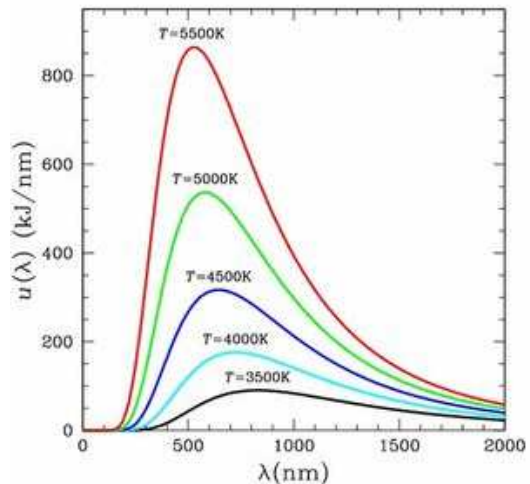


FIG. 2: Measured spectra of wavelengths emitted by a blackbody at different temperatures.

spectrum of the other peaks at about 350 nm (bluish). Using Wien's law, the temperature of the reddish star is

$$T_r = \frac{2.90 \times 10^{-3} \text{ m K}}{700 \times 10^{-9} \text{ m}} = 4140 \text{ K}. \quad (12)$$

The temperature of the bluish star will be double because its peak wavelength is half; just to check

$$T_b = \frac{2.90 \times 10^{-3} \text{ m K}}{350 \times 10^{-9} \text{ m}} = 8280 \text{ K}. \quad (13)$$

For a blackbody the Steffan-Boltzmann equation reads

$$L = \sigma AT^4, \quad (14)$$

where $\sigma = 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. Thus, the power radiated per unit of area from a star is proportional to the fourth power of the Kelvin temperature. 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) [3].

V. HR DIAGRAM

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 [4]. On the HR diagram, the horizontal axis shows the temperature T , whereas the vertical axis the luminosity L , each star

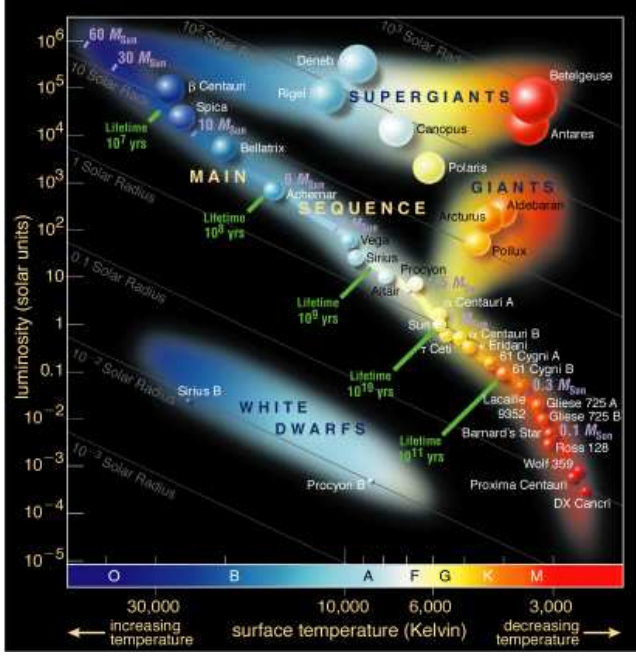


FIG. 3: Hertzsprung-Russell diagram. The vertical axis depicts the inherent brightness of a star, and the horizontal axis depicts the surface temperature increasing from right to left. The spectral class is also indicated. You can see that the Sun is a G-star.

is represented by a point on the diagram shown in Fig. 3. Most of the stars fall along the diagonal band termed the main sequence. Starting at the lowest right, we find the coolest stars, redish 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. Stars that fall on this diagonal band are called main-sequence stars. There are also stars that fall above the main sequence. Above and to the right we find extremely large stars, with high luminosity but with low (redish) 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.

VI. DISTANCE TO A STAR USING HR

Suppose that a detailed study of a certain star suggests that it most likely fits on the main sequence of the HR diagram. Its measured apparent brightness is $I = 1 \times 10^{-12} \text{ W m}^{-2}$, and the peak wavelength of its spectrum is $\lambda_p \approx 600 \text{ nm}$. We can first find the temperature using Wien's law and then estimate the absolute luminosity using the HR diagram; namely,

$$T \approx \frac{2.9 \times 10^{-3} \text{ m K}}{600 \times 10^{-9} \text{ m}} \approx 4800 \text{ K}. \quad (15)$$

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 Eq. (7) we can estimate its distance from us,

$$\begin{aligned} D &= \sqrt{\frac{L}{4\pi I}} \\ &\approx \sqrt{\frac{10^{26} \text{ W}}{4\pi \cdot 10^{-12} \text{ W m}^{-2}}} \\ &\approx 3 \times 10^{18} \text{ m}, \end{aligned} \quad (16)$$

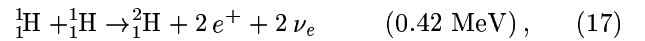
or equivalently 300 ly.

VII. 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, in today's class we will follow the process of stellar evolution from the birth to the death of a star, as we have theoretically reconstructed it.

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 was 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:

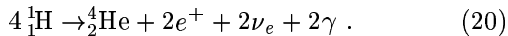


and



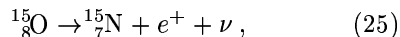
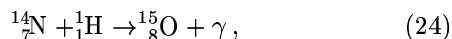
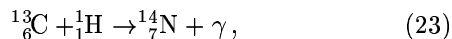
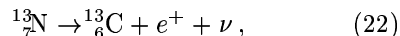
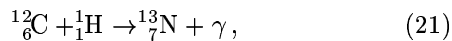
¹ The word “burn” is put in quotation marks because these high-temperature fusion reactions occur via a *nuclear* process, and must not be confused with ordinary burning in air, which is a *chemical* reaction, occurring at the *atomic* level (and at a much lower temperature).

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 ${}^4_2\text{He}$ 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 ${}^3_2\text{He}$ for the third reaction. So the total 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:



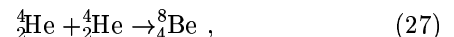
It is easily seen that no carbon is consumed in this cycle (see first and last equations) and that the net effect is the same as the *pp* cycle. The theory of the *pp* cycle and the carbon cycle as the source of energy for the Sun and the stars was first worked out by Hans Bethe in 1939 [5].

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. 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, if it is a star like our Sun, and it is expected to remain there 10 billion years (10^{10} yr). Although most of stars are billions of years old, there is

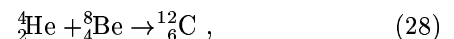
evidence that stars are actually being born at this moment in the Eagle Nebula.

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.

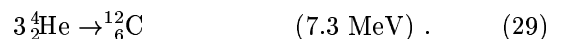
If the star is like our Sun, or larger, further fusion can occur. As the star's outer envelope expands, its core is shrinking and heating up. When the temperature reaches about 10^8 K, even helium nuclei, in spite of their greater charge and hence greater electrical repulsion, can then reach each other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



and



with the emission of two γ -rays. The two reactions must occur in quick succession (because ${}^8_4\text{Be}$ is very unstable), and the net effect is



This fusion of helium causes a change in the star which moves rapidly to the "horizontal branch" of the HR diagram. Further fusion reactions are possible, with ${}^4_2\text{He}$ fusing with ${}^{12}_6\text{C}$ to form ${}^{16}_8\text{O}$. In very massive stars, higher Z elements like ${}^{20}_{10}\text{Ne}$ or ${}^{24}_{12}\text{Mg}$ can be made. This process of creating heavier nuclei from lighter ones (or by absorption of neutrons at higher Z) is called nucleosynthesis.

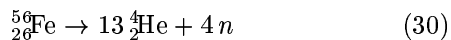
The final fate of the star depends on its mass. Stars can lose mass as parts of their envelope drift off into space. Stars born with a mass less than about 8 solar masses eventually end up with a residual mass less than about 1.4 solar masses, which is known as the Chandrasekhar limit [6]. For them, no further fusion energy can be obtained. The core of such low mass star (original mass

$\lesssim 8M_{\odot}$) contracts under gravity; the outer envelope expands again and the star becomes an even larger red giant. Eventually the outer layers escape into space, the core shrinks, the star cools, descending downward in the HR diagram, becoming a white dwarf. A white dwarf contracts to the point at which the electron clouds start to overlap, but collapses no further because of the Pauli exclusion principle (no two electrons can be in the same quantum state). Arriving at this point is called electron degeneracy. A white dwarf continues to lose internal energy by radiation, decreasing in temperature and becoming dimmer until its light goes out. It has then become a cold dark chunk of ash.

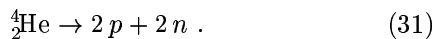
Stars whose residual mass is greater than the Chandrasekhar limit are thought to follow a quite different scenario. A star with this great mass can contract under gravity and heat up even further. In the range $T = 2.5 - 5 \times 10^9$ K, nuclei as heavy as ${}^{56}_{26}\text{Fe}$ and ${}^{58}_{28}\text{Ni}$ can be made. As massive red supergiants age, they produce “onion layers” of heavier and heavier elements in their interiors. 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.

Elements heavier than Ni are thought to form mainly by neutron capture, particularly in supernova explosions. Large number of free neutrons, resulting from nuclear reactions, are present inside highly evolved stars and they can readily combined with, say, a ${}^{56}_{26}\text{Fe}$ nucleus to form (if three are captured) ${}^{59}_{26}\text{Fe}$ which decays to ${}^{59}_{27}\text{Co}$. The ${}^{59}_{27}\text{Co}$ can capture neutrons, also becoming neutron rich and decaying by β^- to next higher Z element and so on. The highest Z elements are thought to form by such neutron capture during supernova (SN) explosions when hordes of neutrons are available.

Yet at these extremely high temperatures, well above 10^9 K, the kinetic energy of the nuclei is so high that fusion of elements heavier than iron is still possible even though the reactions require energy input. But the high energy collisions can also cause the breaking apart of iron and nickel nuclei into helium nuclei and eventually into protons and neutrons



and



These are energy-requiring reactions, but at such extremely high temperature and pressure there is plenty of energy available, enough even to force electrons and protons together to form neutrons in inverse β decay



Eventually, the iron core reaches the Chandrasekhar mass. When something is this massive, not even electron

degeneracy pressure can hold it up. As the core collapses under the huge gravitational forces, protons and electrons are pushed together to form neutrons and neutrinos (even though neutrinos don’t interact easily with matter, at these extremely high densities, they exert a tremendous outward pressure), and the star begins to contract rapidly towards forming an enormously dense neutron star [7]. The outer layers fall inward when the iron core collapses. When the core stops collapsing (this happens when the neutrons start getting packed too tightly – neutron degeneracy), the outer layers crash into the core and rebound, sending shock waves outward. These two effects – neutrino outburst and rebound shock wave – cause the entire star outside the core to be blow apart in a huge explosion: a type II supernova! It has been suggested that the energy released in such a catastrophic explosion could form virtually all elements in the periodic table.² The presence of heavy elements on Earth and in our solar system suggests that our solar system formed from the debris of a supernova.

The core of a neutron star contracts to the point at which all neutrons are as close together as they are in a nucleus. That is, the density of a neutron star is on the order of 10^{14} times greater than normal solids and liquids on Earth. A neutron star that has a mass $\sim 1.5M_{\odot}$ would have a diameter of only about 20 km.

If the final mass of a neutron star is less than about $3 M_{\odot}$, its subsequent evolution is thought to be similar to that of a white dwarf. If the mass is greater than this, the star collapses under gravity, overcoming even the neutron exclusion principle [8]. The star eventually collapses to the point of zero volume and infinite density, creating what is known as a “singularity” [9]. As the density increases, the paths of light rays emitted from the star are bent and eventually wrapped irrevocably around the star. Any emitted photons are trapped into an orbit by the intense gravitational field; they will never leave it. Because no light escapes after the star reaches this infinite density, it is called a black hole.

Binary X-ray sources are places to find strong black hole candidates [10]. 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 (black hole boundary) and crashes into the singularity. Thus we can see 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

² Type Ia supernovae are different. They are believed to be binary stars, one of which is a white dwarf that pulls mass from its companion. When the total mass reaches the Chandrasekhar limit, the star begins to collapse and then explodes as a supernova.

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. When looking for black hole candidates, all these things are taken into account. Many X-ray satellites have scanned the skies for X-ray sources that might be possible black hole candidates. For a hole with the mass of the sun, the density is about 10^{19} kg/m^3 . Such a density is about the highest that can be created through gravitational collapse. Observationally, the lightest black hole candidates are about $6 M_{\odot} \sim 6 \times 10^{30} \text{ kg}$ [3].

VIII. THE OLBERS PARADOX

The XVI century finally saw what came to be a watershed in the development of Cosmology. In 1543 Nicolas 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.

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 Johannes 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, note that if absorption is neglected, the apparent luminosity of a star of absolute luminosity L at a distance r will be $l = L/4\pi r^2$. If the number density of such stars is a constant n , then the number of stars at distances r between r and $r + dr$ is $dN = 4\pi n r^2 dr$, so the total radiant energy density due to all stars is

$$\begin{aligned} \rho_s &= \int l dN = \int_0^\infty \left(\frac{L}{4\pi r^2} \right) 4\pi n r^2 dr \\ &= Ln \int_0^\infty dr. \end{aligned} \quad (33)$$

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

In order to avoid this paradox, both Loys de Chéseaux (1744) [11] and Heinrich Olbers (1826) [12] postulated the existence of an interstellar medium that absorbs the light from very distant stars responsible for the divergence of the integral in Eq. (33). 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.

IX. THE EXPANSION OF THE UNIVERSE

Even though at first glance the stars seem motionless we are going to see that this impression is illusory. 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 $\sim 10^{10} \text{ km}$. 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 $56 \times 10^{12} \text{ km}$; 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 (see Fig. 4). 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.

The observations that we will discuss in this class 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.

Our knowledge of the expansion of the universe rests entirely on the fact that astronomers are able to measure the motion of a luminous body in a direction directly along the line of sight much more accurately than they can measure 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.



FIG. 4: Barnard's star moves against the background of more distant stars faster than any other star that we know about. It moves at the amazing speed of 10.4 seconds of arc/yr, which corresponds to only one lunar diameter every 173 yr. The image shows the region of sky where Barnard's star has been during the past 50 yr. The dimensions of this image are 27×14 minutes of arc. The full moon has a diameter that averages about 30 minutes of arc, so this image has a width approximately equal to that of the full moon. North is "up." Barnard's motion has a direction that is 5 degrees clockwise from north (i.e., position angle = 5 degrees).

A. Doppler Effect

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 Steven Weinberg in "The First Three Minutes." 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.

Suppose that wave crests leave a light source at regular intervals separated by a period T . If the source is moving at velocity $V \ll c$ away from the observer, then during

the time between successive crests the source moves a distance VT . This increases the time required for the wave crest to get from the source to the observer by an amount VT/c , where c is the speed of light. Therefore, the time between arrival of successive wave crests at the observer is

$$T' \approx T + \frac{VT}{c}. \quad (34)$$

The wavelength of the light upon emission is

$$\lambda = cT \quad (35)$$

and the wavelength when the light arrives is

$$\lambda' = cT'. \quad (36)$$

Hence the ratio of these wavelengths is

$$\frac{\lambda'}{\lambda} = \frac{T'}{T} \approx 1 + \frac{V}{c}. \quad (37)$$

The same reasoning applies if the source is moving toward the observer, except that V is replaced with $-V$.³ For example, the galaxies of the Virgo cluster are moving away from our Galaxy at a speed of about 10^3 km/s. Therefore, the wavelength λ' of any spectral line from the Virgo cluster is larger than its normal value λ by a ratio

$$\frac{\lambda'}{\lambda} = \frac{T'}{T} \approx 1 + \frac{10^3 \text{ km/s}}{3 \times 10^5 \text{ km/s}} \approx 1.0033. \quad (38)$$

This effect was apparently first pointed out for both light and sound waves by Johann Christian Doppler in 1842. Doppler thought that this effect might explain the different colors of stars. The light from the stars that happen to be moving away from the Earth would be shifted toward longer wavelengths, and since red light has a wavelength longer than the average wavelength of visible light, such stars might appear more red than average. Similarly, light from stars that happen to be moving toward the Earth would be shifted toward shorter wavelengths, so that the star might appear unusually blue. It was soon pointed out by Buys-Ballot and others that the Doppler effect has essentially nothing to do with the color of a star – it is true that the blue light from a receding star is shifted toward the red, but at the same time some of the star's normally invisible ultraviolet light is shifted into the blue part of the visible spectrum, so the overall color hardly changes. As we discussed in the previous class, stars have different colors chiefly because they have different surface temperatures.

³ For source and observer moving away from each other, Special Relativity's [13] correction leads to $\lambda'/\lambda = \sqrt{(1+V/c)/(1-V/c)}$, where λ is the emitted wavelength as seen in a reference frame at rest with respect to the source and λ' is the wavelength measured in a frame moving with velocity V away from the source along the line of sight; for relative motion toward each other, $V < 0$ in this formula.

B. Hubble Law

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, Joseph 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. 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 William 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. 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., 30 km/s).

In the late 1920's, Hubble discovered that the spectral lines of galaxies were shifted towards the red by an amount proportional to their distances [14]. 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. In terms of the redshift $z \equiv (\lambda' - \lambda)/\lambda$, the “linear” Hubble law can be written as

$$z \approx (H_0/c) r, \quad (39)$$

where c is the speed of light, H_0 is the present value of the Hubble constant and r the distance to the galaxy.⁴ For small velocities ($V \ll c$), from Eq. (37) the Doppler

redshift is $z \approx V/c$. Therefore, $V \approx H_0 r$, which is the most commonly used form of Hubble law. The present day Hubble expansion rate is $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$, where $h = 0.71^{+0.04}_{-0.03}$.

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 the English astrophysicist Edward Arthur Milne.

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 A , B , and C , strung out in a straight line, as shown in Fig. 5. Suppose that the distance between A and B is the same as the distance between B and C . Whatever the speed of B as seen from A , the Cosmological principle requires that C should have the same speed relative to B . But note that C , which is twice away from A as is B , is also moving twice as fast relative to A as is B . We can add more galaxies in our chain, always with the result that the speed of recession of any galaxy relative to any other is proportional to the distance between them.

As often happens in science, this argument can be used both forward and backward. Hubble, in observing a proportionality between the distances of galaxies and their speeds of recession, was indirectly verifying the Cosmological Principle. Contrariwise, we can take the Cosmological Principle for granted on *a priori* grounds, and deduce the relation of proportionality between distance and velocity. In this way, through the relatively easy measurement of Doppler shifts, we are able to judge the distance of very remote objects from their velocities.

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 (including M31 and M33), 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

⁴ 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 $z > 0$ then $\lambda' > \lambda$ and we speak of a redshift; if $z < 0$ then $\lambda' < \lambda$, and we speak of a blueshift.

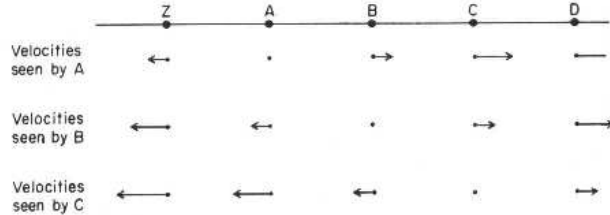


FIG. 5: Homogeneity and the Hubble Law: A string of equally spaced galaxies $Z, A, B, C, \dots$ are shown with velocities as measured from A or B or C indicated by the lengths and directions of the attached arrows. The principle of homogeneity requires that the velocity of C as seen by B is equal to the velocity of B as seen by A , adding these two velocities gives the velocity of C as seen by A , indicated by an arrow twice as long. Proceeding in this way we can fill out the whole pattern of velocities shown in the figure. As can be seen the velocities obey the Hubble law: the velocity of any galaxy, as seen by others is proportional to the distance between them. This is the only pattern of velocities consistent with the principle of homogeneity.

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.

D. The Cosmic Microwave Background

The Cosmological Principle has observational support of another sort, apart from the measurements of the Doppler shifts. After making due allowances for the distortions due to our own Galaxy and the rich nearby cluster of galaxies in the constellation of Virgo, the universe seems remarkably isotropic; that is, it looks the same in all directions. Now, if the universe is isotropic around us, it must also be isotropic about every typical galaxy. However, any point in the universe can be carried into another point by a series of rotations around fixed centers, so if the universe is isotropic around every point, it is necessary also homogeneous. In what follows we will discuss how the observation of the cosmic microwave background (CMB) provides convincing evidence for an isotropic universe.

The expansion of the universe seems to suggest that typical objects in the universe were once much closer together than they are right now. This is the idea for the basis that the universe began about 13.7 billion years ago as an expansion from a state of very high density and temperature known affectionately as the Big Bang.

The Big Bang was not an explosion, because an explosion blows pieces out into the surrounding space. Instead, the Big Bang was the start of an expansion of space itself. The volume of the observable universe was very small at the start and has been expanding ever since. The initial tiny volume of extremely dense matter is not to be thought of as a concentrated mass in the midst of a much larger space around it. The initial tiny but dense volume was the universe – the entire universe. There would not have been anything else. When we say that the universe was once smaller than it is now, we mean that the average separation between galaxies (or other objects) was less. Therefore, it is the *size of the universe itself* that has increased since the Big Bang.

In 1964, Arno Penzias and Robert Wilson were experiencing difficulty with what they assumed to be background noise, or “static,” in their radio telescope [15]. Eventually, they became convinced that it was real and that it was coming from outside the Galaxy. They made precise measurements at wavelength $\lambda = 7.35$ cm, in the microwave region of the electromagnetic spectrum. The intensity of this radiation was found initially not to vary by day or night or time of the year, nor to depend on the direction. It came from all directions in the universe with equal intensity, to a precision of better than 1%. It could only be concluded that this radiation came from the universe as a whole.

The intensity of this CMB as measured at $\lambda = 7.35$ cm corresponds to a blackbody radiation at a temperature of about 3 K. When radiation at other wavelengths was measured, the intensities were found to fall on a blackbody curve, corresponding to a temperature of 2.725 K.

The CMB provides strong evidence in support of the Big Bang, and gives us information about conditions in the very early universe. In fact, in the late 1940s, George Gamow calculated that the Big Bang origin of the universe should have generated just such a CMB [16].

To understand why, let us look at what a Big Bang might have been like. The temperature must have been extremely high at the start, so high that there could not have been any atoms in the very early stages of the universe. Instead the universe would have consisted solely of radiation (photons) and a plasma of charged electrons and other elementary particles. The universe would have been opaque – the photons in a sense “trapped,” travelling very short distances before being scattered again, primarily by electrons. Indeed, the details of the CMB provide strong evidence that matter and radiation were once in thermal equilibrium at very high temperature. As the universe expanded, the energy spread out over an increasingly larger volume and the temperature dropped. Only when the temperature had fallen to about 3,000 K was the universe cool enough to allow the combination of nuclei and electrons into atoms. (In the astrophysical literature this is usually called “recombination,” a singularly inappropriate term, for at the time we were considering the nuclei and electrons had never in the previous history of the universe been combined into atoms!)

The sudden disappearance of electrons broke the thermal contact between radiation and matter, and the radiation continued thereafter to expand freely.

At the moment this happened, the energy in the radiation field at various wavelengths was governed by the conditions of the thermal equilibrium, and was therefore given by the Planck blackbody formula, Eq. (10), for a temperature equal to that of the matter $\sim 3,000$ K. In particular, the typical photon wavelength would have been about one micron, and the average distance between photons would have been roughly equal to this typical wavelength.

What has happened to the photons since then? Individual photons would not be created or destroyed, so the average distance between photons would simply increase in proportion to the size of the universe, i.e., in proportion to the average distance between typical galaxies. But we saw that the effect of the cosmological redshift is to pull out the wavelength of any ray of light as the universe expands; thus the wavelength of any individual photon would also simply increase in proportion to the size of the universe. The photons would therefore remain about one typical length apart, just as for blackbody radiation.

Before proceeding we will pursue this line of argument quantitatively. From Eq. (10) we can obtain the Planck distribution that gives the energy du of a blackbody radiation per unit volume, in a narrow range of wavelengths from λ to $\lambda + d\lambda$,

$$du = \frac{8\pi hc}{\lambda^5} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (40)$$

For long wavelengths, the denominator in the Planck distribution may be approximated by

$$e^{hc/\lambda kT} - 1 \simeq hc/\lambda kT, \quad (41)$$

Hence, in this wavelength region,

$$du = \frac{8\pi kT}{\lambda^4} d\lambda. \quad (42)$$

This is the Rayleigh-Jeans formula. If this formula held down to arbitrarily small wavelengths, $du/d\lambda$ would become infinite for $\lambda \rightarrow 0$, and the total energy density in the blackbody radiation would be infinite. Fortunately, as we saw before, the Planck formula for du reaches a maximum at a wavelength $\lambda = 0.014052 hc/kT$ and then falls steeply off for decreasing wavelengths. The total energy density in the blackbody radiation is

$$u = \int_0^\infty \frac{8\pi hc}{\lambda^5} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (43)$$

Integrals of this sort can be looked up in standard tables of definite integrals; the result gives the Stefan-Boltzmann law

$$u = \frac{8\pi^5 (kT)^4}{15(hc)^3} = 7.56464 \times 10^{-15} (T/K)^4 \text{ erg/cm}^3. \quad (44)$$

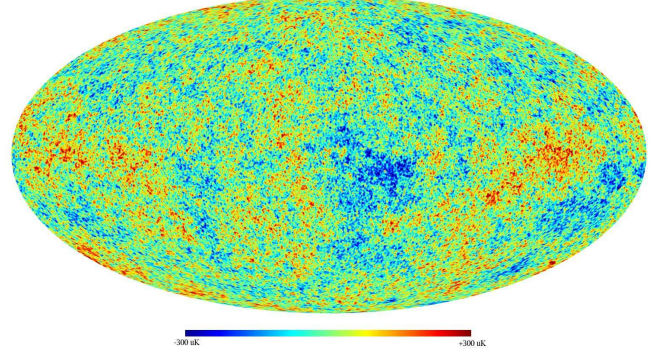


FIG. 6: The CMB over the entire sky, color-coded to represent differences in temperature from the average 2.725 K: the color scale ranges from $+300 \mu\text{K}$ (red) to $-200 \mu\text{K}$ (dark blue), representing slightly hotter and colder spots (and also variations in density.) Results are from the WMAP satellite. The angular resolution is 0.2° [17].

(Recall that $1 \text{ J} \equiv 10^7 \text{ erg} = 6.24 \times 10^{18} \text{ eV}$.)

We can easily interpret the Planck distribution in terms of quanta of light or photons. Each photon has an energy $E = hc/\lambda$. Hence the number dn of photons per unit volume in blackbody radiation in a narrow range of wavelengths from λ to $\lambda + d\lambda$ is

$$dn = \frac{du}{hc/\lambda} = \frac{8\pi}{\lambda^4} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (45)$$

Then the total number of photons per unit volume is

$$\begin{aligned} n &= \int_0^\infty dn \\ &= 8\pi \left(\frac{kT}{hc} \right)^3 \int_0^\infty \frac{x^2 dx}{e^x - 1}, \end{aligned} \quad (46)$$

where $x = hc/(\lambda kT)$. The integral cannot be expressed in terms of elementary functions, but it can be expressed as an infinite series

$$\int_0^\infty \frac{x^2 dx}{e^x - 1} = 2 \sum_{j=1}^\infty \frac{1}{j^3} \approx 2.4. \quad (47)$$

Therefore, the number photon density is

$$\begin{aligned} n &= 60.42198 \left(\frac{kT}{hc} \right)^3 \\ &= 20.28 \left(\frac{T}{\text{K}} \right)^3 \text{ photons cm}^{-3}, \end{aligned} \quad (48)$$

and the average photon energy is

$$\langle E_\gamma \rangle = u/n = 3.73 \times 10^{-16} (T/K) \text{ erg}. \quad (49)$$

Now, let's consider what happens to blackbody radiation in an expanding universe. Suppose the size of the universe changes by a factor f , for example, if it doubles

in size, then $f = 2$. As predicted by the Doppler effect, the wavelengths will change in proportion to the size of the universe to a new value $\lambda' = f\lambda$. After the expansion, the energy density du' in the new wavelength range λ' to $\lambda' + d\lambda'$ is less than the original energy density du in the old wavelength range $\lambda + d\lambda$, for two different reasons:

(i) Since the volume of the universe has increased by a factor of f^3 , as long as no photons have been created or destroyed, the numbers of photons per unit volume has decreased by a factor of $1/f^3$.

(ii) The energy of each photon is inversely proportional to its wavelength, and therefore is decreased by a factor of $1/f$. It follows that the energy density is decreased by an overall factor $1/f^3 \times 1/f = 1/f^4$:

$$du' = \frac{1}{f^4} du = \frac{8\pi hc}{\lambda^5 f^4} d\lambda \frac{1}{e^{hc/\lambda kT} - 1}. \quad (50)$$

If we rewrite Eq. (50) in terms of the new wavelengths λ' , it becomes

$$du' = \frac{8\pi hc}{\lambda'^5} d\lambda' \frac{1}{e^{hc f/\lambda' kT} - 1}, \quad (51)$$

which is exactly the same as the old formula for du in terms of λ and $d\lambda$, except that T has been replaced by a new temperature $T' = T/f$. Therefore, we conclude that freely expanding blackbody radiation remains described by the Planck formula, but with a temperature that drops in inverse proportion to the scale of expansion.

The existence of the thermal CMB gives strong support to the idea that the universe was extremely hot in its early stages. As can be seen in Fig. 6, the background is very nearly isotropic, supporting the isotropic and homogeneous models of the universe. Of course one would expect some small inhomogeneities in the CMB that would provide “seeds” around which galaxy formation could have started. These tiny inhomogeneities were first detected by the COBE (Cosmic Background Explorer) [18] and by subsequent experiments with greater detail, culminating with the WMAP (Wilkinson Microwave Anisotropy Probe) results [19].

X. HOMOGENEOUS AND ISOTROPIC FRIEDMANN-ROBERTSON-WALKER UNIVERSES

In 1917 Albert Einstein presented a model of the universe based on his theory of General Relativity [20]. It described a geometrically symmetric (spherical) space with finite volume but no boundary. In accordance with the Cosmological Principle, the model was homogeneous and isotropic. It was also static: the volume of the space did not change.

In order to obtain a static model, Einstein had to introduce a new repulsive force in his equations [21]. 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

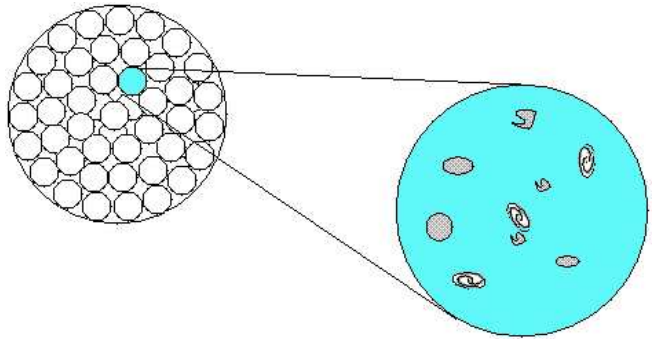


FIG. 7: Spherical region of galaxies with a larger radius than the distance between clusters of galaxies, but smaller radius than any distance characterizing the universe as a whole.

to be static was then reasonable. 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.

The St. Petersburg physicist Alexander Friedmann [22] 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 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. 7. We also assume $\Lambda = 0$.) The mass of this sphere is its volume times the cosmic mass density,

$$M = \frac{4\pi r^3}{3} \rho. \quad (52)$$

We can now consider the motion of a galaxy of mass m at the edge of our spherical region. According to Hubble's law, its velocity will be $V = Hr$ and the corresponding kinetic energy

$$K = \frac{1}{2} m V^2. \quad (53)$$

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

$$U = -\frac{GMm}{r} = -\frac{4\pi m r^2 \rho G}{3}, \quad (54)$$

where $G = 6.67 \times 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ s}^{-2}$ is Newton's constant of gravitation. Hence, the total energy is

$$E = K + U = \frac{1}{2}mV^2 - \frac{GMm}{r}, \quad (55)$$

which has to remain constant as the universe expands. The value of ρ corresponding to $E = 0$ is called the critical density ρ_c . We have,

$$\begin{aligned} E &= \frac{1}{2}mH^2r^2 - \frac{GMm}{r} \\ &= \frac{1}{2}mH^2r^2 - Gm\frac{4\pi}{3}r^2\rho_c \\ &= mr^2 \left(\frac{1}{2}H^2 - \frac{4\pi}{3}G\rho_c \right) = 0 \end{aligned} \quad (56)$$

whence

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (57)$$

The density parameter Ω is defined as $\Omega = \rho/\rho_c$.

Now consider two points at separation r , such that their relative velocity is V . Let $R(t)$ be a time dependent quantity representing the scale of the universe. If R increases with time, all distances, including those between galaxies, will grow. Then

$$r = \frac{R(t)}{R(t_0)} r_0, \quad (58)$$

and

$$V = \dot{r} = \frac{\dot{R}(t)}{R(t_0)} r_0, \quad (59)$$

where dots denote derivative with respect to t . Therefore, the Hubble constant is

$$H = \frac{V}{r} = \frac{\dot{R}(t)}{R(t)}. \quad (60)$$

From the conservation of mass it follows that $\rho_0 R_0^3 = \rho R^3$. Using Eq. (57) for the critical density one obtains

$$\Omega = \frac{8\pi G}{3} \frac{\rho_0 R_0^3}{R^3 H^2}. \quad (61)$$

The deceleration expansion is described by the deceleration parameter q defined as

$$q = -\frac{R\ddot{R}}{\dot{R}^2}. \quad (62)$$

The deceleration parameter describes the change of expansion $\dot{R}$. The additional factors have been included in

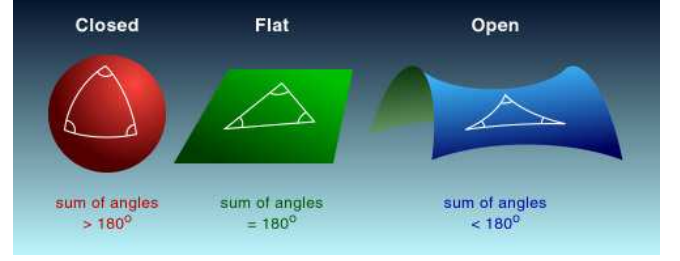


FIG. 8: The two dimensional analogues of the Friedmann models. A spherical surface, a plane, and a pseudo-sphere. Note that the global geometry of the universe affects the sum of angles of a triangle.

order to make it dimensionless, i.e., independent of the choice of units of length and time.

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 M 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.

The $E = 0$ model corresponds to the “flat” Friedmann model, so-called Einstein-de Sitter model. 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. Finally, if $\rho < \rho_c$, 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.

What is meant by a curve space? To answer this question, recall that our normal method of viewing the world is via Euclidean plane geometry. In Euclidean geometry there are many axioms and theorems we take for granted. Non-Euclidean geometries which involve curved space have been independently imagined by Carl Friedrich Gauss (1777-1855), Janos Bolyai (1802 - 1860), and Nikolai Ivanovich Lobachevski (1793-1856). Let us try to understand the idea of a curve space by using two dimensional surfaces.

Consider for example, the two-dimensional surface of a sphere. It is clearly curved, at least to us who view it from outside – from our three dimensional world. But how do the hypothetical two-dimensional creatures determine whether their two-dimensional space is flat (a

plane) or curved? One way would be to measure the sum of the angles of a triangle. If the surface is a plane, the sum of the angles is 180° . But if the space is curved, and a sufficiently large triangle is constructed, the sum of the angles would not be 180° . To construct a triangle on a curve surface, say the sphere of Fig. 8, we must use the equivalent of a straight line: that is the shortest distance between two points, which is called a geodesic. On a sphere, a geodesic is an arc of great circle (an arc in a plane passing through the center of the sphere) such as Earth's equator and longitude lines. Consider, for example, the triangle whose sides are two longitude lines passing from the north pole to the equator, and the third side is a section of the equator. The two longitude lines form 90° angles with the equator. Thus, if they make an angle with each other at the north pole of 90° , the sum of the angles is 270° . This is clearly not Euclidean space. Note, however, that if the triangle is small in comparison to the radius of the sphere, the angles will add up to nearly 180° , and the triangle and space will seem flat. On the saddlelike surface, the sum of the angles of a triangle is less than 180° . Such a surface is said to have negative curvature.

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, Howard Percy Robertson [23] and Arthur Geoffrey Walker [24], 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.

If the universe had a positive curvature, 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 or negative, the universe would be open. It could just go on forever.

In an expanding universe, the galaxies were once much nearer to each other. If the rate of expansion had been unchanging, the inverse of the Hubble constant, $t_{\text{age}} = H_0^{-1}$, would represent the age of the universe. In Friedmann-Robertson-Walker models, however, the expansion is gradually slowing down (i.e., $q < 0$), and thus the Hubble constant gives an upper limit on the age of the universe, $t_{\text{age}} \approx 14$ Gyr. Of course, if $\Lambda \neq 0$ this upper limit for the age of the universe no longer holds.

In an expanding universe the wavelength of radiation is proportional to R , like all other lengths. If the wavelength at the time of emission, corresponding to the scale factor R , is λ , then it will be λ_0 when the scale factor has increased to R_0 : $\lambda_0/\lambda = R_0/R$. The redshift is

$$z = \frac{\lambda_0 - \lambda}{\lambda} = \frac{R_0}{R} - 1; \quad (63)$$

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

XI. THE DARK SIDE OF THE UNIVERSE

According to the standard Big Bang model, the universe is evolving and changing. Individual stars are being created, evolving and dying as white dwarfs, neutron stars, and black holes. At the same time, the universe as a whole is expanding. One important question is whether the universe will continue to expand forever. Just before the year 2000, cosmologists received a surprise. As we discussed in the previous class, gravity was assumed to be the predominant force on a large scale in the universe, and it was thought that the expansion of the universe ought to be slowing down in time because gravity acts as an attractive force between objects. But measurements on type Ia supernovae unexpectedly showed that very distant (high z) supernovae were dimmer than expected [25]. That is, given their great distance d as determined from their low brightness, their speed V determined from the measured z was less than expected according to Hubble's law. This suggests that nearer galaxies are moving away from us relatively faster than those very distant ones, meaning the expansion of the universe in more recent epochs has sped up. This acceleration in the expansion of the universe seems to have begun roughly 5 billion years ago, 8 to 9 Gyr after the Big Bang. What could be causing the universe to accelerate in its expansion? One idea is a sort of quantum field, so-called “quintessence” [26]; another possibility suggests an energy latent in space itself (vacuum energy) related to the cosmological constant. Whatever it is, it has been given the name of “dark energy.”

We will now re-examine the problem of the galaxy at the edge of the massive sphere undergoing expansion to include the dark energy effect. The galaxy will be affected by a central force due to gravity and the cosmological force,

$$m\ddot{r} = -\frac{4\pi G r^3 \rho m}{3r^2} + \frac{1}{3}m\Lambda r, \quad (64)$$

or

$$\ddot{r} = -\frac{4\pi G r \rho}{3} + \frac{1}{3}\Lambda r. \quad (65)$$

In this equations, the radius r and the density ρ are changing with time. They may be expressed in terms of the scale factor R :

$$r = (R/R_0) r_0, \quad (66)$$

where R is defined to be R_0 when the radius $r = r_0$;

$$\rho = (R_0/R)^3 \rho_0, \quad (67)$$

where the density $\rho = \rho_0$ when $R = R_0$. Substituting Eqs. (66) and (67) into Eq. (65), one obtains

$$\ddot{R} = -\frac{a}{R^2} + \frac{1}{3}\Lambda R, \quad (68)$$

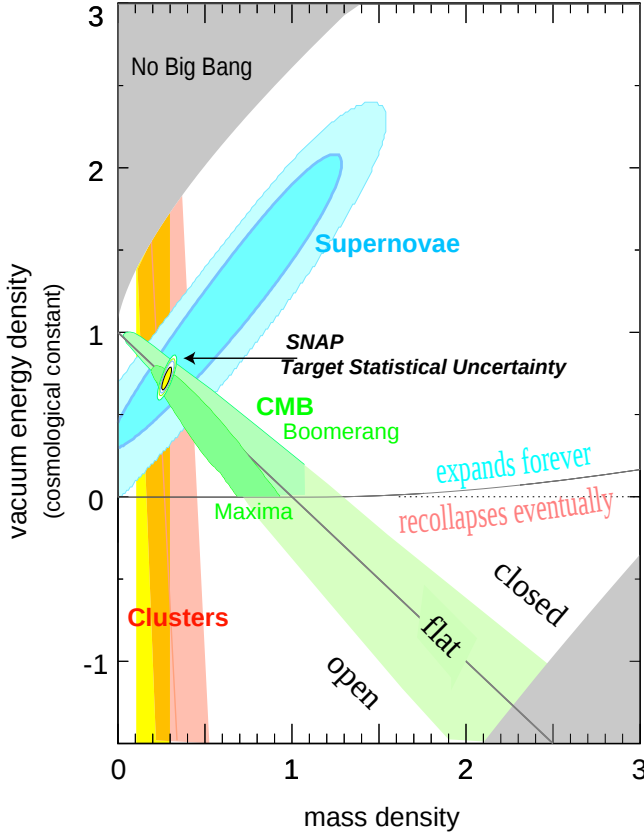


FIG. 9: Shown are three independent measurements of the cosmological parameters (Ω_Λ, Ω_m). The high-redshift supernovae [27], galaxy cluster abundance [28] and the CMB [29] converge nicely near $\Omega_\Lambda = 0.7$ and $\Omega_m = 0.3$. The upper-left shaded region, labeled “no Big Bang,” indicates bouncing cosmologies for which the universe has a turning point in its past [30]. The lower right shaded region corresponds to a universe which is younger than long-lived radioactive isotopes [31], 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 [32].

where $a = 4\pi G R_0^3 \rho_0 / 3$. If Eq. (68) is multiplied on both sides by $\dot{R}$, the left hand side yields

$$\dot{R}\ddot{R} = \frac{1}{2} \frac{d(\dot{R}^2)}{dt}, \quad (69)$$

and thus Eq. (68) takes the form

$$d(\dot{R}^2) = -\frac{2a}{R^2} dR + \frac{2}{3} \Lambda R dR. \quad (70)$$

We now define $R_0 = R(t_0)$. Integrating Eq. (70) from t_0 to t gives

$$\dot{R}^2 - \dot{R}_0^2 = 2a \left(\frac{1}{R} - \frac{1}{R_0} \right) + \frac{1}{3} \Lambda (R^2 - R_0^2). \quad (71)$$

The constants $\dot{R}_0$ and a can be eliminated in favor of the Hubble constant H_0 and the density parameter $\Omega_0 \equiv \Omega_m$.

Because $\rho_c = 3H_0^2/8\pi G$,

$$2a = 8\pi G R_0^3 \rho_0 / 3 = H_0^2 R_0^3 \rho_0 / \rho_c = H_0^2 R_0^3 \Omega_0, \quad (72)$$

where $\Omega_0 = \rho_0/\rho_c$. Using Eq.(72) and $\dot{R}_0 = H_0 R_0$ in Eq. (71), and defining the vacuum energy density $\Omega_\Lambda = \Lambda/(3H_0^2)$, one obtains

$$\frac{\dot{R}^2}{H_0^2 R_0^2} = \Omega_0 \frac{R_0}{R} + \Omega_\Lambda \left(\frac{R}{R_0} \right)^2 + 1 - \Omega_0 - \Omega_\Lambda \quad (73)$$

as the basic differential equation governing $R(t)$. As shown in Fig. 9, the data from the WMAP survey and other recent experiments agree well with theories when they input dark energy as providing 70% of the energy in the universe, and when the total energy density ρ equals the critical density ρ_c . Interestingly, this ρ cannot be only baryonic matter (atoms are 99.9% baryons – protons and neutrons – by weight). WMAP data indicate that the amount of normal baryonic matter in the universe is only 4% of the critical density. What is the other 96%?

There is a strong evidence for a significant amount of nonluminous matter in the universe referred to as dark matter. For example, observations of the rotation of galaxies suggest that they rotate as they had considerably more mass than we can see. From Eq. (3) it is easily seen that the speed of stars revolving around the Galactic center,

$$v = \sqrt{GM/r}, \quad (74)$$

depends on the galactic mass M and their distance from the center r . Observations show that stars farther from the Galactic center revolve much faster than expected from visible matter, suggesting a great deal of invisible matter. Similarly, observations of the motions of galaxies within clusters also suggest they have considerably more mass than can be seen. 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. The evolution of the scale factor for an accelerating universe as derived from Eq. (73) assuming 70% of dark energy, 26% of dark matter, and 4% of baryons, is shown in Fig. 10.

XII. GRAVITATIONAL REDSHIFT

We have seen that the global geometry of spacetime may (in principle) be non-Euclidean; but how does the universe look locally (i.e., on a small scale)? According to Einstein’s theory, spacetime is curved near massive bodies. We might think of space as being like a thin rubber sheet: if a heavy weight is hung from it, it curves. The weight corresponds to a huge mass that causes space (space itself!) to curve. Thus, in Einstein’s theory we do not speak of the “force” of gravity acting on bodies. Instead we say that bodies and light rays move as they do because spacetime is curved. A body at rest or moving

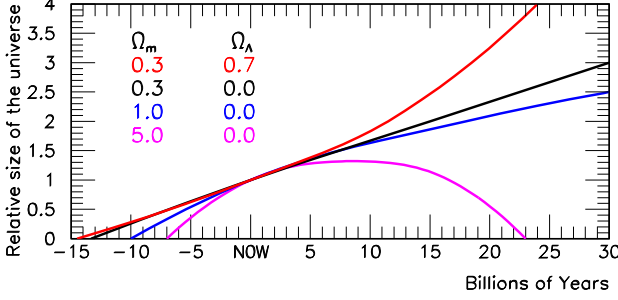


FIG. 10: The three possibilities for expansion in Friedmann-Robertson-Walker models, plus a fourth possibility that includes dark energy.

slowly near a great mass would follow a geodesic towards that body.

The extreme curvature of spacetime could be produced by a black hole. A black hole, as we previously learned, is so dense that even light cannot escape from it. To become a black hole, a body of mass M must undergo gravitational collapse, contracting by gravitational self attraction to within a radius called the Schwarzschild radius [33],

$$r_s = \frac{2GM}{c^2}, \quad (75)$$

where G is the gravitational constant and c the speed of light. The Schwarzschild radius also represents the event horizon of the black hole. By event horizon we mean the surface beyond which no signals can ever reach us, and thus inform us of events that happen. As a star collapses toward a black hole, the light it emits is pulled harder and harder by gravity, but we can still see it. Once the matter passes within the event horizon the emitted light cannot escape, but is pulled back in by gravity. Thus, a light ray propagating in a curve spacetime would be redshifted by the gravitational field. For example, the redshift of radiation from the surface of a star of radius R and mass M is

$$z_g = \frac{1}{\sqrt{1 - r_s/R}} - 1. \quad (76)$$

The gravitational redshift of the radiation from the galaxies is normally insignificant [34].

XIII. ENERGY DENSITY OF STARLIGHT

To see how modern cosmological models avoid the Olbers paradox, we discuss next how the apparent luminosity of a star l and the number of sources per unit volume become corrected due to the expansion of the universe. First, note that each photon emitted with energy $h\nu_1$ will be red-shifted to energy $h\nu_1 R(t_1)/R(t_0)$ and photons emitted at time intervals δt_1 will arrive at time

intervals $\delta t_1 R(t_0)/R(t_1)$, where t_1 is the time the light leaves the source, and t_0 is the time the light arrives at Earth. Second, it is easily seen that the total emitted power by a source at a distance r_1 at the time of emission will spread over a sphere of surface area $4\pi R^2(t_0)r_1^2$. Thus, the apparent luminosity is

$$l = L \left[\frac{R(t_1)}{R(t_0)} \right]^2 \frac{1}{4\pi R^2(t_0) r_1^2}. \quad (77)$$

Now, we assume that at time t_1 there are $n(t_1, L)$ sources per unit volume with absolute luminosity between L and $L + dL$. The proper volume element of space is $dV = R^3(t_1) dr^1 \sin \theta_1 d\theta_1 d\phi_1$, so the number of sources between r_1 and $r_1 + dr_1$ with absolute luminosity between L and $L + dL$ is

$$dN = 4\pi R^3(t_1) r_1^2 n(t_1, L) dr_1 dL. \quad (78)$$

The coordinates r_1 and t_1 are related via $r_1 = r(t_1)$, where $r(t)$ is defined by

$$\int_t^{t_0} \frac{dt'}{R(t')} \equiv \int_0^{r(t)} dr. \quad (79)$$

Differentiation of Eq. (79) leads to $dr_1 = -dt_1/R(t_1)$ and so Eq. (78) can be re-written as

$$dN = 4\pi R^2(t_1) n(t_1, L) r_1^2 |dt_1| dL. \quad (80)$$

The total energy density of starlight is therefore

$$\rho_s = \int l dN = \int_{-\infty}^{t_0} \mathcal{L}(t_1) [R(t_1)/R(t_0)]^4 dt_1, \quad (81)$$

where $\mathcal{L}$ is the proper luminosity density

$$\mathcal{L}(t_1) \equiv \int n(t_1, L) dL. \quad (82)$$

In the hot Big Bang model there is obviously no paradox, because the integral in Eq. (81) is effectively cut off at a lower limit $t_1 = 0$, and the integrand vanished at $t_1 = 0$, roughly like $R(t_1)$. In other words, the modern explanation of the paradox is that the stars have only existed for a finite time, so the light from very distant stars has not yet reached us. Rather than providing the world to be finite in space, as was suggested by Kepler in 1610, the Olbers paradox has shown it to be of a finite age.

XIV. LOOKBACK TIME

When we look out from the Earth, we look back in time. Any other observer in the universe would see the same view. The farther an object is from us, the earlier in time the light we see left it. Figure 11 shows our Earth point of view, looking out in all directions back toward the Big Bang. The brief period (earliest 380,000 yr) when the universe was opaque, because radiation was

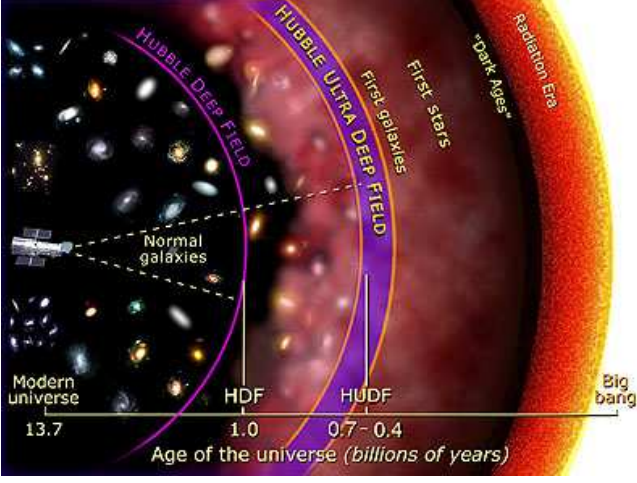


FIG. 11: Observatories such as the Hubble Space Telescope (HST) and the Chandra X-ray Observatory can look back into time, detecting light and other radiation that started out from stars and galaxies billions of years ago. The Ultra Deep Field (UDF) is a deeper, more detailed version of the iconic Hubble Deep Field (HDF) images obtained in 1995 and 1998. The UDF was acquired with a more powerful camera than the original deep field images, and more telescope time was devoted to it, enabling Hubble to capture light from the faintest and most distant objects ever observed. The results give us a view back to a time when the universe was just 5 percent of its present age, about 700 million years after the Big Bang. Future telescopes, such as the James Webb Space Telescope (JWST), will be able to see even farther back in time, to the first stars. Astronomers estimate that these early stars appeared about 300 million years after the Big Bang.

trapped in the “early plasma,” is indicated by an orange band. No matter which direction we look, our view of the very early universe is blocked by this wall of plasma - we can only see as far as its surface, called “the surface of last scattering,” but not into it. Wavelengths from there are red-shifted by $z \approx 1000$.

We now discuss how the parameters of the universe change when looking back toward the Big Bang. As we have seen, in the matter-dominated era, the total mass within a co-moving sphere of radius $R(t)$ is proportional to the number of nuclear particles within the sphere, and it must remain constant (i.e., $4\pi \rho(t) R^3(t)/3 = \text{constant}$). Consequently, $\rho(t) \propto R^{-3}(t)$. Nevertheless, if the mass density is dominated by the mass equivalent to the energy of radiation (the radiation-dominated era), then $\rho(t)$ is proportional to the fourth power of the temperature; and because the temperature varies like $R^{-1}(t)$, $\rho(t) \propto R^{-4}(t)$.

Let us assume that at a time t a typical galaxy of mass m is at a distance $R(t)$ from some arbitrarily chosen galaxy, say the Milky Way. From Eq. (55) we see that

$$E = mR^2(t) \left[\frac{1}{2} H^2(t) - \frac{4}{3} \pi \rho(t) G \right]. \quad (83)$$

Now, $\rho(t)$ increases as $R(t) \rightarrow 0$ at least as fast as $R^{-3}(t)$,

and thus $\rho(t)R^2(t)$ grows at least as fast as $R^{-1}(t)$ for $R(t)$ going to zero. Therefore, in order to keep the energy E constant, the two terms in brackets must nearly cancel each other, so that for $R(t) \rightarrow 0$, $H^2(t) \rightarrow 8\pi\rho(t)G/3$. The characteristic expansion time t_{exp} is the reciprocal of the Hubble constant, hence

$$t_{\text{exp}}(t) \equiv \frac{1}{H(t)} = \sqrt{\frac{3}{8\pi\rho(t)G}}. \quad (84)$$

However, the Hubble constant is proportional to $\sqrt{\rho}$, and therefore we conclude that $H \propto R^{-n/2}(t)$. The velocity of a typical galaxy is then

$$V(t) = H(t)R(t) = \dot{R} \propto R^{1-n/2}(t). \quad (85)$$

Integration of Eq. (85) leads to the relation

$$t_1 - t_2 = \frac{2}{n} \left[\frac{R(t_1)}{V(t_1)} - \frac{R(t_2)}{V(t_2)} \right], \quad (86)$$

or equivalently,

$$t_1 - t_2 = \frac{2}{n} \left[\frac{1}{H(t_1)} - \frac{1}{H(t_2)} \right]. \quad (87)$$

We can express $H(t)$ in terms of $\rho(t)$ and find that

$$t_1 - t_2 = \frac{2}{n} \sqrt{\frac{3}{8\pi G}} \left[\frac{1}{\sqrt{\rho(t_1)}} - \frac{1}{\sqrt{\rho(t_2)}} \right]. \quad (88)$$

Hence, whatever the value of n , the time elapsed is proportional to the change in the inverse square root of the density. This general result can also be expressed more simply by saying that the time required for the density to drop to a value ρ from some value very much greater than ρ is

$$t = \frac{2}{n} \sqrt{\frac{3}{8\pi G \rho}} = \begin{cases} 1/2 t_{\text{exp}} & \text{radiation - dominated} \\ 2/3 t_{\text{exp}} & \text{matter - dominated} \end{cases},$$

where we have neglected the second term in Eq. (88) because $\rho(t_2) \gg \rho(t_1)$.

In summary, we have shown that the time required for the density of the universe to drop to a value ρ from much higher earlier values is proportional to $1/\sqrt{\rho}$, while the density $\rho \propto R^{-n}$. The time is therefore proportional to $R^{n/2}$, or equivalently

$$R \propto t^{2/n} = \begin{cases} t^{1/2} & \text{radiation - dominated era} \\ t^{2/3} & \text{matter - dominated era} \end{cases}. \quad (89)$$

This remains valid until the kinetic and potential energies have both decreased so much that they are beginning to be comparable to their sum, the total energy.

The universe then has a sort of horizon, which shrinks rapidly as we look back toward the beginning. No signal can travel faster than the speed of light, so at any time we can only be affected by events occurring close enough so

that a ray of light would have had time to reach us since the beginning of the universe. Any event that occurred beyond this distance could as yet have no effect on us – it is beyond the horizon. Note that as we look back towards the beginning, the distance to the horizon shrinks faster than the size of the universe. As can be read in Eq. (89) the size of the universe is proportional to the one-half or two-thirds power of the time, whereas the distance to the horizon is simply proportional to the time. Therefore, for earlier and earlier times, the horizon encloses a smaller and smaller portion of the universe. This implies that at a sufficiently early time any given “typical” particle is beyond the horizon.

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 discuss the current state of the art in High Energy Physics.

XV. ELEMENTARY PARTICLES

A. The Four Forces in Nature

Since the years after World War II, particle accelerators have been a principal means of investigating the structure of nuclei. The accelerated particles are projectiles that probe the interior of the nuclei they strike and their constituents. An important factor is that faster moving projectiles can reveal more detail about the nuclei. The wavelength of the incoming particles is given by de Broglie’s wavelength formula $\lambda = h/p$, showing that the greater the momentum p of the bombarding particle, the shorter the wavelength and the finer the detail that can be obtained. (Here h is Planck’s constant.)

The accepted model for elementary particle physics today views quarks and leptons as the basic constituents of ordinary matter. To understand our present-day view we need to begin with the ideas up to its formulation. While the classic discoveries of Thomson [35] (the electron) and Rutherford [36] (the proton) opened successive doors to subatomic and nuclear physics, particle physics may be said to have started with the discovery of the positron in cosmic rays by Carl Anderson at Pasadena in 1932 [37], verifying Paul Dirac’s almost simultaneous prediction of its existence [38]. By the mid 1930s, it was recognized that all atoms can be considered to be made up of neutrons, protons and electrons. The structure of matter seemed fairly simple in 1935 (with a total of six elementary particles: the proton, the neutron, the electron, the positron, the neutrino and the photon), but in the decades that followed, hundreds of other elementary particles were discovered. The properties and interactions of these particles, and which ones should be considered as basic or elementary, is the essence of *elementary*

particle physics.

In 1935 Hideki Yukawa predicted the existence of a new particle that would in some way mediate the strong nuclear force [39]. In analogy to photon exchange that mediates the electromagnetic force, Yukawa argued that there ought to be a particle that mediates the strong nuclear force – the force that holds nucleons together in the nucleus. Just as the photon is called the quantum of the electromagnetic force, the Yukawa particle would represent the quantum of the strong nuclear force. Yukawa predicted that this new particle would have a mass intermediate between that of the electron and the proton. Hence it was called a meson (meaning “in the middle”). We can make a rough estimate of the mass of the meson as follows. For a nucleon at rest to emit a meson would require energy (to make the mass) that would have to come from nowhere, and so such a process would violate conservation of energy. But the uncertainty principle allows non-conservation of energy of an amount ΔE if it occurs only for a time Δt given by

$$\Delta E \Delta t \geq \hbar = h/2\pi . \quad (90)$$

We set $\Delta E = mc^2$, the energy needed to create the mass m of the meson. Now conservation of energy is violated only as long as the meson exists, which is the time Δt required for the meson to pass from one nucleon to the other. If we assume that the meson travels at relativistic speeds, close to the speed of light c , then Δt would be at most about $\Delta t = d/c$, where $d \approx 1.5 \times 10^{-15}$ cm is the maximum distance that can separate interacting nucleons. Thus, replacing in Eq. (90) we have

$$mc^2 \approx 2.2 \times 10^{-11} \text{ J} = 130 \text{ MeV} . \quad (91)$$

The mass of the predicted meson is thus very roughly 130 MeV/ c^2 or about 250 times the electron mass, $m_e \approx 0.51 \text{ MeV}/c^2$. (Note, incidentally, that the electromagnetic force that has an infinite range, $d = \infty$, requires a massless mediator, which is indeed the case of the photon.)

Just as photons can be observed as free particles, as well as acting in an exchange, so it was expected that mesons might be observed directly. Such a meson was searched on the cosmic radiation that enters the Earth’s atmosphere. In 1937 a new particle was discovered whose mass is 106 MeV (207 times the electron mass) [40]. This is quite close to the mass predicted, but this new particle called the muon, did not interact strongly with matter. Then it cannot mediate the strong nuclear force if it does not interact by means of the strong nuclear force. Thus, the muon, which can have either a + or – charge and seems to be nothing more than a very massive electron, is not the Yukawa particle. The particle predicted by Yukawa was finally found in 1947 [41]. It is called the pion (π). It comes in three charged states, +, –, or 0. The π^+ and π^- have a mass of 140 MeV and the π^0 a mass of 135 MeV. All three interact strongly with matter. Soon after their discovery in cosmic rays, pions

TABLE I: Relative strength for two protons in a nucleus of the four forces in nature.

Type	Relative Strength	Field Particle
Strong Nuclear	1	gluons
Electromagnetic	10^{-2}	photon
Weak Nuclear	10^{-6}	$W^\pm Z^0$
Gravitational	10^{-38}	graviton (?)

were produced in the laboratory using a particle accelerator [42]. Reactions observed included $pp \rightarrow pp\pi^0$ and $pp \rightarrow pn\pi^+$. The incident proton from the accelerator must have sufficient energy to produce the additional mass of the pion.

Yukawa's theory of pion exchange as the carrier of the strong force is now out of date, and has been replaced by *quantum chromodynamics* in which the basic entities are quarks, and the carriers of the strong force are gluons. However, the basic idea of Yukawa's theory, i.e., that forces can be understood as the exchange of particles remains valid.

There are four known types of force – or interaction – in nature. The electromagnetic force is carried by the photon, the strong force by gluons. What about the other two: the weak nuclear force and gravity? These two are also mediated by particles. The particles that transmit the weak force are referred to as the W^+ , W^- , and Z^0 , and were detected in 1983 [43]. The quantum (or carrier) of the gravitational force is called the graviton, and has not yet been observed.

A comparison of the four forces is given in Table I, where they are listed according to their (approximate) relative strengths. Note that although gravity may be the most obvious force in daily life (because of the huge mass of the Earth), on a nuclear scale it is the weakest of the four forces, and its effect at the particle level can nearly always be ignored.

B. Particle's Zoo

In the decades following the discovery of the pion, a great many other subnuclear particles were discovered. One way of arranging the particles in categories is according to their interactions, since not all particles interact by means of all four of the forces known in nature (though all interact via gravity).

The gauge bosons include the gluons, the photon and the W and Z particles; these are the particles that mediate the strong, electromagnetic, and weak interactions respectively. The leptons are particles that do not interact via the strong force but do interact via the weak nuclear force. The leptons include the electron e^- , the electron neutrino ν_e , the muon μ^- , the muon neutrino ν_μ , the tau τ^- , the tau neutrino ν_τ and their corresponding antiparticles (e^+ , $\bar{\nu}_e$, μ^+ , $\bar{\nu}_\mu$, τ^+ , $\bar{\nu}_\tau$). The third category of particles is the hadron. Hadrons are those particles

that interact via the strong nuclear force. Therefore they are said to be strongly interacting particles. They also interact via the other forces, but the strong force predominates at short distances. The hadrons include the proton, the neutron, the pion, the kaon K , the Λ , and a large number of other particles. Hadrons can be divided into two subgroups: baryons, which are those particles that have baryon number $B = +1$ (or $B = -1$ in case of their antiparticles), and mesons, which have $B = 0$.

Conservation laws are indispensable in ordering this subnuclear world. The laws of conservation of energy, of momentum, of angular momentum, of electric charge, and baryon number are found to hold precisely in all particle interactions. Also useful are the conservation laws for the three lepton numbers (for a given lepton l , l^- and ν_l are assigned $L_l = +1$ and l^+ and $\bar{\nu}_l$ are assigned $L_l = -1$, whereas all other particles have $L_l = 0$), associated with weak interactions including decays. Nevertheless, an important result has come to the fore in our young XXI century: neutrinos can occasionally change into one another in certain circumstances, a phenomenon called neutrino flavor oscillation [44]. This result suggests that neutrinos are not massless particles and that the lepton numbers L_e , L_μ , and L_τ are not perfectly conserved. The sum $L_e + L_\mu + L_\tau$, however, is believed to be always conserved.

The lifetime of an unstable particle depends on which force is more active in causing the decay. When a stronger force influences the decay, that decay occurs more quickly. Decays caused by the weak force typically have lifetimes of 10^{-13} s or longer (W and Z are exceptions). Decays via the electromagnetic force have much shorter lifetimes, typically about 10^{-16} to 10^{-19} s and normally involve a photon. Hundreds of particles have been found with very short lifetimes, typically 10^{-23} s, that decay via the strong interaction. Their lifetimes are so short that they do not travel far enough to be detected before decaying. The existence of such short-lived particles is inferred from their decay products. In 1952, using a beam of π^+ with varying amounts of energy directed through a hydrogen target (protons), Enrico Fermi [45] found that the number of interactions (π^+ scattered) when plotted versus the pion kinetic energy had a prominent peak around 200 MeV. This led Fermi to conclude that the π^+ and proton combined momentarily to form a short-lived particle, before coming apart again, or at least that they resonated together for a short time. This “new particle” – so called the Δ^{++} – is referred to as a resonance. Since then many other resonances have been found. The width of these excited states is an interesting application of the uncertainty principle. If the particle only lives 10^{-23} s, then its mass will be uncertain by an amount $\Delta E \approx \hbar/\Delta t \approx 10^{-11}$ J ≈ 100 MeV, which is what is observed. Indeed the lifetimes of $\sim 10^{-23}$ s for such resonances are inferred by the reverse process, i.e., from the measured width being ~ 100 MeV.

In the early 1950s, the newly found particles K^0 -meson and Λ^0 -baryon were found to behave strangely in two

ways. (a) They were always produced in pairs; for example the reaction $\pi^- p \rightarrow K^0 \Lambda^0$ occurred with high probability, but the similar reaction $\pi^- p \rightarrow K^0 n$ was never observed to occur. (b) These strange particles, as they came to be called, were produced via the strong interaction (i.e., at high rate), but did not decay at a fast rate characteristic of the strong interaction (even though they decay into strongly interacting particles).

To explain these observations, a new quantum number, *strangeness*, and a new conservation law, *conservation of strangeness* were introduced. By assigning the strangeness numbers $S = +1$ for the kaon and $S = -1$ for the Λ , the production of strange particles in pairs was explained. Antiparticles were assigned opposite strangeness from their particles. Note that for the reaction $\pi^- p \rightarrow K^0 \Lambda^0$, both the initial and final states have $S = 0$ and hence strangeness is conserved. However, for $\pi^- p \rightarrow K^0 n$, the initial state has $S = 0$ and the final state has $S = 1$, so strangeness would not be conserved; and this reaction is not observed.

To explain the decay of strange particles, it is postulated that strangeness is conserved in the strong interaction, but is not conserved in the weak interaction. Therefore, strange particles are forbidden by strangeness conservation to decay to non-strange particles of lower mass via the strong interaction, but could decay by means of the weak interaction at the observed longer lifetimes of 10^{-10} to 10^{-8} s. The conservation of strangeness was the first example of a partially conserved quantity.

C. The “Standard Model”

All particles, except the gauge bosons, are either leptons or hadrons. The principal difference between these two groups is that hadrons interact via the strong interaction, whereas the leptons do not.

The six leptons are considered to be truly elementary particles because they do not show any internal structure, and have no measurable size. (Attempts to determine the size of the leptons have put an upper limit of about 10^{-18} m.)

On the other hand, there are hundreds of hadrons and experiments indicate they do have internal structure. In the early 1960s, Murray Gell-Mann [46] and George Zweig [47] proposed that none of the hadrons are truly elementary, but instead are made up of combinations of three pointlike entities called, somewhat whimsically, quarks. Today quarks are considered the truly elementary particles like leptons. The three quarks originally proposed were labeled, u , d , s and have the names up, down, and strange; respectively. At present the theory has six quarks, just as there are six leptons – based on a presumed symmetry in nature. The other three quarks are called charmed, bottom, and top (labeled c , b , and t). The names apply also to new properties of each that distinguish the new quarks from the old quarks and which (like strangeness) are conserved in strong, but

TABLE II: Quark quantum numbers: charge Q , baryon number B , strangeness S , charm c , “beauty” or bottomness b , and “truth” or topness t

name	symbol	Q	B	S	c	b	t
up	u	$\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	0
down	d	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	0	0
strange	s	$-\frac{1}{3}e$	$\frac{1}{3}$	-1	0	0	0
charm	c	$\frac{2}{3}e$	$\frac{1}{3}$	0	1	0	0
bottom	b	$-\frac{1}{3}e$	$\frac{1}{3}$	0	0	-1	0
top	t	$\frac{2}{3}e$	$\frac{1}{3}$	0	0	0	1

not in weak interactions. The properties of the quarks are given in Table II. Antiquarks have opposite signs of electric charge, baryon number, strangeness, charm, bottomness, and topness.

All hadrons are considered to be made up of combinations of quarks, and their properties are described by looking at their quark content. Mesons consist of a quark-antiquark pair. For example, a π^+ meson is a $u\bar{d}$ combination: note that for $u\bar{d}$ pair $Q = (2/3 + 1/3)e = e$, $B = 1/3 - 1/3 = 0$, $S = 0 + 0$, as they must for the π^+ ; and a K^+ = $u\bar{s}$, with $Q = 1$, $B = 0$, $S = 1$. Baryons on the other hand, consist of three quarks. For example, a neutron is $n = ddu$, whereas an antiproton is $\bar{p} = \bar{u}\bar{u}\bar{d}$.

The truly elementary particles are the leptons, the quarks, and the gauge bosons. Matter is made up mainly of fermions which obey the Pauli exclusion principle, but the carriers of the forces are all bosons (which do not).

Quarks are fermions with spin 1/2 and therefore should obey the exclusion principle. Yet for three particular baryons ($\Delta^{++} = uuu$, $\Delta^- = ddd$, and $\Omega^- = sss$), all three quarks would have the same quantum numbers, and at least two quarks have their spin in the same direction because there are only two choices, spin up ($m_s = +1/2$) or spin down ($m_s = -1/2$). This would seem to violate the exclusion principle!

Not long after the quark theory was proposed, it was suggested that quarks have another property called color, or “color charge” (analogous to the electric charge). The distinction between the six quarks (u, d, s, c, b, t) was referred to as flavor. Each of the flavors of quark can have three colors usually designated red, green, and blue. The antiquarks are colored antired, antigreen, and antiblue. Baryons are made up of three quarks, one with each color. Mesons consist of a quark-antiquark pair of a particular color and its anticolor. Both baryons and mesons are thus colorless or white. Because the color is different for each quark, it serves to distinguish them and allows the exclusion principle to hold. Even though quark color was originally an *ad hoc* idea, it soon became the central feature of the theory determining the force binding quarks together in a hadron.

Each quark is assumed to carry a color charge, analogous to the electric charge, and the strong force between quarks is referred to as the color force. This theory of the

strong force is called quantum chromodynamics (QCD). The particles that transmit the color force are called the gluons. There are eight gluons, all massless and all have color charge. Thus gluons have replaced mesons as particles responsible for the strong (color) force.

One may wonder what would happen if we try to see a single quark with color by reaching deep inside a hadron. Quarks are so tightly bound to other quarks that extracting one would require a tremendous amount of energy, so much that it would be sufficient to create more quarks. Indeed, such experiments are done at modern particle colliders and all we get is not an isolated quark, but more hadrons (quark-antiquark pairs or triplets). This property of quarks, that they are always bound in groups that are colorless, is called confinement. Moreover, the color force has the interesting property that, as two quarks approach each other very closely (or equivalently have high energy), the force between them becomes small. This aspect is referred to as asymptotic freedom [48].

The weak force, as we have seen, is thought to be mediated by the W^+ , W^- , and Z^0 particles. It acts between the “weak charges” that each particle has. Each elementary particle can thus have electric charge, weak charge, color charge, and gravitational mass, although one or more of these could be zero.

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 anti-particle, whereas the 3 neutrinos and antineutrinos have only one chirality (neutrinos are Left-handed and antineutrinos are Right-handed).⁵ 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.

One important aspect of on-going research is the attempt to find a unified basis for the different forces in nature. A so-called gauge theory that unifies the weak and electromagnetic interactions was put forward in the 1960s by Sheldon Lee Glashow [49], Abdus Salam [50], and Steven Weinberg [51]. In this electroweak theory, the weak and electromagnetic forces are seen as two different manifestations of a single, more fundamental *electroweak* interaction. The electroweak theory has had many successes, including the prediction of the $W^\pm$

particles as carriers of the weak force, with masses of $81 \pm 2 \text{ GeV}/c^2$ in excellent agreement with the measured values of $80.482 \pm 0.091 \text{ GeV}/c^2$ [52] (and similar accuracy for the Z^0), and the prediction of the mass of the top quark $m_t = 174.1^{+9.7}_{-7.6} \text{ GeV}/c^2$ in remarkable agreement with the top mass measured at the Tevatron, $m_t = 178.0 \pm 4.3 \text{ GeV}/c^2$ [53]. The combination of the electroweak theory plus QCD for the strong interaction is referred today as the *Standard Model* (SM).

Another intriguing aspect of particle physics is to explain why the W and Z have large masses rather than being massless like the photon. Electroweak theory suggests an explanation by means of the Higgs boson, which interacts with the Z and the W to “slow them down.” In being forced to go slower than the speed of light, they must acquire mass. The search for the Higgs boson has been a priority in particle physics. So far, searches have excluded a Higgs lighter than $115 \text{ GeV}/c^2$ [55]. Yet it is expected to have a mass no larger than $200 \text{ GeV}/c^2$. We are narrowing in on it.

Summing up, the SM is conceptually simple and contains a description of the elementary particles and forces. The SM particles are 12 spin-1/2 fermions (6 quarks and 6 leptons), 4 spin-1 gauge bosons and a spin-0 Higgs boson. Seven of the 16 particles (charm, bottom, top, tau neutrino, W , Z , gluon) were predicted by the SM before they were observed experimentally! There is only one particle predicted by the SM which has not yet been observed [54].

D. Beyond the Standard Model

With the success of the unified electroweak theory, attempts are being made to incorporate it and QCD for the strong (color) force into a so-called grand unified theory (GUT). One type of GUT has been worked out, in which leptons and quarks belong to the same family (and are able to change freely from one type to the other) and the three forces are different aspects of the underlying force [56]. The unity is predicted to occur, however, only on a scale less than about 10^{-32} m , corresponding to an extremely high energy of about 10^{16} GeV . If two elementary particles (leptons or quarks) approach each other to within this unification scale, the apparently fundamental distinction between them would not exist at this level, and a quark could readily change to a lepton, or vice versa. Baryon and lepton numbers would not be conserved. The weak, electromagnetic, and strong (color) force would blend to a force of a single strength.

The unification of the three gauge coupling constants of the SM is sensitive to the particle content of the theory. Indeed if the minimal supersymmetric extension of the SM [57] is used, the match of the gauge coupling constants becomes much more accurate [58]. It is commonly believed that this matching is unlikely to be a coincidence. Supersymmetry (SUSY) predicts that interactions exist that would change fermions into bosons

⁵ 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. By convention for rotation, a standard clock, tossed with its face directed forwards, has Left-handed chirality.

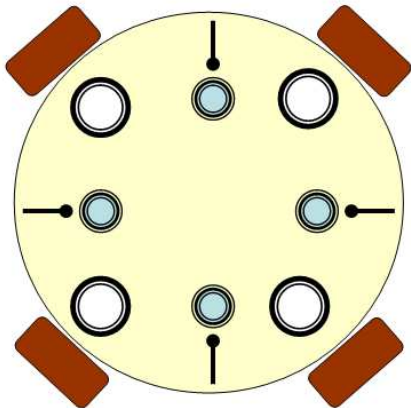


FIG. 12: Symmetry around a table.

and vice versa [59], and that all known fermions have a supersymmetric boson partner. Thus, for each quark we know (a fermion), there would be a squark (a boson) or SUSY quark. For every lepton there would be a slepton. Likewise for every gauge boson (photons and gluons for example), there would be a SUSY fermion. But why haven't all these particles been detected. The best guess is that SUSY particles may be heavier than their conventional counterparts, perhaps too heavy to have been produced in today's accelerators. *Until a supersymmetric particle is found, and this may be possible in the now coming-on-line Large Hadron Collider (LHC), SUSY is just an elegant guess.*

What happens between the unification distance of 10^{-32} m and more normal (larger) distances is referred to as symmetry breaking. As an analogy, consider a table that has four identical place settings as shown in Fig. 12. The table has several kinds of symmetries. For example, it is symmetric to rotations of 90° degrees; that is, the table will look the same if everyone moved one chair to the left or to the right. It is also north-south symmetric and east-west symmetric, so that swaps across the table do not affect the way the table looks. It also does not matter whether any person picks up the spoon to the left of the plate or the spoon to the right. However, once the first person picks up either spoon, the choice is set for all the rest of the table as well. The symmetry has been broken. The underlying symmetry is still there – the blue glasses could still be chosen either way – but some choice must be made, and at that moment, the symmetry of the diners is broken.

Since unification occurs at such tiny distances and huge energies, the theory is difficult to test experimentally. But it is not completely impossible. One testable prediction is the idea that the proton might decay (via, for example $p \rightarrow \pi^0 e^+$) and violate conservation of baryon number. This could happen if two quarks approached to within 10^{-31} m of each other. This is very unlikely at nor-

mal temperature and energy; consequently the decay of a proton can only be an unlikely process. In the simplest form of GUT, the theoretical estimate of the proton lifetime for the decay mode $p \rightarrow \pi^0 e^+$ is about 10^{31} yr, and this has just come within the realm of testability. Proton decays have still not been seen, and experiments put a lower limit on the proton lifetime for the above mode to be about 10^{33} yr [60], somewhat greater than the theoretical prediction. This may seem a disappointment, but on the other hand, it presents a challenge. Indeed more complex GUTs are not affected by this result.

Even more ambitious than GUTs are the attempts to also incorporate gravity, and thus unify all forces of nature into a single theory. Superstring theory [61], in which elementary particles are imagined not as points, but as one-dimensional strings (perhaps 10^{-35} m long), is at present the best hope for unification of all forces.⁶

XVI. THE EARLY UNIVERSE

In today's class, 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 may be helpful to consult Fig. 13 as we go along.

We begin at a time only a minuscule fraction of a second after the Big Bang: it is thought that prior to 10^{-35} s, perhaps as early as 10^{-44} s, the four forces of nature were unified – the realm Superstring theory. This is an unimaginably short time, and predictions can be only speculative.⁷ The temperature would have been about 10^{32} K, corresponding to “particles” moving about every which way with an average kinetic energy of

$$K \approx kT \approx \frac{1.4 \times 10^{-23} \text{ J/K } 10^{32} \text{ K}}{1.6 \times 10^{-10} \text{ J/GeV}} \approx 10^{19} \text{ GeV}, \quad (92)$$

where we have ignored the factor $2/3$ in our order of mag-

⁶ A point worth noting at this juncture: Very recently a new framework with a diametrically opposite viewpoint was put forward. The new premise suggests that the weakness of gravity may be evidence from large extra spatial compactified dimensions [62]. This is possible because SM particles are confined to a 4-dimensional world (corresponding to our apparent universe) and only gravity spills into the higher dimensional spacetime. Hence, if this picture is correct, gravity is not intrinsically weak, but of course appears weak at relatively large distances of common experience because its effects are diluted by propagation in the extra dimensions. The distance at which the gravitational and electromagnetic forces might have equal strength is unknown, but a particularly interesting possibility is that it is around 10^{-19} m, the distance at which electromagnetic and weak forces are known to unify to form the electroweak force. This would imply a fundamental Planck mass, $M_* \sim M_W \sim 1 \text{ TeV}$, at the reach of experiment [63].

⁷ Though for the moment this is a matter of conjecture, it is appealing that Superstring theory predicts a concrete candidate for the quintessence field that can drive the acceleration of the universe in the present epoch [64].

nitude calculation. At $t = 10^{-44}$ s, a kind of “phase transition” is believed to have occurred during which the gravitational force, in effect, “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 the universe entered the grand unified era. There were no distinctions between quarks and leptons; baryon and lepton numbers were not conserved. Very shortly thereafter, as the universe expanded considerably and the temperature had dropped to about 10^{27} K, there was another phase transition and the strong force condensed out at about 10^{-35} s after the Big Bang. Now the universe was filled with a “soup” of leptons and quarks. The quarks were initially free, but soon began to “condense” into mesons and baryons. With this confinement of quarks, the universe entered the hadron era.

About this time, the universe underwent an incredible exponential expansion, increasing in size by a factor of 10^{40} or 10^{50} in a tiny fraction of a second, perhaps 10^{-32} s. The usefulness of this inflationary scenario is that it solved major problems with earlier Big Bang models, such as explaining why the universe is flat, as well as the thermal equilibrium to provide the nearly uniform CMB [65]. Inflation is now a generally accepted aspect of the Big Bang theory.

After the very brief inflationary period, the universe would have settled back into its more regular expansion. The universe was now a “soup” of leptons and hadrons. We can think of this soup as a grand mixture of particles and antiparticles, as well as photons – all in roughly equal numbers – colliding with one another frequently and exchanging energy.

By the time the universe was only about a microsecond (10^{-6} s) old, it had cooled to about 10^{13} K, corresponding to an average kinetic energy of 1 GeV, and the vast majority of hadrons disappeared. 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! To explain our present world, which consists mainly of matter with very little antimatter in sight, we must suppose that earlier in the universe, perhaps around 10^{-35} s after the Big Bang, a slight excess of quarks over antiquarks was formed. This

would have resulted in a slight excess of nucleons over antinucleons, and it is these “leftover” nucleons that we are made of today. The excess of nucleons over antinucleons was about one part in 10^9 . Earlier, during the hadron era there should have been about as many nucleons as photons. After it ended, the “leftover” nucleons thus numbered only about one nucleon per 10^9 photons, and this ratio has persisted to this day. Protons, neutrons, and all other heavier particles were thus tremendously reduced in number by about 10^{-6} s after the Big Bang. The lightest hadrons, the pions, were the last ones to go, about 10^{-4} s after the Big Bang. Lighter particles (those whose threshold temperature are below 10^{11} K), including electrons, positrons, and neutrinos, were the dominant forms of matter, and the universe entered the lepton era. The universe was so dense that even the weakly interacting (anti)neutrinos (that can travel for years through lead bricks without being scattered) were kept in thermal equilibrium with the electrons, positrons and photons by rapid collision with them and with each other.

When the temperature of the universe is 3×10^{10} K, 0.11 s have elapsed. Nothing has changed qualitatively – the contents of the universe are still dominated by electrons, positrons, neutrinos, antineutrinos, and photons, all in thermal equilibrium, and all high above their threshold temperatures. By the time the first second has past (certainly the most eventful second in history!), the universe has cooled to about 10^{10} K. The average kinetic energy is 1 MeV. This is still sufficient energy to create electrons and positrons and balance their annihilation reactions, since their masses correspond to about 0.5 MeV. However, about this time the decreasing density and temperature have increased the mean free path of neutrinos and antineutrinos so much that they are beginning to behave like free particles, no longer in thermal equilibrium with the electrons, positrons, or photons. From now on neutrinos will cease to play an active role in our discussion, except that their energy will continue to provide part of the source of the gravitational field of the universe.

After 13.8 s, the temperature $\sim 3 \times 10^9$ K had dropped sufficiently so that e^+e^- could no longer be formed. Annihilation ($e^+e^- \rightarrow$ photons) continued. Electrons and positrons have now also disappeared from the universe, except for a slight excess of electrons over positrons (later to join with nuclei to form atoms). Thus, about 14 seconds after the Big Bang, the universe entered the radiation era. Its major constituents were photons and neutrinos, but neutrinos partaking only in the weak force, rarely interacted.

As long as thermal equilibrium was preserved, the total entropy remained fixed. The entropy per unit volume is found to be: $S \propto N_T T^3$, where N_T is the effective number of species (total number of degrees of freedom) of particles in thermal equilibrium whose threshold temperatures are below T . In order to keep the total entropy constant, S must be proportional to the inverse cube of the size

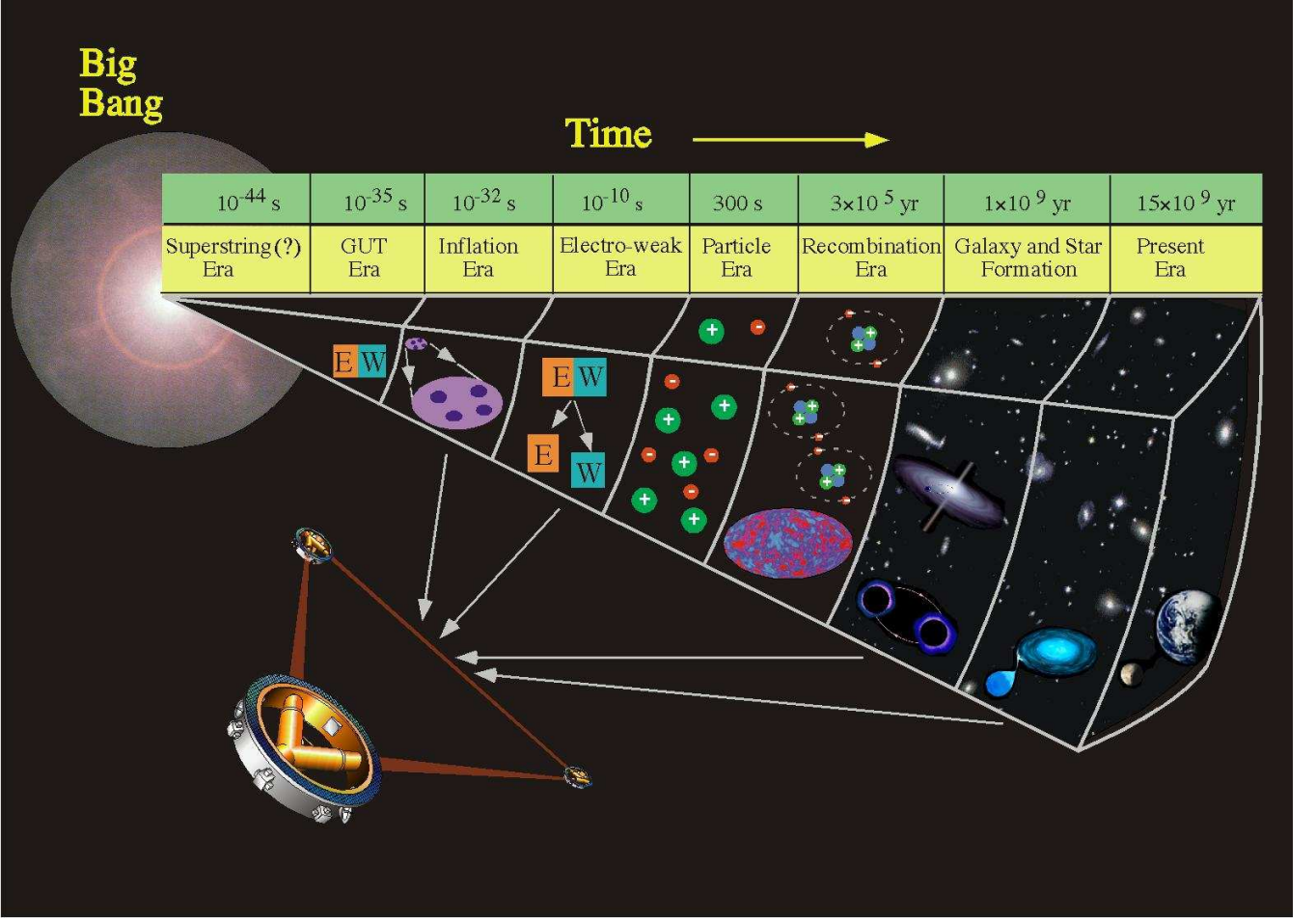


FIG. 13: Compressed graphical representation of events as the universe expanded and cooled after the Big Bang. The Laser Interferometer Space Antenna (LISA) mission, whose launch is envisaged for 2013, is also shown. LISA will be a trio of spacecrafts orbiting the Sun trying to observe gravitational waves by using laser interferometry over astronomical distances.

of the universe. That is, if R is the separation between any pair of typical particles, then $SR^3 \propto N_T T^3 R^3 = \text{constant}$.

Just before the annihilation of electrons and positrons (at about 5×10^9 K) the neutrinos and antineutrinos had already gone out of thermal equilibrium with the rest of the universe, so the only abundant particles in equilibrium were the electrons, positrons and photons. The effective total number of particle species before annihilation were $N_{\text{before}} = 6$. On the other hand, after the annihilation of electrons and positrons, the only remaining abundant particles in equilibrium were photons. Hence the effective number of particle species was $N_{\text{after}} = 2$. It follows from the conservation of entropy that

$$6 (TR)^3|_{\text{before}} = 2 (TR)^3|_{\text{after}}. \quad (93)$$

That is, the heat produced by the annihilation of electrons and positrons increases the quantity TR by a factor

$$\frac{(TR)|_{\text{after}}}{(TR)|_{\text{before}}} = 3^{1/3} \simeq 1.4. \quad (94)$$

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

$$(T_\nu R)|_{\text{after}} = (T_\nu R)|_{\text{before}} = (TR)|_{\text{before}}. \quad (95)$$

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

$$\left(\frac{T}{T_\nu}\right)|_{\text{after}} = \frac{(TR)|_{\text{after}}}{(T_\nu R)|_{\text{after}}} \simeq 1.4. \quad (96)$$

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

$$u_\nu/u_\gamma = 3^{-4/3} \quad 3 = 0.7. \quad (97)$$

From Eq. (44) we obtain the photon energy density at photon temperature T ; hence the total energy density after electron positron annihilation is

$$u_\nu + u_\gamma = 1.7u_\gamma \simeq 1.3 \times 10^{-14} (T/\text{K})^4 \text{ erg/cm}^3. \quad (98)$$

We can convert this to an equivalent mass density,

$$\rho = u/c^2 \simeq 1.22 \times 10^{-35} (T/\text{K})^4 \text{ g/cm}^3. \quad (99)$$

We have seen that the baryonic matter is about 4% of the total energy in the universe $\rho_b \approx 0.04 \rho_c = 0.04 \times 10^{-26} \text{ kg/m}^3$. Using the proton as typical baryonic matter ($m_p = 1.67 \times 10^{-27} \text{ kg}$), this implies that the number density of baryons is $n_b = 0.24 \text{ nucleons/m}^3$. From Eqs. (48) and (97) we see that the neutrino number density is about 10^9 that of nucleons, i.e., $n_\nu = 2.4 \times 10^8 \text{ neutrinos/m}^3$. By assuming that neutrinos saturate the dark matter density we can set an upper bound on the neutrino mass

$$m_\nu < \frac{0.26 \rho_c}{n_\nu} \approx \frac{2.6 \times 10^{-27} \text{ kg/m}^3}{2.4 \times 10^8 \text{ neutrino/m}^3} \approx 10^{-35} \frac{\text{kg}}{\text{neutrino}} \times \frac{9.315 \times 10^8 \text{ eV}/c^2}{1.67 \times 10^{-27} \text{ kg}} \sim 6 \text{ eV}/c^2. \quad (100)$$

Meanwhile, during the next few minutes, crucial events were taking place. Beginning about 2 or 3 minutes after the Big Bang, nuclear fusion began to occur. The temperature had dropped to about 10^9 K , corresponding to an average kinetic energy $\langle K \rangle \approx 100 \text{ keV}$, where nucleons can strike each other and be able to fuse, but now cool enough so that newly formed nuclei would not be broken apart by subsequent collisions. Deuterium, helium, and very tiny amounts of lithium nuclei were probably made. Because the universe was cooling too quickly, larger nuclei were not made. After only a few minutes, probably not a quarter of an hour after the Big Bang, the temperature dropped far enough that nucleosynthesis stopped, not to start again for millions of years (in stars). Thus, after the first hour or so of the universe, matter consisted mainly of bare nuclei of hydrogen (about 75%) and helium (about 25%) and electrons. Nevertheless radiation continues to dominate.

The evolution of the early universe is almost complete. The next important event is presumed to have occurred 380,000 years later. The universe had expanded to about 1/1000 of its present size, and the temperature had cooled to about 3000 K. The average kinetic energy of nuclei, electrons, and photons was less than 1 eV. Since ionization energies of atoms are $\mathcal{O}(\text{eV})$, as the temperature dropped below this point, electrons could orbit the bare nuclei and remain there (without being ejected by collisions), thus forming atoms. With the birth of atoms, the photons – which had been continually scattering from free electrons – now became free to spread unhindered throughout the universe, i.e., the photons became decoupled from matter. The total energy contained in radiation had been decreasing (lengthening in wavelength as the universe expanded), and the total energy contained in matter became dominant. The universe was said to have become matter-dominated. As the universe con-

tinued to expand the photons cooled further, to 2.7 K today, forming the CMB we detect from everywhere in the universe.

After the birth of atoms, stars and galaxies begin to form – presumably by self-gravitation around mass concentrations (inhomogeneities). Stars began to form about 200 million years after the Big Bang, galaxies after almost 10^9 yr . The universe continued to evolve until today, some 13.7 billion years later [66].

XVII. WIMPs

At present, the only dark matter particle which is known to exist from experiment is the neutrino. As we learned in the last class, neutrinos decoupled from thermal equilibrium while still relativistic, constituting a hot dark matter candidate. Their contribution to the universe's matter balance is similar to that of light, but neutrinos play a secondary role. The role is however identifiable: because of their large mean free path, they prevent the smaller structures in the cold dark matter from fully developing. This is visible in the observed distribution of galaxies shown in Fig. 14. Simulations of structure formation with varying amounts of matter in the neutrino component can match to a variety of observations of today's sky, including measurement of galaxy-galaxy correlations and temperature fluctuations on the surface of last scattering. These analyses suggest a more restrictive limit for neutrino masses than the one derived in Eq. (100), namely $m_\nu \lesssim 1 \text{ eV}$.

The simplest model for cold dark matter consists of WIMPs - weakly interacting massive particles. Generic WIMPs were once in thermal equilibrium, but decoupled while strongly non-relativistic. The most promising

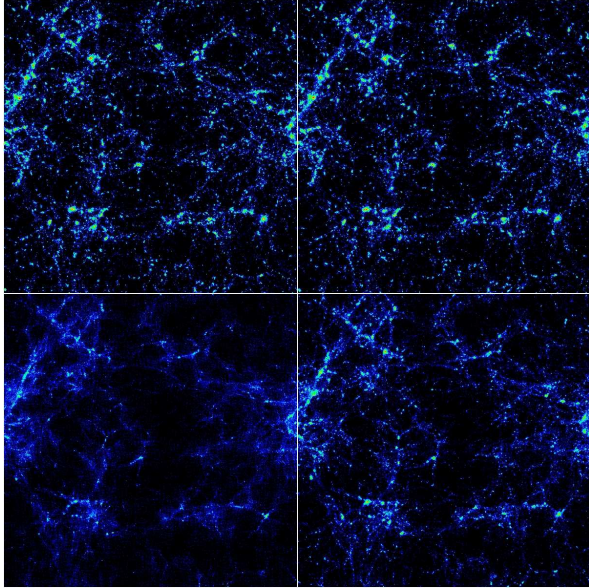


FIG. 14: Simulations of structure formation with varying amounts of matter in the neutrino component, i.e., varying neutrino mass: (top left) massless neutrinos, (top right) $m_\nu = 1$ eV, (bottom left) $m_\nu = 7$ eV, (bottom right) $m_\nu = 10$ eV.

WIMP candidate is probably the lightest SUSY particle (hereafter denoted by χ) [67]. The relic abundance of χ 's is determined by its thermally average annihilation cross section at freezeout. At high temperatures the number density of χ 's is roughly the same as the number density of photons, but as the temperature drops below the mass M_χ , the number density of WIMPs drops exponentially. This continues until the total annihilation cross section is no longer large enough to maintain equilibrium and the WIMP number density “freezes out.” The largest annihilation cross section in the early universe is expected to be roughly $\sim M_\chi^{-2}$. This implies that very massive WIMPs have such a small annihilation cross section that their present abundance would be too large. Indeed, to account for the observed non-baryonic matter the thermally average annihilation cross section at freezeout has to be larger than about 1 pb, yielding $M_\chi < 10$ TeV.

Many approaches have been developed to attempt to detect dark matter. Such endeavors include direct detection experiments which hope to observe the scattering of dark matter particles with the target material of the detector and indirect detection experiments which are designed to search for the products of WIMP annihilation into gamma-rays, anti-matter and neutrinos. The detection sensitivity of current experiments has been improving at a steady rate, and new dark matter hunters have been proposed for the coming years [68].

XVIII. MULTI-MESSENGER ASTRONOMY

A. The Photon Window

Conventional astronomy expands 18 decades in photon wavelengths, from 10^4 cm radio waves to 10^{-14} cm γ rays of GeV energy. The images of the Galactic Plane shown in Fig. 15 summarize the different wave-bands. Our rudimentary understanding of the GeV γ -ray sky was greatly advanced in 1991 with the launch of the Energetic Gamma Ray Experiment Telescope (EGRET) on board of the Compton Gamma Ray Observatory (CGRO). The science returns from EGRET observations exceed pre-launch expectations. In particular, the number of previously known GeV γ ray sources increased from 1–2 dozen to the 271 listed in the 3rd EGRET Catalog [69]. However, of this multitude of sources, only about 100 have been definitively associated with known astrophysical objects. Therefore, most of the γ ray sky, as we currently understand it, consists of unidentified objects. One of the reasons that such a small fraction of the sources were identified is the size of the typical γ -ray error box of EGRET, that was about $1^\circ \times 1^\circ$, an area that contains several candidate sources preventing straightforward identification. This leaves intriguing puzzles for the next generation of GeV γ ray instruments to uncover [70]. 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. 16). 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 comouflaging 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 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 that we discuss next.

B. Cosmic Rays

In 1912 Victor 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 [72]. His discovery of increased radiation at high altitude revealed that we are bombarded by ionizing particles from above. These cosmic ray (CR) particles are now known to consist primarily of protons, helium, carbon, nitrogen and other heavy ions up to iron.

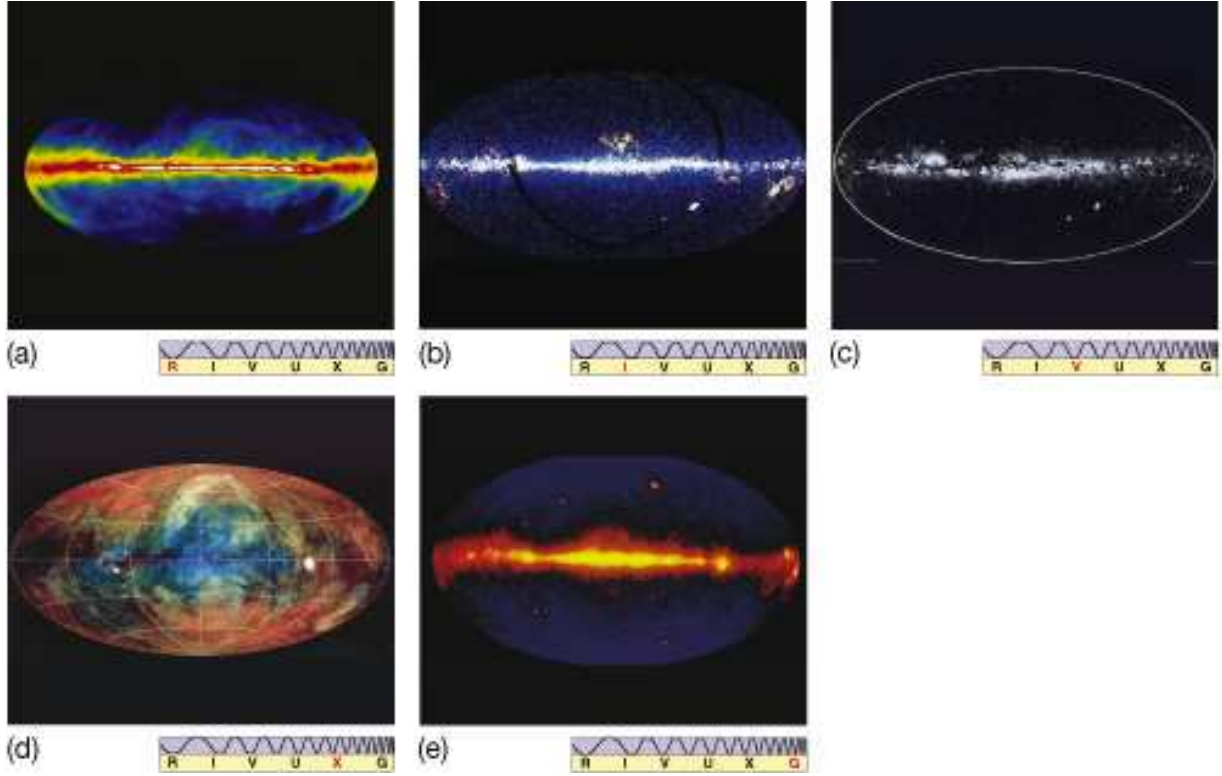


FIG. 15: The Milky Way as it appears at (a) radio, (b) infrared, (c) visible, (d) X-ray, and (e) γ -ray wavelengths. Each frame is a panoramic, view covering the entire sky. The center of the Galaxy, which lies in the direction of the constellation Sagittarius is at the center of each map. (NRAO; NASA; Lund Observatory; K. Dennerl and W. Voges; NASA.)

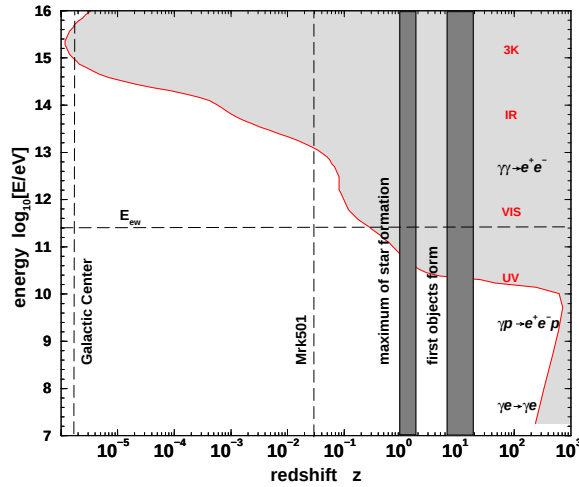


FIG. 16: Mean interaction length for photons on the ultraviolet (UV), visible (VIS), infrared (IR), and microwave backgrounds. The electroweak scale is indicated by a dashed line. The redshifts of the star formation epoch and the famous γ -ray source Markarian 501 are also indicated [71].

Below 10^{14} eV the flux of particles is sufficiently large that individual nuclei can be studied by detectors carried aloft in balloons or satellites. From such direct experi-

ments 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.

Above 10^{14} eV, 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, Pierre Auger concluded from the size of extensive air showers that the spectrum extends up to and perhaps beyond 10^{15} eV [73]. Nowadays substantial

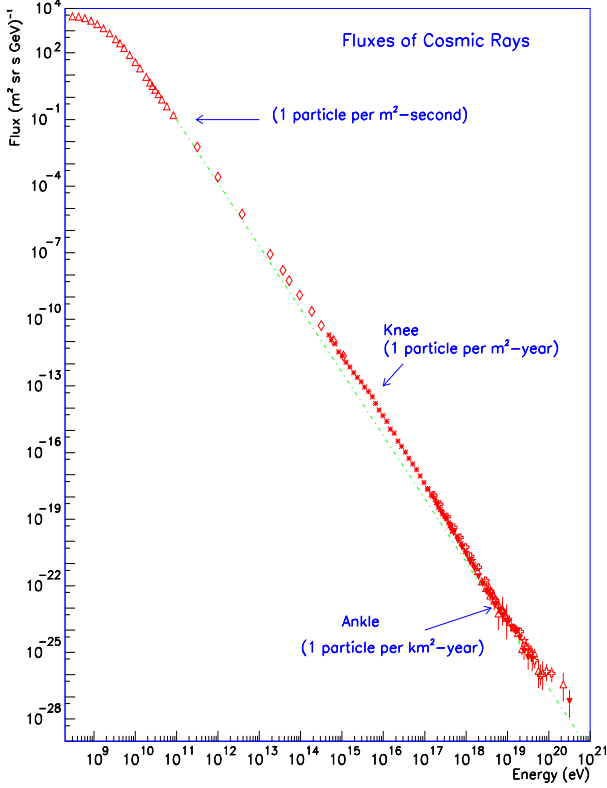


FIG. 17: Compilation of measurements of the differential energy spectrum of cosmic rays. The dotted line shows an E^{-3} power-law for comparison. Approximate integral fluxes (per steradian) are also shown.

progress has been made in measuring the extraordinarily low flux ($\sim 1 \text{ event km}^{-2} \text{ yr}^{-1}$) above 10^{19} eV . Continuously running experiments using both arrays of particle detectors on the ground and fluorescence detectors which track the cascade through the atmosphere, have detected events with primary particle energies somewhat above 10^{20} eV [74].

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. 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: $R > R_{\text{gyro}} = E/B$, i.e. the accelerating magnetic field must contain the particle's orbit. By dimensional analysis, this condition yields a maximum energy $E = \Gamma B R$. 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 $E < \Gamma B M$.

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 efficiencies matching the dimensional limit. It is highly questionable that nature can achieve this feat. Theorists can imagine acceleration in shocks with an efficiency of perhaps 1 – 10%.

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.

In contrast to the irregular shape of the isotropic electromagnetic background spectrum from, say, $10^8 - 10^{20} \text{ Hz}$, the CR energy spectrum above 10^9 eV can be described by a series of power laws, with the flux falling about 3 orders of magnitude for each decade increase in energy (see Fig. 17). In the decade centered at $\sim 10^{15.5} \text{ eV}$ (the knee) the spectrum steepens from $E^{-2.7}$ to $E^{-3.0}$. This feature, discovered around 40 years ago, is still not consistently explained. The spectrum steepens further to $E^{-3.3}$ above $\sim 10^{17.7} \text{ eV}$ (the dip) and then flattens to $E^{-2.7}$ at $\sim 10^{18.5} \text{ eV}$ (the ankle). Within the statistical uncertainty of the data, which is large around 10^{20} eV , the tail of the spectrum is consistent with a simple extrapolation at that slope to the highest energies, possibly with a hint of a slight accumulation around $10^{19.5} \text{ eV}$. A very widely held interpretation of the ankle is that above $10^{18.5} \text{ eV}$ a new population of CRs with extragalactic origin begins to dominate the more steeply falling Galactic population. The extragalactic component seems to be dominated by protons [75].

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^{19} eV 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 ,

$$\theta \simeq \frac{d}{R_{\text{gyro}}} = \frac{dB}{E}, \quad (101)$$

where d is the distance to the source. Scaled to units

relevant to the problem,

$$\frac{\theta}{0.1^\circ} \simeq \frac{(d/\text{Mpc})(B/\text{nG})}{E/10^{20.5} \text{ eV}}. \quad (102)$$

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 $\gtrsim 10^{18}$ eV have a boosted $c\tau_n$ sufficiently large to serve as Galactic messengers.⁸ The decay mean free path of a neutron is $c\Gamma_n\bar{\tau}_n = 9.15(E_n/10^9 \text{ GeV}) \text{ kpc}$, the lifetime being boosted from its rest-frame value, $\bar{\tau}_n = 886 \text{ s}$, 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.

At very high energies, however, the universe becomes opaque to the propagation of cosmic rays. Shortly after the discovery of the CMB, Greisen, Zatsepin, and Kuzmin (GZK) [77] pointed out that this photonic molasses makes the universe opaque to protons of sufficiently high energy, i.e., protons with energies beyond the photopion production threshold,

$$\begin{aligned} E_{p\gamma\text{CMB}}^{\text{th}} &= \frac{m_\pi(m_p + m_\pi/2)}{E_{\gamma\text{CMB}}} \\ &\approx 10^{20} \left(\frac{E_{\gamma\text{CMB}}}{10^{-3} \text{ eV}} \right)^{-1} \text{ eV}, \end{aligned} \quad (103)$$

where m_p (m_π) denotes the proton (pion) mass and $E_{\gamma\text{CMB}} \sim 10^{-3} \text{ eV}$ is a typical CMB photon energy. After pion production, the proton (or perhaps, instead, a neutron) emerges with at least 50% of the incoming energy. This implies that the nucleon energy changes by an e -folding after a propagation distance $\lesssim (\sigma_{p\gamma} n_\gamma y)^{-1} \sim 15 \text{ Mpc}$. Here, $n_\gamma \approx 400 \text{ cm}^{-3}$ is the number density of the CMB photons, $\sigma_{p\gamma} > 0.1 \text{ mb}$ is the photopion production cross section, and y is the average energy fraction (in the laboratory system) lost by a nucleon per interaction. Therefore, if ultra-high energy cosmic rays originate at cosmological distances, the net effect of their interactions would yield a pile-up of particles around $10^{19.6} \text{ eV}$ with the spectrum dropping sharply thereafter. This so-called GZK cutoff has been recently observed [78], suggesting that cosmic ray astronomy (if possible) would be limited to $10^{19} \lesssim E/\text{eV} \lesssim 10^{20}$ [79].

Cosmic accelerators are also cosmic beam dumps producing secondary photons and neutrino beams. Particles accelerated near black holes pass through intense radiation fields or dense clouds of gas leading to production of secondary photons and neutrinos that accompany the

primary cosmic ray beam. The target material, whether a gas or photons, is likely to be sufficiently tenuous so that the primary beam and the photon beam are only partially attenuated. Neutrinos propagate in straight lines and they can reach the Earth from far distance sources, but as we will discuss in the next class, these messengers need large detectors with sensitivity to weak interactions.

C. 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.

Neutrinos were made in staggering numbers at the time of the Big Bang. Like the CMB photons, the relic neutrinos now possess little kinetic energy due to expansion of the universe. There are expected to be at least 114 neutrinos per cubic centimeter, averaged over all space. There could be many more on Earth because of condensation of neutrinos, now moving slowly under the gravitational pull of our galaxy. As of now, we only have a lower limit on the total mass in this free-floating, ghostly gas of neutrinos, but even so it is roughly equivalent to the total mass of all the visible stars in the universe.

These relic neutrinos would be wonderful to observe, and much thought has gone into seeking ways to do so [80]. The problem is that the probability of neutrinos interacting within a detector decreases with the square of the neutrino's energy, for low energies. And even in the rare case when the neutrino does react in the detector the resulting signal is frustratingly miniscule. Nobody has been able to detect these lowest-energy neutrinos as of yet. Prospects are not good for a low-energy neutrino telescope, at least in the near future.

Next best are neutrinos from the nuclear burning of stars. Here we are more fortunate, as we have the Sun close-by, producing a huge flux of neutrinos, which has been detected by several experiments [81]. A forty year mystery persists in the deficit of about one half of the numbers of neutrinos detected compared to expectations, the so-called ‘‘Solar Neutrino Problem’’ [82]. This deficit is now thought probably to be due to neutrino oscillations.

A marvelous event occurred at 07:35:41 GMT on 23 February 1987, when two detectors in deep mines in the US (the IMB experiment [83]) and Japan (the

⁸ Neutron astronomy from nearby extragalactic sources may also be possible [76].

Kamiokande experiment [84]) recorded a total of 19 neutrino interactions over a span of 13 seconds. Two and a half hours later (but reported sooner) astronomers in the Southern Hemisphere saw the first Supernova to be visible with the unaided eye since the time of Kepler, 250 years ago, and this occurred in the Large Magellanic Clouds at a distance of some 50 kiloparsecs (roughly 150,000 light years). As we have seen, supernovae of the gravitational collapse type, occur when elderly stars run out of nuclear fusion energy and can no longer resist the force of gravity. The neutrinos wind up carrying off most of the in-fall energy, some 10% of the total mass-energy of the inner part of star of about $1.4M_\odot$ masses. Approximately 3×10^{53} ergs are released with about 10^{58} neutrinos over a few seconds. This is a staggering thousand times the solar energy released over its whole lifetime! The awesome visible fireworks consist of a mere one thousandth of the energy release in neutrinos. From this spectacular beginning to neutrino astronomy followed many deductions about the nature of neutrinos, such as limits on mass, charge, gravitational attraction, and magnetic moment.

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 machinery which accelerates cosmic rays will inevitably also produce neutrinos, guaranteeing that high energy neutrinos surely arrive to us from the cosmos. The burning question for would-be neutrino astronomers is, however, are there enough neutrinos to detect?

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. The lead project, called DUMAND was canceled in 1995 on account of slow progress and budget difficulties, but managed to make great headway in pioneering techniques, studying backgrounds, and exploring detector designs.

High energy neutrinos are detected by observing the Cherenkov radiation from secondary particles produced by neutrinos interacting inside large volumes of highly transparent ice or water, instrumented with a lattice of photomultiplier tubes. To visualize this technique, consider an instrumented cubic volume of side L . Assume, for simplicity, that the neutrino direction is perpendicular to a side of the cube (for a realistic detector the exact geometry has to be taken into account as well as the arrival directions of the neutrinos). To a first approximation, a neutrino incident on a side of area L^2 will be detected provided it interacts within the detector volume, i.e. within the instrumented distance L . That probability is

$$P = 1 - \exp(-L/\lambda_\nu) \simeq L/\lambda_\nu, \quad (104)$$

with $\lambda_\nu = (\rho N_A \sigma_\nu)^{-1}$. Here ρ is the density of the ice or water, N_A Avogadro's number and σ_ν the neutrino-nucleon cross section. A neutrino flux $\mathcal{F}$ (neutrinos per cm^2 per second) crossing a detector with cross sectional area A ($\simeq L^2$) facing the incident beam, will produce

$$\mathcal{N} = ATP\mathcal{F} \quad (105)$$

events after a time T . In practice, the quantities A , P and $\mathcal{F}$ depend on the neutrino energy and $\mathcal{N}$ is obtained by a convolution over neutrino energy above the detector threshold.⁹

The Antarctic Muon And Neutrino Detector Array (AMANDA), using a 1-mile-deep Antarctic ice area as a Cherenkov detector, has operated for more than 5 years. The AMANDA group has reported detection of upcoming atmospheric neutrinos produced in cosmic ray showers, a demonstration of feasibility [85]. IceCube, the successor experiment to AMANDA, would reach the sensitivity close to the neutrino flux anticipated to accompany the highest energy cosmic rays, dubbed the Waxman-Bahcall bound [86]. This telescope, which is currently being deployed near the Amundsen-Scott station, comprises a cubic-kilometer of ultra-clear ice about a mile below the South Pole surface, instrumented with long strings of sensitive photon detectors which record light produced when neutrinos interact in the ice [87]. Companion experiments in the deep Mediterranean are moving into construction phase.

With the Sun and supernova neutrino observations as proofs of concepts, next generation neutrino experiments will also scrutinize their data for new particle physics, from the signatures of dark matter to the evidence for superstring theory.

Summing up, it seems likely that real high energy neutrino astronomy with kilometer scale projects is only a few years away. Meanwhile, underground detectors wait patiently for the next galactic supernova. In the very long run, as has been the case with every venture into new parts of the electromagnetic spectrum, one can be sure that neutrino astronomy will teach us many new and unexpected wonders as we open a new window upon the universe [88].

D. Gravitational Waves

Ever since Isaac Newton in the XVII century, we have learned that gravity is a force that acts immediately on

⁹ The "effective" telescope area A is not strictly equal to the geometric cross section of the instrumented volume facing the incoming neutrino because even neutrinos interacting outside the instrumented volume may produce a sufficient amount of light inside the detector to be detected. In practice, A is therefore determined as a function of the incident neutrino direction by simulation of the full detector, including the trigger.

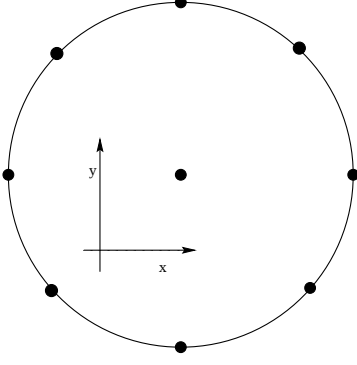


FIG. 18: Initial configuration of test particles on a circle of radius L before a gravitational wave hits them.

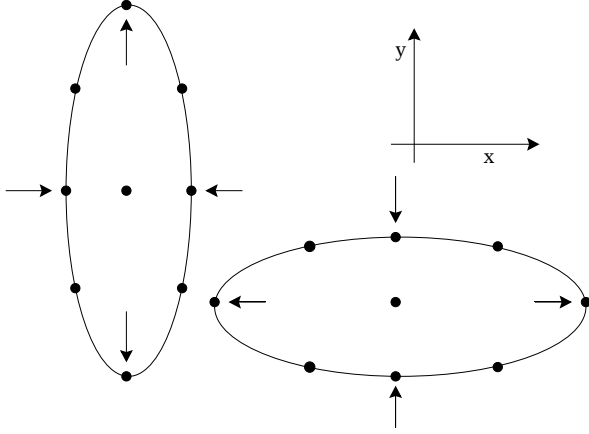


FIG. 19: The effect of a plus-polarized gravitational wave on a ring of particles. The amplitude shown in the figure is roughly $h = 0.5$. Gravitational waves passing through the Earth are many billion billion times weaker than this.

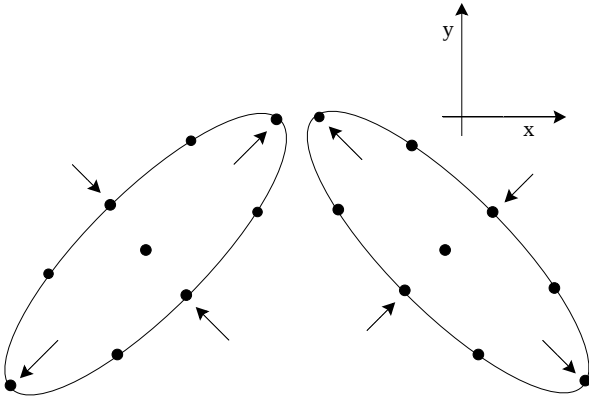


FIG. 20: The effect of cross-polarized gravitational waves on a ring of particles.

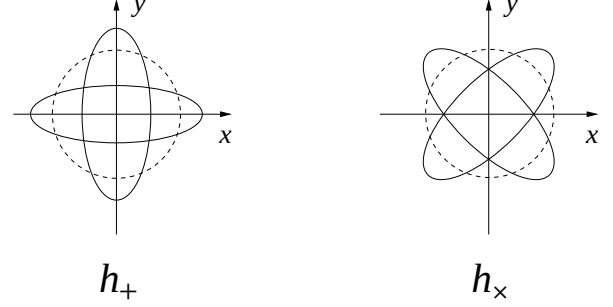


FIG. 21: Two linearly independent polarizations of a gravitational wave are illustrated by displaying their effect on a ring of free particles arrayed in a plane perpendicular to the direction of the wave. The figure shows the distortions in the original circle that the wave produces if it carries the 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.

an object. In Einstein theory of General Relativity, however, gravity is not a "force" at all, but a curvature in space [20]. 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. Propagating 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. 18. 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. 19 and 20. 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 ($h_\times$) 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. 21.

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. The power given off by a binary system of masses M_1 and M_2 separated a distance R is

$$P = -\frac{32}{\pi} \frac{G^4}{c^5} \frac{(M_1 M_2)^2 (M_1 + M_2)}{R^5}. \quad (106)$$

For the Earth-Sun system R is very large and M_1 and M_2 are relatively very small, yielding

$$P = -\frac{32}{\pi} \frac{(6.7 \times 10^{-11} \frac{\text{m}^3}{\text{kg s}^2})^4}{(3 \times 10^8 \text{ m/s})^5} \frac{(6 \times 10^{24} \text{ kg } 2 \times 10^{30} \text{ kg})^2 (6 \times 10^{24} \text{ kg} + 2 \times 10^{30} \text{ kg})}{(1.5 \times 10^{11} \text{ m})^5} = 313 \text{ W}. \quad (107)$$

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 $3.86 \times 10^{26} \text{ W}$. 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 $2.7 \times 10^{33} \text{ J}$. 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 Russell Hulse and Joe Taylor [89]. 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 car-

ried 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, $h \approx 10^{-20}$. 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) [90]. 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.

Even with such long arms, a gravitational wave will only change the distance between the ends of the arms by about 10^{-17} m at most. This is still only a fraction of the width of a proton. Nonetheless, LIGO's interferometers are now running routinely at an even better sensitivity level. As of November 2005, sensitivity had reached the primary design specification of a detectable strain of one part in 10^{21} over a 100 Hz bandwidth.

All detectors are limited at high frequencies by shot noise, which occurs because the laser beams are made up of photons. If there are not enough photons arriving in a given time interval (that is, if the laser is not intense enough), it will be impossible to tell whether a measurement is due to real data, or just random fluctuations in the number of photons.

All ground-based detectors are also limited at low frequencies by seismic noise, and must be very well isolated from seismic disturbances. Passing cars and trains, falling logs, and even waves crashing on the shore hundreds of miles away are all very significant sources of noise in real interferometers.

Space-based interferometers, such as LISA, are also being developed. LISA's design calls for test masses to be placed five million kilometers apart, in separate spacecrafts with lasers running between them, as shown in Fig. 13. The mission will use the three spacecrafts arranged in an equilateral triangle to form the arms of a giant Michelson interferometer with arms about 5 million kilometers long. As gravitational waves pass through the array, they slowly squeeze and stretch the space between the spacecrafts. Although LISA will not be affected by seismic noise, it will be affected by other noise sources, including noise from cosmic rays and solar wind and (of course) shot noise.

As we have seen in our look back through time, after the GUT era, the universe underwent a period of fantastic growth. This inflationary phase concluded with a violent conversion of energy into hot matter and radiation. This reheating process also resulted in a flood of gravitational waves. Because the universe is transparent to the propagation of gravitational waves, LISA will be able to search for these relic ripples probing energy scales far beyond the eV range, associated to the surface of last scattering [91].

XIX. QUANTUM BLACK HOLES

As we have seen, black holes are the evolutionary endpoints of massive stars that undergo a supernova explo-

sion 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 highest 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/m³, 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 $\sim 10^{-35}$ 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 [92]. 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.¹⁰ As the black holes radiate, they lose mass and

¹⁰ One can alternatively think of the emitted particles as coming from the singularity inside the black hole, tunneling out through

so will eventually evaporate completely and disappear. The evaporation is generally regarded as being thermal in character,¹¹ with a temperature inversely proportional to its mass M_{BH} ,

$$T_{\text{BH}} = \frac{1}{8\pi G M_{\text{BH}}} = \frac{1}{4\pi r_s}, \quad (108)$$

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 [95]. 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 [95]

$$A = 27\pi r_s^2. \quad (109)$$

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

$$\frac{d\dot{N}_i}{dQ} = \frac{\sigma_s}{8\pi^2} Q^2 \left[\exp\left(\frac{Q}{T_{\text{BH}}}\right) - (-1)^{2s} \right]^{-1} \quad (110)$$

per degree of particle freedom i . The change of variables $u = Q/T$, brings Eq. (110) into a more familiar form,

$$\dot{N}_i = \frac{27\Gamma_s T_{\text{BH}}}{128\pi^3} \int \frac{u^2}{e^u - (-1)^{2s}} du. \quad (111)$$

This expression can be easily integrated using

$$\int_0^\infty \frac{z^{n-1}}{e^z - 1} dz = \Gamma(n) \zeta(n) \quad (112)$$

and

$$\int_0^\infty \frac{z^{n-1}}{e^z + 1} dz = \frac{1}{2^n} (2^n - 2) \Gamma(n) \zeta(n), \quad (113)$$

yielding

$$\dot{N}_i = f \frac{27\Gamma_s}{128\pi^3} \Gamma(3) \zeta(3) T_{\text{BH}}, \quad (114)$$

where $\Gamma(x)$ ($\zeta(x)$) is the Gamma (Riemann zeta) function and $f = 1$ ($f = 3/4$) for bosons (fermions).¹² Therefore, the black hole emission rate is found to be

$$\dot{N}_i \approx 7.8 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (115)$$

$$\dot{N}_i \approx 3.8 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (116)$$

$$\dot{N}_i \approx 1.9 \times 10^{20} \left(\frac{T_{\text{BH}}}{\text{GeV}} \right) \text{ s}^{-1}, \quad (117)$$

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

$$\frac{\dot{M}_{\text{BH}}}{dQ} = - \sum_i c_i \frac{\sigma_s}{8\pi^2} \frac{Q^3}{e^{Q/T_{\text{BH}}} - (-1)^{2s}}, \quad (118)$$

where c_i is the number of internal degrees of freedom of particle species i . A straightforward calculation yields

$$\dot{M}_{\text{BH}} = - \sum_i c_i \tilde{f} \frac{27\Gamma_s}{128\pi^3} \Gamma(4) \zeta(4) T_{\text{BH}}^2, \quad (119)$$

where $\tilde{f} = 1$ ($\tilde{f} = 7/8$) for bosons (fermions). Assuming that the effective high energy theory contains approximately the same number of modes as the SM (i.e., $c_{s=1/2} = 90$, and $c_{s=1} = 27$), we find

$$\frac{dM_{\text{BH}}}{dt} = 8.3 \times 10^{73} \text{ GeV}^4 \frac{1}{M_{\text{BH}}^2}. \quad (120)$$

Ignoring thresholds, i.e., assuming that the mass of the black hole evolves according to Eq. (120) during the entire process of evaporation, we can obtain an estimate for the lifetime of the black hole,

$$\tau_{\text{BH}} = 1.2 \times 10^{-74} \text{ GeV}^{-4} \int M_{\text{BH}}^2 dM_{\text{BH}}. \quad (121)$$

¹¹ Indeed both the average number [92] and the probability distribution of the number [94] of outgoing particles in each mode obey a thermal spectrum.

¹² 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 $\forall s > 1$. Using these two definitions one can now verify Eq. (47).

Using $\hbar = 6.58 \times 10^{-25}$ GeV s, Eq. (121) can then be re-written as

$$\begin{aligned}\tau_{\text{BH}} &\simeq 2.6 \times 10^{-99} (M_{\text{BH}}/\text{GeV})^3 \text{ s} \\ &\simeq 1.6 \times 10^{-26} (M_{\text{BH}}/\text{kg})^3 \text{ yr} .\end{aligned}\quad (122)$$

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 [96]. 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 [97].

XX. HOMEWORKS

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

2. 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 (a) the apparent brightness I of the Sun, and (b) the absolute luminosity L of the Sun.

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

4. (a) In a forest there are n trees per hectare, evenly spaced. The thickness of each trunk is D . What is the distance of the wood not seen for the trees? (b) How is this related to the Olbers paradox?

5. According to special relativity, the Doppler shift is given by

$$\lambda' = \lambda \sqrt{(1 + V/c)/(1 - V/c)} , \quad (123)$$

where λ is the emitted wavelength as seen in a reference frame at rest with respect to the source, and λ' is the wavelength measured in a frame moving with velocity V away from the source along the line of sight. Show that the Doppler shift in wavelength is $z \approx V/c$ for $V \ll c$.

6. 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.]

7. Astronomers have recently measured the rotation of gas around what might be a supermassive black hole of about 2 billion solar masses at the center of the Galaxy. If the radius of the Galactic center to the gas cloud is 60 light-years, what Doppler shift $\Delta\lambda/\lambda$ do you estimate they saw?

8. Find the photon density of the 2.7 K background radiation.

9. What is the maximum sum-of-the-angles for a triangle on a sphere?

10. Estimate the age of the Universe using the Einstein-de Sitter approximation.

11. To explore the distribution of charge within nuclei, very-high-energy electrons are used. Electrons are used rather than protons because they do not partake in the strong nuclear force, so only the electric charge is investigated. Experiments at the Stanford linear accelerator were using electrons with 1.3 GeV of kinetic energy to obtain the charge distribution of the bismuth nucleus. What is the wavelength, and hence the expected resolution for such a beam of electrons.

12. In the rare decay $\pi^+ \rightarrow e^+ + \nu_e$, what is the kinetic energy of the positron? Assume that the π^+ decays from rest and that neutrinos are massless.

13. Show, by conserving momentum and energy, that it is impossible for an isolated electron to radiate only a single photon.

14. Are any of the following reactions/decays possible? For those forbidden, explain what conservation law is violated. (a) $\mu^+ \rightarrow e^+ + \bar{\nu}_\mu$, (b) $\pi^- p \rightarrow K^0 n$ (c) $\pi^+ p \rightarrow n\pi^0$, (d) $p \rightarrow n e^+ \nu_e$, (e) $\pi^+ p \rightarrow p e^+$, (f) $p \rightarrow e^+ \nu_e$, and (g) $\pi^- p \rightarrow K^0 \Lambda^0$.

15. (a) Symmetry breaking occurs in the electroweak theory at about 10^{-18} m. Show that this corresponds to an energy that is on the order of the mass of the $W^\pm$. (b) Show that the so-called unification distance of 10^{-32} m in a grand unified theory is equivalent to an energy of about 10^{16} GeV.

16. An experiment uses 3300 tons of water waiting to see a proton decay of the type $p \rightarrow \pi^0 e^+$. If the experiment is run for 4 yr without detecting decay, estimate a lower limit on the proton lifetime.

17. What is the temperature that corresponds to 1.8 TeV collisions at the Fermilab collider? To what era in cosmological history does this correspond?

18. Estimate the minimum energy of a cosmic ray proton to excite the Δ resonance scattering off CMB photons.

19. Evaluate the prospects for neutrino detection at the km^3 IceCube facility. Estimate the total number of ultra-high energy events ($10^8 \text{ GeV} < E_\nu < 10^{11} \text{ GeV}$) expected to be detected at IceCube during its lifetime of 15 years, if the cosmic neutrino flux is [86]

$$\mathcal{F}_\nu(E_\nu) \simeq 6.0 \times 10^{-8} (E_\nu/\text{GeV})^{-2} \text{ GeV}^{-1} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} , \quad (124)$$

and the neutrino-nucleon cross section rises with energy as [98]

$$\sigma(E_\nu) \simeq 6.04 (E_\nu/\text{GeV})^{0.358} \text{ pb} . \quad (125)$$

Assume that for this energy range the Earth becomes completely opaque to the propagation of neutrinos. [Hint: $1\text{b} = 10^{-28} \text{ m}^2$].

20. (a) 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 $R = 6GM_{\text{BH}}/c^2$. (b) If the kinetic energy of the neutron star orbiting the black hole is about $7 \times 10^{47} \text{ J}$, how much time will it take the neutron star to fall into the black hole?

21. Hawking has calculated quantum mechanically that a black hole will emit particles as if it were a hot body with temperature T proportional to its surface gravity. Since the surface gravity is inversely proportional to the black hole mass M , and the emitting area is proportional to M^2 , the luminosity or total power emitted is proportional to AT^4 , or equivalently $L \propto M^{-2}$. Show that as M decreases at this rate, the black hole lifetime will be proportional to M^3 . Estimate the lifetime of a black hole with a temperature $T = 10^{12} \text{ K}$ corresponding to $M = 10^{12} \text{ kg}$, i.e., about the mass of a mountain.

22. The spectrum of black holes scales like $1/E^{-3}$, just as the ultra high energy cosmic ray background is observed to do above the so-called “knee.” Using the upper bound on the primordial black hole density, estimate an upper limit on the black hole contribution to the high energy cosmic ray spectrum. Can primordial black holes be the sources of the observed cosmic rays?

Solutions:

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.(a) The apparent brightness is $I = 1.3 \times 10^3 \text{ W/m}^2$.

2.(b) The absolute luminosity is $L = 3.7 \times 10^{26} \text{ W}$.

3. The angular width is the inverse tangent of the diameter of our Galaxy divided by the distance to Andromeda,

$$\phi = \arctan \left[\frac{\text{Galaxy diameter}}{\text{Distance to Andromeda}} \right] \approx 2.4^\circ . \quad (126)$$

For the Moon we obtain,

$$\phi = \arctan \left[\frac{\text{Moon diameter}}{\text{Distance to Moon}} \right] \approx 0.52^\circ . \quad (127)$$

Therefore, the galaxy width is about 4.5 times the Moon width.

4.(a) 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

$$s(x + dx) = s(x) + [1 - s(x)] n D dx , \quad (128)$$

which gives a differential equation for s :

$$\frac{ds(x)}{dx} = [1 - s(x)] n D . \quad (129)$$

This is a separable equation which can be integrated:

$$\int_0^s \frac{ds}{1-s} = \int_0^x n D dx . \quad (130)$$

This yields the solution

$$s(x) = 1 - e^{-n D x} . \quad (131)$$

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. The corresponding probability density is its derivative ds/dx . The mean free path λ is the expectation of this distribution

$$\lambda = \int_0^\infty x \frac{ds(x)}{dx} dx = \frac{1}{n D} . \quad (132)$$

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.

4.(b) 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

$$s(x) = 1 - e^{-n A x} , \quad (133)$$

where $\lambda = (n A)^{-1}$. For example, if there were one sun per cubic parsec, the mean free path would be $1.6 \times 10^4 \text{ pc}$. 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.

5. To show that the Doppler shift in wavelength is $\Delta\lambda/\lambda \approx V/c$ for $v \ll c$ we use the binomial expansion:

$$\begin{aligned} \lambda' &= \lambda (1 + V/c)^{1/2} (1 - V/c)^{-1/2} \\ &\approx \lambda \left[1 + \frac{V}{2c} + \mathcal{O}\left(\frac{V^2}{c^2}\right) \right] \left[1 - \left(-\frac{V}{2c}\right) + \mathcal{O}\left(\frac{V^2}{c^2}\right) \right] \\ &\approx \lambda [1 + V/c + \mathcal{O}(V^2/c^2)] . \end{aligned} \quad (134)$$

6. 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

$$4\lambda = \lambda\sqrt{(1+V/c)/(1-V/c)} \Rightarrow V = 0.88 c. \quad (135)$$

7. We assume that gravity causes a centripetal acceleration on the gas. Using Newton's law,

$$F_{\text{gravity}} = G \frac{m_{\text{gas}} M_{\text{BH}}}{r^2} = \frac{m_{\text{gas}} V_{\text{gas}}^2}{r}, \quad (136)$$

we solve for the speed of the rotating gas, $V_{\text{gas}} = 6.84 \times 10^5$ m/s. The Doppler shift as compared to the light coming from the center of the Galaxy is then $\Delta\lambda/\lambda \approx V_{\text{gas}}/c \approx 2 \times 10^{-3}$.

8. Substituting the CMB temperature in Eq. (48) we obtain $n \approx 400$ photons cm^{-3} .

9. The limiting value for the angles in a triangle on a sphere is 540° . Imagine drawing an equilateral triangle near the north pole, enclosing the north pole. If that triangle were small, the surface would be approximate flat, and each angle on the triangle would be 60° . Then imagine "stretching" each side of the triangle down towards the equator, while keeping sure that the north pole stayed inside the triangle. The angle at each vertex of the triangle would expand, with a limiting value of 180° . The three 180° angles in the triangle would add up to 540° .

10. Setting $\Omega_0 = 1$ and $\Omega_\Lambda = 0$ in Eq. (73) we obtain a differential equation,

$$\dot{R}^2 = H_0^2 R_0^3 / R, \quad (137)$$

that can be easily integrated

$$R = (3H_0 t/2)^{2/3} R_0. \quad (138)$$

Then, for $R = R_0$ we obtain

$$t_0 = \frac{2}{3} \frac{1}{H_0} \sim 10 \text{ Gyr}. \quad (139)$$

11. The momentum of the electron is given by the relativistic identity $E^2 = p^2 c^2 + m^2 c^4$. Each electron has an energy of 1300 MeV, which is about 2500 times the mass of the electron $m_e \approx 0.51 \text{ MeV}/c^2$, thus we can ignore the second term and solve for $p = E/c$. Therefore the de Broglie wavelength is $\lambda = h/p = hc/E$,

$$\begin{aligned} \lambda &= \frac{6.63 \times 10^{-34} \text{ Js}}{1.3 \times 10^9 \text{ eV}} \frac{3 \times 10^8 \text{ m/s}}{1.6 \times 10^{-19} \text{ J/eV}} \\ &\simeq 0.96 \times 10^{-15} \text{ m}. \end{aligned} \quad (140)$$

This resolution, of approximately 1 fm, is on the order of the size of a nucleus.

12. We use conservation of energy,

$$m_{\pi^+} c^2 = E_{e^+} + E_\nu, \quad (141)$$

and momentum. Since the pion decays from rest, the momentum before the decay is zero. Thus, the momentum after the decay is also zero, and so the magnitude of momenta of the positron and the neutrino are equal, $p_{e^+} = p_\nu$. Therefore, $p_{e^+}^2 c^2 = p_\nu^2 c^2$. The positron and neutrino momenta are obtained from the relativistic identity, yielding $E_{e^+}^2 - m_{e^+}^2 c^4 = E_\nu^2$. (Recall neutrinos are taken as massless particles). From Eq. (141) we first find

$$E_{e^+}^2 - m_{e^+}^2 c^4 = (m_{\pi^+} c^2 - E_{e^+})^2 \quad (142)$$

and then solve for

$$E_{e^+} = \frac{1}{2} m_{\pi^+} c^2 + \frac{m_{e^+}^2 c^2}{2 m_{\pi^+}}. \quad (143)$$

Finally the total kinetic energy of the positron is

$$\begin{aligned} K_{e^+} &= \frac{m_{\pi^+} c^2}{2} - m_{e^+} c^2 + \frac{m_{e^+}^2 c^2}{2 m_{\pi^+}} \\ &\simeq \frac{139.6 \text{ MeV}}{2} - 0.511 \text{ MeV} + \frac{0.511 \text{ MeV}}{2 \times 139.6 \text{ MeV}} \\ &\simeq 69.3 \text{ MeV}. \end{aligned} \quad (144)$$

13. We work in the rest frame of the isolated electron, so that it is initially at rest. Energy conservation leads to

$$m_e c^2 = K_e + m_e c^2 + E_\gamma. \quad (145)$$

Hence, $K_e = -E_\gamma$ (i.e., $K_e = E_\gamma = 0$). Because the photon has no energy, it does not exist. Consequently, it has not been emitted.

14.(a) The reaction is forbidden because lepton number is not conserved.

14.(b) The reaction is forbidden because strangeness is not conserved.

14.(c) The reaction is forbidden because charge is not conserved.

14.(d) The reaction is forbidden because energy is not conserved.

14.(e) The reaction is forbidden because lepton number is not conserved.

14.(f) The reaction is forbidden because baryon number is not conserved.

14.(g) The reaction is possible via the strong interaction.

15.(a) We use Eq. (90) to estimate the mass of the particle based on the given distance,

$$mc^2 \approx \frac{hc}{2\pi d} = 1.98 \times 10^{11} \text{ eV} \approx 200 \text{ GeV}. \quad (146)$$

This value is of the same order of magnitude as the mass of the $W^\pm$.

15.(b) We take the de Broglie wavelength, $\lambda = h/p$, as the unification distance. The energy is so high that we assume $E = pc$, and thus $E_{\text{GUT}} = 3.1 \times 10^{15} \text{ GeV}$. A similar result ($E_{\text{GUT}} = 2 \times 10^{16} \text{ GeV}$) can be obtained using Eq. (90). Therefore, this energy is the amount that could be violated in conservation of energy if the universe were the size of the unification distance.

16. As with radioactive decay, the number of decays is proportional to the number of parent species (N), the time interval Δt , and the decay constant λ ,

$$\Delta N = -\lambda N \Delta t, \quad (147)$$

where the minus sign means that N is decreasing. The rate of decay is then $\Delta N/\Delta t = -\lambda N$, or equivalently,

$$\frac{dN}{dt} = -\lambda N \Rightarrow \ln \left(\frac{N}{N_0} \right) = e^{-\lambda t}, \quad (148)$$

where N_0 is the number of protons present at time $t = 0$ and N is the number remaining after a time t . Therefore, $N = N_0 e^{-\lambda t}$. The half-lifetime is defined as the time for which 1/2 of the original amount of protons in a given sample have decayed,

$$T_{1/2} = \frac{\ln 2}{\lambda} \Rightarrow \lambda = \frac{\ln 2}{T_{1/2}}. \quad (149)$$

Hence,

$$\Delta N = -\frac{\ln 2}{T_{1/2}} N \Delta t; \quad (150)$$

thus, for $\Delta N < 1$ over the four-year trial,

$$T_{1/2} > N \Delta t \ln 2. \quad (151)$$

To determine N , we note that each molecule of water H_2O contains 10 protons. So one mole of water (18 g, 6×10^{23} molecules) contains 6×10^{24} protons in 18 g of water, or about 3×10^{26} protons per kg. One ton is 10^3 kg , so the chamber contains

$$\begin{aligned} N &= 3.3 \times 10^6 \text{ kg } 3 \times 10^{26} \text{ protons/kg} \\ &\approx 10^{33} \text{ protons}. \end{aligned} \quad (152)$$

Then a very rough estimate for the lower limit on the proton half-life is $T_{1/2} > 10^{33} \text{ yr } 0.7 \approx 3 \times 10^{33} \text{ yr}$. It should be stressed that this naïve calculation does not take into account either the required study of background events or a statistical analysis of small

signals [99].

17. We can approximate the temperature from the kinetic energy relationship given in Eq. (92),

$$\begin{aligned} T &= \frac{2K}{3k} \\ &= \frac{2 \cdot 1.8 \times 10^{12} \text{ eV } 1.6 \times 10^{-19} \text{ J/eV}}{3 \cdot 1.38 \times 10^{-23} \text{ J/K}} \\ &\approx 1.4 \times 10^{16} \text{ K}. \end{aligned} \quad (153)$$

From Sec. XVI, we see that this temperature corresponds to the hadron era.

18. The average square center-of-mass energy in a $p\gamma_{\text{CMB}}$ collision can be expressed in the Lorentz invariant form (note that the cosine of the angle between the particles average away)

$$s = m_p^2 + 2 E_p E_{\gamma_{\text{CMB}}}, \quad (154)$$

where $E_{\gamma_{\text{CMB}}} \sim 10^{-3} \text{ eV}$ is the typical CMB photon energy and E_p is the energy of the incoming proton. The Δ^+ decays according to

$$s = m_p^2 + m_\pi^2 + 2m_\pi m_p. \quad (155)$$

Thus, from Eqs. (154) and (155), we obtain $E_{p\gamma_{\text{CMB}}}^{\text{th}} \approx 10^{20} \text{ eV}$.

19. The total number of target nucleons present in one km^3 of ice is: $N_T = \rho_{\text{ice}} N_A V \sim 6 \times 10^{38}$, where $\rho_{\text{ice}} \sim 1 \text{ g/cm}^3$ is the density of ice, $N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$ is Avogadro's number, and $V = 10^{15} \text{ cm}^3$ is the total volume. Assuming the Earth is completely opaque to the propagation of neutrinos (i.e., an effective aperture for detection at IceCube of $2\pi \text{ sr}$) the total number of events expected in $T = 15 \text{ yr}$ is

$$\begin{aligned} \mathcal{N} &= 2\pi N_T T \int_{10^8 \text{ GeV}}^{10^{11} \text{ GeV}} \mathcal{F}(E_\nu) \sigma(E_\nu) dE_\nu \\ &\approx 7. \end{aligned} \quad (156)$$

20. Substituting in Eq. (106) we obtain $P = 2.4 \times 10^{47} \text{ W}$. At this emission rate the neutron star will fall into the black hole in $t \approx 2.9 \text{ s}$.

21. Substituting in Eq. (122) we obtain $\tau_{\text{BH}} \sim 1.5 \times 10^{10} \text{ yr}$, about the present age of the universe.

22. The upper bound on the primordial black hole density yields an upper limit on the black hole contribution to the high energy cosmic ray background as observed on Earth given by

$$dJ/dE \simeq [10^{16} \text{ eV}^2/\text{m}^2 \text{ s sr}]/(E \text{ eV})^3. \quad (157)$$

The details of this tedious calculation have been published in [100]. Comparing to the observed background shown in Fig. 17,

$$dJ/dE \simeq [10^{24.8} \text{ eV}^2/\text{m}^2 \text{ s sr}]/(E \text{ eV})^3, \quad (158)$$

we find that the black signal is down by about 9 orders of magnitude.

-
- [1] E. Hubble, *The Realm of Nebulae*, (Yale University Press, New Haven, 1936; reprinted by Dover Publications, Inc., New York, 1958).
 - [2] A rigorous discussion of all the topics discussed in this course can be found in S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972). For a comprehensive book on introductory physics, which covers the essential prerequisite reading material to follow the course, see e.g., D. C. Giancoli, *Physics* (6th Edition, Pearson Prentice Hall, NJ, 2005).
 - [3] 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).
 - [4] E. Hertzsprung, *Astron. Nachr.* **196**, 201 (1913); H. N. Russell, *Science* **37**, 651 (1913).
 - [5] H. A. Bethe, *Phys. Rev.* **55**, 434 (1939).
 - [6] S. Chandrasekhar, *Mon. Not. Roy. Astron. Soc.* **95**, 207 (1935).
 - [7] J. R. Oppenheimer and G. M. Volkoff, *Phys. Rev.* **55**, 374 (1939).
 - [8] J. R. Oppenheimer and H. Snyder, *Phys. Rev.* **56**, 455 (1939).
 - [9] R. Penrose, *Phys. Rev. Lett.* **14**, 57 (1965); S. Hawking, *Phys. Rev. Lett.* **15**, 689 (1965); S. Hawking, *Proc. Roy. Soc. Lond. A* **294**, 511 (1966); S. Hawking, *Proc. Roy. Soc. Lond. A* **295**, 490 (1966); S. Hawking, *Proc. Roy. Soc. Lond. A* **300**, 187 (1967); S. W. Hawking and R. Penrose, *Proc. Roy. Soc. Lond. A* **314**, 529 (1970).
 - [10] A. G. W. Cameron, *Nature* **229**, 178 (1971); R. E. Wilson, *Astrophys. J.* **170**, 529 (1971).
 - [11] 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.
 - [12] H. W. M. Olbers, *Bode's Jahrbuch*, 111 (1826); reprinted by Dickson, *op. cit.*, Appendix I.
 - [13] A. Einstein, *Annalen Phys.* **17**, 891 (1905) [*Annalen Phys.* **14**, 194 (2005)].
 - [14] E. Hubble, *Proc. Nat. Acad. Sci.* **15**, 168 (1927).
 - [15] A. A. Penzias and R. W. Wilson, *Astrophys. J.* **142**, 419 (1965).
 - [16] G. Gamow, *Phys. Rev.* **70**, 572 (1946); R. A. Alpher, H. Bethe and G. Gamow, *Phys. Rev.* **73**, 803 (1948); G. Gamow, *Rev. Mod. Phys.* **21**, 367 (1949).
 - [17] G. Hinshaw *et al.* [WMAP Collaboration], arXiv:astro-ph/0603451.
 - [18] G. F. Smoot *et al.*, *Astrophys. J.* **396**, L1 (1992); C. L. Bennett *et al.*, *Astrophys. J.* **464**, L1 (1996) [arXiv:astro-ph/9601067].
 - [19] D. N. Spergel *et al.* [WMAP Collaboration], *Astrophys. J. Suppl.* **148**, 175 (2003) [arXiv:astro-ph/0302209]; D. N. Spergel *et al.* [WMAP Collaboration], arXiv:astro-ph/0603449.
 - [20] A. Einstein, *Annalen Phys.* **49**, 769 (1916) [*Annalen Phys.* **14**, 517 (2005)].
 - [21] A. Einstein, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1917**, 142 (1917).
 - [22] A. Friedmann, *Z. Phys.* **10**, 377 (1922); A. Friedmann, *Z. Phys.* **21**, 326 (1924).
 - [23] H. P. Robertson, *Astrophys. J.* **82**, 284 (1935); H. P. Robertson, *Astrophys. J.* **83**, 187, 257 (1936).
 - [24] A. G. Walker, *Proc. Lond. Math. Soc.* (2), **42** 90 (1936).
 - [25] A. G. Riess *et al.* [Supernova Search Team Collaboration], *Astron. J.* **116**, 1009 (1998) [arXiv:astro-ph/9805201]; S. Perlmutter *et al.* [Supernova Cosmology Project Collaboration], *Astrophys. J.* **517**, 565 (1999) [arXiv:astro-ph/9812133].
 - [26] B. Ratra and P. J. E. Peebles, *Phys. Rev. D* **37**, 3406 (1988); I. Zlatev, L. M. Wang and P. J. Steinhardt, *Phys. Rev. Lett.* **82**, 896 (1999) [arXiv:astro-ph/9807002].
 - [27] R. A. Knop *et al.* [Supernova Cosmology Project Collaboration], *Astrophys. J.* **598**, 102 (2003) [arXiv:astro-ph/0309368].
 - [28] S. W. Allen, R. W. Schmidt and A. C. Fabian, *Mon. Not. Roy. Astron. Soc.* **334**, L11 (2002) [arXiv:astro-ph/0205007].
 - [29] A. E. Lange *et al.* [Boomerang Collaboration], *Phys. Rev. D* **63**, 042001 (2001) [arXiv:astro-ph/0005004]; A. Balbi *et al.*, *Astrophys. J.* **545**, L1 (2000) [Erratum-ibid. **558**, L145 (2001)] [arXiv:astro-ph/0005124].
 - [30] S. M. Carroll, W. H. Press and E. L. Turner, *Ann. Rev. Astron. Astrophys.* **30**, 499 (1992).
 - [31] B. S. Meyer and D. N. Schramm, *Astrophys. J.* **311**, 406 (1986).
 - [32] G. Aldering *et al.* [SNAP Collaboration], arXiv:astro-ph/0209550.
 - [33] K. Schwarzschild, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916**, 424 (1916) [arXiv:physics/9912033].
 - [34] For further details see e.g., C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, (W. H. Freeman & Co., San Francisco, 1973).
 - [35] J. J. Thomson, *Philosophical Magazine Fifth Series* **44**, 293 (1897).
 - [36] E. Rutherford, *Philosophical Magazine Sixth Series* **37**, 537 (1919).
 - [37] C. D. Anderson, *Science* **76**, 238 (1932); C. D. Anderson, *Phys. Rev.* **43**, 491 (1933).
 - [38] P. A. M. Dirac, *Proc. Roy. Soc. Lond. A* **126**, 360 (1930); P. A. M. Dirac, *Proc. Roy. Soc. Lond. A* **133**, 60 (1931).
 - [39] H. Yukawa, *Proc. Phys. Math. Soc. Jap.* **17**, 48 (1935).
 - [40] J. C. Street and E. C. Stevenson, *Phys. Rev.* **52**, 1003 (1937).
 - [41] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini and C. F. Powell, *Nature* **159**, 694 (1947); C. M. G. Lattes, G. P. S. Occhialini and C. F. Powell, *Nature* **160**,

- 453 (1947); C. M. G. Lattes, G. P. S. Occhialini and C. F. Powell, *Nature* **160**, 486 (1947).
- [42] E. Gardner and C. Lattes, *Science* **107**, 270 (1948); W. Barkas, E. Gardner, and C. Lattes, *Phys. Rev.* **74**, 1558 (1948); J. Richardson, *Phys. Rev.* **74**, 1720 (1948); H. York, B. Moyer and R. Bjorklund, *Phys. Rev.* **76**, 187 (1949); J. Steinberger, W. Panofsky and J. Steller, *Phys. Rev.* **78**, 802 (1950).
- [43] G. Arnison *et al.* [UA1 Collaboration], *Phys. Lett. B* **122**, 103 (1983). M. Banner *et al.* [UA2 Collaboration], *Phys. Lett. B* **122**, 476 (1983); G. Arnison *et al.* [UA1 Collaboration], *Phys. Lett. B* **126**, 398 (1983); P. Bagnaia *et al.* [UA2 Collaboration], *Phys. Lett. B* **129**, 130 (1983); G. Arnison *et al.* [UA1 Collaboration], *Phys. Lett. B* **129**, 273 (1983).
- [44] See e.g., Y. Fukuda *et al.* [Super-Kamiokande Collaboration], *Phys. Rev. Lett.* **81**, 1562 (1998) [arXiv:hep-ex/9807003]; Y. Fukuda *et al.* [Super-Kamiokande Collaboration], *Phys. Rev. Lett.* **82**, 1810 (1999) [arXiv:hep-ex/9812009]; S. Fukuda *et al.* [Super-Kamiokande Collaboration], *Phys. Lett. B* **539**, 179 (2002) [arXiv:hep-ex/0205075]; M. H. Ahn *et al.* [K2K Collaboration], *Phys. Rev. Lett.* **90**, 041801 (2003) [arXiv:hep-ex/0212007]; T. Araki *et al.* [KamLAND Collaboration], *Phys. Rev. Lett.* **94**, 081801 (2005) [arXiv:hep-ex/0406035].
- [45] H. L. Anderson, E. Fermi, E. A. Long and D. E. Nagle, *Phys. Rev.* **85**, 936 (1952).
- [46] M. Gell-Mann, *Phys. Lett.* **8**, 214 (1964).
- [47] G. Zweig, CERN-TH/1964-401 and CERN-TH/1964-412, unpublished.
- [48] D. J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973); H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973).
- [49] S. L. Glashow, *Nucl. Phys.* **22**, 579 (1961).
- [50] A. Salam, in *Elementary Particle Physics*, N. Svartholm, ed. (Nobel Symposium No. 8, Almqvist & Wiksell, Stockholm, 1968), p.367.
- [51] S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967).
- [52] B. Abbott *et al.* [D0 Collaboration], *Phys. Rev. Lett.* **84**, 222 (2000) [arXiv:hep-ex/9909030].
- [53] V. M. Abazov *et al.* [D0 Collaboration], *Nature* **429**, 638 (2004) [arXiv:hep-ex/0406031].
- [54] For further details see e.g., F. Halzen and A. D. Martin, *Quarks And Leptons: An Introductory Course In Modern Particle Physics*, (New York, Usa: Wiley, 1984).
- [55] See e.g., V. M. Abazov *et al.* [D0 Collaboration], *Phys. Rev. Lett.* **97**, 151804 (2006) [arXiv:hep-ex/0607032]; A. Abulencia *et al.* [CDF Collaboration], *Phys. Rev. Lett.* **97**, 081802 (2006) [arXiv:hep-ex/0605124]; V. M. Abazov *et al.* [D0 Collaboration], *Phys. Rev. Lett.* **97**, 161803 (2006) [arXiv:hep-ex/0607022]; S. Schael *et al.* [ALEPH Collaboration], *Eur. Phys. J. C* **47**, 547 (2006) [arXiv:hep-ex/0602042].
- [56] H. Georgi and S. L. Glashow, *Phys. Rev. Lett.* **32**, 438 (1974).
- [57] S. Dimopoulos and H. Georgi, *Nucl. Phys. B* **193**, 150 (1981).
- [58] S. Dimopoulos, S. Raby and F. Wilczek, *Phys. Rev. D* **24**, 1681 (1981).
- [59] J. Wess and B. Zumino, *Nucl. Phys. B* **70**, 39 (1974); J. Wess and B. Zumino, *Phys. Lett. B* **49**, 52 (1974).
- [60] M. Shiozawa *et al.* [Super-Kamiokande Collaboration], *Phys. Rev. Lett.* **81**, 3319 (1998) [arXiv:hep-ex/9806014].
- [61] M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory Vol. 1: Introduction*, (Cambridge University Press, 1987); M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory Vol. 2: Loop Amplitudes, Anomalies And Phenomenology*, (Cambridge University Press 1987).
- [62] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, *Phys. Lett. B* **429**, 263 (1998) [arXiv:hep-ph/9803315]; I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, *Phys. Lett. B* **436**, 257 (1998) [arXiv:hep-ph/9804398]. A similar framework was introduced in L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 3370 (1999) [arXiv:hep-ph/9905221].
- [63] M. Acciarri *et al.* [L3 Collaboration], *Phys. Lett. B* **470**, 281 (1999) [arXiv:hep-ex/9910056]; B. Abbott *et al.* [D0 Collaboration], *Phys. Rev. Lett.* **86**, 1156 (2001) [arXiv:hep-ex/0008065]; C. D. Hoyle, U. Schmidt, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, D. J. Kapner and H. E. Swanson, *Phys. Rev. Lett.* **86**, 1418 (2001) [arXiv:hep-ph/0011014]; L. A. Anchordoqui, J. L. Feng, H. Goldberg and A. D. Shapere, *Phys. Rev. D* **65**, 124027 (2002) [arXiv:hep-ph/0112247]; A. Abulencia *et al.* [CDF Collaboration], *Phys. Rev. Lett.* **97**, 171802 (2006) [arXiv:hep-ex/0605101].
- [64] L. Anchordoqui, H. Goldberg, S. Nawata and C. Nunez, arXiv:0704.0928 [hep-ph].
- [65] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
- [66] For further details see e.g., E. W. Kolb and M. S. Turner, *The Early Universe*, *Front. Phys.* **69**, 1 (1990).
- [67] H. Goldberg, *Phys. Rev. Lett.* **50**, 1419 (1983); J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive and M. Srednicki, *Nucl. Phys. B* **238**, 453 (1984).
- [68] For further details see e.g., G. Bertone, D. Hooper and J. Silk, *Phys. Rept.* **405**, 279 (2005) [arXiv:hep-ph/0404175].
- [69] R. C. Hartman *et al.* [EGRET Collaboration], *Astrophys. J. Suppl.* **123** (1999) 79.
- [70] 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).
- [71] J. G. Learned and K. Mannheim, *Ann. Rev. Nucl. Part. Sci.* **50**, 679 (2000).
- [72] V. F. Hess, *Phys. Z.* **13**, 1804 (1912).
- [73] P. Auger, R. Maze, T. Grivet-Meyer, *Comptes Rendus* **206**, 1721 (1938); P. Auger, P. Ehrenfest, R. Maze, J. Daudin, Robley, and A. Fréon, *Rev. Mod. Phys.* **11**, 288 (1939).
- [74] D. J. Bird *et al.*, *Astrophys. J.* **441**, 144 (1995).
- [75] D. J. Bird *et al.* [HRES Collaboration], *Phys. Rev. Lett.* **71**, 3401 (1993).
- [76] L. A. Anchordoqui, H. Goldberg and T. J. Weiler, *Phys. Rev. Lett.* **87**, 081101 (2001) [arXiv:astro-ph/0103043].
- [77] K. Greisen, *Phys. Rev. Lett.* **16**, 748 (1966); G. T. Zatsepin and V. A. Kuzmin, *JETP Lett.* **4**, 78 (1966) [*Pisma Zh. Eksp. Teor. Fiz.* **4**, 114 (1966)].
- [78] R. Abbasi *et al.* [HiRes Collaboration], arXiv:astro-ph/0703099.
- [79] For further details see e.g., L. Anchordoqui, T. Paul, S. Reucroft and J. Swain, *Int. J. Mod. Phys. A* **18**, 2229 (2003) [arXiv:hep-ph/0206072].
- [80] T. J. Weiler, *Phys. Rev. Lett.* **49**, 234 (1982).
- [81] P. Anselmann *et al.* [GALLEX Collaboration], *Phys.*

- Lett. B **285**, 376 (1992); P. Anselmann *et al.* [GALLEX Collaboration], Phys. Lett. B **342**, 440 (1995); Y. Fukuda *et al.* [Kamiokande Collaboration], Phys. Rev. Lett. **77**, 1683 (1996); Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **81**, 1158 (1998) [Erratum-ibid. **81**, 4279 (1998)] [arXiv:hep-ex/9805021]; Y. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **82**, 2430 (1999) [arXiv:hep-ex/9812011]; S. Fukuda *et al.* [Super-Kamiokande Collaboration], Phys. Rev. Lett. **86**, 5651 (2001) [arXiv:hep-ex/0103032].
- [82] R. J. Davis, D. S. Harmer and K. C. Hoffman, Phys. Rev. Lett. **20**, 1205 (1968); J. N. Bahcall, N. A. Bahcall and G. Shaviv, Phys. Rev. Lett. **20**, 1209 (1968).
- [83] R. M. Bionta *et al.*, Phys. Rev. Lett. **58**, 1494 (1987).
- [84] K. Hirata *et al.* [KAMIOKANDE-II Collaboration], Phys. Rev. Lett. **58**, 1490 (1987);
- [85] E. Andres *et al.*, Nature **410**, 441 (2001).
- [86] E. Waxman and J. N. Bahcall, Phys. Rev. D **59**, 023002 (1999) [arXiv:hep-ph/9807282].
- [87] J. Ahrens *et al.* [IceCube Collaboration], Nucl. Phys. Proc. Suppl. **118**, 388 (2003) [arXiv:astro-ph/0209556].
- [88] For further details see e.g., F. Halzen, Phys. Rept. **333**, 349 (2000); L. Anchordoqui and F. Halzen, Annals Phys. **321**, 2660 (2006) [arXiv:hep-ph/0510389].
- [89] R. A. Hulse and J. H. Taylor, Astrophys. J. **195**, L51 (1975).
- [90] A. Abramovici *et al.*, Science **256**, 325 (1992).
- [91] For further details see e.g., K. S. Thorne, Rev. Mod. Phys. **52**, 285 (1980). K. S. Thorne, Rev. Mod. Phys. **52**, 299 (1980).
- [92] S. W. Hawking, Nature **248**, 30 (1974); S. W. Hawking, Commun. Math. Phys. **43** (1975) 199.
- [93] J. B. Hartle and S. W. Hawking, Phys. Rev. D **13** (1976) 2188.
- [94] L. Parker, Phys. Rev. D **12**, 1519 (1975); R. M. Wald, Commun. Math. Phys. **45**, 9 (1975); S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
- [95] D. N. Page, Phys. Rev. D **13** (1976) 198.
- [96] D. N. Page and S. W. Hawking, Astrophys. J. **206**, 1 (1976).
- [97] S. W. Hawking, Phys. Rev. D **13**, 191 (1976).
- [98] L. A. Anchordoqui, A. M. Cooper-Sarkar, D. Hooper and S. Sarkar, Phys. Rev. D **74**, 043008 (2006) [arXiv:hep-ph/0605086].
- [99] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998) [arXiv:physics/9711021].
- [100] J. Oliensis and C. T. Hill, Phys. Lett. B **143**, 92 (1984).